THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

Epigenetic Age Prediction and Rejuvenation

Daniel J. Simpson

PhD in Precision Medicine

Doctorate in the College of Medicine & Veterinary Medicine

University of Edinburgh

Primary supervisor: Tamir Chandra

Secondary supervisors: Riccardo Marioni and Chris Ponting

Year of Presentation: 2021

Main text: 27,413 Words

Ey up Jimmy, check it out man!

I did it. Miss ya bud.

Abstract

Ageing is a complex, multi-faceted process that afflicts all humans. It invariably increases susceptibility to a range of diseases such as cancer and neurological disorders. Drugs that mimic calorie restriction show promise in slowing down ageing, but very few treatments appear to be able to actively reverse ageing. Partially reprogrammed stem cells have shown potential as an anti-ageing therapy when used to safely rejuvenate mice without tumour incidence. The question remained as to what exactly occurred at a cellular level. Were a subpopulation of cells dedifferentiating, or partially dedifferentiating and causing a rejuvenative effect by being more stem-like? Or, were the cells epigenetically rejuvenated, where cells became more youthful without loss of somatic cell identity? To test either of these hypotheses, two biomarkers were required to track (i) biological ageing and (ii) dedifferentiation state.

By analysing a previously published dataset of fibroblasts dedifferentiating to induced pluripotent stem cells (iPSCs) over a 49-day time-course, I helped assess the dynamics of cellular ageing. Epigenetic age was used a proxy for biological age, while RNA microarray data was used to assess the state of dedifferentiation (ie. by comparing fibroblast specific gene expression with pluripotency gene markers). Partially reprogrammed cells (between days 7 and 15 of dedifferentiation) declined in predicted age (also known as epigenetic age, eAge), while somatic cell identity was maintained. This shows that loss of somatic gene expression and epigenetic age follow different kinetics, suggesting that they can be uncoupled and a possible "safe period" exists where

rejuvenation can be achieved with a minimized risk of cancer.

While epigenetic clocks appear to confer biological age in many respects, their true underlying function remains a mystery, and the precise aspects of ageing they capture is unclear. For example, differences between epigenetic age and chronological age that are associated with ageing disease states, could be caused by biological and technical biases. Biological biases can arise from mutations affecting the DNA methylation machinery, resulting in global sweeps in methylation. Technical biases may arise from errors in bisulphite (BS) conversion, which could cause a slight overestimation in percentage methylation and therefore alter eAge estimates.

To explore how robust the epigenetic clocks are to sweeps of global methylation, incremental increases and decreases of global methylation were simulated in a large cohort. I showed that epigenetic clocks are not impervious to gradual, global changes in methylation. I also showed how discrete alterations in methylation state can cause a significant difference in eAge compared to a control group, which conceivably could occur in experiments testing rare genetic diseases.

I also present an epigenetic clock based on average methylation over genomic regions, rather than individual CpGs. This clock provides a more robust method of predicting age, which may pave the way for more accurate age predictors using mouse RRBS data.

This thesis has demonstrated that epigenetic clocks are invaluable tools for exploring health-span extending therapies. However, caution must be taken when analysing epigenetic data, as mutations and technical issues may confound analysis. Nonetheless, epigenetic clocks have shown great potential in the molecular ageing field. By understanding the precise nature of eAge, avenues to achieve therapeutic anti-ageing therapies may also be achieved.

Lay Summary

Ageing is a condition that impacts everyone. A growing debate in science is whether to classify ageing as a disease, which would imply that it is curable. Growing evidence suggests that ageing can indeed be slowed by calorie restricting diets. Drugs that mimic these diets also have the potential to slow ageing. There is a growing body of evidence that suggests ageing can even be reversed using the same developmental genes that can artificially create stem cells. Stem cells have a lot of medical potential, however, using them to treat people can lead to cancer. My work has helped show that turning on these developmental genes for 7-15 days in skin cells reduces biological age. Their cell type remains the same, which minimises cancer risk. I have also helped develop a technique to better assess the biological age of mice, which will open avenues for testing anti-aging drugs. Finally, I explored the tools used to predict biological age, and how they could be improved.

Signed Declaration

I, Daniel J. Simpson, declare that:

(a) the thesis has been composed by the student, and

(b) either that the work is the student's own, or, if the student has been a member of a research group, that the student has made a substantial contribution to the work, such contribution being clearly indicated, and

(c) that the work has not been submitted for any other degree or professional qualification except as specified, and

(d) that any included publications are the student's own work, except where indicated throughout the thesis and summarised and clearly identified on the declarations page of the thesis.

Please see "Publications" (pg. x) and "Thesis Structure and Contributions" (pg. viii) for full lists of publications that have been included in this thesis and contributions made to this body of work.

05/04/2021

Acknowledgements

I would first like to express my utmost gratitude to my supervisor Tamir Chandra. Not only has he been the best supervisor anyone could ask for, he has also been a most supportive friend during these four and a half years! He has been available as much as possible, whether it is to give advice and moral support, or to critically assess whatever Marvel/Star Wars media has just been released. I feel privileged to be able to say that I have thoroughly enjoyed my PhD, and for that, I absolutely have Tamir to thank.

I would also like to thank my co-secondary supervisor Riccardo Marioni for teaching me about R, statistics, and of course helping me wrap my head around linear regression and the epigenetic clocks. My thanks also go to Nelly Olova, who has been amazing to work with, particularly for the Aging Cell manuscript. Her input on the global methylation project has been greatly appreciated. I would also like to thank everyone in the Chandra lab (particularly Neil Robertson and Nattaphong Rattanavirotkul) for their advice and support.

I would like to thank my parents too. Obviously I wouldn't have been able to do this without them, but they really did go above and beyond providing the support I needed to complete this PhD. For that, I will be eternally grateful.

I would like to take this opportunity to thank my friends who have helped make sure I didn't accidentally leave a swear word or grammar mistake in my thesis. Samuel Jellicoe, who at this point is more like a brother to me. His support and preservation of my sanity for over 19 years has been invaluable and continues to pull me through. Evangelos Kanonidis, who has been my comrade at arms

during the past four and a half years since he's also been doing his thesis and also close to submitting. To your very good health, my friend. Jamie Jordan, who's enthusiasm for science and continued interest in my work reminds me why I chose science as a career in the first place. And to all my other friends and family both from school, university and at the IGMM who have reassured me over the years, I give my thanks.

Last but not least...well okay, maybe least. But I promised I would include him in my acknowledgements. The damn mouse I spent three months hunting in my flat during lockdown. This damn mouse. It persistently kept finding ways into the flat when I'd seal all possible entries with steel wool, it was impervious to sonic deterrents, and I saw with my own eyes that it was smart enough to avoid and leap over the traps I set. By this point I wondered if this mouse should've been doing a PhD instead of me. It forced me to get creative with the arrangements of the traps I set (I'd lay four or five in a row per entry point), and eventually I chased it into a group of traps. It survived with broken leg, and I released it into the wild. I hope it survived the injury, this worthy adversary. Perhaps it was clever enough to find a way. Anyway, I included him here in the acknowledgments because it reminded me, during a time of utter hopelessness, that perseverance, patience and creativity can solve most problems (as corny as it sounds). If I could take on that damn mouse, then I could finish my bloody thesis!

Thesis Structure and Contributions

This thesis is made up of an introduction (Chapter 1), a methods section (Chapter 3), three results chapters (Chapters 4-6) and an overall discussion chapter (Chapter 7). Each results chapter is structured with its own introduction, results and discussion. All projects were supervised by my primary supervisor Tamir Chandra and my co-secondary supervisor Riccardo Marioni.

Chapter 1 is a literature review describing the theories of ageing, the development of epigenetic clocks, and potential rejuvenation strategies. Chapter 1.4 of the literature review has been submitted to Aging Cell for publication, of which we are awaiting a response. Chapter 3 details the methods used for each results chapter, as well as an overview of overall methods used.

Results Chapter 4 was published in Aging Cell (Olova et al. 2018a). I was co-first author on this project with Nelly Olova. I conducted the bioinformatics analysis and generated age predictions and figures. Figure 4.1 was created exclusively by Nelly Olova based on a figure previously published by Ohnuki et al. 2014. I also wrote the initial draft of the manuscript and helped with revisions. This manuscript won the Aging Cell best paper prize of 2019, and I presented our work at the the Anatomical Society Conference on the 8th January 2021. It should be noted that Chapter 4 was published in 2018. Some of the references have been updated with more recent manuscripts. Related studies that have been published since that link to this project are discussed at the end of the thesis in Chapter 7.

Results Chapter 5 is a project exploring how robust the epigenetic clocks are

to global changes in methylation. This project is being prepared for submission, where I am first author. I conducted all bioinformatic analysis for this project.

Results Chapter 6 is a project I started in 2019 and have worked on it with other Masters students. Xiaoxiao Xie initially showed that reduced representation bisulphite (RRBS) datasets vary greatly in coverage of CpG sites from various mouse clocks. Qian Zhou and I conducted the primary analysis together, where we developed various versions of the regional clock using different combinations of datasets to deliver the best possible age predictor. He is co-first author with me for this project, which we are currently preparing for submission.

Publications

Below is a list of publications that I have been involved in, all of which can be found in the appendix. The majority of them are single-cell transcriptomic projects where I was developing my bioinformatics skills.

1. Karen J Mackenzie et al. (July 2017). "cGAS surveillance of micronuclei links genome instability to innate immunity." *Nature* 548.7668, pp. 461–465. ISSN: 1476-4687. DOI: `10.1038/nature23449`

   I was sixth author in this publication, where I conducted the single-cell RNA seq analysis (Appendix. 1).

2. Juan Carlos Lopez-Baez et al. (Feb. 2018). "Wilms Tumor 1b defines a wound-specific sheath cell subpopulation associated with notochord repair." *eLife* 7. ISSN: 2050-084X. DOI: `10.7554/eLife.30657`

   I was second author in this publication, where I conducted single-cell RNA seq analysis (Appendix. 2).

3. N Olova et al. (Nov. 2018a). "Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity." *Aging cell*. DOI: `10.1111/acel.12877`

   I was co-first author in this project with Nelly Olova. This publication has been used for Chapter 4 (Appendix. 3).

4. Bryan R. Conway et al. (Sept. 2020). "Kidney Single-Cell Atlas Reveals Myeloid Heterogeneity in Progression and Regression of Kidney Disease". *Journal of the American Society of Nephrology* 31.12, ASN.2020060806.

I was fifth author on this publication, where I provided the initial code to Eoin O'Sullivan to help him get started with single-cell RNA seq analysis (Appendix. 4).

List of Abbreviations

5hmC	5-hydroxy-methyl-cytosine
5mC	5-methyl-cytosine
A	adenine
ARCH	age-related clonal haematopoesis
bp	base pair
BS	bisulphite
BS-seq	bisulphite sequencing
C	cytosine
chAge	chronological age
COPD	chronic obstructive pulmonary disease
CpG	5'—C—phosphate—G—3
CR	calorie restriction
DHA	docosahexaenoic acid
DHEA	dehydroepiandrosterone
DNA	deoxyribonucleic acid
DNAm	DNA methylation
DNAmTL	DNAm telomere length
DNMT	DNA methyl-transferase
DSB	double-stranded break
eAge	epigenetic age
EEAA	extrinsic epigenetic age acceleration
EGFP	enhanced green fluorescent protein
EPM	epigenetic pacemaker
ESC	embryonic stem cell
EWAS	epigenome-wide association study
FMD	fasting-mimicking diet
FOXO	forkhead box
G	guanine
GEO	gene expression omnibus
GH	growth hormone
GS	generation scotland
GS1	generation scotland batch 1
GWAS	genome-wide association study
H3K36	histone H3 lysine 36
HBOT	hyperbaric oxygen therapy
HDF	human dermal fibroblast
hESC	human embryonic stem cell
HGPS	Hutchinson-Gilford Progeria Syndrome
HSC	haematopoetic stem cells
hTERT	human telomerase reverse transcriptase
IEAA	intrinsic epigenetic age acceleration
IF	intermittent fasting
IGF	insulin-like growth factor
iPSC	induced pluripotent stem cell
IS	immune system
kb	kilo base pair

LASSO	least absolute shrinkage operator
LTL	leukocyte telomere length
MAD	mean absolute deviation
MAE	median absolute error
MBD4	methyl-CpG binding domain protein 4
mo	month
mTOR	mechanistic target of rapamycin
mTORC	mechanistic target of rapamycin complex
MuSC	muscle stem cells
MuSCs	mouse muscle stem cells
OSK	OCT4, SOX2, KLF4
OSKM	OCT4, SOX2, KLF4, c-MYC
OSKMLN	OCT4, SOX2, KLF4, c-MYC + LIN28 + NANOG
PCR	polymerase chain reaction
PedBE	Pediatric-Buccal-Epigenetic
PoA	Pace-of-Ageing
RBC	regional blood clock
RGC	retinal ganglion cell
rhGH	recombinant human growth hormone
RMSE	root mean squared error
RMTC	regional multi-tissue clock
RNA	ribonucleic acid
RNA-seq	RNA sequencing
ROS	reactive oxygen species
RRBS	reduced representation bisulphite sequencing
RS	replicative senescence
S&B	Skin & Blood
SASP	senescence associated secretory phenotype
SC	stem cell
SCAP	senescent cell anti-apoptotic pathway
scM&T-seq	Single-cell DNA methylation and transcriptome sequencing
SCNT	somatic cell nuclear transfer
scRNA-seq	single-cell RNA sequencing
SE	standard error
sjTREC	signal-joint T cell receptor excision circles
T	thymine
TET	ten-eleven translocation
U	uracil
VEGF	vascular endothelial growth factor
VIM	vimentim
WGBS	whole genome bisulphite sequencing

List of Figures

List of Tables

Contents

Chapter 1

Literature Review

1.1 Theories and Causes of Ageing

Ageing is a multifaceted and complex process that leads to increased susceptibility to an extensive range of diseases, such as cardiovascular, metabolic and neurodegenerative varieties, as well as many forms of cancer (López-Otín et al. 2013; Brunet and Berger 2014; Benayoun, Pollina, and Brunet 2015). The hallmarks of ageing include genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, de-regulated nutrient sensing, mitochondrial dysfunction, accumulation of senescent cells, stem cell exhaustion, and altered intercellular communication (reviewed in López-Otín et al. 2013).

Many theories regarding the cause of ageing fall into either of two categories: programmed ageing or damage/error theories. The programmed theories imply that ageing follows a biological schedule that is perhaps a continuation of a programme regulating growth and development (Jin 2010). An example of this is Wnt signalling, which governs cell proliferation, cell fate choice, apoptosis and cell polarity during development (Schlessinger, Hall, and Tolwinski 2009; Clevers and Nusse 2012; Clevers, Loh, and Nusse 2014). It also controls stem cell maintenance in adults and appears to promote mechanistic target of rapamycin

(mTOR) mediated epidermal stem cell exhaustion and ageing (Castilho et al. 2009). Error/damage theories emphasise that environmental stressors induce cumulative damage to living organisms over time and cause ageing (Jin 2010). Examples of damage that contributes to ageing includes DNA mutations and genomic damage (such as DNA double stranded breaks, DSBs), which could be caused by reactive oxygen species (ROS), radiation, and other genotoxins (chemicals or agents that cause DNA or chromosomal damage; Phillips and Arlt 2009; Afanas'ev 2010; Jin 2010; White and Vijg 2016).

1.1.1   Programmed Theories of Ageing

One of the earliest propositions promoting a programmed theory of ageing was by August Weismann. He reasoned that ageing removes older individuals that are inactive and no longer viable to reproduce, and would otherwise consume resources that would better benefit a younger population (Weismann 1889). This theory has been ill favoured, since starvation, predation and other environmental threats endanger animals before reaching old age, hence a gene network designed to undergo ageing would be unlikely to surface via evolution (Kirkwood 2005). The programmed theory of ageing would also imply ageing is a favourable trait evolutionarily selected to benefit the species, however, no mutations have been observed that appear to "turn off" ageing. Therefore, it is unlikely there are specific gene networks that promote ageing (Kirkwood 2005). One counter argument would be animals that exist with rapid life cycles that experience sexual maturity and old age in a short time, eg. the African krillfish has a lifespan of 12 weeks (Kim, Nam, and Valenzano 2016).

Weismann also theorised that heredity is transmitted via a lineage of immortal cells, termed "germline" (Weismann 1889; West et al. 2019), a theory which was validated by the discovery and study of primordial germ cells (Chiquoine 1954; Reik, Dean, and Walter 2001; Lehmann 2012). The distinction between germline and somatic cells is now termed the "Weismann barrier".

More specifically, it is defined as the loss of replicative immortality that is present in early embryogenesis (somatic restriction). It can be observed not only in complex organisms (eg. mammals) as the difference between germline and soma, but also between organisms. For example, simple unicellular organisms display replicative immortality (Florea 2017), and primitive invertebrates such as planaria (Elliott and Sánchez Alvarado 2013) and hydra (Martínez 1998; Schaible et al. 2015) display no evidence of ageing. Moreover, axolotls live their adulthood in an arrested larval stage, and have a profound regenerative potential in many different tissues such as the heart, forebrain and jaw (West et al. 2019). Hence, they are able to live "below" the Weismann barrier and retain regenerative capabilities. Therefore, ageing could be theorised as a deregulation of development pathways. Indeed, developmental pathways are dysregulated during ageing (such as Wnt and Notch), which could be caused by cumulative epigenetic alterations in stem cells (Castilho et al. 2009; Ermolaeva et al. 2018).

1.1.2 Error/Damage Theories of Ageing

No definitive gene pathway has been found that directly and uniquely governs ageing. Error/damage theories of ageing might therefore explain why these developmental gene pathways are involved in ageing. One error/damage theory that has become prominent and underpins many modern theories of ageing is "mutation accumulation" by Peter Medawar (Medawar 1952). It theorises that late-acting deleterious alleles arising from *de novo* germline mutations accumulate as natural selection wanes with age (Medawar 1952; Kirkwood 2005; Cohen and Holmes 2014). After all, natural selection would be almost powerless to select against mutations that are deleterious in late-life, since the individual has likely already reproduced and passed on their genes. Hence, late-acting, deleterious mutations would accumulate over generations, and any individual that lives long enough would experience these mutations as "ageing" (Kirkwood 2005). This means ageing might have evolved indirectly, and could be

3

falsely perceived as a "programmed" process.

One theory which expands on mutation accumulation is "antagonistic pleiotropy". According to this theory, the genes that drive ageing are initially selected because they give reproductive advantages in early life (Williams 1957). Antagonistic pleiotropy and the mutation accumulation theories are not mutually exclusive, both of which have provided the groundwork for many evolution and ageing theories (Kirkwood and Austad 2000; Partridge and Gems 2002; Kirkwood 2005). Indeed, evidence for antagonistic pleiotropy is growing, as more variants that increase fitness in early age but not in late life are identified (Austad and Hoffman 2018; Byars and Voskarides 2019; Austad and Hoffman 2019).

Another prominent ageing theory is "disposable soma". It assumes that the error/damage theory is true, and that ageing is an evolutionary trade-off between reproduction, growth, and DNA repair (Kirkwood 1977; Kirkwood 2005). In other words, it states that resources for somatic maintenance are limited, and therefore reproduction is prioritised (Kirkwood 2005; Schultz and Sinclair 2016). While the end result is similar to antagonistic pleiotropy (prioritising reproductive fitness/youth at the cost of longevity), the process of disposable soma is more intrinsic and stochastic (Kirkwood 1977; Kirkwood 2005).

While evidence for the ageing theories previously mentioned (particularly antagonistic pleiotropy) are building, they are heavily genetics based. It is possible that the damage accumulation (eg. by ROS and other genotoxins) not only causes somatic mutations, but also contributes to epigenetic and chromatin changes that result in loss of stem cell functionality and age-related diseases (Ermolaeva et al. 2018). The following literature review will explore ageing in the context of epigenetics (particularly DNA methylation) and potential rejuvenation therapies that operate through epigenetic remodelling.

1.2 Epigenetic Drift - DNA Methylation Changes with Age

Ageing can be observed not only in an individual's physiology, but also at a cellular and molecular level (Benayoun, Pollina, and Brunet 2015). While many ageing theories have been developed from a genetics point of view, age-related changes are also reflected in the epigenome; the entirety of non-coding genomic modifications that regulate gene expression. Alterations to the distribution of epigenetic markers such as 5-methyl-cytosine (5mC) throughout the genome affect gene expression in different tissues during ageing (D'Aquila et al. 2013). Hence, an emerging hypothesis is that the epigenetic landscape has biological sensors that reflect not only cellular identity, but also cell health and vitality (Benayoun, Pollina, and Brunet 2015).

Epigenetic modifications such as 5mC are one of the most common epigenetic modifications. Throughout this thesis, DNA methylation (DNAm) will refer to 5mC specifically. DNAm occurs non-randomly in the genome and has a dramatic impact on chromatin, by condensing DNA into a compact structure thus preventing gene transcription (Geiman and Robertson 2002; Sierra, Fernández, and Fraga 2015). DNAm can persist throughout cell divisions via the activity of DNA methyl-transferase (DNMT) 1, which binds to hemimethylated CpG sites and methylates the cytosine on the complementary strand (Gruenbaum, Cedar, and Razin 1982; Bestor and Ingram 1983; Li, Bestor, and Jaenisch 1992; Vilkaitis et al. 2005). *De novo* methylation is conducted by DNMT3A and DNMT3B, which are particularly important for initialising the epigenome for development after fertilisation (Okano et al. 1999; Reik, Dean, and Walter 2001; Brunet and Berger 2014).

Demethylation is conducted in either a passive manner (eg. through inhibition of DNMT1 during cell replication; Wolffe, Jones, and Wade 1999; Mayer et al. 2000), or actively via methyl-CpG binding domain protein 4 (MBD4; Hendrich

et al. 1999) or ten-eleven translocation (TET) dioxygenases (Jin et al. 2014; Ichiyama et al. 2015). Global demethylation can also occur through somatic mutations in *DNMT3A/B* which may contribute to age related conditions such as clonal haematopoesis (Robertson et al. 2019; Bond et al. 2019). Age related changes in the epigenetic landscape are collectively referred to as epigenetic drift, be it changes stochastic in nature or otherwise (Zheng, Widschwendter, and Teschendorff 2016; Ermolaeva et al. 2018).

The first indication that DNAm changes with age was observed in humpback salmon by Berdyshev et al. (1967). The same group later showed that DNAm decreases (hypomethylates) with age in rats (Vanyushin et al. 1973). Global hypomethylation was first confirmed in humans by Wilson et al. in 1987 (Wilson et al. 1987). Initially, it was theorised that the general loss of methylation leads to the relaxation of gene expression with age and to a dysdifferentiated phenotype (ZS.-Nagy, Culter, and Semesi 1988; Unnikrishnan et al. 2018). One of the first studies comparing the DNAm profile of monozygotic twins revealed that they are epigenetically indistinguishable during the early years of life (Fraga et al. 2005). However, adult twins showed divergent patterns (hyper- and hypomethylation of specific regions), and this divergence increased with age (Fraga et al. 2005; Zheng, Widschwendter, and Teschendorff 2016). A longitudinal familial study also showed global changes, particularly hypomethylation, occurs with ageing (Bjornsson et al. 2008). Multiple studies using thin-layer chromatography (Wilson et al. 1987) and high-pressure liquid chromatography (Singhal, Mays-Hoopes, and Eichhorn 1987; Fuke et al. 2004; Liu et al. 2011) showed a global decrease in methylation with age in mouse and humans. However, recent next-generation sequencing experiments have shown no global changes in methylation with age in various mouse tissues (Raddatz et al. 2013; Sun et al. 2014; Cole et al. 2017), and skin (Hadad et al. 2016) and cortex (Lister et al. 2013) in humans.

A recent review assessed many of the studies listed here and concluded that global demethylation does occur in most, if not all cells/tissues as an animal ages

(Unnikrishnan et al. 2018). However, they also state that specific, age-related hypomethylation may occur in specific regions of the genome and in specific cells (e.g. certain cell types or senescent cells), even when global DNAm levels are unchanged. For example, CpG islands bound by polycomb proteins undergo *de novo* methylation with age at key tissue types (Maegawa et al. 2010; Klutstein et al. 2017), while lamin-associated domains demethylate with age (Dmitrijeva et al. 2018; Dor and Cedar 2018). Other studies have shown that age-associated DNAm hyper- and hypomethylation that occur in the genome are non-random and independent of tissue/cell-type (Zheng, Widschwendter, and Teschendorff 2016). For example, age-related hypomethylation occurs at strong enhancers and active promoters (Day et al. 2013). Age-associated hypermethylation, on the other hand, is more likely to occur at polycomb repressive complex 2 marked sites in multiple tissue types (Teschendorff et al. 2010; Horvath et al. 2012; Day et al. 2013; West et al. 2013) or at bivalent (poised) chromatin domains (Rakyan et al. 2010).

Global hypomethylation (Feinberg and Vogelstein 1983; Goelz et al. 1985) and specific hypermethylation of promoter CpG islands (Esteller et al. 2001; Esteller 2002; Herman and Baylin 2003) are also hallmarks of cancer cells. This phenomenon is observed in replicative senescence (RS) cells, and a hypothesis has emerged that the epigenome of tumour cells is a result of cells that might escape RS, and that this epigenomic configuration might promote tumorigenesis (Cruickshanks et al. 2013; Zane, Sharma, and Misteli 2014; Xie, Baylin, and Easwaran 2019). However, Xie et al. have showed that the epigenome of RS cells and tumour cells are quite different, and that the tumour epigenentic landscape is not derived from RS epigenetic changes as previously suspected (Xie et al. 2018; Xie, Baylin, and Easwaran 2019). Xie et al. also showed that cancer risk associated with epigenetic alterations are more likely to come from ageing cells accumulating stochastic methylation changes, rather than RS cells (Xie et al. 2018; Xie, Baylin, and Easwaran 2019).

In addition to cancer, global demethylation has been associated with other age-related diseases, such as neurodegenerative disease and cardiovascular disease (Baccarelli et al. 2010; Chouliaras et al. 2013; Xiao, Wang, and Kong 2019). Centenarians have a delayed onset of age-related disease, which be due to a specific DNAm profile where disease-associated CpGs are being repressed (Hitt et al. 1999; Engberg et al. 2009; Andersen et al. 2012; Heyn et al. 2012; Gentilini et al. 2013; Xiao et al. 2015; Xiao, Wang, and Kong 2019). Methylation levels of specific CpGs (associated with *PDE4C* and *CLCN6*) have also been associated with longevity (Lin et al. 2016). What remains unclear is if there is any causal relationship between these age-related methylation changes and ageing.

Epigenetic drift can also be seen in adult stem cell populations, particularly haematopoetic and mesenchymal stem and progenitor cells, resulting in heterogeneous and dysfunctional cell populations. What role does epigenetic drift play in the ageing of stem cells? And how does an ageing stem cell population impact an ageing organism as a whole?

1.3 Stem Cells in Ageing

Stem cells are undifferentiated cells that are able to give rise to any cell of an organism (reviewed in Zakrzewski et al. 2019). They have the ability to self-renew (ie. stem cells can divide and remain pluripotent) and are responsible for giving rise to an organism during embryonic development and maintaining the organism during adulthood (He, Nakada, and Morrison 2009; Zakrzewski et al. 2019). Young stem cells tend to have fairly uniform epigenomes. However, as they age, stochastic errors in DNA methylation maintenance create epigenetic mosaicism among the ageing stem cell population (reviewed in Shibata 2009; Issa 2014). For example, in the haematopoetic system, many age-related DNAm alterations observed are uniform regardless of cell type

(Yuan et al. 2015). This indicates that these methylation changes originate from a stem or early progenitor cell level which in turn causes lineage bias, reduced stem cell function and increased risk of developing haematologic cancer (Kirschner et al. 2017; Ermolaeva et al. 2018; Kurosawa and Iwama 2020). This phenomenon, referred to as age-related clonal haematopoesis (ARCH), can result from mutations in genes that regulate the epigenome, such as *DNMT3A*, *TET2* and *ASXL1* (Ley et al. 2010; Kim et al. 2013; Yasuda et al. 2014; Herold et al. 2017; Jeong et al. 2018; Abelson and Wang 2018; Robertson et al. 2019; Kurosawa and Iwama 2020).

Stem cells age in many unique ways depending on their cell type. Certain cell types proliferate faster with age (eg. haematopoetic and intestinal stem cells), while others are slower (eg. satellite, neuronal, melanocyte and germline stem cells, reviewed in Schultz and Sinclair 2016). Hair follicle stem cells are an exception in that they do not change in proliferation rate (Giangreco et al. 2008; Rittié et al. 2009), but rather have a longer "rest" (telogen) phase which can result in alopecia (Keyes et al. 2013). Regardless of tissue type, ageing stem cells share a number of physiological changes, including telomere shortening and cellular senescence (Schultz and Sinclair 2016). Other interesting aspects of ageing stem cells are DNA damage and epigenetic dysregulation, and the degree to which they are intertwined.

It is possible that stem cell ageing is driven by damage from genotoxic agents (eg. radiation exposure) that cause DSBs which result in mutations that reduce function. Indeed, there is evidence accumulating that DSBs may drive ageing (White and Vijg 2016; Gorbunova and Seluanov 2016). For example, mutations in DSB repair genes lead to premature ageing (Vogel et al. 1999; Shiloh and Kastan 2001; Espejel et al. 2004; Liu et al. 2005; Mostoslavsky et al. 2006). In addition, a recent preprint showed that DSBs cause a disruption in the epigenome which leads to an accelerated rate of ageing in mice (Hayano et al. 2019). DNA replication is a possible source of genotoxic damage during ageing (Flach et al.

2014; Walter et al. 2015). The more stem cells divide, the higher the likelihood of tumorigenesis developing in the daughter tissue (Tomasetti and Vogelstein 2015). Stem cells can avoid possible damage via DNA replication by entering quiescence (a non-cycling state) for prolonged periods of time (White and Vijg 2016; Tsai 2016). Unfortunately during quiescence, stem cells are only able to repair DSBs via non homologous end joining (Schultz and Sinclair 2016; Biechonski et al. 2018), a repair mechanism which directly ligates DSBs without a homologous template, which increases mutagenesis risk (Chang et al. 2017; Biechonski et al. 2018). This is because homologous recombination (an error free mechanism of repairing DNA), is conducted primarily during S-phase in cycling cells (Karanam et al. 2012). Hence, if a stem cell is damaged while it is quiescent, then any daughter cells/tissues might be compromised in function, which could lead to ageing and promote age-related disorders such as ARCH, and cancer.

1.4 DNA Methylation as a Biomarker and Predictor of Age

DNAm has emerged as one of the most efficient biomarkers to predict biological age (Benayoun, Pollina, and Brunet 2015; Jylhävä, Pedersen, and Hägg 2017). In the past decade a large number of age predictors utilising DNAm have been developed. These DNAm age predictors (more commonly known as epigenetic clocks) are created using CpGs that have tractable changes with age. The majority of these clocks are built using penalised regression models (such as elastic net (Zou and Hastie 2005) or LASSO (Tibshirani 1997)), which select a group of CpGs that have a linear relationship with age in a given training dataset (Horvath and Raj 2018). In other words, key CpGs whose age-related hyper- and hypomethylation correlates with age, are selected and weighted in a linear model. The result is an equation, whereby chronological age can be estimated based on the percentage methylation at these key CpG sites in a given sample

(see Methods, pg. 62).

Epigenetic clocks have become increasingly diverse, with each predictor capturing different aspects of ageing. The expanding repertoire of clocks enable the study of ageing and rejuvenative approaches quantitatively. This section (1.4) will give an overview of the growing toolbox of epigenetic clocks to inform which approach might be best suited to a scientific question.

1.4.1 Do Epigenetic Clocks Predict Biological Age?

Epigenetic clocks have proven themselves to be accurate at predicting chronological age (chAge), which is commonly referred to as DNAm age or epigenetic age (eAge). When epigenetic clocks first emerged, a fundamental question arose; if eAge deviates significantly from chAge, is this difference due to inaccuracies of the clock itself, or caused by biological factors (eg. genetics, disease status, environment)? In other words, are these clocks able to predict biological age? This difference between eAge and chAge is referred to as age acceleration, and can be calculated as the mean absolute deviation (MAD) or median absolute deviation between eAge and chAge (Horvath 2013), or as the residual from the linear regression between eAge and chAge (Horvath and Raj 2018).

There are two other measures of age acceleration that can be used when predicting the eAge of blood (Horvath et al. 2016a). Intrinsic eAge acceleration (IEAA) adjusts for both chAge and blood cell counts, resulting in an estimate unaffected by either variable(Horvath et al. 2016a; Quach et al. 2017). Extrinsic eAge acceleration (EEAA) integrates known age-related changes in blood cell count before adjusting for chAge (Horvath et al. 2016a; Quach et al. 2017). IEAA can be thought of as a cell-intrinsic measure of ageing, while EEAA is a measurement of ageing in terms of the immune system (Horvath et al. 2016a; Quach et al. 2017).

For many of the eAge predictors, age acceleration is associated with a

number of age-related diseases and conditions. For example, patients with Down's syndrome (Horvath et al. 2015a), HIV (Horvath and Levine 2015), obesity (Horvath et al. 2014), Huntington's disease (Horvath et al. 2016b), Werner syndrome (Maierhofer et al. 2017) and Sotos syndrome (Martin-Herranz et al. 2019) tend to exhibit increased age acceleration. eAge acceleration has also been associated with physical and cognitive fitness (Marioni et al. 2015b; Breitling et al. 2016; Quach et al. 2017) and neuropathy (Levine et al. 2015; Lu et al. 2017; please see Horvath and Raj 2018 and Declerck and Vanden Berghe 2018 for comprehensive lists of age acceleration-associated conditions). Variation in epigenetic ageing rates between individuals have been shown to significantly depend on sex and race/ethnicity (Horvath et al. 2016a; McCartney et al. 2019). Smoking has been associated with an increase in eAge in airway cells and lung tissue (4.9 and 4.3 years, respectively; Wu et al. 2019), and smoking during pregnancy might have an effect on eAge in offspring (Simpkin et al. 2016). The number of studies associating eAge acceleration with diseases, phenotypes and environmental interventions that appear to affect ageing, emphasises eAge as a candidate metric for biological age (Wang et al. 2017). However, it remains unclear if age-acceleration/clock CpGs are causative of ageing and age-related diseases (see Section 1.4.6). The limitations of eAge as a proxy for biological age remain to be fully established. Most epigenetic clocks described in the following subsection used Illumina DNAm array-based technology and are summarised in **Table 1.1**.

1.4.2 DNAm Array-Based Epigenetic Clocks

Early Epigenetic Age Predictors

The first epigenetic clocks incorporated relatively few CpG sites and samples in their training datasets, in comparison to later clocks. Bocklandt et al., for example created a clock from 68 samples (34 twin pairs) that predicts age in saliva with an

Table 1.1: Epigenetic clocks based on Illumina human DNA methylation arrays. Age-associated CpGs are selected and weighted in a linear model, resulting in epigenetic age predictors (epigenetic clocks). Error (years) is based on mean absolute deviation (MAD) unless otherwise stated.

Clock	No. CpGs	Error (Years)	Method Used to Find Age-Associated CpGs	No. of Samples in Training	Age Range of Training	Cell Types/Tissue Used for Training	Additional Functional Tissues/Cells	Reference
Bocklandt	88	5.2	CpGs with q < 0.05 & absolute corr > 0.57 with age	68 (34 twin pairs)	21-55	Saliva	-	Bocklandt et al. (2011)
Koch & Wagner	5	11	Pavlidis Template Matching	150	16-72	Fibroblasts, keratinocytes, epithelial, peripheral blood	Saliva, breast organoid	Koch & Wagner (2011)
Passage Number	6	-	Pavlidis Template Matching	-	-	Fibroblasts, mesenchymal stem cells	-	Koch et al. (2012)
Horvath (Pan-Tissue)	353	Median Absolute Deviance 3.6	Elastic net regression	8000	0-100	51 different tissues/cell types including blood, brain, muscle	-	Horvath (2013)
Skin & Blood (S&B)	391	No overall MAD for all tissues /cell types	Elastic net regression	896	0-94	Fibroblasts, keratinocytes, buccal cells, endothelial cells, lymphoblastoid, skin, blood, saliva	Brain, neurons, glia, liver, bone	Horvath et al. (2018)
Zhang (Elastic Net)	514	RMSE 2.04	Elastic net regression	13,661	2-104	Whole blood, saliva	Breast, liver, adipose, muscle, endometrium	Zhang et al. (2019)
Zhang (BLUP)	319,607	RMSE ~ 2.04	Best linear unbiased prediction	13,661	2-104	Whole blood, saliva	-	Zhang et al. (2019)
Hannum	71	RMSE 4.9	FDR to filter significant CpGs then elastic net	482	19-101	Whole blood	-	Hannum et al. (2013)
Weidner (102 CpG)	102	3.3	CpGs selected by pearson corr (r > 0.85 or r < −0.85)	575	0-78	Whole blood	-	Weidner et al. (2014)
Weidner (99 CpG)	99	4.1	CpGs derived from 102 previous CpGs in Weidner et al. (2014)	656	19-101	Whole blood	-	Weidner et al. (2014)
Weidner/Lin (3 CpG)	3	7.6	Three CpGs selected from 102 previous CpGs, recursive feature elimination	656	19-101	Whole blood	-	Weidner et al. (2014), Lin et al. (2016)
Boroni Skin	2,266	RMSE 4.98	Elastic net regression	249	18-95	Dermis, epidermis, whole skin	-	Boroni et al. (2020)
Pediatric-Buccal-Epigenetic (PedBE)	94	0.35	Elastic net regression	1,032	0-19.5	Buccal epithelial cells	-	McEwen et al. (2019)

average accuracy of 5.2 years (Bocklandt et al. 2011). Koch and Wagner used five CpG sites and predicted age in multiple cell types, but with lower accuracy than the Bocklandt clock (MAD = 11 years; Koch and Wagner 2011). The same lab produced a six CpG clock that could track passage number in fibroblast cell cultures, regardless of original donor age (Koch et al. 2012). After these initial studies, epigenetic clocks grew in complexity in terms of number of samples, tissues, and CpGs implemented.

Multi-Tissue Age Predictors

The first multi-tissue age predictor (referred to as the Horvath or Pan-Tissue clock) utilised 353 CpGs and has a mean error of 3.6 years, which at that time was unprecedented for any biomarker/age predictor (Horvath 2013).   The training dataset used to construct the clock comprised of 8000 samples from 82 studies, including 51 healthy tissues and cell types. The size of the training data was a step-change in clock design. Hence, the Horvath clock gained popularity in the scientific community since it can predict age in multiple tissues using a relatively small number of CpGs (compared to the rest of the epigenome), and revealed that tissues may age at different rates. For example, brain tissue appears to age slower relative to other tissues in the body, according to the Horvath clock (Horvath 2013; Horvath et al. 2015c).

The association between age acceleration and health/disease status was first shown with the Horvath clock in obesity (Horvath et al. 2014), and has since become an established tool to assess biological age (Horvath et al. 2015a; Horvath and Levine 2015; Horvath et al. 2014; Horvath et al. 2016b; Maierhofer et al. 2017; Martin-Herranz et al. 2019). The Horvath clock has shown some limitations with particular tissues and age-associated disease conditions. One of the most severe premature ageing syndromes, Hutchinson-Gilford Progeria Syndrome (HGPS), did not exhibit age acceleration according to the Horvath clock (Horvath 2013).   Children with multifocal developmental dysfunctions

(syndrome X), who appear to age slower, do not decelerate in eAge (Walker et al. 2015). However, being rare genetic disorders, both studies were limited in the number of individuals tested.

The Horvath clock does not work reliably on cultured cells, particularly fibroblasts (Horvath 2013; Horvath et al. 2018; Horvath et al. 2019). Replicative senescence in primary fibroblasts is a widely used model system in cellular ageing (Hayflick and Moorhead 1961; Hayflick 1965; Chandra and Kirschner 2016). More recently, Horvath et al. developed an epigenetic clock that predicts the age of human fibroblasts, keratinocytes, buccal cells, endothelial cells, lymphoblastoid cells, skin, blood, and saliva samples, better than the original Horvath clock (Horvath et al. 2018). This clock, known as the skin and blood (S&B) clock, is able to predict both *in vivo* and *in vitro* tissues accurately (Horvath et al. 2018; Horvath et al. 2019). The S&B clock also detected a modest, yet significant age acceleration in HGPS samples (Horvath et al. 2018).

The Zhang clock, while primarily trained to work on blood, is able to predict the ages of breast, liver, adipose, and muscle tissue as accurately as the Horvath clock (Zhang et al. 2019). This clock also outperformed both the Horvath and Hannum clocks in predicting blood age. It is unclear why the the Zhang clock seems to transcend tissue type, while other clocks trained on single tissues do not. It is set apart by the size of its training data with over 13,000 samples.

Tissue-Specific Age Predictors

A number of CpG clocks have been developed for single tissues, aiming at an increased accuracy for a given cell type or specialised applications. Multiple clocks have been developed for blood, the first of which was the Hannum clock (Hannum et al. 2013; Weidner et al. 2014; Zhang et al. 2017b; Horvath and Raj 2018; Zhang et al. 2019). A later study found 102 CpG sites that can predict age in blood, 99 of which were adapted for a separate clock that works on the Illumina 450K array (Weidner et al. 2014). It was demonstrated that three of the

102 CpGs alone (selected by recursive feature elimination) can predict age in arrays and pyrosequenced samples (Weidner et al. 2014; Lin et al. 2016). A minimal approach such as this that uses as few CpGs as possible, is a sought after technique for affordable use in clinical and forensic fields (see Chapter 1.4.3, pg. 17).

Inaccuracies in epigenetic clocks are apparent when predicting the age of younger individuals (under 20 years old Simpkin et al. 2016). This might be due to insufficient numbers of young individuals in training datasets, or due to the linear models used to construct the epigenetic clocks (see Chapter 1.4.5, pg. 22). The Pediatric-Buccal-Epigenetic (PedBE) clock was developed for use in 0-20 year olds and trained on a large number of buccal swab samples (1,032, aged 0-19.5 years old; McEwen et al. 2019). This clock performs well (MAD = 0.35 years) and is an example of how the accuracy of epigenetic clocks can be improved not only by targeting specific tissues, but also specific age groups.

A study by Boroni et al. has produced an accurate skin age predictor, based on 2266 CpGs (one of the largest number of CpGs used to create an eAge clock) selected by elastic net regression (Boroni et al. 2020). It was trained on dermis, epidermis and whole skin biopsies (40, 99 and 110 samples respectively) and had a root mean squared error (RMSE) of 4.98 when tested on an external validation dataset of whole skin biopsies (by comparison, the Horavth and SB clocks had RMSEs of 15.74 and 7.64 respectively; Boroni et al. 2020). This study again highlights the power of tissue specific age predictors, which makes one wonder if the large number of CpGs in this particular clock is beneficial? Multiple studies have shown few highly age-associated CpGs are enough to predict eAge (see "Minimised CpG Clocks", next section), hence a large number CpGs would be presumably redundant. It is possible that the large number of CpGs increase clock robustness and account for the age-related epigenetic heterogeneity in skin between different individuals (Bormann et al. 2016; Boroni et al. 2020), which potentially confounds age prediction.

