

The influence of cellular transitions on breast cancer development



Linda J. van Weele

The influence of cellular transitions on breast cancer development

Linda J. van Weele

The research described in this thesis was performed at Stanford University within the framework of the Cancer, Stem Cells & Developmental Biology PhD program, Utrecht University, the Netherlands.

ISBN: 978-94-6419-508-8

Cover art: Allorah Bowen

Layout: Linda van Weele

Printed by: Gildeprint

Copyright © 2022 by Linda Jacqueline van Weele. All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means, without prior permission of the author. The copyright of the publications remains with the publishers.

The influence of cellular transitions on breast cancer development

De invloed van cellulaire transitie op de ontwikkeling van borstkanker

(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

woensdag 25 mei 2022

des middags te 2.15 uur

door

Linda Jacqueline van Weele

geboren op 22 januari 1986
te Rotterdam

Promotoren: Prof. dr. J.E. van Rheenen

Prof. dr. M.F. Clarke

Contents

Chapter 1	7
General introduction	
Thesis outline	
Chapter 2	17
Depletion of <i>Trp53</i> , <i>p16^{Ink4a}</i> , and <i>p19^{Arf}</i> does not promote self-renewal in the mammary gland but amplifies proliferation induced by TNF α	
Chapter 3	37
Mesenchymal tumor cells drive adaptive resistance of <i>Trp53</i> ^{-/-} breast tumor cells to inactivated mutant <i>Kras</i>	
Chapter 4	63
A cell-intrinsic role for TLR2-MYD88 in intestinal and breast epithelia and oncogenesis	
Chapter 5	93
Single-cell transcriptomics of 20 mouse organs creates a <i>Tabula Muris</i>	
Chapter 6	123
A single-cell transcriptomic atlas characterizes ageing tissues in the mouse	
Chapter 7	155
Discussion	
Addendum	171
Nederlandse samenvatting	
Curriculum Vitae	
List of publications	
Dankwoord/acknowledgements	

Chapter 1

General introduction Thesis outline

Linda J. van Weele¹

¹Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of
Medicine, Stanford, CA, USA

The mammary gland

Unique to mammals is the presence of the mammary gland. The mammary gland develops to provide offspring with milk and is fully developed at the start of puberty. Hormones and growth stimulate the proliferation of epithelial cells mainly at the terminal end buds (TEBs), located at the tips of the expanding ducts. The expansion leads to the formation of the epithelial ductal system that branches out into the entire mammary fat pad. During pregnancy, the mammary epithelium expands dramatically and forms alveolar cells that produce and secrete milk during lactation. Upon weaning, the mammary gland involutes through apoptosis and tissue remodeling to go back to a state that is similar to the pre-pregnancy state (Fig. 1). Since the epithelial ductal system and development of the mouse mammary gland is highly similar to the human breast, the mouse mammary gland is a suitable model to study human breast development and cancer. There are differences as well. Where humans have one pair of mammary glands, mice have five. Furthermore, the human mammary gland develops with more complexity, with ductal tree displaying a higher degree of branching than the mouse mammary tree (1, 2). Another difference is found in the surrounding stroma, in humans the epithelium is associated with fibrous connective tissue while in the mouse the stroma is adipocyte-rich with only little fibrous connective tissue (3). In this thesis, I focus on the mouse mammary gland as a model to deepen our understanding of normal and abnormal development.

Cell types

The mammary gland consists of two primary lineages: the luminal and the basal population. The luminal population contains ductal and alveolar cells. The ductal cells line the lumen of the mammary ducts. The basal population contains myoepithelial cells, contractile cells that push the milk through the ducts. The basal population also contains the mammary stem cells (4, 5). Using the fluorescence-activated cell sorter (FACS), the basal and luminal population can be distinguished using cell surface markers such as heat stable antigen (CD24) and integrin $\alpha 6$ (CD49f) (Fig. 2). Instead of CD24, epithelial cell adhesion molecule (EPCAM) is also frequently used to identify and separate the luminal and basal populations (6).

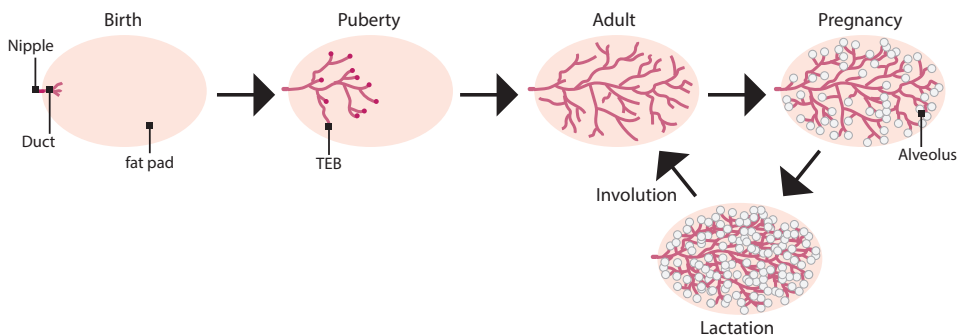


Figure 1. Schematic of postnatal mammary gland development. The epithelial ductal system branches out through the entire fat pad during puberty. During pregnancy, the epithelial cells commence proliferating again leading to the development of alveolar cells and further expansion of the ductal system. After lactation, the gland involutes and returns to a state that is similar to the pre-pregnancy state.

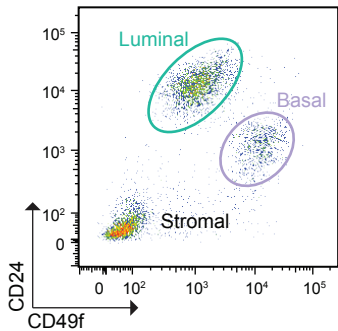


Figure 2. Flow cytometry plot of mouse mammary gland. After selecting for live cells and selecting out cells expressing lineage markers for hematopoietic cells (CD45, Ter119) and endothelial cells (CD31); CD24 and CD49f enable the separation of the mammary epithelial luminal and basal population.

Cell markers

The identification of different populations using cell surface markers, has enabled the separation of live cell populations using FACS. Live cells can be further studied using functional assays, to analyze and determine their function *in vivo* and *in vitro*. Other methods, such as bulk RNA sequencing and single cell RNA sequencing (scRNA-seq), enable the identification of different populations in an unbiased way. Using scRNA-seq, we and others have published atlases for a variety of organs and tissues in the adult mouse (7, 8). When samples are enriched for mammary epithelial cells - using the FACS sorting strategy shown in Figure 2 - separate groups of basal cells, endothelial cells, luminal cells, and stromal cells are confirmed with sc-RNAseq (Fig. 3A). Using commonly used cell surface markers for the different populations, we are able to distinguish the various populations of the mammary gland (Fig. 3B, C). The benefit of enriching for cells of interest using FACS prior to sequencing is the increased specificity for the identification of cell populations of interest. In addition, a cell-selective method like FACS enables the identification of rare cells within the mammary gland more easily. The downside is that you do not get a complete image of all the cell types that constitute, support, and interact directly with the tissue of interest. To also capture a more complete image of the different cell types present in the mammary gland environment, microfluidic droplets were used to capture the expression profile of single cells prior to sequencing. When doing so, we find that non-epithelial cells make up the largest part of the mammary gland. Other than epithelial cells; stromal cells, endothelial cells, and various hemopoietic cell types (B cells, T cells, macrophages) are detected (Fig. 4A). As the mouse ages, the fraction each population contributes to the mammary gland changes (9). The fractions of basal cells and stromal cells go up, the fraction of T cells goes down (Fig. 4B). Of note, even when aiming to capture the individual genetic information from all cell types in the mammary gland environment, not all cell types present in the mammary fat pad are detected. During the processing steps needed to provide cells as a single cell suspension needed for scRNA-seq, cells are lost. For example, although the mouse mammary fat pad is adipocyte-rich, fat cells are not detected in scRNA-seq as they are removed during tissue processing.

Self-renewal

The expansion of the mammary gland is tightly regulated by the mammary stem cells and the stem cells' ability to self-renew. Self-renewal enables stem cells to divide without changing the identity or properties of the stem cell. This contrasts with progenitors. Progenitors are produced by stem cells and are also able to divide, however progenitors have limited proliferation capacity hence their state is short-lived. Eventually the cell division of a progenitor is finite resulting in two daughter cells with a specialized function within the body (Fig. 5). In a tissue, stem cells and progenitors have the important role of creating a balance between producing sufficient new

Chapter 1

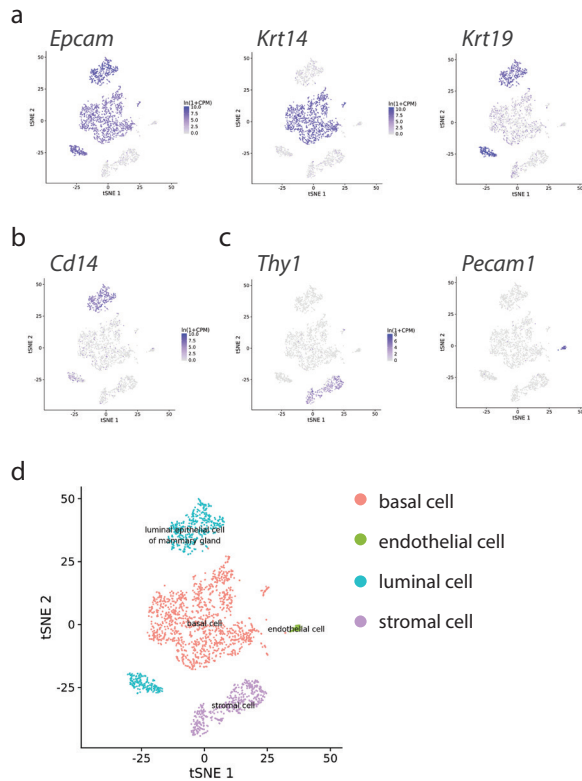


Figure 3. Mouse mammary epithelial enriched scRNA-seq. Data from 4 female mice (age: 3 months) are shown. Basal, luminal, and stromal cells were enriched for using FACS prior to sequencing. Data is accessible using the database <https://tabula-muris.ds.czbiohub.org/>. (a) tSNE plot of populations showing basal, endothelial, luminal and stromal cells. (b) Markers indentifying non-epithelial mammary gland populations. *Thy1* marks stromal cells, *Pecam1* marks endothelial cells. (c) Markers indentifying epithelial mammary gland populations. *Epcam* marks all epithelial cells (luminal and basal), *Krt14* the basal population, *Krt19* the luminal population, and *Cd14* the luminal progenitor population.

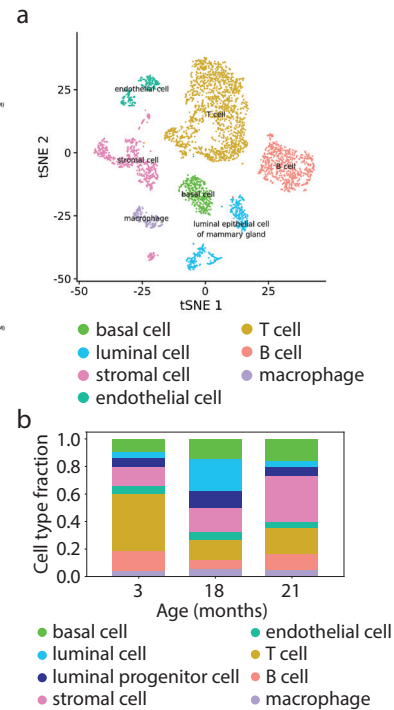


Figure 4. Mammary gland scRNA-seq data of adult mice, cells seperated using microfluidic droplets. Data from 2 female mice per age group (ages: 3, 18, and 21 months) are shown. No population enrichment took place prior to sequencing. Data is accessible using the database <https://tabula-muris.ds.czbiohub.org/>. (a) tSNE plot showing basal, luminal, stromal, endothelial, and hemopoietic (T cells, B cells, macrophages) populations. (b) Changes in cell type fractions of mice of different ages (3 months, 18 months, 21 months).

cells to maintain the integrity of a tissue while preventing excessive proliferation that results in tumorigenesis. Such homeostasis is regulated by both cell intrinsic as well as cell extrinsic signaling. In some tissues, there is a continues turnover of differentiated cells and as such, new cells need to be produced continuously. Other tissues have mostly dormant stem cells, cellular turnover is rare. In a third group, including the mammary gland, new specialized cells are produced in spurts. In the mammary gland this happens during puberty, pregnancy, and lactation (10).

Self-renewal regulation

Different expression patterns between stem cells, progenitors, and differentiated cells, are often rooted in the cells' epigenetic makeup (11). One protein involved with epigenetic regulation is BMI1. BMI1 is a chromatin remodeler that is involved with

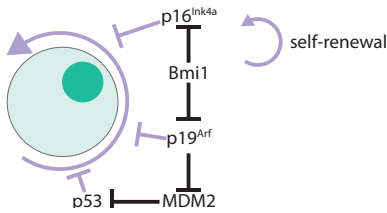


Figure 5. Schematic overview of cellular hierarchy. A stem cell is characterized by its ability to self-renew, creating a copy of itself. Stem cell division can also lead to the creating of progenitor cells and/or specialized cells. Progenitor cell division can lead to a more lineage-committed progenitor and/or to specialized cells. Unlike a stem cell, progenitors can only divide a limited number of times. Once a cell is specialized, the cell will not divide again.

the repression of differentiation. As such, it is required for the proper self-renewal of a variety of tissue stem cells that include the hematopoietic and the mammary epithelial stem cells (12-15). One critical BMI1 target is the *Cdkn2a* locus, encoding the genetic codes for *p16^{INK4A}* and *p14^{ARF}* in humans and *p16^{Ink4a}* and *p19^{Arf}* in mice. In response to stress signals, *p16^{Ink4a}* indirectly halts proliferation and can induce cellular senescence while *p19^{Arf}* inhibits Mdm2 resulting in the release of p53. Subsequently, p53 triggers cell-cycle arrest or apoptosis (16, 17). Bmi1 suppresses the transcription of the *Cdkn2a* genes, hereby enabling self-renewal (Fig. 6). In the mouse hemopoietic system, the simultaneous deletion of p53, *p16^{Ink4a}*, and *p19^{Arf}*, is sufficient for multipotent progenitors to also acquire the ability to self-renew (18).

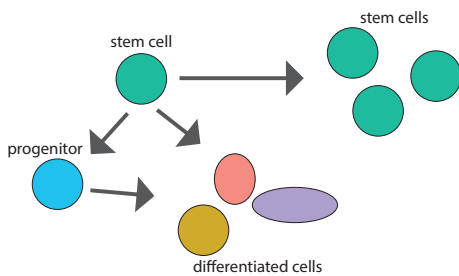


Figure 6. Schematic of the role of Bmi1 in stem cell self-renewal. Bmi1 enables stem cell self-renewal by epigenetically suppressing the transcription of *p16^{Ink4a}* and *p19^{Arf}*. *p19^{Arf}* inhibits MDM2, enabling the release of p53. *p16^{Ink4a}*, *p19^{Arf}*, and p53 all suppress events that enable stem cells to self-renew.

Tumor suppressor genes and oncogenes

Before discovering the role of *p16^{INK4A}/p16^{Ink4a}*, *p14^{ARF}/p19^{Arf}*, and *P53/p53* in the limiting of non-stem cell self-renewal, these genes were identified as tumor suppressors genes. Unaltered, tumor suppressors are responsible for events that limit oncogenesis, such as limiting proliferation and stimulating apoptosis. Unaltered, oncogenes have functions opposing those of tumor suppressors, such as stimulating growth and proliferation. Both are essential for normal homeostasis, and they are regulated by cell intrinsic and cell extrinsic signals. However, in cancer such signaling is dysregulated resulting in the stimulation of cancer-aiding processes such as resisting cell death, continuous proliferative signaling, and the evasion of growth suppressors (19). Genetic changes, which are either mutations or changes in copy number, can alter genes, such as tumor suppressor genes and oncogenes, in such a way that they lose their sensitivity to respond to normal signals and in turn start supporting oncogenesis.

The RAS/RAF/MEK/ERK pathway

The RAS/RAF/MEK/ERK pathway is frequently involved with pro-tumor signaling. The activation of the cascade starts with the activation of a receptor tyrosine kinase (RTK). This leads to the recruitment of guanine nucleotide exchange-factors (GEFs)

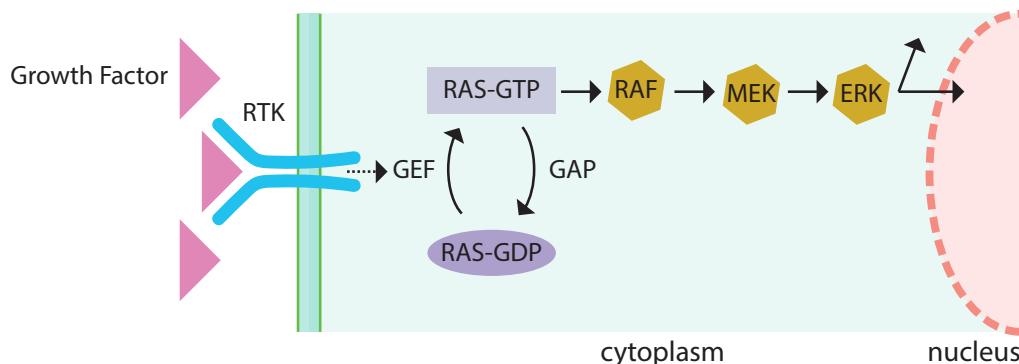


Figure 7. Schematic showing the RAS signaling pathway. Upon growth factor binding to an RTK, RAS is activated by the exchange of GDP to GTP by a GEF. This leads to the activation of RAS. RAS activates the domino kinase cascade of RAF, MEK, and ERK. ERK enables the phosphorylation of substrates both in the cytosol and in the nucleus.

that catalyze the exchange from a GDP to a GTP in small guanosine triphosphatase (GTPase) RAS, thereby activating RAS. The inactivation reverse is mediated by GTPase activating proteins (GAPs), switching active GTP-bound to inactive GDP-bound RAS (20). The regulation of RAS by GEFs and GAPs happens at the plasma membrane, where GEFs and GAPs are recruited to modulate RAS activity. RAS activity regulation is finetuned by the existence of various GEFs and GAPs (21). Once activated, RAS activates a mitogen-activated protein kinase (MAPK) cascade consisting of the serine/threonine kinases RAF, MEK, and ERK. The first step is the activation of RAF by RAS. RAF in turn catalyzes the phosphorylation of MEK, whereafter MEK catalyzes the phosphorylation of ERK. Active ERK catalyzes the phosphorylation of a wide variety of substrates - both in the cytosol and the nucleus. Ultimately, the signaling cascade leads to the transcription of genes that are predominantly involved with survival and proliferation (Fig. 7) (22). As such, this pathway is often overactivated in cancer and many of its components have been identified as oncogenes. Therefore, dysregulated RAS/RAF/MEK/ERK signaling is targeted by multiple drugs in various stages of development, including drugs already in use in the clinic.

The oncogene RAS

RAS is the mostly commonly mutated oncogene, with gain-of-function missense mutations detected in 25% of all human cancers. Therefore, medicinally inactivating RAS has been a topic of research for multiple decades. There are 4 RAS proteins: HRAS, NRAS, KRAS4A, and KRAS4B. KRAS4A and KRAS4B are two KRAS isoforms deriving from alternative RNA splicing, in most tissues KRAS4B is the dominant isoform. In cancer, the three RAS mutation hotspots are located at G12, G13, and Q61. Among the RAS proteins, KRAS is most frequently mutated in cancer accounting for 85% of mutated RAS. Mutations in KRAS are associated with poor clinical outcome. In KRAS, G12 is the most mutated site and the G12D substitution is most common, followed by G12V, and G12C with variability among different cancers (23). With the ability to personalize patient treatment, specifically targeting mutated KRAS variants has recently been shown to work with the approval of a first drug that targets KRAS^{G12C} in patients with KRAS^{G12C}-mutated advanced non-small cell lung cancer. Approvals of other KRAS^{G12C}-targeting drugs are expected to follow

(24-26). The highest prevalence (more than 30% of cases) of *KRAS* mutations are found in pancreatic ductal adenocarcinoma, colorectal adenocarcinoma, and lung adenocarcinoma (23). In contrast, in breast cancer, mutations in *KRAS* are uncommon (27, 28). Nevertheless, *KRAS* is a breast cancer driver gene and mutations in *KRAS* are associated with poor clinical outcome in breast cancer (28, 29). Not only mutations influence a protein. Other genetic alterations, such as copy number changes, can also upregulate oncogene expression and increase the oncogene's signaling output. When combining all breast cancer subtypes, amplifications and activating mutations in *KRAS* are equally common in breast cancer (28). However, when breast cancer patients are stratified by molecular subtype, in 32% of patients with the aggressive subtype basal-like breast cancer *KRAS* amplifications are detected in addition to the amplification of other *RAS/RAF/MEK/ERK* pathway players (30).

REFERENCES

1. Fu NY, Nolan E, Lindeman GJ, Visvader JE. Stem Cells and the Differentiation Hierarchy in Mammary Gland Development. *Physiol Rev.* 2020;100(2):489-523.
2. McNally S, Stein T. Overview of mammary gland development: a comparison of mouse and human. *Mammary gland development.* 2017:1-17.
3. Parmar H, Cunha GR. Epithelial-stromal interactions in the mouse and human mammary gland in vivo. *Endocr Relat Cancer.* 2004;11(3):437-58.
4. Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, et al. Generation of a functional mammary gland from a single stem cell. *Nature.* 2006;439(7072):84-8.
5. Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, et al. Purification and unique properties of mammary epithelial stem cells. *Nature.* 2006;439(7079):993-7.
6. Shehata M, Teschendorff A, Sharp G, Novic N, Russell IA, Avril S, et al. Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res.* 2012;14(5):R134.
7. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell.* 2018;172(5):1091-107.e17.
8. Consortium TM, coordination O, coordination L, processing Oca, sequencing Lpa, analysis Cd, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature.* 2018;562(7727):367-72.
9. Consortium TM. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature.* 2020;583(7817):590-5.
10. Fuchs E, Blau HM. Tissue Stem Cells: Architects of Their Niches. *Cell Stem Cell.* 2020;27(4):532-56.
11. Clarke MF. Clinical and Therapeutic Implications of Cancer Stem Cells. *N Engl J Med.* 2019;380(23):2237-45.
12. Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature.* 2003;423(6937):255-60.
13. Park IK, Qian D, Kiel M, Becker MW, Pihalja M, Weissman IL, et al. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature.* 2003;423(6937):302-5.
14. Rizo A, Olthof S, Han L, Vellenga E, de Haan G, Schuringa JJ. Repression of BMI1 in normal and leukemic human CD34(+) cells impairs self-renewal and induces apoptosis. *Blood.* 2009;114(8):1498-505.
15. Pietersen AM, Evers B, Prasad AA, Tanger E, Cornelissen-Steijger P, Jonkers J, et al. Bmi1 regulates stem cells and proliferation and differentiation of committed cells in mammary epithelium. *Curr Biol.* 2008;18(14):1094-9.
16. Lowe SW, Sherr CJ. Tumor suppression by Ink4a-Arf: progress and puzzles. *Curr Opin Genet Dev.* 2003;13(1):77-83.
17. Salama R, Sadaie M, Hoare M, Narita M. Cellular senescence and its effector programs. *Genes Dev.* 2014;28(2):99-114.
18. Akala OO, Park IK, Qian D, Pihalja M, Becker MW, Clarke MF. Long-term haematopoietic reconstitution by Trp53^{-/-}p16Ink4a^{-/-}p19Arf^{-/-} multipotent progenitors. *Nature.* 2008;453(7192):228-32.
19. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144(5):646-74.
20. Simanshu DK, Nissley DV, McCormick F. RAS Proteins and Their Regulators in Human Disease. *Cell.* 2017;170(1):17-33.
21. Cherfils J, Zeghouf M. Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol Rev.* 2013;93(1):269-309.
22. Plotnikov A, Zehorai E, Procaccia S, Seger R. The MAPK cascades: signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochim Biophys Acta.* 2011;1813(9):1619-33.
23. Hobbs GA, Der CJ, Rossman KL. RAS isoforms and mutations in cancer at a glance. *J Cell Sci.* 2016;129(7):1287-92.

Chapter 1

24. Hallin J, Engstrom LD, Hargis L, Calinisan A, Aranda R, Briere DM, et al. The KRAS G12C Inhibitor MRTX849 Provides Insight toward Therapeutic Susceptibility of KRAS-Mutant Cancers in Mouse Models and Patients. *Cancer Discov.* 2020;10(1):54-71.
25. Hong DS, Fakih MG, Strickler JH, Desai J, Durm GA, Shapiro GI, et al. KRAS G12C Inhibition with Sotorasib in Advanced Solid Tumors. *N Engl J Med.* 2020;383(13):1207-17.
26. Salgia R, Pharaon R, Mambetsariev I, Nam A, Sattler M. The improbable targeted therapy: KRAS as an emerging target in non-small cell lung cancer (NSCLC). *Cell Rep Med.* 2021;2(1):100186.
27. Rochlitz CF, Scott GK, Dodson JM, Liu E, Dollbaum C, Smith HS, et al. Incidence of activating ras oncogene mutations associated with primary and metastatic human breast cancer. *Cancer Res.* 1989;49(2):357-60.
28. Galiè M. RAS as Supporting Actor in Breast Cancer. *Front Oncol.* 2019;9:1199.
29. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell.* 2018;173(2):371-85.e18.
30. Network CGA. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61-70.

Thesis outline

Actively self-renewing tissues need to maintain a delicate balance between providing sufficient cells to maintain the integrity of a tissue and preventing excessive proliferation resulting in tumorigenesis. Consequently, tumors inhibit signals that limit proliferation and stimulate signals that promote survival. A common mechanism that enables tumor growth are changes in tumor suppressor genes and oncogenes. In this dissertation, I aim to further uncover the relationship between mammary gland self-renewal, proliferation, and oncogenesis. Furthermore, we develop a mouse model to increase our understanding of drug resistance.

In **chapter 1** I introduce the concepts discussed in this thesis. Namely, I introduce the mammary gland, the concepts of self-renewal and proliferation, and the role of tumor suppressors and oncogenes in breast cancer precision oncology in breast cancer treatment. Furthermore, I show how a large single cell RNA sequencing project can provide us information on the different cell types present in the mammary gland. Lastly, I introduce the RAS/RAF/MEK/ERK pathway with an emphasis on KRAS, the most commonly mutated oncogene in cancer.

In **chapter 2** we show that depleting the mammary gland of tumor suppressors *Trp53*, *p16^{Ink4a}*, and *p19^{Arf}* amplifies TNF α -induced proliferation of mammary epithelial cells.

In **chapter 3** we describe a novel breast cancer relapse mouse model that mimics the appearance of chemoresistance to mutated Kras inhibition. A subset of *Trp53*^{-/-} mice develops breast tumors upon the activation of oncogenic *Kras*^{G12D}. Once *Kras*^{G12D} is inactivated, the tumors declines followed by remission and eventually relapses. We show that the relapsed tumor has transitioned from an epithelial to a mesenchymal phenotype and has active MAPK/ERK signaling, the pathway activated by Kras.

In **chapter 4** we report that the innate immune system receptor TLR2 also has a cell-intrinsic role in the mammary gland and the intestine. We show that components of the TLR2 pathway - including TLR2, the TLR2 adaptor protein MYD88, and the TLR2 co-receptor CD14 - stimulate the regeneration of both tissues. Furthermore, blocking the TLR2 pathway slows down tumor formation in breast cancer and in intestinal adenoma.

In **chapter 5** we share the work of creating Tabula Muris, a single-cell transcriptomic compendium of the tissues and organs of the *Mus musculus*, the mouse. In **chapter 6**, we continue this work by adding the single-cell transcriptomic data collected at different time points, aiming to provide a public resource of the aging mouse.

In **chapter 7** I discuss the findings presented in this thesis.

Chapter 2

Depletion of *Trp53*, *p16^{Ink4a}*, and *p19^{Arf}* does not promote self-renewal in the mammary gland but amplifies proliferation induced by TNF α

Linda J. van Weele¹ • Ferenc A. Scheeren^{1,3} • Shang Cai^{1,4}
• Angera H. Kuo¹ • Dalong Qian¹ • William H.D. Ho^{1,2} •
Michael F. Clarke¹

Adapted from: *Stem Cell Reports* volume 16, pages 228-236 (2021)

¹Institute for Stem Cell Biology and Regenerative Medicine, School of Medicine, Stanford University, Stanford, CA, USA. ²Department of Stem Cell Biotechnology, California State University Channel Islands, Camarillo, CA, USA. ³Current address: Department of Medical Oncology, Leiden University Medical Center, Leiden, The Netherlands. ⁴Current address: Westlake University, Shilongshan St #18, Xihu District, Hangzhou, Zhejiang Province, China.

The mammary epithelium undergoes several rounds of extensive proliferation during the female reproductive cycle. Its expansion is a tightly regulated process, fueled by the mammary stem cells and these cells' unique property of self-renewal. Sufficient new cells have to be produced to maintain the integrity of a tissue, but excessive proliferation resulting in tumorigenesis needs to be prevented. Three well-known tumor suppressors, p53, p16^{Ink4a}, and p19^{Arf}, have been connected to the limiting of stem cell self-renewal and proliferation. Here we investigate the role of these three genes in the regulation of self-renewal and proliferation of mammary epithelial cells. Using mammary epithelial specific mouse models, we demonstrate that p53, p16^{Ink4a}, and p19^{Arf} do not play a significant role in the limitation of normal mammary epithelium self-renewal and proliferation while in the presence of inflammatory cytokine TNF α , *Trp53*^{-/-} p16^{Ink4a}^{-/-} p19^{Arf}^{-/-} mammary basal cells exhibit amplified proliferation.

The epigenetic regulator BMI1 represses differentiation in various tissue-specific stem cells such as hematopoietic stem cells (1-3) and mammary stem cells (4). BMI1 targets the *Cdkn2a* locus which contains the reading frames for p16^{INK4a} and p14^{ARF} in humans and p16^{Ink4a} and p19^{Arf} in mice that play an essential role in the induction of senescence and cell cycle arrest. p16^{Ink4a} inhibits cell cycle progression from G1 to S phase. P14^{ARF}/p19^{Arf} inhibits MDM2 leading to activation of p53. P53, p16^{Ink4a}, and p19^{Arf} respond to stress signals and induce cell-cycle arrest and apoptosis. Their ability to halt cell division prevents tissues from uncontrolled proliferation that can lead to tumor growth (5, 6).

Within the hematopoietic system, simultaneously deleting p16^{Ink4a}, p19^{Arf}, and *Trp53* leads to a ten-fold increase in the number of cells able to self-renew and provides multipotent progenitors with self-renewing capacity, a property normally only attributed to stem cells (7). In the mammary gland, various groups have deleted *Trp53* or the *Cdkn2a* locus, but not both, and examined the effect on self-renewal (4, 8-10). To our knowledge, no studies have investigated the effect, especially on self-renewal, of deleting these three genes concurrently in the mammary gland.

Here, using our epithelial specific mouse model, we demonstrate that p53, p16^{Ink4a}, and p19^{Arf} do not play a significant role in the regulation of mammary epithelium self-renewal and proliferation in C57BL/6J mice. Using a primary transplantation assay to study short-term self-renewal, depletion of the three tumor suppressors does not affect the number of regenerating cells nor their proliferative capacity in normal physiologic conditions. Long-term self-renewal, evaluated using a secondary transplantation assay, also shows no effect on proliferation. Additionally, tumor suppressor deletion does not affect long-term self-renewal in mammary epithelial cells compared to the control; however, *Trp53*^{-/-} cells show a decrease in regenerating cells compared to *Trp53*^{-/-} p16^{Ink4a}^{-/-} p19^{Arf}^{-/-} cells. RNA expression analysis suggests TNF α -dependent signaling is upregulated in *Trp53*^{-/-} p16^{Ink4a}^{-/-} p19^{Arf}^{-/-} mammary epithelial basal cells. *In vitro*, we find that TNF α increases organoid formation and proliferation in wild type and *Trp53*^{-/-} p16^{Ink4a}^{-/-} p19^{Arf}^{-/-} basal MECs; however, compared to wild type MECs, *Trp53*^{-/-} p16^{Ink4a}^{-/-} p19^{Arf}^{-/-} basal MECs are sensitized to TNF α -induced proliferation.

RESULTS

Generation of $p16^{Ink4a}$ and $p19^{Arf}$ knockout (KO) and/or $Trp53$ KO mouse models

To investigate the effect of p53, $p16^{Ink4a}$, and $p19^{Arf}$ on self-renewal and proliferation, we generated multiple loss-of-function mouse models. Female mice with homozygous germline combined deletions of *Cdkn2a* and *Trp53* are inviable. Therefore, we crossed the *Cdkn2a* null mouse (11) with a conditional *Trp53^{fllox}* (*Trp53^f*) mouse (12). Loss of *Trp53* was induced by the *Krt14-Cre* transgene (13), *Krt14* is expressed in all MECs during embryonic development (14, 15). To visually mark cells that lost *Trp53*, we introduced the *mT/mG* transgene (16). Cre excises *mT/mG*'s *tdTomato* allele and activates the expression of *GFP* simultaneously indicating the excision of *Trp53*. We used flow cytometry to confirm the efficiency of specificity of *Krt14-Cre*. In accordance with a previous publication (17), flow cytometry showed that 99.8% of all GFP⁺ cells are in the luminal and basal populations (Fig. S1A and S1B). Subsequently, real-time qPCR confirmed the absence of *Trp53* in basal and luminal GFP⁺ populations (Fig. S1C). The various crosses lead to the four following mouse models which are used in these studies: 1) WT; *Krt14-Cre mT/mG*, 2) *Trp53* KO; *Krt14-Cre mT/mG Trp53^{flf}*, 3) *Cdkn2a* KO; *Krt14-Cre mT/mG p16^{Ink4a}-/- p19^{Arf}-/-*, 4) *Trp53* & *Cdkn2a* KO; *Krt14-Cre mT/mG Trp53^{flf} p16^{Ink4a}-/- p19^{Arf}-/-* (Fig. 1A).

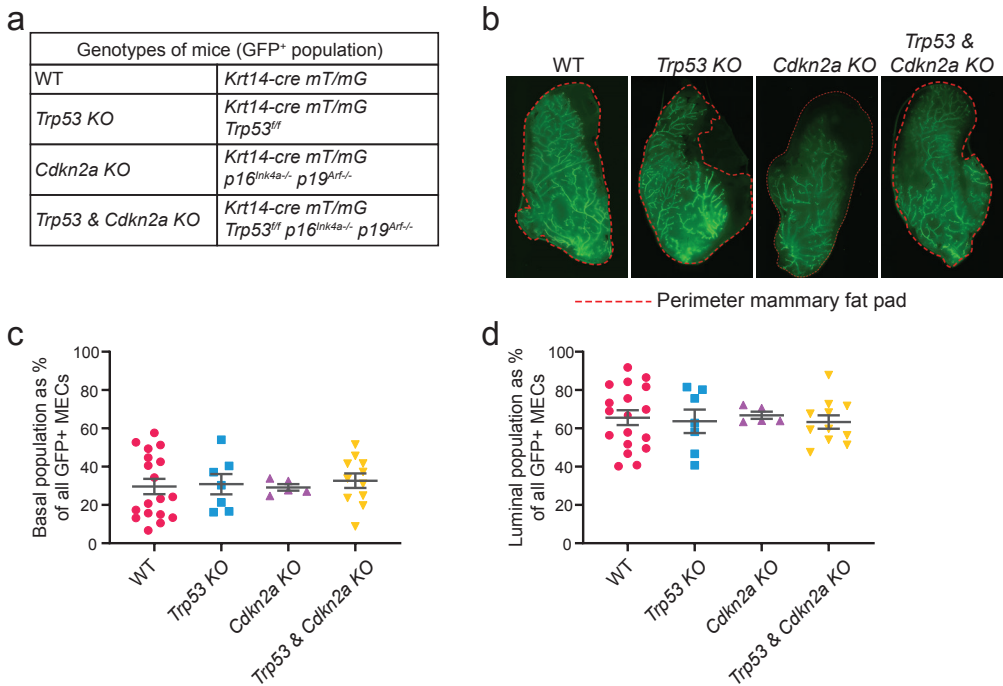


Figure 1. MECs develop similarly in WT, *Trp53* KO, *Cdkn2a* KO, and *Trp53* & *Cdkn2a* KO mice (a) Tabular overview of the abbreviations used for the four different transgenic mice described in this study. In Figure S1, the gating strategy used for FACS through this study (Fig. S1A) and an analysis of the efficiency of the Cre-lox system (Fig. S1B, S1C) are displayed. (b) Representative fluorescent images of WT, *Trp53* KO, *Cdkn2a* KO, and *Trp53* & *Cdkn2a* KO mammary glands. Mammary epithelial cells are marked by GFP. Fat pads were freshly dissected from mice and photographed. (c, d) Flow cytometry analysis shows the percentage of WT, *Trp53* KO, *Cdkn2a* KO, and *Trp53* & *Cdkn2a* KO basal (c) and luminal (d) cells. Data are represented as mean ± SEM, each individual symbol represents data derived from one mouse. (P > 0.05 for all comparisons; one-way ANOVA)

Endogenous MECs develop similarly in all mutant mouse models

The mammary glands of female adult WT, *Trp53* KO, *Cdkn2a* KO, and *Trp53* & *Cdkn2a* KO mice showed no differences in branch development (Fig. 1B). Moreover, the percentage of cells contributing to the basal and luminal populations was equal as determined by flow cytometry (Fig. 1C and 1D). These data show depleting MECs of *p16^{Ink4a}* and *p19^{Arf}* and/or *Trp53* neither results in differences in the phenotypic duct formation nor affects the ratio of basal to luminal cells.

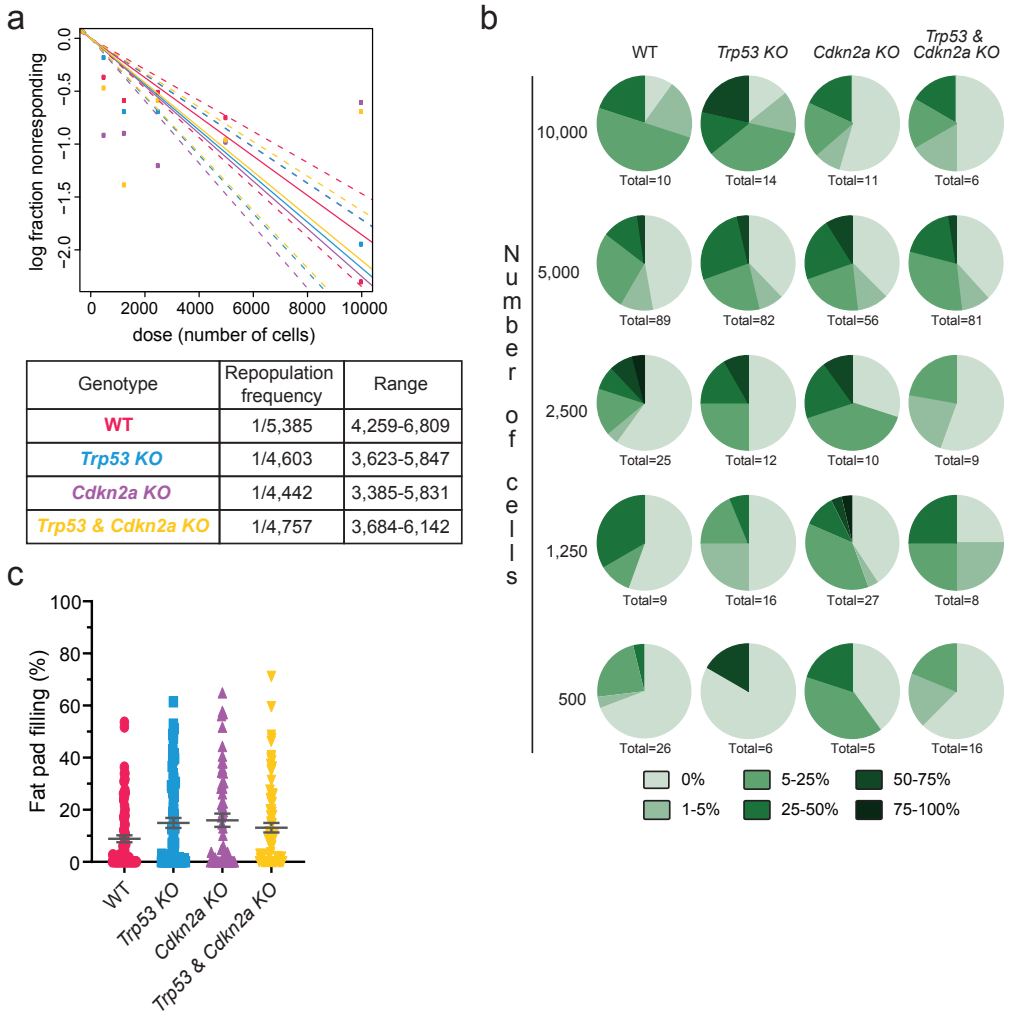


Figure 2. WT, *Trp53* KO, *Cdkn2a* KO, and *Trp53* & *Cdkn2a* KO MECs display equal short-term repopulation potential An overview of this experiment is shown in Figure S2. (a) ELDA results of the limiting dilution transplantation assay. A minimum of 11 donor mice were used for each genotype and data derive from a minimum of 4 separate experiments for each genotype. ($P > 0.05$ for all comparisons)(b, c) Overview of the data displayed in (a). The number of cells injected, the total number of transplantations performed, and the percentage donor outgrowths occupied in recipient fat pads are shown. In (b) data is grouped in pie charts. In (c) the size of the individual outgrowths after 5,000 cells were injected is shown. Data are represented as mean \pm SEM, each individual symbol represents one transplant. ($P > 0.05$ for all comparisons; one-way ANOVA) Similar plots as displayed in (c) for other injected cell numbers are shown in Figure S3.

MECs in all mutant mouse models display equal short-term regenerative and proliferative efficiency in primary transplantation assay

To determine the frequency of cells able to reconstitute the mammary gland, we performed a mammary gland limiting dilution transplantation assay in syngeneic mice (Fig. S2). The repopulation frequency is comparable between MECs derived from WT (1/5,385), *Trp53* KO (1/4,603), *Cdkn2a* KO (1/4,442), and the *Trp53* & *Cdkn2a* KO (1/4,757) glands (Fig. 2A, 2B). Moreover, analysis of the size of the regenerated donor glands also showed no differences between the different mouse models (Fig. 2C, Fig. S3A-S3D). This demonstrates that p53, p16^{Ink4a}, and p19^{Arf} do not affect the ability of MECs to regenerate or proliferate in a primary transplantation assay.

The absence of *Trp53* reduces secondary transplant repopulation frequency compared to *Trp53* & *Cdkn2a* KO MECs

To investigate if the deletion of *Trp53* and/or the *Cdkn2a* locus leads to changes in the ability to self-renew and/or proliferate long-term, secondary transplantations were performed (Fig. S2). WT, *Cdkn2a* KO, and *Trp53* & *Cdkn2a* KO MECs showed a similar frequency of long-term self-renewing cells, as did WT, *Trp53* KO, and *Cdkn2a* KO MECs. Meanwhile, *Trp53* & *Cdkn2a* KO MECs showed a significantly higher regeneration frequency than *Trp53* KO MECs (Fig. 3A); therefore, the ablation of

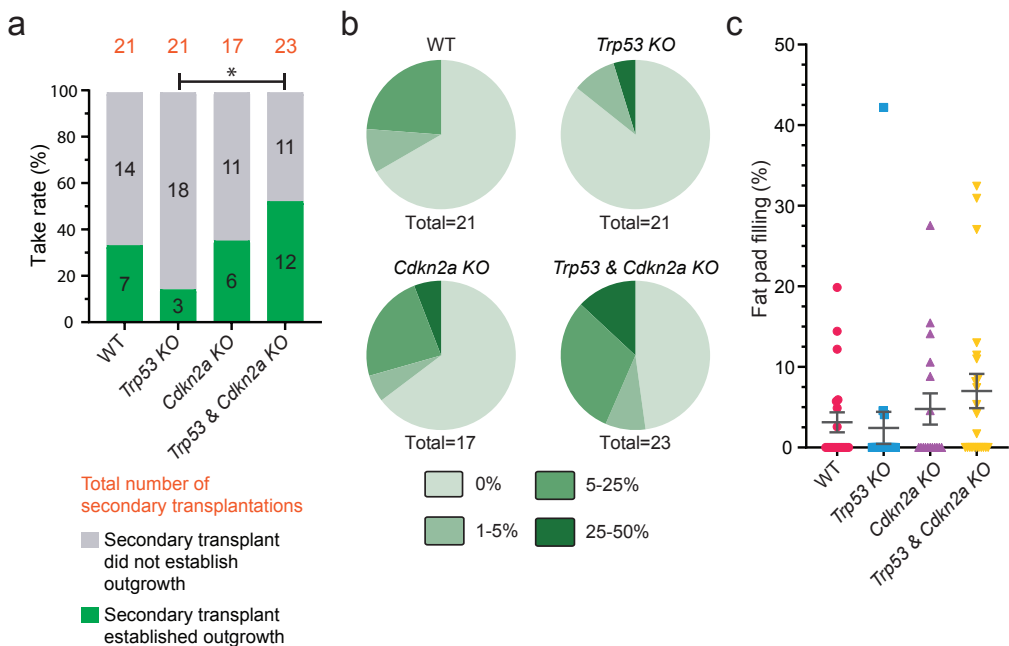


Figure 3. Long-term renewal is reduced in the absence of *Trp53* and rescued by the depletion of p16^{Ink4a} and p19^{Arf} An overview of this experiment is shown in Figure S2. (a) The take rate of the secondary transplantation assay. Total number of secondary transplants shown in orange, positive outgrowths are shown in green, no outgrowths are shown in grey. (* P < 0.05; Fisher's exact test) (b, c) Overview of the data displayed in (a). The total number of transplantations and the percentage donor outgrowths occupied in recipient fat pads are shown. In (b) data is grouped in pie charts. In (c) the size of individual outgrowths is shown. Data are represented as mean ± SEM, each individual symbol represents one transplant. (P > 0.05 for all comparisons; one-way ANOVA)

Chapter 2

Cdkn2a appears to rescue long-term self-renewal capacity in the *Trp53* & *Cdkn2a* KO MECs. The deletion of any of the tumor suppressor genes did not affect the rate of proliferation (Fig. 3B and 3C).

***Trp53* & *Cdkn2a* KO luminal progenitor MECs do not regenerate when transplanted**

In the hematopoietic system, *Trp53* & *Cdkn2a* KO multipotent progenitors acquire the capacity to self-renew and regenerate the blood system in transplantation experiments (7). To investigate if in the mammary gland *Trp53* & *Cdkn2a* KO progenitors also possess such capabilities, we used the fluorescence-activated cell sorter (FACS) to select for progenitors and transplant them in syngeneic mice. Prior to this experiment, we corroborated published data showing that the luminal cells encompass a progenitor cell population that expresses the cell surface marker CD14 (18). *In vitro*, luminal CD14⁺ cells form more organoids than CD14⁻ luminal cells, confirming CD14 enriches strongly for luminal progenitors (Fig. S4A). Subsequently, we performed a limiting dilution assay with WT and *Trp53* & *Cdkn2a* KO luminal CD14⁺ progenitors, none of these transplantations resulted in the regeneration of the mammary epithelium (Fig. S4B). Thus, p53, p16^{Ink4a}, and p19^{Arf} do not or are not sufficient to inhibit luminal CD14⁺ progenitors from self-renewing in transplantation assays.

***Trp53* & *Cdkn2a* KO basal MECs are sensitized to TNF α -induced proliferation**

Lastly, we compared gene expression between WT and *Trp53* & *Cdkn2a* KO mice. Considering luminal progenitors do not regenerate *in vitro* (Fig. S4B), we focused on basal MECs. WT and *Trp53* & *Cdkn2a* KO basal MECs were sorted using FACS and their respective RNA was extracted and submitted for RNA microarray analysis (Fig. 4A). Gene Set Enrichment Analysis (GSEA) revealed that TNF α signaling via NF- κ B was enhanced in *Trp53* & *Cdkn2a* KO basal MECs (Fig. 4B, Table S1). TNF α is a cytokine released by immune cells, predominantly by macrophages. Upon binding to a receiving cell, it can stimulate cell survival or cell death depending on other actors in the cell. When the TNF α signaling cascade activates the transcription factor complex NF- κ B, NF- κ B stimulates cellular survival by increasing proliferation and inflammation (19). To investigate if TNF α affects the growth of basal MECs, we exposed WT and *Trp53* & *Cdkn2a* KO basal MECs to different concentrations of TNF α in an organoid formation assay. Exposure to TNF α led to an increase in organoid number and size, regardless of genotype (Fig. 4C-4E). The number of organoids formed upon TNF α exposure was similar between WT and *Trp53* & *Cdkn2a* KO basal MECs and did not change as the concentration of TNF α increased (Fig. 4D). Notably, *Trp53* & *Cdkn2a* KO basal MEC organoids grew significantly larger than their WT counterpart when exposed to 100 ng/ml TNF α (Fig. 4E). Thus, while TNF α stimulates the formation and proliferation of both WT and *Trp53* & *Cdkn2a* KO basal MEC organoids equally, organoids derived from *Trp53* & *Cdkn2a* KO basal MECs were larger in size.

DISCUSSION

In this report, we set out to understand the molecular mechanisms that regulate self-renewal and proliferation in the epithelium of the mammary gland. We focused on the role of p53, p16^{Ink4a}, and p19^{Arf} as they have been reported to play an important role in the suppression of hematopoietic self-renewal and proliferation (7). Unlike the hematopoietic system, we find that p53, p16^{Ink4a}, and p19^{Arf} do not regulate short-

Tumor suppressors affect mammary gland proliferation

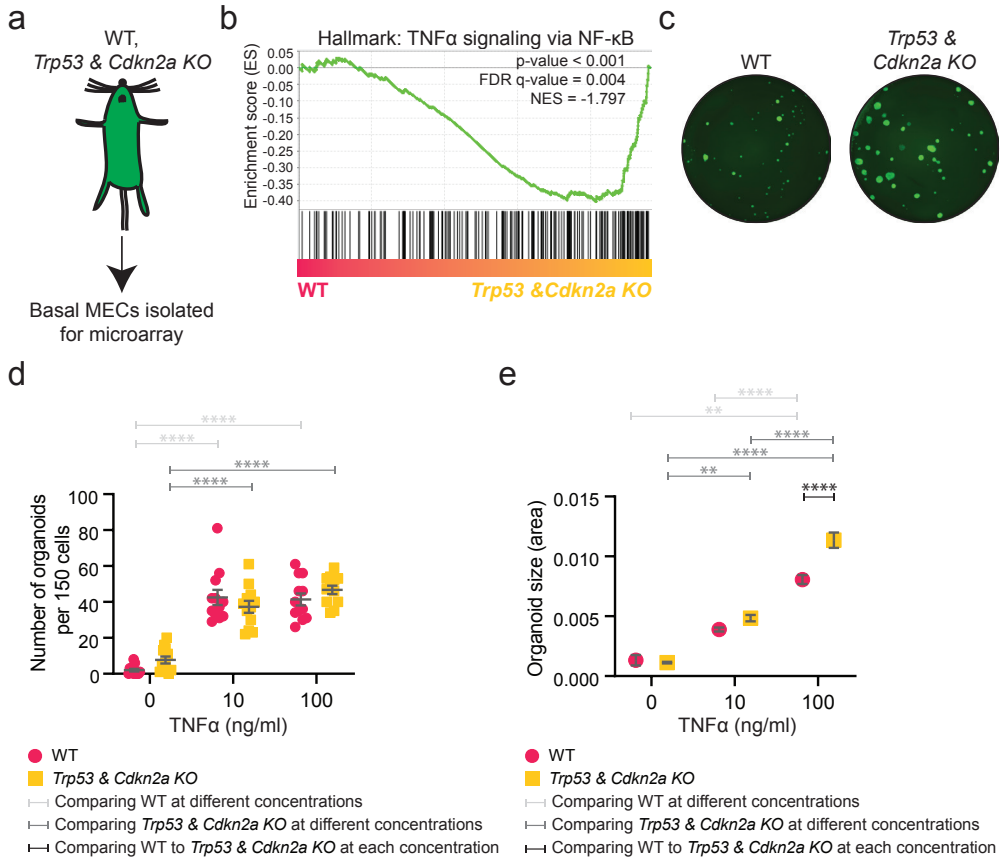


Figure 4. *Trp53* & *Cdkn2a* KO basal MECs are sensitized to TNF α -induced proliferation (a) An overview of WT and *Trp53* & *Cdkn2a* KO basal MECs that were sorted using FACS to isolate RNA for RNA microarray expression analysis. (b) GSEA results of the Hallmark gene set TNF α via NF- κ B of WT and *Trp53* & *Cdkn2a* KO basal MECs, 3 biological replicates of each genotype were analyzed. All significant GSEA results are shown in Table S1. (c) Fluorescent image of an *in vitro* organoid formation assay showing 150 WT or *Trp53* & *Cdkn2a* KO basal cells treated with TNF α . Photos shown are examples of wells treated with 100 ng/ml TNF α for 2 weeks. (d, e) Plots show the number (d) and size (e) of organoids that had formed after 150 WT and *Trp53* & *Cdkn2a* KO basal MECs were cultured for two weeks in the presence of 0, 10, or 100 ng/ml TNF α . Data from 4 biological replicates and 3 technical replicates are displayed. In (d), the number of organoids counted in each well is displayed (n=12 wells). In (e), the symbol shows the mean organoid size. Data are represented as mean \pm SEM. Two-way ANOVA compares WT basal MECs in the presence of different concentrations of TNF α in light grey, *Trp53* & *Cdkn2a* KO basal MECs at different concentrations of TNF α in dark grey, and WT to *Trp53* & *Cdkn2a* KO basal MECs at each concentration of TNF α in black. (* P < 0.05; ** P < 0.01; **** P < 0.0001; two-way ANOVA)

term self-renewal or proliferation in C57BL/6J mice. Interestingly, we found that *Trp53* KO MECs in these mice show a lower long-term regeneration ability than *Trp53* & *Cdkn2a* KO MECs. This is in line with the report that loss of *p16^{Ink4a}*, rather than *Trp53*, contributes to the self-renewal of quiescent mammary stem cells (20). On the other hand, long-term self-renewal is similar between *Trp53* KO, WT, and *Cdkn2a* KO MECs. Furthermore, ablation of any of these tumor suppressor genes does not affect proliferation in secondary transplantations. Therefore, we conclude that overall, p53, p16^{Ink4a}, and p19^{Arf}, do not play a significant role in mammary gland self-renewal or

Chapter 2

are not sufficient to regulate mammary gland self-renewal.

Previously, other groups have examined the effects of these three genes on mammary epithelium self-renewal using transplantation assays. In accordance with what we report, Pietersen et al. found that WT and *Cdkn2a* KO MECs have the same number of self-renewing cells (4). The three groups that studied p53 and mammary gland self-renewal, found an increase in self-renewing cells in the absence of *Trp53*. Two groups used different mouse strains in their studies (8, 10). It has been widely reported that the use of different mouse strains can affect results (21). A third group also reported on the increase of self-renewal in *Trp53*^{-/-} MECs, using a C57BL/6J *Trp53* null mouse (9), opposed to our conditional C57BL/6J *Trp53* KO model, where *Trp53* is only absent in a subset of cells. Moreover, after digestion Cicalese et al. transplanted all remaining cells of the mammary fat pad, including *Trp53* KO stromal cells, whereas we used FACS to select for MECs. In the case of cancer, stromal cells can influence tumor growth through p53 expression (22). This raises the possibility that the loss of p53 in stromal cells influences mammary stem cell frequency. It would be interesting to investigate the repopulation frequency of *Trp53* null and conditional *Trp53* KO glands in a wild type and *Trp53* null environment to answer this question. To corroborate the reliability of our data: we used FACS to select for GFP⁺ MECs, whereby once transplanted, the color marker facilitated the identification of transplanted donor cells among the colorless cells in recipient mice.

Within our mouse model we make use of the *Krt14-Cre* transgenic mouse. It is important to note that normal mammary development has been reported in this transgenic strain (17). However, *Cre* expression can result in unintentional effects. For example, using a different *Krt14-Cre* inducible mouse, tetraploid keratinocytes form in the skin (23). To limit potential misinterpretation of our data, all four mouse models in this study expressed *Krt14-Cre* and at least one target loxP site. Moreover, our WT (1/5,385) limiting dilution data show a similar frequency to C57BL/6J wild type cells (1/9,045) (24), noting that due to the absence of a mammary epithelial marker, Scheeren et al. included stromal cells resulting in a decreased frequency compared to our WT mice.

Unlike the hematopoietic multipotent progenitor, we find that mammary luminal progenitors do not acquire the ability to self-renew. It remains possible we did not transplant enough cells to find a regenerating *Trp53* & *Cdkn2a* KO CD14⁺ luminal progenitor, but if one does arise in the mutant progenitor it is very rare. Another possibility is the existence of a different mammary progenitor cell type, not marked by luminal CD14⁺, which is sensitive to the ablation of *Trp53*, *p16^{Ink4a}*, and *p19^{Arf}*. If this is true, unlike in the hematopoietic system, such a mutated cell does not affect self-renewal significantly, since the repopulation frequency is not affected in primary transplants.

Comparing WT and *Trp53* & *Cdkn2a* KO basal MEC transcriptional data, we found an enhancement of TNF α signaling via the NF- κ B pathway in the *Trp53* & *Cdkn2a* KO basal MECs. We validated these results by exposing WT and *Trp53* & *Cdkn2a* KO basal MECs to TNF α *in vitro*. TNF α exposure strongly increased the number of basal MECs growing out into organoids and their ability to proliferate. While TNF α equally

Tumor suppressors affect mammary gland proliferation

affected organoid number, it resulted in enhanced proliferation in *Trp53* & *Cdkn2a* KO basal MECs. In the rat mammary gland, TNF α induces cellular proliferation as well (25). In mice, TNF α exposure activates NF- κ B, enabling the epithelial cells of the liver, the hepatocytes, to be extensively passaged *in vitro* and even subsequently engrafted *in vivo* (26). Molecularly, p16^{Ink4a}, p19^{Arf}, and p53 can interact with NF- κ B. p16^{Ink4a} and p19^{Arf} can suppress the transcriptional activity of the NF- κ B complex (27). The literature on p53 and NF- κ B is more extensive, describing how p53 can modulate NF- κ B (28). In contrast, mutant p53 can disrupt modulation, increasing and strengthening NF- κ B activity to promote proliferation and tumorigenesis (29). NF- κ B activation is involved with various cancer hallmarks, but most notably with inflammation, often induced by TNF α (30). TNF α is highly expressed in macrophages and macrophages are part of the stromal niche of the mammary gland. Eliminating macrophages impairs mammary gland development and mammary stem cell self-renewal (31). Our data indicate that losing tumor suppressors p53, p16^{Ink4a}, and p19^{Arf} makes mammary cells more prone to excessive proliferation in the presence of inflammation. Therefore, our data underscore the importance of controlling inflammation to slow down the mutation-prone overproliferation and the subsequent oncogenic transformation when mammary stem cells lose their genome-guard p53 and senescence-triggers p16^{Ink4a} and p19^{Arf}.

EXPERIMENTAL PROCEDURES

Animal care and use

Mice purchased from The Jackson Laboratory: *Trp53*^{fl/fl} (Stock #008462), *Krt-14-Cre* (Stock #004782), *mT/mG* (Stock #007676), weaning-aged transplantation recipients C57BL/6J (#000664). From the NCI Mouse Repository: *Cdkn2a* null (Stock #01XB1). All mice were females and backcrossed into the C57/BL6J background for at least six generations. All mice used for this study were maintained at the Stanford Veterinary Service Center in accordance with the guidelines of the Administrative Panel on Laboratory Animal Care (APLAC #10868).

Tissue processing and FACS

8-14-week-old virgin female mice were euthanized and fat pads surgically resected. Tissue was mechanically dissociated. Tissue used for transplantation assays was digested in DMEM/F12 (catalog #12634028, Gibco) with collagenase and hyaluronidase (catalog #07912, STEMCELL Technologies Inc.) and DNase I (Catalog #LS002139, Worthington) for 2 hr with gentle pipetting every 30 min. Tissue used for the TNF α organoid growth assay was digested in DMEM/F12 with gentle collagenase and hyaluronidase (catalog #07919, STEMCELL Technologies Inc.) overnight. The remainder of the protocol was the same for all experiments and has been previously described (35) with minor revisions. Briefly, lysis of red blood cells in ACK (catalog #BW10548E Lonza) for 5 min, was followed by 5 min treatment with pre-warmed 0.25% Trypsin-EDTA (catalog #25-510, GenClone), followed by treatment with pre-warmed dispase (catalog #07913, STEMCELL Technologies Inc.) and DNase I (catalog #LS002139, Worthington) for 2 min. Ultimately, the cells were filtered through a 40 μ m mesh and washed and resuspended in HBSS+2%FBS+PSA. For FACS analysis and sorting, mammary single cells were stained with CD45 (clone #30-F11, Biolegend), CD31 (clone #390, Biolegend), TER-119 (clone #TER119, Biolegend), CD49f

Chapter 2

(clone #GoH3, Biolegend), CD24 (clone #M1/69, Biolegend), CD14 (clone #Sa14-2, Biolegend). To enrich for epithelial cells preceding FACS, most of lineage was depleted on the autoMACS Pro Separator (catalog #130-092-545, Miltenyi Biotec) using microbeads for CD45 (catalog #130052301, Miltenyi Biotec), CD31 (catalog #130097418, Miltenyi Biotec), and TER119 (catalog #130049901, Miltenyi Biotec) according to manufacturer's instructions. Flow cytometry was performed with a 100 μ m nozzle on a BD Flow cytometry Aria II with Flow cytometry Diva software (BD Biosciences). Data analysis was done using FlowJo V10 (FlowJo). For all experiments, debris and cell doublets were excluded using side scatter and forward scatter profiles (area and width). Dead cells were eliminated using DAPI (catalog #32670, Sigma).

Transplantations

Using FACS, lineage (CD45/CD31/TER119) negative, GFP+ cells were sorted in HBSS+2%FBS+PSA media and counted using a hemocytometer. Cells were resuspended with 50% Matrigel (catalog #356234, Corning). If sorted for specific populations, it is marked in the main text. For each transplant, 10 μ l was injected into cleared fat pads of weaning-age mice (21-25 days). Transplantation procedure has been previously described (20). After 7- 20 weeks, transplants were dissected and photographed using a fluorescent microscope. The repopulation frequency was calculated using ELDA (36). As for secondary transplantations, first 5,000 cells were transplanted in a primary transplantation assay. Primary transplants that took up more than 10% of the recipient fat pad, were individually digested, and used for secondary transplantations. After 7 weeks the secondary transplants were dissected and photographed using a fluorescent microscope.

Mammary organoid growth assay

For the organoid assays, 40 μ l growth factor reduced Matrigel (catalog #356230, Corning) was plated in a 96 well plate and solidified at 37 °C for 10 min. For the luminal CD14 assay, growth factor reduced Matrigel was mixed prior to plating with 10,000 L1-Wnt3a feeder cells (generous gift from Dr. Roel Nusse) that were administered a 40Gy dose of X-ray irradiation. MECs were plated on top in 150 μ l culture media. Culture media for the luminal CD14 organoid assay has previously been described (20). In short, the culture media used contained DMEM/F12 (catalog #12634028, Gibco) + 2% FBS + PSA + B27 (Catalog #12587010, Gibco) + 10 mM HEPES (catalog #15630080, Gibco) + 10 ng/ml EGF (Catalog #315-09, PeproTech) + 10 μ M ROCK inhibitor Y27632 (catalog #Y0503, Sigma) + Noggin (catalog #6057-NG, R&D Systems) + N2 (catalog #17502001, Gibco); culture media was replaced twice-a-week. For the TNF α organoid growth assay, culture media was made with DMEM/F12 + 2% FBS + PSA + B27 + 10 mM HEPES + Glutamax (catalog #35050061, Gibco) + N2; culture media was enriched with Recombinant Murine TNF α (catalog #315-01A, PeproTech), TNF α was reconstituted according to manufacturer's instructions and was freshly added to the organoids four times a week. Culture media was replaced twice-a-week. All plates were maintained in a 37 °C incubator at 5% CO₂ for two weeks.

Real-time PCR

5,000 cells of various MEC populations were directly sorted into RNeasy Protect (catalog #76526, Qiagen). RNA was extracted using the RNeasy micro kit (catalog #74004, Qiagen), according to manufacturer's instructions. Reverse transcription of RNA to

Tumor suppressors affect mammary gland proliferation

cDNA was done using SuperScript III First Strand Synthesis kit (catalog #18080051, Invitrogen) according to the manufacturer's instructions. cDNA was preamplified 20 cycles using SybrGreen master mix (catalog #4364346, Applied Biosystems) and Sybergreen primers for Actb and Trp53, sequences listed below. Preamplified cDNA was used for real-time PCR, with Sybergreen primers for Actb and Trp53, sequences listed below. All data was normalized to Actb. Using the 7900HT Real Time PCR system (catalog #4329001, Applied Biosystems). Data was analyzed by SDS2.4 software, Excel and Prism 8. Sybrgreen primers: Actb: Forward GGCTGTATTCCCCTCCATCG Reverse CCAGTTGGTAACAATGCCATGT; Trp35: Forward CTCTCCCCCGCAAAGAAAA Reverse CGGAACATCTCGAAGCGTTA.

RNA Microarray expression analysis

15,000-60,000 cells of basal MECs were directly sorted into RNAprotect (catalog #76526, Qiagen) using FACS. RNA was extracted using the RNeasy micro kit (catalog #74004, Qiagen), according to manufacturer's instructions. Library preparation, hybridization and scanning were all performed by Stanford protein and nucleic acid facility (PAN facility) using the Mouse Gene 2.0 ST Array (Affymetrix). Data was analyzed using Transcriptome Analysis Console Software (ThermoFisher Scientific). Affymetrix gene IDs were converted to Official Gene Symbols using DAVID's Gene ID Conversion Tool v6.8 (37). Using MS Excel, data lacking an associated gene name were deleted from the resulting list. Patterns in gene expression were analyzed using GSEA v4.0.1 (32, 33) and the hallmark gene sets (34). Data was deposited in GEO.

Quantification and statistical analysis

To determine if there were differences among cells from different genotypes or variations in responses to treatments, one-way or two-way ANOVA was calculated using GraphPad Prism 8. In case of a binary outcome, the Fisher's exact test was calculated using GraphPad Prism 8. Statistical analysis was performed only done if data from at least 3 biological replicates was present. Exact number of biological and technical replicates as well as statistical parameters are given in the figures and corresponding figure legends. For analysis of limiting dilution data, ELDA software was used (36). The expected frequencies are shown, as well as the 95% confidence interval (the lower and upper values are shown). To count organoid number and size, images taken from each experiment were processed simultaneously. Fluorescent images were first opened in Photoshop CS6, threshold function was used to remove background noise and to only show organoids to allow for automated counting and size analysis. Subsequently, using ImageJ v1.51, Analyze Particles was used to count and measure the size of the organoids. No statistical method was used to predetermine the size of the sample, experiments were not randomized or blinded. Researchers were blinded when assessing outcomes of take rate and outgrowth size of donor MECs, the mouse ear tag was used for identification.

Data and code availability

Microarray expression data (GEO: GSE137573) are available in the Gene Expression Omnibus.

AUTHOR CONTRIBUTIONS

Conceptualization, L.J.v.W., F.A.S., and M.F.C.; Methodology, L.J.v.W., F.A.S., S.C.,

Chapter 2

and M.F.C.; Investigation, L.J.v.W., D.Q., W.H.D.H.; Formal Analysis, L.J.v.W.; Writing - Original Draft, L.J.v.W.; Writing - Review & Editing, all authors; Funding Acquisition, M.F.C.

ACKNOWLEDGEMENTS

This work was supported by Breast Cancer Research Foundation grant (BCRF-18-027), Ludwig Cancer Research Foundation, and by the CIRM Bridges award (to W.H.D.H.). Some research was performed on a FACSriaal that was purchased using NIH S10 shared instrumentation grant (1S10RR02933801). We thank Amy Doan for help with mouse maintenance and the Stanford Veterinary Service Center for providing mouse care; Dr. Lauren E. Grosberg for help with statistical analysis; Dr. Roel Nusse for providing L1-Wnt3a feeder cells; Patty Lovelace, Jennifer Ho, Stephen Weber, Catherine Carswell Crumpton, and Cheng Pan for management of the flow cytometry facility; Natalia Kosovilka and the Stanford PAN facility for microarray services.

