

© Copyright 2026

Hunter Lee Colegrove

Clonal dynamics of somatic evolution and gene therapy in cancer predisposition syndromes

Hunter Lee Colegrove

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee:

Alison F. Feder, Chair

Raymond J. Monnat Jr.

Rosa Ana Risques

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Clonal dynamics of somatic evolution and gene therapy in cancer predisposition syndromes

Hunter Lee Colegrove

Chair of the Supervisory Committee:

Alison F. Feder

Department of Genome Sciences

Preventing cancer before malignancy is established requires identifying which early somatic events drive cancer progression. This problem is especially important in inherited cancer predisposition syndromes, where germline pathogenic variants put individuals at elevated cancer risks compared to the general population. In this dissertation, I first investigate how an inherited *TP53* pathogenic variant alters the somatic evolutionary dynamics across normal tissues in a family with Li-Fraumeni syndrome, a cancer predisposition syndrome associated with high cancer penetrance and early onset. Using ultra-deep duplex sequencing, I show that a germline *TP53* variant is associated with parallel evolution of a second-hit *TP53* hotspot mutation across organ sites and with elevated mutagenesis in regions of the genome that evolve neutrally in the general population. Then, I investigate whether in situ gene correction reduces cancer risk in Fanconi Anemia, a bone marrow failure and cancer predisposition syndrome, using a computational agent-based model of the oral mucosa. I show that cellular competition between gene corrected and uncorrected cells in an epithelial tissue context is a central determinant of therapeutic success and that gene therapy has the potential to suppress the accumulation of pro-tumorigenic mutations in a cancer predisposed context. Together, this work advances understanding of how germline variation shapes pre-malignant somatic evolution across tissue contexts and provides a quantitative framework for therapeutic prevention in individuals with inherited cancer risk.

TABLE OF CONTENTS

	Page
List of Figures	iv
Chapter 1: Introduction	1
1.1 Cancer progression is an evolutionary process	1
1.1.1 Somatic mosaicism	1
1.1.2 Pre-malignant clonal expansion	4
1.1.3 Evolutionary trajectories toward cancer progression	5
1.2 Inherited driver mutations shorten the path to tumorigenesis.	6
1.2.1 Cancer predisposition syndromes.	6
1.3 Open questions in the field of cancer predisposition syndromes.	8
1.4 Technologies to understand clonal evolution in the context of cancer predisposition syndromes	9
1.4.1 In silico modeling approaches.	9
1.4.2 Sequencing-based approaches	10
1.5 Preventative therapeutic approaches for cancer predisposition syndromes . . .	12
1.5.1 Natural reversion and therapeutic strategies.	12
1.6 Thesis motivation.	13
1.7 Previous publication and co-authorship of dissertation content	14
Chapter 2: Ultra-deep duplex sequencing reveals unique features of somatic evolution in the normal tissues of a family with Li-Fraumeni syndrome .	15
2.1 Introduction	15
2.2 Results	17
2.2.1 Ultra-deep sequencing of a family carrying p.R181H and healthy controls	17

2.2.2	Germline p.R181H does not affect most CHIP clonal expansion but is associated with increased mutation frequency in the mutagenesis panel and reduced positive selection in <i>TP53</i>	20
2.2.3	Chemotherapy increases <i>TP53</i> clonal expansions in the blood of a germline p.R181H carrier and non-carrier	24
2.2.4	Mutation frequency and signatures vary across tissue contexts	28
2.2.5	<i>TP53</i> R248Q/W is a mutational hotspot recurrently found in several tissues.	32
2.2.6	Chromosomal but not tissue-context accounts for differences in selection pressures in <i>TP53</i>	33
2.3	Discussion	34
2.4	Materials & Methods.	39
2.4.1	Sample collection	39
2.4.2	Duplex sequencing and variant calling	40
2.4.3	Mutation classification and pathogenicity	41
2.4.4	Testing for selection.	42
2.4.5	Regression analysis	44
2.4.6	Mutational spectrum and signature analysis.	44
2.4.7	Data access.	45
2.4.8	Code availability.	45
2.5	Competing interest statements.	46
2.6	Funding.	46
2.7	Author contributions.	46
2.8	Acknowledgements	47
Chapter 3:	Epithelial competition determines gene therapy potential to suppress Fanconi Anemia oral cancer risk	48
3.1	Introduction	48
3.2	Results	51
3.2.1	<i>In silico</i> model of gene correction in the oral epithelia	51
3.2.2	Gene correction requires a strong proliferative advantage to spread over clinically relevant time periods	52
3.2.3	Optimizing transgene delivery for maximal corrected tissue coverage	55
3.2.4	Gene correction reduces <i>TP53</i> ⁻ clonal expansion in genomically unstable FA epithelium	59

3.3	Discussion	63
3.4	Acknowledgements	69
3.5	Materials & Methods.	69
3.5.1	Overview of the model	69
3.5.2	Interrogating clonal expansion of corrected cells in an FA background 72	
3.5.3	Lineage competition between gene-corrected cells and <i>TP53</i> mutants .	72
3.5.4	Parametrizing the proliferative advantage and mutation rate of <i>TP53</i> [−] mutations	74
3.5.5	Code availability.	79
Chapter 4:	Conclusions	80
4.1	Importance of predicting malignant transformation and early cancer detection 80	
4.2	Major goals for the study of cancer predisposition syndromes.	81
4.2.1	Background specific effects across contexts	82
4.2.2	Therapeutic considerations	86
4.3	Concluding remarks	90
Appendix A:	Supplementary figures for Chapter 2	91
Appendix B:	Supplementary figures for Chapter 3	97
Bibliography		143

LIST OF FIGURES

Figure Number		Page
2.1	Ultra-deep sequencing of somatic mutations in subjects with and without Li-Fraumeni Syndrome.	18
2.2	Ultra-deep characterization of CHIP mutation and selection in the blood of individuals with and without Li-Fraumeni Syndrome.	21
2.3	Ultra-deep characterization of <i>TP53</i> mutation and selection in the blood of individuals with and without Li-Fraumeni Syndrome.	25
2.4	Multi-tissue ultra-deep characterization of <i>TP53</i> mutation and selection in an individual with Li-Fraumeni Syndrome and esophageal adenocarcinoma. . . .	30
3.1	Modeling gene therapy in FA HSCs and oral epithelium.	49
3.2	Mucosal gene therapy depends on a strong proliferative advantage of gene-corrected cells.	54
3.3	Higher corrected cell numbers and optimal spacing can increase the likelihood of tissue replacement despite clonal interference.	57
3.4	Gene correction reduces spread of <i>TP53</i> mutations through FA tissue.	61
A.1	Duplex depth comparison across genomic regions and samples.	91
A.2	CHIP coding regression of MF and age by gene.	92
A.3	CHIP coding mutational burden association with age, chemotherapy, and LFS. 93	
A.4	Comparison of pathogenicity of blood mutations in CHIP/AML genes in subjects with and without <i>TP53</i> p.R181H germline mutation.	94
A.5	Loss of heterozygosity (LOH) in Li-Fraumeni patient samples based on p.R181H allele frequency.	94
A.6	Concordance of <i>TP53</i> mutations identified in blood samples and tissues from LFS01.	95
A.7	Size and tissue distribution of <i>TP53</i> mutations observed in three or more samples.	96

B.1	Gene correction reduces spread of $TP53$ mutations through FA tissue over time.	97
B.2	Elevated mutation rates increase the number of $TP53^-$ clones, while increased persistence coefficients increase $TP53^-$ clone sizes.. . . .	98
B.3	Basal cell division rates in the default HomeostaticEpidermis model.. . . .	99
B.4	Parameterizing the proliferative advantage and mutation rate of $TP53$ mutations.	100

ACKNOWLEDGMENTS

Thank you, Siokwil

The completion of this dissertation marks the culmination of an arduous and rewarding five year journey. Starting this PhD, I wanted to push myself out of my comfort zone and develop myself as a scientist. Well, it turned out that I didn't waste any time in that regard. I joined a completely computational lab without knowing how to write code. Those technical skills did eventually develop, but more importantly so did the resilience to keep pushing and learning new things, even when progress felt slow. None of this would have been possible without the support of so many people and I am incredibly grateful for all of the help along the way.

I want to start by acknowledging my research advisor, Alison Feder. Alison, thank you for your patience on me as a trainee and always pushing me to maintain the standard you knew I was capable of reaching. Your dedication to my success has shined through with your consistently extensive and thoughtful feedback on my work. I'll always be grateful for your role in my development as a scientist. I'm also proud to have been a part of the lab's early growth and I look forward to the unbounded heights that the Feder Research Group will accomplish in the future.

I'm also thankful to the rest of the members of the Feder lab: Elena Romero, Yingnan Gao, Alex Robertson, Iris Jia, Linh Tran, Sam Hart, Allie Kreitman, Yirui Chen, and Dylan Clark for the engaging lab meetings, fun work environment, and Taste of India outings.

To my supervisory committee members: Gavin, Ivana, Neda, Ray, and Rosana, thank you for your insightfulness on my work and the time you invested to help guide me on my path forward. I would also like to thank Ray and Rosana for their contributions and leadership as collaborators on my dissertation projects. I am very grateful for the mentorship you both provided and I always appreciated the curiosity and positive mindset you brought to our meetings.

To the Genome Sciences department, there's so many of you to thank that made these last five years such an enjoyable experience. Brian Giebel, Maureen Larsen, the entirety of GS-IT, your contributions to the students and department are especially appreciated. Cheers to the 2021 graduate student cohort, the IMA basketball pick-up squad, and the many friends in GS that I've made along the way.

Thank you to the Native community here at UW, especially to Cera, Cheyne, and Josh for our lunch meet-ups. To AISES, SACNAS, and Native Forward for encouraging my growth as a scientist through mentorship, financial support, and fostering an extensive professional network. Thank you to my tribe, the Quinault Indian Nation, for supporting my higher education.

I would also like to acknowledge those involved in my scientific pursuits prior to grad school. Thank you Rena, Maria, and Julie, you were the key early mentors that shaped my foundations as a scientist and encouraged me to continue on this path. Professor Terefe and my Seattle Central math professors, thank you for reinvigorating my academic spirit.

Thank you to my family and friends, your support has been more important than I could ever express. To my grandparents, who make the world feel steady, even when I feel like I'm at sea. To my parents, your everburning encouragement has kept my chin up over and over again and I'm grateful for who you've raised me to be. To Guinness and Reptar, thank you for the reminder that sometimes you need to just put on a record and be in the present. And to those who aren't here to step over the finish line with me, just know I've

carried you with me the whole way.

And thanks finally, and always and forever, to my beautiful wife Briyana for putting yourself between me and the world time and again. Without you repeatedly picking me up, I would have long since fallen on my face and stayed there. None of this without you.

Chapter 1

INTRODUCTION

Cancer is fundamentally an evolutionary process, driven by the accumulation of somatic mutations and selection of the cells harboring those mutations. Certain pathogenic germline variants can reshape this evolutionary trajectory starting from birth, resulting in an increased risk of developing cancer. However, how these germline variants alter this evolutionary trajectory toward malignancy and how any altered dynamics can be therapeutically actionable remain difficult challenges. This dissertation examines how germline variants causing cancer predisposition syndromes reshapes somatic evolution and examines ways in which these dynamics can be modified for cancer prevention.

1.1 Cancer progression is an evolutionary process

1.1.1 Somatic mosaicism

Throughout normal aging, individual cells acquire somatic mutations during DNA replication and in response to endogenous and exogenous mechanisms. When a mutation arises in a cell, it is passed on to its progeny during cell division. This process generates a new lineage of cells containing that mutation, referred to here as a somatic clone. As mutations continue to arise over time, new clones are continuously generated within tissues and

the genetic composition of a given tissue becomes heterogeneous. This mixture of clones, termed somatic mosaicism, represents the cumulative outcome of cellular evolution over the span of an individual's life [1–5].

The dynamics of somatic mosaicism are constituted by two critical factors: mutation, which generates genetic variation, and selection, which impacts what happens to that variation over time. The mutations contributing to this mosaicism are diverse in type and can include genetic alterations such as single nucleotide variants (SNVs), small insertions and deletions, copy-number alterations, and larger structural variants. The processes by which these mutations are introduced vary and can leave characteristic signatures within the genome, especially among SNVs [6]. These mutational signatures can be associated with external forces such as ultraviolet radiation, tobacco smoking, therapeutic agents, or intrinsic cellular processes such as aging-related clock-like mutations or oxidative damage [1–3, 6–11].

Once these mutations arise, the clones harboring them may expand, shrink or even disappear over time. In many cases, mutations have little or no impact on the behavior of the cell [12, 13], and the clone sizes fluctuate without consistent directionality due to stochastic drift processes [14–16]. Neutral mutation clones, especially at small clone sizes, are the most susceptible to these chance fluctuations.

In contrast, mutations that change the cell's ability to survive or divide compared to neighboring cells are subject to positive selection, a stronger force affecting clone sizes. If a mutation confers a selective benefit compared to its neighbors, the clone carrying that mutation can increase in size through clonal expansion [17–19]. Clonal expansion in normal tissues has been most extensively studied in the context of mutations in the blood, a process known as clonal hematopoiesis of indeterminate potential (CHIP) [20, 21]. Large-scale population studies have identified somatic mutations in genes commonly mutated in myeloid malignancies in more than 10% of individuals older than 65, but rare in individu-

als younger than 40 [22–24]. Importantly, CHIP is a strong risk factor in the development of subsequent hematologic cancer [22, 23, 25]. These findings established clonal expansions in otherwise normal tissue as a common feature of both human aging and disease risk.

Following the characterization of CHIP, many studies have investigated somatic mosaicism in other non-cancerous contexts. Martincorena et al. characterized the mutational landscape of histologically normal human eyelid skin and found it to contain hundreds of expanded cancer-associated mutant clones within a given 1 cm² skin area of individuals over the age of 55 [1]. Although cancer-associated mutations in histologically normal skin had been observed previously [26, 27], the extent to which the mutations riddled the skin was not expected and reshaped the field’s perspective about early cancer progression. This pattern is also not unique to the harsh sun-exposed conditions of the skin. Similarly, in the normal esophagus of older individuals, 40-80% of the esophageal epithelium was found to be remodeled by mutated clones [2, 5].

However, the persistence and expansion of mutant clones are not entirely dependent on the nature of the mutation alone. Tissue architecture can spatially constrain the ability of clonal expansion, such as in stratified squamous epithelia [28–32] or in the ductal compartments of mammary tissue [33, 34]. Similarly, as multiple clones begin to arise and expand, clones will come into contact and compete with each other for limited space and resources, resulting in further restrained expansion [29, 30, 35]. Further, the exposure to cytotoxic agents such as chemotherapy can not only cause mutations directly, but also impact the competitive landscape of clonal evolution by altering the selective pressures [9, 10, 36–38]. Thus, the combination of mutation, drift, selection, and tissue architecture are among some of the key processes contributing to the types of somatic mutations we observe in somatic tissues.

1.1.2 *Pre-malignant clonal expansion*

The recent efforts to characterize this somatic variation across multiple human tissues have revealed that not only do many histologically normal tissues contain extensive repopulation by mutated cells, but much of the underlying clonal expansions occur from genes recurrently altered in cancer [1–3, 11, 15, 39]. This observation suggests that the early stages of oncogenic evolution occur long before overt malignancy and appear to develop from a background of pre-existing mutant lineages.

Although the extent to which these pre-cancerous mutations populate normal tissues was not appreciated until recently, the foundation of this idea dates back several decades. Field cancerization, first proposed by Slaughter et al., found histologically abnormal, but non-cancerous tissue frequently surrounded tumors in oral epithelia [40]. This suggested that cancer did not arise from a single isolated transformational event, but within a large ‘field’ of cells biologically primed for malignant transformation. Modern genomic detection has refined this concept by demonstrating that these fields correspond to the somatic clonal expansions of pathogenic driver mutations [41, 42]. These clonal expansions increase the population of cells carrying cancer-associated mutations, and thereby increase the probability that subsequent genetic and/or epigenetic alterations will accumulate on that lineage. This stepwise progression of accumulating mutations is the evolutionary basis for the earliest stages of cancer progression [43–45].

However, the presence of clonally expanded driver mutations do not deterministically result in malignant tumor formation, where clones can persist and expand for years without any histologically abnormal neoplastic growth [1, 2, 26, 46–48]. Additionally, not all mutations that undergo pervasive clonal expansion in normal tissues confer an increased risk of malignant transformation. *NOTCH1* mutations have strong selective advantages in the context of squamous epithelium and are frequently observed in normal esophageal tissue, yet are observed less commonly in esophageal cancers [2, 47]. As such, expanding

NOTCH1 clones may even exhibit protective effects against the emergence of malignant clones and the progression of esophageal tumors [35, 47]. The observation that not all clonally expanded mutations uniformly result in malignancy raises a critical question of how certain lineages acquire the additional alterations necessary for malignant progression.

1.1.3 Evolutionary trajectories toward cancer progression

An early key foundation to understanding how step-wise mutational events lead to cancer progression is from Knudson’s two-hit hypothesis [49]. Knudson proposed that tumor suppressor genes typically require two “hits” to fully inactivate. This helped to explain the occurrence of both inherited and sporadically occurring forms of retinoblastoma. This framework allowed for an individual to inherit a mutation in a cancer-associated gene, but would still require a second-hit prior to malignant transformation and would still be able to explain sporadic cases, where both hits occur somatically.

This initial stepwise framework has been further characterized in the context of colorectal cancer progression, which was originally described to follow a reproducible sequence of mutations involving genes such as *APC*, *KRAS*, and *TP53* [44, 45, 50]. Fearon and Vogelstein associated additional genetic mutations with incrementally reshaped cell behavior leading to progressively more malignant phenotypes [44]. Although colorectal cancer progression does not deterministically follow the same genetic path to malignant progression, and other cancer types can also occur through various evolutionary trajectories, there are biological mechanisms that are commonly altered among diverse cancer types.

Across developing cancers, mutations frequently occur in three broad functional gene classes that include DNA repair genes, tumor suppressors, and oncogenes [51]. Mutations affecting DNA repair genes, such as those in the FA/BRCA pathways, are responsible for the repair of DNA intrastrand cross-links and homologous recombination repair, where these hindered repair mechanisms lead to increased genomic instability [52–54]. Muta-

tions affecting tumor suppressors impair cellular safeguards against malignant transformation, including cell-cycle arrest, senescence, and apoptosis. A notable example is *TP53*, the most widely mutated gene across all human cancers [55, 56]. This allows cells to proliferate in the presence of DNA damage, when the cell would otherwise halt this process. Conversely, oncogenes, such as mutated KRAS, promote proliferative signaling directly, leading to uncontrolled growth [57].

Through the gradual accumulation of these and other genomic alterations, evolving cell populations begin to develop shared functional behaviors that define the earliest stages of cancer progression. Commonly defined functional capabilities on the way to cancer progression include: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing/accessing vasculature, activating invasion and metastasis, reprogramming cellular metabolism, and avoiding immune destruction, commonly known as the "hallmarks of cancer" [58].

1.2 Inherited driver mutations shorten the path to tumorigenesis

1.2.1 Cancer predisposition syndromes

The stepwise evolutionary progression of cancer can be accelerated when a mutation in a driver gene is inherited. Rather than requiring the initial acquisition of a mutation in a cell somatically, inheriting a variant results in all cells of the individual containing this variant and thus put all cells on the first step toward cancer progression at the beginning of the individual's life.

Since inheriting a pathogenic variant results in one key pathway already compromised, fewer secondary events may be necessary for tumorigenic transformation and it can predispose multiple separate locations within the body to concurrently begin cancer progression [59, 60]. This hereditary cancer accelerant, or cancer predisposition syndrome, in-

creases the risk of cancer development and developing multiple primary tumors, reduces the age of onset, and is responsible for approximately 5-10% of all cancers [61].

Some of the most well characterized examples of cancer predisposition syndromes include hereditary breast and ovarian cancer syndrome caused by germline mutations in *BRCA1* or *BRCA2* and Lynch syndrome caused by mutations in DNA mismatch repair genes [62–66]. Other notable high cancer penetrance syndromes such as Li-Fraumeni syndrome, caused by mutations in *TP53*, and Fanconi Anemia caused by mutations in the FA/BRCA pathway are associated with a wide array of cancer types and especially early onset [67–72]. Also, due to the functional importance of the pathways altered in these syndromes, there are often additional clinical manifestations that include but are not limited to cancer risk. This can include congenital abnormalities [72, 73] or bone marrow failure [71, 72, 74]. Importantly, it can also impact treatment decisions [75–77]. For example, chemotherapy and radiation may induce additional mutational burden that can impact an individual’s ability to tolerate standard of care therapy or result in increased risks of subsequent primary tumors [9, 10, 36, 37, 75, 76]. Moreover, different mutation types within the same gene can result in a variety of phenotypic severity, such as cancer penetrance or age of onset [68, 70, 78, 79].

The unique features associated with cancer predisposition syndromes mean they serve as a valuable resource and model for studying the earliest stages of cancer progression. In the general population, tracking mutational accumulation on the evolutionary path toward cancer requires identifying relatively small cell populations of initial clonal expansions and attempting to observe subsequent mutations within those. Whereas in cancer predisposition syndromes, starting all cells at the same controlled first mutation, allows for a unique snapshot into how early progression may vary within these clones. However, the relatively rare number of individuals with these syndromes in the population make these comparisons difficult and is currently an open field of research.

1.3 Open questions in the field of cancer predisposition syndromes

Despite decades of characterization of many cancer predisposition syndromes, there are many open questions regarding how the underlying germline causal variant of these syndromes affects somatic evolution and how to best prevent cancer progression.

It is unclear how specific germline variants may shift the mutation rate and selection pressures within otherwise normal tissues and how they may affect aging related processes such as CHIP. Additionally, it is important to understand how this somatic evolution differs across tissue specific environments, as certain tissues are more susceptible to cancer development in certain cancer predisposition syndromes, but follow different risks in others.

It is also unclear the extent to which chemo- and radiotherapies may affect mutagenesis or selection within individuals with cancer predisposition syndromes. In Li-Fraumeni syndrome cancer therapies can both aid in the elimination of present tumors, but also increase risk of subsequent therapy-related tumors, especially in the formation of acute myeloid leukemia [76]. In individuals without a germline *TP53* variant, *TP53* clonal spread under the selective pressure of cytotoxic therapy is the primary driver of therapy-related acute myeloid leukemia (t-AML) [9, 36–38, 80]. In the context of LFS, where one copy of *TP53* is already partially or fully inactivated, it is unclear if this selective pressure similarly acts on second-hit *TP53* mutations, or if other mechanisms drive t-AML in LFS, such as clonal expansion of additional mutated driver genes.

Similarly, because *TP53* is a tumor suppressor, second-hit inactivation is typically required for tumor progression, but it is unclear if selection favors specific types of second-hits over others. For example, LOH events such as full gene or chr17p deletions are the most common second-hit in LFS tumors [81–83] and that likely suggests a differential selective pressure compared to second-hit point mutations, even if those correspond to highly

disrupted proteins.

1.4 Technologies to understand clonal evolution in the context of cancer predisposition syndromes

Studying clonal evolution in the context of cancer predisposition syndromes is challenging because early somatic events occur at very low variant allele frequencies and long before clinical manifestations of disease. Additionally, the rarity of affected individuals, combined with the clinical invasiveness, expense, and technical difficulty of normal tissue analysis has resulted in relatively little investigation in this area (although see [81, 84, 85]). To progress, experimental and computational methods are required that can detect rare somatic mutations in normal tissues and that can provide a quantitative framework to understand them at scale.

1.4.1 In silico modeling approaches

The direct experimentation of somatic evolution in humans and model organisms requires high-depth sampling across normal tissues. The long timescales, high cost, and invasive sampling can be logistically infeasible. The rarity of individuals with predisposition syndromes compounds the difficulty to understand clonal evolution in this context. As a result, computational models have provided key mechanistic understanding of clonal evolution and early tumor progression.

Computational models provide a way to explore otherwise inaccessible clonal evolutionary dynamics through simulation across many cellular generations. These models can allow for estimation of important parameters central to somatic evolution, such as the selective advantage conferred by driver mutations [86]. Such parameters would otherwise be difficult to directly measure as selection occurs over many cellular generations and would require longitudinal sampling in a given tissue site. For example, Beerenwinkel et al. ap-

plied a Wright-Fisher population genetic approach to find the selective advantage of driver mutations, which modeled stochastic mutational accumulation of driver mutations in a growing colonic adenoma [86]. This approach showed that many genes containing small selective advantages ($\sim 1\%$) produced waiting times to malignancy that aligned with the decadal timescales in colorectal cancer progression [86].

Beyond estimating evolutionary parameters, computational models can provide a means to understand the timing of mutational events that shape clonal outcomes. Modeling the emergence and expansion of revertant patches in epidermolysis bullosa, an inherited skin disease, revealed that the high frequency of observed revertant patches can be explained by the revertant mutations arising after development, but having a sufficient advantage compared to the background cells to overtake areas of the skin cell population [87].

However, many of these models assume that cells act in a well-mixed population and do not account for spatial structure of tissues, where the physical organization of cells may impact their growth. Spatial modeling techniques such as hybrid agent-based models can combine many important aspects to more accurately understand the dynamics of cellular evolution in a given tissue environment [88]. Spatial modeling in epithelial tissues have shown that clonal expansion is largely shaped by interactions with neighboring cells and are constrained by tissue structure and homeostatic processes [29, 30, 35, 88–90]. As a result, even heavily favored clones may have limitations to their expansion.

1.4.2 Sequencing-based approaches

Direct detection of early somatic evolution requires technologies capable of detecting variants that are present in only a fraction of cells within a sample. Next-generation sequencing (NGS) is capable of detecting variants within heterogeneous mixtures of cells and has been paramount to understanding the later stages of cancer progression, inherited mu-

tations, and somatic mosaicism in aging individuals [1, 2, 51, 56]. Large-scale tumor sequencing efforts have catalogued cancer-specific driver mutations and in conjunction with bioinformatic tools have enabled the analysis of mutation and selection processes acting within tumors [91, 92].

However, NGS suffers from errors introduced throughout the preparation and sequencing of DNA fragments. These errors occur at rates of $\sim 1\%$ and obscure the ability to detect mutations at frequencies below these error rates as true mutations become indistinguishable from the errors [93, 94]. To fully understand the extent of somatic evolution happening far before tumor formation, these low frequency mutations need to be reliably detected as early arising mutations start in a single cell and thus begin at low variant allele frequencies.

Consequently, error-corrected sequencing methods have been developed to improve variant detection at low frequencies [46, 93, 94]. Duplex sequencing reduces standard NGS errors by tagging both strands of a DNA molecule, which allows a consensus sequencing read to be constructed using information from both strands [94]. This enables the detection of true mutations that are present in complement on both strands of the DNA molecule, whereas a mutation present on only one strand can be subsequently deemed an error. By accounting for this complementary aspect of the DNA molecule, error rates can be reduced to less than 1 error per 10^9 nucleotides sequenced. This substantial reduction in error rate justifies sequencing at much higher depths ($>10,000\times$) and allows low frequency mutations to be confidently distinguished from sequencing errors. In particular, duplex sequencing methods have found widespread *TP53* mutations at frequencies undetectable by standard NGS approaches across normal tissues and with age-associated increases [46, 95, 96]. These approaches provide direct observation into the earliest stages of somatic evolution, making them a critical tool to understand early tumorigenic events in the general public and within cancer predisposition syndromes. They may also prove

useful as a future framework to understand how therapeutic interventions alter these early events.

1.5 Preventative therapeutic approaches for cancer predisposition syndromes

1.5.1 Natural reversion and therapeutic strategies

A remarkable parallel to therapeutic cancer prevention within cancer predisposition syndromes is the somatic and spontaneous correction of an inherited pathogenic variant that can occur naturally in some cells. This phenomenon called revertant mosaicism, or ‘natural gene therapy’ has been observed in Lesch-Nyhan syndrome, Fanconi Anemia, genodermatoses such as epidermolysis bullosa, and other inherited disorders [87, 97–108]. Revertant mosaicism is often associated with less severe disease phenotypes and reduced subsequent cancer risks.

Since inherited cancer predisposition syndromes are often monogenic disorders with a majority of causal variants being single nucleotide substitutions [68, 109], various molecular mechanisms can result in revertant mosaicism. For example, in Fanconi Anemia, reversion can occur via back mutation, gene conversion, intragenic crossover, and second-site mutation [101–104]. The incidence of revertant mosaicism within inherited disorders can be substantial with up to 25% of Fanconi Anemia patients and 35% of patients with epidermolysis bullosa, implicating a potential selective force for the repeated emergence and persistence of these mutations [108, 110]. Notably, revertant cells found in the hematologic compartment of individuals with FA are capable of conferring a proliferative advantage to repopulate blood cell lineages and reduce disease risks [110, 111]. An early clinical piece of evidence to support this was found in monozygotic twins with FA, where a revertant HPSC in one twin repopulated the hematopoietic compartments of both siblings, both containing the same revertant mutation [111]. A revertant proliferative advantage is not

restricted to just blood cells, where the open nature of the tissue may allow for cellular expansion with less architectural limitations than other tissues. For example, in the spatially structured epidermis, revertant mutations in epidermolysis bullosa not only arise, but expand to form phenotypically normal epidermal patches that can reach sizes large enough to be visualized clinically and improve patient well-being [87, 107].

The promise of natural gene therapy has inspired progress toward therapeutic-induced reversion. The monogenicity associated within inherited disorders makes for an attractive target for gene therapy and the conferred proliferative advantage of revertant cells may thus amplify the corrective potential and not require immediate reversion of all cells to be efficacious. Excitingly, the first successful engraftment of gene corrected hematopoietic stem cells into an FA patient managed to show the proliferative advantage of cells, where lineages expanded over time via the restoration of DNA damage repair, as well as a halt in bone marrow failure progression [112–116]. The complete regeneration of the epidermis in a patient with junctional epidermolysis bullosa has further shown the therapeutic potential of gene correcting cells to treat inherited disorders, where the majority of tissue was grafted to the patient and resulted in stable skin that maintained over long-term follow-up [117]. These *ex vivo* approaches, where the patient’s own cells are corrected and reintroduced, are paving an exciting path for *in vivo* delivery strategies that are poised to reshape the future of gene therapy to a wider range of tissue sites and inherited disorders.

1.6 Thesis motivation

In this thesis, I tackle how cancer predisposition syndromes reshape somatic mutation accumulation and clonal selection prior to malignancy using both spatial computational modeling and sequencing-based approaches. I focus primarily on two complementary questions: 1. How do germline driver mutations alter somatic mutation and clonal selection

in normal tissues? 2. Can we leverage tissue clonal dynamics to maximize cancer prevention in normal tissue? In answer to the first, I analyze multi-tissue duplex sequencing data from a Li-Fraumeni syndrome family to define how an inherited *TP53* variant alters somatic evolution in normal tissues. In answer to the second, I develop a quantitative agent-based modeling framework to evaluate how gene correction strategies can reshape the competitive dynamics within epithelial tissue and suppress cancer-associated mutational accumulation in Fanconi Anemia oral mucosa.

1.7 Previous publication and co-authorship of dissertation content

This dissertation incorporates text and materials from the following previously published and pre-print manuscripts [118, 119], with contributions from all authors:

- Hunter L. Colegrove, Marianne E. Dubard-Gault, Henry Marshall, Brendan F. Kohn, Thomas H. Smith, Zachary K. Norgaard, Fang Yin Lo, Elizabeth K. Schmidt, Jacob E. Higgins, Charles C. Valentine, Desiree A. Marshall, John I. Clark, Eric Q. Konnick, Jesse J. Salk, Marshall S. Horwitz, Raheleh Rahbari, Alison F. Feder, and Rosa Ana Risques. Ultra-deep duplex sequencing reveals unique features of somatic evolution in the normal tissues of a family with Li-Fraumeni syndrome. *bioRxiv*, January 2026. doi: 10.64898/2026.01.12.699071
- Hunter L. Colegrove, Raymond J. Monnat Jr, and Alison F. Feder. Epithelial competition determines gene therapy potential to suppress Fanconi Anemia oral cancer risk. *PLOS Computational Biology*, 21(9):e1012915, September 2025. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1012915

Chapter 2

ULTRA-DEEP DUPLEX SEQUENCING REVEALS UNIQUE FEATURES OF SOMATIC EVOLUTION IN THE NORMAL TISSUES OF A FAMILY WITH LI-FRAUMENI SYNDROME

2.1 Introduction

Li-Fraumeni syndrome (LFS) is a cancer predisposition syndrome caused by an inherited or de novo germline pathogenic variant in *TP53*. Individuals with LFS are at risk of developing multiple cancer types starting early in life, with an up to 24 times higher incidence of cancer than the general population [67, 69, 78]. The increased cancer risk is due to the higher likelihood of completely inactivating *TP53* via a second-hit mutation or loss of heterozygosity, given that the first copy is already mutated in all the cells in the organism [56]. Despite the high cancer prevalence, relatively little is known about the early somatic evolutionary processes that predate and ultimately lead to LFS carcinogenesis, partly because LFS is rare and partly because most sequencing efforts profile cancers rather than normal tissue [81, 120]. Recent deep-sequencing studies capable of detecting cellular clones composed of only a few cells have revealed new insights into the mutational and selective forces operating across diverse healthy tissues in the general population, shedding light on how somatic variation arises and can later drive cancer development [7, 15, 46, 121]. Sim-

ilar efforts to quantify somatic evolutionary processes in individuals with LFS could lead to a better understanding of their risk of cancer development and inform diagnostic and monitoring strategies.

Deep sequencing of normal tissue in individuals with LFS offers a unique approach to understand how a single altered copy of *TP53* affects somatic evolution across tissues. Somatic *TP53* mutations are a common first hit across numerous cancer types in the general population [51], and they may accelerate mutagenesis or clonal selection elsewhere in the genome. Because all cells in individuals with LFS carry a germline *TP53* variant, studying somatic evolution in these individuals allows a comprehensive examination of how second-hit mutations arise and expand. Further, when *TP53* pathogenic variants are present in the germline, carrier cells exist in all anatomical locations, providing an opportunity to understand how subsequent somatic evolutionary processes differ across tissue contexts and predispose LFS individuals to specific cancer types [67].

However, the detection of somatic mutations in normal tissue is challenging because those mutations are rare and escape detection by conventional sequencing methods. Duplex sequencing is an error correction sequencing method that enables ultradeep sequencing ($> 10,000\times$) and unprecedented resolution for mutation detection [94]. Here, we use duplex sequencing of multiple genomic regions in blood DNA from a family of individuals carrying the *TP53* germline pathogenic variant c.542G>A (p.Arg181His or p.R181H hereafter) and a cohort of unaffected individuals to identify differential patterns of somatic mutation and selection in cells with and without *TP53* p.R181H. Among inherited *TP53* pathogenic variants, p.R181H is notable for its relatively high prevalence, partial loss-of-function phenotype, reduced cancer penetrance, and later cancer onset compared to other LFS-associated pathogenic variants [70, 79, 122, 123]. Despite attenuated effects, individuals with germline partial loss-of-function *TP53* variants such as p.R181H maintain substantially increased cancer risk compared to the general population, developing cancer at

younger ages, especially breast cancer [124]. The p.R181H variant has also been associated with cancers outside the main LFS spectrum, including prostate, colorectal, and thyroid [125]. In the LFS family studied here, the proband developed esophageal cancer at age 34 and succumbed to metastatic disease. Multi-tissue collection at autopsy and blood samples procured from family members and healthy donors provided a valuable opportunity to investigate mutagenesis and clonal expansions in the context of the *TP53* p.R181H germline variant.

2.2 Results

2.2.1 Ultra-deep sequencing of a family carrying p.R181H and healthy controls

The study included an individual with LFS who carried the germline *TP53* p.R181H pathogenic variant and died of metastatic esophageal carcinoma at age 34, as well as members of his family. Ultradeep DNA duplex sequencing of autopsy-collected tissues provided a unique opportunity to investigate how the germline p.R181H status influences somatic evolution and clonal dynamics in blood and solid tissues. Our cohort included the proband (LFS01), two family members carrying the same pathogenic variant and not known to have cancer (LFS02, LFS03), one relative without the pathogenic variant (REL01), and seven additional unrelated healthy controls of variable age between 25 and 76 ((CON01-CON07), Figure 2.1A, Supplemental Table S1).

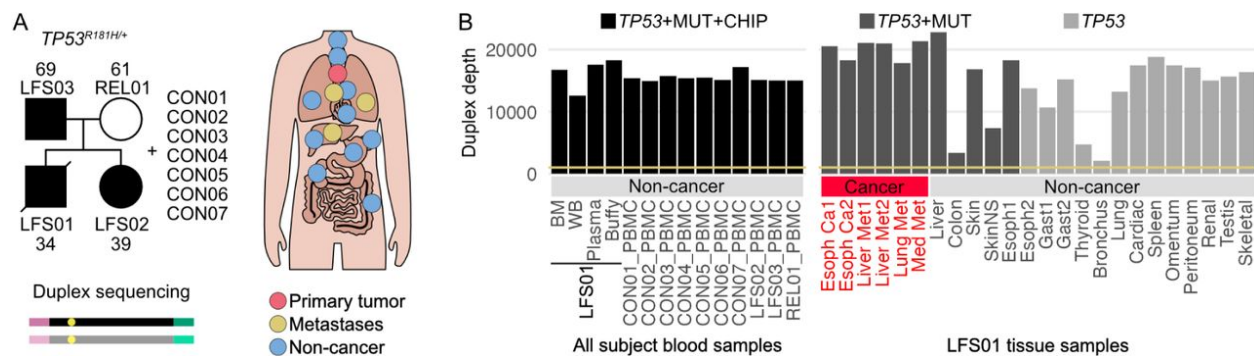


Figure 2.1: Ultra-deep sequencing of somatic mutations in subjects with and without Li-Fraumeni Syndrome.

Figure 2.1: (A) The study included a Li-Fraumeni patient (LFS01, proband), his family (REL01, LFS01-02), and 7 unrelated controls (CON01-07). The pedigree indicates germline *TP53* p.R181H pathogenic variant carriers and ages at sampling. Blood from all individuals, along with primary esophageal (red), metastatic (yellow), and non-cancer (blue) tissues collected from the proband at autopsy, were analyzed with ultra-deep duplex sequencing, which uses double stranded molecular tags that enable mutation calling in both strands of DNA independently. *Tissues not shown in the diagram include skeletal muscle (right pectoralis), sun-exposed skin (left shin), non-sun-exposed skin (left buttock), omentum, peritoneum, and testis. **(B)** Duplex sequencing was performed with 3 different panel combinations depending on sample: *TP53* panel only, *TP53* panel with the mutagenesis panel (MUT), or *TP53* panel combined with both the MUT and the AML-29 Assay (CHIP) panel. Duplex depth (calculated as the mean number of duplex reads per position) is shown for each sample. For some tissues, two samples (labeled 1 or 2) were analyzed. Yellow line represents 1000x duplex depth sample threshold. BM: bone marrow; WB: whole blood; PBMC: peripheral blood mononuclear cells; Esoph: esophagus; Ca: cancer; Met: metastasis; Med: mediastinal; SkinNS: non-sun-exposed skin; Gast: gastric tissue.

Hematopoietic samples were collected from all study participants: buffy coat, plasma, whole blood, and bone marrow from LFS01 and peripheral blood mononuclear cells (PBMCs) from all other subjects (Figure 2.1B, Supplemental Table S2). Low frequency mutations were detected on all hematopoietic samples using ultradeep duplex sequencing with two targeted panels: 1) TwinStrand's AML-29 Assay [126], a panel of 29 genes recurrently mutated in acute myeloid leukemia (AML) and clonal hematopoiesis of indeterminate potential (CHIP, 59 kb footprint, Supplemental Table S3), referred to as CHIP genes hereafter;

and 2) Twinstrand’s Human Mutagenesis Assay, which contains 20 genomic regions selected to be an unbiased footprint of human genomic sequence contexts, avoiding repeats and homopolymers, and with no known role in human cancer (48 kb footprint). The mean duplex consensus sequencing depth (duplex depth) for hematopoietic samples was 15,683X (12,606X - 18,282X, Figure 2.1B). Most targeted genes and regions showed consistently high depth across all samples (Supplemental Figure A.1A-B, Supplemental Table S4). The extremely high depth of sequencing with these target panels enabled us to characterize how the p.R181H variant impacts clonal evolution and mutagenesis in the blood. Unless otherwise specified, all reported sequencing depths and mutation frequencies refer to duplex consensus sequences.

We also conducted extensive multi-tissue sampling in LFS01 (Figure 2.1A). Two primary esophageal tumor samples and four metastatic samples across three metastatic sites (liver, lung, and mediastinal) were collected. In addition, an extensive range of 18 non-cancerous samples spanning 16 tissues were collected (Supplemental Table S2). Duplex sequencing of *TP53* was performed in all samples, and a subset of 11 tissues (all tumor and metastases plus five non-cancer samples: esophagus, liver, colon, sun-exposed skin, and non-sun-exposed skin) were sequenced with the mutagenesis panel. The mean duplex depth was 15,264X. Four tissue samples had reduced coverage with depths between 2,118-7,376X (thyroid, bronchus, colon, and non-sun-exposed skin, Figure 2.1B, Supplemental Table S4), which was due to DNA loss during library preparation. Across all individuals, tissue types, and sequencing panels, we identified 1,992 somatic coding mutations in CHIP genes including *TP53* and 6,487 somatic mutations in the mutagenesis panel (Supplemental Tables S5-S7). Sequencing also confirmed that all tissues from affected individuals were heterozygous for *TP53* R181H.

2.2.2 Germline p.R181H does not affect most CHIP clonal expansion but is associated with increased mutation frequency in the mutagenesis panel and reduced positive selection in TP53

We first examined if the germline p.R181H variant altered the number of mutations acquired in CHIP genes and the degree to which cells carrying those mutations could clonally expand. To do so, we analyzed duplex sequencing of 29 genes associated with CHIP and/or AML in hematopoietic samples, using buffy coat from LFS01 as the closest comparable sample to the PBMCs analyzed from all other subjects. The number of mutations in the coding regions of CHIP genes varied across individuals, with more mutations found in samples from older individuals and in the blood of LFS01 (Figure 2.2A). However, this excess of mutations could be due to multiple confounders, including higher sequencing depth and exposure to chemotherapy. To control for differences in sequencing depth, for each sample we calculated CHIP Mutation Frequency (MF) as the number of coding mutations identified in CHIP genes divided by the total number of nucleotides sequenced in the coding regions of those genes. CHIP MF significantly increased with age (Figure 2.2B). The association with age was also significant for mutations in *DNMT3A* and *TET2* (Supplemental Figure A.2), which are some of the most common CHIP genes in the population [127]. Multivariate analysis demonstrated that CHIP MF was associated with age and chemotherapy but not with the germline p.R181H variant (Figure 2.2C, Supplemental Table S8). To determine if the p.R181H variant affected not just the number of mutant clones but also their degree of clonal expansion, we quantified CHIP Mutation Burden (MB), which takes into consideration the number of mutations and the clone size. CHIP MB was calculated as the total number of duplex reads with coding mutations in any CHIP gene divided by the total number of coding nucleotides sequenced. CHIP MB was significantly associated with age in univariate and multivariate models but it was not associated with chemotherapy or the p.R181H variant (Supplemental Figure A.3, Supple-

mental Table S8).

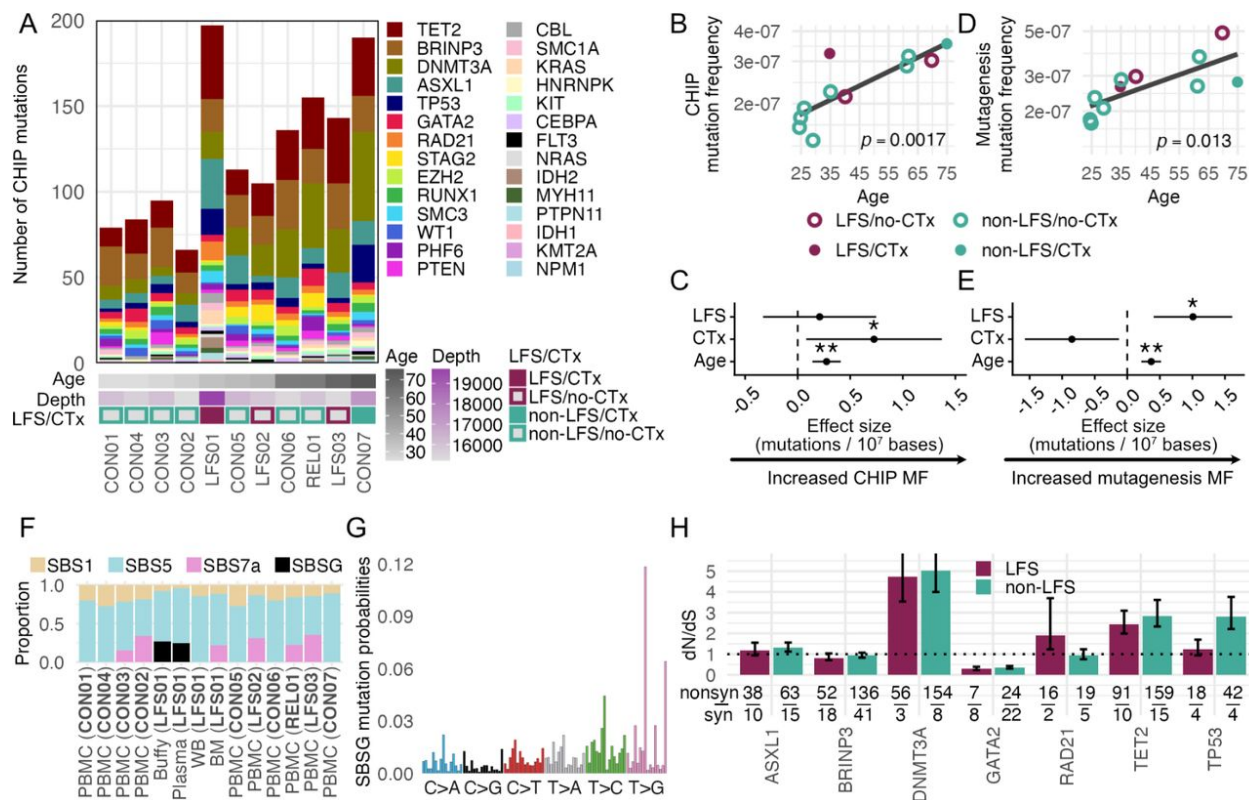


Figure 2.2: Ultra-deep characterization of CHIP mutation and selection in the blood of individuals with and without Li-Fraumeni Syndrome.

