



**Universiteit
Leiden**
The Netherlands

An engineering approach to decode immune responses

Bresser, K.

Citation

Bresser, K. (2023, November 15). *An engineering approach to decode immune responses*. Retrieved from <https://hdl.handle.net/1887/3663147>

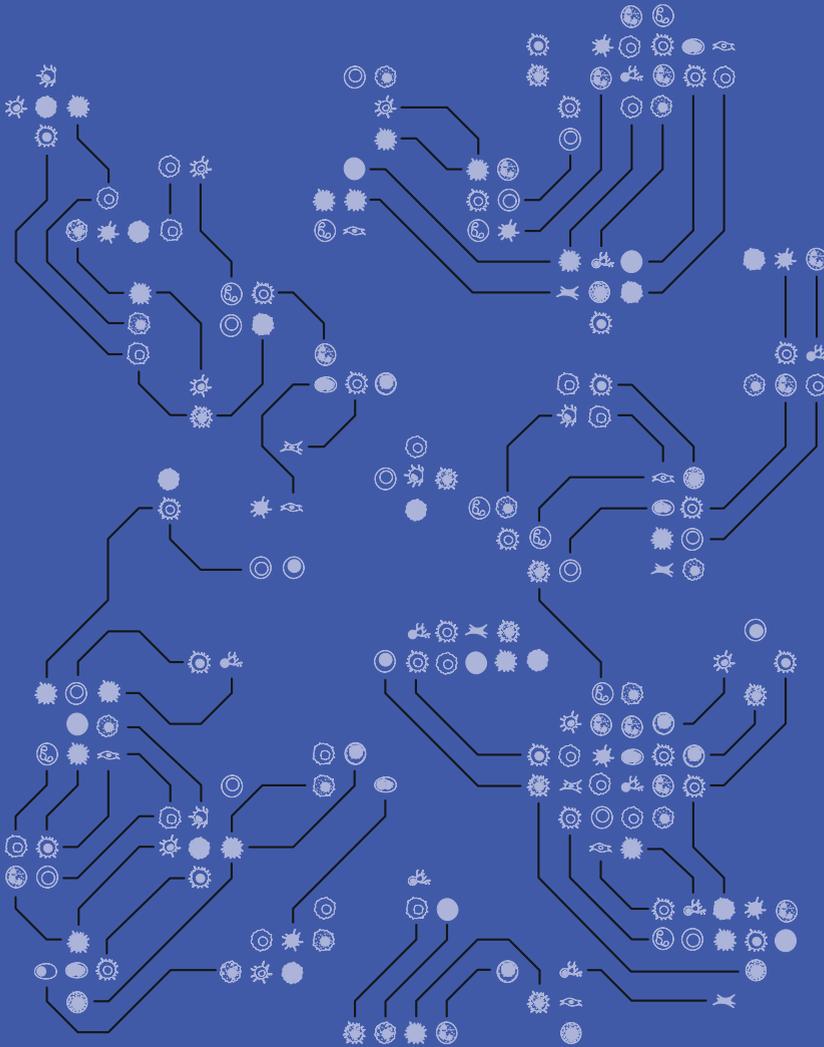
Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3663147>

Note: To cite this publication please use the final published version (if applicable).

An engineering approach to decode immune responses



Kaspar Bresser

An engineering approach to decode immune responses

Kaspar Bresser

Cover design: Bas Stoker

Printed by: Ridderprint | www.ridderprint.nl

ISBN: 978-94-6483-493-2

The research described in this thesis was performed at the division of Molecular Oncology and Immunology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital (NKI-AVL). This work was financially supported by the Oncode Institute.

Copyright © 2023 Kaspar Bresser. All rights reserved.

No part of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without prior permission of the author and the publisher holding the copyright of the articles.

Pictograms of cells were adapted from BioRender images.

Printing of this thesis was financially supported by the NKI-AVL and Leiden University

An engineering approach to decode immune responses

Proefschrift

ter verkrijging van
de graad van doctor aan Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 15 November 2023
klokke 16:15 uur

door

Kaspar Bresser

Promotor

Prof. Dr. A. N. M. Schumacher

Co-promoter

Dr. F. A. Scheeren

Leden promotiecommissie

Prof. Dr. J. J. C. Neefjes

Prof. Dr. S. I. van Kasteren

Prof. Dr. J Keyser-Borst

Prof. Dr. D. Amsen

University of Amsterdam, Amsterdam

Dr. M. C. Wolkers

University of Amsterdam, Amsterdam

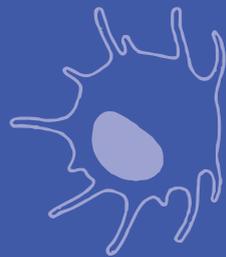
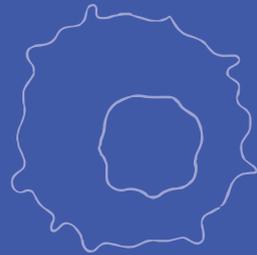
Table of contents

Chapter 1.	Scope of the thesis	6
Chapter 2.	A mouse model that is immunologically tolerant to reporter and modifier proteins	12
Chapter 3.	Gene and protein sequence features augment HLA class I ligand predictions	28
Chapter 4.	Replicative history marks transcriptional and functional disparity in the CD8+ T cell memory pool	50
Chapter 5.	QPCTL regulates macrophage and monocyte abundance and inflammatory signatures in the tumor microenvironment	100
Chapter 6.	Discussion	134
Chapter 7.	Appendices	148

Chapter 1.

Scope of the thesis

OR: "How I sciented the science."



Throughout this thesis, I have endeavored to apply an engineer's mindset in my pursuit to better understand the marvelously convoluted immune system. In doing so, my colleagues and I have generated a number of new 'hardware' (i.e., genetically engineered) tools and 'software' modules (i.e., custom analyses and models) that enable the investigation of several otherwise difficult-to-study concepts. Although we have used these modules here to study immune responses, I hope they may be utilized as tools and approaches to crack outstanding questions in other fields of research. As the work described in this thesis focuses on various aspects of the immune response, I will first briefly touch upon the general organization of the immune system, and subsequently elaborate on the topics relevant for each individual chapter.

The immune system may be viewed as a sophisticated apparatus that is tasked with the identification and eradication of foreign entities. This process occurs through meticulous collaboration between the innate and adaptive branches of the immune system. The innate branch consists of a large number of different cell types (e.g., macrophages, granulocytes and natural killer cells) that collectively recognize a wide variety of common pathogen- and danger-associated molecules. Innate immune cells are generally able to respond quickly upon pathogen encounter, providing a first layer of protection to a nascent infection. In addition, a specialized subset of innate cells, referred to as antigen-presenting cells, are able to leave the site of infection and travel to lymphoid tissues where they can trigger the second branch of the immune system. This adaptive branch comprises T and B cells that are able to recognize foreign antigens in a highly specific manner. During their development, each newly minted T or B cell is endowed with a unique antigen receptor, which determines its antigen specificity. Although the approximate diversity of these antigen receptors present in any given individual is still unresolved, it is likely to be in the order of billions. As a result of this immense diversity, the collective repertoire of T and B cells is able to recognize and respond to any pathogen that is encountered during a lifetime. In addition, due to this massive antigen receptor diversity, T and B cells of a given antigen specificity initially exist in low frequencies. When such 'naïve' lymphocytes recognize their cognate antigen, they transition to an activated 'effector' state and progressively differentiate into various distinct functional subsets that can combat the pathogen. When the pathogen has been subdued, the adaptive immune response enters its final stage, in which the activated lymphocytes will transition to a 'memory' state. Memory lymphocytes are long-lived, potentially persisting for a lifetime, and provide enhanced protection when the same pathogen is encountered later in life. Collectively, through the combined action of all these cell types, the immune system is able to provide tailored responses against many different pathogens.

The recognition and rejection of foreign entities that enter the body is essential to maintain homeostasis. However, in a variety of biomedical studies it is desired to introduce genetically modified cells (e.g., transgenic cells that carry genes encoding fluorescent proteins) into experimental animals such as mice. Due to the exogenous nature of these proteins, such cell transfer experiments are often

plagued by confounding effects caused by the immunogenicity of the transplanted cells. In **chapter 2**, we develop a mouse model in which immune recognition of a large series of reporter proteins is abrogated, thereby providing a solution to this issue. We offer this model (and the methodology through which it was generated) as a tool to the community, and hope it will allow others to perform experiments that would otherwise be impossible.

The antigen presentation machinery continuously samples the intracellular proteome, ultimately leading to the presentation of peptides bound to HLA class I at the cell membrane. These surface-presented peptide-HLA class I complexes are collectively referred to as the HLA class I ligandome, and provides a ‘snapshot’ of the cellular proteome for scrutiny by T cells. Spontaneous alterations to the proteome—as would be the case during viral infections or through genetic mutations accumulated during tumorigenesis—can result in the addition of foreign peptides to the surface-presented HLA class I ligandome. In turn, such peptides can be recognized by antigen-specific T cells, resulting in their activation and subsequent destruction of the aberrant cells. In **chapter 3**, we set out to better understand the process of peptide selection by the antigen presentation machinery. In this effort we demonstrate that genetically encoded sequence features inform on the likelihood of proteins to yield HLA class I ligands. Importantly, these sequence features can be integrated into a classification model, thereby improving the prediction of HLA class I ligands. The improved predictive models that we generate in this chapter may be of value in studies in which the precise identification of an HLA class I ligandome is required, such as the selection of (neo)antigens for cancer immunotherapy.

The T cell pool can be subdivided into 2 major lineages; CD4⁺ and CD8⁺ T cells. Whereas CD4⁺ T cells provide a supportive function during the immune response, CD8⁺ T cells directly seek and destroy infected cells. Upon recognition of their cognate epitope, CD8⁺ T cells enter a phase of rapid clonal expansion, resulting in the generation of a large pool of cytotoxic effector T cells that can combat the infection. A key feature of the CD8⁺ T cell response is the formation of long-lived central memory T cells (T_{CM}) after antigen clearance. This specialized CD8 T cell subset is able maintain itself long-term through homeostatic cell division and possesses a heightened capacity to mount a secondary cytotoxic response upon antigen re-encounter. In **chapter 4**, we study the relationship between the cell state and function of memory CD8⁺ T cells, and the extent of clonal expansion that those cells undergo during an immune response. To this end, we engineer a genetic reporter system that exploits low-probability mutations that occur during cell division to induce the expression of a fluorescent protein, allowing one to ‘record’ the extent of prior proliferation within a cell population of interest. Combining this system with single-cell transcriptomics, we find that the T_{CM} pool is comprised of subsets that have either divided little or extensively, and that this extent of prior division is associated with heightened expression of multipotency- or effector-associated transcripts, respectively. Importantly, we show that the capacity to re-expand into a new wave of cytotoxic cells upon antigen re-encounter is skewed toward memory T cells that had divided little

during the primary response. Our observations in chapter 4 are in support of a model in which during the primary response a sub-group of effector T cells adopts a quiescent phenotype and maintain a less differentiated cell state. These ‘sleeper T cells’ possess superior replicative capacity upon re-infection and therefore represent an important pillar of adaptive immunity.

As I noted above, T cells only constitute a fraction of the total immune response, functioning alongside a large variety of immune and non-immune cells (e.g., fibroblasts and endothelial cells) that collectively determine the progression of the response. The complex interplay between all these cell types is highly apparent in chronic illnesses that the immune system is unable to resolve, as is the case in solid cancers. During tumorigenesis the constantly growing tissue is often infiltrated by immune cells that subsequently engage in a vast network of immune interactions within the tumor microenvironment. In the event that the cancer cells are not eradicated, the tumor microenvironment eventually reaches a state of homeostasis in which immune cells are unable to recognize—or respond to—the foreignness of the tumor. A relatively young field of research, immuno-oncology, has aspired to understand and manipulate these immune interactions in the tumor microenvironment, aiming to tip the scales in favor of the immune system, allowing it to eliminate the mutated cells. In **chapter 5**, we demonstrate that the enzyme glutaminyl-peptide cyclotransferase-like protein (QPCTL) acts as a pleiotropic modifier of the tumor microenvironment. Using syngeneic tumor models, we observe that genetic deletion of QPCTL increases the quantity of tumor-infiltrating macrophages, favors the differentiation of cancer-associated fibroblasts known for inflammatory function, and rewires a TGF- β dominated environment to an IFN γ dominated one. Importantly, we show that the combination of this ablation of QPCTL activity with an immune-activating agent (anti-PD-L1 treatment) can result in delay of tumor growth and enhanced survival in mice. The findings presented in this chapter suggest that abrogation of QPCTL activity can sensitize the tumor microenvironment to immune-activating agents, and provide support for the development of QPCTL inhibitors.

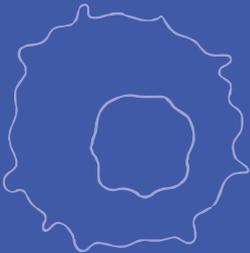
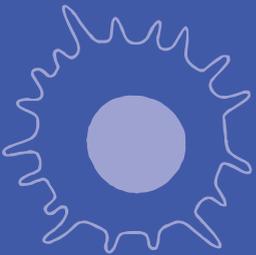
Finally, in **chapter 6**, I will discuss two separate topics that captured my interest while working on this thesis. In the first part of this chapter, I will focus on a concept that has gained a great deal of popularity in the T cell field, the presence of stem cell like behavior in the memory T cell pool. In the second part of this chapter, I will switch gears to a more societal aspect of science, and discuss the manner in which scientific data is recorded and made available to the community.



Chapter 2.

A mouse model that is immunologically tolerant to reporter and modifier proteins

OR: "How we tricked a mouse into believing it's a jellyfish."



Kaspar Bresser^{1,*}, Feline E Dijkgraaf^{1,*}, Colin EJ Pritchard², Ivo J Huijbers², Ji-Ying Song³, Jan C Rohr⁴, Ferenc A Scheeren⁵ and Ton N Schumacher^{1,6}

Published in Communications Biology, May 2020

(1) Division of Molecular Oncology & Immunology, Onco Institute, The Netherlands Cancer Institute, Amsterdam, The Netherlands

(2) Mouse Clinic for Cancer and Aging research (MCCA) Transgenic Facility, The Netherlands Cancer Institute, Amsterdam, The Netherlands

(3) Animal Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

(4) Center for Chronic Immunodeficiency, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany; Center for Pediatrics and Adolescent Medicine, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany.

(5) Department of Medical Oncology, Leiden University Medical Center, Leiden, The Netherlands.

(6) Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands.

(* These authors contributed equally to this work

Abstract

Reporter proteins have become an indispensable tool in biomedical research. However, exogenous introduction of these reporters into mice poses a risk of rejection by the immune system. Here we describe the generation, validation and application of a multiple reporter protein tolerant ‘Tol’ mouse model that constitutively expresses an assembly of shuffled reporter proteins from a single open reading frame. We demonstrate that expression of the *Tol* transgene results in the deletion of CD8⁺ T cells specific for a model epitope, and substantially improves engraftment of reporter-gene transduced T cells. The Tol strain provides a valuable mouse model for cell transfer and viral-mediated gene transfer studies, and serves as a methodological example for the generation of poly-tolerant mouse strains.

Introduction

Following the advent of standardized genetic editing techniques, researchers have isolated a large collection of reporter and modifier proteins (RPs and MPs, respectively) from a variety of species that have since been instrumental to characterize a wide range of biological processes¹. For example, RPs are frequently used to tag endogenous proteins or to track the behavior of individual cells *in vivo*^{2,3}. In addition, MPs (e.g. Cre recombinase and Cas9) are commonly applied to influence *in vivo* cell behavior through the induction of defined genetic alterations⁴. However, as RPs and MPs are almost invariably derived from non-mammalian species, exogenous introduction of these proteins into immunocompetent mice poses the risk of immunological rejection by host-derived T cells. In line with this, multiple cases of immunological rejection of cells expressing RPs, such as firefly luciferase and eGFP, have been reported⁵⁻⁹. Moreover, major histocompatibility complex (MHC) class I-restricted epitopes of luciferase and GFP have been identified, underlining their capacity to induce CD8⁺ T cell responses⁵⁻⁸. Even in the absence of complete immunological rejection of RP- or MP-modified cells, experimental outcomes may potentially be subtly biased through the action of such undesirable immune responses.

To overcome this problem, we set out to engineer a transgene that—once introduced into the genome of a mouse model of choice—prevents the generation of immune responses against RP- or MP-modified cells through the physiological self-tolerance mechanisms. Specifically, CD4⁺ and CD8⁺ T cells that carry a self-reactive T cell receptor (TCR) may be inactivated or deleted through multiple mechanisms. First, during their maturation in the thymus, T cells that carry a self-reactive TCR are deleted through negative selection. Second, auto-reactive T cells that escape deletion through central tolerance are kept in check by a mechanism referred to as peripheral tolerance that requires antigen encounter outside of the primary lymphoid organs⁹. In addition, self-reactive CD4⁺ T cells may develop into regulatory T cells, and play a crucial role in such peripheral tolerance¹⁰. Thus, in order to generate a poly-tolerant mouse model, constitutive expression of foreign antigens in both the thymus and throughout peripheral tissues would be preferred. However, organism-wide

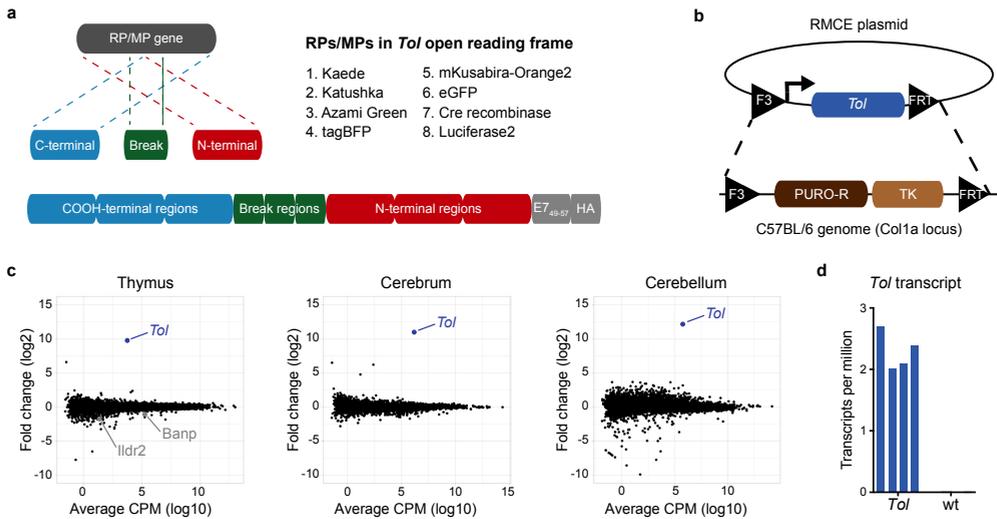


Fig. 1. Generation and validation of the *Tol* mouse model. **a**, Top left: cartoon depicting the strategy for reporter gene fragmentation, applied for each RP/MP gene included in the *Tol* ORF. Top right: reporter proteins included in the *Tol* ORF. Bottom: schematic overview of gene fragment placement in the *Tol* ORF. Note that the *Tol* ORF encodes the HPV E7₄₉₋₅₇ epitope in the COOH-terminal region of the chimeric *Tol* protein. **b**, Cartoon depicting the targeting strategy of the *Tol* ORF into the *Col1a1* locus through recombinase-mediated cassette exchange. **c**, Differential gene expression analysis of indicated organs of *Tol* mice relative to WT mice. Scatterplots indicate average log₂ fold changes and average counts per million. Genes with a statistically significant change in expression in *Tol* mice are indicated in grey, and the *Tol* transcript is indicated in blue. **d**, Expression levels of the *Tol* transcript in thymi of *Tol* and WT mice. Data shown (**c-d**) is aggregated from n=4 (*Tol*) and n=4 (WT) mice. RP = reporter protein, MP = modifier protein, RMCE = recombinase-mediated cassette exchange, PURO-R = puromycin resistance gene, TK = thymidine kinase gene, CPM = counts per million.

expression of proteins such as eGFP or Cre is incompatible with their intended use as cell-type specific reporters or modifiers.

Results

Generation and validation of the ‘*Tol*’ mouse model

In order to engineer a multiple reporter protein tolerant mouse model, we designed a large chimeric open reading frame (ORF) that encodes 6 fluorescent proteins, firefly luciferase, and Cre-recombinase, all in a scrambled format. Specifically, in order to prevent functional expression of the introduced proteins, each individual gene was split into two fragments, and the resulting set of 16 gene fragments was subsequently assembled in a scrambled order. In addition, to ensure tolerance toward potential T cell epitopes present at the gene breakpoints, the 60bp region surrounding each split site was added (**Fig. 1a**). Finally, to be able to test whether tolerance was induced against epitopes throughout the artificial protein, a control CD8⁺ T cell epitope derived from the HPV E7 protein (HPV E7₄₉₋₅₇) was placed at the COOH-terminus. This resulted in a 7,359 bp chimeric ‘*Tol*’ ORF, encoding a protein of ± 275 kDa that covers seven RP and one MP (**Supplementary Fig. 1**).

Recognition of RP- and MP-derived epitopes is dependent on the host MHC haplotype^{11,12}, and risk of immunological rejection of cells modified with individual RPs and MPs thus varies between mouse strains. To test if the *Tol* cassette could be applied to induce tolerance against proteins of

interest, we set out to assess its functionality in the C57BL/6 strain, a widely used mouse strain in immunological research. In addition, MHC binding predictions indicated that the Tol protein contains 78 potential MHC ligands predicted to bind with high affinity (i.e. netMHC4.0 percentile rank <1%) to the C57BL/6 MHC haplotypes (**Supplementary Data 1**). To establish the transgenic C57BL/6 strain, the *Tol* gene was targeted to the *Col1a1* locus of embryonic stem cells (ESCs) via recombinase mediated cassette exchange (**Fig. 1b**)^{13,14}. Successfully modified ESCs were then injected into blastocysts and transferred to pseudopregnant foster mice. The resulting F0 generation was tested for transgene presence (**Supplementary Fig. 2a**), and mice with sufficient chimerism were bred to obtain experimental cohorts of heterozygous ('Tol') and wild-type littermate controls ('WT').

Due to its large size and scrambled design, the Tol protein is expected to be unfolded. Cellular stress induced by accumulation of unfolded proteins has previously been shown to be associated with a variety of pathologies and neurological disorders¹⁵. To test whether expression of the *Tol* transgene would affect murine development, histopathological evaluation of tissues from Tol and WT mice was performed, revealing no evidence for transgene-induced pathology (**Supplementary Table 1**). As a second test for cellular stress, or potential deregulation of other pathways, that may be induced by *Tol* expression, RNA-sequencing was performed on cerebral and cerebellar tissue of Tol and WT animals. In addition, transcriptional activity in thymic tissue, the site of T cell deletion, was compared between Tol-transgenic and WT mice. Hierarchical clustering analysis of the top expressed genes across tissues showed that WT and Tol mice did not typically cluster separately, suggesting that the *Tol* transgene did not induce major transcriptional changes (**Supplementary Fig. 2b**). Subsequent differential gene expression analysis likewise revealed no significant differentially expressed genes, aside from two marginally downregulated genes in the thymus (**Fig. 1c** and **Supplementary Data 2**). Importantly, the *Tol* transcript could be detected in all organs tested (**Fig. 1c, d**), and aligned reads spanned the entire transgene, demonstrating that the full-length ORF was transcribed (**Supplementary Fig. 2c**).

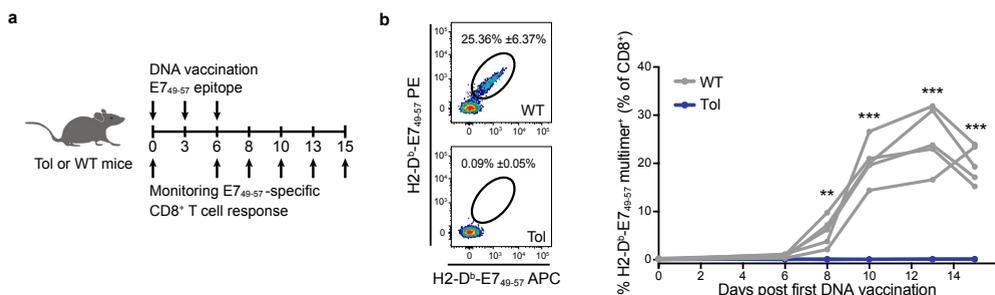


Fig. 2. Complete loss of E7₄₉₋₅₇ immunogenicity in Tol mice. **a**, Schematic overview of vaccination and monitoring strategy. **b**, Left: Representative flow cytometry plots depicting E7₄₉₋₅₇-specific CD8⁺ T cells at day 13 post DNA vaccination in WT and Tol mice. Right: E7₄₉₋₅₇-specific CD8⁺ T cell response in blood of WT (gray) and Tol (blue) mice at the indicated timepoints post first DNA vaccination. Lines indicate individual mice. Cells are gated on IR-dye⁺ CD8⁺ lymphocytes. Depicted data are representative of two independent experiments, including at least 5 mice per group. *P* values were determined by Repeated Measures two-way ANOVA with Sidak's multiple comparisons test to test the significance at each time point. **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

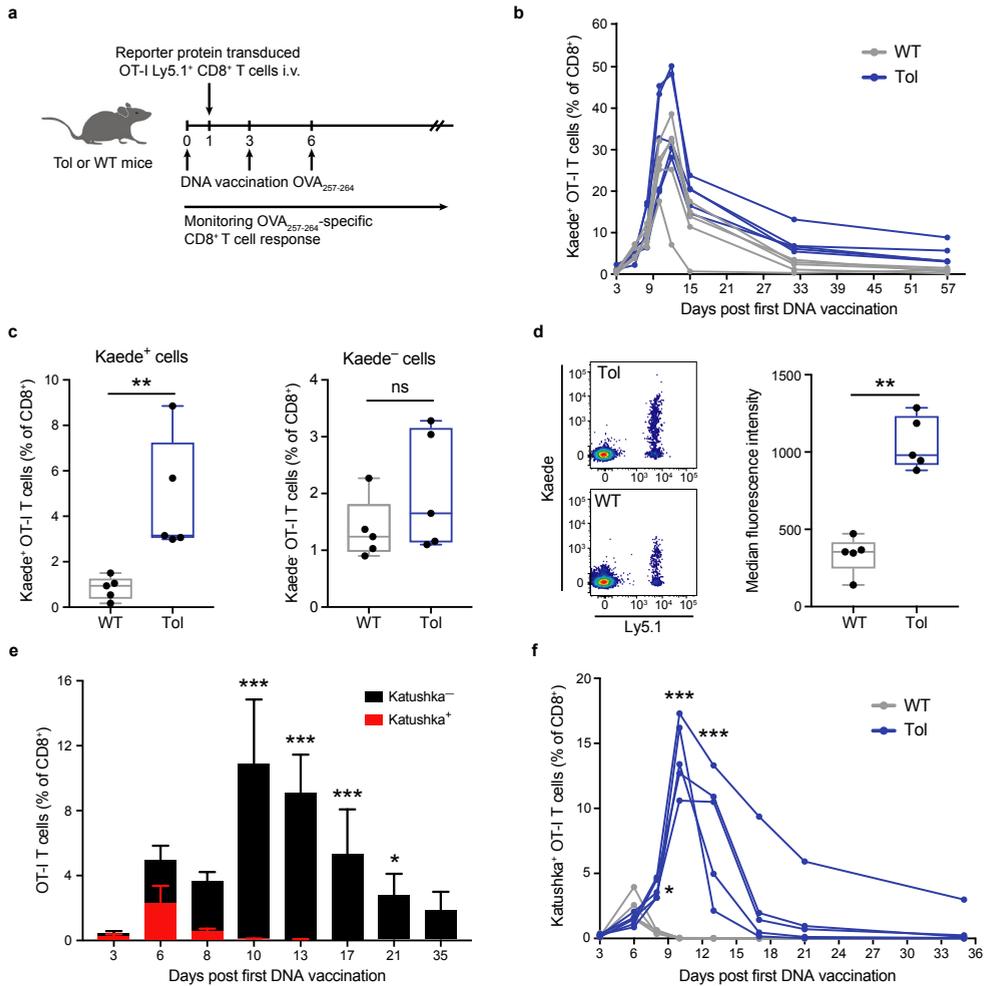


Fig. 3. Enhanced engraftment of Kaede and Katushka expressing cells in Tol mice. **a**, Schematic overview of experimental design. **b**, Percentage of Kaede⁺ OT-I T cells in the blood of WT (gray) and Tol (blue) mice at the indicated timepoints post first DNA vaccination. **c**, Percentage of Kaede⁺ (left) or Kaede⁻ (right) OT-I cells in WT and Tol mice at day 57 after first DNA vaccination. **d**, Left: representative flow cytometry plots depicting Kaede expression (Ly5.1⁺) cells at day 57 after first DNA vaccination. Right: median fluorescence intensity (MFI) of Kaede⁺ OT-I cells in WT and Tol mice at day 57 post first DNA vaccination. **e**, Percentage of Katushka⁺ (red) and Katushka⁻ (black) OT-I cells in WT animals at the indicated timepoints after first DNA vaccination. **f**, Percentage Katushka⁺ OT-I cells in WT and Tol mice at the indicated timepoints after first DNA vaccination. Cells (**b-f**) are gated on IR-dye⁺ CD8⁺ lymphocytes, and plots depict n=5 (Tol) and n=5 (WT) mice. Depicted data are representative of one (**b-d**) or two (**e-f**) independent experiments. Box-plots (**c, d**) depict interquartile range, with whiskers representing minimum and maximum values. Dots (**c, d**) and lines (**b, f**) indicate individual mice. Error bars (**e**) represent SD. *P* values were determined by either two-tailed Mann-Whitney signed rank test (**c, d**) or Repeated Measures two-way ANOVA with Sidak's multiple comparisons test (**e, f**). **P* < 0.05; ***P* < 0.01; ****P* < 0.001, ns = not significant.

Immunological tolerance toward the E7₄₉₋₅₇ epitope in Tol mice

Having observed the presence of the full-length *Tol* transcript in thymic tissue of Tol mice, we next investigated whether T cell tolerance toward epitopes encoded by the transcript was successfully induced. To this purpose, WT and Tol animals were vaccinated with a DNA vector encoding the

HPV E7₄₉₋₅₇ epitope, and the development of CD8⁺ T cell responses against the E7₄₉₋₅₇ epitope was subsequently monitored in peripheral blood by peptide-MHC multimer staining (**Fig. 2a**). Following DNA vaccination, a large population of E7₄₉₋₅₇-specific CD8⁺ T cells was detected in blood of WT mice, with peak E7₄₉₋₅₇-specific CD8⁺ T cell frequencies up to 30% of the total CD8⁺ T cell population (**Fig. 2b, c**). In stark contrast, no E7₄₉₋₅₇-specific CD8⁺ T cells were detected above background in *Tol*-transgene positive animals at any timepoint post vaccination (**Fig. 2b, c**). These data indicate strict tolerance of Tol mice toward a T cell epitope that is present at the 3' of the *Tol* ORF, indicating full-length translation of the protein and successful processing by the antigen presentation machinery.

Enhanced engraftment of fluorescent protein expressing OT-I T cells in Tol animals

To assess whether expression of the shuffled reporter gene fragments in Tol mice resulted in tolerance toward the parental non-shuffled reporter proteins, CD8⁺ Ly5.1⁺ OT-I T cells—specific for the OVA₂₅₇₋₃₆₄ peptide—were transduced with a retrovirus encoding the green-to-red photo-switchable reporter protein Kaede. The resulting Kaede⁺ and Kaede⁻ OT-I T cells were then transferred into Tol and WT mice, and OT-I T cell frequencies were boosted by administration of an OVA₂₅₇₋₃₆₄-epitope encoding DNA vaccine (**Fig. 3a**). In the first days following adoptive cell transfer, the numbers of Kaede⁺ T cells increased rapidly and similarly in both WT and Tol mice (day 10, ratio Tol/WT = 1.25, $p = 0.421$, **Fig. 3b**). Notably, at later time points Kaede⁺ OT-I T cells formed a readily detectable memory T cell population in Tol mice, but declined to near undetectable numbers in WT mice (day 57, ratio Tol/WT = 5.6, $p = 0.008$, **Fig. 3b, c**), suggestive of immunological rejection of the introduced Kaede⁺ cells. As a control, Kaede⁻ T cells were detected in comparable numbers in WT and Tol mice (day 57, ratio Tol/WT = 1.5, $p = 0.309$). In addition, remaining Kaede⁺ T cells in WT mice showed a significantly lower expression of the Kaede protein than Kaede⁺ T cells present in Tol mice ($p = 0.008$, **Fig. 3d**), indicating that immune-mediated rejection can both lead to a reduction in cell quantities and selection of cells with lower transgene expression. To assess whether expression of the *Tol* transgene resulted in long-term tolerance toward Kaede, recipients of Kaede⁺ and Kaede⁻ OT-I T cells were vaccinated and then challenged by a secondary vaccination 155 days post adoptive cell transfer (**Supplementary Fig. 3a**). In Tol mice, Kaede⁺ cells again expanded to large numbers after receiving a secondary stimulus, emphasizing ongoing tolerance toward the transferred cells (**Supplementary Fig. 3b**). In contrast, Kaede⁺ T cells exhibited a stunted proliferative burst, and expressed only low levels of the fluorescent protein in WT animals (**Supplementary Fig. 3c**). The inability to mount a robust secondary response was directly related to Kaede expression, as Kaede⁻ T cells were able to expand in WT mice (**Supplementary Fig. 3d**).

We next examined the engraftment potential of OT-I T cells that expressed a third foreign entity, the rapid-folding red-fluorescent protein Katushka, in Tol and WT recipient mice. Adoptive transfer of Katushka⁺ cells induced a vigorous anti-fluorochrome immune response in WT mice, as reflected by the rapid disappearance of Katushka⁺ OT-I T cells, but not Katushka⁻ T cells (ratio Kat⁻/Kat⁺ cells: day 6 = 1.2; day 10 = 385.4, **Fig. 3e**). Importantly, the observed clearance of Katushka-expressing

cells was abrogated by *Tol* expression, as shown by the presence of an approximately 500-fold larger pool of Katushka⁺ T cells in Tol mice as compared to WT littermates at day 10 post first DNA vaccination ($p = 0.008$, **Fig. 3f**).

To comprehensively evaluate whether cells expressing any of the fluorescent proteins (FPs) encoded by the *Tol* transgene could engraft in Tol mice, we concurrently transferred OT-I T cells that had been transduced with four different FPs (Katushka, BFP, AzamiGreen or mKO2) into either Tol or WT mice. In line with the prior analyses, Katushka⁺ OT-I T cells disappeared from the circulation of WT mice as early as day 10 post-transfer, whereas robust engraftment of Katushka expressing cells was observed in Tol mice (**Supplementary Fig 4**). Analysis of modified T cell frequencies for the other three FPs showed the presence of FP-modified cells in blood of both WT and Tol mice, with a trend toward enhanced engraftment in Tol mice (**Supplementary Fig 4**). Thus, cells modified with 5 different FPs (Kaede, Katushka, BFP, AzamiGreen or mKO2) stably engraft in Tol mice, whereas engraftment of cells expressing at least 2 of these transgenes is severely impaired in WT mice.

Discussion

Manipulation and detection of cellular pathways using modifier and reporter proteins forms a cornerstone of animal research; however, the xenogenic nature of these proteins introduces a substantial risk of recognition by the adaptive immune system. In this work, we describe an approach to fuse multiple RPs/MPs into a large shuffled ORF, thereby perturbing their functionality while maintaining all potentially immunogenic epitopes. Importantly, we show that this approach allows expression in the thymus plus peripheral tissues, that the encoded fusion protein is fully translated, and that this protein translation confers immunological tolerance toward all epitopes tested. Interestingly, comparison of the fate of Katushka⁺ and Kaede⁺ cells in WT mice underlines that immune recognition can either lead to the rapid clearance of transgene expressing cells (i.e. Katushka), or to the gradual selection of cells with lower transgene expression (i.e. Kaede). This second mode of immune rejection may go unnoticed in many experimental settings, thereby representing a hidden confounder. Engraftment of BFP, AzamiGreen and mKO2 transduced OT-I T cells was also observed but was only marginally improved relative to WT mice, suggesting that these proteins have a low immunogenicity in the C57BL/6 strain used in this study. As antigen presentation is MHC restricted, the value of the *Tol* transgene for these FPs may be more profound in other mouse strains, as e.g. also exemplified by the preferential immunogenicity of eGFP in Balb/c mice relative to C57BL/6 strains^{6,16,17}.

Our study provides a methodological framework through which shuffled transgenes are applied to avoid undesirable immune responses against exogenously introduced proteins. Both the *Tol* ORF and the C57BL/6 Tol strain used in this work are available upon reasonable request (from the corresponding author and the Netherlands Cancer Institute Transgenics Core Facility, respectively). Furthermore, creation of a resource of mouse strains that are tolerant for additional reporter and modifier proteins used in biomedical research would be of considerable value.

Methods

Generation of *Tol* open reading frame and vaccination plasmids

The *Tol* ORF was designed as described in Supplementary Fig. 1 and synthesized by Genscript. The full-length ORF was subcloned into a recombinase-mediated cassette exchange (RMCE) compatible vector, producing the pF3-CAG-*Tol*-FRT plasmid (Addgene ID: 141349). In the resultant vector, the *Tol* ORF was positioned downstream of the CAG promoter and flanked by F3 and FRT recombination sites. The entire F3-CAG-*Tol*-FRT cassette was sequence verified by Sanger sequencing. *Katushka*, *Kaede*, *AzamiGreen*, *mKO2* and *tagBFP* ORFs were shuttled directly into the multiple cloning site of pMP71 via Gibson cloning to generate pMP71-*Katushka*, pMP71-*Kaede*, pMP71-*AzamiGreen*, pMP71-*mKO2* and pMP71-*tagBFP*, respectively. Generation of pVAX-E7₄₉₋₆₇ has been described previously²². pVAX-SIINFEKL (Addgene ID: 141350) was generated by Gibson cloning to produce scarless fusions between SIINFEKL, several CD4⁺ T cell epitopes (HELP), and the KDEL-signal peptide.

Mice

C57BL/6-Ly5.1 and OT-I mice were obtained from Jackson Laboratories and crossed to obtain C57BL/6-Ly5.1-OT-I donor mice for adoptive transfer experiments. All animals were maintained and bred in the animal department of The Netherlands Cancer Institute and used for experimentation at 7-14 weeks. All animal experiments were approved by the Animal Welfare Committee of the NKI, in accordance with national guidelines.

Generation of *Tol* transgenic mice

The *Tol* expression cassette was introduced into a locus 3' to the *Col1a1* locus¹⁴ using recombinase-mediated cassette exchange (RMCE)¹⁸. In brief, a C57BL/6J ES cell line was derived from blastocysts¹⁹ and an F3-Puro-deltaTK-FRT cassette was targeted into the ESCs by homologous recombination, as described¹⁴. The pF3-CAG-*Tol*-FRT plasmid was co-transfected with pCAGGS-FLPe into the B6J-RMCE ES cells, followed by selection in medium containing flaluridine. Clones were screened by PCR for correct and complete integration of the F3-*Tol*-FRT cassette into the *Col1a1* locus. Correctly modified ESCs were then injected into B6/NTAC blastocysts, and injected blastocysts were transferred into pseudopregnant B6CBAF1/JRj foster mice, as described previously^{13,19}. Resulting chimeras were tested for chimerism by quantitative-PCR (Q-PCR). Positive animals (>0.4 RQ) were crossed with C57BL/6JRj mice (Janvier) to generate experimental cohorts. Presence of the *Tol* cassette was confirmed in experimental cohorts by PCR using the forward 5'-GGAAAGAATCACAACCTTACG-3' and reverse 5'-AGAGCATTTTCGGTTGAGGCC-3' primers. The *Tol* strain described in this communication is available upon request from the Netherlands Cancer Institute Transgenics Core Facility.

Histopathology

For histopathological analyses, 2 mm-thick hematoxylin-eosin stained sections were prepared from formalin-fixed, paraffin-embedded murine tissues such as skin, spleen, thymus, lymph nodes, liver, pancreas, gastrointestinal tract, heart, lung, kidneys, testes, ovaries, accessory sex glands, bone marrows (sternum and extremity), and muscles. Sections were evaluated and scored by an animal pathologist blinded to animal genotype.

RNA sequencing

RNA was extracted from the indicated frozen tissues using the RNeasy Mini Kit (Qiagen). Whole transcriptome sequencing samples were prepared with the TruSeq Stranded mRNA Kit (Illumina). Single-end 65 bp sequencing was performed on a HiSeq 2500 System (Illumina). Transcript abundance was calculated using Salmon 0.14.1²⁰, based on the GRCm38 transcriptome build. The *Tol* transcript sequence was added before read alignment. Differential gene expression analysis was performed using EdgeR 3.9²¹.

Production of retroviral supernatants

Phoenix-E packaging cells were seeded at 1.4×10^6 cells per 10 cm dish in IMDM supplemented with 8% FCS, glutamax, penicillin/streptomycin (Gibco). After 24 hours, medium was refreshed and cells were transfected using FuGene6 (Roche), following the manufacturer's protocol. 48 hours post transfection, viral supernatant was harvested and passed through a 0.22 μm filter (Sigma). Filtered retroviral supernatant was snap-frozen in liquid nitrogen and stored at -80 C.

Generation and adoptive transfer of Kaede, Katushka, BFP, AzamiGreen or mKO2 positive OT-I splenocytes

Spleens from C57BL/6;Ly5.1;OT-I mice were passed through 70 μm strainers to obtain a single cell suspension. Splenocytes were then seeded at 6×10^6 cells/ml in RPMI 1640 supplemented with 8% FCS, glutamax, penicillin/streptomycin, pyruvate, non-essential amino acids, HEPES, β -mercapto-ethanol (Gibco), 2 $\mu\text{g/ml}$ concanavalin A (Calbiochem) and 1 ng/ml IL-7 (Peprotech). After 48 hours, T cells were harvested and re-seeded on 24-well plates coated with Retronectin (Takara) at 1.5×10^6 cells per well in RPMI 1640 supplemented with 8% FCS, glutamax, penicillin/streptomycin, pyruvate, non-essential amino acids, HEPES, β -mercapto-ethanol and 50 IU IL-2 (Novartis). Retrovirus containing supernatant was added and cells were transduced by spin-infection at 2,000 rpm for 90 minutes. After 24 hours, cells were harvested and washed twice in PBS. Subsequently, T cells were resuspended in HBSS (Gibco) and 1×10^6 CD8⁺ cells were intravenously injected via the tail vein.

DNA vaccination

One day prior to vaccination with DNA encoding MHC-II class restricted helper epitopes²², plus either the OVA₂₅₇₋₂₆₄ epitope ('Help-OVA') or the E7₄₉₋₆₇ epitope ('Help-E7'), hair of mice was removed from hind legs using Veet depilation cream (Reckitt Benckiser). Primary DNA vaccination was subsequently performed on day 0, 3 and 6, as previously described²³. In brief, a droplet of 15 µl of a 2 µg/µl DNA solution in 10 mmol/L Tris pH 8.0 and 1 mmol/L EDTA pH 8.0 was applied on both the inside and outside of the leg, and vaccination was performed using a rotary tattoo device with a sterile disposable 9-needle bar (MT.DERM) oscillating at a frequency of 100 Hz for 1 min with a needle depth of 1 mm. For secondary vaccinations, mice received a single DNA tattoo with 20 µl of the 2 µg/µl plasmid solution on both the inside and outside of the leg, at >155 days after start of primary vaccination.

Analysis of antigen-specific T cell responses by flow cytometry

Blood samples were collected from the tail vein at the indicated time points. Erythrocytes were removed by treatment with NH₄Cl buffer, and cells were washed and then stained with anti-CD45.1-APC (A20, Thermo Fisher Scientific), anti-CD8-PerCp-Cy5.5 (eBioH35-17.2, Thermo Fisher Scientific), near-IR dye (Thermo Fisher Scientific) and indicated MHC multimers. MHC multimers were produced in-house by UV-induced ligand exchange and subsequent labeling with BV421, APC or PE (Thermo Fisher Scientific), as described previously²⁴. Flow cytometry data were acquired on a Fortessa (BD Biosciences) and analyzed in FlowJo (version 10.4.2) according to the gating strategy shown in Supplementary Fig. 5.

Statistics and reproducibility

Statistical analyses were performed in Prism (GraphPad), comparing groups of mice with the two-tailed Mann-Whitney test. Results were regarded as statistically significant at a *P*-value of <0.05.

Data availability

Mouse cDNA sequences (GRCm38) used for Salmon alignment were downloaded from Ensembl (<http://www.ensembl.org/biomart/martview>). RNA-seq data have been deposited, and are available from the Gene Expression Omnibus under accession number GSE147757. All source data underlying the graphs and charts presented in the main figures are presented in Supplementary Data 3.

Acknowledgements

We would like to thank M. Toebe for generation of MHC multimers, and L. Kroese, R. Bin Ali and T. Braumuller for technical assistance in generation and validation of the Tol mouse model. In addition, we thank the NKI animal facility, animal pathology, flow cytometry and genomics

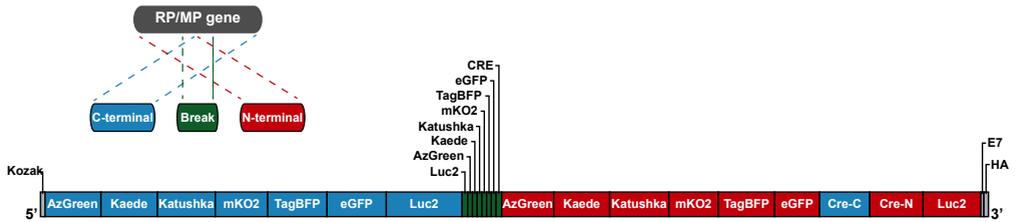
core facility for technical support and members of the Schumacher and Haanen laboratories for discussions. Finally, we would like to thank M. E. Hoekstra for providing illustrations used in the figures. This work was supported by Institute for Chemical Immunology (ICI) grant 003 and ERC AdG Life-His-T (to T.N.S.).

Author contributions

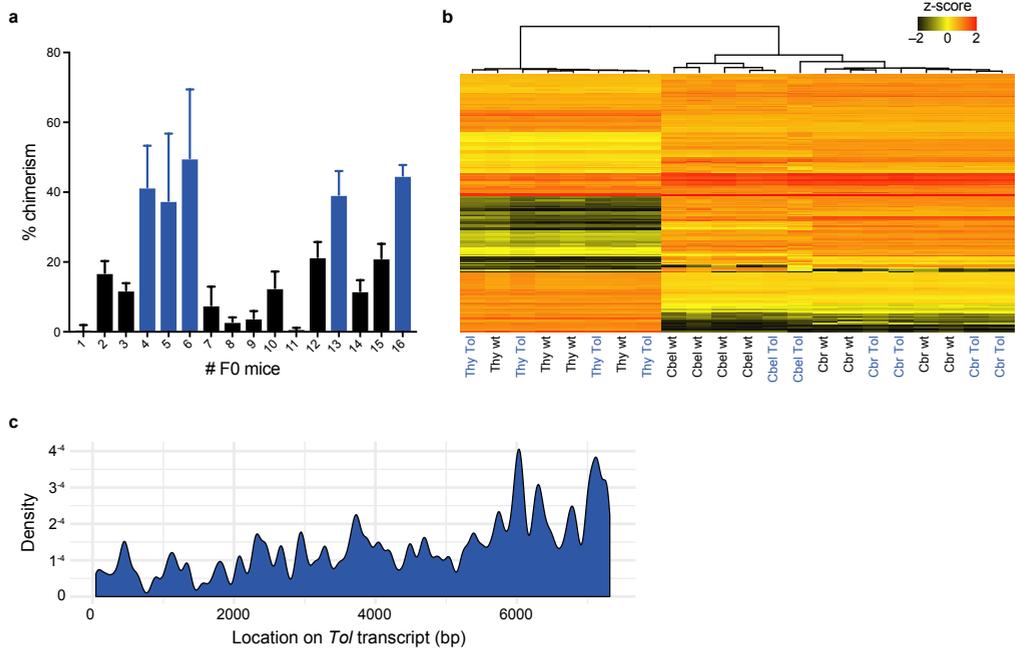
K.B. generated the *Tol* targeting vector and other vectors, and performed monitoring of T cell responses. F.E.D. performed adoptive cell transfer and vaccination experiments, and performed monitoring of T cell responses. J.-Y.S. evaluated histopathological data. J.C.R., K.B., F.E.D. and F.A.S. contributed to design of the *Tol* ORF. I.J.H. and C.E.J.P. generated the genetically engineered Tol mouse model. K.B., F.E.D., F.A.S. and T.N.S. contributed to experimental design and prepared the manuscript with input of all co-authors.

References

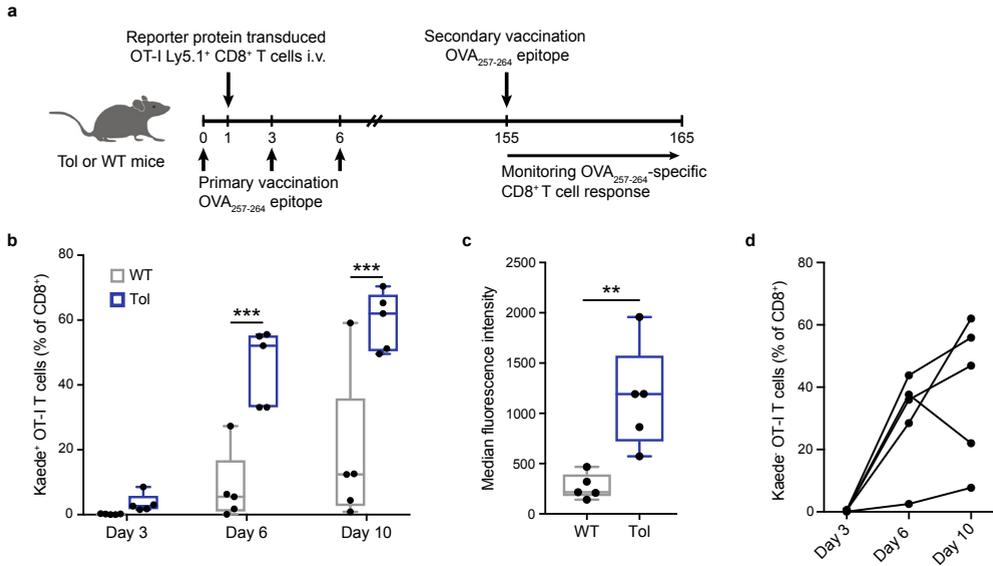
1. Ghim, C.-M., Lee, S. K., Takayama, S. & Mitchell, R. J. The art of reporter proteins in science: past, present and future applications. *BMB Reports* 43, 451–460 (2010).
2. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* 263, 802–805 (1994).
3. Shaner, N. C. et al. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature biotechnology* 22, 1567–1572 (2004).
4. Orban, P. C., Chui, D. & Marth, J. D. Tissue- and site-specific DNA recombination in transgenic mice. *PNAS* 89, 6861–6865 (1992).
5. Davey, G. M. et al. Identification of a MHC I-restricted epitope of DsRed in C57BL/6 mice. *Mol. Immunol.* 53, 450–452 (2013).
6. Han, W. G. H., Unger, W. W. J. & Wauben, M. H. M. Identification of the immunodominant CTL epitope of EGFP in C57BL/6 mice. *Gene Ther.* 15, 700–701 (2008).
7. Gambotto, A. et al. Immunogenicity of enhanced green fluorescent protein (EGFP) in BALB/c mice: identification of an H2-Kd-restricted CTL epitope. *Gene Ther.* 7, 2036–2040 (2000).
8. Limberis, M. P., Bell, C. L. & Wilson, J. M. Identification of the murine firefly luciferase-specific CD8 T-cell epitopes. *Gene Ther.* 16, 441–447 (2009).
9. Xing, Y. & Hogquist, K. A. T-cell tolerance: central and peripheral. *Cold Spring Harb Perspect Biol* 4, a006957–a006957 (2012).
10. Wyss, L. et al. Affinity for self antigen selects Treg cells with distinct functional properties. *Nature immunology* 17, 1093–1101 (2016).
11. Stripecke, R. et al. Immune response to green fluorescent protein: implications for gene therapy. *Gene Ther.* 6, 1305–1312 (1999).
12. Sellers, R. S., Clifford, C. B., Treuting, P. M. & Brayton, C. Immunological variation between inbred laboratory mouse strains: points to consider in phenotyping genetically immunomodified mice. *Vet. Pathol.* 49, 32–43 (2012).
13. Huijbers, I. J. et al. Using the GEMM-ESC strategy to study gene function in mouse models. *Nat Protoc* 10, 1755–1785 (2015).
14. Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* 44, 23–28 (2006).
15. Hetz, C. & Saxena, S. ER stress and the unfolded protein response in neurodegeneration. *Nat Rev Neurol* 13, 477–491 (2017).
16. Aoyama, N. et al. Transgenic mice that accept Luciferase- or GFP-expressing syngeneic tumor cells at high efficiencies. *Genes Cells* 23, 580–589 (2018).
17. Skelton, D., Satake, N. & Kohn, D. B. The enhanced green fluorescent protein (eGFP) is minimally immunogenic in C57BL/6 mice. *Gene Ther.* 8, 1813–1814 (2001).
18. Prosser, H. M., Koike-Yusa, H., Cooper, J. D., Law, F. C. & Bradley, A. A resource of vectors and ES cells for targeted deletion of microRNAs in mice. *Nature biotechnology* 29, 840–845 (2011).
19. Huijbers, I. J. et al. Rapid target gene validation in complex cancer mouse models using re-derived embryonic stem cells. *EMBO Mol Med* 6, 212–225 (2014).
20. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* 14, 417–419 (2017).
21. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010).
22. Ahrends, T. et al. CD27 Agonism Plus PD-1 Blockade Recapitulates CD4+ T-cell Help in Therapeutic Anticancer Vaccination. *Cancer research* 76, 2921–2931 (2016).
23. Bins, A. D. et al. A rapid and potent DNA vaccination strategy defined by in vivo monitoring of antigen expression. *Nature medicine* 11, 899–904 (2005).
24. Toebes, M., Rodenko, B., Ovaa, H. & Schumacher, T. N. M. Generation of peptide MHC class I monomers and multimers through ligand exchange. *Curr Protoc Immunol Chapter* 18, Unit 18.16–18.16.20 (2009).



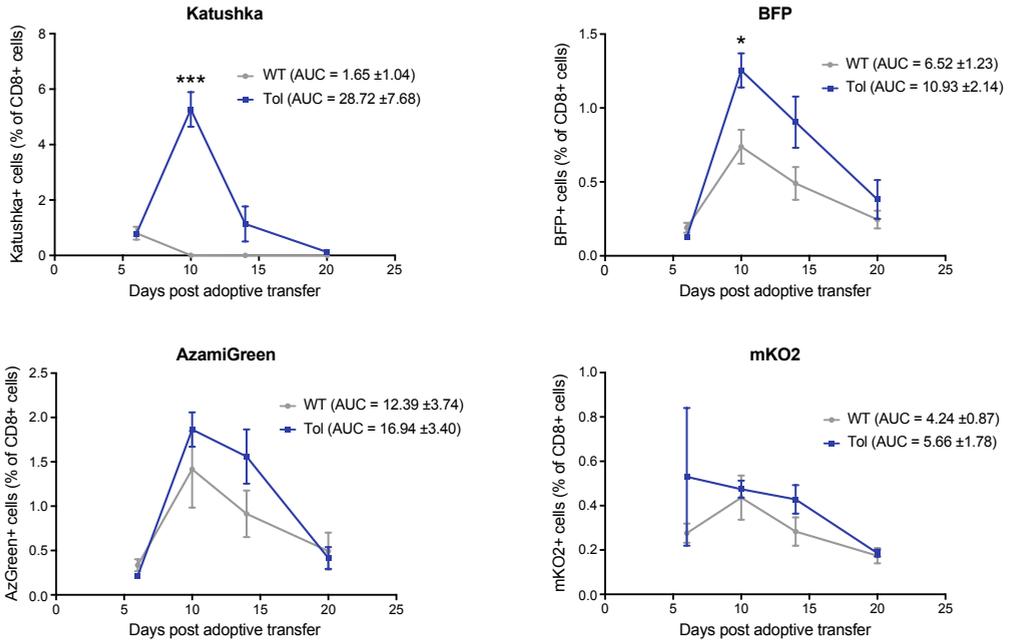
Supplementary Fig. 1. Design of the Tol open reading frame. Cartoon depiction of the Tol open reading frame (ORF), with the position of each gene fragment depicted. AzGreen = Azami Green, mKO2 = mKusabira Orange2, TagBFP = TagBlue fluorescent protein, eGFP = enhanced green fluorescent protein, Luc2 = luciferase2, E7 = HPV E749-57 epitope.



