On the origin of cancer:

Mutation accumulation in healthy and precancerous adult stem cells

Myrthe Jager

ISBN: 978-94-93019-90-4
Design and layout: Myrthe Jager
Photos: Bianca Kamlag van den Winkel (Cover and pages 8, 30, 72, 98, 136 & 198)
and Myrthe Jager (Pages 160 & 212)
Printed by: Proefschriftmaken

On the origin of cancer:

Mutation accumulation in healthy and precancerous adult stem cells

Over de oorsprong van kanker:
Mutatie accumulatie in gezonde en premaligne adulte stamcellen
(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 donderdag 11 oktober 2018 des middags te 12.45 uur

door

Myrthe Jager

geboren 28 juni 1990
te Groningen

Promotor: Prof. dr. ir. E.P.J.G. Cuppen

Copromotor: Dr. R van Boxtel

Contents

The best is yet to be

Chapter 1

Introduction

Myrthe Jager[1], Ruben van Boxtel[2] and Edwin Cuppen[1]

1 Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht University, Universiteitsweg 100, 3584, CG, Utrecht, The Netherlands
2 Princess Máxima Center for Pediatric Oncology, 3584 CT Utrecht, The Netherlands

Introduction

The genome contains all hereditary information that is required to build, maintain, and reproduce an organism. This information is present in every cell and can be passed on to progeny. From plants, to yeast, to humans, all reproducing organisms have a genome (*1*) and the molecular structure is strikingly similar. The genetic information is divided across multiple DNA molecules (chromosomes), each of which is made up of two long strings of four bases (nucleotides): adenine (A), thymine (T), cytosine (C), and guanine (G). The two strings of nucleotides are coiled around each other, forming a double helix, in which an A is always opposite to a T and a C always opposite to a G (*2*). In total, the human genome consists of 12 billion of these bases divided across 46 chromosomes, 23 of which are inherited from the father and 23 from the mother (*3*). Only 1-2% of the nucleotides reside within the coding regions of ~20,000 genes and are translated into proteins (*4*–*6*). Although the function of the remainder of the genome is not fully understood, it is clear that a (large) fraction of these regions is involved in regulating how much of a certain protein is eventually produced at what moment and in which cell (gene expression regulation) (*7*). It is crucial that the genome sequence remains stable during life. A single alteration (mutation) in the DNA code can change the entire structure or the expression of a protein. Changes in the genetic code can thereby affect the health (the fitness) of cells and can ultimately contribute to the development of age-associated diseases, such as cancer (*8*, *9*).

Cancer is the clonal outgrowth of transformed cells, through uncontrolled cellular proliferation and/or reduced cell death (apoptosis). In addition to changes in proliferation and apoptosis, multiple other cellular processes are disturbed in cancer cells (*10*). For example, cancer cells can stimulate the growth of blood vessels (angiogenesis), which enables them to receive enough nutrition required for accelerated growth (*11*). Furthermore, cancer cells often acquire characteristics that allow them to escape their environment and metastasize to another tissue (*12*). Cancer remains a major cause of death, with over 8 million deaths per year worldwide (*13*, *14*). Although the mortality rate has been decreasing in the past decades (*15*), the incidence is expected to rise tremendously, predominantly due to westernization of lifestyle in non-western countries (*16*). Therefore, it will become increasingly important to develop strategies aimed to prevent the development of cancer, in order to reduce the number of deaths by cancer.

To develop prevention strategies it is, among other things, important to understand how and why a tumor develops (tumorigenesis/oncogenesis). Well-known risk factors for developing cancer are old age, genetic predisposition, exposure to sunlight, alcohol consumption, tobacco smoking, viral infections, and

obesity (*17–21*). However, for the majority of these risk factors it remains unclear why they contribute to the development of (specific types of) cancer (*22*). In addition, the exposure to risk factors alone cannot explain the extreme variation in cancer incidence across different tissues in humans (*22–24*). Recently, several new techniques were developed that allow us to gain more insight into the processes that precede (and might cause) oncogenic transformation. Here, I will provide an overview of these techniques, and describe several hypotheses on how risk factors can contribute to tumor development.

Measuring mutations through next-generation sequencing

In 2002, the first drafts of the human genome sequence were published (*3*, *25*). Not long after, so-called next-generation sequencing platforms started to emerge, allowing massively-parallel whole-genome sequencing at reduced costs (*26–28*). These developments opened up avenues for studying mutation accumulation both prior to and after tumor initiation. We can now extract DNA from cells, determine the sequence of the bases in this DNA ("sequence") and compare ("map") the sequence of the genome to the reference genome. Differences in the sequence between a sample and the reference genome are flagged as variants or mutations.

There are several mutation types that can be distinguished in next-generation sequencing data, ranging from single nucleotide alterations to gross chromosomal alterations (*29*, *30*). The smallest mutations are base substitutions, in which single bases are substituted by another base, for example: a cytosine changes into an adenine (C > A). Single (or a small strings of) bases can also be deleted or duplicated, resulting in small insertions and deletions (indels). Larger deletions or duplications of at least 100 bases are called structural variations. DNA sequences can be inverted or translocated to another part of the genome as well, which are two other types of structural variations (*31*). Finally, changes in the number of chromosomes (aneuploidies) can also occur. All deletions and amplifications of genomic sequences change the copy number of a genomic region. These copy number variations are a distinct class of mutations in comparison to base substitutions, translocations, and inversions, which are balanced or copy-number neutral mutations.

The majority of the variants that are detected when a sample is mapped to the reference genome (at least 4-5 million positions) are in fact variations in the sequence of the human genome between individuals, so-called germline polymorphisms (*32–34*). These polymorphisms are for instance responsible for differences in phenotypic traits, like hair color (*35*, *36*), but they can also be responsible for susceptibility to diseases, like cardiovascular diseases and cancer (*37–40*). When we compare the mutations in a tumor genome to a the mutations in the genome of a healthy tissue

sample of the same individual, we can identify mutations that have accumulated in the cancer cells specifically (since the germline polymorphisms are present in the genomes of both samples, we can filter them out), which we call somatic mutations.

Tumor driver and passenger mutations

Tumor genomes have often accumulated many mutations (Figure 1) and, therefore, genome instability is considered to be one of the hallmarks of cancer (*8*, *10*). Depending on tumor type and age of onset, tumors can carry 32 up to (at least) 1,280,000 point mutations per genome (*41–43*). Melanomas have a notoriously unstable genome with a median of over 32,000 base substitutions per genome, whereas childhood cancers are typically quite stable with a median of less than 320 base substitutions per genome (*41*, *43*, *44*). In addition to base substitutions, tumor genomes often carry many copy number variations. Typically, ~30% of the genome is hit by a copy number variation in tumor cells (*45*). The majority of these duplications and deletions is small, at a median length of 1.8 million base pairs, but tumors also frequently loose and/or gain entire chromosome arms or chromosomes (*45*). Similar to base substitutions, the number of copy number variations differs depending on tumor type. For example, dedifferentiated liposarcomas carry an average of 120 copy number variations, whereas Myxoid liposarcomas typically carry less than 10 copy number variations (*45*). Even within one tumor type, there are substantial differences in mutational loads (*43*, *46*, *47*). These differences in mutational loads indicate that different processes underlie the development of each tumor.

The vast majority of the mutations detected in tumor genomes are so-called passenger mutations, which did not provide a selective advantage to cells and did not contribute to the development of cancer, but simply accumulated in a cell as neutral bystanders (*48*, *49*). In striking contrast to the high mutational load detected in tumor genomes, it is estimated that only 4 single nucleotide changes in a cell in the human body can drive the development of cancer (so-called tumor driver mutations) (*48–50*). Larger mutations, like structural variations and aneuploidies, can even have more tumorigenic potential, as they can encompass multiple genes (*51*). In line with this, aneuploidies frequently occur early during tumorigenesis and might even represent tumor driver mutations (*52–54*).

Tumor driver mutations can provide an outgrowth benefit to a cell, for example by increasing the proliferative rate or enabling cells to evade apoptosis (*29*, *50*). These driver mutations typically hit any of the ~400 known cancer driver genes (*55*), of which two subclasses are defined. Oncogenes are genes in which oncogenic gain-of-function mutations - predominantly base substitutions, duplications, or translocations - increase the protein levels or result in increased or constitutive

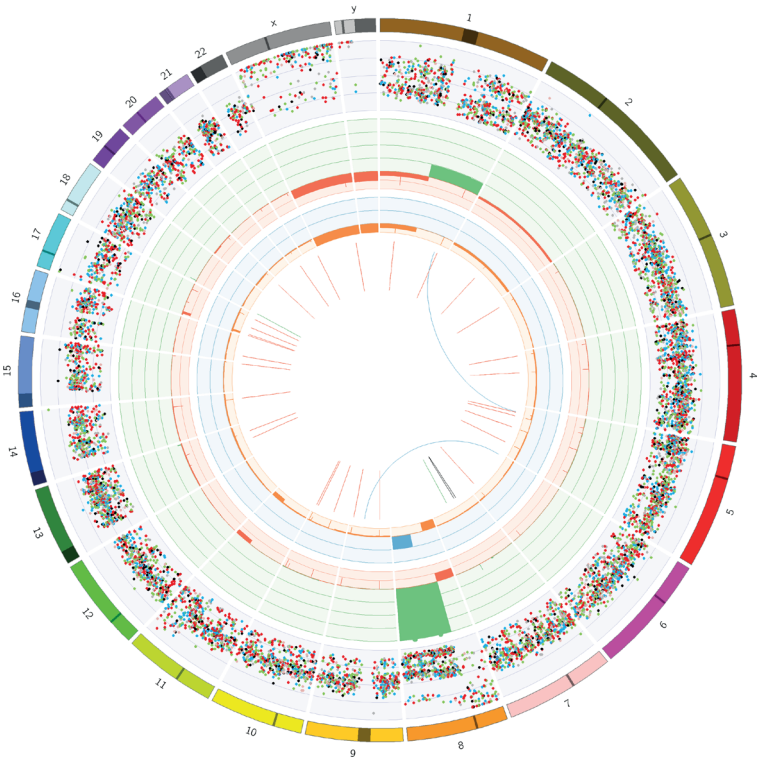


Figure 1. Circos plot depicting the mutational landscape of a liver cancer. The outer circle depicts the chromosomes. Each additional circle represents a mutation type: the inner circle shows the structural variations, in which red lines represent deletions, blue lines represent translocations, green lines represent duplications and black lines represent inversions. This particular tumor has acquired ~30 deletions, two translocations, two duplications, and one inversion. The next two circles depict the minor allele copy number and total copy number respectively. Blue and green blocks represent copy number gains and red and orange blocks represent copy number losses. This sample has 6 copies of one arm of chromosome 8, while it has only 1 copy of the other arm, for example. The next circle depicts the variant allele frequencies of the base substitutions and indels. Small insertions are shown in yellow, small deletions in red, and the colors of base substitutions are similar as was used in (*43*). This samples carries 10,596 base substitutions and 1,484 indels.

activity of proteins (*49*). Well-studied examples of oncogenes include *KRAS* and *c-MYC*. base substitutions in the *KRAS* gene are detected in 22% of all tumors and can result in constitutive activity of KRAS by changing a specific amino acid in the catalytic domain, which increases the proliferative rate of cells (*56*, *57*). Massive copy number amplifications of the *c-MYC* oncogene, which are also observed frequently in tumor genomes, can drive proliferation through increased expression of c-*MYC* (*58–60*). In contrast to oncogenes, tumor suppressors are usually hit by two (epi-) genomic mutations - e.g. base substitutions or deletions - that interfere with the

anti-oncogenic activity of these proteins (*49*). The best studied tumor-suppressor is, without a doubt, *TP53*. Mutations in this gene are detected in almost all tumor types at varying (5 - 80%) frequency (*61*, *62*). *TP53* mutations are also an important driver for childhood cancer and germline mutations in this gene increase cancer risk before the age of 30 by ~50 times (*41*).

The ability to detect driver events has shed light on the causes of cancer. We can start to dissect why a certain cell gave rise to a tumor in any patient by determining the driver mutations in their tumor. This has led to the identification of signaling pathways that are important in the development of cancer (*63*), which can be used to predict the prognosis of a patient (*64*, *65*). Furthermore, the ability to detect driver mutations and disturbed signaling pathways in tumors has had major implications on cancer treatment decisions. With every sequencing effort it is becoming more evident that each tumor carries a unique set of mutations. We can now specifically treat patients based on the flaws in their cancer cells, also known as targeted treatment, precision medicine, or personalized treatment (*66*, *67*). Leukemias that carry a *Bcr-Abl* fusion gene are now, for instance, successfully treated with Gleevec/imatinib (*68*).

Nevertheless, there are also tumors without mutations in a known cancer gene (*69*). This phenomenon is especially evident in childhood cancers, where currently less than 50% of the tumors harbor a mutation in a known cancer gene (*41*). The lack of presence of driver mutations may indicate that we have not yet discovered all cancer genes (*50*). In line with this, new cancer genes are still being discovered (*43*, *69*). Alternatively, changes on top of the genome sequence (the epigenome), which cannot be picked up by conventional genome sequencing, might play a pivotal role in the development of cancer, as these changes can alter the expression of proteins (*70*, *71*). Finally, the lack of driver mutations in some tumors also reflects the challenge of pinpointing the few driver mutations among the typically very high number of passenger mutations. This is complicated even further by the fact that it is not always possible to predict the effect of a genomic mutation on protein function or on activity of a signaling pathway.

Mutational processes

Although searching for driver mutations provides valuable insight into the pathways underlying tumorigenesis and the characteristics of the tumor, it is not always informative of the processes that contributed to driver mutation accumulation. Additionally, it can only help elucidate why cancer arose at the cellular level, but it is not providing answers as to why cancer arose at the organismal level. Hence, we still cannot answer the question why lifestyle choices, such as chronic alcohol

consumption, can contribute to cancer development. Another way of investigating why a cell became a cancer cell is by looking at all somatic (both driver and passenger) mutations that have accumulated in the genome, as these collectively reflect the mutational processes that have been active in a cell (*43*).

Genomes are continuously challenged by substances and processes that can damage the DNA (Figure 2) (*72*, *73*). Some of these mutagenic substances (mutagens) are produced by the cell itself (endogenous) as by-products of cellular processes, like reactive oxygen species (ROS) and alkylating agents (*72*, *74*, *75*). DNA polymerases can also introduce mutations in the genome during replication (*76*, *77*) and the genomic integrity is challenged by spontaneous chemical reactions as well, including spontaneous deamination of methylated cytosines (*78*). Finally, exogenous mutagens (mutagens that are not from the cell within) like tobacco smoke or chemotherapeutic drugs can damage the genome (*72*, *74*, *75*). Each of these mutagenic processes induce specific lesions in the DNA. For example, UV-light causes covalent linkage between two adjacent pyrimidines (thymines and/or cytosines), so-called pyrimidine-dimers (*79*, *80*). ROS, on the other hand, gives rise to the formation of 8-oxoguanine through oxidative damage of normal guanines (*81*–*83*). Some mutagens, like ionizing radiation, can even introduce double strand breaks in the chromosomes (*84*, *85*).

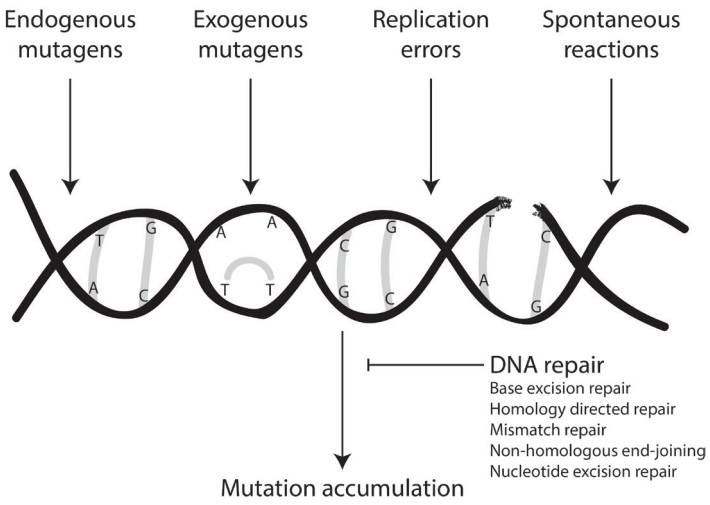


Figure 2. Mutation accumulation is a two-step process. The genome is challenged by various mutagenic processes, which induce lesions in the genome (e.g. thymine dimers or single-strand breaks). DNA repair pathways attempt to resolve these lesions and restore the genomic structure and sequence. If DNA repair pathways fail, or when a lesion escapes repair and becomes fixed during replication, a mutation can be introduced in the genome sequence.

As a consequence of the combination of spontaneous reactions and exposure to mutagens, the genome is estimated to acquire ~100,000 lesions each day (*72*, *78*, *86*). This means that each base in the human genome is hit by a lesion in any of the 3.7×10^{13} cells every 0.00007 seconds (*87*). Without any intrinsic defense mechanisms against mutagenic processes, cells would not be able to cope with the cellular stress that is inevitably associated with this high number of lesions. Cells therefore employ various DNA repair pathways to counteract mutagens, and to maintain a stable genome. The main DNA repair pathways are: Base excision repair (BER), Nucleotide excision repair (NER), Mismatch repair (MMR), Homology directed repair (HDR), and Non-homologous end joining (NHEJ) (*72*, *73*). Each DNA repair pathway is involved in the repair of specific lesions caused by a specific mutagenic process, although there is redundancy. HDR and NHEJ can both repair double-strand breaks in the helix (*72*, *73*). MMR is involved in removing mismatches that arise during replication (*73*). BER and NER can both repair lesions to single bases, for example as a consequence of oxidative damage (*72*, *73*). Finally, NER can also repair larger helix-distorting lesions (*72*, *73*).

Although the vast majority of the lesions is repaired correctly, some lesion escape repair and become fixed during replication or are incorrectly repaired, thereby introducing a mutation in the genome (*72*). Each mutation is thus a direct consequence of a combination of activity of a mutagenic process and a DNA repair pathway (Figure 2). As both of these processes act quite sequence-specific, somatic mutation catalogs can provide a functional readout of the mutational processes that have been active in cells.

Signatures of mutational processes

One can gain insight into the past activity of mutational processes in cells by looking at the mutation types. For example, deamination of methylated cytosines primarily introduces C > T changes, whereas ROS cause C > A changes (*43*, *81*). When a genome shows an increased accumulation of C > A changes, this may indicate that a cell has endured higher ROS levels. However, many more mutational processes are obviously active than can be identified by just looking at the six mutation types. For instance, both spontaneous deamination of methylated cytosines and activity of APOBEC cytidine deaminases can introduce C > T changes in the genome (*43*). Several additional features, therefore, need to be taken into account to fully distinguish the mutational processes that have been active in cells, such as: the genomic location of the mutation, the size of the mutation, transcriptional strand bias, replication strand asymmetry, and the sequence context of the mutated bases (*43*, *69*).

In 2013, systematic analyses of tumor genomes identified 21 “signatures” of

mutational processes. These mutational signatures are based on base substitution types and the immediate sequence context of the base substitution (mutational profiles) (*43*). A few years later, the number of signatures was expanded to 30 signatures of base substitutions and 6 rearrangement signatures of mutational processes in cancer genomes (*69*, *88*). In the near future, a new release will most likely involve 49 signatures of base substitutions, 11 signatures of tandem base substitutions, and 17 indel signatures (*89*). The underlying etiology of some of the signatures is known (Figure 3). For example, a high contribution of Signature 4 mutations in the genome is associated with exposure to tobacco smoking, while a high contribution of Signature 29 is associated with exposure to tobacco chewing (*43*). Signature 7 is characterized by C > T mutations at TpCpN (N = any base) trinucleotides and has been linked to exposure to UV-light (*43*). Signature 1, on the other hand, is characterized by C > T mutations at NpCpG trinucleotides and has been linked to spontaneous deamination of methylated cytosines (*43*). In addition to mutagen exposure, some of the signatures have been linked to deficiency of certain DNA repair pathways. Signature 3 is reflective of HDR-deficiency (*90*, *91*) and many signatures display a transcriptional strand bias in gene bodies, which could indicate that the underlying mutagenic damage can be repaired by transcription-coupled NER (*43*). Furthermore, several signatures have been linked to MMR-deficiency (*92*).

Knowledge on the activity of mutational processes and characteristics of associated mutation patterns could potentially improve the treatment choice for cancer patients, as it can provide information on which DNA repair pathways may have been (in)active (*94*). As mentioned, patients with a deficiency in HDR have a high contribution of mutation Signature 3 (*90*, *91*). Presence of this signature could therefore indicate that these patients might respond well to PARP inhibitors or platinum drugs, which are therapies that have been shown to be effective in HDR-deficient tumors (*95*–*98*). Mutational patterns can also create insight into which drugs should *not* be used to treat a patient. Children with a genetic predisposition to cancer due to an inherited mutation in a DNA repair gene should not be treated with a drug that specifically targets this pathway, as this would sensitize all cells in their body to the treatment, and increases the chance of general cytotoxic effects. Taken together, if we identify which DNA repair pathways are defective in a tumor, this can further guide treatment decision.

In addition to providing new treatment options, these signatures have enabled us to gain more insight into the mutational processes that are active before, during, and after oncogenic transformation. However, it is difficult to distinguish the mutations that occurred prior to oncogenesis from the mutations that occurred within a tumor. Furthermore, in-depth mechanistic insight is still lacking for the

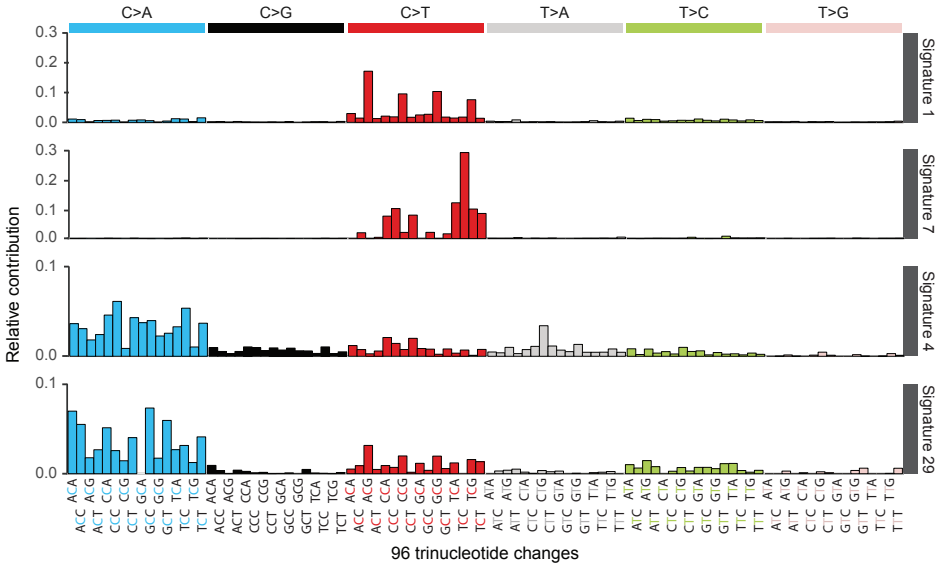


Figure 3. Mutational profiles of 4 mutational signatures that have been associated with mutagenic processes. Each graph shows the relative contribution of the indicated context-dependent mutation types to the mutational profile. Signature 1 and 7, associated with deamination of methylated cytosines and exposure to UV-light respectively, both have a high contribution of C>T changes. However, the 96-type mutational profile shows that the sequence around the mutated base differs substantially. Signature 4 and 29 are both caused by exposure to tobacco, but some small differences in the relative contribution of these mutation types allows us to distinguish tobacco smoking (Signature 4) from tobacco chewing (Signature 29). Figure generated from https://cancer.sanger.ac.uk/cosmic/signatures using the R package MutationalPatterns (*93*).

majority of the mutational signatures (*92*, *99*) and efforts to identify the underlying etiology of mutational signatures have mainly been focused on linking tumor mutation data to DNA repair-deficiency/mutagen exposure. Yet, tumor genomes are typically highly unstable (*8*, *10*) and multiple mutational processes have been active in a cell's life history (*43*, *69*), which complicates associating certain processes to these mutational signatures. Measuring mutations prior to oncogenesis in the cell-of-origin of cancer can help identify which mutational processes underlie which signatures under normal, homeostatic conditions and can ultimately help elucidate which mutational processes contribute to tumorigenesis.

Adult stem cells

Tissue-specific adult stem cells (ASCs) are believed to be a cells-of-origin of cancer (*100–102*). These cells are multipotent stem cells that reside in an organ and can change (differentiate) into other cell types in the organ. In the small intestine, for example, highly proliferative stem cells compensate for the loss of differentiated

epithelial cells by producing additional differentiated cells (*103*). Although genome maintenance is important in all cells, it is especially crucial to maintain genomic stability in stem cells, as these cells are long-lived and can pass on mutations that they acquired in their genome to a high number of progeny cells, including daughter stem cells. Mutations in the genomes of ASCs may therefore have a larger effect on tissue fitness than mutations in differentiated cells. Consistently, mutations in the genomes of ASCs are not only believed to contribute to the development of cancer, but DNA damage in the genomes of ASCs has also been shown to contribute to dysfunctioning and depletion of ASCs in tissues, one of the hallmarks of aging (*9*, *22*, *72*, *102*, *104*).

For a long time, it was difficult to study genome stability in tissue-specific ASCs, due to a lack of specific markers to enrich for human stem cells. Recently, however, a new culturing technique was developed that allows enrichment of ASCs *in vitro* (Figure 4) (*105*, *106*). In short, single LGR5+ ASCs can be grown in Matrigel or BME as 3D organoids (mini-organs). By adding components and growth factors like Wnt3A, R-spondin-1, and EGF to the culture medium, the niche (microenvironment) of the ASCs is mimicked, which allows the selection and the expansion of stem cells specifically. Addition of other components to the culture medium, such as Nicotinamide, allows long-term expansion of these organoid cultures (*105*). The organoid technology was first developed for human ASCs of the colon and small intestine (*105*, *106*), but has been adapted to other tissues as well, including liver (*107*, *108*), prostate (*109*), pancreas (*108*), stomach (*110*), fallopian tube (*111*), and breast (*112*). Furthermore, it is also possible to culture organoids from tumor biopsies (tumoroids) of several tissues, including pancreatic cancer (*113*), breast cancer (*112*), liver cancer (*114*), and prostate cancer (*109*).

Organoids hold great promise for (fundamental) research and have several clinical applications (*115–117*). Firstly, organoids can be used for regenerative medicine purposes, as they provide us with a (seemingly) limitless cell source that can be genetically modified and differentiated into the desired cell type (*115–117*). Secondly, organoids can be used to study and predict drug responses (*115*). The response of intestinal organoids derived from cystic fibrosis (CF) patients to anti-CF treatment is, for example, indicative of the patient's drug response (*118*). Similarly, by monitoring the effect of anti-tumor drugs on patient-derived tumor organoid cultures, it might be possible to predict drug response in a patient-specific manner. Simultaneous treatment of patient-derived healthy organoid cultures might even facilitate the prediction of unwanted side-effects on healthy adjacent tissue, which can be used to adapt treatment dosage. Moreover, organoids can facilitate systematic drug screens to determine the correlation between presence

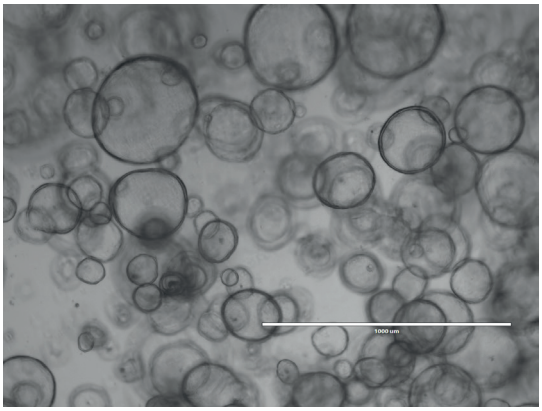


Figure 4. Microscope image of a human small intestinal organoid culture. The large spheres are organoids, consisting of multiple cells. Since the organoids grow in 3D in Matrigel, some of them are out of focus.

of mutational signatures and drug responses (*112*). Finally, organoids provide a novel tool for studying oncogenesis. Sequential induction of oncogenic mutations in healthy intestinal organoid cultures, for example, showed that loss of *APC* and *TP53* is sufficient to induce aneuploidies (*119*). Furthermore, knock-out of DNA repair genes by using CRISPR-Cas9 in organoid cultures can provide insight into the mutational mechanisms that underlie tumorigenesis (*120*).

Hypotheses on the origin of cancer

Several hypotheses have already been generated on why risk factors can contribute to the development of cancer. Firstly, some of the risk factors can cause an increase in the genome-wide mutation rate, thereby increasing the mere chance of hitting a driver mutation, which can be sufficient for tumor formation (*101*). For example, melanoma genomes show a high number of C > T changes (*121*, *122*), which can be caused by exposure to UV-light (*79*, *80*). The majority of the tumor driver mutations in melanomas are C > T changes as well (*122*), suggesting that UV-light-induced mutagenesis underlies the accumulation of (some of) these driver mutations. Similarly, a reduced potential to resolve lesions induced by UV-light also predisposes individuals to the development of cancer. Xeroderma pigmentosum patients are at high risk of developing skin cancer, due to germline mutations in genes involved in NER, which is the main DNA repair pathway involved in repairing pyrimidine dimers (*72*, *123–126*).

However, an increased mutational load is not sufficient to drive cancer development in all tissues (*101*, *127*). In addition to increasing the mutational load, cellular selection by the micro-environment might play a role in tumor formation. Healthy ASCs grow faster than tumor cells in the organoid cultures (*115*), suggesting that there is some type of selection against tumor cells in this normal niche-mimicking environment. However, by changing the culture conditions, we can favour

the outgrowth of the tumor organoids (*115*, *128*). Exposure to risk factors might also induce changes in the cellular micro-environment of cells, which could ultimately reduce the selection against, or might even favour the clonal outgrowth of, cells with oncogenic mutations. Consistently, the microenvironment is known to play a tumor-promoting role after tumor initiation (*129*). Inflammation in the skin due to exposure to UV-light, for example, promotes the formation of metastases, irrespective of the mutagenic potential of UV-light (*130*).

Finally, although not the focus of this thesis, it should be noted that changes on the DNA, other than changes in the sequence itself, are also known to contribute to tumorigenesis (*70*, *71*). The epigenome (meaning 'on top of the genome') can regulate gene expression and, for instance, plays a crucial role in establishing and maintaining cell-type specific gene expression profiles from the same genome in distinct cell types. Changes in the epigenome can therefore change gene expression. In cancer cells, it has been shown that enhanced methylation of the promoter of tumor-suppressor genes is associated with a reduced expression of these genes, which may contribute to cancer development (*70*, *71*). Epigenetic changes can also influence genome stability, both directly and indirectly. Spontaneous deamination of methylated cytosines causes G:T mismatches which can result in C>T changes at NpCpG sequences (*88*). This mutational process is believed to underlie Signature 1, of which significant contribution is found in the genomes of almost all tumor types, including in cancer driver genes (*43*). Epigenetic modifications can indirectly affect genome stability by changing the expression of DNA repair proteins, and thereby reducing the DNA repair capacity of a cell. Epigenetic inactivation of the MMR gene *MLH1* has, for example, been shown to decrease the DNA repair capacity of MMR in colorectal cancer, which causes microsatellite instability (*131–133*), and hypermethylation of the promoter of DNA repair gene *MGMT* can result in oncogenic mutations in the *KRAS* oncogene (*134*). Potentially, exposure to risk factors can change the epigenetic landscape of cells, thereby actively contributing to tumor formation.

These three hypotheses can explain why mutagen exposure and DNA repair deficiency can cause cancer. In fact, multiple of these mechanisms probably act cooperatively in the development of each tumor. Nevertheless, a lot of work remains to be done. For the majority of mutagens, it is still unclear why these increase the chance of developing cancer. Furthermore, the mutational consequences of deficiency of many DNA repair pathways have not been identified yet, although not due to a lack of effort (*135*, *136*). Systematic analyses of the mutational processes in tissue-specific ASCs prior to tumor initiation could help shed light onto the origin of cancer.

Thesis outline

The research presented in this thesis is a combination of the development of novel methods (**Chapter 2 and 6**) and the utilization of these methods to measure mutation accumulation in the genomes of various ASC types (**Chapter 3, 4, and 5**). **Chapter 2** describes a protocol to catalogue mutations that have accumulated in the genome of single ASCs during life or during culturing (Figure 5). In this protocol, the organoid culturing technique is utilized to select and to clonally expand single ASCs, until sufficient DNA can be obtained to perform whole-genome sequencing and to reliably measure mutations present in the cell-of-origin. The protocol is described for human liver but can be adapted to any tissue of which organoids can be generated. The protocol described in Chapter 2 is subsequently used in Chapters 3, 4, and 5 to study mutation accumulation in various ASC types. In **Chapter 3**, we compare mutation accumulation in human ASCs from liver, small intestine, and colon during life. **Chapter 4** describes the mutational consequences of loss of a single DNA repair pathway (NER) on the genomic stability of mouse ASCs during life and human ASCs during culturing. In **Chapter 5**, the mutational consequences of alcohol abuse on human ASCs of the liver are studied. Finally, in **Chapter 6**, we developed a new

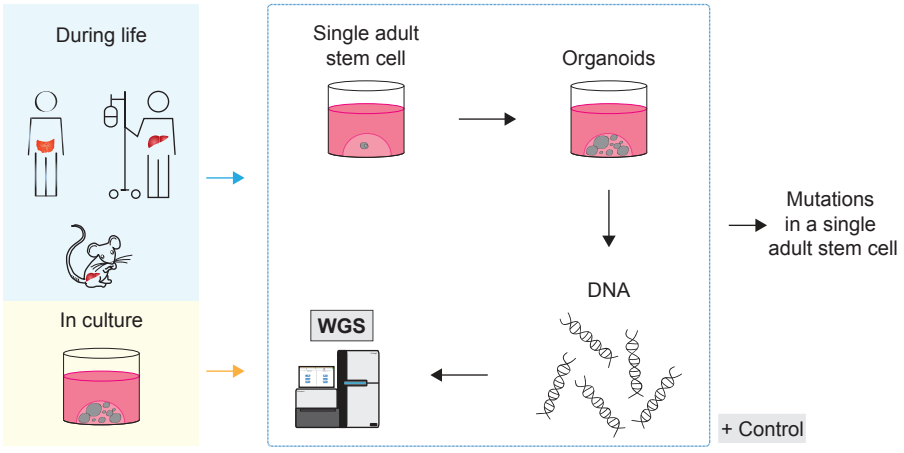


Figure 5. Schematic overview of the technique described in **Chapter 2,** and used in **Chapters 3, 4, and 5**. Single stem cells can be isolated from healthy human tissues, tissues of patients, tissues of disease models, and organoid cultures. These stem cells are expanded as organoid cultures, until sufficient material is generated to perform whole-genome sequencing. All mutations present in the cell-of-origin of this culture will be present in all cells, and therefore we can pick these up reliably. Mutations that occurred during the expansion from a single cell to an organoid culture can be filtered out, based on a low occurrence of this mutation in the entire organoid culture. By sequencing a control sample, we can identify and exclude germline variants. This technique enables the measurement of mutations that accumulated in the genome of a single ASC during life and in culture with high accuracy.

organoid culturing technique for human and mouse pancreatic and pancreatic cancer organoids. In the future, this new organoid culture could be used to study mutation accumulation in pancreatic cancer. These chapters combined provide novel insight into mutation accumulation in ASCs prior to oncogenic transformation. Furthermore, the consequences of several mutational processes are elucidated, which can improve our understanding of tumorigenesis and ultimately may contribute to better cancer prevention and treatment strategies.

ACKNOWLEDGEMENTS

The authors thank Melvin van Staalduinen and Roos Jager for providing (textual) comments.

REFERENCES

1. A. Hiyoshi, K. Miyahara, C. Kato, Y. Ohshima, Does a DNA-less cellular organism exist on Earth? *Genes Cells*. **16**, 1146–1158 (2011).
2. J. D. Watson, F. H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. **171**, 737–738 (1953).
3. E. S. Lander *et al.*, Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921 (2001).
4. E. S. Lander, Initial impact of the sequencing of the human genome. *Nature*. **470**, 187–197 (2011).
5. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature*. **431**, 931–945 (2004).
6. C. Willyard, New human gene tally reignites debate. *Nature*. **558**, 354–355 (2018).
7. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature*. **489**, 57–74 (2012).
8. D. Hanahan, R. A. Weinberg, The Hallmarks of Cancer. *Cell*. **100**, 57–70 (2000).
9. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The Hallmarks of Aging. *Cell*. **153**, 1194–1217 (2013).
10. D. Hanahan, R. A. Weinberg, Hallmarks of Cancer: The Next Generation. *Cell*. **144**, 646–674 (2011).
11. D. Hanahan, J. Folkman, Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell*. **86**, 353–364 (1996).
12. G. P. Gupta, J. Massagué, Cancer metastasis: building a framework. *Cell*. **127**, 679–695 (2006).
13. J. Ferlay *et al.*, Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*. **136**, E359–E386 (2014).
14. B. W. Stewart, C. P. Wild, *World Cancer Report 2014* (2014).
15. M. Quaresma, M. P. Coleman, B. Rachet, 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971–2011: a population-based study. *Lancet*. **385**, 1206–1218 (2015).
16. F. Bray, A. Jemal, N. Grey, J. Ferlay, D. Forman, Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. *Lancet Oncol.* **13**, 790–801 (2012).
17. V. Bouvard *et al.*, A review of human carcinogens—Part B: biological agents. *Lancet Oncol.* **10**, 321–322 (2009).
18. B. Secretan *et al.*, A review of human carcinogens—Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *Lancet Oncol.* **10**, 1033–1034 (2009).
19. F. El Ghissassi *et al.*, A review of human carcinogens—Part D: radiation. *Lancet Oncol.* **10**, 751–752 (2009).
20. E. R. Fearon, Human cancer syndromes: clues to the origin and nature of cancer. *Science*. **278**, 1043–1050 (1997).

21. G. Danaei *et al.*, Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet*. **366**, 1784–1793 (2005).

22. C. Tomasetti, B. Vogelstein, Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*. **347**, 78–81 (2015).

23. J. Ferlay *et al.*, Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer*. **136**, E359–86 (2015).

24. Cancer Statistics Review, 1975-2015 - SEER Statistics, (available at https://seer.cancer.gov/csr/1975_2015/).

25. J. C. Venter *et al.*, The sequence of the human genome. *Science*. **291**, 1304–1351 (2001).

26. M. Barba, H. Czosnek, A. Hadidi, Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses*. **6**, 106–136 (2014).

27. M. Margulies *et al.*, Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. **437**, 376–380 (2005).

28. J. Shendure *et al.*, Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*. **309**, 1728–1732 (2005).

29. M. R. Stratton, P. J. Campbell, P. A. Futreal, The cancer genome. *Nature*. **458**, 719–724 (2009).

30. J. A. Veltman, H. G. Brunner, De novo mutations in human genetic disease. *Nat. Rev. Genet.* **13**, 565–575 (2012).

31. C. Alkan, B. P. Coe, E. E. Eichler, Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12**, 363–376 (2011).

32. 1000 Genomes Project Consortium *et al.*, A global reference for human genetic variation. *Nature*. **526**, 68–74 (2015).

33. R. E. Mills *et al.*, An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* **16**, 1182–1190 (2006).

34. A. J. Iafrate *et al.*, Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).

35. E. A. Grimes, P. J. Noake, L. Dixon, A. Urquhart, Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci. Int.* **122**, 124–129 (2001).

36. P. G. Hysi *et al.*, Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat. Genet.* (2018), doi:10.1038/s41588-018-0100-5.

37. K. Ozaki, T. Tanaka, Genome-wide association study to identify SNPs conferring risk of myocardial infarction and their functional analyses. *Cell. Mol. Life Sci.* **62**, 1804–1813 (2005).

38. J. M. Rodríguez-Pérez *et al.*, HHIPL-1 (rs2895811) gene polymorphism is associated with cardiovascular risk factors and cardiometabolic parameters in Mexicans patients with myocardial infarction. *Gene* (2018), doi:10.1016/j.gene.2018.04.030.

39. M. Podralska *et al.*, Genetic variants in ATM, H2AFX and MRE11 genes and susceptibility to breast cancer in the polish population. *BMC Cancer*. **18**, 452 (2018).

40. S. N. Stacey *et al.*, A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat. Genet.* **43**, 1098–1103 (2011).

41. S. N. Gröbner *et al.*, The landscape of genomic alterations across childhood cancers. *Nature*. **555**, 321–327 (2018).

42. X. Ma *et al.*, Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*. **555**, 371–376 (2018).

43. L. B. Alexandrov *et al.*, Signatures of mutational processes in human cancer. *Nature*. **500**, 415–421 (2013).

44. M. S. Lawrence *et al.*, Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. **499**, 214–218 (2013).

45. R. Beroukhim *et al.*, The landscape of somatic copy-number alteration across human cancers. *Nature*. **463**, 899–905 (2010).

46. S. F. Roerink *et al.*, Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature*. **556**, 457–462 (2018).

47. A. Fujimoto *et al.*, Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500–509 (2016).

48. S. De, S. Ganesan, Looking beyond drivers and passengers in cancer genome sequencing data. *Ann. Oncol.* **28**, 938–945 (2017).

49. D. A. Haber, J. Settleman, Cancer: drivers and passengers. *Nature*. **446**, 145–146 (2007).

50. I. Martincorena *et al.*, Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*. **171**, 1029–1041.e21 (2017).

51. F. Blokzijl *et al.*, Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*. **538**, 260–264 (2016).

52. P. Duesberg, R. Li, Multistep carcinogenesis: a chain reaction of aneuploidizations. *Cell Cycle*. **2**, 202–210 (2003).

53. B. Vogelstein *et al.*, Genetic Alterations during Colorectal-Tumor Development. *N. Engl. J. Med.* **319**, 525–532 (1988).

54. Z. Kan *et al.*, Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* **23**, 1422–1433 (2013).

55. P. A. Futreal *et al.*, A census of human cancer genes. *Nat. Rev. Cancer*. **4**, 177–183 (2004).

56. B. Stolze, S. Reinhart, L. Bulllinger, S. Fröhling, C. Scholl, Comparative analysis of KRAS codon 12, 13, 18, 61, and 117 mutations using human MCF10A isogenic cell lines. *Sci. Rep.* **5**, 8535 (2015).

57. I. A. Prior, P. D. Lewis, C. Mattos, A comprehensive survey of Ras mutations in cancer. *Cancer Res.* **72**, 2457–2467 (2012).

58. T. I. Zack *et al.*, Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).

59. K. J. Wu *et al.*, Direct activation of TERT transcription by c-MYC. *Nat. Genet.* **21**, 220–224 (1999).

60. Z. E. Stine, Z. E. Walton, B. J. Altman, A. L. Hsieh, C. V. Dang, MYC, Metabolism, and Cancer. *Cancer Discov.* **5**, 1024–1039 (2015).

61. M. Olivier, M. Hollstein, P. Hainaut, TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.* **2**, a001008 (2010).

62. A. Petitjean *et al.*, Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mutat.* **28**, 622–629 (2007).

63. R. Sever, J. S. Brugge, Signal transduction in cancer. *Cold Spring Harb. Perspect. Med.* **5** (2015), doi:10.1101/cshperspect.a006098.

64. A. F. T. Ribeiro *et al.*, Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia. *Blood*. **119**, 5824–5831 (2012).

65. A. M. Vannucchi *et al.*, Mutations and prognosis in primary myelofibrosis. *Leukemia*. **27**, 1861–1869 (2013).

66. A. Ashworth, C. J. Lord, J. S. Reis-Filho, Genetic Interactions in Cancer Progression and Treatment. *Cell.* **145**, 30–38 (2011).

67. L. C. Brody, Treating Cancer by Targeting a Weakness. *N. Engl. J. Med.* **353**, 949–950 (2005).

68. M. W. N. Deininger, Specific Targeted Therapy of Chronic Myelogenous Leukemia with Imatinib. *Pharmacol. Rev.* **55**, 401–423 (2003).

69. S. Nik-Zainal *et al.*, Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. **534**, 47–54 (2016).

70. P. A. Jones, S. B. Baylin, The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).

71. A. P. Feinberg, M. A. Koldobskiy, A. Göndör, Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat. Rev. Genet.* **17**, 284–299 (2016).

72. J. H. J. Hoeijmakers, DNA damage, aging, and cancer. *N. Engl. J. Med.* **361**, 1475–1485 (2009).

73. T. Iyama, D. M. Wilson 3rd, DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair* . **12**, 620–636 (2013).

74. J. A. Swenberg *et al.*, Endogenous versus Exogenous DNA Adducts: Their Role in Carcinogenesis, Epidemiology, and Risk Assessment. *Toxicol. Sci.* **120**, S130–S145 (2010).

75. R. C. Gupta, W. K. Lutz, Background DNA damage for endogenous and unavoidable exogenous carcinogens: a basis for spontaneous cancer incidence? *Mutat. Res.* **424**, 1–8 (1999).

76. P. V. Shcherbakova *et al.*, Unique error signature of the four-subunit yeast DNA polymerase epsilon. *J. Biol. Chem.* **278**, 43770–43780 (2003).

77. J. M. Fortune *et al.*, Saccharomyces cerevisiae DNA polymerase delta: high fidelity for base substitutions but lower fidelity for single- and multi-base deletions. *J. Biol. Chem.* **280**, 29980–29987 (2005).

78. T. Lindahl, Instability and decay of the primary structure of DNA. *Nature*. **362**, 709–715 (1993).

79. G. P. Pfeifer, Y.-H. You, A. Besaratinia, Mutations induced by ultraviolet light. *Mutat. Res./Fundam. Mol. Mech. Mutag.* **571**, 19–31 (2005).

80. G. P. Pfeifer, Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment. *Photochem. Photobiol.* **65**, 270–283 (1997).

81. A. P. Grollman, M. Moriya, Mutagenesis by 8-oxoguanine: an enemy within. *Trends Genet.* **9**, 246–249 (1993).

82. B. Epe, Genotoxicity of singlet oxygen. *Chem. Biol. Interact.* **80**, 239–260 (1991).

83. B. van Loon, E. Markkanen, U. Hübscher, Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. *DNA Repair* . **9**, 604–616 (2010).

84. L. H. Thompson, Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. *Mutat. Res.* **751**, 158–246 (2012).

85. A. Mehta, J. E. Haber, Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* **6**, a016428 (2014).

86. G. A. Garinis, G. T. J. van der Horst, J. Vijg, J. H. J. Hoeijmakers, DNA damage and ageing: new-age ideas for an age-old problem. *Nat. Cell Biol.* **10**, 1241–1247 (2008).

87. E. Bianconi *et al.*, An estimation of the number of cells in the human body. *Ann. Hum. Biol.* **40**, 463–471 (2013).

88. L. B. Alexandrov *et al.*, Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).

89. L. Alexandrov *et al.*, The Repertoire of Mutational Signatures in Human Cancer (2018), , doi:10.1101/322859.

90. H. Davies *et al.*, HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 (2017).

91. P. Polak *et al.*, A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat. Genet.* **49**, 1476–1486 (2017).

92. M. Petljak, L. B. Alexandrov, Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis*. **37**, 531–540 (2016).

93. F. Blokzijl, R. Janssen, R. van Boxtel, E. Cuppen, MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).

94. B. B. Campbell *et al.*, Comprehensive Analysis of Hypermutation in Human Cancer. *Cell.* **171**, 1042–1056.e10 (2017).

95. H. Farmer *et al.*, Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*. **434**, 917–921 (2005).

96. H. E. Bryant *et al.*, Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*. **434**, 913–917 (2005).

97. P. C. Fong *et al.*, Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* **361**, 123–134 (2009).

98. M. L. Telli *et al.*, Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clin. Cancer Res.* **22**, 3764–3773 (2016).

99. L. B. Alexandrov, M. R. Stratton, Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet.*

Dev. **24**, 52–60 (2014).

100. N. Barker *et al.*, Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature*. **457**, 608–611 (2009).

101. L. Zhu *et al.*, Multi-organ Mapping of Cancer Risk. *Cell*. **166**, 1132–1146.e7 (2016).

102. P. D. Adams, H. Jasper, K. L. Rudolph, Aging-Induced Stem Cell Mutations as Drivers for Disease and Cancer. *Cell Stem Cell*. **16**, 601–612 (2015).

103. N. Barker *et al.*, Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature*. **449**, 1003–1007 (2007).

104. D. J. Rossi, C. H. M. Jamieson, I. L. Weissman, Stems Cells and the Pathways to Aging and Cancer. *Cell*. **132**, 681–696 (2008).

105. T. Sato *et al.*, Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology*. **141**, 1762–1772 (2011).

106. T. Sato *et al.*, Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature*. **459**, 262–265 (2009).

107. M. Huch *et al.*, In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature*. **494**, 247–250 (2013).

108. L. Broutier *et al.*, Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. *Nat. Protoc.* **11**, 1724–1743 (2016).

109. J. Drost *et al.*, Organoid culture systems for prostate epithelial and cancer tissue. *Nat. Protoc.* **11**, 347–358 (2016).

110. S. Bartfeld *et al.*, In Vitro Expansion of Human Gastric Epithelial Stem Cells and Their Responses to Bacterial Infection. *Gastroenterology*. **148**, 126–136.e6 (2015).

111. M. Kessler *et al.*, The Notch and Wnt pathways regulate stemness and differentiation in human fallopian tube organoids. *Nat. Commun.* **6**, 8989 (2015).

112. N. Sachs *et al.*, A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity. *Cell*. **172**, 373–386.e10 (2018).

113. S. F. Boj *et al.*, Organoid models of human and mouse ductal pancreatic cancer. *Cell*. **160**, 324–338 (2015).

114. L. Broutier *et al.*, Human primary liver cancer-derived organoid cultures for disease modeling and drug screening. *Nat. Med.* **23**, 1424–1435 (2017).

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

116. J. Drost, H. Clevers, Translational applications of adult stem cell-derived organoids. *Development*. **144**, 968–975 (2017).

117. T. Sato, H. Clevers, Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. *Science*. **340**, 1190–1194 (2013).

118. J. F. Dekkers *et al.*, A functional CFTR assay using primary cystic fibrosis intestinal organoids. *Nat. Med.* **19**, 939–945 (2013).

119. J. Drost *et al.*, Sequential cancer mutations in cultured human intestinal stem cells. *Nature*. **521**, 43–47 (2015).

120. J. Drost *et al.*, Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science*. **358**, 234–238 (2017).

121. M. F. Berger *et al.*, Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature*. **485**, 502–506 (2012).

122. E. Hodis *et al.*, A landscape of driver mutations in melanoma. *Cell*. **150**, 251–263 (2012).

123. L. Daya-Grosjean, C. Robert, C. Drougard, H. Suarez, A. Sarasin, High mutation frequency in ras genes of skin tumors isolated from DNA repair deficient xeroderma pigmentosum patients. *Cancer Res.* **53**, 1625–1629 (1993).

124. K. H. Kraemer, M. M. Lee, A. D. Andrews, W. C. Lambert, The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer. The xeroderma pigmentosum paradigm. *Arch. Dermatol.* **130**, 1018–1021 (1994).

125. G. Yang, D. Curley, M. W. Bosenberg, H. Tsao, Loss of xeroderma pigmentosum C (Xpc) enhances melanoma photocarcinogenesis in Ink4a-Arf-deficient mice. *Cancer Res.* **67**, 5649–5657 (2007).

126. C. Li *et al.*, Polymorphisms in the DNA repair genes XPC, XPD, and XPG and risk of cutaneous melanoma: a case-control analysis. *Cancer Epidemiol. Biomarkers Prev.* **15**, 2526–2532 (2006).

127. I. Martincorena *et al.*, Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. **348**, 880–886 (2015).

128. M. van de Wetering *et al.*, Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell*. **161**, 933–945 (2015).

129. T. L. Whiteside, The tumor microenvironment and its role in promoting tumor growth. *Oncogene*. **27**, 5904–5912 (2008).

130. T. Bald *et al.*, Ultraviolet-radiation-induced inflammation promotes angiotropism and metastasis in melanoma. *Nature*. **507**, 109–113 (2014).

131. J. G. Herman *et al.*, Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6870–6875 (1998).

132. X. Li *et al.*, MLH1 promoter methylation frequency in colorectal cancer patients and related clinicopathological and molecular features. *PLoS One*. **8**, e59064 (2013).

133. K. Imai, H. Yamamoto, Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. *Carcinogenesis*. **29**, 673–680 (2008).

134. M. Esteller *et al.*, Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is associated with G to A mutations in K-ras in colorectal tumorigenesis. *Cancer Res.* **60**, 2368–2371 (2000).

135. X. Zou *et al.*, Validating the concept of mutational signatures with isogenic cell models. *Nat. Commun.* **9**, 1744 (2018).

136. B. Meier *et al.*, C. elegans whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome Res.* **24**, 1624–1636 (2014).

Zooming in on mutations

Chapter 2

Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures

Myrthe Jager[1,#], Francis Blokzijl[1,#], Valentina Sasselli[2], Sander Boymans[1], Roel Janssen[1], Nicolle Besselink[1], Hans Clevers[2], Ruben van Boxtel[1,3,†] and Edwin Cuppen[1,†]

1 Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
2 Hubrecht Institute for Developmental Biology and Stem Cell Research, KNAW and University Medical Center Utrecht, Utrecht, the Netherlands
3 Princess Máxima Center for Pediatric Oncology, 3584 CT Utrecht, The Netherlands
# These authors contributed equally to this work
† These authors contributed equally to this work

Adapted from: Nature Protocols 2018 Jan; 13(1): 59-78

ABSTRACT

Characterization of mutational processes in adult stem cells (ASCs) will improve our understanding of aging-related diseases, such as cancer and organ failure, and may ultimately help prevent the development of these diseases. Here, we present a method for cataloging mutations in individual human ASCs without the necessity of using error-prone whole-genome amplification. Single ASCs are expanded *in vitro* into clonal organoid cultures to generate sufficient DNA for accurate whole-genome sequencing (WGS) analysis. We developed a data-analysis pipeline that identifies with high confidence somatic variants that accumulated *in vivo* in the original ASC. These genome-wide mutation catalogs are valuable resources for the characterization of the underlying mutational mechanisms. In addition, this protocol can be used to determine the effects of culture conditions or mutagen exposure on mutation accumulation in ASCs *in vitro*. Here, we describe a protocol for human liver ASCs that can be completed over a period of 3–4 months with hands-on time of ~5 d.

INTRODUCTION

Mutagenic processes continuously challenge the genomic integrity of cells, resulting in the accumulation of somatic mutations during life[1]. Mutations acquired in the genomes of long-lived ASCs are thought to have the largest impact on the fitness of a tissue, as these mutations are propagated to both self-renewing and differentiating progeny cells. Consistently, many have proposed that ASCs are the cells of origin in cancer[2,3,4]. Characterization of the processes that drive mutation accumulation in normal ASCs is therefore pivotal to understanding tissue homeostasis and the development of aging-associated diseases such as cancer and organ failure.

To measure somatic mutations in physiologically normal ASCs, several challenges must be addressed. First, as most healthy tissues are polyclonal, each somatic event is present in only a small population of cells, and therefore, it is not possible to detect these events through bulk sequencing. To quantitatively determine all somatic mutations that accumulate during life, single cells must be assessed. Second, as the majority of an organ is composed of differentiated cells, ASCs must be selected before sequencing. However, antibodies that allow the enrichment of living human ASCs through cell sorting have not yet been generated for most tissue types, and, therefore, different isolation methods are required. Third, as mutations are usually distributed nonrandomly across the genome[1], it is important to use an unbiased sequencing technique to measure complete and representative mutational patterns in single ASCs. Finally, the somatic mutation load in ASCs is typically low[1,5]. Hence, it is important to use a sequencing technique that can sequence genomes of single cells without introducing high numbers of sequencing errors or biases

associated with whole-genome amplification methods, which might overshadow the low numbers of somatic mutations in ASCs.

Recently, a 3D culture technology was developed that facilitates long-term *in vitro* expansion of primary ASCs as organoid cultures[6]. Here, we describe a protocol that uses the organoid culture technology to specifically select single ASCs, clonally expand these cells *in vitro* and perform WGS analysis. This protocol allows, for the first time, measurement of genome-wide mutations that have accumulated in single ASCs of different human tissues, with a high confirmation rate (~91%)[1]. Genome-wide mutation patterns can be subsequently characterized to identify and study the mutagenic and DNA repair processes that have shaped the observed mutational landscapes[7,8]. The entire protocol, described here using liver ASCs as an example, can be adapted for all other tissues that can be cultured as organoids[9].

Development of the protocol

In this protocol, we use the organoid technology to select and expand single ASCs into clonal organoid cultures to facilitate WGS analysis. Single cells are isolated by sorting organoid cell suspensions using flow cytometry; of these single cells, only ASCs can give rise to long-term organoid cultures. These clonal ASC cultures are subsequently expanded until enough DNA can be isolated for WGS analysis (Figs. 1 and 2). As the whole genome is surveyed, both single-nucleotide variations and larger chromosomal aberrations can be studied.

All mutations present in the original ASC are inherited by each cell in the organoid culture and can therefore be accurately detected using bulk sequencing (Fig. 3a). We developed a data-analysis pipeline to generate high-quality catalogs of somatic variants with a confirmation rate of ~91%[1]. Clonality of the cultures can be verified in the WGS data by assessing the variant allele frequency (VAF) of the somatic mutations. Heterozygous mutations present in the cell of origin will be present in all cells of the organoid culture and, therefore, should have a VAF of ~0.5 in the WGS data. Subclonal mutations, which must have been introduced *in vitro* after the single-cell step, can be excluded on the basis of their lower VAF in the WGS data (Fig. 3). This quality-control step is critical, as the amount of somatic mutations in one ASC can only be accurately determined in purely clonal cultures using our approach.

This protocol can be used to catalog mutations that have accumulated in single ASCs either during life (Route A), or during culturing under normal or test conditions, such as in the presence of specific mutagens or drugs (Route B) (Fig. 1). Using this protocol, we have shown that ASCs from human liver, small intestine and colon acquire ~40 novel mutations per year, and that the mutation spectra differ between tissues[1] (Route A). A similar approach was used to obtain somatic variants

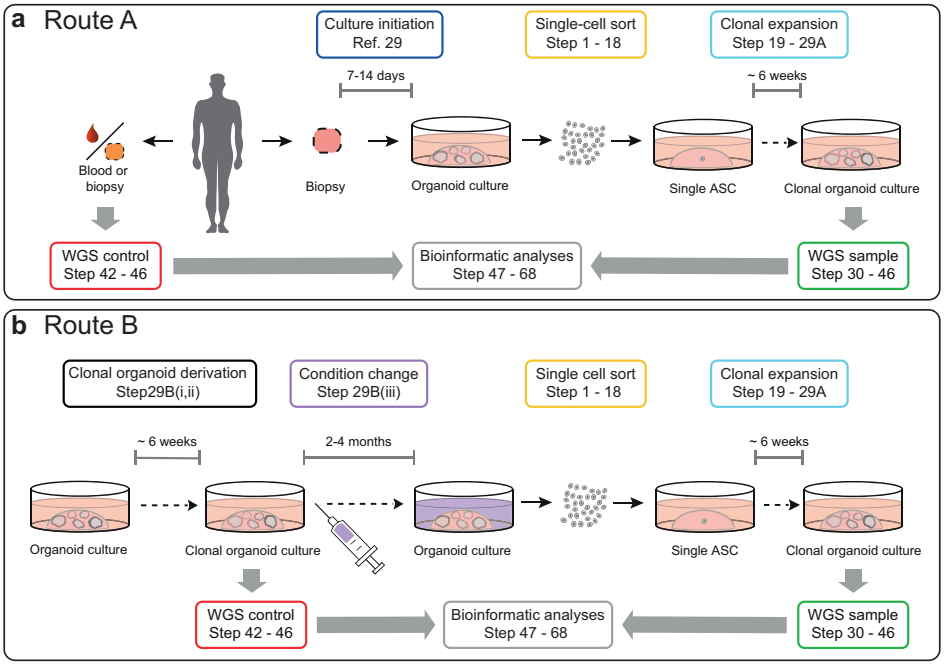


Figure 1. Schematic overview of the experimental setup. (A) Route A depicts the procedure for determining mutations that have accumulated in genomes of ASCs during life. Initially, polyclonal organoid cultures are derived from a biopsy, and expanded for 7–14 d under stem cell conditions. Single cells are subsequently isolated using flow cytometry, and plated to generate clonal organoid cultures. The clonal cultures are expanded for ~6 weeks to obtain sufficient DNA for WGS analysis. A control sample of a different tissue origin (blood or a polyclonal biopsy) is also subjected to WGS to exclude germline variants. (B) Route B depicts the procedure for determining mutations that have accumulated *in vitro* in the genomes of ASCs during organoid culturing. First, a clonal organoid culture is initiated, e.g., through Steps 1–28 of this protocol or through CRISPR/Cas9-mediated gene knockouts followed by clonal selection. This organoid culture is expanded under a culture condition of interest that allows the maintenance of viable ASCs for ~3–5 months. Subsequently, a clonal organoid culture is derived by performing a second single-ASC expansion step. Both the first and second clonal culture are subjected to WGS analysis to identify the variants that were specifically introduced during culturing.

of normal cells to reconstruct a developmental lineage tree in mice[5]. In addition, we applied this protocol to determine the genome stability of human liver stem cells after long-term organoid culture under normal conditions (Route B)[10].

For Route A, a bulk organoid culture is first derived from a tissue biopsy. This culture initiation step is essential to enrich for ASCs. After 1–2 weeks of culturing under conditions that promote stem cell growth, single cells are isolated by flow cytometry and clonally expanded (Fig. 1a). DNA isolated from the clonal organoid culture is subjected to WGS analysis to catalog somatic mutations. A blood sample or a polyclonal biopsy of a different tissue origin taken from the same individual is

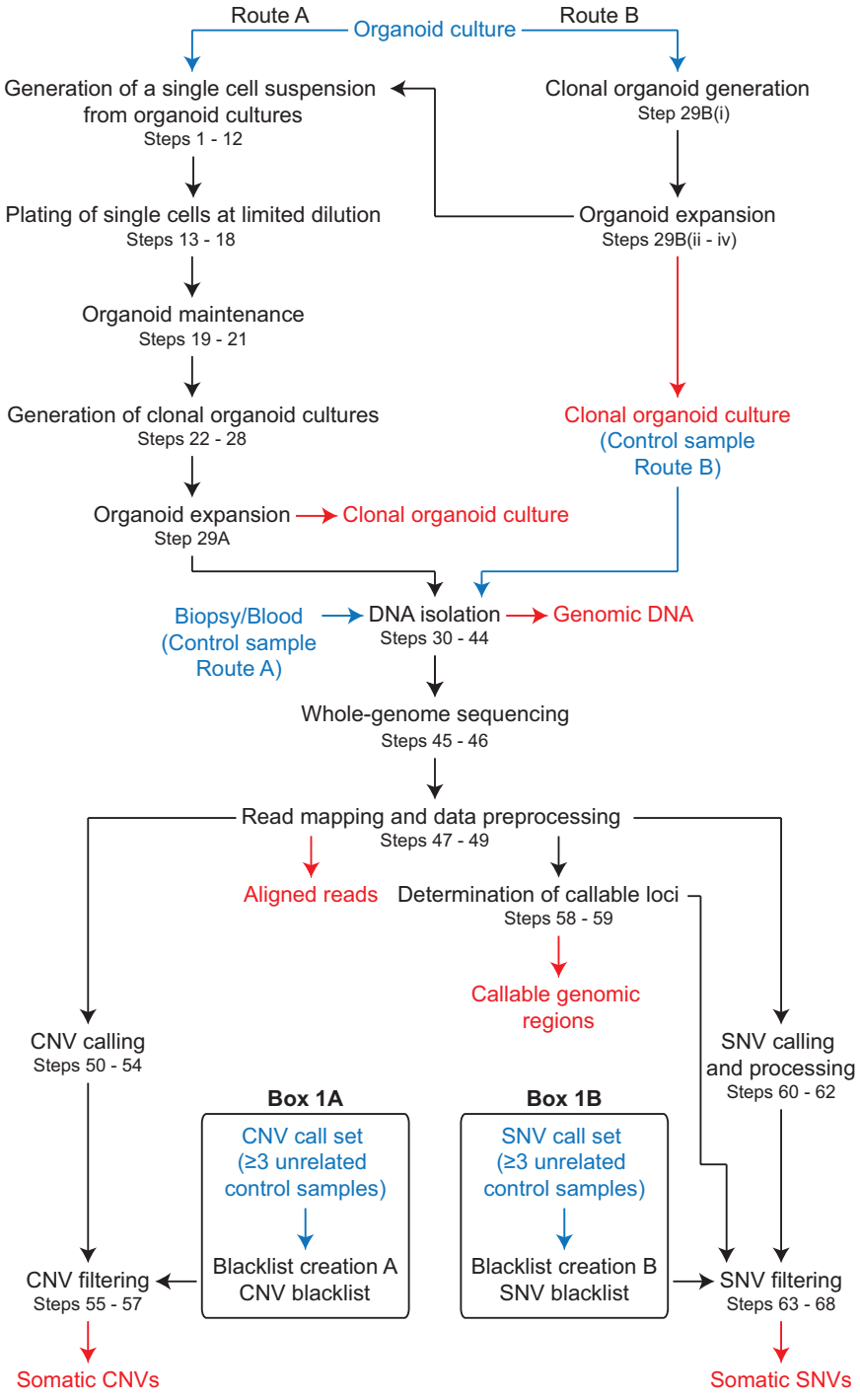


Figure 2. Schematic diagram of the protocol. The required input samples and input files are depicted in blue. The output of the protocol is depicted in red.

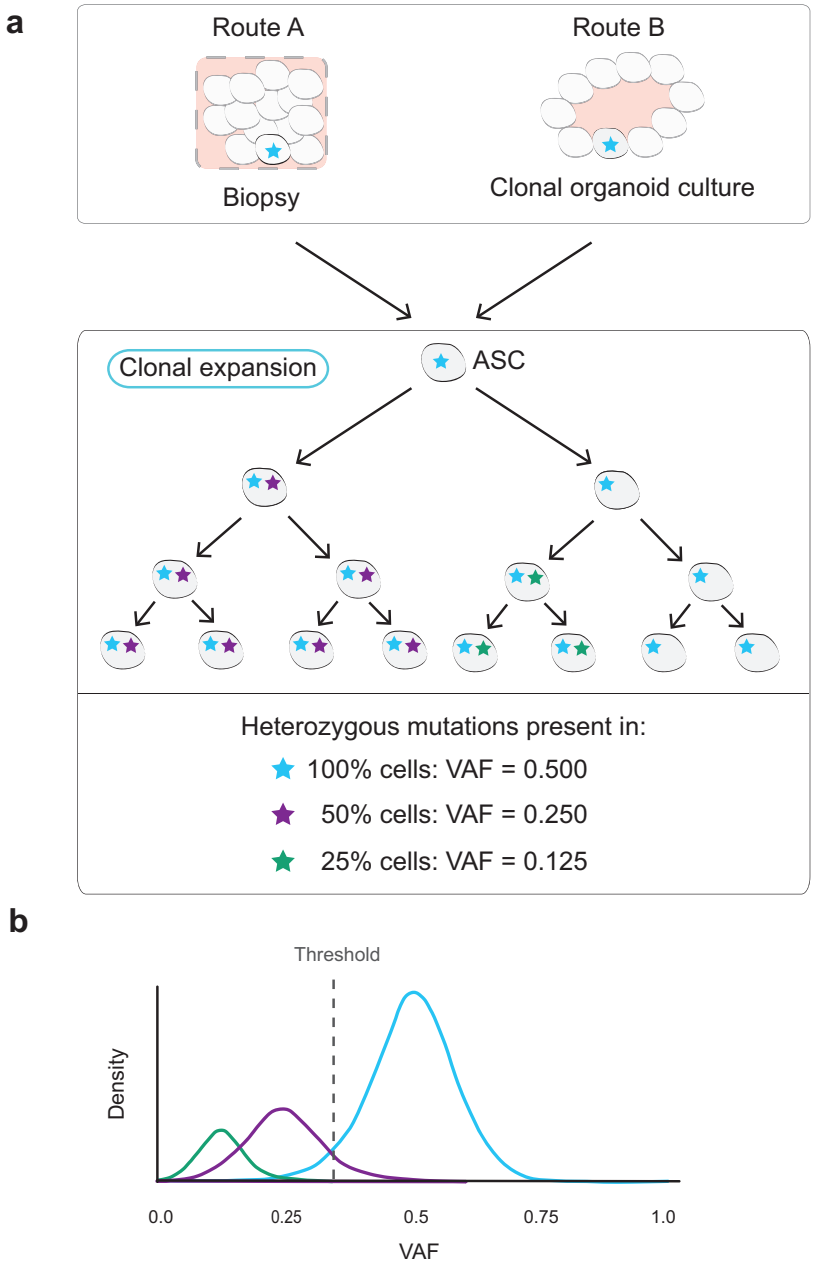


Figure 3. Theoretical variant allele frequency (VAF) of mutations in a clonal organoid culture. (A) A single ASC expands *in vitro* into an organoid culture. All heterozygous mutations that were present in the ASC of origin (represented by a blue star) will be inherited by all cells of the culture (100%), and therefore, they have a VAF of 0.5 in the WGS data. Mutations that are introduced during culturing, after the single-cell step (represented by purple and green stars), will have a VAF ≤0.25. (B) Theoretical VAF density plot of the mutations depicted in (A) A VAF-filtering threshold of 0.3 is applied to discard mutations that are introduced after the single-cell sort.

also subjected to WGS to identify and exclude germline variants. The majority of the variants that remain after filtering have accumulated during life, whereas only a small fraction of the variants were introduced during the culture initiation (see Experimental design)[1].

For Route B, a clonal organoid culture is expanded for 3–5 months, after which a second single-cell step is performed (Fig. 1b). This time period is required for normal ASCs under standard culture conditions to acquire enough mutations for downstream analysis with sufficient statistical power. Both clonal organoid cultures are subjected to WGS analysis to catalog mutations. To obtain the mutations that were specifically acquired between the two clonal steps, mutations identified in the second clonal culture are filtered for those already present in the first clonal culture.

Applications

The method described here can be used to detect the mutations that have accumulated during life in single ASCs of all tissues that can be cultured as organoids (Route A). Currently, organoids can be derived from many different human tissues, including small intestine, prostate, liver, breast and pancreas[9]. Culture protocols for new ASC types are constantly being developed[10,11,12], and, therefore, the number of tissues that can be studied using this technique is expected to increase in the future.

Somatic mutation catalogs of normal human ASCs will facilitate fundamental studies into *in vivo* mutagenesis, which is critical for understanding and possibly preventing the development of aging-associated diseases. Characterizing mutation accumulation in ASCs of various tissues can reveal tissue-specific mutation characteristics that might explain the variation in susceptibility to certain diseases among organs, such as cancer[1]. Moreover, organoids derived from donors of different ages allow the study of the relationship between age, mutation accumulation and disease incidence[1].

In addition to normal ASCs, this protocol can be used to study the mutation load of ASCs from abnormal tissue. For example, generating multiple clonal organoid cultures from a single tumor enables studies of tumor heterogeneity. Moreover, organoids derived from tissues of precancerous conditions, such as chronic inflammation[13], can elucidate aberrant mutational processes that might explain the increased cancer risk. This approach can also be used to assess, e.g., whether progeria patients, who suffer from pathologies that resemble early aging, have a different mutational load than healthy individuals[14]. Furthermore, it is possible to culture ASCs of other species, including mouse and rat[6,15], allowing the use of previously established knockout models or functional *in vivo* assays to study the effects of, e.g., caloric restriction on the mutation load in mice[16].

Route B of this protocol provides a platform for various *in vitro* studies and also for genetic safety testing for regenerative medicine applications. For instance, we have used this protocol to determine the effects of *in vitro* expansion under normal culture conditions on the genomic integrity of liver and intestinal ASCs[1,10]. Organoids hold great promise as a cellular source for regenerative medicine[9]. However, mutation accumulation in their genomes may induce oncogenic potential, which poses a major risk for the use of stem cells in regenerative medicine. Using the protocol described here, we have shown that ASCs maintain a high degree of genome stability in organoid cultures[10]. Nonetheless, as mutations do accumulate *in vitro*, future clinical use of organoids should be accompanied by routine genetic safety testing.

Recently, meta-analyses of human tumor WGS data have revealed 30 so-called 'mutational signatures'[17]. These patterns of mutations are thought to reflect single mutational mechanisms, but the etiology of the majority of these remains unknown. *In vitro* studies of clonal organoids can be performed to detect the mutations that arise under the influence of specific mutagenic substances in human ASCs. Likewise, it is possible to knock out a DNA repair pathway of interest (through CRISPR-Cas9 technology[18]) and assess the mutational consequences on the ASC genomic integrity. These functional assays will therefore help to link DNA damage and DNA repair mechanisms to mutational signatures. The diverse applications of this protocol will advance research into aging, cancer, DNA repair and stem cells in a human setting.

Limitations of the protocol

In addition to its unique strengths, this technique has some limitations. To obtain enough DNA for whole-genome sequencing, cells must self-renew and remain undifferentiated to give rise to an expandable clonal organoid culture. The stemness of the bulk organoid culture will therefore influence the efficacy of the protocol. We did not observe a correlation between the outgrowth potential of human liver ASCs and the age of the donors (Supplementary Fig. 1; Supplementary Table 1). In our experience, the plating efficiency of human liver cells is ~10–20%, and ~44% of the organoids successfully grow after clonal picking (Supplementary Table 1). Hence, a substantial amount of cells either did not tolerate the single-cell step or lacked stem cell properties required to generate an organoid culture. Despite these limitations, we have successfully established organoid cultures after flow cytometry from ~92% of the human liver organoid bulk cultures by picking multiple (~4) organoid cultures from each sample (Supplementary Table 1).

As mentioned previously, a reliable quantification of the genome-wide

somatic mutation load in a single ASC relies on the clonal origin of an organoid culture. However, the clonality of a culture can be determined only during bioinformatic analyses, after a substantial amount of effort has already been invested in generating and sequencing the organoid culture. We optimized this protocol to increase the chances of generating an organoid culture that originated from a single cell, e.g., by testing which dilutions are most likely to result in ~1–2 organoids per well, thereby reducing the chance of two organoids fusing together. As a result, ~93% of the sequenced human liver organoid cultures were found to be clonal after full analysis (Supplementary Table 1).

Comparison with other methods

Several other techniques can be used to determine genome-wide mutation accumulation in single ASCs during life. Mutations that were accumulated in the cell of origin and the most predominant subclones of a tumor can be measured by bulk sequencing, owing to the clonal origin of a tumor. Two mutational signatures have been successfully linked to aging before tumor formation using this approach[19]. However, cancers are characterized by high levels of genomic instability, often lack important DNA repair pathways and can be exposed to extreme environmental stress[20]. These factors complicate the use of tumor sequencing to characterize somatic mutations that accumulate before disease onset. Furthermore, ultra-deep sequencing of normal tissue that naturally displays a high degree of clonality, such as blood or small patches of skin, can be an effective strategy for detecting somatic mutations in noncancerous cells[21,22,23]. However, few normal tissues harbor large clonal cell populations, and extremely deep sequencing (~500×) is usually required to accurately catalog somatic variants[23].

The method described here has important advantages over other currently available sequencing approaches, as it is uniquely suited to accurately measuring mutations throughout the whole genome, specifically of single human ASCs, and does not require preamplification of the DNA before WGS analysis. Somatic mutations are typically distributed nonrandomly across the genome and, for instance, are depleted in genic regions[1,24]. Representative mutational patterns can therefore be constructed only through accurate genome-wide sequencing. Consequently, reporter assays, whole-exome sequencing and targeted deep sequencing are less preferable approaches for determining mutational patterns, as they are highly biased toward genic regions and may therefore yield incomplete and unrepresentative mutational patterns.

Notably, ASCs from organs such as liver, small intestine and colon acquire only ~36 novel SNVs genome wide per year[1]. Human ASCs also maintain a remarkably

stable genome *in vitro* and accumulate only 20–100 mutations per month[10]. Single-cell sequencing, on the basis of whole-genome amplification, albeit a very promising technique, introduces too many errors, such as allele and amplification biases, which will overshadow the low numbers of somatic mutations routinely observed in ASCs[25]. In the future, advances in single-cell sequencing might sufficiently decrease the error rate and thereby allow the generation of high-quality somatic mutation catalogs of single cells. If, in addition, it is possible to sort single ASCs (using, e.g., ASC-specific antibodies), a combination of these techniques may be used to determine genome-wide mutation patterns in human ASCs in the future. Still, our approach has the unique benefit that functional follow-up experiments can be performed on the exact same cellular material used for WGS analyses.

For Route B, induced pluripotent stem cell cultures derived from differentiated cells such as skin fibroblasts can be differentiated into ASCs and cultured as organoids[26,27]. As long as the cell culture is clonal, the bioinformatics protocol presented here can be applied to determine the mutation load in the cell of origin. However, if tissue biopsies are available, ASC cultures are preferable, as they eliminate the need for reprogramming and differentiation into ASCs, which may be associated with increased levels of genomic instability[28].