REFERENCES

1. Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature*. 2003;423(6937):255-60.
2. Park IK, Qian D, Kiel M, Becker MW, Pihalja M, Weissman IL, et al. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature*. 2003;423(6937):302-5.
3. Rizo A, Olthof S, Han L, Vellenga E, de Haan G, Schuringa JJ. Repression of BMI1 in normal and leukemic human CD34(+) cells impairs self-renewal and induces apoptosis. *Blood*. 2009;114(8):1498-505.
4. Pietersen AM, Evers B, Prasad AA, Tanger E, Cornelissen-Steijger P, Jonkers J, et al. Bmi1 regulates stem cells and proliferation and differentiation of committed cells in mammary epithelium. *Curr Biol*. 2008;18(14):1094-9.
5. Charni M, Aloni-Grinstein R, Molchadsky A, Rotter V. p53 on the crossroad between regeneration and cancer. *Cell Death Differ*. 2017;24(1):8-14.
6. Lowe SW, Sherr CJ. Tumor suppression by Ink4a-Arf: progress and puzzles. *Curr Opin Genet Dev*. 2003;13(1):77-83.
7. Akala OO, Park IK, Qian D, Pihalja M, Becker MW, Clarke MF. Long-term haematopoietic reconstitution by Trp53^{-/-}p16Ink4a^{-/-}p19Arf^{-/-} multipotent progenitors. *Nature*. 2008;453(7192):228-32.
8. Chiche A, Moumen M, Petit V, Jonkers J, Medina D, Deugnier MA, et al. Somatic loss of p53 leads to stem/progenitor cell amplification in both mammary epithelial compartments, basal and luminal. *Stem Cells*. 2013;31(9):1857-67.
9. Cicalese A, Bonizzi G, Pasi CE, Faretta M, Ronzoni S, Giulini B, et al. The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells. *Cell*. 2009;138(6):1083-95.
10. Tao L, Roberts AL, Dunphy KA, Bigelow C, Yan H, Jerry DJ. Repression of mammary stem/progenitor cells by p53 is mediated by Notch and separable from apoptotic activity. *Stem Cells*. 2011;29(1):119-27.
11. Serrano M, Lee H, Chin L, Cordon-Cardo C, Beach D, DePinho RA. Role of the INK4a locus in tumor suppression and cell mortality. *Cell*. 1996;85(1):27-37.
12. Jonkers J, Meuwissen R, van der Gulden H, Peterse H, van der Valk M, Berns A. Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nat Genet*. 2001;29(4):418-25.
13. Dassule HR, Lewis P, Bei M, Maas R, McMahon AP. Sonic hedgehog regulates growth and morphogenesis of the tooth. *Development*. 2000;127(22):4775-85.
14. Sun P, Yuan Y, Li A, Li B, Dai X. Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. *Histochem Cell Biol*. 2010;133(2):213-21.
15. Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, et al. Distinct stem cells contribute to mammary gland development and maintenance. *Nature*. 2011;479(7372):189-93.
16. Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L. A global double-fluorescent Cre reporter mouse. *Genesis*. 2007;45(9):593-605.
17. Mitchell EH, Serra R. Normal mammary development and function in mice with Irf8 deleted in MMTV- and K14-Cre expressing cells. *Cilia*. 2014;3(1):4.
18. Asselin-Labat ML, Sutherland KD, Vaillant F, Gyorki DE, Wu D, Holroyd S, et al. Gata-3 negatively regulates the tumor-initiating capacity of mammary luminal progenitor cells and targets the putative tumor suppressor caspase-14. *Mol Cell Biol*. 2011;31(22):4609-

Tumor suppressors affect mammary gland proliferation

- 22.
19. Ting AT, Bertrand MJM. More to Life than NF- κ B in TNFR1 Signaling. *Trends Immunol.* 2016;37(8):535-45.
20. Cai S, Kalisky T, Sahoo D, Dalerba P, Feng W, Lin Y, et al. A Quiescent Bcl11b High Stem Cell Population Is Required for Maintenance of the Mammary Gland. *Cell Stem Cell.* 2017;20(2):247-60.e5.
21. Rivera J, Tessarollo L. Genetic background and the dilemma of translating mouse studies to humans. *Immunity.* 2008;28(1):1-4.
22. Bar J, Moskovits N, Oren M. Involvement of stromal p53 in tumor-stroma interactions. *Semin Cell Dev Biol.* 2010;21(1):47-54.
23. Janbandhu VC, Moik D, Fässler R. Cre recombinase induces DNA damage and tetraploidy in the absence of loxP sites. *Cell Cycle.* 2014;13(3):462-70.
24. Scheeren FA, Kuo AH, van Weele LJ, Cai S, Glykofridis I, Sikandar SS, et al. A cell-intrinsic role for TLR2-MYD88 in intestinal and breast epithelia and oncogenesis. *Nat Cell Biol.* 2014;16(12):1238-48.
25. Ip MM, Shoemaker SF, Darcy KM. Regulation of rat mammary epithelial cell proliferation and differentiation by tumor necrosis factor-alpha. *Endocrinology.* 1992;130(5):2833-44.
26. Peng WC, Logan CY, Fish M, Anbarchian T, Aguisanda F, Álvarez-Varela A, et al. Inflammatory Cytokine TNF α Promotes the Long-Term Expansion of Primary Hepatocytes in 3D Culture. *Cell.* 2018;175(6):1607-19.e15.
27. Rocha S, Campbell KJ, Perkins ND. p53- and Mdm2-independent repression of NF-kappa B transactivation by the ARF tumor suppressor. *Mol Cell.* 2003;12(1):15-25.
28. Perkins ND. The diverse and complex roles of NF- κ B subunits in cancer. *Nat Rev Cancer.* 2012;12(2):121-32.
29. Weisz L, Damalas A, Lontos M, Karakaidos P, Fontemaggi G, Maor-Aloni R, et al. Mutant p53 enhances nuclear factor kappaB activation by tumor necrosis factor alpha in cancer cells. *Cancer Res.* 2007;67(6):2396-401.
30. Taniguchi K, Karin M. NF- κ B, inflammation, immunity and cancer: coming of age. *Nat Rev Immunol.* 2018;18(5):309-24.
31. Chakrabarti R, Celià-Terrassa T, Kumar S, Hang X, Wei Y, Choudhury A, et al. Notch ligand Dll1 mediates cross-talk between mammary stem cells and the macrophageal niche. *Science.* 2018;360(6396).
32. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34(3):267-73.
33. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-50.
34. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1(6):417-25.
35. Prater, M. Shehata, M. Watson, C. J. Stingl, J. Enzymatic dissociation, flow cytometric analysis, and culture of normal mouse mammary tissue. *Methods Mol Biol* 2013;946:395-409.
36. Hu, Y. and Smyth, G.K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods* 2009;15;347(1-2):70-8.
37. Huang, D.W., Sherman, B.T., and Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat Protoc* 2009;4(1), 44-57.

SUPPLEMENTARY EXPERIMENTAL PROCEDURES

Animal care and use

Trp53^{fl/fl} (Stock #008462), *Krt-14-Cre* (Stock #004782), and *mT/mG* (Stock #007676) mice were purchased from The Jackson Laboratory. *Cdkn2a* null (Stock #01XB1), were purchased from the NCI Mouse Repository. Mice were backcrossed into the C57/BL6J background for at least six generations. For transplantation assay recipients, weaning-aged (21-25 days) C57BL/6J females (#000664) were purchased from The Jackson Laboratory. Only female mice were used. All mice used for this study were maintained at the Stanford Veterinary Service Center in accordance with the guidelines of the Administrative Panel on Laboratory Animal Care (APLAC #10868).

Tissue processing and flow cytometry

8-14-week-old virgin female mice were euthanized and fat pads surgically resected. Tissue was mechanically dissociated. Tissue used for transplantation assays was digested in DMEM/F12 (catalog #12634028, Gibco) with collagenase and hyaluronidase (catalog #07912, STEMCELL Technologies Inc.) and DNase I (Catalog #LS002139, Worthington) for 2 hr with gentle pipetting every 30 min. Tissue used for the TNF α organoid growth assay was digested in DMEM/F12 with gentle collagenase and hyaluronidase (catalog #07919, STEMCELL Technologies Inc.) overnight. The remainder of the protocol was the same for all experiments and has been previously described (35) with minor revisions. Briefly, lysis of red blood cells in ACK (catalog #BW10548E Lonza) for 5 min, was followed by 5 min treatment with pre-warmed 0.25% Trypsin-EDTA (catalog #25-510, GenClone), followed by treatment with pre-warmed dispase (catalog #07913, STEMCELL Technologies Inc.) and DNase I (catalog #LS002139, Worthington) for 2 min. Ultimately, the cells were filtered through a 40 μ m mesh and washed and resuspended in HBSS+2%FBS+PSA. For FACS analysis and sorting, mammary single cells were stained with CD45 (clone #30-F11, Biolegend), CD31 (clone #390, Biolegend), TER-119 (clone #TER119, Biolegend), CD49f (clone #GoH3, Biolegend), CD24 (clone #M1/69, Biolegend), CD14 (clone #Sa14-2, Biolegend). To enrich for epithelial cells preceding FACS, most of lineage was depleted on the autoMACS Pro Separator (catalog #130-092-545, Miltenyi Biotec) using microbeads for CD45 (catalog #130052301, Miltenyi Biotec), CD31 (catalog #130097418, Miltenyi Biotec), and TER119 (catalog #130049901, Miltenyi Biotec) according to manufacturer's instructions. Flow cytometry was performed with a 100 μ m nozzle on a BD Flow cytometry Aria II with Flow cytometry Diva software (BD Biosciences). Data analysis was done using FlowJo V10 (FlowJo). For all experiments, debris and cell doublets were excluded using side scatter and forward scatter profiles (area and width). Dead cells were eliminated using DAPI (catalog #32670, Sigma).

Transplantations

Using FACS, lineage (CD45/CD31/TER119) negative, GFP⁺ cells were sorted in HBSS+2%FBS+PSA media and counted using a hemocytometer. Cells were resuspended with 50% Matrigel (catalog #356234, Corning). If sorted for specific populations, it is marked in the main text. For each transplant, 10 μ l was injected into cleared fat pads of weaning-age mice (21-25 days). Transplantation procedure has been previously described (20). After 7-20 weeks, transplants were dissected and photographed using a fluorescent microscope. The repopulation frequency was calculated using ELDA (36). As for secondary transplantations, first 5,000 cells were transplanted in a primary transplantation assay. Primary transplants that took up more than 10% of

the recipient fat pad, were individually digested, and used for secondary transplantations. After 7 weeks the secondary transplants were dissected and photographed using a fluorescent microscope.

Mammary organoid growth assay

For the organoid assays, 40 μ l growth factor reduced Matrigel (catalog #356230, Corning) was plated in a 96 well plate and solidified at 37 °C for 10 min. For the luminal CD14 assay, growth factor reduced Matrigel was mixed prior to plating with 10,000 L1-Wnt3a feeder cells (generous gift from Dr. Roel Nusse) that were administered a 40Gy dose of X-ray irradiation. MECs were plated on top in 150 μ l culture media. Culture media for the luminal CD14 organoid assay has previously been described (Cai et al., 2017). In short, the culture media used contained DMEM/F12 (catalog #12634028, Gibco) + 2% FBS + PSA + B27 (Catalog #12587010, Gibco) + 10 mM HEPES (catalog #15630080, Gibco) + 10 ng/ml EGF (Catalog #315-09, PeproTech) + 10 μ M ROCK inhibitor Y27632 (catalog #Y0503, Sigma) + Noggin (catalog #6057-NG, R&D Systems) + N2 (catalog #17502001, Gibco); culture media was replaced twice-a-week. For the TNF α organoid growth assay, culture media was made with DMEM/F12 + 2% FBS + PSA + B27 + 10 mM HEPES + Glutamax (catalog #35050061, Gibco) + N2; culture media was enriched with Recombinant Murine TNF α (catalog #315-01A, PeproTech), TNF α was reconstituted according to manufacturer's instructions and was freshly added to the organoids four times a week. Culture media was replaced twice-a-week. All plates were maintained in a 37 °C incubator at 5% CO₂ for two weeks.

Real-time PCR

5,000 cells of various MEC populations were directly sorted into RNeasy Protect (catalog #76526, Qiagen). RNA was extracted using the RNeasy micro kit (catalog #74004, Qiagen), according to manufacturer's instructions. Reverse transcription of RNA to cDNA was done using SuperScript III First Strand Synthesis kit (catalog #18080051, Invitrogen) according to the manufacturer's instructions. cDNA was preamplified 20 cycles using SybrGreen master mix (catalog #4364346, Applied Biosystems) and Sybergreen primers for *Actb* and *Trp53*, sequences listed below. Preamplified cDNA was used for real-time PCR, with Sybergreen primers for *Actb* and *Trp53*, sequences listed below. All data was normalized to *Actb*. Using the 7900HT Real Time PCR system (catalog # 4329001, Applied Biosystems). Data was analyzed by SDS2.4 software, Excel and Prism 8. Sybergreen primers: *Actb*: Forward GGCTGTATCCCTCCATCG Reverse CCAGTTGGTAAACAATGCCATGT; *Trp53*: Forward CTCTCCCCCGCAAAGAAAA Reverse CGGAACATCTCGAAGCGTTA.

RNA Microarray expression analysis

15,000-60,000 cells of basal MECs were directly sorted into RNeasy Protect (catalog #76526, Qiagen) using FACS. RNA was extracted using the RNeasy micro kit (catalog #74004, Qiagen), according to manufacturer's instructions. Library preparation, hybridization and scanning were all performed by Stanford protein and nucleic acid facility (PAN facility) using the Mouse Gene 2.0 ST Array (Affymetrix). Data was analyzed using Transcriptome Analysis Console Software (ThermoFisher Scientific). Affymetrix gene IDs were converted to Official Gene Symbols using DAVID's Gene ID Conver-

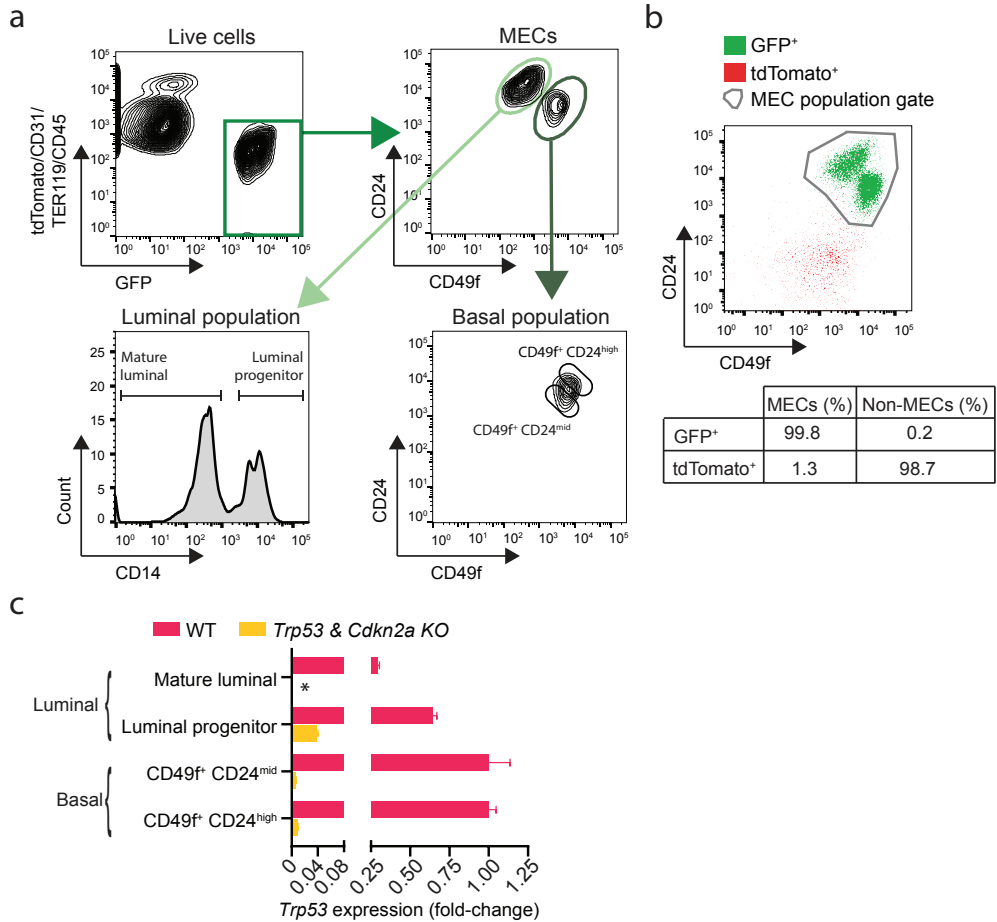
Chapter 2

sion Tool v6.8 (37). Using MS Excel, data lacking an associated gene name were deleted from the resulting list. Patterns in gene expression were analyzed using GSEA v4.0.1 (32, 33) and the hallmark gene sets (34). Data was deposited in GEO.

Quantification and statistical analysis

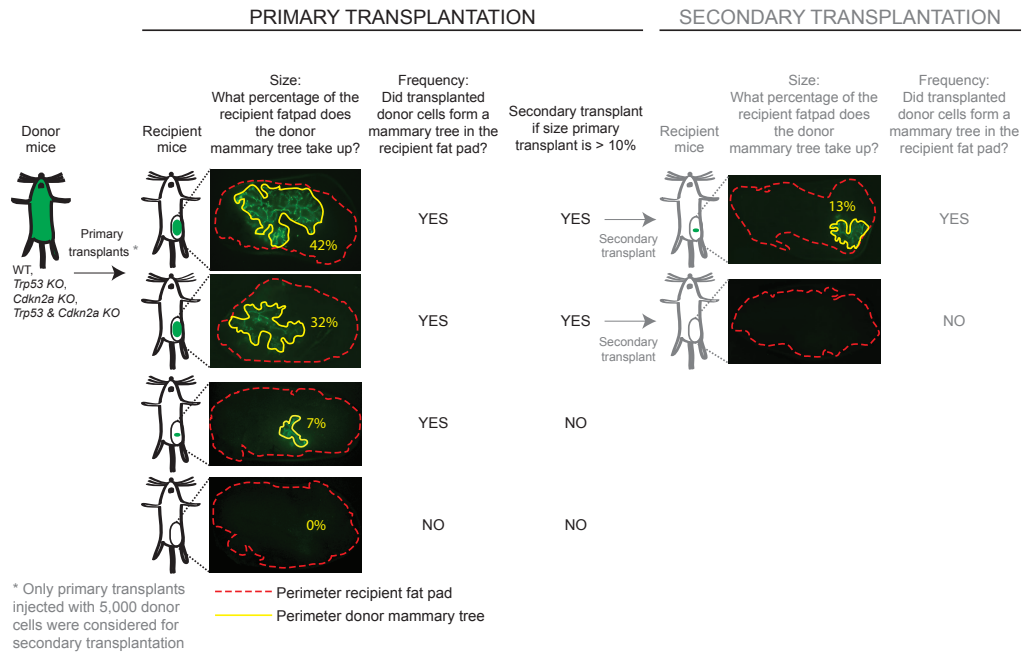
To determine if there were differences among cells from different genotypes or variations in responses to treatments, one-way or two-way ANOVA was calculated using GraphPad Prism 8. In case of a binary outcome, the Fisher's exact test was calculated using GraphPad Prism 8. Statistical analysis was performed only done if data from at least 3 biological replicates was present. Exact number of biological and technical replicates as well as statistical parameters are given in the figures and corresponding figure legends. For analysis of limiting dilution data, ELDA software was used (36). The expected frequencies are shown, as well as the 95% confidence interval (the lower and upper values are shown). To count organoid number and size, images taken from each experiment were processed simultaneously. Fluorescent images were first opened in Photoshop CS6, threshold function was used to remove background noise and to only show organoids to allow for automated counting and size analysis. Subsequently, using ImageJ v1.51, Analyze Particles was used to count and measure the size of the organoids. No statistical method was used to predetermine the size of the sample, experiments were not randomized or blinded. Researchers were blinded when assessing outcomes of take rate and outgrowth size of donor MECs, the mouse ear tag was used for identification.

SUPPLEMENTARY FIGURES

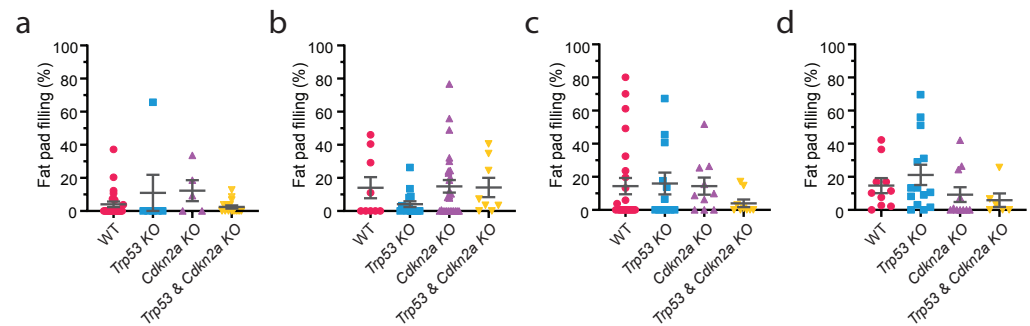


Supplementary Figure 1. *Trp53* expression is absent in MECs in inducible *Trp53* KO mouse model

(a) Flow cytometry gating strategy used throughout this study. GFP⁺ cells are selected from viable cells, eliminating lineage (CD31/TER119/CD45). CD49f and CD24 markers distinguish the luminal and basal MEC populations. CD14 differentiates between a luminal progenitor and a mature luminal population. CD49f⁺ CD24^{high} and CD49f⁺ CD24^{mid} mark two basal populations. (b) Flow cytometry analysis of *Trp53* & *Cdkn2a* KO MECs showing the efficiency of the *Krt14-Cre* and *mT/mG* transgenes. Cells that were successfully targeted by Cre and express GFP, are shown in green. Cells that did not get targeted by Cre and continue to express tdTomato, are shown in red. In the FACS plot, lineage- DAPI- cells are shown, within the gate the epithelial basal and luminal populations and the remaining non-epithelial cells outside of the gate. (c) Real-time qPCR of *Trp53* expression in various MEC populations of the WT and *Trp53* & *Cdkn2a* KO mammary glands. *Trp53* expression was normalized against *Actb* and expression values are set relative to *Trp53* expression in the WT CD49f⁺ CD24^{high} population. Asterisk (*) indicates *Trp53* was undetectable in sample. Data are represented as mean ± SEM, data from 3 technical replicates are shown.

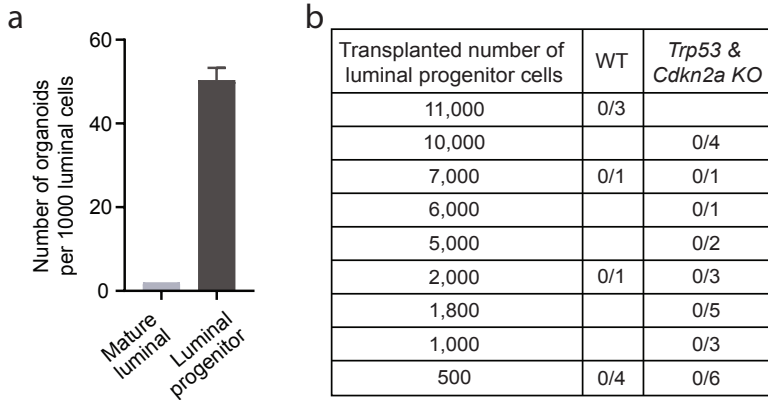


Supplementary Figure 2: Schematic of primary and secondary transplantation workflow and readout
Mammary fat pads of donor mice are digested, sorted by FACS, counted, and donor cells are injected into the cleared fat pad of syngeneic wild type recipient mice. After a minimum of seven weeks, glands are analyzed. Glands are analyzed for frequency (are injected donor cells able to regenerate yes or no?) and size (the percentage of the recipient fat pad the donor mammary tree fills). Only primary transplants that started with the injection of 5,000 and grew out to occupy at least 10% of the recipient fat pad, were digested and used for the secondary transplantation experiment.



Supplementary Figure 3: WT, *Trp53* KO, *Cdkn2a* KO, and *Trp53* & *Cdkn2a* KO MECs display equal proliferative capacity (a-d) Additional comparisons of the sizes of the primary transplants. The size of individual outgrowths is shown, grouped per number of cells injected. Data are represented as mean \pm SEM, each individual symbol represents one transplant. Data are shown for 500 (a), 1,250 (b), 2,500 (c), and 10,000 (d) injected cells. ($P > 0.05$ for all comparisons; one-way ANOVA)

Tumor suppressors affect mammary gland proliferation



Supplementary Figure 4: *Trp53* & *Cdkn2a* KO luminal progenitors do not regenerate *in vivo* (a) *In vitro* organoid formation assay. Comparison of the number of organoids forming after plating mature luminal (luminal CD14⁻) and luminal progenitor (luminal CD14⁺) cells. Data are represented as mean ± SEM, three technical replicates are shown. (b) Limiting dilution transplantation assay with WT and *Trp53* & *Cdkn2a* KO luminal progenitors. The numbers show the Take Rate/Total Number of Transplants. Data are derived from 2 WT and 5 *Trp53* & *Cdkn2a* KO donor mice.

Gene sets enriched in WT basal MECs			
Name gene set	NES	NOM p-val	FDR q-val
FATTY_ACID_METABOLISM	1.899503	<0.001	0.001871
G2M_CHECKPOINT	1.814166	<0.001	0.004539
OXIDATIVE_PHOSPHORYLATION	1.69856	<0.001	0.016057
E2F_TARGETS	1.554148	0.001563	0.046384
ESTROGEN_RESPONSE_EARLY	1.516312	0.004823	0.055867
ESTROGEN_RESPONSE_LATE	1.509459	0.003125	0.049083

Gene sets enriched in <i>Trp53</i> & <i>Cdkn2a</i> KO basal MECs			
Name gene set	NES	NOM p-val	FDR q-val
INTERFERON_ALPHA_RESPONSE	-2.57537	<0.001	0
INTERFERON_GAMMA_RESPONSE	-2.392	<0.001	0
TNFA_SIGNALING_VIA_NFKB	-1.79737	<0.001	0.00423
WNT_BETA_CATENIN_SIGNALING	-1.53162	0.028504	0.030907
IL6_JAK_STAT3_SIGNALING	-1.52458	0.012626	0.026912
HYPOXIA	-1.473	0.004988	0.036912

Supplementary Table 1: GSEA hallmark gene sets analysis results GSEA results using the hallmark gene sets, comparing basal WT to *Trp53* & *Cdkn2a* KO basal MECs. The normalized enrichment score (NES), Nominal p value (NOM p-val), and false discovery rate (FDR q-val) of each gene set enriched at an FDR q-value < 0.05 are shown. Data in table represent the values of 3 biological replicates in each group.

Chapter 3

Mesenchymal tumor cells drive adaptive resistance of *Trp53*^{-/-} breast tumor cells to inactivated mutant *Kras*

Linda J. van Weele¹ • Sabra I. Djomehri¹ • Shang Cai^{1,5} • Jane Antony¹ • Shaheen S. Sikandar^{1,6} • Dalong Qian¹ • William H.D. Ho^{1,2} • Robert West³ • Ferenc A. Scheeren⁴ • Michael F. Clarke¹

Adapted from: *Molecular Oncology*, doi:10.1002/1878-0261.13220, accepted for publication (2022)

¹Institute for Stem Cell Biology and Regenerative Medicine, School of Medicine, Stanford University, Stanford, CA 94305, USA. ²Department of Stem Cell Biotechnology, California State University Channel Islands, Camarillo, CA 93012, USA. ³Department Pathology, Stanford University Medical Center, Palo Alto, CA 94304, USA. ⁴Department of Medical Oncology, Leiden University Medical Center, Leiden RC 2300, The Netherlands. ⁵Current address: Westlake University, Shilongshan St #18, Xihu District, Hangzhou, Zhejiang Province, China. ⁶Current address: Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, CA 95064, USA.

As precision medicine increases the response rate of treatment, tumors frequently bypass inhibition and reoccur. In order for treatment to be effective long term, the mechanisms enabling treatment adaptation need to be understood. Here, we report a mouse model that, in the absence of p53 and the presence of oncogenic *Kras*^{G12D}, develops breast tumors. Upon inactivation of *Kras*^{G12D}, tumors initially regress and enter remission. Subsequently, the majority of tumors adapt to the withdrawal of *Kras*^{G12D} expression and return. *Kras*^{G12D}-independent tumor cells show a strong mesenchymal profile with active RAS-RAF-MEK-ERK (MAPK/ERK) signaling. Both *Kras*^{G12D}-dependent and *Kras*^{G12D}-independent tumors display a high level of genomic instability, and *Kras*^{G12D}-independent tumors harbor numerous amplified genes that can activate the MAPK/ERK signaling pathway. Our study identifies both epithelial-mesenchymal transition (EMT) and active MAPK/ERK signaling in tumors that adapt to oncogenic *Kras*^{G12D} withdrawal in a novel *Trp53*^{-/-} breast cancer mouse model. To achieve long-lasting responses in the clinic to RAS-fueled cancer, treatment will need to focus in parallel on obstructing tumors from adapting to oncogene inhibition.

Both tumor suppressor TP53 and oncogene KRAS are commonly altered in cancer and are known drivers of breast cancer. While *TP53* is the most commonly mutated gene in breast cancer, *KRAS* is more frequently mutated in other cancer types (1). Nevertheless, introducing the expression of mutant *KRAS*^{G12D} in normal human mammary epithelium is sufficient to induce tumor formation (2). Furthermore, the RAS-RAF-MEK-ERK (hereafter referred to as: MAPK/ERK) signaling pathway is frequently overactivated in all types of breast cancer, particularly in the molecular subtype basal-like breast cancer. In basal-like breast cancer, amplifications of MAPK/ERK pathway genes are common and *KRAS* amplifications are detected in 32% of tumors (3, 4). Basal-like breast cancer frequently overlaps with a triple-negative breast cancer (TNBC) diagnosis. Due to the lack of targeted or hormonal therapy for TNBC patients, the TNBC subtype has the worst prognosis with a 4-year relative survival of 77.0% (5).

To study the effect of oncogenic *Kras* expression in breast cancer *in vivo*, we chose to use an *KRAS*^{G12D}-inducible mouse model. The expression of *KRAS*^{G12D} mimics the overactivation of the MAPK/ERK pathway, as the continuously active *KRAS*^{G12D} is the initial driver of this pathway. In addition, we aimed to shed light on what mechanisms tumors use to adapt to targeted therapies. Mutated *KRAS* has been a target of drug development for multiple decades due to its high prevalence in several cancers. Last year, the FDA approved the first drug targeting mutant *KRAS*. This *KRAS*^{G12C} inhibitor is approved for patients with *KRAS*^{G12C} advanced non-small cell lung cancer and other *KRAS*^{G12C} inhibitor are currently in clinical trials (6-8). However, data released from drugs targeting *KRAS*^{G12C}, showed that tumors find ways to bypass *KRAS*^{G12C} inhibition (7, 9). Here we report on a novel breast cancer *Trp53*^{-/-} *KRAS*^{G12D}-inducible mouse model that develops tumors that also adapt to *Kras* inhibition. We find that the tumors that are able to overcome oncogene withdrawal adopt a mesenchymal identity and reactivate the MAPK/ERK pathway. This novel immunocompetent mouse model helps us understand the mechanisms that tumors deploy to bypass *KRAS* inhibitors and could provide a valuable resource for determining effective methods to treat patients more durably.

RESULTS

Generation of a *Trp53*^{-/-} *Kras*^{G12D}-inducible mouse model

To study the mechanisms of tumors bypassing *Kras*^{G12D} inhibition *in vivo*, we used the *MMTV-rtTA TetO-Kras*^{G12D} (*iKras*) mouse. In *MMTV-rtTA TetO-Kras*^{G12D} mice, the expression of *Kras*^{G12D} is activated in the mammary gland in the presence of doxycycline (dox) (10-12). Since in human breast cancer, *KRAS* genetic alterations frequently co-occur with *TP53* genetic alterations (Fig. S1A), we crossed the *MMTV-rtTA TetO-Kras*^{G12D} mouse with the *Krt14-Cre mT/mG Trp53*^{fl/fl} (*Trp53* KO) mouse. In the *Trp53* KO mouse, epithelium-specific Cre recombinase (Cre) activity triggers the deletion of *Trp53* (13, 14). In parallel, Cre inactivates the expression of *tdTomato* and activates the expression of *GFP* in the *mT/mG* allele (15). We have described the *Trp53* knockout (KO) mouse model previously (16). The various crosses lead to the two following mouse models which are used in this study: 1) *Trp53* KO *iKras*^{G12D}; *Krt14-Cre mT/mG Trp53*^{fl/fl} *MMTV-rtTA TetO-Kras*^{G12D} (Fig. S1B) and 2) *iKras*^{G12D}; *Krt14-Cre mT/mG MMTV-rtTA TetO-Kras*^{G12D}. In brief, the *Trp53* KO *iKras*^{G12D} mouse mammary gland does not express *Trp53* and oncogenic *Kras*^{G12D} expression is induced in the presence of doxycycline. The *iKras*^{G12D} mouse only has the inducible *Kras*^{G12D} construct.

A subset of *Trp53*^{-/-} tumors bypasses *Kras*^{G12D} withdrawal and resumes proliferation

Upon the expression of *Kras*^{G12D}, 25.8% (8 out of 31) of *Trp53* KO *iKras*^{G12D} mice developed mammary gland tumors, while none of the *iKras*^{G12D} mice developed

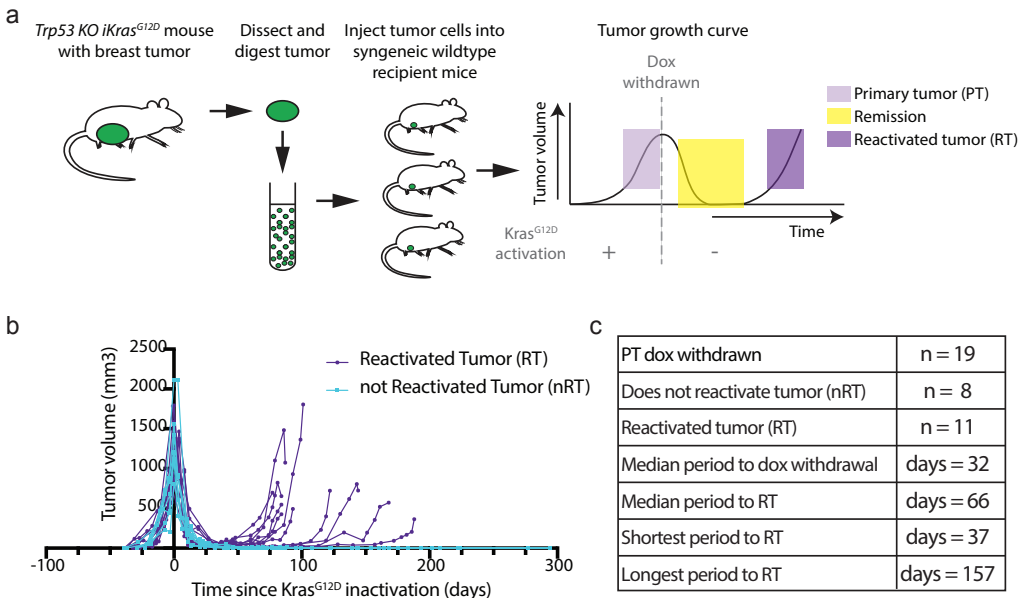


Figure 1. Mammary tumor formation in the *Trp53* KO *iKras*^{G12D} mouse model (a) Schematic showing the *Trp53* KO *iKras*^{G12D} mouse model and the tumor growth curve. (b) Tumor growth curves of tumors where *Kras*^{G12D} activation led to tumor growth, and *Kras*^{G12D} inactivation (day 0) led to either a reactivated tumor (RT, n=11) or led to a not reactivated tumor (nRT, n=8). (c) Table showing the data displayed in B. Median period to dox withdrawal is time between the injection of tumor cells to dox withdrawal. Median period to RT is the time between dox withdrawal and a minimum of doubling of the tumor in size after remission with continued growth thereafter.

Chapter 3

mammary gland tumors (0 out of 15) (Fig. S1C). Analyzing the expression of the epithelial cell surface marker EpCAM by flow cytometry in the 8 *Trp53* KO *iKras*^{G12D} tumors, we found 3 types of *Kras*^{G12D} tumors: a predominantly EpCAM^{high} tumor, a predominantly EpCAM^{low} tumor, and a mixed tumor with similar proportions of EpCAM^{high} and EpCAM^{low} cells. To study the impact of *Kras*^{G12D} withdrawal, we injected *Trp53* KO *iKras*^{G12D} tumors into syngeneic recipient mice. The EpCAM^{low} and the mixed tumor types did not dependent on the inducible *Kras*^{G12D} protein and continued growing upon dox withdrawal. Among the 6 predominantly EpCAM^{high} tumors, 4 established new tumors in syngeneic mice of which 2 also did not dependent on *Kras*^{G12D} expression. This study focuses on the 2 remaining EpCAM^{high} tumors: tumors that declined rapidly upon dox withdrawal, thereby showing to initially dependent on *Kras*^{G12D} expression for tumor formation (Primary Tumor; PT) (Fig. 1A, Fig. S1C). After a median period of 66 days, nearly 60% (11 out of 19) of the tumors in remission had adapted to the absence of *Kras*^{G12D} and reactivated tumor growth (Reactivated Tumor; RT) (Fig. 1B,C). We confirmed transgenic *Kras*^{G12D} expression was not restored in RTs (Fig. S2A) and the successful excision of *Trp53* by PCR in all tumors was confirmed (Fig. S2B). The PTs and RTs formed in the *Trp53* KO *iKras*^{G12D} mouse model enabled us to study how mammary gland tumors overcome the withdrawal of oncogene expression. Taken together, these results show that tumors adapt to *Kras*^{G12D} inactivation.

Reactivated tumors are enriched for EMT markers

To learn more about the PT and the RT, we analyzed EpCAM and CD49f expression on both tumor types and on tumors six days after dox withdrawal, using flow cytometry. In the absence of dox, there was a rapid reduction in the number of EpCAM^{high} cells. Once the tumor resumed to grow again, the tumor consisted predominantly of EpCAM^{low} tumor cells (Fig. 2A,B). The change in EpCAM expression between *Trp53* KO *iKras*^{G12D} primary and reactivated tumors, suggests a switch in tumor type since EpCAM is an epithelial cell marker usually absent in mesenchymal tumors. In support, H&E staining shows changes in tumor histology: PTs displayed an epithelial, invasive phenotype and RTs a stromal, spindle-like, mesenchymal-like, invasive phenotype (Fig. 2C, Fig. S2C). To study the differences between PTs and RTs in more detail, we performed RNA-seq (Fig. S3A,B). Gene set enrichment analysis (GSEA) displayed multiple differences between the two tumor types. The strongest enrichment was found in the epithelial-mesenchymal transition (EMT) hallmark gene set in RTs (Fig. 3A). Moreover, the gene ontology annotations enriched for were cell adhesion and extracellular organization in PTs (Fig. 3B) and developmental processes in RTs (Fig. 3C). In addition, multiple epithelial genes clustered in the PTs and mesenchymal genes in the RTs (Fig. 3D). Lastly, a single sample GSEA (ssGSEA) with a signature set developed to distinguish different tumor types in mice (17), confirmed that the PTs expressed an EMT down signature while the RTs expressed an EMT up signature (Fig. S3C). Together, flow cytometry, histology, and RNA-Seq data showed a change from a dominant epithelial phenotype in the PT to a dominant mesenchymal phenotype in the RT.

Reactivated tumors display active MAPK/ERK signaling

Since the GSEA analysis showed no enrichment of *Kras* signaling in PTs, we looked deeper into the activity of the MAPK/ERK pathway in both tumor types. The MAPK/

ERK signaling cascade commences when a phosphorylated receptor tyrosine kinase (RTK) catalyzes the activation of GTPase Ras. Subsequently, this leads to a catalyzation cascade of related protein-serine/threonine kinases ultimately steering transcription in the nucleus (Fig. 4A). Focusing on a group of genes known to give accurate predictions of MAPK/ERK pathway activity (18), we found no change in the activity of the MAPK/ERK pathway in RTs (Fig. 4B). To confirm this observation, we stained for phosphorylated ERK1/2 (pERK1/2), a marker of activated MAPK/ERK signaling. In the normal adult mammary gland, the vast majority of mammary epithelial cells did not express pERK1/2. Unlike PT and the RT cancer cells, where pERK1/2 was abundant (Fig. 4C, Fig. S4). Quantification showed that the RTs contained a higher number (Fig. 4D) but lower per cell presence of pERK1/2 (Fig. 4E). Cumulatively, similar amounts of pERK1/2 were present in the PTs and RTs (Fig. 4F). Together, this data showed that in the process of bypassing the loss of *Kras*^{G12D} expression, RT cells are either able to maintain and/or reactivate MAPK/ERK signaling (Fig. 4G).

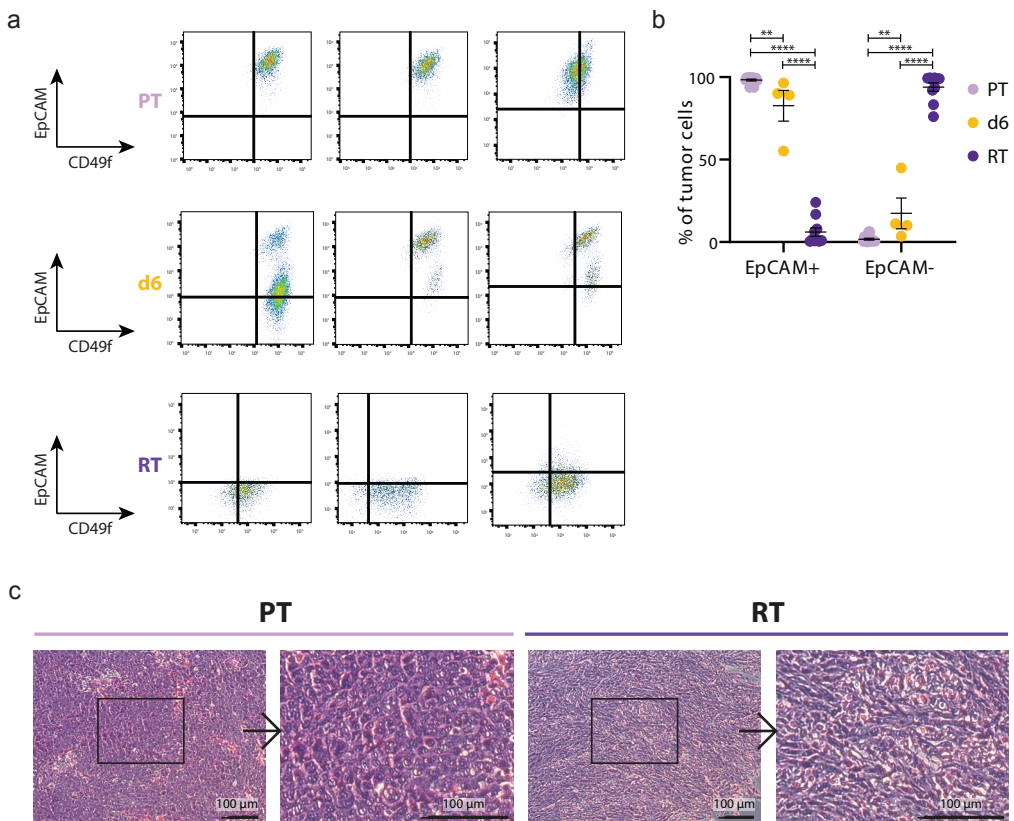


Figure 2. The RT loses EpCAM expression upon *Kras*^{G12D} inactivation and becomes mesenchymal-like (a) Representative FACS plots of 3 PT, 3 RT, and 3 six days after dox withdrawal (d6) tumors. EpCAM and CD49f expression of fluorescent tumor cells, lineage (Ter119/CD31/CD45)- DAPI-. PT n=3, d6 n=3, RT n=3. (b) The percentage of EpCAM^{high} and EpCAM^{low} cells in the PT, d6, and RT tumors. PT n=14, d6 n=4, RT n=10. Mean ± SEM is shown (2way ANOVA, ** p<0.005, **** p<0.0001). (c) Pathology, H&E staining. PTs have an epithelial, invasive phenotype. RT have a stromal, spindle-like, mesenchymal-like, invasive phenotype. Scale bar is set at 100 μm.

The development of tumor reactivation cannot be explained by mutational data

Next, we investigated if genetic changes are responsible for the observed restoration of MAPK/ERK signaling in RTs. Therefore, we performed whole exome sequencing (WES) on two PTs with each three matched RTs. All tumor samples acquired multiple mutations that continued to accumulate in the RTs (Fig. 5A, Table S1). Mutations were mostly single nucleotide variants (SNVs) (Fig. S5A). The nonsynonymous mutations were mostly missense mutations (Fig. S5B). Comparing the mutational profiles of the matched tumors, not all mutations found in the PT were present in the matched RTs and RTs showed different mutations among themselves (Fig. 5B). suggested that each PT was a heterogeneous tumor with multiple clones that have acquired different mutations. Next, aiming to identify mutations that could be responsible for the *Kras*^{G12D}-independent activation of MAPK/ERK signaling, we looked for overlap in genes with nonsynonymous mutations in various samples (Fig. S5C). Specifically, we selected altered genes that were present in at least one RT of each mouse and absent in either PT. However, the 7 genes matching this description (*Clasrp*, *Cyp3a44*, *Gpr18*, *Lrp2*, *Map1b*, *Spen*, *Ttn*), are neither known breast cancer or pan-cancer driver genes

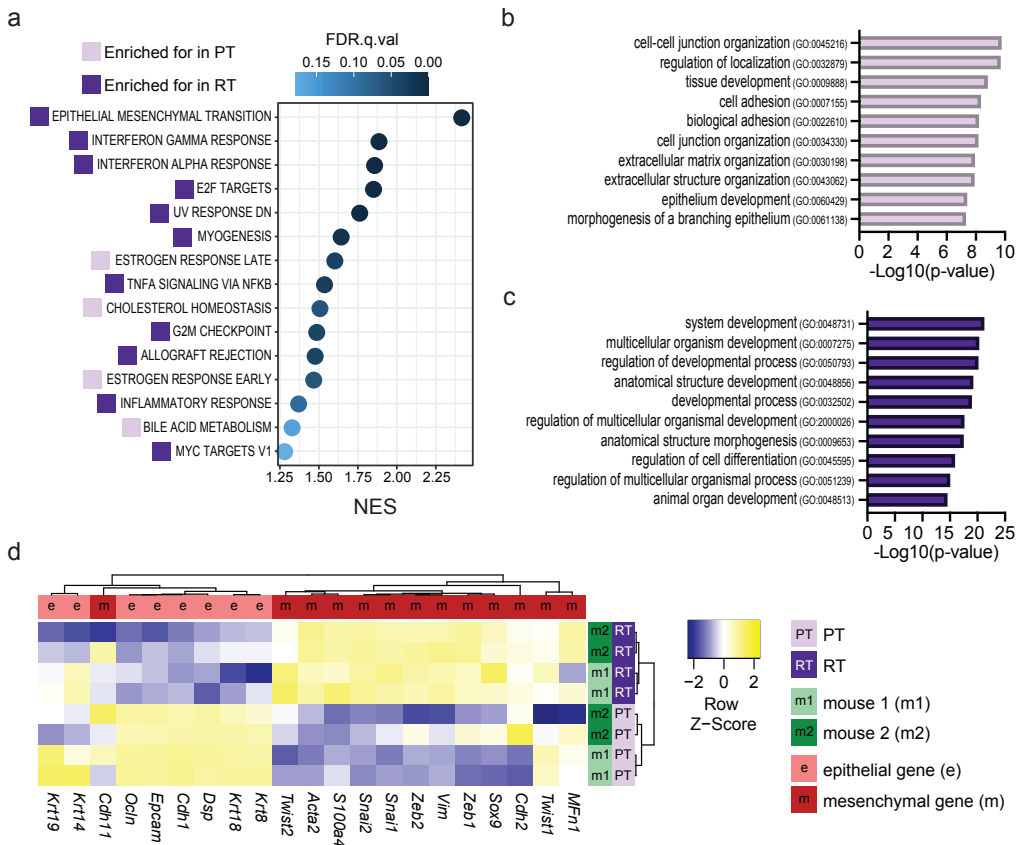


Figure 3. RTs display a strong EMT transition profile Comparison of the transcriptomes of 4 PTs and 4 RTs. (a) GSEA hallmark gene sets upregulated in PTs and RTs (nom. p-value < 0.05, FDR q-value < 0.25). (b) Top 10 gene ontology terms enriched in PT. (c) Top 10 gene ontology terms enriched in RT. (d) Heat map of the expression of epithelial and mesenchymal marker genes.

nor directly involved with the activation of MAPK/ERK signaling (Fig. S5D, Table S2). Of note, the two PTs did not have any mutated genes in common, therefore acquired mutations by the PTs do not explain why these two mice developed breast cancer in

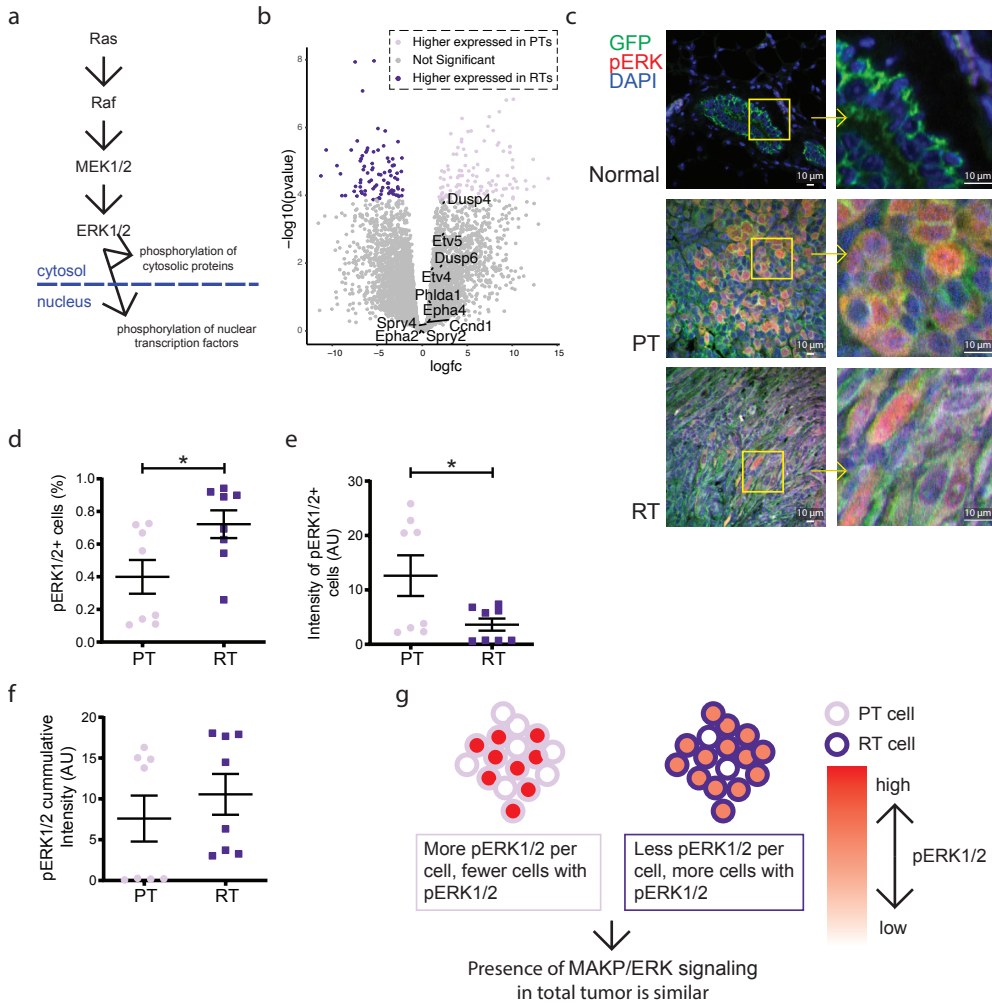


Figure 4. RTs activate MAPK/ERK signaling (a) Simplified schematic of the MAPK/ERK phosphorylation signal transduction cascade. (b) RNA-seq data showing the expression of the MAPK Pathway Activity Score (MAPS) genes in PTs and RTs. PT n=4, RT n=4. (c) Immunofluorescence staining for pERK1/2 (red), GFP (green), and DAPI (blue) of tissue deriving from a normal *Krt14-Cre mT/mG* mammary gland, a PT, and a RT. Scale bar is set at 10 μm . Images of a biological replicate are shown in Supplementary Fig 5. (d) The percentage of cells expressing pERK1/2 in PTs and RTs. Quantification of each 4 images of 2 biological controls of PTs and RTs are shown. Data are represented as mean \pm SEM. (Unpaired t test, * P < 0.05) (e) The intensity of pERK1/2 staining in pERK1/2 positive cells. Quantification of each 4 images of 2 biological controls of PTs and RTs are shown. Data are represented as mean \pm SEM. (Unpaired t test, * P < 0.05) (f) The cumulative intensity of pERK1/2 staining in the tumors. Quantification of each 4 images of 2 biological controls of PTs and RTs are shown. Data are represented as mean \pm SEM. (Unpaired t test, n.s.) (g) Model: constitutive active Kras^{G12D} results in strong MAPK/ERK signaling in a subset of the PT cells. RT cells do not reach as strong of a phosphorylation cascade per cell as in some PT cells. However, since a large percentage of individual RT cells activate MAPK/ERK signaling, the result for both tumor types as a whole is similar MAPK/ERK signaling.

Chapter 3

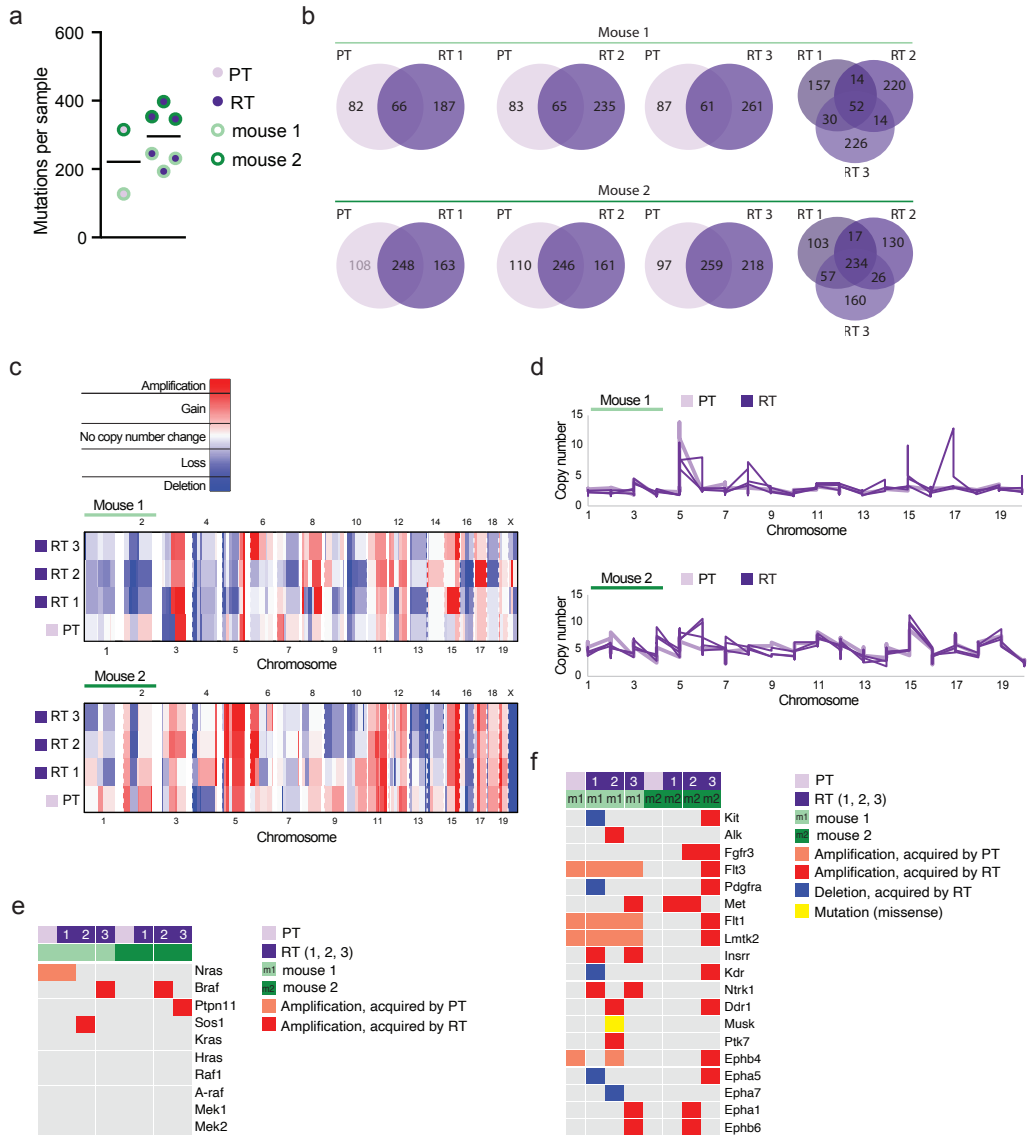


Figure 5. Tumors are genetically unstable and acquire multiple alterations that can activate MAPK/ERK signaling (a) Number of total acquired mutations by each PT and RT. The black horizontal line shows the median. (b) Venn diagrams showing the overlap and divergence of mutations acquired by a PT and its matched RTs (RT 1, 2, 3) and the overlap between mutations acquired by RTs originating from the same PT (mouse 1, mouse 2) (Venn diagram tool: (96)). (c) Copy number profiles with thresholds of ± 0.25 for gains/losses and ± 1 for amplifications/deletions are shown for PT and matched RTs compared to their matched normal genome. Heatmaps show PTs and their 3 matched RTs (RT 1, 2, 3) in both mice (mouse 1, mouse 2) separately. Red indicates gain/amplified and blue loss/deleted regions. (d) Absolute copy number profiles genome-wide for PTs and the 3 matched RTs are shown. Absolute copy numbers are computed based on logR values derived from genomic segments using Sequenza. The profiles are shown separately for the 2 mice (mouse 1, mouse 2). (e) Oncoplot showing genes directly involved with MAPK/ERK signaling, 4 out of 11 have genetic alterations. RTs (RT 1, 2, 3) are shown directly next to their matched PTs. (f) Oncoplot showing RTK genes with genetic alterations. RTs (RT 1, 2, 3) are shown directly next to their matched PTs.

contrast to the majority of *Trp53* KO *iKras*^{G12D} mice. Using our dataset, nonsynonymous mutations did not play a clear role in tumor adaptation.

Various MAPK/ERK pathway genes are amplified in reactivated tumors

Following the analysis of the tumors' mutational profiles, we looked at the copy number alteration (CNA) landscape of the PTs and matched RTs. CNA data showed a high level of chromosome instability (Fig. 5C, Fig. S6). The widespread increase in copy number was already present in PTs (Fig. 5D). On an individual sample level, we looked for amplified or deleted genes that are either part of the MAPK/ERK signaling pathway or could activate the MAPK/ERK signaling pathway, such as RTKs. Many of these genes were amplified in one or multiple RTs, while showing no copy number alterations in their matched PTs and no deletions in any of the other samples (Fig. 5E,F). Hence, the amplification of various components of the MAPK/ERK signaling pathway could provide a mechanism of how tumors restore or maintain active MAPK/ERK signaling.

DISCUSSION

After decades of research, the FDA approved the first KRAS inhibitor last year, targeting oncogenic KRAS^{G12C}. Although a huge milestone, many tumors use mechanisms to adapt to KRAS^{G12C} inhibition, resulting in only temporary relief for patients (7, 9). To study what mechanisms tumors deploy to adapt to *Kras* inactivation, we developed the *Trp53* KO *iKras*^{G12D} mouse model. The *Trp53* KO *iKras*^{G12D} mouse model spontaneously presented with tumors that have found methods to bypass *Kras*^{G12D} inactivation after a period of remission. Since C57BL/6J *iKras* mice did not develop mammary gland tumors, the absence of *Trp53* in the *Trp53* KO *iKras* mice played a role in the enablement of tumorigenesis. Previous research showed that it is context specific if the sole expression of oncogenic *Ras* is sufficient to induce tumorigenesis, is context specific (19-23). The absence of tumor formation is attributed to induced senescence in cells expressing high levels of oncogenic *Ras*, regulated by tumor suppressor genes such as *Trp53* and *Cdkn2a* (21, 22). In a mammary gland *Cdkn2a* KO mouse model, the expression of *Hras*^{G12V} resulted in bigger tumors in *Cdkn2a* KO tumors than in *Cdkn2a* wildtype tumors (22). In contrast to our data, a recent study found that the sole expression of *Kras*^{G12D} is sufficient to induce mammary tumor formation in the mouse mammary gland. Using a MMTV-tTA TetO-*Kras*^{G12D} mouse Rädler et al. and observed mammary tumor formation in 100% of mice with an average latency of 160 ± 41 days (23). There are some important differences between the two studies. We used the Tet-On system, activating the expression of *Kras*^{G12D} in adult mice, while Rädler et al. used the Tet-Off system, where *Kras*^{G12D} is expressed continuously until actively turned off. In addition, the two studies worked with different mouse strains. We backcrossed our mice into the C57BL/6J background, Rädler et al. created their mouse in the FVB/N genetic background, a commonly used strain in cancer research and for the creation of novel transgenic models. We chose the C57BL/6 mouse strain due to its low susceptibility to spontaneous tumor formation (24-28), which is a likely explanation why only a minority of *Trp53* KO *iKras* and none of the *iKras* mice developed mammary tumors. This is different in the FVB/N strain, at 14 months of age 26% of female wildtype FVB/N mice was tumor-bearing (26). A direct comparison between FVB/N and C57BL/6J breast tumor mouse models has been done with the use of the PyMT transgene, a popular transgenic breast cancer

Chapter 3

mouse model due to a short latency in primary mammary tumor development and metastasis to the lungs (29). When FVB/N-PyMT are crossed with only one generation of C57BL/6J mice, the latency of primary and metastatic tumor development stretches out significantly (25) and the latency period further increases when PyMT is backcrossed for at least 5 generation of C57BL/6J mice (30). Other benefits of using the C57BL/6J strain, are the availability of a high-quality genomic data (31) and the absence of an immune response that could clear GFP or tdTomato expressing cells (32-37). Low susceptibility to spontaneous tumor formation, the availability of genomic reference data, and the tolerability of fluorescent proteins were important reasons for us to choose the C57BL/6J strain.

In our *Trp53* KO *iKras* model, we observed that 58% of tumors reoccur. These tumors showed a strong EMT signature that is absent in the PTs. Other studies have shown a connection between EMT and resistance to RAS inhibition using cell lines. In human breast immortalized cell lines, EMT-marker *ZEB1* overcomes *HRAS* induced senescence (38). Lung and pancreatic cancer cell lines that depend on *KRAS* expression are uniformly epithelial while *KRAS*-independent cell lines were not (39). *KRAS*^{G12C}-mutant lung cancer cell lines that displayed both intrinsic and acquired resistance to an *KRAS*^{G12C} inhibitor had undergone EMT (40). In addition, EMT is thought to play an important role in tumor cell plasticity. Reports showed that the transition from an epithelial to a mesenchymal tumor can stimulate chemoresistance (41, 42). In the HER2/neu-inducible breast cancer mouse model, recurrent tumors displayed an EMT-signature (43) and an EMT shift has been observed in a portion lung cancer patient samples that acquired resistance to EGFR inhibitors (44). Together, our data and published studies suggest that tumor cell plasticity - changing a cell's phenotype from an epithelial to a mesenchymal phenotype - fuels tumor cell with a mechanism to escape suppression (45).

In accordance with what we observed in our mouse model, activating mutations and amplification of MAPK/ERK signaling are also observed in patients that acquired resistance to *KRAS*^{G12C} inhibitors (46, 47). Although a direct correlation between the increased copy numbers of multiple upstream and downstream players of *Kras* and the continuation or reactivation of MAPK/ERK signaling we observed, still needs to be shown. Furthermore, upregulated pERK has also been detected in colorectal *KRAS*^{G12C}-mutant cancer cell lines treated with *KRAS*^{G12C} inhibitors, after a short period of downregulated pERK (48). The *KRAS*^{G12C} inhibitors currently tested or approved for patients, represent the first generation of *KRAS* inhibitors. Next generation inhibitors are already being developed. For example, drugs that can inhibit the active GTP-bound form of *KRAS*^{G12C} (47, 49, 50). Furthermore, combination therapy targeting other ERK/MAPK players simultaneously could benefit patients. Preclinical data have shown that combining a *KRAS*^{G12C} inhibitor with an EGFR antibody or an SHP2 inhibitor - two proteins active upstream of RAS - diminished tumor growth in *KRAS*^{G12C}-mutant cancers (48, 51).

In breast cancer, 13% of endocrine-resistant advanced breast cancer develops genetic alterations in MAPK/ERK pathway genes, including *KRAS* (52). Our data showed that also *Kras*^{G12D}-independent breast tumors displayed active MAPK/ERK signaling, this would be in line with the importance of the MAPK/ERK pathway for tumor growth.

The phenomenon of tumors with active MAPK/ERK signaling as a mechanism of adaptation, has been reported for other MAPK/ERK cascade players too. For example, colon cancer patients with the *BRAF*^{V600E} mutation showed limited response to *BRAF*^{V600E} inhibitors. *BRAF*^{V600E}-mutated colorectal cell lines showed that feedback mechanisms upregulate upstream EGFR signaling (53, 54). Another mechanism of adaptation is found in defiance of MEK1/2 inhibitors, inhibitor-treated colorectal and lung cancer cell lines displayed intrachromosomal amplification of mutant *KRAS* or *BRAF* or upregulation of upstream RTK signaling (55, 56). In case of *KRAS*^{G12C} inhibitors, *KRAS*^{G12C}-mutated lung cancer cell lines initially entered quiescence. However, a subset of tumor cells adapted and quickly resumes proliferation (57). In mutant *KRAS* pancreatic ductal adenocarcinoma cell lines, *KRAS* inhibition was well tolerated by the tumor (58). *KRAS*^{G12C} drug resistant cells displayed both active MAPK/ERK signaling and active PI3K-AKT-mTOR signaling, a second pathway that can be activated by *KRAS* (40, 59). These examples suggest that targeting multiple components of the MAPK/ERK pathway may provide a strategy for eliminating or at least delaying tumor adaptation. In the clinic, this strategy has already proven to be successful for some patients with *BRAF*^{V600}-mutant metastatic melanoma as therapy combining a MEK1/2 inhibitor with a *BRAF*^{V600E} inhibitor was successfully applied (60). We endeavored to test if inhibiting multiple MAPK/ERK players simultaneously can prevent RTs from growing in vitro. Unfortunately, the RT cells did not form organoids in vitro, hence we did not succeed in testing this hypothesis. Perhaps this is due to the mesenchymal phenotype of RTs, colorectal cancer samples with a mesenchymal phenotype fail to form organoids as well (61).

Another observation was the high genomic instability of the PTs and RTs. Genomic instability is an important hallmark of cancer. In human cancers, genomic instability occurs in 88% of tumors and correlates with TP53 mutations (62). In breast cancer, aneuploidy is correlated with poor clinical outcome and TP53 mutations are ubiquitous in aneuploid breast tumors (63-65). Breast cancer mouse models that do not directly act through p53 inhibition, displayed few or no CNAs (66). In addition, *Kras*^{G12D} lung and pancreas tumor mouse models also developed widespread CNAs (67). Thus, mouse models - such as the *Trp53* KO *iKras*^{G12D} model described here - that develop spontaneous tumors with CNAs provide an important model for human cancer. In line with other *Trp53* deficient breast cancer models (66), both the *Trp53* KO *iKras*^{G12D} primary and reactivated tumors presented with high rates of genomic instability.

The present study aimed to understand what mechanisms tumor cells deploy to adapt to mutant *Kras* inactivation in vivo. In our novel breast cancer mouse model, we found that once *Kras*^{G12D} expression was inactivated, tumors decreased and entered remission, followed by reactivated tumor growth. RT cells had transitioned from an epithelial to a mesenchymal phenotype and displayed active MAPK/ERK signaling. Tumors presented with high genomic instability and RTs showed amplifications in multiple genes associated with MAPK/ERK signaling. Our data suggests that the genomic instability contributes to emergence of mutated *Kras*-independence via amplifications of various components of the MAPK/ERK signaling pathway. These findings are relevant to therapeutics targeting RAS in human clinical trials.

METHODS

Animal care and use

Mice purchased from The Jackson Laboratory (Ben Harbor, ME, USA): Trp53^{fl/fl} (Stock #008462), Krt14-Cre (Stock #004782), mT/mG (Stock #007676), C57BL/6J (#000664). The *MMTV-rtTA* and *TetO-Kras^{G12D}* mice have been described (10, 11). All mice were backcrossed into the C57/BL6J background for at least six generations. Only female, adult mice were included in the experiments. Dox was administered through dox-impregnated food pellets (625 mg/kg, catalog #TD.01306, Envigo, Indianapolis, IN, USA). All mice used for this study were maintained at the Stanford Veterinary Service Center in accordance with the guidelines of the Administrative Panel on Laboratory Animal Care (APLAC #10868).