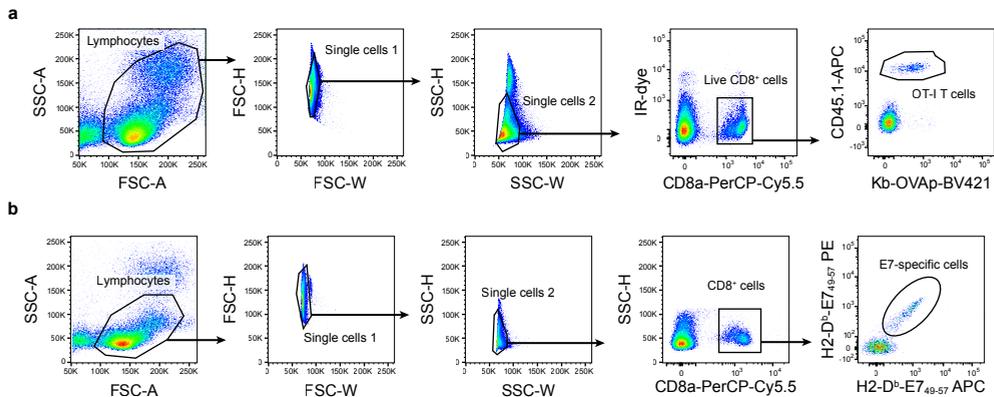
Supplementary Fig. 2. Validation of the Tol mouse model. **a**, Bar chart depicting the percentage of chimerism in F0 animals, as determined by quantitative PCR. Blue bars represent animals used for subsequent breeding. **b**, Hierarchical clustering analysis performed on RNA sequencing data from thymus, cerebrum and cerebellum of WT (black) and Tol (blue) mice. Complete-linkage clustering was performed on the top 1,000 most highly expressed genes in the dataset. Thy = thymus, Cbel = cerebellum, Cbr = cerebrum. **c**, Density of aligned reads along the Tol ORF. Data shown in (b-c) are representative of n=4 (Tol) and n=4 (WT) mice. PURO-R = puromycin resistance gene, TK = thymidine kinase gene, CPM = counts per million.



Supplementary Fig. 3. Enhanced recall potential of Kaede+ OT-I T cells in Tol mice. **a**, Schematic overview of adoptive cell transfer, vaccination and monitoring strategy. **b**, **c**, Percentage and MFI of Kaede+ OT-I T cells after secondary vaccination. Note that both the number of Kaede+ OT-I cells and the Kaede expression level are increased in Tol mice relative to WT mice. **d**, Percentage of Kaede+ OT-I T cells after secondary vaccination. Cells (b-d) are gated on IR-dye-CD8+ lymphocytes, and data depict n=5 (Tol) and n=5 (WT) mice. Box-plots (b, c) depict interquartile range, with whiskers representing minimum and maximum values. Dots (b, c) and lines (d) indicate individual mice. P values were determined by either Repeated Measures two-way ANOVA with Sidak's multiple comparisons test (b) or two-tailed Mann-Whitney signed rank test (c). **P < 0.01; ***P < 0.001.



Supplementary Fig. 4. Analysis of engraftment of Katushka, BFP, AzamiGreen and mKO2 modified T cells in WT and Tol mice. Percentage FP+ OT-I T cells detected in the blood of WT and Tol mice at indicated time-points. Lines represent group means of Tol (n=5) and WT (n=7) mice, error bars depict standard errors of the mean. Mean area under the curve (AUC) is given for each group. P values were determined by Repeated Measures two-way ANOVA with Sidak's multiple comparisons test to test the significance at each time point. *P < 0.05; **P < 0.01; ***P < 0.001.

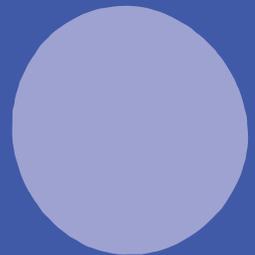
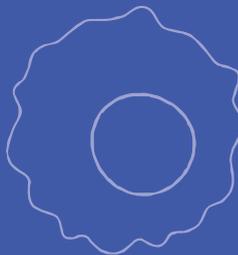
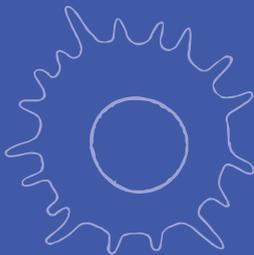
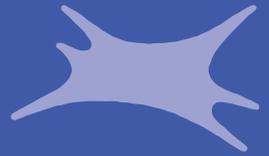
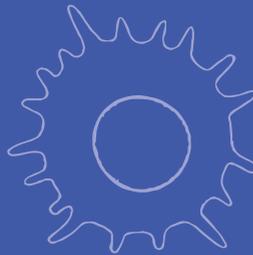
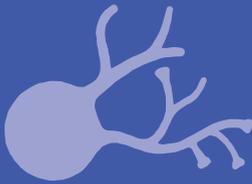


Supplementary Fig. 5. Gating strategy for the detection of antigen-specific T cell responses in blood samples by flow cytometry. **a**, Gating strategy applied to detect transferred OT-I T cell populations. **b**, Gating strategy applied to detect endogenous E7-specific T cell responses.

Chapter 3.

Gene and protein sequence features augment HLA class I ligand predictions

OR: "An 8-year project that was basically performed in 4 months."



Kaspar Bresser^{1,2}, Benoît P Nicolet³, Anita Jeko⁴, Wei Wu⁴, Fabricio Loayza-Puch⁵, Reuven Agami⁶, Albert JR Heck⁴, Monika C Wolkers³, and Ton N Schumacher^{1,2,#}

Unpublished findings

(1) Department of Molecular Oncology and Immunology, Netherlands Cancer Institute, Oncode Institute, Amsterdam, The Netherlands

(2) Department of Hematology, Leiden University Medical Center, Leiden, The Netherlands

(3) Department of Hematopoiesis, Sanquin Research, Landsteiner Laboratory, Amsterdam UMC, University of Amsterdam, Oncode Institute, Amsterdam, The Netherlands.

(4) Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands.

(5) Translational Control and Metabolism, German Cancer Research Center (DKFZ), Heidelberg, Germany.

(6) Division of Oncogenomics, Oncode Institute, The Netherlands Cancer Institute

Abstract

The sensitivity of malignant tissues to T cell-based cancer immunotherapies is dependent on the presence of targetable HLA class I ligands on the tumor cell surface. Peptide intrinsic factors, such as HLA class I affinity, likelihood of proteasomal processing, and transport into the ER lumen have all been established as determinants of HLA ligand presentation. However, the role of sequence features at the gene and protein level as determinants of epitope presentation has not been systematically evaluated. To address this, we performed HLA ligandome mass spectrometry on patient-derived melanoma lines and used this data-set to evaluate the contribution of 7,124 gene and protein sequence features to HLA sampling. This analysis revealed that a number of predicted modifiers of mRNA and protein abundance and turn-over, including predicted mRNA methylation and protein ubiquitination sites, inform on the presence of HLA ligands. Importantly, integration of gene and protein sequence features into a machine learning approach augments HLA ligand predictions to a comparable degree as predictive models that include experimental measures of gene expression. Our study highlights the value of gene and protein features to HLA ligand predictions.

Introduction

Spontaneous or immunotherapy-induced recognition and destruction of malignant tissues by the T cell-based immune system is, to a large extent, dependent on presentation of HLA class I bound peptides to antigen-specific CD8⁺ T cells¹⁻³. Consequently, the composition of the pool of peptide-HLA class I complexes at the cell surface—or the HLA class I ligandome—strongly determines the ‘visibility’ of tumor cells to CD8⁺ cytotoxic T cells. Understanding the various factors that define the composition of this HLA ligandome is thus of major value for cancer immunotherapy.

The HLA class I ligandome is primarily generated through the intracellular degradation of proteins by the proteasome, and subsequent translocation of peptide fragments into the ER lumen by the transporter associated with antigen processing (TAP). These peptides can undergo further trimming by ER-resident aminopeptidases, bind to the peptide-binding groove of HLA class I molecules, and finally traffic to the cell surface to be presented to the immune system^{4,5}. The number of peptides that can theoretically be generated from the human proteome is vast, adding up to approximately 10⁷ distinct peptides for 9-meric species alone⁶. This large space poses a substantial challenge in the prediction of the HLA ligandome of a cell population of interest. Over the past decades, significant advances have been made in reducing this complexity, primarily by focusing on characteristics of the peptide itself or its surrounding sequence. Specifically, HLA class I ligands bind to the peptide-binding groove of HLA class I through shared ‘anchor’ residues, a feature that has been leveraged in

the development of predictive algorithms^{7,8}. In addition, the predictable cleavage preference of the proteasome⁹ has been used to improve epitope prediction accuracy^{10,11}.

Beyond local sequence characteristics, a number of protein-level features are expected to play an important role in the generation of HLA binding peptides, for instance by tuning protein abundance and turn-over¹²⁻¹⁴. In prior work, transcriptome measurements have been used as a proxy for protein expression to aid HLA ligand predictions. However, mRNA and protein abundance correlate poorly in most mammalian cells¹⁵⁻¹⁷, primarily due to post-transcriptional regulation. Such post-transcriptional regulation includes the activity of RNA-binding proteins and non-coding RNA species, and sequence intrinsic features (e.g. GC content and codon usage), which can affect the translational output of mRNAs^{18,19}. Furthermore, post-translational modifications, including ubiquitination and glycosylation, are known to modulate protein abundance, localization, and turn-over rates^{20,21}, and may thereby influence epitope sampling.

In this study, we aimed to examine the potential value of gene and protein sequence features in the prediction of the HLA class I ligands. Implementing a machine learning approach, we show that the performance of such predictions can be improved through the addition of sequence features. Importantly, predictive models that make use of such features achieve the same level of predictive power as models that incorporate experimental measurements of gene-expression levels, and the predictive value of these features was generalizable to external data. Our data exemplify how the 'hard-coded' information of gene and protein sequence features can be exploited to infer a cell's proteomic content and its derivatives.

Results

Identification of human melanoma HLA ligandomes

To investigate putative determinants of the HLA ligandome, we performed LC-MS on pan-HLA immunoprecipitates of three melanoma lines (**Fig. 1A**), resulting in the identification of 18,819 peptides derived from 6,286 proteins at a false discovery rate of <1%. The length distribution of the LC-MS detected peptides closely matched that of known melanoma-derived HLA ligands (IEDB²², **Fig. 1B**), with the vast majority of peptides consisting of 9-meric species. Examination of positional frequencies of each amino acid revealed strong usage biases at position 2 and 9 (**Fig. 1C-D**). To assess whether this observed amino acid enrichment was explained by the known ligand preference of the HLA class I haplotypes expressed by these tumor lines, 9-meric peptide sequences from each melanoma line were clustered using the GibbsCluster algorithm²³. This analysis revealed dominant motifs present in each of the HLA ligandomes that closely matched the corresponding HLA haplotype consensus binding motifs for 11/11 HLA A and B alleles and 5/6 HLA C alleles (**Fig. 1E, Supplementary Figure 1**). In addition, HLA class I binding affinity predictions showed that the majority of LC-MS detected peptides (61.5–91.2%) were predicted to form ligands for at least 1 of the expressed HLA alleles (**Fig. 1F, Supplementary Figure 1**).

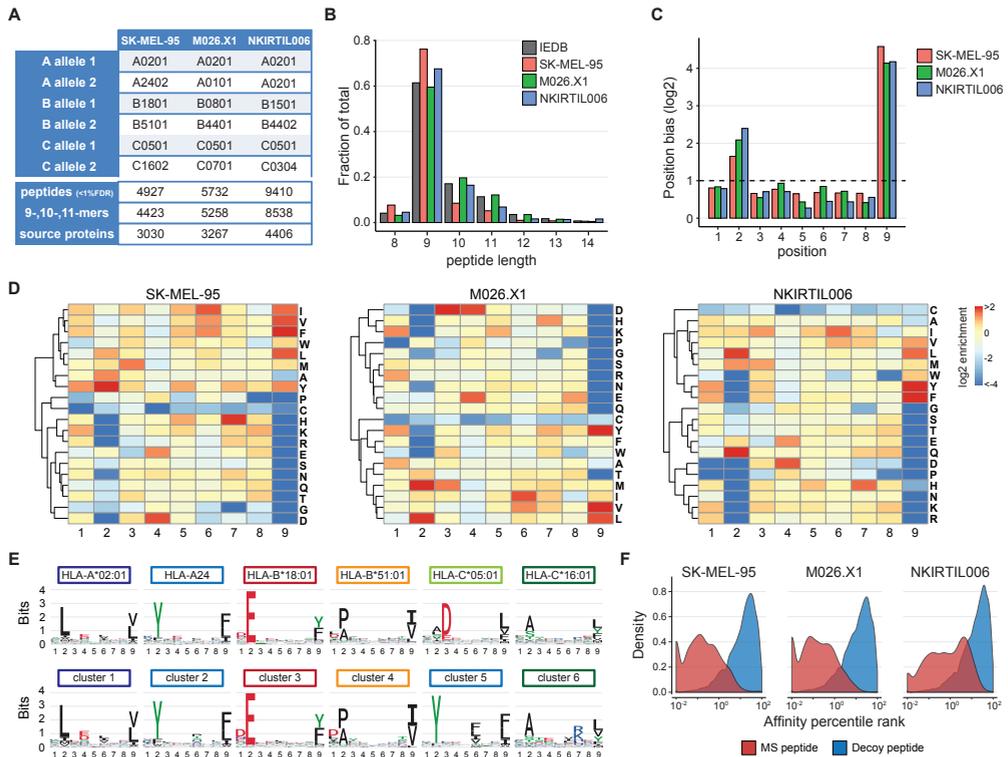


Figure 1. Identification of HLA ligandomes. (A) HLA class I haplotype of the melanoma lines used, and number of peptides and source proteins identified. (B) Peptide length distribution of each LC-MS dataset, compared to the peptide length distribution of known melanoma-derived HLA class I ligands deposited to IEDB. (C-D) Enrichment of indicated amino acids, relative to amino acid occurrence in the proteome, at each position of all 9-meric species in the datasets. Summary depicting the median of the absolute enrichment values of all amino acids for each position (C), and heatmaps visualizing hierarchical clustering of amino acid enrichment (E) are shown. (G) Sequence logos of all 9-mer ligands deposited to IEDB for the HLA class I alleles expressed by SK-MEL-95 (top), and the sequence logos of 6 peptide clusters obtained using the GibbsCluster algorithm (bottom). The number of clusters was constrained to the number of expressed HLA class I alleles. (H) Affinity percentile rank scores of LC-MS detected peptides, compared to randomly drawn peptides from transcribed genes of each melanoma line (decoy peptides).

Gene and protein features inform on HLA sampling

Gene and protein sequence features, such as post-transcriptional or post-translational modification sites, have been shown to provide information on mRNA or protein abundance^{19,24–26}. To determine whether such features can be employed to predict the presence of HLA ligands within the proteome, we made use of a library of 7,124 sequence features. This feature library includes codon and amino acid usage, RNA-binding motifs from 142 RNA-binding proteins (RBPs), predicted miRNA (miR) binding scores, and RNA modification sites that were separately identified in the 5' UTR, 3' UTR and coding sequence²⁷. Predicted post-translational modification (PTM) sites, such as ubiquitination, acetylation, and malonylation sites were additionally included. Of note, this sequence feature library comprises predicted mRNA and protein modification sites, rather than any experimental measurement of such modifications in the cell systems used.

To assess whether individual sequence features can inform of HLA sampling, we first explored

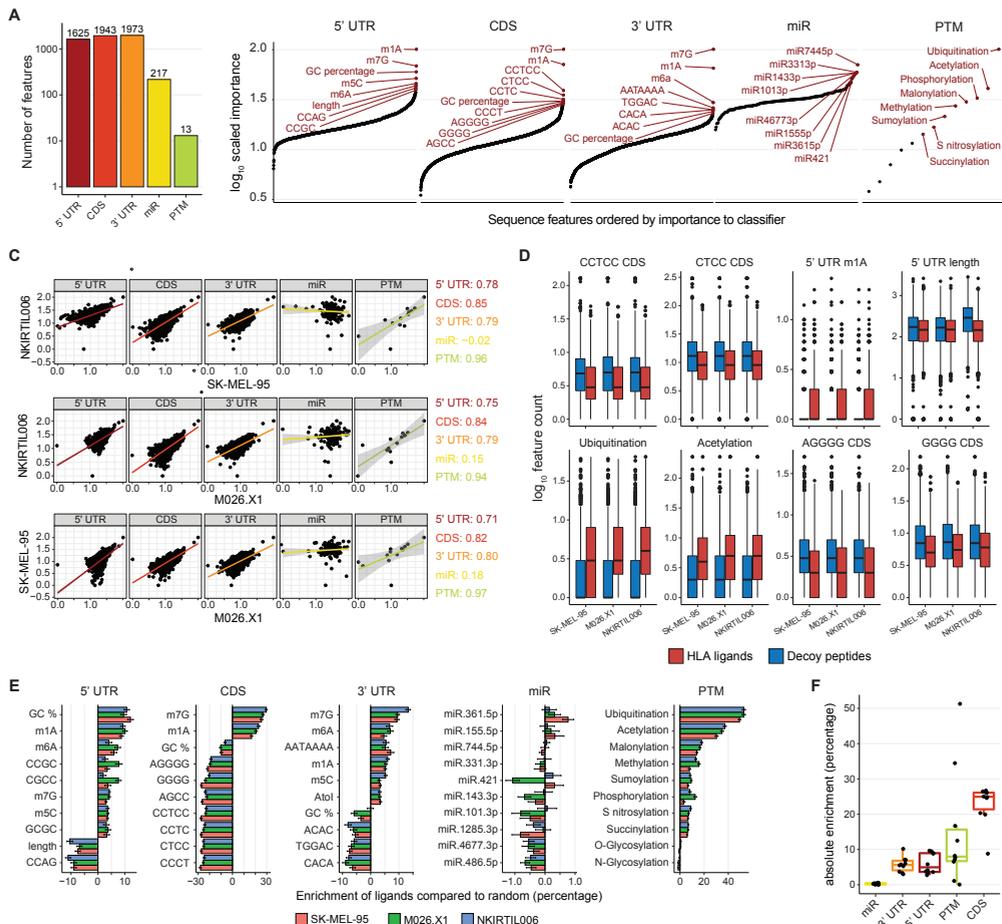


Figure 2. Sequence features inform on HLA sampling. Random forest models were trained using HLA ligandome data from each melanoma line and using individual classes of gene and protein sequence features to identify HLA ligands. **(A)** Sequence feature classes used to fit random forest classifiers for each melanoma line. Values indicate number of features per class. **(B)** Mean importance of sequence features to the models of each class. Feature importance represents the mean decrease in accuracy when that sequence feature is not included in the model, all importance scores are re-scaled per random forest model to a 0-100 scale. Dots indicate individual features. **(C)** Comparison of the importance of all sequence features across the individual random forest models. Dots indicate individual features, linear regressions are shown as colored lines, and 95% confidence intervals as greyed areas. Colored text denotes the respective Pearson correlation coefficients. **(D)** Comparison of sequence feature occurrence between 500 LC-MS detected HLA ligands and the same number of decoy peptides. Selected sequence features are shown. Boxplots indicate group median and 25th and 75th percentiles, whiskers indicate the interquartile range multiplied by 1.5, and dots signify individual peptides. **(E)** HLA ligands and decoy peptides were either ranked at random or by the indicated sequence feature, and the number of HLA ligands in the top 50 ranked peptides was quantified. Data depict the relative increase in HLA ligands as compared to random, bars indicate the mean percentage increase of 50 bootstraps, error bars depict 95% confidence intervals. **(F)** Comparison of averaged absolute enrichment values between feature classes. Boxplots indicate group median and 25th and 75th percentiles, whiskers indicate the interquartile range multiplied by 1.5, and dots signify individual features.

the contribution of a subset of features that could be assigned to five major feature classes (5' UTR, CDS, 3' UTR, miR binding, PTM) and that displayed a substantial degree of variance across the proteome (**Fig. 2A**, 5,771 out of 7,124 features in the library). A set of 2,000 HLA ligands was drawn from each tumor line and supplemented with a 2-fold excess of decoy peptides that were randomly sampled from the transcribed genome. This dataset was then used to train individual Random Forest classifiers for each tumor line and each sequence feature class, which were subsequently used to determine the importance of these sequence features to each of the obtained classifiers (**Fig. 2B**, **Supplementary Figure 2A**, showing normalized importance plots and RF metrics). The importance of sequence features was highly consistent between the different melanoma ligandome datasets, indicating that a shared set of features reliably informed on the presence of HLA ligands (**Fig. 2C**). Furthermore, direct comparison of the occurrence of high-importance sequence features within source proteins of HLA ligands and decoy peptides revealed significant differences for a set of sequence features (**Supplementary Figure 2**). For example, HLA ligands were preferentially sampled from proteins that contained a higher number of predicted sites for ubiquitination and acetylation, two PTMs that can regulate targeted proteasomal degradation and protein stability²⁸⁻³⁰ (**Fig. 2D**). Predicted N1-methyladenosine (m1A) sites within the 5' UTR were also enriched in the mRNA of source proteins of HLA ligands, an effect that appears consistent with the prior observation of enhanced translation efficiency of m1A-modified mRNA molecules²⁵. In contrast, 5'UTR length and occurrence of G-rich motifs in the CDS, features that have previously been suggested to negatively impact mRNA levels and translation, respectively^{31,32}, were negatively associated with the presence of HLA ligands (**Fig. 2D**).

To understand the ability of individual sequence features to contribute to HLA ligand prediction in a more quantitative manner, a custom enrichment score was calculated for each of the selected features (see methods). In brief, the set of HLA ligands and decoy peptides was sorted by the occurrence of each feature or was arranged in a random manner. Subsequently, the quantity of HLA ligands present in the top 50% ranked peptides was compared between these two cases, reflecting the benefit of each feature when used as a single determinant. In concordance with the prior analysis (**Fig. 2C**), miR binding site quantities exhibited no detectable bias toward HLA ligands or decoy peptides. In contrast, sequence features from the other classes showed a consistent capacity to enrich or deplete for the presence of HLA ligands (**Fig. 2E**). The most prominent associations were observed in the CDS and PTM classes (**Fig. 2F**), with some features increasing the number of ligands detected by more than 20%. Notably, computed m1A and N7-methylguanosine (m7G) sites were predictive of the presence of HLA ligands in the protein product irrespective of their location within the coding sequence or the untranslated regions (**Fig. 2E**), an observation that aligns with their general translation-enhancing capacity^{25,33}. Intriguingly, even though GC content was consistently informative on HLA sampling, its directionality was context dependent (positively correlated in the 5' UTR and negatively correlated in the 3'UTR and CDS), in line with prior reports suggesting that GC content may influence mRNA levels differently depending on location^{6,17,34}. Together, the above analyses show that gene and protein sequence features can individually inform on the presence of HLA ligands.

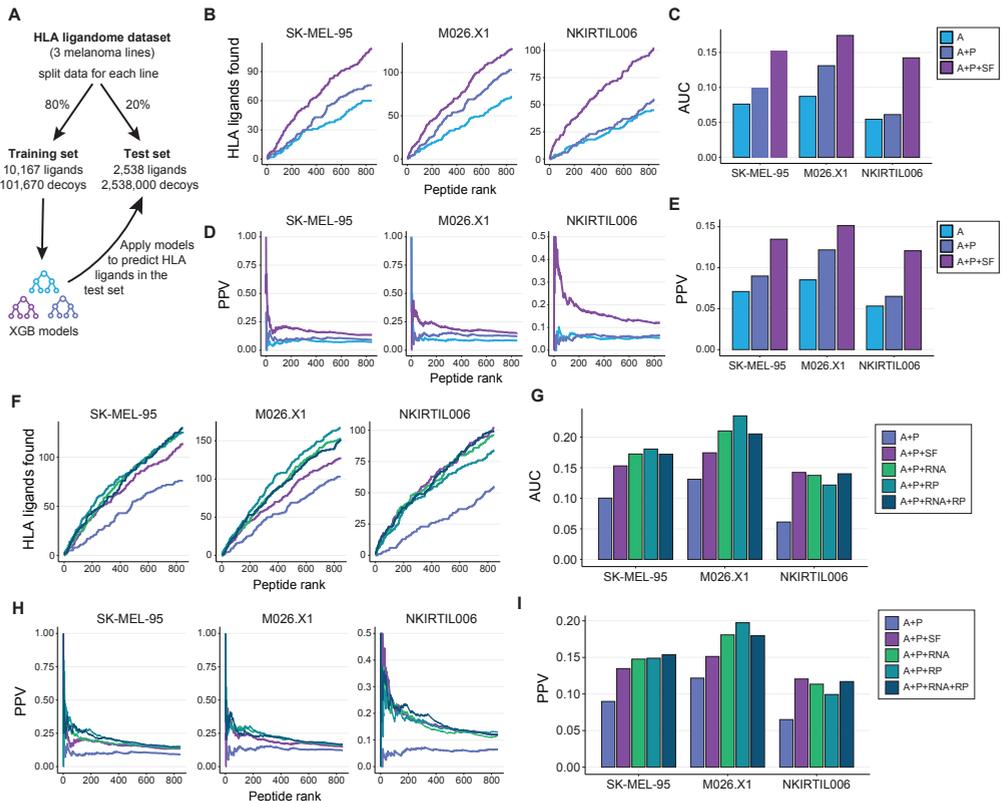


Figure 3. Value of sequence features in HLA ligand predictions. (A) The melanoma line data set was split into a training set and test set at an 80/20 ratio. The training set was used to build XGB classifiers using different combinations of features. (B-C) Number of true HLA ligands observed in the top 0.1% of predicted peptides from the matched melanoma line test set by each of the indicated models. Line graphs depicting the cumulative sum (B) and bar charts depicting AUCs (C) are shown. (D-E) Positive predictive value (PPV) at each peptide rank within the top 0.1% of predicted peptides from the melanoma line test set by each of the indicated models. (F-G) Quantity of true HLA ligands observed in the top 0.1% of predicted peptides from the melanoma line test set by each of the indicated models. Line graphs depicting the cumulative sum (F) and bar charts depicting AUCs (G) are shown. (H-I) Positive predictive value (PPV) at each peptide rank within the top 0.1% of predicted peptides from the melanoma line test set by each of the indicated models. Features used to build classifiers were predicted HLA class I affinity (A), predicted proteasomal processing (P), transcript abundance (RNA), ribosome occupancy (RP), and the sequence feature library (SF).

Sequence features augment HLA ligand predictions

Having shown that individual sequence features can inform on HLA sampling, we next assessed whether these features can be leveraged to improve HLA ligand prediction models. To this end, the melanoma HLA class I ligand dataset was divided into a training set (80%) and test (20%) set that were supplemented with a 4-fold and 1,000-fold excess of decoy peptides, respectively. To evaluate the added value of sequence features to classical HLA ligand prediction methods, such as netMHC (HLA affinity) and netChop (proteasomal processing), the training set was used to generate multiple XGBoost³⁵ classifier models (Fig. 3A), each integrating a different set of explanatory variables. As reported previously^{8,11}, both computed HLA affinity and proteasomal processing were strongly predictive of HLA sampling (Supplementary Figure 2C, D). Importantly, applying the

obtained models to predict HLA ligands in the test set revealed that the classifier that included sequence feature information consistently and substantially outperformed the models that lacked this information. Specifically, the model including sequence features consistently ranked true HLA ligands at a higher position (**Fig. 3B-C**) and increased the positive predictive value by approximately 1.5-fold (**Fig. 3D-E**).

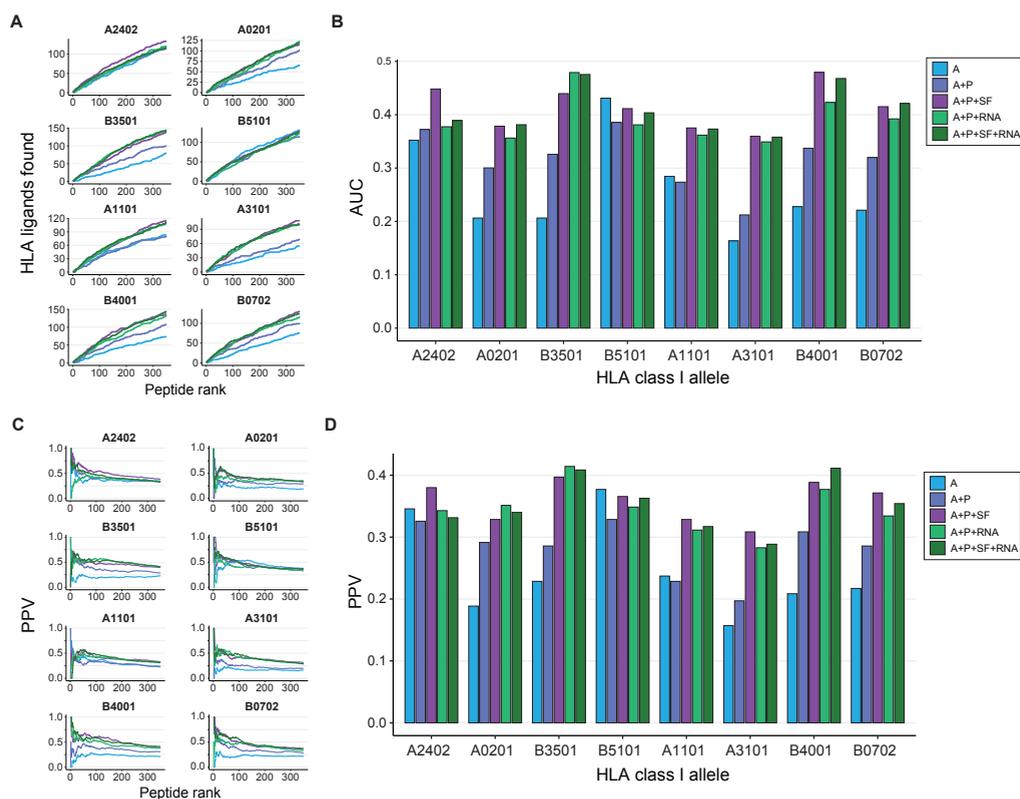
To determine whether the sequence features that were highly informative of HLA sampling when testing separate feature classes (figure 2) were also substantially contributing to the XGBoost classifier, the importance of those features was examined. This assessment revealed that the top scoring features in the prior analyses, such as predicted ubiquitination, acetylation, m7G, and m1A sites, were likewise dominant contributing factors in the XGBoost classifier (**Supplementary Figure 2D**). Furthermore, after predicted affinity and proteasomal processing, PTM and CDS features were generally assigned the highest importance scores (**Supplementary Figure 2E**), underlining their significance in HLA ligand prediction.

Sequence features can match ‘wet lab’ measures of gene transcription and translation

mRNA abundance measurements have been applied in prior studies to enhance the quality of HLA ligand prediction classifiers^{36,37}, and a number of recent efforts have established analysis of ribosome occupancy as a measure of active protein translation^{38,39}. To understand the value of protein and gene sequence features relative to these experimental measurements of either gene expression or ribosome occupancy, we generated mRNA sequencing (transcript abundance) and ribosome profiling (protein translation activity) datasets for each of the melanoma lines. Consistent with previous reports, both measures of gene expression were strongly indicative of HLA sampling and the combination of transcript level data and ribosome occupancy data offered little additional benefit (**Supplementary Figure 3A-E**). Next, to directly compare the predictive value of sequence features versus these ‘wet-lab’ measures of gene transcription and protein translation, additional XGBoost classifiers were trained that included these metrics. Application of this new set of models on the test dataset revealed that the XGBoost model that included sequence features was able to predict true HLA ligands at an equal potency as models that incorporated wet-lab measurements of gene transcription and protein translation (**Fig. 3F-I**). Interestingly, addition of wet-lab measurements of gene transcription and protein translation in a model that contained sequence features did not consistently improve predictiveness (**Supplementary Figure 3F-G**), indicating that the predictive value of sequence features and wet-lab measurements is largely redundant. Together, these data show that gene and protein sequence features jointly provide a similar degree of information on HLA ligandome composition as experimentally obtained expression levels.

Sequence feature-based HLA ligand classifiers are generalizable to external data

To understand whether the value of sequence features in HLA ligandome prediction is generalizable,



we subsequently tested whether the above observations could be extended to another tumor type, other HLA class I alleles, and to an independent dataset. To this purpose, we assessed the performance of the different classifiers on an HLA ligandome data set obtained from multiple mono-allelic human B lymphoblastoid cell lines³⁷, focusing on HLA ligandome data for the 4 most common HLA class A and B alleles⁴⁰, plus HLA ligandome datasets for 4 common HLA A and B alleles that were absent from the original melanoma dataset. From each of these ligandomes, 350 9-meric HLA ligands were sampled and supplemented with a 1,000-fold excess of decoy peptides, resulting in 8 individual test datasets. Application of the classifiers trained on internal ligandome data to these external test sets showed that models that included sequence features outperformed predicted HLA affinity- and processing-based models in 7 out of 8 datasets (**Fig. 4A-D**). Furthermore, the performance of the sequence feature-based classifier was equal to that of XGBoost models that included the matched gene expression information from the external data sets (**Fig. 4F-G**), demonstrating that the value of gene and protein feature-based prediction models is generalizable across cell types and datasets.

Discussion

Gene and protein sequence features represent a class of ‘hard-coded’ regulators of protein expression, influencing this process at many different levels. In this study, we leveraged a large set of such gene and protein features to assess their contribution to the composition of the HLA ligandome. Through this effort, we demonstrate that sequence features can augment HLA ligand predictions, and that the predictive gain obtained in these models is equal to that of models that incorporate experimentally obtained gene expression and protein translation data.

While not formally assessed here, it is expected that at least some of the sequence features contribute to HLA ligand predictions by providing a proxy for protein abundance. This notion is supported by the observation that sequence features such as mRNA region length, GC content, and post-transcriptional modifications can be used to help predict protein levels¹⁷. Furthermore, predicted RNA methylation sites were among the features that were most prominently associated with the presence of HLA ligands, an observation that may be explained by their known modulatory effect on both mRNA stability and translation efficiency^{25,26,33,41}. In addition to features involved in mRNA regulation and translation, our data reveal that the predicted occurrence of several PTM sites informed on the presence of HLA ligands. For ubiquitination, the PTM that displayed the highest predictive value, its positive association with HLA ligand yield may be caused by an enhanced accessibility to proteasomal degradation^{20,30,42}. For other PTMs that were predictive of HLA ligands, such as methylation and acetylation, their involvement in specific pathways is less well understood^{21,43,44}. Our data provide correlative evidence that these modifications influence availability of proteins to the antigen processing machinery but further work will be required to formally test this.

Improvement of HLA ligand prediction approaches remains an active field of research, with the aim to, for example, allow the more precise selection of cancer (neo)antigens for therapeutic purposes^{36,37,45,46}. Because of its generalizable nature and lack of requirement for direct transcriptome measurements, we envision that the approach described here will be of value in these efforts.

Methods

Patient-derived melanoma cell lines

SK-MEL-95 and M026.X1⁴⁷ were a kind gift from Daniel Peeper (Netherlands Cancer Institute). SK-MEL-95 was originally established in the Memorial Sloan Kettering Cancer Center. NKIRTL006 was established in house⁴⁸.

Cell culture

Patient-derived melanoma cell lines were cultured in RPMI (Gibco) supplemented with 8% fetal calf serum (FCS, Sigma), 100 U/ml penicillin (Gibco) and 100 µg/ml streptomycin (Gibco) at 37

°C and 5% CO₂. For mRNA sequencing and ribosome profiling, cell lines were cultured to a density of 70-90% on 150mm Corning tissue-culture treated culture dishes (Merck). For HLA ligandome LC-MS, approximately 1·10⁹ cells were cultured in Corning CELLSTACK Culture Chambers (Corning, 05-539-096).

HLA class I peptide isolation and LC-MS/MS

HLA class I-associated peptides were isolated by immunoprecipitation of HLA class I complexes using the mouse monoclonal IgG2a antibody W6/32, as described previously⁴⁹. Peptides were eluted from HLA class I protein molecules using a 10% acetic acid (v/v) solution, and subsequently separated using a 10 kDa molecular weight cutoff filter. Obtained solution was then desalted into 3 fractions using in-house made c18 STAGE (STop And Go Extraction) tips, eluted with 20%, 30% and 50% acetonitrile, respectively. The resulting fractions were injected on an Agilent 1290 system using a 120-min gradient coupled to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). Fractions 1 and 2 were injected in triplicate, whereas fraction 3 was injected in duplicate. The LC system comprised of a 20 × 0.1 mm i.d. trapping column (Reprosil C18, 3 μm; Dr. Maisch) and a 50 × 0.005 cm i.d. analytical column (Poroshell 120 EC-C18; 2.7 μm). An LC resolving gradient of 13 to 43% Solvent B (80% acetonitrile, 20% water, 0.1% formic acid) was used. The Top Speed method was enabled for fragmentation, where the most abundant precursor ions were selected in a 3 s cycle for data-dependent EThcD. MS1 and MS2 spectra were acquired at a resolution of 60,000 (FWHM at 400 m/z) and 15,000 (FWHM at 400 m/z), respectively. RF lens voltage was set to 60. Dynamic exclusion of 18s was used. Peptide precursors of charges 2 to 6 were fragmented, if accumulated to a minimum intensity of 4·10⁵ within 50 ms. In MS2, a maximum injection time of 250ms was used with a minimum intensity filter of 5·10⁴.

HLA class I peptide analysis

RAW data files were analyzed using the Proteome Discoverer 1.4 software package (Thermo Fisher Scientific). MS/MS scans were searched against the human Swissprot reviewed database (accessed in September 2015; 20,203 entries), with no enzyme specificity using the SEQUEST HT search engine. Precursor ion and MS/MS tolerances were set to 10 ppm and 0.05 Da. Methionine oxidation was set as variable modification. The peptides-to-spectrum matches were filtered for precursor tolerance 5 ppm, <1% FDR using Percolator, XCorr >1.7, and peptide rank 1. Peptides that were between 8 and 14 amino acid long were selected for further analysis. The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD036277.

Replicate injections displayed an overlap of approximately 70% (shared between at least 2 replicates). Consistent with their shared tissue origin, a large part of peptides detected across the melanoma lines mapped to a core group of proteins (47.6% shared between at least 2 lines). In contrast, the MS detected peptides exhibited a small degree of overlap (12.4% shared between at least 2 lines), in line with their difference in HLA haplotype.

mRNA sequencing

Cells were cultured to an approximate density of 80%, and $1 \cdot 10^7$ cells were subsequently dissociated using a cell-scraper in cold (4 °C) PBS, centrifuged for 10 minutes at 300x g, and snap-frozen in liquid nitrogen. RNA was extracted from the frozen pellets using the RNeasy Mini Kit (Qiagen). Whole transcriptome sequencing samples were prepared using the TruSeq Stranded mRNA Kit (Illumina). Single-end 65 bp sequencing was performed on a HiSeq 2500 System (Illumina). Obtained reads were aligned to the GRCh38 reference (gencode release 21) using STAR aligner (version 2.5.2b), and transcripts were quantified using Salmon (version 0.7.0). Transcript counts belonging to a single consensus coding sequence were summed.

Ribosome profiling

Cells were cultured to an approximate density of 80%, and $5 \cdot 10^7$ cells were subsequently treated with 100 µg/ml cycloheximide for 5 minutes at 37 °C. Cells were then washed once in cold (4 °C) PBS containing 100 µg/ml cycloheximide, dissociated using a cell-scraper in cold (4 °C) PBS supplemented with 100 µg/ml cycloheximide, centrifuged for 10 minutes at 300x g, and snap-frozen in liquid nitrogen. Frozen pellets were resuspended in lysis buffer (20 mM Tris-HCl, pH 7.8, 100 mM KCl, 10 mM MgCl₂, 1% Triton X-100, 2 mM DTT, 100 µg/ml cycloheximide, 1× Complete protease inhibitor), and incubated on ice for 20 minutes. Lysates were sheared using a 26G needle, centrifuged for 10 minutes at 1,300x g, and supernatants were transferred to a clean tube. Supernatants were treated with 2 U/µl of RNase I (Ambion) for 45 min at room temperature, with rotation. Next, lysates were fractionated on a linear sucrose gradient (7–47%) using the SW-41Ti rotor (Beckman Coulter) at 221,633x g for 2 hours at 4 °C, without brake. Obtained sucrose gradients were then divided in 14 fractions, and fractions 7–10 (cytosolic ribosomes) were pooled and treated with PCR grade proteinase K (Roche) in 1% SDS to release ribosome protected fragments. The resulting fragments were subsequently purified using Trizol reagent (Invitrogen) and precipitated in the presence of glycogen, following the manufacturer's instructions. For library preparation, RNA was gel-purified on a denaturing 10% polyacrylamide urea (7 M) gel. A section corresponding to 25 to 36 nucleotides—the region that comprises the majority of the ribosome-protected RNA fragments—was excised, and purified through ethanol precipitation. RNA fragments were then 3'-dephosphorylated using T4 polynucleotide kinase (New England Biolabs) for 6 hours at 37°C in 2-(N-morpholino)ethanesulfonic acid (MES) buffer (100 mM MES-NaOH pH 5.5, 10 mM MgCl₂, 10 mM β-mercaptoethanol, 300 mM NaCl). The 3' adaptor was added using T4 RNA ligase 1 (New England Biolabs) for 2.5 hours at 37°C. Ligation products were 5'-phosphorylated with T4 polynucleotide kinase for 30 minutes at 37 °C, and the 5' adaptor was added using T4 RNA ligase 1 for 2 hours at 37 °C. Sequencing was performed on a HiSeq 2500 System (Illumina). Ribosome occupancy was calculated using the Ribomap pipeline⁵⁰, and was aligned to the GRCh38 reference (gencode release 21). Counts belonging to a single consensus coding sequence were summed.

Characterization of LC-MS detected peptides

For comparison of peptide length distributions, known melanoma HLA class I ligands were downloaded from the IEDB web-interface (<https://www.iedb.org>) in June 2021 using the following search filters: Epitope – Any; Assay Outcome – Positive; MHC restriction – Class I; Host – Human; Disease – Melanoma.

To assess the amino acid positional biases of the LC-MS detected peptides, the dataset was filtered for 9-meric species, and the occurrence of each amino acid on each peptide position was tallied. As a reference, all expressed proteins (TPM > 0 in the mRNAseq dataset) were selected for each melanoma line, and the number occurrences of each amino acid was calculated. Amino acid enrichment was then defined as the fraction by which an amino acid occurred at a certain position divided by the fraction by which that amino acid occurred in the reference. The positional bias was defined as the median of the absolute amino acid enrichment values for each peptide position. For binding motif analyses, 9-meric peptide sequences from each melanoma line were clustered using GibbsCluster 2.0 (command line options set to: -g 3-7 -C -D 4 -I 1 -S 5 -T -j 2 -c 1 -k 25), with the number of clusters for each melanoma line set to the number of alleles expressed by that line. Sequence logos were generated using the R package ggseqlogo. To generate reference sequence logos, all known human 9mers for each of the shown HLA class I alleles were downloaded from IEDB in June 2021.

Peptide database construction

To investigate characteristics of HLA class I ligands, a database consisting of LC-MS detected peptides (i.e., true HLA ligands) and not-detected peptides (referred to as decoy peptides) was constructed. To this end, HLA binding scores to the HLA alleles of each melanoma line were calculated for all 9-, 10-, and 11-mers in the human proteome (GRCh38, gencode release 21) using netMHCpan 4.0. Processing scores were calculated using netChop 3.1. Separate databases were generated for each melanoma line by filtering on peptides derived from expressed proteins (TPM > 0 in the mRNAseq dataset), and assigning each peptide the highest affinity rank score out of the expressed HLA alleles. LC-MS detected peptides were then assigned as ‘true HLA ligands’ and the remainder of all peptides as ‘decoy peptides’. When this database was sampled for analyses, equal peptide length distributions were maintained between true HLA ligands and decoy peptides.

Feature library construction

5' UTR, coding region (CDS) and 3' UTR nucleotide sequences were downloaded from ENSEMBL BiomaRt (release 104; accessed September 2021) for all protein-coding transcripts. RNA-binding protein motifs were acquired from ATtract⁵¹ (accessed June 2021) and filtered for human RBPs (142 RBPs; 2,271 motifs). In each transcript region (e.g., 5' UTR, CDS, 3' UTR), motifs were searched and counted using a custom script (see GitHub project), and GC content and nucleotide length were computed. Also included in the sequence feature library were: Codon usage (applying coRdon⁵²), amino acid usage within the CDS, miR-DB⁵³ miRNA seed scores (accessed August 2021 and filtered for miRNA expressed immune cells, based on previous analysis by Juzenas

et al.⁵⁴), sequence homology between Human and Zebrafish (*Danio rerio*, obtained through Ensembl BiomaRt), predicted mRNA modification site occurrence per transcript region (obtained from the RMVar database⁵⁵, accessed at <https://rmvar.renlab.org/> in September 2021), and predicted post-translational modification (Acetylation, Amidation, Hydroxylation, Malonylation, Methylation, N-linked_Glycosylation, O-linked_Glycosylation, Palmitoylation, Phosphorylation, S-nitrosylation, Succinylation, Sumoylation, Ubiquitination) site occurrence (obtained from the dbPTM database⁵⁶, accessed at <https://awi.cuhk.edu.cn/dbPTM/> in June 2021).

Importance assessment of sequence feature classes

To assess the ability of sequence features to inform on HLA sampling, features belonging to five major classes (5' UTR, CDS, 3' UTR, miR binding and PTM) were extracted from the sequence feature library. The 5' UTR, CDS, 3' UTR classes were filtered based on their variance across the proteome using the `nearZeroVar` function in the `caret` R package (setting cutoffs at: `freqRatio < 500` and `percentUnique > 0.05`). All putative miR binding sites and PTMs in the library were used in the analysis. The number of features left after filtering are shown in Figure 2A. 2,000 true HLA ligands and 4,000 decoy peptides were sampled from the peptide database of each melanoma line, and subsequently used to train individual Random Forest models for each melanoma line and each feature class to predict true HLA ligands (15 models in total). The Random Forest models were generated using the R packages `randomForest` and `caret`, using 10-fold cross validation optimizing the ROC metric. Number of trees in each forest was set to 5,000 and minimum terminal node size was set to 2. The `mtry` parameter was set to . Feature importance (i.e., mean decrease in accuracy) was calculated using the `varImp` function from the R package `caret`.

Analyses examining HLA ligand enrichment potential of individual sequence features (Figure 2D-F, Supplementary Figure 2B) were focused on the 10 most important features in each class (defined as the highest mean importance score of the models trained for that feature class), and were performed using 3,389 true HLA ligands and 13,556 decoy peptides per tumor line. For the analysis presented in Figure 2E-F, a custom enrichment metric was calculated. In brief, 30% of the data was sampled and peptides were ranked either by the occurrence of a sequence feature or at random. In both cases the total number of true HLA ligands within the top 50% ranked peptides was tallied. Next, the percentage increase in true HLA ligands was calculated comparing the sequence feature ranked case versus the randomly ranked case. This process was performed for all sequence features in the analyses, and was repeated 50 times.

XGBoost classifiers

The number of experimentally detected HLA ligands from each melanoma line was down-sampled to the number of HLA ligands in the smallest dataset to ensure each melanoma line had equal weight during the analyses. The sampled data was split into a training (80%) and a test (20%) set, and these sets were supplemented with a 4-fold or 1,000-fold excess of decoy peptides. XGBoost models were generated using the R packages `xgboost` and `caret`, using 2-times 10-fold cross validation optimizing

the accuracy metric. Learning rate was set to 0.3, minimum loss reduction was set to 1.0, maximum tree depth was set to 1, sub-sampling ratio of features for each tree was set to 0.5, minimum sum of instance weight needed in a terminal leaf was set to 0.9, number of rounds was set to 1,000.

External HLA ligandome data

Transcriptomic data was accessed from the Gene Expression Omnibus (GEO) at GSE131267 and was aligned to the GRCh38 reference (gencode release 21) using Salmon (quasi-mapping mode, version 0.7.0). Mean transcript counts were calculated between replicates, and transcripts belonging to a single consensus coding sequence were summed. HLA ligands from the Sarkizova study³⁷ were downloaded from the publisher's website. This dataset was filtered for 9-meric peptides and peptides obtained from the mono-allelic cell lines expressing A2402, A0201, B3501, B5101, A1101, A3101, B4001 or B0702. To generate a decoy peptide pool, affinity binding ranks of all 9-meric peptides in the expressed proteome (TPM > 0) of the mono-allelic cell lines were calculated for each selected HLA allele. Processing scores were calculated using netChop 3.1. For each of the mono-allelic cell lines, 350 true HLA ligands and 350,000 decoy peptides were sampled.

Author contributions (CRediT taxonomy)

Kaspar Bresser: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization. Benoit P Nicolet: Conceptualization, Methodology, Formal Analysis, Writing – Review & Editing. Anita Jeko: Formal Analysis, Methodology, Investigation. Wei Wu: Data Curation. Fabricio Loayza-Puch: Investigation, Methodology, Resources. Reuven Agami: Methodology, Resources, Supervision. Albert JR Heck: Conceptualization, Methodology, Resources, Supervision, Funding Acquisition. Monika C Wolkers: Methodology, Writing – Review & Editing, Supervision. Ton N Schumacher: Conceptualization, Methodology, Writing – Review & Editing, Supervision, Funding Acquisition.

Data availability

Transcriptomic and ribosome profiling data presented in this manuscript have been deposited to GEO and can be accessed under the accession number GSE211000. Mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD036277. All statistical source data of the figures presented in this study are provided with this paper. Transcriptomic data of the Sarkizova study³⁷ was accessed from GEO under the accession number GSE131267. HLA ligands from the Sarkizova study³⁷ were downloaded from the publisher's website. Any additional data supporting the findings of this study are available from the corresponding authors upon request.

R scripts that were used to produce the main and extended data figures in the manuscript are available from GitHub (https://github.com/kasbress/HLA_ligandome_feature).

References

1. Tran, E., Robbins, P. F. & Rosenberg, S. A. 'Final common pathway' of human cancer immunotherapy: targeting random somatic mutations. *Nat Immunol* **18**, 255–262 (2017).
2. Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).
3. Wells, D. K. *et al.* Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. *Cell* **183**, 818–834.e13 (2020).
4. Antoniou, A. N., Powis, S. J. & Elliott, T. Assembly and export of MHC class I peptide ligands. *Curr Opin Immunol* **15**, 75–81 (2003).
5. Kloetzel, P. M. Antigen processing by the proteasome. *Nat Rev Mol Cell Biol* **2**, 179–187 (2001).
6. Rao, X., Costa, A. I. C. A. F., van Baarle, D. & Kesmir, C. A comparative study of HLA binding affinity and ligand diversity: implications for generating immunodominant CD8+ T cell responses. *J Immunol* **182**, 1526–1532 (2009).
7. Nielsen, M. *et al.* Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* **12**, 1007–1017 (2003).
8. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**, 511–517 (2016).
9. Kisselev, A. F., Callard, A. & Goldberg, A. L. Importance of the different proteolytic sites of the proteasome and the efficacy of inhibitors varies with the protein substrate. *J Biol Chem* **281**, 8582–8590 (2006).
10. Gomez-Perosanz, M., Ras-Carmona, A., Lafuente, E. M. & Reche, P. A. Identification of CD8+ T cell epitopes through proteasome cleavage site predictions. *BMC Bioinformatics* **21**, 484 (2020).
11. Nielsen, M., Lundegaard, C., Lund, O. & Kesmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57**, 33–41 (2005).
12. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J. & Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics* **14**, 658–673 (2015).
13. Juncker, A. S. *et al.* Systematic characterisation of cellular localisation and expression profiles of proteins containing MHC ligands. *PLoS One* **4**, e7448 (2009).
14. Milner, E., Barnea, E., Beer, I. & Admon, A. The turnover kinetics of major histocompatibility complex peptides of human cancer cells. *Mol Cell Proteomics* **5**, 357–365 (2006).
15. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* **21**, 630–644 (2020).
16. Cuadrado, E. *et al.* Proteomic Analyses of Human Regulatory T Cells Reveal Adaptations in Signaling Pathways that Protect Cellular Identity. *Immunity* **48**, 1046–1059.e6 (2018).
17. Nicolet, B. P. & Wolkers, M. C. *Limited but gene-class specific correlation of mRNA and protein expression in human CD8+ T cells.* <http://biorxiv.org/lookup/doi/10.1101/2020.04.21.053884> (2020) doi:10.1101/2020.04.21.053884.
18. Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* **7**, 548 (2011).
19. Vogel, C. *et al.* Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol Syst Biol* **6**, 400 (2010).
20. Komander, D. & Rape, M. The ubiquitin code. *Annu Rev Biochem* **81**, 203–229 (2012).
21. Narita, T., Weinert, B. T. & Choudhary, C. Functions and mechanisms of non-histone protein acetylation. *Nat Rev Mol Cell Biol* **20**, 156–174 (2019).
22. Vita, R. *et al.* The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res* **47**, D339–D343 (2019).
23. Andreatta, M., Alvarez, B. & Nielsen, M. GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res* **45**, W458–W463 (2017).
24. Nicolet, B. P., Zandhuis, N. D., Lattanzio, V. M. & Wolkers, M. C. Sequence determinants as key regulators in gene expression of T cells. *Immunol Rev* **304**, 10–29 (2021).
25. Li, X. *et al.* Base-Resolution Mapping Reveals Distinct m1A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol Cell* **68**, 993–1005.e9 (2017).
26. Malbec, L. *et al.* Dynamic methylome of internal mRNA N7-methylguanosine and its regulatory role in translation. *Cell Res* **29**, 927–941 (2019).
27. Karagkouni, D. *et al.* DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Research* **46**, D239–D245 (2018).
28. Caron, C., Boyault, C. & Khochbin, S. Regulatory cross-talk between lysine acetylation and ubiquitination: role in the control of protein stability. *Bioessays* **27**, 408–415 (2005).
29. Jeong, J. W. *et al.* Regulation and destabilization of HIF-1alpha by ARD1-mediated acetylation. *Cell* **111**,

- 709–720 (2002).
30. Wilkinson, K. D. *et al.* Metabolism of the polyubiquitin degradation signal: structure, mechanism, and role of isopeptidase T. *Biochemistry* **34**, 14535–14546 (1995).
 31. Mirihana Arachchilage, G., Hetti Arachchilage, M., Venkataraman, A., Piontkivska, H. & Basu, S. Stable G-quadruplex enabling sequences are selected against by the context-dependent codon bias. *Gene* **696**, 149–161 (2019).
 32. Rao, Y. S., Wang, Z. F., Chai, X. W., Nie, Q. H. & Zhang, X. Q. Relationship between 5' UTR length and gene expression pattern in chicken. *Genetica* **141**, 311–318 (2013).
 33. Zhang, L.-S. *et al.* Transcriptome-wide Mapping of Internal N7-Methylguanosine Methylome in Mammalian mRNA. *Mol Cell* **74**, 1304–1316.e8 (2019).
 34. Courel, M. *et al.* GC content shapes mRNA storage and decay in human cells. *Elife* **8**, e49708 (2019).
 35. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016). doi:10.1145/2939672.2939785.
 36. Abelin, J. G. *et al.* Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity* **46**, 315–326 (2017).
 37. Sarkizova, S. *et al.* A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat Biotechnol* **38**, 199–209 (2020).
 38. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
 39. Blevins, W. R. *et al.* Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. *Sci Rep* **9**, 11005 (2019).
 40. Solberg, O. D. *et al.* Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum Immunol* **69**, 443–464 (2008).
 41. Wang, X. *et al.* N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388–1399 (2015).
 42. Ravid, T. & Hochstrasser, M. Diversity of degradation signals in the ubiquitin-proteasome system. *Nat Rev Mol Cell Biol* **9**, 679–690 (2008).
 43. Pang, C. N. I., Gasteiger, E. & Wilkins, M. R. Identification of arginine- and lysine-methylation in the proteome of *Saccharomyces cerevisiae* and its functional implications. *BMC Genomics* **11**, 92 (2010).
 44. Lee, J. M. *et al.* EZH2 generates a methyl degron that is recognized by the DCAF1/DDB1/CUL4 E3 ubiquitin ligase complex. *Mol Cell* **48**, 572–586 (2012).
 45. Jarchum, I. Putting a number on neoepitope quality. *Nat Biotechnol* **36**, 151 (2018).
 46. The problem with neoantigen prediction. *Nat Biotechnol* **35**, 97 (2017).
 47. Kemper, K. *et al.* BRAF(V600E) Kinase Domain Duplication Identified in Therapy-Refractory Melanoma Patient-Derived Xenografts. *Cell Rep* **16**, 263–277 (2016).
 48. Kelderman, S. *et al.* Antigen-specific TIL therapy for melanoma: A flexible platform for personalized cancer immunotherapy. *Eur J Immunol* **46**, 1351–1360 (2016).
 49. Marino, F. *et al.* Arginine (Di)methylated Human Leukocyte Antigen Class I Peptides Are Favorably Presented by HLA-B*07. *J Proteome Res* **16**, 34–44 (2017).
 50. Wang, H., McManus, J. & Kingsford, C. Isoform-level ribosome occupancy estimation guided by transcript abundance with Ribomap. *Bioinformatics* **32**, 1880–1882 (2016).
 51. Giudice, G., Sánchez-Cabo, F., Torroja, C. & Lara-Pezzi, E. ATTRACT—a database of RNA-binding proteins and associated motifs. *Database* **2016**, baw035 (2016).
 52. Elek, A., Kuzman, M. & Vlahovicek, K. coRdon: Codon Usage Analysis and Prediction of Gene Expressivity. (2018).
 53. Chen, Y. & Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Research* **48**, D127–D131 (2020).
 54. Juzenas, S. *et al.* A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Research* **45**, 9290–9301 (2017).
 55. Luo, X. *et al.* RMVar: an updated database of functional variants involved in RNA modifications. *Nucleic Acids Research* **49**, D1405–D1412 (2021).
 56. Huang, K.-Y. *et al.* dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. *Nucleic Acids Res* **47**, D298–D308 (2019).