EXPERIMENTAL DESIGN

The method described here can be used to catalog mutation accumulation in any adult stem cell type that can be expanded into a clonal organoid culture. We have successfully applied this method previously to study mutational patterns in human liver, small intestine, colon and breast, and mouse liver and small intestine[1,5,10]. This protocol describes the entire procedure for human liver organoids. To apply this method to other ASC types, the culture steps can be easily adapted using culture conditions described for other tissues (e.g., pancreas[29], prostate[11], stomach[30], fallopian tube[31], small intestine and colon[32]). Bioinformatic analyses are identical for all ASC types.

Obtaining DNA from clonal organoid cultures

Before starting this protocol, one should derive an organoid culture from a human liver biopsy as described in Broutier *et al.*[29] for Route A, or thaw a previously established organoid culture for Route B. We have previously shown that ~5–25 *de novo* mutations occur *in vitro* in liver stem cells during 1 week of culture[1,10]. This indicates that the single ASC expansion step should be performed shortly after culture initiation, and the condition change should be applied quickly after the first clonal expansion step for Route A and Route B, respectively.

To create a clonal organoid culture, it is critical to generate a single-cell suspension from this bulk organoid culture (Steps 1–12). For example, plating 100 truly single cells is preferred over plating 10,000 doublets. If you retrieve <10,000 single cells, simply skip (some of) the highest cell dilutions when plating single cells at a limited dilution. Organoids usually start to appear 3–7 d after the single-cell sort. Until this point, we add ROCK inhibitor to the medium to inhibit cell death by anoikis[33]. After ~3 weeks, the organoids have grown large enough to be picked, and should be visible to the naked eye. To improve the success rate of obtaining clonal organoid cultures, we advise regular monitoring of the single cells with an inverted microscope after plating. This procedure will allow discrimination between purely clonal organoids and organoids that are potentially derived from multiple cells, for example, by fusion of two small organoids.

It is important to pick organoids from the lowest cell dilution possible, to reduce the chance of expanding a nonclonal organoid culture. Furthermore, it is important to pick multiple clones (at least four), as ~50% of the organoids stop growing before they expand sufficiently. We always pick organoids with tweezers and rupture the organoids by slicing them with needles (Steps 22–27). Alternatively, an organoid can be picked and ruptured by pipette. To do so, use a P20 pipette to aspirate the organoid in order to lift it directly from the BME droplet and transfer it to 200 µl of Adv+++ medium in a 24-well plate. Then rupture the organoid by holding the 24-well plate at a 45° angle and pipetting the organoid up and down with a P200 pipette in the Adv+++ medium at the bottom of the well until there are at least six pieces. Spin the pieces down and plate them in a new droplet of BME. The disadvantages of this alternative approach are that (i) the pieces of ruptured organoids tend to stick to the inside of a pipette tip and (ii) some of the pieces might be lost in the centrifugation step. Therefore, this alternative approach is suitable only when there is an abundance of organoids growing in low cell dilutions.

The clonal organoids are subsequently expanded in a 1:4 ratio every week to obtain at least eight wells in a 24-well plate after 3 weeks. Some organoid cultures grow slightly more slowly and should be expanded in a 1:2 or 1:3 ratio for 4 weeks. We always isolate DNA from 6 wells (in a single 24-well plate) of organoids; the remaining 2 wells are used for cryopreservation. For DNA isolation, we use the Qiagen Genomic tip protocol (Steps 37–43), but we have also successfully isolated DNA using the QIAsymphony (https://www.qiagen.com/at/resources/resourcedetail?id=8bc88dad-4140-467e-a1f0-e390fd193865&lang=en). DNA isolation should yield at least 1 µg of DNA. This is sufficient material for both WGS and future validation experiments. If the yield is <1 µg, organoid expansion and DNA isolation should be performed again.

Control samples

To distinguish variants that were specifically introduced during life (Route A) or during culture (Route B), it is essential to also subject a suitable control sample to WGS analysis. For Route A, we recommend using blood or a polyclonal biopsy from a different tissue of the same donor to exclude germline variants. If this material is not available, it is also possible to use a part of the (polyclonal) liver biopsy used for the generation of the organoid cultures as a control. However, it should be noted that somatic events that are shared between the ASC and the biopsy might be falsely filtered out. Although this caveat may potentially result in a slight increase of the false negatives, previous experiments indicate that -at least for human liver- a negligible number of somatic variants is missed[1]. For Route B, we routinely use a cell pellet of the initial culture as a control to exclude variants that were already present before the application of the test condition of interest. DNA isolation of the control sample should be performed using the same procedure as the clonal organoid cultures. In this way, DNA isolation biases, which have been reported to introduce copy-number differences, will be minimized[34].

Variant calling and filtering

WGS is performed on the Illumina HiSeq 2500 or HiSeq X systems according to standard protocols (https://www.illumina.com/). Copy-number variants (CNVs) are called using Control-FREEC[35] and filtered on the basis of size, evidence in the control sample and a CNV blacklist (Box 1, Step 1A). The copy-number profiles are checked for large ploidy deviations from 2, as these would have an impact on the default filtering of the single-nucleotide variants (SNVs). SNV discovery is performed according to basic protocols 1–2 of the Genome Analysis Toolkit (GATK) best practices workflow for germline single-nucleotide polymorphisms (SNPs) and indels in whole genomes[36]. Subsequently, a custom Single Nucleotide Variant Filtering pipeline (SNVFI available at https://github.com/UMCUGenetics/SNVFI) is applied to generate a catalog of high-quality somatic SNVs (Steps 63–68). The SNV call set is filtered on the basis of several quality parameters, and positions in the Single Nucleotide Polymorphism Database v137.b37 (ref. 37) are excluded. In addition, a protocol-specific SNV blacklist is created by collecting positions that were found to be variable in at least three unrelated individuals (Box 1, Step 1B). We recommend excluding these recurrent positions, as they represent either unknown SNPs or recurring sequencing and/or calling artifacts. Furthermore, somatic variants are selected by excluding events with evidence in the control sample (alternative depth >0). Finally, by default, variants are discarded with a VAF <0.3 to exclude the remaining sequencing noise and mutations

that were introduced *in vitro* after the single-cell sort (Fig. 3b), as this is optimal for ~30× WGS data. The VAF threshold can be adjusted for data with different coverage and/or ploidy states.

Callable region of the genome

To generate a high-quality set of somatic variants, it is critical to determine the 'callable regions' for each sample that was sequenced. To this end, the coverage and read quality at each genomic position is assessed to determine the genomic regions in which variants can be called. A variant call in the organoid culture can represent either a germline variant present in all cells of the individual or a somatic event in the ASC of origin. If there is also evidence for this variant in the control sample, this is most likely a germline variant. If, however, a mutation is called in the organoid culture, but this position is not callable in the control sample, we cannot determine whether it is a somatic or a germline event. Therefore, only variants at positions that are callable in both the organoid and the control sample can be evaluated as somatic variants.

In addition, the callable area is critical for downstream quantitative analyses. For example, callable regions might differ between samples, depending on the method applied for DNA extraction, the library preparation procedure, quality of the sequencing run and the overall sequencing depth. To compare the number of mutations between two samples, the possible differences in the callable areas must be accounted for. For this purpose, the number of mutations that were called in a sample can be extrapolated to the whole genome using the total length of callable regions specific to that sample. Moreover, determination of the callable area is required if you want to test for depletion or enrichment of mutations in certain genomic regions. Without adjusting the analysis for the callable area, you might, for example, observe a depletion of mutations in a certain genomic region (e.g., centromeres) that is solely a result of low coverage in that region and therefore does not represent an actual depletion of mutations.

Level of expertise needed

Execution of the entire procedure requires personnel experienced in several research areas. First, someone experienced in culturing organoids is recommended. Although, in principle, it would be possible to execute this protocol without experience in culturing organoids, organoids are more demanding than standard cell line cultures and require tailored care. Second, the cell sorting should be executed precisely to reduce the chance of generating nonclonal cultures. The method, therefore, requires someone highly experienced in operating a cell sorter. Personnel experienced

in whole-genome sequencing are also essential. These latter two tasks can be outsourced to a fluorescence-activated cell sorting (FACS) and sequencing facility or a service provider. After sequencing, a bioinformatician should be able to process the FASTQ files and perform standard quality control (QC), mapping and variant calling. Finally, basic command line skills are necessary to run the variant-filtering pipeline.

MATERIALS

Reagents

- Human liver organoid cultures, prepared as described in ref. 29
  Caution: The use of human material for these experiments should be approved by the relevant medical ethical committee, and informed consent should be provided by the donors and/or relatives.
  Caution: Human material can be pathogenic and should be handled with gloves.
- A control sample; either blood (Route A) or a polyclonal biopsy (Route A), or the initial organoid culture (Route B)
  Caution: The use of human material for these experiments should be approved by the relevant medical ethical committee, and informed consent should be provided by the donors and/or relatives.
  Caution: Human material can be pathogenic and should be handled with gloves.
  Critical: The control samples can be stored at −80 °C for at least 1 year, or at −20 °C for 1 month.

Reagents for organoid culture

- A 83-01 (Tocris Bioscience, cat. no. 2939)
- Advanced DMEM/F-12 (Thermo Fisher Scientific, cat. no. 12634028)
- Animal-free recombinant human EGF (hEGF; PeproTech, cat. no. AF-100-15)
- B-27 supplement minus vitamin A, 50× (Thermo Fisher Scientific, cat. no. 12587010)
- BME2 RGF Cultrex Pathclear (BME; Amsbio, cat. no. 3533-010-02)
- BSA (Sigma-Aldrich, cat. no. A9418)
- DMSO (Sigma-Aldrich, cat. no. D5879)
- Ethanol, 70% (vol/vol) (Klinipath, cat. no. 4070-9010)
  Caution: Ethanol is highly flammable.
- Forskolin (Tocris Bioscience, cat. no. 1099)
- Gastrin I, human (Sigma-Aldrich, cat. no. G9145)
- GlutaMAX Supplement (Thermo Fisher Scientific, cat. no. 35050038)
- HEPES, 1 M (Thermo Fisher Scientific, cat. no. 15630056)
- hES Cell Cloning & Recovery Supplement, 1,000× (Stemgent, cat. no. 01-0014-500)
- N-2 Supplement, 100× (Thermo Fisher Scientific, cat. no. 17502048)
- *N*-acetyl-L-cysteine (Sigma-Aldrich, cat. no. A9165)
- Nicotinamide (Sigma-Aldrich, cat. no. N0636)
- PBS (Thermo Fisher Scientific, cat. no. 14190-094)
- Penicillin–streptomycin, 10,000 U/ml (Thermo Fisher Scientific, cat. no. 15140122)
- Primocin (Invivogen, cat. no. ant-pm-1)

- Recombinant Human FGF-10 (PeproTech, cat. no. 100-26)
- Recombinant Human HGF (insect derived; PeproTech, cat. no. 100-39)
- Recombinant Human Noggin (PeproTech, cat. no. 120-10C)
- RspoI-conditioned medium, produced as described in Box 1 in ref. 29
  Critical: Recombinant Human R-spondin-1 (PeproTech, cat. no. 120-38) is also commercially available, but we have no experience with this in our laboratory.
- TrypLE Express Enzyme, 1×, phenol red (Thermo Fisher Scientific, cat. no. 12605010)
- Wnt-conditioned medium, produced as described in Box 1 in ref. 29
  Critical: Recombinant Human Wnt-3a protein with carrier (R&D Systems, cat. no. 5036-WN-010) is also commercially available, but we have no experience with this in our laboratory.
- Y-27632 dihydrochloride, ROCK inhibitor (Abmole, cat. no. M1817)

Reagents for DNA isolation and WGS

- Ethanol absolute, ~100% (Merck Millipore, cat. no. 1009832500)
  Caution: Absolute ethanol is highly flammable.
- Genomic DNA Buffer Set (Qiagen, cat. no. 19060)
- Isopropanol (Merck Millipore, cat. no. 1096342500)
  Caution: Isopropanol is highly flammable.
- Proteinase K (Qiagen, cat. no. 19133)
- Reagents as described in the cBot System Guide (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/cbot/cbot-system-guide-15006165-02.pdf)
- Reagents as described in the HiSeq X System Guide (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/hiseqx/hiseq-x-system-guide-15050091-04.pdf)
- Reagents as described in the TruSeq Nano DNA Library Prep Reference Guide (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_tr useq/truseqnanodna/truseq-nano-dna-library-prep-guide-15041110-d.pdf)
- RNase A (Qiagen, cat. no. 19101)
- Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, cat. no. Q32850)
- TE buffer, pH 8.0, low EDTA (Thermo Fisher Scientific, cat. no. 12090015)
- TruSeq Nano DNA Library Prep Kit (Illumina, cat. no. FC-121-4001)

Equipment

- Accu-jet pro Pipette Controller (BrandTech Scientific, cat. no. 26330)
- Cell sorter (Becton Dickinson, model no. BD FACSAria IIU Three-Laser System)
- DNA LoBind microcentrifuge tubes, 1.5 ml, 2.0 ml (Eppendorf, cat. nos. 022431021, 022431048)
- Falcon tubes, 15 ml, 50 ml (Greiner Bio-One, cat. nos. 188271, 227261)
- Filter tips, P10, P20, P200 and P1000 (Greiner Bio-One, cat. nos. 771288, 774288, 739288 and 740288)
- Glass Pasteur pipette, 230-mm, unplugged (VWR, cat. no. 612-2300)
- Microcentrifuge (Sigma-Aldrich, model no. 1-15P)

- Pipetman Classic, P2, P20, P200 and P1000 (Gilson, cat. nos. F144801, F123600, F123601, and F123602)
- Serological pipettes, 2 ml, 5 ml, 10 ml and 25 ml (Sarstedt, cat. nos. 86.1252.001, 86.1253.001, 86.1254.001 and 86.1685.001)
- Sorvall Legend XT centrifuge (Thermo Fisher Scientific, cat. no. 75004505)
- Sorvall Legend XTR centrifuge (Thermo Fisher Scientific, cat. no. 75004520)
- Vortex-Genie 2 (Scientific Industries, cat. no. SI-0256)
- Water bath, 37 °C (Grant, model no. SAP12)
- Water bath, 50 °C (Julabo, model no. TW20)

Equipment for organoid culture

- 4-Well Nunc cell culture-treated multidish, (Thermo Fisher Scientific, cat. no. 176740)
- 24-Well cell culture plate (Greiner Bio-One, cat. no. 662160)
- 24-Well cell culture plate for suspension culture (Greiner Bio-One, cat. no. 662102)
- BioSafety cabinet (Telstar, cat. no. EN 12469)
- CO_2 cell culture incubator, 37 °C and 5% CO_2 (Panasonic, cat. no. MCO-18AIC)
- Falcon round-bottom tube with cell strainer snap cap, 5 ml (Corning, cat. no. 352235)
- Forceps (VWR, cat. no. 232-2122)
- Inverted microscope (Zeiss, model no. Axiovert 25)
- Microlance hypodermic needle, 25 gauge, orange, 16 mm (Becton Dickinson, cat. no. 300600)
- Safety laboratory gas burner, Fuego basic (W$_L$D-TEC, cat. no. 8.201.000)
- Stereomicroscope (Nikon, model no. SMZ800N)
- Surgical scalpel blade, no. 21 (Swann-Morton, cat. no. 0207)

Equipment for DNA isolation and WGS

- cBOT cluster generator for Illumina sequencing (Illumina, cBOT system)
- DNA sequencer (Illumina, model no. HiSeq X Ten)
- Equipment as described in the cBot System Guide (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/cbot/cbot-system-guide-15006165-02.pdf)
- Equipment as described in the HiSeq X System Guide (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/hiseqx/hiseq-x-system-guide-15050091-04.pdf)
- Equipment as described in the TruSeq Nano DNA Library Prep Reference Guide (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqnanodna/truseq-nano-dna-library-prep-guide-15041110-d.pdf)
- Genomic-tip 20/G (Qiagen, cat. no. 10223)
- HiSeq X Ten Reagent Kit (Illumina, cat. no. FC-501-2501)
- Thermomixer, 55 °C (Eppendorf, cat. no. 5384000012)
- Qubit 2.0 fluorometer (Thermo Fisher Scientific, cat. no. Q32866)
- Qubit assay tubes (Thermo Fisher Scientific, cat. no. Q32856)

Equipment for data processing and analysis

- A computer with a Unix or Unix-like system, such as GNU/Linux or MacOS X, and as much memory and computing power as possible. A minimum of 8 GB of RAM is required.
- Our timing estimates are based on a computer that can run 10 threads in parallel.

Software

- SNVFI (https://github.com/UMCUGenetics/SNVFI)
- Java v1.7 (https://www.java.com/en/download/)
- BWA v0.7.5a (ref. 38; http://bio-bwa.sourceforge.net/)
- GATK v3.4-46 (ref. 39; https://software.broadinstitute.org/gatk/)
- Picard v1.141 (https://broadinstitute.github.io/picard/)
- Sambamba v0.5.8 (ref. 40; http://lomereiter.github.io/sambamba/)
- R v3.2.2 (ref. 41; https://www.r-project.org/)
- Bedtools v2.25.0 (ref. 42; http://bedtools.readthedocs.io/en/latest/)
- SAMtools v1.3 (ref. 43; http://samtools.sourceforge.net/)
- Control-FREEC v2.7 (ref. 35; http://boevalab.com/FREEC/)

Reagent Setup

- Adv+++ medium: Supplement advanced DMEM/F-12 with 1% (vol/vol) GlutaMAX, 10 mM HEPES and 1% (vol/vol) penicillin–streptomycin. Store the Adv+++ medium at 4 °C for up to 1 month.
- PBS with 0.1% (wt/vol) BSA: Supplement 1× PBS with 0.1% (wt/vol) BSA and filter sterilize the solution. Store the solution at 4 °C for a maximum of 1 month or at −80 °C for up to 1 year.
- A83-01: Dissolve 10 mg of A83-01 in 4,750 µl of DMSO, and store 50-µl aliquots at −20 °C. Use this stock solution at a ratio of 1:1,000 for up to 2 months, and avoid freeze–thaw cycles.
- hEGF: Dissolve 1 mg of animal-free recombinant hEGF in 2 ml of PBS with 0.1% (wt/vol) BSA, and store 5-µl aliquots at −20 °C. Use this stock solution at a ratio of 1:10,000 for up to 2 months, and avoid freeze–thaw cycles.
- Forskolin: Dissolve 10 mg of Forskolin in 2,440 µl of DMSO, and store 50-µl aliquots at −20 °C. Use this stock solution at a ratio of 1:1,000 for up to 2 months, and avoid freeze–thaw cycles.
- Gastrin: Dissolve 1 mg of gastrin I (human) in 4,800 µl of PBS, and store 5-µl aliquots at −20 °C. Use this stock solution at a ratio of 1:10,000 for up to 2 months, and avoid freeze–thaw cycles.
- *N*-acetyl-L-cysteine: Dissolve 1 g of *N*-acetyl-L-cysteine in 12.27 ml of H_2O, filter sterilize the solution and store 150-µl aliquots at −20 °C. Use this stock solution at a ratio of 1:400 for up to 2 months, and avoid freeze–thaw cycles.
- Nicotinamide: Dissolve 1.2 g of nicotinamide in 10 ml of PBS, filter sterilize the solution and store 0.5-ml aliquots at −20 °C. Use this stock solution at a ratio of 1:100 for up to 2 months, and avoid freeze–thaw cycles.
- FGF-10: Dissolve 100 µg of recombinant human FGF-10 in 1 ml of PBS with 0.1% (wt/vol) BSA, and store 50-µl aliquots at −20 °C. Use this stock solution at a ratio of 1:1,000 for up to 2 months, and avoid freeze–thaw cycles.
- HGF: Dissolve 100 µg of recombinant human HGF (insect derived) in 1 ml of PBS with 0.1% (wt/vol) BSA, and store 15-µl aliquots at −20 °C. Use this stock solution at a ratio of 1:4,000 for

up to 2 months, and avoid freeze–thaw cycles.

- Noggin: Dissolve 100 µg of recombinant human Noggin in 1 ml of Adv+++, and store 50-µl aliquots at −20 °C. Use this stock solution at a ratio of 1:1,000 for up to 2 months, and avoid freeze–thaw cycles.
- ROCK inhibitor: Dissolve 50 mg of Y-27632 dihydrochloride in 1.5 ml of H_2O, filter sterilize the solution and store 50-µl aliquots at −20 °C. Use this stock solution at a ratio of 1:10,000 for up to 2 months, and avoid freeze–thaw cycles.
- Human liver expansion medium: Human liver expansion medium is Adv+++ medium with 10% (vol/vol) RspoI-conditioned medium, 1× B27 minus vitamin A, 1× N2, 10 mM nicotinamide, 1.25 mM N-acetyl-L-cysteine, 1× Primocin, 5 µM A83-01, 10 µM forskolin, 100 ng/ml FGF-10, 25 ng/ml HGF, 10 nM gastrin I and 50 ng/µl hEGF.
  Critical: Store the human liver expansion medium at 4 °C for a maximum of 1 week.
- Human liver single-cell medium: Human liver single-cell medium is human liver expansion medium with 10 µM ROCK inhibitor.
  Critical: Store the human liver single-cell medium at 4 °C for a maximum of 1 week.
- Human liver establishment medium: Human liver establishment medium is human liver single-cell medium with 30% (vol/vol) Wnt-conditioned medium, 100 ng/ml Noggin and 1× hES Cell Cloning & Recovery Supplement.
  Critical: Store the human liver establishment medium at 4 °C for a maximum of 1 week.

PROCEDURE

Generation of a single-cell suspension from organoid cultures

Timing: 1.5 h

Critical: Making a single-cell solution is an essential step in the procedure.

1 | Remove the human liver expansion medium from at least three wells (24-well plate) or six wells (48-well plate) of human liver organoid culture without disturbing the BME droplets.

2 | Prewash a P1000 filter tip in cold TrypLE to make sure that the organoids are less likely to stick to the inside of the pipette tip.

3 | Add cold TrypLE (250 µl per well of a 48-well plate or 500 µl per well of a 24-well plate) to the wells. Using a P1000 pipette and the prewashed P1000 filter tip from Step 2, pipette the TrypLE up and down until the BME is fully resuspended in TrypLE. Transfer the organoid solution to a 15-ml tube. Combine three wells (24-well plate) or six wells (48-well plate) in one 15-ml tube at most. 1.5 ml of TrypLE fits into a 230-mm glass Pasteur pipette, and therefore, all organoids can be ruptured properly in Steps 5–7.

4 | Reduce the size of the tip of a glass Pasteur pipette by ∼50% by holding it in a flame. Cool the Pasteur pipette by pipetting cold Adv+++ medium after narrowing the tip; this also primes the pipette, ensuring that the organoids are less likely to stick to the inside of the Pasteur pipette.

5 | Rupture the organoids by pipetting the organoid solution up and down five times using a glass Pasteur pipette with a narrowed tip.

6 | Incubate the solution for 5 min at 37 °C in a water bath. Then pipette the solution up and down five times, again using a glass Pasteur pipette with a narrowed tip.

7 | Repeat Step 6 two times.

8 | Check the cell suspension using an inverted microscope. The cell suspension should predominantly consist of single cells.

Troubleshooting

9 | Add 10 ml of cold Adv+++ medium to each 15-ml tube, and centrifuge the medium at 450*g* for 5 min at room temperature (RT; ~20 °C).

10 | Using a P1000 pipette and a P200 pipette, carefully remove the supernatant without disturbing the cell pellet (the cell pellet is often not visible).

Critical step: Take great care not to aspirate the cell pellet.

11 | Resuspend the cells in 300 µl of human liver establishment medium, and transfer the cell suspension through the 35-µM filter top of a Falcon round-bottom tube with a cell strainer snap cap into the Falcon round-bottom tube. Keep the cells on ice, and proceed with the cell sort.

Critical step: The cells should be sorted as soon as possible (preferably within 30 min) to minimize cellular stress.

12 | By using a cell sorter, sort 10,000 single cells into a 1.5-ml Eppendorf tube with 500 µl of cold liver establishment medium. Sort the cells with an 85-micron nozzle and 45-psi sample pressure, and a 488-nm (50 mW) laser. The drop delay is determined by using BD Accudrop Beads and the auto drop delay feature in the FACSDiva software. Single cells should be selected in the FACSDiva software on the basis of forward- and side-scatter characteristics. A FACS plot depicting the gating strategy is shown in Figure 4.

Critical step: Isolating 100 single stem cells is preferred over isolating 10,000 likely single cells, which may still contain doublets.

Troubleshooting

Plating of single cells at a limited dilution

Timing: 1 h

Critical: Place a normal 24-well cell culture plate at 37 °C in the cell culture incubator for at least 24 h (up to 1 month) before plating single cells. We advise against using a 24-well cell culture plate for suspension culture in Steps 13–18 of this protocol, as the BME droplets are more likely to dissociate from the surface of this type of cell culture plate (Steps 19–21).

Critical: BME becomes solid at room temperature, so thaw BME on ice and keep the BME on ice while using it.

13 | Centrifuge the 1.5-ml Eppendorf tube with the single-cell suspension from Step 12 at 450*g* for 5 min at RT.

14 | Using a P1000 pipette and a P200 pipette, carefully remove the supernatant without disturbing the cell pellet.

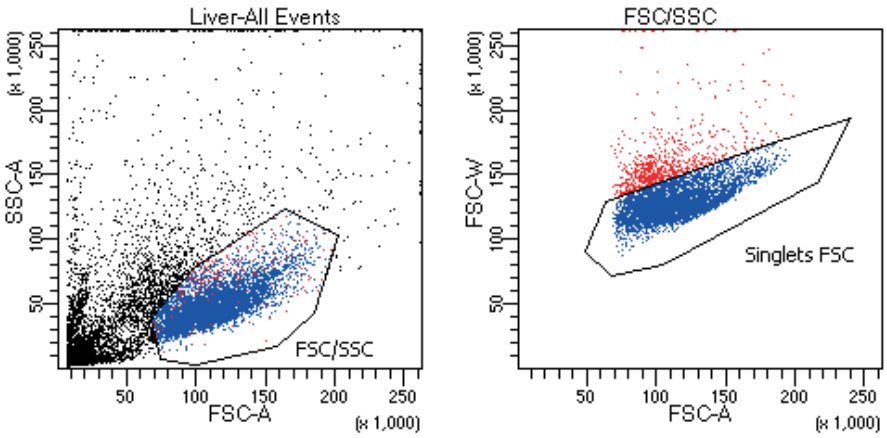


Figure 4. FACS plots depicting gating strategies for sorting single cells from a human liver organoid culture in Step 12. These plots are generated by the FACSDiva software. Each dot represents a cell or clump of cells. Single cells are selected on the basis of the side-scatter area (SSC-A) and the forward-scatter area (FSC-A) (left panel). The forward-scatter width (FSC-W) and FSC-A are used to exclude doublets (right panel). Red dots represent likely doublets. Blue dots represent single cells.

Critical step: Take great care not to aspirate the cell pellet.

15 | Resuspend the cell pellet in 104 µl of BME. If the FACS yielded <10,000 cells, add a lower volume of BME to reach a cell concentration of 96 cells/µl (the 'bulk cell suspension').

Critical step: Avoid creating air bubbles in the BME while pipetting, as these bubbles can burst at 37 °C in the cell culture incubator.

16 | Plate the cells at increasing dilutions, as depicted in Figure 5. Briefly, transfer 1,440 cells (15 µl) from the bulk cell suspension to 273 µl of BME, mix thoroughly using a pipette and plate 6 × 20-µl droplets (5 cells/µl) in 3 wells of a 24-well culture plate. Add 112 µl of BME to the remaining dilution, mix thoroughly using a pipette and plate 6 × 20-µl droplets (3 cells/µl) in 3 wells of a 24-well culture plate. Add 320 µl of BME to the remaining dilution, mix thoroughly using a pipette and plate 12 × 20-µl droplets (1 cell/µl) in 6 wells of a 24-well culture plate. Add 240 µl of BME to the remaining dilution, mix thoroughly using a pipette and plate the BME in 20-µl droplets (0.5 cells/µl) in 10 wells of a 24-well culture plate. Finally, plate the remainder of the 96 cells/µl concentration in droplets of 20 µl in 2 wells of a 24-well culture plate.

Critical step: To minimize adherence of human liver stem cells to the plastic of the cell culture plate, ensure that the BME droplets do not touch the edges of the well. Pipette 2 × 20-µl droplets of BME per well, as indicated in the schematic overview (Fig. 5b).

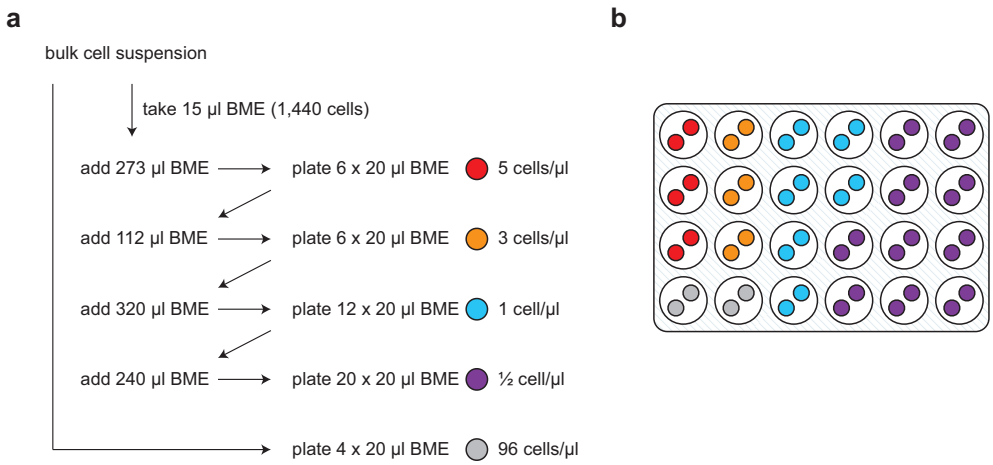


Figure 5. Plating cells at a limited dilution. (A) Schematic diagram of the pipetting steps in Step 16. (B) Schematic overview of the cell dilutions in a 24-well culture plate after performing Step 16. Two 20-µl drops of BME (depicted as small colored circles) should be pipetted into each well (depicted as large white circles). After plating according to the scheme in (A), you will have 2 wells with 96 cells/µl (gray), 3 wells with 5 cells/µl (red), 3 wells with 3 cells/µl (orange), 6 wells with 1 cell/µl (blue) and 10 wells with 0.5 cell/µl (purple).

17 | Let the BME set by incubating the 24-well plate for 30 min at 37 °C in the cell culture incubator.

18 | Cover the BME droplets with 500 µl of human liver establishment medium per well, and put the plate in the cell incubator at 37 °C.

Organoid maintenance

Timing: 3 weeks, 4 h hands-on

Critical: During the organoid maintenance phase of this protocol (Steps 19–21), change the medium carefully. BME droplets can dissociate from the cell culture plate, especially after several weeks. To improve the success rate for obtaining clonal organoid cultures, we advise regular monitoring of the single cells after plating them with an inverted microscope. We recommend localizing at least 40 single cells within a few hours after the single-cell sort, and monitoring whether these cells form clonal organoids at least twice a week, until they have grown enough to be picked.

19 | Change the medium after 3–4 d to human liver single-cell medium. Remove the medium carefully without disturbing the BME, cover the BME droplets with 500 µl of human liver single-cell medium per well and put the plate in the cell incubator at 37 °C.

20 | Change the medium after 3–4 d (~1 week after single-cell sort) to human liver expansion medium. Remove the medium carefully without disturbing the BME, cover the BME droplets with 500 µl of human liver expansion medium per well and put the

plate in the cell incubator at 37 °C. Usually, organoids begin to appear 3–7 d after the single-cell sort.

Troubleshooting

21 | Change the medium every 3–4 d until the organoids are easy to see with the naked eye. Remove the medium carefully without disturbing the BME, cover the BME droplets with 500 µl of human liver expansion medium per well and put the plate in the cell incubator at 37 °C. After ~2 weeks (~3 weeks after the FACS), organoids are usually ready for the next steps. If the organoid culture grows slowly, continue to change the medium every 3–4 d for a maximum of 3 additional weeks (~1.5 months after the FACS), and continue with the next steps.

Troubleshooting

Generation of clonal organoid cultures

Timing: 1 h

22 | Place a 4-well cell culture plate on ice, and allow it to cool for at least 5 min.

23 | Pipette one 20-µl droplet of cold BME into the center of a well, and leave the plate on ice.

24 | Pick an organoid from the lowest dilution possible. If possible, pick an organoid from a BME droplet that contains only one organoid to reduce the chance of accidentally picking up multiple organoids at the same time. Use clean (cleaned with 70% (vol/vol) ethanol) forceps and a stereomicroscope. Gently go around the organoid to detach it from the BME, and simply pick it up. Transfer the organoid to the droplet of BME in a 4-well plate. Make sure that the organoid is properly transferred to the droplet (organoids tend to stick to the forceps).

Troubleshooting

25 | Put the 4-well plate on a stereomicroscope. Use two small needles to cut the organoid into smaller pieces, as you would cut a steak. Continue until you have 5 or 6 pieces. Then quickly put the 4-well plate back on ice.

Caution: The needles are sharp. Be careful while cutting human organoids, to prevent the potential transfer of infectious diseases from the donor to yourself.

Critical step: Work quickly, as BME becomes solid at room temperature. After slicing the organoid, double-check that the organoid pieces are not on the bottom of the well, as they will adhere and potentially differentiate. Move the organoid pieces up in the BME, using the needles to make sure that they do not adhere to the bottom of the plate.

Troubleshooting

26 | Repeat Steps 23–25 at least three times, until four organoids are passaged into four separate wells.

Critical step: Clean the forceps thoroughly with 70% (vol/vol) ethanol after picking an organoid, and use clean needles for each organoid, to prevent the mixing of clonal cultures.

27 | Let the BME set by incubating the 4-well plate upside down for 30 min at 37 °C

in the cell culture incubator.

Critical step: Put the plate upside down in the incubator, to prevent the large organoid pieces from adhering to the bottom of the plate.

28 | Cover the BME droplets with 500 µl of human liver medium per well, and put the plate in the cell incubator at 37 °C.

Organoid expansion

Timing: 3 weeks–5.5 months, 3–22 h hands-on

Critical: Organoid cultures can be expanded every ~7 d on 24-well plates for suspension culture. Refresh the medium every 3–4 d during this time. Do not passage the organoids too sparsely; we recommend passaging the human liver organoids at a ratio of 1:4. However, some clones grow slightly more slowly and should be passaged in a 1:2 or 1:3 ratio.

29 | Use option A to catalog mutations that have accumulated during life (Route A), and option B to catalog mutations that have accumulated during culture (Route B).

a. Route A

Timing: 3 weeks, 3 h hands-on

i. Expand the organoid culture as described in Steps 8–15 of ref. 29, until you can plate the organoids in at least 8 wells (24-well plate). These steps include disrupting the BME and resuspending the organoids in Adv+++, spinning the solution down at 100–200*g*, rupturing the organoids by using a Pasteur pipette with a narrowed tip, adding cold Adv+++ and spinning the solution down at 200–250*g*, aspirating the supernatant, resuspending the organoid solution in an appropriate volume of BME, plating the BME and overlaying the organoids in BME with human liver expansion medium.

ii. Use at least 6 wells for DNA isolation, and cryopreserve 2 wells (24-well plate). The protocol for cryopreservation of human liver organoids is described in Box 3 of ref. 29.

b. Route B

Timing: 3.5-5.5 months of additional culture time, 22 h hands-on

i. Generate a clonal organoid culture, e.g., through Steps 1–28 of this protocol or using CRISPR/Cas9 (ref. 18).

ii. Continue to culture the organoids as described in Step 29A(i), until you have at least 9 wells of organoid culture (24-well plate). Use at least 6 wells for DNA isolation, and cryopreserve 2 wells (24-well plate).

iii. Culture the remaining well(s) of the organoid culture for an additional 2–4 months. During this time period, it is possible to change the culture conditions by adding mutagens or drugs to the medium and testing their effect on the genomic stability of the ASCs in culture.

iv. After these additional months of culture, return to Step 1 and repeat Steps 1–29A(i–ii).

DNA isolation

Timing: 2 d, 4 h hands-on

30 | Remove the human liver medium from at least six wells (24-well plate) without disturbing the BME droplets.

31 | Prewash a P1000 filter tip in cold Adv+++ to make sure that the organoids are less likely to stick to the inside of the pipette tip.

32 | Add 250 µl of cold Adv+++ to each well. Using a P1000 pipette and the prewashed P1000 filter tip from Step 31, pipette the Adv+++ up and down until the BME is fully resuspended in Adv+++ medium. Transfer the organoid solution to a 15-ml tube. Combine 6 wells at most (24-well plate) per 15-ml tube. 1.5-ml of TrypLE fits into a 230-mm glass Pasteur pipette, and, therefore, all organoids can be ruptured properly at Step 34.

33 | Prepare a glass Pasteur pipette as described in Step 4.

34 | Rupture the organoids into small pieces by pipetting the organoid solution up and down approximately ten times using a glass Pasteur pipette with a narrowed tip.

35 | Add 10 ml of cold Adv+++ medium to the organoid solution, and centrifuge the solution at 250*g* for 5 min at RT.

36 | Remove the supernatant without disturbing the cell pellet.

Troubleshooting

Pause point: The cell pellet for DNA isolation can be stored at −20 °C for a month.

37 | To further prepare the organoid samples for DNA isolation, follow Steps 1 and 2 of the Genomic-tip 20/G protocol section titled 'Protocol: Preparation of Tissue Samples', in the *Qiagen Genomic DNA Handbook* (available at https://www.qiagen.com/gb/resources/resourcedetail?id=d2b85b26-16dd-4259-a3a7-a08cbd2a08a3&lang=en).

38 | Add 2 ml of Buffer G2 (with RNase A) and 0.1 ml of proteinase K to each organoid pellet.

39 | Incubate the sample(s) at 55 °C for 2 h. Vortex the samples (Vortex-Genie 2: speed 5) for ~5 seconds every 15 min during incubation.

40 | Centrifuge the sample(s) at 5,000*g* for 10 min at 4 °C, and transfer the supernatant to a new 15-ml tube. Discard the pellet, and use the supernatant in the next step of the protocol.

41 | To isolate DNA from the organoid samples, follow Steps 1–4, 5B and 6B of the Genomic tip 20/G protocol section titled 'Protocol: Isolation of Genomic DNA from Blood, Cultured Cells, Tissue, Yeast, or Bacteria using Genomic-tips', in the *Qiagen Genomic DNA Handbook*. Resuspend the cell pellet in 100 µl of TE buffer (with low

EDTA) at Step 6B.

Pause point: When resuspended in TE buffer (with low EDTA), DNA can be stored at −20 °C for decades.

42 | Prepare control samples for DNA extraction according to the Genomic tip 20/G protocol, as described in the *Qiagen Genomic DNA Handbook*, using option A to isolate DNA from tissue biopsy, option B to isolate DNA from blood or option C to isolate DNA from a control organoid sample:

A. Preparation of tissue biopsy for DNA isolation

i. Follow Steps 1–4 of the Genomic tip 20/G protocol section titled 'Protocol: Preparation of Tissue Samples'.

ii. Incubate the samples overnight at 55 °C.

B. Preparation of a blood sample for DNA isolation

i. Isolate DNA from blood according to the Genomic tip 20/G protocol section titled 'Protocol: Preparation of Blood Samples'.

C. Preparation of a clonal control organoid sample for DNA isolation

i. Follow Steps 37–39 of this protocol.

43 | Isolate DNA from all control samples as described in Steps 40 and 41.

Pause point: When resuspended in TE buffer (with low EDTA), DNA can be stored at −20 °C for decades.

44 | Measure the DNA concentration by using a Qubit fluorometer. DNA isolation should yield at least 1 µg per sample.

Troubleshooting

Whole-genome sequencing

Timing: 5 d, 2 d hands-on

45 | Prepare DNA libraries from 200 ng of genomic DNA using the TruSeq Nano DNA Library Kit according to the TruSeq Nano DNA Library Prep Reference Guide (available at https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqnanodna/truseq-nano-dna-library-prep-guide-15041110-d.pdf). Mix 100.5 µl of sample purification beads (provided with the TruSeq Nano DNA Library Kit) with 83.5 µl of PCR-grade water in Step 2 of the 'Remove large DNA fragments' section in the 'Repair Ends and Select Library Size' section to enrich for a 450-bp insert size.

Troubleshooting

46 | Pool a maximum of 8 libraries, and apply them to the flow cell using the cBOT system according to the cBot System Guide (available at https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/cbot/cbot-system-guide-15006165-02.pdf). Sequence the libraries on an Illumina HiSeq X Ten DNA sequencer for 2×100-bp paired-end sequencing to ~30× base coverage with the HiSeq X Ten Reagent Kit according to the HiSeq

X System Guide (available at https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/hiseqx/hiseq-x-system-guide-15050091-04.pdf).

Troubleshooting

Read mapping and data preprocessing

Timing: 48 h, 4 h hands-on

47 | Follow the GATK best practices[36] basic protocol 1, Steps 1–4 to map the raw reads (FASTQ files) to the reference genome using BWA[38] to create an *aligned_reads.sam* file.

48 | Using the following commands, mark duplicate reads with Sambamba[40] rather than Picard in order to speed up the processing time. This creates a BAM file, *dedup_reads.bam*, with all the original reads with duplicate reads marked.

```
$ sambamba view -S -t 10 -f bam aligned_reads.sam > aligned_reads.bam
$ sambamba sort -t 10 -m 32GB aligned_reads.bam -o sorted_reads.bam
$ sambamba markdup -t 10 sorted_reads.bam dedup_reads.bam
```

49 | Follow the GATK best practices[36] basic protocol 1, Steps 6–10 to perform indel realignment and base quality score recalibration. This creates a BAM file, *recal_reads.bam*, with all the original reads with better alignments around indels and accurate variant quality scores.

CNV calling

Timing: 2.5 h, 45 min hands-on

50 | Download the mappability track for the reference genome (in this example, hg19) created by GEM[44], and use the file *out100m2_hg19.gem* for a read length of 100 bp (Illumina) and up to 2 mismatches, using the following command:

```
$ wget https://xfer.curie.fr/get/nil/7hZIk1C63h0/hg19_len100bp.tar.gz
$ tar -axvf hg19_len100bp.tar.gz
```

51 | Create a config file for Control-FREEC run. Supply the mappability track (Step 50) and a text file with the chromosome lengths of the reference genome (*hg19.GATK.len.txt*), using the following command:

```
$ cat freec_config.txt
[general]
chrLenFile=hg19.GATK.len.txt
ploidy=2
samtools=<path to samtools directory>
chrFiles=<path to directory with chromosomes fasta files>
window=1000
```

```
maxThreads=8
outputDir=<path to outdir>
gemMappabilityFile=out
100m2_hg19.gem
[sample]
mateFile=recal_reads.bam
inputFormat=BAM
mateOrientation=FR
```

52 | Run Control-FREEC to create a **_CNVs* file with coordinates of predicted copy-number alterations, and a **_ratio.txt* file with ratios and predicted copy-number alterations for each window (1 kb) in the genome, as follows:

```
$ freec -conf freec_config.txt
```

53 | Plot the copy-number profile by running the makeGraph R script from Control-FREEC using the following command:

Critical step: Check if there are no large copy-number alterations by manually inspecting the copy-number profile and comparing it with unrelated samples. A ploidy of 2 is critical for the default settings of the SNV filtering in Step 65.

Troubleshooting

```
$ wget https://github.com/BoevaLab/FREEC/
raw/246df00589dca6800df97081f919f2e2177f7ebb/scripts/makeGraph.R$
cat makeGraph.R | R --slave --args 2 *_ratio.txt
```

54 | Rename the CNVs call file to a BED file format, as follows:

```
$ mv clone1_CNVs.txt clone1_CNVs.bed
```

CNV filtering

Timing: 15 min

55 | Filter out CNV calls that are within 1 kb (FREEC window size that was used) of CNVs that are listed in a CNV blacklist, using the following command. A CNV blacklist with recurrent CNVs can be created as described in Box 1, step 1A.

```
$ bedtools window -w 1000 -v -a clone1_CNVs.bed -b CNV_blacklist_
merged.bed > clone1_CNVs_filtered.bed
```

56 | Using the following command, select somatic events by excluding CNVs that overlap with CNVs in control samples:

```
$ bedtools intersect -v -a clone1_CNVs_filtered.bed -b control_CNVs.
bed > clone1_CNVs_filtered_somatic.bed
```

57 | Using the following command, filter out CNVs with size <50 kb:

```
$ cat clone1_CNVs_filtered_somatic.bed | awk '{if(($3-$2)>50000)
print}' > clone1_CNVs_filtered_somatic_clean.bed
```

Determination of callable loci

Timing: 4 h, 10 min hands-on

58 | Run the GATK CallableLoci to determine which positions in the genome are callable. This creates a file, *callableloci_all.bed*, with the callable status of each base in the genome. Then save all callable regions to *callableloci.bed*, as follows.

```
$ java -jar GenomeAnalysisTK.jar -T CallableLoci -R reference.
fa -I recal_reads.bam -o callableloci_all.bed --summary
callableloci-summary.txt --minBaseQuality 0 --minMappingQuality 0
--maxFractionOfReadsWithLowMAPQ 1 --minDepth 20$ grep "CALLABLE"
callableloci_all.bed > callableloci.bed
```


59 | Intersect the callable BED file for the organoid culture and the control sample using bedtools, as follows. This creates a BED file with the genomic regions that are callable in both the organoid culture and the control sample. In case multiple organoid cultures are derived from one individual, the intersection should be performed for each control–clone pair.

```
$ bedtools intersect -a clone1_callableloci.bed -b control_
callableloci.bed > clone1_control_callableloci.bed
```

SNV calling and processing

Timing: 28 h, 1 h hands-on

60 | Follow the GATK best practices basic protocol 2 (ref. 36) to perform the per-sample calling using HaplotypeCaller, followed by joint genotyping of all samples that belong to the same individual (control sample + all clonal organoid samples). This creates a multisample VCF file, *raw_variants.vcf*, that contains all the positions that HaplotypeCaller evaluated to be potentially variant, both SNVs and indels.

61 | Extract the SNVs from the call set with GATK SelectVariants, as follows:

```
$ java -jar GenomeAnalysisTK.jar -T SelectVariants -V raw_variants.
vcf -R reference.fa -selectType SNP -o raw_SNVs.vcf
```

62 | Add quality filters to the SNV call set using GATK VariantFiltration, as follows. This creates a VCF file, *filtered_SNVs.vcf*, that contains all the original SNVs, annotated in the FILTER column with either "PASS" or the name of the filter that was failed.

```
$ java -jar GenomeAnalysisTK.jar -T VariantFiltration -R reference.
fa -V raw_SNVs.vcf --filterExpression "MQ0 >=4 & ((MQ0 / (1.0 * DP))
> 0.1)" -filterName "HARD_TO_VALIDATE" --filterExpression "QUAL <
30.0" -filterName "VeryLowQual" --filterExpression "QUAL > 30.0 & QUAL
< 50.0" -filterName "LowQual" --filterExpression "QD < 1.5" -filterName
"LowQD" -o filtered_SNVs.vcf
```

SNV filtering

Timing: 30 min

63 | Using the following command, remove SNV positions that coincide with an insertion or deletion by excluding positions with an asterisk in the ALT column:

```
$ grep -v ",\*" filtered_SNVs.vcf > clean_SNVs.vcf
```

64 | Using the following command, create the config file with all the paths to the directories of the tools that are needed to run SNVFI and set the maximum number of threads that can be used by SNVFI:

```
$ cat SNVFI.config
SNVFI_ROOT=<path to SNFVI install directory>
BIOVCF_PREFIX=<path wherein the bio-vcf program can be found>
TABIX_PREFIX=<path wherein the tabix programs can be found>
VCFTOOLS_PREFIX=<path wherein the vcftools programs can be found>
R_PREFIX=<path wherein the R program can be found>
RSCRIPT=<path to SNVFI_filtering_R.R R-script>
MAX_THREADS=<maximum number of threads used by SNFVI>SGE="NO"
```

65 | Using the following command, create the INI file that contains the settings of the SNVFI pipeline. Indicate the path to the input VCF file, the 1-based sample number of the organoid culture (subject sample) in the VCF file and that of the control sample of the same individual. In addition, indicate the path to the blacklist files (optional); a *SNV_blacklist.vcf* file can be created as described in Box 1, step 1B; the *dbSNP137.vcf* file contains the variants in the Single Nucleotide Polymorphism Database v137.b37[37]. We advise using the following SNVFI filtering settings for WGS data with ~30× overall coverage and a ploidy of 2: base call quality >100, minimum coverage at variant position >20 and VAF-filtering threshold of 0.3. These settings can be adjusted for data with different coverage and/or ploidy states.

```
$ cat SNVFI_run1.ini
SNV=clean_SNVs.vcf
SUB=<Clone sample number in vcf (1-based)>
CON=<Control sample number in vcf (1-based)>
OUT_DIR=<Output directory>
BLACKLIST=(
'SNV_blacklist.vcf'
'dbSNP137.vcf'
);
QUAL=100
COV=20
VAF=0.3
CLEANUP=YES
```

66 | Execute SNVFI as follows to create a VCF file, *clonesample_controlsample_final.vcf*, which contains the positions that pass all SNVFI filter steps, and a PDF file, *clonesample_controlsample_VAF.pdf*, with the VAF plot of all positions that remain before VAF filtering, which can be used to determine the clonality of the sample (Fig. 6). In addition, a TXT file, *clonesample_controlsample_filter_count.txt*, is created, that

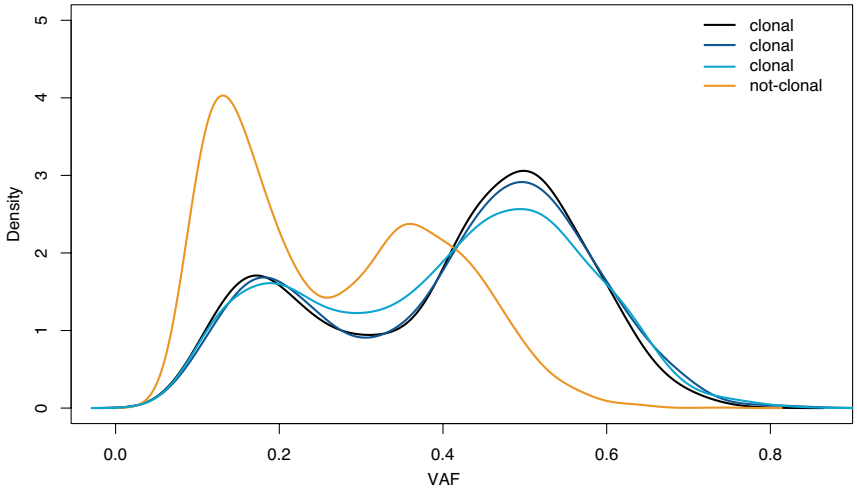


Figure 6. Variant allele frequency (VAF) density plots for 4 organoid cultures. A VAF density plot of all SNVs that remain before VAF filtering in Step 66 is automatically generated by SNVFI. These VAF density plots can be used to determine the clonality of a culture. An organoid culture with nonclonal origin is depicted in orange.

contains the total number of mutations that are retained after each filtering step of SNVFI.

```
$ sh SNVFI_run.sh SNVFI.config SNVFI_run1.ini
```

67 | Select mutations in genomic regions that are callable in both the clone and the control sample, using the following command:

```
$ bedtools intersect -a clone1_control_final.vcf -b clone1_control_
callableloci.bed > clone1_SNVs_callable.vcf
```

68 | Validate the clonal origin of the sequenced ASC culture by assessing the VAF plot output of SNVFI.

Critical step: A distribution around VAF = 0.5 is to be expected for heterozygous somatic variants (Fig. 6). A distribution around VAF = 0.2 is to be expected for subclonal mutations and sequencing noise. When the rightmost peak is shifted to the left (VAF <0.45), this indicates that the organoid culture did not arise from a single stem cell. Exclude all nonclonal cultures from the analysis. Of note, somatic homozygous variants (VAF ~1) are not expected without selection or loss of heterozygosity.

Troubleshooting

Box 1: Blacklist creation for variant filtering

Timing: 30 min

Follow option A to create a CNV blacklist or option B to create a SNV blacklist.

(A) CNV blacklist creation

(i) Collect at least three unfiltered CNV call sets of control samples of different and unrelated individuals that were obtained using Steps 50–54 of the PROCEDURE.

(ii) Using the following command, merge the regions in each CNV BED file:

```
$ bedtools merge-i control_CNVs.bed > control_merged_CNVs.bed
```

(iii) Using the following command, append the merged control CNV BED files:

```
$ cat *control_merged_CNVs.bed > ALL_controls_CNVs.bed
```

(iv) Sort the BED file, using the following command:

```
$ sort -k1,1 -k2,2n ALL_controls_CNVs.bed > ALL_controls_CNVs_
sorted.bed
```

(v) Using the following command, count the number of samples that have a CNV for each base pair in the genome. Supply the text file with the chromosome lengths of the reference genome (*hg19.GATK.len.txt*).

```
$ bedtools genomecov -i ALL_controls_CNVs_sorted.bed -g hg19.GATK.
len.txt -bg > genome_CNV_recurrency.bed
```

(vi) Select all genomic positions that have a CNV event in at least two samples, and merge them using the following command. This creates a BED file, *CNV_blacklist.bed*.

```
$ cat genome_CNV_recurrency.bed | awk '{if ($4>=2) print}' > CNV_
blacklist.bed
```

(B) SNV blacklist creation

(i) Collect at least three unfiltered SNV call sets of control samples of different and unrelated individuals that were obtained using Steps 60 and 61.

(ii) Using the following commands, compress the VCF files using bgzip and index them with tabix:

```
$ bgzip sample1_control_raw_SNVs.vcf$ tabix -p vcf sample1_control_
raw_SNVs.vcf.gz
```

(iii) Using bcftools, create a VCF file with all positions that occur in at least two samples, with the following command. This creates a VCF file, *SNV_blacklist.vcf*.

```
$ bcftools isec -n +2 *control_raw_SNVs.vcf.gz > SNV_blacklist.vcf
```

TROUBLESHOOTING

Step	Problem	Possible reason	Solution
Generating a single cell suspension from organoid cultures			
8	The single cell suspension still contains cell clumps	Inefficient dissociation	Increase dissociation time. Incubate the cell suspension 5 more min at 37°C in the water bath and pipette up and down five times using a glass pasteur pipette with a narrowed tip again.
12	Yield of 0 < 100 cells after FACS	Not enough organoids in the wells used for the dissociation with TrypLE	Combine more wells to get enough organoids.

		Organoids stick to pasteur pipette	Prewash the pasteur pipette in TrypLE or Adv+++ by pipetting up and down five times prior to pipetting the organoids.
		Cells were not dissociated sufficiently	Check the cell suspension using the inverted microscope in step 8. The cell suspension should predominantly consist of single cells. Otherwise, increase dissociation time. Incubate the cell suspension 5 more min at 37°C in the water bath and pipette up and down five times using a glass pasteur pipette with a narrowed tip again.
		Cell pellet was aspirated	Carefully aspirate the supernatant from the single cell pellets, as these pellets are often not visible to the naked eye. Remove the supernatant by using a P1000 pipette and a P200 pipette and leave 5 µl of supernatant behind on top of the cell pellet in the eppendorf.
	Yield of 100 < 10,000 cells after FACS	Possible reasons as described above	Proceed with the protocol. Resuspend the pellet in a smaller volume of BME in step 15 and/or skip the highest cell dilutions in step 16. Usually, organoids will still appear in the low cell dilutions.
Organoid maintenance			
20	No organoids grew from single cells	Organoids are still growing, but not visible yet	Wait at least 14 days after the cell sort. Then check the highest cell dilutions for organoids.
		Stem cells are differentiating or quiescent	Verify whether unsorted bulk cultures can grow using the same culture medium. If so, restart the entire protocol, preferably with another bulk organoid culture. If the bulk cultures also stop growing, check which component in the medium might be responsible for this.
Generating clonal organoid cultures			
24	Organoid is not transferred to the new BME droplet	Organoid floats in the medium	Suck up the organoid using your P200 from the medium and transfer it to the BME droplet in the 4 wells plate. Be careful not to dilute the BME with too much culturing medium, as it will not set at 37°C.
		Organoid sticks to forceps	Using a needle, gently detach the organoid from the forceps and transfer it to the BME droplet in the 4 wells plate.
		Organoids are too small to pick by forceps	Wait a few more days and try again.

25	Organoid cannot be cut into smaller pieces	Organoid moves around in BME	Practice makes perfect. Hold down the organoid with one needle. Slice the organoid using another needle.
		Organoid is too small	Wait a few more days and try again.
DNA isolation			
36	No cell pellet	As described above at step 12	As described above at step 12.
44	< 1 µg DNA	Not enough material used for DNA isolation	Perform DNA isolation again using a larger amount of input material: a larger biopsy/amount of blood as input for the control sample or more than six wells of a 24 well plate of organoid culture.
Whole-genome sequencing			
45	No library or insufficient quality of library	DNA is not pure	Check the purity of the DNA on the Nanodrop 2000 (Thermo Fisher Scientific, cat. no. ND-2000) prior to sequencing. Pure DNA has a 260/280 value of > 1.8 and a 260/230 value of > 2.0. If the DNA is impure, incubate the sample with lysis buffer, Proteinase K and RNase A, according to the DNeasy Blood & Tissue Handbook of Qiagen. Subsequently purify your DNA using 2 volumes AMPure XP beads (Beckman Coultier, cat. no. A63880).
		DNA was fragmented before library preparation	Check the size of the DNA by Gel electrophoresis (~0.8% agarose gel) prior to sequencing. Genomic DNA should hardly move through the gel and one band should be visible near the well. If you see a smear, your DNA is fragmented. Perform a size-selection using 2 volumes AMPure XP beads (Beckman Coultier, cat. no. A63880).
46	Not ~equal amount of reads for all 8 libraries	Differences in amplification efficiency between libraries	Assess amplification efficiency of libraries through a qPCR, as described in the TruSeq® Nano DNA Library Prep Reference Guide at 'Quantify libraries'.
CNV calling			
53	Ploidy deviates from 2	Mapping artefacts	Check copy number profiles of an unrelated sample. If they show the same pattern, these typically small events can be ignored as these are not in your CallableLoci.

		Copy number variant	1. Exclude CNV regions from analysis and continue with default SNV filtering. 2. Loosen the SNV filtering criteria by lowering the VAF threshold and depth threshold in step 65. The VAF threshold should not be lower than 0.1 as these variants are likely sequencing errors[46]. 3. Use e.g. the Battenberg algorithm[2] or SomVarIUS[47] to account for local ploidy state prior to SNV filtering and adjust SNV filtering criteria accordingly.
SNV filtering			
68	Many variants with low VAF	Subject and control mixed up	Check samples in filename of final.vcf and filter_ count.txt file (output step 66).
		Sequencing noise	Sequencing noise typically has a low VAF. If you can still distinguish the noise peak from the peak at VAF = 0.5 (see Figure 5), you can continue. Otherwise, the sequencing noise might interfere with your data analysis. Try to figure out why the sequencing data is noisy. Perhaps use another sequencing platform if you are not using Illumina.
		Many subclonal mutations	This could indicate that the mutation rate is higher in culture, e.g. due to minor differences in culturing conditions. We advise you to assess the *in vitro* mutation rate to reassure that the organoids did not accumulate many mutations prior to the single cell step. As long as there is a peak at VAF = 0.5, you can use this sample.
	Peak at VAF ~ 1	Loss of heterozygosity or selection	Assess CNV calls (steps 50 - 57) at positions of SNVs with VAF ~ 1, to check for copy number losses.
		Germline variants were not excluded: wrong control sample used	Check samples in filename of final.vcf.
	No peak at VAF = 0.5	The organoid culture was not clonal	If you have isolated the DNA of multiple clones, sequence another clone and repeat the data analysis (steps 47 - 68). Otherwise, generate additional clonal organoid cultures and repeat the data analysis (steps 1 - 68).

Table 1. Troubleshooting table.

TIMING

Steps 1–12, generation of a single-cell suspension from organoid cultures: 1.5 h
Steps 13–18, plating of single cells at a limited dilution: 1 h
Steps 19–21, organoid maintenance: 3 weeks, 4 h hands-on
Steps 22–28, generation of clonal organoid cultures: 1 h

Step 29A, organoid expansion to catalog mutations acquired during life (Route A): 3 weeks, 3 h hands-on
Step 29B, organoid expansion to catalog mutations acquired during culture (Route B): 3.5–5.5 months, 22 h hands-on
Steps 30–44, DNA isolation: 2 d, 4 h hands-on
Steps 45 and 46, whole-genome sequencing: 5 d, 2 d hands-on
Steps 47–49, read mapping and data preprocessing: 48 h, 4 h hands-on
Steps 50–54, CNV calling: 2.5 h, 45 min hands-on
Steps 55–57, CNV filtering: 15 min
Steps 58 and 59, determination of callable loci: 4 h, 10 min hands-on
Steps 60–62, SNV calling and processing: 28 h, 1 h hands-on
Steps 63–68, SNV filtering: 30 min
Box 1, blacklist creation: 30 min

ANTICIPATED RESULTS

This protocol is uniquely suited for measuring mutation accumulation in single human ASCs during life (Route A) and during culture (Route B). This protocol yields both data and biological resources (Table 2). Using existing methods, the somatic mutation catalogs (CNV and SNV) can be subjected to in-depth mutational pattern analyses to study the DNA damage and repair mechanisms that have shaped the genomes of the single ASCs[45,46]. In addition, the callable genome size is determined, which is

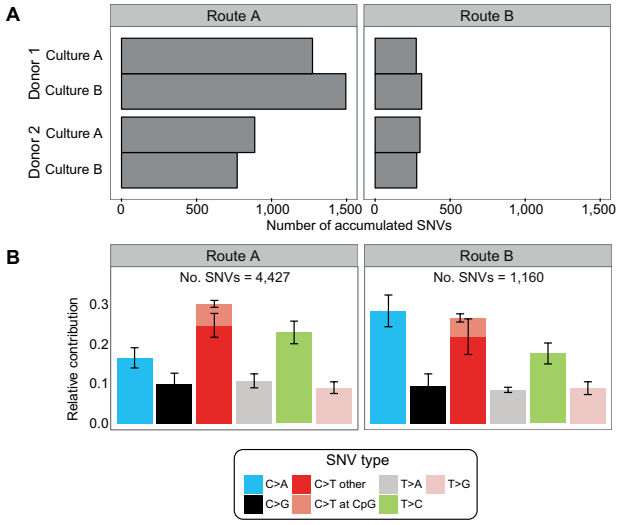


Figure 7. SNV accumulation measurements of four human liver ASCs of two donors aged 46 (Donor 1) and 30 years (Donor 2) for Routes A and B. (A) The number of somatic SNVs that accumulated during life (Route A) and during 3 months of organoid culturing under normal conditions (Route B). (B) Relative contribution of the indicated base substitution types to the mutation spectrum for Routes A and B. Data are represented as the mean relative contribution of each mutation type over all ASCs (n = 4) for each route. Error bars represent the standard deviation. The total number of SNVs is indicated. This figure is adapted with permission from ref. 1, Nature Publishing Group.

Type	Description	Applications
Clonal organoid culture	Cryopreserved clonal organoid lines	Functional follow-up assays such as: RNA Sequencing, ChIP-Sequencing, Western Blot and Immunohistochemistry
Genomic DNA	Genomic DNA from clonal organoid lines and reference sample	• Variant validations • Long-read sequencing • Bisulfite sequencing
Somatic SNVs (VCF file)	Genome-wide catalogue of SNVs that accumulated in a single ASC.	Mutational pattern/mechanism analysis
Somatic CNVs (BED file)	Genome-wide catalogue of CNVs that accumulated in a single ASC.	Mutational pattern/mechanism analysis
Aligned reads (BAM file)	The aligned reads for each sample.	Complex structural variation calling
Callable genomic regions (BED file)	Callable genomic regions for each sample-control pair.	Statistical analysis of genomic distribution of mutations

Table 2. Overview results.

important for downstream quantitative analyses, e.g., to allow comparison between mutation loads of different samples. The WGS data can be used to perform other bioinformatic analyses of, e.g., complex structural variation and large chromosomal aberrations[47,48]. In addition, cryopreserved passages can be thawed to perform functional follow-up studies on the same clonal organoid culture, such as RNA sequencing or immunohistochemistry[29]. Cryopreserved passages can also be used to perform additional Route B assays, such as chemical and genetic perturbations. Isolated DNA can be resequenced to validate identified variants using targeted deep sequencing or to study complex structural variation using a long-read sequencing platform[49].

Approximately 92% of the human liver organoid cultures give rise to new organoid cultures after flow cytometry, and ~93% of these cultures are clonal (Supplementary Table 1). For Route A, we show the illustrative results from measuring mutation accumulation in four liver ASCs. ASCs from a donor of 30 years contained on average 830 somatic SNVs, and those from a donor of 55 years contained on average 1,384 SNVs. The majority of the mutations are C>T and T>C substitutions (Fig. 7). The somatic mutation type and rate are similar for other normal human liver ASCs, regardless of age or gender of the donor[1]. Liver ASCs acquire ~36 SNVs per year (95% confidence interval is 11.9–60.1). Extensive validations showed an overall confirmation rate of ~91% of the final SNV catalogs. In 4 out of 10 liver ASCs, a somatic CNV was observed that was introduced during life, predominantly copy-

number gains[1].

For Route B, human liver ASCs acquire on average 290 SNVs during 3 months of culture under normal conditions (Fig. 7a). We observed no selection for SNVs *in vitro*. We detected only one nonsynonymous SNV in four human liver organoid cultures after 3 months of culture[10]. During culture (Route B), relatively more C>A and fewer T>C substitutions are induced, as compared with Route A (Fig. 7b). This indicates that different DNA damage and/or repair processes are operative during life and culture. We did not observe induction of CNVs *in vitro* in liver ASCs under normal culture conditions.

ACKNOWLEDGEMENTS

The authors thank J. de Ligt for his input on the CNV analysis, and J.F. van Velzen for his input on the FACS procedures. This study was financially supported by a Zenith grant from the Netherlands Genomics Initiative (935.12.003) and funding from the NWO Zwaartekracht program Cancer Genomics.nl to E.C., and funding from Worldwide Cancer Research (WCR, grant no. 16-0193) to R.v.B.

AUTHOR CONTRIBUTIONS

M.J., F.B., H.C., R.v.B. and E.C. wrote the manuscript. M.J., R.v.B. and V.S. developed the wet lab protocol, and N.B. tested the protocol. The bioinformatics pipeline was developed by F.B. and R.v.B., implemented by F.B. and tested by S.B. and R.J.

COMPETING INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).
2. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009)
3. Zhu, L. *et al.* Multi-organ mapping of cancer risk. *Cell* **166**, 1132–1146.e7 (2016).
4. Adams, P.D., Jasper, H. & Rudolph, K.L. Aging-induced stem cell mutations as drivers for disease and cancer. *Cell Stem Cell* **16**, 601–612 (2015).
5. Behjati, S. *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 (2014).
6. Sato, T. *et al.* Single Lgr5 stem cells build crypt–villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
7. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
8. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15**, 585–598 (2014).
9. Clevers, H. Modeling development and disease with organoids. *Cell* **165**, 1586–1597 (2016).
10. Huch, M. *et al.* Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**, 299–312 (2015).
11. Drost, J. *et al.* Organoid culture systems for prostate epithelial and cancer tissue. *Nat. Protoc.* **11**, 347–358 (2016).
12. Boj, S.F. *et al.* Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**, 324–338 (2015).

13. Sun, B., Beicheng, S. & Michael, K. Obesity, inflammation, and liver cancer. *J. Hepatol.* **56**, 704–713 (2012).

14. Hoeijmakers, J.H.J. DNA damage, aging, and cancer. *N. Engl. J. Med.* **361**, 1475–1485 (2009).

15. Kuijk, E.W. *et al.* Generation and characterization of rat liver stem cell lines and their engraftment in a rat model of liver failure. *Sci. Rep.* **6**, 22154 (2016).

16. Vermeij, W.P. *et al.* Restricted diet delays accelerated ageing and genomic stress in DNA-repair-deficient mice. *Nature* **537**, 427–431 (2016).

17. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

18. Schwank, G. & Clevers, H. CRISPR/Cas9-mediated genome editing of mouse small intestinal organoids. *Methods Mol. Biol.* **1422**, 3–11 (2016).

19. Alexandrov, L.B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).

20. Fitzgerald, D.M., Hastings, P.J. & Rosenberg, S.M. Stress-induced mutagenesis: implications in cancer and drug resistance. *Annu. Rev. Cancer Biol.* **1**, 119–140 (2017).

21. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).

22. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).

23. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).

24. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).

25. Gawad, C., Koh, W. & Quake, S.R. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**, 175–188 (2016).

26. Takasato, M., Minoru, T., Er, P.X., Chiu, H.S. & Little, M.H. Generation of kidney organoids from human pluripotent stem cells. *Nat. Protoc.* **11**, 1681–1692 (2016).

27. Huang, S.X.L. *et al.* Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 84–91 (2014).

28. Ronen, D. & Benvenisty, N. Genomic stability in reprogramming. *Curr. Opin. Genet. Dev.* **22**, 444–449 (2012).

29. Broutier, L. *et al.* Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. *Nat. Protoc.* **11**, 1724–1743 (2016).

30. Bartfeld, S. *et al. In vitro* expansion of human gastric epithelial stem cells and their responses to bacterial infection. *Gastroenterology* **148**, 126–136.e6 (2015).

31. Kessler, M. *et al.* The Notch and Wnt pathways regulate stemness and differentiation in human fallopian tube organoids. *Nat. Commun.* **6**, 8989 (2015).

32. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).

33. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.* **25**, 681–686 (2007).

34. van Heesch, S. *et al.* Systematic biases in DNA copy number originate from isolation procedures. *Genome Biol.* **14**, R33 (2013).

35. Boeva, V. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).

36. Van der Auwera, G.A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10. (2013).

37. Sherry, S.T. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

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

39. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

40. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

41. Gentleman, R. *R Programming for Bioinformatics* (CRC Press, 2008).

42. Quinlan, A.R. BEDTools: The Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–11.12. (2014).

43. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

44. Derrien, T. *et al.* Fast computation and applications of genome mappability. *PLoS One* **7**, e30377 (2012).

45. Blokzijl, F., Janssen, R., Van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Preprint at *bioRxiv* http://doi.org/10.1101/071761 (2017).

46. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J. & Stratton, M.R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).

47. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).

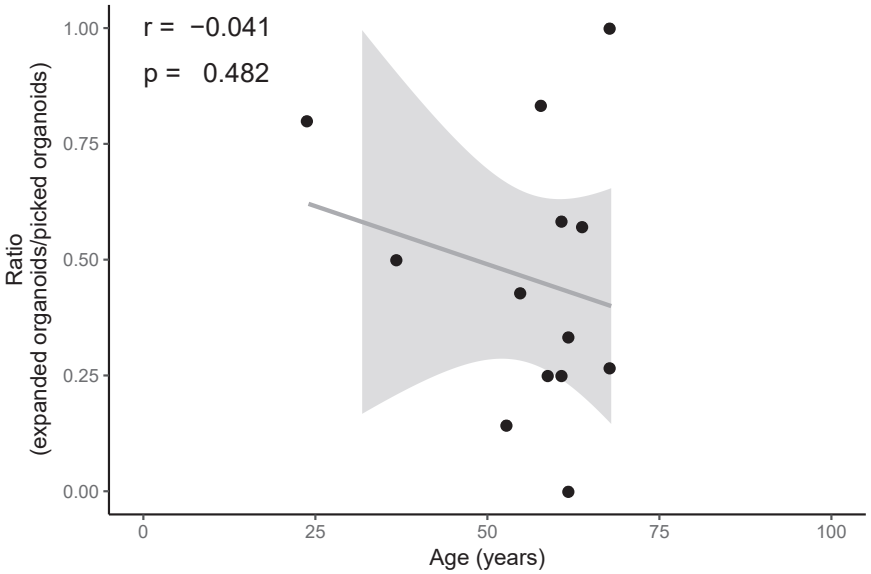
48. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).

49. Huddleston, J. *et al.* Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.* **27**, 677–685 (2016).

50. Chen, L., Liu, P., Evans, T.C. Jr. & Ettwiller, L.M. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science* **355**, 752–756 (2017).

51. Smith, K.S. *et al.* SomVarIUS: somatic variant identification from unpaired tissue samples. *Bioinformatics* **32**, 808–813 (2016).

SUPPLEMENTAL FIGURES AND TABLES



Supplemental figure S1. Outgrowth potential for organoid formation after picking clonal organoids. Each dot represents a single human donor. There is no correlation between the age of the human donor and the number of picked organoids that were expanded (correlation = -0.041, p-value = 0.482).

Donor	Age (years)	Nr. of picked organoids (step 24-25)	Nr. of expanded organoids (after step 29)	Nr. of sequenced organoids (step 45-46)	Nr. of confirmed clonal organoids (after step 66)
a	24	5	4	1	1
b	37	4	2	1	1
c	53	7	1	1	1
d	55	7	3	1	1
e	58	6	5	1	1
f	59	4	1	1	1
g	61	4	1	1	0
h	61	12	7	2	2
i	62	4	0	NA	NA
j	62	9	3	NA	NA
k	64	7	4	2	2
l	68	15	4	2	2
m	68	4	4	2	2
Total		**88**	**39**	**15**	**14**
Percentage*		**44%**		**93%**	

NA = Not applicable: no clones were sequenced for this donor.
* These percentages indicate the success rate for generating organoid cultures (44%) and the number of cultures that were confirmed to be clonal after sequencing (93%), respectively.

Supplemental table S1. Outgrowth potential and clonality of organoid cultures.

A continuous flow of mutagenic stress

Chapter 3

Tissue-specific mutation accumulation in human adult stem cells during life

Francis Blokzijl[1,2], Joep de Ligt[1,2,#], Myrthe Jager[1,2,#], Valentina Sasselli[2,#], Sophie Roerink[3,#], Nobuo Sasaki[2], Meritxell Huch[2,7], Sander Boymans[1,2], Ewart Kuijk[1,2], Pjotr Prins[2] , Isaac J. Nijman[2], Inigo Martincorena[3], Michal Mokry[4], Caroline L. Wiegerinck[4], Sabine Middendorp[4], Toshiro Sato[2], Gerald Schwank[2], Edward E.S. Nieuwenhuis[4], Monique M.A. Verstegen[5], Luc J.W. van der Laan[5], Jeroen de Jonge[5], Jan N.M. IJzermans[5], Robert G. Vries[6], Marc van de Wetering[2], Michael R. Stratton[3], Hans Clevers[2], Edwin Cuppen[1,2] and Ruben van Boxtel[1,2]

1 Center for Molecular Medicine, Cancer Genomics Netherlands, Department of Genetics, University Medical Center Utrecht, Heidelberglaan 100, 3584CX Utrecht, The Netherlands
2 Hubrecht Institute for Developmental Biology and Stem Cell Research, KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584CT Utrecht, The Netherlands
3 Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB101SA, UK
4 Department of Pediatrics, University Medical Center Utrecht, Lundlaan 6, 3584 EA Utrecht, The Netherlands
5 Department of Surgery, Erasmus MC-University Medical Center, 3000 CA Rotterdam, the Netherlands
6 Foundation Hubrecht Organoid Technology (HUB), Uppsalalaan 8, 3584CT Utrecht, The Netherlands
7 Wellcome Trust/Cancer Research UK Gurdon Institute, Wellcome Trust/MRC Stem Cell Institute and Department of Physiology, Development and Neuroscience, University of Cambridge, Tennis Court Road, CB2 1QN Cambridge, UK
# Equal contribution

Adapted from: Nature 2015 Jan; 538(7624): 260-264

ABSTRACT

The gradual accumulation of genetic mutations in human adult stem cells (ASCs) during life is associated with various age-related diseases, including cancer[1,2]. Extreme variation in cancer risk across tissues was recently proposed to depend on the lifetime number of ASC divisions, owing to unavoidable random mutations that arise during DNA replication[1]. However, the rates and patterns of mutations in normal ASCs remain unknown. Here we determine genome-wide mutation patterns in ASCs of the small intestine, colon and liver of human donors with ages ranging from 3 to 87 years by sequencing clonal organoid cultures derived from primary multipotent cells[3,4,5]. Our results show that mutations accumulate steadily over time in all of the assessed tissue types, at a rate of approximately 40 novel mutations per year, despite the large variation in cancer incidence among these tissues[1]. Liver ASCs, however, have different mutation spectra compared to those of the colon and small intestine. Mutational signature analysis reveals that this difference can be attributed to spontaneous deamination of methylated cytosine residues in the colon and small intestine, probably reflecting their high ASC division rate. In liver, a signature with an as-yet-unknown underlying mechanism is predominant. Mutation spectra of driver genes in cancer show high similarity to the tissue-specific ASC mutation spectra, suggesting that intrinsic mutational processes in ASCs can initiate tumorigenesis. Notably, the inter-individual variation in mutation rate and spectra are low, suggesting tissue-specific activity of common mutational processes throughout life.