Figure 2.2: (A) Number of coding mutations in CHIP and/or AML-associated genes by individual. Heatmap indicates age and average duplex sequencing depth across CHIP panel genes. Li-Fraumeni Syndrome (LFS) and chemotherapy history (CTx) are indicated by the color of the squares. (B,D) Linear regressions of coding CHIP mutation frequency (MF) (B) and mutagenesis panel MF (D) with age. Circles are colored by LFS status and filled circles indicate CTx history. Age-associated baseline regression is fit using all individuals and linear regression p-value is indicated. (C,E) Multiple regressions of coding CHIP MF (C) and mutagenesis panel MF (E) using age (scaled by decade), LFS, and CTx status as covariates. MF was scaled per 10,000,000 bases. Asterisks indicate significance value (* $p < 0.05$, ** $p < 0.01$) and error bars indicate 95% confidence intervals. (F) Proportion of mutations associated with a mutational signature using mSigHDP by subject and tissue. (G) Mutational spectra of SBSG. The full 96 trinucleotide substitution spectra are shown, but only the six base substitution classes are labeled. (H) dN/dS grouped by gene and germline *TP53* p.R181H status. Error bars indicate the interquartile range of dN/dS values obtained through bootstrap resampling. Error bars extend beyond the plotted range for *DNMT3A*. Fraction indicates the number of observed non-synonymous (top) and synonymous (bottom) mutations used in calculation. Dotted line indicates absence of selection (dN/dS=1).

In addition to CHIP genes, hematopoietic samples were also sequenced with a mutagenesis panel covering twenty 2.4kb targets avoiding genomic regions under strong selection to serve as a genome-representative cross section of mutational contexts [128]. Somatic mutations identified in those regions were used to calculate a MF (see Methods) called “mutagenesis MF” to distinguish it from CHIP MF. Mutagenesis MF was significantly associated with age in univariate and multivariate models and, interestingly, it was also associated with the germline p.R181H variant but not with chemotherapy (Figure 2.2D-E, Supplemental Table S8). Overall, these results indicate that the germline p.R181H variant does not globally elevate the number or size of CHIP clones but it is associated with more mutations across the genome (Figure 2.2E).

Given that the p.R181H variant status was associated with elevated mutagenesis MF, we considered that the presence or intensity of active mutational signatures may differ based on carrier status. We performed de novo single base signature extraction using Mutational Signature Hierarchical Dirichlet Process (msigHDP), a Bayesian model employed

for de novo extraction of mutational signatures, on the mutations identified with the CHIP and mutagenesis panels (Figure 2.2F). Additional hematopoietic samples were analyzed for mutational signatures including plasma, whole blood, and bone marrow from LFS01. Extracted signatures resembled COSMIC signatures SBS1, SBS5, SBS7a and one signature (reported here as SBSG, Figure 2.2G) previously linked to 5-fluorouracil chemotherapeutic exposure [10]. Clocklike signatures SBS1 and SBS5 were detected in all samples, with similar relative contribution in p.R181H carriers vs non-carriers (mean SBS1 and SBS5 proportion in carriers are 10.5% and 63.8% vs 19.5% and 71.6% in non-carriers, respectively). SBS7a was discovered in five individuals. SBS7a is caused by exposure to ultraviolet light and potentially reflects skin resident T cells recirculating into the blood [11]. Two samples from LFS01 (buffy coat and plasma) showed evidence of chemotherapeutic signature SBSG, but no other p.R181H carrier-specific mutational signatures were detected across the affected individuals. However, we further tested for differences between the underlying mutational profiles found in p.R181H carriers and non-p.R181H carriers using the randomization test AMSD and found a small but significant difference between signatures based on carrier status (cosine similarity of 0.919, $p = 0.032$, [129]), potentially driven by the presence of SBSG in LFS01 or the absence of SBS7a. We therefore conclude that while the p.R181H variant may be associated with minor alterations in mutational processes, no single mutational signature clearly explains the difference.

Next, we examined if the p.R181H variant might affect positive selection in specific CHIP-associated genes, even if it did not appear to associate significantly with clonal expansions across the panel as a whole. For genes with at least 10 mutations in p.R181H carriers and non-carriers, we computed dN/dS in both groups to look for evidence of carrier status-specific positive selection. dN/dS revealed positive selection in *DNMT3A* and *TET2* and negative selection in *GATA2*, but these findings were comparable between p.R181H carriers and non-carriers. Two genes, however, had non-overlapping bootstrapped interquar-

tile ranges between the two groups: *RAD21* and *TP53* (Figure 2.2H). *RAD21* showed more evidence for positive selection among p.R181H carriers whereas *TP53* showed more evidence for positive selection among p.R181H non-carriers.

Nevertheless, pathogenicity analyses did not show significant differences in pathogenicity scores or percentage of pathogenic mutations for each gene between p.R181H carriers and non-carriers (Supplemental Figure A.4). The gene with the most pathogenic mutations was *DNMT3A* and the gene with the least pathogenic mutations was *GATA2*, supporting the dN/dS-based evidence of strong, but p.R181H independent, positive and negative selection in these genes, respectively (Figure 2.2H). While positive selection for *DNMT3A* is well-recognized [127], the negative selection of *GATA2* has not been observed in population-based somatic sequencing and highlights a striking divergence in selective pressures across genes operating within healthy tissues. Nevertheless, the selection dynamics on CHIP genes in the blood is mostly unchanged by germline p.R181H status with the exception of *RAD21* mutations, which are suggestive of positive selection in p.R181H carriers, and *TP53* mutations, which appear to have reduced positive selection in p.R181H carriers.

2.2.3 Chemotherapy increases TP53 clonal expansions in the blood of a germline p.R181H carrier and non-carrier

Given the central role of *TP53* in LFS and the observed differences in *TP53* dN/dS between LFS and non-LFS groups, we further characterized the individual somatic *TP53* mutations identified in blood within this cohort.

Each individual carried between 2 and 20 *TP53* coding mutations in blood, with the highest counts observed in the two individuals with previous chemotherapy exposure (LFS01 and CON07, Figure 2.3A). Most mutations appeared to be pathogenic based on mutation type and the AlphaMissense algorithm (see Methods, Figure 2.3A) and they clustered in

the DNA binding domain of the protein (75% compared to 46% expected, binomial test, $p = 10^{-5}$ Figure 2.3B-C). The observed missense mutations, especially those seen on more than one duplex read, were enriched for higher pathogenicity scores compared to the distribution of scores for all possible *TP53* single-nucleotide substitutions (Figure 2.3D). These results suggest that the *TP53* mutations observed in the blood mostly correspond to positively selected pathogenic somatic mutations.

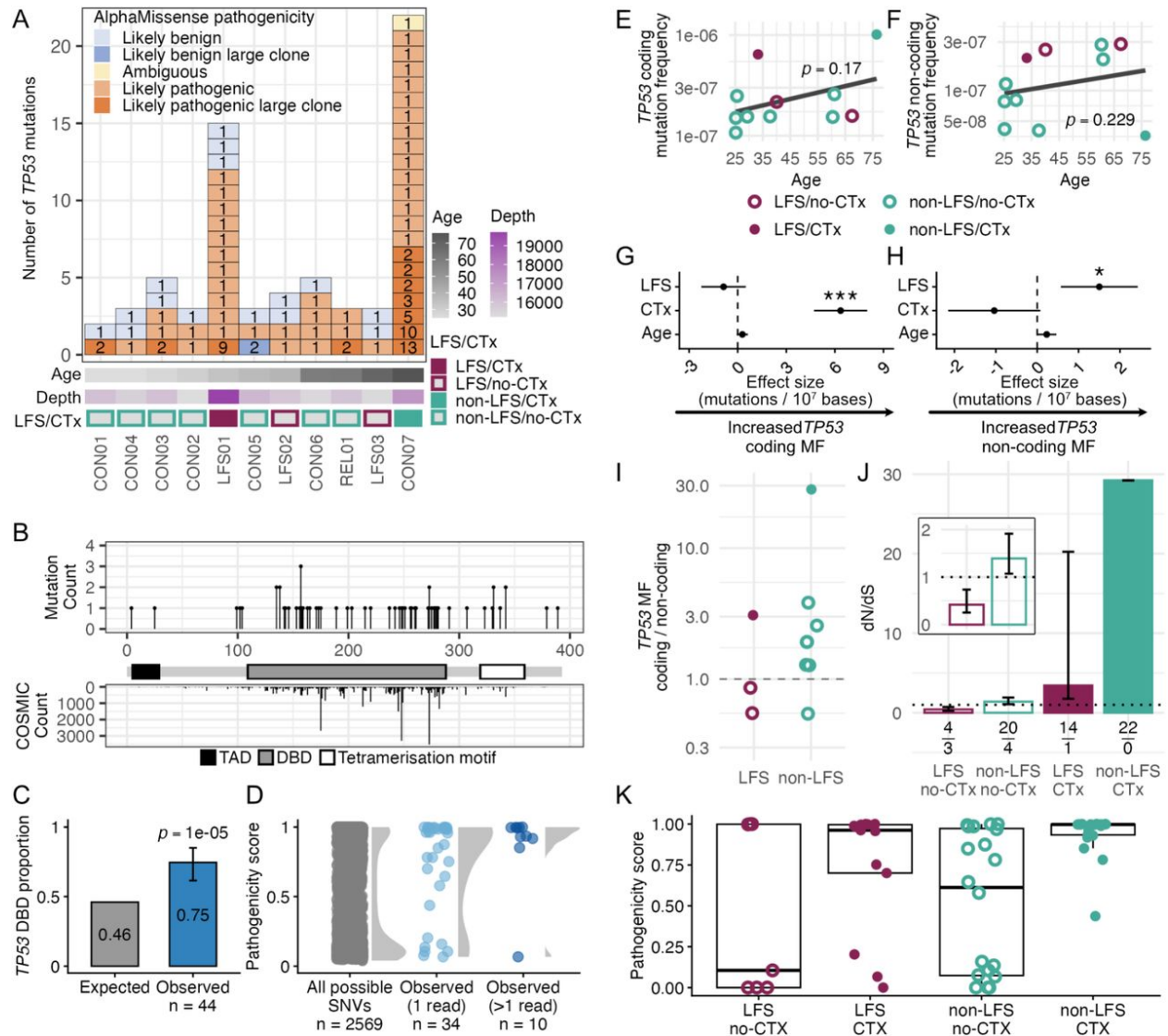


Figure 2.3: Ultra-deep characterization of *TP53* mutation and selection in the blood of individuals with and without Li-Fraumeni Syndrome.

Figure 2.3: **(A)** Number, size, and pathogenicity of coding *TP53* mutations by individual. Each box represents a unique *TP53* mutation within an individual. Box color indicates pathogenicity based on mutation type and AlphaMissense scores (see Methods). The number inside of each box represents the number of duplex reads supporting the mutation. Mutations seen in more than 1 duplex read are considered large clones. Heatmap indicates age and average duplex sequencing depth of *TP53*. Li-Fraumeni Syndrome (LFS) and chemotherapy history (CTx) are indicated by the color of the squares. **(B)** Single nucleotide variant (SNV) counts within the *TP53* coding region across all individuals (top) and mutation counts in the COSMIC database (bottom) by codon position. Major functional regions shown: topologically associating domain (TAD), DNA-binding domain (DBD), and tetramerization motif. **(C)** Expected and observed SNV proportion within the DNA-binding domain of *TP53*. Expected proportion corresponds to the DNA-binding domain's base pair proportion of the total *TP53* coding region. Binomial test p-value and error bars representing 95% confidence interval are shown. **(D)** AlphaMissense pathogenicity scores of all possible *TP53* missense mutations, missense mutations observed in one duplex read, and missense mutations observed in > 1 duplex read. **(E,F)** Linear regressions of coding *TP53* mutation frequency (MF) **(E)** and non-coding *TP53* MF **(F)** with age. Circles are colored by LFS status and filled circles indicate CTx history. Age-associated baseline regression is fit using all individuals and linear regression p-value is indicated. **(G,H)** Multiple regressions of coding *TP53* MF **(G)** and non-coding *TP53* MF **(H)** using age (scaled by decades), LFS, and CTx status as covariates. For MF calculations, sequencing depth was scaled per 10,000,000 bases. Asterisks indicate significance value (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$) and error bars indicate 95% confidence intervals. **(I)** Ratio of *TP53* coding and non-coding MF by LFS status. **(J)** *TP53* dN/dS grouped by germline p.R181H status and CTx status. Error bars indicate the interquartile range of dN/dS values obtained through bootstrap resampling. Fraction indicates the number of observed non-synonymous (top) and synonymous (bottom) mutations used in calculation. Inset shows a magnified view of the no-CTx groups only. **(K)** AlphaMissense pathogenicity scores of observed *TP53* SNVs by germline p.R181H status and CTx status.

Given that observed *TP53* coding mutations appear to be positively selected, we next calculated *TP53* coding and non-coding MF in blood samples (the latter counting only non-coding mutations captured by the *TP53* panel, Supplemental Table S9) to better resolve the contribution of selection and baseline mutagenesis respectively. Similar to the CHIP-panel results in aggregate, *TP53* coding and non-coding MF increased with age, although not significantly (Figure 2.3E-F). In multivariate models, *TP53* coding MF was only significantly associated with chemotherapy whereas *TP53* non-coding MF was asso-

ciated with p.R181H status (Figure 2.3G-H, Supplemental Table S8), mirroring the associations observed in the CHIP panel more broadly. To enable a better understanding of the relationship between *TP53* coding and non-coding MF based on LFS status, we calculated the ratio of *TP53* coding to non-coding MF for each individual (Figure 2.3I). We observed that chemotherapy increases the ratio regardless of LFS status, concordant with chemotherapy promoting the expansion of *TP53* mutant clones. Additionally, we observed that most individuals without LFS and no chemotherapy had a ratio >1 whereas the two individuals with LFS and no chemotherapy had a ratio <1 .

These results agree with the prior observation of reduced positive selection of *TP53* mutations in the blood of LFS individuals (Figure 2.2F) and indicate that, to properly compare dN/dS between groups, the individuals with chemotherapy should be considered separately. Thus, we recalculated the dN/dS for *TP53* mutations in blood by separating LFS and chemotherapy status (Figure 2.3J). We observed that chemotherapy greatly increased dN/dS across both LFS and non-LFS relative to their respective no chemotherapy groups. This supports chemotherapy as a promoter of the expansion of clones with *TP53* pathogenic mutations. Without the confounding of chemotherapy, we observed reduced positive selection in LFS individuals compared to non-LFS individuals (Figure 2.3J). Consistent with the dN/dS result, pathogenicity scores demonstrated high levels for individuals with chemotherapy and lower levels for individuals without chemotherapy with the lowest proportion of *TP53* pathogenic mutations found in LFS individuals without chemotherapy (Figure 2.3K).

In summary, the two LFS individuals without chemotherapy (LFS02 and LFS03) had high levels of non-coding mutations and no comparable increase of *TP53* coding mutations. Only 4/7 (57%) of the *TP53* coding mutations found were non-synonymous compared with 20/24 (83%) in non-LFS patients (20/24). While the difference was not statistically significant, it explains the lower dN/dS values observed in these individuals.

Overall, these results suggest that the p.R181H germline variant might be associated with an increase in mutagenesis but a reduction in the positive selection of *TP53* pathogenic clones. Unfortunately, the strong selective action of chemotherapy precludes confirmation of this finding in the blood of the treated patient.

2.2.4 Mutation frequency and signatures vary across tissue contexts

A unique feature of this study was the availability of both normal and cancerous tissue from 22 different anatomical locations (Figure 2.1A-B), which enabled us to explore *TP53* somatic evolution across different tissues in the context of a germline p.R181H pathogenic variant. We discovered considerable variation in the number of somatic *TP53* mutations in different tissues (Figure 2.4A). The majority of these mutations were missense, many of which were classified as pathogenic by AlphaMissense (Figure 2.4A). Sun-exposed skin contained more than twice as many *TP53* mutations as the next most mutated sample. Many of these mutations were CC>TT dinucleotide variants, likely related to UV-induced mutagenesis (Figure 2.4B). Four samples had disproportionately low sequencing depth compared to the rest (Figure 2.4A), which likely was responsible for their low mutation count. These samples were discarded for mutation frequency analyses. For the rest of the samples, we calculated *TP53* coding and non-coding MF (Figure 2.4C, Supplemental Table S10). With the exception of sun-exposed skin, samples had relatively low and homogeneous levels of non-coding MF (mean: 1.37×10^{-7}). *TP53* coding MF was low and comparable to non-coding MF in bone marrow, peritoneum, testis, kidney and some cancerous samples. In contrast, *TP53* coding MF was elevated in sun-exposed skin, peripheral blood samples (whole blood, buffy coat and plasma), and normal esophagus, which was the tissue of origin of the cancer. Interestingly, the patient had chronic gastritis and *TP53* coding MF was also elevated in the two normal gastric samples. The primary and metastatic cancer samples did not have higher *TP53* coding MF compared to healthy tissues. All

cancer and metastatic samples, however, had loss of heterozygosity (LOH) of the wild-type *TP53* allele as reflected by variant allele frequencies >0.5 for the germline p.R181H variant (Supplemental Figure A.5). These results indicate that the loss of the wild-type allele likely drove tumorigenesis in this individual. LOH-driven carcinogenesis may also explain the lower coding MF of the primary tumor compared to non-cancer esophagus as biallelic *TP53* inactivation may eliminate the selective advantage for additional *TP53* pathogenic mutations. Several non-cancer tissue samples (lung, bronchi, esophagus, omentum, and gastric tissue) also showed variant allele frequencies >0.5 for p.R181H which might reflect early LOH expansion or, most likely, cancer dissemination (micrometastases) in those tissues, which are in proximal location to the primary and metastatic sites.

Mutational signature extraction and decomposition across a subset of the sampled tissues further supported variation in active mutational processes in different tissue contexts (Figure 2.4D). All tissues contained clocklike mutational signatures SBS1 and SBS5, with SBS5 being more prevalent than SBS1 in all tissues. As expected, mutations in sun-exposed skin were dominated by UV-associated signatures SBS7a/b/d. Both primary esophageal cancer samples, four metastatic cancer samples and one healthy esophageal sample showed contributions from SBS17a/b, which have previously been linked to esophageal carcinoma and premalignant Barrett’s esophagus [130]. SBS17b has also been associated with 5-FU treatment [9], making it difficult to unambiguously disentangle the observed drivers of mutagenesis. The high resemblance in mutational signatures between replicate samples (two esophageal cancer samples and two liver metastases) demonstrates the robustness of our approach for high resolution mutation detection and signature extraction and decomposition.

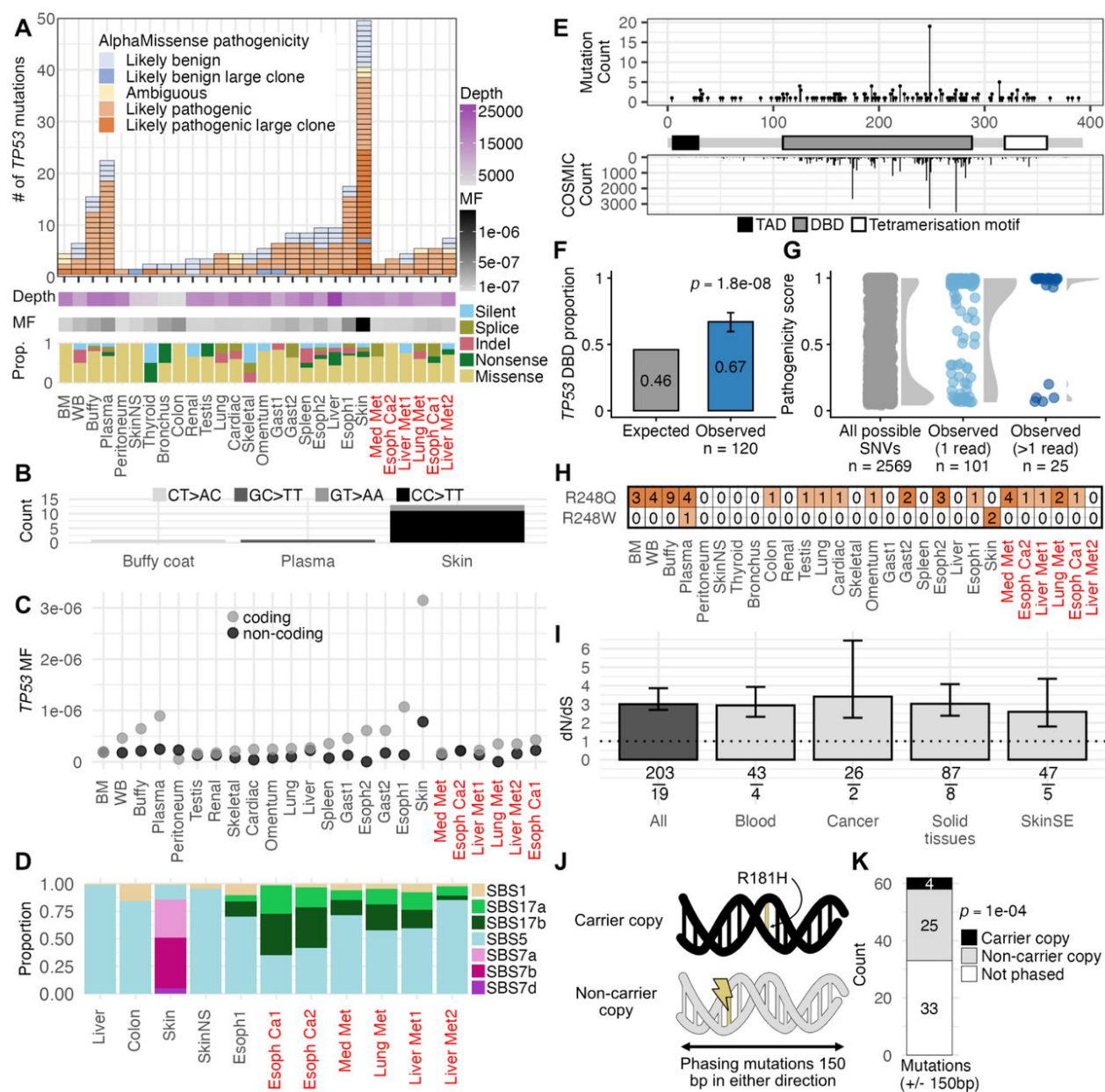


Figure 2.4: Multi-tissue ultra-deep characterization of *TP53* mutation and selection in an individual with Li-Fraumeni Syndrome and esophageal adenocarcinoma.

Figure 2.4: (A) Number, size, and pathogenicity of coding *TP53* mutations by tissue in LFS01. Each box represents a unique *TP53* mutation within a tissue. Box color indicates pathogenicity based on mutation type and AlphaMissense score (see Methods). Mutations seen in more than 1 duplex read are considered large clones. Heatmaps indicate duplex depth (purple) and mutation frequency (MF, black). Bar plot represents the proportion of *TP53* mutation type by tissue. Cancer samples indicated in red text. (B) *TP53* dinucleotide variant counts by tissue in which they were observed. (C) *TP53* coding and non-coding MF by tissue. (D) Proportion of mutations associated with a mutational signature using mSigHDP. Only samples sequenced with *TP53*+MUT panel were included. Cancer samples indicated in red text. (E) Single nucleotide variant (SNV) counts within the *TP53* coding region across all tissues in LFS01 (top) and mutation counts in the COSMIC database (bottom) by codon position. Major functional regions shown: topologically associating domain (TAD), DNA-binding domain (DBD), and tetramerization motif. (F) Expected and observed SNV proportion within the DNA-binding domain of *TP53*. Expected proportion corresponds to the DNA-binding domain's base pair proportion of the total *TP53* coding region. Binomial test p-value and error bars representing 95% confidence interval are shown. (G) AlphaMissense pathogenicity scores of all possible *TP53* SNVs, SNVs observed in one duplex read, and SNVs observed in > 1 duplex read. (H) Size and tissue location of *TP53* mutations by amino acid substitution at codon position 248. The number inside of each box represents the number of duplex reads associated with that given mutation. Cancer samples indicated in red text. (I) *TP53* dN/dS grouped by tissue categories, with all tissues aggregated within All. Error bars indicate the interquartile range of dN/dS values obtained through bootstrap resampling. Fraction indicates the number of observed non-synonymous (top) and synonymous (bottom) mutations used in calculation. (J) Phasing analysis schematic. The chromosome copy carrying p.R181H and non-carrier copy may acquire new somatic mutations (yellow). When a somatic mutation is observed on the same duplex read as the germline mutation site, the allelic phase relative to p.R181H can be determined. (K) Number of somatic *TP53* mutations phased on the carrier copy, non-carrier copy, or not capable of being phased. Binomial test significance between alleles indicated with p-value.

2.2.5 *TP53 R248Q/W is a mutational hotspot recurrently found in several tissues*

A powerful attribute of examining multiple tissues from a single individual is the ability to characterize evolutionary processes of particular genetic alterations occurring independently in matched genetic backgrounds, potentially indicating their selective importance. We therefore examined if the observed mutations preferentially clustered in particular regions of *TP53*. The majority of discovered SNVs were observed in the DNA-binding domain, similar to *TP53* mutations in COSMIC as a whole (Figure 2.4E), our findings in blood (Figure 2.3B), and in prior studies of somatic *TP53* mutations in normal tissue [39, 46, 131]. The enrichment of *TP53* mutations in the DNA-binding domain was statistically significant (Figure 2.4F) indicating positive selection of functionally disruptive mutations as a second hit in the context of a germline p.R181H background. In addition, observed mutations were enriched for higher AlphaMissense scores than expected compared to the distribution of scores for all possible *TP53* single-nucleotide substitutions, especially if the mutations were supported by more than one duplex read, indicating the expansion of pathogenic clones (Figure 2.4G). Strikingly, we found that 18/28 tissue samples contained a mutation at codon 248, the second most widely-mutated *TP53* hotspot in the COSMIC database (Figure 2.4E,H). We observed two different amino acid identities (p.R248Q and p.R248W), both of which are inactivating and dominant negative (21,31,32).

R248Q was observed more frequently than R248W across tissue contexts, although the plasma sample contained both simultaneously (Figure 2.4H). While these mutations remained at low variant allele frequencies that would be undetectable by standard sequencing methods, 9/19 were observed in >1 duplex read, indicating some degree of clonal expansion. While it is possible for leukocyte clones to infiltrate the tissues and be sampled and detected by ultra-deep sequencing, it is unlikely for a clone of such low variant allele frequency (<0.0008) to repeatedly be sampled across different tissues. In addition, we found little evidence that biological or technical contamination was responsible for

this widespread mutational sharing, as most other mutations in the blood were not found across tissues (Supplemental Figure A.6). In general, no other mutation showed such a widespread presence across tissues. The second most frequent mutation was S314F, which was found only in four tissues (Supplemental Figure A.7). These results broadly suggest convergent evolution favoring disruption of R248 across diverse tissue contexts in the background of a germline p.R181H variant.

2.2.6 Chromosomal but not tissue-context accounts for differences in selection pressures in TP53

We next examined if *TP53* mutations were subject to different selective pressures across tissue contexts. To ensure sufficiently large sample sizes for robust computation of dN/dS, we aggregated samples into the larger anatomical categorizations of blood, solid tissues, sun-exposed skin, and cancer (Supplemental Table S11). Consistent with our blood findings for LFS01 (Figure 2.3J), we found significant elevation of dN/dS across multiple tissue contexts, suggesting positive selection for secondary *TP53* mutations (Figure 2.4I). Given the negative selection for *TP53* mutations observed in the blood of p.R181H carriers without chemotherapy, it is likely the positive selection observed in these samples to be a consequence of chemotherapy exposure. The magnitude of dN/dS was similar across tissue groupings, although these groupings may mask more localized tissue effects.

Finally, using the large sample size of *TP53* mutations observed across tissues in all LFS individuals within our cohort, we asked if the chromosomal context of *TP53* mutations affected their ability to rise to observable frequencies. Specifically, we hypothesized that there might be greater positive selection for *TP53* mutations occurring on the chromosomal copy not already carrying p.R181H, as opposed to the partially inactivated p.R181H carrier copy (Figure 2.4J). The relatively short length of duplex reads limited the number of somatic mutations that could be confidently phased to 29. Of these muta-

tions, significantly more (25/29; binomial test, $p = 0.0001$) occurred on the chromosomal copy not already carrying p.R181H, indicating increased selection for second hit mutations affecting the non-carrier *TP53* copy (Figure 2.4K). This significant allelic imbalance supports ongoing selection for second-hit *TP53* inactivating mutations in cells with p.R181H.

2.3 Discussion

Ultra-deep sequencing of normal tissues has revealed that somatic evolution is widespread [15, 39, 46]. Measuring mutational and selective processes across tissues, sexes, environmental exposures and populations is providing new understanding of the mechanisms driving cancer formation [7, 121, 132]. Here, we apply this ultra-deep sequencing approach to investigate somatic evolution in a family of individuals with Li-Fraumeni Syndrome caused by the germline *TP53* pathogenic variant p.R181H, one of the most prevalent mutations causing cancer predisposition [79, 124, 133]. We find that the germline p.R181H variant is associated with increased mutational frequency in the mutagenesis panel but not in the coding regions of CHIP and/or AML-related genes. A stark exception, however, is the study proband, who had a large number of mutations in CHIP panel genes likely related to chemotherapy exposure, which precluded disambiguation of the effect of the germline mutation vs the cancer treatment. Despite that limitation, we observed an interesting pattern of second hit mutations in the proband with recurrent pathogenic substitutions in p.R248 across tissues. We also confirm expected patterns of somatic evolution in p.R181H carriers and non-carriers that suggest our analyses are robust, including the finding of chemotherapy mutational signatures in blood, UV light signatures in sun-exposed skin, increased mutation frequency with age, and positive selection in *DNMT3A* and *TET2*, the most common CHIP genes [22, 24]. In addition, we report the novel observation of strong negative selection in *GATA2*, which is consistent with its essential role in hematopoietic

stem cell survival [134] and implies that even low-frequency deleterious variants are likely purged before clonal expansion can occur. Conversely, the enrichment of *DNMT3A* mutations mirrors its well-established role as an early driver in CHIP. Together, these findings underscore that the mutational landscape in healthy individuals already reflects lineage-defining constraints that foreshadow later MDS/AML evolution, and that this landscape does not appear to be perturbed by the germline *TP53* p.R181H variant.

The increased frequency of mutations in the mutagenesis panel for LFS carriers indicates that p.R181H may affect mutagenesis. The primary phenotypic alterations of R181H mutants are their inability to transactivate p53 target genes and initiate cell cycle arrest, although these effects may be lessened in a heterozygous state [133]. A potential result of this could be the accumulation of mutations in cells that would otherwise not proceed through division, although it is unclear why this would lead to an observed increase in the frequency of mutations in the mutagenesis panel, but not the CHIP panel. One possibility is that increased mutagenesis affects both panels, but the patterns in coding regions are obscured by the dynamics of selection. This would be consistent with the findings of reduced positive selection for *TP53* in the relatives with germline p.R181H although it is unclear whether this would apply to other genes as well. Sequencing of more p.R181H carriers without chemotherapy exposure as well as different tissue samples and genes would be key to further explore these findings.

We found divergent strengths of positive selection on *TP53* mutations between the proband LFS01 and their relatives LFS02 and LFS03, all p.R181H carriers. The relatives showed reduced positive selection in *TP53*, as evidenced by a deficit of coding mutations, $dN/dS < 1$, and lower predicted mutational pathogenicity compared to LFS01 and the non-p.R181H carriers profiled. In contrast, LFS01 showed abundant evidence for positive selection in *TP53* both in the blood and across tissues. A likely driver of these divergent dynamics is that LFS01 is the only relative to have received chemotherapy, which has been

extensively shown to be associated with *TP53* clonal expansions [36–38, 80]. Finding reduced positive selection for *TP53* mutations among untreated individuals carrying the p.R181H variant may point to the importance of LOH rather than secondary SNVs to inactivate the second copy of *TP53*. It has been reported that loss of heterozygosity of *TP53* is an early and near-ubiquitous event in LFS cancers [81] and LOH was indeed the second hit in the tumor of the study proband. While LOH might be the critical second hit in the progression to cancer, we did find some evidence that SNV-based *TP53* mutations are important in p.R181H carriers. We found a surprising recurrence of the p.R248Q/W mutations and we demonstrate that most SNVs occur in the alternate allele carrying the WT R181 amino acid. These results suggest that second hit *TP53* mutations do initiate clonal expansions but to what extent this is restricted to the selective pressure of chemotherapy is unknown.

Profiling cancerous and healthy tissue samples from distinct anatomical locations revealed some expected differences in the mutational and selective patterns of different organs in a p.R181H donor. Specifically, we observed greater intensity of CC>TT dinucleotide variants associated with UV-induced mutagenesis in the skin, and cancer-specific mutational signatures in tumor samples. We also observed that the biopsies with the highest levels of *TP53* mutation frequency, aside from sun-exposed skin, were those procured from the esophagus and the stomach. Interestingly, the patient had chronic acid reflux, which is a risk factor for esophageal cancer, and chronic gastritis. Both conditions are associated with a carcinogenic process driven by *TP53* mutations, although in the esophagus, this typically involves Barrett’s metaplasia [135, 136] and in the stomach it is often related to *Helicobacter pylori* infection [137, 138]. However, pathological assessment of the tissues did not identify Barrett’s esophagus or *H. pylori* infection in this patient. Nevertheless two independent biopsies of these two tissues showed high levels of *TP53* mutation supporting the link between chronic tissue damage and *TP53* clonal expansion. We could

speculate that the germline p.R181H variant might have contributed to a faster carcinogenic process in the esophagus in the context of preexisting epithelial damage, but functional studies would be needed to test this hypothesis. Regarding selection, we did not observe differences across tissues using dN/dS calculations. However, these calculations were performed by grouping mutations into tissue categories due to sample size considerations, which might have hidden associations present only in specific anatomical locations or cell types. In addition, these associations might also be overshadowed by the selective effect of chemotherapy. Access to normal tissue of LFS subjects without chemotherapy and sampling of multiple biopsies will be key to solve these questions in future studies.

Examining multiple tissues from a single individual also revealed widespread parallelism in the clonal evolutionary processes, especially for *TP53* mutations p.R248Q/W which was found in >60% of LFS01 tissues. These mutations are dominant negative, associated with increased cellular proliferation and reduced apoptotic response [55, 126, 139, 140], and are widely mutated in human cancers. Given their prevalence in cancer, it may be unsurprising to observe their simultaneous expansions in multiple normal tissues. However, we did not see similar expansions at other putatively dominant negative *TP53* hotspots (i.e., R175H, Y220C, M237I, R273H, R282W), potentially pointing to a distinct interaction between p.R181H and p.R248Q/W. While this finding could be the result of technical or biological cross-contamination across samples, both possibilities seem unlikely given SNP quality control, the lack of other shared mutations across tissues, and the low variant allele frequency of p.R248Q/W in all samples. It would be of interest to see whether similar preferences for second hits are observed for other *TP53* germline pathogenic variants.

The study had some limitations. First, while we had control samples for blood, we did not have control samples for the normal tissue, and available studies with comparable sequencing depths in concordant tissues are non-existent. Second, LFS01 had chemother-

apy, confounding potential selective dynamics in a p.R181H background. Inclusion of a chemotherapy exposed control (CON07) was useful to partially delineate the selective effects in blood, but further studies would benefit from sampling normal tissue in LFS patients prior to chemotherapy. Blood, buccal swabs, urine, and biopsies from colonoscopies are all feasible procedures for normal tissue collection. Lastly, tissues were only sequenced for *TP53* and a subset for the mutagenesis panel. Now that sequencing is more affordable, extended panels are likely to provide more comprehensive information. Additionally, we only profiled one LFS pathogenic germline variant. While this is one of the more common germline *TP53* pathogenic variants, it also is unusual in its reduced penetrance, atypical cancer profile, and later age onset of cancer [70, 79, 122, 123]. Profiling somatic evolution in people with LFS and different germline *TP53* pathogenic variants will be necessary to understand the generalizability of our findings related to the increased non-coding mutagenesis, the lessening of *TP53* mutation selection, and the preference for second hits in p.R248. Other germline variants with higher penetrance might lead to different phenotypes that could be exploited to personalize surveillance and cancer prevention.

In summary, we demonstrate how ultra-deep sequencing can reveal the evolutionary processes operating in healthy and cancerous tissues of individuals carrying the germline *TP53* p.R181H variant. This study paves the way for other analyses in LFS patients and other cancer predisposition syndromes. We discovered distinct features of mutagenesis and selection in LFS tissues, which reflect underlying biological processes that are incompletely understood and deserve further investigation.

2.4 *Materials & Methods*

2.4.1 *Sample collection*

The study includes samples collected at autopsy from a male patient known to have LFS, who died from metastatic esophageal cancer at the age of 34 (23 normal tissues, 6 tumor tissues) (Figure 2.1, Supplemental Tables S1 and S2). The study also includes blood samples (PBMCs) collected from three family members and from 7 unrelated control donors (Figure 2.1, Supplemental Tables S1 and S2). Samples from the LFS-affected patient were collected through the willd body program at UW. The collection and analysis of samples from the patient, family members and donors was approved under IRB study number STUDY00020611.

Anatomic samples from the LFS-affected patient were collected at autopsy and immediately frozen at -80°C with an estimated post-mortem interval of 22h. The patient had a smoking history, chronic acid reflux, and chronic gastritis. Gastric tissue immunohistochemistry was negative for *Helicobacter pylori* and there was no evidence of goblet cell metaplasia or Barrett’s esophagus. The main tumor was moderately differentiated (grade II of III) adenocarcinoma in esophageal/gastric junction in a background of glandular dysplasia and ulceration. There was a satellite nodule in the cardia of the stomach 10 mm away from the lower border of the main cancer. The nodule had high-grade glandular dysplasia which is a precancerous lesion to adenocarcinoma. The patient had received chemotherapy consisting of folinic acid/5-fluorouracil/oxaliplatin (FOLFOX) for six cycles, followed by treatment with folinic acid/5-fluorouracil/oxaliplatin, and achieved a partial response. As disease progressed, he was treated with a regimen consisting of ramucirumab/docetaxel/zoledronic acid and received palliative radiotherapy for metastatic bone lesions. None of the other subjects in the study had cancer except for the 76-year-old control donor, who had a history of non-Hodgkin lymphoma and was treated with two cycles

of cyclophosphamide/doxorubicin/vincristine/prednisone (CHOP).

Frozen solid tissue samples were manually macrodissected using tissue punchers for a target size of 3mm³ or $\sim$ 25 mg of tissue. Single use tools were used to avoid cross-contamination between samples. Tissue DNA was extracted using the DNeasy Blood and Tissue Qiagen kit following the tissue protocol. Hematopoietic samples from the patient (whole blood, buffy coat, plasma, bone marrow) and PBMCs from family members and donors were also extracted with the DNeasy Blood and Tissue Qiagen kit following the blood protocol (buffy coat) or the cell protocol (rest of samples). DNA was quantified with a Thermo Qubit and quality checked with an Agilent TapeStation Genomic Tape.

2.4.2 Duplex sequencing and variant calling

Libraries for duplex sequencing were prepared using TwinStrand kits (V1 chemistry) with 500ng input DNA. Libraries were captured with 3 different panel combinations depending on the sample type (Supplemental Table S2). All solid tissue samples were captured with the *TP53* panel, which includes the coding region and adjacent non-coding areas of the *TP53* gene ($\sim$ 3,000 bp). The tumor tissues and a subset of normal tissues were captured with the *TP53* panel and a mutagenesis panel (TwinStrand’s Human Mutagenesis Assay), which includes probes covering 20 human genome regions spanning across 20 chromosomes designed to be an unbiased representation of sequence contexts and covering no loci with known selective advantages (48,000 bp, [128]). All blood samples were captured with the mutagenesis panel and a CHIP panel that includes the full coding sequence or hotspots of 29 genes recurrently mutated in adult AML and CHIP (TwinStrand’s AML-29 Assay, $\sim$ 59,000 bp). The CHIP panel includes *TP53* probes. Briefly, samples were digested with enzymatic fragmentation, end-repaired, A-tailed, ligated to TS adapters, amplified, and captured by hybridization with the corresponding probe panels. Two rounds of capture were performed to increase on-target reads. Then libraries were amplified, quantified, and

pooled for sequencing on an Illumina NovaSeq 6000 using S4 flow cells. Raw reads were processed with the DNAnexus platform for duplex sequencing analysis with TwinStrand software version 3.20.1, which uses a variant caller based on VarDict [141]. The pipeline produced duplex error-corrected VCF and depth by position files, which were used for downstream analyses.

2.4.3 Mutation classification and pathogenicity

VCF file outputs were converted to MAF files for each sample with Vcf2Maf v1.6.22 using the Ensembl Variant Effect Predictor (VEP) tool version 112 to assign variant consequences. For each mutation, the variant allele frequency was determined by calculating the ratio of mutant duplex reads to the total duplex depth at the corresponding mutated position. Variants were discarded if they were in positions with depth lower than 1000 duplex reads, a no call fraction greater than 0.1, a variant allele frequency greater than 0.3 (corresponding to germline polymorphisms), or located in repeat regions defined using UCSC RepeatMasker and regions associated with known sequencing artifacts (bedfile provided at https://github.com/Colegrove/LiFraumeni_evolution/inputs/BEDs/LiFraumeni.mask.bed). To account for sequencing artifacts from potentially problematic regions, variants were discarded if they appeared in three or more samples and the DNAnexus-hosted pipeline annotated them as having 8 or more mismatches in the duplex read containing the mutation, suggesting a consistent pattern of mismatches. To account for potential SNP cross-contamination, variants at frequency greater than 0.3 in any sample were excluded from all other samples. Variants located within coding exons or extending up to two nucleotides into adjacent splice sites and indels that overlapped coding regions were considered coding variants. Mutation frequencies (MF) were calculated for the mutagenesis panel, CHIP coding mutations, *TP53* coding mutations, and *TP53* non-coding mutations as the number of unique mutations in each of those genomic regions divided by the total number of du-

plex nucleotides sequenced in those regions. We also calculated coding MF for each of the genes in the CHIP panel. Mutation burden (MB) was calculated for coding genes within the CHIP panel as the number of mutant duplex bases divided by the total sequencing depth across all coding positions in aggregate. All postprocessing analyses unless noted otherwise were performed with RStudio version 2025.09.1 using R version 4.5.1 and packages as indicated in Supplemental Table S12.

Variants in *TP53* were assigned a pathogenic annotation and score based on AlphaMissense [142] and mutation type. For missense variants, pathogenicity was annotated using AlphaMissense, which assigns a pathogenicity score between 0 and 1 such that values >0.564 are considered Likely Pathogenic, ≥ 0.34 but ≤ 0.564 are considered Ambiguous, and scores <0.34 are Likely Benign. Nonsense variants, indels, splice mutations, and multiple nucleotide substitution variants were additionally assigned as Likely Pathogenic and assigned a score of 1. Synonymous mutations were assigned Likely Benign and a score of 0. Large clones were defined as mutations supported by more than one mutant duplex read, as each duplex read represents an independent DNA molecule.

2.4.4 Testing for selection

To assess evidence for selection on somatic mutations, we calculated gene-specific dN/dS ratios for LFS and non-LFS groups within CHIP panel genes. Synonymous and nonsynonymous SNVs were aggregated by gene and LFS status. Nonsynonymous mutations included missense, nonsense, and splice region mutations. The number of possible synonymous and nonsynonymous mutations for each gene was determined using release 115 of the Ensembl Variant Effect Predictor (VEP) web tool within the coding regions of the sequenced gene area. For mutations with multiple VEP annotated consequences, the most severe consequence as ranked by Ensembl was used. To enable dN/dS calculations even in the absence of synonymous mutations, for each gene tested we added a pseudocount

equal to the expected proportion of non-synonymous mutations to the numerator and the expected proportion of synonymous mutations to the denominator. For each gene and group, we resampled observed mutations with replacement 5,000 times to generate a bootstrapped dN/dS distribution. For the *TP53* dN/dS analysis within LFS01’s tissues, samples were aggregated into broader anatomical categories (blood, solid tissue, sun-exposed skin, and cancer) to ensure sufficient sample sizes (Supplemental Table S11).

Phasing of *TP53* somatic mutations relative to the germline p.R181H allele was performed at the duplex consensus sequence-level using BAM files in Python version 3.12.3 with the pysam package v0.22.1. Somatic SNVs and indels located within 150 base pairs of the germline p.R181H site (hg38 chr17:7675070) were selected as candidate variants for phasing, corresponding approximately to the maximum duplex read length. If at least one duplex read spanned both the variant and the p.R181H site, we used the base identity of the p.R181H site to assign the variant to the LFS carrier or non-carrier *TP53* copy. Otherwise, the variant was designated as unphased. For variants containing more than one spanning duplex read, all duplex reads that could be phased were in agreement on the phase.

To contextualize patterns of mutation recurrence and see if mutations in the blood of all subjects and within the tissues of LFS01 overlapped with established recurrently mutated sites such as hotspots, we compared our observed *TP53* mutation distribution to reference data from the Catalogue of Somatic Mutations in Cancer (COSMIC) database v102 (GRCh38). COSMIC mutation data including missense, nonsense, and synonymous substitutions were downloaded as counts per amino acid mutation across all tumor types. We grouped entries by amino acid position and summed counts to generate total codon mutation frequencies. These frequencies were used to create reference lollipop plots aligned to the *TP53* SNVs across all LFS01 samples or across all subject blood samples.