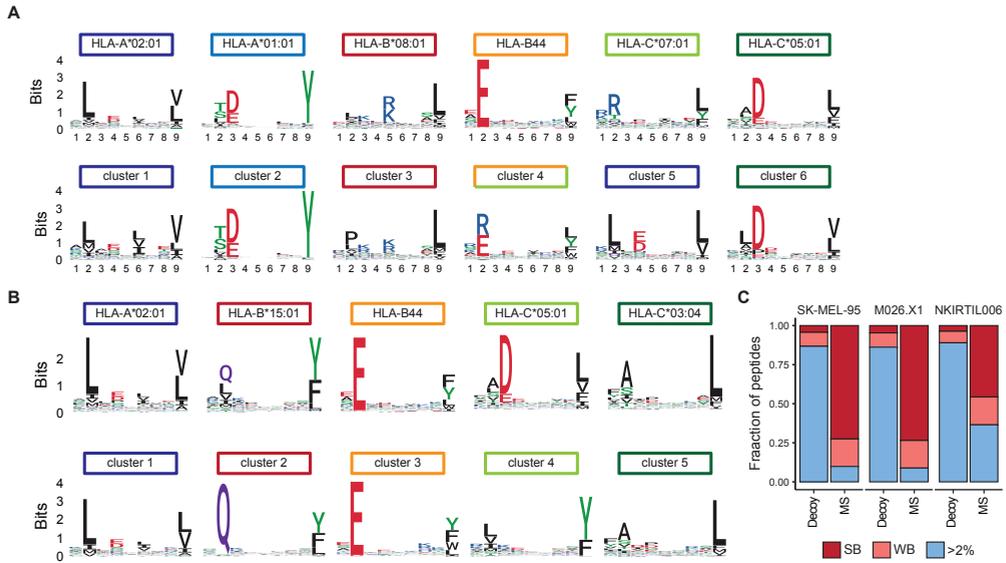


Figure S1. Identification of HLA class I ligandomes (related to Figure 1). (A–B) Sequence logos of the HLA class I alleles (top panels) expressed by M026.X1 (A) and NKIRTL006 (B), and the sequence logos of peptide clusters obtained by the GibbsCluster algorithm (bottom panels). The number of clusters was constrained to the number of expressed HLA class I alleles. (C) Fraction of peptides predicted to have a <math><0.5</math> (strong binders; SB), <math><2.0</math> (weak binders; WB), or >2.0 percentile rank binding affinity within LC-MS detected peptides and decoy peptides, as determined by netMHCpan4-0.

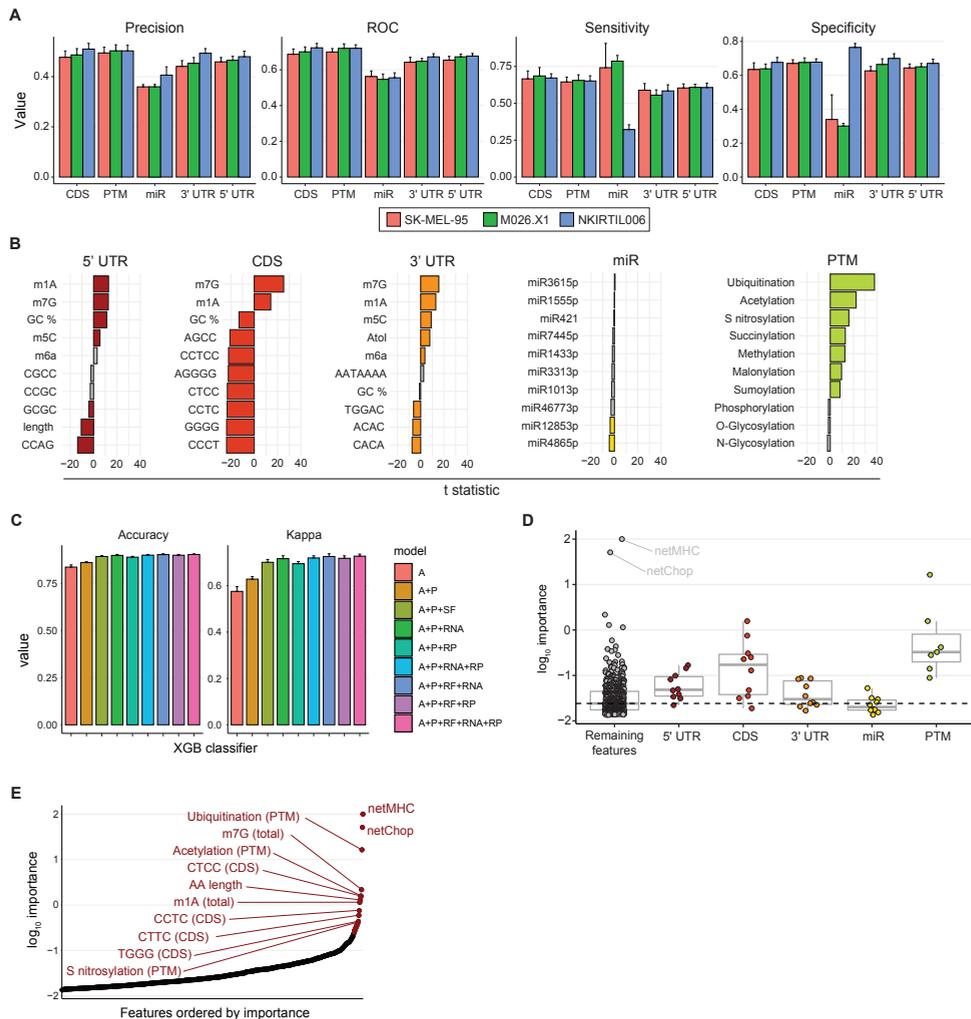


Figure S2. Sequence features inform on HLA sampling (related to Figure 2). (A) Random forest models were trained using data from each melanoma line to classify HLA ligands, using individual classes of gene and protein sequence features. Bar graphs indicate out-of-bag model performance. (B) t statistics obtained comparing counts of indicated sequence features between LC-MS detected and decoy peptides. Top 10 features with the highest importance to each random forest model are shown. (C) Out-of-bag model performance of XGBoost classifiers generated in Figure 3. (D) Feature importance of the top 10 features in each class identified in Figure 2E, compared to the feature importance of the remaining features in the library. Dashed line indicated the median importance of all features in the A+P+SF XGBoost model. Boxplots indicate group median and 25th and 75th percentiles, whiskers indicate the interquartile range multiplied by 1.5, and dots signify individual features. (E) Importance of features to the A+P+SF XGBoost model. The top 12 features are highlighted.

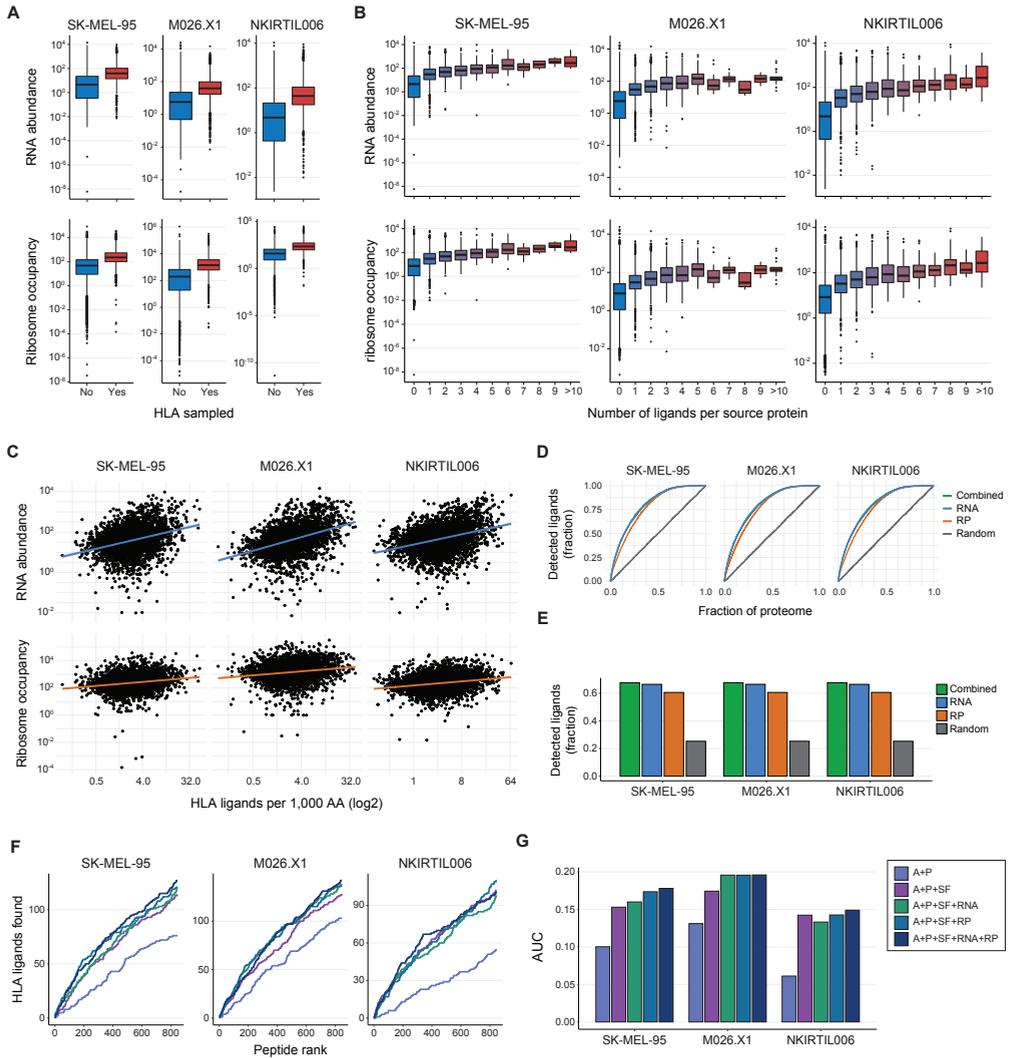
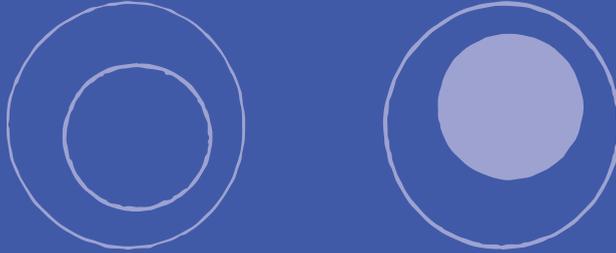


Figure S3. Association of HLA ligandome composition with RNA abundance and ribosome occupancy. (A) RNA abundance (top) and ribosome occupancy (bottom) of proteins for which either HLA ligands were or were not detected by LC-MS. (B) RNA abundance (top) and ribosome occupancy (bottom) of source proteins, binned by the number of HLA ligands detected by LC-MS. (C) HLA sampling density of each source protein, calculated as the number of detected HLA ligands per 1,000 amino acids, plotted against their respective RNA abundance (top) and ribosome occupancy (bottom). (D-E) Predictiveness of the assessed expression metrics. Source proteins were ranked by either RNA abundance, ribosome occupancy or by a randomly generated metric (obtained by shuffling RNA abundance data). In addition, a combined ranking was obtained by averaging the rankings of RNA abundance and ribosome occupancy metrics. Line plots (D) depict the fraction of detected HLA ligands from that melanoma line as a function of the fraction of the analyzed proteome (cumulative protein length). Bar charts (E) depict the fraction of HLA ligands observed within the top quartile of the proteome (cumulative protein length). (F-G) Number of HLA ligands observed in the top 0.1% of predicted peptides from the melanoma line test sets by each of the indicated models. Line graphs depicting the cumulative sum (F) and bar charts depicting the AUCs (G) are shown.





Chapter 4. Replicative history marks transcriptional and functional disparity in the CD8+ T cell memory pool

OR: "Building a unnecessarily complex machine
to approximate if a cell is sleeping or not."



Kaspar Bresser^{1,*}, Lianne Kok^{1,*}, Arpit C. Swain², Lisa A. King^{1,3}, Laura Jacobs¹, Tom S. Weber^{4,5}, Leïla Perié⁶, Ken R. Duffy⁷, Rob J. de Boer², Ferenc A. Scheeren⁸ and Ton N. Schumacher^{1,9}

Published in Nature Immunology, May 2022

(1) Division of Molecular Oncology & Immunology, Oncode Institute, The Netherlands Cancer Institute, Amsterdam, The Netherlands.

(2) Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands.

(3) Present address: Department of Medical Oncology, Amsterdam UMC, Vrije Universiteit Amsterdam, Cancer Center Amsterdam, Amsterdam, The Netherlands.

(4) The Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria, Australia.

(5) The Department of Medical Biology, The University of Melbourne, Melbourne, Victoria, Australia.

(6) Institut Curie, Université PSL, Sorbonne Université, CNRS UMR168, Laboratoire Physico Chimie Curie, Paris, France.

(7) Hamilton Institute, Maynooth University, Maynooth, Ireland.

(8) Department of Dermatology, Leiden University Medical Center, Leiden, Netherlands

(9) Department of Hematology, Leiden University Medical Center, Leiden, The Netherlands

(* These authors contributed equally to this work

Abstract

Clonal expansion is a core aspect of T cell immunity. However, little is known with respect to the relationship between replicative history and the formation of distinct CD8⁺ memory T cell subgroups. To address this issue, we developed a genetic-tracing approach, termed the DivisionRecorder, that reports the extent of past proliferation of cell pools in vivo. Using this system to genetically ‘record’ the replicative history of different CD8⁺ T cell populations throughout a pathogen-specific immune response, we demonstrate that the central memory T cell (T_{CM}) pool is marked by a higher number of prior divisions than the effector memory T cell pool, due to the combination of strong proliferative activity during the acute immune response and selective proliferative activity after pathogen clearance. Furthermore, by combining DivisionRecorder analysis with single cell transcriptomics and functional experiments, we show that replicative history identifies distinct cell pools within the T_{CM} compartment. Specifically, we demonstrate that lowly divided T_{CM} display enriched expression of stem-cell-associated genes, exist in a relatively quiescent state, and are superior in eliciting a proliferative recall response upon activation. These data provide the first evidence that a stem cell like memory T cell pool that reconstitutes the CD8⁺ T cell effector pool upon reinfection is marked by prior quiescence.

Introduction

The CD8⁺ T cell compartment serves to provide protection against intracellular pathogens and also acts as a modifier of cancer growth. Upon antigen encounter, naïve T cells (T_N) undergo extensive gene-expression alterations, while entering a highly proliferative state, dividing every 4h to 6h^{1,2} in mice. This phase of clonal expansion gives rise to a phenotypically and functionally diverse pool of effector T cells (T_{EFF}) that exceeds its precursor population size by >10,000-fold^{3,4}. Unlike T_N , these T_{EFF} have the capacity to disseminate to peripheral tissues, and scan for and kill infected or transformed cells. Upon antigen clearance, around 95% of the T_{EFF} pool succumbs to apoptosis, leaving behind a small long-lived pool of memory T cells (T_M) that is equipped to provide long-term protection against recurring pathogens.

The central role of proliferation in the T cell response has inspired many to study the relationship between replication and T cell state. While earlier work hinted that memory precursor T cells have undergone limited clonal expansion^{5,6}, more recent work studying acute T cell responses in human subjects demonstrated that T_M , as a whole, are derived from precursor cells that have undergone an extensive number of divisions⁷. Furthermore, prior work has shown that cell cycle speed can differ substantially between phenotypically distinct T cell subsets at different time-points in the T cell response. Specifically, central memory T cells (T_{CM}), a subgroup of memory cells that are endowed

with a high level of multipotency, have been documented to undergo homeostatic proliferation after pathogen clearance, while effector memory T cells (T_{EM}) have a low turnover rate^{8,9}. In contrast, during the effector phase, a T_{CM} -like state has been linked to lower division speed and reduced clonal burst size compared to their T_{EM} -like and terminally differentiated counterparts¹⁰⁻¹³.

The phase-dependent association of proliferative activity within specific cell states, in combination with the reported phenotypic instability of certain T cell subsets^{14,15}, makes it difficult to deduce the replicative history (i.e., the cumulative number of prior divisions) of different memory T cell populations, and the possible relationship between such replicative history and functional properties. Here, we develop a genetic-tracing approach—termed DivisionRecorder—that allows the measurement of prior division of cell pools over extensive rounds of division, and apply this approach to determine to what extent replicative history identifies distinct memory T cell states and behaviors. In this effort, we focus on three central issues: (1) What are the differences in replicative history between (precursor-) T_{CM} and T_{EM} in the effector and memory phase? (2) Is there heterogeneity in prior division within the T_{CM} pool? (3) If so, does replicative history of cells within the T_{CM} pool predict their capacity to mount a secondary T cell response?

Results

Division-linked genetic labeling of cell pools

The genome contains a large number of hypervariable short tandem nucleotide repeats (STRs) that accumulate intra-allelic length mutations through DNA polymerase slippage during cell division. Such slippage mutations in endogenous STRs have been used to study lineage trees in various organisms and tissues^{16,17}, and synthetic STRs have previously been employed in a probabilistic labeling approach to define stem cells in the intestinal epithelium and the mammary gland^{18,19}. To investigate the replicative history of memory T cells, we engineered a synthetic STR-reporter system to continuously ‘record’ proliferation in cell pools. This genetically encoded system, termed DivisionRecorder, utilizes a synthetic STR domain to achieve a division-linked low-probability acquisition of a fluorescent mark (**Fig. 1a**). The DivisionRecorder consists of two separate elements: (1) a retroviral-vector encoded module that contains a synthetic STR linked to an out-of-frame CRE recombinase gene; (2) A CRE-activity reporter module that irreversibly induces the expression of a red fluorescent protein (RFP). In its base configuration, all cells that contain the DivisionRecorder only express GFP (hereafter referred to as DR^{GFP} cells). As cells undergo successive divisions, slippage mutations that occur within the synthetic STR yield in-frame variants of the downstream CRE recombinase gene at a fixed, division-dependent, probability (p). The resulting CRE activity induces an irreversible activation of the RFP gene, giving rise to GFP^+RFP^+ cells (hereafter referred to as DR^{RFP}) that pass this genetically encoded label on to subsequent generations, resulting in a cumulative increase in the DR^{RFP} cell fraction within the DivisionRecorder⁺ (DR^+ , i.e., the sum of DR^{GFP} and DR^{RFP}) population as the cell pool expands (**Fig. 1b, Supplementary Note 1**). Importantly, when p is small (< 0.01) the DivisionRecorder yields a near-linear relationship between

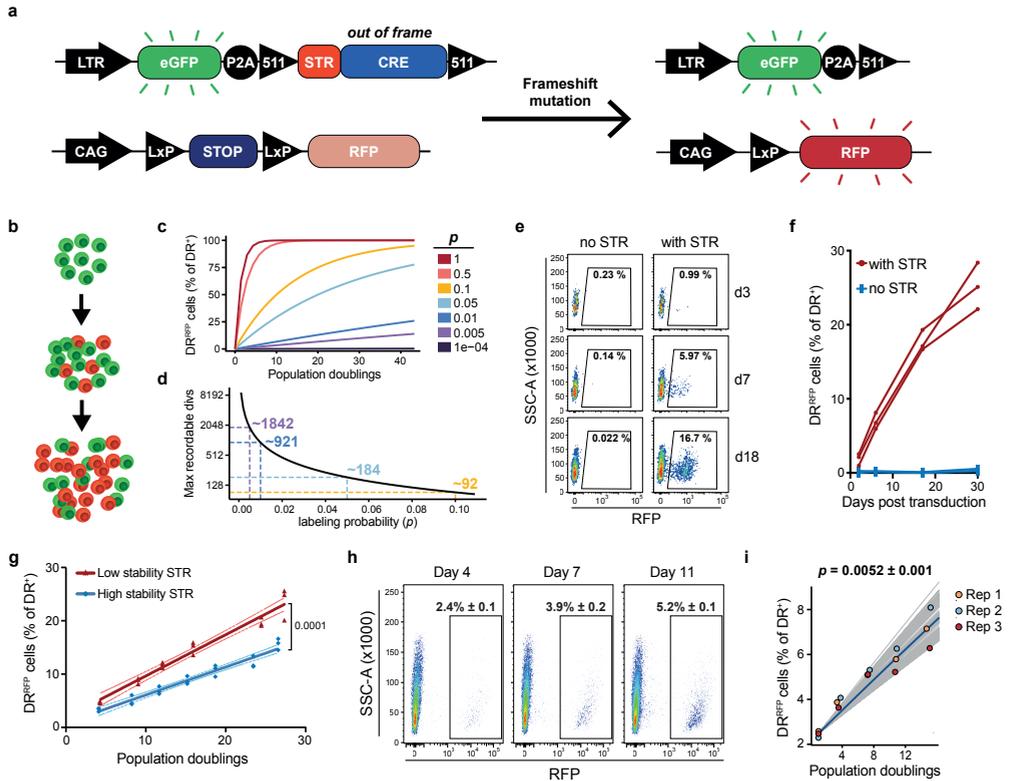


Fig. 1. DivisionRecorder activation is a proxy for replicative history. **a**, Schematic overview of the DivisionRecorder system. **b**, Cartoon depicting progressive DivisionRecorder activation in a proliferating cell pool. **c**, Simulation of the minimal ODE model (See **Supplementary Note 2** for detailed description and equations), depicting DR^{RFP} acquisition as a function of population doublings for the indicated values of DR^{RFP} acquisition probability (p). **d**, Maximal number of theoretically recordable population doublings, approximated by calculating the amount of division events required to reach a 99% DR^{RFP} population. Approximate maximums for selected values of p are indicated, colors correspond to legend in panel **c**. **e-f**, Percentage of DR^{RFP} cells over time in cultured DivisionRecorder⁺ (DR^+) CRE-activity reporter HEK 293T cells ($n=3$ replicates per group) in which the CRE recombinase gene was preceded by either a stable nucleotide region (indicated as “no STR”) or a repeat of 24 guanines (indicated as “with STR”). Representative plots (**e**) and summarizing line graphs (**f**) are shown. Lines connect experimental replicates **g**, Percentage of DR^{RFP} cells across population doublings in DR^+ CRE-activity reporter HEK 293T cells ($n=3$ replicates per group) in which the CRE recombinase gene was preceded by either a low stability STR ([G]24) or a high stability STR ([CA]30). Dots indicate individual samples, lines represent fitted linear regression, dotted lines indicate bounds of the 95% confidence interval. **h-i**, Percentage of DR^{RFP} cells across population doublings in immortalized DR^+ mouse embryonic fibroblasts. Representative flow cytometry plots (**h**) and summarizing graph (**i**) are shown. Best fits of the minimal ODE model are depicted (100 bootstraps per experimental replicate, **Supplementary Note 2**). Blue line represents the median of the bootstraps, grey lines represent individual fits, dots indicate experimental measurements ($n=3$ replicates). p indicates the estimated DR^{RFP} acquisition probability. Depicted experimental data are representative of at least two independent experiments. P values (**g**) were determined by two-sided ANCOVA test.

the DR^{RFP} fraction and the average number of divisions over dozens of population doublings (**Fig. 1c**)²⁰, thereby allowing analysis of replicative history—at the population level—far beyond what can be achieved with classical cell labeling dyes²¹ (**Fig. 1d**).

To test the utility of the DivisionRecorder, we established a reporter cell-line carrying a lox-STOP-lox-RFP cassette. Following retroviral introduction of the GFP-STR-CRE module, a progressive increase in DR^{RFP} cells was observed over time, whereas no label acquisition was observed when the

STR was replaced with a stable DNA sequence (**Fig 1e, f**). Moreover, the rate at which DR^{RFP} cells accumulated was dependent on the sequence stability of the STR^{22,23}, underpinning that p is linked to the likelihood of STR slippage (**Fig 1g**). Similarly, upon introduction of the DivisionRecorder into immortalized embryonic fibroblasts from the Ai9 mouse strain—that carry an endogenous lox-STOP-lox-RFP cassette²⁴—a low and predictable DR^{RFP} acquisition was observed, with a [G]33 STR conferring a p of 0.0052 ± 0.00074 (**Fig. 1h, i**), thereby enabling the measurement of replicative history over many cell divisions (in theory >1,500 population doublings, **Fig. 1d**).

To test whether the DivisionRecorder can be used as a proxy for replicative history in the CD8⁺ T cell compartment in vivo, we generated Ai9;OT-I mice, in which all T cells recognize the OVA₂₅₇₋₂₆₄ epitope, thereby allowing examination of T cell pools in the context of equal TCR affinity. Ai9;OT-I T cells were isolated, modified with the DivisionRecorder to obtain DR⁺ OT-I T cells, transferred into *Listeria monocytogenes*-OVA (*Lm*-OVA) infected mice, and the fraction DR^{RFP} cells was measured over time (**Fig. 2a**). At early time-points post cell transfer (d1-d4), a rapid increase in DR^{RFP} cells was observed (**Fig. 2b, c**), coinciding with the proliferative burst of the antigen-specific CD8⁺ T cell pool. To determine whether the observed accumulation of DR^{RFP} cells formed an accurate measure of prior cell division, DR⁺ OT-I T cells were stained with CellTrace Violet (CTV) prior to cell

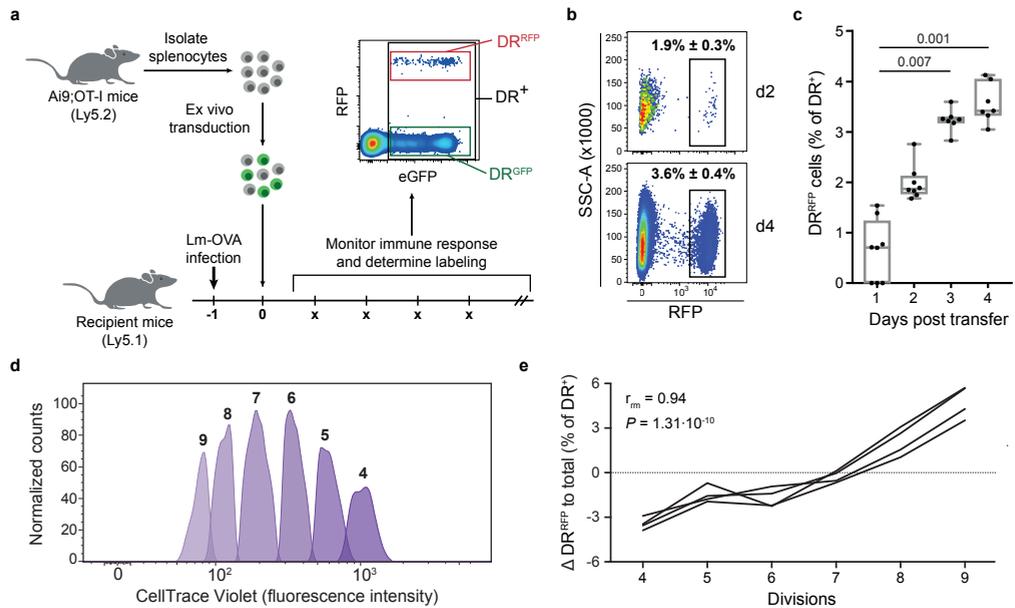


Fig. 2. The DivisionRecorder can be applied to study T cell division kinetics in vivo. **a**, Overview of experimental setup. **b-c**, DR⁺ OT-I T cells were transferred into recipient mice 24 hours post infection with *Lm*-OVA. Spleen samples were analyzed for the percentage of DR^{RFP} cells at day 1-4 post cell transfer. Representative pseudo-color density plots (**b**), and boxplots (**c**) in which the boxes indicate group median and 25th/75th percentiles, whiskers represent min/max, dots represent individual samples ($n=8$ mice for day 1 and 2; $n=7$ mice for day 3 and 4). **d-e**, CTV-stained OT-I T cells were retrovirally transduced with the DivisionRecorder and transferred into recipient mice ($n=4$) 24 hours post infection with *Lm*-OVA. 48 hours post-transfer, splenic DR⁺ OT-I T cells were assessed for CTV dilution (**d**), and the percentage of DR^{RFP} cells within each division peak was analyzed (**e**). All depicted data are representative of at least two independent experiments, lines and symbols indicate individual mice or samples. P values were determined by two-sided Kruskal-Wallis test, with Dunn's multiple comparisons test (**c**), or two-sided repeated measurement correlation test (**e**).

transfer. Notably, analysis of the fraction DR^{RFP} cells within cell pools with different degrees of CTV dilution revealed a close correlation (**Fig. 2d, e**, $r_{\text{fm}} = 0.94$), providing direct evidence that in vivo DR^{RFP} acquisition reflects the extent of past division in the CD8⁺ T cell pool. In conclusion, these data establish that the DivisionRecorder allows the long-term measurement of division history in cell pools in vivo, in a way that is compatible with down-stream methodologies such as single cell sequencing (see below).

CD8⁺ T_{CM} cells are derived from replicative mature T cells

Having validated the utility of the DivisionRecorder to record T cell division, we next sought to determine the replicative history of the total CD8⁺ T_M pool relative to that of the T_{EFF} pool. Analysis of the size of the DR⁺ OT-I T cell compartment in blood following *Lm*-OVA infection showed the characteristic rapid expansion phase, with T cell numbers peaking around day 6, and subsequent contraction into a stable memory pool (**Fig. 3a**). Notably, DR^{RFP} cells remained detectable following formation of T cell memory, thus allowing analysis of replicative history at late time points after infection (**Fig. 3b**).

In case T_M would primarily be derived from T cells that had undergone limited proliferation upon primary antigen encounter, the fraction of DR^{RFP} cells would be expected to decay during the contraction phase, due to the decline in the number of clonally expanded T_{EFF} (**Extended Data Fig. 1, Supplementary Note 3**). However, analysis of DR^{RFP} frequencies in blood demonstrated that the fraction of DR^{RFP} cells did not decline, but instead continued to increase during contraction and memory phase (an increase of 2.07% ± 0.77% between day 13 and 59, **Fig. 3c**). This increase in DR^{RFP} frequencies post pathogen-clearance was not restricted to T cell responses induced by *Lm*-OVA infection, but was also observed upon infection with LCMV-OVA²⁵ (**Fig. 3d**), and was not due to anatomical redistribution of cells with distinct division histories, as the fraction of DR^{RFP} cells increased concurrently in peripheral blood and the primary sites of *Lm*-OVA infection (spleen/liver; **Fig. 3e, f**). Thus, in line with work by Akondy *et al.*⁷, our results support the notion of a replicative ‘mature’, rather than ‘nascent’, CD8⁺ T_M pool, and extends this observation beyond the peripheral blood compartment to the sites of infection.

It has been well documented that T_{CM} are able to maintain the memory pool through infrequent homeostatic cell division^{15,26,27}, and recent work has shown that precursor-T_{CM} slow down their replicative cycle early during the expansion phase¹⁰, suggesting limited clonal expansion of these cells during the early phase of the T cell response. However, it is difficult to translate cell-cycle activity at a given time-point into cumulative proliferative history, and we therefore wished to directly test the relationship between cell state (e.g., T_{CM} or T_{EM}) and replicative history during different stages of the T cell response. To this end, the fraction of DR^{RFP} cells within the T_M pool was calculated at varying expression levels of proteins associated with either multipotency or terminal differentiation (**Fig. 3g**). This analysis revealed a positive correlation between replicative history and the expression of the T_{CM}-associated proteins CD27 ($r_{\text{fm}} = 0.81$, $P = 6.2 \cdot 10^{-14}$) and CD62L ($r_{\text{fm}} = 0.62$, $P = 5.6 \cdot 10^{-7}$)^{15,28,29}, and a negative relationship between prior division and the expression

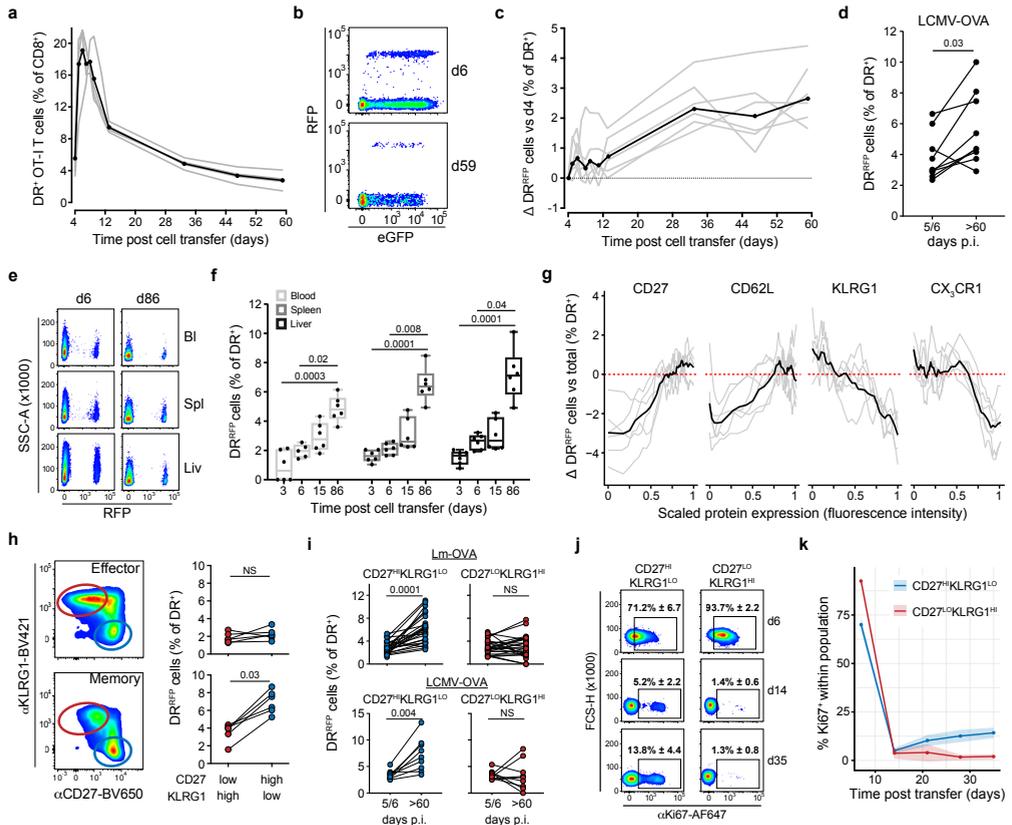


Fig. 3. The multipotent memory T cell pool is formed by replicative ‘mature’ cells. **a-c**, Kinetics of DR⁺ OT-I T cells (**a**) and the percentage of DR^{REP} relative to day 4 (**c**) in response to *Lm*-OVA, measured in peripheral blood (n=6 mice). Representative flow cytometry plots (**b**) showing DR^{REP} and DR^{GFP} frequencies at indicated time points, and line graphs (**a**, **c**) depicting kinetics of single mice (grey) and group median (black). **d**, DR^{REP} percentages within blood at day 5/6 (T_{EFF}) and day >60 (T_M) following LCMV-OVA infection (n=7). **e**, Representative plots depicting DR^{REP} frequencies in blood (Bl), spleen (Spl) and liver (Liv). **f**, Percentage of DR^{REP} detected in indicated organs of recipient mice at the indicated time points (n=6 mice per time point; response to *Lm*-OVA). Boxplots indicate group median and 25th/75th percentiles, whiskers represent min/max, dots represent individual samples. **g**, Moving average of surface marker expression level on splenic DR⁺ cells plotted against the percentage of DR^{REP} within each window during memory (day 86; n=6), means are shown in black. DR^{REP} percentages within each window are corrected for the total percentage of DR^{REP} detected in that sample. **h**, Gating strategy (left) and DR^{REP} percentages (right) of CD27^{hi}KLRG1^{lo} and CD27^{lo}KLRG1^{hi} cells in spleen during effector (d6, top) and memory phase (d86, bottom; n=6) in response to *Lm*-OVA. **i**, DR^{REP} percentages within the CD27^{hi}KLRG1^{lo} and CD27^{lo}KLRG1^{hi} cell populations in blood, comparing effector (day 5/6) and memory (day >60) phases. Data shown for *Lm*-OVA (top; n=22) and LCMV-OVA (bottom; n=7) infections. Lines connect individual mice. **j-k**, Ki67 expression by CD27^{hi}KLRG1^{lo} and CD27^{lo}KLRG1^{hi} OT-I cells in blood in response to *Lm*-OVA. Representative flow cytometry plot (**j**), and line graphs (**k**) where solid lines indicate population means, shaded areas indicate 95% confidence interval (n=11 mice). All depicted data are representative of at least two independent experiments, lines and symbols indicate individual mice or samples. *P* values were determined by two-sided Kruskal-Wallis test with Dunn’s multiple comparisons test (**f**), or two-sided Wilcoxon’s signed-rank test (**d**, **h**, **i**).

of the T_{EM}-associated proteins KLRG1 ($r_{\text{mm}} = -0.83$, $P = 9.0 \cdot 10^{-15}$) and CX₃CR1 ($r_{\text{mm}} = -0.75$, $P = 4.5 \cdot 10^{-11}$)^{14,15,30}. Likewise, defining multipotent T_{CM} and terminally differentiated T_{EM} subsets by joint expression or absence of CD62L and CD27, respectively, (**Extended Data Fig. 2a**), and further partitioning based on the expression of KLRG1 or CX₃CR1, revealed a positive association between division history and a less differentiated cell state (**Extended Data Fig. 2b**). Furthermore,

the division history of CD27^{HI}KLRG1^{LO} T_{CM} present in lymph nodes equaled that of T_{CM} in the spleen, implying that division history is dictated by cell state rather than anatomical location (**Extended Data Fig. 2c**).

Next, to delineate at which point the divergence in replicative history between T cells with a T_{CM}-like multipotent and T_{EM}-like terminally differentiated phenotype developed, we assessed the link between phenotypic marker expression and DR^{RFP} fractions throughout the T cell response. Notably, replicative history varied minimally across T_{EFF} cell states at the peak of the antigen-specific T cell response (d6 post transfer, **Fig. 3h**, **Extended Data Fig. 2d-f**), followed by selective accumulation of DR^{RFP} within the CD27^{HI}KLRG1^{LO} early-T_{CM} pool directly after the peak of the expansion phase (**Fig. 3h, i**, **Extended Data Fig. 1g**), due to continued replicative activity of this subset (**Fig. 3j, k**). The observation that the division history of CD27^{LO}KLRG1^{HI} T cells stays constant post effector phase (**Fig. 3i**) suggests that, in addition to the previously documented lack of proliferative activity of this cell pool^{15,26,31}, this terminally differentiated subset also does not receive significant replenishment by the replicative active CD27^{HI}KLRG1^{LO} T cell pool (**Extended Data Fig. 1h**). The substantial number of divisions that we observe in the CD27^{HI}KLRG1^{LO} cell pool at the peak of the response appears at odds with the proposed limited clonal expansion of precursor-T_M. However, these observations may either be reconciled by the reported trans-differentiation between T_{EFF} cell states^{14,15,30}, or by the fact that a reduced proliferative activity may form a property of only a small part of the memory precursor pool^{10,11,32}. In summary, the above data indicate that the high amount of prior division of the T_{CM} pool results from both strong proliferative activity during the effector phase and selective proliferative activity after pathogen clearance.

Replicative history identifies distinct T_{CM} cell states

Increasing evidence suggests that the T_{CM} pool is highly heterogeneous in terms of both gene expression profiles and prior and ongoing replicative behavior^{14,15,33}, providing an incentive to test for possible associations between division history and transcriptional states within this cell pool. To this end, we carried out single-cell mRNA sequencing (scRNAseq) on DR^{GFP} and DR^{RFP} memory OT-I T cells (75-85 days post *Lm*-OVA infection; **Extended Data Fig. 3**). In addition, to test whether DR⁺ OT-I T_M assume the same spectrum of transcriptional states as non-modified T cells, we performed scRNAseq on OT-I T_M that were generated through adoptive transfer of a small number (2,000) of naïve OT-I T cells followed by *Lm*-OVA infection 24 hours later. DR⁺ OT-I and unmodified OT-I memory T cells were jointly grouped into 23 transcriptionally distinct MetaCells (MCs)³⁴ that included 4 T_{EM} and 19 T_{CM} MCs based on the expression of a small set of multipotency- and effector-associated genes (**Fig. 4a,b**). Notably, while memory T cells derived from small numbers of unmodified OT-I T cells showed a proportionally greater contribution to T_{EM} MCs—consistent with the relationship between precursor frequency and T_{EM} formation³⁵—DR⁺ OT-I T cells and unmodified OT-I T cells were equal in their potential to yield the 19 distinct T_{CM} MCs (**Extended Data Fig. 4**), indicating that the introduction of the DivisionRecorder did not measurably impact the ability of T cells to differentiate into different T_{CM} states.

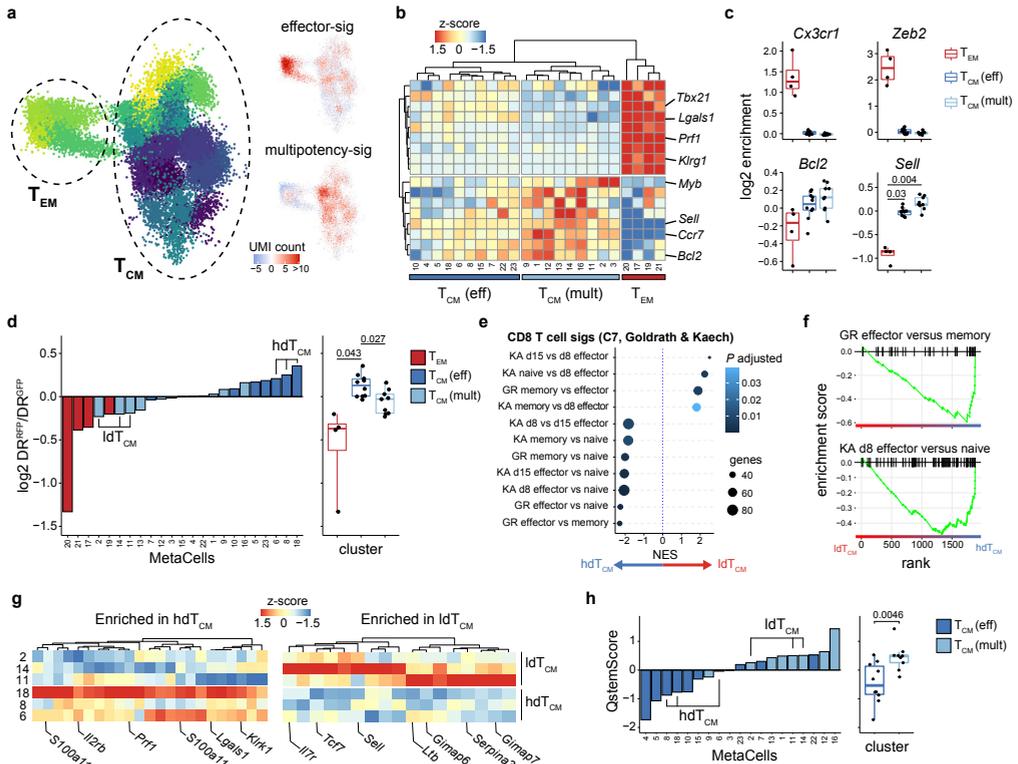


Fig. 4 Replicative history identifies distinct transcriptional states within the T_{CM} pool. Single cell transcriptomic profiling of DR⁺ T cells obtained from spleen in memory phase (Day 75 and 85 post *Lm*-OVA infection). **a**, 2D projection of all profiled cells, colors indicate MCs (left), or relative expression of effector- and multipotency-associated genes (right). Gene list in **Supplementary Table 1**. **b**, Hierarchical clustering of MCs by their expression of effector- and multipotency-associated genes used in A. MCs are divided into 3 clusters based on Euclidean distance. **c**, Expression of selected genes by each MC cluster. **d**, DR^{REP}/DR^{FP} ratio within each MC, depicted as waterfall plot (left) and boxplot (right). **e-f**, Enrichment of gene signatures from MSigDB (C7, collections deposited by Goldrath (GR) and Kaech (KA), **Supplementary Table 2**) by gene set enrichment analysis comparing IdT_{CM} and hdT_{CM} (**e**), and enrichment plots (**f**) of 2 representative gene sets. **g**, Heatmaps depicting genes involved in immune function that were significantly ($P < 0.05$) depleted (left) or enriched (right) within IdT_{CM} (See **Extended Data Figure 4d**, **Supplementary Table 3**). Selected genes are annotated, complete gene lists in **Supplementary Table 4**. **h**, QstemScore of all T_{CM} MCs depicted as waterfall plot (left) and boxplot (right). QstemScore is based on marker genes of quiescent stem cells (**Supplementary Table 5**)³⁰, see methods for calculation. Data depicted were accumulated in two independent experiments (3–4 mice per experiment). Boxplots (**c**, **d**, **h**) indicate group median and 25th/75th percentiles, whiskers indicate the interquartile range multiplied by 1.5, dots signify individual MCs. The phenotype clusters T_{CM}^{EM} , T_{CM} (eff) and T_{CM} (mult) contain 4, 9 and 10 MCs, respectively. P values were determined by two-sided Tukey's HSD test (**c**), two-sided Student's T test with false-discovery rate correction (**d**, **h**), the FGS algorithm followed by the Benjamini-Hochberg procedure (**e**), or two-sided Wilcoxon Rank Sum test with Bonferroni correction (**g**). Significant P values (< 0.05) are indicated in the plots.

Amongst the observed T_{CM} MCs, two transcriptionally distinct subgroups could be identified (**Fig. 4b**). Specifically, while all T_{CM} showed the expected high expression of *Bcl2*, *Sell* and *Cd27*, and minimal expression of *Cx3cr1*, *Zeb2*, *Gzma* and *Prdm1* (**Fig. 4c**, **Extended Data Fig. 5a**), a dichotomy was observed in the expression of multipotency-associated (e.g. *Myb*, *Ccr7*) and effector-associated (e.g. *Tbx21*, *Lgals1*) genes within the T_{CM} pool (denoted as T_{CM} (mult) and T_{CM} (eff), respectively in the figures; **Fig. 4b**, **Extended Data Fig. 5a**). Next, we assessed the relation between transcriptional state and replicative history within the memory T cell pool. In line with the flow

cytometry data, the replicative history of T_{CM} —as a whole—exceeded that of T_{EMP} , thereby validating the scRNAseq approach. Strikingly, T_{CM} enriched for effector genes had overall higher DR^{RFP}/DR^{GFP} ratios compared to T_{CM} enriched for multipotency genes, demonstrating that stemness-related transcriptomic features are inversely associated with division history within the T_{CM} pool (**Fig. 4d**). Correspondingly, comparison of the three T_{CM} MCs with the highest and lowest level of prior division (hdT_{CM} and ldT_{CM} , respectively) revealed that ldT_{CM} were marked by the expression of key multipotency-associated genes, including *Tcf7*, *Sell*, *Myb* and *Eomes*, and several survival factors (*Gimap* and *Birc* family members, **Extended Data Fig. 5b, c**). Moreover, one ldT_{CM} MC was highly enriched for transcripts involved in inhibitory function (*Lag3*, *Cd160*, *Tox*), suggesting a possible analogy with the inhibitory signaling-dependent T_{CM} -precursor subset identified by Johnnidis *et al.*³³ (**Extended Data Fig. 5c**). In contrast, hdT_{CM} commonly expressed genes related to terminal differentiation, such as *Lgals1* and *S100* family members, and showed increased transcript levels for cytotoxicity-associated genes (*Nkg7*, *Ctsw*; **Extended Data Fig. 5b, c**). This link between replicative history and a multipotency versus effector-associated gene expression signature within the T_{CM} pool was further validated by differential gene expression analysis and gene set enrichment analysis (**Fig. 4e-g**, **Extended Data Fig. 5d**). In line with this association, *ex vivo* antigen stimulation of T_{CM} harvested from *Lm*-OVA memory mice showed that T_{CM} that had undergone more prior divisions were more likely to degranulate and less likely to produce IL-2, as compared to their less divided T_{CM} counterparts (**Extended Data Fig. 5e, f**).

The observed divergence in replicative history between distinct T_{CM} states potentially reflects the selective quiescence of a subset of T_{CM} with a less differentiated state. Of note, ldT_{CM} showed reduced expression of *Myc* targets and genes involved in cell metabolism (**Extended Data Fig. 5g**), suggesting that these cells exist in a transcriptionally-enforced replicative quiescent state. To test for such a transcriptional state, we scored the expression of a core gene set of quiescent stem cells from various tissues³⁶ (hereafter referred as QstemScore). Notably, T_{CM} that showed increased expression of multipotency-associated genes were marked by a higher QstemScore than T_{CM} with increased expression of effector-associated genes. (**Fig. 4h**). Moreover, variation in QstemScore could also be detected in gp33-specific P14 T_{CM} from an external data-set³⁷, and those P14 T_{CM} that prominently expressed this gene set transcriptionally resembled the multipotency-signature^{HI}, effector-signature^{LO} OT-I ldT_{CM} described here (**Extended Data Fig. 6**). Together, these data suggest a link between T_{CM} quiescence and the expression of multipotency-associated genes, driving the divergence in replicative history between distinct T_{CM} states.

To directly test whether replicative behavior in the T_{CM} pool is associated with a multipotency-associated state and relates to the functional capacity of T_{CM} to re-expand upon secondary activation, we established a DivisionRecorder-independent, CTV-based serial-transfer approach (**Fig. 5a**). Naïve OT-I and GFP;OT-I T cells were transferred into primary recipients that were subsequently exposed to *Lm*-OVA infection. At day 30 post-infection, early memory T cells were harvested, CTV labeled and transferred into infection-matched secondary recipients. 75 days later, CTV^{HI} (div0-2) and CTV^{LO} (div5+) T_{CM} were isolated, and the resulting T_{CM} populations were then profiled by scRNAseq, or transferred at a 1:1 ratio into tertiary recipients that were subsequently challenged

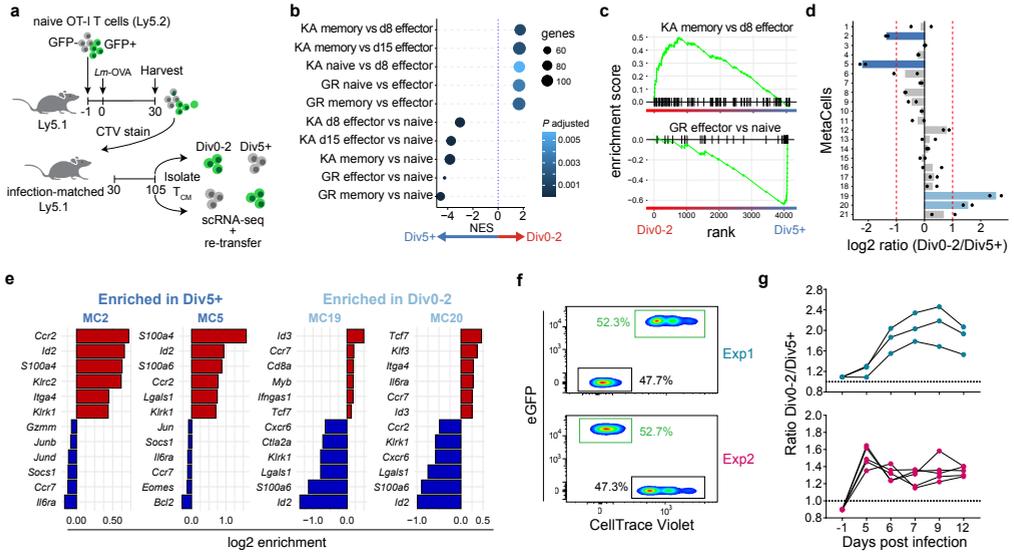


Fig. 5 Replicative history is linked to recall-potential within the T_{CM} pool. **a**, Experimental setup. Primary recipient mice received $5 \cdot 10^5$ naïve OT-I and $5 \cdot 10^5$ naïve GFP;OT-I T cells. 30 days after *Lm*-OVA challenge, CD8 T cells were enriched, labelled with CTV and transferred into infection-matched secondary recipient mice (1 primary recipient per secondary recipient). At d105 post infection, splenic CD27⁺KLRG1⁺ memory T cells that had either divided 0-2 or 5+ times and were either GFP⁺ or GFP⁻ were isolated by FACS. **b**, Enrichment of gene signatures from MsigDB (C7, collections deposited by Goldrath (GR) and Kaech (KA), **Supplementary Table 2**) between Div0-2 and Div5+ cells. Top and bottom 5 pathways are depicted. **c**, Enrichment plots of representative pathways detected by gene set enrichment analysis. **d**, Ratio of normalized counts between Div0-2 and Div5+ cells within each MC separately calculated for GFP⁺ and GFP⁻ populations. Bars indicate averages, dots indicate ratios of either GFP⁺ or GFP⁻ OT-I T cells. Red dotted lines indicate a fold change of 2. **e**, Waterfall plots depicting top and bottom 6 marker genes for selected MCs, filtered for genes involved in immune function (**Supplementary Table 3**). **f**, Flow cytometry plots depicting pre-transfer mixes of Div0-2 and Div5+ T_{CM}. **g**, 8,000-12,000 total memory T cells as described in **f** were transferred into infection-naïve mice, following *Lm*-OVA challenge 24 hours later. Ratios between Div0-2 and Div5+ derived cells was determined from peripheral blood samples at indicated days post infection. Lines connect populations from individual mice (Experiment 1 n = 3; Experiment 2 n = 5). Depicted scRNAseq data was collected from 4 mice, data describing recall potential was obtained from 8 mice. *P* values were determined by the FGSEA algorithm followed by the Benjamini-Hochberg procedure (**e**).

with *Lm*-OVA. Strikingly, comparison of quiescent (div0-2) cells and proliferative (div5+) T_{CM} by gene set enrichment analysis revealed a clear negative association between quiescence and an effector-like transcriptional state, while quiescence was positively associated with multipotency-associated gene expression (**Fig. 5b, c, Extended Data Fig. 7a**). Likewise, inspection of MCs (**Extended Data Fig. 7b-e**) that were enriched in the div0-2 cells, showed a prominent expression of multipotency-associated genes (*Myb*, *Tcf7*, *Id3*), whereas those enriched in div5+ cells showed increased expression of effector-associated genes (*Id2*, *S00a4*, *Lgals1*) (**Fig. 5d, e**). Furthermore, comparison of the expansion potential of div0-2 and div5+ T_{CM} demonstrated that quiescent T_{CM} were superior in generating offspring upon renewed infection (**Fig. 5f, g**), further demonstrating that replicative heterogeneity in the T_{CM} pool is both linked to transcriptional state and functionality.