Tissue processing

Mice bearing tumors were euthanized and tumors were resected. The Fluorescent status of tumors was confirmed under the microscope. For immunohistochemistry, small chunks were saved in 4% PFA exchanged for 70% EtOH the next day. Remainder of the tumor was mechanically dissociated and digested in DMEM/F12 (catalog #12634028, Gibco, Langley, OK, USA) with collagenase and hyaluronidase (catalog #07912, STEMCELL Technologies Inc., Vancouver, Canada) and DNase I (catalog #LS002139, Worthington, Columbus, OH, USA) for 2.5 h with gentle pipetting every 30 min. The remainder of the digestion protocol has previously been described elsewhere (16, 68). Cells were stained for flow cytometry or resuspended with 50% Matrigel (catalog #356234, Corning, Corning, NY) prior to injection into syngeneic recipient mice.

FACS analysis and cell sorting

For FACS analysis and sorting, tumor single cells were stained with CD45 (clone #30-F11, Biolegend, San Diego, CA), CD31 (clone #390, Biolegend), TER-119 (clone #TER119, Biolegend), EpCAM (clone #G8.8, Biolegend), CD24 (clone #M1/69, Biolegend). Debris and cell doublets were excluded using side scatter and forward scatter profiles (area and width), dead cells were excluded using DAPI (catalog #32670, Sigma, St. Louis, MO, USA). For RT qPCR or RNA-seq, cells were directly sorted into RNeasy Protect (catalog #76526, Qiagen, Germantown, MD). To collect DNA (either for PCR or for WES), cell pellets were frozen down at -80 °C. FACS data was analyzed in FlowJo (v10).

Tumor injection and volume measurements in vivo

Cells were injected into the fourth abdominal fat pad by subcutaneous injection at the base of the nipple of female C57BL/6J mice. Mice were anesthetized by a constant flow of oxygen and 2% isoflurane during procedure. Tumor size was measured by caliper. Tumor volume was calculated with the formula: $\frac{4}{3} \times \pi \times (h \times w^2) / 8$, wherein h = height and w = width (69). Tumors that had increased in size on at least two consecutive occasions were considered to have reoccurred. Data was visualized in GraphPad Prism (v8).

Histology: H&E, IHC, quantification

Tissue processing, embedding, and staining with hematoxylin and eosin was done by Stanford's Comparative Medicine Animal Histology Services. Images were acquired

Breast cancer adaptive resistance to mutant *Kras*

by the BZ-X800 fluorescent microscope (Keyence, IL, USA). For IHC, Sections were deparaffinized, dehydrated, and microwaved for 20 min at 95 °C in Sodium Citrate Buffer (10mM Sodium Citrate, 0.05% Tween 20, pH 6.0) for antigen retrieval. Tissue sections were incubated overnight at 4 °C with primary antibodies diluted in phosphate buffered saline (PBS) + 5% goat serum; anti-phospho-ERK1/2 (catalog# 4370S, Cell Signaling Technology, Danvers, MA, USA) at a 1:100 dilution and anti-GFP antibody (catalog# ab13970, Abcam, Cambridge, MA, USA) at a 1:500 dilution. Samples were subsequently washed with PBS and were incubated with secondary antibodies diluted in PBS + 5% goat serum for 1h at RT; goat anti-chicken alexa fluor 488 (catalog # A-11039, Invitrogen, Waltham, MA, USA) and donkey anti-rabbit alexa fluor 594 (catalog # R37119, Invitrogen), both diluted at 1:500. All the immunofluorescence sections and cells were mounted in ProLong Gold with DAPI (catalog# P36931, ThermoFisher, Waltham, MA, USA). Images were acquired with the 20× magnification objective by the LSM 710 Meta confocal microscope (Carl Zeiss, Göttingen, Germany) or by the 10× or 20× magnification objectives BZ-X800 fluorescent microscope (Keyence, Itasca, IL, USA). Images were processed and quantified using ImageJ. 10× or 20× objective images were analyzed for total count of positive staining and intensity. Threshold was set standardized to negative staining controls at 20%. Images were converted to 8-bit for binary analysis of Intensity (using the Analyze -> Measure function). For counts of positive staining, nuclei counts were first established using the Analyze -> Analyze particles function on the DAPI channel. Particle count was set to diameter 1µm < x < 10µm to identify nuclei and rule out background staining. The Voronoi function was used on the staining channels (GFP and RFP) to isolate staining per cell and then the Analyze -> Measure function was applied. The value for mean intensity was calculated using this ratio, and also used to establish positive counts. N=3 technical replicates were considered per experiment for statistical analysis and validated in another independent biological experiment. Significance for statistical analysis was set at p < 0.05.

RT qPCR

Taking samples stored in RNAprotect, RNA was extracted using the RNeasy micro kit (catalog #74004, Qiagen), according to manufacturer's instructions. For reverse transcription to cDNA the SuperScript III First Strand Synthesis kit (catalog #18080051, Invitrogen) was used, according to the manufacturer's instructions. After 20 rounds of preamplification with Sybr Green master mix (catalog #4364346, Applied Biosystems, Waltham, MA, USA) and Sybergreen primers for *Kras* and *Kras*^{G12D}, real-time PCR was done using the 7900HT Real Time PCR system (Applied Biosystems). Data was analyzed by SDS2.4 software and GraphPad Prism (v8). All data was normalized to *Gapdh*. Sybr Green primers: *Kras* forward: GCAGGGTTGGCCTT ACA T; *Kras* reverse: ATGCGTCGCCACATTGAAT; *Kras*^{G12D} forward: CAAGGACAAGGTGTACAGTTATGTGACT; *Kras*^{G12D} reverse: GGCATCTGCTCCTGCTTTTG; *Gapdh* forward: AGGTCGGTGTGAACGGATTG *Gapdh* reverse: TGTAGACCATGTAGTTGAGGTCA (10).

RNA sequencing

Taking samples stored in RNAprotect, RNA was extracted using the RNeasy micro kit (catalog #74004, Qiagen), according to manufacturer's instructions. At least 300 ng RNA was shared for input. Library preparation, sequencing, and initial quality

Chapter 3

control was performed by Novogene (Beijing, China). Briefly, mRNA was enriched using oligo(dT) beads followed by library generation using the NEBNext Ultra™ RNA Library Prep Kit for Illumina (NEB), according to manufacturer's instructions. To allow for sequencing different samples at the same time, adaptor sequences were added to each library. Libraries were sequenced on the NovaSeq 6000 (Illumina, San Diego, CA, USA) generating 150-bp paired-end reads. Per sample, 6 GB of raw data was generated.

RNA sequencing alignment and processing

The sequencing data were uploaded to the Galaxy web platform (69), and we used the public server at usegalaxy.org to analyze the data, following the transcriptomics tutorial provided (71-73). Briefly, sequence quality control was done multiple times using FASTQC (Babraham Institute) and MultiQC (74). Trim Galore! (Babraham Institute) to trim the reads, HISAT2 to map reads to the mm10 reference genome (75), featureCounts to count the aligned reads (76), limma-voom to filter out lowly expressed genes and provide a list of differentially expressed genes and normalized counts (77).

GSEA, ssGSEA, and visualisation RNA-seq data

Patterns in gene expression were analyzed using GSEA v4.0.1 (78, 79) with the Hallmark Gene sets (80). ssGSEA (v10.0.3) analysis (79, 81) was done through GenePattern (82), using gene sets created by Hollern et al. (17). Figures were created with R (4.0.0-4.0.4) in RStudio (1.3-1.4) using packages gplots, ggplots, ggplots2, heatmap.plus, dplyr, forcats, and svglite.

Whole exome sequencing

For tumor samples, FACS-sorted tumor cells were used. For normal matched controls, tail DNA was used. To isolate genomic DNA (gDNA), cells/tails were digested at 65 °C overnight in 200 ul DirectPCR Lysis Reagent (catalog# 102-T, Viagen Biotech, Los Angeles, CA, USA) mixed with 0.5 ul Proteinase K (catalog# P8107S, NEB, Ipswich, MA, USA). Lysis was inactivated at 95°C for 10 min. gDNA was isolated using the Genomic DNA Clean & Concentrator Kit (catalog #4065D, Zymo Research, Irvine, CA, USA), according to manufacturer's instructions. To obtain sufficient DNA for WES, whole genome amplification was done using the REPLI-g mini kit (catalog# 150023, Qiagen), according to manufacturer's instructions. At least 1 µg of gDNA was shared for input. Library preparation, sequencing, and initial quality control was performed by Novogene. Briefly, gDNA was randomly fragmented by sonication (Covaris, Woburn, MA, USA) to DNA fragments of 180-280 bp followed by library generation and exome capture using SureSelectXT Mouse All Exon (Agilent, Santa Clara, CA, USA), according to manufacturer's instructions. Purification was done using AMPure XP (Beckman Coulter, Brea, CA, USA), according to manufacturer's instructions. Quantification was done with the High Sensitivity DNA assay (Agilent) on the Bioanalyzer 2100 (Agilent), according to manufacturer's instructions. Libraries were sequenced on the NovaSeq 6000 (Illumina) generating 150-bp paired-end reads. Per sample, 10 GB of raw data was generated.

WES alignment and processing

The sequencing data were uploaded to the Galaxy web platform (70), and we used

the public server at usegalaxy.org to check the quality of the data using FASTQC (Babraham Institute) and MultiQC (73). The remainder of the analysis was done on the Stanford Sherlock cluster. Trim Galore! (Babraham Institute) to trim the reads, reads were aligned to the mm10 reference genome with BWA-MEM (83), duplicates were marked and removed by MarkDuplicates (Picard, GATK 4.1.4.1) (84).

Mutation calling and analysis

Mutect2 called mutations, FilterMutectCalls marked filtered calls, and SelectVariants filtered the output (GATK 4.1.4.1) (84). Effects of variants were determined using Ensembl's VEP v101 (85), converted to maf using vcf2maf v1.6.19 (86), and summarized using maftools v0.9.30 (87).

CNA analysis

Copy number analysis was performed on WES data of 10 tumor samples using Sequenza 3.0.0 software package (88), including matched normal tissue to improve the specificity of results. Cellularity and ploidy values were estimated, and somatic CNAs were detected and visualized per sample by calculating the depth ratio (\log_2 ratio) of each segment. Thresholds for determining CNA state for copy number gains/losses were defined as ± 0.25 and amplifications/deletions ± 1 , with p-value threshold being 0.05, and the results were visualized using copynumber package 1.30.0 in R (89). Copy number data are shown in Table S3. Annotation of recurrent CNAs was performed with a custom script in R (4.0.4) using biomaRt (2.46.3) (90).

PCR

gDNA was isolated as for WES. PCR primers used to determine if *Trp53* was floxed out or not: p53-int1-fwd: CACAAAAACAGGTTAAACCCA; p53-int10-fwd AAGGGGTATGAGGGACAAGG; p53-int10-rev: GAAGACAGAAAAGGGGAGGG (91).

Data availability

Human breast cancer data was accessed through cBioportal.org (92, 93) on April 27, 2021, and included data with both mutational and CNA data (n=5087) from the TCGA PanCancer atlas, the METABRIC study, and MSKCC (52, 94-96). RNA-seq data that support the findings in this study are openly available in the Gene Expression Omnibus at NCBI (www.ncbi.nlm.nih.gov/geo/) under accession number GSE174441. The nucleotide sequence data that support the findings in this study are openly available in the Sequence Read Archive at NCBI (www.ncbi.nlm.nih.gov/sra) under accession number PRJNA730907.

AUTHOR CONTRIBUTIONS

L.J.v.W., S.C., F.A.S., and M.F.C. conceived the study and designed the experiments. L.J.v.W., S.I.D., and M.F.C. wrote the manuscript. L.J.v.W. performed and analyzed most of the experiments. S.I.D. analyzed the CNA data. S.S.S. performed the fluorescence IHC. J.A. quantified the fluorescence IHC images. D.Q. and W.H.D.H. aided with mouse genotyping and tissue processing. R.W. analyzed the H&E histology data.

ACKNOWLEDGEMENTS

This work was supported by Breast Cancer Research Foundation grant BCRF-18-027

Chapter 3

(to M.F.C.), Department of Defense grant W81XWH-13-1-0281 (to M.F.C.), Ludwig Cancer Research Foundation (to M.F.C.), and the CIRM Bridges award (to W.H.D.H.). Some research was performed on a FACSriaal that was purchased using NIH S10 shared instrumentation grant (1S10RR02933801). We thank Dr. Podsypanina for sharing the MTB-TOR mouse; the Stanford Veterinary Service Center for providing mouse care; Sopheak Sim, Dr. Tomer Kalisky, and Dr. Rosalynn Upton for help with preliminary experiments; Dr. Zoe June Assaf for advice on the WES data analysis; Catherine Carswell Crumpton and Cheng Pan for management of the flow cytometry facility; Doreen Wu in the Stanford Animal Histology Services for help with preparation of histologic specimens; Margaret Cuadro for administrative assistance; the Stanford Sherlock cluster for offering the environment to do bioinformatic analysis.

REFERENCES

1. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*. 2018;173(2):371-85.e18.
2. Nguyen LV, Pellacani D, Lefort S, Kannan N, Osako T, Makarem M, et al. Barcoding reveals complex clonal dynamics of de novo transformed human mammary cells. *Nature*. 2015;528(7581):267-71.
3. Network CGA. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70.
4. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol*. 2007;8(5):R76.
5. Howlader N, Cronin KA, Kurian AW, Andridge R. Differences in Breast Cancer Survival by Molecular Subtypes in the United States. *Cancer Epidemiol Biomarkers Prev*. 2018;27(6):619-26.
6. Hallin J, Engstrom LD, Hargis L, Calinisan A, Aranda R, Briere DM, et al. The KRAS G12C Inhibitor MRTX849 Provides Insight toward Therapeutic Susceptibility of KRAS-Mutant Cancers in Mouse Models and Patients. *Cancer Discov*. 2020;10(1):54-71.
7. Hong DS, Fakih MG, Strickler JH, Desai J, Durm GA, Shapiro GI, et al. KRAS G12C Inhibition with Sotorasib in Advanced Solid Tumors. *N Engl J Med*. 2020;383(13):1207-17.
8. Salgia R, Pharaon R, Mambetsariev I, Nam A, Sattler M. The improbable targeted therapy: KRAS as an emerging target in non-small cell lung cancer (NSCLC). *Cell Rep Med*. 2021;2(1):100186.
9. Skoulidis F, Li BT, Dy GK, Price TJ, Falchook GS, Wolf J, et al. Sotorasib for Lung Cancers with KRAS p.G12C Mutation. *N Engl J Med*. 2021;384(25):2371-81.
10. Fisher GH, Wellen SL, Klimstra D, Lenczowski JM, Tichelaar JW, Lizak MJ, et al. Induction and apoptotic regression of lung adenocarcinomas by regulation of a K-Ras transgene in the presence and absence of tumor suppressor genes. *Genes Dev*. 2001;15(24):3249-62.
11. Gunther EJ, Belka GK, Wertheim GB, Wang J, Hartman JL, Boxer RB, et al. A novel doxycycline-inducible system for the transgenic analysis of mammary gland biology. *FASEB J*. 2002;16(3):283-92.
12. Podsypanina K, Politi K, Beverly LJ, Varmus HE. Oncogene cooperation in tumor maintenance and tumor recurrence in mouse mammary tumors induced by Myc and mutant Kras. *Proc Natl Acad Sci U S A*. 2008;105(13):5242-7.
13. Jonkers J, Meuwissen R, van der Gulden H, Peterse H, van der Valk M, Berns A. Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nat Genet*. 2001;29(4):418-25.
14. Dassule HR, Lewis P, Bei M, Maas R, McMahon AP. Sonic hedgehog regulates growth and morphogenesis of the tooth. *Development*. 2000;127(22):4775-85.
15. Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L. A global double-fluorescent Cre reporter mouse. *Genesis*. 2007;45(9):593-605.
16. van Weele LJ, Scheeren FA, Cai S, Kuo AH, Qian D, Ho WHD, et al. Depletion of Trp53 and Cdkn2a Does Not Promote Self-Renewal in the Mammary Gland but Amplifies Proliferation Induced by TNF- α . *Stem Cell Reports*. 2021;16(2):228-36.
17. Hollern DP, Swiatnicki MR, Andrechek ER. Histological subtypes of mouse mammary tumors reveal conserved relationships to human cancers. *PLoS Genet*. 2018;14(1):e1007135.
18. Wagle MC, Kirouac D, Klijn C, Liu B, Mahajan S, Junttila M, et al. A transcriptional MAPK Pathway Activity Score (MPAS) is a clinically relevant biomarker in multiple cancer types. *NPJ Precis Oncol*. 2018;2(1):7.
19. Guerra C, Mijimolle N, Dhawahir A, Dubus P, Barradas M, Serrano M, et al. Tumor induction by an endogenous K-ras oncogene is highly dependent on cellular context. *Cancer Cell*. 2003;4(2):111-20.
20. Tuveson DA, Shaw AT, Willis NA, Silver DP,

Breast cancer adaptive resistance to mutant Kras

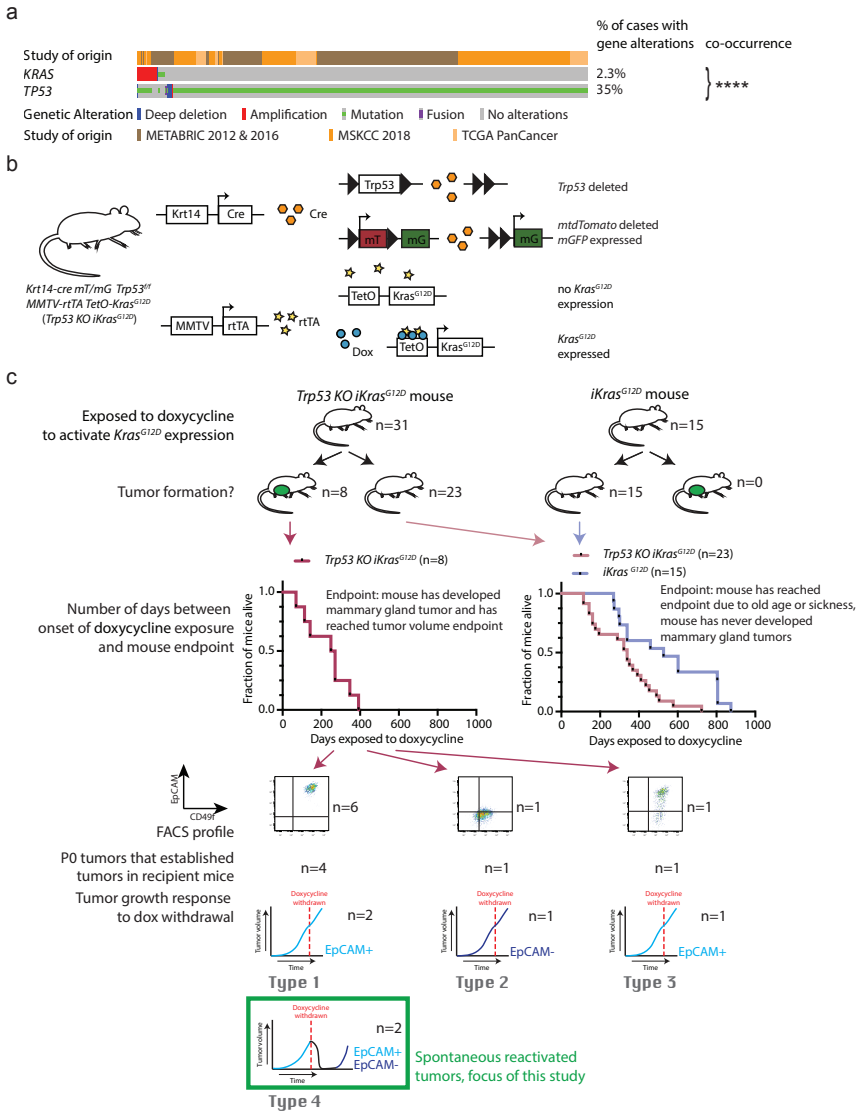
- Jackson EL, Chang S, et al. Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. *Cancer Cell*. 2004;5(4):375-87.
21. Braig M, Lee S, Loddenkemper C, Rudolph C, Peters AH, Schlegelberger B, et al. Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature*. 2005;436(7051):660-5.
 22. Sarkisian CJ, Keister BA, Stairs DB, Boxer RB, Moody SE, Chodosh LA. Dose-dependent oncogene-induced senescence in vivo and its evasion during mammary tumorigenesis. *Nat Cell Biol*. 2007;9(5):493-505.
 23. Rädler PD, Wehde BL, Triplett AA, Shrestha H, Shepherd JH, Pfefferle AD, et al. Highly metastatic claudin-low mammary cancers can originate from luminal epithelial cells. *Nat Commun*. 2021;12(1):3742.
 24. Heston WE, Vlahakis G. Mammary tumors, plaques, and hyperplastic alveolar nodules in various combinations of mouse inbred strains and the different lines of the mammary tumor virus. *Int J Cancer*. 1971;7(1):141-8.
 25. Lifsted T, Le Voyer T, Williams M, Muller W, Klein-Szanto A, Buetow KH, et al. Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. *Int J Cancer*. 1998;77(4):640-4.
 26. Mahler JF, Stokes W, Mann PC, Takaoka M, Maronpot RR. Spontaneous lesions in aging FVB/N mice. *Toxicol Pathol*. 1996;24(6):710-6.
 27. Regua AT, Arrigo A, Doheny D, Wong GL, Lo HW. Transgenic mouse models of breast cancer. *Cancer Lett*. 2021;516:73-83.
 28. Rittinghausen S, Dungworth DL, Ernst H, Mohr U. Naturally occurring pulmonary tumors in rodents. *Respiratory System*: Springer; 1996. p. 183-206.
 29. Guy CT, Cardiff RD, Muller WJ. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol*. 1992;12(3):954-61.
 30. Davie SA, Maglione JE, Manner CK, Young D, Cardiff RD, MacLeod CL, et al. Effects of FVB/NJ and C57Bl/6J strain backgrounds on mammary tumor phenotype in inducible nitric oxide synthase deficient mice. *Transgenic Res*. 2007;16(2):193-201.
 31. Sarsani VK, Raghupathy N, Fiddes IT, Armstrong J, Thibaud-Nissen F, Zinder O, et al. The Genome of C57Bl/6J "Eve", the Mother of the Laboratory Mouse Genome Reference Strain. G3 (Bethesda). 2019;9(6):1795-805.
 32. Denaro M, Oldmixon B, Patience C, Andersson G, Down J. EGFP-transduced EL-4 cells from tumors in C57BL/6 mice. *Gene Ther*. 2001;8(23):1814-5.
 33. Skelton D, Satake N, Kohn DB. The enhanced green fluorescent protein (eGFP) is minimally immunogenic in C57BL/6 mice. *Gene Ther*. 2001;8(23):1813-4.
 34. Gossa S, Nayak D, Zinselmeyer BH, McGavern DB. Development of an immunologically tolerated combination of fluorescent proteins for in vivo two-photon imaging. *Sci Rep*. 2014;4:6664.
 35. Juneja VR, McGuire KA, Manguso RT, LaFleur MW, Collins N, Haining WN, et al. PD-L1 on tumor cells is sufficient for immune evasion in immunogenic tumors and inhibits CD8 T cell cytotoxicity. *J Exp Med*. 2017;214(4):895-904.
 36. Davey GM, Mueller SN, van Vliet C, Gigowski M, Zaid A, Davies B, et al. Identification of a MHC I-restricted epitope of DsRed in C57BL/6 mice. *Mol Immunol*. 2013;53(4):450-2.
 37. Ansari AM, Ahmed AK, Matsangos AE, Lay F, Born LJ, Marti G, et al. Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments. *Stem Cell Rev Rep*. 2016;12(5):553-9.
 38. Morel AP, Ginestier C, Pommier RM, Cabaud O, Ruiz E, Wicinski J, et al. A stemness-related ZEB1-MSRB3 axis governs cellular pliancy and breast cancer genome stability. *Nat Med*. 2017;23(5):568-78.
 39. Singh A, Greninger P, Rhodes D, Koopman L, Violette S, Bardeesy N, et al. A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival. *Cancer Cell*. 2009;15(6):489-500.
 40. Adachi Y, Ito K, Hayashi Y, Kimura R, Tan TZ, Yamaguchi R, et al. Epithelial-to-mesenchymal Transition is a Cause of Both Intrinsic and Acquired Resistance to KRAS G12C Inhibitor in KRAS G12C-Mutant Non-Small Cell Lung Cancer. *Clin Cancer Res*. 2020;26(22):5962-73.
 41. Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature*. 2015;527(7579):525-30.
 42. Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong ST, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature*. 2015;527(7579):472-6.
 43. Moody SE, Perez D, Pan TC, Sarkisian CJ, Portocarrero CP, Sterner CJ, et al. The transcriptional repressor Snail promotes mammary tumor recurrence. *Cancer Cell*. 2005;8(3):197-209.
 44. Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med*. 2011;3(75):75ra26.
 45. Tulchinsky E, Demidov O, Kriajevska M, Barlev NA, Imyanitov E. EMT: A mechanism for escape from EGFR-targeted therapy in lung cancer. *Biochim Biophys Acta Rev Cancer*. 2019;1871(1):29-39.

Chapter 3

46. Awad MM, Liu S, Rybkin II, Arbour KC, Dilly J, Zhu WW, et al. Acquired Resistance to KRAS. *N Engl J Med.* 2021;384(25):2382-93.
47. Tanaka N, Lin JJ, Li C, Ryan MB, Zhang J, Kiedrowski LA, et al. Clinical Acquired Resistance to KRAS. *Cancer Discov.* 2021;11(8):1913-22.
48. Amodio V, Yaeger R, Arcella P, Cancelliere C, Lamba S, Lorenzato A, et al. EGFR Blockade Reverts Resistance to KRAS. *Cancer Discov.* 2020;10(8):1129-39.
49. Schulze CJ, Bermingham A, Choy TJ, Cregg JJ, Kiss G, Marquez A, et al., editors. Tri-complex inhibitors of the oncogenic, GTP-bound form of KRAS (G12c) overcome RTK-mediated escape mechanisms and drive tumor regressions in vivo. *MOLECULAR CANCER THERAPEUTICS*; 2019: AMER ASSOC CANCER RESEARCH 615 CHESTNUT ST, 17TH FLOOR, PHILADELPHIA, PA
50. Nichols R, Schulze C, Bermingham A, Choy T, Cregg J, Kiss G, et al. A06 tri-complex inhibitors of the oncogenic, GTP-bound form of KRASG12C overcome RTK-mediated escape mechanisms and drive tumor regressions in preclinical models of NSCLC. *Journal of Thoracic Oncology.* 2020;15(2):S13-S4.
51. Ryan MB, Fecce de la Cruz F, Phat S, Myers DT, Wong E, Shahzade HA, et al. Vertical Pathway Inhibition Overcomes Adaptive Feedback Resistance to KRAS. *Clin Cancer Res.* 2020;26(7):1633-43.
52. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasani N, et al. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell.* 2018;34(3):427-38.e6.
53. Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, et al. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature.* 2012;483(7387):100-3.
54. Corcoran RB, Ebi H, Turke AB, Coffee EM, Nishino M, Cogdill AP, et al. EGFR-mediated re-activation of MAPK signaling contributes to insensitivity of BRAF mutant colorectal cancers to RAF inhibition with vemurafenib. *Cancer Discov.* 2012;2(3):227-35.
55. Little AS, Balmanno K, Sale MJ, Newman S, Dry JR, Hampson M, et al. Amplification of the driving oncogene, KRAS or BRAF, underpins acquired resistance to MEK1/2 inhibitors in colorectal cancer cells. *Sci Signal.* 2011;4(166):ra17.
56. Kitai H, Ebi H, Tomida S, Floros KV, Kotani H, Adachi Y, et al. Epithelial-to-Mesenchymal Transition Defines Feedback Activation of Receptor Tyrosine Kinase Signaling Induced by MEK Inhibition in KRAS-Mutant Lung Cancer. *Cancer Discov.* 2016;6(7):754-69.
57. Xue JY, Zhao Y, Aronowitz J, Mai TT, Vides A, Qeriqi B, et al. Rapid non-uniform adaptation to conformation-specific KRAS(G12C) inhibition. *Nature.* 2020;577(7790):421-5.
58. Chen JK, Taipale J, Cooper MK, Beachy PA. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev.* 2002;16(21):2743-8.
59. Misale S, Fatherree JP, Cortez E, Li C, Bilton S, Timonina D, et al. KRAS G12C NSCLC Models Are Sensitive to Direct Targeting of KRAS in Combination with PI3K Inhibition. *Clin Cancer Res.* 2019;25(2):796-807.
60. Long GV, Flaherty KT, Stroyakovskiy D, Gogas H, Levchenko E, de Braud F, et al. Dabrafenib plus trametinib versus dabrafenib monotherapy in patients with metastatic BRAF V600E/K-mutant melanoma: long-term survival and safety analysis of a phase 3 study. *Ann Oncol.* 2017;28(7):1631-9.
61. Koppens MA, Bounova G, Cornelissen-Steijger P, de Vries N, Sansom OJ, Wessels LF, et al. Large variety in a panel of human colon cancer organoids in response to EZH2 inhibition. *Oncotarget.* 2016;7(43):69816-28.
62. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell.* 2018;33(4):676-89.e3.
63. Xu J, Huang L, Li J. DNA aneuploidy and breast cancer: a meta-analysis of 141,163 cases. *Oncotarget.* 2016;7(37):60218-29.
64. Pfister K, Pipka JL, Chiang C, Liu Y, Clark RA, Keller R, et al. Identification of Drivers of Aneuploidy in Breast Tumors. *Cell Rep.* 2018;23(9):2758-69.
65. Minussi DC, Nicholson MD, Ye H, Davis A, Wang K, Baker T, et al. Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature.* 2021;592(7853):302-8.
66. Ben-David U, Ha G, Khadka P, Jin X, Wong B, Franke L, et al. The landscape of chromosomal aberrations in breast cancer mouse models reveals driver-specific routes to tumorigenesis. *Nat Commun.* 2016;7:12160.
67. Chung WJ, Daemen A, Cheng JH, Long JE, Cooper JE, Wang BE, et al. mutant genetically engineered mouse models of human cancers are genomically heterogeneous. *Proc Natl Acad Sci U S A.* 2017;114(51):E10947-E55.
68. Prater M, Shehata M, Watson CJ, Stingl J. Enzymatic dissociation, flow cytometric analysis, and culture of normal mouse mammary tissue. *Methods Mol Biol.* 2013;946:395-409.
69. Westphal K, Leschner S, Jablonska J, Loessner H, Weiss S. Containment of tumor-colonizing bacteria by host neutrophils. *Cancer Res.* 2008;68(8):2952-60.
70. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 2018;46(W1):W537-W44.
71. Doyle M, Phipson B, Dashnow H. 1: RNA-Seq reads to counts (Galaxy Training Materials) 2021 [updated 20201 March 12. Available from:

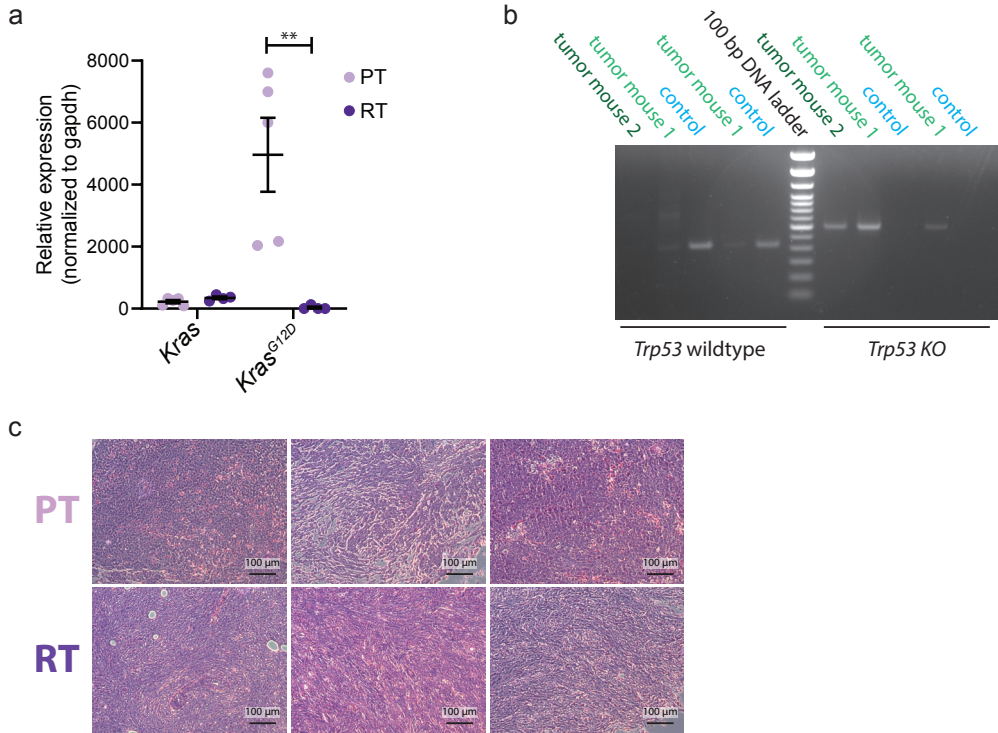
- <https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/rna-seq-reads-to-counts/tutorial.html#citing-this-tutorial>.
72. Doyle M, Phipson B, Maksimovic J, Trigos A, Ritchie M, Dashnow H, et al. 2: RNA-seq counts to genes (Galaxy Training Materials) 2021 [updated 2021 March 23. Available from: <https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/rna-seq-counts-to-genes/tutorial.html>.
 73. Batut B, Hiltmann S, Bagnacani A, Baker D, Bhardwaj V, Blank C, et al. Community-Driven Data Analysis Training for Biology. *Cell Syst*. 2018;6(6):752-8.e1.
 74. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-8.
 75. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357-60.
 76. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-30.
 77. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29.
 78. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267-73.
 79. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-50.
 80. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-25.
 81. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009;462(7269):108-12.
 82. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet*. 2006;38(5):500-1.
 83. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics. 2013.
 84. GA VdA, BD OC. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. 1st Edition ed: O'Reilly Media; 2020.
 85. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.
 86. Kandoth C. mskcc/vcf2maf: vcf2maf v1.6.19 2020 [
 87. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28(11):1747-56.
 88. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol*. 2015;26(1):64-70.
 89. Nilsen K, Liestøl K, Van Loo P, Moen Vollan HK, Eide MB, Rueda OM, et al. Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics*. 2012;13:591.
 90. Bartha Á, Györfy B. Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools Available in Clinical Oncology. *Cancers (Basel)*. 2019;11(11).
 91. Marino S, Vooijs M, van Der Gulden H, Jonkers J, Berns A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev*. 2000;14(8):994-1004.
 92. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-4.
 93. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
 94. TCGA Research Network, <https://www.cancer.gov/tcga>.
 95. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-52.
 96. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun*. 2016;7:11908.
 97. VIB / UGent Bioinformatics & Evolutionary Genomics, <http://bioinformatics.psb.ugent.be/webtools/Venn/.xs>
 98. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. *Cancer Cell*. 2018;33(4):690-705.e9.

SUPPLEMENTARY FIGURES

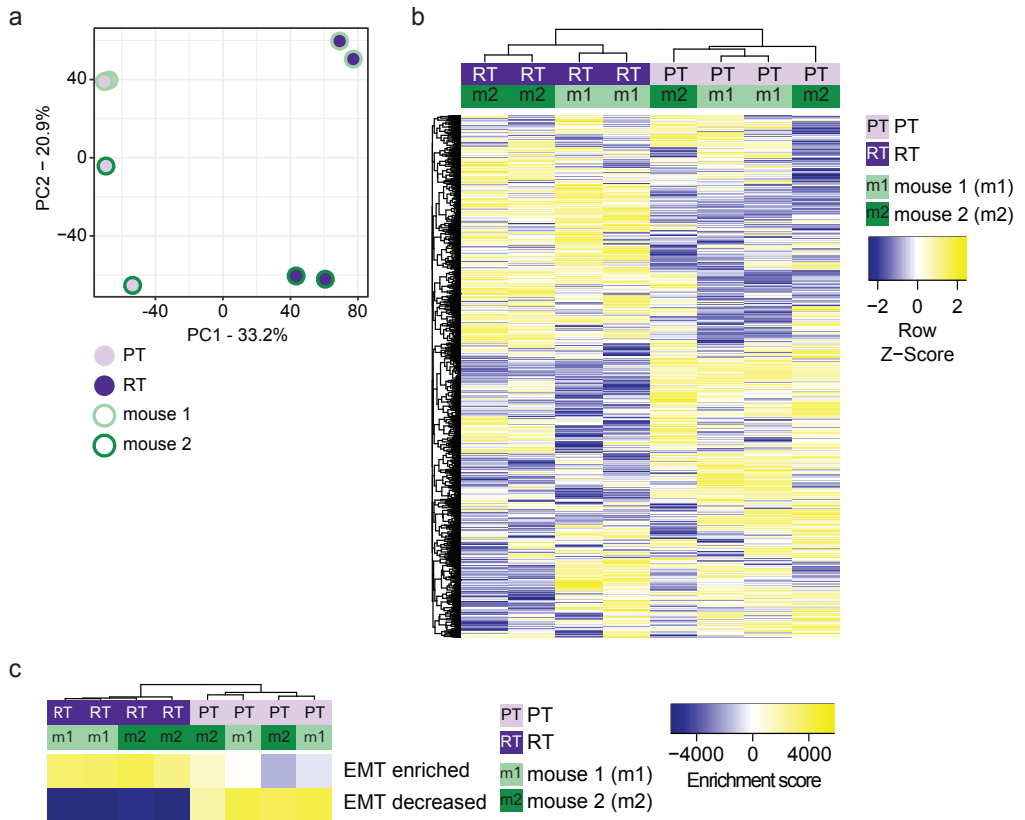


Supplementary Figure 1. Tumor formation in *Trp53 KO iKras^{G12D}* mice (a) Oncoplot of human breast cancer samples showing genetic alterations in the *KRAS* and *TP53* genes, n= 4925. Only patient samples with genetic alterations in at least one of these two genes are shown. Data is compiled from 3 studies (51, 93-95). (Co-occurrence genetic alterations in both genes q-value = < 0.001) (b) The various components of the mouse model. In cell expression *Krt14*, such as all mammary gland epithelial cells during development, *Cre* is expressed. *Cre* deletes part of the *Trp53* gene, thereby eliminating the expression of *Trp53*. *Cre* also deletes the gene coding for membrane-bound *tdTomato* and a STOP codon inhibiting membrane-bound *GFP* expression, resulting in the expression of membrane-bound *GFP*. The *MMTV* promoter is expressed in the mammary epithelium. *rtTA* can only bind to the *TetO* promoter in the presence of *dox*, resulting in the expression of *Kras^{G12D}*. Therefore, in the absence of *dox* *Kras^{G12D}* is not expressed. (c) Tumor frequency and survival plots of all *Trp53 KO iKras^{G12D}* and *iKras^{G12D}* mice in this study. FACS profiles and tumor growth profiles upon transgenic *Kras^{G12D}* activation and inactivation by *dox* in *Trp53 KO iKras^{G12D}* mice that developed mammary gland tumors. The focus of this study are tumors that initially go into remission upon *Kras^{G12D}* inactivation, followed by *Kras^{G12D}*-independent reactivation.

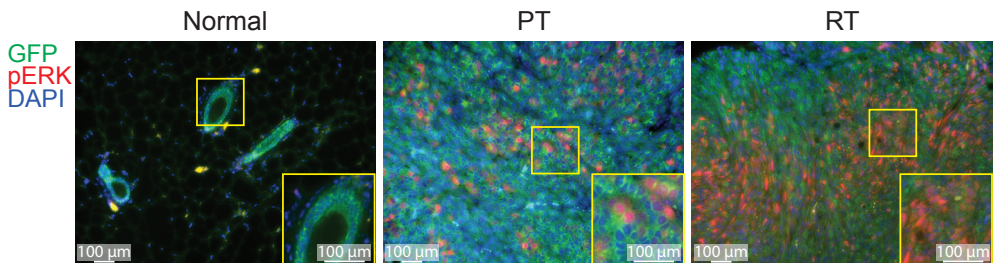
Breast cancer adaptive resistance to mutant *Kras*



Supplementary Figure 2. Verification of the *Trp53* KO *iKras^{G12D}* mouse model (a) qPCR data showing the expression of *Kras* and *Kras^{G12D}* in PTs (n=5) and in RTs (n=4). Data are represented as mean \pm SEM. (Unpaired t test, ** P <0.01) (b) PCR data confirming the presence (*Trp53* wildtype) or absence (*Trp53* KO) of the *Trp53* allele in *Trp53* KO *iKras^{G12D}* tumor cells, control is wildtype tail DNA. (c) Additional H&E staining from PT and RT tumors, in addition to slides shown in Fig. 2C. Scale bar is set at 100 μ m.

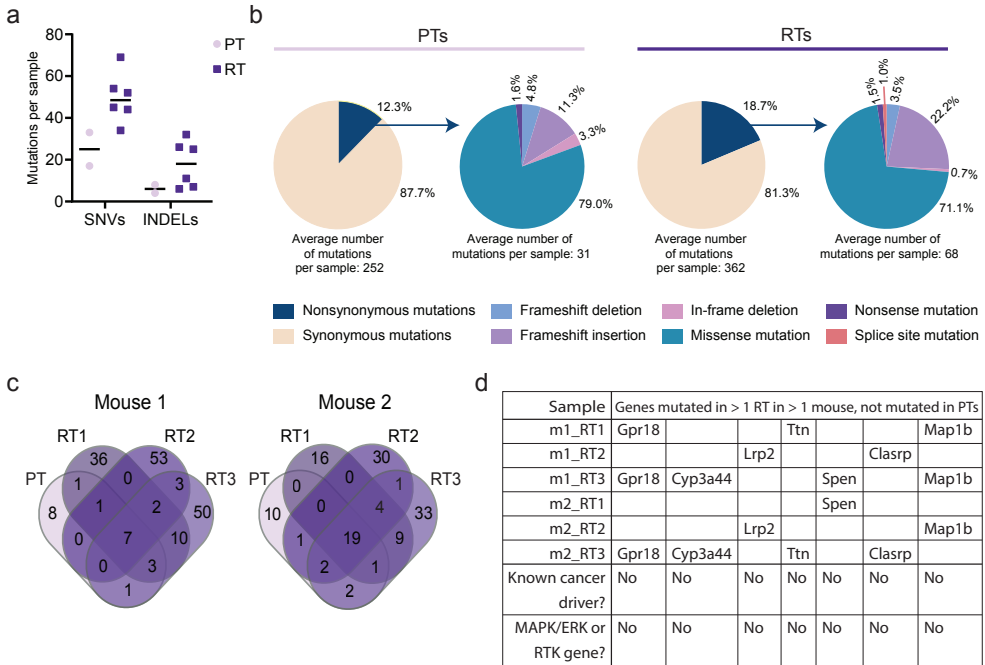


Supplementary Figure 3. RT and PT samples cluster and RNA-seq confirms EMT profile in RT tumors (a) PCA of PT (4) and RT (4) RNA-seq samples. (b) Heatmap of RNA-seq expression profiles showing all filtered and normalized counts. (c) ssGSEA gene set developed to analyze mouse tumor histology. Gene sets for EMT enriched and EMT decreased are shown. Gene sets were developed by Hollern et al. (17).

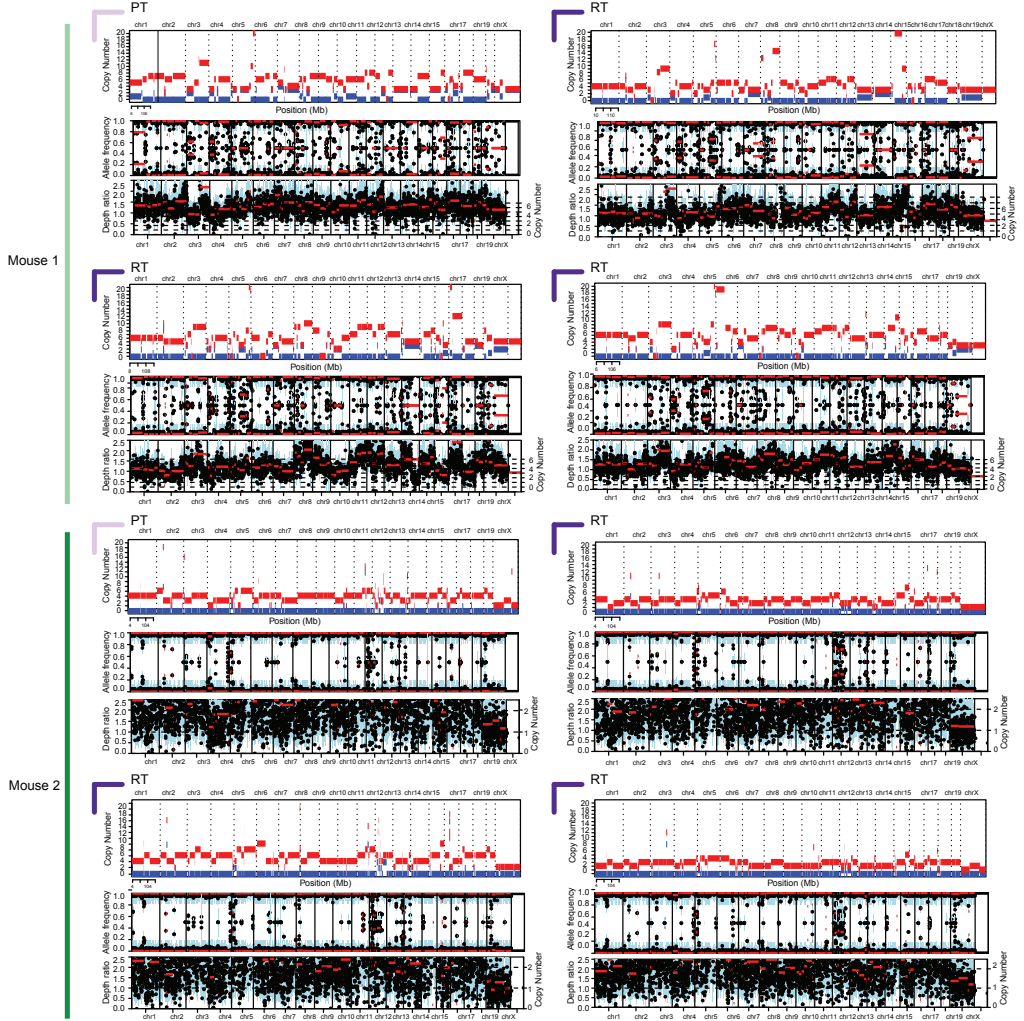


Supplementary Figure 4 RTs activates MAPK/ERK signaling, additional data Figure in addition to Fig. 5C, showing a second set of biological samples in an experiment performed separately, using a different fluorescent microscope for imaging. Immunofluorescence staining for pERK1/2 (red), GFP (green), and DAPI (blue) of tissue deriving from a normal *Krt14-Cre mT/mG* mammary gland, a PT, and a RT. Scale bar is set at 100 μm.

Breast cancer adaptive resistance to mutant Kras



Supplementary Figure 5 Analysis mutational data (a) Number of SNVs and INDELs acquired by each PT and RT. The black horizontal line shows the median. (b) Pie chart showing the average percentage of nonsynonymous and synonymous mutations in PTs and RTs and the average percentage of the different types of nonsynonymous mutations in PTs and RTs. (c) Venn diagram of overlap between genes with nonsynonymous mutations in PT and 3 matched RTs (RT1-RT3), mouse 1 and mouse 2 (Venn diagram tool: (96)). (d) Overview of genes with nonsynonymous mutations in at least 1 RT (RT1/RT2/RT3) of each mouse (m1/m2) and no mutations in either PT. The resulting 7 genes were compared to MAPK/ERK genes, RTK genes, pan-cancer, and breast cancer driver genes (1, 3, 97).



Supplementary Figure 6 Allele-specific copy number analysis for 2 PTs and 6 RTs against their matched normal counterparts Analysis was performed using the Sequenza algorithm for paired tumor-normal WES data to visualize CNAs in each sample as described by raw copy number, depth ratio (logR) and b-allele frequency profiles. Red and blue represent the overall copy number and minor allele, respectively.

Supplementary information is available online.

Table S1 Overview mutations

Table S2 Overview of genes with nonsynonymous mutations and information on their human homolog pan- and breast cancer driver status

Table S3 Unannotated CNA data

Chapter 4

A cell-intrinsic role for TLR2-MYD88 in intestinal and breast epithelia and oncogenesis

Ferenc A. Scheeren^{1, 2 *} • Angera H. Kuo^{1 *} • **Linda J. van Weele**¹ • Shang Cai¹ • Iris Glykofridis² • Shaheen S. Sikandar¹ • Maider Zabala¹ • Dalong Qian¹ • Jessica S. Lam¹ • Darius Johnston¹ • Jens P Volkmer¹ • Debashis Sahoo^{1, 8} • Matt van de Rijn⁴ • Fred Dirbas⁵ • George Somlo⁶ • Tomer Kalisky^{7, 9} • Michael E. Rothenberg¹ • Michael F. Clarke^{1, 3}

Adapted from: *Nature Cell Biology* volume 16, pages 1238–1248 (2014)

¹Stanford Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, 265 Campus Drive, Stanford, California 94305, USA. ²The Netherlands Cancer Institute, 1066 CX, Amsterdam, The Netherlands. ³Department of medicine, Division of Oncology, Stanford University, Stanford, California 94305, USA. ⁴Department of Pathology, ⁵Department of Surgery, Stanford University, Stanford, California 94305, USA. ⁶Department of Medical Oncology, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, California 91010, USA. ⁷Department of Bioengineering and Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA. ⁸Current address: Department of Pediatrics, University of California at San Diego, La Jolla, California 92093, USA. ⁹Current address: Faculty of Engineering, Bar-Ilan University, Ramat Gan 52900, Israel. *co-first authors

Contribution: L.J.v.W. performed and analyzed in vitro experiments, maintained mouse colonies, edited the manuscript

It has been postulated that there is a link between inflammation and cancer. Here we describe a role for cell-intrinsic toll-like receptor-2 (TLR2; which is involved in inflammatory response) signalling in normal intestinal and mammary epithelial cells and oncogenesis. The downstream effectors of TLR2 are expressed by normal intestinal and mammary epithelia, including the stem/progenitor cells. Deletion of MYD88 or TLR2 in the intestinal epithelium markedly reduces DSS-induced colitis regeneration and spontaneous tumour development in mice. Limiting dilution transplantations of breast epithelial cells devoid of TLR2 or MYD88 revealed a significant decrease in mammary repopulating unit frequency compared with the control. Inhibition of TLR2, its co-receptor CD14, or its downstream targets MYD88 and IRAK1 inhibits growth of human breast cancers *in vitro* and *in vivo*. These results suggest that inhibitors of the TLR2 pathway merit investigation as possible therapeutic and chemoprevention agents.

Microbial pathogens trigger inflammatory response through pattern-recognition receptors, resulting in the activation of the innate immune system (1, 2, 3, 4, 5). Toll-like receptors (TLRs) are among the best characterized pattern-recognition receptors. In addition to microbial ligands, a number of endogenous ligands, including intracellular components released after apoptosis and cell death, can stimulate TLRs. This suggests a role for TLRs other than directly recognizing and responding to microbes (2, 4, 5). CD14 is a GPI-linked protein that is found on the surface of many TLR4- and/or TLR2-expressing cells (6, 7, 8, 9). Once ligands bind to CD14, signal transduction takes place through TLR2 and TLR4.

Inflammatory response can play a role in cancer (for example, TLR2-induced inflammatory response in myeloid cells can affect tumour progression and metastasis) (2, 3, 4, 5, 6, 10, 11, 12, 13, 14). Moreover, myeloid differentiation primary response protein-88 (MYD88), an adaptor protein for TLRs and IL-1R in innate immune cells (15), is crucial for spontaneous intestinal tumour formation¹⁴ and has activating oncogenic mutations in diffuse large B cell lymphoma and chronic lymphocytic leukaemia (16, 17). Other examples linking inflammation and cancer are: chronic gastritis due to *Helicobacter pylori* infection is a major risk for gastric cancer and chronic colitis often leads to colorectal cancer (18, 19, 20, 21).

Transplantation and lineage-tracing data suggested the presence of multi-potent mammary stem cells (MSCs) and progenitors and showed that a single MSC is able to reconstitute a functional mammary gland (22, 23, 24, 25, 26, 27, 28, 29, 30). MSCs reside at the top of the mammary cellular proliferation hierarchy; however, molecular pathways regulating their self-renewal are not well characterized.

The intestinal epithelium consists of long-lived stem cells (resides near the crypt bottom) with self-renewal capacity. Two types of stem cell are proposed in small intestinal crypts: cycling crypt base columnar cells and quiescent + 4 cells. As in the breast, regulators of stem cell self-renewal in intestinal epithelium are poorly understood (2, 3, 6, 8, 10, 11, 12, 13, 14, 15, 19, 21, 31, 32, 33, 34, 35, 36, 37, 38, 39). We show that TLR2/CD14 signalling through MYD88 is an important pathway for regeneration of intestinal epithelial cells after inflammation damage, regeneration of mammary epithelium, and growth of some cancers. Our data reveal that inhibitors of

the TLR pathway warrant therapeutic exploration in the treatment or chemoprevention of breast and colon cancer.

RESULTS

Epithelial-specific TLR2-MYD88 signalling in the intestinal tract

We first examine the role of TLR2 signalling in the intestine because this pathway has been linked to homeostasis of intestinal epithelial cells (39, 40). Flow cytometry analysis of intestinal cells isolated from *Lgr5*-GFP mice demonstrated that *Lgr5*-GFP^{pos}EpCAM^{pos}CD45^{neg} cells, which are enriched for intestinal stem cells, expressed slightly higher levels of TLR2 as compared with *Lgr5*-GFP^{pos}EpCAM^{pos}CD45^{neg} cells (Fig. 1a). This shows that TLR2 is expressed by the LGR5^{pos} intestinal stem cells.

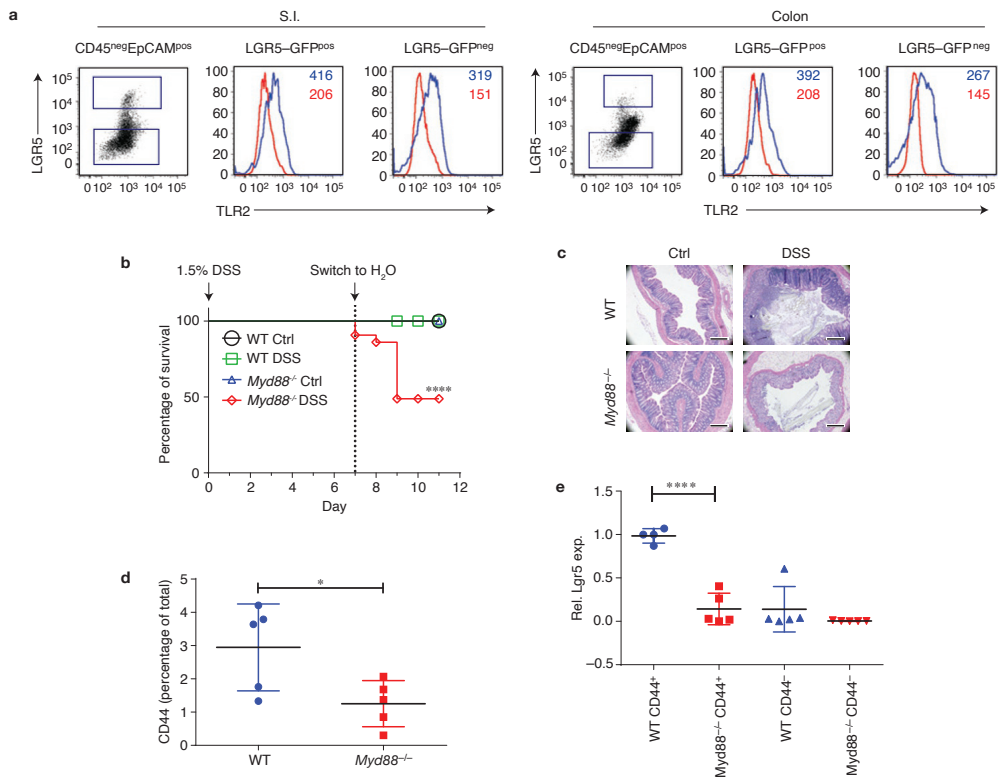


Figure 1: TLR2 and MYD88 are functional in murine intestinal epithelial cells. (a) Representative flow cytometry histogram of TLR2 on small intestine (S.I.) and colon (DAP1^{neg} CD45^{neg} EpCAM^{pos}) *Lgr5*-GFP^{pos} cells; red is the isotype control for each specific sub-population, and blue is the TLR2 staining. Median fluorescence is shown for each histogram. Data are representative of 4 mice. (b) WT and *Myd88*^{-/-} mice were given normal drinking water or 1.5% DSS in drinking water for 7 days (WT Ctrl: N = 9 mice; WT DSS: N = 28, *Myd88*^{-/-} Ctrl: N = 8; *Myd88*^{-/-} DSS: N = 43). Mice were changed to normal drinking water on day 8. Survival was monitored for 11 days. These data are combined from 3 different experiments. **** P < 0.0001. (c) Representative photo of H&E staining of colons from WT and *Myd88*^{-/-} mice at day 11 (scale bars, 0.5 mm). Data are representative of 8 mice. (d) Percentage of CD44 on colonic cells as determined by flow cytometry between DSS-treated WT (N = 4) and *Myd88*^{-/-} (N = 5) mice. * P = 0.0337. (e) qPCR for *Lgr5* expression on epithelial CD44^{pos} and CD44^{neg} sorted populations of WT (N = 4) and *Myd88*^{-/-} (N = 5) mice. **** P < 0.0001. Values represent mean ± s.d. Student's unpaired t-test for independent samples was used.

The epithelium of the intestines is constantly self-renewing and will quickly regenerate after injury. To study the regenerative capacity of TLR2 signalling in colonic epithelium, we subjected *Myd88*^{-/-} mice to dextran sodium sulphate (DSS) to induce acute colitis. In this model, administration of DSS in drinking water may result in weight loss, bloody diarrhoea and eventual death. These clinical manifestations are the results of epithelial barrier dysfunction, degeneration and necrosis with consequent recruitment and activation of inflammatory cells in the colon (41, 42, 43). Withdrawal of DSS in drinking water induces the regeneration of the colon that includes rapid crypt hyperplasia and fission. *Myd88*^{-/-} and wild-type (WT) mice were treated with 1.5% DSS for 7 days (Day 0 to Day 7) to induce acute colitis and then switched back to normal drinking water to promote regeneration (Day 8 to Day 11). Consistent with published data (40, 44), *Myd88*^{-/-} mice are more susceptible to DSS-induced colitis as exhibited by higher morbidity as compared with WT (Fig. 1b). Furthermore, DSS-treated mice from the regeneration period (Day 11) revealed that *Myd88*^{-/-} had more severe crypt loss and deterioration of crypt-villus architecture, and less crypt hyperplasia and fission (Fig. 1c). Flow cytometry analysis of the colon epithelial cells (EpCAM^{pos}) from DSS-treated mice revealed a decrease in expression of CD44, which labels the intestinal crypt base that contains stem cells and progenitors, in *Myd88*^{-/-} versus WT mice (Fig. 1d). Expression of *Lgr5*, which marks the proliferative intestinal stem cell, has been shown to be important for intestinal regeneration after injury (1, 45, 46, 47). To ascertain whether the expression of *Lgr5* is affected, we performed *Lgr5* qPCR on flow-sorted EpCAM^{pos}CD44^{pos} cells from DSS-treated mice on Day 9. The expression of *Lgr5* is significantly reduced in DSS-treated *Myd88*^{-/-} versus WT mice (Fig. 1e). These data suggest that the TLR-MYD88 pathway may play a role in the expansion of the intestinal LGR5^{pos} stem cell pool critical for after injury.

TLR2-MYD88 protects *Apc*^{min/+} mice from adenomas

LGR5^{pos} intestinal stem cells are reported to be the cells of origin of adenoma initiated by adenomatous polyposis coli (*Apc*) mutations (48, 49, 50, 51, 52, 53). Moreover, it has been shown that ablation of MYD88 protects *Apc*^{min/+} mice from developing intestinal tract tumours (14, 54, 55, 56). Confirming that TLR2 is involved in the protection conferred by MYD88, *Tlr2*^{-/-}*Apc*^{min/+} mice have reduced mortality and polyp formation compared with *Apc*^{min/+} littermates (Fig. 2a, b). Deletion of *Tlr2-Myd88* in the mouse could protect mice from adenoma formation indirectly through the innate immune system, directly through effects on epithelial cells, or both.

To assess whether there is an intestinal epithelial-specific component, we investigated whether deletion of MYD88 in epithelial cells, which should dampen all TLR signalling, can also protect *Apc*^{min/+} mice from adenoma formation. We developed *Myd88*^{fl/fl}*Vil-CRE*^{ert2}*Apc*^{min/+} mice, and induced *CRE*^{ert2} with tamoxifen at the weaning age. Consistent with reports on the mortality of *Apc*^{min/+} mice, un-induced *Myd88*^{fl/fl}*Vil-CRE*^{ert2}*Apc*^{min/+} mice developed intestinal tumours and died within the age of 5–6 months. When MYD88 was ablated in a *Vil-CRE*^{ert2}-specific manner, median survival of *Apc*^{min/+} mice went from 149 to 259 days (Fig. 2c). After 16 weeks of tamoxifen induction, the number of polyps was extensively reduced when MYD88 was ablated as compared with the control (Fig. 2d). These data suggest that the TLR-signalling component MYD88, at least partly, affects the epithelial stem cells and contributes in

a cell-intrinsic manner to the mortality and morbidity caused by inactivation of Apc in intestinal tumours.

TLR2-CD14-MYD88 complex in the mammary epithelium

We postulated that other epithelia, such as the breast epithelium, might also contain a stem cell compartment that uses the TLR pathway. We began by addressing whether TLR2 was expressed on murine mammary epithelial cells (MECs). Using published markers of murine MECs (mammary repopulating units (MRUs: CD24^{pos}, CD49^{fpos}), myo-epithelial cells (MYOs: CD24^{dim}CD49^{fpos}), luminal progenitor cells (Ma-CFCs: CD24^{pos}CD49^{fneg}CD61^{pos}) and mature luminal cells (EPIs: CD24^{pos}, CD49^{fmin}, CD61^{neg}); refs 24, 25, 57, 58, 59), we found that TLR2 was expressed in all populations (Fig. 3a). We next confirmed expression of TLR2 and its downstream adaptor MYD88 by qPCR on flow-sorted populations. Increased *Tlr2* expression was found in the Ma-CFC, MRU and MYO populations versus the mature luminal population. *Myd88* expression was similar in all populations (Supplementary Fig. 1a, b).

To investigate whether TLR2 has a functional role in MECs, we examined the effect of known TLR2 ligands (Pam3CSK4 activates TLR2/1; Malp-2 activates TLR2/6) in

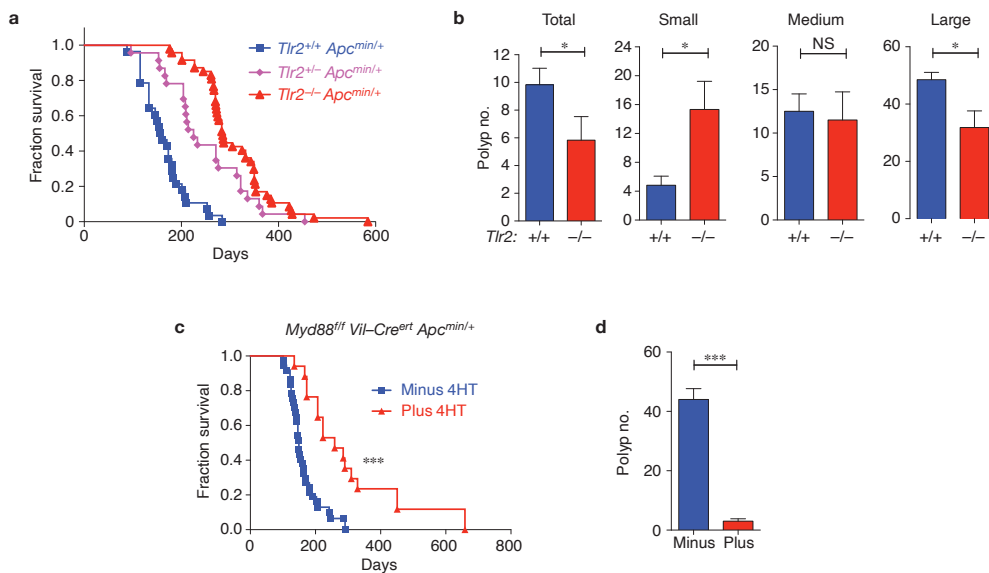


Figure 2: TLR2 and MYD88 protect *Apc*^{min/+} mice from adenomas. (a) Kaplan-Meier survival curve showing *Tlr2*^{+/+}*Apc*^{min/+} (N = 28 mice), *Tlr2*^{+/-}*Apc*^{min/+} (N = 23) *P* < 0.001 and *Tlr2*^{-/-}*Apc*^{min/+} (N = 47) *P* < 0.001. Median survival rates of *Tlr2*^{+/+}*Apc*^{min/+}, *Tlr2*^{+/-}*Apc*^{min/+} and *Tlr2*^{-/-}*Apc*^{min/+} mice are 157, 225 and 285 days, respectively. (b) Total number of visible polyps (* *P* = 0.024) and number of small (* *P* = 0.028), medium (NS, not significant) and large polyps (* *P* = 0.05) in the S.I. and colon was quantified in age (16–20 weeks old)-matched *Tlr2*^{+/+}*Apc*^{min/+} (blue; N = 12) and *Tlr2*^{-/-}*Apc*^{min/+} (red; N = 6) mice. Analysis of the visible polyps in the intestinal tract showed that *Tlr2* deficiency resulted in a significantly decreased tumour burden with decreased large polyps and increased smaller-sized polyps as compared with wild-type. (c) Kaplan-Meier survival curve showing *Myd88*^{fl/fl} *Vil-Cre*^{ert2} *Apc*^{min/+} mice not treated with 4HT (blue, N = 35) and mice treated with 4HT (red, N = 15), *P* < 0.001. (d) Total number of visible polyps in S.I. and colon was quantified at 12 weeks after 4HT injection, *** *P* < 0.001 (N = 4 mice). Values represent mean ± s.d. Student's unpaired t-test for independent samples was used.