2.4.5 Regression analysis

First, to examine the relationships between coding mutation frequency (MF), mutagenesis MF, or mutation burden (MB) and age, we fit linear regression models. For each subject, we calculated coding and mutagenesis MF as the number of unique coding or mutagenesis mutations across all genes in the CHIP panel divided by the total number of coding or mutagenesis nucleotides sequenced and MB as the number of mutant coding bases divided by the total number of coding nucleotides sequenced. Linear models in the form $Y = 1 + X$ were fit separately for coding and mutagenesis MF or MB (Y) against age in years (X) using ordinary least squares regression implemented through the `lm()` function in R.

To further account for confounding variables, we performed multiple linear regression. The number of unique coding mutations and the number of mutant coding or mutagenesis bases across all genes in the CHIP panel were modeled separately with age (in decades), LFS status, and CTx status as covariates. We fit models in the form $Y = 1 + age + LFS + CTx$ where LFS and CTx were treated as binary covariates using Gaussian generalized linear models with an identity link function as implemented in `glm()` in R. Analyses were also performed separately for each gene.

2.4.6 Mutational spectrum and signature analysis

For a signature agnostic comparison between mutational spectra, we used AMSD v0.3.0 [129] performed using the “mean” option between p.R181H carrier and non-carrier groups and excluding samples with previous exposure to chemotherapy.

Mutational signatures were extracted to evaluate mutagenic processes in p.R181H carrier and non-carrier individuals. Only solid tissues sequenced with *TP53*+MUT panels were included in signature analysis and all hematopoietic samples. A separate signature extraction was performed for solid tissue samples and for hematopoietic samples. In both

instances, single base substitutions were initially classified by their 96 trinucleotide contexts for pyrimidine-mutated base pairs using SigProfilerMatrixGenerator v.1.3.3. Mutational signatures were then extracted in each case using a Bayesian Hierarchical Dirichlet Process (mSig HDP, [143]). Tissue samples were pre-annotated as either ‘non-cancerous’ or ‘malignant,’ to facilitate hierarchical clustering of samples to enhance the accuracy of signature extraction [143]. Extracted signatures from hematopoietic samples were decomposed using a recently published reference signature database for blood and chemotherapy-associated mutational signatures [10, 144]. Four mutational signatures were extracted across hematopoietic samples: SBS1, SBS5, SBS7a, and SBSG. Extracted mutational signatures from solid tissue samples were decomposed using SigProfiler Extractor v.1.1.21 with COSMIC v3.4 reference database [144].

Seven mutational signatures were extracted from solid tissue samples: SBS1, SBS5, SBS7a/b/d, and SBS17a/b. Although LFS01 had received 5-Fluorouracil chemotherapy, the signature for this agent (SBSG) was not incorporated into tissue analysis due to its similarity with SBS17a and SBS17b, which are known to be involved in esophageal adenocarcinoma [130].

2.4.7 Data access

All data associated with this study are provided in the paper or the Supplemental Materials. Sequencing data from this study will be made available through the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject>) prior to publication.

2.4.8 Code availability

The code to reproduce all analyses can be found at .

2.5 Competing interest statements

THS, ZKN, JEH, CCV, EKS, and JJS are equity holders in TwinStrand Biosciences Inc. JEH, CCV, and JJS are named inventors on one or more patents owned by TwinStrand Biosciences Inc. EKS, JJS, and RAR are named inventors on one or more patents owned by the University of Washington and licensed to TwinStrand Biosciences Inc. and for which receive royalties. THS, ZKN, JEH, CCV, EKS, and JJS are previous employees of TwinStrand Biosciences Inc. RAR was a consultant and equity holder at TwinStrand Biosciences Inc. No disclosures were reported by the other authors.

2.6 Funding

National Institutes of Health grant R01 CA259384 (RAR)

National Institutes of Health grant DP2 CA280623 (AFF)

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

2.7 Author contributions

JJS, MSH, MEDG, and RAR conceived and designed the study; HLC, HM, RR, AF, and RAR designed the analysis; MEDG, DAM, EQK, JJS, MSH, and RAR contributed to sample acquisition; THS, ZKN, FYL, EKS, JEH, and CCV processed samples and performed sequencing and initial data analysis; HLC, HM, BFK, AF, and RAR analyzed the data; HLC, HM, RR, AF, and RAR interpreted the data; HLC, AF, and RAR wrote the article; JJS, MSH, and RR reviewed the article and provided feedback. All authors read and approved the final version of the manuscript.

2.8 Acknowledgements

We deeply thank the patient, the family, and the donors that provided samples, without whom this research would not have been possible. We thank the Willed Body Program at the University of Washington School of Medicine, the gastrointestinal oncology team at Fred Hutchinson Cancer Center who cared for this patient, the University of Washington autopsy service team who performed the pan tissue collection, and the University of Washington Medical Scientist Training Program classes of 2020 and 2021 (CONJ 511: Genomic Dissection) who provided excellent discussion about the case. We also thank Lauren Santos, program coordinator under Dr. Dubard-Gault at Fred Hutchinson Cancer Center, who helped consent individuals for this project and E. Jonlin for regulatory support.

Chapter 3

EPITHELIAL COMPETITION DETERMINES GENE THERAPY POTENTIAL TO SUPPRESS FANCONI ANEMIA ORAL CANCER RISK

3.1 Introduction

Fanconi Anemia (FA) is an inherited genetic disorder and cancer predisposition syndrome linked to increased risk of bone marrow failure, leukemia, and epithelial cancers. FA arises from biallelic pathogenic mutations in any of 23 *FANC* genes that work in concert as a pathway (the ‘FA pathway’) to detect and coordinate the repair of many types of DNA damage including DNA interstrand crosslinks and double-strand breaks [52, 53, 72]. This compromised repair capacity is associated with an elevated risk of certain cancers, particularly head and neck squamous cell carcinomas (HNSCC). FA patients experience HNSCC rates several hundred to over a thousand fold greater than the general population [62, 71, 145–147], and a corresponding elevation in mucosal anogenital squamous carcinomas. These cancers often contain pathogenic mutations in *TP53*, a crucial tumor suppressor gene [148–150], and are genomically unstable. Further, they are difficult to treat via standard care chemotherapy and radiation due to the FA-associated constitutional DNA repair defect [62, 77, 151, 152].

Given these challenges, corrective gene therapy represents a promising strategy for

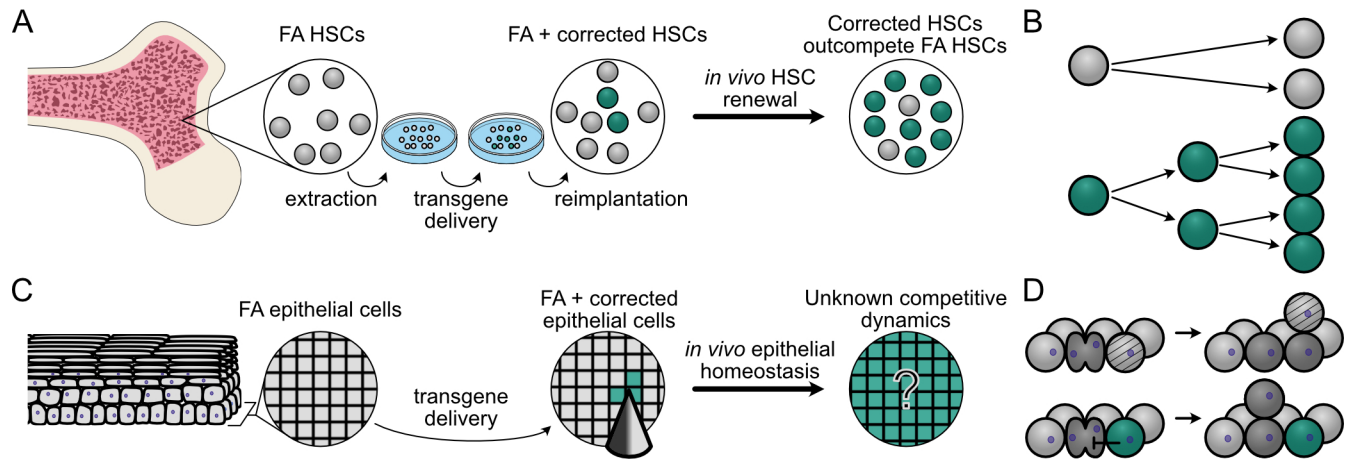


Figure 3.1: Modeling gene therapy in FA HSCs and oral epithelium. (A) Hematopoietic stem cell (HSC) gene therapy in FA. FA HSCs (grey) are harvested from FA patient bone marrow, corrected ex vivo (green), then re-implanted to expand in vivo. (B) Corrected HSCs (green) have a proliferative advantage compared to FA HSCs (grey) and outcompete uncorrected HSCs over time in vivo. (C) Oral mucosal gene therapy in FA. In situ gene delivery (here via microneedles) into oral epithelium to correct the underlying FANC gene defect. Over time, corrected cells (green) may or may not clonally expand within FA oral epithelium to replace FA host cells (grey). (D) In our model, FA epithelial cells (light grey) are displaced (light grey with bands) by dividing FA cells (dark grey, top), in contrast to gene-corrected epithelial cells (green) that can resist displacement by dividing neighbors (dark grey) to persist and expand in the basal layer proliferative niche (bottom).

the prevention of FA-associated HNSCC. This strategy has already shown success in FA-associated bone marrow failure, where gene-corrected hematopoietic stem cells (HSCs) can reverse marrow failure and suppress the risk of developing leukemia: HSCs are removed, the specific mutant *FANC* gene is corrected ex vivo, and corrected cells are re-implanted into the bone marrow (Figure 3.1A). Recent FA gene therapy clinical studies have found that gene-corrected FA HSCs possess a proliferative advantage over uncorrected FA HSCs (Figure 3.1B, [112–114, 116, 153]). As a result, small numbers of implanted corrected HSCs have been shown to expand over time to reverse FA-associated bone marrow failure.

The success of HSC gene correction in treating marrow failure suggests that oral gene

correction may provide a parallel way to prevent FA-associated HNSCC. However, gene correcting mucosal epithelial cells may prove more challenging than HSCs due to the constraints of tissue architecture. Epithelial tissues are tightly organized, which could limit corrected cell expansion when compared with HSCs in bone marrow (Figure 3.1C). Numerous theoretical and empirical studies have suggested that cells growing under spatial restriction grow more slowly than their unstructured counterparts [28, 29, 154–157]. However, cellular lineages can clearly expand within epithelial tissues, e.g., those lineages harboring cancer driver mutations [29, 47, 48], or spontaneous revertant mutations in the context of heritable skin diseases such as epidermolysis bullosa [87, 105, 107]. Thus understanding the degree of proliferative advantage required for FA-corrected oral epithelial cells to expand and persist will be essential in devising mucosal gene correction protocols for FA epithelium.

A second major challenge of preventative gene therapy against FA-associated HNSCCs is that oral epithelial cells must be corrected *in situ* rather than removed and reimplanted (Figure 3.1C). This will require identifying efficient mucosal gene correction protocols that favor the expansion and persistence of gene-corrected epithelial cells. One promising new approach is to adapt microneedle arrays originally developed for mucosal vaccination to deliver gene correction reagents [158–163]. Microneedle arrays can deliver customizable doses of correction reagents in different spatial configurations, but the optimal array configurations and microneedle properties required to achieve efficient delivery will need to be determined.

Computational modeling provides an efficient framework in which to integrate both different delivery approaches and the effect of epithelial structure on gene correction to optimize therapeutic success. Agent-based models have a long history of providing valuable insight into cell-cell competition and the emergent spatial dynamics shaping tissue behavior [16, 88, 90]. We therefore extended an existing agent-based model of squamous epithe-

lium [30] to include gene correction delivered by arrays of diffusible microneedles in order to answer several key questions: 1) what degree of proliferative advantage must corrected cells possess to persist and spread effectively in the oral epithelium; 2) how do different gene correction efficiencies and spatially-structured delivery strategies such as microneedle arrays affect long-term gene correction success; and 3) can mucosal gene correction prevent the clonal spread of pathogenic *TP53* mutations and the progression to oral cancer in individuals with FA? Our results provide a framework for understanding the dynamics of gene therapy in FA oral epithelium, and how best to use mucosal gene correction to reduce cancer risk and improve FA patient health outcomes.

3.2 Results

3.2.1 *In silico* model of gene correction in the oral epithelia

In order to simulate the impact of gene therapy in epithelial tissues we employed the HomeostaticEpidermis model, a three-dimensional lattice-based hybrid cellular automaton originally developed to model epidermal dynamics over human lifespans [30]. In this model, each epithelial cell is an independent agent whose rate of division and death are determined by a diffusible growth factor emanating from the basal layer. Dividing cells reside on a basal layer where the resultant progeny are displaced upward, undergo terminal differentiation and are eventually shed at the epithelial surface. This model robustly captures key epithelial features including structure, homeostasis with cell turnover, and realistic clonal expansion of cells that resist death or differentiation. We modified cellular turnover rates to adapt the epidermal model to better recapitulate properties of the oral mucosa (see Discussion and Materials and Methods).

We further adapted the HomeostaticEpidermis model to explore FA oral gene therapy by introducing simulated gene correction to confer a potential proliferative advantage

over uncorrected FA cells. In stratified squamous epithelium, clonal expansion may result from a bias in basal layer cell fate decisions that favors the production of progenitor over differentiated cell progeny [164–166]. While the mechanisms driving a potential epithelial proliferative advantage in the context of FA gene correction are unknown, evidence from FA HSC correction experiments shows that a selective advantage can be driven by resistance to apoptosis, increased survival in the presence of DNA damaging agents, and longer-term maintenance of self-renewal capacity [53, 72, 167–170]. All of these factors would contribute to longer-term residence of corrected cells in the basal layer, allowing them to retain proliferative capacity. We therefore modeled a proliferative advantage as a persistence coefficient, p_{corr} , which represents the probability that a corrected cell resists displacement out of the basal layer during neighboring cell divisions (Figure 3.1D). This mirrors the approach from Schenck et al. for modeling the proliferative advantage of *NOTCH1* mutations in skin [30], and permitted us to explore the potential effect of gene correction in an FA background, where gene-corrected cells had anywhere from no advantage (or a neutral $p_{corr} = 0$) to an extremely strong advantage ($p_{corr} = 1$). We first investigated the proliferative advantage necessary to promote reliable expansion of corrected cellular lineages in an FA mucosal background.

3.2.2 Gene correction requires a strong proliferative advantage to spread over clinically relevant time periods

We hypothesized that gene-corrected cells will require a proliferative advantage compared to uncorrected FA cells to spread through FA oral mucosa. To set baseline expectations for cellular competition without a proliferative advantage, we first considered the case in which a corrected cell has equivalent division behavior to the FA background: that is, when $p_{corr} = 0$. We simulated 0.67 mm² (100 x 100 cells) FA tissue sections and delivered gene correction to 10 cells clustered in the tissue center (mirroring single microneedle de-

livery, Figure 3.2A). Because the median age at which FA patients develop solid tumors is ~ 30 years [71, 147], we reasoned that corrected cells would need to displace FA cells in the first decade or two after gene correction to be most clinically impactful. We thus followed simulated tissue sections for up to 50 years to determine the fate of corrected cells and their descendants (Figure 3.2A). We considered three fates: all corrected cells could be displaced from the basal layer (i.e., “loss”); corrected cells could overtake 80% or more of the basal layer (i.e., “confluence”); or neither of these outcomes might be reached (i.e., “ongoing” dynamics). Note, our use of the word confluence parallels its standard definition as the percentage of a culture dish covered by a cellular population, as we are interested in the state of gene correction reaching widespread basal layer coverage (a metric for therapeutic success). Under neutral conditions in which corrected cells have no replicative advantage over uncorrected cells ($p_{corr} = 0$), nearly all simulations (97%) led to the loss of corrected cells by 50 years (Figure 3.2B) with most corrected cell loss occurring within months of correction (Figure 3.2C). In the remaining 3% of simulations, neutral corrected patches remained, but were of modest size with the largest corrected clone reaching only 0.15 mm^2 . Thus, in the absence of a proliferative advantage, gene correction will not be able to efficiently convert FA epithelium.

We next investigated how increasing the proliferative advantage might enable more efficient spread of corrected cells through epithelium on clinically-relevant timelines. While we have strong evidence that *FANC* gene reversions possess a proliferative advantage in non-epithelial contexts (e.g., HSCs), we have little direct knowledge about the degree of proliferative advantage conferred by *FANC* gene correction in epithelial tissues. We therefore investigated a wide range of persistence coefficients ranging from unlikely to guaranteed persistence ($p_{corr} = 0.001, 0.01, 0.1, 0.2, 0.5, 1$) to identify how strong a proliferative advantage must be to ensure durable spread. More biological context for the strength of these persistence coefficients is given in Supplemental Table 1. In all experiments, FA cells

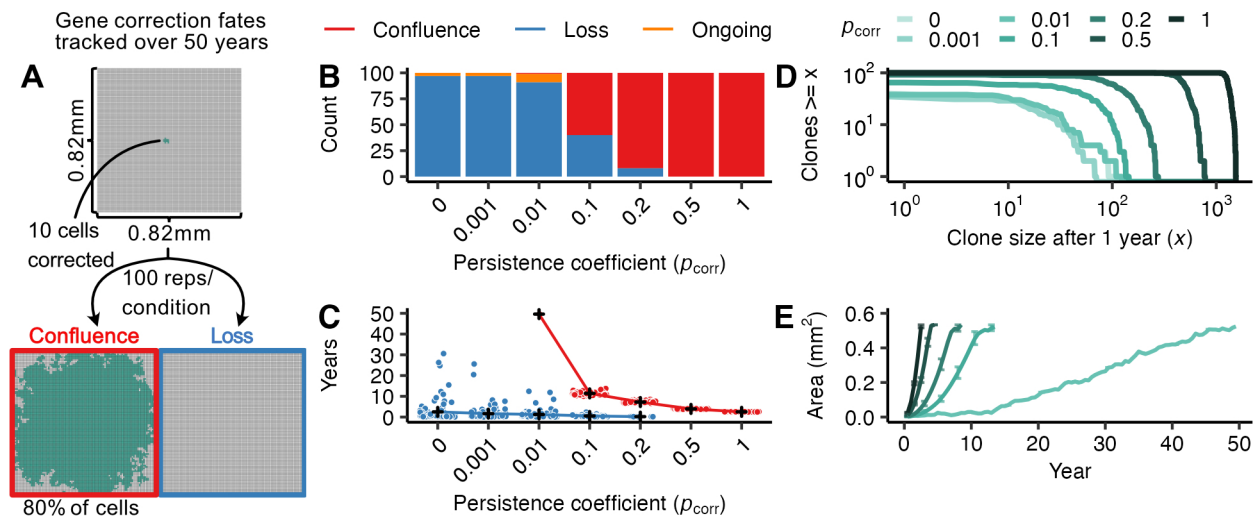


Figure 3.2: Mucosal gene therapy depends on a strong proliferative advantage of gene-corrected cells.

(A) 10 gene-corrected cells (green) were tracked over 50 years in a 0.67 mm² FA mucosal tissue section (100 replicate simulations per condition). The gene correction fates were tracked to identify simulations reaching confluence (80% of basal layer cells corrected, red box) or loss (no remaining gene-corrected cells, blue box). (B) Number of simulations achieving confluence (red), loss (blue), or ongoing expansion (orange) at 50 years as a function of persistence coefficient (p_{corr} = 0, 0.001, 0.01, 0.2, 0.5, 1). (C) Time at which simulations reached confluence or loss as a function of persistence coefficient, with mean times for each p_{corr} indicated in black (+) and outcome of individual simulations as points. Gene-corrected patches not reaching confluence or loss by 50 years are not plotted. (D) Number and size of gene-corrected patches at 1 year as a function of persistence coefficient p_{corr} . (E) Average area of confluent gene-corrected patches as a function of time and persistence coefficient p_{corr} . Bars indicate interquartile ranges.

did not resist displacement (i.e., had a persistence coefficient of 0).

As expected, larger persistence coefficients decreased the probability of corrected cell loss (Figure 3.2B), with a $p_{corr} \geq 0.1$ required to avoid most loss events. When loss did occur at large persistence coefficients, it occurred early with 88% of losses within the first year after gene correction (Figure 3.2C). This is analogous to the establishment frequency in population genetics, where once beneficial mutations reach a certain frequency, loss becomes increasingly unlikely. In contrast, gene correction with persistence coefficients closer to neutrality can persist for years or decades before loss though rarely expand to any significant degree. Full clone size distributions one year after correction are shown in Figure 3.2D as a function of p_{corr} .

For gene-corrected cells that escape loss, larger persistence coefficients also increased the speed of corrected cell spread. Corrected cells with smaller persistence coefficients ($p_{corr} < 0.1$) that escaped loss almost never reached confluence in the tissue sections by 50 years. In contrast, gene correction with $p_{corr} \geq 0.1$ that avoided early loss always reached confluence by 50 years (Figure 3.2B). As the persistence coefficient increased, corrected cell confluence was achieved at progressively earlier times (Figure 3.2C), with $p_{corr} = 1$ growing at approximately five times the speed of $p_{corr} = 0.1$ ($0.25 \pm 0.027 \text{ SE } mm^2/year$ versus $0.048 \pm 2.0 \cdot 10^{-3} \text{ SE } mm^2/year$, Figure 3.2E). These results indicate that effective oral mucosal gene correction must confer a substantial proliferative advantage ($p_{corr} \geq 0.1$) to ensure persistence and spread on clinically-relevant timescales.

3.2.3 *Optimizing transgene delivery for maximal corrected tissue coverage*

Having demonstrated above that individual gene-corrected patches with a strong proliferative advantage could spread locally, we next considered how patch spread was determined by transgene dose and the spatial distribution of corrected cells, where microneedle delivery allows gene dose, spacing, and in-tissue diffusibility to be modified as part of micronee-

dle design and fabrication.

We first investigated how a greater number of corrected cells, equated here to transgene dose, affected the probability of corrected cell loss and time to confluence when delivered in a single compact patch (Figure 3.3A). As expected, correcting more cells decreased the probability of loss across persistence coefficients (Figure 3.3B), where 30 as opposed to 10 corrected cells decreased the probability of loss for $p_{corr} = 0.1$ from 50% to 10%, but not the time to confluence under favorable conditions (Figure 3.3C). This reflects that the time needed to grow from 3 cells to 30 is short, compared to the longer period needed to replace a majority of cells in a target tissue of size 0.67 mm^2 (circa 10 years for $p_{corr} = 0.1$). Avoiding early corrected cell loss is the primary advantage conferred by a larger dose at a single injection site.

We next investigated how transgene diffusion from a microneedle injection site might improve corrected cell spread by distributing a fixed number of corrected cells over a larger tissue area. We reasoned that more spatially distributed delivery might allow corrected cells to more effectively compete for space and replace FA cells. We modeled transgene diffusion from a point of delivery with an approximation to Brownian diffusion. The diffusion constant D (see Materials and Methods) was examined for $D \in (2, 10, 20)$ to capture low, intermediate and high degrees of spatial diffusion from 10 initial corrected cells (Figure 3.3D). Surprisingly, increased diffusibility did not affect the probability of corrected patch loss or its rate of spread (Figure 3.3EF). One explanation is that early stochastic loss of individual corrected cells is more likely driven by being displaced by neighboring cells (regardless of correction status), rather than a failure to displace corrected neighboring cells when dividing. These results indicate that efforts to enable wider transgene diffusion from single microneedles is unlikely to improve the spread of corrected cell patches.

A key advantage of microneedles is their ability to be arrayed in different configurations to optimize transgene delivery. In light of finding that transgene diffusibility min-

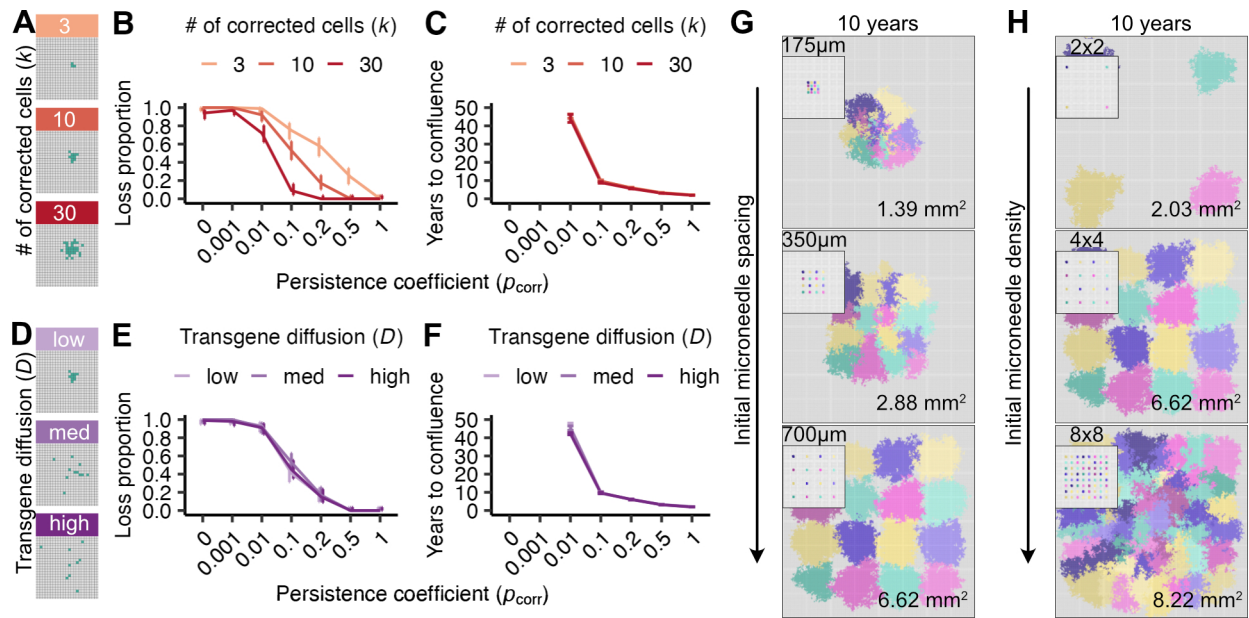


Figure 3.3: Higher corrected cell numbers and optimal spacing can increase the likelihood of tissue replacement despite clonal interference. Gene-corrected patch loss probability and time to confluence as a function of corrected cell number (**A-C**) and spatial delivery distribution (**D-F**). 100 simulations/condition were run for each persistence coefficient p_{corr} , with error bars indicating binomial errors (**B,E**) or interquartile ranges (**C, F**). (**A**) Initial spatial distribution of gene-corrected cells ($k = 3, 10, 30$ cells, in green) with a constant diffusion coefficient ($D = 2$). Probability of gene-corrected patch loss (**B**) and time to confluence (**C**) as a function of persistence coefficient (p_{corr}) and initial corrected cell number (k). (**D**) Initial spatial distribution of ten gene-corrected cells (green) as a function of diffusion coefficient ($D = 2, 10, 20$), corresponding to low, medium and high diffusion). Probability of gene-corrected patch loss (**E**) and time to confluence (**F**) as a function of persistence coefficient (p_{corr}) and diffusion coefficient (D). Visualizations of gene-corrected cell patches ten years after correction as a function of microneedle spacing (**G**) or density (**H**) on 10.67 mm^2 tissue sections. Initial arrayed delivery shown in panel insets, and each color represents the descendants of a single microneedle that corrects $k = 30$ cells with $D = 2$. Areas of corrected cell patches at ten years are quantified in the lower right corner of each tissue section.

imally affected the rate of spread from single needles, we investigated how microneedle arrays could be used to optimize transgene delivery and tissue correction over a 10 year period.

We first asked how needle spacing might be configured to avoid competition and promote spread within a tissue section. We reasoned that individual microneedles could be designed to minimize loss, with interference between microneedle sites a potentially more important determinant of the spread of corrected cells as illustrated in Figures 3D-F. We investigated the impact of needle spacing on transgene delivery, corrected cell persistence and rate of spread using a 4x4 array of 16 low diffusibility ($D = 2$) microneedles in which each needle corrected 30 cells to confer a strong proliferative advantage ($p_{corr} = 0.1$). The needle spacings we explored, of 175 μm , 350 μm , and 700 μm , reflect commonly engineered interneedle distances, where our target for gene correction and spread was a 10.67 mm^2 tissue section (Figure 3.3G). We found that the most widely spaced microneedle arrays corrected more than five times as much tissue area in a 10 year period than the most tightly spaced arrays for a given number of microneedles. In tightly spaced arrays, corrected cells at the perimeter expanded more than those in the interior; this apparent competition among corrected cells was not observed in the most widely spaced arrays. Thus appropriate transgene delivery spacing has the potential to minimize interference or competition of corrected cells and maximize the area of corrected tissue.

Microneedle arrays can be engineered with substantially higher densities than the ones we examined above. Although tightly spaced arrays led to a slower rate of confluence, they achieved greater localized tissue coverage than did widely spaced arrays (Figure 3.3G). We found that further increasing needle densities led to more complete conversion of a given tissue area, but did not limit the approach to confluence (Figure 3.3H). For example, increasing microneedle density to 8x8 on a fixed backing size of 4.43 mm^2 increased tissue correction by $\sim 25\%$, whereas decreasing microneedle density to 2x2 on the same backing

patch size decreased the tissue correction by $\sim 70\%$. These results indicate that dense microneedle array transgene delivery in conjunction with a strong proliferative advantage can provide widespread tissue level correction even when promoting potential spatial competition among corrected cells and clones. Thus our model should be useful for testing a broad array of potential gene correction arrays to optimize gene delivery under different clinical constraints (see Discussion).

3.2.4 Gene correction reduces $TP53^-$ clonal expansion in genomically unstable FA epithelium

If FA gene-corrected cells can spread through an FA oral epithelium, how might this counter HNSCC risk or progression? An early stage in the development of FA-associated and sporadic HNSCCs is the clonal expansion of pathogenic mucosal $TP53$ mutations [148–150]. We thus asked if FA gene correction could counter the spread of $TP53$ mutations in oral epithelium. In our model, mutations in a single copy of $TP53$ can occur in either corrected (here denoted $FANC^+$) or FA (here denoted $FANC^-$) cells according to mutation rates $FANC^+$ and $FANC^-$ (Figure 3.4A), with $TP53$ mutations conferring persistence coefficient advantages of $p_{FANC^+}^{TP53^-}$ and $p_{FANC^-}^{TP53^-}$, respectively. As before, corrected $FANC^+$ cells without $TP53$ mutations have a persistence coefficient of p_{corr} , so corrected $FANC^+$ cells with a $TP53$ mutation have a total persistence coefficient of $p_{corr} + p_{FANC^+}^{TP53^-}$ (Figure 3.4B). We estimated a mutation rate (μ_{FANC^+}) and persistence coefficient ($p_{FANC^+}^{TP53^-}$) of $TP53$ mutations in a $FANC^+$ background based on the rate of occurrence and size of $TP53$ clones in healthy human esophageal tissue (see Materials and Methods). In the absence of direct measures, we initially assumed that the rates of occurrence and clonal expansion of $TP53$ mutations were similar in $FANC^+$ and $FANC^-$ tissues (i.e., $\mu_{FANC^+} = \mu_{FANC^-}$ and $p_{FANC^+}^{TP53^-} = p_{FANC^-}^{TP53^-}$). We reasoned that even if FA gene-corrected and uncorrected cells produced the same number of $TP53$ mutations with comparable proliferative advantages,

gene correction might make $TP53$ mutant cells less able to displace $FANC^+$ as opposed to $FANC^-$ cells, thus restricting their expansion potential.

To test this idea, we asked if gene correction could limit $TP53$ mutational prevalence in FA tissue when $TP53$ mutation rates and proliferative advantages were comparable for $FANC^-$ and $FANC^+$ cells (Figure 3.4CF). We simulated 0.33 mm^2 (70×70 cells) tissue sections in which a single microneedle gene corrected 30 cells ($D = 2$) with persistence coefficients of $p_{corr} = 0.01$ or $p_{corr} = 0.1$ at the tissue center, then followed these sections up to 50 years. We compared these corrected tissue sections to uncorrected tissue section controls, and also plotted $TP53^-$ expansion in healthy esophageal tissue as a reference [2]. We found that $TP53$ mutations reached similar tissue frequencies over 46 years regardless of gene correction status (Figure 3.4CF), as might be expected under an additive fitness model. Gene correction restricted $TP53^-$ spread in $FANC^-$ cells, while mutations still arose and expanded within corrected $FANC^+$ cells leading to similar overall tissue frequencies (compare cyan $FANC^+/TP53^-$ cells with $p_{corr} = 0.1$ in Figure 3.4F tissue sections to uncorrected magenta $FANC^-/TP53^-$ cells). Gene correction might be beneficial if $FANC^-/TP53^-$ double mutants are more likely to progress to cancer than $FANC^+/TP53^-$ mutants. However, this is unlikely to lower the overall $TP53^-$ burden if $TP53$ mutation rates and proliferative advantages are comparable between corrected and uncorrected tissue. Full 46-year tissue coverage trajectories from these simulations are shown in Supplemental Figure B.1A-C, and demonstrate that $TP53^-$ tissue coverage over time is similar for each value of p_{corr} .

The above results assumed that the $TP53$ mutation rate and mutant proliferative advantage are identical regardless of $FANC$ genotype. However, the dramatically increased risk of HNSCC in FA individuals likely reflects genomic instability in FA that could generate and promote the spread of pathogenic mutations through FA epithelia. We therefore investigated how FA genotype affects the rate of $TP53$ mutation generation and spread,

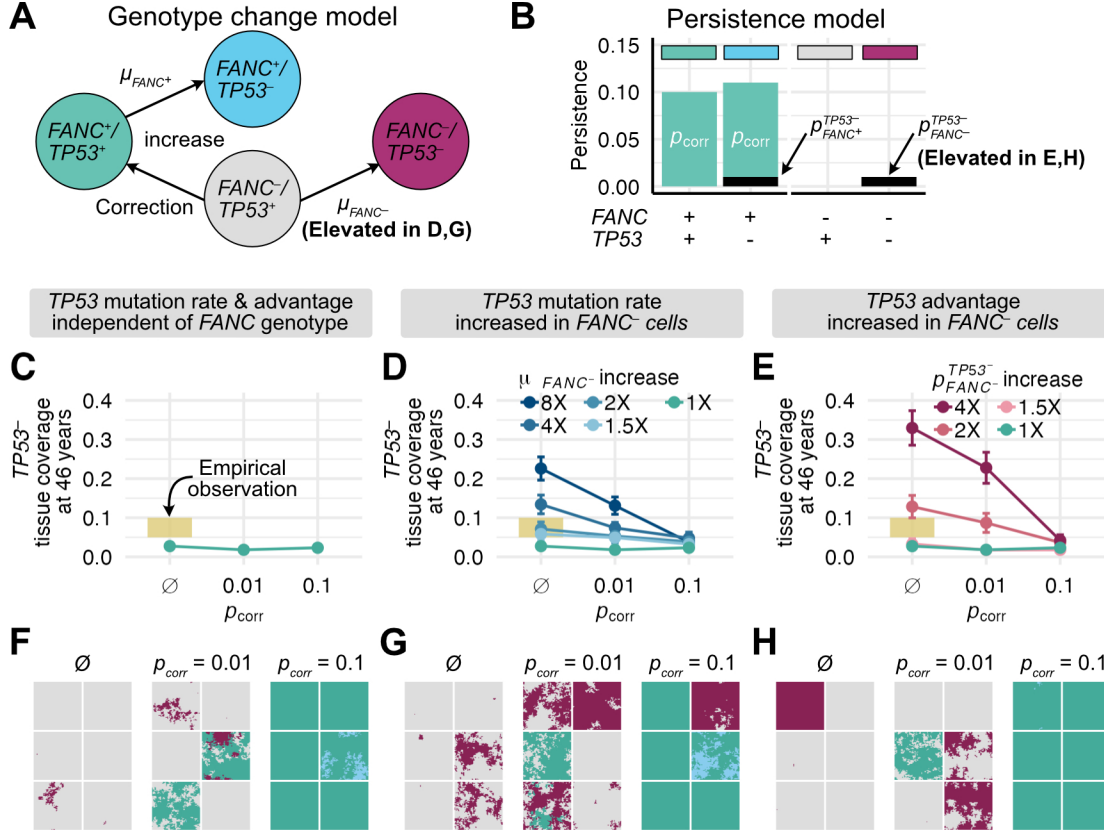


Figure 3.4: Gene correction reduces spread of $TP53$ mutations through FA tissue. (A) Model of four cellular genotypes at two loci: FA cells ($FANC^-/TP53^+$) in grey; FA cells with $TP53$ mutations ($FANC^-/TP53^-$) in magenta; gene-corrected cells ($FANC^+/TP53^+$) in green; gene-corrected cells with $TP53$ mutations ($FANC^+/TP53^-$) in cyan. Arrows indicate transitions that can occur via mutation or gene correction. (B) Genotype-specific persistence coefficients for (A). (C-H) $TP53^-$ tissue coverage as a function of gene correction at 46 years in 0.33 mm^2 simulated tissue sections, comparing no correction ($\emptyset$) to correction with $p_{corr} = 0.01$ or $p_{corr} = 0.1$. Panels used 100 simulations at a given persistence coefficient except in panels C,F where 300 simulations were performed. C-E show the proportion of tissue coverage at 46 years with (C) comparable $TP53$ mutation rates and persistence coefficients regardless of FA genotype, (D) an elevated $TP53$ mutation rate in $FANC^-$ cells ($\mu_{FANC^-} = m * \mu_{FANC^+}$, $m = (1.5, 2, 4, 8)$) and equivalent $TP53$ persistence coefficients or (E) equal $TP53$ mutation rates in $FANC^-$ and $FANC^+$ cells with elevated $TP53$ persistence coefficients in $FANC^-$ cells ($p_{FANC^-}^{TP53^-} = r * p_{FANC^+}^{TP53^-}$, $r = (1.5, 2, 4)$). Error bars represent standard errors. Yellow bars represent $TP53$ tissue coverage range from Martincorena et al. [2] for normal esophagus. (F-H): six representative tissue sections at 46 years from C-E, with colors corresponding to genotypes shown in (A). (G) Consequence of an 8-fold increase in μ_{FANC^-} , and (H) a 4-fold increase in $p_{FANC^-}^{TP53^-}$.

and thus could determine the outcome of FA gene correction.

We first considered that $FANC^-$ cells have a higher mutation rate than $FANC^+$ cells [97, 98, 171]. As there have been a range of values reported for this increase (up to 8-30x), we reran the experiments above to determine outcomes with $TP53$ mutation rates in uncorrected $FANC^-$ cells (μ_{FANC^-}) of 1.5-, 2-, 4-, or 8-fold higher than in corrected $FANC^+$ cells (μ_{FANC^+} , Figure 3.4DG). In the absence of FA gene correction, $FANC^-$ increased by a factor of eight led to a corresponding increase in the tissue burden of $TP53^-$ cells to cover 23% of tissue sections at 46 years, or nearly 10-fold higher than the coverage estimate of 2.7% under the background mutation rate (Figure 3.4D). While elevated mutation rates produced more $TP53^-$ clones, individual clone sizes remained similar to those under baseline mutation rates (Supplemental Figure B.2AB). We found that by introducing gene correction with a small persistence coefficient ($p_{corr} = 0.01$, Figure 3.4D), $TP53^-$ tissue coverage was modestly reduced compared to uncorrected tissue (a reduction of 42% in simulations where $FANC^-$ cells had a 8x elevated mutation rate). In contrast, gene correction with a larger persistence coefficient ($p_{corr} = 0.1$, Figure 3.4D) resulted in tissue sections with almost no additional $TP53$ mutations above baseline simulations that used background mutation rates. This effect depended on gene-corrected $FANC^+$ cells replacing more mutable $FANC^-$ cells: in the 2% of $p_{corr} = 0.1$ simulations in which gene corrected cells were stochastically lost, $FANC^-/TP53^-$ cells expanded as if they were in uncorrected tissue (see panel inset in Figure 3.4G). Full 46-year tissue coverage trajectories from these simulations are shown in Supplemental Figure B.1D-F.

We next considered that $TP53$ mutations might confer a stronger proliferative advantage in FA cells than in gene-corrected cells (i.e., $p_{FANC^+}^{TP53^-} < p_{FANC^-}^{TP53^-}$). The basis of this assumption is observed non-additive effects on the rate of cell turnover and tumor formation in cells harboring both $TP53$ mutations and FA mutations (see Discussion, [74, 172–174]). We parameterized possible fitness advantages using human lymphoblastoid and

mouse HSC experimental data (see Materials and Methods). To test the impact of this effect, we reran the experiments above with identical $FANC^-$ and $FANC^+$ mutation rates, but with $p_{FANC^-}^{TP53^-}$ increased by a factor of 1.5-, 2-, or 4-fold compared to $p_{FANC^+}^{TP53^-}$ (Figure 3.4EH). In the absence of gene correction, when $p_{FANC^-}^{TP53^-}$ was increased by a factor of 4, there was a corresponding increase in the tissue burden of $TP53^-$ cells to cover 33% of tissue sections at 46 years, compared with 2.7% under the background proliferative advantage. In contrast to the mutation rate experiments, increased $TP53^-$ coverage was driven primarily by larger clones in conjunction with a more modest increase in clone number (Supplemental Figure B.2CD). As was the case for gene correction affecting mutation rates, gene correction with a small persistence coefficient ($p_{corr} = 0.01$, Figure 3.4E) modestly reduced the $TP53^-$ tissue burden by $\sim 31\%$ among cells with a 4x elevated proliferative advantage compared to uncorrected tissue. Gene correction with a large persistence coefficient ($p_{corr} = 0.1$, Figure 3.4E) resulted in tissue sections with almost no excess $TP53^-$ mutational coverage beyond baseline simulations in which $FANC^-$ and $FANC^+$ cells have identical persistence coefficients associated with $TP53$ loss. Again, the rapid spread of gene-corrected cells largely abrogated the proliferative advantage of $TP53$ mutations in a $FANC^-$ background. Full 46-year tissue coverage trajectories are shown in Supplemental Figure B.1G-I. These results collectively indicate that FA-directed mucosal gene therapy has the potential to reduce $TP53$ -associated HNSCCs in FA, and that likely mechanisms include a suppression of FA-associated genomic instability and the strong selective pressure for loss of $TP53$ function.

3.3 Discussion

We extended an agent-based model of squamous epithelial clonal competitive dynamics [30] to study epithelial gene correction in oral mucosa. We chose this approach over the

more analytically-tractable branching-process models widely used to study tumorigenesis [175, 176] because it permitted us to more faithfully capture biologically important spatial constraints on the spread of mutations and gene correction in an epithelial context. These constraints are critical to understand in order to develop highly effective gene correction protocols for FA or other epithelial diseases, as spatial and non-spatial growth patterns have well-characterized diverging dynamics [28, 29, 154–157]. This approach recapitulated certain results well-appreciated from branching process models (i.e., correcting more cells limits stochastic loss, cells with small fitness advantages can persist at low frequencies at length), but it also provided practical observations into potential spread rates of corrected patches, optimal strategies for designing correction arrays (see more below), and the role of interference both among patches and with pathogenic clones, all of which rely on incorporating spatial competition explicitly.

A compelling biological rationale for extending the HomeostaticEpidermis (HE) model is the many shared features of oral mucosa and epidermis. These tissues are contiguous at the proximal end of the GI tract, and share common embryological origins, many biological features and pathogenic mechanisms. Alterations in progenitor cell dynamics in these epithelia can promote both pre-neoplastic and neoplastic disease by facilitating the expansion of mutant clones [29, 35, 47, 48, 177, 178]. Nevertheless, tissue-specific differences do exist between the oral mucosa and the epidermis, which could affect specific quantitative conclusions reached by our model, especially related to rate of spread of gene correction through tissue. To minimize these effects, we modified the HE model to incorporate oral mucosa-specific data where they exist (e.g., on oral mucosal turnover, cellular density) and further note that many epidermis-specific features (e.g., keratinization, presence of hair follicles, certain wound healing properties) are not represented in the HE model and thus do not impact our findings.

Implementation of our adapted HE model provided several important insights into

oral mucosal dynamics, and the potential to counter cancer risk in individuals with FA:

- 1) gene correction needs to confer a strong proliferative advantage to escape stochastic corrected cell loss, and to promote tissue replacement on clinically-relevant decadal timescales;
- 2) tissue replacement success can be maximized by increasing delivery dose and by distributing tissue delivery spacing to speed tissue replacement while reducing competitive interference among corrected lineages; and 3) gene correction has the substantial potential to reduce the high risk of HNSCC in FA if it lowers the mutation rate and/or proliferative advantage of oncogenic mutations. Each of these results and their translational importance are discussed in greater detail below.

We found that gene corrected cells will require a strong proliferative advantage (i.e., $p_{corr} > 0.1$) and rapid spread within the oral epithelium for effective FA oral gene correction. Whether FA gene correction can confer a strong proliferative advantage has been most convincingly demonstrated in the hematopoietic system [112–114, 116, 153], and is plausible though less well-documented in oral mucosa. The potential for persistence and rapid spread in related squamous epithelia is exemplified by heritable skin diseases such as epidermolysis bullosa, where patients may display mutationally revertant skin patches 2 cm² or larger, even at ages as young as 8, and with geometries inconsistent with emergence during development [87, 105, 107]. Similar data exist from individuals with ‘ichthyosis-with-confetti’, who have large revertant skin patches caused by mitotic recombination that can reach up to 4 cm² in size and have been documented in individuals as young as 18 [106]. Comparably rapid tissue replacement may occur in the oral mucosa, which has faster measured cell turnover than does epidermis [164, 179–181]. If FA gene correction confers only a weak intrinsic mucosal fitness advantage, this might be augmented by engineering additional genetic alterations. For example, recent studies have found that *NOTCH1* mutations can confer a fitness advantage that promotes clonal expansion while minimally affecting cancer risk [47].