MAIN

It has not yet been possible to measure somatic mutation loads in ASCs from specific human tissues. However, such knowledge could be valuable in understanding tissue homeostasis and repair capacities as well as ASC vulnerabilities to extrinsic factors. The accumulation of mutations as life progresses is thought to underlie the genesis of age-related diseases such as cancer[6] and organ failure[2]. Mutations acquired in the genomes of multipotent ASCs are believed to have the largest impact on the mutational load of tissues, owing both to their potential for self-renewal and capacity to propagate mutations to their daughter cells[1,2]. Consistently, cancer-initiating mutations in intestinal ASCs lead to tumour formation within weeks, whereas these mutations fail to drive intestinal adenomas when induced in differentiated cells[7]. Unavoidable random mutations that arise during DNA replication in normal ASCs have recently been proposed to impart a large influence on cancer risk[1]. Consequently, tissues with a high ASC turnover would show higher cancer incidence when compared to tissues with low ASC proliferation rates[1,8]. However, computational modelling has suggested that the variation in ASC proliferation rate alone cannot exclude extrinsic

risk factors as important determinants of organ-specific cancer incidence[9]. Yet, the number of mutations that accumulate during the lifespan of normal human ASCs with different turnover rates has, to date, not been directly determined and compared. To understand tissue homeostasis and tissue-specific susceptibility to cancer and ageing-associated diseases it is important to assess mutation accumulation in ASCs of different tissues.

Here, we experimentally define ASCs as those cells that give rise to long-term organoid cultures and have the potential to differentiate into multiple tissue-specific cell types[3,4,5]. To catalogue the *in vivo*-acquired somatic mutations in individual normal human ASC genomes, we used an *in vitro* system to expand single ASCs into epithelial organoids, which reflect the genetic make-up of the original ASC (Extended Data Fig. 1a and Methods). This procedure allowed us to obtain sufficient DNA for accurate whole-genome sequencing (WGS) analysis, while circumventing the high noise levels associated with single-cell DNA amplification[10]. We assessed ASCs from the small intestine, colon and liver, tissues that differ greatly in proliferation rate and cancer risk[1]. Cancer incidence is much higher in the colon compared to the small intestine and liver[1]. We sequenced 45 independent clonal organoid cultures derived from 19 donors ranging in age from 3 to 87 years (Extended Data Table 1). In addition, we sequenced a blood or polyclonal biopsy sample of each donor to identify and exclude germline variants. Subclonal mutations, which must have been introduced *in vitro* after the single-cell step, were discarded based on their low variant-allele frequency (Extended Data Figs 1b–d, 2 and Methods). Overall, we identified 79,790 heterozygous clonal somatic point mutations and subsequent extensive validations showed an overall confirmation rate of approximately 91% (Extended Data Figs 1, 3).

A positive correlation (*t*-test linear mixed model; $P < 0.05$) between the number of somatic point mutations and the age of the donor could be observed for all organs (Fig. 1a and Extended Data Fig. 4), indicating that ASCs gradually accumulate mutations with age, independent of tissue type. Notably, we found that the annual mutation rate in ASCs was in the same range for all assessed tissues, despite the dissimilar cancer incidence in these tissues; ASCs of the colon, small intestine and liver accumulate around 36 mutations per year (95% confidence intervals are 26.9–50.6, 25.8–43.6 and 11.9–60.1, respectively; Fig. 1b). The mutation spectra in small intestinal and colon ASCs were very similar, but differed markedly from liver (Fig. 1c). Notably, the mutation spectrum within tissues did not differ between young and elderly donors (Extended Data Fig. 5).

Genome-wide mutation patterns in the ASCs provide insights into the mutational and DNA repair processes that are active in different organs[11]. Using non-negative matrix factorization[12], we extracted three mutational process signatures (Fig.

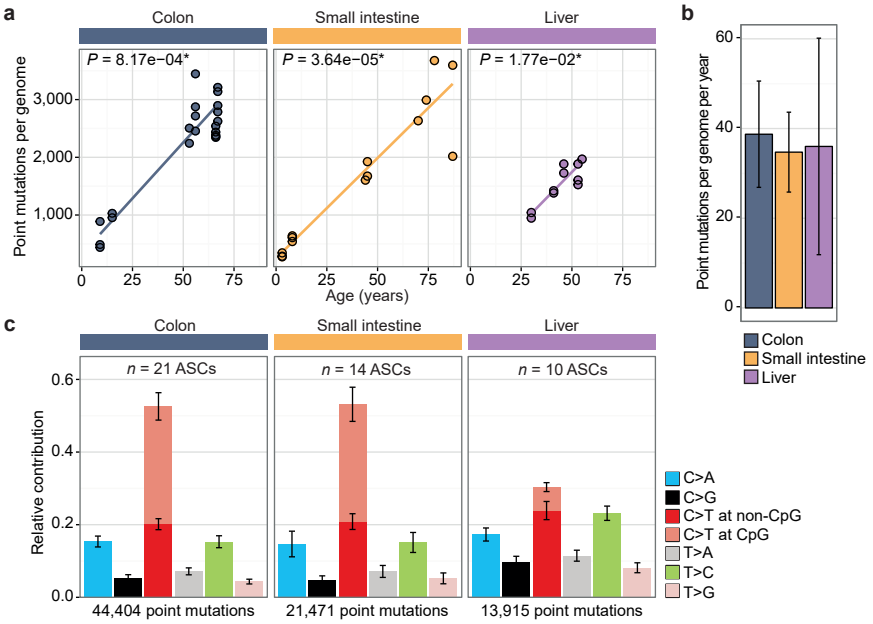


Figure 1. Age-associated accumulation of somatic point mutations in human ASCs.(A) Correlation of the number of somatic point mutations in each ASC type examined (extrapolated to the whole autosomal genome) with age of the donors per tissue. Each data point represents a single ASC. The *P* values of the age effects in the linear mixed model (two-tailed *t*-test) are indicated for each tissue. The sample sizes for colon, small intestine and liver ASCs are 6, 9 and, 5 donors, with, in total, 21, 14 and 10 ASCs, respectively. (B) Somatic mutation accumulation rate per tissue as estimated by the linear mixed models in (A). Error bars represent the 95% confidence intervals of the slope estimates. (C) Relative contribution of the indicated mutation types to the point mutation spectrum for each tissue type. Data are represented as the mean relative contribution of each mutation type over all ASCs per tissue type ($n = 21$, 14 and 10 for colon, small intestine and liver, respectively) and error bars represent standard deviation. The total number of identified somatic point mutations per tissue is indicated.

2a and Methods). All of these signatures were previously described in a pan-cancer analysis[11]. Signature A (corresponding to signature 5 in ref. 11), characterized by T:A to C:G transitions, was the main contributor to the mutation spectrum observed in the liver and was also clearly present in the small intestine and colon (Fig. 2). Although the underlying mutational process remains unknown, the number of mutations attributed to this signature that accumulate with age resembles a linear trend in all tissues (Fig. 2b). This suggests that this signature represents a universal genomic ageing mechanism (that is, a chemical process acting on DNA molecules) independent of cellular function or proliferation rate.

The majority of the somatic mutations observed in small intestinal and colon ASCs could be attributed to signature B (corresponding to signature 1A in ref. 11), which is characteristic of spontaneous deamination of methylated cytosine residues

into thymine at CpG sites (Fig. 2a). The resulting T:G mismatch can be effectively repaired, but the mutation is incorporated if DNA replication occurs before the repair is initiated[13]. In line with this, high rates of signature B mutations are observed in many cancer types of epithelial origin with high cell turnover[13]. This process showed a minimal contribution to the age-related mutational load in liver ASCs (Fig. 2c), which is likely to reflect the relatively low division rate of these cells during life. Finally, contribution of a third signature, signature C (corresponding to signature 18 in ref. 11), was minimal in all tissues and did not correlate with age (Fig. 2b). Sequential clonal ASC expansions in culture followed by WGS analysis showed that *in vitro*-induced mutations are predominantly characterized by this signature (Extended Data Fig. 6 and Methods).

Signature B mutations were strongly associated with the timing of replication and predominantly present in late-replicating DNA (Extended Data Fig. 7) even though the majority of CpG dinucleotides are located in early-replicating DNA. This bias suggests that this mutagenic process is more active in late-replicating DNA or, alternatively, that replication-coupled repair shows reduced activity in late-replicating

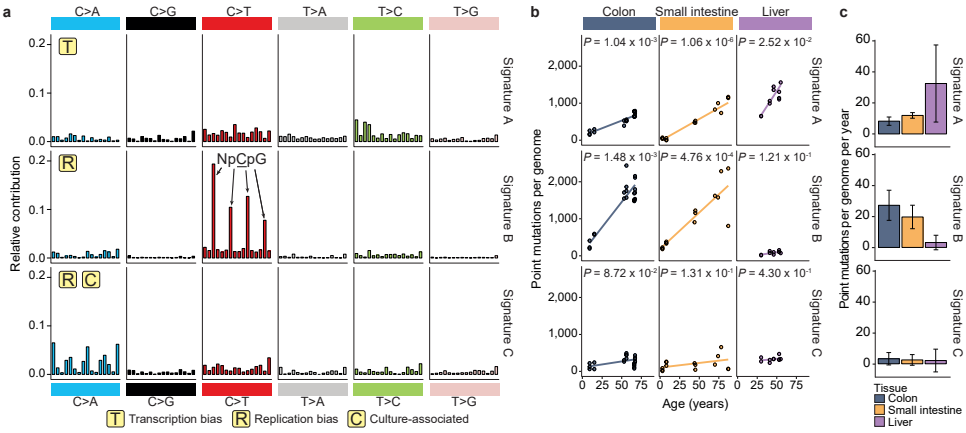


Figure 2. Signatures of mutational processes in human ASCs and their tissue-specific contribution. (A) Contribution of context-dependent mutation types to the three mutational signatures that were identified by non-negative matrix factorization (NMF) analysis of the somatic mutation collection observed in the ASCs across all assessed tissues. The contribution of each trinucleotide (order is similar to that in ref. 11) to each signature is shown. For each signature, the presence of transcriptional-strand bias, DNA-replication-timing bias and/or association with the culture system is indicated. (B) Absolute contribution of each mutational signature type (extrapolated to the whole autosomal genome) plotted against the age of the donors for each tissue. Each data point represents a single ASC. The *P* values of the age effects per tissue are shown (linear mixed model, two-tailed *t*-test). (C) Signature-specific mutation rate per year per genome for each tissue as estimated by the linear mixed model in (B). Error bars represent the 95% confidence intervals of the slope estimates.

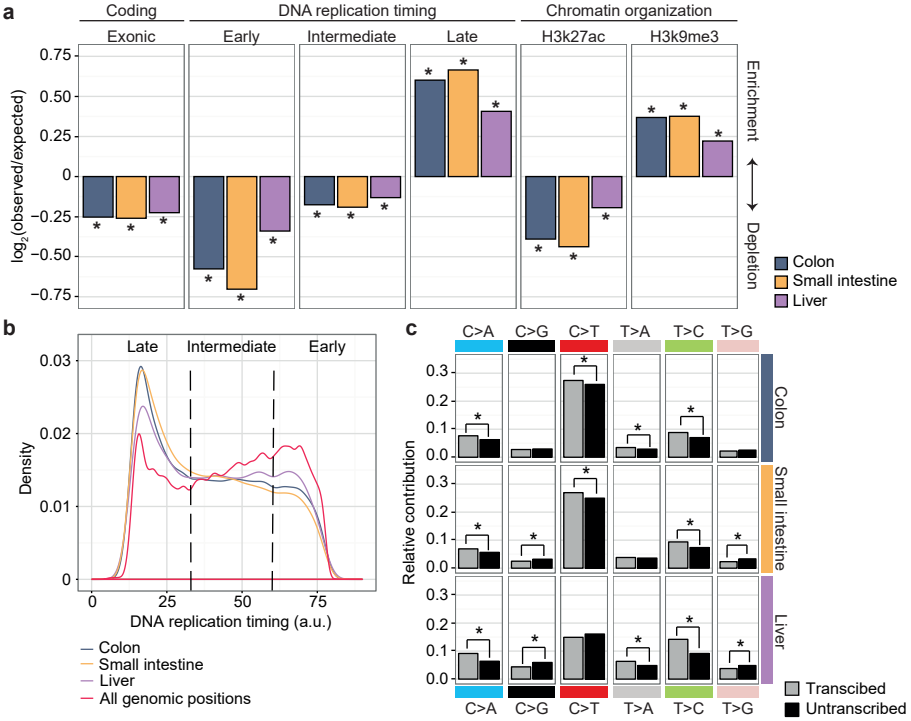


Figure 3. Non-random genomic distribution of somatic point mutations in ASCs. (A) Enrichment and depletion of somatic point mutations in the indicated genomic regions for each tissue. The $\log_2$ ratio of the number of observed and expected point mutations indicates the effect size of the enrichment or depletion in each region. *$P < 0.05$, one-sided binomial test. (B) Distribution of DNA replication timing for all genomic positions and the somatic point mutations detected in human ASCs per tissue. (C) Relative contribution of each point-mutation type on the transcribed and untranscribed strand for each tissue. *$P < 0.05$, two-sided Poisson test.

DNA[14]. Consequently, somatic mutations in small intestine and colon ASCs were strongly enriched in late-replicating DNA and depleted in early-replicating DNA (Fig. 3a, b). In addition, somatic point mutations in small intestine and colon ASCs were depleted in H3K27ac (histone H3 acetyl Lys27)-associated DNA and enriched in H3K9me3 (histone H3 trimethyl Lys9)-associated DNA (Fig. 3a), similar to patterns previously observed in cancer[15]. As genic regions are predominantly located in early-replicating DNA and open chromatin, we observed a depletion of mutations in exonic sequences (Fig. 3a). This demonstrates that genome-wide mutation rates and spectra cannot be reliably estimated using mutation discovery in reporter genes[16], such as the T-lymphocyte *HPRT* cloning assay[17], or by deep sequencing of genic regions[18,19,20,21]. To test whether the depletion of coding mutations was caused by selection against cells with damaging mutations, we calculated the ratio of non-synonymous to synonymous mutations (d*N*/d*S*) taking into account the mutation

spectra and sequence composition (see Methods)[18]. We did not observe negative selection for non-synonymous mutations (Extended Data Fig. 7f), arguing against the negative selection of cells with damaging protein-coding mutations.

In liver ASCs, somatic mutations are more randomly distributed throughout the genome and are less associated with replication timing or chromatin status (Fig. 3a). Nevertheless, a comparable depletion of exonic mutations was observed in all tissues (Fig. 3a), suggesting that liver ASCs use different mechanisms to maintain genetic integrity in functionally relevant regions. Signature A, the most predominant in liver ASCs, shows little bias towards DNA-replication-timing dynamics, but a pronounced transcriptional-strand bias[11] (Extended Data Fig. 7), consistent with activity of transcription-coupled repair[22]. In line with this, point mutations in the genic regions of the assessed ASCs showed a significant transcriptional strand bias, exemplified by the more frequent occurrence of T:A to C:G transitions on the transcribed strand compared to the untranscribed strand (Fig. 3c).

Our results indicate that a stable balance between the degree of DNA damage and the subsequent repair is maintained throughout life in various ASC types, since mutations accumulate steadily and display a constant mutation spectrum. Earlier work in mice using mutation-discovery in a *LacZ* reporter gene, showed major age-related changes in mutation spectra in different tissues[23]. The difference between these observations could be explained by the comprehensive genome-wide analysis applied here to ASCs, whereas reporter assays assess specific genes predominantly in differentiated cells. Although variation in tissue-specific mutation spectra in mice has been reported previously[23,24,25], we observed a difference in both mutation rate and spectrum in human cells (Extended Data Fig. 8). This indicates that mutation data derived from mice are not necessarily suitable for interpreting mutational processes and their consequences in humans.

Although we analysed cells from many different donors without controlling for lifestyle differences or gender, the point-mutation rate and spectrum were highly similar between individuals within organs. This suggests that incidental exposure to environmental mutagenic factors has minimal effect on the point-mutation landscapes in normal ASCs of the organs we assessed. Cell-intrinsic mutational processes, such as deamination-induced mutagenesis in rapidly cycling ASCs, seem to be more important determinants of point-mutation load. Indeed, many colorectal cancer mutations in the driver genes *APC*, *TP53*, *SMAD4* and *CTNNB1* are C:G to T:A transitions at CpG dinucleotides, whereas liver cancer driver mutations in the same genes have a completely different spectrum (Fig. 4a). However, ASCs of the colon and small intestine show very similar age-related mutation characteristics, although cancer incidence is extremely low in the human small intestine[1,9]. In addition to

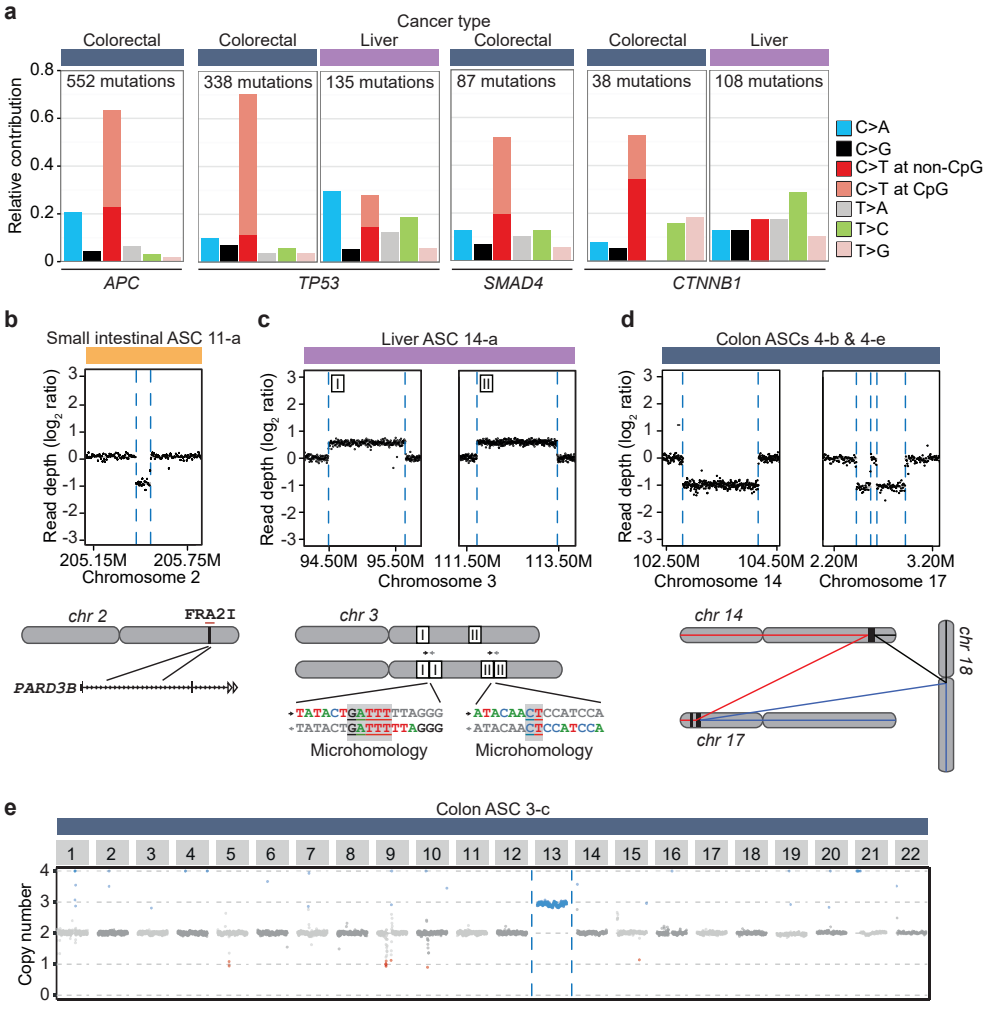


Figure 4. Cancer-associated mutation spectra in driver genes and structural variation in normal ASCs. (A) Spectrum of point mutations in cancer driver genes *APC*, *TP53*, *SMAD4* and *CTNNB1* identified in colorectal and liver cancer. The total number of somatic point mutations per gene per cancer type is indicated. (B) Read-depth analysis indicating a relatively small deletion (~90 kb) located within a common fragile site (*FRA2I*) in intestinal ASC 11-a. Each point represents the $\log_2$ value of the GC-corrected read-depth ratio per 5-kb window. Dashed lines indicate breakpoint regions; a schematic representation of the identified structural variant with associated genomic and breakpoint features is depicted below. (C) Two large (>1 Mb) tandem duplications identified in liver ASC 14-a with microhomology at the breakpoints; duplications are indicated in the schematic representation of the identified structural variants below the graph. (D) A complex structural variation (an unbalanced translocation involving 3 chromosomes) identified in colon ASCs 4-b and 4-e. Coloured lines in the schematic below show the predicted derivative chromosomes. (E) Read-depth analysis indicating a trisomy of chromosome 13 in colon ASC 3-c. Each data point represents the median chromosome copy number per 500-kb bin plotted over the genome, with alternating colours for each successive chromosome.

somatic point mutations, we evaluated the presence of somatic structural variants (Fig. 4b–e and Extended Data Table 2). We detected small deletions (91–443 kb) in 3 out of 14 small intestinal ASCs and a larger deletion (2 Mb) in one ASC. Notably, colon ASCs showed complex and larger chromosomal instability in 4 out of 15 colon ASCs, including a complex translocation (Fig. 4d) and a trisomy (Fig. 4e). These events are characteristic of segregation errors that can occur during cell division, and are a hallmark of many colorectal cancers[26]. In addition, other factors, such as tissue clonality or external agents may also contribute to the difference in cancer incidence between colon and small intestine.

Here we have shown that ASCs of organs with different cancer incidences gradually accumulate mutations at similar rates, but that the mutation profiles are tissue-specific. In the ASCs of the tissues assessed here, mutation accumulation is primarily driven by a combination of proliferation-dependent mutation incorporation following spontaneous deamination of methylated cytosine residues and another process with a currently unknown underlying molecular mechanism. Notably, the former intrinsic, unavoidable mutational process can cause the same types of mutation as those observed in cancer driver genes. We have shown that, at least in colon ASCs, this class of mutations could have a role in driving tumorigenesis.

ACKNOWLEDGEMENTS

The authors would like to thank the gastroenterologists of the UMCU/Wilhelmina Children's Hospital and Diakonessen Hospital for obtaining human duodenal and colon biopsies and R. Eijkemans for his advice on the statistical analyses. This study was financially supported by a Zenith grant of the Netherlands Genomics Initiative (935.12.003) to E.C., the NWO Zwaartekracht program Cancer Genomics.nl and funding of Worldwide Cancer Research (WCR no. 16-0193) to R.B. We declare no competing financial interests.

AUTHOR CONTRIBUTIONS

C.L.W., S.M. and E.E.S.N. obtained duodenal biopsies. N.S., M.M., E.E.S.N., M.M.A.V. and J.J. obtained colon biopsies. M.M.A.V., L.J.W.L., J.J. and J.N.M.I. obtained human liver biopsies. M.J., V.S., N.S., M.H., E.K., C.L.W., T.S., G.S. and R.B. performed ASC culturing. M.W. performed cell sorting. S.R., M.R.S., E.C. and R.B. performed sequencing. F.B., J.L., S.B., P.P., I.J.N., I.M. and R.B. performed bioinformatic analyses. F.B, R.G.V., H.C., E.C. and R.B. were involved in the conceptual design of the study. F.B., H.C., E.C. and R.B. wrote the manuscript.

METHODS

No sample-size estimate was calculated before the study was executed. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Human tissue material

Endoscopic, colorectal and duodenal biopsy samples were obtained from individuals of different ages that had been admitted for suspected inflammation. One individual (donor 1) showed no inflammation during colonoscopy, but was later diagnosed with microscopic colitis. The other individuals were found to be healthy based on standard histological examination. Endoscopic biopsies were performed at the University Medical Center Utrecht and the Wilhelmina Children's Hospital. The patients' informed consent was obtained and this study was approved by the ethical committee of University Medical Center

Utrecht. Additionally, normal tissue was isolated from resected colon segments at >5 cm distance from a tumour in three colorectal cancer patients (donors 3, 4 and 19). The colonic tissues were obtained at The Diakonessen Hospital Utrecht with informed consent and the study was approved by the ethical committee. Liver biopsies (0.5–1 cm^3) were obtained from donor livers during transplantations performed at the Erasmus Medical Center, Rotterdam. Both liver and colon biopsies were obtained from donor 18. The Medical Ethical Council of the Erasmus MC approved the use of this material for research purposes, and informed consent was provided by all donors and/or relatives.

Establishment of clonal ASC cultures

Dissociated colon and small intestinal crypts were isolated from the biopsies and cultured for 1 - 2 weeks under conditions that are optimal for stem-cell proliferation, as previously described[5]. Liver cells were isolated from human liver biopsies and cultured as previously described[3]. From these cultures, single cells were sorted by flow cytometry and clonally expanded (Extended Data Fig. 1a). Clonal ASC cultures were subsequently established by manual picking of individual organoids derived from single cells and *in vitro* expansion for a period of ~6 weeks.

Whole-genome sequencing and read alignment

DNA libraries for Illumina sequencing were generated using standard protocols (Illumina) from 200 ng - 1 µg of genomic DNA isolated from the clonally expanded ASC cultures with genomic tips (Qiagen). The libraries were sequenced with paired-end (2 × 100 bp) runs using Illumina HiSeq 2500 sequencers to a minimal depth of 30× base coverage. Samples of donors 1, 2, 3, 4, 10, 12, 13, 15, 16, 18 and 19 were sequenced using Illumina HiSeq X Ten sequencers to equal depth. The reference samples, blood or biopsy, were sequenced similarly. Sequence reads were mapped against human reference genome GRCh37 using Burrows–Wheeler Aligner v0.5.9 mapping tool[27] with settings 'bwa mem -c 100 –M'. Sequence reads were marked for duplicates using Sambamba v0.4.7 (ref. 28) and realigned per donor using Genome Analysis Toolkit (GATK) IndelRealigner v2.7.2 and sequence read-quality scores were recalibrated with GATK BaseRecalibrator v2.7.2. Alignments from different libraries of the same ASC culture were combined into a single BAM file.

Point mutation calling

Raw variants were multi-sample (per donor) called using the GATK UnifiedGenotyper v2.7.2 (ref. 29) and GATK-Queue v2.7.2 with default settings and additional option 'EMIT_ALL_CONFIDENT_SITES'. The quality of variant and reference positions was evaluated using GATK VariantFiltration v2.7.2 with options '–filterExpression "MQ0 ≥4 && ((MQ0 / (1.0 * DP)) > 0.1)"–filterName "HARD_TO_VALIDATE"–filterExpression "QUAL < 30.0 "–filterName "VeryLowQual"–filterExpression "QUAL > 30.0 && QUAL < 50.0 "–filterName "LowQual"–filterExpression "QD < 1.5 "–filterName "LowQD"'.

Point mutation filtering

To obtain high-quality catalogues of somatic point mutations, we applied a comprehensive filtering procedure (Extended Data Fig. 1b). We considered variants that were passed by VariantFiltration and had a GATK phred-scaled quality score ≥100. Subsequently, for each ASC culture, we considered the positions with a base coverage of at least 20× in both the culture and the reference sample (blood or biopsy). Furthermore, we only regarded variants at autosomal chromosomes. We excluded variant positions that overlapped with single-nucleotide polymorphisms (SNPs) in the SNP database (dbSNP) v137.b37 (ref. 30). Furthermore, we excluded all positions that were found to be variable in at least two of three unrelated individuals (that is, donor 5, 6 and X (not in study)) to exclude recurrent sequencing artefacts. To obtain somatic point mutations, we filtered out all variants with any evidence of the alternative allele in the reference sample. We validated the clonal origin of the sequenced ASC cultures by analysing the variant allele frequencies (VAFs) of the somatic mutations. Two cultures (donor 14, cell b and donor 17, cell c) showed a shift in the peak of the somatic heterozygous mutations to the left, indicating that they did not arise from a single stem cell, and were therefore excluded from the analysis (Extended Data Fig. 2). Finally, for all cultures we excluded point mutations with a VAF < 0.3 to exclude mutations that were

potentially induced *in vitro* after the (first) clonal step (Extended Data Fig. 1b–d). The number of mutations that passed each filtering step for the samples of donor 5 and 6 is depicted in Extended Data Fig. 1c. The overlap of the point mutations between ASCs of the same donor is depicted in Extended Data Fig. 4d.

Validations of point mutations

We evaluated our mutation filtering procedure by independent validations of 374 pre-selected positions that were either discarded or passed during filtering using amplicon-based next-generation sequencing. To this end, primers were designed ~250 nucleotides 5′ and 3′ from the candidate point mutations to obtain amplicons of ~500 bp (primer sequences available upon request). These regions were PCR-amplified for both the organoid cultures and reference samples of donor 5 and 6, using 5 ng genomic DNA, 1× PCR Gold Buffer (Life Technologies), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP and 1 unit of AmpliTaq Gold (Life Technologies) in a final volume of 10 µl. This which was held at 94 °C for 60 s followed by 15 cycles at 92 °C for 30 s, 65 °C for 30 s (with a decrement of 0.2 °C per cycle) and 72 °C for 60 s; followed by 30 cycles of 92 °C for 30 s, 58 °C for 30 s and 72 °C for 60 s; with a final extension at 72 °C for 180 s. The PCR products were pooled and barcoded per culture. Illumina sequence libraries were generated according to the manufacturer's protocol. Subsequently, the libraries were pooled and sequenced using the MiSeq platform (2 × 250 bp) to an average depth of ~100×. Alignment and variant-calling was performed as described above. For each ASC we evaluated those positions with at least 20× coverage for both culture and reference sample, and defined positive positions as those with a call in culture, with a VAF ≥ 0.3 and no call in the reference sample. Subsequently, we determined the number of confirmed negatives of the positions that were filtered out for each filter step (Extended Data Fig. 1d). Moreover, we determined the number of confirmed positive of the positions that passed all filters (Extended Data Fig. 1e, f).

Assessment of effects of *in vitro* culturing on ASC mutation load

We expanded 10 initial clonal organoid cultures from small intestine and liver for a further 3–5 months (equivalent to ~20 weekly passages), upon which we isolated single cells and subjected them to clonal expansion to obtain sufficient DNA for WGS (Extended Data Fig. 6a). This approach allowed us to catalogue the mutations that accumulated in single ASCs during the culturing period between the two clonal steps. To this end, we selected the somatic point mutations that were unique to the sub-clonal cultures and not present in the corresponding original clonal cultures and therefore acquired during the *in vitro* expansion. We evaluated the specificity of our mutation-discovery procedure by determining the confirmation rate of the mutations identified in the original clone in the corresponding subclone. Only positions that had a coverage of ≥20× in both the original clonal and corresponding subclonal culture as well as in the reference sample were evaluated. On average, 91.1% ± 4.87 (mean ± s.d). of these point mutations were confirmed in the subclonal cultures (Extended Data Fig. 3).

Correlation between ASC somatic point mutation accumulation and age

The surveyed area per ASC was calculated as the number of positions coverage ≥20× in both culture and the reference sample. The percentage of the whole non-N autosomal genome (GCRh37: 2,682,655,440 bp) that is surveyed in each ACS is depicted in Extended Data Table 1. For each ASC the total number of identified somatic point mutations was extrapolated to the whole non-N autosomal genome using its surveyed area. Subsequently, a linear mixed-effects regression model was fitted to estimate the effect of age on the number of somatic point mutations for each tissue using the nlme R package[31,32], in which 'donor' is modelled as a random effect to resolve the non-independence that results from having multiple measurements per donor. A two-tailed *t*-test was performed to test whether the slope is significantly different from zero (that is to say, whether the fixed age effect in the linear mixed model is statistically significant). The intercept of the regression lines with the *y* axis represents the somatic mutations present at birth (that have accumulated in the tissue lineage during prenatal development) plus the noise levels in the data and the mutations that have accumulated during the first week(s) of culturing proceeding the clonal step (see above). Since all cells were assessed in a similar manner, noise levels will be comparable and therefore will not bias the mutation rate (slope) estimates. The slope of the regression line was used to estimate the fixed age effect on somatic point mutation rate per tissue.

To exclude the possibility that differences in surveyed areas between ASCs bias our results, we performed the age correlation and spectrum analyses on a subset of mutations that are located in genomic regions that are surveyed (≥20×) in all samples in this study. This consensus surveyed area comprises 38.2% of the autosomal non-N genome and both the mutation rate and spectra were highly similar to those in Fig. 1c (Extended Data Fig. 4a–c), indicating that the differences in surveyed areas between the clones do not bias our conclusions.

Definition of genomic regions

To generate a conserved DNA replication timing profile for the human genome, we downloaded 16 Repli-seq data sets from the ENCODE project[33] at the University of California, SantaCruz (UCSC) genome browser[34] (GRCh37/hg19). The data consisted of Wavelet-smoothed values per 1-kb bin throughout the genome for 15 different cell lines (BJ, BG02ES, GM06990, GM12801, GM12812, GM12813, GM12878, HeLa-S3, HepG2, HUVEC, IMR90, K562, MCF-7, NHEK and SK-N-SH). We considered the median values of all cell lines per bin, thereby excluding cell-specific values. We arbitrarily divided the genome into early- (≥60), intermediate- (>33 & <60) and late- (≤33) replicating bins (Fig. 3b). To generate a conserved chromatin-association profile for the human genome, we downloaded data containing the H3K9me3 signal per 25-nucleotide bin throughout the genome for 22 different cell lines (A549, AG04450, DND41, GM12878, H1-hESC, HeLa-S3, HepG2, HMEC, HSMM, HSMMt, HUVEC, K562, monocytes-CD14+_RO1746, NH-A, NHDF-Ad, NHEK, NHLF, osteoblasts, MCF-7, NT2-D1, PBMC and U2OS) and the H3K27ac signal for 9 different cell lines (CD20+_RO01794, DND41, H1-hESC, HeLa-S3, HSMM, monocytes-CD14+_RO1746, NH-A, NHDF and osteoblasts). Data were downloaded from the ENCODE project[33] at the UCSC browser[34] (GRCh37/hg19) and the median values of all cell lines per bin were calculated. Next, we determined the distribution of the fractions of all bins (genome-wide). According to the shape of the resulting graph, we considered bins with an H3K9me3 value ≥4, or an H3K27ac value ≥2, as associated with that chromatin mark. Finally, exonic sequences were defined as all exonic regions reported in Ensembl v75 (GCRh37)[35].

Enrichment or depletion of point mutations in genomic regions

We determined whether somatic point mutations were enriched or depleted in the genomic regions described above. To this end, we determined how many point mutations were observed in each genomic region for each donor. Next, we calculated the number of bases that were surveyed in each genomic region and calculated the expected number of point mutations by multiplying this surveyed length with the genome-wide point-mutation frequency. The $\log_2$(observed/expected) of the mutations in the genomic regions was used as a measure of the effect size of the depletion or enrichment. One-tailed binomial tests were performed to calculate the statistical significance of deviations from the expected number of mutations in the genomic regions using pbinom[31]; $P < 0.05$ was considered significant.

Mutational signatures

The occurrences of all 96-trinucleotide changes were counted for each ASC and averaged per donor. Three mutational signatures were extracted using NMF[36]. To determine the replication bias of signatures, we determined whether the point mutations were located in an intermediate, early or late replicating region (as defined above) using GenomicRanges[37] and repeated the NMF on a 288 count matrix (96 trinucleotides × 3 replication timing regions). Similarly, we looked at transcriptional strand bias by performing NMF on a 192 count matrix (96 trinucleotides × 2 strands). To this end, we selected all point mutations that fall within gene bodies and checked whether the mutated C or T was located on the transcribed or non-transcribed strand. We defined the transcribed units of all protein coding genes based on Ensembl v75 (GCRh37)[35] and included introns and untranslated regions.

Selection analysis (d*N*/d*S*)

The d*N*/d*S* ratio was determined as described previously[18]. In brief, we used 192 rates, one for each of the possible trinucleotide changes in both strands. For each substitution type, we counted the number of potential synonymous and non-synonymous mutations in the protein-coding sequences of the human genome, using the longest DNA coding sequence as the reference sequence for each gene. Poisson

regression was used to obtain maximum-likelihood estimates and confidence intervals of the normalized ratio of non-synonymous versus synonymous mutations (d*N*/d*S* ratio). The d*N*/d*S* ratio was tested against neutrality (d*N*/d*S* = 1) using a likelihood-ratio test.

Comparison of mouse and human intestinal ASCs mutation loads

Intestinal ASCs were isolated from the proximal part of the small intestine of randomly chosen ~2-year-old mice (one male and one female) carrying the Lgr5-EGFP-Ires-CreERT2 allele (mice were C57BL/6 background) by sorting for GFP^{high} cells. Subsequently, three *Lgr5*-positive cells per animal were clonally expanded as described[4]. All experiments were approved by the Animal Care Committee of the Royal Dutch Academy of Sciences according to the Dutch legal ethical guidelines. DNA isolated from the intestinal ASC cultures isolated from mouse 1 were sequenced with paired-end (75 and 35 bp) runs using SOLiD 5500 sequencers (Life Technologies) to an average depth of ~18× base coverage. Intestinal ASC cultures of mouse 2 were sequenced using Illumina HiSeq 2500 sequencers as described above. Sequence reads were aligned using Burrows–Wheeler Aligner to the mouse reference genome (NCBIM37) and point mutations were called using the GATK UnifiedGenotyper v2.7.2 as described above. Post-processing filters for the intestinal ASCs of mouse 1 (analysed by SOLiD sequencing) were as follows: a minimum depth of 10×, variant uniquely called in one intestinal stem cell without more than one alternative allele found at the same position in the other ASCs of the same mouse, a GATK a phred-scaled quality score ≥100, variant absent in mouse 2, variant position absent in the dbSNP (build 128) and a VAF ≥ 0.25. Post-processing filters for the intestinal ASCs of mouse 2 (analysed by Illumina sequencing) were as described above for the human mutation data.

Cancer-associated mutation spectra analysis in driver genes

Mutations identified in the indicated genes in colorectal or liver cancers were downloaded from cBioPortal (http://www.cbioportal.org/). Only point mutations that resulted in a missense, nonsense or splice-site mutation were considered.

CNV detection

To detect copy-number variations (CNVs), BAM files were analysed for read-depth variations by CNVnator v0.2.7 (ref. 39) with a bin size of 1 kb and Control-FREEC v6.7[40] with a bin size of 5 kb. Highly variable regions, defined as harbouring germline CNVs in at least three control samples, were excluded from the analysis. To obtain somatic CNVs, we excluded CNVs for which there was evidence in the reference sample (blood/biopsy) of the same individual. Resulting candidate CNV regions were assessed for additional structural variants on the paired-end and split-read level through DELLY v0.3.3 (ref. 41). Based on these results, we excluded five candidate CNV regions as mapping artefacts on the read-depth level and acquired base-pair accuracy of the involved breakpoints for the other events. This also revealed the tandem orientation of the duplication events and the complex structural variation in the colon sample.

Reported gene definitions (Extended Data Table 2) are based on Ensembl v75 (GCRh37)[35]. Common fragile sites overlapping the events were detected using existing definitions[42]. LINE/SINE elements within 100 bp of the breakpoints were determined with the repeat element annotation[43] from the UCSC genome browser[34] GCRh37 (retrieved 26 October 2015).

DATA ACCESS

The human sequencing data have been deposited at the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/) under accession numbers EGAS00001001682 and EGAS00001000881. The mouse sequencing data have been deposited at the European Nucleotide Archive (http://www.ebi.ac.uk/ena/) under accession number ERP005717.

CODE AVAILABILITY

All code and filtered vcf files are freely available under a MIT License at https://wgs11.op.umcutrecht.nl/mutational_patterns_ASCs/ and https://github.com/CuppenResearch/MutationalPatterns/.

COMPETING INTERESTS

The authors declare no competing financial interests.

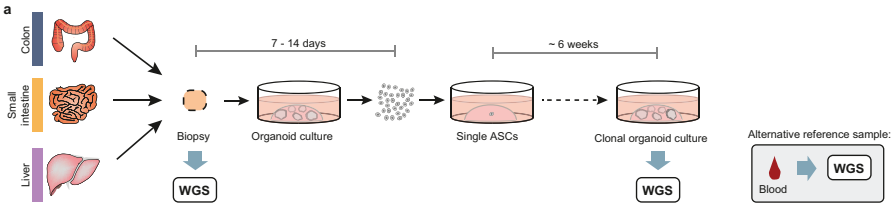
REFERENCES

1. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015)
2. Rossi, D. J., Jamieson, C. H. M. & Weissman, I. L. Stems cells and the pathways to aging and cancer. *Cell* **132**, 681–696 (2008)
3. Huch, M. *et al.* Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**, 299–312 (2015)
4. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009)
5. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011)
6. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009)
7. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009)
8. Milholland, B., Auton, A., Suh, Y. & Vijg, J. Age-related somatic mutations in the cancer genome. *Oncotarget* **6**, 24627–24635 (2015)
9. Wu, S., Powers, S., Zhu, W. & Hannun, Y. A. Substantial contribution of extrinsic risk factors to cancer development. *Nature* **529**, 43–47 (2016)
10. Hou, Y. *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148**, 873–885 (2012)
11. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013)
12. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Reports* **3**, 246–259 (2013)
13. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015)
14. Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, 81–84 (2015)
15. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012)
16. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010)
17. Finette, B. A. *et al.* Determination of *HPRT* mutant frequencies in T-lymphocytes from a healthy pediatric population: statistical comparison between newborn, children and adult mutant frequencies, cloning efficiency and age. *Mutat. Res.* **308**, 223–231 (1994)
18. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015)
19. Xie, M. *et al.* Age-related cancer mutations associated with clonal hematopoietic expansion. *Nat. Med.* **20**, 1472–1478 (2014)
20. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014)
21. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014)
22. Pleasance, E. D. *et al.* A comprehensive

catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010)

23. Dollé, M. E. T., Snyder, W. K., Dunson, D. B. & Vijg, J. Mutational fingerprints of aging. *Nucleic Acids Res.* **30**, 545–549 (2002)
24. Dollé, M. E., Snyder, W. K., Gossen, J. A., Lohman, P. H. & Vijg, J. Distinct spectra of somatic mutations accumulated with age in mouse heart and small intestine. *Proc. Natl Acad. Sci. USA* **97**, 8403–8408 (2000)
25. Behjati, S. *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 (2014)
26. Fearon, E. R. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.* **6**, 479–507 (2011)
27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009)
28. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015)
29. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011)
30. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001)
31. R Core Team. *R: A language and environment for statistical computing* ; http://www.r-project.org/ (2015)
32. Pinheiro J *et al. nlme: Linear and Nonlinear Mixed Effects Models.* https://cran.r-project.org/web/packages/nlme/nlme.pdf (2016)
33. ENCODE Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2013)
34. Rosenbloom, K. R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670–D681 (2015)
35. Cunningham, F. *et al.* Ensembl 2015. *Nucleic Acids Res* . **43**, D662–D669 (2015)
36. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010)
37. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013)
38. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* . **21**, 974–984 (2011)
39. Boeva, V. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012)
40. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012)
41. Le Tallec, B. *et al.* Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Reports* **4**, 420–428 (2013)
42. Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418–420 (2000)

SUPPLEMENTAL FIGURES AND TABLES



b

FILTER STEPS

1. Multi-sample called
2. Surveyable loci
3. No common SNPs
4. No recurring positions
5. No evidence reference
6. VAF ≥0.3 & Call

c

MUTATIONS PASSING FILTERS

	ASCs 5-a	5-b	5-c	6-a	6-b	6-c
1. Multi-sample called	4,494,378	4,494,378	4,494,378	4,563,975	4,563,975	4,563,975
2. Surveyable loci	3,232,027	3,101,003	3,211,703	3,646,190	3,465,866	3,594,650
3. No common SNPs	76,899	74,560	76,336	92,212	85,268	89,451
4. No recurring positions	33,515	31,976	33,156	40,834	37,272	39,520
5. No evidence reference	976	915	970	2,011	1,868	1,972
6. VAF ≥0.3 & Call	246	245	304	591	506	614

d

VALIDATIONS OF MUTATIONS FILTERED OUT

	ASCs 5-a	5-b	5-c	6-a	6-b	6-c	Total	Unique positions
Incorrectly filtered	0	1*	0	0	0	0	1*	1*
Correctly filtered	17	20	18	22	24	22	123	41
Incorrectly filtered	0	0	0	1	1	1	3	1
Correctly filtered	14	16	14	14	13	14	85	25
Incorrectly filtered	0	1	0	0	0	0	1	1
Correctly filtered	14	13	16	15	16	17	91	22
Incorrectly filtered	0	0	0	0	0	0	0	0
Correctly filtered	0	0	0	0	0	0	0	0
Incorrectly filtered	0	0	1	0	0	0	1	1
Correctly filtered	30	65	58	35	62	56	306	156

e

VALIDATIONS OF MUTATIONS PASSING ALL FILTERS

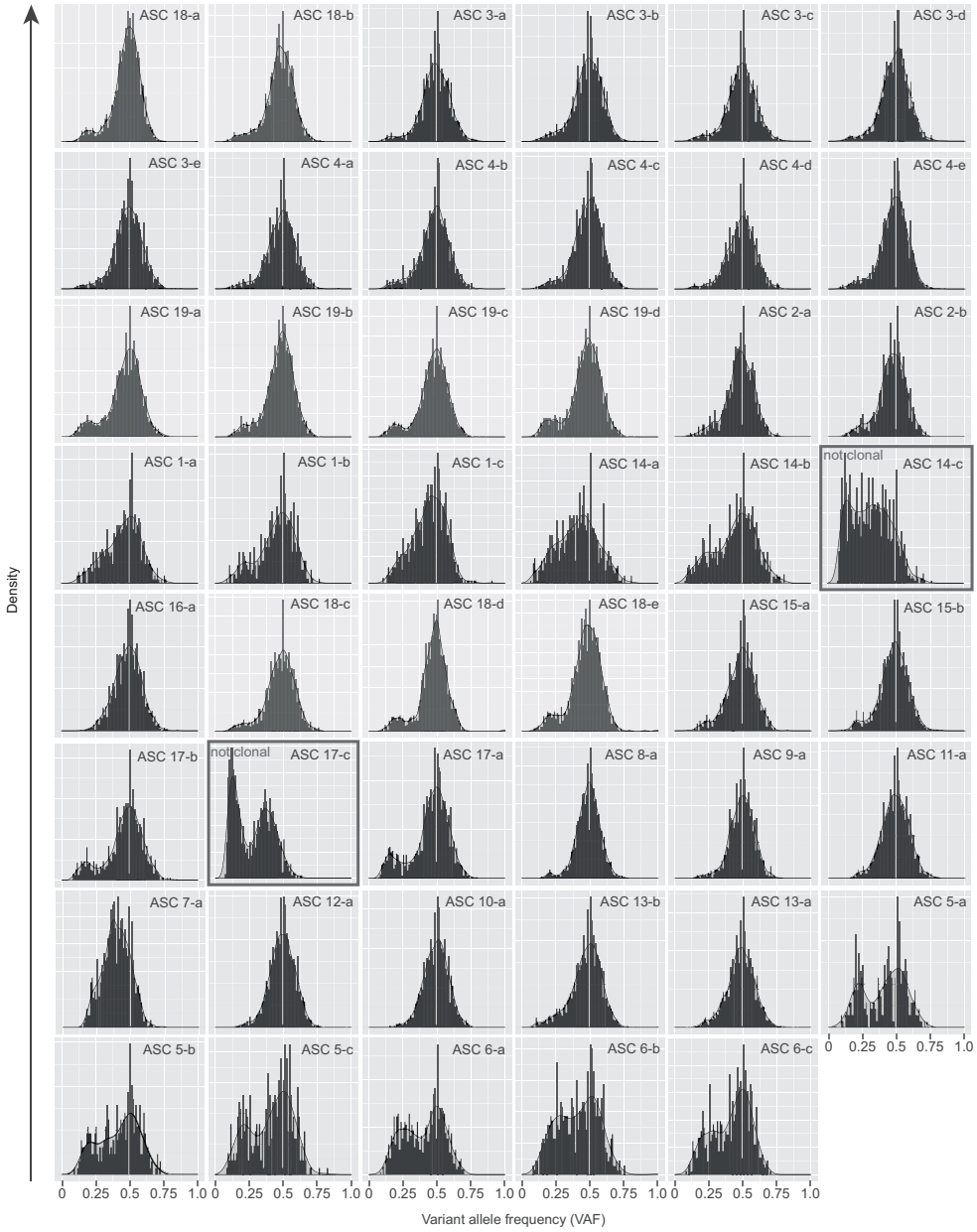
	ASCs 5-a	5-b	5-c	6-a	6-b	6-c	Total
Total positions tested	43	15	19	38	14	22	151
Confirmed mutations	37	11	18	34	10	21	131
Not confirmed	6	4	1	4	4	1	20

f

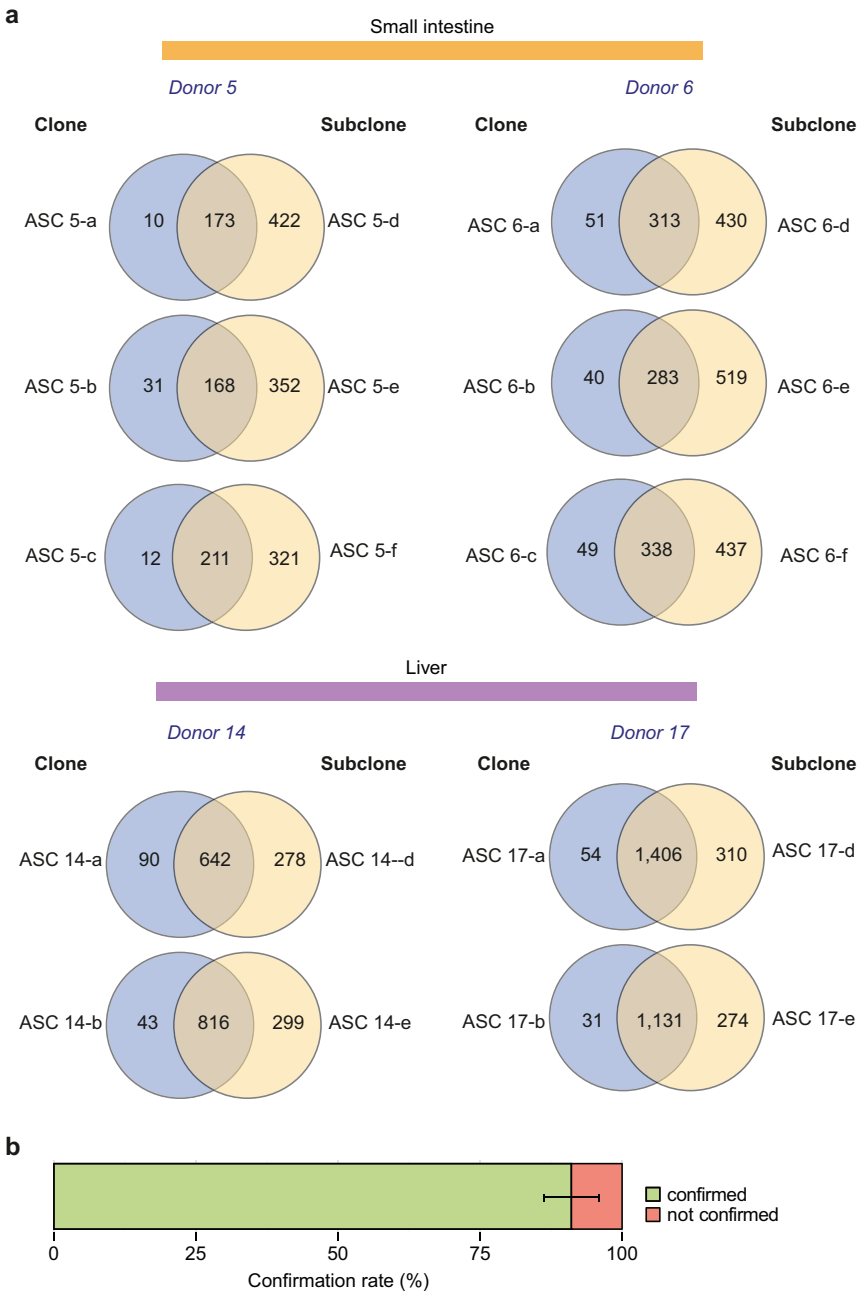
MANUAL INSPECTION OF NOT CONFIRMED POSITIONS

	ASCs 5-a	5-b	5-c	6-a	6-b	6-c	Total
True positive	1	1	1	1	2	0	6
False positive	3	3	0	2	1	1	10
Missed in germline	2	0	0	1	1	0	4

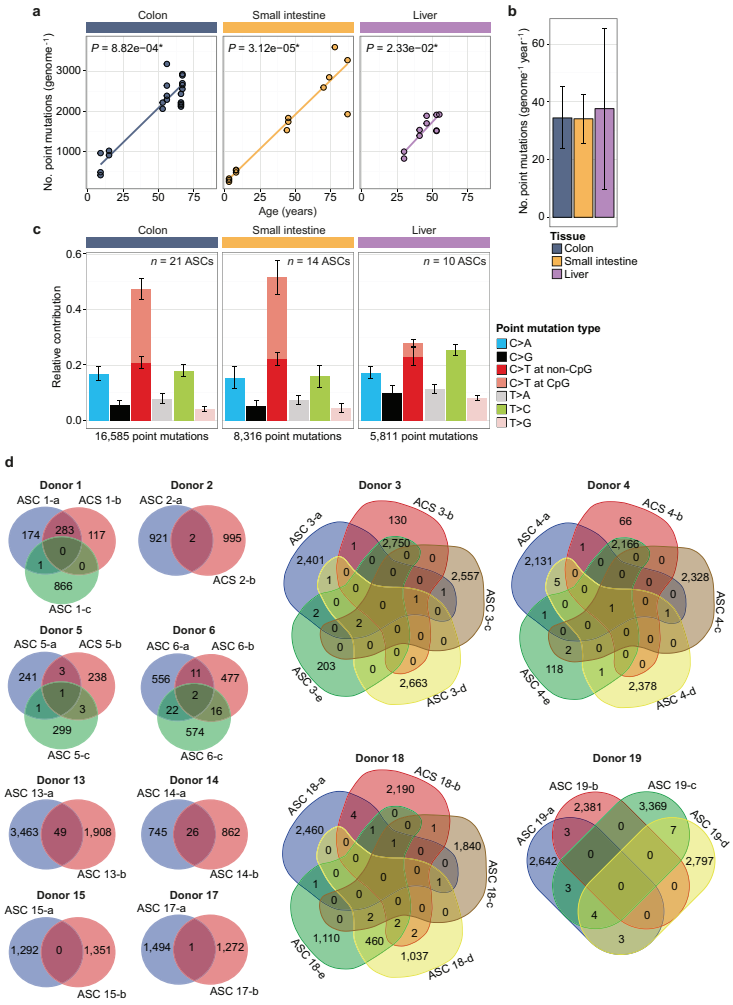
Supplemental figure S1. Cataloguing somatic mutation loads in human ASCs. (A) Schematic overview of the experimental setup to determine somatic mutations in individual human ASCs. Colon, small intestine and liver biopsies were cultured in bulk for 1-2 week(s) before single cells were sorted and clonally expanded until enough DNA could be isolated for WGS analysis. WGS of the clonal organoid culture allows for cataloguing of somatic variants in the original ASCs that gave rise to the clonal cultures that were acquired during life and the first 7–14 days of culturing. Biopsy or blood was sequenced as a reference sample. (B) Filter steps to obtain somatic mutations in ASCs. (C) Number of point mutations that pass each corresponding filter step in (A) for each ASC culture of donors 5 and 6. (D) Independent validations of mutations that were filtered out by amplicon-based resequencing. The asterisk indicates a position that is not located in the surveyed areas of the assessed ASCs in the original experiment, which is corrected for in all analyses. (E) Independent validations of mutations that passed all filters by amplicon-based re-sequencing. Confirmed positions are defined as those with a call in the indicated ASC with a VAF ≥ 0.3 and without a call in the corresponding reference sample. (F) Qualification of unconfirmed positions based on manual inspection. True-positive positions are positions that were correctly called, but for which the VAF threshold was not met in the validation experiment. False-positive positions are positions without evidence in the validation experiment or are noisy. 'Missed in germline' are positions that were called in the reference sample in the validation experiment.



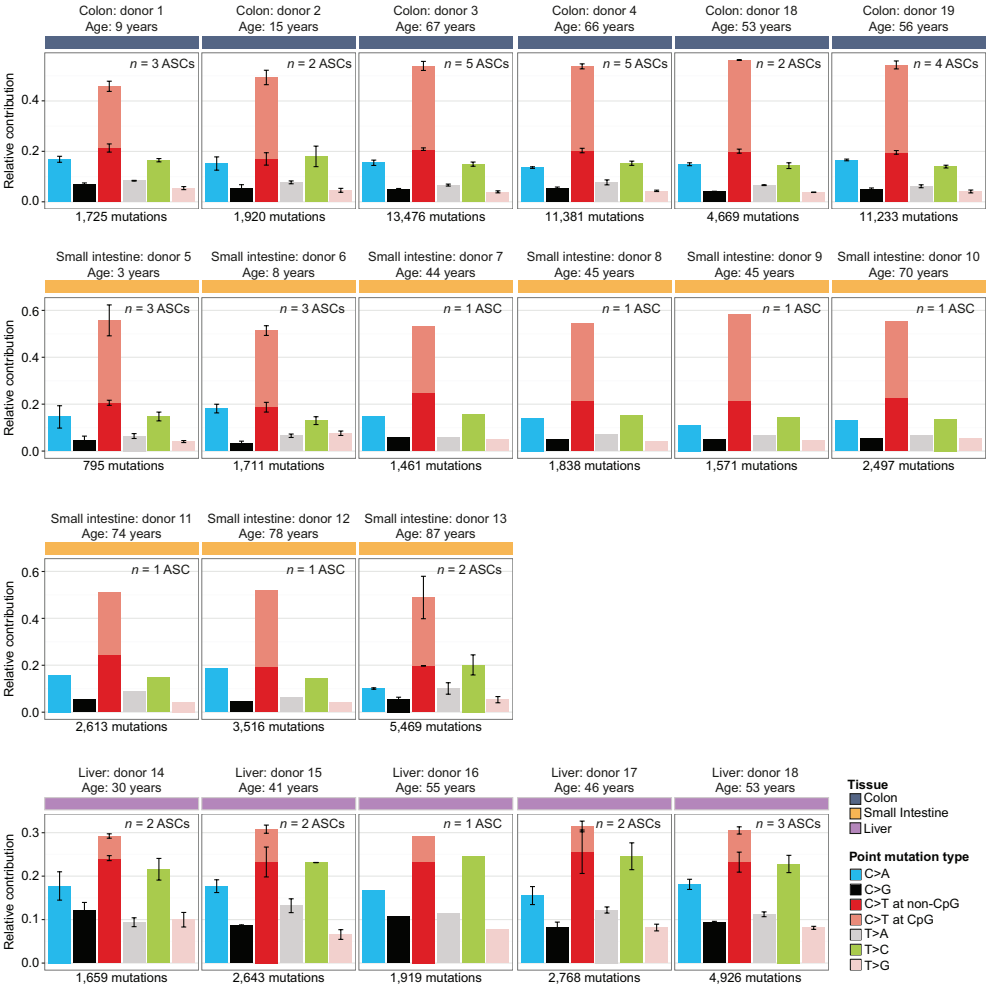
Supplemental figure S2. Variant allele frequency distribution plot for each assessed ASC. A distribution plot of the VAFs of all somatic mutations that remain before filtering for the VAF in filter step 6 (Extended Data Fig. 1b). Clonal heterozygous somatic mutations form a peak around VAF = 0.5. A threshold of VAF ≥ 0.3 was used to obtain somatic mutations that were clonal in the organoid cultures and therefore present in the original cloned ASCs (see Methods). Mutations acquired after the single ASC expansion step are subclonal (that is, not present in all cells of the clonal culture) and have lower VAFs. Two samples (donor 14, ASC 14-b and donor 17, ASC 17-c) showed a shift in the main VAF peak to the left, indicating that these cultures did not arise from a single ASC and were therefore excluded from the study.



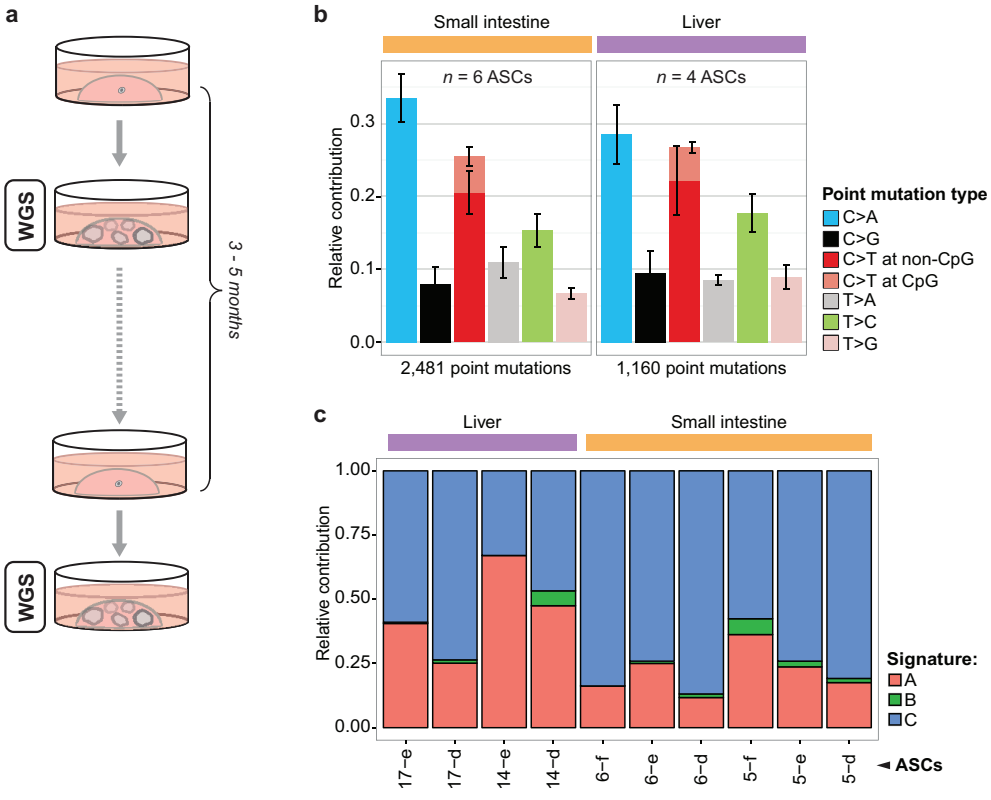
Supplemental figure S3. Confirmation rate of somatic point mutations. (A) Overlap of somatic point mutations between the clonal organoid cultures and corresponding subcloned cultures depicted in Extended Data Fig. 6. (B) Confirmation rate of point mutations, which were observed in the original cloned culture, in the corresponding subcloned culture. Data are represented as the mean percentage of confirmed point mutations over all clone–subclone pairs indicated in (A) ($n = 10$) and error bars represent s.d.



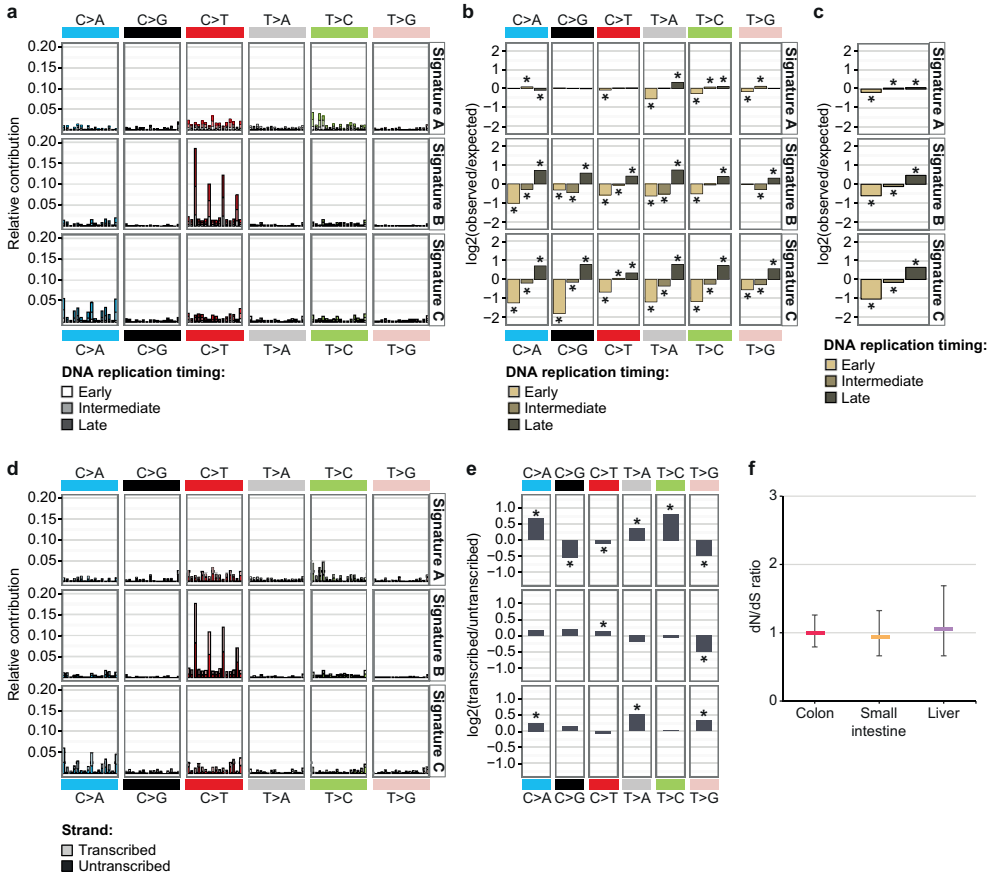
Supplemental figure S4. Somatic mutation loads in consensus-surveyed area and overlap of point mutations between ASCs from the same donor. (A) Correlation of the number of somatic point mutations per ASC, which were observed in the genomic regions that were surveyed (for example, a base coverage of at least 20× in both the clonal culture and the reference sample; Methods) in all the ASCs, with the age of the donors per tissue indicated. This consensus-surveyed area comprises 38.2% of the non-N autosomal genome. Each data point represents a single ASC. Indicated are the *P* values of the age effects in the linear mixed model (two-tailed *t*-test) for each tissue. The sample sizes for colon, small intestine and liver are 6, 9 and 5 donors and 21, 14 and 10 ASCs, respectively. (B) Somatic mutation accumulation rate per tissue as estimated by the linear mixed models in (A). Error bars represent the 95% confidence intervals of the slope estimates. (C) Relative contribution of the indicated mutation types to the point mutation spectra in the consensus-surveyed area per tissue type. Data are represented as the mean relative contribution of each mutation type over all ASCs per tissue type ($n = 21$, 14 and 10 for colon, small intestine and liver, respectively); error bars represent s.d. The total number of identified somatic point mutations per tissue is shown. (D) Overlap of the somatic point mutations between ASCs of the same donor. The number of point mutations, observed in the total surveyed area per ASC, that are shared between the assessed ASCs of the same donor is indicated.



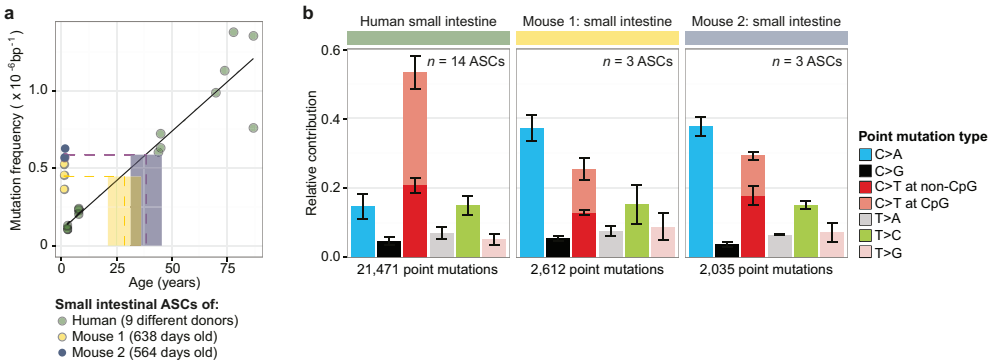
Supplemental figure S5. Point-mutation spectrum per donor. Relative contribution of the different types of point mutation to the spectrum of each donor. Data are represented as the mean relative contribution of each mutation type when multiple ASCs were measured per donor (the number *n* of ASC per donor is depicted for each donor) and error bars represent standard deviation. Indicated are the age of the donors, the total number of point mutations used to determine each spectrum and the tissue type.



Supplemental figure S6. Mutation patterns associated with long-term *in vitro* expansion of ASCs. (A) Schematic overview of the experimental setup to catalogue mutations associated with the organoid culture system. Clonal small intestinal and liver organoid cultures (Extended Data Fig. 1a) were cultured for 3–5 months. A second clonal expansion step was subsequently performed, followed by WGS analysis, to catalogue all the mutations that were present in the subcloned ASCs. To obtain mutations that were specifically acquired during culturing, mutations in the original clonal cultures were subtracted from those observed in the corresponding second subcloned cultures. (B) Relative contribution of the indicated mutation types to the point mutation spectra specifically observed *in vitro* per tissue type. Data are represented as the mean relative contribution of each mutation type over all subcloned ASCs per tissue type ($n = 6$ and 4 for small intestine and liver, respectively) and error bars represent s.d. Indicated are the total number of identified somatic point mutations, which were specifically acquired between the two clonal expansion steps indicated in (A), per tissue. (C), Relative contribution of the mutational signatures depicted in Fig. 2a, which explain the mutation spectra depicted in (B).



Supplemental figure S7. Non-random distribution of mutational signatures throughout the genome. (A) Context- and replication-timing-dependent mutation spectrum of the three mutational signatures depicted in Fig. 2a. Indicated is the contribution of each trinucleotide to the signatures (order is similar as in ref. 11), subdivided into the fraction of the trinucleotide-change present in early, intermediate or late replicating genomic regions. (B) $\log_2$ ratio of the observed and expected number of mutations per indicated base substitution (summed over all trinucleotides) in early-, intermediate- and late-replicating genomic regions for each of the signatures depicted in (A). $\log_2$ ratio indicates the effect size of the bias and asterisks indicate significant DNA-replication-timing bias ($P < 0.05$, binomial test). c, $\log_2$ ratio of the total number of observed and expected mutations in early-, intermediate- and late-replicating genomic regions for each signature depicted in (A). $\log_2$ ratio indicates the effect-size of the bias and asterisks indicate significant DNA replication timing bias ($P < 0.05$, binomial test). (D) Context- and transcriptional-strand-dependent mutation spectrum of the three mutational signatures depicted in Fig. 2a. Indicated is the contribution of each trinucleotide to the signatures (order is similar to that in ref. 11), subdivided into the fraction of the trinucleotide-change present on the transcribed and untranscribed strand. (E) $\log_2$ ratio of the number of mutations on the transcribed and untranscribed strand per indicated base substitution for each signature depicted in (D). $\log_2$ ratio indicates the effect size of the bias and asterisks indicate significant transcriptional strand bias ($P < 0.05$, binomial test). (F) The d*N*/d*S* ratio for all protein-coding somatic point mutations observed in all ASCs per tissue type. Error bars indicate 95% confidence intervals (likelihood ratio test).



Supplemental figure S8. Comparison of mutation loads between intestinal ASCs derived from human and mouse. (A) Mutation frequency in mouse intestinal ASCs is compared to the linear fit, describing the relationship between the mutation frequency in human intestinal ASCs and age of the donor. Indicated by the dotted lines are the mean mutation frequencies over all ASCs per mouse ($n = 3$) and the corresponding age of human linear fit. (B) Relative contribution of the indicated mutation types to the point mutation spectra for all assessed human intestinal ASCs and for each mouse. Data are represented as the mean relative contribution of each mutation type over all the ASCs per indicated category ($n = 14$, 3 and 3 for human, mouse 1 and mouse 2, respectively), error bars indicate s.d.

Extended Data Table 1 | Overview of somatic point mutations detected in ASCs

ASC	Donor	Age (years)	Gender	Tissue	Surveyed genome (%)*	No. point mutations†
1-a	1	9	Female	Colon	93.8	458
1-b	1	9	Female	Colon	91.2	400
1-c	1	9	Female	Colon	97.6	867
2-a	2	15	Male	Colon	96.8	923
2-b	2	15	Male	Colon	96.8	997
18-a	18	53	Male	Colon	98.5	2,468
18-b	18	53	Male	Colon	98.1	2,201
19-a	19	56	Male	Colon	97.7	2,655
19-b	19	56	Male	Colon	97.1	2,384
19-c	19	56	Male	Colon	98.2	3,383
19-d	19	56	Male	Colon	97.8	2,811
4-a	4	66	Female	Colon	90.7	2,140
4-b	4	66	Female	Colon	95.3	2,234
4-c	4	66	Female	Colon	95.7	2,332
4-d	4	66	Female	Colon	93.9	2,386
4-e	4	66	Female	Colon	96.1	2,289
3-a	3	67	Female	Colon	91.8	2,409
3-b	3	67	Female	Colon	91.8	2,882
3-c	3	67	Female	Colon	91.9	2,561
3-d	3	67	Female	Colon	92.0	2,667
3-e	3	67	Female	Colon	92.0	2,957
5-a	5	3	Female	Small intestine	89.0	246
5-b	5	3	Female	Small intestine	85.5	245
5-c	5	3	Female	Small intestine	88.5	304
6-a	6	8	Female	Small intestine	97.1	591
6-b	6	8	Female	Small intestine	93.5	506
6-c	6	8	Female	Small intestine	96.2	614
7-a	7	44	Male	Small intestine	91.1	1,461
8-a	8	45	Male	Small intestine	95.5	1,838
9-a	9	45	Male	Small intestine	93.9	1,571
10-a	10	70	Female	Small intestine	94.8	2,497
11-a	11	74	Male	Small intestine	87.3	2,613
12-a	12	78	Female	Small intestine	95.6	3,516
13-a	13	87	Male	Small intestine	97.7	3,512
13-b	13	87	Male	Small intestine	97.0	1,957
14-a	14	30	Male	Liver	81.3	771
14-b	14	30	Male	Liver	85.2	888
15-a	15	41	Female	Liver	93.5	1,292
15-b	15	41	Female	Liver	95.1	1,351
17-a	17	46	Female	Liver	79.4	1,495
17-b	17	46	Female	Liver	73.7	1,273
18-c	18	53	Male	Liver	97.9	1,845
18-d	18	53	Male	Liver	98.5	1,504
18-e	18	53	Male	Liver	98.2	1,577
16-a	16	55	Male	Liver	97.5	1,919

* Percentage of the non-N autosomal genome with >= 20x coverage in both ASC culture and reference sample.
† Number of somatic point mutations detected within surveyed genome.

Supplemental table S1. Overview of somatic point mutations detected in ASCs

Extended Data Table 2 | Identified somatic structural variations in ASCs

Copy Number Variants

Sample	Tissue	Chr	Start	Stop	Size	Type	No. genes	Fragile site	Microhomology	Genes at breakpoint	LINE/SINE
ASC 14-a	Liver	3	94,491,729	95,651,811	1,160,082	gain	5	-	5 bp	-|-	L1MC1|-
ASC 14-a	Liver	3	111,726,406	113,471,637	1,745,231	gain	46	-	2 bp	TAGLN3|ATP6V1A	L1MC1|L1M5
ASC 16-a	Liver	9	50,763,759	141,213,431	90,449,672	gain	1,472	-	NA	NA	NA
ASC 18-e	Liver	7	132,751,706	133,009,202	257,496	gain	2	-	0 bp	CHCHD3|EXOC4	MIR|L1PA6
ASC 18-d	Liver	5	59,125,105	59,718,364	593,259	loss	1	-	0 bp	PDE4D|PDE4D	-|L1PA6
ASC 8-a	Small intestine	5	3,815,936	3,908,819	92,883	loss	0	-	2 bp	-|-	-|-
ASC 11-a	Small intestine	2	205,420,067	205,511,877	91,810	loss	1	FRA2I	1 bp	PARD3B|PARD3B	AluSx|L1ME3B
ASC 13-a	Small intestine	11	63,974,352	66,222,668	2,248,316	loss	163	-	3 bp	FERMT3|-	-|L1M4b
ASC 13-b	Small intestine	1	5,878,566	6,321,750	443,184	loss	13	FRA1A	1 bp	-|-	THE1B|-
ASC 1-c	Colon	3	60,700,662	61,199,328	498,666	loss	4	FRA3B	1 bp	FHIT|FHIT	L1PA3|L1PA3
ASC 3-c	Colon	13	0	115,169,878	115,169,878	gain	1,217	-	NA	NA	NA
ASC 4-b&e	Colon	14	102,805,595	104,172,376	1,366,781	loss	57	-	NA	NA	NA
ASC 4-b&e	Colon	17	2,429,169	2,572,747	143,578	loss	5	-	CTTG ins	-|PAFAH1B1	AluJo|AluSq
ASC 4-b&e	Colon	17	2,634,433	2,927,007	292,574	loss	4	-	NA	NA	NA

Unbalanced Translocations

Sample	Tissue	Chr (1)	Position (1)	Chr (2)	Position (2)	Type	No. genes	Fragile site	Microhomology	Genes at breakpoint	LINE/SINE
ASC 4-b&e	Colon	14	102,805,595	17	2,634,145	translocation	NA	-	4 bp	ZNF839|-	-|-
ASC 4-b&e	Colon	14	104,172,376	18	18,518,987	translocation	NA	-	0 bp	XRCC3|-	-|ALR Alpha
ASC 4-b&e	Colon	17	2,927,007	18	18,518,987	translocation	NA	-	0 bp	RAF1GAP2|-	L1PA8|ALR Alpha

No. genes, number of genes overlapping the event; fragile site, common fragile sites overlapping the event; microhomology, number of bases of microhomology observed at breakpoints; genes at breakpoint, gene bodies affected by the breakpoint; LINE/SINE elements, observed elements within 100 bp of the breakpoint.