Chapter 4

an *in vitro* colony-forming assay. Basal cells (CD24^{dim/pos}CD49f^{pos}) and luminal cells (CD24^{pos}CD49f^{neg/dim}) were sorted and plated on Matrigel with or without TLR2 stimulation (Fig. 3b). Each ligand increased colony formation of both luminal and basal cells. More importantly, HMGB1, an endogenous TLR2 ligand that is released

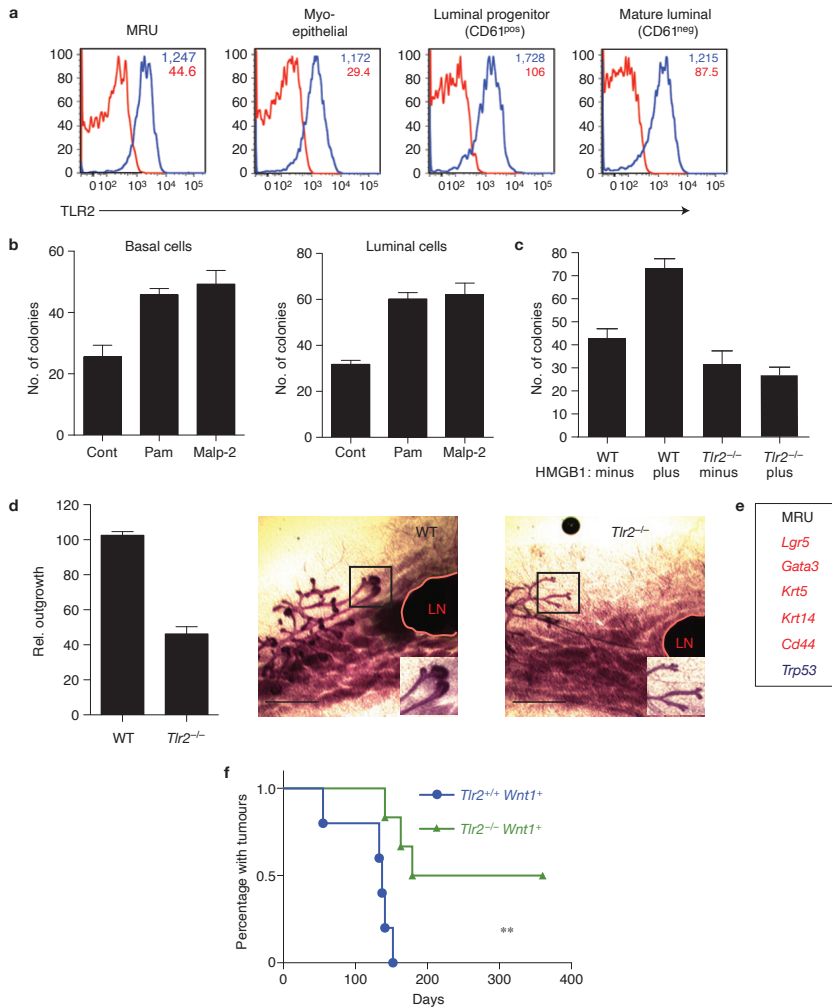


Figure 3: TLR2 is functionally expressed in normal murine mammary epithelial cells. (a) Representative flow cytometry histogram showing staining for TLR2 (red) and the isotype control (blue) for each specific sub-population. Median fluorescence is shown for each histogram. Data are representative of 4 mice. (b) Matrigel *in vitro* colony-forming capacity of luminal and basal cells with Pam3CSK4 and Malp-2. Basal cells: Cont (control) versus Pam: $P = 0.0035$; Cont versus Malp-2; $P = 0.007$. Luminal cells: Cont versus Pam: $P = 0.0001$; Cont versus Malp-2: $P = 0.001$. (c) Sorted Ma-CFCs from WT mice and Tlr2^{-/-} mice in a 2D colony-forming assay with or without HMGB1. WT control versus WT HMGB1: $P = 0.0001$. Experiments in b,c were repeated twice with $N = 4$ mice in total. (d) Carmine Alum-stained inguinal mammary glands of 22-day-old mice, WT and Tlr2^{-/-} mice. Lymph node (LN). Representative photograph is shown; data are representative of 8 mice. Scale bar, 1 mm. (e) List of differentially expressed genes based on single-cell PCR of MRUs ($N = 2$). (f) Kaplan-Meier survival curve showing Tlr2^{+/+}Wnt1⁺ ($N = 6$ mice) and Tlr2^{-/-}Wnt1⁺ ($N = 6$ mice) ** $P < 0.03$. Values represent mean \pm s.d. Student's unpaired t-test for independent samples was used.

on cell death, increased colony formation of flow-sorted Ma-CFCs from a WT mouse but not from a *Tlr2*^{-/-} mouse (Fig. 3c). Without HMGB1, WT and *Tlr2*^{-/-} Ma-CFCs had similar levels of colony formation (Fig. 3c and Supplementary Fig. 1c). Next, we determined whether TLR2 has a functional role *in vivo* by analysing ductal outgrowth in mammary glands of 4-6-week-old *Tlr2*^{-/-} mice, which corresponds to puberty. WT mice showed a ductal tree with terminal end buds almost reaching the lymph nodes at day 22 post-partum. In contrast, the development of the age-matched ductal tree in the *Tlr2*^{-/-} mice was severely inhibited (Fig. 3d) and only a few terminal end buds were observed. *Tlr2*^{-/-} epithelial ducts showed decreased invasion of the fat pad at day 22 post-partum but not at week 6-8. In addition, flow cytometry analysis and immunohistochemistry of 6-week-old mammary glands and mammary whole-mounts of lactating WT and *Tlr2*^{-/-} mice showed no appreciable difference (Supplementary Fig. 2a-c).

As for intestine, deletion of *Tlr2* could affect MECs by a cell-extrinsic manner, cell-intrinsic manner, or both. To determine whether the MRU compartment was affected by the absence of TLR2 in a cell-intrinsic manner, we performed transplantation experiments. We transplanted sorted lineage-negative (linneg) cells from WT or *Tlr2*^{-/-} mice into recipient WT mice and found that *Tlr2*^{-/-} mice have decreased frequency of MRUs as determined by the outgrowth of the mammary epithelial tree (Fig. 4a and Supplementary Fig. 3a). Furthermore, limiting dilution transplantation of sorted MRUs of *Tlr2*^{-/-} and secondary transplants of outgrowths from the MRU transplants also exhibited decreased outgrowth of the mammary epithelial tree as compared with WT (Fig. 4b-d). The outgrowth defects occurred in mice whose immune systems are WT suggesting that there is a decreased number of MRUs in *Tlr2*^{-/-} mice and TLR signalling has a cell-intrinsic component that is not solely due to signalling defects of innate immune cells.

The role of the TLR2 co-activator CD14 in MECs was investigated next (7, 9). The expression of CD14 in different populations of murine MECs was measured (Supplementary Fig. 2d). Consistent with published results we find that luminal cells consist of a CD14^{pos} and a CD14^{neg} population²³. Luminal CD14^{pos} cells have been described to be progenitors as this population has the capacity to form colonies *in vitro*. Intriguingly, MYOs as well as MRUs also expressed CD14, although at a lower level. As for WT and *Tlr2*^{-/-} mice, flow cytometry analysis and immunohistochemistry of 6-week-old mammary glands and mammary whole-mounts of lactating *Cd14*^{-/-} mice showed no appreciable difference (Supplementary Fig. 2a-c). Limiting dilutions of sorted linneg and MRUs revealed that there is a reduction in MRU frequency when *Cd14* is ablated in the MECs (Fig. 4a-c and Supplementary Fig. 3a). TLR4 was also expressed in all 4 populations (Supplementary Fig. 2d). Our transplant studies of *Tlr4*^{-/-} mice revealed a decreased MRU frequency (Supplementary Fig. 3b, c), although not as significant as that observed for *Tlr2*^{-/-} and *Cd14*^{-/-} mice. Together these data suggest that TLR2, CD14 and TLR4 are functionally expressed on murine MECs and that they affect the MRU function.

To confirm that MYD88 has a function in the MRU compartment independent of signalling from innate immune cells, we analysed *Myd88*^{-/-} MECs of cytokeratin-14 (K14)-CreposMyd88^{f/f} mice. Analysis of the K14-Cre mouse revealed that both

Chapter 4

luminal and myo-epithelial cells arise from K14pos-expressing progenitors (28). Limiting dilution transplantation of sorted linneg cells from K14-CreposMyd88f/f and K14-CrenegMyd88f/f mice revealed a 24-fold reduction in the MRU frequency when MYD88 is ablated in MECs (Fig. 4e, f). These data suggest that MYD88 could play a

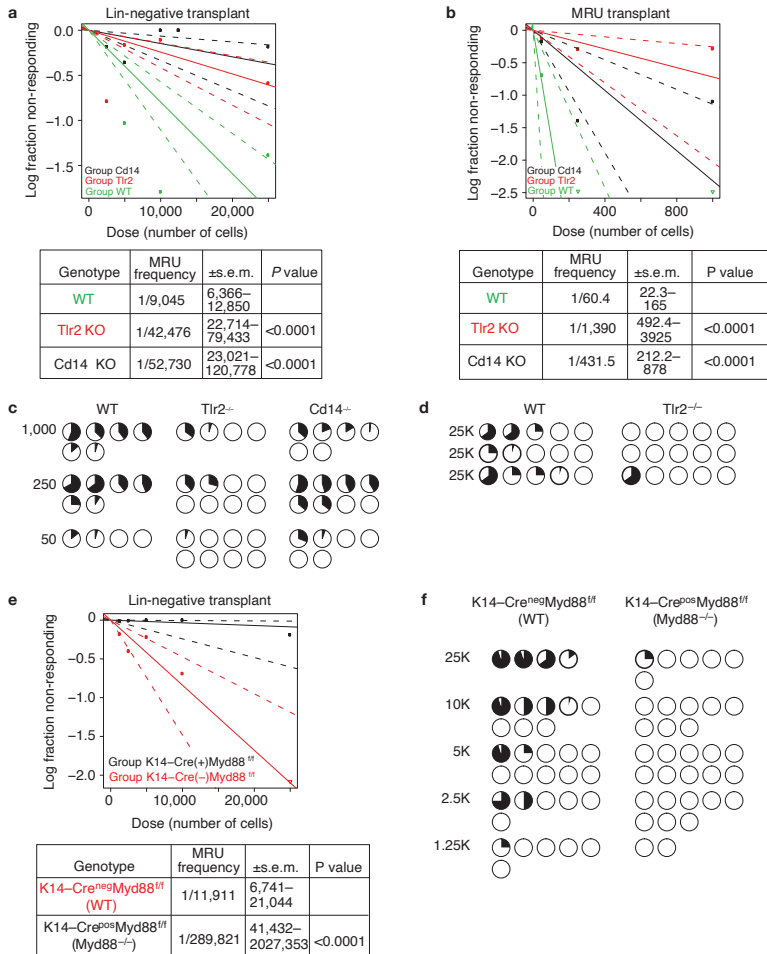


Figure 4: Limiting dilutions of mammary epithelial cells. (a) Limiting dilution transplant of *Tlr2*^{-/-} and *Cd14*^{-/-}lin^{neg} MECs. ELDA graph and analysis showing stem cell (MRU) frequencies. Data for: WT (N = 75 samples) pooled from 5 + independent experiments; *Tlr2*^{-/-} (N = 75 samples) pooled from 5 + independent experiments, P < 0.0001; *Cd14*^{-/-} (N = 46 samples) pooled from 4 independent experiments, P < 0.0001. (b,c) Limiting dilutions transplant of *Tlr2*^{-/-} and *CD14*^{-/-} sorted MRUs including ELDA graph and analysis of MRU frequencies and pie charts showing mammary tree coverage of the fatpad of individual recipients in each group. Data for WT (N = 16 samples), *Tlr2*^{-/-} (N = 20 samples), P < 0.0001 and *Cd14*^{-/-} (N = 20 samples), P < 0.0001, pooled from 3 independent experiments. (d) Pie charts for secondary transplants of MECs from primary MRU transplants of WT and *Tlr2*^{-/-} mice. Secondary transplants were done using Lin^{neg} cells (25K, 10–12 weeks after primary transplants of mice from a). Data for WT (N = 15 samples) and *Tlr2*^{-/-} (N = 15 samples) pooled from 3 independent experiments. Student's unpaired t-test for independent samples was used; P = 0.007. (e,f) Limiting dilution transplant of *Myd88*-deficient lin^{neg} MECs using *K14-Cre*^{neg}*Myd88*^{f/f} (WT), *K14-Cre*^{pos}*Myd88*^{f/f} (*Myd88*^{-/-}) including pie charts showing mammary tree coverage of the fatpad in individual recipients in each group. ELDA graph and analysis showing MRU frequency. Data for WT (N = 34 samples) and *Myd88*^{-/-} (N = 34 samples) pooled from 4 independent experiments. P < 0.0001.

role in the function of mammary stem cells. Immunohistochemistry of the Myd88-deficient mammary glands demonstrated that the mammary gland remains intact (Supplementary Fig. 2e, f).

IL-1R1 and IL-18R are also upstream of MYD88 (15); hence, we examine whether they have a role in the decreased transplant efficiency of *Tlr2*^{-/-} and *Cd14*^{-/-} cells. Flow cytometry analysis showed that IL-1R1 is expressed on MRUs and Ma-CFCs whereas IL-18R is expressed only on the Ma-CFCs. (Supplementary Fig. 2d). Limiting dilution transplantation of sorted linneg cells from *Il-1r1*^{-/-} and *Il-18r*^{-/-} mice did not show a decreased MRU frequency (Supplementary Fig. 3d). These results suggest that although MYD88 is a downstream effector of TLR2, CD14, IL-1R1 and IL-18R signalling, the signal for efficient function of MRU is through the TLR2/CD14 pathway.

Single-cell analysis of MECs

Owing to a significant reduction of MRUs in *Tlr2*^{-/-} MECs, we reasoned that these cells may express altered levels of genes involved in self-renewal. We examined the expression of a number of genes involved in self-renewal in WT and *Tlr2*^{-/-} cells using single-cell qPCR. MRUs, MYOs, Ma-CFCs and EPIs were double sorted for single-cell qPCR. Hierarchical clustering was performed according to the similarity of their gene expression. Although similar clusters were observed a number of genes that are involved in self-renewal were differentially expressed in the MRUs between WT and *Tlr2*^{-/-} mice (Fig. 3e and Supplementary Fig. 4). The upregulation of *Trp53* prompted us to investigate whether the *Ink4a/Arf* locus is functionally involved in the reduction of MRU in the *Tlr2*^{-/-} mice. As previously published, *Ink4a/Arf* deficiency did not affect the MRU frequency, suggesting that *INK4A/ARF* levels in WT stem and progenitor cells are negligible¹⁸. Importantly, *Ink4a/Arf* deficiency resulted in a rescue of MRU frequency of the *Tlr2*^{-/-} cells (Supplementary Figs 3e, f). These data suggest that MRUs are affected by TLR2.

TLR2 in development of mammary tumours

To determine whether TLR2 affects *in vivo* breast cancer development, MMTV-*Wnt1* transgenic mice were bred on a *Tlr2*^{-/-} background. *Tlr2*^{-/-}MMTV-*Wnt1* (*Tlr2*^{-/-}*Wnt1*) mice have a markedly reduced tumour formation compared with *Tlr2*^{+/+}MMTV-*Wnt1* (*Tlr2*^{+/+}*Wnt1*) littermates (Fig. 3f). Median tumour-free days of *Tlr2*^{+/+}*Wnt1* and *Tlr2*^{-/-}*Wnt1* are 137 and 269 days respectively. These data suggest that TLR2 is involved in *de novo* mammary tumour formation.

TLR signalling in human breast and colon cancer

Given the possible role of TLRs in normal murine colon and breast epithelial stem cells, and in murine intestinal and breast tumours, we investigated whether it might play a role in human cancer. Genomic copy number alterations of the TLR2 downstream effectors, IRAK1 and IRAK4, are common; amplifications are found in breast (IRAK1, 23.8%) and colon (IRAK1, 18%; IRAK4, 39.7%) cancer (www.broadinstitute.org/tumorscape). In addition, analysis of TLR2 in 482 breast tumours and 240 colon tumours from publicly available data sets (cBio Cancer Genomics Portal and Sanger Catalogue of Somatic Mutations in Cancer) revealed 1 case of a nonsense mutation (R230X) in breast cancer and 4 cases of nonsense mutations (E230X, E283X (2X), E481X) in colon cancer.

Chapter 4

As IRAK1 and IRAK4 amplifications are common in colon and breast cancer, respectively, we set out to determine the role of the TLR pathway in human breast cancer. Flow cytometry analysis revealed that in contrast to normal human (Fig. 5a)

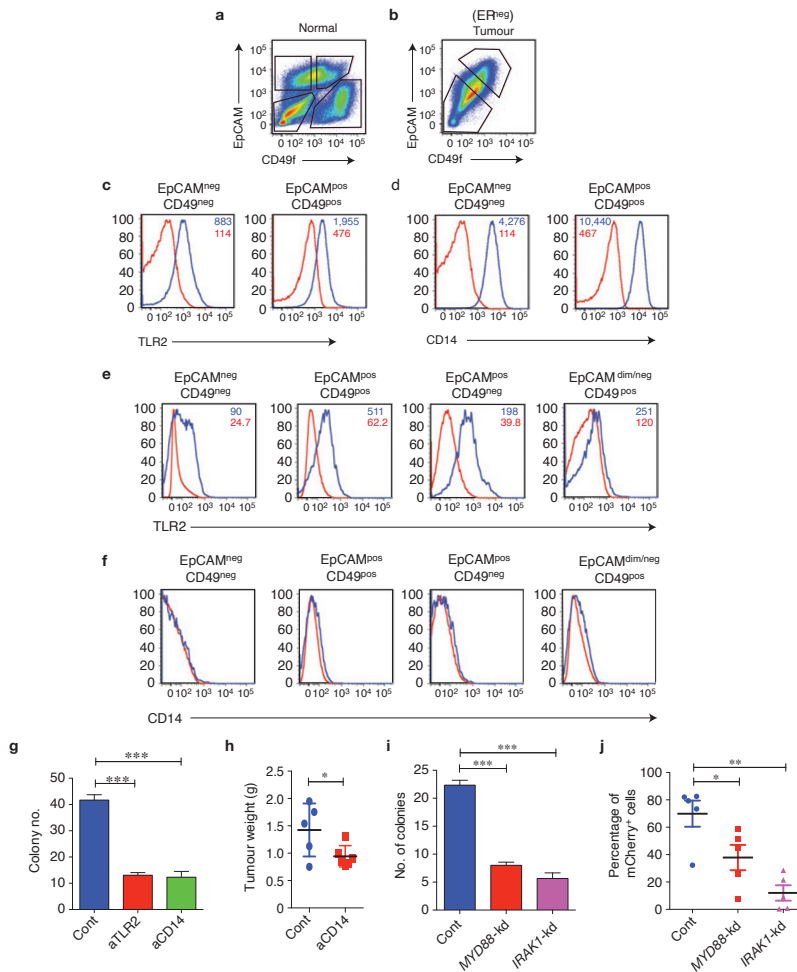


Figure 5: TLR2 expression on human ER^{neg} breast cancer influences *in vitro* colony formation. (a,b) Representative flow cytometry analysis of normal breast epithelial cells (a) and of primary ER^{neg} breast tumours (b). (c,d) Representative histograms for TLR2 (c) and CD14 (d) of the outlined populations in (b). Red: isotype control; blue: TLR2 or CD14 staining. (e,f) Representative histograms for TLR2 (e) and CD14 (f) of the outlined populations in (a). Red: isotype control; blue: TLR2 or CD14 staining. (a-f) Data are representative of more than 10 independent patient samples. (g) *In vitro* colony-forming capacity of a primary ER^{neg} tumour. Neutralizing anti-TLR2 and anti-CD14 blocked clonogenicity; data are shown for one representative tumour of 4 different primary tumours. $N = 3$ samples. $***P < 0.001$. (h) Neutralizing anti-CD14 inhibits *in vivo* tumour growth in a breast tumour xenotransplantation model. Data shown are for one representative of 2 experiments with different primary tumours. Control (Cont), $N = 5$ mice; anti-CD14 (aCD14), $N = 5$ mice; $*P = 0.04$. (i) Knockdown of *MYD88* or *IRAK1* resulted in decreased clonogenicity in primary ER^{neg} tumour. ($N = 3$ samples); $***P < 0.001$. Data shown are for a representative of 2 different primary tumours. (j) *In vivo* effects of *MYD88* or *IRAK1* knockdown in primary ER^{neg} tumour ($N = 5$ mice for each condition). $*P = 0.0425$, $**P = 0.0008$. Data shown are for a representative of 2 different primary tumours. Each group consists of 5 mice. All values in this figure represent mean \pm s.d. Student's unpaired t-test for independent samples was used.

and mouse MECs (Supplementary Fig. 1), which have distinct luminal and basal phenotypes, many human ER^{neg} breast tumours contain a homogeneous population of EpCAM^{pos}CD49f^{pos} cells (Fig. 5b). Similar to normal murine Ma-CFCs, which contain the highest expression of TLR2 (Fig. 3a) and CD14 (ref. 23; Supplementary Fig. 2d), the EpCAM^{pos}CD49f^{pos} population of ER^{neg} tumour cells also has a high expression of TLR2 and CD14 (Fig. 5c, d). TLR2 and CD14 were also expressed on normal MECs (Fig. 5e, f). To determine whether TLR2 and CD14 are needed for the growth of EpCAM^{pos}CD49f^{pos} cells from human ER^{neg} breast tumours, we cultured these cells *in vitro* on Matrigel, and treated them with a neutralizing antibody against TLR2 or CD14. The TLR2 and CD14 neutralizing antibody blocked colony formation of ER^{neg} tumour cells (Fig. 5g and Supplementary Fig. 5a). Furthermore, a CD14 neutralizing antibody blocked tumour growth of two independent ER^{neg} breast cancer xenografts *in vivo* (Fig. 5h). Short hairpin RNA (shRNA)-mediated knockdown of TLR2 in two breast cell lines similarly decreased clonogenic outgrowth (Supplementary Figs 5b and 6c).

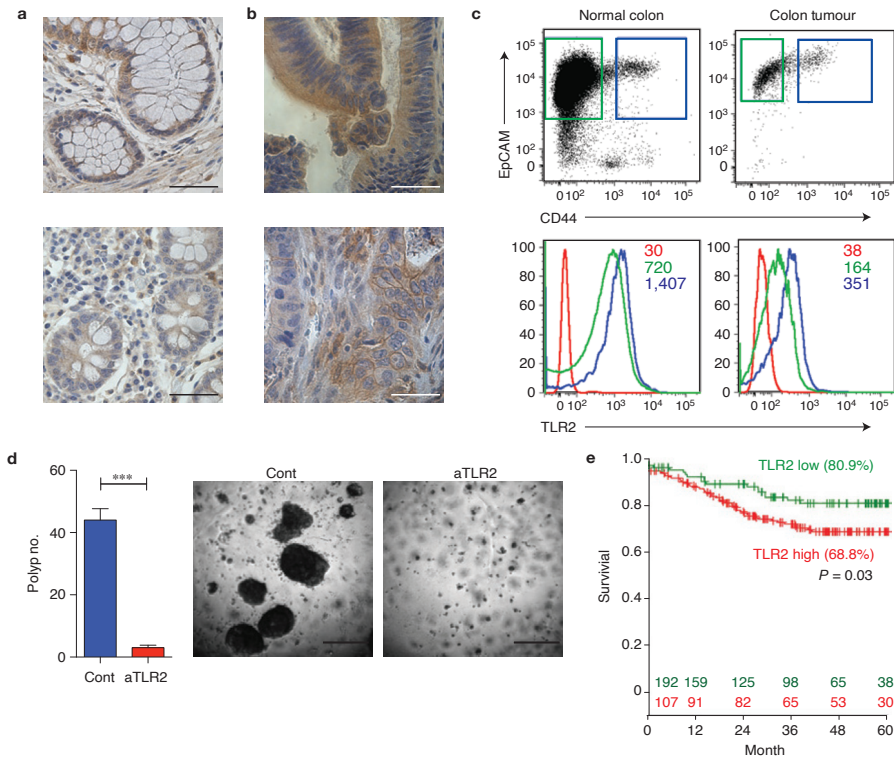


Figure 6: TLR2 expression on colon cancer influences *in vitro* colony formation. (a, b) Immunohistochemistry for TLR2 in normal colon (a) and colon cancer (b). Data are representative of more than 10 independent patient samples (scale bar, 50 μ m). (c) Representative histograms of flow cytometry analysis of a primary colon cancer and normal colon epithelial cells for TLR2; in red: isotype control; in blue: CD44⁻ cells; in green: CD44⁺ cells. Data are representative of 4 independent patient samples. (d) *In vitro* colony-forming capacity of a primary colon cancer. Neutralizing anti-TLR2 blocked clonogenicity. $N = 3$ independent experiments of one tumour. Data are for one representative tumour of 3 different tumours; *** P value is 0.0008. Values represent mean \pm s.d. Student's unpaired t-test for independent samples was used. (Scale bar, 500 μ m.) (e) Kaplan-Meier analysis of the probability of overall survival according to TLR2 expression. TLR2 high ($N = 107$ patients), TLR2 low ($N = 192$ patients); data pooled from 2 independent data sets. $P = 0.03$.

Next, MYD88 and IRAK1, which participate in TLR signalling downstream of TLR2, were investigated. Cells isolated from human breast ER^{neg} xenograft tumours and breast cancer cell lines were treated with shRNAs against MYD88 and IRAK1, then cultured on Matrigel to determine their ability to form colonies *in vitro*. Inhibition of MYD88 and IRAK1 resulted in decreased clonogenicity (Fig. 5i and Supplementary Fig. 5d, e) and decreased NF- κ B activity as indicated by IL-1 β expression (Supplementary Fig. 5f). The MYD88 and IRAK1 shRNAs also resulted in a negative selection in the formation of *in vivo* xenograft tumours derived from ER^{neg} breast cancer cells (Fig. 5j). Together these data support the notion that TLR2 and CD14 are expressed on some ER^{neg} breast cancers and that the TLR2-CD14-MYD88-IRAK1 pathway is important for tumorigenicity.

Finally, because our data suggest that the TLR2 pathway plays an important role in the formation of pre-cancerous polyps/adenomas in the *Apc*^{min/+} mouse model, we wanted to evaluate the role of the TLR2 pathway in human primary tumours. Immunohistochemistry and flow cytometry analysis showed that normal colon epithelial and colon cancer cells express TLR2 (Fig. 6a-c and Supplementary Fig. 6). To ascertain whether TLR2 functions in colon cancer, we cultured primary colon tumour cells *in vitro* on Matrigel and treated the cells with a TLR2 neutralizing antibody. The TLR2 neutralizing antibody blocked colony formation in primary colon cancer (Fig. 6d). Furthermore, we found in a multivariate analysis from two independent data sets that high TLR2 expression is significantly associated with worse overall survival in colon cancer (Fig. 6e). Together, these findings show that the TLR2 pathway is important in murine intestinal tumorigenesis as well as in human colon cancer.

As mutational studies of *Drosophila* TLR (27) and mammalian TLR4 (22) showed that truncated TLR mutants can be constitutively active, we assessed whether nonsense mutations found in public data sets of human colon and breast cancers can activate TLR2 signalling. We first examined whether these cancer mutations can lead to the expression of a truncated TLR2 from a downstream alternative start site (M490; Fig. 7a). We generated WT TLR2 (wtTLR2) and mutant TLR2-E283X (Δ TLR2) vectors fused to YFP for easy visualization and transfected them into HEK293T cells. YFP was detected with both the wtTLR2-YFP fusion as well as the Δ TLR2-YFP fusion, showing that the Δ TLR2-YFP mutant was transcribed and translated (Fig. 7b). As TLR2 signals through the NF- κ B pathway, we tested whether the deletion mutant could induce activation of NF- κ B. Transient expression of Δ TLR2-YFP caused an increased NF- κ B activation as compared with wtTLR2-YFP, confirming that this deletion mutant is a gain-of-function mutant leading to constitutive activation of NF- κ B (Fig. 7c). In addition, we used 9-induced genome editing to create a similar truncated TLR2 gene in HEK293T cells (19, 21) (Fig. 7d and Supplementary Fig. 7). This resulted in an increase in expression of *IL-1 β* (indicates NF- κ B activation), *Axin2* and *Lgr5* (suggests an increase in Wnt signalling; Fig. 7e). All of these data suggest that truncated TLR2 can be constitutively active and that the TLR2 nonsense mutations found in human tumour es can potentially result in active TLR2 signalling.

DISCUSSION

Although previous studies have shown a role for TLR2 and MYD88 in tumour growth

TLR2-MYD88 in intestinal and breast epithelia and oncogenesis

and progression (14, 39) and a tumour-cell-extrinsic role for TLR2 and MYD88 (6, 31), this study suggests cell-intrinsic TLR2-CD14-MYD88 signalling in colon and breast epithelial cells; however, this does not exclude a potential role for cell-extrinsic signalling through the same pathway. The suggested epithelial-cell-intrinsic role is demonstrated by using intestinal- and mammary-specific *MyD88*-deletion mice and by transplanting MECs into WT recipients. All of these data suggest that there is a cell-intrinsic component of TLR2-CD14-MyD88 signalling in epithelial cells.

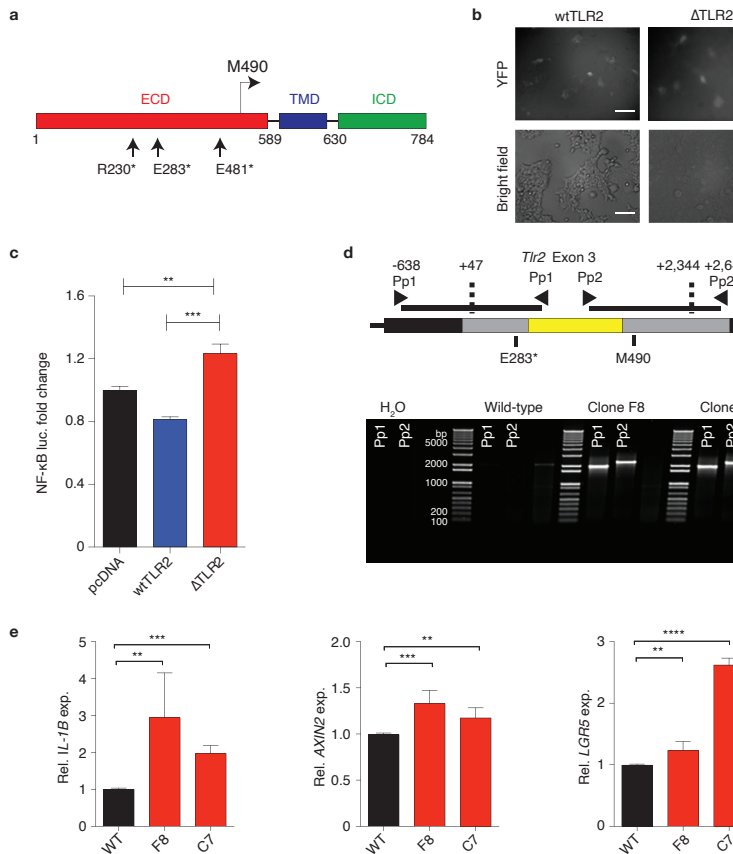


Figure 7: Analysis of TLR2 mutations. (a) Structure of the TLR2 gene. Extracellular domain (ECD), transmembrane domain (TMD) and intracellular domain (ICD). The nonsense mutations and the alternative in frame start-site (M490) are shown. (b) Protein expression of wtTLR2-YFP and Δ TLR2-YFP mutant as determined by fluorescence of YFP by fluorescence microscopy. (Scale bar, 100 μ m.) (c) Δ TLR2-YFP mutant constitutively activate NF- κ B as determined with an NF- κ B-driven luciferase reporter; $N = 6$ independent examples per condition; $**P < 0.01$; $***P < 0.001$. Values represent mean \pm s.d. Student's unpaired t -test for independent samples was used. (d) Strategy of the genome modification using Cas9 to induce double-stranded breaks in the *Tlr2* locus, and template for homology redirected repair (left arm starts at + 46 and right arm ends at + 2344 with an antisense CMV-Cherry-pA cassette in between. Primer pairs (Pp1 and Pp2) consist of 1 primer specific for the HR construct and 1 specific for the untargeted TLR2 locus. PCR analysis showing insertion of the targeting vector by homologous recombination. (e) Real-time qPCR for *IL-1B*, *AXIN2* and *LGR5* on 2 genome-modified clones with genome-edited *Tlr2*. $N = 3$ independent experiments; *Il-1b* expression, $**P = 0.02$, $***P = 0.0001$; *Axin2* expression, $**P = 0.01$, $***P = 0.003$; *Lgr5* expression, $**P = 0.01$; $****P = 0.0000001$. Values represent mean \pm s.d. Student's unpaired t -test for independent samples was used.

Chapter 4

In MECs, our analysis of mutant mice suggests that the TLR2 pathway is important for normal mammary gland development and the self-renewal property of MRUs. In colon epithelial cells, our data suggest that the TLR2-MYD88 pathway is important in the expansion of the stem and progenitor cell pool during regeneration after injury as demonstrated by the DSS colitis animal model. Hence, TLR2 signalling can potentially play an important role in regulating self-renewal independent of its role in inflammation or in haematopoietic cells (39).

It would be intriguing to know which endogenous TLR2 ligand(s) can activate TLR2 signalling in stem cells. Interestingly, many endogenous TLR2 ligands are secreted by dying cells. Apoptosis is a crucial step during normal mammary development, especially lumen formation (41, 43), as well as during intestinal cell differentiation (44). Furthermore, the intestinal epithelium demonstrates a remarkable capacity for responding to injury or increased apoptosis by upregulating the self-renewal capacity of stem cells. It is possible that TLR2 signalling could play a role in such stem cell plasticity. For example, HMGB1, which is released during apoptosis and cell death, has been shown to induce a TLR2-dependent inflammatory response in haematopoietic cells (45, 46, 47). In addition, we provide evidence that HMGB1 induces epithelial cell clonogenicity, thereby linking apoptosis and TLR2 signalling.

As inflammation in response to pathogens leads to collateral damage and death to normal cells, one could speculate that these ligands simultaneously stimulate innate immune cells and proliferation of normal stem cells to facilitate the repair of the tissue damage induced by the inflammatory response. This increased proliferation would be accompanied by an increased mutation rate, providing a direct molecular link of TLR signalling to inflammation and cancer. In support of this, we show that inactivation of TLR, specifically in intestinal epithelial cells, significantly reduces the development of cancers in an *Apc* mutant mouse model.

Pathways used by normal stem cells are frequently used by cancer cells (49, 51). Several lines of evidence suggest that TLR signalling drives proliferation of at least a subset of cancers. Large-scale sequencing of human tumours has shown that although there are certain genes that are commonly mutated in tumours, there are also a large number of mutations that occur at much lower frequency. A better insight into the role of uncommon genetic variants/mutation will have implications for the detection and management of cancer. This has recently been shown by the finding that rare *PPM1D* mutations lead to a predisposition to develop breast and ovarian cancer (52). In breast cancer, *IRAK1* is frequently amplified in breast cancer and is in the peak of the amplicon (www.broadinstitute.org/tumorscape), which suggest that activation of the TLR pathway might occur in a significant number of patients. In support of this, our data clearly show that TLR signalling drives proliferation of both colon and ER^{neg} breast cancer cells *in vitro*. In addition, our data demonstrate a low rate of mutations that can result in constitutive TLR2 signalling. Importantly, blocking TLR signalling in an ER^{neg} tumour *in vivo* by a CD14 neutralizing antibody inhibits tumour growth. This demonstrates that the pathway is used by tumours *in vivo*. Recently, NF- κ B has been shown to directly activate Wnt signalling, which can potentially lead to tumour initiation⁴⁸. Our data show that in epithelial cells of the intestinal tract and the mammary, TLR2-MYD88 affects the expression of Wnt target genes (for example,

Lgr5 and Cd44). This potentially explains the reduced tumorigenicity in Tlr2^{-/-} or Myd88-deficient MMTV-Wnt1 and Apc^{min/+} models.

ER^{neg} breast cancers have a tendency to relapse and metastasize, and are associated with recurrence, metastasis and a high mortality rate (14, 54, 55, 56). The data presented here suggest that there may be a role in targeting the TLR2-MYD88-CD14-IRAK1 signalling pathway to develop potential drugs to treat this subset of breast cancer patients who generally have a poorer prognosis. Different TLRs (for example, TLR2 and TLR4) are potentially active in these tumours but share common downstream effectors such as MYD88 and the co-receptor CD14. Inhibiting these targets may provide a more generalized inhibition of tumour growth. However, as such inhibitors are likely to also inhibit the pathway in innate immune cells, it is difficult to predict the effects on tumours by concurrent inhibition of the pathway in epithelial cells and innate immune cells. Recent studies demonstrate that blocking tumour cell CD47, which allows tumours to evade the innate immune system, seems to have therapeutic potential in breast cancer (57). Hence, to preserve an innate immune response against the tumour, a combinatorial therapeutic strategy that combines neutralizing antibodies against CD14 (or IRAK1 kinase inhibitors) with a neutralizing antibody against CD47 may be effective. In this combination, inhibition of TLR signalling could directly inhibit tumour cell growth, whereas inactivation of CD47 could potentially activate an innate immune response against the tumour through a TLR-independent pathway (58).

METHODS

Mice

Tlr2^{-/-} mice (B6.129-Tlr2^{tm1Kir}/J), Cd14^{-/-} mice (B6.129S-Cd14^{tm1Frm}/J), Tlr4^{-/-} mice (B6.B10ScN-Tlr4^{lps-del}/JthJ), Il-1r1^{-/-} mice (B6.129S7-Il1r1^{tm1mx}/J), Il-18r1^{-/-} mice (B6.129P2-Il18r1^{tm1AKi}/J), Myd88^{-/-} mice (B6.129P2(SJL)-Myd88^{tm1.1Defr}/J), Myd88^{flox} mice (B6.129P2(SJL)-Myd88^{tm1Defr}/J), K14-Cre transgenic mice (Tg(KRT14-cre)1Amc/J), Apc^{min} mice (cBL/6J-ApcMin/J), FVB.Cg-Tg(Wnt1)1Hev/J, C57BL/6-TgN(ACTbEGFP)1Osb (pCx-GFP) and Nod/Scid/Il2r^{-/-} (NSG) mice (NOD.Cg-Prkdc^{scid}IL2Rg^{tm1Wjl}/SzJ) were purchased from the Jackson Laboratory. Villin-Cre^{ERT2} (60) was previously described. K14-Cre transgenic mice were on a mixed background and were crossed to mice with a C57BL/6J background for 3 generations. Recipient mice in the transplant assays for wild-type, Tlr2^{-/-} and Cd14^{-/-} cells were C57BL/6J mice and for the K14-Cre Myd88f/f mice we used NSG mice. All mice used for this study were maintained at the Stanford Animal Facility in accordance with the guidelines of the animal care use committee.

Preparation of single-cell suspensions of tissues

Mice were euthanized and all fat pads surgically resected. Tissue was digested in DMEM/F12 for 2 h, and then processed as previously described (25, 60, 61). Briefly, mechanically dissociated mammary glands were treated for two hours with collagenase and hyaluronidase (StemCell Technologies) followed by lysis of red blood cells in ACK (NH₄CL), followed by 1-2 min treatment with pre-warmed 0.25% trypsin EDTA (Invitrogen), followed by prewarmed dispase (StemCell Technologies) plus DNaseI (Sigma) for 2 min, and filtration through a 70 µm mesh and washed with flow cytometry buffer (HBBS, 2% FCS, PSA). Single-cell samples of murine colon and

Chapter 4

small intestine were prepared as described previously (16, 61).

For human samples, informed consent was obtained after the approval of protocols by the Stanford University and City of Hope Institutional Review Boards. Human breast specimens, and primary or xenograft tumours were mechanically dissociated into >1-2 mm³ pieces with a razor blade and digested at 37 °C with collagenase and hyaluronidase, in Advanced DMEM/F12 (Invitrogen) with 2 mM Glutamax (Invitrogen), 120 µg ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin, 0.25 µg ml⁻¹ amphotericin-B (PSA) and incubated for 4-12 h. At the end of the incubation, cells were treated with ACK to lyse the red blood cells followed by a short incubation in dispase and DNaseI. Cells were filtered through a 70 µm nylon mesh and washed with flow cytometry buffer (HBBS, 2% FCS, PSA). Preparation of human colon tumours was performed as described previously (62, 63).

DSS-induced colitis

Colitis was induced by 1.5% (w/v) dextran sodium sulphate (DSS; relative molecular mass 36,000-50,000; MP Biochemicals) added to the drinking water for 7 days. Colons were collected for histology or experiment at day 9 to day 11.

In vivo transplants

For mouse transplants, double-sorted live lin^{neg} cells (CD45^{neg}CD31^{neg}TER119^{neg}DAPI^{neg}) or MRU population (Lin^{neg}CD24^{dim}CD49^{high}) was collected in staining media and resuspended with 50% Matrigel (BD Biosciences). Per transplant 10 µl was injected into cleared fat pads of weaning-age mice (21-28 days), of indicated genotype, as previously described²⁵. All transplants were allowed to grow for 6-10 weeks before analysis.

Tumour cell engraftment and treatment

Human tumour cells were suspended in staining media containing 50% Matrigel and injected into the fourth abdominal fat pad by subcutaneous injection at the base of the nipple of female NSG mice. Two weeks after tumour injection mice received 100 mg anti-CD14 (clone 60bca from ATCC) every other day until the end of the experiment. The experiment was performed twice with 5 mice per group.

In vitro colony-forming assays

For the human and mouse 3D *in vitro* colony assay, 96-well ultralow attachment round-bottom plates (BD) were prepared with a feeder layer of irradiated L-WNT3a cells mixed with 40 µl of growth factor reduced-Matrigel (BD) per well. Cells were re-suspended in media and transferred onto the solidified Matrigel mix. Medium used to culture the human cells is as follows: advanced Dulbecco's modified Eagle medium/F12 (Invitrogen), 10% FBS, 2.5% growth factor-reduced Matrigel (BD), 10 µM Y-27632 (sigma), 10 ng ml⁻¹ EGF (R&D), 100 ng ml⁻¹ Noggin (R&D), 250 ng ml⁻¹ RSPO-1 (R&D), 1X B27 (Invitrogen), 1X N2 (Invitrogen) 120 µg ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin (P/S). Mouse 2D *in vitro* colony assay cells were cultured in Epicult B medium (StemCell Technologies) with 5% serum in the presence of 13,000 cm² irradiated NIH-3T3 cells. After 24-48 h, the medium was replaced with serum-free Epicult B, and colonies were counted 7 days later. Anti-TLR2 antibodies were clone T2.5 (10 µg ml⁻¹; eBioscience catalogue number 14-9024-82) and TL2.1 (10

$\mu\text{g ml}^{-1}$; Biologend catalogue number 309710). Purified CD14 antibody hybridoma clone 60bca ($10 \mu\text{g ml}^{-1}$; ATCC catalogue number HB-247) was used for *in vitro* assays. TLR2 ligands: Pam3CSK4 ($25 \mu\text{g ml}^{-1}$, TLR1/2, Imgenex), Malp-2 (50 ng ml^{-1} , TLR2/6, Imgenex), HMGB1 (50 ng ml^{-1} , TLR2, Sigma-Adrich). HT29, MDA-MB-231, MCF7, MDA-MB-468 and 293T cells were cultured with DMEM supplemented with 10% fetal bovine serum, 100 U penicillin and $100 \mu\text{g ml}^{-1}$ streptomycin.

Flow cytometry

To reduce nonspecific binding, cells suspended in staining buffer were blocked on ice for 10 min with rat IgG (Sigma) 10 mg ml^{-1} at 1:1,000. Cells were then stained, in the dark, on ice for 30-60 min with optimal antibody concentrations, which were determined by titration experiments. Mouse antibodies include CD45-biotin (clone 30-F11, BD catalogue number 553078); $1 \mu\text{g ml}^{-1}$, TER119-biotin (clone TER-119, BD catalogue number 553672); $1 \mu\text{g ml}^{-1}$, CD31-biotin (clone 390, BD catalogue number 558737); $1 \mu\text{g ml}^{-1}$, CD24-Fitc (clone M1/69, eBioscience catalogue number 11-0242-85); $0.5 \mu\text{g ml}^{-1}$, TLR2 PeCy7 (clone T2.5, eBioscience catalogue number 25-9024-80); $2 \mu\text{g ml}^{-1}$, EpCAM-APC-Cy7 (clone G8.8, Biologend catalogue number 118218); $1 \mu\text{g ml}^{-1}$, TLR4 PeCy7 (clone SA15-21, Biologend catalogue number 145408); $1 \mu\text{g ml}^{-1}$, IL-18R1 Alexa Fluor 647 (clone BG, Biologend catalogue number 132903); $1 \mu\text{g ml}^{-1}$, CD49f-APC (clone GoH3, Biologend catalogue number 313615); $1 \mu\text{g ml}^{-1}$, CD61 PE (clone 2C9.G2, Biologend catalogue number 104308); $2 \mu\text{g ml}^{-1}$, CD14 Apc-Cy7 (clone Sa14-2, Biologend catalogue number 123318); $1 \mu\text{g ml}^{-1}$, streptavidin APC-Cy7 (Biologend catalogue number 405208); $1 \mu\text{g ml}^{-1}$ and streptavidin PacificBlue (Invitrogen catalogue number S-11222); $1 \mu\text{g ml}^{-1}$. Lineage cells consist of CD45-, TER119- and CD31-positive cells in mouse. Human antibodies used include: EpCAM-Alexa Fluor 488 (clone 9C4, Biologend catalogue number); $1 \mu\text{g ml}^{-1}$, CD49f-APC (clone GoH3, Biologend catalogue number); $1 \mu\text{g ml}^{-1}$, CD14 Apc-Cy7 (clone M5E2, Biologend catalogue number); $1 \mu\text{g ml}^{-1}$, CD10 PeCy7/Apc-Cy7 (clone H110a, Biologend catalogue number); $1 \mu\text{g ml}^{-1}$, H-2Kd biotin/Pacific Blue (clone SF1-1.1, Biologend catalogue number); $1 \mu\text{g ml}^{-1}$, TLR2 PeCy7 (clone T2.5, eBioscience catalogue number); $2 \mu\text{g ml}^{-1}$, CD31 biotin (clone WM59, BD Biosciences catalogue number); $1 \mu\text{g ml}^{-1}$, CD45 biotin (clone HI30, BD Biosciences catalogue number); $1 \mu\text{g ml}^{-1}$, IL-1R1 Alexa Fluor 647 (clone 35F5, BD Biosciences catalogue number); $1 \mu\text{g ml}^{-1}$ from BD Biosciences. In primary human tissue lineage cells consist of cells positive for CD45 and CD31; in xenograft lineage cells consist of H-2Kd-positive cells. Flow cytometry was performed with a $100 \mu\text{m}$ nozzle on a BD Flow cytometry Aria II with Flow cytometry Diva software. Data analysis was performed using Flowjo. For all experiments, side scatter and forward scatter profiles (area and width) were used to eliminate debris and cell doublets. Dead cells were eliminated by excluding 4',6-diamidino-2-phenylindole (DAPI)-positive cells (Molecular Probes).

Plasmids

Sequences for human MyD88 shRNA1, IRAK1 shRNA1 and IRAK1 shRNA4 were obtained from L. M. Staudt (NCI, NIH, USA) (16, 64) and cloned into pSUPER. The H1-shRNAsette (EcoR1-Xho1) was sub-cloned into pCDH1 (EcoR1-Sal1) upstream of EF1a-Cherry. Lentiviral constructs against human TLR2 were purchased from Sigma, and included pLKO.1 puro vector (no. 1 TRCN0000057021 and no. 2 TRCN0000057019).

Chapter 4

shRNA-pLKO.1-puro control was used as a negative control. Lenti-viruses were produced as described previously⁶³. pcDNA3 TLR2-YFP (Addgene plasmid 13016) was used for PCR-based mutagenesis to obtain pcDNA3 Δ TLR2-YFP. First PCR: T7 FW primer plus Δ TLR2 Rev (5'-CAGTCATCAAACCTATAATTCTAACAATCC-3') and TLR2 WT(1705-5'-GGGATGGAGAGTCACACAGG-3')Rev plus Δ TLR2 FW (5'-GGATTGTTAGAATTATAGTTTGATGACTG-3'); second PCR: T7 FW primer and TLR2 WT(1705)Rev. PCR product was digested with BamH1 and Ecor1 and cloned into pcDNA3 TLR2-YFP and was subsequently sequenced thoroughly. Phusion high-fidelity DNA polymerase (NEB) was used to perform the TLR2-mutation-based PCR.

The PX459 (pSp(BB)-2A-Puro; ref. 65), obtained from Addgene, was used to clone in the sgRNA seq (5'-GGCCTAACAATTGCAAAGGT-3'). Homologous recombination sette was ordered as a gBlock from IDT; TLR2 locus from + 47 to + 2344 (exon 3) was cloned using TopoTA cloning kit (Invitrogen). CMV-Cherry-pA was cloned in the Hpa1 site in the anti-sense orientation as determined by sequencing. Primer pairs: Pp1 FW 5'-CTTGGTCTGCCTCGAGTTTC-3'; Pp1 Rev 5'-GAAAGTCCCGTTGATTTTGG-3'; Pp2 FW 5'-CCCCGTAATGCAGAAGAAGA-3'; Pp2 Rev 5'-GCAACCAATCCCTTGATA-3'.

Luciferase reporter assay

HEK293T cells were seeded a day before transfection. All transfections were carried out with Fugene 6 (Promega) according to the manufacturer's instructions. Cells were transfected with NF- κ B luciferase expression construct (Promega), pRL-TK Renilla luciferase vector (Promega) plus pcDNA3 TLR2-YFP, pcDNA3 Δ TLR2-YFP or pcDNA3 empty control. Luciferase activities were measured using the Dual-Luciferase Reporter Assay System (Promega) and normalized to Renilla luciferase activity. Three independent samples were used in triplicate for each experiment. Data are from two independent experiments.

Statistical analysis

Differences between groups were analysed using unpaired Student's t-tests. Error bars represent standard deviations (\pm s.d.). For mouse survival curves, differences between curves were analysed using the log-rank (Mantel-Cox) test. For limiting dilution analyses, the frequency of mammary repopulating units was calculated using ELDA software⁶⁴. Expected frequencies are reported, as well as the 95% confidence intervals (lower and upper values are indicated). No statistical method was used to predetermine sample size, experiments were not randomized and investigators were not blinded to allocation during experiments. Investigators were blinded when assessing the outcome of animal experiments (animals had only an ear tag no. for identification) during: pathology assessment of DSS-treated colons, and flow cytometry analysis of different groups of treated mice.

Single-cell gene expression.

This was performed as described previously (25, 60, 61). In short, double-sorted single cells were sorted into individual wells of 96-well plates containing 5 μ l lysis buffer (CellsDirect qRT-PCR mix; Invitrogen) and 2U SuperasIn. After reverse transcription and pre-amplification using multiplexed PCR, reactions were loaded (Hamilton StarLET pipetting robot) on a Fluidigm microfluidic chip. The microfluidic chips were run on the BioMark real-time PCR reader (Fluidigm), and loaded chips underwent

thermocycling and fluorescent quantification according to the manufacturer's instructions. Positive or negative associations among pairs of genes were tested by Spearman correlation, and P values were calculated using $n = 10,000$ permutations. The following TaqMan assays (Applied Biosystems) were used: Lgr5, Mm00438890_m1; Gata3, Mm00484683_m1; Krt5, Mm00503549_m1; Krt14, Mm00516870_mH; CD44, Mm01277161_m1; Actb, Mm00607939_s1; Trp53, Mm01731287_m1; Krt17, Mm01306857_mH; TLR2, Mm00442346_m1 ;MyD88, Mm0044338; Krt9, Mm01701806_m1.

The following (intron spanning) primers were used for quantitative real-time PCR using SYBR green: Axin2 FW 5'-CTCCCCACCTTGAATGAAGA-3'; Axin2 Rev 5'-TGGCTGGTGCAAAGACATAG-3'; LGR5 FW 5'-CTTCCAACCTCAGCGTCTTC-3'; LGFR5 Rev 5'-TTTCCCGCAAGACGTAAGTCTC-3'; β -actin FW 5'-GGATGCAGAAGGAGATCACTG-3'; β -actin Rev 5'-CGATCCACACGGAGTACTTG-3'; IL-1B FW 5'-AGCTGATGGCCTAACAGA-3'; IL-1B Rev 5'-GCATCTTCTCAGCTTGCC-3'; TLR2 FW 5'-GGCGTTCTCTCAGGTGACTG-3'; TLR2 Rev 5'-CTTCCTGGAGAGGCTGATG-3'; MYD88 FW 5'-TTGAGGAGGATTGCCAAAAG-3'; MYD88 Rev 5'-CATCTCTGCACAACTGGA-3'; IRAK1 FW 5'-CTCTGACCAGCCAAGGTCTC-3'; IRAK1 Rev 5'-GCCCGAGGAGTACATCAAGA-3'.

ACKNOWLEDGEMENTS

This study was supported by the National Institutes of Health (NCI), the Breast Cancer Research Foundation, the Ludwig Institute, The California Institute for Regenerative Medicine and the Department of Defense (DOD). F.A.S. was supported by NWO-Rubicon grant, a fellowship from the Dutch Cancer Society and by a seed grant of the organization My Blue Dots. We thank T.N. Schumacher and M.A. Child for scientific input, S. Sim for her assistance with single-cell PCR assays, P. Lovelace for her assistance with flow cytometry and K. Montgomery for IHC. Some research was performed on a FACS Aria that was purchased using NIH S10 Shared Instrumentation Grant (1S10RR02933801) funds.

CONTRIBUTIONS

F.A.S. and A.H.K. performed, designed and analysed research and wrote the paper; L.J.v.W., S.C., M.Z., S.S.S., I.G., D.J. and M.E.R. performed, designed and analysed research; D.Q. and J.S.L. performed research; J.P.V., T.K., S.R.Q., D.S. and M.v.d.R. designed and analysed research; F.M.D. and G.S. provided critical reagents; M.F.C. designed research and wrote paper.

REFERENCES

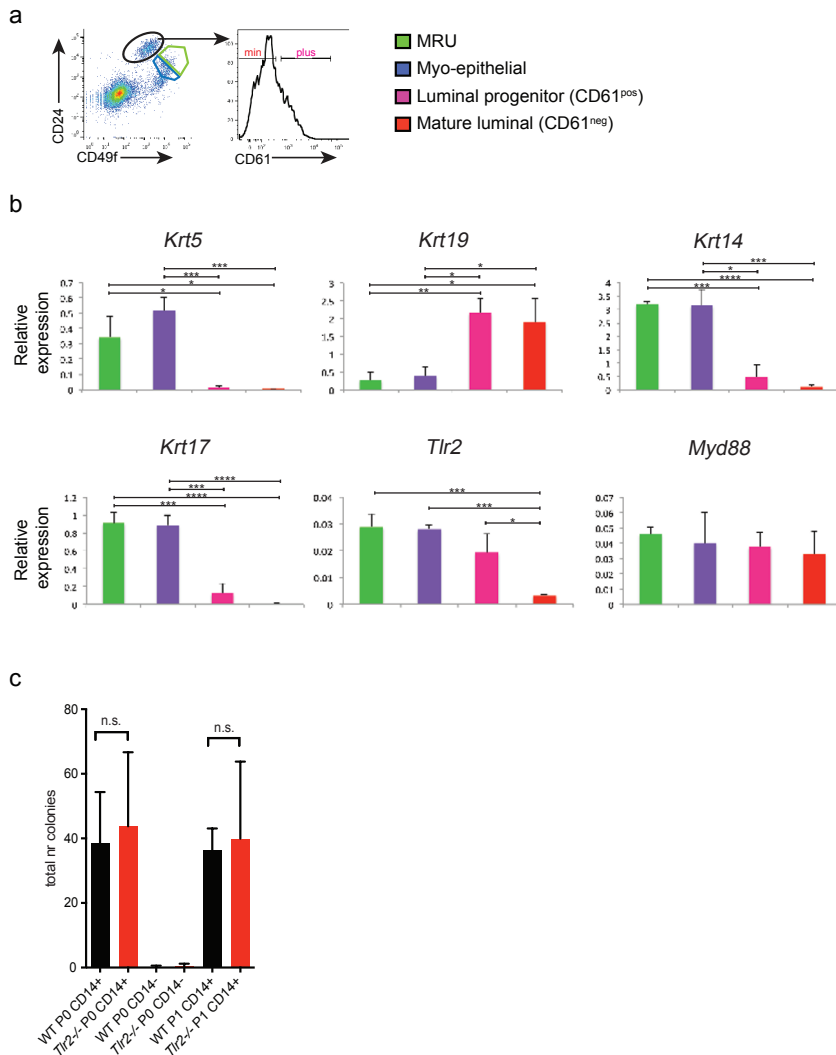
1. Metcalfe, C., Kljavin, N. M., Ybarra, R. & de Sauvage, F. J. Stem cells are indispensable for radiation-induced intestinal regeneration. *Cell Stem Cell* 14, 149-159 (2014).
2. Yu, L., Wang, L. & Chen, S. Endogenous toll-like receptor ligands and their biological significance. *J. Cell. Mol. Med.* 14, 2592-2603 (2010).
3. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* 124, 783-801 (2006).
4. West, X. Z. et al. Oxidative stress induces angiogenesis by activating TLR2 with novel endogenous ligands. *Nature* 467, 972-976 (2010).
5. Park, J. S. High mobility group box 1 protein interacts with multiple Toll-like receptors. *Am. J. Physiol. Cell Physiol.* 290, C917-C924 (2005).
6. Kim, S. et al. Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate

Chapter 4

- metastasis. *Nature* 457, 102-106 (2009).
7. Yang, R-B. et al. Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* 395, 284-288 (1998).
 8. Lee, M. S. & Kim, Y-J. Signaling pathways downstream of pattern-recognition receptors and their cross talk. *Annu. Rev. Biochem.* 76, 447-480 (2007).
 9. Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J. & Mathison, J. C. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science* 249, 1431-1433 (1990).
 10. Clevers, H. At the crossroads of inflammation and cancer. *Cell* 118, 671-674 (2004).
 11. Rakoff-Nahoum, S. & Medzhitov, R. Toll-like receptors and cancer. *Nat. Rev. Cancer* 9, 57-63 (2009).
 12. Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* 420, 860-867 (2002).
 13. Balkwill, F. & Mantovani, A. Inflammation and cancer: back to Virchow? *Lancet* 357, 539-545 (2001).
 14. Rakoff-Nahoum, S. & Medzhitov, R. Regulation of spontaneous intestinal tumorigenesis through the adaptor protein MyD88. *Science* 317, 124-127 (2007).
 15. Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annu. Rev. Immunol.* 21, 335-376 (2003).
 16. Ngo, V. N. et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature* 470, 115-119 (2011).
 17. Puente, X. S. et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 475, 101-105 (2011).
 18. Pietersen, A. M. et al. Bmi1 regulates stem cells and proliferation and differentiation of committed cells in mammary epithelium. *Curr. Biol.* 18, 1094-1099 (2008).
 19. Cong, L. et al. Multiplex genome engineering using CRISPR/ systems. *Science* 339, 819-823 (2013).
 20. Uemura, N. et al. *Helicobacter pylori* infection and the development of gastric cancer. *N. Engl. J. Med.* 345, 784-789 (2001).
 21. Mali, P. et al. RNA-guided human genome engineering via Φ . *Science* 339, 823-826 (2013).
 22. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* 388, 394-397 (1997).
 23. Asselin-Labat, M-L. et al. Gata-3 negatively regulates the tumor-initiating capacity of mammary luminal progenitor cells and targets the putative tumor suppressor p53. *Mol. Cell Biol.* 31, 4609-4622 (2011).
 24. Shackleton, M. et al. Generation of a functional mammary gland from a single stem cell. *Nature* 439, 84-88 (2006).
 25. Stingl, J. et al. Purification and unique properties of mammary epithelial stem cells. *Nature* 439, 993-997 (2006).
 26. Kordon, E. C. & Smith, G. H. An entire functional mammary gland may comprise the progeny from a single cell. *Development* 125, 1921-1930 (1998).
 27. Schneider, D. S., Hudson, K. L., Lin, T. Y. & Anderson, K. V. Dominant and recessive mutations define functional domains of Toll, a transmembrane protein required for dorsal-ventral polarity in the *Drosophila* embryo. *Genes Dev.* 5, 797-807 (1991).
 28. Keymeulen, A. V. et al. Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 479, 189-193 (2011).
 29. Rios, A. C., Fu, N. Y., Lindeman, G. J. & Visvader, J. E. In situ identification of bipotent stem cells in the mammary gland. *Nature* 506, 322-327 (2014).
 30. Plaks, V. et al. Lgr5-expressing cells are sufficient and necessary for postnatal mammary gland organogenesis. *Cell Rep.* 3, 70-78 (2013).
 31. Grivennikov, S. I. et al. Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. *Nature* 491, 254-258 (2012).
 32. Sangiorgi, E. & Capecchi, M. R. Bmi1 is expressed in vivo in intestinal stem cells. *Nat. Genet.* 40, 915-920 (2008).
 33. Tian, H. et al. A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* 478, 255-259 (2011).
 34. Montgomery, R. K. et al. Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *Proc. Natl Acad. Sci. USA* 108, 179-184 (2011).
 35. Takeda, N. et al. Interconversion between intestinal stem cell populations in distinct niches. *Science* 334, 1420-1424 (2011).
 36. Powell, A. E. et al. The Pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. *Cell* 149, 146-158 (2012).
 37. Muñoz, J. et al. The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent ' + 4' cell markers. *EMBO J.* 31, 3079-3091 (2012).
 38. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 449, 1003-1007 (2007).
 39. Tye, H. et al. STAT3-driven upregulation of TLR2 promotes gastric tumorigenesis independent of tumor inflammation. *Cancer Cell* 22, 466-478 (2012).
 40. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* 118, 229-241 (2004).
 41. Humphreys, R. C. et al. Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis.

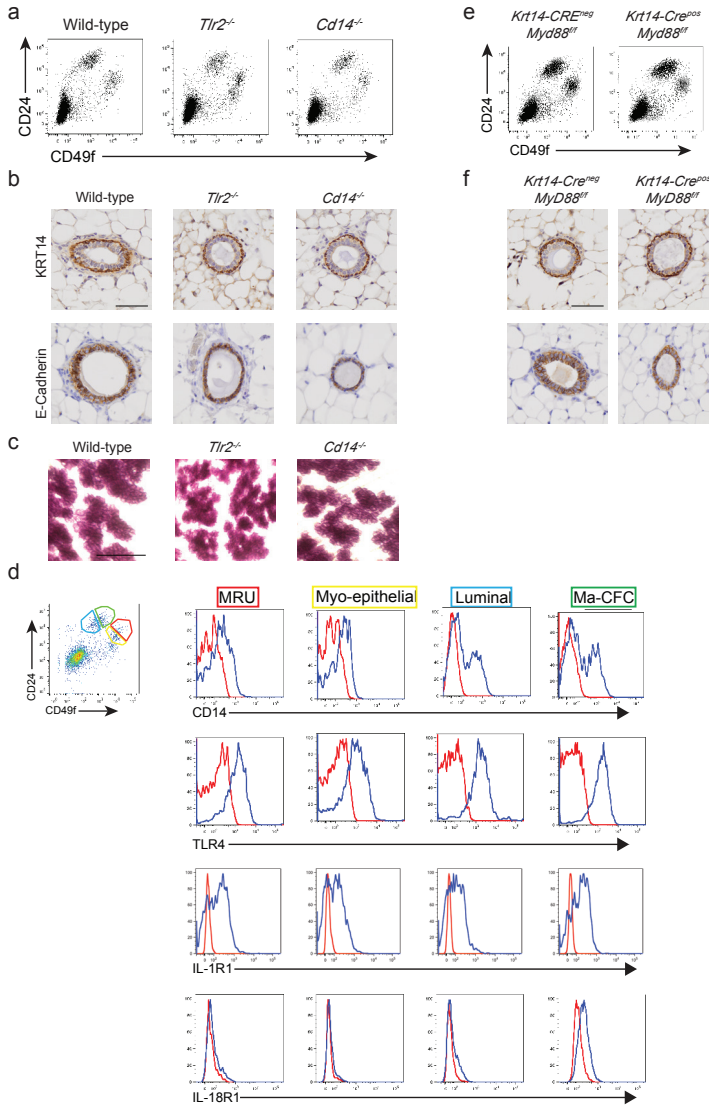
- Development 122, 4013-4022 (1996).
42. Kitajima, S., Takuma, S. & Morimoto, M. Changes in colonic mucosal permeability in mouse colitis induced with dextran sulfate sodium. *Exp. Anim.* 48, 137-143 (1999).
 43. Mailloux, A. et al. BIM regulates apoptosis during mammary ductal morphogenesis, and its absence reveals alternative cell death mechanisms. *Dev. Cell* 12, 221-234 (2007).
 44. Hall, P. A., Coates, P. J., Ansari, B. & Hopwood, D. Regulation of cell number in the mammalian gastrointestinal tract: the importance of apoptosis. *J. Cell Sci.* 107, 3569-3577 (1994).
 45. Bianchi, M. E. & Manfredi, A. A. High-mobility group box 1 (HMGB1) protein at the crossroads between innate and adaptive immunity. *Immunol. Rev.* 220, 35-46 (2007).
 46. Park, J. S. et al. Involvement of toll-like receptors 2 and 4 in cellular activation by high mobility group box 1 protein. *J. Biol. Chem.* 279, 7370-7377 (2004).
 47. Urbanaviciute, V. et al. Induction of inflammatory and immune responses by HMGB1-nucleosome complexes: implications for the pathogenesis of SLE. *J. Exp. Med.* 205, 3007-3018 (2008).
 48. Schwitalla, S. et al. Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* 152, 25-38 (2012).
 49. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* 414, 105-111 (2001).
 50. Barker, N. et al. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* 457, 608-611 (2009).
 51. Shackleton, M., Quintana, E., Fearon, E. R. & Morrison, S. J. Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell* 138, 822-829 (2009).
 52. Ruark, E. et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. *Nature* 493, 406-410 (2012).
 53. Schepers, A. G. et al. Lineage tracing reveals Lgr5 + stem cell activity in mouse intestinal adenomas. *Science* 337, 730-735 (2012).
 54. Dent, R. et al. Pattern of metastatic spread in triple-negative breast cancer. *Breast Cancer Res. Treat.* 115, 423-428 (2009).
 55. Bauer, K. R., Brown, M., Cress, R. D., Parise, C. A. & Caggiano, V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer* 109, 1721-1728 (2007).
 56. Carey, L. A. et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 295, 2492-2502 (2006).
 57. Willingham, S. B. et al. The CD47-signal regulatory protein alpha (SIRP α) interaction is a therapeutic target for human solid tumors. *Proc. Natl Acad. Sci. USA* 109, 6662-6667 (2012).
 58. Chao, M. P. et al. Calreticulin is the dominant pro-phagocytic signal on multiple human cancers and is counterbalanced by CD47. *Sci. Transl. Med.* 2, 63ra94 (2010).
 59. Asselin-Labat, M-L. et al. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat. Cell Biol.* 9, 201-209 (2007).
 60. el Marjou, F. et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* 39, 186-193 (2004).
 61. Rothenberg, M. E. et al. Identification of a cKit(+) colonic crypt base secretory cell that supports Lgr5(+) stem cells in mice. *Gastroenterology* 142, 1195-1205.e6 (2012).
 62. Dalerba, P. et al. Phenotypic characterization of human colorectal cancer stem cells. *Proc. Natl Acad. Sci. USA* 104, 10158-10163 (2007).
 63. Shimonono, Y. et al. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell* 138, 592-603 (2009).
 64. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* 347, 70-78 (2009).
 65. Ran, F. A. et al. Genome engineering using the CRISPR-9 system. *Nat. Protoc.* 8, 2281-2308 (2013).

SUPPLEMENTARY FIGURES

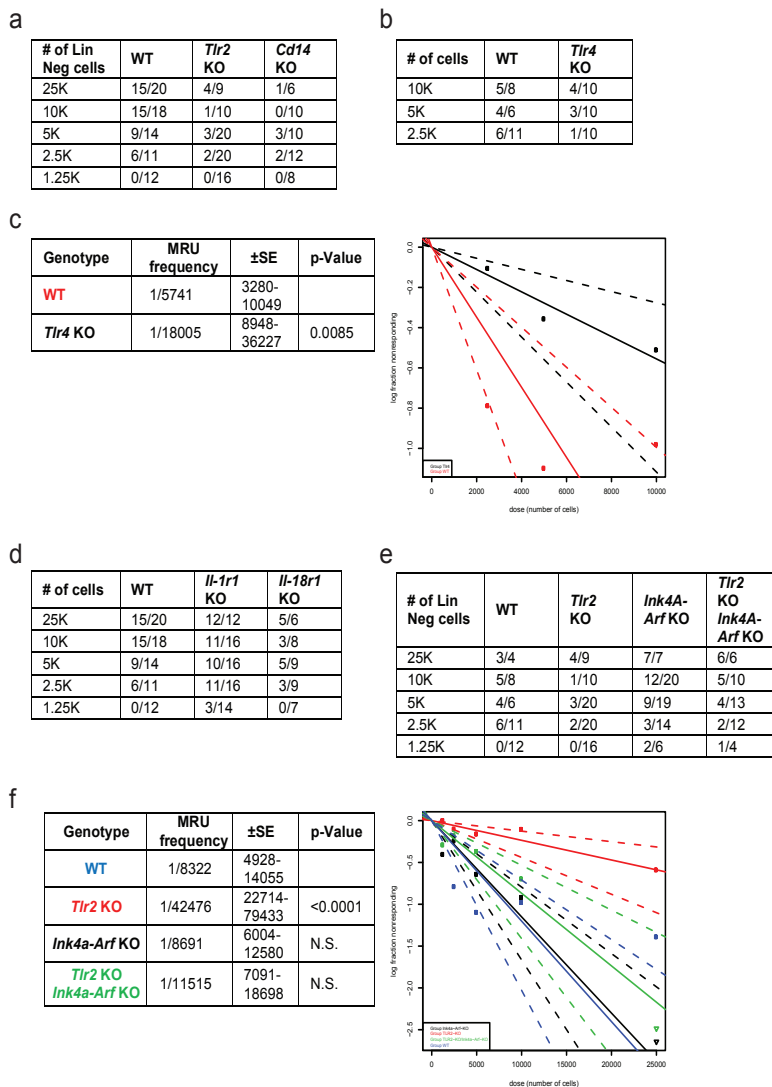
**Supplementary Figure 1: TLR2 and MyD88 expression in murine mammary epithelial subpopulations.**

(a) Representative FACS dot plot and histogram of CD24, CD49f and CD61. Data are representative of 6 mice. (b) Quantitative rt-PCR on MEC populations as indicated in a. Basal markers: Krt5, Krt14, Krt17; Luminal marker: Krt19; $N = 3$ mice; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.005$; **** $P < 0.001$. Error bars represent s.e.m. One-tail unpaired t -test analysis was used. (c) Ma-CFCs were sorted based on expression of CD24^{high}CD49f^{low/neg} luminal phenotype in combination with CD14. 200 cells luminal cells sorted for CD14^{pos} or CD14^{neg} were plated out on matrigel in triplicates. After 7-12 days colonies were counted and passaged as a single cell suspension again ($N = 4$ mice). 4 independent experiments were performed and the average is shown, $P = 0.7, 0.9$ (NS: non-significant). Values represent mean \pm s.d. Student's unpaired t -test for independent samples was used.