Re-expansion potential of T_{CM} is linked to prior division

Having observed a link between prior division and recall potential in adoptive transfer experiments,

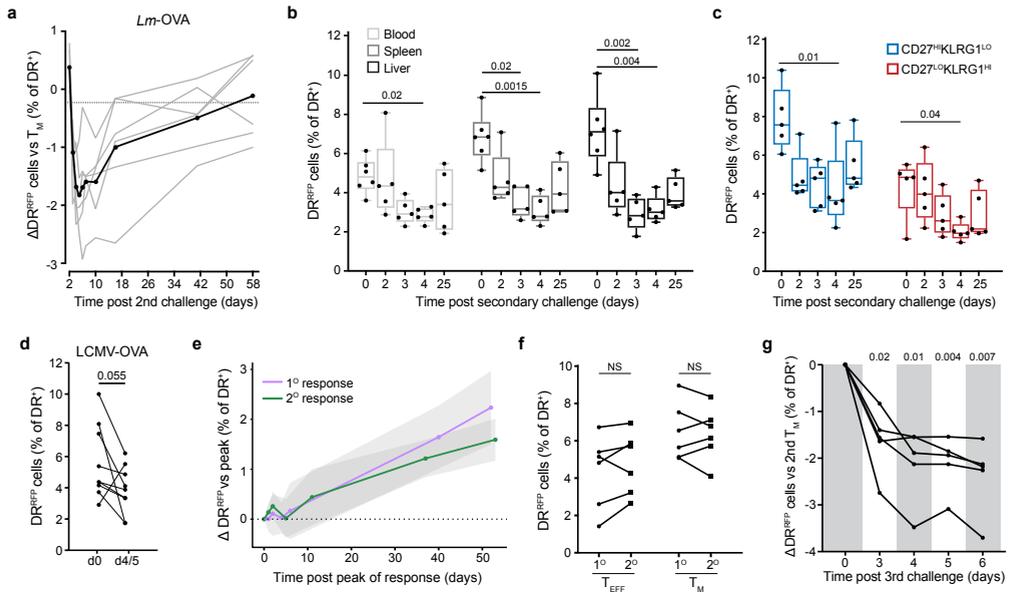


Fig. 6. The secondary T_{EFF} pool is predominantly generated by previously quiescent memory T cells. **a**, Kinetics of the percentage of DR^{RFP} cells in blood upon secondary *Lm*-OVA infection. Values are relative to the DR^{RFP} percentage within the respective memory pools ($n=6$ mice), black line represents group mean. **b-c**, DR^{RFP} percentages in indicated organs (**b**), or within splenic $CD27^{LO}KLRG1^{HI}$ and $CD27^{HI}KLRG1^{LO}$ populations (**c**) at indicated time points ($n=6$ mice per time point) post-secondary infection. Boxplots indicate group median and 25th/75th percentiles, whiskers represent min/max, dots represent individual samples. **d**, DR^{RFP} percentages in blood at memory (day >60) and at the peak of the secondary response (day 4/5 post-recall). Memory pools were generated with LCMV-OVA, recall infection was performed with *Lm*-OVA. **e**, DR^{RFP} acquisition in blood following primary and secondary infection. Values are relative to DR^{RFP} percentage at the peak of the primary or secondary response. Lines represent group medians ($n=6$ mice per group), greyed areas represent 95% confidence intervals. **f**, DR^{RFP} percentages in blood during effector and memory phases of the primary and secondary responses. Lines connect data of individual mice ($n=6$). **g**, DR^{RFP} percentages in blood ($n=5$ mice) upon tertiary infection. Mice were challenged twice with *Lm*-OVA with a >60 days interval, and subsequently infected with LCMV-OVA >60 days post-secondary infection. Depicted data are representative of at least 2 independent experiments. *P* values were determined by two-sided Kruskal-Wallis test with Dunn's multiple comparisons test (**b, c**), two-sided Wilcoxon signed-rank test (**d, f**), or repeated-measures one-way ANOVA followed by Dunnett correction (**g**).

we set out to verify this relationship without disruption of the T_M niche, through re-challenge of recipient mice carrying DR^+ memory OT-IT cells. In case the capacity for renewed expansion would primarily be restricted to replicative quiescent T_{CM} cells, the fraction of DR^{RFP} cells should show an initial decay upon reinfection—due to the increased preponderance of offspring derived from this previously quiescent population—followed by a gradual recovery throughout the contraction phase, as a result of novel division-dependent label acquisition. Notably, analysis of the fraction of DR^{RFP} T cells in blood revealed a steep decline during the first days post-secondary infection, followed by a gradual recovery during secondary memory formation (**Fig. 6a, Supplementary Note 4**). This transient reduction in the DR^{RFP} fraction was observed in multiple anatomical compartments (blood, spleen, liver), occurred independent of cell phenotype, and was also observed in LCMV-OVA induced T_M pools responding to secondary challenge (**Fig. 6b-d**). Of note, DR^{RFP} cell accumulation during the secondary contraction phase occurred at a comparable rate as during the primary response (**Fig. 6e**), yielding a secondary T_M pool that—despite extensive renewed clonal expansion—had undergone a similar number of divisions as the initial memory pool (**Fig. 6f**, median fold difference

= 1.03). Thus, the replicative histories of the T_{EFF} and T_M pools of the secondary T cell response mimic those of the primary T cell response, supporting the notion that the secondary expansion wave is mounted by a group of T_{CM} that has undergone limited prior division. Furthermore, this low-division T_{CM} pool is able to repeatedly reconstitute the effector T cell pool, as the same decrease in the fraction of DR^{RFP} cells was observed upon tertiary infection of mice (**Fig. 6g**).

To determine whether the observed data are consistent with re-expansion being driven by a memory T cell subset that becomes quiescent early in the immune response, we simulated T cell responses in which a fraction of T_{CM} precursors acquires replicative quiescence during the primary T cell response (see **Supplementary Note 5, Extended Data Fig. 8a**). Specifically, T cell responses were simulated that yielded quiescent T cells at a frequency of either $\pm 0.1\%$ or $\pm 1\%$ of the T_{EFF} pool, resulting in T_M pools in which quiescent T_{CM} accounted for ± 3 and ± 25 percent of the memory population (**Fig. 7a**). Modeling of DR^{RFP} labeling rates during recall responses in which the potential to re-expand was either abruptly lost as a function of the number of prior divisions (fun 1 and 2), or was lost more gradually across division history (fun 3), demonstrated that the transient drop in DR^{RFP} fractions is only consistent with models in which the capacity to re-expand is restricted to cells that have undergone limited clonal expansion (**Fig. 7a, b**). Furthermore, the stringency of this relation is strongly dependent on the relative size of the quiescent T_{CM} pool (**Fig. 7b**).

Taken together, our data establish that replicative state is not homogeneously distributed within

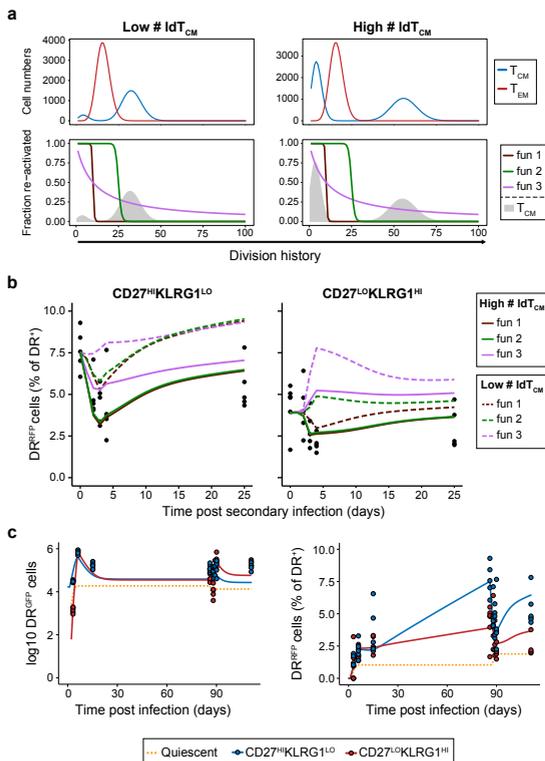


Fig. 7. Modelled T cell responses are consistent with the presence of a replication-competent quiescent T_{CM} population. **a**, Division history of T_{CM} and T_{EM} pools generated by modelled T cell responses (see **Supplementary Note 5**) during which a high (capped at 1% of the T_{EFF} pool size) or low (capped at 0.1% of the T_{EFF} pool size) fraction of T cells acquire quiescence during the effector phase (top). 3 re-expansion functions were used to restrict which fraction of T_{CM} with a given number of prior divisions will re-expand during recall (bottom). For reference, the division history of T_{CM} is shown as a shaded area. **b**, Modelled DR^{RFP} percentages within the $CD27^{LO}KLRG1^{HI}$ and $CD27^{HI}KLRG1^{LO}$ populations during secondary responses, with each re-expansion function applied to a memory pool containing either a high or low number of quiescent T_{CM} . Black dots indicate experimental measurements. **c**, Best fit of the modelled T cell response (number of quiescent T cells capped to 1% of T_{EFF}) experimental data obtained from spleen, depicting either cell numbers (left) or DR^{RFP} percentages (right). See **Supplementary Note 5** for details. Lines indicate the modeled populations; dots indicate experimental measurements

the T_{CM} pool and is linked to distinct transcriptional and functional properties. Specifically, our observations are consistent with a dichotomy in the T_{CM} pool, in which a self-renewing T_{CM} population maintains the T_M pool but marginally contributes to secondary expansion, and a replication-competent quiescent T_{CM} population is required to form the T_{EFF} pool that arises upon renewed infection (**Fig. 7c**, **Extended Data Fig. 8, 9**).

Discussion

Here, we report the development and application of the DivisionRecorder to dissect the replicative history of cell pools *in vivo*. We show that this approach allows longitudinal examination of division history, and how it may be combined with technologies such as flow cytometry and scRNAseq to couple cell state to division history. In the application presented here, the DivisionRecorder requires viral transduction to introduce one of its modules. While this did not significantly disrupt cell behavior in our study, development of a fully germline encoded DivisionRecorder system will be attractive, for instance, to follow replicative behavior of cell pools that are not amenable to adoptive transfer.

Using the DivisionRecorder, we demonstrate that, as a whole, the multipotent $CD8^+$ T cell pool has undergone substantial proliferation at the peak of the expansion phase, and continues to proliferate following pathogen clearance, resulting in a cumulative replicative age of the T_{CM} pool that exceeds that of the T_{EFF} and T_{EM} pool. Previous work has shown that a fraction of $CD62L^{HI}$ precursor- T_M divide at a lower rate than terminally differentiated effector subsets^{10,11,32}. In line with this, we observed a lower fraction of $Ki67^{HI}$ cells within the multipotent effector pool compared to the terminally differentiated pool, early post infection. At the same time, our data indicate that this difference does not result in a reduced cumulative number of past divisions within the entire $CD62L^{HI}$ T_{EFF} pool. Conceivably, these findings may be reconciled by the ability of highly proliferative $CD62L^{LO}$ T_{EFF} to phenotypically convert to a less differentiated $CD62L^{HI}$ state^{14,15,30}. Alternatively, the precursor- T_{CM} pool may harbor a heterogeneity in replicative history that is not revealed by the phenotypic markers used.

In line with the latter possibility, by combining the DivisionRecorder with scRNAseq we reveal that, while the T_{CM} pool has undergone substantial prior division as a whole, replicative history is heterogeneous within this pool and is associated with specific transcriptional states. First, our data demonstrate the presence of T_{CM} that bear transcriptional similarities to T_{EM} cells but, in contrast to T_{EM} , remain highly proliferative in the absence of inflammation (**Extended Data Fig. 9**). Second, we identify a population of quiescent T_{CM} that expresses reduced levels of effector-associated genes, and high levels of pro-survival genes and genes associated with quiescent stem cells³⁶. Several recent studies have reported the early emergence of $TCF-1^{HI}$ and $CD62L^{HI}$ effector cells that develop into memory T cells exhibiting stemness features^{38,39}. Moreover, Johnnidis *et al.*³³ propose early expression of inhibitory receptors as a mechanism preserving hallmark memory features. Although these early T cell subsets bear similarities to the quiescent T_{CM} observed here, further investigations into the developmental origin of distinct T_{CM} states are necessary to better understand the lineage

relationships between the T_{CM} states described here, and those present during the early phases of the T cell response.

A hallmark of immunological memory is the ability to efficiently generate a new wave of T_{EFF} upon renewed infection. Our data demonstrate that this ability is predominantly confined to a subgroup of replicative nascent T_M cells. The combined observations of a less differentiated quiescent T_{CM} population, and the reconstitution of the secondary and tertiary T_{EFF} pool by the output of these nascent progenitors, make a compelling argument for the presence of a bona fide stem cell population within the T_M pool. A growing body of work has examined a stem cell-like memory T cell (T_{SCM}) population^{40,41}, generally using cell phenotype to enrich and study these cells *ex vivo*. Using a function-driven, phenotype-agnostic, approach that does not require removal of cells from their niche, we observe a cell behavior that fits the profile of stem cell-like memory T cells *in situ*.

In high turnover tissues, such as the bone marrow^{42,43}, the intestinal epithelium^{44,45} and skin epidermis^{46,47}, two distinct behaviors of multipotent progenitor cells have been described: Actively dividing cells that promote normal tissue homeostasis, and quiescent cells that have been documented to break their dormancy upon tissue injury and exhibit profound re-population capacity^{42,45,48,49}. We propose that the two T_{CM} behaviors we describe provide the T cell compartment with the same capacity for renewal. Thus, the T cell pool can be viewed as an autonomous tissue that abides by organizing principles akin to those of the hematopoietic system and solid organs.

Methods

DivisionRecorder vector generation

In order to prevent expression of Cre recombinase during bacterial cloning, a synthetic intron—containing a splice donor, a branch site, a pyridine rich region, and a splice acceptor—was inserted into the Cre gene through three-fragment isothermal assembly. To prevent low level Cre translation occurring from alternative start sites, two ATG codons (position 78 and 84) were replaced by TGT codons. Finally, the Cre start codon was replaced by an EcoRI-spacer-XhoI site, to facilitate subsequent introduction of synthetic STRs. To generate the DivisionRecorder vector, two lox511 sites were introduced into the multiple cloning site of the pMX retroviral vector. Subsequently, an eGFP gene and the modified Cre recombinase gene were introduced directly upstream and downstream of the 5' lox511 site, respectively. Finally, a P2A element was inserted directly in between the eGFP gene and the 5' Lox511 site. Together, this resulted in a cassette comprising from 5' to 3': Kozak, an eGFP gene, a P2A site, a lox511 site, an EcoRI restriction site, spacer, an XhoI restriction site, a Cre recombinase gene, and a lox511 site. In its base configuration, Cre recombinase is out of frame. Synthetic STR domains were ordered as oligonucleotides (Invitrogen) and subsequently dimerized. STR dimers were inserted via the EcoRI and XhoI sites. Full sequences of all oligonucleotides are supplied in **Supplementary Table 6**.

Cre-activity reporter vector generation

LoxP sites were introduced into the multiple cloning site of the pCDH-CMVp-MCS-PGK-BlastR vector. In addition, a Katushka open reading frame was introduced, resulting in a vector containing from 5' to 3'; The CMV promoter, a floxed scrambled open reading frame, a Katushka open reading frame, the PGK promoter, and a blasticidin resistance gene.

Establishment of cell lines

The Cre-activity reporter cell line used in **Figure 1** was generated by retroviral transduction of HEK 293T cells (ATCC) with the Cre-activity reporter plasmid and subsequent Blasticidin selection (2 µg/ml, InvivoGen). Transduced cells were seeded at 1% confluency, and resulting single cell-derived colonies were transferred to individual wells. Clones were then examined for efficiency of induction of Katushka expression upon transfection with Cre recombinase, and the best-performing clone was selected. Cre-activity reporter cells were cultured in IMDM (Gibco) supplemented with 8% fetal calf serum (FCS, Sigma), 100 U/ml penicillin (Gibco), 100 µg/ml streptomycin (Gibco) and 2 mM Glutamax (Gibco). A mouse embryonic fibroblast (MEF) cell line from the Ai9 mouse strain was generated by modification of E14.5 embryonic fibroblasts with a retroviral vector encoding short-hairpin RNA directed against the p53 mRNA. Resultant cells were cultured in IMDM supplemented with 8% FCS, 100 U/ml penicillin, 100 µg/ml streptomycin and 2 mM Glutamax.

Mice

C57BL/6J-Ly5.1, OT-I, UBC-GFP and Ai9 mice were obtained from Jackson Laboratories, and strains were maintained in the animal department of The Netherlands Cancer Institute (NKI). Ai9 and OT-I, and UBC-GFP and OT-I mice were crossed to obtain the Ai9;OT-I and GFP;OT-I strains, respectively. Between 5-10 mice, both male and female, of the age of 6 to 15 weeks were used for each experiment. All animal experiments were approved by the Animal Welfare Committee of the NKI, in accordance with national guidelines.

Generation of DivisionRecorder⁺ OT-I T cells

Platinum-E cells (Cell Biolabs Inc) cultured in IMDM supplemented with 8% FCS, 100 U/ml penicillin, 100 µg/ml streptomycin, and 2 mM Glutamax were transfected with the DivisionRecorder vector using FuGeneTM6 (Promega). Retroviral supernatant was harvested 48h after transfection and stored at -80°C. Splens from Ai9;OT-I mice were harvested and mashed through a 70 µm strainer (Falcon) into a single cell suspension and resulting splenocytes were subsequently treated with NH₄Cl to remove erythrocytes. Subsequently, splenocytes were cultured in T cell medium (RPMI (Gibco Life Technologies) with 8% FCS, 100 U/ml penicillin, 100 µg/ml streptomycin, Glutamax, 10mM HEPES (pH 7.4), MEM Non-Essential Amino Acids (Gibco), 1mM Sodium pyruvate (Gibco), 50 µM 2-mercaptoethanol, supplemented with 1 ng/ml recombinant murine IL-7 (PeproTech) and 2 µg/mL ConcanavalinA (Merck)). After 48h, splenocytes were re-seeded on RetroNectin (Takara)

coated plates in T cell medium supplemented with 60 IU/mL human IL-2 and DivisionRecorder virus, and were centrifuged for 90min at 400g to allow spinfection. Virus concentration was chosen such that a transduction efficiency of approximately 10-15% was achieved, in order to minimize the occurrence of multiple retroviral integrations (**Supplementary Note 6**). Cells were harvested 24h later and a small aliquot was stained with anti-CD8-PerCPy5.5, anti-Vb5-PeCy7, anti-CD45.2-AF700 and DAPI to determine the fraction viable OT-I T cells (DAPI-CD8⁺Vb5⁺CD45.2⁺) by flow cytometry (Fortessa, BD Bioscience), which generally was around ~80%. CD8⁺Vb5⁺CD45.2⁺ cells that expressed GFP were considered as DivisionRecorder⁺ OT-I cells. Within the initial population of DivisionRecorder⁺ OT-I cells, the fraction of cells that already showed reporter activation (as inferred by tdTomato expression) 24h after transduction was consistently between 0.4 and 0.8%. Activated splenocytes were prepared for adoptive transfer (see below).

Infection, adoptive transfer and cell recovery

C57BL/6J-Ly5.1 mice were infected with 5,000-10,000 CFU of a recombinant *Listeria monocytogenes* strain that expresses ovalbumin or with 5,000 PFU artLCMV-OVA²⁵. Approximately 24h later, infected mice received 5,000-40,000 DivisionRecorder⁺ OT-I T cells through intravenous tail vein injection. To analyze OT-I T cell responses in peripheral blood over time, 25-50 μ L blood samples were obtained from the tail vein at the indicated time points, and were treated with NH₄Cl supplemented with 0.2 mg/ml grade-II DNaseI (Roche) to remove erythrocytes (see Methods, Flow Cytometry). To obtain spleen and liver samples, mice were sacrificed, organs were harvested, and single cell suspensions were prepared by means of mashing through a 100 μ M or 70 μ M strainer (Falcon), respectively. Subsequently, erythrocytes were removed by treatment with NH₄Cl. To purify leukocytes from single cell suspensions of liver tissue, cell suspensions were separated over a 37.5% Percoll (Sigma) density gradient. Obtained blood, spleen and liver samples were further processed for flow cytometric analysis, scRNA-sequencing or functional in vitro assays, as indicated. Samples were monitored for the occurrence of retroviral silencing; which was not observed in any of the examined samples (**Supplementary Note 7**)

Validation of DivisionRecorder functionality

To assess the ability of the DivisionRecorder to faithfully report on the replicative history of T cell populations using dilution of cell dyes as a reference, as described in **Fig. 2d-e**, we employed an experimental approach that was optimized to obtain sufficient DR^{RFP} events within the limited number of cell divisions that can be followed using cell dyes such as CTV (i.e., by transferring a high number of cells modified at a high transduction efficiency). Conclusions from this experiment are restricted to the validation of the functionality of the DivisionRecorder in dividing CD8⁺ T cells. Splenic CD8⁺ T cells were isolated using the Mouse CD8 T Lymphocyte Enrichment Set (BD Biosciences) and were subsequently stained with CellTrace™ Violet (ThermoFisher). Next, cells were activated for 16h in T cell medium supplemented with 0.05 μ g/mL SIINFEKL peptide and 60 IU/mL IL-2. Following this activation step, cells were seeded onto RetroNectin® (Takara

Bio) coated plates and were transduced with DivisionRecorder virus by spinfection for 4h in the presence of IL-2 and SIINFEKL peptide. Analysis of CellTrace™ Violet signal by flow cytometry indicated that the cells had not undergone a full cell division post labeling. Subsequently, 6×10^6 OT-I T cells were transferred into *Lm*-OVA infected recipients. Spleens were harvested 48h after adoptive transfer, processed into single cell suspensions and prepared for flow cytometric analysis. In order to accurately determine the fraction of DR^{RFP} cells per division during the initial stages of the proliferative burst when cumulative switching rate is still low, analysis of a large number of DivisionRecorder⁺ OT-I T cells events is required. For this reason, a transduction efficiency of ~60% was chosen in these experiments, instead of the 10-15% transduction efficiency used in other experiments. Note that a high transduction efficiency will result in the more frequent occurrence of cells that carry multiple retroviral integrations. The presence of cells with multiple integrations will result in a higher, yet stable, DR^{RFP} acquisition rate, as compared to the experimental set-up used in the remainder of the study.

Ex vivo analysis of degranulation and cytokine secretion potential of memory T cells

Spleens were harvested from recipient mice at >60 days post-infection, and CD8 T cells were isolated using the Mouse CD8 T Lymphocyte Enrichment Set (BD Biosciences). Following isolation, T cells were plated at 10^6 cells per well in 96-well U bottom plates in T cell medium supplemented with 0.05 µg/mL SIINFEKL peptide to selectively activate OVA-specific T cells. Following a 4hr incubation, capacity of indicated T cell populations to either produce the indicated cytokines or to degranulate was assessed. To allow analysis of cytokine production, Brefeldin A (GolgiPlug™, BD Biosciences) was added 30 minutes after initiation of T cell stimulation. To allow analysis of degranulation, T cell medium was supplemented with anti-CD107a and anti-CD107b antibodies at the initiation of T cell stimulation, and Brefeldin A (GolgiPlug™, BD Biosciences) and Monensin (GolgiStop™, BD Biosciences) were added 30 minutes after initiation of T cell stimulation. At the end of the T cell stimulation period, cells were stained for KLRG1 and CD27 and prepared for flow cytometric analysis (see below).

Flow cytometric analysis

Cells were taken up in PBS (Invitrogen) supplemented with 0.5% bovine serum albumin (BSA, Fisher Scientific), and stained with antibodies directed against the indicated cell surface proteins (1:200 dilution), for 30min on ice. To allow detection of intracellular cytokine production, cells were fixed and permeabilized with CytoFix/CytoPerm™ (BD Biosciences) according to the manufacturer's protocol and subsequently stained using antibodies against IL-2, TNFα and IFNγ. To detect intranuclear Ki-67 expression, the Foxp3/Transcription factor Staining buffer set (eBioscience) was used. See **Supplementary Table 7** for list of antibodies used in the study. All samples were acquired on a BD LSR Fortessa™ (BD Bioscience); DR^{GFP} and DR^{RFP} cells were identified as CD8⁺Vβ5⁺CD45.2⁺GFP⁺tdTomato⁻ and CD8⁺Vβ5⁺CD45.2⁺GFP⁻tdTomato⁺,

respectively. Flow cytometry data analysis was performed using FlowJo V10. An example of the used gating strategy is depicted in **Extended Data Fig. 10**.

For the moving average analysis depicted in **Fig. 3g** and **Extended Data Fig. 2e**, CD8⁺Vβ5⁺CD45.2⁺GFP⁺ events were exported and further processed using the R package FlowCore⁵⁰. In brief, outlier events (i.e., antibody aggregates/cell doublets) were removed, fluorescence intensities of each of the cell surface proteins were normalized using an inverse hyperbolic sine transformation and subsequently scaled between 0 and 1. To obtain the depicted moving averages, the fraction of DR^{RFP} cells was calculated within windows that each contained 10% of total cells, starting with the 10% of cells with the lowest expression levels for the indicated marker, and with subsequent windows moving up by steps of 2.5%.

Single cell RNA sequencing and data analysis of DivisionRecorder modified cells

The scRNAseq dataset of DivisionRecorder modified and unmodified OT-I memory T cells was obtained in two independent experiments, comprising 11 mice in total (See **Extended Data Fig. 3**). Experiment 1 included 3 mice containing DR⁺ memory T cells (mouse 1-3), which were processed in a single batch. Experiment 2 included 4 mice containing DR⁺ memory T cells (mouse 4-7) and 4 mice containing memory T cells derived from naïve OT-I T cells (unmodified, mouse 8-11), which were processed in two separate batches (batch 1: mouse 4-5 and mouse 8-9, batch 2: mouse 6-7 and mouse 10-11).

Spleens of DivisionRecorder⁺ OT-I T cell recipient mice (n=7) or naïve OT-I T cell recipient mice (n=4) were harvested >65 days post-infection. Splenocytes were stained with fluorochrome-conjugated antibodies directed against CD8, CD45.2 and Vβ5 (See **Supplementary Table 7**), to allow purification of transferred cells by FACS using the BD FACSAria™ Fusion Flow Cytometer (BD Biosciences). DR⁺ cells were subsequently FACS purified based on their expression of RFP and GFP. Following the isolation of DR^{GFP} and DR^{RFP} memory T cells by FACS (FACSAria Fusion, BD Biosciences), obtained cell populations were barcode-labeled with distinct anti-mouse TotalSeq™ Hashtag antibodies (TotalSeq-A0301-0306, Biolegend), and pooled, with an equal number of cells from each mouse to form the total pool of cells for scRNA-sequencing. If the amount of sorted DR^{RFP} cells from a particular sample was limited, it was pooled together with another DR^{RFP} sample to reduce cell loss during cell hashing (as indicated in **Extended Data Fig. 3**). Single-cell RNA isolation and library preparation was performed according to the manufacturer's protocol of the 10X Genomics Chromium™ Single Cell 3' kit, and the cDNA library was sequenced on a NextSeq™ 550 Sequencing System (Illumina). Cumulative data tallied to a total of ±15,000 cells. Feature-barcode matrices were generated using the Cell Ranger software of the 10X Genomics Chromium™ pipeline. Cells that could be ascribed to multiple samples or to no sample (inferred from the detection of multiple or no Hash tags), cells with a transcript (UMI) count lower than 1,500 and cells with a mitochondrial-gene fraction higher than 0.12 were excluded from downstream analysis. Next, cells were further filtered based on gene counts, setting upper and lower

thresholds separately for each sample-batch to control for differences in sequencing depth (gene-count-thresholds: Experiment 1 [1,200-3,000], experiment 2 batch 1 [800 -2,500], experiment 2 batch 2 [1,000-3,000]). Subsequent analysis of the remaining 11,767 cells was performed using the Seurat⁵¹ and MetaCell³⁴ R packages.

To examine enrichment or depletion of DR^{RFP} cells within the different MetaCells, cell counts were first normalized across hashtags. Data obtained from the different mice were subsequently aggregated and used to calculate the ratio of DR^{RFP} versus DR^{GFP} cells in each MetaCell. The immune signature gene list used in several analyses was composed of gene clusters involved, or proposed to be involved in T cell function. The full gene list is described in **Supplementary Table 3**.

Differential gene-expression testing was performed using the FindMarkers function (Wilcoxon Rank Sum test) implemented in Seurat, comparing all IdT_{CM} to all hdT_{CM}. Significantly differentially expressed genes ($P < 0.05$) were subsequently used for gene-set enrichment analysis using the R package fgsea⁵², testing for enriched gene-sets from the C7 immunologic or the H Hallmark gene-sets from Molecular Signatures Database (only including sets that consisted of >10 genes). Results from this analysis were filtered for collections deposited by Kaech and Goldrath (**Supplementary Table 2**), focusing on relevant CD8⁺ T cell biology.

To calculate the QstemScore, the log₂ enrichment values of genes that were positively or negatively associated with stem cell quiescence (**Supplementary Table 5**) were first summed within each MetaCell resulting in a positive and a negative score. QstemScore was then obtained by subtracting the negative-score from the positive-score.

Re-analysis of LCMV specific memory T cell scRNAseq dataset

Single cell transcriptomes from P14 memory T cells (harvested from spleen at day 90 post infection) were obtained from the Gene Expression Omnibus (accession GSE131847, sample GSM3822202). All single cells from this dataset were clustered applying the MetaCell algorithm. Next, T_{CM} MetaCells were determined based on the expression levels of core effector- and multipotency-related genes (**Supplementary Table 1**). QstemScores were then calculated for each of the T_{CM} MetaCells, and the 2 highest and 2 lowest scoring MetaCells were selected. Pearson correlations were subsequently calculated between each of these 4 T_{CM} MetaCells and all of the T_{CM} MetaCells from the OT-I dataset described here.

CTV-based serial transfer experiment and analysis

Spleens from OT-I and GFP;OT-I mice were harvested and CD8⁺ T cells were isolated using the Mouse CD8 T Lymphocyte Enrichment Set (BD Biosciences) according to the manufacturer's protocol. The obtained cells were mixed in a 1:1 ratio and transferred to 4 primary recipient C57BL/6J-Ly5.1 mice (1.5×10^6 T cells per recipient), and 24 hours later recipients were infected with 5,000-10,000 CFU *Lm*-OVA. 30 days post-infection, spleens and lymph nodes were harvested and CD8⁺ T cells were enriched using the Mouse CD8 T Lymphocyte Enrichment Set (BD

Biosciences), replacing the supplied antibody-cocktail with a mixture of anti-mouse CD19, CD20 and CD4 biotinylated antibodies (used 1:200 each, See **Supplementary Table 7** for information on antibody clones). The enriched cell pool was subsequently stained with CellTrace™ Violet (ThermoFisher) and re-transferred into 4 infection-matched secondary C57BL/6J-Ly5.1 recipients. 74 days after secondary transfer (104 days post-infection) spleens and lymph nodes were harvested from the secondary recipients and stained with anti-mouse KLRG1-PE, CD27-APC, and CD45.2-AF700 (See **Supplementary Table 7** for information on antibody clones). Next, stained cell-pools were first enriched for transferred cells (i.e., CD45.2⁺) through FACS using the BD FACSAria™ Fusion Flow Cytometer (BD Biosciences), and subsequently sorted again to obtain 4 populations of T_{CM} based on both GFP expression and CTV dilution: KLRG1⁺CD27⁺GFP⁺Division0-2, KLRG1⁺CD27⁺GFP⁺Division5+, KLRG1⁻CD27⁺GFP⁺Division0-2, KLRG1⁻CD27⁺GFP⁺Division5+. These cell pools were then further processed for tertiary transfer or single-cell RNA sequencing.

For tertiary transfer, GFP⁺Division0-2 cells were mixed 1:1 with the GFP⁺Division5+ cells (experiment 1), or GFP⁺Division0-2 cells were mixed 1:1 with GFP⁻Division5+ cells (experiment 2), thereby controlling for potential confounding effects of the donor strain. Next, 10,000 cells of either obtained cell pool were transferred in naive tertiary recipient C57BL/6J-Ly5.1 mice (3 mice for experiment 1, 4 mice for experiment 2). 24 hours later recipients were infected with 10,000 CFU *Lm*-OVA and the ratio of GFP⁺ over GFP⁻ cells within the transferred population (Ly5.2⁺) in blood was monitored by flow-cytometry over time.

For scRNAseq analysis, cell pools obtained by cell-sorting were barcode-labeled with distinct anti-mouse TotalSeq™ Hashtag antibodies (TotalSeq-A0301-0304, Biolegend), and subsequently pooled. Single-cell mRNA isolation and library preparation was performed according to the manufacturer's protocol of the 10X Genomics Chromium™ Single Cell 3' kit, and the cDNA library was sequenced on a NextSeq™ 550 Sequencing System (Illumina). Feature-barcode matrices were generated using the Cell Ranger software of the 10X Genomics Chromium™ pipeline, resulting in 13,064 single-cell transcriptomes. Cells that could be ascribed to multiple samples or to no sample (inferred from the detection of multiple or no Hashtags), cells with a transcript (UMI) count lower than 2,000 and cells with a mitochondrial-gene fraction higher than 0.12 were excluded from downstream analysis. Finally, cells with a gene-count of >2,800 were additionally excluded from further analysis. Subsequent analysis of the remaining 9,702 cells was performed using the Seurat⁵¹ and MetaCell³⁴ R packages.

Differential gene-expression testing was performed using the FindMarkers function (Wilcoxon Rank Sum test) implemented in Seurat, comparing all CTV^{HI} (division0-2) cells to all CTV^{LO} (division5+) cells. Significantly differentially expressed genes ($P < 0.05$) were subsequently used for gene-set enrichment analysis using the R package fgsea⁵², testing for enriched gene-sets from the C7 immunologic gene-sets (only including sets that consisted of >10 genes). Results from this analysis were filtered for collections deposited by Kaech and Goldrath (**Supplementary Table 2**), focusing on relevant CD8⁺ T cell biology.

For the MetaCell-based analysis, the number of cells within each hashtag-MetaCell combination

was counted, and subsequently normalized to 1,000 cells within each hashtag. The ratios of CTV^{HI} over CTV^{LO} was then calculated separately for the GFP;OT-I and OT-I derived cells.

Statistical analysis

Flow cytometric data was acquired using BD FACSDiva (v8.0) software. Flow cytometric data was analyzed using Flowjo (v10.4.2), R (v6.3.1, 'Action of the Toes'), and FLOWCore (v1.52.1). Single cell RNA sequencing data was analyzed using R (v 6.3.1), Seurat (v3.1.1), and MetaCell (v0.3.41). Data was visualized using Graphpad (V8.4.1, Prism software) and GGplot (v3.2.1). No statistical methods were used to pre-determine sample sizes, and sample sizes were chosen based on those reported in previous publications^{13,53}. Data distribution was assumed to be normal but this was not formally tested. Mice were stratified according to age and sex where appropriate. Data collection and analysis were not performed blind to the conditions of the experiments. No data points were excluded from the analyses.

Materials

All commercially available reagents are listed in **Supplementary Table 8**.

Data availability

Transcriptomic data presented in the manuscript have been deposited to the Gene Expression Omnibus (GEO), and can be accessed under the GEO accessions GSE169154 and GSE184947. The gp33-specific P14 T cell scRNAseq dataset was retrieved from GEO (accession GSE131847, sample GSM3822202). All statistical source data of the figures presented in the present study are provided with this paper. Indicated gene sets used in gene set enrichment analyses were retrieved from the Molecular Signatures Database (MSigDB) at <http://www.gsea-msigdb.org/gsea/msigdb>. Any additional data supporting the findings of this study are available from the corresponding authors upon request.

Code availability

R scripts that were used to produce the main and extended data figures in the manuscript are available from GitHub (https://github.com/kasbress/DivisionRecorder_analysis).

Acknowledgements

We would like to thank Monika C Wolkers (Sanquin, Amsterdam), Carmen Gerlach (Karolinska Institute, Stockholm) and Klaas van Gisbergen (Sanquin, Amsterdam) for helpful discussions regarding experimental procedures and sharing biological material, and Doron Merkler (University of Geneva, Geneva) for kindly providing the artLCMV-OVA. In addition, we would like to thank the NKI Genomics Core Facility and Flow Cytometry Core Facility for providing experimental

support. This work was supported by ERC AdG Life-his-T (Grant agreement ID: 268733) to T.N.S. and an NWO grant (ALWOP.265) to R.J.d.B.

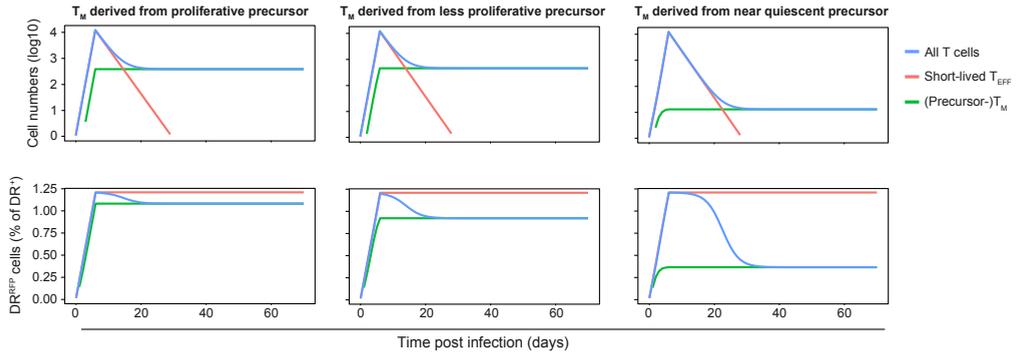
Author contributions

The study was designed by K.B., L.K., F.A.S. and T.N.S., and supervised by T.N.S. and F.A.S.; K.B. and L.K. jointly performed, analyzed, and visualized all experimental work included in the manuscript; F.A.S. and K.B. designed and developed the retroviral DivisionRecorder construct. L.A.K. and L.J. performed optimization and validation experiments integral to the design of the DivisionRecorder; A.C.S. and R.J.D.B. performed mathematical modelling, together with T.S.W., L.P. and K.R.D.; K.B. and L.K. wrote the manuscript with the input of co-authors; T.N.S. and F.A.S. critically reviewed and revised the manuscript.

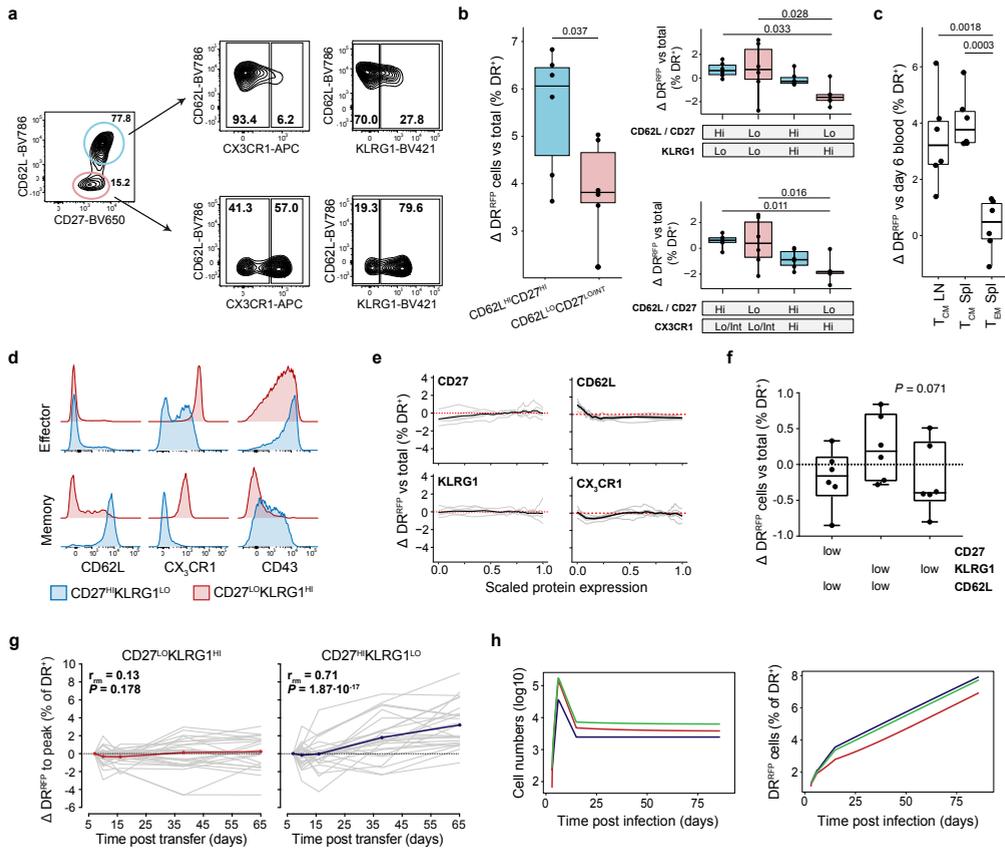
References

1. Hwang, L. N., Yu, Z., Palmer, D. C. & Restifo, N. P. The in vivo expansion rate of properly stimulated transferred CD8+ T cells exceeds that of an aggressively growing mouse tumor. *Cancer Res.* **66**, 1132–1138 (2006).
2. Yoon, H., Kim, T. S. & Braciale, T. J. The Cell Cycle Time of CD8+ T Cells Responding In Vivo Is Controlled by the Type of Antigenic Stimulus. *PLoS ONE* **5**, e15423 (2010).
3. Obar, J. J., Khanna, K. M. & Lefrançois, L. Endogenous Naive CD8+ T Cell Precursor Frequency Regulates Primary and Memory Responses to Infection. *Immunity* **28**, 859–869 (2008).
4. Blattman, J. N. *et al.* Estimating the precursor frequency of naive antigen-specific CD8 T cells. *J. Exp. Med.* **195**, 657–664 (2002).
5. Buchholz, V. R., Schumacher, T. N. M. & Busch, D. H. T Cell Fate at the Single-Cell Level. *Annu. Rev. Immunol.* **34**, 65–92 (2016).
6. Restifo, N. P. & Gattinoni, L. Lineage relationship of effector and memory T cells. *Curr. Opin. Immunol.* **25**, 556–563 (2013).
7. Akondy, R. S. *et al.* Origin and differentiation of human memory CD8 T cells after vaccination. *Nature* **552**, 362–367 (2017).
8. Sarkar, S. *et al.* Functional and genomic profiling of effector CD8 T cell subsets with distinct memory fates. *J. Exp. Med.* **205**, 625–640 (2008).
9. Obar, J. J. & Lefrançois, L. Early signals during CD8 T cell priming regulate the generation of central memory cells. *J. Immunol. Baltim. Md 1950* **185**, 263–272 (2010).
10. Kretschmer, L. *et al.* Differential expansion of T central memory precursor and effector subsets is regulated by division speed. *Nat. Commun.* **11**, 113 (2020).
11. Kinjyo, I. *et al.* Real-time tracking of cell cycle progression during CD8+ effector and memory T-cell differentiation. *Nat. Commun.* **6**, 6301 (2015).
12. Buchholz, V. R. *et al.* Disparate individual fates compose robust CD8+ T cell immunity. *Science* **340**, 630–635 (2013).
13. Gerlach, C. *et al.* Heterogeneous differentiation patterns of individual CD8+ T cells. *Science* **340**, 635–639 (2013).
14. Herndler-Brandstetter, D. *et al.* KLRG1+ Effector CD8+ T Cells Lose KLRG1, Differentiate into All Memory T Cell Lineages, and Convey Enhanced Protective Immunity. *Immunity* **48**, 716–729.e8 (2018).
15. Gerlach, C. *et al.* The Chemokine Receptor CX3CR1 Defines Three Antigen-Experienced CD8 T Cell Subsets with Distinct Roles in Immune Surveillance and Homeostasis. *Immunity* **45**, 1270–1284 (2016).
16. Reizel, Y. *et al.* Colon stem cell and crypt dynamics exposed by cell lineage reconstruction. *PLoS Genet.* **7**, e1002192 (2011).
17. Shlush, L. I. *et al.* Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability. *Blood* **120**, 603–612 (2012).
18. Kozar, S. *et al.* Continuous clonal labeling reveals small numbers of functional stem cells in intestinal crypts and adenomas. *Cell Stem Cell* **13**, 626–633 (2013).
19. Davis, F. M. *et al.* Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. *Nat. Commun.* **7**, 13053 (2016).
20. Weber, T. S., Perić, L. & Duffy, K. R. Inferring average generation via division-linked labeling. *J. Math. Biol.* **73**, 491–523 (2016).
21. Tempany, J. C., Zhou, J. H., Hodgkin, P. D. & Bryant, V. L. Superior properties of CellTrace Yellow™ as a division tracking dye for human and murine lymphocytes. *Immunol. Cell Biol.* **96**, 149–159 (2018).
22. Lai, Y. The Relationship Between Microsatellite Slippage Mutation Rate and the Number of Repeat Units. *Mol. Biol. Evol.* **20**, 2123–2131 (2003).
23. Koole, W., Schäfer, H. S., Agami, R., van Haften, G. & Tijsterman, M. A versatile microsatellite instability reporter system in human cells. *Nucleic Acids Res.* **41**, e158–e158 (2013).
24. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
25. Kallert, S. M. *et al.* Replicating viral vector platform exploits alarmin signals for potent CD8+ T cell-mediated tumour immunotherapy. *Nat. Commun.* **8**, 15327 (2017).
26. Wherry, E. J. *et al.* Lineage relationship and protective immunity of memory CD8 T cell subsets. *Nat. Immunol.* **4**, 225–234 (2003).
27. Becker, T. C. *et al.* Interleukin 15 Is Required for Proliferative Renewal of Virus-specific Memory CD8 T Cells. *J. Exp. Med.* **195**, 1541–1548 (2002).
28. Graef, P. *et al.* Serial transfer of single-cell-derived immunocompetence reveals stemness of CD8(+) central memory T cells. *Immunity* **41**, 116–126 (2014).
29. Olson, J. A., McDonald-Hyman, C., Jameson, S. C. & Hamilton, S. E. Effector-like CD8+ T Cells in the Memory Population Mediate Potent Protective Immunity. *Immunity* **38**, 1250–1260 (2013).

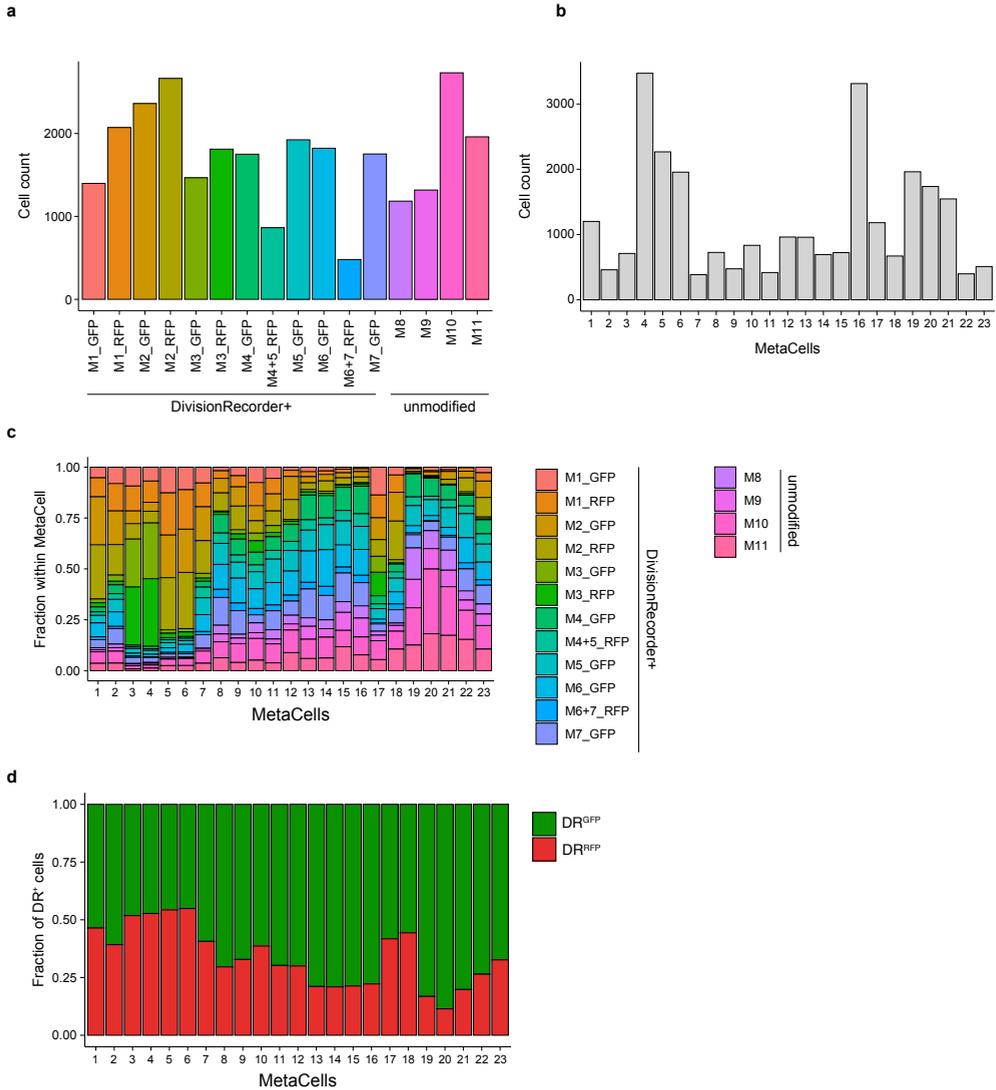
30. Youngblood, B. *et al.* Effector CD8 T cells dedifferentiate into long-lived memory cells. *Nature* **552**, 404–409 (2017).
31. Voehringer, D. *et al.* Viral infections induce abundant numbers of senescent CD8 T cells. *J. Immunol. Baltim. Md 1950* **167**, 4838–4843 (2001).
32. Lin, W.-H. W. *et al.* CD8 + T Lymphocyte Self-Renewal during Effector Cell Determination. *Cell Rep.* **17**, 1773–1782 (2016).
33. Johnnidis, J. B. *et al.* Inhibitory signaling sustains a distinct early memory CD8⁺ T cell precursor that is resistant to DNA damage. *Sci. Immunol.* **6**, eabe3702 (2021).
34. Baran, Y. *et al.* MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol.* **20**, 206 (2019).
35. Badovinac, V. P., Haring, J. S. & Harty, J. T. Initial T cell receptor transgenic cell precursor frequency dictates critical aspects of the CD8(+) T cell response to infection. *Immunity* **26**, 827–841 (2007).
36. Cheung, T. H. & Rando, T. A. Molecular regulation of stem cell quiescence. *Nat. Rev. Mol. Cell Biol.* **14**, 329–340 (2013).
37. Kurd, N. S. *et al.* Early precursors and molecular determinants of tissue-resident memory CD8⁺ T lymphocytes revealed by single-cell RNA sequencing. *Sci. Immunol.* **5**, eaaz6894 (2020).
38. Grassmann, S. *et al.* Early emergence of T central memory precursors programs clonal dominance during chronic viral infection. *Nat. Immunol.* **21**, 1563–1573 (2020).
39. Pais Ferreira, D. *et al.* Central memory CD8⁺ T cells derive from stem-like Tcf7hi effector cells in the absence of cytotoxic differentiation. *Immunity* **53**, 985–1000.e11 (2020).
40. Gattinoni, L. *et al.* Wnt signaling arrests effector T cell differentiation and generates CD8⁺ memory stem cells. *Nat. Med.* **15**, 808–813 (2009).
41. Gattinoni, L. *et al.* A human memory T cell subset with stem cell-like properties. *Nat. Med.* **17**, 1290–1297 (2011).
42. Laurenti, E. *et al.* CDK6 Levels Regulate Quiescence Exit in Human Hematopoietic Stem Cells. *Cell Stem Cell* **16**, 302–313 (2015).
43. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
44. Schepers, A. G., Vries, R., van den Born, M., van de Wetering, M. & Clevers, H. Lgr5 intestinal stem cells have high telomerase activity and randomly segregate their chromosomes. *EMBO J.* **30**, 1104–1109 (2011).
45. Yan, K. S. *et al.* The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 466–471 (2012).
46. Clayton, E. *et al.* A single type of progenitor cell maintains normal epidermis. *Nature* **446**, 185–189 (2007).
47. Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nat. Med.* **11**, 1351–1354 (2005).
48. An, Z. *et al.* A quiescent cell population replenishes mesenchymal stem cells to drive accelerated growth in mouse incisors. *Nat. Commun.* **9**, 378 (2018).
49. Sugimura, R. *et al.* Noncanonical Wnt Signaling Maintains Hematopoietic Stem Cells in the Niche. *Cell* **150**, 351–365 (2012).
50. Hahne, F. *et al.* flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* **10**, 106 (2009).
51. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
52. Korotkevich, G. *et al.* Fast gene set enrichment analysis. <http://biorxiv.org/lookup/doi/10.1101/060012> (2016) doi:10.1101/060012.
53. Kok, L. *et al.* A committed tissue-resident memory T cell precursor within the circulating CD8⁺ effector T cell pool. *J. Exp. Med.* **217**, e20191711 (2020).



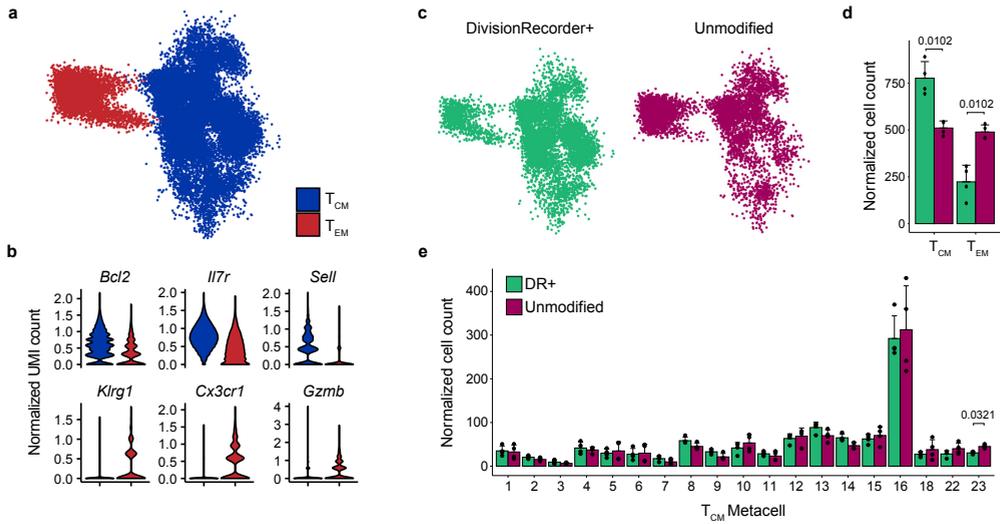
Extended Data Fig. 1. Simulation of different scenarios of memory T cell formation. Simulated data depicting a responding antigen-specific T cell population (blue), comprised of T_{EFF} undergoing clonal expansion and subsequent contraction (red), plus memory precursor T cells (MP, green) that develop into T_{IM} . Activated T_{EFF} are modeled to divide rapidly for 6 days (expansion phase), die at a fixed rate throughout the response, and can differentiate into MP cells only during the expansion phase. Cell numbers (top row) and DR^{REP} percentages (bottom row) are shown for 3 scenarios: (left) T_{EFF} can give rise to MP cells during the entire expansion phase, irrespective of the number of prior divisions, (middle) only T_{EFF} that have gone through at most 24 divisions can give rise to MP cells, or (right) only T_{EFF} that have gone through at most 10 divisions can give rise to MP cells. Note the strong decay in DR^{REP} percentage that is observed during memory formation in case T cell memory is founded by T cells that have undergone few divisions. See **Supplementary Note 3** for detailed description and equations.



Extended Data Fig. 2. Evaluation of the division history of T cell subsets throughout a response to *Lm*-OVA. **a**, Gating strategy used to identify indicated T_M populations (d86) in spleen samples. **b**, DR^{RFP} percentages within splenic T_M populations (n=6 mice) as identified in panel a. **c**, DR^{RFP} percentages within the CD27^{HI}KLRG1^{LO} T_{CM} subset in spleen and lymph nodes (LN) and within the CD27^{LO}KLRG1^{HI} T_{EM} subset in spleen. **d**, Cell surface expression of CX₃CR1, CD62L, and CD43 within splenic CD27^{LO}KLRG1^{HI} and CD27^{HI}KLRG1^{LO} populations at the peak of the T_{EFF} phase (day 6 post infection) and in memory phase (day 86 post infection). **e**, Moving-average of surface marker expression of splenic DR^{RFP} OT-I T cells during effector phase (day 6), depicted as in Fig. 3g. **f**, Boxplots depicting DR^{RFP} percentages within T_{EFF} (day 6 post infection) subsets in spleen (n=6 mice), relative to the total DR^{RFP} percentage. **g**, Kinetics of DR^{RFP} percentages within CD27^{LO}KLRG1^{HI} (left) and CD27^{HI}KLRG1^{LO} (right) DR^{RFP} OT-I T cell populations in blood. Values are relative to the percentage of DR^{RFP} cells detected at the peak of the response (day 6). Grey lines represent individual mice (n = 22), red and blue lines indicate group mean. **h**, Simulation of the phenotype model (See **Supplementary Note 5** for details) illustrating a scenario in which conversion of CD27^{HI}KLRG1^{LO} to CD27^{LO}KLRG1^{HI} cells occur only after the peak of the response at a low rate. Depicted are the overall cell numbers (left), and the percentage DR^{RFP} cells of DR^{RFP} OT-I T cells in CD27^{HI}KLRG1^{LO} cells (blue), CD27^{LO}KLRG1^{HI} cells (red) and the total T cell population (green). Note that in this scenario the fraction DR^{RFP} within the terminally differentiated CD27^{LO}KLRG1^{HI} population would increase to almost twice the experimentally observed frequency. All depicted data are representative of at least two independent experiments. Boxplots (**c**, **d**, **g**) represent group median and 25th/75th percentiles, whiskers indicate the interquartile range multiplied by 1.5 (**c**, **d**) or min/max (**g**), dots indicate individual samples. *P* values were determined by one-way ANOVA followed by Tukey's HSD post-hoc test (**c** and **d**), two-sided Student's *T* test (**c**), two-sided repeated measurement correlation test (**h**), or two-sided Friedman test (**g**). All significant (< 0.05) *P* values are indicated in the plots.