Our model predicts that gene correction will be most beneficial at suppressing cancer risk in FA if it reduces genomic instability or counters the spread of pathogenic epithelial mutations. There is a strong biological rationale for these suppositions. FA cells accumulate chromosomal aberrations due to a DNA damage sensing and repair deficit [54, 73, 148, 182], and tumors that arise on this background are further mutated and chromosomally rearranged [148]. Complementing the FA defect suppresses these phenotypes, whether by gene correction [112–114, 116, 153] or spontaneous genetic reversion [99, 100, 102, 103, 183–185], in FA patient HSCs. The restoration of FA function reduces DNA damage markers, restores chemoresistance and suppresses genomic instability. Moreover, FA gene correction also has the potential to suppress the spread of pathogenic mutations and tumor progression. The plausibility of these additional effects has been observed in *Fancc2/TP53*, *Fancc/TP53* and *Fancc/TP53* double knock-out mice: all have significantly lower rates of cancer-free survival than either single mutant or control mice for epithelial as well as non-epithelial tumors [173, 174, 186]. The importance of the FA-*TP53* interaction is illustrated in mouse HSCs, where *TP53* elimination on a *Fancc* mutant background increases cellular turnover rates to a greater extent than on a *Fancc*-competent background [172]. A potential mechanism for this increased fitness advantage is that persistent DNA damage in FA activates *TP53* to arrest cell cycling for DNA damage repair [74, 187]. *TP53* loss, in contrast, confers a growth advantage by suppressing proliferation arrest to allow continued cell division in genomically unstable or mutant cells [74, 172, 188]. Gene therapy to restore FA pathway function could thus suppress multiple FA-dependent mechanisms that promote genomic instability, and that provide strong selective pressure for the emergence of *TP53* mutations.

Our oral mucosal epithelial agent-based model could be extended or improved in several ways. First, we could more explicitly model the basal layer stem cell hierarchy. This would likely improve the biological accuracy of the model without substantially altering

global tissue dynamics [30]. Second, our use of a persistence coefficient could be improved using experimental data that quantify the survival benefits of gene correction in an FA background, when such data become available in an epithelial context. In FA-corrected HSCs, proliferative advantage appears to reflect a combination of reduced apoptosis, increased survival after DNA damage, and/or improved self-renewal capabilities [167–169]. Better characterization of these potential mechanisms in FA epithelia should improve modeling insight and predictive outcomes. Third, we used existing related data outside FA oral mucosa to fit certain model parameters. We fit the fitness benefits of *TP53* mutations using clonal expansion data from the esophagus. Oral and esophageal mucosa are contiguous and related, but we note that *TP53* mutations can expand differentially across more distinct tissue types [1, 2]. We also approximated the potential fitnesses of epithelial *TP53*/FA double mutants and FA mutation rate elevation using non-epithelial cell type data (see Methods), although we explored a wide range of potential interaction effects for both of these experiments (Fig 4D and 4E). Direct measures of these parameters in their exact tissue contexts would improve specific quantitative model predictions, but would be unlikely to qualitatively change the patterns that we observe.

More sophisticated genetic models of FA or other diseases could also be developed to provide additional mechanistic insight while capturing between-individual differences. For example, we explored only the effects of single deleterious *TP53* mutations as some *TP53* mutations exhibit dominant-negative effects with heterozygous clonal expansion in mice [48, 189, 190]. A second way to more accurately model FA would be to incorporate loss of function of one of the 23 FA genes with loss of function variants in *ADH5* and *ALDH2*. These two genes are key members of aldehyde detoxification pathways that limit the toxicity of formaldehyde and acetaldehyde, both common and unavoidable sources of endogenous and exogenous DNA damage that activates *TP53*-dependent DNA damage signaling and FA pathway activity [191]. Our modeling approach can be extended to predict a range

of fitness effects of these inputs in both FA and FA gene-corrected cells. A final note is that our modeling framework can accommodate a wide range of new data from many gene delivery platforms in addition to microneedle-mediated gene delivery [158–163].

Gene therapy represents a promising approach to treat or prevent a growing number of disease states and predispositions. Computational modeling, as illustrated here, can provide a useful framework that identifies important determinants of therapeutic success across many diseases and their target cells, tissues or organs. A critical next step in determining the potential of FA oral epithelial gene therapy is to quantify the proliferative advantage of gene-corrected cells in FA epithelia. It may be possible to estimate proliferative advantages now for a subset of FA patients using deep sequencing of oral mucosal biopsies, and from these data estimate reversion rates and the likely revertant proliferative advantage for FA-revertant cells. Several new mouse models of FA (see, e.g., [186]) should soon allow modeling predictions to be tested directly by lineage tracking to quantify the fitness effect of FA gene correction. These mouse models have the advantage of adding spatial distribution data on the spread of FA-corrected cells. All of these approaches should provide new insight into the competitive dynamics of FA gene-corrected cells.

A final important point is that many gene therapy or editing protocols are now focused on in vivo gene delivery, where success will depend on integrating target cell and tissue type-specific features with delivery method-specific details [192, 193]. Our approach allows many of these protocol-, patient- and disease-specific details to be captured and explored *in silico*, and to aid the design of clinical protocols with the highest likelihood of success [194]. Modeling also has the potential to minimize excessive animal testing and patient sampling. Key protocol details (delivery vehicle, delivery route, dose and timeline) can be identified and then tracked to measure protocol effectiveness. Clinical data from small n trials - or even single patients - can then be used to improve protocol design, implementation and tracking of patient-specific outcomes (see, e.g., the recent $n = 1$ in vivo

editing of CPS1 deficiency [195]). All of these efforts should substantially improve the design, implementation and eventual success of gene therapy protocols to treat or prevent a wide range of human diseases.

3.4 *Acknowledgements*

We acknowledge useful discussion with Kim Woodrow, Craig Dorrell, Markus Grompe, members of the Feder lab at UW and helpful comments from three anonymous reviewers. This work was supported by a Research Award from the Fanconi Cancer Foundation to RJM and by 1DP2CA280623 to AFF.

3.5 *Materials & Methods*

3.5.1 *Overview of the model*

Model Framework: We extended a cellular automaton model of epithelial cell division and homeostasis in the Hybrid Automata Library [89], HomeostaticEpidermis (HE) [30], to simulate the dynamics of oral epithelial cells and corrective gene therapy. Briefly, HE simulates cells as independent agents on a three dimensional lattice that divide and die according to the available concentration of a growth factor diffusing from the basal layer (see [30] for full parameterization of this growth factor gradient). During a cellular division event, one daughter cell occupies the location of the parent cell, while the other is positioned either vertically above or in one of the four orthogonal neighboring positions within the basal layer Von Neumann neighborhood. The placement of this second daughter cell is governed by the division location probability ω , where $\omega = 0.2518617$ represents the probability of vertical placement, and each orthogonal direction has a probability of $(1 - \omega)/4$. If a daughter cell occupies a neighboring position within the basal layer, the displaced cell

stratifies out of the basal layer and all cells above it are displaced upwards. The epithelial dynamics of this study focus only on the basal layer, as once a cell stratifies to the suprabasal layers, cells do not de-differentiate and re-enter the basal layer. Cells can mutate during division, which we examine in the latter half of our study. This model reproduces key behaviors of epithelial tissues, including the distribution of mutated clone sizes over time. As such, most existing model parameters are kept at their default values to preserve homeostatic tissue function. New functions and parameters specific to our model are described below.

In our model, all cells begin with mutant *FANC* gene function (i.e., an FA phenotype), as expected in an inherited disorder. Throughout most of the analyses, this FA phenotype does not affect the normal maintenance of homeostatic equilibrium or cellular behavior, although in section Lineage competition between gene-corrected cells and *TP53* mutants, we describe how uncorrected FA cells may affect genome stability and control of *TP53*⁻ mediated clonal expansions.

Adaptation of epidermal model for the oral epithelium: We sought to preserve homeostatic behavior of the HE model while accounting for the increased turnover rate of oral epithelial cells compared to epidermal cells. Under the default parameters in the HE model, basal cells divide on average 0.4 times per week (Supplemental Figure B.3). However, basal cells in the oral epithelium and esophagus divide approximately twice per week [48, 164, 179]. We therefore rescaled time in our model such that each reported year of simulation time in the oral epithelium corresponds to 4.5 years of simulation time in the HE model. To reflect the higher density of cells in the oral epithelium versus epidermal tissue, tissue section sizes are determined using a density of 15,000 basal layer cells/mm² [29, 164], instead of 10,000 basal layer cells/mm² in the HE model [30, 196].

Gene Correction: To mimic *in vivo* microneedle gene delivery that corrects FA oral epithelial cells, we allowed FA cells to be transformed into corrected cells by exposure to a simulated transgene at a pre-specified treatment time. We apply the *in silico* transgene delivery to the center of simulated tissue sections except when otherwise noted at the start of the simulation. Transgene spread through the tissue is governed by an approximation to Brownian diffusion away from the injection site. Specifically, for each cellular position on the tissue basal layer, we determine the relative probability of correction by integrating the PDF of a bivariate Gaussian distribution over the cell's x and y coordinate boundaries on the lattice. We then sample k cell coordinates for correction from this probability distribution without replacement. To simulate a transgene injection with greater or lesser diffusion, we examine Gaussian distributions with different variances (D) centered on the injection site coordinates (x_{inj}, y_{inj}) as follows:

$$\mu = \begin{pmatrix} x_{inj} \\ y_{inj} \end{pmatrix}, \Sigma = \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix}. \quad (3.1)$$

Corrected cells can have altered dynamics in the basal layer that may allow them to spread, in analog to the ability of gene-corrected cells to spread in the hematopoietic compartment. In stratified squamous epithelium, clonal expansion occurs due to an imbalance in the fate of progenitor cells in the basal layer [29, 164, 166]. We model this mechanism via a persistence coefficient (p_{corr}), which prevents a cell from being displaced by its neighbor cell's division. Specifically, when a cell divides and randomly chooses to displace a neighboring cell possessing a persistence coefficient p_{corr} , the neighboring cell will not be displaced vertically with probability p_{corr} , and will be displaced vertically with probability $1 - p_{corr}$. If the neighboring cell is not displaced, the dividing cell instead places its daughter cell vertically and does not displace any neighboring cells. This mirrors the implementation of *NOTCH1* mutations in Schenck et al. [30]. FA cells are assumed to have

$p_{corr} = 0$ throughout.

3.5.2 *Interrogating clonal expansion of corrected cells in an FA background*

To investigate the spread of gene correction through the oral epithelium, we simulated FA tissue sections of size 0.67 mm^2 (100 cells x 100 cells). We introduced a single corrected patch of k cells with persistence coefficient p_{corr} in the center of the basal layer of the tissue as described above and recorded tissue states longitudinally for up to 50 years at six month intervals. Unless otherwise noted, $k = 10$ and $D = 2$, reflecting a tightly clustered patch of few corrected cells. Simulations were stopped if no corrected cells remained in the basal layer ('loss') or if the corrected patch of cells occupied 80% of the basal layer ('confluence'). Simulations reaching neither endpoint in 50 years were designated as 'ongoing'. We ran 100 correction replicates for $p_{corr} = (0, 0.001, 0.01, 0.1, 0.2, 0.5, 1.0)$ and recorded the proportion of simulations in which correction was lost or reached confluence and the time at which that event occurred. We repeated these analyses examining different numbers of corrected cells ($k = 3, 10, 30$) and different degrees of transgene diffusion ($D = 2, 10, 20$). Clone size distributions were determined by computing the number of cells associated with each of the corrected patches for each persistence coefficient after one year. To determine the expansion rates of corrected patches, we examined only simulation replicates reaching confluence and converted the number of corrected cells at each timepoint (six-month intervals) into area assuming each cell was $66.7 \text{ } \mu\text{m}^2$ [29, 164].

3.5.3 *Lineage competition between gene-corrected cells and TP53 mutants*

We investigated how gene correction might decrease the prevalence of *TP53* mutant cells in the oral epithelium of people with FA. To run these analyses, we expanded our model to permit cells to acquire *TP53* inactivating mutations at cell division in both *FANC*⁻ (FA)

and $FANC^+$ (corrected) cells. Each cell division, each $FANC^-$ and $FANC^+$ cell acquires an expected number of $TP53$ inactivating mutations, μ_{FANC^-} and μ_{FANC^+} , respectively. We determined the probability of $TP53$ inactivation to be $7.48 \cdot 10^{-7} \text{division}^{-1}$ as the product of the normalized genewide $TP53$ mutation rate $2.99 \cdot 10^{-6} \text{division}^{-1}$ [30, 197] and the probability of gene inactivation conditional on mutation as determined from deep mutational scanning data (~ 0.25 , [198]). More information on this calculation is available in ‘Parametrizing the proliferative advantage and mutation rate of $TP53^-$ mutations’. A Poisson number of inactivating mutations with $\lambda = \mu_{FANC^-}$ or $\lambda = \mu_{FANC^+}$ is drawn for each cell division and cells gain a fitness advantage if an inactivating mutation occurs. Acquiring multiple mutations in $TP53$ did not alter clonal fitness beyond a single hit (see Discussion). $FANC^-$ and $FANC^+$ cells containing a $TP53$ inactivating mutation had an additional increase in their persistence coefficient of $p_{FANC^-}^{TP53^-}$ and $p_{FANC^+}^{TP53^-}$, respectively. Initially, we assumed that $p_{FANC^-}^{TP53^-} = p_{FANC^+}^{TP53^-} = 0.01$, based on model fitting to the prevalence of $TP53$ mutations in noncancerous tissue (see section ‘Parametrizing the proliferative advantage and mutation rate of $TP53^-$ mutations’ for full details).

We considered three experimental conditions in which gene correction could slow the rate at which $TP53$ mutations accumulated in tissue:

1. If $FANC^+$ cells were more difficult to displace than $FANC^-$ cells, but were otherwise identical in terms of the $TP53$ mutation rate and clonal expansion speed. In this case, we examined simulation runs in which $FANC^- = FANC^+$ and $p_{FANC^-}^{TP53^-} = p_{FANC^+}^{TP53^-}$.
2. If $FANC^+$ cells were more genetically stable than $FANC^-$ cells and thus acquired fewer $TP53$ mutations. In this case, we examined simulations in which $\mu_{FANC^-} = m \cdot \mu_{FANC^+}$ and $p_{FANC^-}^{TP53^-} = p_{FANC^+}^{TP53^-}$, where $m = (1.5, 2, 4, 8)$.
3. If $FANC^-$ cells permitted a faster rate of clonal expansion of $TP53$ inactivating

mutations than $FANC^+$ cells. In this case, we examined simulations in which $\mu_{FANC^-} = \mu_{FANC^+}$ and $p_{FANC^-}^{TP53^-} = r \cdot p_{FANC^+}^{TP53^-}$, where $r = (1.5, 2, 4)$.

We conducted 300 replicates of experimental condition 1 and 100 replicates of experimental conditions 2 and 3, both with and without gene correction. Among simulations with gene correction, we examined small and large persistence coefficients for corrected cells ($p_{corr} = 0.01$ and $p_{corr} = 0.1$, respectively). We ran simulations on 0.33 mm^2 (70×70 cells) tissue sections for up to 50 years with $k = 30$ initially corrected cells distributed with $D = 2$. Simulations without correction ran uninterrupted for the full 50 years. Simulations with correction were run for up to 50 years, but were stopped if all corrected cells were lost from the basal layer. This stoppage criterion reflects our assumption that once all corrected cells are lost from the basal layer, the tissue behaves similarly to one never having undergone correction. To address variation in replicate numbers caused by early stoppage, mutational characteristics for samples in which gene correction was lost were sampled from a random replicate that did not undergo correction at each timepoint. We then quantified the proportion of the tissue with at least one $TP53$ mutation at six month intervals for up to 46 years.

3.5.4 Parametrizing the proliferative advantage and mutation rate of $TP53^-$ mutations

To quantify how gene correction could disrupt $TP53$ mutation clonal expansions in FA oral epithelium, we first parameterized $TP53^-$ mutation rates ($FANC^+$) and proliferative advantages ($p_{FANC^+}^{TP53^-}$) in the context of our model. We estimated these parameters via comparison to a published study of $TP53$ mutational expansion in healthy (i.e., non-FA) esophagus [2], another non-keratinized stratified squamous epithelium that is spatially-contiguous with the oral mucosa.

Martincorena et al. [2] collected 2 mm^2 sections of esophageal tissue from study par-

ticipants of different ages and deep sequenced each section to measure the number of unique *TP53* mutations and their corresponding variant allele frequencies (VAFs). Because 2 mm² tissue sections are computationally expensive to simulate under the HE model, we simulated smaller tissue sections (70 x 70 cells or ~ 0.33 mm²) and compared them to a spatially downsampled version of [2], intended to mimic results from Martincorena et al. if smaller sections were sequenced.

To approximate downsampling of the clinically-derived 2 mm² samples, we first spatially reconstructed possible tissue sections that reflect the mutation data observed in each full 2 mm² sample (i.e., each tissue section at a given age has i distinct *TP53* mutations indexed 1, ..., i , at frequencies $f_1, \dots, f_i$). With our expected cellular density, each 2 mm² tissue section is represented by a 173 x 173 cellular grid. We then place i mutations in cells on the grid such that each mutation is spatially contiguous and at its observed frequency, f_i . Specifically, starting with the most frequent mutation, we choose a random starting point on the grid to assign the mutation and then mutate a random neighboring cell of the currently mutated cells to extend the spatially contiguous mutational patch until the mutation reaches its observed frequency f_i . This process is repeated serially for each additional mutation observed in a tissue section, with the additional constraint that if the frequency of total mutations reaches 0.5 or more, the mutation that would exceed this threshold is instead assigned to an existing mutant clone background of higher frequency. After reconstructing potential tissue sections that could produce the correct number of *TP53* mutations and their VAFs in the 2 mm² sections, we randomly sampled a 0.33 mm² (70 x 70 cell) square subsection from each larger reconstructed sample and counted the number of mutations and their VAFs in this subsection. To align with Martincorena et al. [2], who excluded mutations sampled at frequencies below 0.0018 (108 genomic copies/(29929 cells * 2 genomic copies/cell)) in 2 mm² tissue sections, we applied the equivalent threshold in our 0.33 mm² sections, corresponding to a frequency 0.011 (i.e.,

108 genomic copies/(4900 cells * 2 genomic copies/cell)).

To identify the *TP53* persistence coefficient value that most accurately reflects the mutation patterns observed in empirical data, we simulated 70x70 tissue sections across a range of p_{FANC+}^{TP53-} values and compared the simulated data with the downsampled empirical observations. Martincorena et al. [2] collected 844 samples from nine donors of varying ages. We examined the 560 samples from the six donors aged 20 to 55 to match the correction timescales profiled in this study. 177 samples contained detectable *TP53* mutations, which were spatially reconstructed and then downsampled as described above. Only *TP53* nonsense, missense and synonymous SNV mutations were included, with non-SNV and splice mutations excluded. Total empirical mutational prevalences (i.e., the number of unique *TP53* mutations per tissue) were computed by combining the downsampled tissue sections and the tissue sections containing no *TP53* mutations.

We conducted 100 simulation replicates for each $p_{FANC+}^{TP53-} = (0, 0.007, 0.01, 0.0125, 0.015)$ and matched two metrics between the simulated and observed data: the number of unique *TP53* mutations in the tissue and the average mutational VAF of an observed *TP53* mutation. Using the default mutation rate of the HE model resulted in an excess number of *TP53* mutations compared to empirical data, even when these mutations conferred no proliferative advantage (Supplemental Figure B.4A). To better align the number of mutations in a tissue sample with observed data, we reasoned that the mutation rate producing mutations capable of expanding was likely lower than expected from the gene length and average mutation rate alone. To compute an adjusted mutation rate, we referenced deep mutational scanning data of *TP53* and found that approximately 25% of *TP53* mutations were expected to confer a selective benefit [198]. We therefore reduced the model's default expected number of *TP53* mutations by a factor of four, and found that this reduced mutation rate ($7.48 \cdot 10^{-7} \text{division}^{-1}$) was capable of producing mutation counts and sizes that roughly matched Martincorena et al. [2], as shown in Supplemental Figure B.4B.

To compare the number of distinct mutations arising, we calculated the squared differences between the mean number of mutations in the downsampled empirical data and the mean number of mutations in the simulated data for each timepoint and for each p_{FANC+}^{TP53-} value. We then averaged these squared differences across all timepoints and identified which p_{FANC+}^{TP53-} value minimized the mean squared error (Supplemental Figure B.4D).

To find the p_{FANC+}^{TP53-} value best able to produce empirical *TP53* mutational size distributions (i.e. VAFs), we took a two step approach. First, we fit exponential distributions that described the simulated VAFs for each p_{FANC+}^{TP53-} value separately at each timepoint. We fit these distributions using the R `fitdist` function within the `fitdistrplus` package [199]. Before fitting, we recentered the simulated VAFs by subtracting off the lower limit of detection frequency (0.011) to better conform to the shape of the exponential distribution (i.e., bounded below at 0). Second, we then used the best fit at each time point and for each simulated p_{FANC+}^{TP53-} to compute the probability that an interval encompassing each empirical downsampled VAF measurement was drawn from each of the best fit distributions. Specifically, for an observed VAF measurement f and best fit parameter λ for a given age group and p_{FANC+}^{TP53-} , we computed the probability of the observed data as:

$$p(f|\lambda) = \int_{f-\epsilon}^{f+\epsilon} \lambda e^{-\lambda x} dx \quad (3.2)$$

We used $\epsilon = 0.0002$ although we found that results were robust to our choice of small ϵ . As when fitting the distributions initially to find the best fit λ s, we recentered the observed VAFs by subtracting off the lower limit of detection before evaluating the probability expressions above. We then summed the log probabilities for a given p_{FANC+}^{TP53-} across all age groups to find the p_{FANC+}^{TP53-} that generates simulated data that best conforms to the empirical data through time (Supplemental Table 2, Supplemental Figure B.4CE). We did not perform exponential best fit calculations to $p_{FANC+}^{TP53-} = 0$ as there was only one ob-

served mutation for the age ranges 20-23 and 24-27. Two outlier measurements diverging substantially from the overall temporal signal were excluded when determining the best fit parameters: the number of mutations in a heavy smoker aged 44-47 and the variant allele frequencies among ages 20-23 (Supplemental Figure B.4BC).

We found that $p_{FANC+}^{TP53-} = 0.01$ both minimized the squared error for the number of mutations and minimized the negative log-likelihood fit to the mutational VAFs, so we proceeded with $p_{FANC+}^{TP53-} = 0.01$ as our proliferative advantage of *TP53* mutations in healthy tissue.

Due to interactions between the FA and *TP53* pathways, loss of *TP53* function in cells already carrying *FANC* mutations is likely to confer a more extreme growth phenotype (i.e., $p_{FANC-}^{TP53-} > p_{FANC+}^{TP53-}$, see Discussion). To parameterize p_{FANC-}^{TP53-} , we drew on both *in vitro* and *in vivo* data from human lymphoblastoid and mouse HSC experiments [74, 172] which provide consistent, quantitative evidence that *TP53* inactivation enhances proliferation to a greater degree in FA cells than non-FA cells:

1. FA and non-FA lymphoblastoid cell lines, both with and without *TP53* knockdown, were exposed to a pulse of interstrand crosslinking agents then grown in culture for 72 hours [74]. In non-FA cells, *TP53* knockdown increased growth by 17%, but in FA cells, *TP53* knockdown increased growth by 40%. These values were extracted from time point H72 in Figure 3C of the original publication [74] using WebPlotDigitizer [200]. This suggests $p_{FANC-}^{TP53-} \approx 0.40/0.17 \cdot p_{FANC+}^{TP53-} = 2.35 \cdot p_{FANC+}^{TP53-}$.
2. HSCs from mice constructed with all combinations of *Fanca*^{+/+} or *Fanca*^{-/-} and *Tp53*^{-/-} or *Tp53*^{+/+} were tracked for 20 weeks [172]. The depletion of HSCs was measured to track proliferative exhaustion due to more rapid cycling. Addition of *Tp53*^{-/-} to a *Fanca*^{+/+} background resulted in a decrease in abundance of 35% between weeks 8 and 20, while addition of *Tp53*^{-/-} to a *Fanca*^{-/-} background re-

sulted in a decrease in abundance of 86% between weeks 8 and 20. These values were extracted from Figure 1D of the original publication [172] using WebPlotDigitizer [200]. This suggests $p_{FANC-}^{TP53-} \approx 0.86/0.35 \cdot p_{FANC+}^{TP53-} = 2.45 \cdot p_{FANC+}^{TP53-}$, although provides less direct evidence than the growth experiments described above.

The consistency between these two studies, in different cell types from different organisms, prompted our use of an increased proliferative advantage (i.e., r) in the range of ~ 2.4 . Specifically, we tested $r = (1.5, 2, 4)$ to explore potential increases or decreases in proliferative advantage that may accompany an epithelial context.

3.5.5 Code availability

The code to reproduce these analyses can be found at

https://github.com/federlab/HomeostaticEpithelium_FAcorrection.

Chapter 4

CONCLUSIONS

4.1 Importance of predicting malignant transformation and early cancer detection

Early cancer detection is one of the most effective means to improve survivability across human cancers, yet up to 50% of cancers are not detected until a late stage [201, 202].

Although imaging based detection methods have improved the ability to detect smaller lesions, they remain insufficiently sensitive for the earliest lesions and population based screening at intervals necessary for tracking are unfeasible for most cancers [201, 202].

However, because cancer progresses through a gradual evolutionary process, there remains an opportunity to identify molecular evidence prior to or at the earliest stages of neoplastic growth and enable therapeutic intervention. Translating this opportunity into clinically useful biomarkers remains a central challenge of the field. A key early event during tumorigenesis is the clonal expansion of a pathogenic mutation, however, the early detection of pathogenic mutations alone is an inadequate indication of malignant progression. Sequencing across normal tissues such as blood, skin, esophagus, colon, and uterine lavage has found that cancer-associated mutations are widespread and can persist for decades, but malignant transformation of those mutant clones is comparatively rare [1, 2, 26, 46–48]. This suggests that early detection methods must extend beyond somatic driver de-

tection and account for the relevant genetic and microenvironmental context in which the somatic mutations arise and spread. Progression risk can be shaped by the background genetic context [203], tissue architecture [33], epigenetic mechanisms [204–208], and quantifying pre-malignant cancer risks will require understanding the interplay of these and other factors. Cancer predisposition syndromes, where a germline genetic variant is fixed across all cells, offer a powerful system to improve the tractability of understanding these interactions.

4.2 Major goals for the study of cancer predisposition syndromes

Inherited cancer predisposition syndromes (CPS) provide a unique window to understand how genetic variation influences the progression of cancer as all cells share the same pathogenic variant amidst an otherwise heterogeneous somatic background. As a result, CPSs have served as foundational model systems for cancer biology that often generalize to sporadic cancer in the general public [62]. One important example comes from *BRCA1/2*-associated hereditary breast and ovarian cancer syndromes, where studies of *BRCA1* and *BRCA2* linked inherited cancer risk to impaired homologous recombination mediated DNA repair [63–65]. This mechanism provided the rationale for synthetic lethal targeting of *BRCA*-deficient tumors with PARP inhibitors, now a common biomarker guided therapeutic strategy across *BRCA*-mutant and homologous recombination deficient cancers [63–65]. Although cancer predisposition syndromes serve as a generalizable model for sporadic cancer progression, there is an urgent clinical need to develop therapeutic prevention strategies to address the lifelong elevated cancer risk compared to the general population. This need is particularly heightened in syndromes where the same germline defect increasing cancer risk also limits the therapeutic options after malignancy develops, which can require reductions in dose intensity or avoiding specific therapies entirely. Individuals with Fan-

congenital Anemia poorly tolerate DNA-damaging chemotherapy and radiation, and those with Li-Fraumeni Syndrome face elevated risks of therapy-related secondary malignancies [71, 75–77, 146, 209]. These constraints highlight the importance for developing interventions to suppress cancer risks. Bringing the future scientific and clinical potential of investigating CPSs to fruition will include two important research directions: 1) How does an inherited pathogenic variant alter the mutagenic and selective forces acting across contexts (i.e., different tissues, epigenetic and transcriptomic states, cytotoxic exposures, immune microenvironment)? 2) Can these signals be leveraged for predicting malignant progression and be therapeutically targeted?

4.2.1 Background specific effects across contexts

What was learned and impact to the field: Chapter 2 of this dissertation outlines the collective effort from myself and collaborators to investigate somatic evolution within Li-Fraumeni Syndrome (LFS), a cancer predisposition syndrome caused by germline *TP53* pathogenic variants. LFS serves as one of the most broadly impactful CPSs as *TP53* is the most widely mutated gene across all human cancers [56]. My work specifically conducted somatic genetic profiling of a family with a germline *TP53* p.R181H variant, one of the most prevalent germline variants characterized in LFS [79, 124, 125]. This work demonstrated that a germline p.R181H *TP53* variant increases mutagenesis in regions of the genome not expected to be under positive or negative selection in blood compared to non-carriers, providing direct evidence that background genetic variation shapes the somatic evolutionary process. This aligns with findings in other CPSs such as Shwachman-Diamond syndrome [210] and other germline variants associated with increased acute myeloid leukemia risk, where germline variants altered the somatic mutation selection landscape in blood [203, 211–213].

Additionally, this project extends our knowledge of how germline genetic variation im-

pacts somatic evolution across multiple tissue contexts and presents the first, to my knowledge, multi-tissue ultra-deep sequencing based characterization of non-cancerous tissues in LFS. Characterizing somatic evolution across pre-malignant solid tissues within CPSs has been relatively limited, but has been informative for determining if inherited risk leaves detectable evolutionary signatures prior to cancer development [62, 81, 84, 85]. Whole genome sequencing of normal intestinal crypts of individuals carrying germline variants in *MUTYH* had elevated somatic mutations in normal tissue by a factor of greater than 2-fold compared to crypts without the germline variant and had characteristic mutational signatures associated with the variant [84]. Conversely, sequencing of colorectal crypts in Lynch Syndrome found similar mutational processes and mutation rates to individuals without Lynch Syndrome, but identified DNA mismatch repair deficiency as a potential pre-malignant biomarker [85]. While these studies demonstrate the value of characterizing somatic evolution in pre-malignant tissues, the field has largely focused within the dominant tissue at risk within a given syndrome. Many CPSs increase cancer risk across multiple organs and additional multi-tissue studies are needed to determine whether inherited pathogenic variants reshape somatic evolution in tissue dependent ways.

We found that tissue specific contexts were a primary determinant of the mutational patterns found across LFS solid tissue samples. Mutational signatures differed by tissue, particularly in skin and blood, and cancer sample signatures were consistent with sporadic cancers of similar tissue origin. This suggests that germline *TP53* p.R181H germline variants do not universally impose detectable mutational signatures across tissues and instead the local exposures of the tissue environment are largely responsible for mutagenic patterns. Our results suggest that the germline variant may provide a stronger influence on the selective pressures governing clonal expansion than the mutational processes themselves.

Interestingly, we showed evidence that the p.R181H germline variant altered selec-

tive patterns tissue-wide in the LFS individual including preferential mutation on the WT *TP53* allele and the dominant-negative hotspot (p.R248Q/W) across most tissues. The repeated emergence of p.R248Q/W is notable, considering the large number of other inactivating mutations that are possible, including other hotspots with similar dominant-negative properties. Although the mechanism underlying this preference is unresolved, future approaches such as computational protein modelling of multimeric complexes containing different mutation combinations may help explain the preferential selection in *TP53* and may provide a framework for understanding the selection of second-hits across other CPSs. While this result needs to be investigated beyond the single individual in this study, it suggests more broadly that specific CPS-associated germline variants may reshape clonal evolution and constrain the evolutionary trajectory of second-hit mutations across multiple contexts.

Interesting challenges and opportunities remaining: Multiple important challenges remain unresolved in this study specifically, and in the general approach of using clinical samples to understand context-dependency in people with an inherited driver mutation.

1. Multi-tissue characterization was performed in a single LFS individual and without comparable non-LFS controls sequenced at comparable duplex sequencing depths, there is a limited ability to distinguish the contribution of the germline variant from the tissue-specific somatic processes that occur with age. The most direct path to resolve this is by expanded cohort sequencing of multiple LFS and non-LFS individuals that span across multiple *TP53* germline variants. Additionally, the multi-tissue sequencing was primarily focused on *TP53*, which limits the ability to detect other important driver genes that may be under selection in LFS. Particularly useful genes to investigate are those involved in Wnt and PI3K/AKT signaling, and homologous recombination due to their prevalence to arise somatically in LFS tumors [81].

More broadly, the limitations of compiling extensive study cohorts of rare CPSs are partially alleviated with ongoing CPS registries and biobanking efforts [214], along with the NIH SMAHT network, which aims to generate a reference of somatic evolution across normal tissues in the general population [215], and will serve as a non-CPS resource where evolutionary patterns can be compared. Separately, the use of transgenic mouse models of CPS offers advantages that human cohort studies cannot yet provide. In mice, subject numbers can be scaled, longitudinal sampling becomes more feasible, and confounding variables can be controlled in ways not possible in human studies. Additionally, mouse models allow the use of emerging technologies such as lineage recording systems to reconstruct clonal phylogenies [216] and potentially record temporally ordered cell signaling states during pre-malignant tissue maintenance [217, 218].

2. Microenvironmental tissue characterization of LFS normal tissues, such as altered transcriptional or epigenetic states may inform clonal expansion and transformation risks. Liquid biopsy studies in LFS have demonstrated that LFS individuals have a distinguishable methylation status in cell-free DNA compared to the general population, suggesting that germline *TP53* variation alters the environment in ways that could reshape the selective conditions for additional somatic mutations [219, 220]. However, these blood-based signatures reflect cells in aggregate and cannot resolve any spatially heterogeneous signatures within a given tissue. Spatial transcriptomic and epigenomic platforms [221] offer opportunities to bridge this gap by co-localizing genetic clonal expansions with specific tissue niches in which they arise. This approach could provide the resolution to ask whether specific tissue states alter selection in distinct regions that cannot be represented by bulk sequencing alone.

More broadly, microenvironmental differences may provide insight into whether spe-

cific altered epigenetic or transcriptomic states favor the expansion of LOH clones over point mutations, as LOH is the predominant second-hit in LFS and other tumor suppressor driven CPS tumors [48, 81]. Testing this possibility will require more sensitive approaches to detect somatic LOH events at low frequencies and preserve tissue context information. Emerging spatial technologies that could enable this investigation include bioinformatic advances that complement spatial transcriptomic methods by inferring copy number alterations while preserving the tissue gene expression niche [222, 223]. Though long-read sequencing does not yet have the sensitivity to detect small clones in bulk tissue, it can resolve larger structural variants, variant phasing, and copy number architecture beyond what is achievable by short-reads [224–226].

4.2.2 Therapeutic considerations

What was learned and impact to the field: Chapter 3 of this dissertation outlines my effort along with my collaborators to develop a computational framework for therapeutic oral cancer prevention in Fanconi Anemia, a CPS caused by germline variants in the FA pathway. FA is an important model for preventative intervention because affected individuals have a several hundred to over thousand fold greater risk of developing head and neck squamous cell carcinomas compared to the general public and standard cytotoxic therapies are often intractable due to the underlying genetic defect [71, 75, 77, 146, 151, 209].

Gene correction as a preventative approach in CPSs has primarily shown effectiveness by ex vivo correction in the hematopoietic system, including Wiskott-Aldrich syndrome [227, 228], and in FA [112, 113], where lentiviral corrected cells demonstrate a proliferative advantage and repopulate the hematopoietic compartment. Extending gene therapy strategies into solid tissues, however, poses unique challenges, as corrected cells will be

subject to the local cellular competition within structured tissue architectures not present in blood.

Using the agent-based modelling approach of in vivo gene therapy in FA epithelium outlined in Chapter 3, we found that epithelial competition is the central determinant of therapeutic success, where only gene corrected lineages with strong proliferative advantages manage to avoid stochastic loss via displacement of neighboring uncorrected cells. While studies of clonal competition in a therapeutic context of an inherited disorder are limited, this finding is consistent with other modelling approaches of normal somatic evolution, where the extent of pre-malignant clonal expansion was susceptible to tissue architecture and spatial constraints [29, 30, 33, 35]. We further found these constraints could be mitigated through the design of microneedle array delivery platforms, where the spatial configuration of initial correction locations influenced cell spread and corrected coverage potential, providing a quantitative framework for optimizing delivery geometries that has the capability to generalize to other tissue applications.

We also find that corrected epithelial patches with strong proliferative advantages reduce the elevated accumulation of *TP53* mutations in FA to levels aligned with the general population, demonstrating that gene correction has the potential to not only restore gene function, but realign the evolutionary trajectory of cancer-associated mutations in an at risk tissue. This broadly encourages future therapeutic strategies to consider targeting the tissue specific environmental factors shaping selection of clonal expansion in addition to the genetic underpinnings of disease.

Interesting challenges and opportunities remaining: Several important challenges remain to advance the development of preventative therapeutic strategies in CPSs.

1. Unlike the hematopoietic system where ex vivo correction and re-infusion provide direct correction of the target cell population, in situ correction of solid tissues requires

gene delivery systems to perfuse into the target cells within the native tissue context and are thus not universally applicable across tissues. As described in Chapter 3, absorbable microneedle arrays are well suited for clinically accessible mucosal epithelium such as the oral cavity, but more broad application to less accessible tissues will require delivery methods tailored to anatomical sites. Emerging examples such as inhalable gene therapy being developed for lung epithelia in Cystic Fibrosis [229] and oral delivered vectors for gastrointestinal epithelium in Lynch Syndrome and Familial Adenomatous Polyposis [230] are expanding the number of available target sites for correction and the continued development of tissue-specific delivery platforms will be broaden the number of CPSs eligible for preventative gene therapy.

However, successful prevention will depend not only on whether therapeutic approaches perfuse into the appropriate tissue, but if those corrected cells can persist and expand within that context. Corrected lineages modeled in the oral stratified squamous epithelium in Chapter 3 spread largely through lateral displacement of neighboring cells in the basal layer, whereas in the crypt organization within colorectal epithelium, clonal expansion occurs first within an individual crypt and expansion beyond the crypt requires additional mechanisms such as crypt fission [231–234]. Thus in the context of gene correction, high delivery coverage, repeated dosing, or mechanisms that promote processes other than stem cell renewal may be required for adequate prevention depending on the tissue site.

This creates a need for experimental and computational frameworks for evaluating the efficiency of gene therapy delivery and the clonal dynamics of corrected cells within tissue specific contexts. Mouse models of CPSs paired with technologies such as fluorescently labeled markers, lineage tracing, live cell imaging, and deep sequencing can be used to track therapeutic correction efficiency and how those corrected clones are retained and spread over time [179, 186, 216, 218]. These measurements

can then parameterize models like the framework developed in Chapter 3 to test parameters relevant to a specific CPS's tissue context such as delivery dosing and frequency, spatial distribution, and interactions with somatic pathogenic mutations at scales that are difficult to produce empirically. In this way, modeling can serve as a parallel framework to optimize therapy prior to advancing to additional mouse modeling or preclinical studies.

2. While gene therapy approaches offer exciting preventative opportunities for CPSs, combination with non-genetic forms of therapy may further increase efficacy as many CPSs have been associated with altered tissue biology beyond the direct function of the mutated gene in histologically normal epithelium [62, 84, 85, 219, 235]. FA provides a notable example, where distinct methylation signatures have been identified in plasma compared to non-FA individuals and these signatures persist in FA individuals with revertant mosaicism [235]. While it is unclear whether these methylation patterns impact malignant progression, it suggests that restoring FA pathway function may not fully restore all molecular consequences of the FA background.

More broadly, this supports the investigation to what therapeutic cancer prevention strategies can complement gene correction in a CPS context. A notable example is the daily use of aspirin to reduce colorectal cancer risk in Lynch syndrome by over 35% [66] without targeting the genetic defect directly. Other non-genetic approaches such as epigenetically targeted therapies provide one possible approach, where recent work has emerged in oncology by using small molecules to target mechanisms such as DNA methylation, histone modification, and chromatin reorganization [236, 237]. Although these therapies have primarily been applied to established malignancies, they may be useful earlier in cancer development if persistent epigenetic alterations contribute to CPS-associated risk. Similarly, variant specific small molecules such as

rezatapopt, which restores function to the p53 Y220C mutant protein, have shown promise reducing sporadic *TP53* Y220C mutant cancers [238]. This may offer future prophylactic relevance for LFS individuals harboring this variant and could be evaluated in LFS mouse model systems. Although variant specific small molecules are likely limited to those variants which produce stable and targetable proteins, they could expand preventative opportunities for tissue sites where gene replacement is difficult.

4.3 Concluding remarks

Identifying the molecular signals that foretell malignant progression amongst the noise of benign somatic evolution in normal tissues remains one of the central unresolved challenges in oncology. Progression toward this goal will require understanding how genetic background, somatic mutation, epigenetic state, and the tissue specific architecture and microenvironment collectively shape the selective landscape for cancer to progress and develop frameworks to calculate risk. Cancer predisposition syndromes, where a pathogenic germline variant is fixed across all cells, reduce one axis of this complexity, and expanding normal tissue characterization in these populations should be a priority. The mutagenic and selective remodeling revealed through molecular characterization can directly inform prevention strategies for CPS individuals and provide a generalizable framework for sporadic cancer progression.

Appendix A

SUPPLEMENTARY FIGURES FOR CHAPTER 2

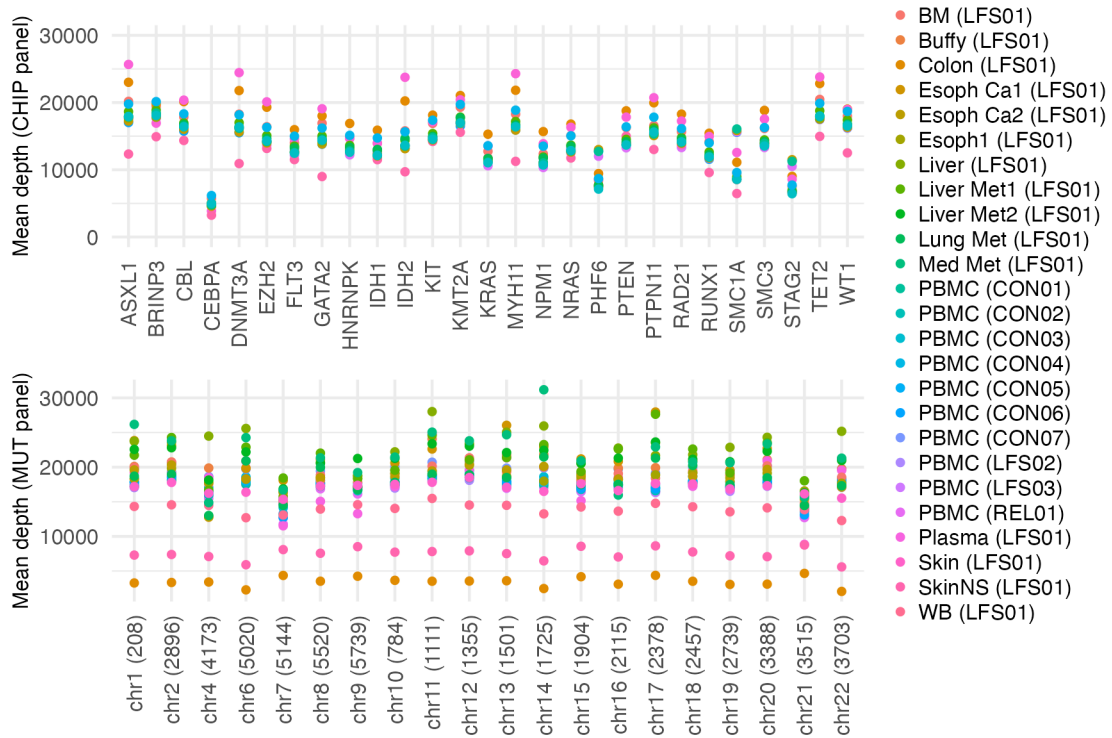


Figure A.1: Duplex depth comparison across genomic regions and samples. (A) Mean duplex sequencing depth of CHIP panel genes colored by sample. **(B)** Mean duplex sequencing depth of MUT panel regions colored by sample. **(C)** Mean duplex depth of *TP53* by sample. BM: bone marrow; WB: whole blood; PBMC: peripheral blood mononuclear cells; Esoph: esophagus; Gast: gastric tissue; SkinNS: non-sun-exposed skin; Med: mediastinal; Met: metastasis; Ca: cancer.