TLR2-MYD88 in intestinal and breast epithelia and oncogenesis

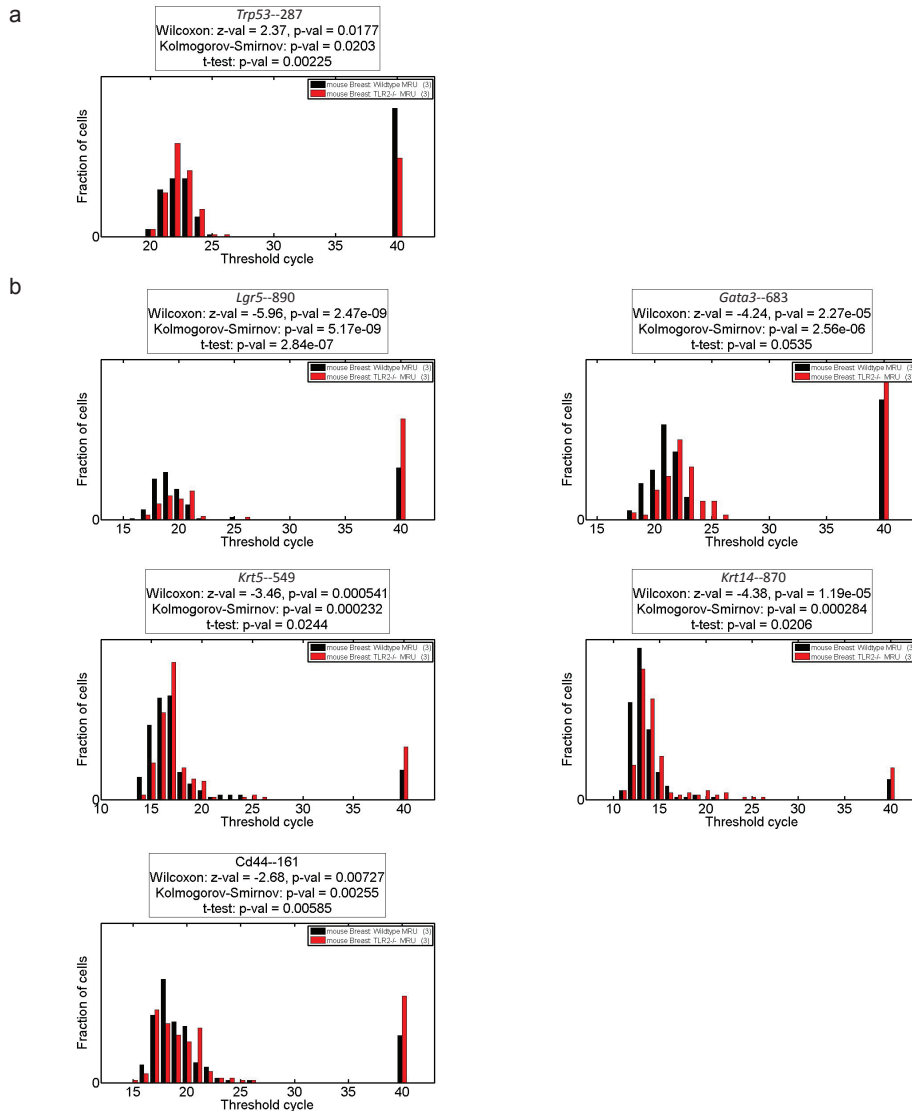


Supplementary Figure 2: Phenotypic characterization of mammary epithelial cells. (a) Flow cytometric analysis of mammary epithelial cells from 6 weeks old mice of wild-type, *Tlr2*^{-/-} and *Cd14*^{-/-} mice. Cells were gated on live and lineage negative cells. (b) Immunohistochemistry on mammary glands from 6 weeks old mice of wild-type, *Tlr2*^{-/-} and *Cd14*^{-/-} mice. Myoepithelial cells (Cytokertain-14 (KRT14)) and luminal cells (E-CADHERIN) are present and properly organized in both wild-type as well as the knock out mammary glands. Scale bar is 50 μ m. (c) Carmine Alum staining on whole mount of lactation of wild-type, *Tlr2*^{-/-} and *Cd14*^{-/-} mice. Scale bar is 1 mm. (d) Flow cytometry analysis of mammary epithelial cells from 6 to 8 weeks old mice. Cells were gated on live and lineage negative and stained for CD24, CD49f and CD14, TLR4, IL-1R1 or IL-18R1. In red is the isotype control for each specific sub-population, in blue is the CD14, TLR4, IL-1R1 or IL-18R1 staining. (e) Flow cytometry analysis of *Myd88*^{-/-} mammary epithelial cells, from 6 weeks old mice of *Krt14-Cre*^{neg}*Myd88*^{fl/fl} and *Krt14-Cre*^{pos}*Myd88*^{fl/fl} mice. Cells were gated on live and lineage negative cells. (f) Immunohistochemistry on mammary glands from 6 weeks old mice of *Krt14-Cre*^{neg}*Myd88*^{fl/fl} and *Krt14-Cre*^{pos}*Myd88*^{fl/fl} mice. Myoepithelial cells (Cytokertain-14 (KRT14)) and luminal cells (E-CADHERIN) are present and properly organized in the *Myd88* knockout mammary gland. Scale bar is 50 μ m. All analyses were done with at least 3 mice.

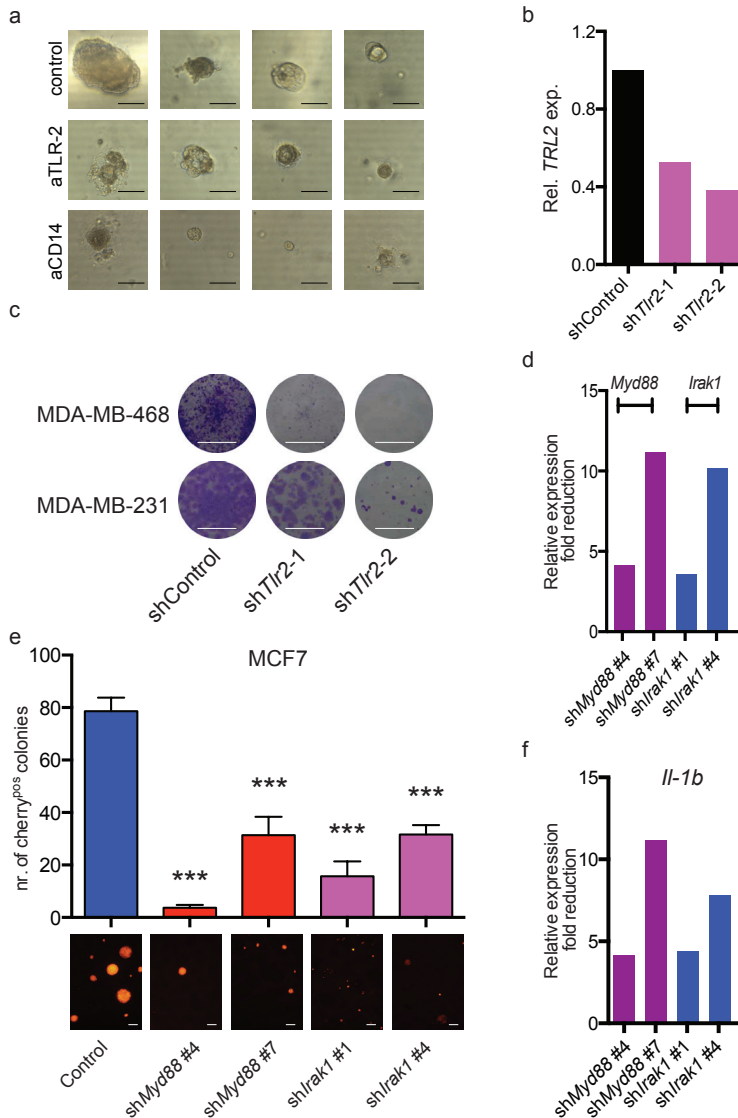


Supplementary Figure 3: Limiting dilutions of mammary epithelial cells. To determine the MRU frequency of each donor genotype we injected sorted cells, as indicated from donor mice of 10 to 14 weeks old mice in cleared mammary fat pads of three weeks old recipients. Recipient mice were syngeneic C57BL/6J mice. Numbers of successful out growths and numbers of total injections are shown for each dilution and genotype. MRU frequency and confidence was determined by ELDA graph and analysis. (a) Raw numbers for limiting dilutions of *Tlr2*^{-/-} and *Cd14*^{-/-} *lin*^{neg} MECs for Fig. 4a. (b,c) Raw numbers for limiting dilutions of *Tlr4*^{-/-} *lin*^{neg} MECs, including ELDA analysis. Data for WT (N = 25 samples), *Tlr4*^{-/-} (N = 30 samples) pooled from 3 independent experiments. P = 0.0085. (d) Raw numbers for limiting dilutions of *Il-1r1*^{-/-} (N = 74 samples) and *Il-18r1*^{-/-} *lin*^{neg} (N = 39 samples) MECs. (e,f) Limiting dilutions of *Ink4a-Arf*^{+/+} *Tlr2*^{+/+} (WT), *Ink4a-Arf*^{+/+} *Tlr2*^{-/-} (*Tlr2* KO), *Ink4a-Arf*^{-/-} *Tlr2*^{+/+} (*Ink4a-Arf* KO) and *Ink4a-Arf*^{-/-} *Tlr2*^{-/-} (*Ink4a-Arf* KO *Tlr2* KO) *lin*^{neg} MECs, including ELDA graph and analysis. Data for WT (N = 41 samples), *Tlr2*^{-/-} KO (N = 75 samples), P < 0.0001, *Ink4a-Arf* KO (N = 66 samples), P = N.S. and *Ink4a-Arf* KO *Tlr2* KO (N = 45 samples), P = N.S. are pooled from 4 independent experiments.

TLR2-MYD88 in intestinal and breast epithelia and oncogenesis

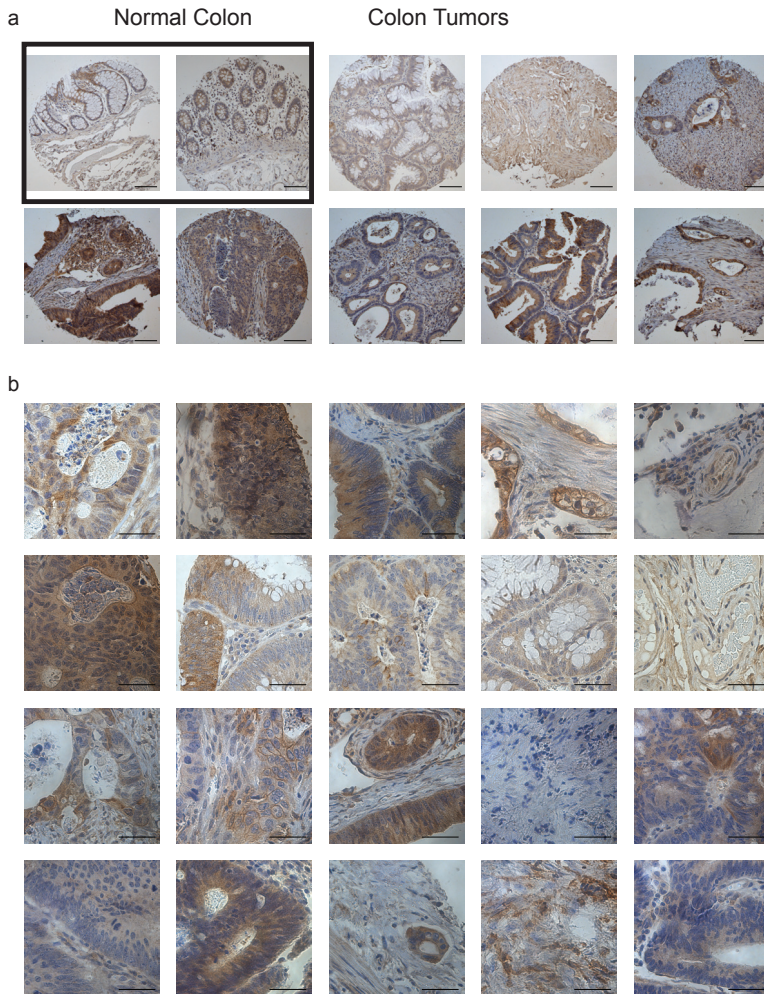


Supplementary Figure 4: Single cell gene expression of MRU. Single cell expression showing expression of wild-type (black) and *Tlr2*^{-/-} (red) MRUs for Fig. 3e. Cells were double sorted and subjected to multiplexed single cell rt-PCR. Normalized Ct values as visualized by comparing the distribution of Ct values in histograms. (a) Gene that was significantly up-regulated in *Tlr2*^{-/-} cells as compared to wild-type cells. (b) Genes that were significantly down-regulated in *Tlr2*^{-/-} as compared to wild-type cells. Each analysis was done on 2 different mice.

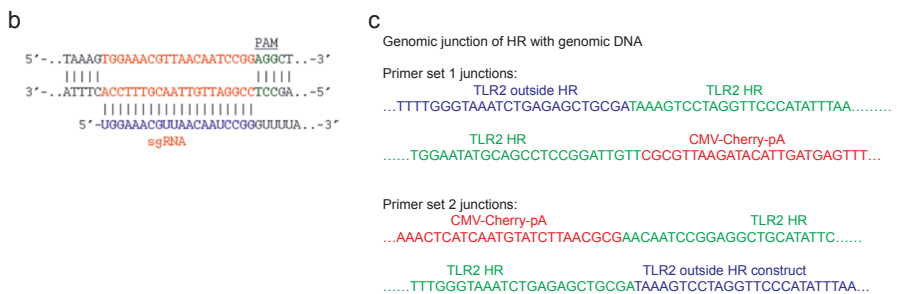
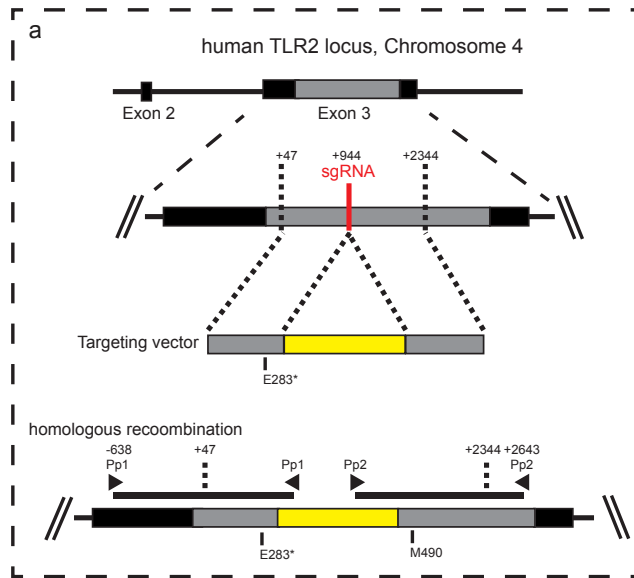


Supplementary Figure 5: Antibody and/or shRNA mediated blockade of TLR2, CD14, MYD88 and IRAK1. (a) Representative examples colony formation on matrigel of primary ER^{neg} breast cancer cells ($N = 4$). Scale bar is 20 μm . Cells were treated with control antibody or with neutralizing TLR2 and CD14 antibodies. (b,c) Two breast cell lines were transduced with the lentivirus to knockdown TLR2 and after puromycin selection cells were plated out and stained with crystal violet blue after 9 days in culture. Data are representative of 2 experiments. Scale bar is 1 cm. (d) MDA-MB-231 cells were transduced with control or indicated shRNA and knock-down efficiency was determined by qPCR. All shRNA mediated knock-down constructs resulted in a 4 to 12x reduction of its target mRNA. Data are representative of 2 experiments. (e) Knock-down of MYD88 and IRAK1 in the breast cell line MCF7. After transduction cells were sorted for Cherry and plated out on matrigel in triplicate to determine colony forming capacity. The number of colonies on matrigel is shown and a representative photo of the colonies is shown. $N = 6$ samples pooled from 2 independent experiments. $***P < 0.001$. Scale bar is 20 μm . Values represent mean \pm s.d. Student's unpaired t -test for independent samples was used. (f) MYD88 and IRAK1 knock down resulted in decreased expression of the NF- κ B target gene IL-1B indicating that the knock down resulted in decreased NF- κ B activity. Data are representative of 2 experiments.

TLR2-MYD88 in intestinal and breast epithelia and oncogenesis



Supplementary Figure 6: TLR2 immunohistochemistry on normal and tumour colon tissue. (a) 10X magnification of immunohistochemistry for TLR2. Boxed is normal colon tissue, the rest are colon tumours. Scale bar is 200 μ m. (b) 40X magnification of immunohistochemistry for TLR2 of colon tumours. Scale bar is 50 μ m.



Supplementary Figure 7: Genome editing the *Tlr2* locus using CAS9. (a) Schematic of the targeted region of human TLR2 locus and targeting strategy. sgRNA cuts around + 944. The homologues recombination (HR) cassette has a left arm (including the E283* mutation) and a right arm. In between there is a CMV Cherry pA cassette (yellow box) in the anti sense orientation. Primer pairs (Pp1 & Pp2) are shown to detect homologues recombination. (b) Region of the sgRNA complementary to the protospacer. (c) Sequenced region shows correct integration of HR construct as determined by sequencing the genomic junctions.

Chapter 5

Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*

The Tabula Muris Consortium¹ [including **Linda J. van Weele**²]

Adapted from: *Nature* volume 562, pages 367-372 (2018)

¹A list of all authors and their affiliations is available online

²Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA

Contribution: for the mammary gland L.J.v.W. designed the experiment, collected and processed the tissue, annotated cell types.

Here we present a compendium of single-cell transcriptomic data from the model organism *Mus musculus* that comprises more than 100,000 cells from 20 organs and tissues. These data represent a new resource for cell biology, reveal gene expression in poorly characterized cell populations and enable the direct and controlled comparison of gene expression in cell types that are shared between tissues, such as T lymphocytes and endothelial cells from different anatomical locations. Two distinct technical approaches were used for most organs: one approach, microfluidic droplet-based 3'-end counting, enabled the survey of thousands of cells at relatively low coverage, whereas the other, full-length transcript analysis based on fluorescence-activated cell sorting, enabled the characterization of cell types with high sensitivity and coverage. The cumulative data provide the foundation for an atlas of transcriptomic cell biology.

The cell is a fundamental unit of structure and function in biology, and multicellular organisms have evolved various cell types with specialized roles. Although cell types have historically been characterized by morphology and phenotype, the development of molecular methods has enabled increasingly precise descriptions of their properties, typically by measuring protein or mRNA expression patterns (1). Technological advances have also expanded measurement multiplexing such that highly parallel sequencing can now enumerate nearly every mRNA molecule in a single cell (2, 3, 4, 5, 6, 7, 8). This approach has provided insights into cell biology and organ composition from various organisms (9, 10, 11, 12, 13, 14, 15, 16, 17, 18). However, although these reports provide valuable characterization of individual organs, it is challenging to compare data collected from different animals by independent labs with varying experimental techniques. It therefore remains unknown whether these data can be synthesized as a more general resource for biology.

Here we report a compendium of cell types from the mouse *Mus musculus*; we refer to this as a *Tabula Muris*, or 'Mouse Atlas'. We analysed several organs from the same mouse, generating a dataset controlled for age, environment and epigenetic effects. This enabled the direct comparison of cell-type composition between organs, and the comparison of shared cell types across organs. The compendium comprises single-cell transcriptomic data from 100,605 cells isolated from 20 organs from three female and four male, C57BL/6JN, three-month-old mice (10–15 weeks), analogous to 20-year-old humans (Fig. 1a). Aorta, bladder, bone marrow, brain (cerebellum, cortex, hippocampus and striatum), diaphragm, fat (brown, gonadal, mesenteric and subcutaneous), heart, kidney, large intestine, limb muscle, liver, lung, mammary gland, pancreas, skin, spleen, thymus, tongue and trachea from the same mouse were immediately processed into single-cell suspensions. All organs were single-cell-sorted into plates using fluorescence-activated cell sorting (FACS), and many were also loaded into microfluidic droplets (see Extended Data and Methods).

All data, protocols, analysis scripts and an interactive data browser are publicly available (for details, see 'Data availability'). This release enables the exact replication of all results, facilitates in-depth analyses not completed here, and provides a comparative framework for future studies using the large variety of murine disease models. Although these data are by no means a complete representation of all mouse

organs and cell types, they provide a first draft attempt to create an organism-wide representation of cellular diversity.

Defining organ-specific cell types

To define cell types, we analysed each organ independently by performing principal component analysis (PCA) on the most variable genes between cells, followed by nearest-neighbour graph-based clustering. We then used cluster-specific gene expression of known markers and genes that are differentially expressed between clusters to assign cell-type annotations to each cluster (Extended Data Figs. 1, 2, Supplementary Table 1). We used a standard annotation method for all organs; step-by-step instructions to reproduce this method are provided in the supplemental Organ Annotation Vignette using the liver as an example. Cell type descriptions and defining genes for each organ are available in the Supplementary Information. For each cluster, we provide annotations in the controlled vocabulary of the cell ontology (19) to facilitate inter-experiment comparisons. Many of these cell types have not previously been obtained in pure populations, and our data provide a wealth of new information on their characteristic gene-expression profiles. Some unexpected discoveries include a potential new role for *Neurog3*, *Hhex* and *Prss53* in the adult pancreas, a cell population expressing *Chodl* in limb muscle, transcriptional heterogeneity of brain endothelial cells, the expression of MHC class II genes by adult mouse T cells, and sets of transcription factors that distinguish cell types across organs.

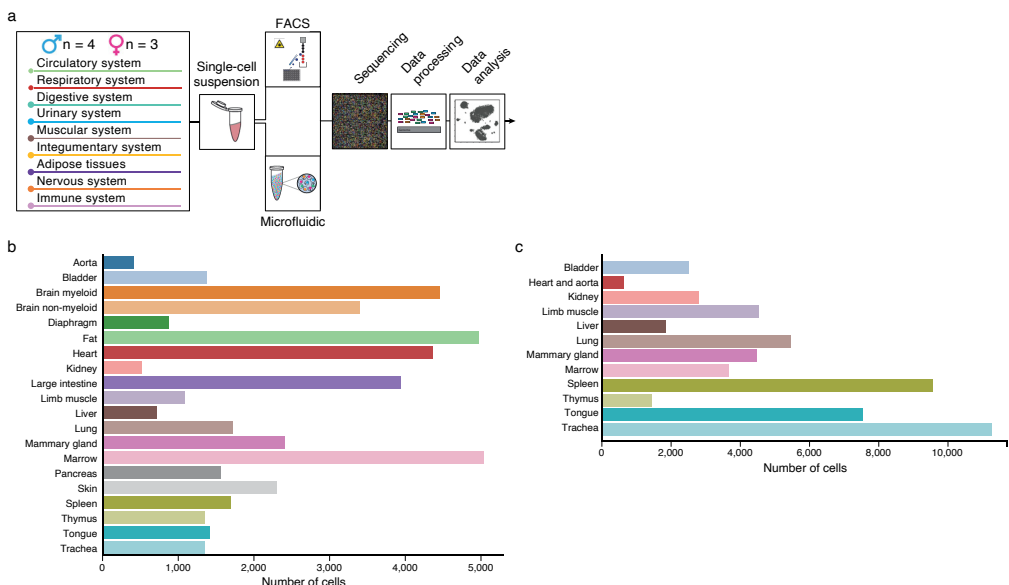


Figure 1: Overview of Tabula Muris. (a) 20 organs from four male and three female mice were analysed. After dissociation, cells were sorted by FACS and, for some organs, captured in microfluidic oil droplets. Cells were lysed, transcriptomes amplified and sequenced, reads mapped, and data analysed. (b) Bar plot showing the number of sequenced cells prepared by FACS from each organ (n = 20 organ types). (c) Bar plot showing the number of sequenced cells prepared by microfluidic droplets from each organ (n = 12 organ types).

Methodological comparison

We performed single-cell RNA-sequencing with two methods: FACS-based cell capture in plates and microfluidic-droplet-based capture (hereafter denoted the FACS method and the microfluidic-droplet method, respectively). To understand the technical biases of each approach, we performed both methods on many organs. Overall, 44,949 cells from the FACS method and 55,656 cells from the microfluidic-droplet method were retained after quality control. Single-cell transcriptomes were sequenced to an average depth of 814,488 reads per cell (FACS) and 7,709 unique molecular identifiers (UMIs) per cell (microfluidic droplet). Comparing methods shows organ-specific differences in the number of cells analysed (Fig. 1b, c), reads per cell (Extended Data Fig. 3a, c) and genes per cell (Extended Data Fig. 3b, d). Furthermore, with both methods the most abundant cell types analysed were epithelial cells and leukocytes, although FACS captured a larger diversity of cell types (Extended Data Fig. 4).

Any individual single-cell sequencing experiment offers only a partial view of cell-type diversity within an organism and gene expression within each cell type. We illustrate the expected variability between methods and experiments by comparing our two measurement approaches to a third method, microwell-seq (20). One notable feature is the variability in the number of genes detected per cell between organs and methods. For example, the median number of genes detected in the bladder is around 4,900 (FACS), 2,900 (microfluidic droplet) and 900 (microwell-seq), whereas in the kidney it is around 1,400 (FACS), 1,900 (microfluidic droplet) and 500 (microwell-seq). In the bladder, liver, lung, mammary gland, trachea, tongue and spleen, nearly twice as many genes are detected per cell with the FACS method compared to the microfluidic-droplet method, whereas the heart and marrow show comparable numbers between the two methods (Extended Data Fig. 5a). This difference is probably not due to sequencing depth, as both FACS and microfluidic-droplet libraries are nearly saturated (Extended Data Fig. 5b). In these comparisons, a gene is considered detected if a single read maps to it, as that is the only value at which reads and UMIs can be treated equally. We also found that the number of detected genes decreases similarly across organs as the read or UMI threshold for a detectable gene is increased (Extended Data Fig. 6).

Next, we investigated whether the three methods agree on the genes defining each cell cluster (Methods). As expected, the FACS and microfluidic-droplet methods show the closest agreement, probably because they used the same biological samples. However, there are several dozen to several hundred genes common to all methods that define each cluster (Extended Data Fig. 7, Supplementary Table 2). This suggests that combining independent datasets can lead to more robust characterizations of gene expression.

Spleen and kidney are two organs for which FACS was performed without marker-based sorting, which enables us to compare the number and relative abundance of different cell types between methods. For those cell types that are captured by both methods, the proportion of each cell type is equivalent (Pearson correlation coefficient: spleen, 0.99; kidney, 0.99). Nonetheless, the microfluidic-droplet method

identified cell types that were missed by the FACS method in both organs, for example kidney mesangial cells, and splenic dendritic and natural killer cells. This is partially explained by cellular abundance and sampling depth (12,333 microfluidic-droplet cells compared with 2,216 FACS cells, Supplementary Table 1), and possibly from cell capture and lysis biases between methods.

As the FACS method captures fewer cells but detects more molecules per cell than the microfluidic-droplet method, we asked whether the two methods agree in their 'bulk' gene-expression profiles for the 33 shared cell populations (Methods). Such gene-expression profiles largely correlate (Pearson correlation coefficient: 0.74–0.90), which suggests that although biases between methods exist, both accurately recapitulate average cell-type gene-expression profiles.

Global clustering across organs

To detect relationships between cells from different organs, we visualized all FACS

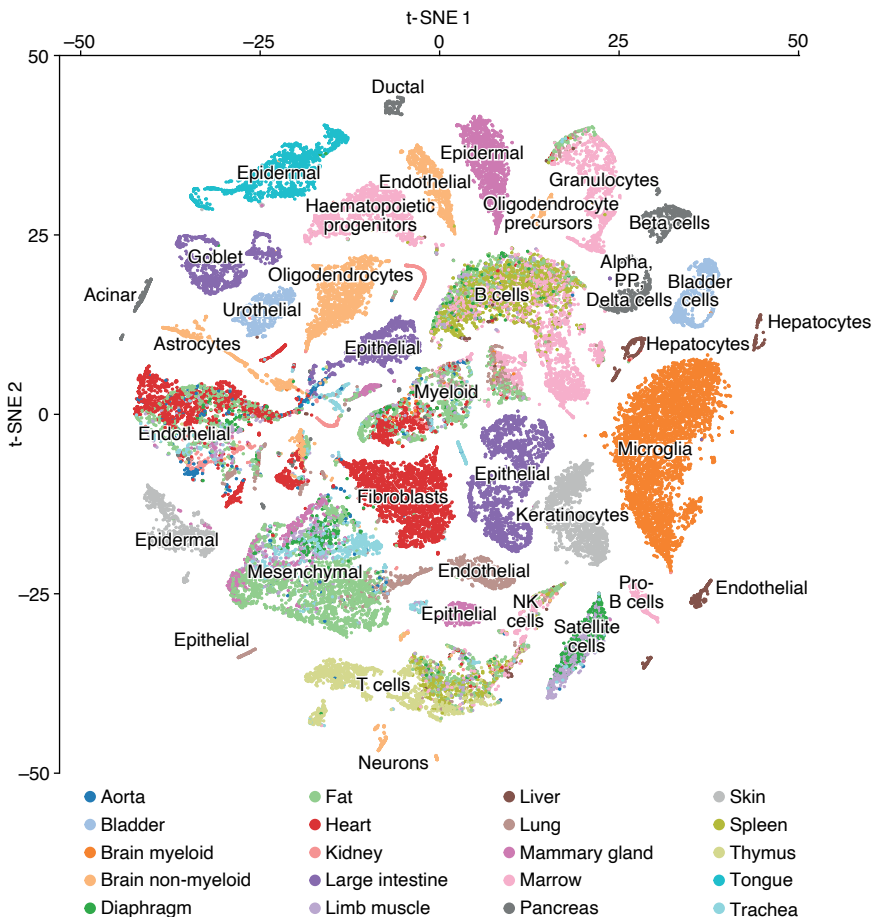


Figure 2: t-SNE visualization of all FACS cells. t-SNE plot of all cells collected by FACS, coloured by organ, overlaid with the predominant cell type composing each cluster; $n = 44,949$ individual cells.

Chapter 5

cells with t-SNE and grouped them with unbiased, graph-based clustering (Fig. 2, Extended Data Fig. 8). As expected, cells from different organs often mixed, with 25 of 54 clusters containing (at least five) cells from distinct organs (Fig. 3). For example, clusters 3 and 48 each contain endothelial cells from five or more organs, and clusters 1 and 24 contain mesenchymal and stromal cells from four or more organs. Cluster 2 contains B cells from fat, limb muscle, lung, spleen, marrow and liver, but also cells annotated as leukocytes and lymphocytes from the thymus, heart and limb muscle. This suggests that the effect of cell type on measured gene expression is stronger than the effect of batch or dissociation protocol.

Cluster co-membership alone, however, is insufficient to conclude that two cell populations from different organs represent the same or similar cell types; at any given resolution, unbiased clustering that groups related cells may also group unrelated cells (21). Therefore, to determine which clusters are composed of related or unrelated cell types, we computed a heterogeneity score for each cluster (Methods), and found low scores for the biologically sensible clusters discussed above (Extended Data Fig. 9). By contrast, the astrocytes and epithelial cells in cluster 53 are as different from one another as two random cells.

In addition to these heterogeneous groups, the clustering reveals small populations of potentially mislabelled cells inside homogenous populations. For example, ten thymus cells in cluster 3 (composed of 2,379 cells) are annotated as 'leukocytes', but they express *Pecam1*, which is an endothelial marker. This is a predictable artefact of the annotation scheme: because entire clusters, rather than individual cells, were annotated in each organ, a sufficiently rare cell type that was algorithmically grouped with a more populous cell type will be misannotated. This seems to occur only for populations smaller than about 30 cells, which comprise less than 4% of the overall dataset, and represents the lower limit of sensitivity in the current release of data interpretation.

The fact that most cells of similar cell types cluster together across organs and biological replicates shows that batch effects are not the main source of variance in the dataset. Our findings also show that manual annotation of cell types is consistent with unbiased transcriptomic clustering for sufficiently large populations. We expect that further development of multi-scale comparison algorithms will facilitate the discovery of both universal and organ-specific gene modules within these shared cell types.

To demonstrate an example of investigating common cell types across organs, we collectively analysed all FACS cells annotated as T cells, which revealed five clusters (Fig. 4). Cluster 0 comprises thymic cells undergoing VDJ recombination characterized by the expression of *Rag1*, *Rag2* and *Dnnt*, and includes uncommitted double-positive T cells (*Cd4+* and *Cd8a+*). Cluster 4 contains predominantly proliferating thymic T cells, which may represent pre-T cells expanding after VDJ recombination. Clusters 1-3 contain mostly single-positive T cells (*Cd4+* or *Cd8a+*). Cluster 3 contains *Cd5^{hi}* thymic T cells that are possibly undergoing positive selection, whereas Cluster 2 contains mostly non-thymic T cells expressing the high-affinity IL2 receptor (encoded

by the genes *Ii2ra* and *Ii2rb*), which suggests that they are activated. Notably, they also express MHC class II genes (*H2-Aa* and *H2-Ab1*). Although this is known in human T cells, MHC class II was previously thought to be restricted to professional antigen-presenting cells in mice²². Finally, Cluster 1 also represents mature T cells, but primarily splenic.



Figure 3: Comparison of cell-type determination. Comparison of cell-type determination as performed by unbiased whole-transcriptome comparison versus manual annotation of clusters by organ-specific experts. The x axis represents clusters from Fig. 2, while the y axis represents manual expert annotation of clusters derived from individual organs analysed independently (Extended Data Fig. 1). The unbiased method discovers relationships between similar cell types found in different organs; in particular T cells from different organs are grouped into a single cluster, B cells from different organs into a different single cluster, and endothelial cells from different organs into a single cluster (regions outlined in blue boxes).

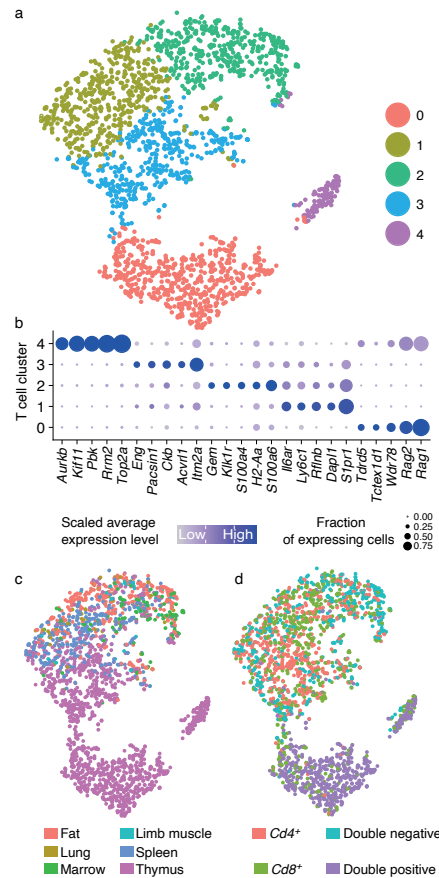


Figure 4: Analysis of all sorted T cells. (a) t-SNE plot of all T cells coloured by cluster membership, highlighting the five identified clusters; $n = 2,847$ individual cells. (b) Dot plot showing level of expression (colour scale) and number of expressing cells (point diameter) within each cluster of T cells. *Rflnb* is also known as *Fam101b*. (c) t-SNE plot of all T cells coloured by organ of origin (fat, lung, marrow, limb muscle, spleen or thymus); $n = 2,847$ individual cells. d, t-SNE plot of all T cells coloured by classification of T cells to four categories based on expression of Cd4 and Cd8 ($Cd4^+$, $Cd8^+$, $Cd4^+Cd8^+$, $Cd4^-Cd8^-$); $n = 2,847$ individual cells.

Global transcription factor analysis

One major goal of defining cell identities is to understand the underlying regulatory networks. We investigated how transcription factors contribute to cell-type identity by

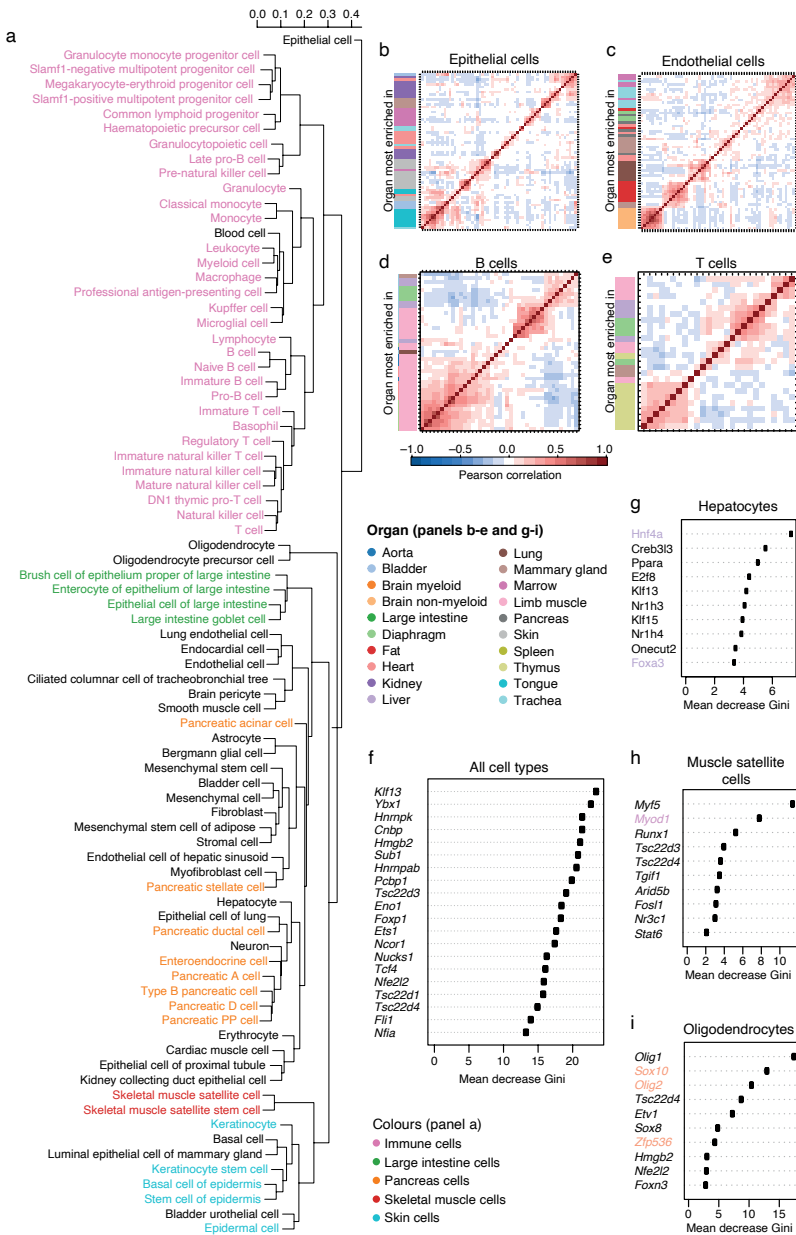


Figure 5: Transcription factor analysis. (a) Dendrogram of cell types constructed with only transcription factors. b-e, Correlograms of top organ-specific transcription factors for epithelial cells (b), endothelial cells (c), B cells (d) and T cells (e). Row colours correspond to the organ of the most-enriched cell type; n = 60 randomly selected cells for each cell type. f, Top 20 transcription factors (mean Gini importance) of the random-forest model when classifying all cell types. g-i, Top 10 transcription factors (mean Gini importance) of the random-forest model when classifying each cell type individually. The coloured genes correspond to transcription factors used in successful reprogramming protocols.

clustering averaged gene-expression profiles for each cell type using only the 1,016 transcription factors expressed in our dataset (Fig. 5a). The resulting dendrogram closely resembles the dendrogram produced using all expressed genes, indicating that transcription factors can be used to reconstruct known cell-ontology relationships between bulk populations (entanglement = 0.11; Extended Data Fig. 10a). By contrast, when we repeated the analysis using cell-surface markers, RNA splicing factors, or the two groups combined (equivalent to a random set of genes), the entanglement was 0.22, 0.25 and 0.34, respectively, which suggests that none of these molecular classes define cell type to the extent that transcription factors do.

We then analysed organ-specific transcription factors by performing correlation analysis on shared cell types between organs (23) (epithelial cells, endothelial cells, B cells and T cells; Fig. 5b–e, Extended Data Fig. 10b–i). To understand which transcription factors were most informative for specifying cell types, we performed variable selection using random forest models (Methods) and determined that 136 transcription factors are needed to simultaneously define all cell types across all organs (Fig. 5f, Supplementary Table 3). We then determined the transcription factor sets that distinguish each individual cell type from all other cells. These sets vary substantially in size (from 2 to 813 transcription factors) and are not necessarily unique to each cell type (Fig. 5g–i, Supplementary Table 4).

A possible application for such transcription factor networks is the design of reprogramming protocols. Indeed, the transcription factors used in published methods are found in the cell-type-specific transcription factors sets we discovered (Supplementary Table 5). For some cell types, such as hepatocytes, satellite cells and oligodendrocytes, those reprogramming factors are the top variables segregating cell types (Fig. 5g–i). In fact, for nearly all reprogramming protocols the transcription factors used also specified the targeted cell type in our data (Supplementary Table 5), which suggests that our data can inform novel reprogramming schemes.

DISCUSSION

A key challenge for single-cell studies is to understand transcriptomic changes caused by dissociation. A previous study showed that quiescent limb-muscle satellite cells activate upon dissociation and consequently express immediate early genes and other dissociation-related markers (24). We clearly observed these markers in several organs including limb muscle (Extended Data Fig. 11), but many showed little evidence of cellular activation. Therefore, the dissociation-related satellite-cell markers are not universal, and organs probably display unique dissociation-related expression profiles. Importantly, the presence of such changes in gene expression does not prevent the identification of cell type or the comparison of cell types across organs.

Another challenge for single-cell studies is experimental design amid the choice of several technologies. Droplet-based technologies offer certain advantages in the discovery of rare cell types or states, for example when many cells (tens of thousands) are required to reconstruct whole-organism architecture and developmental lineages (25, 26). FACS-based methods generate high coverage over small cellular populations

Chapter 5

(tens to thousands), and are beneficial for enriching specific or rare cell types, and for studying subtle heterogeneity involving lowly expressed genes (27), alternative splicing (15) and sequence variation analysis (28). There are opportunities to combine the two methods, such as by running sorted cells on a microfluidic-droplet platform, which could potentially accommodate both cell-type enrichment and cost factors.

Recently, a complementary scRNA-seq study across mouse organs was published (20). Those data contained four times as many cells and included several sample types not present in our data, such as neonatal and fetal organs, cell lines, and young adult ovary, peripheral blood, placenta, prostate, small intestine, stomach, testis and uterus. However, our FACS data contain four times as many genes per cell, and we analysed several organs not present in the other dataset (20), such as aorta, four brain regions, diaphragm, four fat types, four adult heart chambers, adult telogen and anagen skin, tongue and trachea. Additionally, several features of our study facilitate replication and cross-experiment analysis: all data, analysis and code are freely available; our web portal enables one to query gene expression in all organs simultaneously; we annotated cell types using standard cell ontology terms, thereby enabling cross-organ and cross-experiment analyses; age and sex are controlled in our data by collecting all organs from the same mice; both sexes are represented for all organs in our data; organs were perfused, enabling the analysis of tissue-resident immune cells; and full-length transcript data make possible transcription factor, splice variant, and sequence variant analyses.

In conclusion, we have created a compendium of single-cell transcriptional measurements across 20 mouse organs. This *Tabula Muris*, or 'Mouse Atlas', has many uses, including the discovery of new putative cell types, the discovery of novel gene expression in known cell types, and the ability to compare cell types across organs. It will also serve as a reference of healthy young adult organs, which can be used as a baseline for current and future mouse models of disease. Although it is not an exhaustive characterization of all mouse organs, it does provide a rich dataset of the most highly studied organs in biology. The *Tabula Muris* provides a framework and description of many of the most populous and important cell populations within the mouse, and represents a foundation for future studies across a multitude of diverse physiological disciplines.

METHODS

Mice and organ collection

Four 10-15 week old male and four virgin female C57BL/6JN mice were shipped from the National Institute on Aging colony at Charles River (housed at 67-73 °F) to the Veterinary Medical Unit (VMU; housed at 68-76 °F) at the VA Palo Alto (VA). At both locations, mice were housed on a 12-h light/dark cycle, and provided food and water ad libitum. The diet at Charles River was NIH-31, and Teklad 2918 at the VA VMU. Littermates were not recorded or tracked, and mice were housed at the VA VMU for no longer than 2 weeks before euthanasia. Before tissue collection, mice were placed in sterile collection chambers at 8 am for 15 min to collect fresh fecal pellets. After anaesthetization with 2.5% v/v Avertin, mice were weighed, shaved, and blood was drawn via cardiac puncture before transcatheter perfusion with 20 ml

PBS. Mesenteric adipose tissue was then immediately collected to avoid exposure to the liver and pancreas perfusate, which negatively affects cell sorting. Isolating viable single cells from both the pancreas and the liver of the same mouse was not possible; therefore, two males and two females were used for each. Whole organs were then dissected in the following order: large intestine, spleen, thymus, trachea, tongue, brain, heart, lung, kidney, gonadal adipose tissue, bladder, diaphragm, limb muscle (tibialis anterior), skin (dorsal), subcutaneous adipose tissue (inguinal pad), mammary glands (fat pads 2, 3 and 4), brown adipose tissue (interscapular pad), aorta and bone marrow (spine and limb bones). Organ collection concluded by 10 am. After single-cell dissociation as described below, cell suspensions were either used for FACS of individual cells into 384-well plates, or for preparation of the microfluidic droplet library. All animal care and procedures were carried out in accordance with institutional guidelines approved by the VA Palo Alto Committee on Animal Research.

Tissue dissociation and sample preparation

Specific protocols for each tissue are described in the Supplementary Information.

Sample size, randomization and blinding

No sample size choice was performed before the study. Randomization and blinding were not performed: the authors were aware of all data and metadata-related variables during the entire course of the study.

SINGLE-CELL METHODS

Lysis plate preparation

Lysis plates were created by dispensing 0.4 µl lysis buffer (0.5 U Recombinant RNase Inhibitor (Takara Bio, 2313B), 0.0625% Triton™ X-100 (Sigma, 93443-100ML), 3.125 mM dNTP mix (Thermo Fisher, R0193), 3.125 µM Oligo-dT30VN (Integrated DNA Technologies, 5'AAGCAGTGGTATCAACGCAGAGTACT30VN-3') and 1:600,000 ERCC RNA spike-in mix (Thermo Fisher, 4456740)) into 384-well hard-shell PCR plates (Bio-Rad HSP3901) using a Tempest liquid handler (Formulatrix). 96-well lysis plates were also prepared with 4 µl lysis buffer. All plates were sealed with AlumaSeal CS Films (Sigma-Aldrich Z722634) and spun down (3,220g, 1 min) and snap-frozen on dry ice. Plates were stored at -80 °C until sorting.

FACS

After dissociation, single cells from each organ and tissue were isolated into 384- or 96-well plates via FACS. Most organs were sorted into 384-well plates using SH800S (Sony) sorters. Heart and liver were sorted into 96-well plates and cardiomyocytes were hand-picked into 96-well plates. Limb muscle and diaphragm were sorted into 384-well plates on an Aria III (Becton Dickinson) sorter. The last two columns of each 384 well plate were intentionally left as blanks. For most organs, single cells were selected with forward scatter, and dead cells and common cell types were excluded with a single colour channel. Combinations of fluorescent antibodies were used for most organs to enrich for rare cell populations (see Supplementary Information), but some were stained only for viable cells. Colour compensation was used whenever necessary. On the SH800, the highest purity setting ('Single cell') was used for all but the rarest cell types, for which the 'Ultrapure' setting was used. Sorters were calibrated

Chapter 5

using FACS buffer every day before collecting any cells, and also after every eight sorted plates. For a typical sort, 1–3 ml of pre-stained cell suspension was filtered, vortexed gently, and loaded onto the FACS machine. A small number of cells were flowed at low pressure to check cell and debris concentrations. The pressure was then adjusted, flow paused, the first destination plate unsealed and loaded, and sorting started. If a cell suspension was too concentrated, it was diluted using FACS buffer or 1X PBS. For some cell types, such as hepatocytes, 96-well plates were used because it was not possible to sort individual cells accurately into 384-well plates. Immediately after sorting, plates were sealed with a pre-labelled aluminium seal, centrifuged, and flash frozen on dry ice. On average, each 384-well plate took 8 min to sort.

cDNA synthesis and library preparation

cDNA synthesis was performed using the Smart-seq2 protocol (7, 8). In brief, 384-well plates containing single-cell lysates were thawed on ice followed by first-strand synthesis. 0.6 μl of reaction mix (16.7 U μl^{-1} SMARTScribe Reverse Transcriptase (Takara Bio, 639538), 1.67 U μl^{-1} Recombinant RNase Inhibitor (Takara Bio, 2313B), 1.67X First-Strand Buffer (Takara Bio, 639538), 1.67 μM TSO (Exiqon, 5'-AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG-3'), 8.33 mM dithiothreitol (Bioworld, 40420001-1), 1.67 M Betaine (Sigma, B0300-5VL) and 10 mM MgCl_2 (Sigma, M1028-10X1ML)) was added to each well using a Tempest liquid handler. Reverse transcription was carried out by incubating wells on a ProFlex 2 \times 384 thermal-cycler (Thermo Fisher) at 42 °C for 90 min, and stopped by heating at 70 °C for 5 min.

Subsequently, 1.5 μl of PCR mix (1.67X KAPA HiFi HotStart ReadyMix (Kapa Biosystems, KK2602), 0.17 μM IS PCR primer (IDT, 5'-AAGCAGTGGTAT CAACGCAGAGT-3'), and 0.038 U μl^{-1} Lambda Exonuclease (NEB, M0262L)) was added to each well with a Mantis liquid handler (Formulatrix), and second-strand synthesis was performed on a ProFlex 2x384 thermal-cycler by using the following program: 1) 37 °C for 30 min, 2) 95 °C for 3 min, 3) 23 cycles of 98 °C for 20 s, 67 °C for 15 s and 72 °C for 4 min, and 4) 72 °C for 5 min.

The amplified product was diluted with a ratio of 1 part cDNA to 10 parts 10 mM Tris-HCl (Thermo Fisher, 15568025), and concentrations were measured with a dye-fluorescence assay (Quant-iT dsDNA High Sensitivity kit; Thermo Fisher, Q33120) on a SpectraMax i3x microplate reader (Molecular Devices). Sample plates were selected for downstream processing if the mean concentration of blanks (ERCC-containing, non-cell wells) was greater than 0 ng μl^{-1} , and, after linear regression of the values obtained from the Quant-iT dsDNA standard curve, the R² value was greater than 0.98. Sample wells were then selected if their cDNA concentrations were at least one standard deviation greater than the mean concentration of the blanks. These wells were reformatted to a new 384-well plate at a concentration of 0.3 ng μl^{-1} and a final volume of 0.4 μl using an Echo 550 acoustic liquid dispenser (Labcyte).

Illumina sequencing libraries were prepared as described previously (14). In brief, tagmentation was carried out on double-stranded cDNA using the Nextera XT Library Sample Preparation kit (Illumina, FC-131-1096). Each well was mixed with 0.8 μl Nextera tagmentation DNA buffer (Illumina) and 0.4 μl Tn5 enzyme (Illumina), then

incubated at 55 °C for 10 min. The reaction was stopped by adding 0.4 µl Neutralize Tagment Buffer (Illumina) and centrifuging at room temperature at 3,220g for 5 min. Indexing PCR reactions were performed by adding 0.4 µl of 5 µM i5 indexing primer, 0.4 µl of 5 µM i7 indexing primer, and 1.2 µl of Nextera NPM mix (Illumina). PCR amplification was carried out on a ProFlex 2x384 thermal cycler using the following program: 1) 72 °C for 3 min, 2) 95 °C for 30 s, 3) 12 cycles of 95 °C for 10 s, 55 °C for 30 s and 72 °C for 1 min, and 4) 72 °C for 5 min.

Library pooling, quality control and sequencing

After library preparation, wells of each library plate were pooled using a Mosquito liquid handler (TTP Labtech). Pooling was followed by two purifications using 0.7x AMPure beads (Fisher, A63881). Library quality was assessed using capillary electrophoresis on a Fragment Analyzer (AATI), and libraries were quantified by qPCR (Kapa Biosystems, KK4923) on a CFX96 Touch Real-Time PCR Detection System (Biorad). Plate pools were normalized to 2 nM and equal volumes from 10 or 20 plates were mixed together to make the sequencing sample pool. A PhiX control library was spiked in at 0.2% before sequencing.

Sequencing libraries from 384-well and 96-well plates

Libraries were sequenced on the NovaSeq 6000 Sequencing System (Illumina) using 2 × 100-bp paired-end reads and 2 × 8-bp or 2 × 12-bp index reads with either a 200- or 300-cycle kit (Illumina, 20012861 or 20012860).

Microfluidic droplet single-cell analysis

Single cells were captured in droplet emulsions using the GemCode Single-Cell Instrument (10x Genomics), and scRNA-seq libraries were constructed as per the 10x Genomics protocol using GemCode Single-Cell 3' Gel Bead and Library V2 Kit. In brief, single cell suspensions were examined using an inverted microscope, and if sample quality was deemed satisfactory, the sample was diluted in PBS with 2% FBS to a concentration of 1000 cells per µl. If cell suspensions contained cell aggregates or debris, two additional washes in PBS with 2% FBS at 300g for 5 min at 4 °C were performed. Cell concentration was measured either with a Moxi GO II (Orfio Technologies) or a haemocytometer. Cells were loaded in each channel with a target output of 5,000 cells per sample. All reactions were performed in the Biorad C1000 Touch Thermal cycler with 96-Deep Well Reaction Module. 12 cycles were used for cDNA amplification and sample index PCR. Amplified cDNA and final libraries were evaluated on a Fragment Analyzer using a High Sensitivity NGS Analysis Kit (Advanced Analytical). The average fragment length of 10x cDNA libraries was quantitated on a Fragment Analyzer (AATI), and by qPCR with the Kapa Library Quantification kit for Illumina. Each library was diluted to 2 nM, and equal volumes of 16 libraries were pooled for each NovaSeq sequencing run. Pools were sequenced with 100 cycle run kits with 26 bases for Read 1, 8 bases for Index 1, and 90 bases for Read 2 (Illumina 20012862). A PhiX control library was spiked in at 0.2 to 1%. Libraries were sequenced on the NovaSeq 6000 Sequencing System (Illumina).

Data processing

Sequences from the NovaSeq were de-multiplexed using bcl2fastq version

Chapter 5

2.19.0.316. Reads were aligned using to the mm10plus genome using STAR version 2.5.2b with parameters TK. Gene counts were produced using HTSEQ version 0.6.1p1 with default parameters, except 'stranded' was set to 'false', and 'mode' was set to 'intersection-nonempty'. Sequences from the microfluidic droplet platform were de-multiplexed and aligned using CellRanger version 2.0.1, available from 10x Genomics with default parameters.

Clustering

Standard procedures for filtering, variable gene selection, dimensionality reduction and clustering were performed using the Seurat package version 2.2.1. A detailed worked example, including the mathematical formulae for each operation, is in the Organ Annotation Vignette. The parameters that were tuned on a per-tissue basis (resolution and number of principal components (PCs)) can be viewed in the tissue-specific Rmd files available on GitHub. For each tissue and each sequencing method (FACS and microfluidic droplet), the following steps were performed:

1. Cells were lexicographically sorted by cell ID to ensure reproducibility.
2. Cells with fewer than 500 detected genes were excluded. (A gene counts as detected if it has at least one read mapping to it). Cells with fewer than 50,000 reads (FACS) or 1,000 UMI (microfluidic droplet) were excluded.
3. Counts were log-normalized for each cell using the natural logarithm of 1 + counts per million (for FACS) or 1 + counts per ten thousand (for microfluidic droplet).
4. Variable genes were selected using a threshold (0.5) for the standardized log dispersion, in which the standardization was performed separately according to binned values of log mean expression.
5. The variable genes were projected onto a low-dimensional subspace using principal component analysis. The number of principal components was selected on the basis of inspection of the plot of variance explained.
6. A shared-nearest-neighbours graph was constructed on the basis of the Euclidean distance in the low-dimensional subspace spanned by the top principal components. Cells were clustered using a variant of the Louvain method that includes a resolution parameter in the modularity function¹³.
7. Cells were visualized using a 2-dimensional t-distributed Stochastic Neighbour Embedding of the PC-projected data.
8. Cell types were assigned to each cluster using the abundance of known marker genes. Plots showing the expression of the markers for each tissue appear in the Extended Data.
9. When clusters appeared to be mixtures of cell types, they were refined either by increasing the resolution parameter for clustering or subsetting the data and rerunning steps 3–7.

Asimilar analysis was done globally for all FACS-processed cells and for all microfluidic-droplet-processed cells to produce an unbiased clustering.

Heterogeneity score.

Let C be a cluster, decomposed into annotated cell types $C=T_1U\cdots UT_kC=T_1U\cdots UT_k$. For each pair of cell types T_i, T_j , we compute the average distance between their members: $d_{ij} = \frac{1}{|T_i||T_j|} \sum_{x \in T_i, y \in T_j} |x-y|$. The heterogeneity score C is the maximum of those distances over cell types i with at least five cells. For the FACS data, the

vector x for a cell is the PC-projection from step 5 above. Extended Data Fig. 9 contains heat maps of the cell-type distance matrix d_{ij} for select clusters and a bar plot of the heterogeneity scores for all clusters containing several cell types.

Differential expression overlap analysis

For FACS and microfluidic droplet data, differential expression analysis for each organ was performed using a Wilcoxon rank-sum test as implemented in the 'FindAllMarkers' function of the Seurat package. Differential expression was performed between cell ontology groups and resulted in a list of differentially expressed genes ($\ln(\text{FoldChange}) > 0.25$) between each cell ontology group and all other ontology groups of the same organ. For microwell-seq we used the corresponding published lists for each cell type and for every organ. We then assessed the overlap of those lists between the three methods. As the nomenclature is not identical, the analysis was performed between cell types that could be matched with a certain degree of confidence between the three methods (Supplementary Table 2).

Correlating bulk gene expression profiles

For the 33 cell populations shared between FACS and microfluidic droplets, the average gene-expression profile of each population was calculated. The quality of such a bulk gene-expression profile depends on the total number of detected molecules. FACS detects more molecules per cell, but fewer cells. Microfluidic droplets detect fewer molecules per cell, but more cells. To assess the agreement between methods on annotated cell types, Pearson correlation was used on the log expression profiles of each shared cell population. (Only genes present at 1 count per million or greater in at least one of the datasets were considered. A pseudocount of 1 count per million was added before taking logarithms.)

Calculation of dissociation scores

For each organ, principal component analysis was performed on a subset of 140 dissociation-related genes (23). The first principal component was used as the 'dissociation score' as it corresponds to the variance within these genes.

Defining cell type-enriched transcription factors

Transcription factors were defined as the 1,140 genes annotated by the Gene Ontology term 'DNA binding transcription factor activity', downloaded from the Mouse Genome Informatics database (<http://www.informatics.jax.org.laneproxy.stanford.edu/mgihome/GO/project.shtml>, accessed on 10 November 2017). Cell types were defined as unique combinations of cell ontology and organ annotation (for example, Lung_Endothelial_cell). All analyses were performed on the full dataset, except the correlograms for which the data was subsampled by randomly selecting 60 cells from each cell type. Enriched transcription factors were defined by the Seurat FindMarkers function with the Wilcoxon significance test for the target cell type against the all of the rest of the cell types combined. These were filtered by $p_{\text{val}} < 10^{-3}$, $\text{avg_diff} > 0.2$, $\text{pct.1} - \text{pct.2} > 0.1$ (per cent detected difference > 0.1), and $\text{pct.1} > 0.3$ (detected in $> 30\%$ of target cells).

Cell-type comparisons between methods using cell ontology classes

We used the OntologyX R package family version 2.4 (libraries `ontologyIndex`, `ontologyPlot`, and `ontologySimilarity`) to draw the representative cell ontology dendrograms (function `onto_plot`). To compute the tanglegram (function `tanglegram` from `dendextend` R package version 1.8) we used the dendrogram created from all expressed genes as the reference for comparisons to the dendrograms produced using particular gene ontology cellular functions (transcription factors, cell surface markers, RNA splicing factors). The entanglement scores were calculated using the `step2side` method (function `untangle` from `dendextend` R package). Entanglement is a measure of alignment between two dendrograms. The entanglement score ranges from 0 (exact alignment) to 1 (no alignment)²⁹.

Defining transcription factor networks with random forests

We used random forests (a classifier that combines many single decision trees) to calculate the importance of each gene for defining cell types³⁰. The `varSelRF` R package version 0.7-8 uses the out-of-bag error as the minimization criterion and carries out variable elimination with random forests by successively eliminating the least important variables (with importance as returned from the random forest analysis). The algorithm iteratively fits random forests, at each iteration building a new forest after discarding those variables (genes) with the smallest variable importance; the selected set of genes is the one that yields the smallest out-of-bag error rate. This leads to the selection of small sets of non-redundant variables.

Code availability

All code used for analysis is available on GitHub (<https://github.com/czbiohub/tabula-muris>).

ACKNOWLEDGEMENTS

We thank Sony Biotechnology for making an SH800S instrument available for this project. Some of the cell sorting/flow cytometry analysis for this project was performed using a Sony SH800S instrument in the Stanford Shared FACS Facility. Some FACS experiments used instruments in the VA Flow Cytometry Core, which is supported by the US Department of Veterans Affairs, Palo Alto Veterans Institute for Research and the National Institutes of Health. This work was supported by the Chan Zuckerberg Biohub, NIH Grant DP1 AG053015 and the NOMIS Foundation (T.W.-C.) as well as partly by the Stanford Islet Research Core in the Stanford Diabetes Research Center (P30 DK116074). We thank A. McGeever for contributions to the design of the Tabula Muris web portal.

CONTRIBUTIONS

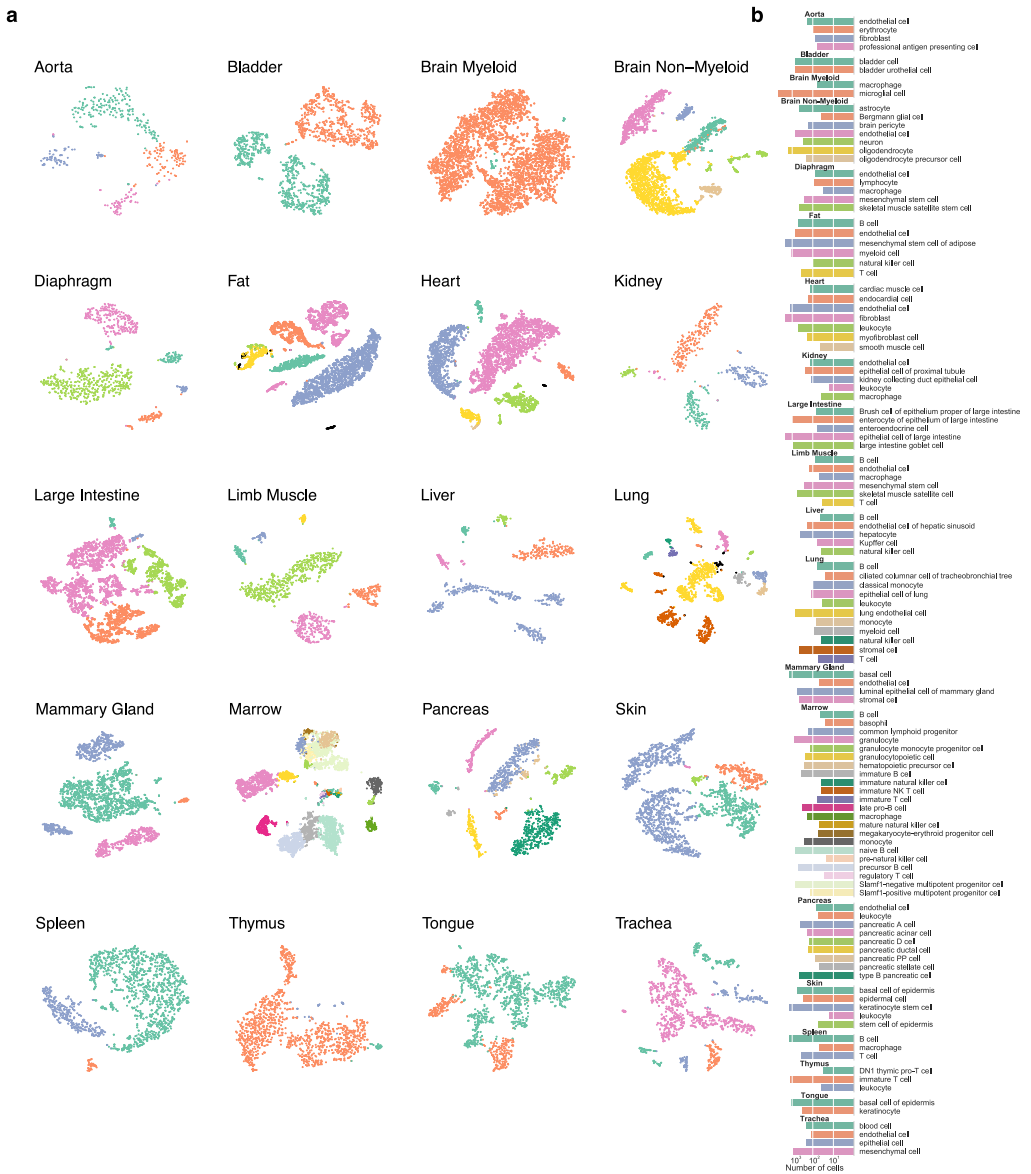
An overview of contributions is available online.

REFERENCES

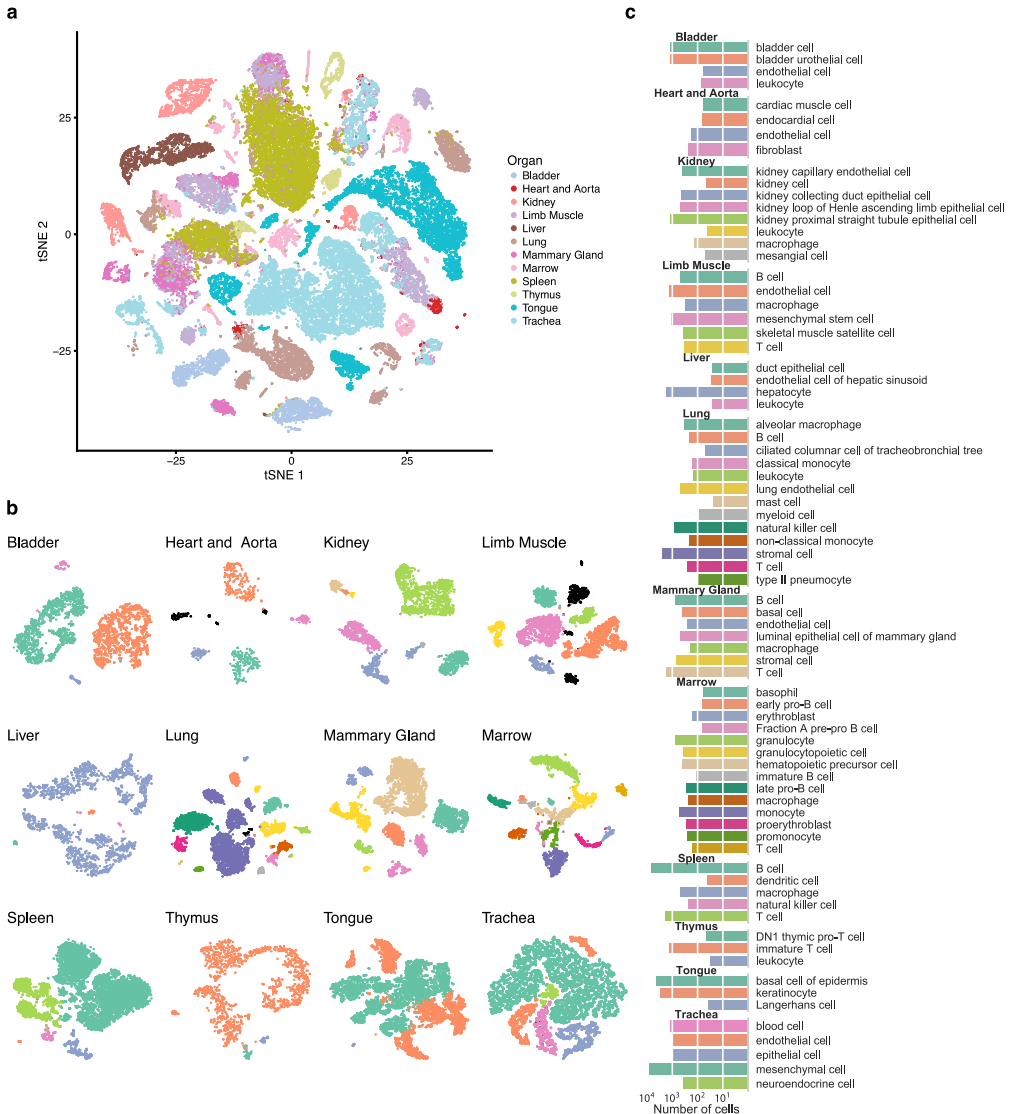
1. Alberts, B. et al. *Essential Cell Biology* (W.W. Norton & Company, New York, 2016).
2. Guo, G. et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 18, 675-685 (2010).
3. Dalerba, P. et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* 29, 1120-1127 (2011).
4. Thorsen, T., Roberts, R. W., Arnold, F. H. & Quake, S. R. Dynamic pattern formation in a vesicle-generating microfluidic device. *Phys.*

- Rev. Lett. 86, 4163–4166 (2001).
5. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214 (2015).
 6. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201 (2015).
 7. Ramsköld, D. et al. Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* 30, 777–782 (2012).
 8. Wu, A. R. et al. Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* 11, 41–46 (2014).
 9. Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509, 371–375 (2014).
 10. Enge, M. et al. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell* 171, 321–330.e14 (2017).
 11. Halpern, K. B. et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* 542, 352–356 (2017).
 12. Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* 551, 333–339 (2017).
 13. Villani, A.-C. et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 356, eaah4573 (2017).
 14. Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* 112, 7285–7290 (2015).
 15. Gokce, O. et al. Cellular taxonomy of the mouse striatum as revealed by single-cell RNA-seq. *Cell Rep.* 16, 1126–1137 (2016).
 16. Usoskin, D. et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* 18, 145–153 (2015).
 17. Zeisel, A. et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347, 1138–1142 (2015).
 18. Li, H. et al. Classifying *Drosophila* olfactory projection neuron subtypes by single-cell RNA sequencing. *Cell* 171, 1206–1220.e22 (2017).
 19. Bakken, T. et al. Cell type discovery and representation in the era of high-content single cell phenotyping. *BMC Bioinformatics* 18 (Suppl 17), 559 (2017).
 20. Han, X. et al. Mapping the mouse cell atlas by microwell-seq. *Cell* 172, 1091–1107.e17 (2018).
 21. Freytag, S., Tian, L., Lonnstedt, I., Ng, M. & Bahlo, M. Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data. *F1000Res* 7, 1297 (2018).
 22. Holling, T. M., Schooten, E. & van Den Elsen, P. J. Function and regulation of MHC class II molecules in T-lymphocytes: of mice and men. *Hum. Immunol.* 65, 282–290 (2004).
 23. Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* 74, 016110 (2006).
 24. van den Brink, S. C. et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* 14, 935–936 (2017).
 25. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* 556, 108–112 (2018).
 26. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357, 661–667 (2017).
 27. Liu, Z. et al. Single-cell transcriptomics reconstructs fate conversion from fibroblast to cardiomyocyte. *Nature* 551, 100–104 (2017).
 28. Darmanis, S. et al. Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep.* 21, 1399–1410 (2017).
 29. Kassambara, A. Practical guide to cluster analysis in R: unsupervised machine learning 1st edn (CreateSpace, North Charleston, 2017).
 30. Díaz-Uriarte, R. & Alvarez de Andrés, S. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 7, 3 (2006).