Supplemental table S2. Identified somatic structural variations in ASCs

A villain within

Chapter 4

Deficiency of nucleotide excision repair explains mutational signature observed in cancer

Myrthe Jager[1,#], Francis Blokzijl[1,#], Ewart Kuijk[1], Maria Vougioukalaki[2], Roel Janssen[1], Nicolle Besselink[1], Sander Boymans[1], Joep de Ligt[1], Jan Hoeijmakers[2,3], Joris Pothof[2], Ruben van Boxtel[1,3,†] and Edwin Cuppen[1,†]

1 Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht University, Universiteitsweg 100, 3584, CG, Utrecht, The Netherlands
2 Erasmus Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands
3 Princess Máxima Center for Pediatric Oncology, 3584 CT Utrecht, The Netherlands
# These authors contributed equally to this work
† These authors contributed equally to this work

Manuscript in preparation

ABSTRACT

Nucleotide excision repair (NER) is one of the main DNA repair pathways that protect cells against genomic damage. Disruption of this pathway can contribute to the development of cancer and accelerate aging. Tumors deficient in NER are more sensitive to cisplatin treatment. Characterization of the mutational consequences of NER-deficiency may therefore provide important diagnostic opportunities. Here, we analyzed the somatic mutational profiles of adult stem cells (ASCs) from NER-deficient *Ercc1*$^{-/\Delta}$ mice, using whole-genome sequencing analysis of clonally derived organoid cultures. Our results indicate that NER-deficiency increases the base substitution load in liver, but not in small intestinal ASCs, which coincides with a tissue-specific aging-pathology observed in these mice. The mutational landscape changes as a result of NER-deficiency in ASCs of both tissues and shows an increased contribution of Signature 8 mutations, which is a pattern with unknown etiology that is recurrently observed in human cancers. The scattered genomic distribution of the acquired base substitutions indicates that deficiency of global-genome NER (GG-NER) is responsible for the altered mutational landscape. In line with this, we also observed increased Signature 8 mutations in a GG-NER-deficient human organoid culture in which *XPC* was deleted using CRISPR-Cas9 gene-editing. Elevated levels of Signature 8 mutations may therefore serve as a novel biomarker for NER-deficiency and could improve personalized cancer treatment strategies.

INTRODUCTION

The genome is continuously exposed to mutagenic processes, which can damage the DNA and can ultimately result in the accumulation of mutations. To counteract these processes, cells exploit multiple DNA repair pathways that each repair specific lesions. Deficiency of these pathways can contribute to cancer initiation and progression. To increase insight into the cellular processes that underlie mutation accumulation, such as DNA repair deficiency, genome-wide mutational patterns of tumors can be characterized (Alexandrov et al. 2013; Nik-Zainal et al. 2016). To date, systematic analyses of tumor genomes have revealed 30 signatures of base substitutions and 6 rearrangement signatures of mutational processes in cancer genomes (Alexandrov et al. 2013; Nik-Zainal et al. 2016). These mutational signatures may have important diagnostic value. For example, several signatures are associated with BRCA1/2 inactivity and can consequently be predictive for a response to PARP inhibition or cisplatin treatment (Waddell et al. 2015; Davies et al. 2017).

Although for some signatures the underlying molecular process (Kim et al. 2016; Alexandrov et al. 2013, 2016) or involved DNA repair pathway (Kim et al. 2016; Davies et al. 2017; Alexandrov et al. 2013) is known, in-depth mechanistic

insight is still lacking for the majority of the mutational signatures (Petljak and Alexandrov 2016). Efforts to link mutational processes to specific signatures have mainly focused on associating mutation data from tumors to mutagen exposure and DNA repair-deficiency. Yet, tumors are genomically highly unstable and typically multiple processes have contributed to mutation accumulation (Alexandrov et al. 2013; Nik-Zainal et al. 2016), which hampers the identification of the processes that cause specific mutational signatures. We recently developed an approach for measuring mutations in non-cancerous adult stem cells (ASCs), by combining organoid culturing technology with whole-genome sequencing (WGS) (Jager et al. 2018; Drost et al. 2017). This method can be used to determine the mutations that have accumulated during life and during culturing. Tissue-specific ASCs maintain a highly stable genome both *in vivo* and *in vitro*, and therefore provide a stable system to study mutational processes in detail (Blokzijl et al. 2016; Huch et al. 2015). Furthermore, ASCs constitute a relevant cell source to study mutational patterns, as these cells are believed to be the cell-of-origin for specific types of cancer (Barker et al. 2009; Zhu et al. 2016; Adams et al. 2015).

Using this technique, we set out to determine the mutational consequences of deficiency of Nucleotide excision repair (NER). NER is one of the main cellular DNA repair pathways (Iyama and Wilson 2013), and consists of two subpathways: global-genome NER (GG-NER), which repairs bulky helix-distorting lesions throughout the genome, and transcription-coupled NER (TC-NER), which resolves RNA polymerase blocking lesions during transcription (Iyama and Wilson 2013; Marteijn et al. 2014; Hoeijmakers 2009). Somatic mutations in *ERCC2*, a key factor of NER, were previously associated with Signature 5 in urothelial tumors (Kim et al. 2016). However, NER has been suggested to underlie multiple mutational signatures, based on large-scale tumor mutation analyses (Alexandrov et al. 2013), and not all NER-deficient tumors are characterized by a high Signature 5 contribution (Kim et al. 2016). This suggests that NER-deficiency might be associated with other mutational signatures as well.

To characterize the mutational consequences of NER-deficiency, we studied mutagenesis in *Ercc1*$^{-/\Delta}$ mice and *XPC*-knockout (*XPC*^{KO}) organoids. ERCC1 plays a crucial role in the core NER pathway involving both GG-NER and TC-NER (Kirschner and Melton 2010; Iyama and Wilson 2013; Sijbers et al. 1996a; Aboussekhra et al. 1995), in crosslink repair (Rahn et al. 2010), and in single strand annealing (SSA) of double strand breaks (Al-Minawi et al. 2008). *ERCC1* is mutated in ~4.5% of all human tumors, especially skin and liver cancer (http://dcc.icgc.org), and single nucleotide polymorphisms in *ERCC1* have been linked to an increased risk of developing colorectal cancer (Ni et al. 2014). *Ercc1*$^{-/\Delta}$ mice are hemizygous for a single truncated *Ercc1* allele, which largely corrupts protein function (Dollé et al. 2011; Weeda et al.

1997) and results in decreased NER-activity (Su et al. 2012). *Ercc1*$^{-/\Delta}$ mice have a reduced lifespan as a result of progeroid-like symptoms and live five times shorter than wild-type (WT) littermates (Dollé et al. 2011; Vermeij et al. 2016). The livers of *Ercc1*$^{-/\Delta}$ mice display various aging-like characteristics and pathology (Dollé et al. 2011; Gregg et al. 2012; Niedernhofer et al. 2006; Weeda et al. 1997), whereas, other organs, such as the small intestine, do not show an obvious pathological phenotype. Thus the consequences of loss of ERCC1 differ considerably between tissues, although the reason for this remains unclear. XPC is involved in the recognition of bulky DNA adducts in the GG-NER pathway specifically (Puumalainen et al. 2015; Iyama and Wilson 2013). Germline mutations in this gene cause Xeroderma Pigmentosum, a disorder characterized by enhanced sensitivity to UV-light and development of various cancer types at an early age (Sands et al. 1995; Melis et al. 2008; Dupuy and Sarasin 2015).

Here, we combined the organoid culture system with *in vivo* and *in vitro* knockout models, providing us with the unique opportunity to characterize the genome-wide mutational consequences of NER-deficiency in a stable genetic background. Both quantitative and qualitative differences were identified, creating novel insight into the molecular processes underlying mutation accumulation, cancer, and aging.

RESULTS

Loss of NER protein ERCC1 increases the number of base substitutions in liver, but not in small intestinal mouse ASCs

To characterize the mutational consequences of NER-deficiency, we generated clonal organoid cultures from single liver and small intestinal ASCs of three female *Ercc1*$^{-/\Delta}$ mice and three female WT littermates (Fig. 1A). The tissues were harvested at the age of 15 weeks, which is the time point at which *Ercc1*$^{-/\Delta}$ mice generally start to die as a consequence of early aging pathologies (Vermeij et al. 2016). WGS analysis of DNA isolated from the clonal organoid cultures allows for reliable determination of the somatic mutations that were accumulated during life in the original ASCs (Blokzijl et al. 2016; Jager et al. 2018). Subclonal mutations acquired after the single-cell-step will only be present in a subpopulation of the cells and are filtered out based on a low allele frequency (Jager et al. 2018). We also sequenced the genomes of polyclonal biopsies from the tail of each mouse, which served as control samples to exclude germline variants.

To determine transcriptome profiles, we performed RNA sequencing on one clonal organoid culture from each tissue of each mouse. As expected, *Ercc1* is significantly differentially expressed ($P < 0.05$, negative binomial test) between

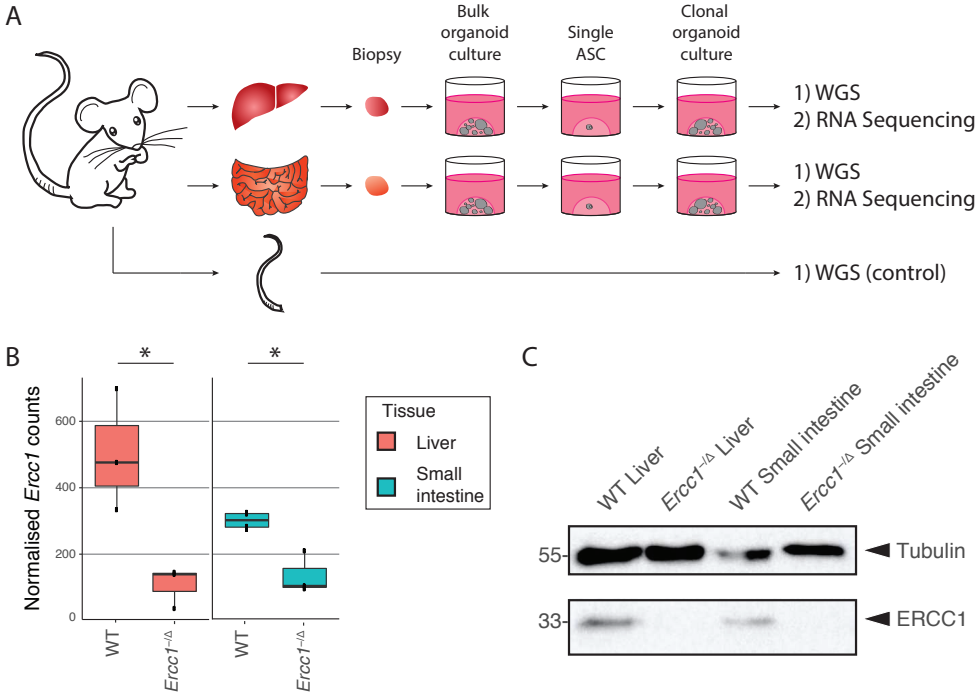


Figure 1. Experimental setup and tissue-specific expression of *Ercc1* in mouse ASCs. (A) Schematic overview of the experimental setup used to determine the mutational patterns in single ASCs from the liver and small intestine of mice. Biopsies from the liver and small intestine of six 15-week-old female mice (three *Ercc1*$^{-/\Delta}$ mice and three WT littermates) were cultured in bulk for ~1.5 week to enrich for ASCs. Subsequently, clonal organoids were derived from these bulk organoid cultures and expanded for approximately 1 month, until there were enough cells to perform both WGS and RNA sequencing. As a control sample for filtering germline variants, a biopsy of the tail of each mouse was also subjected to WGS. (B) Boxplots depicting normalized mRNA counts of *Ercc1* in ASC organoid cultures from liver and small intestine of *Ercc1*$^{-/\Delta}$ mice (n = 3 and n = 3, respectively) and WT littermates (n = 3 and n = 4, respectively). Asterisks represent significant differences ($P < 0.05$, negative binomial test). (C) Western blot analysis of ERCC1 in *Ercc1*$^{-/\Delta}$ and WT small intestinal and liver mouse organoids.

WT and *Ercc1*$^{-/\Delta}$ in both liver and small intestinal ASCs (Fig. 1B), confirming the anticipated effects of the *Ercc1* mutations at the mRNA level. While there is some *Ercc1* expression in *Ercc1*$^{-/\Delta}$ ASCs, the C-terminal domain of ERCC1 is essential in ERCC1-XPF complex formation and disruption of this interaction reduces the stability of ERCC1 protein (Tripsianes et al. 2005; de Laat 1998; Sijbers et al. 1996b). Indeed, ERCC1 protein is not detectable by immunoblotting in *Ercc1*$^{-/\Delta}$ organoid cultures of both tissues (Fig. 1C). No other DNA repair genes were differentially expressed between WT and *Ercc1*$^{-/\Delta}$ ASCs (Supplemental File S1). Notably, the expression of 8 out of 9 core NER genes, including *Ercc1*, is higher in WT liver ASCs than WT small intestinal ASCs (Supplemental Fig. S1, Supplemental Table S1).

WGS analysis on the clonally-expanded organoid cultures revealed 4,238 somatic base substitutions in the autosomal genome of 11 clonal ASC samples (Fig. 2A; Supplemental Table S2). Liver ASCs of WT mice acquired 19.5 ± 4.1 (mean ± standard deviation) base substitutions per week. This rate is similar in ASCs of the small intestine, at 16.1 ± 3.1 mutations per week, and is in line with the observation that human liver and intestinal ASCs have similar mutation accumulation rates *in vivo* (Blokzijl et al. 2016). Loss of ERCC1 induced a twofold increase (45.5 ± 3.0 base substitutions per week) in the number of base substitutions in ASCs of the liver (Fig. 2A, Supplemental Fig. S2A). However, we did not observe a different mutation rate in small intestinal ASCs of *Ercc1*$^{-/\Delta}$ mice (21.0 ± 4.9 base substitutions per week) compared with WT small intestinal ASCs (Fig. 2A, Supplemental Fig. S2A). We also observed a significant increase in the number of double base substitutions in liver ASCs lacking ERCC1 (q < 0.05, t-test, FDR correction; Fig. 2B, Supplemental Fig. S2B, Supplemental Table S3). *Ercc1*$^{-/\Delta}$ liver ASCs acquire 0.49 ± 0.06 double base substitutions per week, while WT liver ASCs acquire only 0.05 ± 0.04 double base substitutions per week. Again, we did not observe this difference between WT and mutant ASCs of the small

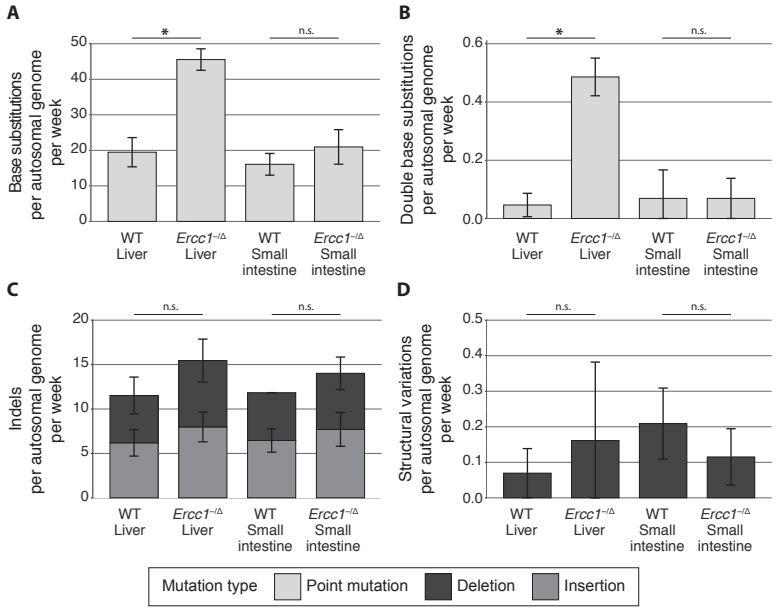


Figure 2. Somatic mutation rates in the genomes of ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice. (A) Base substitutions, (B) double base substitutions, (C) indels, and (D) SVs acquired per autosomal genome per week in ASCs of WT liver (n = 3), *Ercc1*$^{-/\Delta}$ liver (n = 3), WT small intestine (n = 2), and *Ercc1*$^{-/\Delta}$ small intestine (n = 3). Error bars represent standard deviations. Asterisks represent significant differences (q < 0.05, two-sided t-test, FDR correction). n.s. : non-significant (q ≥ 0.05, two-sided t-test, FDR correction).

intestine (0.07 ± 0.10 and 0.07 ± 0.07 per week, respectively). The increased number of double base substitutions in the liver ASCs remained significant after normalizing for the total number of base substitutions (q < 0.05, t-test, FDR correction; Supplemental Fig. S2C), indicating a liver-specific enrichment of double base substitutions in *Ercc1*$^{-/\Delta}$ ASCs compared with WT.

In addition to the 4,238 base substitutions, we identified 2,116 small insertions and deletions (indels) and 21 larger deletions (≥100 bp) in the autosomal genome of the 11 clonal ASC samples (Supplemental Table S2). As opposed to the base substitutions, we observed similar indel numbers for WT and *Ercc1*$^{-/\Delta}$ ASCs of both tissues (Fig. 2C, Supplemental Fig. S2D). Of note, identification of indels is more challenging than base substitutions, and as a result, these calls might contain more false positives. ASCs in the small intestine and liver of the mice acquire approximately 13.3 ± 3.4 indels per week, independent of *Ercc1* mutation status. Likewise, loss of ERCC1 did not influence the number or type of structural variations (SVs) in ASCs of the small intestine and the liver (Fig. 2D, Supplemental Fig. S2E, Supplemental Table S4). Each mouse ASCs carried 0 - 6 deletions (median length of 539 bp; Supplemental Table S4). Finally, a genome-wide copy-number profile was generated to identify chromosomal gains and losses. These profiles indicated that all WT and *Ercc1*$^{-/\Delta}$ ASCs were karyotypically stable during life (Supplemental Fig. S3). Nevertheless, some subclonal aneuploidies were detected in a WT as well as an *Ercc1*$^{-/\Delta}$ liver organoid sample, which are most likely culturing artefacts that occurred *in vitro* after the clonal step irrespective of *Ercc1* mutation status.

Loss of NER protein ERCC1 induces Signature 8 mutations in mouse ASCs

To further dissect the mutational consequences of NER-deficiency, we characterized the mutation spectra in the mouse ASCs. Regardless of tissue-type, the mutation spectra of all assessed ASCs are predominantly characterized by C:G > A:T mutations and C:G > T:A mutations (Fig. 3A). However, the mutation spectra of NER-proficient and NER-deficient ASCs differed significantly for both tissues (q < 0.05, X^2-test, FDR correction). Indeed, there are some notable differences, such as an increased contribution of T:A > A:T mutations in *Ercc1*$^{-/\Delta}$ ASCs compared with WT ASCs (Fig. 3A).

To gain insight into these differences, we generated 96-channel mutational profiles of all ASCs (Supplemental Fig. S4) and assessed the contribution of each COSMIC mutational signature (http://cancer.sanger.ac.uk/cosmic/signatures) to the average 96-channel mutational profile (centroid) per group (Supplemental Fig. S6B). We could reconstruct the original centroids well with the 30 COSMIC signatures, as the reconstructed centroids are highly similar to the original centroids for all four

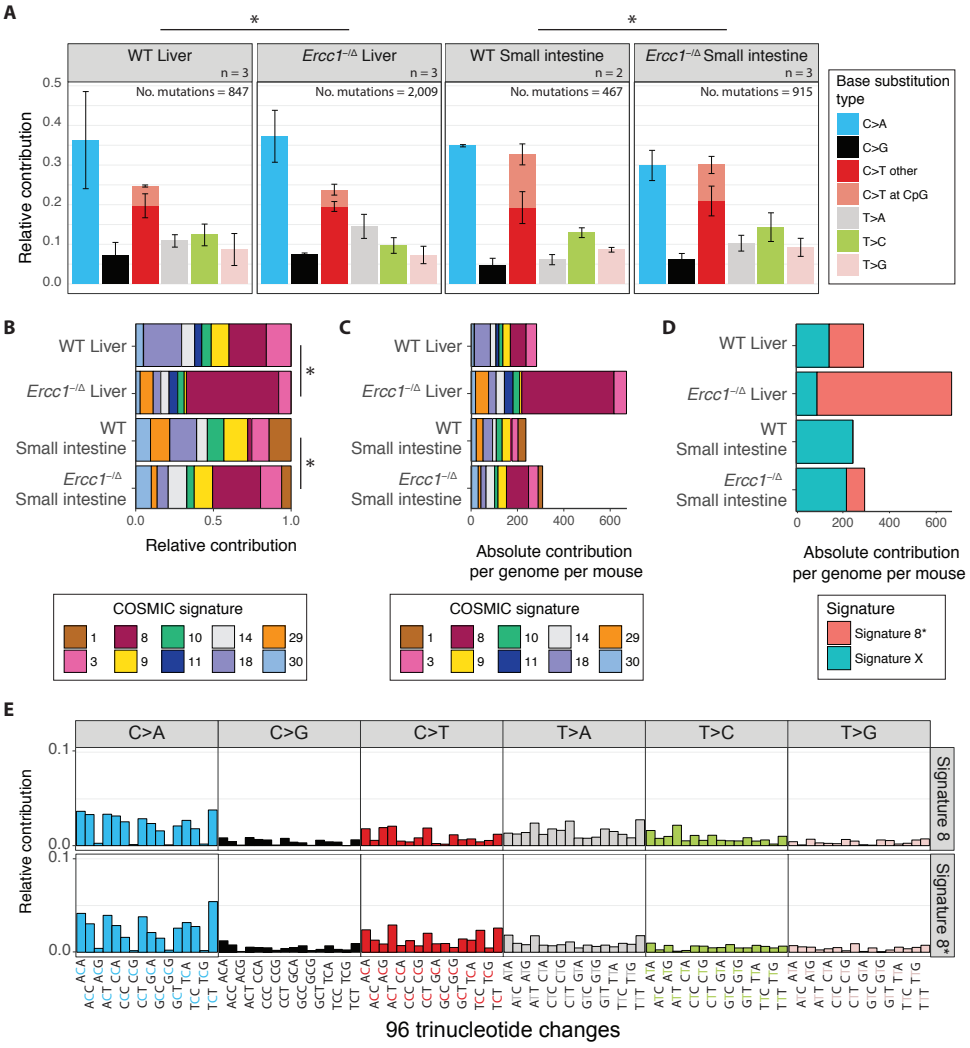


Figure 3. Mutational patterns of base substitutions acquired in the genomes of ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice. (A) Mean relative contribution of the indicated mutation types to the mutation spectrum for each mouse ASC group. Error bars represent standard deviations. The total number of mutations, and total number of ASCs (n) per group is indicated. Asterisks indicate significant differences in mutation spectra ($q < 0.05$, X^2-test, FDR correction). (B) Relative contribution of the indicated COSMIC mutational signatures to the average 96-channel mutational profiles of each mouse ASC group. Asterisks indicate significantly different signature contributions, *P* values where obtained using a bootstrap resampling approach (Methods, Supplemental Fig. S6E-F) (C) Absolute contribution of the indicated COSMIC mutational signatures to the average 96-channel mutational profiles of each mouse ASC group. (D) Absolute contribution of two mutational signatures that were identified by non-negative matrix factorization (NMF) analysis to the average 96-channel mutational profiles of each mouse ASC group. (E) Relative contribution of each indicated context-dependent base substitution type to mutational Signature 8 and Signature 8*.

ASC groups (average cosine similarity = 0.95, Supplemental Fig. S6A). Although the mutational profiles of the ASC groups are comparable in the 96 dimensions (average cosine similarity = 0.90, Supplemental Fig. S5), the contribution of the COSMIC signatures is significantly different between NER-proficient and NER-deficient ASC groups for both liver and small intestine ($d > d_{WT\_0.05}$ *and* $d > d_{MUT\_0.05}$, bootstrap resampling method, see Methods, Supplemental Fig. S6).

We subsequently reconstructed the 96-channel mutational profiles using the top 10 contributing COSMIC mutational signatures (Fig. 3B-C; Supplemental Fig. S6). We could reconstruct the centroids comparably well using this subset of 10 COSMIC signatures (average cosine similarity = 0.95, Supplemental Fig. S6A). The 96-channel mutational profiles of NER-deficient liver ASCs not only closely resemble Signature 8 (cosine similarity of 0.92; Supplemental Fig. S7), but Signature 8 can almost fully explain the increase in base substitutions in NER-deficient liver ASCs (Fig. 3B-C). The number of Signature 8 mutations is also increased in all small intestinal ASCs of *Ercc1*$^{-/\Delta}$ mice compared with WT small intestinal ASCs (Fig. 3B-C). This finding shows that NER-deficiency can result in elevated numbers of Signature 8 mutations in ASCs, regardless of tissue-type.

In addition, we performed an unbiased signature analysis by extracting two mutational signatures *de novo* from the mouse mutation catalogs using non-negative matrix factorization (NMF) (Supplemental File S2, Supplemental Fig. S8). One of the identified signatures, Signature X, contributes approximately 100 mutations to liver ASCs and 200 mutations to small intestinal ASCs, in both WT and *Ercc1*$^{-/\Delta}$ mice (Fig. 3D), suggesting that this signature represents a mutational process that is generally active in mouse ASCs. In line with this, Signature X is highly similar to 96-channel mutational profiles of ASCs of the small intestine of old mice (Behjati et al. 2014) (cosine similarity = 0.95, Supplemental Fig. S8B). As expected, this mouse signature is not similar to any of the known COSMIC signatures identified in human tumor sequencing data (Supplemental Fig. S8B). The other signature, Signature 8*, is highly similar to COSMIC Signature 8 (cosine similarity = 0.91; Fig. 3E) and has an increased contribution in *Ercc1*$^{-/\Delta}$ liver ASCs compared with WT (Fig. 3D; Supplemental Fig. S9C). Moreover, the contribution of Signature 8* mutations is also increased in *Ercc1*$^{-/\Delta}$ small intestinal ASCs in comparison to WT small intestinal ASCs (Fig. 3D; Supplemental Fig. S9C). These findings confirmed that NER-deficiency results in base substitutions that show a 96-channel profile similar to COSMIC Signature 8.

Mutations are distributed non-randomly throughout the genome in cancer cells and in human ASCs (Schuster-Böckler and Lehner 2012; Blokzijl et al. 2016). NER is one of the pathways that is suggested to underlie this non-random distribution of mutations (Perera et al. 2016; Zheng et al. 2014). Firstly, NER-activity

has been linked to a local enrichment of mutations at gene promoters (Perera et al. 2016). However, we do not observe any significant differences in the depletion of mutations in promoters, promoter-flanking, and enhancer regions between NER-proficient and -deficient ASCs (Supplemental Fig. S9A). Secondly, TC-NER results in a depletion of mutations in expressed genes, as this pathway repairs lesions on the transcribed strand during transcription (Pleasance et al. 2010). Mutations are indeed depleted in genic regions of NER-proficient WT mouse ASCs, but the depletion is not significantly different in NER-deficient ASCs (n.s., Poisson test, FDR correction; Supplemental Fig. S9A). Moreover, the average expression levels of genes in which the somatic mutations are located do not differ between *Ercc1*$^{-/\Delta}$ and WT ASCs (n.s., *t*-test, FDR correction; Supplemental Fig. S9B), suggesting that *Ercc1*$^{-/\Delta}$ ASCs do not accumulate more mutations in expressed genes. Finally, there are no obvious changes in transcriptional strand bias, although the mutation numbers are too low to be conclusive (Supplemental Fig. S9C). NER-deficiency thus influences both the mutation load and mutation type, but not the genomic distribution of the observed base substitutions in mouse ASCs, suggesting that the contribution of TC-NER in the observed mutational consequences is minimal in these cells.

Loss of GG-NER protein XPC induces Signature 8 mutations in human ASCs

To identify a potential causal relationship between NER-deficiency and Signature 8 in human ASCs, we generated a human GG-NER deficient *XPC*^{KO} ASC using CRISPR-Cas9 gene-editing in a human small intestinal organoid culture (Fig. 4A). After confirming absence of XPC protein (Fig. 4B), we passaged the *XPC*^{KO} clones for ± 2 months to allow the accumulation of sufficient mutations for downstream analyses. Subsequently, we derived subclonal cultures of single ASCs and expanded these until sufficient DNA could be isolated for WGS. This approach allowed us to catalog the mutations that specifically accumulated between the two clonal expansion steps in the absence of XPC (Supplemental Fig. S10A) (Drost et al. 2017; Blokzijl et al. 2016; Jager et al. 2018). As a control, WGS data of three previously-established *XPC*^{WT} organoid cultures of the same human donor was used (Blokzijl et al. 2016).

Similar to the *Ercc1*$^{-/\Delta}$ mouse ASCs, loss of XPC in human ASCs induced an increase in the genome-wide number of base substitutions acquired per week. (3-fold increase; Fig. 4C, Supplemental Fig. S10B, Supplemental Table S5). In addition, the number of double base substitutions acquired per week was approximately 17 times higher (Fig. 4D, Supplemental Table S5, Supplemental Table S6). We did not observe a marked change in the genomic distribution of acquired mutations as a result of *XPC* deletion in human ASCs, nor a change in transcriptional strand bias (Supplemental Fig. S10C-D). In total, approximately 39% of the increase in base

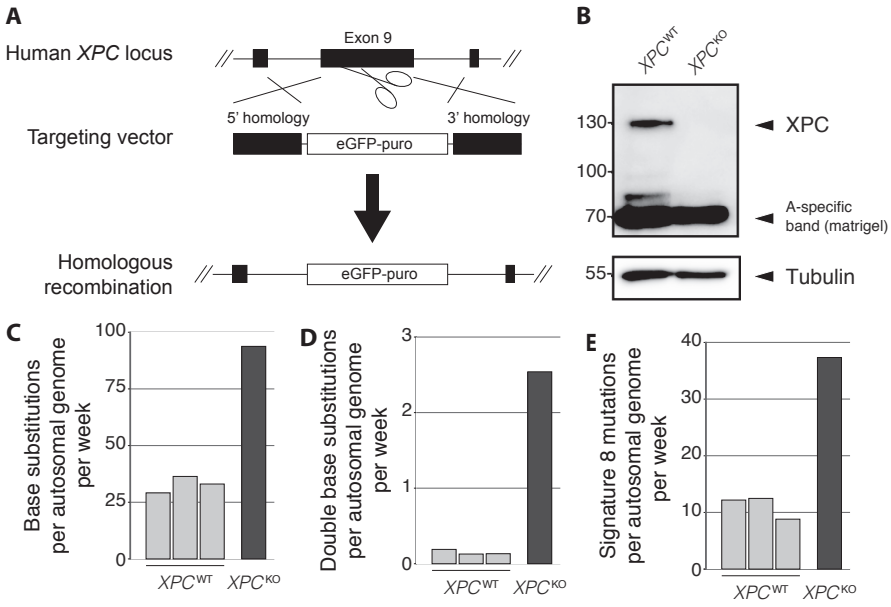


Figure 4. Mutational consequences of *XPC*^{KO} in human intestinal organoid cultures *in vitro*. (A) Targeting strategy for the generation of *XPC*^{KO} organoid cultures using CRISPR-Cas9 gene-editing. (B) Western blot analysis of XPC in human *XPC*^{WT} and *XPC*^{KO} organoids. (C) Number of base substitutions, (D) double base substitutions, and (E) Signature 8 mutations acquired per autosomal genome per week in human *XPC*^{WT} ASCs (n = 3) and an *XPC*^{KO} ASC (n = 1) *in vitro*.

substitutions in the *XPC*^{KO} ASC can be explained by Signature 8 (Fig. 4E, Supplemental Fig. S10B), confirming that NER-deficiency can cause an increase in the number of Signature 8 mutations, independent of tissue-type or species.

DISCUSSION

We exploited mouse knockouts, organoid culturing, CRISPR-Cas9 gene-editing, WGS, and mutational signature analyses to study the genome-wide mutational consequences of NER-deficiency in individual ASCs of human and mice. Our results show that loss of ERCC1 induces a significant increase in the accumulation of base substitutions in liver ASCs, but not in small intestinal ASCs *in vivo*. Interestingly, the mutational increase coincides with the tissue-specific pathological aging phenotype observed in *Ercc1*$^{-/\Delta}$ mice (Dollé et al. 2011; Gregg et al. 2012). A possible explanation for this difference between tissues is that liver ASCs might be more dependent on DNA repair facilitated by ERCC1 compared with small intestinal ASCs, e.g. as a result of tissue-specific mutagen exposure. In line with this, WT liver ASCs show a higher basal expression of *Ercc1* and other NER genes compared with WT small intestinal ASCs. However, the transcription levels of DNA repair components do not necessarily

reflect DNA repair-activity, due to post-transcriptional regulation (Naipal et al. 2015). Alternatively, liver and small intestinal ASCs might cope differently with unrepaired DNA damage as a result of loss of ERCC1, such as the utilization of alternative DNA repair mechanisms, like translesion synthesis (TLS) polymerases, to bypass polymerase-blocking lesions, or differential induction of apoptosis or senescence.

ERCC1 is involved in multiple DNA repair pathways, including TC-NER, GG-NER, SSA, and crosslink repair. Previously, it has been shown that SSA- and crosslink repair-deficiencies result in increased number of indels and SVs in mice, whereas NER-deficiency introduces base substitutions (Dollé et al. 2006). Since we only observe an increase in base substitutions, NER-deficiency is likely responsible for the mutational consequences of loss of ERCC1 in liver ASCs *in vivo*. If TC-NER-deficiency underlies the differential mutation accumulation, this would be reflected by an increase in mutations in expressed genes in *Ercc1*$^{-/\Delta}$ mice. However, WT and *Ercc1*$^{-/\Delta}$ cells show a similar depletion of mutations in genes, indicating that the observed mutational consequences of impaired ERCC1 is rather an effect of defective GG-NER. In line with this, we show that GG-NER-deficiency can also induce an increase in the number of base substitutions in a human small intestinal organoid culture that is deleted for GG-NER component *XPC*. More specifically, the increased base substitution load can be largely explained by an increased contribution of Signature 8 in both systems. In line with our observations, a mutational signature similar to Signature 8 has been shown to increase with age in the neurons of NER-deficient patients (Lodato et al. 2017).

Until now, the etiology of Signature 8 was unknown (http://cancer.sanger.ac.uk/cosmic/signatures). As Signature 8 mutations are also detected in healthy human and mouse ASCs (Fig. 3B, Fig. 4E), this signature most likely represents a mutagenic process that is generally active in normal cells. Signature 8 is characterized by C:G > A:T mutations and is associated with double base substitutions (Alexandrov et al. 2013; Nik-Zainal et al. 2016). C:G > A:T mutations have been linked to several processes, including oxidative stress (Kamiya et al. 1995; Degtyareva et al. 2013). Consistently, organoid culturing causes mutations indicative of high oxidative stress (Blokzijl et al. 2016). Interestingly, NER has been suggested to play a role in the repair of tandem DNA lesions that result from oxidative stress (Bergeron et al. 2010; Cadet et al. 2012). If left unrepaired, these lesions can block regular DNA polymerases, but can be bypassed by error-prone TLS polymerases, resulting in increased incorporation of tandem mutations (Cadet et al. 2012). Moreover, it has been shown that oxidative stress results in increased induction of double base substitutions in NER-deficient human fibroblasts (Lee 2002). In line with this, we observe a significant increase in the double base substitution load in mouse liver ASCs and a similar trend in the

human ASC culture as a result of NER-deficiency, yet the number of double base substitutions is much lower than single base substitutions. Thus Signature 8 could reflect oxidative DNA damage bypassed by TLS.

Although NER-deficiency does not affect the base substitution load in the mouse small intestine, it does result in an increased contribution of Signature 8 mutations. This is in clear contrast to mouse liver ASCs, where NER-deficiency has both a qualitative and quantitative consequence on the accumulation of base substitutions. More specifically, the absolute contribution of Signature 8 mutations is similar in WT liver and *Ercc1*$^{-/\Delta}$ small intestinal ASCs. This clearly demonstrates that DNA-repair deficiency can have tissue-specific consequences, but also indicates that the absolute contribution of Signature 8 mutations should be compared to the basal contribution in the same tissue in order to detect NER-deficiency.

We did not observe a notable contribution of signatures that have been previously observed in liver cancer in ASCs of *Ercc1*$^{-/\Delta}$ livers (http://cancer.sanger.ac.uk/cosmic/signatures) (Supplemental Fig. S6B). This finding suggests that the mutational processes that underlie these signatures are only active after oncogenic transformation, or that mutagen exposure in liver cancer (progenitor) cells is different from *in vivo* mouse ASCs and *in vitro* human ASCs. Liver cancer-specific Signature 24, for example, is associated with Aflatoxin intake (Huang et al. 2017), a substance to which our mice and organoids were not exposed. In addition, Signature 1 and Signature 5, which have been previously associated with age (Blokzijl et al. 2016; Alexandrov et al. 2015), did not have an increased contribution in the ASCs of progeroid *Ercc1*$^{-/\Delta}$ mice. Finally, a high contribution of mutational Signature 5 has been linked to the presence of somatic mutations in *ERCC2*, a key factor in both TC-NER and GG-NER, in human urothelial cancer (Kim et al. 2016; Iyama and Wilson 2013). As mentioned however, we did not observe an increase Signature 5 contribution in ASCs without ERCC1 or XPC. This difference in mutational consequences could reflect various differences between these systems, such as different effects of the mutations on protein function, distinct roles of the proteins, or tumor- and/or tissue-specific activity of mutagenic damage and/or DNA repair processes. In our study, we deleted specific NER components in an otherwise normal genetic background, providing us with the unique opportunity to characterize the direct mutational consequences of NER-deficiency.

Determination of the NER-capacity of tumors can be important for precision medicine, as it has been shown that tumors with mutations in NER genes (Stubbert et al. 2010; Van Allen et al. 2014; Amable 2016; Zhang et al. 2017), and tumors with low expression of *ERCC1* (Olaussen et al. 2006; Li et al. 2000; Amable 2016) are sensitive to cisplatin treatment. However, translation of these findings into the clinical

setting has been challenging, because connecting tumor biopsy mRNA levels and immunohistochemistry measurements to NER-activity remains an unresolved issue (Bowden 2014), and interpreting the effects of mutations in DNA repair genes on NER-capacity is challenging. Rather than looking for the presence of causal events, mutational catalogs can be used as a functional readout of NER-capacity in tumors. Here, we show that NER-deficiency can induce an increase in Signature 8 mutations in both mouse and human ASCs. Signature 8 has an overall prevalence of 2% in sequenced human tumors, is found in medulloblastoma (Alexandrov et al. 2013) and contributes to the mutational profile of the majority of breast cancer tumors (Nik-Zainal et al. 2016; Alexandrov et al. 2013). Our results show that an increase in the number of Signature 8 mutations with respect to the normal number of Signature 8 mutations in a cancer type might serve as a novel biomarker for (GG-)NER-deficiency and has the potential to guide treatment decision. Additional clinical studies will be required to demonstrate the predictive value of these mutations for treatment response.

ACKNOWLEDGEMENTS

The authors would like to thank the the animal caretakers of the Erasmus MC for taking care of the mice and the Utrecht Sequencing Facility for providing the sequencing service and data. Utrecht Sequencing Facility is subsidized by the University Medical Center Utrecht, Hubrecht Institute and Utrecht University. This study was financially supported by the NWO Zwaartekracht program Cancer Genomics.nl.

AUTHOR CONTRIBUTIONS

M.J., E.K., M.V., N.B., and R.B. performed organoid culturing. N.B. and R.B. generated western blots and sequenced the organoid cultures. M.J., F.B., R.J., S.B., J.L., and R.B. performed bioinformatic analyses. M.J., F.B., E.K., J.H., J.P., R.B., and E.C. were involved in the conceptual design of this study. M.J., F.B., R.B., and E.C. wrote the manuscript.

METHODS

Mouse tissue material

Ercc1$^{-/\Delta}$ mice were generated and maintained as previously described (Vermeij et al., 2016). Briefly, by crossing *Ercc1*$^{\Delta/+}$ (C57BL6J or FVB background) with *Ercc1*$^{+/-}$ mice (FVB or C57BL6J background), *Ercc1*$^{-/\Delta}$ mice were generated in a uniform F1 C57BL6J/FVB hybrid background. Wild type F1 littermates were used as controls. Animals were housed in individually ventilated cages under specific pathogen-free conditions in a controlled environment (20–22 °C, 12 h light : 12 h dark cycle). Experiments were performed in accordance with the Principles of Laboratory Animal Care and with the guidelines approved by the Dutch Ethical Committee in full accordance with European legislation.

We used three 15-week old female *Ercc1*$^{-/\Delta}$ mice and three female WT littermates for our experiments. Tails were harvested and stored at -20°C. Livers and small intestines were harvested and kept on ice in Adv+++ medium (Advanced DMEM/F-12 with 1% GlutaMAX, HEPES 10 mM and 1% penicillin/streptomycin) for a few hours until further processing.

Human tissue material

Endoscopic biopsies were performed at the University Medical Center Utrecht and the Wilhelmina Children's Hospital. The patients' informed consent was obtained and this study was approved by the ethical committee of University Medical Center Utrecht.

Generation of clonal *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures

Single liver ASCs were isolated from livers as described previously (Kuijk et al. 2016). Liver organoid cultures were initiated by culturing the liver ASCs in BME overlaid with mouse liver culture initiation medium (50% Adv+++ medium, 35% WNT3A conditioned medium (produced in house), 5% NOGGIN conditioned medium (produced in house), 5% RSPOI conditioned medium (produced in house), 1x B27 without retinoic acid, 1x N2, 1x Primocin, 10mM Nicotinamide, 0.625mM N-acetylcysteine, 100ng/ml FGF-10, 10µM ROCKi, 50 ng/ml HGF, 10nM Gastrin, and 50ng/ml hEGF). 1.5 week after culture initiation, clonal organoid liver cultures were generated and expanded according to protocol (Jager et al. 2018) in mouse liver expansion medium (90% Adv+++ medium, 5% RSPOI conditioned medium (produced in house), 1x B27 without retinoic acid, 1x N2, 1x Primocin, 10mM Nicotinamide, 0.625mM N-acetylcysteine, 100ng/ml FGF-10, 50 ng/ml HGF, 10nM Gastrin, and 50ng/ml hEGF).

Crypts were isolated from small intestines as described previously (Sato et al. 2009). Small intestinal organoid cultures were initiated by culturing the small intestinal ASCs in matrigel overlaid with mouse small intestine medium (50% WNT3A conditioned medium (produced in house), 30% Adv+++ medium, 10% NOGGIN conditioned medium (produced in house), 10% RSPOI conditioned medium (produced in house), 1x B27, 1x hES Cell Cloning & Recovery Supplement, 1x Primocin, 10µM ROCKi, 1.25mM N-acetylcysteine, and 50ng/ml hEGF). Clonal small intestinal organoid cultures were generated by picking single organoids manually and clonally expanding these organoid cultures according to protocol in mouse small intestine medium (Jager et al. 2018). Culture expansion failed for the small intestine of mouse WT1.

Generation of a clonal and subclonal *XPC*^{KO} organoid culture

Clonal *XPC*^{KO} organoid cultures were generated from a small intestinal bulk organoid culture derived previously (Blokzijl et al. 2016) using the CRISPR-Cas9 gene-editing technique as described in (Drost et al. 2017). One clonal human *XPC*^{KO} organoid culture was obtained and cultured for 72 days in human small intestinal organoid medium (50% WNT3A conditioned medium (produced in house), 30% Adv+++ medium, 20% RSPOI conditioned medium (produced in house), 1x B27, 1x Primocin, 1.25mM N-acetylcysteine, 0.5µM A83-01, 10µM SB202190, 100ng/ml recombinant Noggin, and 50ng/ml hEGF). Subsequently, a subclonal culture was derived according to protocol (Jager et al. 2018).

Western blot

Protein samples from mouse organoid cultures were collected in Laemmli buffer and measured using the Qubit^{tm} 3.0 Fluorometer (Thermo Fisher Scientific) with the Qubit^{tm} Protein Assay Kit (Thermo Fisher Scientific, Q33211). Protein samples from human organoid cultures were collected in Laemmli buffer and measured using a Lowry protein assay. 30µg of protein per sample was run on a 10% SDS page gel. Subsequently, the proteins were transferred to a nitrocellulose membrane. After transfer, the membrane was blocked for 1 hour using 5% ELK (Campina) at room temperature and subsequently incubated overnight with the primary antibody (ERCC1: Abcam, ab129267; XPC: Cell Signaling Technology; #12701). Secondary antibody was incubated 1 hour at room temperature, and subsequently proteins were visualized using the Amersham ECL Western blotting analysis system (GE Healthcare, RPN2109) and the Amersham Imager 600 system (GE Healthcare).

RNA sequencing and differential expression analysis of *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures

For each mouse (three *Ercc1*$^{-/\Delta}$ mice and three WT littermates), we performed RNA sequencing on one clonal organoid culture from the liver and the small intestine. An additional small intestinal organoid clone was sequenced of mice WT2 and WT3 to increase the amount of replicates for differential expression analysis, as culture expansion failed for the small intestine of WT1. Total RNA was collected in TRIzol and purified from all organoid cultures using the Qiasymphony (Qiagen). RNA libraries for Illumina sequencing were generated from 50 ng of poly-A selected mRNA using the Neoprep (Illumina) and sequenced 2 x 75 bp paired-end to approximately 3300 Million base pairs per sample with the Illumina NextSeq 500 at the Utrecht Sequencing Facility.

RNA sequencing reads were mapped with STAR v.2.4.2a to the mouse reference genome GRCm38. The BAM files were sorted with Sambamba v0.5.8 and reads were counted with HTSeq-count version 0.6.1p1 (default settings) to exons as defined in GRCm38v70.gtf (Ensembl). Non-uniquely mapped reads were not counted. Subsequently, DESeq v1.28.0 was used to normalize counts. DESeq nbinomTest was used to test for differential expression (1) of *Ercc1* between *Ercc1*$^{-/\Delta}$ and WT liver ASCs, (2) of *Ercc1* between *Ercc1*$^{-/\Delta}$ and WT small intestinal ASCs, (3) of 83 other DNA repair genes (Casorelli et al. 2006) between *Ercc1*$^{-/\Delta}$ and WT liver ASCs, and (4) between *Ercc1*$^{-/\Delta}$ and WT small intestinal ASCs, and (5) of 9 NER genes between the WT liver and WT small intestinal ASCs. Differentially expressed genes with $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing correction) were considered significant.

WGS and read alignment

DNA was isolated from mouse liver organoid cultures and mouse control (tail) samples using the genomic tip 20-G kit (Qiagen) and from mouse small intestinal organoid samples and the human *XPC*^{KO} sample using the Qiasymphony (Qiagen). DNA libraries for Illumina sequencing were generated from 200 ng genomic DNA using standard protocols (Illumina) and sequenced 2 x 100 bp paired-end to 30X base coverage with the Illumina HiSeq Xten at the Hartwig Medical Foundation. The sequence reads of *XPC*^{KO} were mapped to the GRCh37 human reference genome using using the Burrows–Wheeler Aligner (BWA) v0.7.5a (Li and Durbin 2009), with settings '-t 4 -c 100 -M'. The mapped data of clonal *XPC*^{WT} organoids was previously generated in the study ('donor_id' 6) (Blokzijl et al. 2016). The sequence reads of the mouse ASCs were mapped to the GRCm38 mouse reference genome using using the Burrows–Wheeler Aligner (BWA) v0.7.5a (Li and Durbin 2009), with settings '-t 4 -c 100 -M'. The WGS data of the tails confirmed that the *Ercc1*$^{-/\Delta}$mice have compound heterozygous mutations in *Ercc1* and the WT littermates do not (Supplemental Fig. S11).

Callable genome

The callable genome was defined for all sequenced samples using the GATK CallableLoci tool v3.4.46 (Van der Auwera et al. 2013) with default settings and additional optional parameters 'minBaseQuality 10', 'minMappingQuality 10', 'maxFractionOfReadsWithLowMAPQ 20', and 'minDepth 20'. 'CALLABLE' regions were extracted from every output file. Subsequently, genomic regions that were callable (1) in the mouse organoid clone and the control (tail) sample, and (2) in the human organoid clone, subclone, and control (blood) were intersected to define a genomic region that is surveyed in all samples that were compared. Approximately 90 ± 1% of the autosomal genome was surveyed in every mouse clone (Supplemental Table S2), and 73 - 88% of the autosomal genome was surveyed in each human subclone (Supplemental Table S5).

Base substitution and indel calling

For both human and mouse samples, base substitutions and indels were multi-sample called with GATK HaplotypeCaller v3.4.46 with default settings and additional options '-stand_call_conf 30 -stand_emit_conf 15' and GATK Queue v3.4.46. For mouse samples the quality of the calls was assessed using GATK VariantFiltration v3.4.46 with options 'QD < 2.0, MQ < 40.0, FS > 60.0, HaplotypeScore > 13.0, MQRankSum < -12.5, ReadPosRankSum < -8.0' for base substitutions and 'QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0' for indels, with additional options 'clusterSize 3' and 'clusterWindowSize 35'. For human samples the quality of the calls was assessed using GATK VariantFiltration v3.4.46 with options 'QD < 2.0, MQ < 40.0, FS > 60.0, HaplotypeScore > 13.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1), DP < 5, QUAL < 30, QUAL >= 30.0 && QUAL < 50.0, SOR > 4.0' for base substitutions and 'QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0, MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1), DP < 5, QUAL < 30.0, QUAL >= 30.0 && QUAL < 50.0, SOR > 10.0' for indels, with additional options 'clusterSize 3' and 'clusterWindowSize 10.

Base substitution filtering

To obtain high-quality catalogs of somatic base substitutions, we applied a comprehensive filtering procedure. For the mouse samples, we only considered positions on the autosomal genome that were

callable (see "Callable genome") in both the organoid and control (tail) sample. We excluded positions at which indels were called, as these positions likely represent false-positive base substitution calls. Furthermore, we only included positions with a 'PASS' flag by GATK VariantFiltration, a GATK phred-scaled quality score ≥ 100, a sample-level genotype quality of 99 in the organoid culture and ≥ 10 in the control (tail) sample, and a coverage of ≥ 20X in the organoid and the tail sample. We subsequently excluded variants with any evidence in another organoid sample or control (tail) sample of the same mouse to remove germline variants. To exclude potentially missed germline events, we also removed positions that have any evidence in the organoid and/or control samples of the other mice. Finally, we excluded positions with a variant allele frequency (VAF) < 0.3 in the organoid sample to exclude mutations that were induced after the clonal step.

For the human samples, we only considered positions on the autosomal genome that were callable (see "Callable genome") in the control (blood) sample, clonal organoid and subclonal organoid culture. We considered mutations with a 'PASS' flag by GATK VariantFiltration and a GATK phred-scaled quality score ≥ 100. For both the clonal and subclonal organoid cultures, all variants with evidence in the control (blood) sample were excluded, to remove germline variants. To exclude potentially missed germline events, we removed positions that are in the Single Nucleotide Polymorphism Database v137. b3730, or in a blacklist with positions that are recurrent in unmatched individuals (BED-file available upon request). Subsequently, for both the clonal and subclonal cultures, all variants with a VAF < 0.3 were excluded. Finally, the resulting somatic base substitution catalogs of the clonal and subclonal cultures were compared and all events unique to the subclonal organoid were considered to be accumulated after the *XPC* deletion, that is: between the two sequential clonal expansion steps.

Clonality of organoid cultures

We validated whether the organoid samples were clonal based on the VAF of somatic base substitutions, before the final filter step (VAF < 0.3). Each cell acquires its own set of somatic mutations and the reads supporting a mutation will be diluted in the WGS data of non-clonal samples, resulting in a low VAF. After extensive filtering of somatic base substitutions, liver organoid samples from WT1, WT2, and *Ercc1*$^{-/\Delta}$2 showed a shift in the VAK-peak away from 0.5 and therefore these samples were excluded from further analyses (Supplemental Fig. S12). An additional liver organoid culture from these mice was sequenced and these samples were confirmed to be clonal (Supplemental Fig. S12).

Double base substitutions

We selected base substitutions from the filtered variant call format (VCF) files that were called on consecutive bases in the mouse or human reference genome. The double base substitutions were subsequently manually checked in the Integrative Genomics Viewer (IGV) to exclude double base substitutions present in the control sample, and/or with many base substitutions or indels in the region, as these are (likely) false positives.

Indel filtration of *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures

We only considered positions on the autosomal genome that were callable (see "Callable genome") and had a sequencing depth of ≥ 20X in both the organoid sample and the control (tail) sample. We excluded positions that overlap with a base substitution. Furthermore, we only considered positions with a filter 'PASS' from VariantFiltration, a GATK phred-scaled quality score > 250 and a sample-level genotype quality of 99 in both the organoid sample and the control (tail) sample. We subsequently excluded Indels that are located within 50 base pairs of an indel called in another organoid sample and indels with any evidence in another organoid sample or a control (tail) sample. Finally, we excluded positions with a VAF < 0.3 in the organoid sample.

SV calling and filtration of *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures

SVs were called with DELLY v0.7.2 with settings 'type DEL DUP INV TRA INS', 'map-qual 1', 'mad-cutoff 9', 'min-flank 13', and 'geno-qual 5' (Rausch et al. 2012). We only considered SVs of at least 100 bp on the autosomal chromosomes that were called with a filter 'PASS', and a sample-specific genotype quality of

at least 90 in the organoid culture and the control sample. We subsequently excluded positions with any evidence in the control (tail) sample. The filtered SVs were finally checked manually in IGV to reduce false-positives and we excluded SVs present in the tail sample, with no visible change in the read-depth (for duplications and deletions), and/or with many base substitutions in the region.

Genome-wide copy number profiles of *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures

To generate a virtual karyotype, genome-wide copy number states were determined using FreeC v7.2 with settings 'ploidy 2', 'window 1000' and 'telocentromeric 50000' (Boeva et al. 2012). Subsequently, the average copy number across bins of 500,000 bp was calculated and plotted to assess genome stability.

Base substitution types

We retrieved the base substitution types from all the filtered VCF files, converted them to the 6 types of base substitutions that are distinguished by convention, and generated a mutation spectrum (the C>T changes at NpCpG sites are considered separately from C>T changes at other sites) for the four ASC groups (*Ercc1*$^{-/\Delta}$ liver, *Ercc1*$^{-/\Delta}$ small intestine, WT liver, and WT small intestine), as well as *XPC*^{KO}, *XPC*^{WT}1, *XPC*^{WT}2, and *XPC*^{WT}3 ASCs. X^2-tests were performed to determine whether the mutation spectra differ significantly between (1) mouse WT and *Ercc1*$^{-/\Delta}$ liver ASCs, and (2) mouse WT and *Ercc1*$^{-/\Delta}$ small intestinal ASCs. *P* values were corrected for multiple testing using Benjamini-Hochberg FDR correction, and differences in mutation rates between *Ercc1*$^{-/\Delta}$ and WT mouse ASCs with $q < 0.05$ were considered significant.

We retrieved the sequence context for all base substitutions to generate the 96-channel mutational profiles for each assessed ASC. Subsequently, the centroid of the 96-channel mutational profiles was calculated per mouse ASC group. Pairwise cosine similarities of all 96-channel mutational profiles and of all centroids were computed. We also calculated the cosine similarities of the 96-channel mutational profiles and centroids with all 30 COSMIC mutational signatures (http://cancer.sanger.ac.uk/cosmic/signatures) (Supplemental Fig. S7). These analyses were performed with the R package MutationalPatterns (Blokzijl et al. 2018).

***De novo* mutational signature extraction**

We extracted two signatures using non-negative matrix factorization (NMF) from the 96-channel mutational profiles of the mouse ASCs. Although the number of base substitutions is low for this dimension reduction approach, it does provide an unbiased method to characterize the mutational processes that have been active in the ASCs. Subsequently, we computed the absolute contribution of these *de novo* extracted signatures to the centroids of the mouse ASC groups. We also calculated the cosine similarity of these two mutational signatures to the 30 COSMIC mutational signatures (http://cancer.sanger.ac.uk/cosmic/signatures) and to the 96-channel centroid of six small intestinal ASCs from two old mice that was published previously (Behjati et al. 2014). These analyses were performed with MutationalPatterns (Blokzijl et al. 2018).

Quantification of the contribution of COSMIC mutational signatures to the 96-channel mutational profiles

We estimated the contribution of the 30 COSMIC mutational signatures (http://cancer.sanger.ac.uk/cosmic/signatures) to the centroids of each mouse ASC group and to the 96-channel mutational profiles of the human organoids using MutationalPatterns (Blokzijl et al. 2018) (Supplemental Fig. S6B, Supplemental Fig. S10B). We ranked the COSMIC signatures based on the total contribution of these signatures to the centroids of the mouse samples. Next, we iteratively reconstructed the centroids of the ASC groups, first using the top 2 COSMIC signatures, and in each iteration the next COSMIC signature was included until all 30 signatures were used. The cosine similarity was calculated between the original and the reconstructed centroid for each mouse ASC group (Supplemental Fig. S6A). As expected, the addition of more signatures increases the similarity of the reconstructed centroids with the original centroids, but after 10 COSMIC signatures the cosine similarities plateau (Supplemental Fig. S6A). Therefore, we used the signature contribution with this subset of 10 COSMIC signatures to the centroids of the four ASC groups (Fig. 3B-C).

Determination of the statistical significance of differences in signature contributions

A bootstrap resampling - similar to that performed in (Zou et al. 2018) - was applied to generate 7,000 replicas of the 96-channel mutational profile of each WT liver ASC (n = 3), which yielded 21,000 WT liver replicas in total. Subsequently, 3 replicas were randomly selected and the relative contribution of 30 COSMIC signatures was determined for their centroid. Euclidean distance d_{WT} was calculated between the relative signature contributions of the replicas centroid and that of the original centroid. This was repeated 10,000 times to construct a distribution of d_{WT} (Supplemental Figure 6C). Next, the threshold distance with *P* value = 0.05, $d_{WT\_0.05}$, was identified. The same approach was taken to generate 7,000 replicas of each Ercc1$^{-/\Delta}$ (MUT) liver ASC (n = 3) and construct a distribution of d_{MUT} (Supplemental Figure 6C). The Euclidean distance *d* between the relative signature contributions of the original WT and Ercc1$^{-/\Delta}$ liver centroids were considered to be significantly different when $d > d_{MUT}$ *and* $d > d_{WT}$. Similarly, bootstrap distributions were generated for WT and Ercc1$^{-/\Delta}$ (MUT) small intestine (Supplemental Figure 6D), with the exception that for the generation of the d_{MUT} distribution only 2 replicas were randomly selected in each permutation, as there are only 2 WT small intestinal ASC samples in the original set. Finally, we repeated the same analyses for the relative contributions of the subset of 10 COSMIC signatures for both liver (Supplemental Figure 6E) and small intestine (Supplemental Figure 6F).

Enrichment or depletion of base substitutions in genomic regions

To test whether the base substitutions appear more or less frequently than expected in genes, promoters, promoter-flanking, and enhancer regions, we loaded the UCSC Known Genes tables as TxDb objects for Mm10 (Team BC and Maintainer 2016) and Hg19 (Carlson and Maintainer 2015), and the regulatory features for Mm10 and Hg19 from Ensembl using biomaRt (Durinck et al. 2005, 2009). We tested for enrichment or depletion of base substitutions in the genomic regions per ASC group (*Ercc1*$^{-/\Delta}$ liver, *Ercc1*$^{-/\Delta}$ small intestine, WT liver, WT small intestine, *XPC*^{KO} and *XPC*^{WT}) using a one-sided Binomial test with MutationalPatterns (Blokzijl et al. 2018)**,** which corrects for the surveyed genomic areas (Supplemental Fig. S9A, Supplemental Fig. S10C). Two-sided Poisson tests were performed to test for significant differences in the ratio of base substitutions within a genomic region divided by the total number of base substitutions between (1) mouse WT and *Ercc1*$^{-/\Delta}$ liver ASCs, (2) mouse WT and *Ercc1*$^{-/\Delta}$ small intestinal ASCs, and (3) human *XPC*^{KO} and human *XPC*^{WT} ASCs (Supplemental Fig. S9A, Supplemental Fig. S10C). Within species, differences in mutation rates with $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing correction) were considered significant.

To test whether base substitutions occur more frequently in more highly expressed genes in the NER-deficient mouse ASCs, we first selected base substitutions that occurred within genes in the mouse ASCs. Per ASC group, we next determined the average Reads Per Kilobase per Million mapped reads (RPKM) of these genes. Two-sided *t*-tests were performed to test for significant difference in the average expression of genes that carry a somatic mutation between (1) mouse WT and *Ercc1*$^{-/\Delta}$ liver ASCs, and (2) mouse WT and *Ercc1*$^{-/\Delta}$ small intestinal ASCs (Supplemental Fig. S9B). Differences in gene expression distributions with $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing correction) were considered significant.

Transcriptional strand bias of base substitutions

For the base substitutions within genes we determined whether the mutations are located on the transcribed or the non-transcribed strand. To this end, we determined whether the mutated "C" or "T" base is on the same strand as the gene definition, which is untranscribed, or the opposite strand, which is transcribed. We generated a 192-channel mutational profile per ASC group with the relative contribution of each mutation type with separate bars for the mutations on the transcribed and untranscribed strand, and calculated the significance of the strand bias using a two-sided Poisson test with MutationalPatterns (Supplemental Fig. S9C, Supplemental Fig. S10D) (Blokzijl et al. 2018). Furthermore, we performed two-sided Poisson tests to test whether there is a significant difference in strand bias per mutation type between (1) mouse WT and *Ercc1*$^{-/\Delta}$ liver ASCs, (2) mouse WT and *Ercc1*$^{-/\Delta}$ small intestinal ASCs, and (3) human *XPC*^{KO} and human *XPC*^{WT} ASCs (Supplemental Fig. S9C, Supplemental Fig. S10D). Within species,

differences in strand bias with an adjusted P-value $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing correction) were considered significant.

Calculation and comparison of mutation rates

To calculate the mutation rates per genome per week, we quantified the number of somatic base substitutions, double nucleotide mutations, indels, and SVs for each mouse ASC. Moreover, we quantified the number of base substitutions, double base substitutions and Signature 8 mutations for the human ASCs. All event counts were extrapolated to the entire autosomal genome using the callable genome length (see" Callable genome") for both mouse and human ASCs to correct for differences in the surveyed genome. Subsequently, the mutation rates were calculated by dividing the extrapolated number of mutations by the number of weeks in which the mutations were accumulated (WT and *Ercc1*$^{-/\Delta}$ mouse organoids: 16 weeks (15 weeks during life and 1 week *in vitro*); *XPC*^{WT} human organoids: 20.6 weeks; *XPC*^{KO} human organoids 10.3 weeks). To determine the proportion of additionally accumulated mutations in the *XPC*^{KO} culture that can be attributed to Signature 8 in human ASCs, we first calculated the increase in base substitutions and the increase in Signature 8 mutations of *XPC*^{KO} compared to *XPC*^{WT}1, *XPC*^{WT}2, and *XPC*^{WT}3 separately. We then divided the increase in Signature 8 mutations by the total increase in base substitutions.

Two-tailed *t*-tests were performed to determine whether the mutation rates differ significantly between (1) mouse WT and *Ercc1*$^{-/\Delta}$ liver ASCs, and (2) mouse WT and *Ercc1*$^{-/\Delta}$ small intestinal ASCs. Of note, these tests assume that the data is normally distributed. Differences in mutation rates between *Ercc1*$^{-/\Delta}$ and WT mouse ASCs with $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing correction) were considered significant.

DATA ACCESS

The sequencing data of the mouse samples have been deposited at the European Nucleotide Archive under accession number ERP021379. The sequencing data of the human samples have been deposited at the European Genome-Phenome archive under accession numbers EGAS00001001682 and EGAS00001002681. Filtered VCF files are freely available at https://wgs11.op.umcutrecht.nl/NERdeficiency/.

DISCLOSURE DECLARATION

The authors have nothing to disclose.

REFERENCES

Aboussekhra A, Biggerstaff M, Shivji MK, Vilpo JA, Moncollin V, Podust VN, Protić M, Hübscher U, Egly JM, Wood RD. 1995. Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell* **80**: 859–868.

Adams PD, Jasper H, Lenhard Rudolph K. 2015. Aging-Induced Stem Cell Mutations as Drivers for Disease and Cancer. *Cell Stem Cell* **16**: 601–612.

Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, Stratton MR. 2015. Clock-like mutational processes in human somatic cells. *Nat Genet* **47**: 1402–1407.

Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, Totoki Y, Fujimoto A, Nakagawa H, Shibata T, et al. 2016. Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**: 618–622.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, et al. 2013. Signatures of mutational processes in human cancer. *Nature* **500**: 415–421.

Al-Minawi AZ, Saleh-Gohari N, Helleday T. 2008. The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. *Nucleic Acids Res* **36**: 1–9.

Amable L. 2016. Cisplatin resistance and opportunities for precision medicine. *Pharmacol Res* **106**: 27–36.

Barker N, Ridgway RA, van Es JH, van de Wetering

M, Begthel H, van den Born M, Danenberg E, Clarke AR, Sansom OJ, Clevers H. 2009. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**: 608–611.

Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, Martincorena I, Petljak M, Alexandrov LB, Gundem G, et al. 2014. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**: 422–425.

Bergeron F, Auvré F, Radicella JP, Ravanat J-L. 2010. HO* radicals induce an unexpected high proportion of tandem base lesions refractory to repair by DNA glycosylases. *Proc Natl Acad Sci U S A* **107**: 5528–5533.

Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, et al. 2016. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**: 260–264.

Blokzijl F, Janssen R, van Boxtel R, Cuppen E. 2018. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* **10**: 33.

Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E. 2012. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**: 423–425.

Bowden NA. 2014. Nucleotide excision repair: Why is it not used to predict response to platinum-based chemotherapy? *Cancer Lett* **346**: 163–171.

Cadet J, Ravanat J-L, TavernaPorro M, Menoni H, Angelov D. 2012. Oxidatively generated complex DNA damage: tandem and clustered lesions. *Cancer Lett* **327**: 5–15.

Carlson M, Maintainer BP. 2015. *TxDb.Hsapiens. UCSC.hg19.knownGene: Annotation package for TxDb object(s)*.

Casorelli I, Tenedini E, Tagliafico E, Blasi MF, Giuliani A, Crescenzi M, Pelosi E, Testa U, Peschle C, Mele L, et al. 2006. Identification of a molecular signature for leukemic promyelocytes and their normal counterparts: Focus on DNA repair genes. *Leukemia* **20**: 1978–1988.

Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, Ramakrishna M, Martin S, Boyault S, Sieuwerts AM, et al. 2017. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med* **23**: 517–525.

Degtyareva NP, Heyburn L, Sterling J, Resnick MA, Gordenin DA, Doetsch PW. 2013. Oxidative stress-induced mutagenesis in single-strand DNA occurs primarily at cytosines and is DNA polymerase zeta-dependent only for adenines and guanines. *Nucleic Acids Res* **41**: 8995–9005.

de Laat W. 1998. Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucleic Acids Res* **26**: 4146–4152.

Dollé MET, Busuttil RA, Garcia AM, Wijnhoven S, van Drunen E, Niedernhofer LJ, van der Horst G, Hoeijmakers JHJ, van Steeg H, Vijg J. 2006. Increased genomic instability is not a prerequisite for shortened lifespan in DNA repair deficient mice. *Mutat Res* **596**: 22–35.

Dollé MT, Kuiper R, Roodbergen M, Robinson J, de Vlugt S, Wijnhoven SP, Beems RB, de la Fonteyne L, de With P, van der Pluijm I, et al. 2011. Broad segmental progeroid changes in short-lived Ercc1 −/Δ7 mice. *Pathobiology of Aging & Age-related Diseases* **1**: 7219.

Drost J, van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, de Ligt J, Behjati S, Grolleman JE, van Wezel T, et al. 2017. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science*. http://dx.doi.org/10.1126/science.aao3130.

Dupuy A, Sarasin A. 2015. DNA damage and gene therapy of xeroderma pigmentosum, a human DNA repair-deficient disease. *Mutat Res* **776**: 2–8.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**: 3439–3440.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**: 1184–1191.

Gregg SQ, Gutiérrez V, Robinson AR, Woodell

T, Nakao A, Ross MA, Michalopoulos GK, Rigatti L, Rothermel CE, Kamileri I, et al. 2012. A mouse model of accelerated liver aging caused by a defect in DNA repair. *Hepatology* **55**: 609–621.

Hoeijmakers JHJ. 2009. DNA damage, aging, and cancer. *N Engl J Med* **361**: 1475–1485.

Huang MN, Yu W, Teoh WW, Ardin M, Jusakul A, Ng AWT, Boot A, Abedi-Ardekani B, Villar S, Myint SS, et al. 2017. Genome-scale mutational signatures of aflatoxin in cells, mice, and human tumors. *Genome Res* **27**: 1475–1486.

Huch M, Gehart H, van Boxtel R, Hamer K, Blokzijl F, Verstegen MMA, Ellis E, van Wenum M, Fuchs SA, de Ligt J, et al. 2015. Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**: 299–312.

Iyama T, Wilson DM 3rd. 2013. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair* **12**: 620–636.

Jager M, Blokzijl F, Sasselli V, Boymans S, Besselink N, Janssen R, Clevers H, van Boxtel R, Cuppen E. 2018. Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures. *Nat Protoc* **13**: 59.

Kamiya H, Murata-Kamiya N, Koizume S, Inoue H, Nishimura S, Ohtsuka E. 1995. 8-Hydroxyguanine (7,8-dihydro-8-oxoguanine) in hot spots of the c-Ha-ras gene: effects of sequence contexts on mutation spectra. *Carcinogenesis* **16**: 883–889.

Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Kwiatkowski DJ, Rosenberg JE, Van Allen EM, D'Andrea A, Getz G. 2016. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat Genet* **48**: 600–606.

Kirschner K, Melton DW. 2010. Multiple roles of the ERCC1-XPF endonuclease in DNA repair and resistance to anticancer drugs. *Anticancer Res* **30**: 3223–3232.

Kuijk EW, Rasmussen S, Blokzijl F, Huch M, Gehart H, Toonen P, Begthel H, Clevers H, Geurts AM, Cuppen E. 2016. Generation and characterization of rat liver stem cell lines and their engraftment in a rat model of liver failure. *Sci Rep* **6**. http://dx.doi.org/10.1038/srep22154.

Lee D-H. 2002. Oxidative DNA damage induced by copper and hydrogen peroxide promotes CG->TT tandem mutations at methylated CpG dinucleotides in nucleotide excision repair-deficient cells. *Nucleic Acids Res* **30**: 3566–3573.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li Q, Yu JJ, Mu C, Yunmbam MK, Slavsky D, Cross CL, Bostick-Bruton F, Reed E. 2000. Association between the level of ERCC-1 expression and the repair of cisplatin-induced DNA damage in human ovarian cancer cells. *Anticancer Res* **20**: 645–652.

Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, Kwon M, Sherman MA, Vitzhum CM, Luquette LJ, Yandava C, et al. 2017. Aging and neurodegeneration are associated with increased mutations in single human neurons. http://dx.doi.org/10.1101/221960.

Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. 2014. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol* **15**: 465–481.

Melis JPM, P.M. Melis J, W.P. Wijnhoven S, Beems RB, Roodbergen M, van den Berg J, Moon H, Friedberg E, van der Horst GTJ, H.J. Hoeijmakers J, et al. 2008. Mouse Models for Xeroderma Pigmentosum Group A and Group C Show Divergent Cancer Phenotypes. *Cancer Res* **68**: 1347–1353.

Naipal KAT, Raams A, Bruens ST, Brandsma I, Verkaik NS, Jaspers NGJ, Hoeijmakers JHJ, van Leenders GJLH, Pothof J, Kanaar R, et al. 2015. Attenuated XPC Expression Is Not Associated with Impaired DNA Repair in Bladder Cancer. *PLoS One* **10**: e0126029.

Niedernhofer LJ, Garinis GA, Raams A, Lalai AS, Robinson AR, Appeldoorn E, Odijk H, Oostendorp R, Ahmad A, van Leeuwen W, et al. 2006. A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature* **444**: 1038–1043.

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M,

Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, et al. 2016. Landscape of somatic mutations in 560 breast cancer whole genome sequences. *Nature* **534**: 47.

Ni M, Zhang W-Z, Qiu J-R, Liu F, Li M, Zhang Y-J, Liu Q, Bai J. 2014. Association of ERCC1 and ERCC2 polymorphisms with colorectal cancer risk in a Chinese population. *Sci Rep* **4**. http://dx.doi.org/10.1038/srep04112.

Olaussen KA, Dunant A, Fouret P, Brambilla E, André F, Haddad V, Taranchon E, Filipits M, Pirker R, Popper HH, et al. 2006. DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *N Engl J Med* **355**: 983–991.

Perera D, Poulos RC, Shah A, Beck D, Pimanda JE, Wong JWH. 2016. Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. *Nature* **532**: 259–263.

Petljak M, Alexandrov LB. 2016. Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis* **37**: 531–540.

Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin M-L, Ordóñez GR, Bignell GR, et al. 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**: 191–196.

Puumalainen M-R, Rüthemann P, Min J-H, Naegeli H. 2015. Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation. *Cell Mol Life Sci* **73**: 547–566.

Rahn JJ, Adair GM, Nairn RS. 2010. Multiple roles of ERCC1-XPF in mammalian interstrand crosslink repair. *Environ Mol Mutagen* **51**: 567–581.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**: i333–i339.

Sands AT, Abuin A, Sanchez A, Conti CJ, Bradley A. 1995. High susceptibility to ultraviolet-induced carcinogenesis in mice lacking XPC. *Nature* **377**: 162–165.

Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, van Es JH, Abo A, Kujala P, Peters PJ, et al. 2009. Single Lgr5 stem cells build crypt□villus structures in vitro without a mesenchymal niche. *Nature* **459**: 262–265.

Schuster-Böckler B, Lehner B. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**: 504–507.

Sijbers AM, de Laat WL, Ariza RR, Biggerstaff M, Wei Y-F, Moggs JG, Carter KC, Shell BK, Evans E, de Jong MC, et al. 1996a. Xeroderma Pigmentosum Group F Caused by a Defect in a Structure-Specific DNA Repair Endonuclease. *Cell* **86**: 811–822.

Sijbers AM, van der Spek PJ, Odijk H, van den Berg J, van Duin M, Westerveld A, Jaspers NG, Bootsma D, Hoeijmakers JH. 1996b. Mutational analysis of the human nucleotide excision repair gene ERCC1. *Nucleic Acids Res* **24**: 3370–3380.

Stubbert LJ, Smith JM, McKay BC. 2010. Decreased transcription-coupled nucleotide excision repair capacity is associated with increased p53- and MLH1-independent apoptosis in response to cisplatin. *BMC Cancer* **10**: 207.

Su Y, Orelli B, Madireddy A, Niedernhofer LJ, Scharer OD. 2012. Multiple DNA Binding Domains Mediate the Function of the ERCC1-XPF Protein in Nucleotide Excision Repair. *J Biol Chem* **287**: 21846–21855.

Team BC, Maintainer BP. 2016. *TxDb.Mmusculus.UCSC.mm10.knownGene: Annotation package for TxDb object(s)*.

Tripsianes K, Folkers G, Ab E, Das D, Odijk H, Jaspers NGJ, Hoeijmakers JHJ, Kaptein R, Boelens R. 2005. The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure* **13**: 1849–1858.

Van Allen EM, Mouw KW, Kim P, Iyer G, Wagle N, Al-Ahmadie H, Zhu C, Ostrovnaya I, Kryukov GV, O'Connor KW, et al. 2014. Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* **4**: 1140–1153.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**: 11.10.1–33.

Vermeij WP, Dollé MET, Reiling E, Jaarsma D, Payan-Gomez C, Bombardieri CR, Wu H, Roks AJM, Botter SM, van der Eerden BC, et al. 2016. Restricted diet delays accelerated ageing and genomic stress in DNA-repair-deficient mice. *Nature* **537**: 427–431.

Waddell N, Pajic M, Patch A-M, Chang DK, Kassahn KS, Bailey P, Johns AL, Miller D, Nones K, Quek K, et al. 2015. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**: 495–501.

Weeda G, Donker I, de Wit J, Morreau H, Janssens R, Vissers CJ, Nigg A, van Steeg H, Bootsma D, Hoeijmakers JHJ. 1997. Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. *Curr Biol* **7**: 427–439.

Zhang R, Jia M, Xue H, Xu Y, Wang M, Zhu M, Sun M, Chang J, Wei Q. 2017. Genetic variants in ERCC1 and XPC predict survival outcome of non-small cell lung cancer patients treated with platinum-based therapy. *Sci Rep* **7**: 10702.

