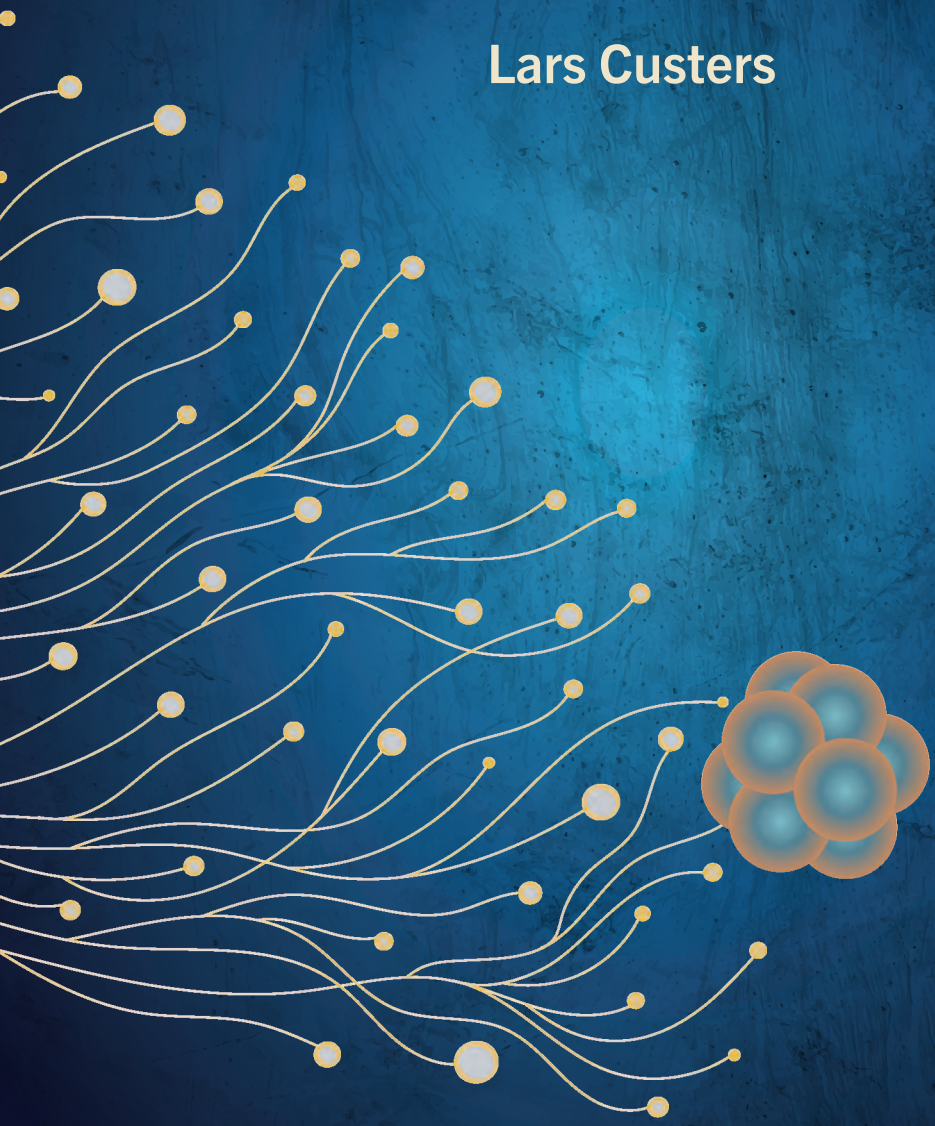
From Derailed Development to Childhood Cancer

Using organoid models to study Malignant Rhabdoid Tumors

Lars Custers



From Derailed Development to Childhood Cancer

Using Organoid Models to study Malignant Rhabdoid Tumors

Lars Custers

ISBN	987-90-393-7404-7
Copyright	© Lars Custers, all rights reserved.
Cover illustration	Designed by Lars Custers, adapted from vectors downloaded from vecteezy.com and designed by Just Call Me Acar and vector.satrian.
Layout	Lars Custers, aided by Elise van Reen.
Printing	R&R GROUP grafische dienstverleners
	info@rrgroup.nl

From Derailed Development to Childhood Cancer

Using Organoid Models to Study Malignant Rhabdoid Tumors

Van ontspoorde ontwikkeling tot kinderkanker

Een onderzoek naar rhabdoïde tumoren met organoïde modellen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 15 oktober 2021 des middags te 12.15 uur

door

Lars Martinus Catharina Custers

geboren op 28 augustus 1991
te Venlo

Promotor Prof. dr. J.C. Clevers

Copromotor Dr. J. Drost

"Freedom is the freedom to say that two plus two make four.
If that is granted, all else follows."

George Orwell

Examination committee

Prof. dr. M. Maurice

Prof. dr. F.C.P. Holstege

Prof. dr. E.M. Hol

Prof. dr. L.H.J. Looijenga

Prof. dr. E.P.J.G. Cuppen

Contents

CHAPTER 1

General introduction and outline of the thesis

Lars Custers[1,2]

[1]Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands

[2]Oncode Institute, the Netherlands

Pediatric cancer

Pediatric cancer has seen a significant increase in patient survival rates over the last decades, promoted by advances in biomedical research and improved therapeutic strategies[1]. Nonetheless, pediatric cancer is still the leading disease-related cause of death in children[2], and for a subset of childhood cancer types therapeutic innovation has stalled. Further research efforts are required to accelerate the development of effective treatment options and ultimately find a cure for all types of childhood cancer.

Childhood tumors are relatively rare, with an incidence of 1 in ~7000 children per year[3] (~550 patients in the Netherlands). They encompass a heterogeneous spectrum of cancer types including central nervous system tumors, retinoblastoma, renal tumors, germ cell tumors, leukemias, lymphomas, neuroblastomas, sarcomas of bone and soft tissue, rhabdoid tumors and other rare cancers[4]. Most of the tumors that arise in children never or rarely occur in adults. They are characterized by distinct genetic features and unique underlying oncogenic mechanisms[4,5]. Insights gained from adult malignancies therefore cannot be directly translated to childhood cancer biology.

Recent profiling of pediatric cancer genomes has provided a comprehensive overview of the genetic drivers underpinning malignant transformation[4,6]. However, the clinical implications and underlying molecular mechanisms often remain elusive. Moreover, subtypes can be classified based on epigenetic profiles[7,8], increasing the complexity even further. A combination of the low incidence of pediatric cancers and the substantial genetic and epigenetic heterogeneity implies that, for instance in the Netherlands, of each tumor subtype only a small number is diagnosed each year, which complicates both clinical and pre-clinical studies. International collaborations are thus crucial to overcome these limitations.

Childhood cancers are fundamentally different than adult tumors. Genomic instability and an accumulation of genetic alterations, induced by mutagen exposure and/or aging, is thought to be essential for adult tumorigenesis[9]. In contrast, most pediatric cancer genomes harbor relatively few somatic mutations and only involve a single or a few genetic driver events that are often unique to childhood malignancies[4,6]. A subset of childhood cancers is indicated to initiate in prenatal life and are therefore commonly regarded as products of aberrant embryonic development[10,11]. A recurrent characteristic of these so-called embryonal tumors is a block in cellular maturation that retains the cells in a malignant embryonic and proliferative state[12]. This embryonic profile is absent in adult cells and may therefore serve as a specific vulnerability of childhood cancer cells. Altogether, pediatric cancer is distinct from adult malignancies in many perspectives[10-12]. Accordingly, to further improve pediatric patient survival, treatment strategies are required that specifically target the unique oncogenic properties of childhood cancer cells.

The genetic simplicity that characterizes childhood tumors should not be confused with the complexity of underlying molecular mechanisms. Tumorigenesis is highly

dependent on the cellular context, meaning that the oncogenic effect of a genetic lesion is often restricted to a specific cell type and developmental time[13]. As embryonic development is highly dynamic and involves continuously changing cellular identities, it can be challenging to identify the cell-of-origin and define the hijacked developmental programs that drive pediatric cancer. The developmental pathways directing embryogenesis and cell state transitions are for a major part controlled by epigenetic gene regulation. Interestingly, the genetic drivers of childhood cancer are enriched for alterations affecting transcription factors or epigenetic modifiers[10]. Epigenetic dysregulation is therefore thought to be at the root of many pediatric malignancies. This is particularly apparent for malignant rhabdoid tumors (MRTs), which is the main focus of the research presented in this Thesis. MRT is driven by a single recurrent genetic aberration, bi-allelic loss of *SMARCB1* or in rare cases *SMARCA4*[7,8,14-16]. Both are subunits of the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodeler complex[17]. MRT is therefore a compelling example of how a single genetic event during embryonic development can block cellular maturation and induce malignant transformation by epigenetic mechanisms.

Malignant rhabdoid tumors

MRT is an aggressive and highly metastatic malignancy that has been particularly challenging to treat. Survival rates (15-50%) are lagging behind compared to other pediatric tumors (e.g. Wilms tumor: >90%)[18-22]. The majority of patients are exposed to intense multimodal therapy that involves a combination of surgery, high-dose chemo- and radiotherapy[18,20]. Despite the harsh treatment strategy, most tumors become therapy resistant. Accordingly, MRT remains one of the most lethal pediatric cancers. The patients that do survive encounter sever treatment-related late effects, including cardiovascular disease, second malignancies, psychological disorders, and many more[23]. Hence, MRT patients are in desperate need of effective targeted therapy.

MRT incidence peaks in early childhood (<3 years)[18-20]. Its occurrence is anatomically heterogenous, as it can arise in numerous sites in the body. MRTs predominantly localize to the central nervous system (66%[24]) wherein they are referred to as Atypical Teratoid Rhabdoid Tumors (AT/RT). Outside the central nervous system, MRTs are classified as extracranial rhabdoid tumors (ECRT), arising most frequently in the kidney, but also in other soft tissues such as the liver and the head and neck[18]. As MRT can be both a brain and a solid tumor, it crosses clinical disciplines, which has hampered the study of MRT as a whole. MRT is considered as a rare pediatric cancer, as prevalence is relatively low across childhood tumors (0.9-2% of renal cancers; 1.5-2.1% of brain cancers), but substantial (~15%) among infant cancers (diagnosed <1 year)[18-20].

Historically, MRT was diagnosed by immunohistochemical analysis of tumor tissue, assessing cell morphology and immunostainings for markers such as vimentin and cytokeratin[25]. Characteristic for MRT cells is their rhabdoid morphology,

which is featured by large eccentric nuclei with prominent nucleoli and a cytoplasm with eosinophilic inclusions[25-27]. However, a diagnosis based on these immunohistochemical criteria was far from conclusive as rhabdoid features can also be observed in other central nervous system tumors[24,28]. Therefore, in the past, MRT were often misdiagnosed, or were classified as a subtype of another tumor entity (e.g. Wilms tumor[29]). Only after the rise of systematic next-generation sequencing of cancer genomes, a uniform and distinctive feature of MRT was identified[15,16,30], which is the bi-allelic loss of the tumor suppressor genes *SMARCB1* (>95%) or *SMARCA4* (<5%). Accordingly, INI-1 (protein product encoded by *SMARCB1*) and BRG1 (encoded by *SMARCA4*) immunostaining is now performed routinely to confirm MRT diagnosis[31].

MRT genetics and epigenetic subclassification

MRT is characterized as genetically stable, as the mutational burden is relatively low compared to other (adult) malignancies[16]. Unique to MRT is that it harbors a single recurrent genetic driver event. Comprehensive analyses of MRT genomes have not identified additional recurrent mutations other than those found in *SMARCB1* or *SMARCA4*[14,32,33]. The spectrum of genetic lesions affecting the *SMARCB1* gene includes loss-of-function mutations as well as large and focal deletions of chromosome arm 22q[7,16]. As for many tumor suppressor genes, the second *SMARCB1* allele is often affected by copy-number neutral loss of heterozygosity[7,16]. Approximately one-third of MRT patients carry a germline *SMARCB1* mutation, which is acquired *de novo* for the majority of cases[34]. This condition is referred to as Rhabdoid Tumor Predisposition Syndrome, which predisposes to the development of often more than one primary tumor and implies an even poorer prognosis[35].

Although genetically uniform, MRTs are clinically heterogeneous. A small subset of MRT patients responds favorably to therapy. However, the biological basis for this clinical heterogeneity is not firmly established. Transcriptomic analyses of AT/RT tissues allowed the identification of subgroups, wherein the expression of subgroup markers *BMP4* or *ASCL1* were suggested to associate with patient survival[36,37]. However, the small size of the studied tumor cohorts is a limitation that makes the clinical use of these markers questionable. Nonetheless, these attempts set the starting point for a more comprehensive molecular characterization of MRTs using larger cohorts.

As MRTs are genetically rather homogeneous, tumor heterogeneity was hypothesized to establish by epigenetic mechanisms. Extensive studies of DNA methylation patterns demonstrated that AT/RTs can be classified into molecular subgroups with distinctive epigenetic profiles[8,38]. Within the AT/RT entity, three subgroups could be identified, referred to as MYC, sonic hedgehog (SHH) and tyrosinase (TYR)[8]. DNA methylation profiles were found to correlate with tumor transcriptomes, which enabled the identification of subgroup specific signaling pathways. The AT/RT-SHH subgroup was defined by an active SHH and NOTCH signaling pathway, as genes such as *MYCN, GLI2* and *ASCL1* were highly expressed. The AT/RT-MYC subgroup was

characterized by high expression of the *MYC* oncogene and a cluster of *HOX* genes, whereas the AT/RT-TYR subgroup showed upregulation of melanosomal markers *MITF*, *TYR* and *DCT*[8]. Similarly, a cohort of ECRTs was classified into two groups according to microRNA, gene expression and DNA methylation profiles[32]. Analysis of a combined cohort of AT/RT and ECRT samples showed that DNA methylation profiles of ECRTs and AT/RT-MYC were largely shared and distinct from AT/RT-SHH and -TYR subgroups, indicating a close molecular relationship of central nervous system and extracranial MRTs[7] (**Table 1**). Interestingly, the type of *SMARCB1* mutations were found to be differently distributed between MRT subgroups, which could after all indicate a possible genetic contribution to MRT heterogeneity[7,8]. Furthermore, recent investigations demonstrated that the rare *SMARCA4*-mutant MRT are molecularly distinct from *SMARCB1*-deficient MRT and should therefore be regarded as a separate subgroup[39].

DNA methylation is known to have a significant impact on transcription, as it is associated with repressive chromatin that silences the expression of genes[40]. Subgroup-specific DNA methylation patterns are therefore thought to result in the divergent transcriptional profiles that define MRT identity. MRTs were found to exhibit aberrant levels of global DNA methylation[7,8], which is a feature that has been recognized to significantly contribute to oncogenesis for other tumors[40]. Interestingly, AT/RT-MYC and ECRT subgroups exhibited decreased methylation, whereas AT/RT-SHH and -TYR genomes were strongly hypermethylated compared to other pediatric brain tumors and healthy brain tissues[7]. The increased methylation levels mostly concentrated on gene promoter regions, including the promoters of potential tumor suppressors. As expected, mRNA levels of these genes were low, which suggests an epigenetic mechanism of tumor suppressor silencing in MRT. Other methylation elements known to influence gene expression are so-called partially methylated domains (PMD), which exhibit disordered methylation patterns that also associate with repressive chromatin[41]. PMDs are highly variable between tissues and a prominent feature in tumor methylomes[42,43]. The number of PMDs strongly varied between subgroups, in which the hypomethylated AT/RT-MYC subgroup showed the highest number of PMDs[7,8]. MRT heterogeneity is suggested to be at least partly mediated by the abundance and distribution of hypermethylated regions and PMDs[7,8].

On the other hand, chromatin regions with low DNA methylation levels, referred to as DNA methylation valleys, are associated with active chromatin and were shown to correlate with the acetylation of histone 3 at lysine 27 (H3K27ac), a marker of active genomic regulatory regions, in MRT tissues[7,8]. Comparison of active regulatory regions between patient tissues enabled the identification of subgroup-specific enhancer elements. Enhancers are distal gene regulatory regions that are primary determinants of lineage and cell type specification. A subclass of enhancers referred to as super-enhancers, was linked to the overexpression of subgroup-specific signature genes such as *GLI2*, *MITF* and the *HOX*-cluster[7,8]. Aberrant activation of super-enhancers can significantly contribute to oncogenesis, as has been shown for other tumor

Table 1 | Overview of defining features of MRT subgroups.
Adapted from Chun et al[7].

	ECRT	AT/RT		
Subgroup	Renal, extra-renal	MYC	SHH	TYR
Gene signature	Mesenchymal-like		Neural-like	
Enhancers	HOXC, MYC		GLI2, PTCH2	TYR, MITF
Global DNA methylation	Hypomethylation many PMDs		Hypermethylation No/few PMDs	
SMARCB1 mutations	Focal deletions	Broad deletions	Non-sense SNVs	Focal deletions
Cytotoxic T-cell infiltration	++++	+++++	+	++

entities[44], which could possibly also apply to MRTs. The distinct enhancer landscapes and correlating transcriptomes of each MRT subgroup could be matched to programs of embryonic development[7,8,38]. AT/RT-SHH and -TYR resembled a neural profile, whereas AT/RT-MYC and ECRT displayed mesenchymal features[7].

Altogether, *SMARCB1* loss is the genetic driver for each MRT subgroup, but the diverging chromatin landscape of MRTs suggests that the precise mechanism underlying MRT development may be subgroup specific. However, these underlying mechanisms remain largely elusive.

The SWI/SNF complex and gene regulation

In the nucleus of a cell, genomic DNA does not exist as a 'naked' molecule but is organized by association with DNA binding proteins to condense into a complex structure referred to as chromatin. The fundamental unit of chromatin is the nucleosome, which consists of DNA that is tightly wrapped around an octamer of histone proteins. Adjacent nucleosomes are coiled into high-order chromatin structures to form the DNA fibers that make up a chromosome. The tight association of DNA and histones prevents the binding of gene regulatory proteins such as transcription factors, which consequently inhibits the transcription of genes[45]. The compact structure that makes DNA inaccessible is referred to as heterochromatin. To activate gene expression, nucleosomes require to be actively mobilized to increase chromatin accessibility and expose the DNA to the transcriptional machinery required for mRNA production. Nucleosome occupancy is dynamically regulated by a cellular process referred to as chromatin remodeling[46,47]. The complex process of chromatin remodeling involves an interplay of epigenetic modifiers, transcription factors and chromatin remodeler complexes to shape chromatin topology and subsequently activate or inactive gene expression. The chromatin machinery can specifically modulate the accessibility of gene promoters and enhancers, and in such a way control the transcriptional profile that determines cell phenotype. Altogether, chromatin remodelers have been shown

to play a central role in gene regulation, which is essential for the proper function of all cellular processes.

One of the key chromatin remodelers is the SWI/SNF complex, which regulates chromatin accessibility by an ATP-dependent eviction of nucleosomes[46]. Recruitment of SWI/SNF at gene promoters and enhancers mostly correlates with transcription activation[17]. The mammalian SWI/SNF complexes, also known as BAF complexes, exist in many forms, as subunit composition is diverse and variable[46,48,49]. Broadly, BAF complexes can be subdivided into three categories based on subunit composition: canonical BAF (cBAF), polybromo-associated BAF (pBAF) and non-canonical BAF (ncBAF)[48]. Within the different subfamilies many subtle variants can occur. The catalytic subunit with ATPase activity that is essential in all three subfamilies is either SMARCA4 or SMARCA2, which are mutually exclusive. Other core subunits of cBAF include SMARCB1, SMARCC1, SMARCC2, and SMARCD1. Furthermore, an increasing number of proteins have been identified as BAF subunits, which can be integrated into the complex in a variety of combinations. How subunit composition precisely affects the core chromatin remodeling function of BAF is not entirely clear. However, distinct genomic binding patterns can be observed for the different BAF subfamilies[48-51]. Therefore, BAF complex diversification is hypothesized to determine binding specificity, allowing for a complex-specific localization at the chromatin.

Furthermore, BAF complex composition is dependent on cellular context, as the assembly of BAF subunits was found to differ across lineages and cell types. Subunit diversification is most apparent during embryonic development[52]. The divergent transcriptional programs required for lineage-specification are indicated to be regulated by distinct BAF subtypes. For instance, embryonic stem cells (ESCs) harbor a BAF subtype (esBAF) that is essential for maintaining expression of pluripotency and self-renewal genes[53]. When *in vitro*-cultured ESCs are subjected to differentiation protocols by changing the medium composition, BAF subunit composition changes, which repositions the complex to genes driving differentiation. This process is well studied in neural development, in which the BAF complex recombines upon transition from ESCs to neural progenitor cells to upregulate genes involved in neural lineage commitment[52,54,55]. Subunit composition changes again upon terminal neuronal differentiation to induce cell cycle exit[52,54,55]. Overall, the BAF chromatin remodeler has evolved to gain the extensive complexity that is required for tissue specific gene regulation.

Aberrant chromatin remodeling in MRT development

Not surprisingly, aberrant function of the SWI/SNF complex has frequently been linked to disease. Nearly 25% of all human cancers harbor SWI/SNF subunit mutations[56]. The role of chromatin remodeling in tumor development is most clearly exemplified in MRT, as the sole genetic cause of MRT is loss of *SMARCB1*. To better understand how

a mutated chromatin remodeler can cause malignant transformation, knowledge on SMARCB1 function as a subunit of the SWI/SNF complex is crucial.

SMARCB1 was demonstrated to be essential in maintaining complex integrity of cBAF[49,51,57]. SMARCB1 loss abrogated the assembly of cBAF complexes, but on the contrary enhanced formation of a residual chromatin remodeler complex identified as ncBAF. The SMARCB1-independent ncBAF also occurs in non-malignant cells. In ESCs, ncBAF was found to be essential for maintaining pluripotency[50], but was functionally different from esBAF[53]. ncBAF complexes localized to distinct genomic regions, which were enriched for CTCF binding sites and topologically associating domain (TADs) boundaries, suggesting a role for ncBAF in chromatin architecture[50]. Similar binding patterns were observed in MRT cells, in which the residual ncBAF complex was found to be specifically maintained at super-enhancers close to genes driving processes involved in self-renewal[49,51,57].

These findings propose that MRT tumorigenesis is driven by aberrant binding of a residual ncBAF complex to oncogenic super-enhancers (**Fig. 1**). In addition, SMARCB1 loss prevents the recruitment of a functional cBAF complex, which in normal cells would bind to so-called typical enhancers that are required for differentiation[51,58]. Consequently, MRT cells are thought to be locked in an embryonic and proliferative state that is likely responsible for tumorigenesis. Further investigations are required to uncover the specific developmental pathways deregulated by defective SWI/SNF chromatin remodeling.

Current therapeutic approaches for MRT

Understanding the contribution of SMARCB1 to SWI/SNF complex assembly and its function in gene regulation has provided valuable insights into potential therapeutic vulnerabilities. For instance, the putative oncogenic role of the SMARCB1-independent ncBAF, that is enhanced in MRT, suggests that MRT may be sensitive to inhibition of residual SWI/SNF chromatin remodeling function. Indeed, knock-down or pharmacological inhibition of SMARCA4 in MRT cells was shown to induce a proliferation arrest and differentiation phenotype[59,60], suggesting that MRT growth is dependent on ncBAF. Recent work by Wang *et al.* and Michel *et al.* focused on identifying ncBAF-specific subunits that could potentially serve as a vulnerability in MRT[49,57]. Analysis of BAF complex perturbations showed that BRD9 is synthetically lethal with SMARCB1 loss, which proposes BRD9 as potential therapeutic target in MRT[49,57,61].

The transcriptional programs driving embryonic development are controlled by a fine balance between the SWI/SNF complex and other epigenetic regulators[62]. When this balance is impaired because of SMARCB1 loss, the epigenetic regulators that cooperate with the SWI/SNF complex may become a therapeutic vulnerability, as exemplified for BRD9[49,57]. Following that line of thought, pharmacological (re)activation of the genomic regulatory elements that are silenced because of SMARCB1 loss may

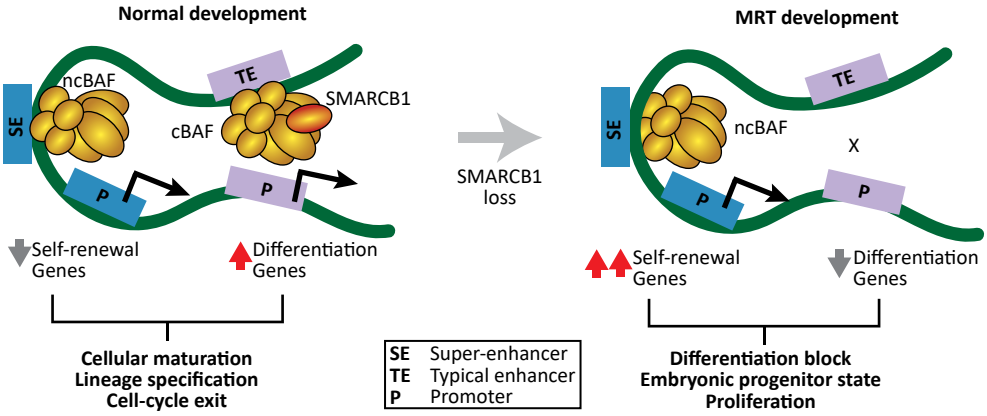


Figure 1 | Effects of SMARCB1 loss on gene regulation.
Adapted from Wang et al[51].

induce a bypass of the maturation block underlying MRT. For instance, inhibition of transcriptional repressors such as histone deacetylases (HDACs) or the Polycomb Repressive Complex 2 (PRC2) catalytic subunit EZH2, was found to be sufficient to de-repress differentiation genes and induce MRT cell maturation[63-66]. Targeting epigenetic modifiers in MRT therefore appears to be a promising avenue to counter the malignant effects of defective SWI/SNF chromatin remodeling and restore the epigenetic balance required for proper gene regulation.

Uncovering the oncogenic signaling pathways that are activated downstream of *SMARCB1* loss is key to design novel targeted therapy for MRT. Reconstitution of *SMARCB1* in MRT cell-lines has shown to induce a cell cycle arrest and differentiation phenotype[67]. Investigation of SMARCB1 targets revealed that MRT development and proliferation is dependent on the p16-CyclinD1/CDK4/CDK6-Rb axis[68-71]. Accordingly, AURKA and CDK4/6 have been proposed as potential therapeutic targets[72,73]. MRT signaling pathways can further be characterized by high activity of receptor tyrosine kinases (RTKs). Activating mutations for RTKs are absent in MRT. However, multiple RTKs were found to be overexpressed or aberrantly phosphorylated[38,74-76]. Genetic perturbation and drug screens in MRT cell-lines have identified multiple RTKs as MRT-specific vulnerabilities[75,77]. RTK inhibitors have shown to effectively reduce MRT growth *in vitro*, including drugs targeting EGFR, IGFR, VEGFR, MET, PDGFR and FGFR[74,75,77-80]. However, a limitation is that the efficacy of each RTK inhibitor was found to be highly heterogeneous across MRT cell-lines[75].

MRT subgroups are thought to be driven by distinct oncogenic signaling pathways, which is likely to translate to differential drug sensitivities. To define such subgroup-specific vulnerabilities, Torchia *et al.* interrogated MRT transcriptomes for differentially expressed genes, which predicted a subgroup-specific effect of BMP, NOTCH and PDGFRB inhibitors[38]. These findings indicate that although MRT are grouped as one

tumor entity, treatment strategies may need to be adapted for each subgroup. Both AT/RT-MYC and ECRT subgroups are characterized by high expression of the *MYC* oncogene, which is indicated to be an essential driver of tumor growth[81]. Up to date, it has been challenging to inhibit MYC pharmacologically because of its undruggable protein structure[82]. Nonetheless, alternatives that indirectly target MYC pathways have been extensively investigated to device effective therapies against MYC-driven tumors. Such therapies include inhibitors targeting the transcriptional machinery (e.g. BRD4 and CDK9) or protein metabolism, which have shown efficacy in MRT cell-lines[81,83-87]. The pathways driving tumor growth for the AT/RT-SHH and -TYR subgroups are less obvious, but reports have indicated that targets such as GLI1 and MITF are druggable with potential efficacy in MRT cells[8,88].

Despite indications of MRT-vulnerabilities in pre-clinical studies, there are no therapeutics yet that have shown significant clinical efficacy. Extensive target discovery and validation is therefore still highly required to provide better therapeutic options and ultimately improve the chances of survival for MRT patients. Furthermore, MRT subtype-specific vulnerabilities remain unclear, possibly due to limited functional interrogation of the putative oncogenic pathways driving growth in the distinct MRT subgroups. Therefore, deciphering the SMARCB1-dependent signaling pathways across different MRT subgroups is crucial for the identification of patient-specific therapeutic targets.

MRT models

Genetically engineered mouse models

Modelling tumorigenesis in mice has significantly contributed to our understanding of cancer origin and development. Over the years, multiple genetically engineered mouse models (GEMMs) have been established in which the *Smarcb1* gene was genetically modified to induce complete loss of protein (**Fig. 2**)[89-96], mimicking MRT genetics. However, MRT GEMMs have been variably successful. Loss of *Smarcb1* was found to be embryonically lethal when induced in the early phases of embryogenesis[90,92,95], which complicated modelling attempts. Crucial insights into MRT origin came from a GEMM wherein *Smarcb1* inactivation could be temporally controlled, which allowed for assessment of *Smarcb1* function during the sequential stages of embryogenesis[91]. The GEMM demonstrated that timing of *Smarcb1* loss determines tumor phenotype. Late induction in adult mice lead to lymphoma formation, whereas early *Smarcb1* inactivation between E6-E10 generated central nervous system MRTs with full penetrance, strongly indicating that MRT initiation is limited to a specific window of embryogenesis. Furthermore, the neural crest lineage was found to be particularly susceptible to MRT formation. Specific *Smarcb1* inactivation in early neural crest generated central nervous system MRTs that resembled the different patient AT/RT subgroups[96].

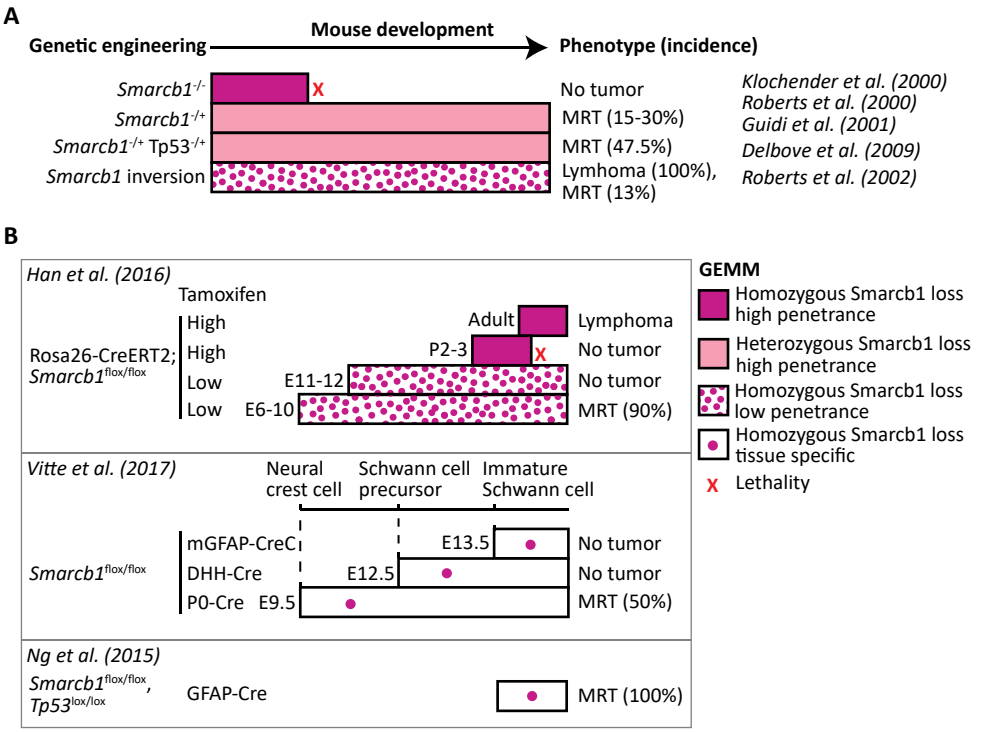


Figure 2 | Schematic representation of *Smarcb1* loss in genetically engineered mouse models and corresponding tumor phenotypes.
Inactivation of *Smarcb1* in germline **(A)** or conditional **(B)**. Adapted from Vitte et al[96].

Altogether, developmental timing and cell-type were shown to be critical determinants of MRT development (**Fig. 2**). The origin of MRT was pinpointed to the early embryonic development of the neural crest lineage. However, for these GEMMs, MRT heterogeneity could only partially be modeled in mice, as extracranial tumors were poorly represented[91,96]. For instance, rhabdoid tumors of the kidney or liver were never observed. Thus, the exact cell-of-origin of ECRTs remains an open question.

SMARCB1 loss in human embryonic stem cells

To study the effect of *SMARCB1* loss in a human developmental context, representative *in vitro* models of embryonic development are required. Embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs) have proven representative models of embryonic development[97]. They can be subjected to differentiation protocols that recapitulate the processes of embryonic differentiation and can generate cell-types across germ layers. ESCs have previously been used to model pediatric tumorigenesis[98]. Particularly for MRT, for which the cell-of-origin is not firmly established, ESCs were a suitable model, as the tumorigenic effect of *SMARCB1* loss can be functionally

interrogated throughout the sequential stages of ESC differentiation. Accordingly, Langer *et al.* demonstrated that a *SMARCB1* knockdown specifically impaired differentiation of the neural lineage, whereas endodermal and mesodermal cells were unaffected, strengthening the hypothesis that MRT arises during neural development[99]. Furthermore, Terada *et al.* showed that *SMARCB1* loss in iPSCs induced MRT-like tumors upon xenotransplantation in mice[100]. To be noted, these modelling attempts were predominantly performed in a *TP53*-mutant background[100], which is not a valid representation of patient MRT genetics and therefore not the most optimal model. Altogether, the specific cellular context required for MRT development was recapitulated in human *in vitro* models, which further highlighted the essential role of *SMARCB1* in neural differentiation.

Patient-derived MRT models

Progress in our understanding of tumor biology has been largely dependent on pre-clinical models. Patient-derived models of MRT include cancer cell-lines, which have significantly contributed to novel molecular insights on MRT over the years. Cell-lines can be expanded and manipulated with relative ease, which allows for efficient and scalable experiments. However, the initial establishment of cell-lines is an inefficient process[101]. Therefore, for rare tumors such as MRT, it has been challenging to generate a representative collection of *in vitro* models that fully covers tumor heterogeneity. An additional limitation is that many cancer cell-lines fail to faithfully represent the cellular biology and heterogeneity of the tumor tissue they were derived from, which complicates the translation of pre-clinical findings to effective treatment options[102].

Advancements in 3D culture technology have significantly improved pre-clinical cancer models. A culture system that has rapidly evolved over the last decade are the 3D stem cell-based models referred to as organoids[103]. Organoids are established by seeding cells extracted from embryonic or adult tissues into a 3D matrix. Culture conditions are customized to recapitulate the micro-environment of the *in vivo* stem cell niche. By providing a suitable growth factor mix, the stem cells can self-renew and generate the different cell-types that make up an organ. Organoids can self-organize into polarized 3D structures that recapitulate *in vivo* cell-cell interactions, and therefore serve as a representative model of organ homeostasis[103].

Following the development of organoid models for normal cells, the technology was also applied to malignant tissues, which has provided organoid models for an increasing number of tumor entities[104]. In contrast to cancer cell-lines, tumor organoids have been shown to phenotypically and genetically resemble the original tissue and better maintain cellular heterogeneity[104]. In general, the establishment of organoids from tumor material is relatively efficient, which facilitates the generation of large collections of patient-derived tumor models, often termed organoid biobanks, that better capture tumor heterogeneity. Furthermore, tumor organoids potentially have improved predictive power for therapy response, which has been demonstrated for multiple tumor entities over the years[105-109]. Drug efficacy can be tested on organoids

for cohorts of individual patients representing a variety of tumor subgroups, to better understand patient- and subgroup-specific drug vulnerabilities. Additionally, drug toxicity can be assessed by testing the compounds on organoids derived from normal tissue of the same individual, which is crucial to determine treatment specificity.

Up to date, patient-derived tumor organoid models for pediatric cancer are still scarce but show great promise. Calandrini *et al.* established an organoid biobank that represents the most prevalent pediatric renal tumor subtypes[110]. They were the first to establish organoids from MRT tissue, which were demonstrated to representatively match the primary tissue on a morphological, (epi)genetic and transcriptomic level[110]. Therefore, to identify potential therapeutic targets with high confidence, MRT organoids appear as suitable models for screening approaches (e.g. high-throughput drug or CRISPR-Cas9 knock-out screens). Furthermore, MRT organoids can be exploited to better understand the downstream effects of *SMARCB1* loss and how SWI/SNF chromatin remodeling function is altered in MRT.

Tracing the origin of childhood cancer

MRTs have shown to retain features of embryonic cells and are therefore thought to arise as a consequence of aberrant embryogenesis[100,111,112]. The maintenance of an embryonic profile indicates that MRT cells are halted in development, which is a common theme across embryonal tumors[11]. Investigations into tumor origin are essential to increase our understanding of the processes driving embryonal tumorigenesis. A direct comparison of normal and tumor development can reveal the malignant embryonic pathways responsible for tumor growth and can therefore aid the design of targeted treatment.

Defining tumor cell identity by transcriptomic comparison

Clues about the origin and development of embryonal tumors can be derived from tumor transcriptomes, as tumor cells may partially retain the transcriptional profile of its embryonic origin[112-115]. Recent advances in single cell genomics have led to a rapid expansion of cell atlases of embryonic development[116]. This enabled the mapping of tumor cells to developmental trajectories, which can define tumor differentiation states and in such a way provide indications from which cell the tumor possibly arises from[112-115]. Analysis of single-cell transcriptomes of tumor tissues allows for a precise quantitative readout of cancer cell state. However, large patient cohorts of single cell datasets are usually unavailable up to date and therefore tumor to normal comparisons are often performed using bulk transcriptomes. A limitation is that bulk mRNA-seq averages the various cell-type specific gene expression patterns that are present in the tumor. Nonetheless, recent advancements in bioinformatics led to so-called deconvolution methods that computationally assess cell proportions in bulk tumor data using a single cell reference and in such a way more precisely define tumor identity[117]. Application of such deconvolution methods to pediatric tumor transcriptomes demonstrated that embryonal tumors generally recapitulate

transcriptional programs unique to fetal cells[112-115,118], supporting the hypothesis of an embryonic maturation block.

To assess the cellular identity of pediatric brain tumors, Jessa *et al.* recently generated a detailed map of mouse brain development by extensive single-cell mRNA sequencing of fetal brain tissue from different developmental time-points (E10-P14)[112]. Fetal brain signals could be quantified for a variety of childhood brain tumors by deconvolution of bulk mRNA profiles. For instance, a projection of medulloblastoma transcriptomes onto the different lineages of the brain could stratify medulloblastoma subgroups by developmental state. The correlation of medulloblastoma subgroups to distinct lineages suggests that medulloblastoma may originate from different neural progenitor cells. Similarly, AT/RT transcriptomes of different subgroups were assessed, which revealed that AT/RT cells do not specifically map to any cell of the developing brain. However, when using a mouse reference dataset of early development (E6.5-E8.5) representing the gastrulation phase, AT/RT gene signatures could be matched to a broad range of early mesodermal and neuroectodermal cells. In agreement with analyses by Johann *et al.*[8], the developmental identity of AT/RTs was found to be subgroup specific.

Overall, it has been challenging to pinpoint the origin of MRT by transcriptional analyses. Results have been ambiguous in the sense that similarities were found with a broad range of early neural and mesenchymal cells. A popular hypothesis is that MRTs arise from different cells-of-origin, which would explain the large diversity of tumor locations and the heterogeneous tumor phenotypes. Alternatively, MRTs could arise from a migrating multipotent cell that can acquire both neural and mesenchymal identities.

Genetic lineage tracing as a tool to reveal patient tumor origin

Although MRT *in vitro* and *in vivo* modelling approaches have established the tumor suppressor function of *SMARCB1*, direct evidence of patient tumor origin can only be obtained through a technique referred to as genetic lineage tracing[119]. Genetic lineage tracing is a method that can track cell populations in retrospect using acquired somatic mutations. From conception onwards, for every cell division, a small number of somatic mutations accumulate in each cells genome which are passed on to its progeny[119]. Therefore, if two cells share a close common ancestor, they would inherit the somatic genome of the same ancestor cell, which would consequently be largely overlapping[119]. A group of cells that is the progeny of the same ancestor cell is referred to as a clonal expansion. On the other hand, if two cells share a distant common ancestor, the multitude of cell divisions that have occurred independently would generate somatic mutations that are unique to each cell. By exploiting this concept, somatic genomes can be analyzed to build a tree that shows the relative phylogenetic distance of cells or tissues.

Genetic lineage tracing applied to embryonic development allows for a construction of developmental lineages[120], wherein early embryonic mutations are shared for a substantial part of the embryo and late embryonic mutations are unique to specific lineages, organs or tissues (**Fig. 3a**). Accordingly, cells from different germ layers can be distinguished by somatic genome as they separated early in development[119]. Embryonal tumors are thought to arise from a single embryonic cell. Therefore, the somatic genome of a tumor can be compared to normal tissues from the same individual to interrogate which tissue the tumor is most closely related to, based on the shared somatic mutations (**Fig. 3b**). Following that rationale, whole-genome sequencing of tumor and patient-matching normal material and subsequent phylogenetic analysis can reveal the tissue-of-origin[119,121].

Genetic lineage tracing of pediatric tumors is limited by tissue availability, as accessibility to both tumor and normal material from the same individual is rare. Therefore, experiments are often performed on tissues extracted postmortem. Alternatively, tissues can be retrieved through surgical tumor resection, in which tumor but also surrounding normal tissues are removed, which is a standard procedure for pediatric renal tumors. This allowed Coorens *et al.* to investigate the origin of Wilms tumors, which was found to have initiated in the fetal kidney parenchyma[121]. As the root of MRT is not firmly established, similar strategies could be applied to evidently reveal the cell-of-origin of patient MRT.

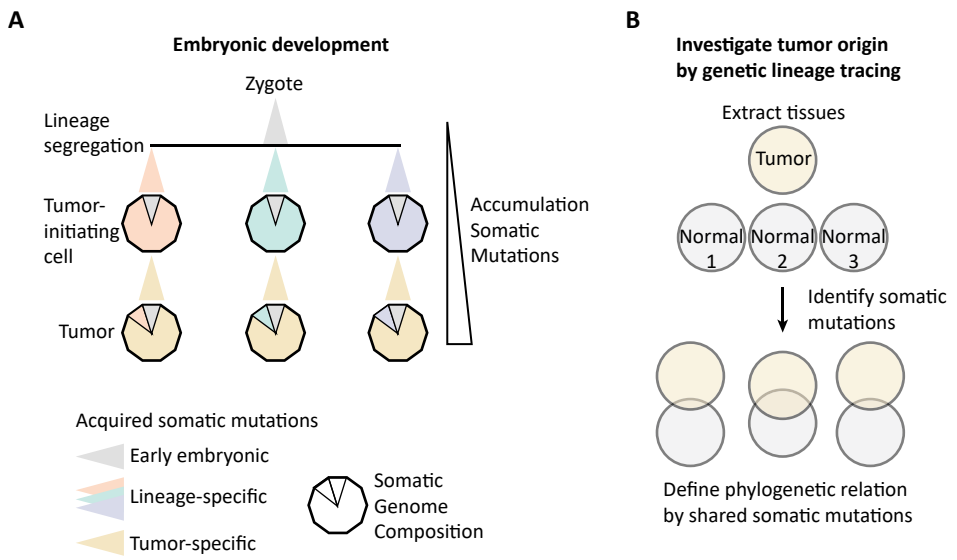


Figure 3 | Genetic lineage tracing as a tool to reveal patient tumor origin.
(A) Somatic mutations acquired during embryonic development act as natural genetic barcodes.
(B) Tumor origin can be identified by phylogenetic analysis of somatic mutations shared with normal tissues.

Thesis outline

Treatment outcome for MRT patients remains poor up to date. Therapeutic strategies still rely on chemotherapeutics, to which most MRTs quickly develop resistance. Improvement may come from therapy that is specifically designed to target the unique oncogenic properties of MRT cells, which are still largely unexplored. Investigating the downstream consequences of *SMARCB1* loss may identify the signaling pathways responsible for malignant growth, that can potentially serve as novel therapeutic targets. Furthermore, an accurate comparison of MRT to embryonic development may reveal the developmental pathways that are 'hijacked' in MRT, which may be exploited for treatment. A crucial first step is to define the origin of MRT, which has proven challenging over the years. It is still not firmly established in which developmental context MRT arises.

In this Thesis, we aimed to define the origin of MRT, investigate the SMARCB1-dependent differentiation pathways underpinning MRT development, and translate our findings to potential therapeutic targets. This General introduction (**Chapter 1**) is followed by a Review (**Chapter 2**[122]) that outlines different *in vitro* modelling approaches used for studying embryonal tumorigenesis. In **Chapter 3**[123], we investigated the cellular identity of adult and pediatric renal tumors, by mapping bulk tumor transcriptomes to single cell references of normal adult and fetal cells. We demonstrate that the cellular signals derived from MRT transcriptomes comprise a mix of early mesoderm and neural crest, which may point towards the cell MRT is derived from. In parallel, we sought out to investigate MRT origin by alternative approaches (**Chapter 4**[124]). We established phylogenetic relations of tumor and patient-matching normal tissues by comparing shared somatic mutations, which provided proof that MRT is closely related to the neural crest. Furthermore, we examined MRT differentiation trajectories by reverting *SMARCB1* loss in MRT organoids, which suggested that MRT cells are blocked along a differentiation path from neural crest to mesenchyme. Quantitative transcriptional predictions indicated mTOR and HDAC inhibition to mirror *SMARCB1* reconstitution, which we further validated experimentally. In **Chapter 5**, we examined changes in chromatin accessibility and interactions induced by *SMARCB1* reconstitution in MRT organoids, which uncovered *SMARCB1* to be essential for the regulation of transcriptional enhancers. In **Chapter 6**, we report the results of CRISPR-Cas9 knock-out screens in MRT organoids, which revealed MRT-specific vulnerabilities. Our findings demonstrate that an unbiased genetic screening approach using more physiological cancer organoid models can provide novel insights into tumor dependencies. And finally, we place our findings in a broader perspective in a general discussion (**Chapter 7**).

References

1. Pui, C. H., Gajjar, A. J., Kane, J. R., Qaddoumi, I. A. & Pappo, A. S. Challenging issues in pediatric oncology. *Nat Rev Clin Oncol* **8**, 540-549, doi:10.1038/nrclinonc.2011.95 (2011).

2. Viale, P. H. The American Cancer Society's Facts & Figures: 2020 Edition. *J Adv Pract Oncol* **11**, 135-136, doi:10.6004/jadpro.2020.11.2.1 (2020).

3. Steliarova-Foucher, E. *et al.* International incidence of childhood cancer, 2001-10: a population-based registry study. *Lancet Oncol* **18**, 719-731, doi:10.1016/S1470-2045(17)30186-9 (2017).

4. Grobner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321-327, doi:10.1038/nature25480 (2018).

5. Jones, D. T. W. *et al.* Molecular characteristics and therapeutic vulnerabilities across paediatric solid tumours. *Nat Rev Cancer* **19**, 420-438, doi:10.1038/s41568-019-0169-x (2019).

6. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371-376, doi:10.1038/nature25795 (2018).

7. Chun, H. E. *et al.* Identification and Analyses of Extra-Cranial and Cranial Rhabdoid Tumor Molecular Subgroups Reveal Tumors with Cytotoxic T Cell Infiltration. *Cell Rep* **29**, 2338-2354 e2337, doi:10.1016/j.celrep.2019.10.013 (2019).

8. Johann, P. D. *et al.* Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes. *Cancer Cell* **29**, 379-393, doi:10.1016/j.ccell.2016.02.001 (2016).

9. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* **11**, 220-228, doi:10.1038/nrm2858 (2010).

10. Filbin, M. & Monje, M. Developmental origins and emerging therapeutic opportunities for childhood cancer. *Nat Med* **25**, 367-376, doi:10.1038/s41591-019-0383-9 (2019).

11. Marshall, G. M. *et al.* The prenatal origins of cancer. *Nat Rev Cancer* **14**, 277-289, doi:10.1038/nrc3679 (2014).

12. Behjati, S., Gilbertson, R. J. & Pfister, S. M. Maturation Block in Childhood Cancer. *Cancer Discov* **11**, 542-544, doi:10.1158/2159-8290.CD-20-0926 (2021).

13. Chen, X., Pappo, A. & Dyer, M. A. Pediatric solid tumor genomics and developmental pliancy. *Oncogene* **34**, 5207-5215, doi:10.1038/onc.2014.474 (2015).

14. Hasselblatt, M. *et al.* High-resolution genomic analysis suggests the absence of recurrent genomic alterations other than SMARCB1 aberrations in atypical teratoid/rhabdoid tumors. *Genes Chromosomes Cancer* **52**, 185-190, doi:10.1002/gcc.22018 (2013).

15. Hasselblatt, M. *et al.* SMARCA4-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. *Acta Neuropathol* **128**, 453-456, doi:10.1007/s00401-014-1323-x (2014).

16. Lee, R. S. *et al.* A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. *J Clin Invest* **122**, 2983-2988, doi:10.1172/JCI64400 (2012).

17. Wilson, B. G. & Roberts, C. W. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* **11**, 481-492, doi:10.1038/nrc3068 (2011).

18. Brennan, B., Stiller, C. & Bourdeaut, F. Extracranial rhabdoid tumours: what we have learned so far and future directions. *Lancet Oncol* **14**, e329-336, doi:10.1016/S1470-2045(13)70088-3 (2013).

19. Fruhwald, M. C., Biegel, J. A., Bourdeaut, F., Roberts, C. W. & Chi, S. N. Atypical teratoid/rhabdoid tumors-current concepts, advances in biology, and potential future therapies. *Neuro Oncol* **18**, 764-778, doi:10.1093/neuonc/nov264 (2016).

20. Hoffman, L. M. *et al.* Advancing biology-based therapeutic approaches for atypical teratoid rhabdoid tumors. *Neuro Oncol* **22**, 944-954, doi:10.1093/neuonc/noaa046 (2020).

21. Nemes, K. *et al.* Clinical and genetic risk factors define two risk groups of extracranial malignant rhabdoid tumours (eMRT/RTK). *Eur J Cancer* **142**, 112-122, doi:10.1016/j.ejca.2020.10.004 (2021).

22. Reinhard, H. *et al.* Rhabdoid tumors in children: prognostic factors in 70 patients diagnosed in Germany. *Oncol Rep* **19**, 819-823 (2008).

23. Oeffinger, K. C. *et al.* Chronic health conditions in adult survivors of childhood cancer. *N Engl J Med* **355**, 1572-1582, doi:10.1056/NEJMsa060185 (2006).

24. Finetti, M. A., Grabovska, Y., Bailey, S. & Williamson, D. Translational genomics of malignant rhabdoid tumours: Current impact and future possibilities. *Semin Cancer Biol* **61**, 30-41, doi:10.1016/j.semcancer.2019.12.017 (2020).

25. Rorke, L. B., Packer, R. J. & Biegel, J. A. Central nervous system atypical teratoid/rhabdoid tumors of infancy and childhood: definition of an entity. *J Neurosurg* **85**, 56-65, doi:10.3171/jns.1996.85.1.0056 (1996).

26. Itakura, E. *et al.* Subcellular distribution of cytokeratin and vimentin in malignant rhabdoid tumor: three-dimensional imaging with confocal laser scanning microscopy and double immunofluorescence. *Mod Pathol* **14**, 854-861, doi:10.1038/modpathol.3880401 (2001).

27. Louis, D. N. *et al.* The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* **114**, 97-109, doi:10.1007/s00401-007-0243-4 (2007).

28. Bourdeaut, F. *et al.* hSNF5/INI1-deficient tumours and rhabdoid tumours are convergent but not fully overlapping entities. *J Pathol* **211**, 323-330, doi:10.1002/path.2103 (2007).

29. Beckwith, J. B. & Palmer, N. F. Histopathology and prognosis of Wilms tumors: results from the First National Wilms' Tumor Study. *Cancer* **41**, 1937-1948, doi:10.1002/1097-0142(197805)41:5<1937::aid-cncr2820410538>3.0.co;2-u (1978).

30. Versteege, I. *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394**, 203-206, doi:10.1038/28212 (1998).

31. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* **131**, 803-820, doi:10.1007/s00401-016-1545-1 (2016).

32. Chun, H. E. *et al.* Genome-Wide Profiles of Extra-cranial Malignant Rhabdoid Tumors Reveal Heterogeneity and Dysregulated Developmental Pathways. *Cancer Cell* **29**, 394-406, doi:10.1016/j.ccell.2016.02.009 (2016).

33. Kieran, M. W. *et al.* Absence of oncogenic canonical pathway mutations in aggressive pediatric rhabdoid tumors. *Pediatr Blood Cancer* **59**, 1155-1157, doi:10.1002/pbc.24315 (2012).

34. Eaton, K. W., Tooke, L. S., Wainwright, L. M., Judkins, A. R. & Biegel, J. A. Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors. *Pediatr Blood Cancer* **56**, 7-15, doi:10.1002/pbc.22831 (2011).

35. Sredni, S. T. & Tomita, T. Rhabdoid tumor predisposition syndrome. *Pediatr Dev Pathol* **18**, 49-58, doi:10.2350/14-07-1531-MISC.1 (2015).

36. Birks, D. K. *et al.* High expression of BMP pathway genes distinguishes a subset of atypical teratoid/rhabdoid tumors associated with shorter survival. *Neuro Oncol* **13**, 1296-1307, doi:10.1093/neuonc/nor140 (2011).

37. Torchia, J. *et al.* Molecular subgroups of atypical teratoid rhabdoid tumours in children: an integrated genomic and clinicopathological analysis. *Lancet Oncol* **16**, 569-582, doi:10.1016/S1470-2045(15)70114-2 (2015).

38. Torchia, J. *et al.* Integrated (epi)-Genomic Analyses Identify Subgroup-Specific Therapeutic Targets in CNS Rhabdoid Tumors. *Cancer Cell* **30**, 891-908, doi:10.1016/j.ccell.2016.11.003 (2016).

39. Holdhof, D. *et al.* Atypical teratoid/rhabdoid tumors (ATRTs) with SMARCA4 mutation are molecularly distinct from SMARCB1-deficient cases. *Acta Neuropathol* **141**, 291-301, doi:10.1007/s00401-020-02250-7 (2021).

40. Robertson, K. D. DNA methylation and human disease. *Nat Rev Genet* **6**, 597-610, doi:10.1038/nrg1655 (2005).

41. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322, doi:10.1038/nature08514 (2009).

42. Brinkman, A. B. *et al.* Partially methylated domains are hypervariable in breast cancer and fuel widespread CpG island hypermethylation. *Nat Commun* **10**, 1749, doi:10.1038/s41467-019-09828-0 (2019).

43. Salhab, A. *et al.* A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains. *Genome Biol* **19**, 150, doi:10.1186/s13059-018-1510-5 (2018).

44. Sur, I. & Taipale, J. The role of enhancers in cancer. *Nat Rev Cancer* **16**, 483-493, doi:10.1038/nrc.2016.62 (2016).

45. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207-220, doi:10.1038/s41576-018-0089-8 (2019).

46. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407-422, doi:10.1038/nrm.2017.26 (2017).

47. Valencia, A. M. & Kadoch, C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. *Nat Cell Biol* **21**, 152-161, doi:10.1038/s41556-018-0258-1 (2019).

48. Mashtalir, N. *et al.* Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* **175**, 1272-1288 e1220, doi:10.1016/j.cell.2018.09.032 (2018).

49. Michel, B. C. *et al.* A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol* **20**, 1410-1420, doi:10.1038/s41556-018-0221-1 (2018).

50. Gatchalian, J. *et al.* A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. *Nat Commun* **9**, 5139, doi:10.1038/s41467-018-07528-9 (2018).

51. Wang, X. *et al.* SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. *Nat Genet* **49**, 289-295, doi:10.1038/ng.3746 (2017).

52. Cenik, B. K. & Shilatifard, A. COMPASS and SWI/SNF complexes in development and disease. *Nat Rev Genet* **22**, 38-58, doi:10.1038/s41576-020-0278-0 (2021).

53. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* **106**, 5181-5186, doi:10.1073/pnas.0812889106 (2009).

54. Lessard, J. *et al.* An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201-215, doi:10.1016/j.neuron.2007.06.019 (2007).

55. Staahl, B. T. *et al.* Kinetic analysis of npBAF to nBAF switching reveals exchange of SS18 with CREST and integration with neural developmental pathways. *J Neurosci* **33**, 10348-10361, doi:10.1523/JNEUROSCI.1258-13.2013 (2013).

56. Mittal, P. & Roberts, C. W. M. The SWI/SNF complex in cancer - biology, biomarkers and therapy. *Nat Rev Clin Oncol* **17**, 435-448, doi:10.1038/s41571-020-0357-3 (2020).

57. Wang, X. *et al.* BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. *Nat Commun* **10**, 1881, doi:10.1038/s41467-019-09891-7 (2019).

58. Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun* **8**, 14648, doi:10.1038/ncomms14648 (2017).

59. Chasse, M. H. *et al.* Mithramycin induces promoter reprogramming and differentiation of rhabdoid tumor. *EMBO Mol Med* **13**, e12640, doi:10.15252/emmm.202012640 (2021).

60. Erkek, S. *et al.* Comprehensive Analysis of Chromatin States in Atypical Teratoid/Rhabdoid Tumor Identifies Diverging Roles for SWI/SNF and Polycomb in Gene Regulation. *Cancer Cell* **35**, 95-110 e118, doi:10.1016/j.ccell.2018.11.014 (2019).

61. Kramer, K. F., Moreno, N., Fruhwald, M. C. & Kerl, K. BRD9 Inhibition, Alone or in Combination with Cytostatic Compounds as a Therapeutic Approach in Rhabdoid Tumors. *Int J Mol Sci* **18**, doi:10.3390/ijms18071537 (2017).

62. Kadoch, C., Copeland, R. A. & Keilhack, H. PRC2 and SWI/SNF Chromatin Remodeling Complexes in Health and Disease. *Biochemistry* **55**, 1600-1614, doi:10.1021/acs.biochem.5b01191 (2016).

63. Alimova, I. *et al.* Inhibition of EZH2 suppresses self-renewal and induces radiation sensitivity in atypical rhabdoid teratoid tumor cells. *Neuro Oncol* **15**, 149-160, doi:10.1093/neuonc/nos285 (2013).

64. Kerl, K. *et al.* The histone deacetylase inhibitor SAHA acts in synergism with fenretinide and doxorubicin to control growth of rhabdoid tumor cells. *BMC Cancer* **13**, 286, doi:10.1186/1471-2407-13-286 (2013).

65. Knutson, S. K. *et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc Natl Acad Sci U S A* **110**, 7922-7927, doi:10.1073/pnas.1303800110 (2013).

66. Muscat, A. *et al.* Low-Dose Histone Deacetylase Inhibitor Treatment Leads to Tumor Growth Arrest and Multi-Lineage Differentiation of Malignant Rhabdoid Tumors. *Clin Cancer Res* **22**, 3560-3570, doi:10.1158/1078-0432.CCR-15-2260 (2016).

67. Albanese, P., Belin, M. F. & Delattre, O. The tumour suppressor hSNF5/INI1 controls the differentiation potential of malignant rhabdoid cells. *Eur J Cancer* **42**, 2326-2334, doi:10.1016/j.ejca.2006.03.028 (2006).

68. Betz, B. L., Strobeck, M. W., Reisman, D. N., Knudsen, E. S. & Weissman, B. E. Re-expression of hSNF5/INI1/BAF47 in pediatric tumor cells leads to G1 arrest associated with induction of p16ink4a and activation of RB. *Oncogene* **21**, 5193-5203, doi:10.1038/sj.onc.1205706 (2002).

69. Oruetxebarria, I. *et al.* P16INK4a is required for hSNF5 chromatin remodeler-induced cellular senescence in malignant rhabdoid tumor cells. *J Biol Chem* **279**, 3807-3816, doi:10.1074/jbc.M309333200 (2004).

70. Alarcon-Vargas, D. *et al.* Targeting cyclin D1, a downstream effector of INI1/hSNF5, in rhabdoid tumors. *Oncogene* **25**, 722-734, doi:10.1038/sj.onc.1209112 (2006).

71. Tsikitis, M., Zhang, Z., Edelman, W., Zagzag, D. & Kalpana, G. V. Genetic ablation of Cyclin D1 abrogates genesis of rhabdoid tumors resulting from Ini1 loss. *Proc Natl Acad Sci U S A* **102**, 12129-12134, doi:10.1073/pnas.0505300102 (2005).

72. Geoerger, B. *et al.* A Phase I Study of the CDK4/6 Inhibitor Ribociclib (LEE011) in Pediatric Patients with Malignant Rhabdoid Tumors, Neuroblastoma, and Other Solid Tumors. *Clin Cancer Res* **23**, 2433-2441, doi:10.1158/1078-0432.CCR-16-2898 (2017).

73. Mosse, Y. P. *et al.* A Phase II Study of Alisertib in Children with Recurrent/Refractory Solid Tumors or Leukemia: Children's Oncology Group Phase I and Pilot Consortium (ADVL0921). *Clin Cancer Res* **25**, 3229-3238, doi:10.1158/1078-0432.CCR-18-2675 (2019).

74. Darr, J., Klochendler, A., Isaac, S., Geiger, T. & Eden, A. Phosphoproteomic analysis reveals Smarcb1 dependent EGFR signaling in Malignant Rhabdoid tumor cells. *Mol Cancer* **14**, 167, doi:10.1186/s12943-015-0439-5 (2015).

75. Oberlick, E. M. *et al.* Small-Molecule and CRISPR Screening Converge to Reveal Receptor Tyrosine Kinase Dependencies in Pediatric Rhabdoid Tumors. *Cell Rep* **28**, 2331-2344 e2338, doi:10.1016/j.celrep.2019.07.021 (2019).

76. Wohrle, S. *et al.* Fibroblast growth factor receptors as novel therapeutic targets in SNF5-deleted malignant rhabdoid tumors. *PLoS One* **8**, e77652, doi:10.1371/journal.pone.0077652 (2013).

77. Chauvin, C. *et al.* High-Throughput Drug Screening Identifies Pazopanib and Clofilium Tosylate as Promising Treatments for Malignant Rhabdoid Tumors. *Cell Rep* **21**, 1737-1745, doi:10.1016/j.celrep.2017.10.076 (2017).

78. Obaid, H. *et al. In Vitro* Investigation Demonstrates IGFR/VEGFR Receptor Cross Talk and Potential of Combined Inhibition in Pediatric Central Nervous System Atypical Teratoid Rhabdoid Tumors. *Curr Cancer Drug Targets* **20**, 295-305, doi:10.2174/1568009619666191111153049 (2020).

79. Singh, A. *et al.* Profiling pathway-specific novel therapeutics in preclinical assessment for central nervous system atypical teratoid rhabdoid tumors (CNS ATRT): favorable activity of targeting EGFR- ErbB2 signaling with lapatinib. *Mol Oncol* **7**, 497-512, doi:10.1016/j.molonc.2013.01.001 (2013).

80. Wong, J. P. *et al.* Dual Targeting of PDGFRalpha and FGFR1 Displays Synergistic Efficacy in Malignant Rhabdoid Tumors. *Cell Rep* **17**, 1265-1275, doi:10.1016/j.celrep.2016.10.005 (2016).

81. Alimova, I. *et al.* Inhibition of MYC attenuates tumor cell self-renewal and promotes senescence in SMARCB1-deficient Group 2 atypical teratoid rhabdoid tumors to suppress tumor growth *in vivo*. *Int J Cancer* **144**, 1983-1995, doi:10.1002/ijc.31873 (2019).

82. Chen, H., Liu, H. & Qing, G. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduct Target Ther* **3**, 5, doi:10.1038/s41392-018-0008-7 (2018).

83. Carugo, A. *et al.* p53 Is a Master Regulator of Proteostasis in SMARCB1-Deficient Malignant Rhabdoid Tumors. *Cancer Cell* **35**, 204-220 e209, doi:10.1016/j.ccell.2019.01.006 (2019).

84. Howard, T. P. *et al.* Rhabdoid Tumors Are Sensitive to the Protein-Translation Inhibitor Homoharringtonine. *Clin Cancer Res* **26**, 4995-5006, doi:10.1158/1078-0432.CCR-19-2717 (2020).

85. Moreno, N. *et al.* Combined BRD4 and CDK9 inhibition as a new therapeutic approach in malignant rhabdoid tumors. *Oncotarget* **8**, 84986-84995, doi:10.18632/oncotarget.18583 (2017).

86. Morin, A. *et al.* Proteasome inhibition as a therapeutic approach in atypical teratoid/rhabdoid tumors. *Neurooncol Adv* **2**, vdaa051, doi:10.1093/noajnl/vdaa051 (2020).

87. Tran, H. M. *et al.* Upregulation of Protein Synthesis and Proteasome Degradation Confers Sensitivity to Proteasome Inhibitor Bortezomib in Myc-Atypical Teratoid/Rhabdoid Tumors. *Cancers (Basel)* **12**, doi:10.3390/cancers12030752 (2020).

88. Kerl, K. *et al.* Arsenic trioxide inhibits tumor cell growth in malignant rhabdoid tumors *in vitro* and *in vivo* by targeting overexpressed Gli1. *Int J Cancer* **135**, 989-995, doi:10.1002/ijc.28719 (2014).

89. DelBove, J. *et al.* Inactivation of SNF5 cooperates with p53 loss to accelerate tumor formation in Snf5(+/-);p53(+/-) mice. *Mol Carcinog* **48**, 1139-1148, doi:10.1002/mc.20568 (2009).

90. Guidi, C. J. *et al.* Disruption of Ini1 leads to peri-implantation lethality and tumorigenesis in mice. *Mol Cell Biol* **21**, 3598-3603, doi:10.1128/MCB.21.10.3598-3603.2001 (2001).

91. Han, Z. Y. *et al.* The occurrence of intracranial rhabdoid tumours in mice depends on temporal control of Smarcb1 inactivation. *Nat Commun* **7**, 10421, doi:10.1038/ncomms10421 (2016).

92. Klochendler-Yeivin, A. *et al.* The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. *EMBO Rep* **1**, 500-506, doi:10.1093/embo-reports/kvd129 (2000).

93. Ng, J. M. *et al.* Generation of a mouse model of atypical teratoid/rhabdoid tumor of the central nervous system through combined deletion of Snf5 and p53. *Cancer Res* **75**, 4629-4639, doi:10.1158/0008-5472.CAN-15-0874 (2015).

94. Roberts, C. W., Galusha, S. A., McMenamin, M. E., Fletcher, C. D. & Orkin, S. H. Haploinsufficiency of Snf5 (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice. *Proc Natl Acad Sci U S A* **97**, 13796-13800, doi:10.1073/pnas.250492697 (2000).

95. Roberts, C. W., Leroux, M. M., Fleming, M. D. & Orkin, S. H. Highly penetrant, rapid tumorigenesis through conditional inversion of the tumor suppressor gene Snf5. *Cancer Cell* **2**, 415-425, doi:10.1016/s1535-6108(02)00185-x (2002).

96. Vitte, J., Gao, F., Coppola, G., Judkins, A. R. & Giovannini, M. Timing of Smarcb1 and Nf2 inactivation determines schwannoma versus rhabdoid tumor development. *Nat Commun* **8**, 300, doi:10.1038/s41467-017-00346-5 (2017).

97. Liu, G., David, B. T., Trawczynski, M. & Fessler, R. G. Advances in Pluripotent Stem Cells: History, Mechanisms, Technologies, and Applications. *Stem Cell Rev Rep* **16**, 3-32, doi:10.1007/s12015-019-09935-x (2020).

98. Funato, K., Major, T., Lewis, P. W., Allis, C. D. & Tabar, V. Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation. *Science* **346**, 1529-1533, doi:10.1126/science.1253799 (2014).

99. Langer, L. F., Ward, J. M. & Archer, T. K. Tumor suppressor SMARCB1 suppresses super-enhancers to govern hESC lineage determination. *Elife* **8**, doi:10.7554/eLife.45672 (2019).

100. Terada, Y. *et al.* Human Pluripotent Stem Cell-Derived Tumor Model Uncovers the Embryonic Stem Cell Signature as a Key Driver in Atypical Teratoid/Rhabdoid Tumor. *Cell Rep* **26**, 2608-2621 e2606, doi:10.1016/j.celrep.2019.02.009 (2019).

101. Kamb, A. What's wrong with our cancer models? *Nat Rev Drug Discov* **4**, 161-165, doi:10.1038/nrd1635 (2005).

102. Gillet, J. P., Varma, S. & Gottesman, M. M. The clinical relevance of cancer cell lines. *J Natl Cancer Inst* **105**, 452-458, doi:10.1093/jnci/djt007 (2013).

103. Clevers, H. Modeling Development and Disease with Organoids. *Cell* **165**, 1586-1597, doi:10.1016/j.cell.2016.05.082 (2016).

104. Drost, J. & Clevers, H. Organoids in cancer research. *Nat Rev Cancer* **18**, 407-418, doi:10.1038/s41568-018-0007-6 (2018).

105. Ganesh, K. *et al.* A rectal cancer organoid platform to study individual responses to chemoradiation. *Nat Med* **25**, 1607-1614, doi:10.1038/s41591-019-0584-2 (2019).

106. Ooft, S. N. *et al.* Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. *Sci Transl Med* **11**, doi:10.1126/scitranslmed.aay2574 (2019).

107. Tiriac, H. *et al.* Organoid Profiling Identifies Common Responders to Chemotherapy in Pancreatic Cancer. *Cancer Discov* **8**, 1112-1129, doi:10.1158/2159-8290.CD-18-0349 (2018).

108. Vlachogiannis, G. *et al.* Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* **359**, 920-926, doi:10.1126/science.aao2774 (2018).

109. Yao, Y. *et al.* Patient-Derived Organoids Predict Chemoradiation Responses of Locally Advanced Rectal Cancer. *Cell Stem Cell* **26**, 17-26 e16, doi:10.1016/j.stem.2019.10.010 (2020).

110. Calandrini, C. *et al.* An organoid biobank for childhood kidney cancers that captures disease and tissue heterogeneity. *Nat Commun* **11**, 1310, doi:10.1038/s41467-020-15155-6 (2020).

111. Deisch, J., Raisanen, J. & Rakheja, D. Immunohistochemical expression of embryonic stem cell markers in malignant rhabdoid tumors. *Pediatr Dev Pathol* **14**, 353-359, doi:10.2350/10-09-0902-OA.1 (2011).

112. Jessa, S. *et al.* Stalled developmental programs at the root of pediatric brain tumors. *Nat Genet* **51**, 1702-1713, doi:10.1038/s41588-019-0531-7 (2019).

113. Filbin, M. G. *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **360**, 331-335, doi:10.1126/science.aao4750 (2018).

114. Gojo, J. *et al.* Single-Cell RNA-Seq Reveals Cellular Hierarchies and Impaired Developmental Trajectories in Pediatric Ependymoma. *Cancer Cell* **38**, 44-59 e49, doi:10.1016/j.ccell.2020.06.004 (2020).

115. Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599, doi:10.1126/science.aat1699 (2018).

116. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science* **370**, doi:10.1126/science.aba7721 (2020).

117. Avila Cobos, F., Vandesompele, J., Mestdagh, P. & De Preter, K. Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics* **34**, 1969-1979, doi:10.1093/bioinformatics/bty019 (2018).

118. Hovestadt, V. *et al.* Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**, 74-79, doi:10.1038/s41586-019-1434-6 (2019).

119. Baron, C. S. & van Oudenaarden, A. Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat Rev Mol Cell Biol* **20**, 753-765, doi:10.1038/s41580-019-0186-3 (2019).

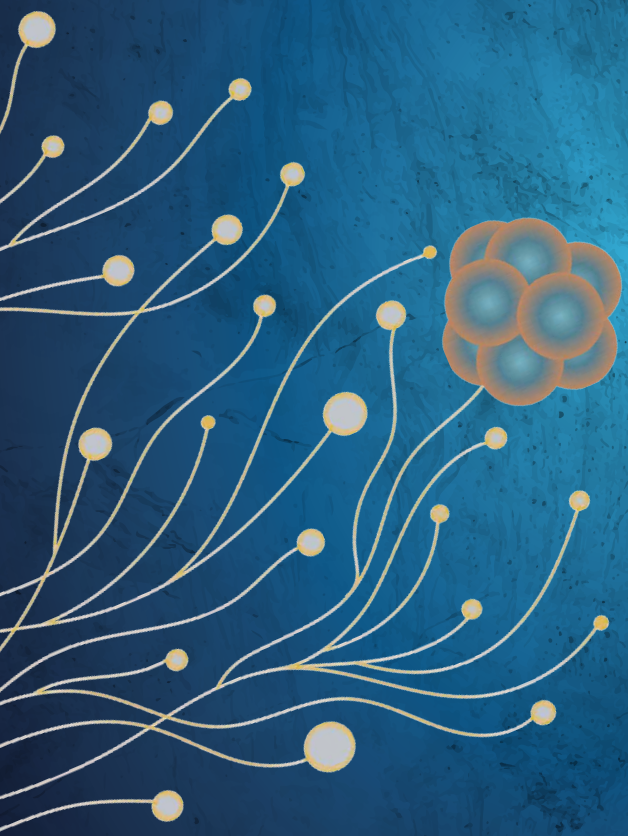
120. Lodato, M. A. *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94-98, doi:10.1126/science.aab1785 (2015).

121. Coorens, T. H. H. *et al.* Embryonal precursors of Wilms tumor. *Science* **366**, 1247-1251, doi:10.1126/science.aax1323 (2019).

122. Custers, L., Paassen, I. & Drost, J. *In vitro* Modeling of Embryonal Tumors. *Front Cell Dev Biol* **9**, 640633, doi:10.3389/fcell.2021.640633 (2021).

123. Young, M. D. et al. Single cell derived mRNA signals across human kidney tumors. Nat Commun 12, 3896, doi:10.1038/s41467-021-23949-5 (2021).

124. Custers, L. *et al.* Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nat Commun* **12**, 1407, doi:10.1038/s41467-021-21675-6 (2021).



CHAPTER 2

In vitro modeling of embryonal tumors

Lars Custers[1,2,†], Irene Paassen[1,2,†], Jarno Drost[1,2*]

***Front. Cell Dev. Biol.*, 26 February 2021**

[1]Princess Máxima Center for Pediatric Oncology, The Netherlands.

[2]Oncode Institute, The Netherlands

[†] These authors contributed equally to this work.

[*] These authors supervised this work.

Abstract

A subset of pediatric tumors affects very young children and are thought to arise during fetal life. A common theme is that these embryonal tumors hijack developmental programs, causing a block in differentiation and, as a consequence, unrestricted proliferation. Embryonal tumors, therefore typically maintain an embryonic gene signature not found in their differentiated progeny. Still, the processes underpinning malignant transformation remain largely unknown, which is hampering therapeutic innovation. To gain more insight into these processes, *in vitro* and *in vivo* research models are indispensable. However, embryonic development is an extremely dynamic process with continuously changing cellular identities, making it challenging to define cells-of-origin. This is crucial for the development of representative models, as targeting the wrong cell or targeting a cell within an incorrect developmental time window can result in completely different phenotypes. Recent innovations in *in vitro* cell models may provide more versatile platforms to study embryonal tumors in a scalable manner. In this review, we outline different *in vitro* models that can be explored to study embryonal tumorigenesis and for therapy development.

Introduction

Cancer is the leading disease-related cause of death in children (Siegel *et al.*, 2016; Cunningham *et al.*, 2018). A significant subset of pediatric tumors occurs in early childhood, suggestive of an origin in prenatal life (Marshall *et al.*, 2014). These so-called embryonal tumors are thought to develop as a consequence of aberrant development. However, for many embryonal tumors the processes driving tumorigenesis remain unknown. Whereas, adult cancers develop by a progressive accumulation of mutations over many years (Stratton *et al.*, 2009), embryonal tumors are typically characterized by a relatively low mutational burden and only a few genetic events to drive tumorigenesis (Vogelstein *et al.*, 2013; Gröbner *et al.*, 2018; Rahal *et al.*, 2018; Kattner *et al.*, 2019). The few genetic alterations that do occur likely cause fetal cells to maintain a progenitor-like state and prohibit differentiation. This maturation block has been suggested to prime cells for malignant transformation (Chen *et al.*, 2015; Puisieux *et al.*, 2018; Rahal *et al.*, 2018; Jessa *et al.*, 2019). To better understand the processes underpinning embryonal tumorigenesis, a direct comparison between normal and tumor development is key. Gene expression profiling of fetal tissues with single cell resolution has provided more insights into the developmental trajectories driving embryogenesis. Comparison of such profiles with tumor gene expression signatures have defined the cellular identity of several embryonal tumors, possibly pointing to their cellular origin (Boeva *et al.*, 2017; Young *et al.*, 2018, 2020; Hovestadt *et al.*, 2019; Jessa *et al.*, 2019; Vladoiu *et al.*, 2019). Yet, in many cases these studies are merely correlative and lack subsequent functional validation. To do so, representative *in vitro* and *in vivo* preclinical models are crucial.

Genetically engineered mouse models (GEMMs) have been the golden standard for finding the cellular origin of cancers, by introducing tumor driver events in putative tumor-initiating cells (Visvader, 2011; Marshall *et al.*, 2014). Although GEMMs have provided important insights into tumorigenesis, several drawbacks limit their potential as a representative model of embryonal tumors. Embryonic development is an extremely dynamic process with continuously changing cellular identities, which makes it very challenging to target the right cell at the right time. For instance, homozygous loss of the Wilms tumor driver gene *Wt1* was shown to be embryonically lethal in mice (Kreidberg *et al.*, 1993), whereas a specific *Wt1* ablation at E11.5 in a small fraction of nephron progenitor cells resulted in Wilms tumor formation (Hu *et al.*, 2011; Berry *et al.*, 2015; Huang *et al.*, 2016). Moreover, GEMM generation is time consuming and mouse development does not fully recapitulate human embryogenesis (Navin *et al.*, 2010, 2011; Blakeley *et al.*, 2015; Theunissen and Jaenisch, 2017). The development of new *in vitro* cell models increasingly recapitulating the complexity of organogenesis will open new avenues for the development of novel, relevant embryonal tumor models. In this review, we discuss the currently available *in vitro* models to study embryonal tumorigenesis as well as the discovery of new therapeutic strategies.

Cell Lines of Fetal Origin

A broad range of cell lines has been established over the last decades. Cell lines are easy to maintain and typically do not consume many resources, which allows for fast and parallel modeling of multiple tumor driver events. This is particularly useful to interrogate the complex genetics underlying heterogeneous tumor phenotypes. One such tumor is neuroblastoma, which is characterized by a variety of driver events, including *MYCN* amplification and *ALK* mutations (Johnsen *et al.*, 2019). To study neuroblastoma initiation, models of its embryonic origin, neural crest (Johnsen *et al.*, 2019), are required. *In vitro* murine neural crest models can be generated by extraction of neural tubes from mouse embryos, which are subsequently placed in a culture dish to initiate the migration of neural crest cells onto the plate (Maurer *et al.*, 2007; Olsen *et al.*, 2017). The neural crest cells lose their multipotency over time *in vitro* (6–10 cell divisions) (Stemple and Anderson, 1992) and are, therefore only suitable for short-term experiments. However, multipotency can be maintained by exogenous c-Myc expression. Accordingly, Maurer *et al.* (2007) generated the JoMa1 neural crest cell line, which was established from mouse embryos carrying the inducible c-MycER transgene, enabling tamoxifen-inducible c-Myc expression and maintenance of multipotency. In both the JoMa1 cell line (Schulte *et al.*, 2013) and non-genetically modified neural crest cells (Olsen *et al.*, 2017), overexpression of *MycN* was proven sufficient to generate neuroblastoma upon transplantation in immune-deficient mice. Other murine neural crest-derived neuroblastoma models accommodate oncogenic variants of *Alk* or *Phox2b*, which was shown to impair neural crest development and inhibit sympathoadrenal differentiation processes (Reiff *et al.*, 2010; Schulte *et al.*, 2013; Montavon *et al.*, 2014). However, murine neural crest

development has been shown to be different from human in many aspects (O'Rahilly and Müller, 2007; Betters *et al.*, 2010). Cohen *et al.* (2020), therefore developed a mouse-human chimera to study neuroblastoma formation in a human setting. Human iPSC-derived neural crest cells were injected *in utero* into gastrulating mouse embryos to form a human neural crest lineage in mice. For neuroblastoma modeling, the neural crest cells were subsequently genetically engineered with inducible expression constructs of *MYCN* and an oncogenic variant of *ALK*. Upon induction, mice developed tumors characteristic of patient neuroblastoma, and tumor transcriptomes resembled neuroblastoma patients more closely than GEMMs. Interestingly, injections subcutaneously lead to tumor formation but without expression of neuroblastoma markers (Cohen *et al.*, 2020). These findings suggest that human neural crest cells serve as a more representative model than mouse, but only when generated in the appropriate developmental context and orthotopic environment.