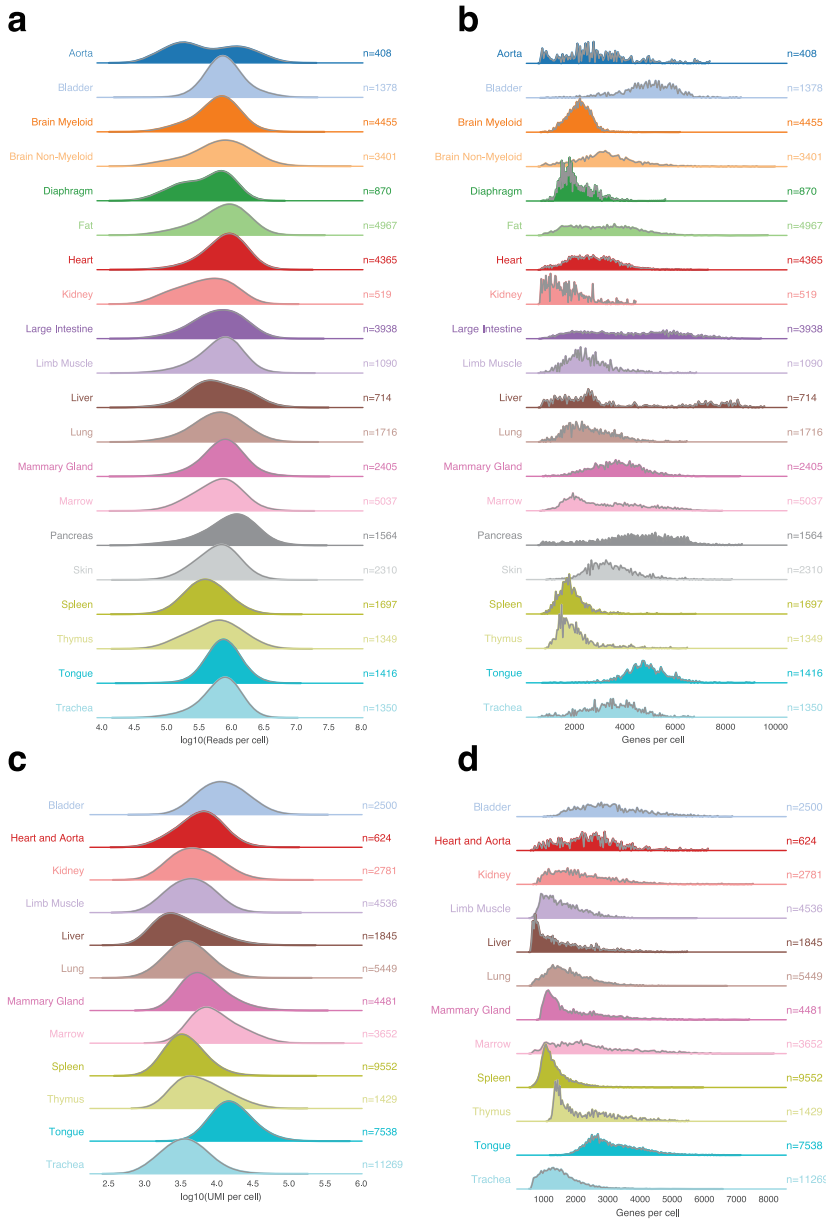
EXTENDED DATA FIGURES AND TABLES



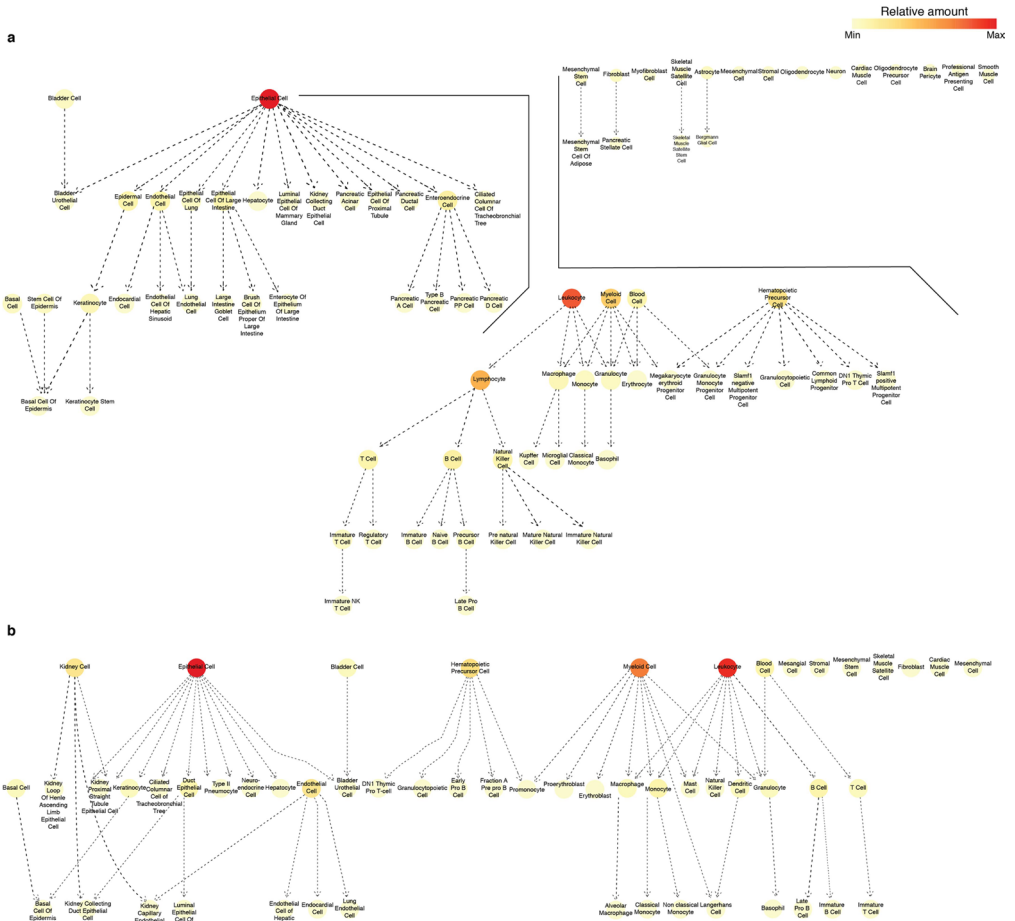
Extended Data Figure 1: The number and type of FACS cells that compose each organ. (a) Cells for each organ visualized with t-SNE, coloured by cell type. Cell types were determined by differential gene expression of known markers between clusters. (b) Bar plots quantifying the number of each annotated cell type. Cell type colours match their respective t-SNE plot.



Extended Data Figure 2: The number and type of microfluidic cells that compose each organ. (a) t-SNE plot of all cells collected by the microfluidic-droplet method, coloured by organ, overlaid with the predominant cell type that composes each cluster. (b) Cells for each organ visualized with t-SNE, coloured by cell type. Cell types were determined by differential gene expression of known markers between clusters. (c) Bar plots quantifying the number of each annotated cell type. Cell type colours match their respective t-SNE plot.

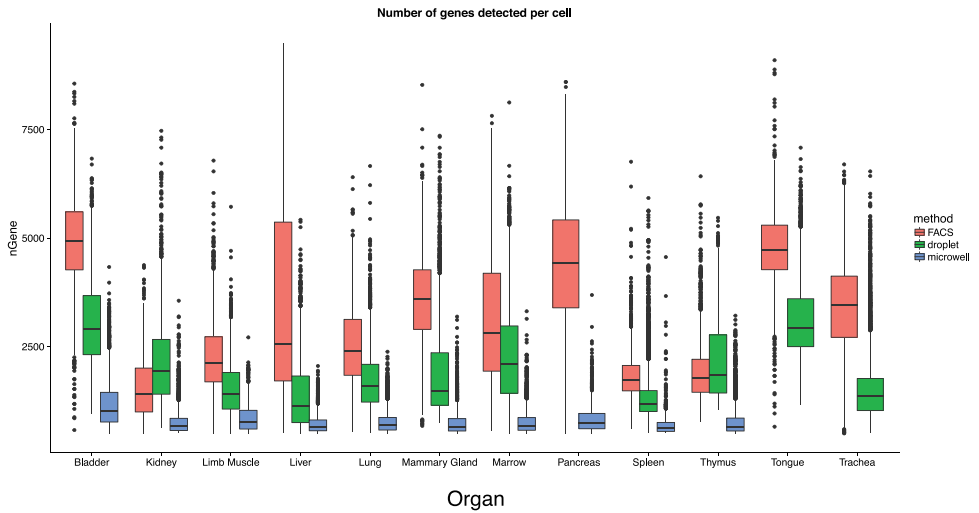


Extended Data Figure 3: The number of reads, UMIs and genes detected per cell for each organ. (a, c) Histograms for each organ of the number of reads per cell (FACS) (a) and UMIs per cell (microfluidic droplet) (c). (b, d) Histogram of the number of genes detected per cell for each organ from the FACS method (b), and the microfluidic-droplet method (d).

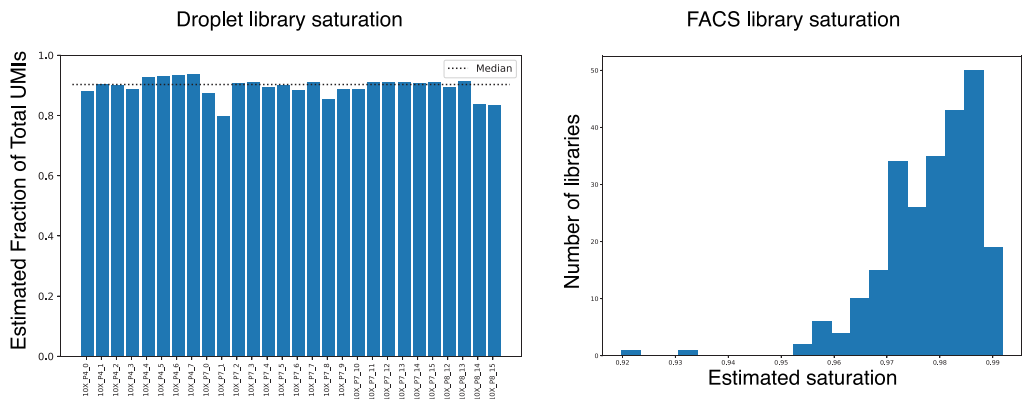


Extended Data Figure 4: Graphical representation of cell ontology class representation. (a, b) Datasets from the FACS method (a) and the microfluidic-droplet method (b), coloured by the relative amount of each cell type in each dataset.

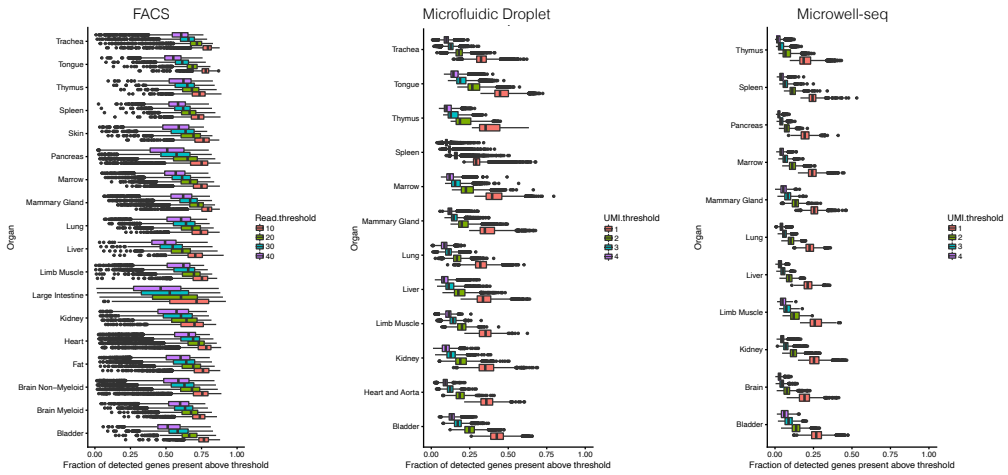
a



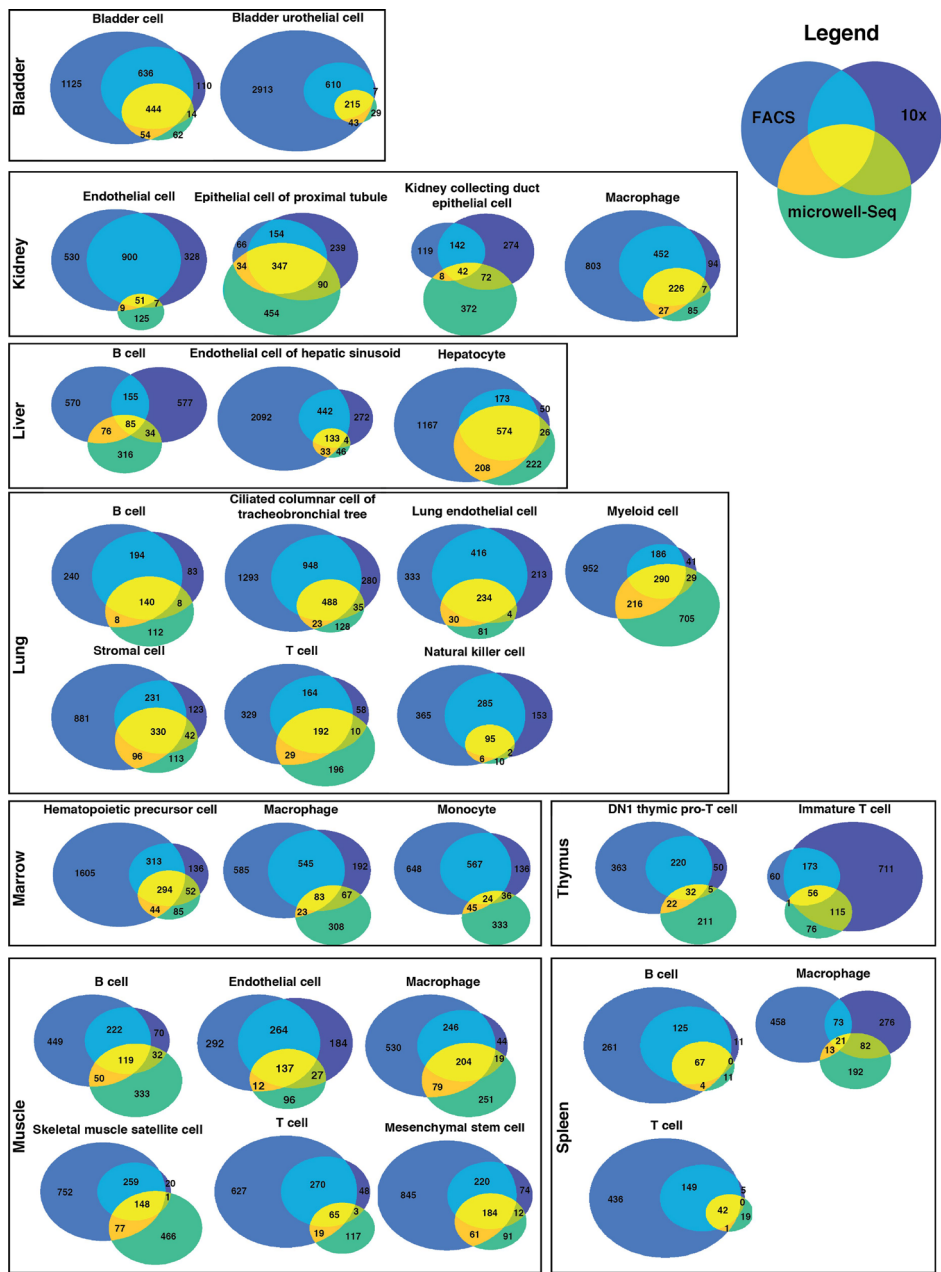
b



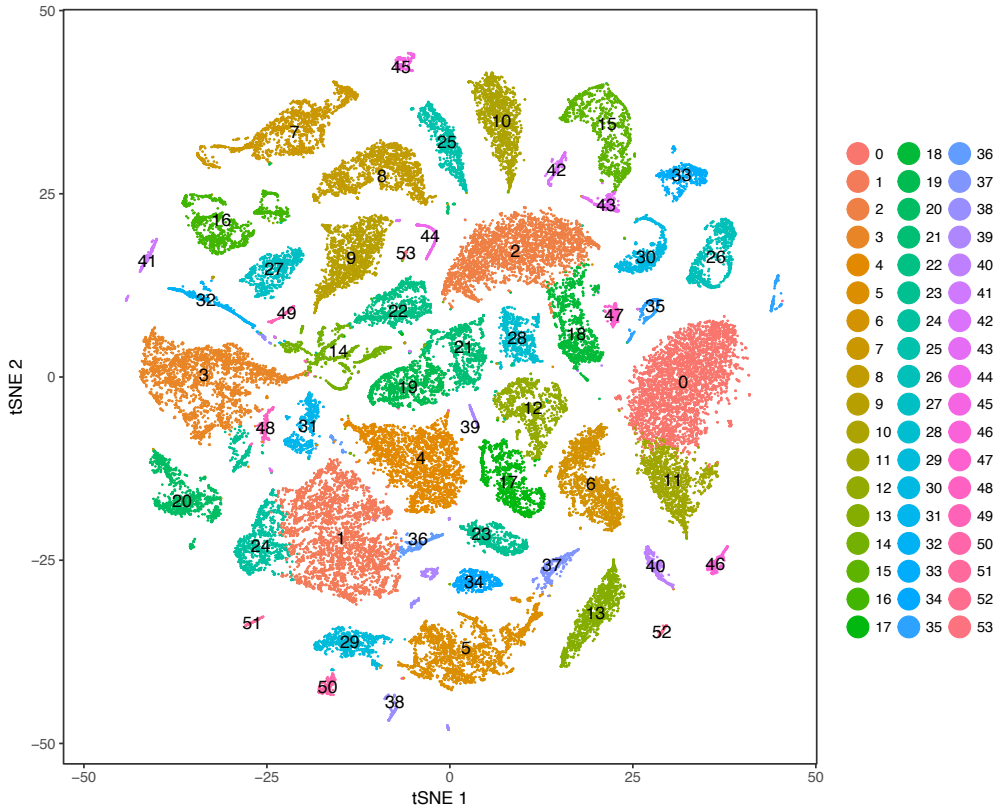
Extended Data Figure 5: Methodological comparison of detected genes and library saturation. (a) The number of genes detected (threshold of >0 reads or UMIs per cell) by FACS (red; $n = 21,105$ individual cells), microfluidic-droplet (green; $n = 55,032$ individual cells) and microwell-seq (blue; $n = 25,891$ individual cells) methods. (b) Library saturation fraction for all microfluidic-droplet libraries. Dotted horizontal line demarcates the median saturation (around 0.9). (c) Library saturation for all FACS libraries. Saturation was calculated using the number of detected genes while downsampling the number of reads per library. Summary statistics are contained in Supplementary Table 6



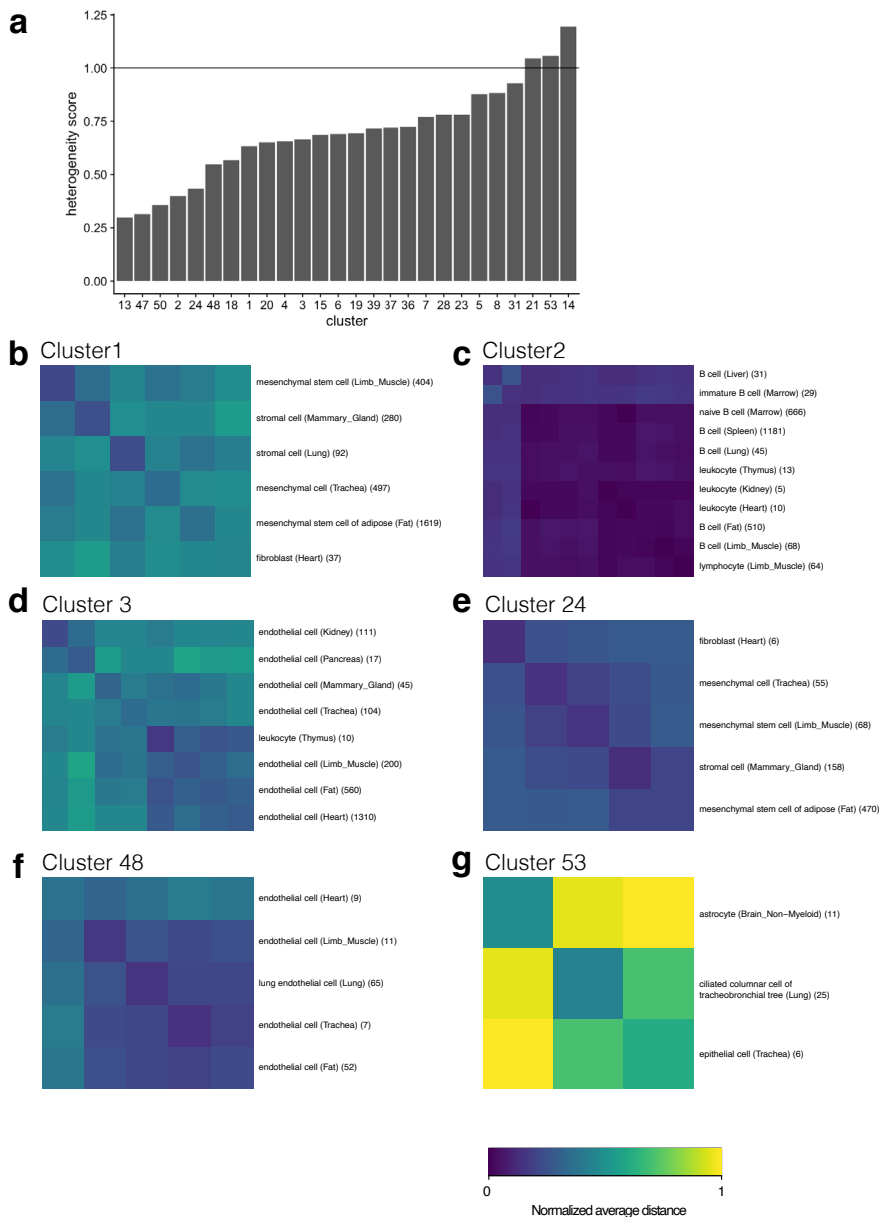
Extended Data Figure 6: The number of detected genes decreases similarly across organs as the read or UMI threshold is increased. Fraction of all detected genes (defined as >0 reads or UMIs) for each cell, across all organs, detected at increasing read or UMI thresholds for FACS (left; n = 44,949 individual cells), microfluidic-droplet (middle; n = 55,656 individual cells), and microwell-seq (right; n = 28,372 individual cells) methods. Summary statistics are contained in Supplementary Table 6



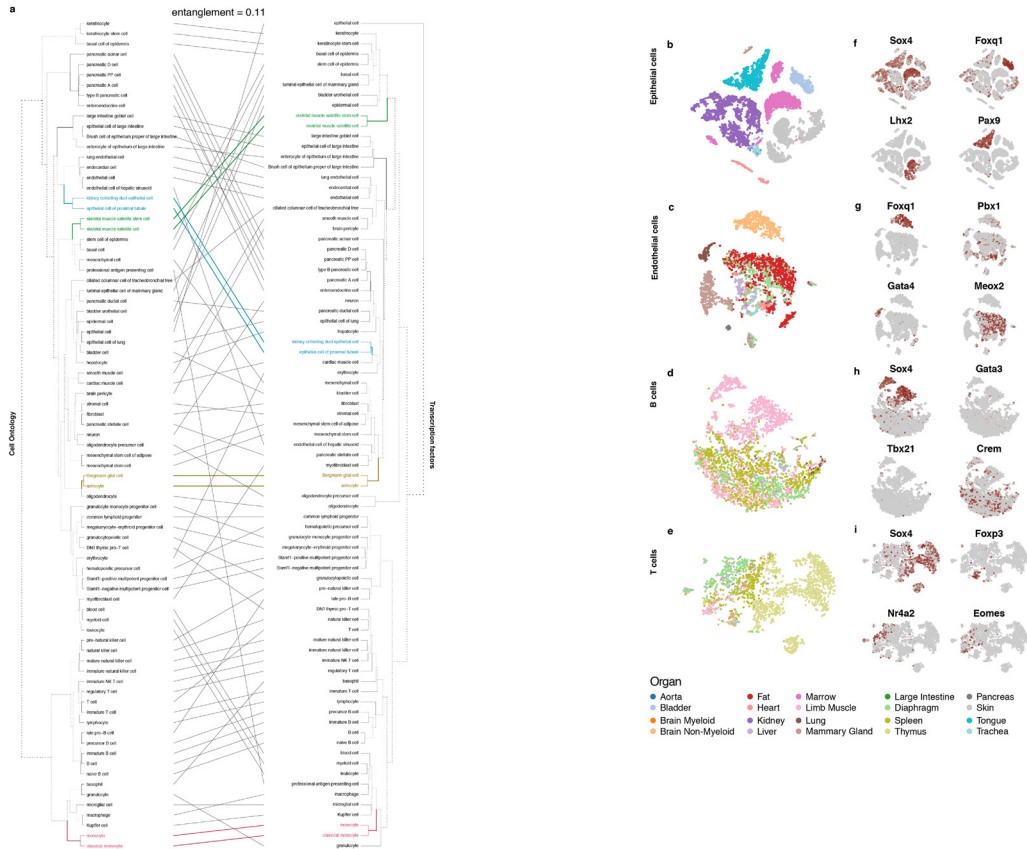
Extended Data Figure 7: The number of differentially expressed genes for each cell type that are common between methods. Venn diagrams showing the overlap between differentially expressed genes for each common cell type across the three methods (FACS, microfluidic-droplet and microwell-seq). Plotted data are provided in tabular form in Supplementary Table 2.



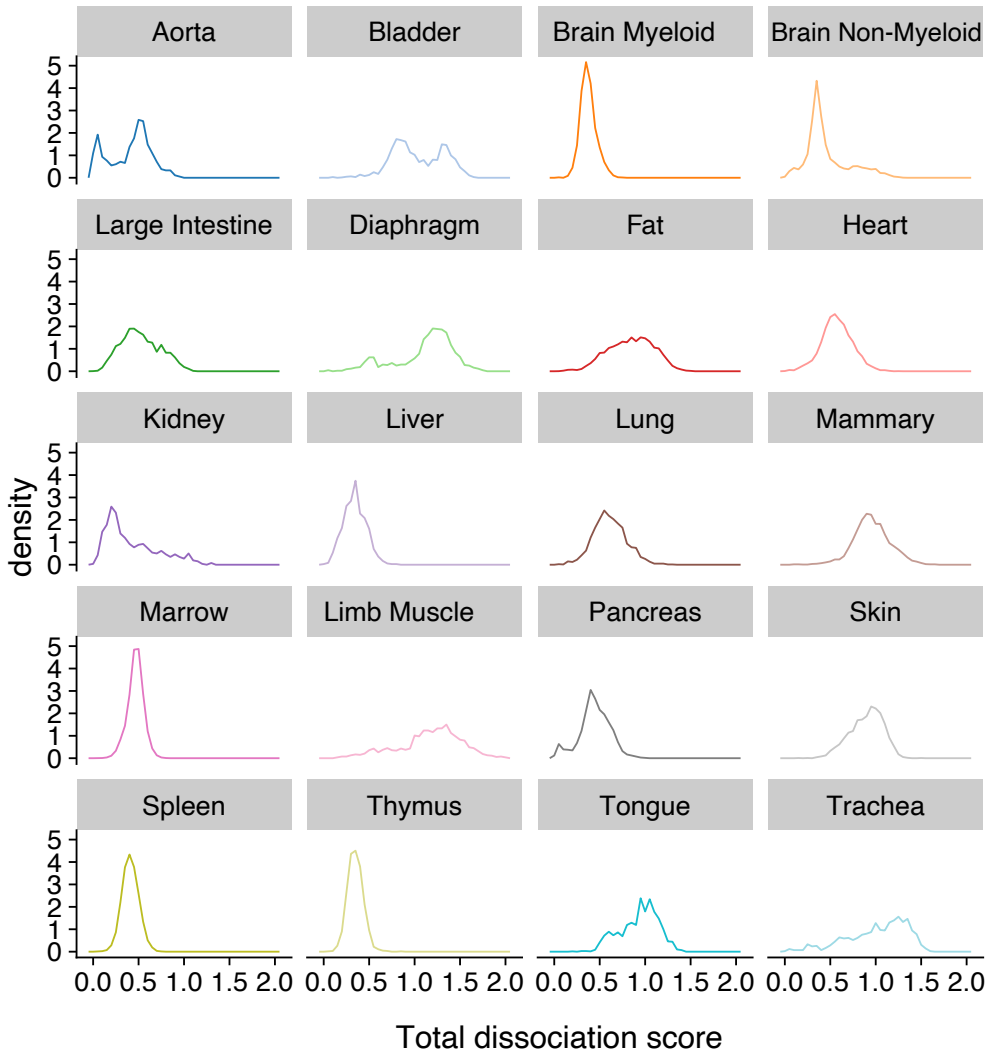
Extended Data Figure 8: t-SNE visualization of all FACS cells by cluster ID. n = 44,949 individual cells. Clusters are discussed in the text and further analysed in Fig. 3.



Extended Data Figure 9: Metrics of cluster heterogeneity. (a) Bar plot showing the heterogeneity score for each cluster containing several cell types. b-g, Heat maps showing the average between-cell-type distances within select clusters, normalized so that the average distance between pairs of FACS cells is 1, clipped to a max of 1, for clusters 1 (b), 2 (c), 3 (d), 24 (e), 48 (f) and 53 (g).



Extended Data Figure 10: Contribution of transcription factors to cell identity. (a) Tanglegram contrasting the dendrogram obtained using all expressed genes with one obtained using only the expression of transcription factors. The solid lines indicate segments that did not change position during the alignment between the two trees, and the dotted lines correspond to dendrogram branches reordered during the entanglement calculations. The colours indicate the branches for which identical leaves are aligned in both dendrograms. b-e, t-SNE visualization of epithelial (b), endothelial (c), B cells (d) and T cells (e), coloured by organ. f-i, t-SNE visualization of epithelial (f), endothelial (g) B cell (h) and T cell (i) expression of select transcription factors (from grey, low, to red, high). In b-i, n = 60 randomly selected cells for each cell type.



Extended Data Figure 11: Dissociation-induced gene-expression scores for each organ analysed with FACS. The dissociation score for each organ represents the magnitude of the first principal component of the 140 dissociation-associated genes from ref. 24. The y axis shows the probability density of the normalized histogram.

Supplementary information is available online.

Supplementary Table 1

Number of cells belonging to each annotated cell type across all organs for FACS and microfluidic droplets.

Supplementary Table 2

Cell type comparisons and lists of differentially expressed genes common between methods (FACS, droplet, microwell-Seq).

Supplementary Table 3

Random forest results for simultaneously defining all cell types with TF expression. This table consists of 4 tabs. The first tab summarizes the most important variables for the model. The second tab indicates all the unique cell types being classified. To build this model we used a data subset of 10 cells for each of the unique cell types. The third tab contains the classification confusion matrix and the fourth tab the summary of the average classification error per cell type.

Supplementary Table 4

Random forest results for defining each individual cell type with TF expression. In this model, each unique cell type (112 total) was compared to all other cell types. The numbers are the mean decrease Gini score. The value NA in each cell means that the transcription factor corresponding to that row does not contribute the classification model of the cell corresponding to that column.

Supplementary Table 5

Literature review of the current successful reprogramming protocols in the mouse and respective comparisons with the TF's expression in the FACS dataset.

Supplementary Table 6

Summary statistics for Extended Data Figures 5 and 6.

Chapter 6

A single-cell transcriptomic atlas characterizes ageing tissues in the mouse

The Tabula Muris Consortium¹ [including **Linda J. van Weele**²]

Adapted from: *Nature* volume 583, pages 590-595 (2020)

¹A list of all authors and their affiliations is available online. ²Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA

Contribution: for the mammary gland, L.J.v.W. designed the experiment, collected and processed the tissue, annotated cell types.

Ageing is characterized by a progressive loss of physiological integrity, leading to impaired function and increased vulnerability to death (1). Despite rapid advances over recent years, many of the molecular and cellular processes that underlie the progressive loss of healthy physiology are poorly understood (2). To gain a better insight into these processes, here we generate a single-cell transcriptomic atlas across the lifespan of *Mus musculus* that includes data from 23 tissues and organs. We found cell-specific changes occurring across multiple cell types and organs, as well as age-related changes in the cellular composition of different organs. Using single-cell transcriptomic data, we assessed cell-type-specific manifestations of different hallmarks of ageing—such as senescence (3), genomic instability (4) and changes in the immune system (2). This transcriptomic atlas—which we denote *Tabula Muris Senis*, or ‘Mouse Ageing Cell Atlas’—provides molecular information about how the most important hallmarks of ageing are reflected in a broad range of tissues and cell types.

We performed single-cell RNA sequencing on more than 350,000 cells from male and female C57BL/6JN mice belonging to six age groups, ranging from 1 month (the equivalent of human early childhood) to 30 months (the equivalent of a human centenarian)(Fig. 1a). For all mice, we prepared single-cell suspensions of the bladder, bone marrow, brain (cerebellum, cortex, hippocampus and striatum), fat (brown, gonadal, mesenteric and subcutaneous), heart and aorta, kidney, large intestine, limb muscle and diaphragm, liver, lung, mammary gland, pancreas, skin, spleen, thymus, tongue and trachea. Data were collected for all six age groups using a microfluidic droplet method (droplet); the 3-month, 18-month and 24-month time points were also analysed using single cells sorted in microtitre well plates (fluorescence-activated cell sorting; FACS) (Extended Data Figs. 1–3, Supplementary Tables 1, 2). Owing to technical constraints, not every tissue was analysed at all time points; a complete list is provided in Extended Data Fig. 4a. The droplet data enable large numbers of cells to be analysed using 3' end counting, whereas the FACS data allow for higher-sensitivity measurements over smaller numbers of cells as well as enabling the acquisition of sequence information across the entire transcript length. The analysis of multiple organs from the same mouse enables us to obtain data that is controlled for age, environment and epigenetic effects.

Data from the 3-month time point—which has previously been published and constitutes the *Tabula Muris* (5)—represents approximately 20% of the cells in the entire dataset, and was used as a basis from which to perform semi-automated cell-type annotation of the data from the additional time points (Fig. 1b, Extended Data Fig. 4b). Using this approach, we were able to automatically annotate more than 70% of the cells. All the automated cell annotations were reviewed and approved by human experts, and the remaining cells were annotated by hand, creating one of the largest manually curated single-cell transcriptomic resources in existence. Many of these cell types have not previously been obtained in pure populations, and these data provide a wealth of new information about their characteristic gene expression profiles. Out of 529,823 total cells sequenced, 110,824 cells for FACS and 245,389 cells for droplet passed our strict filtering criteria (Extended Data Fig. 4b) and were annotated (Extended Data Fig. 2a, b), which was carried out separately for each

tissue and method. The remaining cells are also included in the online dataset but were not used for further analysis here. To investigate whether cell annotations were consistent across the entire organism, we used the bbknn batch-alignment algorithm (6) to correct for method-associated batch effects (Supplementary Table 3). After batch correction, we clustered all cells using an unbiased, graph-based clustering approach (7, 8) (Fig. 1c, d) and assessed the co-occurrence of similarly annotated cells in the same clusters. For example, cells annotated as B cells or endothelial cells tend to occupy the same clusters irrespective of their tissue of origin or the method with which they were processed (Fig. 1e, f, Extended Data Fig. 1g-l).

The *Tabula Muris Senis* enables the discovery of ageing-related changes in specific cell types. Single-cell data enables us to resolve whether gene expression changes

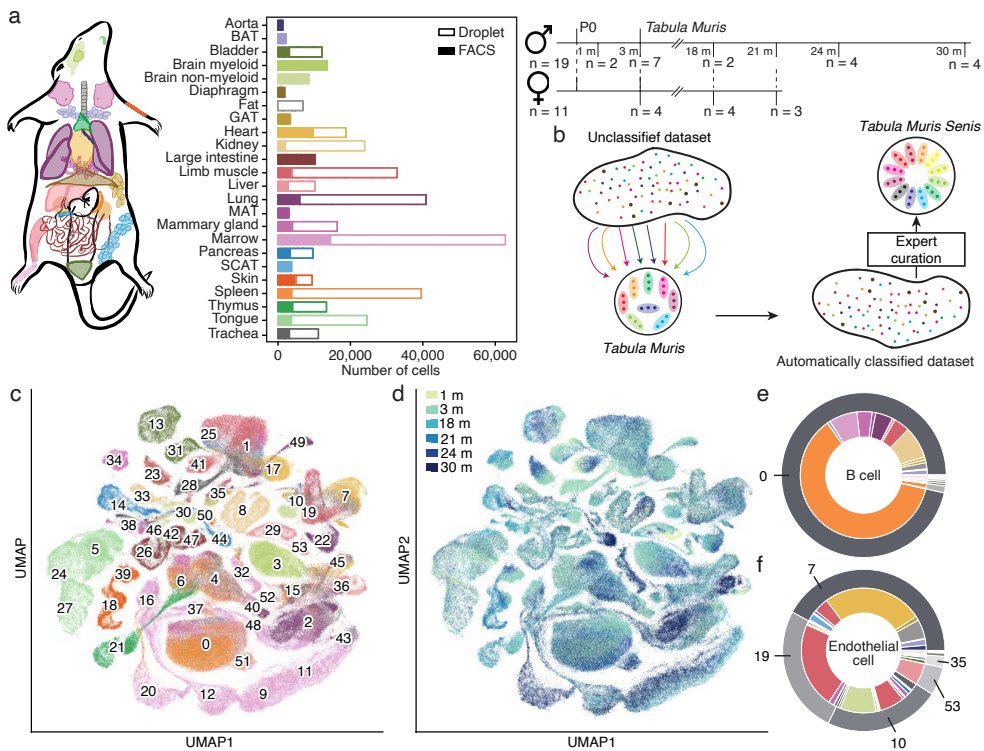


Figure 1: Overview of the *Tabula Muris Senis*. (a) A total of 23 organs from 19 male and 11 female mice were analysed at 6 different time points. The bar plot shows the number of sequenced cells per organ prepared by FACS ($n = 23$ organs) and by microfluidic droplets ($n = 16$ organs). For the droplet dataset the fat sub-tissues were processed together (Fat = BAT + GAT + MAT + SCAT; BAT, brown adipose tissue; GAT, gonadal adipose tissue; MAT, mesenteric adipose tissue; SCAT, subcutaneous adipose tissue). (b) Annotation workflow. Data were clustered together across all time points. We used the *Tabula Muris* (3-month time point) as a reference for the automated pipeline and the annotations were manually curated by tissue experts. (c, d) Uniform manifold approximation and projection (UMAP) plot of all cells, coloured by organ and overlaid with the Louvain cluster numbers (c) and age (d); $n = 356,213$ individual cells. See Extended Data Fig. 4c, d for the colour dictionaries. e, f, B cells (e) and endothelial cells (f) independently annotated for each organ cluster together by unbiased whole-transcriptome Louvain clustering, irrespective of the organ in which they were found.

Chapter 6

observed in bulk experiments are due to changes in gene expression in each cell of the population, or whether the gene expression in each cell stays constant but the number of cells of that type changes, or both. In a global analysis of gene expression changes using the *Tabula Muris Senis* and bulk RNA sequencing from tissues (9), we observed that—in many cases—changes in gene expression are due to both changes in the numbers of cells in a population and changes in the gene expression levels in each cell (Extended Data Fig. 5a, b). As one specific example, we investigated how the fraction of cells that express *Cdkn2a* changes with age. The expression of *Cdkn2a* and its protein product p16 is one of the most frequently used markers of senescence (10) and is an important hallmark of ageing (11). The proportion of cells expressing *Cdkn2a* more than doubled in older mice compared with younger mice according to analysis by both FACS (Fig. 2a) and droplet (Fig. 2b) methods; this was accompanied by a twofold increase in the expression levels of p16 in cells in which it was expressed (Fig. 2c, d). Notably, in 30-month-old mice the fraction of cells that expressed p16 was smaller than in 24-month-old mice, perhaps because long-living animals have a slower rate of senescence. Using a list of previously characterized senescence markers (12, 13, 14, 15), we plotted the fraction of cells expressing each marker across all age groups (Supplementary Table 4). *Cdkn2a* has the highest correlation between ageing and the fraction of cells in which it is expressed; other genes with positive correlation include *E2f2* (16), *Lmnb1* (17, 18), *Tnf* and *Itgax* (19). For some genes, including members of the Sirt family (*Sirt3*, *Sirt4* and *Sirt5*), the fraction of cells in which they were expressed was found to decrease with age; this is consistent with previous literature finding that sirtuins—the protein products of Sirt genes—are essential in delaying cellular senescence (20, 21).

The cellular composition of each tissue type tends to vary with age. We investigated changes in the cellular composition of tissues for which data from at least three time points was available (Supplementary Table 5). Because dissociation does not affect all cell types in a tissue equally, changes in the relative composition of a given cell type with age are more meaningful than comparing proportions of different cell types at a single age (22, 23, 24). The bladder shows pronounced changes in cell-type composition with age (Fig. 2e): whereas the mesenchymal compartment of this tissue decreases by a factor of three over the lifetime of the mouse (Fig. 2e, left), the urothelial compartment increases by a similar amount (Fig. 2e, right). The observation that the proportion of bladder urothelial cells increases with age is concordant with known age-related urothelial changes (25). Using differential gene expression analysis to assess overall changes in tissues with age, we found that stromal-associated genes (*Col1a1*, *Col1a2*, *Col3a1* and *Dcn*) are downregulated while epithelial-associated genes (*Krt15*, *Krt18* and *Sfn*) are upregulated, supporting the compositional observations (Fig. 2f, Supplementary Table 6). The decline of the endothelial population suggests that bladder ageing in mice might be associated with lower organ vascularization, which is consistent with recent findings (26, 27) and with the observed downregulation of vasculature-associated genes *Htra1* and *Fos* (Fig. 2f, Supplementary Table 6). The increase in the leukocyte population could indicate an inflammatory tissue microenvironment, a common hallmark of ageing that is consistent with literature on overactive bladders (28) and is supported by a significant overexpression of *Lgals3*, *Igfbp2* and *Ly6d* across the tissue (Fig. 2f, Supplementary

Table 6), as well as by the overexpression of genes associated with immune response—such as *Tnfrsf12a* and *Cdkn1a*—in both bladder (mesenchymal) cells and bladder urothelial cells (Supplementary Table 6). Moreover, when comparing across ages, we observed that old leukocytes show increased expression of pro-inflammatory markers—such as *Cd14*, *Lgals3* and *Tnfrsf12a*—and decreased expression of anti-

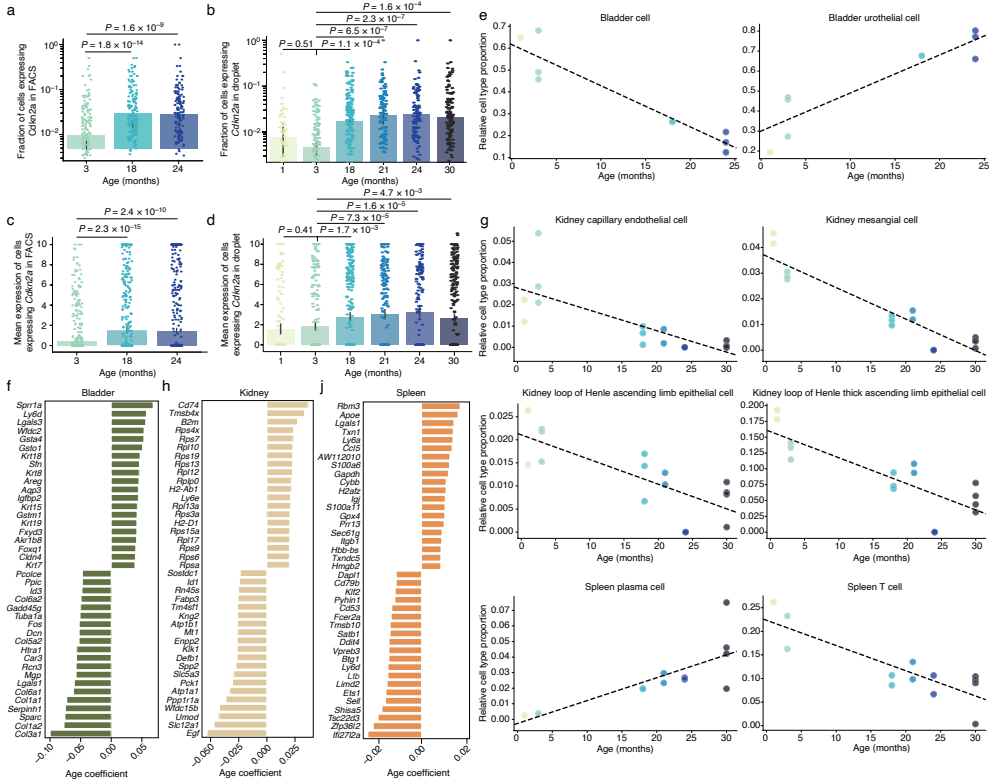


Figure 2: Cellular changes during ageing. (a, b) Bar plot showing the fractions of cells expressing *Cdkn2a* in each age group for FACS (a) and droplet (b) experiments. (c, d) Bar plot of the median expression of *Cdkn2a* for the cells that do express the gene at each age group for FACS (c) and droplet (d) experiments. The y axis shows log-transformed and scaled values. All data are expressed as mean \pm s.d. with individual data points shown. P values were obtained using a Mann-Whitney-Wilcoxon rank-sum two-sided test. $n = 44,518, 34,027$ and $31,551$ individual cells for FACS at 3, 18 and 24 months, respectively; $n = 25,980, 45,602, 44,645, 35,828, 37,660$ and $55,674$ individual cells for droplet at 1, 3, 18, 21, 24 and 30 months, respectively. (e, g, i) The relative abundances of bladder cells (e, left) and bladder urothelial cells (e, right), kidney capillary endothelial cells (g, top left), mesangial cells (g, top right), loop of Henle ascending limb epithelial cells (g, bottom left) and loop of Henle thick ascending limb epithelial cells (g, bottom right), and spleen plasma cells (i, left) and T cells (i, right) change significantly with age. $P < 0.05$ and $r^2 > 0.7$ for a hypothesis test with the null hypothesis that the slope is zero, using two-sided Wald test with t-distribution of the test statistic. (f, h, j) Top 20 upregulated and top 20 downregulated genes in bladder (f), kidney (h) and spleen (j) computed using MAST (51), treating age as a continuous covariate while controlling for sex and technology. Genes were classified as significant under a false-discovery rate (FDR) threshold of 0.01 and an age coefficient threshold of 0.005 (corresponding to an approximately 10%-fold change). For f, $n = 970, 3,804, 2,739$ and $3,864$ individual cells for bladder at 1, 3, 18 and 24 months, respectively; for h, $n = 2,488, 2,832, 3,806, 2,257, 6,384$ and $5,713$ individual cells for kidney at 1, 3, 18, 21, 24 and 30 months, respectively; for j, $n = 2,986, 8,839, 7,141, 6,395, 5,245$ and $8,946$ individual cells for spleen at 1, 3, 18, 21, 24 and 30 months, respectively. The P values for the cell type compositional changes are shown in Supplementary Table 5.

Chapter 6

inflammatory markers such as *Cd9* and *Cd81* (Supplementary Table 6).

Age-related changes in the kidney include a decrease in the relative abundance of mesangial cells, capillary endothelial cells, loop of Henle ascending limb epithelial cells and loop of Henle thick ascending limb epithelial cells (Fig. 2g). Both mesangial cells and capillary endothelial cells are core glomerular cells, and the reduction in their relative abundances with age (Fig. 2g, top)—together with a tissue-wide reduction of *Egf* and *Atp1a1* expression (Fig. 2h, Supplementary Table 6)—suggest an impaired glomerular filtration rate (29, 30). Notably, local *Atp1a1* expression increases with age in both capillary endothelial cells and mesangial cells, suggesting that a compensation mechanism counteracts the effects of the declining proportion of these cells with age. This finding is reinforced by the results of differential gene expression analysis, suggesting that the expression of *Umod*—which encodes uromodulin, the most abundant protein in urine (31)—is also reduced across the tissue. The protein uromodulin is produced by the epithelial cells that line the thick ascending limb, and therefore—given the relative decrease in the proportion of epithelial cells in the ascending and the thick ascending limb—our results suggest that normal kidney functions are impaired (32) (Fig. 2g, bottom, Fig. 2h, Supplementary Table 6). As with *Atp1a1*, we see that the expression of *Umod* increases in a cell type that becomes less abundant with age, leading to an overall reduction of its expression in the organ.

In the spleen, the proportion of T cells decreases with age while the relative amount of plasma cells increases (Fig. 2i). This is supported by the upregulation of B cell and plasma cell marker genes (*Cd79a* and *Jchain* (also known as *Igj*), respectively; Fig. 2j, Supplementary Table 6) and the downregulation of *Cd3d* (Fig. 2j, Supplementary Table 6). Similarly, in the mammary gland we observed a decline of the T cell population (Extended Data Fig. 5c). Age-related decline of T cell populations has been associated with increased risk of infectious disease and cancer (33), and our results suggest that such a decline might also occur in the spleen and the mammary gland. Moreover, genes encoding AP1 transcription factors (34) (*Junb*, *Jund* and *Fos*) were upregulated with age (Extended Data Fig. 5d, Supplementary Table 6), consistent with the observation that normal involution of the mammary gland is accompanied by increased expression of this gene family (35).

The tissue composition of the liver also changes with age. We observed an age-related decrease in the relative number of hepatocytes (Extended Data Fig. 6a-d), which is supported by the reduction in the expression of albumin (*Alb*; Extended Data Fig. 6e, Supplementary Table 6). Differential gene expression analysis revealed an increased immune signature, as illustrated by the overexpression of *H2-Aa*, *H2-Ab1*, *H2-D1*, *H2-Eb1*, *Cd74*, *Lyz2* and others (Extended Data Fig. 6e). Previous findings suggested that pro-inflammatory macrophages drive cellular senescence, and identified the gene *Il1b* as showing markedly different expression in the liver with age (12) (Extended Data Fig. 6f). We performed *in situ* RNA staining in liver Kupffer cells (Extended Data Fig. 6g) with *Clec4f* and found that the number of *Clec4f*-positive cells does not change with age, consistent with the results of the tissue composition analysis (Extended Data Fig. 6h, Supplementary Table 7). However, when co-staining with *Il1b*, we found that the number of cells expressing both *Clec4f* and *Il1b* increased with age

(Extended Data Fig. 6h–j). The expression of *Il1b* is low under normal physiological conditions (36). Specific blocking of *IL1-R1* in hepatocytes has been shown to attenuate cell death after injury, supporting the idea that increased expression of *Il1b* in Kupffer cells is typically a poor prognostic (37). Liver sinusoidal endothelial cells (LSECs) have a unique role in immune defence, being the main carriers of the mannose receptor (MRC1) in this organ (38) (Extended Data Fig. 6k). We identified increased expression of *Mrc1* with age in Kupffer cells, whereas the overall expression of *Mrc1* in liver endothelial cells decreased with age (Supplementary Table 6). By performing in situ RNA staining for *Mrc1* alongside the classical LSEC marker *Pecam1* (Extended Data Fig. 6l, Supplementary Table 7), we found that the number of *Mrc1*-expressing LSECs increased with age (Extended Data Fig. 6m–o). Although *Mrc1* expression did not increase with age in LSECs (Supplementary Table 6), the overall number of cells expressing *Mrc1* did increase significantly with age (Extended Data Fig. 6n). LSECs have been found to have a reduced endocytic capacity in aged livers, and it has been suggested that LSECs proliferate after injury or that bone-marrow-derived LSEC progenitors are recruited to the liver. This suggests that changes in the gene signatures of LSECs with age are closely linked with the function of these cells in immune response.

Genomic instability is among the most widely studied hallmarks of ageing (1), and full-length transcript data enables analysis of the accumulation of somatic mutations with age. We used the Genome Analysis ToolKit (39) to identify single-nucleotide polymorphisms across all FACS samples simultaneously (40, 41) (Supplementary Table 8). We focused on genes that were expressed in at least 75% of cells for each age group within a particular tissue, and observed an age-related increase in the number

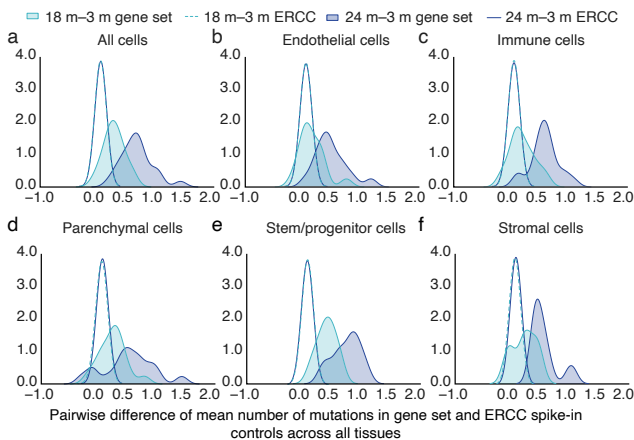


Figure 3: Mutational burden across tissues in ageing mice. (a–f), Distribution of the difference of the mean mutation in the gene set (and ERCC spike-in controls) per cell between 24 months and 3 months and between 18 months and 3 months for all tissues and cells (a) and with the cell types split into five functional groups: endothelial cells (b), immune cells (c), parenchymal cells (d), stem/progenitor cells (e) and stromal cells (f). Filled and solid line distributions correspond to the mean mutation difference in gene set. White and dashed line distributions correspond to the mean mutation difference in ERCC spike-in controls. Note that the mean mutation difference in ERCC spike-in controls overlaps for both age groups. The y axis corresponds to the kernel density estimation (arbitrary units).

Chapter 6

of mutations across all of the organs we analysed (Fig. 3, Extended Data Fig. 7a, c, e) – tongue and bladder were the most affected. We controlled for sequencing coverage and gene expression levels (Extended Data Fig. 8a, c, e), and verified that the number of mutations exceeded those expected from amplification and sequencing errors, which can be estimated using External RNA Controls Consortium (ERCC) controls that were spiked into each well (42) (Fig. 3, Extended Data Figs. 7b, d, f, 8b, d, f). Despite the fact that it is difficult to infer absolute genome-wide mutation rates from the transcriptome, which is known to overstate apparent mutational rates for various reasons (42), the observed trend is a useful indirect estimate of mutational frequency and genome stability.

Ageing also affects the immune system (2), so we analysed clonal relationships between B cells and T cells throughout the organism. We computationally reconstructed the sequence of the B cell receptor and the T cell receptor for B cells and T cells present in the FACS data using singlecell-ige and TraCeR, respectively (43, 44). B cell receptors were assembled for 6,050 cells (Fig. 4a, Extended Data Fig. 9a) and T cell receptors for 6,000 cells (Fig. 4b, Extended Data Fig. 9b). The number of cells with assembled B cell receptors was 1,818 for 3-month-old, 1,356 for 18-month-old and 2,876 for 24-month-old mice. We parsed the singlecell-ige (43) output to define B cell clonotypes on the basis of the sequence of the assembled B cell receptor (Supplementary Table 9) and found that, whereas most of the cells at 3 months were not part of a clone (9% were part of a clonal family), the number of B cells belonging to a clonotype doubled at 18 months (20%) when compared to 3 months, and doubled again from 18 months to 24 months (around 38%). The number of cells with assembled T cell receptors was roughly equal between 3-month-old, 18-month-old and 24-month-old mice (2,076, 2,056 and 1,868 cells, respectively). Clonotype assignment is part of the output obtained by TraCeR (44) (Supplementary Table 9). Notably, only around 3% (55 out of 1,895) of the cells at 3 months were part of a clone. At 18 months and 24 months, this value increased to around 23% (479 out of 2,056) and around 20% (348 out of 1,780) of the cells, respectively, again indicating an increase in clonality of the T cell repertoire at later ages. These changes in clonality for both B cell and T cell repertoires are noteworthy, because they suggest that the immune system of a 24-month-old mouse is less likely to respond to new pathogens. This corroborates literature reports suggesting that older individuals have a higher vulnerability to new infections and lower benefits from vaccination compared with younger individuals (45, 46).

Finally, we computed an overall diversity score to identify which cell types were more susceptible to changes with age (Extended Data Fig. 10). The diversity score is computed as the Shannon entropy of the cluster assignment and then regressed against age to provide a P value (see Methods). We observed significant changes in diversity for cells of the immune system that originate from the brain and the kidney (Fig. 4c, Extended Data Fig. 11a, b). These results were not confounded by the number of genes expressed per cell (Extended Data Fig. 11c, d). In brain myeloid microglial cells, the majority of young (3 month) microglia occupy clusters 1 and 6, whereas old (18 month and 24 month) microglia constitute the vast majority of cells in clusters 10, 12 and 14 (Fig. 4d). Trajectory analysis suggests that young microglia

go through an intermediate state, represented by the clusters mostly occupied by 18-month microglial cells, before acquiring the signature of old microglia (Extended Data Fig. 11e). Clusters 10, 12 and 14 mainly comprise 18-month and 24-month microglia. These cells upregulate major histocompatibility complex (MHC) class I genes (*H2-D1*, *H2-K1* and *B2m*), along with genes associated with degenerative disease (for example, *Fth1*) (47, 48). When compared with clusters 1 and 6—which contain mostly 3-month microglia—gene expression in clusters 10, 12 and 14 is enriched with interferon responsive or regulatory genes (for example, *Oas12*, *Oas1a*, *Ifit3*, *Rtp4*, *Bst2*, *Stat1*, *Irf7*, *Ifitm3*, *Usp18*, *Ifi204* and *Ifit2*), suggesting an expansion of this small pro-inflammatory subset of microglia in the ageing brain (49). Moreover, the list of differentially expressed genes between ‘young’ and ‘old’ clusters resembles the previously reported Alzheimer’s disease-specific microglial signature (47), with 55 out of the top 200 differential expressed genes shared between the two differential gene expression lists (Fig. 4e, Supplementary Table 10). Regarding kidney macrophages, we found two clusters for which the composition changed markedly with age. Cluster 10 is primarily composed of cells from 1-month-old and 3-month-old mice, whereas cluster 13 is mostly composed of cells from 18-month-old, 21-month-old, 24-month-old and 30-month old mice (Fig. 4f). Differential gene expression revealed that cluster 10 is enriched for an M2-macrophage gene signature (for example, *Ii10*, *H2-Eb1*, *H2-Ab1*, *H2-Aa*, *Cd74*, *C1qa*, *Cxcl16*, *Hexb*, *Cd81*, *C1qb* and *Cd72*) whereas cluster

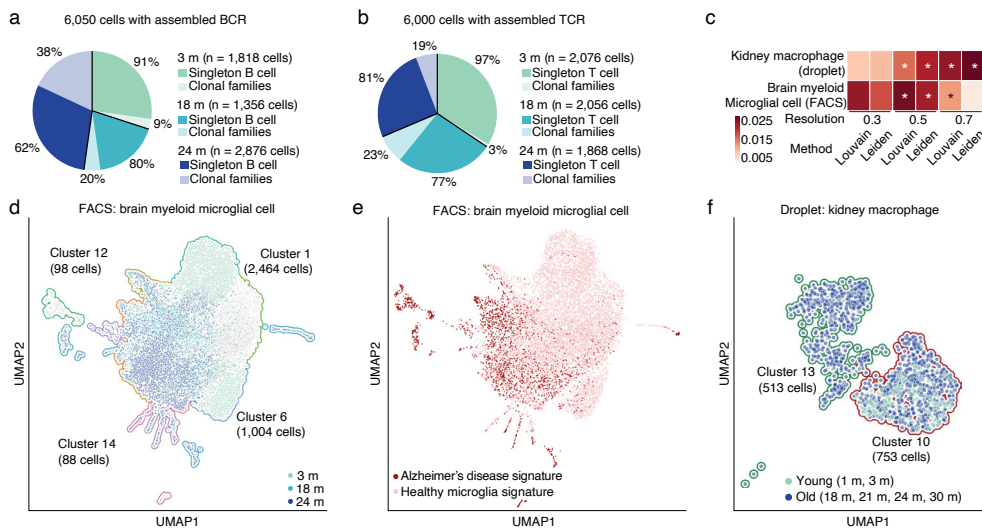


Figure 4: The ageing immune system. (a, b) B cell (a) and T cell (b) clonal families. The pie charts show the proportion of singleton B cells and B cells and that are part of clonal families (a) and singleton T cells and T cells that are part of clonal families (b) at 3 months, 18 months and 24 months. See Extended Data Fig. 9 for clonal networks. BCR, B cell receptor; TCR, T cell receptor. c, Diversity scores for the two cell types for which this score changes significantly with age. d, UMAP plot of the brain myeloid microglial cell Leiden clusters (numbers) coloured by age. The relative cell composition of faded clusters does not change with age; that of the coloured clusters does change with age. e, UMAP plot of the brain myeloid microglial cells when scored using the microglia Alzheimer’s disease signature (Supplementary Table 10). n = 4,532, 4,461 and 4,424 individual microglia cells for brain myeloid at 3 months, 18 months and 24 months, respectively. f, UMAP plot of the kidney macrophage Leiden clusters (numbers) coloured by age group. n = 62, 139, 264, 105, 284 and 553 individual macrophage cells for kidney at 1, 3, 18, 21, 24 and 30 months, respectively.

Chapter 6

13 resembles a M1-proinflammatory macrophage state (50) (for example, *Hp*, *Itgal*, *Msrb1* and *Gngt2*) (Extended Data Fig. 11f, Supplementary Table 10).

The *Tabula Muris Senis* is a comprehensive resource for the cell biology community that offers a detailed molecular and cell-type-specific portrait of ageing. We view such a cell atlas as an essential companion to the genome: the genome provides a blueprint for the organism, but does not explain how genes are used in a cell-type-specific manner or how the usage of genes changes over the lifetime of the organism. The cell atlas provides a deep characterization of phenotype and physiology and serves as a reference for understanding many aspects of the changes in cell biology that occur in mammals during their lifespan.

METHODS

All data, protocols, analysis scripts and an interactive data browser are publicly available.

EXPERIMENTAL PROCEDURES

Mice and organ collection

Male and virgin female C57BL/6JN mice were shipped from the National Institute on Ageing colony at Charles River (housed at 19–23 °C) to the Veterinary Medical Unit (VMU; housed at 20–24 °C) at the VA Palo Alto (VA). At both locations, mice were housed on a 12 h/12 h light/dark cycle and provided with food and water ad libitum. The diet at Charles River was NIH-31, and at the VA VMU was Teklad 2918. Littermates were not recorded or tracked, and mice were housed at the VA VMU for no longer than 2 weeks before euthanasia, with the exception of mice older than 18 months, which were housed at the VA VMU beginning at 18 months of age. Before tissue collection, mice were placed in sterile collection chambers at 8:00 for 15 min to collect fresh fecal pellets. After anaesthetization with 2.5% v/v Avertin, mice were weighed, shaved, and blood was drawn via cardiac puncture before transcatheterial perfusion with 20 ml PBS. Mesenteric adipose tissue was then immediately collected to avoid exposure to the liver and pancreas perfusate, which negatively affects cell sorting. Isolating viable single cells from both the pancreas and the liver of the same mouse was not possible; therefore, two males and two females were used for each. Whole organs were then dissected in the following order: large intestine, spleen, thymus, trachea, tongue, brain, heart, lung, kidney, gonadal adipose tissue, bladder, diaphragm, limb muscle (tibialis anterior), skin (dorsal), subcutaneous adipose tissue (inguinal pad), mammary glands (fat pads 2, 3 and 4), brown adipose tissue (interscapular pad), aorta and bone marrow (spine and limb bones). Organ collection concluded by 10:00. After single-cell dissociation as described below, cell suspensions were used either for FACS of individual cells into 384-well plates, or for preparation of the microfluidic droplet library. All animal care and procedures were carried out in accordance with institutional guidelines approved by the VA Palo Alto Committee on Animal Research.

Tissue dissociation and sample preparation

All tissues were processed as previously described (5).

Sample size, randomization and blinding

No sample size choice was performed before the study. Randomization and blinding were not performed: the authors were aware of all data and metadata-related variables during the entire course of the study.

Single-cell methods

All protocols used in this study are described in detail elsewhere (5). These include: preparation of lysis plates; FACS sorting; cDNA synthesis using the Smart-seq2 protocol (52, 53); library preparation using an in-house version of Tn5 (54, 55); library pooling and quality control; and sequencing. For further details please refer to <https://doi.org/10.17504/protocols.io.2uwgexe>.

Microfluidic droplet single-cell analysis

Single cells were captured in droplet emulsions using the GemCode Single-Cell Instrument (10x Genomics) and scRNA-seq libraries were constructed as per the 10x Genomics protocol using GemCode Single-Cell 3' Gel Bead and Library V2 Kit. In brief, single cell suspensions were examined using an inverted microscope, and if sample quality was deemed satisfactory, the sample was diluted in PBS with 2% FBS to a concentration of 1,000 cells per μl . If cell suspensions contained cell aggregates or debris, two additional washes in PBS with 2% FBS at 300g for 5 min at 4°C were performed. Cell concentration was measured either with a Moxi GO II (Orflo Technologies) or a haemocytometer. Cells were loaded in each channel with a target output of 5,000 cells per sample. All reactions were performed in the Biorad C1000 Touch Thermal cycler with 96-Deep Well Reaction Module. Twelve cycles were used for cDNA amplification and sample index PCR. Amplified cDNA and final libraries were evaluated on a Fragment Analyzer using a High Sensitivity NGS Analysis Kit (Advanced Analytical). The average fragment length of 10x cDNA libraries was quantitated on a Fragment Analyzer (AATI), and by qPCR with the Kapa Library Quantification kit for Illumina. Each library was diluted to 2 nM, and equal volumes of 16 libraries were pooled for each NovaSeq sequencing run. Pools were sequenced with 100 cycle run kits with 26 bases for Read 1, 8 bases for Index 1, and 90 bases for Read 2 (Illumina 20012862). A PhiX control library was spiked in at 0.2 to 1%. Libraries were sequenced on the NovaSeq 6000 Sequencing System (Illumina).