1.4.3 Minimised CpG Clocks

Most of the epigenetic clocks mentioned above rely on Illumina Infinium arrays. The price of these arrays limits the applicability of eAge technology in drug-discovery. Drug-discovery pipelines require the multiplexing of thousands of samples, but not necessarily the accuracy of the arrays. Here we will discuss clocks relying on fewer CpGs (minimised clocks), which have the potential to be upscaled or run at a lower sample cost. The forensics field has developed multiple minimised clocks using strong age-associated CpGs (such as ELOVL2 and FHL2; Garagnani et al. 2012; Bacalini et al. 2017), and are designed for common tissues found at crime scenes, such as blood, saliva, buccal swabs and semen (**Table 1.2**). Some clocks have also been trained on deceased individuals and have proven effective at predicting their age at death (Dias et al. 2020a; Dias et al. 2020c; Dias et al. 2020b).

Minimised clocks use a variety of technologies such as the Qiagen platform for pyrosequencing (referred to as pyrosequencing from here), which is more cost-effective for profiling the methylation of select CpGs. The Weidner 3 CpG clock (see Chapter 1.4.2, pg. 15) for example, can predict age in blood samples using pyrosequencing (Weidner et al. 2014), but over-predicts age in saliva (a common source of DNA at crime scenes) by 14.6 years on average (Eipel et al. 2016). When adapted for saliva by adding two additional buccal-specific age-associated CpGs, eAge prediction was improved (Eipel et al. 2016).

Pyrosequencing had its own limitations such as multiplexing; allowing a large number of samples and CpGs to be pooled and sequenced in a single run, however, new approaches increasing multiplex capabilities in pyrosequencing are emerging (Fleckhaus and Schneider 2020). Another assay, termed SNaPshot, can multiplex 10 CpG sites (Thermo Fisher 2020), and is used for many minimised epigenetic clocks.

The use of minimised clocks in forensics is just developing and for most clocks cross-validation is missing (Cho et al. 2017). However, the clock by

Table 1.2: Minimized CpG epigenetic clocks. Epigenetic clocks created using a low number of CpGs (typically under 10), usually from preselected CpGs/regions known to have high age-correlation. Error is based on mean absolute deviation (MAD).

Clock	Sequencing	No. CpGs	Error (Years)	Method Used to Find Age-Associated CpGs	No. of Samples in Training	Age Range of Training	Cell Types/Tissue Used for Training	Reference
Weidner 3 CpG	Bisulphite pyrosequencing	3	4.5	3 CpGs selected from 102 previous CpGs by recursive feature elimination	82	0-78	Whole blood	Weidner et al. (2014)
Eipel Buccal	Bisulphite pyrosequencing	5	5.1	3 CpGs from Weidner et al. (2014), plus two additional buccal specific CpGs	55	1-85	Saliva	Eipel et al. (2016)
Zbieć-Piekarska (ZP) Clock	Bisulphite pyrosequencing	5	3.9	8 CpGs from Hannum et al. (2013) then multivariate linear regression	420	2-75	Peripheral blood	Zbieć-Piekarska et al. (2015)
Cho Model 2	Bisulphite pyrosequencing	5	4.2	Similar CpGs to ZP clock (same associated genes but different CpGs), trained in multivariate regression model	100	20-74	Whole blood	Cho et al. (2017)
Jung - Blood	SNaPShot	5	3.5	CpGs used by Cho Model 2 (with different C1orf132 CpG) retrained in multivariate linear model	100	~19-70	Whole blood	Jung et al. (2019)
Jung - Saliva	SNaPShot	5	3.6	""	100	~19-70	Saliva	Jung et al. (2019)
Jung - Buccal Swab	SNaPShot	5	4.3	""	100	~19-70	Buccal epithelial cells	Jung et al. (2019)
Jung - Mixed Tissue	SNaPShot	5	3.8	""	300	~19-70	Whole blood, saliva, buccal epithelial cells	Jung et al. (2019)
Dias - Deceased Clock	Bisulphite PCR	5	8.8	PCR of CpGs from previous studies, trained in multivariate linear model	51 (Deceased)	24-86	Blood	Dias et al. (2020a)
Dias - Multi-Locus Model	Bisulphite PCR	4	5.4	Using CpGs/regions previously used in Dias et al. (2020a), trained in multivariate linear model	53	1-95	Peripheral blood	Dias et al. (2020b)
Dias - Blood (5 CpG)	SNaPShot	5	4.3	Same CpGs used by Jung et al. (2019), retrained in multivariate linear model	59	1-94	Peripheral blood	Dias et al. (2020c)
Dias - Blood (3 CpG)	SNaPShot	3	4.8	3 of 5 CpGs used by Jung et al. (2019) were retrained in multivariate linear model	59	1-94	Peripheral blood	Dias et al. (2020c)

Zbieć-Piekarska et al. has been validated and adapted in other studies. It is based on 5 out of 8 CpGs previously identified by Hannum et al. as showing the strongest age-association (Hannum et al. 2013; Zbieć-Piekarska et al. 2015), and has a standard error of 4.5 years and an MAD of 3.9 years. The genes associated with these CpGs are *ELOVL2, C1orf132, TRIM59, KLF14* and *FHL2*. A clock based on five CpGs (located near the same previous genes) was created by Cho et al., and has been validated in multiple tissues (Cho et al. 2017; Jung et al. 2019; Dias et al. 2020c). These CpGs not only operate adequately with SNaPshot assays as tissue-specific age predictors, but also as a multi-tissue age predictor for common forensic tissues (blood, saliva and buccal swab) (Jung et al. 2019). Three of the CpG sites (near *ELOVL2, FHL2* and *C1orf132*) have also proven sufficient to predict age efficiently (Dias et al. 2020c). These studies have demonstrated the versatility and accuracy predictors based on a few select CpGs can have, and might be good candidates to increase the scale of eAge prediction.

1.4.4 Composite Epigenetic Clocks as Predictors of Morbidity and Mortality

Epigenetic clocks have proven capable of estimating not only chAge, but also time-to-death. Marioni et al. first showed that the higher difference between eAge and chAge, the greater the risk of all-cause mortality (mortality independent of health status, known genetic factors, and lifestyle factors; Marioni et al. 2015a). This finding was further validated in other studies (Chen et al. 2016; Christiansen et al. 2016). Positive age acceleration was also shown to predict cause-specific mortality in cancer and cardiovascular disease (Perna et al. 2016). These mortality associations were found using clocks that were not designed to directly predict mortality. Various composite approaches have been developed, whereby CpGs that correlate with metrics of physiological or cellular ageing (eg. cholesterol or protein abundance) are used to construct a clock to

Table 1.3: Composite and mortality epigenetic clocks. All clocks below are composite clocks, ie. CpGs that correlate with physiological or cellular ageing are used to create a biological age predictor (except the Zhang Mortality Clock, where mortality data was directly regressed on DNAm).

Clock	No. CpGs	Method Used to Obtain CpGs	No. of Samples in Training	Reference
PhenoAge	513	Elastic net	9,926	Levine et al. (2018)
GrimAge	1,113	Elastic net	1,731	Lu et al. (2019a)
Zhang Mortality Clock	10	LASSO Cox regression	548	Zhang et al. (2017a)
DunedinPoAm	46	Elastic net	810	Belsky et al. (2020)
Telomere Clock	140	Elastic net	2,256	Lu et al. (2019b)

predict age (**Table 1.3**). These clocks were build with the potential of capturing more of age-relevant biology than clocks trained on chAge alone.

The first composite biomarker age predictor (created independent of DNAm data) was based on survey data from NHANES III, a large cohort (9,926 samples) with over 23 years of mortality data (Levine 2013). This predictor incorporated ten biomarkers that significantly correlated with age; C-reactive protein, serum creatinine, glycated haemoglobin, systolic blood pressure, serum albumin, total cholesterol, cytomegalovirus optical density, serum alkaline phosphatase, forced expiratory volume, and serum urea nitrogen (Levine 2013). Using a similar process, Levine et al. combined chAge plus nine other biomarkers. The resulting phenotypic clock was regressed on DNA methylation data using elastic net regression, resulting in 513 CpGs forming the DNAm PhenoAge clock (Levine et al. 2018). This clock predicts all-cause mortality, cancer, healthspan, physical functioning, and Alzheimer's disease more accurately than previous age predictors (Levine et al. 2018).

The GrimAge clock developed by Lu et al., uses the methylation of CpGs associated with smoking (pack-years) and levels of 7 plasma proteins previously associated with mortality (Ridker et al. 2003; Ignjatovic et al. 2011), as surrogates for physiological risk factors (Lu et al. 2019a). The age acceleration of GrimAge was not only found to be associated with age related conditions and lifestyle factors, but outperformed previous attempts at predicting time-to-death,

time-to-coronary heart disease and time-to-cancer (Lu et al. 2019a). A significant association has been shown between GrimAge acceleration and lifelong trauma, but not childhood trauma (Katrinli et al. 2020), which is consistent with other studies showing Hannum (Wolf et al. 2016) and Horvath (Yang et al. 2020) clock age accelerations in post-traumatic stress disorder. GrimAge acceleration is also significantly associated with cortical atrophy (Katrinli et al. 2020), shorter pregnancy periods and lower birthweight (Ross et al. 2020).

Zhang et al. created a mortality-specific predictor, where they performed an epigenome wide association study (EWAS) on a cohort with up to 14 years follow up data. 58 CpGs were found that correlate with all-cause mortality, from which a predictor was constructed using only ten of the CpGs (Zhang et al. 2017a). 48 of the CpGs identified had been associated with smoking, alcohol consumption, diabetes and cancer, some of which were also found in previous EWAS studies (Travers et al. 2013; Nilsson et al. 2014; Chambers et al. 2015; Gao et al. 2015; Teschendorff et al. 2015; Al Muftah et al. 2016; Zhang et al. 2017a).

A DNAm telomere length (DNAmTL) estimator was created by Lu et al. 2019b, where leukocyte telomere length (LTL) was regressed against blood methylation data. This resulted in 140 LTL-associated CpGs forming the DNAmTL estimator (Lu et al. 2019b). Not only does the DNAmTL predict LTL accurately, but it also demonstrates stronger predictive power of lifespan, time-to-coronary heart disease, time-to-congestive heart failure and smoking history compared to normal LTL.

Variability in early-life environmental exposures has been proposed as one of the main confounders of mortality clocks (Bell et al. 2019; Hillary et al. 2020). Belsky et al. addressed this directly, by analysing rate of change of 18 blood-chemistry and organ-system-function in a cohort with the same birth year and birth place (Belsky et al. 2015; Hillary et al. 2020). Termed "Pace-of-Ageing" (PoA), this measure formed the basis of the DunedinPoAm clock, a proxy

approach with PoA regressed on DNAm (Belsky et al. 2020). In other words, DunedinPoAm aims to provide the rate of biological ageing at a single-time-point of a person (Belsky et al. 2020).

The Marioni lab compared the performance of six of the epigenetic age/mortality predictors mentioned (Horvath, Hannum, PhenoAge, GrimAge, DNAmTL and DunedinPoAm) in terms of lifespan and disease prediction, on the Generation Scotland cohort (Hillary et al. 2020). GrimAge overall had the best performance; it predicted the prevalence of chronic obstructive pulmonary disease (COPD) and the incidence of multiple diseases, including COPD, type-2 diabetes and cardiovascular disease. GrimAge also outperformed other clocks for predicted death in terms of all-cause mortality, after adjustment for lifestyle risk factors. Another recent study also showed GrimAge outperforms Horvath, Hannum and PhenoAge clocks at predicting all-cause mortality and age-related clinical phenotypes (McCrory et al. 2020). However, DunedinPoAm did reveal faster rates of biological ageing associated with lung cancer and COPD. PhenoAge and DNAmTL also showed associations with disease incidence for type 2 diabetes and ischemic heart disease respectively. Hence, composite clocks can use DNAm to predict non-DNAm traits, which in turn can be used as additional variables to accurately predict biological age, disease status and mortality.

1.4.5 Inaccuracies and Tick-Rate of Epigenetic Clocks

The Horvath clock is the most widely used clock for it's accuracy, versatility and the accumulated knowledge we have of its behaviour from previous studies (see Chapter 1.4.1, pg. 11). As with the Horvath clock, most clocks that followed after were also built on penalised linear regression models. However, are there intrinsic inaccuracies in the Horvath clock, and the approach used to construct epigenetic clocks?

El Khoury et al. analysed previously published DNAm datasets and found

that both the Horvath and Hannum clocks systematically underestimate the age of older individuals (El Khoury et al. 2019). If age acceleration is dependent on chAge itself, biological interpretation of age acceleration at very old age becomes difficult. Centenarian peripheral blood mononuclear cells are predicted 8.6 years younger than their chAge, according to the Horvath clock (Horvath et al. 2015b). Similar findings were also found in analysis of cerebellum tissue from supercentenarians (Horvath et al. 2015c). The interpretation has been that the younger age predicted for centenarians reflects survival bias, where the lower biological age enabled the centenarians to live long. However, with the clock possibly underpredicting older age systematically, this assumption might need to be reexamined.

While negative age acceleration (eAge predicted lower than chAge) was highest in the cerebellum, this underestimation was also observed in other tissues (including blood) from multiple datasets (El Khoury et al. 2019; Marioni et al. 2019; Martin-Herranz et al. 2019). It was also found that when accounting for age as a cofactor, the correlation between age acceleration and amyloid plaque load in brain tissue is attenuated (El Khoury et al. 2019), which is inconsistent with previous findings (Levine et al. 2015). It is possible that 5-hydroxymethyl cytosine (an epigenetic modification more prevalent in brain tissue and indistinguishable from 5mC after bisulphite conversion) could cause age prediction offset in brain tissue (Lunnon et al. 2016; El Khoury et al. 2019). However, 5hmC is not prevalent in blood, and therefore does not explain the negative age acceleration in blood detected by Marioni et al. and El Khoury et al. (Marioni et al. 2019; El Khoury et al. 2019). These alterations in predictive accuracy of the clock in older individuals could be due intrinsic changes in the rate of biological ageing during certain time points. The rate of change, or "tick" rate, was explored earlier in the Horvath clock study (Horvath 2013). By looking at the weighted averages of the 353 CpGs compared with chAge, the tick rate was exponential between 0 and 20 years old, after which it continued linearly. As

such, the Horvath clock applies a logarithmic transformation to ages <20 years, while the linear model is unaltered for ages >20 years (Horvath 2013; Snir, Farrell, and Pellegrini 2019). The study suggested that a higher organismal growth and cell division rate at early age might explain the initial acceleration in ageing (Horvath 2013). A later study found a faster eAge tick rate during puberty in girls (Binder et al. 2018). However, no decrease in the tick rate of older subjects was observed (Horvath 2013), which could be due to a lack of older individuals in the training dataset used to construct the Horvath clock (El Khoury et al. 2019). Differences in tick rate could also be sex-specific. The Horvath, Hannum, and Zbieć-Piekarska clocks show slightly faster ageing in men than women (Bergsma and Rogaeva 2020).

A recent study found that simple multiple linear regression outperforms sophisticated machine learning techniques (Lau and Fung 2020). However, if there is indeed a non-linear progression of age acceleration, then other models might be worth exploring to predict eAge. Support vector regression has been suggested as an alternative to penalised linear regression (Xu et al. 2015; Aliferi et al. 2018). The epigenetic pacemaker (EPM) is an algorithm where predicted age follows a logarithmic trend (Snir, VonHoldt, and Pellegrini 2016; Snir, Farrell, and Pellegrini 2019). Whether EPM or other non-linear models predict eAge in centenarians more accurately has not been determined.

1.4.6   Underlying Mechanism of the Epigenetic Clock

The Horvath clock gained the attention of the scientific community due to the fact that age can be predicted in multiple tissues using a relatively small number of CpGs (compared to the rest of the epigenome; Horvath 2013). The fact that such a clock can be constructed provokes the question, is there a functional significance that correlates these CpGs with age in multiple tissues? If ageing is a phenomena that we are "programmed" to undergo, then are these CpGs an integral part of that machinery? To understand the nature of eAge/epigenetic

clocks, we must understand the aspects of physiological ageing they capture, the CpGs that constitute these clocks, and any causative relationship with ageing.

What Aspects of Physiological Ageing Does eAge Capture?

eAge acceleration (eAge higher than chAge) or deceleration (eAge lower than chAge) reflects most diseases or environmental factors (eg. smoking) that appear to increase or decrease ageing at a physiological level (Horvath et al. 2014; Horvath 2015; Horvath and Levine 2015; Horvath et al. 2016b; Marioni et al. 2015a; Chen et al. 2016; Simpkin et al. 2016; Maierhofer et al. 2017; Horvath et al. 2018; Martin-Herranz et al. 2019; Wu et al. 2019; Higgins-Chen et al. 2020). However, caution should be considered when inferring the biological significance of age acceleration. DNAm age prediction is a relatively new field, and the true nature of how age acceleration is mechanistically associated with biological age is not fully understood. Does eAge reflect or measure known physiological/cellular ageing phenomena (eg. telomere length, senescence)?

Similar eAge among the tissues of an individual person suggests that eAge is not a measure of cellular proliferation, since different tissues have variable proliferation rates and frequency (Horvath 2013; Horvath et al. 2015c; Horvath et al. 2019). Indeed, multiple studies have shown that while eAge correlates with cell passage number, the Horvath age predictor does not correlate with cell division since it can track eAge in non-proliferative tissues (eg. neuronal cells; Horvath 2013; Yang et al. 2016; Horvath et al. 2019). A mitotic clock (EpiTOC) has been developed specifically to track cell divisions, and measured acceleration of this clock correlates with cancer status (Yang et al. 2016). It would be intuitive to assume that eAge reflects other known aspects of ageing such as senescence, since an increase in senescence cells is a key hallmark of ageing (López-Otín et al. 2013; Horvath et al. 2019). However, this is not the case as both replicative and damage induced senescence do not correlate with increased eAge (Lowe, Horvath, and Raj 2016; Horvath et al. 2019). In addition,

human telomerase reverse transcriptase (hTERT) expressing cells continue to epigenetically age despite never being able to enter replicative senescence (Kabacik et al. 2018). Leukocyte telomere length (LTL) erosion is one of the first biological phenomena that showed potential as biomarkers of ageing (Harley, Futcher, and Greider 1990; Hastie et al. 1990; Lindsey et al. 1991; Frenck, Blackburn, and Shannon 1998), and could be a physiological sign of ageing that correlates with eAge. However, like cellular proliferation and senescence, multiple studies have shown that eAge has no association with telomere length (Lowe, Horvath, and Raj 2016; Marioni et al. 2016; Kabacik et al. 2018; Horvath et al. 2019; Cypris et al. 2020).

A plausible alternative is that eAge is governed by cellular differentiation. As stem cells divide during development, they differentiate into different cell types as the embryo matures, which could be reflected by changes in eAge. One study tested the influence of tissue identity on eAge by growing keratinocytes in a media that encourages differentiation. No increase of eAge was observed in the differentiating keratinocytes compared to the non-differentiating, proliferating keratinocytes (Horvath et al. 2019). A separate study transdifferentiated fibroblasts to neurons via miRNAs. The reprogrammed neurons not only had a similar eAge as the donor fibroblasts but also similar telomere length, oxidative stress and DNA damage (Huh et al. 2016). This means that direct reprogramming, ie. changing cellular identity without OSKM-mediated dedifferentiation, had no effect on eAge. These studies give credence to the idea that cellular identity and eAge (biological age), are indeed separate.

It has been hypothesised that eAge related changes are reflected in intracellular alterations and changes in cell composition in a subset of cells termed "clock cells" (Horvath and Raj 2018). eAge might therefore capture the loss of somatic cells in some tissues (Horvath and Raj 2018) or the loss of stem cells, which do decline during ageing (Hernando-Herraez et al. 2019). A caveat is that eAge can be captured in neuronal cells, which are terminally differentiated

cells and lack a stem cell pool (Horvath 2013; Horvath et al. 2015c; Horvath and Raj 2018). It also possible that eAge measures aspects of age-related epigenetic drift/deregulation (Yu et al. 2020), which is discussed in Chapter 5 pg. 101. The precise aspects of physiological ageing that eAge captures (if any) remains a mystery. But further investigations into genes associated with eAge/clock CpGs and associations with other ageing biomarkers (see Chapter 1.5, pg. 34) may disclose clues to the true nature of eAge.

Clock CpGs with Possible Functions in Ageing

As previously mentioned (see Chapter 1.4, pg. 10), DNAm became apparent as a potential biomarker of ageing with the discovery of strong age-associated CpGs, such as those in the CpG islands of ELOVL2, FHL2 and PENK1 (Garagnani et al. 2012; Bacalini et al. 2017). The CpGs neighbouring ELOVL2 strongly hypermethylate with age (Garagnani et al. 2012) and have been used in multiple forensic clocks (see Chapter 1.4.3, pg. 1.4.3). The function of ELOVL2 in terms of ageing has been unclear, until now. ELOVL2 is an enzyme involved in elongation of long-chain polyunsaturated fatty acids, and also in the production of docosahexaenoic acid (DHA). DHA is the main polyunsaturated fatty acid in the retina and brain, and is necessary for healthy retinal function. Chen et al. showed that the *Elovl2* promoter is more highly methylated in the retina of aged mice, and that demethylation of this site recovers age-related decline in visual function via increased expression of *Elovl2* (Chen et al. 2020). This is one of few studies to show a causative link of age-associated CpGs with phenotypic ageing. With ELOVL2 in itself being a strong biomarker for ageing in multiple tissues in both human and mouse (Garagnani et al. 2012; Hannum et al. 2013; Bacalini et al. 2017; Slieker et al. 2018; Chen et al. 2020), it is conceivable that it has more substantial, functional roles in ageing yet to be discovered. It is also possible other epigenetic clock CpGs may contribute to the ageing phenotype.

Transcriptional Associations with eAge

One approach to finding functional components of the epigenetic clock (ie. CpGs that have a causal link with ageing) is to analyse gene expression changes that correlate with age acceleration or with the methylation of age-associated CpGs. In the Horvath clock, a significant number of the 193 CpGs that hypermethylate with age are located in poised (bivalent) promoters (Horvath 2013). Poised chromatin has both repressive and expressive modifications, which means these promoters could be transcriptionally active. Of the 160 CpGs that hypomethylate with age from the 353 Horvath CpGs, a significant number are located in either weak promoters or strong enhancer regions (Horvath 2013). However, linking the activity of age-related CpGs with specific gene expression is a difficult task (Jung and Pfeifer 2015; Zheng, Widschwendter, and Teschendorff 2016; Yin et al. 2017; Horvath and Raj 2018), mainly because the epigenetic state of cells in any given tissue is heterogeneous. This means dual transciptomic and epigenetic sequencing at a single cell level is required to achieve a clear functional link between the two (Horvath and Raj 2018). Single-cell DNA methylation and transcriptome sequencing (scMT-seq) achieves exactly this, by separating the RNA and DNA of single cells, and sequencing the RNA using Smart-seq2 and the DNA by single-cell bisulphite sequencing (Angermueller et al. 2016). With more multi-omic approaches such as this being developed and becoming more commonplace, age related changes between epigenetic clock CpGs and gene expression may one day be revealed.

In fact, a recent study by Hernando-Herraez et al. used scMT-seq to assess ageing in mouse muscle stem cells (MuSCs). They isolated young and old quiescent MuSCs and determined that epigenetic drift (specifically, stochastic methylation heterogeneity at promoters) is associated with age-associated transcriptional heterogeneity (Hernando-Herraez et al. 2019). They also predicted eAge by aggregating single cells by individual (two young and two old mice, with 35 cells per individual). Their age predictor performed accurately on

the young MuSCs, however, their old MuSCs had a similar eAge to the young samples ($\sim$10 weeks, while the chAge of the old MuSCs were $\sim$100 weeks). To compensate for this error, they estimated eAge using different combinations of cells and permutations, by removing 5% of cells and calculating eAge of the subsequent sample. The old MuSCs were still $\sim$90 weeks lower than the chronological age (Hernando-Herraez et al. 2019). It is possible that the negative age acceleration could be explained by the quiescent state of the old MuSCs, however, multiple reports have shown that the Horvath and S&B clocks do not correlate with cell proliferation (Horvath 2013; Yang et al. 2016; Horvath et al. 2019). It is possible that the age predictor developed by Hernando-Herraez et al. captures aspects of cell proliferation that other age predictors do not.

A genome wide association study (GWAS) is a method that could reveal genes that regulate eAge by finding genetic polymorphisms that correlate with eAge. A GWAS of cerebellum tissue found variants near an mTOR complex 2 gene (MLST8) and in an RNA-helicase gene (DHX57) that are associated with age acceleration. Many genes associated with cerebellar age acceleration also had overlap with neurodegenerative conditions such as Alzheimer's disease (Lu et al. 2016). Another GWAS revealed that one of the loci associated with intrinsic eAge acceleration (IEAA, see Chapter 1.4.1, pg. 11) co-locates with hTERT (Lu et al. 2018). Variants of hTERT were found that associated with both IEAA and longer telomeres. Moreover, it was showed *in vitro* that higher hTERT expression (which is normally associated with cellular longevity) increases eAge. Control cells passaged with no hTERT experienced an initial increase in eAge after 33 days in culture, that eventually plateaued. These findings further enforce that eAge is not governed by cell division, replicative senescence or telomere length per se (Lowe, Horvath, and Raj 2016; Marioni et al. 2016; Kabacik et al. 2018; Horvath et al. 2019; Cypris et al. 2020), since short telomeres are indicative of high proliferation and triggers replicative senescence. This paradoxial result could explain the previously reported rapid rate of epigenetic ageing during embryonic

development and early postnatal life, which are periods of fast organismal growth coupled with high hTERT expression and cell division (Hiyama and Hiyama 2007; Simpkin et al. 2016; Simpkin et al. 2017; Lu et al. 2018).

Another approach involves screening for developmental disorders that cause an acceleration or deceleration of eAge. This was conducted by Martin-Herranz et al., who screened 367 genetic disorders, and found that Sotos syndrome significantly accelerated eAge (Martin-Herranz et al. 2019). Sotos syndrome is caused by a loss-of-function mutation in *NSD1*, which encodes a histone H3 lysine 36 (H3K36) methyltransferase (Kurotaki et al. 2002; Choufani et al. 2015). Methylated H3K36 recruits DNMT3A/B and promotes methylation of surrounding regions. The authors hypothesise that H3K36 methylation machinery might break down with age. The NSD1 mutation Martin-Herranz et al. observed might simulate an ageing affect that occurs naturally. An updated study with more samples (particularly of Sotos syndrome) is required to corroborate their findings (Martin-Herranz et al. 2019).

Overall, it remains unclear if the methylation status of clock CpGs has a causative influence on ageing. *MLST8* and *DHX57* variants have been associated with age acceleration. The demethylation of CpGs associated with *ELOVL2* (used in minimised clocks) appears to improve vision in mice (Chen et al. 2020) and is one of the few examples of a causative age-associated clock CpG. Aberrant methylation of these sites with age could have an influence on vision or other age-related outcomes yet to be shown.

1.4.7 Mouse and Non-Human Epigenetic Age Predictors

Since the advent of DNAm age prediction for humans, age predictors have been created for other species; mice (**Table 1.4**), rats (Horvath et al. 2020b; Levine et al. 2020), dogs (Thompson et al. 2017; Wang et al. 2020), wolves (Thompson et al. 2017), humpback whales (Polanowski et al. 2014), chimpanzees (Ito et al. 2018; Guevara et al. 2020), marmosets (Horvath et al. 2020a), naked mole

Table 1.4: Mouse epigenetic clocks. All clocks below were trained on mouse RRBS data (with the exception of Wang et al., which also used WGBS).

Clock	Number of CpGs	Correlation (R^2)	Method Used to Find Age-Associated CpGs	Number of Samples in Training Data	Age Range of Training Samples (Months)	Cell Types/Tissue Used for Training	Reference
Wang	107	0.91	Elastic net	148	0.2-26	Liver	Wang et al. (2017)
Petkovich	90	> 0.90	Elastic net	141	3-35	Partial blood	Petkovich et al. (2017)
Stubbs Multi-Tissue	329	0.7	Elastic net	129	0.2-9.5	Liver, lung, heart, muscle, spleem, cerebellum, cortex	Stubbs et al. (2017)
Meer	435	0.89	Elastic net	~333	0.2–35	Blood, heart, cortex, liver, lung, muscle, spleen, cerebellum, pro B cells, follicular B cells	Meer et al. (2018)
Thompson All CpGs (Ridge)	582	0.79	Ridge Regression	893	0-30	Various tissues including adipose, blood, kidney, liver, lung, muscle, spleen	Thompson et al. (2018)
Thompson All CpGs (Elastic Net)	582	0.82	Elastic net	893	0-30	""	Thompson et al. (2018)
Thompson Conserved CpGs (Ridge)	273	0.64	Ridge Regression	893	0-30	""	Thompson et al. (2018)
Thompson Conserved CpGs (Elastic Net)	273	0.68	Elastic net	893	0-30	""	Thompson et al. (2018)

rats (Lowe et al. 2020), sea bass (Anastasiadi and Piferrer 2020), and zebrafish (Mayne et al. 2020) (see **Table 1.5** for a list of non-human/mouse epigenetic clocks).

In 2017, three mouse epigenetic clocks were developed primarily using reduced representation bisulphite sequencing (RRBS) data. Wang et al. 2017 used 148 CpGs from liver tissue (using both RRBS and whole genome bisulphite data, WGBS), and found a moderate conservation of age related CpGs between human and mouse. Their clock also showed an age reduction for calorie restriction, rapamycin and Prop1[df/df] dwarfism (which results in lifespan extension up to 1.5 fold; Brown-Borg et al. 1996; Wang et al. 2017; Cole et al. 2017). Petkovich et al. built a mouse epigenetic clock using 90 CpGs from blood,

Table 1.5: Studies that have developed epigenetic clocks for non-human or mouse species.

Study	Species	Platform
Horvath et al. (2020b)	Rat, human	HorvathMammalMethylChip40
Levine et al. (2020)	Rat	RRBS
Sugrue et al. (2020)	Sheep, human	HorvathMammalMethylChip40
Thompson et al. (2017)	Dogs, wolves	RRBS
Polanowski et al. (2014)	Humpback whale	Bisulphite pyrosequencing
Ito et al. (2018)	Chimpanzee	Bisulphite pyrosequencing
Guevara et al. 2020	Chimpanzee, human	Human Illumina 850K array
Lowe et al. (2020)	Naked mole rat	Bisulphite PCR
Anastasiadi & Piferrer	Seabass	Multiplex bisulphite sequencing
Mayne et al. (2020)	Zebrafish	RRBS
Horvath et al. (2020a)	Marmoset	HorvathMammalMethylChip40
Wang et al. (2020)	Mouse, dogs	Syntenic Bisulfite Sequencing

and detected that calorie restriction reduces eAge according to their clock (Petkovich et al. 2017). The first mouse multi-tissue age predictor was constructed based on 329 unique CpGs with a median absolute error of 3.33 weeks, mainly trained on young and middle aged mice (0.2-9.5 months) (Stubbs et al. 2017). A recent multi-tissue age predictor in mouse has been developed by Meer et al. that uses 435 CpGs, and predicts age across a wide age-range (1-35 months; Meer et al. 2018). It operates on multiple tissues including blood, liver, brain and heart (Meer et al. 2018). Thompson et al. created four mouse RRBS clocks to compare statistical methods (Thompson et al. 2018). They found the most accurate clock resulting from using elastic net regression. However, only the ridge regression based clocks were able to replicate age reduction in the Prop1 [df/df] dwarf mice, while all the clocks were able to show an effect with calorie restriction (Thompson et al. 2018).

The Wang, Stubbs and Petkovich mouse clocks mentioned here show little overlap in CpGs utilised (Field et al. 2018). This is probably due to the variability of RRBS data, where the regional genome coverage differs between protocols and enzymes used, rather than different statistical methods applied (Field et al.

2018; Thompson et al. 2018). Transferability of these clocks to datasets outside of the original studies has therefore been difficult (see Results Chapter 6, pg. 103). WGBS at a high enough coverage for eAge prediction is expensive, and most mouse clocks are trained on RRBS. Another alternative is has been developed by FOXO BioScience, who have collaborated with Van Andel Institute and Illumina to create a cost-effective Infinium Mouse Methylation Array (FOXO BioScience 2020).

Indeed, other studies have created a similar custom array to accurately predict age in model organisms. Currently available as a preprint, the Horvath lab have published an epigenetic clock that works on both rats and humans (Horvath et al. 2020b). This was created using a custom Illumina methylation array called the HorvathMammalMethylChip40, made up of 36,000 CpGs conserved among 50 mammalian species. The MAE for human and rat data was 0.03, and a correlation of 0.95. Three single tissue clocks were also created for rat liver, brain and blood, as well as a multi-tissue clock combining all three tissues (Horvath et al. 2020b). Another preprint has been released of a sheep epigenetic clock, using the same array, with a median error of 5.1 months (~3.5-4.2 % of expected sheep lifespan). The study reported that castrated sheep had a higher age acceleration than age-matched controls, and a dual human and sheep clock was constructed with an additional 1,848 human samples. (Sugrue et al. 2020).

A rat clock has also been developed using 134 RRBS whole blood samples (Levine et al. 2020). Elastic net selected 68 CpGs, and had a correlation of r = 0.9 in their test dataset. It also appears to work in mice, where it predicted reduced age acceleration after calorie restriction.

Many age associated CpGs are conserved between different species (Horvath 2013; Wang et al. 2017; Horvath et al. 2020b; Wang et al. 2020), which means that pan-species clocks might be possible. For example, an epigenetic clock has been created using 394 CpGs from modules of developmental genes with conserved, age-related methylation changes, between mouse, human and dogs

(Wang et al. 2020). If developed, a pan-species clock might give more information regarding biological ageing, regardless of species.

1.5 Non-DNAm-Based Biomarkers and Age Predictors

While DNAm has proven to be the most accurate predictor of biological age, other interesting biomarkers have been used to construct age predictors. It remains unclear whether DNAm predictors constructed differently captures the same components of ageing. Non-DNAm biomarkers might capture processes DNAm is inert to (**Table 1.6**). Among the earliest biomarkers of ageing was leukocyte telomere length (LTL; Harley, Futcher, and Greider 1990; Hastie et al. 1990; Lindsey et al. 1991; Frenck, Blackburn, and Shannon 1998). Indeed, multiple predictive clocks have been developed using telomere length (Tsuji et al. 2002; Karlsson et al. 2008; Hewakapuge et al. 2008; Ren et al. 2009; Weidner et al. 2014), and short telomere length in adults is associated with increased mortality rates (Goglin et al. 2016). The 4977-bp deletion in human mitochondrial DNA has been explored as a potential biomarker and predictor of ageing in human skeletal muscle, heart and brain (Meissner et al. 1999; Meissner et al. 2008b), however, a direct correlation could not be shown due to the heterogeneity of the mutations (Meissner et al. 1999; Meissner et al. 2008b; Saeed, Berlin, and Cruz 2012). Progressive loss of the Y chromosome in men with age has potential as a biomarker of ageing (Jacobs et al. 1963; Guttenbach, Schakowski, and Schmid 1994; Barros et al. 2020). Another biomarker that has been developed into several age predictors are signal-joint T cell receptor excision circles (sjTREC). sjTREC are DNA segments that result from T-cell receptor gene rearrangement, and their levels correlate negatively with age in blood. Multiple sjTREC clocks have been constructed, with standard errors ranging from 7.35 to 9.42 years (Zubakov et al. 2010; Ou et al. 2012; Cho et al.

Table 1.6: Studies that have developed non-DNAm age predictors.

Clock Type	Studies
Leukocyte telomere length (LTL)	Tsuji et al. (2002), Karlsson et al. (2008), Hewakapug et al. (2008), Ren et al. (2009), Weidner et al. (2014)
Signal-joint T cell receptor excision circles (sjTREC)	Zubakov et al. (2010), Ou et al. (2012), Cho et al. (2014), Ibrahim et al. (2016), Cho et al. (2017)
Transcriptomic	Peters et al. (2015)
Blood Plasma	Tanaka et al. (2018), Lehallier et al. (2019)
Cellular Morphology	Phillips et al. (2017)
Microbiome	Galkin et al. 2020

2014; Ibrahim, Gaballah, and Rashed 2016; Cho et al. 2017). In fact, a combination of sjTREC and CpG methylation provided an accurate age predictor with a mean absolute deviation of 3.31 years (Cho et al. 2017).

Transcription-based age prediction has also shown potential. Large scale changes in gene expression with age were initially shown in a small selection of genes (*LRRN3, CD27, GRAP, CCR6, VAMP5* and *CD248*) which were used to differentiate between younger (age <65 years) and older subjects (age >=75 years; Harries et al. 2011; Holly et al. 2013). Senescence associated genes have also shown interesting associations with ageing. For example, p16INK4a monotonically increases with age, ie. it increases exponentially during young age and eventually plateaus (Tsygankov et al. 2009). Additionally, seven specific senescence associated secretory phenotype (SASP) proteins have been shown to correlate highly with age and adverse post-operation outcomes, and could predict these outcomes better than chronological age (Schafer et al. 2020).

Peters et al. have shown progress developing a transcriptional clock. After performing whole-blood gene expression (RNA microarray-based) meta-analysis with 14,983 individuals, 1,497 genes were identified that are differentially expressed during ageing. Furthermore, they found 12,280 CpG sites whose methylation state correlated with these age-associated expression changes.

They then performed ridge-regression between chAge and gene expression of 11,908 genes to generate a transcriptomic clock. The age acceleration between transcriptomic age and chAge correlated with age-associated features such as blood pressure, cholesterol levels, fasting glucose, and body mass index (Peters et al. 2015). A separate study found 2213 genes that are differentially expressed in blood over time, with the most differentially expressed genes affected by a subset of 77 putative, causative genes (Bryois et al. 2017). These genes could potentially be used to construct a transcriptomic clock.

Proteomic clocks are an alternative age predictor that has recently been demonstrated with blood plasma proteins (Tanaka et al. 2018; Lehallier et al. 2019). The first plasma-based proteomic clock used 240 healthy adults aged 22-93 years old, and found 217 proteins that significantly correlated with age (Tanaka et al. 2018). Elastic net regression was performed on these 217 plasma proteins to create a predictor (76 proteins after penalised regression) with a correlation of 0.94 with chAge (Tanaka et al. 2018). In a second study, 2925 plasma proteins were measured from 4265 adults ranging in age from 18-95 (Lehallier et al. 2019). 373 of these proteins were used to construct a proteomic clock to predict age and healthspan (Lehallier et al. 2019). Interestingly, while some of the plasma proteins identified in this study progressed linearly, some had a non-linear relationship with age. Waves of proteomic-plasma changes with age were also associated with age-associated diseases and phenotypic traits.

Recently, Phillip et al. used a combination of biophysical and bio-molecular assays to find the best cellular parameters to predict chAge in primary dermal fibroblasts (Phillip et al. 2017). Since these assays relied on high-throughput single-cell technologies, cellular heterogeneity was also analysed by measuring cell-to-cell variation of biophysical and bio-molecular features, which showed potential as a predictor of age. Phillip et al. then found that cell roughness peak, combined with nuclear size, provided the one of the best predictors of chAge (Phillip et al. 2017). While this study is greatly limited by it's cohort size (9

samples in training dataset), it highlights the benefits of quantifying non-traditional measurements of age, such as biophysics, which has been sparsely investigated in an ageing context. Another recently developed and unusual age predictor uses microflora taxonomic profiles to predict age (Galkin et al. 2020).

These novel approaches have proven that alternate methods such as proteomic age prediction is possible. Perhaps looking at DNAm in combination with non-DNAm based biomarkers will broaden our understanding and predictive power of biological ageing and mortality. Composite clocks such as PhenoAge and GrimAge (see Chapter 1.4.4, pg. 19) are first steps in that direction. Transcription clocks may reveal regulators of biological ageing, for example, if key ageing genes are found to be linked with eAge either by correlating with age acceleration or directly with methylation changes of key clock CpGs.

1.6 Potential Rejuvenation Strategies

One of the most exciting applications of eAge predictors is to screen potential molecules and compounds for use in anti-ageing therapies. Many recent studies have been dedicated to the slowing, and even the reversal of ageing, both of which were long assumed impossible. These rejuvenation strategies can be categorised into one of four wide-ranging groups; systemic (blood) factors, metabolic alterations, senescent cell ablation and cellular reprogramming (Mahmoudi, Xu, and Brunet 2019).

1.6.1 Blood Factors

Blood factors have shown rejuvenative properties through heterochronic parabolic studies, whereby the circulatory systems of a young and old organism are fused. Exposure of aged mice to young serum was initially shown to increase the regenerative capacity of muscle stem cells and hepatocytes by

restoring Notch signalling and decreasing Wnt signalling pathways (Conboy et al. 2005; Brack et al. 2007). These findings are critical in showing that despite having an aged phenotype, these cells still have a regenerative capacity given the correct environment to promote tissue regeneration (Conboy et al. 2005; Brack et al. 2007).

Since these initial findings, heterochronic parabiosis has been shown to ameliorate age-related hallmarks in cardiomyocytes (Loffredo et al. 2013), skeletal muscle (Sinha et al. 2014; Katsimpardi et al. 2014), brain function (Villeda et al. 2011; Villeda et al. 2014; Katsimpardi et al. 2014; Smith et al. 2015), the central nervous system (Ruckh et al. 2012) and bone regeneration (Baht et al. 2015). It also appears that factors in blood plasma, rather than blood cells themselves, are responsible for these rejuvenative effects (Ruckh et al. 2012; Villeda et al. 2014). Human blood plasma was able to restore mouse hippocampal function, with the candidate anti-ageing factor acting here being TIMP2, a metalloproteinase inhibitor (Castellano et al. 2017).

However, a possible explanation of heterochronic parabiosis is that young blood dilutes pro-ageing factors present in older blood (Mahmoudi, Xu, and Brunet 2019). These pro-ageing factors discovered through heterochronic parabiosis experiments include eotaxin (Villeda et al. 2011) and β_2-microglobulin (Smith et al. 2015). Another pro-ageing example is excessive Wnt signalling, which promotes skeletal muscle to aberrantly convert to fibrous connective tissue, a hallmark of ageing muscle (Brack et al. 2007). The involvement of Wnt in ageing is supported by another report where Wnt signalling induces activation of the mTOR pathway, which in turn promoted epidermal stem cell exhaustion and ageing (Castilho et al. 2009).

These studies have shown that old blood contains factors that promote ageing, while young blood contains factors which impede ageing. However, the direct interactions between these factors and the signalling pathways they are involved in still need to be fully discerned before clinical application (Mahmoudi,

Xu, and Brunet 2019). A single-cell RNA sequencing (scRNA-seq) study of 13 organs from old mice five weeks after undergoing heterochronic parabiosis, has been released as a preprint (Pálovics et al. 2020). It revealed that young blood reverses global loss of gene expression with age, particularly genes encoding components of the electron transport chain. This highlights mitochondrial restoration as a key component of heterochronic parabiosis. Multiple cell types, including haematopoetic stem cells, were also rejuvenated. Once peer-reviewed, this study may form the basis of further exploration into the precise understanding of blood born factors that promote ageing or rejuvenation.

1.6.2 Metabolic Alterations

Metabolic alterations primarily involve the manipulation of dietary intake which in turn offset ageing by improving stress response and autophagy (Mahmoudi, Xu, and Brunet 2019). Dietary restriction is one of the main examples, of which there are many different kinds (Brandhorst et al. 2015).