Another embryonal tumor entity where differences between human and mouse models of tumorigenesis were observed is retinoblastoma. The common driver event of retinoblastoma is loss of *RB1* during retinal development (Dimaras *et al.*, 2015). Retinoblastoma modeling using GEMMs has proven challenging, as engineering of *Rb1*-deficient mice resulted in embryonic lethality (Lee *et al.*, 1992; Wikenheiser-Brokamp, 2006) and retina-specific depletion of *Rb1* was required. However, in contrast to human, mouse retinal cells were proven insensitive to *Rb1* depletion and required additional knock-outs of tumor suppressors *p107* or *p130* for retinoblastoma development (Robanus-Maandag *et al.*, 1998; Dannenberg *et al.*, 2004; MacPherson *et al.*, 2004). To generate human models of retinal development, Xu *et al.* (2014) isolated human fetal retinal cells post-fertilization retaining all retinal precursor cell types (RPCs) at distinct maturation states. Depletion of *RB1* within the different RPCs indicated post-mitotic cone-precursors to be most prone to develop into retinoblastoma, based on its ability to form tumors with expression of retinoblastoma markers upon xenografting in mice (Xu *et al.*, 2014). Furthermore, *RB1* loss in matured retinal cells did not induce retinoblastoma, validating that tumor initiation is restricted to a specific cell within retinal development.

Overall, *in vitro* modeling of retinoblastoma and neuroblastoma in human and mouse fetal cell cultures uncovered that fundamental differences between mice and human development can impede representative modeling of embryonal tumors.

Pluripotent Stem Cell-Derived Cell Lines

Classical cell lines are typically composed of a single type of progenitor-like cell representing a specific germ layer (i.e., endoderm, ectoderm, mesoderm, neural crest). Culture models still capable of generating the different germ layers give the opportunity to model embryonal tumors of which it is not yet clear from which lineage they arise, or which seem to arise across the boundaries of the different germ layers. Current *in vitro* models capable of recapitulating these different

developmental trajectories include pluripotent stem cells (PSCs) such as embryonic stem cells (ESCs) and induced PSCs (iPSCs), which can self-renew and be subjected to differentiation protocols that enforce all germ layers (Liu G. *et al.*, 2020). PSCs can be stably maintained in culture and are permissive for genetic manipulation (Liu G. *et al.*, 2020). With the development of effective differentiation protocols, PSCs can mirror embryonic development and therefore serve as a valuable model to study tumorigenesis. iPSCs are generated through the forced dedifferentiation of somatic cells, which thereby regain pluripotency. The molecular mechanisms that underly this reprogramming show significant similarities with the processes driving a subset of the embryonal germ cell tumors (GCTs) (Oosterhuis and Looijenga, 2019), including yolk sac tumors, embryonal carcinomas, and teratomas. GCTs encompass a diverse group of cancer entities that arise from cells of the early embryo or germ line (Oosterhuis and Looijenga, 2019). Interestingly, somatic mutations play a minor role as drivers of GCT development. Tumors are thought to arise by epigenetic deregulation of the cell-of-origin or aberrant stem cell niche factors (Oosterhuis and Looijenga, 2019). The developmental potency of the cell-of-origin can be reprogrammed through increased expression of well-known pluripotency factors, such as NANOG and OCT4 (De Jong and Looijenga, 2006; Thomas *et al.*, 2011). Xenograft studies have shown that iPSCs and ESCs are intrinsically tumorigenic (Ben-David and Benvenisty, 2011). Upon xenografting, iPSCs develop into a benign GCT referred to as teratoma or in some cases more malignant GCTs, dependent on the reprogramming method applied (Lee *et al.*, 2013). These findings indicate that maintaining an early embryonic cellular context is, by itself, sufficient for tumor initiation. Although PSC tumorigenicity is a limitation for its potential application in regenerative medicine, iPSCs and ESCs can on the other hand serve as *in vitro* models of GCTs.

A major class of genes mutated in childhood as well as adult cancers are subunits of the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodeling complex (Wilson and Roberts, 2011; Shain and Pollack, 2013). The role of this complex in embryonal tumors is clearly exemplified in malignant rhabdoid tumors (MRT), which are characterized by the complete loss of SWI/SNF subunit *SMARCB1* (95% of cases) or *SMARCA4* (5% of cases) (Lee *et al.*, 2012; Hasselblatt *et al.*, 2014). To study MRT initiation, *SMARCB1* was knocked down in hESCs using RNA interference (Langer *et al.*, 2019). The differentiation capacity of hESCs was subsequently assessed, demonstrating that *SMARCB1* inhibition specifically repressed neural induction, whereas mesodermal and endodermal lineage induction was not affected (Langer *et al.*, 2019). In culture conditions inducing neural differentiation, SMARCB1 was shown to be essential for increased chromatin accessibility at neural differentiation genes and silencing of pluripotency-related super-enhancers (Wang *et al.*, 2017; Langer *et al.*, 2019). Furthermore, *SMARCB1*-null iPSCs that were transplanted into mice were able to generate MRT (Terada *et al.*, 2019). Interestingly, iPSCs that had further progressed to neural progenitor cells (NPCs) generated tumors without rhabdoid features. These results show a lineage-specific role for *SMARCB1 in vitro*, validating recently

developed MRT GEMMs wherein *Smarcb1* loss-induced rhabdoid tumor development was demonstrated to be limited to a specific developmental time and lineage (Han *et al.*, 2016; Vitte *et al.*, 2017).

A different layer of epigenetic regulation affected in embryonal tumors is the post-translational modification of histone tails, which enables a rapid switch between active or repressive histone marks to dynamically regulate gene expression during development. Mutations in histones are specifically characterized in a subset of pediatric gliomas. In diffuse intrinsic pontine glioma (DIPG), nearly 80% of cases have a missense mutation in the histone 3.3 gene (*H3F3A*), causing a substitution of methionine for lysine 27 (H3K27M) (Khuong-Quang *et al.*, 2012; Schwartzentruber *et al.*, 2012; Wu *et al.*, 2012). The origin of DIPG was indicated to lie in early neural development (Filbin *et al.*, 2018; Sun *et al.*, 2019), presumably making NPCs derived from ESCs a suitable model for tumor initiation. In line with this, overexpression of the H3K27M mutant in NPCs resulted in increased proliferation (Funato *et al.*, 2014). Interestingly, introduction of the mutation was ineffective in uninduced ESCs or mature astrocytes. For a majority of DIPG cases, H3K27M mutations are typically co-occurring with amplification of *PDGFRA* and loss of *TP53* (Khuong-Quang *et al.*, 2012). Combined introduction of these three genetic events in NPCs induced more extensive neoplastic features, generating DIPG when transplanted in mice (Funato *et al.*, 2014). This combination of mutations prohibited early NPCs to differentiate to astrocytes (Funato *et al.*, 2014), explaining the observed maturation block in DIPG.

MRT and DIPG modeling approaches using ESCs and iPSCs have demonstrated that a specific cellular context is required for malignant transformation, meaning that tumorigenesis is restricted to a specific developmental time and fetal cell type.

Pluripotent Stem Cell-Derived Organoids

Recent innovations in three-dimensional (3D) culture technology, such as organoids, has opened new opportunities for generating additional representative models of embryonal tumors. Organoids can be derived from adult (ASC) or pluripotent stem cells. They typically capture the cellular and genetic heterogeneity of native tissue and recapitulate cellular hierarchy and dynamics to a large extent, which is most likely a consequence of their 3D architecture (Clevers, 2016). Therefore, 3D organoid cultures seem to better recapitulate organ morphogenesis (Clevers, 2016).

Following that rationale, 3D retinal organoids were established from hESCs or iPSCs, allowing for more comprehensive studies of retinoblastoma initiation in human cells (Zhong *et al.*, 2014; Kuwahara *et al.*, 2015). Loss of *RB1* in retinal organoids showed a dysregulation of retinal maturation processes, impairing differentiation toward photoreceptors, ganglion, and bipolar cells (Zheng *et al.*, 2020). However, the depletion of *RB1* was not sufficient for retinoblastoma initiation as the organoids did not fully recapitulate the retinoblastoma cell phenotype. In addition, transplantation of *RB1*-null organoids into immune-deficient mice did not result in retinoblastoma

formation (Zheng *et al.*, 2020). In contrast, Liu H. *et al.* (2020) utilized an alternative hESC-derived retinal organoid model, in which *RB1* depletion did successfully generate tumors upon xenografting and better resembled patient retinoblastoma. These findings illustrate that the finetuning of retinal organoid establishment can affect the outcome of *RB1* depletion, possibly due to differences in cellular composition and the presence or absence of the cell-of-origin. These studies further highlight the specific cellular context required for retinoblastoma initiation and point out a possible limitation of PSC-derived models, as they may not be able to generate the full extent of cell-types found *in vivo*.

A frequent source of embryonal tumors is the embryonic brain. Human brain development can be mimicked by differentiation of PSCs to neural progenitor cells. In culture, they can self-organize into cerebral or cerebellar organoids containing different cell types in a polarized structure (Muguruma *et al.*, 2015; Luo *et al.*, 2016). Embryonic cerebellar organoids have been successfully used to model pediatric brain tumors, including medulloblastoma and rhabdoid tumors (Ballabio *et al.*, 2020; Parisian *et al.*, 2020). Organoid cultures can be utilized to introduce tumorigenic mutations in a systematic manner, as shown for cerebral organoids (Bian *et al.*, 2018), demonstrating the potential for high-throughput *in vitro* tumor modeling. Furthermore, cerebellar organoids can be exploited to decipher tumor subtype-specific processes. Medulloblastoma, among other embryonal tumor entities, is classified into subtypes based on the oncogenic activation of specific signaling pathways (Cavalli *et al.*, 2017). The medulloblastoma subgroup 3 (*MYC* amplified subgroup) was successfully modeled in cerebellar organoids by combination of *MYC* and *OTX2* or *GFI1* overexpression (Ballabio *et al.*, 2020). The genetically modified cerebellar organoids showed increased proliferation and enrichment for progenitor cells, indicative of a differentiation block. Upon transplantation into mice, medulloblastomas developed resembling subgroup 3 tumors based on marker genes and DNA methylation patterns. Other medulloblastoma subtypes, likely arising from distinct neural differentiation trajectories, have not been modeled *in vitro* up to date (Gibson *et al.*, 2010; Grammel *et al.*, 2012; Hovestadt *et al.*, 2019). To do so, tumor initiation models composed of different neural lineages may be required.

Overall, the development of embryonic organoid cultures has provided relevant models of embryonal tumorigenesis. By approaching *in vivo* physiology, human organoids may serve as a promising alternative for time- and labor-intensive *in vivo* studies.

Reverse Tumor Modeling and Differentiation Therapy

Relieving the differentiation block underpinning embryonal tumor development could potentially serve as a therapeutic approach (i.e., maturation therapy). To develop such therapies, the differentiation block must first be defined, which can be achieved through reverse tumor modeling by, for instance, reverting the oncogenic driver in cultured tumor cells. Following this principle, inhibition of N-MYC in *MYCN*-

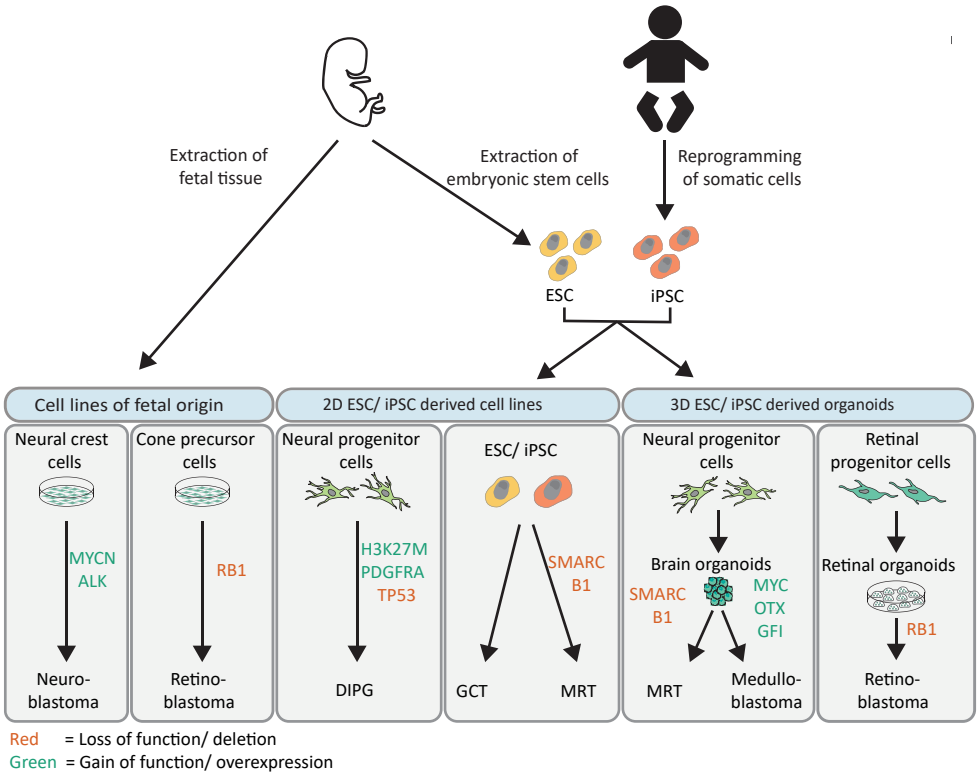


Figure 1 | Overview of embryonal tumor modeling techniques.
Illustration that summarizes the different *in vitro* approaches used to model embryonal tumors (MRT = malignant rhabdoid tumor; DIPG = diffuse intrinsic pontine glioma, GCT = germ cell tumor). *In vitro* tumor models are grouped by their source: cell lines of fetal origin, 2D embryonic stem cell (ESC) derived or induced pluripotent stem cell derived (iPSC) cell lines and 3D ESC/iPSC derived organoids. The gene-editing of tumor driver events is indicated (red = loss of function/deletions; green = gain of function/overexpression). Permission to reuse and Copyright: Medical illustrations used in in this figure were modified from Servier Medical Art, licensed under a Creative Commons Attribution 3.0 Generic License.

amplified neuroblastoma cell lines induced a differentiation morphology as well as upregulation of neural differentiation genes (Kang *et al.*, 2006; Henriksen *et al.*, 2011; Jiang *et al.*, 2011; Westermark *et al.*, 2011; Hossain *et al.*, 2013). Differentiation phenotypes were also observed upon genetic manipulation of medulloblastoma models (Liu *et al.*, 2017; Cheng *et al.*, 2020; Zagozewski *et al.*, 2020), and MRT models (Betz *et al.*, 2002; Nakayama *et al.*, 2017; Wang *et al.*, 2017). These studies show that reversal of the genetic driver can transform tumor cells to a more mature cell state, possibly reflecting the matured cell type it would have become, had it not become cancerous. Genetic repair of driver genes is not feasible at present (Dunbar *et al.*, 2018). An alternative strategy is to induce differentiation pharmacologically. For instance, experiments performed in MRT models with *SMARCB1* re-expression

Table 1 | *In vitro* embryonic tumor initiation models and differentiation therapies.

Tumor	Origin	*In vitro* models	Differentiation therapy
Neuroblastoma	Neural crest cells (NCCs)	*MYCN* overexpression in mouse primary NCCs (Olsen *et al.*, 2017)	Retinoic acid treatment (Lone *et al.*, 2016; Westerlund *et al.*, 2017)
		MYCN / *ALK-F1174L* overexpression in a mouse NC cell-line (Schulte *et al.*, 2013)	HDAC inhibitors (Frumm *et al.*, 2013; Hahn *et al.*, 2008)
		Mouse-human chimeras with *MYCN* over-expression in iPSC-derived hNCCs (Cohen *et al.*, 2020)	EZH2 inhibitors (L. Chen *et al.*, 2018)
		Engineering human 1p36 deletions in mouse NCCs (García-López *et al.*, 2020)	
MRT	Neural crest cells (NCCs)	*SMARCB1* knockout in iPSCs (Terada *et al.*, 2019)	HDAC inhibitors (Muscat *et al.*, 2016)
		SMARCB1 knockdown in ESCs (Langer *et al.*, 2019)	EZH2 inhibitors (Knutson *et al.*, 2013)
		SMARCB1 knockout in cerebellar organoids (Parisian *et al.*, 2020)	
Medulloblas-toma	Neural progenitor cells	*MYC* overexpression in cerebellar organ-oids (Ballabio *et al.*, 2020)	Retinoic acid treatment (Patties *et al.*, 2016)
		MYCN overexpression in neuroepithelial stem cells (M. Huang *et al.*, 2019)	BET-bromodomain inhibitors (Bandopadhayay *et al.*, 2019)
			SHH inhibitors (Ocasio *et al.*, 2019)
			EZH2 inhibitors (Cheng *et al.*, 2020)
DIPG	Oligodendrocyte precursor cells	H3K27M mutations in hESC derived NPCs (Funato *et al.*, 2014)	HDAC inhibitors (Anastas *et al.*, 2019
		ACVR1 mutations in neurospheres (Hoeman *et al.*, 2019)	BET-bromodomain inhibitors (Mohammad *et al.*, 2017)
Retinoblastoma	Cone precursor cells	*RB1* depletion in fetal retinal cell cultures (Xu *et al.*, 2014)	
		RB1 depletion in hESC derived retinal organoids (Zheng *et al.*, 2020, Hui Liu *et al.*, 2020)	

identified EZH2 and BRD9 as promising therapeutic targets (Erkek *et al.*, 2019; Wang *et al.*, 2019). Moreover, aberrant epigenetic regulation is often causal of the malignant embryonic state of pediatric cancer cells (Lawlor and Thiele, 2012), potentially explaining the sensitivity of different embryonal tumors to drugs targeting epigenetic modifiers (Table 1). Treatment of *in vitro* pediatric tumor models with differentiation agents can recapitulate the effects achieved by driver reversal. However, a durable effect of differentiation therapy can only be acquired through induction of an irreversible growth arrest. As *in vivo* studies have shown, single agent treatment may not suffice to induce terminal differentiation and that combination therapy is required to do so (Hahn *et al.*, 2008; Botrugno *et al.*, 2009; Westerlund *et al.*, 2017; Chen *et al.*, 2018). A powerful tool to identify new (combinations of) drugs are high-throughput drug screens performed on *in vitro* tumor models. Organoids directly derived from patient tumor tissue could provide such models, as they have been shown to closely resemble its parental tissue (Drost and Clevers, 2018). Confirming their potential, an increasing number of reports described that tumor organoids are predictive for patient drug response (Tiriac *et al.*, 2018; Vlachogiannis *et al.*, 2018; Ganesh *et al.*, 2019; Ooft *et al.*, 2019; Yao *et al.*, 2020). Recently, the organoid technology was also successfully applied to several pediatric cancers, including embryonal tumors such as

MRT and Wilms tumors (Schutgens *et al.*, 2019; Calandrini *et al.*, 2020). The efficient establishment and cryopreservation of tumor organoid models from primary patient tissue allows for the generation of large patient cohorts stored in organoid biobanks. This is seemingly of particular interest for rare tumors, such as embryonal tumors, for which research material is scarce. In conclusion, the generation of novel and more representative *in vitro* embryonal tumor models is key for the improvement of differentiation therapeutics.

Discussion

In this review, we have attempted to outline the rapidly developing field of *in vitro* embryonal tumor models and discussed their added value to embryonal tumor research (**Figure 1** and **Table 1**). Still, each model has its intrinsic limitations. For instance, fetal cells can be extracted and cultured from fetal tissues (Xu *et al.*, 2014), but in many cases they do not represent the continuously changing cellular identities found during embryonic development. Alternatively, iPSCs or ESCs cell lines can be deployed to recapitulate these dynamics. Still, even though the spectrum of differentiation protocols is rapidly expanding, many embryonic cell types found *in vivo* cannot yet be captured *in vitro*. Additionally, *in vitro* cultures of ESCs or iPSCs have been shown to be susceptible to "spontaneous" malignant transformation, which can complicate the interpretation of modeling experiments (Ben-David and Benvenisty, 2011). Furthermore, 2D cultures do not capture 3D tissue architecture (Pampaloni *et al.*, 2007). These limitations have been to some extent improved in 3D organoid cultures, which better capture the cell-cell interactions found during embryonic organogenesis (Clevers, 2016). The development of mouse-human chimeras has highlighted the role of the microenvironment in tumor progression (Cohen *et al.*, 2020) and reveals a promising opportunity to bridge the gap of *in vitro* and *in vivo* tumor modeling, as mouse-human chimeras have the advantage of having human cells combined with an *in vivo* murine microenvironment. A good representation of patient tumor evolution remains challenging in *in vitro* models. In patients, tumors originate from a single tumor-initiating cell, wherein a genetic driver event induces aberrant signaling pathways that provide a cell with competitive advantages. Continuous selection of such cells (clonal selection) is thought to form the basis of tumor initiation, progression, and heterogeneity (Navin *et al.*, 2010, 2011). *In vitro* models typically do not reflect the environmental conditions causing clonal selection, as culture conditions are only a simplified version of *in vivo* signaling complexity. Embryonal tumors maintain a fetal identity, which is no longer present in matured tissues (Orbach *et al.*, 2013; Marshall *et al.*, 2014). The characterization of developmental programs in embryonal tumors can therefore give crucial insights into the processes underpinning malignant growth. Single cell transcriptome profiling of tumors and developing tissues has proven to be a promising tool to reveal such processes, which could potentially serve as therapeutic targets (Filbin *et al.*, 2018; Zhang *et al.*, 2019). Similar methods can also be applied to *in vitro* models recapitulating embryonal tumorigenesis, as demonstrated for the

retinoblastoma organoid model generated by Liu H. *et al.* (2020), which has the advantage that it allows for a direct comparison of normal and tumor development.

Although many *in vitro* embryonic cell-derived tumor models have been established over the years, the spectrum is biased toward ectoderm-derived tumors. It seems a matter of time before mesoderm- or endoderm-derived *in vitro* tumor models (e.g., Wilms tumor and hepatoblastoma) will be developed, as the number of culture systems for fetal tissues is rapidly expanding (Low *et al.*, 2019; Ooms *et al.*, 2020; Hendriks *et al.*, 2021).

We are only just beginning to understand the complexity of embryonal tumor development. Although capturing this complexity in a single *in vitro* model might not be feasible, further development of representative *in vitro* cell models recapitulating at least part of it is crucial to gain further insight into the fundamental processes underpinning malignant growth and the development of new therapeutic strategies.

Author Contributions

LC, IP, and JD wrote the manuscript. JD supervised the work. All authors approved the manuscript for publication.

Funding

We are grateful for support from the European Research Council (ERC) starting Grant 850571 (JD), the Dutch Cancer Society (KWF)/Alpe d'HuZes Bas Mulder Award to JD (KWF/Alpe d'HuZes, 10218), and the Children Cancer-free Foundation (KiKa #338, LC; base funding, IP).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

ASC, adult stem cell; DIPG, diffuse intrinsic pontine glioma; ESC, embryonic stem cell; GCT, germ cell tumor; GEMM, genetically engineered mouse model; H3K27M, histone 3 methionine for lysine 27 substitution; iPSC, induced pluripotent stem cell; MRT, malignant rhabdoid tumor; NPC, neural progenitor cell; NCC, neural crest cell; RPC, retinal precursor cell; SWI/SNF, SWItch/Sucrose Non-Fermentable.

References

Anastas, J. N., Zee, B. M., Kalin, J. H., Kim, M., Guo, R., Alexandrescu, S., *et al.* (2019). Re-programing chromatin with a bifunctional LSD1/HDAC inhibitor induces therapeutic differentiation in DIPG. *Cancer Cell* 36, 528–544. e10. doi: 10.1016/j.ccell.2019.09.005

Ballabio, C., Anderle, M., Gianesello, M., Lago, C., Miele, E., Cardano, M., *et al.* (2020). Modeling medulloblastoma *in vivo* and with human cerebellar organoids. *Nat. Commun.* 11:583. doi: 10.1038/s41467-019-13989-3

Bandopadhayay, P., Piccioni, F., O'Rourke, R., Ho, P., Gonzalez, E. M., Buchan, G., *et al.* (2019). Neuronal differentiation and cell-cycle programs mediate response to BET-bromodomain inhibition in MYC-driven medulloblastoma. *Nat. Commun.* 10:2400. doi: 10.1038/s41467-019-10307-9

Ben-David, U., and Benvenisty, N. (2011). The tumorigenicity of human embryonic and induced pluripotent stem cells. *Nat. Rev. Cancer* 11, 268–277. doi: 10.1038/nrc3034

Berry, R. L., Ozdemir, D. D., Aronow, B., Lindström, N. O., Dudnakova, T., Thornburn, A., *et al.* (2015). Deducing the stage of origin of Wilms' tumours from a developmental series of Wt1-mutant mice. *DMM Dis. Models Mechan.* 8, 903–917. doi: 10.1242/dmm.018523

Betters, E., Liu, Y., Kjaeldgaard, A., Sundström, E., and García-Castro, M. I. (2010). Analysis of early human neural crest development. *Dev. Biol.* 344, 578–592. doi: 10.1016/j.ydbio.2010.05.012

Betz, B. L., Strobeck, M. W., Reisman, D. N., Knudsen, E. S., and Weissman, B. E. (2002). Re-expression of hSNF5/INI1/BAF47 in pediatric tumor cells leads to G1 arrest associated with induction of p16ink4a and activation of RB. *Oncogene* 21, 5193–5203. doi: 10.1038/sj.onc.1205706

Bian, S., Repic, M., Guo, Z., Kavirayani, A., Burkard, T., Bagley, J. A., *et al.* (2018). Genetically engineered cerebral organoids model brain tumor formation. *Nat. Methods* 15, 631–639. doi: 10.1038/s41592-018-0070-7

Blakeley, P., Fogarty, N. M. E., Del Valle, I., Wamaitha, S. E., Hu, T. X., Elder, K., *et al.* (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development (Cambridge)* 142, 3151–3165. doi: 10.1242/dev.123547

Boeva, V., Louis-Brennetot, C., Peltier, A., Durand, S., Pierre-Eugène, C., Raynal, V., *et al.* (2017). Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* 49, 1408–1413. doi: 10.1038/ng.3921

Botrugno, O. A., Santoro, F., and Minucci, S. (2009). Histone deacetylase inhibitors as a new weapon in the arsenal of differentiation therapies of cancer. *Cancer Lett.* 280, 134–144. doi: 10.1016/j.canlet.2009.02.027

Calandrini, C., Schutgens, F., Oka, R., Margaritis, T., Candelli, T., Mathijsen, L., *et al.* (2020). An organoid biobank for childhood kidney cancers that captures disease and tissue heterogeneity. *Nat. Commun.* 11:1310. doi: 10.1038/s41467-020-15155-6

Cavalli, F. M. G., Remke, M., Rampasek, L., Peacock, J., Shih, D. J. H., Luu, B., *et al.* (2017). Intertumoral heterogeneity within medulloblastoma subgroups. *Cancer Cell* 31, 737–754.e6. doi: 10.1016/j.ccell.2017.05.005

Chen, L., Alexe, G., Dharia, N. V., Ross, L., Iniguez, A. B., Conway, A. S., *et al.* (2018). CRISPR-Cas9 screen reveals a MYCN-amplified neuroblastoma dependency on EZH2. *J. Clin. Invest.* 128, 446–462. doi: 10.1172/JCI90793

Chen, X., Pappo, A., and Dyer, M. A. (2015). Pediatric solid tumor genomics and developmental pliancy. *Oncogene* 34, 5207–5215. doi: 10.1038/onc.2014.474

Cheng, Y., Liao, S., Xu, G., Hu, J., Guo, D., Du, F., *et al.* (2020). NeuroD1 dictates tumor cell differentiation in medulloblastoma. *Cell Rep.* 31:107782. doi: 10.1016/j.celrep.2020.107782

Clevers, H. (2016). Modeling development and disease with organoids. *Cell* 165, 1586–1597. doi: 10.1016/j.cell.2016.05.082

Cohen, M. A., Zhang, S., Sengupta, S., Ma, H., Bell, G. W., Horton, B., *et al.* (2020). Formation of human neuroblastoma in mouse-human neural crest chimeras. *Cell Stem Cell* 26, 579–592.e6. doi: 10.1016/j.stem.2020.02.001

Cunningham, R. M., Walton, M. A., and Carter, P. M. (2018). The major causes of death in children and adolescents in the United States. *N. Eng. J. Med.* 379, 2468–2475. doi: 10.1056/nejmsr1804754

Dannenberg, J. H., Schuijff, L., Dekker, M., Van Der Valk, M., and Te Riele, H. (2004). Tissue-specific tumor suppressor activity of retinoblastoma gene homologs p107 and p130. *Genes Dev.* 18, 2952–2962. doi: 10.1101/gad.322004

De Jong, J., and Looijenga, L. H. J. (2006). Stem cell marker OCT3/4 in tumor biology and germ cell tumor diagnostics: history and future. *Crit. Rev. Oncogen.* 12, 171–203. doi: 10.1615/CritRevOncog.v12.i3-4.10

Dimaras, H., Corson, T. W., Cobrinik, D., White, A., Zhao, J., Munier, F. L., *et al.* (2015). Retinoblastoma. *Nat. Rev. Dis. Primers* 1:1502. doi: 10.1038/nrdp.2015.21

Drost, J., and Clevers, H. (2018). Organoids in cancer research. *Nat. Rev. Cancer* 18, 407–418. doi: 10.1038/s41568-018-0007-6

Dunbar, C. E., High, K. A., Joung, J. K., Kohn, D. B., Ozawa, K., and Sadelain, M. (2018). Gene therapy comes of age. *Science* 359:eaan4672. doi: 10.1126/science.aan4672

Erkek, S., Johann, P. D., Finetti, M. A., Drosos, Y., Chou, H. C., Zapatka, M., *et al.* (2019). Comprehensive analysis of chromatin states in atypical teratoid/rhabdoid tumor identifies diverging roles for SWI/SNF and polycomb in gene regulation. *Cancer Cell* 35, 95–110.e8. doi: 10.1016/j.ccell.2018.11.014

Filbin, M. G., Tirosh, I., Hovestadt, V., Shaw, M. L., Escalante, L. E., Mathewson, N. D., *et al.* (2018). Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* 360, 331–335. doi: 10.1126/science.aao4750

Frumm, S. M., Fan, Z. P., Ross, K. N., Duvall, J. R., Gupta, S., Verplank, L., *et al.* (2013). Selective HDAC1/HDAC2 inhibitors induce neuroblastoma differentiation. *Chem. Biol.* 20, 713–725. doi: 10.1016/j.chembiol.2013.03.020

Funato, K., Major, T., Lewis, P. W., Allis, C. D., and Tabar, V. (2014). Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation. *Science* 346, 1529–1533. doi: 10.1126/science.1253799

Ganesh, K., Wu, C., O'Rourke, K. P., Szeglin, B. C., Zheng, Y., Sauvé, C. E. G., *et al.* (2019). A rectal cancer organoid platform to study individual responses to chemoradiation. *Nat. Med.* 25, 1607–1614. doi: 10.1038/s41591-019-0584-2

García-López, J., Wallace, K., Otero, J. H., Olsen, R., Wang, Y., and dong (2020). Large 1p36 deletions affecting arid1a locus facilitate mycn-driven oncogenesis in neuroblastoma. *Cell Rep.* 30, 454–464.e5. doi: 10.1016/j.celrep.2019.12.048

Gibson, P., Tong, Y., Robinson, G., Thompson, M. C., Currle, D. S., Eden, C., *et al.* (2010). Subtypes of medulloblastoma have distinct developmental origins. *Nature* 468, 1095–1099. doi: 10.1038/nature09587

Grammel, D., Warmuth-Metz, M., Von Bueren, A. O., Kool, M., Pietsch, T., Kretzschmar, H. A., *et al.* (2012). Sonic hedgehog-associated medullobla stoma arising from the cochlear nuclei of the brainstem. *Acta Neuropathol.* 123, 601–614. doi: 10.1007/s00401-012-0961-0

Gröbner, S. N., Worst, B. C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V. A., *et al.* (2018). The landscape of genomic alterations across childhood cancers. *Nature* 555, 321–327. doi: 10.1038/nature25480

Hahn, C. K., Ross, K. N., Warrington, I. M., Mazitschek, R., Kanegai, C. M., Wright, R. D., *et al.* (2008). Expression-based screening identifies the combination of histone deacetylase inhibitors and retinoids for neuroblastoma differentiation. *Proc. Natl. Acad. Sci. U S A.* 105, 9751–9756. doi: 10.1073/pnas.0710413105

Han, Z. Y., Richer, W., Fréneaux, P., Chauvin, C., Lucchesi, C., Guillemot, D., *et al.* (2016). The occurrence of intracranial rhabdoid tumours in mice depends on temporal control of Smarcb1 inactivation. *Nat. Commun.* 7:10421. doi: 10.1038/ncomms10421

Hasselblatt, M., Nagel, I., Oyen, F., Bartelheim, K., Russell, R. B., Schüller, U., *et al.* (2014). SMARCA4-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. *Acta Neuropathol.* 128, 453–456. doi: 10.1007/s00401-014-1323-x

Hendriks, D., Artegiani, B., Hu, H., Chuva, de Sousa Lopes, S., and Clevers, H. (2021). Establishment of human fetal hepatocyte organoids and CRISPR–Cas9-based gene knockin and knockout in organoid cultures from human liver. *Nat. Protocols* 16, 182–217. doi: 10.1038/s41596-020-00411-2

Henriksen, J. R., Haug, B. H., Buechner, J., Tømte, E., Løkke, C., Flaegstad, T., *et al.* (2011). Conditional expression of retrovirally delivered anti-MYCN shRNA as an *in vitro* model system to study neuronal differentiation in MYCN-amplified neuroblastoma. *BMC Dev. Biol.* 11:1. doi: 10.1186/1471-213X-11-1

Hoeman, C. M., Cordero, F. J., Hu, G., Misuraca, K., Romero, M. M., Cardona, H. J., *et al.* (2019). ACVR1 R206H cooperates with H3.1K27M in promoting diffuse intrinsic pontine glioma pathogenesis. *Nat. Commun.* 10:1023. doi: 10.1038/s41467-019-08823-9

Hossain, M. M., Banik, N. L., and Ray, S. K. (2013). N-Myc knockdown and apigenin treatment controlled growth of malignant neuroblastoma cells having N-Myc amplification. *Gene* 529, 27–36. doi: 10.1016/j.gene.2013.07.094

Hovestadt, V., Smith, K. S., Bihannic, L., Filbin, M. G., Shaw, M. K. L., Baumgartner, A., *et al.* (2019). Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* 572, 74–79. doi: 10.1038/s41586-019-1434-6

Hu, Q., Gao, F., Tian, W., Ruteshouser, E. C., Wang, Y., Lazar, A., *et al.* (2011). Wt1 ablation and Igf2 upregulation in mice result in wilms tumors with elevated ERK1/2 phosphorylation. *J. Clin. Invest.* 121, 174–183. doi: 10.1172/JCI43772

Huang, L., Mokkapati, S., Hu, Q., Ruteshouser, E. C., Hicks, M. J., and Huff, V. (2016). Nephron Progenitor but not stromal progenitor cells give rise to wilms tumors in mouse models with β-Catenin activation or Wt1 ablation and Igf2 upregulation. *Neoplasia (United States)* 18, 71–81. doi: 10.1016/j.neo.2015.12.001

Huang, M., Tailor, J., Zhen, Q., Gillmor, A. H., Miller, M. L., Weishaupt, H., *et al.* (2019). Engineering genetic predisposition in human neuroepithelial stem cells recapitulates medulloblastoma tumorigenesis. *Cell Stem Cell* 25, 433–446.e7. doi: 10.1016/j.stem.2019.05.013

Jessa, S., Blanchet-Cohen, A., Krug, B., Vladoiu, M., Coutelier, M., Faury, D., *et al.* (2019). Stalled developmental programs at the root of pediatric brain tumors. *Nat. Genet.* 51, 1702–1713. doi: 10.1038/s41588-019-0531-7

Jiang, R., Xue, S., and Jin, Z. (2011). Stable knockdown of MYCN by lentivirus-based RNAi inhibits human neuroblastoma cells growth *in vitro* and *in vivo*. *Biochem. Biophys. Res. Commun.* 410, 364–370. doi: 10.1016/j.bbrc.2011.06.020

Johnsen, J. I., Dyberg, C., and Wickström, M. (2019). Neuroblastoma—a neural crest derived embryonal malignancy. *Front. Mol. Neurosci.* 12:9. doi: 10.3389/fnmol.2019.00009

Kang, J. H., Rychahou, P. G., Ishola, T. A., Qiao, J., Evers, B. M., and Chung, D. H. (2006). MYCN silencing induces differentiation and apoptosis in human neuroblastoma cells. *Biochem. Biophys. Res. Commun.* 351, 192–197. doi: 10.1016/j.bbrc.2006.10.020

Kattner, P., Strobel, H., Khoshnevis, N., Grunert, M., Bartholomae, S., Pruss, M., *et al.* (2019). Compare and contrast: pediatric cancer versus adult malignancies. *Cancer Metastasis Rev.* 38, 673–682. doi: 10.1007/s10555-019-09836-y

Khuong-Quang, D. A., Buczkowicz, P., Rakopoulos, P., Liu, X. Y., Fontebasso, A. M., Bouffet, E., *et al.* (2012). K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta Neuropathol.* 124, 439–447. doi: 10.1007/s00401-012-0998-0

Knutson, S. K., Warholic, N. M., Wigle, T. J., Klaus, C. R., Allain, C. J., Raimondi, A., *et al.* (2013). Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc. Natl. Acad. Sci. U S A.* 110, 7922–7927. doi: 10.1073/pnas.1303800110

Kreidberg, J. A., Sariola, H., Loring, J. M., Maeda, M., Pelletier, J., Housman, D., *et al.* (1993). WT-1 is required for early kidney development. *Cell* 74, 679–691. doi: 10.1016/0092-8674(93)90515-R

Kuwahara, A., Ozone, C., Nakano, T., Saito, K., Eiraku, M., and Sasai, Y. (2015). Generation of a ciliary margin-like stem cell niche from self-organizing human retinal tissue. *Nat. Commun.* 6:6286. doi: 10.1038/ncomms7286

Langer, L. F., Ward, J. M., and Archer, T. K. (2019). Tumor suppressor SMARCB1 suppresses super-enhancers to govern hESC lineage determination. *eLife* 8:e45672. doi: 10.7554/eLife.45672

Lawlor, E. R., and Thiele, C. J. (2012). Epigenetic changes in pediatric solid tumors: promising new targets. *Clin. Cancer Res.* 18, 2768–2779. doi: 10.1158/1078-0432.CCR-11-1921

Lee, A. S., Tang, C., Rao, M. S., Weissman, I. L., and Wu, J. C. (2013). Tumorigenicity as a clinical hurdle for pluripotent stem cell therapies. *Nat. Med.* 19, 998–1004. doi: 10.1038/nm.3267

Lee, E. Y. H. P., Chang, C. Y., Hu, N., Wang, Y. C. J., Lai, C. C., Herrup, K., *et al.* (1992). Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis. *Nature* 359, 288–294. doi: 10.1038/359288a0

Lee, R. S., Stewart, C., Carter, S. L., Ambrogio, L., Cibulskis, K., Sougnez, C., *et al.* (2012). A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. *J. Clin. Invest.* 122, 2983–2988. doi: 10.1172/JCI64400

Liu, G., David, B. T., Trawczynski, M., and Fessler, R. G. (2020). Advances in pluripotent stem cells: history, mechanisms, technologies, and applications. *Stem Cell Rev. Rep.* 16, 3–32. doi: 10.1007/s12015-019-09935-x

Liu, H., Sun, Q., Sun, Y., Zhang, J., Yuan, H., Pang, S., *et al.* (2017). MELK and EZH2 cooperate to regulate medulloblastoma cancer stem-like cell proliferation and differentiation. *Mol. Cancer Res.* 15, 1275–1286. doi: 10.1158/1541-7786.MCR-17-0105

Liu, H., Zhang, Y., Zhang, Y.-Y., Li, Y.-P., Hua, Z.-Q., Zhang, C.-J., *et al.* (2020). Human embryonic stem cell-derived organoid retinoblastoma reveals a cancerous origin. *Proc. Natl. Acad. Sci.* 117, 33628–33638. doi: 10.1073/pnas.2011780117

Lone, A. M., Dar, N. J., Hamid, A., Shah, W. A., Ahmad, M., and Bhat, B. A. (2016). Promise of retinoic acid-triazolyl derivatives in promoting differentiation of neuroblastoma cells. *ACS Chem. Neurosci.* 7, 82–89. doi: 10.1021/acschemneuro.5b00267

Low, J. H., Li, P., Chew, E. G. Y., Zhou, B., Suzuki, K., Zhang, T., *et al.* (2019). Generation of human PSC-Derived kidney organoids with patterned nephron segments and a *de novo* vascular network. *Cell Stem Cell* 25, 373–387.e9. doi: 10.1016/j.stem.2019.06.009

Luo, C., Lancaster, M. A., Castanon, R., Nery, J. R., Knoblich, J. A., and Ecker, J. R. (2016). Cerebral organoids recapitulate epigenomic signatures of the human fetal brain. *Cell Rep.* 17, 3369–3384. doi: 10.1016/j.celrep.2016.12.001

MacPherson, D., Sage, J., Kim, T., Ho, D., McLaughlin, M. E., and Jacks, T. (2004). Cell type-specific effects of Rb deletion in the murine retina. *Genes Dev.* 18, 1681–1694. doi: 10.1101/gad.1203304

Marshall, G. M., Carter, D. R., Cheung, B. B., Liu, T., Mateos, M. K., Meyerowitz, J. G., *et al.* (2014). The prenatal origins of cancer. *Nat. Rev. Cancer* 14, 277–289. doi: 10.1038/nrc3679

Maurer, J., Fuchs, S., Jäger, R., Kurz, B., Sommer, L., and Schorle, H. (2007). Establishment and controlled differentiation of neural crest stem cell lines using conditional transgenesis. *Differentiation* 75, 580–591. doi: 10.1111/j.1432-0436.2007.00164.x

Mohammad, F., Weissmann, S., Leblanc, B., Pandey, D. P., Højfeldt, J. W., Comet, I., *et al.* (2017). EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat. Med.* 23, 483–492. doi: 10.1038/nm.4293

Montavon, G., Jauquier, N., Coulon, A., Peuchmaur, M., Flahaut, M., Bourloud, K. B., *et al.* (2014). Wild-type ALK and activating ALK-R1275Q and ALK-F1174L mutations upregulate Myc and initiate tumor formation in murine neural crest progenitor cells. *Oncotarget* 5, 4452–4466. doi: 10.18632/oncotarget.2036

Muguruma, K., Nishiyama, A., Kawakami, H., Hashimoto, K., and Sasai, Y. (2015). Self-organization of polarized cerebellar tissue in 3D culture of human pluripotent stem cells. *Cell Rep.* 10, 537–550. doi: 10.1016/j.celrep.2014.12.051

Muscat, A., Popovski, D., Jayasekara, W. S. N., Rossello, F. J., Ferguson, M., Marini, K. D., *et al.* (2016). Low-dose histone deacetylase inhibitor treatment leads to tumor growth arrest and multi-lineage differentiation of malignant rhabdoid tumors. *Clin. Cancer Res.* 22, 3560–3570. doi: 10.1158/1078-0432.CCR-15-2260

Nakayama, R. T., Pulice, J. L., Valencia, A. M., McBride, M. J., McKenzie, Z. M., Gillespie, M. A., *et al.* (2017). SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. *Nat. Genet.* 49, 1613–1623. doi: 10.1038/ng.3958

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., *et al.* (2011). Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–95. doi: 10.1038/nature09807

Navin, N., Krasnitz, A., Rodgers, L., Cook, K., Meth, J., Kendall, J., *et al.* (2010). Inferring tumor progression from genomic heterogeneity. *Genome Res.* 20, 68–80. doi: 10.1101/gr.099622.109

Ocasio, J., Babcock, B., Malawsky, D., Weir, S. J., Loo, L., Simon, J. M., *et al.* (2019). scRNA-seq in medulloblastoma shows cellular heterogeneity and lineage expansion support resistance to SHH inhibitor therapy. *Nat. Commun.* 10:5829. doi: 10.1038/s41467-019-13657-6

Olsen, R. R., Otero, J. H., García-López, J., Wallace, K., Finkelstein, D., Rehg, J. E., *et al.* (2017). MYCN induces neuroblastoma in primary neural crest cells. *Oncogene* 36, 5075–5082. doi: 10.1038/onc.2017.128

Ooft, S. N., Weeber, F., Dijkstra, K. K., McLean, C. M., Kaing, S., van Werkhoven, E., *et al.* (2019). Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. *Sci. Trans. Med.* 11:eaay2574. doi: 10.1126/scitranslmed.aay2574

Ooms, A. H. A. G., Calandrini, C., de Krijger, R. R., and Drost, J. (2020). Organoid models of childhood kidney tumours. *Nat. Rev. Urol.* 17, 311–313. doi: 10.1038/s41585-020-0315-y

Oosterhuis, J. W., and Looijenga, L. H. J. (2019). Human germ cell tumours from a developmental perspective. *Nat. Rev. Cancer* 19, 522–537. doi: 10.1038/s41568-019-0178-9

O'Rahilly, R., and Müller, F. (2007). The development of the neural crest in the human. *J. Anatomy* 211, 335–351. doi: 10.1111/j.1469-7580.2007.00773.x

Orbach, D., Sarnacki, S., Brisse, H. J., Gauthier-Villars, M., Jarreau, P. H., Tsatsaris, V., *et al.* (2013). Neonatal cancer. *Lancet Oncol.* 14, e609–e620. doi: 10.1016/S1470-2045(13)70236-5

Pampaloni, F., Reynaud, E. G., and Stelzer, E. H. K. (2007). The third dimension bridges the gap between cell culture and live tissue. *Nat. Rev. Mol. Cell Biol.* 8, 839–845. doi: 10.1038/nrm2236

Parisian, A. D., Koga, T., Miki, S., Johann, P. D., Kool, M., Crawford, J. R., *et al.* (2020). SMARCB1 loss interacts with neuronal differentiation state to block maturation and impact cell stability. *Genes Dev.* 34, 1316–1329. doi: 10.1101/gad.339978.120

Patties, I., Kortmann, R. D., Menzel, F., and Glasow, A. (2016). Enhanced inhibition of clonogenic survival of human medulloblastoma cells by multimodal treatment with ionizing irradiation, epigenetic modifiers, and differentiation-inducing drugs. *J. Exp. Clin. Cancer Res.* 35:1. doi: 10.1186/s13046-016-0376-1

Puisieux, A., Pommier, R. M., Morel, A. P., and Lavial, F. (2018). Cellular pliancy and the multistep process of tumorigenesis. *Cancer Cell* 33, 164–172. doi: 10.1016/j.ccell.2018.01.007

Rahal, Z., Abdulhai, F., Kadara, H., and Saab, R. (2018). Genomics of adult and pediatric solid tumors. *Am. J. Cancer Res.* 8, 1356–1386.

Reiff, T., Tsarovina, K., Majdazari, A., Schmidt, M., Del Pino, I., and Rohrer, H. (2010). Neuroblastoma Phox2b variants stimulate proliferation and dedifferentiation of immature sympathetic neurons. *J. Neurosci.* 30, 905–915. doi: 10.1523/JNEUROSCI.5368-09.2010

Robanus-Maandag, E., Dekker, M., Van Der Valk, M., Carrozza, M. L., Jeanny, J. C., Dannenberg, J. H., *et al.* (1998). p107 is a suppressor of retinoblastoma development in pRB-deficient mice. *Genes Dev.* 12, 1599–1609. doi: 10.1101/gad.12.11.1599

Schulte, J. H., Lindner, S., Bohrer, A., Maurer, J., De Preter, K., Lefever, S., *et al.* (2013). MYCN and ALKF1174L are sufficient to drive neuroblastoma development from neural crest progenitor cells. *Oncogene* 32, 1059–1065. doi: 10.1038/onc.2012.106

Schutgens, F., Rookmaaker, M. B., Margaritis, T., Rios, A., Ammerlaan, C., Jansen, J., *et al.* (2019). Tubuloids derived from human adult kidney and urine for personalized disease modeling. *Nat. Biotechnol.* 37, 303–313. doi: 10.1038/s41587-019-0048-8

Schwartzentruber, J., Korshunov, A., Liu, X. Y., Jones, D. T. W., Pfaff, E., Jacob, K., *et al.* (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* 482, 226–231. doi: 10.1038/nature10833

Shain, A. H., and Pollack, J. R. (2013). The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS One* 8:e55119. doi: 10.1371/journal.pone.0055119

Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer statistics, 2016. *CA: Cancer J. Clin.* 66, 7–30. doi: 10.3322/caac.21332

Stemple, D. L., and Anderson, D. J. (1992). Isolation of a stem cell for neurons and glia from the mammalian neural crest. *Cell* 71, 973–985. doi: 10.1016/0092-8674(92)90393-Q

Stratton, M. R., Campbell, P. J., and Futreal, P. A. (2009). The cancer genome. *Nature* 458, 719–724. doi: 10.1038/nature07943

Sun, Y., Xu, C., Pan, C., Chen, X., Geng, Y., Wu, Y., *et al.* (2019). Diffuse intrinsic pontine gliomas exhibit cell biological and molecular signatures of fetal hindbrain-derived neural progenitor cells. *Neurosci. Bull.* 35, 216–224. doi: 10.1007/s12264-018-00329-6

Terada, Y., Jo, N., Arakawa, Y., Sakakura, M., Yamada, Y., Ukai, T., *et al.* (2019). Human pluripotent stem cell-derived tumor model uncovers the embryonic stem cell signature as a key driver in atypical teratoid/rhabdoid tumor. *Cell Rep.* 26, 2608–2621.e6. doi: 10.1016/j.celrep.2019.02.009

Theunissen, T. W., and Jaenisch, R. (2017). Mechanisms of gene regulation in human embryos and pluripotent stem cells. *Development (Cambridge)* 144, 4496–4509. doi: 10.1242/dev.157404

Thomas, J., Adegboyega, P., Iloabachie, K., Mooring, J. W., and Lian, T. (2011). Sinonasal teratocarcinosarcoma with yolk sac elements: a neoplasm of somatic or germ cell origin? *Annals Diagn. Pathol.* 15, 135–139. doi: 10.1016/j.anndiagpath.2010.01.004

Tiriac, H., Belleau, P., Engle, D. D., Plenker, D., Deschênes, A., Somerville, T. D. D., *et al.* (2018). Organoid profiling identifies common responders to chemotherapy in pancreatic cancer. *Cancer Discovery* 8, 1112–1129. doi: 10.1158/2159-8290.CD-18-0349

Visvader, J. E. (2011). Cells of origin in cancer. *Nature* 469, 314–322. doi: 10.1038/nature09781

Vitte, J., Gao, F., Coppola, G., Judkins, A. R., and Giovannini, M. (2017). Timing of Smarcb1 and Nf2 inactivation determines schwannoma versus rhabdoid tumor development. *Nat. Commun.* 8:300. doi: 10.1038/s41467-017-00346-5

Vlachogiannis, G., Hedayat, S., Vatsiou, A., Jamin, Y., Fernández-Mateos, J., Khan, K., *et al.* (2018). Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* 359, 920–926. doi: 10.1126/science.aao2774

Vladoiu, M. C., El-Hamamy, I., Donovan, L. K., Farooq, H., Holgado, B. L., Sundaravadanam, Y., *et al.* (2019). Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* 572, 67–73. doi: 10.1038/s41586-019-1158-7

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., and Kinzler, K. W. (2013). Cancer genome landscapes. *Science* 340, 1546–1558. doi: 10.1126/science.1235122

Wang, X., Lee, R. S., Alver, B. H., Haswell, J. R., Wang, S., Mieczkowski, J., *et al.* (2017). SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. *Nat. Genet.* 49, 289–295. doi: 10.1038/ng.3746

Wang, X., Wang, S., Troisi, E. C., Howard, T. P., Haswell, J. R., Wolf, B. K., *et al.* (2019). BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. *Nat. Commun.* 10:1881. doi: 10.1038/s41467-019-09891-7

Westerlund, I., Shi, Y., Toskas, K., Fell, S. M., Li, S., Surova, O., *et al.* (2017). Combined epigenetic and differentiation-based treatment inhibits neuroblastoma tumor growth and links HIF2α to tumor suppression. *Proc. Natl. Acad. Sci. U S A.* 114, E6137–E6146. doi: 10.1073/pnas.1700655114

Westermark, U. K., Wilhelm, M., Frenzel, A., and Henriksson, M. A. (2011). The MYCN oncogene and differentiation in neuroblastoma. *Sem. Cancer Biol.* 21, 256–266. doi: 10.1016/j.semcancer.2011.08.001

Wikenheiser-Brokamp, K. A. (2006). Retinoblastoma family proteins: insights gained through genetic manipulation of mice. *Cell. Mol. Life Sci.* 63, 767–780. doi: 10.1007/s00018-005-5487-3

Wilson, B. G., and Roberts, C. W. M. (2011). SWI/SNF nucleosome remodellers and cancer. *Nat. Rev. Cancer* 11, 481–492. doi: 10.1038/nrc3068

Wu, G., Broniscer, A., McEachron, T. A., Lu, C., Paugh, B. S., Becksfort, J., *et al.* (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat. Genet.* 44, 251–253. doi: 10.1038/ng.1102

Xu, X. L., Singh, H. P., Wang, L., Qi, D. L., Poulos, B. K., Abramson, D. H., *et al.* (2014). Rb suppresses human cone-precursor-derived retinoblastoma tumours. *Nature* 514, 385–388. doi: 10.1038/nature13813

Yao, Y., Xu, X., Yang, L., Zhu, J., Wan, J., Shen, L., *et al.* (2020). Patient-Derived organoids predict chemoradiation responses of locally advanced rectal cancer. *Cell Stem Cell* 26, 17–26.e6. doi: 10.1016/j.stem.2019.10.010

Young, M. D., Mitchell, T. J., Custers, L., Margaritis, T., Morales, F., Kwakwa, K., *et al.* (2020). Single cell derived mRNA signals across human kidney tumors. *BioRxiv* [preprint] doi: 10.1101/2020.03.19.998815

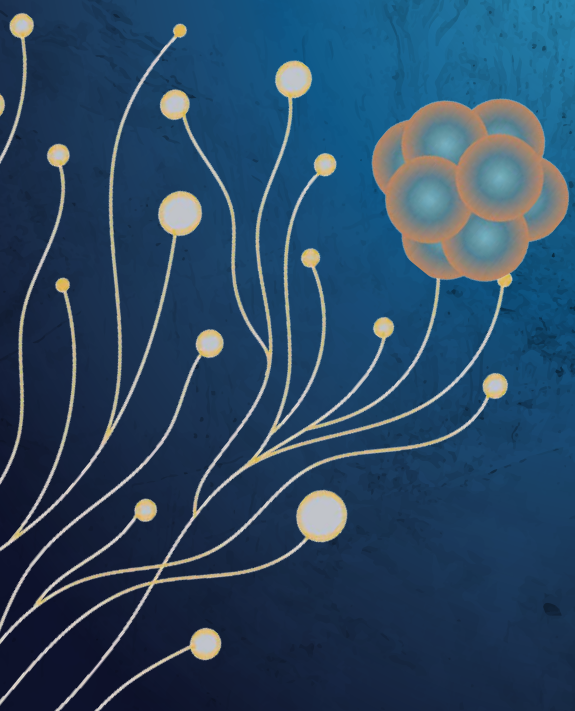
Young, M. D., Mitchell, T. J., Vieira Braga, F. A., Tran, M. G. B., Stewart, B. J., Ferdinand, J. R., *et al.* (2018). Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* 361, 594–599. doi: 10.1126/science.aat1699

Zagozewski, J., Shahriary, G. M., Morrison, L. C., Saulnier, O., Stromecki, M., Fresnoza, A., *et al.* (2020). An OTX2-PAX3 signaling axis regulates Group 3 medulloblastoma cell fate. *Nat. Commun.* 11:3627. doi: 10.1038/s41467-020-17357-4

Zhang, L., He, X., Liu, X., Zhang, F., Huang, L. F., Potter, A. S., *et al.* (2019). Single-Cell transcriptomics in medulloblastoma reveals tumor-initiating progenitors and oncogenic cascades during tumorigenesis and relapse. *Cancer Cell* 36, 302–318.e7. doi: 10.1016/j.ccell.2019.07.009

Zheng, C., Schneider, J. W., and Hsieh, J. (2020). Role of RB1 in human embryonic stem cell-derived retinal organoids. *Dev. Biol.* 462, 197–207. doi: 10.1016/j.ydbio.2020.03.011

Zhong, X., Gutierrez, C., Xue, T., Hampton, C., Vergara, M. N., Cao, L. H., *et al.* (2014). Generation of three-dimensional retinal tissue with functional photoreceptors from human iPSCs. *Nat. Commun.* 5:4047. doi: 10.1038/ncomms5047



CHAPTER 3

Single cell derived mRNA signals across human kidney tumors

Matthew D Young[†*1], Thomas J Mitchell[†1,2,3], Lars Custers[†4,5], Thanasis Margaritis[4], Francisco Morales-Rodriguez[4,5], Kwasi Kwakwa[1], Eleonora Khabirova[1], Gerda Kildisiute[1], Thomas RW Oliver[1,2], Ronald R de Krijger[4,6], Marry M van den Heuvel-Eibrink[4], Federico Comitani[7], Alice Piapi[8], Eva Bugallo-Blanco[8], Christine Thevanesan[8], Christina Burke[8], Elena Prigmore[1], Kirsty Ambridge[1], Kenny Roberts[1], Felipe A Vieira Braga[9], Tim HH Coorens[1], Ignacio Del Valle[8], Anna Wilbrey-Clark[1], Lira Mamanova[1], Grant D Stewart[2,3], Vincent J Gnanapragasam[2,3,10], Dyanne Rampling[11], Neil Sebire[12], Nicholas Coleman[2,13], Liz Hook[2,13], Anne Warren[2], Muzlifah Haniffa[1,14,15], Marcel Kool[4,16,17], Stefan M Pfister[16,17,18,] John C Achermann[8], Xiaoling He[19,20], Roger A Barker[19,20], Adam Shlien[7,21,22], Omer A Bayraktar[1], Sarah A Teichmann[1,23], Frank C Holstege[4], Kerstin B Meyer[1], Jarno Drost[4,5*], Karin Straathof[8,11*], Sam Behjati[1,2,24*]

***Nature Communications,* 23 June 2021**

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

[2]Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK.

[3]Department of Surgery, University of Cambridge, Cambridge, CB2 0QQ, UK.

[4]Princess Máxima Center for Pediatric Oncology, 3584 CS Utrecht, The Netherlands.

[5]Oncode Institute, 3584 CS Utrecht, The Netherlands.

[6]Department of Pathology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

[7]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada.

[8]UCL Great Ormond Street Hospital Institute of Child Health, London WC1N 1EH, UK.

[9]Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands.

[10]Cambridge Urology Translational Research and Clinical Trials office, Cambridge Biomedical Campus Cambridge CB2 0QQ University of Cambridge

[11]Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK.

[12]NIHR Great Ormond Street Hospital BRC and Institute of Child Health, London WC1N 3JH, UK.

[13]Department of Pathology, University of Cambridge, Cambridge, CB2 1QP, UK.

[14]Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, NE1 4LP, UK.

[15]Intitute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, NE1 4HH, UK.

[16]Hopp Children´s Cancer Center Heidelberg (KiTZ), 69120 Heidelberg Germany.

[17]German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Division of Pediatric Neurooncology, 69120 Heidelberg, Germany.

[18]Heidelberg University Hospital, Department of Pediatric Hematology and Oncology, 69120 Heidelberg, Germany.

[19]MRC-WT Cambridge Stem Cell Institute, University of Cambridge, CB2 0QQ, Cambridge, UK.

[20]Department of Clinical Neuroscience, University of Cambridge, Cambridge, CB2 0QQ, UK.

[21]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada.

[22]Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada.

[23]Cavendish Laboratory, University of Cambridge, JJ Thomson Ave, Cambridge CB3 0HE, UK.

[24]Department of Paediatrics, University of Cambridge, Cambridge, CB2 0QQ, UK.

†These authors contributed equally.

*These authors jointly supervised this work.

Abstract

Tumor cells may share some patterns of gene expression with their cell of origin, providing clues into the differentiation state and origin of cancer. Here, we study the differentiation state and cellular origin of 1300 childhood and adult kidney tumors. Using single cell mRNA reference maps of normal tissues, we quantify reference "cellular signals" in each tumor. Quantifying global differentiation, we find that childhood tumors exhibit fetal cellular signals, replacing the presumption of "fetalness" with a quantitative measure of immaturity. By contrast, in adult cancers our assessment refutes the suggestion of dedifferentiation towards a fetal state in most cases. We find an intimate connection between developmental mesenchymal populations and childhood renal tumors. We demonstrate the diagnostic potential of our approach with a case study of a cryptic renal tumor. Our findings provide a cellular definition of human renal tumors through an approach that is broadly applicable to human cancer.

Introduction

As cancer cells evolve from normal cells, they may retain patterns of messenger RNA (mRNA) characteristic of the cell of origin. In such cases, the cancer cell transcriptome may contain information that can identify the cancer cell of origin, its differentiation state, or trajectory towards a cancer cell. It is therefore conceivable that tumor transcriptomes can be used to identify the cells from which tumors arise and test fundamental hypotheses regarding tumor's differentiation states, such as the "fetalness" of childhood tumors or the dedifferentiation of adult tumors towards a fetal state.

Single cell transcriptomics allows for a direct quantitative comparison to be made between single tumor and relevant normal cell transcriptomes. For example, single cell transcriptomes identified that a specific subtype of proximal tubular cells are the normal cell correlate of clear cell renal cell carcinoma (ccRCC) cells[1]. Such experiments can also reveal more precise information about normal cells within the tumor microenvironment. However, the high resource requirements of single cell transcriptomics preclude investigations of large patient cohorts, which are required to study rare subtypes, test the generalizability of such signals and determine associations with clinical parameters. An alternative approach is to identify the presence of single cell derived mRNA signals in bulk tumor transcriptomes, utilizing normal single cell transcriptomes as a reference. Smaller numbers of single cancer cell experiments can then be used to validate cellular signals identified.

Tumor bulk transcriptomes for most types of human cancer have been generated in the context of cancer genomics efforts of recent years, such as those conducted by the *International Cancer Genome Consortium* (ICGC) and *The Cancer Genome Atlas* (TCGA)[2,3]. Single cell reference data, generated by efforts collectively known as

the *Human Cell Atlas*[4,5], have begun to provide quantitative transcriptional definitions of the normal cells that constitute the developing and mature human kidneys[1,6,7,8,9,10]. By combining these bulk tumor transcriptome databases with single cell reference data, we may therefore be able to identify single cell signals in bulk transcriptomes across large cohorts of kidney tumors.

Here, we study normal single cell mRNA signals in bulk kidney tumor transcriptomes ($n = 1258$; **Fig. 1A**, **Supplementary Data 1**) and validate our findings using targeted single cell experiments ($n = 10$, **Fig. 1A**, **Supplementary Table 1**). There are three central aims of our analyses. Firstly, we test the fundamental presumption that childhood renal tumors exhibit fetal cell signals whilst adult tumors dedifferentiate towards a fetal state. Next, we define for each tumor type its normal cell correlate which may represent its cell of origin and provide diagnostic cues. Finally, we explore the tumor micro-environment across different tumor types.

Results

An integrated single cell reference map of the kidney

The nephron is the functional unit of the kidney and together with its associated vasculature and support cells make up the majority of kidney cells. The nephron is derived from the mesoderm and forms from a combination of mesenchymal cell populations that mature into the epithelial cells of the nephron via mesenchymal to epithelial transition (MET)[11]. To precisely define these mesenchymal populations and the populations they mature into, we created a refined fetal kidney reference map combining previously generated[1,8] and newly generated human fetal kidney single cell data (**Fig. 1B, S1**).

This reference revealed 4 key mesenchymal populations: mesenchymal progenitor cells (MPCs), cap mesenchyme (CM), and two populations of specialized interstitial cells: smooth muscle-like cells (ICa), and cortical stromal cells (ICb) (**Fig. 1B, S1A–C**)[8,11]. The cap mesenchyme condenses on the ureteric bud and forms the tubular structures of the nephron via mesenchymal to epithelial transition. The mesenchymal cells which do not form cap mesenchyme and remain in the interstitial space form interstitial support cells for the nephron, such as mesangial cells. The final mesenchymal population, which we termed mesenchymal progenitor cells, was not present in sufficient numbers to be reported in earlier single cell transcriptomic studies of the developing kidney. These MPCs are enriched for early time points (**Fig. 1C**), strongly resemble mesenchymal cells in the fetal adrenal (**Supplementary Fig. 1D**)[12], and both populations resemble primitive mesodermal populations in the post gastrulation mouse embryo (**Supplementary Fig. 1D**)[13]. Developmentally, both the adrenal cortex and the kidney are derived from the same mesodermal lineage.

We combined this refined map of the developing kidney with previously generated maps of the mature kidney[1], the developing adrenal gland[12], and the post-gastrulation

mouse[13] (**Fig. 1A**). Together these provide a complete single cell reference map of the kidney across developmental time.

Quantification of reference cellular mRNA signals in bulk transcriptomes

Our single cell reference map of the kidney provides a cellular mRNA signal for each population of cells. To measure the abundance of these reference cellular signals in bulk tumor transcriptomes, we devised a method that fits raw bulk mRNA counts for the entire transcriptome—not just marker genes—to a weighted linear combination of transcriptomic signals derived from reference single cell data.

A number of bulk deconvolution tools exist that aim to identify the cellular composition of bulk tissues using a single cell reference[14,15,16]. However, the aim of our analysis was not to identify and quantify the number of cells present in the microenvironment, but to identify the major cellular signals (or transcriptional programs) used by tumor cells. As such, we do not expect any of our single cell reference populations to exactly match the tumor cells' transcriptome. We therefore designed our method to identify the major transcriptional signals (defined using single cell data) present in bulk transcriptomic data, with the expectation that the provided reference signals would not perfectly match the transcriptomes of the cells in the tumor. We term this approach "cellular signal analysis" to differentiate it from "deconvolution analysis", the inference of the cellular composition of bulk transcriptomes.

To validate our approach we applied cellular signal analysis and published deconvolution methods, MuSiC[15] and BSeq-SC[14] to human bulk transcriptomes of known origin: purified normal leucocytes, pre-B cell leukemia, and peripheral blood mononuclear cells[17,18]. For this comparison we used a reference that combined single cell transcriptomes from peripheral blood cells with a negative control population; proximal tubular kidney cells. As proximal tubular cells are completely absent from the source material for these bulk transcriptomes, an ideal method would not assign any contribution from this population.

We first considered those bulk transcriptomes which we expected to be well described by the provided reference (e.g., bulk B cell transcriptomes should be well modeled by a B cell signal). All three methods identified the correct cellular signal in most cases (**Fig. 1D, S2A**). However, MuSiC found a small but implausible renal tubular signal in most bulk transcriptomes, which was mostly absent from cellular signal analysis. This difference was even more pronounced for bulk transcriptomes where an adequate reference was not available (**Fig. 1E, S2B**). Here, MuSiC identified a substantial contribution from renal proximal tubular cells to pre-B cell leukemia, while BSeq-SC was unable to differentiate pre-B cell leukemia transcriptomes from normal mature B cells (**Supplementary Fig. 2B**). Cellular signal analysis identified pre-B cell leukemia as most similar to B cells, but with a substantial part of the signal unexplained by the given reference (**Fig. 1E, Supplementary Fig. 2B**).

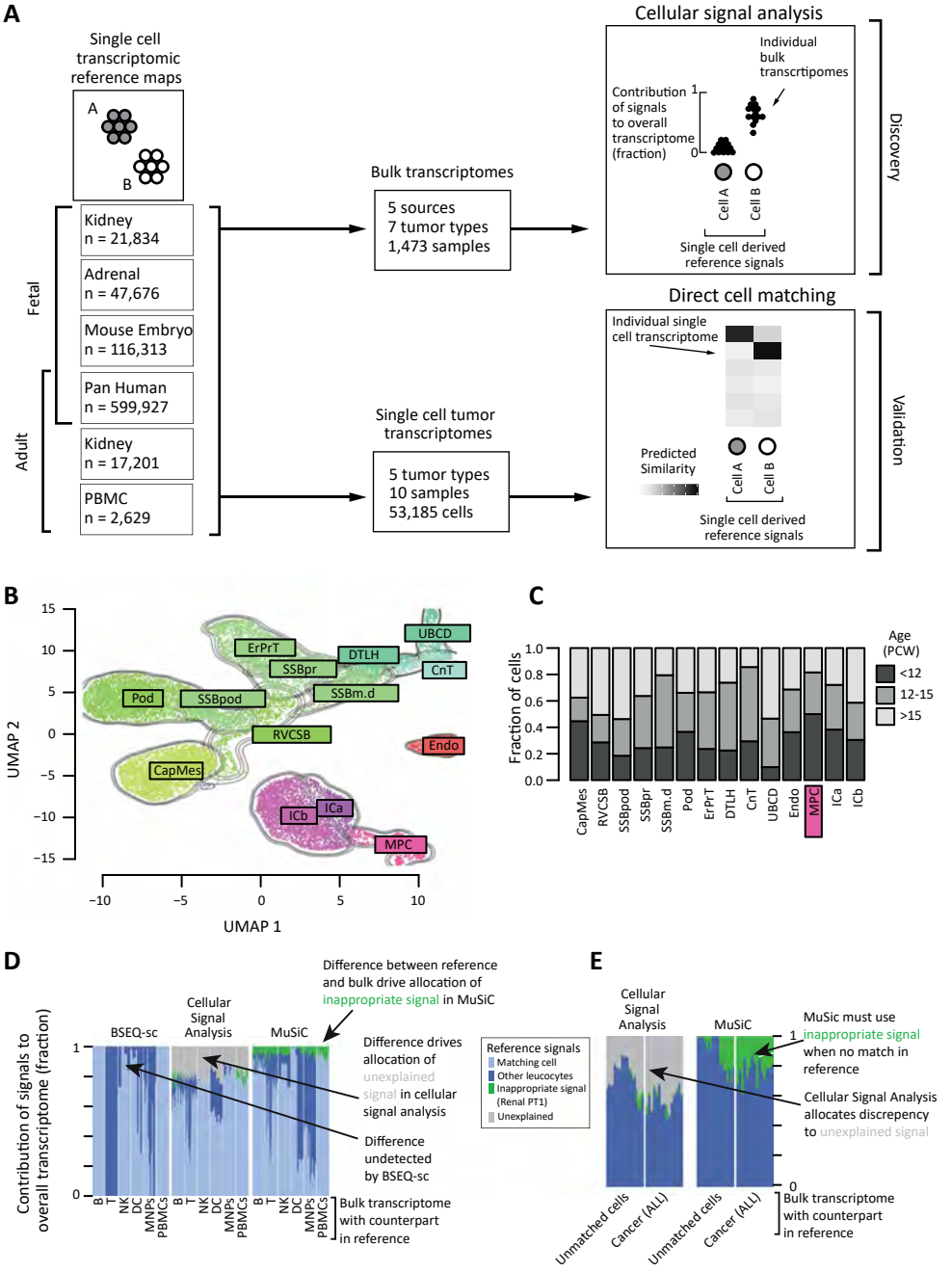


Fig. 1 | Methodology overview and validation.
(A) Overview of methodology: Single cell reference atlases (left) define cellular signals. These are used to calculate the contribution of each cellular signal to bulk transcriptomes (top, **Supplementary Data 1**), where signal contributions are normalized to give a score between 0 and 1 for each bulk

As a further test, we applied all three methods to 766 ccRCC transcriptomes from The Cancer Genome Atlas[19] to assess whether the known cellular identity of these cancer cells could be identified. Cellular signal analysis best identified the signal of a specific proximal tubular cell population as the predominant cell signal in ccRCC bulk cancer transcriptomes (**Supplementary Fig. 3**).

To test the methods on a more traditional "deconvolution" metric, we applied cellular signal analysis and MuSiC to 100 pseudo-bulk transcriptomes constructed from the reference single cell data. We then estimated how accurately the known number of cells of each type that was used to construct the pseudo-bulk transcriptome could be recovered (**Supplementary Fig. 2C, D**). This comparison found that while cellular signal analysis had reasonable accuracy, MuSiC was consistently the best performing method (**Supplementary Fig. 2C, D**). This highlights that the cost of the flexibility built into cellular signal analysis in accommodating unexplained signals is lower accuracy in determining the cellular composition of bulk transcriptomes.

Taken together, these comparisons demonstrate the need for a bespoke approach to identify the main cellular signals in bulk transcriptomes where the reference data is incomplete. Cellular signal analysis quantifies the inadequacy of the reference through the allocation of "unexplained signal". Mathematically, this "unexplained signal" represents an intercept term, included to limit the assignment of spurious signals when a bulk transcriptome differs from all signals in the reference (see Methods). As the reference becomes progressively less suited to the bulk transcriptome being modeled, the "unexplained signal" contribution becomes steadily larger (**Supplementary Fig. 2C**).

transcriptome, cellular signal pair (top right). These findings are validated by comparing the same cellular signals (left) to single cell tumor transcriptomes (bottom, **Supplementary Table 1**), where logistic regression generates a similarity score for each single cell transcriptome, cellular signal pair (bottom right). **(B)** Combined fetal kidney reference map: Contours and colors indicate the labeled cell type. *CapMes* Cap Mesenchyme, *RVCSB* Renal vesicle and comma-shaped body, *SSBpod* S-shaped body podocyte, *SSBpr* S-shaped body proximal tubules, *SSBm.d* S-shaped body medial and distal, *Pod* Podocytes, *ErPrT* Early proximal tubules, *DTLH* Distal tubule and loop of Henle, *UBCD* Ureteric Bud and collecting duct, *CnT* Connecting tubules, *Endo* Endothelium, *ICa* Interstitial cells a (smooth muscle), *ICb* Interstitial cells b (stromal), *MPC* Mesenchymal progenitor cells. **(C)** Age distribution of fetal kidney populations: Bar heights indicate fraction of cell type (as in B) by fetal age (color) in post conception weeks. **(D)** Benchmarking with match in provided reference: Comparison of two "bulk deconvolution" methods (BSEQ-sc and MuSiC) to cellular signal analysis, using bulk transcriptomes for which a good match exists in the reference single cell dataset. Bars height represent signal contributions from an immune cell and proximal tubular cell (PT1, included as a negative control) reference set in explaining bulk transcriptome from peripheral blood or flow sorted cells as indicated by the x-axis, bar color, and legend. "Matching cell" indicates a contribution from the expected signal (e.g., NK cell signal in NK bulk transcriptomes). See also Supplementary Fig. 2A. **(E)** Benchmarking with no match in provided reference: As in D, except bulk transcriptomes are flow sorted immune cells not in the reference (labeled "Unmatched cells") or pre-B cell acute lymphoblastic leukemias (labeled "Cancer ALL"). See also Supplementary Fig. 2B. Source data are available as a Source Data file.

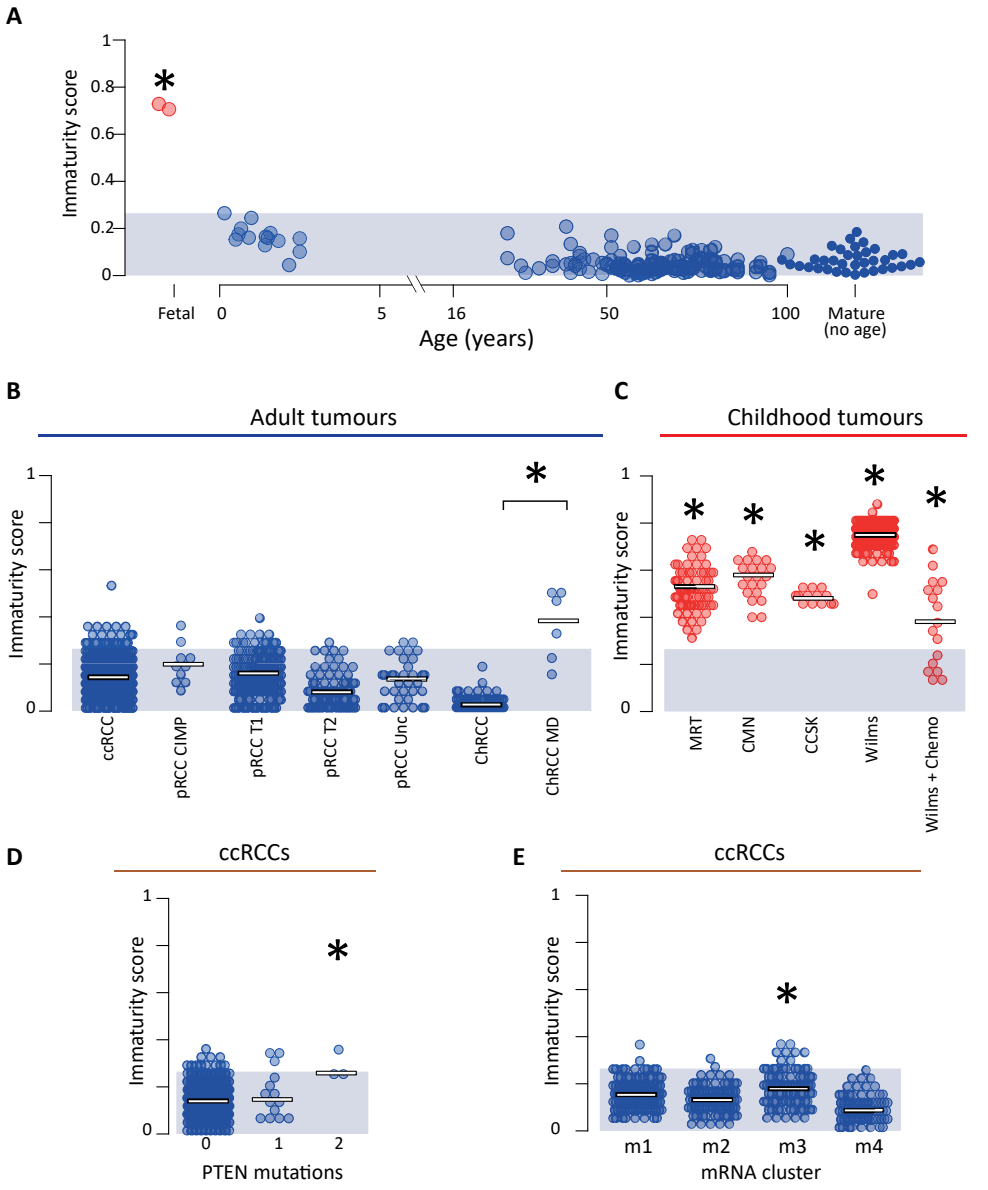


Figure 2 | Immaturity score.
(A) Normal kidney: From 201 bulk transcriptomes from normal kidney an immaturity score was calculated by fitting each bulk transcriptome using a combined mature and fetal kidney cellular signal reference. The immaturity score is the total normalized signal contribution from fetal kidney in each bulk transcriptome (y-axis). The x-axis shows sample age, with unknown age on the right and fetal samples on the left in red. The shaded blue area indicates the range of maturity scores across all normal post-natal transcriptomes. The star indicates that fetal samples have maturity scores significantly higher than normal samples (p=0.015, two-sided Wilcoxon rank sum test). **(B)** Adult renal tumors: Immaturity score (as in A) for 853 adult renal tumors, with normal immaturity score range shown by blue shading. The metabolically divergent subtype

Childhood tumors, but not adult tumors, exhibit a fetal transcriptome

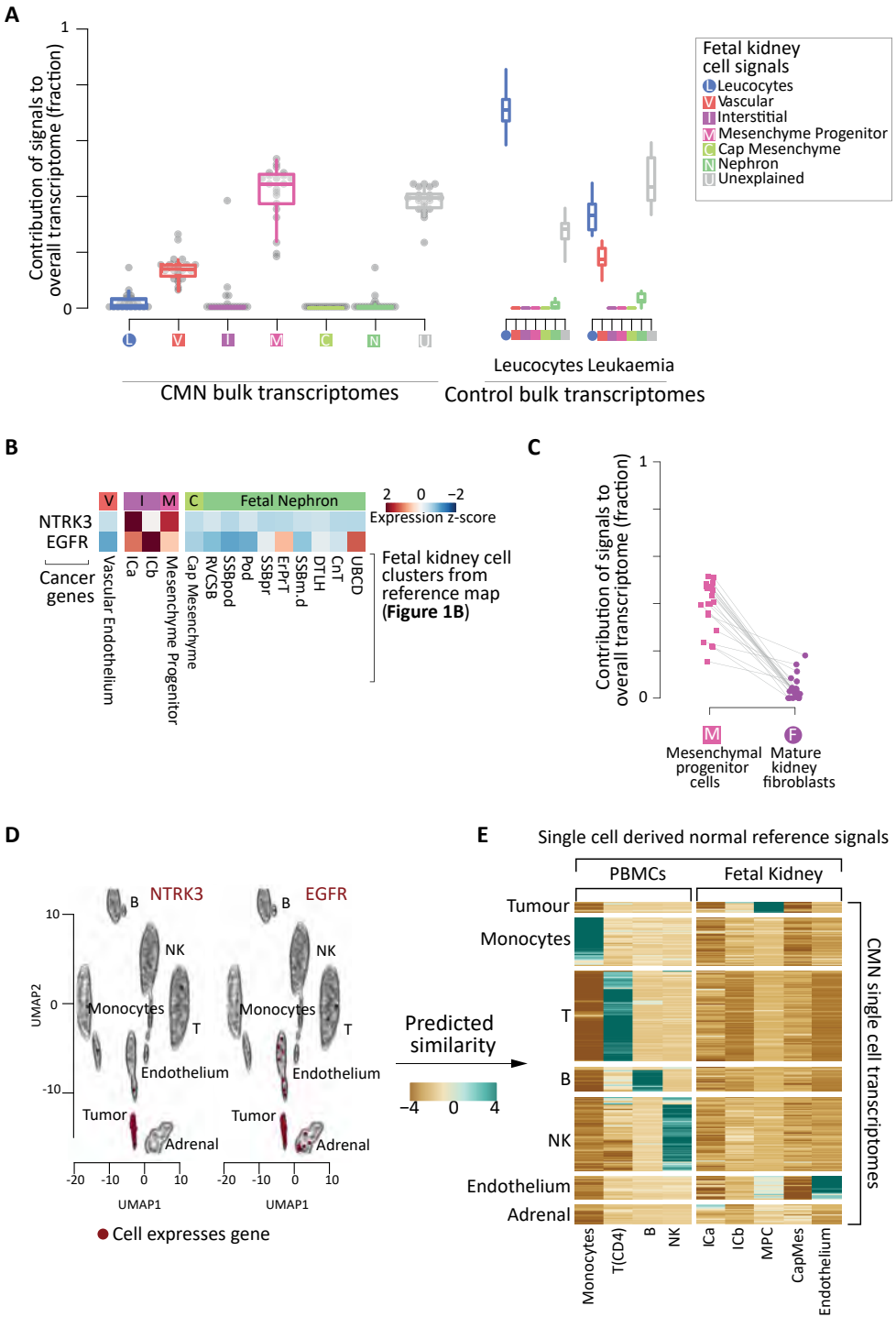
For each tumor, we determined whether it exhibited a fetal or mature (i.e., post-natal) transcriptome, to guide the choice of reference in subsequent analyses. This analysis also enabled us to test two fundamental hypotheses about the differentiation state of tumors—that childhood tumors represent fetal cell types and that adult cancers, especially epithelial malignancies, dedifferentiate towards a fetal state. We define dedifferentiation to be the reversion of a mature cell to a fetal state, at the level of the whole transcriptome.