In situ RNA hybridization and quantification

In situ RNA hybridization was performed using the Advanced Cell Diagnostics RNAscope Multiplex Fluorescent Detection kit v2 (323110, Bio-Techne) according to the manufacturer's instructions. Staining of mouse liver specimens was performed using 5- μm paraffin-embedded thick sections. Mouse livers were fixed in 10% formalin buffer saline (HT501128, Sigma-Aldrich) for 24 h at room temperature before paraffin embedding. For multiplex staining the following probes were used; Clec4f (Mm-Clec4f 480421, Il1b (Mm-Il1b 316891-C2), Pecam1 (Mm-Pecam-1 316721), Mrc1 (Mm-Mrc1 437511-C3). Slides were counterstained with Prolong gold antifade reagent with DAPI (P36931, Life Technologies). Mounted slides were imaged on a Leica DM6 B fluorescent microscope (Leica Biosystems). Image quantification was performed using the starfish open source image-based transcriptomics pipeline (see Starfish: Open Source Image Based Transcriptomics and Proteomics Tools, <http://>

github.com/spacetx/starfish and ref. 58).

COMPUTATIONAL METHODS

Data extraction

Sequences from the NovaSeq were de-multiplexed using bcl2fastq v.2.19.0.316. Reads were aligned to the mm10plus genome using STAR v.2.5.2b with parameters TK. Gene counts were produced using HTSEQ v.0.6.1p1 with default parameters, except 'stranded' was set to 'false', and 'mode' was set to 'intersection-nonempty'. Sequences from the microfluidic droplet platform were de-multiplexed and aligned using Cell Ranger v.2.0.1, available from 10x Genomics with default parameters.

Data pre-processing

Gene count tables were combined with the metadata variables using the Scanpy (56) Python package v.1.4.2. We removed genes that were not expressed in at least 3 cells and then cells that did not have at least 250 detected genes. For FACS we removed cells with fewer than 5,000 counts, and for the droplet method we removed cells with fewer than 2,500 unique molecular identifiers (UMIs). The data was then normalized using size factor normalization such that every cell has 10,000 counts and log transformed. We computed highly variable genes using default parameters and then scaled the data to a maximum value of 10. We then computed principal component analysis, neighbourhood graph and clustered the data using Louvain (7) and Leiden (8) methods. The data was visualized using UMAP projection. When performing batch correction to remove the technical artefacts introduced by the technologies, we replaced the neighbourhood graph computation with bbknn (6). Step-by-step instructions to reproduce the pre-processing of the data are available from GitHub.

Cell type annotation

To define cell types we analysed each organ independently but combining all ages. In brief, we performed principal component analysis on the most variable genes between cells, followed by Louvain and Leiden graph-based clustering. Next we subset the data for 3 months (*Tabula Muris* (5)) and computed how many cell types map to each individual cluster. For the clusters that had a single 1:1 mapping (cluster:cell type) we propagated the annotations for all ages; in case there is a 1:many mapping we flagged that cluster for manual validation. Step-by-step instructions to reproduce this method are available from GitHub. For each cluster, we provide annotations in the controlled vocabulary of the cell ontology (57) to facilitate inter-experiment comparisons. Using this method, we were able to annotate automatically (around 1 min per tissue) more than 70% of the dataset. The automatic annotations were then reviewed by each of the tissue experts leading to a fully curated dataset for all the cell types in the *Tabula Muris Senis*.

Tissue cell composition analysis

For each tissue and age, we computed the relative proportion of each cell type. Next we used `scipy.stats.linregress` to regress the relative tissue-cell type changes against age and considered significant the changes with $P < 0.05$ for a hypothesis test with the null hypothesis that the slope is zero, using two-sided Wald test with t-distribution of the test statistic and a $r^2 > 0.5$.

Differential gene expression

We performed differential gene expression analysis on each tissue with a well-powered sample size (more than 100 cells in both young (1 month and 3 month) and old (18 months, 21 months, 24 months and 30 months) age groups). We used a linear model (51) treating age as a numerical variable while controlling for sex and technology. We applied a false-discovery rate (FDR) threshold of 0.01 and an age coefficient threshold of 0.005 (corresponding to an approximately 10% fold change).

Comparison between bulk and single-cell datasets

The differential gene analysis was defined on a per tissue basis. First, we investigated genes on the basis of the single-cell data. We considered only cells from male mice and perform our analysis on the $\log(1 + \text{counts per million (cpm)})$ transformed single-cell count matrices. Note that normalization of the single-cell data was done on a per cell basis. We defined two groups of cells on the basis of age: young cells with age ≤ 3 months (Y) and old cells with age >3 months (O). For each gene we compute the \log_2 fold change of cell and read counts between O and Y. We defined cell count as the fraction of cells that express the gene. Similarly, we defined read count as the mean read count of the gene in the cells that express it. The calculated \log_2 fold-changes of a gene reflect its expression changes with ageing within the single-cell data. Next we analysed each gene on the basis of the bulk data. We computed the Spearman correlation (ρ) of bulk DESeq2 normalized gene expression with ageing. We defined two groups of genes on the basis of the bulk data, increasing with age $\rho > 0.7$ (U) and decreasing with age $\rho < -0.7$ (D). Finally, we compared the \log_2 fold-changes based on the single-cell data between the bulk data defined groups U and D. Specifically, we ran a Wilcoxon-Mann-Whitney test to understand whether \log_2 fold-changes of cell or read counts could distinguish between the two groups. We used the U statistic for effect size.

T cell processing

We used TraCeR44 v.0.5 to identify T cell clonal populations. We ran tracer assemble with-species Mmus set. We then ran tracer summarize with -species Mmus to create the final results. We used the following versions for TraCeR dependencies: igblast v.1.7.0, kallisto v.0.43.1, Salmon v.0.8.2, Trinity v.2.4.0, GRCm38 reference genome. Step-by-step instructions to reproduce the processing of the data are available from GitHub.

B cell processing

We used singlecell-ige (43) v.eafb6d126cc2d6511faae3efbd442abd7c6dc8ef (<https://github.com/dcroote/singlecell-ige>) to identify B cell clonal populations. We used the default configuration settings, except we set the species to mouse. Step-by-step instructions to reproduce the processing of the data are available from GitHub.

Mutation analysis

We used samtools (59) v.1.9 and GATK (39) v.4.1.1.0 for mutation analysis. We used samtools faidx to create our index file. Then we used GATK CreateSequenceDictionary

Chapter 6

and GRCm38, as the reference, to create our sequence dictionary. Next we used GATK AddOrReplaceReadGroups to create a single read group using parameters -RGID 4 -RGLB lib1 -RGPL illumina -RGPU unit1 -RGSM 20. Finally we used GATK HaplotypeCaller to call the mutations. We disabled the following read filters: MappingQualityReadFilter, GoodCigarReadFilter, NotSecondaryAlignmentReadFilter, MappedReadFilter, MappingQualityAvailableReadFilter, Non-zeroReferenceLengthAlignmentReadFilter, NotDuplicateReadFilter, PassesVendorQualityCheckReadFilter, and WellformedReadFilter, but kept all other default settings. The results were summarized per gene in the form of a mutation count per cell table. We started by removing genes mutated in more than 60% of cells, to eliminate the possible bias of germline mutations. Then for each tissue we selected genes expressed in at least 75% of the cells for all the time points to avoid confounding the mutation results with differential gene expression associated with age. Next we computed the average number of mutations in the gene set (or ERCC spike-in controls) per cell and also the average number of raw counts (Supplementary Table 8) and plotted the different distributions. Step-by-step instructions to reproduce the processing of the data are available from GitHub.

Trajectory analysis

We used partition-based graph abstraction (PAGA (60)) to reconstruct the ageing trajectory in brain microglial cells. Step-by-step instructions to reproduce the processing of the data are available from GitHub.

Diversity score

The raw FACS or droplet dataset were used as the input. We filtered genes expressed in fewer than 5 cells, filtered cells if expressing fewer than 500 genes and discarded cells with total number of counts less than 5,000. Next we performed size factor normalization such that every cell had 1×10^4 counts and performed a \log_{1p} transformation. This was followed by clustering, in which we clustered every tissue and every tissue-cell type for every mouse separately using 6 different configurations: resolution parameters (0.3, 0.5, 0.7) \times clustering method (Louvain, Leiden). This is to provide a robust clustering result. For each combination (each tissue-mouse and each tissue-cell_type-mouse), we computed the clustering diversity score as the Shannon entropy of the cluster assignment. We then regressed the diversity score against age to detect the systematic increase or decrease of clustering diversity with respect to age. FDR was used to correct for multiple comparisons. A tissue or a tissue-cell type was selected if the slope was consistent (having the same sign) in all six clustering configurations and at least two out of six clustering configurations had FDR < 0.3 . For each selected tissue or tissue-cell type, a separate UMAP was computed using cells from all mice for visualization using Leiden clustering with resolution parameter 0.7.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The entire dataset can be explored interactively at <http://tabula-muris-senis>.

ds.czbiohub.org/. Gene counts and metadata are available from figshare (<https://doi.org/10.6084/m9.figshare.8273102.v2>) and the Gene Expression Omnibus under accession code GSE132042; the raw data files are available from a public AWS S3 bucket (<https://registry.opendata.aws/tabula-muris-senis/>).

Code availability

The code used for the analysis is available from GitHub at <https://github.com.laneproxy.stanford.edu/czbiohub/tabula-muris-senis>.

ACKNOWLEDGEMENTS

We thank Sony Biotechnology for making an SH800S instrument available for this project. Some of the cell sorting and flow cytometry analysis for this project was done on a Sony SH800S instrument in the Stanford Shared FACS Facility. Some FACS experiments were performed with instruments in the VA Flow Cytometry Core, which is supported by the US Department of Veterans Affairs (VA), Palo Alto Veterans Institute for Research (PAVIR), and the National Institutes of Health (NIH). This work was supported by the Chan Zuckerberg Biohub, Department of Veterans Affairs grant IK6 BX004599 (T.W.-C.) and NIH/NIA DP1 grant AG053015 (T.W.-C.). We thank B. Tojo for the artwork, and C. Xu and J. Batson for discussions.

CONTRIBUTIONS

An overview of contributions is available online.

REFERENCES

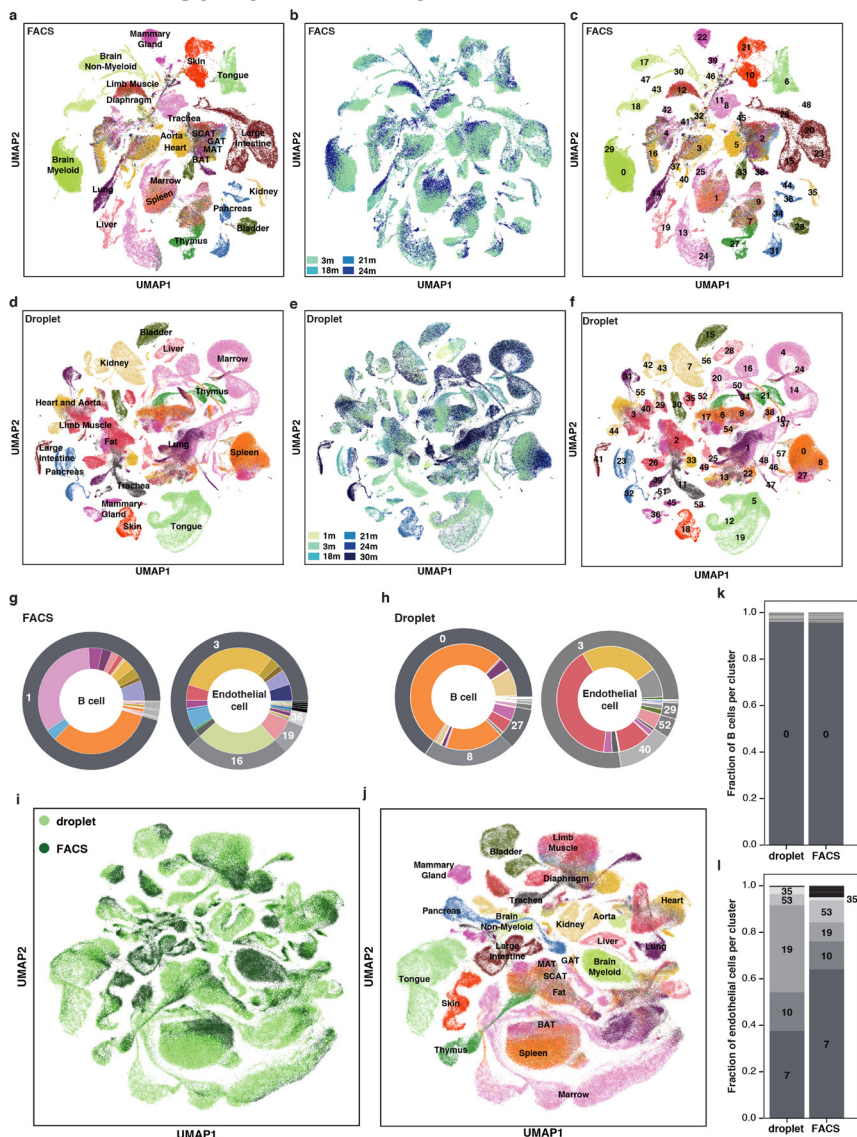
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* 153, 1194-1217 (2013).
- Nikolich-Zugich, J. The twilight of immunity: emerging concepts in aging of the immune system. *Nat. Immunol.* 19, 10-19 (2018).
- Campisi, J. Aging, cellular senescence, and cancer. *Annu. Rev. Physiol.* 75, 685-705 (2013).
- Vijg, J. & Suh, Y. Genome instability and aging. *Annu. Rev. Physiol.* 75, 645-668 (2013).
- The Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 562, 367-372 (2018).
- Polański, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* 36, 964-965 (2020).
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* P10008 (2008).
- Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* 9, 5233 (2019).
- Schaum, N. et al. The murine transcriptome reveals global aging nodes with organ-specific phase and amplitude. Preprint at <https://www.biorxiv.org/content/10.1101/662254v1> (2019).
- Rayess, H., Wang, M. B. & Srivatsan, E. S. Cellular senescence and tumor suppressor gene p16. *Int. J. Cancer* 130, 1715-1725 (2012).
- Hernandez-Segura, A., Nehme, J. & Demaria, M. Hallmarks of cellular senescence. *Trends Cell Biol.* 28, 436-453 (2018).
- Covarrubias, A. J. et al. Aging-related inflammation driven by cellular senescence enhances NAD consumption via activation of CD38+ pro-inflammatory macrophages. Preprint at <https://www.biorxiv.org/content/10.1101/609438v2> (2019).
- Nagano, T. et al. Identification of cellular senescence-specific genes by comparative transcriptomics. *Sci. Rep.* 6, 31758 (2016).
- Carnero, A. in *Cell Senescence. Methods in Molecular Biology (Methods and Protocols)* Vol. 965 (eds Galluzzi, L., Vitale, I., Kepp, O. & Kroemer, G.) (Humana, 2013).
- Wang, A. S. & Dreesen, O. Biomarkers of cellular senescence and skin aging. *Front. Genet.* 9, 247 (2018).
- Vernier, M. et al. Regulation of E2Fs and senescence by PML nuclear bodies. *Genes Dev.* 25, 41-50 (2011).
- Dreesen, O. et al. Lamin B1 fluctuations have differential effects on cellular proliferation and senescence. *J. Cell Biol.* 200, 605-617 (2013).
- Shah, P. P. et al. Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev.* 27, 1787-1799 (2013).
- Li, P. et al. The inflammatory cytokine TNF- α promotes the premature senescence of rat nucleus pulposus cells via the PI3K/Akt

Chapter 6

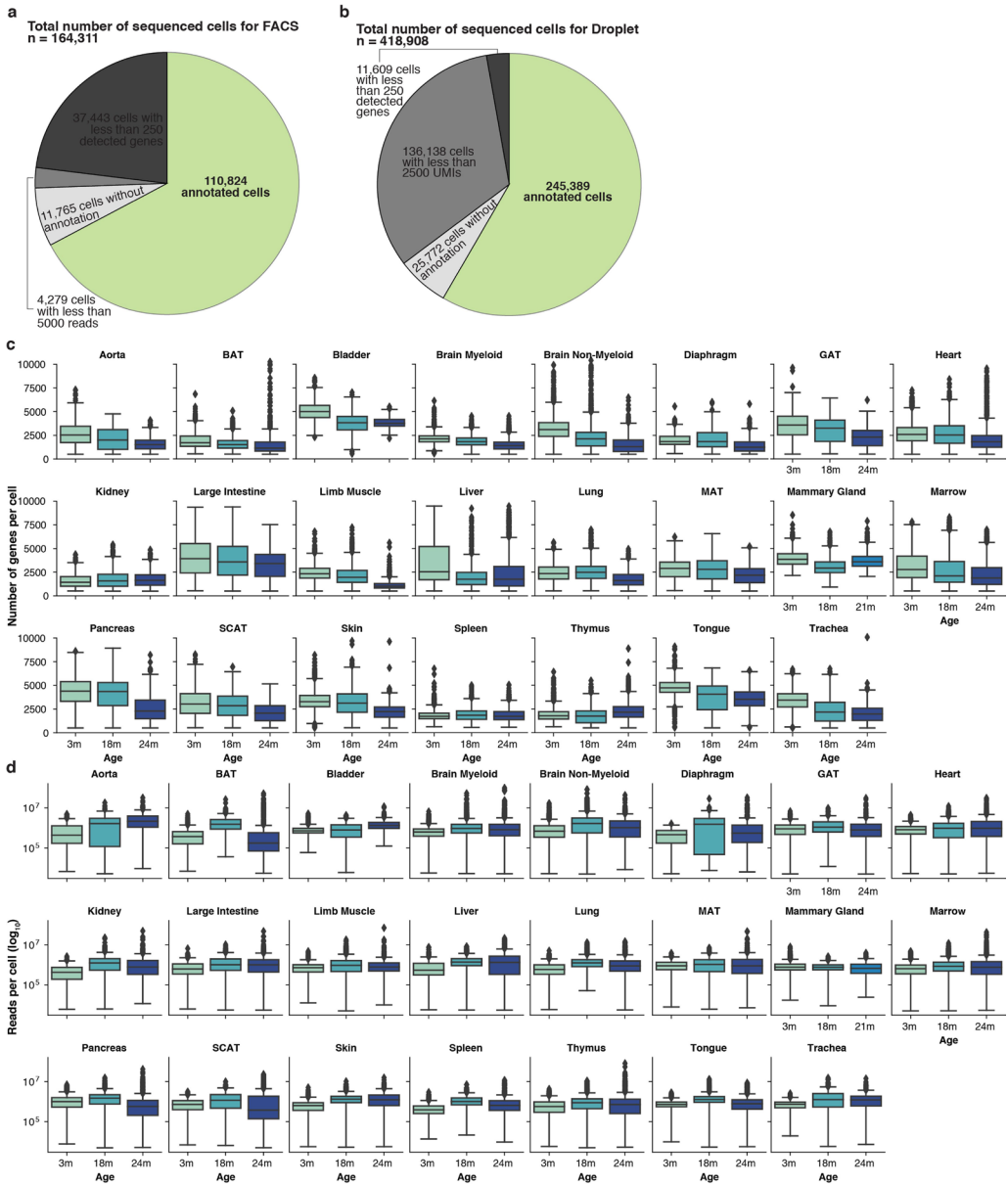
- signaling pathway. *Sci. Rep.* 7, 42938 (2017).
20. Saunders, L. R. & Verdin, E. Sirtuins: critical regulators at the crossroads between cancer and aging. *Oncogene* 26, 5489-5504 (2007).
 21. Lee, S.-H., Lee, J.-H., Lee, H.-Y. & Min, K.-J. Sirtuin signaling in cellular senescence and aging. *BMB Rep.* 52, 24-34 (2019).
 22. van den Brink, S. C. et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* 14, 935-936 (2017).
 23. Tung, P.-Y. et al. Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* 7, 39921 (2017).
 24. Nguyen, Q. H., Pervolarakis, N., Nee, K. & Kessenbrock, K. Experimental considerations for single-cell RNA sequencing approaches. *Front. Cell Dev. Biol.* 6, 108 (2018).
 25. Daly, D. M. et al. Age-related changes in afferent pathways and urothelial function in the male mouse bladder. *J. Physiol. (Lond.)* 592, 537-549 (2014).
 26. Burmeister, D. M., AbouShwareb, T., Bergman, C. R., Andersson, K.-E. & Christ, G. J. Age-related alterations in regeneration of the urinary bladder after subtotal cystectomy. *Am. J. Pathol.* 183, 1585-1595 (2013).
 27. Andersson, K.-E., Boedtker, D. B. & Forman, A. The link between vascular dysfunction, bladder ischemia, and aging bladder dysfunction. *Ther. Adv. Urol.* 9, 11-27 (2017).
 28. Suskind, A. M. The aging overactive bladder: a review of aging-related changes from the brain to the bladder. *Curr. Bladder Dysfunct. Rep.* 12, 42-47 (2017).
 29. Zhang, D. et al. Downregulation of ATP1A1 promotes cancer development in renal cell carcinoma. *Clin. Proteomics* 14, 15 (2017).
 30. Isaka, Y. Epidermal growth factor as a prognostic biomarker in chronic kidney diseases. *Ann. Transl. Med.* 4, S62 (2016).
 31. Devuyt, O., Olinger, E. & Rampoldi, L. Uromodulin: from physiology to rare and complex kidney disorders. *Nat. Rev. Nephrol.* 13, 525-544 (2017).
 32. Tokonami, N. et al. Uromodulin is expressed in the distal convoluted tubule, where it is critical for regulation of the sodium chloride cotransporter NCC. *Kidney Int.* 94, 701-715 (2018).
 33. Palmer, S., Albergante, L., Blackburn, C. C. & Newman, T. J. Thymic involution and rising disease incidence with age. *Proc. Natl Acad. Sci. USA* 115, 1883-1888 (2018).
 34. Shen, Q. et al. The AP-1 transcription factor regulates postnatal mammary gland development. *Dev. Biol.* 295, 589-603 (2006).
 35. Girmius, N., Edwards, Y. J. K. & Davis, R. J. The cJUN NH2-terminal kinase (JNK) pathway contributes to mouse mammary gland remodeling during involution. *Cell Death Differ.* 25, 1702-1715 (2018).
 36. Tan, Q. et al. The role of IL-1 family members and Kupffer cells in liver regeneration. *BioMed Res. Int.* 2016, 6495793 (2016).
 37. Gehrke, N. et al. Hepatocyte-specific deletion of IL1-RI attenuates liver injury by blocking IL-1 driven autoinflammation. *J. Hepatol.* 68, 986-995 (2018).
 38. Liu, Y., Gardner, C. R., Laskin, J. D. & Laskin, D. L. Classical and alternative activation of rat hepatic sinusoidal endothelial cells by inflammatory stimuli. *Exp. Mol. Pathol.* 94, 160-167 (2013).
 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. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491-498 (2011).
 41. Auwera, G. A. et al. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11.10.1-11.10.33 (2013).
 42. Zook, J. M., Samarov, D., McDaniel, J., Sen, S. K. & Salit, M. Synthetic spike-in standards improve run-specific systematic error analysis for DNA and RNA sequencing. *PLoS ONE* 7, e41356 (2012).
 43. Croote, D., Darmanis, S., Nadeau, K. C. & Quake, S. R. High-affinity allergen-specific human antibodies cloned from single IgE B cell transcriptomes. *Science* 362, 1306-1309 (2018).
 44. Stubbington, M. J. T. et al. T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* 13, 329-332 (2016).
 45. Goronzy, J. J. & Weyand, C. M. Understanding immunosenescence to improve responses to vaccines. *Nat. Immunol.* 14, 428-436 (2013).
 46. Goronzy, J. J. & Weyand, C. M. Successful and maladaptive T cell aging. *Immunity* 46, 364-378 (2017).
 47. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* 169, 1276-1290.e17 (2017).
 48. Li, Q. et al. Developmental heterogeneity of microglia and brain myeloid cells revealed by deep single-cell RNA sequencing. *Neuron* 101, 207-223.e10 (2019).
 49. Hammond, T. R. et al. Single-cell RNA sequencing of microglia throughout the mouse lifespan and in the injured brain reveals complex cell-state changes. *Immunity* 50, 253-271.e6 (2019).
 50. Jablonski, K. A. et al. Novel markers to delineate murine M1 and M2 macrophages. *PLoS ONE* 10, e0145342 (2015).
 51. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.*

- 16, 278 (2015).
52. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096-1098 (2013).
 53. Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* 112, 7285-7290 (2015).
 54. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* 24, 2033-2040 (2014).
 55. Hennig, B. P. et al. Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3 (Bethesda)* 8, 79-89 (2018).
 56. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15 (2018).
 57. Diehl, A. D. et al. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semantics* 7, 44 (2016).
 58. McQuin, C. et al. CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* 16, e2005970 (2018).
 59. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079 (2009).
 60. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* 20, 59 (2019).

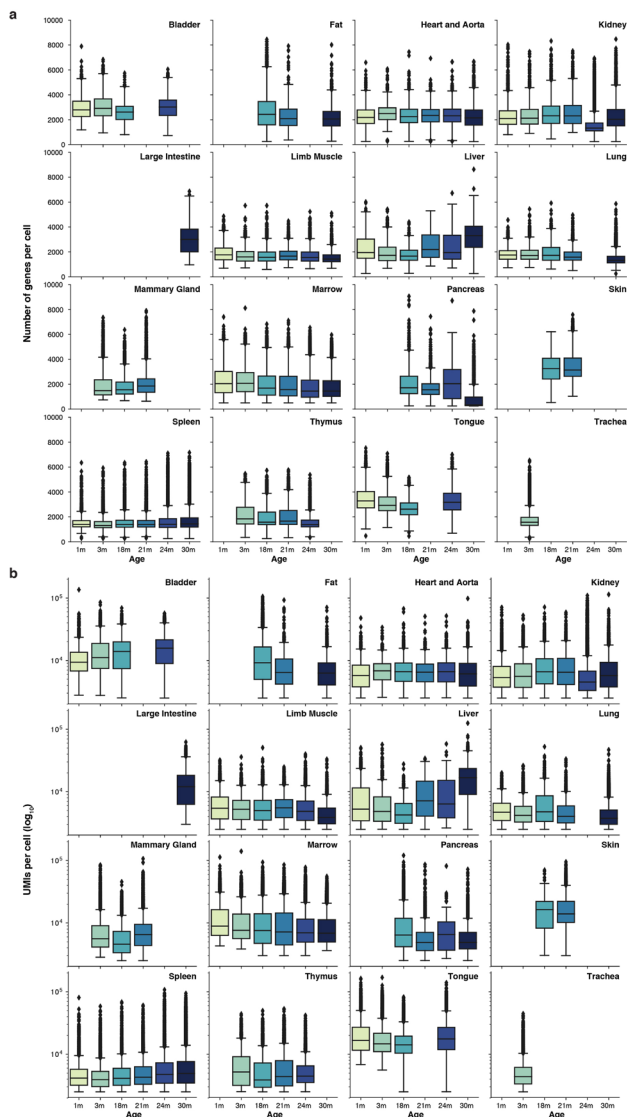
EXTENDED DATA FIGURES AND TABLES



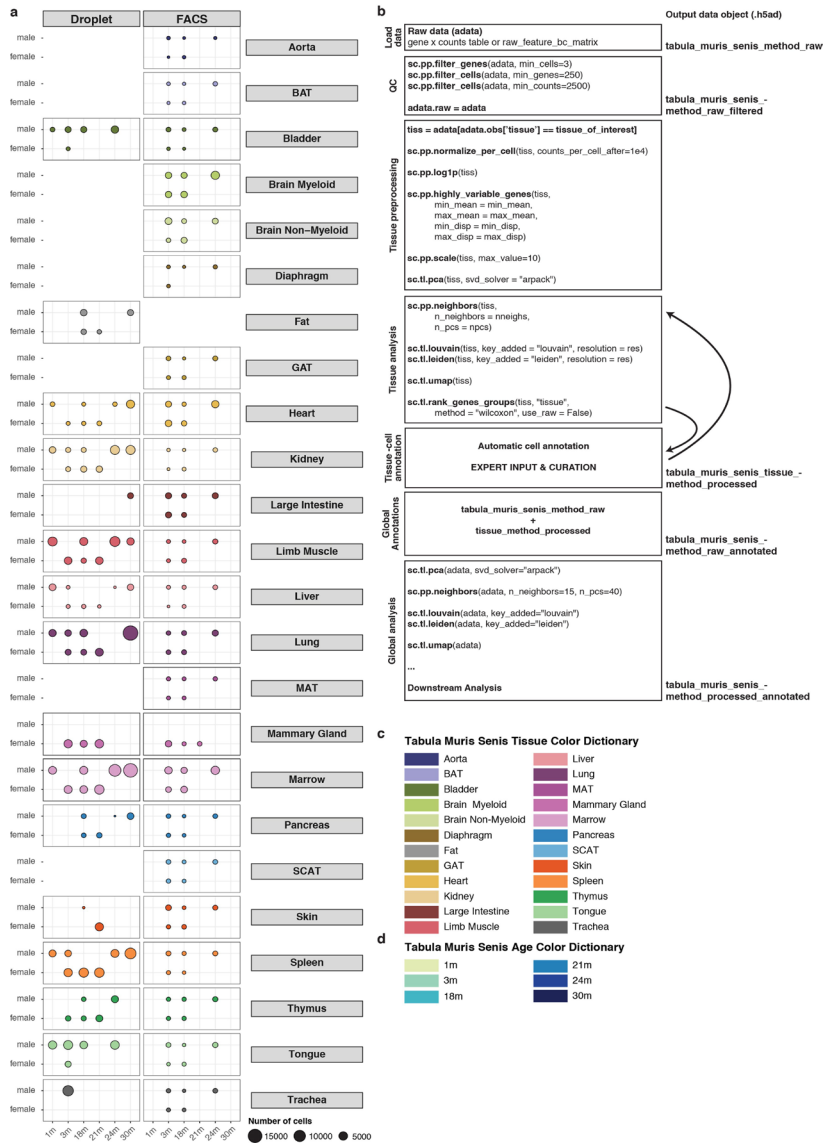
Extended Data Figure 1: UMAP visualizations of the whole *Tabula Muris Senis*. (a, b) UMAP plot of all cells collected for FACS coloured by tissue (a) or age (b). (c) UMAP plot of all cells collected by FACS, coloured by organ (Extended Data Fig. 4c), overlaid with the Louvain cluster numbers. $n = 110,824$ individual cells for FACS. (d, e) UMAP plot of all cells collected for droplet coloured by organ (d) or age (e). f, UMAP plot of all cells collected by droplet, coloured by organ (Extended Data Fig. 4c), overlaid with the Louvain cluster numbers. $n = 245,389$ individual cells for droplet. (g) B cells (left) and endothelial cells (right) in FACS independently annotated for each organ cluster together by unbiased whole-transcriptome Louvain clustering, irrespective of the organ from which they originated. (h) B cells (left) and endothelial cells (right) in droplet independently annotated for each organ cluster together by unbiased whole-transcriptome Louvain clustering, irrespective of the organ from which they originated. (i, j) UMAP plot of all cells collected coloured by method (i) or tissue (j). $n = 356,213$ individual cells for FACS and droplet combined. (k, l) B cells (k) and endothelial cells (l) cluster together by unbiased whole-transcriptome Louvain clustering, irrespective of the technology by which they were found.



Extended Data Figure 2: *Tabula Muris Senis* quality control statistics overall summary and detailed for the FACS dataset. (a) Pie chart with the summary statistics for FACS. (b) Pie chart with the summary statistics for droplet. (c) Box plot of the number of genes detected per cell for each organ and age for FACS. (d) Box plot of the number of reads per cell (log-scale) for each organ and age for FACS. For (c, d) all data are expressed as mean \pm s.d. Individual data points (black diamonds) correspond to outliers outside of the quantile distribution. The sample size (number of cells for each tissue and age) is available in Supplementary Table 1.

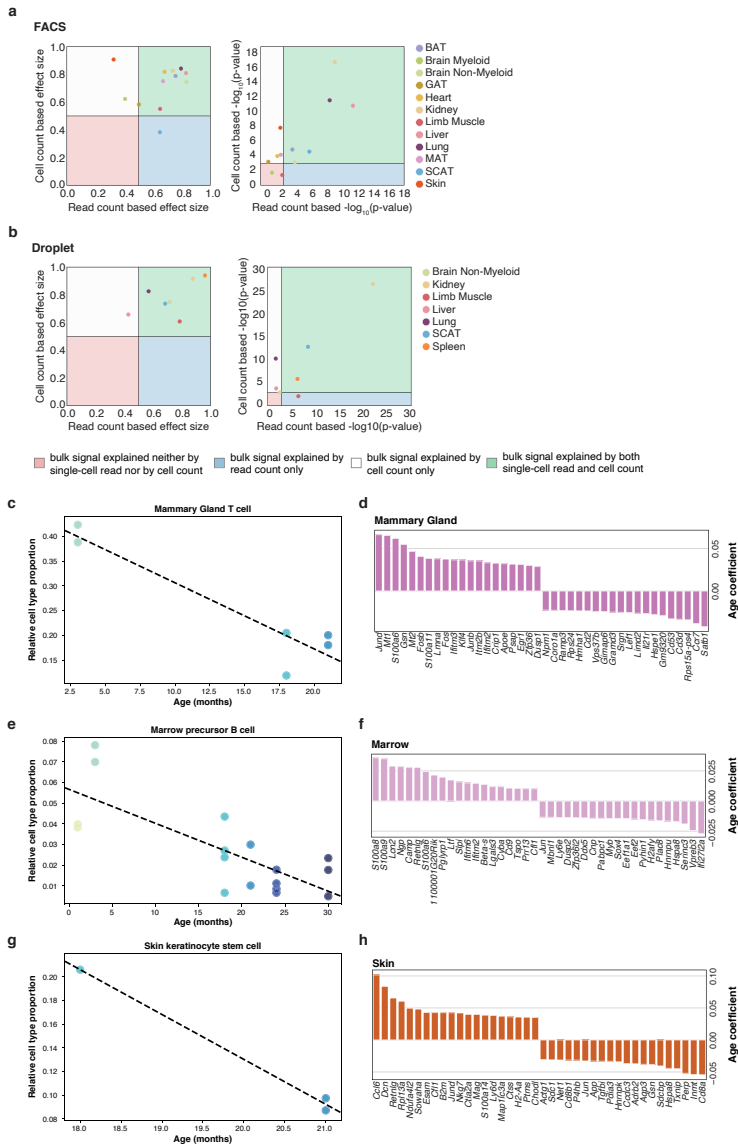


Extended Data Figure 3: *Tabula Muris Senis* quality control statistics detailed for the droplet dataset. (a) Box plot of the number of genes detected per cell for each organ and age for droplet. (b) Box plot of the number of UMIs per cell (log scale) for each organ and age for droplet. All data are expressed as mean \pm s.d. Individual data points (black diamonds) correspond to outliers outside of the quantile distribution. The sample size (number of cells for each tissue and age) is available in Supplementary Table 2.



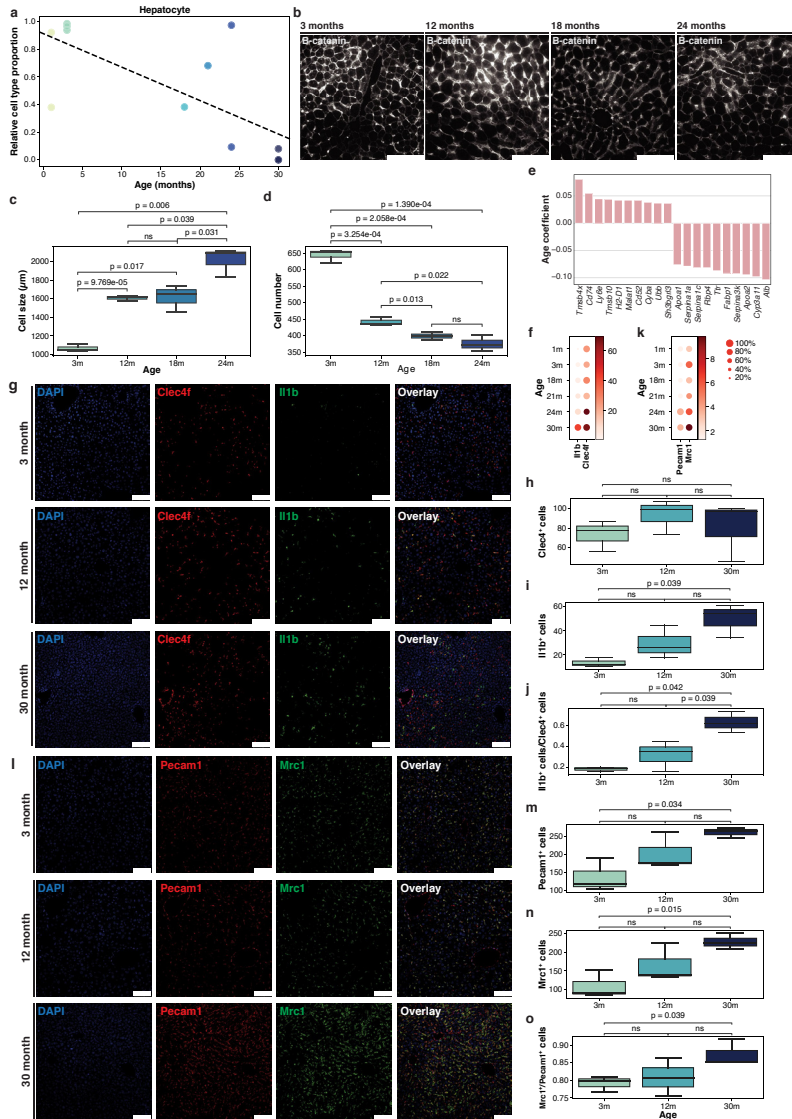
Extended Data Figure 4: Number of cells in *Tabula Muris Senis* across age, sex, tissue and technology and schematic of the data processing. (a) Balloon plot showing the number of sequenced cells per sequencing method per organ per sex per age. (b) Schematic of the analysis workflow. (c, d) *Tabula Muris Senis* colour dictionary for organs and tissues (c) and ages (d).

Chapter 6

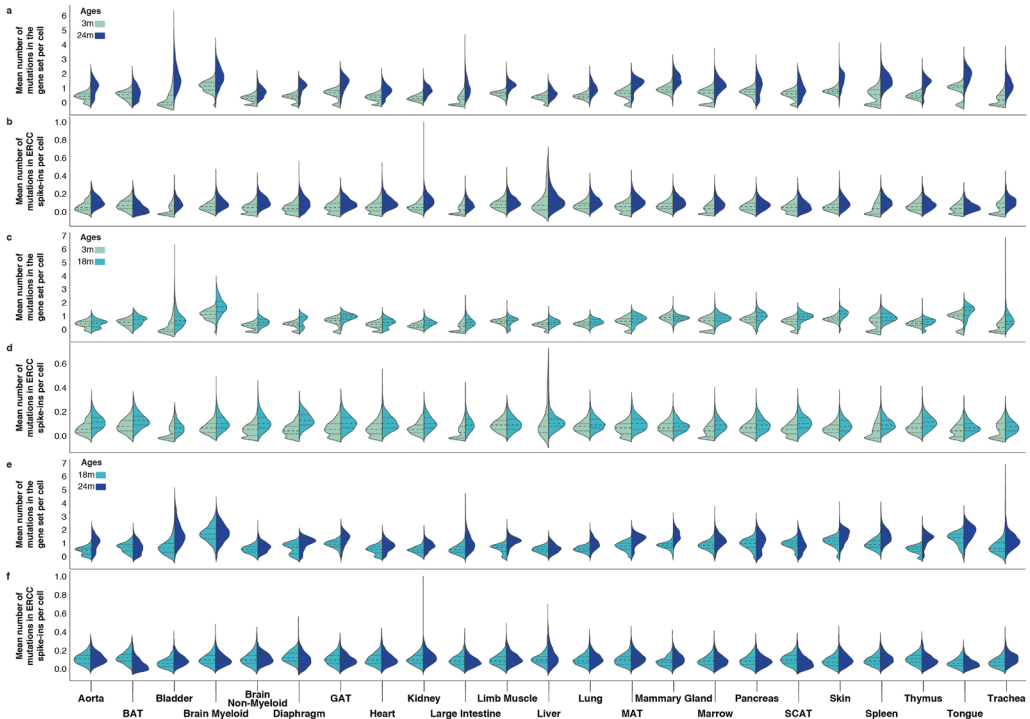


Extended Data Figure 5: See next page for caption.

Extended Data Figure 5: Comparison of bulk and single-cell datasets and tissue cell compositions. (a, b) Ageing patterns from bulk and single-cell data are consistent. Strong changes in bulk gene expression with ageing can be either explained by cell or read count-based changes in single-cell data FACS (a) and droplet (b). Two-sided Wilcoxon-Mann-Whitney indicates that single-cell data based log₂ fold-changes of cell or read counts distinguish between up and down regulated genes in bulk data. $n = 110,824$ individual cells for FACS and $n = 245,389$ individual cells for droplet. (c) Mammary gland T cell relative abundances change significantly with age ($P < 0.05$ and $r^2 > 0.7$ for a hypothesis test with the null hypothesis that the slope is zero, using two-sided Wald test with t-distribution of the test statistic). (d) Top 20 upregulated and downregulated genes in mammary gland computed using MAST (51), treating age as a continuous covariate while controlling for sex and technology. Genes were classified as significant under an FDR threshold of 0.01 and an age coefficient threshold of 0.005 (corresponding to an approximately 10% fold change). $n = 6,393, 3,635$ and $5,549$ individual cells for mammary gland at 3, 18 and 21 months, respectively. (e) Relative abundances of marrow precursor B cells change significantly with age ($P < 0.05$ and $r^2 > 0.7$ for a hypothesis test with the null hypothesis that the slope is zero, using two-sided Wald test with t-distribution of the test statistic). (f) Top 20 upregulated and downregulated genes in marrow computed using MAST (51), treating age as a continuous covariate while controlling for sex and technology. Genes were classified as significant under an FDR threshold of 0.01 and an age coefficient threshold of 0.005 (corresponding to an approximately 10% fold change). $n = 3,027, 8,559, 11,496, 5,216, 12,943$ and $13,496$ individual cells for marrow at 1, 3, 18, 21, 24 and 30 months, respectively. (g) Relative abundances of skin keratinocyte stem cells change significantly with age ($P < 0.05$ and $r^2 > 0.7$ for a hypothesis test with the null hypothesis that the slope is zero, using two-sided Wald test with t-distribution of the test statistic). (h) Top 20 upregulated and downregulated genes in skin computed using MAST (51), treating age as a continuous covariate while controlling for sex and technology. Genes were classified as significant under an FDR threshold of 0.01 and an age coefficient threshold of 0.005 (corresponding to an approximately 10% fold change). $n = 2,346, 1,494, 4,352$ and $1,122$ individual cells for skin at 3, 18, 21 and 24 months, respectively. The P values for the cell type compositional changes are shown in Supplementary Table 5.

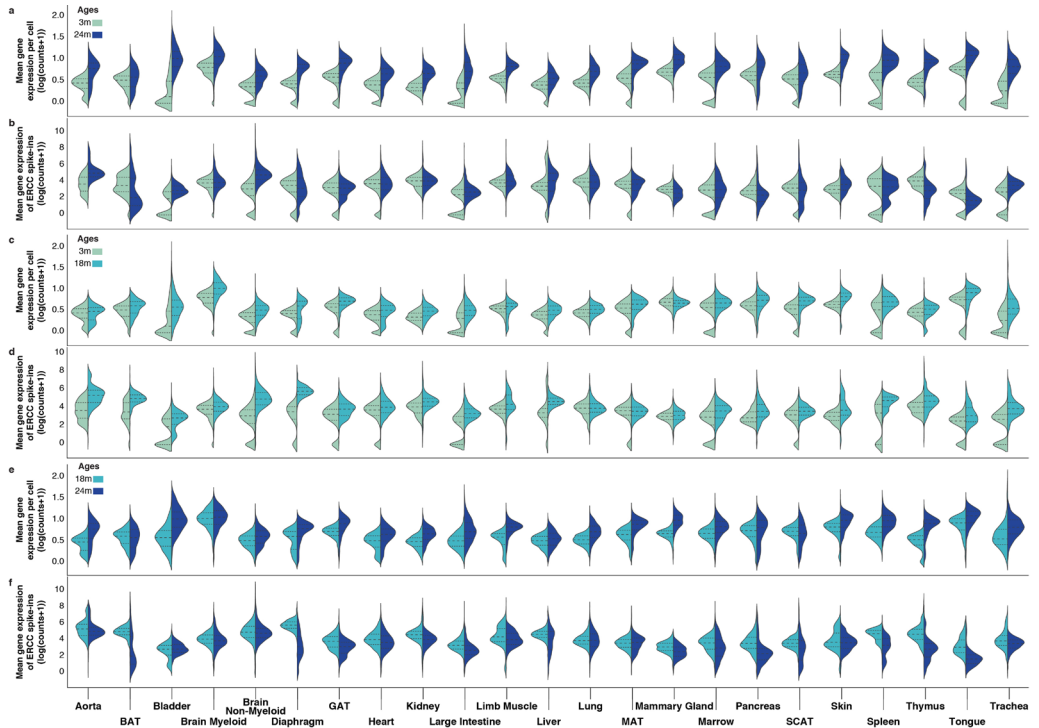


Extended Data Figure 6: Cellular changes during ageing in the liver. (a) Relative abundances of liver hepatocytes change significantly with age ($P < 0.05$ and $r^2 > 0.7$ for a hypothesis test with the null hypothesis that the slope is zero, using two-sided Wald test with t-distribution of the test statistic). $n = 2,791, 2,832, 3,806, 2,257, 6,384$ and $5,713$ individual cells for liver 1, 3, 18, 21, 24 and 30 months, respectively. The P values for the cell type compositional changes are shown in Supplementary Table 5. b-d, Bright-field imaging of hepatocytes at different ages (b) and respective quantification (c, d). e, Top 10 upregulated and downregulated genes in liver computed using MAST (51), treating age as a continuous covariate while controlling for sex and technology. Genes were classified as significant under an FDR threshold of 0.01 and an age coefficient threshold of 0.005 (corresponding to an approximately 10% fold change). The sample size is the same as for a, f, k, Gene expression of *Il1b* and *Clec4f* (f) and *Pecam1* and *Mrc1* (k) in the liver droplet dataset for the six ages. g-j, Staining of Kupfer cells at different ages (g) and respective quantification (h-j). l-o, Staining of liver endothelial cells at different ages (l) and respective quantification (m-o). Scale bars, 100 μm . For c, d, h-j and m-o, all data are expressed as mean \pm s.d. and P values were obtained using a Welch's test. The sample size for each group is available in Supplementary Table 7.

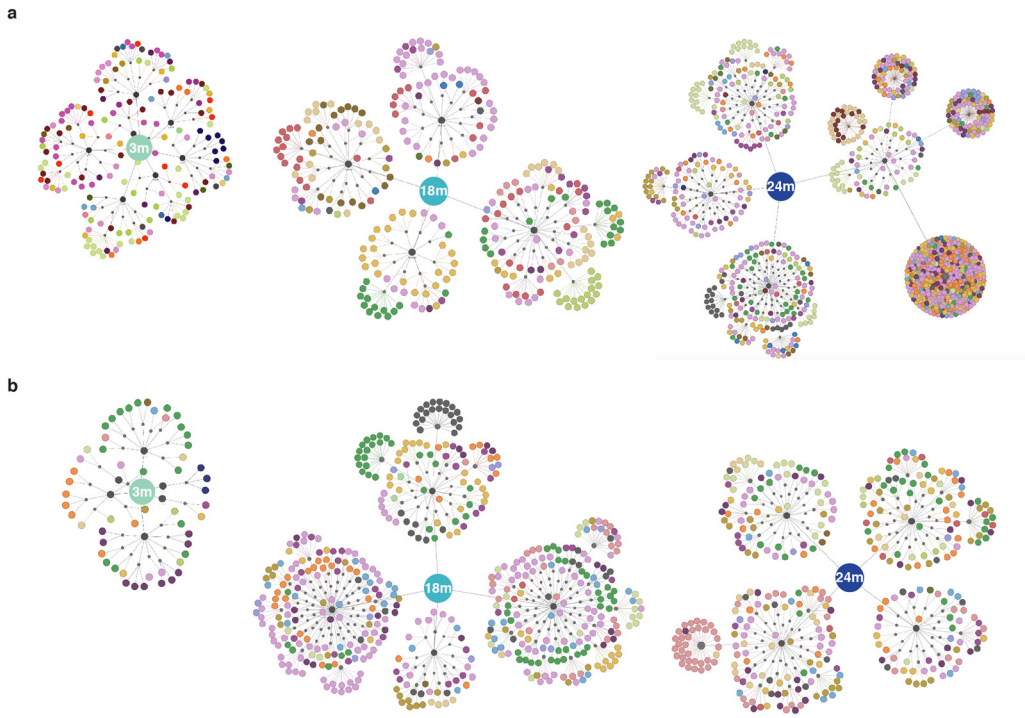


Extended Data Figure 7: Mean number of somatic mutations with age. (a-f) Mean number of somatic mutations in genes (a, c, e) and ERCC spike-in controls (b, d, f) across all tissues per age group (3 months and 24 months (a, b), 3 months and 18 months (c, d), 18 months and 24 months (e, f)). Mutations are presented as the mean number of mutations per gene or ERCC spike-in per cell.

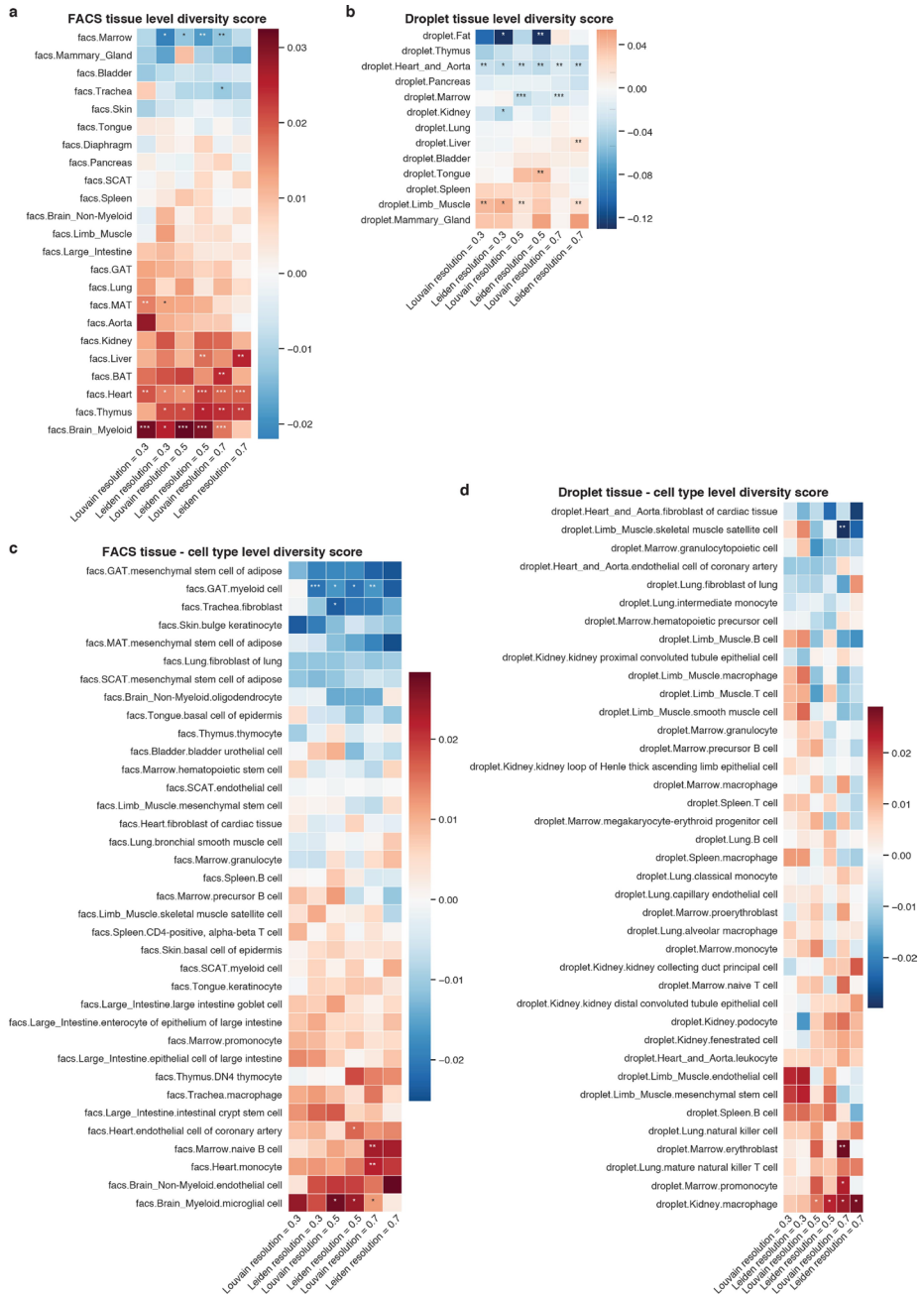
Chapter 6



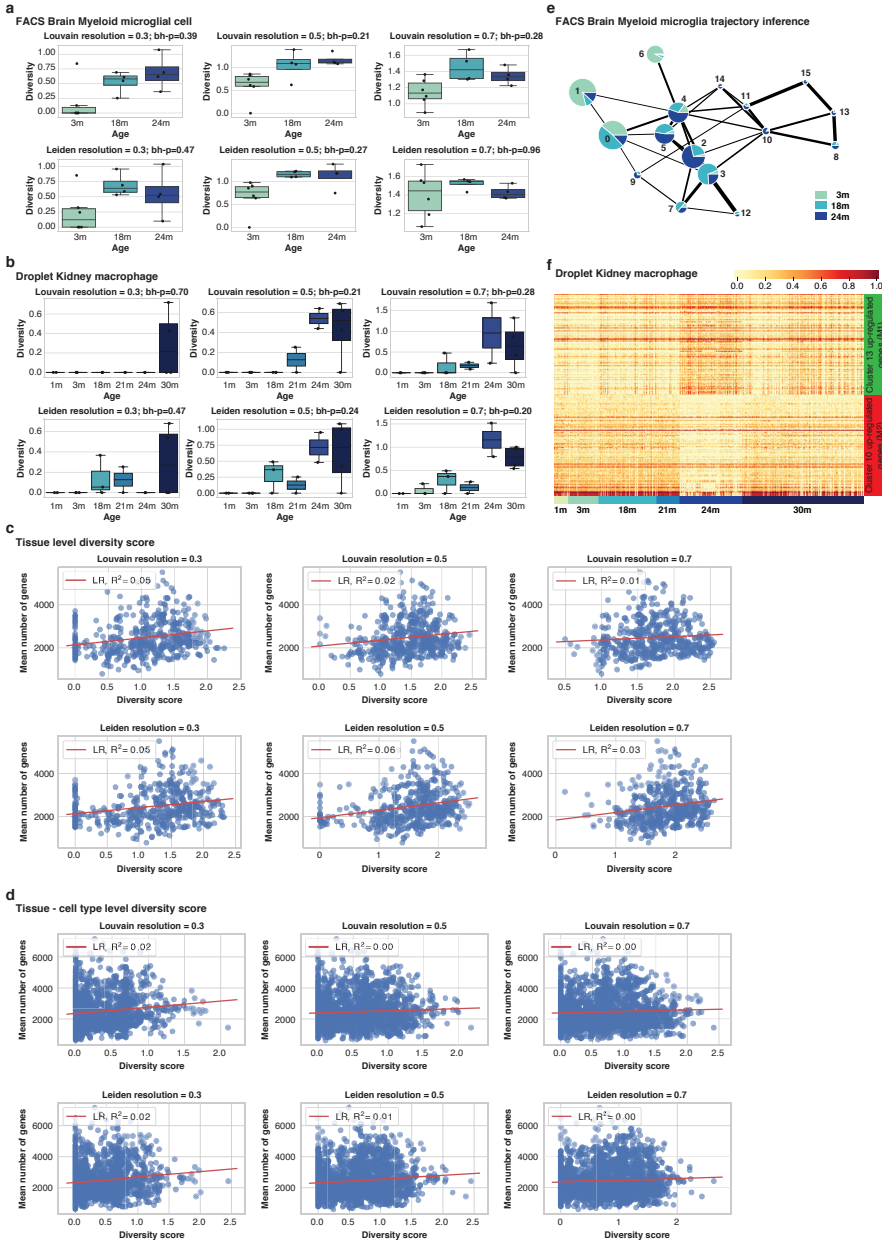
Extended Data Figure 8: Gene (raw) expression with age. (a-f) Gene raw expression (a, c, e) and ERCC spike-in control raw expression (b, d, f) across all tissues per age group (3 months and 24 months (a, b), 3 months and 18 months (b, d), 18 months and 24 months (c, e)). Raw expression are presented as the mean number of counts per gene or ERCC spike-in control per cell.



Extended Data Figure 9: Immune repertoire clonality analysis. (a) B-cell clonal families. For each time point, the clonal families are represented in a tree structure for which the central node is age. Connected to the age node there is an additional node (dark grey) that represents each mouse and the clonal families are depicted for each mouse. For each clonal family, cells that are part of that family are coloured by the organ of origin. (b) T-cell clonal families. For each time point, clonal families are represented in a tree structure for which the central node is age. Connected to the age node there is an additional node (dark grey) that represents each mouse and the clonal families are depicted for each mouse. For each clonal family, cells that are part of that family are coloured by the organ of origin.



Extended Data Figure 10: Diversity score summary. (a, b) Heat map summary of the overall tissue diversity score for FACS (a) and droplet (b). (c, d) Heat map summary of the tissue cell-type diversity score for FACS (c) and droplet (d).



Extended Data Figure 11: Diversity score and differential gene expression analysis for brain myeloid and kidney. (a, b) Diversity score at different cluster resolutions for FACS brain myeloid microglia cell (a) and droplet kidney macrophage (b). n = 14 mice for a and n = 16 mice for (b). All data are expressed as quantiles. The P values were obtained using a linear regression and two-sided F-test, adjusted for multiple comparison using the Benjamini-Hochberg procedure (that is, bh-p value). (c, d) Diversity score correlation with the number of genes expressed per tissue (c) or tissue cell-type (d). The red line corresponds to the linear regression curve. (e) Trajectory analysis for a brain myeloid microglia cell. (f) Heat map showing differential gene expression analysis of cluster 10 (mostly young macrophages) compared with cluster 13 (mostly old macrophages). For the complete gene list, see Supplementary Table 10.

Supplementary information is available online.

Supplementary Information

This file contains a full list of authors in the Tabula Muris Consortium

Reporting Summary

Supplementary Table 1: Summary of the FACS dataset. (a) Number of cells grouped by age, sex, mouse id and tissue. (b) Number of cells grouped by tissue, cell ontology class and age. (c) Number of cells grouped by Louvain cluster number, cell ontology class, tissue and age. (d) Number of cells grouped by cell ontology class, Louvain cluster number, tissue and age. (e) Fraction of cells in each Louvain cluster per cell ontology class and tissue. (f) Fraction of cells in each Louvain cluster per tissue. (g) Fraction of cells in each Louvain cluster per cell ontology class.

Supplementary Table 2: Summary of the droplet dataset. (a) Number of cells grouped by age, sex, mouse id and tissue. (b) Number of cells grouped by tissue, cell ontology class and age. (c) Number of cells grouped by Louvain cluster number, cell ontology class, tissue and age. (d) Number of cells grouped by cell ontology class, Louvain cluster number, tissue and age. (e) Fraction of cells in each Louvain cluster per cell ontology class and tissue. (f) Fraction of cells in each Louvain cluster per tissue. (g) Fraction of cells in each Louvain cluster per cell ontology class.

Supplementary Table 3: Summary of the batch corrected dataset (FACS and droplet). (a) Number of cells grouped by age, sex, mouse id and tissue. (b) Number of cells grouped by tissue, cell ontology class and age. (c) Number of cells grouped by Louvain cluster number, cell ontology class, tissue and age. (d) Number of cells grouped by cell ontology class, Louvain cluster number, tissue and age. (e) Fraction of cells in each Louvain cluster per cell ontology class and tissue. (f) Fraction of cells in each Louvain cluster per tissue. (g) Fraction of cells in each Louvain cluster per cell ontology class. (h) Adjusted rand score before and after bbknn batch correction for method (droplet, FACS).

Supplementary Table 4: Cellular fraction changes for senescence markers. This Supplementary Table supports Figure 2a-d. We performed linear regression analysis to identify for which senescence marker genes the proportion of cells expressing that marker significantly (p -value <0.05) changed with age for a hypothesis test whose null hypothesis is that the slope is zero, using two-sided Wald Test with t-distribution of the test statistic. Correlation coefficients were computed using the Pearson's correlation metrics.

Supplementary Table 5: Cellular fraction changes. This Supplementary Table supports Figure 2e,g,i; Extended Data Figure 5c,e,g and Extended Data Figure 6a. We performed linear regression analysis to identify which cellular relative abundances change significantly with age (p -value <0.05 and $r^2>0.7$) for a hypothesis test whose null hypothesis is that the slope is zero, using two-sided Wald Test with t-distribution of the test statistic. Correlation coefficients were computed using the Pearson's correlation metrics.

Supplementary Table 6: Differential gene expression analysis. This Supplementary Table supports Figure 2f,h,j; Extended Data Figure 5d,f,h and Extended Data Figure 6e. The differential gene expression was performed using MAST (57). For a more detailed explanation please refer to the Methods section.

Supplementary Table 7: Quantification of Liver in-situ staining. This Supplementary Table supports Figure 6b-d,g-j and l-o. fov stands for field of view.

Supplementary Table 8: Summary statistics for the GATK analysis. This Supplementary Table supports Figure 3 and Extended Data Figures 7,8. *Cell* is the unique cell identifier; *ercc* is the average number of mutations per cell found in the ERCC spike-in, *adata* is the average number of mutations per cell in the gene set of the tissue; *ercc_raw_counts* are the average number of ERCC spike-in counts per cell and *ercc_counts* are the $\log(\text{ercc_raw_counts}+1)$; *adata_raw_counts* are the average number of gene counts per cell and *adata_counts* are the $\log(\text{adata_raw_counts}+1)$; *tissue*, *age* and *cell_ontology_class* are the metadata of the respective cell id and *agemum* is the age as a numerical variable; *functional_annotations* is a categorical variable binning each cell type as endothelial, immune, parenchymal, stem cell/progenitor or stromal.

Supplementary Table 9: B-cell and T-cell repertoire analysis raw data. This table supports Figure 4a, b and Extended Data Figure 9.

Supplementary Table 10: Differential gene expression for the tissue cell type whose diversity significantly changes with age. (a) FACS brain myeloid microglia differentially upregulated genes between clusters 10, 12 and 14 versus clusters 1 and 6. (b) FACS brain myeloid microglia differentially upregulated genes between clusters 1 and 6 versus clusters 10, 12 and 14. (c) Droplet kidney macrophage differentially upregulated genes between cluster 13 and cluster 10. (d) Droplet kidney macrophage differentially upregulated genes between cluster 10 and cluster 13. (e) Alzheimer's disease microglia signature from (47). This Supplementary Table supports Figure 4d-f and Extended Data Figure 11f. Differential gene expression was computed using Wilcoxon rank-sum adjusted for multiple comparison using the Benjamini-Hochberg procedure.

Chapter 7

Discussion

Linda J. van Weele¹

¹Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA

Chapter 7

The mammary gland is a dynamic tissue. Fueled by hormones, it changes during different phases of life. During pregnancy, mammary gland cells differentiate that provide offspring with milk during lactation. Upon weaning, these cells undergo apoptosis, enabling the tissue to quickly return to a pre-pregnancy state. This cycle is repeated each pregnancy. The mammary gland is also a common site for oncogenesis, breast cancer is estimated to affect 1 in 7 women at some point in their lives in both the Netherlands as well as in the United States. Cancer develops when cells lose sensitivity to the signals that regulate tissue homeostasis and start proliferating uncontrollably, negatively affecting the tissue(s) around them. In this thesis, my aim was to increase our understanding of the transition of a normal cell to an oncogenic cell. In **chapters 2 and 4**, we looked at intracellular changes resulting in a cell's changed behavior in response to its microenvironment. In **chapter 3**, we studied tumor cellular plasticity upon the inactivated expression of an oncogene. In **chapters 5 and 6**, we provided scientists with a framework of what a normal expression profile looks like, by assembling an atlas of the transcriptome of the mouse at different stages of life, aiming to help us unravel how expression is altered in a state of disease.

In this discussion, I will further elaborate on the topics presented in each chapter and discuss common themes among the different chapters. Namely, I will touch up on the epithelial-to-mesenchymal (EMT) transition, an example of cellular plasticity that we encountered in our research. Furthermore, I will discuss how on one hand, the timing and location of a mutation are deterministic for the prospect of a tumor to develop, and on the other hand, developments in personalized medicine and tumor sequencing provide avenues to surpass the tissue of origin as the main criteria for treatment decision. Preclinical research plays an important role in the improvement of patient stratification. Lastly, I will discuss preclinical models and aim to explain how the results in this thesis are not always in line with other published observations.

Cellular plasticity in tumors: EMT

Cancer plasticity has been observed *in vivo*. For example, in a breast cancer mouse model fluorescently labeled tumors grew and shrank in succession, showing that the fate of tumor formation changed continuously (1). One type of cellular plasticity is EMT. EMT plays an important role in embryonic development. During embryogenesis, epithelial cells lose their cell adherent properties to transition into migratory mesenchymal cells. EMT is also observed in cancer. In cancer, it was initially thought to be an essential component for the successful dissemination and secondary colonization of solid tumor cells (2). More recently, the contribution of EMT to malignancy has proven to be more complex. The mesenchymal identity may help cells to migrate away from the primary tumor site, but the mesenchymal state does not necessarily result in higher metastatic success (3). Cells that have lost their epithelial identity and express mesenchymal genes, even showed a decreased capacity to form tumors (4). Reports have shown that EMT does induce chemoresistance that contributes to malignancy (5, 6). In **chapter 3**, we observed the EMT signature in the relapse tumor. However, signs of metastases to the liver or lung were absent, two common sites of breast cancer metastasis (7). Of note, the observation of metastasis is usually rare in mouse models (8). Thus, the absence of metastasis in our mouse model might not be representative for human cancer. So how can does some data

link the mesenchymal signature to metastasis and other data does not? A possible explanation is the plasticity between the epithelial and mesenchymal states. EMT and MET - the reverse process of mesenchymal-to-epithelial transition - are not binary processes, various intermediate states exist (9).

Timing and occurrence of mutations determine risk malignancy will develop

The use of genetic sequencing has enabled us to distinguish different molecular subtypes in cancer. This knowledge has led to the development of targeted therapies, designed to interfere with specific proteins and cellular processes that a tumor depends on. Commonly altered proteins that are predicted or have been shown to drive cancer, are called cancer drivers. However, the sole presence of such a mutated gene in a tissue does not guarantee the development of a tumor. An example is found in colon cancer. KRAS, APC, and TP53 are cancer drivers in colon adenocarcinoma (10). Crypts carrying an oncogenic mutation in *Kras*, *Apc*, or *Trp53* expand in the normal mouse intestine (11, 12). However, the stem cells that carry these oncogenic mutations are regularly replaced by stem cells lacking such mutations and no tumors are formed (11). Here, the replacement of mutant cells is a process fueled by the epithelial cells, not by the immune system (13), and has also been reported in the skin (14). Another example is a study where canine epithelial kidney cells with a wildtype copy of *Trp53* outcompete the same cells with mutant *Trp53* unless oncogenic *Ras* is activated (15). Other colon cancer research has shown that the order of mutation acquisition is important for initiation and the pace of tumor development (16, 17). These studies suggest that the solo acquisition of a cancer driver mutation is not always sufficient to lead to malignancy, other circumstances play a role too. *TP53* mutations have been studied in a wide array of tumors and are thought to get acquired later in tumor development (18).