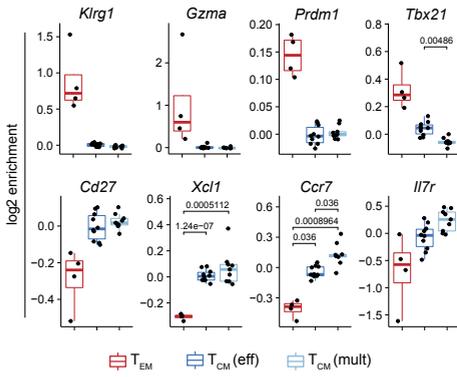


Extended Data Fig. 3. Single cell mRNA sequencing of DivisionRecorder⁺ and unmodified memory T cells. Single cell mRNA sequencing was performed on DivisionRecorder modified and unmodified OT-I memory T cells (Day 75 and 85 post *Lm*-OVA infection), isolated from spleens (n=7 mice with DR⁺ memory T cells; n=4 with unmodified memory T cells). Obtained data were aggregated from two independent experiments (Experiment 1: M1-3; Experiment 2: M4-11). All cells were jointly analysed and clustered. **a**, Cell count per sample. **b**, Total cell count per MC. **c**, Sample composition of each MC. **d**, Relative contribution of DR^{GFP} and DR^{RFP} to the total DR⁺ pool within each MC.

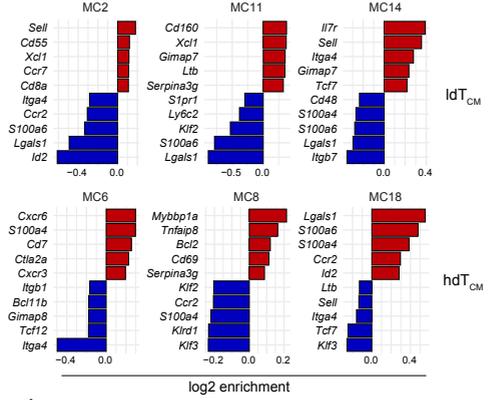


Extended Data Fig. 4. T_{CM} transcriptional states are preserved in DR⁺ OT-I T cells. Comparison of transcriptional states of splenic memory T cells generated by either DivisionRecorder modified, or unmodified OT-I T cells (Day 75 and 85 post *Lm*-OVA infection). **a-b**, Memory OT-I T cells cluster into T_{CM} (blue) and T_{EM} (red). 2D projection colored by subset (**a**), and violin plots depicting normalized UMI counts of selected genes (**b**) are shown. **c**, 2D projection of either DR⁺ (left) or unmodified (right) memory OT-I T cells. **d**, Contribution of DR⁺ and unmodified memory T cells to the T_{CM} and T_{EM} subsets. **e**, Contribution of DR⁺ and unmodified OT-I T cells to the 19 MCs that jointly make up the T_{CM} subset. Dots indicate individual mice ($n=3$ per condition). Note that all T_{CM} states are generated in near-equal proportions by DR⁺ and unmodified memory T cells. Depicted scRNAseq data was obtained from 6 individual mice, and was aggregated from 2 independent experiments. P values were determined by two-sided Student's T test followed by Bonferroni correction for multiple testing (**d** and **e**). P values < 0.05 are indicated.

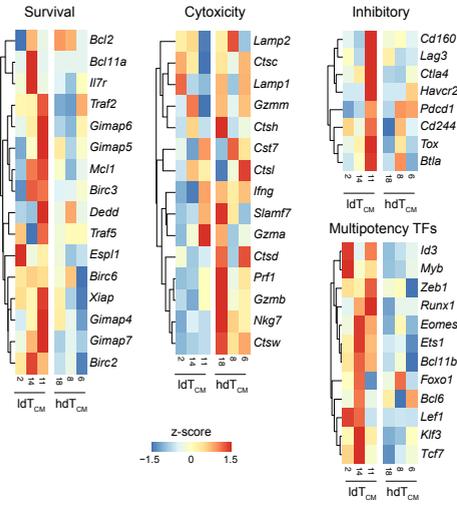
a



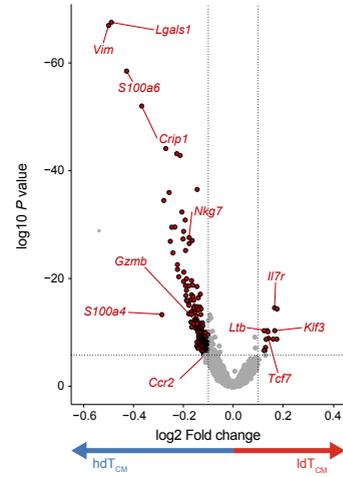
b



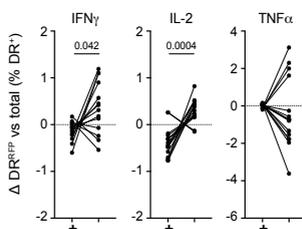
c



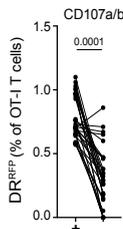
d



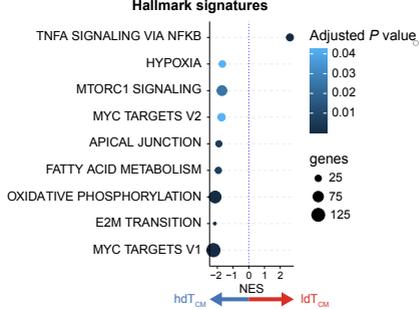
e



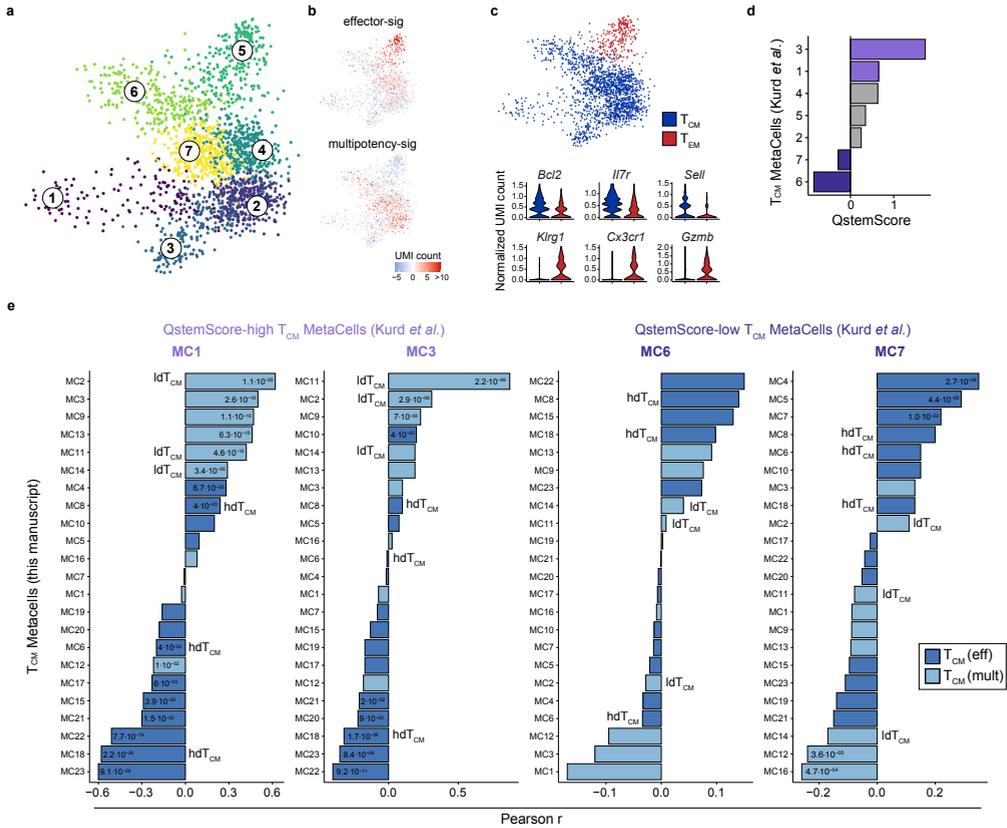
f



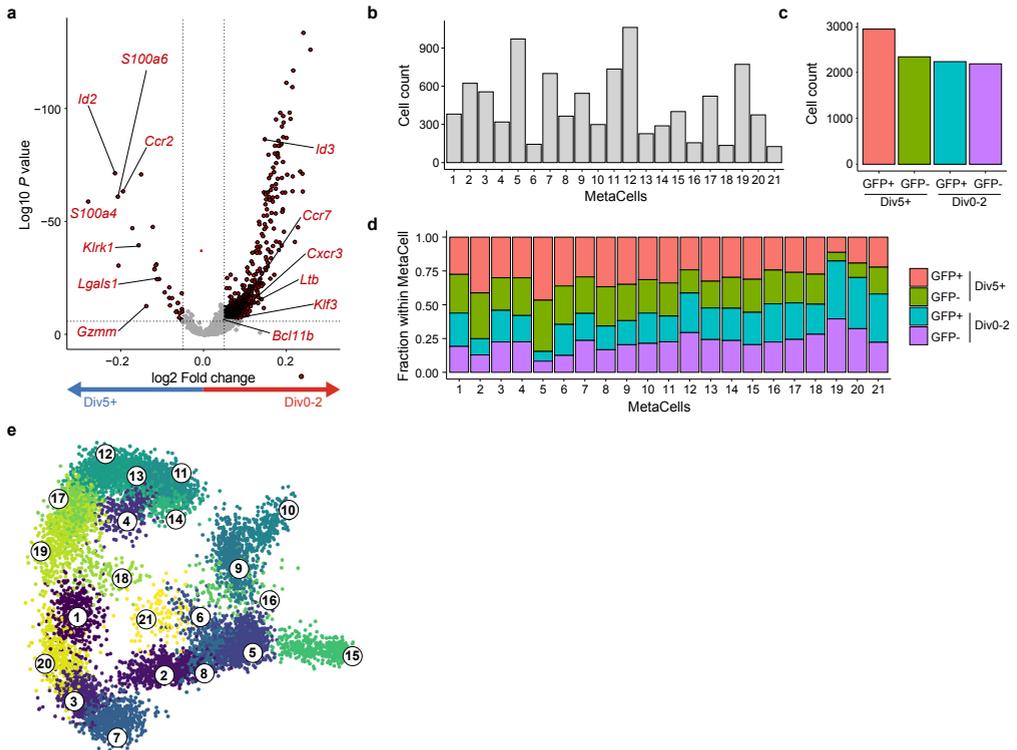
g



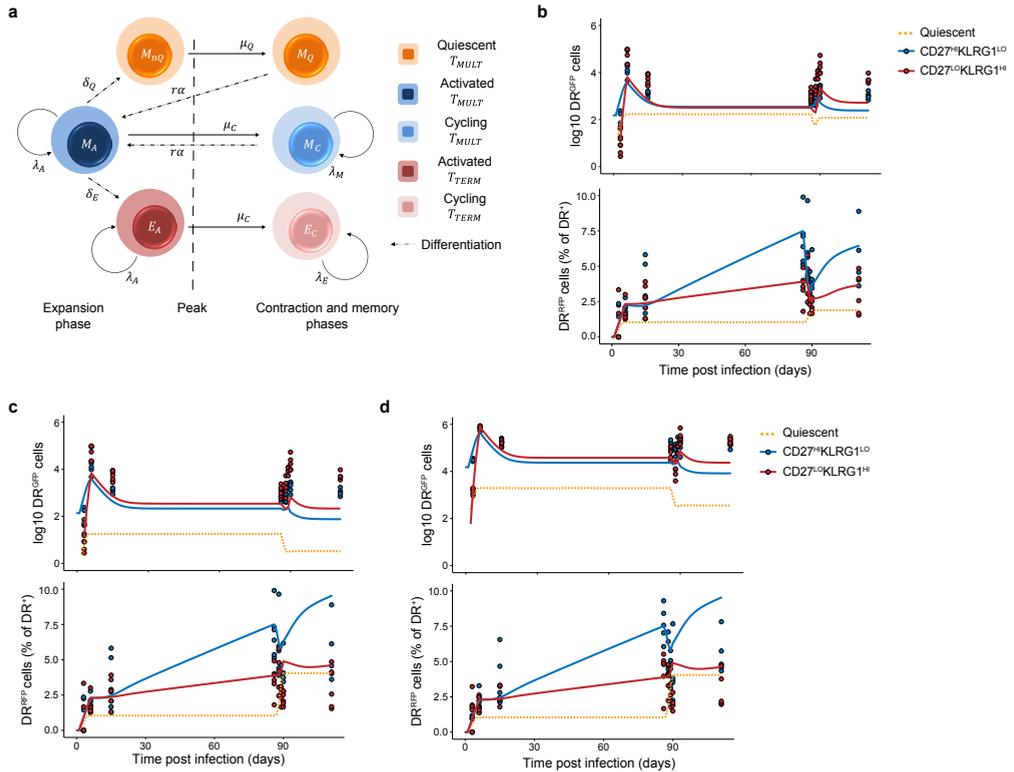
Extended Data Fig. 5. Replicative history identifies distinct transcriptional states within the T_{CM} pool. Single cell transcriptomic profiling of DR⁺ T cells obtained from spleen in memory phase (Day 75 and 85 post *Lm*-OVA infection). **a**, Log2 enrichment of selected genes in each MC cluster. Boxplots indicate group median and 25th/75th percentiles, whiskers indicate the interquartile range multiplied by 1.5, dots signify individual MCs. The phenotype clusters T_{EM}^+ , T_{CM} (eff) and T_{CM} (mult) contain 4, 9 and 10 MCs, respectively. For definition of T_{CM} (eff) and T_{CM} (mult), see Fig. 4B. **b**, Top and bottom marker genes of ldT_{CM} (Top, MC2, 11, 14) and hdT_{CM} (Bottom, MC6, 8, 18), see Fig. 4D for ldT_{CM} and hdT_{CM} definitions. **c**, Heatmaps depicting z-score transformed enrichment values of genes related to cell survival (left), cytotoxicity and effector function (middle), inhibitory markers (top-right), and transcription factors involved in T cell multipotency (bottom-right). Expression is depicted for the 3 ldT_{CM} and 3 hdT_{CM} MCs. **d**, Volcano plot depicting differentially expressed genes in ldT_{CM} versus hdT_{CM} . Significantly (adjusted P-value < 0.05) differentially expressed genes are depicted in red. Selected genes are highlighted. **e**, Cytokine release of CD27^{hi}KLRG1^{lo} DR⁺ T cells (isolated from spleen at day >60 post infection) 4 hours post *ex vivo* stimulation. Percentage DR^{REP} cells within cytokine producers (+) and non-producers (-), relative to the average DR^{REP} percentage within each sample, is depicted. Lines connect individual *ex vivo* stimulated samples (n=12), obtained from 3 mice. **f**, *Ex vivo* degranulation of CD27^{hi}KLRG1^{lo} DR⁺ T cells (isolated from spleen at day >60 post infection) 4 hours post *ex vivo* stimulation. Percentage DR^{REP} cells within the CD107a/b positive (+) or negative (-) cell populations is depicted. Lines connect individual samples *ex vivo* stimulated samples (n=17), obtained from 5 mice. **g**, Enrichment of gene signatures from MSigDB (Hallmark) by gene set enrichment analysis comparing ldT_{CM} and hdT_{CM} . Data depicted was accumulated in two independent experiments (3-4 mice per experiment). P values were determined by Tukey's HSD test (**a**), Wilcoxon Rank Sum test with Bonferroni correction (**d**), two-sided Wilcoxon signed-rank test (**e**, **f**), the FGSEA algorithm followed by the Benjamini-Hochberg procedure (**g**). P values < 0.05 are indicated.



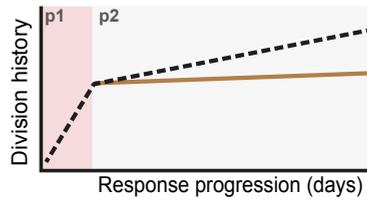
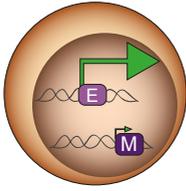
Extended Data Fig. 6. gp33-specific P14 T_{CM} with increased expression of genes associated with replicative quiescence resemble OT-I IdT_{CM} . Re-analysis of scRNAseq profiled splenic of P14 memory T cells, published in Kurd *et al.* (Kurd *et al.*, *Science Immunology*, 2020). **a-b**, 2D projection of P14 memory T cells 90 days post LCMV infection, colors indicate individual MCs (**a**), or the relative expression of effector- and multipotency-associated genes (**b**). Gene list in **Supplementary Table 1**. **c**, P14 memory T cells cluster into T_{CM} (blue) and T_{EM} (red). 2D projection colored by subset (top), and violin plots depicting normalized UMI counts of selected genes (bottom) are shown. **d**, QstemScore of all T_{CM} MCs in the Kurd *et al.* dataset. **e**, Pearson correlations between the Kurd *et al.* P14 T_{CM} MCs that score high (MC1, 3) or low (MC6, 7) for QstemScore, and all OT-I T_{CM} MCs described here. Data are depicted as waterfall plots, asterisks indicate significant correlations. T_{CM} (eff), T_{CM} (mult), IdT_{CM} and hdT_{CM} MCs are defined in **Figure 4**. P values were determined by two-sided Pearson correlation test followed by Bonferroni correction (e). P values < 0.05 are indicated in the plots.



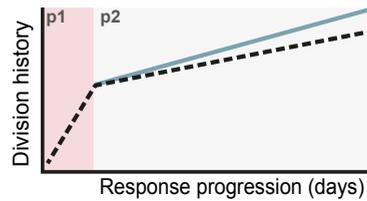
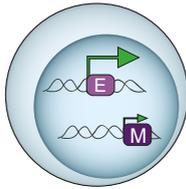
Extended Data Fig. 7. Single cell mRNA sequencing analysis of highly divided and less divided splenic T_{CM} . **a**, Volcano plot depicting differentially expressed genes in Div0-2 versus Div5+ T_{CM} . Significantly differentially expressed genes (Adjusted $P < 0.05$) are depicted in red. Selected immune-related genes are highlighted. **b**, Cell count per MC. **c**, Number of sequenced cells per sample included in the analysis. **d**, Sample composition of each MC. **e**, 2D projection, colors indicate different MCs. Depicted scRNAseq data was collected from 4 individual mice. P values were determined by Wilcoxon Rank Sum test with Bonferroni correction (**a**).



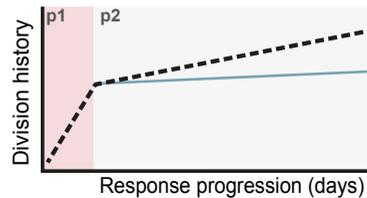
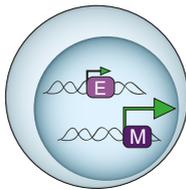
Extended Data Fig. 8. Modelled T cell responses are consistent with the presence of a replication-competent quiescent T_{CM} population. **a**, Cartoon of the phenotype model depicting phenotypes, the considered interactions among them and the parameters associated with the interactions. Arrows indicate various events occurring during the response, such as cell division (denoted with λ), differentiation to a different phenotype (denoted with δ), cell death during contraction (denoted with μ), and recruitment toward the secondary response during recall infection (denoted with r). Subscripts indicate the phenotype of the cell that the parameter is affecting. Full list of parameters can be found in **Supplementary Note 5**. **b-d**, Best fit of the modelled T cell response to the experimental measurements depicting either cell numbers (top plot in each panel), or DR^{REP} percentages (bottom plot in each panel). The total number of quiescent T cells generated was either capped at 1% (**b**) or 0.1% (**c, d**) of the T_{EFF} pool. Lines depict the modeled populations; Dots indicate the experimental measurements obtained from peripheral blood (**b, d**) or spleen (**c**). See **Supplementary Note 5** for more details and calculations. Experimental data points are representative of at least two independent experiments, dots indicate individual mice (n=6 mice per time point).

T_{EM}


Action upon reinfection	
Proliferation	—
Cytotoxicity	↑↑↑

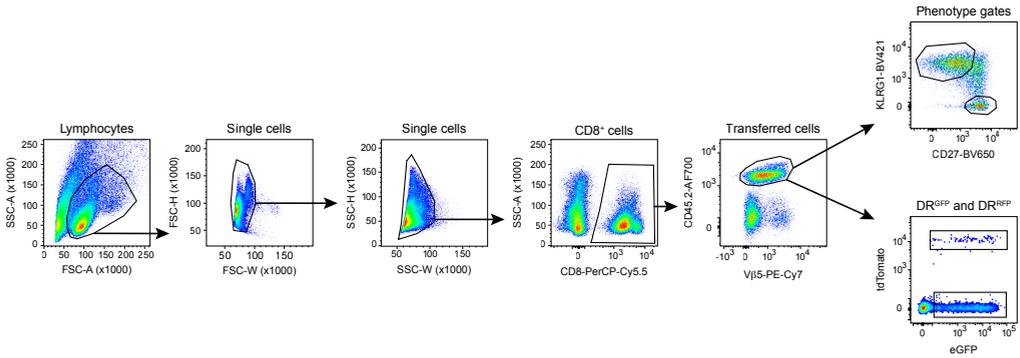
 proliferative T_{CM}


Action upon reinfection	
Proliferation	↑
Cytotoxicity	↑

 quiescent T_{CM}


Action upon reinfection	
Proliferation	↑↑↑
Cytotoxicity	—

Extended Data Fig. 9. Model describing replicative behaviors in the CD8⁺ memory T cell pool. Upon infection, antigen-specific CD8⁺ T cells activate and rapidly expand (phase 1, p1). Following pathogen clearance (p2), a subset of memory T cells continues to divide, resulting in a progressive increase in the replicative history of the overall T cell memory pool (dotted line). Within this population, three separate behaviors of transcriptionally disparate memory T cell pools can be distinguished. **Top**) Terminally differentiated T_{EM} cells that cease division after the inflammation phase (p1) and that are marked by high transcription of effector- and minimal expression of multipotency-associated genes ([E], [M]). Upon reactivation, these cells exert rapid effector functions, but lack the potential to re-expand. **Middle**) A subgroup of T_{CM} that continues to proliferate in the memory phase, exhibits diminished levels of multipotency-associated transcripts, and that abundantly expresses effector-associated genes. Although the functionality of these cells upon reinfection requires further study, their heightened expression of effector-associated genes suggests that these cells exert cytotoxic activity upon reinfection. The contribution of these cells to the secondary T_{EFF} pool is limited. **Bottom**) A subgroup of T_{CM} cells that shows low expression of effector-associated genes but increased expression of multipotency-associated genes, and that exists in a near-quiescent state after the inflammation phase. Upon renewed infection, this cell pool is primarily responsible for the generation of a new wave of secondary T_{EFF} . Based on their transcriptional profile, these cells are expected to have limited immediate cytotoxic functions.



Extended Data Fig. 10. Gating strategy. General gating applied to flow cytometry data presented in the study. Single lymphocytes were first selected using morphology gates, and were subsequently gated on CD8⁺ T cells and transferred OT-I T cells (Vβ5⁺CD45.2⁺). Next, DR^{RFP} and RF^{GFP} could be directly selected, or first separated by phenotype depending on the analysis. The data presented here was analyzed from blood of a recipient of DR⁺ cells, and was acquired 6 days post infection with *Lm*-OVA. Phenotype gates other than those shown here are defined in their respective figures.

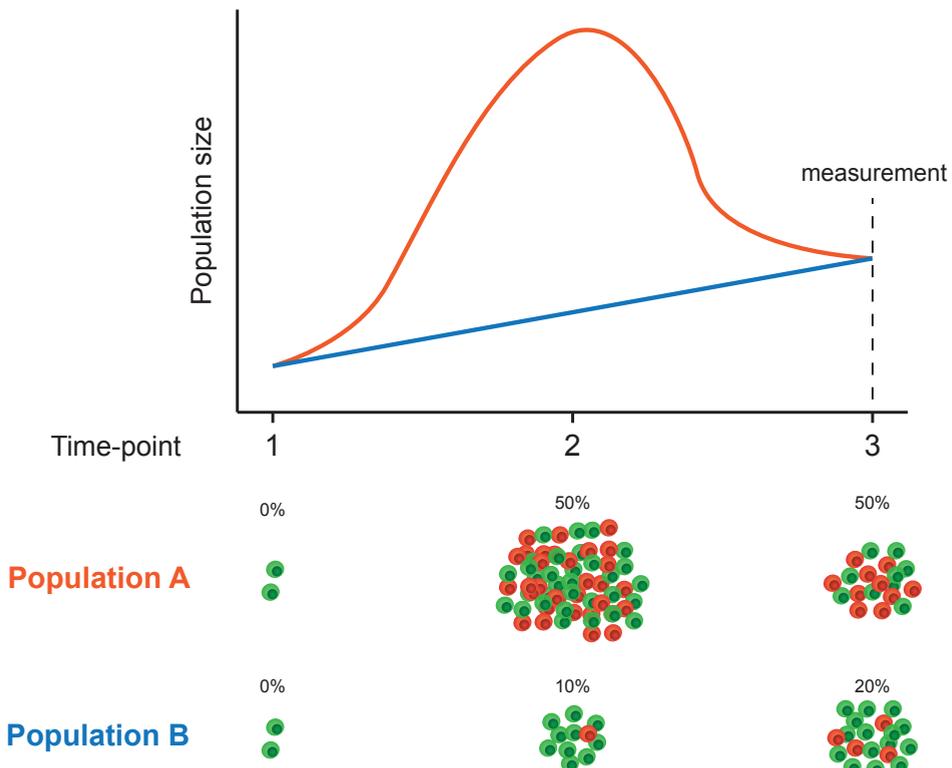
Supplementary Notes

Supplementary Note 1

Division-coupled stochastic labeling as a proxy of division history

If we would consider a hypothetical T cell response comprising of two independent populations (A and B; see cartoon below): Both populations start dividing at time-point 1 with the same amount of RFP label. Population A undergoes a large amount of expansion, and subsequently stops dividing and contracts significantly. Population B undergoes a low level of proliferation, but experiences no contraction. At the time-point of measurement (time-point 3), population A and B have an equal size.

Due to the high level of proliferation in population A, it accumulates a larger amount of RFP⁺ cells at time-point 2 as compared to its counterpart. As population A stops dividing, it stops accumulating RFP⁺ cells. Next, because RFP labeling occurred stochastically, contraction will occur to the same extent in the labeled and the unlabeled cell pool. As a result, the two equally sized populations that are analyzed at time-point 3 will contain different fractions of RFP⁺ cells, which reflect the difference in division history between these populations.



Supplementary Note 2

Minimal ODE model

We detail the results from Weber et al¹, as originally shown for a branching process for a system of ordinary differential equations (ODEs). As in Weber et al¹, the average generation number is defined as the mean of the generation numbers of all the cells in the population.

For $\frac{dN}{dt} = (\lambda - d)N$ cells dividing at a rate λ /day and dying at a rate d /day, according to the ODE with time can be written as

$$G' = (2 - k)\lambda G + k\lambda(1 - p)G - \lambda G - dG \quad (1)$$

$$R' = k\lambda pG + 2\lambda R - \lambda R - dR$$

the fraction of DR^{RFP} labeled cells, f_R at time t is

$$f_R(t) = 1 - f_G(0)e^{-k\lambda pt} \quad (2)$$

here, p is the labelling probability and k is the number of daughter cells that get labeled. $k = 1$ is the asymmetric case when only one daughter cell can be labeled during cell division, and $k = 2$ is the symmetric case when both daughters can get labeled during cell division.

In such a model the average generation number, $\mu_t = 2\lambda t$, is independent of the death rate². For this ODE, the relationship between the fraction of DR^{RFP} labeled cells in a population and its average generation number at some time-point t is

$$\mu_t = 2\lambda t = -\frac{2}{kp} \ln\left(\frac{f_G(t)}{f_G(0)}\right) \quad (3)$$

MEF experiment described in Fig. 1h-i

Linear regression on the cell number data from the MEF experiment was used to infer the division rate λ (**Fig. I**). Using the same minimal ODE and assuming no cell death, the estimate for the division rate in the MEF data is $\lambda = 0.698$ /day. The 95% confidence interval for the fit is 0.673-0.723.

With the division rate known, **Eq. (2)** was fitted to the fraction of DR^{RFP} labeled cells to estimate the switching probability p . **Fig. 1i** shows the fits of 300 bootstraps (100 per MEF experiment) on the MEF data. The basic statistics of the switching probability estimates from these fits are $p = 0.0053$ (mean), 0.0052 (median), 0.0043-0.0063 (95% CI).

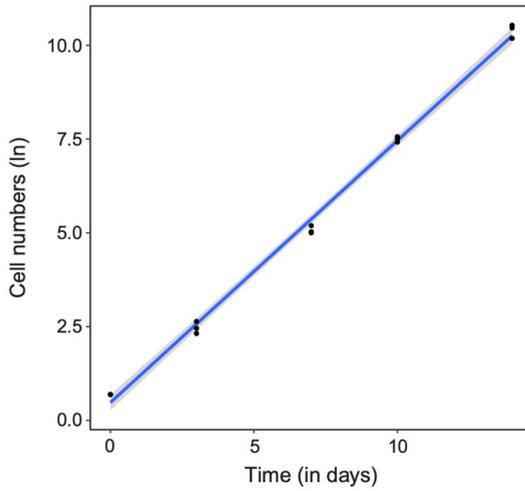


Figure 1: Immortalized DivisionRecorder⁺ (DR⁺) mouse embryonic fibroblasts were cultured, counted, and analyzed every 3–4 days. Natural log of the number of DR⁺ cells is shown for three experimental replicates. Black dots represent the experimental data, the best fit of the linear regression is depicted by the blue line, greyed area represents the 95% confidence interval. The slope of this regression line is the division rate of the cells in the MEF experiment.

Supplementary Note 3

Expected DR^{RFP} fraction in the memory T cell pool

In Extended Data Fig. 1 we model the expected DR^{RFP} fractions in the memory T cell pool for a several scenarios where memory is generated from a subset of T cells that have undergone different a different number of divisions during the effector phase.

Consider the clonal expansion and subsequent contraction of a T cell population with two phenotypes: activated A cells and quiescent Q cells. Activated cells divide at a rate $\lambda_A = 2/\text{day}$ for 6 days (expansion phase), die at a rate $d_A = 0.2/\text{day}$ throughout, and can differentiate into quiescent cells at a rate α_i during the expansion phase. One daughter cell of a dividing unlabeled cell (DR^{GFP}, X_G) can become permanently labeled (DR^{RFP}, X_R) with a switching probability $p = 0.0013$. The labelling is genetic i.e., the daughters of a labeled cell cannot be unlabeled. In the model below, $i \geq 0$ denotes the division number.

$$\begin{aligned}
 A'_{G_{i+1}} &= \lambda_A A_{G_i} + (1-p)\lambda_A A_{G_i} - (\alpha_i + \lambda_A + d_A)A_{G_{i+1}} \\
 A'_{R_{i+1}} &= p\lambda_A A_{G_i} + 2\lambda_A A_{R_i} - (\alpha_i + \lambda_A + d_A)A_{R_{i+1}} \\
 Q'_{G_{i+1}} &= \alpha_i A_{G_{i+1}} \\
 Q'_{R_{i+1}} &= \alpha_i A_{R_{i+1}}
 \end{aligned} \tag{4}$$

We analyze two variants of the model. In the first variant, activated cells can only become quiescent when they have completed less than or equal to n divisions. In the second variant, we allow quiescent cells to be formed from activated cells throughout the expansion phase regardless of their prior division number (referred to as 'all'). To create a similar number of quiescent cells in all cases we adjust the rate at which quiescent cells are formed. We depict two examples of the first variant in **Extended Data Fig. 1** (left and middle panels): $n = 10$ and $n = 24$ (i.e., in the left panel we set $\alpha_i = 0.25/\text{day}$ when $i < 10$ (and $t < 6$ days), and $\alpha_i = 0$ otherwise, and in the middle panel we set $\alpha_i = 0.1/\text{day}$ when $i < 24$ (and $t < 6$ days), and $\alpha_i = 0$ otherwise). The second variant shown in **Extended Data Fig. 1** (right panel) has the lowest rate at which quiescent cells are formed, $\alpha_i = 0.05/\text{day}$ for all i .

By numerical integration of Eq. (4), we show in **Extended Data Fig. 1** that if the population that is persisting during the memory phase were composed of quiescent cells only, the percentage of DR^{RFP} labeled cells would decrease after the peak, regardless of when quiescent cells appear. Naturally, the decrease in the percentage of DR^{RFP} labeled cells after the peak is smaller when more quiescent cells are formed during the expansion phase. Note that we would not obtain much more quiescent cells if we would increase α_i in the $n = 10$ scenario because a too large α_i cripples the expansion of the activated cells.

Supplementary Note 4

Reduction in DR^{RFP} fractions during recall responses (Fig. 6)

In **Figure 6** we apply the DivisionRecorder to assess the replicative recall potential of high- and low-division memory T cells, *in situ*. Based on the reduction in the fraction of DR^{RFP} cells that we reproducibly observe early upon recall, we conclude that secondary T_{EFF} cells predominantly derive from low-division T_{CM}, a conclusion that is in line with the data obtained using a CTV-based serial transfer approach (**Fig. 5**).

As a potential alternative explanation for the observed reduction in DR^{RFP} fractions, it could be proposed that recall responses would be based on the output of only a very small pool of memory T cells, and the numerical dominance of GFP-positive cells over RFP-positive cells (92.5% versus 7.5%, respectively) would make it likely that such cells would all be GFP-positive, even if replicative recall potential was identical for GFP-positive cells and RFP-positive cells. To determine how small the responding cell pool would have to be to achieve a reproducible drop in the fraction of DR^{RFP} cells without occasional 'jackpot events', in which one of the early responders would be RFP-positive (thereby resulting in a secondary T_{EFF} pool that is largely RFP-positive, something that is not experimentally observed), we modelled memory pools with 7.5% RFP-positive cells, responding to a secondary infection, allowing various numbers of these cells to expand, and then assessed the DR^{RFP} fractions within the resulting T_{EFF} pools (**Figure II**). This analysis indicates that secondary T_{EFF} pools generated from a very small precursor pool (< 20 cells) would show a reduction in DR^{RFP} fraction in the majority of mice, even if replicative recall potential would be equal between low-division and high-division T_{CM} populations. However, models that assume such a tight bottleneck do show the occurrence of jackpot events, an observation that is inconsistent with the experimental data (**Fig. 6**). In addition, an assumption of a responding cell pool of < 20 cells is inconsistent with the widely held view that recall responses are more rapid because of the larger pool of responding cells. Based on this analysis, we conclude that the observed drop in labeling rate during recall responses cannot be explained by T cell expansion during recall responses being driven by a very small pool of reactivated T cells.

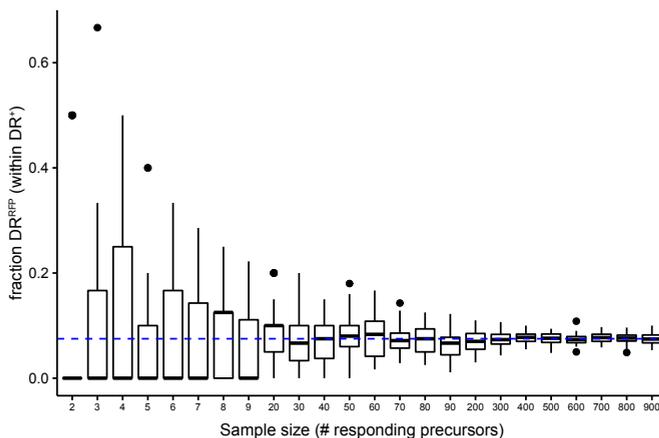


Figure II. Random sampling of precursor cells initiating the recall response. Expected DR^{RFP} fractions in secondary T_{EFF} pools (vertical axis) when a random sample of responding cells of a particular size (horizontal axis) is selected from a memory population that contains 7.5% DR^{RFP} cells (marked by horizontal dashed line). Selected responding cells expand to an equal degree to form the secondary T_{EFF} pool. Simulations were repeated 39 times (i.e., the number of mice in which recall experiments were performed in this manuscript). Boxplots indicate group median and 25th/75th percentiles, whiskers indicate the interquartile range multiplied by 1.5. Outliers are shown as dots.

Supplementary Note 5

Phenotype model

In the main text, we argue that the observed drop in DR^{RFP} percentages is due to the preferential recruitment of lowly divided cells into subsequent responses. Here, we provide support for this proposition by demonstrating that a similar transient drop in DR^{RFP} frequencies is observed in mechanistic mathematical models that have a preferential recruitment of lowly divided cells into the secondary response. In the modeled T cell responses described below we consider two major phenotypic subsets; CD27^{HI}KLRG1^{LO} (hereafter referred to as T_{MULT}) and CD27^{LO}KLRG1^{HI} (hereafter referred to as T_{TERM}).

For an experiment in which DR^{RFP} label flow is asymmetrical and permanent, the DR^{RFP} accumulation in a population can be modelled by Eq. (5); see the cartoon in Extended Data Fig. 8a. We model five phenotypes: clonally expanding activated T_{MULT}s (M_A), clonally expanding activated T_{TERM}s (E_A), cycling T_{MULT}s (M_C), cycling T_{TERM}s (E_C) and quiescent T_{MULT}s (M_Q, M_{nQ}). The variable M_{nQ} (for newQ) keeps track of newly formed quiescent cells to prevent them from becoming re-activated during the same expansion phase. These cells become M_Q during the contraction and memory phase and can be re-activated during the secondary expansion phase.

Upon encountering antigen, a naïve T cell becomes activated, starts dividing and gives rise to a continuum of phenotypically different populations. The phenotype model (**Extended Data Fig. 8a**) broadly classifies this continuum into the five phenotypes introduced above. An activated T_{MULT} can either differentiate into an activated T_{TERM} or can stop dividing and become a quiescent T_{MULT}. After the peak of the response, a small part of the activated T_{MULT}s and T_{TERM}s becomes cycling T_{MULT}s and long-lived T_{TERM}s, respectively. During the recall response, we model a scenario in which a fraction of the cells engages in renewed expansion. Upon secondary antigen encounter (set at day 86 post primary infection, in concordance with the recall experiment presented in **Fig. 6b**), the model undergoes a second sequence of expansion, contraction, and memory formation.

To examine the scenario in which lowly divided cells are preferentially recruited, it was important to know the division history of cells prior to the re-expansion. We, therefore, formulated a division-indexed model (similar to **Eq. 4**) to track the number of divisions of the DR^{GFP} and DR^{RFP} cells of each phenotype over time. The model is described in full in the supplementary R codes. For readability we here present a collapsed version of the model, which can be obtained by summing over the division numbers as well as the DR^{GFP} and DR^{RFP} cells (compare the full model equations shown in **Eq. (6)** to the collapsed equation in **Eq. (5)** for a single phenotype, E_A):

$$\begin{aligned}
\frac{dM_A}{dt} &= \alpha(i)r(M_C + M_Q) + 2\lambda_A M_A - (\lambda_A + d_A + \delta_E(t) + \delta_Q(t) + \mu_C)M_A \\
\frac{dE_A}{dt} &= \delta_E(t)M_A + 2\lambda_A E_A - (\lambda_A + d_A + \mu_C)E_A \\
\frac{dM_C}{dt} &= \mu_C M_A + 2\lambda_M M_C - (\lambda_M + d_M + \alpha(i)r + (1 - \alpha(i))d_C)M_C \\
\frac{dE_C}{dt} &= \mu_C E_A + 2\lambda_E E_C - (\lambda_E + d_E + d_C)E_C \\
\frac{dM_{nQ}}{dt} &= \delta_Q(t)M_A - \mu_Q M_{nQ} \\
\frac{dM_Q}{dt} &= \mu_Q M_{nQ} - (\alpha(i)r + (1 - \alpha(i))d_C)M_Q \\
\delta_E(t) &= \delta_{E_{max}}(1 - e^{-e_a t_{exp}}) \\
\delta_Q(t) &= \delta_{Q_{max}} e^{-m_q t_{exp}} \\
\alpha_i &= \left(1 + \left(\frac{i}{h}\right)^m\right)^{-1} ; i = 1, \dots, n
\end{aligned} \tag{5}$$

where t_{exp} (for time in expansion) is the time since the most recent challenge. The fraction of cells that is expected to be re-activated, $0 \leq \alpha(i) \leq 1$, is a function of the division number, i . Each phenotype is actually indexed by the number of divisions completed, i , and is subdivided into an unlabeled (DR^{GFP}, R) and a labeled (DR^{RFP}, G) subtype. For instance, the full division-indexed model for activated T_{TERM}s is:

$$\begin{aligned}
\frac{dE_{A_{G_i}}}{dt} &= \delta_E(t)M_{A_{G_i}} + (2 - p)\lambda_A E_{A_{G_{i-1}}} - (\lambda_A + d_A + \mu_C)E_{A_{G_i}} \\
\frac{dE_{A_{R_i}}}{dt} &= \delta_E(t)M_{A_{R_i}} + p\lambda_A E_{A_{G_{i-1}}} + 2\lambda_A E_{A_{R_{i-1}}} - (\lambda_A + d_A + \mu_C)E_{A_{R_i}}
\end{aligned} \tag{6}$$

where p is the switching probability. A complete list of division-indexed phenotype specified equations can be found in the R code. Below, we explain the step-wise parameter estimation procedure of this model.

Fitting the model and estimating the parameters

Two phases of immune response were inferred from the experimental data: the expansion phase (day 0-6 for the primary response and day 86-90 for the secondary response), and the non-expansion phase, i.e., the contraction and memory phases (day 6-86 for the primary response and day 90-111 for the secondary response). The percentages of DR^{RFP} cells in blood and spleen were found to be highly similar. Additionally, the ratio of the number of DR^{GFP} cells in spleen to that in blood was similar across all time points. The blood and spleen data were therefore fitted simultaneously under the assumption that both compartments are well-mixed (**Fig. 7a-c, Extended Data Fig. 8**).

Our primary interest was to find a realistic division history of all phenotypes on day 86, to subsequently test whether the preferential recruitment of lowly divided cells can explain the kinetics of labeling rate during recall responses. The number of free model parameters was reduced to 6 using a few simplifying assumptions:

1. The division rate during the expansion phase is the same for T_{MULT} s and T_{TERM} s,
2. The death rate of cells during the expansion phase is negligible,
3. The formation of quiescent cells decreases with time post antigen encounter,
4. The rate at which T_{TERM} s are formed increases with time (instead of division number),
5. Only 5% of the activated cells survive after the expansion phase, i.e., $\mu_c = \frac{d_A}{19}$, and
6. T_{MULT} s and T_{TERM} s maintain constant numbers during the memory phase, i.e., $\lambda_M = d_M$ and $\lambda_E = d_E$.

The free parameters were estimated by fitting the collapsed model sequentially to the data obtained during the primary expansion and non-expansion phases. First, the initial number of cells that were activated among the grafted cells in blood and spleen, the division rate of the activated cells, and the differentiation rate of the activated cells into the different phenotypes were estimated using the experimental data obtained during the expansion phase. Next, these estimated parameters were used to estimate the division rates of the cycling cells from the experimental data obtained during the non-expansion phase. Because the estimation of the re-activation function was infeasible, as it would require fitting the full division-indexed model to the data, this function was tuned manually to obtain an optimal description of the data with the full model. The estimated parameters for the best description of the data are listed in **Table S1**.

Expansion of adoptively transferred DR⁺ cells is expected to occur after a short delay, covering both the time required to identify an antigen-positive APC and to initiate cell division after TCR triggering. This delay was fixed to 1 day as our dataset lacked the appropriate information for this parameter to be estimated. The loss rate of activated cells was fixed to d_A 0.3/day. The rate at which the formation of quiescent cells declines was tuned such that only 1% of the quiescent population

formed during the primary response would be DR^{RFP} . As antigen-experience is known to influence secondary memory formation, e.g. the expansion phase occurring on a shorter time-scale and a considerably slower contraction phase, the rates during the secondary response differ from those during the primary response³. Therefore, m_q and μ_c and were set to 0.5/day and $\frac{d_A}{4}$, respectively.

The maximum rate at which quiescent cells were formed was fixed to different values

($\delta_{Q_{max}} \in \{0.01, 0.1\}$) to generate different numbers of quiescent cells ($10^3, 10^4$) at the peak of the primary response (day 6) (**Fig. 7a**). Unsurprisingly, formation of a larger number of quiescent cells during the primary response, resulted in a larger drop in DR^{RFP} frequencies during secondary expansion (**Fig. 7b**). Higher numbers of quiescent T_{CM} also generated secondary responses that were higher in magnitude, providing a better explanation of the data (**Extended Data Fig. 8b-d**). Three different re-activation functions, signifying either an abrupt (fun 1-2) or gradual loss (fun 3) of re-expansion potential based on the number of prior divisions, were tested (**Fig. 7a**). The experimental data was only congruent with scenarios where re-expansion potential was restricted to cells that had undergone limited clonal expansion (**Fig. 7a-b**). Furthermore, higher numbers of quiescent T_{CM} correlated positively with larger drops in DR^{RFP} frequencies upon re-expansion (**Fig. 7b**).

Parameter	Description	Value	Units	On (1)/Off (0) during phases	
				Expansion	Non-expansion
t_{on}	Time delay before the first division	1 (F)	days	-	-
λ_A	Division rate of activated T_{MULT} and T_{TERM} cells	0.89	/day	1	0
d_A	The rate at which (re-) activated T_{MULT} and T_{TERM} cells leave circulation	0.3 (F)	/day	0	1
d_C				1	0
δ_{Qmax}	Maximum differentiation rate of the activated T_{MULT} into quiescent T_{MULT} cells	0.1 (F)	/day	1	0
m_q	The rate at which $\delta_Q(t)$ changes	0.25 (F)	/day	-	-
δ_{Emax}	Maximum differentiation rate of the activated T_{MULT} into activated T_{TERM} cells	2 (F)	/day	1	0
e_a	The rate at which $\delta_E(t)$ changes	0.15	/day	-	-
$\lambda_M = d_M$	Division and death rate of the cycling T_{MULT} cells	0.3	/day	1	1
$\lambda_E = d_E$	Division and death rate of the cycling T_{TERM} cells	0.04	/day	1	1
r	The rate of re-activation of quiescent and cycling T_{MULTS}	1 (F)	/day	1	0
h	Cut-off of the fraction re-activated function	25 (F)	-	-	-
m	Slope of the fraction re-activated function	30 (F)	-	-	-
$M_{AG}(0)$	Initial number of naïve cells at day 0	16748.72 (Spleen)	Cells (avg.)	-	-
		149.30 (Blood)			

Table S1: The estimated parameters for the best fit of the phenotype model to the data. The parameter values in this table were obtained by fitting the phenotype model to the blood and spleen data simultaneously using the pseudorandom-search algorithm (see pseudoOptim) in the modFit function of the FME R package⁴. F signifies that these parameters were set to a fixed value.

Supplementary Note 6

Analysis of DivisonRecorder single integration frequency

As the DivisonRecorder is retrovirally introduced into the genome, a fraction of the modified cells may carry multiple integrations. As each DivisonRecorder has an independent probability of slippage—and hence creation of an in-frame *Cre* gene—during cell division, this means that cells that contain more than 1 integration will have a 'faster clock', resulting in the more rapid labeling of these cells. However, as the DivisonRecorder is applied as a population-based metric, and conclusions are based on comparison of different timepoints and/or different cell populations within individual mice, the presence of a fraction of cells with multiple integration events will not influence the interpretation of the obtained data. Nevertheless, to minimize variation in the fraction of cells with >1 integration event between experiments, we aimed for a low and standardized transduction efficiency, in which the occurrence of multiple integration events will be minor. To determine which fraction of single integrations could be expected as a factor of transduction efficiency, Ai9 mouse embryonic fibroblasts (MEFs) or ex vivo activated Ai9;OT-I cells were transduced with a mixture of 2 retroviruses encoding either GFP or Katushka. The fraction of single and double-positive cells could subsequently be used to estimate the relationship between transduction efficiency and the percentage of single integrations. This analysis shows that at a transduction efficiency of ~10-15% (the transduction efficiency used for in vivo experiments), approximately 85-90% of the modified cells contains a single integration, and this percentage is comparable between the two cell types assessed (**Figure III**). Thus, the large majority of switch events we observe in our experiments derives from cells carrying a single reporter.

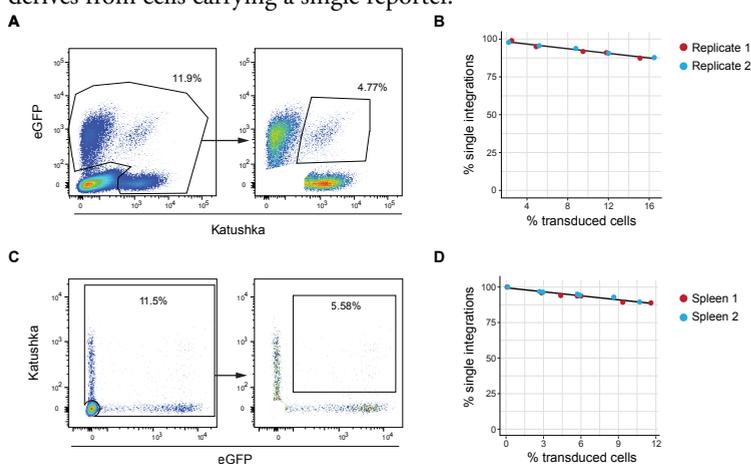


Figure III. Frequency of multiple retroviral integration events. GFP and Katushka encoding retroviruses were mixed 1:1, and subsequently used to transduce either Ai9 MEF cells (**A-B**) or ex vivo activated Ai9;OT-I cells (**C-D**). Ai9 MEF and Ai9;OT-I cells were assessed for fluorescent protein expression at day 7 or 24 hours post transduction, respectively. The percentage of cells carrying a single integration was calculated as $100 - (2 \times \text{fraction GFP}^+\text{Katushka}^+ \text{ cells})$ within the transduced Ai9 MEF cell population. **A**) Gating strategy to determine the percentage of GFP⁺Katushka⁺ cells within the transduced Ai9 MEF cell population. **B**) Plot depicting the percentage of Ai9 MEF cells carrying a single retroviral integration at different transduction efficiencies. **C**) Gating strategy to determine the percentage of GFP⁺Katushka⁺ cells within the transduced Ai9;OT-I cell population. **D**) Plot depicting the percentage of Ai9;OT-I cells carrying a single retroviral integration at different transduction efficiencies. Depicted data was obtained in a single experiment consisting of two experimental replicates. Dots indicate individual samples, lines represent a linear regression fitted to the data points (**B, D**).

Supplementary Note 7

Retroviral silencing of the DivisionRecorder does not occur

As the GFP-Cre module of the DivisionRecorder is introduced retrovirally, there is a potential risk of retroviral transcriptional silencing or attenuation. Such silencing events could influence interpretation of the data, as the fraction of DR^{RFP} cells over DR^{GFP} cells may become skewed. As only the GFP-Cre module is retrovirally introduced, whereas the RFP reporter that is switched on upon Cre activity is germline encoded, the extent of retroviral silencing can be experimentally determined by measuring the occurrence of cells that do show RFP expression (and hence did at some point express the GFP-Cre module) but lack GFP expression. As depicted in **Figure IV**, virtually no RFP⁺GFP⁻ cells are observed within recipient mice, either during the acute phase or in the memory phase, demonstrating that retroviral silencing is extremely rare.

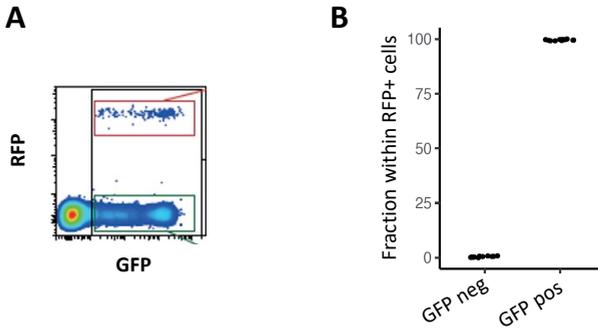
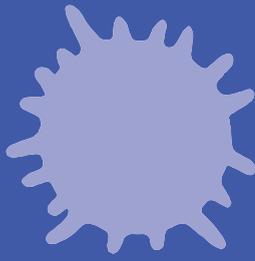
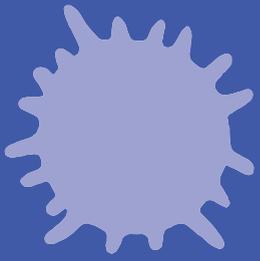


Figure IV. GFP expression of RFP⁺ cells. LM-OVA infected recipient mice received 20,000 DR⁺ cells and the occurrence of GFP⁻RFP⁺ cells was assessed in spleen on day 6 and 86 after adoptive cell transfer. **A)** Flow cytometry plot, gated on CD45.2⁺ cells, depicting RFP and GFP expression. **B)** Fraction of GFP⁻ and GFP⁺ cells within the RFP⁺ cell population (n=12 mice).

References Supplementary Notes

1. Weber, T. S., Perié, L. & Duffy, K. R. Inferring average generation via division-linked labeling. *J Math Biol* **73**, 491–523 (2016).
2. De Boer, R. J. & Perelson, A. S. Quantifying T lymphocyte turnover. *J. Theor. Biol.* **327**, 45–87 (2013).
3. Masopust, D., Ha, S.-J., Vezys, V. & Ahmed, R. Stimulation History Dictates Memory CD8 T Cell Phenotype: Implications for Prime-Boost Vaccination. *J Immunol* **177**, 831–839 (2006).
4. Soetaert, K. & Petzoldt, T. Inverse Modelling, Sensitivity and Monte Carlo Analysis in R Using Package **FME**. *J. Stat. Soft.* **33**, (2010).



Chapter 5.

QPCTL regulates macrophage and monocyte abundance and inflammatory signatures in the tumor microenvironment

OR: "Practising pronouncing glutaminy-peptide cyclotransferase in the bathroom mirror before every group meeting"

Kaspar Bresser^{1*}, Meike E. W. Logtenberg^{1*}, Mireille Toebes¹, Natalie Proost², Justin Sprengers², Bjorn Siteur², Manon Boeije², Lona J. Kroese³ and Ton N. Schumacher^{1,2}

Published in Oncoimmunology, March 2022

(1) Division of Molecular Oncology & Immunology, Onco Institute, The Netherlands Cancer Institute, Amsterdam, The Netherlands

(2) Preclinical Intervention Unit, Mouse Clinic for Cancer and Ageing, The Netherlands Cancer Institute, Amsterdam, The Netherlands

(3) Transgenic Facility, Mouse Clinic for Cancer and Aging Research, The Netherlands Cancer Institute, Amsterdam, The Netherlands.

(*) These authors contributed equally

Abstract

The enzyme glutamyl-peptide cyclotransferase-like protein (QPCTL) catalyzes the formation of pyroglutamate residues at the NH₂-terminus of proteins, thereby influencing their biological properties. A number of studies have implicated QPCTL in the regulation of chemokine stability. Furthermore, QPCTL activity has recently been shown to be critical for the formation of the high affinity SIRP α binding site of the CD47 “don’t-eat-me” protein. Based on the latter data, interference with QPCTL activity—and hence CD47 maturation—may be proposed as a means to promote anti-tumor immunity. However, the pleiotropic activity of QPCTL makes it difficult to predict the effects of QPCTL inhibition on the tumor microenvironment (TME). Using a syngeneic mouse melanoma model, we demonstrate that QPCTL deficiency alters the intratumoral monocyte-to-macrophage ratio, results in a profound increase in the presence of pro-inflammatory cancer-associated fibroblasts (CAFs) relative to immunosuppressive TGF- β 1-driven CAFs, and leads to an increased IFN and decreased TGF- β transcriptional response signature in tumor cells. Importantly, the functional relevance of the observed TME remodeling is demonstrated by the synergy between QPCTL deletion and anti PD-L1 therapy, sensitizing an otherwise refractory melanoma model to anti-checkpoint therapy. Collectively, these data provide support for the development of strategies to interfere with QPCTL activity as a means to promote tumor-specific immunity.