We calculated the immaturity by fitting each bulk transcriptome to a combined reference set composed of cellular signals from both mature and fetal kidney reference populations. The immaturity score was the fractional contribution of the developmental signals to the bulk transcriptome. Using this approach, we established a reference range of mature normal kidneys (**Fig. 2A**). We demonstrated the validity of this range by scoring fetal kidney transcriptomes which lay significantly outside the mature range ($p = 0.015$, Wilcoxon rank sum test).

We next calculated the same maturity score for individual tumors, which showed a clear signal of "fetalness" across all types of childhood kidney tumors (**Fig. 2B, C**). Although all childhood kidney tumors had a significant enrichment for developmental cellular signals, pretreated Wilms tumor had a significantly lower score than other childhood kidney tumors, including untreated Wilms. The comparison between treated and untreated Wilms suggests that chemotherapy reduces the developmental signal in Wilms tumor, a notion we explore in detail in a later section.

A significant developmental signal was absent from almost all adult tumors (**Fig. 2B**). This suggests that global "dedifferentiation" to a developmental state does not occur in adult kidney tumors. One obvious exception to the ubiquitous lack of a strong developmental signal in adult tumors ($p < 10^{-4}$, Wilcoxon rank sum test) was a cohort of lethal chromophobe RCC, classified previously as metabolically divergent due to their comparatively low expression of genes associated with the Krebs cycle, electron transport chain, and the AMPK pathway[19].

of Chromophobe renal cell carcinomas have a significantly different maturity score as indicated by the star (p = 5.6 × 10–6 two-sided Wilcoxon rank sum test). **(C)** Childhood renal tumors: Immaturity score (as in A) for 287 childhood renal tumors, with normal immaturity score range shown by blue shading. Each type of childhood tumor had a significantly different maturity score than post-natal normal tissue kidneys (p < 2.2 × 10–16 (MRT), 6.8 × 10–14 (CMN), 1.8 × 10–9 (CCSK), < 2.2 × 10–16 (Wilms), and 10–10 (Wilms + Chemo) two-sided Wilcoxon rank sum test). **(D)** ccRCCs by *PTEN* mutation status: Immaturity score for clear cell renal cell carcinomas as calculated in A, split by *PTEN* mutation status (0 = wild type, 1 = mono-allelic loss, 2 = bi-allelic loss). The star indicates that bi-allelic loss is a significant predictor of higher immaturity score (p = 0.003, two-sided t-test with multiple hypothesis correction). **(E)** ccRCCs by transcriptional group: Immaturity score for clear cell renal cell carcinomas as calculated in A, split by transcriptomic subgroups20. The star indicates that samples in m3 have a significantly lower immaturity score (p = 2.9 × 10–6, two-sided t-test with multiple hypothesis correction). Source data are available as a Source Data file.


A
Contribution of signals to overall transcriptome (fraction)
1
0
Fetal kidney cell signals
Leucocytes
Vascular
Interstitial
Mesenchyme Progenitor
Cap Mesenchyme
Nephron
Unexplained
L V I M C N U
CMN bulk transcriptomes
Leucocytes Leukaemia
Control bulk transcriptomes
B
V I M C Fetal Nephron
NTRK3
EGFR
2 0 −2
Expression z-score
Cancer genes
Vascular Endothelium
ICa
ICb
Mesenchyme Progenitor
Cap Mesenchyme
RVCSB
SSBpod
Pod
SSBpr
ErPrT
SSBm.d
DTLH
CnT
UBCD
Fetal kidney cell clusters from reference map (Figure 1B)
C
Contribution of signals to overall transcriptome (fraction)
1
0
M
F
Mesenchymal progenitor cells
Mature kidney fibroblasts
D
NTRK3
EGFR
B
NK
Monocytes
T
Endothelium
Tumor
Adrenal
UMAP2
10
0
-10
-20 -10 0 10
UMAP1
Cell expresses gene
Predicted similarity
−4 0 4
E
Single cell derived normal reference signals
PBMCs
Fetal Kidney
Tumour
Monocytes
T
B
NK
Endothelium
Adrenal
CMN single cell transcriptomes
Monocytes
T(CD4)
B
NK
ICa
ICb
MPC
CapMes
Endothelium

Motivated by this observation, we tested whether other clinical markers such as somatic genotype, morphology, or molecularly defined subgroup were predictive of immaturity score. We found that clear cell renal cell carcinomas with two independent somatic mutations in *PTEN* had a significantly higher immaturity score (**Fig. 2D**; t-test, FDR < 0.01). As with lethal chromophobe tumors, *PTEN* mutated ccRCCs conferred a far worse prognosis, with all samples belonging to the TCGA defined m3/ccB subgroup with the worst prognostic outcome of all groups[20]. Investigating further, we found an association between immaturity score and the m3 transcriptional subgroup (**Fig. 2E**; t-test, FDR < 0.01). No other clinical covariate had a statistically significant association with immaturity score at a 1% significance level (**Supplementary Tables 2 and 3**).

Congenital mesoblastic nephroma resembles mesenchymal progenitor cells

Congenital mesoblastic nephroma (CMN) is a renal tumor of infants that has low metastatic potential. There are two morphological subtypes of CMN, classical and cellular variants[21]. Cell signal analysis in CMN bulk transcriptomes ($n = 18$) revealed a uniform signal of mesenchymal progenitor cells across tumors (**Fig. 3A, S4**), irrespective of morphological subtype. Of note, these mesenchymal progenitor cells were characterized by expression of *NTRK3* and *EGFR* genes (**Fig. 3B**), the principal oncogenes that drive CMN through activating structural variants[22]. To verify that this signal was not a generic consequence of fibroblast-like cells, we repeated the analysis of bulk CMN transcriptomes using a developmental reference combined with mature fibroblasts. This comparison revealed the same match to mesenchymal progenitor cells, with a low contribution from mature fibroblasts (**Fig. 3C**).

Figure 3 | Congenital mesoblastic nephromas.
(A) Composition of bulk CMNs: The relative contribution of single cell derived signals from fetal kidney in explaining the bulk transcriptomes of 18 congenital mesoblastic nephromas (CMNs) along with control leucocyte and ALL populations. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is represented by a single point and boxplots shows the distribution with median (middle line), 1st and 3rd quartiles (box limits) and 1.5 times the inter-quartile range (whiskers). Each signal type is abbreviated and colored as per the legend, with squares for fetal and circles for mature. CMN samples are shown on the left and control samples on the right, where "Leukocytes" are bulk transcriptomes from flow sorted leukocytes and "Leukemia" represent B-precursor acute lymphoblastic leukemia. **(B)** Expression of CMN cancer genes in fetal kidney: Expression of CMN driver genes (rows) in reference fetal kidney single cell RNA-seq populations (columns), scaled to mean 0 and standard deviation 1 in each row (i.e., z-transformed). **(C)** Comparing mesenchymal progenitor cell signals to mature fibroblasts: All 18 CMN bulk transcriptomes were analyzed using a reference signal set including both fetal kidney cells and the fibroblasts from mature kidney. This figure shows the comparison of their inferred contribution to each transcriptome for each sample (y-axis), with lines joining points representing the same sample. **(D)** Expression of CMN marker genes: tSNE map of 4,416 single cell transcriptomes from a CMN biopsy, where contours indicate clusters of cells of the type labeled. Cells positive for *NTRK3* (left) and *EGFR* (right) are colored red. *B* B cell, *T* T cell, *DC* dendritic cell, *NK* NK cell, *NKT* NKT cell. **(E)** Comparison of single cell CMN to fetal kidney: Comparison of clusters of cells from D. (rows) with fetal kidney and leucocyte reference populations (columns). For each CMN cluster/reference population pair a log-similarity score was calculated using logistic regression (see Methods). Positive log-similarity scores represent a high probability of similarity between the reference and test cluster. Source data are available as a Source Data file.

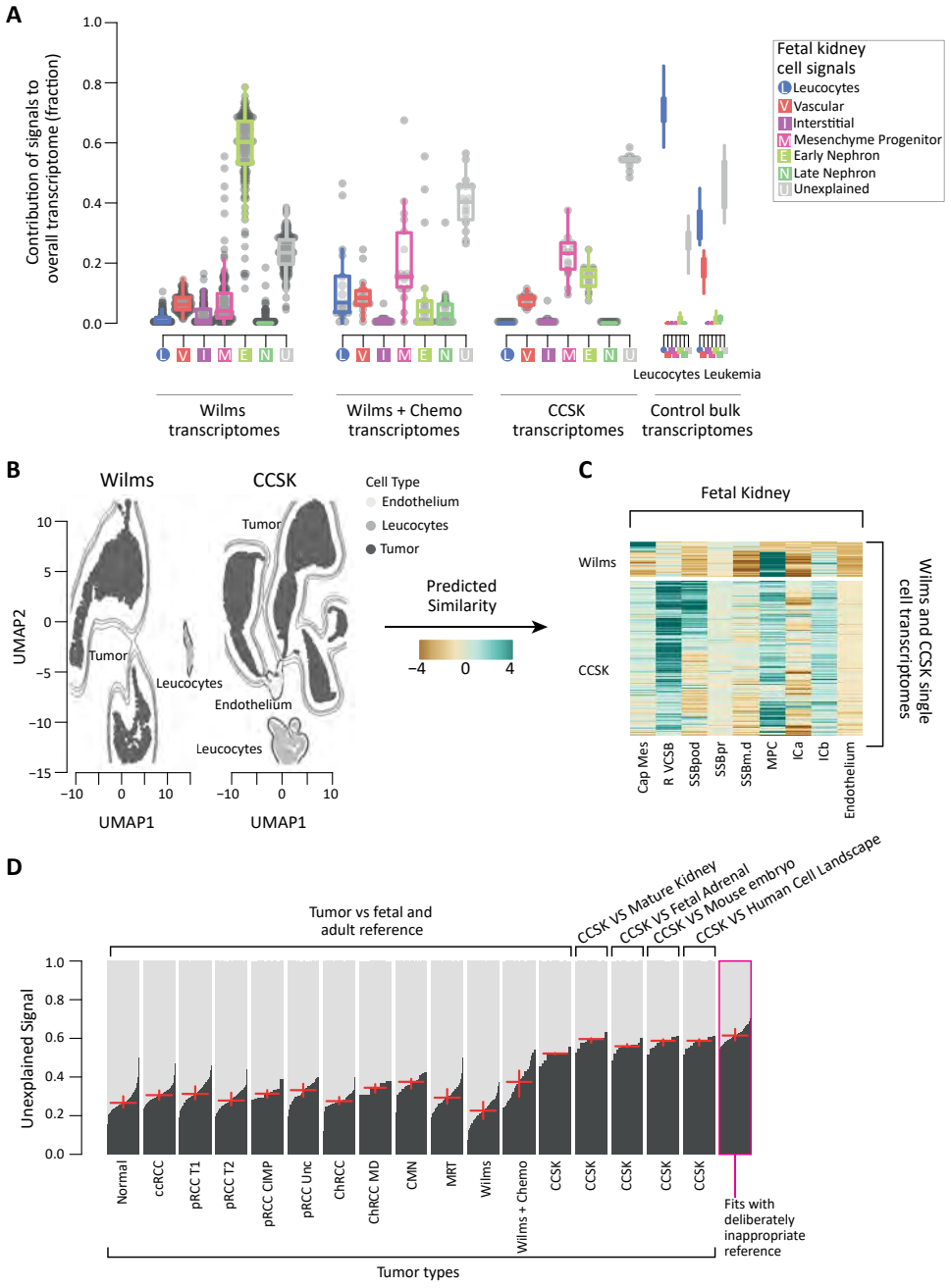


Figure 4 | Wilms tumor and clear cell sarcoma of the kidney.
(A) Bulk Wilms tumor and CCSK compared to fetal kidney: The relative contribution of single cell derived signals from fetal kidney in explaining the bulk transcriptomes of 137 nephroblastomas (Wilms tumors) and 13 CCSKs along with control populations. The relative contribution of each signal to each bulk RNA-seq

To validate this mesenchymal stem cell signal in CMN, we subjected cells dissociated from a fresh CMN tumor specimen, to single cell mRNA sequencing using the Chromium 10x platform. We annotated single cells based on literature derived marker genes (**Fig. 3D, S5**) and compared to single cell clusters of normal fetal kidneys using previously developed quantitative approaches[1]. This comparison revealed that CMN tumor cells matched the same mesenchymal progenitor cell population, validating the cell signal seen in bulk tumor tissue (**Fig. 3E**).

Wilms tumor, clear cell sarcoma of the kidney and the effect of treatment

Wilms tumor is the most common childhood kidney cancer and is thought to arise from aberrant cells of the developing nephron. Clear cell sarcoma of the kidney (CCSK) is a rare, at times aggressive childhood renal cancer that is treated as a high risk Wilms tumor in clinical practice[23]. We assessed the cellular signals in bulk transcriptomes from treatment-naive CCSK, high risk treatment-naive Wilms, and intermediate risk Wilms post chemotherapy. Cellular signal analysis revealed a largely uniform early nephron signal (cap mesenchyme, comma-shaped body, S-shaped body) in the treatment-naive Wilms cohort (**Fig. 4A, S6**). By comparison, the post-treatment cohort had a much-reduced contribution from the early nephron, instead containing a mixture of tubular, early nephron, and mesenchymal signals with a relatively high unexplained signal fraction (**Fig. 4A, S6**). Previous work utilizing single cell data from post-chemotherapy Wilms tumors identified the same lack of cap mesenchyme signal identified by our analysis of bulk transcriptomes[1]. The CCSK transcriptomes showed a mixture of mesenchymal and early nephron signals, with an extremely high unexplained signal fraction (**Fig. 4A, S6**).

To validate the cap mesenchyme signal in treatment-naive Wilms, we generated single cell mRNA transcriptomes from one fresh, treatment-naive sample. Annotation of this data revealed two proliferating populations (**Fig. 4B, S7**). Comparison to fetal

sample is shown by the y-axis. Each signal/sample combination is represented by a single point and boxplots shows the distribution with median (middle line), 1st and 3rd quartiles (box limits) and 1.5 times the inter-quartile range (whiskers). Each signal type is abbreviated and colored as per the legend, with squares for fetal and circles for mature. Wilms/CCSK samples are shown on the left and control samples on the right, where "Leukocytes" are bulk transcriptomes from flow sorted leukocytes and "Leukemia" represent B-precursor acute lymphoblastic leukemia. **(B)** UMAP of CCSK and Wilms single cell transcriptomes: Points represents cell transcriptomes from Wilms tumor (left) or CCSK (right), with shading, contours, and labels indicate cell type. **(C)** Comparison of CCSK and Wilms transcriptomes to reference signals: Similarity of transcriptomes from B (rows) to fetal kidney reference signals (columns), where color indicates logit similarity. **(D)** Comparison of unexplained signal contribution to CCSKs and other tumor types: For each group of samples, the unexplained signal is calculated using the reference set of signals given at the top (e.g., fetal kidney). The unexplained signal fractions are shown by black bars, sorted in increasing order, with the red horizontal line showing the median value and the vertical line the range between the 25th and 75th percentiles. CCSK samples were fitted using 5 different reference sets (fetal and mature kidney, mature kidney only, fetal adrenal, mouse embryo, and the pan-tissue human cell landscape). The final group on the right, represents samples fitted using inappropriate references. This population serves as a calibration of the expected level of unexplained signal when the bulk transcriptome is not explained by any of the provided reference signals. Source data are available as a Source Data file.

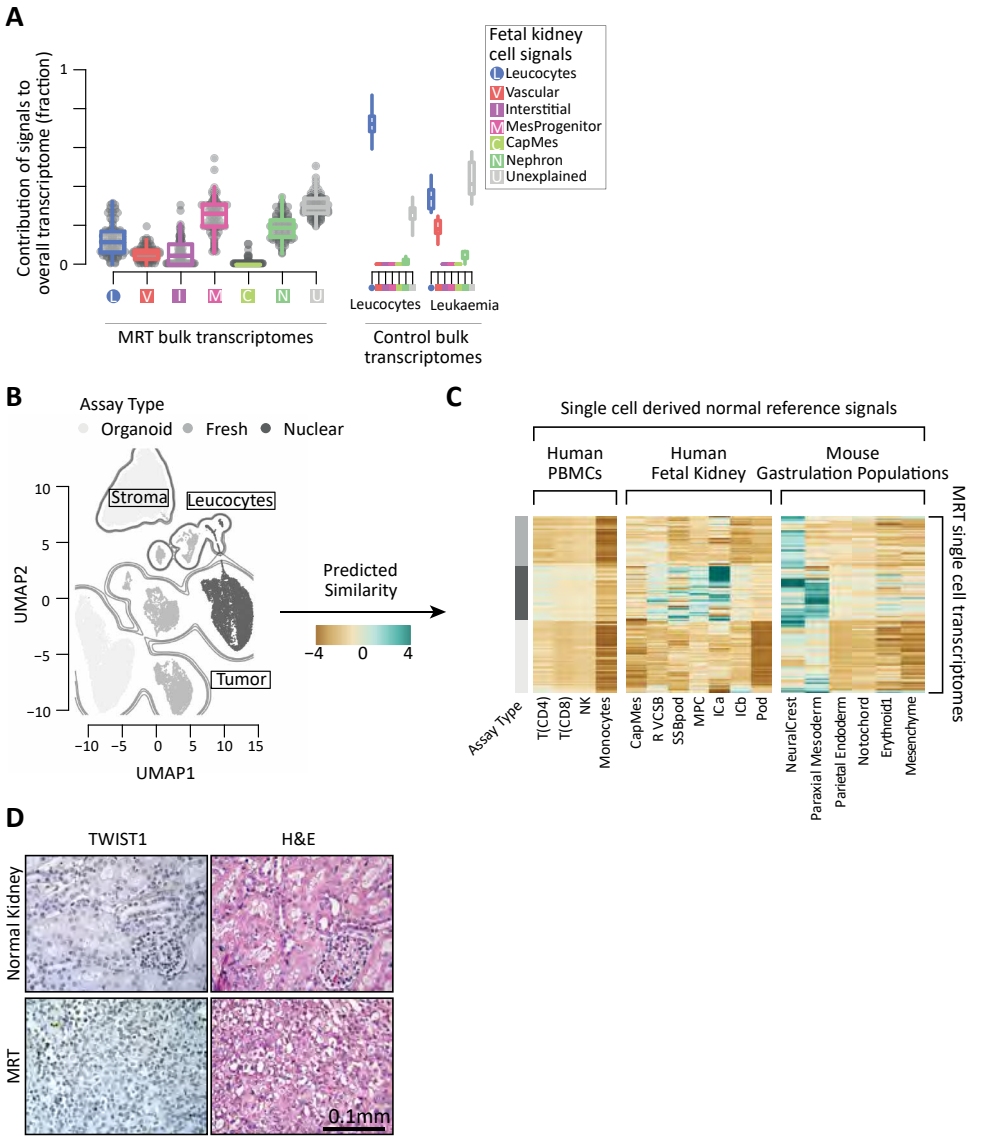


Figure 5 | Malignant rhabdoid tumors.
(A) Bulk MRTs compared to fetal kidney: The relative contribution of single cell derived signals from fetal kidney in explaining the bulk transcriptomes of 65 malignant rhabdoid tumors (MRTs) along with control populations. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is represented by a single point and boxplots shows the distribution with median (middle line), 1st and 3rd quartiles (box limits) and 1.5 times the inter-quartile range (whiskers). Each signal type is abbreviated and colored as per the legend, with squares for fetal and circles for mature. MRT samples are shown on the left and control samples on the right, where "Leukocytes" are bulk transcriptomes from flow sorted leukocytes and "Leukemia" represent B-precursor acute lymphoblastic leukemia. **(B)** UMAP of single cell MRT transcriptomes: Each dot represents a single transcriptome from

kidney showed that one of these populations strongly matched the cap mesenchyme, validating its presence in treatment-naive Wilms tumor (**Fig. 4C**). The second population exhibited a strong match to mesenchymal progenitor cells (**Fig. 4C**).

To further investigate the origins of CCSK we generated single nuclear transcriptomes from 2 archival samples and single cell transcriptomes from one fresh sample (**Fig. 4B, S8**). In contrast to Wilms tumor, all CCSK tumor cells matched multiple mesenchymal and early nephron populations (**Fig. 4C**). Although the matching populations were consistent with the results of cell signal analysis on bulk CCSK transcriptomes (**Fig. 4A**), the match to multiple reference populations at the single cell level suggests that CCSK transcriptomes represent a transcriptional state that is intermediate between multiple mesenchymal populations in the developing kidney. To test the possibility that the true normal cell correlate for CCSKs was not in the fetal kidney, we next matched CCSK bulk transcriptomes against mature kidney, fetal adrenal, developing mouse, and the pan-tissue human cell landscape[24]. In each of these comparisons, the unexplained signal explained at least 50% of the CCSK bulk transcriptomes, a much higher fraction than any other tumor type (**Fig. 4D**). This unexplained signal fraction was comparable to the level obtained from a deliberately inappropriate comparison of flow sorted B cell bulk transcriptomes compared to the non-immune developing kidney (**Fig. 4D**). In aggregate, these data suggest that CCSKs represent transcriptionally grossly distorted renal mesenchymal cells.

Malignant rhabdoid tumors exhibit signals of neural crest and early mesenchyme

Malignant rhabdoid tumor (MRT) is an aggressive, often fatal childhood cancer, that typically affects the kidney but may also occur at other sites. It is considered to be the extracranial counterpart of the CNS tumor, atypical teratoid/rhabdoid tumor (AT/RT). The principal, usually sole, driver event in MRT and AT/RT is biallelic inactivation of *SMARCB1*. In previous analyses of microRNA profiles, MRTs co-clustered with a range of tissues: neural crest derived tumors, cerebellum, and synovial sarcoma[25].

Assessing fetal renal single cell signals in 65 MRT bulk transcriptomes yielded a mesenchymal progenitor cell signal (**Fig. 5A, S9**). However, the nephron and unexplained signal fractions were also high, indicating that tumor cells only moderately resemble this reference population. To investigate further, we studied

either tumor/tubular derived organoid cells (white), fresh tissue MRTs cells (gray) or archival MRT nuclei (black). Contours indicate tumor cells, stroma, and leucocytes as labeled. **(C)** Log similarity of single cell MRT cells to fetal kidney and developing mouse: Comparison of the transcriptomes in B to cellular signals defined from single cell reference transcriptomes. The reference population is indicated on the x-axis and the gray bar on the left indicates the technology each cell was derived from. Each row corresponds to a single transcriptome from B. The color scheme encodes the logit similarity score for each cell against each reference population (see Methods). **(D)** Immunohistochemistry of TWIST1 in MRT and normal kidney: Staining of a region of normal kidney and MRT tissue for TWIST1. The MRT image shows a part of the tissue selected for being TWIST1 positive, there were large sections of tumor tissue that were also TWIST1 negative. All normal kidney tissue was TWIST1 negative. This experiment was repeated 3 times and the scale bar (bottom-right) indicates 0.1 mm. Source data are available as a Source Data file.

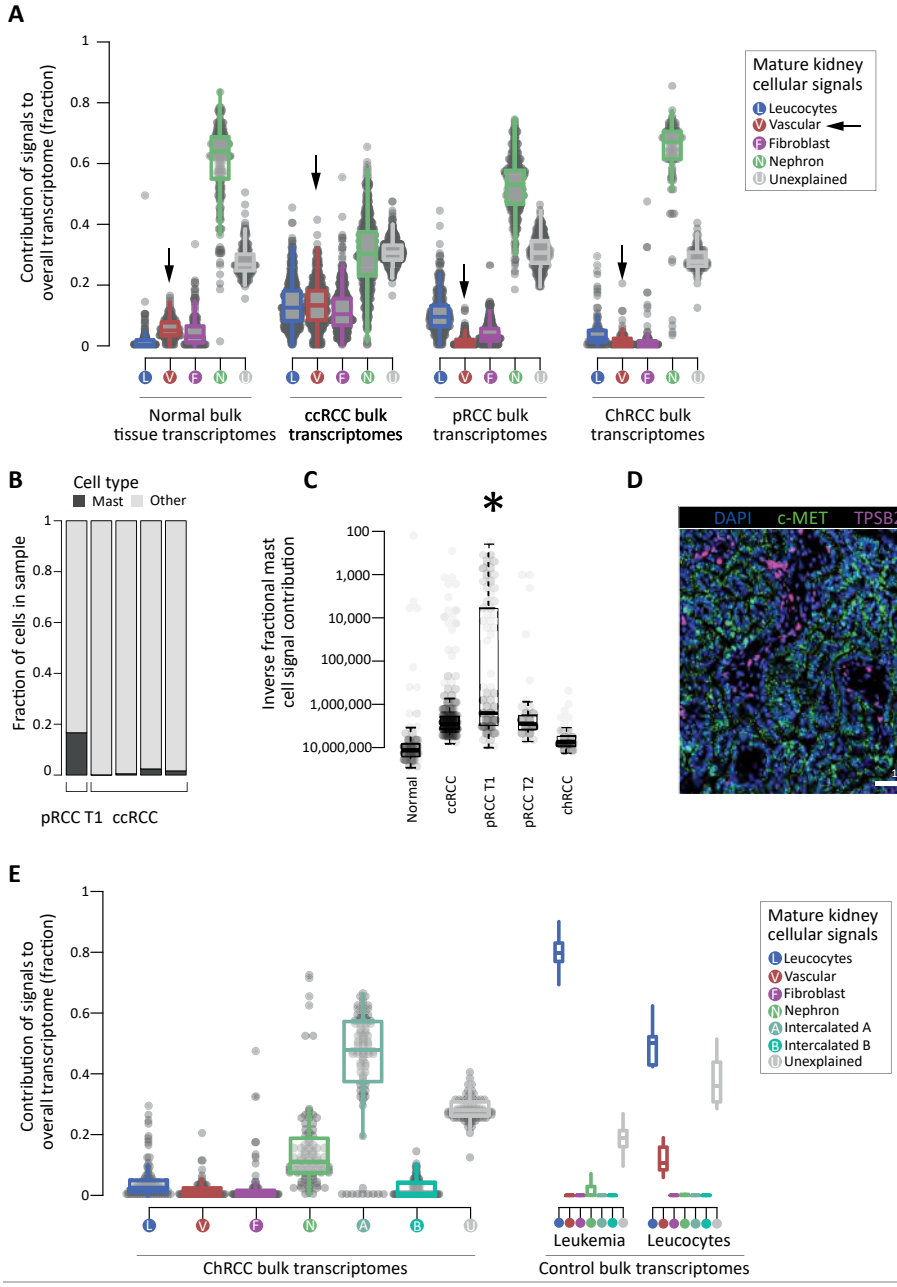


Figure 6 | Adult kidney tumors.
(A) Bulk renal cell carcinomas compared to mature kidney: The relative contribution of single cell derived signals from fetal kidney in explaining the bulk transcriptomes of 171 normal kidney biopsies, 500 clear cell renal cell carcinomas (ccRCC), 274 papillary renal cell carcinomas (pRCC), and 81 chromophobe renal cell carcinomas (ChRCC), along with control populations. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. Each signal/sample combination is represented by a single point and boxplots shows the distribution with median (middle line), 1st and 3rd quartiles (box limits) and 1.5

MRT single cell transcriptomes, derived from an MRT expanded by a primary organoid culture[26] (see Methods), from nuclear mRNA sequencing, and from fresh tissue MRT cells (**Fig. 5B, S10**). Comparison to our fetal kidney reference revealed that MRT cell transcriptomes did not show any consistent match (**Fig. 5C**). This may indicate that the mesenchymal progenitor cell signal obtained in bulk represents a signal of the broad embryological lineage of the tumor, rather than a cell type.

We therefore compared MRT cells against published reference cell populations of gastrulation embryos generated from mice[13], a developmental stage that is not accessible to study in humans. Although there were differences between and within samples, all produced a match to neural crest and/or early mesodermal/mesenchymal populations (**Fig. 5C**). To validate this early mesodermal signal, we performed immunohistochemistry for the presence of a protein specific to paraxial mesoderm, TWIST1. Consistent with its expression in a subset of cells by single cell mRNA sequencing, occasional MRT cells exhibited TWIST1 staining, whilst no protein was detected in normal kidney (**Fig. 5D, S11**). Overall our data show that MRTs do not exclusively exhibit mRNA signals of either neural crest or mesenchyme cells. Instead, our findings point at a hybrid state of MRTs, representing mRNA features of both, neural crest and mesenchyme, suggesting that MRTs may come from early mesoderm or form along the differentiation trajectory of neural crest to mesenchyme.

Adult tumors represent specific tubular cells

As discussed above, our analyses confirmed a previous finding that the predominant single cell signal in the most common types of adult renal cancer, clear cell RCC (ccRCC) and papillary RCC (pRCC), was derived from a specific subtype of proximal tubular cells, termed PT1 cell (**Supplementary Fig. 3**)[1]. In addition, cell signal analysis also revealed some properties of the tumor microenvironment. We found a prominent vascular endothelial signal in ccRCCs (**Fig. 6A, S12**), but not in pRCCs. The downstream effects in RCC of inactivation of the von Hippel–Lindau gene and upregulation of vascular endothelial growth factors are well documented[27]. The

times the inter-quartile range (whiskers). Each signal type is abbreviated and colored as per the legend, with squares for fetal and circles for mature. **(B)** Mast cell fraction in single cell RCC samples: Bar height indicates mast cell fraction (black) or other cell fraction (gray) in 5 single cell RCC expriments (x-axis labels). **(C)** Mast cell signals in bulk RCC transcriptomes: Inverse of mast cell fraction for bulk transcriptomes (dots) of type given on x-axis. Boxplots show the distribution median (middle line), 1st and 3rd quartiles (box limits), and 1.5 times the inter-quartile range beyond the box-limits (whiskers) and the star indicates that mast cell signals are significantly higher in pRCC T1 type tumors than pRCC T2 (two-sided Wilcoxon rank-sum test, $p = 1.5 \times 10–5$). **(D)** smFISH validation: An example section of single molecule fluorescence in-situ hybridization imaging of a pRCC T1 tumor section. Nuclei are stained blue with dapi and expression of the tumor marker *MET* is shown in green and the mast cell marker TPSB2 in purple. See Supplementary Table 4 for a quantification of smFISH applied to pRCC T1/T2 and ccRCC tumors. smFISH imaging was performed on one tumor section from each of pRCC T1, pRCC T2, and ccRCC. The scale bar (bottom-right) indicates approximately 100 µm. **(E)** Bulk chromophobe renal cell carcinomas compared to mature kidney: The same as A, but for 81 chromophobe renal cell carcinomas (ChRCC) bulk transcriptomes. Source data are available as a Source Data file.

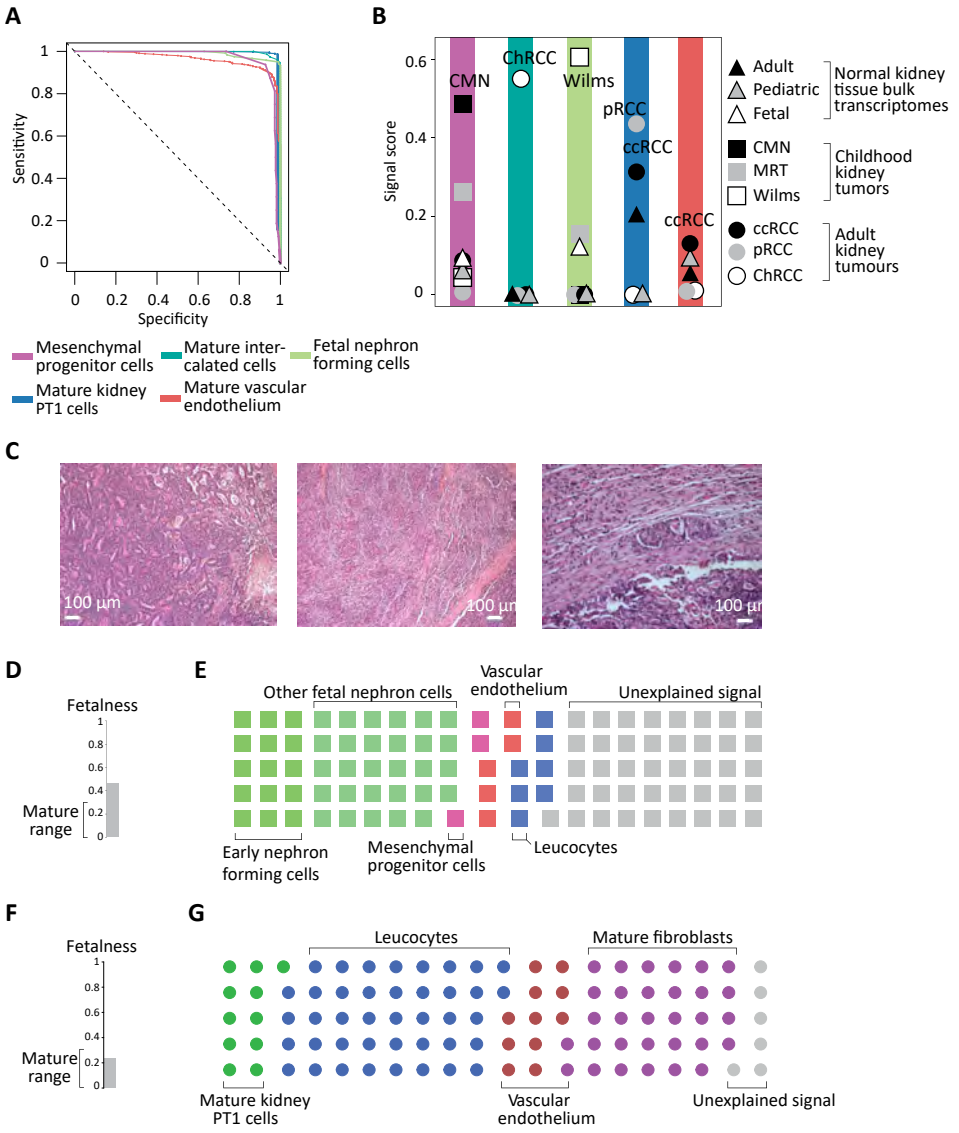


Figure 7 | Clinical utility of cellular signal analysis.
(A) Sensitivity/Specificity of signals in classifying tumor types: Curves showing the sensitivity and specificity of using the scores defined by the color scheme to classify tumors by type at different cut-offs. The different score and tumor type pairs are: fetal interstitial cells and CMN (light blue), intercalated cells and ChRCCs (dark blue), developing nephron and Nephroblastoma (light green), PT1 and ccRCCs/pRCCs (dark green), and mature vascular and ccRCCs (red). **(B)** Median reference contribution by tumor type: Each point represents the median score for the group of samples indicated by the combination of shape (tissue type, see legend) and shading (score type, as in A). **(C)** Histology image of unclassified childhood renal tumor: The tumor mostly compromised pleomorphic epithelioid cells that formed tubules, papillae, glands and nests, as well as more solid areas with spindled cells and clefting similar to that of synovial sarcoma. Patchy tumor necrosis was apparent. Some areas showed smaller, more uniform cells lining narrow tubular

prominent difference in the endothelial signal provides a read-out of this pathway, further explaining why anti-angiogenic treatments appear to be more effective in ccRCCs than in pRCCs[28].

Continuing our investigation of the tumor microenvironment, we observed mast cells to be over-represented in single cell data derived from pRCCs (**Fig. 6B**). Performing cellular signal analysis revealed a high contribution of mast cell signal in a subset of tumors, significantly enriched for type 1 pRCC tumors ($p < 1e\text{-}4$, Wilcoxon rank-sum test; **Fig. 6C**). This finding was further validated by single molecule fluorescence in-situ hybridization (smFISH), which found a higher fraction of mast cells a type 1 pRCC sample, than type 2 or ccRCC (**Fig. 6D, Supplementary Table 4**).

Previous analyses of chromophobe cell renal cell carcinoma (ChRCC) have shown that ChRCC exhibit expression profiles of collecting duct cells[29]. Controversy exists as to whether the normal cell correlate of ChRCC is the type A or type B intercalated cells[30]. This is in part due to ChRCC retaining expression of both canonical markers of intercalated cells, SLC4A1 and SLC26A4 respectively (**Supplementary Fig. 13**). Using cell signal analysis, which considers the entire transcriptomes of type A and type B cells, rather than just two markers, revealed a uniform type A signal across all chromophobe tumors (**Fig. 6E, S14**), bar the lethal variant of so-called metabolically divergent tumors (**Fig. 2B, S14**). The proliferation and active remodeling of type A cells has been demonstrated under conditions of systemic acidosis[31], lending further credence to their possible status as the cell of origin for ChRCCs.

Single cell signals provide diagnostic clues

An overarching finding of our study was that each tumor type possesses a particular pattern of cellular signals that were uniform in, and specific to, bulk transcriptomes from individual tumor types. Accordingly, cellular signal assessment of bulk transcriptomes may provide sensitive and specific diagnostic clues. To test this proposition, we assessed how accurately the tumor type of each sample in our data could be determined based only on its cellular signals. We found that the prevalence of the most common cellular signal for each type could be used to infer the tumor type of each bulk transcriptome (**Fig. 7A, B, S15**). As further validation, we applied this approach to an independent cohort of Wilms tumors. All were correctly identified as childhood tumors and had cellular signals consistent with Wilms tumor (**Supplementary Fig. 16**).

structures, resembling adenomatous perilobar nephrogenic rests. Overall, the morphology and ancillary tests were inconclusive. Scale bars at bottom of each image indicate approximately 100 µm. **(D)** Immaturity score for unclassified childhood renal tumor: Calculated as in Fig. 2, with score range for normal post-natal kidney indicated on left. **(E)** Summary of signal contribution from fetal and mature kidney to unclassified childhood renal tumor: Each color represents the signal type labeled and fraction of squares of each type matches the signal contribution. **(F)** Immaturity score for childhood renal cell carcinoma: As in D. **(G)** Summary of signal contribution from fetal and mature kidney to childhood renal cell carcinoma: As in E but for a transcriptome derived from renal cell carcinoma fit using a mature kidney signal set. Source data are available as a Source Data file.

We next examined cellular signals in the bulk transcriptome of a histologically undefinable metastatic primary renal tumor from an 11-year-old boy. Following resection, the tumor was examined histologically, both locally and by international reference renal pathologists (**Fig. 7C**). A definitive diagnosis could not be reached although an adult type renal cell carcinoma was favored. Nevertheless, the child was treated as a Wilms-like tumor, with cytotoxic chemotherapy and radiotherapy, following nephrectomy. He remains in complete remission two years following diagnosis, thus retrospectively suggesting a diagnosis of a Wilms-like tumor, as adult type kidney carcinomas do not respond to cytotoxic treatment.

We performed bulk mRNA sequencing on tumor specimens from this patient. Assessment of mRNA signals in bulk tissue suggested that the tumor exhibited a fetal transcriptome with cellular signals consistent with a Wilms-like tumor (**Fig. 7D, E**). The transcriptional diagnosis of a Wilms-like tumor was further substantiated by analyses of whole genome sequences. The tumor harbored classical somatic changes of Wilms, namely canonical *CTNNB1* and *KRAS* hotspot mutations and uniparental disomy of 11p (**Supplementary Fig. 17**). By comparison, when we assessed single cell signals of an adult-type ccRCC that developed in a 15 year old adolescent, we found an overall mature transcriptome. Furthermore, the tumor exhibited the PT1 signal of ccRCC as well as a stark vascular endothelial signal typical of ccRCC (**Fig. 7F, G**).

Discussion

We have determined normal cell signals in the major types of human renal tumors. This has enabled us to replace the approximate notion of the "fetalness" of childhood renal tumors with quantitative transcriptional evidence that the entire spectrum of pediatric renal tumors represent an aberrant developmental state. At the same time, our analyses question the suggestion that adult, epithelial-derived kidney cancers revert to a fetal state at the whole transcriptome level (i.e., "dedifferentiate"). Importantly, when we found transcriptional evidence of dedifferentiation in adult tumors, it conferred a dismal prognosis. Furthermore, among childhood tumors we found examples of cell signals representing differentiation trajectories, such as the neural crest to mesenchyme conversion in MRT, validating our recent finding based on phylogenetic and differentiation studies[32]. By contrast, the different types of adult tumors resembled specific renal tubular cells.

A central question that our findings raise is whether mRNA signals point to the cell of origin of tumors. When the similarity between mRNA signals and specific cell types was high, as found in most tumor types, this may be a plausible proposition. For example, in CMN, which typically occurs within the first weeks of life, our analysis identified an early mesenchymal progenitor cell population, characterized by the disease-defining oncogenes of CMN, as the likely cell of origin of CMN. In some tumors, transformation may entirely distort and obliterate gene expression profiles

of the cell of origin. We found CCSK transcriptomes to represent such an extreme modification of the transcriptome of the developing kidney.

A further finding of our study was that within each category, the majority of tumors exhibited remarkably uniform cellular signals. That is, despite a high diversity in clinical outcome, tumors of the same type almost universally had the same dominant cellular signal (**Fig. 7A, B**). This indicates that there are overarching transcriptional features, beyond individual gene markers, that unite tumor entities despite underlying intra- and inter-tumor genetic heterogeneity. Therefore, cellular signals of renal tumors may lend themselves as diagnostic adjuncts, as illustrated here by our ability to resolve the identity of a histologically unclassifiable childhood tumor. Moreover, the cellular transcriptome itself may represent a therapeutic target that transcends individual patients, if we had tools available to manipulate transcription in a predictable manner. This may be a particularly attractive approach for targeting transcriptional states of fetal cells retained in childhood cancer that are absent from normal post-natal tissues.

Overall our findings attach specific cell labels to human renal tumors that are underpinned by quantitative molecular data obtained from single cell mRNA sequences, independent of the interpretation of marker genes. As reference data from single cell transcriptomes expand through efforts such as the *Human Cell Atlas*, it will be feasible to annotate existing large repositories of tumor bulk transcriptomes, to derive a cellular transcriptional definition of human cancer.

Methods

Ethics statement

Human kidney and tumour tissues were collected through studies approved by UK NHS research ethics committees. Patients or guardians provided informed written consent for participation in this study as stipulated by the study protocols. These studies have the following references: NHS National Research Ethics Service reference 03/018 (DIAMOND study; adult kidney tissues); NHS National Research Ethics Service reference 16/EE/0394 (pediatric tissues); NHS National Research Ethics Service reference 96/085 (fetal tissues). Additional fetal tissue was provided by the Joint MRC / Wellcome Trust-funded (grant # 099175/Z/12/Z) Human Developmental BiologyResource (HBDR, http://www.hdbr.org; (10)), with appropriate maternal written consent and approval from the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee. HDBR is regulated by the UK Human Tissue Authority (HTA;www.hta.gov.uk) and operates in accordance with the relevant HTA Codes of Practice. Fetal tissues from both sources were obtained from terminations and ranged from 7 to 18 post conception weeks (**Supplementary Table 1**). Organoids were generated from human tissue as approved by the medical ethics committee of the Erasmus Medical Center (Rotterdam, the Netherlands).

Tissue processing and data generation

10X single cell sequencing of fresh tissue and bulk sequencing of DNA/RNA

Fresh tissues were processed to generate single suspensions for processing on the Chromium 10X controller (V2/3 3' chemistry), as previously described[1]. The MRT and normal kidney tissues organoids were derived and maintained as previously described[26]. Libraries were produced according to the manufacturer's instructions and sequenced on an Illumina HiSeq4000 device. Sequencing of bulk RNA and DNA was performed as previously described[1].

Cell-Seq2 experiments

Following resection, a random piece was selected from viable tumour tissue, minced, and viably frozen. On the day of the sorting, the sample was thawed and dissociated into a single-cell suspension in AdDF+++ (Advanced DMEM/F12 containing 1× Glutamax, 10 mM HEPES and antibiotics) containing Collagenase 1a (1mg/mL, Sigma, C9407) and DNase (0.25µg/mL, Stemcell), supplemented with Rho-kinase inhibitor Y-27632 (10 µM, Abmole). The samples were digested on an orbital shaker for 30 min at 37 °C. The suspension was washed first with AdDF+++ and next with MACS buffer (PBS pH 7.2+2mM EDTA +0.5% Bovine Serum Albumine), followed each time by centrifugation at 300×g. Viable single cells were sorted based on forward/side scatter properties and DAPI/DRAQ5 staining using FACS (MoFlo Astrios EQ, Beckman Coulter) into 384-well plates (Biorad) containing 10 µl mineral oil (Sigma) and 50 nl of RT primers.

10X single nuclei sequencing

Single nuclei were isolated from frozen tissue using a glass dounce homogeniser. Samples were homogenised in buffer A (Sucrose 0.25 M, BSA 10 mg/ml, $MgCl_2$0.005 M, protease inhibitors and RNAse inhibitors RNAseIn – 0.12 U/ul and Superasin 0.06 U/ul), using ~25 strokes with the "loose" pestle and ~20 strokes with the "tight" pestle. Nuclei were cleaned up using a 30% Percol gradient and resuspended in buffer B (Sucrose 0.32 M, BSA 10 mg/ml, CaCl2 3 mM, MgAc2 2 mM, EDTA 0.1 mM, Tris-HCl 10 mM, DTT 1mM in the presence of protease and RNAse inhibitors as in buffer A).

Nuclei were mixed 1:1 with Tryphan blue and counted using a disposable haemocytometer, then diluted to the appropriate concentration. Nuclei were loaded on to the 10X Chromium controller as per the Chromium Single Cell 3 Reagent Kits v3 User Guide, targeting to recover 5000 nuclei. Post GEM-RT cleanup, cDNA amplification and 3' gene expression library construction were carried out according to the user guide. The resulting libraries were sequenced on the Novaseq platform.

Immunohistochemistry of MRT tissue

Immunohistochemistry was performed on 3–4 µm sections of tissue fixed in 4% paraformaldehyde, dehydrated and embedded in paraffin according to standard protocols. Sections were subjected to H&E and immunohistochemical staining using antibodies for INI-1 (BD Transduction Laboratories, 612111, 1:400) or TWIST (Abcam, ab50581, 1:500). Counterstaining was performed using Mayer's Hematoxylin (1:3 dilution). The Leica DMi8 microscope was used for imaging.

RNAscope smFISH and immunohistochemistry

FFPE tissue sections of 5 µm thickness were processed using a Leica BOND RX to automate staining with the RNAscope Multiplex Fluorescent Reagent Kit v2 Assay and RNAscope 4-plex Ancillary Kit for Multiplex Fluorescent Reagent Kit v2 (Advanced Cell Diagnostics, Bio-Techne) in combination with immunohistochemistry (IHC) for PECAM1[33]. Owing to intense tissue autofluorescence, samples were treated with a photobleaching procedure based upon that of Lin *et al.*[34]. It was observed that photobleaching prior to RNAscope probe hybridisation adversely affected smFISH staining, presumably due to loss of RNA integrity in the alkaline solution. Therefore, photobleaching was conducted following RNAscope probe and tree amplification reagents (AMP1/2/3) but before channel-specific HRP reagents and fluorophores. Initial automated processing included baking at 60°C for 30 minutes and dewaxing, as well as heat-induced epitope retrieval at 95°C for 15 minutes in buffer ER2 and digestion with Protease III for 15 minutes. Following RNAscope probe and AMP hybridisation according to the manufacturer's instructions, slides were briefly rinsed in PBS and then subjected to photobleaching. Slides were incubated horizontally in 4.5% hydrogen peroxide, 24 mM sodium hydroxide in PBS in a Nunc Square BioAssay Dish atop a white light box for 30 minutes. Slides were thoroughly rinsed with sterile deionised water and Leica BOND wash before RNAscope staining was resumed

with the sequential development of three probe channels using tyramide signal amplification with Opal 570, Opal 650 (both Akoya Biosciences) and TSA-biotin (TSA Plus Biotin Kit, Perkin Elmer) and streptavidin-conjugated Atto 425 (Sigma Aldrich). Finally, IHC was carried out, beginning with a blocking step of 1 hour in Primary Antibody Diluent (Leica), followed by rabbit anti-PECAM1 (Abcam ab28364) at 1:600 at room temperature for 2 hours, and then HRP goat anti-rabbit IgG (Thermo G21234) at 1:1000 at room temperature for 1 hour. Both antibodies were diluted in Primary Antibody Diluent. IHC signal was developed using Opal 520 (Akoya).

Stained sections were imaged with a Perkin Elmer Opera Phenix High-Content Screening System, in confocal mode with 1 µm z-step size, using a 20× water-immersion objective (NA 0.16, 0.299 µm/pixel). Channels: DAPI (excitation 375 nm, emission 435-480 nm), Atto 425 (ex. 425 nm, em. 463-501 nm), Opal 520 (ex. 488 nm, em. 500- 550 nm), Opal 570 (ex. 561 nm, em. 570-630 nm), Opal 650 (ex. 640 nm, em. 650-760 nm).

Basic data processing and quality control

Mapping of DNA reads

DNA sequencing reads were aligned to the GRCh 37d5 reference genome using the Burrows-Wheeler transform (BWA-MEM)[35]. Sequencing depth at each base was assessed using Bedtools coverage v2.24.0.

Substitution calling

Single base somatic substitutions were called using an in-house version of CaVEMan v1.11.2 (Cancer Variants through Expectation Maximization)[36]. CaVEMan compares sequencing reads from tumor and matched normal samples and uses a naïve Bayesian model and expectation-maximization approach to calculate the probability of a somatic variant at each base (https://github.com/cancerit/CaVEMan). Small insertions and deletions (indels) were called using an in-house version of Pindel (v2.2.2; github.com/cancerit/cgpPindel). Post-processing filters required that the following criteria were met to call a somatic substitution:

1. At least a third of the reads calling the variant had a base quality of 25 or higher.
2. If coverage of the mutant allele was less than 8, at least one mutant allele was detected in the first 2/3 of the read.
3. Less than 5% of the mutant alleles with base quality ≥ 15 were found in the matched normal.
4. Bidirectional reads reporting the mutant allele.
5. Not all mutant alleles reported in the second half of the read.
6. Mean mapping quality of the mutant allele reads was ≥ 21.

7. Mutation does not fall in a simple repeat or centromeric region.
8. Position does not fall within a germline insertion or deletion.
9. Variant is not reported by ≥ 3 reads in more than one percent of samples in a panel of approximately 400 unmatched normal samples.
10. A minimum 2 reads in each direction reporting the mutant allele.
11. At least 10-fold coverage at the mutant allele locus.
12. Minimum variant allele fraction 5%.
13. No insertion or deletion called within a read length (150bp) of the putative substitution.
14. No soft-clipped reads reporting the mutant allele.
15. Median BWA alignment score of the reads reporting the mutant allele ≥ 140.

The following variants were flagged for additional inspection for potential artefacts, germline contamination or index-jumping event:

1. Any mutant allele reported within 150bp of another variant.
2. Mutant allele reported in >1% of the matched normal reads.

Copy number detection in bulk DNA

The ascatNGS algorithm (v4.0.1)[37] was used to estimate tumor purity and ploidy and to construct copy number profiles prior to running the Battenberg algorithm (v2.2.5) (github.com/cancerit/cgpBattenberg) to allow for tumor subclonality.

Bulk RNA mapping and quantification

Where possible, we have processed all bulk RNA-seq data using the exact same pipeline as the recount2 project[38]. That is, we used RAIL-RNA to produce counts of bases aligned to each gene in each sample[39]. Counts were then converted to fragments aligned to genes by dividing counts by the average fragment length for the sample. This approach allowed us to combine our in-house data with any dataset processed by the recount2 project, in particular the TCGA and GTEX projects.

The length for each gene was calculated as the sum of unique exonic bases for all transcripts associated with each gene. We used the gencode v25 GTF annotation[40] and GRCh38 human reference genome.

In order to run the recount2 pipeline, we required access to the sequencer output (BAM files or fastq). In some cases, we only had access to processed data, either in the form of raw fragment counts, or transcripts per million (TPM). TPMs were converted to fragment counts by multiplying by 1 million, rounding to an integer and assigning each gene an effective length of 1.

Where TPM values were needed for direct comparison of gene expression, we calculated TPM values from fragment counts by dividing by gene length, then normalizing the counts/bp by forcing them to sum to 1,000,000 across all genes in a sample.

Single cell RNA mapping, quantification, quality control and normalization

Single cell RNA-seq data were quantified using the 10X software package cellranger (version 2.0.2 for V2 chemistry, 3.0.2 for V3 chemistry) to map sequencing data to version 2.1.0 of the build of the GRCh38 reference genome supplied by 10X.

Data were normalized for sequencing depth by dividing by the total number of UMIs in each cell and then transformed to a log scale for each cell using the Seurat version 3.1.4[41] NormalizeData function. That is, the transformed data, y, is given by:

$$y_{gc} = \log\left(1 + F\frac{x_{gc}}{\sum_g x_{gc}}\right)$$

where x is the UMI count matrix with g indexing gene and c indexing the cell. F is the Seurat "scale.factor" parameter (which we left at the default value of 10,000).

Doublets were determined using scrublet[42] and ambient RNA contamination was removed with SoupX[43]. To filter lower quality cells, we performed high resolution clustering (Seurat graph-based clustering with resolution = 10) and filtered any cell which:

1. had greater than 5% expression originating from mitochondrial genes
2. was marked as a doublet
3. expressed fewer than 500 distinct transcripts
4. or belonged to a cluster where greater than 50% of cells failed one of 1-3.

The rationale behind this approach was to conservatively remove cells with a very similar transcriptome to cells which have failed QC.

To prevent similarity to reference maps (e.g. fetal kidney) being driven by cell cycle state, we also removed any cell with evidence of being in S or G2M phase. We determined the cell cycle phase by scoring each cell based on panel of genes specific to each phase using the Seurat CellCycleScoring function. We also removed all leucocytes from each tissues reference map.

Analysis of processed data

Derivation of color scheme

In deriving a color scheme to represent the different types of cellular signal used in this paper, we started by designating a series of hue ranges to represent each tissue type. These hue ranges were then further sub-divided to represent more specific cell types. To separate fetal and mature versions of the same cell type, we used different

values of the "value" parameter in hue, saturation, value color space to represent fetal (0.9) and mature (0.7) cell signals. Finally, we set the saturation value to 0.6 by default and allowed this to vary as necessary to emphasize differences between cell types with otherwise similar colors. This color scheme is summarized in **Supplementary Fig. 18**.

We also constructed a color scheme for each sample type in this study. We used light/pastel colors to represent non-tumor or control samples and solid colors for tumors. We used the same color to represent Neuroblastoma and ChRCC tumors as they were never referenced in the same plot. This color scheme is summarized in **Supplementary Fig. 19**.

Dimension reduction and cluster generation of single cell RNA data

Following normalization, we identified genes with high variability using the Seurat FindVariableGenes function. This function calculates the mean expression and dispersion for each gene, then groups genes into bins (of size 20) by their mean expression and identifies any gene for which the z-score calculated from the dispersion exceeds some cut-off. We used the default cut-off of z=1 and mean expression in the range 0.1 to 8.

The normalized data was scaled to have mean 0 and standard deviation 1 and principle component analysis was performed using the variable genes identified together with any gene that we identified as being potentially biologically interesting (regardless of its variability in the data).

We determined the optimal number of principle components (PCs) using molecular cross validation (https://github.com/constantAmateur/MCVR)[44]. We used these to construct a two-dimensional representation of the data using either tSNE[45,46] or UMAP[47]. This representation was then used only to visualize the data.

Clusters were identified using the community identification algorithm as implemented in the Seurat "FindClusters" algorithm. We used the number of PCs determined above as input to this method and set the resolution parameter to 1. We chose this value of the resolution parameter as it produced a number of clusters that was large enough to capture most of the important biological variability but not so large as to make detailed manual scrutiny of each cluster impractical. All other parameters were set to the function defaults.

Annotation of fetal kidney single cell data

To create the fetal kidney reference, we combined the raw 10X output from previous studies[1,8] together with data from 4 additional fetal kidneys. The combined data were quantified and clustered as described above, with the exception that the clustering resolution parameter was set to 2 to obtain a more granular annotation.

To annotate these clusters, we used a previously published detailed annotation of the fetal kidney as a reference[8]. That is, we first trained a logistic regression model on

just the "PloS" data[8]. In training this model we used the elastic net regularization procedure with alpha=0.99 to produce strong regularization but prevent strongly co-linear genes being excluded. This model was fit using the "glmnet" R package[48].

To obtain regression coefficients specific to each cluster in our training data we fit a series of N binomial logistic regression models, where N is the number of clusters in the training data (i.e., one-versus-rest binomial logistic regression). To prevent the observed frequencies of cells (which we do not expect to accurately reflect the true abundances in situ) from biasing the regression coefficients we use an offset for each model given by,

$$\log\left(\frac{f}{1-f}\right)$$

where f is the fraction of cells in the cluster being trained.

In each case, we performed 10-fold cross validation and selected the regularization co-efficient, lambda, to be as large as possible (i.e., as few non-zero coefficients as possible) such that the cross validated accuracy was within 1 standard deviation of the minimum.

These models were then used to calculate a predicted similarity for each cell in the combined fetal kidney data set. In calculating the predicted values, an offset of 0 was used. Softmax normalization was not used to allow for the possibility that cell types were present in the combined reference not present in the "PLoS" map. Clusters with a similarity of less than 1 (logit scale) to any of the reference data were labelled as "undecided".

Following the application of the logistic regression model, we elected to merge categories in the reference data that were commonly found in the same clusters. Specifically, we combined:

- NPCa, NPCb, and NPCc categories into CapMes.
- RVCSBa, RVCSBb into RVCSB.
- ErPrT and SSBpr into ErPrT.

We removed Leu, Prolif, PTA and Mes as no cluster contained a majority of these cells. Each cluster was then annotated with whichever of the reference categories had the highest similarity score averaged across all cells in the cluster. This procedure left one cluster as "Undecided" (that is, most cells in these clusters could not be allocated unambiguously to one of the reference populations). Closer inspection of this cluster revealed in to be an early mesenchymal population, which we labeled as MPC for mesenchymal progenitor cells as discussed in the manuscript.

Additional reference signal sets

In addition to the above annotated single cell data sets, cellular reference signals were also taken from additional data sets:

- A mature kidney single cell reference map[1].

The 10x demonstration PBMC data set[49], annotated as described here (https://satijalab.org/seurat/v3.0/pbmc3k_tutorial.html). This data set was used to define a set of leucocyte signals that were added to all other reference maps.

- Fetal adrenal reference map[12]
- Whole embryo mouse data[13]
- A pan-tissue human reference from the human cell landscape publication[24]

Annotation of congenital mesoblastic nephroma single cell data

Single cell transcriptomes derived from a congenital mesoblastic nephroma were processed into clusters as described above. Following clustering, we assigned a cell type to each cluster using marker genes identified as in previous work[1,43].

To confirm that the cluster without expression of other known markers represented CMN tumor cells, we investigated the expression of ETV6, NTRK3 and EGFR in this cluster. CMN is known to be driven by an activating rearrangement between ETV6 and NTRK3, which results in a fusion product with the 5' end of ETV6 fused to the 3' end of NTRK3. As the 10X assay measures expression using a 3' enrichment strategy, CMN tumor cells should show a high level of NTRK3 expression. This is indeed what we find, with the cluster marked tumor expressing NTRK3 more than 10 times more strongly than the next closest cluster. By contrast, ETV6 was not highly expressed in this cluster, as would be expected. Finally, EGFR is known to be highly expressed by tumor cells and was also found to be highly expressed in the cluster designated as tumor.

Annotation of congenital malignant rhabdoid tumor cell data

For MRT single cell data, clusters were determined and marker genes identified as described above for CMN. Common non-tumor populations were annotated based on well known markers (**Supplementary Fig. 10**). Tumor cells were identified by loss of the *SMARCB1*, except for MRT2 for which the molecular diagnostic workup did not identify mutation of *SMARCB1*. For this sample, tumor cells were identified based on the loss expression of other members of the SWI/SNF *SMARCA2* and *SMARCA4*.

Cell similarity inference using single cell data

To measure the similarity of a target single cell transcriptome to a reference single cell data-set we used the methodology based on logistic regression outlined in detail[1]. Briefly, we train a logistic regression model with elastic net regularization (alpha=0.99)

on the reference training set. We then use this trained model to infer a similarity score for each cell in the query data set for each cell type in the reference data.

Softmax normalization was not used to allow for the possibility that some cells in the query data set do not resemble any of the cell types in the reference data set. Predicted logits were averaged within each cluster in the query dataset. This approach was implemented using the "glmnet" package in R[48].

Similarity to mouse data

In assessing the closest match to the organoid MRT data, we considered a comparison to a mouse reference dataset[13]. We performed the similarity analysis as described above, with the only difference being that we limit the analysis to orthologous genes as determined by ENSEMBL biomart.

Sensitivity and specificity of samples to particular cellular signals

To perform the sensitivity and specificity analysis (**Fig. 7A-B**) we first constructed a set of cellular signals that were indicative of a particular tumor type. These were MPCs and CMN, Intercalated cells and ChRCC, nephrogenic cells (i.e., cap mesenchyme, primitive vesicle, and ureteric bud) and Wilms tumor, PT1 and ccRCC/pRCC, and mature vasculature and ccRCC. We calculated all of these scores for every sample in our data set and then evaluated the sensitivity and specificity at different cut-offs for each score to construct the sensitivity/specificity curves in **Fig. 7A** (ROC curve).

Quantification of smFISH images

The tiled images exported from the Phenix were illumination corrected andstitched together into large, multi-channel fluorescence images by a specialized tool supplied by Perkin-Elmer. These images were then analysed in the Qupath[50] Bioimage Analysis program. First, all nuclei were segmented using Qupath's cell detection algorithm, then each nucleus was expanded by 3 microns to estimate the area covered by the entire cell. The subcellular spot detection option was then used to detect all fluorescent spots in all the RNA-SCOPE channels for the size range of 1-8 square microns. Each spot was automatically assigned to a detected cell. The detection data for each image was then exported as a .csv file for further analysis.

Method to quantify single cell derived signals in bulk transcriptomes

Data preparation

Bulk RNA-seq data

Each bulk RNA-seq sample required two pieces of information: fragment counts per gene and the effective length of each gene. As described above, gene counts for this paper were mostly generated using the Rail-RNA pipeline[39] and effective gene lengths as the sum of unique exonic bases per gene. However, fragment counts and effective gene lengths can be calculated in any way.

Single cell reference data

To calculate reference single cell signals, single cell data must first be clustered and annotated. Cells are then grouped together by annotation and raw counts summed across all cells within a group. Summed counts are then normalized to sum to 1 across all genes so that a reference signal is defined as a vector, S,

$$S = \{s_1, s_2, \ldots, s_m\}\, s.t. \sum_{g=1}^{m} s_g = 1$$

where there are m genes.

Reference signals can be constructed in any way (e.g., including batch correction for combined data sets), so long as the final signal can be normalized such that

$$\sum_g s_g = 1 \sum_g s_g = 1.$$

Model fit

The aim of the signal assignment method is to infer how much of each of the different reference signals best explains the supplied bulk transcriptome. That is, we are aiming to solve for the values of beta in,

$$y_{gp} = \beta_{0p} + \beta_{1p} s_{g1} + \beta_{2p} s_{g2} + \cdots + \beta_{np} s_{gn}$$

where $y_{gp} y_{gp}$ is the fragment counts for gene g in sample p, s_{gc} is the reference signal c for gene g and $\beta_{cp} \beta_{cp}$ is the contribution of signal c to sample p. Note that the term $\beta_{0p} \beta_{0p}$ represents the contribution the intercept term for sample p, which is equivalent to the inclusion of an additional flat signature for which $s_g = \text{constant } \forall g$ $s_g = \text{constant } \forall g$. The inclusion of this intercept term provides a measure of the extent to which the reference signal set is inappropriate for the sample given.

In order to efficiently calculate the contributions of the same set of reference signals simultaneously, we formulated the following model,

$$\beta_{cp} = e^{z_{cp}}$$

where c represents the signal and p the sample as above. We solve for $z_{cp} z_{cp}$ instead of $\beta_{cp} \beta_{cp}$ directly as this formulation ensures that the contributions of each signal are always strictly positive. Next, we calculate the expression for gene g and sample p implied by these values of $\beta_{cp} \beta_{cp}$

$$\lambda_{gp} = S_{gc} \beta_{cp}$$

where g represents the gene, and then calculate

$$\lambda'_{gp} = \lambda_{gp} l_{gp}$$

where $l_{gp} l_{gp}$ represents the effective length of gene g in sample p. This modification by length is necessary as the fragment counts by gene created by bulk RNA-seq

are proportional to the length of a gene. Finally, a joint negative log likelihood is calculated as

$$-\log \mathcal{L} = \sum_{p} \sum_{g} w_g(\lambda'_{gp} - y_{gp} \log \lambda'_{gp})$$

where w_g is an optional gene penalty applied to genes deemed to be biologically less important. In this paper we set w_g to 1 for all genes except a set of housekeeping genes that are set to 0.5 and metabolic genes set to 0. It is this equation that is minimized with respect to z_{cp} in order to solve for the values of $\beta_{cp}\beta_{cp}$. This log-likelihood is simply the Poisson log-likelihood for a Poisson distribution with mean $\lambda'_{gp}\lambda'_{gp}$ (see section below for a discussion of this choice of distribution).

This model is directly specified using the tensorflow framework[51], which allows the efficient minimization of this equation utilizing graphics processors and multi-core machines. The optimization is performed using stochastic gradient decent using Adam[52]. We require two termination conditions to be met before the optimization is terminated:

1. the fractional decrease in the log-likelihood must be less than some tolerance parameter.
2. the fractional change in Q must be less than the same tolerance parameter.

Q is defined as the sum of the sigmoid transformation of z_{cp} and roughly measures the number of signals with non-zero contributions to the fit. Without this second termination condition, optimization would terminate with the coefficients of many signals given small but non-zero values as these non-zero values barely shift the total log-likelihood.

Post processing

Having obtained optimized values for $\beta_{cp}\beta_{cp}$, we next normalize these values by first modifying the intercept term to

$$\beta'_{0p} = \frac{\beta_{0p}}{m}$$

which make the modified intercept term equivalent to fitting an additional signal with a completely flat profile. Following this modification, we then normalize the values of $\beta_{cp}\beta_{cp}$ to sum to 1 for each sample. These normalized values represent the relative contribution of each signal to each sample and are the values reported throughout this manuscript. The final normalization step essentially controls for differences in bulk RNA-seq library size and makes the coefficients comparable across samples.

As an additional measure of the goodness we re-fit the above model with only the intercept term and then calculate,

$$pR^2 = 1 - \frac{\log \mathcal{L}_{\text{full}}}{\log \mathcal{L}_{\text{int}}}$$

which is, a McFadden's pseudo R-squared value[53] pR^2 given by 1 minus the ratio of the log-likelihood of the full model fit over the model fit with only the intercept term.

In order to aid with interpretation of the fit, the contribution from similar cellular signals is often aggregated before being presented in the Figures and Supplementary Figures. For example, there are multiple endothelial signals in the mature kidney, but for simplicity and readability we have combined them throughout this study.

Quantification of goodness of fit

One of the central aims of cell signal analysis compared to deconvolution methods is to allow for the possibility of a mismatch between the reference and the cells present in the bulk transcriptome. We allow for this through the inclusion of an intercept term in the model and quantify the mismatch by the relative contribution of the intercept and the fraction of the variance explained by the reference as measured by pR^2. To understand the intuition behind this choice it is useful to consider a simple linear model with and without an intercept.

Supposing we have bulk transcriptomes composed of two cell populations A and B. As a reference we have single cell transcriptomes from population A, but not B. Let us now consider how a linear model with and without an intercept term behaves. That is, for each bulk transcriptome we fit two models:

$$Y_g = \beta_A R_g \qquad (1)$$

$$Y_g = \beta_A R_g + \beta_0 \qquad (2)$$

Where Y_g is the expression in the bulk transcriptome for gene g, R is the reference signal at gene g derived from single cell transcriptomes population A, and the ß terms are estimated using linear regression. If Y_g is composed only of cells from population A, models (1) and (2) are identical (i.e., $ß_0$=0). Let us now consider what happens if a small number of cells from population B are added to Y_g. Further assume that cell types A and B are unrelated cell types with uncorrelated transcriptomes. The maximum likelihood estimator for beta in model (1) is:

$$\hat{\beta}_A = \frac{< YR >}{< R^2 >}$$

where < > denotes an average across all genes. For model (2) the maximum likelihood values of ß are:

$$\hat{\beta}_A = \frac{< YR > - < Y >< R >}{< R^2 > - < R >^2} = \frac{\text{cov}(Y,R)}{\text{cov}(R,R)}$$

$$\hat{\beta}_0 = < Y > - \hat{\beta}_A < R >$$

where cov(...) indicates the covariance of the two variables. Write $Y_g = n_A R_g + n_B S_g$ $Y_g = n_A R_g + n_B S_g$, where S_g is the cell signal reference for population B and n_A and n_B give the number of cells from each population. It can then be shown that the difference in ß in the simple population A only model (n_B=0) and the full model with both populations is the same as fitting a model with n_A=0. That is,

$$\Delta\hat{\beta}_A = n_B\left(\frac{< SR > - < S >< R >}{< R^2 > - < R >^2}\right) = \frac{\mathrm{cov}(S,R)}{\mathrm{cov}(R,R)}$$

$$\Delta\hat{\beta}_0 = n_B\left(< S > - \Delta\hat{\beta}_A < R >\right)$$

Where delta represent the difference between the model with and without a contribution from population B (i.e., the difference between models where n_B=0 and n_B>0). The equivalent formula for model (1) is,

$$\Delta\hat{\beta}_A = n_B\left(\frac{< SR >}{< R^2 >}\right)$$

From this it can immediately be seen that the effect of adding a cell population not accounted for in the reference (population B in this example) in model (1) is to increase the contribution from the unrelated population present in the reference (population A, $\Delta ß_A$>0). By contrast, for model (2) the contribution from the unrelated reference population (A) remains unchanged ($\Delta ß_A$=0) and the intercept term is increased to model the unaccounted-for population (B).

Of course, our model is not a linear model, but is based on a constrained generalized linear model. Furthermore, the above argument assumes that population B is unrelated to A. If population B has some correlation/covariance with A, then the fit for the population with which it is correlated will be increased by an amount proportional to the degree of correlation (see formula above). However, it is straightforward to show that (assuming Y and R are strictly positive) model (2) will always increase $ß_A$ by less than model (1). Thus, the inclusion of the intercept term will always improve the ability of the model to handle mismatches between the reference cell signals and the transcriptomes of the cells that make up each bulk transcriptome.

One thing the above toy example makes clear is that while the intercept term accounts for and quantifies variation not in the reference to an extent, it is not a complete solution. To further identify cases where there is a mismatch between the reference and the observed data we make use of the goodness of fit statistic, pR^2 define above. Continuing with the linear model analogy, this metric will identify cases where Y is modified by a perturbation δ that can both increase and decrease expression. An example of this would be the changes to a normal cell's transcriptome as it transforms into a cancer. Suppose that the perturbation was not correlated with the reference and has an average value of zero. It follows that the effect of this perturbation on the model fit parameters is,

$$\Delta\hat{\beta}_A = \frac{\mathrm{cov}(\delta, R)}{\mathrm{cov}(R, R)} = \frac{0}{\mathrm{cov}(R, R)} = 0$$

$$\Delta\hat{\beta}_0 = n_B\left(< \delta > - \Delta\hat{\beta}_A < R > \right) = n_B(0 - 0 < R >) = 0$$

However, although the fit does not change, the perturbation will decrease the total likelihood of the model and increase the value of pR^2. In combination, the intercept term and pseudo R-squared metric provide a quantification of the mismatch between reference cell signals and bulk transcriptomes

Benchmarking

We compared our method to two other methods: MuSiC[15] and BSeq-SC[14]. In both cases we used the default settings recommended by each method. As MuSiC requires a reference containing multiple cells derived from multiple samples, we were unable to create include leucocytes as part of our reference panel for **Fig. 1D-E** for MuSiC. BSeq-SC required marker genes for each population in the reference. To generate markers for each reference population we identified genes significantly enriched in the target population with a binomial test, using the quickMarkers function in the SoupX R package[43]. We further refined this set of markers by requiring that markers be expressed in at least 40% of cells within the cluster they mark and less than 5% of all other clusters. With these requirements, some clusters, notably the PT1 cluster, did not have any marker genes and so received a value of 0 in the BSeq-SC fit.

Calibration of intercept term

To assess the range of contributions from the intercept term (i.e., the "unexplained signal") when no appropriate reference is provided we constructed a series of inappropriate fits. In each case we selected a cell signal reference set that we knew was inappropriate to the set of samples being considered. The range of intercept values in this fit then gives a quantitative range that is indicative of how much weight is given to the intercept when no appropriate reference is present.

Choice of Poisson distribution

The choice of the Poisson distribution as the likelihood model at first glance seem a curious one, given that the Negative Binomial distribution (of which the Poisson distribution is a specific case) is widely used to model both bulk RNA-seq and single cell RNA-seq data. However, some reflection reveals that this choice is actually well justified.

We wish to evaluate the probability of observing a particular number of fragment counts in a bulk RNA-seq experiment, given that this experiment is composed of the addition of signals from a collection of single cell derived transcriptomic signals. To do this, we need to know how likely a particular set of fragment counts is, given a fixed contribution from each of the single cell signals.

Let us assume that the number of counts for a gene g in cell type c in a single cell RNA-seq experiment (with a fixed number of reads) can be well modelled by a negative binomial distribution with mean μ and over-dispersion φ. That is, the variance of this distribution is given by,

$$\sigma_{gc}^2 = \mu_{gc} + \mu_{gc}^2 \phi_{gc}$$

The distribution we are interested in, is then the distribution resulting from the sum of $\{N_0, N_1, N_2, ..., N_k\}$ random samples from the set of negative binomial distributions representing cell types {1,2,...,k}. That is, the distribution we are interested in is given by the sum of negative binomial distributions.

The moment generating function for the compound distribution is then given by,

$$M_{\text{comp}}(t) = \prod_{c \in C} \left(1 + \mu_c \phi_c (1 - e^t)\right)^{-\frac{1}{\phi_c}}$$

where C is the set of signals summed to form the compound distribution. This moment generating function completely specifies the compound distribution. However, we can use the method of moments to approximate this compound distribution with another Negative Binomial distribution with mean μ and over-dispersion φ. To do this, observe that the first and second moment of the compound distribution are,

$$E(X) = \sum_{c \in C} \mu_c$$

$$E(X^2) = \sum_{c \in C} \mu_c + \left(\sum_{c \in C} \mu_c\right)^2 + \sum_{c \in C} \mu_c^2 \phi_c$$

while the first and second moments of a negative binomial distribution with mean μ and over-dispersion φ are,

$$E(X) = \mu$$

$$E(X^2) = \mu + \mu^2 + \mu^2 \phi$$

Using the method of moments, this implies that the negative binomial approximation to the compound distribution has,

$$\mu = \sum_{c \in C} \mu_c$$

$$\phi = \sum_{c \in C} \left(\frac{\mu_c}{\mu}\right)^2 \phi_c$$

That is, the mean of the compound distribution equals the sum of the means of each component distribution (as expected). The over-dispersion of the compound

distribution is equal to the weighted sum of the component distributions. Closer consideration of the equation for the compound over-dispersion reveals that the over-dispersion of the compound distribution is almost always considerably less than the average over-dispersion of its component distributions.

For example, consider the case where all distributions have approximately the same mean and rewrite the compound over-dispersion as,

$$\phi = \sum_{c \in C} \left(\frac{\mu_c}{< \mu_c >} \right)^2 \frac{\phi_c}{N^2}$$

where angle brackets denote an average and N is the number of elements in C. Assuming the ratio in brackets is close to 1 gives,

$$\phi = \frac{< \phi_c >}{N}$$

So, in the case of distributions with similar means, the over-dispersion of the compound distribution is always N times less than the mean over-dispersion of the individual distributions. Consequently, as the number of distributions being summed over increases, the over-dispersion goes to zero and the compound Negative Binomial distribution approaches a Poisson distribution. The more general case where the means of the component distributions are not all similar is more complex, but in the limit of many distributions, the compound over-dispersion still approaches 0.

This result justifies the use of a Poisson distribution as the likelihood model in our fitting procedure. Although the individual signals from which the fit is derived are negative binomially distributed, the distribution of their sum is Poisson distributed. It may be that an extension of the Poisson model used here may prove useful, to model effects such as uncertainty in the effective length of genes for example, but it is not required to accurately represent the compound distribution on which our model depends.

Data availability

Raw nucleotide sequences for single cell data newly generated for this study are available in the European Genome-Phenome Archive under restricted access with accession codes EGAD00001004304 (CMN), EGAD00001006296 (MRT organoids), EGAD00001007498 (CCSK CEL-Seq2), and EGAD00001007572 (Wilms, other CCSK, and other MRT), access can be obtained by contacting the Data Access Committees EGAC00001001146 (CCSK CEL-Seq2) or EGAC00001000205 (everything else). We additionally utilised publicly available single cell kidney data from previous work[1,8,54]. The raw nucleotide data for which are available in the European Genome-Phenome Archive under restricted access with accession codes EGAS00001002171, EGAS00001002486, EGAS00001002325, and EGAS00001002553, and in the Human Cell Atlas Data Portal with project ID abe1a013-af7a-45ed-8c26-f3793c24a1f4.

Raw nucleotide sequences for bulk transcriptomic data newly generated for this study are available in the European Genome-Phenome Archive under restricted access with accession codes EGAS00001002487 and EGAS00001002534, access can be obtained by contacting the Data Access Committee EGAC00001000205. We also utilised publicly available bulk transcriptomes from: congenital mesoblastic nephroma[22], Wilms tumor[55], fetal kidney[56], the Therapeutically Applicable Research to Generate Effective Treatments (https://ocg.cancer.gov/programs/target) initiative, phs000218, available at https://portal.gdc.cancer.gov/projects, data generated by the TCGA Research Network, available at https://www.cancer.gov/tcga, and The Genotype-Tissue Expression (GTEx) Project as mapped by the recount2 project[38].

Mapped count data (i.e., tables of counts) are available as **Supplementary Data 2** (bulk transcriptomes) and **Supplementary Data 3** (single cell transcriptomes). This includes both newly generated data in this study and data obtained from public repositories. Sample metadata, including references to the source from which this data was obtained are listed for each unit of data in **Supplementary Data 1** (bulk transcriptomes) and **Supplementary Table 2** (single cell transcriptomes). All patient samples generated in this study are listed in **Supplementary Table 5**. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

Code availability

As an annex to the supplementary methods, we have provided all source code used in generating the results, figures, and tables used in this study as Supplementary Software 1. The purpose of these code files is to provide additional details as to how we implemented the analyses described in the Methods section. We provide the code necessary to run cellular signal analysis, along with some documentation and an example dataset online at https://github.com/constantAmateur/cellSignalAnalysis.

Acknowledgements

We acknowledge funding from Wellcome (Fellowships to S.B., K.S., J.C.A. 209328/Z/17/Z, core funding to the Sanger Institute, strategic award 211276/Z/18/Z). Additional support was provided by the St Baldrick's Foundation (S.B.), Great Ormond Street Hospital Children's Charity (J.C.A.), Great Ormond Street Hospital Biomedical Research Centre, Olivia Hodson Cancer Fund (K.S.), CRUK (Fellowship to T.J.M.), National Institute for Health Research funded Cambridge Biomedical Research Centre (R.A.B., A.Y.W). The views expressed are those of the authors and not necessarily those of the National Health Service, National Institute for Health Research, or Department of Health.

The CUH adult renal cancer sampling had infrastructure support from the Urological Malignancies Programme which is part of the CRUK Cambridge Centre, funded by Cancer Research UK Major Centre Award C9685/A25117, and supported by the NIHR

Cambridge BRC. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. F.C.H. acknowledges funding from the ERC advanced grant DynaMech-671174. We thank the Máxima FACS facility for sorting, Single Cell Discoveries for library preparations, and Single Cell Genomics facilities for help with scRNA analysis. J.D. acknowledges funding from the European Research Council (ERC) starting grant 850571, Dutch Cancer Society (KWF/Alpe d'HuZes Bas Mulder Award; KWF/Alpe d'HuZes, #10218), Foundation Children Cancer Free (KiKa #338, L.C.), Oncode Institute. We thank Dr Amos Burke for his contribution. We are grateful to our patients, young and old, for participating in our study.