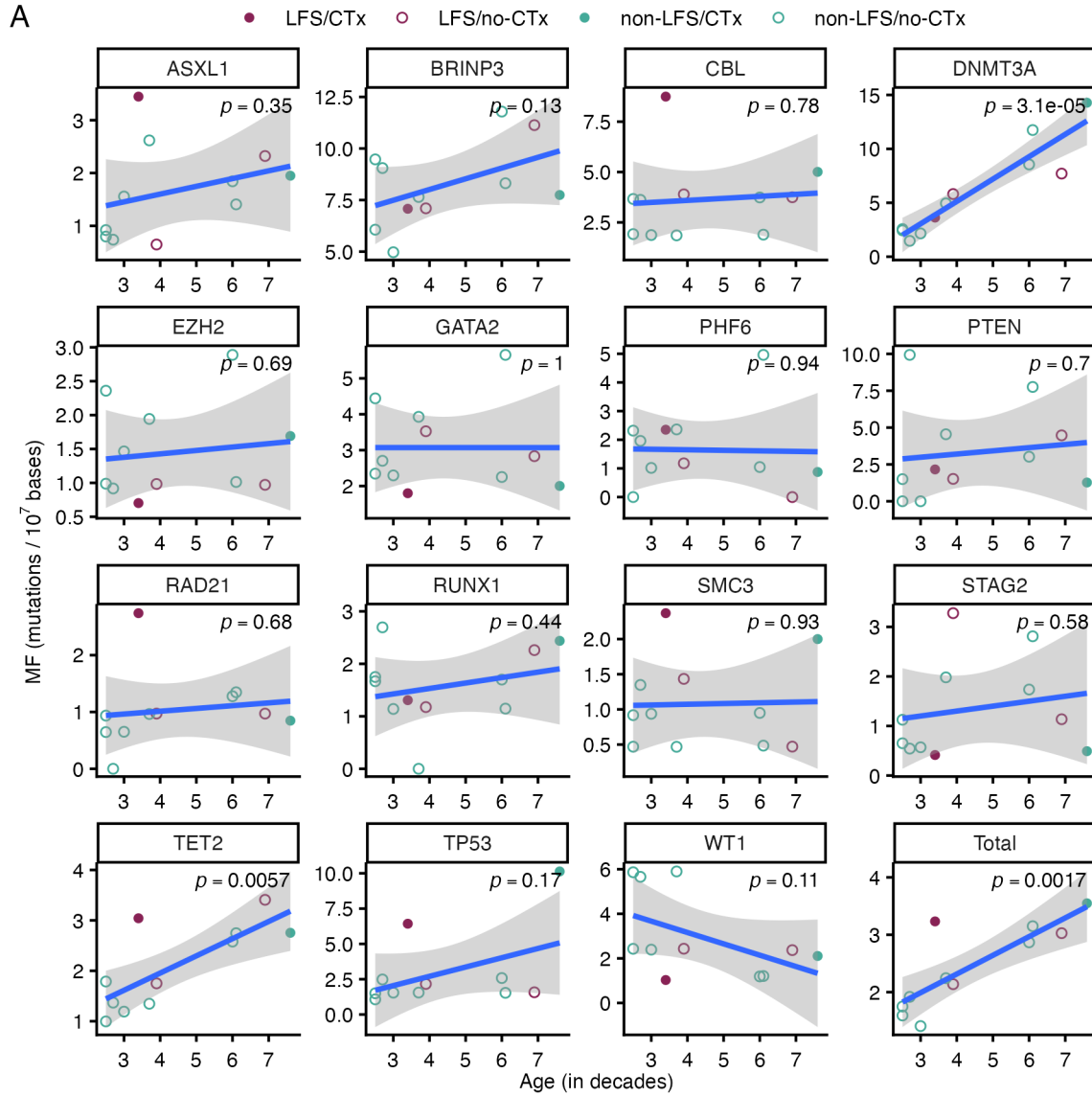


Figure A.2: CHIP coding regression of MF and age by gene. Linear regressions of coding mutation frequencies (MF) in CHIP genes by age. Circles are colored by LFS (maroon) and non-LFS (teal) status and filled circles indicate chemotherapy (CTx) treatment. Age-associated baseline regression (blue line) is fit using all individuals and linear regression p-value is indicated. Grey areas indicate the confidence intervals of the regression line. Mutation frequency is measured as mutations per 10,000,000 bp sequenced.

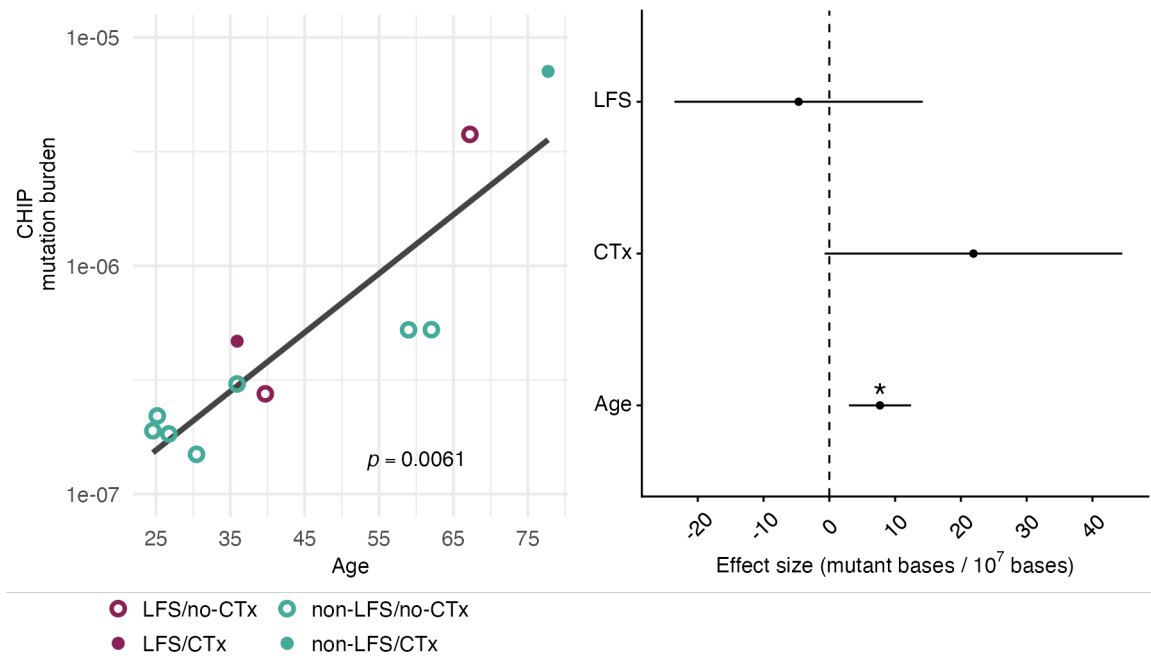


Figure A.3: CHIP coding mutational burden association with age, chemotherapy, and LFS. (A) Linear regressions of coding CHIP mutation burden (MB) with age. Circles are colored by LFS (maroon) and non-LFS (teal) status and filled circles indicate chemotherapy (CTx) treatment. Age-associated baseline regression is fit using all individuals. **(B)** Multiple regression of coding CHIP MB using age (scaled by decades), LFS, and CTx status as covariates. MB was scaled per 10,000,000 bases. Asterisks indicate significance value (* $p < 0.05$, ** $p < 0.01$) and error bars indicate 95% confidence intervals.

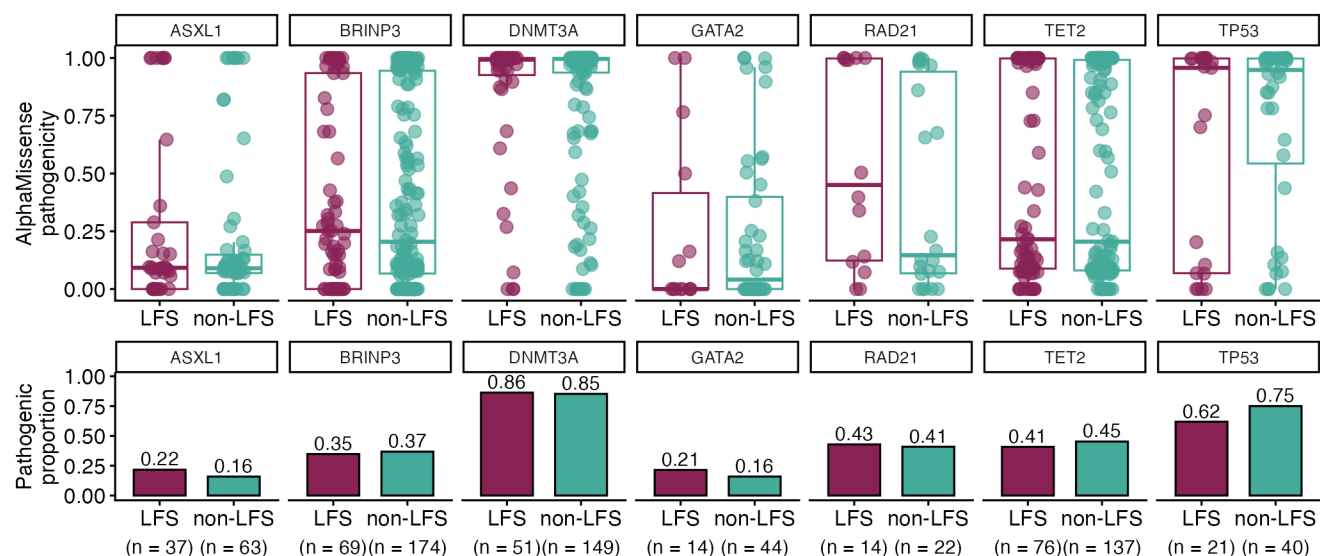


Figure A.4: Comparison of pathogenicity of blood mutations in CHIP/AML genes in subjects with and without *TP53* p.R181H germline mutation. (A) AlphaMissense pathogenicity scores of observed CHIP SNVs by germline p.R181H status (LFS: maroon; non-LFS: teal) within individual gene. Box plots indicate the median and interquartile range. **(B)** Percentage of pathogenic mutations from **A**. Pathogenicity is determined based on mutation type and AlphaMissense scores (see Methods).

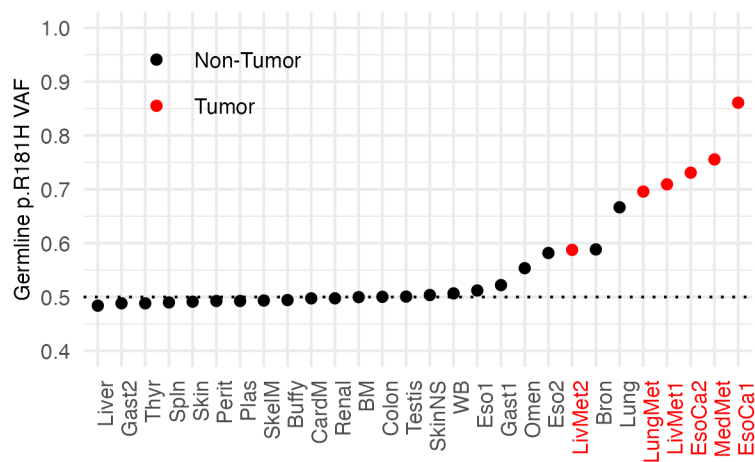


Figure A.5: Loss of heterozygosity (LOH) in Li-Fraumeni patient samples based on p.R181H allele frequency. Variant allele frequency (VAF) of the p.R181H germline variant across LFS01 samples. Cancer samples indicated in red text.

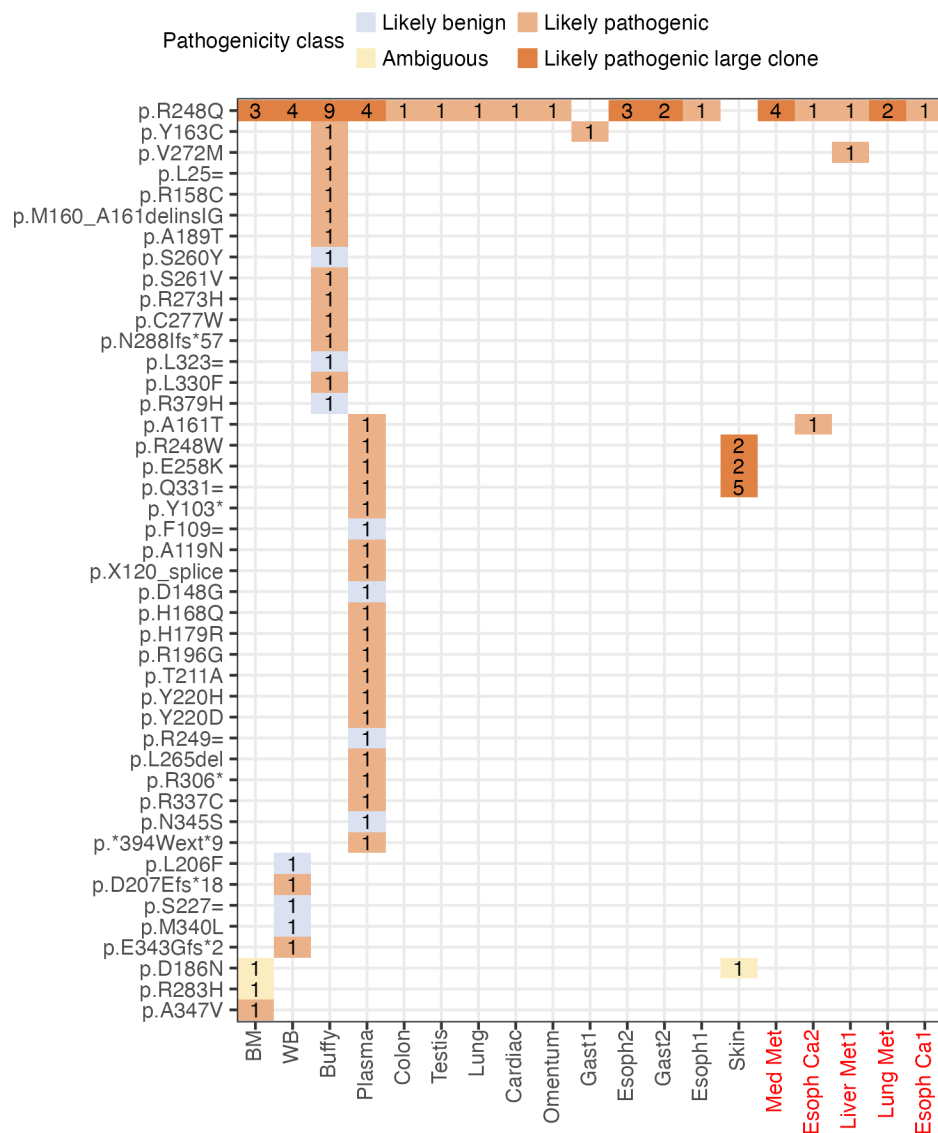


Figure A.6: Concordance of *TP53* mutations identified in blood samples and tissues from LFS01. The plot includes all *TP53* mutations identified in bone marrow (BM), whole blood (WB), buffy coat (buffy), and plasma. Box color is determined by AlphaMissense pathogenicity score. Number inside of each box represents the number of duplex reads associated with that given mutation. Mutations with a mutant duplex read count greater than 1 are deemed a large clone and are colored a darker orange or blue depending on pathogenicity status. Cancer samples indicated in red text.

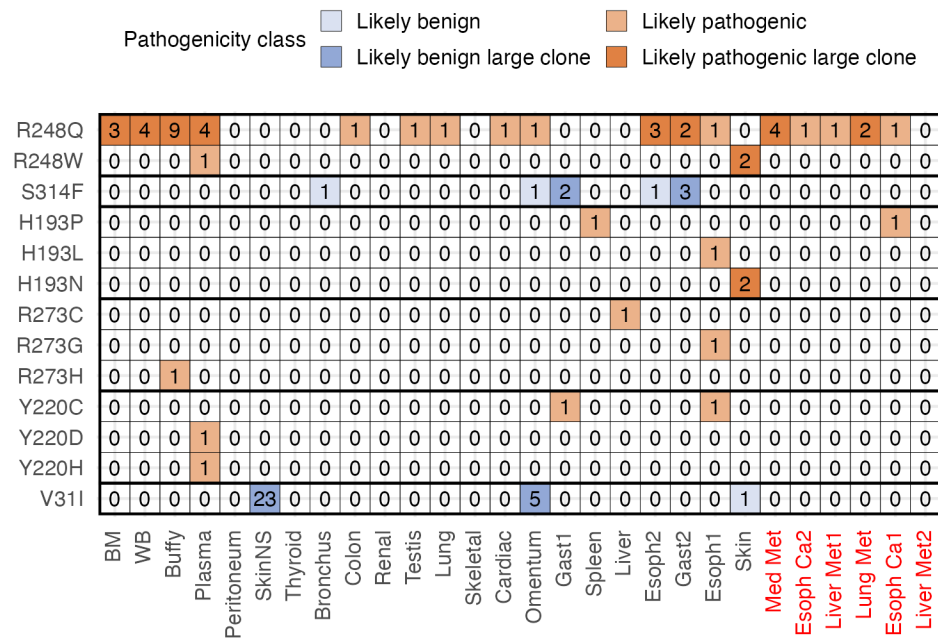


Figure A.7: Size and tissue distribution of *TP53* mutations observed in three or more samples. Box color is determined by AlphaMissense pathogenicity score (likely pathogenic: orange; likely benign: blue). Number inside of each box represents the number of duplex reads associated with that given mutation. Mutations with a mutant read count greater than 1 are deemed a large clone and are colored a darker orange or blue depending on pathogenicity status. Cancer samples indicated in red text.

Appendix B

SUPPLEMENTARY FIGURES FOR CHAPTER 3

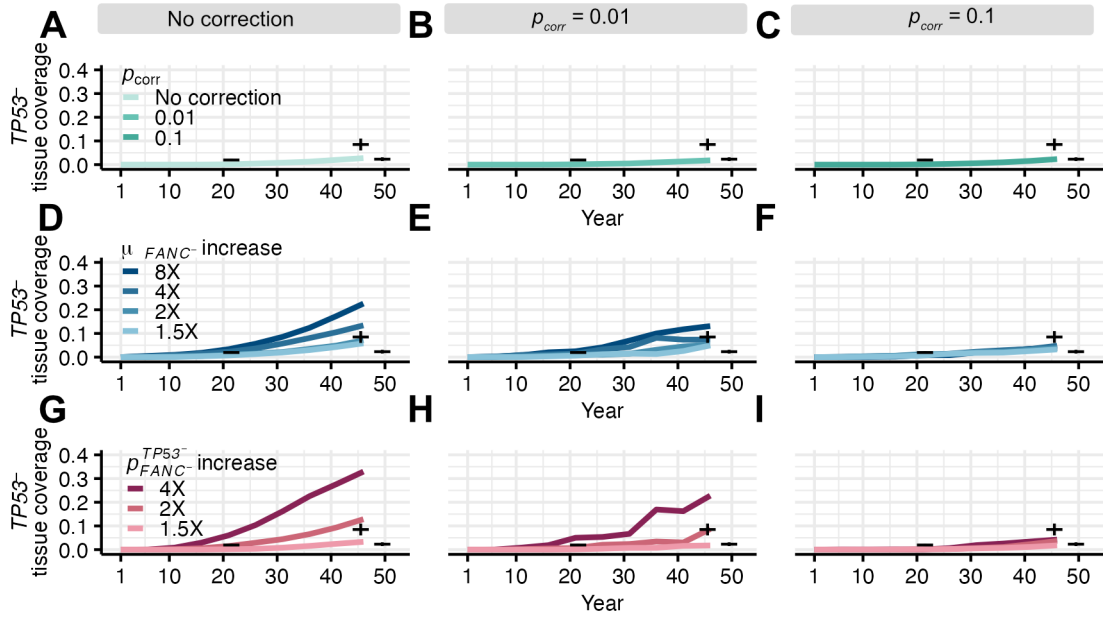


Figure B.1: Gene correction reduces spread of $TP53$ mutations through FA tissue over time. (A-I) $TP53^-$ tissue coverage tracked over time for 46 years on 0.33 mm^2 simulated tissue sections with or without single gene correction ($k = 30$ cells and $D = 2$). Columns represent conditions with no correction, $p_{corr} = 0.01$ and $p_{corr} = 0.1$. Rows represent the experimental conditions as described in Fig 4. Black crosses represent $TP53$ tissue coverage ranges from Martincorena et al. [2] normal esophagus, with vertical lines indicating the observed range. 300 simulations were performed for each condition within each persistence coefficient for **A-C** and 100 simulations were performed for **D-I**.

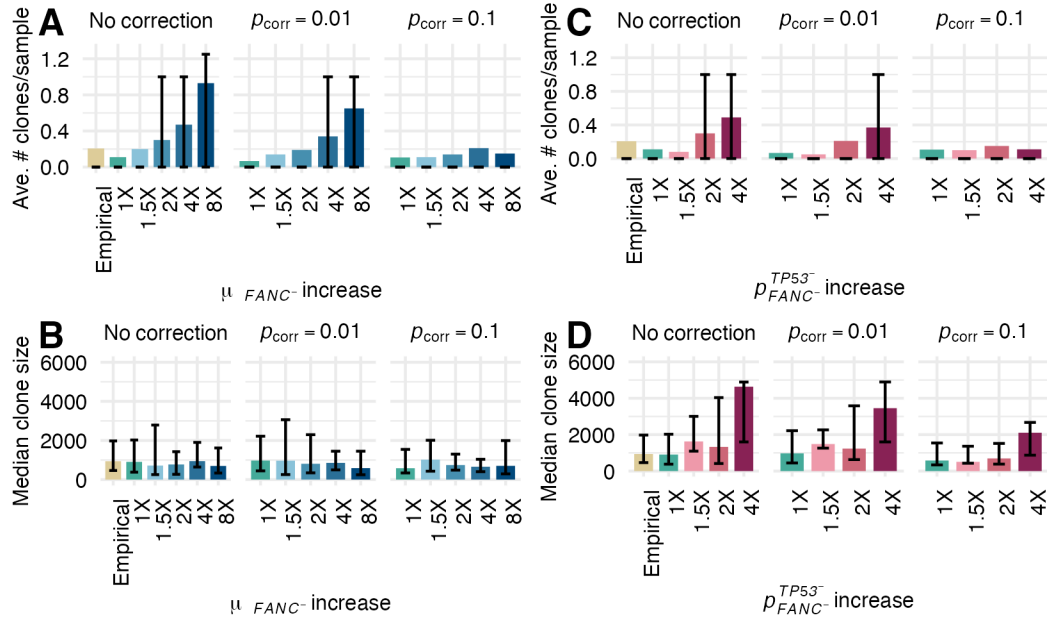


Figure B.2: Elevated mutation rates increase the number of $TP53^-$ clones, while increased persistence coefficients increase $TP53^-$ clone sizes. (A-D) The average number and size of $TP53^-$ clones after 46 years on 0.33 mm^2 simulated tissue sections with ($p_{corr} = 0.01$ or 0.1) or without gene correction. The 1X condition in each panel (green) represents simulations where $TP53$ mutation rates and persistence coefficients are equal across $FANC^-$ and $FANC^+$ cells. The empirical condition in each panel (yellow) represents data from downsampled normal esophageal tissue in ages 36-55 from Martincorena et al. [2]. 300 tissue simulations of the 1X condition and 100 tissue simulations of all other conditions were performed. Bars indicate interquartile ranges. **(A)** Average number of distinct $TP53^-$ clones when the $TP53$ mutation rate is elevated in $FANC^-$ cells ($\mu_{FANC^-} = m * \mu_{FANC^+}$, $m = (1.5, 2, 4, 8)$). **(B)** Median $TP53^-$ clone size under the same conditions as **(A)**. **(C)** Average number of distinct $TP53^-$ clones when $TP53$ persistence coefficients are elevated in $FANC^-$ cells ($p_{FANC^-}^{TP53} = r * p_{FANC^+}^{TP53}$, $r = (1.5, 2, 4)$). **(D)** Median $TP53^-$ clone size under the same conditions as **(C)**.

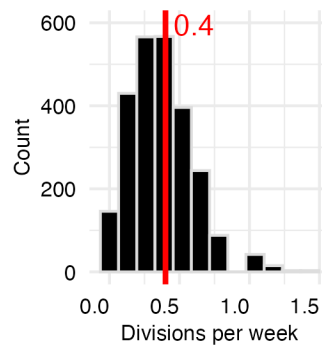


Figure B.3: Basal cell division rates in the default HomeostaticEpidermis model. Distribution of average cell division rates across basal layer positions in the HomeostaticEpidermis model over 100 model timesteps on a 50x50 cell grid. The red line indicates the mean division rate (0.4 divisions per week).

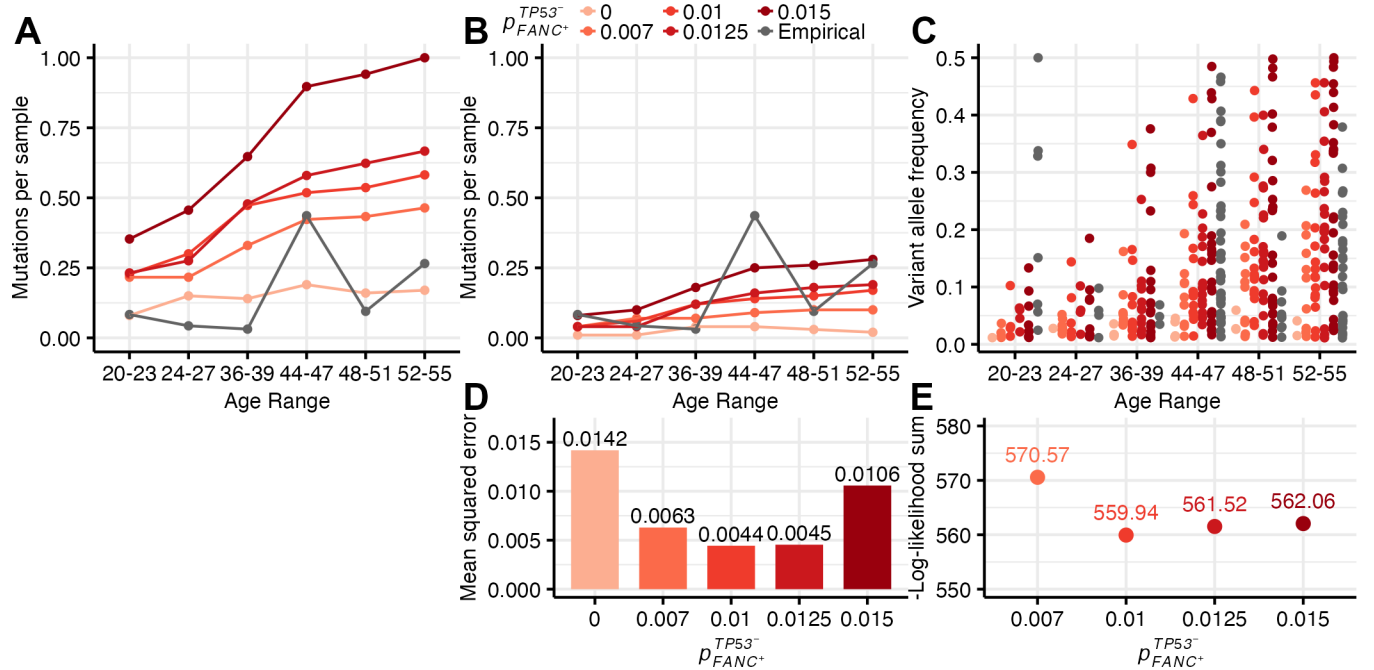


Figure B.4: Parameterizing the proliferative advantage and mutation rate of *TP53* mutations. (A) The number of *TP53* mutations per 0.33 mm² tissue sample in simulations with varying persistence coefficients ($p_{FANC+}^{TP53-} = 0, 0.007, 0.01, 0.0125, 0.015$) using the default mutation rate of the HomeostaticEpidermis model (colored by persistence coefficient), compared to empirical data (grey). 100 tissue simulations were performed for each age range and persistence coefficient. (B) The number of *TP53* mutations per sample in simulations with varying persistence coefficients using the calibrated mutation rate ($FANC^{+}$). (C) Variant allele frequencies of mutation meeting the lower limit of detection in simulated (colored by persistence coefficient) and empirical (grey) tissue. (D) Mean squared error between the simulated and empirical data for the number of *TP53* mutations across various p_{FANC+}^{TP53-} values. Age range 44-47 was excluded from the mean squared error computation. (E) Negative log-likelihood comparisons of empirical variant allele frequencies against simulated distributions across p_{FANC+}^{TP53-} values. Age range 20-23 was excluded from negative log-likelihood computation.

BIBLIOGRAPHY

- [1] Iñigo Martincorena, Amit Roshan, Moritz Gerstung, Peter Ellis, Peter Van Loo, Stuart McLaren, David C. Wedge, Anthony Fullam, Ludmil B. Alexandrov, Jose M. Tubio, Lucy Stebbings, Andrew Menzies, Sara Widaa, Michael R. Stratton, Philip H. Jones, and Peter J. Campbell. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, 348(6237):880–886, May 2015. doi: 10.1126/science.aaa6806.
- [2] Iñigo Martincorena, Joanna C. Fowler, Agnieszka Wabik, Andrew R. J. Lawson, Federico Abascal, Michael W. J. Hall, Alex Cagan, Kasumi Murai, Krishnaa Mahbubani, Michael R. Stratton, Rebecca C. Fitzgerald, Penny A. Handford, Peter J. Campbell, Kourosh Saeb-Parsy, and Philip H. Jones. Somatic mutant clones colonize the human esophagus with age. *Science*, 362(6417):911–917, November 2018. doi: 10.1126/science.aau3879.
- [3] Henry Lee-Six, Sigurgeir Olafsson, Peter Ellis, Robert J. Osborne, Mathijs A. Sanders, Luiza Moore, Nikitas Georgakopoulos, Franco Torrente, Ayesha Noorani, Martin Goddard, Philip Robinson, Tim H. H. Coorens, Laura O’Neill, Christopher Alder, Jingwei Wang, Rebecca C. Fitzgerald, Matthias Zilbauer, Nicholas Coleman, Kourosh Saeb-Parsy, Inigo Martincorena, Peter J. Campbell, and Michael R. Stratton. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature*, 574(7779):532–537, October 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1672-7.
- [4] Jan Vijg. Somatic mutations, genome mosaicism, cancer and aging. *Current Opinion in Genetics & Development*, 26:141–149, June 2014. ISSN 0959-437X. doi: 10.1016/j.gde.2014.04.002.
- [5] Akira Yokoyama, Nobuyuki Kakiuchi, Tetsuichi Yoshizato, Yasuhito Nannya, Hiromichi Suzuki, Yasuhide Takeuchi, Yusuke Shiozawa, Yusuke Sato, Kosuke Aoki, Soo Ki Kim, Yoichi Fujii, Kenichi Yoshida, Keisuke Kataoka, Masahiro M. Nakagawa, Yoshikage Inoue, Tomonori Hirano, Yuichi Shiraishi, Kenichi Chiba, Hiroko Tanaka, Masashi Sanada, Yoshitaka Nishikawa, Yusuke Amanuma, Shinya Ohashi, Ikuo Aoyama, Takahiro Horimatsu, Shin’ichi Miyamoto, Shigeru Tsunoda, Yoshiharu Sakai, Maiko Narahara, J. B. Brown, Yoshitaka Sato, Genta Sawada, Koshi Mimori, Sachiko Minamiguchi, Hironori Haga, Hiroshi Seno, Satoru Miyano, Hideki Makishima, Manabu Muto, and Seishi Ogawa. Age-related remodelling of oesophageal

epithelia by mutated cancer drivers. *Nature*, 565(7739):312–317, January 2019. ISSN 1476-4687. doi: 10.1038/s41586-018-0811-x.

- [6] Ludmil B. Alexandrov, Serena Nik-Zainal, David C. Wedge, Samuel A. J. R. Aparicio, Sam Behjati, Andrew V. Biankin, Graham R. Bignell, Niccolò Bolli, Ake Borg, Anne-Lise Børresen-Dale, Sandrine Boyault, Birgit Burkhardt, Adam P. Butler, Carlos Caldas, Helen R. Davies, Christine Desmedt, Roland Eils, Jónunn Erla Eyfjörð, John A. Foekens, Mel Greaves, Fumie Hosoda, Barbara Hutter, Tomislav Ilicic, Sandrine Imbeaud, Marcin Imielinski, Natalie Jäger, David T. W. Jones, David Jones, Stian Knappskog, Marcel Kool, Sunil R. Lakhani, Carlos López-Otín, Sancha Martin, Nikhil C. Munshi, Hiromi Nakamura, Paul A. Northcott, Marina Pajic, Elli Papaemmanuil, Angelo Paradiso, John V. Pearson, Xose S. Puente, Keiran Raine, Manasa Ramakrishna, Andrea L. Richardson, Julia Richter, Philip Rosenstiel, Matthias Schlesner, Ton N. Schumacher, Paul N. Span, Jon W. Teague, Yasushi Totoki, Andrew N. J. Tutt, Rafael Valdés-Mas, Marit M. van Buuren, Laura van ’t Veer, Anne Vincent-Salomon, Nicola Waddell, Lucy R. Yates, Jessica Zucman-Rossi, P. Andrew Futreal, Ultan McDermott, Peter Lichter, Matthew Meyerson, Sean M. Grimmond, Reiner Siebert, Elías Campo, Tatsuhiro Shibata, Stefan M. Pfister, Peter J. Campbell, and Michael R. Stratton. Signatures of mutational processes in human cancer. *Nature*, 500(7463):415–421, August 2013. ISSN 1476-4687. doi: 10.1038/nature12477.
- [7] Ferriol Calvet, Raquel Blanco Martinez-Illescas, Ferran Muiños, Maria Tretiakova, Elena S. Latorre-Esteves, Jeanne Fredrickson, Maria Andrianova, Stefano Pellegrini, Axel Rosendahl Huber, Joan Enric Ramis-Zaldivar, Shuyi Charlotte An, Elana Thieme, Brendan F. Kohn, Miguel L. Grau, Abel Gonzalez-Perez, Nuria Lopez-Bigas, and Rosa Ana Risques. Sex and smoking bias in the selection of somatic mutations in human bladder. *Nature*, pages 1–9, October 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09521-x.
- [8] Sergey Senkin, Sarah Moody, Marcos Díaz-Gay, Behnoush Abedi-Ardekani, Thomas Cattiaux, Aida Ferreira-Iglesias, Jingwei Wang, Stephen Fitzgerald, Mariya Kazachkova, Raviteja Vangara, Anh Phuong Le, Erik N. Bergstrom, Azhar Khandekar, Burçak Otlı, Saamin Cheema, Calli Latimer, Emily Thomas, Joshua Ronald Atkins, Karl Smith-Byrne, Ricardo Cortez Cardoso Penha, Christine Carreira, Priscilia Chopard, Valérie Gaborieau, Pekka Keski-Rahkonen, David Jones, Jon W. Teague, Sophie Ferlicot, Mojgan Asgari, Surasak Sangkhathat, Worapat Attawet-tayanon, Beata Świątkowska, Sonata Jarmalaite, Rasa Sabaliauskaite, Tatsuhiro Shibata, Akihiko Fukagawa, Dana Mates, Viorel Jinga, Stefan Rascu, Mirjana Mijuskovic, Slavisa Savic, Sasa Milosavljevic, John M. S. Bartlett, Monique Albert, Larry Phouthavongsy, Patricia Ashton-Prolla, Mariana R. Botton, Brasil Silva Neto, Stephania Martins Bezerra, Maria Paula Curado, Stênio de Cássio Zequi, Rui Manuel Reis, Eliney Ferreira Faria, Nei Soares de Menezes, Renata Spagnoli Ferrari, Rosamonde E. Banks, Naveen S. Vasudev, David Zaridze, Anush Mukeriya,

- Oxana Shangina, Vsevolod Matveev, Lenka Foretova, Marie Navratilova, Ivana Holcatova, Anna Hornakova, Vladimir Janout, Mark P. Purdue, Nathaniel Rothman, Stephen J. Chanock, Per Magne Ueland, Mattias Johansson, James McKay, Ghislaine Scelo, Estelle Chanudet, Laura Humphreys, Ana Carolina de Carvalho, Sandra Perdomo, Ludmil B. Alexandrov, Michael R. Stratton, and Paul Brennan. Geographic variation of mutagenic exposures in kidney cancer genomes. *Nature*, 629(8013):910–918, May 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07368-2.
- [9] Oriol Pich, Ferran Muiños, Martijn Paul Lolkema, Neeltje Steeghs, Abel Gonzalez-Perez, and Nuria Lopez-Bigas. The mutational footprints of cancer therapies. *Nature Genetics*, 51(12):1732–1740, December 2019. ISSN 1546-1718. doi: 10.1038/s41588-019-0525-5.
- [10] Emily Mitchell, My H. Pham, Anna Clay, Rashesh Sanghvi, Nicholas Williams, Sandra Pietsch, Joanne I. Hsu, Nina Friesgaard Øbro, Hyunchul Jung, Aditi Vedi, Sarah Moody, Jingwei Wang, Daniel Leongamornlert, Michael Spencer Chapman, Ellie Dunstone, Anna Santarsieri, Alex Cagan, Heather E. Machado, E. Joanna Baxter, George Follows, Daniel J. Hodson, Ultan McDermott, Gary J. Doherty, Inigo Martincorena, Laura Humphreys, Krishnaa Mahbubani, Kourosh Saeb Parsy, Koichi Takahashi, Margaret A. Goodell, David Kent, Elisa Laurenti, Peter J. Campbell, Raheleh Rahbari, Jyoti Nangalia, and Michael R. Stratton. The long-term effects of chemotherapy on normal blood cells. *Nature Genetics*, 57(7):1684–1694, July 2025. ISSN 1546-1718. doi: 10.1038/s41588-025-02234-x.
- [11] Heather E. Machado, Emily Mitchell, Nina F. Øbro, Kirsten Kübler, Megan Davies, Daniel Leongamornlert, Alyssa Cull, Francesco Maura, Mathijs A. Sanders, Alex T. J. Cagan, Craig McDonald, Miriam Belmonte, Mairi S. Shepherd, Felipe A. Vieira Braga, Robert J. Osborne, Krishnaa Mahbubani, Iñigo Martincorena, Elisa Laurenti, Anthony R. Green, Gad Getz, Paz Polak, Kourosh Saeb-Parsy, Daniel J. Hodson, David G. Kent, and Peter J. Campbell. Diverse mutational landscapes in human lymphocytes. *Nature*, 608(7924):724–732, August 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-05072-7.
- [12] Iñigo Martincorena, Keiran M. Raine, Moritz Gerstung, Kevin J. Dawson, Kerstin Haase, Peter Van Loo, Helen Davies, Michael R. Stratton, and Peter J. Campbell. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*, 171(5):1029–1041.e21, November 2017. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2017.09.042.
- [13] Austin L. Hughes. Near Neutrality. *Annals of the New York Academy of Sciences*, 1133(1):162–179, 2008. ISSN 1749-6632. doi: 10.1196/annals.1438.001.
- [14] Benjamin D. Simons and Hans Clevers. Strategies for Homeostatic Stem Cell Self-Renewal in Adult Tissues. *Cell*, 145(6):851–862, June 2011. ISSN 0092-8674. doi: 10.1016/j.cell.2011.05.033.

- [15] Nobuyuki Kakiuchi and Seishi Ogawa. Clonal expansion in non-cancer tissues. *Nature Reviews Cancer*, 21(4):239–256, April 2021. ISSN 1474-1768. doi: 10.1038/s41568-021-00335-3.
- [16] Ketevan Chkhaidze, Timon Heide, Benjamin Werner, Marc J. Williams, Weini Huang, Giulio Caravagna, Trevor A. Graham, and Andrea Sottoriva. Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data. *PLOS Computational Biology*, 15(7):e1007243, July 2019. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1007243.
- [17] Mel Greaves and Carlo C. Maley. Clonal evolution in cancer. *Nature*, 481(7381):306–313, January 2012. ISSN 1476-4687. doi: 10.1038/nature10762.
- [18] Peter C. Nowell. The Clonal Evolution of Tumor Cell Populations. *Science*, 194(4260):23–28, October 1976. doi: 10.1126/science.959840.
- [19] Inigo Martincorena. Somatic mutation and clonal expansions in human tissues. *Genome Medicine*, 11(1):35, May 2019. ISSN 1756-994X. doi: 10.1186/s13073-019-0648-4.
- [20] Siddhartha Jaiswal and Benjamin L. Ebert. Clonal hematopoiesis in human aging and disease. *Science*, 366(6465):eaan4673, November 2019. doi: 10.1126/science.aan4673.
- [21] David P. Steensma, Rafael Bejar, Siddhartha Jaiswal, R. Coleman Lindsley, Mikkael A. Sekeres, Robert P. Hasserjian, and Benjamin L. Ebert. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*, 126(1):9–16, July 2015. ISSN 0006-4971. doi: 10.1182/blood-2015-03-631747.
- [22] Giulio Genovese, Anna K. Kähler, Robert E. Handsaker, Johan Lindberg, Samuel A. Rose, Samuel F. Bakhoun, Kimberly Chambert, Eran Mick, Benjamin M. Neale, Menachem Fromer, Shaun M. Purcell, Oscar Svantesson, Mikael Landén, Martin Höglund, Sören Lehmann, Stacey B. Gabriel, Jennifer L. Moran, Eric S. Lander, Patrick F. Sullivan, Pamela Sklar, Henrik Grönberg, Christina M. Hultman, and Steven A. McCarroll. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine*, 371(26):2477–2487, December 2014. ISSN 0028-4793. doi: 10.1056/NEJMoa1409405.
- [23] Siddhartha Jaiswal, Pierre Fontanillas, Jason Flannick, Alisa Manning, Peter V. Grauman, Brenton G. Mar, R. Coleman Lindsley, Craig H. Mermel, Noel Burt, Alejandro Chavez, John M. Higgins, Vladislav Moltchanov, Frank C. Kuo, Michael J. Kluk, Brian Henderson, Leena Kinnunen, Heikki A. Koistinen, Claes Ladenvall, Gad Getz, Adolfo Correa, Benjamin F. Banahan, Stacey Gabriel, Sekar Kathiresan, Heather M. Stringham, Mark I. McCarthy, Michael Boehnke, Jaakko Tuomilehto,

- Christopher Haiman, Leif Groop, Gil Atzmon, James G. Wilson, Donna Neuberg, David Altshuler, and Benjamin L. Ebert. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine*, 371(26):2488–2498, December 2014. ISSN 0028-4793. doi: 10.1056/NEJMoa1408617.
- [24] Christopher S. Marnell, Alexander Bick, and Pradeep Natarajan. Clonal hematopoiesis of indeterminate potential (CHIP): Linking somatic mutations, hematopoiesis, chronic inflammation and cardiovascular disease. *Journal of Molecular and Cellular Cardiology*, 161:98–105, December 2021. ISSN 0022-2828. doi: 10.1016/j.yjmcc.2021.07.004.
- [25] Julia T. Warren and Daniel C. Link. Clonal hematopoiesis and risk for hematologic malignancy. *Blood*, 136(14):1599–1605, October 2020. ISSN 0006-4971. doi: 10.1182/blood.2019000991.
- [26] Alan S. Jonason, Subrahmanyam Kunala, Gary J. Price, Richard J. Restifo, Henry M. Spinelli, John A. Persing, David J. Leffell, Robert E. Tarone, and Douglas E. Brash. Frequent clones of p53-mutated keratinocytes in normal human skin. *Proceedings of the National Academy of Sciences*, 93(24):14025–14029, November 1996. doi: 10.1073/pnas.93.24.14025.
- [27] Allon M. Klein, Douglas E. Brash, Philip H. Jones, and Benjamin D. Simons. Stochastic fate of p53-mutant epidermal progenitor cells is tilted toward proliferation by UV B during preneoplasia. *Proceedings of the National Academy of Sciences*, 107(1):270–275, January 2010. doi: 10.1073/pnas.0909738107.
- [28] M. D. Lynch, C. N. S. Lynch, E. Craythorne, K. Liakath-Ali, R. Mallipeddi, J. N. Barker, and F. M. Watt. Spatial constraints govern competition of mutant clones in human epidermis. *Nature Communications*, 8(1):1119, October 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-00993-8.
- [29] Bartomeu Colom, Maria P. Alcolea, Gabriel Piedrafita, Michael W. J. Hall, Agnieszka Wabik, Stefan C. Dentro, Joanna C. Fowler, Albert Herms, Charlotte King, Swee Hoe Ong, Roshan K. Sood, Moritz Gerstung, Inigo Martincorena, Benjamin A. Hall, and Philip H. Jones. Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. *Nature Genetics*, 52(6):604–614, June 2020. ISSN 1546-1718. doi: 10.1038/s41588-020-0624-3.
- [30] Ryan O. Schenck, Eunjung Kim, Rafael R. Bravo, Jeffrey West, Simon Leedham, Darryl Shibata, and Alexander R. A. Anderson. Homeostasis limits keratinocyte evolution. *Proceedings of the National Academy of Sciences*, 119(35):e2006487119, August 2022. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2006487119.
- [31] Wengeng Zhang, Eva Remenyik, Daniel Zeltermann, Douglas E. Brash, and Norbert M. Wikonkal. Escaping the stem cell compartment: Sustained UVB exposure

- allows p53-mutant keratinocytes to colonize adjacent epidermal proliferating units without incurring additional mutations. *Proceedings of the National Academy of Sciences*, 98(24):13948–13953, November 2001. doi: 10.1073/pnas.241353198.
- [32] Dennis L. Chao, J. Thomas Eck, Douglas E. Brash, Carlo C. Maley, and E. Georg Luebeck. Preneoplastic lesion growth driven by the death of adjacent normal stem cells. *Proceedings of the National Academy of Sciences*, 105(39):15034–15039, September 2008. doi: 10.1073/pnas.0802211105.
- [33] Jeffrey West, Ryan O. Schenck, Chandler Gatenbee, Mark Robertson-Tessi, and Alexander R. A. Anderson. Normal tissue architecture determines the evolutionary course of cancer. *Nature Communications*, 12(1):2060, April 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-22123-1.
- [34] Marta Ciwinska, Hendrik A. Messal, Hristina R. Hristova, Catrin Lutz, Laura Bornes, Theofilos Chalkiadakis, Rolf Harkes, Nathalia S. M. Langedijk, Stefan J. Hutten, Renée X. Menezes, Jos Jonkers, Stefan Prekovic, Benjamin D. Simons, Colinda L. G. J. Scheele, and Jacco van Rheenen. Mechanisms that clear mutations drive field cancerization in mammary tissue. *Nature*, 633(8028):198–206, September 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07882-3.
- [35] B. Colom, A. Herms, M. W. J. Hall, S. C. Dentre, C. King, R. K. Sood, M. P. Alcolea, G. Piedrafita, D. Fernandez-Antoran, S. H. Ong, J. C. Fowler, K. T. Mahbubani, K. Saeb-Parsy, M. Gerstung, B. A. Hall, and P. H. Jones. Mutant clones in normal epithelium outcompete and eliminate emerging tumours. *Nature*, 598(7881):510–514, October 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03965-7.
- [36] Kelly L. Bolton, Ryan N. Ptashkin, Teng Gao, Lior Braunstein, Sean M. Devlin, Daniel Kelly, Minal Patel, Antonin Berthon, Aijazuddin Syed, Mariko Yabe, Catherine C. Coombs, Nicole M. Caltabellotta, Mike Walsh, Kenneth Offit, Zsofia Stadler, Diana Mandelker, Jessica Schulman, Akshar Patel, John Philip, Elsa Bernard, Gunes Gundem, Juan E. Arango Ossa, Max Levine, Juan S. Medina Martinez, Noushin Farnoud, Dominik Glodzik, Sonya Li, Mark E. Robson, Choonsik Lee, Paul D. P. Pharoah, Konrad H. Stopsack, Barbara Spitzer, Simon Mantha, James Fagin, Laura Boucai, Christopher J. Gibson, Benjamin L. Ebert, Andrew L. Young, Todd Druley, Koichi Takahashi, Nancy Gillis, Markus Ball, Eric Padron, David M. Hyman, Jose Baselga, Larry Norton, Stuart Gardos, Virginia M. Klimek, Howard Scher, Dean Bajorin, Eder Paraiso, Ryma Benayed, Maria E. Arcila, Marc Ladanyi, David B. Solit, Michael F. Berger, Martin Tallman, Montserrat Garcia-Closas, Nilanjan Chatterjee, Luis A. Diaz, Ross L. Levine, Lindsay M. Morton, Ahmet Zehir, and Elli Papaemmanuil. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nature Genetics*, 52(11):1219–1226, November 2020. ISSN 1546-1718. doi: 10.1038/s41588-020-00710-0.