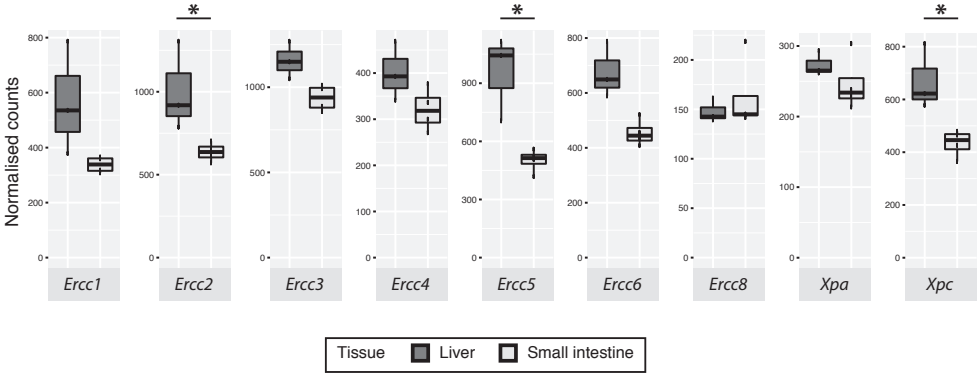
Zheng CL, Wang NJ, Chung J, Moslehi H, Sanborn JZ, Hur JS, Collisson EA, Vemula SS, Naujokas A, Chiotti KE, et al. 2014. Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell Rep* **9**: 1228–1234.

Zhu L, Finkelstein D, Gao C, Shi L, Wang Y, López-Terrada D, Wang K, Utley S, Pounds S, Neale G, et al. 2016. Multi-organ Mapping of Cancer Risk. *Cell* **166**: 1132–1146.e7.

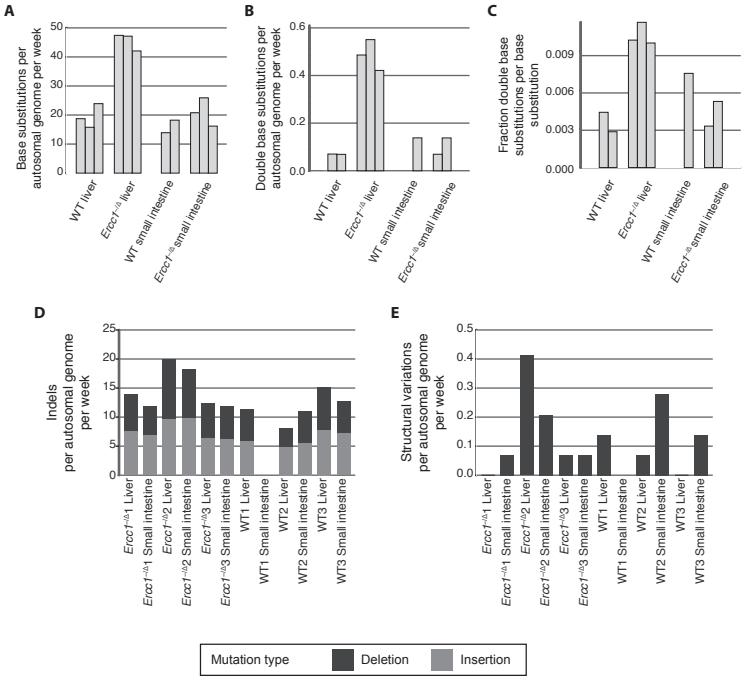
Zou X, Owusu M, Harris R, Jackson SP, Loizou JI, Nik-Zainal S. 2018. Validating the concept of mutational signatures with isogenic cell models. *Nat Commun* **9**: 1744.

SUPPLEMENTAL FIGURES AND TABLES

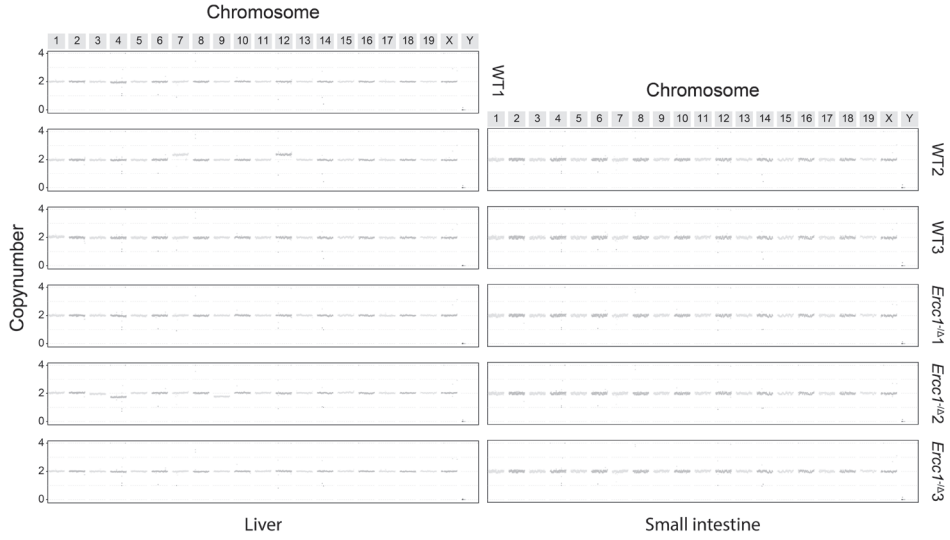
Supplemental files S1 and S2 are available upon request.



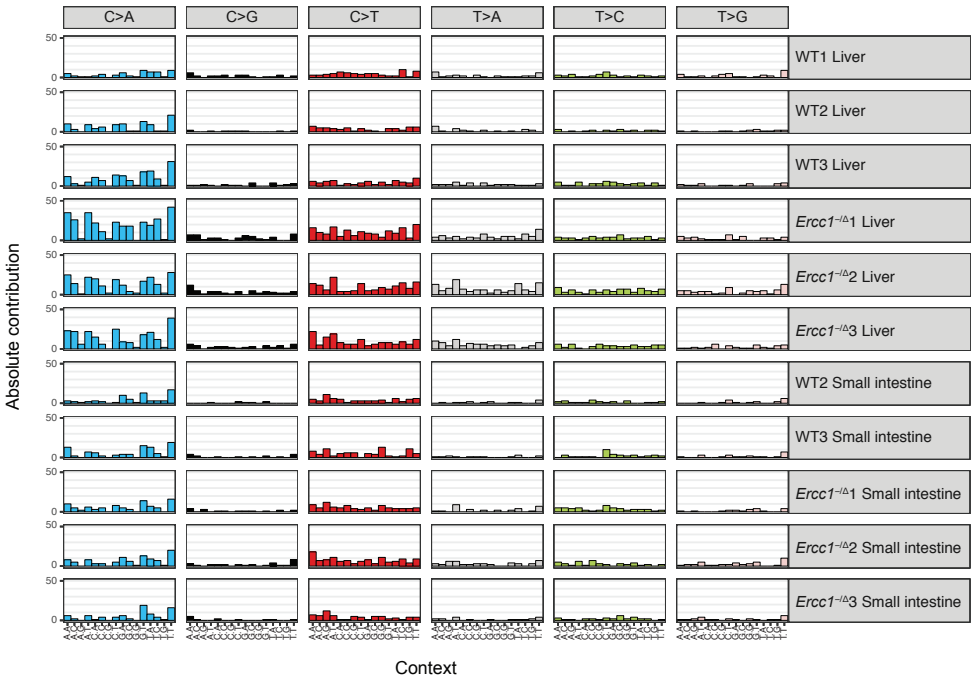
Supplemental figure S1. Boxplots of normalized mRNA counts of 9 core NER genes in WT mouse ASCs from liver (n = 3) and small intestine (n = 4). Asterisks represent significant differential expression (q < 0.05, two-sided t-test, FDR correction).



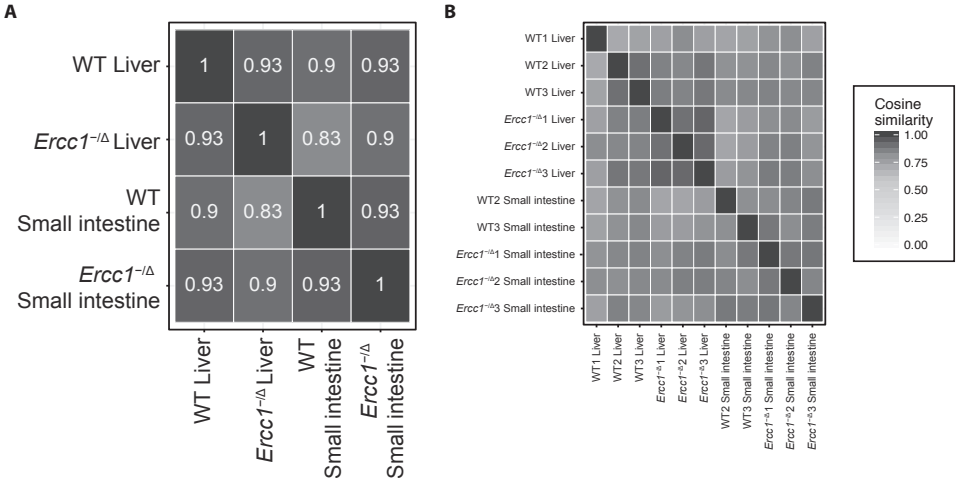
Supplemental figure S2. Somatic mutation rates in the genomes of single ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice. (A) Base substitutions, (B) double base substitutions, (C) fraction double base substitutions/base substitutions, (D) indels, and (E) SVs acquired per autosomal genome per week in single ASCs from WT liver, *Ercc1*$^{-/\Delta}$ liver, WT small intestine, and *Ercc1*$^{-/\Delta}$ small intestine.



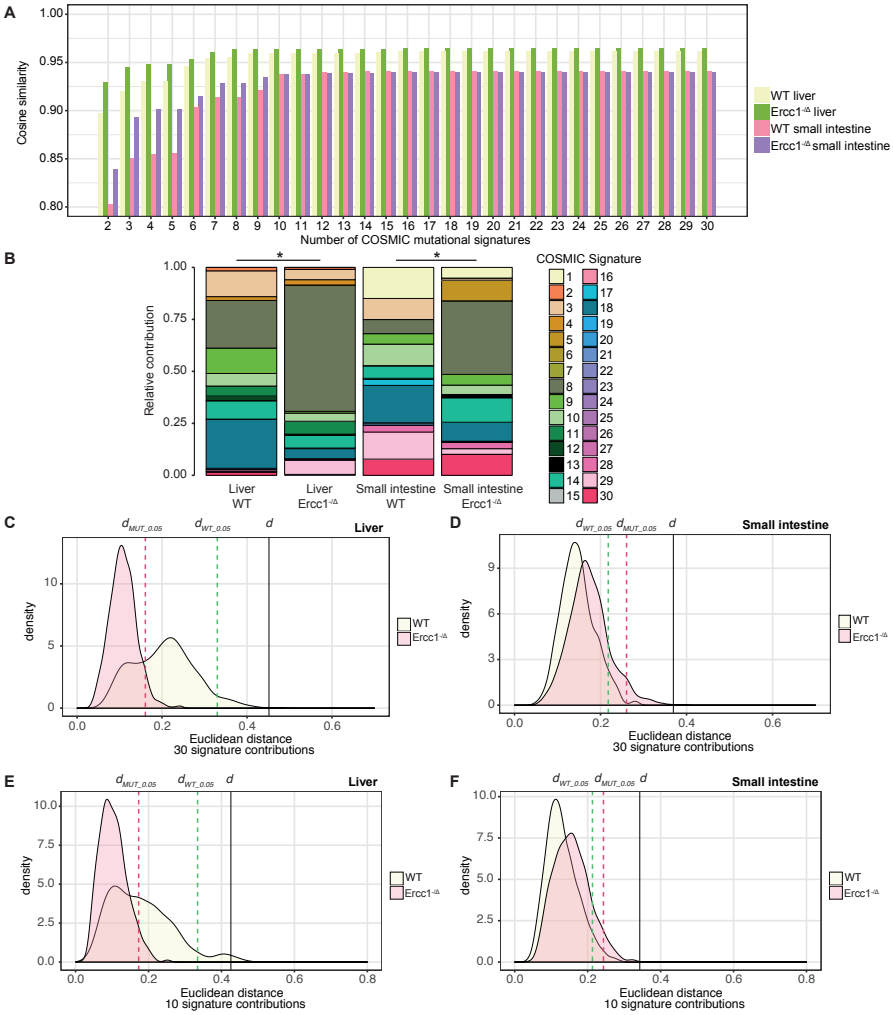
Supplemental figure S3. Genome-wide copy-number profiles of single ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice.



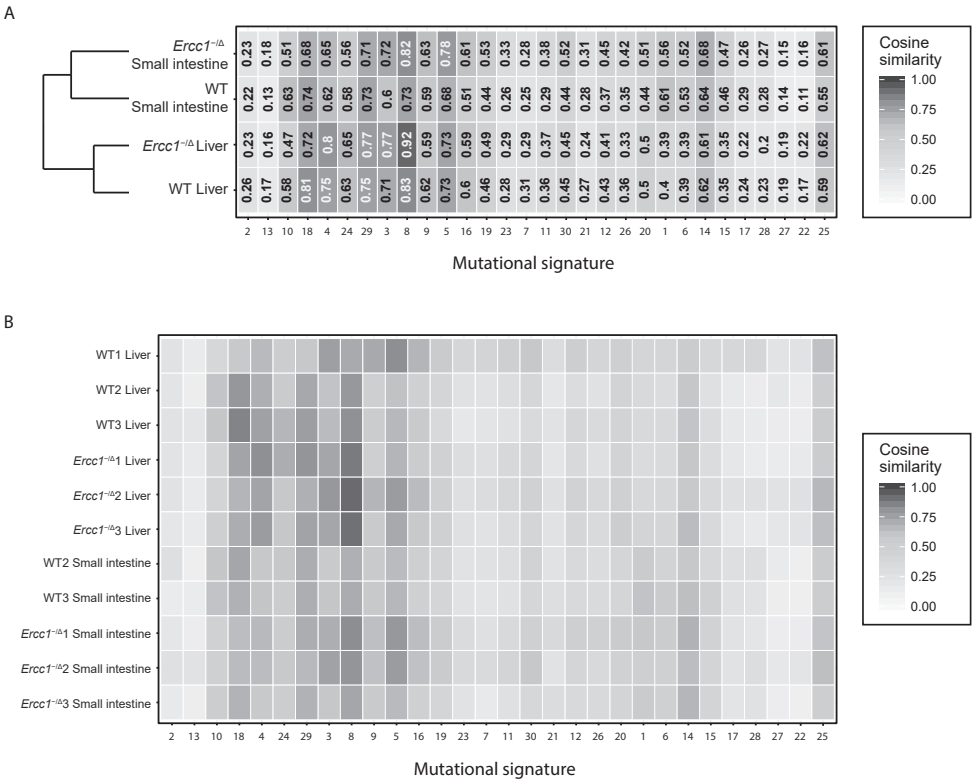
Supplemental figure S4. Absolute contribution of each indicated context-dependent base substitution type to the mutational profiles of ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice.



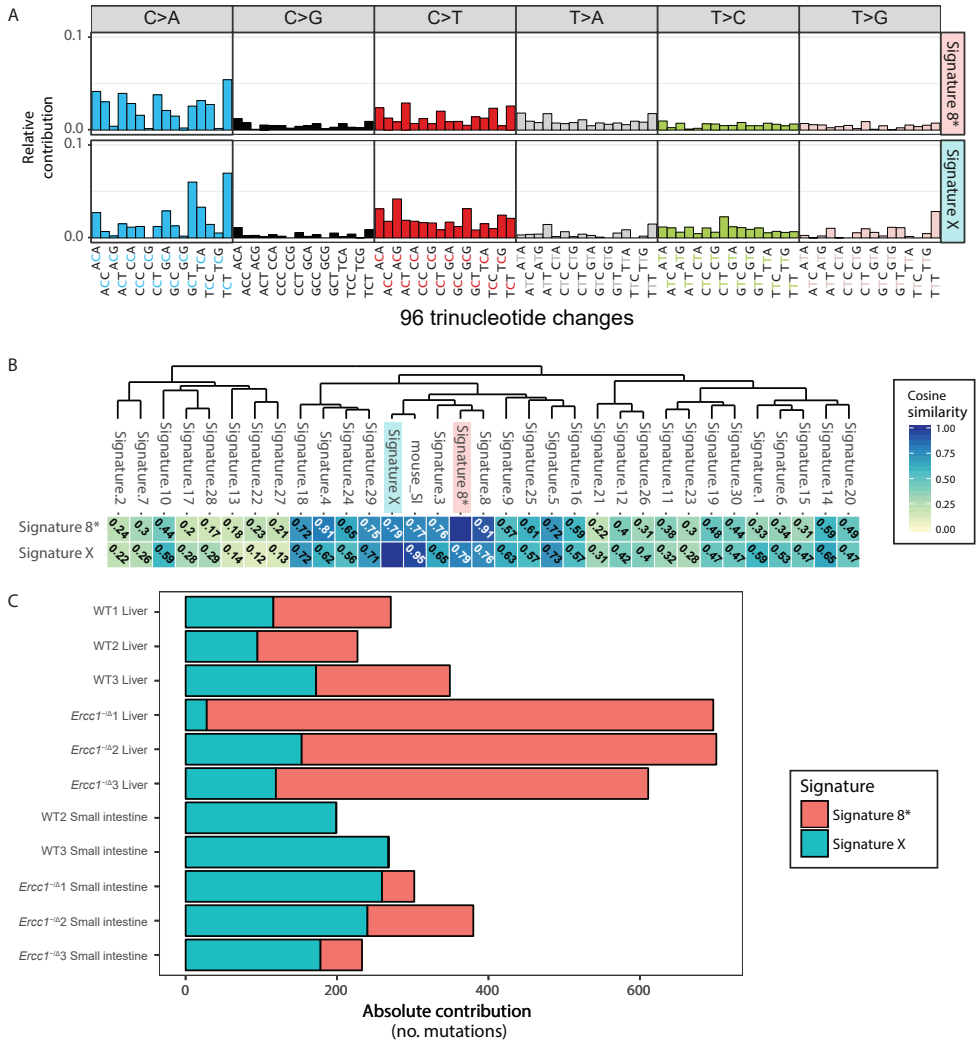
Supplemental figure S5. Similarity between mutational profiles. (A) Cosine similarity between the mutational profiles of all indicated mouse ASC groups (B) Cosine similarity between the mutational profiles of all mouse ASCs.



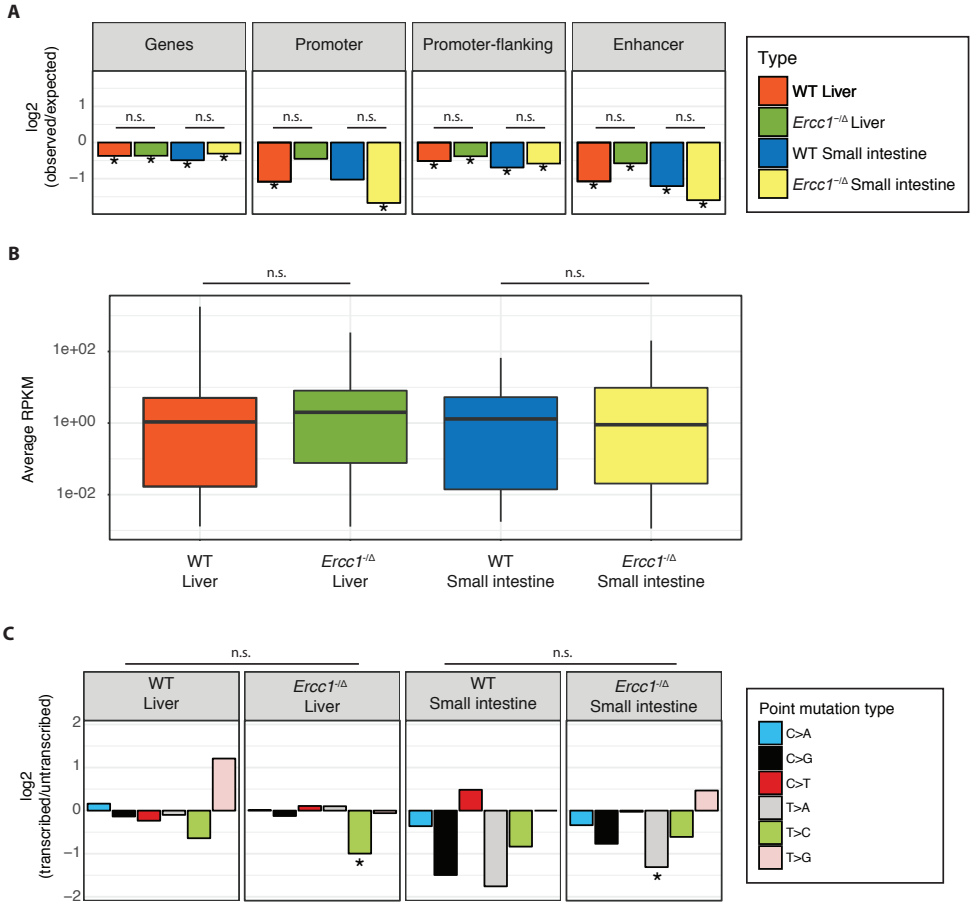
Supplemental figure S6. (A) Cosine similarity between the original centroids and those that were reconstructed using the indicated number of COSMIC signatures for each ASC group. (B) Relative contribution of the 30 COSMIC mutational signatures to the centroids of each ASC group. Asterisks indicate significantly different signature contributions, as determined using the bootstrap distributions depicted in c and d. (C) A bootstrap resampling approach was taken to construct a population of WT and *Ercc1*$^{-/\Delta}$ samples. Subsequently, 3 replicas were randomly selected and the relative signature contributions were calculated for the centroid of the replicas. The Euclidean distance was calculated between this contribution vector and that of the original centroid. This was repeated 10,000 times to construct a distribution of distances d_{WT} for WT (shown in green). Similarly a distribution of d_{MUT} was generated for *Ercc1*$^{-/\Delta}$ (shown in red). The green dashed line indicates the distance where *P* value = 0.05, $d_{WT\_0.05}$, and red dashed line indicates the distance where *P* value = 0.05, $d_{MUT\_0.05}$. The distance between the relative signature contributions of the original WT and *Ercc1*$^{-/\Delta}$ centroids, *d*, is indicated with a black line. The signature contributions of WT and *Ercc1*$^{-/\Delta}$ are considered to be different when $d > d_{MUT\_0.05}$ and $d > d_{WT\_0.05}$. Similarly, bootstrap distributions were generated for small intestine (D), and for relative contributions of a subset of 10 COSMIC signatures for both liver (E) and small intestine (F).



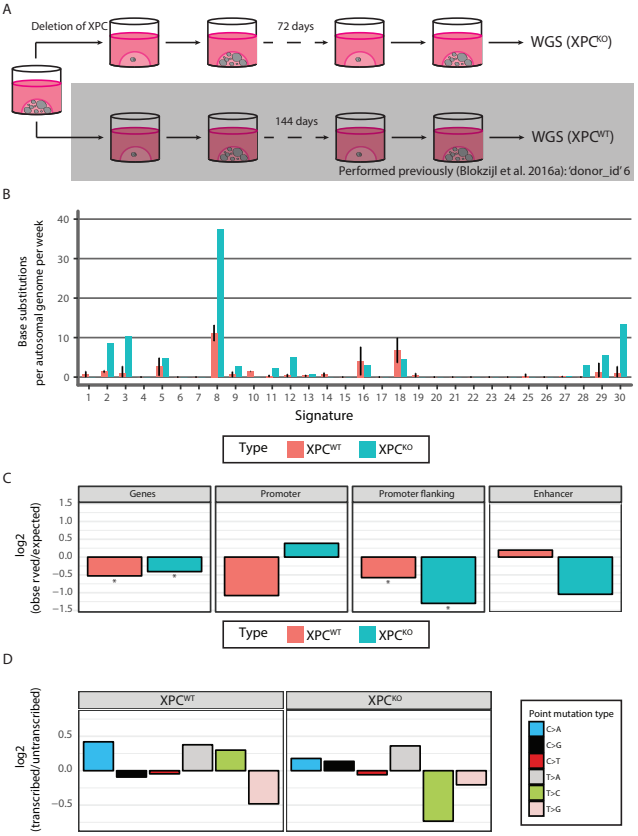
Supplemental figure S7. Similarity between mutational profiles and mutational signatures. Cosine similarity between the mutational profiles of and each COSMIC mutational signatures. (A) per indicated mouse ASC group (B) per mouse ASC. The signatures have been ordered according to hierarchical clustering (complete linkage) using the cosine similarity between signatures, such that similar signatures are displayed close together. The samples are hierarchically clustered (complete linkage) in (A) using the Euclidean distance between the vectors of cosine similarities with the signatures.



Supplemental figure S8. Mutational signatures in mouse adult stem cells (ASCs). (A) Relative contribution of each indicated context-dependent base substitution type to the two mutational signatures that were extracted by non-negative matrix factorization (NMF) of the 96-channel mutational profiles of the mouse ASCs. (B) Cosine similarity of the two mutational signatures depicted in (A) to the 30 current COSMIC mutational signatures and to a centroid mutational profile detected in small intestinal ASCs of old mice (data published previously, Behjati et al. 2014). The signatures have been ordered according to hierarchical clustering (complete linkage) using the cosine similarity between signatures, such that similar signatures are displayed close together. (C) Absolute contribution of the two mutational signatures depicted in (A) to the mutational profiles of the sequenced mouse ASCs.

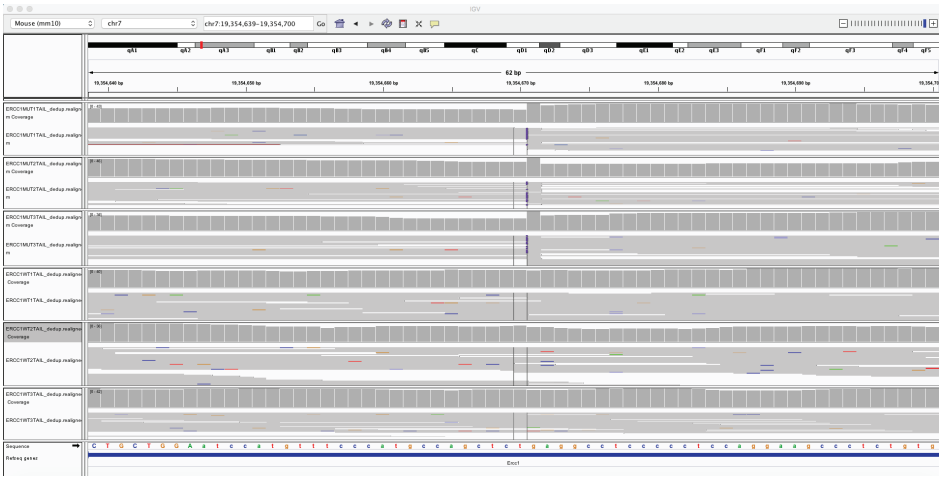


Supplemental figure S9. Genomic distribution of somatic base substitutions the genomes of ASCs from WT liver (n = 3), *Ercc1*$^{-/\Delta}$ liver (n = 3), WT small intestine (n = 2), and *Ercc1*$^{-/\Delta}$ small intestine (n = 3). (A) Depletion of somatic base substitutions in genes, promoter, promoter-flanking regions, and enhancers for each indicated ASC group. The log2 ratio of the number of observed and expected base substitutions indicates the effect size of the depletion in each region. Asterisks represent significant depletions per indicated ASC group ($P < 0.05$, Binomial test, one-sided). n.s. : denotes non-significant differences in depletion between ASC groups ($q \geq 0.05$, Poisson test, two-sided). (B) Boxplots of the Reads Per Kilobase per Million mapped reads (RPKM) values of the genes in which a somatic SNV was detected per ASC group. n.s. : denotes non-significant differences in mean expression levels between ASC groups ($q \geq 0.05$, *t*-test, two-sided). (C) Transcriptional strand bias of base substitutions in genic regions. Log2 ratio of the number of mutations on the transcribed and untranscribed strand per indicated point mutation type for each sample. Asterisks represent significant strand asymmetries per indicated ASC group ($P < 0.05$, Poisson test, two-sided). n.s. : denotes non-significant differences in strand asymmetry between ASC groups ($q \geq 0.05$, Poisson test, two-sided).

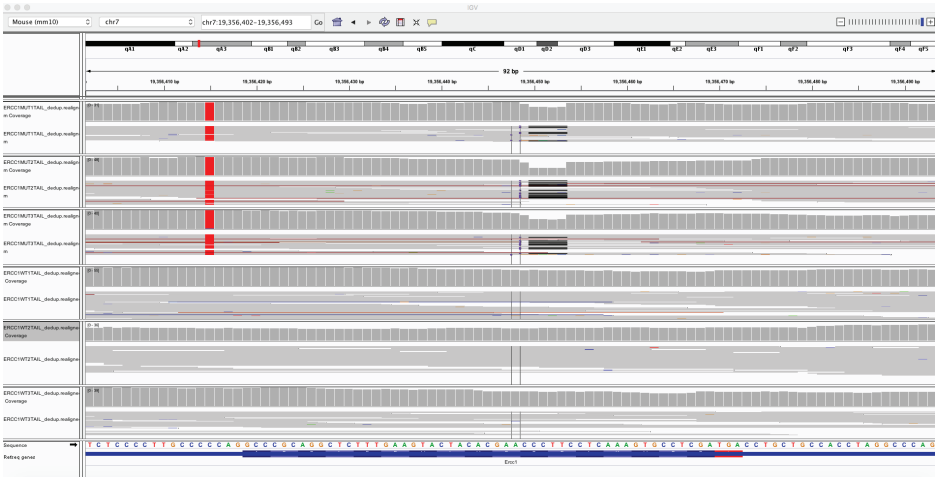


Supplemental figure S10. Mutational consequences of deletion of *XPC* in human ASCs *in vitro*. (A) Schematic overview of the experimental setup used to determine the mutational consequences of KO of *XPC* in single ASCs. A clonal *XPC*^{KO} organoid culture was generated from a human organoid culture through CRISPR-Cas9 gene-editing. These organoids were cultured for 72 days to allow accumulation of sufficient mutations to perform downstream analyses. Subsequently, a subclonal organoid culture was derived from this clonal organoid culture and expanded until there was enough material to perform WGS. As a control sample for filtering germline variants, we used a blood sample that was genome sequenced previously (Blokzijl et al. 2016a). The mutational patterns in the genome of *XPC*^{KO} ASCs were compared to mutational patterns observed previously in *XPC*^{WT} ASCs from the same human donor (Blokzijl et al. 2016a). (B) Contribution of the COSMIC mutational signatures to the mutational profile of *XPC*^{KO} and mean contribution of the COSMIC mutational signatures to the mutational profiles of *XPC*^{WT} ASCs. Error bars represent standard deviations. (C) Depletion/enrichment of base substitutions in genes, promoter, promoter-flanking regions, and enhancers. The log2 ratio of the number of observed and expected number of base substitutions indicates the effect size of the depletion/enrichment in each region. Asterisks represent significant depletions and enrichments per indicated ASC group ($P < 0.05$, Binomial test, one-sided). n.s. : denotes non-significant differences in depletion and/or enrichment between ASC groups ($q \geq 0.05$, Poisson test, two-sided). (D) Transcriptional strand bias of base substitutions in genic regions. Log2 ratio of the number of mutations on the transcribed and untranscribed strand per indicated point mutation type for each sample. Asterisks represent significant strand asymmetries per indicated ASC group ($P < 0.05$, Poisson test, two-sided). n.s. : denotes non-significant differences in strand asymmetry between ASC groups ($q \geq 0.05$, Poisson test, two-sided).

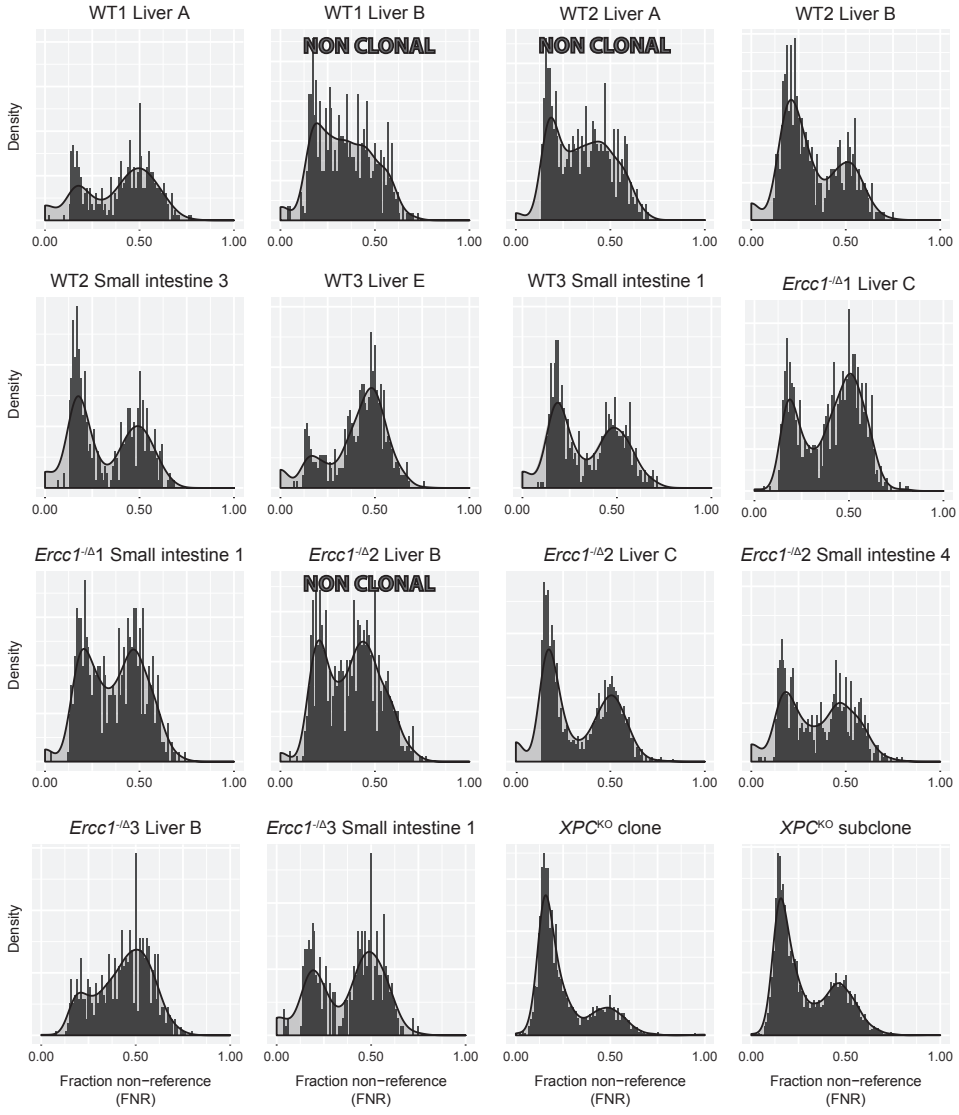
A



B



Supplemental figure S11. IGV screenshots of mutations in the *Ercc1* gene in WT and *Ercc1*$^{-/\Delta}$ mice. (A) *Ercc1*$^{-}$allele and (B) *Ercc1*$^{\Delta}$ allele in the WGS data of the tails of all WT and *Ercc1*$^{-/\Delta}$ mice.



Supplemental figure S12. Distribution plot of the variant allele frequencies (VAFs) of all identified somatic base substitutions that remain before VAF ≥ 0.3 filtering for each ASC.

Ensembl gene ID	Gene symbol	log2FoldChange (WT SI/WT liver)	p-value	Adjusted p-value
ENSMUSG00000026048	*Ercc5*	-0.930	0.000	0.002
ENSMUSG00000030094	*Xpc*	-0.628	0.013	0.040
ENSMUSG00000030400	*Ercc2*	-0.656	0.009	0.040
ENSMUSG00000054051	*Ercc6*	-0.571	0.029	0.052
ENSMUSG00000003549	*Ercc1*	-0.747	0.027	0.052
ENSMUSG00000024382	*Ercc3*	-0.303	0.217	0.304
ENSMUSG00000022545	*Ercc4*	-0.321	0.236	0.304
ENSMUSG00000028329	*Xpa*	-0.151	0.529	0.595
ENSMUSG00000021694	*Ercc8*	0.138	0.601	0.601

SI = small intestine

Supplemental table S1. The log2 fold-change in expression of 9 core NER genes between WT small intestinal ASCs and WT liver ASCs.

Mouse	Tissue	Callable genome (%)	No. Base substitutions*	No. Double base substitutions*	No. Small insertions*	No. Small deletions*	No. Structural variations*
Ercc1 $^{-/\Delta}$1	Liver	90.0%	683	7	111	90	0
Ercc1 $^{-/\Delta}$1	Small intestine	90.2%	300	1	101	72	1
Ercc1 $^{-/\Delta}$2	Liver	90.8%	685	8	142	149	6
Ercc1 $^{-/\Delta}$2	Small intestine	90.7%	376	2	143	122	3
Ercc1 $^{-/\Delta}$3	Liver	89.0%	599	6	92	84	1
Ercc1 $^{-/\Delta}$3	Small intestine	90.0%	233	0	90	80	1
WT1	Liver	90.4%	271	0	86	78	2
WT1	Small intestine	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
WT2	Liver	89.1%	225	1	69	46	1
WT2	Small intestine	89.4%	199	0	79	77	4
WT3	Liver	90.7%	347	1	113	107	0
WT3	Small intestine	90.5%	264	2	107	78	2

* Observed number of mutations within the callable genome

Supplemental table S2. Overview of somatic base substitutions, indels, and structural variations detected in mouse ASCs.

Mouse	Tissue	Chromosome	Position	Type
Ercc1 $^{-/\Delta}$1	Liver	1	19963547-19963548	CC>AT
Ercc1 $^{-/\Delta}$1	Liver	2	75869242-75869243	GG>AA
Ercc1 $^{-/\Delta}$1	Liver	2	97605553-97605554	GC>CT
Ercc1 $^{-/\Delta}$1	Liver	2	151610833-151610834	GG>TT
Ercc1 $^{-/\Delta}$1	Liver	10	49132495-49132496	TC>AA
Ercc1 $^{-/\Delta}$1	Liver	10	55373184-55373185	CT>TA
Ercc1 $^{-/\Delta}$1	Liver	17	3467736-3467737	GG>TT
Ercc1 $^{-/\Delta}$1	SI	17	83387913-83387914	GG>AA
Ercc1 $^{-/\Delta}$2	Liver	3	157520204-157520205	AA>TG
Ercc1 $^{-/\Delta}$2	Liver	5	59202560-59202561	GA>TT
Ercc1 $^{-/\Delta}$2	Liver	5	102337631-102337632	AG>GA
Ercc1 $^{-/\Delta}$2	Liver	6	111610726-111610727	AA>GG
Ercc1 $^{-/\Delta}$2	Liver	9	101601114-101601115	AC>GA
Ercc1 $^{-/\Delta}$2	Liver	10	40895371-40895372	TC>GT
Ercc1 $^{-/\Delta}$2	Liver	11	107811666-107811667	GC>TT
Ercc1 $^{-/\Delta}$2	Liver	14	50134277-50134278	TG>GT
Ercc1 $^{-/\Delta}$2	SI	4	18177691-18177692	AC>TT
Ercc1 $^{-/\Delta}$2	SI	13	54599043-54599044	CC>TT
Ercc1 $^{-/\Delta}$3	Liver	2	54244172-54244173	CC>AT
Ercc1 $^{-/\Delta}$3	Liver	4	39336415-39336416	CA>AC
Ercc1 $^{-/\Delta}$3	Liver	6	116403190-116403191	TC>GA
Ercc1 $^{-/\Delta}$3	Liver	11	70529426-70529427	TC>AA
Ercc1 $^{-/\Delta}$3	Liver	14	63208917-63208918	GC>AA
Ercc1 $^{-/\Delta}$3	Liver	15	14643203-14643204	TC>AA
WT2	Liver	10	107687236-107687237	AG>TT
WT3	Liver	1	13942075-13942076	AG>GT
WT3	SI	5	54402194-54402195	GT>TC
WT3	SI	14	13541219-13541220	CA>AT

Supplemental table S3. Double point mutations acquired in the genomes of WT and *Ercc1*$^{-/\Delta}$ mouse ASCs.

Mouse	Tissue	Chromosome	Start	End	Size (bp)	Type
Ercc1$^{-/\Delta}$1	SI	14	98382845	98383374	529	deletion
Ercc1$^{-/\Delta}$2	Liver	11	4307381	4308024	643	deletion
Ercc1$^{-/\Delta}$2	Liver	11	96366839	96367238	399	deletion
Ercc1$^{-/\Delta}$2	Liver	15	14954694	14961303	6609	deletion
Ercc1$^{-/\Delta}$2	Liver	15	82986523	82989502	2979	deletion
Ercc1$^{-/\Delta}$2	Liver	16	3744900	3745261	361	deletion
Ercc1$^{-/\Delta}$2	Liver	19	25020360	25021085	725	deletion
Ercc1$^{-/\Delta}$2	SI	3	108934215	108934569	354	deletion
Ercc1$^{-/\Delta}$2	SI	4	88438548	88439859	1311	deletion
Ercc1$^{-/\Delta}$2	SI	6	49000048	49000651	603	deletion
Ercc1$^{-/\Delta}$3	Liver	15	98807785	98833375	25590	deletion
Ercc1$^{-/\Delta}$3	SI	5	36712689	36713090	401	deletion
WT1	Liver	4	152179670	152523647	343977	deletion
WT1	Liver	17	52028043	52028582	539	deletion
WT2	Liver	19	14877486	14877950	464	deletion
WT2	SI	4	145065976	145066394	418	deletion
WT2	SI	5	41625866	41635804	9938	deletion
WT2	SI	5	41625866	41687721	61855	deletion
WT2	SI	17	35644289	35644686	397	deletion
WT3	SI	2	84747100	84747615	515	deletion
WT3	SI	6	139571588	139571908	320	deletion

bp = base pairs

Supplemental table S4. SVs acquired in the genomes of WT and *Ercc1*$^{-/\Delta}$ mouse ASCs.

Sample	No. weeks in culture	Callable genome (%)	No. Base substitutions*	No. Double base substitutions*
XPC^{WT}1	20.6	76.9%	467	3
XPC^{WT}2	20.6	75.9%	572	2
XPC^{WT}3	20.6	73.4%	503	2
XPC^{KO}	10.3	88.0%	895	23

* Observed number of mutations within the callable genome

Supplemental table S5. Overview of somatic base substitutions, indels, and structural variations detected in human ASCs.

Sample	Chromosome	Position	Type
XPC^{WT}1	5	28200738-28200739	TG>CA
XPC^{WT}1	14	86428191-86428192	TC>AA
XPC^{WT}1	14	87517131-87517132	CC>AT
XPC^{WT}2	1	218692585-218692586	CA>AT
XPC^{WT}2	8	89594139-89594140	TC>GA
XPC^{WT}3	3	147536195-147536196	AG>GA
XPC^{WT}3	4	29048020-29048021	CC>AA
XPC^{KO}	1	69460060-69460061	GA>AC
XPC^{KO}	1	234938262-234938263	CA>TT
XPC^{KO}	2	7379071-7379072	TA>GT
XPC^{KO}	2	45682863-45682864	TT>AA
XPC^{KO}	2	57337763-57337764	TT>AA
XPC^{KO}	2	98545278-98545279	AC>GA
XPC^{KO}	2	99366427-99366428	TG>AA
XPC^{KO}	2	144580563-144580564	TG>CT
XPC^{KO}	2	171495577-171495578	TC>GA
XPC^{KO}	3	67720855-67720856	TA>AG
XPC^{KO}	3	139045761-139045762	GG>AT
XPC^{KO}	4	136997624-136997625	GG>AA
XPC^{KO}	4	189948523-189948524	TC>AA
XPC^{KO}	6	75466827-75466828	GA>TT
XPC^{KO}	6	104498768-104498769	GT>AA
XPC^{KO}	6	129297075-129297076	GT>AA
XPC^{KO}	8	36646090-36646091	GT>AA
XPC^{KO}	11	119899524-119899525	AC>TT
XPC^{KO}	12	52842015-52842016	AC>GA
XPC^{KO}	13	51476415-51476416	TT>GC
XPC^{KO}	19	9559875-9559876	TC>GA
XPC^{KO}	19	45595513-45595514	AC>TT
XPC^{KO}	22	46424298-46424299	TC>AT

Supplemental table S6. Double point mutations acquired in the genomes of *XPC*^{WT} and *XPC*^{KO} human ASCs.

A fish a day keeps the doctor away,

a wine a day keeps the doctor coming

Chapter 5

Effect of chronic alcohol use on mutation accumulation in precancerous cirrhotic liver adult stem cells

Myrthe Jager[1], Ewart Kuijk[1], Ruby Lieshout[2], Mauro Locati[1], Nicolle Besselink[1], Roel Janssen[1], Sander Boymans[1], Jeroen de Jonge[2], Jan IJzermans[2], Michael Doukas[3], Monique Verstegen[2], Ruben van Boxtel[1,4], Luc van der Laan[2] and Edwin Cuppen[1,#]

1 Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht University, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands
2 Department of Surgery, Erasmus Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands
3 Department of Pathology, Erasmus Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands
4 Princess Máxima Center for Pediatric Oncology, 3584 CT Utrecht, The Netherlands

Manuscript in preparation

ABSTRACT

Alcohol consumption is a risk factor for the development of liver cancer. Nevertheless, the underlying mechanisms remain unclear. Here, we determined the mutational consequences of chronic alcohol use on the genomes of liver stem cells prior to cancer development. Surprisingly, no change in mutation rate or spectrum was observed in these genomes. Analysis of the trunk mutations in an alcohol-related liver tumor by multi-site whole-genome sequencing confirms the absence of alcohol-induced mutational signatures. However, we do see an enrichment of non-silent mutations in cancer genes, including a negative regulator of the EGF-pathway, *PTPRK*. Our results indicate that the carcinogenic effects of alcohol are likely indirect. We speculate that the tissue environment in the cirrhotic liver may provide a fertile ground for cells with oncogenic mutations.

MAIN

Alcohol consumption is an important risk factor for the development of various cancer types, including hepatocellular carcinoma (HCC), and causes an estimated 400,000 cancer-related deaths each year (*1–5*). In spite of the clear link between alcohol and tumorigenesis, the underlying mechanism remains debated. Alcohol consumption might directly drive the development of cancer through an increased mutation accumulation in the genome (*6*). Consistently, the first metabolite of ethanol, acetaldehyde, is highly carcinogenic and can cause mutations such as tandem base substitutions (*7–9*). Acetaldehyde can also contribute to the formation of reactive oxygen species (ROS) (*10*, *11*), which can be mutagenic (*12*, *13*). Furthermore, alcohol consumption has been reported to be associated with a modest increase in specific mutations, Signature 16 mutations, in esophageal and liver cancer (*14–17*). However, the observed increase in the base substitution load due to alcohol consumption in human cancer (*14*, *15*), seems too small to account for the massively enhanced cancer risk. Furthermore, an increased mutation load alone is not sufficient to drive liver cancer in mice (*18*). This suggests that additional, non-mutational factors also play a pivotal role in the link between alcohol consumption and tumor formation (*19*), especially in the liver.

To determine whether alcohol consumption has mutational consequences prior to the development of liver cancer, we sequenced the genomes of eight independent clonal organoid cultures derived from liver biopsies of five cirrhotic, but non-cancerous, livers from patients with a known history of chronic alcohol use that were undergoing a liver transplantation (further referred to as 'alcoholic livers'; Suppl. table S1). To determine and exclude germline variations, we sequenced the blood of all five patients as well. As shown previously, this approach allows the identification

of mutations that accumulated in single adult stem cells (ASCs) during life with high confidence (*20*, *21*). To gain insight into the mutational consequences of alcohol consumption, the somatic mutation catalogs in alcoholic livers were compared to the mutation catalogs that were obtained previously from whole genome sequencing (WGS) data of five healthy liver donors (*21*). Furthermore, five clonal liver organoid cultures derived from four additional healthy liver donors with ages ranging from 24 to 68 were added to these analyses, to increase the number of healthy liver donors and to obtain age-matched healthy controls (Suppl. table S1).

In total, we analyzed 42,093 base substitutions that accumulated in 23 ASCs derived from 9 healthy and 5 alcoholic livers (Suppl. table S1). Consistent with previous observations (*21*), somatic base substitutions accumulate linearly with age in healthy liver ASCs (two-tailed *t*-test, linear mixed model; $P < 0.05$) (Fig. 1A). Healthy liver ASCs acquire ~39.4 (95% confidence interval: 30.5 - 48.3) somatic base substitutions each year. The mutation load in alcohol liver ASCs is similar to, and within the 95% confidence interval of, age-matched healthy liver ASCs (Fig. 1A). In addition, alcohol consumption does not affect the number of tandem base substitutions acquired in the genomes of liver ASCs (Suppl. Fig. S1). This indicates that alcohol consumption does not have a direct effect on the base substitution load in liver ASCs.

Genome-wide patterns of base substitutions reflect activity of mutational processes that have been active in cells (*22*). To identify if excessive alcohol consumption changes these profiles in liver ASCs, we performed in-depth mutational analyses. The mutational profiles of healthy liver ASCs are characterized by a high contribution of C:G > A:T, C:G > T:A, and T:A > C:G mutations (Fig. 1B-C; Suppl. Fig. S2). Interestingly, the mutational profiles of alcoholic liver ASCs are highly similar to the mutational profiles of healthy liver ASCs (cosine similarity = 0.99), suggesting that chronic alcohol use does not affect the mutational processes in liver ASCs. To identify whether any of the known mutational signatures (*22*, *23*) can be associated with alcohol use, we calculated the contribution of the cancer signatures to the mutational profiles of all ASCs and subsequently determined whether the contribution differs between healthy and alcoholic liver ASCs. Signature 5 and Signature 16 can explain the majority of the accumulated mutations in both healthy and alcoholic liver (Suppl. Fig. S3). However, in contrast to previously reported observations in liver tumors (*14*, *15*), we do not observe a significant increase in Signature 16 mutations in alcoholic liver ASCs in comparison to healthy ASCs (*t*-test, FDR correction; n.s.).

In addition to base substitutions, we also determined the accumulation of copy number alterations (CNAs) in the genomes of the liver ASCs. The majority of the healthy liver ASCs did not acquire any CNAs (Suppl. table S1) (*21*). Similarly, none of the assessed alcoholic liver ASCs acquired a CNA (Suppl. table S1). At the

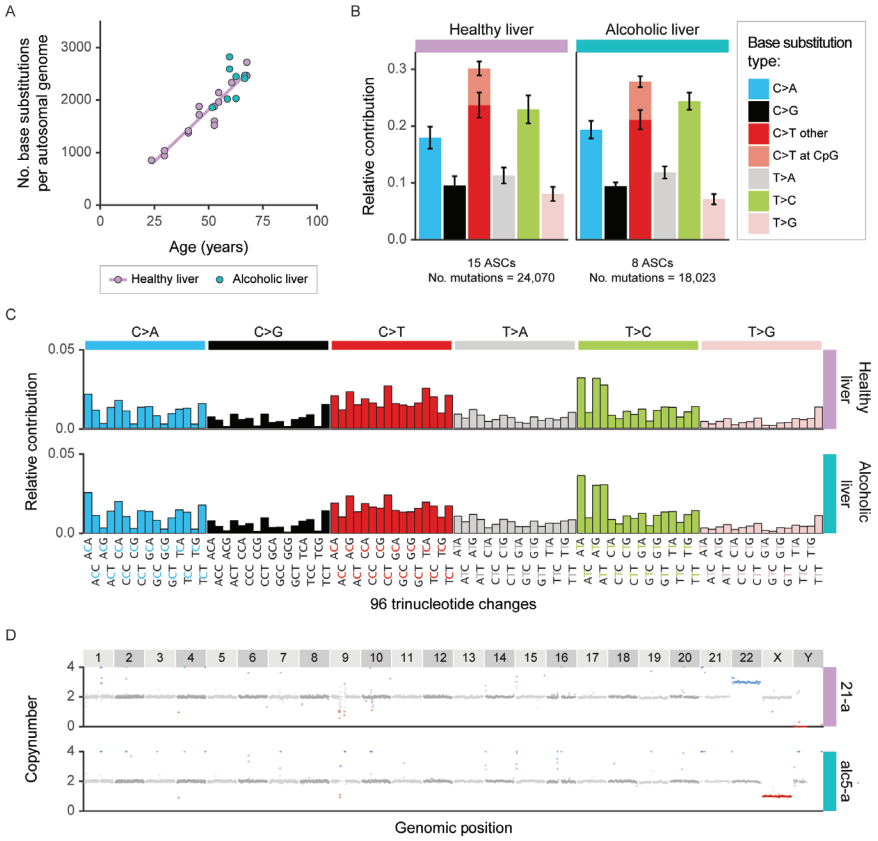


Figure 1. Mutation accumulation in the genomes of healthy and alcoholic liver stem cells. (A) The number of base substitutions acquired in the autosomal genomes of 15 healthy and 8 alcoholic liver ASCs derived from 9 and 5 donors, respectively. Each stem cell is represented by a data point. In the healthy liver, a linear accumulation of mutations with age was observed (two-tailed *t*-test, linear mixed model; $P = 1.56 \times 10^{-5}$), indicated by the purple trendline. (B) Mutation spectra of the accumulated base substitutions in the genomes of healthy and alcoholic liver ASCs. Mean relative contribution of the indicated base substitution types is depicted. Error bars represent standard deviations. Number of mutations is indicated. (C) Relative contribution of the 96 context-dependent base substitution types to the mutational profiles of healthy and alcoholic liver stem cells. Number of mutations as indicated in (B). (D) Genome-wide copy number profiles of ASC 21-a and alc5-a. Red data points indicate a copy number state < 2, grey data points represent a copy number state of 2, and blue data points indicate a copy number state > 2.

chromosomal level, however, we observed a trisomy in a liver ASC from a 68-year-old healthy female and a chromosome Y gain in a liver ASC from a 67-year-old alcoholic male (Fig. 1D). Similar to colon ASCs (*21*), but irrespective of alcohol consumption, this shows that aneuploidies occur in liver ASCs of older individuals. Taken together, these results strongly suggest that chronic alcohol consumption does not contribute to the development of HCC through an altered base substitution or CNA accumulation in the liver prior to oncogenesis.

A possible explanation for the absence of a correlation between alcohol consumption and mutational patterns is that the cells that we have sequenced are too early in the precancerous state. Therefore, we also sequenced five biopsies across a 13 cm HCC of a 60-year-old male with a history of chronic alcohol use and identified mutations that were shared by all biopsies (Fig. 2A; Suppl. Fig. S4). This approach allows the identification of mutational processes that have been active in a most recent common ancestor (MRCA) both prior to tumor formation and in the early stages of tumor development (Fig. 2A). As a control sample, we sequenced a non-tumor biopsy adjacent to the tumor, to identify and exclude germline mutations. In total, we identified 19,200 unique somatic base substitutions across all five HCC biopsies (Suppl. table S2; Suppl. Fig. S4B). Analysis of the mutations shared by all biopsies (i.e. trunk mutations) revealed that the MRCA of these biopsies accumulated 7,203 base substitutions (Suppl. table S2; Suppl. Fig. S4B), two CNVs (Suppl. table

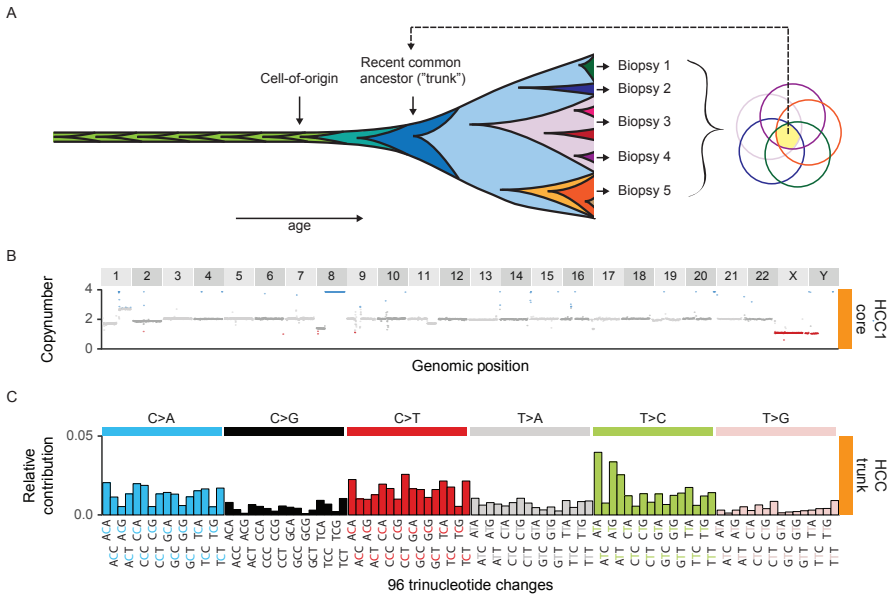


Figure 2. Measuring mutations of a recent common ancestor of an HCC in a patient with a history of alcohol abuse. (A) Clonal sweeps occur during life, both prior to cancer development and after a cell has accumulated sufficient driver mutations to facilitate tumorigenesis. The mutations that accumulated in a recent common ancestor (the "trunk") can be identified, by taking multiple biopsies and identifying the mutations that are shared by all biopsies (yellow). We took five biopsies across a 13 cm HCC to identify the trunk mutations. (B) Genome-wide copy number profile of one of the five HCC biopsies (HCC1-core), which is representative for all HCC biopsies (Suppl. Fig. S4C). Red data points indicate a copy number state < 2, grey data points represent a copy number state of 2, and blue data points indicate a copy number state > 2. (C) Relative contribution of the 96 context-dependent base substitution types to the mutational profile of the 7,203 trunk mutations that accumulated in the recent common ancestor of an HCC.

S3), and several chromosomal aneuploidies (Fig. 2B; Suppl. Fig. S4C). The trunk base substitutions have a variant allele frequency (VAF) of ~0.5 after adjusting for the estimated tumor percentage in each biopsy (Suppl. Fig. S5, Suppl. Fig. S6), underlining that these somatic mutations are clonal in each sample and that the biopsies share a MRCA that accumulated 7,203 heterozygous base substitutions.

The number of CNVs is similar to the observed number of CNVs in healthy liver ASCs (*21*). However, the base substitution load is ~2.5 times higher than expected for a 60-year-old individual and, in contrast to the precancerous ASCs, this MRCA of the HCC already carried several chromosome arm gains and losses (Fig. 2B). To determine whether the increased mutational load is a direct consequence of alcohol consumption, or rather something that is inherently linked to early stages of tumorigenesis we compared the mutation catalog of the HCC trunk mutations to publicly available mutation catalogs of HCC which are predominantly linked to viral hepatitis infection (*16*, *24*). Cancers related to chronic exposure to mutagens, such as UV-light and tobacco, typically show very high mutational loads and very specific mutation types (*22*). If alcohol would contribute to the development of cancer in a similar manner as these mutagens, one would therefore expect a high mutation load in comparison to liver cancers with another cause, such as viral infection, and a notable change in mutational profiles. However, the base substitution load in each biopsy (9,347 - 10,830; Suppl. table S2) is similar to the mutational load in liver cancers which are predominantly caused by HBV and HCV infection (*16*). Furthermore, the mutational profile of the clonal base substitutions in the HCC is highly similar to healthy and alcoholic liver ASCs (Fig. 2C; cosine similarity of 0.97 and 0.98, respectively). Finally, copy number gains and losses, especially of chromosome 1 and 8, are frequently observed in HCCs due to HBV (*24*). These results confirm the initial observations in ASCs that alcohol consumption itself does not introduce mutations in the genome of liver cells.

The high mutational load in the MRCA might reflect that this cell is not the cell-of-origin (Fig. 2A). Potentially, a clonal sweep occurred after a substantial amount of mutations already accumulated in the MRCA. Alternatively, it should be noted that tissue-specific liver ASCs might not be the cell-of-origin for HCC and the mutation rate in the cancer-initiating cells might, in fact, be higher. Nevertheless, as the cancer-initiating cells are exposed to the same mutagenic damage as the liver ASCs (and show the same mutational signatures), our results still point towards an indirect non-mutational mechanism of increased alcohol-induced cancer risk.

In the liver, development of HCC is almost always preceded by chronic inflammation and cirrhosis (*19*), and this extrinsic damage appears to be required for the formation of liver cancer, at least in mice (*18*). This shows that the cirrhotic

microenvironment may play a crucial role in the development of liver cancer. Potentially, the inflamed, cirrhotic microenvironment drives liver cancer by creating a fertile ground for cells with oncogenic driver mutations (*18*, *25*, *26*). To identify whether alcohol consumption induces changes in selection of cells, we next analyzed the genomic distribution of the acquired base substitutions. Although positive selection of cells with driver mutations has been shown to play a role to tumorigenesis, these selective advantages are not numerous and are easily overlooked by analyzing the gross genomic distribution of mutations (*25*). Indeed, base substitutions are significantly depleted in functional genomic regions including genes and enhancers in healthy liver ASCs, alcoholic liver ASCs, and the MRCA of the HCC (Fig. 3A). Furthermore the normalized ratio of non-synonymous to synonymous mutations (d*N*/d*S*) is ~1 in all assessed cell types (Fig. 3B). This suggests that there is no general change in selection against more deleterious mutations. However, we observe a subtle enrichment of potential driver mutations in alcoholic liver ASCs (Fig. 3C; Table 1), although it should be noted that the number of mutations is low.

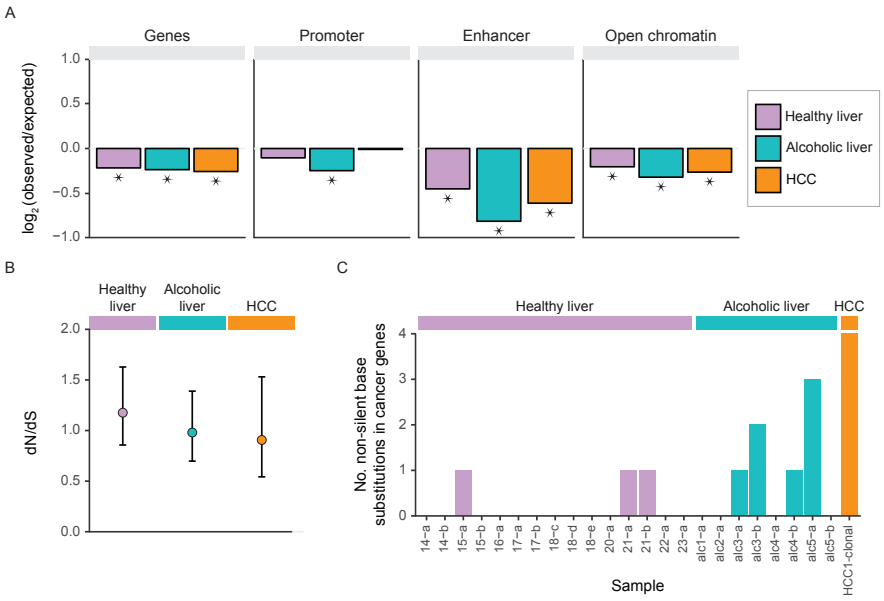


Figure 3. Genome-wide distribution of the somatic base substitutions acquired in the genomes of healthy liver stem cells, alcoholic liver stem cells, and a recent common ancestor of an HCC. (A) The effect size of the depletion of somatic base substitutions in genes, promoters, enhancers, and open chromatin regions for the indicated sample types. Asterisks indicate significant depletion. (B) d*N*/d*S* of the somatic base substitutions in genes in the indicated sample types. Data points represent the Maximum-likelihood estimates and error bars represent the 95% confidence intervals. (C) Number of nonsynonymous and nonsense base substitutions in cancer genes in each indicated sample.

Only one in three healthy liver ASCs has acquired a non-silent base substitution in a cancer census gene that is hit by dominant mutations in cancer (https://cancer.sanger.ac.uk/census). In alcoholic liver ASCs, on the other hand, we observe a total of seven non-silent base substitutions in these cancer genes across 8 ASCs. Two alcoholic liver ASCs even acquired multiple non-silent hits in cancer genes (Fig. 3C; Table 1). It is estimated that only 4 nonsynonymous base substitutions in cancer genes can drive the development of liver cancer (*25*). Consistently, we identified four nonsynonymous mutations in cancer genes in the MRCA of the HCC (Fig. 3C; Table 1). The modest increase in non-silent mutations in cancer genes in alcoholic liver ASCs may, therefore, indicate that these cells are already moving towards development of HCC.

Notably, the cancer gene *PTPRK* is hit by heterozygous nonsense mutations in two independent alcoholic liver ASCs, which is significantly more than expected based on the background mutation rate adjusted for the mutational profile (Table 1; likelihood ratio test, FDR correction; q = 0.02) (*25*). None of the healthy ASCs of the liver, small intestine, or colon, however, acquired a mutation in *PTPRK* (*21*). This enrichment of *PTPRK* mutations in alcoholic liver ASCs indicates that these mutations might provide an outgrowth benefit in alcoholic livers. PTPRK is a receptor-type tyrosine phosphatase, which dephosphorylates tyrosine residues of the EGFR (*27*). Reduced expression of *PTPRK* results in increased EGFR phosphorylation and EGF-signaling, and ultimately enhances cellular proliferation (*27–29*). To identify whether the accumulated heterozygous nonsense mutations in *PTPRK* affect the expression

Gene symbol	Sample type*	Samples	Mutation type	q (nonsynonymous)	q (nonsense)
ATP1A1	Healthy	15-a	Missense	1,00	1,00
PAX7	Healthy	21-a	Missense	1,00	1,00
PREX2	Healthy	21-b	Missense	1,00	1,00
PTPRK	ALC	alc3-a, alc5-a	Nonsense, Nonsense	1,00	0,02
ALK	ALC	alc3-b	Missense	1,00	1,00
CACNA1D	ALC	alc3-b	Missense	1,00	1,00
ZNF331	ALC	alc4-b	Missense	1,00	1,00
CUX1	ALC	alc5-a	Missense	1,00	1,00
TERT	ALC	alc5-a	Missense	1,00	1,00
CD274	HCC	HCC1-clonal	Missense	1,00	1,00
CIITA	HCC	HCC1-clonal	Missense	1,00	1,00
KLF4	HCC	HCC1-clonal	Missense	1,00	1,00
MUC1	HCC	HCC1-clonal	Missense	1,00	1,00

* ALC = alcoholic liver ASC ; Healthy = healthy liver ASC ; HCC = trunk mutations hepatocellular carcinoma

Table 1. Nonsynonymous and nonsense mutations in cancer genes observed in healthy liver ASCs, alcoholic liver ASCs, and the MRCA of an HCC. q values (likelihood ratio test, FDR correction) indicate significant enrichment of non-silent base substitutions within genes.

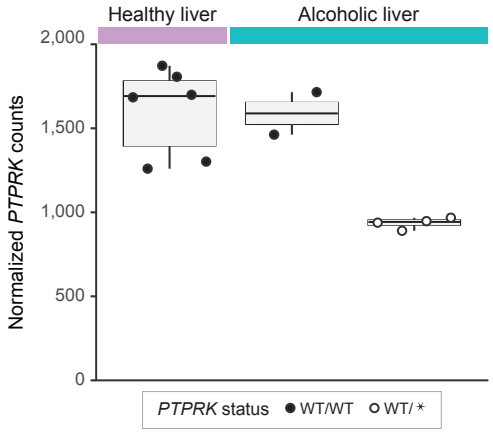


Figure 4. Boxplots of normalized *PTPRK* mRNA counts in organoid cultures of alcoholic $PTPRK^{WT/*}$ livers and healthy and alcoholic $PTPRK^{WT/WT}$ liver cultures. Normalized counts were calculated for duplicate measures of 2 alcoholic $PTPRK^{WT/*}$, 3 healthy $PTPRK^{WT/WT}$, and 1 alcoholic $PTPRK^{WT/WT}$ organoid cultures. Each data point represents a single measurement. WT = wildtype, * = nonsense mutation.

of *PTPRK*, we next performed RNA-sequencing. Indeed, the nonsense mutations resulted in a significantly reduced expression of *PTPRK* (Fig. 4; $P < 0.05$, negative binomial test), indicating a gene dosage effect. Single-nucleotide polymorphisms that cause enhanced EGF-signaling have already been shown to increase the risk of developing HCC (*26*, *30*, *31*). Potentially, nonsense mutations in *PTPRK* can contribute to the development of HCC by enhancing EGF-signaling as well, although further research should be conducted to identify the significance of these findings in human cancer.

Taken together, our data indicates that an increased positive selection of cells carrying potential driver mutations, including these nonsense mutations in *PTPRK*, might be occurring in precancerous, cirrhotic livers. This could ultimately drive the development of HCC as a consequence of alcohol consumption through the following mechanism. Chronic damage to the liver due to alcohol consumption causes apoptosis and necrosis of cells in the liver, including hepatocytes, which leads to liver inflammation (*32*). Subsequently, tissue-specific ASCs, which are normally quiescent, are required to proliferate to aid in the regeneration of the liver (*33*). Oncogenic mutations that have accumulated in the genomes of ASCs by random chance in both healthy and alcoholic livers, could potentially enable these cells to proliferate more quickly. As this proliferative advantage is especially important in the inflamed, cirrhotic liver, these mutations might thus allow clonal outgrowth of ASCs specifically in the cirrhotic livers. Consequently, the number of cells that carry driver mutations is simply higher in alcoholic livers and, hence, the chance of an additional driver mutation occurring in a cell which already has accumulated driver mutation(s) is higher. In conclusion, we propose that chronic alcohol consumption creates an inflamed, cirrhotic liver tissue environment, which may provide a fertile ground for cells with oncogenic mutations. As alcohol consumption has been linked

to epigenetic changes in the liver as well (*34*, *35*), future research should be directed to identify whether these mechanisms might work cooperatively in inflamed tissues.

Inflammatory diseases, such as inflammatory bowel disease and pancreatitis, increase the risk of developing cancer in multiple tissues (*36*). However, it was believed that this link was at least partially established by a direct induction of mutations in the genome (*37*). The results presented here might indicate that inflammation is more directly involved in the development of cancer in these diseases. Furthermore, changes in the microenvironment due to alcohol consumption might contribute to the development of cancer in other tissues as well, such as the breast and the esophagus. Our findings illustrate that reversal of the inflammatory phenotype that precedes cirrhosis and HCC might be sufficient to prevent the development of HCC.

ACKNOWLEDGEMENTS

The authors would like to thank the Utrecht Sequencing Facility and the UBEC for sequencing and for input on the bioinformatic analyses, respectively. The UBEC is subsidized by the University Medical Center Utrecht and the Utrecht Sequencing Facility is subsidized by the University Medical Center Utrecht, Hubrecht Institute, and Utrecht University. This study was financially supported by the research program InnoSysTox (project number 114027003), by the Netherlands Organisation for Health Research and Development (ZonMw), by the Dutch Cancer Society (project number 10496) and is part of the Oncode Institute, which is partly financed by the Dutch Cancer Society and was funded by the gravitation program CancerGenomiCs.nl from the Netherlands Organisation for Scientific Research (NWO). We thank the Hartwig Medical Foundation (Amsterdam, The Netherlands) for generating, analyzing and providing access to reference whole genome sequencing data of the Netherlands population.

AUTHOR CONTRIBUTIONS

R.L., J.J., J.I., and M.D. collected liver biopsies. M.J., E.K., and N.B. performed organoid culturing. N.B. isolated the RNA and sequenced the organoid cultures. M.J., M.L., R.J., and S.B. performed bioinformatic analyses. M.J., E.K., M.V., R.B., L.L., and E.C. were involved in the conceptual design of this study. M.J. and E.C. wrote the manuscript. E.K. and R.B. provided textual comments. R.B., L.L., and E.C. supervised this study.

METHODS

Human tissue material

All human specimens were obtained in the Erasmus Medical Center Rotterdam. Liver biopsies from healthy liver donors and patients with alcoholic cirrhosis were obtained during liver transplantation procedures. The biopsies were collected in cold organ preservation fluid (Belzer UW Cold Storage Solution, Bridge to Life, London, UK) and transported and stored at 4˚C until use. The liver and tumor biopsies from the hepatocellular carcinoma patient were gathered from a resected specimen and stored at -80ºC until use. The acquisition of these liver and tumor biopsies for research purposes was approved by the Medical Ethical Committee of the Erasmus Medical Center (MEC-2014-060 and MEC-2013-143). Informed consent was provided by all patients involved.

The biopsies of the HCC were sliced into slices of 6µm. Subsequently, the tumor percentage of both ends of each biopsy was determined using HE staining (Suppl. Fig. S6). The tumor percentage of the biopsies was considered to be the average of these two values. The remaining slices were used for long-term storage at -80ºC and for DNA isolation.

Generation of clonal liver organoid cultures from human liver biopsies

We derived organoid cultures from healthy and alcoholic liver tissue material as described previously (*38*,

39), by dissociating liver biopsies into single cell solutions using human liver digestion solution (EBSS supplemented with 1 mg/ml collagenase type 1A and 0.1 mg/ml DNaseI) and plating these into BME overlaid with culture medium. After 2-3 days, organoids started to appear in the BME. One week after isolation, the organoids were passaged to remove blood cells from the BME. The cultures were maintained for approximately 14 days after isolation, to enrich for ASCs. Subsequently, clonal organoid cultures were generated from these organoid cultures as described previously (*20*) (**Chapter 2**: steps 1 - 29), by FACS-sorting single cells, plating these in limiting dilution in BME overlaid with culture medium, and picking single clonal organoids 2 - 3 weeks after the FACS-sort. The organoid cultures were expanded for several weeks, until there was enough material to isolate enough DNA to perform WGS.

Whole-genome sequencing and read alignment

DNA was isolated from all organoid cultures, blood samples, and tissue biopsies using the Qiasymphony (Qiagen). Whole-genome sequencing libraries were generated from 200 ng of genomic DNA according to standard protocols (Illumina). The organoid cultures and control samples were sequenced paired-end (2 x 100 bp) to a depth of at least 30X coverage on the Illumina HiSeq Xten. The HCC biopsies were sequenced paired-end (2 x 100 bp) to a depth of at least 60X coverage on the Illumina HiSeq Xten, which is required to identify the somatic base substitutions in the tumor cells, as the biopsies contain ~50% healthy cells (Suppl. Fig. S6). Whole-genome sequencing was performed at the Hartwig Medical Foundation in Amsterdam, the Netherlands. Using the Burrows–Wheeler Aligner (BWA) tool v0.7.5a (*40*) with settings '-t 4 -c 100 -M', the sequence reads were mapped to the human reference genome GRCh37.

Copy number alteration calling and filtering

For the organoid cultures, CNA catalogs were obtained and filtered according to protocol (*20*) (**Chapter 2**: steps 50 - 57), by using using FreeC v2.7 (*41*). BED-file of blacklist positions is available upon request. For the HCC biopsies, structural variants were called using Manta v.1.1.0 (*42*) with standard settings. We only considered structural variations of at least 150 base pairs in autosomal the genome with a manta filter 'PASS'. Subsequently, the mutation catalogs of all five biopsies were intersected with a window of 500 bp to obtain the trunk CNAs using bedtools (*43*).

All CNA calls were inspected manually in the Integrative Genomics Viewer (IGV) to exclude false-positives with no visible change in read-depth. The breakpoints were identified manually in IGV. Finally, the number of genes within the deletions was obtained from http://genome.ucsc.edu/.

Genome-wide copy number profiles

Genome-wide copy number profiles of the ASCs were estimated by using the output of the FreeC calls obtained in 'Copy number alteration calling and filtering' prior to filtering. Subsequently, we calculated the mean copy number across 500,000 bp bins to determine genome stability. Genome-wide copy number profiles of the HCC biopsies and the adjacent liver biopsy were obtained in a similar manner.

Base substitution calling and filtering

For the organoid cultures, high-confidence base substitution catalogs were obtained by filtering GATK v3.4-46 (*44*) variant calls according to protocol (*20*) (**Chapter 2:** steps 58 - 68), with additional removal of variants with a sample-specific genotype quality < 10 in the control sample, and positions with a sample-specific genotype quality < 99 in the organoid clone sample. BED-file of blacklist positions is available upon request. All organoids showed a peak at a base substitution VAF of 0.5, indicating that the organoid samples are clonal (Suppl. Fig. S7). We downloaded publicly available variant call format (VCF) files and surveyed bed files of healthy liver ASCs from donors 14 - 18 from https://wgs11.op.umcutrecht.nl/mutational_patterns_ASCs/ to allow the comparison between healthy and alcoholic liver ASCs.