Author Contributions

M.D.Y. and S.B. conceived of the experiment and wrote the manuscript. M.D.Y. performed analyses, aided by T.J.M., E.K., G.K., and T.H.H.C. I.D.V. and J.C.A. provided expertise on adrenal gland analysis. L.C. performed organoid experiments with F.A.V.B. T.R.W.O., N.S., D.R., N.C., L.H., R.R.K., A.W. provided pathological expertise. F.C., M.M.H.E., and A.S. provided clinical data. A.P., E.B.B., F.M., C.T., C.B., G.D.S., V.J.G., M.H., M.K., S.M.P., O.A.B., K.R., K.K, F.C.H., J.D., F.C.H., E.P., K.A., contributed to discussions and / or data. S.A.T., T.M., F.C.H., F.M., J.D., R.R.K, contributed fetal and tumor single cell data, together with K.B.M., R.A.B., X.H., A.W.C, L.M. S.B. and M.D.Y. directed the study, in conjunction with K.S. (single cell cancer work) and J.D. (organoid work).

Competing interests

The authors declare no competing interests.

References

1. Young, M. D. *et al.* Single cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).

2. A, B., P, W. & Jc, Z. SnapShot: TCGA-Analyzed Tumors. *Cell* vol. 173 https://pubmed.ncbi.nlm.nih.gov/29625059/ (2018).

3. International Cancer Genome Consortium *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).

4. Behjati, S., Lindsay, S., Teichmann, S. A. & Haniffa, M. Mapping human development at single-cell resolution. *Dev. Camb. Engl.* **145**, (2018).

5. A, R. *et al.* The Human Cell Atlas. *eLife* vol. 6 https://pubmed.ncbi.nlm.nih.gov/29206104/ (2017).

6. Lake, B. B. *et al.* A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nat. Commun.* **10**, 2832 (2019).

7. Arazi, A. *et al.* The immune cell landscape in kidneys of patients with lupus nephritis. *Nat. Immunol.* **20**, 902–914 (2019).

8. Hochane, M. *et al.* Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development. *PLoS Biol.* **17**, (2019).

9. Wang, P. *et al.* Dissecting the Global Dynamic Molecular Profiles of Human Fetal Kidney Development by Single-Cell RNA Sequencing. *Cell Rep.* **24**, 3554-3567.e3 (2018).

10. Menon, R. *et al.* Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Dev. Camb. Engl.* **145**, (2018).

11. Brunskill, E. W. *et al.* Atlas of gene expression in the developing kidney at microanatomic resolution. *Dev. Cell* **15**, 781–791 (2008).

12. Kildisiute, G. *et al.* Report – Tumor to normal single cell mRNA comparisons reveal a pan-neuroblastoma cancer cell. *bioRxiv* 2020.06.22.164301 (2020) doi:10.1101/2020.06.22.164301.

13. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).

14. M, B. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell systems* vol. 3 https://pubmed.ncbi.nlm.nih.gov/27667365/ (2016).

15. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).

16. Kang, K. *et al.* A novel computational complete deconvolution method using RNA-seq data. *bioRxiv* 496596 (2018) doi:10.1101/496596.

17. Medeiros, B. C. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N. Engl. J. Med.* **360**, 1787; author reply 1787-1788 (2009).

18. Monaco, G. *et al.* RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep.* **26**, 1627-1640.e7 (2019).

19. Ricketts, C. J. *et al.* The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep.* **23**, 313-326.e5 (2018).

20. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).

21. Vujanć, G. M. & Charles, A. K. Renal tumours of childhood: an update. *Pathology (Phila.)* **40**, 217–227 (2008).

22. Wegert, J. *et al.* Recurrent intragenic rearrangements of EGFR and BRAF in soft tissue tumors of infants. *Nat. Commun.* **9**, (2018).

23. Dome, J. S. *et al.* Advances in Wilms Tumor Treatment and Biology: Progress Through International Collaboration. *J. Clin. Oncol.* **33**, 2999–3007 (2015).

24. X, H. *et al.* Construction of a human cell landscape at single-cell level. *Nature* vol. 581 https://pubmed.ncbi.nlm.nih.gov/32214235/ (2020).

25. Chun, H.-J. E. *et al.* Genome-Wide Profiles of Extra-cranial Malignant Rhabdoid Tumors Reveal Heterogeneity and Dysregulated Developmental Pathways. *Cancer Cell* **29**, 394–406 (2016).

26. Calandrini, C. *et al.* An organoid biobank for childhood kidney cancers that captures disease and tissue heterogeneity. *Nat. Commun.* **11**, 1310 (2020).

27. O, I., Ap, L., C, J., Wg, K. & Ma, G. Negative regulation of hypoxia-inducible genes by the von Hippel-Lindau protein. *Proceedings of the National Academy of Sciences of the United States of America* vol. 93 https://pubmed.ncbi.nlm.nih.gov/8855223/ (1996).

28. Nm, T. *et al.* A phase 2 trial of sunitinib in patients with advanced non-clear cell renal cell carcinoma. *European urology* vol. 62 https://pubmed.ncbi.nlm.nih.gov/22771265/ (2012).

29. Davis, C. F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).

30. Lindgren, D. *et al.* Cell-Type-Specific Gene Programs of the Normal Human Nephron Define Kidney Cancer Subtypes. *Cell Rep.* **20**, 1476–1489 (2017).

31. Welsh-Bacic, D., Nowik, M., Kaissling, B. & Wagner, C. A. Proliferation of acid-secretory cells in the kidney during adaptive remodelling of the collecting duct. *PloS One* **6**, e25240 (2011).

32. Custers, L. *et al.* Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nat. Commun.* **12**, 1407 (2021).

33. Roberts, K. & Bayraktar, O. A. Automation of Multiplexed RNAscope Single-Molecule Fluorescent In Situ Hybridization and Immunohistochemistry for Spatial Tissue Mapping. in *In Situ Hybridization Protocols* (eds. Nielsen, B. S. & Jones, J.) 229–244 (Springer US, 2020). doi:10.1007/978-1-0716-0623-0_15.

34. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7**, e31657 (2018).

35. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **26**, 589–595 (2010).

36. Jones, D. *et al.* cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Curr. Protoc. Bioinforma.* **56**, 15.10.1-15.10.18 (2016).

37. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16910–16915 (2010).

38. Collado-Torres, L. *et al.* Reproducible RNA-seq analysis using recount2. *Nat. Biotechnol.* **35**, 319–321 (2017).

39. Nellore, A. *et al.* Rail-RNA: scalable analysis of RNA-seq splicing and coverage. *Bioinforma. Oxf. Engl.* **33**, 4033–4040 (2017).

40. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).

41. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

42. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281-291.e9 (2019).

43. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data. *bioRxiv* 303727 (2020) doi:10.1101/303727.

44. Batson, J., Royer, L. & Webber, J. Molecular Cross-Validation for Single-Cell RNA-seq. *bioRxiv* 786269 (2019) doi:10.1101/786269.

45. Maaten, L. van der. Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).

46. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

47. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv180203426 Cs Stat* (2018).

48. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).

49. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

50. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).

51. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *ArXiv160304467 Cs* (2016).

52. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs* (2017).

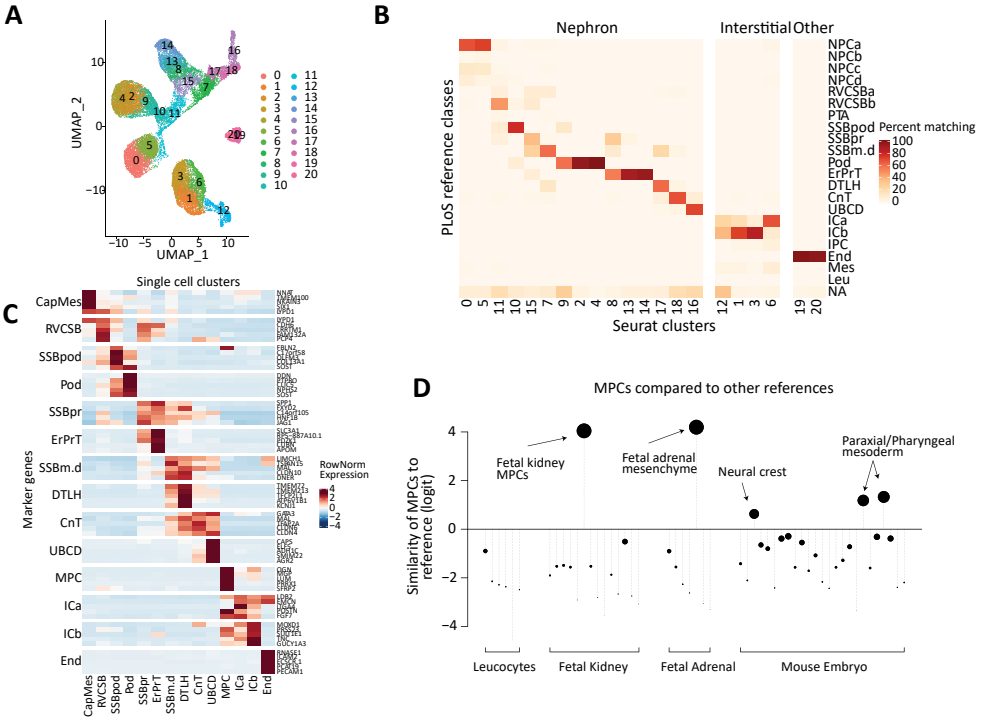
53. Mcfadden, D. *Frontiers in Economics*. (1974).

54. Stewart, B. J. *et al.* Spatiotemporal immune zonation of the human kidney. *Science* **365**, 1461–1466 (2019).

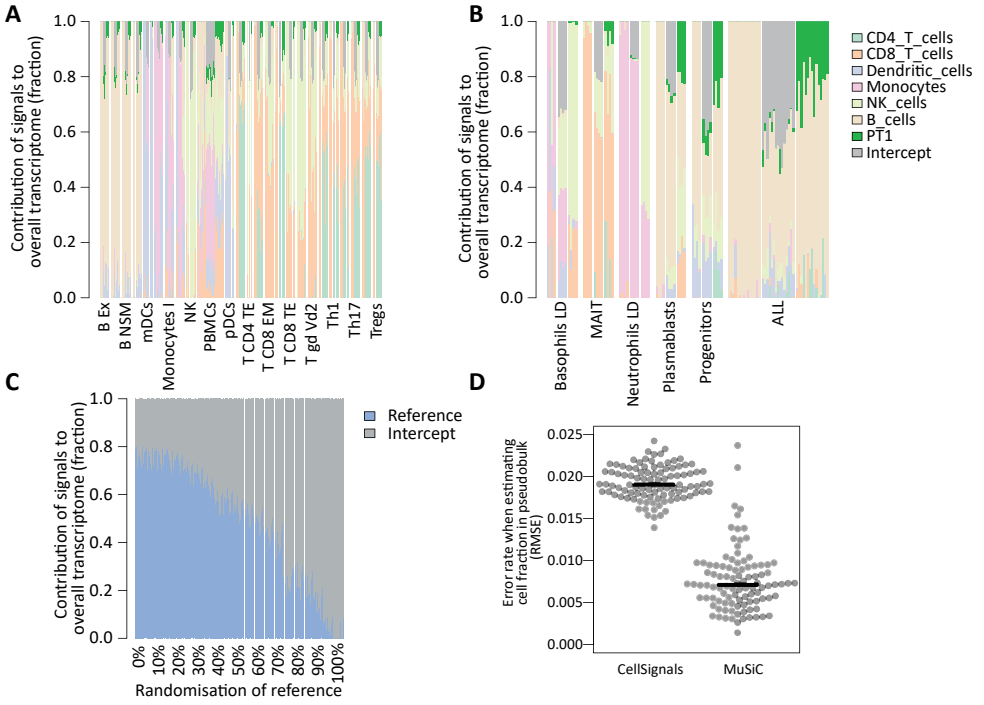
55. Coorens, T. H. H. *et al.* Embryonal precursors of Wilms tumor. *Science* **366**, 1247 (2019).

56. Gerrard, D. T. *et al.* An integrative transcriptomic atlas of organogenesis in human embryos. *eLife* **5**, e15657 (2016).

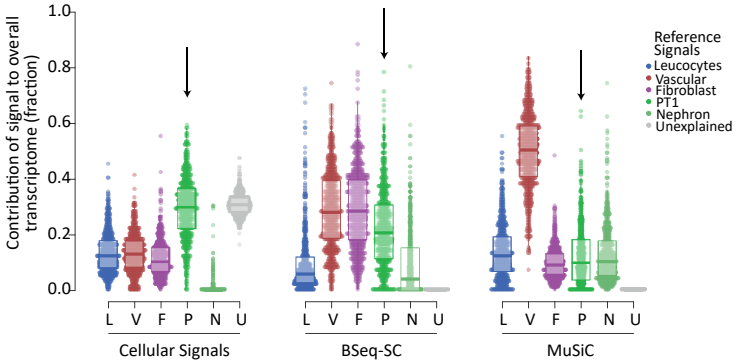
Supplementary information



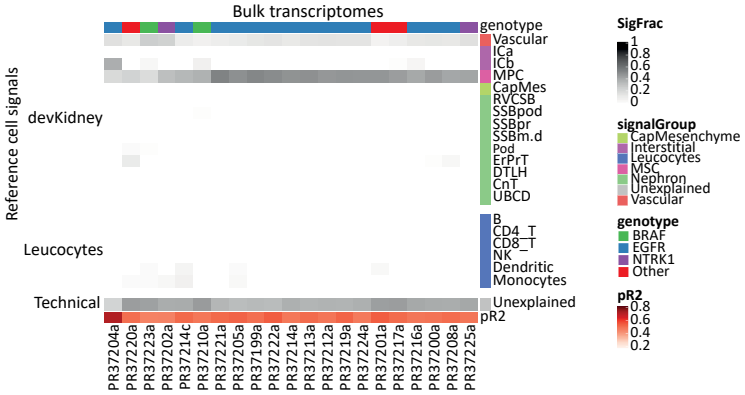
Supplementary Figure 1 | Fetal kidney reference map.
(A) UMAP of fetal kidney clusters – Reduced dimension representation (UMAP) of the transcriptomes of 21,834 fetal kidney cells where each point represents a cell and nearby cells have similar transcriptomes. Each cell is labelled by a color corresponding to the cluster to which it belongs, as indicated by the legend on the right. Additionally, a label for each cluster is placed at the mean position of cells belonging to each cluster. Clusters are generated as discussed in the methods section. **(B) Comparison of clusters to PLoS annotation** – Each cell is assigned an annotation from1 based on logistic regression, or NA if the similarity score is less than 1 (logit) or multiple populations have similarity greater than 1. The x-axis shows the clusters from **A,** the yaxis the reference populations and the color scale the fraction of cells assigned to each reference class. Note the high proportion of NAs in cluster 12. **(C) Marker genes defining cell types** – For each annotated cell population in **Fig. 1B**, the top 5 algorithmically determined marker genes are shown. The color scheme indicates the average normalized expression of the cells in the cluster indicated on the x-axis, z-scaled across all cell types to have mean 0 and standard deviation 1. **(D) Similarity of MPCs to other tissues** – Similarity score (logits) calculated by logistic regression trained on the fetal kidney reference to MPCs. Note that the extremely high similarity to the equivalent population in the fetal adrenal. Source data are available as a Source Data file.



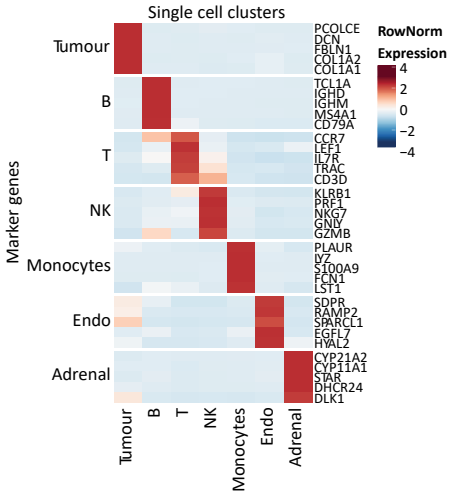
Supplementary Figure 2 | Benchmarking of cellular signal analysis.
(A) Benchmarking using bulk transcriptomes with match in reference – The same comparison as in **Fig. 1D** but where each reference cell signal is shown individiually rather than grouped into "matching" and "non-matching". Each bar represents a bulk transcriptome, with the different colors indicating the relative contribution from each reference cell signal (see legend). Transcriptomes are grouped into blocks of the same type as shown on the x-axis and within each block the results of deconvolution using BSeq-SC, cellular signal analysis, and MuSiC are shown in separate blocks from left to right. **(B) Benchmarking using bulk transcriptomes with no match in reference** – The same comparison as in **Fig. 1E** but where each reference cell signal is shown individiually rather than grouped into "matching" and "non-matching". Each bar represents a bulk transcriptome, with the different colors indicating the relative contribution from each reference cell signal (see legend). Transcriptomes are grouped into blocks of the same type as shown on the x-axis and within each block the results of deconvolution using BSeq-SC, cellular signal analysis, and MuSiC are shown in separate blocks from left to right. **(C) Response of unexplained signal (Intercept) to incomplete reference** – Bulk transcriptomes from flow sorted B cells are fit using cellular signal analysis using a reference consisting of a B cell signal only. This reference is then progressively randomized to decrease its suitability for the bulk B cell transcriptomes as indicated by the label on the x-axis. In each case, the resulting fraction of the signal that is allocated to either the modified reference, or the intercept term (the "unexplained signal") is shown by the size of the colored bars. **(D) Recovering composition of bulk tissue from bulk transcriptomes** – Using the same single cell PBMC data used as a reference in panels A and B, 100 pseudobulk transcriptomes were generated by random sampling. These pseudobulk transcriptomes were then deconvoluted with both cell signal analysis and MuSiC using the same single cell data as a reference. Finally, the abundance of cell types in each pseudobulk transcriptome was estimated and compared to the true proportions to calculate the root mean squared error (RMSE) for each sample. The RMSE values for all 100 simulations are shown on the y-axis, and the deconvolution method is shown on the x-axis. The black line indicates the median RMSE for each method. Source data are available as a Source Data file.



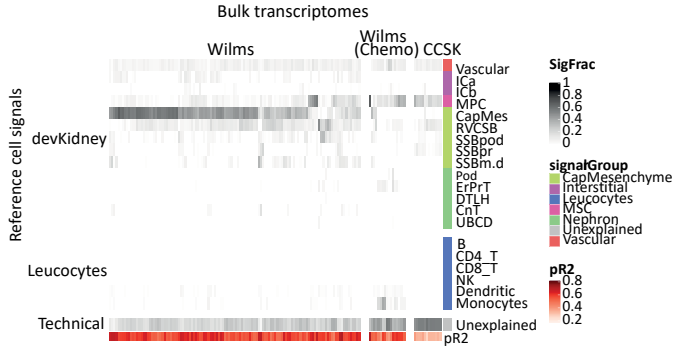
Supplementary Figure 3 | Ability of different methods to recover known RCC signal.
We applied our cellular signal method, BSEQ-sc and MuSiC to a population of clear cell and papillary cell renal cell carcinomas, using reference signals derived from the mature normal kidney. This population of tumors is known to resemble the PT1 population of proximal tubular cells and have a strong vascular component. The relative contribution of each signal to each bulk RNA-seq sample is shown by the y-axis. The results for the three methods are split into blocks, with cellular signals on the left, BSEQ-sc in the middle, and MuSiC on the right. Each signal type is labelled with an abbreviation and colored as shown by the legend on the right. Signals are marked with a square for fetal kidney and circle for mature kidney. For our method, there is an additional signal which represents the unexplained signal for each sample (see Methods). Contributions due to vascular signals, fibroblast signals, leucocyte signals, and all non-proximal tubular nephron signals are aggregated together. Each signal/sample combination is represented by a single point and the distribution of relative signal contributions to the bulk transcriptomes are summarized with boxplots. Arrows mark the PT1 population which these tumors are known to resemble. Boxplots show the distribution median (middle line), 1st and 3rd quartiles (box limits), and 1.5 times the inter-quartile range beyond the box-limits (whiskers). Source data are available as a Source Data file.



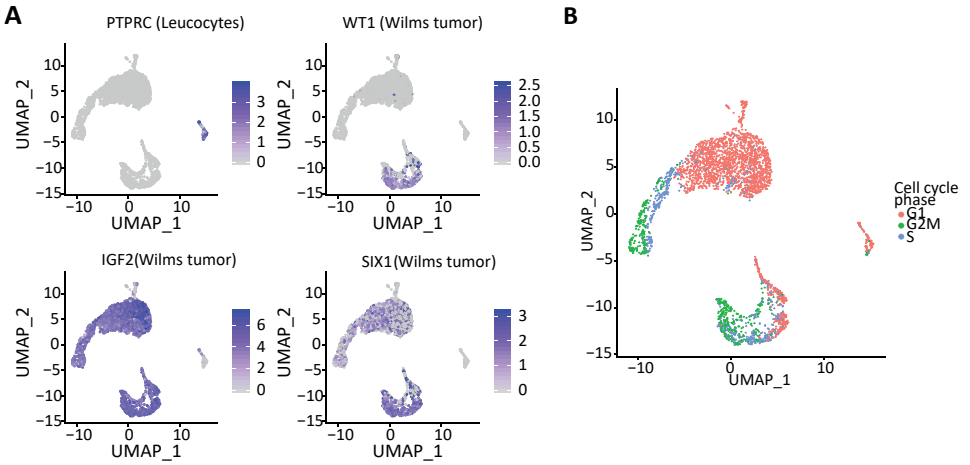
Supplementary Figure 4 | CMNs resembles interstitial cells.
The same data as in **Fig. 3A**, but presented in heatmap form. Each column represents a sample and each row a collection of reference signal. The shading in each cell represents the relative contribution of that collection of reference signals to explaining the bulk transcriptome of that sample. The CMN samples are then further sub-divided by genotype. Sample IDs are printed below each column. The pR2 column represents a pseudo-R squared value for each sample, calculated as 1 minus the ratio of the log likelihoods of the full model to a model consisting of only the intercept term.Source data are available as a Source Data file.



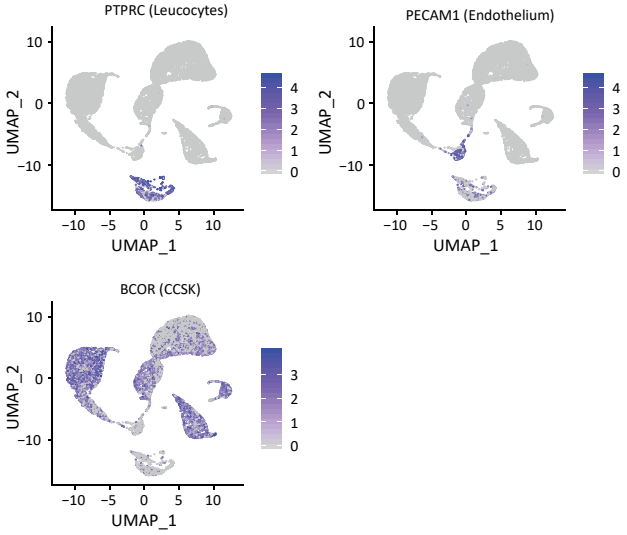
Supplementary Figure 5 | Key markers of cell types of CMN.
For each annotated cell population in **Fig. 3D**, the top 5 algorithmically determined marker genes are shown. The color scheme indicates the average normalized expression of the cells in the cluster indicated on the x-axis, z-scaled across all cell types to have mean 0 and standard deviation 1. Source data are available as a Source Data file.



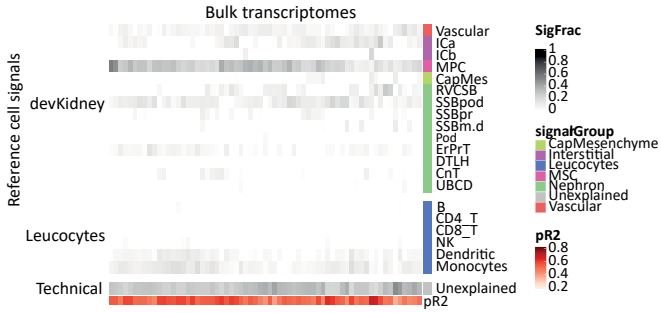
Supplementary Figure 6 | Wilms/Nephroblastoma and CCSK fit using fetal kidney signals.
The same data as in **Fig. 4A**, but presented in heatmap form. Each column represents a sample and each row a collection of reference signal. The shading in each cell represents the relative contribution of that collection of reference signals to explaining the bulk transcriptome of that sample. The pR2 column represents a pseudo-R squared value for each sample, calculated as 1 minus the ratio of the log likelihoods of the full model to a model consisting of only the intercept term. Source data are available as a Source Data file.



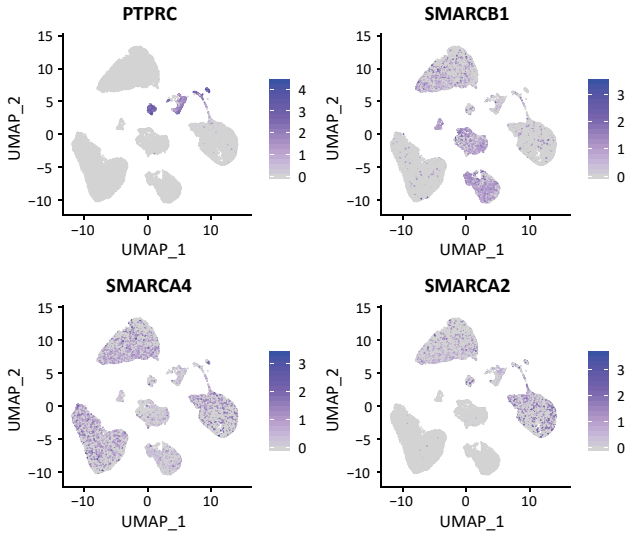
Supplementary Figure 7 | Annotation of single cell Wilms data.
(A) Expression of key markers – Reduced dimension representation (UMAP) of the transcriptomes of Wilms cells, colored by log-normalized expression of the gene in the panel title. The cell type that a gene is a marker of is shown in brackets. **(B) Phase of cell cycle** – The same data as panel **A**, but colored by inferred cell cycle phase as indicated by the legend on the right. Source data are available as a Source Data file.



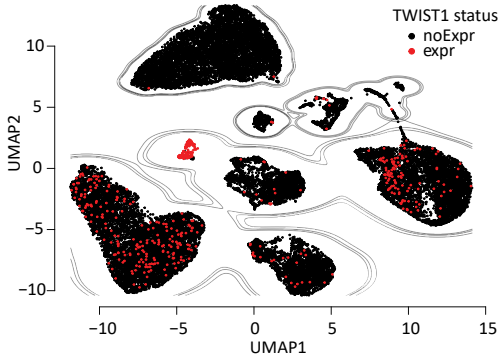
Supplementary Figure 8 | Annotation of single cell CCSK data.
Reduced dimension representation (UMAP) of the transcriptomes of CCSK cells, colored by log-normalized expression of the gene in the panel title. The cell type that a gene is a marker of is shown in brackets. Source data are available as a Source Data file.



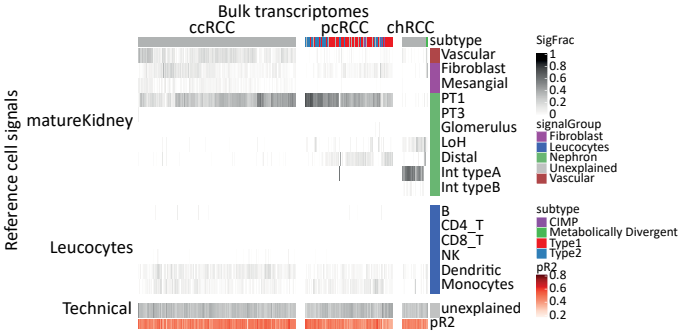
Supplementary Figure 9 | MRTs fit using fetal kidney signals.
The same data as in **Fig. 5A**, but presented in heatmap form. Each column represents a sample and each row a collection of reference signal. The shading in each cell represents the relative contribution of that collection of reference signals to explaining the bulk transcriptome of that sample. The pR2 column represents a pseudo-R squared value for each sample, calculated as 1 minus the ratio of the log likelihoods of the full model to a model consisting of only the intercept term. Source data are available as a Source Data file.



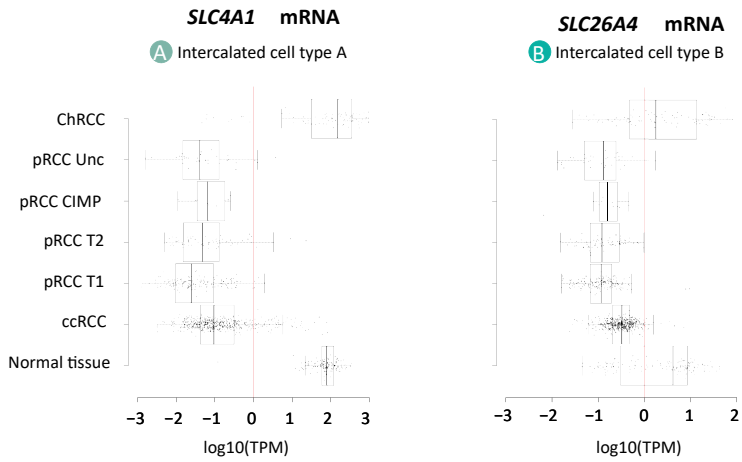
Supplementary Figure 10 | Annotation of single cell MRT data.
Reduced dimension representation (UMAP) of the transcriptomes of MRT cells, colored by log-normalized expression of the gene in the panel title. Source data are available as a Source Data file.



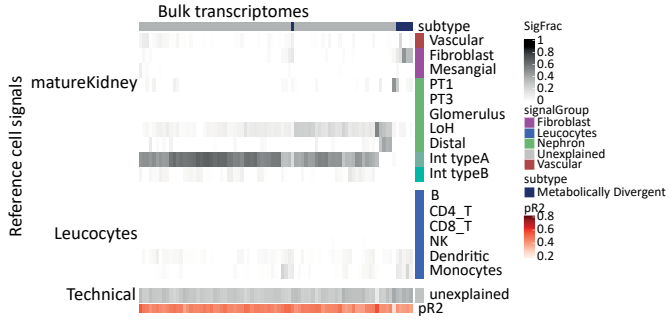
Supplementary Figure 11 | TWIST1 positive cells in MRT.
Reduced dimension representation (UMAP) of the transcriptomes of MRT cells (as in **Fig. 5B**), with cells with any expression of *TWIST1* colored red and those without colored black. Source data are available as a Source Data file.



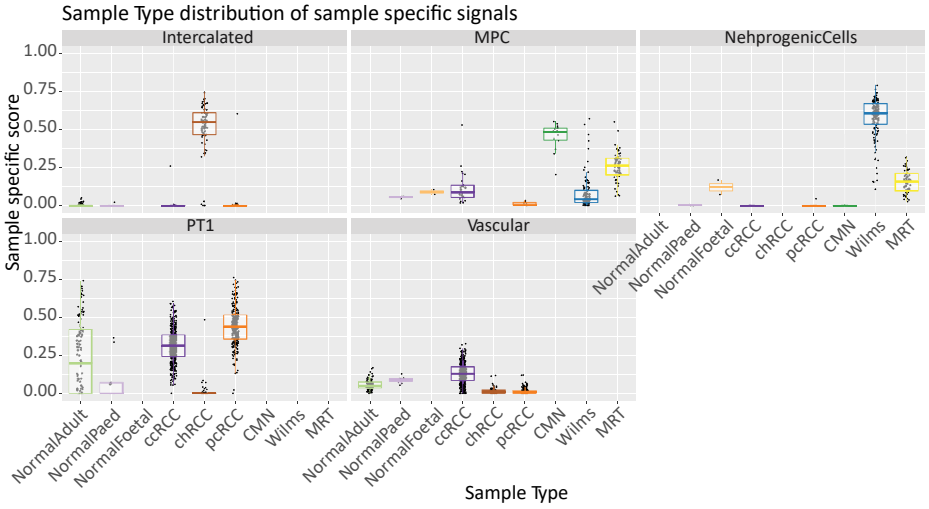
Supplementary Figure 12 | All RCCs fit using mature kidney signals.
The same data as in **Fig. 6A**, but presented in heatmap form. Each column represents a sample and each row a collection of reference signal. Subtypes of tumor are indicated by the annotation row at the top of the figure. The shading in each cell represents the relative contribution of that collection of reference signals to explaining the bulk transcriptome of that sample. The pR2 column represents a pseudo-R squared value for each sample, calculated as 1 minus the ratio of the log likelihoods of the full model to a model consisting of only the intercept term. Source data are available as a Source Data file.



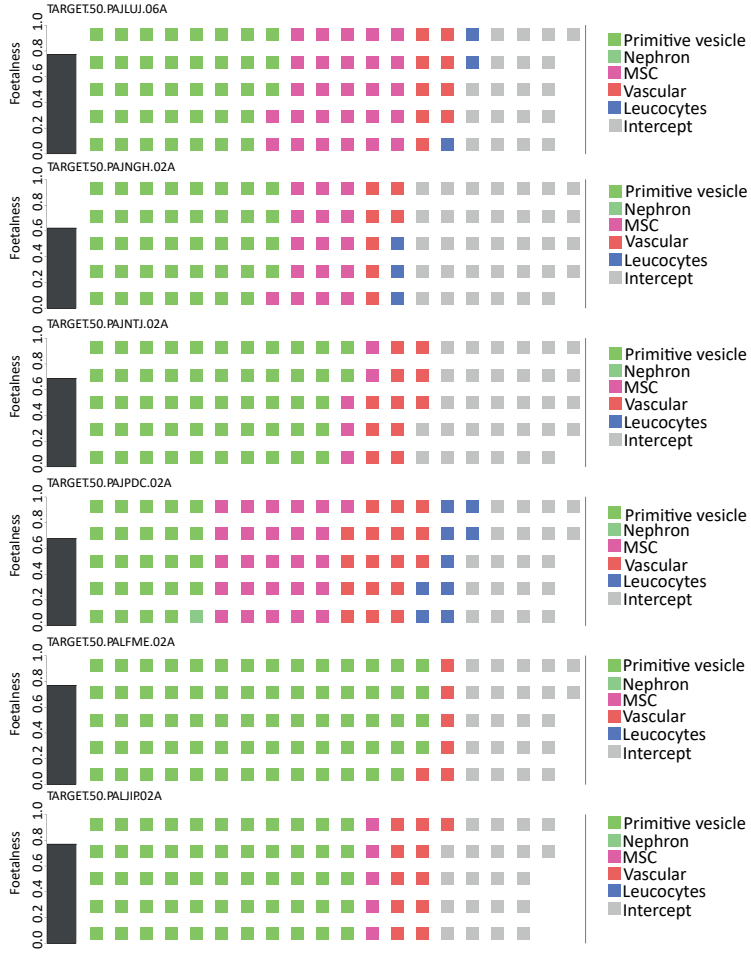
Supplementary Figure 13 | Expression of markers of Type A and B intercalated cells in bulk transcriptomes.
Boxplots showing expression of canonical markers of Type A (left) and B (right) intercalated cells in different populations of bulk kidney tumor and normal transcriptomes. Expression values are normalized to transcripts per million (TPM) and log10 transformed and shown on the x-axis, with each point representing a sample and a boxplot showing the distribution of expression values for that sample type. A red line marks 0 on the log transformed expression scale. Boxplots show the distribution median (middle line), 1st and 3rd quartiles (box limits), and 1.5 times the nter-quartile range beyond the box-limits (whiskers). Source data are available as a Source Data file.



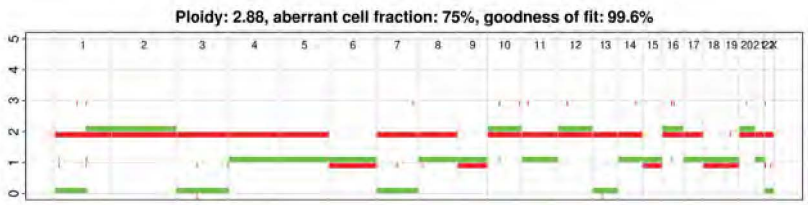
Supplementary Figure 14 | ChRCC fit using mature kidney signals.
The same data as in **Fig. 6E**, but presented in heatmap form. Each column represents a sample and each row a collection of reference signal. Subtypes of tumor are indicated by the annotation row at the top of the figure. The shading in each cell represents the relative contribution of that collection of reference signals to explaining the bulk transcriptome of that sample. The pR2 column represents a pseudo-R squared value for each sample, calculated as 1 minus the ratio of the log likelihoods of the full model to a model consisting of only the intercept term. Source data are available as a Source Data file.



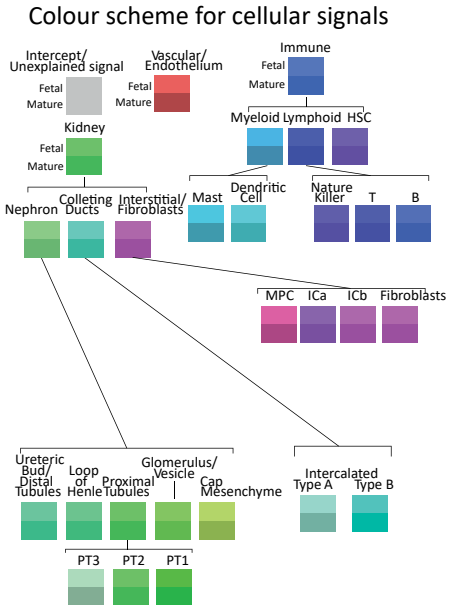
Supplementary Figure 15 | Specificity of signal contributions to tumor types.
Each panel shows the distribution of the contribution of different "scores" calculated by aggregating the contribution from various signals, across a range of different tumor and normal samples. Each point represents a sample, with its y-axis value giving its score, with the score type given by the panel. Samples are broken down by sample/tumor type as shown on the x-axis, with each sample type given a different color. Box-plots show the distribution of the scores within each sample type. This is a more detailed version of the data shown in **Fig. 7B**. Boxplots show the distribution median (middle line), 1st and 3rd quartiles (box limits), and 1.5 times the inter-quartile range beyond the box-limits (whiskers). Source data are available as a Source Data file.



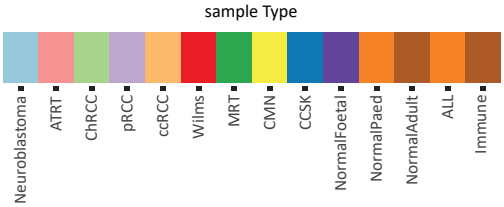
Supplementary Figure 16 | Additional validation of tumor type inference.
The fetalness score and cell signal contribution for 6 non-primary Wilms tumors calculated as in **Fig. 7D-G**. The contribution from each cell signal to each bulk transcriptome was multiplied by 100 and rounded down to the nearest integer, which then determined the number of squares shown for each sample. The type of signal each square represents is indicated by the legend on the right. Source data are available as a Source Data file.



Supplementary Figure 17 | CN profile of unknown childhood renal tumor.
Copy number profile of the unknown childhood renal tumor (SangerProject1700_PR40309a). The colored lines represent the major (red) and minor (green) alleles copy number (y-axis) along the chromosomes as shown at the top of the plot.



Supplementary Figure 18 | Color scheme for cell types.
Overview of the color scheme used throughout the figures in this paper to represent different cell types. Cell types are organized hierarchically, with fetal/ mature versions of the same cell type represented by colors of the same hue and saturation but different value. The exception to this rule is the unexplained signal color which is the same in both a fetal and mature context.



Supplementary Figure 19 | Color scheme for sample types.
Overview of the color scheme used throughout the figures in this paper to represent different sample types. Where colors have been duplicated they either never occur in the same context, or it is obvious from the context which color refers to which sample type.

Supplementary Table 1 | Single cell manifest.
A table giving a description of each single cell experiment included in this study. In the age column, PCW stands for post conception weeks. For each row, the raw count data is available in **Supplementary Data 3** in a folder with name given by the "channelLabel" column.

channelLabel	patientLabel	sampleType	disease	age	sort	dataSource	assay
4834STDY7002873	Foetus15	Normal	Normal	12 PCW	CD45+	DOI: 10.1126/science.aat1699	10X 3'
4834STDY7002874	Foetus15	Normal	Normal	12 PCW	CD45-	DOI: 10.1126/science.aat1699	10X 3'
4834STDY7002875	Foetus16	Normal	Normal	8+1 PCW	CD45+	DOI: 10.1126/science.aat1699	10X 3'
4834STDY7002876	Foetus16	Normal	Normal	8+1 PCW	CD45-	DOI: 10.1126/science.aat1699	10X 3'
4834STDY7002881	Foetus17	Normal	Normal	9+1 PCW	None	DOI: 10.1126/science.aat1699	10X 3'
4834STDY7002885	Foetus17	Normal	Normal	9+1 PCW	CD45+	DOI: 10.1126/science.aat1699	10X 3'
4834STDY7002886	Foetus17	Normal	Normal	9+1 PCW	CD45-	DOI: 10.1126/science.aat1699	10X 3'
FCAImmP7462242	Foetus35	Normal	Normal	7+6 PCW	CD45+	DOI: 10.1126/science.aat5031	10X 3'
FCAImmP7462243	Foetus35	Normal	Normal	7+6 PCW	CD45-	DOI: 10.1126/science.aat5031	10X 3'
FCAImmP7528292	Foetus38	Normal	Normal	13+6 PCW	CD45+	DOI: 10.1126/science.aat5031	10X 3'
FCAImmP7528293	Foetus38	Normal	Normal	13+6 PCW	CD45-	DOI: 10.1126/science.aat5031	10X 3'
FCAImmP7555849	Foetus41	Normal	Normal	16+6 PCW	CD45+	DOI: 10.1126/science.aat5031	10X 3'
FCAImmP7555850	Foetus41	Normal	Normal	16+6 PCW	CD45-	DOI: 10.1126/science.aat5031	10X 3'
FCAImmP7579214	Foetus45	Normal	Normal	13+6 PCW	CD45+	DOI: 10.1126/science.aat5031	10X 3'
FCAImmP7579215	Foetus45	Normal	Normal	13+6 PCW	CD45-	DOI: 10.1126/science.aat5031	10X 3'
w09	w09	Normal	Normal	NA	NA	DOI: 10.1371/journal.pbio.3000152	10X 3'
w11	w11	Normal	Normal	NA	NA	DOI: 10.1371/journal.pbio.3000152	10X 3'
w13	w13	Normal	Normal	NA	NA	DOI: 10.1371/journal.pbio.3000152	10X 3'
w16	w16	Normal	Normal	NA	NA	DOI: 10.1371/journal.pbio.3000152	10X 3'
w18	w18	Normal	Normal	NA	NA	DOI: 10.1371/journal.pbio.3000152	10X 3'
4602STDY7685338	PD42187	Tumour	CMN	NA	None	EGAS00001002325	10X 3'
4602STDY7685339	PD42187	Tumour	CMN	NA	None	EGAS00001002325	10X 3'
5640STDY7891141	MRT1	NormalOrganoid	MRT	NA	None	EGAS00001003386	10X 3'
5640STDY7891142	MRT1	TumourOrganoid	MRT	NA	None	EGAS00001003386	10X 3'
NB8350521	PD48777	Tumour	Wilms	NA	None	EGAD00001007572	10X 3'
NB8350522	PD48777	Tumour	Wilms	NA	CD45-	EGAD00001007572	10X 3'
NB8350523	PD48777	Tumour	Wilms	NA	CD45-	EGAD00001007572	10X 3'
NB8368876	PD48777	Tumour	Wilms	NA	None	EGAD00001007572	10X 3'
NB8368877	PD48777	Tumour	Wilms	NA	CD45-	EGAD00001007572	10X 3'
NB8711738	CCSK1	Tumour	CCSK	NA	NA	EGAD00001007572	Nuclei 10X 3'
NB8711739	CCSK1	Tumour	CCSK	NA	NA	EGAD00001007572	Nuclei 10X 3'
NB8711740	CCSK2	Tumour	CCSK	NA	NA	EGAD00001007572	Nuclei 10X 3'
NB8711741	CCSK2	Tumour	CCSK	NA	NA	EGAD00001007572	Nuclei 10X 3'
NB8711742	CCSK3	Tumour	CCSK	NA	NA	EGAD00001007572	Nuclei 10X 3'
TM211	1015T1	Tumour	CCSK	NA	NA	EGAD00001007498	CEL -Seq2
TM212	1015T1	Tumour	CCSK	NA	NA	EGAD00001007498	CEL -Seq2
NB8350518	PD47704	Tumour	MRT	NA	None	EGAD00001007572	10X 3'
NB8350519	PD47704	Tumour	MRT	NA	CD45-	EGAD00001007572	10X 3'
NB8350520	PD47704	Tumour	MRT	NA	CD45-	EGAD00001007572	10X 3'
NB8368300	PD47704	Tumour	MRT	NA	None	EGAD00001007572	10X 3'
NB8368301	PD47704	Tumour	MRT	NA	CD45-	EGAD00001007572	10X 3'
NB8711736	MRT3	Tumour	MRT	NA	NA	EGAD00001007572	Nuclei 10X 3'
NB8711737	MRT3	Tumour	MRT	NA	NA	EGAD00001007572	Nuclei 10X 3'

Supplementary Table 2 | Linear model of immaturity by genotype.
Generated using "summary(fit)" in R, where fit is a linear model to predict the immaturity score of RCCs using sample genotype as covariates. Covariates are coded such that the intercept term corresponds to the wild type of non-tumor biopsy.

coefficient	Estimate	Std. Error	t value	Pr(>|t|)	qVal
(Intercept)	0.043420795	0.007086455	6.127293879	1.34E-09	1.05E-08
splitccRCC	0.097857036	0.007328719	13.35254347	3.22E-37	1.16E-35
splitpRCC CIMP	0.142529785	0.023561658	6.049225545	2.14E-09	1.28E-08
splitpRCC T1	0.108900758	0.008940814	12.1801836	1.09E-31	1.97E-30
splitpRCC T2	0.0410931	0.010917653	3.763913481	0.00017818	0.000916353
splitpRCC Unc	0.085075024	0.0139183	6.112458078	1.46E-09	1.05E-08
splitchRCC	-0.016045051	0.010752786	-1.492176164	0.13600536	0.30601206
splitchRCC MD	0.271093348	0.041082107	6.598818053	7.09E-11	8.51E-10
hitTypeTP53singleHit	0.021047117	0.011677218	1.802408447	0.0718175	0.215452499
hitTypeTP53doubleHit	0.000687359	0.021635183	0.031770449	0.974662196	0.974662196
hitTypeBAP1singleHit	-0.013486051	0.008870115	-1.520391994	0.128765671	0.30601206
hitTypeBAP1doubleHit	-0.069745425	0.049447887	-1.410483414	0.158744444	0.317488888
hitTypeCDKN2AsingleHit	0.024653553	0.025087408	0.982706237	0.32601761	0.510288433
hitTypeVHLsingleHit	0.006883926	0.006049754	1.137885346	0.255472876	0.459851177
hitTypeVHLdoubleHit	-0.007308534	0.019379223	-0.377132486	0.706164553	0.85176515
hitTypePBRM1singleHit	0.002144084	0.006294524	0.340626871	0.733464435	0.85176515
hitTypePBRM1doubleHit	-0.006196931	0.029184362	-0.212337396	0.831892171	0.907518732
hitTypeSETD2singleHit	0.016482962	0.008195675	2.011178013	0.044606071	0.145983506
hitTypeSETD2doubleHit	-0.008720474	0.025086187	-0.347620563	0.728206894	0.85176515
hitTypeKDM5CsingleHit	-0.015657612	0.011737054	-1.334032499	0.182532552	0.345851151
hitTypeKDM5CdoubleHit	0.103279661	0.050626133	2.040046403	0.04163883	0.145983506
hitTypePTENsingleHit	0.009161375	0.011702299	0.782869655	0.43391075	0.650866126
hitTypePTENdoubleHit	0.136496255	0.040337911	3.38382063	0.000745889	0.003356502
hitTypeMTORsingleHit	-0.007429789	0.011244737	-0.66073476	0.508952457	0.732891538
hitTypeMTORdoubleHit	0.023895949	0.049156597	0.486118856	0.627001896	0.806145295
hitTypePIK3CAsingleHit	-0.017528165	0.016367274	-1.070927538	0.284490719	0.487698376
hitTypePIK3CAdoubleHit	-0.119272346	0.071174849	-1.675765351	0.094133438	0.260677213
hitTypeMETsingleHit	0.013717814	0.013809371	0.993369928	0.320798196	0.510288433
hitTypeMETdoubleHit	0.010675768	0.049227288	0.216866867	0.828361523	0.907518732
hitTypeFAT1singleHit	0.000809979	0.012684051	0.063858047	0.949097517	0.974662196
hitTypeNF2singleHit	0.036993869	0.017774285	2.081313973	0.037688938	0.145983506
hitTypeKDM6AsingleHit	-0.030125498	0.019992242	-1.506859435	0.132199554	0.30601206
hitTypeKDM6AdoubleHit	-0.003423503	0.050729383	-0.067485599	0.946210199	0.974662196
hitTypeSMARCB1singleHit	-0.028075701	0.019652678	-1.428594183	0.153469791	0.317488888
hitTypeNFE2L2singleHit	-0.009935681	0.018581713	-0.534702086	0.592988685	0.790651579
hitTypeSTAG2singleHit	-0.009271449	0.016090918	-0.576191427	0.564630661	0.7817963

Supplementary Table 3 | Linear model of immaturity by other features.
Generated using "summary(fit)" in R, where fit is a linear model to predict the immaturity score of ccRCCs using sample genotype as covariates. Covariates are coded such that the intercept term corresponds to unknown mRNA group, unknown miRNA group, grade G1, and stage I.

coefficient	Estimate	Std. Error	t value	Pr(>|t|)	qVal
(Intercept)	0.163510869	0.018861217	8.669158019	6.91E -17	1.11E -15
mRNA1	0.023556367	0.010863869	2.168322037	0.030629906	0.08167975
mRNA2	0.001214761	0.011694473	0.103874825	0.917312529	0.993826308
mRNA3	0.061739964	0.011961718	5.161463027	3.61E -07	2.89E -06
mRNA4	-0.028420201	0.01216071	-2.3370511	0.019851151	0.068526977
miRNA1	0.033287926	0.01161994	2.864724376	0.004358656	0.023246165
miRNA2	-0.008154287	0.011496292	-0.709297152	0.478488096	0.695982686
miRNA3	-0.010115854	0.010740307	-0.941858989	0.346743598	0.554789757
miRNA4	0.015538685	0.011823078	1.314267285	0.18939056	0.336694329
gradeG2	-0.025578429	0.019110589	-1.33844272	0.181392447	0.336694329
gradeG3	-0.044934636	0.019467088	-2.308236131	0.02141468	0.068526977
gradeG4	-0.042089242	0.021361856	-1.970298941	0.049384495	0.112878847
gradeGX	-0.001147628	0.066465078	-0.017266634	0.986231154	0.993826308
stageStage II	0.000917833	0.010061156	0.091225427	0.927351929	0.993826308
stageStage III	-9.37E -05	0.007659532	-0.012228755	0.990248243	0.993826308
stageStage IV	7.37E -05	0.009513506	0.007741727	0.993826308	0.993826308

Supplementary Table 4 | Mast cell fraction for RCCs
Quantification of mast cell prevalence from smFISH of different tumors. In the column "tType" pRCC1/2 represents papillary cell renal cell carcinoma type 1/2.

tType	nCells	nMast	mastFrac
pRCC2	1215037	8598	0.007076328
ccRCC	645654	48862	0.075678304
pRCC1	686715	57156	0.083231035

Supplementary Table 5 | Patient samples
A list of all the patient samples used in this study, giving disease type and assay performed.

ID	Disease	Assay
PD42187	CMN	SingleCellTranscriptomics
MRT1	MRT	SingleCellTranscriptomics
PD48777	Wilms	SingleCellTranscriptomics
CCSK1	CCSK	SingleCellTranscriptomics
CCSK2	CCSK	SingleCellTranscriptomics
CCSK3	CCSK	SingleCellTranscriptomics
1015T1	CCSK	SingleCellTranscriptomics
PD47704	MRT	SingleCellTranscriptomics
MRT3	MRT	SingleCellTranscriptomics
PR36165	Wilms	BulkTranscriptomics
PR37104	ccRCC	BulkTranscriptomics
PR37228	ccRCC	BulkTranscriptomics
PR37272	Wilms	BulkTranscriptomics
PR37276	Wilms	BulkTranscriptomics
PR40710	Wilms	BulkTranscriptomics
PR40712	Wilms	BulkTranscriptomics
PR40713	Wilms	BulkTranscriptomics
PR40715	Wilms	BulkTranscriptomics
PR40716	Wilms	BulkTranscriptomics
PR40717	Wilms	BulkTranscriptomics
PR40718	Wilms	BulkTranscriptomics
PR40719	Wilms	BulkTranscriptomics
PR40720	Wilms	BulkTranscriptomics
PR40722	Wilms	BulkTranscriptomics
PR40726	Wilms	BulkTranscriptomics
PR40727	Wilms	BulkTranscriptomics
PR40728	Wilms	BulkTranscriptomics
PR40729	Wilms	BulkTranscriptomics
PR37201	CMN	BulkTranscriptomics
PR37205	CMN	BulkTranscriptomics
PR37217	CMN	BulkTranscriptomics
PR37220	CMN	BulkTranscriptomics
PR37223	CMN	BulkTranscriptomics
PR37199	CMN	BulkTranscriptomics
PR37200	CMN	BulkTranscriptomics
PR37204	CMN	BulkTranscriptomics
PR37208	CMN	BulkTranscriptomics
PR37212	CMN	BulkTranscriptomics
PR37213	CMN	BulkTranscriptomics
PR37214	CMN	BulkTranscriptomics
PR37216	CMN	BulkTranscriptomics
PR37219	CMN	BulkTranscriptomics
PR37221	CMN	BulkTranscriptomics
PR37222	CMN	BulkTranscriptomics
PR37224	CMN	BulkTranscriptomics
PR37210	CMN	BulkTranscriptomics
PR37225	CMN	BulkTranscriptomics
PR37202	CMN	BulkTranscriptomics
PR40309	CrypticRenalTumor	BulkTranscriptomics

CHAPTER 4

Somatic mutations and single cell transcriptomes reveal the root of malignant rhabdoid tumours

Lars Custers[1,2†], Eleonora Khabirova[3†], Tim H.H. Coorens[3†], Thomas R.W. Oliver[3,4,5], Camilla Calandrini[1,2], Matthew D. Young[3], Felipe A. Vieira Braga[6], Peter Ellis[3], Lira Mamanova[3], Heidi Segers[7], Arie Maat[1], Marcel Kool[1,8,9], Eelco W. Hoving[1], Marry M. van den Heuvel-Eibrink[1], James Nicholson[4,10], Karin Straathof[11,12], Liz Hook[4,5], Ronald R. de Krijger[1,13], Claire Trayers[4], Kieren Allinson[4], Sam Behjati[3,4,10*], Jarno Drost[1,2*]

Nature Communications, **3 March 2021**

[1]Princess Máxima Center for Pediatric Oncology, 3584CS, Utrecht, the Netherlands

[2]Oncode Institute, 3584CS, Utrecht, the Netherlands

[3]Wellcome Sanger Institute, Hinxton, CB10 1SA, UK

[4]Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK

[5]Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK

[6]Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

[7]Department of Pediatric Hemato-Oncology, University Hospital Leuven, Leuven, Belgium

[8]Hopp Children's Cancer Center (KiTZ), 69120 Heidelberg, Germany

[9]Division of Pediatric Neurooncology, German Cancer Research Center DKFZ and German Cancer Consortium DKTK, 69120 Heidelberg, Germany

[10]Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK

[11]UCL Great Ormond Street Hospital Institute of Child Health Biomedical Research Centre, London, UK

[12]Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK

[13]Department of Pathology, University Medical Center Utrecht, 3584CX Utrecht, the Netherlands

[†] These authors contributed equally to this work.

[*] These authors jointly supervised this work.

Abstract

Malignant rhabdoid tumour (MRT) is an often lethal childhood cancer that, like many paediatric tumours, is thought to arise from aberrant fetal development. The embryonic root and differentiation pathways underpinning MRT are not firmly established. Here, we study the origin of MRT by combining phylogenetic analyses and single-cell mRNA studies in patient-derived organoids. Comparison of somatic mutations shared between cancer and surrounding normal tissues places MRT in a lineage with neural crest-derived Schwann cells. Single cell mRNA readouts of MRT differentiation, which we examine by reverting the genetic driver mutation underpinning MRT, SMARCB1 loss, suggests that cells are blocked *en route* to differentiating into mesenchyme. Quantitative transcriptional predictions indicate that combined HDAC and mTOR inhibition mimic MRT differentiation, which we confirm experimentally. Our study defines the developmental block of MRT and reveals potential differentiation therapies.

Introduction

Malignant rhabdoid tumours (MRT) are soft tissue cancers that predominantly affect infants. Although they may arise in any body part, MRT usually form in isolation or synchronously in the kidney and the brain (where they are referred to as atypical teratoid / rhabdoid tumours, AT/RT). MRT, especially metastatic MRT, remain one of the most lethal childhood cancers, even following intense multimodal treatment. The sole driver event of MRT is the occurrence of biallelic mutations in the genes encoding SMARCB1 (INI1, 95% of cases) or SMARCA4 (5% of cases), the core subunits of the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin-remodelling complex[1-3]. In about one third of cases, one of the variants is present in the germline, thus predisposing children to the development of MRT[4].

Like most childhood cancers[5], MRT are thought to arise during embryogenesis, a notion that has recently been substantiated in studies of mouse models of *Smarcb1* loss. Rhabdoid tumours, albeit the majority being AT/RT, only developed in these mice when *Smarcb1* was inactivated during very early embryogenesis, but not at later fetal stages or in adult animals. Renal MRT were never observed[6]. Analyses of bulk, and more recently of single cell transcriptomes, suggest that MRT retain an overall fetal transcriptome with neural as well as mesenchymal signals[7-10]. These findings suggest as a plausible source of rhabdoid tumours the ectoderm-derived neural crest, which is uniquely capable of generating cell types across the boundaries of the germ layers, mesoderm and ectoderm.

The fetal origin of MRT may be exploitable therapeutically by promoting differentiation of MRT along developmental pathways. The possibility of devising differentiation treatments for childhood cancer has recently gained traction with the advent of high throughput single cell assays[5,10-12]. Single cell transcriptomic readouts enable precise,

comprehensive and quantitative comparisons of cancer cells to the transcriptional changes underpinning normal cellular development, thus potentially revealing therapeutic avenues for promoting cellular maturation.

Here, we define the developmental root of MRT and reveal opportunities for differentiation therapy, combining phylogenetic analyses of tumours and surrounding normal tissues, single cell mRNA readouts, and perturbation experiments in patient-derived MRT organoids.

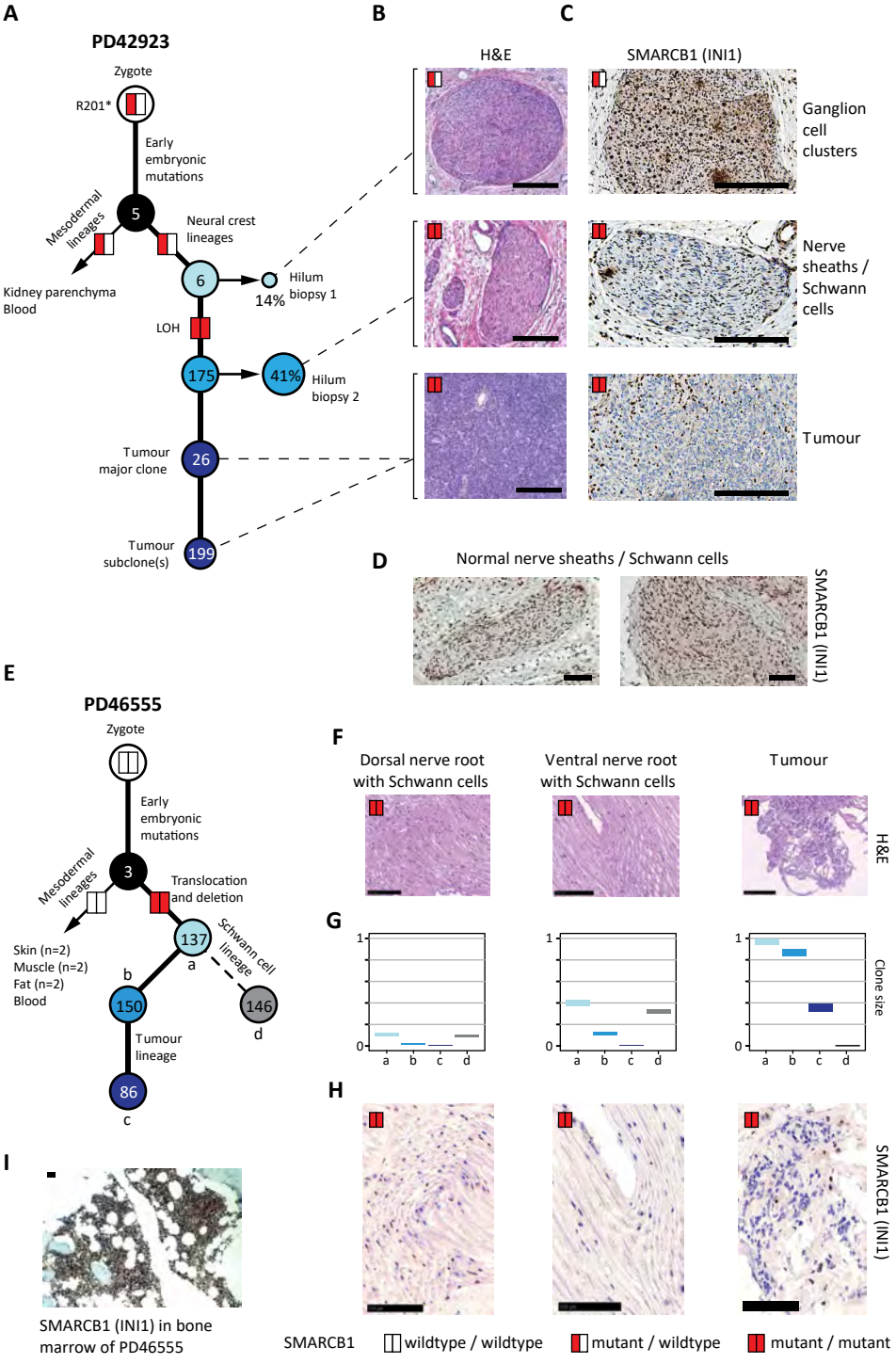
Results

Malignant rhabdoid tumours are phylogenetically related to neural crest-derived tissues

The starting point of our investigation were phylogenetic analyses of MRT, to establish whether the origin of MRT lies in the neural crest in humans. We have previously shown that it is feasible to reconstruct the developmental relationship between childhood tumours and normal tissues from the distribution of somatic mutations across tissues[13]. Applying these principles to MRT, we used DNA whole genome sequencing (WGS) to study two cases of MRT along with corresponding normal tissues.

First, we examined tissue obtained from a child presenting with the most common type of extracranial rhabdoid tumour, renal MRT. The child carried a pathogenic germline *SMARCB1* mutation. We performed WGS of tumour (n=2), blood cells, kidney parenchyma, and renal hilar tissue (n=2) (**Fig. 1a**, **Supplementary Table 1**). Using an established variant calling pipeline[13-16], we determined somatic variants in each tissue, from which we derived phylogenetic relations between tumour and normal tissues. The possibility of observing shared mutations due to tumour contamination was addressed by histological examination and by quantitative assessment through a mixture model (Methods). The key finding in this first case was that some, but not all somatic mutations of the tumour, were present in hilar tissues, occupied by ganglion cells and Schwann cells (**Fig. 1a, b, Supplementary Fig. 1**). Both these cell types are derived from the neural crest. However, there were no shared somatic mutations between tumour and blood or kidney parenchyma, bar ubiquitous early embryonic mutations (**Fig. 1a, Supplementary Fig. 1**). These findings place MRT on an ectodermal, neural crest-derived lineage with Schwann cells, distant from mesodermal blood and kidney parenchymal lineages.

Examining shared mutations between tumour and hilar tissues more closely, we found that one hilar biopsy, occupied mainly by ganglion cells, shared only a small number (n=6) of variants with the tumour, whereas the second, composed of Schwann cells, shared 175 mutations with the tumour (**Fig. 1a, b, Supplementary Data 1**) including copy number-neutral loss of heterozygosity of *SMARCB1* (**Supplementary Fig. 2**). To verify this finding, we performed immunohistochemistry of the *SMARCB1* protein, INI1 (**Fig. 1c**). As predicted from the distribution of mutations, the first hilar


A
PD42923
Zygote
R201*
Early embryonic mutations
5
Mesodermal lineages
Neural crest lineages
Kidney parenchyma
Blood
6
Hilum biopsy 1
14%
LOH
175
41%
Hilum biopsy 2
Tumour major clone
26
Tumour subclone(s)
199
B
C
H&E
SMARCB1 (INI1)
Ganglion cell clusters
Nerve sheaths / Schwann cells
Tumour
D
Normal nerve sheaths / Schwann cells
SMARCB1 (INI1)
E
PD46555
Zygote
Early embryonic mutations
3
Mesodermal lineages
Translocation and deletion
Skin (n=2)
Muscle (n=2)
Fat (n=2)
Blood
137
Schwann cell lineage
a
b
150
146
d
Tumour lineage
86
c
F
Dorsal nerve root with Schwann cells
Ventral nerve root with Schwann cells
Tumour
H&E
G
Clone size
1
0
a b c d
H
SMARCB1 (INI1)
I
SMARCB1 (INI1) in bone marrow of PD46555
SMARCB1
wildtype / wildtype
mutant / wildtype
mutant / mutant

biopsy showed only occasional INI1 negative cells, consistent with a heterozygous germline mutation of *SMARCB1*. By contrast, the Schwann cells of the second hilar biopsy, which should have stained ubiquitously and intensely positive for INI1 (**Fig. 1d**), did not exhibit INI1 staining (**Fig. 1c**), consistent with biallelic loss of *SMARCB1* predicted from the somatic genome of this tissue.

Next, we examined tissues obtained *post mortem* from a child who succumbed to an MRT of the cervical spine. The tumour bulk was situated ventrally in the extradural space. The child did not carry germline *SMARCB1* mutations. No early mosaic (i.e. present in blood) variant affecting *SMARCB1* was discovered in this patient. We studied tumour tissue along with nine normal tissues: skin (n=2), fat (n=2), muscle (n=2), blood, dorsal and ventral nerve roots (**Fig. 1e, Supplementary Table 1**). Pursuing the same analyses as before, we found that the tumour was somatically related to neural crest-derived Schwann cells sampled in nerve roots (**Fig. 1e, f, Supplementary Fig. 1, Supplementary Data 1**), but not to any other normal tissue. The clonal composition underlying this phylogenetic relation was complex. Based on variant allele frequencies and distribution of mutations, we were able to discern four clones (**Fig. 1g**), two of which were shared between Schwann cells and tumour. In addition, the tumour and Schwann cell lineages possessed a private clone each, alluding to a sustained potential of tumour and Schwann cells for subclonal diversification. Analysis of copy number variants revealed a biallelic loss of *SMARCB1* in tumour and both nerve roots (**Fig. 1e**, **Supplementary Fig. 2, Supplementary Data 1**), which again we were able to validate through INI1 staining. INI1 negative Schwann cells were more readily found in the ventral root, consistent with the larger clone sizes in this tissue (~40% vs ~20%, **Fig. 1h, i**). Together, these observations provide the most direct evidence yet that human MRT is phylogenetically related to the neural crest lineage and firmly places its origin in fetal life.

Figure 1 | MRT are phylogenetically closely related to neural crest-derived Schwann cells.
(A) Phylogenetic tree representing the somatic genetic relation of a renal MRT and normal tissues. Percentages: clone size in tissues. Numbers inside circles: mutation burden within cluster. Red or white coloured rectangles: *SMARCB1* mutations status (red = mutant; white = wild type). LOH, loss of heterozygosity. H&E **(B)** staining of biopsies and INI1 **(C)** immunostaining, showing INI1 negative Schwann cells in hilum biopsy 2. Scale bars = 100 μm. **(D)** Pattern of positive INI1 staining in Schwann cells of normal nerve sheath from control hilar regions of two independent donors. Scale bars = 100 μm. **(E)** Phylogenetic tree representing the somatic genetic relation of an extradural (spinal) MRT and normal tissues. Embryonic clusters of mutations are denoted a to d. The annotation otherwise follows **(A)**. H&E staining **(F)**, clone size of the different mutational clusters (a-d) **(G)**, and INI1 immunostaining **(H)** of dorsal nerve root, ventral nerve root, tumour and **(I)** bone marrow of the same donor. The latter showing positive INI1 staining. Scale bars = 100 μm.

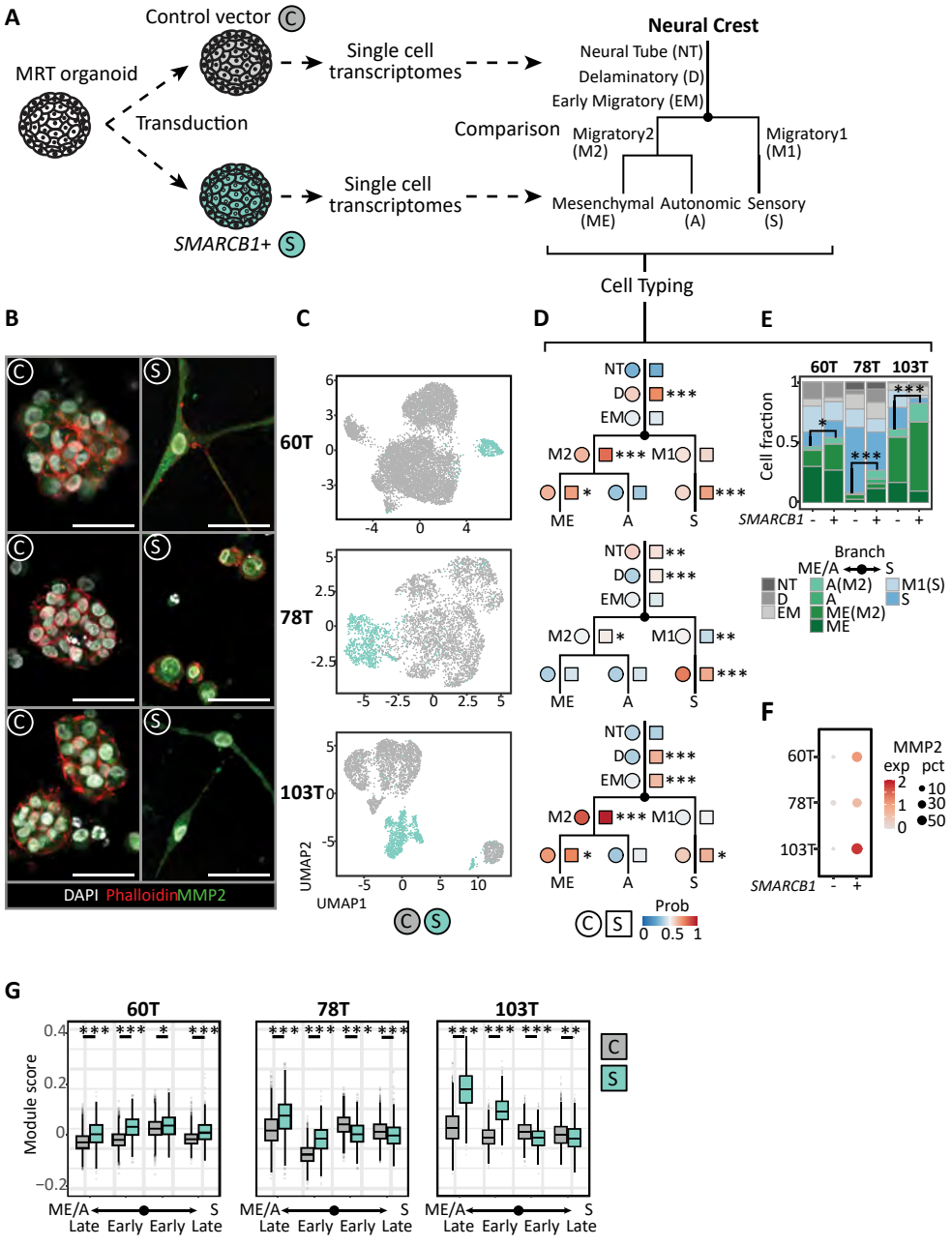


Figure 2 | *SMARCB1* reconstitution drives MRT differentiation.
(A) Schematic representation of *SMARCB1* reconstitution in patient-derived MRT organoids and subsequent single cell transcriptome comparison to fetal mouse neural tube and neural crest cell types. Branching tree represents differentiation trajectories of mouse neural crest. Abbreviations are indicated. **(B)** Representative immunofluorescence images of MRT control (C) and *SMARCB1*+ (S) organoids. White: DAPI (nuclei), red: phalloidin (membranes), green: MMP2 (mesenchymal marker). Scale bars equal 50 µm. **(C)** UMAP

SMARCB1 reconstitution drives MRT differentiation

To establish the differentiation stage of MRT within neural crest development, we studied the consequences of reversing loss of *SMARCB1*, the principal genetic driver of MRT. As a model of MRT, we utilised patient-derived MRT organoids, which have been shown to faithfully recapitulate the genetic, transcriptional, and epigenetic features of primary MRT tissue[17]. We reconstituted *SMARCB1* expression in three MRT organoid cultures[17] (60T, 78T and 103T; **Supplementary Table 2**) by lentiviral transduction with either a control or *SMARCB1* expression plasmid (**Fig. 2a, Supplementary Fig. 3a**). DNA methylation profiles of MRT organoids resembled those of primary MRT tissue, irrespective of *SMARCB1* status (**Supplementary Fig 3b**). Reconstitution of *SMARCB1* expression induced a proliferation arrest in all MRT cultures (**Supplementary Fig. 3c**) with a morphological transformation of cells (**Fig. 2b, Supplementary Fig. 4a, c**). While both 60T and 103T transformed from a grape-like to a neural- or fibroblast-like morphology with long extensions protruding from the cell body, 78T stopped proliferating without an apparent morphological change.

To assess the transcriptional profiles underpinning these phenotypic changes, we subjected organoid cultures, without and with *SMARCB1* re-expression, to single cell mRNA sequencing (10x Genomics Chromium platform, n=16,133 cells post

representation of single cells from MRT control (gray) or *SMARCB1*+ (green) organoid lines (60T control/*SMARCB1*+: 8059/425 cells, 78T control/*SMARCB1*+: 3195/806 cells, 103T control/*SMARCB1*+: 2694/953 cells). **(D)** Dot plots represent similarity of MRT control (circles) or *SMARCB1*+ (squares) cells to neural crest differentiation trajectories. Colours represent the average probability (prob) that the MRT cells are similar to the indicated neural crest cell type (predicted similarity score estimated by logistic regression[12]). Changes in similarity score between control and *SMARCB1*+ cells were assessed for cell types with average similarity score > 0.5. P-values were calculated using an unpaired Student's t test (two-tailed): * <1e-3, ** <1e-9, *** <1e-15 (-log10(p-value): 60T D=45, S=27, M2=66, ME=3.7; 78T NT=9, D=54, M1=14, S=22, M2=4.4; 103T D=198, EM=40, S=7.8, M2=314, ME=3.2). **(E)** Stacked bar plot represents relative frequencies of single cell annotations for MRT control (-) and *SMARCB1*+ (+) organoids, showing a consistent conversion of neural to mesenchymal signals. Cell type annotation was assigned for each single cell based on the highest similarity score. Colours represent neural crest cell types depicted in Fig. 2a. Cell type migratory2 (M2) was assigned as either migratory mesenchyme (ME(M2)) or migratory autonomic (A(M2)) based on the highest similarity score. The relative frequency of the mesenchymal/autonomic (ME/A) branch was compared between control and *SMARCB1*+ organoids for each patient line. P-values were calculated using a chi-square test: * <0.01, *** <1e-15 (p-value: 60T=0.0048; 78T=4.9e-48; 103T=1.0e-32). **(F)** Dot plot shows expression levels (exp) of mesenchymal marker *MMP2* for MRT control (-) and *SMARCB1*+ (+) organoids for each patient line. Colour-code from gray to red refers to average *MMP2* transcript levels (unique molecular identifier (UMI)). Dot size refers to the percentage of cells (pct) showing *MMP2* expression. **(G)** Box plot representation of gene module scores for MRT control (gray) and *SMARCB1*+ (green) single cells (n = 60T control/*SMARCB1*+: 8059/425 cells; 78T control/*SMARCB1*+: 3195/806 cells; 103T control/*SMARCB1*+: 2694/953 cells), showing consistent upregulation of mesenchymal/autonomic differentiation genes for *SMARCB1*+ cells. Box plots indicate median (middle line), 25th and 75th percentile (box). Whiskers represent the range excluding outliers (dot). Module scores were generated by averaging gene expression levels per set of genes. Gene sets include marker genes for either sensory (S) or mesenchymal/autonomic (ME/A) differentiation branches, distinguishing early and late differentiation genes. Module scores were assessed by comparing control and *SMARCB1*+ cells. P-values were calculated using an unpaired Student's t test (two-tailed): * <1e-3, ** <1e-9, *** <1e-15 (-log10(p-value) ME/A late 60T=28, 78T=64, 103T=Inf; ME/A early 60T=77, 78T=134, 103T=Inf; S early 60T=5.6, 78T=72, 103T=54; S late 60T=28, 78T=16, 103T=11).

filtering). Cell cycle profiles generated from single cell transcriptomes confirmed the growth arrest induced by *SMARCB1*, with 78T showing the least penetrant effect (**Supplementary Fig. 3d**). UMAP clustering revealed that, as expected, most transcriptomic variance of MRT single cells can be explained by donor, as cells first separate by patient line (**Supplementary Fig. 3e**). After transduction, the majority of cells expressing *SMARCB1* segregated into independent cell clusters for each patient line (**Fig. 2c, Supplementary Fig. 5a-c).** This segregation was not explained by batch effects of individual cultures, as unsuccessfully transduced cells co-clustered with cells from cultures transduced with the control plasmid (**Supplementary Fig. 3f**).

As our phylogenetic analyses revealed that MRT are neural crest-derived (**Fig. 1**), we subsequently assessed the similarity of MRT cells with and without *SMARCB1* re-expression relative to single cell signals of (murine) neural crest development[18] using logistic regression[12] (**Fig. 2a**). At baseline (i.e. no *SMARCB1* re-expression), MRT organoid transcriptomes primarily resembled mesenchymal and neural cells (**Fig. 2d, Supplementary Fig. 3g**), as previously shown[7,8,10]. Additionally, each individual patient line exhibited different signals of neural crest differentiation stages (**Fig. 2d, Supplementary Fig. 3g**). Patient lines 60T and 103T primarily resembled mesenchymal cells, whereas 78T exhibited a more neural signal. Examining cellular mRNA profiles upon *SMARCB1* reconstitution, MRT cells appeared consistently more differentiated. That is, they resembled their normal counterpart more strongly, as similarity to most neural crest cell types increased (**Fig. 2d, Supplementary Fig. 3g**). Further assessment of cell type similarity showed that *SMARCB1* reconstitution promotes a neural to mesenchyme conversion that is consistent among all three MRT organoid cultures (**Fig. 2e**). These results were validated using a second independent mouse single cell mRNA reference of early neural and mesenchymal cell types[19] (**Supplementary Fig. 3h**). In agreement, analysis of neural crest differentiation genes showed a significant upregulation of mesenchymal markers (**Fig. 2b, f, g, Supplementary Fig. 4b-d, Supplementary Data 2**).

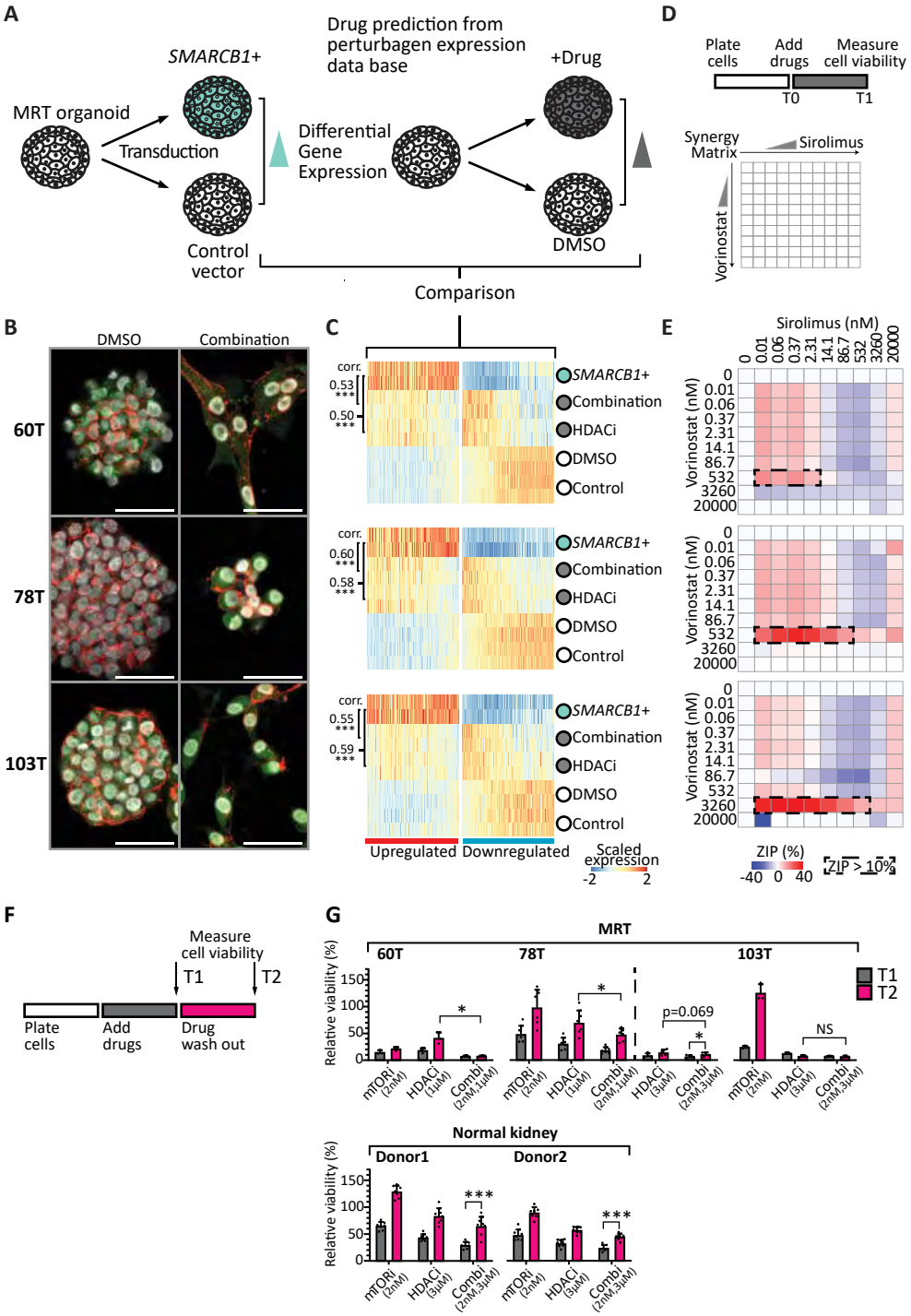
Additional cell typing was performed for each single cell cluster separately to evaluate intra-organoid heterogeneity (**Supplementary Fig. 5a-f, Supplementary Data 2**), showing that single cell clusters exhibited variable neural crest differentiation signals. However, *SMARCB1*+ clusters consistently induced a relative gain of mesenchymal differentiation signal, with the exception of minor cluster 60T_S2, which retained a more neural identity. Altogether, our findings place MRT on a developmental trajectory of neural crest to primarily mesenchyme conversion, which is promoted by *SMARCB1* reconstitution.