- [37] Catherine C. Coombs, Ahmet Zehir, Sean M. Devlin, Ashwin Kishtagari, Aijazuddin Syed, Philip Jonsson, David M. Hyman, David B. Solit, Mark E. Robson, José Baselga, Maria E. Arcila, Marc Ladanyi, Martin S. Tallman, Ross L. Levine, and Michael F. Berger. Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell*, 21(3):374–382.e4, September 2017. ISSN 1934-5909. doi: 10.1016/j.stem.2017.07.010.
- [38] Koichi Takahashi, Daisuke Nakada, and Margaret Goodell. Distinct landscape and clinical implications of therapy-related clonal hematopoiesis. *The Journal of Clinical Investigation*, 134(19), October 2024. ISSN 0021-9738. doi: 10.1172/JCI180069.
- [39] Julia Matas, Brendan Kohn, Jeanne Fredrickson, Kelly Carter, Ming Yu, Ting Wang, Xianrong Gui, Thierry Soussi, Victor Moreno, William M. Grady, Miguel A. Peinado, and Rosa Ana Risques. Colorectal Cancer Is Associated with the Presence of Cancer Driver Mutations in Normal Colon. *Cancer Research*, 82(8):1492–1502, April 2022. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-21-3607.
- [40] Danely P. Slaughter, Harry W. Southwick, and Walter Smejkal. “Field cancerization” in oral stratified squamous epithelium. Clinical implications of multicentric origin. *Cancer*, 6(5):963–968, 1953. ISSN 1097-0142. doi: 10.1002/1097-0142(195309)6:5<963::AID-CNCR2820060515>3.0.CO;2-Q.
- [41] Boudewijn J. M. Braakhuis, C. René Leemans, and Ruud H. Brakenhoff. A genetic progression model of oral cancer: current evidence and clinical implications. *Journal of Oral Pathology & Medicine*, 33(6):317–322, 2004. ISSN 1600-0714. doi: 10.1111/j.1600-0714.2004.00225.x.
- [42] Kit Curtius, Nicholas A. Wright, and Trevor A. Graham. An evolutionary perspective on field cancerization. *Nature Reviews Cancer*, 18(1):19–32, January 2018. ISSN 1474-1768. doi: 10.1038/nrc.2017.102.
- [43] Leslie Foulds. The natural history of cancer. *Journal of Chronic Diseases*, 8(1):2–37, July 1958. ISSN 0021-9681. doi: 10.1016/0021-9681(58)90039-0.
- [44] Eric R. Fearon and Bert Vogelstein. A genetic model for colorectal tumorigenesis. *Cell*, 61(5):759–767, June 1990. ISSN 0092-8674. doi: 10.1016/0092-8674(90)90186-I.
- [45] Bert Vogelstein and Kenneth W. Kinzler. Cancer genes and the pathways they control. *Nature Medicine*, 10(8):789–799, August 2004. ISSN 1546-170X. doi: 10.1038/nm1087.
- [46] Jesse J. Salk, Kaitlyn Loubet-Senear, Elisabeth Maritschnegg, Charles C. Valentine, Lindsey N. Williams, Jacob E. Higgins, Reinhard Horvat, Adriaan Vanderstichele, Daniela Nachmanson, Kathryn T. Baker, Mary J. Emond, Emily Loter, Maria Tretiakova, Thierry Soussi, Lawrence A. Loeb, Robert Zeillinger, Paul Speiser,

- and Rosa Ana Risques. Ultra-Sensitive *TP53* Sequencing for Cancer Detection Reveals Progressive Clonal Selection in Normal Tissue over a Century of Human Lifespan. *Cell Reports*, 28(1):132–144.e3, July 2019. ISSN 2211-1247. doi: 10.1016/j.celrep.2019.05.109.
- [47] Emilie Abby, Stefan C. Dentro, Michael W. J. Hall, Joanna C. Fowler, Swee Hoe Ong, Roshan Sood, Albert Herms, Gabriel Piedrafita, Irina Abnizova, Christian W. Siebel, Moritz Gerstung, Benjamin A. Hall, and Philip H. Jones. Notch1 mutations drive clonal expansion in normal esophageal epithelium but impair tumor growth. *Nature Genetics*, 55(2):232–245, February 2023. ISSN 1546-1718. doi: 10.1038/s41588-022-01280-z.
 - [48] Kasumi Murai, Stefan Dentro, Swee Hoe Ong, Roshan Sood, David Fernandez-Antoran, Albert Herms, Vasiliki Kostiou, Irina Abnizova, Benjamin A. Hall, Moritz Gerstung, and Philip H. Jones. p53 mutation in normal esophagus promotes multiple stages of carcinogenesis but is constrained by clonal competition. *Nature Communications*, 13(1):6206, October 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-33945-y.
 - [49] Alfred G. Knudson. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proceedings of the National Academy of Sciences*, 68(4):820–823, April 1971. doi: 10.1073/pnas.68.4.820.
 - [50] Chay Paterson, Hans Clevers, and Ivana Bozic. Mathematical model of colorectal cancer initiation. *Proceedings of the National Academy of Sciences*, 117(34):20681–20688, August 2020. doi: 10.1073/pnas.2003771117.
 - [51] Francisco Martínez-Jiménez, Ferran Muiños, Inés Sentís, Jordi Deu-Pons, Iker Reyes-Salazar, Claudia Arnedo-Pac, Loris Mularoni, Oriol Pich, Jose Bonet, Hanna Kranas, Abel Gonzalez-Perez, and Nuria Lopez-Bigas. A compendium of mutational cancer driver genes. *Nature Reviews Cancer*, 20(10):555–572, October 2020. ISSN 1474-1768. doi: 10.1038/s41568-020-0290-x.
 - [52] Andrew J. Deans and Stephen C. West. DNA interstrand crosslink repair and cancer. *Nature Reviews Cancer*, 11(7):467–480, July 2011. ISSN 1474-1768. doi: 10.1038/nrc3088.
 - [53] Joshi Niraj, Anniina Färkkilä, and Alan D. D’Andrea. The Fanconi Anemia Pathway in Cancer. *Annual Review of Cancer Biology*, 3(Volume 3, 2019):457–478, March 2019. ISSN 2472-3428. doi: 10.1146/annurev-cancerbio-030617-050422.
 - [54] Xavier Renaudin and Filippo Rosselli. The FANC/BRCA Pathway Releases Replication Blockades by Eliminating DNA Interstrand Cross-Links. *Genes*, 11(5):585, May 2020. ISSN 2073-4425. doi: 10.3390/genes11050585.

- [55] Magali Olivier, Monica Hollstein, and Pierre Hainaut. TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harbor Perspectives in Biology*, 2(1):a001008, January 2010. ISSN , 1943-0264. doi: 10.1101/cshperspect.a001008.
- [56] Lawrence A. Donehower, Thierry Soussi, Anil Korkut, Yuexin Liu, Andre Schultz, Maria Cardenas, Xubin Li, Ozgun Babur, Teng-Kuei Hsu, Olivier Lichtarge, John N. Weinstein, Rehan Akbani, and David A. Wheeler. Integrated Analysis of *TP53* Gene and Pathway Alterations in The Cancer Genome Atlas. *Cell Reports*, 28(5):1370–1384.e5, July 2019. ISSN 2211-1247. doi: 10.1016/j.celrep.2019.07.001.
- [57] Dharendra K. Simanshu, Dwight V. Nissley, and Frank McCormick. RAS Proteins and Their Regulators in Human Disease. *Cell*, 170(1):17–33, June 2017. ISSN 0092-8674. doi: 10.1016/j.cell.2017.06.009.
- [58] Douglas Hanahan. Hallmarks of cancer—Then and now, and beyond. *Cell*, 189(8):2254–2277, April 2026. ISSN 0092-8674. doi: 10.1016/j.cell.2025.12.049.
- [59] James Whitworth, Philip S. Smith, Jose-Ezequiel Martin, Hannah West, Andrea Luchetti, Faye Rodger, Graeme Clark, Keren Carss, Jonathan Stephens, Kathleen Stirrups, Chris Penkett, Rutendo Mapeta, Sofie Ashford, Karyn Megy, Hassan Sha-keel, Munaza Ahmed, Julian Adlard, Julian Barwell, Carole Brewer, Ruth T. Casey, Ruth Armstrong, Trevor Cole, Dafydd Gareth Evans, Florentia Fostira, Lynn Greenhalgh, Helen Hanson, Alex Henderson, Jonathan Hoffman, Louise Izatt, Ajith Kumar, Ava Kwong, Fiona Lalloo, Kai Ren Ong, Joan Paterson, Soo-Mi Park, Rakefet Chen-Shtoyerman, Claire Searle, Lucy Side, Anne-Bine Skytte, Katie Snape, Emma R. Woodward, Timothy Aitman, Hana Alachkar, Sonia Ali, Louise Allen, David Allsup, Gautum Ambegaonkar, Julie Anderson, Richard Antrobus, Ruth Armstrong, Gavin Arno, Gururaj Arumugakani, Sofie Ashford, William Astle, Antony Attwood, Steve Austin, Chiara Bacchelli, Tamam Bakchoul, Tadbir K. Bariana, Helen Baxendale, David Bennett, Claire Bethune, Shahnaz Bibi, Maria Bitner-Glindzicz, Marta Bleda, Harm Boggard, Paula Bolton-Maggs, Claire Booth, John R. Bradley, Angie Brady, Matthew Brown, Michael Browning, Christine Bryson, Siobhan Burns, Paul Calleja, Natalie Canham, Jenny Carmichael, Keren Carss, Mark Caulfield, Elizabeth Chalmers, Anita Chandra, Patrick Chinnery, Manali Chitre, Colin Church, Emma Clement, Naomi Clements-Brod, Virginia Clowes, Gerry Coghlan, Peter Collins, Victoria Cookson, Nichola Cooper, Paul Corris, Amanda Creaser-Myers, Rosa DaCosta, Louise Daugherty, Sophie Davies, John Davis, Minka De Vries, Patrick Deegan, Sri V. V. Deevi, Charu Deshpande, Lisa Devlin, Eleanor Dewhurst, Peter Dixon, Rainer Doffinger, Natalie Dormand, Elizabeth Drewe, David Edgar, William Egner, Wendy N. Erber, Marie Erwood, Marie Erwood, Tamara Everington, Remi Favier, Helen Firth, Debra Fletcher, Frances Flinter, Amy Frary, Kathleen Freson, Bruce Furie, Abigail Furnell, Daniel Gale, Alice Gardham, Michael

Gattens, Neeti Ghali, Pavandeep K. Ghataorhe, Rohit Ghurye, Simon Gibbs, Kimberley Gilmour, Paul Gissen, Sarah Goddard, Keith Gomez, Pavel Gordins, Stefan Graf, Stefan Gräf, Daniel Greene, Alan Greenhalgh, Andreas Greinacher, Sofia Grigoriadou, Detelina Grozeva, Scott Hackett, Charaka Hadinnapola, Rosie Hague, Matthias Haimel, Csaba Halmagyi, Tracey Hammerton, Daniel Hart, Grant Hayman, Johan W. M. Heemskerk, Robert Henderson, Anke Hensiek, Yvonne Henskens, Archana Herwadkar, Simon Holden, Muriel Holder, Susan Holder, Fengyuan Hu, Anna Huis in't Veld, Aarnoud Huissoon, Marc Humbert, Jane Hurst, Roger James, Stephen Jolles, Dragana Josifova, Rashid Kazmi, David Keeling, Peter Kelleher, Anne M. Kelly, Fiona Kennedy, David Kiely, Nathalie Kingston, Ania Koziell, Deepa Krishnakumar, Taco W. Kuijpers, Taco Kuijpers, Dinakantha Kumararatne, Manju Kurian, Michael A. Laffan, Michele P. Lambert, Hana Lango Allen, Hana Lango-Allen, Allan Lawrie, Sara Lear, Melissa Lees, Claire Lentaigne, Ri Liesner, Rachel Linger, Hilary Longhurst, Lorena Lorenzo, Eleni Louka, Rajiv Machado, Rob Mackenzie Ross, Robert MacLaren, Eamonn Maher, Jesmeen Maimaris, Sarah Mangles, Ania Manson, Rutendo Mapeta, Hugh S. Markus, Jennifer Martin, Larahmie Masati, Mary Mathias, Vera Matser, Anna Maw, Elizabeth McDermott, Coleen McJannet, Stuart Meacham, Sharon Meehan, Karyn Megy, Sarju Mehta, Michel Michaelides, Carolyn M. Millar, Shahin Moledina, Anthony Moore, Nicholas Morrell, Andrew Mumford, Sai Murng, Elaine Murphy, Sergey Nejentsev, Sadia Noorani, Paquita Nurden, Eric Oksenhendler, Shokri Othman, Willem H. Ouwehand, Willem H. Ouwehand, Sofia Papadia, Soo-Mi Park, Alasdair Parker, John Pasi, Chris Patch, Joan Paterson, Jeanette Payne, Andrew Peacock, Kathelijne Peerlinck, Christopher J. Penkett, Joanna Pepke-Zaba, David Perry, David J. Perry, Val Pollock, Gary Polwarth, Mark Ponsford, Waseem Qasim, Isabella Quinti, Stuart Rankin, Julia Rankin, F. Lucy Raymond, Paula Rayner-Matthews, Karola Rehnstrom, Evan Reid, Christopher J. Rhodes, Michael Richards, Sylvia Richardson, Alex Richter, Irene Roberts, Matthew Rondina, Elisabeth Rosser, Catherine Roughley, Noémi Roy, Kevin Rue-Albrecht, Crina Samarghitean, Alba Sanchis-Juan, Richard Sandford, Saikat Santra, Ravishankar Sargur, Sinisa Savic, Gwen Schotte, Sol Schulman, Harald Schulze, Richard Scott, Marie Scully, Suranjith Seneviratne, Carrock Sewell, Olga Shamardina, Debbie Shipley, Ilenia Simeoni, Suthesh Sivapalaratnam, Kenneth G. C. Smith, Aman Sohal, Laura Southgate, Simon Staines, Emily Staples, Hannah Stark, Hans Stauss, Penelope Stein, Jonathan Stephens, Kathleen Stirrups, Sophie Stock, Jay Suntharalingam, Kate Talks, Yvonne Tan, Jecko Thachil, James Thaventhiran, Ellen Thomas, Moira Thomas, Dorothy Thompson, Adrian Thrasher, Marc Tischkowitz, Catherine Titterton, Cheng-Hock Toh, Mark Toshner, Carmen Treacy, Richard Trembath, Salih Tuna, Wojciech Turek, Ernest Turro, Chris Van Geet, Marijke Veltman, Julie Vogt, Julie von Ziegenweldt, Anton Vonk Noordegraaf, Emma Wakeling, Ivy Wanjiku, Timothy Q. Warner, Evangeline Wassmer, Hugh Watkins, Christopher Watt, Andrew Webster, Steve Welch, Sarah Westbury, John Wharton, Deborah Whitehorn, Martin Wilkins, Lisa Willcocks, Cather-

- ine Williamson, Geoffrey Woods, Geoff Woods, John Wort, Nigel Yeatman, Patrick Yong, Tim Young, Ping Yu, Marc D. Tischkowitz, and Eamonn R. Maher. Comprehensive Cancer-Predisposition Gene Testing in an Adult Multiple Primary Tumor Series Shows a Broad Range of Deleterious Variants and Atypical Tumor Phenotypes. *The American Journal of Human Genetics*, 103(1):3–18, July 2018. ISSN 0002-9297. doi: 10.1016/j.ajhg.2018.04.013.
- [60] Junne Kamihara, Huma Q. Rana, and Judy E. Garber. Germline TP53 Mutations and the Changing Landscape of Li–Fraumeni Syndrome. *Human Mutation*, 35(6): 654–662, 2014. ISSN 1098-1004. doi: 10.1002/humu.22559.
- [61] Preetha Anand, Ajaikumar B. Kunnumakara, Chitra Sundaram, Kuzhuvelil B. Harikumar, Sheeja T. Tharakan, Oiki S. Lai, Bokyung Sung, and Bharat B. Aggarwal. Cancer is a Preventable Disease that Requires Major Lifestyle Changes. *Pharmaceutical Research*, 25(9):2097–2116, September 2008. ISSN 1573-904X. doi: 10.1007/s11095-008-9661-9.
- [62] Michele Carbone, Sarah T. Arron, Bruce Beutler, Angela Bononi, Webster Cavenee, James E. Cleaver, Carlo M. Croce, Alan D’Andrea, William D. Foulkes, Giovanni Gaudino, Joanna L. Groden, Elizabeth P. Henske, Ian D. Hickson, Paul M. Hwang, Richard D. Kolodner, Tak W. Mak, David Malkin, Raymond J. Monnat, Flavia Novelli, Harvey I. Pass, John H. Petrini, Laura S. Schmidt, and Haining Yang. Tumour predisposition and cancer syndromes as models to study gene–environment interactions. *Nature Reviews Cancer*, 20(9):533–549, September 2020. ISSN 1474-1768. doi: 10.1038/s41568-020-0265-y.
- [63] Hannah Farmer, Nuala McCabe, Christopher J. Lord, Andrew N. J. Tutt, Damian A. Johnson, Tobias B. Richardson, Manuela Santarosa, Krystyna J. Dillon, Ian Hickson, Charlotte Knights, Niall M. B. Martin, Stephen P. Jackson, Graeme C. M. Smith, and Alan Ashworth. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, 434(7035):917–921, April 2005. ISSN 1476-4687. doi: 10.1038/nature03445.
- [64] Helen E. Bryant, Niklas Schultz, Huw D. Thomas, Kayan M. Parker, Dan Flower, Elena Lopez, Suzanne Kyle, Mark Meuth, Nicola J. Curtin, and Thomas Helleday. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*, 434(7035):913–917, April 2005. ISSN 1476-4687. doi: 10.1038/nature03443.
- [65] Christopher J. Lord and Alan Ashworth. PARP inhibitors: Synthetic lethality in the clinic. *Science*, 355(6330):1152–1158, March 2017. doi: 10.1126/science.aam7344.
- [66] John Burn, Harsh Sheth, Faye Elliott, Lynn Reed, Finlay Macrae, Jukka-Pekka Mecklin, Gabriela Möslein, Fiona E. McRonald, Lucio Bertario, D. Gareth Evans, Anne-Marie Gerdes, Judy W. C. Ho, Annika Lindblom, Patrick J. Morrison, Jem

- Rashbass, Raj Ramesar, Toni Seppälä, Huw J. W. Thomas, Kirsi Pylvänäinen, Gillian M. Borthwick, John C. Mathers, D. Timothy Bishop, Alex Boussioutas, Carole Brewer, Jackie Cook, Diana Eccles, Anthony Ellis, Shirley V. Hodgson, Jan Lubinski, Eamonn R. Maher, Mary EM Porteous, Julian Sampson, Rodney J. Scott, and Lucy Side. Cancer prevention with aspirin in hereditary colorectal cancer (Lynch syndrome), 10-year follow-up and registry-based 20-year data in the CAPP2 study: a double-blind, randomised, placebo-controlled trial. *The Lancet*, 395(10240):1855–1863, June 2020. ISSN 0140-6736, 1474-547X. doi: 10.1016/S0140-6736(20)30366-4.
- [67] Kate A. McBride, Mandy L. Ballinger, Emma Killick, Judy Kirk, Martin H. N. Tattersall, Rosalind A. Eeles, David M. Thomas, and Gillian Mitchell. Li-Fraumeni syndrome: cancer risk assessment and clinical management. *Nature Reviews Clinical Oncology*, 11(5):260–271, May 2014. ISSN 1759-4782. doi: 10.1038/nrclinonc.2014.41.
- [68] Gaëlle Bougeard, Mariette Renaux-Petel, Jean-Michel Flaman, Camille Charbonnier, Pierre Fermey, Muriel Belotti, Marion Gauthier-Villars, Dominique Stoppa-Lyonnet, Emilie Consolino, Laurence Brugières, Olivier Caron, Patrick R. Benusiglio, Brigitte Bressac-de Paillerets, Valérie Bonadona, Catherine Bonaïti-Pellié, Julie Tinat, Stéphanie Baert-Desurmont, and Thierry Frebourg. Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. *Journal of Clinical Oncology*, 33(21):2345–2352, July 2015. ISSN 0732-183X. doi: 10.1200/JCO.2014.59.5728.
- [69] Phuong L. Mai, Ana F. Best, June A. Peters, Rosamma M. DeCastro, Payal P. Khincha, Jennifer T. Loud, Renée C. Bremer, Philip S. Rosenberg, and Sharon A. Savage. Risks of first and subsequent cancers among TP53 mutation carriers in the National Cancer Institute Li-Fraumeni syndrome cohort. *Cancer*, 122(23):3673–3681, 2016. ISSN 1097-0142. doi: 10.1002/cncr.30248.
- [70] Christian P. Kratz, Claire Freycon, Kara N. Maxwell, Kim E. Nichols, Joshua D. Schiffman, D. Gareth Evans, Maria I. Achatz, Sharon A. Savage, Jeffrey N. Weitzel, Judy E. Garber, Pierre Hainaut, and David Malkin. Analysis of the Li-Fraumeni Spectrum Based on an International Germline TP53 Variant Data Set: An International Agency for Research on Cancer TP53 Database Analysis. *JAMA Oncology*, 7(12):1800–1805, December 2021. ISSN 2374-2437. doi: 10.1001/jamaoncol.2021.4398.
- [71] Blanche P. Alter, Neelam Giri, Sharon A. Savage, and Philip S. Rosenberg. Cancer in the National Cancer Institute inherited bone marrow failure syndrome cohort after fifteen years of follow-up. *Haematologica*, 103(1):30–39, January 2018. ISSN 1592-8721. doi: 10.3324/haematol.2017.178111.
- [72] Grzegorz Nalepa and D. Wade Clapp. Fanconi anaemia and cancer: an intricate relationship. *Nature Reviews Cancer*, 18(3):168–185, March 2018. ISSN 1474-1768. doi: 10.1038/nrc.2017.116.

- [73] A. Malcolm R. Taylor, Cynthia Rothblum-Oviatt, Nathan A. Ellis, Ian D. Hickson, Stefan Meyer, Thomas O. Crawford, Agata Smogorzewska, Barbara Pietrucha, Corry Weemaes, and Grant S. Stewart. Chromosome instability syndromes. *Nature Reviews Disease Primers*, 5(1):1–20, September 2019. ISSN 2056-676X. doi: 10.1038/s41572-019-0113-0.
- [74] Raphael Ceccaldi, Kalindi Parmar, Enguerran Mouly, Marc Delord, Jung Min Kim, Marie Regairaz, Marika Pla, Nadia Vasquez, Qing-Shuo Zhang, Corinne Pondarre, Régis Peffault de Latour, Eliane Gluckman, Marina Cavazzana-Calvo, Thierry Leblanc, Jérôme Larghero, Markus Grompe, Gérard Socié, Alan D. D’Andrea, and Jean Soulier. Bone Marrow Failure in Fanconi Anemia Is Triggered by an Exacerbated p53/p21 DNA Damage Response that Impairs Hematopoietic Stem and Progenitor Cells. *Cell Stem Cell*, 11(1):36–49, July 2012. ISSN 1934-5909. doi: 10.1016/j.stem.2012.05.013.
- [75] Rex H. Lee, Hyunseok Kang, Sue S. Yom, Agata Smogorzewska, Daniel E. Johnson, and Jennifer R. Grandis. Treatment of Fanconi Anemia–Associated Head and Neck Cancer: Opportunities to Improve Outcomes. *Clinical Cancer Research*, 27(19):5168–5187, October 2021. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-21-1259.
- [76] Juliette Thariat, Francois Chevalier, Daniel Orbach, Luc Ollivier, Pierre-Yves Marcy, Nadege Corradini, Arnaud Beddok, Nicolas Foray, and Gaelle Bougeard. Avoidance or adaptation of radiotherapy in patients with cancer with Li-Fraumeni and heritable TP53-related cancer syndromes. *The Lancet Oncology*, 22(12):e562–e574, December 2021. ISSN 1470-2045, 1474-5488. doi: 10.1016/S1470-2045(21)00425-3.
- [77] David I. Kutler, Krupa R. Patel, Arleen D. Auerbach, Jennifer Kennedy, Francis P. Lach, Erica Sanborn, Marc A. Cohen, William I. Kuhel, and Agata Smogorzewska. Natural history and management of Fanconi anemia patients with head and neck cancer: A 10-year follow-up. *The Laryngoscope*, 126(4):870–879, 2016. ISSN 1531-4995. doi: 10.1002/lary.25726.
- [78] Kelvin César de Andrade, Payal P Khincha, Jessica N Hatton, Megan N Frone, Talia Wegman-Ostrosky, Phuong L Mai, Ana F Best, and Sharon A Savage. Cancer incidence, patterns, and genotype–phenotype associations in individuals with pathogenic or likely pathogenic germline *TP53* variants: an observational cohort study. *The Lancet Oncology*, 22(12):1787–1798, December 2021. ISSN 1470-2045. doi: 10.1016/S1470-2045(21)00580-5.
- [79] Kelvin C. de Andrade, Natasha T. Strande, Jung Kim, Jeremy S. Haley, Jessica N. Hatton, Megan N. Frone, Payal P. Khincha, Gretchen M. Thone, Uyenlinh L. Mirshahi, Cynthia Schneider, Heena Desai, James T. Dove, Diane T. Smelser, Arnold J. Levine, Kara N. Maxwell, Douglas R. Stewart, David J. Carey, and Sharon A. Savage. Genome-first approach of the prevalence and cancer phenotypes of pathogenic

- or likely pathogenic germline *TP53* variants. *Human Genetics and Genomics Advances*, 5(1):100242, January 2024. ISSN 2666-2477. doi: 10.1016/j.xhgg.2023.100242.
- [80] Kevin T. Nead, Taebeom Kim, LiJin Joo, Tina L. McDowell, Justin W. Wong, Ireneaeus C. C. Chan, Elizabeth Brock, Jing Zhao, Ting Xu, Chad Tang, Chang-Lung Lee, Jun-ichi Abe, Kelly L. Bolton, Zhongxing Liao, Paul A. Scheet, and Steven H. Lin. Impact of cancer therapy on clonal hematopoiesis mutations and subsequent clinical outcomes. *Blood Advances*, 8(19):5215–5224, October 2024. ISSN 2473-9529. doi: 10.1182/bloodadvances.2024012929.
- [81] Nicholas Light, Mehdi Layeghifard, Ayush Attery, Vallijah Subasri, Matthew Zatzman, Nathaniel D. Anderson, Rupal Hatkar, Sasha Blay, David Chen, Ana Novokmet, Fabio Fuligni, James Tran, Richard de Borja, Himanshi Agarwal, Larissa Waldman, Lisa M. Abegglen, Daniel Albertson, Jonathan L. Finlay, Jordan R. Hansford, Sam Behjati, Anita Villani, Moritz Gerstung, Ludmil B. Alexandrov, Gino R. Somers, Joshua D. Schiffman, Varda Rotter, David Malkin, and Adam Shlien. Germline TP53 mutations undergo copy number gain years prior to tumor diagnosis. *Nature Communications*, 14(1):77, January 2023. ISSN 2041-1723. doi: 10.1038/s41467-022-35727-y.
- [82] Ozge Ceyhan-Birsoy, Pier Selenica, M Herman Chui, Gowtham Jayakumaran, Ryan Ptashkin, Maksym Misyura, Umut Aypar, Sowmya Jairam, Ciyu Yang, Yirong Li, Nikita Mehta, Yelena Kemel, Erin Salo-Mullen, Anna Maio, Margaret Sheehan, Ahmet Zehir, Maria Carlo, Alicia Latham, Zsofia Stadler, Mark Robson, Kenneth Offit, Marc Ladanyi, Michael Walsh, Jorge S Reis-Filho, and Diana Mandelker. Paired Tumor-Normal Sequencing Provides Insights Into the TP53-Related Cancer Spectrum in Patients With Li-Fraumeni Syndrome. *JNCI: Journal of the National Cancer Institute*, 113(12):1751–1760, December 2021. ISSN 0027-8874. doi: 10.1093/jnci/djab117.
- [83] Hany Ariffin, Pierre Hainaut, Anna Puzio-Kuter, Soo Sin Choong, Adelyne Sue Li Chan, Denis Tolkunov, Gunaretnam Rajagopal, Wenfeng Kang, Leon Li Wen Lim, Shekhar Krishnan, Kok-Siong Chen, Maria Isabel Achatz, Mawar Karsa, Jannah Shamsani, Arnold J. Levine, and Chang S. Chan. Whole-genome sequencing analysis of phenotypic heterogeneity and anticipation in Li-Fraumeni cancer predisposition syndrome. *Proceedings of the National Academy of Sciences*, 111(43):15497–15501, October 2014. doi: 10.1073/pnas.1417322111.
- [84] Philip S. Robinson, Laura E. Thomas, Federico Abascal, Hyunchul Jung, Luke M. R. Harvey, Hannah D. West, Sigurgeir Olafsson, Bernard C. H. Lee, Tim H. H. Coorens, Henry Lee-Six, Laura Butlin, Nicola Lander, Rebekah Truscott, Mathijs A. Sanders, Stefanie V. Lensing, Simon J. A. Buczacki, Rogier ten Hoopen, Nicholas Coleman, Roxanne Brunton-Sim, Simon Rushbrook, Kourosh Saeb-Parsy, Fiona

- Laloo, Peter J. Campbell, Iñigo Martincorena, Julian R. Sampson, and Michael R. Stratton. Inherited MUTYH mutations cause elevated somatic mutation rates and distinctive mutational signatures in normal human cells. *Nature Communications*, 13(1):3949, July 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-31341-0.
- [85] Bernard C. H. Lee, Philip S. Robinson, Tim H. H. Coorens, Helen H. N. Yan, Sigurgeir Olafsson, Henry Lee-Six, Mathijs A. Sanders, Hoi Cheong Siu, James Hewinson, Sarah S. K. Yue, Wai Yin Tsui, Annie S. Y. Chan, Anthony K. W. Chan, Siu Lun Ho, Peter J. Campbell, Inigo Martincorena, Simon J. A. Buczacki, Siu Tsan Yuen, Suet Yi Leung, and Michael R. Stratton. Mutational landscape of normal epithelial cells in Lynch Syndrome patients. *Nature Communications*, 13(1):2710, May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29920-2.
- [86] Niko Beerenwinkel, Tibor Antal, David Dingli, Arne Traulsen, Kenneth W. Kinzler, Victor E. Velculescu, Bert Vogelstein, and Martin A. Nowak. Genetic Progression and the Waiting Time to Cancer. *PLOS Computational Biology*, 3(11):e225, November 2007. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0030225.
- [87] Peter C. van den Akker, Anna M. G. Pasmooij, Hans Joenje, Robert M. W. Hofstra, Gerard J. te Meerman, and Marcel F. Jonkman. A "late-but-fitter revertant cell" explains the high frequency of revertant mosaicism in epidermolysis bullosa. *PLOS ONE*, 13(2):e0192994, February 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0192994.
- [88] Jeffrey West, Mark Robertson-Tessi, and Alexander R. A. Anderson. Agent-based methods facilitate integrative science in cancer. *Trends in Cell Biology*, 33(4):300–311, April 2023. ISSN 0962-8924, 1879-3088. doi: 10.1016/j.tcb.2022.10.006.
- [89] Rafael R. Bravo, Etienne Baratchart, Jeffrey West, Ryan O. Schenck, Anna K. Miller, Jill Gallaher, Chandler D. Gatenbee, David Basanta, Mark Robertson-Tessi, and Alexander R. A. Anderson. Hybrid Automata Library: A flexible platform for hybrid modeling with real-time visualization. *PLOS Computational Biology*, 16(3):e1007635, March 2020. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1007635.
- [90] Jill A. Gallaher, Susan C. Massey, Andrea Hawkins-Daarud, Sonal S. Noticewala, Russell C. Rockne, Sandra K. Johnston, Luis Gonzalez-Cuyar, Joseph Juliano, Orlando Gil, Kristin R. Swanson, Peter Canoll, and Alexander R. A. Anderson. From cells to tissue: How cell scale heterogeneity impacts glioblastoma growth and treatment response. *PLOS Computational Biology*, 16(2):e1007672, February 2020. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1007672.
- [91] Matthew H. Bailey, Collin Tokheim, Eduard Porta-Pardo, Sohini Sengupta, Denis Bertrand, Amila Weerasinghe, Antonio Colaprico, Michael C. Wendl, Jaegil

Kim, Brendan Reardon, Patrick Kwok-Shing Ng, Kang Jin Jeong, Song Cao, Zixing Wang, Jianjiong Gao, Qingsong Gao, Fang Wang, Eric Minwei Liu, Loris Mularoni, Carlota Rubio-Perez, Niranjana Nagarajan, Isidro Cortés-Ciriano, Daniel Cui Zhou, Wen-Wei Liang, Julian M. Hess, Venkata D. Yellapantula, David Tamborero, Abel Gonzalez-Perez, Chayaporn Suphavitai, Jia Yu Ko, Ekta Khurana, Peter J. Park, Eliezer M. Van Allen, Han Liang, Samantha J. Caesar-Johnson, John A. Demchok, Ina Felau, Melpomeni Kasapi, Martin L. Ferguson, Carolyn M. Hutter, Heidi J. Sofia, Roy Tarnuzzer, Zhining Wang, Liming Yang, Jean C. Zenklusen, Jia-shan (Julia) Zhang, Sudha Chudamani, Jia Liu, Laxmi Lolla, Rashmi Naresh, Todd Pihl, Qiang Sun, Yunhu Wan, Ye Wu, Juok Cho, Timothy DeFreitas, Scott Frazer, Nils Gehlenborg, Gad Getz, David I. Heiman, Jaegil Kim, Michael S. Lawrence, Pei Lin, Sam Meier, Michael S. Noble, Gordon Saksena, Doug Voet, Hailei Zhang, Brady Bernard, Nyasha Chambwe, Varsha Dhankani, Theo Knijnenburg, Roger Kramer, Kalle Leinonen, Yuexin Liu, Michael Miller, Sheila Reynolds, Ilya Shmulevich, Vesteinn Thorsson, Wei Zhang, Rehan Akbani, Bradley M. Broom, Apurva M. Hegde, Zhenlin Ju, Rupa S. Kanchi, Anil Korkut, Jun Li, Han Liang, Shiyun Ling, Wenbin Liu, Yiling Lu, Gordon B. Mills, Kwok-Shing Ng, Arvind Rao, Michael Ryan, Jing Wang, John N. Weinstein, Jiexin Zhang, Adam Abeshouse, Joshua Armenia, Debyani Chakravarty, Walid K. Chatila, Ino de Bruijn, Jianjiong Gao, Benjamin E. Gross, Zachary J. Heins, Ritika Kundra, Konnor La, Marc Ladanyi, Augustin Luna, Moriah G. Nissan, Angelica Ochoa, Sarah M. Phillips, Ed Reznik, Francisco Sanchez-Vega, Chris Sander, Nikolaus Schultz, Robert Sheridan, S. Onur Sumer, Yichao Sun, Barry S. Taylor, Jioajiao Wang, Hongxin Zhang, Pavana Anur, Myron Peto, Paul Spellman, Christopher Benz, Joshua M. Stuart, Christopher K. Wong, Christina Yau, D. Neil Hayes, Joel S. Parker, Matthew D. Wilkerson, Adrian Ally, Miruna Balasundaram, Reanne Bowlby, Denise Brooks, Rebecca Carlsen, Eric Chuah, Noreen Dhalla, Robert Holt, Steven J. M. Jones, Katayoon Kasaian, Darlene Lee, Yussanne Ma, Marco A. Marra, Michael Mayo, Richard A. Moore, Andrew J. Mungall, Karen Mungall, A. Gordon Robertson, Sara Sadeghi, Jacqueline E. Schein, Payal Sipahimalani, Angela Tam, Nina Thiessen, Kane Tse, Tina Wong, Ashton C. Berger, Rameen Beroukhim, Andrew D. Cherniack, Carrie Cibulskis, Stacey B. Gabriel, Galen F. Gao, Gavin Ha, Matthew Meyerson, Steven E. Schumacher, Juliann Shih, Melanie H. Kucherlapati, Raju S. Kucherlapati, Stephen Baylin, Leslie Cope, Ludmila Danilova, Moiz S. Bootwalla, Phillip H. Lai, Dennis T. Maglinte, David J. Van Den Berg, Daniel J. Weisenberger, J. Todd Auman, Saianand Balu, Tom Bodenheimer, Cheng Fan, Katherine A. Hoadley, Alan P. Hoyle, Stuart R. Jefferys, Corbin D. Jones, Shaowu Meng, Piotr A. Mieczkowski, Lisle E. Mose, Amy H. Perou, Charles M. Perou, Jeffrey Roach, Yan Shi, Janae V. Simons, Tara Skelly, Matthew G. Soloway, Donghui Tan, Umadevi Veluvolu, Huihui Fan, Toshinori Hinoue, Peter W. Laird, Hui Shen, Wanding Zhou, Michelle Bellair, Kyle Chang, Kyle Covington, Chad J. Creighton, Huyen Dinh, HarshaVardhan Doddapaneni, Lawrence A. Donehower, Jennifer Drummond, Richard A. Gibbs,

Robert Glenn, Walker Hale, Yi Han, Jianhong Hu, Viktoriya Korchina, Sandra Lee, Lora Lewis, Wei Li, Xiuping Liu, Margaret Morgan, Donna Morton, Donna Muzny, Jireh Santibanez, Margi Sheth, Eve Shinbrot, Linghua Wang, Min Wang, David A. Wheeler, Liu Xi, Fengmei Zhao, Julian Hess, Elizabeth L. Appelbaum, Matthew Bailey, Matthew G. Cordes, Li Ding, Catrina C. Fronick, Lucinda A. Fulton, Robert S. Fulton, Cyriac Kandoth, Elaine R. Mardis, Michael D. McLellan, Christopher A. Miller, Heather K. Schmidt, Richard K. Wilson, Daniel Crain, Erin Curley, Johanna Gardner, Kevin Lau, David Mallery, Scott Morris, Joseph Paulauskis, Robert Penny, Candace Shelton, Troy Shelton, Mark Sherman, Eric Thompson, Peggy Yena, Jay Bowen, Julie M. Gastier-Foster, Mark Gerken, Kristen M. Leraas, Tara M. Lichtenberg, Nilsa C. Ramirez, Lisa Wise, Erik Zmuda, Niall Corcoran, Tony Costello, Christopher Hovens, Andre L. Carvalho, Ana C. de Carvalho, José H. Fregnani, Adhemar Longatto-Filho, Rui M. Reis, Cristovam Scapulatempo-Neto, Henrique C. S. Silveira, Daniel O. Vidal, Andrew Burnette, Jennifer Eschbacher, Beth Hermes, Ardene Noss, Rosy Singh, Matthew L. Anderson, Patricia D. Castro, Michael Ittmann, David Huntsman, Bernard Kohl, Xuan Le, Richard Thorp, Chris Andry, Elizabeth R. Duffy, Vladimir Lyadov, Oxana Paklina, Galiya Setdikova, Alexey Shabunin, Mikhail Tavobilov, Christopher McPherson, Ronald Warnick, Ross Berkowitz, Daniel Cramer, Colleen Feltmate, Neil Horowitz, Adam Kibel, Michael Muto, Chandrajit P. Raut, Andrei Malykh, Jill S. Barnholtz-Sloan, Wendi Barrett, Karen Devine, Jordonna Fulop, Quinn T. Ostrom, Kristen Shimmel, Yingli Wolinsky, Andrew E. Sloan, Agostino De Rose, Felice Giuliente, Marc Goodman, Beth Y. Karlan, Curt H. Hagedorn, John Eckman, Jodi Harr, Jerome Myers, Kelinda Tucker, Leigh Anne Zach, Brenda Deyarmin, Hai Hu, Leonid Kvecher, Caroline Larson, Richard J. Mural, Stella Somiari, Ales Vicha, Tomas Zelinka, Joseph Bennett, Mary Iacocca, Brenda Rabeno, Patricia Swanson, Mathieu Latour, Louis Lacombe, Bernard Têtu, Alain Bergeron, Mary McGraw, Susan M. Staugaitis, John Chabot, Hanina Hibshoosh, Antonia Sepulveda, Tao Su, Timothy Wang, Olga Potapova, Olga Voronina, Laurence Desjardins, Odette Mariani, Sergio Roman-Roman, Xavier Sastre, Marc-Henri Stern, Feixiong Cheng, Sabina Signoretti, Andrew Berchuck, Darell Bigner, Eric Lipp, Jeffrey Marks, Shannon McCall, Roger McLendon, Angeles Secord, Alexis Sharp, Madhusmita Behera, Daniel J. Brat, Amy Chen, Keith Delman, Seth Force, Fadlo Khuri, Kelly Magliocca, Shishir Maithel, Jeffrey J. Olson, Taofeek Owonikoko, Alan Pickens, Suresh Ramalingam, Dong M. Shin, Gabriel Sica, Erwin G. Van Meir, Hongzheng Zhang, Wil Eijckenboom, Ad Gillis, Esther Korpershoek, Leendert Looijenga, Wolter Oosterhuis, Hans Stoop, Kim E. van Kessel, Ellen C. Zwarthoff, Chiara Calatozzolo, Lucia Cuppini, Stefania Cuzzubbo, Francesco DiMeco, Gaetano Finocchiaro, Luca Mattei, Alessandro Perin, Bianca Pollo, Chu Chen, John Houck, Pawadee Lohavanichbutr, Arndt Hartmann, Christine Stoehr, Robert Stoehr, Helge Taubert, Sven Wach, Bernd Wullich, Witold Kykler, Dawid Murawa, Maciej Wiznerowicz, Ki Chung, W. Jeffrey Edenfield, Julie Martin, Eric Baudin, Glenn Bubley, Raphael Bueno, Assunta De Rienzo, William G. Richards, Steven Kalkanis, Tom

Mikkelsen, Houtan Noushmehr, Lisa Scarpace, Nicolas Girard, Marta Aymerich, Elias Campo, Eva Giné, Armando López Guillermo, Nguyen Van Bang, Phan Thi Hanh, Bui Duc Phu, Yufang Tang, Howard Colman, Kimberley Evason, Peter R. Dottino, John A. Martignetti, Hani Gabra, Hartmut Juhl, Teniola Akeredolu, Serghei Stepa, Dave Hoon, Keunsoo Ahn, Koo Jeong Kang, Felix Beuschlein, Anne Breggia, Michael Birrer, Debra Bell, Mitesh Borad, Alan H. Bryce, Erik Castle, Vishal Chandan, John Cheville, John A. Copland, Michael Farnell, Thomas Flotte, Nasra Giamma, Thai Ho, Michael Kendrick, Jean-Pierre Kocher, Karla Kopp, Catherine Moser, David Nagorney, Daniel O'Brien, Brian Patrick O'Neill, Tushar Patel, Gloria Petersen, Florencia Que, Michael Rivera, Lewis Roberts, Robert Smallridge, Thomas Smyrk, Melissa Stanton, R. Houston Thompson, Michael Torben-son, Ju Dong Yang, Lizhi Zhang, Fadi Brimo, Jaffer A. Ajani, Ana Maria Angulo Gonzalez, Carmen Behrens, Jolanta Bondaruk, Russell Broaddus, Bogdan Czer-niak, Bitá Esmali, Junya Fujimoto, Jeffrey Gershenwald, Charles Guo, Alexander J. Lazar, Christopher Logothetis, Funda Meric-Bernstam, Cesar Moran, Lois Ramon-detta, David Rice, Anil Sood, Pheroze Tamboli, Timothy Thompson, Patricia Tron-coso, Anne Tsao, Ignacio Wistuba, Candace Carter, Lauren Haydu, Peter Hersey, Valerie Jakrot, Hojabr Kakavand, Richard Kefford, Kenneth Lee, Georgina Long, Graham Mann, Michael Quinn, Robyn Saw, Richard Scolyer, Kerwin Shannon, An-drew Spillane, Jonathan Stretch, Maria Synott, John Thompson, James Wilmott, Hikmat Al-Ahmadie, Timothy A. Chan, Ronald Ghossein, Anuradha Gopalan, Dou-glas A. Levine, Victor Reuter, Samuel Singer, Bhuvanesh Singh, Nguyen Viet Tien, Thomas Broudy, Cyrus Mirsaidi, Praveen Nair, Paul Drwiega, Judy Miller, Jen-nifer Smith, Howard Zaren, Joong-Won Park, Nguyen Phi Hung, Electron Kebe-bew, W. Marston Linehan, Adam R. Metwalli, Karel Pacak, Peter A. Pinto, Mark Schiffman, Laura S. Schmidt, Cathy D. Vocke, Nicolas Wentzensen, Robert Worrell, Hannah Yang, Marc Moncrieff, Chandra Goparaju, Jonathan Melamed, Harvey Pass, Natalia Botnariuc, Irina Caraman, Mircea Cernat, Inga Chemencedji, Adrian Clipca, Serghei Doruc, Ghenadie Gorincioi, Sergiu Mura, Maria Pirtac, Irina Stancul, Di-ana Tcaciuc, Monique Albert, Iakovina Alexopoulou, Angel Arnaout, John Bartlett, Jay Engel, Sebastien Gilbert, Jeremy Parfitt, Harman Sekhon, George Thomas, Doris M. Rassl, Robert C. Rintoul, Carlo Bifulco, Raina Tamakawa, Walter Urba, Nicholas Hayward, Henri Timmers, Anna Antenucci, Francesco Facciolo, Gianluca Grazi, Mirella Marino, Roberta Merola, Ronald de Krijger, Anne-Paule Gimenez-Roqueplo, Alain Piché, Simone Chevalier, Ginette McKercher, Kivanc Birsoy, Gene Barnett, Cathy Brewer, Carol Farver, Theresa Naska, Nathan A. Pennell, Daniel Raymond, Cathy Schilero, Kathy Smolenski, Felicia Williams, Carl Morrison, Jef-frey A. Borgia, Michael J. Liptay, Mark Pool, Christopher W. Seder, Kerstin Junker, Larsson Omberg, Mikhail Dinkin, George Manikhas, Domenico Alvaro, Maria Con-siglia Bragazzi, Vincenzo Cardinale, Guido Carpino, Eugenio Gaudio, David Chesla, Sandra Cottingham, Michael Dubina, Fedor Moiseenko, Renumathy Dhanasekaran, Karl-Friedrich Becker, Klaus-Peter Janssen, Julia Slotta-Huspenina, Mohamed H.