For the HCC biopsies, base substitutions were called by using Strelka v1.0.14 with settings 'SkipDepthFilters = 0', 'maxInputDepth = 250', 'depthFilterMultiple = 3.0', 'snvMaxFilteredBasecallFrac = 0.4', 'snvMaxSpanningDeletionFrac = 0.75', 'indelMaxRefRepeat = 1000', 'indelMaxWindowFilteredBasecallFrac = 0.3', 'indelMaxIntHpolLength = 14', 'ssnvPrior = 0.000001', 'sindelPrior = 0.000001', 'ssnvNoise = 0.0000005', 'sindelNoise = 0.000001', 'ssnvNoiseStrandBiasFrac = 0.5', 'minTier1Mapq = 20', 'minTier2Mapq

= 5', 'ssnvQuality_LowerBound = 10', 'sindelQuality_LowerBound = 10', 'isWriteRealignedBam = 0', and 'binSize = 25000000'. We only considered variations with a filter 'PASS. Subsequently, the mutation catalogs of all five biopsies were intersected to obtain the trunk base substitutions using bedtools (*43*). We only considered base substitutions on the autosomal genome that did not overlap with an indel call. Positions that were detected at least 5 times in 1,762 Dutch individuals were removed from these catalogs using the Hartwig Medical Foundation Pool of Normals (HMF-PON) version 2 (available upon request), to exclude Dutch germline variations. Only 138 mutations are found in 4 biopsies, whereas we detect 7,203 mutations in five biopsies (Suppl. Fig. S4), indicating that the majority of the trunk mutations were identified successfully.

To exclude that the observed differences in mutational loads and mutational spectra are a consequence of the differences between the filtering pipelines, we also filtered all alcohol liver ASCs using our 'HCC filtering pipeline'. We did not observe a striking difference in mutational load or mutational profiles between the alcoholic liver samples using both filtering pipelines (Suppl. Fig. S8).

Tumor adjusted allele frequencies

The VAFs of the shared mutations (the trunk mutations) were calculated for each biopsy. Subsequently, we calculated the tumor-adjusted variant allele frequency (TAF) per biopsy, in which the VAF is divided by the tumor-fraction. Most biopsies showed a peak at a TAF of 0.5 (Suppl. Fig. S5), indicating that these mutations are clonal in each sample and that the biopsies share a recent common ancestor.

Base substitution and tandem base substitution rate in liver ASCs

The number of base substitutions in the genomes of liver ASCs was obtained from the VCF files and extrapolated to the non-N autosomal genome (2,682,655,440 bp) of GRCh37 using the callable/surveyed genome size obtained in 'Base substitution calling and filtering'. To identify whether the number of somatic base substitutions acquired in the genomes of liver ASCs are correlated with the age of the donor, we fitted a linear mixed-effects regression model using the nlme R package, as described previously (*21*). The donor was modelled as a random effect in this model, to account for the fact that multiple ASCs were sequenced from the same donor. Two-tailed *t*-test were performed to determine whether the correlation was significant. The accumulation of base substitutions did not correlate significantly with age in the alcoholic liver ASCs, most likely due to the fact that the age-range is much smaller in these donors. Therefore, we obtained the 95% confidence interval of the healthy liver ASCs from the output of the linear mixed-effects regression model and determined whether the number of somatic base substitutions acquired in the genomes of the alcoholic liver ASCs are within this 95% confidence interval.

To identify tandem base substitutions, we extracted base substitutions that were called on two consecutive bases in the GRCh37 human reference genome from the VCF files. Similar to single base substitutions, we extrapolated this number to the non-N autosomal genome and determined whether the number of tandem base substitutions was correlated with the age of the donor using a linear mixed effects regression model. As the number of tandem base substitutions did not significantly correlate with age in the alcoholic liver ASCs, we determined whether these mutation numbers are within the 95% confidence interval of the healthy liver ASCs.

Mutational pattern analysis

Mutation types were extracted from the VCF files and the mutational profiles were generated by retrieving the sequence context of each mutation. For the healthy and alcoholic liver ASCs, we calculated an 'average' mutational profile. Pairwise cosine similarities of these average mutational profiles and of the mutational profile of the trunk mutations of the HCC were calculated, to identify how similar these profiles are.

We reconstructed the mutational profiles of each liver ASCs using the 30 known mutational signatures. Next, we selected 20 mutational signatures that had a contribution of > 100 base substitutions across all liver ASCs and determined the relative contribution of these 20 signatures to the mutational profiles (Suppl. Fig. S3). A mutation rate per year was computed for each of these signatures per ASC. Two-sided *t*-tests were performed to identify significant differences in the number of mutations attributed to each signature per year between healthy and alcoholic liver ASCs. Differences with $q < 0.05$ (Benjamini-

Hochberg FDR multiple-testing correction) were considered significant. All mutational pattern analyses were performed using the MutationalPatterns R package (*45*).

Genomic distribution of somatic base substitutions

The promoter, enhancer, and open chromatin regions of hg19 were obtained from Ensembl using biomaRt (*46*, *47*) and the genic regions of hg19 were loaded from UCSC Known Genes tables as TxDb object (*48*). To determine whether the somatic base substitutions are non-randomly distributed, we tested for enrichment and depletion of base substitutions in these regions with a one-sided Binomial test, corrected for the callable regions per sample. For the HCC trunk mutations, the callable regions were obtained by defining callable loci per biopsy using the GATK CallableLoci tool v3.4.46 (*49*), with additional optional parameters 'minBaseQuality 10', 'minMappingQuality 10', 'maxFractionOfReadsWithLowMAPQ 20', and 'minDepth 15'. Subsequently, these files were intersected to obtain the regions that are callable in all biopsies. 96.79% of the non-N autosomal genome was callable in all six biopsies.

Two-sided poisson tests were performed to estimate significant differences in depletion/enrichment in all genomic regions between the healthy liver ASCs, the alcoholic liver ASCs, and the trunk mutations of the HCC. Differences with $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing correction) were considered significant. All mutational pattern analyses were performed using the MutationalPatterns R package (*45*).

d*N*/d*S* and Identification of non-silent mutations in cancer genes

d*N*/d*S* ratios were computed using the *dNdScv* R package (*25*). Briefly, this package computes the (local) background mutation rates and sequence composition of genes to calculate the background mutation rate for each gene. A likelihood ratio test is subsequently performed to identify genes that are significantly hit by non-silent mutations. The output of the *dNdScv* package was used to identify non-silent (nonsynonymous, nonsense, and splice site) mutations in cosmic cancer genes. For this analysis, we only considered 409 out of 719 cancer genes, which are categorized as 'tier 1' cancer genes (genes with sufficient evidence of being a cancer driver) and have a dominant effect, as we only have mono-allelic hits. The list of cosmic cancer genes was obtained from https://cancer.sanger.ac.uk/cosmic/census.

RNA sequencing

Organoid cultures of three healthy donors (18-c, 21-b, and 22-a) and three alcoholic organoids, of which 2 with a nonsense mutation in *PTPRK* (alc3-a and alc5-a) and one without any mutations in *PTPRK* (alc-3b), were cultured for 1 day either in presence or absence of hEGF in the culture medium. Subsequently, cells were collected in 0.5 ml Trizol. Total RNA was isolated using the QiaSymphony SP with the QiaSymphony RNA kit (Qiagen, 931636). mRNA sequencing libraries were generated from 50 ng total RNA using the Illumina Neoprep TruSeq stranded mRNA library prep kit (Illumina, NP-202-1001). RNA libraries were sequenced paired-end (2 x 75 bp) on the Nextseq500 to > 20 million reads per sample.

RNA sequencing reads were mapped to the human reference genome GRCh37 with STAR v.2.4.2a (*50*). The BAM-files were indexed using Sambamba v0.5.8 Subsequently, reads were counted using HTSeq-count 0.6.1p1 and read counts were normalized using DESeq v1.28.0. DESeq nbinomTest was used to test for differential expression of *PTPRK* between the organoids with a nonsense *PTPRK* mutation and the other organoids.

DATA ACCESS

Whole-genome sequencing data is available EGAS00001002983.

REFERENCES

1. V. Bagnardi *et al.*, Alcohol consumption and site-specific cancer risk: a comprehensive dose–response meta-analysis. *Br. J. Cancer*. **112**, 580–593 (2014).
2. B. W. Stewart, C. P. Wild, *World Cancer Report 2014* (2014).
3. P. Boffetta, M. Hashibe, C. La Vecchia, W. Zatonski, J. Rehm, The burden of cancer

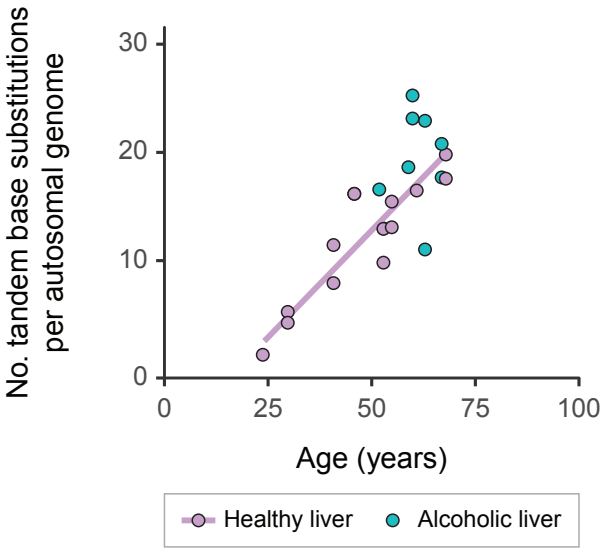
attributable to alcohol drinking. *International Journal of Cancer*. **119**, 884–887 (2006).

4. World Health Organization, *Global Status Report on Alcohol and Health* (World Health Organization, 2014).
5. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, International Agency for Research on Cancer, *Alcohol Consumption and Ethyl Carbamate* (World Health Organization, 2010).
6. A. Mizumoto *et al.*, Molecular Mechanisms of Acetaldehyde-Mediated Carcinogenesis in Squamous Epithelium. *Int. J. Mol. Sci.* **18** (2017), doi:10.3390/ijms18091943.
7. G. Obe, H. Ristow, Mutagenic, cancerogenic and teratogenic effects of alcohol. *Mutat. Res.* **65**, 229–259 (1979).
8. A. Helander, K. Lindahl-Kiessling, Increased frequency of acetaldehyde-induced sister-chromatic exchanges in human lymphocytes treated with an aldehyde dehydrogenase inhibitor. *Mutat. Res. Lett.* **264**, 103–107 (1991).
9. T. Matsuda, M. Kawanishi, S. Matsui, T. Yagi, H. Takebe, Specific tandem GG to TT base substitutions induced by acetaldehyde are due to intra-strand crosslinks between adjacent guanine bases. *Nucleic Acids Res.* **26**, 1769–1774 (1998).
10. M. Tamura, H. Ito, H. Matsui, I. Hyodo, Acetaldehyde is an oxidative stressor for gastric epithelial cells. *J. Clin. Biochem. Nutr.* **55**, 26–31 (2014).
11. G. Novitskiy, K. Traore, L. Wang, M. A. Trush, E. Mezey, Effects of ethanol and acetaldehyde on reactive oxygen species production in rat hepatic stellate cells. *Alcohol. Clin. Exp. Res.* **30**, 1429–1435 (2006).
12. A. P. Grollman, M. Moriya, Mutagenesis by 8-oxoguanine: an enemy within. *Trends Genet.* **9**, 246–249 (1993).
13. B. van Loon, E. Markkanen, U. Hübscher, Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. *DNA Repair* . **9**, 604–616 (2010).
14. J. Chang *et al.*, Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. *Nat. Commun.* **8**, 15290 (2017).
15. K. Schulze *et al.*, Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* **47**, 505–511 (2015).
16. A. Fujimoto *et al.*, Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500–509 (2016).
17. E. Letouzé *et al.*, Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* **8**, 1315 (2017).
18. L. Zhu *et al.*, Multi-organ Mapping of Cancer Risk. *Cell*. **166**, 1132–1146.e7 (2016).
19. H. K. Seitz, F. Stickel, Molecular mechanisms of alcohol-mediated carcinogenesis. *Nat. Rev. Cancer*. **7**, 599–612 (2007).
20. M. Jager *et al.*, Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures. *Nat. Protoc.* **13**, 59–78 (2018).
21. F. Blokzijl *et al.*, Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*. **538**, 260–264 (2016).
22. L. B. Alexandrov *et al.*, Signatures of mutational processes in human cancer. *Nature*. **500**, 415–421 (2013).
23. S. Nik-Zainal *et al.*, Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. **534**, 47–54 (2016).
24. Z. Kan *et al.*, Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* **23**, 1422–1433 (2013).
25. I. Martincorena *et al.*, Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*. **171**, 1029–1041.e21 (2017).
26. V. Hernandez–Gea, S. Toffanin, S. L. Friedman, J. M. Llovet, Role of the Microenvironment in the Pathogenesis and Treatment of Hepatocellular Carcinoma. *Gastroenterology*. **144**, 512–527 (2013).
27. Y. Xu, L.-J. Tan, V. Grachtchouk, J. J. Voorhees, G. J. Fisher, Receptor-type protein-tyrosine

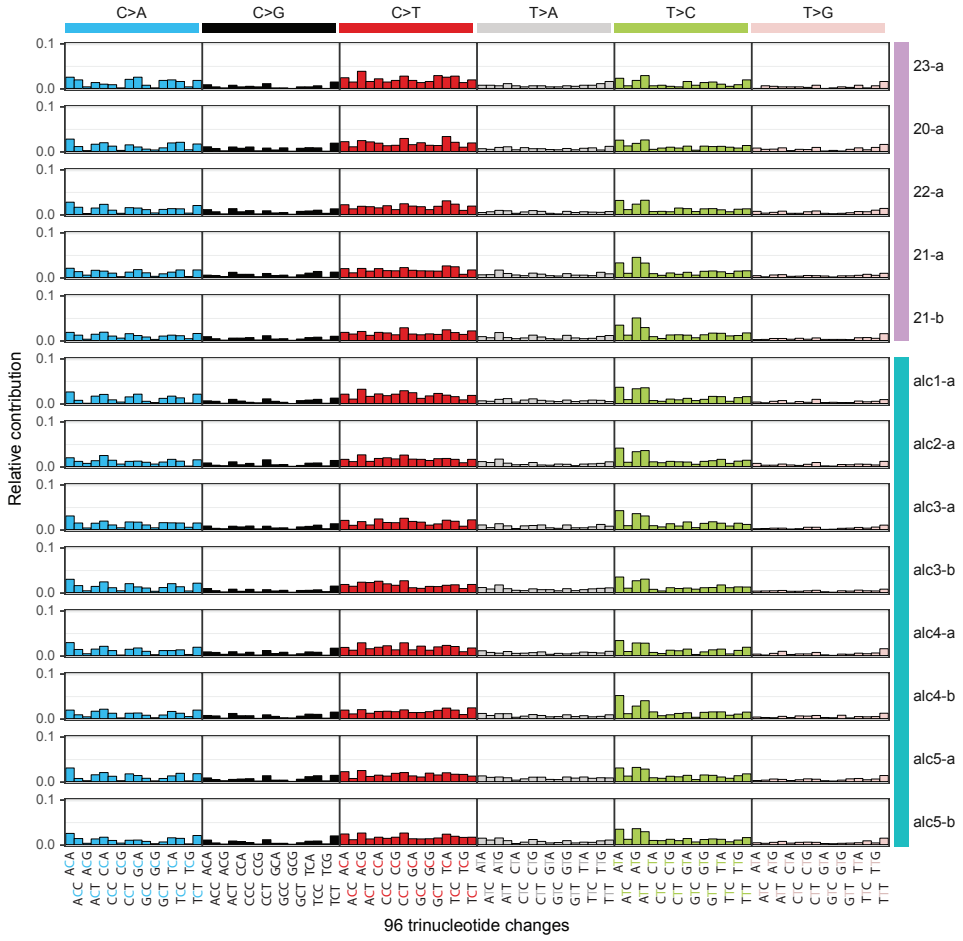
phosphatase-kappa regulates epidermal growth factor receptor function. *J. Biol. Chem.* **280**, 42694–42700 (2005).

28. P.-H. Sun, L. Ye, M. D. Mason, W. G. Jiang, Protein tyrosine phosphatase kappa (PTPRK) is a negative regulator of adhesion and invasion of breast cancer cells, and associates with poor prognosis of breast cancer. *J. Cancer Res. Clin. Oncol.* **139**, 1129–1139 (2013).
29. J. R. Flavell *et al.*, Down-regulation of the TGF-beta target gene, PTPRK, by the Epstein-Barr virus encoded EBNA1 contributes to the growth and survival of Hodgkin lymphoma cells. *Blood*. **111**, 292–301 (2008).
30. J.-H. Zhong *et al.*, Epidermal Growth Factor Gene Polymorphism and Risk of Hepatocellular Carcinoma: A Meta-Analysis. *PLoS One*. **7**, e32159 (2012).
31. K. K. Tanabe *et al.*, Epidermal growth factor gene functional polymorphism and the risk of hepatocellular carcinoma in patients with cirrhosis. *JAMA*. **299**, 53–60 (2008).
32. T. Luedde, N. Kaplowitz, R. F. Schwabe, Cell death and cell death responses in liver disease: mechanisms and clinical relevance. *Gastroenterology*. **147**, 765–783.e4 (2014).
33. W.-Y. Lu *et al.*, Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nat. Cell Biol.* **17**, 971–983 (2015).
34. Z. Herceg, A. Paliwal, Epigenetic mechanisms in hepatocellular carcinoma: how environmental factors influence the epigenome. *Mutat. Res.* **727**, 55–61 (2011).
35. S. D. Shukla, R. W. Lim, Epigenetic effects of ethanol on the liver and gastrointestinal system. *Alcohol Res.* **35**, 47–55 (2013).
36. A. Mantovani, P. Allavena, A. Sica, F. Balkwill, Cancer-related inflammation. *Nature*. **454**, 436–444 (2008).
37. T. Shimizu, H. Marusawa, Y. Endo, T. Chiba, Inflammation-mediated genomic instability: roles of activation-induced cytidine deaminase in carcinogenesis. *Cancer Sci.* **103**, 1201–1206 (2012).
38. L. Broutier *et al.*, Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation. *Nat. Protoc.* **11**, 1724–1743 (2016).
39. M. Huch *et al.*, Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell.* **160**, 299–312 (2015).
40. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).
41. V. Boeva *et al.*, Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*. **28**, 423–425 (2012).
42. X. Chen *et al.*, Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. **32**, 1220–1222 (2016).
43. A. R. Quinlan, BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinformatics*. **47**, 11.12.1–34 (2014).
44. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
45. F. Blokzijl, R. Janssen, R. van Boxtel, E. Cuppen, MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).
46. S. Durinck *et al.*, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. **21**, 3439–3440 (2005).
47. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
48. M. Carlson, B. P. Maintainer, *TxDb.Hsapiens. UCSC.hg19.knownGene: Annotation package for TxDb object(s)* (2015).
49. G. A. Van der Auwera *et al.*, From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*. **43**, 11.10.1–33 (2013).
50. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

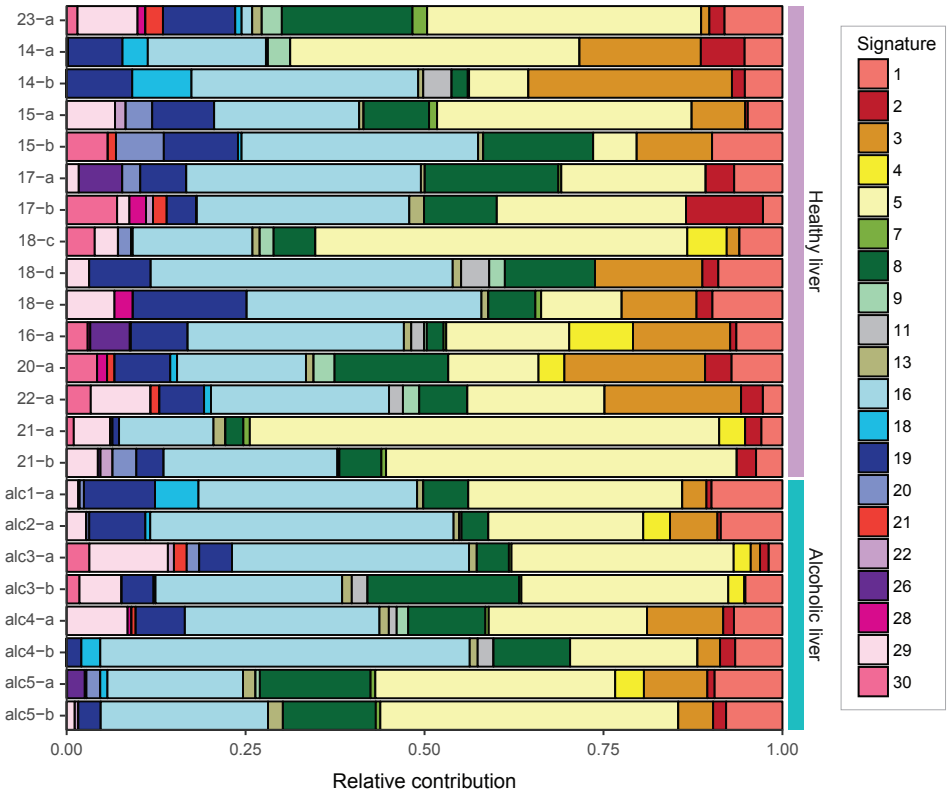
SUPPLEMENTAL FIGURES AND TABLES



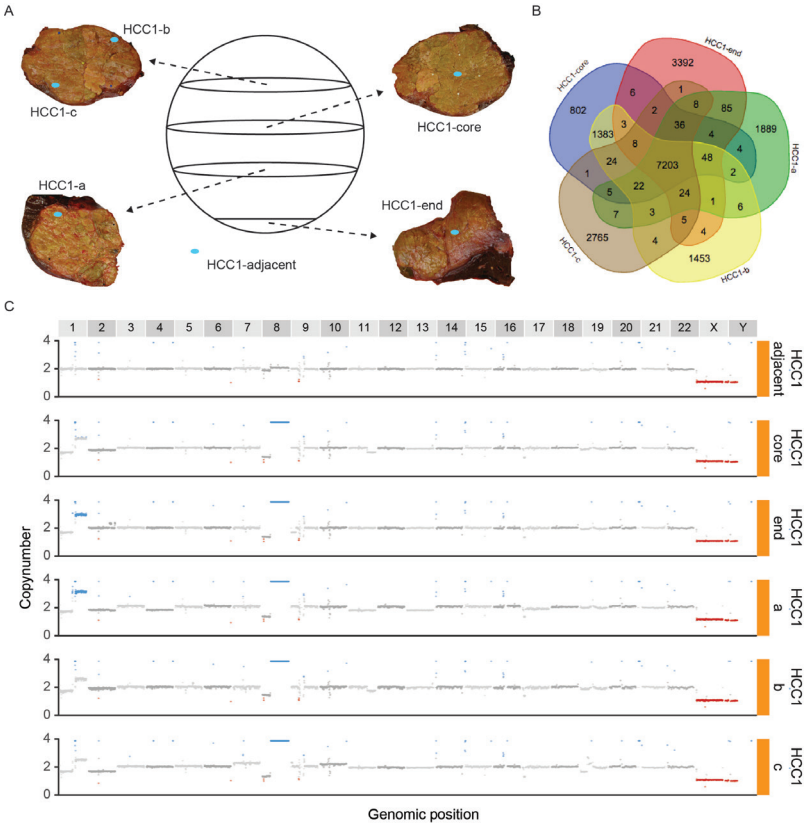
Supplemental figure S1. The number of tandem base substitutions acquired in the autosomal genomes of 15 healthy and 8 alcoholic liver ASCs derived from 9 and 5 donors, respectively. Data points represent single stem cells. A linear accumulation of tandem base substitutions with age was observed in the healthy liver (two-tailed *t*-test, linear mixed model; $P = 2.01 \times 10^{-4}$), indicated by the purple trendline. All tandem base substitution numbers in the genomes of the alcoholic liver ASCs are within the 95% confidence interval of the healthy liver ASCs, except for alc4-b, which is slightly lower (11.3 tandem base substitutions at 63 years).



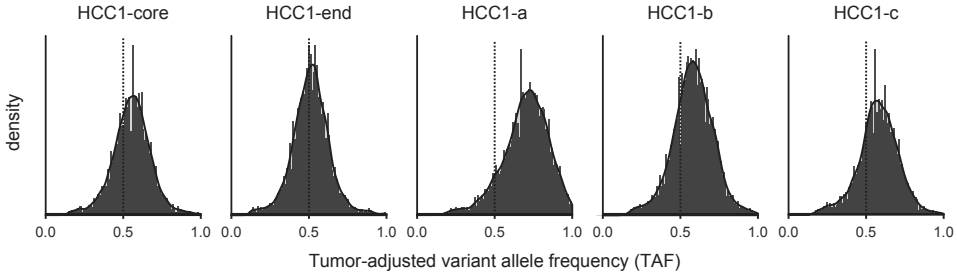
Supplemental figure S2. Mutational profiles of the acquired base substitutions in the genomes of healthy and alcoholic liver ASCs. In this figure, we only displayed the mutational profiles of the new samples. Base substitution spectra of the remaining healthy liver ASCs can be obtained from Extended Data Figure 5 of (*21*).



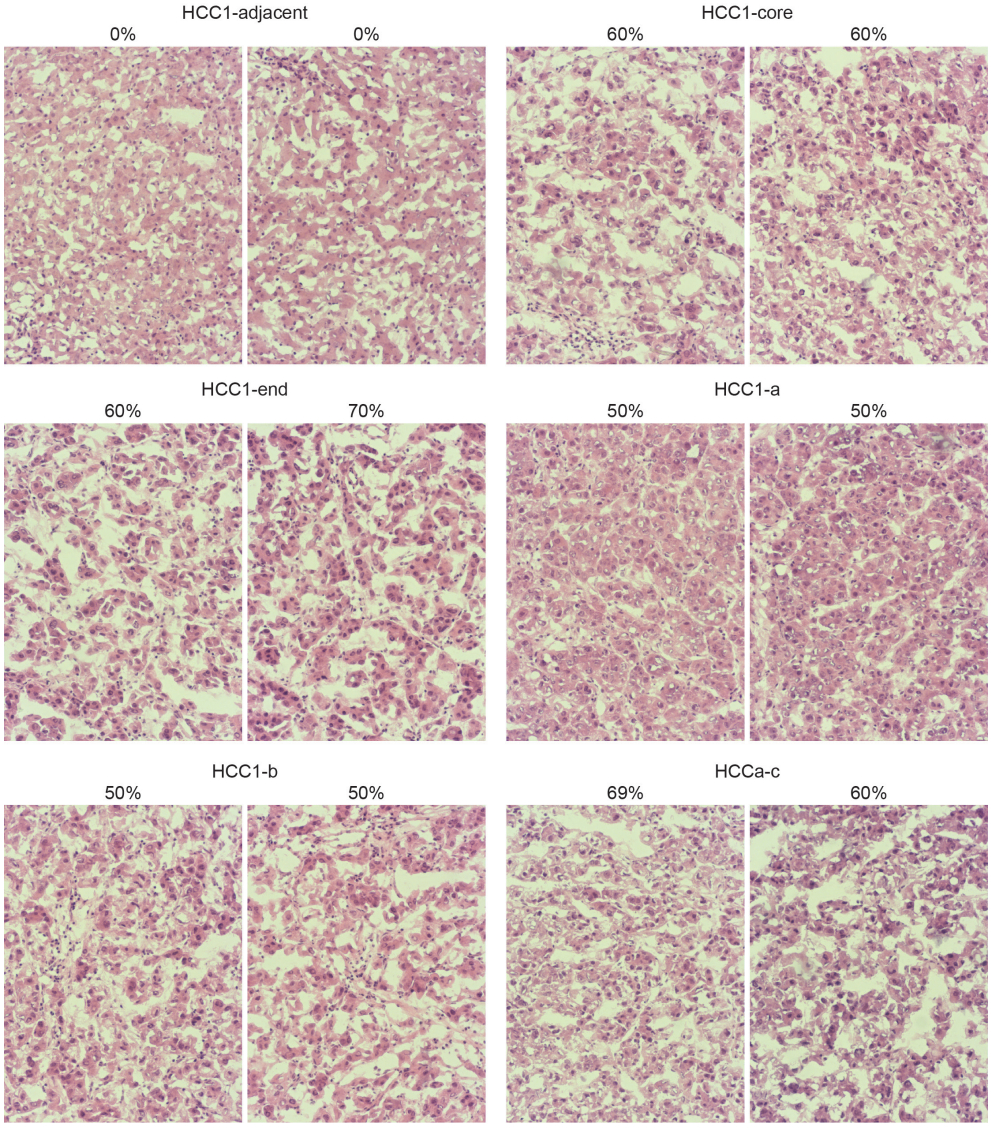
Supplemental figure S3. Relative contribution of the 30 COSMIC mutational signatures to the mutational profiles of the somatic base substitutions acquired in the genomes of healthy and alcoholic liver ASCs.



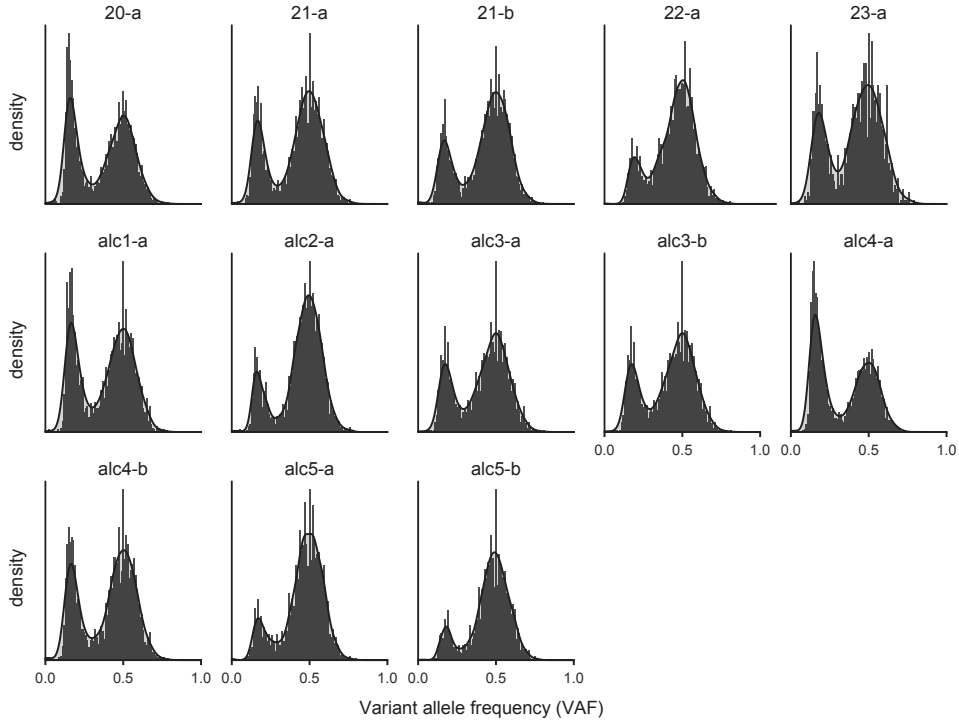
Supplemental figure S4. Measuring the accumulation of mutations in biopsies of an alcohol-related HCC. (A) Schematic depiction of the tumor (large circle) and the biopsies that were taken (blue dots) from 4 slices of the tumor (ovals). (B) Venn diagram of somatic base substitutions identified in five biopsies of the alcohol-related HCC. Venn diagram was created using http://bioinformatics.psb.ugent.be/webtools/Venn/. (C) Genome-wide copy number profiles of all five HCC biopsies and of healthy adjacent tissue. Red data points indicate a copy number state < 2, grey data points represent a copy number state of 2, and blue data points indicate a copy number state > 2.



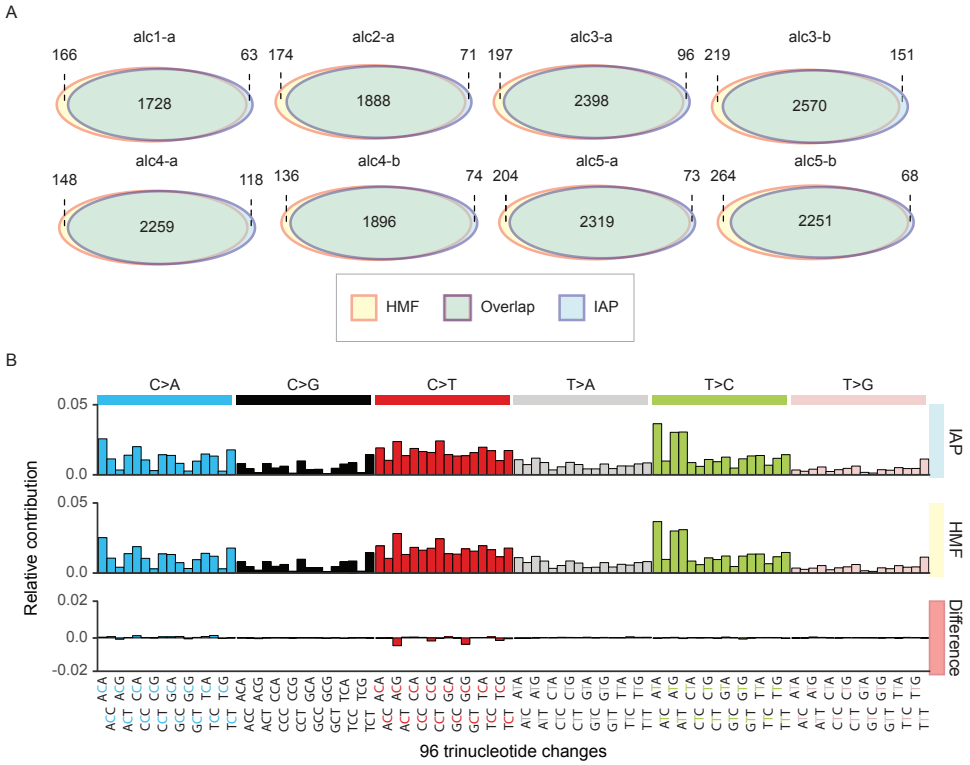
Supplemental figure S5. Distribution of the variant allele frequencies of the 7,203 somatic base substitutions that are shared by all five HCC biopsies, per biopsy, adjusted for the estimated tumor percentage per biopsy. Dotted lines indicate a tumor-adjusted variant allele frequency (TAF) of 0.5.



Supplemental figure S6. HE-staining of slices of the ends of five HCC biopsies and of a healthy adjacent biopsy. Tumor percentage (indicated above each picture) was estimated by the pathology department of the UMC Utrecht (the Netherlands) based on these stainings.



Supplemental figure 7. Variant allele frequency (VAF) distribution of the somatic base substitutions acquired in the genomes of healthy and alcoholic liver ASCs before filtering for VAF ≥ 0.3. In this figure, VAF plots of new samples are only displayed. VAF plots of the remaining healthy liver ASCs can be found in Extended Data Figure 2 of (*21*).



Supplemental figure 8. Comparison between the base substitution catalogs of the alcoholic liver ASCs obtained using two independent variant calling pipelines. The Illumina sequencing pipeline (IAP) (https://github.com/UMCUGenetics/IAP) was used to identify base substitutions in the healthy and alcoholic liver ASCs, whereas the Hartwig Medical Foundation (HMF) pipeline (https://github.com/hartwigmedical) was used to detect base substitutions in the five HCC biopsies. (A) Venn diagrams showing the overlapping variant calls between mutation catalogs of IAP and HMF per alcoholic liver ASC. A variant was considered to overlap when it was called on the same genomic position. (B) Relative contribution of the 96 context-dependent base substitution types to the mutational profiles of mutations obtained using IAP and HMF, and to the difference between these two profiles. To generate these plots, mutational profiles of 8 alcoholic liver ASCs were combined into a single mutational profile per filtering pipeline.

Sample	Donor	Age	Gender	Sample type	Surveyed genome (%)	Base substitutions	Tandem base substitutions	CNAs*
23-a	23	24	Male	Liver	95.6	822	2	0
14-a	14	30	Male	Liver	81.4	771	4	2
14-b	14	30	Male	Liver	85.2	888	5	0
15-a	15	41	Female	Liver	93.5	1,292	11	0
15-b	15	41	Female	Liver	95.1	1,351	8	0
17-a	17	46	Female	Liver	79.4	1,495	13	0
17-b	17	46	Female	Liver	73.7	1,273	12	0
18-c	18	53	Male	Liver	97.9	1,845	10	0
18-d	18	53	Male	Liver	98.5	1,504	13	1
18-e	18	53	Male	Liver	98.3	1,577	13	1
16-a	16	55	Male	Liver	97.5	1,919	13	1
20-a	20	55	Female	Liver	96.2	2,066	15	0
22-a	22	61	Male	Liver	96.5	2,259	16	0
21-a	21	68	Female	Liver	96.2	2,625	19	0
21-b	21	68	Female	Liver	96.5	2,383	17	0
alc1-a	alc1	52	Male	Alcoholic liver	96.0	1,791	16	0
alc2-a	alc2	59	Male	Alcoholic liver	96.6	1,959	18	0
alc3-a	alc3	60	Female	Alcoholic liver	96.0	2,494	22	0
alc3-b	alc3	60	Female	Alcoholic liver	96.2	2,721	24	0
alc4-a	alc4	63	Male	Alcoholic liver	96.8	2,377	22	0
alc4-b	alc4	63	Male	Alcoholic liver	96.7	1,970	11	0
alc5-a	alc5	67	Male	Alcoholic liver	96.7	2,392	20	0
alc5-b	alc5	67	Male	Alcoholic liver	95.9	2,319	17	0

*CNA = copy number alteration; see extended data table 2 of Blokzijl *et. al* (2016) for type and size of CNA

Supplemental table S1. Somatic base substitutions, tandem base substitutions, and copy number alterations acquired in the genomes of healthy and alcoholic liver ASCs during life.

Sample	Donor	Age	Gender	Sample type	Base substitutions
HCC1-clonal	HCC1	60	Male	Clonal HCC	7,203
HCC1-core	HCC1	60	Male	HCC biopsy	9,553
HCC1-end	HCC1	60	Male	HCC biopsy	10,830
HCC1-a	HCC1	60	Male	HCC biopsy	9,347
HCC1-b	HCC1	60	Male	HCC biopsy	10,193
HCC1-c	HCC1	60	Male	HCC biopsy	10,118

Supplemental table S2. Somatic base substitutions identified in five biopsies of one alcohol-related HCC.

Sample	Donor	Age	Gender	Sample type	Chr.*	Start	Stop	Size	Type	Genes
HCC1-clonal	HCC1	60	Male	Clonal HCC	10	60262954	60270868	7,915	Deletion	0
HCC1-clonal	HCC1	60	Male	Clonal HCC	17	2837706	7887745	5,050,039	Deletion	545

* Chr = chromosome

Supplemental table S3. Somatic copy number alterations detected in a recent common ancestor of an alcohol-related HCC.

New beginnings

Chapter 6

Organoid models of human and mouse ductal pancreatic cancer

Sylvia Boj,[1,2,#] Chang-Il Hwang,[3,4,#] Lindsey Baker,[3,4,#] Iok In Christine Chio,[3,4,#] Dannielle Engle,[3,4,#] Vincenzo Corbo,[3,4,#] Myrthe Jager,[1,#] Mariano Ponz-Sarvise,[3,4] Hervé Tiriac,[3,4] Mona Spector,[3,4] Ana Gracanin,[1,2] Tobiloba Oni,[3,4,5] Kenneth Yu,[3,4,6,7] Ruben van Boxtel,[1] Meritxell Huch,[1,8] Keith Rivera,[3] John Wilson,[3] Michael Feigin,[3,4] Daniel Öhlund,[3,4] Abram Handly-Santana,[4,9] Christine Ardito-Abraham,[3,4] Michael Ludwig,[3,4] Ela Elyada,[3,4] Brinda Alagesan,[3,4,10] Giulia Biffi,[3,4] Georgi Yordanov,[4,9] Bethany Delcuze,[3,4] Brianna Creighton,[3,4] Kevin Wright,[3,4] Youngkyu Park,[3,4] Folkert Morsink,[11] Quintus Molenaar,[12] Inne Borel Rinkes,[12] Edwin Cuppen,[1] Yuan Hao,[3] Ying Jin,[3] Isaac Nijman,[1] Christine Iacobuzio-Donahue,[6] Steven Leach,[6] Darryl Pappin,[3] Molly Hammell,[3] David Klimstra,[13] Olca Basturk,[13] Ralph Hruban,[14] George Johan Offerhaus,[11] Robert Vries,[1,2] Hans Clevers[1,*] and David Tuveson[3,4,6,*]

1 Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW), University Medical Centre Utrecht and CancerGenomics.nl, 3584CT Utrecht, the Netherlands
2 Foundation Hubrecht Organoid Technology (HUB), 3584CT, Utrecht, the Netherlands
3 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
4 Lustgarten Foundation Pancreatic Cancer Research Laboratory, Cold Spring Harbor, NY 11724, USA
5 Graduate Program in Molecular and Cellular Biology, Stony Brook University, Stony Brook, NY 11794, USA
6 Rubenstein Center for Pancreatic Cancer Research, MSKCC, New York, NY 10065, USA
7 Weill Medical College at Cornell University, New York, NY 10065, USA
8 Current address: Gurdon Institute-University of Cambridge, Tennis Court Road Cambridge, CB2 1QN, UK
9 Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
10 Graduate Program in Genetics, Stony Brook University, Stony Brook, NY 11794, USA
11 Department of Pathology, University Medical Centre Utrecht, 3584 CX Utrecht, The Netherlands
12 Department of Surgery, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands
13 Department of Pathology, MSKCC, New York, NY 10065, USA
14 The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA
Co-first authors
*Correspondence

Adapted from: Cell 2016 Oct; 160(1-2): 324-338

SUMMARY

Pancreatic cancer is one of the most lethal malignancies due to its late diagnosis and limited response to treatment. Tractable methods to identify and interrogate pathways involved in pancreatic tumorigenesis are urgently needed. We established organoid models from normal and neoplastic murine and human pancreas tissues. Pancreatic organoids can be rapidly generated from resected tumors and biopsies, survive cryopreservation, and exhibit ductal- and disease-stage-specific characteristics. Orthotopically transplanted neoplastic organoids recapitulate the full spectrum of tumor development by forming early-grade neoplasms that progress to locally invasive and metastatic carcinomas. Due to their ability to be genetically manipulated, organoids are a platform to probe genetic cooperation. Comprehensive transcriptional and proteomic analyses of murine pancreatic organoids revealed genes and pathways altered during disease progression. The confirmation of many of these protein changes in human tissues demonstrates that organoids are a facile model system to discover characteristics of this deadly malignancy.

INTRODUCTION

Mortality due to pancreatic cancer is projected to surpass that of breast and colorectal cancer by 2030 in the United States (Rahib et al., 2014, Siegel et al., 2013). This dire scenario reflects an aging population, the improvement of outcomes for breast and colorectal cancer patients, the advanced stage at which most patients with pancreatic cancer are diagnosed, and the lack of durable treatment responses in pancreatic cancer patients. Indeed, effective therapeutic strategies for patients with pancreatic ductal adenocarcinoma (PDA) have been difficult to identify (Abbruzzese and Hess, 2014).

The therapeutic resistance of PDA has been explored in a variety of cell culture and animal model systems, with clinically actionable findings encountered only occasionally (Villarroel et al., 2011). Patient-derived xenografts (PDXs) have yielded insights into PDA, but their generation requires a large amount of tissue, and they take multiple months to establish (Kim et al., 2009, Rubio-Viqueira et al., 2006). Genetically engineered mouse models (GEMMs) of PDA have also been generated as a parallel system for fundamental biological investigation and preclinical studies (Pérez-Mancera et al., 2012). These GEMMs accurately mimic the pathophysiological features of human PDA, including disease initiation from preinvasive pancreatic intraepithelial neoplasms (PanINs) (Hingorani et al., 2003, Pérez-Mancera et al., 2012) and were used to discover that PDA possesses a deficient vasculature that impairs drug delivery (Erkan et al., 2009, Jacobetz et al., 2013, Koong et al., 2000, Olive et al., 2009, Provenzano et al., 2012). Although GEMMs have informed PDA therapeutic

development (Beatty et al., 2011, Frese et al., 2012, Neesse et al., 2014), they are expensive and time consuming (Pérez-Mancera et al., 2012). In addition, both human PDA and GEMMs exhibit an extensive stromal component that decreases the neoplastic cellularity, making it difficult to isolate and characterize the epithelium-derived malignant cells in pancreatic neoplastic tissues.

To study neoplastic cells, dissociated human tumors are often grown in two-dimensional (2D) culture conditions (Sharma et al., 2010), which do not support growth of untransformed, nonneoplastic pancreatic cells. Three-dimensional (3D) culture strategies have been developed to study normal, untransformed cells but so far have only allowed minimal propagation (Agbunag and Bar-Sagi, 2004, Lee et al., 2013, Means et al., 2005, Rovira et al., 2010, Seaberg et al., 2004). A comprehensive 3D cell culture model of murine and human PDA progression would facilitate investigation of genetic drivers, therapeutic targets, and diagnostics for PDA.

To address this deficiency, we sought to generate normal and neoplastic pancreatic organoids by modifying approaches we previously pioneered to culture intestinal (Sato et al., 2009), gastric (Barker et al., 2010), colon carcinoma (Sato et al., 2011), hepatic (Huch et al., 2013b), pancreatic (Huch et al., 2013a), and prostatic organoids (Gao et al., 2014, Karthaus et al., 2014). We developed 3D organoids from normal and malignant murine pancreatic tissues and used this model system to investigate PDA pathogenesis. Pancreatic organoids derived from wild-type mice and PDA GEMMs accurately recapitulate physiologically relevant aspects of disease progression *in vitro*. Following orthotopic transplantation, organoids from wild-type mouse normal pancreata are capable of regenerating normal ductal architecture, unlike other 3D model systems. We further developed methods to generate pancreatic organoids from normal and diseased human tissues, as well as from endoscopic needle biopsies. Following transplantation, organoids derived from murine and human PDA generate lesions reminiscent of PanIN and progress to invasive PDA. Finally, we demonstrate the utility of organoids to identify molecular pathways that correlate with disease progression and that represent therapeutic and diagnostic opportunities.

RESULTS

Murine Pancreatic Ductal Organoids Expressing Oncogenic Kras Recapitulate Features of PanINs

Recently, we derived continuously proliferating, normal pancreatic organoids from adult murine ductal cells (Huch et al., 2013a). We optimized this approach to generate models of PDA progression. We manually isolated small intralobular ducts and established organoid cultures from C57Bl/6 mouse normal pancreata and pancreatic

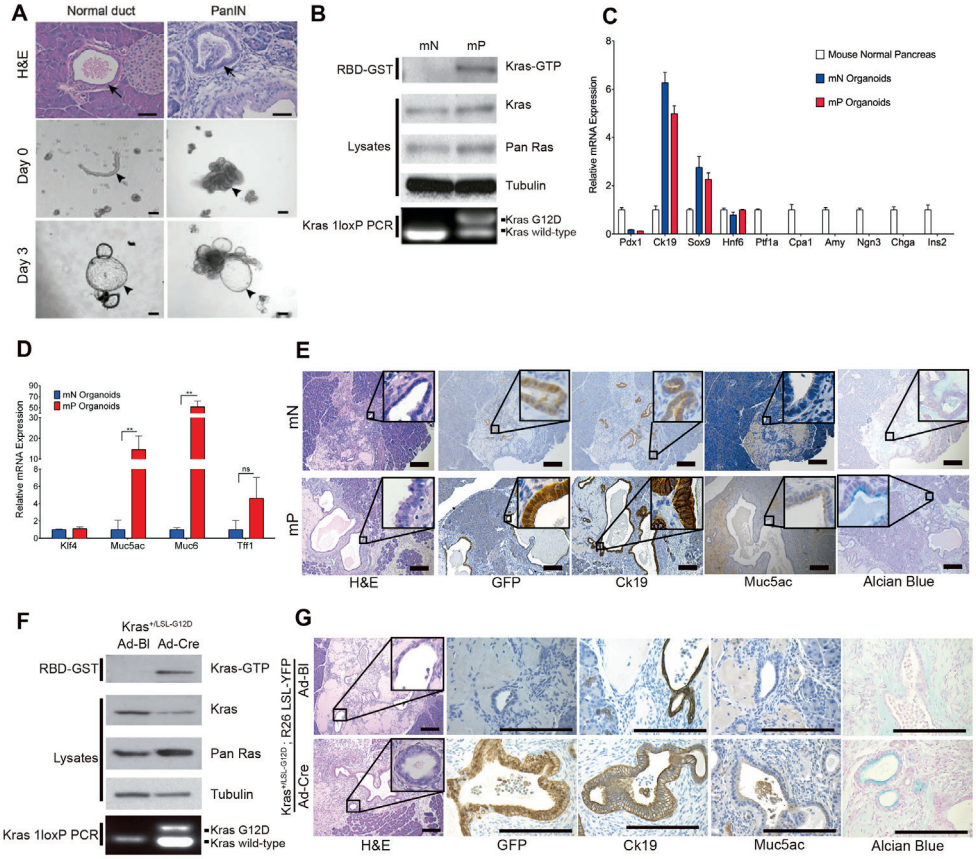


Figure 1. Oncogenic *Kras*^{G12D} Expression in Pancreatic Ductal Organoids Is Sufficient to Induce Preinvasive Neoplasms. (A) Hematoxylin and eosin (H&E) staining of murine pancreatic tissue used to prepare organoids (top). Arrows indicate mouse normal or PanIN ductal structures. Ducts embedded in Matrigel immediately following isolation (middle) and organoids 3 days postisolation (bottom). Arrowheads mark isolated ducts and growing organoids. Scale bars, 50 µm. (B) Immunoblots for Kras, pan Ras, Kras-GTP by RBD-GST pull-down, and Tubulin in mN and mPanIN (mP) organoids. PCR confirmation of Cre-mediated recombination of the *Kras*$^{LSL-G12D}$ allele (bottom). (C) qRT-PCR of ductal (*Pdx1*, *Ck19*, *Sox9*, and *Hnf6*), acinar (*Ptf1a*, *Cpa1*, and *Amy*), and endocrine (*Ngn3*, *Chga*, and *Ins2*) lineage markers in mN and mP organoids. Means of three biological replicates are shown. Error bars indicate SEMs. Values were normalized to mouse normal pancreas. (D) qRT-PCR of genes indicative of PanIN lesions (*Muc5ac*, *Muc6*, *Tff1*, and *Klf4*) in mN and mP organoids. Values were normalized to mN organoids. Means of three biological replicates are shown. Error bars indicate SEMs. $^{**}p < 0.01$ by two-tailed Student's *t* test. (E) H&E, Alcian blue staining, and immunohistochemistry (IHC) of orthotopic, syngeneic transplants of GFP-transduced mN and mP organoids. Scale bars, 200 µm. (F) Immunoblots for Kras, pan Ras, Kras-GTP by RBD-GST pull-down, and tubulin in *Kras*$^{+/LSL-G12D}$ organoids transduced with adenoviral-Cre (Ad-Cre) or adenoviral-blank (Ad-Bl). PCR confirmation of Cre-mediated recombination of the *Kras*$^{LSL-G12D}$ allele (bottom). (G) H&E, Alcian blue staining, and IHC of orthotopic syngeneic transplants of organoids transduced with Ad-Bl (*Kras*$^{+/LSL-G12D}$; *R26*$^{LSL-YFP}$) and Ad-Cre (*Kras*$^{+/G12D}$; *R26*^{YFP}) 2 weeks posttransplant. Scale bars, 200 µm.

tissues that contained low-grade murine PanIN (mPanIN-1a/b) from *Kras*$^{+/LSL\text{-}G12D}$; *Pdx1-Cre* ("KC") mice (Figure 1A). KC mice develop a spectrum of preinvasive ductal lesions that mirror human PanINs and, upon aging, stochastically develop primary and metastatic PDA (Hingorani et al., 2003). Ducts from KC pancreata were often larger and exhibited higher grades of dysplasia compared to those from wild-type mice (Figure 1A). After 1–3 days in culture, organoid growth was observed from isolated ducts (Figure 1A). We created a collection of 10 murine normal (mN) and 9 PanIN (mP) organoid cultures that we have continuously propagated for over 20 passages and successfully cryopreserved (Table S1A available online). mP organoids exhibited recombination of the conditional *Kras*$^{LSL\text{-}G12D}$ allele and higher levels of Kras-GTP when compared to mN organoids (Figure 1B).

To determine the contribution of different pancreatic lineages to the organoids, we evaluated the expression of pancreatic lineage markers in these cultures. Genes associated with the ductal lineage (*Ck19* and *Sox9*) (Cleveland et al., 2012) were enriched in the mN and mP organoids compared to total pancreatic tissues, which contain relatively few ductal cells (Figure 1C). In addition, the mP organoids upregulated genes indicative of a PanIN disease state (*Muc5ac*, *Muc6*, and *Tff1*) relative to mN, with no difference in *Klf4* (Figure 1D) (Prasad et al., 2005). GFP-transduced mN and mP organoids were orthotopically transplanted into syngeneic C57Bl/6 or *Nu/Nu* mice. mN organoids quickly formed ductal structures comprised of simple cuboidal cells that persisted for up to 1 month (n = 9/27 transplants) but were not observed after 2 months (n = 0/13 transplants) (Figure 1E and Table S1B). In comparison, mP organoids formed small cysts lined with a single layer of simple cuboidal ductal cells interspersed with mucin-containing columnar epithelial cells. Although we could not demonstrate that the mP transplants were contiguous with the native ductal system, they resembled preinvasive mPanIN (Figure S1C). These dysplastic epithelial cells persisted for 2 months or longer (n = 16/18 transplants), were GFP and Ck19 positive, expressed the mPanIN-associated mucin Muc5ac, and stained prominently with Alcian blue (Figure 1E and Table S1C). In addition, when compared to mN transplants, mP transplants had increased proliferation and a robust stromal response, which are characteristics of autochthonous mPanIN tissue (Figures S1A–S1C). The ability of transplanted mP organoids to form lesions with many of the features of mPanINs demonstrates the utility of this system as a model for early pancreatic neoplasia.

Multiple cellular origins have been proposed for the development of PDA, with the pancreatic acinar cell hypothesized to be a major contributor to PDA initiation (De La O et al., 2008, Gidekel Friedlander et al., 2009, Guerra et al., 2003, Habbe et al., 2008, Kopp et al., 2012, Morris et al., 2010, Sawey et al., 2007). However,

recent studies have suggested that transformation of pancreatic ductal cells can also give rise to PDA (Pylayeva-Gupta et al., 2012, Ray et al., 2011, von Figura et al., 2014). Acinar cells isolated from wild-type pancreata are unable to form organoids in our conditions (Huch et al., 2013a). Therefore, our pancreatic ductal organoid system offers a unique opportunity to determine whether ductal cells can give rise to mPanIN. To assess whether expression of oncogenic Kras in pancreatic ductal organoids is sufficient to induce mPanIN formation *in vivo*, we derived organoids from ducts harboring the conditional $Kras^{LSL\text{-}G12D}$ allele (Hingorani et al., 2003). Following activation of Kras by adenoviral-*Cre* (Ad-*Cre*) infection, $Kras^{G12D}$ organoids maintained expression of genes specific to ductal cells and not acinar or endocrine lineages (Figures S1D and S1E). Recombination of the $Kras^{LSL\text{-}G12D}$ allele was confirmed by PCR, and levels of GTP-bound Kras were increased relative to control-infected organoids (Figure 1F). In addition, expression of $Kras^{G12D}$ resulted in the upregulation of genes associated with human PanIN (Figure S1F). The $Kras^{G12D}$-expressing organoids demonstrated increased proliferation relative to control organoids (Figure S1G). Finally, $Kras^{G12D}$ organoids formed mPanIN-like structures with columnar cell morphology when implanted orthotopically into syngeneic mice (Figure 1G). This morphology contrasted with the normal-appearing ductal architecture formed by transplanting $Kras^{+/LSL\text{-}G12D}$ organoids or wild-type mN (Figures 1E and 1G). The ability of mPanIN-like structures to develop from $Kras^{G12D}$-expressing ductal organoids following transplantation demonstrates that ductal cells are also competent to form mPanINs.

Tumor-Derived Organoids Provide a Model for Murine PDA Progression

We prepared pancreatic ductal organoids from multiple murine primary tumors (mT) and metastases (mM) from KC and $Kras^{+/LSL\text{-}G12D}$; $Trp53^{+/LSL\text{-}R172H}$; *Pdx1-Cre* ("KPC") mice, which develop mPDA more rapidly than KC mice (Figures 2A and Table S2A) (Hingorani et al., 2005). mT and mM organoids exhibited recombination of the $Kras^{LSL\text{-}G12D}$ allele, as well as increased levels of Kras-GTP and Kras protein (Figure 2B). mT and mM organoids had increased levels of S6 phosphorylation, but not of Erk or Akt phosphorylation (Figure 2B).

Orthotopic transplantation of mT organoids initially generated low- and high-grade lesions that resembled mPanIN (Figure 2C and Table S2B). Over longer periods of time (1–6 months), transplants developed into invasive primary and metastatic mPDA (Figure 2C and Table S2B). mT organoids engrafted with a similar efficiency upon orthotopic transplantation in *Nu/Nu* mice (91.7%) compared to C57Bl/6 mice (85%), but disease progression was accelerated in *Nu/Nu* hosts (Table S2B). Although most mT organoid transplants required several months to

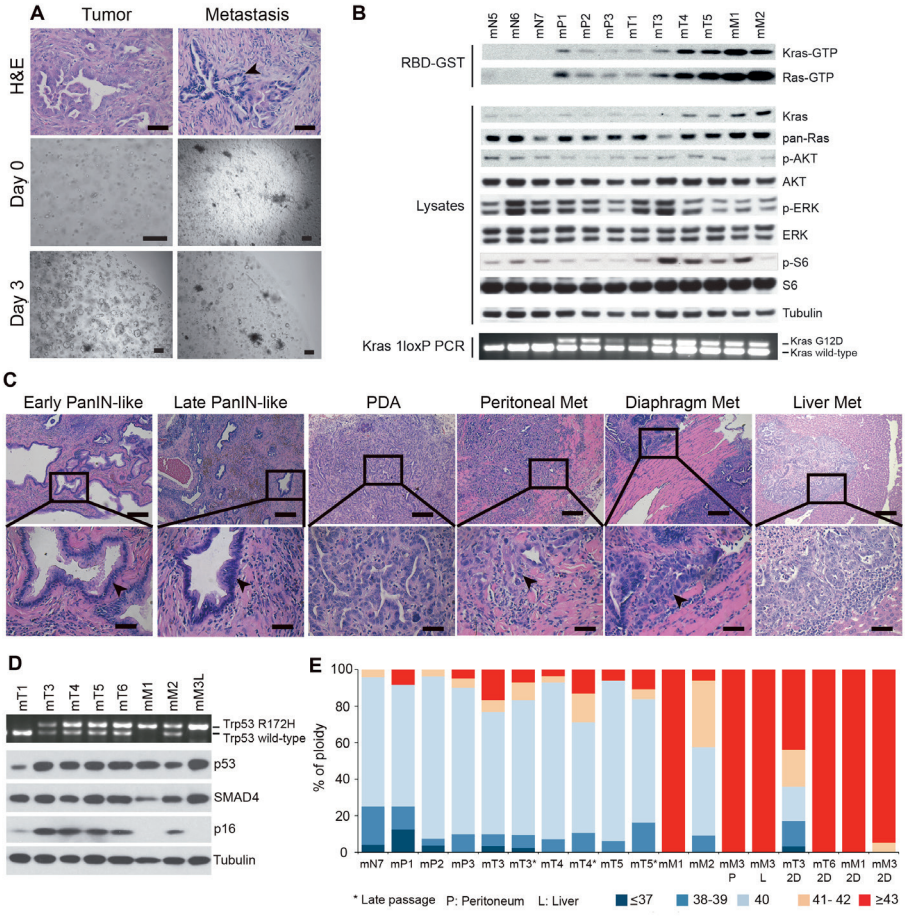


Figure 2. Modeling Murine PDA Progression with Tumor- and Metastasis-Derived Organoids. (A) H&E staining of murine tissue from which tumor and metastasis organoids were derived (top). Arrowhead indicates metastasis. Scale bars, 50 μm. Digested murine tissues embedded in Matrigel immediately following isolation (middle) and organoids 3 days postisolation (bottom). Scale bars, 200 μm. (B) Immunoblots of selected signaling effectors, Kras-GTP and Ras-GTP by RBD-GST pull-down, and tubulin. PCR confirmation of *Kras*$^{LSL\text{-}G12D}$ recombination in mP, mT, and mM organoids (bottom). (C) H&E staining of tumors and metastases (Met) derived from mT organoid orthotopic transplants. Scale bars, 200 μm (top) and 50 μm (bottom). (D) Loss of heterozygosity of the wild-type *Trp53* allele determined by PCR (top) and immunoblot analysis of Trp53, Smad4, p16, and Tubulin. mM3L, derived from a liver metastasis. (E) Karyotypes of organoids and monolayer (2D) cell lines.

progress from early mPanIN-like lesions to invasive and metastatic cancer (Figure 2C and Table S2B), mM organoids rapidly formed invasive mPDA within 1 month (Table S2C). The ability of organoid transplants to reproduce the discrete stages of disease progression contrasts with the rapid formation of advanced mPDA following transplantation of 2D cell lines (Figures S2A–S2C) (Olive et al., 2009).

Tumors derived from transplanted mT and mM organoids exhibited prominent stromal responses and resembled autochthonous tumors from KPC mice (Figure S2A) (Olive et al., 2009). This stromal response is often absent in tumors formed from 2D cell lines (Figure S2A) (Olive et al., 2009). Low vascular density and high vessel-to-tumor distance were also observed, demonstrating the close resemblance of the organoid transplantation models to autochthonous mPDA, in contrast to transplanted 2D cell lines (Figures S2A–S2C) (Olive et al., 2009).

Loss of heterozygosity (LOH) for *Trp53* has been reported as a common feature of mPDA based on studies of 2D cell lines (Hingorani et al., 2005). Therefore, we assayed for *Trp53* LOH in our murine 3D organoids. All mT organoids prepared from KPC tumors maintained expression of p16, did not exhibit *Trp53* LOH, and maintained a stable karyotype, whereas most mM organoids lost the wild-type *Trp53* allele and were aneuploid (Figures 2D, 2E, and S2D). We generated 2D cell lines from mT and mM organoids but found that mN and mP organoids were unable to propagate in 2D. mT1 was derived from a KC mouse PDA, lacks the mutant *Trp53* allele, and was also unable to propagate in 2D. All mT-derived 2D cell lines exhibited *Trp53* LOH and were aneuploid (Figures 2E and S2D).

To determine whether organoids are suitable for genetic cooperation experiments, shRNAs targeting p53 and p16/p19 were introduced into mP organoids (Figure S2E). Although the proliferation of mP organoids increased upon knockdown of either p53 or p16/p19 (Figure S2G), only p53 knockdown enabled 2D growth and colony formation (Figure S2F; data not shown). Also, only p53 knockdown promoted progression of mP organoid transplants to invasive carcinoma within 3 months (Figure S2H). This contrasts with a previous report that Kras mutation and biallelic loss of p16/p19 promoted mPDA (Aguirre et al., 2003, Bardeesy et al., 2006) and may reflect differences in the genetic system or the initiating cellular compartment. Nevertheless, the cooperation between p53 depletion and oncogenic Kras demonstrates that organoids are a facile system to evaluate genetic mediators of PDA progression.

Human Pancreatic Organoids Model PanIN to PDA Progression

We modified our culture conditions to support the propagation of human normal and malignant pancreatic tissues. Isolation of ductal fragments was not always feasible because some normal pancreatic tissue samples were predigested in preparation for islet transplantation. Therefore, we directly embedded digested material into Matrigel. This approach achieved an isolation efficiency of 75%–80% for human normal (hN) organoids (Figures 3A and S3 and Table S3). hN organoids require transforming growth factor β (TGF-β) pathway inhibitors (A83-01 and Noggin), R-Spondin1 and

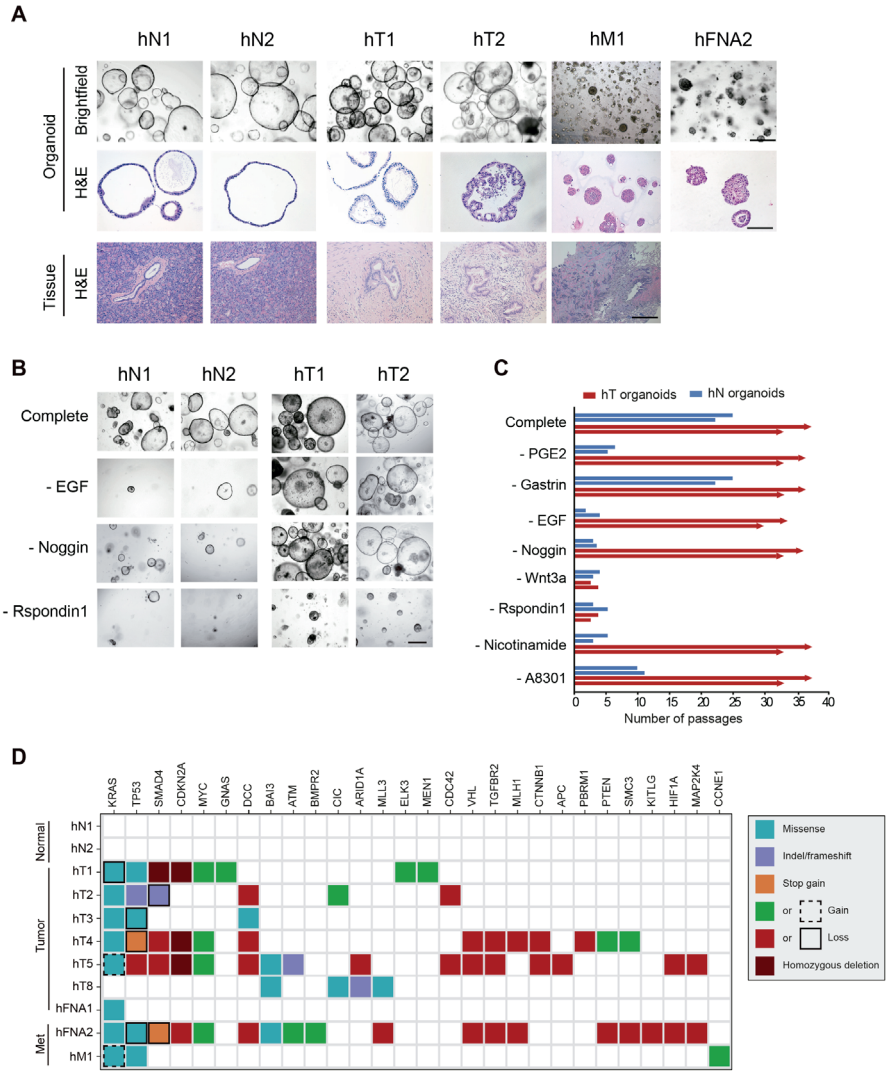


Figure 3. Human Pancreatic Ductal Organoids Recapitulate Features of Normal and Neoplastic Ducts. (A) Representative images (top) and H&E staining (middle) of human organoid cultures established from normal tissues (hN1-2), resected primary tumors (hT1-2), a resected metastatic lung lesion (hM1), and a fine-needle aspiration biopsy of a metastatic lesion (hFNA2). H&E staining of the resected tissues from which the organoids were derived (bottom). Scale bars, 500 µm (top), 250 µm (middle), and 500 µm (bottom). (B) Representative images of hN and hT organoids cultured for 2 weeks (1 passage) in human complete media or in human complete media lacking the indicated factors. Scale bars, 500 µm. (C) Number of passages hN and hT organoids could be propagated in the absence of the indicated factors. (D) Targeted sequencing analysis of human organoids. Genes altered in more than one sample and/or known to be mutated in PDA are shown. If multiple mutations were found in a gene, only one mutation per gene is shown. Color key for the type of genetic alterations is shown. Met indicates organoids derived from metastatic samples.

Wnt3a-conditioned media, EGF, and PGE2 for propagation (Figures 3B and 3C). Unlike mN organoids, which have unlimited propagation in culture, hN organoids ceased proliferating after 20 passages or ~6 months but could be cryopreserved.

We adapted the methods described above to accommodate the extensive desmoplastic reaction in freshly resected PDA specimens and generated human tumor-derived organoids (hT) (Figures 3A and S3 and Table S3). hT organoids could be passaged indefinitely and cryopreserved (Figure 3C). The establishment of hT organoids had efficiencies of 75% (n = 3/4) and 83% (n = 5/6) in the Netherlands and USA, respectively (Table S3). The first specimen that failed to generate an organoid culture was obtained from a patient that had undergone neo-adjuvant chemotherapy, and histologic examination of this specimen revealed extensive necrosis. The second specimen that did not generate an organoid culture was predominantly composed of stromal cells, without sufficient viable tumor cells to establish a culture. Although the hN organoids had a simple, cuboidal morphology, the hT organoids had differing degrees of dysplastic tall columnar cells, resembling low-grade PanINs (Figures 3A). hT organoids tolerated the withdrawal of certain growth factors from the media (Figures 3B and 3C).

85% of pancreatic cancer patients are ineligible for surgical resection of their tumors (Ryan et al., 2014). Therefore, we determined whether hT organoids could be generated from the limited amount of cellular material provided by endoscopic biopsies using fine needle aspirations (FNA). Initial attempts to generate organoids from FNA biopsies were hampered by loss of cellular material during digestion. Upon optimization of these conditions, human FNA biopsy organoids (hFNA) were generated from two specimens that were not dissociated prior to suspension in Matrigel (Figures 3A and S3 and Table S3). This approach is broadly applicable to PDA patients and enables serial sampling.

Targeted sequencing of 2,000 cancer-associated genes was performed on hN and hT organoids. As expected, no mutations were detected in the hN organoid cultures. These analyses identified oncogenic *KRAS* mutations in the majority of tumor-derived samples (n = 8), as well as mutations in *TP53* (n = 7), *SMAD4* (n = 5), and *CDKN2A* (n = 4) (Figure 3D and Table S4). We also noted amplification of known oncogenes, such as *MYC* (n = 4), and loss of tumor suppressors, including *TGFBR2* (n = 3) and *DCC* (n = 5). Importantly, the same *KRAS* mutations observed in several hT organoids were confirmed in the primary PDA from which they were derived (Table S4). The allele frequency of oncogenic *KRAS* variants in hT1–hT5 and hFNA2 ranged from ~50–100%. In contrast, the $KRAS^{G12V}$ allele frequency in hFNA1 was only 1% (Table S4), which may result from coexistence of wild-type ductal cells. Although *KRAS* mutations were not detected in hT8 (Figure 3D and Table S4), the presence

of mutations in known PDA genes (*ARID1A* and *MLL3*) suggests that hT8 contains malignant cells (Table S4).

To further characterize the cell types present in primary PDA organoids, we evaluated the expression of pancreatic lineage markers. hN and hT organoids expressed markers of ductal cells, but not other pancreatic lineages (Figure 4A). The karyotypes of hT organoids were highly aneuploid, whereas the hN organoids were predominantly and stably diploid (Figure 4B). The PDA-associated biomarker CA19-9 (Makovitzky, 1986) was also elevated in hT relative to hN organoids (Figure 4C). The hN and hT organoids are therefore reflective of normal and neoplastic human pancreatic ductal cells and offer a model system to explore pancreatic cancer biology in the more genetically complex background of human cancer.

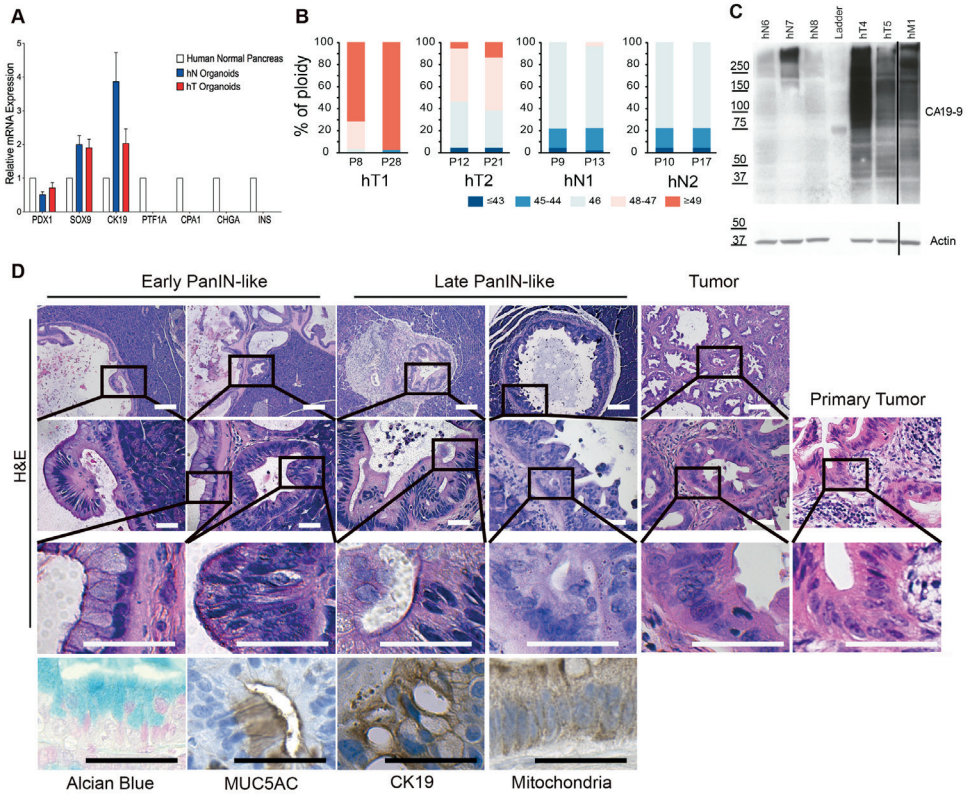


Figure 4. Molecular Characterization and Orthotopic Transplantation of Human Organoids. (A) qRT-PCR of pancreas lineage markers in hN (n = 3) and hT (n = 4) organoids. Mean expression levels were normalized to total pancreas. Error bars indicate SEMs. (B) Karyotyping of human organoids (2 hN, 2hT) at the indicated passages (P). (C) CA19-9 and actin levels in hN, hT, or hM organoids. The solid line indicates noncongruent lanes. (D) H&E, Alcian blue staining, and IHC of orthotopic hT2 transplants and the primary tumor. Scale bars, 200 μm (top two panels) and 50 μm (bottom two panels).

Following orthotopic transplantation into *Nu/Nu* mice, hN organoids produced normal ductal structures at low efficiency (n = 2/23), whereas hT organoids efficiently generated a spectrum of low- and high-grade, extraductal PanIN-like lesions within 1 month (n = 9/12) (Figures 4D and S4A and Table S4D). The hT-derived transplants initially formed well-defined hollow lesions lined by a single layer of columnar epithelial cells with apical mucin and basally located, relatively uniform nuclei. The nuclei were small and lacked the pleomorphism and hyperchromasia often seen in invasive PDA. These lesions progressed over several months to infiltrative carcinoma comprised of poorly defined and invasive glands (Figures 4D and S4A and Table S4). A prominent desmoplastic reaction was present in hT-derived PanIN-like structures and PDA, including the deposition of a collagen-rich stroma and the recruitment of αSMA-positive cells (Figure S4B). The mutation or loss of *TP53* or *SMAD4* in hT1 and hT2 was also detected by IHC in these tumors (Figure S4C and Table S4). Overall, hT organoids represent a transplantable model of human pancreatic cancer progression.

Gene Expression Analysis of Murine Pancreatic Ductal Organoids Implicates Candidate Genes in PDA Progression

The mouse organoids were prepared from syngeneic mice, offering the ability to discern gene expression changes in organoids and determine whether these changes correlate with PDA progression. We harvested RNA from mN (n = 7), mP (n = 6), and mT (n = 6) organoids and generated strand-specific RNA-sequencing (RNA-seq) libraries. Sequences were mapped to the mm9 version of the mouse genome, and relative transcript abundances (transcripts per million) of 29,777 mouse genes were determined (Table S5). Principal component analysis revealed that mN organoids were distinct from mP and mT organoids (Figure 5A and Table S5).

Genes whose levels differed significantly among mN, mP, and mT organoids were identified. 772 genes were found downregulated and 863 genes upregulated in mP relative to mN organoids (Figure 5B and Table S5). When mT organoids were compared to mN organoids, 2,721 genes were downregulated and 2,695 were upregulated. In addition, 823 genes were downregulated and 640 genes were upregulated in mT relative to mP organoids. Distinct patterns of gene expression were found in the data set (Figure 5C). The majority of genes differentially expressed in mP relative to mN organoids changed in a similar manner in mT relative to mN organoids (Figure 5D). However, a much larger cohort of genes changed in expression in mT relative to mN than in mP relative to mN organoids (Figure 5D), suggesting that mP organoids represent an intermediate state between mN and mT organoids.