Mimicking *SMARCB1* reconstitution pharmacologically

Reconstitution of *SMARCB1* to drive differentiation of MRT would appear to be an attractive, non-cytotoxic treatment strategy. However, reinstating *SMARCB1* expression genetically in children is not feasible at present. An alternative strategy is to find agents that mimic the changes induced by *SMARCB1* re-expression. Using bulk mRNA-seq, we defined a *SMARCB1*+ transcriptional program based on genes upregulated upon *SMARCB1* re-expression in our three MRT organoid cultures (**Supplementary Fig. 6a, Supplementary Data 3**). We could validate the *SMARCB1*+ program in MRT tissue and found a positive correlation with *SMARCB1* expression levels in normal tissues (**Supplementary Fig. 6b**). To explore therapeutic avenues, we searched a publicly available perturbation data base[20] for drugs that induce expression changes of *SMARCB1* reconstitution (**Fig. 3a**). This analysis identified a variety of HDAC and mTOR inhibitors (**Supplementary Fig. 6c**) as the top hits. Interestingly, HDAC inhibitors have previously been identified for treatment of rhabdoid tumours by orthogonal approaches[21].

We tested the phenotypic and transcriptional effects of these agents, alone or in combination, across the three MRT organoid cultures. HDAC inhibition alone induced a morphological transformation akin to *SMARCB1* reconstitution (**Fig. 3b, Supplementary Fig. 7a**). Furthermore, there was a significant correlation between gene expression changes of bulk culture transcriptomes of *SMARCB1* re-expression and HDAC inhibition (**Fig. 3c**, **Supplementary Fig. 6d**). Inhibition of mTOR signalling primarily constrained organoid growth, which, however, was readily reversible by drug washout (**Fig. 3f, g**). Combination of HDAC and mTOR inhibition induced the phenotypic and transcriptional changes of *SMARCB1* reconstitution / HDAC inhibition as well as a marked proliferation arrest (**Fig. 3b, c, Supplementary Fig. 7a**). The action of HDAC and mTOR inhibition was synergistic, as corroborated by assessment of the two drugs in dose-response matrices (**Fig. 3d, e, Supplementary Fig. 8a**). Furthermore, the combined effects of the drugs on viability were more durable than single agent treatment. On its own, anti-proliferation effects of each drug were readily reversible upon washout (bar HDAC inhibition in MRT organoid 103T). By contrast, combination treatment had more lasting effects on proliferation in all tested MRT organoid cultures (**Fig. 3f, g, Supplementary Fig. 7b**). While regrowth of 60T and 103T was completely diminished after drug wash-out, 78T showed minor regrowth, which could relate to the remnant proliferating cells that were also observed upon *SMARCB1* reconstitution (**Supplementary Fig. 3d**). To determine whether MRT are in particular sensitive to HDACi and mTOR inhibition, we tested the sensitivity of normal kidney organoids[22] to both drugs. Normal kidney organoids were significantly more resistant to single agents as well as combination treatment, and in contrast to MRT, showed significant regrowth upon washout of drug combination (**Fig. 3f, g, Supplementary Fig 7b, Supplementary Fig. 8b-d**).

Mechanistically, the longevity of the effects of combined HDAC and mTOR inhibition may be mediated through interference with MYC, as MYC-driven cancer cell lines


A
MRT organoid
Transduction
SMARCB1+
Control vector
Differential Gene Expression
Drug prediction from perturbagen expression data base
+Drug
DMSO
Comparison
D
Plate cells
Add drugs
Measure cell viability
T0
T1
Synergy Matrix
Sirolimus
Vorinostat
B
DMSO
Combination
60T
78T
103T
C
corr.
0.53 ***
0.50 ***
0.60 ***
0.58 ***
0.55 ***
0.59 ***
SMARCB1+
Combination
HDACi
DMSO
Control
Upregulated
Downregulated
Scaled expression
-2
2
E
Sirolimus (nM)
0
0.01
0.06
0.37
2.31
14.1
86.7
532
3260
20000
Vorinostat (nM)
ZIP (%)
-40 0 40
ZIP > 10%
F
Measure cell viability
T1
T2
Plate cells
Add drugs
Drug wash out
G
MRT
60T
78T
103T
Relative viability (%)
150
100
50
0
*
p=0.069
NS
mTORi (2nM)
HDACi (1μM)
HDACi (3μM)
Combi (2nM,1μM)
Combi (2nM,3μM)
T1
T2
Normal kidney
Donor1
Donor2
***

seem to be particularly susceptible to this drug combination[23]. In our experiments, we interrogated hallmark pathways and perturbation gene sets[24], which validated our MRT *SMARCB1*+ model, as we found enrichment for SWI/SNF-related perturbation gene sets upon *SMARCB1* reconstitution (e.g. SNF5 (*SMARCB1*) and subunits of polycomb repressive complexes[25]) (**Supplementary Fig. 6e, f**). Further, we found that MYC target genes were strongly downregulated upon *SMARCB1* re-expression (**Supplementary Fig. 6e, h**). This was mimicked by combined HDAC / mTOR inhibition, significantly more strongly than by single agent treatment (**Supplementary Fig. 6i**). Further examination of differentially expressed genes showed that identified pathways are largely shared between *SMARCB1* reconstitution and combination treatment (**Supplementary Fig. 6g, Supplementary Data 3**). Together, these analyses identify combined HDAC / mTOR inhibition as pharmacological mimics of *SMARCB1* reconstitution that prohibit proliferation and induce differentiation in MRT.

Figure 3 | Combined HDAC / mTOR inhibition mirrors *SMARCB1* reconstitution.
(A) Overview of methodology used for discovery of potential differentiation therapeutics. **(B)** Representative immunofluorescence images of MRT organoids treated with DMSO control or a combination of vorinostat (HDACi, 1μM) and sirolimus (mTORi, 2nM). White: DAPI (nuclei), red: phalloidin (membranes), green: MMP2 (mesenchymal marker). Scale bars equal 50 μm. **(C)** Heatmaps represent gene expression values (n = 2 independent experiments) of MRT control or *SMARCB1*+ organoids, or MRT organoids treated with vorinostat (HDACi, 1μM) or both vorinostat and sirolimus (combination,1μM/2nM). Heatmaps are subset for genes differentially expressed upon *SMARCB1* re-expression (**Supplementary Data 3**). Genes are ordered by the average mRNA changes induced by *SMARCB1* re-expression and treatment. Colour-code represents gene expression values scaled by gene. Pearson correlation coefficients (corr.) were generated by comparing mRNA changes induced by either *SMARCB1* re-expression or HDACi / combination treatment. P-values are indicated for Pearson's correlation tests (two-tailed): *** <1e-15 (-log10(p-value): Combi 60T=217, 78T=Inf, 103T=221; HDACi 60T=268, 78T=306, 103T=192). **(D)** Schematic overview of the dose response matrix setup to find synergy between HDAC (vorinostat) and mTOR (sirolimus) inhibitors in MRTs. **(E)** Graphs show zero interaction potency (ZIP) scores that indicate either synergistic (red) or antagonistic (blue) effects of combination treatment. ZIP scores are generated by calculating the observed deviation from a reference model that assumes drugs are non-interacting (synergy when ZIP > 10%[51]). The dashed rectangles highlight the drug concentration ranges where synergy between the two drugs is the strongest. Source data are provided as a Source Data file. **(F)** Schematic overview of the regrowth assay. **(G)** Bar graphs represent cell viability values normalized to timepoint 1 (T1) DMSO controls for each MRT or normal kidney organoid line. Mean and SD (error bars) of independent experiments (dot) are indicated (n = 60T/103T: 3, 78T mTOR/HDAC1μM/Combi2nM1μM: 6, 78T HDAC3μM/Combi2nM3μM: 4. normal kidney: 7). Each independent experiment is an average of 4 technical replicates. Source data are provided as a Source Data file. Additional effect of combination treatment on cell viability was determined by comparing combination (T2) with HDACi (T2) treatment. Regrowth capability was assessed by comparing T2 to T1. P-values were calculated using a paired ratio Student's t test (two-tailed): * <0.05, ** <0.01, *** <0.001 (p-value: Combi 1μM T1 vs HDACi 1μM T1 60T=0.020, 78T=0.012; Combi 3μM T2 vs Combi 3μM T1 78T=0.013, normal kidney donor1=2.5e-5, donor2=1.8e-5).

Discussion

We investigated the origin of MRT by combining phylogenetic and transcriptional analyses with experiments in model systems. Our findings indicate that MRT arise from the neural crest *en route* to differentiating to mesenchyme and suggest combined mTOR / HDAC inhibition as a therapeutic hypothesis.

Previous investigations into the origin of MRT have built on transcriptional and epigenetic analyses of MRT. Despite the genetic uniformity of intra- (AT/RT) and extracranial rhabdoid tumours, with *SMARCB1* loss being the sole recurrent genetic driver, such analyses have revealed phenotypic, transcriptomic and epigenetic variation in rhabdoid tumours, collectively showing neural and / or mesenchymal differentiation of MRT[7,8,26,27]. Our phylogenetic analyses now firmly place the root of MRT in a lineage with neural crest derived Schwann cells. The varied phenotype of MRT may thus be explained by the ability of neural crest lineages to generate cells across the boundaries of mesoderm and ectoderm. Modelling attempts of rhabdoid tumours in mice have shown that the timing of *Smarcb1* loss during development is critical for the formation of tumours[6,28]. It is interesting to note that in our study some morphologically normal Schwann cells partly harboured the somatic genome of MRT, including homozygous loss of *SMARCB1* in some cells. This would suggest that *SMARCB1* loss on its own may not suffice to generate tumours, or to prevent normal cellular differentiation. This proposition is further corroborated by the wide spectrum of tumours, including more benign entities such as Schwannomas, that pathogenic germline mutations in *SMARCB1* (and related genes) predispose to[29]. Taken together, therefore, factors other than embryological timing of disruption of *SMARCB1* would seem to influence tumour formation in humans.

A unique feature that distinguishes childhood from adult cancers is the fetal origin of paediatric tumours. The developmental programs that underpin the aberrant differentiation of cancer cells may lend themselves as therapeutic target[5]. A precedent for this notion is the clinical use of retinoic acid derivatives as maturation treatment of neuroblastoma (https://clinicaltrials.gov/ct2/show/NCT01704716). It is, however, challenging to devise maturation treatments, as conventional readouts employed in high throughput genetic or drug screens, such as viability or proliferation, may not necessarily capture differentiation states. We employed a novel strategy here, by defining, in quantitative molecular terms, the target state of MRT differentiation, as defined through *SMARCB1* reconstitution. We then used in silico matching of target state and drug effects to search for agents that mimic genetic *SMARCB1* reconstitution. Our approach represents a specific (biased) search for agents that elicit a predefined transcriptional effect (in our case MRT maturation). Therefore, although it lacks the power of unbiased drug screens to discovering the unknown, our approach can be a focused path to drug discovery, when the target state can be defined in quantitative, molecular terms.

Studies of rare tumours, such as MRT, that rely on access to fresh material and detailed sampling are invariably small, even when conducted across large consortia such as ours. Therefore, as the size of the biobank of MRT organoids grows with time, we will have to re-examine our observations in larger cohorts. Moreover, it will be important to examine the generalisability of our findings to other tumours driven by biallelic loss of *SMARCB1* and other members of the BAF chromatin-remodelling complex, in particular AT/RT which are considered to be the intracranial counterpart of MRT.

MRT remains one of the most aggressive childhood cancers, which often rapidly progresses despite intense cytotoxic treatment. It would therefore seem attractive to immediately try combined mTOR / HDAC inhibition in the treatment of MRT. However, what may seem to be a harmless differentiation agent *in vitro*, may have severe adverse effects on postnatal development in children[5]. For instance, a phase II trial on using a sonic hedgehog (SHH) inhibitor in children with SHH medulloblastoma was terminated because of the induction of widespread growth plate fusions[30]. Nevertheless, we would suggest that the therapeutic hypothesis our findings formulate, merits further evaluation. More broadly, our study defines a nimble blueprint for quantitative approaches to the discovery of maturation targets which is applicable across childhood cancer.

Methods

Ethics statement

Tumour and normal tissues used for genetic lineage tracing (**Supplementary Table 1**) were obtained as part of the SIOP2001 study approved by the medical ethical committees of the institutes involved (Ethical Committee University Leuven (Belgium), Medical ethical committee of the Erasmus Medical Centre Rotterdam (the Netherlands), or were obtained from the Children's Cancer and Leukaemia Group (UK) Tissue Bank. Informed written consent was provided by all patients and/or guardians.

Whole genome sequencing

DNA was extracted from either formalin-fixed paraffin-embedded (FFPE) tissues or fresh frozen tumour or tissue samples. Short insert (500bp) genomic libraries were constructed and 75 base pair (FFPE) or 150 base pair (frozen) paired-end sequencing clusters were generated on either the Illumina HiSeq X or the Illumina NovaSeq platform according to Illumina no-PCR library protocols. An overview of samples and sequencing variables, including the average sequence coverage, is shown in **Supplementary Table 1**.

DNA sequence processing and mutation calling

DNA sequences were aligned to the GRCh37d5 reference genome using the Burrows-Wheeler algorithm (BWA-MEM)[31]. Single-nucleotide variants (SNVs) were called against the reference genome using CaVEMan[32]. Beyond the standard post-processing filters of CaVEMan, we removed variants affected mapping artefacts associated with BWA-MEM by setting the median alignment score of reads supporting a mutation as greater than or equal to 140 (ASMD>=140) and requiring that fewer than half of the reads were clipped (CLPM=0).

Across all samples from one patient and their parents, we recounted the SNVs that were called in any sample from the patient, using minimum values for read mapping quality (30) and base quality (25). Germline and somatic variants were distinguished using a one-sided binomial exact test on the number of variant reads and depth present in the matched tissue samples. Resulting p-values were corrected for multiple testing with the Benjamini-Hochberg method[33] and a cut-off was set at $q < 10^{-5}$. Variants were also filtered out if they were called in a region of consistently low (<15x) or high depth (>50x) across all samples from one patient. These thresholds were halved for the X and Y chromosomes in the male patient, PD42923. Using a beta-binomial model of a site-specific error rate as previously employed[16], we distinguished true presence of SNVs from support due to sequencing errors and noise. All shared SNVs were further visually inspected using the genome browser, Jbrowse[34]. See **Supplementary Data 1** for SNV calls of PD42923 and PD46555, respectively. Copy number variants (CNVs) were called using ASCAT[35] and Battenberg[36]. Structural variants (SVs) were called using BRASS[37] (**Supplementary Data 1**).

Clustering and classification of SNVs

To reconstruct the clonal composition of normal tissues, we employed an N-dimensional Bayesian mixture model based on the Dirichlet process[38]. Briefly, SNVs were clustered based on their distribution of variant supporting reads and total depth across all (N) samples. Therefore, it naturally accounts for differences in coverage between samples and does not rely on hard VAF cut-offs. Cellular contributions of mutational clusters were reconciled into a tree based on the pigeon hole principle, i.e. the sum of contributions in daughter branches can never exceed that of the parental one.

Organoid culture

MRT organoids were previously established and characterized[17]. In brief, organoid line 60T, 78T and 103T (**Supplementary Table 2**) were seeded in growth factor-reduced BME (Trevigen) and cultured in organoid medium (Advanced DMEM/F12 (Gibco) containing 1x Glutamax, 10 mM HEPES, and antibiotics (AdDF+++), supplemented with 1.5% B27 supplement (Gibco), 10% R-spondin-conditioned medium, EGF (50 ng/ml, Peprotech), FGF-10 (100 ng/ml, Peprotech), N-acetylcysteine (1.25 mM, Sigma), Rho-kinase inhibitor Y-27632 (10 µM, Abmole), and A83-01 (5 µM, Tocris Bioscience) as described[17]. For *SMARCB1* re-expression, MRT organoids were lentivirally transduced[39] with pLKO.1-UbC-luciferase-blast[40] or pLKO.1-UbC-hSMARCB1-blast lentiviruses. After two days, 10 µg/ml blasticidin was added to the culture medium. For DNA methylation profiling, immunofluorescence and mRNA sequencing experiments (bulk as well as single cell), organoids were harvested four days after lentiviral infection. For cell viability measurements, organoids were harvested seven days after lentiviral infection and viability was measured using CellTiter-Glo 3D reagent (Promega) according to manufacturer's instructions on a Fluorstar Omega plate reader. HDACi in MRT organoids was performed using 1µM entinostat (SelleckChemicals) or 1µM vorinostat (MedChemExpress), added two days after seeding single cells. For HDACi and mTORi immunofluorescence and mRNA-seq experiments, 1µM vorinostat and/or 2nM sirolimus (MedChemExpress) were added to the organoid cultures two days after seeding single cells. Cells were harvested four days after addition of the drugs.

Histology

Tissues were fixed in 4% paraformaldehyde, dehydrated and embedded in paraffin. Immunohistochemistry was performed according to standard protocols on 3–4 µm sections. Sections were subjected to H&E and immunohistochemical staining with the following primary antibodies: INI1 (BD Transduction Laboratories, 612111, 1:400).

Immunofluorescence

Immunofluorescence experiments were performed as described[41], using DAPI (Thermo Fisher, D9542, 1:1000), Alexa Fluor 647 Phalloidin (Thermo Fisher, A22287, 1:200) and primary antibody MMP2 (Thermo Fisher, MA5-13590, 1:500). High-resolution

3D imaging was performed in µ-Slide 8 Well chambers (IBIDI, 80826) using the Leica SP8 confocal microscope and a 20x water immersion objective. Organoids were imaged in 3D by acquiring z-stacks, which were visualized by maximum intensity projections. Acquisition settings for MMP2 were fixed across experiments.

DNA methylation profiling

Genomic DNA was extracted from MRT organoids using the ReliaPrep gDNA Tissue Miniprep System (Promega) according to manufacturer's protocol. DNA methylation profiles were established using Illumina Human MethylationEPIC BeadChip arrays at the German Cancer Research Center (DKFZ) Genomics and Proteomics Core Facility according to manufacturer's protocol. DNA methylation data were analysed as described[26]. Organoid methylation data were co-clustered with previously published DNA methylation data of AT/RT and MRT[7].

Single cell mRNA sequencing

MRT organoids were dissociated into single cell suspensions using TrypLE Express (ThermoFisher) supplemented with Y-27632 (10 µM, Abmole) and subsequently processed on the Chromium 10X Genomics controller as previously described[10]. Cell viability was estimated using Trypan Blue solution (Thermo Fisher, 15250061), ranging from 38-65% viable cells. For 78T and 103T samples, single cells were mixed pre-loading in a one to one ratio. Libraries were produced according to the manufacturer's protocol and sequenced on Illumina platforms (HiSeq4000 or NextSeq500).

Mapping, quantification, quality control and demultiplexing of single cell mRNA-seq data

Raw sequencing data was aligned to the reference genome (GRCh38) and quantified with cellranger 2.0.2 pipeline[42] for 60T samples (sequencing was done with Chromium single cell 3' v2 chemistry) and cellranger 3.0.3 pipeline for mixed 78T and 103T samples (sequencing was done with Chromium single cell 3' v3 chemistry), producing a table of counts of unique molecular identifiers (UMI) for each single cell and gene with sequencing depth (average reads/cell) for 60T control (35k), *SMARCB1*+ (490k), 78T & 103T mixed control (34k) and *SMARCB1*+ (44k). Demultiplexing of 78T and 103T was performed with demuxlet software[43] with default parameters using genotype data of patient line 78T and 103T[17]. Gene expression matrices were further processed with python package scanpy version 1.4.4.post1[44]. Poor quality cells were filtered out based on a low number of genes (<200 detected genes) and high mitochondrial content (>10% for 60T and >20% for 78T and 103T). Cells were filtered based on *SMARCB1* transcript counts. For MRT control samples, cells were removed with *SMARCB1*>0. For MRT *SMARCB1*+ samples, cells were removed with *SMARCB1*=0. Doublets were detected and excluded with python package scrublet version 0.2.1[45] with default parameters. Demultiplexing of organoid lines 78T and 103T was performed based on cluster assignment provided by demuxlet. Additionally, single cell outliers for 78T and 103T were removed based on low coverage (detected UMIs < 1500) or high counts

(more than 60000 for 78T and 50000 for 103T). All the scripts used for filtering and quality control are listed in the "Code availability" section. Sample metadata are provided in **Supplementary Table 3**.

Clustering and module scores

For UMAP visualization and single cell clustering, raw gene expression matrices of filtered cells were processed using R package Seurat version 3.1.4 and standard analysis pipelines[46]. Downstream analysis was performed for all single cells combined (**Fig. 2f, g, Supplementary Fig. 3d, e**) or separately for each patient line (**Fig. 2c, Supplementary Fig. 4b, 5a-f**) and included the following steps. Data normalization of raw counts was performed using R package sctransform[47] version 0.2.1 integrated in Seurat. The single command "SCTransform" was applied, which replaces the Seurat commands "NormalizeData", "ScaleData" and "FindVariableFeatures". Counts were normalized using the standard normalization method "LogNormalize" with scaling factor 10,000. Confounding sources of variation were removed by regressing out mitochondrial content ("PercentageFeatureSet" for gene names starting with "MT"). Single cells were scored for cell cycle ("CellCycleScoring") which was additionally regressed out to correct for unwanted cell cycle effects. Subsequently, we performed principal component analysis ("RunPCA") using highly variable features (3000 genes using default parameters), followed by dimensionality reduction ("RunUMAP"). Inclusion of principal components was based on a scree plot ("ElbowPlot") by placing the cut-off at the elbow of the curve (Combined: 1-18, 60T: 1-15, 78T: 1-17, 103T: 1-15). Clusters were determined by Louvain graph-based clustering, performed separately for each patient line ("FindNeighbours" and "FindClusters"). The optimal clustering resolution was identified by calculation of average silhouette widths for a series of resolutions (**Supplementary Fig. 5b**). Differentially expressed genes for control and *SMARCB1*+ single cells were determined using "FindMarkers" (**Supplementary Data 2**). Marker genes defining single cell clusters were determined using "FindAllMarkers" (**Supplementary Data 2**). P-values were calculated using a Wilcoxon Rank Sum test and were corrected for multiple testing. Module scores were calculated ("AddModuleScore") on a combined normalised dataset of all patient lines for inter-organoid comparisons (**Fig. 2g**), or separately for each patient line for assessment of single cell clusters (**Supplementary Fig. 5f**). Neural crest branch specific gene sets include marker genes that separate sensory and autonomic/mesenchymal branches, further specifying late genes (upregulated in differentiated cells) and early genes (upregulated in progenitor and differentiated cells), as described[18]. Mouse genes without human orthologue were excluded. All the scripts used for clustering and module scores are listed in the "Code availability" section.

Logistic regression

The probability that the transcriptome for each single MRT cell is similar to each mouse neural crest cluster was estimated with logistic regression in R, as described previously[12]. The fetal mouse neural crest dataset[18] was extracted from GEO (accession

number GSE129114) selecting for the following cells: Wnt1 E8.5 whole embryo, Wnt1 E9.5 trunk and Wnt1 E10.5 tail. Cluster assignment for each cell in the combined dataset was extracted from supplementary table 9[18]. The second dataset of fetal mouse organogenesis[19] was extracted from GEO (accession number GSE119945), selecting cells from early mesenchymal, glial and neural trajectories with the following cell annotations (Main_cell_type field in GSE119945_cell_annotate.csv.gz file): Early mesenchyme, Intermediate Mesoderm, Connective tissue progenitors, Schwann cell precursor, Sensory neurons, Neural Tube, Neural progenitor cells, Radial glia. Mouse gene symbols were converted to their human orthologues using biomaRt version 2.40.4 in R. Raw transcript counts for human genes that mapped to multiple mouse genes were excluded from further analysis. The logistic regression model was trained on the neural crest dataset with R package cv.glmnet, and was used to generate probability scores for the MRT cells.

Bulk mRNA sequencing

Total RNA was extracted using Trizol reagent (ThermoFisher) and quality-checked using Bioanalyzer2100 RNA Nano 6000 chips (Agilent). Sequencing libraries were prepared using the NEBNext® Ultra™ RNA Library Prep Kit (New England Biolabs) according to manufacturer's protocol. Stranded paired-end sequencing (PE150) was performed on the Illumina HiSeq or Illumina NovaSeq platform by Novogene (Hong Kong).

For data analysis, TruSeq3 adapters were removed from reads using Trimmomatic version 0.36.5, followed by sliding window trimming to trim low-quality bases (<20 average quality over 4 bases). Unpaired reads were removed for subsequent steps. Reads were mapped to the reference genome (GRCh38) using STAR version 2.6.0 and assigned using featureCounts version 1.6.3 based on gene annotation GENCODE version 28. Gene expression changes were calculated using the R package DESeq2 version 1.22.1, and differentially expressed genes were determined by the Wald significance test with multiple-testing correction (FDR < 0.01 and fold change > 2) (**Supplementary Data 3**). The overlap of differentially expressed genes for three patient lines was tested using a multi-set hypergeometric test[48]. For *SMARCB1*+ program validation, z-scores were generated from gene expression data acquired from a paediatric renal tumour biobank[17] and downloaded from the GTEx portal (https://gtexportal.org/home/datasets, GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct.gz). For principal component analysis (PCA), gene counts were normalized by variance stabilizing transformation (VST, DESeq2), and principal components were generated using prcomp in R. For heatmaps, VST counts were corrected for batch effect using R package limma version 3.38.3. Heatmaps were generated using R package pheatmap version 1.0.12, scaling was applied by gene. Gene set enrichment analysis[49] for hallmark and perturbation[24] gene sets was performed using the R package clusterProfiler version 3.10.1. Genes were ranked according to *SMARCB1*-induced mRNA changes that were averaged for 60T, 78T and 103T. For the

comparison of *SMARCB1* re-expression and combination treatment, unordered gene sets (generated by intersection of differentially expressed genes, **Supplementary Data 3**) were submitted for hallmark pathway enrichment analysis.

Connectivity map

A connectivity map was generated by submitting the top 100 genes upregulated by *SMARCB1* re-expression in MRT organoids to the online CLUE query tool[20]. Input genes were selected based on significantly upregulated genes overlapping for 60T, 78T and 103T. Genes were ranked based on average gene expression changes. Connectivity map drugs were ranked by the median percentage of similarity.

Western blot

Western blot on MRT organoids was performed as described[50] using the following antibodies:

SMARCB1/INI1 A-5 (Santa Cruz Biotechnology, sc-166165, 1:1000), beta-tubulin H235 (Santa Cruz Biotechnology, sc-9104, 1:1000).

Drug testing and synergy assays

MRT organoids were harvested and washed in ice-cold AdDF+++. Organoids were subsequently filtered using a 70 µm cell strainer (Falcon) and resuspended in 5% BME in medium. Next, ~500 organoids were plated using the Multi-drop™ Combi Reagent Dispenser on repellent black 384-well plates (Corning) to which medium with drugs were added using the Tecan D300e Digital Dispenser. Four technical replicates were included in each experiment. Five days after adding drugs, cell viability was measured using CellTiter-Glo 3D reagent (Promega) according to manufacturer's instructions. Results were normalized to the DMSO vehicle (100%). Dose response curves were generated by nonlinear regression (curve fit) using GraphPad Prism v8.0.2 Synergy scores were generated by R package synergyfinder version 1.8.0 using the ZIP method[51].

Regrowth assay

MRT and normal kidney organoids were dissociated into single cells using mechanical disruption and TrypLE Express (Invitrogen, 12605036), and 5000 cells were seeded in 5µl BME droplets in two separate flat-bottom 96-well plates (Corning). Drugs were added either one (78T), two (60T and 103T) or three (normal kidney donor 1 and 2) days after seeding to normalize for growth rate. Vorinostat and sirolimus concentration were selected based on the highest synergy score for each patient line. Sirolimus concentration was fixed for all experiments (2nM), while vorinostat concentration varied (60T: 1µM, 78T: 1µM, 103T: 3µM). Both concentrations of vorinostat (1µM & 3µM) were tested for 78T and normal kidney organoids. Five days after addition of the drugs, cell viability was measured in the first plate using CellTiter-Glo 3D reagent (Promega) according to manufacturer's instructions. In the second plate, medium was

exchanged for medium without drugs. Five days after drug removal, cell viability was measured. Results were normalized to the DMSO vehicle of T1 (100%).

Statistics and reproducibility

For **Fig. 1**, genetic lineage tracing experiments were performed for n=2 independent donors. Images of **Fig. 1b-d, i** correspond to the regions that were sampled for genetic lineage tracing experiments, which were imaged before and after the collection of material (n=2). Representative images of **Fig. 1d** were derived from n=2 independent donors, for which at least n=3 areas with nerve tissue were assessed for each donor. For **Fig. 2b** and **Supplementary Fig. 4a, c, d** representative images were derived from n=3 independent immunofluorescence experiments. The western blot image in **Supplementary Fig. 3a** was representative of n=3 independent experiments. For **Fig. 3b**, representative images were derived from n=2 independent immunofluorescence experiments. For **Supplementary Fig. 7a-b**, representative images were derived from at least n=3 independent experiments.

Data and materials availability

Raw sequencing data have been deposited in the European Genome-phenome Archive (EGA, www.ebi.ac.uk/ega/). Accession numbers are EGAD00001006574 (bulk mRNA-seq; https://ega-archive.org/datasets/EGAD00001006574) and EGAD00001006296 (WGS & scRNA-seq; https://ega-archive.org/datasets/EGAD00001006296). The data is available under restricted access. Access can be granted by contacting biobank@prinsesmaximacentrum.nl (EGAD00001006574) or datasharing@sanger.uk.ac (EGAD00001006296). DNA methylation data are available under GEO (www.ncbi.nlm.nih.gov/geo/) accession number GSE161814. Processed scRNA-seq data (derived from raw sequencing data deposited under accession number EGAD00001006296) are available at https://github.com/kheleon/mrt-paper[52]. Content includes filtered_gene_bc_matrices and filtered_feature_bc_matrix folders from cellranger output. The fetal mouse neural crest dataset[18] was extracted from GEO (www.ncbi.nlm.nih.gov/geo/, accession number GSE129114). The second dataset of fetal mouse organogenesis[19] was extracted from GEO (www.ncbi.nlm.nih.gov/geo/, accession number GSE119945). Gene expression data of the paediatric renal tumour biobank[17] was extracted from EGA (www.ebi.ac.uk/ega/, accession number EGAD00001005318 https://ega-archive.org/datasets/EGAD00001005318 and EGAD00001005319 https://ega-archive.org/datasets/EGAD00001005319). Gene expression data of normal tissues was extracted from GTEx portal (https://gtexportal.org/home/datasets, GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct.gz). The remaining data are available within the Article, Supplementary Information, Supplementary Data and Source Data provided with this paper or are available from the authors upon request.

Code availability

Jupyter notebooks for processing of scRNA-seq data are available at https://github.com/kheleon/mrt-paper[52].

Acknowledgements

We thank the clinical teams who facilitated our research, as well as the Human Research Tissue Bank at Addenbrooke's hospital and Domenico Castigliego (UMC Utrecht) for immunohistochemical analyses. We thank the CCLG Tissue Bank for access to samples, and contributing CCLG Centres, including members of the ECMC Paediatric Network. The CCLG Tissue Bank is funded by Cancer Research UK and CCLG. J.D. received funding from the European Research Council (ERC) starting grant 850571 (J.D.), the Dutch Cancer Society (KWF)/Alpe d'HuZes Bas Mulder Award to J.D. (KWF/Alpe d'HuZes, 10218), Foundation Children Cancer Free (KiKa 338, L.C.) and Nikai 4 Life. S.B. was supported by the St. Baldrick's Foundation (Robert J. Arceci International Award to S.B.) and the Wellcome Trust (fellowship to S.B.; Sanger core funding). We are profoundly indebted to the children and their families who, at a most difficult time, agreed to participate in our research.

Author information

S.B. and J.D. conceived of the study. L.C. performed *in vitro* studies. E.K. performed transcriptional analyses, aided by L.C. Phylogenetic analyses were conducted by T.H.H.C. Pathological expertise was provided by T.R.W.O, L.H., C.T., R.R.K. C.C. performed drug assays. Statistical expertise was provided by M.D.Y. L.C. and F.A.V.B. prepared samples for single cell analyses. P.E. and A.M. were involved in tissue processing. Samples were curated and/or experiments were performed by H.S., L.M., E. W. H., M.H.E., J.N. K.S., and K.A. M.K. performed DNA methylation assays. S.B. supervised phylogenetic and transcriptional analyses. J.D. supervised experimental work. S.B. and J.D. wrote the manuscript and directed this work.

Ethics declarations

Competing interests

None.

References

1. Biegel, J. A. *et al.* Germ-line and acquired mutations of INI1 in atypical teratoid and rhabdoid tumors. *Cancer Res* **59**, 74-79 (1999).

2. Hasselblatt, M. *et al.* SMARCA4-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. *Acta Neuropathol* **128**, 453-456, doi:10.1007/s00401-014-1323-x (2014).

3. Versteege, I. *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394**, 203-206, doi:10.1038/28212 (1998).

4. Eaton, K. W., Tooke, L. S., Wainwright, L. M., Judkins, A. R. & Biegel, J. A. Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors. *Pediatr Blood Cancer* **56**, 7-15, doi:10.1002/pbc.22831 (2011).

5. Filbin, M. & Monje, M. Developmental origins and emerging therapeutic opportunities for childhood cancer. *Nat Med* **25**, 367-376, doi:10.1038/s41591-019-0383-9 (2019).

6. Han, Z. Y. *et al.* The occurrence of intracranial rhabdoid tumours in mice depends on temporal control of Smarcb1 inactivation. *Nat Commun* **7**, 10421, doi:10.1038/ncomms10421 (2016).

7. Chun, H. E. *et al.* Identification and Analyses of Extra-Cranial and Cranial Rhabdoid Tumor Molecular Subgroups Reveal Tumors with Cytotoxic T Cell Infiltration. *Cell Rep* **29**, 2338-2354 e2337, doi:10.1016/j.celrep.2019.10.013 (2019).

8. Chun, H. E. *et al.* Genome-Wide Profiles of Extra-cranial Malignant Rhabdoid Tumors Reveal Heterogeneity and Dysregulated Developmental Pathways. *Cancer Cell* **29**, 394-406, doi:10.1016/j.ccell.2016.02.009 (2016).

9. Gadd, S., Sredni, S. T., Huang, C. C., Perlman, E. J. & Renal Tumor Committee of the Children's Oncology, G. Rhabdoid tumor: gene expression clues to pathogenesis and potential therapeutic targets. *Lab Invest* **90**, 724-738, doi:10.1038/labinvest.2010.66 (2010).

10. Young, M. D. *et al.* Single cell derived mRNA signals across human kidney tumors. *bioRxiv*, 2020.2003.2019.998815, doi:10.1101/2020.03.19.998815 (2020).

11. Filbin, M. G. *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **360**, 331-335, doi:10.1126/science.aao4750 (2018).

12. Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599, doi:10.1126/science.aat1699 (2018).

13. Coorens, T. H. H. *et al.* Lineage-Independent Tumors in Bilateral Neuroblastoma. *N Engl J Med* **383**, 1860-1865, doi:10.1056/NEJMoa2000962 (2020).

14. Olafsson, S. *et al.* Somatic Evolution in Non-neoplastic IBD-Affected Colon. *Cell* **182**, 672-684 e611, doi:10.1016/j.cell.2020.06.036 (2020).

15. Yoshida, K. *et al.* Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature* **578**, 266-272, doi:10.1038/s41586-020-1961-1 (2020).

16. Coorens, T. H. H. *et al.* Embryonal precursors of Wilms tumor. *Science* **366**, 1247-1251, doi:10.1126/science.aax1323 (2019).

17. Calandrini, C. *et al.* An organoid biobank for childhood kidney cancers that captures disease and tissue heterogeneity. *Nat Commun* **11**, 1310, doi:10.1038/s41467-020-15155-6 (2020).

18. Soldatov, R. *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**, doi:10.1126/science.aas9536 (2019).

19. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502, doi:10.1038/s41586-019-0969-x (2019).

20. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e1417, doi:10.1016/j.cell.2017.10.049 (2017).

21. Muscat, A. *et al.* Low-Dose Histone Deacetylase Inhibitor Treatment Leads to Tumor Growth Arrest and Multi-Lineage Differentiation of Malignant Rhabdoid Tumors. *Clin Cancer Res* **22**, 3560-3570, doi:10.1158/1078-0432.CCR-15-2260 (2016).

22. Schutgens, F. *et al.* Tubuloids derived from human adult kidney and urine for personalized disease modeling. *Nat Biotechnol* **37**, 303-313, doi:10.1038/s41587-019-0048-8 (2019).

23. Simmons, J. K. *et al.* Cooperative Targets of Combined mTOR/HDAC Inhibition Promote MYC Degradation. *Mol Cancer Ther* **16**, 2008-2021, doi:10.1158/1535-7163.MCT-17-0171 (2017).

24. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).

25. Erkek, S. *et al.* Comprehensive Analysis of Chromatin States in Atypical Teratoid/Rhabdoid Tumor Identifies Diverging Roles for SWI/SNF and Polycomb in Gene Regulation. *Cancer Cell* **35**, 95-110 e118, doi:10.1016/j.ccell.2018.11.014 (2019).

26. Johann, P. D. *et al.* Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes. *Cancer Cell* **29**, 379-393, doi:10.1016/j.ccell.2016.02.001 (2016).

27. Torchia, J. *et al.* Integrated (epi)-Genomic Analyses Identify Subgroup-Specific Therapeutic Targets in CNS Rhabdoid Tumors. *Cancer Cell* **30**, 891-908, doi:10.1016/j.ccell.2016.11.003 (2016).

28. Vitte, J., Gao, F., Coppola, G., Judkins, A. R. & Giovannini, M. Timing of Smarcb1 and Nf2 inactivation determines schwannoma versus rhabdoid tumor development. *Nat Commun* **8**, 300, doi:10.1038/s41467-017-00346-5 (2017).

29. Biegel, J. A., Busse, T. M. & Weissman, B. E. SWI/SNF chromatin remodeling complexes and cancer. *Am J Med Genet C Semin Med Genet* **166C**, 350-366, doi:10.1002/ajmg.c.31410 (2014).

30. Robinson, G. W. *et al.* Irreversible growth plate fusions in children with medulloblastoma treated with a targeted hedgehog pathway inhibitor. *Oncotarget* **8**, 69295-69302, doi:10.18632/oncotarget.20619 (2017).

31. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

32. Jones, D. *et al.* cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Curr Protoc Bioinformatics* **56**, 15 10 11-15 10 18, doi:10.1002/cpbi.20 (2016).

33. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289-300, doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x (1995).

34. Buels, R. *et al.* JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol* **17**, 66, doi:10.1186/s13059-016-0924-1 (2016).

35. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* **107**, 16910-16915, doi:10.1073/pnas.1009843107 (2010).

36. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994-1007, doi:10.1016/j.cell.2012.04.023 (2012).

37. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54, doi:10.1038/nature17676 (2016).

38. Brunner, S. F. *et al.* Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538-542, doi:10.1038/s41586-019-1670-9 (2019).

39. Koo, B. K. *et al.* Controlled gene expression in primary Lgr5 organoid cultures. *Nat Methods* **9**, 81-83, doi:10.1038/nmeth.1802 (2011).

40. Fumagalli, A. *et al.* Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proc Natl Acad Sci U S A* **114**, E2357-E2364, doi:10.1073/pnas.1701219114 (2017).

41. Dekkers, J. F. *et al.* High-resolution 3D imaging of fixed and cleared organoids. *Nat Protoc* **14**, 1756-1771, doi:10.1038/s41596-019-0160-8 (2019).

42. Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).

43. Kang, H. M. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol* **36**, 89-94, doi:10.1038/nbt.4042 (2018).

44. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).

45. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289, doi:10.1016/j.cels.2018.11.005 (2019).

46. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).

47. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**, 296, doi:10.1186/s13059-019-1874-1 (2019).

48. Wang, M., Zhao, Y. & Zhang, B. Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep* **5**, 16923, doi:10.1038/srep16923 (2015).

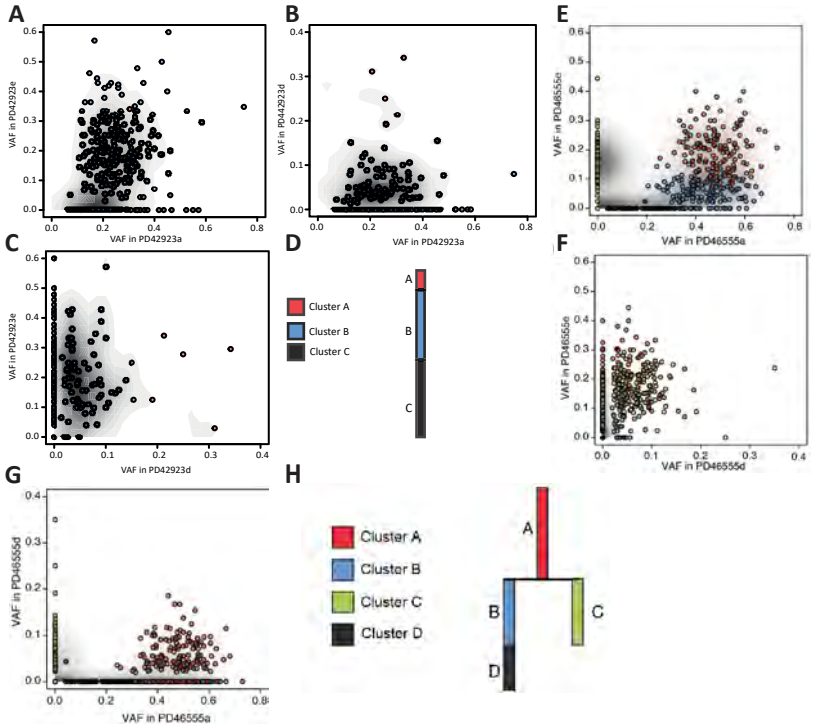
49. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

50. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43-47, doi:10.1038/nature14415 (2015).

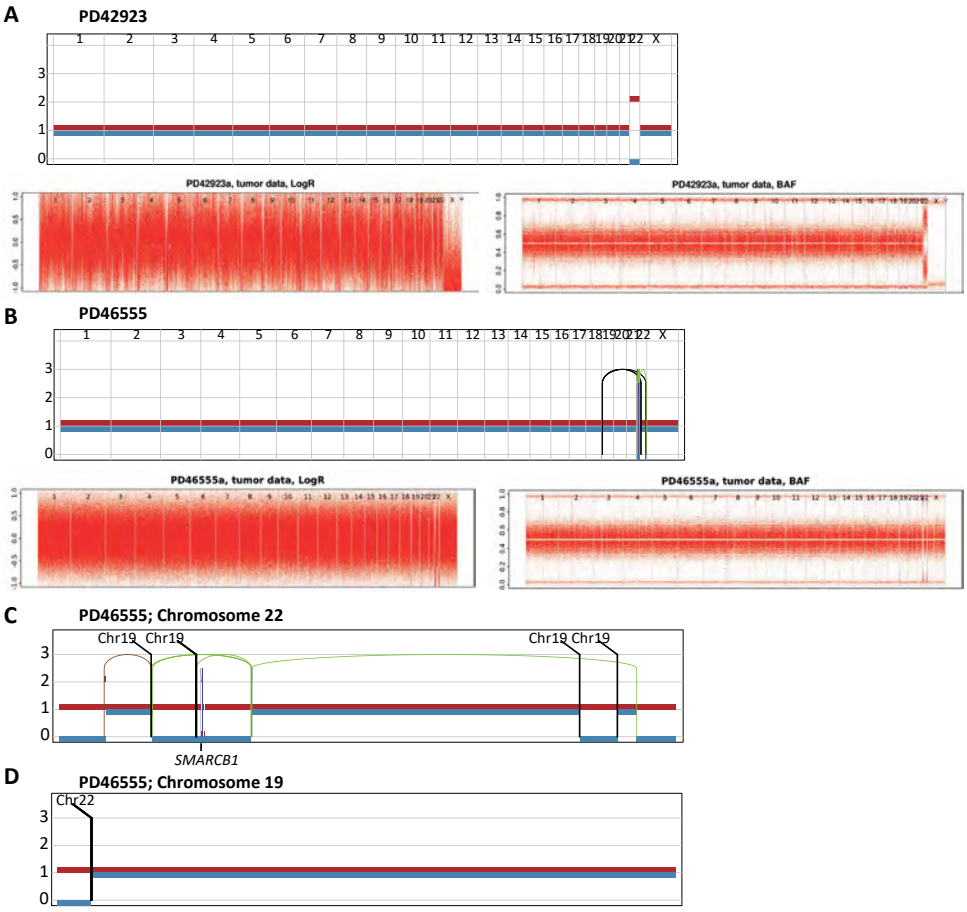
51. Yadav, B., Wennerberg, K., Aittokallio, T. & Tang, J. Searching for Drug Synergy in Complex Dose-Response Landscapes Using an Interaction Potency Model. *Comput Struct Biotechnol J* **13**, 504-513, doi:10.1016/j.csbj.2015.09.001 (2015).

52. Custers, L. *et al.* Somatic mutations and single cell transcriptomes reveal the root of malignant rhabdoid tumours. Zenodo. https://doi.org/10.5281/zenodo.4446540 (2020)

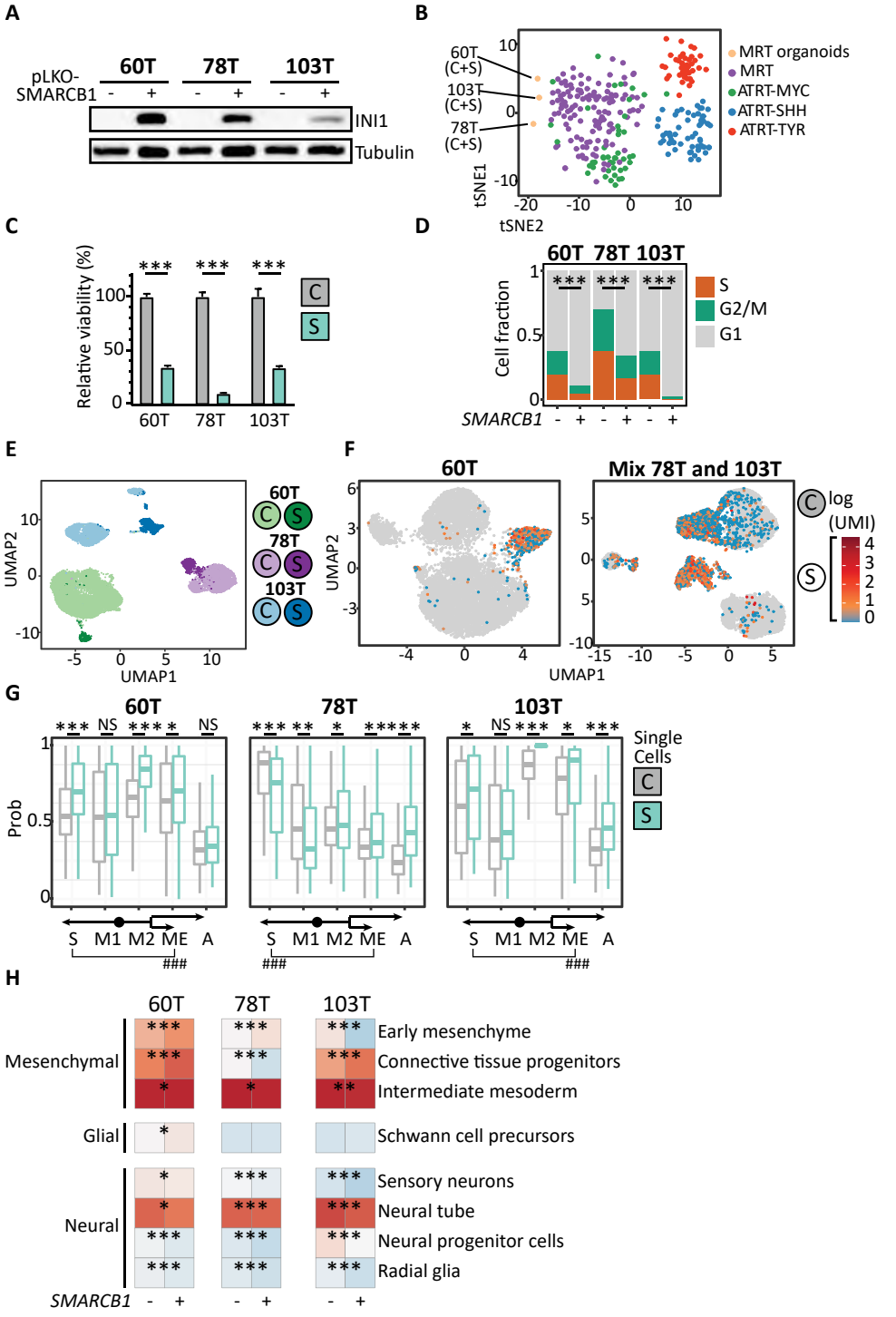
Supplementary information



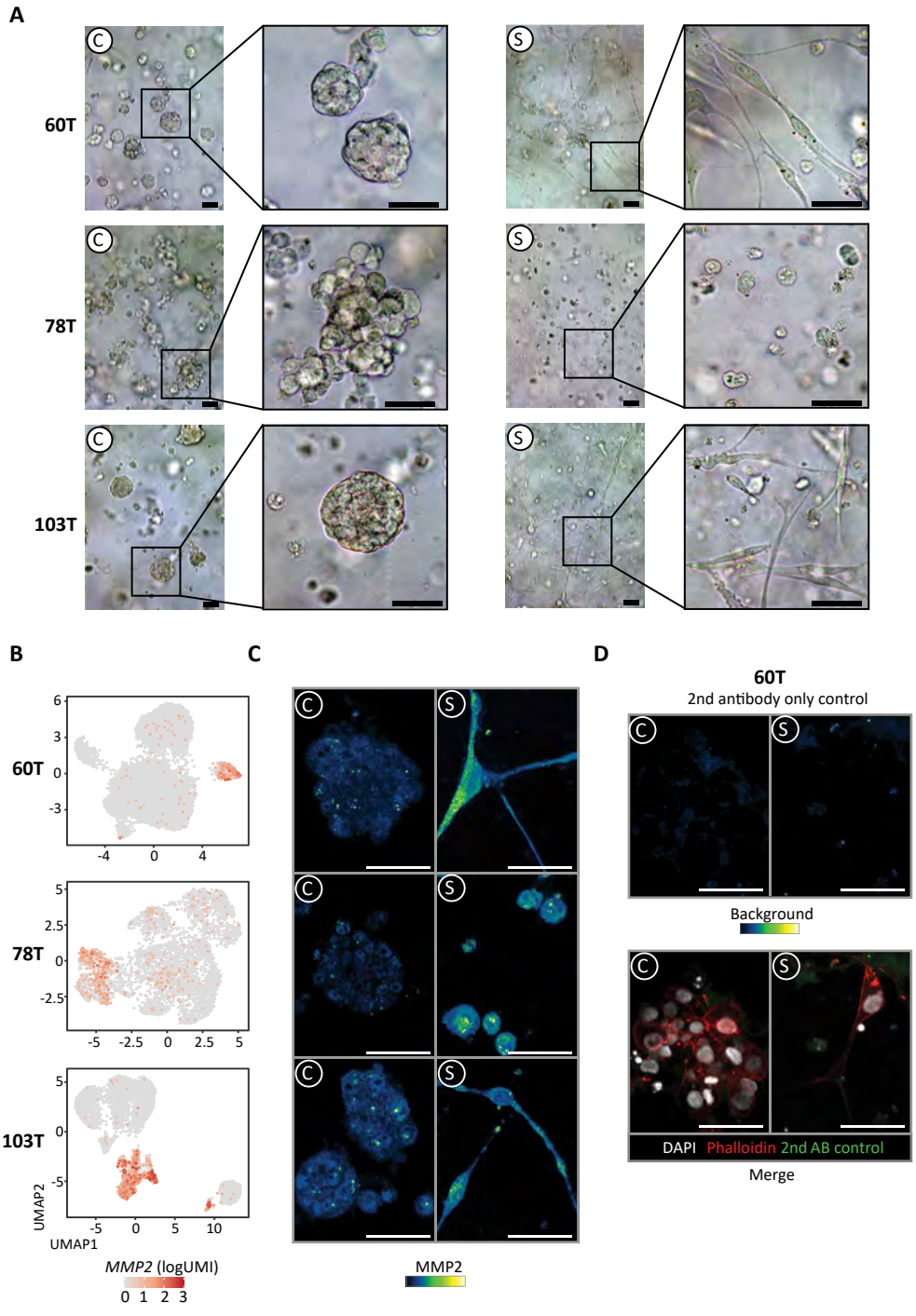
Supplementary Figure 1 | Clonal reconstruction of normal tissues in PD42923 and PD46555. **(A)** - **(D)** Clustering of SNVs across PD42923a, -d, and -e revealed four clusters contributing to their clonal architecture. Comparison of VAFs of these mutations between the different samples of PD42923 **(A-C)** reveals the phylogenetic ordering of these clusters **(D)**. **(E)** - **(H)** Clustering of SNVs across PD46555a, -d, and -e revealed four clusters contributing to their clonal architecture. Comparison of VAFs of these mutations between the different samples of PD46555 **(E-G)** reveals the phylogenetic ordering of these clusters **(H)**. The possibility that we observed tumour mutations in normal tissue because of contamination with tumour cells was addressed in three ways. Firstly, specialist pathologists reviewed sections of the pieces of tissues that we sequenced. In the hilar samples there were no tumour cells seen. In the nerve root samples, occasional tumour cells were seen. However, the frequency of these was small and could not account for the clone size observed in nerve root samples. PD46555g (cervical spine sample) was found to be heavily contaminated with tumour cells (>10%) upon inspection by the pathologists. This fraction outnumbered the INI1 negative normal cells, and hence any genomic signal of mutational sharing is likely derived from the contaminant. Therefore, we have excluded this sample from further analysis. Secondly, we performed INI1 staining to validate the expected loss of INI1 staining in normal Schwann cells, as predicted from the genomic data. This confirmed that there were morphologically normal looking Schwann cells that lacked INI1 staining in all samples that exhibited biallelic loss of SMARCB1 genomically. Thirdly, we applied a Dirichlet process mixture modelling to define the cluster composition of the normal tissues. Tumour contamination would manifest by a single cluster, corresponding to the clonal tumour somatic mutations, being present in the normal at the contamination rate. However, we observed that the clonal tumour mutations fall into two clusters with a differential contribution to the normal tissues. This would be inconsistent with a single, clonal contaminant. In addition, such contamination would not explain mutations that are private to the normal tissues and not found in tumour tissues, of which there were many.



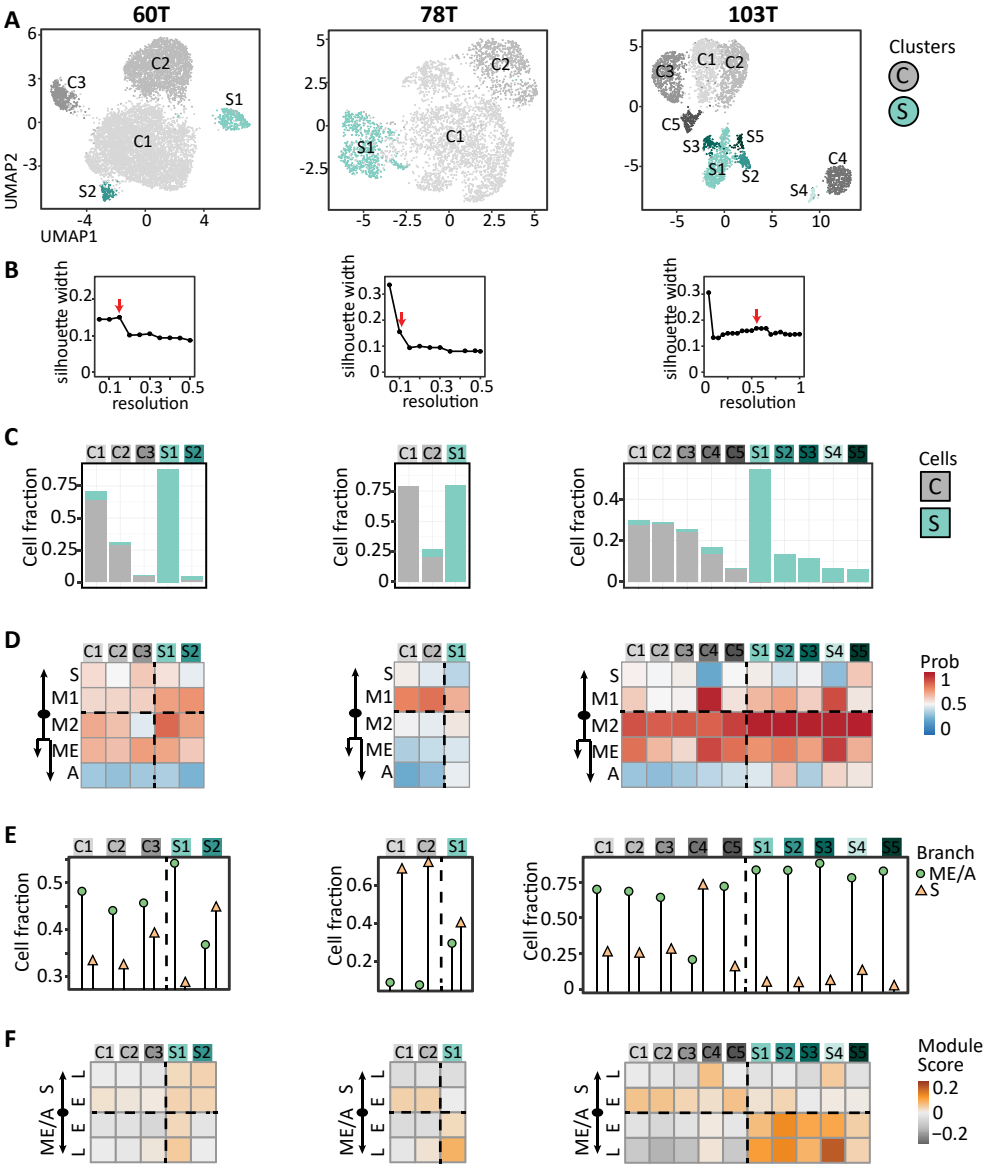
Supplementary Figure 2 | Overview of CNVs and SVs in two MRT samples.
Copy number variants (CNVs) and structural variants (SVs) for the malignant rhabdoid tumours of PD42923 **(A)** and PD46555 **(B)**, genome-wide. Major and minor copy numbers are displayed in red and blue, respectively. Black lines denote interchromosomal translocations, green inversions, brown tandem-duplications, and blue deletions. The genome-wide distributions of the log-transformed normalised read depth (logR) and B-allele frequency (BAF) are displayed below the copy number landscapes. In both cases, chromosome 22, the location of SMARCB1, is affected. PD42923 displays a copy number neutral loss of heterozygosity of chromosome 22. For PD46555, enhanced representations of chromosome 22 **(C)**, and chromosome 19 **(D)** reveal complex rearrangements affecting the tumour.


A
pLKO-
SMARCB1
60T
78T
103T
INI1
Tubulin
B
60T (C+S)
103T (C+S)
78T (C+S)
tSNE1
tSNE2
MRT organoids
MRT
ATRT-MYC
ATRT-SHH
ATRT-TYR
C
Relative viability (%)
C
S
D
60T 78T 103T
Cell fraction
S
G2/M
G1
SMARCB1
E
UMAP2
UMAP1
60T
78T
103T
F
60T
Mix 78T and 103T
log (UMI)
G
Prob
Single Cells
S M1 M2 ME A
H
Mesenchymal
Glial
Neural
Early mesenchyme
Connective tissue progenitors
Intermediate mesoderm
Schwann cell precursors
Sensory neurons
Neural tube
Neural progenitor cells
Radial glia
SMARCB1

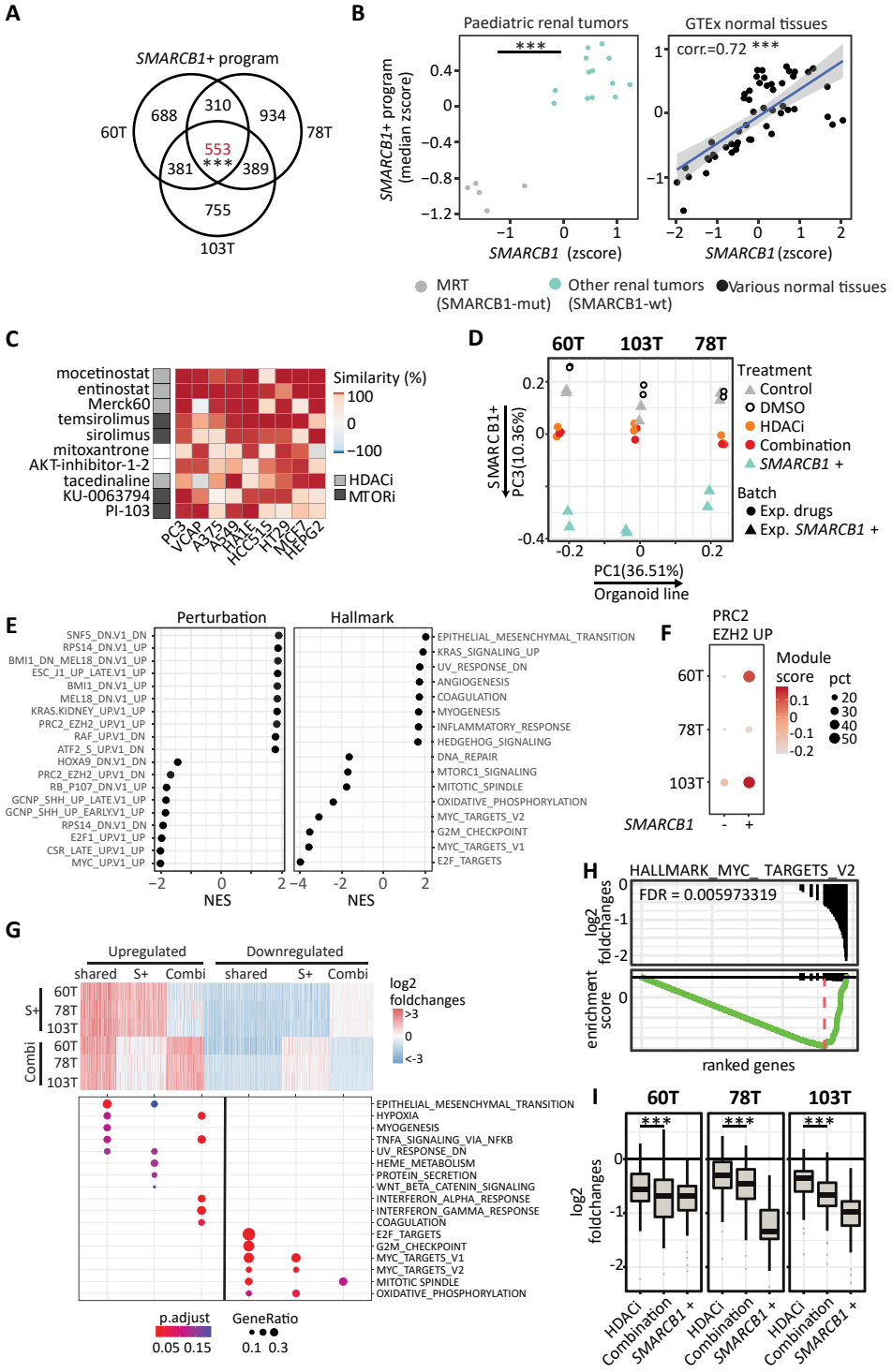
Supplementary Figure 3 | *SMARCB1* reconstitution in organoids as model for MRT differentiation. **(A)** Western blot analysis of INI1 expression in MRT organoids lentivirally transduced with either a control (-) or *SMARCB1* expression (+) plasmid. Tubulin protein levels are used as loading control. Molecular weight markers are indicated in kilodalton (kDa). Source data are provided as a Source Data file. **(B)** t-SNE representation of DNA methylation profiles of MRT control (C) and *SMARCB1*+ (S) organoids compared to MRT and AT/RT (subtypes MYC, SHH, TYR) tissue, showing that MRT organoids cluster closely to MRT tissue independent of *SMARCB1* expression. t-SNE coordinates of MRT control and *SMARCB1*+ organoids completely overlap. Different patient lines are indicated. **(C)** Bar graphs represent cell viability of MRT *SMARCB1*+ (green) relative to MRT control organoids (gray). Mean and SD (error bars) of 4 technical replicates (dot) are indicated. Source data are provided as a Source Data file. Growth was assessed by comparing control to *SMARCB1*+ organoids. P-values were calculated using an unpaired Student's t test (two-tailed): * <0.05, ** <0.01, *** <0.001 (p-value: 60T=2.1e-8, 78T=1.4e-8, 103T=2.5e-6). **(D)** Stacked bar plot represents relative frequencies of single cells from MRT control (-) or *SMARCB1*+ (+) organoids annotated for cell cycle phase (derived from single cell transcriptomes). Colours distinguish G1 (gray), G2/M (green) or S (orange) phase. Frequencies of cell cycle phase annotations were compared between control and *SMARCB1*+ organoids for each patient line. P-values were calculated using a chi-square test (two-tailed): *** <1e-15 (-log10(p-value): 60T=27, 78T=77, 103T=96). **(E)** UMAP representation of single cells from MRT organoid lines 60T (green, control/*SMARCB1*+: 8059/425 cells), 78T (purple, control/*SMARCB1*+: 3195/806 cells) and 103T (blue, control/*SMARCB1*+: 2694/953 cells) distinguishing control (light) and *SMARCB1*+ (dark) cells. **(F)** UMAP representation of single cells separated by sample: 60T (control/*SMARCB1*+: 8059/638 cells) or mix 78T and 103T (control/*SMARCB1*+: 7214/3389 cells; demultiplexing was performed as described in Methods). For these UMAPs, cells were not filtered for *SMARCB1* expression and thus include non-transduced cells, demonstrating that unsuccessfully transduced cells in the *SMARCB1*+ sample cluster with control-transduced cells thereby excluding batch effects. Colour distinguishes the batch of MRT control cells (gray) from the batch of MRT *SMARCB1*+ cells (colour-code from blue to red refers to *SMARCB1* transcript levels (unique molecular identifier (UMI)). *SMARCB1*-negative cells were filtered out from *SMARCB1*+ samples for all subsequent analysis. **(G)** Box plots represent single cell similarity scores (n = 60T control/*SMARCB1*+: 8059/425 cells; 78T control/*SMARCB1*+: 3195/806 cells; 103T control/*SMARCB1*+: 2694/953 cells) for cell types of the mesenchyme/autonomic (ME/A) or sensory (S) branch (illustrated in Fig. 2a). Box plots indicate median (middle line), 25th and 75th percentile (box). Whiskers represent the range excluding outliers. Mesenchyme and sensory similarity were compared for control cells to identify the major neural crest cell type signal at baseline. P-values were calculated using a paired Student's t test (two-tailed): ### <1e-15 below figure to indicate cell type with highest average similarity score (-log10(p-value): 60T=100, 78T=Inf, 103T=52). Additionally, similarity scores were compared between control and *SMARCB1*+ cells. P-values were calculated using an unpaired Student's t test (two-tailed): * <1e-3, ** <1e-9, *** <1e-15 (exact p-values are indicated in Supplementary Table 4). **(H)** Heatmaps represent similarity of MRT control (-) and *SMARCB1*+ (+) cells to a mouse organogenesis cell type reference[19], comparing early mesenchymal and neural cell type similarities. Colours represent the average probability (prob) that the MRT cells are similar to the indicated cell type (predicted similarity score estimated by logistic regression[12]). Changes in similarity scores between control and *SMARCB1*+ cells were assessed. P-values were calculated using an unpaired Student's t test (two-tailed): * <1e-3, ** <1e-9, *** <1e-15 (exact p-values are indicated in Supplementary Table 4).


A
C
S
60T
78T
103T
B
C
D
60T
2nd antibody only control
Background
DAPI Phalloidin 2nd AB control
Merge
UMAP1
UMAP2
MMP2 (logUMI)
0 1 2 3
MMP2

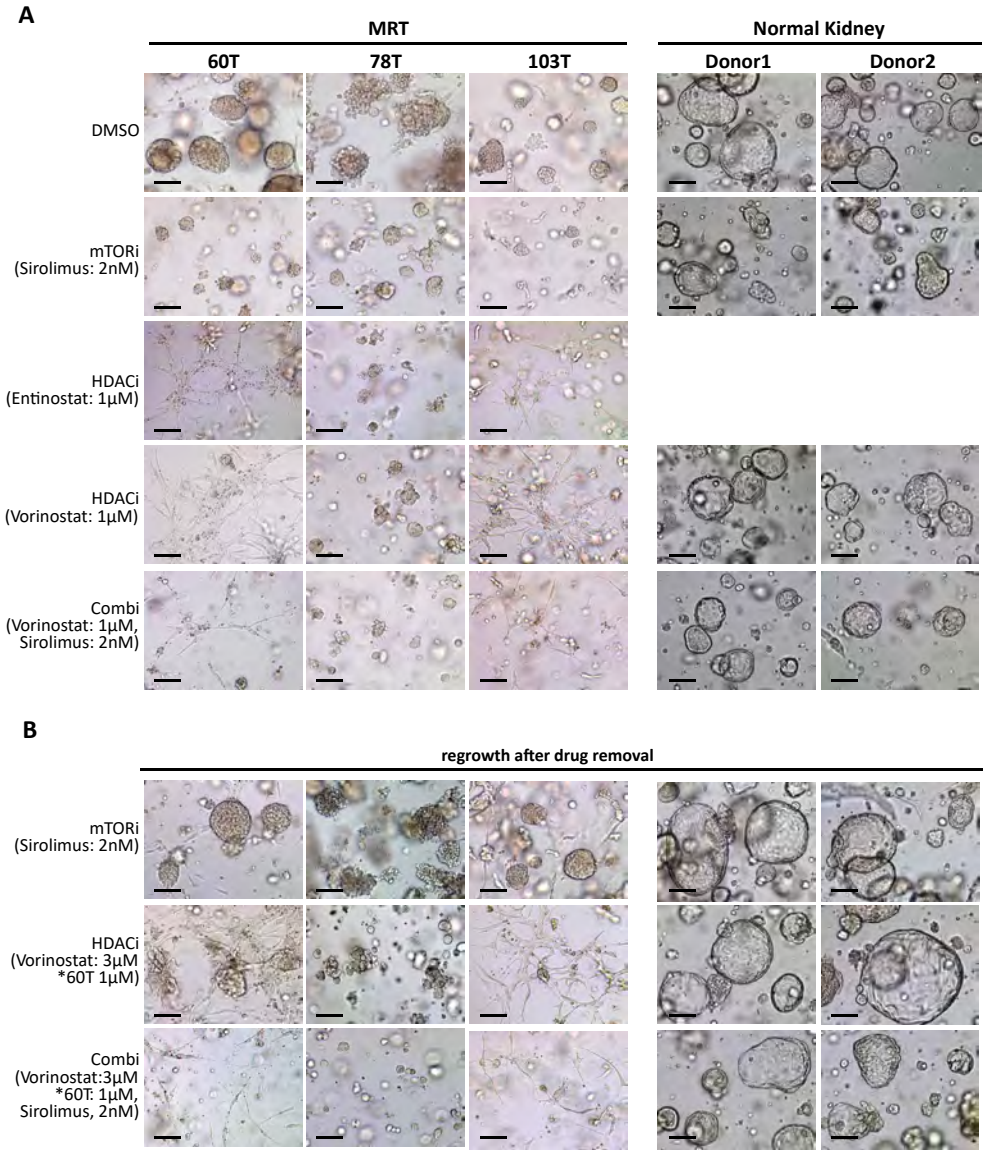
Supplementary Figure 4 | Morphological transformation of MRT organoids upon SMARCB1 reconstitution.
(A) Representative brightfield images of MRT control (C) and SMARCB1+ (S) organoid lines including zoom in. Scale bars equal 100 µm. **(B)** UMAP representation of single cells from MRT control and *SMARCB1*+ organoid lines. Colour-code from gray to red refers to *MMP2* transcript levels (unique molecular identifier (UMI)). **(C)** Representative immunofluorescence images of MRT control (C) and *SMARCB1*+ (S) organoids stained for MMP2. Merged images are shown in Fig. 2b. Colour-code represents immunofluorescent signal intensity. Scale bars equal 50 µm. **(D)** Representative immunofluorescence images of 60T control (C) and *SMARCB1*+ (S) organoids only incubated with the secondary antibody, to determine background signal (top panel). Colour-code represents immunofluorescent signal intensity. Bottom panel shows a merge of DAPI (white; nuclei), phalloidin (red; membranes) and background signal (green). Scale bars equal 50 µm.


A
60T
78T
103T
Clusters
C
S
UMAP2
UMAP1
B
silhouette width
resolution
C
Cell fraction
Cells
D
Prob
E
Branch
ME/A
S
F
Module Score

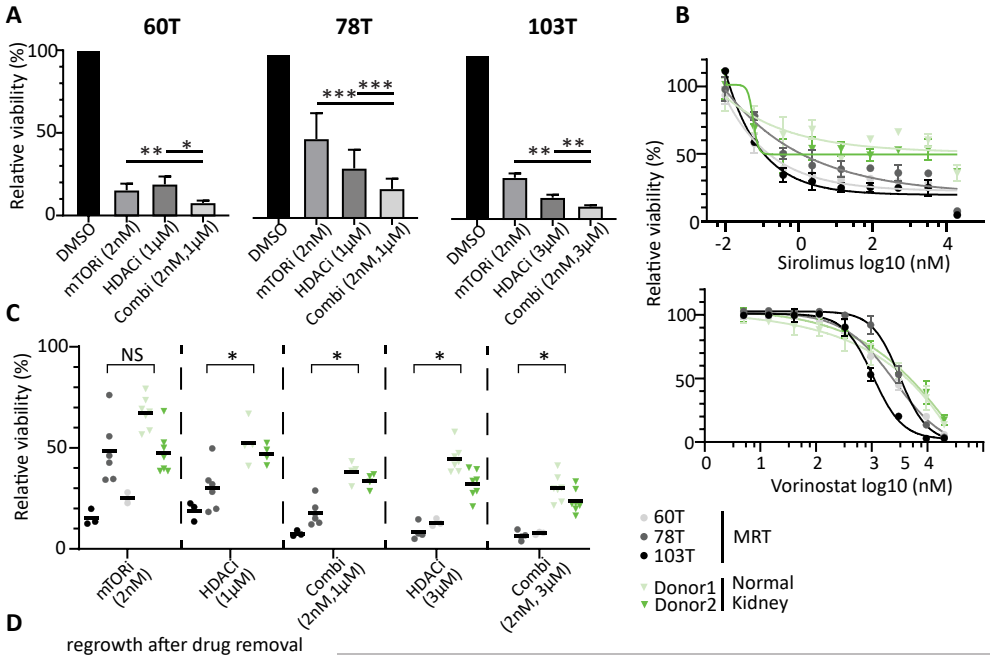
Supplementary Figure 5 | Intra-organoid heterogeneity of neural crest signals.
(A) UMAP representation of single cells from MRT control and *SMARCB1+* organoid lines showing cluster assignment. Cells are coloured for either control (gray: light to dark) or *SMARCB1+* (green: light to dark) dominant clusters, which are numbered according to cluster size. (60T C1-3/S1-2: 5,209/2,350/425/375/125 cells, 78T C1-2/S1: 2,624/725/652 cells, 103T C1-5/S1-5: 769/768/664/395/172/521/126/110/64/58 cells) **(B)** Scatter plot depicts a series of resolutions used for the Louvain-based clustering of MRT control and *SMARCB1+* single cells, and the corresponding quantification of average silhouette width, which was used to determine optimal clustering resolution. The resolution with the highest average silhouette width (arrow) was used for subsequent analyses. **(C)** Stacked bar plots represent relative frequencies of control (gray) and *SMARCB1+* (green) single cells for each cluster, showing a segregation of cells upon *SMARCB1* reconstitution. **(D)** Heatmaps show average similarity score per cluster dominant for either MRT control (gray) or *SMARCB1+* (green) cells. Clusters were compared to cell types from the mesenchymal/autonomic or sensory branch, showing heterogeneity of neural crest signals within and between patient lines. Abbreviations are indicated in Fig. 2a. **(E)** Graphs represent relative frequencies of cell type annotations for MRT control (gray) or *SMARCB1+* (green) dominant clusters. Cell type annotation was assigned for each single cell based on the highest similarity score. Cell type annotations were grouped into the mesenchymal/autonomic (ME/A; green circle) or sensory (S; orange triangle) differentiation branch. **(F)** Heatmaps represent average gene module scores for MRT control (gray) or *SMARCB1+* (green) dominant clusters. Module scores were generated by averaging gene expression levels per set of genes. Gene sets include marker genes for either sensory (S) or mesenchymal/autonomic (ME/A) differentiation branches, distinguishing early (E) and late (L) differentiation genes.


A
SMARCB1+ program
60T
78T
103T
688
310
934
553
***
381
389
755
B
Paediatric renal tumors
GTEx normal tissues
SMARCB1+ program (median zscore)
SMARCB1 (zscore)
corr.=0.72 ***
MRT (SMARCB1-mut)
Other renal tumors (SMARCB1-wt)
Various normal tissues
C
mocetinostat
entinostat
Merck60
temsirolimus
sirolimus
mitoxantrone
AKT-inhibitor-1-2
tacedinaline
KU-0063794
PI-103
Similarity (%)
HDACi
MTORi
PC3
VCAP
A375
A549
HALE
HCC515
HT29
MCF7
MEPG2
D
60T
103T
78T
Treatment
Control
DMSO
HDACi
Combination
SMARCB1 +
Batch
Exp. drugs
Exp. SMARCB1 +
SMARCB1+ PC3(10.36%)
PC1(36.51%)
Organoid line
E
Perturbation
Hallmark
NES
F
PRC2 EZH2 UP
Module score
pct
SMARCB1
G
Upregulated
Downregulated
shared
S+
Combi
log2 foldchanges
p.adjust
GeneRatio
H
HALLMARK_MYC_TARGETS_V2
FDR = 0.005973319
log2 foldchanges
enrichment score
ranked genes
I
60T
78T
103T
log2 foldchanges
HDACi
Combination
SMARCB1 +

Supplementary Figure 6 | Finding drugs mimicking *SMARCB1*-induced MRT differentiation.
(A) Venn diagram showing the overlap of genes (red) upregulated by *SMARCB1* in MRT organoids, further referred to as the *SMARCB1*+ program. The intersect was assessed for three independent patient lines. P-value was calculated using a multi-set exact test (one-tailed)[48]: *** <1e-15 (p-value: 0) **(B)** Scatter plot showing the relative expression (median z-score) of the *SMARCB1*+ program versus *SMARCB1* expression levels (z-score) for tissues from a paediatric renal tumour biobank[17] (left) or normal tissues (right). The expression of the *SMARCB1*+ program was compared between *SMARCB1* mutant (mut) tissues (MRT; gray; n = 5) and *SMARCB1* wild-type tissues (wt; other renal tumours; green; n = 13). P-values were calculated using an unpaired Student's t test (two-tailed): *** <0.001 (p-value=5.5e-10). In normal tissues (black), the relationship of *SMARCB1* expression levels and the *SMARCB1*+ program was assessed. Correlation coefficient (corr.) and p-value were calculated using Pearson's correlation: *** <0.001 (p-value=2.2e-9). **(C)** Represented is the top 10 of drugs that mimic mRNA changes of *SMARCB1* re-expression, extracted from the CLUE database[20]. Colour-code represents the percentage of similarity for each cell line (x-axis) and drug treatment (y-axis) to *SMARCB1* re-expression. Drugs are annotated as HDAC (dark-gray) or mTOR (light-gray) inhibitors. **(D)** Principal component analysis (PCA) of bulk mRNA-seq samples shows the transcriptional effect of indicated drug treatment (different colours) on MRT organoids, compared with MRT control (gray) or *SMARCB1*+ (green) organoids (n = 2 independent experiments). PC1 and PC3 separate samples by organoid line and drug treatment/*SMARCB1* re-expression. Batch refers to either the experiment with *SMARCB1* re-expression (*SMARCB1*+) or drug treatment (drugs). **(E)** Gene set enrichment analysis (GSEA) for hallmark and perturbation gene sets, using genes ranked by the mRNA changes induced by *SMARCB1*. The top 10 of statistically significant gene sets (adjusted p-value <0.01) is presented for up- and downregulated genes, showing the normalized enrichment score (NES) on the x-axis. P-values were calculated using a permutation test (two-tailed) and were corrected for multiple-testing. **(F)** Genes repressed by polycomb repressive complex subunit EZH2 were upregulated by *SMARCB1* (Supplementary Fig. 6e). This result was confirmed in the scRNA-seq experiment. Dot plot shows gene module scores (average gene expression level for PRC2_EZH2 gene set) for MRT control (-) and *SMARCB1*+ (+) organoids for each patient line. Colour-code from gray to red refers to average module score. Dot size refers to the percentage of cells (pct) expressing the gene module. **(G)** Enrichment analysis for hallmark pathways using unordered sets of differentially expressed genes upon *SMARCB1* re-expression (S+) or combination treatment, either shared or exclusive. The top 5 of significantly enriched terms (adjusted p-value <0.25) are presented (bottom). P-values were calculated using a Fischer's exact test (one-tailed) and were corrected for multiple-testing. Changes in mRNA levels for these genes are visualized in the heatmap (top) for each patient line. **(H)** Gene set enrichment analysis for MYC target genes, with bars representing average mRNA changes of *SMARCB1* re-expression in MRT organoids (top). Enrichment score is visualized by the ranking of genes based on these mRNA changes, showing enrichment for MYC targets with genes downregulated by *SMARCB1* re-expression (bottom). **(I)** Boxplots represent mRNA expression changes of MYC target genes (n = 58 genes) induced by either drug treatment or *SMARCB1* re-expression for each MRT organoid line. Box plots indicate median (middle line), 25th and 75th percentile (box). Whiskers represent the range excluding outliers (dot). Additional effect of combination treatment on MYC target genes (n=58 genes) was assessed by comparing combination with HDACi treatment. P-values were calculated using a paired Student's t test (two-tailed): *** <0.001 (p-value: 60T=4.9e-6, 78T=9.7e-10, 103T=2.6e-14).