Dietary Restrictions

Calorie restriction (CR) is one of the most common dietary restrictions, whereby calorie consumption is reduced without triggering malnutrition (Weindruch et al. 1986; Montano 2014). The effect of CR increasing lifespan has been known for over 70 years (McCay, Crowell, and Maynard 1935; Weindruch et al. 1986; Weindruch et al. 1988; Bordone and Guarente 2005). In rats, it has been shown to ameliorate age-related hallmarks in the epigenome (Kim et al. 2016) and oxidative stress in peripheral nerves (Opalach et al. 2010). A two year trial of CR in humans (n=34, control = 19) showed a slowed metabolic rate, along with reduced oxidative stress (Redman et al. 2018).

Intermittent fasting (IF) is another dietary example, which involves one or more fasting day per week (0-800 kcal per fast day), and is commonly structured as 5:2 ratio of normal to diet days per week (Barnard, Snowdon, and Hewitson 2018).

Studies with rodents have shown that IF can promote resilience against cancer, diabetes, heart disease and neuro-degeneration (reviewed in Longo and Mattson 2014). In humans, IF has shown to not only promote weight loss, but also proven as efficient as CR at improving insulin sensitivity and blood pressure (Harvie et al. 2011).

Some dietary restrictions are capable of rejuvenation even when initiated in mid-life or later. The periodic fasting-mimicking diet (FMD) is one example, which is designed to achieve the effects of fasting while providing essential micronutrients and minimising the burden of fasting. In humans, FMD is structured as follows; ~1090 kcal is consumed on the first day (10% protein, 56% fat, 34% carbohydrate), and days 2-5 ~725kcal are consumed (9% protein, 44% fat, 47% carbohydrate) (Brandhorst et al. 2015). Clinical trials in humans showed FMD benefited test subjects by reducing body weight, cardiovascular risk, and IGF-1 expression, while increasing ketone body production (Brandhorst et al. 2015; Wei et al. 2017). FMD was able to increase median life span, reduce cancer incidence in middle aged mice (Brandhorst et al. 2015), and reduce risk of diabetes in young mice (Wei et al. 2018).

Another diet that may have rejuvenative effects when initiated later in life is the ketogenic diet (Newman et al. 2017; Roberts et al. 2017). This diet involves eating a normal caloric intake, whilst also minimising carbohydrate consumption (>50g) (Paoli and Bosco 2015; Mahmoudi, Xu, and Brunet 2019). Like the fasting methods previously mentioned, the ketogenic diet initiates ketosis, which shifts the body towards metabolising fatty acids for energy rather than carbohydrates (Paoli and Bosco 2015; Newman et al. 2017; Roberts et al. 2017).

mTOR and Rapamycin

What all these diets mentioned have in common is that they manipulate nutrient sensing pathways, particularly those involving mTOR and insulin/insulin-like growth factor (IGF), which downstream extends longevity (Weindruch et al.

1986; Weindruch et al. 1988; Blüher, Kahn, and Kahn 2003; Holzenberger et al. 2003; Pearson et al. 2008; Johnson, Rabinovitch, and Kaeberlein 2013). mTOR forms the catalytic core of two protein complexes; mTORC1 (characterised primarily by the presence of RAPTOR), and mTORC2 (characterised primarily by the presence of RICTOR) (Johnson, Rabinovitch, and Kaeberlein 2013). mTORC1 promotes cellular proliferation, growth and survival via anabolic processes such as lipid, ribosomal and protein biosynthesis, mitochondrial metabolism, and also catabolic functions such as inhibition of autophagy (Laplante and Sabatini 2009; Zoncu, Efeyan, and Sabatini 2011; Watanabe, Wei, and Huang 2011). Rapamycin is a drug which acts by binding with FKBP12 (a subunit of mTORC1), which in turn inhibits mTORC1 by interrupting interactions between mTOR and RAPTOR (another key subunit of mTORC1; Johnson, Rabinovitch, and Kaeberlein 2013; Kazyken et al. 2019). Inhibiting mTOR via rapamycin as a drug intervention extends lifespan in mice by acting as a dietary restriction mimic (Harrison et al. 2009; Chen et al. 2009; Miller et al. 2011; Anisimov et al. 2011; Neff et al. 2013). As well as increased longevity, rapamycin promotes autophagy, a process the decline of which is associated with a number age-related diseases such as diabetes, cancer, cardiovascular and neurodegenerative diseases (Cuervo 2008; Mizushima et al. 2008; Johnson, Rabinovitch, and Kaeberlein 2013; Mahmoudi, Xu, and Brunet 2019; Singh et al. 2019). Rapamycin and the mTOR signalling pathway have since become major candidates in anti-ageing therapy (Johnson, Rabinovitch, and Kaeberlein 2013).

mTORC2 also promotes cellular proliferation and survival, however its overall function is less clearly defined than mTORC1 (Kazyken et al. 2019). While mTORC2 is not directly influenced by rapamycin, it is theorised that chronic exposure can sequester mTOR from mTORC2, which leads to activation of forkhead box (FOXO) genes (Johnson, Rabinovitch, and Kaeberlein 2013; Singh et al. 2019). FOXO genes (which are also activated by low insulin/IGF signalling) in turn control a cascade of genes that promote autophagy and stress resistance

(Johnson, Rabinovitch, and Kaeberlein 2013; Singh et al. 2019).

Sirtuins and Resveratrol

Sirtuins are a set of genes correlated with nutrient sensing that, when upregulated, can delay the onset of age-related diseases (Pfluger et al. 2008; Banks et al. 2008; Kanfi et al. 2010; Hubbard and Sinclair 2014; Kugel et al. 2016) and increase longevity (Kanfi et al. 2010; Satoh et al. 2013; Mercken et al. 2014; Mitchell et al. 2014). Sirtuins are highly conserved NAD^+-dependent deacylases involved in a variety of processes, including transcription regulation (eg. FOXO genes), cell survival, metabolism regulation, DNA repair, circadian rhythm regulation and inflammation (reviewed in Kane and Sinclair 2018). However, there is controversy surrounding the role of sirtuins in longevity. Firstly, their role in CR is still debated. Since CR enhances NAD^+ production, and also converts nicotinamide (a sirtuin inhibitor) to niacin. It is therefore reasonable to hypothesise that CR increases sirtuin activation and plays a roll in CR mediated longevity (Dang 2014). Indeed, deletion of *SIR2* (a homolog of *SIR1*) blocks CR-induced life extension in yeast (Kaeberlein, McVey, and Guarente 1999), with similar results reported in worms (Rogina and Helfand 2004) and fruit flies (Wang et al. 2006). While the direct role of sirtuins in CR is still being elucidated, other studies have shown that sirtuins mediate the rejuvenative effects of CR in skeletal and cardiac muscle (reviewed in Zullo et al. 2018).

Resveratrol is a drug that has been routinely studied for life extending properties. This compound can be extracted from red wine, and has shown life extension in yeast, worms, flies and fish (Howitz et al. 2003; Wood et al. 2004). While studies have shown that it does not extend life in mice fed on normal diets (Baur et al. 2006; Pearson et al. 2008; Miller et al. 2011), mice fed on a calorie high diet appear to have life extension, improved mitochondrial function and protection from diet-induced obesity and insulin resistance (Baur et al. 2006; Lagouge et al. 2006). Resveratrol appears to induce similar transcriptional

changes, ageing retardation, and improved insulin mediated glucose uptake in muscle as CR (Barger et al. 2008). It has been shown that sirtuins are activated by resveratrol (Dai et al. 2010; Lakshminarasimhan et al. 2013) which is postulated as the reason why resveratrol promotes life extension, however, this activation has not been shown *in vivo* and the life extension effects also differ between worm and fly strains (Baur et al. 2006). Hence, the exact role of resveratrol in sirtuin activation is still debated (Callaway 2010; Dang 2014), and has been the subject of controversy. In 2008, the pharmaceutical company GlaxoSmithKline bought Sirtris for 720 million dollars, a company that was researching the anti-ageing properties of reseveratrol and its role in activating sirtuins (Callaway 2010). The clinical trial for the proprietary version of resveratrol was later cancelled by Sirtris (Popat et al. 2013), and the company was dismantled and absorbed by GlaxoSmithKline (Ledford 2013; Carroll 2015).

1.6.3 Senescent Cell Ablation

Accumulation of Senescent Cells and Ageing

Cellular senescence is a cell-intrinsic mechanism that irreversibly and stably halts cell division in response to intrinsic and extrinsic stressors such as DNA-damage, telomere shortening, oxidative stress, oncogenic and mitogenic stimuli (Hernandez-Segura, Nehme, and Demaria 2018; Herranz and Gil 2018). Initially, this phenomenon was identified in cultured fibroblasts that when left to divide undergo a finite number of cell divisions until they stop (known as the "Hayflick limit"; Hayflick and Moorhead 1961). This limited replicative capacity of cultured cells (also known as replicative senescence) suggested a mechanism of ageing at a cellular level, which would therefore contribute to organismal ageing (Rattanavirotkul, Kirschner, and Chandra 2020). Indeed, senescent cells accumulate in aged tissue as less immune cells (eg. macrophages, natural killer cells and T-cells) are present to clear them (Schmeer et al. 2019), which is

associated with decline in the adaptive immune system with age (McElhaney and Effros 2009). Senescent cells secrete inflammatory factors (a state known as senescence-associated secretory phenotype, SASP) which over time can damage nearby cells, hence they attract immune system (IS) cells for clearance (Campisi 2013; Schmeer et al. 2019). This precludes the possibility that senescent cells are produced at a higher rate in ageing tissue, which would increase cellular damage, oncogenic mutations, and/or senescence-inducing events (Sedelnikova et al. 2004; Wang et al. 2009; Hewitt et al. 2012; Campisi 2013). SASP inhibitors such as rapamycin, metformin, or JAK1/2 inhibitors are also of interest for reducing age/senescence related inflammation, however, their precise properties and side affects in regards to SASP is still not fully understood (Kirkland and Tchkonia 2017).

Whether or not the accumulation of senescent cells (and therefore accumulation of SASP) is causative or a bi-product of ageing is only recently becoming apparent (Campisi 2013). Clearance of p16 (cyclin-dependent kinase inhibitor and known marker of senescence) positive cells in progeroid mice, both throughout life and in later life, delayed age related pathologies and extended lifespan (Baker et al. 2011). Later, the same experiment was conducted with naturally aged mice which also extended lifespan, delayed tumourigenesis and attenuated age-related degeneration of organs (Baker et al. 2016). These studies indicate that SASP might directly contribute to ageing. However, it cannot be ruled out that ageing and senescence might mutually influence each other through a positive feedback loop, exacerbating tissue damage and an aged phenotype (Schmeer et al. 2019).

Senolytics

The clearance of senescent cells as a potential therapeutic treatment has stimulated research into compounds that can selectively kill them. Senolytics are such drugs, which are able to specifically kill senescent cells by inducing

apoptosis (Kirkland and Tchkonia 2017). The first senolytics were discovered by Zhu et al. 2015. They hypothesised that since senescent cells are resistant to apoptosis (Wang 1995), specific senescent cell anti-apoptotic pathways (SCAPs) exist and are up-regulated to resist apoptosis. Small interfering RNA were used to screen for and inhibit these SCAPs (Zhu et al. 2015), and were targeted for drug interventions promoting apoptosis in senescent cells. The most promising drugs revealed in this study were dasatinib and quercetin, which act as kinase inhibitors promoting apoptosis and normally used in cancer treatments (Zhu et al. 2015).   Since then, a large number of senolytics have been discovered, however, their full function and applicability in anti-ageing therapy needs to be assessed (Mahmoudi, Xu, and Brunet 2019).

A key aspect of ageing is a decline in stem cell population (as mentioned in Chapter 1.3.), most notably due to senescence. Ablation of senescent cells in aged stem cell populations is crucial for effective rejuvenation to ensure that healthy tissue can be renewed. Recent senolytics research showed that the removal of senescent cells is enough to spur stem cell propagation. ABT263 for example, is a senoyltic which not only cleared senescent haematopoetic stem cells (HSCs) of both aged and irradiated mice, but also rejuvenated their HSCs (Chang et al. 2016).   Ruxolitinib (a JAK1/2 inhibitor) showed promise rejuvenating fat progenitors in mice by clearing senescent cells present (Xu et al. 2015).

A treatment that has shown senyolytic properties and potential for combating ageing is hyperbaric oxygen therapy (HBOT; Hachmo et al. 2020). HBOT utilises 100% oxygen in an environmental pressure greater than one absolute atmospheres to increase the concentration of oxygen dissolved in bodily tissues (Hachmo et al. 2020; Amir and Shai 2020). Certain HBOT-induced hyperoxic environments mimic the effects of hypoxia, and can promote stem cell proliferation, mitochondrial biogenesis, neurogenesis and angiogenesis, and expression of hypoxia induced factor, vascular endothelial growth factor (VEGF)

and SIRT (Amir and Shai 2020; Hachmo et al. 2020). Repeated, intermittent HBOT significantly reduced the number of senescent immune system cells and also increased telomere length of immune system cells in healthy adults aged 65+ (Hachmo et al. 2020). It also up-regulated anti-oxidant genes in response to increased ROS, which with repeat exposures becomes protective (Amir and Shai 2020; Hachmo et al. 2020). Since high ROS and mitochondrial dysfunction have been associated with ageing (López-Otín et al. 2013; Singh et al. 2019), it is possible that HBOT rejuvenates via increased anti-oxidants and better regulation of ROS (Cimino et al. 2012; Amir and Shai 2020; Hachmo et al. 2020). Unfortunately, telomere length was the only proxy used by Hachmo et al. to measure biological age, hence further studies (involving DNAm and eAge for example) are needed to quantify and confirm any rejuvenative properties of HBOT.

The clearance of senescent cells through treatments and drugs such as senolytics, have shown promise as an anti-ageing therapies. They may prove highly effective when combined with other rejuvenative techniques (e.g. CR mimics). However, more research is needed to properly assess the beneficial and detrimental side affects of senolytics before they can be used clinically (Mahmoudi, Xu, and Brunet 2019).

1.6.4 Cellular Reprogramming and Rejuvenation

Single Cell Nuclear Transfer

In 1957, Conrad Waddington postulated that once a cell is fully differentiated, it cannot revert back to a pluripotent state (Waddington 1957). The first evidence that cellular differentiation is malleable came shortly after with the development of somatic cell nuclear transfer (SCNT; Gurdon, Elsdale, and Fischberg 1958; Gurdon 1962), where a somatic cell nucleus is transferred into an enucleated, unfertilized egg cell and divides to form an embryo that is genetically identical

to the donor cell. This demonstrated that cells retain the information to become pluripotent and that a somatic cell can, in essence, be rejuvenated (Takahashi and Yamanaka 2016).

Initial cloning experiments with SCNT were conducted with frogs (Gurdon, Elsdale, and Fischberg 1958; Gurdon 1962). SCNT as a cloning process gained publicity when it was used to create the first even cloned mammal, "Dolly" the sheep. An SCNT-derived artificial sheep zygote was implanted into a surrogate mother, and resulted in the birth of a viable cloned sheep genetically identical to the initial donor (Wilmut et al. 1997). One of the first questions raised was regarding the "age" of Dolly's cells (Marión and Blasco 2010). Did the biological age of Dolly's cells match her chAge, or the chAge of her somatic donor? Indeed, the premature death of Dolly (aged 6.5 years), with normal life expectancy of 12 years for Dolly's sheep breed, combined with developing osteoarthritis (Sinclair et al. 2016; Burgstaller and Brem 2017) raised concerns regarding Dolly's biological age.

Telomere length was one of the main biomarkers available to measure age when Dolly was first created (Harley, Futcher, and Greider 1990; Hastie et al. 1990; Lindsey et al. 1991; Frenck, Blackburn, and Shannon 1998). Analysis of Dolly's cells revealed that the telomeres were actually shorter by ~20% compared to age-matched control sheep (Shiels et al. 1999). This observation initially suggested that SCNT does not reset biological age to zero (Marión and Blasco 2010). However, analysis of telomeres of other SCNT-derived sheep (including sheep derived from the same cell line as Dolly) and animals (eg. mice) had normal telomere lengths for their respective age groups (Wakayama et al. 2000; Lanza et al. 2000; Tian, Xu, and Yang 2000; Betts 2001; Clark et al. 2003; Sinclair et al. 2016). Indeed, a recent study showed that SCNT of telomerase haplo-insufficient cells restores telomere length (Sung et al. 2014). The exact reason Dolly had such anomalous health conditions remains a mystery, but it is clear that the reprogramming qualities of the ovum hold

47

rejuvenative potential that could be applied in anti-ageing therapy.

Induced Pluripotent Stem Cells

Groundbreaking work by Takahashi and Yamanaka in 2006 further proved that somatic cell identity is indeed re-writable. They showed that over-expression of four transcription factors (Oct3/4, Sox2, Klf4 and c-Myc, now referred to as the "Yamanaka factors" or "OSKM" factors) rearranges the epigenetic landscape and converts somatic cells to a pluripotent state (Takahashi and Yamanaka 2006). Since the creation of induced pluripotent stem cells (iPSCs) *in vitro*, it has become clear that cellular identity is dictated by epigenetic changes, rather than by loss or alterations of genomic DNA (Singh and Zacouto 2010; Koche et al. 2011). The process of generating iPSCs has been optimised over the years, and has also been achieved via chemical induction, rather than gene expression, in mouse cells (reviewed in Ji et al. 2016; Takeda et al. 2018).

Partial Reprogramming and Epigenetic Rejuvenation

iPSCs offer the promise of directed, personalised therapy (ie. iPSCs grown from patient cells, minimising incompatibility) for diseases that are currently incurable, such as neurodegenerative diseases of the central nervous system, heart infarction, diabetes mellitus, and also liver, lung, and kidney disease varieties (Moradi et al. 2019). However, ethical and safety considerations have to be met before iPSCs can be implemented for *in vivo* procedures (Singh et al. 2015; Takahashi and Yamanaka 2016; Moradi et al. 2019), primarily regarding cancer risk. Direct injection of human embryonic stem cells (hESCs) invariably leads to cancer in mice due to their high proliferation rate and substantial differentiation potential (Hentze et al. 2009; Abad et al. 2013; Ohnishi et al. 2014; Moradi et al. 2019).

It was presumed that as a cell reverts to a pluripotent state the aged epigenome is also reset to zero (Singh and Zacouto 2010; Manukyan and Singh

2012). Indeed, the Horvath clock confirmed that hESCs and iPSCs have an eAge of zero (Horvath 2013). Moreover, the telomeres of iPSCs are longer than the parent differentiated cells, and are comparable in length to telomeres of control hESCs (Marion et al. 2009; Lapasset et al. 2011). Telomere resetting also occurs when reprogramming senescent somatic cells from both centinarians, supercentinarians (albeit at a lower incidence) and HGPS samples (Lapasset et al. 2011; Lee et al. 2020). Oxidative stress, gene expression profiles and mitochondrial metabolism are also reset and indistinguishable from hESCs (Lapasset et al. 2011). If the reversal of age can be uncoupled from complete dedifferentiation, a viable rejuvenation strategy might exist, a process termed "epigenetic rejuvenation" (Singh and Zacouto 2010; Manukyan and Singh 2012; Rando and Chang 2012).

To achieve epigenetic rejuvenation via reprogramming factors, one must look at the intermediate states during dedifferentiation, where cells have started to epigenetically change (presumably de-age), but have not yet fully dedifferentiated (ie. lost somatic identity; Singh and Zacouto 2010; Manukyan and Singh 2012). Partially reprogrammed cells are such examples, which are isolated between days 7 and 11 during OSKM-induced dedifferentiation but have not yet lost their cellular identity (Tanabe et al. 2013). Thus, partial reprogramming is a method of using OSKM factors to revert aged cells to a younger state. The epigenetic rejuvenation potential of partial reprogramming with OSKM factors was initially revealed by expressing OSKM+LIN28 in human senescent fibroblasts, which led to restoration of the high mobility of histone protein 1β by day 9, a key characteristic of young fibroblasts (Manukyan and Singh 2014). Ocampo et al. further demonstrated that partial reprogramming, achieved by transient, periodic induction of OSKM (OSKM expressed for 2 days, then not expressed for 5 days), ameliorates signs of ageing without complete loss of cellular identity (Ocampo et al. 2016). They conducted partial reprogramming first on progeroid mouse fibroblasts and alleviated

age-associated hallmarks, such as DNA damage, nuclear envelope damage, dysregulation of histone modifications, stress and senescence associated factors, and mitochondrial-associated reactive oxygen species (ROS) production. Similar rejuvenation of dysregulated histone modifications was also observed when partial reprogramming was conducted on high-passage human fibroblasts (derived from iPSCs). Moreover, they applied partial reprogramming *in vivo* to progeroid mice, and found that not only was their lifespan extended, but also no cancer or teratomas developed. When repeated in naturally aged mid-life mice, regenerative capacity of muscle and pancreas after injury was improved, as well as glucose tolerance (Ocampo et al. 2016).

1.6.5 Treatments That Reduced Biological Age According to Epigenetic Clocks

Drug Interventions

eAge offers a unique proxy for biological age that allows us to quantifiably measure the extent of rejuvenation of an anti-ageing treatment. Therefore, existing candidate therapies and new drugs can be validated experimentally, with the increase or decrease in eAge providing a measure of their efficacy. For example, vitamin D-sufficient individuals have a lower eAge acceleration according to the Horvath clock, and longer LTL (Chen et al. 2019; Vetter et al. 2020). Epigenetic clocks applied to mouse models are also effective at testing anti-ageing drug candidates. Wang et al. demonstrated this *in vivo* with their mouse liver age predictor. They found that CR, rapamycin and Prop1$^{df/df}$ dwarfism (which also results in up to 1.5 fold life extension) caused significant age deceleration compared to wild type, untreated age-matched control mice (Wang et al. 2017). Other studies have also shown CR to reduce age acceleration in mice according to a mouse (Petkovich et al. 2017) and rat epigenetic clock (Levine et al. 2020).

Horvath et al.   showed that Rapamycin reduces eAge in humans keratinocytes, according to the SB clock (Horvath et al. 2019). Furthermore, they showed that rapamycin in combination with Y-27632 (a rho inhibitor which promotes cell division) reduces eAge without inhibiting cell proliferation. Both the rapamycin and rapamycin+Y-27632 treated cells entered replicative senescence with no increase of eAge. This means the action by which rapamycin reduces eAge is unrelated to cell proliferation or senescence suppression.

Boroni et al. applied their own skin-specific age predictor (see Chapter 1.4.2, pg. 15) to three day rapamycin treated fibroblasts from a HGPS patient (Boroni et al. 2020). They found no significant change in eAge. Cells were incubated with rapamycin for three days, which might not be enough time induce a reduction in eAge. When repeated with skin biopsies (from healthy patients) rather than cell cultures, rapamycin did reduce the eAge of a number of samples. However, the difference in eAge between treated and non-treated skin biopies was not significant. Boroni et al. also treated fibroblasts with ABT-263 (a senolytic) for three days. While the difference in eAge from healthy control cells was not significant, the treated samples exhibited increased eAge the higher the senolytic concentration, despite a decrease in senescent cells and expression of age-associated genes (Boroni et al. 2020).   While it is not clear why this occurred, it does highlight that the mechanisms controlling the epigenetic clock (or at least, in the case of the Boroni et al. skin clock) do not necessarily correlate directly with age related gene expression changes.

While the number of eAge related anti-ageing studies conducted with cell cultures and mouse models is growing, the number conducted in human clinical trials is sparse. Nonetheless, a preliminary study was conducted by the Horvath lab, whereby a "immune intervention" was applied to nine nominally healthy men aged 51-65 over the course of a year (Fahy et al. 2019).   The immune intervention had three components; recombinant human growth hormone (rhGH), dehydroepiandrosterone (DHEA) and metformin.   rhGH is the main

component, which has been shown to have thymotrophic and immune restorative properties (Kelley, Brief, and Westley 1986; Napolitano et al. 2008; Plana et al. 2011). DHEA is an abundant steroid that declines with age, however its role as an anti-ageing supplement and potential side effects is still debated (Samaras et al. 2013). Metformin is a calorie restriction mimic that has shown anti-ageing properties in model organisms, and reduces incidence of age-related disease in humans such as cancer (Soukas, Hao, and Wu 2019). Both DHEA and metformin were also added to minimise any "diabetogenic" effect of rhGH, since rhGH can increase insulin levels (Marcus et al. 1990). There was a $\sim$1.5 year mean eAge decrease from baseline age after 1 year of treatment (Fahy et al. 2019). The rate of eAge decrease was -1.6 years/year between 0 and 9 months, and -6.5 years/year between 9 and 12 months for all clocks tested (Horvath, Hannum, PhenoAge and GrimAge). A reduction of fat in the thymus (a sign of thymic rejuvenation) was observed, and a decrease in CD38+ monocytes (increases are associated with age-related inflammation; Camacho-Pereira et al. 2016), and an increase in lymphocytes (Fahy et al. 2019). While the treatment shows potential as an anti-ageing/immunosenescent therapy, a larger study has to be conducted to validate Fahy et al.'s findings, particularly the mechanism of action of their treatment. Growth hormone (GH) has a large variety of biological effects such as energy homeostasis and metabolism (Vijayakumar, Yakar, and LeRoith 2011), hence the eAge rejuvenation observed here might be due to activity independent of restoring the immune system. In addition, this study highlights the need to carefully explore the role of GH in ageing. Current theories show that decreased GH lead to decreased IGF-1 and insulin signalling, which downstream deactivates mTOR and promotes longevity (Singh et al. 2019). While this has been shown in many animal models such as mouse and worm, their roles in human ageing are unclear (reviewed in Junnila et al. 2013; Reddy and Chaiban 2017; Vitale et al. 2019). It appears that human longevity is more related to insulin sensitivity rather than IGF-1 levels, and the analysis of IGF-1

levels is confounded by its inverse proportionality to insulin sensitivity (Vitale et al. 2019; Fahy et al. 2019). Hence, the effects of IGF-1 and insulin need to be untangled first before rhGH can be used as an anti-ageing therapy (Fahy et al. 2019).

Yamanaka Factor-Induced Rejuvenation

Since Ocampo et al. transiently expressed OSKM in mice (see Chapter 1.6.4, pg. 46), partial reprogramming has become an exciting avenue for rejuvenative research. Unfortunately, eAge prediction for mouse was unavailable for Ocampo et al., hence the exact extent of rejuvenation by partial reprogramming *in vivo* has not yet been determined. Understanding the kinetics of partial reprogramming is essential as an anti-ageing therapy to avoid tumorigenesis. Sarkar et al. conducted a similar experiment to Ocampo et al. in adult human dermal fibroblasts and endothelial cells (Sarkar et al. 2020). They transiently expressed OSKM+LIN28+NANOG (OSKMLN) for four days in aged samples then analysed gene expression and methylation two days after interruption (Ocampo et al. by comparison used a doxycyclin-inducable system and forced expression 2-4 days in cell cultures). Sarkar et al. compared both before and after treatment eAge and expression with young samples. According to the Horvath clock, the OSKMLN treatment significantly reduced age in both the fibroblasts (mean age acceleration = -1.84) and endothelial cells (mean age acceleration = -4.94), although the effect was more pronounced in the latter. In both tissue types, RNA-seq analysis revealed that treated cells were transcriptionally comparable to younger cells rather than the original age cells. This could also be seen to a certain extent in a PCA analysis where in both tissues, treated cells tend to cluster closer to young cells than the original aged cells. Rejuvenative effects were also observed in analysis of other markers in terms of heterochromatin (HP1γ), lamina (LAP2α), proteosomal activity, autophagosome formation and mitochondrial ROS. In addition, expression of cell

identity markers was maintained in the treated cells, meaning no loss of somatic identity occurred (Sarkar et al. 2020). The main caveat however, is the limited sample size; three young (25-35 years), three aged (60-90 years) and therefore three treated samples. A repeat experiment with a larger sample size is required to correctly assess the rejuvenative potential of partial reprogramming. It should also be noted that the Horvath clock does not reliably predict the eAge of fibroblasts (Horvath 2013; Horvath et al. 2018; Horvath et al. 2019). Boroni et al. developed their own skin-specific eAge predictor (see Chapter 1.4.2, pg. 15) and tested it on Sarkar et al's OSKML treated fibroblast dataset (Boroni et al. 2020). According to their skin clock, three of the four samples declined in eAge after treatment, while according to the Horvath clock, only two of the samples actually declined in eAge after treatment. Hence, caution should be taken regarding which epigenetic clock is chosen for a given experiment.

Stem cell decline in functionality and proliferation is an important hallmark and potential cause of ageing (see Chapter 1.3, pg. 8). Can partial reprogramming rejuvenate old stem cells? Sarkar et al. also tested this, by transplanting young, old, and transient OSKMLN-treated old mouse-derived skeletal MuSCs into injured muscles of immunocompromised mice. They observed an improved regenerative ability in the OSKMLN-treated old MuSCs, comparable to that of the young MuSCs, and no teratomas or neoplastic legions developed. Improved muscle function was also observed in the muscles grafted with OSKMLN-treated old MuSCs compared to untreated old MuSCs. Sarkar et al. repeated the experiment with old (60-80 years) human MuSCs, and found that they also had a higher proliferative capacity than the untreated cells, and comparable to young human MuSCs (Sarkar et al. 2020). Taken together, these results provide further evidence that partial reprogramming has potential as an anti-ageing therapy, however, more testing is still required to understand the precise kinetics of the process.

An approach similar to partial reprogramming has recently been used to

restore the vision of mice (Lu et al. 2020). Lu et al. showed that ectopic expression of *Oct4, Sox2* and *Klf4* (OSK) stimulated axon regeneration in an optic-nerve-crush-injury mouse model and also recovers vision in a glaucoma mouse model. The regeneration process requires active DNA demethylation mediated by TET1 and TET2. *c-Myc* was excluded from their treatment to avoid teratoma formation, as it is an oncogene that reduces lifespan in mice (Hofmann et al. 2015). Unfortunately, the regeneration results could not be replicated in aged mice, only the injury model which is not representative of age-induced advance vision loss. Unlike Ocampo et al. who cyclically induced OSKM expression (continuous expression for 2 days out of 7 days), Lu et al. expressed OSK continuously, and showed that 10-18 months of continuous expression did not increase tumour incidence. However, no direct measurement of cell identity or extent of dedifferentiation (eg. somatic or pluripotency genes) was performed. This makes it difficult to confirm epigenetic rejuvenation is occurring, since it is possible that cells are becoming more stem-like and are regenerating the tissue, rather than an age related rejuvenation occurring. This distinction is crucial, otherwise long term cancer development is still possible if cells are becoming more stem-like.

Two epigenetic clocks were used to assess the extent of rejuvenation in their experiments. The retinal ganglion cells (RGCs) eAge were analysed with the ribosomal clock, which was developed for mice and uses DNAm of ribosomal DNA to quantify age. The ribosomal clock revealed that while non-OSK-treated RGCs increased in eAge after injury, the OSK treatment mitigated this effect. The second eAge predictor used was created in the same study using 38 RRBS mouse samples varying in age from 1 to 30 months old, and showed a significant eAge decrease in the treated samples. However, they also included injured mice and OSK-treated samples in their training. There is the possibility that these samples, particularly the OSK-treated samples, might bias the clock to predicting OSK-treated samples. Normally epigenetic clocks are constructed using controls

only, so that deviations due to a disease or treatment are more apparent. The fact that they use two separate clocks for two different experiments within the same study is not optimal, however, it is understandable since RRBS data does not always capture all CpGs required for eAge prediction at the necessary coverage (see Chapter 5, pg. 89; Field et al. 2018; Thompson et al. 2018).

Chapter 2

Aims and Hypothesis of Thesis

Epigenetic age prediction has emerged as a powerful tool that has revolutionised the field of molecular gerontology. It can be used as a proxy to quantify biological age and test the efficacy of anti-ageing interventions. Partial reprogramming has shown promise as a treatment to safely reverse ageing whilst retaining the ability to revert to or maintain original cell identity, both *in vivo* (Ocampo et al. 2016; Lu et al. 2020; Alle et al. 2021) and *in vitro* (Olova et al. 2018a; Sarkar et al. 2020; Gill et al. 2021). However, key questions regarding this treatment remain. To what extent are dedifferentiation and epigenetic rejuvenation intertwined? Do partially reprogrammed cells decrease in biological age and to what extent? What are the kinetics of epigenetic rejuvenation in relation to tissue identity during partial reprogramming?

In Chapter 4, I worked with Nelly Olova to answer some of these questions. I analysed a previously published dataset of fibroblasts reprogramming to iPSCs. We expected that epigenetic age would decline to zero based on previous work (Horvath 2013). However, it remained unclear whether eAge would steadily decline over the time-course, or suddenly drop at a particular time-point. We also were not sure if gene expression trajectories (somatic and pluripotency genes) would match the decline trajectory in eAge. We hypothesised two possible outcomes:

1. As epigenetic age decreases, fibroblast expression decreases at the same rate

2. Decrease in epigenetic age and fibroblast expression follow different trajectories over the time-course.

While eAge is becoming more commonplace as a metric for biological age, caution must be considered when inferring the biological significance of age acceleration. eAge acceleration correlates with disease states, but the precise nature of physiological ageing captured remains unclear (see Chapter 1, Section 1.4.6). Whether the change of methylation is causal to ageing remains to be shown and herein lies a caveat studying diseases or interventions that directly affect DNAm. Studying a process that interacts with DNAm might alter age prediction, without visibly altering the actual ageing trajectory. For example, it is possible that a global increase or decrease in methylation caused either by technical errors (Olova et al. 2018b) or mutations in oncogenes (such as *DNMT3A* or *TET2*; Robertson et al. 2019), could result in false positive shifts of eAge.

To investigate how robust the epigenetic clocks are to global changes in methylation (see Chapter 5), I tested a number of scenarios using DNAm data from the Generation Scotland (GS) study (Smith et al. 2006; Smith et al. 2013). I hypothesised that a global increase or decrease in DNAm will alter eAge prediction. It was not clear whether a global increase in methylation will increase age, or vice versa. The null hypothesis here would be that no change in eAge occurs, which would mean the epigenetic clocks are able to compensate for global changes in DNAm.

Another advancement in the testing of anti-ageing therapies is the development of epigenetic clocks for mice (Wang et al. 2017; Petkovich et al. 2017; Stubbs et al. 2017; Meer et al. 2018; Thompson et al. 2018; Han et al. 2018; Han et al. 2020). The majority of these clocks have been developed for reduced representation bisulphite sequencing (RRBS) data. Unfortunately there

is high variability in the CpGs captured, because regional genome coverage differs between protocols and enzymes used (Field et al. 2018; Thompson et al. 2018). Hence, transferability of mouse clocks to external datasets (i.e. datasets outside of the data used to train the clocks) results in less accurate age prediction. To improve the accuracy of RRBS-based age prediction, myself and Qian Zhao aimed to create mouse clocks that use mean percentage methylation of genomic regions (e.g. 2 kb), rather than individual CpGs, as an input. We expected that these age predictors would outperform other published mouse clocks when tested on external datasets.

Overall, this thesis aims, using epigenetic clocks and transcriptional data, to assess the potential of partial reprogramming as an anti-ageing therapy. It also aims to improve our understanding of eAge prediction and the tools available to measure eAge, in the hopes of more accurately capturing biological age.

Chapter 3

Methods

3.1 Commonly Used Techniques

Computation for each chapter was conducted using programming languages including R and Bash. Specific programme versions are stated in each chapter. Many of the chapters have techniques in common, which are evalutated below.

3.1.1 Bisuphite Conversion and DNAm Processing

All results chapters in this thesis analyse DNAm data from other studies. Briefly, DNAm data are the result of bisulphite (BS) conversion, a process where unmethylated cytosines (C) are converted to uracil (U, which are amplified as thymine, T, residues after PCR), and methylated C residues remain protected (Frommer et al. 1992). In reduced representation bisuphite sequencing (RRBS) experiments (which are typically used for mouse epigenetic age predictors, see Chapter 6), BS-converted DNA is digested by MSP1 which cuts genomic DNA at CpG sites, regardless of their methylation state (Meissner et al. 2005; Gu et al. 2011; Baheti et al. 2016). The resulting fragments are enriched with CpG sites (capturing ~1% of the genome) and processed with next-generation sequencing techniques (Meissner et al. 2008a; Gu et al. 2011).

Illumina-based microarrays (BeadChips) such as the 27K, 450K and EPIC

HumanMethylation arrays are typically used for human epigenetic age prediction since they robustly measure methylation at similar CpG sites (slight differences between each array; Field et al. 2018). These arrays contain approximately 27,578, 482,421 and 850,000 CpG sites respectively (Pidsley et al. 2013; Pidsley et al. 2016). Briefly, BS-converted DNA is applied to an Illumina BeadChip, where the CpGs hybridise to matching probes within a chip. Each probe is an oligonucliotide (approx. 50 bp) attached to 3-micron beads inside the microwells of the BeadChip (Illumina 2015a).

The 450K and EPIC arrays are the most commonly used for epigenetic clocks, both of which use two bead types; Infinium I and Infinium II (Illumina 2015b). Infinium I has two probe types per CpG; one that is designed to bind to unmethylated DNA (T present at CpG site), the other to methylated DNA (C present at CpG site). If unmethylated DNA binds to a probe designed for unmethylated DNA then ends will complement, and the probe will extend by one fluorescently tagged nucleotide (Illumina 2015b). If a probe designed for methylated CpGs binds to an unmethylated CpG (or vice versa), the sequence will not extend by 1 bp. The BeadChip is then scanned and the ratio between methylated and unmethylated DNA per locus is calculated. Infinium II uses one bead type, where the bound CpG to an oligonucleotide, and is determined as methylated or unmethylated depending on if the single fluorescently-tagged nucleotide added is an adenine (A) or a guanine (G). If a G is added, then it has bound to a C, meaning the site is methylated. If an A is added, then the site present is a T, meaning the CpG is unmethylated (Illumina 2015b). The resulting signal (β value) is a ratio between methylated and unmethylated signals which range from 0 (absolutely unmethylated) to 1 (absolutely methylated; Horvath and Raj 2018).

3.1.2 Penalised Linear Regression

As reviewed in Chapter 1.4, epigenetic clocks are constructed using CpGs that have tractable changes with age. These CpGs are selected using penalised regression models such as elastic net (Zou and Hastie 2005) or least absolute shrinage operator (LASSO; Tibshirani 1997), which selects a group of CpGs that have a linear relationship with age in a given training dataset without overfitting (Horvath and Raj 2018; Field et al. 2018). The resulting linear model is as follows:

$$eAge = I + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_m X_m \text{ (3.1)}$$

where $eAge$ represents the epigenetic age, I is the intercept value where the linear model would meet the y-axis, β represents the average methylation score of a particular CpG or genomic region, and X is the coefficient/weight (Hepp et al. 2016; Field et al. 2018). Hence, eAge can be predicted by inputting the β values from a sample into equation 3.1.

3.1.3 Calculating adj. R^2

R^2 is a statistical model derived from analyses based on the general linear model, such as linear regression. It represents the proportion of variance in the outcome variable (eg. chAge) which is explained by the predictor variables (eg. eAge) (Miles 2014). In other words, we are investigating what percentage of variance in chAge can be explained by DNAm/eAge and to find the strength of correlation between the two. For this type of analysis the adjusted R^2 is typically used as it corrects for population level analysis:

$$Adj.R^2 = 1 - (1 - R^2)((N - 1))/((N - k - 1)) \text{ (3.2)}$$

where N represents the number of samples, and k represents the number of predictors (Miles 2014).

3.2 Methods for Chapter 4: Partial Reprogramming Induces Steady Decline in Epigenetic Age Before Loss of Somatic Identity

3.2.1 Overview of the Ohnuki et al. Experimental Setup and Datasets

450K DNA methylation array and gene expression microarray data of full HDF reprogramming time-course was obtained from GSE54848. A schematic of experimental setup and time points is provided in **Fig. 4.1**. Briefly, HDF cells were transfected with EGFP-labelled OSKM on day 0 and cultured in virus-containing medium for 24 hours, then replaced by 10% FBS-containing medium for 8 days before replacing with human ESC medium. EGFP (+) cells, representing the population of successfully transfected cells, which permanently express the OSKM factors, were sorted by flow cytometry on day 3. Intermediate reprogrammed cells positive for the pluripotency marker TRA-1-60 were sorted by magnetic activated cell sorting on days 7, 11, 15, 20 and 28 post-transfection. Day 28-sorted TRA-1-60 (+) cells were further expanded and samples collected three more times on each seventh day, i.e. on days 35, 42 and 49. Thus, sorted and collected cells at each time point were subjected to both gene expression and DNA methylation array analysis. Microarray gene expression (data available as LOG2 transformed) was performed for three to four replicates per data point, whilst DNA methylation data was performed for two to three replicates per time point.

3.2.2 Predicting eAge

The pre-processed 450K DNA methylation array matrix of average methylation per CpG site of the full HDF reprogramming time-course was obtained from

GSE54848 (downloaded using getGEO function from GEOquery package) and uploaded to the online DNA methylation age calculator to assess eAge:

https://labs.genetics.ucla.edu/horvath/dnamage/ (Horvath 2013). Data processing including Horvath's normalisation was performed according to tutorial guidelines. Missing CpG values were imputed by Horvath's online DNAm age calculator. During QC, around 1600 CpGs were lost, therefore methylation data for each time point contained 26,987 CpG sites out of the suggested 28,587 CpGs, a fact unlikely to have any significant impact on the normalisation or age prediction. PhenoAge, Skin & Blood, Hannum, Weidner 99 and 3 CpG age predictors were applied to average methylation values. Missing CpG values were imputed as zero before applying these age predictors. All ages presented in the manuscript are calculated eAges, no actual ages of HDF donors were available.

3.2.3   Methylation Age Trajectories

For the Horvath multi-tissue age predictor, a 'broken stick' model with two linear sections was constructed to chart overall change in DNA methylation age over time between the three HDF cell lines. A linear mixed model was then specified with a random intercept term for each replicate. A variable break point was set between the minimum and maximum day, plus and minus a small constant (3 days), respectively. The predicted values from the regression models were plotted against the measurement day. For all other age predictor plots (**Fig. 4.3**), mean eAge was calculated for all samples at each time point (2-3 samples depending on the time point) and plotted against time during the time-course. Standard deviation for eAge was also calculated and plotted as error bars at each time point.

3.2.4 Gene Clusters and Trajectories

For each gene in a category (e.g. pluripotent gene list), a loess curve with a span of 0.5 was fitted with the predicted values extracted at each time point. The predicted values were then normalised within each gene to a value of 1 at the first time point and a value of 0 and the last time point (and vice versa for the pluripotent genes). K-means clustering for longitudinal data was applied to determine the optimal number of trajectories within each gene category. All analyses were performed in R 3.4.3, using the kml (Genolini et al. 2015), lme4 (Bates et al. 2015), and lmerTest (Kuznetsova, Brockhoff, and Bojesen Christensen 2016) packages.

3.3 Methods for Chapter 5: How Robust are Epigenetic Clocks to Global Methylation Changes?

All computation was conducted in R-3.6.1. All figures were plotted using ggplot2 (ver. 3.3.3; Wickham 2016).

3.3.1 Data Used for Analysis

Generation Scotland (GS) is a large cohort (23,960 individuals) where 94.2% have at least one other first-degree family member participating in the study (as previously described in Smith et al. 2006; Smith et al. 2013; Hillary et al. 2020). Batch one of Generation Scotland (GS1, n=5100, aged 18-94) was used for our study to minimise batch effects. DNA methylation levels were quantified with the Illumina HumanMethylationEPIC BeadChip Array on blood samples from GS participants.

3.3.2   Sampling from Generation Scotland

Samples for each scenario were obtained randomly from GS using the sample_n function from dplyr ver 1.0.5 (Wickham et al. 2021). Test and control groups do not overlap in samples, neither do test groups within each scenario. Analysis was conducted using CpGs for the Horvath, Hannum and S&B clocks only, rather than using all the sites in the EPIC array.