Introduction

Regulation of immune cell activity at sites of infection or cancer growth frequently occurs through a balance of signals that are received by immune activating and immune inhibitory receptors¹. For example, while activation of myeloid cells, including neutrophils, macrophages and monocytes, often occurs through Fc receptor signaling, such activation can be prevented through the simultaneous engagement of ITIM/ITSM-containing inhibitory receptors such as SIRP α . Specifically, binding of the “don’t-eat-me” signal CD47, which is widely expressed on hematopoietic and non-hematopoietic cells, to the SIRP α receptor on myeloid cells has been shown to result in decreased myeloid effector function, including suppression of target cell phagocytosis by macrophages and tumor cell killing by neutrophils²⁻⁴.

The inhibitory capacity of CD47 is dependent on the maturation of its SIRP α binding site by the ER-resident enzyme QPCTL^{5,6}. Similar to its secreted family member QPCT, QPCTL catalyzes the cyclization of N-terminal glutamine and glutamic acid residues on target proteins into a pyroglutamate residue (pGlu)^{7,8}. As shown by structural analysis, the pGlu residue at the N-terminus of CD47 contributes to the interaction surface with SIRP α ⁹; and, through genetic screening, it was shown that the activity of QPCTL is critical for the formation of this residue, making this enzyme a key regulator of the high-affinity CD47-SIRP α binding site⁶. In line with this, prevention

of pGlu formation on CD47, either by genetic knock-out or small molecule inhibition, leads to reduced SIRP α binding and increased macrophage- and neutrophil-dependent killing of antibody-opsionized target cells. Based on its role in regulating CD47-SIRP α signaling, and the possibility to develop small molecule inhibitors of enzymatic activity, QPCTL forms a potentially interesting target in cancer immunotherapy.

In addition to CD47, chemokines such as CCL2 and CX₃CL1 have been identified as QPCTL and/or QPCT substrates¹⁰⁻¹². The formation of the N-terminal pGlu on CCL2 was shown to increase its *in vivo* activity, both by conferring resistance to aminopeptidases and by increasing its capacity to induce CCR2 signaling¹⁰. Likewise, pGlu-modified CX₃CL1 appears to show an increased capacity to promote CX₃CR1 signaling *in vitro*¹². Finally, around 600 human proteins harbor a N-terminal glutamine or glutamic acid residue after the predicted signal peptide cleavage site, and it is plausible that additional QPCTL/QPCT substrates exist amongst this group of proteins¹³.

Because of its known or potential role in the post-translational modification of different immune- or tumor cell-related molecules, it is difficult to predict the overall effects of QPCTL inhibition on the tumor microenvironment (TME), and the poor pharmacokinetics of available QPCT/QPCTL inhibitors such SEN177⁶ has precluded evaluation of such effects by small-molecule inhibition. To address this question, we have generated a QPCTL-deficient mouse model and combined it with QPCTL-deficient tumor cells, to map the effects of QPCTL deficiency on either cellular compartment on the composition of the TME. The obtained data reveal that QPCTL deficiency results in a skewing of the macrophage-monocyte ratio, causes an approximately 20-fold change in the balance between TGF- β -producing myofibroblastic cancer-associated fibroblasts (myCAFs) and cytokine-secreting inflammatory CAFs (iCAFs), and shifts tumor cells from a TGF- β -responding to an IFN-responding state. Collectively, these changes convert the TME to pro-inflammatory environment that sensitizes tumors to anti-PD-L1 therapy. Together, these results suggest that therapeutic manipulation of QPCTL activity may synergize with current cancer immunotherapies.

Results

Genetic deletion of QPCTL does not significantly alter the immune compartment

Previous work has indicated that genetic deletion of QPCTL can be used to study its function *in vivo*^{10,14}; however, the possible effects of QPCTL inactivation on the TME has not been studied. To determine how blockade of QPCTL activity alters the TME, we applied CRISPR-Cas9 mediated gene-editing to generate a QPCTL-deficient (QPCTL^{-/-}) C57BL/6 mouse strain that is compatible with commonly used tumor models. To first determine whether QPCTL deficiency results in abrogation of pGlu formation on CD47, peripheral blood cells from QPCTL^{-/-} mice and WT littermates were stained with mouse SIRP α and an anti-mouse CD47 antibody that recognizes CD47 independent of pyroglutamate formation⁶. As compared to WT littermates, blood cells of

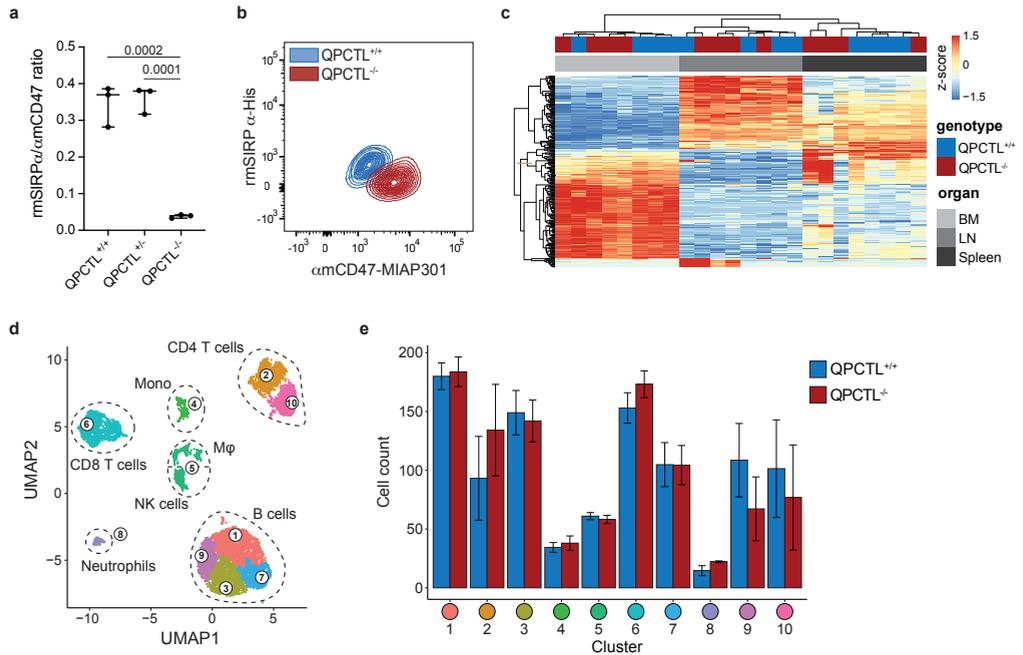


Figure 1. Generation and characterization of QPCTL-deficient mice. (a) Ratio of recombinant mouse (rm) SIRP α -His and anti-mouse (am)CD47 antibody (clone MIAP301) binding to blood cells from QPCTL^{+/+}, QPCTL^{+/-}, and QPCTL^{-/-} mice, as measured by flow cytometry. Dots depict the ratio of rmSIRP α -His/amCD47-MIAP301 mean fluorescence intensity (MFI) on blood cells from individual mice, group medians are indicated and whiskers represent min/max, $n = 3$ mice per group. (b) Flow cytometry plot depicting data described in a) for blood cells from a representative QPCTL^{+/+} and QPCTL^{-/-} mouse. (c) Heatmap depicting hierarchical clustering performed on the 1,000 most differentially expressed genes in bone marrow (BM), lymph node (LN) and spleen samples from QPCTL^{+/+} and QPCTL^{-/-} mice. (d-e) Unbiased Euclidean distance-based clustering of immune cells obtained from spleens of QPCTL^{+/+} and QPCTL^{-/-} mice. UMAP 2-dimensional projection (d) depicts the obtained clusters. Cell counts of both genotypes within each cluster are depicted (e). Bars indicate group means, error bars represent standard error of the mean. P values were determined by one-way ANOVA followed by Tukey's HSD test (a) or by two-sided Student's T test with Bonferroni correction for multiple testing (e). Significant P values (< 0.05) are indicated in the plots. Data are representative of 3 independent experiments (a-b), or were obtained in a single experiment (c-d). UMAP, Uniform Manifold Approximation and Projection.

QPCTL-deficient mice displayed significantly decreased SIRP α binding, thereby providing the first evidence that QPCTL is also a critical CD47 modifier *in vivo* (Fig. 1a-b).

To understand how QPCTL deficiency influences steady-state immune cell frequencies and gene expression, QPCTL^{-/-} and QPCTL^{+/+} littermates were subjected to histopathological, transcriptomic and flow cytometric analysis. Histopathological assessment of QPCTL^{-/-} mice revealed no significant morphological aberrations relative to littermate controls (Supplementary Data 1), and gene expression analysis of spleen, lymph nodes and bone marrow revealed no genotype-specific transcriptional changes (Fig. 1c, Supplementary Fig 1a), indicating that QPCTL deficiency does not result in major alterations in steady-state immune activity. Likewise, no substantial differences in cell counts or immune cell population frequencies were observed in blood (Supplementary Fig. 1b-c). In spleen, a modest increase in the frequency of NK cells of total non-myeloid cells and a decrease in the fraction of activated cells of total CD4⁺ T cells was observed in QPCTL^{-/-} mice, but

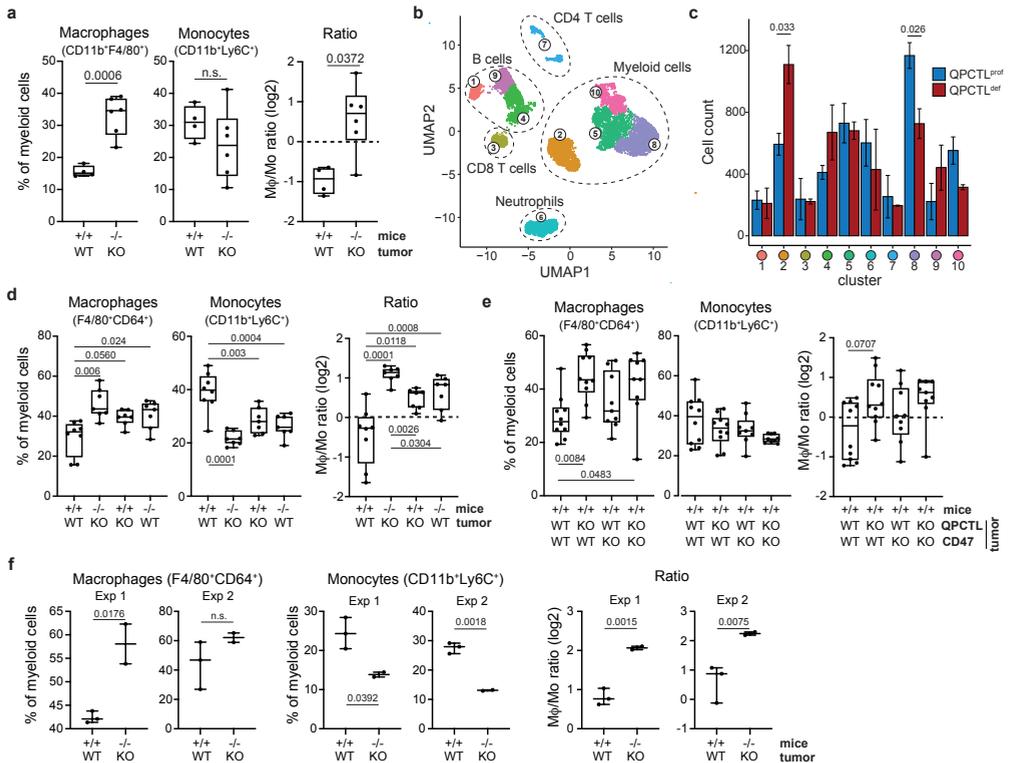


Figure 2. Tumor and host QPCTL deficiency alters the M ϕ -Mo ratio in the TME. (a) Frequency of macrophages and monocytes of myeloid cells (CD11b⁺), and M ϕ -Mo ratio, in the TME of QPCTL^{+/+} ($n = 4$) and QPCTL^{-/-} ($n = 6$) mice inoculated with QPCTL-WT or QPCTL-KO B16F10 melanoma cells, respectively. Tumors were analyzed between 14–16 days post inoculation. (b) UMAP visualizing 30,000 cells sampled from the data shown in a. 5,000 cells were randomly drawn from each sample ($n = 3$ mice per group) prior to analysis. Colors indicate clusters obtained by Euclidean distance-based hierarchical clustering, cluster phenotype is shown in **Supplementary Figure 2f**. (c) Contribution of cells from QPCTL^{+/+} ($n = 3$) and QPCTL^{-/-} ($n = 3$) TMEs to each cluster shown in b. Bars indicate group means, error bars represent standard error of the mean. (d) Frequency of macrophages and monocytes of myeloid cells (CD11b⁺), and M ϕ -Mo ratio, in the TME of QPCTL^{+/+} and QPCTL^{-/-} mice inoculated with either QPCTL-WT or QPCTL-KO B16F10 melanoma cells ($n = 7$ –8 per group). Tumors were analyzed between 12–14 days post inoculation. (e) Frequency of macrophages and monocytes of myeloid cells (CD11b⁺), and M ϕ -Mo ratio, in the TME of QPCTL^{+/+} mice inoculated with QPCTL-WT, QPCTL-KO, CD47-KO, or CD47/QPCTL double-KO (dKO) B16F10 cells. Tumors were analyzed between 14–16 days post inoculation. (f) Frequency of macrophages and monocytes of myeloid cells (CD11b⁺), and M ϕ -Mo ratio, in the TME of QPCTL^{+/+} and QPCTL^{-/-} mice inoculated with QPCTL-WT and QPCTL-KO MC38 cells, respectively. Data from 2 independent experiments are shown ($n = 5$ per experiment). Tumors were analyzed at 22 (experiment 1) or 29 (experiment 2) days post inoculation. Dots indicate measurements from individual mice, group medians are indicated and whiskers represent min/max. P values were determined by two-sided Student's T test without (a, f) or with Bonferroni correction for multiple testing (c), or by one-way ANOVA followed by Tukey's HSD test (d, e). Significant P values (< 0.05) are indicated in the plots. For all boxplots, dots represent individual mice, group median and 25th/75th percentiles are indicated by the box, whiskers indicate min/max. Data are representative of at least 2 independent experiments (a, d, f), or were obtained in a single experiment (b, c, e). UMAP, Uniform Manifold Approximation and Projection.

no significant changes in other immune cell type frequencies were identified (**Supplementary Fig. 1d**). The absence of substantial differences in immune cell frequencies was corroborated by unbiased hierarchical clustering of cells obtained from QPCTL^{+/+} and QPCTL^{-/-} spleen samples (**Fig. 1d** and **Supplementary Fig. 1e**).

QPCTL-deficiency alters macrophage-monocyte-ratios in the TME

To test whether systemic QPCTL deficiency influences immune cell infiltration in the TME, QPCTL^{+/+} and QPCTL^{-/-} mice were inoculated with wild-type (QPCTL-WT) and QPCTL knock-out (QPCTL-KO) B16F10 melanoma cells (**Supplementary Fig. 2a**), respectively, and TMEs were analyzed by flow cytometry 14–16 days post inoculation. QPCTL-proficient and -deficient tumors grew with similar kinetics and were similarly infiltrated by large numbers of myeloid cells (**Supplementary Fig. 2b-c**). Importantly, within the myeloid subset, QPCTL-deficient TMEs exhibited a significant higher frequency of macrophages and a substantially increased macrophage-monocyte (M ϕ -Mo) ratio (**Fig. 2a**). Although no other significant changes within the immune infiltrate could be detected (**Supplementary Fig. 2d**), an increase in the frequency of B cells, a trend toward a decrease in monocytes, and a decrease in the frequency of CD4⁺ T cells within the non-myeloid immune cell subset was observed in peripheral blood samples from QPCTL-deficient tumor-bearing mice (**Supplementary Fig. 2e**). Also, when immune infiltrates of QPCTL-deficient and -proficient TMEs were analyzed through unbiased hierarchical clustering, an increase in F4/80⁺ cells (macrophages) and a decrease in Ly6C^{high} cells (monocytes) in QPCTL-deficient TMEs was observed (**Fig. 2b-c, Supplementary Fig. 2f**).

To determine whether the increased M ϕ -Mo-ratio could be attributed to a lack of QPCTL activity in either host or tumor cells, QPCTL^{+/+} and QPCTL^{-/-} mice were inoculated with either QPCTL-WT or QPCTL-KO melanoma tumor cells (**Supplementary Fig. 3a**). Both tumor and host QPCTL deficiency led to an increased M ϕ -Mo-ratio, but the most profound increase in M ϕ -Mo-ratios was observed when QPCTL activity was lacking in both cell compartments (**Fig. 2d, Supplementary Fig. 3b-c**). In blood, the most pronounced differences in immune cell frequencies were found when comparing tumor-bearing versus non-tumor-bearing animals—independent of QPCTL activity—emphasizing that QPCTL deficiency does not impact the systemic immune compartment in a major way (**Supplementary Fig. 3d**). To explore to what extent loss of pGlu-modified CD47 contributed to the altered intra-tumoral M ϕ -Mo balance, wild-type mice were inoculated with QPCTL-, CD47- or double-KO tumor cells. Absence of CD47 resulted in M ϕ frequencies that were numerically higher than observed in recipients of WT B16 tumor cells, but to a lower extent than observed in recipients of QPCTL-KO cells, and did not significantly alter M ϕ -Mo-ratios (**Fig. 2e**). This suggests that, if the CD47-modifying activity of QPCTL contributes to the altered myeloid cell composition in these tumors, it likely plays a minor role.

To test whether the role of QPCTL as a modifier of the TME extended to other tumor models, QPCTL^{+/+} and QPCTL^{-/-} mice were inoculated with QPCTL-WT and QPCTL-KO MC38 colon carcinoma cells, respectively. Following tumor outgrowth, a profound increase in M ϕ -Mo-ratio was observed in QPCTL-deficient TMEs (**Fig. 2f, Supplementary Fig. 4a-b**), while no significant changes were detected in other immune cell subsets (data not shown). Together, these data indicate that QPCTL affects the TME composition in at least two different tumor models, and that combined tumor and host QPCTL deficiency leads to a significant increase in M ϕ -Mo-ratio.

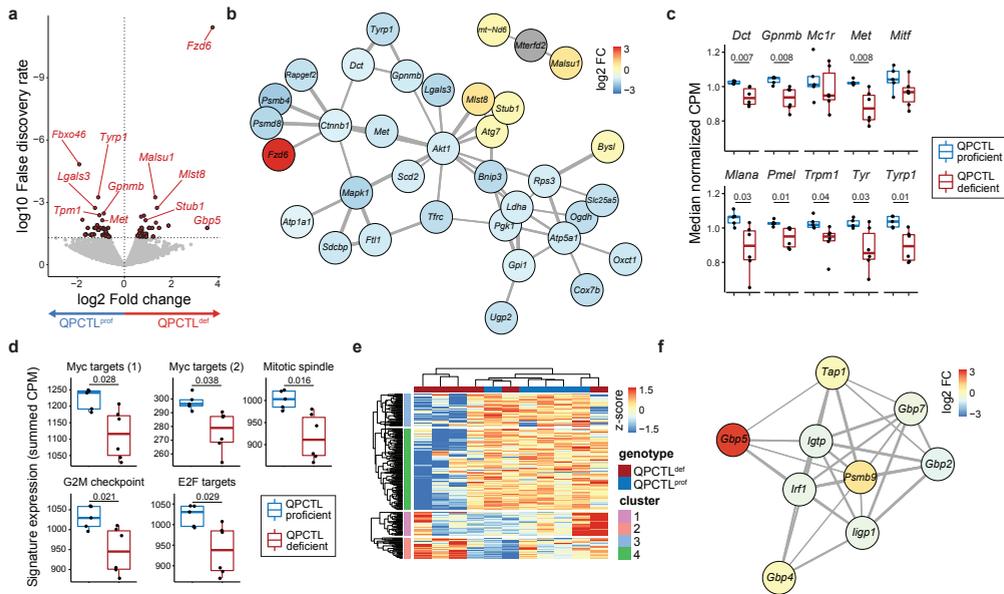


Figure 3. QPCTL deficiency results in suppression of melanogenesis and cell metabolism. mRNA sequencing was performed on sorted CD45-negative cells from QPCTL-proficient ($n = 5$) and QPCTL-deficient ($n = 6$) B16F10 TMEs. Tumors were harvested at day 14 post inoculation. **(a)** Differential gene expression analysis comparing CD45-negative cells obtained from QPCTL deficient versus QPCTL-proficient TMEs. Genes with a false discovery rate (FDR) < 0.05 are indicated in red. Selected genes are indicated in the plot. **(b)** Network analysis (StringDB) performed on all significantly (FDR < 0.05) differentially expressed genes. Genes with a medium interaction strength (> 0.4) are included. Line thickness indicates interaction strength. Nodes are colored based on log₂ fold differences obtained in **a**. **(c)** Transcript abundance of selected genes in the melanogenesis pathway. Boxplots indicate group median and 25th/75th percentiles, whiskers indicate the interquartile range multiplied by 1.5, dots signify individual samples. **(d)** Signature expression of cell cycle-associated hallmark signatures from MSigDB, calculated as the summed CPM of all genes within each signature. Boxplots indicate group median and 25th/75th percentiles, whiskers indicate the interquartile range multiplied by 1.5, dots signify individual samples. **(e)** Hierarchical clustering of the 1,000 most differentially expressed genes across all samples, depicted as a row-normalized heatmap. **(f)** Network analysis (StringDB) performed on genes from cluster 2 **(e)**. Genes with a medium interaction strength (> 0.4) are included. Line thickness indicates interaction strength. Nodes are colored based on log₂ fold differences obtained in **a**. P values were determined by one-way ANOVA followed by Tukey's HSD test **(c, d)**. Significant P values (< 0.05) are indicated in the plots. Data are representative of 2 independent experiments. CPM, counts per million; MSigDB, Molecular Signatures Database.

QPCTL deficiency is associated with suppressed intra-tumoral melanogenesis and cell metabolism

Having established that QPCTL deficiency is associated with an alteration in intra-tumoral immune cell composition, we set out to investigate the effect of QPCTL deficiency on the tumor cell and stromal cell compartment of the TME. RNA sequencing of CD45-negative cell fractions from QPCTL-deficient and -proficient TMEs showed that QPCTL deficiency resulted in differential expression of a substantial set of genes (**Fig. 3a**, **Supplementary Fig. 5a**), and network analysis revealed that expression of multiple genes involved in melanogenesis (*Dct*, *Tyrp1*, *Gpnmb*) was reduced in QPCTL-deficient melanomas (**Fig. 3b**). Assessment of the expression level of a broader set of genes involved in melanogenesis likewise showed dampening of this pathway (**Fig. 3c**). Interestingly, these transcriptional changes coincided with the functional abrogation of melanin production, as evidenced by a loss of pigmentation of QPCTL-deficient tumors (**Supplementary**

Fig. 5b). Expression of a network of genes involved in cell cycle (*Mapk1*, *Akt1*) and cell metabolism (*Pgk1*, *Atp5a1*, *Oxct1*) was additionally found to be decreased in CD45-negative cells in the QPCTL-deficient setting (**Fig. 3b, d, Supplementary Fig. 5c**). While dampened expression of these gene sets was consistently observed, effect sizes were small and had no discernible effect on tumor outgrowth (**Supplementary Fig. 2b-c**).

To further explore putative transcriptional alterations in the CD45-negative compartment as a result of QPCTL deletion, obtained transcriptomes were clustered based on the top 1,000 most variable genes, revealing a cluster of genes that was enriched in 3 out of 6 QPCTL-deficient samples (gene cluster 2, **Fig. 3e**). Network analysis performed on this cluster showed that this gene set contained a small network comprised of IFN induced genes (*Gbp* family members, *Tap1*, *Irf1*; **Fig. 3f**). Interestingly, previous studies have shown that IFN γ can act as a suppressor of melanogenesis in B16F10 melanoma cells^{15–17}, a notion that potentially links this IFN signature to the observed decrease in tumor pigmentation.

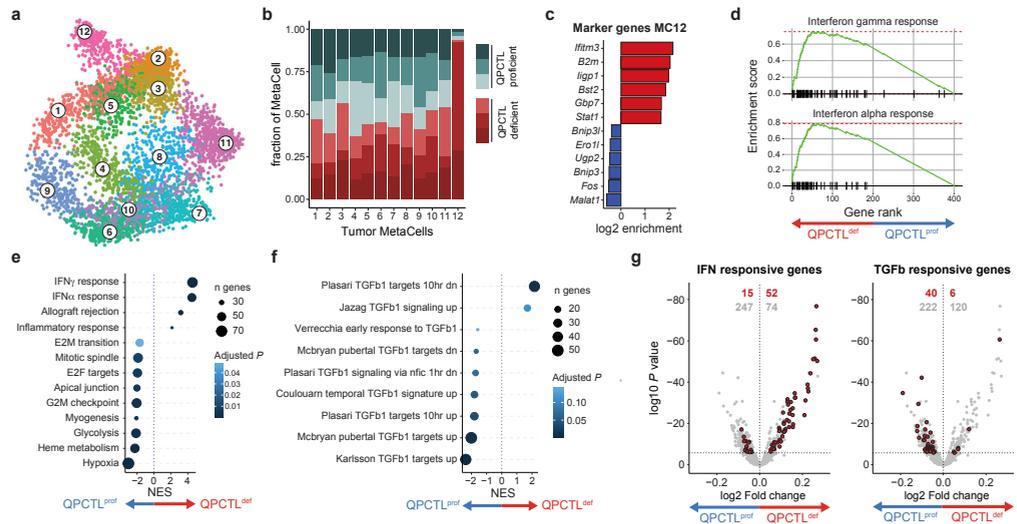


Figure 4. QPCTL deficiency leads to an increased IFN- and decreased TGF-b-response signature in tumor cells. scRNA sequencing was performed on sorted live cells from QPCTL-proficient ($n = 3$) and QPCTL-deficient ($n = 3$) B16F10 TMEs. Tumors were harvested at day 14 post inoculation. (a) 2-dimensional MetaCell projection of the tumor cell compartment. Single cells are colored by MetaCell. (b) Stacked barchart depicting the sample composition of each tumor cell MetaCell. Cell counts from each sample were normalized to 1,000 cells. (c) Enrichment of marker genes (6 highest and lowest expressed) in tumor cell MetaCell 12. (d) Gene set enrichment analysis performed on the top and bottom 200 genes expressed by MC12 (see **Supplementary Fig. 7b**). Gene-enrichment plots for the IFN γ and IFN α response gene-sets are depicted. (e-f) Differential gene expression analysis comparing tumor cells derived from QPCTL-proficient and QPCTL-deficient TMEs, followed by gene set enrichment analysis using either hallmark (e) or immunologic (f) gene sets from MSigDB. Results obtained from the immunologic signature gene sets were filtered for those containing "TGFb". Gene sets with a $P < 0.05$ are shown. (g) Volcano plots depicting differential gene expression analysis. Horizontal line indicates an adjusted P value cutoff of 0.05. IFN (left) or TGF-b (right) signature genes are highlighted in red (see **Supplementary Table 1** for signature genes). Red numbers denote quantity of significant differentially expressed genes within the signature, grey numbers denote the quantity of remaining differentially expressed genes. Depicted data were obtained in a single experiment, consisting of 6 mice. NES, normalized enrichment score; MSigDB, Molecular Signatures Database.

Single cell transcriptomic profiling reveals remodeling of the tumor micro-environment by QPCTL deficiency

The above data indicate that QPCTL inactivation affects both immune cell composition of the TME and the transcriptome of non-immune cells at the tumor site. With the aim to potentially link these two observations we inoculated QPCTL^{+/+} and QPCTL^{-/-} mice with QPCTL-WT and QPCTL-KO B16F10 cells, respectively, and performed scRNAseq of both immune cells and non-immune cells from the resulting QPCTL-deficient and -proficient TMEs. Applying the MetaCell algorithm¹⁸ on 13,093 transcriptomes derived from 6 TMEs (3 QPCTL-proficient, 3 QPCTL-deficient), showed 3 transcriptionally divergent cell supertypes, reflecting immune cells (*Ptprc*, *Irgam*), fibroblasts (*Col1a1*, *Acta2*), and tumor cells (*Mlana*, *Pmel*) (**Supplementary Fig. 6**). To identify cell type-specific changes that accompany QPCTL deficiency, each of these supertypes was subsequently re-clustered and analyzed separately.

To investigate which of the transcriptional changes observed in the CD45-negative compartment could be mapped to the tumor cell compartment, this supertype was re-grouped into transcriptionally disparate MCs (**Fig. 4a, Supplementary figure 7a**), and the relative contribution of cells from either QPCTL-deficient mice or QPCTL-proficient samples to the different MCs was examined. Strikingly, 1 MC (MC12) was nearly exclusively observed in QPCTL-deficient samples (**Fig. 4b**). Examination of the marker genes of MC12 showed a prominent presence of IFN induced transcripts *Ifitm3*, *B2m*, *Bst2* and *H2-D1* (**Fig. 4c**), and gene-set enrichment analysis performed on MC12 marker genes identified both IFN γ and IFN α response as the strongest enriched gene-sets (**Fig. 4d, Supplementary figure 7b-c**). IFN γ and IFN α/β response signatures show a high level of overlap¹⁹⁻²¹, making it difficult to assign the observed response to either cytokine. To potentially deconvolute these signatures, and diagnose other putative cytokine response signatures, the CytoSig model¹⁹ was applied to MC12. In line with the GSEA, this analysis detected transcription response signatures of all IFN types, with the highest score being observed for IFN γ (**Supplementary figure 7d**). CytoSig additionally identified a reduction in TGF- β signaling in MC12, and retrospective analysis of the bulk RNAseq data revealed reduced expression of TGF- β responsive genes in QPCTL-deficient TMEs (**Supplementary figure 7e**). Furthermore, differential gene expression analysis showed that QPCTL deficiency was associated with a reduced TGF- β and an increased IFN responsive signature (**Fig. 4e-g**) across all MCs, potentially indicating either an altered abundance of—or sensitivity to—these cytokines. Notably, in vitro sensitivity of B16F10 cells to both TGF- β and IFN γ was unaltered by QPCTL deficiency (**Supplementary figure 7f-i**), arguing in favor of an altered abundance of these cytokines in the TME.

Next, the immune cell compartment was grouped into 11 MetaCells (MCs), classified as either CD3⁺ lymphocytes (*Cd8a*, *Cd3e*), dendritic cells (*Ccr7*, *H2-Aa*), or macrophages/monocytes (*Adgre1*, *Fcgr1*; **Fig. 5a-b, Supplementary Fig. 8a**). The macrophage/monocyte (M ϕ /Mo) MCs could be further subdivided into 2 groups that were marked by high expression of either *Ccr2* and *Irga4* (M ϕ /Mo-1 subgroup) or *Ms4a7* and *Pf4*^{fl} (M ϕ /Mo-2 subgroup; **Fig. 5b-c**), suggestive of a blood- versus tissue-derived origin²²⁻²⁵. Analysis of the contribution of cells from QPCTL-deficient and -proficient TMEs to individual MCs indicated that QPCTL deficiency changed the relative abundance of the

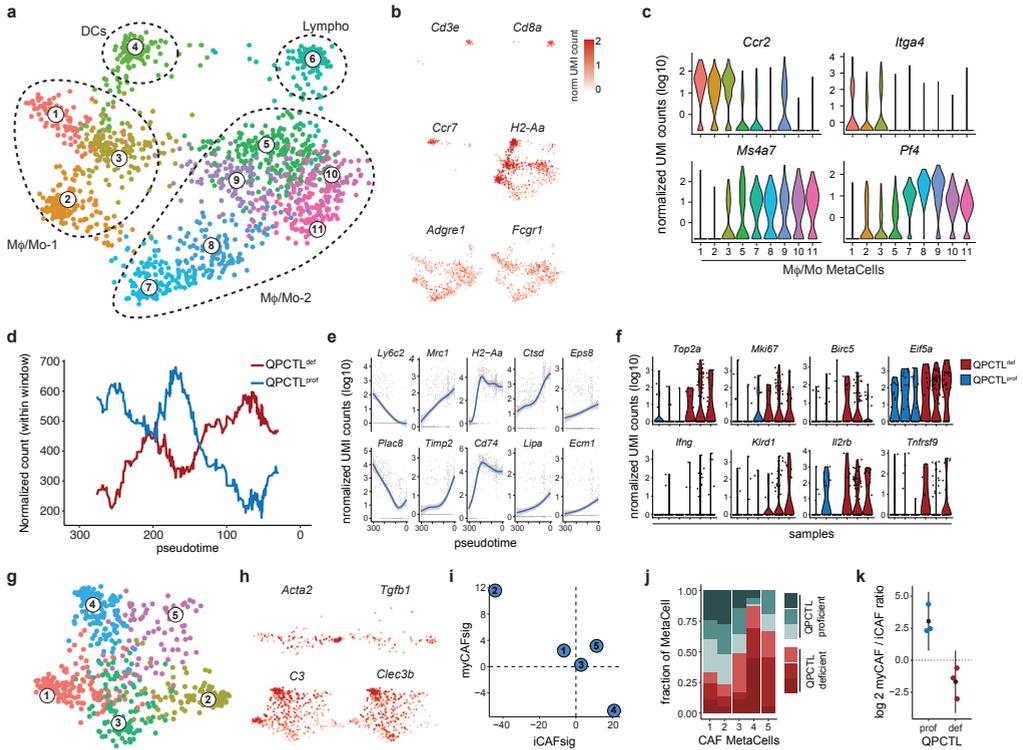


Figure 5. QPCTL deficiency alters the immune cell compartment and CAF polarization in the TME. scRNA sequencing was performed on sorted live cells from QPCTL-proficient ($n = 3$) and QPCTL-deficient ($n = 3$) B16F10 TMEs. Tumors were harvested at day 14 post inoculation. **(a, b)** 2-dimensional MetaCell projection of the immune cell compartment. Single cells are colored by MetaCell **(a)**, or normalized UMI count **(b)** of selected genes. **(c)** Violin plots depicting normalized UMI counts of selected genes across M ϕ /Mo MCs. **(d, e)** SlingShot Trajectory analysis performed on M ϕ /Mo subset 1 (MC1, 2 and 3). **(d)** QPCTL-deficient or QPCTL-proficient TMEs replicates were pooled, and normalized cell counts were tallied within windows of 60 cell wide, sliding 1 cell per frame. Lines indicate normalized cell counts within each window. **(e)** Normalized UMI counts of selected genes that are significantly associated with pseudotime. Blue lines indicate general additive linear models, grey areas indicate confidence intervals, grey dots represent single cells. **(f)** Violin plots depicting normalized UMI counts of selected genes within the CD3⁺ lymphoid cell MetaCell (MC6). **(g, h)** 2-dimensional MetaCell projection of the fibroblast compartment. Single cells are colored by MetaCell **(g)**, or normalized UMI count **(h)** of selected genes. **(i)** Enrichment of iCAF and myCAF signatures (**Supplementary Table 2**) in each CAF MetaCell. Signature values represent summed log₂ transformed enrichment values, calculated using the MetaCell algorithm. **(j)** Stacked barchart depicting sample composition of each CAF MetaCell. Cell counts from each sample were normalized to 1,000 cells. **(k)** myCAF/iCAF ratio detected in QPCTL-proficient and -deficient TMEs. Colored dots indicate individual mice, black dots indicate means, whiskers indicate the standard deviation. Depicted data were obtained in a single experiment, consisting of 6 mice. iCAF, inflammatory cancer-associated fibroblast; myCAF, myofibroblastic cancer-associated fibroblast; UMI, unique molecular identifier.

different cell states that jointly comprised the M ϕ /Mo-1 subgroup (**Supplementary Fig. 8b**). As intra-tumoral M ϕ /Mo cells can exist within a continuum of transcriptional cell states^{26,27}, pseudotime analysis was performed on M ϕ /Mo subgroup 1 to examine if such a continuum could be observed, and whether this was linked to QPCTL deficiency. This analysis demonstrated a strong continuous association between pseudotime (i.e. cell state) and sample-origin (**Fig. 5d**). To investigate the transcriptional changes underlying this association, genes were clustered based on their expression kinetics across pseudotime (**Supplementary Fig. 8c**), revealing a gradual loss of expression of Mo-related genes *Ly6c2* and *Plac8* across pseudotime (**Fig. 5e**), with the lowest expression levels

found in the area that contained the highest fraction of cells from QPCTL TMEs. At the same time, an increase in transcripts linked to ‘inflammation-resolatory’ Mf’s (*Mrc1*, *Timp2*), antigen-presentation (*H2-Aa*, *Cd74*) and Mf effector function (*Eps8*, *Ctsd*, *Ecm1*, *Lipa*)^{28–31} was observed. Consistent with these findings, Mf/Mo cells that were more dominant in QPCTL-deficient samples had reduced expression of monocyte-associated transcripts, and displayed transcriptional similarity to previously identified tumor-associated macrophage subsets³² (**Supplementary Fig. 8d**). Together with the observed Mf-Mo skewing (**Figure 2**), these data argue in favor of a model in which QPCTL deficiency in the TME leads to transcriptional changes that drive Mo-to-Mf conversion. Moreover, the intra-tumoral macrophages identified by flow cytometry (**Figure 2**) exhibited high-level surface expression of MHC class II (*H2-Aa*), CD206 (*Mrc1*) and CCR2 (*Ccr2*), linking these cells to the Mf/Mo population identified through scRNAseq (**Supplementary Fig. 8e**).

Congruent with the flow cytometric analysis, no differences in the frequencies of CD3⁺ lymphoid cells derived from QPCTL-deficient versus QPCTL-proficient TMEs could be observed in the scRNAseq dataset. However, as activated lymphoid cells are potent producers of IFN, we queried whether transcriptional features associated with lymphocyte activation were detected more frequently in lymphocytes from QPCTL-deficient TMEs. Interestingly, expression of genes associated with TCR-triggering (*Ifng*, *Il2rb* and *Tnfrsf9*) and cell cycle activity (*Top2a*, *Mki67*, *Birc5*) was detected more frequently in cells derived from the QPCTL-deficient tumors (**Fig. 5f**). Likewise, lymphocytes derived from QPCTL-deficient TMEs showed increased expression of the *Ccl3*, *Ccl4*, and *Ccl5* chemokines, and of the T cell activation-related genes *Ly6a*, *Nkg7*, and *Gzmb* (**Supplementary Fig. 8f**). While the increase in *Ifng* gene expression in lymphocytes in QPCTL-deficient samples was only modest, the parallel observation of other aspects of lymphocyte activation in these samples suggests that these cells may, at least in part, be responsible for the IFN responsive signature that is observed in the tumor cell compartment.

Diverse subsets of cancer-associated fibroblasts (CAFs) that possess distinct immunomodulatory functions have been reported in the TME of different cancer types³³, and two highly distinct populations— termed TGF- β -producing myfibroblastic CAFs (myCAF) and IL-1-driven inflammatory CAFs (iCAF)—have been identified in a recent set of cross-species studies^{34,35}. As a reduced TGF- β response signature was identified as one of the characteristics of tumor cells in QPCTL-deficient tumors, we next asked whether QPCTL deficiency affected CAF polarization. MetaCell-based clustering within the fibroblast cell supertype resulted in 5 transcriptionally distinct MCs (**Fig. 5g**, **Supplementary Fig. 8g**). Assessment of transcripts known to be involved in CAF function showed that MC2 was enriched for *Tgfb1* transcripts and several myCAF markers (e.g. *Acta2*, *Itgb1*), whereas MC4 exhibited more pronounced expression of genes involved in functional inhibition of TGF- β (*Ltbp1*, *Dcn*) and multiple iCAF markers (*C3*, *Clec3b*; **Fig. 5g**, **Supplementary Fig. 8h**). In line with this, analysis of signature enrichment-scores pertaining to these two subsets showed that MC2 and MC4 scored the highest for either the myCAF or iCAF signature, respectively (**Fig. 5h**). Moreover, increased surface expression of the myCAF and iCAF markers CD29/ITGB1 and Ly6C was detected on MC2 and MC4, respectively (**Supplementary Fig. 8i**). Strikingly, MC2 and MC4 displayed the highest depletion and enrichment in QPCTL-deficient and -proficient

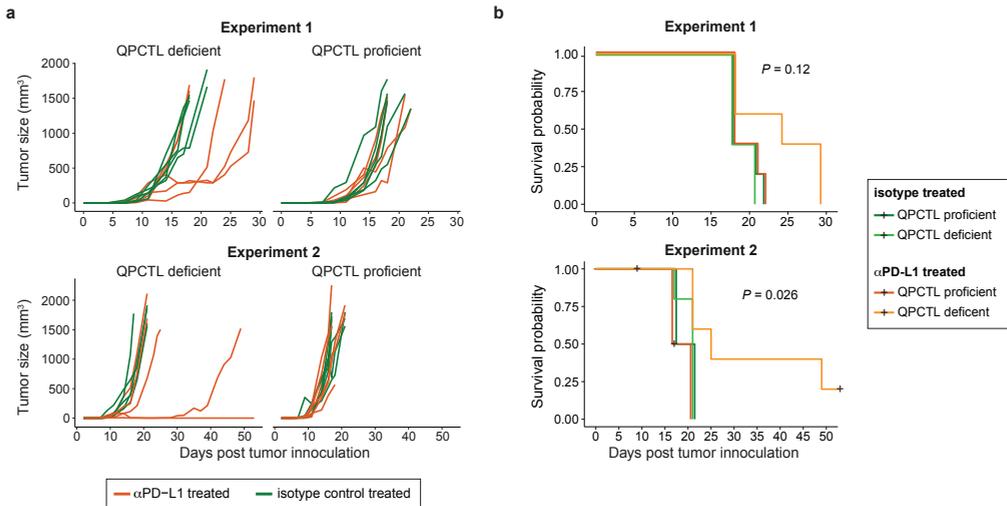


Figure 6. QPCTL deficiency sensitizes the tumor microenvironment to anti-PD-L1 treatment. QPCTL^{+/+} and QPCTL^{-/-} mice were inoculated with QPCTL-WT and QPCTL-KO B16F10 melanoma cells, respectively. Each group subsequently received either αPD-L1 or isotype control antibody treatment at day 7, 9 and 11 post tumor inoculation. (a) Tumor growth curves, assessed until day 50 post tumor inoculation. Lines represent individual mice. Data from two experiments are depicted ($n = 5$ per group). (b) Survival probabilities of mice treated with αPD-L1 or isotype control antibody in a QPCTL-proficient and -deficient setting. Black plus-signs indicate censored events. Data from two independent experiments are depicted ($n = 5$ per group). Global P values were determined by log-rank test (b). Data from 2 independent experiments are depicted.

TMEs (Fig. 5i), resulting in a 20-fold increase in the iCAF/myCAF ratio in QPCTL-deficient TMEs (Fig. 5j). Thus, in the absence of QPCTL activity, polarization of fibroblasts toward TGF- β -producing myCAFs is reduced in favor of the more pro-inflammatory iCAFs.

QPCTL deficiency enhances susceptibility of B16F10 tumors to anti-PD-L1 treatment

The B16F10 melanoma commonly shows a poor response to single agent PD1/PD-L1 checkpoint blockade^{36,37}. Having observed that QPCTL deficiency alters the TME to a more proinflammatory state, we hypothesized that QPCTL deficiency may modulate the sensitivity of B16F10 tumors to such PD-1/PD-L1 blocking therapies. To test this, QPCTL^{+/+} and QPCTL^{-/-} mice were inoculated with QPCTL-WT and QPCTL-KO B16F10 cells, respectively. Upon tumor formation (6 days post-inoculation) mice were treated with either an anti-PD-L1 or isotype control antibody. In QPCTL-proficient animals, tumor growth progressed rapidly and was not influenced by anti-PD-L1 therapy (Fig. 6a). In contrast, anti-PD-L1 therapy did result in improved tumor control in QPCTL-deficient mice (Fig. 6a), and led to a prolonged survival in approximately 50% of animals (Fig. 6b), providing direct evidence that the TME modulatory effect of QPCTL removal has functional consequences.

Discussion

QPCTL activity is known to influence the properties of a number of molecules that are active in the TME and may potentially influence additional—as of yet unidentified—substrates. To obtain a global view of the cumulative effects of QPCTL activity on the host's immune response to tumor growth, we made use of a QPCTL^{-/-} mouse model in combination with syngeneic QPCTL-KO tumor cell lines. We conclude that inactivation of QPCTL alters M ϕ -Mo abundance, increases IFN pathway activity relative to TGF- β pathway activity, and leads to a profound increase in iCAFs relative to myCAF in the TME (**Supplementary Fig. 9a**). In line with the observed skewing of the TME to a pro-inflammatory state that is induced by QPCTL deficiency, we demonstrate that such deficiency leads to the sensitization of B16F10 melanomas to anti-PD-L1 treatment.

The current study has the following limitations: 1) Germline deletion of QPCTL may potentially lead to developmental alterations that influence the host's response to tumor challenge, e.g. affecting the capacity of certain CAF or immune subsets to differentiate, independent of QPCTL activity during tumor outgrowth. However, the absence of clear phenotypic alterations at baseline, and the fact that an increased M ϕ -Mo-ratio was observed in wild-type mice challenged with QPCTL-deficient tumor cells, argue against this possibility. 2) In the present study we have aimed to model the effects of depletion of QPCTL activity on the tumor micro-environment, whereas glutaminyl cyclase inhibitors will, based on the similarity of their active sites, likely inhibit both QPCTL and QPCT activity^{38,39}. Such inhibition of QPCT may be relevant as siRNA-mediated suppression of QPCT has been shown to reduce expression of CCL2, CX₃CL1 and CD54/ICAM¹². In future work, dual inactivation of QPCTL and QPCT may form a means to test this.

The depletion of monocytes that we observe in QPCTL-deficient TMEs may potentially be explained by a decreased functionality of the CCL2-CCR2 signaling axis. In pre-clinical models of breast cancer, the CCL2-CCR2 axis has been shown to influence the abundance of monocytes in primary tumors⁴⁰ and metastatic lesions⁴¹. Furthermore, monocyte recruitment was found to be reduced after thioglycolate challenge of mice that were either QPCTL-deficient or treated with QPCT/QPCTL inhibitors¹⁰. However, it is important to note that at high concentrations, pyroGlu-CCL2 and unmodified CCL2 demonstrate similar chemotactic activity¹⁰, and the effect of impaired pyro-glutamylated CCL2 will therefore depend on local concentrations.

Contrary to expectations, we observed a relative increase in macrophage frequencies in QPCTL-deficient TMEs. Transcriptomic profiling of intra-tumoral M ϕ /Mo cells revealed that these macrophages expressed monocyte-associated molecules (e.g. *Ccr2*, *Itga4*) and pseudotime analysis suggests the existence of intermediate M ϕ /Mo cell states, together arguing in favor of their monocytic origin. In prior work, abrogation of the CCR2-CCL2 signaling axis in monocytes has been shown to strongly reduce the accumulation of intra-tumoral macrophages⁴². Based on these data, we propose that the boosting of intra-tumoral macrophages by QPCTL inactivation occurs through a mechanism that is independent of CCL2, and is potentially driven by an accelerated monocyte-to-macrophage differentiation program.

Having observed a number of independent alterations in the TME that are induced by QPCTL deficiency, it is of interest to speculate on the possible causal relationship between these individual changes. One possible scenario (**Supplementary Fig. 9b**) is that suppression of the CCR2-CCL2 axis due to the lack of QPCTL activity causes an early reduction in the influx of monocytes, which have been shown to form a major source of TGF- β ⁴³. Such an initial deficit in TGF- β abundance could potentially limit myCAF polarization and favor differentiation toward iCAFs, relieving TGF- β -driven suppressive effects on myeloid and lymphoid effector cells. Notably, TGF- β can act as a suppressor of IFN γ production by NK cells^{44,45} and CD8⁺ T cells^{46,47}, and increased abundance of activation-associated transcripts was observed in lymphocytes from QPCTL-deficient TMEs, together suggesting that this cell pool may be the source of the observed IFN-response signature. Genetic ablation experiments (**Figure 2**) indicated that CD47 does not play a significant mechanistic role in the TME-modulatory effects of QPCTL in the B16 melanoma model. However, as the CD47/SIRP α axis acts primarily through inhibition of activating signals of ITAM-containing receptors, such as activating Fc receptors on myeloid cells, it is plausible that the observed synergy between QPCTL deficiency and treatment with an opsonizing anti-PD-L1 antibody is at least partially mediated through its effect on the CD47 pathway.

In summary, our data provide evidence that removal of QPCTL activity can shift the TME from an immunosuppressive (monocyte skewed, myCAF, TGF- β) towards a pro-inflammatory (macrophage skewed, iCAF, IFN) milieu, and acts synergistically with anti-PD-L1 therapy to enhance tumor control and survival. If this TME-remodeling effect can also be achieved through pharmacological inhibition of QPCTL activity in human cancers, such inhibitors may offer potential in combination treatment strategies that include checkpoint blocking antibodies and/or tumor-opsonizing antibodies.

Methods

Generation of transgenic mice

C57/Bl6JR mice were obtained from Janvier. QPCTL^{-/-} mice carrying an 811bp deletion in exon 2 of the *Qpctl* gene were generated on the C57BL/6JRj background using pronuclear microinjection in mice zygotes with a CRISPR/Cas9 mixture (50 ng/ml Cas9 RNA and 25 ng/ml sgRNA, in water). The sgRNA was targeted to the second exon of the *Qpctl* gene (5'-GCACAATCAATAAGGGACGC-3'). QPCTL^{-/-} mice and QPCTL^{+/+} mice were identified by PCR using the following primers: Fwd_KO (5'-GTTTTAGGGATGGATGCCGC-3'), located before the 811bp deletion, Fwd_WT (5'-GGACTCCTAGTAGGCAACGG-3'), located in the 811bp deletion, and Rev (5'-GGCTGTTTTGGGATCTTCGG-3'), located after the 811bp deletion.

Evaluation of mouse blood cell counts

Whole blood of mice was collected by heart puncture and total cell counts were determined using a

DxH500 Hematology Analyzer (Beckman Coulter).

Peripheral blood collection and preparation

Whole blood of mice was collected into heparin-coated tubes by heart puncture or tail vein puncture at indicated time points. Samples were incubated twice for 5 minutes in erythrocyte-lysis buffer (0.15M NH_4Cl , 10mM KHCO_3 , 0.1 mM EDTA, pH 7.4), and washed once in staining buffer (0.5% BSA in PBS). Cells were then used for antibody staining, as described below.

Cell lines

B16F10 cells and MC38-AMS cells were kindly provided by D. Peeper (Division of Molecular Oncology & Immunology, Oncode Institute, The Netherlands Cancer Institute, Amsterdam, The Netherlands). The MC38-AMS cell line is a variant of the MC38 cell line available from Kerfast. Whole exome sequencing was performed to compare the MC38-AMS and MC38-Kerfast line, and data have been uploaded to the Sequence Read Archive. B16F10 and MC38-AMS cells were cultured in DMEM (Gibco) supplemented with 10% FCS and penicillin-streptomycin. Cells were cultured at 37 °C and 5% CO_2 .

CRISPR/Cas9-mediated generation of CD47 and QPCTL knockout cells

To generate QPCTL- knockout (KO), CD47-KO, and WT control B16F10 cell lines, cells were transfected with pLentiCRISPR v.2 vector encoding sgRNA targeting the murine QPCTL (5'-TATTGATTGTGCGACCCCG-3') or CD47 (5'-AGCAACAGCGCCGCCCAA-3') gene, or left untransfected. Culture medium of transfected cells was supplemented the next day with 2 $\mu\text{g ml}^{-1}$ puromycin for at least 2 days. Selected cells were expanded, and subsequently sorted on the basis of $\alpha\text{mCD47-MIAP301}^{\text{LO}}$ mSIRP α -Fc $^{\text{LO}}$ phenotype (in case of CD47 knockout), or $\alpha\text{mCD47-MIAP301}^{\text{HI}}$ mSIRP α -Fc $^{\text{LO}}$ phenotype (in case of QPCTL knockout), in order to obtain bulk knockout populations. WT control B16F10 cells were sorted based on morphology gating only. Next, single cells were isolated and expanded, and approximately 50 knockout or wild-type clones were pooled to obtain pure knockout or wild-type populations. To generate CD47/QPCTL double KO (dKO) cell lines, B16F10 QPCTL KO cell lines were transfected with pLentiCRISPR v.2 vector encoding sgRNA targeting the murine CD47 gene. 1 day after transfection, culture medium was supplemented with 2 $\mu\text{g ml}^{-1}$ puromycin for at least 2 days. Single cells were isolated and expanded, and 12 clones were pooled to obtain knockout populations. To generate QPCTL-KO and control MC38-AMS cell lines, cells were transduced with pLentiCRISPR v.2 vector encoding sgRNA targeting the murine QPCTL gene or a non-targeting control gRNA. 2 days after transduction, culture medium was supplemented with 2 $\mu\text{g ml}^{-1}$ puromycin for at least 4 days. Next, single cells were isolated and expanded, and 12 knock-out or control clones were pooled to obtain cell populations for further use. Gene disruption was validated by sequence analysis of the relevant gene locus by TIDE⁴⁸ analysis and, in case of CD47, by flow cytometry.

Tumor challenge

To analyze the effect of QPCTL deficiency in both host and tumor cells or in host cells only, 8- to 25-week-old male and female QPCTL^{-/-} or wild-type QPCTL^{+/+} littermate controls were injected with 2×10^5 of the indicated B16F10 tumor cell line in a 100 μ L solution of PBS (Lonza) and Matrigel (Corning) (1:1) in the right flank on day 0. To analyze the effect of tumor cell CD47 deficiency or CD47 and QPCTL-double deficiency, 9–12-week-old C57/Bl6JR (female; Janvier) were injected with 2×10^5 of the indicated B16F10 tumor cell line. To analyze the effect of QPCTL deficiency in both host and tumor cells in MC38-AMS tumors, 8–25-week-old QPCTL^{-/-} or wild-type QPCTL^{+/+} littermate controls were injected with 5×10^5 of the indicated MC38-AMS tumor cell line. Tumors were measured 3 times a week, and mice were sacrificed 13–17 days (B16F10 tumors) or 21–29 days (MC38 tumors) after tumor challenge. Mice with a tumor volume equal or below 40mm³ were excluded and tumors used for subsequent flow cytometry analyses ranged from 75–1436 mm³ (B16F10 tumors) or 112.5 - 786.5mm³ (MC38 tumors).

TME single-cell preparation

Tumors were fragmented on ice and were subsequently digested in DMEM (10 ml per tumor) supplemented with collagenase IV (2 mg ml⁻¹, Sigma Aldrich) and DNase I (50 μ g ml⁻¹, Sigma Aldrich) for 30 min at 37 C. Subsequently, 40 ml DMEM supplemented with 10% FCS was added per tumor, and cell suspensions were passed through 100 μ m filters. Next, samples were incubated for 5 minutes in erythrocyte-lysis buffer (0.15M NH₄Cl, 10mM KHCO₃, 0.1 mM EDTA, pH 7.4), and washed once in staining buffer (0.5% BSA in PBS). Tumor single cell suspensions were then counted and used for antibody staining.

Flow cytometry

Cell surface CD47 was assessed by staining of blood immune cells with the anti-mouse CD47 antibody MIAP301 at a dilution of 1:100 or 1:200 plus His-tagged recombinant mouse SIRP α (rmSIRP α -His) (4, 12 or 36 μ g ml⁻¹), in PBS containing 0.5% (w/v) BSA (Sigma) and 0.2% (w/v) sodium azide (Sigma) (FACS buffer) for 30 min at room temperature, protected from light. After two washes with FACS buffer, cells were stained with a fluorochrome-labeled anti-His antibody at a dilution of 1:100 or 1:200 for 30 min at 4 °C in FACS buffer, while protected from light. Cells were then washed with FACS buffer, and DAPI, propidium iodide, or 7-AAD Viability Staining Solution (eBioscience) was added to allow dead cell exclusion. Antibodies used to analyze immune cells in tumor single cell suspensions are listed in **Supplementary Table 3**. Measurements were performed on an LSRII, LSRFortessa, or FACSCantoII (BD Biosciences). Data were analyzed using FACS Diva software (BD Biosciences) and FlowJo software.

Unbiased flow cytometry data analysis

Samples were preprocessed using FlowJo software, compensating for spectral overlap, selecting IR-Dye⁻CD45⁺ single cells and removing outlier cells. Further analysis was performed in R, implementing the FlowCore package⁴⁹. Samples were subsampled to obtain 10,000 or 30,000 total

cells for spleen or tumor analysis, respectively. Next, a logicle (biexponential) transformation was applied to the measured fluorescence intensities. Uniform Manifold Approximation and Projection (UMAP) was used for dimension reduction, and subsequently used for hierarchical clustering by Euclidean distance (Ward's method). Relative contributions of cells derived from QPCTL^{-/-} and QPCTL^{+/+} samples to each of the clusters was then assessed.