The glycosyltransferase *Gcnt3* and putative protein disulfide isomerase *Agr2*

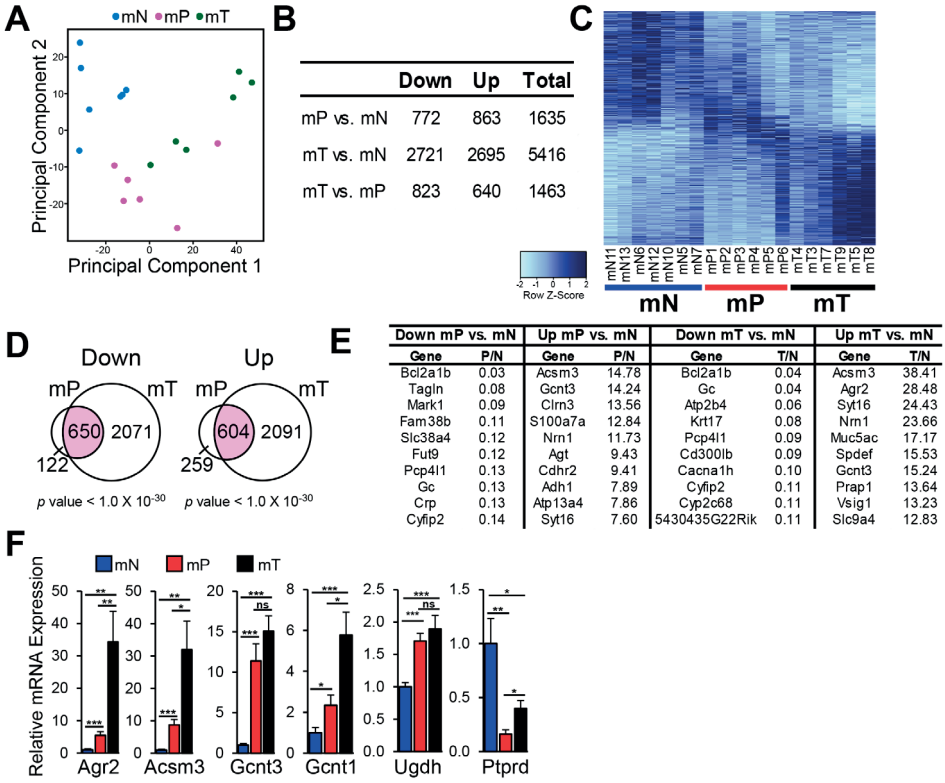


Figure 5. Gene Expression Analysis of Murine Organoids Reveals Genetic Changes Correlated with Pancreatic Cancer Progression. (A) Principal component analysis of gene expression data for mN, mP, and mT organoids. (B) The number of genes differentially expressed (DESeq adjusted *p* value < 0.05) among mN (n = 7), mP (n = 6), and mT (n = 6) organoids. (C) Heatmap showing relative expression levels using Z score normalization among mN, mP, and mT organoids. Color key of Z score is shown. (D) Venn diagrams show overlap of genes significantly differentially expressed in mP and mT relative to mN organoids. The *p* values for overlaps were determined by two-tailed Fisher's exact test. (E) Genes with the largest fold changes in mP or mT relative to mN organoids. (F) qRT-PCR validation of mN, mP and mT organoid gene expression changes. Values were normalized to mean levels in mN organoids. n = 8 mN, 7 mP, and 8 mT organoid cultures. Error bars indicate SEMs. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and ns, not significant by two-tailed Student's *t* test.

were among the most upregulated genes in both mP and mT organoids and have been demonstrated to be elevated in human PDA (Figure 5E) (Dumartin et al., 2011, Zhao et al., 2014). The most upregulated gene in both mP and mT relative to mN organoids was the acyl-CoA synthetase *Acsm3* (Figure 5E). RNA-seq results were confirmed by qRT-PCR for 35 out of 40 genes (Table S5), including the upregulation of *Agr2*, *Acsm3*, *Gcnt1*, *Gcnt3*, and *Ugdh* and the downregulation of *Ptprd* in mP and mT organoids (Figure 5F and Table S5). Among the genes upregulated in mP and mT relative to mN organoids, *Gcnt1*, *Gcnt3*, *Acsm3*, *Agr2*, *Syt16*, *Nt5e*, and

Ugdh were upregulated following the Ad-Cre-induced expression of oncogenic *Kras*^{G12D}, suggesting that these genes are activated downstream of mutant *Kras*^{G12D} (Figure S5A). To determine whether organoid RNA-seq profiles resembled gene expression patterns *in vivo*, we compared our organoid RNA-seq data to a published transcription profile of murine pancreatic tumors upon *Kras*^{G12D} inactivation (Ying et al., 2012). Genes differentially expressed upon inactivation of oncogenic Kras overlapped significantly with those up or downregulated in mP or mT relative to mN organoids (Figure S5B). These analyses demonstrate the ability of the organoid system to identify molecular alterations associated with PDA progression.

Proteomic Alterations in Murine Pancreatic Ductal Organoids Predict Pathways Associated with PDA Progression

As an orthogonal method to investigate molecular alterations in murine pancreatic organoids, we characterized the global proteomes of mN (n = 5), mP (n = 4), and mT (n = 5) organoids. Protein lysates were processed using amine-reactive isobaric tags for relative and absolute quantification (iTRAQ) mass spectrometry (Wiese et al., 2007). Samples were run in four 8-plex experiments and merged using an approach that normalizes the data to common samples included across all experiments (Extended Experimental Procedures). Upon merging, 6,051 unique protein isoforms were quantified in all samples. We applied linear regression modeling on the normalized intensity peak values and identified 710 protein isoform expression changes between mN and mP organoids (Figure 6A). 1,047 protein isoforms changed expression between mN and mT organoids, and 63 differentially expressed proteins were identified between mP and mT (Figure 6A). The relatively small number of protein expression changes identified between mP and mT organoids reflects their biological similarity (Figure S6A).

mN organoids showed unique proteomic profiles from their mP and mT counterparts (Figures 6B and 6C). To compare the proteomic and RNA-seq data, we collapsed the unique protein isoforms into their corresponding 4,155 genes. Some protein expression changes (e.g., 123/150 for downregulated and 96/151 for upregulated mP proteins) did not reflect corresponding transcriptional changes, indicating that protein stability may play a role in cancer progression, particularly in mP organoids (Figure 6D). Nonetheless, the proteomic data validated many of the expression changes identified by RNA-seq (Figure 6D), including upregulation of Gcnt3, Agr2, and Ugdh (Table S6). Additionally, of the 1,599 genes whose expression levels changed in mT relative to mN organoids that were measured by mass spectrometry, 301 (19%) showed corresponding protein changes (Figure 6D).

Gene Set Enrichment Analysis (GSEA) on the RNA-seq and proteomic data

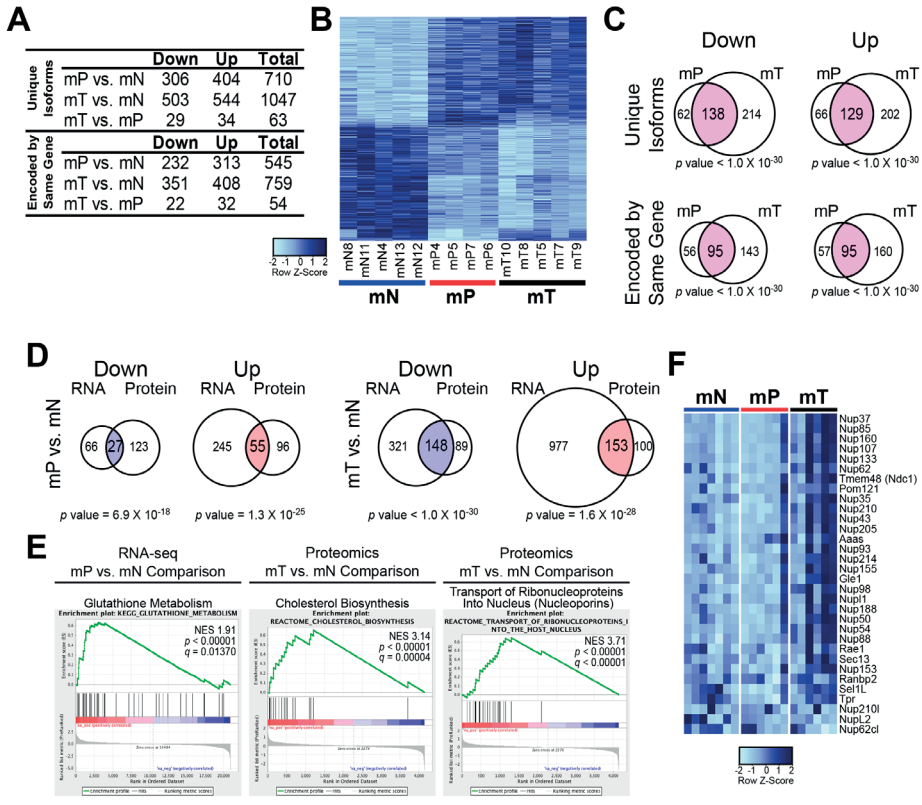


Figure 6. Proteomic Profiling of Murine Organoids Uncovers Molecular Pathways Linked to Pancreatic Cancer Progression. (A) Protein expression changes by iTRAQ proteomic analysis of murine organoids. Both unique protein isoforms and protein isoforms encoded by the same gene are included (adjusted *p* value < 0.1 by linear regression analysis). (B) Heatmap of unique protein isoforms that differ (adjusted *p* value < 0.05) among mN, mP, and mT organoids. Color key of the Z score is shown. (C) Venn diagrams showing overlaps between proteins differentially expressed (p < 0.05) in mP and mT relative to mN organoids. *p* values for overlaps were determined by two-tailed Fisher's exact test. (D) Venn diagrams showing overlaps between genes and proteins found differentially expressed by RNA-seq and proteomic analyses (adjusted $p < 0.05$). *p* values for the overlaps were determined by two-tailed Fisher's exact test. (E) Molecular pathways found enriched by GSEA analysis of RNA-seq and proteomic data. Normalized enrichment scores (NESs), *p* and *q* values are shown. (F) Heatmap showing relative gene expression levels of nucleoporins in mN, mP, and mT organoids determined by RNA-seq. Color key of the Z score is shown.

(Subramanian et al., 2005) revealed elevated expression of genes and proteins involved in glutathione metabolism and biological oxidations in mP relative to mN organoids (Figures 6E, S6B, and S6C and Table S7), which is consistent with elevations in reactive oxygen species metabolism previously reported in *Kras*^{G12D} cells (DeNicola et al., 2011, Ying et al., 2012). Enrichment of proteins involved in glutathione metabolism was also found in mT relative to mN organoids (Table S7). Additionally, we identified a significant positive enrichment of proteins involved in the steroid

biosynthesis, cholesterol biosynthesis, one carbon pool by folate, and pyrimidine metabolism pathways (Figures 6E, S6B, and S6C and Table S7), which is consistent with an earlier report (Ying et al., 2012). Similar pathways were enriched in mP relative to mN organoids (cholesterol biosynthesis, one carbon pool by folate, and pyrimidine metabolism) (Figures S6B and S6C and Table S7), whereas fatty acid metabolism and TCA cycle/respiratory electron transport pathways were downregulated (Figure S6C and Table S7). The increase in anabolic and decrease in catabolic pathways suggest that complex alterations in fatty acid and nucleotide metabolism occur during PDA progression.

Interestingly, we also found broad upregulation of the nucleoporin family at both the RNA and protein levels in the mT relative to mN organoids (Figures 6E and 6F and Table S6). The individual nucleoporins NUP214, NUP153, and NUPL1 were previously identified in shRNA dropout screens in PDA cell lines (Cheung et al., 2011, Shain et al., 2013). Furthermore, amplification of NUP153 was detected in one human PDA cancer cell line, and elevation of NUP88 was detected in human primary PDA (Cheung et al., 2011, Gould et al., 2000, Shain et al., 2013). This systematic analysis of molecular alterations in pancreatic organoids implicates nuclear transport as a pathway correlated with pancreatic cancer progression.

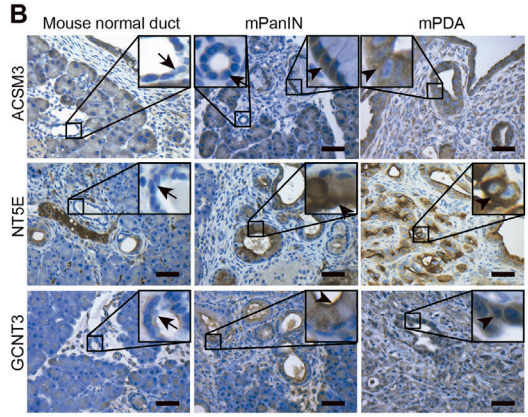
In Vivo Mouse and Human Validation of Candidates Associated with PDA Progression in Organoids

To demonstrate that the mouse organoid culture system represents a biological resource for the accurate discovery of genes associated with PDA progression, we selected 16 genes upregulated in mT organoids for validation in primary tissue specimens by IHC and immunofluorescence (IF) (Figure 7A). These 16 genes included enzymes, membrane proteins, structural proteins, and secreted ligands, which could represent candidate biomarkers and therapeutic targets. Of the 14 antibodies that generated a detectable signal on murine pancreatic tissue sections, 13 antibodies confirmed the increased expression of the candidate protein in mPanIN and mPDA lesions in concordance with the RNA-seq and proteomic data (Figures 7A, 7B, and S7A). 11 of the 13 candidate antibodies were compatible for evaluation in human tissues, and 7 of these candidates were upregulated in human PDA when compared to normal pancreatic ductal tissues (Figures 7C, 7D, and S7B). The high expression of many of these markers was recapitulated in orthotopic transplants of hT organoids into *Nu/Nu* mice (Figure 7C). These results indicate that the organoid culture system accurately models PDA progression and can serve as a resource for the discovery and genetic dissection of pathways driving human pancreatic tumorigenesis.

A

Gene name	Mouse normal duct	mPanIN	mPDA
ACSM3	+	++	+++
NT5E	-	++	+++
UGDH	-	+	+++
GCNT1	-	+++	++
GCNT3	+	++	++
FGF1	-	++	++
AGT	-	+	++
TNS4	+	++	+++
LIN28	+	+++	++
ADH1	-	+	++
DDC	-	+	++
PCSK6	-	+	+
GCAT	+++	++	++
AGR2	+	+++	+++

B



C

Gene name	Human normal duct	Human PDA	hT Transplants
ACSM3	+	+++	+++
NT5E	+	+++	+++
GCNT3	+	+++	++
UGDH	-	+++	+++
GCNT1	+	++	+
ADH1	++	+++	+
TNS4	+	+++	++

D

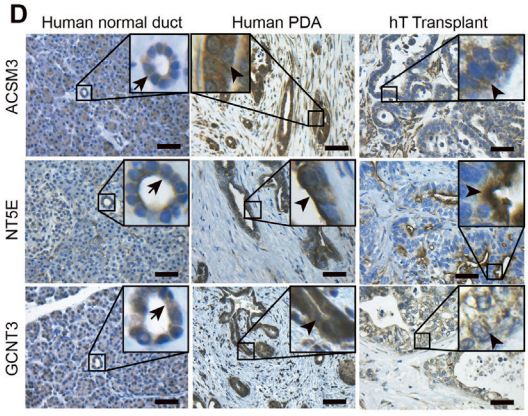


Figure 7. Increased Levels of ACSM3, NT5E, and GCNT3 Correlate with Mouse and Human PDA Progression. (A) IHC analysis of 14 candidate genes in mouse adjacent normal ducts, mPanIN and mPDA. Differential expression is indicated as − (negative), + (weak), ++ (moderate), or +++ (strong). Only the ductal component of the normal pancreas was scored. (B) IHC analysis of Acsm3, Nt5e, and Gcnt3 in mouse normal ducts, mPanIN and mPDA tissues. Arrow indicates adjacent normal ducts in mPanIN tissues. Arrowhead indicates mPanIN or mPDA. Scale bars, 50 μm. (C) IHC analysis of seven candidate genes in human normal pancreas, hT orthotopic transplants, and PDA tissues. Differential expression is indicated as − (negative), + (weak), ++ (moderate), or +++ (strong). Only the ductal component of the normal pancreas was scored. (D) IHC analysis of ACSM3, NT5E, and GCNT3 in human normal pancreas and PDA tissues. Arrow indicates normal ducts, and arrowhead indicates PDA. Scale bars, 50 μm.

DISCUSSION

We have established pancreatic organoids as a tractable and transplantable system to probe the molecular and cellular properties of neoplastic progression in mice and humans. In contrast to prior reports (Agbunag and Bar-Sagi, 2004, Rovira et al., 2010, Seaberg et al., 2004), our culture conditions prevent the rapid exhaustion of normal ductal cells *in vitro* and generate a normal ductal architecture following orthotopic transplantation. Importantly, the ability to passage and transplant both normal and

neoplastic ductal cells enables a detailed analysis of molecular pathways and cellular biology that is not possible when neonatal pancreatic fragments are propagated in air-liquid interfaces or when induced pluripotent cells are employed (Agbunag and Bar-Sagi, 2004, Kim et al., 2013, Li et al., 2014). Our finding that nucleoporins are broadly upregulated in the neoplastic murine organoids, coupled with the known associations of nucleoporins to cell proliferation and cell transformation, presents a class of proteins to investigate in pancreatic cancer progression (Gould et al., 2000, Köhler and Hurt, 2010). Furthermore, the ability to systematically characterize human pancreatic cancer organoids that lack KRAS mutations, such as hT8, will reveal driver genes for PDA. Finally, because organoids can be readily established from small patient biopsies, they should hasten the development of personalized approaches for pancreatic cancer patients.

ACKNOWLEDGEMENTS

We thank Peter Kapitein and Jan Schuurman from Inspire 2 Live for helping to establish the collaboration between D.A.T. and H.C. We also thank H. Begthel and J. Korving for technical assistance. This work was performed with assistance from the CSHL Proteomic, Histology, DNA Sequencing, Antibody, and Bioinformatics Shared Resources, which are supported by the Cancer Center Support Grant 5P30CA045508. D.A.T. is a distinguished scholar of the Lustgarten Foundation and Director of the Lustgarten Foundation-designated Laboratory of Pancreatic Cancer Research. D.A.T. is also supported by the Cold Spring Harbor Laboratory Association, the Carcinoid Foundation, PCUK, and the David Rubinstein Center for Pancreatic Cancer Research at MSKCC. In addition, we are grateful for support from the following: Stand Up to Cancer/KWF (H.C.), the STARR foundation (I7-A718 for D.A.T.), DOD (W81XWH-13-PRCRP-IA for D.A.T.), the Sol Goldman Pancreatic Cancer Research Center (R.H.H.), the Italian Ministry of Health (FIRB - RBAP10AHJ for V.C.), Sociedad Española de Oncología Médica (SEOM for M.P.S.), Louis Morin Charitable Trust (M.E.F.), the Swedish Research Council (537-2013-7277 for D.Ö.), The Kempe Foundations (JCK-1301 for D.Ö.) and the Swedish Society of Medicine (SLS-326921, SLS-250831 for D.Ö.), the Damon Runyon Cancer Research Foundation (DRG-2165-13 for I.I.C.C.), the Human Frontiers Science Program (LT000403/2014 for E.E.), the Weizmann Institute of Science Women in Science Award (E.E.), the American Cancer Society (PF-13-317-01-CSM for C.M.A.A.), the Hearst Foundation (A.H.S.), and the NIH (5P30CA45508-26, 5P50CA101955-07, 1U10CA180944-01, 5U01CA168409-3, and 1R01CA190092-01 for D.A.T.; CA62924 for R.H.H.; CA134292 for S.D.L.; 5T32CA148056 for L.A.B. and D.D.E.; and CA101955 UAB/UMN SPORE for L.A.B.). In addition, S.F.B. and M.H. are supported by KWF/PF-HUBR 2007-3956, A.G is supported by EU/232814-StemCellMark, and R.G.J.V. is supported by GenomiCs.nl (CGC). M.J., R.B., and E.C. are supported by the CancerGenomics.nl (NWO Gravitation) program. Ralph Hruban receives royalty payments from Myriad Genetics for the PalB2 inventions. Hans Clevers and Meritxell Huch have patents pending and granted on the organoid technology.

AUTHOR CONTRIBUTIONS

S.F.B. initiated the project, developed the methods for isolating mouse and human organoids, and characterized human organoids (Figures 1A, 3, 4B, 4D, and S4C and Tables S3 and S4). C.-I.H. developed transplantation models for organoids and performed shRNA knockdown and histological and karyotypic analyses (Figures 1A, 1E, 1G, 2A, 2C–2E, 4D, 7, S1A–S1C, S2A–S2F, S2H, S4A, S4B, S7A, and S7B and Tables S1, S2, and S4). L.A.B. performed RNA-seq on mouse organoids and analyzed RNA-seq and proteomic data (Figures 5, 6, S5B, and S6 and Tables S5, S6, and S7). I.I.C.C. conducted proteomic evaluation of mouse organoids and analyzed proteomic data (Figures 6 and S6C). D.D.E. developed mouse organoid methods and evaluated CA19-9 levels in human organoids (Figures 1A, 2A, 4C, S3, and S6 and Table S6). V.C.

developed human organoid methods, performed molecular analyses of organoids, and prepared material for DNA-sequencing and sequencing of Kras (Figures 1C, 1D, 3A, 3D, 4A, S1D–S1F, and S5A and Tables S3, S4, and S5). M.J. performed and analyzed the DNA sequencing of human organoids (Figure 3D and Table S4). Mouse and human organoid preparation and characterization was performed by M.P.-S., H.T., M.S.S., T.O., D.Ö., A.H.-S., C.M.A.-A., M.L., E.E., B.A., M.E.F., G.N.Y., G.B., B.D., B.C., K.W., K.H.Y., Y.P., M. Huch, A.G., F.H.M.M., and S.D.L. Sequencing analyses were performed by Y.H., Y.J., M. Hammell, I.J.N., E.C., and R.v.B. Pathological analyses were performed by G.J.O., R.H.H., D.S.K., O.B., and C.I.-D. Surgical resections and tissue dissection were performed by I.Q.M. and I.H.B.R. Proteomic development was performed by D.J.P., K.D.R., and J.P.W. Overall study management was conducted by D.A.T., H.C., and R.G.J.V. S.F.B., D.D.E., L.A.B., M.E.F., C.H., H.T., V.C., M.P.S., R.G.J.V., H.C., I.I.C.C., and D.A.T. contributed to manuscript writing.

METHODS

Animals

Trp53$^{+/LSL\text{-}R172H}$, *Kras*$^{+/LSL\text{-}G12D}$, and *Pdx1-Cre* strains in C57Bl/6 background were interbred to obtain *Pdx1-Cre; Kras*$^{+/LSL\text{-}G12D}$ (KC) and *Pdx1-Cre; Kras*$^{+/LSL\text{-}G12D}$*; Trp53*$^{+/LSL\text{-}R172H}$ (KPC) mice (Hingorani et al., 2005). The *R26*$^{LSL\text{-}YFP}$ strain was interbred to get the desired genotype. C57Bl/6 and athymic *Nu/Nu* mice were purchased from Charles River Laboratory and Jackson Laboratory. All animal experiments were conducted in accordance with procedures approved by the IACUC at Cold Spring Harbor Laboratory (CSHL).

Murine Pancreatic Ductal Organoid Culture

Detailed procedures to isolate normal pancreatic ducts have been described previously (Huch et al., 2013a). In brief, normal and preneoplastic pancreatic ducts were manually picked after enzymatic digestion of pancreas with 0.012% (w/v) collagenase XI (Sigma) and 0.012% (w/v) dispase (GIBCO) in DMEM media containing 1% FBS (GIBCO) and were seeded in growth-factor-reduced (GFR) Matrigel (BD). For tumors and metastases, bulk tissues were minced and digested overnight with collagenase XI and dispase and embedded in GFR Matrigel.

Human Specimens

Pancreatic cancer tissues and adjacent normal pancreas were obtained from patients undergoing surgical resection at the University Medical Centre Utrecht Hospital, Memorial Sloan-Kettering Cancer Center (MSKCC), MD Anderson Cancer Center (MDACC), and Weill Cornell Medical College (WCMC). Normal pancreatic tissue was also obtained from islet transplant programs at the University of Illinois at Chicago and University of Miami Miller School of Medicine. All human experiments were approved by the ethical committees of the University Medical Centre Utrecht or the IRBs of MSKCC, MDACC, WCMC, and CSHL. Written informed consent from the donors for research use of tissue in this study was obtained prior to acquisition of the specimen. Samples were confirmed to be tumor or normal based on pathological assessment.

Human Pancreatic Tumor and Normal Organoid Culture

Tumor tissue was minced and digested with collagenase II (5 mg/ml, GIBCO) in human complete medium (see below) at 37°C for a maximum of 16 hr. The material was further digested with TrypLE (GIBCO) for 15 min at 37°C, embedded in GFR Matrigel, and cultured in human complete medium (AdDMEM/F12 medium supplemented with HEPES [1×, Invitrogen], Glutamax [1×, Invitrogen], penicillin/streptomycin [1×, Invitrogen], B27 [1×, Invitrogen], Primocin [1 mg/ml, InvivoGen], N-acetyl-L-cysteine [1 mM, Sigma], Wnt3a-conditioned medium [50% v/v], RSPO1-conditioned medium [10% v/v, Calvin Kuo], Noggin-conditioned medium [10% v/v] or recombinant protein [0.1 μg/ml, Peprotech], epidermal growth factor [EGF, 50 ng/ml, Peprotech], Gastrin [10 nM, Sigma], fibroblast growth factor 10 [FGF10, 100 ng/ml, Preprotech], Nicotinamide [10 mM, Sigma], and A83-01 [0.5 μM, Tocris]). Normal samples were processed as above, except that the collagenase digestion was done for a maximum of 2 hr in the presence of soybean trypsin inhibitor (1 mg/ml, Sigma). Following digestion, cells were embedded in GFR Matrigel and cultured in human complete medium with the addition of PGE2 (1 μM, Tocris).

Transplantation

For the orthotopic engraftment of mouse and human organoids, mice were anesthetized using Isoflurane, and Ketoprofen (5 mg/kg), which was subcutaneously administered. An incision was made in the left abdominal side. Organoids (approximately $1x10^6$ cells/mouse) were prepared either from cultures or from cryopreserved stocks. In the case of cryopreserved stocks, organoids were thawed in HEPES (1x, Invitrogen), Glutamax (1x, Invitrogen), and penicillin/streptomycin (1x, Invitrogen), in AdDMEM/F12 media and stabilized for 4 hr at 37°C in 5% CO_2. Organoids were washed with ice-cold PBS, physically broken into pieces by triturating through fire-polished glass Pasteur pipettes, and finally resuspended in 50 µl of Matrigel (Matrigel, BD) diluted 1:1 with cold PBS. The organoid suspension was injected into the tail region of the pancreas using insulin syringes (29 Gauge). Successful injection was verified by the appearance of a fluid bubble without signs of intraperitoneal leakage. The abdominal wall was sutured with absorbable Vicryl suture (Ethicon), and the skin was closed with wound clips (CellPoint Scientific Inc.). Mice were euthanized at the indicated time points.

Proliferation Assay

Organoids were dissociated into single cells by first triturating them in media through a fire-polished glass pipette, and then by enzymatic dissociation with 2 mg/ml dispase dissolved in TrypLE (Life Technologies), until the organoids appeared as single cells under the microscope. Cells were counted, and diluted to 10 cells/µL in a mixture of complete media, Rho Kinase inhibitor Y-27632 (10.5 µM final concentration, Sigma), and Growth factor-reduced Matrigel (GFR-Matrigel, 10% final concentration). 100 µl of this mixture (1000 cells per well) was plated in 96-well plates (Nunc), whose wells had been previously coated with a bed of GFR-Matrigel to prevent attachment of the cells to the bottom of the plate. Cell viability was measured every 24 hr using the CellTiter-Glo assay (Promega) and SpectraMax I3 microplate reader (Molecular Devices). Five replicate wells per time point were used. Luminescence data were analyzed with GraphPad Prism.

Karyotyping

Colcemid (50 mg/ml) was added to organoid cultures, and cultures were incubated for at least 1 hr at 37°C in 5% CO_2. After dissociating pancreatic organoids into single cells as described for the proliferation assay, cells were incubated with a hypotonic solution (KCl, 0.075 M) for 20 min. Cells were then fixed with methanol/glacial acetic acid (3:1) and dropped onto glass slides. Giemsa staining (KaryoMax, GIBCO) was performed according to manufacturer's recommendations. At least 20 or 50 metaphase-arrested cells were counted for mouse or human cells, respectively.

Retroviral Production and Infection in Organoids

PGK-Neo-IRES-EGFP or PGK-Neo-IRES-nuclear EGFP retroviruses were produced in ecotropic Phoenix cells, concentrated with RetroX concentrator (Clontech), and resuspended with organoid culture media supplemented with Y-27632 (10 mM, Sigma). To knock-down expression of p53 and p16/p19, short hairpin RNAs against p53 (#1224) and p16/p19 (#478) were inserted in a miR-30 backbone driven by the MSCV promoter (Premsrirut et al., 2011). Organoid infections were performed as described previously (Koo et al., 2012). In brief, $5x10^4$ single cells were resuspended with concentrated retrovirus and spinoculated at 600 RCF for 1 hr at room temperature. Two days after infection, cells were treated with 1 mg/ml G418 (GIBCO) for selection.

Colony Formation Assay

Organoids were dissociated to single cells as described above. 2,000 cells were plated into 6 well plates in triplicate and cultured for 2 weeks in DMEM supplemented with 10% FBS. Cells were fixed with 95% ethanol and stained with 0.5% crystal violet.

Adenoviral Infection

Ductal organoids were prepared from 2-3 month-old *KRAS$^{+/LSL-G12D}$; R26$^{LSL-YFP}$* mice and propagated in GFR-Matrigel for at least 3 passages before infection with 500 pfu/cell of AdCMV-Cre or AdCMV-Blank

(University of Iowa Gene Transfer Vector Core Facility). Spinoculation was performed as described above.

Histology

Tissues were fixed in 10% neutral buffered formalin and embedded in paraffin. Sections were subjected to H&E, Alcian Blue, Nuclear Fast Red, and Masson's Trichrome staining as well as immunohistochemical staining. The following primary antibodies were used for immunohistochemical staining: GFP (D5.1, Cell Signaling) 1:200; CK19 (TromaIII, developed by Rolf Kemler, Max-Planck Institute of Immunobiology, Freiberg, Germany, and obtained from the Hybridoma Bank at the University of Iowa, Iowa City, Iowa, USA) 1:1000; Muc5ac (45M1, Abcam) 1:400; human-specific Mitochondria (MAB1273, Millipore) 1:300; human-specific CAM5.2 (B&D) 1:100; p53 (Thermo Scientific, clone D07_DP53-12) 1:2000; SMAD4 (Santa Cruz, clone B7) 1:300; Smooth Muscle Actin alpha (Ab5694, Abcam); Collagen I (Ab34710, Abcam) 1:100; Ki-67 (Ab15580, Abcam) 1:500; CD31 (Ab28364, Abcam) 1:50, GCNT3 (PA5-24455, Thermo Scientific), 1:200; TNS4 (sc-98530, Santa Cruz), 1:500; AGT (R&D, AF6966), 1:50; ADH1 (Ab108203, Abcam), 1:500; GCNT1 (sc-130143, Santa Cruz), 1:50; LIN28A (NBP1-49537, Novus), 1:500; FGF1 (sc-1884, Santa Cruz), 1:100; PCSK6 (ab110144, Abcam), 1:200; DDC (ab3905, Abcam), 1:500; NT5E (13160, Cell Signaling), 1:200; GCAT (ab181094, Abcam), 1:200; UGDH (sc-137058, Santa Cruz), 1:200; ACSM3 (10168-2-AP, Protein Tech), 1:500; AGR2 (13062, Cell Signaling), 1:100; and E-cadherin (610181, BD Biosciences), 1:500. Staining intensity was semiquantitatively scored as: - (negative), + (weak), ++ (moderate), or +++ (strong). Only the ductal component of the normal pancreas was scored. Measurement of the mean vascular density and distance between tumor cells and vessels was performed as described previously (Olive et al., 2009).

Genotyping

Kras and *p53* 1loxP PCR was performed as described previously (Hingorani et al., 2005, Olive et al., 2004).

Western Blot Analysis and Kras GTP Pull-Down Assay

Standard techniques were employed for immunoblotting of mouse organoids. Organoids were quickly harvested using cold PBS on ice. Organoids were then lysed with 25 mM HEPES, pH 7.5; 150 mM NaCl; 1% NP-40; 10 mM MgCl2; 1 mM EDTA; 2% Glycerol. Protein lysates were separated in 4%–12% Bis-Tris NuPage gels (Life Technologies). Western blots were probed with the following antibodies: p53 (FL-393, Santa Cruz), p16 (H-156, Santa Cruz), SMAD4 (EP618Y, Abcam); Tubulin (2148, Cell Signaling); phospho-ERK1/2 (4370, Cell Signaling), pan-ERK1/2 (4695, Cell Signaling), phospho-Akt (4060, Cell Signaling), pan-Akt (4685, Cell Signaling), phospho-ribosomal S6 (4858, Cell Signaling), and S6 Ribosomal Protein (Cell Signaling #2317). The levels of Kras-GTP were determined using the Kras activation assay (Cell Biolabs) according to the manufacturer's instructions. The following antibodies were used: mouse Kras (Santa-Cruz); pan-Ras (Cell signaling).

Human organoid immunoblots were carried out on lysates from hN and hT organoids cultured in the presence of PGE2. The organoids were harvested by incubating the organoids in Cell Recovery Solution (Corning), rotating for 1 – 2 hr at 4°C. Organoids were washed two times in ice cold PBS and then lysed using 1% Triton X-100; 150 mM NaCl; 5 mM EDTA; 50 mM Tris, pH 7.5; and protease inhibitors (Roche). Protein lysates were separated on a 4%–12% Bis-Tris NuPage gel (Life Technologies). Western blots were probed with CA19-9 (1116-NS-19-9 Hybridoma) and Pan-Actin (D18C11, Cell Signaling) antibodies.

Quantitative RT-PCR

RNA was extracted from cell cultures or freshly isolated tissues using TRIzol reagent (Invitrogen), followed by column-based purification with the PureLink RNA Mini Kit (Ambion). cDNA was synthesized using 1 μg of total RNA and TaqMan Reverse Transcription Reagents (Applied Biosystems). All targets were amplified (40 cycles) using gene-specific Taqman primers and probe sets (Applied Biosystems) on a 7900HT Real time-PCR instrument (Applied Biosystems). Relative gene expression quantification was performed using the ΔΔCt method with the Sequence Detection Systems Software, Version 1.9.1 (Applied Biosystems). Expression levels were normalized by *Hprt*.

RNA-Sequencing of Murine Organoids

For each organoid line, 4-6 wells of organoids from a 24 well plate were harvested in 1 ml of TRIzol reagent and flash-frozen. All lines were at passage 3 or 4 post-isolation. RNA was extracted using the TRIzol Plus RNA Purification Kit (LifeTechnologies) per manufacturer's instructions. RNA samples were treated on column with PureLink DNase (Life Technologies). The quality of purified RNA samples was determined using a Bioanlyzer 2100 (Agilent) with an RNA 6000 Nano Kit. RNAs with RNA Integrity Number (RIN) values greater than 9.0 were used to generate sequencing libraries. Libraries were generated from 1 µg of total RNA using a TruSeq Stranded Total RNA Kit with Ribo-zero human/mouse/rat (Illumina # RS-122-2201) per manufacturer's instructions. For the final amplification, 11 cycles of PCR were used. Libraries were quality checked and quantified using a Bioanalyzer 2100 (Agilent) with a DNA 1000 Kit. Equimolar amounts of libraries were pooled and subjected to paired-end, 101 base-pair sequencing at the Cold Spring Harbor DNA Sequencing Next Generation Shared Resource using an Illumina HiSeq 2000. Two libraries were pooled per lane of sequencing.

RNA-Sequencing Data Analysis

The quality of reads was assessed using the FASTQC program (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Reads were mapped to the mm9 version of the mouse genome and transcript abundances were estimated using the RSEM program (version 1.2.11) (Li and Dewey, 2011), which internally calls the Bowtie alignment program (v1.1.0) (Langmead et al., 2009). Up to 3 mis-matches were allowed in the first 25 base-pairs, and no more than 200 multiple alignments were allowed. Differential expression targets were determined by using the DESeq2 (v1.4.1) program (Anders and Huber, 2010) (http://dx.doi.org/10.1101/002832,http://dx.doi.org/10.1101/002832) available in R/Bioconductor (R version 3.1.0). An adjusted p value < 0.05, determined by DESeq2, was used to identify differentially expressed genes.

Additional RNA-Seq and Proteomic Analysis

Principal component analysis was performed using R/Bioconductor functions. Pathway enrichment analysis was performed on a set of curated canonical pathways from KEGG, Reactome and Biocarta available in the molecular signatures database (MSigDB) by using the GseaPreranked tool available in GSEA software (Mootha et al., 2003, Subramanian et al., 2005). Heatmaps were generated using the gplots2 package in R. For Venn Diagrams, the significance of the overlaps was determined using a two-tailed Fisher's exact test. For comparison of our data with the expression study of Ying et al., data from the latter study were analyzed using GEO2R, and genes with at least one probe with a p value < 0.05 were considered significantly up- or downregulated. Only genes measured in both studies were compared.

Mutation Analysis of the Cancer Minigenome

The samples were analyzed in two separate batches. Batch 1 contained samples hT1, hT2, hT3, hN1 and hN2. Single nucleotide variants (SNVs) and small insertions or deletions (indels) were detected by targeted sequencing of a designed "Cancer mini-genome" version 2 consisting of the exons of ~2000 cancer-associated genes (gene list is available upon request) (Vermaat et al., 2012). Barcoded fragment libraries were prepared from ~500ng of isolated DNA as previously described (Harakalova et al., 2011). Libraries were pooled and subsequently enriched in-solution for the Cancer mini-genome (Vermaat et al., 2012) using the Sureselect target enrichment system (Agilent). Enriched libraries were sequenced to a median depth of ~250x with paired-end runs using a SOLiD 5500xl sequencer (Life Technologies).

Sequence reads were mapped to the human reference genome (GRCh37/hg19), using the Burrows-Wheeler Aligner (BWA) v0.5.9 mapping tool (Li and Durbin, 2009) with settings '-c –l 25 –k 2 –n 10'. Variant calling was performed using a custom pipeline, which identifies variants with at least 10x coverage, an allele frequency of 0.15, support from at least 4 independent reads and multiple (> 2) occurrences in the seed sequence (e.g., the first 25 most accurately mapped bases of the read). All the identified candidate variant positions were subsequently genotyped in a multi-sample format using SAMtools v.0.1.18 mpileup (Li et al., 2009), to enable detection of possible low-frequency variants.

Batch 2 contained samples hFNA1, hFNA2, hT4, hT5, hT6, hT7, hT8, and hM1. SNVs and indels were detected by targeted sequencing of a designed "Cancer mini-genome" version 3 consisting of the exons of ~2000 cancer-associated genes (gene list is available upon request) (Vermaat et al., 2012).

Barcoded fragment libraries were prepared from 100-500ng of isolated DNA according to the KAPA DNA library preparation protocol for Illumina (KAPA Biosystems). Libraries were pooled and subsequently enriched in-solution for the Cancer mini-genome (Vermaat et al., 2012) using the Sureselect target enrichment system (Agilent). Enriched libraries were sequenced to a median depth of ~350x with single-end runs using an Illumina Nextseq 500 sequencer (Illumina).

Sequence reads were mapped to the human reference genome (GRCh37/hg19), using the Burrows-Wheeler Aligner (BWA) Maximal Exact Matches (MEM) v0.7.5a mapping tool (Li and Durbin, 2009, Li et al., 2009) with settings '-c 100 -m'. Variant calling was performed using the GATK haplotype caller v.3.2-2 with 'best practices' settings. Variants calls that did not pass the filter were discarded.

For all samples, candidate variants that overlapped with positions in the Single Nucleotide Polymorphism Database (dbSNP 137.b37) were filtered out, unless these variant positions also overlapped with mutations in the Catalogue Of Somatic Mutations In Cancer database (COSMIC). Second, candidate variants that overlapped variants present in the Genome of the Netherlands (GoNL) were filtered out (Genome of the Netherlands, 2014). Furthermore, we only considered variants with a predicted effect on the amino acid sequence (such as nonsynonymous mutations), a base coverage of at least 20x and an alternative allele frequency of at least 0.2. Additionally, SNVs that were called in > 60% of the samples (per sequencing run) were not considered. Indels that were called in more than one sample were also filtered out. Finally, we only considered mutations with a COSMIC ID and novel mutations in known pancreatic cancer genes present in both Cancer mini-genome designs (Campbell et al., 2010, Jones et al., 2008, Wu et al., 2011) (Table S4). Mutations in four known PDAC genes (*KRAS*, *P53*, *SMAD4*, and *CDKN2A*) that showed allele frequencies of the variants below 0.2 were manually inspected using the IGV tool (http://www.broadinstitute.org/igv/).

Coverage of pancreatic cancer genes (Table S4) was assessed using CoDeCZ (Nijman et al., 2014). We compared the normalized coverage of each exon in each sample to the median normalized coverage in a comparable reference set. A copy number change was considered amplification when the coverage of at least 45% of consequent exons in one gene was at least 3 MADs higher than the median coverage in the reference samples. A copy number change was considered a deletion when the coverage of at least 45% of consequent exons in one gene was at least 3 MADs lower than the median coverage in the reference samples. A copy number change was considered a homozygous deletion when maximally 1 read mapped to a gene in a sample, and at least 300 reads mapped to this gene in at least four other organoid samples.

Sanger Sequencing

DNA was extracted from human organoids using the DNAeasy blood and tissue kit (QIAGEN) and quantified with the Nanodrop Spectrophotometer (Thermo). Primers for amplification and sequencing of exon 1 and 2 of the KRAS gene were designed using the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), based on the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) reference sequence files (Gene ID: 3845). PCR conditions were as follows: 94°C for 2 min; 3 cycles of 94°C for 15 s, 64°C for 30 s, 70°C for 30 s; 3 cycles of 94°C for 15 s, 61°C for 30 s, 72°C for 30 s; 3 cycles of 94°C for 15 s, 58°C for 30 s, 72°C for 30 s; and 35 cycles of 94°C for 15 s, 57°C for 30 s, and 72°C for 30 s, followed by 72°C for 5 min and 4°C thereafter. PCR products were purified using QIAquick PCR purification kit and sequenced by capillary electrophoresis using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA) on 3730 DNA Analyzer, ABI capillary electrophoresis system (Applied Biosystems). Sequence traces were analyzed using the Codon Code Aligner (www.codoncode.com/aligner/). To enrich for neoplastic cells primary PDA tissues were subjected to laser-capture microdissection. The laser capture microdissection and laser pressure catapulting (LMPC) technique was performed with the Palm CombiSystem of primary tissues (LMDPT, Carl Zeiss MicroImaging) equipped with an Axiovert 200 M Zeiss inverted microscope (Carl Zeiss AG) and a 3-chip charge-coupled-device (3CCD) color camera HV-D30 (Hitachi Kokusai Electric Inc.). A 40x objective was used to catapult the selected tumor area. The microdissected material was catapulted into AdhesiveCap 200 opaque cap (Carl Zeiss) and the DNA was isolated using QIAamo DNA micro Kit (QIAGEN) according to manufacturer's instructions.

Proteomic Analysis of Murine Organoids

Lysis, Tryptic Digestion, and iTRAQ Labeling

Pelleted cells were lysed with 300 µl of lysis buffer (10 mM HEPES, pH 8.0; 0.5 mM EDTA; 1% NP-40; 0.1% SDS). 5 µl each of protease inhibitor cocktail 1 (Sigma-Aldrich), phosphatase inhibitor cocktail 2 (Sigma-Aldrich), and phosphatase inhibitor cocktail 3 (Sigma-Aldrich) were added to each sample. The lysate was sonicated for 30 s and aspirated through a 25 gauge needle five times. The lysate was then centrifuged at 14,000 rpm for 10 min at 4°C to pellet any insoluble material. The supernatant was then removed and placed into a fresh 1.5 ml centrifuge tube. A BCA assay (Pierce) was performed on 8 µl of lysate to determine protein concentration. 80 µg of protein from each sample was then brought to 40 µl final volume with 100 mM triethylammonium bicarbonate buffer (TEAB), pH 7.8. *Tris*(2-carboxyethyl) phosphine (TCEP) was added to a final concentration of 5 mM, and the samples were heated to 55°C for 20 min and allowed to cool to room temperature. Methyl methanethiosulfonate (MMTS) was added to a final concentration of 10 mM, and the samples were incubated at room temperature for a further 20 min to complete blocking of free sulfhydryl groups. Alkylated proteins were then precipitated by the addition of 4x sample volume of methanol, followed by 2x sample volume of chloroform and 3x volume of water. The samples were then incubated at −20°C for 2 hr and centrifuged at 14,000 for 10 min at 4°C. The upper liquid layer was removed and discarded without disturbing the pellet. Methanol was added to 3x the original sample volume, the sample vortexed and then centrifuged at 14,000 rpms for 10 min at 4°C. The entire supernatant was removed and discarded without disturbing the pellet, and the pellet was allowed to dry. The precipitated proteins were reconstituted with 50 µl of 100 mM TEAB, and 1% Protease max surfactant added to a final concentration of 0.1%. The pellet was then sonicated until completely dissolved. 2 µg of sequencing grade trypsin (Promega) was then added to the samples and they were digested overnight at 37°C and dried *in vacuo*. Peptides were reconstituted in 50 µl of 0.5 M TEAB/70% isopropanol and labeled with 8-plex iTRAQ reagent for 2 hr at room temperature essentially according to Ross et al. (2004). Labeled samples were then acidified to pH 4 using formic acid, combined and concentrated *in vacuo* until ~10 µl remained.

Two-Dimensional Fractionation

Peptides were fractionated using a high-low pH reverse phase separation strategy adapted from Gilar et al. (2005). For the first (high pH) dimension, peptides were fractionated on a 10 cm x 1.0 mm column packed with Gemini 3u C18 resin (Phenomenex, Ventura, CA) at a flow rate of 100 µl/min. Mobile phase A consisted of 20 mM ammonium formate pH 10 and mobile phase B consisted of 90% acetonitrile/20 mM ammonium formate pH 10. 100 µg of total peptide was reconstituted with 50 µl of mobile phase A and the entire sample injected onto the column. Peptides were separated using a 35 min linear gradient from 5% B to 70% B and then increasing mobile phase to 95% B for 10 min. Fractions were collected every minute for 80 min and were then combined into 22 fractions using the concatenation strategy described by Wang et al. (2011). An estimated 1 µg of peptide from each of the 22 fractions was then separately injected into the mass spectrometer using capillary reverse phase LC at low pH, described below.

Capillary LC Mass Spectrometry

An Orbitrap Velos Pro mass spectrometer (Thermo Scientific), equipped with a nano-ion spray source was coupled to an EASY-nLC system (Thermo Scientific). The nano-flow LC system was configured with a 180-µm id fused silica capillary trap column containing 3 cm of Aqua 5-µm C18 material (Phenomenex), and a self-pack PicoFrit 100-µm analytical column with an 8-µm emitter (New Objective, Woburn, MA) packed to 15cm with Aqua 3-µm C18 material (Phenomenex). Mobile phase A consisted of 2% acetonitrile/0.1% formic acid and mobile phase B consisted of 90% acetonitrile/ 0.1% formic Acid. 3 µl of each sample dissolved in mobile phase A, was injected through the autosampler onto the trap column. Peptides were then separated using the following linear gradient steps at a flow rate of 400 nl/min: 5% B for 1 min, 5% B to 35% B over 70 min, 35% B to 75% B over 15 min, held at 75% B for 8 min, 75% B to 8% B over 1 min and the final 5 min held at 8% B.

Eluted peptides were directly electrosprayed into the Orbitrap Velos Pro mass spectrometer with the application of a distal 2.3 kV spray voltage and a capillary temperature of 275°C. Each full-scan mass spectrum (Res = 60,000; 380-1700 *m/z*) was followed by MS/MS spectra for the top 12 masses.

High-energy collisional dissociation (HCD) was used with the normalized collision energy set to 35 for fragmentation, the isolation width set to 1.2 and activation time of 0.1. A duration of 70 s was set for the dynamic exclusion with an exclusion list size of 500, repeat count of 1 and exclusion mass width of 10 ppm. We used monoisotopic precursor selection for charge states 2+ and greater, and all data were acquired in profile mode.

Database Searching

Peaklist files were generated by Mascot Distiller (Matrix Science). Protein identification and quantification was carried using Mascot 2.4 (Perkins et al., 1999) against the Uniprot Human sequence database (88,698 sequences; 35,138,129 residues). Methylthiolation of cysteine and N-terminal and lysine iTRAQ modifications were set as fixed modifications, methionine oxidation and deamidation (NQ) as variable. Trypsin was used as cleavage enzyme with one missed cleavage allowed. Mass tolerance was set at 30 ppm for intact peptide mass and 0.3 Da for fragment ions. Search results were rescored to give a final 1% FDR using a randomized version of the same Uniprot Human database. Protein-level iTRAQ ratios were calculated as intensity weighted, using only peptides with expectation values < 0.05. Global ratio normalization (summed) was applied across all iTRAQ channels. Protein enrichment was then calculating by dividing sample protein ratios by the corresponding control sample channel.

Proteomics Data Analysis and Merging

For differential comparison of proteomics data, the limma package was used (Smyth, 2004). An adjusted *p* value < 0.05, determined by limma, was used to identify differentially expressed protein isoforms.

The proteomics data were collected across 4 8-plex iTRAQ experiments as follows:

Exp8a	**Exp8a**	**Exp8a**	**Exp8a**	**Exp8a**	**Exp8a**	**Exp8a**	**Exp8a**
113	114	115	116	117	118	119	121
N4	N8	P2	P3	P4	T4	T5	T7
Exp8b	**Exp8b**	**Exp8b**	**Exp8b**	**Exp8b**	**Exp8b**	**Exp8b**	**Exp8b**
113	114	115	116	117	118	119	121
N4	N8	P5	P6	P7	T8	T9	T10
Exp8c	**Exp8c**	**Exp8c**	**Exp8c**	**Exp8c**	**Exp8c**	**Exp8c**	**Exp8c**
113	114	115	116	117	118	119	121
N5	N11	P5	P6	P7	T8	T9	T10
Exp8d	**Exp8d**	**Exp8d**	**Exp8d**	**Exp8d**	**Exp8d**	**Exp8d**	**Exp8d**
113	114	115	116	117	118	119	121
N12	N13	P4	P7	T7	T8	M1	M2

To merge separate iTRAQ experiments, we considered only proteins identified and quantified in *all* experiments a-d. Although shared channels were treated identically in sample handling (including labeling and total amount of protein loaded) the absolute detection levels of protein and peptides often shows variability between separate MS runs, requiring normalization of signal intensity.

Merging across experiments was accomplished by systematically holding two or three samples constant between experiments a – d ("shared channels"). Only proteins identified and quantified in all experiments were merged. Although shared channels were identically treated, the same protein will not in general have the same quantitative iTRAQ value in replicate experiments. Differences in measured quantitative values are due to 1) different sampling (i.e., differences in the peptides observed between experiments, this a function of data dependent acquisition); and 2) different time points of sampling (i.e., the point of chromatographic elution of the peptide under observation, affecting the amount of sample and possible contaminants observed by the detector). These differences will be unique to each observation and protein.

Non-negative multiplicative normalization coefficients *a*, *b*, *c* and *d* were calculated for each

individual protein from the shared channels of all experiments (vectors A, B, C and D, each consisting of solely the channels shared with other experiments A_1, A_2, A_3...) such that 1) the sum of squared error after normalization between replicate channels is minimized; and 2) the total reporter ion counts for all shared experiments before and after normalization is held constant. These coefficients are then applied to all channels within the experiment. That is,

$$\min_{a,b,c,d} \sum \left((a\mathbf{A} - b\mathbf{B})^2 + (a\mathbf{A} - c\mathbf{C})^2 + (a\mathbf{A} - d\mathbf{D})^2 + \cdots + (c\mathbf{C} - d\mathbf{D})^2 \right)$$

$$\text{where}$$

$$\sum (\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D}) = \sum (a\mathbf{A} + b\mathbf{B} + c\mathbf{C} + d\mathbf{D})$$

As a quantitative check, for the 6051 proteins identified and quantified, the average CV between all replicate normalized channels was 5.61%, within the range of error of quantitative mass spectrometric techniques and within the linear range of detection for an Orbitrap-class machine. Most coefficients (> 80%) were within two-fold of each other, indicating that observation was, for mass spectrometric measurement, relatively reproducible.

DATA ACCESS

All RNA-seq data are available at Gene Expression Omnibus (GEO) under accession number GSE63348. The proteomic raw data are available at PeptideAtlas under accession number PASS00625. The targeted DNA-sequencing data are available at EMBL European Nucleotide Archive under the accession number ERP006373

CONFLICTS OF INTEREST

Dr. Ralph Hruban receives royalty payments from Myriad Genetics for the PalB2 inventions.
Dr. Hans Clevers and Meritxell Huch have patents pending and granted on the organoid technology.

REFERENCES

Abbruzzese, J.L., and Hess, K.R. (2014). New option for the initial management of metastatic pancreatic cancer? J Clin Oncol *32*, 2405-2407.

Agbunag, C., and Bar-Sagi, D. (2004). Oncogenic K-ras drives cell cycle progression and phenotypic conversion of primary pancreatic duct epithelial cells. Cancer research *64*, 5659-5663.

Aguirre, A.J., Bardeesy, N., Sinha, M., Lopez, L., Tuveson, D.A., Horner, J., Redston, M.S., and DePinho, R.A. (2003). Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. Genes Dev *17*, 3112-3126.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Bardeesy, N., Aguirre, A.J., Chu, G.C., Cheng, K.H., Lopez, L.V., Hezel, A.F., Feng, B., Brennan, C., Weissleder, R., Mahmood, U., *et al.* (2006). Both p16(Ink4a) and the p19(Arf)-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. Proc Natl Acad Sci U S A *103*, 5947-5952.

Barker, N., Huch, M., Kujala, P., van de Wetering, M., Snippert, H.J., van Es, J.H., Sato, T., Stange, D.E., Begthel, H., van den Born, M., *et al.* (2010). Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro. Cell Stem Cell *6*, 25-36.

Beatty, G.L., Chiorean, E.G., Fishman, M.P., Saboury, B., Teitelbaum, U.R., Sun, W., Huhn, R.D., Song, W., Li, D., Sharp, L.L., *et al.* (2011). CD40 agonists alter tumor stroma and show efficacy against pancreatic carcinoma in mice and humans. Science *331*, 1612-1616.

Campbell, P.J., Yachida, S., Mudie, L.J., Stephens, P.J., Pleasance, E.D., Stebbings, L.A., Morsberger, L.A., Latimer, C., McLaren, S., Lin, M.L., *et al.* (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature *467*, 1109-1113.

Cheung, H.W., Cowley, G.S., Weir, B.A., Boehm, J.S., Rusin, S., Scott, J.A., East, A., Ali, L.D., Lizotte, P.H., Wong, T.C., *et al.* (2011). Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc Natl Acad Sci U S A *108*, 12372-12377.

Cleveland, M.H., Sawyer, J.M., Afelik, S., Jensen, J., and Leach, S.D. (2012). Exocrine ontogenies: On the development of pancreatic acinar, ductal and centroacinar cells. Seminars in Cell & Developmental Biology *23*, 711-719.

De La, O.J., Emerson, L.L., Goodman, J.L., Froebe, S.C., Illum, B.E., Curtis, A.B., and Murtaugh, L.C. (2008). Notch and Kras reprogram pancreatic acinar cells to ductal intraepithelial neoplasia. Proc Natl Acad Sci U S A *105*, 18907-18912.

DeNicola, G.M., Karreth, F.A., Humpton, T.J., Gopinathan, A., Wei, C., Frese, K., Mangal, D., Yu, K.H., Yeo, C.J., Calhoun, E.S., *et al.* (2011). Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. Nature *475*, 106-109.

Dumartin, L., Whiteman, H.J., Weeks, M.E., Hariharan, D., Dmitrovic, B., Iacobuzio-Donahue, C.A., Brentnall, T.A., Bronner, M.P., Feakins, R.M., Timms, J.F., *et al.* (2011). AGR2 is a novel surface antigen that promotes the dissemination of pancreatic cancer cells through regulation of cathepsins B and D. Cancer research *71*, 7091-7102.

Erkan, M., Reiser-Erkan, C., Michalski, C.W., Deucker, S., Sauliunaite, D., Streit, S., Esposito, I., Friess, H., and Kleeff, J. (2009). Cancer-stellate cell interactions perpetuate the hypoxia-fibrosis cycle in pancreatic ductal adenocarcinoma. Neoplasia *11*, 497-508.

Frese, K.K., Neesse, A., Cook, N., Bapiro, T.E., Lolkema, M.P., Jodrell, D.I., and Tuveson, D.A. (2012). nab-Paclitaxel potentiates gemcitabine activity by reducing cytidine deaminase levels in a mouse model of pancreatic cancer. Cancer Discov *2*, 260-269.

Gao, D., Vela, I., Sboner, A., Iaquinta, P.J., Karthaus, W.R., Gopalan, A., Dowling, C., Wanjala, J.N., Undvall, E.A., Arora, V.K., *et al.* (2014). Organoid cultures derived from patients with advanced prostate cancer. Cell *159*, 176-187.

Genome of the Netherlands, C. (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nature genetics *46*, 818-825.

Gidekel Friedlander, S.Y., Chu, G.C., Snyder, E.L., Girnius, N., Dibelius, G., Crowley, D., Vasile, E., DePinho, R.A., and Jacks, T. (2009). Context-dependent transformation of adult pancreatic cells by oncogenic K-Ras. Cancer Cell *16*, 379-389.

Gilar, M., Olivova, P., Daly, A.E., and Gebler, J.C. (2005). Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. Journal of separation science *28*, 1694-1703.

Gould, V.E., Martinez, N., Orucevic, A., Schneider, J., and Alonso, A. (2000). A novel, nuclear pore-associated, widely distributed molecule overexpressed in oncogenesis and development. Am J Pathol *157*, 1605-1613.

Guerra, C., Mijimolle, N., Dhawahir, A., Dubus, P., Barradas, M., Serrano, M., Campuzano, V., and Barbacid, M. (2003). Tumor induction by an endogenous K-ras oncogene is highly dependent on cellular context. Cancer Cell *4*, 111-120.

Habbe, N., Shi, G., Meguid, R.A., Fendrich, V., Esni, F., Chen, H., Feldmann, G., Stoffers, D.A., Konieczny, S.F., Leach, S.D., *et al.* (2008). Spontaneous induction of murine pancreatic intraepithelial neoplasia (mPanIN) by acinar cell targeting of oncogenic Kras in adult mice. Proc Natl Acad Sci U S A *105*, 18913-18918.

Harakalova, M., Mokry, M., Hrdlickova, B., Renkens, I., Duran, K., van Roekel, H., Lansu, N., van Roosmalen, M., de Bruijn, E., Nijman, I.J., *et al.* (2011). Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. Nat Protoc *6*, 1870-1886.

Hingorani, S.R., Petricoin, E.F., Maitra, A., Rajapakse, V., King, C., Jacobetz, M.A., Ross, S., Conrads, T.P., Veenstra, T.D., Hitt, B.A., *et al.* (2003). Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. Cancer Cell *4*, 437-450.

Hingorani, S.R., Wang, L., Multani, A.S., Combs, C., Deramaudt, T.B., Hruban, R.H., Rustgi, A.K., Chang, S., and Tuveson, D.A. (2005). Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer Cell *7*, 469-483.

Huch, M., Bonfanti, P., Boj, S.F., Sato, T., Loomans, C.J., van de Wetering, M., Sojoodi, M., Li, V.S., Schuijers, J., Gracanin, A., *et al.* (2013a). Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. Embo J, 2708-2721.

Huch, M., Dorrell, C., Boj, S.F., van Es, J.H., Li, V.S., van de Wetering, M., Sato, T., Hamer, K., Sasaki, N., Finegold, M.J., *et al.* (2013b). In vitro expansion of

single Lgr5+ liver stem cells induced by Wnt-driven regeneration. Nature *494*, 247-250.

Jacobetz, M.A., Chan, D.S., Neesse, A., Bapiro, T.E., Cook, N., Frese, K.K., Feig, C., Nakagawa, T., Caldwell, M.E., Zecchini, H.I., *et al.* (2012). Hyaluronan impairs vascular function and drug delivery in a mouse model of pancreatic cancer. Gut, 112-120.

Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., *et al.* (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science *321*, 1801-1806.

Karthaus, W.R., Iaquinta, P.J., Drost, J., Gracanin, A., van Boxtel, R., Wongvipat, J., Dowling, C.M., Gao, D., Begthel, H., Sachs, N., *et al.* (2014). Identification of multipotent luminal progenitor cells in human prostate organoid cultures. Cell *159*, 163-175.

Kim, J., Hoffman, John P., Alpaugh, R.K., Rhim, Andrew D., Reichert, M., Stanger, Ben Z., Furth, Emma E., Sepulveda, Antonia R., Yuan, C.-X., Won, K.-J., *et al.* (2013). An iPSC Line from Human Pancreatic Ductal Adenocarcinoma Undergoes Early to Invasive Stages of Pancreatic Cancer Progression. Cell Rep *3*, 2088-2099.

Kim, M.P., Evans, D.B., Wang, H., Abbruzzese, J.L., Fleming, J.B., and Gallick, G.E. (2009). Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice. Nat Protoc *4*, 1670-1680.

Kohler, A., and Hurt, E. (2010). Gene regulation by nucleoporins and links to cancer. Mol Cell *38*, 6-15.

Koo, B.K., Stange, D.E., Sato, T., Karthaus, W., Farin, H.F., Huch, M., van Es, J.H., and Clevers, H. (2012). Controlled gene expression in primary Lgr5 organoid cultures. Nature methods *9*, 81-83.

Koong, A.C., Mehta, V.K., Le, Q.T., Fisher, G.A., Terris, D.J., Brown, J.M., Bastidas, A.J., and Vierra, M. (2000). Pancreatic tumors show high levels of hypoxia. Int J Radiat Oncol Biol Phys *48*, 919-922.

Kopp, J.L., von Figura, G., Mayes, E., Liu, F.F., Dubois, C.L., Morris, J.P.t., Pan, F.C., Akiyama, H., Wright, C.V., Jensen, K., *et al.* (2012). Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. Cancer Cell *22*, 737-750.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Lee, J., Sugiyama, T., Liu, Y., Wang, J., Gu, X., Lei, J., Markmann, J.F., Miyazaki, S., Miyazaki, J., Szot, G.L., *et al.* (2013). Expansion and conversion of human pancreatic ductal cells into insulin-secreting endocrine cells. eLife *2*, e00940.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics *12*, 323.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, X., Nadauld, L., Ootani, A., Corney, D.C., Pai, R.K., Gevaert, O., Cantrell, M.A., Rack, P.G., Neal, J.T., Chan, C.W., *et al.* (2014). Oncogenic transformation of diverse gastrointestinal tissues in primary organoid culture. Nat Med *20*, 769-777.

Makovitzky, J. (1986). The distribution and localization of the monoclonal antibody-defined antigen 19-9 (CA19-9) in chronic pancreatitis and pancreatic carcinoma. An immunohistochemical study. Virchows Archiv B, Cell pathology including molecular pathology *51*, 535-544.

Means, A.L., Meszoely, I.M., Suzuki, K., Miyamoto, Y., Rustgi, A.K., Coffey, R.J., Wright, C.V.E., Stoffers, D.A., and Leach, S.D. (2005). Pancreatic epithelial plasticity mediated by acinar cell transdifferentiation and generation of nestin-positive intermediates. Development *132*, 3767-3776.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., *et al.* (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics *34*, 267-273.

Morris, J.P.t., Cano, D.A., Sekine, S., Wang, S.C., and Hebrok, M. (2010). Beta-catenin blocks Kras-dependent reprogramming of acini into pancreatic cancer precursor lesions in mice. J Clin Invest *120*, 508-520.

Neesse, A., Frese, K.K., Chan, D.S., Bapiro, T.E.,

Howat, W.J., Richards, F.M., Ellenrieder, V., Jodrell, D.I., and Tuveson, D.A. (2014). SPARC independent drug delivery and antitumour effects of nab-paclitaxel in genetically engineered mice. Gut *63*, 974-983.

Nijman, I.J., van Montfrans, J.M., Hoogstraat, M., Boes, M.L., van de Corput, L., Renner, E.D., van Zon, P., van Lieshout, S., Elferink, M.G., van der Burg, M., *et al.* (2014). Targeted next-generation sequencing: a novel diagnostic tool for primary immunodeficiencies. The Journal of allergy and clinical immunology *133*, 529-534.

Olive, K.P., Jacobetz, M.A., Davidson, C.J., Gopinathan, A., McIntyre, D., Honess, D., Madhu, B., Goldgraben, M.A., Caldwell, M.E., Allard, D., *et al.* (2009). Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. Science *324*, 1457-1461.

Olive, K.P., Tuveson, D.A., Ruhe, Z.C., Yin, B., Willis, N.A., Bronson, R.T., Crowley, D., and Jacks, T. (2004). Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. Cell *119*, 847-860.

Perez-Mancera, P.A., Guerra, C., Barbacid, M., and Tuveson, D.A. (2012). What We Have Learned About Pancreatic Cancer from Mouse Models. Gastroenterology, 1079-1092.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551-3567.

Prasad, N.B., Biankin, A.V., Fukushima, N., Maitra, A., Dhara, S., Elkahloun, A.G., Hruban, R.H., Goggins, M., and Leach, S.D. (2005). Gene expression profiles in pancreatic intraepithelial neoplasia reflect the effects of Hedgehog signaling on pancreatic ductal epithelial cells. Cancer research *65*, 1619-1626.

Premsrirut, P.K., Dow, L.E., Kim, S.Y., Camiolo, M., Malone, C.D., Miething, C., Scuoppo, C., Zuber, J., Dickins, R.A., Kogan, S.C., *et al.* (2011). A rapid and scalable system for studying gene function in mice using conditional RNA interference. Cell *145*, 145-158.

Provenzano, P.P., Cuevas, C., Chang, A.E., Goel, V.K., Von Hoff, D.D., and Hingorani, S.R. (2012). Enzymatic targeting of the stroma ablates physical barriers to treatment of pancreatic ductal adenocarcinoma. Cancer Cell *21*, 418-429.

Pylayeva-Gupta, Y., Lee, K.E., Hajdu, C.H., Miller, G., and Bar-Sagi, D. (2012). Oncogenic Kras-induced GM-CSF production promotes the development of pancreatic neoplasia. Cancer Cell *21*, 836-847.

Rahib, L., Smith, B.D., Aizenberg, R., Rosenzweig, A.B., Fleshman, J.M., and Matrisian, L.M. (2014). Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res *74*, 2913-2921.

Ray, K.C., Bell, K.M., Yan, J., Gu, G., Chung, C.H., Washington, M.K., and Means, A.L. (2011). Epithelial tissues have varying degrees of susceptibility to Kras(G12D)-initiated tumorigenesis in a mouse model. PLoS One *6*, e16786.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., *et al.* (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics *3*, 1154-1169.

Rovira, M., Scott, S.G., Liss, A.S., Jensen, J., Thayer, S.P., and Leach, S.D. (2010). Isolation and characterization of centroacinar/terminal ductal progenitor cells in adult mouse pancreas. Proc Natl Acad Sci U S A *107*, 75-80.

Rubio-Viqueira, B., Jimeno, A., Cusatis, G., Zhang, X., Iacobuzio-Donahue, C., Karikari, C., Shi, C., Danenberg, K., Danenberg, P.V., Kuramochi, H., *et al.* (2006). An in vivo platform for translational drug development in pancreatic cancer. Clinical cancer research : an official journal of the American Association for Cancer Research *12*, 4652-4661.

Ryan, D.P., Hong, T.S., and Bardeesy, N. (2014). Pancreatic adenocarcinoma. N Engl J Med *371*, 1039-1049.

Sato, T., Stange, D.E., Ferrante, M., Vries, R.G., Van Es, J.H., Van den Brink, S., Van Houdt, W.J., Pronk, A., Van Gorp, J., Siersema, P.D., *et al.* (2011). Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology *141*, 1762-1772.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., *et al.* (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262-265.

Sawey, E.T., Johnson, J.A., and Crawford, H.C. (2007). Matrix metalloproteinase 7 controls pancreatic

acinar cell transdifferentiation by activating the Notch signaling pathway. Proc Natl Acad Sci U S A *104*, 19327-19332.

Seaberg, R.M., Smukler, S.R., Kieffer, T.J., Enikolopov, G., Asghar, Z., Wheeler, M.B., Korbutt, G., and van der Kooy, D. (2004). Clonal identification of multipotent precursors from adult mouse pancreas that generate neural and pancreatic lineages. Nat Biotechnol *22*, 1115-1124.

Shain, A.H., Salari, K., Giacomini, C.P., and Pollack, J.R. (2013). Integrative genomic and functional profiling of the pancreatic cancer genome. BMC genomics *14*, 624.

Sharma, S.V., Haber, D.A., and Settleman, J. (2010). Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. Nat Rev Cancer *10*, 241-253.

Siegel, R., Naishadham, D., and Jemal, A. (2013). Cancer statistics, 2013. CA Cancer J Clin *63*, 11-30.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology *3*, Article3.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Vermaat, J.S., Nijman, I.J., Koudijs, M.J., Gerritse, F.L., Scherer, S.J., Mokry, M., Roessingh, W.M., Lansu, N., de Bruijn, E., van Hillegersberg, R., *et al.* (2012). Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. Clin Cancer Res *18*, 688-699.

Villarroel, M.C., Rajeshkumar, N.V., Garrido-Laguna, I., De Jesus-Acosta, A., Jones, S., Maitra, A., Hruban, R.H., Eshleman, J.R., Klein, A., Laheru, D., *et al.* (2011). Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. Mol Cancer Ther *10*, 3-8.

von Figura, G., Fukuda, A., Roy, N., Liku, M.E., Morris Iv, J.P., Kim, G.E., Russ, H.A., Firpo, M.A., Mulvihill, S.J., Dawson, D.W., *et al.* (2014). The chromatin regulator Brg1 suppresses formation of intraductal papillary mucinous neoplasm and pancreatic ductal adenocarcinoma. Nat Cell Biol *16*, 255-267.

Wang, Y., Yang, F., Gritsenko, M.A., Wang, Y., Clauss, T., Liu, T., Shen, Y., Monroe, M.E., Lopez-Ferrer, D., Reno, T., *et al.* (2011). Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics *11*, 2019-2026.

Wiese, S., Reidegeld, K.A., Meyer, H.E., and Warscheid, B. (2007). Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics *7*, 340-350.

Wu, J., Jiao, Y., Dal Molin, M., Maitra, A., de Wilde, R.F., Wood, L.D., Eshleman, J.R., Goggins, M.G., Wolfgang, C.L., Canto, M.I., *et al.* (2011). Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. Proc Natl Acad Sci U S A *108*, 21188-21193.

Ying, H., Kimmelman, A.C., Lyssiotis, C.A., Hua, S., Chu, G.C., Fletcher-Sananikone, E., Locasale, J.W., Son, J., Zhang, H., Coloff, J.L., *et al.* (2012). Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. Cell *149*, 656-670.

Zhao, L.L., Zhang, T., Liu, B.R., Liu, T.F., Tao, N., and Zhuang, L.W. (2014). Construction of pancreatic cancer double-factor regulatory network based on chip data on the transcriptional level. Molecular biology reports *41*, 2875-2883.

SUPPLEMENTAL FIGURES AND TABLES

Supplemental tables S1-S7 are available upon request or through https://www.cell.com/action/showImagesData?pii=S0092-8674%2814%2901592-X.

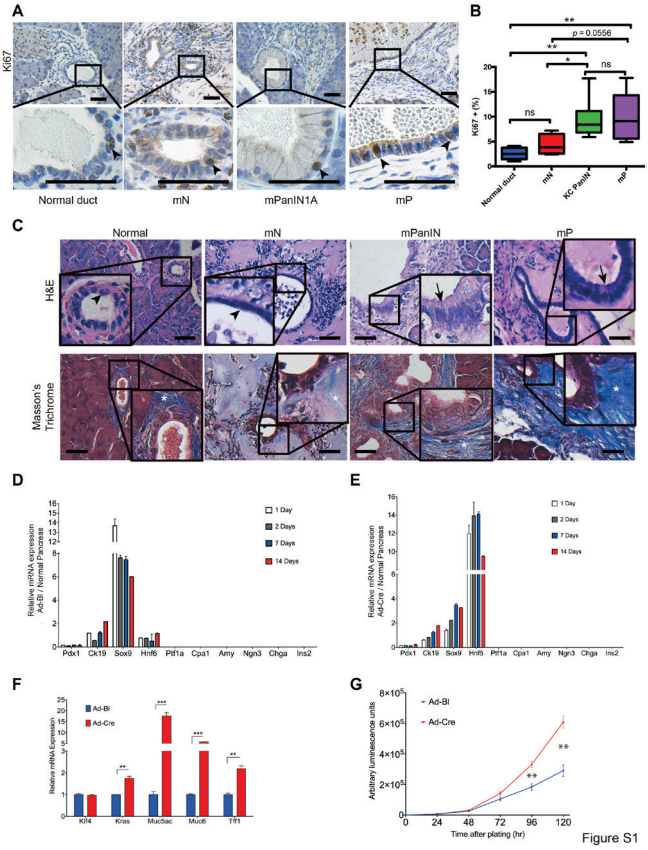


Figure S1

Supplemental figure S1. Cellular and Histological Features of Transplanted mN and mP Organoids, Related to Figure 1. (A) Representative Ki-67 immunohistochemistry (IHC) from pancreatic ductal cells in wild-type mice (n = 5), mN organoids orthotopically transplanted into syngeneic C57Bl/6 pancreata (n = 4), low-grade mPanIN-1a in *Kras*$^{+/LSL\text{-}G12D}$; *Pdx1-cre* (KC) mice (n = 3); and mP organoids orthotopically transplanted into C57Bl/6 pancreata that formed mPanIN-like structures (n = 3). Arrowheads show representative Ki-67 positive cells. Scale bars represent 50 µm. (B) Quantification of Ki67-positive cells from A. Error bars are standard deviations (SDs). *, **, ns: $p < 0.05$, 0.01, or not significant by two-tailed Student's *t* test. (C) Representative pancreatic tissues stained with H&E and Masson's Trichrome from normal pancreata, mN transplants, KC mice and mP transplants (n > 3 mice). Arrowhead shows simple cuboidal epithelium of normal ducts. Arrows show tall columnar cells present in mPanIN and mP transplants. mP and KC pancreata demonstrate a more extensive desmoplastic stroma compared to mN and normal ducts (asterisk). Scale bars represent 50 µm. (D) Expression of pancreatic lineage-specific genes in control-transduced *Kras*$^{+/LSL\text{-}G12D}$ organoids over 2 weeks in culture. (E) Expression of pancreatic lineage-specific genes in Cre-transduced *Kras*$^{+/G12D}$ organoids over 2 weeks in culture. Expression was normalized to normal pancreas. Data represent mean of 3 replicate qRT-PCRs of individual organoid cultures. Error bars are SD. (F) *Kras*^{G12D}-induced expression of genes associated with pre-invasive mPanIN lesions, including Muc5ac, Muc6 and Tff1, quantified by qRT-PCR as presented in Figure 1D. Data represent mean of 3 biological replicates. Error bars show SEM. **, ***: $p < 0.01$, 0.001 by two-tailed Student's *t* test. (G) Proliferation of *Kras*^{G12D}-expressing organoids compared to *Kras*$^{+/LSL\text{-}G12D}$ organoids was measured by cell viability luminescence assays. Error bars indicate standard deviation for 5 replicates. Data are representative of three independently infected organoid cultures. **: $p < 0.01$ by Kolmogorov-Smirnov test.