3.3.3   Age Prediction and Simulating Methylation Changes

Horvath epigenetic age was calculated using the agep function from the R package Bigmelon (Gorrie-Stone et al. 2019). Hannum (Hannum et al. 2013) and S&B (Horvath et al. 2018) ages were calculated in R using code provided by the original studies. A for loop was made in R which applies a relative or absolute change in methylation to a temporary copy of the methylation array (clock sites only). Each clock was applied to the altered data, and the output of ages were stored as a column in a metadata matrix in R.

Relative methylation change was applied by multiplying each mean percentage methylation (β value, varying between 0 (0%) and 1 (100%)) clock CpG by a percentage (eg. β*0.9 or 1.1 to get minus or plus 10% methylation respectively).   Absolute methylation change was calculated by adding or subtracting an absolute percentage value to each clock CpG β (eg. subtracting or adding 0.1 to get minus or plus 10% methylation respectively).   Age acceleration was calculated by subtracting chAge from predicted eAge. For **Fig. 5.1D-E,J-K**, the age acceleration was calculated, then mean average age acceleration for each age group was calclulated and plotted. Only one sample was aged 90 and was excluded from the mean average age acceleration plots in **Fig. 5.1D-E,J-K**. Standard deviation for mean age acceleration was calculated with plotrix 3.8-1 (Lemon 2006). Statistical comparisons between all test and control groups were calculated with the Welch's T-Test in R. Linear regression

lines in **Figures 5.1** and **5.2** were calculated and plotted using ggplot2, and all individual data points were hidden from the plot (except in **Fig. 5.2A,C**).

3.4 Methods for Chapter 6: Robust Age Prediction in Mouse Based on Average Methylation per Genomic Region

3.4.1 Data Collection

The following mouse RRBS datasets were downloaded from the NCBI Gene Expression Omnibus (GEO); GSE120137 (n=548; Thompson et al. 2018), GSE60012 (n=91; Reizel et al. 2015), GSE93957 (n=61; Stubbs et al. 2017), GSE80672 (n=162; Petkovich et al. 2017), GSE85251 (n=64; Reizel et al. 2018) and GSE121141 (n=81; Meer et al. 2018). The annotation information (metadata) of the sequencing samples, such as age and tissue, was downloaded from the NCBI SRA Run Selector. The unit of the chronological age is different among datasets, including month and week. For the convenience of subsequent calculations, all ages in weeks were converted to age in months. The conversion formula is:

$$Age(mo) = Age(wk) \times 7 \div 30.42 \tag{3.3}$$

where $Age(mo)$ is an age in months, $Age(wk)$ is an age in weeks which is multiplied by 7 (week in days), and 30.42 is the average days per month.

3.4.2 Data Processing

The following processes were implemented via bash scripts on Eddie, a high performance computing cluster with a Linux-based operating system provided by the University of Edinburgh.

Read adaptors were removed by TrimGalore (ver 0.5.0, –rrbs) which

combines FastQC (ver 0.11.4) and cutadapt (ver 1.9.1; Krueger 2012). The following processing steps were conducted with Bismark (ver 0.18.1), a utility designed to process and map BS-seq data using packages such as bowtie2 (mapping, ver 2.2.6) and samtools (read processing, ver 1.6; Krueger and Andrews 2011):

1. A bisuphite converted version of GRCm38.p6 reference genome was created by the Bismark genome preparation module (Krueger 2016).

2. The trimmed reads were aligned to the bisulphite converted genome with Bismark (settings: –multicore 2 –phred33-quals -N 0 -L 20).

3. The Bismark methylation extractor module then obtained (from BAM files generated in the second step) the mapped reads and methylation count information (CpG, CHG and CHH contexts). CHG and CHH sites were removed.

4. The bismark2bedGraph module was used to extract the methylation information from CpG_context files and convert them to cov files (bismark.cov.gz) which contain six columns: (1) chromosome number; (2) start position; (3) end position; (4) percentage of methylated reads in total reads (β score); (5) the number of methylated reads; (6) the number of unmethylated reads (Krueger 2016).

3.4.3 Coverage Assessment

To compare whether there was a difference in coverage between all downloaded datasets (see "Methods: Data Collection" for list of datasets), the number of reads from the cov files was extracted (using bash and python scripts). The coverage of each CpG site was calculated as follows:

$$coverage = methylated\,read\,counts + unmethylated\,read\,counts$$

(3.4)

The coverage of clock CpGs (Stubbs and Meer clocks) was extracted and organised into a matrix, which was loaded into R (ver 4.0.3). Each column represented a CpG site and each row represented a sample. Coverage of clock sites in different datasets was plotted using ggplot2 (ver. 3.3.3; Wickham 2016) in R (**Fig. 6.1**).

3.4.4 Applying Published BS-seq Clocks to Various Datasets

Two single tissue (Wang and Petkovitch) and three multi-tissue (Stubbs, Meer and Thompson) mouse BS-seq clocks were tested on various datasets. For each clock, the CpG sites were extracted from a given dataset the clock would be applied to (**Table 6.2**). Each clock was ran according to scripts and instructions from their respective publications, eg. the toRun_Imputation.R script provided by Stubbs et al. (2017) was used to run their clock (Wang et al. 2017; Petkovich et al. 2017; Stubbs et al. 2017; Meer et al. 2018; Thompson et al. 2018).

To evaluate the results, the adjusted R squared (adj. R^2) and median absolute error (MAE) were calculated, and linear regression lines were illustrated with ggplot2 in R (Wickham 2016). MAE is the median absolute error between epigenetic age and the chronological age, where if a test dataset has an MAE of X, then the age acceleration will differ by less than X in 50% of the samples (Horvath 2013). The formula is as follows:

$$MAE = median\left(|predicted\,age - chronological\,age|\right) \quad (3.5)$$

3.4.5 Data Filtering

Before constructing an epigenetic clock, the data needed to be filtered. We choose 3-100 reads per CpG as the filter criteria since sites with less than 3 reads are likely caused by sequencing errors or mapping errors caused by SNPs. Sites with more than 100 reads could be due to amplification bias.

3.4.6   Genome Segmentation

Genome segmentation was conducted using SeqMonk (ver 1.47.0; Andrews 2007), an interactive desktop application with a suite of tools for analysing mapped genomic data. GRCm38_v100 was selected as the reference genome. Next, the filtered .cov files were imported into SeqMonk, "Define Probe" was selected in "Data", and then "Running Windows Generator" was selected to set the window/region size (probe size) and step size.   After setting the probe, "Bisulphite methylation over features" was selected from "Quantitation Pipelines" to calculate the average methylation level per window/region. The formula of the average methylation level is:

$$methylation\,score = (\sum methylation\,counts)/(\sum total\,counts)\,(3.6)$$

where $counts$ are the number of corresponding reads in a specific region. This produced a counts matrix, where samples are columns and rows are genomic regions, which was exported as a text file. Each dataset (see Methods: Data Collection 67) was segmented for all window sizes (1-9kb) and were used for both training and testing the region clocks.

3.4.7   Epigenetic Clock Construction Using LASSO

Each text file of region counts per region clock iteration (generated by SeqMonk) was loaded into R. All percentage methylation values for each region were divided by 100, so that 0-1 would represent 0% to 100% respectively. Regions with NaN (ie. no reads recorded) were removed). Least Absolute Shrinkage and Selection Operator (LASSO) regression model from the glmnet package (ver 4.1-1; Friedman, Hastie, and Tibshirani 2010) was applied. A list of regions, their coefficients, and an intercept value is then generated.

Nine regional blood clocks (RBCs) and regional multi-tissue clocks (RMTCs) were generated (1-9 kb region sizes, **Table 6.3**). Whole blood from GSE80672 (training=141, test=21; Petkovich et al. 2017) was used to construct the RBC.

Liver, muscle, spleen, lung, heart, cerebellum and hippocampus tissue from GSE60012 (Reizel et al. 2015) and GSE85251 (Reizel et al. 2018) was used to construct the RMTC (training=118, test=13). Only tissues with samples at two different time points minimum were used, excluding tissues that only had samples at single time points. The multi-tissue and blood datasets were randomly partitioned into 90% training and 10 % test using the caret package (ver 6.0-86; Kuhn 2020) in R.

Each clock was applied to a particular dataset using a for loop written in R. It extracts each required CpG region (generated from LASSO penalised regresson) from each sample in the counts matrix (generated from SeqMonk) and multiplies it with its corresponding weight (generated from the LASSO penalised regression) as according to equation (3.1). Each iteration of the RBC was applied to the blood 10% test dataset (adj. R^2 values were plotted in **Fig. 6.2B**, left panel) and the Thompson et al. (2018) 3 mo blood samples (GSE120137) (adj. R^2 and MAE values were recorded in **Table 6.3**). All iterations of the RMTC was applied to Stubbs et al. (2017) (GSE93957) and Thompson et al. (2018) (GSE120137), and their total adj R^2 and MAE were recorded in **Table 6.3**. Each clock generated was then applied to the 90% training and 10% test data, and their adj. R^2 values were plotted in **Fig. 6.2B**, right panel.

Chapter 4

Partial Reprogramming Induces Steady Decline in Epigenetic Age Before Loss of Somatic Identity

4.1 Introduction

The human ageing process is accompanied by multiple degenerative diseases. Our understanding of such ageing related disorders is, nevertheless, fragmented, and the existence and nature of a general underlying cause are still much debated (Faragher 2015; Gladyshev and Gladyshev 2016). The generation of induced pluripotent stem cells (iPSCs) allows the reprogramming of somatic cells back to an embryonic stem cell (ESC) like state with an unlimited regenerative capacity. This has led to multiple strategies for tissue replacement in degenerative diseases (Takahashi et al. 2007). Clinical application of iPSCs however, is at its infancy (Singh et al. 2015; Takahashi and Yamanaka 2016; Moradi et al. 2019), and the potency of iPSCs bears risks, not least cancer induction. For example, *in vivo* experiments with iPSCs have shown that continuous expression of Yamanaka factors (Oct4, Sox2, Klf4 and c-Myc, thus OSKM) in adult mice invariably leads to cancer (Hentze et al. 2009; Abad et al. 2013; Ohnishi et al. 2014; Moradi et al.

2019).

To avoid this risk, a parallel concept of epigenetic rejuvenation has been proposed: the ageing process in cells can be reversed whilst avoiding dedifferentiation (Singh and Zacouto 2010; Manukyan and Singh 2012; Rando and Chang 2012). In other words, an old dysfunctional heart cell could be rejuvenated without the need for it to be passed through an embryonic/iPSC state. The concept of epigenetic rejuvenation requires that rejuvenation and dedifferentiation each follow a distinct pathway. Nevertheless, it is not well understood whether rejuvenation and dedifferentiation are invariably intertwined, or instead whether it is possible to manipulate age without risking dedifferentiation.

The epigenetic rejuvenation potential of partial reprogramming with OSKM factors was previously shown by the forced expression of OSKM+LIN28 in senescent human fibroblasts, which led to recovering the high mobility of histone protein 1β by day 9, a feature characteristic for young fibroblasts (Manukyan and Singh 2014). Ocampo et al. further demonstrated that partial reprogramming by transient cyclic induction of OSKM ameliorates signs of ageing and extends lifespan in progeroid mice, with no resulting teratoma formation (Ocampo et al. 2016). This established partial reprogramming as a promising candidate intervention for age-related disease. Estimating epigenetic age (eAge), which is currently the most promising proxy for biological age (Jylhävä, Pedersen, and Hägg 2017; Wagner 2017), was, however, not possible to measure in mice at the time of the Ocampo study. This has left the nature (ie. dedifferentiation/rejuvenation) of the described cellular changes unexplored:

1. Does the epigenetic remodelling seen truly reflect rejuvenation (i.e. a reduction in cellular/tissue age)? If so, can we observe a decrease in eAge in partially reprogrammed human cells?

2. What is the extent of rejuvenation upon reaching a partially reprogrammed state (e.g. years of eAge decrease)?

3. What are the dynamics of dedifferentiation in early reprogramming?

A major obstacle in understanding the relationship between differentiation and ageing has been our inability to accurately measure cellular age with a high correlation to the chronological age (chAge) of the organism. However, over the last five years a number of age predictors have been developed, the most accurate of which utilise DNA methylation (known as epigenetic clocks; Bocklandt et al. 2011; Koch and Wagner 2011; Hannum et al. 2013; Horvath 2013; Weidner et al. 2014; Levine et al. 2018; Lu et al. 2019a; Zhang et al. 2019), with the first multi-tissue age-predictor being the most widely applicable and used (r=0.96). Referred to as the "Horvath clock", this clock shows the highest correlation to chAge, predicting the age (or eAge) of multiple tissues with a median error of 3.6 years (Horvath 2013). eAge is distinct from and poorly correlated with other age-related biomarkers, such as cellular proliferation, senescence and telomere length, which have been shown to correlate independently with eAge (Lowe, Horvath, and Raj 2016; Marioni et al. 2016; Kabacik et al. 2018; Horvath et al. 2019; Cypris et al. 2020). Moreover, an acceleration of eAge as measured by the Horvath clock is associated with a higher risk of all-cause mortality (Marioni et al. 2015a; Christiansen et al. 2016; Perna et al. 2016)), premature ageing syndromes (Down and Werner) (Horvath et al. 2015a; Maierhofer et al. 2017), frailty and menopause (Breitling et al. 2016; Levine et al. 2016). All of these studies suggest that eAge may capture a degree of biological ageing.

To delineate the crosstalk of epigenetic rejuvenation and de-differentiation, here we analysed the dynamics of eAge and somatic gene expression during the course of iPSC reprogramming of human dermal fibroblasts (HDFs). We observe onset of a continuous decline of eAge after day 3 from induction with Yamanaka factors, accompanied by decline, but not loss of somatic gene expression. Our results suggest a "safe" window within the reprogramming time-course, during which epigenetic rejuvenation might be achieved without de-differentiation.

4.2 Results

4.2.1 Epigenetic Age Shows a Continuous Decline After Day 3 of Reprogramming

To understand the dynamics of eAge during reprogramming, we applied Horvath's multi-tissue age predictor over a previously published reprogramming time-course on human dermal fibroblasts (HDFs; Horvath 2013; Ohnuki et al. 2014). After OSKM transfection, successfully transformed subpopulations were isolated and analysed at regular time points during 49-days for gene expression and DNA methylation (detailed schematic shown in **Fig. 4.1**). Epigenetic rejuvenation, ie. decrease of eAge, commenced between days 3 and 7 after OSKM transduction in the partially reprogrammed TRA-1-60 (+) cells (characterised in Tanabe et al. 2013) and continued steadily until day 20, when eAge was stably reset to zero (**Fig. 4.2**). A broken stick model (comprising two linear regressions joined at a break-point), showed a good fit to the observed data starting from day 3, and measured a steady decrease with 3.8 years per day until day 20 (SE 0.27, P = 3.8x10-7) (**Fig. 4.2**). The TRA-1-60 (+) cell populations at days 7 and 11 have been previously characterised as 'partially reprogrammed' for their high expression of pluripotency markers but also high reversion rates towards somatic state (Tanabe et al. 2013). Therefore, the observed eAge decline at days 7 and 11 suggests that partial reprogramming can indeed be considered a rejuvenation mechanism in human cells.

Horvath's multi-tissue age predictor is the most accurate and widely used for various cell types and tissues (Wagner 2017). Nevertheless, we calculated eAge from alternative DNA methylation-based age predictors: four tissue-specific clocks (Hannum et al. 2013; Weidner et al. 2014; Horvath et al. 2018), one that incorporates clinical measures, called PhenoAge (Levine et al. 2018), and individual CpGs previously correlated with age (Garagnani et al. 2012). All clocks consistently reached the point of reset to their iPSC eAge at day 20,
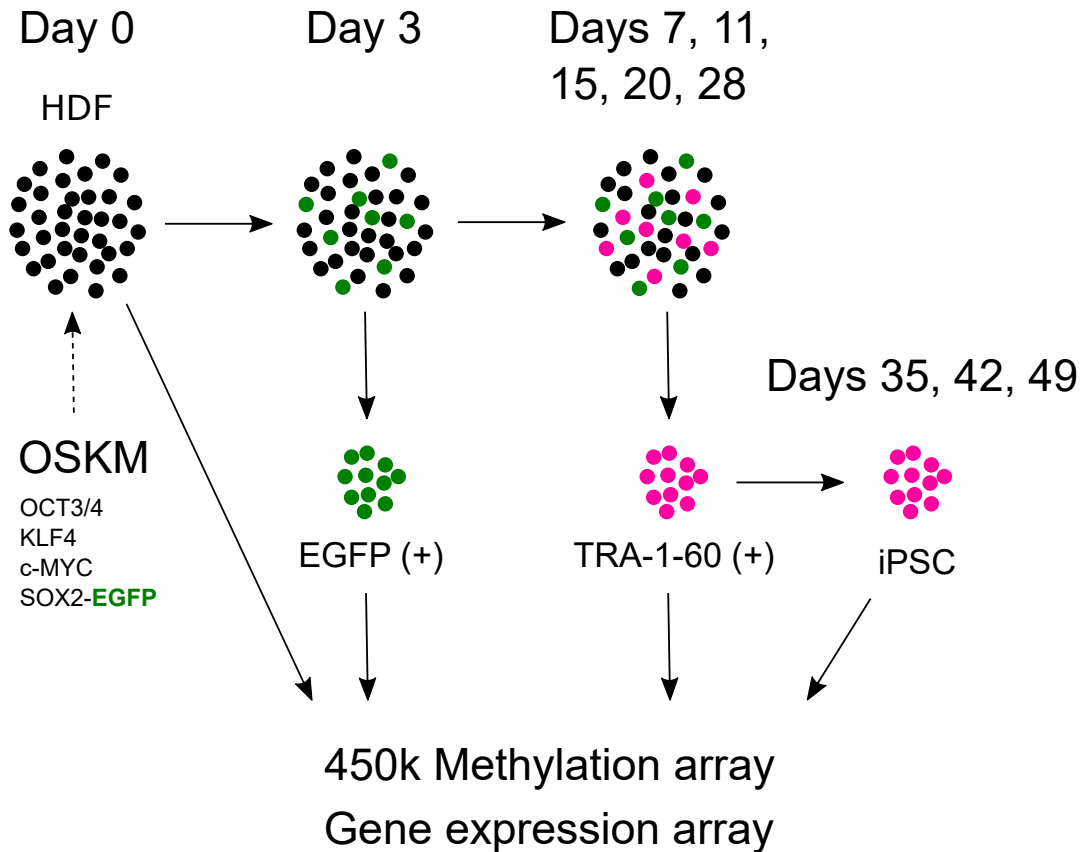
**Figure 4.1: Schematic overview of the experimental setup of Ohnuki et
al iPSC time-course and dataset time points.** HDF cells were transfected
with EGFP-labelled OSKM on day 0 and cultured in virus-containing medium
for 24 hours, then replaced by 10% FBS-containing medium for 8 days before
replacing with human ESC medium. EGFP (+) cells, representing the population
of successfully transfected cells, were sorted by flow cytometry on day 3.
Intermediate reprogrammed cells positive for the human pluripotency marker
TRA-1-60 were sorted by magnetic activated cell sorting on days 7, 11, 15, 20 and
28 post-transfection. Day 28-sorted TRA-1-60 (+) cells were further expanded
and samples collected three more times on each seventh day, i.e. on days 35,
42 and 49. The sorted and collected cells at each time point were subjected to
both gene expression and CpG methylation array sequencing. Microarray gene
expression was performed for three to four replicates per data point, whilst DNA
methylation was performed for two to three replicates per time point. Both 450K
DNA methylation array and gene expression microarray datasets were obtained
from GSE54848 (Ohnuki et al. 2014). Figure created by Nelly Olova, based on
Fig. 1A in Ohnuki et al. 2014.
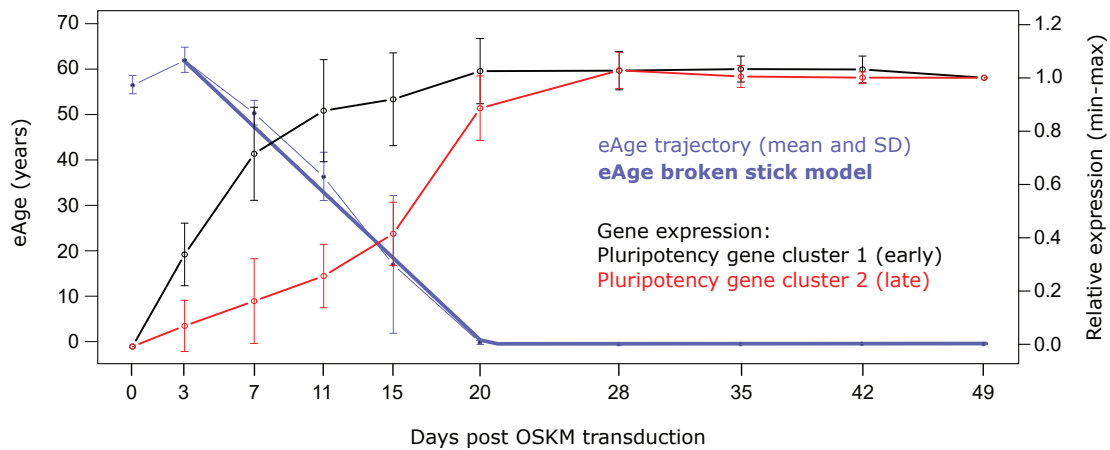
**Figure 4.2: Dynamics of eAge and pluripotent gene expression in a 49-day
HDF reprogramming time-course.** Left Y axis: eAge trajectory of Horvath's
multi-tissue age predictor calculated from DNA methylation arrays from the
following cell populations: day 0 (HDFs), day 3 (OSKM-expressing EGFP (+)
HDFs), day 7, 11, 15, 20 and 28 (human pluripotency marker TRA-1-60 (+) cells
at intermediate stages of reprogramming), and fully reprogrammed iPSCs from
days 35, 42 and 49. Data was fit with a broken stick model composed of two
linear sections. Error bars represent SD. Measured rate (years per day) of eAge
decrease [day 3 - day 20] = -3.8, SE 0.27, P = 3.8x10-7. Right Y axis: Composite
gene expression trajectories of key pluripotency markers, statistically clustered as
per Genolini et al. 2016. Microarray expression data was obtained for the same
time points and cell populations as for eAge. Relative expression values were
LOG2 transformed and presented as arbitrary units starting from '0' for 'day 0' to
'1' for 'day 49'. Error bars represent SD.
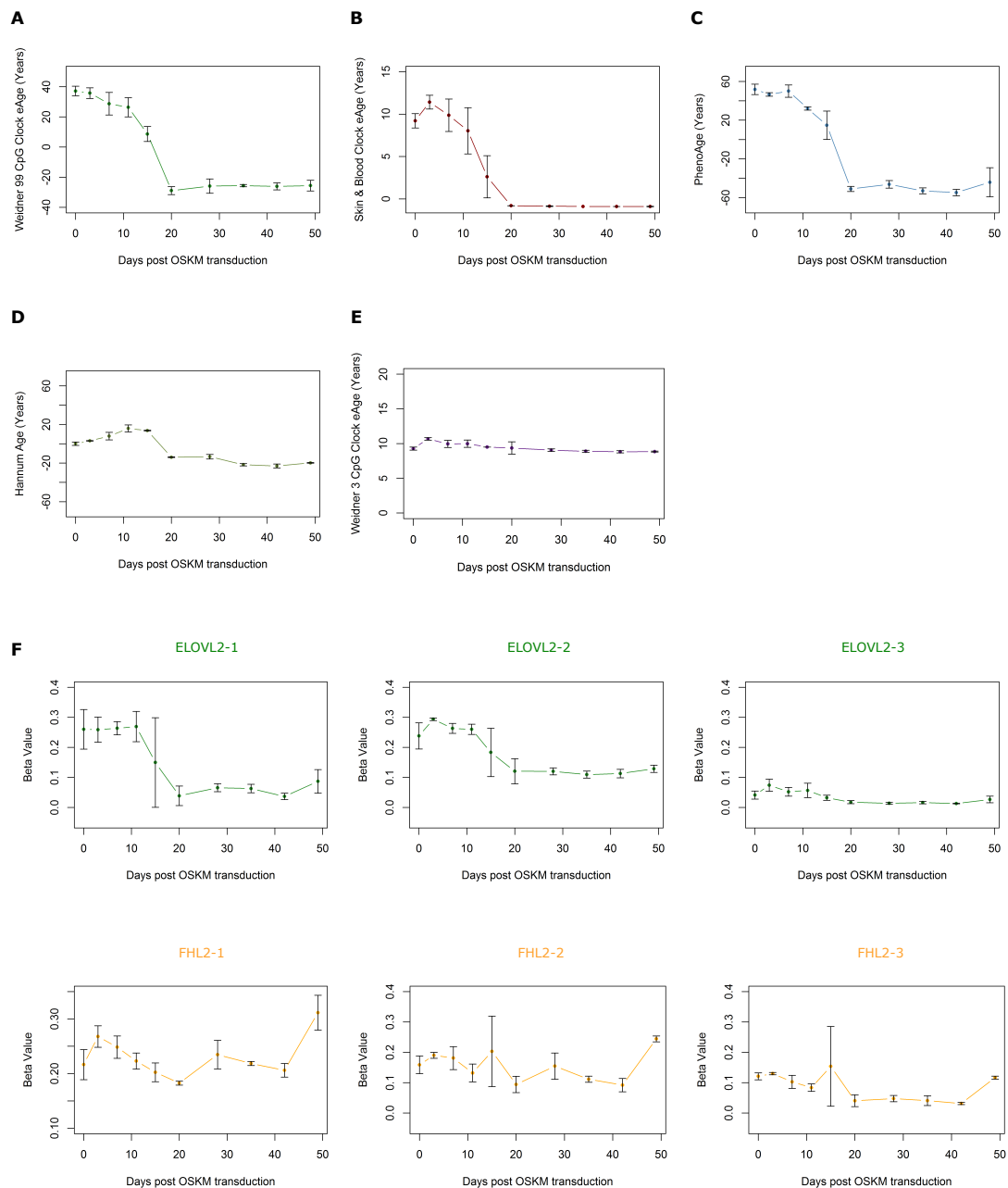
Figure 4.3: eAge trajectories of different DNA methylation-based epigenetic clocks. (**A**) Weidner 99 CpG blood-based epigenetic clock (Weidner et al. 2014); (**B**) Skin blood clock (Horvath et al. 2018); (**C**) PhenoAge (Levine et al. 2018); (**D**) Hannum blood-based epigenetic clock (Hannum et al. 2013); (**E**) Weidner 3 CpG epigenetic clock (Weidner et al. 2014); (**F**) Individual CpG age predictors found in CpG islands within the ELOVL2 and FHL2 genes (Garagnani et al. 2012).

despite the cells not being fully reprogrammed before day 28 (Ohnuki et al.
2014) (Ohnuki et al. 2014) (**Fig. 4.3**). Again, eAge showed a steady decline from
day 3 to day 20 in the skin blood and Weidner 99 CpG clocks, PhenoAge
declined from day 7 to day 20, while the Hannum and Weidner 3 CpG clocks did
not produce informative trajectories.   Overall, eAge values and 'years' of
decrease varied between the clocks (actual chAge of HDF donors was not
available for reference; **Fig. 4.3**). The highest age associated individual CpG
(ELOVL2's cg16867657) showed a similar trajectory to the Horvath eAge
decline, however, the remaining CpGs produced inconsistent trajectories (**Fig.
4.3**). The observed differences are not surprising, given the alternative clocks
were validated for blood (Hannum et al. 2013; Weidner et al. 2014), forensic
applications (Horvath et al. 2018), whole organisms (Levine et al. 2018) or
various tissues as for the individual CpGs (Garagnani et al. 2012).

4.2.2 Loss of Somatic Gene Expression is Uncoupled from Rejuvenation Dynamics and Occurs Step-Wise.

In Ocampo et al. (2016), partial reprogramming was achieved after just two days
of OKSM induction in mice carrying an inducible OSKM transgene. However,
such 'secondary' systems for direct reprogramming are known to have up to
50-fold higher efficiency and accelerated kinetics in comparison to virally
transduced *in vitro* systems (Wernig et al. 2008). To facilitate comparison to
other systems and associate eAge with intermediate states in the
reprogramming trajectory we compared it to gene expression measured in the
same samples. We analysed corresponding microarray expression data for 19
well-established pluripotency marker genes (**Table. 4.1** and **Fig. 4.4**) as a proxy
for reaching a mature pluripotent state (Ginis et al. 2004; Boyer et al. 2005; Cai
et al. 2006; Mallon et al. 2013; Galan et al. 2013). We statistically clustered the
expression patterns of those genes (Genolini et al. 2015), which resulted in two
composite trajectories.   These trajectories followed previously described

Table 4.1: List of pluripotency and fibroblast marker genes used in gene expression clusters. Key pluripotent marker genes were selected from Ginis et al. (2004); Cai et al. (2006); Mallon et al. (2013); Galan et al. (2013); Boyer et al. (2005). Fibroblast marker genes were selected from Kalluri Zeisberg (2006); Zhou et al. (2016); Janmaat et al. (2015); Pilling et al. (2009); Chang et al. (2014); Goodpaster et al. (2008); MacFadyen et al. (2005).

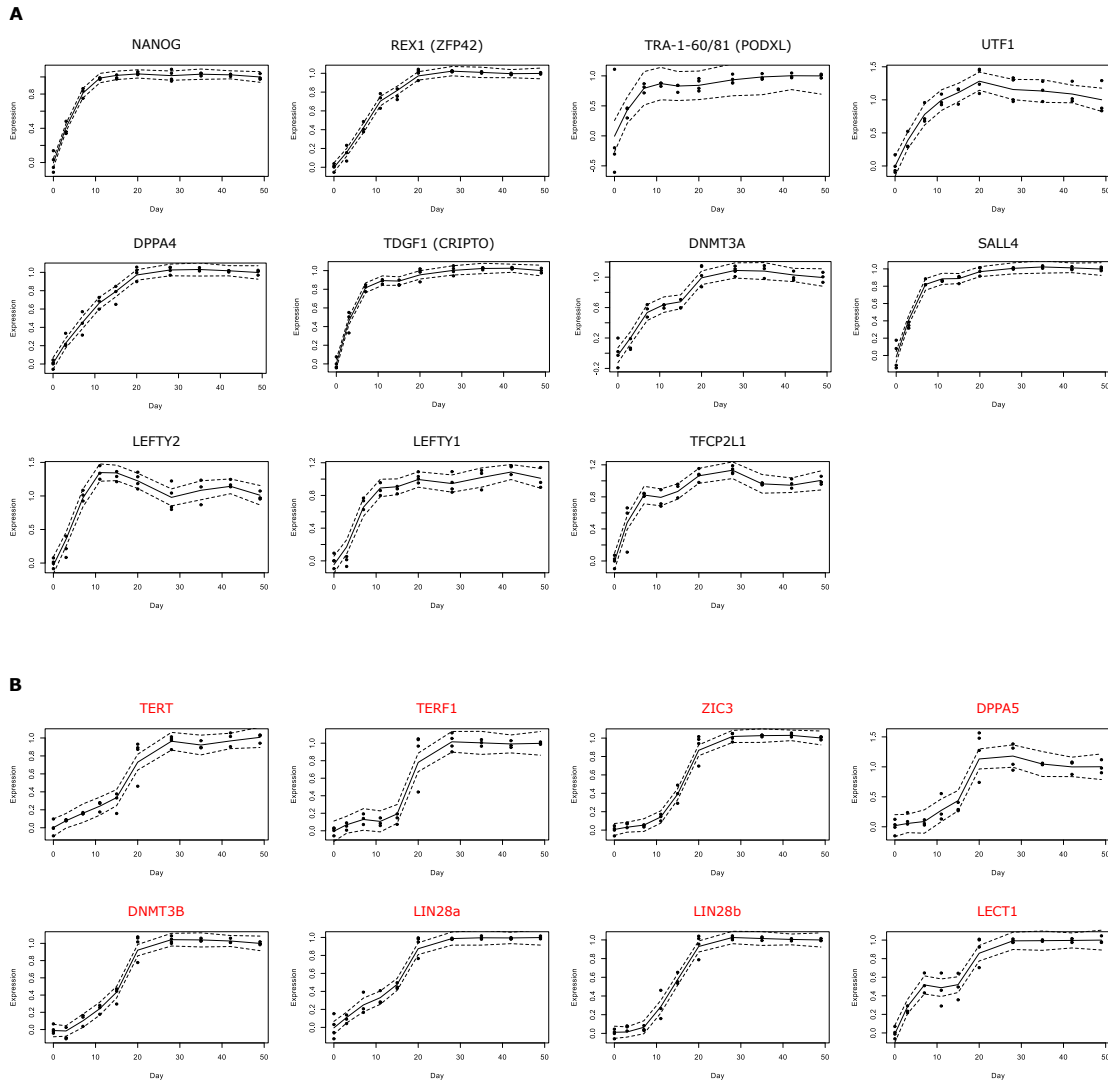| Marker	Gene	Protein name	Accession	Cluster
Pluripotency	NANOG	Nanog homeobox	A_23_P204640	1 (early)
Pluripotency	REX1 (ZFP42)	Zinc Finger Protein 42	A_23_P395582	1 (early)
Pluripotency	TRA-1-60/81 (PODXL)	Podocalyxin	A_23_P215060	1 (early)
Pluripotency	UTF1	Undifferentiated embryonic cell transcription factor 1	A_33_P3294217	1 (early)
Pluripotency	DPPA4	Developmental pluripotency associated 4	A_23_P380526	1 (early)
Pluripotency	TDGF1 (CRIPTO)	Teratocarcinoma-derived growth factor 1	A_23_P366376	1 (early)
Pluripotency	SALL4	Spalt like transcription factor 4	A_23_P109072	1 (early)
Pluripotency	LEFTY1	Left-right determination factor 1	A_23_P160336	1 (early)
Pluripotency	LEFTY2	Left-right determination factor 2	A_23_P137573	1 (early)
Pluripotency	DNMT3A	DNA methyl-transferase 3A	A_23_P154500	1 (early)
Pluripotency	TFCP2L1	Transcription factor CP2 like 1	A_23_P5301	1 (early)
Pluripotency	TERF1	Telomeric repeat binding factor (NIMA-interacting) 1	A_23_P216149	2 (late)
Pluripotency	DPPA5	Developmental pluripotency associated 5	A_32_P233950	2 (late)
Pluripotency	TERT	Telomerase reverse transcriptase	A_23_P110851	2 (late)
Pluripotency	ZIC3	Zic family member 3	A_23_P327910	2 (late)
Pluripotency	LIN28a	LIN28 homolog A	A_23_P74895	2 (late)
Pluripotency	LIN28b	LIN28 homolog B	A_33_P3220615	2 (late)
Pluripotency	LECT1	Leukocyte cell derived chemotaxin 1	A_23_P25587	2 (late)
Pluripotency	DNMT3B	DNA methyl-transferase 3B	A_23_P28953	2 (late)
Fibroblast	COL3A1	Pro-collagen α2(III)	A_24_P935491	1
Fibroblast	FSP-1	Fibroblast surface protein	A_23_P94800	1
Fibroblast	TGFB3	Transforming growth factor beta 3	A_23_P88404	1
Fibroblast	TGFB2	Transforming growth factor beta 2	A_24_P402438	1
Fibroblast	COL1A2	Pro-collagen α2(I)	A_24_P277934	2
Fibroblast	ITGA1	Integrin a1b1 (VLA-1)	A_33_P3353791	2
Fibroblast	DDR2	Discoidin-domain-receptor-2	A_23_P452	2
Fibroblast	P4HA3	Prolyl 4-hydroxylase	A_24_P290286	2
Fibroblast	THY1	Thy-1 cell surface antigen; CD90	A_33_P3280845	2
Fibroblast	FAP	Fibroblast activation protein	A_23_P56746	2
Fibroblast	CD248	Endosialin, TEM1	A_33_P3337485	2
Fibroblast	VIM	Vimentin	A_23_P161190	2
Fibroblast	COL1A1	Pro-collagen α1(I)	A_33_P3304668	3
Fibroblast	ITGA5	Integrin a5b1	A_23_P36562	3
Fibroblast	P4HA1	Prolyl 4-hydroxylase	A_33_P3214481	3
Fibroblast	P4HA2	Prolyl 4-hydroxylase	A_33_P3394933	3
Fibroblast	TGFB1	Transforming growth factor beta 1	A_24_P79054	3
Fibroblast	HSP47	Serpin family H member 1, SERPINH1	A_33_P3269203	-
Fibroblast	CD34	Hematopoietic progenitor cell antigen	A_23_P23829	-

Figure 4.4: Expression of key pluripotency markers in a 49-day HDF reprogramming time-course. Individual expression dynamics of Cluster 1 genes (early pluripotency markers) in (a) and Cluster 2 genes (late expressing pluripotency markers) in (b). Values are LOG2 transformed and normalised between 0 and 1 for 'day 0' and 'day 49', respectively, based on the average values between biological replicates for each time point. Dotted line marks CI. Gene label colours correspond to cluster colours in Fig. 4.2
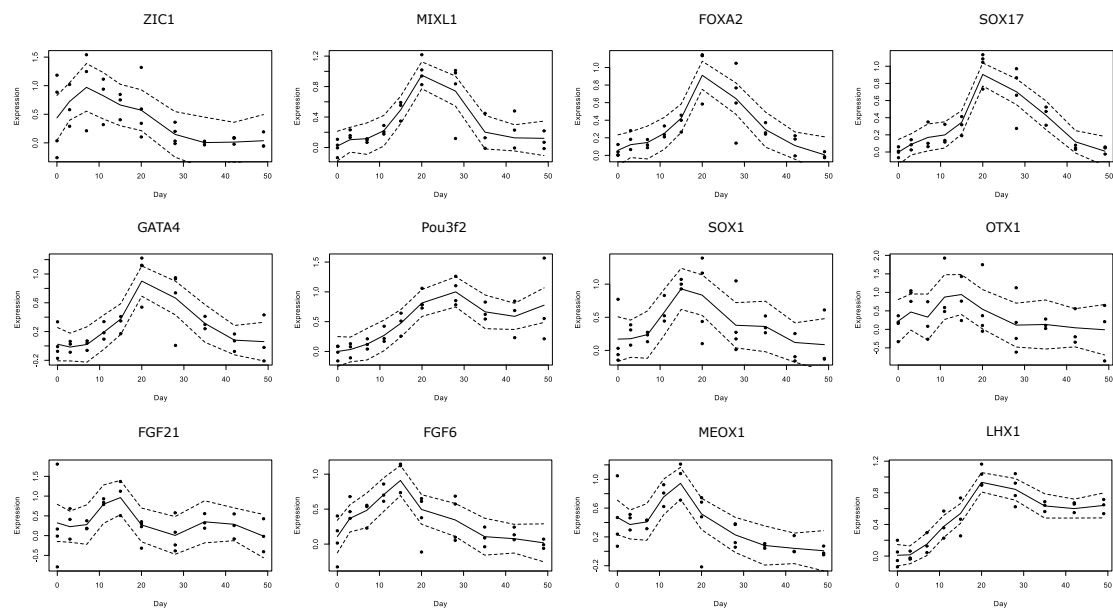
Figure 4.5: Expression trajectories of key developmental genes in a 49-day HDF reprogramming time-course. Values are LOG2 transformed and normalised between 0 and 1 for the 'minimum' and 'maximum' value respectively, based on the average values between biological replicates for each time point. Dotted line marks CI.

expression dynamics of early (cluster 1) and late (cluster 2) activated pluripotency genes (**Fig. 4.2**; Buganim et al. 2012; Tanabe et al. 2013; Chung et al. 2014; Takahashi and Yamanaka 2016). Pluripotency gene cluster 1 included *NANOG, SALL4, ZFP42, TRA-1-60, UTF1, DPPA4* and *LEFTY2*, and their expression increased dramatically within the first 10 days and then established stable pluripotency expression levels by day 20. In contrast, pluripotency gene cluster 2 (containing late expressing genes such as *LIN28, ZIC3* and *DNMT3B*) elevated expression slower and reached stable pluripotency levels by day 28 (Tanabe et al. 2013; Chung et al. 2014). Interestingly, eAge reset to zero at the same time that the genes in cluster 1 reached their pluripotent state levels, which temporally precedes full pluripotency. This also coincided with a peak in expression of a number of embryonic developmental genes between days 15 and 20, and might suggest that the reset marks a point where the cells reach an embryonic-like state but are not yet fully pluripotent (**Table. 4.1** and **Fig. 4.5**). In summary, eAge decline is observed well within the

first wave of pluripotency gene expression.

Therapeutic partial reprogramming will depend on rejuvenation with minimal dedifferentiation, which carries the risk of malignancies. We studied the dynamics of fibroblast gene down-regulation as a proxy for the loss of somatic cell identity. The individual trajectories of 19 commonly used fibroblast marker genes (MacFadyen et al. 2005; Kalluri and Zeisberg 2006; Goodpaster et al. 2008; Pilling et al. 2009; Chang, Li, and Guo 2014; Janmaat et al. 2015; Zhou et al. 2016 **Table. 4.1** and **Fig. 4.6**) clustered into three composite expression patterns, two of which (clusters 2 and 3) went into an immediate decline after OSKM induction (**Fig. 4.7**). However, one fibroblast-specific cluster (cluster 1) remained stable in its expression for the first 15 days. Interestingly, after day 7, fibroblast-specific gene expression in clusters 2 and 3 stopped declining and plateaued until day 15, coinciding with a peak in expression of senescence markers between days 11 and 15 (**Fig. 4.8**). Vimentin (VIM), for example, remained at 60% of maximal expression until day 15 of reprogramming, similarly to FAP, CD248 and COL1A2 in cluster 2 (**Fig. 4.6**). After day 15, fibroblast gene expression declined rapidly in all three clusters, and only by day 35 had all reached ESC expression levels, marking a complete loss of somatic identity (**Fig. 4.7**). Cluster 1, which contains the well described indicators of fibroblast identity FSP1, COL3A1 and TGFB2/3 (Kalluri and Zeisberg 2006), showed the slowest decline, and was also the last to reach ESC expression levels. In summary, we found that a number of fibroblast specific genes maintained high expression levels until day 15, by which time a substantial drop in eAge has been observed.