IFN γ and TGF- β sensitivity of B16F10 cell lines

For IFN γ sensitivity testing, QPCTL-WT or QPCTL-KO cell lines were seeded on 6-well plates at 50,000 cells per well, incubated at 37 °C for 3 hours to allow cells to adhere, and subsequently treated with indicated amounts of IFN γ (Mouse IFN-gamma Recombinant Protein, ThermoFisher Scientific) for 21 hours. Cells were then harvested with trypsin-EDTA (Gibco), washed twice with PBS, and examined either through flow-cytometry or western blotting. For TGF- β sensitivity testing, QPCTL-WT or QPCTL-KO cell lines were seeded on 6-well plates at 50,000 cells per well and incubated at 37 °C for 16 hours. Next, cells were pre-incubated at 37 °C for 4 hours in culture medium containing 0.2% FCS, and subsequently incubated at 37 °C with indicated quantities of recombinant mouse TGF- β 1 protein (R&D systems) for 1 hour in culture medium containing 0.2% FCS. Cells were then harvested with trypsin-EDTA (Gibco), washed twice with PBS, and examined through western blotting. For flow-cytometry cells were stained with anti-PD-L1-BV421 and anti-H2-Kb-PE (both 1:100 dilution, see Supplementary Table 3) for 15 minutes at room temperature, washed twice with FACS buffer, and analyzed on an LSRFortessa (BD Biosciences). For western blot analyses, cells were incubated on ice for 30 minutes in 200 ml RIPA buffer (1% Triton X100, 0.1% Sodium deoxycholate, 0.1% SDS, 1 mM EDTA, 10 mM Tris pH 8, 140 mM NaCl) supplemented with Halt™ Protease and Phosphatase Inhibitor Cocktail (ThermoFisher Scientific), followed by pulse mixing on a Vortex Genie (Scientific Industries). Lysates were then centrifuged at 20,000x *g* for 20 minutes at 4 °C and protein concentrations in the resulting supernatants were determined using Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific) according to manufacturer's protocol. Next, equal amounts of protein were processed using the Novex NuPAGE Electrophoresis system (Thermo Fisher Scientific) and Trans-Blot Turbo Transfer system (Bio-Rad) according to the manufacturers' instructions. Membranes were blocked using Western Blocking Reagent (Roche) for 1 hours, and subsequently stained overnight at 4 °C with indicated primary antibodies diluted in Western Blocking Reagent (see Supplementary Table 3 for antibody information) followed by 2 hours at 4 °C with either anti-rabbit or anti-mouse secondary antibodies conjugated to HRP (see Supplementary Table 3 for antibody information). Western blots were imaged using the ChemiDoc MP Imaging System (Bio-Rad).

RNA sequencing

RNA was extracted from the indicated frozen tissues using the RNeasy Mini Kit (Qiagen). Cell populations isolated by FACS were washed once in PBS, and subsequently lysed in RLT buffer (Qiagen). Whole transcriptome sequencing samples were prepared with the TruSeq Stranded

mRNA Kit (Illumina). Paired-end 50bp sequencing was performed on a NovaSeq 6000 system (S1 flowcell, Illumina), obtaining an average of 18×10^6 reads per sample. Reads were aligned to the pre-built GRCm38 genome_snp_tran reference using HISAT2⁵⁰, and transcript counts were obtained using an in-house generated pipeline (GenSum, <https://github.com/NKI-GCF/gensum>). Differential gene expression analysis was performed using the edgeR package⁵¹. Network analysis was performed using the stringDB database, applying the igraph package for visualization.

Single cell RNA sequencing analyses

Single-cell digests of QPCTL^{-/-} and QPCTL^{+/+} TMEs were generated as outlined above. Cells were stained with IR-Dye for dead cell exclusion and with anti-mouse TotalSeqä Hashtag antibodies (TotalSeq-A0301-06, Biolegend), pooled in equal numbers, and were single-cell sorted on a BD Fusion cell sorter. Single-cell RNA isolation and library preparation was performed according to the manufacturer's protocol of the 10X Genomics Chromiumä Single Cell 3' kit, and the cDNA library was sequenced on the NextSeq™550 Sequencing System (Illumina). A total of $\pm 3.7 \times 10^8$ reads resulted in the detection of 14,888 cells with a median of 3,344 detected genes per cell. Feature-barcode matrices were generated using the Cell Ranger software of the 10X Genomics Chromium pipeline. Further processing was subsequently performed using the MetaCell¹⁸ and Seurat R packages⁵². Cells that contained less than 500 UMIs or had a mitochondrial transcript fraction of >0.2 were removed. Next, variable genes across the dataset were identified with a normalized variance/mean threshold at 0.1 and a down-sampled coverage threshold at 80, yielding 1,021 genes. These genes were subsequently used as anchors to search for gene-gene correlations across the dataset, and genes with correlations of >0.1 were included. The obtained genes were then clustered into 50 separate gene-modules, and each was annotated manually (**Supplementary Table 4**).

To identify the major cell types, a feature-gene list was compiled of gene-modules that contained marker genes for various cell types (modules 10, 17, 19, 21, 24, 26, 29, 30, 33, 39, 40, 44, 47, 48, 49; Supplementary Table 3), and these feature-genes were used to generate MetaCells. The obtained MetaCells were then classified as either immune cells, fibroblasts, or tumor cells, as shown in **Supplementary figure 6**. MetaCells that contained significant expression of marker genes from multiple cell types were identified as 'doublet MetaCells', and excluded from further analysis.

Subsequent analysis was performed within each individual cell type. In brief, doublet detection was performed using the HTODemux function of Seurat, setting the positive quantile at 0.99. Cells containing either a high amount of UMIs (UMI-thresholds: Immune [10,000], fibroblast [11,000], tumor cell [30,000]) or gene-counts (gene-count-thresholds: Immune [3,000], fibroblast [4,000], tumor cell [5,700]) were considered doublets and excluded. Feature genes used for cell type-specific MetaCell generation were obtained using the mcell_gset_filter_varmean and mcell_gset_filter_cov functions implemented in the Metacell package. These features genes were filtered for genes involved in cell cycle (gene-module 7 and 20) and ribosomal proteins (gene-module 2).

For all plots showing normalized UMI counts, a center log ratio normalization was applied, as implemented in the Seurat package⁵². To calculate sample fractions within MetaCells, cell counts

were first normalized to 10,000 cells within each sample-hashtag to allow comparisons.

Pseudotime analysis was performed using the Slingshot algorithm⁵³. Gene-level general additive models were fitted to feature-genes used for MetaCell generation applying the fitGAM function from the TradeSeq R package⁵⁴, setting knots at 5. Only genes that associated significantly (adjusted P value < 0.05) with pseudotime were used in subsequent analysis. Genes were then clustered based on expression kinetics across pseudotime based on Euclidean distance (**Supplementary Table 5**). To assess sample composition across pseudotime, the 3 replicates from QPCTL^{-/-} or QPCTL^{+/+} TMEs were analyzed together. To allow pooling of replicates, cell counts were normalized to the total number of cell counts within each sample. Normalized cell counts were then tallied within windows of 60 cell-codes wide, sliding 1 cell-code per frame. Differential gene expression analysis was performed using the FindMarkers function implemented in Seurat. Wilcoxon Rank Sum test was used to obtain log₂ fold changes.

For comparison of Immune MetaCell 1–3 to the external monocyte/macrophage cell clusters, the scRNAseq dataset from Gubin *et al.*³² was retrieved from the Gene Expression Omnibus (GSE119352). The external data was subsequently filtered for monocyte/macrophage cell clusters (Mac_s1–5), and normalized through centered log-ratio transformation. Differentially expressed genes within each cluster was then determined by comparing each cluster to all others applying Wilcoxon Rank Sum test (FindAllMarks function, Seurat R package). 25 marker genes were then selected for each MetaCell, defined as the 25 most enriched genes within that MetaCell. Each of these gene-sets was then used to compute a similarity score with each of the Mac_s clusters. Each similarity score was calculated by filtering the Wilcoxon Rank Sum test results of a given Mac_s cluster for a marker gene-set, followed by a weighted sampling of the log₂ transformed fold change values (sampling 10,000 times, with replacement, weighted by the MetaCell gene-enrichment value), and finally averaging (median) the obtained values.

Pathology

For histopathological analyses, 2 μm -thick hematoxylin-eosin-stained sections were prepared from formalin-fixed, paraffin-embedded murine tissues, including skin, spleen, thymus, lymph nodes, liver, pancreas, gastrointestinal tract, heart, lung, kidneys, testes, ovaries, accessory sex glands, bone marrow (sternum and extremity), and muscles. Sections were evaluated and scored by an animal pathologist blinded to animal genotype.

Statistical analysis

All statistical analyses were performed either with R (V4.0.5, ‘Shake and Throw’) or Graphpad (V8.4.1, Prism software). All statistical test were two-sided, unless otherwise indicated. Differences were considered statistically significant if $P < 0.05$. The n values used to calculate statistics, the type of replicates and the relevant significant P values are noted in the figure legends.

Ethical compliance

All animal experiments were approved by the Animal Welfare Committee of the Netherlands Cancer Institute (NKI), in accordance with national guidelines. Animals were maintained in the animal department of the NKI, housed in individually ventilated cage systems under specific-pathogen-free conditions, and received food and water freely. Mice were used at 8-25 weeks of age.

Data availability

Transcriptomic data presented in the manuscript have been deposited to the Gene Expression Omnibus, and can be accessed as series GSE180201. Exome data for MC38-Kerafast and MC38-AMS have been deposited to the Sequence Read Archive, and can be accessed as project PRJNA753254. R scripts used to produce key figures in the manuscript have been submitted to GitHub (https://github.com/kasbress/QPCTL_Project).

Acknowledgments

We thank members of the Schumacher lab, Ferenc Scheeren (Leiden University Medical Centre) and Leila Akkari (Netherlands Cancer Institute) for helpful discussions, and the Netherlands Cancer Institute – Antoni van Leeuwenhoek (NKI/AVL) Transgenic unit, Preclinical Intervention Unit and flow facility for technical support and input. We thank the NKI animal pathology facility for technical support, and Sjoerd Klarenbeek for histopathological assessment of QPCTL^{-/-} mice and discussions. We thank the NKI genomics core facility for processing, and providing input on, sequencing data presented in this work, and for analysis of the MC38-AMS exome data. This work was supported by ERC AdG SENSIT (grant agreement ID 742259) to T.N.S.

Author contributions

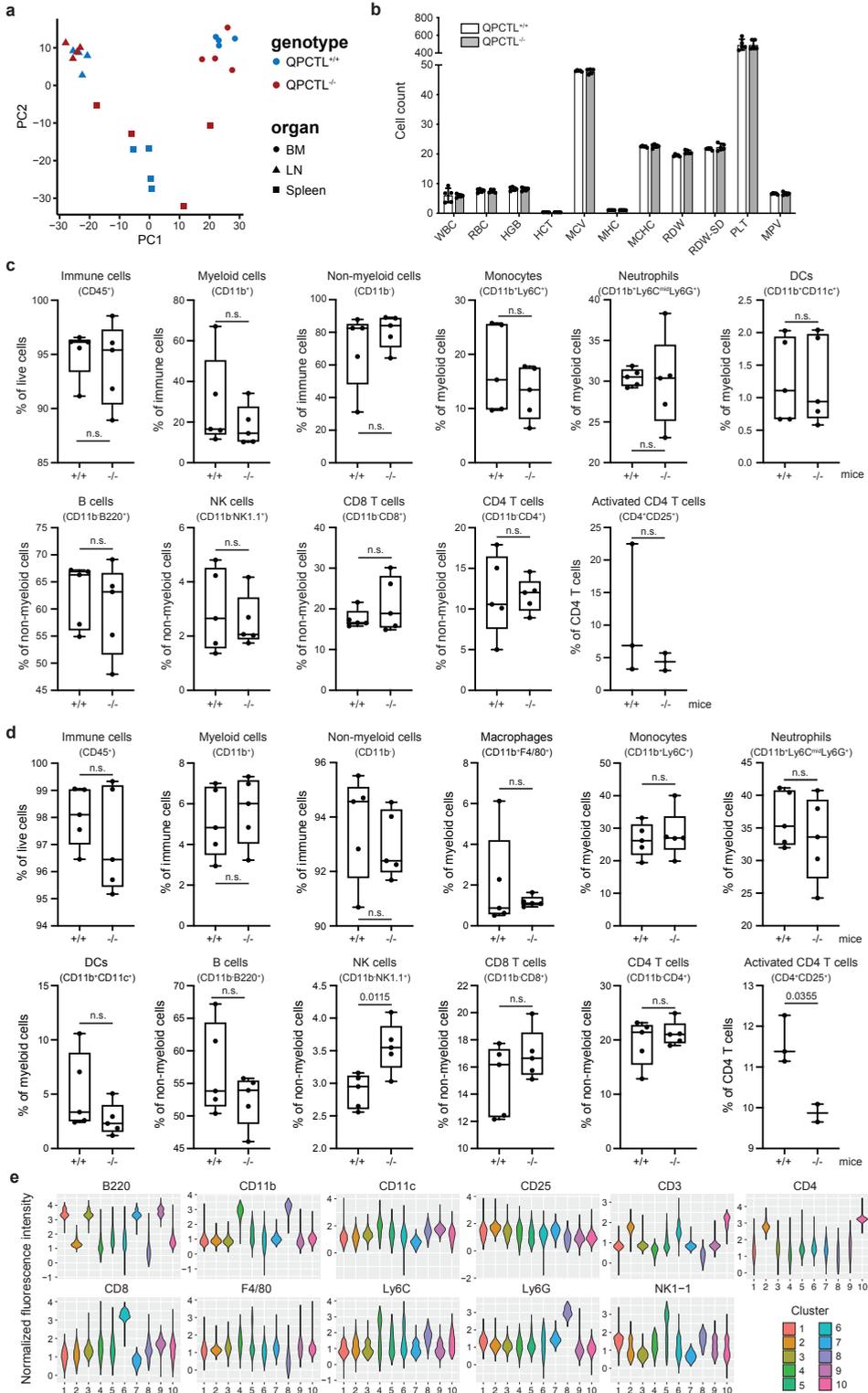
M.E.W.L. and K.B. conceived the project, designed and performed experiments, interpreted data, curated data and co-wrote the manuscript. M.T. designed and performed experiments. N.P, J.S, B.S and M.B. performed experiments. L.K established the QPCTL^{-/-} strain. T.N.S. conceived the project, designed experiments, interpreted data and co-wrote the manuscript.

References

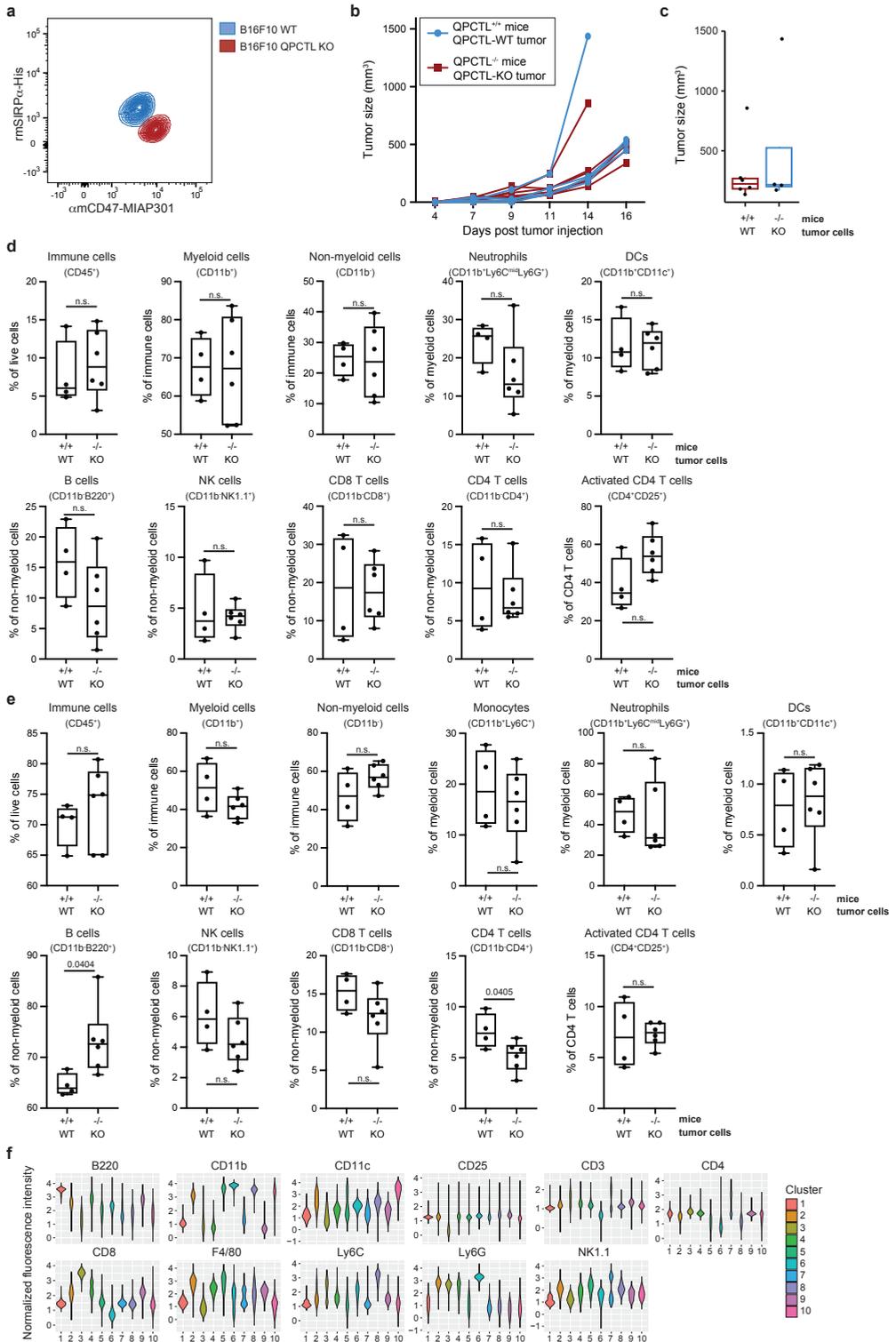
1. Rumpret, M. et al. Functional categories of immune inhibitory receptors. *Nat Rev Immunol* 20, 771–780 (2020).
2. Jaiswal, S. et al. CD47 Is Upregulated on Circulating Hematopoietic Stem Cells and Leukemia Cells to Avoid Phagocytosis. *Cell* 138, 271–285 (2009).
3. Majeti, R. et al. CD47 Is an Adverse Prognostic Factor and Therapeutic Antibody Target on Human Acute Myeloid Leukemia Stem Cells. *Cell* 138, 286–299 (2009).
4. Zhao, X. W. et al. CD47–signal regulatory protein- α (SIRP α) interactions form a barrier for antibody-mediated tumor cell destruction. *Proc National Acad Sci* 108, 18342–18347 (2011).
5. Cynis, H. et al. Isolation of an Isoenzyme of Human Glutaminyl Cyclase: Retention in the Golgi Complex Suggests Involvement in the Protein Maturation Machinery. *J Mol Biol* 379, 966–980 (2008).
6. Logtenberg, M. E. W. et al. Glutaminyl cyclase is an enzymatic modifier of the CD47- SIRP α axis and a target for cancer immunotherapy. *Nat Med* 25, 612–619 (2019).

7. Schilling, S. et al. Identification of Human Glutaminy Cyclase as a Metalloenzyme POTENT INHIBITION BY IMIDAZOLE DERIVATIVES AND HETEROCYCLIC CHELATORS*. *J Biol Chem* 278, 49773–49779 (2003).
8. Stephan, A. et al. Mammalian glutaminy cyclases and their isoenzymes have identical enzymatic characteristics. *FEBS J* 276, 6522–6536 (2009).
9. Hatherley, D. et al. Paired Receptor Specificity Explained by Structures of Signal Regulatory Proteins Alone and Complexed with CD47. *Mol Cell* 31, 266–277 (2008).
10. Cynis, H. et al. The isoenzyme of glutaminy cyclase is an important regulator of monocyte infiltration under inflammatory conditions. *Embo Mol Med* 3, 545–558 (2011).
11. Chen, Y.-L. et al. Inhibition of glutaminy cyclase attenuates cell migration modulated by monocyte chemoattractant proteins. *Biochem J* 442, 403–412 (2012).
12. Kehlen, A. et al. N-terminal pyroglutamate formation in CX3CL1 is essential for its full biologic activity. *Bioscience Rep* 37, (2017).
13. Mair, B. et al. High-throughput genome-wide phenotypic screening via immunomagnetic cell sorting. *Nat Biomed Eng* 3, 796–805 (2019).
14. Becker, A. et al. IsoQC (QPCTL) knock-out mice suggest differential substrate conversion by glutaminy cyclase isoenzymes. *Biol Chem* 397, 45–55 (2016).
15. Natarajan, V. T. et al. IFN- γ signaling maintains skin pigmentation homeostasis through regulation of melanosome maturation. *Proc Natl Acad Sci USA* 111, 2301–6 (2014).
16. Zhou, J., Ling, J., Wang, Y., Shang, J. & Ping, F. Cross-talk between interferon-gamma and interleukin-18 in melanogenesis. *J Photochem Photobiology B Biology* 163, 133–143 (2016).
17. Son, J., Kim, M., Jou, I., Park, K. C. & Kang, H. Y. IFN- γ inhibits basal and α -MSH-induced melanogenesis. *Pigm Cell Melanoma R* 27, 201–8 (2013).
18. Baran, Y. et al. MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol* 20, 206 (2019).
19. Jiang, P. et al. Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat Methods* 18, 1181–1191 (2021).
20. Liu, S.-Y., Sanchez, D. J., Aliyari, R., Lu, S. & Cheng, G. Systematic identification of type I and type II interferon-induced antiviral factors. *Proc National Acad Sci* 109, 4239–4244 (2012).
21. Gough, D. J. et al. Functional Crosstalk between Type I and II Interferon through the Regulated Expression of STAT1. *Plos Biol* 8, e1000361 (2010).
22. Zimmerman, K. A. et al. Single-Cell RNA Sequencing Identifies Candidate Renal Resident Macrophage Gene Expression Signatures across Species. *J Am Soc Nephrol Jasn* 30, 767–781 (2019).
23. Qian, J. et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res* 30, 745–762 (2020).
24. Mattioli, I. et al. The macrophage tetraspan MS4A4A enhances dectin-1-dependent NK cell-mediated resistance to metastasis. *Nat Immunol* 20, 1012–1022 (2019).
25. Bowman, R. L. et al. Macrophage Ontogeny Underlies Differences in Tumor-Specific Education in Brain Malignancies. *Cell Reports* 17, 2445–2459 (2016).
26. Arlauckas, S., Oh, N., Li, R., Weissleder, R. & Miller, M. A. Macrophage imaging and subset analysis using single-cell RNA sequencing. *Nanotheranostics* 5, 36–56 (2021).
27. Cheng, S. et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* 184, 792–809.e23 (2021).
28. Chen, Y.-J. et al. Eps8 Protein Facilitates Phagocytosis by Increasing TLR4-MyD88 Protein Interaction in Lipopolysaccharide-stimulated Macrophages. *J Biol Chem* 287, 18806–18819 (2012).
29. Fan, W. et al. ECM1 Prevents Activation of Transforming Growth Factor β , Hepatic Stellate Cells, and Fibrogenesis in Mice. *Gastroenterology* 157, 1352–1367.e13 (2019).
30. Huang, S. C.-C. et al. Cell-intrinsic lysosomal lipolysis is essential for alternative activation of macrophages. *Nat Immunol* 15, 846–855 (2014).
31. Yu, T. et al. Modulation of M2 macrophage polarization by the crosstalk between Stat6 and Trim24. *Nat Commun* 10, 4353 (2019).
32. Gubin, M. M. et al. High-Dimensional Analysis Delineates Myeloid and Lymphoid Compartment Remodeling during Successful Immune-Checkpoint Cancer Therapy. *Cell* 175, 1014–1030.e19 (2018).
33. Mhaidly, R. & Mechta-Grigoriou, F. Fibroblast heterogeneity in tumor micro-environment: Role in immunosuppression and new therapies. *Semin Immunol* 48, 101417 (2020).
34. Öhlund, D. et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med* 214, 579–596 (2017).
35. Elyada, E. et al. Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov* 9, 1102–1123 (2019).
36. Wculek, S. K. et al. Effective cancer immunotherapy by natural mouse conventional type-1 dendritic cells bearing dead tumor antigen. *J Immunother Cancer* 7, 100 (2019).

37. Sánchez-Paulete, A. R. et al. Cancer Immunotherapy with Immunomodulatory Anti-CD137 and Anti-PD-1 Monoclonal Antibodies Requires BATF3-Dependent Dendritic Cells. *Cancer Discov* 6, 71–79 (2015).
38. Jimenez-Sanchez, M. et al. siRNA screen identifies QPCT as a druggable target for Huntington's disease. *Nat Chem Biol* 11, 347–354 (2015).
39. Lues, I. et al. A phase 1 study to evaluate the safety and pharmacokinetics of PQ912, a glutaminyl cyclase inhibitor, in healthy subjects. *Alzheimer's Dementia Transl Res Clin Interventions* 1, 182–195 (2015).
40. Arwert, E. N. et al. A Unidirectional Transition from Migratory to Perivascular Macrophage Is Required for Tumor Cell Intravasation. *Cell Reports* 23, 1239–1248 (2018).
41. Qian, B.-Z. et al. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* 475, 222–225 (2011).
42. Tymoszyk, P. et al. In situ proliferation contributes to accumulation of tumor-associated macrophages in spontaneous mammary tumors. *Eur J Immunol* 44, 2247–2262 (2014).
43. Gupta, A. et al. Isoform specific anti-TGF β therapy enhances antitumor efficacy in mouse models of cancer. *Commun Biology* 4, 1296 (2021).
44. Laouar, Y., Sutterwala, F. S., Gorelik, L. & Flavell, R. A. Transforming growth factor- β controls T helper type 1 cell development through regulation of natural killer cell interferon- γ . *Nat Immunol* 6, 600–607 (2005).
45. Tauriello, D. V. F. et al. TGF β drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature* 554, 538–543 (2018).
46. Park, B. V. et al. TGF- β 1-Mediated Smad3 Enhances PD-1 Expression on Antigen-Specific T Cells in Cancer. *Cancer Discov* 6, 1366–1381 (2016).
47. Thomas, D. A. & Massagué, J. TGF- β directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. *Cancer Cell* 8, 369–380 (2005).
48. Brinkman, E. K., Chen, T., Amendola, M. & Steensel, B. van. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42, e168 (2014).
49. Hahne, F. et al. flowCore: a Bioconductor package for high throughput flow cytometry. *Bmc Bioinformatics* 10, 106 (2009).
50. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37, 907–915 (2019).
51. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinform Oxf Engl* 26, 139–40 (2009).
52. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29 (2021).
53. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *Bmc Genomics* 19, 477 (2018).
54. Berge, K. V. den et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* 11, 1201 (2020).

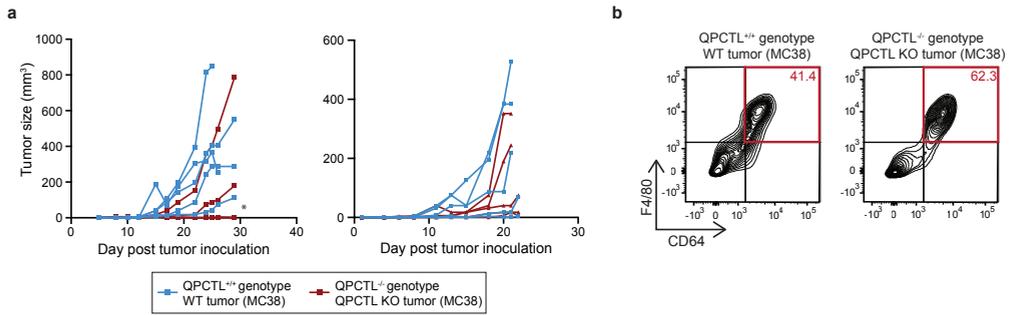


Supplementary Figure 1 (on previous page). Phenotypic analysis of QPCTL KO mice (related to Figure 1). (a) Principal component analysis performed on the 1,000 most differentially expressed genes in bone marrow (BM), lymph node (LN) and spleen samples from QPCTL^{+/+} and QPCTL^{-/-} mice. The first two components are plotted. (b) Quantification of indicated parameters in peripheral blood ($n = 5$ mice per group). Dots depict data from individual mice, bars represent group means, error bars indicate standard deviation. (c) Quantification of indicated immune cell populations as a frequency of total live, myeloid (CD11b⁺), non-myeloid (CD11b⁻) or total CD4 T cells (CD3⁺CD4⁺), in blood of QPCTL^{+/+} and QPCTL^{-/-} mice. Dots depict individual mice, boxplots indicate group median and 25th/75th percentiles, whiskers indicate min/max. $n = 3$ (QPCTL^{+/+}) or $n = 2$ (QPCTL^{-/-}) mice for activated CD4 T cells; $n = 5$ mice per group for all other immune cell subtypes. (d) Quantification of indicated immune cell populations as a frequency of total live, myeloid (CD11b⁺), non-myeloid (CD11b⁻), or total CD4 T cells (CD3⁺CD4⁺), in spleen of QPCTL^{+/+} and QPCTL^{-/-} mice. Dots depict individual mice, boxplots indicate group median and 25th/75th percentiles, whiskers indicate min/max. $n = 3$ (QPCTL^{+/+}) or $n = 2$ (QPCTL^{-/-}) mice for activated CD4 T cells; $n = 5$ mice per group for all other immune cell subtypes. (e) Violin plots depicting expression of indicated markers by the cell clusters described in Fig. 1d. *P* values were determined by two-sided Student's *T* test (c, d). Data were obtained in single experiments. WBC, white blood cell; RBC, red blood cell; HGB, hemoglobin; HCT, hematocrit; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; RDW, red cell distribution width; RDW-SD, red cell distribution width - size distribution; PLT, platelet; PV, mean platelet volume.

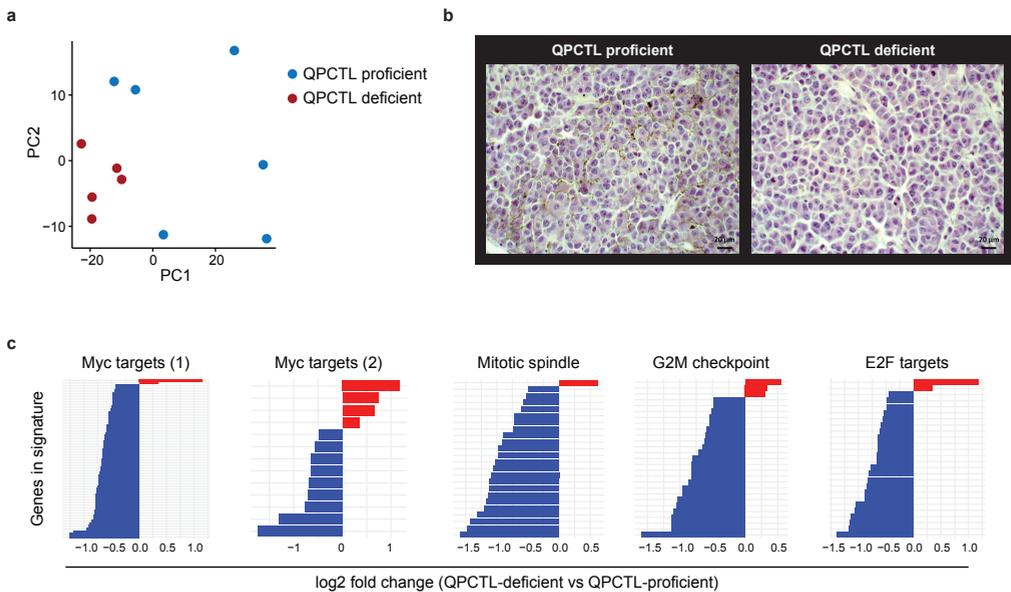


Supplementary Figure 2 (on previous page). Effect of QPCTL deficiency on the TME and blood cell compartment in tumor-bearing animals (related to Figure 2). (a) Representative flow cytometry plot depicting rmSIRP α -His and α mCD47 antibody (clone MIAP301) binding to QPCTL-WT and QPCTL-KO B16F10 cells. (b-f) QPCTL^{+/+} ($n = 4$) or QPCTL^{-/-} ($n = 6$) mice were inoculated with B16F10 WT and B16F10 QPCTL KO tumor cells, respectively. Mice were sacrificed between day 14–16 post inoculation. (b, c) Tumor growth curves and tumor sizes at day 14 post inoculation. Boxplots indicate group median and 25th/75th percentiles, whiskers indicate the interquartile range multiplied by 1.5, dots signify individual samples. (d) Quantification of indicated immune cell populations as a frequency of total live, total immune (CD45⁺), myeloid (CD11b⁺), non-myeloid (CD11b⁻) or total CD4 T cells (CD3⁺CD4⁺), in QPCTL-proficient or -deficient TMEs. Dots depict individual mice, boxplots indicate group median and 25th/75th percentiles, whiskers indicate min/max. (e) Quantification of indicated immune cell populations as a frequency of total live, total immune (CD45⁺), myeloid (CD11b⁺), non-myeloid (CD11b⁻) or total CD4 T cells (CD3⁺CD4⁺), in blood of tumor bearing mice in a QPCTL-proficient or -deficient setting. Dots depict individual mice, boxplots indicate group median and 25th/75th percentiles, whiskers indicate min/max. (f) Violin plots depicting marker expression within the obtained clusters described in **Fig. 2b**. P values were determined by two-sided Student's T test (d, e). Data are representative of at least 2 independent experiments.

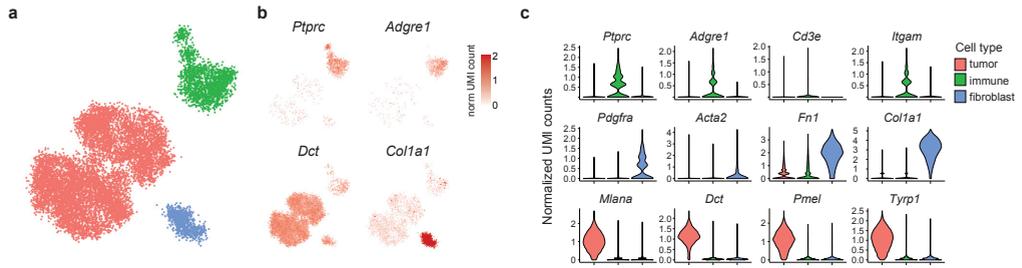
Supplementary Figure 3 (on previous page). Effect of QPCTL deficiency in tumor and host cell compartments (related to Figure 2). QPCTL^{+/+} and QPCTL^{-/-} mice were inoculated with either QPCTL-WT or QPCTL-KO B16F10 melanoma cells ($n = 7-8$ per group). Mice were sacrificed between 12–14 days post tumor inoculation. **(a)** Tumor growth curves. **(b)** Representative flow cytometry plots of data described in **Fig. 2d**, depicting macrophages (F4/80⁺CD64⁺ cells) amongst total myeloid (CD11b⁺) cells in the TME. Numbers depict the percentage macrophages within the myeloid cell gate. **(c)** Quantification of indicated immune cell populations as a frequency of total immune (CD45⁺), myeloid (CD11b⁺) or non-myeloid (CD11b⁻) cells, in the indicated TMEs. Dots depict individual mice, boxplots indicate group median and 25th/75th percentiles, whiskers indicate min/max. **(d)** Quantification of indicated immune cell populations as a frequency of total immune (CD45⁺), myeloid (CD11b⁺) or non-myeloid (CD11b⁻) cells, in blood before tumor inoculation (0 days), or 11 days after tumor inoculation. Dots depict individual mice, boxplots indicate group median and 25th/75th percentiles, whiskers indicate min/max. *P* values were determined by one-way ANOVA followed by Tukey's HSD test **(c)**. Significant *P* values (< 0.05) are indicated in the plots. Data are representative of at least 2 independent experiments **(a-c)**, or were obtained in a single experiment **(d)**.



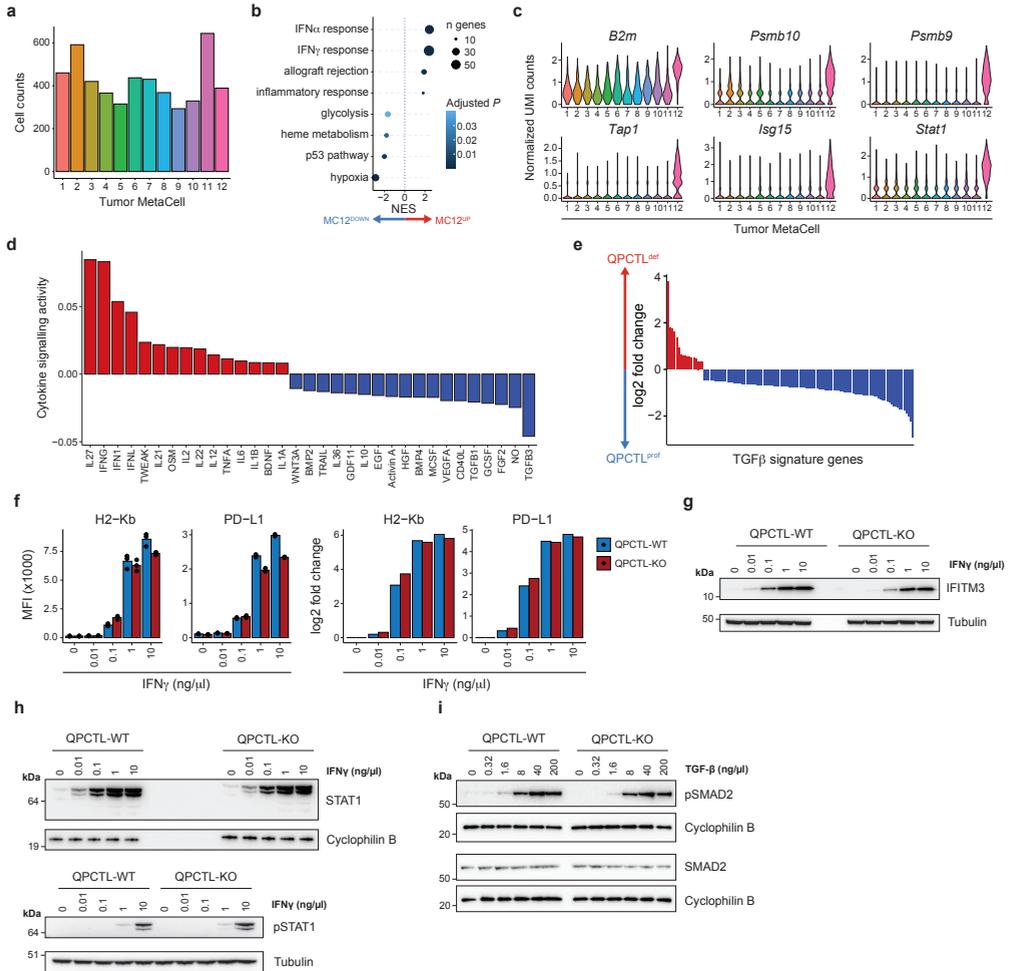
Supplementary Figure 4. Effect of QPCTL deficiency on TME composition (related to Figure 2). QPCTL^{+/+} and QPCTL^{-/-} mice were inoculated with QPCTL-WT and QPCTL-KO MC38 cells, respectively. Data from 2 independent experiments are shown ($n = 5$ per experiment). Mice were sacrificed at 22 (experiment 1) or 29 (experiment 2) days post tumor inoculation. (a) Tumor growth curves in two independent experiments. Asterisk (*) indicates 3 overlapping lines. (b) Representative flow cytometry plots of data described in **Fig. 2f**, depicting macrophages (F4/80⁺CD64⁺) amongst total myeloid (CD11b⁺) cells in the TME. Numbers depict the percentage macrophages within the myeloid cell gate. Data were obtained in 2 independent experiments.



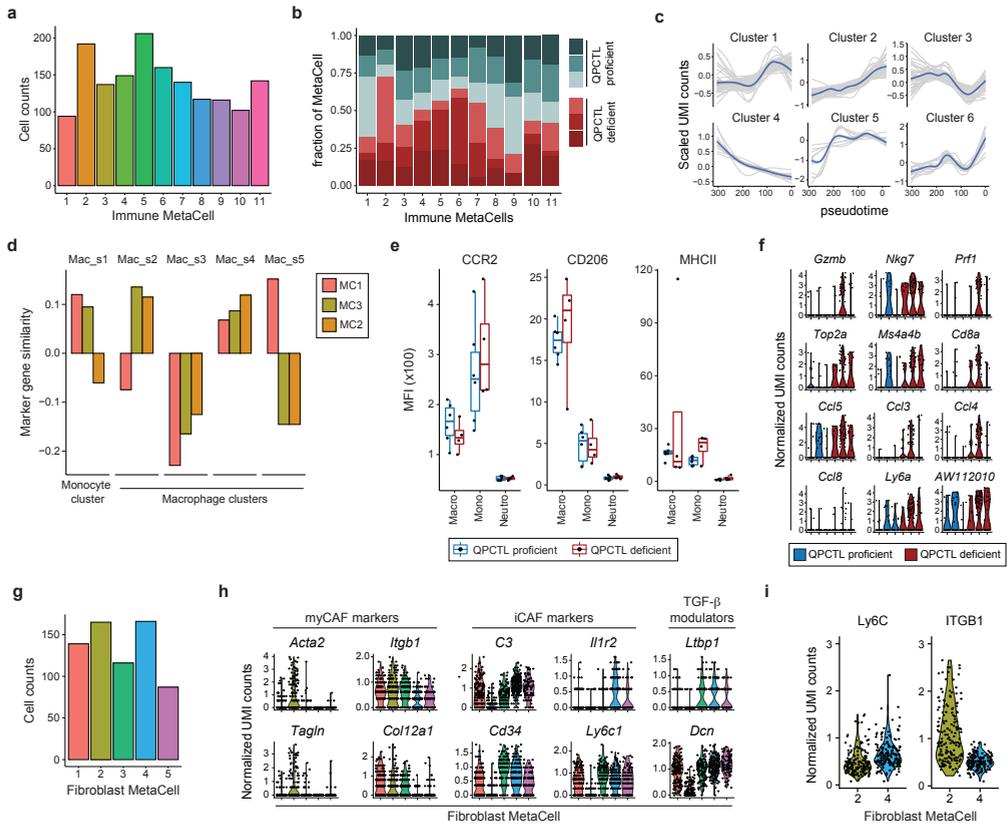
Supplementary Figure 5. Effect of QPCTL deficiency on tumor melanogenesis and cell cycle-related gene expression (related to Figure 3). mRNA sequencing was performed on sorted CD45-negative cells from QPCTL-proficient ($n = 5$) and QPCTL-deficient ($n = 6$) B16F10 TMEs. Tumors were harvested at day 14 post inoculation. (a) Principal component analysis performed on the 1,000 most differentially expressed genes across all samples. (b) Representative hematoxylin-eosin stained sections from QPCTL-proficient and -deficient TMEs. Note the presence of melanin signal (brown) in the QPCTL-proficient, but not the QPCTL-deficient sample. (c) Waterfall plots depicting log₂ fold change values of genes from indicated hallmark pathways obtained from MSigDB. Only significantly ($P < 0.05$) differentially expressed genes are included. Data are representative of 2 independent experiments. PC, principal component; MSigDB, Molecular Signatures Database.



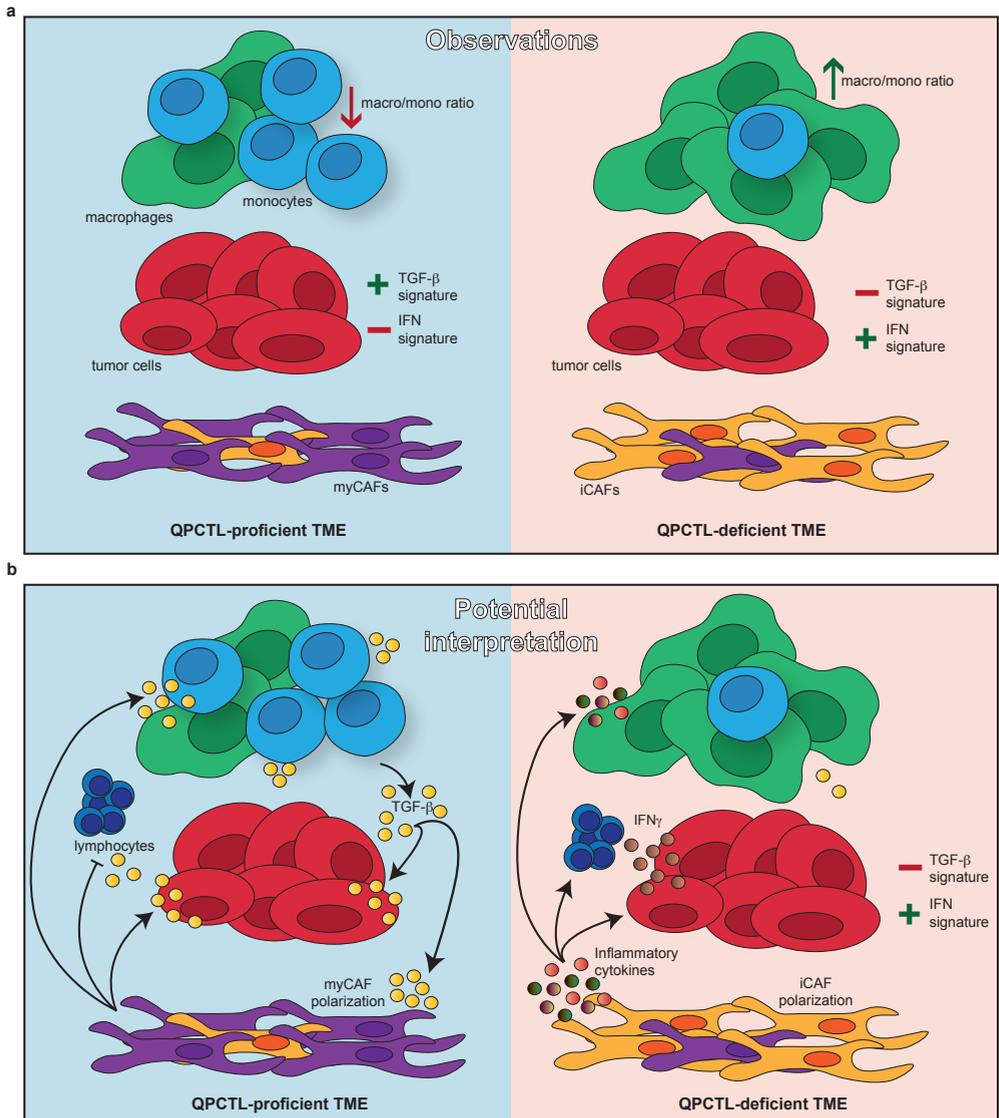
Supplementary Figure 6. Characteristics of the TME cell supertypes identified by single cell RNA sequencing (related to Figure 4 and 5). scRNA sequencing was performed on sorted live cells from QPCTL-proficient ($n = 3$) and QPCTL-deficient ($n = 3$) B16F10 TMEs. Tumors were harvested at day 14 post inoculation. **(a, b)** 2-dimensional MetaCell projection of all cells analyzed by scRNAseq. Single cells are colored by MetaCell **(a)**, or normalized UMI count **(b)** of selected genes. **(c)** Violin plots depicting gene expression of selected cell type-specific genes within the major clusters obtained. Depicted data were obtained in a single experiment, consisting of 6 mice.



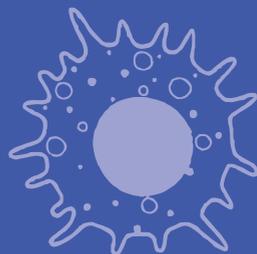
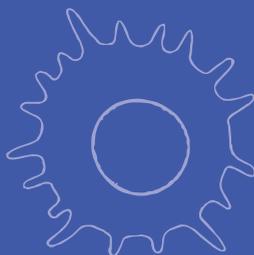
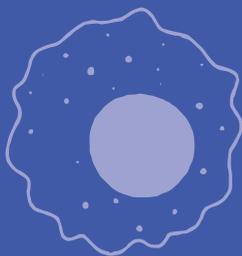
Supplementary Figure 7. QPCTL deficiency leads to an increased IFN- and decreased TGF- β -response signature in tumor cells (related to Figure 4). (a-d) scRNA sequencing was performed on sorted live cells from QPCTL-proficient ($n = 3$) and QPCTL-deficient ($n = 3$) B16F10 TMEs. Tumors were harvested at day 14 post inoculation. (a) Absolute cell counts per tumor cell MetaCell included in the analysis. (b) Gene set enrichment analysis performed on the top and bottom 200 genes expressed by MC12. Hallmark gene sets from MSigDB were used in the analysis, and only significant gene sets ($P < 0.05$) were included in the plot. (c) Normalized UMI counts of selected IFN responsive genes within each tumor MetaCell. (d) Cytokine signaling activity in tumor cell MC12 as predicted by the CytoSig algorithm (Jiang *et al.*, Nature Methods, 2021). (e) Analysis of bulk RNAseq data (Fig. 3). Log₂ fold changes of TGF- β associated transcripts between QPCTL-deficient and -proficient samples. (f) H2-Kb and PD-L1 cell surface expression of QPCTL-WT and QPCTL-KO B16F10 lines after a 20-hour incubation with the indicated amounts of IFN γ . For fold-change plots, group means were calculated and normalized to the MFI detected in the control treated samples. (g-h) Protein levels of IFITM3 (g), STAT1 and phosphorylated STAT1 (h) detected in cell lysates obtained from QPCTL-WT and QPCTL-KO B16F10 lines after a 20-hour incubation with indicated amounts of IFN γ . (i) Phosphorylation state of SMAD2 detected in cell lysates obtained from QPCTL-WT and QPCTL-KO B16F10 lines after a 1-hour incubation with indicated amounts of TGF- β . Data were obtained in a single experiment, consisting of 6 mice (a-e), or are representative of 2 independent experiments (f-i).



Supplementary Figure 8 QPCTL deficiency alters the immune cell compartment and CAF polarization in the TME (related to Figure 5). (a-d and f-i) scRNA sequencing was performed on sorted live cells from QPCTL-proficient ($n = 3$) and QPCTL-deficient ($n = 3$) B16F10 TMEs. Tumors were harvested at day 14 post inoculation. (a) Absolute cell counts per immune MetaCell included in the analysis. (b) Stacked barchart depicting the sample composition of each immune MetaCell. Cell counts from each sample were normalized to 1,000 cells. (c) Gene clusters obtained through hierarchical clustering of gene expression kinetics across pseudotime. Graphs depict general additive models fitted for each gene in grey. Blue lines represent average trends for each cluster. Note that gene cluster 2 and 4 exhibit a strong positive and negative association with pseudotime, respectively. Single gene examples from these clusters are depicted in **Figure 4e**. (d) Comparison of marker-gene expression between the MetaCells comprising M ϕ /Mo subset 1 (MC1, 2 and 3) and the M ϕ /Mo cell clusters described by Gubin and colleagues (Mac_s_1–5; Gubin *et al.*, Cell, 2018). The 20 most enriched genes of each MetaCell were selected, and similarity to the Mac_s clusters was calculated as the sum of log₂ transformed enrichment of those selected genes within each Mac_s cluster. See methods for details on analysis. (e) Comparison of cell surface expression levels of indicated proteins on macrophages, monocytes and neutrophils assessed in QPCTL-proficient ($n = 6$) or -deficient ($n = 4$) TMEs, analyzed at day 14 post tumor inoculation. (f) Top 12 genes with the highest relative expression in CD3⁺ lymphocytes (MC6) from QPCTL-deficient compared to QPCTL-proficient TMEs. Violin plots depicting normalized UMI counts. (g) Absolute cell counts per CAF MetaCell included in the analysis. (h) Violin plots depicting normalized UMI counts of selected genes across CAF MetaCells. (i) cell surface expression levels of indicated proteins on CAF MC2 and MC4, as assessed using barcoded antibodies. Data were obtained in a single experiment, consisting of 6 mice (a-d and f-i), or are representative of 2 independent experiments (e).



Supplementary Figure 9. Graphical abstract. (a) Cartoon depiction of the findings in the study. **(b)** Potential scenario that connects the reported findings.



Chapter 6. Discussion

OR: "Lets hear some answers to
questions that nobody asked."



While making my first steps into the academic world I encountered a host of tenacious dogmas, enlightening opinions, and close-to-ideal-but-not-so-perfect systems. Although this journey has caused me to consider, and re-consider, many different topics included in this thesis, I will use this this chapter to dive a bit deeper into two—perhaps somewhat unrelated—themes that have fascinated me during my work. First, I will consider the current theories on stemness in the CD8⁺ memory T cell population and offer my own view on this topic, a view that was strongly shaped by our work in chapter 4. Second, I will discuss the current structure for academic data-sharing that should allow for the re-use (or re-purposing) of data and findings. In chapter 3 through 5 I sought to use data from others to enrich our own observations, but found that this was not a simple feat, in part forming the basis for my interest in the matter. In the final section of this part of the discussion, I will give an example of some of the current problems with the accessibility of scientific data, and opine on a number of ways to improve the manner in which data is shared across the academic community.

Stemness in the CD8⁺ memory T cell pool

My colleagues and I ended the discussion of **chapter 4** with the notion that the organizing principles of the T_{CM} pool shares some similarities with the stem cell compartments found in solid tissues. Stemness in the CD8⁺ memory T cell pool has been widely discussed in the literature, with the concept of stemness taking many shapes and forms. Although the premise of true stem cell activity is, at first glance, somewhat counterintuitive for a highly differentiated cell type such as T cells, I do feel that this concept provides a helpful framework when considering lineage relationships in the T cell pool.

In the following sections I will outline some influential studies investigating stemness in the CD8⁺ T cell memory pool, describe where, in my view, these studies have placed our understanding of this concept, and additionally suggest a few topics that I believe require further consideration.

Stemness in the context of T cell immunity

Stemness is commonly defined as the capacity of a cell pool to allow both *self-renewal* (duplicating oneself in relative perpetuity) and *differentiation* (regenerating a functional tissue)¹. In many anatomical compartments, this property is restricted to a small subset of multipotent cells capable of differentiating into the specialized cells that make up the tissue²⁻⁴; a process that is accompanied with the progressive loss of stemness. This concept also holds true for the majority of immune cells, such as monocytes and neutrophils, which are continuously replenished by hematopoietic stem cells (HSCs) in the bone marrow. However, due to their adaptive nature, T and B cells cannot rely on this replenishment model. Specifically, the naïve T cell pool comprises an immense variety of T cell receptor (TCR) clonotypes, generated through random re-arrangement of gene-fragments. Upon pathogen encounter, relevant antigen-specific T cell clonotypes are selected to expand, differentiate and combat the ongoing infection, and subsequently establish a long-lived memory pool. This memory T cell pool retains the capacity to repeatedly differentiate and expand upon multiple cycles of infection, but must do so independent of de novo generation from the bone marrow, in order to retain the critical clonotype information. This highlights an interesting question in the developmental hierarchy of T cell memory; *How to allow for successive rounds of proliferation and differentiation without reinforcements from HSCs?*

An attractive hypothesis for the maintenance of T cell memory is that, analogous to other tissues, stemness or multipotency is restricted to a minute subset. Evidence in favor of this model came from a series of studies describing a small subset of memory T cells that existed in a multipotent naïve-like state, and this population has been coined ‘memory stem cells’ (or T_{SCM})⁵⁻⁷. This T_{SCM} population possessed superior proliferative potential and retained a high level of multipotency upon TCR stimulation. Furthermore, the T_{SCM} pool was phenotypically similar to the T_{CM} pool apart from the peculiar retention of CD45RA, a protein that has been extensively used by immunologists as a mark for naïve T cells. Transcriptional profiling put these antigen-experienced cells at the apex of the memory T cell pool hierarchy^{7,8}, placing them somewhere between naïve and central memory T cells.

Although the presence of T_{SCM} provides a parsimonious model for developmental hierarchy in the memory T cell pool, this model does not fully fit with adoptive cell transfer studies showing that the ability to give rise to secondary effector pools is abundantly present in the T_{CM} pool^{9,10}. This stem-like capacity of T_{CM} was elegantly demonstrated by Graef *et al.*¹¹, showing that single CD62L⁺ T_{CM} cells (that had not been specifically selected for the expression of T_{SCM} associated markers) were able to reconstitute a functional T cell pool throughout multiple successive rounds of single cell transfer. This study thereby established that stemness is a characteristic shared across many cells in the T_{CM} pool.