Supplementary Figure 7 | HDACi and MTORi combination treatment recapitulates MRT *SMARCB1*+ organoid morphology.
(A) Representative brightfield images of MRT and normal kidney organoid lines treated with vehicle (DMSO), sirolimus (2nM; mTORi), entinostat (1µM; HDACi), vorinostat (1µM; HDACi) or the combination of sirolimus (2nM) and vorinostat (1µM). Scale bars denote 100 µm. **(B)** Representative brightfield images of MRT and normal kidney organoid lines after wash-out of drugs (T2) showing that combination treatment induces a durable differentiation phenotype in MRT. Scale bars denote 100 µm.



Supplementary Figure 8 | Growth inhibitory effects of HDACi and MTORi are MRT specific.

(A) Bar graphs represent the cell viability of MRT organoids after drug treatment (T1) relative to vehicle control (DMSO). Mean and SD (error bars) of independent experiments (dot) are indicated (n = 60T/103T: 3; 78T: 6). Each independent experiment is an average of 4 technical replicates. Source data are provided as a Source Data file. The additional effect of drug combination is assessed by comparing combination treatment with either mTORi (2nM) or HDACi (1μM or 3μM). P-values were calculated using an unpaired Student's t test (two-tailed): * <0.05, ** < 0.01, *** <0.001. (p-value: Combi vs mTORi 60T=0.008, 78T=4.5e-7, 103T=0.003; vs HDACi 60T=0.02, 78T=0.0002, 103T=0.005) **(B)** Dose response curves of sirolimus (mTORi, left) and vorinostat (HDACi, right) on MRT (gray) and normal kidney (green) organoid lines. Each dot and error bar represent the mean and SD of two independent experiments (each independent experiment is an average of 4 technical replicates). Source data are provided as a Source Data file. **(C)** Graph shows the cell viability of MRT (gray) and normal kidney (green) organoids after drug treatment (T1) relative to vehicle control (DMSO). Organoids were treated with either sirolimus (2nM), vorinostat (1μM or 3μM) or the combination. Each dot represents an independent experiment (n = 60T/103T: 3, 78T mTOR/HDAC1μM/Combi2nM1μM: 6, 78T HDAC3μM/Combi2nM3μM: 4, normal kidney: 7), which is an average of 4 technical replicates. Source data are provided as a Source Data file. The mean is indicated as a horizontal line for each organoid line. Average cell viability of MRT and normal kidney organoids were compared. P-values were calculated using an unpaired Student's t test (two-tailed): test with donor 1 * <0.05, ** < 0.01, *** <0.001, test with donor 2 # <0.05, ## < 0.01, ### <0.001 (exact p-values are indicated in Supplementary Table 4). **(D)** Graph shows the cell viability of MRT and normal kidney organoids after drug washout (T2) normalized to timepoint 1 (T1) DMSO controls. Each dot represents an independent experiment (n = 60T/103T: 3, 78T: 4, normal kidney: 7), which is an average of 4 technical replicates. Source data are provided as a Source Data file. The mean is indicated as a horizontal line for each organoid line. Regrowth capability was assessed by comparison of average cell viability of MRT and normal kidney organoids. P-values were calculated using an unpaired Student's t test (two-tailed): test with donor 1 * <0.05, ** < 0.01, *** <0.001, test with donor 2 # <0.05, ## < 0.01, ### <0.001 (p-value: Donor1 vs 60T=0.0007, vs 78T=0.0002, vs 103T=0.0006; Donor2 vs 60T=2.9e-6, vs 78T=1.3e-6, vs 103T=2.4e-6).

Supplementary Table 1 | Overview of WGS data.

Case	Sample	Sample type	FFPE	WGS Coverage	Platform
PD42923	PD42923a	Tumour	N	43.68	X10
PD42923	PD42923b	Blood	N	31.46	X10
PD42923	PD42923c	Kidney	N	31.74	X10
PD42923	PD42923d	Ganglion	Y	25.47	X10
PD42923	PD42923e	Nerve	Y	26.19	X10
PD42923	PD42923f	Tumour	Y	22.22	X10
PD42923	PD42923g	Kidney	Y	25.61	X10
PD46555	PD46555a	Tumour	N	36.98	Novaseq
PD46555	PD46555d	Dorsal nerve roots	N	31.34	Novaseq
PD46555	PD46555e	Ventral nerve roots	N	36.65	Novaseq
PD46555	PD46555f	Spinal cord	N	35.86	Novaseq
PD46555	PD46555g	Upper cervical spine	N	27.26	Novaseq
PD46555	PD46555h	Spinal cord	N	29.7	Novaseq
PD46555	PD46555i	Posterior dura	N	40.78	Novaseq
PD46555	PD46555j	Fat	N	28.48	Novaseq
PD46555	PD46555k	Fat	N	34.98	Novaseq
PD46555	PD46555l	Muscle	N	34.45	X10
PD46555	PD46555m	Muscle	N	29.79	Novaseq
PD46555	PD46555r	Skin	N	39.51	Novaseq
PD46555	PD46555s	Skin	N	30.62	Novaseq
PD46555	PD46555w	Blood	N	30.77	Novaseq

Supplementary Table 2 | MRT organoid features.

Patient line	Diagnosis	Source
60T	Malignant Rhabdoid Tumor of the Kidney	Primary Tumor
78T	Malignant Rhabdoid Tumor of the Kidney	Primary Tumor
103T	Malignant Rhabdoid Tumor of the Kidney	Primary Tumor

Supplementary Table 3 | Sample processing info scRNA-seq.

Sample name	Original ID	Patient line	Details	Gender	10X kit	average reads/cell
MRT_60T_CTRL	5640STDY7891142	60T	MRT Control	male	3' v2	35970
MRT_60T_SMB	5640STDY7891143	60T	MRT *SMARCB1* +	male	3' v2	490249
MRT_78and103_CTRL	snRNA10x-Lars-CTRL-17579	78T and 103T mixed	MRT Control	78T=male,103T=female	3' v3	34798
MRT_78and103_SMB	snRNA10x-Lars-SMB1-17580	78T and 103T mixed	MRT *SMARCB1* +	78T=male,103T=female	3' v3	44561

Sample name	sequencing platform	Genome reference	cellranger version	%mito	SMARCB1 filter	gene cutoff
MRT_60T_CTRL	Illumina HiSeq 4000	GRCh38 1.2.0	2.0.2	10	SMARCB1 > 0	none
MRT_60T_SMB	Illumina HiSeq 4000	GRCh38 1.2.0	2.0.2	10	SMARCB1 == 0	none
MRT_78and103_CTRL	NextSeq 500	GRCh38 1.2.0	3.0.3	20	SMARCB1 > 0	1500
MRT_78and103_SMB	NextSeq 500	GRCh38 1.2.0	3.0.3	20	SMARCB1 == 0	1500

Sample name	EGA ID
MRT_60T_CTRL	EGAN00002133134
MRT_60T_SMB	EGAN00002133135
MRT_78and103_CTRL	EGAN00002580391, EGAN00002580392, EGAN00002580393, EGAN00002580394
MRT_78and103_SMB	EGAN00002580397, EGAN00002580398

Supplementary Table 4 | Exact p-values.

Supplementary Fig. 3g	**SMARCB1 vs Control**		
Cell Type	**60T**	**78T**	**103T**
Sensory	4.58133E-28	3.14728E-22	1.75E-08
Migratory.1	0.2275684	3.315E-15	0.08191271
Migratory.2	5.89544E-67	0.00003463	0
Mesenchyme	0.0002235	4.20523E-12	0.0006274
Autonomic	0.02985853	7.49681E-93	1.3824E-57
Supplementary Fig. 3h	**SMARCB1 vs Control**		
Cell Type	**60T**	**78T**	**103T**
Early mesenchyme	1.48789E-35	2.00176E-18	4.454E-173
Connective tissue progenitors	1.59887E-80	2.41191E-68	2.6248E-127
Intermediate mesoderm	0.01167182	6.39985E-07	3.7136E-12
Schwann cell precursors	5.15326E-07	0.3710319	0.4429936
Sensory neurons	0.000473307	1.88498E-23	6.4646E-141
Neural tube	0.007759082	1.6638E-27	7.40169E-21
Neural progenitor cells	7.6889E-15	1.73252E-35	5.59331E-73
Radial glia	1.20211E-16	1.41149E-24	6.9577E-111
Supplementary Fig. 8c	**Comparison vs normal kidney donor 1**		
Drug	**60T**	**78T**	**103T**
mTORi (2nM)	5.78324E-06	0.016856109	2.35666E-05
HDACi (1µM)	0.003957308	0.014547196	NA
Combi (2nM, 1µM)	0.000197062	0.000667108	NA
HDACi (3µM)	NA	5.37192E-06	5.74512E-05
Combi (2nM, 3µM)	NA	9.53401E-05	0.000535419
	Comparison vs normal kidney donor 2		
Drug	**60T**	**78T**	**103T**
mTORi (2nM)	0.001134997	0.927179112	0.008510203
HDACi (1µM)	0.000573236	0.025865327	NA
Combi (2nM, 1µM)	9.5596E-05	0.001887569	NA
HDACi (3µM)	NA	0.000168	0.001769478
Combi (2nM, 3µM)	NA	0.000388825	0.001892323

CHAPTER 5

SMARCB1 loss drives patient-specific super-enhancer formation in malignant rhabdoid tumors

Lars Custers[1,2†], Ning Qing Liu[2,3†], Irene Paassen[1,2], Hans Teunissen[2,3], Elzo de Wit[2,3*], Jarno Drost[1,2*]

Manuscript in preparation

[1]Princess Máxima Center for Pediatric Oncology, 3584CS, Utrecht, the Netherlands

[2]Oncode Institute, the Netherlands

[3]Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, the Netherlands

† These authors contributed equally to this work

* These authors jointly supervised this work

Abstract

Derailed epigenetic regulation of gene expression is emerging as a prominent driver of many pediatric cancer types. This is particularly evident for pediatric malignancies uniquely defined by mutations in regulators of the epigenome, including malignant rhabdoid tumors (MRTs). MRTs harbor a single recurrent genetic driver event, the bi-allelic inactivation of *SMARCB1*, a subunit of the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodeler complex. Aberrant chromatin remodeling is therefore thought to drive MRT. To gain insight into the gene regulatory mechanisms that underly MRT development, we reconstituted *SMARCB1* in patient-derived MRT organoids. By interrogating chromatin dynamics, we found SMARCB1 to be essential for the regulation of enhancer activity. Integration of gene expression profiles, chromatin dynamics, and 3D genome architecture revealed that while the SMARCB1-dependent transcriptional programs of different MRTs are similar, robust differences in enhancer landscape can be found. Our study shows that loss of *SMARCB1* drives oncogene expression by affecting enhancer utilization in a patient-specific manner.

Introduction

Comprehensive analysis of childhood cancer genomes revealed that, in comparison to the vast majority of adult cancers, the mutational burden of pediatric tumors is relatively low[1,2]. Many pediatric malignancies harbor a single or a few genetic driver events[1,2], which frequently affect genes encoding for epigenetic modifiers[1,2]. Accordingly, derailed epigenetic gene regulation is emerging as a prominent driver of many pediatric cancer types. This is particularly evident for malignant rhabdoid tumors (MRTs), which are characterized by a single recurrent genetic driver event, the bi-allelic loss of *SMARCB1*, or in rare cases *SMARCA4*[3-7]. Both are subunits of the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodeler complex[8]. Aberrant chromatin remodeling is therefore thought to drive MRT development.

The genomic simplicity that features MRT does not correlate with tumor malignancy. MRTs are aggressive and often lethal cancers for which effective treatment options are lacking up to date[9,10]. They can arise inside the central nervous system (CNS), in which they are referred to as Atypical Teratoid/Rhabdoid Tumors (AT/RTs)[10] but are also diagnosed in the kidney or soft tissues, which are collectively termed as extracranial rhabdoid tumors (ECRTs)[9]. Furthermore, AT/RT subgroups can be defined based on DNA methylation patterns, commonly referred to as MYC, sonic hedgehog (SHH) and tyrosinase (TYR)[6]. The ECRTs are epigenetically and transcriptionally similar to the AT/RT-MYC subgroup, which illustrates the close molecular relationship of extracranial and CNS MRT[3]. MRT predominantly strikes in early childhood but already initiates during prenatal life[11] and is therefore considered to be a product of aberrant

embryonic development. In line with this, we recently provided evidence that ECRTs originate from the developing neural crest lineage[12].

Loss of *SMARCB1* during embryonic development is indicated to block cellular maturation, thereby retaining cells in a malignant embryonic and proliferative state[13-15]. Nevertheless, the gene regulatory mechanisms affected by *SMARCB1* loss that contribute to MRT development remain largely unknown. Investigations into SMARCB1-dependent chromatin dynamics may uncover the oncogenic signaling pathways downstream of *SMARCB1* loss, that may serve as potential therapeutic targets in MRT. Recent work demonstrated that SMARCB1 is essential for the activation of typical enhancers and suppression of super-enhancers in MRT cell-lines[16]. Aberrant enhancer regulation is therefore suggested to underly MRT development.

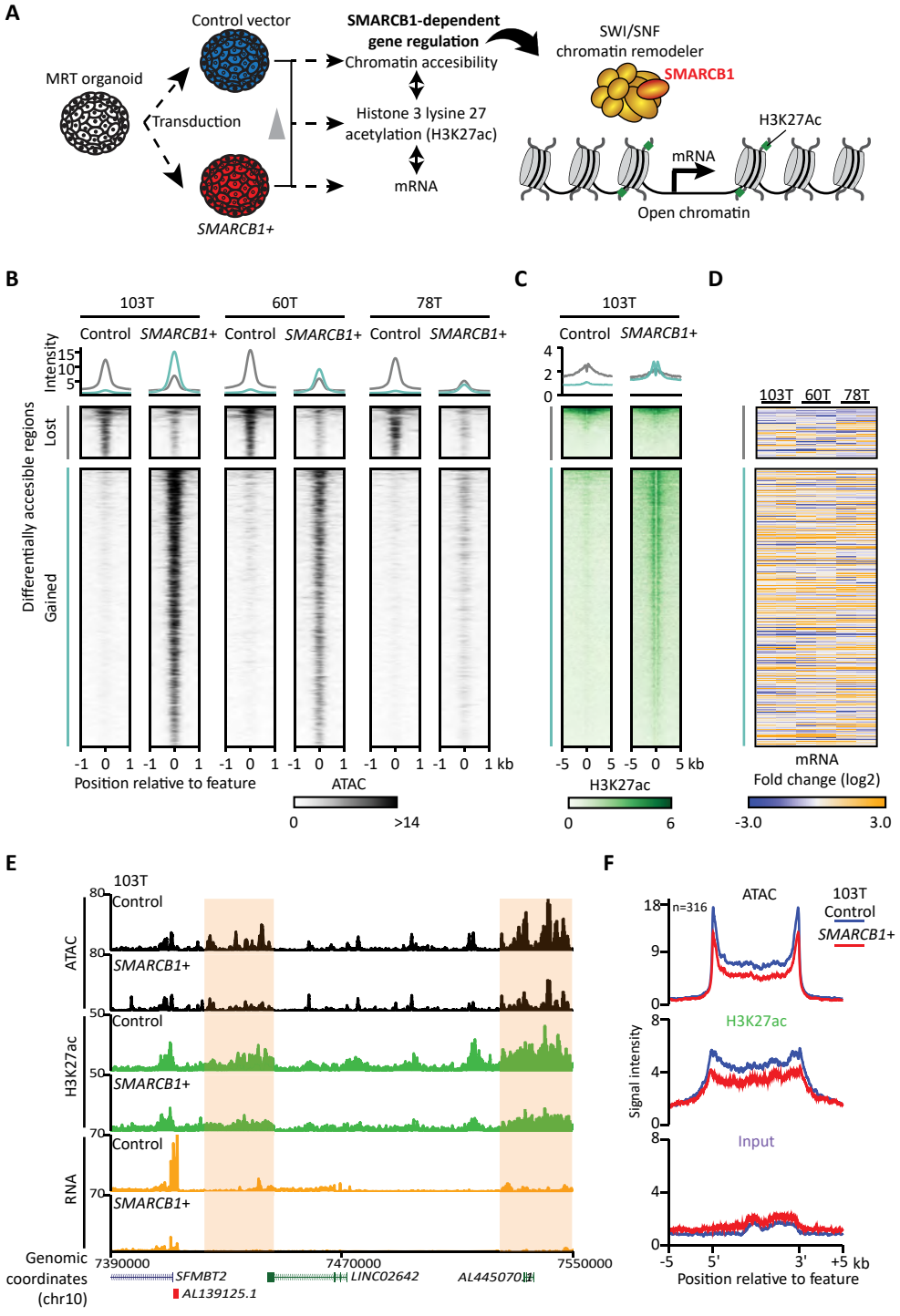
Here, we sought out to investigate the SMARCB1-dependent enhancer landscapes across MRT patients. Recently, renal MRT organoid models were established that closely resemble the tumor tissue they were derived from on a morphological, transcriptomic, and (epi)genetic level[17], and therefore represent suitable models for studying chromatin regulation in a patient-specific manner. Accordingly, we interrogated SMARCB1-dependent chromatin dynamics by reconstituting *SMARCB1* in MRT organoids[12], which indicated SMARCB1 to be essential for enhancer regulation, in line with previous findings[16]. Integration of gene expression profiles, chromatin dynamics, and 3D genome architecture revealed that while the SMARCB1-dependent transcriptional programs of different MRTs are similar, robust differences in enhancer landscape can be found. We identified patient-specific super-enhancers that, in absence of SMARCB1, were indicated to drive expression of the *MYC* oncogene, which is required for MRT proliferation.

Results

Analysis of chromatin dynamics in MRT reveals SMARCB1-dependent enhancer regulation

In this study, we used three different patient-derived MRT organoid models (60T, 78T and 103T), which were previously established and characterized in depth[17]. To study the effects of SMARCB1 on gene regulation, we lentivirally transduced either a control or *SMARCB1* expression (*SMARCB1*+) construct in MRT organoids (**Fig. 1a**). Reconstitution of *SMARCB1* induces a cell cycle arrest and differentiation phenotype that mirrors neural crest differentiation to mesenchyme, which we previously described in detail[12].

The cellular process of lineage specification is primarily governed by genomic regulatory regions that control the expression of differentiation genes[18]. In the eukaryotic genome, cis-regulatory regions are featured by open chromatin, which is modulated by chromatin remodelers such as the SWI/SNF complex (**Fig. 1a**)[8]. To interrogate SMARCB1 function in MRT differentiation, we measured chromatin accessibility


A
Control vector
MRT organoid
Transduction
SMARCB1+
SMARCB1-dependent
gene regulation
Chromatin accesibility
Histone 3 lysine 27
acetylation (H3K27ac)
mRNA
SWI/SNF
chromatin remodeler
SMARCB1
H3K27Ac
mRNA
Open chromatin
B
103T
60T
78T
Control
SMARCB1+
Intensity
Differentially accesible regions
Lost
Gained
Position relative to feature
ATAC
C
H3K27ac
D
mRNA
Fold change (log2)
E
ATAC
H3K27ac
RNA
Genomic
coordinates
(chr10)
SFMBT2
LINC02642
AL445070
AL139125.1
F
n=316
Signal intensity
Input
Position relative to feature

in MRT control and *SMARCB1*+ organoids using ATAC-seq[19] (**Fig. 1a**). By principal component analysis, we found that most variance could be explained by *SMARCB1* reconstitution, which was consistent across the three different MRT patient models (**Supplementary Fig. 1a**). Further, we identified differentially accessible regions (DARs) by comparing MRT *SMARCB1*+ to control organoids (**Supplementary Fig. 1b**). By examination of the DARs shared between the three MRT patient models, we found that *SMARCB1* reconstitution predominantly enhanced chromatin accessibility (**Fig. 1b, Supplementary Fig. 1b**). The gain of open chromatin associated with increased acetylation of histone 3 at lysine 27 (H3K27ac) (**Fig. 1c**), a marker of active genomic regulatory regions[20,21] (**Fig. 1a**). The open chromatin mostly localized to intergenic or intronic regions (**Supplementary Fig. 1c**). Accordingly, these findings indicate SMARCB1 to activate typical distal enhancers. To assess the effect of enhancer activation on transcription, we compared changes in chromatin accessibility with gene expression patterns in MRT control and *SMARCB1*+ organoids (**Supplementary Fig. 1a**) and linked the DARs to the nearest gene transcription start site (TSS). This revealed that SMARCB1-induced changes in open chromatin at typical enhancers significantly correlates with mRNA expression levels of nearby genes (**Fig. 1d, Supplementary Fig. 1d**).

Next, we assessed the genomic regions that lost chromatin accessibility upon *SMARCB1* reconstitution, for which only a minor subset was featured by H3K27ac in MRT control cells (**Fig. 1b, c**). The DARs that gained chromatin accessibility upon SMARCB1 reconstitution showed sharp peaks of H3K27ac that characterizes typical enhancers (**Fig. 1b, c**). In contrast, the open chromatin specific to MRT control cells exhibited broad H3K27ac peaks (**Fig. 1b, c**). These H3K27ac patterns feature super-

Figure 1 | Analysis of chromatin dynamics in MRT reveals SMARCB1-dependent enhancer regulation.
(A) Visualization of experimental setup and genomic readouts (left) and schematic representation of SMARCB1-dependent gene regulation (right). **(B)** Heatmaps display chromatin accessibility signals (ATAC-seq) for MRT control and *SMARCB1*+ organoids for three patient models (103T, 60T and 78T). The signal intensity represents an average of three independent biological replicates. Heatmaps are subset for genomic sites (rows) that significantly gained (green) or lost (gray) chromatin accessibility upon *SMARCB1* reconstitution in all three patient models. The rows represent 2kb regions centered on chromatin accessibility peaks and ranked by signal intensity. Average signal intensity for regions that lost (gray) or gained (green) chromatin accessibility is indicated (top). **(C)** Heatmaps show H3K27ac signals (standard ChIP-seq) for 103T control and *SMARCB1*+ MRT organoids. The rows represent 10kb genomic regions centered on H3K27ac peaks and are aligned to B. Average signal intensity is indicated (top). **(D)** Heatmap represent mRNA fold changes (RNA-seq) induced by *SMARCB1* reconstitution in MRT organoids (103T, 60T and 78T). Two independent biological replicates are depicted. The genes that are genomically closest to the differentially accessible regions indicated in B are included in the heatmap. **(E)** Representative screenshot of a putative super-enhancer upstream of the *SFMBT2* TSS. Chromatin accessibility, H3K27ac and RNA signal intensities are indicated for 103T control and *SMARCB1*+ MRT organoids. The open chromatin cluster is highlighted (yellow). **(F)** Graphs represent the average signal intensity of chromatin accessibility (top) and H3K27ac (standard ChIP-seq, middle) and input chromatin (bottom) for genomic regions defined as super-enhancers in 103T control MRT cells. Average signal intensities are indicated for 103T control (blue) and *SMARCB1*+ (red) MRT cells. The genomic positions are relative to the 5' and 3' end of the open chromatin cluster.

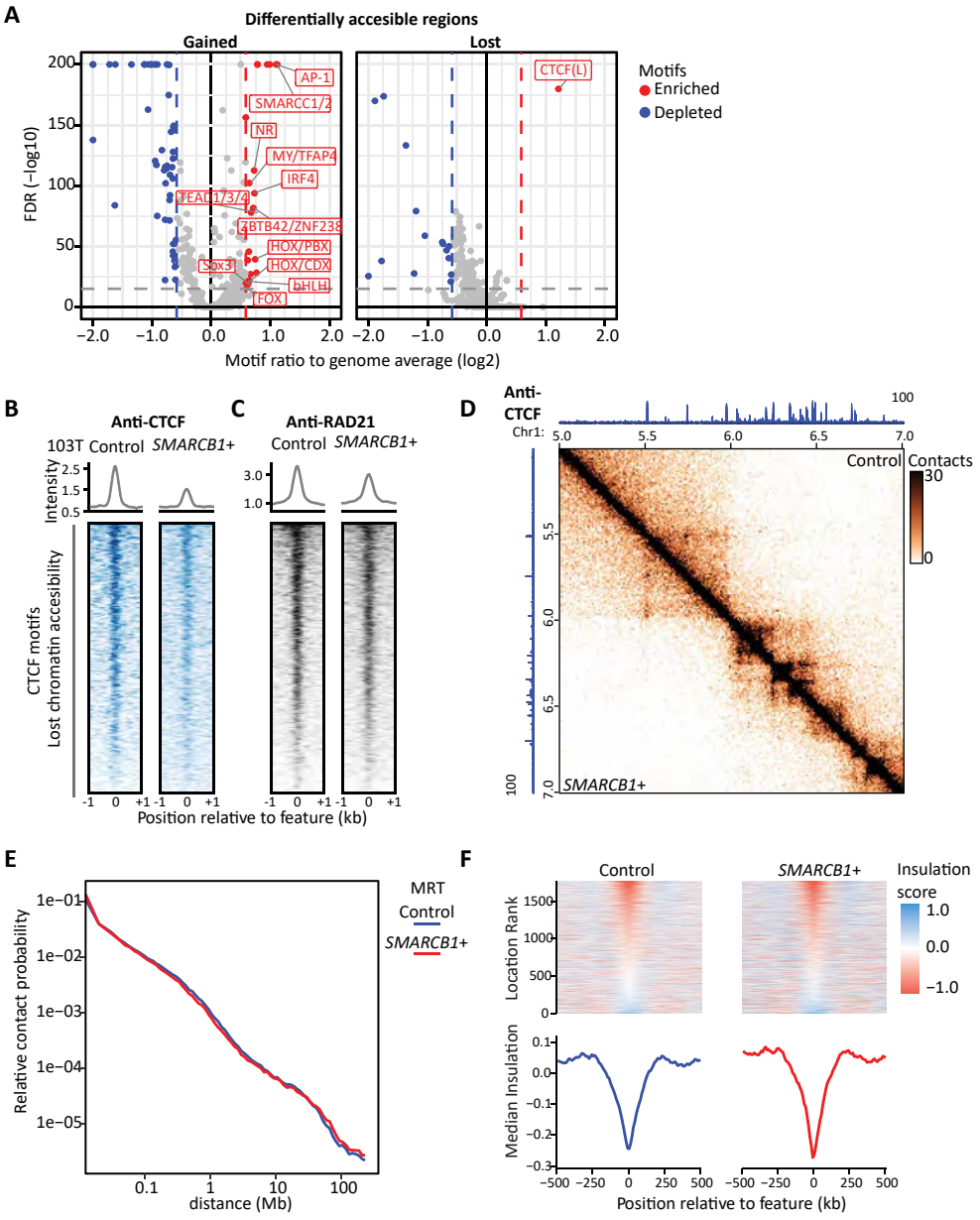


Figure 2 | SMARCB1-dependent CTCF binding does not correlate with 3D genome changes.
(A) Scatter plots represent frequencies of transcription factor binding motifs (TFBM) and corresponding false discovery rates (FDR) calculated using a Fisher's Exact test. Motifs were quantified in genomic regions that gained (left) or lost (right) chromatin accessibility upon *SMARCB1* re-expression (Fig. 1b). Motif frequencies were normalized to the genome average. Significantly enriched motifs are labeled. **(B)** Heatmaps show CTCF and RAD21 **(C)** binding (calibrated ChIP-seq) for 103T control and *SMARCB1+* MRT organoids. The rows represent 2kb genomic regions centered on signal peaks. Rows are subset for genomic sites containing CTCF motifs that lost chromatin accessibility in MRT *SMARCB1+* cells. Average signal intensity

enhancers, which are genomic clusters of multiple enhancers in close proximity[22]. We found super-enhancer activity to be SMARCB1-dependent, as chromatin accessibility and H3K27ac levels decreased upon *SMARCB1* re-expression (**Fig. 1e**). Super-enhancer suppression by SMARCB1 correlated with downregulation of associated genes (**Fig. 1d, Supplementary Fig. 1d**).

Altogether, our data indicate that SMARCB1 suppresses super-enhancers, while activating typical enhancers in MRT cells. These findings are in line with previous studies using MRT cell-lines, demonstrating enhancer regulation to rely on SMARCB1-dependent SWI/SNF complex function[16].

Open chromatin in *SMARCB1*-null MRT cells is enriched for CTCF motifs

Open chromatin allows for binding of transcription factors through DNA-specific interactions and subsequent modulation of transcription[23]. Accordingly, to study the gene regulatory circuits downstream of *SMARCB1* re-expression, we searched for transcription factor binding motifs in the DARs identified by ATAC-seq (**Fig. 1b**). We compared the frequency of each motif to the genome average to determine motif enrichment or depletion, which is suggestive of transcription factor recruitment. As expected[16,24], we identified the SMARCC1 and AP-1 motif to be highly enriched in the genomic regions that became more accessible upon *SMARCB1* reconstitution (**Fig. 2a, Supplementary Fig. 2a, b**). Other enriched motifs included HOX transcription factors, which have previously been implied in ATRT-MYC and ECRT subgroups[3,6], and motifs of transcription factors driving mesenchymal differentiation, including FOXC1[25], TWIST1[25], MYOD1[26] (**Fig. 2a, Supplementary Fig. 2a**). Interestingly, we found a single motif with significantly reduced chromatin accessibility upon *SMARCB1* re-expression, namely the CTCF motif (**Fig. 2a, Supplementary Fig. 2a, b**). Our analysis of binding motifs identified putative transcription factors with potential gene regulatory function downstream of *SMARCB1* reconstitution, which is a starting point for further investigation.

SMARCB1-dependent CTCF binding does not correlate with 3D genome changes

CTCF has been implicated in modulation of chromatin structure[27]. As the CTCF motif is depleted upon *SMARCB1* re-expression, we hypothesized that aberrant chromatin architecture may contribute to MRT development. CTCF is known to be involved in the formation of chromatin loops that bring gene promoters and enhancers in proximity, which promotes transcription[27]. Accordingly, aberrant binding of CTCF in *SMARCB1*-

is indicated (top). **(D)** Representative screenshot of Hi-C map showing chromatin interaction frequencies and corresponding CTCF binding for 103T control and *SMARCB1*+ MRT cells. **(E)** Graph represents the relative contact probability (derived from Hi-C) for CTCF motifs that lost chromatin accessibility upon *SMARCB1* reconstitution. Chromatin contacts are binned for genomic distance. Contact probabilities are indicated for 103T control (blue) and *SMARCB1*+ (red) MRT organoids. **(F)** Graphs show insulation scores (derived from Hi-C) of CTCF motifs that lost chromatin accessibility upon *SMARCB1* reconstitution for 103T control and *SMARCB1*+ MRT cells. The rows (top) represent 1Mb genomic regions centered on the CTCF motifs and ranked by insulation score. Median insulation score is indicated (bottom).

null MRT cells may deregulate the expression of genes involved in tumorigenesis. We therefore set out to assess SMARCB1-dependent changes in chromatin contacts.

To validate SMARCB1-dependent binding of CTCF to the DNA, we performed ChIP-seq for CTCF in 103T control and *SMARCB1*+ MRT organoids. Indeed, we observed that the decreased accessibility to CTCF motifs (**Fig. 2a**) corresponds to loss of CTCF binding (**Fig. 2b**). Of note, the subset of open chromatin that contained CTCF motifs and binding sites in MRT control cells did not exhibit H3K27ac signal (**Supplementary Fig. 2c**), and therefore appear to be mostly non-overlapping with the super-enhancers identified in MRT cells (**Fig. 1c**).

The architectural role of CTCF is closely associated with the cohesin protein complex[27]. Cohesin actively extrudes DNA strands through its ring-shaped structure until it is blocked by convergent CTCF binding sites[27,28]. Therefore, we assessed cohesin binding sites in 103T control and *SMARCB1*+ MRT cells. We performed ChIP-seq for RAD21, a core subunit of the cohesin complex, and found that, in contrast to CTCF, cohesin binding was not significantly altered upon *SMARCB1* re-expression (**Fig. 2c**), which may suggest that chromatin looping is not affected. Next, we investigated the effect of *SMARCB1* reconstitution on the 3D genome. To do so, we subjected 103T control and *SMARCB1*+ MRT organoids to Hi-C. Analysis of chromatin contacts of CTCF-binding sites showed that *SMARCB1* re-expression does not induce significant genome-wide changes of CTCF-anchored chromatin interactions (**Fig. 2d, e**), in line with the unchanged cohesin binding. Furthermore, CTCF can function as a chromatin insulator, mediating the formation of the boundaries of topologically associated domains (TADs)[29]. The strength of these boundaries can be quantified, which is referred to as an insulation score[30]. Genome-wide assessment of insulation scores for CTCF binding sites showed that chromatin insulation is not affected upon loss of CTCF binding (**Fig. 2f**), indicating that the strength of TAD boundaries is unchanged. Thus, our findings suggest that the SMARCB1-induced loss of CTCF binding does not cause genome-wide changes in the 3D folding of the genome.

Patient-specific super-enhancers form SMARCB1-dependent chromatin loops

Super-enhancers have been indicated to drive the expression of genes promoting proliferation in MRT cell-lines[16]. Our findings so far support this hypothesis, as we demonstrated that SMARCB1 is essential to suppress super-enhancer activity (**Fig. 1f**). Distal enhancers regulate transcription by forming chromatin loops to gene promoters[31]. To define these chromatin loops, we evaluated the Hi-C interaction maps in 103T control and *SMARCB1*+ MRT organoids. Analysis of chromatin interactions revealed that the insulation score of super-enhancers was generally decreased upon *SMARCB1* reconstitution (**Fig. 3a**). This suggests that super-enhancers may form aberrant chromatin interactions as a consequence of *SMARCB1* loss in MRT.

Next, we searched for genomic regulatory regions exhibiting SMARCB1-dependent chromatin interactions. One of the regions showing a strong change in chromatin

architecture upon *SMARCB1* reconstitution involved the *MYC* locus (**Fig. 3b**). Using Hi-C, we found that the *MYC* gene formed a chromatin loop to a region ~1Mb downstream (which we referred to as E2) in 103T control MRT cells. This genomic site exhibited super-enhancer features, as it was characterized by a broad cluster of open chromatin and H3K27ac[32] (**Fig. 3d, Supplementary Fig. 3a**). Moreover, enhancer activity was reflected by the transcription of enhancer RNAs (eRNAs)[33] (**Fig. 3e**). The putative gene regulatory loop of E2 to *MYC* was diminished in MRT *SMARCB1*+ cells, as chromatin accessibility, contacts and H3K27ac were reduced (**Fig. 3b, d, Supplementary Fig. 3a**). In addition, both E2 eRNA and *MYC* mRNA levels were downregulated in 103T *SMARCB1*+ MRT cells (**Fig. 3e**). Altogether, these findings indicate the E2 super-enhancer to putatively regulate *MYC* expression in 103T in a SMARCB1-dependent manner.

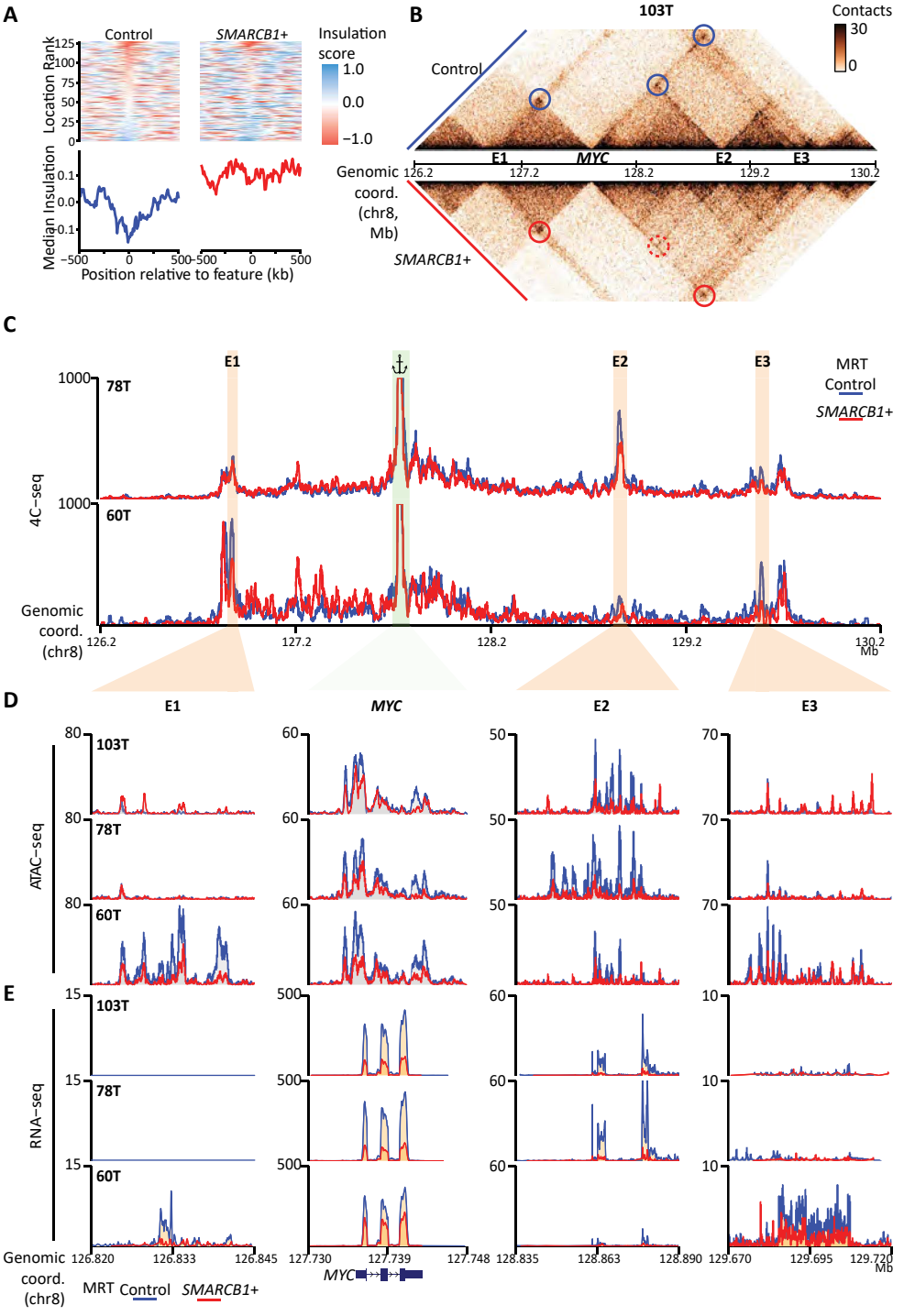
To generalize our findings, we examined chromatin interactions in two other patient-derived MRT organoid cultures (60T and 78T) by deploying high-resolution 4C, with the *MYC* promoter as the viewpoint. As in 103T, we found the E2 super-enhancer looping to the *MYC* promoter in 78T, which was diminished upon *SMARCB1* re-expression (**Fig. 3c**). Interestingly, in 60T, E2 enhancer activity was relatively low compared to 78T and 103T. Moreover, the open chromatin and H3K27ac was not affected by SMARCB1 re-expression (**Fig. 3d, e**). Instead, we identified SMARCB1-dependent long-range chromatin contacts of the *MYC* promoter with two alternative super-enhancers (termed E1 and E3) (**Fig. 3c**). These findings indicate differential use of *MYC* super-enhancers across MRT patient models. Nevertheless, the SMARCB1 dependency of these super-enhancers was uniform, as chromatin contacts, accessibility, enhancer RNAs and *MYC* mRNA levels were consistently reduced upon *SMARCB1* reconstitution (**Fig. 3b-e**).

Thus, our findings indicate SMARCB1-dependent super-enhancers to regulate levels of the *MYC* oncogene through long-range chromatin loops in a patient-specific manner.

Inhibition of BRD9 enforces MYC downregulation and MRT differentiation

Recent investigations indicated that MRT may be driven by aberrant function of a residual non-canonical SWI/SNF complex, which was found to localize at super-enhancers in MRT cells[16,34,35]. Accordingly, we set out to investigate the effect of residual SWI/SNF inhibition on *MYC* super-enhancer function in MRT cells (**Fig. 4a**). To do so, we treated MRT organoids with a pharmacological inhibitor of BRD9, a subunit unique to the non-canonical SWI/SNF complex[34,35]. We observed that, morphologically, MRT cells exhibited a differentiation phenotype alike *SMARCB1* re-expression[12] (**Fig. 4b**). Quantitative RT-PCR (RT-qPCR) further demonstrated that BRD9 inhibition caused a reduction of *MYC* mRNA levels (**Fig. 4c**), suggesting that the non-canonical SWI/SNF complex is essential to maintain *MYC* transcription.

To investigate if MYC depletion is sufficient to induce MRT differentiation, we used RNA interference to knockdown MYC in MRT organoids. We lentivirally transduced


A
Control
SMARCB1+
Location Rank
Median Insulation
Position relative to feature (kb)
Insulation score
B
103T
Control
SMARCB1+
Contacts
E1
MYC
E2
E3
Genomic coord. (chr8, Mb)
C
78T
60T
4C−seq
MRT
Control
SMARCB1+
Genomic coord. (chr8)
D
ATAC−seq
E
RNA−seq
MRT Control SMARCB1+

MRT organoids with an expression construct containing a shRNA sequence targeting *MYC* mRNA or a scrambled control. We confirmed knockdown of *MYC* mRNA levels using RT-qPCR for two different shRNA sequences (**Supplementary Fig. 3b**). Cell proliferation measurements revealed that proliferation of MRT organoids is significantly decreased upon MYC knockdown (**Fig. 4d**) without showing an apparent morphological change (**Fig. 4e**). These results suggest that increased *MYC* expression caused by *SMARCB1* loss is at least partially involved in the malignant growth of MRT.

Super-enhancer activity in MRT is reflected in embryonic cells

Our *SMARCB1* reconstitution experiments outlined the differential usage of *MYC* super-enhancers in different patient-derived MRT organoid cultures. To confirm our findings in primary MRT tissues, we examined H3K27ac profiles at the *MYC* locus for six patient-derived ECRT tissues[16]. In line with our findings in MRT organoids, we identified H3K27ac peaks at the E1, E2 and E3 locus, which was heterogeneously distributed across patient samples in different combinations (**Fig. 5a**). Despite the small sample size, these results strongly suggest that distinct enhancer landscapes drive *MYC* expression in MRT patients.

We have previously demonstrated that MRT arises from the neural crest lineage[12]. Furthermore, we and others have shown that tumors inherit gene expression patterns from the cell they derive from[36-40]. Following that rationale, we interrogated single-cell transcriptomes of murine neural crest development for *Myc* expression patterns[25]. We found that *Myc* mRNA levels were particularly high in neural crest cells that were on a differentiation trajectory to mesenchyme (**Supplementary Fig. 4a**). This observation is in line with our previous report, which demonstrated that MRT arises along a neural crest to mesenchyme differentiation path[12,36].

In that perspective, we speculated that the E1-3 super-enhancers that we identified in MRT (**Fig. 3d, e, Fig. 5a**) may similarly be acquired from the MRT cell-of-origin, the neural crest. Therefore, we investigated H3K27ac patterns of the *MYC* genomic locus in various cultured embryonic cell types. In line with the origin of MRT, only neural crest cells and astrocytes exhibited H3K27ac signal at all three MRT super-enhancer

Figure 3 | Patient-specific super-enhancers form SMARCB1-dependent chromatin loops.
(A) Graphs display insulation scores (derived from Hi-C) of SMARCB1-dependent super-enhancers (Fig. 1f) for 103T control and *SMARCB1*+ MRT cells. The rows represent 1Mb genomic regions centered on the super-enhancers and ranked by insulation score. Median insulation score is indicated (bottom). **(B)** Hi-C map showing chromatin interaction frequencies of the *MYC* locus for 103T control (top) and *SMARCB1*+ (bottom) MRT cells. Chromatin loops are indicated (circles). Dashed circle represents a chromatin loop that is diminished in 103T *SMARCB1*+ MRT cells. The labels E1-3 indicate super-enhancers identified in MRT control cells. **(C)** Line graphs represent chromatin interactions of the *MYC* promoter (viewpoint of 4C-seq) for 60T and 78T control (blue) and *SMARCB1*+ (red) organoids. The labels E1-3 indicate super-enhancers identified in MRT control cells. **(D)** Line graphs show chromatin accessibility signals (ATAC-seq) for genomic regions of *MYC* and distal super-enhancers (E1-3), which is indicated for MRT control (blue) and *SMARCB1*+ (red) cells. **(E)** Line graphs display mRNA levels for *MYC* and enhancer RNA levels for distal super-enhancers (E1-3), which is indicated for MRT control (blue) and *SMARCB1*+ (red) cells.

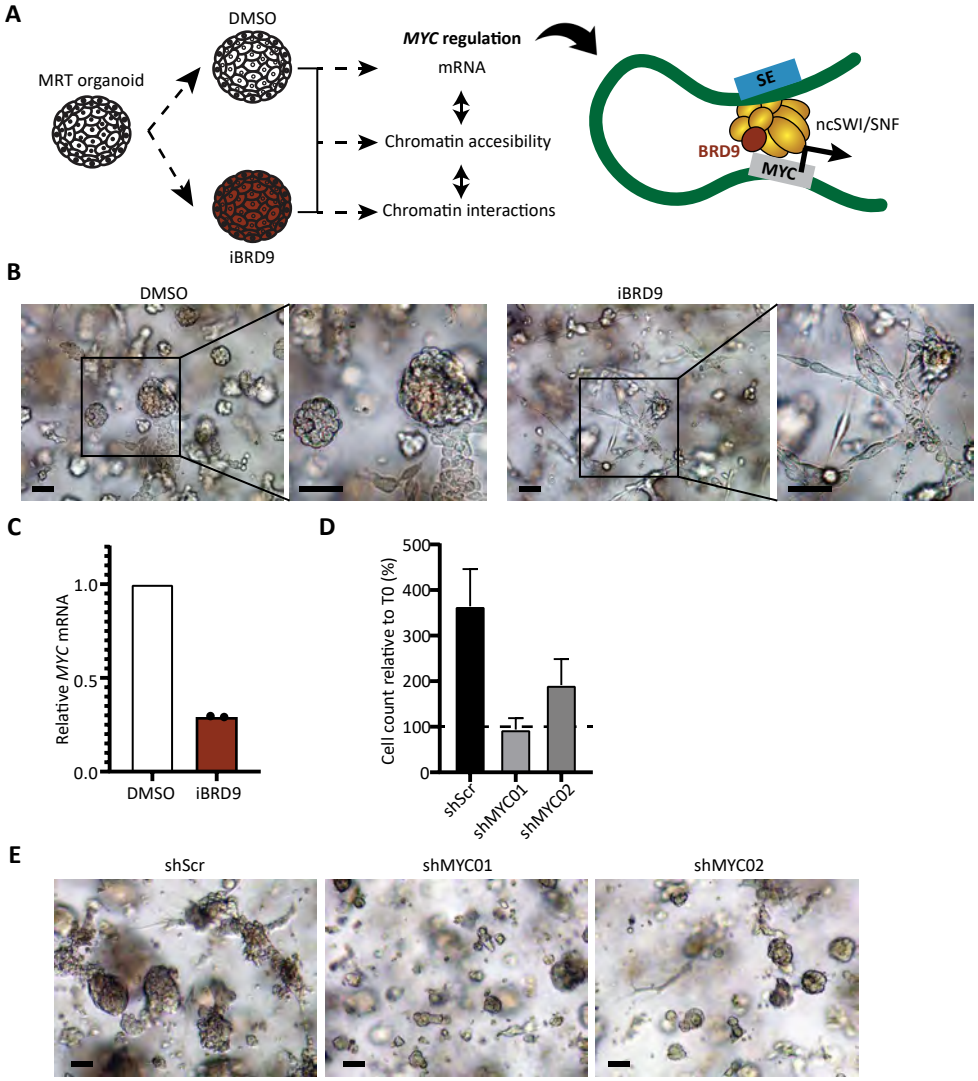


Figure 4 | Inhibition of BRD9 enforces MYC downregulation and MRT differentiation.
(A) Visualization of experimental setup and genomic readouts (left) and schematic representation of putative regulation of *MYC* transcription by the non-canonical SWI/SNF complex (right). **(B)** Representative bright-field images of 103T MRT organoids treated with DMSO (control) or a BRD9 inhibitor (iBRD9). Scale bars denote 100 µm. **(C)** Bar graph depicts mRNA expression levels (RT-qPCR) of *MYC* for 103T MRT organoid treated with DMSO (control) or BRD9 inhibitor (iBRD9). Expression levels are normalized to *GAPDH* mRNA and DMSO control. Bar represents an average of two independent experiments (dot). **(D)** Bar graphs represent growth rates of 103T MRT organoids expressing short hairpin RNAs targeting *MYC* mRNA (shMYC01, shMYC02) or a scrambled control (shScr). Values are relative to the cell number seeded at the starting point of the experiment (T0, 100%). Mean and SD (error bars) of three technical replicates are indicated. **(D)** Representative bright-field images of 103T MRT organoids expressing short hairpin RNAs targeting *MYC* mRNA (shMYC01, shMYC02) or a scrambled control (shScr). Scale bars denote 100 µm.

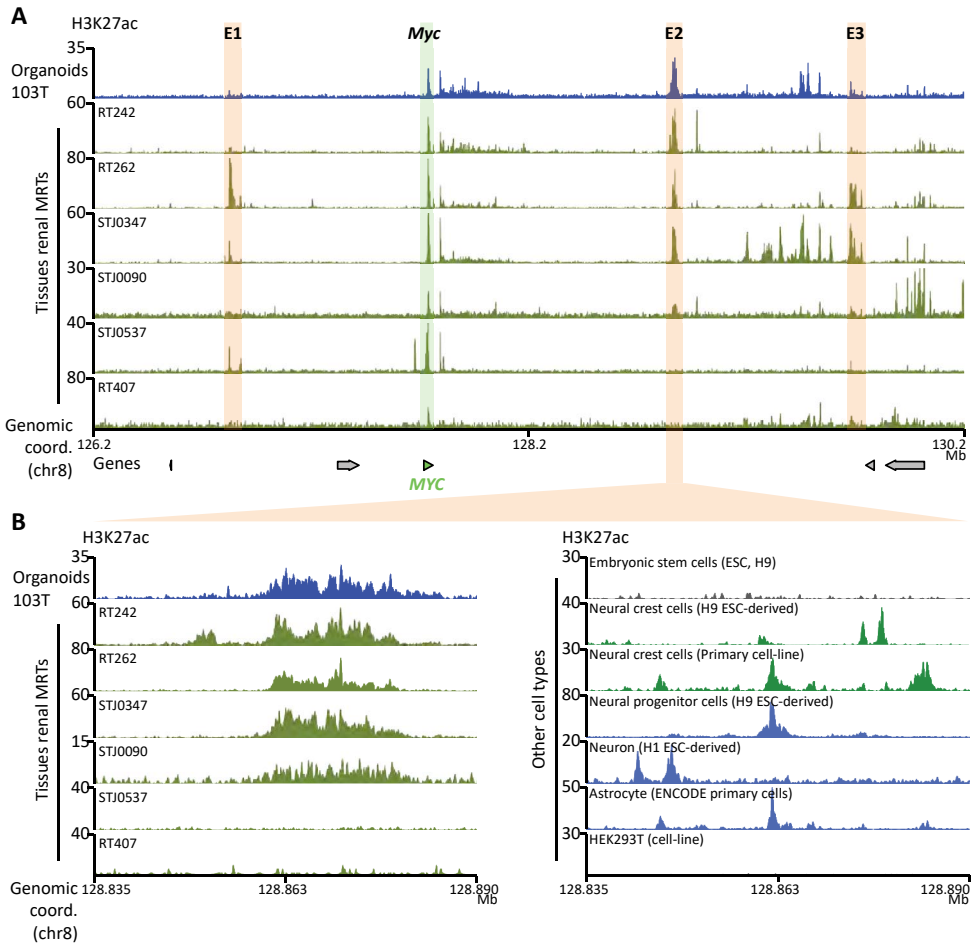


Figure 5 | Super-enhancer activity in MRT is reflected in embryonic cells.
(A) Line graphs show H3K27ac signals (ChIP-seq) of the *MYC* locus for 103T MRT organoids and six renal MRT tissues. Genomic regions of *MYC* (green) and super-enhancers identified in MRT organoids (E1-3, yellow) are highlighted. **(B)** Zoom in of the E2 genomic region showing H3K27ac for 103T MRT organoids and tissues (left) and cultured embryonic cell-types (right).

loci (**Supplementary Fig. 4b**). Remarkably, whereas no significant H3K27ac was detected at E1-3 in H9 ESCs, strong H3K27Ac signals could be detected at all three enhancers in neural crest cells derived from these ESCs. Furthermore, the E2 region exhibited H3K27ac peaks in all cell-types from the neural lineage but not in ESCs or HEK293T cells (**Fig. 5b, Supplementary Fig. 4b**). However, the H3K27ac patterns did not qualify as super-enhancers. Whereas a broad H3K27ac pattern at the E2 region was detected in MRT cells, typical (narrow) enhancer peaks were observed in non-malignant cells (**Fig. 5b**).

Altogether, our comprehensive profiling of chromatin dynamics in organoids and tissues revealed robust differences in the enhancer landscape between MRTs of different patients, which reflected enhancer utilization during neural crest differentiation. Our study thus provides evidence that derailed activity of embryonal enhancers caused by *SMARCB1* loss underpins rhabdoid tumorigenesis.

Discussion

Here, we studied gene regulation in MRT by interrogating SMARCB1-dependent chromatin dynamics and architecture. We defined the enhancer landscape underpinning the MRT differentiation block and revealed a long-range chromatin loop likely regulating expression levels of the *MYC* oncogene.

In contrast to its simple genome, MRTs exhibit an enhancer landscape that is complex and significantly distinct from *SMARCB1* wild-type normal tissues[16]. In that perspective, functional experiments are crucial to decipher the specific genomic regulatory regions that are impaired upon *SMARCB1* loss. Reconstitution of *SMARCB1* in MRT organoids allowed us to interrogate the epigenetic mechanisms that underly the maturation block. Accordingly, we found that SMARCB1-induced differentiation of MRT cells specifically associated with activation of typical enhancers and inhibition of super-enhancers, which supports a previous study by Wang *et al*. that assessed SWI/SNF function in MRT cell-lines[16]. They showed that SMARCB1 is essential for SWI/SNF complex stability. Absence of SMARCB1 abrogated the canonical SWI/SNF complex, which therefore failed to activate typical enhancers that control differentiation[16]. On the other hand, loss of *SMARCB1* was found to enhance the formation of a residual non-canonical SWI/SNF complex[16]. This non-canonical SWI/SNF complex localized to super-enhancers to regulate genes associated with proliferation.

We found that inhibition of the non-canonical SWI/SNF complex by treatment with a chemical inhibitor of BRD9 is sufficient to induce differentiation of MRT cells. We hypothesize that the non-canonical SWI/SNF complex is essential to maintain super-enhancer activity and the differentiation block in MRT. Further investigations are required to assess non-canonical SWI/SNF binding sites in MRT organoids with and without *SMARCB1* and determine if inhibition of BRD9 directly impairs super-enhancer activity and chromatin loops.

We found the CTCF motif to be enriched in the open chromatin of *SMARCB1*-null MRT cells. A role for CTCF in MRT development has been suggested previously by alternative approaches[13,34,41]. For instance, CTCF motifs associated with binding of the non-canonical SWI/SNF complex, which is enhanced in MRT compared to *SMARCB1*-wildtype cells[34]. However, even though the correlative relationship of SMARCB1 and CTCF has become apparent in multiple studies, a functional explanation is still lacking. We showed that CTCF loss does not associate with significant 3D genome changes. Recent reports demonstrated that genome-wide chromatin architecture is only affected when CTCF is near completely depleted[29]. This may relate to our

observations, as we found *SMARCB1* re-expression to merely cause a mild reduction of CTCF binding genome-wide, which did not correlate with decreased cohesin binding. Overall, further investigations are required to uncover the relationship of CTCF and SMARCB1, and to understand how CTCF binding may contribute to MRT development.

We illustrated that, even though *MYC* was consistently downregulated upon *SMARCB1* reconstitution, the specific super-enhancers regulating *MYC* were heterogeneous across MRT organoids and tissues. Of note, the sample size was relatively small and ChIP-seq quality of MRT tissues was not optimal, which calls for further validation. Previous experiments in MRT organoids demonstrated that intra-tumor heterogeneity exists on a transcriptomic level[12]. As we deployed bulk sequencing techniques, our results merely represent an average, and therefore we cannot exclude the occurrence of enhancer heterogeneity on a single-cell level. Recent developments in single-cell genomics allow for interrogation of chromatin dynamics with a single-cell resolution[42], which may provide further insights into intratumoral enhancer plasticity in MRT.

In this report, we focused on the chromatin dynamics controlling *MYC* transcription. However, additional perturbation experiments are required to validate that E1, E2 and E3 super-enhancers are essential to maintain *MYC* mRNA levels in a patient-specific manner. Furthermore, we found that MYC downregulation was not sufficient to induce differentiation. This suggests that there are other genes contributing to the differentiation block in MRT to which similar gene regulatory mechanisms may apply. Additional investigations are required to determine if enhancer plasticity is a direct consequence of *SMARCB1* loss and may therefore be a common phenomenon of the aberrant gene regulatory circuits that feature MRT.

In addition, our findings suggest that the E1-E3 super-enhancers likely regulating *MYC* transcription in MRT are not induced *de novo* upon *SMARCB1* loss, as neural crest cells mirror the MRT enhancer loci. However, the H3K27ac patterns in neural crest cells resemble those of typical enhancers, whereas MRTs exhibit super-enhancer features. In that perspective, we speculate that neural crest cells are epigenetically primed for *MYC* super-enhancer formation upon SMARCB1 loss and may therefore be particularly susceptible for malignant transformation.

Our interrogation of SMARCB1-dependent chromatin dynamics was key for the identification of *MYC* super-enhancers in ECRTs. Yet, MRT are epigenetically heterogeneous, where AT/RT-SHH and -TYR subgroups are characterized by distinct enhancer landscapes and do not exhibit the MYC gene expression signature[3,6]. We propose that a similar experimental approach applied to AT/RT-MYC -SHH and -TYR organoids may uncover subgroup-specific chromatin dynamics and provide handles for therapeutic interventions.

Methods

Organoid culture

MRT organoids were cultured as previously described[26]. For sub-culturing, MRT organoids were mechanically dissociated into single cells and seeded in growth factor-reduced BME (Trevigen). Organoid medium was composed of the following: Advanced DMEM/F12 (Gibco) containing 1x Glutamax, 10 mM HEPES, and antibiotics (AdDF+++), supplemented with 1.5% B27 supplement (Gibco), 10% R-spondin-conditioned medium, EGF (50 ng/ml, Peprotech), FGF-10 (100 ng/ml, Peprotech), N-acetylcysteine (1.25 mM, Sigma), Rho-kinase inhibitor Y-27632 (10 μM, Abmole), and A83-01 (5 μM, Tocris Bioscience). For *SMARCB1* re-expression, MRT organoids were transduced with pLKO.1-UbC-luciferase-blast or pLKO.1-UbC-hSMARCB1-blast lentiviruses, as described[12]. After two days, 10 μg/ml blasticidin was added to the culture medium. MRT cells were harvested (ATAC-seq, ChIP-seq, Hi-C, 4C-seq) four days after transduction.

For BRD9 inhibition experiments, MRT organoids were plated as single-cells and treated with 10μM I-BRD9 (MedChemExpress, HY-18975) one day after plating. Cells were harvested for RNA isolation five days after drug addition.

ATAC-seq

MRT organoids were washed in ice-cold AdDF+++ and viably frozen in Recovery Cell Culture Freezing Medium (ThermoFisher, 12648010). For library preparation, cells were thawed and processed following an established protocol[43]. In short, nuclei were isolated from cells and permeabilized. The isolated nuclei were tagmented using Tn5 transposase (produced in-house), followed by two sequential 9-cycle PCR amplification steps. The resulting DNA fragments (< 700 bp) were purified using SPRI beads (Beckman Coulter). ATAC-seq libraries were sequenced on a HiSeq 2500 (Illumina).

ATAC-seq data were analyzed as previously described[44]. In short, sequencing reads were mapped to the hg38 reference genome using BWA-MEM (version 0.7.15-r1140)[45]. The mapped reads were filtered using SAMtools, discarding reads with mapping quality score <15, as well as optical PCR duplicates. The coverage files were produced using the deepTools (version 3.0) method "normalize to 1X genome coverage". A merged peak list was generated from ATAC-seq data of MRT control and *SMARCB1*+ cells (n=3 independent experiments). The read coverage under the peaks was determined using a HTSeq tool (version 0.9.1). The peaks with >10 reads in each replicate were included for further analysis. The filtered coverage data was normalized using DESeq2 (version 1.18.1) with default parameters. Differential peaks were detected using a Wald test (FDR <0.05 and fold change >2). The alignment of ATAC peaks to genes was done by "a 'basal plus extension' method to link ATAC–seq peaks to their gene targets.

Identification open chromatin clusters

Super-enhancers were detected from ATAC-seq data using the Homer super-enhancer caller, based on an established method[32]. The mapped and filtered reads of the ATAC-seq data were used as input to detect open chromatin clusters. The detected peaks within a 12.5 kb window were merged. The open chromatin clusters were then ranked based on their read coverage and plotted in a reverse ROC curve. The clusters passing the point where the slope is greater than 1 were defined as the super-enhancers.

Motif analysis

We identified and quantified the number of motifs for the peaks specific to MRT control or *SMARCB1*+ samples using GimmeMotifs (version 0.13.1)[46] and the non-redundant cis-bp database (version 3.0). As a background peak set, we used the peaks that were unchanged upon *SMARCB1* re-expression. First, we normalized motif frequencies to the total number of identified motifs in that sample. Then, we calculated the log2-enrichment score of MRT control or *SMARCB1*+ motifs by comparison to motif frequency in the background peak set. The p-value was calculated using the Fisher exact test. To detect the Tn5 integration pattern around a transcription factor motif, we first obtained the genomic coordinates of the motifs of interest from the results of the GimmeMotifs analysis. The pyDNase (https://pythonhosted.org/pyDNase/scripts.html) footprint detector was used to determine the Tn5 integration pattern around a motif using the mode "ATAC-seq". The results were visualized as the average Tn5 integration events.

Standard ChIP-seq and calibrated ChIP-seq (RELACS)

The ChIP-seq experiments were performed according to an established protocol[43]. MRT organoids were washed in ice-cold AdDF+++ and PBS. Subsequently, cells were cross-linked with a final concentration of 1% formaldehyde for 10 min. Glycine (0.2 M final concentration) was used to quench the cross-linking reaction. The cross-linked cells were then lysed and sonicated using Bioruptor Plus sonication device (Diagenode) to obtain ~300 bp chromatin. For ChIP, sonicated chromatin was incubated overnight at 4°C with antibodies that had first been coupled to Protein G beads (ThermoFisher). After incubation, captured chromatin was washed, eluted and de-crosslinked. The released DNA fragments were purified using MiniElute PCR Purification Kit (Qiagen). The ChIP experiments were performed using the following antibodies: CTCF (07-729, Merck Millipore, 5 µl per ChIP), RAD21 (ab154769, Abcam, 2.2 µg per ChIP), and H3K27ac (ab4729, Abcam, 5 µg per ChIP). The purified DNA fragments were prepared using the KAPA HTP Library Preparation Kit (Roche) following manufacturer's instructions. The libraries were sequenced on an Illumina HiSeq 2500 using the single-end 65-cycle mode.

Calibrated ChIP-seq (RELACS) experiments were performed according to an established protocol[47]. Nuclei were isolated and permeabilized with 0.5% SDS. Chromatin was digested in situ using restriction enzyme CviKI-1 (NEB, R0710L) and

barcoded using RELACS custom barcodes. Nuclei from each sample were pooled and lysed via sonication to extract the barcoded chromatin fragments. ChIPs were then performed on the extracted chromatin fragments using the antibodies described above, and a small part of sonicated chromatin was preserved as an input control for late quantitative calibration. Immunoprecipitated chromatin was used for Illumina library preparation (NEBNext Ultra II DNA Library Prep Kit, NEB-E7645) and sequenced on HiSeq 2500 Illumina machine (single-end, read length 65 bp).

ChIP-seq data was analyzed as previously described[44]. In short, sequencing reads were mapped to the hg38 reference genome using BWA-MEM (version 0.7.15-r1140)[45]. The mapped reads were filtered using SAMtools, discarding reads with mapping quality score <15, as well as optical PCR duplicates. The coverage files were produced using the deepTools (version 3.0) method "normalize to 1X genome coverage". Alignment of ChIP–seq signal was performed with deepTools (version 3.0), using the "reference-point" method to align the signal coverage from narrow peaks. The 'scale-regions' method was applied to align the signal coverage from broad regions (super enhancers). Heat maps were generated directly using deepTools (version 3.0). The alignment plots of the super enhancers were generated with aligned matrices that were further processed in R.

RNA-seq

RNA-seq data was previously generated and processed by Custers et al[12]. Fold changes induced by *SMARCB1* re-expression were calculated using DESeq2 with default parameters.

Hi-C

We generated Hi-C data as previously described[48] with minor modifications[44]. In short, MRT organoids (~10 millions cells) were washed in ice-cold AdDF+++ and PBS. Subsequently, cells were cross-linked with a final concentration of 2% formaldehyde for 10 min. Glycine (0.2 M final concentration) was used to quench the cross-linking reaction. The restriction enzyme MboI was used to digest crosslinked DNA in the nucleus. At the restriction overhangs, biotinylated nucleotides were incorporated. Subsequently, overhangs were joined by blunt-end ligation. The ligated DNA was enriched by streptavidin pull-down. A standard end-repair and A-tailing method was used to further prepare the Hi-C libraries, which were sequenced on an Illumina Nova-Seq platform generating paired-end 150 bp reads.

Raw Hi-C sequencing reads were processed using HiC-Pro[49], which includes mapping, identification of valid Hi-C pairs, generation of contact matrices and ICE normalization[50]. HiCCUPS (version 0.9) was used to call chromatin loops. Subsequent analyses were performed in GENOVA, a visualization tool of Hi-C data written in R (http://github.com/deWitLab/GENOVA).

4C-seq

4C was performed as previously described[44]. In short, we used MboI as the first and Csp6I as the second restriction enzyme. The viewpoint was designed at the MYC promoter region using the following primers:

Fw: CTCTTTCCCTACACGACGCTCTTCCGATCTTCTCCCTGGGACTCTTGATC

Rv: ACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGTTTAGCCCTGAGATG

The 4C-seq libraries were sequenced using a NextSeq 550.

Mapping of the sequencing reads was performed according to our 4C mapping pipeline (http://github.com/deWitLab/4C_mapping). 4C data was normalized to 1 million intrachromosomal reads using peakC (http://github.com/deWitLab/peakC)[51].

RNA interference

For MYC knockdown, MRT organoids were infected with pLKO.1-puro lentiviruses. For each condition, 30,000 cells were seeded. After two days, puromycin (InvivoGen, 0.5 μg/ml) was added to the culture medium. Nine days after lentiviral infection, organoids were dissociated into single cells, counted, and harvested for RNA. The following shRNA sequences were used: MYC01: CCCAAGGTAGTTATCCTTAAA; MYC02: ACTGAAAGATTTAGCCATAAT

Quantitative RT-PCR

Total RNA was isolated using the MN Nucleospin RNA kit (Bioke, 740955) according to manufacturer's protocol. RNA was reverse transcribed using the GoScript Reverse Transcriptase kit (Promega, A5001) following manufacturer's instructions. mRNA levels of *MYC* were quantified using the iQ SYBR Green Supermix (Bio-Rad, 1708882) and the following qPCR primers: MYC Fw: GATTCTCTGCTCTCCTCGACG, Rv: GATGTGTGGAGACGTGGCA; GAPDH Fw: TGCACCACCAACTGCTTAGC, Rv: GGCATGGACTGTGGTCATGAG

Single-cell mRNA analysis of murine neural crest

UMAP visualization and gene expression data were acquired using the following webtool created by Soldatov et al.[25]. Single-cell mRNA levels were visualized for *Myc* and *Mki67*.

http://pklab.med.harvard.edu/cgi-bin/R/rook/nc.cranial_E85/index.html

MRT tissues and non-malignant cell H3K27ac analysis

ChIP-seq data was extracted from GEO (https://www.ncbi.nlm.nih.gov/geo/): Astrocyte_NH-A (GSM733763), CNCCs_H9 (GSM1817151), HEK293T (GSM1249889), hESC_H9 (GSM602294), hNCC (GSM2664365), Neuron_H1 (GSM2527494), NPC_H9 (GSM2534037), K_RT242 (GSM1835844), K_RT262 (GSM1835847), K_RT407 (GSM1835850), K_STJ0090 (GSM2400527), K_STJ0347 (GSM2400528), K_STJ0537 (GSM2400529).

References

1. Grobner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321-327, doi:10.1038/nature25480 (2018).

2. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371-376, doi:10.1038/nature25795 (2018).

3. Chun, H. E. *et al.* Identification and Analyses of Extra-Cranial and Cranial Rhabdoid Tumor Molecular Subgroups Reveal Tumors with Cytotoxic T Cell Infiltration. *Cell Rep* **29**, 2338-2354 e2337, doi:10.1016/j.celrep.2019.10.013 (2019).

4. Hasselblatt, M. *et al.* High-resolution genomic analysis suggests the absence of recurrent genomic alterations other than SMARCB1 aberrations in atypical teratoid/rhabdoid tumors. *Genes Chromosomes Cancer* **52**, 185-190, doi:10.1002/gcc.22018 (2013).

5. Hasselblatt, M. *et al.* SMARCA4-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. *Acta Neuropathol* **128**, 453-456, doi:10.1007/s00401-014-1323-x (2014).

6. Johann, P. D. *et al.* Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes. *Cancer Cell* **29**, 379-393, doi:10.1016/j.ccell.2016.02.001 (2016).

7. Lee, R. S. *et al.* A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. *J Clin Invest* **122**, 2983-2988, doi:10.1172/JCI64400 (2012).

8. Wilson, B. G. & Roberts, C. W. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* **11**, 481-492, doi:10.1038/nrc3068 (2011).

9. Brennan, B., Stiller, C. & Bourdeaut, F. Extracranial rhabdoid tumours: what we have learned so far and future directions. *Lancet Oncol* **14**, e329-336, doi:10.1016/S1470-2045(13)70088-3 (2013).

10. Hoffman, L. M. *et al.* Advancing biology-based therapeutic approaches for atypical teratoid rhabdoid tumors. *Neuro Oncol* **22**, 944-954, doi:10.1093/neuonc/noaa046 (2020).

11. Han, Z. Y. *et al.* The occurrence of intracranial rhabdoid tumours in mice depends on temporal control of Smarcb1 inactivation. *Nat Commun* **7**, 10421, doi:10.1038/ncomms10421 (2016).

12. Custers, L. *et al.* Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nat Commun* **12**, 1407, doi:10.1038/s41467-021-21675-6 (2021).

13. Langer, L. F., Ward, J. M. & Archer, T. K. Tumor suppressor SMARCB1 suppresses super-enhancers to govern hESC lineage determination. *Elife* **8**, doi:10.7554/eLife.45672 (2019).

14. Parisian, A. D. *et al.* SMARCB1 loss interacts with neuronal differentiation state to block maturation and impact cell stability. *Genes Dev* **34**, 1316-1329, doi:10.1101/gad.339978.120 (2020).

15. Terada, Y. *et al.* Human Pluripotent Stem Cell-Derived Tumor Model Uncovers the Embryonic Stem Cell Signature as a Key Driver in Atypical Teratoid/Rhabdoid Tumor. *Cell Rep* **26**, 2608-2621 e2606, doi:10.1016/j.celrep.2019.02.009 (2019).

16. Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun* **8**, 14648, doi:10.1038/ncomms14648 (2017).

17. Calandrini, C. *et al.* An organoid biobank for childhood kidney cancers that captures disease and tissue heterogeneity. *Nat Commun* **11**, 1310, doi:10.1038/s41467-020-15155-6 (2020).

18. Ong, C. T. & Corces, V. G. Enhancers: emerging roles in cell fate specification. *EMBO Rep* **13**, 423-430, doi:10.1038/embor.2012.52 (2012).

19. Shashikant, T. & Ettensohn, C. A. Genome-wide analysis of chromatin accessibility using ATAC-seq. *Methods Cell Biol* **151**, 219-235, doi:10.1016/bs.mcb.2018.11.002 (2019).

20. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**, 825-837, doi:10.1016/j.molcel.2013.01.038 (2013).

21. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

22. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat Genet* **47**, 8-12, doi:10.1038/ng.3167 (2015).

23. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet* **15**, 69-81, doi:10.1038/nrg3623 (2014).

24. Mashtalir, N. *et al.* Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* **175**, 1272-1288 e1220, doi:10.1016/j.cell.2018.09.032 (2018).

25. Soldatov, R. *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**, doi:10.1126/science.aas9536 (2019).

26. Albini, S. *et al.* Epigenetic reprogramming of human embryonic stem cells into skeletal muscle cells and generation of contractile myospheres. *Cell Rep* **3**, 661-670, doi:10.1016/j.celrep.2013.02.012 (2013).

27. Braccioli, L. & de Wit, E. CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem* **63**, 157-165, doi:10.1042/EBC20180069 (2019).

28. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* **60**, 676-684, doi:10.1016/j.molcel.2015.09.023 (2015).

29. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922, doi:10.1016/j.cell.2017.05.004 (2017).

30. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244, doi:10.1038/nature14450 (2015).

31. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* **20**, 437-455, doi:10.1038/s41576-019-0128-0 (2019).

32. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

33. Sartorelli, V. & Lauberth, S. M. Enhancer RNAs are an important regulatory layer of the epigenome. *Nat Struct Mol Biol* **27**, 521-528, doi:10.1038/s41594-020-0446-0 (2020).

34. Michel, B. C. *et al.* A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol* **20**, 1410-1420, doi:10.1038/s41556-018-0221-1 (2018).

35. Wang, X. *et al.* BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. *Nat Commun* **10**, 1881, doi:10.1038/s41467-019-09891-7 (2019).

36. Young, M. D. et al. Single cell derived mRNA signals across human kidney tumors. Nat Commun 12, 3896, doi:10.1038/s41467-021-23949-5 (2021).

37. Filbin, M. G. *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **360**, 331-335, doi:10.1126/science.aao4750 (2018).

38. Gojo, J. *et al.* Single-Cell RNA-Seq Reveals Cellular Hierarchies and Impaired Developmental Trajectories in Pediatric Ependymoma. *Cancer Cell* **38**, 44-59 e49, doi:10.1016/j.ccell.2020.06.004 (2020).

39. Jessa, S. *et al.* Stalled developmental programs at the root of pediatric brain tumors. *Nat Genet* **51**, 1702-1713, doi:10.1038/s41588-019-0531-7 (2019).

40. Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599, doi:10.1126/science.aat1699 (2018).

41. Chasse, M. H. *et al.* Mithramycin induces promoter reprogramming and differentiation of rhabdoid tumor. *EMBO Mol Med* **13**, e12640, doi:10.15252/emmm.202012640 (2021).

42. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490, doi:10.1038/nature14590 (2015).

43. Liu, N. Q. *et al.* The non-coding variant rs1800734 enhances DCLK3 expression through long-range interaction and promotes colorectal cancer progression. *Nat Commun* **8**, 14418, doi:10.1038/ncomms14418 (2017).

44. Liu, N. Q. *et al.* WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nat Genet* **53**, 100-109, doi:10.1038/s41588-020-00744-4 (2021).

45. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843-2851, doi:10.1093/bioinformatics/btu356 (2014).

46. van Heeringen, S. J. & Veenstra, G. J. GimmeMotifs: a *de novo* motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics* **27**, 270-271, doi:10.1093/bioinformatics/btq636 (2011).

47. Arrigoni, L. *et al.* RELACS nuclei barcoding enables high-throughput ChIP-seq. *Commun Biol* **1**, 214, doi:10.1038/s42003-018-0219-z (2018).

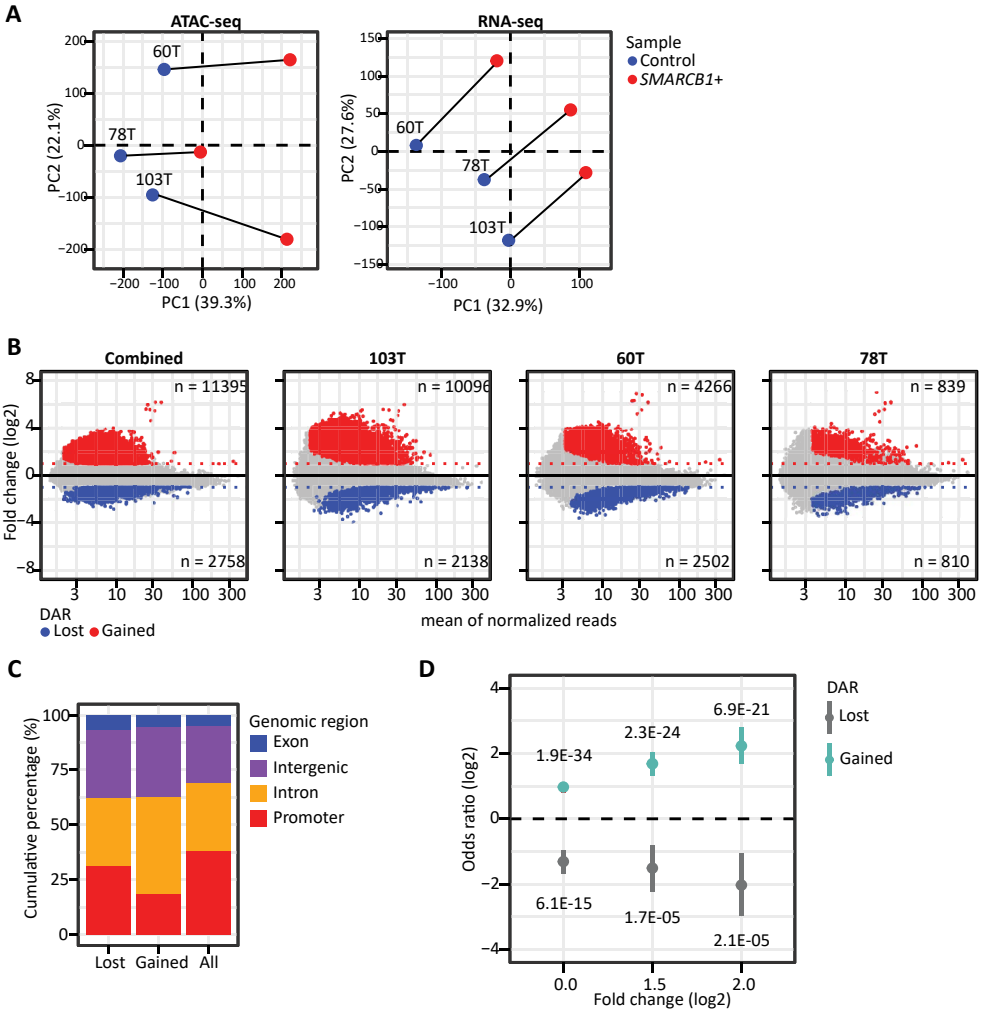
48. Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

49. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).

50. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9**, 999-1003, doi:10.1038/nmeth.2148 (2012).

51. Geeven, G., Teunissen, H., de Laat, W. & de Wit, E. peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res* **46**, e91, doi:10.1093/nar/gky443 (2018).

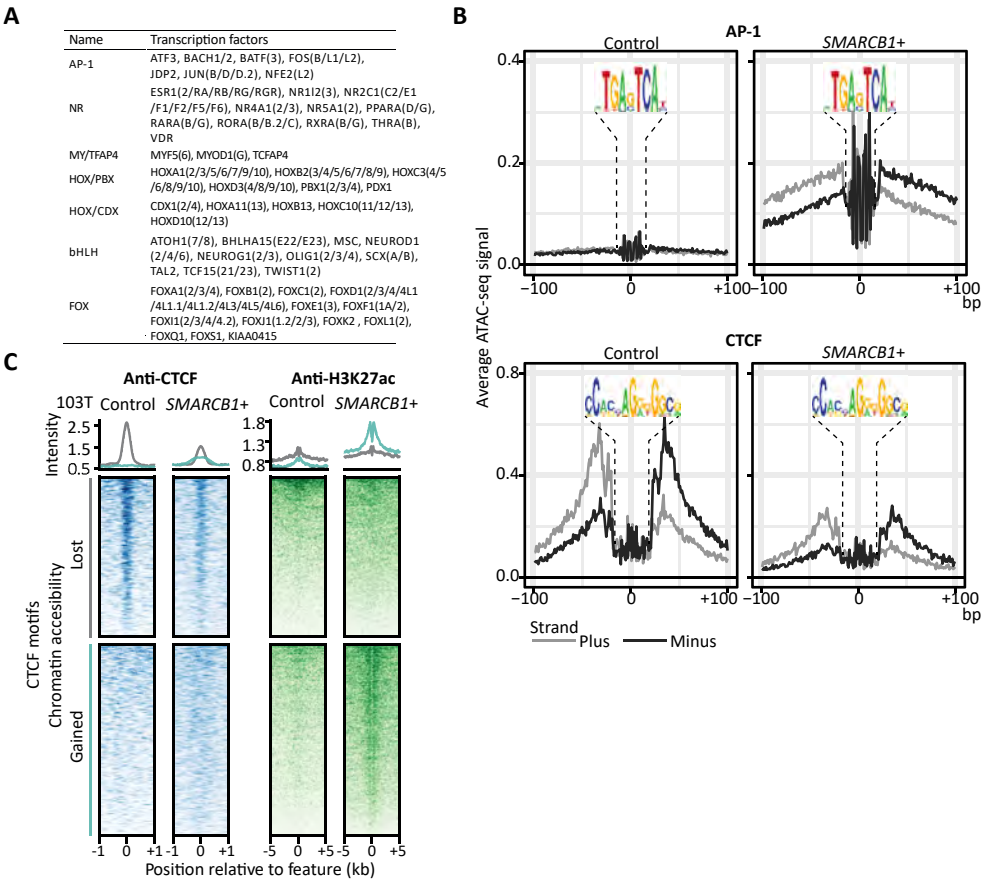
Supplementary information



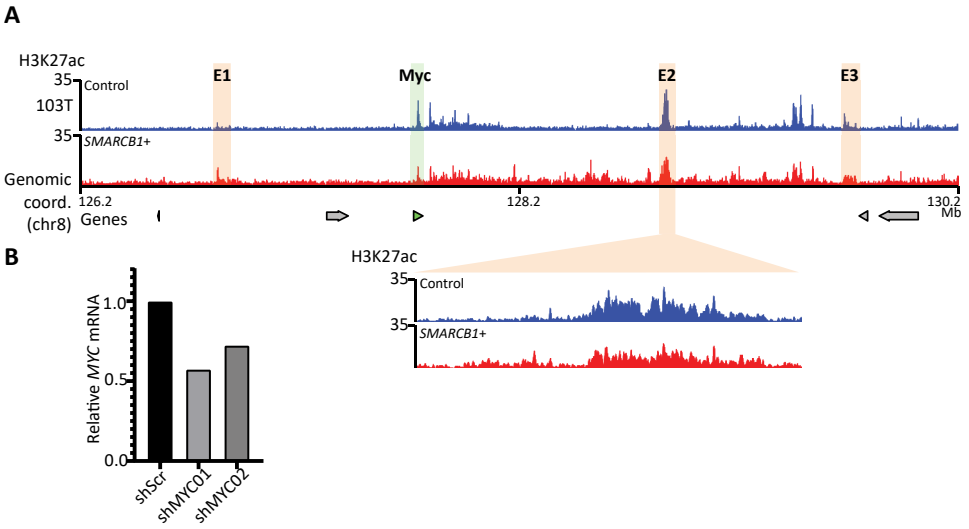
Supplementary Fig. 1 | Characterization of SMARCB1-dependent gene regulatory elements. **(A)** Principal component analysis performed on chromatin accessibility patterns (left) or mRNA levels (right) of MRT control (blue) and *SMARCB1+* (red) organoids. Chromatin accessibility patterns are an average of three independent experiments, whereas mRNA levels are an average of two independent experiments. Organoids models are labeled and connected by line. **(B)** Scatter plot shows chromatin accessibility changes induced by *SMARCB1* re-expression for each MRT organoid line separate or averaged (combined 60T, 78T and 103T). Genomic regions that significantly (FDR <0.05) gained (red) or lost (blue) chromatin accessibility are highlighted. Fold changes represent an average of three independent experiments. **(C)** Stacked bar plot represents relative frequencies of genomic sites localizing to gene promoters, introns, exons or intergenic regions, distinguishing genomic sites that gained or lost chromatin accessibility upon *SMARCB1* reconstitution, or all open chromatin in MRT organoids. **(D)** Odds ratios indicate a significant association of chromatin accessibility and gene expression changes induced by *SMARCB1* re-expression in MRT organoids (Fisher's Exact test). Three different cutoffs of log2 fold changes were used to calculate the association, which increased when the cutoff became more stringent.

A

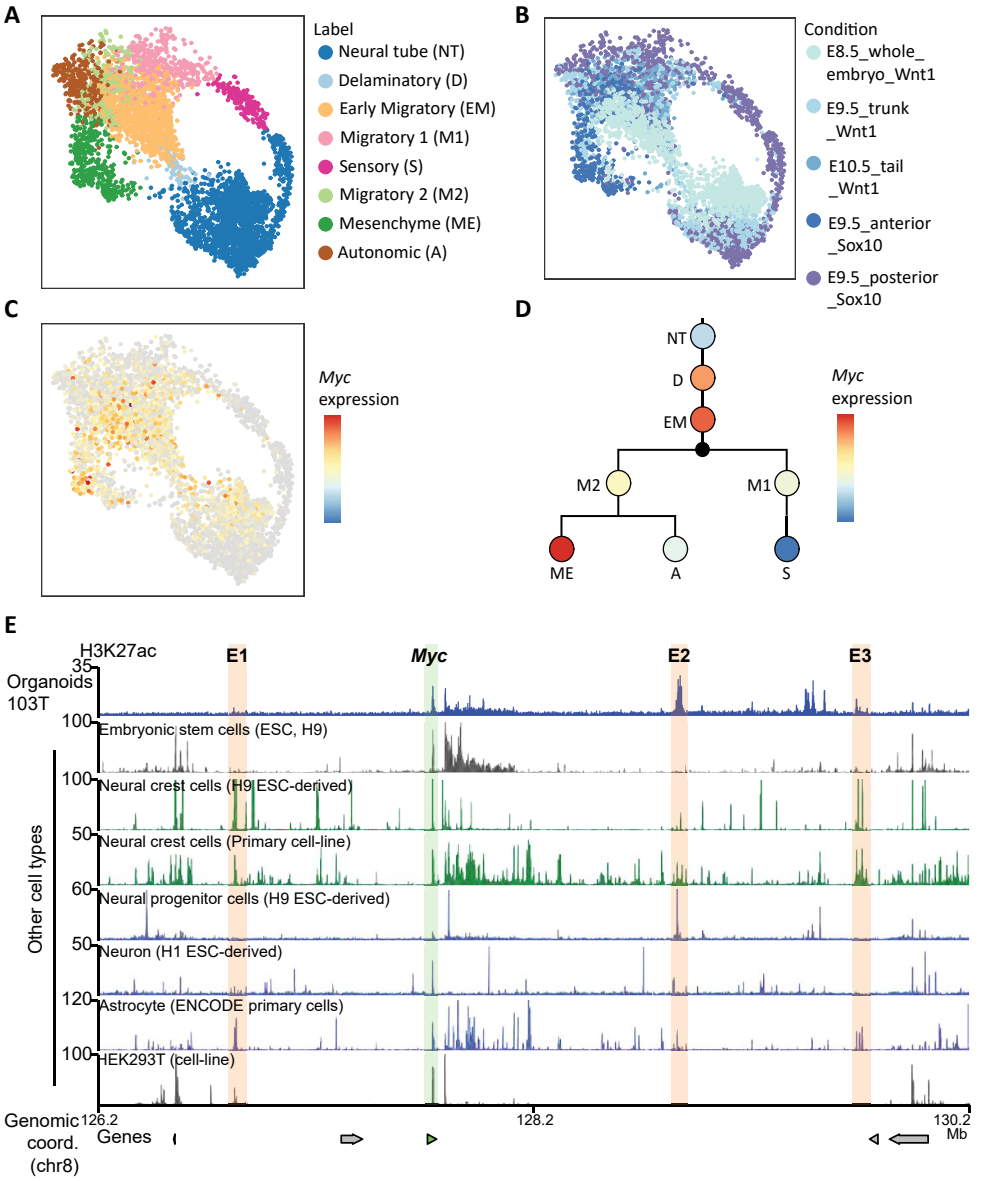
Name	Transcription factors
AP-1	ATF3, BACH1/2, BATF(3), FOS(B/L1/L2), JDP2, JUN(B/D/D.2), NFE2(L2)
NR	ESR1(2/RA/RB/RG/RGR), NR1I2(3), NR2C1(C2/E1 /F1/F2/F5/F6), NR4A1(2/3), NR5A1(2), PPARA(D/G), RARA(B/G), RORA(B/B.2/C), RXRA(B/G), THRA(B), VDR
MY/TFAP4	MYF5(6), MYOD1(G), TCFAP4
HOX/PBX	HOXA1(2/3/5/6/7/9/10), HOXB2(3/4/5/6/7/8/9), HOXC3(4/5 /6/8/9/10), HOXD3(4/8/9/10), PBX1(2/3/4), PDX1
HOX/CDX	CDX1(2/4), HOXA11(13), HOXB13, HOXC10(11/12/13), HOXD10(12/13)
bHLH	ATOH1(7/8), BHLHA15(E22/E23), MSC, NEUROD1 (2/4/6), NEUROG1(2/3), OLIG1(2/3/4), SCX(A/B), TAL2, TCF15(21/23), TWIST1(2)
FOX	FOXA1(2/3/4), FOXB1(2), FOXC1(2), FOXD1(2/3/4/4L1 /4L1.1/4L1.2/4L3/4L5/4L6), FOXE1(3), FOXF1(1A/2), FOXI1(2/3/4/4.2), FOXJ1(1.2/2/3), FOXK2 , FOXL1(2), FOXQ1, FOXS1, KIAA0415



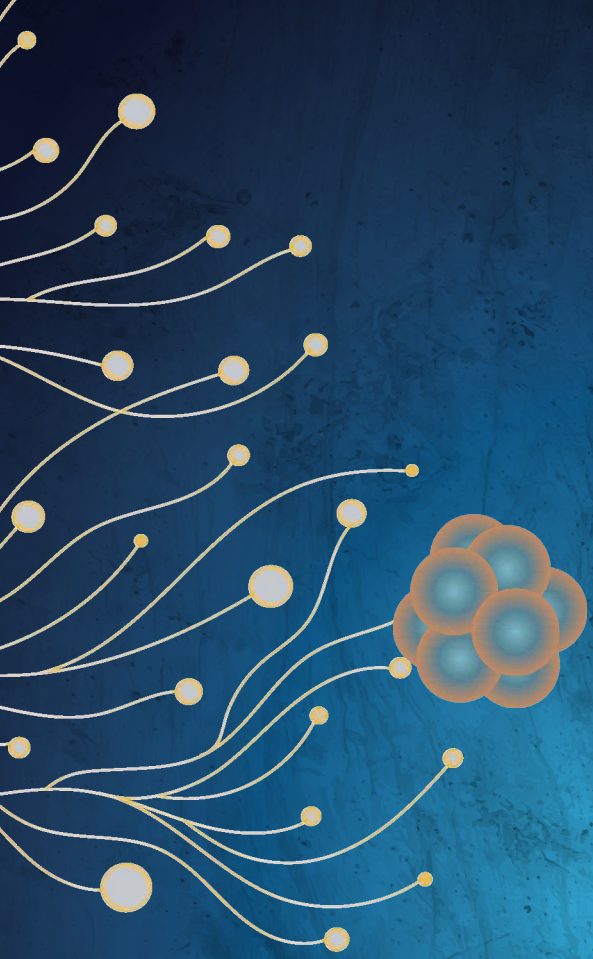
Supplementary Fig. 2 | Transcription factor binding motif analysis.
(A) Table shows transcription factors associated with the motifs in Fig. 2a. **(B)** Line graphs show average ATAC-seq signals of differentially accessible regions containing AP-1 or CTCF binding motifs for MRT control (left) or *SMARCB1*+ (right) organoids (average of 60T, 78T and 103T). The genomic sites are 200bp in length and are centered on the motif. Chromatin accessibility is distinguished for the plus (dark) and minus (light) DNA strand. **(C)** Heatmaps show CTCF and H3K27ac binding (calibrated ChIP-seq) for 103T control and *SMARCB1*+ MRT organoids. The rows represent 2kb genomic regions centered on signal peaks. Rows are subset for genomic sites containing CTCF motifs that lost (gray) or gained (green) chromatin accessibility in MRT *SMARCB1*+ cells. Average signal intensity is indicated (top).



Supplementary Fig. 3 | *MYC* in MRT organoids.
(A) Line graphs show H3K27ac signals (ChIP-seq) of the *MYC* locus for 103T control (blue) and *SMARCB1*+ (red) MRT organoids. Genomic regions of MYC (green) and super-enhancers identified in MRT organoids (E1-3, yellow) are highlighted. **(B)** Bar graph depicts mRNA expression levels (RT-qPCR) of MYC for 103T MRT organoids expressing short hairpin RNAs targeting *MYC* mRNA (shMYC01, shMYC02) or a scrambled control (shScr). Values represent one experiment. Expression levels are normalized to *GAPDH* mRNA and DMSO control. Of note, RNA was harvested at the endpoint of the experiment. We showed that MYC knockdown has a growth inhibitory effect on MRT cells (Fig. 4d). Therefore, we expect that *MYC* mRNA levels are not fully representative of the starting point of the experiment.



Supplementary Fig. 4 | *MYC* in neural crest cells.
(A-C) UMAP representation of single cell transcriptomes derived from murine neural crest generated by Soldatov et al[25]. Single cells are depicted by color according to cell-type in A, mouse model and time of extraction in B, or *Myc* mRNA expression levels in C. **(D)** Branching tree represents differentiation trajectories of mouse neural crest. Abbreviations are indicated. Dot plots represent average *Myc* mRNA expression levels in different neural crest cell types. **(E)** Line graphs show H3K27ac signals (ChIP-seq) of the *MYC* locus for 103T MRT organoids and non-malignant cell-types. Genomic regions of MYC (green) and super-enhancers identified in MRT organoids (E1-3, yellow) are highlighted.



CHAPTER 6

A CRISPR-Cas9 knock-out screen in patient-derived organoids identifies potential vulnerabilities of malignant rhabdoid tumors

Lars Custers[1,2], Maroussia Ganpat[1,2], Francisco Morales-Rodriguez[1,2], Jarno Drost[1,2]

Manuscript in preparation

[1]Princess Máxima Center for Pediatric Oncology, 3584CS, Utrecht, the Netherlands

[2]Oncode Institute, the Netherlands

Abstract

Malignant rhabdoid tumors (MRT) are aggressive cancers of early childhood, characterized by a single recurrent genetic driver event, the bi-allelic loss *SMARCB1*. MRT are difficult to treat and lack actionable genetic events. There is therefore an urgent need for novel effective treatment options. A subset of MRT is thought to be driven by overexpression of *MYC*, which however is not directly druggable up to date. Therapeutic innovation is hampered by the lack of physiologically relevant MRT cell models. Here, we deployed CRISPR-Cas9 knock-out screens in patient-derived MRT organoids, which were demonstrated to faithfully resemble patient MRT, to uncover potential therapeutic targets. We identified the mRNA splicing machinery as a potential MRT-specific dependency, which we confirmed pharmacologically. Furthermore, analysis of MYC-dependent vulnerabilities suggested CDK12 inhibition as an alternative route to target MYC and impair MRT growth. This study demonstrates that an unbiased genetic screening approach using physiologically relevant cancer models can provide novel insights into tumor dependencies, which is urgently required for lethal tumors such as MRT.

Introduction

Extensive characterization of cancer genomes over the years has identified many oncogenes that have been established as suitable targets for treatment[1]. Yet, mutations affecting tumor suppressor genes are not directly druggable. To identify vulnerabilities in tumors harboring such undruggable driver events, a mechanistic understanding of tumor biology is crucial. A particularly lethal pediatric cancer that lacks targetable events is malignant rhabdoid tumor (MRT)[2-4]. MRT is typically diagnosed in early childhood, but already initiates during embryonic development[5-7]. Tumors most frequently present inside the central nervous system (CNS), where they are referred to as Atypical Teratoid Rhabdoid Tumor (AT/RT), or outside the CNS in soft tissues and organs (e.g. kidney, liver), which are grouped as extracranial rhabdoid tumors (ECRTs)[2-4]. MRTs are characterized by a single recurrent driver event, which is the bi-allelic loss of tumor suppressor *SMARCB1* (>95%) or *SMARCA4* (<5%)[8-13]. Both are core subunits of the SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodeler complex. Furthermore, AT/RT can be classified into subgroups based on DNA methylation profiles, referred to as MYC, Sonic hedgehog (SHH) and Tyrosinase (TYR)[8,11]. ECRT epigenetics and transcriptomics are primarily shared with the AT/RT-MYC subgroup, indicating a close molecular relationship[8]. MRT subgroups are featured by distinct signaling pathways and are therefore hypothesized to comprise differential vulnerabilities[14].

Besides *SMARCB1* mutations, MRT are genetically stable with absence of additional recurrent genetic tumor drivers[8-13], which hampers the design of therapies that are based on actionable genetic events (so-called targeted therapy). Therefore, uncovering

signaling pathways downstream of *SMARCB1* loss underpinning tumorigenesis is crucial for the development of novel targeted therapy. Perturbation screens have significantly contributed to define gene function in malignant growth in an unbiased fashion. This includes high-throughput drug screens, which provide a fast platform for assessment of growth inhibitory effects across *in vitro* human cancer models[15]. However, drug screens are limited to compound availability, which hampers the discovery of novel potential targets. The development of the CRISPR-Cas9 genetic toolkit in combination with pooled single guide RNA libraries enabled knock-out screens that allow for assessment of gene essentiality in a genome-wide fashion[16]. Tumor-specific dependencies can be assessed by comparison of gene essentialities across tumor entities, which has been facilitated by Project Achilles that involved the genetic screening of over 800 cancer cell-lines to generate a Cancer Dependency Map (DepMap)[17].

These efforts allowed a comparison of gene dependencies between MRT and other cancer cell-lines[18-21]. However, a drawback of cancer cell-lines is that they often poorly represent patient tumor tissue, which complicates the translation of experimental findings to clinical application[22,23]. This also applies to MRT, as transcriptomic and epigenetic profiling of several MRT cell-lines revealed that, for the majority, similarity to patient MRT was questionable[19]. Recent advancements in cell culture technology have contributed to more accurate modeling of patient tumor physiology, which includes the development of 3D stem-cell based cultures referred to as organoids[24]. Tumor organoid models can be established with high efficiency from patient tumor biopsies and have shown improved resemblance to their native tissue[25]. Recently, organoid models for renal MRT were established that, in contrast to cell-lines, faithfully resembled patient tumor tissue on a morphological, (epi) genetic and transcriptomic level[26].

Loss of *SMARCB1* has been proposed to cause aberrant activation of kinase signaling pathways responsible for MRT growth[19,27]. A majority of pre-existing drugs targets kinases as they have a druggable protein structure that can be effectively inhibited. Therefore, as no effective treatment exists for MRT up to date, we sought out to interrogate the MRT kinome for vulnerabilities that can potentially be exploited for therapy. Here, we defined the essentially of 504 kinases by deploying CRISPR-Cas9 knock-out screens in patient-derived renal MRT organoids.

Results

CRISPR-Cas9 knock-out screens reveal essential kinases in MRT organoids

In this study, we included two organoid lines that were derived from renal MRT, namely 60T and 103T, that have previously been established and characterized in depth[26]. For genetic screening (**Fig. 1a**), both organoid lines were lentivirally transduced with an established LentiCrisprV2 library[28] including 5971 sgRNAs targeting 504

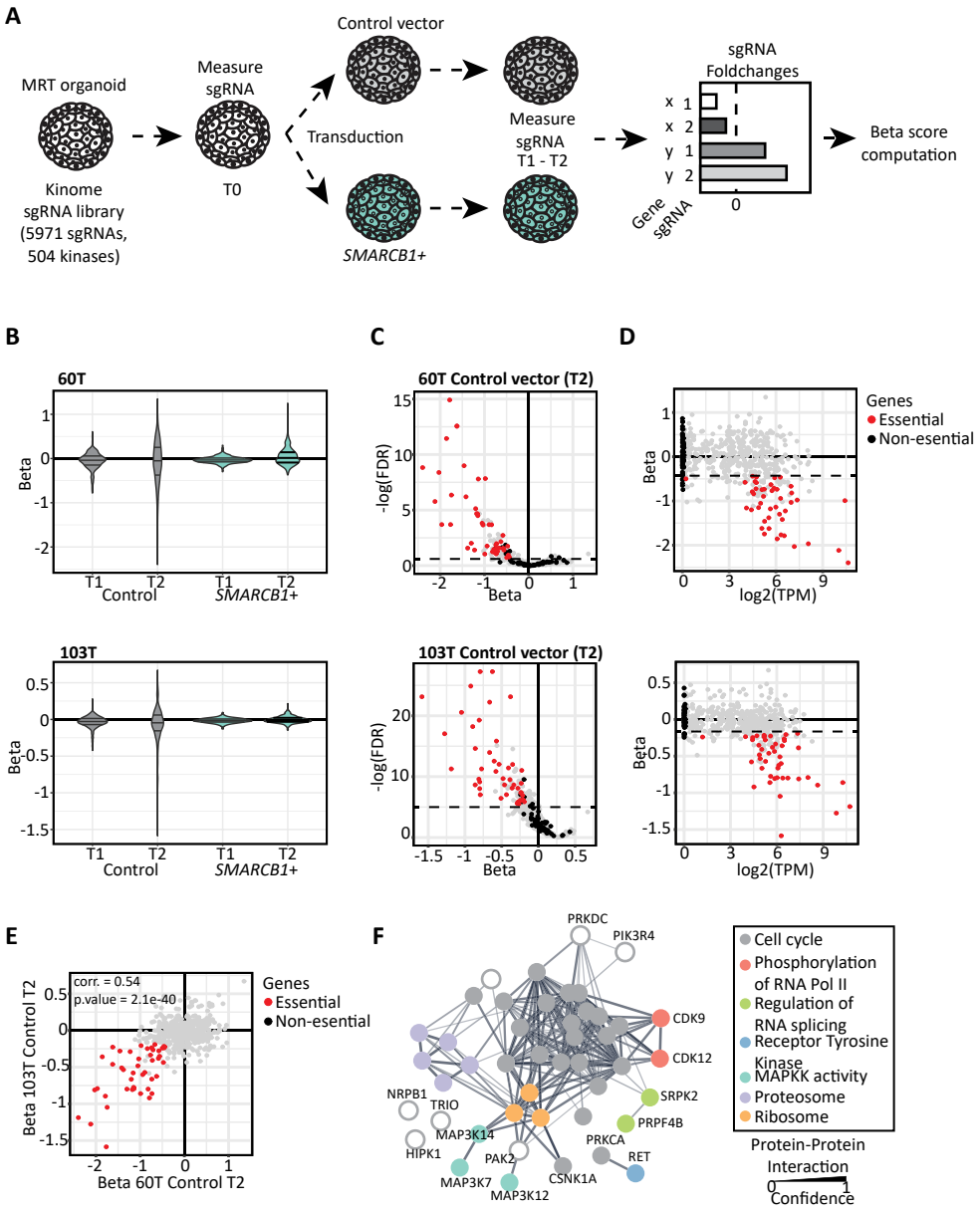


Figure 1 | CRISPR-Cas9 knock-out screens reveal essential kinases in MRT organoids.
(A) Schematic representation of CRISPR-Cas9 knock-out screen targeting 504 kinases in MRT organoids with and without *SMARCB1* reconstitution. **(B)** Violin plot shows the distribution of normalized gene dependency scores (beta) derived from the CRISPR-Cas9 knock-out screens performed in MRT control and *SMARCB1*+ organoid lines. Different timepoints (T1 and T2) are indicated. Beta scores of MRT control organoids at T2 are used for subsequent analysis. **(C)** Scatter plots display beta scores and associated false discovery rates (FDR) calculated by a Wald statistical test. The FDR cutoff used for identification of essential kinases is visualized by the dashed line. Color indicates genes that are not expressed and therefore considered

different kinases, 10 essential genes and 50 non-targeting control sgRNAs. sgRNA distribution was assessed using targeted next generation sequencing of genomic DNA sgRNA integrants. A quantification of sgRNA counts at timepoint zero (T0) exhibited a uniform coverage of sgRNAs, indicating a successful transduction of the sgRNA library into MRT cells (**Supplementary Fig. 1a**). Next, to investigate *SMARCB1*-dependent vulnerabilities, MRT organoids were lentivirally transduced with a control or *SMARCB1* expression construct (**Fig. 1a**). Reconstitution of SMARCB1 induces a growth arrest and differentiation phenotype in MRT organoids that we have previously characterized in detail[5]. Subsequently, MRT organoids with and without *SMARCB1* were cultured and sgRNAs were quantified at two timepoints (T1 and T2) for three biological replicates (**Fig. 1a**, **Supplementary Fig. 1a, b**). The effect of gene knock-out on MRT organoid growth was assessed by calculation of sgRNA fold changes, comparing counts from T1 and T2 samples to the starting point T0. Based on sgRNA fold changes, gene dependency scores (beta) were generated using an established analysis pipeline[29] that involved maximum-likelihood estimation (MLE) correcting for gene copy number variations (**Fig. 1a**). Beta scores are therefore a representation of the sgRNA foldchanges that occur due to the knock-out of the gene they target. The beta score is only significant when multiple sgRNAs targeting that gene are effective, and when this effect is reproducible across the replicate experiments. Negative beta scores correspond to a growth disadvantage and vice versa. To account for gene-independent toxicity of Cas9-induced double strand breaks, we defined non-essential genes by identifying kinases that were not expressed, and normalized gene dependency scores accordingly[17,29].

We observed that, in general, the kinase perturbation effect on MRT organoid growth was enhanced over time (**Fig. 1b**), which agrees with the notion that gene dependency scores correlate with the number of cell doublings[29]. In the proliferating MRT control organoids, the overall effects of kinase knock-out were biased towards a growth disadvantage. In contrast, the MRT *SMARCB1*+ cells predominantly showed sgRNA enrichment over time, suggesting a bypass of the *SMARCB1*-induced growth arrest (**Fig. 1b**).

First, we sought out to identify the kinases for which genetic perturbation had strongest effects on MRT organoid proliferation. We aimed to limit false positives by selecting statistically significant genes (**Fig. 1c**) and by setting a beta score cutoff based on non-essential gene distribution (**Fig. 1d**). Post filtering, we showed that

as non-essential (black) or genes that were identified as essential in both 60T and 103T (red). **(D)** Scatter plots show the correlation of gene expression levels and beta scores. The beta score cutoff (dashed line) used for essential gene identification was set at the 10th percentile of the beta score distribution of non-essential genes. **(E)** Comparison of beta scores of 60T and 103T patient lines shows a significant correlation. Pearson's correlation coefficient (corr.) and associated p-value is indicated. **(F)** STRING protein interaction map that illustrates functional and physical interactions (node) of kinases (dot, gene names are indicated) essential in both 60T and 103T organoid lines. Interaction confidence is indicated by line thickness. Colors depict cellular process related to that kinase.

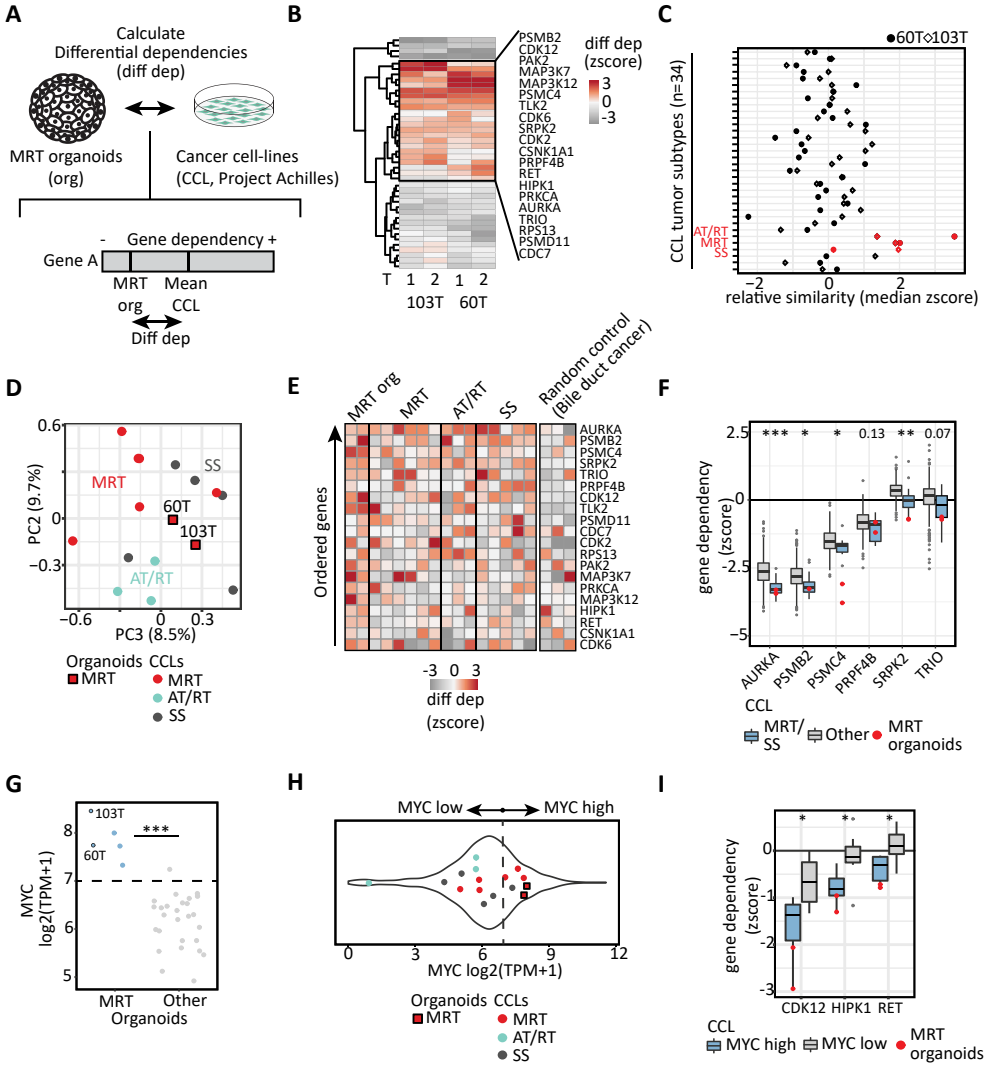


Figure 2 | Comparison of kinase dependencies to cancer cell-lines identifies MRT-specific vulnerabilities.

(A) Schematic visualization of the method used for calculating differential dependency scores (diff dep) across MRT organoids (org) and cancer cell-lines (CCL, Project Achilles[33]). **(B)** Heatmap shows differential dependency scores (gray to red) of kinases identified as essential in MRT organoids (Fig. 1). A positive differential dependency scores means that MRT organoids are particularly vulnerable to genetic perturbation of that kinase compared to other CCLs. Timepoint (T) 1 and 2 of MRT organoids (60T and 103T) are indicated. Genes that represent MRT-organoid specific vulnerabilities are labeled. **(C)** Dot plot represents the similarity of MRT organoid kinase dependencies to n=34 different tumor subtypes. Similarity scores were generated by comparing gene dependencies of MRT organoids at T2 to each CCL included in Project Achilles. These similarity scores were scaled for each CCL and median scores per tumor subtype are indicated for 60T (circle) and 103T (diamond). The tumor subtypes with the highest similarity scores are labeled (red). **(D)** Principal component analysis performed on gene dependency correlations between MRT/

MRT organoid growth is dependent on 45 kinases, which were identified as essential in both patient lines (**Fig. 1c, d, Supplementary Fig. 1c, Supplementary Table 1**). Gene dependency was consistent between 60T and 103T (**Fig. 1e**) and was found to correlate with mRNA levels (**Fig. 1c**), which is a strong predictor of gene essentiality[30]. As expected, the identified essential kinases were not affecting growth of *SMARCB1+* MRT cells in which cell cycle arrest was induced[5] (**Supplementary Fig. 1d**).

Subsequently, we analyzed kinase function by generating a protein interaction network that pointed out essential cellular processes such as the cell cycle, proteasome, ribosome and RNA metabolism (**Fig. 1f**). Furthermore, we found organoid growth to be dependent on receptor tyrosine kinases (RTKs) and MAPKs (**Fig. 1f**), which have previously been implicated as MRT vulnerabilities[8,19,31,32]. Overall, the genetic screening of MRT organoids faithfully identified essential kinases, which formed the basis for further investigations into MRT-specific dependencies.

Comparison of kinase dependencies to cancer cell-lines identifies MRT-specific vulnerabilities

Besides drug efficacy, tumor specificity is crucial to establish an effective therapeutic window. MRT-specific vulnerabilities can be interrogated by comparing gene dependencies between MRT and other cancer models. To identify such differential dependencies, we combined our CRISPR-Cas9 knock-out screening results in MRT organoid models (**Fig. 1**) with the full cancer cell-line dataset (Project Achilles) that includes a large variety of cancer types[33] (**Fig. 2a**). To ensure a valid comparison, we scaled gene dependencies by cell-line/organoid to account for confounder effects such as proliferation rate[17]. This normalization resulted in comparable distributions of gene dependency scores (**Supplementary Fig. 2a**). Furthermore, we validated that common essential genes, which were defined by analysis of cancer cell-lines[17] (Project Achilles), were also essential in MRT organoids (**Supplementary Fig. 2a**). Next, we generated a differential dependency score for each kinase by calculating the gene dependency differences across all cancer cell-lines and MRT organoids. A positive score hence indicates a tumor-specific vulnerability. Analysis of differential dependency

SS organoids and cell-lines. Principal component 2 and 3 are visualized. Variance proportions are indicated for each component. Color depicts tumor subtype. **(E)** Heatmap shows differential dependency scores (gray to red) of genes defined as MRT organoid-specific (Fig. 2b) for MRT/SS organoids and cell-lines. As a control, 3 random cancer cell-lines (bile duct) were included. Genes are ordered by average differential dependency score. **(F)** Boxplots represent gene dependency scores of the top 5 kinases with highest average differential dependency score across MRT/SS organoids and cell-lines (Fig. 2e). MRT/SS organoids and cell-lines (blue) are compared with all other CCLs (gray). MRT organoids are indicated (red dot). False discovery rates (FDR) were calculated using a two-sample Wilcoxon test (two-tailed). FDR: * <0.05, ** <0.01, *** <0.001. **(G)** Dot plot compares *MYC* mRNA levels of MRT (blue) and other pediatric renal tumor organoids (gray). P-value was calculated using an unpaired Student's t test (two-tailed). P-value: *** <0.001. **(H)** Violin plot represents *MYC* mRNA levels of CCLs (Project Achilles) and MRT organoids, distinguishing MYC high and low tumor models. MRT/SS cell-lines and organoids are indicated (dot) and are colored by tumor subtype. **(I)** Boxplots compare gene dependency scores of MYC high (blue) and low (gray) MRT/SS cell-lines and organoids. MRT organoids are indicated (red dot). Statistically significant genes (FDR < 0.1) are presented. FDRs were calculated using a two-sample Wilcoxon test (two-tailed). FDR: * <0.1.

scores suggested that the genetic perturbation of 20 kinases affected MRT organoid growth more than other cancer cell-lines (**Fig. 2b**).

We aimed to generalize these findings to a broader cohort of MRT models. First, we compared gene dependency scores of MRT organoids to a large collection of cancer cell-lines including various tumor types, generating a similarity score by correlation methods. We confirmed that kinome perturbation patterns in MRT organoids were most similar to MRT and AT/RT cell-lines (**Fig. 2c**). Additionally, we found a high resemblance to synovial sarcoma (SS), which is characterized by fusion of *SS18* to *SSX* genes. Expression of this fusion protein prevents the integration of SMARCB1 into the SWI/SNF complex[34]. In that perspective, SS development is thought to be driven by a mechanism that is similar to MRT[21]. Therefore, we also included SS cell-lines for subsequent analyses.

Differences in culturing methods and CRISPR-Cas9 screen designs have been shown to significantly affect the outcome of genetic screens[35]. Thus, we investigated how organoid dependencies compared to those of cell-lines. Clustering of MRT/SS cell-line and organoid dependencies demonstrated that indeed most variance can be explained by cell culture differences, as organoids and cell-lines separated first (**Supplementary Fig. 2b**). Secondly, the tumor models clustered mostly by subtype (**Fig. 2d**), which suggests each tumor entity to comprise distinct kinase dependencies.

As both MRT and SS malignancy is caused by aberrant SWI/SNF function, we aimed to define common differential dependencies within the MRT and SS group. In general, gene dependencies were relatively heterogeneous across MRT and SS models (**Fig. 2e**). Nonetheless, we could identify 5 kinases as shared vulnerabilities (**Fig. 2e, f**), including *AURKA*, and proteasome subunits *PSMB2* and *PSMC4* that have been reported previously as MRT vulnerabilities[36-39]. Additionally, we identified genes such as *SRPK2* and *PRPF4B* which have been described as regulators of mRNA splicing[40].

Growth of ECRT and AT/RT-MYC subgroups is thought to be, at least partly, driven by aberrant expression of the *MYC* oncogene[5,8,41]. As direct pharmacological inhibition of MYC remains challenging up to date[42], we sought out to identify *MYC*-dependent kinase vulnerabilities. First, we defined MYC subgroup resemblance of MRT cell-lines and organoids. Oberlick et al. showed that subgroup classification for a majority of MRT cell-lines is unclear, as DNA methylation and mRNA profiles were found to not fully match those of patient tumor tissue[19]. In contrast, the renal MRT organoids faithfully represented the MYC subgroup epigenetically[5] (**Supplementary Fig. 2c**) and exhibited high *MYC* mRNA levels (**Fig. 2g**). In agreement with the epigenetic profiling, only three out of six MRT cell-lines mirrored the *MYC* overexpression phenotype (**Fig. 2h**). It is not clear if the AT/RT cell-lines included were derived from AT/RT-MYC patient material. However, they did not show MYC overexpression. Accordingly, to better capture ECRT vulnerabilities, we interrogated kinase perturbation effects for MYC dependency. MRT and SS models characterized by high *MYC* expression showed increased sensitivity to the genetic perturbation of *CDK12*, *HIPK1* and *RET* (**Fig. 2i**).

CDK12 has previously been reported as a potential target for MYC-driven tumors, as it was found to regulate *MYC* mRNA processing through phosphorylation of RNA polymerase II[43,44].

Altogether, we identified differential dependencies that had previously been reported, which therefore confirm the validity of our screening results and analysis. Additionally, we uncovered novel kinase essentialities involved in mRNA splicing and *MYC* regulation that could serve as potential therapeutic targets.

MRT-specific dependencies correlate with gene expression and reflect tumor origin

Predicting differential dependencies based on DNA mutations is not applicable to MRT, as besides *SMARCB1* loss, other recurrent genetic events are absent. On the other hand, RNA expression has proven to be a strong indicator of tumor-specific vulnerabilities[30]. Accordingly, we assessed the mRNA levels of the essential kinases identified in the MRT organoids (**Fig. 1**) in various *in vitro* models. Gene expression was significantly higher in MRT compared to the proliferative but non-malignant kidney organoids[26], also in three independent MRT organoid lines that were not included for the genetic screens (**Fig. 3a**). We extended this transcriptomic analysis to the cancer cell-lines included in Project Achilles. For assessment of gene expression levels, we made a distinction between common and MRT organoid-specific vulnerabilities. The relative mRNA levels of the genes that were identified as MRT organoid-specific dependencies (**Fig. 2b**), were significantly and consistently higher in MRT organoids compared to other cancer cell-lines (**Fig. 3b**). In contrast, essential kinases that were defined as common vulnerabilities, did not exhibit statistically significant overexpression (**Fig. 3b**). These findings suggest an association between differential dependency and gene expression.

MRT is an embryonal tumor that arises during neural crest development[5,7]. However, MRT subgroups exhibit differential gene expression signatures, with similarities to neural and mesenchymal lineages[5,8,45]. Transcriptomes of ECRT and AT/RT-MYC subgroups predominantly show mesenchymal features. To assess if these cellular identities are reflected in the MRT-specific vulnerabilities, we correlated differential dependencies with normal tissue gene expression patterns (**Fig. 3c**). To avoid correlations based on tissue proliferation rate, we excluded kinases that clustered with proliferation marker *KI67* (**Supplementary Fig. 3**). Accordingly, differential dependencies were found to correlate most strongly with brain tissue transcriptomes, with a highest score for the spinal cord, which was consistent for both MRT organoid lines (**Fig. 3d**). In agreement with the correlation scores, the relative expression levels of the kinases identified as MRT organoid-specific vulnerabilities were significantly higher than the genes defined as common dependencies, which was specific to spinal cord tissue (**Fig. 3e**).

These findings indicate that MRT-specific vulnerabilities are reflected in the gene expression profiles of ectoderm-derived tissues, which suggests that differential dependencies are related to tumor origin.

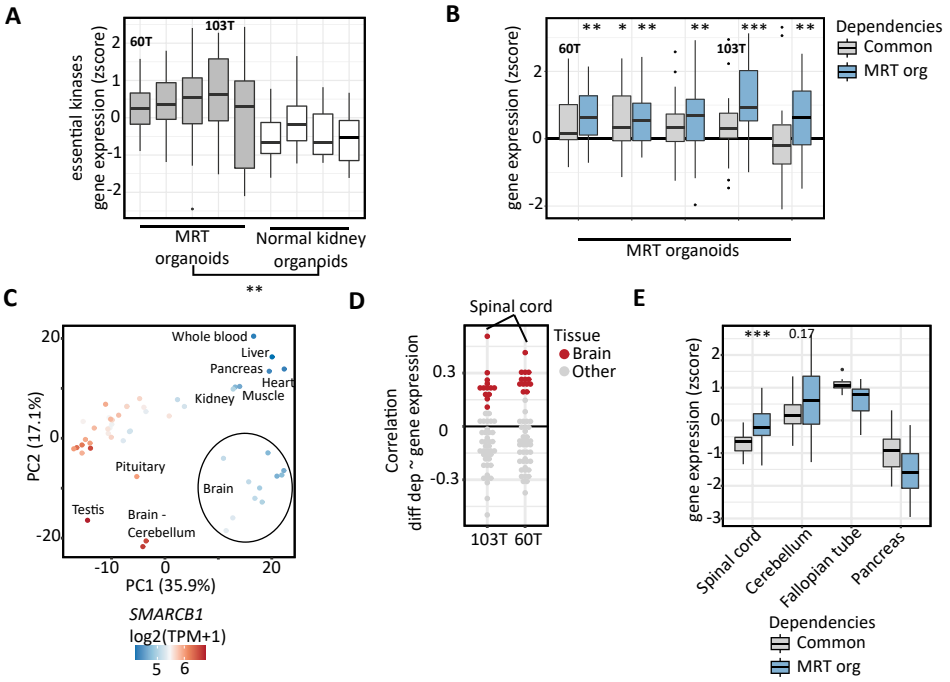


Figure 3 | MRT-specific dependencies correlate with gene expression and reflect tumor origin. **(A)** Boxplots represent relative mRNA expression levels of essential kinases (Fig. 1) comparing MRT and normal kidney organoids. P-value was calculated using an unpaired Student's t test (two-tailed). P-value: ** <0.01. **(B)** Boxplots show mRNA expression levels of essential kinases of MRT organoids relative to CCLs (Project Achilles). Essential kinases are distinguished as common (gray) or MRT organoid specific (MRT org, blue, Fig. 2b) dependencies. Differences in gene expression levels were assessed between MRT organoids and CCLs. P-values were calculated using a two-sample Wilcoxon test (two-tailed). P-value: * <0.05, ** <0.01, *** <0.001. **(C)** Principal component analysis (PCA) performed on the mRNA levels of all kinases included in the CRISPR-Cas9 knock-out screen for a variety of normal tissues (GTEx). Color depicts relative *SMARCB1* expression levels. **(D)** Pearson's correlations were performed to assess the relationship of kinase differential dependency scores in MRT organoids to the kinase mRNA levels in different normal tissues (GTEx). Dot plot represents correlation coefficients for 60T and 103T, indicating brain (red) or other (gray) normal tissues. **(E)** Boxplots represent relative mRNA expression levels of essential kinases, comparing common (gray) with MRT organoid specific (MRT org, blue, Fig. 2b) dependencies in different normal tissues. P-values were calculated using an unpaired Student's t test (two-tailed). P-value: * <0.05, ** <0.01, *** <0.001.

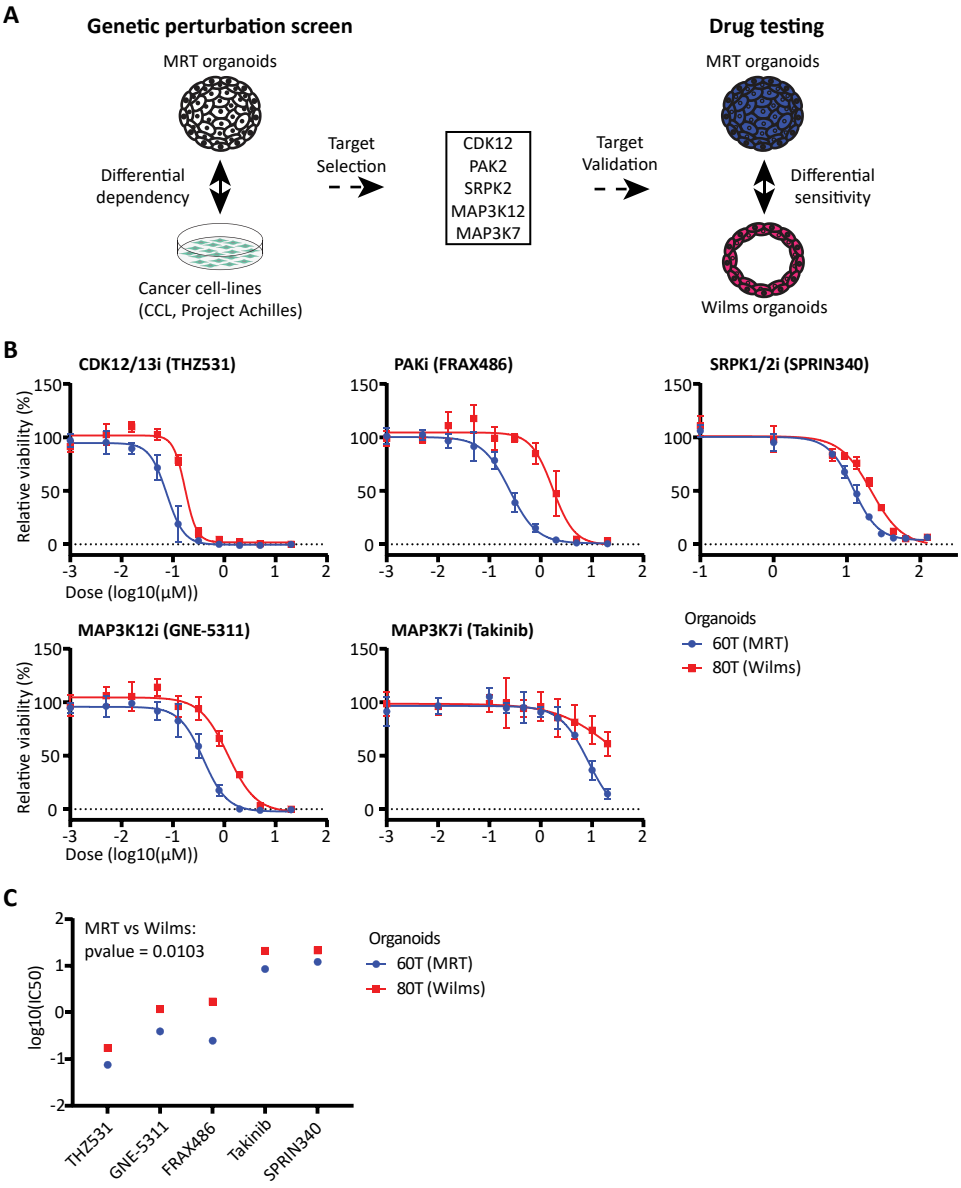


Figure 4 | Pharmacological inhibition of kinase induces MRT-specific effects on cell growth.
(A) Schematic representation of drug target selection and validation. **(B)** Dose-response curves show cell viability relative to DMSO control of MRT (blue dot) and Wilms tumor (red square) organoids. Organoids were treated with inhibitors for CDK12/13 (THZ531), PAK (FRAX486), SRPK1/2 (SPRIN340), MAP3K12 (GNE-5311) and MAP3K7 (Takinib) for ten different drug concentrations. Each dot and error bar represent the mean and SD of 4 technical replicates. **(C)** Half maximal inhibitory concentrations (IC_{50}) were derived from dose-response curves for 60T (MRT, blue dot) and 80T (Wilms, red square). Drug efficacy was compared between 60T and 80T. P-value was calculated using a paired Student's t test (two-tailed).

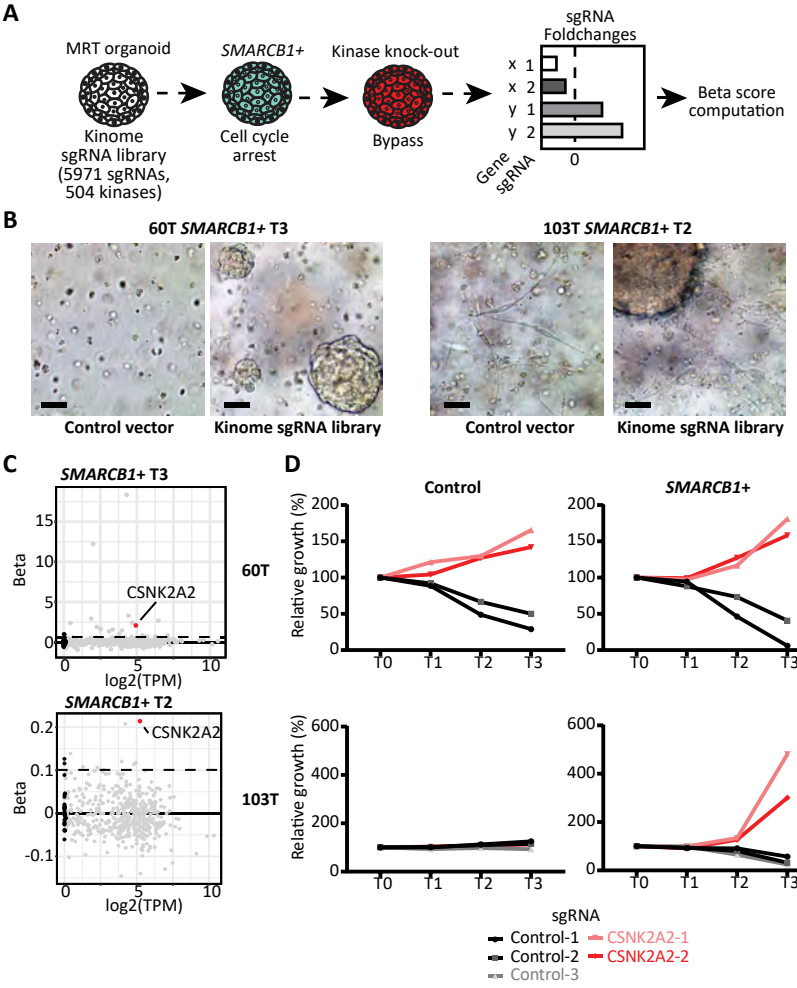


Figure 5 | CSNK2A2 is essential for the MRT cell cycle arrest induced by *SMARCB1*.
(A) Schematic visualization of the CRISPR-Cas9 knock-out screens in MRT *SMARCB1*+ organoids, focusing on the kinase perturbations that putatively cause a bypass of the *SMARCB1*-induced cell cycle arrest. **(B)** Representative bright-field images of MRT *SMARCB1*+ organoids transduced with the kinome sgRNA library or a control vector at timepoint 2 (T2) for 103T, or 3 (T3) for 60T. Scale bars denote 100µm. **(C)** Scatter plots show the correlation of gene expression levels and beta scores. The beta score cutoff (dashed line) used for gene identification was set at the 99th percentile of the beta score distribution of non-essential genes. Color indicates genes that are not expressed and therefore considered as non-essential (black) or genes that were identified as essential for the SMARCB1-induced cell cycle arrest in both 60T and 103T organoids (red). **(D)** Line graph that shows relative growth over time of 60T and 103T control and *SMARCB1*+ organoids transduced with a sgRNA targeting either the *AAVS1* control region (gray, n=3) or the *CSNK2A2* gene (red, n=2). Growth of transduced organoids was assessed by flow cytometry measurements of GFP positive (transduced with sgRNA) relative to GFP negative (untransduced) cells at 4 consecutive timepoints (T0-3). The GFP ratio at timepoint 0 was set at 100%.

Pharmacological inhibition of kinases induces MRT-specific effects on cell growth

To translate our genetic perturbation results to applicable therapeutic options, we set-up a small drug screen to interrogate the growth inhibitory effects of a selection of small-molecule inhibitors (**Fig. 4a**). The targets were selected based on MRT organoid specificity (**Fig. 2b**) and drug availability. Accordingly, we subjected the MRT organoids to pharmacological inhibition of SRPK1/2, which are regulators of the RNA splicing machinery, and CDK12, which is indicated to control *MYC* mRNA processing. Furthermore, we targeted the MAP kinase signaling cascade through inhibition of MAP3K7 and MAP3K12. And finally, we inhibited the PAK family, which are downstream effectors of Rho GTPases. As a control, we used a *SMARCB1* wild-type Wilms tumor organoid line[26], which was cultured under identical conditions. Cell viability was assessed after five days of incubation with the different drugs for a broad concentration range (ten doses), which allowed us to generate dose-response curves (**Fig. 4b**). When comparing dose-response curves of MRT and Wilms organoids, we observed a shift for each drug, which indicates differential drug efficacy. By quantification of the half maximal inhibitory concentrations (IC_{50}), we validated that MRT were consistently more sensitive to the selected inhibitors than Wilms tumor organoids (**Fig. 4c**). The largest therapeutic window could be observed upon PAK inhibition, with a drug efficacy of 85% in MRT vs 15% in Wilms organoids at a concentration of 792nM. Although the organoid cohort is small, the drug screen results indicate that the genetic perturbation effects observed in the MRT organoids can be predictive of drug response.

CSNK2A2 is essential for the MRT cell cycle exit induced by *SMARCB1*

Investigations into the *SMARCB1*-dependent cellular processes driving MRT differentiation can be exploited to uncover potential differentiation therapeutics, as previously described[5]. To uncover these pathways, we sought out to identify kinases that are essential for the *SMARCB1*-induced cell cycle exit in MRT cells. Therefore, we assessed the CRISPR-Cas9 knock-out screens performed in the MRT *SMARCB1*+ organoids for sgRNA enrichment (**Fig. 5a**). A bypass of the growth arrest as a consequence of kinase perturbation thus results in a positive beta score. This was recapitulated in the distribution of gene dependencies, which were biased towards a growth advantage for the MRT *SMARCB1*+ organoids (**Fig. 1b**). To improve assessment of sgRNA enrichment, we included a third timepoint (T3) for 60T (**Supplementary Fig. 1e**). Furthermore, we observed that within the pool of MRT cells that were transduced with the kinome sgRNA library, some clones were resistant to *SMARCB1*+ induced differentiation, and continued growth alike MRT control cells (**Fig. 5b**). Kinase perturbation effects were analyzed as for MRT control organoids, but now focusing on growth advantage. Accordingly, *CSNK2A2* knock-out was indicated to induce a growth advantage in both MRT *SMARCB1*+ organoid lines (**Fig. 5c**).

To independently validate these results, we set-up a growth assay that allows for direct comparison of growth rates of *CSNK2A2* wild-type and knock-out MRT organoids.

First, we lentivirally transduced the MRT cells with an inducible Cas9 expression construct. Next, we performed a second infection at an MOI of ~0.3 with a lentiviral vector that expresses both GFP and a sgRNA targeting *CSNK2A2*. Cas9 expression was induced for ~2 weeks. Accordingly, in the bulk cell population, the putative *CSNK2A2* knock-out cells could be distinguished from the wild-type by GFP expression. Then, we re-expressed *SMARCB1* by lentiviral transduction, which was the starting point (T0) of the assay. To establish growth rates, we quantified the proportion of GFP-positive cells over time by sequential flow cytometry measurements. In parallel, we used sgRNAs targeting a control genomic region (*AAVS1*) to establish the effect of gene independent Cas9 targeting. The growth assays showed that the MRT *SMARCB1*+ cells transduced with the sgRNAs (n=2) targeting *CSNK2A2* gained a growth advantage for both organoid lines (**Fig. 5d**), thereby reproducing the results derived from the CRISPR-Cas9 knock-out screens. The effect of *CSNK2A2* knock-out was *SMARCB1*-dependent in 103T, as the growth rate of 103T control cells were unchanged. In contrast, 60T showed a growth advantage in MRT cells with and without SMARCB1 reconstitution. These findings provide a starting point for further investigations into the role of casein kinase 2 in MRT cell cycle regulation and differentiation pathways.

Discussion

We studied MRT kinase dependencies by performing CRISPR-Cas9 knock-out screens in renal MRT organoids. Our investigations uncovered MRT-specific vulnerabilities that could be effectively targeted by small-molecule inhibitors. Additionally, we interrogated kinase function in MRT differentiation processes, which indicated casein kinase 2 as a putative regulator of *SMARCB1*-induced growth arrest.

Previous genetic perturbation screens in MRT cell-lines have identified multiple RTKs and CDK4/6 as MRT-specific vulnerabilities[19,46]. These findings were partly recapitulated in the MRT organoid models included in this study. For instance, we identified *RET* and *CDK6* as MRT-specific vulnerabilities in both patient lines. Additionally, the genetic perturbation of other RTKs, such as *PDGFRB* and *FGFR1*, induced patient-specific effects. This is in agreement with a previous report demonstrating that RTK dependency is variable across MRT cell-lines[19]. On the other hand, a comparison of MRT organoid and cell-line genetic screening results revealed that significant differences exist. This could relate to experimental factors (e.g. sgRNA library effectiveness) or to intrinsic differences of cell-lines and organoid models. Tumor organoids have been proposed as superior models for interrogating tumor vulnerabilities, as they more faithfully represent native tumor physiology and better predict drug response[25]. Furthermore, a direct comparison of 2D and 3D cultures subjected to CRISPR-Cas9 knock-out screens revealed that gene dependencies largely depend on culturing conditions[35]. Moreover, genetic perturbation effects in 3D cultures were indicated to better capture genes essential for *in vivo* tumor growth[35], which altogether promotes the use of tumor organoid models for genetic screening approaches. However, organoid-based CRISPR-Cas9 knock-out screens remain

challenging, as the handling of large cell numbers, which is required for genome-wide screens, is in general more complicated for organoid cultures.

A limitation of our study is that, so far, the number of MRT organoid lines included was small. Nevertheless, we found that the kinase expression profile of 60T and 103T, which strongly correlated with gene dependency, was recapitulated in three additional patient lines. Furthermore, by extending our analyses to MRT and SS cell-lines, we identified kinases regulating RNA splicing, *SRPK2* and *PRPF4B*, as a uniform MRT vulnerability. What remains to be investigated is how MRT, as a consequence of *SMARCB1* loss, gains such altered dependency on the spliceosome. Besides SWI/SNF function at gene promoters and enhancers, the complex can bind to gene bodies and act as a regulator of alternative splicing[47]. Therefore, we speculate that targeting of the spliceosome is synthetically lethal with aberrant SWI/SNF function. Tumor cell survival has appeared to be highly dependent on RNA splicing for a subset of cancer types[48], which has therefore developed as a promising therapeutic avenue, potentially also applicable to MRT.

A single genetic aberration, *SMARCB1* loss, causes a variety of MRT phenotypes with possibly different vulnerabilities[14], which complicates the development of therapy aimed to uniformly treat all MRT patients. Therefore, a subgroup-specific assessment of potential therapeutic targets may be required. Here, we studied ECRTs that are indicated to be driven by overexpression of *MYC*[41], which we demonstrated to be faithfully recapitulated in the renal MRT organoids in contrast to most MRT cell-lines. We identified *CDK12* as one of the most promising hits. Genetic perturbation induced strongest growth reduction in MRT and SS models with high *MYC* mRNA levels. *CDK12* belongs to the family of kinases that phosphorylate RNA polymerase II. This family further includes *CDK7* and *CDK9*, which were also identified as essential kinases in MRT organoids. However, *CDK7* and *CDK9* are common essential genes as they regulate general processes of transcription. Drugs targeting CDK7 and *CDK9* therefore arguably lack tumor specificity, which we confirmed with our analysis of differential dependency. In contrast, CDK12 is indicated to specifically regulate subsets of genes involved in processes such as DNA damage and stress response[49,50]. Furthermore, CDK12 is suggested to regulate pre-mRNA processing, as it associates with RNA splicing, polyadenylation and termination machinery[49,50]. Interestingly, the best studied example is its pre-mRNA processing function of the *MYC* gene[43]. Accordingly, pharmacological inhibition of CDK12 was shown to deplete MYC[44]. These findings propose a mechanism in which CDK12 inhibition indirectly reduces MYC levels and therefore emerges as a potential therapeutic target for MYC-driven MRT.

Furthermore, we identified MRT-specific dependencies for which inhibitors are unavailable up to date, such as *HIPK1* and *TLK2*. Interestingly, *HIPK1*, which was found to be *MYC*-dependent, is a co-repressor of homeodomain transcription factors, which are specifically upregulated in ECRT and ATRT-MYC[8,11]. In that perspective,

HIPK1 appears to be a promising therapeutic target, which upon further investigation may be considered for drug development.

In addition to our study of MRT vulnerabilities, we interrogated kinase essentiality for the *SMARCB1*-induced MRT differentiation phenotype. The genetic screens in MRT *SMARCB1*+ organoids identified a subunit of casein kinase 2 alpha (*CSNK2A2*), which has been described as a critical mediator of development and differentiation[51,52]. Interestingly, casein kinase 2 has been reported to phosphorylate SMARCA4, a core subunit of the SWI/SNF complex[53,54]. SMARCA4 phosphorylation appeared to be essential to modulate SWI/SNF subunit composition, which is required to induce neural differentiation[52,55]. Following that line of thought, we hypothesize that in addition to SMARCB1 reconstitution into the SWI/SNF complex, phosphorylation of SMARCA4 by casein kinase 2 is essential for MRT differentiation. Further investigations into the role of casein kinase 2 in MRT may provide valuable insights into the differentiation block induced by *SMARCB1* loss, that could potentially be exploited for development of maturation therapy.

Overall, we discovered MRT vulnerabilities using CRISPR-Cas9 knock-out screens in organoids. Further functional experiments that shed light on the molecular mechanisms are required to confirm their true potential as therapeutic target in MRT. Future work is also aimed at intensifying the genetic screening approach to a larger cohort of MRT organoid models and beyond the kinome, to increase our understanding of MRT dependencies, which is urgently needed for development of effective treatment options.

Methods

Organoid culture

MRT organoids were cultured as previously described[26]. For sub-culturing, MRT organoids were mechanically dissociated into single cells and seeded in growth factor-reduced BME (Trevigen). Organoid medium was composed of the following: Advanced DMEM/F12 (Gibco) containing 1x Glutamax, 10 mM HEPES, and antibiotics (AdDF+++), supplemented with 1.5% B27 supplement (Gibco), 10% R-spondin-conditioned medium, EGF (50 ng/ml, Peprotech), FGF-10 (100 ng/ml, Peprotech), N-acetylcysteine (1.25 mM, Sigma), Rho-kinase inhibitor Y-27632 (10 μM, Abmole), and A83-01 (5 μM, Tocris Bioscience).

CRISPR-Cas9 knock-out screen

MRT organoids were mechanically dissociated into single cells and lentivirally transduced[56] by spinfection with an established LentiCRISPRV2-puro kinome sgRNA library[28] at a multiplicity of infection (MOI) of ~0.3 to prevent multiple integrations per cell. As a control, transduction of the LentiCRISPRv2-puro construct without sgRNA sequence was performed in parallel. Transduced MRT cells were subjected to puromycin (InvivoGen, 0.5 μg/ml) selection two days after transduction. At timepoint 0 (T0), which was 7 days after transduction, cells were collected, and a sample was taken for genomic DNA isolation to establish sgRNA representation at the starting point of the genetic screen. The remaining cells were transduced with either pLKO.1-UbC-luciferase-blast[57] or pLKO.1-UbC-hSMARCB1-blast[5] lentiviruses. Transductions were performed n=3 times semi-independently for each condition, scaled to a minimum of 500 cells per sgRNA (n=5971) to ensure optimal library coverage. MRT control and *SMARCB1*+ organoids were subjected to blasticidin (InvivoGen, 10 μg/ml) selection two days after transduction, and subsequently passaged at T1 (10 days post T0) and T2 (17 days post T0). For 60T, an additional timepoint (T3) was included (23 days post T0). For each timepoint, samples were collected for gDNA isolation. For sub-culturing steps, library complexity was maintained at 500 cells per sgRNA. Genomic DNA was isolated from the collected samples using the ReliaPrep gDNA Tissue Miniprep System (Promega) according to manufacturer's protocol. sgRNA sequences were amplified from genomic DNA by standard PCR protocols and subjected to next-generation sequencing on the Illumina NextSeq500 platform.

Analysis of gene essentiality in MRT organoids

The sgRNA sequencing reads generated from the CRISPR-Cas9 knock-out screens were analyzed using the MAGeCK algorithm (version 0.5.8.1), following standard analysis pipelines[29]. First, raw sequencing reads were aligned to a custom sgRNA reference file using "mageck count" with default parameters. Quality control was performed for each sample, considering the number of reads mapped (>3 million reads per sample), the number of zero sgRNA counts and the Gini index which measures the evenness of sgRNA read counts. Samples with >1000 zero count sgRNAs were excluded for

analysis (60T *SMARCB1*+ T3, replicate 1 and 3). The sgRNA count table was used as input for "mageck mle" which identifies essential genes using maximum likelihood estimation. In short, non-targeting control sgRNAs were used to normalize counts and establish the null distribution. Principal component analysis was performed on the replicates using normalized counts to reveal potential outliers. sgRNA fold changes were generated by comparing each sample to the starting point T0. After removal of sgRNA outliers, sgRNA efficacy was estimated, which was incorporated in the final analysis. Gene dependency was assessed by analysis of sgRNA fold changes for each gene to generate beta scores, which indicate a growth advantage (+) or disadvantage (-) upon gene knock-out. Beta scores were normalized for copy number variations (CNV), which have been reported to bias CRISPR-Cas9 screening results[58]. CNV profiles were previously established by whole-genome sequencing of 60T and 103T organoids[26]. P-values were calculated using either a permutation based or Wald statistical test and were corrected for multiple-testing to establish false discovery rate (FDR).

Gene-independent toxicity of Cas9-induced double strand breaks was established by assessment of beta scores of non-expressed genes (TPM < 0.1), which we assumed to be non-essential, as previously reported[17,29]. Beta scores were normalized, so that the median of non-essential genes was zero. For identification of essential kinases, statistical significance and beta scores were considered. The distribution of beta scores for non-essential genes was used to establish cut-offs for FDR (60T Wald FDR < 0.25; 103T Wald FDR < 1e-05) and beta score (set at 10th percentile of non-essential gene distribution: 60T -0.43; 103T -0.16). For assessment of sgRNA enrichment in MRT *SMARCB1*+ organoids, cut-offs were set for FDR<0.1 and beta score (99th percentile of non-essential gene distribution) >0.62 (60T) and >0.10 (103T). The kinases identified as essential in both 60T and 103T were used for subsequent analysis.

Protein-protein interaction map

The physical and functional interactions of essential kinases were studied using the STRING protein-protein association network webtool (https://string-db.org/)[59] and were visualized using Cytoscape (version 3.8.2). Functional annotation was performed manually using the GO-term enrichment analysis incorporated in the STRING webtool.

Differential dependency scores

To assess tumor-specific vulnerabilities, we generated differential dependency scores across tumor models as described[17]. We combined MRT organoid and cancer cell-line (Project Achilles[33]) genetic screening data. Gene dependency scores were scaled within each tumor model (mean=0, SD=1) to account for confounder effects (e.g. proliferation rate). For each kinase, a differential dependency score was generated by calculating the deviation from the average gene dependency score. This deviation was subsequently scaled linearly using the SD of the full dataset. For each kinase, differential dependency was assessed relative to other kinases by applying z score

transformation (mean=0, SD=1). MRT organoid-specific vulnerabilities were identified by unsupervised hierarchical clustering (pheatmap version 1.0.12) of essential kinases using differential dependency scores.

Similarity of MRT organoid gene dependency to cancer cell-lines

A similarity score matrix was generated by inter-comparison of all MRT organoid and cancer cell-line gene dependencies using the R package MAGeCKFlute and command "ResembleDepMap". Similarity scores were scaled to assess relative similarity, and in such a way normalize for factors that bias similarity score (e.g. quality of genetic screen data). The relative similarity score of 60T and 103T to different tumor subtypes was assessed, and the median similarity score of each subtype was indicated. Tumor subtypes with <5 cancer cell-lines were excluded for analysis. Pearson's correlation coefficients were generated for MRT/SS organoids and cell-lines, which was used as input for principal component analysis.

Analysis of kinase mRNA expression levels

For assessment of gene expression levels of pediatric renal tumor and normal kidney organoids we used mRNA-seq data generated from a renal organoid biobank[26]. For cancer cell-lines, mRNA-seq data was derived from the DepMap portal[33]. Normal tissue gene expression data was derived from the GTEx portal (https://gtexportal.org).

Analysis of DNA methylation data

DNA methylation profiling of MRT organoids was performed previously[5]. For MRT subgroup classification, IDAT files were uploaded to the Brain Tumor Methylation Profiler version v11b4 (https://www.molecularneuropathology.org/mnp), which provided a classifier score for each organoid line. A classifier score >0.9 ensures a classification with high confidence.

Drug screening

Organoids were harvested and filtered using a 70 µm cell strainer (Falcon) and subsequently resuspended in medium with 5% BME. Then, organoids were seeded using the Multi-drop™ Combi Reagent Dispenser on black 384-well plates (Corning, ~500 organoids / well). Drugs were added using the Tecan D300e Digital Dispenser. Four technical replicates were included in each experiment. Five days after adding drugs, cell viability was measured using CellTiter-Glo 3D reagent (Promega) according to manufacturer's instructions. Results were normalized to the DMSO vehicle (100%). Dose response curves and IC_{50} values were generated by nonlinear regression (curve fit) using GraphPad Prism v8.0.2.

Compounds used for screening: THZ531 (MCE, HY-103618), GNE-3511 (MCE, HY-12947), FRAX486 (MCE, HY-15542B), Takinib (MCE, HY-103490), SRPIN340 (MCE, HY-13949).

Growth assays

To validate genetic perturbation effects of *CSNK2A2* in MRT *SMARCB1+* organoids we performed growth assays as follows. First, MRT organoids were lentivirally transduced with a doxycyclin-inducible pCW-iCas9-puro construct, and selected with puromycin (InvivoGen, 0.5 µg/ml) two days after transduction. Next, organoids were transduced with pLentiGuide-GFP-Hygro lentiviruses aiming for an MOI of ~0.3. The pLentiGuide-GFP-Hygro construct contained one of the following sgRNA sequences: AAVS1-Control-1 (TGGGACCACCTTATATTCCC), AAVS1-Control-2 (GAGCCACATTAACCGGCCCT), AAVS1-Control-3 (ATTCCCAGGGCCGGTTAATG), CSNK2A2-1 (ACGAGGCTCACGTCCCGAGC), CSNK2A2-2 (CAGGGCCCGGGTCTACGCCG). No antibiotic selection was performed for these cells, to allow for a direct comparison of growth rates of wild-type (GFP-negative) and putative *CSNK2A2* knock-out (GFP-positive) cells. Upon transduction, doxycycline (1 µg/ml) was added to the medium to induce Cas9 expression. Then, 14 days after transduction, a sample of the MRT cells was analyzed using flow cytometry, measuring the proportion of GFP-positive cells, which was the starting point of the growth assay (T0). The remaining cells were transduced with either pLKO.1-UbC-luciferase-blast[57] or pLKO.1-UbC-hSMARCB1-blast[5] lentiviruses. At this point, doxycycline was removed from the medium. MRT control and *SMARCB1+* organoids were subjected to blasticidin (InvivoGen, 10 µg/ml) selection two days after transduction. For three subsequent timepoints (T1-T3) MRT organoids were passaged, and the proportion of GFP-positive cells was measured by flow cytometry for each timepoint to establish growth rates of wild-type versus putative *CSNK2A2* knock-out cells.

References

1. Felsher, D. W. Cancer revoked: oncogenes as therapeutic targets. *Nat Rev Cancer* **3**, 375-380, doi:10.1038/nrc1070 (2003).

2. Brennan, B., Stiller, C. & Bourdeaut, F. Extracranial rhabdoid tumours: what we have learned so far and future directions. *Lancet Oncol* **14**, e329-336, doi:10.1016/S1470-2045(13)70088-3 (2013).

3. Fruhwald, M. C., Biegel, J. A., Bourdeaut, F., Roberts, C. W. & Chi, S. N. Atypical teratoid/rhabdoid tumors-current concepts, advances in biology, and potential future therapies. *Neuro Oncol* **18**, 764-778, doi:10.1093/neuonc/nov264 (2016).

4. Hoffman, L. M. *et al.* Advancing biology-based therapeutic approaches for atypical teratoid rhabdoid tumors. *Neuro Oncol* **22**, 944-954, doi:10.1093/neuonc/noaa046 (2020).

5. Custers, L. *et al.* Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nat Commun* **12**, 1407, doi:10.1038/s41467-021-21675-6 (2021).

6. Han, Z. Y. *et al.* The occurrence of intracranial rhabdoid tumours in mice depends on temporal control of Smarcb1 inactivation. *Nat Commun* **7**, 10421, doi:10.1038/ncomms10421 (2016).

7. Vitte, J., Gao, F., Coppola, G., Judkins, A. R. & Giovannini, M. Timing of Smarcb1 and Nf2 inactivation determines schwannoma versus rhabdoid tumor development. *Nat Commun* **8**, 300, doi:10.1038/s41467-017-00346-5 (2017).

8. Chun, H. E. *et al.* Identification and Analyses of Extra-Cranial and Cranial Rhabdoid Tumor Molecular Subgroups Reveal Tumors with Cytotoxic T Cell Infiltration. *Cell Rep* **29**, 2338-2354 e2337, doi:10.1016/j.celrep.2019.10.013 (2019).

9. Hasselblatt, M. *et al.* High-resolution genomic analysis suggests the absence of recurrent genomic alterations other than SMARCB1 aberrations in atypical teratoid/rhabdoid tumors. *Genes Chromosomes Cancer* **52**, 185-190, doi:10.1002/gcc.22018 (2013).

10. Hasselblatt, M. *et al.* SMARCA4-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. *Acta Neuropathol* **128**, 453-456, doi:10.1007/s00401-014-1323-x (2014).

11. Johann, P. D. *et al.* Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes. *Cancer Cell* **29**, 379-393, doi:10.1016/j.ccell.2016.02.001 (2016).

12. Lee, R. S. *et al.* A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. *J Clin Invest* **122**, 2983-2988, doi:10.1172/JCI64400 (2012).

13. Versteege, I. *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394**, 203-206, doi:10.1038/28212 (1998).

14. Torchia, J. *et al.* Integrated (epi)-Genomic Analyses Identify Subgroup-Specific Therapeutic Targets in CNS Rhabdoid Tumors. *Cancer Cell* **30**, 891-908, doi:10.1016/j.ccell.2016.11.003 (2016).

15. Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* **10**, 188-195, doi:10.1038/nrd3368 (2011).

16. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87, doi:10.1126/science.1247005 (2014).

17. Dempster, J. M. *et al.* Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. *bioRxiv*, 720243, doi:10.1101/720243 (2019).

18. Howard, T. P. *et al.* MDM2 and MDM4 Are Therapeutic Vulnerabilities in Malignant Rhabdoid Tumors. *Cancer Res* **79**, 2404-2414, doi:10.1158/0008-5472.CAN-18-3066 (2019).

19. Oberlick, E. M. *et al.* Small-Molecule and CRISPR Screening Converge to Reveal Receptor Tyrosine Kinase Dependencies in Pediatric Rhabdoid Tumors. *Cell Rep* **28**, 2331-2344 e2338, doi:10.1016/j.celrep.2019.07.021 (2019).

20. Wang, X. *et al.* BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. *Nat Commun* **10**, 1881, doi:10.1038/s41467-019-09891-7 (2019).

21. Michel, B. C. *et al.* A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol* **20**, 1410-1420, doi:10.1038/s41556-018-0221-1 (2018).

22. Kamb, A. What's wrong with our cancer models? *Nat Rev Drug Discov* **4**, 161-165, doi:10.1038/nrd1635 (2005).

23. Gillet, J. P., Varma, S. & Gottesman, M. M. The clinical relevance of cancer cell lines. *J Natl Cancer Inst* **105**, 452-458, doi:10.1093/jnci/djt007 (2013).

24. Clevers, H. Modeling Development and Disease with Organoids. *Cell* **165**, 1586-1597, doi:10.1016/j.cell.2016.05.082 (2016).

25. Drost, J. & Clevers, H. Organoids in cancer research. *Nat Rev Cancer* **18**, 407-418, doi:10.1038/s41568-018-0007-6 (2018).

26. Calandrini, C. *et al.* An organoid biobank for childhood kidney cancers that captures disease and tissue heterogeneity. *Nat Commun* **11**, 1310, doi:10.1038/s41467-020-15155-6 (2020).

27. Darr, J., Klochendler, A., Isaac, S., Geiger, T. & Eden, A. Phosphoproteomic analysis reveals Smarcb1 dependent EGFR signaling in Malignant Rhabdoid tumor cells. *Mol Cancer* **14**, 167, doi:10.1186/s12943-015-0439-5 (2015).

28. Wang, C. *et al.* CDK12 inhibition mediates DNA damage and is synergistic with sorafenib treatment in hepatocellular carcinoma. *Gut* **69**, 727-736, doi:10.1136/gutjnl-2019-318506 (2020).

29. Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc* **14**, 756-780, doi:10.1038/s41596-018-0113-7 (2019).

30. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e516, doi:10.1016/j.cell.2017.06.010 (2017).

31. Meel, M. H. *et al.* MEK/MELK inhibition and blood-brain barrier deficiencies in atypical teratoid/rhabdoid tumors. *Neuro Oncol* **22**, 58-69, doi:10.1093/neuonc/noz151 (2020).

32. Weingart, M. F. *et al.* Disrupting LIN28 in atypical teratoid rhabdoid tumors reveals the importance of the mitogen activated protein kinase pathway as a therapeutic target. *Oncotarget* **6**, 3165-3177, doi:10.18632/oncotarget.3078 (2015).

33. Broad. DepMap 20Q4 Public. *figshare*, doi:https://doi.org/10.6084/m9.figshare.13237076.v4 (2020).

34. Kadoch, C. & Crabtree, G. R. Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell* **153**, 71-85, doi:10.1016/j.cell.2013.02.036 (2013).

35. Han, K. *et al.* CRISPR screens in cancer spheroids identify 3D growth-specific vulnerabilities. *Nature* **580**, 136-141, doi:10.1038/s41586-020-2099-x (2020).

36. Carugo, A. *et al.* p53 Is a Master Regulator of Proteostasis in SMARCB1-Deficient Malignant Rhabdoid Tumors. *Cancer Cell* **35**, 204-220 e209, doi:10.1016/j.ccell.2019.01.006 (2019).

37. Lee, S., Cimica, V., Ramachandra, N., Zagzag, D. & Kalpana, G. V. Aurora A is a repressed effector target of the chromatin remodeling protein INI1/hSNF5 required for rhabdoid tumor cell survival. *Cancer Res* **71**, 3225-3235, doi:10.1158/0008-5472.CAN-10-2167 (2011).

38. Morin, A. *et al.* Proteasome inhibition as a therapeutic approach in atypical teratoid/rhabdoid tumors. *Neurooncol Adv* **2**, vdaa051, doi:10.1093/noajnl/vdaa051 (2020).

39. Tran, H. M. *et al.* Upregulation of Protein Synthesis and Proteasome Degradation Confers Sensitivity to Proteasome Inhibitor Bortezomib in Myc-Atypical Teratoid/Rhabdoid Tumors. *Cancers (Basel)* **12**, doi:10.3390/cancers12030752 (2020).