4.3   Discussion

Epigenetic rejuvenation or the reversal of cellular age, is a promising concept as it could avoid the oncogenic risks associated with dedifferentiation. Here, we
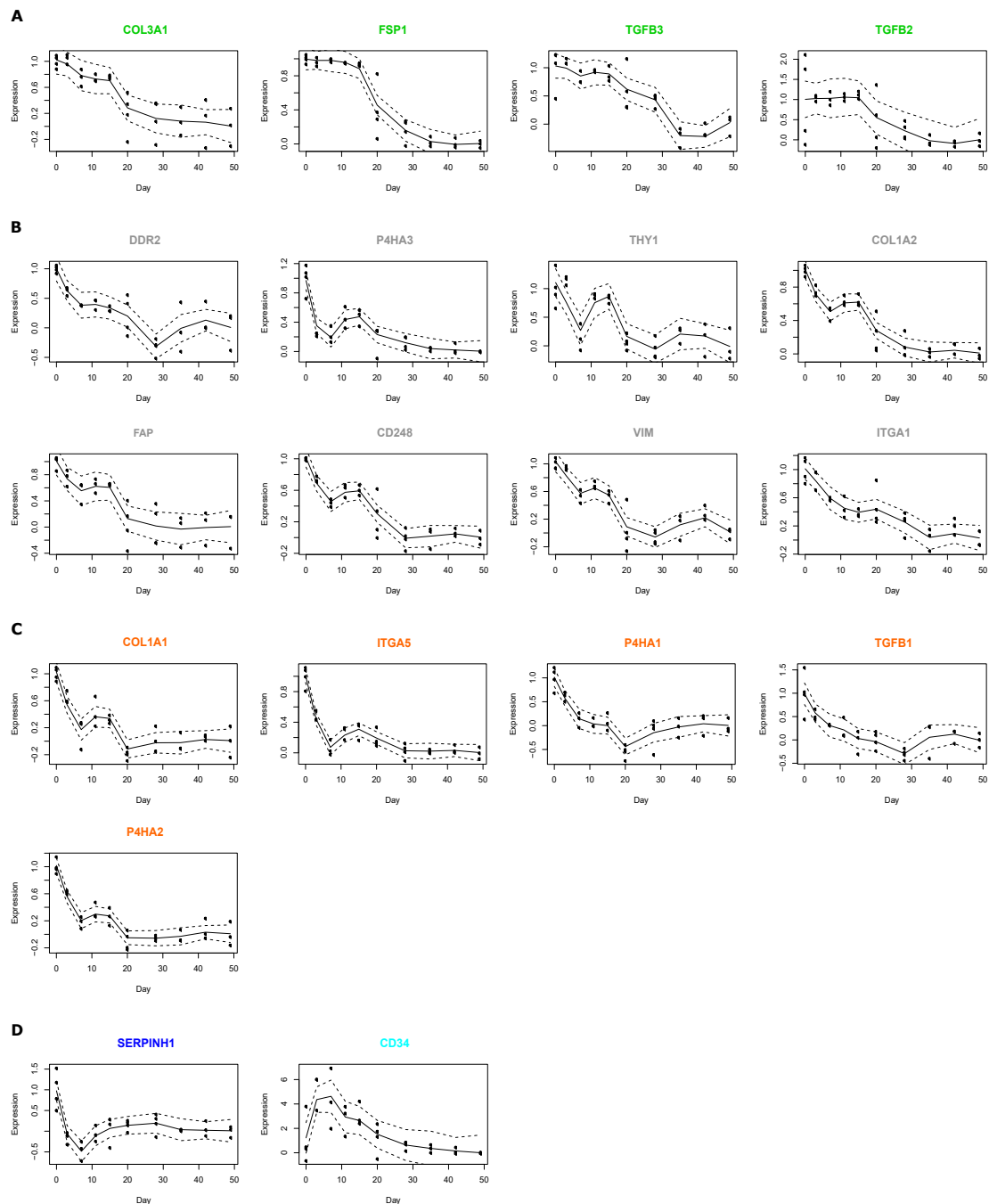
Figure 4.6: Expression of key fibroblast somatic markers in a 49-day HDF reprogramming time-course. Individual expression dynamics of Cluster 1 fibroblast genes (**A**), Cluster 2 (**B**) and Cluster 3 genes (**C**). Values are LOG2 transformed and normalised between 1 and 0 for 'day 0' and 'day 49', respectively, based on the average values between biological replicates for each time point. Dotted line marks CI. Gene label colours correspond to cluster colours in **Fig. 4.7**. SERPINH1 and CD34 expression could not be fit in any of the above clusters and are presented separately (**D**).

84
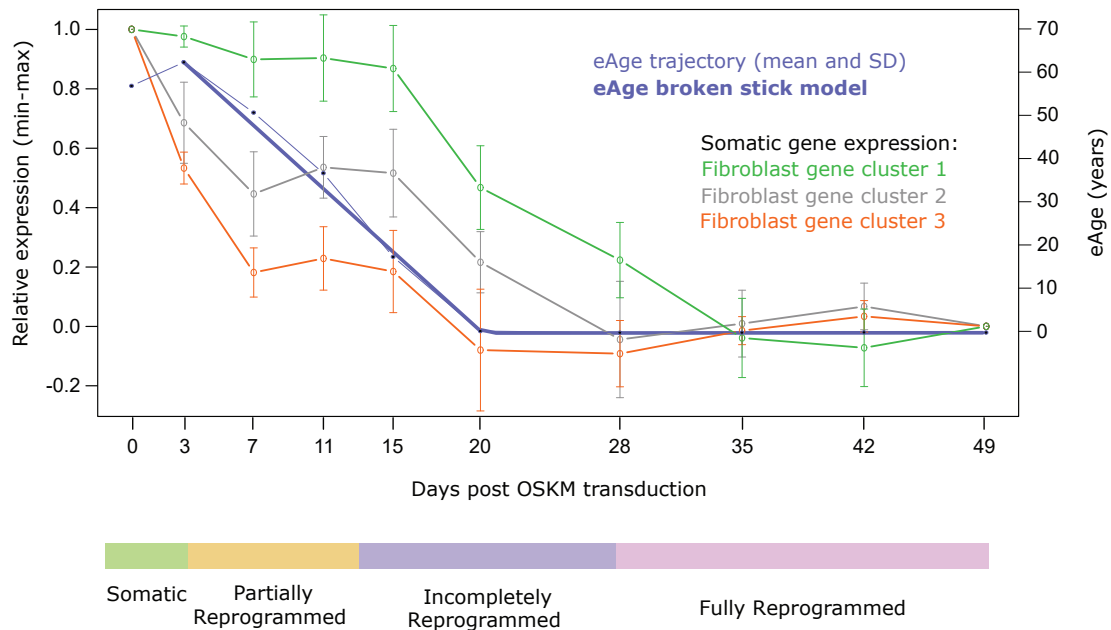
**Figure 4.7: Dynamics of eAge and fibroblast gene expression in a 49-day
HDF reprogramming time-course.** Left Y axis: composite gene expression
trajectories of key fibroblast markers statistically clustered as described for the
pluripotency markers in **Fig. 4.2**. Relative expression values were presented as
arbitrary units starting from '1' for 'day 0' to '0' for 'day 49'.Right Y axis: eAge as
in **Fig. 4.2** left Y axis, without SD.
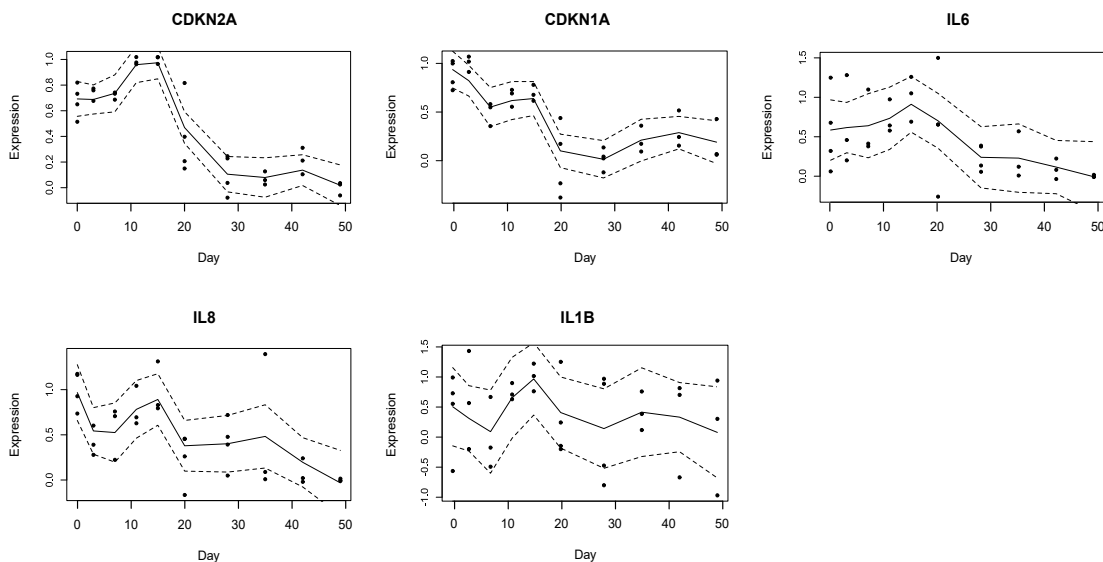
**Figure 4.8: Expression trajectories of key senescence markers in a 49-
day HDF reprogramming time-course.** Individual expression trajectories of 5
senescence associated genes. Values are LOG2 transformed and normalised
between 0 and 1 for the 'minimum' and 'maximum' value, respectively, based on
the average values between biological replicates for each time point. Dotted line
marks CI.

analysed a reprogramming time-course on HDFs and show that eAge declines in partially reprogrammed cells before their somatic identity is entirely lost.

It is well established that partial reprogramming happens within an early, reversible phase during the iPSC reprogramming time-course, which involves the stochastic activation of pluripotency genes. It is followed by a more deterministic maturation phase with predictable order of gene expression changes, where cell fate is firmly bound towards pluripotency (Takahashi and Yamanaka 2016; Smith, Sindhu, and Meissner 2016). Indeed, it has been shown that mouse fibroblasts fail to become iPSC and revert to their original somatic state if OSKM expression is discontinued during the initial stochastic phase (Brambrink et al. 2008; Stadtfeld et al. 2008). Previously, Tanabe et al. (2013) showed that TRA-1-60 (+) cells at reprogramming days 7 and 11 have not yet reached maturation and are partially reprogrammed, but our analysis already shows a decrease in their eAge according to multiple age predictors (**Fig. 4.2, Fig. 4.3** and **Fig. 4.7**). We have also shown that a large proportion of fibroblast marker genes maintain relatively high levels of expression until day 15 (**Fig. 4.7** and Supplementary Figure 5). Nearly unchanged levels of expression on day 15 were previously also shown for a large proportion of somatic genes (Tanabe et al. 2013). Together with increased senescence gene expression between days 11 and 15 (**Fig. 4.8**), this likely contributes to the high propensity of partially reprogrammed TRA-1-60 (+) cells to revert back to somatic phenotype before day 15 in the time-course (Tanabe et al. 2013). Interestingly, the step-wise decline of fibroblast gene expression coinciding with a peak in expression of senescence genes seems to delay the loss of somatic identity but not the expression of pluripotency genes. Taken together, the different dynamics between the step-wise fibroblast expression and the linear decline in eAge further indicate that dedifferentiation and epigenetic rejuvenation can be uncoupled.

One of the main caveats of this experiment is that, while day 7-11 cells have been previously characterised as 'partially reprogrammed' (Tanabe et al. 2013),

it is not clear exactly what proportion of cells are more highly or perhaps fully differentiated. NANOG expression reaches its maximum by day 10 (**Fig. 4.4**), as do other pluripotency markers, which does not eliminate the possibility of a subpopulation of cells being highly or fully dedifferentiated. Single-cell RNA sequencing would help determine which of the sorted TRA-1-60(+) cells are fully or partially differentiated. The only issue would be determining eAge of the same single cells. A preprint has been recently released which predicts epigenetic age at a single cell level (Trapp, Kerepesi, and Gladyshev 2021). This technique would need to be adapted to analyse RNA expression in the same cells (similar to scNMT; Clark et al. 2018) to accurately show both the pluripotency state, level of fibroblast expression (somatic identity) and eAge of a cell.

Our data suggest a window of opportunity within the uncommitted reprogramming phase, where a decline of eAge happens alongside partial maintenance of fibroblast gene expression. A deeper understanding of the kinetics of rejuvenation will be required to master therapeutic partial reprogramming, since any progress of dedifferentiation, even in a small subpopulation, carries the risk of malignancies. Our bulk expression analysis does not allow for a precise definition of the safe rejuvenation boundaries, and further experiments on a single cell level and in *in vivo* conditions are needed to determine a safe epigenetic rejuvenation window in different reprogramming systems. Upon defining safe boundaries, consideration should also be given to the steep decline of eAge, which resets to zero well ahead of the establishment of a pluripotent state, according to a number of age predictors (**Fig. 4.3**). Most likely this marks the point of reaching prenatal or embryonic stage, as suggested by the peak in expression of key developmental genes (**Fig. 4.5**).

The extent of epigenetic rejuvenation in years (human) or months (mouse), which can be achieved through partial reprogramming, also needs further attention and will most likely differ with the different reprogramming systems. The Horvath clock shows up to 10 years of rejuvenation in Ohnuki et al.'s system

by day 7 and another 10+ years by day 11. However, the intrinsic median estimation error of 3.6 years in this age predictor, the varying eAge rejuvenation values between the different age predictors, and the intra-replicate biological variation seen from the large error bars, highlight the need for more experiments and repetitions before this is established with a higher certainty.

Despite the obvious differences in reprogramming kinetics, our results also suggest that the improvements observed by Ocampo et al. (2016) in their OSKM-inducible secondary reprogramming system, might be due to epigenetic rejuvenation. It remains to be shown how stable in time the rejuvenated phenotype is in either of the systems. Further analysis is also needed regarding the effect of partial reprogramming on adult stem cells or premalignant cells, which have already shown a higher propensity of transforming to malignancy (Abad et al. 2013; Ohnuki et al. 2014). It is possible that a premalignant phenotype could be attenuated or amplified by partial reprogramming. In summary, our findings reveal exciting possibilities but also open a number of questions and highlight areas that need further attention.

Chapter 5

How Robust are Epigenetic Clocks to Global Methylation Changes?

5.1   Introduction

Ageing can be characterised as a progressive, physiological deterioration that leads to an increased susceptibility to a range of health outcomes, including cancer, cardiovascular, metabolic and neurodegenerative diseases (López-Otín et al. 2013; Brunet and Berger 2014; Benayoun, Pollina, and Brunet 2015). Manipulation of nutrient-sensing pathways were identified as regulators of ageing.   These include mechanistic target of rapamycin (mTOR) and insulin/insulin-like growth factor (IGF) signalling, which have been related to lifespan extension in mammals such as mice (Weindruch et al. 1986; Blüher, Kahn, and Kahn 2003; Holzenberger et al. 2003; Pearson et al. 2008; Johnson, Rabinovitch, and Kaeberlein 2013). Cellular reprogramming experiments have shown epigenetic rejuvenation in mice and humans might also be possible (Manukyan and Singh 2014; Ocampo et al. 2016; Olova et al. 2018a; Sarkar et al. 2020; Lu et al. 2020). Healthspan extending research has been impeded by the lack of tools to quantify rejuvenation, which requires a biomarker that confers biological ageing more effectively than chronological age (Baker and

Sprott 1988).

DNA methylation (DNAm) is an epigenetic mark maintained by three types of DNA methyltransferase (DNMT): DNMT1, which maintains DNA methylation patterns during DNA replication (Gruenbaum, Cedar, and Razin 1982; Bestor and Ingram 1983; Li, Bestor, and Jaenisch 1992; Vilkaitis et al. 2005); and DNMT3A/B, which *de novo* methylates both unmethylated and hemi-methylated DNA, and help initialise the epigenome after fertilisation (Okano et al. 1999; Jackson et al. 2004; Brunet and Berger 2014). Global demethylation of the epigenome occurs with increased age (Wilson et al. 1987; Fuke et al. 2004; Bjornsson et al. 2008). Changes in expression of DNMT3B and DNMT1 with age have been reported, which could have causal changes to DNAm patterns and affect eAge prediction (Ciccarone et al. 2016). DNMT3A/B also appear important for somatic functions, most likely by conducting *de novo* methylation of sites missed by DNMT1. For example, loss of function DNMT3A can lead to age-related clonal haematopoesis (ARCH), which in turn increases risk of haematological cancers (Ley et al. 2010; Kim et al. 2013; Yasuda et al. 2014; Herold et al. 2017; Jeong et al. 2018; Abelson and Wang 2018; Robertson et al. 2019; Kurosawa and Iwama 2020). Indeed, global hypomethylation (Feinberg and Vogelstein 1983; Goelz et al. 1985) and specific hypermethylation of promoter CpG islands (Esteller et al. 2001; Esteller 2002; Herman and Baylin 2003) are hallmarks of cancer cells.

Over the last decade, various age predictors know as epigenetic clocks have been developed that utilise DNA methylation. Most epigenetic clocks are built using a penalised regression algorithm that selects and weights key CpGs whose hyper- and hypomethylation correlate with age in a linear model across the adult lifespan, resulting in a predictor of age (Horvath 2013; Horvath and Raj 2018). DNAm outperforms all other biomarkers as a predictor of biological age, including leukocyte telomere length (LTL) and age associated gene expression (Benayoun, Pollina, and Brunet 2015; Marioni et al. 2016; Jylhävä, Pedersen,

and Hägg 2017). The average difference between eAge and chronological age
(age acceleration) is associated with a number of age-related conditions and
diseases, eg. Down's syndrome and obesity; Horvath et al. 2014; Horvath et al.
2015a; Marioni et al. 2015b; Levine et al. 2015; Horvath and Levine 2015;
Horvath et al. 2016b; Breitling et al. 2016; Maierhofer et al. 2017; Quach et al.
2017; Horvath et al. 2018; Martin-Herranz et al. 2019. However, it remains
unclear how robust epigenetic clocks are to global changes in methylation.
There are a number of ways that changes detected in age acceleration could be
due to false positives. These biases can be grouped in either one of two broad
categories; biological and technical biases.

Biological biases could involve germline or somatic mutations related to DNA
methylation machinery. For example, we have previously found that ARCH
patients with somatic mutations in DNMT3A and ten-eleven translocation (TET)
2 have an increased epigenetic age (Robertson et al. 2019). However, it is
plausible that mutations in these genes cause global methylation changes that
result in inaccurate age prediction, leading to a false eAge acceleration. It is
therefore also possible that similar mutations might be present in any number of
experiments analysing eAge, and could contribute to a false positive result, or
even a false negative. A point mutation in the PWWP region of DNMT3A causes
aberrant hypermethylation at bivalent chromatin in mice (Sendžikaitė et al.
2019), and could conceivably lead to global hypermethylation. Similar somatic
mutations could manifest not only in DNMTs and TETs, but any number of genes
that interact with them. For example, DNMT3L does not directly methylate CpG
sites, but facilitates DNMT3A/B mediated methylation (Hata et al. 2002; Mao
et al. 2020). Technical biases may arise from certain methodological steps such
as the DNA extraction method (which depending on the method, can affect
methylation capture accuracy by 1-3%; Soriano-Tárraga et al. 2013) or the
efficiency of bisulphite (BS) conversion.   For example, BS-induced DNA
degradation can cause a depletion of unmethylated cytosines, causing an up to

10% overestimation of global methylation (Olova et al. 2018b).

In this study, we analysed how susceptible epigenetic clocks are to global changes in methylation by exploring a variety of scenarios using DNAm data in participants from the Generation Scotland (GS) study (Smith et al. 2006; Smith et al. 2013). We show that epigenetic clocks are not impervious to gradual, global changes in methylation. Discrete alterations in methylation state, caused by potential biological or technical variables, can have a dramatic impact on eAge prediction.

5.2 Results

DNA methylation from 5,100 ostensibly healthy participants in the GS study were gradually hyper- and hypomethylated (between plus and minus 20%), and the Horvath, Hannum and Skin & Blood (S&B) clocks were applied. Relative methylation change was calculated by multiplying each percentage methylation (β) value by a factor (eg. 0.9 or 1.1 to achieve minus or plus 10% respectively, **Fig. 5.1A-F**). Absolute methylation change was calculated by adding or subtracting a percentage value; eg. +/- 0.1 to get plus or minus 10% absolute methylation (**Fig. 5.1G-L**). Hypermethylation overall increases eAge, whereas hypomethylation decreases eAge. There is an inflection point (for example, 20 years old in the Horvath clock) where this relationship tends to reverse, so individuals less than 20 years old are more likely to have a higher eAge with greater hypomethylation (**Fig. 5.1A,D**). This inversion is likely because the Horvath clock applies a logarithmic transformation to ages <20 years (Horvath 2013; Snir, Farrell, and Pellegrini 2019).The mean age acceleration tends to decrease as age increases, regardless of methylation status (**Fig. 5.1D-F,J-L**). The difference in eAge between hyper/hypomethyated samples and unchanged samples appears to increase with age in the relative samples, but not in the absolute samples (**Fig. 5.1D-F,J-L**). When certain CpG sites reach a saturation

point (100% hyper- or hypomethylation), the linear model starts to decay. This
can be seen in the Hannum clock (where positive and negative methylation lines
do not increase or decrease in a regular way compared to the other clocks; **Fig.
5.1B,E**). Absolute changes in methylation also exhibit a similar breakdown of the
linear trend (**Fig. 5.1G-L**), as absolute increases or decreases in methylation are
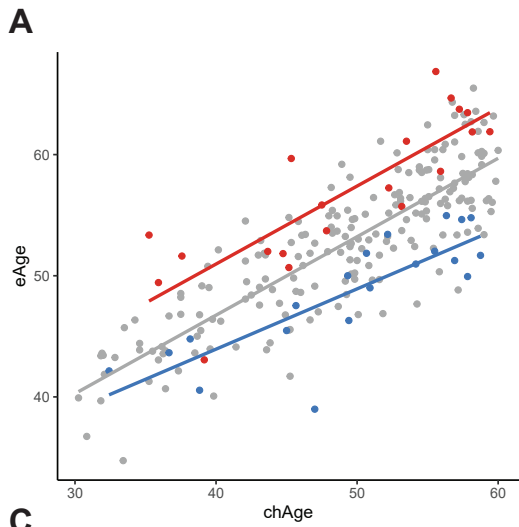more likely to saturate or desaturate/floor certain CpGs.
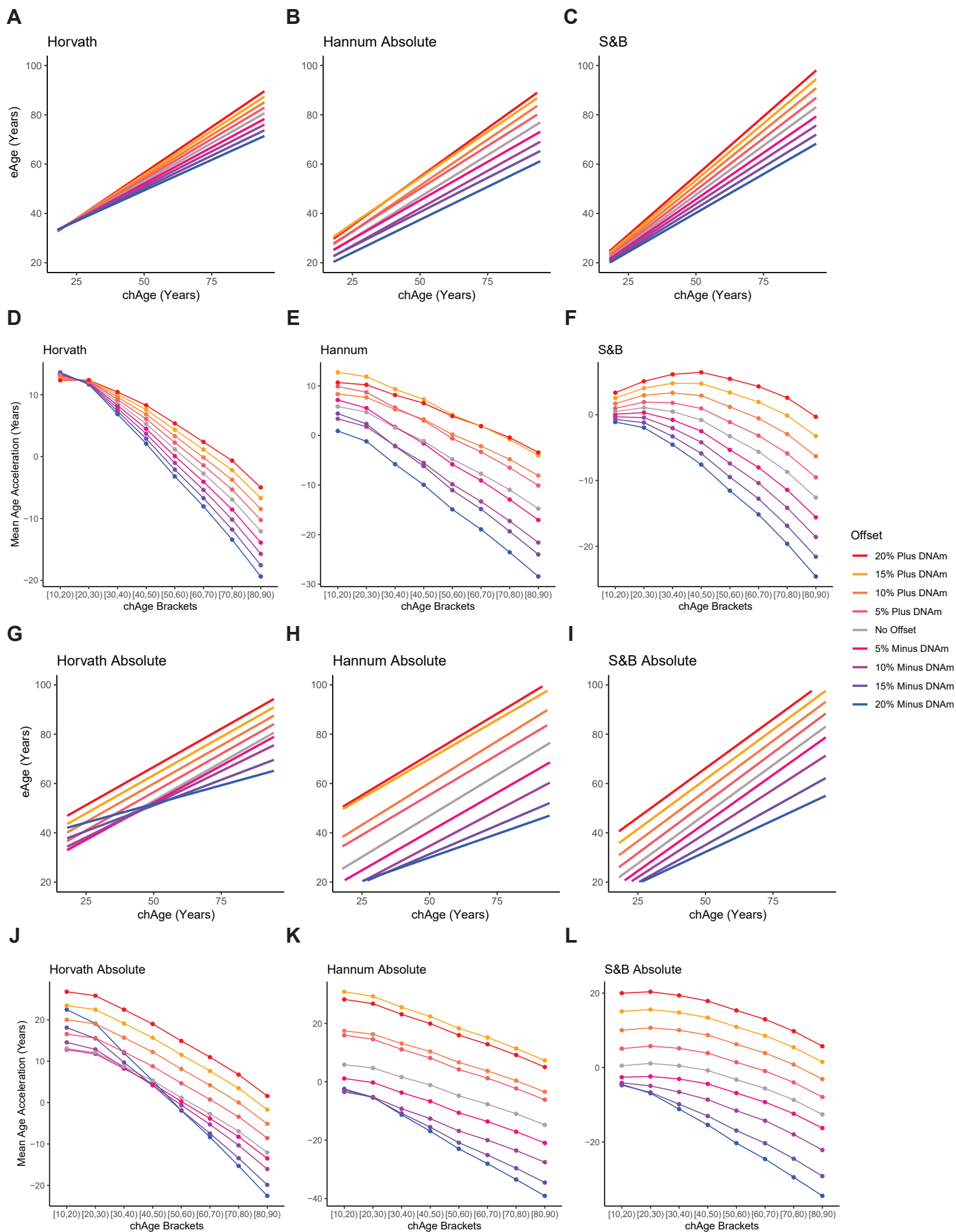
**Figure 5.1: Epigenetic clocks are not resistant to global changes in
methylation.** Statistical comparisons between test and control groups were
calculated with the Welch's T-Test. (**A-C**) Relative methylation changes from
+20% to -20% for Horvath (**A**), Hannum (**B**) and S&B (**C**) clocks, epigenetic age
(eAge) plotted on y-axis against chronological age (chAge) on x-axis. (**D-F**) Mean
age acceleration of each age group in GS1 for Horvath (**D**), Hannum (**E**) and S&B
(**F**) clocks, for relative methylation changes. (**G-I**) Absolute changes in global
methylation from +20% to -20% for Horvath (**G**), Hannum (**H**) and S&B (**I**) clocks.
(**J-L**) Mean age acceleration of each age group in GS1 for Horvath (**J**), Hannum
(**K**) and S&B (**L**) clocks, for absolute methylation changes.

A common caveat to consider in eAge studies is the use of small sample
sizes for rare diseases (eg. under 30; Walker et al. 2015; Horvath et al. 2016b;
Maierhofer et al. 2017; Horvath et al. 2018). Two scenarios were created using
random subsets of GS participants aged 30-60 to depict how global methylation
changes could result in false positive differences in age acceleration. In
Scenario 1, 200 samples were unaltered, representing the control group. Test 1
and Test 2 had 20 samples each, and could represent two random test
populations with a rare genetic disorder. These test groups were globally hyper-
or hypomethylated, which represents any possible biological or technical bias
that might occur (eg. DNA methylation machinery mutations, or errors in BS
conversion). In Test 1, global methylation was increased by 20% and Test 2 was
decreased by 20% (**Fig. 5.2A**). No significant difference in eAge was present
between Test 1 unaltered and control, but there was a nominally significant
difference ($p \leq 0.05$) between Test 2 unaltered and controls (**Fig. 5.2B**). Relative
increases and decreases of 20% caused significant increases and decreases (P
≤ 0.0001) in age acceleration, respectively, compared to the control group (**Fig.
5.2A,B**). An absolute increase of 20% also caused a significant increase (P $\leq$
0.0001) from the control (**Fig. 5.2C,D**). A significant difference (P ≤ 0.001) in
eAge from control was also observed in Test 2 after an absolute decrease of
20% methylation (**Fig. 5.2D**). In both test groups, linear regression analyses
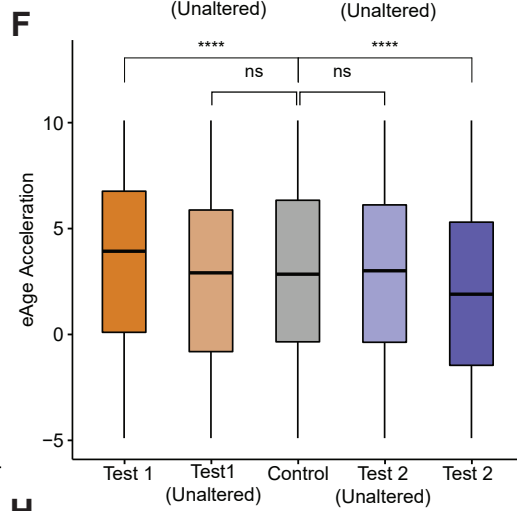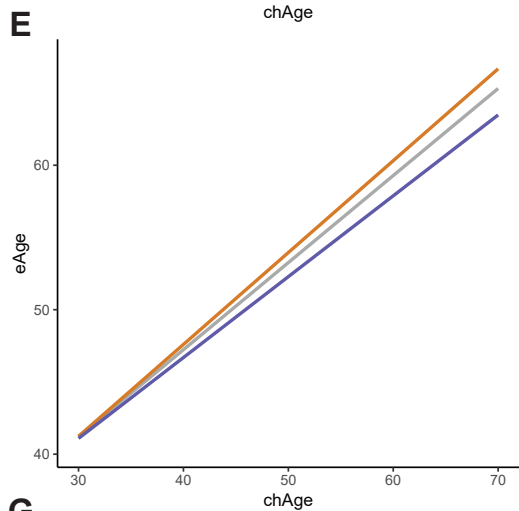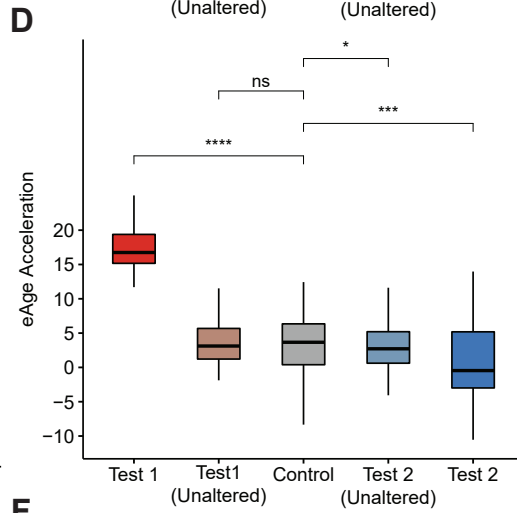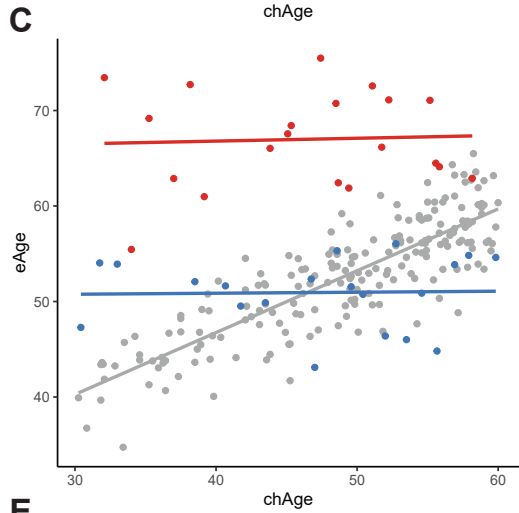show no association between chAge and eAge, due to saturation or

desaturation/flooring of clock CpG sites (**Fig. 5.2C**).

Alterations in the epigenome as large as 20% are extreme and perhaps unlikely. We therefore specified a second scenario to demonstrate how a small change in global methylation can create a significant difference in age acceleration. 2000 control samples and 500 samples for each test group, aged 30-70, were randomly selected from GS (**Fig. 5.2E-H**). No significant difference was present between the unaltered test groups and the control group (**Fig. 5.2F,H**). A relative 5% methylation increase caused a 0.73 years difference in eAge between Test 1 and control, which was statistically significant ($P \leq 0.001$) (**Fig. 5.2E,F**). Similarly, a relative 5% decrease in methylation caused a 0.95 year difference between Test 2 and control which was also statistically significant ($P \leq 0.001$; **Fig. 2F**). Absolute increases and decreases in methylation also caused significant increases and decreases in eAge (respectively) from the control group (**Fig. 5.2G,H**). The absolute 5% methylation increase caused Test 1 eAge acceleration to be 3.27 years higher than the control (**Fig. 5.2H**). An absolute 5% methylation decrease caused Test 2 to decrease by 1.09 years from control, similar to the relative version.
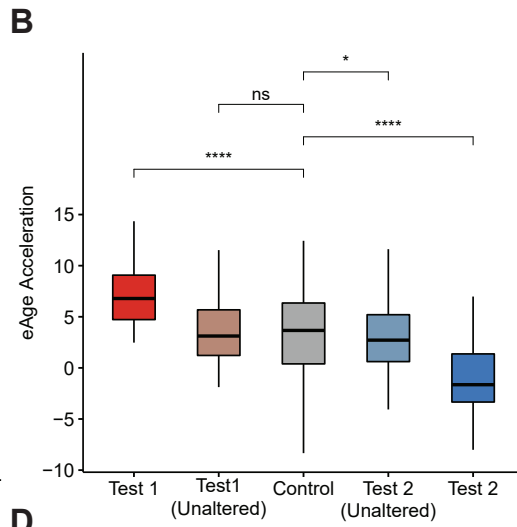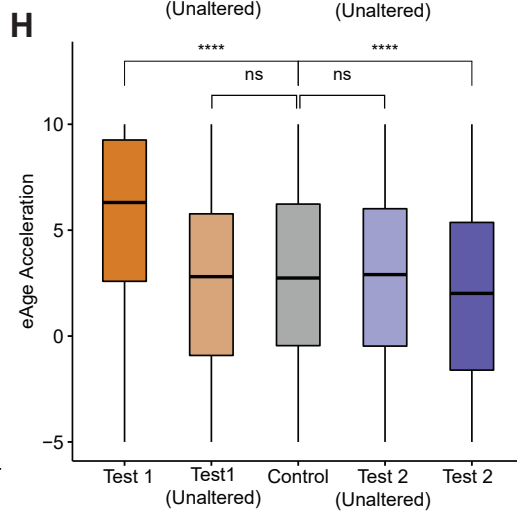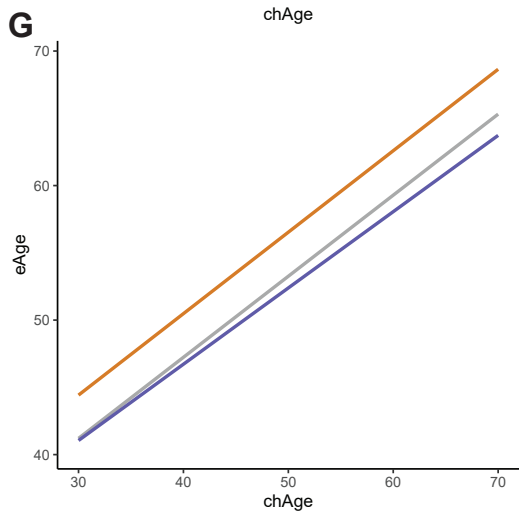
Figure 5.2: Global methylation changes can create false positive results. (**A,B**) First experiment scenario with relative methylation changes; a control group (n=200) and two test groups, Test 1 (plus 20%, n=20) and Test 2 (minus 20%, n=20). (**A**) The Horvath clock was applied to all three groups, Horvath eAge on y-axis against chAge on x-axis. (**B**) Mean age acceleration for relative changes in Scenario 1. The control, Test 1 unaltered and Test 2 unaltered had mean age accelerations of 3.54 +/- 0.30, 4.2 +/- 1.05, and 2.66 +/- 1.02 years respectively. Significant differences (P $\leq$ 0.001) in age acceleration (7.72 +/- 0.99 years Test 1 and -0.86 +/- 1.08 years Test 2) resulted between the two test groups and control. (**C,D**) Absolute methylation change version of scenario 1. The Horvath clock was applied to all three groups (**C**), Horvath eAge on y-axis against chAge on x-axis. (**D**) Mean age acceleration for relative changes in scenario 1. Significant difference (P $\leq$ 0.001) in age acceleration (17.89 +/- 1.04 years Test 1 and 1.38 +/- 1.53 years Test 2) resulted between Test 1 and control, but not between Test 2 and control (**D**). (**E,F**) Second experiment scenario; a control group (n=2000) and two test groups, Test 1 (plus 20%, n=500) and Test 2 (minus 20%, n=500). (**E**) The Horvath clock was applied to all three groups, Horvath eAge on y-axis against chAge on x-axis. (**F**) Mean age acceleration for relative changes in Scenario 2. The control, Test 1 unaltered and Test 2 unaltered had mean age accelerations of 2.81 +/- 0.11, 2.77 +/- 0.52, and 2.82 +/- 0.22 years respectively. Significant differences (P $\leq$ 0.001) in age acceleration (3.54 +/- 0.22 years Test 1 and 1.86 +/- 0.22 years -5% Test 2) resulted between the two test groups and control. (**G,H**) Absolute version of second experiment scenario. The Horvath clock was applied to all three groups (**G**), Horvath eAge on y-axis against chAge on x-axis. Significant differences (P $\leq$ 0.001) in age acceleration (2.81 +/- 0.11 years control, 6.08 +/- 0.22 years Test 1 and 1.72 +/- 0.23 years -5% Test 2) resulted between the two test groups and control (**H**).

5.3 Discussion

Epigenetic clocks have emerged as tools that provide a proxy measure of biological age. However, the precise aspects of physiological ageing they measure is unclear, hence it is still unclear if epigenetic clocks truly capture biological age. Often, the age acceleration of a test dataset is compared with larger control datasets. When testing rare genetic disorders, the number of samples are typically low (eg. under 30; Walker et al. 2015; Horvath et al. 2016b; Maierhofer et al. 2017; Horvath et al. 2018). A significant difference in age acceleration between cases and control may indicate that the condition causes a change in biological age.

We have shown that the epigenetic clock is not robust against global changes. Global hyper- and hypomethylation can increase and decrease eAge, respectively. In addition, these differences appears to be more pronounced with age. We have demonstrated two scenarios that create false positive results due to a large (+/-20%) or a small (+/-5%) methylation change. Biological and technical biases could globally increase or decrease methylation, which would significantly offset eAge prediction and result in false positive results. Somatic variants in DNA methylation machinery (such as DNMT3A) could cause these global changes in methylation which result in accelerated eAge (Robertson et al. 2019). Indeed, mutations in NSD1 (Soto syndrome) have been hypothesised to alter DNAm patterns and affect DNMT3A activity, which therefore increases eAge (Martin-Herranz et al. 2019).

Technical biases can cause a global change in methylation that could alter eAge. Difference in BS conversion between experiments results in discrepancies of methylation readouts. If the conversion rate is too low (due to BS-induced degradation of unmethylated cytosines), then the methylation signal could be overestimated by up to 10% (Olova et al. 2018b). This effect might be a confounding factor when comparing BS experiments between different labs, resulting in batch effects (eg. studies where case samples are assessed by one lab, and control samples are obtained online from a variety of sources Horvath and Levine 2015; Martin-Herranz et al. 2019). Martin-Herranz et al. showed that various DNAm array datasets cluster based on batch, capturing a degree of technical variance that might be due to differences in BS conversion (Martin-Herranz et al. 2019). Sub-optimal BS conversion might also cause false negative results. For example, 160 of the 353 CpGs hypomethylate with age in the Pan-Tissue clock (Horvath 2013). Therefore, it is possible that a low conversion rate would cause these sites to appear hypermethyated and alter the predicted age to hide any resulting increase or decrease with age. Processing test and control samples in the same lab/batch will ensure the same protocol for

conversion and minimise technical biases. Alternatively, when using control and test samples from different labs, batch effect correction and and composite cell correction may minimise technical variance (Martin-Herranz et al. 2019).

Increases and decreases in methylation can saturate or desaturate certain CpGs, at which point the linear model starts to degrade. This might explain why epigenetic clocks do not function properly with certain conditions. For example, global hypomethylation (Feinberg and Vogelstein 1983; Goelz et al. 1985) and hypermethylation of certain promoter CpG islands (Esteller et al. 2001; Esteller 2002; Herman and Baylin 2003) are hallmarks of cancer cells. The epigenetic clock does not show accelerated ageing in cancer cells (Horvath 2015), which might be due to global methylation deregulation occurring in a similar manner to Figure 5.1.

Consistent with other reports for the Horvath and Hannum clocks, we found a lower average epigenetic age in older individuals (**Fig. 5.1D-F,J-L**; El Khoury et al. 2019; Marioni et al. 2019; Martin-Herranz et al. 2019). While a lower predicted age in older individuals could be due to survivor bias (ie. individuals who survive to old age may be likely to be healthier), we see in our data that both global hyper- and hypomethylation decreases age acceleration, but the effect is exacerbated in the hypomethylated samples. It is possible that this decrease in eAge acceleration with increasing age is due to saturation of certain CpGs, which in turn would reduce the "tick rate" of the epigenetic clock (see Chapter 1.4.5). This has been previously hypothesised by El Khoury et al. (2019). They note that out of the ten most influential CpGs of the Horvath clock, three of them might be candidates for CpG saturation since they are normally highly methylated and expected to increase in methylation with age (El Khoury et al. 2019). Therefore, certain CpGs may only work until a certain age before they saturate or desaturate, at which point the linear model degrades and prediction becomes less accurate, as shown in **Figure 5.1**.

While this has been an *in silico* investigation, it would also be interesting to

corroborate our results *in vitro*. Heterozygous knock-out, or knock-down, of DNMT3A/B in human fibroblasts could cause a gradual decrease in global methylation. Alternatively, an extra copy of DNMT3A/B could be inserted, or transiently expressed using a viral vector, to cause a global increase in methylation. Alternatively, chemical treatment using 5-aza-2'-deoxycytidine (a DNMT inhibitor) can induce global methylation (Jones and Taylor 1980; Chuang et al. 2005; Ramos et al. 2015). Knock-down, transient expression or chemical treatment would be the most favourable, as global methylation percentages and eAge can be calculated pre- and post-treatment, and therefore the exact amount of global methylation change can measured for each technique.

Since eAge does not correlate with most physiological signs of ageing, such as cell division, replicative senescence and telomere length (see Chapter 1.4.6 for more detail; Lowe, Horvath, and Raj 2016; Marioni et al. 2016; Kabacik et al. 2018; Horvath et al. 2019; Cypris et al. 2020), it might instead measure aspects of age-related epigenetic drift/deregulation (Yu et al. 2020)? Demethylation can occur in either a passive manner (eg. through inhibition of DNMT1 during cell replication; Wolffe, Jones, and Wade 1999; Mayer et al. 2000), or actively via methyl-CpG binding domain protein 4 (MBD4; Hendrich et al. 1999) or TET enzymes (Jin et al. 2014; Ichiyama et al. 2015). However, both theories are confounded regarding measurable changes in eAge as age increases. There is little evidence to suggest that active processes, such as TET, directly demethylate with age and affect eAge prediction (Wallace 2014; Zhang et al. 2016; Yu et al. 2020). eAge can be measured in nonproliferating tissues (Horvath 2013; Yang et al. 2016; Horvath et al. 2019) which means passive demethyation is less likely. It is possible that actively dividing tissues accumulate somatic mutations in DNA methylation machinery during ageing, which result in epigenetic drift that can be seen as aberrant eAge prediction.

Our study shows that the linear models used to build epigenetic clocks do not compensate for global changes in methylation. Can the linear model be adapted

or changed to improve age prediction? The Horvath clock applies a logarithmic
transformation to ages <20 years, and the linear model remains unaltered for
ages >20 years (Horvath 2013; Snir, Farrell, and Pellegrini 2019). Perhaps
further transformation in old age is required for accurate age prediction? Or
specifically building a clock that excludes highly saturated or desaturated CpGs
will result in a more stable predictor? Alternatively, a compensatory approach
might be required for older age groups to account for saturating/desaturating
CpGs (eg. imputing optimal values below 0 or above 1). It is possible that
biological age is intrinsically non-linear in its progression. Therefore, alternate
models may prove more appropriate to predict age and account for
saturating/desaturating CpGs, such as support vector regression (Xu et al. 2015;
Aliferi et al. 2018), or logarithmic/quadratic models (Snir, VonHoldt, and
Pellegrini 2016; Snir, Farrell, and Pellegrini 2019). Similar approaches may
prove useful in providing accurate age prediction that are immune to global
methylation changes.