So how does this finding fit with the T_{SCM} model? A simple consideration provides some insight here: As demonstrated in **chapter 4**, the T_{CM} pool comprises cells that exhibit a variety of transcriptional states and distinct behaviors. Therefore, the transcriptome analyses of bulk T_{SCM} and T_{CM} that placed T_{SCM} as a more naïve-like subset relative to T_{CM} , may have been confounded by the latter's internal heterogeneity. In line with this possibility, a recent study by Galletti *et al.* showed that the depletion of PD1⁺TIGIT⁺ cells from the T_{SCM} and T_{CM} pools largely eliminated the transcriptional and functional differences between these two cell populations¹². Furthermore, findings from a scRNAseq study of tumor-infiltrating and blood-derived T cells suggest that there is a noteworthy degree of promiscuity in the expression of CD45 isoforms across T cell subsets¹³. These findings could imply that T_{SCM} should not be considered an entirely distinct population as previously imagined, but rather a constituent of the T_{CM} pool that is primarily set apart by its alternative splicing of CD45. Whether the differential expression of distinct splice-forms of CD45 is functionally relevant in memory maintenance or re-expansion potential will be an interesting topic for future endeavors.

In summary, the studies by Galletti and Graef would support a model in which the T_{CM} pool, as a whole, serves as a stem-like reservoir maintaining each T cell clone. While stemness appears to be a shared property of T_{CM} , our findings, in addition to those made by Galletti and colleagues, signify that a degree of specialization is present within this population. This begs the question: If all T_{CM} are equal, are some T_{CM} more equal than others?

Decreasing potential or division of labor

The phenotypic and transcriptional diversity within the T_{CM} population that we observe in **chapter 4** presents us with two intriguing avenues to pursue. First, it will be interesting to explore how this diversity is established. Our observation that prior cell division is correlated with an effector-like transcriptional program is consistent with several models. For instance, this could mean that at some point during the acute phase of the T cell response T_{CM} precursors diverge, with one lineage “deactivating” into a more naïve-like quiescent behavior, whereas the other lineage maintains a more activated state and continues to divide. Alternatively, effector-like T_{CM} cells could derive from a separate lineage of de-differentiated effector T cells^{14,15}, which could explain their transcriptional state and extensive degree of prior proliferation. A strategy to experimentally address this issue would be to leverage the *Klrg1*^{C^{re}} mice developed by Herndler-Brandstetter *et al.*¹⁴, to examine whether the effector-like T_{CM} pool is enriched for cells that have previously expressed the effector T

cell-associated protein KLRG1.

A second valuable direction will be to investigate the functional relevance of T_{CM} diversity upon re-infection. In our efforts we have found substantial differences in transcriptional profiles and re-expansion potential within the T_{CM} pool. This fits with an observation made in the Galletti study¹² that a more differentiated T_{CM} sub-population, that the authors term “pre-exhausted”, exhibits reduced replicative potential. A matter that is currently unresolved is whether this more differentiated state indicates that these cells are of little value or, rather, that the diversity in cell states in the T_{CM} pool is reflective of a division of labor. Some experiments from **chapter 4** and Galletti *et al.* may provide some insight here. Specifically, our effector-like T_{CM} appear to degranulate to a larger extent upon short term ex vivo stimulation. Likewise, Galletti’s “pre-exhausted” T_{CM} contain more accessible chromatin at cytotoxicity-related genomic loci. These observations could indicate a degree of specialization within the T_{CM} pool, in which some cells are more prone to re-expand and others are predisposed to rapidly re-exert effector functions.

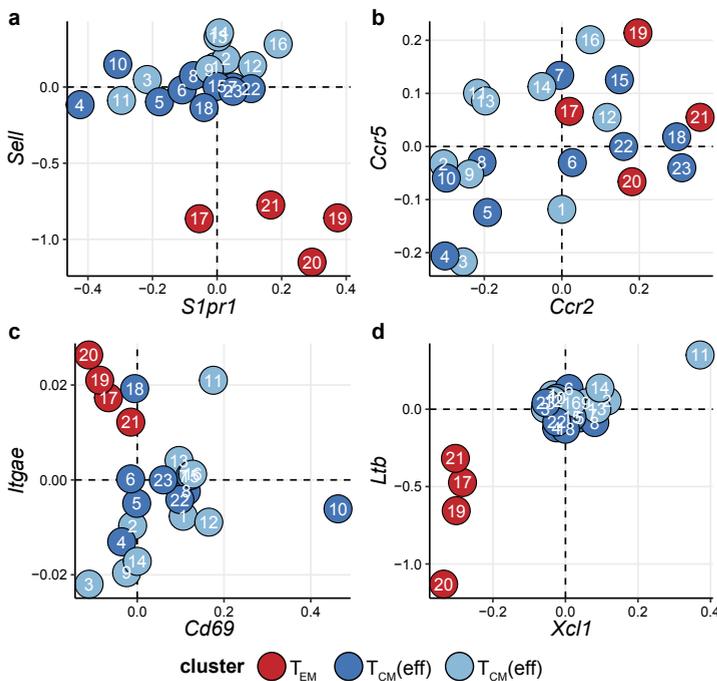


Figure 1. T_{CM} exhibit heterogeneous mRNA levels of various genes involved in cell trafficking and localization. Differential expression of selected genes across the T_{CM} and T_{EM} MetaCells identified in **chapter 4**. (a-d) Log2 transformed gene-expression enrichment is plotted for [a] *Sell* (CD62L) versus *S1pr1* (Sphingosine-1-phosphate receptor 1), [b] *Ccr5* versus *Ccr2*, [c] *Itgae* (CD103) versus *Cd69*, and [d] *Ltb* (Lymphotoxin beta) versus *Xcl1* (Lymphotoxin)

Nature versus nurture at the cellular level

The studies discussed above provide compelling evidence that the T_{CM} pool comprises cells that exist in distinct states, and that differ in their capacity to execute specific functions upon re-activation. But what underlies this disparity during a recall response? Is, for instance, the enhanced re-expansion potential of quiescent T_{CM} fully attributable to their cell state (i.e., their *nature*), or could the biased localization of different T_{CM} types in distinct niches (i.e., *nurture*) play a role?

The lymphoid tissues in which T_{CM} largely reside (such as lymph nodes and spleen) have a complex organization of myeloid, lymphoid, and stromal cells, compartmentalizing these organs into distinct niches. Several studies have highlighted the importance of memory T cell positioning within secondary lymphoid organs to recall responses, and demonstrated a key role for chemokine receptors in this process^{16–18}. However, such studies have generally not assessed whether T_{CM} with different cell states are differentially positioned. Re-examination of the scRNAseq dataset of splenic CD8⁺ memory T cells presented in **chapter 4** offers some clues on this matter. While all T_{CM} expressed high levels of *Sell* transcripts (encoding the lymphoid-tissue entry receptor CD62L), these cells displayed heterogeneous expression of the tissue-egress associated gene *S1pr1* and several chemokine receptors (**Fig. 1a-b**). Several T_{CM} MetaCells additionally differed in their expression of *Cd69* and *Itgae* (encoding CD103), genes classically associated with tissue-resident memory T cells (**Fig. 1c**). Interestingly, one of these MetaCells with elevated *Cd69* and *Itgae* expression was additionally marked by relatively high levels of *Ltb* and *Xcl1* transcripts (**Fig. 1d**). Both of these genes encode secreted factors for which the receptors are present on the myeloid and stromal component of lymphoid tissues^{19,20}. The differential expression of these cell migration and retention-associated genes could therefore imply that these T_{CM} subtypes possess a different affinity toward specific local niches.

As a final note, if such differential positioning of T_{CM} indeed underlies distinct functional outcomes, it would be highly interesting to investigate the stability of these niches. Specifically, are these niches seeded upon memory formation and subsequently remain immutable, or is there a certain degree of plasticity, allowing T_{CM} to move in and out of these niches? Also, can such niches exclusively be seeded during resolution of infection, or is simply the correct expression of specific chemotactic receptors/factors enough? This latter question may be particularly noteworthy, as its answer would strongly affect the manner in which studies using the re-transfer of T cell subsets (that are taken out of their original niche) should be interpreted. By the same token, the mechanism of niche formation could have implications for T cell based cellular therapies, as putative factors necessary for the establishment of niches that ensure long-lasting protective T cell responses may not be sufficiently present. In the event that improper niche formation negatively impacts T cell immunity, such cellular therapies may conceivably be modified to incorporate this component, for instance through the use of adjuvants (e.g., cytokines or chemokines) or forced expression of putative niche-inducing factors through genetic engineering of transferred T cells.

Scientific equity through data sharing

Scientific discovery is, at its core, a community effort, with each new insight being built on the foundation of data that was provided by predecessors. Therefore, I feel it is important that our precious data is viewed not only as a means to an end for our specific question, but first and foremost as a starting point for others. This would entail properly storing and sharing our published findings in a way that is easily accessible and interpretable by others. In **chapters 3, 4 and 5** I sought to validate specific findings using data from others, and found that the habit of proper data storage is still far from commonplace in the scientific community. Although I could usually find plenty of studies that contained experiments useful to my research questions, a large share of this data was either not published alongside the article at all or uploaded in a manner that did not allow for proper re-use.

By no means am I suggesting that all is Fire and Brimstone, in fact, data sharing is a field of lively discussion and steady (albeit slow) innovation. In the next few paragraphs, I will briefly outline the most prominent philosophy for data sharing, then provide an example of a data type for which we are still ‘playing catch-up’, and finally discuss my view on the role of three major stakeholders/components of the scientific community (publishers, repositories, and scientists) in the improvement of data availability.

A philosophy for open science

In practice, the academic field is purposefully unequal, resulting in disparities in the ability of different labs to generate certain types of datasets. Because the source of this inequality is difficult to address (e.g., there cannot and should not be infinite funding for all labs), data sharing provides a way toward more *equitable outcomes*, as everyone would be able to reap the benefits of data obtained by a few. To achieve such an outcome, a number of parties from various disciplines met in Leiden (2014) to discuss the principles of open science. During this workshop, the FAIR principles²¹ were drafted, which represent a general philosophy of data sharing that can be applied broadly in the scientific community. Essentially, for a dataset to be FAIR it needs to be *Findable*, *Accessible*, *Interoperable* and *Re-usable*, factors that are mostly determined by the richness of information on experimental conditions or outcomes of analyses that is shared alongside the data, referred to collectively as metadata.

Say I have found an intriguing gene-expression network in my pet cell type, and now would like to pressure-test these findings in an external RNAseq dataset. First, I should be able to *find* a relevant dataset, meaning the original research article should refer me to the RNAseq data through a permanent link or should be easily found through a query of the relevant repository. In the latter case, rich metadata detailing the experiment is key, as it simplifies discovery through a search engine. Second, I should be able to *access* and download the data freely without the need of going through a paywall or creating some site-specific account, to the extent that privacy regulations permit. Third, the dataset needs to be interpretable (both to machines and humans) so I can integrate it in any analysis—through its *interoperability*—meaning that the files use standardized formats and its

annotations use vocabularies that are widely applied (e.g., genes are identified with gene symbols or ensemble identifiers). Last, I should be able to easily use, or rather *re-use*, the obtained data for its new purpose. To achieve this, rich metadata is again crucial, providing machine-readable sample-level information on both experimental conditions and downstream analyses (e.g., outcomes of a clustering analysis or the code that was applied to generate the manuscript figures). If the data produced by a study is shared in a FAIR manner, a secondary user should be able to use and mine the dataset within a matter of hours, providing the user with a means to start asking questions the original authors have not considered.

The case of single cell RNA sequencing

Single cell sequencing is a good example where FAIR data sharing is highly beneficial. Due to their richness of information, a multitude of questions can be probed in each dataset, making these datasets useful reference points for research lines outside of their original intended purpose. In order to serve as such reference points, it is key that access is provided to both the raw data *and* the results of all down-stream analyses. This means that detailed metadata—*at the cell level*—is crucial, as it allows direct re-use and integration of a study's results for new endeavors.

Unfortunately, assessment of single cell sequencing datasets deposited to the Gene Expression Omnibus (GEO) over the last couple of years shows that the majority of datasets does not include identifiable metadata (**Fig. 1a**). While such data can still be re-analyzed from scratch, a direct comparison to the authors results is severely complicated. The observed inconsistency in the FAIRness of single cell sequencing datasets uploaded to GEO is not entirely surprising, as the information page '*Submitting high-throughput sequence data to GEO*' does not provide information or guidelines on submitting single cell sequencing data (as of this writing). Furthermore, a lot of variability can be found in the data availability guidelines among the different scientific publishers. Some publishers, such as Cell Press²², leave little ambiguity, whereas many other publishers mainly provide either vague or dated guidelines^{23,24}. Interestingly however, these differences do not appear to result in better or worse commitment to data re-usability, as the percentage of depositions that include metadata are comparable between the different publishers (**Fig. 1b, c**).

As a final note, improved FAIRness of scRNAseq data appears to be positively associated with the influence of a manuscript. Specifically, manuscripts that included metadata in their data depositions are generally cited to a higher degree as compared to those that do not (**Fig. 1d**). This trend could potentially indicate that proper data deposition increases the likelihood that others will re-use the data and thus reference the original manuscript in their work.

Where can we improve things?

Publishers. As gatekeepers of peer-reviewed scientific content, academic journals play a pivotal role in FAIR data sharing. As mentioned above, many publisher guidelines on data sharing are written in an *implicit* manner, suggesting various repositories and requesting adherence to community

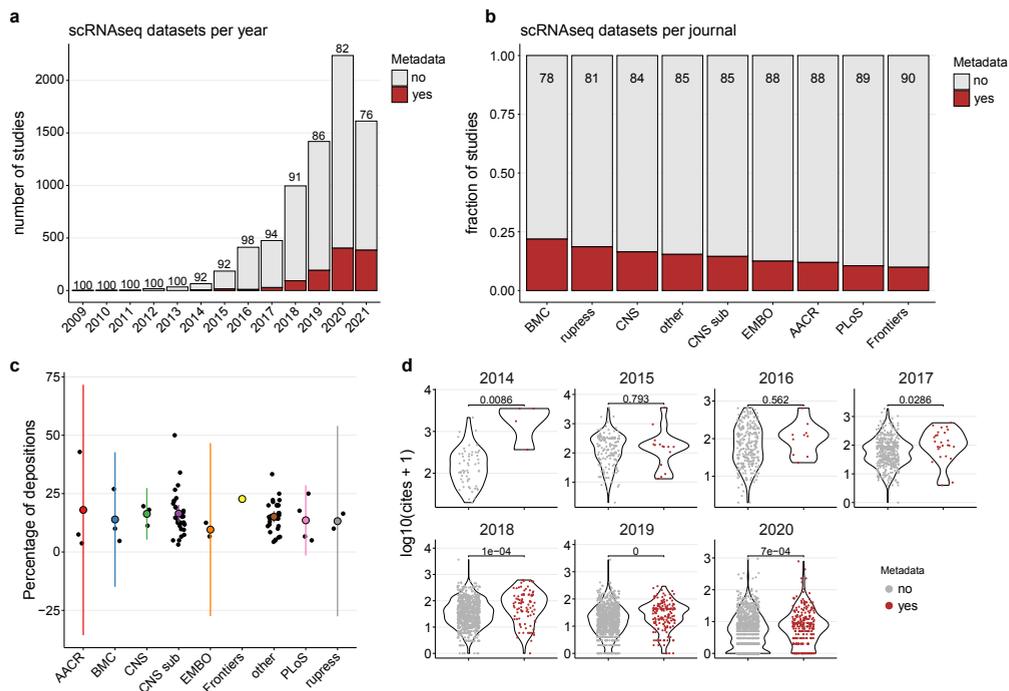


Figure 2. Inclusion of metadata in scRNAseq datasets deposited to GEO. (a) Number of depositions that did or did not include metadata in each year since 2009. Numbers on top of stacks denote the percentage of depositions without metadata. (b) Fraction of depositions that did or did not include metadata per publisher. Numbers on top of stacks denote the percentage of depositions without metadata. (c) Percentage of depositions that did or did not include metadata per publisher. Black dots indicate individual journals, colored dots indicate means, colored lines represent the 95% confidence interval. (d) Number of citations that a manuscript received since publication. Depicted as violin plots, dots indicate individual manuscripts. *P* values indicated in the plots were calculated by Wilcoxon signed-rank test, followed by Holm-Bonferroni correction.

standards such as the FAIR principles. I feel a more effective approach would be to make such guidelines *explicit*. For example, journals could provide a mock-up manuscript in which various commonly used data types are used, showing an impeccable example of how the different data components can be deposited and shared. The same mock-up would ideally be used across multiple journals from the same or different publishers to achieve homogeneous standards.

Peer-review offers another opportunity to ensure that the data underlying the results of the study can be readily assessed. The peer-review system is in place to ensure that published content is valuable to the community, it therefore makes sense to allocate more weight to data availability in the assessment of a manuscript. This could be achieved by either requesting reviewers to include an analysis of the efforts that were made by the authors to adhere to the FAIR principles in their assessment, or appointing a specialized reviewer whom specifically covers this aspect. To simplify this process, a short checklist could be offered to reviewers specifying points of interest. For instance, if both raw and processed data can be found and downloaded easily, and whether field-relevant repositories are used.

Repositories. As data repositories define how results are deposited, these entities are in a crucial position in the data sharing network. In my view, a big leap in the right direction would be for repositories to harmonize their guidelines with the scientific publishers. This would, as noted in the segment above, include the specification of explicit instructions on the contents of a deposition that match the requirements of major publishers and dummy uploads that are linked to mock-up manuscripts.

Furthermore, setting strict requirements for the inclusion of metadata and processed data alongside raw data would be appropriate. This prevents repositories from turning into ‘data dumpsites’, where the findings are *technically* shared, but re-use is severely complicated. Using scRNAseq uploads as an example, this would entail that sequencing results should be supplemented with *at least* one metadata file (e.g., results from clustering or pseudotime analyses) and *at least* one processed data file (e.g., results from gene-set enrichment or custom analyses). Again, implementation of these requirements will work best if they are set in collaboration with the scientific publishers.

Scientists. In my view, both publishers and repositories have a huge influence in shaping the data sharing environment. However, I do not believe that a perfect system can be built if we would have these two bodies policing the FAIRness of all data uploads. They should be here to guide and enable the process, but the ultimate responsibility needs to lie with the ones generating the data, the scientists.

Improving the way that scientists treat their data is not trivial; to be efficacious—at least in the long term—FAIR practices need to be instilled into the culture of the community. This would begin at the university level, for example including primers in the curriculum that discuss best practices on documenting one’s findings. Next, internships provide a perfect microcosm for an applied scientific project, making them crucial moments in teaching scientists-to-be the importance of reporting re-usable data. In practice, the supervisor could provide the student with a system that would make the obtained data FAIR within the lab, and essentially ready to publish. It would also be desirable to integrate the importance of finishing the internship with re-usable data in the grading system, giving this aspect equal weight as, for instance, quality of the practical work.

In large part, the lab culture is defined by its principal investigator. Therefore, it is important for them to provide guidance in, and promote adherence to, a FAIR system of data storage. This would not necessarily require group leaders to micro-manage filing systems used by their scientists, but can simply entail being an advocate for the desired ideals. For example, engaging in discussions on open science during meetings or encouraging diligence in these topics during the final stages of a project.

Concluding remarks

Everything taken together, it will take a concerted effort from all members of the scientific community to shape and—most importantly—maintain a FAIR data sharing network. As I have noted, I feel that a tight collaboration between publishers and repositories will result in a huge step

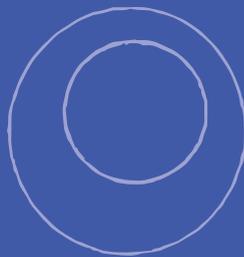
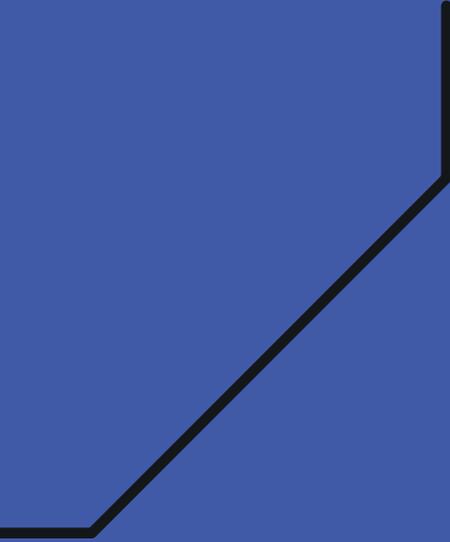
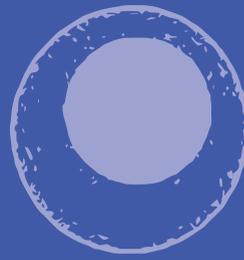
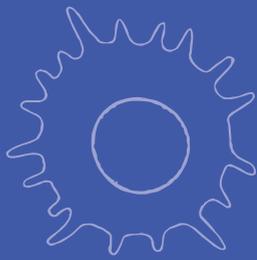
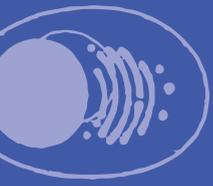
forward. Such a unified front would eliminate much of the ambiguity that currently complicates data deposition. Furthermore, it will likely be key to provide both scientists and publishers with appropriate incentives to implement the FAIR principles. For example, by more prominently integrating adherence to these principles into journal impact metrics, or awarding/penalizing investigators during the allocation of publicly funded research grants.

Technologies and their resulting datasets will continue to evolve, meaning that keeping FAIRness at a constant level represents a Red Queen's race. I feel it is important that we, as a community, do not lose sight of this principle, and remain flexible in the implementation of changes in order to 'keep up'.

References

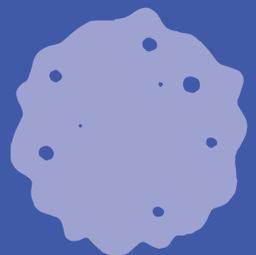
1. Laplane, L. *Cancer stem cells: philosophy and therapies*. (Harvard University Press, 2016).
2. Ito, M. et al. Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nat. Med.* 11, 1351–1354 (2005).
3. Yan, K. S. et al. The intestinal stem cell markers *Bmi1* and *Lgr5* identify two functionally distinct populations. *Proc. Natl. Acad. Sci. U.S.A.* 109, 466–471 (2012).
4. Wilson, A. et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* 135, 1118–1129 (2008).
5. Zhang, Y., Joe, G., Hexner, E., Zhu, J. & Emerson, S. G. Host-reactive CD8⁺ memory stem cells in graft-versus-host disease. *Nat Med* 11, 1299–1305 (2005).
6. Gattinoni, L. et al. Wnt signaling arrests effector T cell differentiation and generates CD8⁺ memory stem cells. *Nat. Med.* 15, 808–813 (2009).
7. Gattinoni, L. et al. A human memory T cell subset with stem cell-like properties. *Nat. Med.* 17, 1290–1297 (2011).
8. Fuertes Marraco, S. A. et al. Long-lasting stem cell-like memory CD8⁺ T cells with a naïve-like profile upon yellow fever vaccination. *Sci. Transl. Med.* 7, (2015).
9. Bouneaud, C., Garcia, Z., Kourilsky, P. & Pannetier, C. Lineage relationships, homeostasis, and recall capacities of central- and effector-memory CD8 T cells in vivo. *J Exp Med* 201, 579–590 (2005).
10. Wherry, E. J. et al. Lineage relationship and protective immunity of memory CD8 T cell subsets. *Nat Immunol* 4, 225–234 (2003).
11. Graef, P. et al. Serial transfer of single-cell-derived immunocompetence reveals stemness of CD8(+) central memory T cells. *Immunity* 41, 116–126 (2014).
12. Galletti, G. et al. Two subsets of stem-like CD8⁺ memory T cell progenitors with distinct fate commitments in humans. *Nat Immunol* 21, 1552–1562 (2020).
13. Zhou, B. et al. Single-cell RNA-Seq reveals the link between CD45 isoforms and tumor-infiltrating T cells heterogeneity in liver cancer. <http://biorxiv.org/lookup/doi/10.1101/2020.03.22.002824> (2020) doi:10.1101/2020.03.22.002824.
14. Herndler-Brandstetter, D. et al. KLRG1⁺ Effector CD8⁺ T Cells Lose KLRG1, Differentiate into All Memory T Cell Lineages, and Convey Enhanced Protective Immunity. *Immunity* 48, 716–729.e8 (2018).
15. Youngblood, B. et al. Effector CD8 T cells dedifferentiate into long-lived memory cells. *Nature* 552, 404–409 (2017).
16. Kastenmüller, W. et al. Peripheral Prepositioning and Local CXCL9 Chemokine-Mediated Guidance Orchestrate Rapid Memory CD8⁺ T Cell Responses in the Lymph Node. *Immunity* 38, 502–513 (2013).
17. Böttcher, J. P. et al. Functional classification of memory CD8⁺ T cells by CX3CR1 expression. *Nat Commun* 6, 8306 (2015).
18. Bangs, D. J. et al. CXCR3 regulates stem and proliferative CD8⁺ T cells during chronic infection by promoting interactions with DCs in splenic bridging channels. *Cell Reports* 38, 110266 (2022).
19. Shou, Y. et al. Redefining the Role of Lymphotoxin Beta Receptor in the Maintenance of Lymphoid Organs and Immune Cell Homeostasis in Adulthood. *Front. Immunol.* 12, 712632 (2021).
20. Crozat, K. et al. The XC chemokine receptor 1 is a conserved selective marker of mammalian cells homologous to mouse CD8 α ⁺ dendritic cells. *Journal of Experimental Medicine* 207, 1283–1292 (2010).
21. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016).
22. Information for Authors: Cell. <https://www.cell.com/cell/authors>.
23. Supplementary information | Nature. <https://www.nature.com/nature/for-authors/supp-info>.
24. Science Journals: editorial policies. <https://www.science.org/content/page/science-journals-editorial-policies>.





Chapter 7. Appendices

OR: "The part that people actually read."



English summary

In my (perhaps somewhat subjective) judgement, the immune system can be considered to be the most fascinating part of the body. This system encompasses a large diversity of cell types and an unbelievably large number of interactions and processes, all of which are the result of an age-old evolutionary arms race against the pathogens that surround us. While compiling this thesis, I have had the pleasure of studying various components of the immune system, and making small contributions to various topics within the field of immunology. Through this summary, I will provide some general background information on the immune system, and then briefly highlight the findings made in each chapter.

The central task of the immune system is to detect and then destroy ‘foreign entities’ that are potentially harmful to the body. For example, the detection and recognition of pathogens (e.g. bacteria or parasites) is based on the principle that these organisms look highly different from us. This recognition occurs mainly in two ways: (1) A *non-specific* manner, involving a large group of immune cells collectively termed the ‘innate’ immune system. These cells often recognize common patterns associated with pathogens, for example certain sugars that are prominently produced by a large family of bacteria and that are not produced by the human body. (2) A *specific* manner, mainly due to the action of T and B cells, collectively referred to as the ‘adaptive’ branch of the immune system. These T and B cells can recognize foreign proteins via specialized receptors that are present on the surface of these cells, the so-called T and B cell receptors, respectively. During the development of T and B cells, those receptors are assembled through a semi-random process, in which each new T or B cell acquires a unique receptor. Due to this process, there is always at least one T and B cell in the body that can *specifically* recognize a protein of a pathogen that is causing an infection. Once those pathogen-specific cells recognize their target, they become activated and begin dividing at a high rate to generate a large number of daughter cells, all of which can attack the pathogen in a targeted manner: B cells through the production of antibodies, and T cells by the production of cytotoxic molecules or substances that alert other cells to the ongoing danger. Ultimately, in most cases, the foreign entity is defeated and cleared through a coordinated collaboration between the many cells that operate in these two branches of the immune system.

Tricking the immune system

While the recognition of foreign proteins by the immune system is normally a great asset, it complicates in biomedical studies that investigate the physiological ‘behavior’ of various cell types. In these experiments, cells are frequently studied in the context of a living organism, such as mice (*Mus musculus*), where it is often crucial that the cells of interest can be distinguished from all other cells. The technological solution generally applied in these cases is to label the cells of interest with a fluorescent protein. This is done by inserting a gene that encodes a fluorescent

protein (often from some type of jellyfish or coral) into the DNA of the cells of interest, after which these cells start producing this protein themselves. After this genetic modification, the cells can be readily distinguished from other cells by their fluorescence. However, when such modified cells are transplanted into a mouse with a competent immune system, those cells are rejected from the host. Not entirely unexpected, since the modified cells now contain large amounts of a foreign protein (e.g., from a jellyfish). For this reason, many types of experiments are difficult to perform, if not completely impossible.

In **chapter 2**, we created a novel genetically modified ‘transgenic’ mouse model to solve this problem. Our approach is based on the following reasoning; modified cells are rejected because the fluorescent protein is foreign to the body. By ensuring that this protein is already present from birth, it will be recognized as part of the organism, and hence rejection may be prevented. To achieve this, we inserted the genetic code of several frequently used fluorescent proteins into the DNA of mouse embryonic stem cells, which we then used to create a transgenic mouse. For obvious reasons, we did not want all cells in this animal to become fluorescent, and we therefore modified the genetic code of the introduced genes such that they were no longer functional. This approach is described in **chapter 2**. In addition, we show that these transgenic animals have indeed become tolerant to a number of fluorescent proteins, allowing cells that are labeled with these proteins to be transplanted without complications. We have made this transgenic mouse strain available to the academic community.

Predicting the ‘appearance’ of a cell

All proteins in a cell have a certain lifespan, after which they are broken down into smaller fragments, generally referred to as peptides. The majority of these peptides are further degraded to their individual amino acids, and these are then recycled in new cellular processes. However, a small fraction of peptides is transported into the lumen of the endoplasmic reticulum, bound to a specialized protein named HLA (Human Leukocyte Antigen) class I, and finally transported to the cell surface to be presented to the immune system. Cytotoxic T cells continuously scan HLA class I-peptide complexes on the surface of cells with their T cell receptor, and in this way monitor whether something is amiss with the cell. For example, if cells are infected with a virus, or if mutated proteins are present (as is the case with many cancers), foreign peptides will be presented. These foreign peptides can subsequently be recognized by a cytotoxic T cell, after which the T cell becomes activated and can clear the affected cells.

The peptides presented via HLA class I thus determine how cells are ‘seen’ by a cytotoxic T cell. Understanding which peptides are presented by HLA class I is therefore of major interest in therapeutic approaches where activation of cytotoxic T cells is desirable, such as vaccination and cancer immune-therapy.

In **chapter 3** we aimed to acquire more insight into this process by directly measuring which peptides were bound to HLA class I in a number of different melanoma lines. Using these data, we compared genes (and proteins) that either did, or did not, yield HLA-bound peptides using a large database of

gene and protein characteristics. This database contained more than 7,000 characteristics, including the occurrence of sequence motifs and potential protein modification sites. We observed that the genes (and proteins) from which presented peptides were derived often contained or lacked certain of these characteristics, with, for instance, a clear predictive power of certain predicted post-translational modifications. We then incorporated the database of protein and gene characteristics into an algorithm that we trained to predict whether or not peptides are presented by HLA. Finally, we were able to show that the incorporation of this information greatly improved the predictive value of such algorithms.

The role of dormant T cells in immunological memory

Once a T cell becomes activated during an infection, it will start to divide at a high rate. The aim of this proliferation is to generate a large number of daughter cells that all share the same T cell receptor, and can thus all specifically recognize and attack cells that present a specific antigen. When the infection has been successfully cleared, the vast majority of these ‘effector T cells’ are no longer needed and die off. A fraction of the pathogen-specific T cells remains alive for many years as a relatively stable population. During a reinfection, these ‘memory T cells’ are able to generate a new wave of effector T cells much faster than during the first infection, thereby offering the body long-term protection against the pathogen. This principle is central to the prophylactic activity of vaccination and, for this reason, the formation and function of memory T cells (and B cells) is a widely studied topic in immunology.

As noted above, cell division is a key feature of the T cell response; however, little is known about the relationship between T cell proliferation during an infection and the formation of T cell memory. In **chapter 4** we set out to study this process *in vivo*. To this end, we developed a synthetic transgenic construct (termed DivisionRecorder) that contains an inactive gene that encodes a fluorescent protein. The DivisionRecorder was designed in such a way that during each cell division there is a small probability that the gene activates, switching the cell to an irreversible fluorescent state. This means that we were able to use the fraction of fluorescent cells within a population as a measure for the amount of proliferation that had occurred in the past. Using this method, we were able to determine the relative number of cell divisions that different groups of T cells had undergone during, and after, infection. From these data it became clear that there is a large degree of heterogeneity in the number of cell divisions that memory T cells have undergone. In addition, we observed that this heterogeneity was associated with distinct cell characteristics. For example, we identified a group of memory T cells that have undergone only few cell divisions during the immune response. Furthermore, these ‘dormant’ memory T cells were found to have the highest potential to rapidly divide during reinfection, making these cells a crucial component of immunological memory.

Priming tumors

Like pathogens, cancer cells can be recognized by the immune system as foreign. Indeed, during its development, a tumor is ‘infiltrated’ by many different types of immune cells. In many cases,

however, the immune system is unable to reject the cancer cells. The subsequently established tumor micro-environment contains a large number of immune-supportive and immune-dampening molecules that exist in a stalemate. For example, fibroblasts may secrete proteins named chemokines that act as signaling molecules to attract different types of immune cells, whereas tumor cells can increase cell surface levels of membrane-bound inhibitory proteins (such as CD47 or PD-L1) that prevent immune cells from attacking the malignancy.

Proteins consist of a chain of distinct amino acids that can—due to their sequence—fold into complex three-dimensional structures. The formation of such structures is crucial for the function of a protein, but is often not sufficient. Additional modifications to the amino acid residues may be necessary for the ‘maturation’ of the protein. An example of such a modification is the cyclization of glutamine or glutamic acid residues located at the start of a protein chain, which is crucial for the function of certain membrane-bound and secreted proteins, including several chemokines and inhibitory protein CD47. The formation of this modification is catalyzed by the enzyme glutaminy-peptide cyclotransferase (QPCTL), therefore making this enzyme a potentially important regulator of the balance between immune-supportive and immune-dampening molecules in the tumor micro-environment.

In **chapter 5** we used genetically modified mice and melanoma cells to investigate tumor growth and characteristics of the tumor micro-environment in a scenario in which QPCTL is rendered inactive. Using this approach, we observed that tumor growth was unaffected, but that the composition of tumors was significantly altered in the absence of QPCTL. Interestingly, we found that these alterations to the tumor micro-environment were indicative of a more inflamed milieu. This led us to conclude that QPCTL-deficient tumors might be more sensitive to additional immunotherapeutic treatments. We confirmed this by combining QPCTL deletion with an immune activating therapy (anti-PD-L1 therapy), a combination treatment that resulted in slower tumor growth and sporadic tumor regression. These findings suggest that QPCTL is an interesting target in the treatment of cancer, as an addition to existing cancer immunotherapies.

Nederlandse samenvatting

Naar mijn (wellicht wat subjectieve) oordeel, is het immuunsysteem het meest fascinerende onderdeel van het lichaam. Dit systeem bestaat uit vele cel typen en een onwaarschijnlijk grote hoeveelheid interacties en processen, die allen het gevolg zijn van een eeuwenoude evolutionaire wapenwedloop met de ziekteverwekkers om ons heen. Tijdens het samenstellen van dit proefschrift heb ik het genoeg gehad om verschillende componenten van het immuunsysteem te bestuderen, en bijdragen te leveren aan diverse onderwerpen binnen de immunologie. Middels deze samenvatting zal ik wat algemene achtergrondinformatie geven over het immuunsysteem, en vervolgens het onderwerp en de bevindingen van elk hoofdstuk kort toelichten.

De centrale taak van het immuunsysteem is om 'lichaamsvreemde entiteiten' die potentieel schadelijk zijn voor het lichaam op te sporen en vervolgens te vernietigen. Het opsporen en herkennen van pathogenen (i.e., ziekteverwekkers zoals bacteriën of parasieten) is gebaseerd op het algemene principe dat deze organismen er anders uitzien dan wij. De herkenning van zulke 'lichaamsvreemdheid' gebeurt overwegend op twee manieren: (1) Een relatief *aspecifieke* manier, waarbij een grote groep immuun cellen is betrokken die collectief het 'angeboren' immuunsysteem genoemd wordt. Deze cellen herkennen vaak algemene patronen die geassocieerd zijn met pathogenen, bijvoorbeeld bepaalde suikerverbindingen die uitsluitend door een grote familie van bacteriën worden geproduceerd en dus lichaamsvreemd zijn voor mensen. (2) Een *specifieke* manier, waarbij er primair een rol is voor de T en B cellen, die gezamenlijk de 'adaptieve' tak van het immuunsysteem vormen. Deze T en B cellen kunnen lichaamsvreemde eiwitten herkennen via speciale eiwitten die aanwezig zijn op hun celmembraan, respectievelijk de zogenaamde T en B cel receptoren. Tijdens de ontwikkeling van T en B cellen worden deze receptoren via een semi-willekeurig proces samengesteld, waardoor elke nieuwe T of B cel een unieke receptor verkrijgt. Door dit proces van diversificatie is er tijdens een infectie altijd wel minstens één T en B cel in het lichaam die het vermogen heeft een eiwit (of een fragment van een eiwit) van het desbetreffende pathogeen te herkennen. Zodra deze pathogeenspecifieke cellen hun doelwit herkennen neemt de activiteit van deze cellen sterk toe en beginnen de cellen in een hoog tempo te delen om zo een grote hoeveel dochtercellen te genereren welke allen—op een doelgerichte manier—het pathogeen kunnen aanvallen: B cellen doen dit laatste via de productie van antilichamen, en T cellen door de productie van cytotoxische moleculen of stoffen die andere cellen alarmeren voor het aanwezige gevaar. Uiteindelijk wordt, in de meeste gevallen, de lichaamsvreemde entiteit verslagen en opgeruimd door een gecoördineerde samenwerking tussen de vele cellen die opereren in deze twee takken van het immuunsysteem.

Omzeilen van herkenning door het immuunsysteem

De herkenning van lichaamsvreemde eiwitten—en de daaropvolgende afstoting—door het

immuunsysteem is normaliter een groot goed. In de biomedische wetenschap kunnen dit soort immunoreacties echter onwenselijk zijn. Een relevant voorbeeld hiervan zijn studies waarin het ‘natuurlijke gedrag’ van celtypen wordt onderzocht met behulp van fluorescente eiwitten. In dit soort experimenten worden cellen vaak bestudeerd in de context van een levend organisme, zoals de muis (*Mus musculus*), waarbij het cruciaal is dat de cellen van interesse goed kunnen worden onderscheiden van alle andere cellen in het lichaam. De oplossing die hiervoor over het algemeen wordt toegepast is om de cellen die bestudeerd worden te markeren met een fluorescent eiwit. Dit wordt gedaan door een gen dat codeert voor een fluorescent eiwit—vaak afkomstig van een kwal of koraal—in te brengen in het DNA van de cellen van interesse, waarna deze cellen zelf dit eiwit gaan produceren. Na deze genetische modificatie kunnen de cellen makkelijk onderscheiden worden van andere cellen op grond van hun fluorescentie. Een fundamenteel probleem bij deze methode is dat als dit type genetisch gemodificeerde cellen naar een muis met een competent immuunsysteem worden getransplanteerd, deze cellen vaak afgestoten worden. Niet geheel onverwacht, aangezien de gemodificeerde cellen nu een grote hoeveelheid lichaamsvreemd eiwit (b.v. afkomstig van een kwal) bevatten. Om deze reden zijn een groot aantal type experimenten moeilijk uit te voeren of geheel onmogelijk.

In **hoofdstuk 2** hebben wij een nieuw type genetisch gemodificeerde ‘transgene’ muis gecreëerd om dit probleem op te lossen. In onze aanpak zijn we uitgegaan van de volgende redenering: De gemodificeerde cellen worden afgestoten omdat het fluorescente eiwit lichaamsvreemd is. Door ervoor zorgen dat dit eiwit al vanaf de geboorte aanwezig is zal het als lichaamseigen herkend worden, en zal afstoting worden voorkomen. Om dit te bereiken hebben we de genetische code van verschillende fluorescente eiwitten ingebracht in het DNA van embryonale stamcellen van een muis, die we vervolgens hebben gebruikt om een transgene muis te creëren. Omdat we vanzelfsprekend niet wilden dat alle cellen in dit dier fluorescent zouden worden, hebben we de genetische code van de ingebrachte genen zo aangepast dat deze niet meer functioneel zijn. Deze methode staat beschreven in **hoofdstuk 2**. In dit hoofdstuk laten we vervolgens zien dat de transgene dieren inderdaad tolerant zijn geworden voor een aantal fluorescente eiwitten, waardoor cellen die met deze fluorescente eiwitten gemarkeerd zijn zonder problemen getransplanteerd kunnen worden. Deze transgene muizen stam hebben we beschikbaar gesteld aan de academische gemeenschap.

Het ‘uiterlijk’ van een cel voorspellen

Alle eiwitten in een cel hebben een bepaalde levensduur, waarna ze worden afgebroken in kleinere fragmenten, ook wel peptiden genoemd. Een groot deel van deze peptiden wordt verder afgebroken tot losse aminozuren, voor hergebruik in nieuwe cellulaire processen. Een kleiner deel wordt echter gebonden aan een gespecialiseerd eiwit genaamd HLA (Human Leukocyte Antigen) klasse I en uiteindelijk vervoerd naar het cel oppervlakte om gepresenteerd te worden aan het immuunsysteem. Cytotoxische T cellen scannen continu HLA-klasse I-peptiden complexen op het oppervlak van cellen met hun T cel receptor, en houden op deze manier in de gaten of er iets mis is met de cel. Wanneer cellen bijvoorbeeld geïnfecteerd zijn door een virus of als er gemuteerde eiwitten aanwezig zijn (zoals het geval is bij veel kankers), dan zullen er lichaamsvreemde peptiden gepresenteerd

worden. Deze lichaamsvreemde peptiden kunnen vervolgens door een T cel herkend worden, waarna de T cel geactiveerd raakt en de aangetaste cellen kan gaan opruimen.

In **hoofdstuk 3** hebben we meer inzicht verkregen in het proces van peptide presentatie door direct te meten welke peptiden gebonden waren aan HLA klasse I in een aantal verschillende melanoomlijnen. Middels deze data konden we genen (en eiwitten) vergelijken die al dan niet HLA-gebonden peptiden opleverden, met behulp van een grote database met gen- en eiwitkenmerken. Deze database bevatte meer dan 7.000 kenmerken, waaronder de frequentie van bepaalde motieven in de DNA-sequentie en mogelijke eiwitmodificaties. Met deze analyse hebben we vastgesteld dat genen (en eiwitten) waarvan gepresenteerde peptiden afkomstig zijn, vaak bepaalde kenmerken bevatten of juist missen. We vonden bijvoorbeeld dat aminozuur sequenties die gemodificeerd kunnen worden met een acetyl- of ubiquitine-groep verrijkt zijn in eiwitten die gepresenteerde peptiden bevatte. Vervolgens hebben we de database met eiwit- en gen-kenmerken opgenomen in een algoritme dat we hebben getraind om HLA klasse I gepresenteerde peptiden te voorspellen. Ten slotte konden we aantonen dat het incorporeren van deze informatie de voorspellende waarde van dergelijke algoritmen sterk verbeterde.

De rol van slapende T cellen in immunologisch geheugen

Op het moment dat een T cel gedurende een infectie geactiveerd raakt, gaat deze in een snel tempo delen. Het doel van dit proces is om een grote hoeveelheid dochtercellen te creëren die allen dezelfde T cel receptor tot expressie brengen, en dus allen specifiek de geïnfecteerde cellen kunnen herkennen en aanvallen. Als de infectie succesvol is afgeweerd sterft het overgrote deel van deze 'effector T cellen' af, ze zijn niet meer nodig. Een fractie van de pathogeen-specifieke T cellen blijft gedurende vele jaren aanwezig in het lichaam als een relatief stabiele populatie. Deze 'geheugen T cellen' kunnen tijdens een her-infectie veel sneller opnieuw een nieuwe golf effector T cellen genereren, en bieden het lichaam hierdoor langdurige bescherming tegen het desbetreffende pathogeen. Dit principe staat bijvoorbeeld centraal in de werking van vaccinatie, en om deze reden is de vorming en werking van geheugen T cellen een breed bestudeerd onderwerp in de immunologie.

Zoals hierboven opgemerkt, is celdeling een belangrijk kenmerk van de T-celrespons; er is echter weinig bekend over de relatie tussen proliferatie tijdens een infectie en de vorming van geheugen T cellen. Om dit proces te bestuderen hebben we in **hoofdstuk 4** een synthetisch transgeen construct (genaamd DivisionRecorder) ontwikkeld dat gebruikt kan worden om celdelingen te meten. De DivisionRecorder bevat een inactief gen dat codeert voor een fluorescerend eiwit. Dit gen is zodanig ontworpen dat er tijdens elke celdeling een kleine kans is dat het gen wordt geactiveerd, waardoor de cel onomkeerbaar fluorescent wordt. Dit betekent dat we de fractie fluorescente cellen binnen een populatie konden gebruiken als maat voor de hoeveelheid proliferatie die in het verleden had plaatsgevonden. Door middel van deze methode konden we relatieve bepalingen uitvoeren van de hoeveelheid celdelingen die verschillende groepen T cellen hadden ondergaan tijdens en na een infectie. Uit deze data werd duidelijk dat er een grote mate van heterogeniteit is in de hoeveelheid celdelingen die geheugen T cellen hebben ondergaan, en tevens vonden we dat deze heterogeniteit

geassocieerd was met specifieke karakteristieken. Zo vonden we dat er een groep geheugen T cellen bestaat die weinig celdelingen hebben ondergaan gedurende de immuunrespons. Deze ‘slapende’ geheugen T cellen bleken het hoogste potentieel te bezitten om tijdens een her-infectie opnieuw in snel tempo te gaan delen, wat deze cellen cruciaal maakt voor het immunologische geheugen.

Tumoren op scherp zetten

Net als ziekteverwekkers kunnen kankercellen door het immuunsysteem als lichaamsvreemd worden herkend. Tijdens de ontwikkeling van een tumor wordt deze ‘geïnfiltreerd’ door veel verschillende soorten immuun cellen. In veel gevallen is het immuunsysteem echter niet in staat de kankercellen af te stoten. De vervolgens tot stand gebrachte micro-omgeving van de tumor bevat een groot aantal immuun-ondersteunende en immuun-dempende moleculen die tezamen verweven zitten in een impasse. Fibroblasten kunnen bijvoorbeeld eiwitten uitscheiden (die chemokinen worden genoemd) die fungeren als signaal moleculen om verschillende soorten immuun cellen aan te trekken, terwijl tumorcellen de hoeveelheid van membraangebonden immuun-remmende eiwitten (zoals CD47 of PD-L1) verhogen die de functionaliteit van immuun cellen kunnen blokkeren.

Eiwitten bestaan uit een aaneenschakeling van verschillende aminozuren die zich—door middel van hun volgorde—in complexe driedimensionale structuren kunnen vouwen. De vorming van zulke structuren is essentieel voor de functie van een eiwit, maar is vaak niet voldoende. Verdere modificaties aan de aminozuurresiduen kunnen nodig zijn voor de ‘maturatie’ van het eiwit. Een voorbeeld van een dergelijke modificatie is de cyclisatie van glutamine- of glutaminezuurresiduen die zich aan het begin van een eiwitketen bevinden, wat cruciaal is voor de functie van bepaalde membraangebonden en uitgescheiden eiwitten, waaronder verschillende chemokinen en immuun-remmend eiwit CD47. De vorming van deze modificatie wordt gekatalyseerd door het enzym glutaminyl-peptide cyclotransferase (QPCTL), waardoor dit enzym een potentieel belangrijke regulator is van de balans tussen immuun-ondersteunende en immuun-dempende moleculen in de micro-omgeving van de tumor.

In **hoofdstuk 5** hebben we gebruik gemaakt van genetisch gemodificeerde muizen en melanoomcellen om de groei en ontwikkeling van tumoren te onderzoeken in een scenario waarin QPCTL inactief is. Middels deze strategie vonden we dat, alhoewel tumorgroei onaangetast was, de compositie van de tumor significant veranderd was. De som van alle veranderingen die we vonden in de afwezigheid van QPCTL duidde erop dat de omgeving binnen de tumor veel meer weg had van een actieve ontsteking. Hieruit concludeerden we dat QPCTL-deficiënte tumoren wellicht sensitiever zouden zijn voor verdere immuuntherapie. Dit bevestigden we door QPCTL-uitschakeling te combineren met een immuun-activerende therapie (anti-PD-L1 therapie), waarop we langzamere tumorgroei en soms zelfs tumorregressie observeerden. Onze bevindingen laten zien dat QPCTL een interessant doelwit is om te inhiberen in combinatie met bestaande immuuntherapieën in de behandeling van kanker.

List of publications

Kaspar Bresser, Feline E. Dijkgraaf, Colin E. J. Pritchard, Ivo J. Huijbers, Ji-Ying Song, Jan C. Rohr, Ferenc A. Scheeren & Ton N. Schumacher. A mouse model that is immunologically tolerant to reporter and modifier proteins. *Communications Biology*, 3, 273 (29 May 2020). <https://doi.org/10.1038/s42003-020-0979-0>

Lianne Kok, Feline E. Dijkgraaf, Jos Urbanus, **Kaspar Bresser**, David W. Vredevoogd, Rebeca F. Cardoso, Leïla Perié, Joost B. Beltman, Ton N. Schumacher. A committed tissue-resident memory T cell precursor within the circulating CD8⁺ effector T cell pool. *Journal of Experimental Medicine*, 217 (10): e20191711 (5 October 2020). doi: <https://doi.org/10.1084/jem.20191711>

Esmee P. Hoefsmit, Elisa A. Rozeman, Trieu My Van, Petros Dimitriadis, Oscar Krijgsman, Jordan W. Conway, Ines Pires da Silva, Jacqueline E. van der Wal, Steven L. C. Ketelaars, **Kaspar Bresser**, Annegien Broeks, Ron M. Kerkhoven, Jason W. Reeves, Sarah Warren, Pia Kvistborg, Richard A. Scolyer, Ellen W. Kapiteijn, Daniel S. Peeper, Georgina V. Long, Ton N. M. Schumacher and Christian U. Blank. Comprehensive analysis of cutaneous and uveal melanoma liver metastases. *The Journal of Immunotherapy of Cancer* (2): e001501 (8 December 2020). doi: 10.1136/jitc-2020-001501.

Kaspar Bresser, Meike E. W. Logtenberg, Mireille Toebes, Natalie Proost, Justin Sprengers, Bjorn Siteur, Manon Boeije, Lona J. Kroese and Ton N. Schumacher. QPCTL regulates macrophage and monocyte abundance and inflammatory signatures in the tumor microenvironment, *OncoImmunology*, 11:1 (17 March 2022), DOI: 10.1080/2162402X.2022.2049486

Kaspar Bresser, Lianne Kok, Arpit C. Swain, Lisa A. King, Laura Jacobs, Tom S. Weber, Leïla Perié, Ken R. Duffy, Rob J. de Boer, Ferenc A. Scheeren and Ton N. Schumacher. Replicative history marks transcriptional and functional disparity in the CD8⁺ T cell memory pool. *Nature Immunology* 23, 791–801 (7 April 2022). <https://doi.org/10.1038/s41590-022-01171-9>

Curriculum vitae

Kaspar Bresser was born on August 18th 1991 in Voorburg, the Netherlands. After completing his secondary education and obtaining a HAVO diploma from the Huygens Lyceum, Kaspar pursued an education in the academic sciences. Joining the 'Hoger Laboratorium Onderwijs' at the Hogeschool Leiden in 2008, he first trained as a general-purpose lab technician, and later specialized in molecular biology. During his time at the Hogeschool Leiden, he had his first experience with the field of tumor immunology in the lab of Thorbald van Hall at the Leiden University Medical Center (LUMC) investigating the potential role of the MHC class II chaperone molecule 'invariant chain' in the occurrence of unconventional T cell epitopes in cancer cells. After his work at the LUMC, Kaspar moved to the lab of Ton Schumacher at the Netherlands Cancer Institute (NKI), where he assisted in the establishment of a T cell receptor gene-capture approach and helped identify various T cells receptors directed against shared tumor antigens. Upon completing his traineeships, Kaspar received Bachelor of applied Science degree from the Hogeschool Leiden in 2012 and subsequently enrolled in the study 'Biomedical Sciences' at the University of Amsterdam (UvA), specializing in Infection and Immunology. During his studies at the UvA, he moved back to the NKI to work on the development of genetic tools that would allow for the identification of repressive genetic elements in the lab of Bas van Steensel, furthering his training in molecular biology and genetic engineering. Next, to conclude his Masters, Kaspar moved to the lab of Hidde Ploegh at the Whitehead Institute for Biomedical Research in Cambridge. Here he worked on several projects centered around various applications of alpaca-derived single-domain antibodies, focusing mainly on an approach to re-direct target proteins to the proteasomal degradation pathway. In 2014 Kaspar graduated cum laude from the UvA and obtained his Master of Science degree. Later that year he returned to the NKI perform his graduate studies, pursuing a diverse set of projects in the field of (onco-)immunology under the supervision of Ton Schumacher. The results of those endeavors are detailed in this thesis. From 2022 onward, Kaspar joined the group of Monika Wolkers to study the intricacies of post-transcriptional regulation in T cells.

Dankwoord

Lianne, ik weet dat je het verschrikkelijk vindt om in de spotlight gezet te worden. Helaas, hier volgen lieve woorden en complimentjes, hou je vast. Jouw aanstekelijke passie voor 'T cell lineages' en 'differentiation states' zijn bepalend geweest voor mijn huidige academische pad. Dit bleek de gateway-drug te zijn that sealed my fate as an immunologist. Ik heb enorm veel geleerd van jouw toffe wetenschappelijke ideeën, elegante experimenten, en genadeloze discussietactieken. Je daagt me constant uit, in de wetenschap en daarbuiten. Je maakt mij moedig genoeg om keuzes te maken die ik eigenlijk heel eng vindt. Dankjewel voor alle dingen.

Ben, we've traversed the scientific jungle side-by-side from the start, so it's only fitting to face my defense with you beside me. You're a great friend and an excellent sparring partner. You've kept me creative and critical throughout our time together, and I hope we'll be able to continue our separate paths together.

Ton, ik kan me niet inbeelden dat ik het bedrijven van wetenschap op een andere plek, of van een ander persoon, beter had kunnen leren. Het meedogenloos trimmen van een groot wetenschappelijk werk naar een pakketje van 3,000 woorden is voor velen een exercitie in cynisme, maar je hebt me laten zien dat het vooral ook een leuke uitdaging kan zijn, ik ga het sparren en brainstormen erg missen, en hoop dat we elkaar nog af en toe kruisen.

Ferenc, dankjewel voor al je bijstand in wat soms voelde als het veilig landen van een brandend vliegtuig. Je hebt me een goede duw in de rug gegeven tijdens het begin van mijn werk, waar ik enorm veel aan gehad heb, ik ben enorm dankbaar om je als mentor en als vriend te hebben, en kijk uit naar nog veel koffietjes en filosoferen over wetenschap.

Mireille en Jos, sfeer-bepalers, gezelligheid-pushers, relativatie-afdwingers. Zonder jullie had alles er ongelofelijk anders uit gezien. Jullie gaven de afdeling de veiligheid van een tweede thuis. Bedankt hiervoor.

Feline en Meike, vrienden en gewaardeerde collega's. Wat was het heerlijk om met jullie wetenschap te mogen bedrijven. We hebben samen toffe projecten gecreëerd, maar vooral enorm veel lol gehad over alles en over niets. It was an honor and a privilege.

Pap en mam, jullie bodemloos vertrouwen in al mijn, soms toch wel twijfelachtige beslissingen, heeft mij altijd de rust gegeven om mijn eigen pad te bewandelen. Dank jullie wel voor het meegeven van een kritische kijk op de wereld om me heen. Dank jullie wel voor alle liefde. Dank jullie wel voor alles.

Anne, Mirjam, Maarten, Max, Steven, vriendjes en mede-filosofen. Dank jullie voor alle inzichten, al het gegiegel, en alle volledig uitgedachte projecten waar nooit een pipet voor is opgetild. Alle dalen en pieken waren grandioos, omdat ik het met jullie mocht delen.

Simone, bedankt voor alle liefde, steun en vriendschap die je met me gedeeld hebt. Je wist een sociaal vuurtje bij me aan te stoken waarvan ik niet wist dat ik 'm had. Daarnaast heb je me door een groot deel van dit thesis heen geloodst, en me bewust gehouden dat er ook een wereld bestaat buiten pipetteren en naar een scherm staren. Dankjewel!

Kim, wanneer kom je nou eens terug van je vakantie in Drenthe... Nee hoor, prima. Over de telefoon kan je me ook prima stabiel houden! Dankjewel dat een paar woorden altijd genoeg zijn om een verhaal over te brengen. En dankjewel voor je bodemloze capaciteit om mijn frustraties aan te horen. Je bent een held.

Bas, mede-zoogdier, mede-observeerder. Dank voor je continue bijdrage aan mijn kijk op het leven. En een bijna even grote dank voor je bijdrage aan het afmaken van dit thesis.