Abdel-Rahman, Dina Aziz, Sue Bell, Colleen M. Cebulla, Amy Davis, Rebecca Duell, J. Bradley Elder, Joe Hilty, Bahavna Kumar, James Lang, Norman L. Lehman, Randy Mandt, Phuong Nguyen, Robert Pilarski, Karan Rai, Lynn Schoenfield, Kelly Senecal, Paul Wakely, Paul Hansen, Ronald Lechan, James Powers, Arthur Tischler, William E. Grizzle, Katherine C. Sexton, Alison Kastl, Joel Henderson, Sima Porten, Jens Waldmann, Martin Fassnacht, Sylvia L. Asa, Dirk Schadendorf, Marta Couce, Markus Graefen, Hartwig Huland, Guido Sauter, Thorsten Schlomm, Ronald Simon, Pierre Tennstedt, Oluwale Olabode, Mark Nelson, Oliver Bathe, Peter R. Carroll, June M. Chan, Philip Disaia, Pat Glenn, Robin K. Kelley, Charles N. Landen, Joanna Phillips, Michael Prados, Jeffry Simko, Karen Smith-McCune, Scott Vandenberg, Kevin Roggin, Ashley Fehrenbach, Ady Kendler, Suzanne Sifri, Ruth Steele, Antonio Jimeno, Francis Carey, Ian Forgie, Massimo Mannelli, Michael Carney, Brenda Hernandez, Benito Campos, Christel Herold-Mende, Christin Jungk, Andreas Unterberg, Andreas von Deimling, Aaron Bossler, Joseph Galbraith, Laura Jacobus, Michael Knudson, Tina Knutson, Deqin Ma, Mohammed Milhem, Rita Sigmund, Andrew K. Godwin, Rashna Madan, Howard G. Rosenthal, Clement Adebamowo, Sally N. Adebamowo, Alex Boussioutas, David Beer, Thomas Giordano, Anne-Marie Mes-Masson, Fred Saad, Therese Bocklage, Lisa Landrum, Robert Mannel, Kathleen Moore, Katherine Moxley, Russel Postier, Joan Walker, Rosemary Zuna, Michael Feldman, Federico Valdivieso, Rajiv Dhir, James Luketich, Edna M. Mora Pinero, Mario Quintero-Aguilo, Carlos Gilberto Carlotti, Jose Sebastião Dos Santos, Rafael Kemp, Ajith Sankarankuty, Daniela Tirapelli, James Catto, Kathy Agnew, Elizabeth Swisher, Jenette Creaney, Bruce Robinson, Carl Simon Shelley, Eryn M. Godwin, Sara Kendall, Cassaundra Shipman, Carol Bradford, Thomas Carey, Andrea Haddad, Jeffrey Moyer, Lisa Peterson, Mark Prince, Laura Rozek, Gregory Wolf, Rayleen Bowman, Kwun M. Fong, Ian Yang, Robert Korst, W. Kimryn Rathmell, J. Leigh Fantacone-Campbell, Jeffrey A. Hooke, Albert J. Kovatich, Craig D. Shriver, John DiPersio, Bettina Drake, Ramaswamy Govindan, Sharon Heath, Timothy Ley, Brian Van Tine, Peter Westervelt, Mark A. Rubin, Jung Il Lee, Natália D. Aredes, Aramaz Mariamidze, Michael S. Lawrence, Adam Godzik, Nuria Lopez-Bigas, Josh Stuart, David Wheeler, Gad Getz, Ken Chen, Alexander J. Lazar, Gordon B. Mills, Rachel Karchin, and Li Ding. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, 173(2):371–385.e18, April 2018. ISSN 0092-8674. doi: 10.1016/j.cell.2018.02.060.

- [92] John G Tate, Sally Bamford, Harry C Jubb, Zbyslaw Sondka, David M Beare, Nidhi Bindal, Harry Boutselakis, Charlotte G Cole, Celestino Creatore, Elisabeth Dawson, Peter Fish, Bhavana Harsha, Charlie Hathaway, Steve C Jupe, Chai Yin Kok, Kate Noble, Laura Ponting, Christopher C Ramshaw, Claire E Rye, Helen E Speedy, Ray Stefancsik, Sam L Thompson, Shicai Wang, Sari Ward, Peter J Campbell, and Simon A Forbes. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, 47(D1):D941–D947, January 2019. ISSN 0305-1048. doi: 10.1093/nar/gky1015.

- [93] Michael W. Schmitt, Scott R. Kennedy, Jesse J. Salk, Edward J. Fox, Joseph B. Hiatt, and Lawrence A. Loeb. Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences*, 109(36):14508–14513, September 2012. doi: 10.1073/pnas.1208715109.
- [94] Scott R. Kennedy, Michael W. Schmitt, Edward J. Fox, Brendan F. Kohn, Jesse J. Salk, Eun Hyun Ahn, Marc J. Prindle, Kawai J. Kuong, Jiang-Cheng Shen, Rosa-Ana Risques, and Lawrence A. Loeb. Detecting ultralow-frequency mutations by Duplex Sequencing. *Nature Protocols*, 9(11):2586–2606, November 2014. ISSN 1750-2799. doi: 10.1038/nprot.2014.170.
- [95] Jeffrey D. Krimmel, Michael W. Schmitt, Maria I. Harrell, Kathy J. Agnew, Scott R. Kennedy, Mary J. Emond, Lawrence A. Loeb, Elizabeth M. Swisher, and Rosa Ana Risques. Ultra-deep sequencing detects ovarian cancer cells in peritoneal fluid and reveals somatic TP53 mutations in noncancerous tissues. *Proceedings of the National Academy of Sciences*, 113(21):6005–6010, May 2016. doi: 10.1073/pnas.1601311113.
- [96] Talayeh S. Ghezelayagh, Brendan F. Kohn, Jeanne Fredrickson, Enna Manhardt, Marc R. Radke, Ronit Katz, Heidi J. Gray, Renata R. Urban, Kathryn P. Pennington, John B. Liao, Kemi M. Doll, Elise J. Simons, Jennifer K. Burzawa, Barbara A. Goff, Paul Speiser, Elizabeth M. Swisher, Barbara M. Norquist, and Rosa Ana Risques. Uterine Lavage Identifies Cancer Mutations and Increased TP53 Somatic Mutation Burden in Individuals with Ovarian Cancer. *Cancer Research Communications*, 2(10):1282–1292, October 2022. ISSN 2767-9764. doi: 10.1158/2767-9764.CRC-22-0314.
- [97] M. Sala-Trepat, J. Boyse, P. Richard, D. Papadopoulo, and E. Moustacchi. Frequencies of HPRT lymphocytes and glycophorin A variants erythrocytes in Fanconi anemia patients, their parents and control donors. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 289(1):115–126, September 1993. ISSN 0027-5107. doi: 10.1016/0027-5107(93)90137-5.
- [98] David J. Araten, David W. Golde, Rong H. Zhang, Howard T. Thaler, Lucia Gargiulo, Rosario Notaro, and Lucio Luzzatto. A Quantitative Measurement of the Human Somatic Mutation Rate. *Cancer Research*, 65(18):8111–8117, September 2005. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-04-1198.
- [99] Ilaria Persico, Ilaria Fiscarelli, Alessandra Pelle, Michela Faleschini, Barbara Pasini, Anna Savoia, and Roberta Bottega. Phenotype reversion as “natural gene therapy” in Fanconi anemia by a gene conversion event. *Frontiers in Genetics*, 14, September 2023. ISSN 1664-8021. doi: 10.3389/fgene.2023.1240758.
- [100] Quinten Waisfisz, Neil V. Morgan, Maria Savino, Johan P. de Winter, Carola G. M. van Berkel, Maureen E. Hoatlin, Leonarda Ianzano, Rachel A. Gibson, Fre Arwert, Anna Savoia, Christopher G. Mathew, Jan C. Pronk, and Hans Joenje. Spontaneous

- functional correction of homozygous Fanconi anaemia alleles reveals novel mechanistic basis for reverse mosaicism. *Nature Genetics*, 22(4):379–383, August 1999. ISSN 1546-1718. doi: 10.1038/11956.
- [101] John J. Gregory, John E. Wagner, Peter C. Verlander, Orna Levran, Sat Dev Batish, Cindy R. Eide, Amy Steffenhagen, Betsy Hirsch, and Arleen D. Auerbach. Somatic mosaicism in Fanconi anemia: Evidence of genotypic reversion in lymphohematopoietic stem cells. *Proceedings of the National Academy of Sciences*, 98(5):2532–2537, February 2001. doi: 10.1073/pnas.051609898.
 - [102] M. Gross, H. Hanenberg, S. Lobitz, R. Friedl, S. Herterich, R. Dietrich, B. Gruhn, D. Schindler, and H. Hoehn. Reverse mosaicism in Fanconi anemia: natural gene therapy via molecular self-correction. *Cytogenetics and Cell Genetics*, 98(2-3):126–135, April 2003. ISSN 0301-0171. doi: 10.1159/000069805.
 - [103] Jean Soulier, Thierry Leblanc, Jérôme Larghero, Hélène Dastot, Akiko Shimamura, Philippe Guardiola, Hélène Esperou, Christèle Ferry, Charlotte Jubert, Jean-Paul Feugeas, Annie Henri, Antoine Toubert, Gérard Socié, André Baruchel, François Sigaux, Alan D. D’Andrea, and Eliane Gluckman. Detection of somatic mosaicism and classification of Fanconi anemia patients by analysis of the FA/BRCA pathway. *Blood*, 105(3):1329–1336, February 2005. ISSN 0006-4971. doi: 10.1182/blood-2004-05-1852.
 - [104] Eileen Nicoletti, Gayatri Rao, Juan A. Bueren, Paula Río, Susana Navarro, Jordi Surrallés, Grace Choi, and Jonathan D. Schwartz. Mosaicism in Fanconi anemia: concise review and evaluation of published cases with focus on clinical course of blood count normalization. *Annals of Hematology*, 99(5):913–924, May 2020. ISSN 1432-0584. doi: 10.1007/s00277-020-03954-2.
 - [105] Marcel F Jonkman, Hans Scheffer, Rein Stulp, Hendri H Pas, Miranda Nijenhuis, Klaas Heeres, Katsushi Owaribe, Leena Pulkkinen, and Jouni Uitto. Revertant Mosaicism in Epidermolysis Bullosa Caused by Mitotic Gene Conversion. *Cell*, 88(4): 543–551, February 1997. ISSN 0092-8674. doi: 10.1016/S0092-8674(00)81894-2.
 - [106] Keith A. Choate, Yin Lu, Jing Zhou, Murim Choi, Peter M. Elias, Anita Farhi, Carol Nelson-Williams, Debra Crumrine, Mary L. Williams, Amy J. Nopper, Alanna Bree, Leonard M. Milstone, and Richard P. Lifton. Mitotic Recombination in Patients with Ichthyosis Causes Reversion of Dominant Mutations in KRT10. *Science*, 330(6000):94–97, October 2010. doi: 10.1126/science.1192280.
 - [107] Peter C. van den Akker, Maria C. Bolling, and Anna M. G. Pasmooij. Revertant Mosaicism in Genodermatoses: Natural Gene Therapy Right before Your Eyes. *Biomedicines*, 10(9):2118, September 2022. ISSN 2227-9059. doi: 10.3390/biomedicines10092118.

- [108] Anna M. G. Pasmooij, Marta Garcia, Maria J. Escamez, A. Miranda Nijenhuis, Antoni Azon, Natividad Cuadrado-Corrales, Marcel F. Jonkman, and Marcela Del Rio. Revertant Mosaicism Due to a Second-Site Mutation in *COL7A1* in a Patient with Recessive Dystrophic Epidermolysis Bullosa. *Journal of Investigative Dermatology*, 130(10):2407–2411, October 2010. ISSN 0022-202X. doi: 10.1038/jid.2010.163.
- [109] Kai J. Rogers, Wenqing Fu, Joshua M. Akey, and Raymond J. Monnat, Jr. Global and disease-associated genetic variation in the human Fanconi anemia gene family. *Human Molecular Genetics*, 23(25):6815–6825, December 2014. ISSN 0964-6906. doi: 10.1093/hmg/ddu400.
- [110] María José Ramírez, Roser Pujol, Juan Pablo Trujillo-Quintero, Jordi Minguillón, Massimo Bogliolo, Paula Río, Susana Navarro, José A. Casado, Isabel Badell, Estela Carrasco, Judith Balmaña, Albert Català, Julián Sevilla, Cristina Beléndez, Bienvenida Argilés, Mónica López, Cristina Díaz de Heredia, Gayatri Rao, Eileen Nicoletti, Jonathan D. Schwartz, Juan A. Bueren, and Jordi Surrallés. Natural gene therapy by reverse mosaicism leads to improved hematology in Fanconi anemia patients. *American Journal of Hematology*, 96(8):989–999, 2021. ISSN 1096-8652. doi: 10.1002/ajh.26234.
- [111] Anuj Mankad, Toshiyasu Taniguchi, Barbara Cox, Yassmine Akkari, R. Keaney Rathbun, Lora Lucas, Grover Bagby, Susan Olson, Alan D’Andrea, and Markus Grompe. Natural gene therapy in monozygotic twins with Fanconi anemia. *Blood*, 107(8):3084–3090, April 2006. ISSN 0006-4971. doi: 10.1182/blood-2005-07-2638.
- [112] Paula Río, Susana Navarro, Wei Wang, Rebeca Sánchez-Domínguez, Roser M. Pujol, José C. Segovia, Massimo Bogliolo, Eva Merino, Ning Wu, Rocío Salgado, María L. Lamana, Rosa M. Yáñez, José A. Casado, Yari Giménez, Francisco J. Román-Rodríguez, Lara Álvarez, Omaira Alberquilla, Anna Raimbault, Guillermo Guenechea, M. Luz Lozano, Laura Cerrato, Miriam Hernando, Eva Gálvez, Raquel Hladun, Irina Giralt, Jordi Barquinero, Anne Galy, Nagore García de Andoín, Ricardo López, Albert Catalá, Jonathan D. Schwartz, Jordi Surrallés, Jean Soulier, Manfred Schmidt, Cristina Díaz de Heredia, Julián Sevilla, and Juan A. Bueren. Successful engraftment of gene-corrected hematopoietic stem cells in non-conditioned patients with Fanconi anemia. *Nature Medicine*, 25(9):1396–1401, September 2019. ISSN 1546-170X. doi: 10.1038/s41591-019-0550-z.
- [113] Paula Río, Josune Zubicaray, Susana Navarro, Eva Gálvez, Rebeca Sánchez-Domínguez, Eileen Nicoletti, Elena Sebastián, Michael Rothe, Roser Pujol, Massimo Bogliolo, Philipp John-Neek, Antonella Lucía Bastone, Axel Schambach, Wei Wang, Manfred Schmidt, Lise Larcher, José C. Segovia, Rosa M. Yáñez, Omaira Alberquilla, Begoña Díez, María Fernández-García, Laura García-García, Manuel Ramírez, Anne Galy, Francois Lefrere, Marina Cavazzana, Thierry Leblanc, Nagore

- García de Andoin, Ricardo López-Almaraz, Albert Catalá, Jordi Barquinero, Sandra Rodríguez-Perales, Gayatri Rao, Jordi Surrallés, Jean Soulier, Cristina Díaz-de Heredia, Jonathan D. Schwartz, Julián Sevilla, Juan A. Bueren, Paula Río, Josune Zubicaray, Susana Navarro, Eva Gálvez, Rebeca Sánchez-Domínguez, Eileen Nicoletti, Elena Sebastián, Michael Rothe, Roser Pujol, Massimo Bogliolo, Philipp John-Neek, Antonella L. Bastone, Axel Schambach, Wei Wang, Manfred Schmidt, Lise Larcher, José C. Segovia, Rosa M. Yáñez, Florencio Pérez-Maroto, Ana de la Cruz, José A. Casado, Yari Giménez, Lara Álvarez, Omaira Alberquilla, Begoña Díez, María Fernández-García, Laura García-García, Ana Gómez, Almudena Galán, Manuel Ramírez, Rocío Salgado, Anne Galy, Francois Lefrere, Marina Cavazzana, Thierry Leblanc, Nagore Garcia de Andoin, Ricardo López-Almaraz, Albert Catalá, Jordi Barquinero, Sandra Rodríguez-Perales, Gayatri Rao, Jordi Surrallés, Jean Soulier, Cristina Díaz-de Heredia, Jonathan D. Schwartz, Julián Sevilla, and Juan A. Bueren. Haematopoietic gene therapy of non-conditioned patients with Fanconi anaemia-A: results from open-label phase 1/2 (FANCOLEN-1) and long-term clinical trials. *The Lancet*, 404(10471):2584–2592, December 2024. ISSN 0140-6736, 1474-547X. doi: 10.1016/S0140-6736(24)01880-4.
- [114] Paula Río, Susana Navarro, Guillermo Guenechea, Rebeca Sánchez-Domínguez, Maria Luisa Lamana, Rosa Yáñez, Jose A. Casado, Parinda A. Mehta, Maria Roser Pujol, Jordi Surrallés, Sabine Charrier, Anne Galy, José C. Segovia, Cristina Díaz de Heredia, Julián Sevilla, and Juan A. Bueren. Engraftment and in vivo proliferation advantage of gene-corrected mobilized CD34+ cells from Fanconi anemia patients. *Blood*, 130(13):1535–1542, September 2017. ISSN 0006-4971. doi: 10.1182/blood-2017-03-774174.
- [115] Paula Río, Néstor W. Meza, África González-Murillo, Susana Navarro, Lara Álvarez, Jordi Surrallés, Maria Castella, Guillermo Guenechea, José C. Segovia, Helmut Hahnenberg, and Juan A. Bueren. In vivo proliferation advantage of genetically corrected hematopoietic stem cells in a mouse model of Fanconi anemia FA-D1. *Blood*, 112(13):4853–4861, December 2008. ISSN 0006-4971. doi: 10.1182/blood-2008-05-156356.
- [116] Miren Lasaga, Paula Río, Amaia Vilas-Zornoza, Nuria Planell, Susana Navarro, Diego Alignani, Beatriz Fernández-Varas, Daniel Mouzo, Josune Zubicaray, Roser M. Pujol, Eileen Nicoletti, Jonathan D. Schwartz, Julián Sevilla, Marina Ainciburi, Asier Ullate-Agote, Jordi Surrallés, Rosario Perona, Leandro Sastre, Felipe Prosper, David Gomez-Cabrero, and Juan A. Bueren. Gene therapy restores the transcriptional program of hematopoietic stem cells in Fanconi anemia. *Haematologica*, 108(10):2652–2663, April 2023. ISSN 1592-8721. doi: 10.3324/haematol.2022.282418.
- [117] Tobias Hirsch, Tobias Rothoeft, Norbert Teig, Johann W. Bauer, Graziella Pellegrini, Laura De Rosa, Davide Scaglione, Julia Reichelt, Alfred Klausegger, Daniela Kneisz, Oriana Romano, Alessia Secone Seconetti, Roberta Contin, Elena Enzo,

- Irena Jurman, Sonia Carulli, Frank Jacobsen, Thomas Luecke, Marcus Lehnhardt, Meike Fischer, Maximilian Kueckelhaus, Daniela Quaglino, Michele Morgante, Silvio Biciato, Sergio Bondanza, and Michele De Luca. Regeneration of the entire human epidermis using transgenic stem cells. *Nature*, 551(7680):327–332, November 2017. ISSN 1476-4687. doi: 10.1038/nature24487.
- [118] Hunter L. Colegrove, Marianne E. Dubard-Gault, Henry Marshall, Brendan F. Kohn, Thomas H. Smith, Zachary K. Norgaard, Fang Yin Lo, Elizabeth K. Schmidt, Jacob E. Higgins, Charles C. Valentine, Desiree A. Marshall, John I. Clark, Eric Q. Konnick, Jesse J. Salk, Marshall S. Horwitz, Raheleh Rahbari, Alison F. Feder, and Rosa Ana Risques. Ultra-deep duplex sequencing reveals unique features of somatic evolution in the normal tissues of a family with Li-Fraumeni syndrome. *bioRxiv*, January 2026. doi: 10.64898/2026.01.12.699071.
- [119] Hunter L. Colegrove, Raymond J. Monnat Jr, and Alison F. Feder. Epithelial competition determines gene therapy potential to suppress Fanconi Anemia oral cancer risk. *PLOS Computational Biology*, 21(9):e1012915, September 2025. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1012915.
- [120] Nabamita Boruah, David Hoyos, Renyta Moses, Ryan Hausler, Heena Desai, Anh N. Le, Madeline Good, Gregory Kelly, Ashvathi Raghavakaimal, Maliha Tayeb, Mohana Narasimhamurthy, Abigail Doucette, Peter Gabriel, Michael J. Feldman, Jinae Park, Miguel Lopez de Rodas, Kurt A. Schalper, Shari B. Goldfarb, Anupma Nayak, Arnold J. Levine, Benjamin D. Greenbaum, and Kara N. Maxwell. Distinct genomic and immunologic tumor evolution in germline TP53-driven breast cancers. *bioRxiv*, April 2024. doi: 10.1101/2024.04.03.588009.
- [121] Andrew R. J. Lawson, Federico Abascal, Pantelis A. Nicola, Stefanie V. Lensing, Amy L. Roberts, Georgios Kalantzis, Adrian Baez-Ortega, Natalia Brzozowska, Julia S. El-Sayed Moustafa, Dovile Vaitkute, Belma Jakupovic, Ayrin Nessa, Samuel Wadge, Marc F. Österdahl, Anna L. Paterson, Doris M. Rassl, Raul E. Alcantara, Laura O’Neill, Sara Widaa, Siobhan Austin-Guest, Matthew D. C. Neville, Moritz J. Przybilla, Wei Cheng, Maria Morra, Lucy Sykes, Matthew Mayho, Nicole Müller-Sienerth, Nicholas Williams, Diana Alexander, Luke M. R. Harvey, Thomas Clarke, Alex Byrne, Jamie R. Blundell, Matthew D. Young, Krishnaa T. A. Mahbubani, Kourosh Saeb-Parsy, Hilary C. Martin, Michael R. Stratton, Peter J. Campbell, Raheleh Rahbari, Kerrin S. Small, and Iñigo Martincorena. Somatic mutation and selection at population scale. *Nature*, pages 1–10, October 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09584-w.
- [122] Oleg Timofeev and Thorsten Stiewe. Rely on Each Other: DNA Binding Cooperativity Shapes p53 Functions in Tumor Suppression and Cancer Therapy. *Cancers*, 13(10):2422, January 2021. ISSN 2072-6694. doi: 10.3390/cancers13102422.

- [123] Boris Klimovich, Nastasja Merle, Michelle Neumann, Sabrina Elmshäuser, Andrea Nist, Marco Mernberger, Daniel Kazdal, Albrecht Stenzinger, Oleg Timofeev, and Thorsten Stiewe. p53 partial loss-of-function mutations sensitize to chemotherapy. *Oncogene*, 41(7):1011–1023, February 2022. ISSN 1476-5594. doi: 10.1038/s41388-021-02141-5.
- [124] Alexander Sun Zhang, Meis Omran, Cecilia Arthur, Helena Malmgren, Rita Barbosa-Matos, Susana Seixas, Carla Oliveira, Emma Tham, and Svetlana Bajalica-Lagercrantz. TP53 p.R181H is enriched in the Swedish cohort (SWEP53) and associated with a distinct breast and prostate phenotype. *Scientific Reports*, 15(1):35033, October 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-22407-2.
- [125] Claire Freycon, Laura Palma, Crystal Budd, Frederic Coulombe, Leora Witkowski, Pierre Hainaut, William D. Foulkes, and Catherine Goudie. Germline p.R181H variant in TP53 in a family exemplifying the genotype-phenotype correlations in Li-Fraumeni syndrome. *Familial Cancer*, 23(4):665–669, November 2024. ISSN 1573-7292. doi: 10.1007/s10689-024-00419-7.
- [126] Ambritha Balasundaram and C. George Priya Doss. Unraveling the Structural Changes in the DNA-Binding Region of Tumor Protein p53 (TP53) upon Hotspot Mutation p53 Arg248 by Comparative Computational Approach. *International Journal of Molecular Sciences*, 23(24):15499, January 2022. ISSN 1422-0067. doi: 10.3390/ijms232415499.
- [127] Michael D. Kessler, Amy Damask, Sean O’Keeffe, Nilanjana Banerjee, Dadong Li, Kyoko Watanabe, Anthony Marketta, Michael Van Meter, Stefan Semrau, Julie Horowitz, Jing Tang, Jack A. Kosmicki, Veera M. Rajagopal, Yuxin Zou, Yariv Hourvas, Arkopravo Ghosh, Christopher Gillies, Joelle Mbatchou, Ryan R. White, Niek Verweij, Jonas Bovijn, Neelroop N. Parikshak, Michelle G. LeBlanc, Marcus Jones, David J. Glass, Luca A. Lotta, Michael N. Cantor, Gurinder S. Atwal, Adam E. Locke, Manuel A. R. Ferreira, Raquel Deering, Charles Paulding, Alan R. Shuldiner, Gavin Thurston, Adolfo A. Ferrando, Will Salerno, Jeffrey G. Reid, John D. Overton, Jonathan Marchini, Hyun M. Kang, Aris Baras, Gonçalo R. Abecasis, and Eric Jorgenson. Common and rare variant associations with clonal haematopoiesis phenotypes. *Nature*, 612(7939):301–309, December 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-05448-9.
- [128] Jonatan Axelsson, Danielle LeBlanc, Habiballah Shojaeisaadi, Matthew J. Meier, Devon M. Fitzgerald, Daniela Nachmanson, Jedidiah Carlson, Alexandra Golubeva, Jake Higgins, Thomas Smith, Fang Yin Lo, Richard Pilsner, Andrew Williams, Jesse Salk, Francesco Marchetti, and Carole Yauk. Frequency and spectrum of mutations in human sperm measured using duplex sequencing correlate with trio-based de novo mutation analyses. *Scientific Reports*, 14(1):23134, October 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-73587-2.

- [129] Samuel F. M. Hart, Nicolas Alcala, Alison F. Feder, and Kelley Harris. A signature-agnostic test for differences between tumor mutation spectra reveals carcinogen and ancestry effects. *bioRxiv*, May 2025. doi: 10.1101/2025.05.15.654154.
- [130] Sujath Abbas, Oriol Pich, Ginny Devonshire, Shahriar A. Zamani, Annalise Katz-Summercorn, Sarah Killcoyne, Calvin Cheah, Barbara Nutzinger, Nicola Grehan, Nuria Lopez-Bigas, Rebecca C. Fitzgerald, and Maria Secrier. Mutational signature dynamics shaping the evolution of oesophageal adenocarcinoma. *Nature Communications*, 14(1):4239, July 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-39957-6.
- [131] Eric Rios-Doria, Elizabeth U. Parker, Brendan F. Kohn, Mindy Pike, Coohleen Coombes, Elena Latorre-Esteves, Daniel J. Reiter, Jeanne Fredrickson, Ronit Katz, Elizabeth M. Swisher, Kemi M. Doll, and Rosa Ana Risques. TP53 somatic evolution in the normal endometrium of Black and White individuals. *Gynecologic Oncology*, 197:1–10, June 2025. ISSN 0090-8258, 1095-6859. doi: 10.1016/j.ygyno.2025.04.002.
- [132] Akira Yokoyama, Koichi Watanabe, Yoshikage Inoue, Tomonori Hirano, Masashi Tamaoki, Kenshiro Hirohashi, Shun Kawaguchi, Yoshihiro Ishida, Yasuhide Takeuchi, Yo Kishimoto, Soo Ki Kim, Chikatoshi Katada, Yasuhito Nannya, Hiroshi Seno, Seishi Ogawa, Manabu Muto, and Nobuyuki Kakiuchi. Somatic mosaicism in the buccal mucosa reflects lifestyle and germline risk factors for esophageal squamous cell carcinoma. *Science Translational Medicine*, 17(796):eadq6740, April 2025. doi: 10.1126/scitranslmed.adq6740.
- [133] Renyta Moses, Alexandra Indeglia, Alison S. Levine, Ryan Hausler, Gregory Kelly, Sven A. Miller, Isabel Anez, Melissa Heller, Rosella Delgado, Caitlin Orr, Wendy Kohlmann, Anne Naumer, Jennie Vagher, Sophie R. Cahill, Luke D. Maese, John Karanicolas, Judy E. Garber, Maureen E. Murphy, and Kara N. Maxwell. Variation at the R181 residue of p53 confers loss of p53 DNA binding cooperativity with the retention of mitochondrial-associated apoptosis. *bioRxiv*, August 2025. doi: 10.1101/2025.08.12.669925.
- [134] Juan Bautista Menendez-Gonzalez, Milica Vukovic, Ali Abdelfattah, Lubaid Saleh, Alhomidi Almotiri, Leigh-anne Thomas, Aloña Agirre-Lizaso, Aleksandra Azevedo, Ana Catarina Menezes, Giusy Tornillo, Sarah Edkins, Kay Kong, Peter Giles, Fernando Anjos-Afonso, Alex Tonks, Ashleigh S. Boyd, Kamil R. Kranc, and Neil P. Rodrigues. Gata2 as a Crucial Regulator of Stem Cells in Adult Hematopoiesis and Acute Myeloid Leukemia. *Stem Cell Reports*, 13(2):291–306, August 2019. ISSN 2213-6711. doi: 10.1016/j.stemcr.2019.07.005.
- [135] Mark Redston, Amy Noffsinger, Anthony Kim, Fahire G. Akarca, Marianne Rara, Diane Stapleton, Laurel Nowden, Richard Lash, Adam J. Bass, and Matthew D. Stachler. Abnormal *TP53* Predicts Risk of Progression in Patients With Barrett’s

- Esophagus Regardless of a Diagnosis of Dysplasia. *Gastroenterology*, 162(2):468–481, February 2022. ISSN 0016-5085. doi: 10.1053/j.gastro.2021.10.038.
- [136] Matthew D. Stachler, Nicholas D. Camarda, Christopher Deitrick, Anthony Kim, Agoston T. Agoston, Robert D. Odze, Jason L. Hornick, Anwasha Nag, Aaron R. Thorner, Matthew Ducar, Amy Noffsinger, Richard H. Lash, Mark Redston, Scott L. Carter, Jon M. Davison, and Adam J. Bass. Detection of Mutations in Barrett’s Esophagus Before Progression to High-Grade Dysplasia or Adenocarcinoma. *Gastroenterology*, 155(1):156–167, July 2018. ISSN 0016-5085. doi: 10.1053/j.gastro.2018.03.047.
- [137] Takahiro Shimizu, Hiroyuki Marusawa, Yuko Matsumoto, Tadashi Inuzuka, Atsuyuki Ikeda, Yosuke Fujii, Sachiko Minamiguchi, Shin’ichi Miyamoto, Tadayuki Kou, Yoshiharu Sakai, Jean E. Crabtree, and Tsutomu Chiba. Accumulation of Somatic Mutations in *TP53* in Gastric Epithelium With *Helicobacter pylori* Infection. *Gastroenterology*, 147(2):407–417.e3, August 2014. ISSN 0016-5085. doi: 10.1053/j.gastro.2014.04.036.
- [138] C. Morgan, G. J. S. Jenkins, T. Ashton, A. P. Griffiths, J. N. Baxter, E. M. Parry, and J. M. Parry. Detection of p53 mutations in precancerous gastric tissue. *British Journal of Cancer*, 89(7):1314–1319, October 2003. ISSN 1532-1827. doi: 10.1038/sj.bjc.6601302.
- [139] Nancy Klemm, Roman R. Schimmer, Nils K. Konrad, Flavian Thelen, Jonas Fullin, Ebru Topçu, Christian Koch, Milena Treacy, Matthew Joseph Leventhal, Marco M. Bühler, Veronika Lysenko, Alexandre P.A. Theocharides, Kari J. Kurppa, Stefan Balabanov, Tuncay Baubec, Andrei V. Krivtsov, Peter G. Miller, Scott A. Armstrong, Benjamin L. Ebert, Markus G. Manz, Cesar Nombela-Arrieta, and Steffen Boettcher. The Prolonged Half-Life of the p53 Missense Variant R248Q Promotes Accumulation and Heterotetramer Formation with Wild-Type p53 to Exert the Dominant-Negative Effect. *Cancer Research*, 85(11):1978–1996, June 2025. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-24-1136.
- [140] Steffen Boettcher, Peter G. Miller, Rohan Sharma, Marie McConkey, Matthew Leventhal, Andrei V. Krivtsov, Andrew O. Giacomelli, Waihay Wong, Jesi Kim, Sherry Chao, Kari J. Kurppa, Xiaoping Yang, Kirsten Milenkowic, Federica Piccioni, David E. Root, Frank G. Rücker, Yael Flamand, Donna Neuberg, R. Coleman Lindley, Pasi A. Jänne, William C. Hahn, Tyler Jacks, Hartmut Döhner, Scott A. Armstrong, and Benjamin L. Ebert. A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. *Science*, 365(6453):599–604, August 2019. doi: 10.1126/science.aax3649.
- [141] Zhongwu Lai, Aleksandra Markovets, Miika Ahdesmaki, Brad Chapman, Oliver Hofmann, Robert McEwen, Justin Johnson, Brian Dougherty, J. Carl Barrett, and Jonathan R. Dry. VarDict: a novel and versatile variant caller for next-generation

- sequencing in cancer research. *Nucleic Acids Research*, 44(11):e108, June 2016. ISSN 0305-1048. doi: 10.1093/nar/gkw227.
- [142] Jun Cheng, Guido Novati, Joshua Pan, Clare Bycroft, Akvilė Žemgulytė, Taylor Applebaum, Alexander Pritzel, Lai Hong Wong, Michal Zielinski, Tobias Sargeant, Rosalia G. Schneider, Andrew W. Senior, John Jumper, Demis Hassabis, Pushmeet Kohli, and Žiga Avsec. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*, 381(6664):eadg7492, September 2023. doi: 10.1126/science.adg7492.
- [143] Mo Liu, Yang Wu, Nanhai Jiang, Arnoud Boot, and Steven G Rozen. mSigHdp: hierarchical Dirichlet process mixture modeling for mutational signature discovery. *NAR Genomics and Bioinformatics*, 5(1):lqad005, March 2023. ISSN 2631-9268. doi: 10.1093/nargab/lqad005.
- [144] S. M. Ashiqul Islam, Marcos Díaz-Gay, Yang Wu, Mark Barnes, Raviteja Vangara, Erik N. Bergstrom, Yudou He, Mike Vella, Jingwei Wang, Jon W. Teague, Peter Clapham, Sarah Moody, Sergey Senkin, Yun Rose Li, Laura Riva, Tongwu Zhang, Andreas J. Gruber, Christopher D. Steele, Burçak Otlı, Azhar Khandekar, Ammal Abbasi, Laura Humphreys, Natalia Syulyukina, Samuel W. Brady, Boian S. Alexandrov, Nischalan Pillay, Jinghui Zhang, David J. Adams, Iñigo Martincorena, David C. Wedge, Maria Teresa Landi, Paul Brennan, Michael R. Stratton, Steven G. Rozen, and Ludmil B. Alexandrov. Uncovering novel mutational signatures by *de novo* extraction with SigProfilerExtractor. *Cell Genomics*, 2(11):100179, November 2022. ISSN 2666-979X. doi: 10.1016/j.xgen.2022.100179.
- [145] Daniel E. Johnson, Barbara Burtneß, C. René Leemans, Vivian Wai Yan Lui, Julie E. Bauman, and Jennifer R. Grandis. Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers*, 6(1):1–22, November 2020. ISSN 2056-676X. doi: 10.1038/s41572-020-00224-3.
- [146] Eunike Velleuer, Elisa Domínguez-Hüttinger, Alfredo Rodríguez, Leonard A. Harris, and Carsten Carlberg. Concepts of multi-level dynamical modelling: understanding mechanisms of squamous cell carcinoma development in Fanconi anemia. *Frontiers in Genetics*, 14, November 2023. ISSN 1664-8021. doi: 10.3389/fgene.2023.1254966.
- [147] Philip S. Rosenberg, Mark H. Greene, and Blanche P. Alter. Cancer incidence in persons with Fanconi anemia. *Blood*, 101(3):822–826, February 2003. ISSN 0006-4971. doi: 10.1182/blood-2002-05-1498.
- [148] Andrew L. H. Webster, Mathijs A. Sanders, Krupa Patel, Ralf Dietrich, Raymond J. Noonan, Francis P. Lach, Ryan R. White, Audrey Goldfarb, Kevin Hadi, Matthew M. Edwards, Frank X. Donovan, Remco M. Hoogenboezem, Moonjung Jung, Sunandini Sridhar, Tom F. Wiley, Olivier Fedrigo, Huasong Tian, Joel Rosiene, Thomas Heineman, Jennifer A. Kennedy, Lorenzo Bean, Rasim O. Rosti,

- Rebecca Tryon, Ashlyn-Maree Gonzalez, Allana Rosenberg, Ji-Dung Luo, Thomas S. Carroll, Sanjana Shroff, Michael Beaumont, Eunike Velleuer, Jeff C. Rastatter, Susanne I. Wells, Jordi Surrallés, Grover Bagby, Margaret L. MacMillan, John E. Wagner, Maria Cancio, Farid Boulad, Theresa Scognamiglio, Roger Vaughan, Kristin G. Beaumont, Amnon Koren, Marcin Imielinski, Settara C. Chandrasekharappa, Arleen D. Auerbach, Bhuvanesh Singh, David I. Kutler, Peter J. Campbell, and Agata Smogorzewska. Genomic signature of Fanconi anaemia DNA repair pathway deficiency in cancer. *Nature*, 612(7940):495–502, December 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-05253-4.
- [149] Hester J. T. van Zeeburg, Peter J. F. Snijders, Thijs Wu, Eliane Gluckman, Jean Soulier, Jordi Surrallés, Maria Castella, Jacqueline E. van der Wal, Johan Wennerberg, Joseph Califano, Eunike Velleuer, Ralf Dietrich, Wolfram Ebell, Elisabeth Bloemena, Hans Joenje, C. René Leemans, and Ruud H. Brakenhoff. Clinical and Molecular Characteristics of Squamous Cell Carcinomas From Fanconi Anemia Patients. *JNCI: Journal of the National Cancer Institute*, 100(22):1649–1653, November 2008. ISSN 0027-8874. doi: 10.1093/jnci/djn366.
- [150] Hiep Tai Nguyen, Weiliang Tang, Andrew L. H. Webster, Jeffrey R. Whiteaker, Christopher M. Chandler, Ricardo Errazquin, Khashayar Roohollahi, Madeline Fritzke, Elizabeth E. Hoskins, Erica Jonlin, Leslie Wakefield, Lucas B. Sullivan, Eleanor Y. Chen, Josephine Dorsman, Ruud Brakenhoff, Amanda G. Paulovich, Markus Grompe, Ramon Garcia-Escudero, Susanne I. Wells, Agata Smogorzewska, and Raymond J. Monnat Jr. Fanconi anemia-isogenic head and neck cancer cell line pairs: A basic and translational science resource. *International Journal of Cancer*, 153(1):183–196, 2023. ISSN 1097-0215. doi: 10.1002/ijc.34506.
- [151] Jiahui Lin and David I. Kutler. Why Otolaryngologists Need to be Aware of Fanconi Anemia. *Otolaryngologic Clinics of North America*, 46(4):567–577, August 2013. ISSN 0030-6665. doi: 10.1016/j.otc.2013.04.002.
- [152] Carlo Dufour and Filomena Pierri. Modern management of Fanconi anemia. *Hematology*, 2022(1):649–657, December 2022. ISSN 1520-4391. doi: 10.1182/hematology.2022000393.
- [153] Sebastian M. Siegner, Laura Ugalde, Alexandra Clemens, Laura Garcia-Garcia, Juan A. Bueren, Paula Rio, Mehmet E. Karasu, and Jacob E. Corn. Adenine base editing efficiently restores the function of Fanconi anemia hematopoietic stem and progenitor cells. *Nature Communications*, 13(1):6900, November 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34479-z.
- [154] Sebastian J. Streichan, Christian R. Hoerner, Tatjana Schneidt, Daniela Holzer, and Lars Hufnagel. Spatial constraints control cell proliferation in tissues. *Proceedings of the National Academy of Sciences*, 111(15):5586–5591, April 2014. doi: 10.1073/pnas.1323016111.

- [155] Harry Eagle and Elliot M. Levine. Growth Regulatory Effects of Cellular Interaction. *Nature*, 213(5081):1102–1106, March 1967. ISSN 1476-4687. doi: 10.1038/2131102a0.
- [156] Maya A. Lewinsohn, Trevor Bedford, Nicola F. Müller, and Alison F. Feder. State-dependent evolutionary models reveal modes of solid tumour growth. *Nature Ecology & Evolution*, 7(4):581–596, April 2023. ISSN 2397-334X. doi: 10.1038/s41559-023-02000-4.
- [157] Robert Noble, Dominik Burri, Cécile Le Sueur, Jeanne Lemant, Yannick Viossat, Jakob Nikolas Kather, and Niko Beerenwinkel. Spatial structure governs the mode of tumour evolution. *Nature Ecology & Evolution*, 6(2):207–217, February 2022. ISSN 2397-334X. doi: 10.1038/s41559-021-01615-9.
- [158] Rachel L. Creighton and Kim A. Woodrow. Microneedle-Mediated Vaccine Delivery to the Oral Mucosa. *Advanced Healthcare Materials*, 8(4):1801180, 2019. ISSN 2192-2659. doi: 10.1002/adhm.201801180.
- [159] Rachel L. Creighton, Kate A. Faber, Carmen I. Tobos, My-Anh Doan, Teri Guo, and Kim A. Woodrow. Oral mucosal vaccination using integrated fiber microneedles. *Journal of Controlled Release*, 367:649–660, March 2024. ISSN 0168-3659. doi: 10.1016/j.jconrel.2024.01.062.
- [160] Koen van der Maaden, Emine Sekerdag, Wim Jiskoot, and Joke Bouwstra. Impact-Insertion Applicator Improves Reliability of Skin Penetration by Solid Microneedle Arrays. *The AAPS Journal*, 16(4):681–684, July 2014. ISSN 1550-7416. doi: 10.1208/s12248-014-9606-7.
- [161] James J. Norman, Jaya M. Arya, Maxine A. McClain, Paula M. Frew, Martin I. Meltzer, and Mark R. Prausnitz. Microneedle patches: Usability and acceptability for self-vaccination against influenza. *Vaccine*, 32(16):1856–1862, April 2014. ISSN 0264-410X. doi: 10.1016/j.vaccine.2014.01.076.
- [162] Yunzhe Ma, Wenqian Tao, Shelly J. Krebs, William F. Sutton, Nancy L. Haigwood, and Harvinder S. Gill. Vaccine Delivery to the Oral Cavity Using Coated Microneedles Induces Systemic and Mucosal Immunity. *Pharmaceutical Research*, 31(9):2393–2403, September 2014. ISSN 1573-904X. doi: 10.1007/s11095-014-1335-1.
- [163] Celia L. McNeilly, Michael L. Crichton, Clare A. Primiero, Ian H. Frazer, Michael S. Roberts, and Mark A. F. Kendall. Microprojection arrays to immunise at mucosal surfaces. *Journal of Controlled Release*, 196:252–260, December 2014. ISSN 0168-3659. doi: 10.1016/j.jconrel.2014.09.028.
- [164] David P. Doupé, Maria P. Alcolea, Amit Roshan, Gen Zhang, Allon M. Klein, Benjamin D. Simons, and Philip H. Jones. A Single Progenitor Population Switches Behavior to Maintain and Repair Esophageal Epithelium. *Science*, 337(6098):1091–1093, August 2012. doi: 10.1126/science.1218835.