Chapter 6

Robust Age Prediction in Mouse Based on Average Methylation per Genomic Region

6.1 Introduction

A number of age predictors have been created based on Illumina human DNAm arrays, which offer precise and reproducible readouts of thousands of CpGs, resulting in reliable age prediction (Bocklandt et al. 2011; Koch and Wagner 2011; Hannum et al. 2013; Horvath 2013; Weidner et al. 2014; Levine et al. 2018; Lu et al. 2019a; Zhang et al. 2019). The predicted age generated by these clocks is referred to as epigenetic age (eAge), and the age acceleration (difference between eAge and chronological age, chAge) is associated with a number of disease states, conditions and all-cause mortality (Horvath et al. 2014; Horvath 2015; Horvath and Levine 2015; Horvath et al. 2016b; Marioni et al. 2015a; Chen et al. 2016; Simpkin et al. 2016; Maierhofer et al. 2017; Horvath et al. 2018; Lu et al. 2019a; Martin-Herranz et al. 2019; Wu et al. 2019; Higgins-Chen et al. 2020). As a result, eAge has become a popular proxy to estimate biological age, however, whether it truly predicts biological age is still

debated (Field et al. 2018).

Epigenetic clocks have also been created for mice, which has paved the way for quantifying the rejuvenation potential of life-span extending drugs and therapies (Wang et al. 2017; Petkovich et al. 2017; Stubbs et al. 2017; Meer et al. 2018; Thompson et al. 2018; Han et al. 2018; Han et al. 2020). Many of these clocks use reduced representation bisulphite sequencing (RRBS) data, where bisuphite-converted DNA is digested by MSP1, an enzyme that cuts DNA at CpG sites, regardless of their methylation state (Meissner et al. 2005; Gu et al. 2011; Baheti et al. 2016). This process ensures fragments are enriched with CpG sites, and is a cheaper alternative to whole genome bisuphite sequencing (WGBS) since it captures only ~1% of the genome (Meissner et al. 2008a; Gu et al. 2011). However, the overlap of CpG sites captured in RRBS and their coverage, differs between datasets due to differences in protocols and enzymes used (Field et al. 2018; Thompson et al. 2018). In addition, many of the bisuphite-sequencing (BS-seq) based clocks were created by pooling multiple datasets together and dividing into test and training data (i.e. no out-of-sample testing is conducted). This means RRBS mouse clocks have limited transferability (i.e. lower accuracy) when applied to studies outside of the original training data of a given clock.

We have developed RRBS mouse clocks that use average methylation over large regions (termed regional epigenetic clocks), rather than individual CpGs. We have demonstrated that a window size between 2 and 6 Kb will generate a robust epigenetic clock that is transferable across multiple RRBS platforms and experiments, and outperforms current RRBS mouse clocks, such as the Stubbs multi-tissue and Petkovich blood clocks.

Table 6.1: Published bisuphite sequencing mouse clocks and the datasets used to train them. Datasets highlighted yellow were used to construct the multi-tissue region clocks (Rizel et al. 2018 was also used but not listed), and the datasets highlighted red were used to construct the blood region clock.

Clock	Reference	Cell Types/Tissue Used for Training	Samples Used for Training	GEO	Reference
Petkovich	Petkovich et al. (2017)	Whole blood	141	GSE80672	Petkovich et al. (2017)
Wang	Wang et al. (2017)	Liver	99	GSE60012	Reizel et al. (2015)
		Liver	4	SRA344045 (WGBS)	Gravina et al. (2016)
		Liver	4	GSE89275 (WGBS)	Wang et al. (2017)
Stubbs Multi-Tissue	Stubbs et al. (2017)	Liver	7	GSE52268	Cannon et al. (2014)
		Liver	50	GSE60012; GSE93957	Reizel et al. (2015); Stubbs et al. (2017)
		Lung	12	GSE93957	Stubbs et al. (2017)
		Heart	12	GSE93957	Stubbs et al. (2017)
		Muscle	18	GSE60012	Reizel et al. (2015)
		Spleen	10	GSE60012	Reizel et al. (2015)
		Cerebellum	8	GSE60012	Reizel et al. (2015)
		Cortex	12	GSE93957	Stubbs et al. (2017)
Meer	Meer et al. (2018)	Whole Blood	125	GSE80672	Petkovich et al. (2017)
		Liver	38	GSE60012	Reizel et al. (2015)
		Muscle	20	GSE60012	Reizel et al. (2015)
		Spleen	8	GSE60012	Reizel et al. (2015)
		Cerebellum	6	GSE60012	Reizel et al. (2015)
		Pro B cells	3	GSE70538	Orlanski et al. (2016)
		Follicular B cells	2	GSE70538	Orlanski et al. (2016)
		Liver	10	GSE84573	McCormick et al. (2017)
		Cortex	5	GSE84573	McCormick et al. (2017)
		Heart	4	GSE84573	McCormick et al. (2017)
		Cortex	13	GSE93957	Stubbs et al. (2017)
		Heart	11	GSE93957	Stubbs et al. (2017)
		Liver	12	GSE93957	Stubbs et al. (2017)
		Lung	14	GSE93957	Stubbs et al. (2017)
		Cortex	15	GSE121141	Meer et al. (2018)
		Heart	18	GSE121141	Meer et al. (2018)
		Liver	16	GSE121141	Meer et al. (2018)
		Lung	12	GSE121141	Meer et al. (2018)

6.2 Results

6.2.1 Analysing Published Mouse RRBS Clocks

We first analysed the coverage of the CpGs of two multi-tissue mouse RRBS clocks (Stubbs and Meer, **Table 6.1**) in various mouse RRBS datasets (**Fig. 6.1A**). Coverage of clock sites vary from dataset to dataset. For example, some sites are consistently higher in coverage than others for all datasets, however, some might be high in one dataset and low in others. Meer et al. (2018) consistently had the highest coverage on all clocks, which is expected since Meer et al. trained their clock using CpGs that were present in most of the datasets they analysed.

We applied five published BS-seq mouse clocks to various publicly available datasets (**Table 6.2, Fig. 6.1B-F**). Clocks applied to their own training data, (eg. Stubbs and Meer applied to Stubbs et al. 2017 data, **Fig. 6.1B,C**) had adj. R^2 values over 0.6 (**Table 6.2**). The main exceptions were the Stubbs and Wang clocks applied to the Reizel et al. (2018) dataset (adj. R^2 = 0.90 and 0.69 respectively), which is likely because this dataset was processed by the same lab of the Reizel et al. (2015) data which was used to train these clocks. However, clocks applied to datasets outside of their training datasets had adj. R^2 values lower than 0.5 (**Table 6.2**). The Stubbs clock was trained on 0-9 mo samples, however, it performs less accurately on 9 mo samples in the Stubbs et al. (2017) data (**Fig. 6.1B**) and 10 mo samples in Thompson et al. (2018) data (**Fig. 6.1D**). The Meer clock performed well on the Thompson et al. (2018) data, and was particularly accurate on blood (adj. R^2=0.89, MAE=1.9, **Fig. 6.1E**). The Petkovitch blood clock had a low age acceleration and accuracy (R^2=0.41, MAE=5.69, **Fig. 6.1F**).
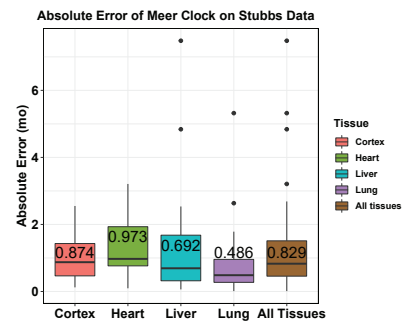
A Coverage on muti-tissue datasets

Stubbs Clock

Meer Clock

Chromosomes: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

B

Stubbs Clock Applied to Stubbs Data

$R^2_{adj} = 0.85$ $P <= 0.001$
$R^2_{adj} = 0.78$ $P <= 0.001$
$R^2_{adj} = 0.87$ $P <= 0.001$
$R^2_{adj} = 0.68$ $P <= 0.001$
$R^2_{adj} = 0.78$ $P <= 0.001$

Absolute Error of Stubbs Clock on Stubbs Data

Cortex 1.050 Heart 0.817 Liver 1.025 Lung 0.372 All Tissues 0.712

Tissue: Cortex, Heart, Liver, Lung, All Tissues

C

Meer Clock Applied to Stubbs Data

$R^2_{adj} = 0.86$ $P <= 0.001$
$R^2_{adj} = 0.81$ $P <= 0.001$
$R^2_{adj} = 0.5$ $P = 0.00202$
$R^2_{adj} = 0.77$ $P <= 0.001$
$R^2_{adj} = 0.45$ $P <= 0.001$

Absolute Error of Meer Clock on Stubbs Data

Cortex 0.874 Heart 0.973 Liver 0.692 Lung 0.486 All Tissues 0.829

Tissue: Cortex, Heart, Liver, Lung, All tissues

D

Stubbs Clock Applied to Thompson Data

$R^2_{adj} = 0.71$ $P <= 0.001$
$R^2_{adj} = 0.73$ $P <= 0.001$
$R^2_{adj} = 0.45$ $P <= 0.001$
$R^2_{adj} = 0.19$ $P = 0.00881$
$R^2_{adj} = 0.33$ $P <= 0.001$

Absolute Error of Stubbs Clock on Thompson's data

Blood 5.072 Liver 6.177 Lung 8.258 Muscle 6.831 All Tissues 6.667

Tissue: Cortex, Heart, Liver, Lung, All Tissues

E

Meer Clock Applied to Thompson Data

$R^2_{adj} = 0.89$ $P <= 0.001$
$R^2_{adj} = 0.64$ $P <= 0.001$
$R^2_{adj} = 0.45$ $P <= 0.001$
$R^2_{adj} = 0.041$ $P = 0.145$
$R^2_{adj} = 0.45$ $P <= 0.001$

Absolute Error of Meer Clock on Thompson Data

Blood 1.904 Liver 3.863 Lung 5.477 Muscle 7.487 All Tissues 3.715

Tissue: Blood, Liver, Lung, Muscle, All tissues

F

Petkovich Clock Applied to Thompson Data

$R^2_{adj} = 0.41$ $P <= 0.001$

Absolute Error of Petkovich Clock Applied to Thompson Data

Blood 5.681

Figure 6.1: Mouse epigenetic clocks applied to various datasets. (**A**) Median coverage of multitissue datasets for CpG sites of two multi-tissue mouse RRBS clocks. Rows are each clock applied to a particular dataset (columns). For each plot, x-axis is a particular CpG represented by a line, y-axis is the median number of reads (coverage) per CpG. Top row; Stubbs clock sites (329 CpGs), Bottom row; Meer clock sites (435 CpGs). (**B-E**) Two multi-tissue clocks applied to multi-tissue datasets. Left panels; epigenetic age (eAge) plotted against chronological age (chAge) in months (mo).   Right panels; absolute error in months.   (**B,C**) Stubbs clock (**B**) and Meer clock (**C**) applied to Stubbs et al. (2017), which was used to train both clocks. (**D,E**) Stubbs clock (**D**) and Meer clock (**E**) applied to Thompson et al. (2018), a dataset outside of their training. (**F**) Petkovitch blood clock applied to 3 mo blood samples (Thompson et al. 2018).

Table 6.2: Bisulphite sequencing mouse clocks applied to various datasets. R^2 and median absolute error (MAE) values result from the clocks applied to all tissues in a given dataset. Datasets highlighted yellow were used to construct the corresponding clock. Datasets in bold were used in the multi-tissue regional clock.

	Mouse clock	Datasets clocks were applied to	Tissue	Total adjusted R squared	Total MAE	GEO accession
Single Tissue Clocks	Petkovich blood	Thompson et al. (2018)	Blood (> 3 months old)	0.41	5.68	GSE120137
		Thompson et al. (2018)	Liver	0.32	5.96	GSE120137
		Meer et al. (2018)	Liver	-0.04	7.09	GSE121141
	Wang liver	Stubbs et al. (2017)	Liver	0.36	1.94	GSE93957
		Reizel et al. (2018)	Liver	0.69	0.87	GSE85251
		Reizel et al. (2015)	Liver	0.62	1.55	GSE60012
Multi-Tissue Clocks		Thompson et al. (2018)	Blood, liver, lung, muscle	0.33	6.67	GSE120137
		Meer et al. (2018)	Liver, lung, heart, cortex	0.13	8.03	GSE121141
	Stubbs	Stubbs et al. (2017)	Liver, lung, heart, cortex	0.78	0.71	GSE93957
		Reizel et al. (2018)	Hepatocyte, lung, heart, hippocampus	0.90	0.33	GSE85251
		Reizel et al. (2015)	Liver, muscle, cerebellum, spleen	0.89	0.23	GSE60012
		Thompson et al. (2018)	Blood, liver, lung, muscle	0.45	3.72	GSE120137
		Meer et al. (2018)	Liver, lung, heart, cortex	0.75	1.83	GSE121141
	Meer	Stubbs et al. (2017)	Liver, lung, heart, cortex	0.71	0.83	GSE93957
		Reizel et al. (2018)	Hepatocyte, lung, heart, hippocampus	0.55	1.36	GSE85251
		Reizel et al. (2015)	Liver, muscle, cerebellum, spleen	0.72	1.16	GSE60012
		Thompson et al. (2018)	Blood, liver, lung, muscle, kidney, adipose	0.78	8.77	GSE120137
		Meer et al. (2018)	Liver, lung, heart, cortex	0.13	4.89	GSE121141
	Thompson	Stubbs et al. (2017)	Liver, lung, heart, cortex	0.45	2.49	GSE93957
		Reizel et al. (2018)	Hepatocyte, lung, heart, hippocampus	0.47	9.60	GSE85251
		Reizel et al. (2015)	Liver, muscle, cerebellum, spleen	0.51	1.40	GSE60012

6.2.2   Generating Regional Epigenetic Clocks

We next developed mouse RRBS clocks based on average methylation of particular region or window sizes (eg. 2kb) to see if the resulting clocks were

more accurate on non-training datasets than published clocks. We segmented the mouse genome into various sizes (1-9kb) using a running window generator, and for each segment size we conducted penalised regression against age to select optimal regions and weights for age prediction (**Fig. 6.2A**). Nine regional blood clocks (RBCs) and regional multi-tissue clocks (RMTCs) were generated (1-9 kb region sizes each, **Table 6.3**). Whole blood from GSE80672 (aged 3-35 mo, training=141, test=21; Petkovich et al. 2017) was used to construct the RBC. Liver, muscle, spleen, lung, heart, cerebellum and hippocampus tissue from GSE60012 (Reizel et al. 2015) and GSE85251 (Reizel et al. 2018) was used to construct the RMTC (aged 0.23-4.6 mo, training=118, test=13). Only tissues with samples at two different time points minimum were used, hence tissues that only had samples at single time points were rejected.

Both the RBC and RMTC had R^2 values higher than 0.85 when applied to their respective training or test data (**Fig. 6.2B**). When the RBCs and RMTCs were applied to datasets outside of their training, we found 2 to 6kb segment sizes produced clocks with the best correlation with age (**Fig. 6.2C, Table 6.3**). Performance tends to fluctuate depending on segment size. The blood clock performed best with a 2kb window when applied to the Thompson et al. (2018) blood data (**Fig. 6.2C**, left panel). The MTC clock performed well at 2kb for both the Stubbs and Thompson multi-tissue datasets (**Fig. 6.2C**, right panel). The MTC clock performed best at 4kb on the Stubbs dataset, and better at 5kb with the Thompson data.
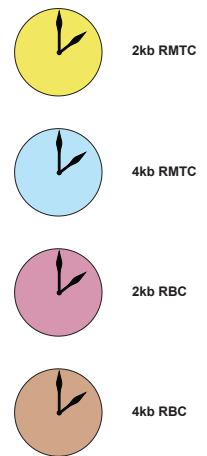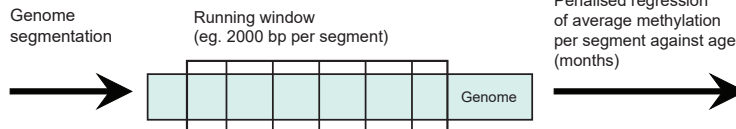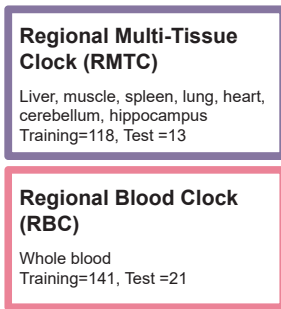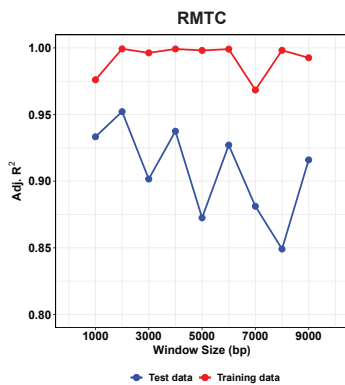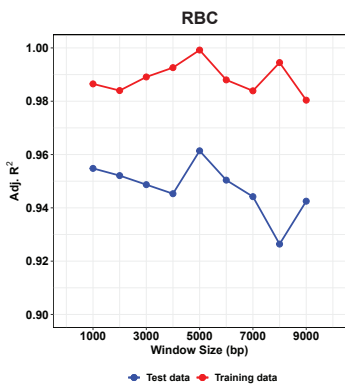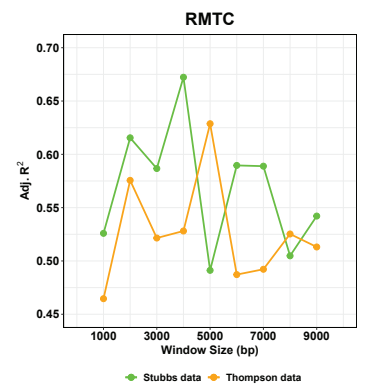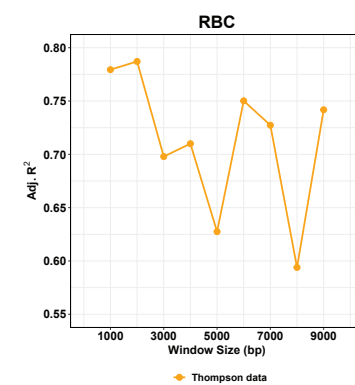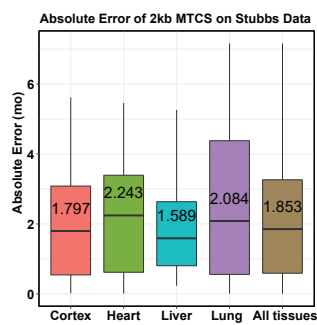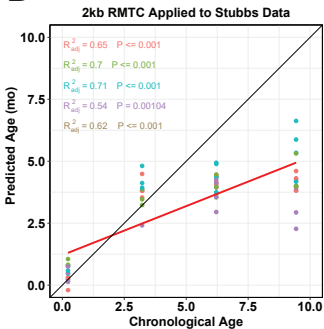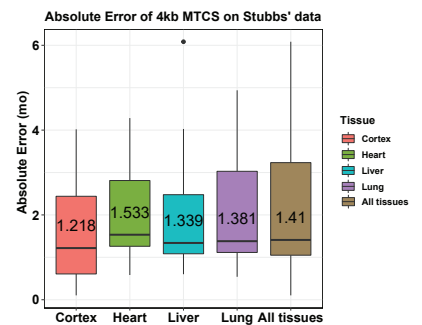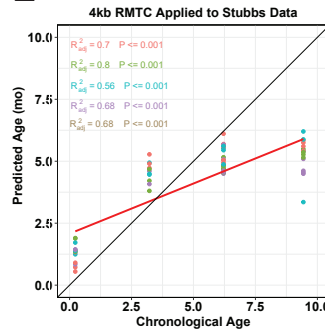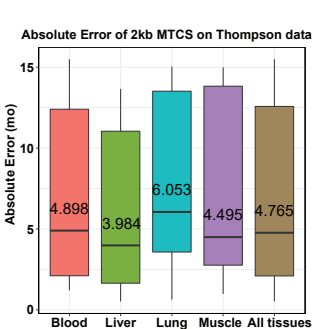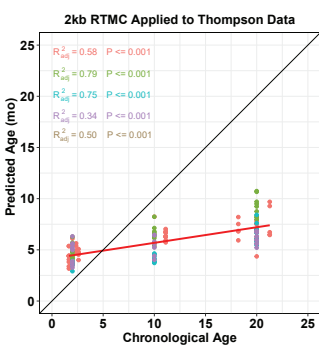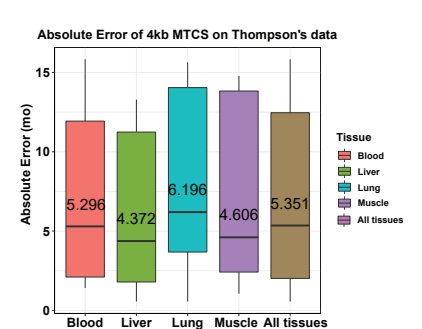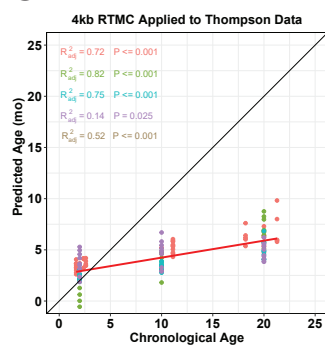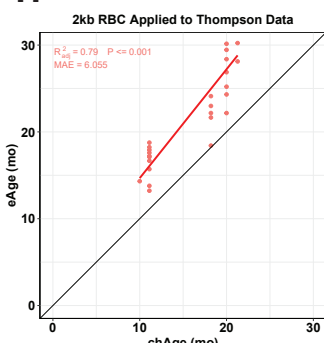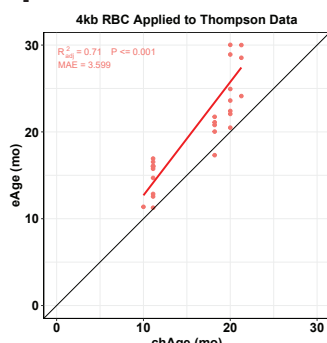
Figure 6.2: 2 and 4 kb genome window sizes provide robust age prediction in out-of-sample datasets. (**A**) Graphical overview of experiment. For both the multi-tissue and single tissue datasets, a running window generator was applied to segment the genome of the RRBS training data into various sizes (eg. 2kb). Segments were trained against age (months, mo) using penalised regression. Four regional epigenetic clocks were produced; a 2kb and 4kb regional multi-tissue clock (RMTC), and a 2kb and 4kb regional blood clock (RBC). (**B,C**) Adjusted R^2 values (y-axis) of each regional clocks (x-axis) epigenetic age correlated with chronological age in each dataset. (**B**) Left panel: Various RBCs applied to the RBC 90% training data (red line) and 10% test data (blue line). Right panel: RMTC applied to the RMTC 90% training data (red line) and 10% test data (blue line). (**C**) Left panel; RBCs applied to Thompson et al. (2018) blood dataset. Right panel; RMTCs applied to Stubbs et al. (2017) and Thompson et al. (2018) multi-tissue datasets. (**D-I**) Regional clocks applied to various datasets. Left panels; epigenetic age (eAge) plotted against chronological age (chAge) in months (mo). Right panels; absolute error in months. 2kb and 4kb RMTCs applied to Stubbs et al. (2017) (**D,E**, respectively) and Thompson et al. (2018) (**F,G**, respectively) datasets. (**H,I**) 2kb (**H**) and 4kb (**I**) RBCs applied to Thompson et al. (2018) 3 month blood datasets.

We chose 2kb and 4kb segment sizes of the RBC and RMTC to apply to the Stubbs and Thompson datasets (**Fig. 6.2C-H**), which were not used to train these clocks (see **Table. 6.3** for all window sizes of RBCs and RMTCs applied to these datasets). The Stubbs et al. (2017) dataset consisted of mouse cortex, heart, liver and lung samples aged 0-9 months old. Overall the 4kb clock had the highest adj. R^2 (0.67) and lowest MAE for all tissue types (**Fig. 6.2D,E**). The Stubbs and Meer clocks were more accurate than the RMTCs applied to the Stubbs et al. (2017) data, which is expected since this data was used to train both clocks. However, the MAE of the 4kb was 1.4 mo for all tissues, which is only ~0.6-0.7 mo greater than the Stubbs and Meer clocks. In addition, the 4kb RMTC performs accurately on cortex tissue (adj. R^2 = 0.7, MAE = 1.2), which was not included in the training data for this clock.

The Thompson et al. (2018) multi-tissue dataset consists of blood, liver, lung and muscle, aged ~2-22 months. Both the 2kb and 4kb RMTCs did not perform as well on this dataset (**Fig. 6.2G,H**) than when applied to the younger samples in the Stubbs dataset, which is not surprising since the RMTCs were trained on

Table 6.3: Regional blood clocks (RBCs) and regional multi-tissue clocks (RMTCs) of various window sizes applied to Stubbs et al. (2017) and Thompson et al. (2018) datasets. Datasets in bold are the regional clocks presented in **Fig. 6.2**.

Datasets	Clocks	Total adjusted R squared	Total MAE
Stubbs et al. (2017) GSE93957	1kb RMTC	0.526	1.635
	2kb RMTC	0.616	1.853
	3kb RMTC	0.587	1.355
	4kb RMTC	0.672	1.410
	5kb RMTC	0.491	1.857
	6kb RMTC	0.590	1.492
	7kb RMTC	0.589	1.482
	8kb RMTC	0.505	1.608
	9kb RMTC	0.542	1.531
Thompson et al. (2018) GSE120137	1kb RMTC	0.417	6.094
	2kb RMTC	0.492	4.425
	3kb RMTC	0.523	5.844
	4kb RMTC	0.519	6.126
	5kb RMTC	0.628	5.948
	6kb RMTC	0.583	4.421
	7kb RMTC	0.503	4.836
	8kb RMTC	0.563	5.408
	9kb RMTC	0.536	4.248
Thompson et al. (2018) (> 3 months Blood) GSE120137	1kb RBC	0.779	6.161
	2kb RBC	0.787	6.055
	3kb RBC	0.698	6.223
	4kb RBC	0.710	3.599
	5kb RBC	0.628	9.767
	6kb RBC	0.750	7.307
	7kb RBC	0.727	5.359
	8kb RBC	0.594	4.309
	9kb RBC	0.742	5.539

young samples. The 2kb segment size was the most optimal for the RMTC when applied to the Thompson dataset. It had the lowest MAE when applied to liver, but also performed well with blood, even though no blood samples were included to train the RMTC. Both the 2k and 4kb RMTCs outperform the Stubbs multi-tissue clock when applied to the Thompson et al. (2018) data (**Fig. 6.1D**). The Meer clock outperformed the RMTCs when applied to the Stubbs et al. (2017) data which is not surprising since this dataset was included to train the Meer clock. The Meer clock outperformed both RMTCs for most tissues in the Thompson et al. (2018) data except muscle. Overall, the RMTCs had higher R^2 values for all

tissues when applied to the Thompson et al. (2018) study than the Meer clock.

Both RBCs had an R^2 higher than 0.70 when applied to the Thompson et al. (2018) <3 month old blood data. While the 2kb blood clock had the highest R^2 out of all the other window sizes, the 4kb clock had the lowest MAE of 3.6 mo (**Table. 6.3, Fig. 6.2H,I**). The 4kb RMTC also outperforms the Petkovitch blood clock applied to the Thompson et al. (2018) dataset (**Fig. 6.1F**) as it had a higher correlation and lower MAE.

6.3   Discussion

Age prediction utilising RRBS data for accurate age prediction has proven difficult due to the uneven coverage of CpG sites captured between different experiments. We have shown that an approach using regions of CpGs can improve the robustness of age prediction. Using the same dataset used to create the Petkovitch blood clock, our RBC outperforms this clock when tested on non-training datasets.

Our RMTCs outperform most multi-tissue predictors when applied to datasets outside of training data. The only clock that appears to outperform the RMTCs is the Meer clock. This is likely because Meer et al. constructed their clock only using CpGs that were present in most of the datasets they analysed (**Fig. 6.1A**; Meer et al. 2018). The RTMCs showed transferability to non-training tissue types (such as cortex and blood). It is possible that age specific genomic regions are enough to capture biological age regardless of tissue type. The RMTCs also performed admirably on the 0-9 mo Stubbs et al. (2017) data, and almost as well as the Stubbs and Meer clocks which were originally trained on this dataset (**Fig. 1B,C, Fig. 2D,E**).

Interestingly, the trend of the predicted age for the RMTCs applied to young mice appears to be logarithmic/quadratic (**Fig. 6.2D,E**. It is possible that eAge progresses in a non-linear fashion. For example, the epigenetic pacemaker model

is an epigenetic clock where the predicted age follows a logarithmic trend (Snir,

VonHoldt, and Pellegrini 2016; Snir, Farrell, and Pellegrini 2019). In addition, the

Horvath clock applies a logarithmic transformation to ages <20 years (Horvath

2013; Snir, Farrell, and Pellegrini 2019). The Horvath and Hannum clocks also

under-predicts the age of older individuals (see Chapter 1.4.5; El Khoury et al.

2019). Therefore, eAge/biological age might have a non-linear relationship with

chAge. Training clocks on specific age ranges (eg. the Stubbs clock and our

RMTCs that are trained on under 9 mo mice) might be a more effective approach.

Alternatively, training a clock assuming a non-linear relationship with age, rather

than linear, might produce a more accurate age predictor.

Targeted methods have been developed for mice to reduce the cost that

comes with RRBS and WGBS. The Wagner lab developed a 3 CpG and 15 CpG

clock for multiple platforms (pyrosequencing, droplet digital PCR and barcoded

bisuphite amplicon sequencing), which offered accurate age prediction at few

CpGs for relatively low cost (Han et al. 2018; Han et al. 2020). Techniques such

as this might be more viable for age prediction since they have a higher

coverage at a few CpGs, whereas RRBS experiments vary in coverage

necessary for CpG-specific age prediction (**Fig. 6.1**). Nonetheless, our regional

genomic clocks have offered a proof in principle that age prediction can be more

effective by accounting for multiple CpGs, which negates the lack of read depth

at key CpGs that would otherwise hinder single CpG-based clocks.

Chapter 7

Discussion

This thesis has covered a number of research questions regarding epigenetics and ageing. Chapter 1 consisted of a literature review of the current theories of ageing, rejuvenation and the quantification of biological age. DNAm remains one of the most accurate biomarkers for ageing, as shown by the accuracy of the epigenetic clocks. However, it remains unclear whether eAge is a true proxy for biological age. For example, some studies have shown no correlation between eAge acceleration and certain diseases or environmental factors, such as type II diabetes (Horvath et al. 2016a; Grant et al. 2017), heroin use (Kozlenkov et al. 2017), or depression (Starnawska et al. 2019). Why eAge acceleration tracks with certain age-related disorders and not others is still not well understood. Difficulties may in part be the fact that there is not a precise, formal definition of "biological age", particularly in the context of epigenetic clocks (Field et al. 2018). Epigenetic age predictors also have limitations predicting particular age ranges. The tick-rate of the epigenetic clock (see Section 1.4.5, pg. 23) appears to be higher for under 20 year olds, hence the Horvath clock applies a logarithmic transformation to ages <20 years, while the linear model remains unaltered for ages >20 years (Horvath 2013; Snir, Farrell, and Pellegrini 2019). The Horvath and Hannum clocks systematically under predict the age of over 60 years olds in blood (El Khoury et al. 2019). This is could be due to survival bias, where the lower eAge reflects a lower biological age, which enabled the

centenarians to live longer. However, since the clocks appear to consistently under predict older age, this assumption is still in question. Alternatively, this discrepancy might be the result of a regression to the mean effect, where very high values (eAges) are underestimated by regression models. It is therefore possible that epigenetic clocks are limited because they are built using linear models. Biological age likely does not follow a linear trend. Hence alternate models such as deep learning and support vector regression (Xu et al. 2015; Aliferi et al. 2018; Levy et al. 2020; Galkin et al. 2020; Galkin et al. 2021), or logarithmic models such as the epigenetic pacemaker (Snir, VonHoldt, and Pellegrini 2016; Snir, Farrell, and Pellegrini 2019) might prove a more effective alternative to linear models. Clocks trained on specific age groups, such as PedBE, are valid alternatives that could prove effective (McEwen et al. 2019).

The approach used by Horvath to develop epigenetic clocks has spawned not only an abundance of similar DNAm age predictors, but also other novel approaches, such as transcriptional (Peters et al. 2015; Bryois et al. 2017), proteomic (Tanaka et al. 2018; Lehallier et al. 2019) and cellular biophysical/biomolecular (Phillip et al. 2017) clocks (see Section 1.5, pg. 34). Indeed, DNAm can be regressed with health co-factors such as smoking and alcohol consumption to produce predictors of complex traits and mortality (McCartney et al. 2018). While DNAm is one of the most accurate and versatile biomarkers for ageing and disease, our understanding of it is still developing. eAge by definition is not the same as chAge, yet epigenetic age predictors are built by training DNAm on chAge in a linear model (Field et al. 2018). Building a highly accurate predictor of chAge does not necessarily equate to an accurate predictor of biological age. Therefore, training eAge on chAge alone is not enough to explain biological age, as demonstrated by composite clocks such as PhenoAge and GrimAge clocks. Composite clocks (see Section 1.4.4) are created using CpGs that correlate with metrics of physiological or cellular ageing (such as c-reactive protein, serum albumin and systolic blood pressure) to train

in a linear model instead of chAge (Levine 2013; Lu et al. 2019a). Hence, looking at DNAm in combination with other non-DNAm based biomarkers may broaden our understanding, predictive power and accuracy of biological ageing and mortality. Composite clocks such as PhenoAge and GrimAge are first steps in that direction. Transcription clocks may reveal regulators of biological ageing, for example, if key ageing genes are found to be linked with eAge either by correlating with age acceleration or directly with methylation changes of key clock CpGs.

Partial reprogramming was discussed as a potential rejuvenation strategy where biological age is reduced by transiently expressed Yamanaka (OSKM) factors, but cell identity is retained. Ocampo et al. (2016) were the first to show this was possible *in vivo*. What remained unclear was the nature of rejuvenation occurring.   Was a subpopulation of cells dedifferentiating, or partially dedifferentiating and producing a rejuvenative effect to surrounding cells by being more stem-like? Alternatively, were the cells epigenetically rejuvenated, meaning that the cells became more youthful without loss of somatic cell identity? To test these hypotheses, two biomarkers were required to track (i) biological ageing and (ii) dedifferentiation state.

In Chapter 4, a reprogramming time-course of fibroblasts dedifferentiating to iPSCs was analysed using a variety of epigenetic clocks and comparing eAge (biological age) with gene expression. Analysing the declining eAge trajectory in combination with fibroblast and pluripotency cell markers helped us determine that eAge does indeed have separate kinetics to dedifferentiation. We also determined that a safe window may exist during dedifferentiation where discontinuation of the process may result in rejuvenated cells. We did not explicitly state boundaries where this window might exist, since our bulk expression analysis did not allow for precise identification of safe rejuvenation boundaries. Experiments at a single cell level and in *in vivo* conditions would help confirm a safe epigenetic rejuvenation window depending on the

reprogramming system. However, we observed a steep decline in eAge and stable expression of somatic genes between days 7 and 15 (**Fig. 4.7**), which means stable rejuvenation could occur if dedifferentiation was discontinued during this time.

A recent preprint might have confirmed this hypothesis, where Gill et al. transiently expressed OSKM in fibroblasts from middle aged donors and discontinued expression after 10-17 days (Gill et al. 2021). Between days 10-17, cells temporarily lose their cell identity and their eAge is reduced. After discontinuation of OSKM treatment, cells revert back to their original cell identity and retain their reduction of eAge. Since our study was published (Olova et al. 2018a), other studies have shown similar results. Sarkar et al. transiently expressed OSKM+LIN28+NANOG (OSKMLN) for four days in aged fibroblasts and endothelial cells, then analysed gene expression and methylation two days after halting treatment (Sarkar et al. 2020). Lu et al. showed that ectopic expression of Oct4, Sox2 and Klf4 (OSK) stimulated axon regeneration and recovered vision in mice (Lu et al. 2020). Both studies showed a reduction in eAge after treatment. While Sarkar et al. confirmed somatic identity was retained in their cell cultures, Lu et al. only analysed whether tumours developed in their mice, which after 10 months they did not. In both cases, no measurement of pluripotency markers (eg. TRA-1-60) was conducted, which means the degree of dedifferentiation of these cells cannot be confirmed. This highlights the importance of sorting for cells expressing pluripotency markers (eg. TRA-1-60, as shown in our study Fig. 4.1) when analysing epigenetic rejuvenation and partial reprogramming. Cells expressing TRA-1-60 are progressing towards a state of pluripotency. When obtained between days 7-11 of dedifferentiation, somatic gene expression is still high, hence these cells are characterised as partially reprogrammed (Tanabe et al. 2013). Without confirming the presence of partially reprogrammed cells, or the degree of pluripotency more specifically, safe implementation of Yamanaka factors as a

rejuvenation therapy cannot be ensured.

As reviewed in Chapter 1.4.6, caution still must be considered when interpreting biological age. It is not yet fully understood what physiological aspects of ageing the epigenetic clocks are measuring. Chapter 5 showed that the epigenetic clock is not resistant to global changes in methylation. Steps must be taken in experiments to account for biological and technical biases that might cause global methylation, and in turn, eAge offsets. Genotyping for somatic mutations in DNA methylation machinery might be one solution. Alternatively, designing an epigenetic clock that does not use, or can account for, saturation or desaturation of CpGs, might result in a more accurate and robust clock. Non-linear approaches to building an epigenetic clock might be a solution, ie. designing a clock that does not assume that eAge/biological age progresses in a linear fashion.

Finally, in Chapter 6, a proof of principle was presented that uses average methylation per region, rather than per individual CpG, to improve the robustness of BS-seq based mouse age prediction. This approach offered a more accurate blood age predictor than the Petkovich blood clock, and a multi-tissue clock more accurate than the Stubbs multi-tissue clock.

Bibliography

Abad, María et al. (Oct. 2013). "Reprogramming in vivo produces teratomas and iPS cells with totipotency features". *Nature* 502.7471, pp. 340–345.

Abelson, Sagi and Jean C.Y. Wang (Nov. 2018). "Age-related clonal hematopoiesis". *Current Opinion in Hematology* 25.6, pp. 441–445.

Afanas'ev, Igor (2010). "Signaling and damaging functions of free radicals in aging-free radical theory, hormesis, and TOR". *Aging and Disease* 1.2, pp. 75–88.

Al Muftah, Wadha A. et al. (2016). "Epigenetic associations of type 2 diabetes and BMI in an Arab population". *Clinical Epigenetics*.

Aliferi, Anastasia et al. (Nov. 2018). "DNA methylation-based age prediction using massively parallel sequencing data and multiple machine learning models". *Forensic Science International: Genetics* 37, pp. 215–226.

Alle, Quentin et al. (May 2021). "A single short reprogramming early in life improves fitness and increases lifespan in old age". *bioRxiv*, p. 2021.05.13.443979.

Amir, Hadanny and Efrati Shai (June 2020). "The hyperoxic-hypoxic paradox". *Biomolecules* 10.6, pp. 1–17.

Anastasiadi, Dafni and Francesc Piferrer (Mar. 2020). "A clockwork fish: Age prediction using DNA methylation-based biomarkers in the European seabass". *Molecular Ecology Resources* 20.2, pp. 387–397.

Andersen, Stacy L. et al. (Apr. 2012). "Health span approximates life span among many supercentenarians: Compression of morbidity at the approximate limit of

life span". *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 67 A.4, pp. 395–405.

Andrews, Simon (2007). "Babraham Bioinformatics - SeqMonk Mapped Sequence Analysis Tool".

Angermueller, Christof et al. (Feb. 2016). "Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity". *Nature Methods* 13.3, pp. 229–232.

Anisimov, Vladimir N. et al. (Dec. 2011). "Rapamycin increases lifespan and inhibits spontaneous tumorigenesis in inbred female mice". *Cell Cycle* 10.24, pp. 4230–4236.

Austad, Steven N and Jessica M Hoffman (Jan. 2018). "Is antagonistic pleiotropy ubiquitous in aging biology?" *Evolution, Medicine, and Public Health* 2018.1, pp. 287–294.

— (Jan. 2019). "Response to genes that improved fitness also cost modern humans: evidence for genes with antagonistic effects on longevity and disease". *Evolution, Medicine, and Public Health* 2019.1, pp. 7–8.

Bacalini, Maria Giulia et al. (2017). "Systemic Age-Associated DNA Hypermethylation of ELOVL2 Gene: In Vivo and in Vitro Evidences of a Cell Replication Process". *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*.

Baccarelli, Andrea et al. (Nov. 2010). "Ischemic heart disease and stroke in relation to blood DNA methylation". *Epidemiology* 21.6, pp. 819–828.

Baheti, Saurabh et al. (Dec. 2016). "Targeted alignment and end repair elimination increase alignment and methylation measure accuracy for reduced representation bisulfite sequencing data". *BMC Genomics* 17.1, p. 149.

Baht, Gurpreet S. et al. (May 2015). "Exposure to a youthful circulaton rejuvenates bone repair through modulation of β-catenin". *Nature Communications* 6.1, pp. 1–10.

Baker, Darren J. et al. (Nov. 2011). "Clearance of p16 Ink4a-positive senescent cells delays ageing-associated disorders". *Nature* 479.7372, pp. 232–236.

Baker, Darren J. et al. (Feb. 2016). "Naturally occurring p16 Ink4a-positive cells shorten healthy lifespan". *Nature* 530.7589, pp. 184–189.

Baker, George T. and Richard L. Sprott (Jan. 1988). "Biomarkers of aging". *Experimental Gerontology* 23.4-5, pp. 223–239.

Banks, Alexander S. et al. (Oct. 2008). "SirT1 Gain of Function Increases Energy Efficiency and Prevents Diabetes in Mice". *Cell Metabolism* 8.4, pp. 333–341.

Barger, Jamie L. et al. (June 2008). "Correction: A Low Dose of Dietary Resveratrol Partially Mimics Caloric Restriction and Retards Aging Parameters in Mice". *PLoS ONE* 3.6.

Barnard, Jordan, Deborah Snowdon, and Lucy Hewitson (Jan. 2018). "Dietary Interventions for Weight Loss and Essential Aspects of Nutrition Post–Bariatric Surgery". In: *Practical Guide to Obesity Medicine*. Elsevier, pp. 315–327.

Barros, Bela et al. (July 2020). "Loss of Chromosome Y and Its Potential Applications as Biomarker in Health and Forensic Sciences". *Cytogenetic and Genome Research* 160.5, pp. 225–237.

Bates, Douglas et al. (Oct. 2015). "Fitting linear mixed-effects models using lme4". *Journal of Statistical Software* 67.1.

Baur, Joseph A. et al. (Nov. 2006). "Resveratrol improves health and survival of mice on a high-calorie diet". *Nature* 444.7117, pp. 337–342.

Bell, Christopher G. et al. (Nov. 2019). "DNA methylation aging clocks: Challenges and recommendations". *Genome Biology* 20.1, pp. 1–24.