Timing and location of *TP53* mutations

TP53 is the most widespread mutated cancer drive gene in cancer (10). However, the acquisition of *TP53* mutations is usually not an early event in tumor development (18). For example, in breast cancer, mutations in *Trp53* exon 5 and 8 do not only correlate with severity in patient tumor grade and nodal status, but also tumor size (19). In our mammary gland-specific *Trp53*^{-/-} mouse model that we crossed with various other transgenes for the studies described in **chapters 2 and 3**, we did not observe breast cancer development. A phenotype that changed in a subset of mice once *Kras*^{G12D} expression was introduced in the mammary gland as well. There is more evidence that the timing of inactivation is meaningful for tumor formation. Patients with the *TP53* mutated inherited condition Li-Fraumeni, develop different types of tumors than patients with *TP53* spontaneous mutations. For example, Li-Fraumeni patients commonly develop ER-positive breast cancer between 18-45 years of age while spontaneous *TP53* mutations are most common in triple-negative breast cancer (20, 21). Interestingly, Li-Fraumeni patients develop a pattern of tumors in different stages of life. Specific tumor types are more common to manifest in childhood, in young adults, or later in life in Li-Fraumeni patients. It has been hypothesized that for certain cancers the sole absence of *TP53* is sufficient to develop, while for others additional genetic alterations need to develop which takes time (22). The widespread tumor development of Li-Fraumeni syndrome is the exception rather

Chapter 7

than the rule among known germline mutations that fuel cancer, as usually cancer-causing germline mutations result in a more constricted cancer pattern. For example, inherited mutations in *BRCA1* and *BRCA2* resulting in breast and ovarian cancer and inherited mutations in *KIT* resulting in GIST (23). The widespread tumor formation in Li-Fraumeni syndrome patients correlates with the notion that *TP53* mutations have been identified as cancer drivers across most cancer types (10).

Tumor cell of origin

Not only the timing and tissue of mutation acquisition matter, also the specific cell type where the tumor first developed plays a role. For example, in studies done in the mouse mammary gland, the *Pik3ca*^{H1047R} substitution mutation results in different breast cancer subtypes depending on the cell type the tumor develops from (24, 25). A possible explanation of why the same mutated cancer driver has a different effect in different tissues are variations in their epigenetic profiles. The epigenetic state is formed during development and by a cell's environment. It determines what genes are expressed and what genes are not. Therefore, a known cancer driver that is mutated but is not or barely expressed, will not have the same effect as if this gene was highly expressed. One such an example is seen when *BRAF*^{V600E}-mutated colon cancer and *BRAF*^{V600E}-mutated melanoma are compared. Drugs inhibiting *BRAF*^{V600E} are highly effective at treating *BRAF*^{V600E}-mutated melanoma (26), but not at treating *BRAF*^{V600E}-mutated colon cancer (27). It was shown that EGFR - a cell surface receptor that is active upstream of BRAF - is highly expressed in the *BRAF*^{V600E}-mutated colon cancer but not in *BRAF*^{V600E}-mutated melanoma. The inhibition of *BRAF*^{V600E} results in a positive feedback loop of EGFR activation. Consequently, colon cancer cells are sensitive to enhanced EGFR activation due to the large presence of EGFR in these cells, unlike the melanoma cells (28). These were examples where either the cell of origin affects the tumor subtype developing or affects the success of treatment. Therefore, solely following tumor sequencing data to find targetable mutations, does not provide a complete picture. However, are there other instances in which different tumor types share the same driver, where patients be treated with the same protocol? To look for this, clinical trials have been designed that take in patients with a wide variety of cancer types that have the targetable mutation in common.

The next step in personalized medicine: does the tumor type matter?

Traditionally, treatment plans are based on the organ a tumor was detected in. Cancer genomics has altered this approach. Sequencing of a tumor has led to the discovery of molecular subtypes in a certain cancers, including breast cancer (29, 30). As described earlier in this discussion, it has also allowed for the identification of cancer driver mutations and subsequently started driving the discovery of targeted drugs focused on taking advantage of a tumor's dependency of cancer drivers. This has resulted in the development of basket clinical trials (23, 31). In a basket clinical trial, the inclusion criteria focus on matching a specific mutation with a treatment plan, instead of a tumor type. The NCI-MATCH Trial (Molecular Analysis for Therapy Choice), also known as MATCH, is such a basket trial. Small groups of patients are assigned a treatment based on genetic detected in their tumors. MATCH continues to enroll patients - with a focus on rare cancers - in various clinical trials. Currently, a handful of trials has already been completed and the results have been shared

(32-38). For example, a trial enrolling patients with the *BRAF*^{V600E} mutation that were treated with a combination of the *BRAF*^{V600E} inhibitor dabrafenib and the MEK1/2 inhibitor trametinib resulted in responses in 7 different cancer types warranting further investigation. This fuels the hope that this treatment plan can work for additional patients than currently included in approvals (36). However, in a basket clinical trials treating patients with *BRAF* mutations other than *BRAF*^{V600E} with trametinib monotherapy, did not show clinical efficacy (34). A French basket trial, comparing targeted therapy in patients with advanced disease with a treatment chosen by their physician, did not result in a difference in progression-free survival among the two groups either (39). The results from these basket clinical trials show us that we need more information than solely matching the driver gene with the targeted therapy to treat patients effectively. We also discuss this phenomenon in **chapter 3**, mutations in cancer drivers are not independent of their context. For example, *RAS* mutations are highly prevalent in pancreatic ductal adenocarcinoma, colorectal adenocarcinoma, multiple myeloma, and various other cancers, but not commonly found in breast tumors (10, 40). However, the pathways that *RAS* is central to, such as MAPK/ERK and PI3K/AKT/mTOR signaling, are frequently overactivated in breast cancer (41-43). This shows that even though inhibiting an actionable mutation can be meaningful across different cancers, there is also more knowledge that needs be gained on tissue-specific and cell-specific cancer signaling. Therefore, preclinical research will need to improve our predictive knowledge to enroll patients more successfully in clinical trials to move the success of personalized medicine forward.

Preclinical model systems

Various model systems were used for the work described in this thesis. The most dominant model system deployed was the mouse. In addition, experiments taking advantage of the strengths of the organoid system, patient-derived xenografts (PDXs), and cell lines were included. All four are used in cancer research. Organoids, PDXs, and mouse models are common model systems in stem cell research as well. All have different advantages and disadvantages that I will outline below.

Cell lines

Cell lines provide a practically indefinite resource of homogenous material. They are easy, quick, and relatively cheap to work with. The environment a cell line is exposed to is highly controllable, enabling to test the effect of a single change on the behavior and phenotype of the cell. Cell lines are usually the best choice to answer questions on the molecular level, such as 'What pathway(s) are affected by intervention A' or 'what protein-protein interactions take place in condition B'. Furthermore, due to the ability to scale up a cell line experiment efficiently and reliably, high-throughput testing - including multiple cell lines simultaneously - can be applied to test for drugs that limit cellular proliferation. Even though cell lines still require animal serum to grow, the use of cell lines is considered more ethical since fewer animals are sacrificed than in experiments using laboratory animals. However, data derived from cell line experiments are not always suitable to model what happens *in vivo* in humans. For cell lines to proliferate continuously, they have been described to acquire genetic alterations not found *in vivo*. Mechanisms inhibiting continuous proliferation are inactivated and additional genetic alterations are acquired with every duplication

Chapter 7

cycle. This results in cells that show different properties from cells found in vivo and it results in a homogenous cell populations, unlike the heterogeneity found in human tissues and tumors (44). A practical problem with the use of cell lines is contamination by other cell lines or pathogens. Although this problem has been reported as early as in 1981 (45), the problem remains pervasive as many research groups do not check for this on a regular basis (46, 47). Albeit costlier, another in vitro system in which cells collect fewer mutations, is less likely to get contaminated by other cell types, and that aims to closer resemble in vivo conditions are organoids.

Organoids

Organoids are in vitro simplified model systems of heterogenous epithelial tissues. Organoids are generated from (induced) pluripotent stem cells, tissue stem cells, or a mix of different primary cells containing tissue stem cells (48). In a three-dimensional culture, cells divide and self-organize to form heterogenous structures. In **chapter 2** we use three-dimensional cultures to study mouse mammary epithelial cells' response to TNF α and in **chapter 4** to study the response of human colon and breast cancer cells to TLR2 and CD14 inhibition. Organoids resemble the physiological state of a tissue or tumor more closely than a homogenous, hypermutated cell line does. Since organoids can be made with freshly collected human cells. Therefore, organoids enable the direct modeling of healthy and sick human cells. Furthermore, the organoid setup still facilitates easy manipulation. Downsides of organoids compared to cell lines are that organoids are costlier, generally require more work to establish and maintain, and have a more definitive lifespan. Furthermore, protocol standardization among different research groups is lacking for organoids, increasing the variability in data outcomes between different research groups (49). The latter is not surprising, as we are still in the midst of improving culture conditions to mimic physiological conditions closely. Organoids lack cells and structures commonly surrounding a tissue in vivo; such as endothelial cells, fibroblasts, immune cells, and the microbiome. In addition, bioengineers work on technical challenges such as optimizing the scaffolding organoids need to form three-dimensional cultures and imitating mechanical stimuli provided in vivo (50). Moreover, there is no standardization among research groups how to verify an organoid culture resembles a tissue, enabling a large degree of plasticity that may only display in specific culture conditions (51). Plasticity has also been observed in organoids derived from human breast tissue. Organoids from both the basal population - shown to contain the mammary stem cells (52, 53) - and the luminal population generate both lineages (54). The plasticity observed in two-dimensional cultures is even more diverse. In long-term culture, human mammary epithelial cells acquired epigenetic and genetic changes enabling the cells to survive in vitro beyond the senescence barrier. Moreover, these cells start expressing breast cancer markers (55-59). In short-term culture, human mammary epithelial cells do not only form luminal and basal cells but also bipotent progenitor/stem cells (60). It is unclear if such plasticity also takes place in normal physiology. Beyond modeling normal and sick tissues, organoids can also play a major role in the field of drug screening. Today preclinical screening is not at the level of two-dimensional cell line screening yet (61). In summary, a lot of progress has been made in organoid culture to model in vivo conditions, especially in certain tissues. Ideally, organoids will one day replace the use of transgenic and PDX animal models (62). However today,

organoid systems are not a substitute yet for the complex network of interactions found in animals.

Mouse models

Just like humans, mice are mammals. However, their lifecycle is much shorter. The gestation period ranges from 18 to 22 days and the ovulation of inbred strains starts between 6 and 8 weeks of age. Furthermore, they are one of the smaller mammal species making care taking relatively affordable. As mice provide a common model system for a wide variety of research, many tools have been developed and much knowledge has been acquired. One such tool has been the development of transgenic mice. Using genetic modification, one or multiple genes can be knocked in, knocked out, replaced by a mutant variant, or genetic constructs can be inserted such as the inducible Tet-on transgene used in **chapter 3**. Due to the use and maintenance of inbred strains with minimal genetic diversity, standards that are maintained by organizations such as the Jackson Laboratory and Charles River, research done among different labs all over the world using the same strain can be compared. Of note, experiments done in different mouse strains can lead to different results, as mentioned later in this discussion. Despite a high level of similarity among mice from one strain, there is always some variation just like in humans. This variation is caused by the high level of complexity of an animal and its environment, that is not found in culture systems. In **chapter 3**, mice from the same C57BL/6J strain with the same transgenes and exposed to the same conditions, did not only display a different level of susceptibility in tumor development but also showed to develop tumors with a different FACS profile and dependency on $Kras^{G12D}$. To take such differences into account, always multiple mice are used per experimental group. Wherefore, also in **chapters 5 and 6** multiple mice per age group are included to provide reliable data. Another important characteristic of syngeneic mouse models is the presence of the immune system. Prolonged inflammation contributes to tumor progress (63) and in immunotherapy the patients' immune system is modulated to drive an antitumor response (64), stressing the importance of the contribution of the immune system to tumor modeling. The immune system also interacts with stem cells, playing a role in tissue regeneration and normal homeostasis (65). As discussed in **chapters 2 and 4**, where the interaction between the immune signaling proteins and self-renewal is discussed. For cancer research in particular, a disadvantage of mouse models is the rarity of metastasis. Even though metastasis is the main reason of cancer-related deaths. Metastasis often is an inefficient process with a long latency, while a mouse has a short lifespan. Furthermore, the mouse is a small animal, therefore primary tumor development develops often too quick for metastases to develop (8). Furthermore, there are many biological and genetic differences between humans and preclinical models such as the mouse. Although mouse and human genes share a high level of homology, at the nucleotide and the gene regulation and expression level there are big differences (66-68). This discrepancy results in inaccurate predictions of the success of a drug in humans. Today, only 12.8% of drugs that move from preclinical models into clinical trials, are approved for use in patients (69). The main reason for failure is due to lack of efficacy and the second most common reason is due to safety issues, such as toxicity (70). Looking for preclinical models that better reflect human tumors, scientists have been looking at PDXs.

PDX

PDX tumors can be established by the direct engraftment of patient material in an immunocompromised mouse. Most commonly, the engraftments are tumor fragments or single-cell suspensions. Using this model system, the development of a human tumor and its response to modulations by drugs or other interventions can be studied in vivo, as we did in **chapter 4**. In breast cancer, PDXs have shown to reflect the pathology, metastasis, and survival rate of the patients that the tumors derived from (71, 72). In a PDX screen of various tumor types where they looked for associations between genotype and treatment response, PDX response was representative for in patient response (73, 74). However, single-cell and bulk sequencing data showed that PDX tumors undergo clonal selection and quickly acquire copy number alterations during passages, copy number alterations that differed from the ones observed in human primary samples (75, 76). Hence, how truthfully PDX models represent human tumor development varies among different studies. One method to improve the representation of PDX models, is by co-transplanting components of the tumor microenvironment. Co-transplantation with human mesenchymal cells increases the likelihood that human breast cancer cells establish themselves in an immunocompromised recipient mouse (71). However, co-transplanted human cells have a limited lifespan, thus the tumor stroma is eventually replaced by mouse stroma (77, 78). To study the interaction between the immune system and cancer, especially in the field of immuno-oncology drug development, scientists are creating protocols to transplant human hemopoietic cells in immunocompromised mice as to establish the human immune system (79-81). To do so, immune cells and tumor cells must be HLA-matched to prevent tumor cell rejection by the human immune cells (82, 83). Other challenges this technique faces: graft versus host disease, low success rate of immune cell engraftment, incomplete development of human immune system, ethical complications in models that use fetal tissue to establish the immune system, and the high cost of this system (84). Ideally, supporting cells and tumor cells are derived from the same patient but this is in practice difficult to achieve. In conclusion, although there is great potential in using humanized mouse models to study tumor development, currently it is not ready to be used on the large scale that is desirable.

Possible explanations for the inconsistencies in research outcomes among different research groups

Among the different preclinical models, standardization amid different research groups is lacking. This leads to variations in outcome. For example, in the data shared in **chapters 2 and 3**, other groups working with transgenic mice had reported different results, even though the same (trans)gene(s) were knocked out or introduced. In **chapter 2** we report no increase in self-renewal in *Trp53* knockout mammary epithelial cells, while other groups have observed an increase in self-renewal in the same cells (85-87). In **chapter 3**, we found that the expression alone of mammary gland-specific oncogenic *Kras*^{G12D} did not result in breast tumor formation, while another group observed breast tumor formation in 100% of mice expressing mammary gland-specific *Kras*^{G1D} (88). Discrepancies in results are reported to originate from the use of different protocols and different mouse inbred strains. However, there is also increasing attention for the influence of the microbiome on result outcomes.

Differences in protocol

Firstly, differences in protocols can have a profound effect on results. Hines et al. shared the work of pinpointing how two collaborating labs on opposite American coasts ended up with different FACS profiles and different organoid viability yields using the same samples and same reagents. They concluded that a more vigorous and shorter digestion process resulted in the loss of breast cancer stem cell marker CD44 expression and a 5 times lower yield of organoids (89). Finding the digestion time sweet spot to optimize for mammary epithelial cell yield and viability, has been discussed since the first publications on mouse mammary stem cell isolation came out (90, 91). Different labs follow different protocols (92). However, there is additional data that support that a slower digestions protocol results in higher mammary epithelial cell viability (93). A slower and milder digestion protocol has also been reported to result in higher cell viability and yield using human primary samples (94). Although not precisely documented, I have also observed higher cell yield using an overnight digestion protocol using gentle collagenase and hyaluronidase (C/H) (catalog #07919, STEMCELL Technologies Inc.) than a shorter protocol that included a 2-hour C/H (catalog #07912, STEMCELL Technologies Inc.) digestion step, supporting the above-mentioned reports (89, 93). In the studies reported on in this thesis, we had already used the 2-hour C/H digestion protocol for most data when we tested the overnight digestion protocol. Therefore, we kept the protocol for similar experiments constant while shifting to the overnight gentle C/H digestion protocol for new experiments. This is noted in the Methods section of **chapter 2**. Another example we find in **chapter 5**, we used a FACS-based method and microfluidic-droplet method to isolate single-cells and subsequently their RNA for single-cell RNAseq (scRNAseq) from 20 mouse organs to create a scRNAseq atlas. A couple of months prior, Han et al. shared a complementary project; using a different method called microwell-seq to capture single cells prior to library preparation (95). A comparison of the data generated by these 3 methods, show that the FACS method detects more genes per cells than the other 2 methods. In case of the mammary gland, the medium number of genes detected per cell was for FACS 3,603, for microfluid droplet 1,483, and for microwell-seq 657 genes. This variation in genes detected was exemplary for other tissues as well. Nevertheless, there are good reasons to choose the microfluid droplet or microwell-seq method instead of FACS. These two techniques are cheaper, quicker, do not require cells to undergo the stress that is inherent to cell sorting by FACS, and show a more complete picture of all cell types present in a tissue since no prior enrichment is done. The benefit of FACS is that one can enrich for populations of interest and provide more in-depth data on lowly expressed genes and rare cell types within those selected populations (96). In addition to differences in cell capture and library preparation between Han et al. and our Tabula Muris publications, Han et al. followed a different protocol for the digestion of the mammary gland. They used a shorter protocol than used by us in **chapters 2-6** or published by other groups (92, 97). Without a side-by-side comparison we do not know what the consequences are. In summary, variation among protocols between different research groups can lead to different outcomes.

Differences in mouse strain

Chapter 7

Secondly, the use of different mouse strains can affect results (98, 99). Especially in the field of cancer research, particular inbred mouse strains are known to be more prone to tumor development than others (100-104). This is amply discussed in **chapters 2 and 3**. One additional example that is not discussed in these chapters, are differences in transplantation efficiency with mouse luminal epithelial cells among different research groups, using different strains and different protocols. In our hands, transplanting C57BL/6J mouse mammary epithelial luminal progenitor cells that are mixed with 50% Matrigel (catalog #356234, Corning) do not grow out upon transplantation in recipient syngeneic mice, as reported in **chapter 2**. This contrasts with two other publications that report that when luminal epithelial mammary cells are transplanted in the presence of Matrigel, cells grow out in recipient mice. However, in the absence of Matrigel, no outgrowths are observed (90, 105). Vaillant et al. worked with the FVB/N mouse strain while Jeselsohn et al. worked with the Balb/c mouse strain. Both groups transplanted luminal mammary epithelial cells, our population was further enriched luminal mammary epithelial progenitor cells. Not only the mouse strain is different among these three groups, but there is also variation in the digestion protocols and the number of cells transplanted. In short, there are numerous differences among these three manuscripts. This makes it difficult to conclusively state why our observations differ: does only one variable or multiple variables contribute? To keep variations caused by using different mouse strains as small as possible, institutions like the Jackson laboratory and Charles River are of vital importance. They maintain the various inbred strains and use different methods to confirm for example C57BL/6J mice stay C57BL/6J. A new inbred mouse strain is created by breeding separately for at least 20 generations (106). Such separate breeding results in genetic and phenotypic differences and new substrains, as observed in C57BL/6 derived substrains (99). However, even when research groups use the same mouse strain, the fact that their mice have spent time in separate facilities can also have an impact on results.

Differences in microbiota

Thirdly, experiments performed in a different facility can affect results. Among animal facilities mouse microbiomes vary (107). One reason of variation in the mammary gland microbiome is diet (108). The microbiome lives in symbiosis with us through our body, including the mammary gland (109). When looking at cancer, human breast tumors display a different microbiome than normal breast tissue (110-112). Using human paired normal and ER-positive breast cancer samples, tumor samples showed a lower load of bacterial DNA that drops further with the advance of stage (110), a difference in microbiota between benign and malignant disease (111), and changes among the different breast cancer subtypes (112). How these distinctions in microbiome affect disease development and outcome, has yet to be unraveled.

Reproducibility is considered to be one of the pillars of science (113). To obtain this, observations need to be considered in context of the model system used, the protocol followed, and most likely also the environment the experiments are performed in. The hope is that as our preclinical models improve and we better our predictions of how a tumor in each individual patient forms and progresses, we will increase our capability to provide each patient with a treatment plan that is optimized for them.

Concluding remarks

The research presented in this thesis provides us with novel insights into the link between essential self-renewal/proliferation and (the early phases of) damaging tumorigenesis. Pathways, first described as essential for the proper functioning of the immune system, also influence typical tissue development and atypical tissue development - such as too much proliferation leading to tumorigenesis - directly in non-immune cells. Another common theme in this thesis is the use of omics, namely transcriptomics and genomics. These technologies have made it possible to make encompassing comparisons between large numbers of cells or samples. Since most research described here encompassed the development and exploration of new mouse models, analyzing the transcriptome was instrumental in providing clues of what mechanisms could be responsible for and should be validated to understand an observed phenotype. Creating and understanding new mouse models requires time and labor. Hence various projects presented here would benefit from follow-up studies. **Chapter 2** triggers an interest in the role of TNF α in mammary gland cellular proliferation and provides clues that TNF α could lead to tumor formation when circumstances allow. The mouse tumor model presented in **chapter 3** should be analyzed in much more detail to answer questions such as: why does only a subset of mice develop tumors? What is the reason different tumor types develop? What changes during tumor decline and remission enable tumor growth reactivation? Answers to these questions are important to prevent tumors from adapting to drug-induced repression of an oncogene such as KRAS. **Chapter 4** provides novel treatment avenues that need more work to determine if these findings can be translated into therapies. And lastly, the setup of the projects described in **chapters 5 and 6** should be - and will be - repeated with other model organisms and human samples to expand its impact for the scientific community.

REFERENCES

- Zomer A, Ellenbroek SI, Ritsma L, Beerling E, Vriskoop N, Van Rheenen J. Intravital imaging of cancer stem cell plasticity in mammary tumors. *Stem Cells*. 2013;31(3):602-6.
- Thiery JP. EMT: An Update. *Methods Mol Biol*. 2021;2179:35-9.
- Beerling E, Seinstra D, de Wit E, Kester L, van der Velden D, Maynard C, et al. Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep*. 2016;14(10):2281-8.
- Sikandar SS, Kuo AH, Kalisky T, Cai S, Zabala M, Hsieh RW, et al. Role of epithelial to mesenchymal transition associated genes in mammary gland regeneration and breast tumorigenesis. *Nat Commun*. 2017;8(1):1669.
- Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature*. 2015;527(7579):525-30.
- Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong ST, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature*. 2015;527(7579):472-6.
- Chen MT, Sun HF, Zhao Y, Fu WY, Yang LP, Gao SP, et al. Comparison of patterns and prognosis among distant metastatic breast cancer patients by age groups: a SEER population-based analysis. *Sci Rep*. 2017;7(1):9254.
- Gómez-Cuadrado L, Tracey N, Ma R, Qian B, Brunton VG. Mouse models of metastasis: progress and prospects. *Dis Model Mech*. 2017;10(9):1061-74.
- Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. *Nature*. 2018;556(7702):463-8.
- Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*. 2018;173(2):371-85.e18.
- Vermeulen L, Morrissey E, van der Heijden M, Nicholson AM, Sottoriva A, Buczacki S, et al. Defining stem cell dynamics in models of intestinal tumor initiation. *Science*. 2013;342(6161):995-8.
- Snippert HJ, Schepers AG, van Es JH, Simons BD, Clevers H. Biased competition between Lgr5 intestinal stem cells driven by oncogenic

- mutation induces clonal expansion. *EMBO Rep.* 2014;15(1):62-9.
13. Kon S, Ishibashi K, Katoh H, Kitamoto S, Shirai T, Tanaka S, et al. Cell competition with normal epithelial cells promotes apical extrusion of transformed cells through metabolic changes. *Nat Cell Biol.* 2017;19(5):530-41.
 14. Brown S, Pineda CM, Xin T, Boucher J, Suzzo KC, Park S, et al. Correction of aberrant growth preserves tissue homeostasis. *Nature.* 2017;548(7667):334-7.
 15. Watanabe H, Ishibashi K, Mano H, Kitamoto S, Sato N, Hoshiba K, et al. Mutant p53-Expressing Cells Undergo Necroptosis via Cell Competition with the Neighboring Normal Epithelial Cells. *Cell Rep.* 2018;23(13):3721-9.
 16. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. *N Engl J Med.* 1988;319(9):525-32.
 17. Takeda H, Wei Z, Koso H, Rust AG, Yew CC, Mann MB, et al. Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor progression. *Nat Genet.* 2015;47(2):142-50.
 18. Rivlin N, Brosh R, Oren M, Rotter V. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes Cancer.* 2011;2(4):466-74.
 19. Olivier M, Langerød A, Carrieri P, Bergh J, Klaar S, Eyfjord J, et al. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. *Clin Cancer Res.* 2006;12(4):1157-67.
 20. Levine AJ, Puzio-Kuter AM, Chan CS, Hainaut P. The Role of the p53 Protein in Stem-Cell Biology and Epigenetic Regulation. *Cold Spring Harb Perspect Med.* 2016;6(9).
 21. Levine AJ, Chan CS, Dudgeon C, Puzio-Kuter A, Hainaut P. The Evolution of Tumors in Mice and Humans with Germline p53 Mutations. *Cold Spring Harb Symp Quant Biol.* 2015;80:139-45.
 22. Levine AJ. p53: 800 million years of evolution and 40 years of discovery. *Nat Rev Cancer.* 2020;20(8):471-80.
 23. Schneider G, Schmidt-Supprian M, Rad R, Saur D. Tissue-specific tumorigenesis: context matters. *Nat Rev Cancer.* 2017;17(4):239-53.
 24. Koren S, Reavie L, Couto JP, De Silva D, Stadler MB, Roloff T, et al. PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. *Nature.* 2015;525(7567):114-8.
 25. Van Keymeulen A, Lee MY, Ousset M, Brohée S, Rorive S, Giraddi RR, et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature.* 2015;525(7567):119-23.
 26. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med.* 2011;364(26):2507-16.
 27. Kopetz S, Desai J, Chan E, Hecht J, O'dwyer P, Lee R, et al. PLX4032 in metastatic colorectal cancer patients with mutant BRAF tumors. *Journal of Clinical Oncology.* 2010;28(15_suppl):3534-.
 28. Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, et al. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature.* 2012;483(7387):100-3.
 29. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature.* 2000;406(6797):747-52.
 30. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012;486(7403):346-52.
 31. Redig AJ, Jänne PA. Basket trials and the evolution of clinical trial design in an era of genomic medicine. *J Clin Oncol.* 2015;33(9):975-7.
 32. Jhaveri K, Wang X, Makker V, Luoh S-W, Mitchell E, Zwiebel J, et al. Ado-trastuzumab emtansine (T-DM1) in patients with HER2-amplified tumors excluding breast and gastric/gastroesophageal junction (GEJ) adenocarcinomas: results from the NCI-MATCH trial (EAY131) subprotocol Q. *Annals of Oncology.* 2019;30(11):1821-30.
 33. Azad NS, Gray RJ, Overman MJ, Schoenfeld JD, Mitchell EP, Zwiebel JA, et al. Nivolumab is effective in mismatch repair-deficient noncolorectal cancers: Results from arm Z1D-A subprotocol of the NCI-MATCH (EAY131) study. *Journal of Clinical Oncology.* 2020;38(3):214.
 34. Johnson DB, Zhao F, Noel M, Riely GJ, Mitchell EP, Wright JJ, et al. Trametinib activity in patients with solid tumors and lymphomas harboring BRAF non-V600 mutations or fusions: results from NCI-MATCH (EAY131). *Clinical Cancer Research.* 2020;26(8):1812-9.
 35. Chae YK, Hong F, Vaklavas C, Cheng HH, Hammerman P, Mitchell EP, et al. Phase II study of AZD4547 in patients with tumors harboring aberrations in the FGFR pathway: results from the NCI-MATCH trial (EAY131) subprotocol W. *Journal of Clinical Oncology.* 2020;38(21):2407.
 36. Salama AK, Li S, Macrae ER, Park J-I, Mitchell EP, Zwiebel JA, et al. Dabrafenib and trametinib in patients with tumors with BRAFV600E mutations: Results of the NCI-MATCH trial subprotocol H. *Journal of Clinical Oncology.* 2020;38(33):3895-904.
 37. Kalinsky K, Hong F, McCourt CK, Sachdev JC, Mitchell EP, Zwiebel JA, et al. Effect of

- capivasertib in patients with an AKT1 E17K-mutated tumor: NCI-MATCH subprotocol EAY131-Y nonrandomized trial. *JAMA oncology*. 2021;7(2):271-8.
38. Cleary JM, Wang V, Heist RS, Kopetz ES, Mitchell EP, Zwiebel JA, et al. Differential Outcomes in Codon 12/13 and Codon 61 NRAS-Mutated Cancers in the Phase II NCI-MATCH Trial of Binimetinib in Patients with NRAS-Mutated Tumors. *Clinical Cancer Research*. 2021;27(11):2996-3004.
 39. Le Tourneau C, Delord JP, Gonçalves A, Gavoille C, Dubot C, Isambert N, et al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet Oncol*. 2015;16(13):1324-34.
 40. Hobbs GA, Der CJ, Rossman KL. RAS isoforms and mutations in cancer at a glance. *J Cell Sci*. 2016;129(7):1287-92.
 41. Network CGA. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70.
 42. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol*. 2007;8(5):R76.
 43. Miricescu D, Totan A, Stanescu-Spinu II, Badoiu SC, Stefani C, Greabu M. PI3K/AKT/mTOR Signaling Pathway in Breast Cancer: From Molecular Landscape to Clinical Aspects. *Int J Mol Sci*. 2020;22(1).
 44. Dai X, Cheng H, Bai Z, Li J. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J Cancer*. 2017;8(16):3131-41.
 45. Nelson-Rees WA, Daniels DW, Flandermeyer RR. Cross-contamination of cells in culture. *Science*. 1981;212(4493):446-52.
 46. MacLeod RA, Dirks WG, Matsuo Y, Kaufmann M, Milch H, Drexler HG. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int J Cancer*. 1999;83(4):555-63.
 47. Capes-Davis A, Theodosopoulos G, Atkin I, Drexler HG, Kohara A, MacLeod RA, et al. Check your cultures! A list of cross-contaminated or misidentified cell lines. *Int J Cancer*. 2010;127(1):1-8.
 48. Huch M, Koo BK. Modeling mouse and human development using organoid cultures. *Development*. 2015;142(18):3113-25.
 49. Kim J, Koo BK, Knoblich JA. Human organoids: model systems for human biology and medicine. *Nat Rev Mol Cell Biol*. 2020;21(10):571-84.
 50. Hofer M, Lutolf MP. Engineering organoids. *Nat Rev Mater*. 2021:1-19.
 51. Fu NY, Nolan E, Lindeman GJ, Visvader JE. Stem Cells and the Differentiation Hierarchy in Mammary Gland Development. *Physiol Rev*. 2020;100(2):489-523.
 52. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med*. 2009;15(8):907-13.
 53. Eirew P, Stingl J, Raouf A, Turashvili G, Aparicio S, Emerman JT, et al. A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability. *Nat Med*. 2008;14(12):1384-9.
 54. Rosenbluth JM, Schackmann RCJ, Gray GK, Selfors LM, Li CM, Boedicker M, et al. Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. *Nat Commun*. 2020;11(1):1711.
 55. Holst CR, Nuovo GJ, Esteller M, Chew K, Baylin SB, Herman JG, et al. Methylation of p16(INK4a) promoters occurs in vivo in histologically normal human mammary epithelia. *Cancer Res*. 2003;63(7):1596-601.
 56. Garbe JC, Bhattacharya S, Merchant B, Bassett E, Swisshelm K, Feiler HS, et al. Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells. *Cancer Res*. 2009;69(19):7557-68.
 57. Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, et al. Defining the cellular precursors to human breast cancer. *Proc Natl Acad Sci U S A*. 2012;109(8):2772-7.
 58. Roy S, Gascard P, Dumont N, Zhao J, Pan D, Petrie S, et al. Rare somatic cells from human breast tissue exhibit extensive lineage plasticity. *Proc Natl Acad Sci U S A*. 2013;110(12):4598-603.
 59. Breindel JL, Skibinski A, Sedic M, Wronski-Campos A, Zhou W, Keller PJ, et al. Epigenetic Reprogramming of Lineage-Committed Human Mammary Epithelial Cells Requires DNMT3A and Loss of DOT1L. *Stem Cell Reports*. 2017;9(3):943-55.
 60. Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc Natl Acad Sci U S A*. 2011;108(19):7950-5.
 61. Brancato V, Oliveira JM, Correlo VM, Reis RL, Kundu SC. Could 3D models of cancer enhance drug screening? *Biomaterials*. 2020;232:119744.
 62. Jackson SJ, Thomas GJ. Human tissue models in cancer research: looking beyond the mouse. *Dis Model Mech*. 2017;10(8):939-42.
 63. Coussens LM, Werb Z. Inflammation and cancer. *Nature*. 2002;420(6917):860-7.
 64. Hiam-Galvez KJ, Allen BM, Spitzer MH. Systemic immunity in cancer. *Nat Rev Cancer*. 2021;21(6):345-59.

Chapter 7

65. Abnave P, Ghigo E. Role of the immune system in regeneration and its dynamic interplay with adult stem cells. *Semin Cell Dev Biol.* 2019;87:160-8.
66. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002;420(6915):520-62.
67. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature.* 2014;515(7527):355-64.
68. Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, et al. Comparison of the transcriptional landscapes between human and mouse tissues. *Proc Natl Acad Sci U S A.* 2014;111(48):17224-9.
69. Yamaguchi S, Kaneko M, Narukawa M. Approval success rates of drug candidates based on target, action, modality, application, and their combinations. *Clin Transl Sci.* 2021;14(3):1113-22.
70. Arrowsmith J, Miller P. Trial watch: phase II and phase III attrition rates 2011-2012. *Nat Rev Drug Discov.* 2013;12(8):569.
71. DeRose YS, Wang G, Lin YC, Bernard PS, Buys SS, Ebbert MT, et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat Med.* 2011;17(11):1514-20.
72. Eyre R, Alférez DG, Spence K, Kamal M, Shaw FL, Simões BM, et al. Patient-derived Mammosphere and Xenograft Tumour Initiation Correlates with Progression to Metastasis. *J Mammary Gland Biol Neoplasia.* 2016;21(3-4):99-109.
73. Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med.* 2015;21(11):1318-25.
74. Izumchenko E, Paz K, Ciznadija D, Sloma I, Katz A, Vasquez-Dunddel D, et al. Patient-derived xenografts effectively capture responses to oncology therapy in a heterogeneous cohort of patients with solid tumors. *Ann Oncol.* 2017;28(10):2595-605.
75. Ben-David U, Ha G, Tseng YY, Greenwald NF, Oh C, Shih J, et al. Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat Genet.* 2017;49(11):1567-75.
76. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature.* 2015;518(7539):422-6.
77. Cassidy JW, Caldas C, Bruna A. Maintaining Tumor Heterogeneity in Patient-Derived Tumor Xenografts. *Cancer Res.* 2015;75(15):2963-8.
78. Morton JJ, Bird C, Refaeli Y, Jimeno A. Humanized Mouse Xenograft Models: Narrowing the Tumor-Microenvironment Gap. *Cancer Res.* 2016;76(21):6153-8.
79. Shultz LD, Saito Y, Najima Y, Tanaka S, Ochi T, Tomizawa M, et al. Generation of functional human T-cell subsets with HLA-restricted immune responses in HLA class I expressing NOD/SCID/IL2r gamma(null) humanized mice. *Proc Natl Acad Sci U S A.* 2010;107(29):13022-7.
80. Lee J, Dykstra B, Spencer JA, Kenney LL, Greiner DL, Shultz LD, et al. mRNA-mediated glycoengineering ameliorates deficient homing of human stem cell-derived hematopoietic progenitors. *J Clin Invest.* 2017;127(6):2433-7.
81. Yao LC, Aryee KE, Cheng M, Kaur P, Keck JG, Brehm MA. Creation of PDX-Bearing Humanized Mice to Study Immuno-oncology. *Methods Mol Biol.* 2019;1953:241-52.
82. Ishikawa F, Yasukawa M, Lyons B, Yoshida S, Miyamoto T, Yoshimoto G, et al. Development of functional human blood and immune systems in NOD/SCID/IL2 receptor {gamma} chain(null) mice. *Blood.* 2005;106(5):1565-73.
83. Wang M, Yao LC, Cheng M, Cai D, Martinek J, Pan CX, et al. Humanized mice in studying efficacy and mechanisms of PD-1-targeted cancer immunotherapy. *FASEB J.* 2018;32(3):1537-49.
84. Eswaraka J, Giddabasappa A. Humanized mice and PDX models. *Patient Derived Tumor Xenograft Models: Elsevier;* 2017. p. 75-89.
85. Cicalese A, Bonizzi G, Pasi CE, Faretta M, Ronzoni S, Giuliani B, et al. The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells. *Cell.* 2009;138(6):1083-95.
86. Tao L, Roberts AL, Dunphy KA, Bigelow C, Yan H, Jerry DJ. Repression of mammary stem/progenitor cells by p53 is mediated by Notch and separable from apoptotic activity. *Stem Cells.* 2011;29(1):119-27.
87. Chiche A, Moumen M, Petit V, Jonkers J, Medina D, Deugnier MA, et al. Somatic loss of p53 leads to stem/progenitor cell amplification in both mammary epithelial compartments, basal and luminal. *Stem Cells.* 2013;31(9):1857-67.
88. Rädler PD, Wehde BL, Triplett AA, Shrestha H, Shepherd JH, Pfefferle AD, et al. Highly metastatic claudin-low mammary cancers can originate from luminal epithelial cells. *Nat Commun.* 2021;12(1):3742.
89. Hines WC, Su Y, Kuhn I, Polyak K, Bissell MJ. Sorting out the FACS: a devil in the details. *Cell Rep.* 2014;6(5):779-81.
90. Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, et al. Purification and unique properties of mammary epithelial stem cells. *Nature.* 2006;439(7079):993-7.
91. Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, et al. Generation of a functional mammary gland from a single stem cell. *Nature.* 2006;439(7072):84-8.

92. Smalley MJ, Kendrick H, Sheridan JM, Regan JL, Prater MD, Lindeman GJ, et al. Isolation of mouse mammary epithelial subpopulations: a comparison of leading methods. *J Mammary Gland Biol Neoplasia*. 2012;17(2):91-7.
93. Gao H, Dong Q, Chen Y, Zhang F, Wu A, Shi Y, et al. Murine mammary stem/progenitor cell isolation: Different method matters? *Springerplus*. 2016;5:140.
94. Zubeldia-Plazaola A, Ametller E, Mancino M, Prats de Puig M, López-Plana A, Guzman F, et al. Comparison of methods for the isolation of human breast epithelial and myoepithelial cells. *Front Cell Dev Biol*. 2015;3:32.
95. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*. 2018;172(5):1091-107.e17.
96. Liu Z, Wang L, Welch JD, Ma H, Zhou Y, Vaseghi HR, et al. Single-cell transcriptomics reconstructs fate conversion from fibroblast to cardiomyocyte. *Nature*. 2017;551(7678):100-4.
97. Sun H, Xu X, Deng C. Preparation of Single Epithelial Cells Suspension from Mouse Mammary Glands. *Bio Protoc*. 2020;10(4):e3530.
98. Rivera J, Tessarollo L. Genetic background and the dilemma of translating mouse studies to humans. *Immunity*. 2008;28(1):1-4.
99. Mekada K, Abe K, Murakami A, Nakamura S, Nakata H, Moriwaki K, et al. Genetic differences among C57BL/6 substrains. *Exp Anim*. 2009;58(2):141-9.
100. Heston WE, Vlahakis G. Mammary tumors, plaques, and hyperplastic alveolar nodules in various combinations of mouse inbred strains and the different lines of the mammary tumor virus. *Int J Cancer*. 1971;7(1):141-8.
101. Lifsted T, Le Voyer T, Williams M, Muller W, Klein-Szanto A, Buetow KH, et al. Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. *Int J Cancer*. 1998;77(4):640-4.
102. Mahler JF, Stokes W, Mann PC, Takaoka M, Maronpot RR. Spontaneous lesions in aging FVB/N mice. *Toxicol Pathol*. 1996;24(6):710-6.
103. Regua AT, Arrigo A, Doheny D, Wong GL, Lo HW. Transgenic mouse models of breast cancer. *Cancer Lett*. 2021;516:73-83.
104. Rittinghausen S, Dungworth DL, Ernst H, Mohr U. Naturally occurring pulmonary tumors in rodents. *Respiratory System*: Springer; 1996. p. 183-206.
105. Vaillant F, Lindeman GJ, Visvader JE. Jekyll or Hyde: does Matrigel provide a more or less physiological environment in mammary repopulating assays? *Breast Cancer Res*. 2011;13(3):108.
106. MICE* COSGNF. A revision of the standardized genetic nomenclature for mice. *Journal of Heredity*. 1963;54(4):159-60.
107. Parker KD, Albeke SE, Gigley JP, Goldstein AM, Ward NL. Microbiome Composition in Both Wild-Type and Disease Model Mice Is Heavily Influenced by Mouse Facility. *Front Microbiol*. 2018;9:1598.
108. Shively CA, Register TC, Appt SE, Clarkson TB, Uberseder B, Clear KYJ, et al. Consumption of Mediterranean versus Western Diet Leads to Distinct Mammary Gland Microbiome Populations. *Cell Rep*. 2018;25(1):47-56.e3.
109. Urbaniak C, Cummins J, Brackstone M, Macklaim JM, Gloor GB, Baban CK, et al. Microbiota of human breast tissue. *Appl Environ Microbiol*. 2014;80(10):3007-14.
110. Xuan C, Shamonki JM, Chung A, Dinome ML, Chung M, Sieling PA, et al. Microbial dysbiosis is associated with human breast cancer. *PLoS One*. 2014;9(1):e83744.
111. Hieken TJ, Chen J, Hoskin TL, Walther-Antonio M, Johnson S, Ramaker S, et al. The Microbiome of Aseptically Collected Human Breast Tissue in Benign and Malignant Disease. *Sci Rep*. 2016;6:30751.
112. Banerjee S, Wei Z, Tian T, Bose D, Shih NNC, Feldman MD, et al. Prognostic correlations with the microbiome of breast cancer subtypes. *Cell Death Dis*. 2021;12(9):831.
113. The challenges of replication. *Elife*. 2017;6.

Addendum

Nederlandse samenvatting

Curriculum Vitae

List of publications

Dankwoord/acknowledgements

Nederlandse samenvatting

In dit proefschrift beschrijf ik mijn promotieonderzoek naar het borstweefsel en borstkanker.

Zoogdieren, zoals de mens, zijn de enige organismen met borstweefsel. Het doel van borstweefsel is het produceren van melk dat pasgeborenen voedt. Weefsels bestaan voor een groot deel uit cellen. Deze cellen staan niet op zichzelf, ze staan in constante communicatie met de rest van het lichaam. Veranderingen in het lichaam, leiden tot veranderingen in de signalen die cellen ontvangen en afgeven. Hormonen zijn een voorbeeld van signaalmoleculen die door het hele lichaam reizen. Tijdens de zwangerschap verandert de hormoonhuishouding wat tot veranderingen in de architectuur van het borstweefsel leidt: het borstweefsel stoomt zich klaar voor melkproductie en melktransport. Om de juiste cellen hiervoor te produceren, beginnen borstweefselcellen zich te vermenigvuldigen en te specialiseren.

Er zijn twee celtypen die zich kunnen vermenigvuldigen en op deze manier gespecialiseerde dochtercellen kunnen voortbrengen: stamcellen en progenitorcellen. Een groot verschil tussen deze twee celtypen is dat stamcellen elke keer als ze het signaal krijgen te moeten delen, ze dit ook daadwerkelijk kunnen doen. Progenitorcellen daarentegen veranderen een klein beetje na elke deling waardoor ze na een paar delingen uiteindelijk het vermogen om te delen verliezen. Voor elke cel geldt dat de celdeling strikt wordt gereguleerd: celdeling is van belang om een weefsel te onderhouden of te veranderen maar als er meer cellen dan nodig worden gemaakt kan dit tot tumorformatie leiden. Wanneer te veel cellen zich vermenigvuldigen, komt dat doordat de communicatie tussen cellen verstoord is of een cel niet meer naar behoren reageert. Een cel ontvangt foutieve signalen dat het moet delen of een cel reageert niet meer op signalen die deling remmen. Want cellen ontvangen niet alleen signalen van buiten, ook binnen een cel vindt er veel communicatie plaats. Een reden waarom de interne communicatie verstoord kan zijn, zijn veranderingen in het DNA. Deze veranderingen heten mutaties. Het DNA van een individu verwerft op veel plekken mutaties door de jaren heen, maar alleen mutaties op specifieke plaatsen in het DNA hebben zo een sterke invloed dat ze tot ziekte kunnen leiden. Deze specifieke locaties zijn vaak onderdeel van een gen, een stuk DNA waar in de cel kopieën van gemaakt worden. Deze kopieën heten RNA. Het grootste gedeelte van de RNA-moleculen geldt als de instructie voor de bouw van eiwitten. Eiwitten zijn complexe moleculen die essentieel zijn voor de opbouw van het lichaam. Ook het grootste gedeelte van de signaalmoleculen zijn eiwitten, zodoende reguleren zij de communicatie zowel tussen cellen als binnen een cel.

Als een mutatie in een gen het functioneren van een eiwit beïnvloedt, kan dit grote gevolgen hebben voor een cel. In het geval van kanker, kijken wetenschappers vaak naar twee groepen genen: tumorsuppressorgenen en oncogenen. Tumorsuppressorgenen bevatten de code voor eiwitten die de groei van een tumor remmen, mutaties in deze genen die het eiwit afremmen of uitschakelen dragen aan tumorgroei bij. Oncogenen bevatten de code voor eiwitten die tumorgroei kunnen stimuleren, gevaarlijke mutaties in deze genen zorgen ervoor dat de eiwitten minder

gevoelig worden voor onderdrukkende signalen en daardoor ongeremd tumorgroei kunnen stimuleren.

In **hoofdstuk 1**, beschrijf ik de cellulaire opbouw van het borstweefsel en hoe deze verandert tijdens verschillende levensfasen van de vrouw, zoals tijdens de pubertijd en een zwangerschap. In mijn proefschrift, hebben we gebruik gemaakt van verschillende modelsystemen, met nadruk op het modelsysteem de muis. Daarom behandel ik in dit hoofdstuk waar het borstweefsel van de mens en de muis gelijk zijn en waar zij verschillen. De opbouw vertoont veel overeenkomsten. De verschillen zitten voornamelijk in het aantal borstklieren (de mens heeft één paar, de muis heeft vijf paren), het netwerk van melkklieren heeft een ietwat andere opbouw, en de cellen die de borstcellen direct omringen en ondersteunen hebben een andere samenstelling. Maar de ontwikkeling en de belangrijkste soorten cellen zijn hetzelfde tussen de twee zoogdieren, waardoor de muis een goed modelsysteem is voor de humane borst. Verder beschrijf ik verschillende technieken die door het hele proefschrift terugkeren. Zoals een techniek genaamd flowcytometrie, wat het meten van stromende cellen betekent. Hierbij gebruiken we eiwitten die gedeeltelijk buiten een cel uitsteken om celtypes te identificeren: een melk-producerende cel heeft een ander profiel dan een spiercel. Een andere techniek die veelvuldig terugkeert is het analyseren van het transcriptoom. Het transcriptoom omvat alle RNA-moleculen in een cel. Dit geeft een beeld van welke eiwitten in een cel aanwezig zijn en in welke hoeveelheid: als er relatief veel RNA-moleculen aanwezig zijn die coderen voor eiwit A en weinig die coderen voor eiwit B, is het aannemelijk dat er meer van eiwit A dan eiwit B aanwezig is in de cel. Wetenschappers hebben betrouwbare en relatief eenvoudige technieken ontwikkeld om het transcriptoom te analyseren, voor eiwitten ligt dit ingewikkelder. Daarom gebruiken we vaak het transcriptoom als indicator voor de eiwitten die aanwezig zijn.

In **hoofdstuk 2**, kijken we naar het effect van de tumorsuppressoreiwitten p53, p16, en p19 op celdeling in de borst. Gezonde tumorsuppressorgenen en oncogenen zijn onmisbaar voor het goed functioneren van een gezonde cel. Zo spelen de tumorsuppressoreiwitten p53, p16, en p19 (dit eiwit heet p14 in de mens) een belangrijke rol in de regulatie van celdeling. Eerder onderzoek heeft laten zien dat in de afwezigheid van deze genen progenitorcellen in het bloed van muizen zich als stamcellen gaan gedragen: ze kunnen veel vaker delen dan dat normaliter mogelijk is voor een progenitorcel. Dit effect is alleen waarneembaar als alle drie de eiwitten tegelijkertijd afwezig zijn. Dus samen remmen p53, p16, en p19 de celdeling in bloedprogenitorcellen. In **hoofdstuk 2**, hebben we onderzocht of deze drie eiwitten even belangrijk zijn voor het controleren van de celdeling in het borstweefsel. Hiervoor hebben we gebruik gemaakt van muismodellen. Onze bevindingen laten zien dat het uitschakelen van p53, p16, en p19 niet genoeg is om de deling van borstweefselcellen te beïnvloeden, dit geldt zowel voor stam- als progenitorcellen. Wel komen we erachter dat zonder deze drie genen cellen gevoeliger zijn voor een groeistimulus die wordt geïnitieerd $TNF\alpha$. $TNF\alpha$ is een eiwit dat we vooral kennen als een onderdeel van het afweersysteem, maar daarnaast kan het ook cellen stimuleren om af te sterven of voort te bestaan. In dit geval stimuleert het dus het voortbestaan van de borstweefselcellen. Kortom, p53, p16, en p19 alleen zijn niet afdoende om borstweefselceldeling te remmen, maar de drie tumorsuppressoreiwitten remmen

wel door TNF α -gestimuleerde celdeling.

In **hoofdstuk 3** beschrijf ik een nieuw borstkankermuismodel. Het bijzondere aan deze muis is dat het model staat voor patiënten die een terugval ervaren. Dit houdt in dat een medicijn in eerste instantie aanslaat, maar dat de tumor zich na verloop van tijd weet aan te passen en op volle kracht terugkeert. In andere woorden, de tumor is ongevoelig geworden voor het medicijn en heeft een nieuwe manier gevonden om weer te kunnen gaan groeien. Voor dit onderzoek heb ik ook de muis gebruikt die het tumorsuppressoreiwit p53 mist, net zoals in **hoofdstuk 2**. Daarnaast is een gemuteerde versie van het oncogen *Kras* aanwezig. Het eiwit *Kras* wordt in gezonde cellen geactiveerd als een cel moet groeien en gaat delen. Wanneer het gemuteerd raakt, kan het overactief worden en daardoor celgroei en -deling stimuleren die ongewenst is, een belangrijk kenmerk van een tumor. In het borstkankermuismodel uit **hoofdstuk 3**, kunnen we de gemuteerde *Kras* weer uit de cel verwijderen, waarop de borsttumor snel kleiner wordt. Maar na een periode afwezig te zijn geweest, begint dezelfde tumor weer te groeien. Ondanks dat de tumor in eerste instantie afhankelijk was van het gemuteerde *Kras*, heeft het zich nu aangepast aan het gemis van het oncogene eiwit en groeit het weer verder. Door de twee tumoren met elkaar te vergelijken - de tumor die voor groei afhankelijk is van het gemuteerde *Kras* en de tumor die niet meer afhankelijk van dit eiwit is - vinden we dat de tumoren van elkaar verschillen. De onafhankelijke tumor heeft eigenschappen die in patiënten vaak kenmerkend zijn voor een slechte prognose. Ook vinden we aanwijzingen dat eiwitten die normaal gesproken *Kras* activeren, in grotere hoeveelheden aanwezig zijn. Dit kan erop duiden dat de tumorgroei die in eerste instantie door de overactieve oncogene *Kras* werd gevoed, nu door partners van *Kras* wordt gevoed. Onze bevindingen laten zien dat als we patiënten willen genezen of langdurig willen helpen, we behandelingen moeten ontwikkelen die voorkomt dat tumoren zich aan medicatie kan ontwringen.

Zoals we in **hoofdstuk 2** zagen dat het afweersysteem eiwit TNF α ook een rol in celgroei speelt, komen we in **hoofdstuk 4** nog een belangrijke afweersysteemspeler tegen die ook een rol in celgroei speelt, genaamd TLR2. TLR2 is onderdeel van een arsenaal van afweercellen, waar het op zoek is naar ziekteverwerkers zodat de afweercellen ze snel onschadelijk kunnen maken. Waar TNF α voornamelijk actief is als product van andere cellen die de borstweefselcellen willen beïnvloeden, is TLR2 onderdeel van de borstweefselcellen. In **hoofdstuk 4** delen we onze bevindingen dat het activeren van TLR2 - en andere eiwitten die in direct contact staan met TLR2 - in borstweefselcellen en darmcellen tot verhoogde celgroei leidt. Daarnaast, leidt het remmen van deze eiwitten tot het remmen van tumorgroei in zowel borstkanker als darmkanker. Dit onderzoek laat zien hoe mechanismen waarvan we dachten dat ze alleen van belang waren voor de werking van het afweersysteem, ook worden gebruikt door gezonde cellen en tumorcellen. Verder laten we zien hoe deze kennis gebruikt kan worden om kanker te behandelen.

Eerder in deze samenvatting noemde ik al het transcriptoom: de verzameling van alle RNA-moleculen in bijvoorbeeld een cel of een weefsel. Om een beter beeld te krijgen van alle verschillende soorten cellen in het lichaam, hebben we met een groep wetenschappers de transcriptomen van 350.000 verschillende cellen geanalyseerd.

De totstandkoming van deze kennisbank wordt gedeeld in de **hoofdstukken 5 en 6**. We hebben hiervoor het modeldier de muis gebruikt. De cellen kwamen van muizen van verschillende leeftijden, zowel mannetjes als vrouwtjes muizen waren onderdeel van dit project. Zodat verschillen die veroorzaakt worden door geslacht of leeftijd kunnen worden onderzocht. Informatie specifiek voor het borstweefsel die uit dit onderzoek naar voren is gekomen staat in **hoofdstuk 1**. Naast het borstweefsel, heeft deze groep onderzoekers nog naar 22 andere weefsels gekeken. Een project zoals hier staat beschreven vereist de betrokkenheid van een grote groep wetenschappers met verschillende achtergronden en is duur om uit te voeren. Maar het is erg waardevol voor de wetenschappelijke gemeenschap, het verschaft immers een uitgebreide kennisatlas van het (muizen)lichaam die door iedereen te raadplegen is.

Curriculum Vitae

Linda Jacqueline van Weele was born in 1986. She attended secondary school at the Stedelijk Gymnasium Arnhem. After a gap year in Argentina, she completed a bachelor's degree majoring in Biomedical Sciences and a minor in Governance at Utrecht University. She attended one semester at Lund University, Sweden. Part of this semester was spent doing research under the supervision of Dr. Ellen Tufvesson, in the laboratory of Prof. Dr. Leif Björner, confirming the reliable detection of asthma biomarkers in non-invasive exhaled breath condensate. Furthermore, one year she partook in the Utrecht University council, representing the university's student population to the university board. Once she completed her bachelor's degree, she continued her studies at Utrecht University enrolling in the master's program *Cancer, Genomics and Developmental Biology*. She completed research projects in three different research groups. For the first project, she worked with a zebrafish model to identify developmental defects in the congenital disease dyskeratosis congenita under the supervision of Dr. Alyson MacInnes at the Hubrecht Institute, Utrecht, the Netherlands. Next, she looked at the role of calcium-dependent proteases - named Calpains - in tumor growth. This project was under the supervision of Dr. Ferenc Scheeren in the laboratory of Prof. Dr. Michael Clarke at Stanford University, United States of America. Lastly, she optimized a SPARC-dependent oncolytic adenovirus under the supervision of Dr. Eduardo Piaggio, in the laboratory of Prof. Dr. Osvaldo Podhajcer, at Instituto Leloir, Buenos Aires, Argentina. She graduated from her master's program in the spring of 2012. In the summer of 2012, she started her PhD in the laboratory of Prof. Dr. Michael Clarke at Stanford University, United States of America, under the supervision of Prof. Dr. Jacco van Rheenen, which resulted in the work described in this dissertation.

List of publications

Published

van Weele LJ et al. (2022) Mesenchymal tumor cells drive escape of relapsed *Trp53*^{-/-} breast tumor cells with inactivated mutant *Kras*. *Molecular Oncology*, doi:10.1002/1878-0261.13220 - in press.

Tabula Muris Consortium [including van Weele LJ] (2022) Molecular hallmarks of heterochronic parabiosis at single cell resolution. *Nature*, 603(7900):309-314.

van Weele LJ et al. (2021) Depletion of *Trp53*, *p16^{Ink4a}*, and *p19^{Arf}* does not promote self-renewal in the mammary gland but amplifies proliferation induced by TNF α . *Stem Cell Reports*, 16(2):228-236.

Tabula Muris Consortium [including van Weele LJ] (2020) Ageing hallmarks exhibit organ-specific temporal signatures. *Nature*, 583(7817):596-602.

Tabula Muris Consortium [including van Weele LJ] (2020) A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, 583(7817):590-595.

Tabula Muris Consortium [including van Weele LJ] (2018) Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, 562(7727):367-372.

Cai S, Kalisky T, Sahoo D, Dalerba P, Feng W, Lin Y, Qian D, Kong A, Yu J, Wang F, Chen EY, Scheeren FA, Kuo AH, Sikandar SS, Hisamori S, van Weele LJ, et al. (2017) A Quiescent Bcl11b High Stem Cell Population Is Required for Maintenance of the Mammary Gland. *Cell Stem Cell*, 20(2):247-260.e5.

Zabala M, Lobo NA, Qian D, van Weele LJ, Heiser D, Clarke MF (2016) Overview: Cancer Stem Cell Self-Renewal. In Liu, H. and Lathia, J.D. (ed.) *Cancer Stem Cells - Targeting the Roots of Cancer, Seeds of Metastasis, and Sources of Therapy Resistance*. Academic Press, pp. 25-58.

Scheeren FA*, Kuo AH*, van Weele LJ, et al. (2014) A cell-intrinsic role for TLR2-MYD88 in intestinal and breast epithelia and oncogenesis. *Nature Cell Biology*, 16:1238-48.

*Co-first authors

Published prior to start of PhD program

Pereboom TC*, van Weele LJ*, Bondt A, and MacInnes AW (2011) A zebrafish model of dyskeratosis congenita reveals hematopoietic stem cell formation failure resulting from ribosomal protein-mediated p53 stabilization. *Blood*, 118: 5458-65. *Co-first authors

Tufvesson E, van Weele LJ, Ekedahl H, and Bjermer L. (2010) Levels of cysteinyl-leukotrienes in exhaled breath condensate are not due to saliva contamination. *The Clinical Respiratory Journal*, 4: 83-8.

Dankwoord/Acknowledgements

For everyone that was involved with my personal and professional life but somehow didn't pick up much of my native tongue (it must have been my poor teaching skills): this is the part people read first when receiving a thesis. Because the time spend in graduate school is so much more than what ended up in this thesis. I learned from you, I laughed with you, you helped me grow and become the person I am today. For that, I'll always be grateful.

Mike, at the end of my rotation you offered me to stay, being supportive of whatever format I had in mind. This also applies to my research; you've offered me lots of freedom to work on my ideas but were there to help me work through my thinking process when needed. I appreciate your efficient way of communicating. You cut to the chase and certainly do not like emails the length of this paragraph. I also want to acknowledge **Christine**, you welcome us with open arms in your house and set an incredible high standard for paella.

Jacco, bedankt dat je het vertrouwen had dat ik dit project van ver met succes zou afronden. Als we elkaar een hele tijd niet gesproken hadden, had ik binnen een dag alweer een mailtje terug. Ook ben jij degene geweest die mij aan Ferenc heeft voorgesteld, zonder die stageplek was ik niet in Mike's lab terecht gekomen.

Puck en **Niels**, leuk dat jullie in mijn promotiecommissie wilden zitten voor deze minder gebruikelijke constructie. Bedankt voor de input die jullie geleverd hebben tijdens onze bijeenkomsten. I would also like to thank **prof. Geert Kops**, **prof. Patrick Derksen**, **prof. Paul Coffe**, and **prof. Inne Borel Rinkes** for taking the time to assess my thesis.

Ferenc, ook nadat je weer 9 tijdzones verderop woonde, ben je nog steeds mijn trouwe mentor gebleven. Je passie voor de wetenschap en een continue stroom aan ideeën, werken aanstekelijk en hebben mij meerdere malen geholpen als ik weer eens op een hobbelig paadje terecht was gekomen. Jij en **Anja** gaven mij zo een warm welkom. Binnen een week had ik een huis, vrienden, een fiets, en kende ik iedereen in SIM1.

Oh, my beloved Clarke lab. Let's start with Bay 2. **Shang**, it was such a luxury to have your kindness and incredible scientific mind just a chair turn away. Speaking from experience, the big group of people that has joined your lab picked an excellent mentor. **Diane**, your drive for team spirit not only got the entire lab healthier, but also (more importantly) made sure that the Turtle Trophy got firmly placed in our bay. **Jane**, you almost decided to move to the next bay. But luckily you decided just in time, there was too much (purely scientific...) fun in our bay to move away. **Zhen**, you're the lab library system, you know if a reagent is present, and you know exactly where it is. **Liz**, so happy you chose to join and were there for the ups and also the numerous downs of graduate school. So much compassion; that time I cried you couldn't help but join me. You graduated a while ago, but it feels like we are ending this journey together. **Mai**, cuando pienso en vos oigo tu risa. Además, pienso a mirar debajo de

nuestras mesas para confirmar que estabas para que pudiéramos hablar (o gritar, los paneles bloquearon bastante de volumen!). **Ange**, how am I ever supposed to get fit again without sharing the suffering with you? Or keep my second place as brightest dressed? **Shaheen**, I was always excited to share new data with you because I knew you would be excited too and share a new perspective. You are a very caring mentor, everyone in your lab is very lucky. **Robert**, we had so many good conversations while both working until late into the evening. Though never as late as **Taichi**, friendly and helpful at all hours of the day. Also, you had the biggest pen collection I have ever seen. **Dalong**, I'll never forget how you were assisting me doing surgeries when all of a sudden, I got really ill. You kindly directly send me home, finishing up what I couldn't finish. The lab was really Mike and you, staying on top of all the nuts and bolts so we could just focus on doing experiments. **Taylor**, you only joined us for a year, but you were so easygoing, I knew I wanted to continue our friendship. **Peggy**, even though we didn't get to see each other in person anymore once the pandemic hit, we exchanged plenty of emails to keep each other updated. **Will**, always a helping hand and excited to learn and contribute to everyone's research. **Sabra**, thanks for all your work for the Kras paper, I wouldn't have been able to finish that project without you.

All I had to do was to open the fridge, someone that stocks up on green juice had to be Californian. **Lisa**, you introduced me to the state's natural wonders - including the Mystery Spot - and its people, which were so interesting I wanted to come back. Maybe I'm not staying because you didn't? There's not one friend I've shared more lunches with than you, **Luhua**. It's so easy to talk with you. I think that's how our friendship started, with a serendipitous run-in that turned into an hour-long conversation. **Zoe**, knowing you were going to be in SF when I moved up there, made the longer commute worth it. It has strengthened our friendship to live (a 15 min drive) around the corner. The same move, that ended our weekly runs, **Lauren**. Catching up while observing the life cycle of Shoreline's geese was always a highlight of my week. I still keep my hopes up we'll be able to reinstate our running tradition one day. **Diane**, you have such a great can-do attitude, breaking glass ceilings left and right. I'll be eternally grateful for your gift of bringing people together. Without your Sonoma trip invitation, my life would have been completely different. **Méryl** and **Dawn**, I got so lucky getting to live with you. Seasoned campers in Pinnacles, a dark & locked hotel late at night in Tahoe, the letter S. **Nick** smoothly joined our shenanigans. **Méryl**, you've been with me, day after day, through most of this journey, I'm so glad you're able to attend and be my paranimf. **Anna**, that tree looks pretty miserable but hey, it has two oranges in it, so you'll surely risk your life, just climb it, then offer the oranges to me. **Rishi**, fresh of the airplane Anja and Ferenc took me to your house, showing me postdocs can be fun people too (and post-postdocs too!). You always make sure everyone is warm, fed & drinking wine. **Amy**, I learned later that was your home too. You always have something kind to say, after seeing you I always feel uplifted. **Aparna**, you show me what a great career in biotech looks like. While eating delicious Indian or Thai food or providing the best view of 4th of July fireworks. **Bojk** en **Jolien**, heerlijk was het om in de zon, onder de palmbomen in onze moedertaal te kunnen praten. **Robert**, my concert buddy. Half asleep, with a major jetlag, you still made it out, days before the pandemic shut down our future music plans. **Snert** en **Rolo**, nu ken ik Californian heimwee ook. Motelkamers met

dino's die niet vijf maal zo groot worden. Gele slakken en enorme bomen. Bijzondere plaatsnamen. Iets wat ongebruikelijke Sinterklaascadeaus.

Marijn, Vera, Rosanna, Hilde, Hedwig, Juliette. Derde kerstdag staat vast rood omrand. **Marijn, Rosanna en Hilde**, jullie zijn zelfs langsgesproken en hebben wat van mijn leven in Californië gezien. **Marijn en Rosanna**, jullie zijn mijn trouwste reizigers, is het al ter sprake gekomen dat ik inmiddels weer ergens anders woon? **Vera**, gesprekken met jou nemen altijd een onverwachte wending. **Selma**, we waren maar een jaar samen in Utrecht maar toch voelt het alsof we onze gehele studietijd op dezelfde plek hebben doorgebracht. Luhua en jij hebben een keer geskypet zodat je mijn Amerikaanse leventje wat beter zou leren kennen, alsof je daar zodoende ook een beetje onderdeel van werd. **Marieke**, elke keer als ik terug in Nederland was wilde je het mij zo makkelijk mogelijk maken om af te spreken zodat het soms zelfs lukte om elkaar een paar keer te zien. Altijd weer meteen als vanouds. **Laurien, Eline, Sophie, Annemarijn, Lotte.** Inmiddels woont nog maar een van ons in Utrecht, maar het is wel de plek waar ik aan deze reis begonnen ben. **Laurien**, leuk om jullie over campus rond te leiden voordat jullie in het kleinste autootje dat op de luchthaven te vinden was in zuidelijke richting vertrokken. Bijzonder dat onze vriendschap hechter is geworden de afgelopen jaren. **Rik**, dankzij jou, heb ik altijd mijn best gedaan om vrienden te worden met mijn huisgenoten, jij hebt mij laten zien dat dat het samenwonen zoveel leuker maakt. **Nicoline**, quien pensaba que nos íbamos ver en tantos distintos países? Apoyarnos unas a otras desde la distancia.

Alyson, I learned so much during my time in your lab. I would not have taken the step to getting a PhD without that positive experience. You populated your lab with leukerds **Tamara en Paul**. Van jullie heb ik geleerd hoe het lab leven eruitziet. Wetenschap bedrijven, dat is een teamsport. Pipetteren doe je met een goed (fout) muzikje op de achtergrond. Wetenschappers, dat zijn leuke mensen met wie je het liefst je vrijdagavond ook nog eens doorbrengt. Leuke mensen zoals **Sjoerd, Laila en Joram**.

Els, ooms & tantes, neefjes & nichtjes, nadat jullie met een grote groep ook bij de uitreiking van mijn masterdiploma aanwezig waren, mogen we eindelijk vieren dat ik ook dit diploma mag behalen. Erg bijzonder dat we zoveel jaren later, nog grotendeels allemaal samen zijn.

Susan, George, Leah and Rob. You're such a loving family, I felt welcome immediately. **Susan and George**, it means a lot you'll be joining me to close off this chapter properly.

Papa en mama, dankzij jullie steun en vertrouwen heb ik dit avontuur aangedurfd en blijf ik nieuwe avonturen aangaan. Ook ben ik dankbaar voor jullie talloze bezoeken, er zijn maar weinig mensen die zoveel van westelijk Amerika hebben gezien als jullie. Tijd om de andere kant van dit land beter te leren kennen! Mam, doordat je meerdere keren op Asher bent komen passen kon ik mijn laatste artikel en proefschrift afronden. Anders zouden we hier nog niet zijn beland. **Esther en Ruben**, bedankt voor de fijne en waardevolle gesprekken. **Es**, je hebt al mijn verhalen over mijn pieken en dalen geduldig aangehoord, begrip tonend dat sommige dingen nou eenmaal anders

lopen dan gepland.

Daniel, I have moved my graduation timeline up at least a dozen times. This wasn't really what you expected, but you stayed supportive and let me figure it out (mostly) on my own timeline. Your boundless energy and push to move forward are inspiring. Life with you is never dull. **Asher**, het is een grote eer om jou van dichtbij te zien groeien en ontwikkelen. Soms benader ik je als een experiment, ik kijk naar jouw data en probeer te voorspellen hoe ik in de toekomst een goede uitslag kan herhalen. Vooral nog, met weinig succes. Ook met jou, zit het leven vol verrassingen.