40. Corkery, D. P., Holly, A. C., Lahsaee, S. & Dellaire, G. Connecting the speckles: Splicing kinases and their role in tumorigenesis and treatment response. *Nucleus* **6**, 279-288, doi:10.1080/19491034.2015.1062194 (2015).

41. Alimova, I. *et al.* Inhibition of MYC attenuates tumor cell self-renewal and promotes senescence in SMARCB1-deficient Group 2 atypical teratoid rhabdoid tumors to suppress tumor growth in vivo. *Int J Cancer* **144**, 1983-1995, doi:10.1002/ijc.31873 (2019).

42. Chen, H., Liu, H. & Qing, G. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduct Target Ther* **3**, 5, doi:10.1038/s41392-018-0008-7 (2018).

43. Davidson, L., Muniz, L. & West, S. 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev* **28**, 342-356, doi:10.1101/gad.231274.113 (2014).

44. Zeng, M. *et al.* Targeting MYC dependency in ovarian cancer through inhibition of CDK7 and CDK12/13. *Elife* **7**, doi:10.7554/eLife.39030 (2018).

45. Young, M. D. et al. Single cell derived mRNA signals across human kidney tumors. Nat Commun 12, 3896, doi:10.1038/s41467-021-23949-5 (2021).

46. Merk, D. J. *et al.* Genome-wide CRISPR and small-molecule screens uncover targetable dependencies in ATRT. *bioRxiv*, 2020.2012.2009.417378, doi:10.1101/2020.12.09.417378 (2020).

47. Batsche, E., Yaniv, M. & Muchardt, C. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol* **13**, 22-29, doi:10.1038/nsmb1030 (2006).

48. Lee, S. C. & Abdel-Wahab, O. Therapeutic targeting of splicing in cancer. *Nat Med* **22**, 976-986, doi:10.1038/nm.4165 (2016).

49. Liu, H., Liu, K. & Dong, Z. Targeting CDK12 for Cancer Therapy: Function, Mechanism, and Drug Discovery. *Cancer Res* **81**, 18-26, doi:10.1158/0008-5472.CAN-20-2245 (2021).

50. Lui, G. Y. L., Grandori, C. & Kemp, C. J. CDK12: an emerging therapeutic target for cancer. *J Clin Pathol* **71**, 957-962, doi:10.1136/jclinpath-2018-205356 (2018).

51. Gotz, C. & Montenarh, M. Protein kinase CK2 in development and differentiation. *Biomed Rep* **6**, 127-133, doi:10.3892/br.2016.829 (2017).

52. Padilla-Benavides, T., Reyes-Gutierrez, P. & Imbalzano, A. N. Regulation of the Mammalian SWI/SNF Family of Chromatin Remodeling Enzymes by Phosphorylation during Myogenesis. *Biology (Basel)* **9**, doi:10.3390/biology9070152 (2020).

53. Padilla-Benavides, T. *et al.* CK2-Dependent Phosphorylation of the Brg1 Chromatin Remodeling Enzyme Occurs during Mitosis. *Int J Mol Sci* **21**, doi:10.3390/ijms21030923 (2020).

54. Padilla-Benavides, T. *et al.* Casein kinase 2-mediated phosphorylation of Brahma-related gene 1 controls myoblast proliferation and contributes to SWI/SNF complex composition. *J Biol Chem* **292**, 18592-18607, doi:10.1074/jbc.M117.799676 (2017).

55. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* **106**, 5181-5186, doi:10.1073/pnas.0812889106 (2009).

56. Koo, B. K. *et al.* Controlled gene expression in primary Lgr5 organoid cultures. *Nat Methods* **9**, 81-83, doi:10.1038/nmeth.1802 (2011).

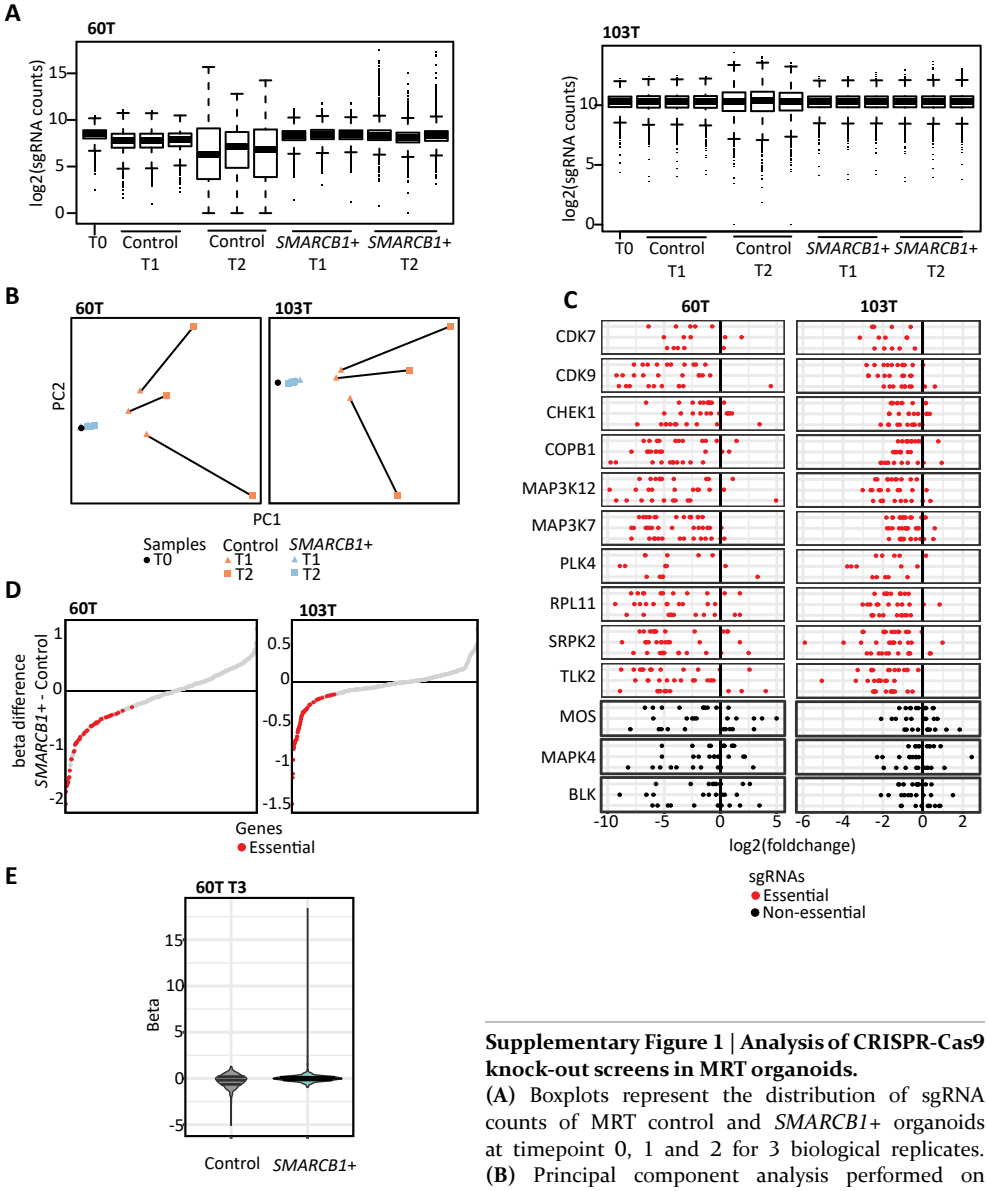
57. Fumagalli, A. *et al.* Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proc Natl Acad Sci U S A* **114**, E2357-E2364, doi:10.1073/pnas.1701219114 (2017).

58. Aguirre, A. J. *et al.* Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov* **6**, 914-929, doi:10.1158/2159-8290.CD-16-0154 (2016).

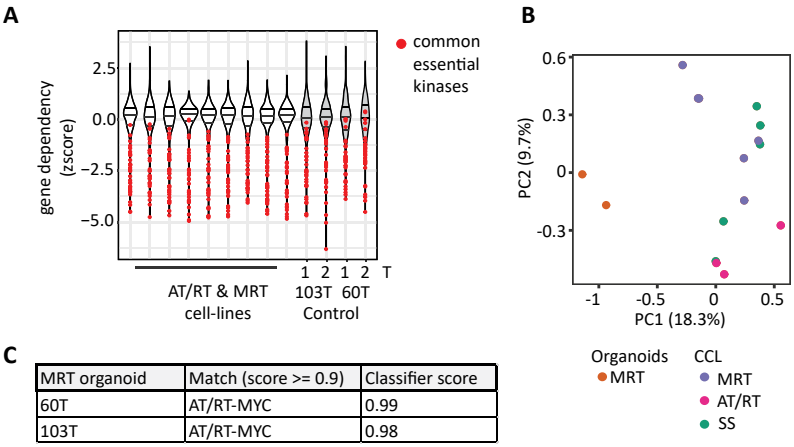
59. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613, doi:10.1093/nar/gky1131 (2019).

60. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469-474, doi:10.1038/nature26000 (2018).

Supplementary information



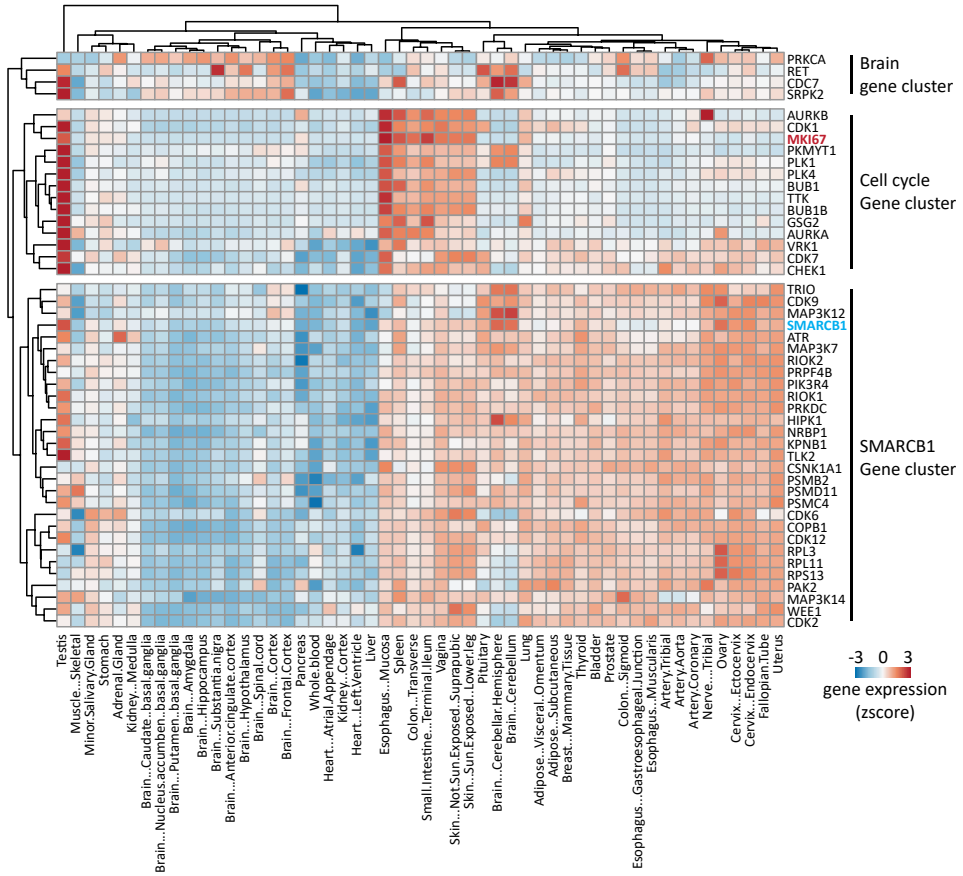
Supplementary Figure 1 | Analysis of CRISPR-Cas9 knock-out screens in MRT organoids.
(A) Boxplots represent the distribution of sgRNA counts of MRT control and *SMARCB1+* organoids at timepoint 0, 1 and 2 for 3 biological replicates. **(B)** Principal component analysis performed on normalized sgRNA counts of the T0 sample (black), MRT control (orange) and *SMARCB1+* (blue) organoids at timepoint 1 (triangle) and 2 (square) for 3 biological replicates. Samples from the same biological replicate are connected by line. **(C)** Dot plot shows sgRNA fold changes of MRT organoids at T2 for a random selection of kinases identified as essential (Fig. 1, n=10, red) or non-essential (n=3, black). **(D)** Dot plot represents beta score differences for MRT control and *SMARCB1+* organoids for each kinase. Kinases are ranked according to the beta difference. Kinases identified as essential in MRT organoids (Fig. 1) are indicated in red. **(E)** Violin plot shows the distribution of normalized gene dependency scores (beta) derived from the CRISPR-Cas9 knock-out screens performed in 60T control and *SMARCB1+* organoids at timepoint 3 (T3).



C

MRT organoid	Match (score >= 0.9)	Classifier score
60T	AT/RT-MYC	0.99
103T	AT/RT-MYC	0.98

Supplementary Figure 2 | Comparison of MRT organoids and cancer cell-lines.
(A) Violin plot shows gene dependency scores of MRT organoids (timepoint 1 and 2) and cell-lines (Project Achilles) that were scaled by tumor model to normalize for confounder effects such as proliferation rate. Common essential genes established from the DepMap genetic screening dataset are indicated in red. **(B)** Principal component analysis performed on gene dependency correlations between MRT/SS organoids and cell-lines. Principal component 1 and 2 are visualized, variance proportions are indicated for each component. Color depicts tumor subtype. **(C)** DNA methylation profiles[5] of 60T and 103T organoids were subjected to the Brain Tumor Methylation Profiler[60] and were classified as the AT/RT-MYC subgroup with high confidence.



Supplementary Figure 3 | Kinase expression patterns across organoids and tissues.
Heatmap represents scaled gene expression values of kinases identified as essential in MRT organoids (Fig. 1) across normal tissues (GTEx median TPM). Genes are classified into three groups by unsupervised hierarchical clustering. Cluster 1 is termed the brain gene cluster, as the kinases are highly expressed in brain tissues. Cluster 2 includes cell-cycle associated genes, as they correlate with proliferation marker *KI67* (red). Genes in cluster 3 correlate with *SMARCB1* (blue) expression levels.

Supplementary Table 1 | Essential kinases identified in 60T and 103T organoids.

Symbol	Gene Name	Beta 60T	Beta 103T	Sig. Gene
AAK1	AP2 associated kinase 1(AAK1)	-0.09	-0.18	103T
ABL2	ABL proto-oncogene 2, non-receptor tyrosine kinase(ABL2)	-0.41	-0.17	103T
ADCK5	aarF domain containing kinase 5(ADCK5)	0.59	-0.24	103T
AKT1	AKT serine/threonine kinase 1(AKT1)	-0.24	-0.37	103T
AKT2	AKT serine/threonine kinase 2(AKT2)	0.33	-0.33	103T
ATR	ATR serine/threonine kinase(ATR)	-0.84	-0.25	60T & 103T
AURKA	aurora kinase A(AURKA)	-1.79	-0.89	60T & 103T
AURKB	aurora kinase B(AURKB)	-0.74	-0.66	60T & 103T
BMPR2	bone morphogenetic protein receptor type 2(BMPR2)	0.05	-0.43	103T
BTK	Bruton tyrosine kinase(BTK)	0.45	-0.26	103T
BUB1	BUB1 mitotic checkpoint serine/threonine kinase(BUB1)	-1.62	-0.46	60T & 103T
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B(BUB1B)	-1.18	-0.81	60T & 103T
CAMK2G	calcium/calmodulin dependent protein kinase II gamma(CAMK2G)	-0.23	-0.31	103T
CAMK4	calcium/calmodulin dependent protein kinase IV(CAMK4)	-0.12	-0.19	103T
CDC7	cell division cycle 7(CDC7)	-1.75	-1.05	60T & 103T
CDK1	cyclin dependent kinase 1(CDK1)	-1.14	-0.79	60T & 103T
CDK11A	cyclin dependent kinase 11A(CDK11A)	-0.24	-0.30	103T
CDK12	cyclin dependent kinase 12(CDK12)	-1.15	-0.77	60T & 103T
CDK2	cyclin dependent kinase 2(CDK2)	-1.86	-0.36	60T & 103T
CDK6	cyclin dependent kinase 6(CDK6)	-1.42	-0.51	60T & 103T
CDK7	cyclin dependent kinase 7(CDK7)	-1.38	-0.54	60T & 103T
CDK9	cyclin dependent kinase 9(CDK9)	-0.94	-0.60	60T & 103T
CHEK1	checkpoint kinase 1(CHEK1)	-0.79	-0.92	60T & 103T
COPB1	coatomer protein complex subunit beta 1(COPB1)	-0.82	-0.80	60T & 103T
CSK	c-src tyrosine kinase(CSK)	-0.38	-0.38	103T
CSNK1A1	casein kinase 1 alpha 1(CSNK1A1)	-0.98	-0.21	60T & 103T
DSTYK	dual serine/threonine and tyrosine protein kinase(DSTYK)	0.08	-0.21	103T
DYRK1A	dual specificity tyrosine phosphorylation regulated kinase 1A(DYRK1A)	0.55	-0.26	103T
FES	FES proto-oncogene, tyrosine kinase(FES)	0.07	-0.21	103T
FGFR1	fibroblast growth factor receptor 1(FGFR1)	0.02	-0.22	103T
FGFR3	fibroblast growth factor receptor 3(FGFR3)	-0.21	-0.20	103T
FGR	FGR proto-oncogene, Src family tyrosine kinase(FGR)	-0.24	-0.24	103T
GAK	cyclin G associated kinase(GAK)	-0.23	-0.57	103T
GRK6	G protein-coupled receptor kinase 6(GRK6)	0.11	-0.22	103T
GSG2	germ cell associated 2, haspin(GSG2)	-0.58	-0.28	60T & 103T
GUCY2F	guanylate cyclase 2F, retinal(GUCY2F)	-0.39	-0.49	103T
HIPK1	homeodomain interacting protein kinase 1(HIPK1)	-0.59	-0.38	60T & 103T
KPNB1	karyopherin subunit beta 1(KPNB1)	-1.97	-0.79	60T & 103T
LCK	LCK proto-oncogene, Src family tyrosine kinase(LCK)	-0.04	-0.20	103T
MAP3K12	mitogen-activated protein kinase kinase kinase 12(MAP3K12)	-1.62	-0.23	60T & 103T
MAP3K14	mitogen-activated protein kinase kinase kinase 14(MAP3K14)	-0.74	-0.23	60T & 103T
MAP3K19	mitogen-activated protein kinase kinase kinase 19(MAP3K19)	-0.11	-0.20	103T
MAP3K5	mitogen-activated protein kinase kinase kinase 5(MAP3K5)	0.29	-0.28	103T
MAP3K7	mitogen-activated protein kinase kinase kinase 7(MAP3K7)	-1.46	-0.28	60T & 103T
MAPK1	mitogen-activated protein kinase 1(MAPK1)	0.59	-0.21	103T

Symbol	Gene Name	Beta 60T	Beta 103T	Sig. Gene
MERTK	MER proto-oncogene, tyrosine kinase(MERTK)	0.20	-0.27	103T
NPR2	natriuretic peptide receptor 2(NPR2)	0.49	-0.24	103T
NRBP1	nuclear receptor binding protein 1(NRBP1)	-0.77	-0.33	60T & 103T
NUAK1	NUAK family kinase 1(NUAK1)	-0.33	-0.19	103T
OXSR1	oxidative stress responsive 1(OXSR1)	-0.17	-0.25	103T
PAK2	p21 (RAC1) activated kinase 2(PAK2)	-1.23	-0.29	60T & 103T
PDGFRB	platelet derived growth factor receptor beta(PDGFRB)	-0.18	-0.20	103T
PDK2	pyruvate dehydrogenase kinase 2(PDK2)	-0.07	-0.21	103T
PEAK1	pseudopodium enriched atypical kinase 1(PEAK1)	0.67	-0.22	103T
PIK3R4	phosphoinositide-3-kinase regulatory subunit 4(PIK3R4)	-0.56	-0.35	60T & 103T
PKMYT1	protein kinase, membrane associated tyrosine/threonine 1(PKMYT1)	-0.44	-0.22	60T & 103T
PLK1	polo like kinase 1(PLK1)	-0.73	-0.86	60T & 103T
PLK4	polo like kinase 4(PLK4)	-1.20	-0.67	60T & 103T
PRKCA	protein kinase C alpha(PRKCA)	-0.67	-0.19	60T & 103T
PRKCZ	protein kinase C zeta(PRKCZ)	-0.01	-0.35	103T
PRKDC	protein kinase, DNA-activated, catalytic polypeptide(PRKDC)	-1.03	-0.39	60T & 103T
PRPF4B	pre-mRNA processing factor 4B(PRPF4B)	-0.71	-0.26	60T & 103T
PSMB2	proteasome subunit beta 2(PSMB2)	-1.77	-1.59	60T & 103T
PSMC4	proteasome 26S subunit, ATPase 4(PSMC4)	-2.03	-0.81	60T & 103T
PSMD11	proteasome 26S subunit, non-ATPase 11(PSMD11)	-1.30	-0.80	60T & 103T
PSMD6	proteasome 26S subunit, non-ATPase 6(PSMD6)	-0.80	-0.59	103T
RET	ret proto-oncogene(RET)	-0.50	-0.24	60T & 103T
RIOK1	RIO kinase 1(RIOK1)	-0.63	-0.63	60T & 103T
RIOK2	RIO kinase 2(RIOK2)	-1.05	-0.48	60T & 103T
RPL11	ribosomal protein L11(RPL11)	-0.99	-0.86	60T & 103T
RPL3	ribosomal protein L3(RPL3)	-2.40	-1.19	60T & 103T
RPS13	ribosomal protein S13(RPS13)	-2.12	-1.28	60T & 103T
RPS6KB1	ribosomal protein S6 kinase B1(RPS6KB1)	-0.37	-0.20	103T
SBK2	SH3 domain binding kinase family member 2(SBK2)	-0.20	-0.19	103T
SCYL3	SCY1 like pseudokinase 3(SCYL3)	0.04	-0.21	103T
SIK1	salt inducible kinase 1(SIK1)	0.24	-0.17	103T
SRPK2	SRSF protein kinase 2(SRPK2)	-0.46	-0.24	60T & 103T
STK25	serine/threonine kinase 25(STK25)	-0.07	-0.27	103T
STK32C	serine/threonine kinase 32C(STK32C)	0.35	-0.24	103T
TBK1	TANK binding kinase 1(TBK1)	0.51	-0.34	103T
TESK1	testis-specific kinase 1(TESK1)	-0.39	-0.27	103T
TLK2	tousled like kinase 2(TLK2)	-1.16	-0.58	60T & 103T
TRIO	trio Rho guanine nucleotide exchange factor(TRIO)	-0.46	-0.22	60T & 103T
TRPM7	transient receptor potential cation channel subfamily M member 7(TRPM7)	0.36	-0.42	103T
TSSK1B	testis specific serine kinase 1B(TSSK1B)	0.00	-0.23	103T
TTBK1	tau tubulin kinase 1(TTBK1)	0.86	-0.22	103T
TTK	TTK protein kinase(TTK)	-0.73	-0.41	60T & 103T
VRK1	vaccinia related kinase 1(VRK1)	-0.48	-0.44	60T & 103T
WEE1	WEE1 G2 checkpoint kinase(WEE1)	-0.90	-0.57	60T & 103T
WNK1	WNK lysine deficient protein kinase 1(WNK1)	-0.34	-0.31	103T

CHAPTER 7

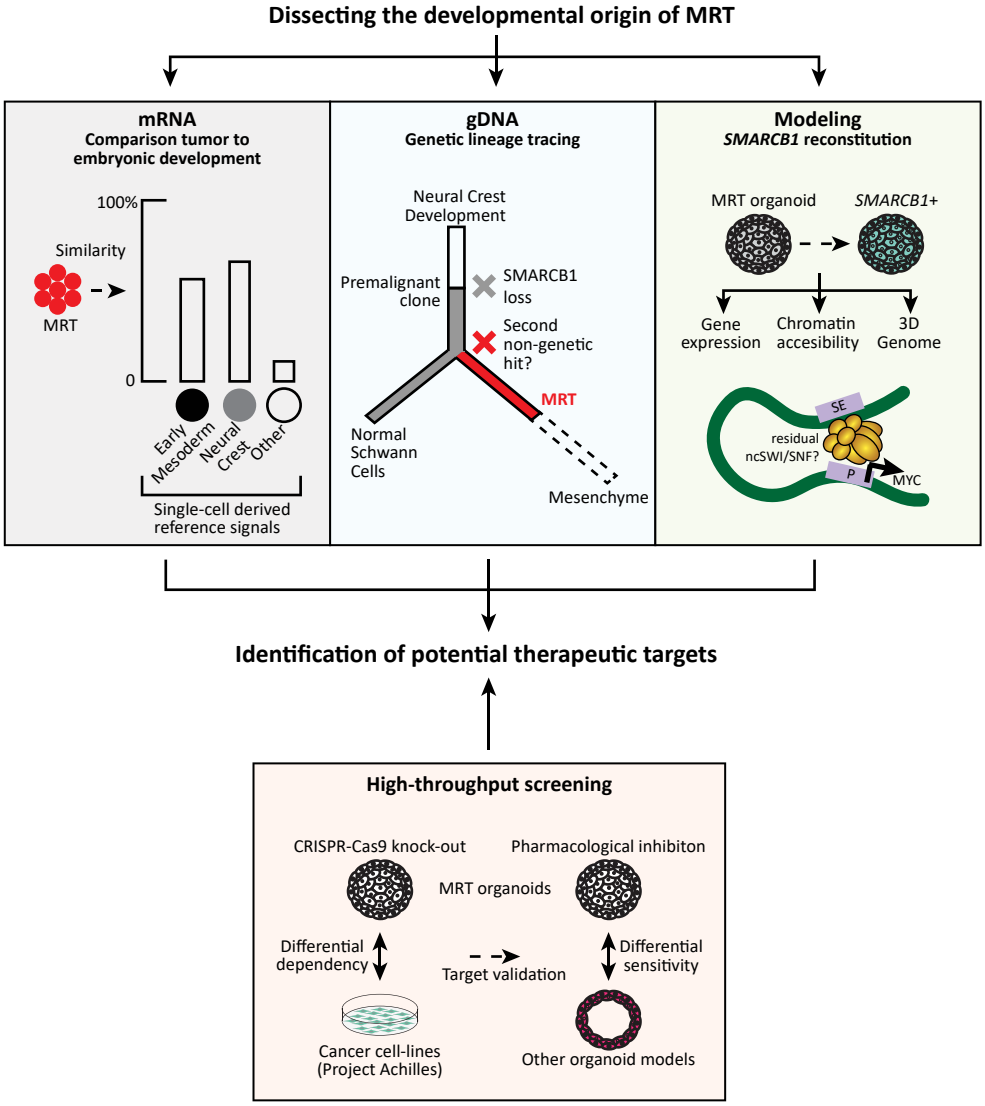
General discussion

Lars Custers[1,2]

[1]Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands

[2]Oncode Institute, the Netherlands

Graphical abstract Thesis



MRT = Malignant Rhabdoid Tumor
ncSWI/SNF = SMARCB1-independent non-canonical SWI/SNF chromatin remodeler complex
P = Gene promoter
SE = Super-enhancer

Pediatric tumor development

Most tumor entities that arise during childhood, never or rarely occur in adults, which complicates the adoption of treatment strategies from adult practice. Therefore, further improvement in the survival of children with cancer is likely to come from therapies specifically designed to target the unique properties of childhood cancer cells. A pediatric malignancy for which effective targeted treatment is currently lacking is malignant rhabdoid tumor (MRT, introduced in **Chapter 1**), which has been the primary focus of our research presented in this Thesis.

In general, cellular differentiation is thought to act as an obstacle for malignant transformation[1]. Accordingly, dedifferentiation has been implicated as one of the first steps of adult tumor progression, often induced by the initial genetic event[2]. Many pediatric tumors arise from embryonic progenitor cells[3,4], which arguably lack such differentiation barrier, and therefore seem to have an increased susceptibility to tumorigenesis. The immaturity of embryonic progenitor cells may contribute to a cellular context in which merely a single or a few genetic events are sufficient for tumor development. However, this permissible cellular context can be highly specific and distinct for each tumor entity[3,4]. The oncogenic effect of individual genetic events may be restricted to a particular cell type at a unique developmental time, which complicates the genetic modelling of embryonal tumors (discussed in **Chapter 2**[5]). Oncogenesis thus relies on a specific cooperation between the genetic event and its embryonic context.

Investigations into the intimate connection between childhood malignancies and embryonic development may identify the dysregulated developmental pathways at the root of tumorigenesis, which may give a novel perspective on childhood tumor vulnerabilities (discussed in **Chapter 2**[5]). For this reason, our studies aimed to identify the cellular origin of MRT and several other childhood malignancies as well as to define the SMARCB1-dependent differentiation pathways underpinning MRT development to uncover novel therapeutic options.

Exploring the origin of childhood cancer using gene expression signatures

Tumor transcriptomes may accommodate clues of the differentiation state and origin of human cancer, as tumor cells may inherit gene expression patterns from the cell they derive from[6-9]. Therefore, we interrogated the differentiation state of adult and childhood renal cancers by mapping bulk tumor transcriptomes to single-cell references of normal adult and fetal cells (**Chapter 3**[10]). Global analysis of tumor differentiation signals revealed that pediatric tumor transcriptomes exhibit features of fetal cells. Our findings indicated a consistent fetal signature across childhood renal cancers, which is supporting evidence of their aberrant developmental state. These fetal signals were, with exception, absent in adult malignancies, which challenges the hypothesis that embryonic reprogramming underlies adult tumorigenesis.

Our approach allows for a comparison of childhood cancer to embryonic development on a whole-transcriptome level, which provided quantitative signals defining renal cancer identity. However, evidence of transcriptional programs in tumor cells is merely correlative and does not necessarily prove cellular origin. Proto-oncogenes and tumor suppressors often have central roles in normal developmental processes. Genetic impairment of such genes can therefore alter tumor differentiation state in such a way that it is not representative of the cell-of-origin[3], which has been illustrated for rhabdomyosarcoma. Activation of Sonic hedgehog signaling in the adipocyte lineage in mice resulted in a skeletal muscle-like tumor that resembled patient rhabdomyosarcoma[11], which exemplifies how tumor phenotype can deviate from the tumor-initiating cell. This phenomenon may also apply to the clear cell sarcomas of the kidney (CCSKs), as we found that CCSK mRNA profiles did not faithfully map to kidney development or other single-cell references.

Furthermore, we found the transcriptional identity to be remarkably uniform within each tumor entity. This suggests that, although inter-tumor heterogeneity is significant on phenotypic, genetic, and epigenetic levels, there may be common over-arching fetal programs driving tumor growth. As these embryonic programs are unique to pediatric cancer cells, they may serve as a therapeutic vulnerability. Following that line of thought, we envision that genes that are essential for embryonic development but non-essential in adult cells, may prove to be effective therapeutic targets for childhood cancer. For instance, gene knockout experiments in mouse models demonstrated that the loss of putative pro-tumorigenic genes *Trim28* or *Twist1* results in embryonic lethality[12,13]. However, when specifically depleted in postnatal life, no obvious abnormal phenotypes exist[12,13], which may therefore indicate a therapeutic opportunity. In addition, the embryonic profile of pediatric tumors may be exploited for immunotherapy, as childhood cancer cells may harbor antigens unique to fetal tissues. For instance, the cell surface protein GD2 was found to be highly expressed in embryonal tumors, and rarely in adult cells[14]. Antibodies targeting GD2 have shown clinical efficacy in neuroblastoma, and may potentially be extended to other embryonal malignancies[15]. Altogether, these findings illustrate how a better definition of the fetal identity of pediatric cancer may aid the identification of novel therapeutic targets.

Furthermore, future work may move beyond mRNA and focus on comparison of tumor and normal cell epigenomes. DNA methylation patterns or gene regulatory landscapes of childhood cancer have been suggested to also derive from their fetal precursor cells[16,17]. A better definition of such programs can therefore provide additional insight on the developmental state of pediatric cancer.

Dissecting the embryonic origin of MRT

Our transcriptomic analysis of renal tumors demonstrated that MRT transcriptomes resemble specific cellular signals of early neural crest, but simultaneously exhibit similarity to mesodermal cell types (**Chapter 3**[10]). To further investigate this ambiguity, we conducted genetic lineage tracing experiments using somatic mutations to uncover patient MRT origin (**Chapter 4**)[18]. Somatic mutations are acquired each cell division, which act as genetic scars that are passed on from the ancestor cell to its progeny upon replication of its genomic DNA[19]. Analysis of shared somatic mutations between tumor and normal tissues can therefore faithfully define phylogenetic relations[20,21]. In contrast to mRNA, genomic DNA is not ambiguous. Somatic mutations are in principle unique and irreversible and can therefore serve as natural barcodes to trace back the origin of a (tumor) cell. Accordingly, we provided definitive proof that human MRTs are derived from the neural crest and thereby excluded a mesodermal origin.

The neural crest is a stream of embryonic cells migrating throughout the embryo to generate a diverse spectrum of cell types[22]. The multipotency of neural crest cells may relate to the heterogeneous transcriptional profiles (e.g. mesenchymal, neural, melanosomal) underlying MRT subgroups[23,24]. On the other hand, the transcriptomic identity of MRT may also be a product of grossly distorted gene regulation induced by *SMARCB1* loss. Following that line of thought, the mesenchymal profile that we identified in MRT transcriptomes may be a direct consequence of *SMARCB1* loss, and not a reflection of tumor origin. This hypothesis can only be tested by functional experiments that interrogate the role of *SMARCB1* in MRT development. Therefore, we reconstituted *SMARCB1* in MRT organoids to determine the effect of *SMARCB1* on tumor differentiation state (**Chapter 4**[18]). Re-expression of *SMARCB1* induced a bypass of the maturation block. We hypothesized that by normalization of *SMARCB1* levels, MRT cells resume the differentiation trajectory that had been halted during embryonic development and differentiate into the cell they were destined to become. We showed that MRT cell differentiation consistently characterized a neural to mesenchyme conversion. Accordingly, we concluded that MRT arises along the path of neural crest differentiation to mesenchymal cells, which explains the presence of both neural and mesenchymal gene expression signatures in MRT.

The differentiation of neural crest to mesenchyme is well studied in craniofacial development, where it generates part of the bone, cartilage and connective tissues that make up the skull[25]. It is conventionally thought that only cranial neural crest cells have the potency to form mesenchyme *in vivo*[26,27]. This concept challenges our indications that extracranial MRT arises from neural crest *en route* to becoming mesenchyme. However, an alternative source of neural crest-derived mesenchyme in the trunk has recently been discovered. Schwann cell precursors (SCPs) are neural crest-derived embryonic cells that migrate along developing peripheral nerves, consecutive to the neural crest migratory wave that occurs at an earlier embryonic timepoint[28]. Initially, SCPs were merely considered as the source of mature Schwann

cells, but further investigations demonstrated that SCPs encompass a multipotency similar to neural crest cells[28]. SCPs have been shown to contribute to embryogenesis by generating neuroendocrine cells, autonomic neurons, melanocytes and various mesenchymal populations[28]. Interestingly, our genetic lineage tracing experiments (**Chapter 4**)[18] showed that the tissue biopsies that were phylogenetically most closely related to MRT primarily consisted of Schwann cells, a derivative of SCPs. Altogether, these observations indicate that extracranial MRT may arise in neural crest-derived SCPs that are primed for mesenchymal differentiation.

The central nervous system variant of MRT, referred to as atypical teratoid/rhabdoid tumor (AT/RT), can be classified into subgroups based on epigenetic and transcriptomic profiles[23,24]. The different subgroups comprise distinct features of either mesenchyme, neurons, or melanocytes, which are all neural crest derivatives. Accordingly, these diverse tumor transcriptomes suggest that AT/RT subgroups may arise from distinct neural crest differentiation trajectories. To investigate AT/RT origin, we propose a similar approach as described in **Chapter 4**[18], which combines phylogenetic analysis of AT/RT tissues with *SMARCB1* reconstitution experiments in AT/RT organoids. Moreover, our study (**Chapter 4**[18]) defines a research strategy that may be more broadly applied to unravel the developmental root of childhood cancer.

MRT develops from a premalignant tissue bed

The evolutionary dynamics of embryonal cancers such as MRT are not fully understood. Tumor evolution does not seem to adhere to the slow and progressive accumulation of mutations that underlies the malignant transformation of adult cells[29]. Adult cancer development often initiates from a premalignant lesion, which is a clonal expansion of normal cells that have gained a proliferative advantage through genetic or epigenetic aberrations[30]. This precancerous tissue is benign but constitutes an increased risk of developing into an invasive cancer upon acquisition of additional driver events. To investigate if such premalignant lesions also exist for pediatric cancer, genomic characterization of the normal tissues from which the tumor develops is required. Accordingly, by genomic analysis of MRT and surrounding normal tissues (**Chapter 4**[18]), we revealed that morphologically normal Schwann cells harbored the same bi-allelic *SMARCB1* mutations as found in the tumor. Thus, *SMARCB1* loss appears to be essential but not sufficient for tumor initiation. In that perspective, the *SMARCB1*-null Schwann cell lineage would qualify as a premalignant clone that requires a second, likely non-genetic, hit for oncogenic progression. Such precursor cells have also been identified for Wilms tumor, which was shown to arise from aberrant clonal expansions of renal tissue[31]. The premalignant cells were characterized by an epigenetic event, hypermethylation of *H19*. Interestingly, in contrast to some adult premalignant tissues, both Wilms and MRT precursor cells appeared normal histologically, which may explain why precursor lesions have not been widely discovered yet across childhood cancers. Further research efforts are required to determine if premalignancies are a common phenomenon for pediatric tumors.

We consider two models of tumor progression that could potentially explain our findings, which have previously been posed by Chen *et al.*[3]. One, SMARCB1 loss epigenetically reprograms the cell-of-origin in a stochastic manner. Accordingly, tumor development only initiates from a clone in which all tumor suppressive epigenetic barriers are stochastically removed. Alternatively, SMARCB1 loss does not reorganize the epigenome of the cell-of-origin, but only exercises its tumorigenic effect when the *SMARCB1*-mutant embryonic cell acquires a tolerant epigenome in a specific developmental context. Both models would hypothetically result in non-malignant *SMARCB1*-null cells, as we have observed (**Chapter 4**[18]).

Altogether, our studies indicate that MRT develops from a *SMARCB1*-null premalignant tissue bed. We propose that there are other factors besides SMARCB1 loss that are required to initiate MRT development. Such requirements may include a specific micro-environment, or cell intrinsic factors such as developmental competence or epigenetic organization. A comprehensive epigenetic comparison of MRT versus *SMARCB1*-null Schwann cells would possibly allow us to define the putative epigenetic hits that follow *SMARCB1* loss.

Developmental pathways as drivers of MRT

It has become apparent that epigenetic dysregulation is at the core of many pediatric cancers[32]. As SMARCB1 is a subunit of the SWI/SNF chromatin remodeler, we hypothesized that the maturation block that underlies MRT development is established by aberrant chromatin remodeling. Therefore, we interrogated changes in chromatin topology by reconstituting *SMARCB1* in MRT organoids (**Chapter 5**), which demonstrated that SMARCB1 is essential for enhancer regulation. Furthermore, by assessment of the 3D genome, we identified a SMARCB1-dependent chromatin loop to putatively regulate oncogenic levels of *MYC*. To our interest, we found that the genomic regions that were characterized as *MYC* super-enhancers in MRT, exhibited features of typical enhancers specifically in neural crest cells. These findings may suggest that the *MYC* enhancer in MRT is not induced *de novo* upon *SMARCB1* loss, but rather upgraded in neural crest progenitors.

Investigations into neural crest development have illustrated that *MYC* is essential for maintaining multipotency and proliferation[33]. Developmental differentiation trajectories frequently involve pivotal roles for the tumor suppressors and proto-oncogenes that underly embryonal tumorigenesis. This is well exemplified for neuroblastoma, that arises because of *MYCN* gene amplification in neural crest cells[34]. In normal neural crest development, *MYCN* is highly expressed in the sympathoadrenal trajectory, where it functions as a driver of cell expansion in the post-migratory neural crest[35]. Similarly, by analysis of single-cell transcriptomes of murine neural crest[27], we found that *MYC* was highly expressed in neural crest cells primed for mesenchymal differentiation (**Chapter 5**). This corresponds to the *SMARCB1*-induced MRT to mesenchyme differentiation that we defined in **Chapter 4**[18]. Based on these findings,

we hypothesize that, in absence of *SMARCB1*, neural crest cells *en route* to becoming mesenchyme fail to diminish *MYC* enhancer activity, which consequently makes them insensitive to differentiation signals, resulting in the maturation block of MRT.

However, *MYC* overexpression is likely not sufficient to generate MRT, as *MYC* gene amplifications never occur. This suggests that there may be additional oncogenes that are aberrantly regulated because of *SMARCB1* loss and are essential for tumorigenesis. In support of this notion, our preliminary results in **Chapter 5** indicate that MYC knockdown is not sufficient to revert the maturation block.

Overall, our findings support a tumor progression model in which the developmental pathways required for malignant transformation are epigenetically primed in the cell-of-origin. Future work may be designated to functionally interrogate the oncogenic properties of super-enhancers in MRT, which may unfold the epigenetic organization required for MRT development. Furthermore, we demonstrated that the super-enhancer landscape controlling *MYC* transcription is heterogeneous across MRT patients (**Chapter 5**). We hypothesize that this enhancer plasticity may be a genome-wide phenomenon in MRT. Although evidence is still thin, we speculate that *SMARCB1* loss creates an epigenetic context that is permissive for the aberrant formation of oncogenic super-enhancers. Further investigations are required to test this hypothesis.

SWI/SNF complex aberrations may constitute common tumor vulnerabilities

A major part of our research approach has been based on analysis of SMARCB1-dependent transcriptional and epigenetic programs in MRT (**Chapter 4 & 5**). In addition to that, we set-up a CRISPR-Cas9 knock-out screen in MRT organoids to identify potential therapeutic targets in an unbiased fashion (**Chapter 6**). Our findings demonstrated that a genetic screening approach using more physiological cancer models (i.e. patient-derived tumor organoids) can provide novel insights into tumor dependencies. Furthermore, we predict that by investigating SWI/SNF-dependent vulnerabilities in MRT, therapeutic targets may be identified that can potentially be extended to other tumors harboring genetic alterations of SWI/SNF subunits, which is over 25% of all cancers[36]. Our findings support the concept that shared vulnerabilities exist across SWI/SNF-mutant tumors. We demonstrated that kinase dependencies were partly overlapping between MRT and synovial sarcoma models (**Chapter 6**), which may relate to the similar SWI/SNF aberrations underpinning both tumor types.

Previous reports have illustrated that residual SWI/SNF complex members or associated epigenetic modifiers may constitute a synthetic lethality in SWI/SNF-mutant tumors[37,38]. Direct inhibition of the residual SWI/SNF complex has emerged as a promising therapeutic option. For instance, targeting of the non-canonical SWI/SNF complex by BRD9 perturbation was shown to be specifically effective in both MRT (**Chapter 5**) and synovial sarcoma[39]. Furthermore, inhibition of EZH2, a subunit of the Polycomb Repressive Complex 2 (PRC2), has shown pre-clinical effectiveness across

multiple SWI/SNF-mutant tumors[38], owing to the strong antagonistic relationship that PRC2 and SWI/SNF have in gene regulation[40]. However, drugs targeting universally expressed epigenetic modifiers may induce broad and therefore aspecific effects, which may be a limitation for treatment. Accordingly, we propose to combine epigenetic drugs that aim to normalize the aberrant gene regulatory mechanisms underlying SWI/SNF-mutant tumors, with therapeutics that target more specific oncogenic signaling pathways, such as the kinases identified in **Chapter 6**. We have further illustrated this concept in **Chapter 4**[18], which demonstrated the enhanced effectiveness of a combination of HDAC and mTOR inhibitors.

Altogether, combination therapy may provide a more feasible therapeutic window and may additionally prevent the rapid development of drug resistance. Future work will reveal if the potential therapeutic targets presented in this thesis (**Chapter 4-6**) are restricted to MRT or if they can be extended to synovial sarcoma and possibly other tumors driven by aberrant SWI/SNF function.

Differentiation therapy

A common characteristic of pediatric cancers that arise in prenatal life is a maturation block that locks the cancer-initiating cell in an embryonic and proliferative state[4,41]. We have demonstrated that the developmental arrest of MRT is reversible by re-expression of *SMARCB1* (**Chapter 4**[18]). By transcriptional predictions, we identified mTOR and HDAC inhibitors as pharmacological mimics of *SMARCB1* reconstitution. Furthermore, we found that at least morphologically, a similar differentiation phenotype in MRT organoids was achieved by BRD9 inhibition (**Chapter 5**).

High-throughput drug screens have proven useful for exploring drug efficacy in an unbiased fashion. However, precise assessment of cell phenotypes using such methods is complicated. Accordingly, for discovery of differentiation therapeutics, an alternative approach is required. Our findings exemplify how transcriptional signatures can faithfully predict therapy response and tumor cell differentiation (**Chapter 4**[18]). We envision that improvement of datasets that capture the transcriptional (or other) response to drugs will enable more precise predictions and in such a way improve the matching of drug to cancer. We have illustrated how side-by-side investigations into tumor and normal development increases our understanding of the maturation block at the root of pediatric cancers, which may guide the discovery of novel differentiation therapeutics.

If differentiation therapy is to succeed, enforcing a terminal cell cycle arrest is essential. We demonstrated that treatment with single agents fails to irreversibly bypass the maturation block in MRT organoids (**Chapter 4**[18]). Differentiation therapy was most effective when combining mTOR and HDAC inhibitors, as it most importantly prevented regrowth. Accordingly, we speculate that dual targeting is key for induction of terminal differentiation[42]. First, it is essential to target the self-renewal capacity of tumor cells. Secondly, to prevent regrowth, we hypothesize that tumor cells require to

be epigenetically reprogrammed so that reversal is prohibited. This can be achieved by targeting epigenetic modifiers such as HDACs, which can open the chromatin at differentiation genes and keep the tumor cell in a matured state. However, considering that MRT arises in early childhood with ongoing postnatal developmental processes, adverse effects of maturation therapeutics should be carefully evaluated.

A common issue in cancer treatment is drug resistance. We therefore assume that differentiation therapy by itself may not be sufficient for the complete remission of MRT. Accordingly, we anticipate a treatment strategy in which cytotoxic drugs are required to decrease most of the tumor mass, and differentiation therapy may specifically target residual cancer stem cells and micro metastasis to prevent relapse. Evidence for this concept has been obtained from neuroblastoma, for which retinoic acid was successfully applied to fight minimal residual disease[43-45]. Overall, enforcing a bypass of the maturation block by pharmacological means appears to be a promising therapeutic avenue in pediatric cancer[42].

Concluding remarks

In conclusion, in this Thesis we have defined the cellular identity of multiple renal tumors, including that of MRT. We provided evidence of MRT origin, gained crucial insights into the differentiation pathways underpinning MRT development, and have identified therapeutic vulnerabilities through an approach that is broadly applicable to childhood cancer. MRT remains a continuous challenge for pediatric oncologists, and for the sake of MRT patients that still face a dismal prognosis, we strive that our findings may contribute to significant therapeutic improvement.

References

1. Puisieux, A., Pommier, R. M., Morel, A. P. & Lavial, F. Cellular Pliancy and the Multistep Process of Tumorigenesis. *Cancer Cell* **33**, 164-172, doi:10.1016/j.ccell.2018.01.007 (2018).

2. Roy, N. & Hebrok, M. Regulation of Cellular Identity in Cancer. *Dev Cell* **35**, 674-684, doi:10.1016/j.devcel.2015.12.001 (2015).

3. Chen, X., Pappo, A. & Dyer, M. A. Pediatric solid tumor genomics and developmental pliancy. *Oncogene* **34**, 5207-5215, doi:10.1038/onc.2014.474 (2015).

4. Marshall, G. M. *et al.* The prenatal origins of cancer. *Nat Rev Cancer* **14**, 277-289, doi:10.1038/nrc3679 (2014).

5. Custers, L., Paassen, I. & Drost, J. In vitro Modeling of Embryonal Tumors. *Front Cell Dev Biol* **9**, 640633, doi:10.3389/fcell.2021.640633 (2021).

6. Filbin, M. G. *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **360**, 331-335, doi:10.1126/science.aao4750 (2018).

7. Gojo, J. *et al.* Single-Cell RNA-Seq Reveals Cellular Hierarchies and Impaired Developmental Trajectories in Pediatric Ependymoma. *Cancer Cell* **38**, 44-59 e49, doi:10.1016/j.ccell.2020.06.004 (2020).

8. Jessa, S. *et al.* Stalled developmental programs at the root of pediatric brain tumors. *Nat Genet* **51**, 1702-1713, doi:10.1038/s41588-019-0531-7 (2019).

9. Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599, doi:10.1126/science.aat1699 (2018).

10. Young, M. D. et al. Single cell derived mRNA signals across human kidney tumors. Nat Commun 12, 3896, doi:10.1038/s41467-021-23949-5 (2021).

11. Hatley, M. E. *et al.* A mouse model of rhabdomyosarcoma originating from the adipocyte lineage. *Cancer Cell* **22**, 536-546, doi:10.1016/j.ccr.2012.09.004 (2012).

12. Rousseaux, M. W. *et al.* Depleting Trim28 in adult mice is well tolerated and reduces levels of alpha-synuclein and tau. *Elife* **7**, doi:10.7554/eLife.36768 (2018).

13. Xu, Y. *et al.* Inducible knockout of Twist1 in young and adult mice prolongs hair growth cycle and has mild effects on general health, supporting Twist1 as a preferential cancer target. *Am J Pathol* **183**, 1281-1292, doi:10.1016/j.ajpath.2013.06.021 (2013).

14. Nazha, B., Inal, C. & Owonikoko, T. K. Disialoganglioside GD2 Expression in Solid Tumors and Role as a Target for Cancer Therapy. *Front Oncol* **10**, 1000, doi:10.3389/fonc.2020.01000 (2020).

15. Dobrenkov, K. & Cheung, N. K. GD2-targeted immunotherapy and radioimmunotherapy. *Semin Oncol* **41**, 589-612, doi:10.1053/j.seminoncol.2014.07.003 (2014).

16. Lin, C. Y. *et al.* Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57-62, doi:10.1038/nature16546 (2016).

17. Mack, S. C. *et al.* Therapeutic targeting of ependymoma as informed by oncogenic enhancer profiling. *Nature* **553**, 101-105, doi:10.1038/nature25169 (2018).

18. Custers, L. *et al.* Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nat Commun* **12**, 1407, doi:10.1038/s41467-021-21675-6 (2021).

19. Woodworth, M. B., Girskis, K. M. & Walsh, C. A. Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat Rev Genet* **18**, 230-244, doi:10.1038/nrg.2016.159 (2017).

20. Baron, C. S. & van Oudenaarden, A. Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat Rev Mol Cell Biol* **20**, 753-765, doi:10.1038/s41580-019-0186-3 (2019).

21. Lodato, M. A. *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94-98, doi:10.1126/science.aab1785 (2015).

22. Bronner, M. E. & Simoes-Costa, M. The Neural Crest Migrating into the Twenty-First Century. *Curr Top Dev Biol* **116**, 115-134, doi:10.1016/bs.ctdb.2015.12.003 (2016).

23. Chun, H. E. *et al.* Identification and Analyses of Extra-Cranial and Cranial Rhabdoid Tumor Molecular Subgroups Reveal Tumors with Cytotoxic T Cell Infiltration. *Cell Rep* **29**, 2338-2354 e2337, doi:10.1016/j.celrep.2019.10.013 (2019).

24. Johann, P. D. *et al.* Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes. *Cancer Cell* **29**, 379-393, doi:10.1016/j.ccell.2016.02.001 (2016).

25. Dash, S. & Trainor, P. A. The development, patterning and evolution of neural crest cell differentiation into cartilage and bone. *Bone* **137**, 115409, doi:10.1016/j.bone.2020.115409 (2020).

26. Rocha, M. *et al.* From head to tail: regionalization of the neural crest. *Development* **147**, doi:10.1242/dev.193888 (2020).

27. Soldatov, R. *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**, doi:10.1126/science.aas9536 (2019).

28. Furlan, A. & Adameyko, I. Schwann cell precursor: a neural crest cell in disguise? *Dev Biol* **444 Suppl 1**, S25-S35, doi:10.1016/j.ydbio.2018.02.008 (2018).

29. Hahn, W. C. & Weinberg, R. A. Rules for making human tumor cells. *N Engl J Med* **347**, 1593-1603, doi:10.1056/NEJMra021902 (2002).

30. Curtius, K., Wright, N. A. & Graham, T. A. An evolutionary perspective on field cancerization. *Nat Rev Cancer* **18**, 19-32, doi:10.1038/nrc.2017.102 (2018).

31. Coorens, T. H. H. *et al.* Embryonal precursors of Wilms tumor. *Science* **366**, 1247-1251, doi:10.1126/science.aax1323 (2019).

32. Filbin, M. & Monje, M. Developmental origins and emerging therapeutic opportunities for childhood cancer. *Nat Med* **25**, 367-376, doi:10.1038/s41591-019-0383-9 (2019).

33. Kerosuo, L. & Bronner, M. E. cMyc Regulates the Size of the Premigratory Neural Crest Stem Cell Pool. *Cell Rep* **17**, 2648-2659, doi:10.1016/j.celrep.2016.11.025 (2016).

34. Johnsen, J. I., Dyberg, C. & Wickstrom, M. Neuroblastoma-A Neural Crest Derived Embryonal Malignancy. *Front Mol Neurosci* **12**, 9, doi:10.3389/fnmol.2019.00009 (2019).

35. Hansford, L. M. *et al.* Mechanisms of embryonal tumor initiation: distinct roles for MycN expression and MYCN amplification. *Proc Natl Acad Sci U S A* **101**, 12664-12669, doi:10.1073/pnas.0401083101 (2004).

36. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet* **45**, 592-601, doi:10.1038/ng.2628 (2013).

37. Centore, R. C., Sandoval, G. J., Soares, L. M. M., Kadoch, C. & Chan, H. M. Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. *Trends Genet* **36**, 936-950, doi:10.1016/j.tig.2020.07.011 (2020).

38. Mittal, P. & Roberts, C. W. M. The SWI/SNF complex in cancer - biology, biomarkers and therapy. *Nat Rev Clin Oncol* **17**, 435-448, doi:10.1038/s41571-020-0357-3 (2020).

39. Michel, B. C. *et al.* A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol* **20**, 1410-1420, doi:10.1038/s41556-018-0221-1 (2018).

40. Kadoch, C., Copeland, R. A. & Keilhack, H. PRC2 and SWI/SNF Chromatin Remodeling Complexes in Health and Disease. *Biochemistry* **55**, 1600-1614, doi:10.1021/acs.biochem.5b01191 (2016).

41. Behjati, S., Gilbertson, R. J. & Pfister, S. M. Maturation Block in Childhood Cancer. *Cancer Discov* **11**, 542-544, doi:10.1158/2159-8290.CD-20-0926 (2021).

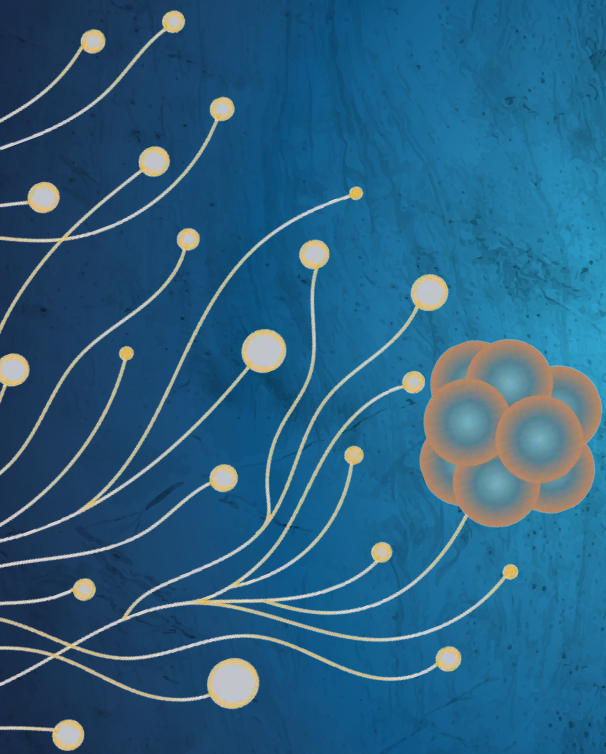
42. de The, H. Differentiation therapy revisited. *Nat Rev Cancer* **18**, 117-127, doi:10.1038/nrc.2017.103 (2018).

43. Matthay, K. K. *et al.* Long-term results for children with high-risk neuroblastoma treated on a randomized trial of myeloablative therapy followed by 13-cis-retinoic acid: a children's oncology group study. *J Clin Oncol* **27**, 1007-1013, doi:10.1200/JCO.2007.13.8925 (2009).

44. Matthay, K. K. *et al.* Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. Children's Cancer Group. *N Engl J Med* **341**, 1165-1173, doi:10.1056/NEJM199910143411601 (1999).

45. Villablanca, J. G. *et al.* Phase I trial of 13-cis-retinoic acid in children with neuroblastoma following bone marrow transplantation. *J Clin Oncol* **13**, 894-901, doi:10.1200/JCO.1995.13.4.894 (1995).

7



Addendum

Nederlandse samenvatting

De afgelopen decennia zijn de overlevingskansen van kinderkankerpatiënten fors gestegen, mede door biomedisch onderzoek en de ontwikkeling van innovatieve behandelingen. Toch is kanker nog steeds de meest voorkomende ziekte-gerelateerde doodsoorzaak bij kinderen, en voor een aantal kankertypes is helaas nog geen effectieve behandeling. Daarnaast veroorzaken veel van de huidige therapieën ernstige bijwerkingen, die de gezondheid van kinderen chronisch kunnen schaden. Er is veel wat we nog niet begrijpen over de ontwikkeling van kinderkanker. Daarom is het belangrijk om onderzoek te intensiveren met het doel om effectievere en meer gerichte therapieën te kunnen ontwikkelen.

Voor een subgroep van tumoren die vooral bij jonge kinderen voorkomt zijn aanwijzingen dat ze al ontstaan voor de geboorte. Verondersteld wordt dat deze tumoren een gevolg zijn van ontspoorde embryonale ontwikkeling. Tijdens de normale embryonale ontwikkeling, dienen voorloper cellen snel te vermenigvuldigen om weefselgroei te bevorderen, om uiteindelijk te differentiëren in de verschillende celtypes die nodig zijn voor goed functionerende organen. Een terugkerend kenmerk van kindertumoren is dat ze vast zitten in dit embryonale voorloper stadium, waardoor ze blijven groeien en niet kunnen matureren en differentiëren. Het embryonaal profiel dat kindertumoren kenmerkt is afwezig in volwassen cellen, en zou daarom een specifieke therapeutische kwetsbaarheid kunnen zijn van kinderkankercellen. Om deze embryonale identiteit beter te begrijpen is het cruciaal om te ontdekken hoe en uit welk celtype kinderkanker ontstaat.

Veel van de kinderkankertumoren hebben mutaties in genen die coderen voor epigenetische regulatoren, dat zijn eiwitten die de structuur van het DNA kunnen modificeren en reguleren. DNA is zeer compact opgevouwen om in de celkern te passen, maar dient lokaal te worden geopend voor het activeren van genen. Dit is een actief en sterk gecontroleerd mechanisme dat nodig is om vrijwel alle processen in een cel te reguleren. Als dit proces wordt verstoord als gevolg van een mutatie kan kanker ontstaan. Een goed voorbeeld hiervan zijn maligne rhabdoïde tumoren (MRT), die gekenmerkt worden door een inactivatie van het *SMARCB1* gen. SMARCB1 is een subeenheid van het SWI/SNF complex, die als functie heeft om de toegankelijkheid van het DNA te reguleren. In andere woorden, om het DNA open of dicht te vouwen. MRT is een zeer agressieve tumor die het meeste voorkomt bij jonge kinderen (<3 jaar) en kan ontstaan in verschillende delen van het lichaam, zoals het centrale zenuwstelsel en de nieren. MRT is een van de meest dodelijke kindertumoren met zeer lage overlevingskansen (15-50%). Er is dus duidelijk behoefte aan meer kennis en betere behandelingsopties.

Dit proefschrift beschrijft onze studies naar de oorsprong van nierkanker en MRT, de moleculaire processen die verantwoordelijk zijn voor de ontwikkeling van MRT, en hoe we deze informatie kunnen vertalen naar nieuwe potentiële therapeutische aangrijpingspunten. Om de mechanismen te ontrafelen die kindertumoren

aandrijven zijn experimentele modelsystemen essentieel. Door recente ontwikkeling van innovatieve kweektechnieken is het mogelijk om tumorcellen te groeien en te manipuleren in het lab (*in vitro*). Een van deze celkweek systemen zijn zogenoemde tumor organoïden, die efficiënt gegenereerd kunnen worden uit patiëntenmateriaal. Deze cellen groeien in een 3D matrix als mini-tumoren en zijn zeer representatief voor het weefsel waar ze vandaan komen. Dit maakt organoïden uiterst geschikt als studiemodel. Organoïden afkomstig van MRT materiaal zijn van grote waarde gebleken voor de experimenten beschreven in dit proefschrift.

In **Hoofdstuk 2** beschrijven we door middel van een literatuurstudie de verschillende *in vitro* modellen die de ontwikkeling van embryonale tumoren na kunnen bootsen in het lab.

In **Hoofdstuk 3** analyseren we de embryonale identiteit en mogelijke oorsprong van niertumoren en MRT op basis van mRNA patronen. Omdat elk celtype verschillende genen tot expressie brengt, heeft elk celtype zijn eigen kenmerkende mRNA profiel. Tumorcellen erven waarschijnlijk een gedeelte van het mRNA profiel van de cel waar ze uit zijn ontstaan. Door quantitatief de mRNA patronen van tumorcellen te vergelijken met verschillende embryonale en volwassen celtypes, hebben we ontdekt dat kinderkanker significant lijkt op embryonale cellen, in tegenstelling tot volwassen kanker. Dit is een sterke aanwijzing dat de verschillende kindertumoren die we hebben geïncludeerd in onze analyse, inderdaad uit embryonale cellen ontstaan.

In **Hoofdstuk 4** zoeken we bewijs voor de oorsprong van MRT met behulp van phylogenetische analyses. Dit is een methode waar we op basis van mutaties in het DNA, die als genetische barcodes fungeren, kunnen achterhalen hoe twee cellen of weefsels aan elkaar gerelateerd zijn. Door mutaties te identificeren en deze te vergelijken tussen tumor en verschillende gezonde weefsels hebben we definitief bewijs kunnen leveren dat MRT afkomstig is van een groep embryonale voorloper cellen, genaamd de “neural crest”.

Verder hebben we organoïden gebruikt om de rol van *SMARCB1* mutaties te bestuderen in de ontwikkeling van MRT. Door MRT organoïden genetisch te manipuleren was het mogelijk om het *SMARCB1* verlies te herstellen door *SMARCB1* opnieuw tot expressie te brengen. Dit resulteerde in een groeistop van MRT organoïden die lijkt op een differentiatie van neural crest naar mesenchymale cellen. Deze resultaten tonen aan dat MRT ontstaat uit embryonale neural crest die op weg zijn om te differentiëren naar mesenchymale cellen.

Het stoppen van MRT groei door herexpressie van *SMARCB1* is op dit moment niet toepasbaar in patiënten. Daarom hebben we gezocht naar medicijnen die de herexpressie van *SMARCB1* kunnen nabootsen. De mRNA veranderingen die plaatsvonden tijdens de differentiatie van MRT organoïden hebben we geanalyseerd en gebruikt om de effectiviteit van verschillende medicijnen te voorspellen. Aan de hand van deze voorspellingen hebben we een combinatie van HDAC en mTOR remmers

getest op MRT organoïden, wat een onomkeerbare groeistop induceerde, en dus mogelijk een effectieve therapie zou kunnen zijn om MRT patiënten te behandelen.

In **Hoofdstuk 5** bestuderen we de rol van SMARCB1 als epigenetische regulator in de ontwikkeling van MRT. We analyseerden de structuur en vouwing van DNA in MRT organoïden, hoe deze verandert tijdens SMARCB1 herexpressie, en wat deze veranderingen voor effect hebben op genexpressie. Op deze manier proberen we te achterhalen welke cellulaire processen afwijken door verlies van SMARCB1, die mogelijk verantwoordelijk zijn voor de kwaadaardige groei van MRT cellen. Onze resultaten tonen aan dat SMARCB1 essentieel is voor het reguleren van zogenoemde "enhancers". Dit zijn stukjes genetische code in het DNA die niet voor een eiwit coderen, maar belangrijk zijn voor het reguleren van genexpressie. Onze resultaten laten zien dat SMARCB1 vermoedelijk nodig is om de enhancers van het *MYC* gen te deactiveren. Het remmen van *MYC* expressie bleek essentieel om MRT groei te stoppen.

In **Hoofdstuk 6** rapporteren we de resultaten van een CRISPR-Cas9 knock-out screen in MRT organoïden, waarmee we MRT-specifieke kwetsbaarheden ontdekte. De CRISPR-Cas9 toolkit is een methode om gericht DNA in cellen te modificeren. Dit maakt het mogelijk om genen uit te schakelen. Door dit op grote schaal voor verschillende genen toe te passen, hebben we achterhaald welke genen verantwoordelijk zijn voor de groei van MRT. Daarna hebben we de eiwitten waarvoor deze genen coderen met medicijnen geremd. Dit resulteerde in een groeistop die specifiek was voor MRT organoïden, en daarom mogelijk een nieuwe effectieve behandeling zou kunnen zijn voor MRT patiënten.

Concluderend, we hebben de embryonale identiteit gedefinieerd van verschillende niertumoren en MRT. We hebben bewijs geleverd voor de cellulaire oorsprong van MRT, cruciale bevindingen gedaan over de differentiatie paden die ten grondslag liggen aan MRT, en therapeutische kwetsbaarheden ontdekt. MRT blijft tot op de dag van vandaag een uitdaging voor kinderoncologen. In het belang van de patiënten die nog steeds een sombere prognose hebben, hopen we dat onze bevindingen bij zullen dragen aan therapeutische innovatie die de overlevingskans van MRT patiënten gaat vergroten.

DNA: Moleculen die het menselijk genoom opmaken en alle genetische code bevatten die nodig is voor elke cel in ons lichaam.

mRNA: Net als DNA, fungeert mRNA als genetische code. mRNA is een boodschapper molecuul dat wordt overgeschreven van DNA en een centrale rol heeft in het vertalen van genen naar eiwitten.

Gen: Een stukje DNA in het genoom dat codeert voor een eiwit.

Epignetische regulatoren: Eiwitten die de structuur van het DNA in een cel kunnen modificeren en reguleren.

Differentiëren: De transformatie van een voorloper naar een gespecialiseerde cel met een specifieke functie (bijvoorbeeld een spier, vet of zenuwcel).

Neural crest: Voorloper cellen die tijdens de embryonale ontwikkeling onder andere het perifere zenuwstelsel vormen.

Mesenchymale cellen: Een verzamelnaam voor cellen die onder andere bot-, kraak- en bindweefsel vormen.

Author contributions per chapter

CHAPTER 1
L.C. wrote the chapter, supervised by J.D.

CHAPTER 2
L.C., I.P., and J.D. wrote the manuscript. J.D. supervised the work.

CHAPTER 3
M.D.Y. and S.B. conceived of the experiment and wrote the manuscript. M.D.Y. performed analyses, aided by T.J.M., E.K., G.K., and T.H.H.C. I.D.V. and J.C.A. provided expertise on adrenal gland analysis. L.C. performed organoid experiments with F.A.V.B. T.R.W.O., N.S., D.R., N.C., L.H., R.R.K., A.W. provided pathological expertise. F.C., M.M.H.E., and A.S. provided clinical data. A.P., E.B.B., F.M., C.T., C.B., G.D.S., V.J.G., M.H., M.K., S.M.P., O.A.B., K.R., K.K, F.C.H., J.D., F.C.H., E.P., K.A., contributed to discussions and / or data. S.A.T., T.M., F.C.H., F.M., J.D., R.R.K, contributed fetal and tumor single cell data, together with K.B.M., R.A.B., X.H., A.W.C, L.M. S.B. and M.D.Y. directed the study, in conjunction with K.S. (single cell cancer work) and J.D. (organoid work).

CHAPTER 4
S.B. and J.D. conceived of the study. L.C. performed *in vitro* studies. E.K. performed transcriptional analyses, aided by L.C. Phylogenetic analyses were conducted by T.H.H.C. Pathological expertise was provided by T.R.W.O, L.H., C.T., R.R.K. C.C. performed drug assays. Statistical expertise was provided by M.D.Y. L.C. and F.A.V.B. prepared samples for single cell analyses. P.E. and A.M. were involved in tissue processing. Samples were curated and/or experiments were performed by H.S., L.M., E. W. H., M.H.E., J.N. K.S., and K.A. M.K. performed DNA methylation assays. S.B. supervised phylogenetic and transcriptional analyses. J.D. supervised experimental work. S.B. and J.D. wrote the manuscript and directed this work.

CHAPTER 5
L.C. performed the in vitro experiments, aided by I.P. N.Q.L. performed the genomic analysis. H.T. performed the library preparations. J.D. and E.W. supervised and directed the work. L.C. wrote the chapter, aided by N.Q.L. and supervised by J.D.

CHAPTER 6
L.C. performed the in vitro experiments, aided by M.G. and F.M.R. L.C. performed the analysis. J.D. supervised and directed the work. L.C. wrote the chapter, supervised by J.D.

CHAPTER 7
L.C. wrote the chapter, supervised by J.D.

List of publications

Matthew D Young[†*], Thomas J Mitchell[†], **Lars Custers**[†], Thanasis Margaritis, Francisco Morales-Rodriguez, Kwasi Kwakwa, Eleonora Khabirova, Gerda Kildisiute, Thomas RW Oliver, Ronald R de Krijger, Marry M van den Heuvel-Eibrink, Federico Comitani, Alice Piapi, Eva Bugallo-Blanco, Christine Thevanesan, Christina Burke, Elena Prigmore, Kirsty Ambridge, Kenny Roberts, Felipe A Vieira Braga, Tim HH Coorens, Ignacio Del Valle, Anna Wilbrey-Clark, Lira Mamanova, Grant D Stewart, Vincent J Gnanapragasam, Dyanne Rampling, Neil Sebire, Nicholas Coleman, Liz Hook, Anne Warren, Muzlifah Haniffa, Marcel Kool, Stefan M Pfister, John C Achermann, Xiaoling He, Roger A Barker, Adam Shlien, Omer A Bayraktar, Sarah A Teichmann, Frank C Holstege, Kerstin B Meyer, Jarno Drost[*], Karin Straathof[*], Sam Behjati[*]. Single-cell derived mRNA signals across human kidney tumors. *Nat Commun* **12**, 3896, doi:10.1038/s41467-021-23949-5 (2021).

Lars Custers[†], Eleonora Khabirova[†], Tim H.H. Coorens[†], Thomas R.W. Oliver, Camilla Calandrini, Matthew D. Young, Felipe A. Vieira Braga, Peter Ellis, Lira Mamanova, Heidi Segers, Arie Maat, Marcel Kool, Eelco W. Hoving, Marry M. van den Heuvel-Eibrink, James Nicholson, Karin Straathof, Liz Hook, Ronald R. de Krijger, Claire Trayers, Kieren Allinson, Sam Behjati[*], Jarno Drost[*]. Somatic mutations and single cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nat Commun* **12**, 1407, doi:10.1038/s41467-021-21675-6 (2021).

Lars Custers[†], Irene Paassen[†], Jarno Drost. *In vitro* modeling of embryonal tumors. *Front Cell Dev Biol* **9**, 640633, doi:10.3389/fcell.2021.640633 (2021).

Camilla Calandrini[†], Frans Schutgens[†], Rurika Oka, Thanasis Margaritis, Tito Candelli, Luka Mathijsen, Carola Ammerlaan, Ravian L. van Ineveld, Sepide Derakhshan, Sanne de Haan, Emmy Dolman, Philip Lijnzaad, **Lars Custers**, Harry Begthel, Hindrik H. D. Kerstens, Lindy L. Visser, Maarten Rookmaaker, Marianne Verhaar, Godelieve A. M. Tytgat, Patrick Kemmeren, Ronald R. de Krijger, Reem Al-Saadi, Kathy Pritchard-Jones, Marcel Kool, Anne C. Rios, Marry M. van den Heuvel-Eibrink, Jan J. Molenaar, Ruben van Boxtel, Frank C. P. Holstege, Hans Clevers[*], Jarno Drost[*]. An organoid biobank for childhood kidney cancers that captures disease and tissue heterogeneity. *Nat Commun* **11**, 1310, doi:10.1038/s41467-020-15155-6 (2020).

Rui Lopes[†], Gozde Korkmaz[†], Sonia Aristin Revilla, Romy van Vliet, Remco Nagel, **Lars Custers**, Yongsoo Kim, Pieter C van Breugel, Wilbert Zwart, Behzad Moumbeini, Zohar Manber, Ran Elkon, Reuven Agami. *CUEDC1* is a primary target of ERα essential for the growth of breast cancer cells. *Cancer Lett* **436**, 87-95, doi:10.1016/j.canlet.2018.08.018 (2018).

[†] These authors contributed equally to this work.

[*] These authors jointly supervised this work.

Curriculum Vitae

Lars Custers was born on August 28th in 1991 in Venlo, the Netherlands. In 2009, he obtained his high school degree (VWO) at College Den Hulster in Venlo, after which he started the Bachelor Biomedical Sciences at Maastricht University. During his Bachelor, he performed an internship at the department of Molecular Cell Biology at Maastricht University, identifying prognostic markers for lung cancer patients. In 2013 he obtained his Bachelor degree and started a gap year in which he did voluntary work as a teacher in Peru. Lars then moved to back to the Netherlands to start his Master Cancer, Stem Cells & Developmental Biology at Utrecht University. During his Master, he performed his first internship in the lab of dr. Rachel Giles at the University Medical Centre Utrecht where he investigated renal disease using RNA interference and microscopy. His second internship was done in the lab of prof. dr. Reuven Agami at the Netherlands Cancer Institute in Amsterdam, in which he studied genomic regulatory regions using CRISPR-Cas9. He obtained his Master degree in 2017, after which he started as a PhD candidate in June 2017. He joined the lab of dr. Jarno Drost at the Princess Maxima Centre in Utrecht. There, he studied the connection between pediatric cancer and embryonic development using organoid models. The results are presented in this Thesis.

Acknowledgements

The work presented in this Thesis could never have been accomplished without the help of the many wonderful people that I encountered during the four years of my PhD. Most importantly, I had an amazing time, both scientifically and socially. As such, I would like to say some words of thanks.

Ten eerste wil ik **Jarno** bedanken voor deze kans. Jouw eindeloze motivatie en enthousiasme hebben me geïnspireerd om echt het beste uit mezelf te halen. Door jou supervisie en betrokkenheid heb ik veel over de wetenschap geleerd, waardoor ik me heb kunnen ontwikkelen tot de onderzoeker die ik nu ben. Toen ik aan dit avontuur begon, was je groep nog relatief klein en het Prinses Maxima Centrum stond nog in de kinderschoenen. Het was geweldig om er vanaf de start te staan, en de groei mee te maken die het ziekenhuis en jouw lab hebben doorgemaakt. Ik wens je het allerbeste voor de toekomst en ik ben er zeker van dat we elkaar weer tegenkomen.

Research is never a single person job. The Drost lab has become a very close and supportive team. This was essential for me to come this far, so big thanks for that. Things that I will miss the most are: the gossip during lunch, the humor and jokes during cell culture sessions, complaining about all and nothing during coffee breaks, the crazy retreats and parties, and much more. Unfortunately, the last 1.5 years have been "slightly" boring on the social side, but still we managed to creatively come together, either online or with 1.5m distance, which I really appreciated.

Special thanks to **Camilla**. I'm glad that we could experience the PhD journey together, which has been challenging from time to time. You boosted my confidence when needed and always showed so much personal interest. I'm really grateful for you being my paranymph! Thank you **Sepi** for all the support. I really enjoyed sharing our passion for food and Roberto's gelato! **Francisco**, thanks for your great lectures on world history, the board game nights, our music alliance in ML2, and your skills in making fun of people (especially Irene and Juliane). Thank you, **Irene** for being Francisco's target. Without joking now, I was really happy to work on our joint projects together. I may not have shown it too much, but I really valued your work and thoughts! **Juliane**, it was a pleasure working with you, always bringing a positive vibe with your humor and laugh, even at times when work was overwhelming. Our joint ML2 sessions were things to look forward to. Laaaaaarsssiiieee, is what I would often hear when I entered the office at the PMC, which was unmistakably **Arianna**. Thank you for all your help and advice in the lab and outside. Your enthusiasm is infectious! When I just started, I was introduced to the lab by **Maroussia**, who was an intern in the lab at the time. Recently, she started as PhD student taking over the same project that she initially started. So, we can say the circle is complete, good luck with everything! When it is borrel time, **Sanne** is there, it was great working and especially drinking together! Always fun having **Yvonne** around, even showing up when heavily pregnant. The same applies to you **Sofia**, love your humor and thanks for all the support! Furthermore, I

would like to thank the rest of the (former) lab members **Kim, Frans, Irene, Dilara, Laura, Chiara, Luca** and **Marloes** for a great time at the PMC!

Next, I would like to thank **Marry, Ronald** and **Eelco** for the pleasant collaborations at the PMC. You were invaluable for the clinical perspective in our research. Furthermore, I greatly valued collaborating with **Elzo** and **Sam** and their teams. My development as a scientist has been greatly influenced by your unique views on biology. **Ning Qing,** it was always a pleasure working with you, as our collaboration was smooth, relaxed and optimistic. I hope we can work together more often in the future.

It is clear that the Princess Maxima Centre is full of nice and talented people, which has made the occasional high workload way easier to handle. There are many people I would like to thank, and I apologize if I forget to mention someone upfront. Firstly, I would like to thank the entire **van Boxtel** group, our joint retreats were so much fun. Special thanks to **Axel**, who is one of my Paranymphs. I've come to know you during our Master program, and from there our careers are running in parallel. It was great fun to organize the joint Drost-Van Boxtel retreat and I enjoyed all our futsal sessions. I must say, you've developed some skills over the years. Good luck in Barcelona, I will definitely come and visit you there! **Ravian**, it was great having someone around with the same sense of humor. I really valued our small collaborations and discussions on science. More importantly, I greatly enjoyed our talks on football, even though you're obviously misguided as a Feyenoord supporter. One of my fellow craft beer lovers, **Rijndert**, thanks for all the great times at the borrels and parties. I hope you can still do all the hikes that you planned in the near future and let me know all about them! **Jeff**, my roommate during several retreats and masterclasses, thanks for everything and especially the memorable Quadruple night! **Wim**, with you around, it was impossible to be grumpy. Also, it was sort of impossible to bail on a pubquiz night or futsal game, as you can be very persuasive on such matters! **Evelyn** and **Margit**, thanks for all the serious and especially the 'not so serious' conversations. You were great colleagues with whom I could discuss about anything. **Eddy**, someone I could always count on to show up. It was nice to finally be able to visit an Ajax game together! Furthermore, I would like to thank the numerous great colleagues that I worked with at the PMC: **Waleed, Michael, Thanasis, Philip, Tito, Femke, Karin, Marc, Karlijn, Eline, Niels, Arianne, Sjors, Rurika, Jurrian, Mark V, Mark v R, Freek, Thomas, Valentina, Wilco, Rico, Jordy, Vera, Florijn, Frank, Hannah, Mirjam, Flavia, Winnie,** and the many more people that I'm probably forgetting here! Additionally, I would like to thank all the PIs and facilities at the PMC for their support and guidance. Importantly, I would like to express my gratitude to all the children and their parents who decided to participate in the studies.

Special thanks to my PhD Committee members **Jop, Hans** and **Frank**, who have given me valuable feedback on my progress over the years. Furthermore, I would like to express my gratitude to the examination committee and all the opponents present at my defense.

I've also met a lot of great people not directly related to my research. Last years would have been way less fun without the Futsal team, the Pubquiz team and all the people joining for the Ardenne weekend. So big thanks to everyone involved!

Daarnaast ben ik altijd gemotiveerd door familie en vrienden. **Koen, Kevin, Falke, Mark, Raoul, Edje, Sam, Lisanne, Tim, Aiman, Angelica, Sibani** en voetbalteam, bedankt voor de welkome afleiding die vaak hoognodig was! **Shahin**, thanks for being such a great neighbour! Verder wil ik mijn oom **Roland** nog specifiek bedanken voor het drukwerk. **Kay, Floortje, Mam, Pap, Jan** en **Jeanne** bedankt voor alle steun, de altijd positieve woorden en interesse in mijn werk. Ik heb met veel plezier een poging gedaan om uit te leggen waar ik met zoveel passie aan gewerkt heb.

Tot slot, degene waar ik deze vier jaar het meest intens mee heb beleefd. Mijn vriendin (en sinds kort verloofde!!) **Elise**, bedankt voor alles. Nog even de cellen eten geven in het weekend. Presentaties maken tot s' avonds laat. Niet op komen dagen voor het avondeten. Dit zijn zomaar een aantal dingen waar ik je echt nooit over heb horen klagen, waar je wel alle recht toe had. Bedankt voor al het begrip, en de continue steun waar ik echt enorm veel aan heb gehad. We gaan spannende tijden tegemoet, maar dat gaat helemaal goed komen samen!