Belsky, D. W. et al. (May 2020). "Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm". *eLife* 9, pp. 1–56.

Belsky, Daniel W. et al. (July 2015). "Quantification of biological aging in young adults". *Proceedings of the National Academy of Sciences of the United States of America* 112.30, E4104–E4110.

Benayoun, Bérénice A., Elizabeth A. Pollina, and Anne Brunet (Oct. 2015). "Epigenetic regulation of ageing: Linking environmental inputs to genomic stability". *Nature Reviews Molecular Cell Biology* 16.10, pp. 593–610.

Berdyshev, G D et al. (1967). "Nucleotide composition of DNA and RNA from somatic tissues of humpback and its changes during spawning". *Biokhimiia (Moscow, Russia)* 32.5, pp. 988–93.

Bergsma, Tessa and Ekaterina Rogaeva (2020). "DNA Methylation Clocks and Their Predictive Capacity for Aging Phenotypes and Healthspan". *Neuroscience Insights* 15.

Bestor, T. H. and V. M. Ingram (1983). "Two DNA methyltransferases from murine erythroleukemia cells: Purification, sequence specificity, and mode of interaction with DNA". *Proceedings of the National Academy of Sciences of the United States of America* 80.181, pp. 5559–5563.

Betts, D. H. (Jan. 2001). "Reprogramming of telomerase activity and rebuilding of telomere length in cloned cattle". *Proceedings of the National Academy of Sciences* 98.3, pp. 1077–1082.

Biechonski, Shahar et al. (Dec. 2018). "Attenuated DNA damage responses and increased apoptosis characterize human hematopoietic stem cells exposed to irradiation". *Scientific Reports* 8.1.

Binder, Alexandra M. et al. (Jan. 2018). "Faster ticking rate of the epigenetic clock is associated with faster pubertal development in girls". *Epigenetics* 13.1, pp. 85–94.

Bjornsson, Hans T. et al. (June 2008). "Intra-individual change over time in DNA methylation with familial clustering". *JAMA - Journal of the American Medical Association* 299.24, pp. 2877–2883.

Blüher, Matthias, Barbara B. Kahn, and C. Ronald Kahn (Jan. 2003). "Extended longevity in mice lacking the insulin receptor in adipose tissue". *Science* 299.5606, pp. 572–574.

Bocklandt, Sven et al. (2011). "Epigenetic predictor of age". *PLoS ONE*.

Bond, Jonathan et al. (July 2019). "DNMT3A mutation is associated with increased age and adverse outcome in adult t-cell acute lymphoblastic leukemia". *Haematologica* 104.8, pp. 1617–1625.

Bordone, Laura and Leonard Guarente (Apr. 2005). "Calorie restriction, SIRT1 and metabolism: Understanding longevity". *Nature Reviews Molecular Cell Biology* 6.4, pp. 298–305.

Bormann, Felix et al. (June 2016). "Reduced ¡scp¿DNA¡/scp¿ methylation patterning and transcriptional connectivity define human skin aging". *Aging Cell* 15.3, pp. 563–571.

Boroni, Mariana et al. (July 2020). "Highly accurate skin-specific methylome analysis algorithm as a platform to screen and validate therapeutics for healthy aging". *Clinical Epigenetics* 12.1.

Boyer, Laurie A. et al. (Sept. 2005). "Core transcriptional regulatory circuitry in human embryonic stem cells". *Cell* 122.6, pp. 947–956.

Brack, Andrew S. et al. (Aug. 2007). "Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis". *Science* 317.5839, pp. 807–810.

Brambrink, Tobias et al. (Feb. 2008). "Sequential Expression of Pluripotency Markers during Direct Reprogramming of Mouse Somatic Cells". *Cell Stem Cell* 2.2, pp. 151–159.

Brandhorst, Sebastian et al. (July 2015). "A Periodic Diet that Mimics Fasting Promotes Multi-System Regeneration, Enhanced Cognitive Performance, and Healthspan". *Cell Metabolism* 22.1, pp. 86–99.

Breitling, Lutz Philipp et al. (Feb. 2016). "Frailty is associated with the epigenetic clock but not with telomere length in a German cohort". *Clinical Epigenetics* 8.1, pp. 1–8.

Brown-Borg, H. M. et al. (1996). "Dwarf mice and the ageing process". *Nature* 384.6604, p. 33.

Brunet, Anne and Shelley L. Berger (2014). "Epigenetics of aging and aging-related disease". *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 69.SUPPL. 1, pp. 17–20.

Bryois, Julien et al. (Apr. 2017). "Time-dependent genetic effects on gene expression implicate aging processes". *Genome Research* 27.4, pp. 545–552.

Buganim, Yosef et al. (Sept. 2012). "Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase". *Cell* 150.6, pp. 1209–1222.

Burgstaller, Jörg Patrick and Gottfried Brem (Aug. 2017). "Aging of Cloned Animals: A Mini-Review". *Gerontology* 63.5, pp. 417–425.

Byars, Sean G and Konstantinos Voskarides (Jan. 2019). "Genes that improved fitness also cost modern humans: evidence for genes with antagonistic effects on longevity and disease". *Evolution, Medicine, and Public Health* 2019.1, pp. 4–6.

Cai, Jingli et al. (Mar. 2006). "Assessing Self-Renewal and Differentiation in Human Embryonic Stem Cell Lines". *Stem Cells* 24.3, pp. 516–530.

Callaway, Ewen (Aug. 2010). "GlaxoSmithKline strikes back over anti-ageing pills". *Nature*.

Camacho-Pereira, Juliana et al. (June 2016). "CD38 Dictates Age-Related NAD Decline and Mitochondrial Dysfunction through an SIRT3-Dependent Mechanism". *Cell Metabolism* 23.6, pp. 1127–1139.

Campisi, Judith (Feb. 2013). "Aging, Cellular Senescence, and Cancer". *Annual Review of Physiology* 75.1, pp. 685–705.

Cannon, Matthew V. et al. (Mar. 2014). "Maternal nutrition induces pervasive gene expression changes but no detectable DNA methylation differences in the liver of adult offspring". *PLoS ONE* 9.3.

Carroll, John (2015). "Controversial resveratrol triggers new headlines–this time for Alzheimer's — FierceBiotech".

Castellano, Joseph M. et al. (Apr. 2017). "Human umbilical cord plasma proteins revitalize hippocampal function in aged mice". *Nature* 544.7651, pp. 488–492.

Castilho, Rogerio M. et al. (Sept. 2009). "mTOR Mediates Wnt-Induced Epidermal Stem Cell Exhaustion and Aging". *Cell Stem Cell* 5.3, pp. 279–289.

Chambers, John C. et al. (2015). "Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: A nested case-control study". *The Lancet Diabetes and Endocrinology*.

Chandra, Tamir and Kristina Kirschner (June 2016). "Chromosome organisation during ageing and senescence". *Current Opinion in Cell Biology* 40, pp. 161–167.

Chang, Howard H.Y. et al. (Aug. 2017). "Non-homologous DNA end joining and alternative pathways to double-strand break repair". *Nature Reviews Molecular Cell Biology* 18.8, pp. 495–506.

Chang, Jianhui et al. (Jan. 2016). "Clearance of senescent cells by ABT263 rejuvenates aged hematopoietic stem cells in mice". *Nature Medicine* 22.1, pp. 78–83.

Chang, Yuqiao, He Li, and Zhikun Guo (Apr. 2014). "Mesenchymal Stem Cell-Like Properties in Fibroblasts". *Cellular Physiology and Biochemistry* 34.3, pp. 703–714.

Chen, Brian H. et al. (2016). "DNA methylation-based measures of biological age: Meta-analysis predicting time to death". *Aging* 8.9, pp. 1844–1865.

Chen, Chong et al. (Nov. 2009). "MTOR regulation and therapeutic rejuvenation of aging hematopoietic stem cells". *Science Signaling* 2.98, ra75–ra75.

Chen, Daniel et al. (Feb. 2020). "The lipid elongation enzyme ELOVL2 is a molecular regulator of aging in the retina". *Aging Cell* 19.2.

Chen, Li et al. (Jan. 2019). "Effects of Vitamin D 3 supplementation on epigenetic aging in overweight and obese african americans with suboptimal

Vitamin D status: A randomized clinical trial". *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 74.1, pp. 91–98.

Chiquoine, A. Duncan (Feb. 1954). "The identification, origin, and migration of the primordial germ cells in the mouse embryo". *The Anatomical Record* 118.2, pp. 135–146.

Cho, Sohee et al. (May 2014). "Age estimation via quantification of signal-joint T cell receptor excision circles in Koreans". *Legal Medicine* 16.3, pp. 135–138.

Cho, Sohee et al. (July 2017). "Independent validation of DNA-based approaches for age prediction in blood". *Forensic Science International: Genetics* 29, pp. 250–256.

Choufani, S. et al. (Dec. 2015). "NSD1 mutations generate a genome-wide DNA methylation signature". *Nature Communications* 6.1, pp. 1–7.

Chouliaras, Leonidas et al. (Sept. 2013). "Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients". *Neurobiology of Aging* 34.9, pp. 2091–2099.

Christiansen, Lene et al. (Feb. 2016). "DNA methylation age is associated with mortality in a longitudinal Danish twin study". *Aging Cell* 15.1, pp. 149–154.

Chuang, Jody C. et al. (Oct. 2005). "Comparison of biological effects of non-nucleoside DNA methylation inhibitors versus 5-aza-2-deoxycytidine". *Molecular Cancer Therapeutics* 4.10, pp. 1515–1520.

Chung, Kyung-Min et al. (Apr. 2014). "Single Cell Analysis Reveals the Stochastic Phase of Reprogramming to Pluripotency Is an Ordered Probabilistic Process". *PLoS ONE* 9.4. Ed. by Austin John Cooney, e95304.

Ciccarone, Fabio et al. (Aug. 2016). "Age-dependent expression of DNMT1 and DNMT3B in PBMCs from a large European population enrolled in the MARK-AGE study". *Aging Cell* 15.4, pp. 755–765.

Cimino, F. et al. (Dec. 2012). "Pulsed high oxygen induces a hypoxic-like response in human umbilical endothelial cells and in humans". *Journal of Applied Physiology* 113.11, pp. 1684–1689.

Clark, A. John et al. (2003). "Proliferative lifespan is conserved after nuclear transfer". *Nature Cell Biology*.

Clark, Stephen J. et al. (Dec. 2018). "ScNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells e". *Nature Communications* 9.1, pp. 1–9.

Clevers, Hans, Kyle M. Loh, and Roel Nusse (Oct. 2014). "An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control". *Science* 346.6205.

Clevers, Hans and Roel Nusse (June 2012). "Wnt/β-catenin signaling and disease". *Cell* 149.6, pp. 1192–1205.

Cohen, A.A. and D.J. Holmes (Jan. 2014). "Evolution and the Biology of Aging". In: *Reference Module in Biomedical Sciences*. Elsevier.

Cole, John J. et al. (2017). "Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions". *Genome Biology*.

Conboy, Irina M. et al. (Feb. 2005). "Rejuvenation of aged progenitor cells by exposure to a young systemic environment". *Nature* 433.7027, pp. 760–764.

Conway, Bryan R. et al. (Sept. 2020). "Kidney Single-Cell Atlas Reveals Myeloid Heterogeneity in Progression and Regression of Kidney Disease". *Journal of the American Society of Nephrology* 31.12, ASN.2020060806.

Cruickshanks, Hazel A. et al. (2013). "Senescent Cells Harbour Features of the Cancer Epigenome". *Nature Cell Biology* 15.12, pp. 1495–1506.

Cuervo, Ana Maria (Dec. 2008). "Autophagy and aging: keeping that old broom working". *Trends in Genetics* 24.12, pp. 604–612.

Cypris, Olivia et al. (Aug. 2020). "PRDM8 reveals aberrant DNA methylation in aging syndromes and is relevant for hematopoietic and neuronal differentiation". *Clinical epigenetics* 12.1, p. 125.

D'Aquila, Patrizia et al. (2013). "Epigenetics and aging". *Maturitas* 74.2, pp. 130–136.

Dai, Han et al. (Oct. 2010). "SIRT1 activation by small molecules: Kinetic and biophysical evidence for direct interaction of enzyme and activator". *Journal of Biological Chemistry* 285.43, pp. 32695–32703.

Dang, Weiwei (2014). "The controversial world of sirtuins". *Drug Discovery Today: Technologies* 12, e9.

Day, Kenneth et al. (Sept. 2013). "Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape". *Genome Biology* 14.9.

Declerck, Ken and Wim Vanden Berghe (Sept. 2018). "Back to the future: Epigenetic clock plasticity towards healthy aging". *Mechanisms of Ageing and Development* 174, pp. 18–29.

Dias, Helena Correia et al. (Mar. 2020a). "Age Estimation Based on DNA Methylation Using Blood Samples From Deceased Individuals". *Journal of Forensic Sciences* 65.2, pp. 465–470.

Dias, Helena Correia et al. (Nov. 2020b). "Age prediction in living: Forensic epigenetic age estimation based on blood samples". *Legal Medicine* 47.

Dias, Helena Correia et al. (June 2020c). "DNA methylation age estimation in blood samples of living and deceased individuals using a multiplex SNaPshot assay". *Forensic Science International* 311, p. 110267.

Dmitrijeva, Marija et al. (2018). "Tissue-specific DNA methylation loss during ageing and carcinogenesis is linked to chromosome structure, replication timing and cell division rates." *Nucleic acids research* 46.14, pp. 7022–7039.

Dor, Yuval and Howard Cedar (2018). "Principles of DNA methylation and their implications for biology and medicine". *www.thelancet.com* 392, p. 777.

Eipel, Monika et al. (May 2016). "Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures". *Aging* 8.5, pp. 1034–1048.

El Khoury, Louis Y. et al. (2019). "Systematic underestimation of the epigenetic clock and age acceleration in older subjects". *Genome Biology* 20.1.

Elliott, Sarah A. and Alejandro Sánchez Alvarado (May 2013). "The history and enduring contributions of planarians to the study of animal regeneration". *Wiley Interdisciplinary Reviews: Developmental Biology* 2.3, pp. 301–326.

Engberg, Henriette et al. (2009). "Centenarians - A useful model for healthy aging? A 29-year follow-up of hospitalizations among 40000 Danes born in 1905". *Aging Cell* 8.3, pp. 270–276.

Ermolaeva, Maria et al. (Sept. 2018). "Cellular and epigenetic drivers of stem cell ageing". *Nature Reviews Molecular Cell Biology* 19.9, pp. 594–610.

Espejel, Silvia et al. (May 2004). "Shorter telomeres, accelerated ageing and increased lymphoma in DNA-PKcs-deficient mice". *EMBO Reports* 5.5, pp. 503–509.

Esteller, Manel (Aug. 2002). "CpG island hypermethylation and tumor suppressor genes: A booming present, a brighter future". *Oncogene* 21.35 REV. ISS. 3, pp. 5427–5440.

Esteller, Manel et al. (2001). "A gene hypermethylation profile of human cancer". *Cancer Research*.

Fahy, Gregory M. et al. (Dec. 2019). "Reversal of epigenetic aging and immunosenescent trends in humans". *Aging Cell* 18.6.

Faragher, Richard G. A. (July 2015). "Should we treat aging as a disease? The consequences and dangers of miscategorisation". *Frontiers in Genetics* 6.JUL, p. 171.

Feinberg, Andrew P. and Bert Vogelstein (1983). "Hypomethylation distinguishes genes of some human cancers from their normal counterparts". *Nature* 301.5895, pp. 89–92.

Field, Adam E et al. (2018). "Molecular Cell Review DNA Methylation Clocks in Aging: Categories, Causes, and Consequences".

Flach, Johanna et al. (Aug. 2014). "Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells". *Nature* 512.7513, pp. 198–202.

Fleckhaus, Jan and Peter M. Schneider (Jan. 2020). "Novel multiplex strategy for DNA methylation-based age prediction from small amounts of DNA via Pyrosequencing". *Forensic Science International: Genetics* 44, p. 102189.

Florea, Michael (Oct. 2017). "Aging and immortality in unicellular species". *Mechanisms of Ageing and Development* 167, pp. 5–15.

FOXO BioScience (2020). "FOXO BioScience, Formerly Known as Life Epigenetics, Announces New Infinium Mouse Methylation Array in Strategic Collaboration with Van Andel Institute".

Fraga, Mario F. et al. (July 2005). "Epigenetic differences arise during the lifetime of monozygotic twins". *Proceedings of the National Academy of Sciences of the United States of America* 102.30, pp. 10604–10609.

Frenck, Robert W., Elizabeth H. Blackburn, and Kevin M. Shannon (May 1998). "The rate of telomere sequence loss in human leukocytes varies with age". *Proceedings of the National Academy of Sciences of the United States of America* 95.10, pp. 5607–5610.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani (2010). "Regularization paths for generalized linear models via coordinate descent". *Journal of Statistical Software*.

Frommer, M. et al. (Mar. 1992). "A genomic sequencing protocol that yields a positive display of 5- methylcytosine residues in individual DNA strands". *Proceedings of the National Academy of Sciences of the United States of America* 89.5, pp. 1827–1831.

Fuke, C. et al. (2004). "Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: An HPLC-based study". *Annals of Human Genetics*.

Galan, Amparo et al. (Apr. 2013). "Defining the Genomic Signature of Totipotency and Pluripotency during Early Human Development". *PLoS ONE* 8.4. Ed. by Ying Xu, e62135.

Galkin, Fedor et al. (June 2020). "Human Gut Microbiome Aging Clock Based on Taxonomic Profiling and Deep Learning". *iScience* 23.6, p. 101199.

Galkin, Fedor et al. (2021). "DeepMAge: A Methylation Aging Clock Developed with Deep Learning". *Aging and disease*, pp. 0-.

Gao, Xu et al. (Oct. 2015). "DNA methylation changes of whole blood cells in response to active smoking exposure in adults: A systematic review of DNA methylation studies". *Clinical Epigenetics* 7.1, p. 113.

Garagnani, Paolo et al. (Dec. 2012). "Methylation of ELOVL2 gene as a new epigenetic marker of age". *Aging Cell* 11.6, pp. 1132–1134.

Geiman, Theresa M. and Keith D. Robertson (2002). "Chromatin remodeling, histone modifications, and DNA methylation - How does it all fit together?" *Journal of Cellular Biochemistry* 87.2, pp. 117–125.

Genolini, Christophe et al. (May 2015). "Kml and kml3d: R packages to cluster longitudinal data". *Journal of Statistical Software* 65.4, pp. 1–34.

Gentilini, Davide et al. (Oct. 2013). "Role of epigenetics in human aging and longevity: Genome-wide DNA methylation profile in centenarians and centenarians' offspring". *Age* 35.5, pp. 1961–1973.

Giangreco, Adam et al. (Apr. 2008). "Epidermal stem cells are retained in vivo throughout skin aging". *Aging Cell* 7.2, pp. 250–259.

Gill, Diljeet et al. (Jan. 2021). "Multi-omic rejuvenation of human cells by maturation phase transient reprogramming". *bioRxiv*, p. 2021.01.15.426786.

Ginis, Irene et al. (May 2004). "Differences between human and mouse embryonic stem cells". *Developmental Biology* 269.2, pp. 360–380.

Gladyshev, Timothy V. and Vadim N. Gladyshev (Dec. 2016). "A Disease or Not a Disease? Aging As a Pathology". *Trends in Molecular Medicine* 22.12, pp. 995–996.

Goelz, Susan E. et al. (Apr. 1985). "Hypomethylation of DNA from benign and malignant human colon neoplasms". *Science* 228.4696, pp. 187–190.

Goglin, Sarah E. et al. (Oct. 2016). "Change in Leukocyte Telomere Length Predicts Mortality in Patients with Stable Coronary Heart Disease from the Heart and Soul Study". *PLOS ONE* 11.10. Ed. by Raymund J. Wellinger, e0160748.

Goodpaster, Tracy et al. (Apr. 2008). "An immunohistochemical method for identifying fibroblasts in formalin-fixed, paraffin-embedded tissue". *Journal of Histochemistry and Cytochemistry* 56.4, pp. 347–358.

Gorbunova, Vera and Andrei Seluanov (June 2016). "DNA double strand break repair, aging and the chromatin connection". *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 788, pp. 2–6.

Gorrie-Stone, Tyler J et al. (Mar. 2019). "Bigmelon: tools for analysing large DNA methylation datasets". *Bioinformatics* 35.6. Ed. by Janet Kelso, pp. 981–986.

Grant, Crystal D. et al. (Dec. 2017). "A longitudinal study of DNA methylation as a potential mediator of age-related diabetes risk". *GeroScience* 39.5-6, p. 475.

Gravina, Silvia et al. (Dec. 2016). "Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome". *Genome Biology* 17.1, p. 150.

Gruenbaum, Yosef, Howard Cedar, and Aharon Razin (1982). "Substrate and sequence specificity of a eukaryotic DNA methylase". *Nature* 295.5850, pp. 620–622.

Gu, Hongcang et al. (Mar. 2011). "Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling". *Nature Protocols* 6.4, pp. 468–481.

Guevara, Elaine E. et al. (Nov. 2020). "Age-associated epigenetic change in chimpanzees and humans". *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1811, p. 20190616.

Gurdon, J. B. (1962). "The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles." *Journal of embryology and experimental morphology* 10.December, pp. 622–640.

Gurdon, J. B., T. R. Elsdale, and M. Fischberg (1958). "Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei". *Nature* 182.4627, pp. 64–65.

Guttenbach, Martina, Renate Schakowski, and Michael Schmid (1994). "Aneuploidy and ageing: sex chromosome exclusion into micronuclei". *Human Genetics* 94.3, pp. 295–298.

Hachmo, Yafit et al. (Nov. 2020). "Hyperbaric oxygen therapy increases telomere length and decreases immunosenescence in isolated blood cells : a prospective trial". *Aging* 12.

Hadad, Niran et al. (2016). "Absence of genomic hypomethylation or regulation of cytosine-modifying enzymes with aging in male and female mice". *Epigenetics and Chromatin*.

Han, Yang et al. (Aug. 2018). "Epigenetic age-predictor for mice based on three CpG sites". *eLife* 7.

Han, Yang et al. (Dec. 2020). "Targeted methods for epigenetic age predictions in mice". *Scientific Reports* 10.1, pp. 1–10.

Hannum, Gregory et al. (2013). "Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates". *Molecular Cell*.

Harley, Calvin B., A. Bruce Futcher, and Carol W. Greider (1990). "Telomeres shorten during ageing of human fibroblasts". *Nature* 345.6274, pp. 458–460.

Harries, Lorna W. et al. (Oct. 2011). "Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing". *Aging Cell* 10.5, pp. 868–878.

Harrison, David E. et al. (July 2009). "Rapamycin fed late in life extends lifespan in genetically heterogeneous mice". *Nature* 460.7253, pp. 392–395.

Harvie, M. N. et al. (May 2011). "The effects of intermittent or continuous energy restriction on weight loss and metabolic disease risk markers: A randomized trial in young overweight women". *International Journal of Obesity* 35.5, pp. 714–727.

Hastie, Nicholas D. et al. (1990). "Telomere reduction in human colorectal carcinoma and with ageing". *Nature* 346.6287, pp. 866–868.

Hata, Kenichiro et al. (2002). "Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice". *Development* 129.8, pp. 1983–1993.

Hayano, Motoshi et al. (Oct. 2019). "DNA Break-Induced Epigenetic Drift as a Cause of Mammalian Aging". *SSRN Electronic Journal*, p. 808659.

Hayflick, L. (Mar. 1965). "The limited in vitro lifetime of human diploid cell strains". *Experimental Cell Research* 37.3, pp. 614–636.

Hayflick, L. and P. S. Moorhead (Dec. 1961). "The serial cultivation of human diploid cell strains". *Experimental Cell Research* 25.3, pp. 585–621.

He, Shenghui, Daisuke Nakada, and Sean J. Morrison (Nov. 2009). "Mechanisms of stem cell self-renewal". *Annual Review of Cell and Developmental Biology* 25, pp. 377–406.

Hendrich, Brian et al. (Sept. 1999). "The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites". *Nature* 401.6750, pp. 301–304.

Hentze, Hannes et al. (May 2009). "Teratoma formation by human embryonic stem cells: Evaluation of essential parameters for future safety studies". *Stem Cell Research* 2.3, pp. 198–210.

Hepp, Tobias et al. (May 2016). "Approaches to regularized regression - A comparison between gradient boosting and the lasso". *Methods of Information in Medicine* 55.5, pp. 422–430.

Herman, James G. and Stephen B. Baylin (Nov. 2003). "Gene Silencing in Cancer in Association with Promoter Hypermethylation". *New England Journal of Medicine* 349.21, pp. 2042–2054.

Hernandez-Segura, Alejandra, Jamil Nehme, and Marco Demaria (June 2018). "Hallmarks of Cellular Senescence". *Trends in Cell Biology* 28.6, pp. 436–453.

Hernando-Herraez, Irene et al. (Dec. 2019). "Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells". *Nature Communications* 10.1, p. 4361.

Herold, S. et al. (July 2017). "Donor cell leukemia: Evidence for multiple preleukemic clones and parallel long term clonal evolution in donor and recipient". *Leukemia* 31.7, pp. 1637–1640.

Herranz, Nicolás and Jesús Gil (Apr. 2018). "Mechanisms and functions of cellular senescence". *Journal of Clinical Investigation* 128.4, pp. 1238–1246.

Hewakapuge, Sudinna et al. (Sept. 2008). "Investigation of telomere lengths measurement by quantitative real-time PCR to predict age". *Legal Medicine* 10.5, pp. 236–242.

Hewitt, Graeme et al. (2012). "Telomeres are favoured targets of a persistent DNA damage response in ageing and stress-induced senescence". *Nature Communications*.

Heyn, Holger et al. (June 2012). "Distinct DNA methylomes of newborns and centenarians". *Proceedings of the National Academy of Sciences of the United States of America* 109.26, pp. 10522–10527.

Higgins-Chen, Albert T. et al. (Aug. 2020). "Schizophrenia and Epigenetic Aging Biomarkers: Increased Mortality, Reduced Cancer Risk, and Unique Clozapine Effects". *Biological Psychiatry* 88.3, pp. 224–235.

Hillary, Robert F. et al. (July 2020). "Epigenetic measures of ageing predict the prevalence and incidence of leading causes of death and disease burden". *Clinical epigenetics* 12.1, p. 115.

Hitt, Rachel et al. (Aug. 1999). "Centenarians: The older you get, the healthier you have been". *Lancet* 354.9179, p. 652.

Hiyama, E. and K. Hiyama (Apr. 2007). "Telomere and telomerase in stem cells". *British Journal of Cancer* 96.7, pp. 1020–1024.

Hofmann, Jeffrey W. et al. (Jan. 2015). "Reduced expression of MYC increases longevity and enhances healthspan". *Cell* 160.3, pp. 477–488.

Holly, Alice C. et al. (Apr. 2013). "Towards a gene expression biomarker set for human biological age". *Aging Cell* 12.2, pp. 324–326.

Holzenberger, Martin et al. (Jan. 2003). "IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice". *Nature* 421.6919, pp. 182–187.

Horvath, Steve (2013). "DNA methylation age of human tissues and cell types". *Genome Biology*.

— (2015). "Erratum to DNA methylation age of human tissues and cell types [Genome Biology, 14, R115, (2013)]". *Genome Biology* 16.1, pp. 1–5.

Horvath, Steve and Andrew J. Levine (Nov. 2015). "HIV-1 Infection Accelerates Age According to the Epigenetic Clock". *Journal of Infectious Diseases* 212.10, pp. 1563–1573.

Horvath, Steve and Kenneth Raj (2018). "DNA methylation-based biomarkers and the epigenetic clock theory of ageing".

Horvath, Steve et al. (Oct. 2012). "Aging effects on DNA methylation modules in human brain and blood tissue". *Genome biology* 13.10, R97.

Horvath, Steve et al. (Oct. 2014). "Obesity accelerates epigenetic aging of human liver". *Proceedings of the National Academy of Sciences of the United States of America* 111.43, pp. 15538–15543.

Horvath, Steve et al. (June 2015a). "Accelerated epigenetic aging in Down syndrome". *Aging Cell* 14.3, pp. 491–495.

Horvath, Steve et al. (2015b). "Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring". *Aging* 7.12, pp. 1159–1170.

Horvath, Steve et al. (2015c). "The cerebellum ages slowly according to the epigenetic clock". *Aging* 7.5, pp. 294–306.

Horvath, Steve et al. (Dec. 2016a). "An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease". *Genome Biology* 17.1, p. 171.

Horvath, Steve et al. (July 2016b). "Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels". *Aging* 8.7, pp. 1485–1512.

Horvath, Steve et al. (July 2018). "Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies". *Aging* 10.7, pp. 1758–1775.

Horvath, Steve et al. (May 2019). "Rapamycin retards epigenetic ageing of keratinocytes independently of its effects on replicative senescence, proliferation and differentiation". *Aging* 11.10, pp. 3238–3249.

Horvath, Steve et al. (Nov. 2020a). "DNA methylation age analysis of rapamycin in common marmosets". *bioRxiv*, p. 2020.11.21.392779.

Horvath, Steve et al. (May 2020b). "Reversing age: dual species measurement of epigenetic age with a single clock". *bioRxiv*, p. 2020.05.07.082917.

Howitz, Konrad T. et al. (Sept. 2003). "Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan". *Nature* 425.6954, pp. 191–196.

Hubbard, Basil P. and David A. Sinclair (Mar. 2014). "Small molecule SIRT1 activators for the treatment of aging and age-related diseases". *Trends in Pharmacological Sciences* 35.3, pp. 146–154.

Huh, Christine J. et al. (Sept. 2016). "Maintenance of age in human neurons generated by microRNA-based neuronal conversion of fibroblasts". *eLife* 5.September2016.

Ibrahim, Samah F., Iman F. Gaballah, and Laila A. Rashed (July 2016). "Age Estimation in Living Egyptians Using Signal Joint T-cell Receptor Excision Circle Rearrangement". *Journal of Forensic Sciences* 61.4, pp. 1107–1111.

Ichiyama, Kenji et al. (Apr. 2015). "The Methylcytosine Dioxygenase Tet2 Promotes DNA Demethylation and Activation of Cytokine Gene Expression in T Cells". *Immunity* 42.4, pp. 613–626.

Ignjatovic, Vera et al. (2011). "Age-related differences in plasma proteins: How plasma proteins change from neonates to adults". *PLoS ONE*.

Illumina (2015a). *Infinium HD Methylation Assay Protocol Guide (15019519)*. Tech. rep.

— (2015b). *Infinium Methylation Coverage*. Tech. rep.

Issa, Jean Pierre (Jan. 2014). "Aging and epigenetic drift: A vicious cycle". *Journal of Clinical Investigation* 124.1, pp. 24–29.

Ito, Hideyuki et al. (Dec. 2018). "Estimation of chimpanzee age based on DNA methylation". *Scientific Reports* 8.1, p. 9998.

Jackson, Melany et al. (Oct. 2004). "Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells". *Molecular and Cellular Biology* 24.20, pp. 8862–8871.

Jacobs, Patricia A. et al. (1963). "Change of human chromosome count distributions with age: Evidence for a sex difference". *Nature* 197.4872, pp. 1080–1081.

Janmaat, C. J. et al. (Dec. 2015). "Human Dermal Fibroblasts Demonstrate Positive Immunostaining for Neuron- and Glia- Specific Proteins". *PLOS ONE* 10.12. Ed. by Mária A. Deli, e0145235.

Jeong, Mira et al. (Apr. 2018). "Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo". *Cell Reports* 23.1, pp. 1–10.

Ji, Pengfei et al. (2016). "Induced Pluripotent Stem Cells: Generation Strategy and Epigenetic Mystery behind Reprogramming". *Stem Cells International* 2016.

Jin, Chunlei et al. (2014). "TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells". *Nucleic Acids Research*.

Jin, Kunlin (2010). "Modern biological theories of aging". *Aging and Disease*.

Johnson, Simon C., Peter S. Rabinovitch, and Matt Kaeberlein (Jan. 2013). "MTOR is a key modulator of ageing and age-related disease". *Nature* 493.7432, pp. 338–345.

Jones, Peter A. and Shirley M. Taylor (May 1980). "Cellular differentiation, cytidine analogs and DNA methylation". *Cell* 20.1, pp. 85–93.

Jung, Marc and Gerd P. Pfeifer (Dec. 2015). "Aging and DNA methylation". *BMC Biology* 13.1, p. 7.

Jung, Sang-Eun et al. (Jan. 2019). "DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from

blood, saliva, and buccal swab samples". *Forensic Science International: Genetics* 38, pp. 1–8.

Junnila, Riia K. et al. (June 2013). "The GH/IGF-1 axis in ageing and longevity". *Nature Reviews Endocrinology* 9.6, pp. 366–376.

Jylhävä, Juulia, Nancy L Pedersen, and Sara Hägg (July 2017). "Biological Age Predictors." *EBioMedicine* 21, pp. 29–36.

Kabacik, Sylwia et al. (Oct. 2018). "Epigenetic ageing is distinct from senescence-mediated ageing and is not prevented by telomerase expression". *Aging* 10.10, pp. 2800–2815.

Kaeberlein, Matt, Mitch McVey, and Leonard Guarente (Oct. 1999). "The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms". *Genes and Development* 13.19, pp. 2570–2580.

Kalluri, Raghu and Michael Zeisberg (May 2006). "Fibroblasts in cancer". *Nature Reviews Cancer* 6.5, pp. 392–401.

Kane, Alice E. and David A. Sinclair (Sept. 2018). "Sirtuins and NAD+ in the development and treatment of metabolic and cardiovascular diseases". *Circulation Research* 123.7, pp. 868–885.

Kanfi, Yariv et al. (Apr. 2010). "SIRT6 protects against pathological damage caused by diet-induced obesity". *Aging Cell* 9.2, pp. 162–173.

Karanam, Ketki et al. (July 2012). "Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase". *Molecular Cell* 47.2, pp. 320–329.

Karlsson, Andreas O. et al. (Aug. 2008). "Estimating human age in forensic samples by analysis of telomere repeats". *Forensic Science International: Genetics Supplement Series* 1.1, pp. 569–571.

Katrinli, Seyma et al. (May 2020). "Evaluating the impact of trauma and PTSD on epigenetic prediction of lifespan and neural integrity". *Neuropsychopharmacology*, pp. 1–8.

Katsimpardi, Lida et al. (May 2014). "Vascular and neurogenic rejuvenation of the aging mouse brain by young systemic factors". *Science* 344.6184, pp. 630–634.

Kazyken, Dubek et al. (June 2019). "AMPK directly activates mTORC2 to promote cell survival during acute energetic stress". *Science Signaling* 12.585, p. 3249.

Kelley, K. W., S. Brief, and H. J. Westley (Aug. 1986). "GH3 pituitary adenoma cells can reverse thymic aging in rats". *Proceedings of the National Academy of Sciences of the United States of America* 83.15, pp. 5663–5667.

Keyes, Brice E. et al. (Dec. 2013). "Nfatc1 orchestrates aging in hair follicle stem cells". *Proceedings of the National Academy of Sciences of the United States of America* 110.51, E4950–E4959.

Kim, Chul Hong et al. (Dec. 2016). "Short-term calorie restriction ameliorates genomewide, age-related alterations in DNA methylation". *Aging Cell* 15.6, pp. 1074–1081.

Kim, Soo Jin et al. (2013). "A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells". *Blood* 122.25, pp. 4086–4089.

Kim, Yumi, Hong Gil Nam, and Dario Riccardo Valenzano (Feb. 2016). "The short-lived African turquoise killifish: An emerging experimental model for ageing". *DMM Disease Models and Mechanisms* 9.2, pp. 115–129.

Kirkland, James L. and Tamara Tchkonia (July 2017). "Cellular Senescence: A Translational Perspective". *EBioMedicine* 21, pp. 21–28.

Kirkwood, T. B.L. (1977). "Evolution of ageing". *Nature* 270.5635, pp. 301–304.

Kirkwood, Thomas B.L. (Feb. 2005). "Understanding the odd science of aging". *Cell* 120.4, pp. 437–447.

Kirkwood, Thomas B.L. and Steven N. Austad (Nov. 2000). "Why do we age?" *Nature* 408.6809, pp. 233–238.

Kirschner, Kristina et al. (May 2017). "Proliferation Drives Aging-Related Functional Decline in a Subpopulation of the Hematopoietic Stem Cell Compartment". *Cell Reports* 19.8, pp. 1503–1511.

Klutstein, Michael et al. (Feb. 2017). "Contribution of epigenetic mechanisms to variation in cancer risk among tissues". *Proceedings of the National Academy of Sciences of the United States of America* 114.9, pp. 2230–2234.

Koch, Carmen M. and Wolfgang Wagner (2011). "Epigenetic-aging-signature to determine age in different tissues". *Aging*.

Koch, Carmen M. et al. (Apr. 2012). "Monitoring of cellular senescence by DNA-methylation at specific CpG sites". *Aging Cell* 11.2, pp. 366–369.

Koche, Richard P. et al. (2011). "Reprogramming factor expression induces rapid and widespread targeted chromatin remodeling". *Cell Stem Cell*.

Kozlenkov, Alexey et al. (June 2017). "DNA Methylation Profiling of Human Prefrontal Cortex Neurons in Heroin Users Shows Significant Difference between Genomic Contexts of Hyper- and Hypomethylation and a Younger Epigenetic Age". *Genes* 8.6, pp. 2–15.

Krueger, Felix (2012). "Babraham Bioinformatics - Trim Galore!"

— (2016). *Bismark Bisulfite Mapper-User Guide-v0.15.0*. Tech. rep. Babraham Institute.

Krueger, Felix and Simon R. Andrews (June 2011). "Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications". *Bioinformatics* 27.11, pp. 1571–1572.

Kugel, Sita et al. (June 2016). "SIRT6 suppresses pancreatic cancer through control of Lin28b". *Cell* 165.6, pp. 1401–1415.

Kuhn, Max (Mar. 2020). "Classification and Regression Training [R package caret version 6.0-86]".

Kurosawa, Shuhei and Atsushi Iwama (Dec. 2020). "Aging and leukemic evolution of hematopoietic stem cells under various stress conditions". *Inflammation and Regeneration* 40.1, pp. 1–10.

Kurotaki, Naohiro et al. (Mar. 2002). "Haploinsufficiency of NSD1 causes Sotos syndrome". *Nature Genetics* 30.4, pp. 365–366.

Kuznetsova, A., P. B. Brockhoff, and R. H. Bojesen Christensen (Oct. 2016). "Tests in Linear Mixed Effects Models [R package lmerTest version 3.1-3]".

Lagouge, Marie et al. (Dec. 2006). "Resveratrol Improves Mitochondrial Function and Protects against Metabolic Disease by Activating SIRT1 and PGC-1α". *Cell* 127.6, pp. 1109–1122.

Lakshminarasimhan, Mahadevan et al. (Mar. 2013). "Sirt1 activation by resveratrol is substrate sequence-selective". *Aging* 5.3, pp. 151–154.

Lanza, Robert P. et al. (2000). "Extension of cell life-span and telomere length in animals cloned from senescent somatic cells". *Science*.

Lapasset, Laure et al. (Nov. 2011). "Rejuvenating senescent and centenarian human cells by reprogramming through the pluripotent state." *Genes & development* 25.21, pp. 2248–53.

Laplante, Mathieu and David M. Sabatini (Oct. 2009). "mTOR signaling at a glance". *Journal of Cell Science* 122.20, pp. 3589–3594.

Lau, Pui Yin and Wing Kam Fung (Nov. 2020). "Evaluation of marker selection methods and statistical models for chronological age prediction based on DNA methylation". *Legal Medicine* 47, p. 101744.

Ledford, Heidi (2013). "GSK absorbs controversial 'longevity' company : Nature News blog". *Nature News Blog*.

Lee, Jieun et al. (May 2020). "Induced pluripotency and spontaneous reversal of cellular aging in supercentenarian donor cells: Induced Pluripotent Stem Cells from a 114-year-old Supercentenarian". *Biochemical and Biophysical Research Communications* 525.3, pp. 563–569.

Lehallier, Benoit et al. (Dec. 2019). "Undulating changes in human plasma proteome profiles across the lifespan". *Nature Medicine* 25.12, pp. 1843–1850.

Lehmann, Ruth (June 2012). "Germline stem cells: Origin and destiny". *Cell Stem Cell* 10.6, pp. 729–739.

Lemon, J (2006). "Plotrix: a package in the red light district of R". *R-News* 6.4, pp. 8–12.

Levine, Morgan et al. (Nov. 2020). "A rat epigenetic clock recapitulates phenotypic aging and co-localizes with heterochromatin". *eLife* 9.

Levine, Morgan E (2013). "Modeling the Rate of Senescence: Can Estimated Biological Age Predict Mortality More Accurately Than Chronological Age?" *BioloGical ScienceS cite journal as: J Gerontol a Biol Sci Med Sci* 68.6, pp. 667–674.

Levine, Morgan E. et al. (2015). "Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning". *Aging* 7.12, pp. 1198–1211.

Levine, Morgan E. et al. (Aug. 2016). "Menopause accelerates biological aging". *Proceedings of the National Academy of Sciences of the United States of America* 113.33, pp. 9327–9332.

Levine, Morgan E. et al. (Apr. 2018). "An epigenetic biomarker of aging for lifespan and healthspan". *Aging* 10.4, pp. 573–591.

Levy, Joshua J. et al. (Mar. 2020). "MethylNet: An automated and modular deep learning approach for DNA methylation analysis". *BMC Bioinformatics* 21.1, p. 108.

Ley, Timothy J. et al. (Dec. 2010). "¡¿DNMT3A¡/¿ Mutations in Acute Myeloid Leukemia". *New England Journal of Medicine* 363.25, pp. 2424–2433.

Li, En, Timothy H. Bestor, and Rudolf Jaenisch (June 1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality". *Cell* 69.6, pp. 915–926.

Lin, Qiong et al. (2016). "DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy". *Aging* 8.2, pp. 394–401.

Lindsey, Janet et al. (Jan. 1991). "In vivo loss of telomeric repeats with age in humans". *Mutation Research DNAging* 256.1, pp. 45–48.

Lister, Ryan et al. (2013). "Global epigenomic reconfiguration during mammalian brain development". *Science*.

Liu, Baohua et al. (July 2005). "Genomic instability in laminopathy-based premature aging". *Nature Medicine* 11.7, pp. 780–785.

Liu, Liang et al. (2011). "Insufficient DNA methylation affects healthy aging and promotes age-related health problems". *Clinical Epigenetics*.

Loffredo, Francesco S. et al. (May 2013). "Growth differentiation factor 11 is a circulating factor that reverses age-related cardiac hypertrophy". *Cell* 153.4, pp. 828–839.

Longo, Valter D. and Mark P. Mattson (Feb. 2014). "Fasting: Molecular mechanisms and clinical applications". *Cell Metabolism* 19.2, pp. 181–192.

Lopez-Baez, Juan Carlos et al. (Feb. 2018). "Wilms Tumor 1b defines a wound-specific sheath cell subpopulation associated with notochord repair." *eLife* 7.

López-Otín, Carlos et al. (June 2013). "The hallmarks of aging". *Cell* 153.6, p. 1194.

Lowe, Donna, Steve Horvath, and Kenneth Raj (2016). "Epigenetic clock analyses of cellular senescence and ageing". *Oncotarget* 7.8, pp. 8524–8531.

Lowe, Robert et al. (Mar. 2020). "DNA methylation clocks as a predictor for ageing and age estimation in naked mole-rats, Heterocephalus glaber". *Aging* 12.5, pp. 4394–4406.