- [165] Elizabeth Clayton, David P. Doupé, Allon M. Klein, Douglas J. Winton, Benjamin D. Simons, and Philip H. Jones. A single type of progenitor cell maintains normal epidermis. *Nature*, 446(7132):185–189, March 2007. ISSN 1476-4687. doi: 10.1038/nature05574.
- [166] Gabriel Piedrafita, Vasiliki Kostiou, Agnieszka Wabik, Bartomeu Colom, David Fernandez-Antoran, Albert Herms, Kasumi Murai, Benjamin A. Hall, and Philip H. Jones. A single-progenitor model as the unifying paradigm of epidermal and esophageal epithelial maintenance in mice. *Nature Communications*, 11(1):1429, March 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-15258-0.
- [167] Kevin P. Battaile, Raynard L. Bateman, Derik Mortimer, Jean Mulcahy, R. Keaney Rathbun, Grover Bagby, William H. Fleming, and Markus Grompe. In Vivo Selection of Wild-Type Hematopoietic Stem Cells in a Murine Model of Fanconi Anemia. *Blood*, 94(6):2151–2158, September 1999. ISSN 0006-4971. doi: 10.1182/blood.V94.6.2151.
- [168] Ángel Raya, Ignasi Rodríguez-Pizà, Guillermo Guenechea, Rita Vassena, Susana Navarro, María José Barrero, Antonella Consiglio, Maria Castellà, Paula Río, Eduard Sleep, Federico González, Gustavo Tiscornia, Elena Garreta, Trond Aasen, Anna Veiga, Inder M. Verma, Jordi Surrallés, Juan Bueren, and Juan Carlos Izpisua Belmonte. Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature*, 460(7251):53–59, July 2009. ISSN 1476-4687. doi: 10.1038/nature08129.
- [169] Sakiko Suzuki, Ronny R. Racine, Nathan A. Manalo, Sharon B. Cantor, and Glen D. Raffel. Impairment of fetal hematopoietic stem cell function in the absence of *Fancd2*. *Experimental Hematology*, 48:79–86, April 2017. ISSN 0301-472X. doi: 10.1016/j.exphem.2016.11.006.
- [170] Suying Liu, Es Vivona, and Peter Kurre. Why hematopoietic stem cells fail in Fanconi anemia: Mechanisms and models. *BioEssays*, 47(1):2400191, 2025. ISSN 1521-1878. doi: 10.1002/bies.202400191.
- [171] Viktoria N. Evdokimova, Reagan K. McLoughlin, Sharon L. Wenger, and Stephen G. Grant. Use of the glycophorin A somatic mutation assay for rapid, unambiguous identification of Fanconi anemia homozygotes regardless of GPA genotype. *American Journal of Medical Genetics Part A*, 135A(1):59–65, 2005. ISSN 1552-4833. doi: 10.1002/ajmg.a.30687.
- [172] Xiaoli Li, Andrew F. Wilson, Wei Du, and Qishen Pang. Cell-Cycle-Specific Function of p53 in Fanconi Anemia Hematopoietic Stem and Progenitor Cell Proliferation. *Stem Cell Reports*, 10(2):339–346, February 2018. ISSN 2213-6711. doi: 10.1016/j.stemcr.2017.12.006.

- [173] Brian Freie, Xiixin Li, Samantha L. M. Ciccone, Kathy Nawa, Scott Cooper, Catherine Vogelweid, Laurel Schantz, Laura S. Haneline, Attilio Orazi, Hal E. Broxmeyer, Suk-Hee Lee, and D. Wade Clapp. Fanconi anemia type C and p53 cooperate in apoptosis and tumorigenesis. *Blood*, 102(12):4146–4152, December 2003. ISSN 0006-4971. doi: 10.1182/blood-2003-03-0971.
- [174] Scott Houghtaling, Laura Granville, Yasmine Akkari, Yumi Torimaru, Susan Olson, Milton Finegold, and Markus Grompe. Heterozygosity for p53 (Trp53+/) Accelerates Epithelial Tumor Formation in Fanconi Anemia Complementation Group D2 (Fancd2) Knockout Mice. *Cancer Research*, 65(1):85–91, January 2005. ISSN 0008-5472. doi: 10.1158/0008-5472.85.65.1.
- [175] P. Armitage and R. Doll. The Age Distribution of Cancer and a Multi-stage Theory of Carcinogenesis. *British Journal of Cancer*, 8(1):1–12, March 1954. ISSN 1532-1827. doi: 10.1038/bjc.1954.1.
- [176] Richard Durrett. Branching Process Models of Cancer. In Richard Durrett, editor, *Branching Process Models of Cancer*, pages 1–63. Springer International Publishing, Cham, 2015. ISBN 978-3-319-16065-8. doi: 10.1007/978-3-319-16065-8_1.
- [177] Kasumi Murai, Greta Skrupskelyte, Gabriel Piedrafitra, Michael Hall, Vasiliki Kostiou, Swee Hoe Ong, Tibor Nagy, Alex Cagan, David Goulding, Allon M. Klein, Benjamin A. Hall, and Philip H. Jones. Epidermal Tissue Adapts to Restrain Progenitors Carrying Clonal *p53* Mutations. *Cell Stem Cell*, 23(5):687–699.e8, November 2018. ISSN 1934-5909. doi: 10.1016/j.stem.2018.08.017.
- [178] Maria P. Alcolea, Philip Greulich, Agnieszka Wabik, Julia Frede, Benjamin D. Simons, and Philip H. Jones. Differentiation imbalance in single oesophageal progenitor cells causes clonal immortalization and field change. *Nature Cell Biology*, 16(6):612–619, June 2014. ISSN 1476-4679. doi: 10.1038/ncb2963.
- [179] Kyle B. Jones, Sachiko Furukawa, Pauline Marangoni, Hongfang Ma, Henry Pinkard, Rebecca D’Urso, Rapolas Zilionis, Allon M. Klein, and Ophir D. Klein. Quantitative Clonal Analysis and Single-Cell Transcriptomics Reveal Division Kinetics, Hierarchy, and Fate of Oral Epithelial Progenitor Cells. *Cell Stem Cell*, 24(1):183–192.e8, January 2019. ISSN 1934-5909. doi: 10.1016/j.stem.2018.10.015.
- [180] Ron Sender and Ron Milo. The distribution of cellular turnover in the human body. *Nature Medicine*, 27(1):45–48, January 2021. ISSN 1546-170X. doi: 10.1038/s41591-020-01182-9.
- [181] Qiuwei Pan, Anna M. Nicholson, Hugh Barr, Lea–Anne Harrison, George D. Wilson, Julia Burkert, Rosemary Jeffery, Malcolm R. Alison, Leendert Looijenga, Wey–Ran Lin, Stuart A. C. McDonald, Nicholas A. Wright, Rebecca Harrison, Maikel P. Pepelenbosch, and Janusz A. Jankowski. Identification of Lineage-Uncommitted, Long-Lived, Label-Retaining Cells in Healthy Human Esophagus and Stomach, and in

- Metaplastic Esophagus. *Gastroenterology*, 144(4):761–770, April 2013. ISSN 0016-5085. doi: 10.1053/j.gastro.2012.12.022.
- [182] Julie Rageul and Hyungjin Kim. Fanconi anemia and the underlying causes of genomic instability. *Environmental and Molecular Mutagenesis*, 61(7):693–708, 2020. ISSN 1098-2280. doi: 10.1002/em.22358.
- [183] Andrew D. Hughes and Peter Kurre. The impact of clonal diversity and mosaicism on haematopoietic function in Fanconi anaemia. *British Journal of Haematology*, 196(2):274–287, 2022. ISSN 1365-2141. doi: 10.1111/bjh.17653.
- [184] J. R. Lo Ten Foe, M. L. Kwee, M. A. Rooimans, A. B. Oostra, A. J. P. Veerman, M. van Weel, R. M. Pauli, N. T. Shahidi, I. Dokal, I. Roberts, C. Altay, E. Gluckman, R. A. Gibson, C. G. Mathew, F. Arwert, and H. Joenje. Somatic Mosaicism in Fanconi Anemia: Molecular Basis and Clinical Significance. *European Journal of Human Genetics*, 5(3):137–148, May 1997. ISSN 1476-5438. doi: 10.1007/BF03405891.
- [185] Rajalakshmi S. Asur, Danielle C. Kimble, Francis P. Lach, Moonjung Jung, Frank X. Donovan, Aparna Kamat, Raymond J. Noonan, James W. Thomas, Morgan Park, Peter Chines, Adrianna Vlachos, Arleen D. Auerbach, Agata Smogorzewska, and Settara C. Chandrasekharappa. Somatic mosaicism of an intragenic duplication in both fibroblast and peripheral blood cells observed in a Fanconi anemia patient leads to milder phenotype. *Molecular Genetics & Genomic Medicine*, 6(1):77–91, 2018. ISSN 2324-9269. doi: 10.1002/mgg3.350.
- [186] Ricardo Errazquin, Angustias Page, Anna Suñol, Carmen Segrelles, Estela Carrasco, Jorge Peral, Alicia Garrido-Aranda, Sonia Del Marro, Jessica Ortiz, Corina Lorz, Jordi Minguillon, Jordi Surrallés, Cristina Belendez, Martina Alvarez, Judith Balmaña, Ana Bravo, Angel Ramirez, and Ramon Garcia-Escudero. Development of a mouse model for spontaneous oral squamous cell carcinoma in Fanconi anemia. *Oral Oncology*, 134:106184, November 2022. ISSN 1368-8375. doi: 10.1016/j.oraloncology.2022.106184.
- [187] Sara Jaber, Eléonore Toufekhtchan, Vincent Lejour, Boris Bardot, and Franck Toledo. p53 downregulates the Fanconi anaemia DNA repair pathway. *Nature Communications*, 7(1):11091, April 2016. ISSN 2041-1723. doi: 10.1038/ncomms11091.
- [188] Jasmine D. Peake, Kalisse I. Horne, Chiaki Noguchi, John P. Gilligan, and Eishi Noguchi. The p53 DNA damage response and Fanconi anemia DNA repair pathway protect against acetaldehyde-induced replication stress in esophageal keratinocytes. *Cell Cycle*, 22(18):2088–2096, September 2023. ISSN 1538-4101. doi: 10.1080/15384101.2023.2261740.
- [189] Minna Piipponen, Pilvi Riihilä, Liisa Nissinen, and Veli-Matti Kähäri. The Role of p53 in Progression of Cutaneous Squamous Cell Carcinoma. *Cancers*, 13(18):4507, January 2021. ISSN 2072-6694. doi: 10.3390/cancers13184507.

- [190] Michael P. Kim and Guillermina Lozano. Mutant p53 partners in crime. *Cell Death & Differentiation*, 25(1):161–168, January 2018. ISSN 1476-5403. doi: 10.1038/cdd.2017.185.
- [191] Meng Wang, Laura T. L. Brandt, Xiaonan Wang, Holly Russell, Emily Mitchell, Ashley N. Kamimae-Lanning, Jill M. Brown, Felix A. Dingler, Juan I. Garaycoechea, Tomoya Isobe, Sarah J. Kinston, Muxin Gu, George S. Vassiliou, Nicola K. Wilson, Berthold Göttgens, and Ketan J. Patel. Genotoxic aldehyde stress prematurely ages hematopoietic stem cells in a p53-driven manner. *Molecular Cell*, 83(14):2417–2433.e7, July 2023. ISSN 1097-2765. doi: 10.1016/j.molcel.2023.05.035.
- [192] Connor A. Tsuchida, Kevin M. Wasko, Jennifer R. Hamilton, and Jennifer A. Doudna. Targeted nonviral delivery of genome editors in vivo. *Proceedings of the National Academy of Sciences*, 121(11):e2307796121, March 2024. doi: 10.1073/pnas.2307796121.
- [193] Krishanu Saha, Erik J. Sontheimer, P. J. Brooks, Melinda R. Dwinell, Charles A. Gersbach, David R. Liu, Stephen A. Murray, Shengdar Q. Tsai, Ross C. Wilson, Daniel G. Anderson, Aravind Asokan, Jillian F. Banfield, Krystof S. Bankiewicz, Gang Bao, Jeff W. M. Bulte, Nenad Bursac, Jarryd M. Campbell, Daniel F. Carlson, Elliot L. Chaikof, Zheng-Yi Chen, R. Holland Cheng, Karl J. Clark, David T. Curiel, James E. Dahlman, Benjamin E. Deverman, Mary E. Dickinson, Jennifer A. Doudna, Stephen C. Ekker, Marina E. Emborg, Guoping Feng, Benjamin S. Freedman, David M. Gamm, Guangping Gao, Ionita C. Ghiran, Peter M. Glazer, Shaoqin Gong, Jason D. Heaney, Jon D. Hennebold, John T. Hinson, Anastasia Khvorova, Samira Kiani, William R. Lagor, Kit S. Lam, Kam W. Leong, Jon E. Levine, Jennifer A. Lewis, Cathleen M. Lutz, Danith H. Ly, Samantha Maragh, Paul B. McCray, Todd C. McDevitt, Oleg Mirochnitchenko, Ryuji Morizane, Niren Murthy, Randall S. Prather, John A. Ronald, Subhojit Roy, Sushmita Roy, Venkata Sabisetti, W. Mark Saltzman, Philip J. Santangelo, David J. Segal, Mary Shimoyama, Melissa C. Skala, Alice F. Tarantal, John C. Tilton, George A. Truskey, Moriel Vandsburger, Jonathan K. Watts, Kevin D. Wells, Scot A. Wolfe, Qiaobing Xu, Wen Xue, Guohua Yi, and Jiangbing Zhou. The NIH Somatic Cell Genome Editing program. *Nature*, 592(7853):195–204, April 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03191-1.
- [194] Shankar Subramaniam, Metin Akay, Mark A. Anastasio, Vasudev Bailey, David Boas, Paolo Bonato, Ashutosh Chilkoti, Jennifer R. Cochran, Vicki Colvin, Tejal A. Desai, James S. Duncan, Frederick H. Epstein, Stephanie Fraley, Cecilia Giachelli, K. Jane Grande-Allen, Jordan Green, X. Edward Guo, Isaac B. Hilton, Jay D. Humphrey, Chris R Johnson, George Karniadakis, Michael R. King, Robert F. Kirsch, Sanjay Kumar, Cato T. Laurencin, Song Li, Richard L. Lieber, Nigel Lovell, Prashant Mali, Susan S. Margulies, David F. Meaney, Brenda Ogle, Bernhard Pals-son, Nicholas A. Peppas, Eric J. Perreault, Rick Rabbitt, Lori A. Setton, Lon-

- nie D. Shea, Sanjeev G. Shroff, Kirk Shung, Andreas S. Tolia, Marjolein C.H. van der Meulen, Shyni Varghese, Gordana Vunjak-Novakovic, John A. White, Raymond Winslow, Jianyi Zhang, Kun Zhang, Charles Zukoski, and Michael I. Miller. Grand Challenges at the Interface of Engineering and Medicine. *IEEE Open Journal of Engineering in Medicine and Biology*, 5:1–13, 2024. ISSN 2644-1276. doi: 10.1109/OJEMB.2024.3351717.
- [195] Kiran Musunuru, Sarah A. Grandinette, Xiao Wang, Taylor R. Hudson, Kevin Briseno, Anne Marie Berry, Julia L. Hacker, Alvin Hsu, Rachel A. Silverstein, Logan T. Hille, Aysel N. Ogul, Nancy A. Robinson-Garvin, Juliana C. Small, Sarah McCague, Samantha M. Burke, Christina M. Wright, Sarah Bick, Venkata Indurthi, Shweta Sharma, Michael Jepperson, Christopher A. Vakulskas, Michael Collingwood, Katie Keogh, Ashley Jacobi, Morgan Sturgeon, Christian Brommel, Ellen Schmaljohn, Gavin Kurgan, Thomas Osborne, He Zhang, Kyle Kinney, Garrett Rettig, Christopher J. Barbosa, Sean C. Semple, Ying K. Tam, Cathleen Lutz, Lindsey A. George, Benjamin P. Kleinstiver, David R. Liu, Kim Ng, Sadik H. Kasim, Petros Giannikopoulos, Mohamad-Gabriel Alameh, Fyodor D. Urnov, and Rebecca C. Ahrens-Nicklas. Patient-Specific In Vivo Gene Editing to Treat a Rare Genetic Disease. *New England Journal of Medicine*, 392(22):2235–2243, June 2025. ISSN 0028-4793. doi: 10.1056/NEJMoa2504747.
- [196] Benjamin D. Simons. Deep sequencing as a probe of normal stem cell fate and preneoplasia in human epidermis. *Proceedings of the National Academy of Sciences*, 113(1):128–133, January 2016. doi: 10.1073/pnas.1516123113.
- [197] Ryan O. Schenck, Gabriel Brosula, Jeffrey West, Simon Leedham, Darryl Shibata, and Alexander R.A. Anderson. Gattaca: Base-Pair Resolution Mutation Tracking for Somatic Evolution Studies using Agent-based Models. *Molecular Biology and Evolution*, 39(4):msac058, April 2022. ISSN 1537-1719. doi: 10.1093/molbev/msac058.
- [198] Eran Kotler, Odem Shani, Guy Goldfeld, Maya Lotan-Pompan, Ohad Tarcic, Anat Gershoni, Thomas A. Hopf, Debora S. Marks, Moshe Oren, and Eran Segal. A Systematic p53 Mutation Library Links Differential Functional Impact to Cancer Mutation Pattern and Evolutionary Conservation. *Molecular Cell*, 71(1):178–190.e8, July 2018. ISSN 1097-2765. doi: 10.1016/j.molcel.2018.06.012.
- [199] Marie Laure Delignette-Muller and Christophe Dutang. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64:1–34, March 2015. ISSN 1548-7660. doi: 10.18637/jss.v064.i04.
- [200] Ankit Rohatgi. WebPlotDigitizer: Version 5.2, 2024. URL <https://automeris.io/>. Version 5.2.

- [201] David Crosby, Sangeeta Bhatia, Kevin M. Brindle, Lisa M. Coussens, Caroline Dive, Mark Emberton, Sadik Esener, Rebecca C. Fitzgerald, Sanjiv S. Gambhir, Peter Kuhn, Timothy R. Rebbeck, and Shankar Balasubramanian. Early detection of cancer. *Science*, 375(6586):eaay9040, March 2022. doi: 10.1126/science.aay9040.
- [202] Muxin Gu, Woody Z. Zhang, Rebecca C. Fitzgerald, Alexander M. Frankell, and George S. Vassiliou. Harnessing genomics for early cancer detection, risk stratification and prevention. *Nature Genetics*, 58(4):704–716, April 2026. ISSN 1546-1718. doi: 10.1038/s41588-026-02505-1.
- [203] Jie Liu, Duc Tran, Liying Xue, Brian J. Wiley, Caitlyn Vlasschaert, Caroline J. Watson, Hamish A. J. MacGregor, Xiaoyu Zong, Irenaeus C. C. Chan, Indrani Das, Md Mesbah Uddin, Abhishek Niroula, Gabriel Griffin, Benjamin L. Ebert, Taralynn Mack, Yash Pershad, Brian Sharber, Michael Berger, Ahmet Zehir, Ryan Ptashkin, Ross L. Levine, Elli Papaemmanuil, Vijai Joseph, Teng Gao, Yelena Kemel, Diana Mandelker, Konrad H. Stopsack, Paul D. P. Pharoah, Semanti Mukherjee, Li Ding, Yin Cao, Matthew J. Walter, Jamie R. Blundell, Nilanjan Chatterjee, Kenneth Offit, Lucy A. Godley, Daniel C. Link, Zsofia K. Stadler, Alexander G. Bick, Pradeep Natarajan, and Kelly L. Bolton. Germline genetic variation impacts clonal hematopoiesis landscape and progression to malignancy. *Nature Genetics*, 57(8):1872–1880, August 2025. ISSN 1546-1718. doi: 10.1038/s41588-025-02250-x.
- [204] Dominik Glodzik, Ana Bosch, Johan Hartman, Mattias Aine, Johan Vallon-Christersson, Christel Reuterswård, Anna Karlsson, Shamik Mitra, Emma Niméus, Karolina Holm, Jari Häkkinen, Cecilia Hegardt, Lao H. Saal, Christer Larsson, Martin Malmberg, Lisa Rydén, Anna Ehinger, Niklas Loman, Anders Kvist, Hans Ehrencrona, Serena Nik-Zainal, Åke Borg, and Johan Staaf. Comprehensive molecular comparison of BRCA1 hypermethylated and BRCA1 mutated triple negative breast cancers. *Nature Communications*, 11(1):3747, July 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-17537-2.
- [205] Xia Li, Xiaoping Yao, Yibaina Wang, Fulan Hu, Fan Wang, Liying Jiang, Yupeng Liu, Da Wang, Guizhi Sun, and Yashuang Zhao. MLH1 Promoter Methylation Frequency in Colorectal Cancer Patients and Related Clinicopathological and Molecular Features. *PLOS ONE*, 8(3):e59064, March 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0059064.
- [206] N. J. Belshaw, G. O. Elliott, R. J. Foxall, J. R. Dainty, N. Pal, A. Coupe, D. Garg, D. M. Bradburn, J. C. Mathers, and I. T. Johnson. Profiling CpG island field methylation in both morphologically normal and neoplastic human colonic mucosa. *British Journal of Cancer*, 99(1):136–142, July 2008. ISSN 1532-1827. doi: 10.1038/sj.bjc.6604432.
- [207] Gerard J. Nuovo, Hidewaki Nakagawa, Kaisa Sotamaa, and Albert de la Chapelle. Hypermethylation of the MLH1 Promoter With Concomitant Absence of Transcript

- and Protein Occurs in Small Patches of Crypt Cells in Unaffected Mucosa From Sporadic Colorectal Carcinoma. *Diagnostic Molecular Pathology*, 15(1):17, March 2006. ISSN 1052-9551. URL https://journals.lww.com/molecularpathology/abstract/2006/03000/hypermethylation_of_the_mlh1_promoter_with.3.aspx.
- [208] K. Kawakami, A. Ruszkiewicz, G. Bennett, J. Moore, F. Grieu, G. Watanabe, and B. Iacopetta. DNA hypermethylation in the normal colonic mucosa of patients with colorectal cancer. *British Journal of Cancer*, 94(4):593–598, February 2006. ISSN 1532-1827. doi: 10.1038/sj.bjc.6602940.
- [209] David I. Kutler, Arleen D. Auerbach, Jaya Satagopan, Philip F. Giampietro, Sat Dev Batish, Andrew G. Huvos, Andy Goberdhan, Jatin P. Shah, and Bhuvanesh Singh. High Incidence of Head and Neck Squamous Cell Carcinoma in Patients With Fanconi Anemia. *Archives of Otolaryngology–Head & Neck Surgery*, 129(1):106–112, January 2003. ISSN 0886-4470. doi: 10.1001/archotol.129.1.106.
- [210] Alyssa L. Kennedy, Kasiani C. Myers, James Bowman, Christopher J. Gibson, Nicholas D. Camarda, Elissa Furutani, Gwen M. Muscato, Robert H. Klein, Kaitlyn Ballotti, Shanshan Liu, Chad E. Harris, Ashley Galvin, Maggie Malsch, David Dale, John M. Gansner, Taizo A. Nakano, Alison Bertuch, Adrianna Vlachos, Jeffrey M. Lipton, Paul Castillo, James Connelly, Jane Churpek, John R. Edwards, Nobuko Hijiya, Richard H. Ho, Inga Hofmann, James N. Huang, Siobán Keel, Adam Lamble, Bonnie W. Lau, Maxim Norkin, Elliot Stieglitz, Wendy Stock, Kelly Walkovich, Steffen Boettcher, Christian Brendel, Mark D. Fleming, Stella M. Davies, Edie A. Weller, Christopher Bahl, Scott L. Carter, Akiko Shimamura, and R. Coleman Lindley. Distinct genetic pathways define pre-malignant versus compensatory clonal hematopoiesis in Shwachman-Diamond syndrome. *Nature Communications*, 12(1): 1334, February 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-21588-4.
- [211] Claire C. Homan, Michael W. Drazer, Kai Yu, David M. Lawrence, Jinghua Feng, Luis Arriola-Martinez, Matthew J. Pozsgai, Kelsey E. McNeely, Thuong Ha, Parvathy Venugopal, Peer Arts, Sarah L. King-Smith, Jesse Cheah, Mark Armstrong, Paul Wang, Csaba Bödör, Alan B. Cantor, Mario Cazzola, Erin Degelman, Courtney D. DiNardo, Nicolas Duployez, Remi Favier, Stefan Fröhling, Ana Rio-Machin, Jeffery M. Klco, Alwin Krämer, Mineo Kurokawa, Joanne Lee, Luca Malcovati, Neil V. Morgan, Georges Natsoulis, Carolyn Owen, Keyur P. Patel, Claude Preudhomme, Hana Raslova, Hugh Rienhoff, Tim Ripperger, Rachael Schulte, Kiran Tawana, Elvira Velloso, Benedict Yan, Erika Kim, Raman Sood, Amy P. Hsu, Steven M. Holland, Kerry Phillips, Nicola K. Poplawski, Milena Babic, Andrew H. Wei, Cecily Forsyth, Helen Mar Fan, Ian D. Lewis, Julian Cooney, Rachel Susman, Lucy C. Fox, Piers Blombery, Deepak Singhal, Devendra Hiwase, Belinda Phipson, Andreas W. Schreiber, Christopher N. Hahn, Hamish S. Scott, Paul Liu, Lucy A. Godley, and Anna L. Brown. Somatic mutational landscape of hereditary hematopoietic malignancies caused by germline variants in RUNX1, GATA2, and

- DDX41. *Blood Advances*, 7(20):6092–6107, October 2023. ISSN 2473-9529. doi: 10.1182/bloodadvances.2023010045.
- [212] Jane E. Churpek, Khateriaa Pyrtel, Krishna-Latha Kanchi, Jin Shao, Daniel Koboldt, Christopher A. Miller, Dong Shen, Robert Fulton, Michelle O’Laughlin, Catrina Fronick, Iskra Pusic, Geoffrey L. Uy, Evan M. Braunstein, Mark Levis, Julie Ross, Kevin Elliott, Sharon Heath, Allan Jiang, Peter Westervelt, John F. DiPersio, Daniel C. Link, Matthew J. Walter, John Welch, Richard Wilson, Timothy J. Ley, Lucy A. Godley, and Timothy A. Graubert. Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. *Blood*, 126(22): 2484–2490, November 2015. ISSN 0006-4971. doi: 10.1182/blood-2015-04-641100.
- [213] Anna L. Brown, Peer Arts, Catherine L. Carmichael, Milena Babic, Julia Dobbins, Chan-Eng Chong, Andreas W. Schreiber, Jinghua Feng, Kerry Phillips, Paul P. S. Wang, Thuong Ha, Claire C. Homan, Sarah L. King-Smith, Lesley Rawlings, Cassandra Vakulin, Andrew Dubowsky, Jessica Burdett, Sarah Moore, Grace McKavanagh, Denae Henry, Amanda Wells, Belinda Mercorella, Mario Nicola, Jeffrey Suttle, Ella Wilkins, Xiao-Chun Li, Joelle Michaud, Peter Brautigan, Ping Cannon, Meryl Altree, Louise Jaensch, Miriam Fine, Carolyn Butcher, Richard J. D’Andrea, Ian D. Lewis, Devendra K. Hiwase, Elli Papaemmanuil, Marshall S. Horwitz, Georges Natsoulis, Hugh Y. Rienhoff, Jr, Nigel Patton, Sally Mapp, Rachel Susman, Susan Morgan, Julian Cooney, Mark Currie, Uday Popat, Tilmann Bochtler, Shai Izraeli, Kenneth Bradstock, Lucy A. Godley, Alwin Krämer, Stefan Fröhling, Andrew H. Wei, Cecily Forsyth, Helen Mar Fan, Nicola K. Poplawski, Christopher N. Hahn, and Hamish S. Scott. RUNX1-mutated families show phenotype heterogeneity and a somatic mutation profile unique to germline predisposed AML. *Blood Advances*, 4(6): 1131–1144, March 2020. ISSN 2473-9529. doi: 10.1182/bloodadvances.2019000901.
- [214] Christina M. Dutzmann, Nathalie E. Palmaers, Lucas J. Müntnich, Farina J. Strüwe, Judith Penkert, Birte Sängler, Beatrice Hoffmann, Anja Karow, Christina Reimer, Tanja Gerasimov, Marena R. Niewisch, and Christian P. Kratz. Research on Rare Diseases in Germany – The cancer predisposition syndrome registry. *Journal of Health Monitoring*, 8(4):17–23, December 2023. ISSN 2511-2708. doi: 10.25646/11828.
- [215] Tim H. H. Coorens, Ji Won Oh, Yujin Angelina Choi, Nam Seop Lim, Boxun Zhao, Adam Voshall, Alexej Abyzov, Lucinda Antonacci-Fulton, Samuel Aparicio, Kristin G. Ardlie, Thomas J. Bell, James T. Bennett, Bradley E. Bernstein, Thomas G. Blanchard, Alan P. Boyle, Jason D. Buenrostro, Kathleen H. Burns, Fei Chen, Rui Chen, Sangita Choudhury, Harsha V. Doddapaneni, Evan E. Eichler, Gilad D. Evrony, Melissa A. Faith, Thomas G. Fazzio, Robert S. Fulton, Manuel Garber, Nils Gehlenborg, Soren Germer, Gad Getz, Richard A. Gibbs, Raquel G. Hernandez, Fulai Jin, Jan O. Korb, Dan A. Landau, Heather A. Lawson, Niall J. Lennon, Heng Li, Yan Li, Po-Ru Loh, Gabor Marth, Michael J. McConnell, Ryan E.

- Mills, Stephen B. Montgomery, Pradeep Natarajan, Peter J. Park, Rahul Satija, Fritz J. Sedlazeck, Diane D. Shao, Hui Shen, Andrew B. Stergachis, Hunter R. Underhill, Alexander E. Urban, Melissa W. VonDran, Christopher A. Walsh, Ting Wang, Tao P. Wu, Chenghang Zong, Eunjung Alice Lee, and Flora M. Vaccarino. The Somatic Mosaicism across Human Tissues Network. *Nature*, 643(8070):47–59, July 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09096-7.
- [216] Dian Yang, Matthew G. Jones, Santiago Naranjo, William M. Rideout, Kyung Hoi (Joseph) Min, Raymond Ho, Wei Wu, Joseph M. Replogle, Jennifer L. Page, Jeffrey J. Quinn, Felix Horns, Xiaojie Qiu, Michael Z. Chen, William A. Freed-Pastor, Christopher S. McGinnis, David M. Patterson, Zev J. Gartner, Eric D. Chow, Trevor G. Bivona, Michelle M. Chan, Nir Yosef, Tyler Jacks, and Jonathan S. Weissman. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell*, 185(11):1905–1923.e25, May 2022. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2022.04.015.
- [217] Junhong Choi, Wei Chen, Anna Minkina, Florence M. Chardon, Chase C. Suiter, Samuel G. Regalado, Silvia Domcke, Nobuhiko Hamazaki, Choli Lee, Beth Martin, Riza M. Daza, and Jay Shendure. A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature*, 608(7921):98–107, August 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04922-8.
- [218] Hanna Liao, Junhong Choi, and Jay Shendure. Molecular recording using DNA Typewriter. *Nature Protocols*, 19(10):2833–2862, October 2024. ISSN 1750-2799. doi: 10.1038/s41596-024-01003-0.
- [219] Derek Wong, Ping Luo, Leslie E. Oldfield, Haifan Gong, Ledia Brunga, Ron Rabinowicz, Vallijah Subasri, Clarissa Chan, Tiana Downs, Kirsten M. Farncombe, Beatrice Luu, Maia Norman, Julia A. Sobotka, Precious Uju, Jenna Eagles, Stephanie Pedersen, Johanna Wellum, Arnavaz Danesh, Stephenie D. Prokopec, Eric Y. Stutheit-Zhao, Nadia Znassi, Lawrence E. Heisler, Richard Jovelin, Bernard Lam, Beatriz E. Lujan Toro, Kayla Marsh, Yogi Sundaravadanam, Dax Torti, Carina Man, Anna Goldenberg, Wei Xu, Patrick Veit-Haibach, Andrea S. Doria, David Malkin, Raymond H. Kim, and Trevor J. Pugh. Early Cancer Detection in Li–Fraumeni Syndrome with Cell-Free DNA. *Cancer Discovery*, 14(1):104–119, January 2024. ISSN 2159-8274. doi: 10.1158/2159-8290.CD-23-0456.
- [220] Tej Pandya, Maria Zagorulya, Michelle M. Leung, Marcellus Augustine, Lydia Y. Liu, Aino-Maija Leppä, Ulysse Baruchel, Sin Wi Ng, Tamara Klockner, Miriam Mugabo, Anthony J. Griffen, Oleg Blyuss, Chrysante S. Iliakis, Amalie Grenov, Kerstin Haase, David C. Muller, Ka Hung Chan, Jincheng Wu, Vernon A. Burk, Neil Wright, Alix Le Marois, Ekaterina Pazukhina, Sophia Ward, Hubert Slawinski, Marc Pelletier, Cian Murphy, Matthew D. Park, Thomas Snoeks, Alejandro Suarez-Bonnet, Simon L. Priestnall, Alexandros Chardas, Charlotte Grieco, Ami Archer, Alpkaan

- Celik, Alejandro Jimenez-Sanchez, Rachel Scott, Hana Zahed, Léa Montégut, Rafael Meza, Clinton H. Durney, Stephen Lam, Takahiro Karasaki, Roel C.H. Vermeulen, Huilei Xu, Pablo Serrano-Fernandez, Tatjana Crnogorac-Jurcevic, Usha Menon, Sophia Apostolidou, Alexey Zaikin, Richard Gunu, Harry J. Whitwell, Zhe Huang, Zonglun Li, Xin Hu, Bo Zhu, Liming Li, María-Dolores Chirlaque, Marcela Guevara, P. Martijn Kolijn, Aghiles Guenoun, Neeloffer Mookherjee, Mattias Johansson, Ziqiao Wang, Nilanjan Chatterjee, Chao-Hua Chiu, Zhengming Chen, Dana Pe'er, Erik Sahai, Saskia Freytag, Andreas Wack, Marc J. Gunter, Miriam Merad, Jianjun Zhang, Christopher Carlsten, Pan-Chyr Yang, Hsuan-Yu Chen, Elizabeth A. Platz, Lindsay M. LaFave, Karl Smith-Byrne, Mariam Jamal-Hanjani, Kevin Litchfield, Nuno R. Nene, Nicholas McGranahan, Eva Grönroos, William Hill, Clare E. Weeden, and Charles Swanton. Plasma signals of lung tumor promotion for molecular cancer prevention. *Cell*, 189(13):3903–3921.e26, June 2026. ISSN 00928674. doi: 10.1016/j.cell.2026.05.005.
- [221] Luyi Tian, Fei Chen, and Evan Z. Macosko. The expanding vistas of spatial transcriptomics. *Nature Biotechnology*, 41(6):773–782, June 2023. ISSN 1546-1696. doi: 10.1038/s41587-022-01448-2.
- [222] Cong Ma, Metin Balaban, Jingxian Liu, Siqi Chen, Michael J. Wilson, Christopher H. Sun, Li Ding, and Benjamin J. Raphael. Inferring allele-specific copy number aberrations and tumor phylogeography from spatially resolved transcriptomics. *Nature Methods*, 21(12):2239–2247, December 2024. ISSN 1548-7105. doi: 10.1038/s41592-024-02438-9.
- [223] Andrew Erickson, Mengxiao He, Emelie Berglund, Maja Marklund, Reza Mirzazadeh, Niklas Schultz, Linda Kvastad, Alma Andersson, Ludvig Bergenstråhle, Joseph Bergenstråhle, Ludvig Larsson, Leire Alonso Galicia, Alia Shamikh, Elisa Basmaci, Teresita Díaz De Ståhl, Timothy Rajakumar, Dimitrios Doultinos, Kim Thrane, Andrew L. Ji, Paul A. Khavari, Firaz Tarish, Anna Tanoglidi, Jonas Maaskola, Richard Colling, Tuomas Mirtti, Freddie C. Hamdy, Dan J. Woodcock, Thomas Helleday, Ian G. Mills, Alastair D. Lamb, and Joakim Lundeberg. Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature*, 608(7922):360–367, August 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-05023-2.
- [224] Hillary Elrick, Carolin M. Sauer, Jose Espejo Valle-Inclan, Katherine Trevers, Melanie Tanguy, Sonia Zumalave, Solange De Noon, Francesc Muyas, Rita Cascão, Angela Afonso, Alistair G. Rust, Fernanda Amary, Roberto Tirabosco, Adam Giess, Timothy Freeman, Alona Sosinsky, Katherine Piculell, David T. Miller, Claudia C. Faria, Greg Elgar, Adrienne M. Flanagan, and Isidro Cortes-Ciriano. SAVANA: reliable analysis of somatic structural variants and copy number aberrations using long-read sequencing. *Nature Methods*, 22(7):1436–1446, July 2025. ISSN 1548-7105. doi: 10.1038/s41592-025-02708-0.

- [225] Kieran O'Neill, Erin Pleasance, Jeremy Fan, Vahid Akbari, Glenn Chang, Katherine Dixon, Veronika Csizmok, Signe MacLennan, Vanessa Porter, Andrew Galbraith, Cameron J. Grisdale, Luka Culibrk, John H. Dupuis, Richard Corbett, James Hopkins, Reanne Bowlby, Pawan Pandoh, Duane E. Smailus, Dean Cheng, Tina Wong, Connor Frey, Yaoqing Shen, Eleanor Lewis, Luis F. Paulin, Fritz J. Sedlazeck, Jessica M. T. Nelson, Eric Chuah, Karen L. Mungall, Richard A. Moore, Robin Coope, Andrew J. Mungall, Melissa K. McConechy, Laura M. Williamson, Kasmintan A. Schrader, Stephen Yip, Marco A. Marra, Janessa Laskin, and Steven J. M. Jones. Long-read sequencing of an advanced cancer cohort resolves rearrangements, unravels haplotypes, and reveals methylation landscapes. *Cell Genomics*, 4(11), November 2024. ISSN 2666-979X. doi: 10.1016/j.xgen.2024.100674.
- [226] Lisa A. Lansdon, Byunggil Yoo, Ayse Keskus, Irina Pushel, Chengpeng Bi, Tanveer Ahmad, Asher Bryant, Adam Walter, Margaret Gibson, Mary Rindler, Weijie Li, Sultan M. Habeebu, Linda D. Cooley, John Herriges, Elena Repnikova, Lei Zhang, Keith J. August, Terrie G. Flatt, Alan S. Gamis, Erin M. Guest, J. Allyson Hays, Maxine Hetherington, Karen Lewing, Tomi Pastinen, Mikhail Kolmogorov, and Midhat S. Farooqi. Proof-of-concept study for the detection of somatic structural variant driver alterations using HiFi long-read sequencing in a pediatric leukemia cohort. *npj Genomic Medicine*, April 2026. ISSN 2056-7944. doi: 10.1038/s41525-026-00560-5.
- [227] Alessandro Aiuti, Luca Biasco, Samantha Scaramuzza, Francesca Ferrua, Maria Pia Cicalese, Cristina Baricordi, Francesca Dionisio, Andrea Calabria, Stefania Giannelli, Maria Carmina Castiello, Marita Bosticardo, Costanza Evangelio, Andrea Assanelli, Miriam Casiraghi, Sara Di Nunzio, Luciano Callegaro, Claudia Benati, Paolo Rizzardi, Danilo Pellin, Clelia Di Serio, Manfred Schmidt, Christof Von Kalle, Jason Gardner, Nalini Mehta, Victor Neduva, David J. Dow, Anne Galy, Roberto Miniero, Andrea Finocchi, Ayse Metin, Pinaki P. Banerjee, Jordan S. Orange, Stefania Galimberti, Maria Grazia Valsecchi, Alessandra Biffi, Eugenio Montini, Anna Villa, Fabio Ciceri, Maria Grazia Roncarolo, and Luigi Naldini. Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome. *Science*, 341(6148): 1233151, August 2013. doi: 10.1126/science.1233151.
- [228] Francesca Ferrua, Maria Pia Cicalese, Stefania Galimberti, Stefania Giannelli, Francesca Dionisio, Federica Barzaghi, Maddalena Migliavacca, Maria Ester Bernardo, Valeria Calbi, Andrea Angelo Assanelli, Marcella Facchini, Claudia Fosati, Elena Albertazzi, Samantha Scaramuzza, Immacolata Brigida, Serena Scala, Luca Basso-Ricci, Roberta Pajno, Miriam Casiraghi, Daniele Canarutto, Federica Andrea Salerio, Michael H. Albert, Antonella Bartoli, Hermann M. Wolf, Rossana Fiori, Paolo Silvani, Salvatore Gattillo, Anna Villa, Luca Biasco, Christopher Dott, Emily J. Culme-Seymour, Koenraad van Rossem, Gillian Atkinson, Maria Grazia Valsecchi, Maria Grazia Roncarolo, Fabio Ciceri, Luigi Naldini, and Alessandro Aiuti. Lentiviral haemopoietic stem/progenitor cell gene therapy for

- treatment of Wiskott-Aldrich syndrome: interim results of a non-randomised, open-label, phase 1/2 clinical study. *The Lancet Haematology*, 6(5):e239–e253, May 2019. ISSN 2352-3026. doi: 10.1016/S2352-3026(19)30021-3.
- [229] Lindsey W. Plasschaert, Kelvin D. MacDonald, and Jeffrey S. Moffit. Current landscape of cystic fibrosis gene therapy. *Frontiers in Pharmacology*, 15, October 2024. ISSN 1663-9812. doi: 10.3389/fphar.2024.1476331.
- [230] Ning-ning Yue, Hao-ming Xu, Jing Xu, Min-zheng Zhu, Yuan Zhang, Cheng-Mei Tian, Yu-qiang Nie, Jun Yao, Yu-jie Liang, De-feng Li, and Li-sheng Wang. Therapeutic potential of gene therapy for gastrointestinal diseases: Advancements and future perspectives. *Molecular Therapy - Oncolytics*, 30:193–215, September 2023. ISSN 2372-7705. doi: 10.1016/j.omto.2023.08.007.
- [231] Carlos Lopez-Garcia, Allon M. Klein, Benjamin D. Simons, and Douglas J. Winton. Intestinal Stem Cell Replacement Follows a Pattern of Neutral Drift. *Science*, 330(6005):822–825, November 2010. doi: 10.1126/science.1196236.
- [232] Ann-Marie Baker, Calum Gabbutt, Marc J. Williams, Biancastella Cereser, Noor Jawad, Manuel Rodriguez-Justo, Marnix Jansen, Chris P. Barnes, Benjamin D. Simons, Stuart AC McDonald, Trevor A. Graham, and Nicholas A. Wright. Crypt fusion as a homeostatic mechanism in the human colon. *Gut*, 68(11):1986–1993, November 2019. ISSN 0017-5749, 1468-3288. doi: 10.1136/gutjnl-2018-317540.
- [233] Anna M. Nicholson, Cora Olpe, Alice Hoyle, Ann-Sofie Thorsen, Teja Rus, Mathilde Colombé, Roxanne Brunton-Sim, Richard Kemp, Kate Marks, Phil Quirke, Shalini Malhotra, Rogier ten Hoopen, Ashraf Ibrahim, Cecilia Lindskog, Meagan B. Myers, Barbara Parsons, Simon Tavaré, Mark Wilkinson, Edward Morrissey, and Douglas J. Winton. Fixation and Spread of Somatic Mutations in Adult Human Colonic Epithelium. *Cell Stem Cell*, 22(6):909–918.e8, June 2018. ISSN 1934-5909. doi: 10.1016/j.stem.2018.04.020.
- [234] Cora Olpe, Doran Khamis, Maria Chukanova, Nefeli Skoufou-Papoutsaki, Richard Kemp, Kate Marks, Cerys Tatton, Cecilia Lindskog, Anna Nicholson, Roxanne Brunton-Sim, Shalini Malhotra, Rogier ten Hoopen, Rachael Stanley, Douglas J. Winton, and Edward Morrissey. A Diffusion-like Process Accommodates New Crypts During Clonal Expansion in Human Colonic Epithelium. *Gastroenterology*, 161(2): 548–559.e23, August 2021. ISSN 0016-5085. doi: 10.1053/j.gastro.2021.04.035.
- [235] Daria Pagliara, Andrea Ciolfi, Lucia Pedace, Sadegheh Haghshenas, Marco Ferilli, Michael A. Levy, Evelina Miele, Claudia Nardini, Camilla Cappelletti, Raissa Relator, Angela Pitisci, Rita De Vito, Simone Pizzi, Jennifer Kerkhof, Haley McConkey, Francesca Nazio, Sarina G. Kant, Maddalena Di Donato, Emanuele Agolini, Marta Matraxia, Barbara Pasini, Alessandra Pelle, Tiziana Galluccio, Antonio Novelli, Tahsin Stefan Barakat, Marco Andreani, Francesca Rossi, Cristina Mecucci, Anna

- Savoia, Bekim Sadikovic, Franco Locatelli, and Marco Tartaglia. Identification of a robust DNA methylation signature for Fanconi anemia. *The American Journal of Human Genetics*, 110(11):1938–1949, November 2023. ISSN 0002-9297, 1537-6605. doi: 10.1016/j.ajhg.2023.09.014.
- [236] Wanlin Dai, Xinbo Qiao, Yuanyuan Fang, Renhao Guo, Peng Bai, Shuang Liu, Tingting Li, Yutao Jiang, Shuang Wei, Zhijing Na, Xue Xiao, and Da Li. Epigenetics-targeted drugs: current paradigms and future challenges. *Signal Transduction and Targeted Therapy*, 9(1):332, November 2024. ISSN 2059-3635. doi: 10.1038/s41392-024-02039-0.
- [237] Hany E. Marei. Epigenetic regulators in cancer therapy and progression. *npj Precision Oncology*, 9(1):206, June 2025. ISSN 2397-768X. doi: 10.1038/s41698-025-01003-7.
- [238] Ecaterina E. Dumbrava, Geoffrey I. Shapiro, Aparna R. Parikh, Melissa L. Johnson, Anthony W. Tolcher, John A. Thompson, Anthony B. El-Khoueiry, Andrae L. Vandross, Shivaani Kummar, Dale R. Shepard, Kim LeDuke, Lisa Sheehan, Leila Alland, Arshad Haque, Deepika Jalota, Marc Fellous, and Alison M. Schram. Phase 1 Study of Rezatapopt, a p53 Reactivator, in TP53 Y220C–Mutated Tumors. *New England Journal of Medicine*, 394(9):872–883, February 2026. ISSN 0028-4793. doi: 10.1056/NEJMoa2508820.