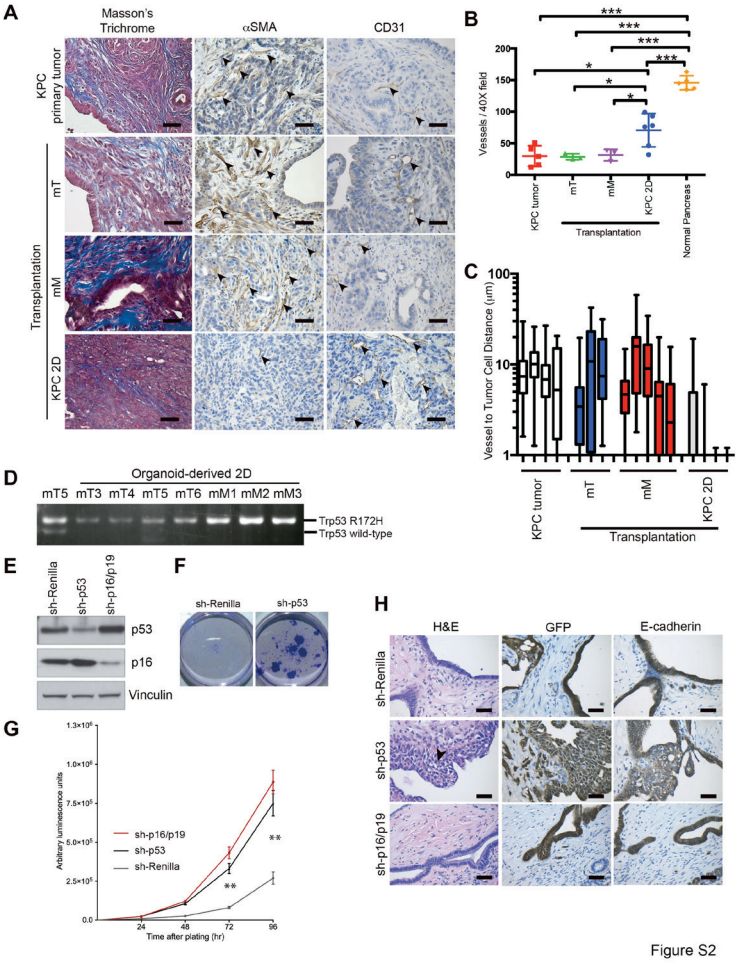
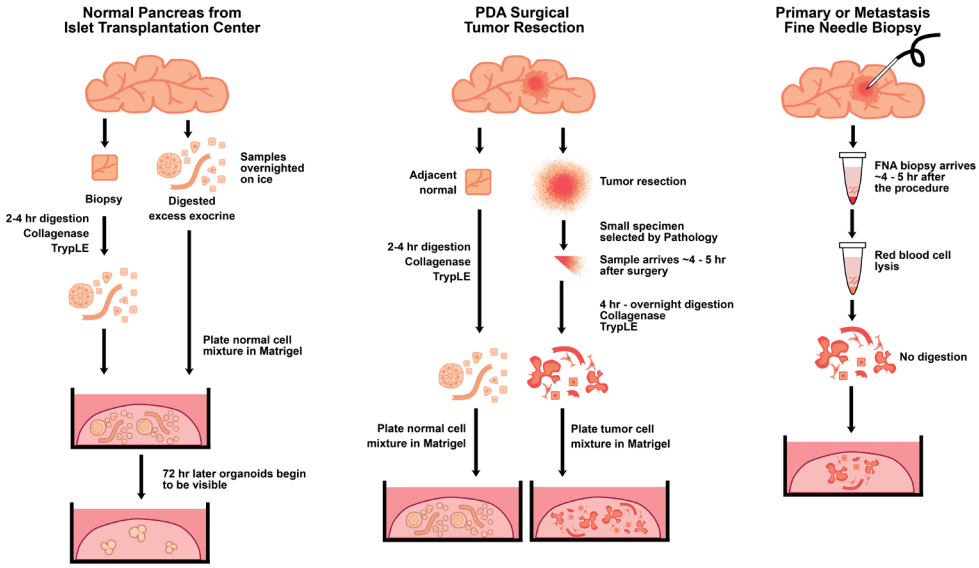


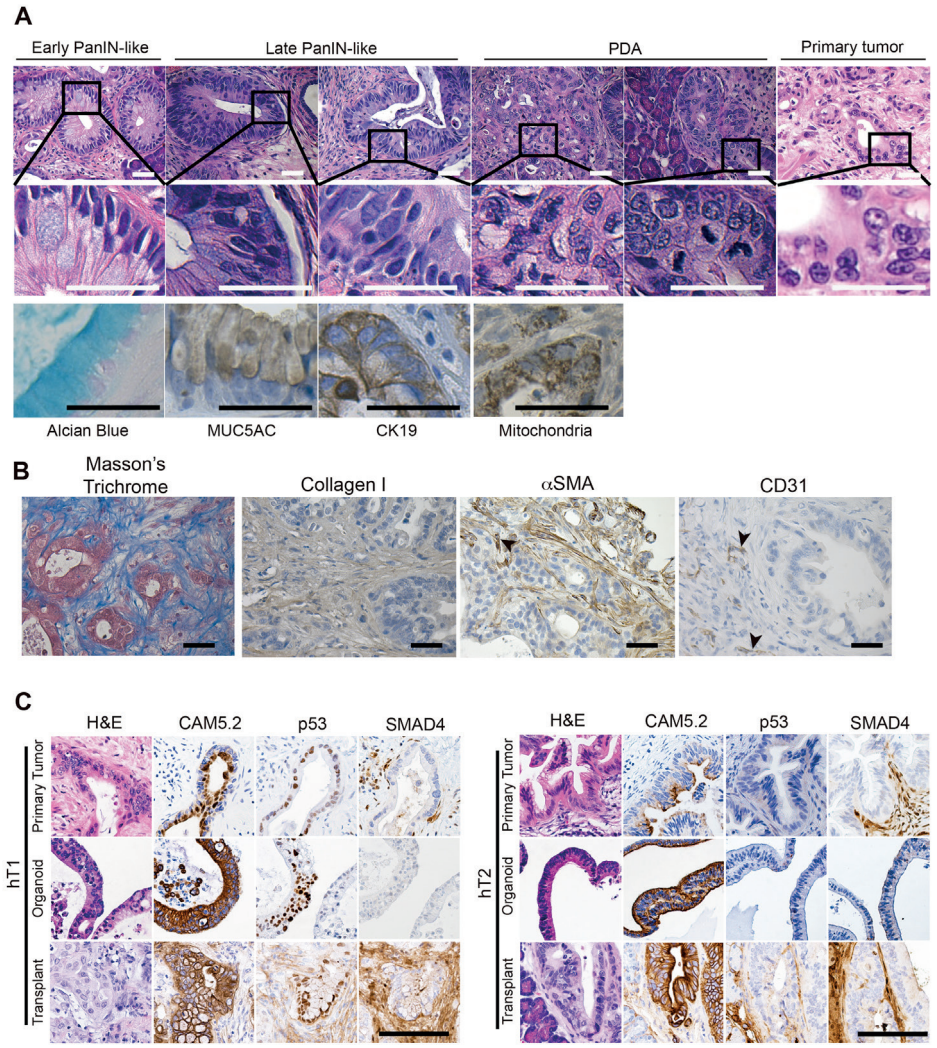
Figure S2

Supplemental figure S2. Histological Features of Transplanted mT and mM Organoids Resemble Autochthonous KPC Tumors, Related to Figure 2. (A) Representative histological analysis of transplanted mT and mM organoids. Transplants were evaluated for stromal reaction by Masson's Trichrome staining and αSMA content (arrowheads). Intratumoral blood vessels were evaluated by CD31 IHC (arrowheads). Scale bars represent 50 μm. (B) Quantification of vascular density in KPC tumors and mT and mM transplants. Data are shown as mean and error bars show SD. *, ***: $p < 0.05$, 0.001 by two-tailed Student's t test. (C) Quantification of vascular-to-neoplastic cell distance in primary KPC tumors, mT, and mM transplanted organoids. Data are shown as box-and-whisker plots, with the edges of the box representing the 25th and 75th percentiles and the whiskers representing the minimum and maximum values. (D) PCR analysis of wild-type and mutant *Trp53* alleles in mT organoid and organoid-derived 2D cultures. (E) Immunoblots of p53 and p16 protein levels, with Vinculin as a loading control. (F) Crystal Violet staining of cells derived from mP organoids transduced with a p53 or Renilla-control shRNA and plated as a monolayer culture. (G) Proliferation of organoids upon depletion of p16/p19 and p53 compared with sh-Renilla control was measured by cell viability luminescence assays. **: $p < 0.01$ by Kolmogorov-Smirnov test. (H) Histologic analyses of shRNA-transduced mP1 organoids orthotopically transplanted into immune-compromised (NSG) mice, including H&E and IHC for GFP and E-cadherin. The transplants were assessed after 1 month. Micro-invasive PDA is denoted by arrowheads. Scale bars represent 50 μm.

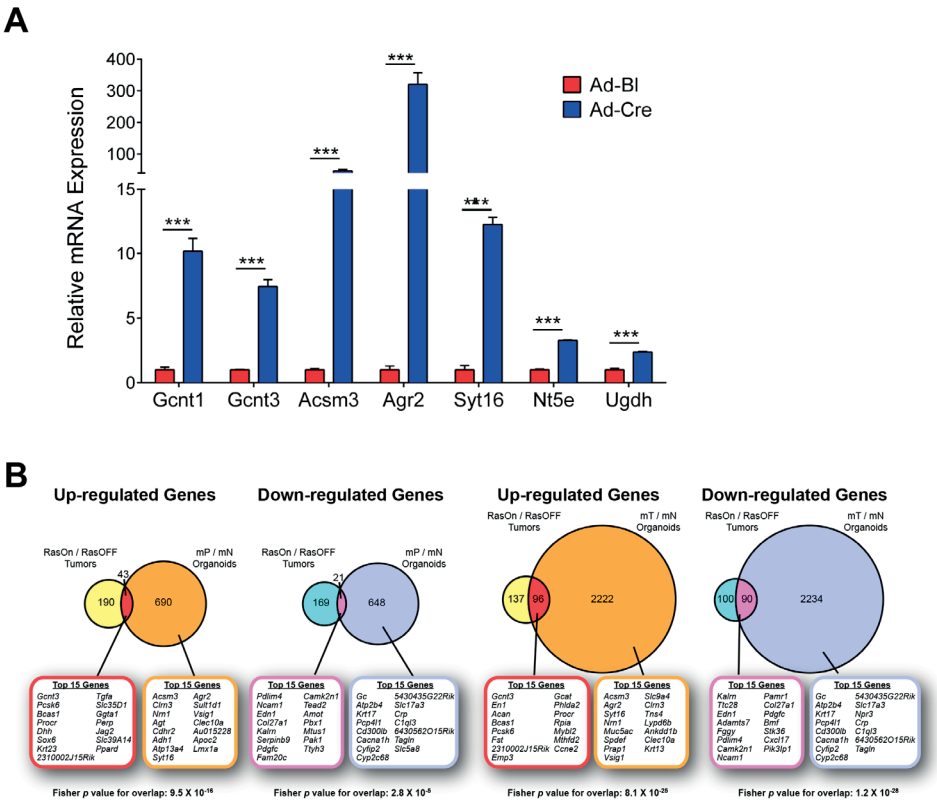
Figure S3



Supplemental figure S3. Schematic Representation of the Generation of Human Organoid Culture, Related to Figure 3. Left panel: schematic representation of human organoid culture from normal pancreas. Middle panel: schematic representation of human tumor organoid culture from resected PDA specimens. Right panel: schematic representation of human tumor organoid culture from fine needle biopsy.



Supplemental figure S4. Histological Characterization of Transplanted hT Organoids, Related to Figure 4. (A) Following orthotopic transplantation into *Nu/Nu* mice, hT4 organoids formed low- and high-grade PanIN-like structures within one month (n = 2/2 mice). PDA was observed at later time points (n = 2/2 mice). Histology of the primary tumor is included (right-most panels). Mucinous metaplasia is highlighted by Alcian Blue staining as well as MUC5AC and CK19 IHC. IHC staining for human mitochondrial protein confirms the human origin of the orthotopic human PanIN-like and PDA cells. Scale bars represent 50 µm. (B) Tumors derived from orthotopically transplanted hT4 organoids were evaluated for stromal reaction by Masson's Trichrome staining, and by IHC for Collagen I content and presence of αSMA-expressing intratumoral fibroblasts (n = 2 tumors for each). The vasculature in transplanted hT4 tumors was evaluated by CD31 IHC. Scale bars represent 50 µm. (C) IHC was performed on hT1 and hT2 organoids and their associated orthotopic transplants for the expression of CAM5.2 (human-specific CK8), TRP53 (p53), and SMAD4. CAM5.2 staining was used to identify cells of human origin in the mouse pancreata. Scale bar represents 50 µm.



Supplemental figure S5. Additional Expression Analysis of Murine Organoids, Related to Figure 5. (A) Activation of *Kras*^{G12D} by adenoviral-Cre infection (AdCre) of *Kras*$^{+/LSL-G12D}$; *R26*$^{LSL-YFP}$ derived organoids induced expression of genes that were identified as upregulated in mP or mT organoids by RNA-seq analysis. Gene levels were quantified by qRT-PCR and normalized to the expression levels of *Kras*$^{+/LSL-G12D}$; *R26*$^{LSL-YFP}$-derived organoids infected with Adeno-blank (AdBl). Mean of 3 biological replicates. Error bars show SEMs. ***: $p < 0.001$ by two-tailed Student's t test. (B) Genes previously reported (Ying et al., 2012) as significantly up- or downregulated in murine pancreatic tumors following silencing of oncogenic *Kras*^{G12D} were compared to genes significantly up- or downregulated in mP and mT organoids. Venn diagrams show overlap between the two datasets. p values for the overlaps, determined by two-tailed Fisher's Exact test, are shown.

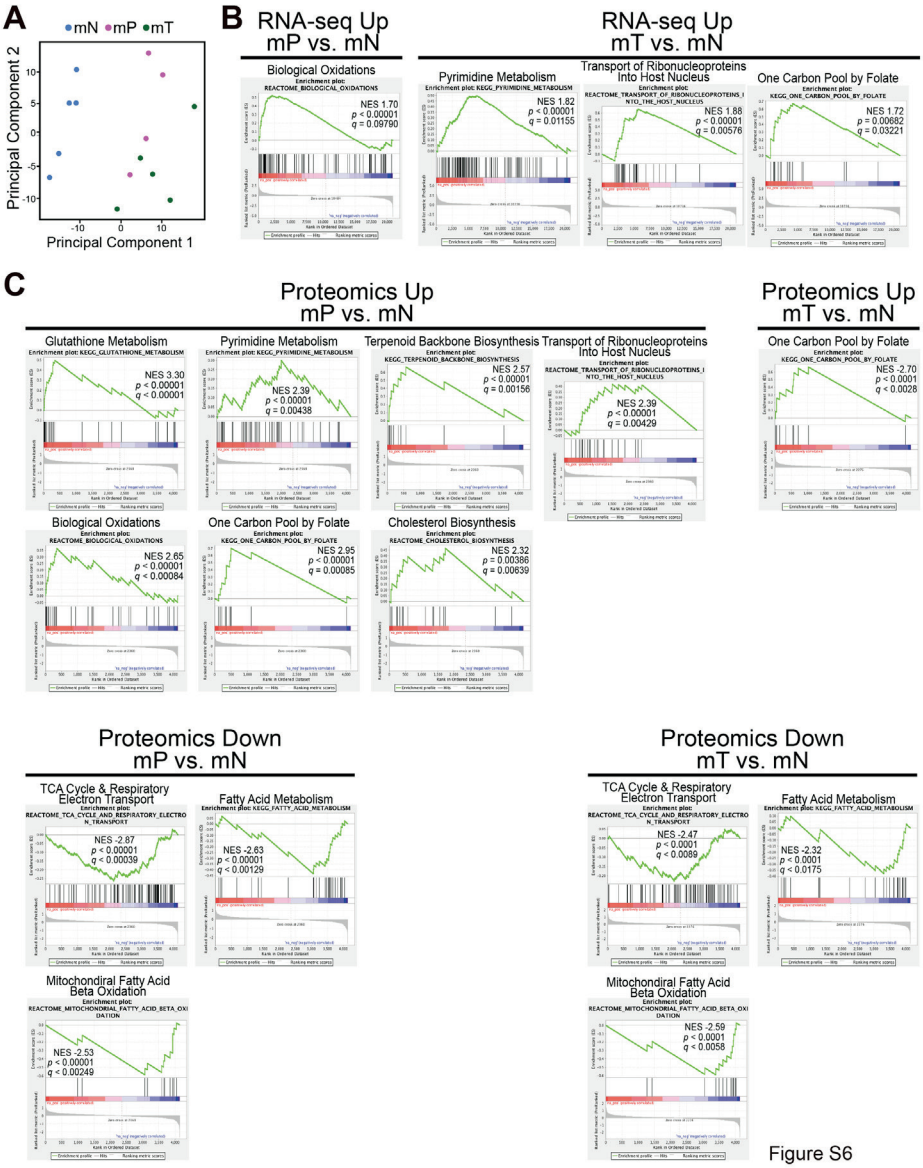
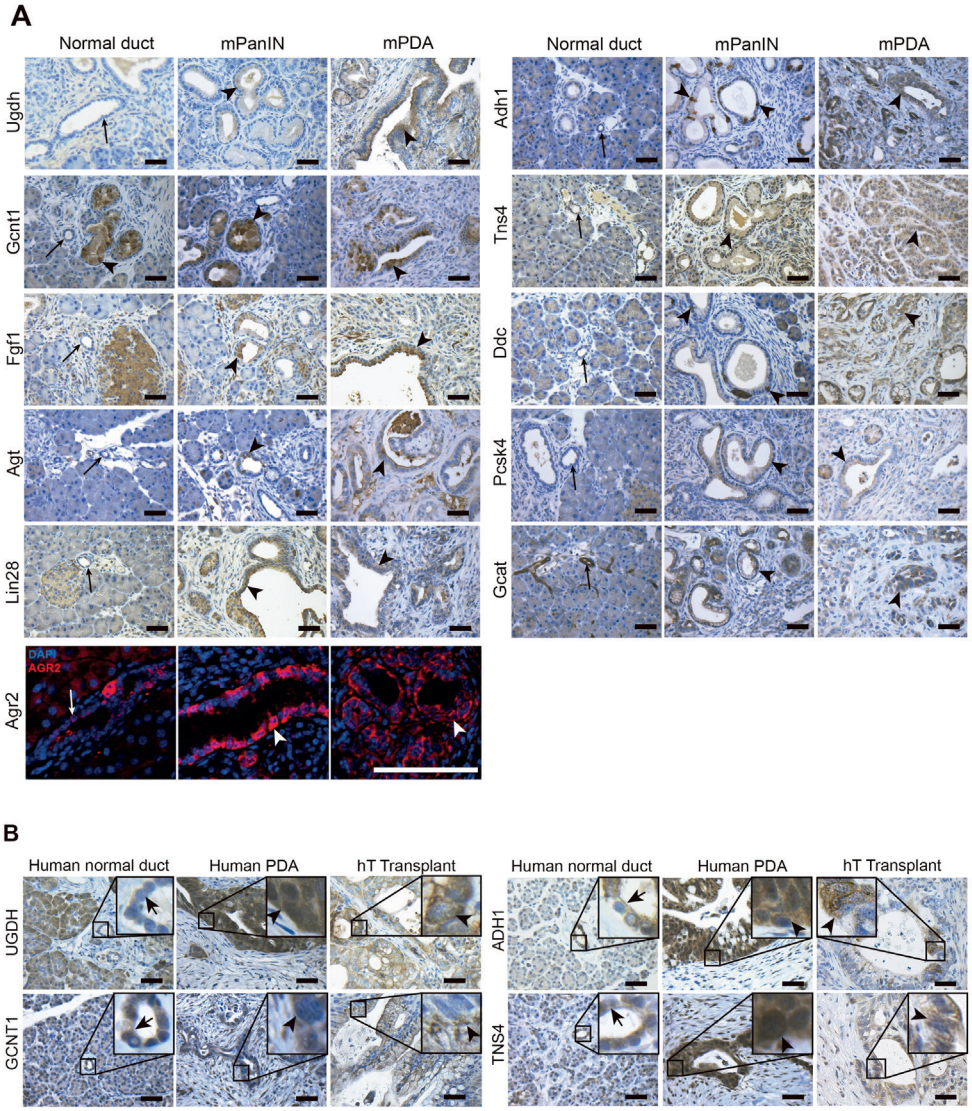


Figure S6

Supplemental figure S6. Additional Characterization of Proteomic Data and Pathway Analysis, Related to Figure 6. (A) Principal component analysis of protein expression data for mN, mP, and mT organoids. (B) Examples of molecular pathways found enriched by GSEA analysis of RNA-seq data. Normalized enrichment scores (NESs), *p* and *q* values are shown. (C) Examples of molecular pathways found enriched by GSEA analysis of proteomic data. NESs, *p* and *q* values are shown.



Supplemental figure S7. Immunohistochemical Analysis of Genes Found to be Differentially Expressed by Either RNA-Seq or Proteomic Analyses, Related to Figure 7. (A) IHC analysis of Ugdh, Gcnt1, Fgf1, Agt, Lin28, Adh1, Tns4, Ddc, Pcsk4, and Gcat; as well as immunofluorescence for Agr2 in mouse PanIN and PDA tissues. Arrows indicate adjacent normal ducts in mPanIN tissues and arrowheads indicate PanIN or PDA. Scale bars represent 50 μm. (B) IHC analysis of UGDH, GCNT1, ADH1 and TNS4 in human normal pancreas and PDA. Arrows indicate normal ducts and arrowheads indicate PDA. Scale bars represent 50 μm.

The road to better understanding continues

Chapter 7

General discussion

Myrthe Jager[1], Ruben van Boxtel[2] and Edwin Cuppen[1]

1 Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht University, Universiteitsweg 100, 3584, CG, Utrecht, The Netherlands
2 Princess Máxima Center for Pediatric Oncology, 3584 CT Utrecht, The Netherlands

Introduction

Cancer is caused by the sequential accumulation of driver mutations in the genome of a single cell, allowing the clonal expansion of this cell (*1*, *2*). Several factors can increase the risk of developing cancer, such as old age, genetic predisposition, exposure to sunlight, alcohol consumption, tobacco smoking, viral infections, and obesity (*3*–*7*). However, for the majority of the risk factors it remains unclear how they contribute to the development of (specific types of) cancer. Furthermore, humans are more likely to develop cancer in certain tissues (*8*, *9*) and the exposure to risk factors alone cannot explain the variation in cancer incidence across tissues (*10*). It has been proposed that the number of stem cell divisions correlates with tissue-specific cancer incidence, suggesting that the majority of cancer can be explained by "bad luck" (*10*, *11*). Highly proliferative cells might incorporate more mutations in their genomes during life, thereby increasing the risk of accumulating specific driver mutations and developing cancer. However, this model has received much criticism in the field, mainly due to the debatable methods that were used (*12*, *13*).

To gain insight into the mutational processes that contribute to the accumulation of driver events, and ultimately to cancer development, genome-wide patterns of the accumulated somatic point mutations can be characterized (*14*). In this thesis, I have investigated mutational processes in tissue-specific adult stem cells (ASCs), which are believed to be the cells-of-origin of many cancer types (*15*–*17*). Using a new tool, described in **Chapter 2** of this thesis (*18*), we characterized the mutational profiles of healthy and precancerous ASCs prior to tumor initiation. In this chapter I will put the findings of this thesis into broader context, and provide potential angles for future research.

Genomes accumulate tumor driver mutations with age

Age is the most important risk factor for cancer and the majority of cancer patients is at least 45 years old (*2*, *9*, *19*, *20*). Sequencing of tumor genomes showed that the number of somatic mutations in a tumor increases with age (*21*), and roughly half of these mutations was predicted to occur prior to tumor initiation (*22*). Consistently, we show that the number of somatic point mutations increases linearly with age in the genomes of physiologically normal ASCs of the small intestine, colon, and liver in **Chapter 3** (*23*). An increased mutational load with age is associated with a higher number of tumor driver mutations by mere chance, which can ultimately be sufficient to drive tumor initiation (*17*).

For a long time, it was unclear which mutational processes cause this increase in the mutational load with age. However, recently 30 signatures of base substitutions and 6 rearrangement signatures have been identified, providing a

useful framework for determining the mutational processes that underlie mutation accumulation (*24*, *25*). We found that the age-dependent accumulation of mutations is caused by a tissue-specific, yet age-independent, combination of two clock-like mutational processes in multiple tissues: Signature 1 and Signature 5 (*23*, *26*) (**Chapter 3**). Signature 1 mutations are believed to be a result of spontaneous deamination of methylated cytosines at CpG dinucleotides (*25*). The etiology of Signature 5 is currently unknown, although the association with a transcriptional strand bias (*25*) suggests that either the mutagenic process or DNA repair pathway is more active during transcription. Since these two signatures are detected in all tumor types (*24*, *25*) (Figure 1), the underlying mutational processes most likely play a general role in the accumulation of mutations during life. Interestingly, the tissue-specific mutation spectra observed in healthy ASCs of colon, small intestine, and liver are very similar to the mutation spectra of driver mutations in tumor suppressor genes of the corresponding tumor type (*23*) (**Chapter 3**). This suggests that the "unavoidable" activity of the mutational processes that drive Signature 1 and Signature 5 mutations might, indeed, stochastically introduce tumor driver mutations during life, thereby contributing to an increased cancer risk with age.

Another important risk factor for the development of various types of cancer is the repeated exposure to extrinsic risk factors, such as UV-light and tobacco smoking. These extrinsic risk factors (or at least some of these extrinsic risk factors) may primarily exert their tumorigenic effect through induction of somatic mutations in the genome as well (*27*). To date, 6 out of 30 COSMIC mutational signatures have been linked to an exposure to exogenous mutagens (*28*). Tumor types in which these mutational signatures are detected almost exclusively have a high median mutational load (Figure 1) (*25*) and the mutation spectra of tumor driver genes in these tumors are similar to the mutational profiles linked to exogenous mutagen exposure (*29*, *30*). This suggests that, similar to endogenous mutagens, (at least some) exogenous mutagens can contribute to an increased cancer risk by increasing the point mutation load in ASCs, thereby generating driver mutations.

The increase in mutational load with age due to mutational processes also explains why patients with a heterozygous germline mutation in a DNA repair gene are predisposed to developing several types of cancer at an earlier age (*31–33*). In theory, only one additional mutation is sufficient to cause deficiency of an entire DNA-repair pathway in these patients, which in turn induces a tissue-specific increase in the mutation rate (**Chapter 4**). Cells in healthy individuals, however, will only become DNA repair-deficient after obtaining two hits in the same DNA repair pathway. To date, 7 out of 30 COSMIC mutational signatures have been linked to DNA repair-deficiency (and 2 additional signatures in an unpublished manuscript

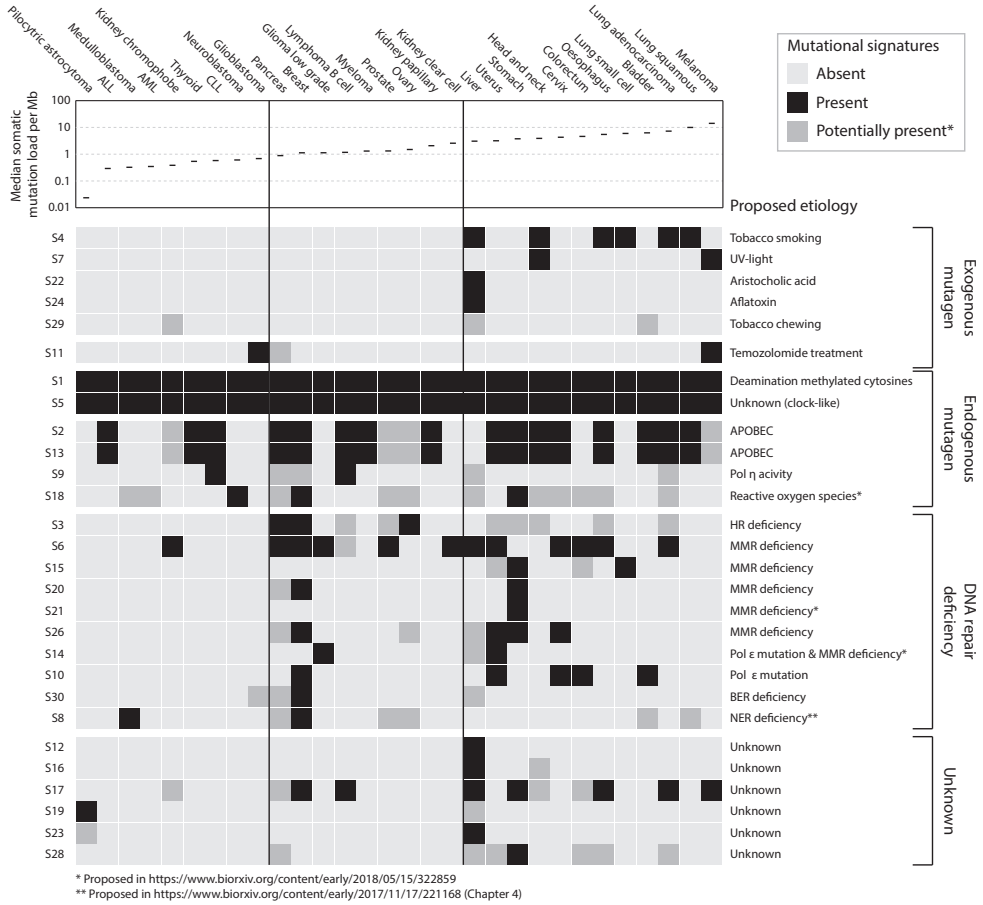


Figure 1. Contribution of the 30 known COSMIC mutational signatures to the point mutations in genomes of 30 tumor types. Tumor types are sorted based on median mutational load (*25*) and signatures are sorted based on proposed etiology (*34*, *36*). Black squares indicate presence of a mutational signature (*36*), dark grey squares indicate proposed presence of a mutational signature (*34*), and light grey squares indicate absence of a mutational signature.

(*34*)), predominantly mismatch repair deficiency (*25*, *28*, *35*).

Nucleotide-excision repair (NER) deficiency can cause a tissue-specific increase in the number of Signature 8 mutations in ASCs of the liver and small intestine (**Chapter 4**). We hypothesized that Signature 8 mutations can be caused by oxidative stress and subsequent mutation incorporation during replication by error-prone translesion synthesis (TLS) polymerases (**Chapter 4**). However, reactive oxygen species have recently been associated with Signature 18 mutations (*34*), rather than Signature 8 mutations. We actually observe a decreased contribution of this signature in NER-deficient ASCs as compared to NER-proficient ASCs. Taken

together, this suggests that Signature 18 mutations might be caused by repair of oxidative lesions by NER and in absence of functional NER, the more error prone TLS takes over, which introduces Signature 8 mutations. Proficient DNA repair pathways can, therefore, also introduce point mutations in the genomes of ASCs during life.

When one orders tumor types based on the median point mutations load (Figure 1), several striking patterns emerge. Most tumor types show a contribution of several mutational processes (Figure 1). This trend is especially evident for liver cancer, breast cancer, pancreatic cancer, and stomach cancer. However, tumors with a high median load are often linked to exposure to exogenous mutagens, whereas tumors with a low median mutation load are almost exclusively caused by the activity of (derailed) endogenous mutational processes. Whether or not these mutational processes occur prior to tumor initiation remains to be investigated. Nevertheless, this pattern may provide some insight into the underlying etiology of signatures with as yet unknown etiology. Following this logic, signatures 12, 16, and 28 are only detected in tumors with a high median load, which may suggest that these are linked to repeated exposure to exogenous mutagens (and/or DNA repair-deficiency). Signatures 17, 19, and 23 on the other hand are found across multiple tumor types independent of the median mutational load, suggesting that these can be caused by (derailed) endogenous mutagenic processes (and/or DNA repair-deficiency).

Taken together, an age-dependent, tissue-specific increase in point mutation load (especially Signature 1 and Signature 5), together with an increased mutation rate due to exposure to exogenous mutagens and/or DNA repair deficiency may be sufficient to drive tumor initiation through stochastic introduction of driver events during life. However, the accumulation of somatic point mutations alone cannot explain tumor initiation entirely (*37*). Although the cancer risk and number of stem cell divisions differ substantially between colon, small intestine, and liver (*17*), we observe a gradual accumulation of point mutations with age at similar rate in all ASC types (~40 point mutations per year) (**Chapter 3**). Furthermore, the mutation spectra of colon and small intestine are highly similar, indicating that the mutational processes in these cells are similar throughout life (**Chapter 3**), yet cancer incidence differs considerably (*10*). We also observe a similar depletion of point mutations in coding regions and open chromatin in all three ASC types in **Chapter 3**, showing that the difference in tumor incidence cannot be explained by an increased number of mutations in functional genomic elements. Furthermore, consumption of alcohol greatly enhances the risk of developing liver cancer (*38*) and, yet, is not linked to an increased point mutation load in liver ASCs (**Chapter 5**). Consistently, alcohol only enhances genomic instability in mouse blood cells that are already DNA repair-deficient or cannot metabolize acetaldehyde (*39*). These results suggest that cancer

is not just caused by the incorporation of mutations during replication, or bad luck, and that additional events are required.

In addition to point mutations, small insertions and deletions (indels) (*40*), structural variations (SVs), and copy number alterations (CNAs) might also contribute to tumor initiation. Due to the fact that indel calling results in a high number of false positives, even after extensive filtering (data not shown), we did not look at the oncogenic potential of indels in this thesis. We sporadically observed chromosomal aneuploidies in liver and colon ASCs of individuals of > 65 years old (**Chapters 3 and 5**), and we observed one complex, unbalanced SV in a colon ASC in **Chapter 3** (*23*). Since the cancer risk is the highest in colon (as compared to the other assessed tissue types) (*17*), and old individuals, our results could indicate that CNAs might also play a role in tumor initiation. Consistently, aneuploidies occur frequently early during tumorigenesis and are believed to be driver events as well (*41–43*). In contrast to these findings however, in **Chapter 5** we show that an increased risk of developing cancer is not necessarily accompanied by an increase in the number of CNAs in precancerous liver ASCs. Therefore, the number of observations needs to be increased to shed light onto these contradicting results, and on a potential role of CNAs in the genomes of ASCs in tumor initiation.

The cellular (micro-)environment drives selection of precancerous cells

In spite of the clear (and very intuitive) association between an increased number of driver events and an increased cancer risk, the introduction of tumor driver events is not sufficient to cause cancer in all tissue types (*17*, *44*). Cancer driver mutations are, for example, frequently observed in normal sun-exposed skin in the human eye lid (*44*). In the livers of mice, conditional introduction of tumor driver events did not cause an increase in cancer incidence (*17*). These observations suggest that, in addition to mutations, non-mutational mechanisms also play a pivotal role in tumor initiation. One of these mechanisms might be the variation in epigenetic regulation of gene expression (*45–47*). A progressive increase of methylation at promoters (*47*) can cause a progressive decrease in the expression of these genes (*46*). As opposed to nonsense mutations, this more gradual process allows cells to slowly adapt to the reduced expression of genes, thereby reducing potential lethality of these changes. This mechanism is suggested to be especially important in the generation of a 'second-hit' in a tumor suppressor gene (*46*). Since the epigenetic landscape differs between tissues and gradually changes with age, epigenetic variation might even explain (part of) the tissue-specific age-dependent cancer risk (*48*, *49*).

Another non-mutational cancer-driving process is the selection for "precancerous" cells (cells with tumor driver mutations in their genomes). In the

human eye lid, clonal patches of cells that carry the same tumor driver mutation are frequently observed, indicative of positive selection of single precancerous cells (*44*). Similarly, driver mutations in *APC* and *KRAS* confer an outgrowth benefit to small intestinal stem cells in mice (*50*). Interestingly, the cellular microenvironment can enhance this positive selection for precancerous cells. Although *TP53* mutations do not necessarily provide a selective benefit in normal mouse colon crypts, inflammation in the colon can enhance clonal outgrowth of TP53-mutated ASCs (*50*). This observation provides an intriguing explanation to the known link between inflammatory diseases, such as inflammatory bowel disease, and increased cancer risk (*51*). Nevertheless, it remained unclear whether this damaged cellular environment would be sufficient to drive cancer, or whether an increase in the mutational load is also required (*17*).

Chronic inflammation appears to be required for liver cancer, and liver cancer rarely occurs in absence of damage to the liver (*17*, *52*). In **Chapter 5** we present a model on how damage to the liver alone might be sufficient to drive liver cancer. Chronic alcohol consumption does not seem to directly affect the somatic mutation rate or mutational patterns in liver ASCs. Yet, we do detect a higher number of potential oncogenic mutations in these stem cells as compared to healthy ASCs, potentially suggesting differences in cellular selection. In contrast to healthy liver ASCs, we even detected multiple driver events in some alcoholic liver stem cells. It is estimated that only 4 nonsynonymous point mutations in cancer genes are required to drive the development of liver cancer (*53*). This suggests that liver ASCs from alcoholics might be moving towards development of liver cancer by changes in the microenvironment, although this still needs to be formally tested.

This leaves the question why the damaged micro-environment selects for these precancerous liver cells. Chronic damage to the liver induces proliferation of the quiescent liver ASCs to aid in liver regeneration (*17*, *54*). Potentially, the increased proliferation of precancerous cells might confer a selective benefit to these cells. Indeed, although damage to the liver is required for tumor formation in adult livers, actively proliferating stem cells in neonatal livers can drive tumor formation when they acquire tumor driver mutations in a normal, non-damaged tissue-environment (*17*). Positive selection of precancerous cells might drive tumorigenesis through the following mechanism. Oncogenic mutations that stochastically accumulate with age, like the *PTPRK* mutations observed in **Chapter 5**, may allow clonal expansion of precancerous ASCs once they are required to proliferate. As a result, a larger fraction of the stem cells within a proliferation-promoting cellular environment have acquired oncogenic mutations, and the chance of a second stochastic driver event happening in a cell that already carries a driver event increases. Ultimately, this positive cellular

selection increases the chance of developing sufficient tumor driver events in a single cell to allow the development of cancer. Even after cancer initiation, these oncogenic mutations may continue to provide a clonal outgrowth benefit to cells, as cancer cells are also highly proliferative (*55*).

This mechanism provides a novel explanation to the observation that tissues with a higher proliferative rate are associated with a higher cancer risk (*10*, *11*): whereas the selective process needs to be activated in quiescent tissues like the liver, it might already be active in proliferative tissues such as the colon. In addition, it indicates that other extrinsic risk factors may drive cancer initiation by changing the micro-environment as well. Exposure to UV-light, for example, also changes the cellular environment prior to tumor initiation (*56*) and can even promote the formation of metastases after tumor initiation by induction of inflammation, irrespective of its mutagenic potential (*57*). In other words, cancer can be partially caused by "bad luck", but one should not underestimate the effects of lifestyle.

Future research

Combined with recent advances in the field, the research described in this thesis provides several angles for future research. Firstly, although we find indications of positive selection of precancerous stem cells in **Chapter 5**, we have not formally tested whether the presence of oncogenic mutations indeed provides a clonal outgrowth benefit in the liver. To this end, one could take multiple adjacent biopsies from a cirrhotic liver, and perform deep sequencing on a panel of cancer genes. The variant allele frequency of oncogenic mutations across the biopsies provides a measure for the extent of clonal outgrowth. Repeated measurements will show how common this selective process is.

Secondly, the assumption that exogenous mutagens primarily cause cancer through mutation accumulation should be revisited. It would be really interesting to measure the mutational consequences of tobacco smoking on the genomes of liver ASCs, as these are not directly exposed to the smoke. Furthermore, the mutational consequences of alcohol consumption in the mouth and esophagus should be elucidated, since the concentration of acetaldehyde is the highest in saliva (*52*). These analyses will increase our understanding of the role of repeated exposure to exogenous mutagens in tissue-specific cancer risk. Furthermore, they can create insight into the underlying etiology of the mutational signatures.

Thirdly, the role of the epigenome in tumor initiation should be elucidated further, as it seems evident that changes in the epigenetic regulation of gene expression can drive tumor initiation. To this end, the transcriptomes (or epigenomes) of healthy and precancerous ASCs need to be compared. In **Chapter 5** of this thesis,

we performed RNA-sequencing on ASC cultures from healthy and alcoholic liver, allowing us to measure the consequences of alcohol use on gene expression. However, we only detected 9 significantly differentially expressed genes between these ASCs (data not shown; available upon request). This could indicate that the epigenome and transcriptome of precancerous liver ASCs are similar to healthy liver ASCs. Alternatively, the epigenome and transcriptome of ASCs during culture might not fully reflect those of ASCs during life. In order to identify whether organoid culturing affects the epigenetic landscape and/or transcriptome, one could FACS-sort GFP-marked Lgr5+ ASCs from mouse tissue and from mouse organoid cultures of the same tissue, and compare the sequencing results obtained with these ASCs.

Future research should also identify whether small insertions and deletions (indels), SVs, and CNAs play a causal role in the development of cancer. To this end, both the accumulation of these mutations in the genomes of tissue-specific ASCs prior to tumor initiation as well as the oncogenic potential of these mutations should be determined. Determination of the mutational load is, however, challenging for all mentioned mutation types due to technical limitations. Both indel calling and SV/CNA calling results in a high number of false positives even after extensive filtering (data not shown) and the high number of false positives can overshadow the low number of true mutations in genomically stable ASCs. Efforts should be done to improve the calling and filtering of indels, SVs, and CNAs. Furthermore, as the number of somatic CNAs and SVs in the genomes of tissue-specific ASCs is typically low, it is especially challenging to pinpoint a potential causal role of these mutations in the development of cancer. A reliable load can only be determined by sequencing the genomes of many ASCs. Alternatively, ultra-deep sequencing of tissue biopsies might provide a most cost-effective approach to determine mutation accumulation prior to tumor initiation (*44*).

Finally, although the characterization of the COSMIC signatures has provided interesting clues into mutagenesis, there are some pitfalls to the use of these signatures to explain mutational processes. One of them is the fact that some signatures are very similar (e.g. Signature 5 and 16), whereas others are very distinct (e.g. Signature 13) (*58*). This can bias the reconstruction of mutational profiles using these 30 signatures. Potentially, generating new mutational signatures that are not necessarily a reflection of the underlying biological process, but which are evenly distinct from each other could improve the clinical application of these signatures. There are, however, additional pitfalls. Firstly, the mutational signatures were identified using a large number of tumor exomes and only a few tumor whole-genomes, which biases the identified mutational signatures towards genes. As we observe a depletion of mutations in genes (**Chapter 3**), the genome-wide mutational patterns

might be different. Therefore, it would be better to extract signatures from exomes and whole-genomes separately. This will also allow researchers to use either exome or genome signatures of mutational processes, based on the type of data that they have. Furthermore, the distribution of cancer types used to extract the mutational signatures is uneven, and therefore the signatures are biased towards certain cancer types, such as breast cancer (*25*). As mutagenic processes can be different between tissues, a more equal representation of each tumor type might provide better signatures. Combined, these improvements to the extraction of signatures could potentially increase the clinical application of the mutational signatures. In the (near) future, new signatures will be released, which include specific signatures for tandem base substitutions and for indels as well (*34*). Although these new signatures are a significant improvement as compared to the old version (e.g. more genome sequences were used to extract the signatures), the mentioned issues have not been resolved yet.

Conclusions

Cancer is a major cause of death, with over 8 million deaths per year worldwide (*8*, *59*). The incidence is expected to rise in the next decades, due to westernization of lifestyle in non-western countries and increased life expectancy (*60*). Therefore, it is becoming even more important to understand the molecular and environmental causes of cancer, to facilitate the development of strategies aimed to prevent the development of cancer. Measuring mutations that accumulated in genomically-stable ASCs can improve our understanding of tumorigenesis, as these cells are believed to be the cell-of-origin of cancer. We developed a novel technique, which provided us with the unique opportunity to measure somatic mutations that occur in the genomes of ASCs during life, prior to the development of cancer. Surprisingly, the point mutation load in ASCs is not linked to a tissue-specific cancer risk, at least for intestine and liver. The accumulation of complex structural variations and the selection of precancerous stem cells (by a tumor-promoting microenvironment), however, might play a pivotal and underestimated role in tumor development. Ultimately, a tissue-specific combination of mutational processes, epigenetic variation, and cellular selection is most likely involved in tumor initiation in each patient. Although many questions still remain unanswered, such as the role of epigenetic variation in cancer risk, the research described in this thesis brings us one step closer to understanding the origin of cancer.

ACKNOWLEDGEMENTS

The authors thank Sjors Middelkamp for providing (textual) comments.

REFERENCES

1. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell*. **144**, 646–674 (2011).
2. P. Armitage, R. Doll, The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer*. **91**, 1983–1989 (2004).
3. V. Bouvard *et al.*, A review of human carcinogens—Part B: biological agents. *Lancet Oncol.* **10**, 321–322 (2009).
4. B. Secretan *et al.*, A review of human carcinogens—Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *Lancet Oncol.* **10**, 1033–1034 (2009).
5. F. El Ghissassi *et al.*, A review of human carcinogens—Part D: radiation. *Lancet Oncol.* **10**, 751–752 (2009).
6. E. R. Fearon, Human cancer syndromes: clues to the origin and nature of cancer. *Science*. **278**, 1043–1050 (1997).
7. G. Danaei *et al.*, Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet*. **366**, 1784–1793 (2005).
8. J. Ferlay *et al.*, Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer*. **136**, E359–86 (2015).
9. Cancer Statistics Review, 1975-2015 - SEER Statistics, (available at https://seer.cancer.gov/csr/1975_2015/).
10. C. Tomasetti, B. Vogelstein, Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*. **347**, 78–81 (2015).
11. C. Tomasetti, L. Li, B. Vogelstein, Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*. **355**, 1330–1334 (2017).
12. A. I. Rozhok, G. M. Wahl, J. DeGregori, A Critical Examination of the "Bad Luck" Explanation of Cancer Risk. *Cancer Prev. Res.* . **8**, 762–764 (2015).
13. C. R. Weinberg, D. Zaykin, Is bad luck the main cause of cancer? *J. Natl. Cancer Inst.* **107** (2015), doi:10.1093/jnci/djv125.
14. S. Nik-Zainal *et al.*, The life history of 21 breast cancers. *Cell*. **149**, 994–1007 (2012).
15. N. Barker *et al.*, Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature*. **457**, 608–611 (2009).
16. P. D. Adams, H. Jasper, K. L. Rudolph, Aging-Induced Stem Cell Mutations as Drivers for Disease and Cancer. *Cell Stem Cell*. **16**, 601–612 (2015).
17. L. Zhu *et al.*, Multi-organ Mapping of Cancer Risk. *Cell*. **166**, 1132–1146.e7 (2016).
18. M. Jager *et al.*, Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures. *Nat. Protoc.* **13**, 59–78 (2018).
19. R. Meza, J. Jeon, S. H. Moolgavkar, E. G. Luebeck, Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16284–16289 (2008).
20. C. Harding, F. Pompei, R. Wilson, Peak and decline in cancer incidence, mortality, and prevalence at old ages. *Cancer*. **118**, 1371–1386 (2012).
21. B. Milholland, A. Auton, Y. Suh, J. Vijg, Age-related somatic mutations in the cancer genome. *Oncotarget*. **6**, 24627–24635 (2015).
22. C. Tomasetti, B. Vogelstein, G. Parmigiani, Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proceedings of the National Academy of Sciences*. **110**, 1999–2004 (2013).
23. F. Blokzijl *et al.*, Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*. **538**, 260–264 (2016).
24. S. Nik-Zainal *et al.*, Landscape of somatic mutations in 560 breast cancer whole genome sequences. *Nature*. **534**, 47 (2016).
25. L. B. Alexandrov *et al.*, Signatures of mutational processes in human cancer. *Nature*. **500**, 415–421 (2013).
26. L. B. Alexandrov *et al.*, Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).

27. J. H. J. Hoeijmakers, DNA damage, aging, and cancer. *N. Engl. J. Med.* **361**, 1475–1485 (2009).

28. M. Petljak, L. B. Alexandrov, Understanding mutagenesis through delineation of mutational signatures in human cancer. *Carcinogenesis*. **37**, 531–540 (2016).

29. E. Hodis *et al.*, A landscape of driver mutations in melanoma. *Cell*. **150**, 251–263 (2012).

30. P. M. K. Westcott *et al.*, The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature*. **517**, 489–492 (2015).

31. H. T. Lynch, C. L. Snyder, T. G. Shaw, C. D. Heinen, M. P. Hitchins, Milestones of Lynch syndrome: 1895-2015. *Nat. Rev. Cancer*. **15**, 181–194 (2015).

32. F. J. Couch, K. L. Nathanson, K. Offit, Two decades after BRCA: setting paradigms in personalized cancer care and prevention. *Science*. **343**, 1466–1470 (2014).

33. A. Dupuy, A. Sarasin, DNA damage and gene therapy of xeroderma pigmentosum, a human DNA repair-deficient disease. *Mutat. Res.* **776**, 2–8 (2015).

34. L. Alexandrov *et al.*, The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv* (2018), p. 322859.

35. J. Drost *et al.*, Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science*. **358**, 234–238 (2017).

36. Cosmic, Signatures of Mutational Processes in Human Cancer, (available at https://cancer.sanger.ac.uk/cosmic/signatures).

37. M. J. Bissell, W. C. Hines, Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat. Med.* **17**, 320–329 (2011).

38. V. Bagnardi *et al.*, Alcohol consumption and site-specific cancer risk: a comprehensive dose–response meta-analysis. *Br. J. Cancer*. **112**, 580–593 (2014).

39. J. I. Garaycoechea *et al.*, Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nature*. **553**, 171–177 (2018).

40. I. Martincorena *et al.*, Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*. **171**, 1029–1041.e21 (2017).

41. P. Duesberg, R. Li, Multistep carcinogenesis: a chain reaction of aneuploidizations. *Cell Cycle*. **2**, 202–210 (2003).

42. B. Vogelstein *et al.*, Genetic Alterations during Colorectal-Tumor Development. *N. Engl. J. Med.* **319**, 525–532 (1988).

43. Z. Kan *et al.*, Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* **23**, 1422–1433 (2013).

44. I. Martincorena *et al.*, Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. **348**, 880–886 (2015).

45. A. P. Feinberg, M. A. Koldobskiy, A. Göndör, Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat. Rev. Genet.* **17**, 284–299 (2016).

46. P. A. Jones, S. B. Baylin, The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).

47. A. E. Teschendorff *et al.*, Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. *Genome Med.* **4**, 24 (2012).

48. M. P. Boks *et al.*, The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One*. **4**, e6767 (2009).

49. B. C. Christensen *et al.*, Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.* **5**, e1000602 (2009).

50. L. Vermeulen *et al.*, Defining stem cell dynamics in models of intestinal tumor initiation. *Science*. **342**, 995–998 (2013).

51. A. Mantovani, P. Allavena, A. Sica, F. Balkwill, Cancer-related inflammation. *Nature*. **454**, 436–444 (2008).

52. H. K. Seitz, F. Stickel, Molecular mechanisms of alcohol-mediated carcinogenesis. *Nat. Rev. Cancer*. **7**, 599–612 (2007).

53. I. Martincorena *et al.*, Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*.

171, 1029–1041.e21 (2017).

54. W.-Y. Lu *et al.*, Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nat. Cell Biol.* **17**, 971–983 (2015).

55. D. Hanahan, R. A. Weinberg, The Hallmarks of Cancer. *Cell.* **100**, 57–70 (2000).

56. M. R. Zaidi *et al.*, Interferon-γ links ultraviolet radiation to melanomagenesis in mice. *Nature.* **469**, 548–553 (2011).

57. T. Bald *et al.*, Ultraviolet-radiation-induced inflammation promotes angiotropism and metastasis in melanoma. *Nature.* **507**, 109–113 (2014).

58. F. Blokzijl, R. Janssen, R. van Boxtel, E. Cuppen, MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).

59. B. W. Stewart, C. P. Wild, *World Cancer Report 2014* (2014).

60. F. Bray, A. Jemal, N. Grey, J. Ferlay, D. Forman, Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. *Lancet Oncol.* **13**, 790–801 (2012).

Wait, there is more?

Addendum

Nederlandse samenvatting

Kanker is een ziekte van het genoom

Het genoom bevat alle erfelijke informatie die nodig is om een organisme te vormen, te laten functioneren en voort te laten planten. De genetische informatie is verdeeld over meerdere DNA moleculen (chromosomen), die elk bestaan uit twee lange kettingen van 4 basen: adenine (A), thymine (T), cytosine (C) en guanine (G). De twee strengen van basen zijn om elkaar heen gedraaid en vormen een dubbele helix, waarin een A altijd tegenover een T zit en een C altijd tegenover een G. Elke cel in het menselijk lichaam bevat 12 miljard basen, verdeeld over 46 chromosomen. In het humane genoom bevinden zich zo'n 20.000 genen. Deze coderen voor eiwitten, de werkpaarden van de cel. Eiwitten voeren taken uit zoals het kopiëren van het DNA en het reguleren van de celdeling.

Het genoom is voortdurend onderhevig aan stoffen en processen die schade kunnen veroorzaken. Die beschadigingen zorgen ervoor dat de informatie in het genoom niet langer toegankelijk is. Het merendeel van de schade wordt gerepareerd door DNA-reparatie processen. Soms gaat dit echter mis, waardoor er veranderingen (mutaties) in de volgorde van de basen in het genoom, de genoomsequentie, ontstaan. Een enkele mutatie kan de structuur van een eiwit veranderen. Dit kan leiden tot een verminderd functioneren van een eiwit en uiteindelijk ook tot verminderd functioneren van een cel. Wanneer er meerdere mutaties (circa 4 mutaties) in specifieke genen in het genoom van 1 cel plaatsvinden, kan dit zelfs leiden tot het ontstaan van kanker. Aangezien de accumulatie van mutaties een continu proces is, neemt het aantal mutaties toe met de leeftijd. Daarom is de kans op kanker groter bij hogere leeftijd. Om kanker beter te kunnen voorkomen en gerichter te kunnen behandelen is het belangrijk om inzicht te verkrijgen in de (mutatie)processen die bijdragen aan het ontstaan van kanker.

Met behulp van "Next-Generation Sequencing"-technieken kunnen we de sequentie van het genoom bepalen (sequencen). Op plekken waar deze sequentie afwijkt van het genoom dat iemand heeft meegekregen bij de geboorte is gedurende het leven een mutatie opgetreden. Het type mutatie en de basen direct voor en na de mutatie, het mutatieprofiel, kunnen inzicht verschaffen in welk proces de mutatie heeft veroorzaakt. Blootstelling aan UV-licht zorgt er bijvoorbeeld voor dat cytosines in thymines veranderen (C > T) op TCN sequenties (TCN > TTN; N = iedere base). In de afgelopen jaren zijn er 30 mutatieprofielen van mutatieprocessen, zogeheten signatures, in kaart gebracht door de genomen van vele tumoren te sequencen. In figuur 3 van **hoofdstuk 1** is bijvoorbeeld te zien hoe deaminatie van gemethyleerde cytosines (Signature 1) onderscheiden kan worden van mutaties door blootstelling aan UV-licht (Signature 7).

Adulte stamcellen

Hoewel het in elke cel belangrijk is dat de genoomsequentie stabiel is, is het vooral cruciaal in adulte stamcellen. Adulte stamcellen zijn 'multipotente' stamcellen die kunnen veranderen in alle celtypen van het weefsel waarin ze zich bevinden. De adulte stamcellen vullen op deze manier het 'celtekort' in een orgaan aan als er cellen doodgaan. Je kunt je voorstellen dat het van groot belang is dat deze stamcellen genomisch stabiel blijven, want als een stamcel een mutatie oploopt bezitten alle nakomelingen van deze stamcel die verandering ook. Bij dikke darmkanker is reeds aangetoond dat adulte stamcellen dé cel kunnen zijn die een tumor vormt. Inzicht in de opeenstapeling van mutaties in genomen van adulte stamcellen kan daarom ons inzicht vergroten in de processen die bijdragen aan het ontstaan van kanker.

Het was echter om meerdere redenen gecompliceerd om het genoom van een gezonde adulte stamcel te bestuderen. Allereerst is het voor de meeste organen (vooralsnog) lastig tot onmogelijk om enkel adulte stamcellen eruit te selecteren. Een weefsel bestaat voornamelijk uit gewone 'functionele' cellen en maar voor een heel klein deel uit stamcellen. Als je lukraak alle cellen zou gaan sequencen, zou je om die reden bijna geen stamcellen sequencen. Er moet dus een methode zijn om adulte stamcellen te selecteren. Bovendien moet je het genoom van een enkele stamcel sequencen om betrouwbaar mutaties op te kunnen sporen. Elke cel loopt gedurende het leven zijn eigen set mutaties op. Als je het genoom van 30 cellen tegelijk sequencet, is het om technische redenen niet mogelijk die mutaties te onderscheiden van de fouten die tijdens het sequencen gemaakt worden. Als je echter het genoom van 1 cel 30 keer sequencet, kan je de mutaties wel goed opsporen, omdat deze vaker voorkomen in de sequencing data dan de sequencing fouten. Er moet dus een manier zijn om 1 adulte stamcel te vermeerderen totdat er genoeg cellen zijn om het genoom betrouwbaar te sequencen.

In **hoofdstuk 2** van dit proefschrift wordt een nieuw protocol beschreven dat het mogelijk maakt om de genoomsequentie van enkele stamcellen te bepalen. Hierbij wordt gebruik gemaakt van zogeheten organoïde kweken (organoids). Organoids zijn 'mini-organen' die in de celkweek in 3D kunnen groeien. Organoids verschaffen een (in theorie) oneindige hoeveelheid cellen voor regeneratieve therapieën. Bovendien zijn adulte stamcellen de enige cellen die een organoïde kweek kunnen vormen, waardoor er selectie voor deze cellen mogelijk is. Organoids kunnen gekweekt worden van allerlei weefsels en zelfs van tumoren. Elk type kweek vereist echter wel specifieke condities om de groei te bevorderen. In **hoofdstuk 6** wordt beschreven hoe gezonde cellen en tumorcellen van de alvleesklier gekweekt kunnen worden als organoids.

In het nieuwe protocol, beschreven in hoofdstuk 2, worden enkele stamcellen

als organoid vermeerderd totdat er genoeg cellen zijn om het hele genoom te sequencen. Vervolgens kunnen de mutaties bepaald worden met behulp van een nieuw ontwikkeld bioinformatisch stappenplan (een bioinformatische pipeline). Mutaties die in de oorspronkelijke stamcel aanwezig waren, zitten in alle cellen en kunnen daarom goed opgespoord worden. Mutaties die in de kweek hebben plaatsgevonden zijn eruit te filteren op basis van de frequentie waarmee deze voorkomen in de sequencing data. De techniek is beschreven voor leverstamcellen en kan toegepast worden op alle adulte stamcellen die als organoid gekweekt kunnen worden.

Genomische stabiliteit van adulte stamcellen

In **hoofdstuk 3** is het nieuwe protocol gebruikt om de genomische stabiliteit van gezonde lever- en (dikke en dunne) darmstamcellen van mensen te bepalen. Hiervoor zijn stamcellen van biopten van donoren geïsoleerd en vervolgens is de genoomsequentie van deze stamcellen vastgesteld. Zoals verwacht, neemt het aantal mutaties toe met de leeftijd. Elk jaar treden ongeveer 40 puntmutaties (mutaties waarbij een enkele base verandert in een andere base) op per stamcel in alle drie de weefsels. In de darm wordt het merendeel van de mutaties veroorzaakt door deaminatie van gemethyleerde cytosines (Signature 1 mutaties). In de lever is een ander proces actief (Signature 5 mutaties). Aangezien mensen een veel grotere kans hebben op het ontwikkelen van darmkanker dan van leverkanker, is het verrassend dat de toename in het aantal mutaties hetzelfde is voor alle drie de weefsels. De aanwezigheid van grotere 'complexere' mutaties in de genomen van dikke darmstamcellen verklaart mogelijk het verschil in incidentie van kanker in de drie hiervoor genoemde weefsels.

In **hoofdstuk 4** zijn de mutationele consequenties van het missen van een DNA-reparatie proces beschreven. Hiervoor zijn stamcellen gesequencet uit de lever en uit de dunne darm van muizen die een DNA-reparatie eiwit missen, te weten *ERCC1*. Dit eiwit speelt een essentiële rol in nucleotide excisie reparatie (NER). In de lever leidt deficiëntie van dit eiwit tot een toename in het aantal puntmutaties. Het spectrum van de mutaties lijkt op Signature 8 in NER-deficiënte levercellen. In de darm leidt het verlies van *ERCC1* niet tot een toename van het aantal mutaties, maar neemt het aantal Signature 8 mutaties eveneens toe. Om te achterhalen of de link tussen dit mutatie signature en NER ook in mensen aanwezig is, is vervolgens de genoomsequentie bepaald van een menselijke dunne darm organoid waarin NER gedeactiveerd is. Ook hier neemt het aantal mutaties toe, waarbij het merendeel van de toename verklaard kan worden door Signature 8. NER-deficiënte tumoren reageren goed op bepaalde behandelingen (zoals cisplatine). Een verhoogd aantal

Signature 8 mutaties in het genoom van een tumor is derhalve mogelijk een indicatie dat cisplatine een goede therapie is voor de desbetreffende patiënt.

Tot slot wordt in **hoofdstuk 5** het effect van alcoholconsumptie op de genomische stabiliteit van leverstamcellen beschreven. Alcohol wordt geassocieerd met een verhoogd risico op verschillende typen kanker, waaronder leverkanker. Echter is het niet bekend waarom alcohol het risico op kanker verhoogt. Om dit te achterhalen, is de genoomsequentie bepaald van leverstamcellen van alcoholisten die een levertransplantatie hebben ondergaan. Vervolgens zijn de verkregen mutatieprofielen vergeleken met de mutatieprofielen van leverstamcellen van gezonde individuen. Alcoholisten hebben een sterk verhoogde kans op het ontwikkelen van leverkanker. Verrassend genoeg heeft de alcoholconsumptie echter geen direct effect op de sequentie van het genoom. Er zijn wel meer stamcellen met mutaties in genen die in tumoren ook vaak een mutatie hebben. Wellicht verandert alcoholconsumptie de cellulaire omgeving van stamcellen, waardoor stamcellen die toevallig een mutatie oplopen in een kankergen gemakkelijker uitgroeien. Uiteindelijk vergroot dit de kans dat meerdere kankergenen in 1 cel mutaties oplopen, hetgeen bijdraagt aan een verhoogd risico op tumorvorming.

In **hoofdstuk 7** van dit proefschrift worden de resultaten samengevat, in een bredere context geplaatst en worden toekomstige uitdagingen besproken. Het werk dat in dit proefschrift beschreven is, illustreert dat een verhoogd risico op kanker niet per se veroorzaakt wordt door een hoger aantal puntmutaties. Zowel de accumulatie van grotere complexere mutaties in de genoomsequentie als de selectie van specifieke cellen door de cellulaire omgeving spelen mogelijk een essentiële rol in het ontstaan van kanker. Aangezien kanker een ziekte is van het genoom, vormt het verkrijgen van meer inzicht in mutatieprocessen een belangrijke stap richting het doorgronden van kanker. Dit zal uiteindelijk resulteren in verbeterde methoden die gericht zijn op de preventie en behandeling van kanker.

Dankwoord

Enfin, daar gaat-ie: promotie-onderzoek doe je niet alleen. Iedereen die mij de afgelopen jaren heeft geholpen, gesteund en bijgestaan ben ik daarom erg dankbaar.

Allereerst: Edwin, bedankt voor alles de afgelopen jaren. Ik heb enorm veel geleerd van jou als begeleider en als wetenschapper. Vooral jouw zakelijke/gefocuste/pragmatische insteek in de wetenschap vind ik inspirerend. Ik ben je dankbaar dat je me de kans hebt geboden om in jouw lab te promoveren, waar naast een goede wetenschappelijke sfeer ook waarde gehecht wordt aan een lekker borreltje (en hamka's) op zijn tijd. Ook al was je soms druk, als PhD student was ik duidelijk een prioriteit. Het was dan ook eerder regel dan uitzondering dat je binnen een paar uur al feedback had gegeven op een nieuw manuscript. Ik wil je ook bedanken voor alle persoonlijk advies die je de afgelopen jaren gegeven hebt. Je hebt me enorm goed begeleid en dankzij jou heb ik het beste uit mezelf kunnen halen.

Ruben, ook jou wil ik bedanken voor alles. Jouw onbegrensde enthousiasme heeft me (vrijwel) moeiteloos door de afgelopen jaren heen geloodst. Bovendien was het fijn dat je zoveel vertrouwen had in mijn capaciteiten, als ik dat soms zelf even minder had. Dat was echt een opsteker. Ik ben je vooral dankbaar dat je me zo goed hebt begeleid tijdens mijn eerste jaren als PhD student, waardoor de 'sprong in het diepe' een stuk minder diep leek. Ik heb heel veel van je geleerd en het belangrijkste wat je me (onbewust) hebt meegegeven is een optimistische blik in het geval van tegenslagen. Heel veel succes met je eigen onderzoeksgroep op het PMC!

Hans Bos en Wouter de Laat, bedankt voor jullie wijze raad tijdens onze jaarlijkse PhD commissie meetings. Het is altijd fijn wanneer iemand (anders dan je promotor of copromotor) zegt dat het normaal is dat je na het eerste jaar nog geen vijf publicaties op zak hebt. I also want to thank my reading committee for taking the time to critically read this dissertation. Thank you Paul Coffer, Susanne Lens, Marcel Tijsterman, Puck Knipscheer, and Hans Bos. Bovendien wil ik ook iedereen bedanken die de afgelopen maanden (delen van) dit proefschrift van feedback hebben voorzien (ook al was de tijd soms krap). Edwin, Ruben, Sjors, Melvin, Roos, Sake en Bianca: bedankt voor alle feedback.

I would also like to thank the entire Cuppen group for the past years. I have thoroughly enjoyed every moment with you, both scientific and social. So thank you Joep, Arne, Mauro, Jerome, Roelof, Anna, Petra, Pjotr, Martin, Ies, Marieke, Ewart, Lisanne, Robin,

Wensi, Wim, Pim, Nico, Esther, Ilse, Annelies, Stef, Terry, Robert, Bastiaan, Marlous, Simone, and Monique. Sommige mensen in het Cuppen lab wil ik nog specifiek bedanken, omdat ik niet zonder jullie kan en/of dit proefschrift er heel anders uit zou hebben gezien zonder jullie. Ewart, ik zou met alle plezier weer 2 maanden fulltime naast je celkweken in 1 kast. Succes met programmeren ;)! Mark, ik hoop dat we ooit weer een lab bench kunnen delen. Silvia and Melissa, thank you for being the best students a supervisor could ever wish for. Nicolle, Sander en Roel, ik kan miljoenen goede dingen over jullie zeggen, maar de volgende vier woorden vatten dit perfect samen: jullie zijn de beste!

Ik wil ook graag mijn mede PhDs in de Cuppen groep bedanken. Roel en Sebas, bedankt voor de gezelligheid en de instructies gedurende mijn eerste jaar. Sharon and Luan, good luck with your PhD. Maar bovenal Francis, Sjors en Judith, zonder jullie waren de afgelopen jaren heel anders (en veel minder gezellig) geweest. Ik hoop dat onze paden in de toekomst weer kruizen. Judith, straks ben je senior PhD student in de Cuppen groep. Dat brengt natuurlijk allerlei verantwoordelijkheden met zich mee, zoals borrels organiseren. Heel veel succes de aankomende jaren! Sjors, ik ga je genuanceerde commentaar en wetenschappelijke inzicht missen. Nog een paar maanden en dan ben jij ook klaar. Ik hoop dat je snel een leuke nieuwe baan vindt. Succes met de laatste loodjes! Francis, het was enorm fijn om het gehele promotie-traject tegelijk met jou bij Edwin te doorlopen. Samen naar New York en the grand city of Doorwerth (waar ik je moeder nog ontmoet heb). En tijdens alle retreats en masterclasses was je mijn vaste roomie. Terwijl we in de tussentijd ook nog een paar manuscripten samen schreven. Ons gedeelde perfectionisme heeft een aantal mooie hoofdstukken voor ons beide opgeleverd ("als je klikt waar ik nu ben… ben je er?...ja dit woord…ik weet niet…het klinkt net niet lekker"). Bedankt voor alles en heel veel succes tijdens je verdere loopbaan!

Ook wetenschappers en collega's buiten de Cuppen groep ben ik dankbaar. Ik wil de Clevers groep bedanken voor alle hulp en leerzame tips over organoids. Bedankt Hans, Valentina, Sylvia, Marc, Karien, Stieneke, Harry, Jeroen en Ana. I thank our collaborators for the nice (and fruitful) collaborations. Together we really gathered some interesting insights into mutation accumulation. Ruby, Luc, Monique, Johanna, Maria, Jan and Joris, I hope you continue to work with the Cuppen group for many years! Thank you fellow PhD students Daniëlle, Glen, Nayia, Maartje, Arianna, Jasmin, Loes, Ivar, Carien, João and many, many, many more (too many people to mention by name) for all the fun retreats, borrels, conversations in the lab, dinners, and masterclasses. Genoomdocenten Marc, Elly, Francis, Bob, Loes, Charlotte, Sasja,

Elianne, Rianne, Laurens en Sietske: bedankt voor alle inspirerende vergaderingen over studenten. Ik had het lesgeven van tevoren toch ietwat onderschat en met hulp van jullie viel het inderdaad allemaal mee. Wigard en Menno, toen ik begon met mijn Master wist ik zéker dat ik geen PhD student wilde worden. De stages die ik onder jullie begeleiding gelopen heb, hebben mijn gedachten doen veranderen. Dank daarvoor!

Tot slot wil ik nog de mensen bedanken die mij altijd oneindig steunen: mijn familie en mijn schoonfamilie. Papa (ja, je staat er toch in!) en mama, jullie hebben mij altijd gezegd dat ik eruit moet halen wat erin zit. Mede dankzij jullie ben ik dan ook een PhD gaan doen. Ik weet dat ik altijd op jullie kan rekenen en dat jullie enorm trots op me zijn. Ik wil jullie danken voor deze oneindige steun! Mam, toen ik vier jaar geleden vol enthousiasme vroeg of ik foto's van jou in mijn proefschrift mocht gebruiken, wist je denk ik niet dat het zoveel werk zou zijn. Ik ben een aantal keer van gedachten veranderd ("mam, ik wil toch iets heel anders"), maar dat maakte je niet veel uit. Het belangrijkste voor jou was dat ik het een mooi proefschrift zou vinden. Het eindresultaat mag er wat mij betreft zijn. Bedankt voor alle input, mails en brainstorm sessies over de lay-out!

Roos en Mick, ook jullie enorm bedankt voor alle support! Deze kwam vooral in de vorm van de broodnodige afleiding, zoals wanneer we samen Nyenrode onveilig gingen maken (gelukkig zorgt Melvin voor genoeg paraplus en Mick voor genoeg tosti's voor de twee hongerige zusjes), of wanneer ik jou, Roos, uit de parkeerplaats van het UMC moest bevrijden. Maar jullie hebben ook heel wat verhalen aan moeten horen over DNA. Ik hoop dat we samen nog veel van dit soort momenten mogen bijschrijven en jullie altijd je pokerface behouden als ik weer over biologie begin! Ed, Marja, Stephen en Nicol, weinig spreekt meer steun en begrip uit dan een familie die klust aan ons huis, terwijl ik achter de computer zit te werken. En naast het feit dat jullie me op deze manier hielpen focussen, hielpen jullie me ook ontspannen door het inbouwen van rustmomenten (kopjes thee, weekendjes weg, etc.). Bedankt hiervoor, maar ook voor alle belangstelling! Ook alle andere familieleden van mij en Melvin (Oma! Beppe! ooms, tantes, neven, nichten, achterneefjes en achternichten): heel erg bedankt voor alle interesse. Jullie hebben mij geïnspireerd om een inleiding te schrijven die jullie hopelijk ook grotendeels begrijpen. De aankomende kerstdiners kan ik weer gewoon aanschuiven (in plaats van mijn cellen in de celkweek een kerstdiner te voeren).

Melvin, lief, ik zou een heel hoofdstuk met dankbetuigingen kunnen vullen voor jou, maar ik hou het kort zodat ik nog wat leuke anekdotes overhoud voor onze trouwgeloften in februari. Heel erg bedankt voor alle steun. Bijvoorbeeld tijdens romantische diners, die ik ruw verstoorde met allerlei "interessante feiten" over leverkanker. Of wanneer je Peaky Blinders moest terugspoelen, omdat ik uit het niets ging vertellen welke verrassende onderzoeksresultaten ik de afgelopen week had verkregen. Of wanneer je me op zondag heen en weer reed naar het UMC, zodat ik even het lab in kon. Het valt niet mee, samenwonen met een PhD student. We hebben veel mijlpalen bereikt de afgelopen jaren (eerste huis, jouw deeltijd Master of Science, tweede huis, 10 jaar samen, etc.) en er zullen er ongetwijfeld nog velen volgen. Ik hou van je!

List of publications

Ressa A, Bosdriesz E, de Ligt J, Mainardi S, Maddalo G, Prahallad A, Jager M, de la Fonteijne L, Fitzpatrick M, Groten S, Altelaar AFM, Bernards R, Cuppen E, Wessels L, Heck AJR. A system-wide approach to monitor responses to synergistic BRAF and EGFR inhibition in colorectal cancer cells. *Mol Cell Proteomics*. 2018 Jul 3.

Jager M*, Blokzijl F*, Sasselli V, Boymans S, Janssen R, Besselink N, Clevers H, van Boxtel R, Cuppen E. Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures. *Nat Protoc*. 2018 Jan;13(1):59-78.
**Equal contribution*

Blokzijl F, de Ligt J*, Jager M*, Sasselli V*, Roerink S*, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, Nijman IJ, Martincorena I, Mokry M, Wiegerinck CL, Middendorp S, Sato T, Schwank G, Nieuwenhuis EE, Verstegen MM, van der Laan LJ, de Jonge J, IJzermans JN, Vries RG, van de Wetering M, Stratton MR, Clevers H, Cuppen E, van Boxtel R. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*. 2016 Oct 13;538(7624):260-264.
**Equal contribution*

Boj SF*, Hwang CI*, Baker LA*, Chio I*, Engle DD*, Corbo V*, Jager M*, Ponz-Sarvise M, Tiriac H, Spector MS, Gracanin A, Oni T, Yu KH, van Boxtel R, Huch M, Rivera KD, Wilson JP, Feigin ME, Öhlund D, Handly-Santana A, Ardito-Abraham CM, Ludwig M, Elyada E, Alagesan B, Biffi G, Yordanov GN, Delcuze B, Creighton B, Wright K, Park Y, Morsink FH, Molenaar IQ, Borel Rinkes IH, Cuppen E, Hao Y, Jin Y, Nijman IJ, Iacobuzio-Donahue C, Leach SD, Pappin DJ, Hammell M, Klimstra DS, Basturk O, Hruban RH, Offerhaus GJ, Vries RG, Clevers H, Tuveson DA. Organoid models of human and mouse ductal pancreatic cancer. *Cell*. 2015 Jan 15;160(1-2):324-38.
**Co-first authors*

van Heesch S, Simonis M, van Roosmalen MJ, Pillalamarri V, Brand H, Kuijk EW, de Luca KL, Lansu N, Braat AK, Menelaou A, Hao W, Korving J, Snijder S, van der Veken LT, Hochstenbach R, Knegt AC, Duran K, Renkens I, Alekozai N, Jager M, Vergult S, Menten B, de Bruijn E, Boymans S, Ippel E, van Binsbergen E, Talkowski ME, Lichtenbelt K, Cuppen E, Kloosterman WP. Genomic and functional overlap between somatic and germline chromosomal rearrangements. *Cell Rep*. 2014 Dec 24;9(6):2001-10.

Kloosterman WP, Tavakoli-Yaraki M, van Roosmalen MJ, van Binsbergen E, Renkens I, Duran K, Ballarati L, Vergult S, Giardino D, Hansson K, Ruivenkamp CA, Jager M, van Haeringen A, Ippel EF, Haaf T, Passarge E, Hochstenbach R, Menten B, Larizza L, Guryev V, Poot M, Cuppen E. Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. *Cell Rep*. 2012 Jun 28;1(6):648-55.

Manuscripts in preparation

Jager M*, Blokzijl F*, Kuijk E, Bertl J, Vougioukalaki M, Janssen R, Besselink N, Boymans S, de Ligt J, Skou Pedersen J, Hoeijmakers J, Pothof J, van Boxtel R, Cuppen E. Deficiency of nucleotide excision repair explains mutational signature observed in cancer. *In preparation*

**These authors contributed equally to this work*

Jager M, Kuijk E, Lieshout R, Locati M, Besselink N, Janssen R, Boymans S, de Jonge J, IJzermans J, Doukas M, Verstegen M, van Boxtel R, van der Laan L, Cuppen E. Effect of chronic alcohol use on mutation accumulation in precancerous cirrhotic liver adult stem cells. *In preparation*

Blokzijl F*, Kuijk E*, Jager M, Chuva de Sousa Lopes S, van Boxtel R, Cuppen E. Higher mutation accumulation rate in human stem cells during fetal development than postnatal life. *In preparation*

**Equal contribution*

Kuijk E, Jager M, Locati M, van Hoeck A, van der Roest B, Korzelius J, Janssen R, Besselink N, Boymans S, van Boxtel R, Cuppen E. Genome-wide mutational impact of the in vitro culture of human pluripotent and adult stem cells. *In preparation*

Curriculum Vitae

Myrthe Jager was born on June 28 1990 in Groningen, the Netherlands. At the age of 1, she moved to Leusden. In 2008, Myrthe graduated from the Stedelijk Gymnasium Johan van Oldenbarnevelt in Amersfoort. In that same year, she started her bachelor 'Biomedische Wetenschappen' at the University of Utrecht. In 2011, she obtained her Bachelor's degree *cum laude* and continued her education with the 'Cancer Genomics and Developmental Biology' Master's programme at the University of Utrecht. Owing to her fascination with DNA, she performed two internships in the field of (epi)genomics. The first in the lab of Prof.dr.ir. Edwin Cuppen at the UMC Utrecht ("Upregulation of immunoglobulin genes in VCFS patients") and the second in the lab of Dr. Menno Creyghton at the Hubrecht Institute ("The epigenetic landscape of haploid embryonic stem cells"). After writing her Master's thesis in the lab of Prof.dr. Susanne Lens at the UMC Utrecht ("Deregulated Aurora B activity and its implications in cancer"), she received her Master's degree in 2013.

Myrthe then started her PhD research on mutations in the genomes of adult stem cells in the lab of Prof.dr.ir. Edwin Cuppen at the Hubrecht Institute and at the UMC Utrecht. The results of this research are published in this thesis. During her PhD, Myrthe was a member of the PhD committee and she also enrolled in the PhD teaching talent programme, which allowed her to obtain her basic teaching qualification (BKO) in 2017. Myrthe will continue her scientific career in the labs of Dr. Wigard Kloosterman and Dr. Jeroen de Ridder at the UMC Utrecht, working on developing and validating the next-generation of liquid biopsy tests for cancer.