Lu, Ake T. et al. (Feb. 2016). "Genetic variants near MLST8 and DHX57 affect the epigenetic age of the cerebellum". *Nature Communications* 7.1, pp. 1–9.

Lu, Ake T. et al. (May 2017). "Genetic architecture of epigenetic and neuronal ageing rates in human brain regions". *Nature Communications* 8.1, pp. 1–14.

Lu, Ake T. et al. (Dec. 2018). "GWAS of epigenetic aging rates in blood reveals a critical role for TERT". *Nature Communications* 9.1, p. 387.

Lu, Ake T. et al. (Jan. 2019a). "DNA methylation GrimAge strongly predicts lifespan and healthspan". *Aging*.

Lu, Ake T. et al. (Aug. 2019b). "DNA methylation-based estimator of telomere length". *Aging* 11.16, pp. 5895–5923.

Lu, Yuancheng et al. (2020). "Reprogramming to recover youthful epigenetic information and restore vision". *Nature* 588.7836, pp. 124–129.

Lunnon, Katie et al. (Feb. 2016). "Variation in 5-hydroxymethylcytosine across human cortex and cerebellum". *Genome Biology* 17.1, p. 27.

MacFadyen, John R. et al. (May 2005). "Endosialin (TEM1, CD248) is a marker of stromal fibroblasts and is not selectively expressed on tumour endothelium". *FEBS Letters* 579.12, pp. 2569–2575.

Mackenzie, Karen J et al. (July 2017). "cGAS surveillance of micronuclei links genome instability to innate immunity." *Nature* 548.7668, pp. 461–465.

Maegawa, Shinji et al. (Mar. 2010). "Widespread and tissue specific age-related DNA methylation changes in mice". *Genome Research* 20.3, pp. 332–340.

Mahmoudi, Salah, Lucy Xu, and Anne Brunet (Jan. 2019). "Turning back time with emerging rejuvenation strategies". *Nature Cell Biology* 21.1, pp. 32–43.

Maierhofer, Anna et al. (2017). "Accelerated epigenetic aging in Werner syndrome". *Aging* 9.4, pp. 1143–1152.

Mallon, Barbara S. et al. (Jan. 2013). "StemCellDB: The Human Pluripotent Stem Cell Database at the National Institutes of Health". *Stem Cell Research* 10.1, pp. 57–66.

Manukyan, Maria and Prim B. Singh (May 2012). "Epigenetic rejuvenation". *Genes to Cells* 17.5, pp. 337–343.

— (May 2014). "Epigenome rejuvenation: HP1β mobility as a measure of pluripotent and senescent chromatin ground states". *Scientific Reports* 4.1, p. 4789.

Mao, Shi-Qing et al. (June 2020). "Genome-wide DNA Methylation Signatures Are Determined by DNMT3A/B Sequence Preferences". *Biochemistry*.

Marcus, Robert et al. (Feb. 1990). "Effects of Short Term Administration of Recombinant Human Growth Hormone to Elderly People*". *The Journal of Clinical Endocrinology & Metabolism* 70.2, pp. 519–527.

Marion, Rosa M. et al. (Feb. 2009). "Telomeres Acquire Embryonic Stem Cell Characteristics in Induced Pluripotent Stem Cells". *Cell Stem Cell* 4.2, pp. 141–154.

Marión, Rosa M. and Maria A. Blasco (Apr. 2010). "Telomere rejuvenation during nuclear reprogramming". *Current Opinion in Genetics and Development* 20.2, pp. 190–196.

Marioni, Riccardo E et al. (Jan. 2019). "Tracking the Epigenetic Clock Across the Human Life Course: A Meta-analysis of Longitudinal Cohort Data". *The Journals of Gerontology: Series A* 74.1, pp. 57–61.

Marioni, Riccardo E. et al. (Dec. 2015a). "DNA methylation age of blood predicts all-cause mortality in later life". *Genome Biology* 16.1, p. 25.

Marioni, Riccardo E. et al. (2015b). "The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936". *International Journal of Epidemiology*.

Marioni, Riccardo E. et al. (Apr. 2016). "The epigenetic clock and telomere length are independently associated with chronological age and mortality". *International Journal of Epidemiology* 45.2, pp. 424–432.

Martin-Herranz, Daniel E. et al. (Aug. 2019). "Screening for genes that accelerate the epigenetic aging clock in humans reveals a role for the H3K36 methyltransferase NSD1". *Genome Biology* 20.1, pp. 1–19.

Martínez, Daniel E. (May 1998). "Mortality patterns suggest lack of senescence in hydra". *Experimental Gerontology* 33.3, pp. 217–225.

Mayer, Wolfgang et al. (Feb. 2000). "Demethylation of the zygotic paternal genome". *Nature* 403.6769, pp. 501–502.

Mayne, Benjamin et al. (Dec. 2020). "A DNA methylation age predictor for zebrafish". *Aging* 12.

McCartney, Daniel L. et al. (Sept. 2018). "Epigenetic prediction of complex traits and death." *Genome biology* 19.1, p. 136.

McCartney, Daniel L. et al. (Dec. 2019). "An epigenome-wide association study of sex-specific chronological ageing". *Genome Medicine* 12.1.

McCay, C. M., M. F. Crowell, and L. A. Maynard (1935). "The effect of retarded growth upon the length of life span and upon the ultimate body size. 1935." *Nutrition (Burbank, Los Angeles County, Calif.)*

McCormick, Helen et al. (Dec. 2017). "Isogenic mice exhibit sexually-dimorphic DNA methylation patterns across multiple tissues". *BMC Genomics* 18.1.

McCrory, Cathal et al. (Nov. 2020). "GrimAge outperforms other epigenetic clocks in the prediction of age-related clinical phenotypes and all-cause mortality". *The Journals of Gerontology: Series A*.

McElhaney, Janet E. and Rita B. Effros (Aug. 2009). "Immunosenescence: what does it mean to health outcomes in older adults?" *Current Opinion in Immunology* 21.4, pp. 418–424.

McEwen, Lisa M. et al. (Oct. 2019). "The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells". *Proceedings of the National Academy of Sciences*, p. 201820843.

Medawar, P. B. (1952). "An Unsolved Problem of Biology". In: *The Uniqueness of the Individual*.

Meer, Margarita V et al. (Nov. 2018). "A whole lifespan mouse multi-tissue DNA methylation clock". *eLife* 7.

Meissner, Alexander et al. (2005). "Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis". *Nucleic Acids Research* 33.18, pp. 5868–5877.

Meissner, Alexander et al. (Aug. 2008a). "Genome-scale DNA methylation maps of pluripotent and differentiated cells". *Nature* 454.7205, pp. 766–770.

Meissner, Christoph et al. (Nov. 1999). "Estimation of age at death based on quantitation of the 4977-bp deletion of human mitochondrial DNA in skeletal muscle". *Forensic Science International* 105.2, pp. 115–124.

Meissner, Christoph et al. (July 2008b). "The 4977 bp deletion of mitochondrial DNA in human skeletal muscle, heart and different areas of the brain: A useful biomarker or more?" *Experimental Gerontology* 43.7, pp. 645–652.

Mercken, Evi M. et al. (Oct. 2014). "SRT2104 extends survival of male mice on a standard diet and preserves bone and muscle mass". *Aging Cell* 13.5, pp. 787–796.

Miles, Jeremy (Sept. 2014). "R Squared, Adjusted R Squared". In: *Wiley StatsRef: Statistics Reference Online*. Chichester, UK: John Wiley & Sons, Ltd.

Miller, Richard A. et al. (Feb. 2011). "Rapamycin, but not resveratrol or simvastatin, extends life span of genetically heterogeneous mice". *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 66 A.2, pp. 191–201.

Mitchell, Sarah J. et al. (2014). "The SIRT1 activator SRT1720 extends lifespan and improves health of mice fed a standard diet". *Cell Reports* 6.5, pp. 836–843.

Mizushima, Noboru et al. (Feb. 2008). "Autophagy fights disease through cellular self-digestion". *Nature* 451.7182, pp. 1069–1075.

Montano, Monty (Jan. 2014). "Translational models, methods and concepts in studies of aging and longevity". In: *Translational Biology in Medicine*. Elsevier, pp. 75–101.

Moradi, Sharif et al. (Nov. 2019). "Research and therapy with induced pluripotent stem cells (iPSCs): Social, legal, and ethical considerations". *Stem Cell Research and Therapy* 10.1, pp. 1–13.

Mostoslavsky, Raul et al. (Jan. 2006). "Genomic instability and aging-like phenotype in the absence of mammalian SIRT6". *Cell* 124.2, pp. 315–329.

Napolitano, Laura A. et al. (Mar. 2008). "Growth hormone enhances thymic function in HIV-1-infected adults". *Journal of Clinical Investigation* 118.3, pp. 1085–1098.

Neff, Frauke et al. (Aug. 2013). "Rapamycin extends murine lifespan but has limited effects on aging". *Journal of Clinical Investigation* 123.8, pp. 1–2.

Newman, John C. et al. (Sept. 2017). "Ketogenic Diet Reduces Midlife Mortality and Improves Memory in Aging Mice". *Cell Metabolism* 26.3, pp. 547–557.

Nilsson, Emma et al. (2014). "Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes". *Diabetes*.

Ocampo, Alejandro et al. (Dec. 2016). "In Vivo Amelioration of Age-Associated Hallmarks by Partial Reprogramming." *Cell* 167.7, pp. 1719–1733.

Ohnishi, Kotaro et al. (Feb. 2014). "Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation". *Cell* 156.4, pp. 663–677.

Ohnuki, Mari et al. (Aug. 2014). "Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential". *Proceedings of the National Academy of Sciences of the United States of America* 111.34, pp. 12426–12431.

Okano, Masaki et al. (Oct. 1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development". *Cell* 99.3, pp. 247–257.

Olova, N et al. (Nov. 2018a). "Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity." *Aging cell*.

Olova, Nelly et al. (Mar. 2018b). "Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data". *Genome Biology* 19.1, p. 33.

Opalach, Katherine et al. (Feb. 2010). "Lifelong calorie restriction alleviates age-related oxidative damage in peripheral nerves". *Rejuvenation Research* 13.1, pp. 65–74.

Orlanski, Shari et al. (May 2016). "Tissue-specific DNA demethylation is required for proper B-cell differentiation and function". *Proceedings of the National Academy of Sciences of the United States of America* 113.18, pp. 5018–5023.

Ou, Xue-ling et al. (Aug. 2012). "Predicting Human Age with Bloodstains by sjTREC Quantification". *PLoS ONE* 7.8. Ed. by Robert Lafrenie, e42412.

Pálovics, Róbert et al. (Nov. 2020). "Molecular hallmarks of heterochronic parabiosis at single cell resolution". *bioRxiv*, p. 2020.11.06.367078.

Paoli, Antonio and Gerardo Bosco (Jan. 2015). "The Ketogenic Mediterranean Diet". In: *The Mediterranean Diet: An Evidence-Based Approach*. Elsevier Inc., pp. 271–280. ISBN: 9780124079427.

Partridge, Linda and David Gems (2002). "Mechanisms of ageing: Public or private?" *Nature Reviews Genetics* 3.3, pp. 165–175.

Pearson, Kevin J. et al. (Aug. 2008). "Resveratrol Delays Age-Related Deterioration and Mimics Transcriptional Aspects of Dietary Restriction without Extending Life Span". *Cell Metabolism* 8.2, pp. 157–168.

Perna, Laura et al. (June 2016). "Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort". *Clinical Epigenetics* 8.1, p. 64.

Peters, Marjolein J. et al. (Oct. 2015). "The transcriptional landscape of age in human peripheral blood". *Nature Communications* 6.1, pp. 1–14.

Petkovich, Daniel A et al. (Apr. 2017). "Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions." *Cell metabolism* 25.4, pp. 954–960.

Pfluger, Paul T. et al. (July 2008). "Sirt1 protects against high-fat diet-induced metabolic damage". *Proceedings of the National Academy of Sciences of the United States of America* 105.28, pp. 9793–9798.

Phillip, Jude M. et al. (June 2017). "Biophysical and biomolecular determination of cellular age in humans". *Nature Biomedical Engineering* 1.7.

Phillips, David H. and Volker M. Arlt (2009). "Genotoxicity: damage to DNA and its consequences." *EXS* 99, pp. 87–110.

Pidsley, Ruth et al. (2013). "A data-driven approach to preprocessing Illumina 450K methylation array data". *BMC Genomics* 14.1.

Pidsley, Ruth et al. (Oct. 2016). "Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling". *Genome Biology* 17.1, p. 208.

Pilling, Darrell et al. (Oct. 2009). "Identification of Markers that Distinguish Monocyte-Derived Fibrocytes from Monocytes, Macrophages, and Fibroblasts". *PLoS ONE* 4.10. Ed. by Laurent Rénia, e7475.

Plana, Montserrat et al. (July 2011). "The reconstitution of the thymus in immunosuppressed individuals restores CD4-specific cellular and humoral immune responses". *Immunology* 133.3, pp. 318–328.

Polanowski, Andrea M. et al. (2014). "Epigenetic estimation of age in humpback whales". *Molecular Ecology Resources* 14.5, pp. 976–987.

Popat, Rakesh et al. (Mar. 2013). "A phase 2 study of SRT501 (resveratrol) with bortezomib for patients with relapsed and or refractory multiple myeloma". *British Journal of Haematology* 160.5, pp. 714–717.

Quach, Austin et al. (2017). "Epigenetic clock analysis of diet, exercise, education, and lifestyle factors". *Aging* 9.2, pp. 419–446.

Raddatz, Günter et al. (2013). "Aging is associated with highly defined epigenetic changes in the human epidermis". *Epigenetics and Chromatin*.

Rakyan, Vardhman K. et al. (Apr. 2010). "Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains". *Genome Research* 20.4, pp. 434–439.

Ramos, María-Paz et al. (Mar. 2015). "DNA demethylation by 5-aza-2-deoxycytidine is imprinted, targeted to euchromatin, and has limited transcriptional consequences". *Epigenetics & Chromatin 2015 8:1* 8.1, pp. 1–18.

Rando, Thomas A and Howard Y Chang (Jan. 2012). "Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock." *Cell* 148.1-2, pp. 46–57.

Rattanavirotkul, Nattaphong, Kristina Kirschner, and Tamir Chandra (Sept. 2020). "Induction and transmission of oncogene-induced senescence". *Cellular and Molecular Life Sciences* 1, p. 3.

Reddy, S. Sethu K. and Joumana T. Chaiban (2017). "The endocrinology of aging: A key to longevity "great expectations"". *Endocrine Practice* 23.9, pp. 1110–1119.

Redman, Leanne M. et al. (Apr. 2018). "Metabolic Slowing and Reduced Oxidative Damage with Sustained Caloric Restriction Support the Rate of Living and Oxidative Damage Theories of Aging". *Cell Metabolism* 27.4, pp. 805–815.

Reik, Wolf, W. Dean, and J. Walter (Aug. 2001). "Epigenetic reprogramming in mammalian development." *Science (New York, N.Y.)* 293.5532, pp. 1089–93.

Reizel, Yitzhak et al. (May 2015). "Gender-specific postnatal demethylation and establishment of epigenetic memory." *Genes & development* 29.9, pp. 923–33.

Reizel, Yitzhak et al. (Dec. 2018). "Postnatal DNA demethylation and its role in tissue maturation". *Nature Communications* 9.1, pp. 1–11.

Ren, Fu et al. (Sept. 2009). "Estimation of Human Age According to Telomere Shortening in Peripheral Blood Leukocytes of Tibetan". *The American Journal of Forensic Medicine and Pathology* 30.3, pp. 252–255.

Ridker, Paul M. et al. (2003). "C-reactive protein, the metabolic syndrome, and risk of incident cardiovascular events: an 8-year follow-up of 14 719 initially healthy American women." *Circulation*.

Rittié, Laure et al. (Dec. 2009). "Hedgehog signaling maintains hair follicle stem cell phenotype in young and aged human skin". *Aging Cell* 8.6, pp. 738–751.

Roberts, Megan N. et al. (Sept. 2017). "A Ketogenic Diet Extends Longevity and Healthspan in Adult Mice". *Cell Metabolism* 26.3, pp. 539–546.

Robertson, Neil A. et al. (Aug. 2019). "Age-related clonal haemopoiesis is associated with increased epigenetic age". *Current Biology* 29.16, R786–R787.

Rogina, Blanka and Stephen L. Helfand (Nov. 2004). "Sir2 mediates longevity in the fly through a pathway related to calorie restriction". *Proceedings of the National Academy of Sciences of the United States of America* 101.45, pp. 15998–16003.

Ross, Kharah M. et al. (Dec. 2020). "Epigenetic age and pregnancy outcomes: GrimAge acceleration is associated with shorter gestational length and lower birthweight". *Clinical Epigenetics* 12.1.

Ruckh, Julia M. et al. (Jan. 2012). "Rejuvenation of regeneration in the aging central nervous system". *Cell Stem Cell* 10.1, pp. 96–103.

Saeed, Maria, Rebecca M. Berlin, and Tracey Dawson Cruz (Nov. 2012). "Exploring the utility of genetic markers for predicting biological age". *Legal Medicine* 14.6, pp. 279–285.

Samaras, Nikolaos et al. (Aug. 2013). "A review of age-related dehydroepiandrosterone decline and its association with well-known geriatric syndromes: Is treatment beneficial?" *Rejuvenation Research* 16.4, pp. 285–294.

Sarkar, Tapash Jay et al. (Dec. 2020). "Transient non-integrative expression of nuclear reprogramming factors promotes multifaceted amelioration of aging in human cells". *Nature Communications* 11.1, pp. 1–12.

Satoh, Akiko et al. (Sept. 2013). "Sirt1 extends life span and delays aging in mice through the regulation of Nk2 Homeobox 1 in the DMH and LH". *Cell Metabolism* 18.3, pp. 416–430.

Schafer, Marissa J. et al. (June 2020). "The senescence-associated secretome as an indicator of age and medical risk". *JCI Insight* 5.12.

Schaible, Ralf et al. (Dec. 2015). "Constant mortality and fertility over age in Hydra". *Proceedings of the National Academy of Sciences of the United States of America* 112.51, pp. 15701–15706.

Schlessinger, Karni, Alan Hall, and Nicholas Tolwinski (Feb. 2009). "Wnt signaling pathways meet Rho GTPases". *Genes and Development* 23.3, pp. 265–277.

Schmeer et al. (Nov. 2019). "Dissecting Aging and Senescence—Current Concepts and Open Lessons". *Cells* 8.11, p. 1446.

Schultz, Michael B. and David A. Sinclair (Jan. 2016). "When stem cells grow old: Phenotypes and mechanisms of stem cell aging". *Development (Cambridge)* 143.1, pp. 3–14.

Sedelnikova, Olga A. et al. (2004). "Senescing human cells and ageing mice accumulate DNA lesions with unrepairable double-strand breaks". *Nature Cell Biology*.

Sendžikaitė, Gintarė et al. (Dec. 2019). "A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice". *Nature Communications* 10.1.

Shibata, D. (Jan. 2009). "Inferring human stem cell behaviour from epigenetic drift". *Journal of Pathology* 217.2, pp. 199–205.

Shiels, P. G. et al. (May 1999). "Analysis of telomere lengths in cloned sheep [9]". *Nature* 399.6734, pp. 316–317.

Shiloh, Yosef and Michael B. Kastan (Jan. 2001). "ATM: Genome stability, neuronal development, and cancer cross paths". *Advances in Cancer Research* 83, pp. 209–254.

Sierra, Marta, Agustín Fernández, and Mario Fraga (Sept. 2015). "Epigenetics of Aging". *Current Genomics* 16.6, pp. 435–440.

Simpkin, Andrew J. et al. (Jan. 2016). "Prenatal and early life influences on epigenetic age in children: A study of mother-offspring pairs from two cohort studies". *Human Molecular Genetics* 25.1, pp. 191–201.

Simpkin, Andrew J. et al. (2017). "The epigenetic clock and physical development during childhood and adolescence: Longitudinal analysis from a UK birth cohort". *International Journal of Epidemiology*.

Sinclair, K. D. et al. (July 2016). "Healthy ageing of cloned sheep". *Nature Communications* 7.1, pp. 1–10.

Singh, Param Priya et al. (2019). "Leading Edge The Genetics of Aging: A Vertebrate Perspective". *Cell* 177, pp. 200–220.

Singh, Prim B. and Fred Zacouto (2010). "Nuclear reprogramming and epigenetic rejuvenation". *Journal of Biosciences* 35.2, pp. 315–319.

Singh, Vimal K. et al. (Feb. 2015). "Induced pluripotent stem cells: Applications in regenerative medicine, disease modeling, and drug discovery". *Frontiers in Cell and Developmental Biology* 3.FEB, p. 2.

Singhal, Ram P., Laura L. Mays-Hoopes, and Gunther L. Eichhorn (1987). "DNA methylation in aging of mice". *Mechanisms of Ageing and Development*.

Sinha, Manisha et al. (May 2014). "Restoring systemic GDF11 levels reverses age-related dysfunction in mouse skeletal muscle". *Science* 344.6184, pp. 649–652.

Slieker, Roderick C. et al. (May 2018). "Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception". *Epigenetics and Chromatin* 11.1, p. 25.

Smith, Blair H et al. (June 2013). "Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness". *International Journal of Epidemiology* 42.3, pp. 689–700.

Smith, Blair H. et al. (Oct. 2006). "Generation Scotland: The Scottish Family Health Study; a new resource for researching genes and heritability". *BMC Medical Genetics* 7.1, p. 74.

Smith, Lucas K. et al. (Aug. 2015). "β2-microglobulin is a systemic pro-aging factor that impairs cognitive function and neurogenesis". *Nature Medicine* 21.8, pp. 932–937.

Smith, Zachary D., Camille Sindhu, and Alexander Meissner (Mar. 2016). "Molecular features of cellular reprogramming and development". *Nature Reviews Molecular Cell Biology* 17.3, pp. 139–154.

Snir, Sagi, Colin Farrell, and Matteo Pellegrini (Sept. 2019). "Human epigenetic ageing is logarithmic with time across the entire lifespan". *Epigenetics* 14.9, pp. 912–926.

Snir, Sagi, Bridgett M. VonHoldt, and Matteo Pellegrini (Nov. 2016). "A Statistical Framework to Identify Deviation from Time Linearity in Epigenetic Aging". *PLoS Computational Biology* 12.11, p. 1005183.

Soriano-Tárraga, Carolina et al. (2013). "DNA Isolation Method Is a Source of Global DNA Methylation Variability Measured with LUMA. Experimental Analysis and a Systematic Review". *PLoS ONE* 8.4, pp. 1–8.

Soukas, Alexander A., Haibin Hao, and Lianfeng Wu (Oct. 2019). "Metformin as Anti-Aging Therapy: Is It for Everyone?" *Trends in Endocrinology and Metabolism* 30.10, pp. 745–755.

Stadtfeld, Matthias et al. (Mar. 2008). "Defining Molecular Cornerstones during Fibroblast to iPS Cell Reprogramming in Mouse". *Cell Stem Cell* 2.3, pp. 230–240.

Starnawska, A. et al. (Sept. 2019). "Epigenome-wide association study of depression symptomatology in elderly monozygotic twins". *Translational Psychiatry* 9.1, pp. 1–14.

Stubbs, Thomas M. et al. (Dec. 2017). "Multi-tissue DNA methylation age predictor in mouse". *Genome Biology* 18.1, p. 68.

Sugrue, V J et al. (Nov. 2020). "Castration delays epigenetic aging and feminises DNA methylation at androgen-regulated loci". *bioRxiv*, p. 2020.11.16.385369.

Sun, Deqiang et al. (2014). "Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal". *Cell Stem Cell*.

Sung, Li Ying et al. (Dec. 2014). "Telomere Elongation and Naive Pluripotent Stem Cells Achieved from Telomerase Haplo-Insufficient Cells by Somatic Cell Nuclear Transfer". *Cell Reports* 9.5, pp. 1603–1609.

Takahashi, Kazutoshi and Shinya Yamanaka (2006). "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors". *Cell* 126.4, pp. 663–676.

— (Mar. 2016). "A decade of transcription factor-mediated reprogramming to pluripotency". *Nature Reviews Molecular Cell Biology* 17.3, pp. 183–193.

Takahashi, Kazutoshi et al. (Nov. 2007). "Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors". *Cell* 131.5, pp. 861–872.

Takeda, Yukimasa et al. (May 2018). "Chemical compound-based direct reprogramming for future clinical applications". *Bioscience Reports* 38.3.

Tanabe, Koji et al. (July 2013). "Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts". *Proceedings of the National Academy of Sciences of the United States of America* 110.30, pp. 12172–12179.

Tanaka, Toshiko et al. (Oct. 2018). "Plasma proteomic signature of age in healthy humans." *Aging cell* 17.5, e12799.

Teschendorff, Andrew E. et al. (Apr. 2010). "Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer". *Genome Research* 20.4, pp. 440–446.

Teschendorff, Andrew E. et al. (July 2015). "Correlation of smoking-associated DNA methylation changes in buccal cells with DNA methylation changes in epithelial cancer". *JAMA Oncology* 1.4, pp. 476–485.

Thermo Fisher (2020). "SNaPshot™ Multiplex Kit".

Thompson, Michael J et al. (Oct. 2018). "A multi-tissue full lifespan epigenetic clock for mice." *Aging* 10.10, pp. 2832–2854.

Thompson, Michael J. et al. (2017). "An epigenetic aging clock for dogs and wolves". *Aging* 9.3, pp. 1055–1068.

Tian, X. C., J. Xu, and X. Yang (2000). "Normal telomere lengths found in cloned cattle". *Nature Genetics*.

Tibshirani, Robert (1997). "The lasso method for variable selection in the cox model". *Statistics in Medicine*.

Tomasetti, Cristian and Bert Vogelstein (Jan. 2015). "Variation in cancer risk among tissues can be explained by the number of stem cell divisions". *Science* 347.6217, pp. 78–81.

Trapp, Alexandre, Csaba Kerepesi, and Vadim N. Gladyshev (Mar. 2021). "Profiling epigenetic age in single cells". *bioRxiv*, p. 2021.03.13.435247.

Travers, Mary E. et al. (2013). "Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets". *Diabetes*.

Tsai, Robert Y.L. (May 2016). "Balancing self-renewal against genome preservation in stem cells: How do they manage to have the cake and eat it too?" *Cellular and Molecular Life Sciences* 73.9, pp. 1803–1823.

Tsuji, Akiko et al. (May 2002). "Estimating age of humans based on telomere shortening". *Forensic Science International* 126.3, pp. 197–199.

Tsygankov, Denis et al. (Sept. 2009). "A quantitative model for age-dependent expression of the p16INK4a tumor suppressor". *Proceedings of the National Academy of Sciences of the United States of America* 106.39, pp. 16562–16567.

Unnikrishnan, Archana et al. (Apr. 2018). "Revisiting the genomic hypomethylation hypothesis of aging". *Annals of the New York Academy of Sciences* 1418.1, pp. 69–79.

Vanyushin, B.F. et al. (1973). "The 5-Methylcytosine in DNA of Rats". *Gerontology* 19.3, pp. 138–152.

Vetter, Valentin Max et al. (2020). "Epigenetic Clock and Leukocyte Telomere Length are Associated with Vitamin D Status, but not with Functional Assessments and Frailty in the Berlin Aging Study II". *The Journals of Gerontology: Series A*.

Vijayakumar, Archana, Shoshana Yakar, and Derek LeRoith (2011). "The intricate role of growth hormone in metabolism". *Frontiers in Endocrinology* 2.SEP.

Vilkaitis, Giedrius et al. (Jan. 2005). "Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase". *Journal of Biological Chemistry* 280.1, pp. 64–72.

Villeda, Saul A. et al. (Sept. 2011). "The ageing systemic milieu negatively regulates neurogenesis and cognitive function". *Nature* 477.7362, pp. 90–96.

Villeda, Saul A. et al. (May 2014). "Young blood reverses age-related impairments in cognitive function and synaptic plasticity in mice". *Nature Medicine* 20.6, pp. 659–663.

Vitale, Giovanni et al. (Feb. 2019). "ROLE of IGF-1 system in the modulation of longevity: Controversies and new insights from a centenarians' perspective". *Frontiers in Endocrinology* 10.FEB, p. 27.

Vogel, Hannes et al. (Sept. 1999). "Deletion of Ku86 causes early onset of senescence in mice". *Proceedings of the National Academy of Sciences of the United States of America* 96.19, pp. 10770–10775.

Waddington, C. H. (1957). *The strategy of the genes: A discussion of some aspects of theoretical biology*. ISBN: 9781317657552.

Wagner, Wolfgang (June 2017). "Epigenetic aging clocks in mice and men". *Genome Biology* 18.1, p. 107.

Wakayama, Teruhiko et al. (2000). "Cloning of mice to six generations". *Nature*.

Walker, Richard F. et al. (2015). "Epigenetic age analysis of children who seem to evade aging". *Aging* 7.5, pp. 334–339.

Wallace, Susan S. (July 2014). "Base excision repair: A critical player in many games". *DNA Repair* 19, pp. 14–26.

Walter, Dagmar et al. (Apr. 2015). "Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells". *Nature* 520.7548, pp. 549–552.

Wang, Chunfang et al. (2009). "DNA damage response and cellular senescence in tissues of aging mice". *Aging Cell.*

Wang, Eugenia (1995). "Senescent Human Fibroblasts Resist Programmed Cell Death, and Failure to Suppress bcl2 Is Involved". *Cancer Research* 55.11.

Wang, Tina et al. (Dec. 2017). "Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment". *Genome Biology* 18.1, p. 57.

Wang, Tina et al. (2020). "Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome". *Cell Systems* 11.2, pp. 176–185.

Wang, Yamei et al. (Sept. 2006). "C. elegans 14-3-3 proteins regulate life span and interact with SIR-2.1 and DAF-16/FOXO". *Mechanisms of Ageing and Development* 127.9, pp. 741–747.

Watanabe, Ryosuke, Liu Wei, and Jing Huang (Apr. 2011). "MTOR signaling, function, novel inhibitors, and therapeutic targets". *Journal of Nuclear Medicine* 52.4, pp. 497–500.

Wei, Min et al. (Feb. 2017). "Fasting-mimicking diet and markers/risk factors for aging, diabetes, cancer, and cardiovascular disease". *Science Translational Medicine* 9.377.

Wei, Siying et al. (Dec. 2018). "Intermittent administration of a fasting-mimicking diet intervenes in diabetes progression, restores β cells and reconstructs gut microbiota in mice". *Nutrition & Metabolism* 15.1, p. 80.

Weidner, Carola I. et al. (Feb. 2014). "Aging of blood can be tracked by DNA methylation changes at just three CpG sites". *Genome Biology* 15.2, R24.

Weindruch, R. et al. (Apr. 1986). "The retardation of aging in mice by dietary restriction: Longevity, cancer, immunity and lifetime energy intake". *Journal of Nutrition* 116.4, pp. 641–6541.

Weindruch, Richard et al. (Mar. 1988). "Influences of Aging and Dietary Restriction on Serum Thymosinαl Levels in Mice". *Journal of Gerontology* 43.2, B40–B42.

Weismann, August (1889). *Essays upon Heredity and Kindred Biological Problems*. Volume 1. Oxford: Clarendon Press.

Wernig, Marius et al. (Aug. 2008). "A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types". *Nature Biotechnology* 26.8, pp. 916–924.

West, James et al. (Apr. 2013). "An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways". *Scientific Reports* 3.

West, Michael D et al. (2019). "Toward a unified theory of aging and regeneration".

White, Ryan R. and Jan Vijg (Sept. 2016). "Do DNA Double-Strand Breaks Drive Aging?" *Molecular Cell* 63.5, pp. 729–738.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*.

Wickham, Hadley et al. (2021). *dplyr: A Grammar of Data Manipulation*.

Williams, George C. (Dec. 1957). "Pleiotropy, Natural Selection, and the Evolution of Senescence". *Evolution* 11.4, p. 398.

Wilmut, I. et al. (Feb. 1997). "Viable offspring derived from fetal and adult mammalian cells". *Nature* 385.6619, pp. 810–813.

Wilson, V. L. et al. (July 1987). "Genomic 5-methyldeoxycytidine decreases with age." *Journal of Biological Chemistry* 262.21, pp. 9948–9951.

Wolf, Erika J. et al. (Jan. 2016). "Accelerated DNA methylation age: Associations with PTSD and neural integrity". *Psychoneuroendocrinology* 63, pp. 155–162.

Wolffe, A. P., P. L. Jones, and P. A. Wade (May 1999). "DNA demethylation". *Proceedings of the National Academy of Sciences of the United States of America* 96.11, pp. 5894–5896.

Wood, Jazon G. et al. (Aug. 2004). "Sirtuin activators mimic caloric restriction and delay ageing in metazoans". *Nature* 430.7000, pp. 686–689.

Wu, Xiaohui et al. (Dec. 2019). "Effect of tobacco smoking on the epigenetic age of human respiratory organs". *Clinical Epigenetics* 11.1, p. 183.

Xiao, Fu Hui, Hao Tian Wang, and Qing Peng Kong (2019). "Dynamic DNA methylation during aging: A "prophet" of age-related outcomes". *Frontiers in Genetics* 10.feburay.

Xiao, Fu Hui et al. (Mar. 2015). "A genome-wide scan reveals important roles of DNA methylation in human longevity by regulating age-related disease genes e0120388". *PLoS ONE* 10.3.

Xie, Wenbing, Stephen B. Baylin, and Hariharan Easwaran (Jan. 2019). "DNA methylation in senescence, aging and cancer". *Oncoscience* 6.1-2, p. 291.

Xie, Wenbing et al. (Feb. 2018). "DNA Methylation Patterns Separate Senescence from Transformation Potential and Indicate Cancer Risk". *Cancer Cell* 33.2, pp. 309–321.

Xu, Cheng et al. (Dec. 2015). "A novel strategy for forensic age prediction by DNA methylation and support vector regression model". *Scientific Reports* 5.1, pp. 1–10.

Yang, Ruoting et al. (May 2020). "A DNA methylation clock associated with age-related illnesses and mortality is accelerated in men with combat PTSD". *Molecular Psychiatry*, pp. 1–11.

Yang, Zhen et al. (Oct. 2016). "Correlation of an epigenetic mitotic clock with cancer risk". *Genome Biology* 17.1, p. 205.

Yasuda, T. et al. (Feb. 2014). "Leukemic evolution of donor-derived cells harboring IDH2 and DNMT3A mutations after allogeneic stem cell transplantation". *Leukemia* 28.2, pp. 426–428.

Yin, Yimeng et al. (May 2017). "Impact of cytosine methylation on DNA binding specificities of human transcription factors". *Science* 356.6337.

Yu, Ming et al. (Feb. 2020). "Epigenetic aging: More than just a clock when it comes to cancer". *Cancer Research* 80.3, pp. 367–374.

Yuan, Tian et al. (2015). "An Integrative Multi-scale Analysis of the Dynamic DNA Methylation Landscape in Aging". *PLoS Genetics* 11.2, pp. 1–21.

Zakrzewski, Wojciech et al. (Dec. 2019). "Stem cells: past, present, and future". *Stem Cell Research & Therapy* 10.1, p. 68.

Zane, Linda, Vivek Sharma, and Tom Misteli (2014). "Common features of chromatin in aging and cancer: Cause or coincidence?" *Trends in Cell Biology* 24.11, pp. 686–694.

Zbieć-Piekarska, Renata et al. (2015). "Development of a forensically useful age prediction method based on DNA methylation analysis". *Forensic Science International: Genetics* 17.February 2016, pp. 173–179.

Zhang, Jinglu et al. (May 2016). "AGE-induced keratinocyte MMP-9 expression is linked to TET2-mediated CpG demethylation". *Wound Repair and Regeneration* 24.3, pp. 489–500.

Zhang, Qian et al. (Dec. 2019). "Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing". *Genome Medicine* 11.1, p. 54.

Zhang, Yan et al. (Mar. 2017a). "DNA methylation signatures in peripheral blood strongly predict all-cause mortality". *Nature Communications* 8, p. 14617.

Zhang, Yan et al. (2017b). "Individual CpG sites that are associated with age and life expectancy become hypomethylated upon aging". *Clinical Epigenetics*.

Zheng, Shijie C., Martin Widschwendter, and Andrew E. Teschendorff (May 2016). "Epigenetic drift, epigenetic clocks and cancer risk". *Epigenomics* 8.5, pp. 705–719.

Zhou, Linli et al. (July 2016). "Dermal fibroblasts induce cell cycle arrest and block epithelial–mesenchymal transition to inhibit the early stage melanoma development". *Cancer Medicine* 5.7, pp. 1566–1579.

Zhu, Yi et al. (2015). "The achilles' heel of senescent cells: From transcriptome to senolytic drugs". *Aging Cell* 14.4, pp. 644–658.

Zoncu, Roberto, Alejo Efeyan, and David M. Sabatini (Jan. 2011). "MTOR: From growth signal integration to cancer, diabetes and ageing". *Nature Reviews Molecular Cell Biology* 12.1, pp. 21–35.

Zou, Hui and Trevor Hastie (Apr. 2005). "Regularization and variable selection via the elastic net". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320.

ZS.-Nagy, Imre, Richard G. Culter, and Imre Semesi (Mar. 1988). "Dysdifferentiation Hypothesis of Aging and Cancer: A Comparison with the Membrane Hypothesis of Aging". *Annals of the New York Academy of Sciences* 521.1 Neuroimmunomo, pp. 215–225.

Zubakov, D. et al. (Nov. 2010). "Estimating human age from T-cell DNA rearrangements". *Current Biology* 20.22, R970–R971.

Zullo, Alberto et al. (Apr. 2018). "Sirtuins as mediator of the anti-ageing effects of calorie restriction in skeletal and cardiac muscle". *International Journal of Molecular Sciences* 19.4.

Appendix

**Appendix 1: cGAS surveillance of micronuclei links
genome instability to innate immunity**

LETTER

cGAS surveillance of micronuclei links genome instability to innate immunity

Karen J. Mackenzie[1]*, Paula Carroll[1]*, Carol-Anne Martin[1], Olga Murina[1], Adeline Fluteau[1], Daniel J. Simpson[1], Nelly Olova[1], Hannah Sutcliffe[1], Jacqueline K. Rainger[1], Andrea Leitch[1], Ruby T. Osborn[1], Ann P. Wheeler[1], Marcin Nowotny[2], Nick Gilbert[1], Tamir Chandra[1], Martin A. M. Reijns[1] & Andrew P. Jackson[1]

DNA is strictly compartmentalized within the nucleus to prevent autoimmunity[1]; despite this, cyclic GMP–AMP synthase (cGAS), a cytosolic sensor of double-stranded DNA, is activated in autoinflammatory disorders and by DNA damage[2–6]. Precisely how cellular DNA gains access to the cytoplasm remains to be determined. Here, we report that cGAS localizes to micronuclei arising from genome instability in a mouse model of monogenic autoinflammation, after exogenous DNA damage and spontaneously in human cancer cells. Such micronuclei occur after mis-segregation of DNA during cell division and consist of chromatin surrounded by its own nuclear membrane. Breakdown of the micronuclear envelope, a process associated with chromothripsis[7], leads to rapid accumulation of cGAS, providing a mechanism by which self-DNA becomes exposed to the cytosol. cGAS is activated by chromatin, and consistent with a mitotic origin, micronuclei formation and the proinflammatory response following DNA damage are cell-cycle dependent. By combining live-cell laser microdissection with single cell transcriptomics, we establish that interferon-stimulated gene expression is induced in micronucleated cells. We therefore conclude that micronuclei represent an important source of immunostimulatory DNA. As micronuclei formed from lagging chromosomes also activate this pathway, recognition of micronuclei by cGAS may act as a cell-intrinsic immune surveillance mechanism that detects a range of neoplasia-inducing processes.

DNA is a key pathogen-associated molecular pattern that is sensed by innate immune receptors in the cytosol and endosomal compartments[8], so strict compartmentalization of cellular DNA in the nucleus and in mitochondria is necessary to avoid sensing of self-DNA[1]. cGAS is an important cytosolic nucleic acid sensor, and double-stranded DNA (dsDNA) is its canonical ligand[9,10]. cGAS activation generates the cyclic dinucleotide cyclic GMP–AMP (cGAMP), which in turn induces a type I interferon response via the adaptor STING (stimulator of interferon genes)[11]. Aberrant recognition of immunostimulatory cytosolic DNA has been implicated in neoplasia and systemic autoinflammatory diseases[12–14], with cGAS- or STING-dependent inflammation associated with mutations in multiple nucleases[15].

One such nuclease, RNase H2, maintains mammalian genome integrity through its role in ribonucleotide excision repair[16], suggesting that endogenous DNA damage may generate the nucleic acid ligands that are sensed by cGAS. Notably, micronuclei occur at a high frequency in *Rnaseh2b*−/− *Trp53*−/− mouse embryonic fibroblasts (MEFs) compared with *Rnaseh2b*+/+ *Trp53*−/− MEFs[16] (Fig. 1a; hereafter referred to as *Rnaseh2b*−/− and *Rnaseh2b*+/+ MEFs, respectively). This led us to consider micronuclei as a potential source of immunostimulatory DNA. Such micronuclei, which have their own nuclear envelope (Fig. 1b), arise during mitosis from lagging chromosomal DNA and chromatin bridges, as a consequence of unresolved genome

instability (Fig. 1c, Supplementary Video 1, Extended Data Fig. 1a, b and Supplementary Information). Increased micronuclei formation was also observed in *Rnaseh2b*A174T/A174T mice (*P* = 0.003; Extended Data Fig. 1c, d), a model for the autoinflammatory disorder Aicardi-Goutières syndrome, confirming that micronuclei arise as a result of RNase H2 deficiency both *in vitro* and *in vivo*, irrespective of p53 status. As the interferon-stimulated gene (ISG) upregulation and proinflammatory response in both *Rnaseh2b*−/− MEFs and *Rnaseh2b*A174T/A174T mice is cGAS- and STING-dependent[5], accumulation of micronuclear DNA correlated with activation of the cGAS and STING pathway. Furthermore, investigation of the subcellular localization of cGAS in *Rnaseh2b*−/− MEFs stably expressing GFP–cGAS established that cGAS was strongly enriched in micronuclei (Fig. 1d; 83.3 ± 1.4% of micronuclei were GFP–cGAS-positive), whereas GFP alone showed no such localization (Extended Data Fig. 1e, f), consistent with cGAS binding to micronuclear DNA.

To determine whether localization of cGAS to micronuclei was a general phenomenon, exogenous DNA damage was induced in GFP–cGAS-expressing MEFs. After 1 Gy irradiation, we observed frequent localization of cGAS to micronuclei (Fig. 1e), along with a cGAS-dependent proinflammatory response. Increased secretion of CCL5 (a robust indicator of cGAS-dependent ISG responses in MEFs[5]) correlated with increased frequency of micronuclei in both *Trp53*+/+ and *Trp53*−/− MEFs (Fig. 1f–h). Furthermore, consistent with an increased tendency to form micronuclei in *Trp53*−/− cells, following irradiation both micronucleus formation (*P* = 0.0078) and CCL5 production (*P* = 0.020) were significantly enhanced compared with *Trp53*+/+ cells. ISG transcripts were also induced (Extended Data Fig. 1g) at levels comparable to those found in previous studies of genotoxic damage[2,4].

In human U2OS osteosarcoma epithelial cells, endogenous cGAS was detected by immunofluorescence in spontaneously formed micronuclei (Extended Data Fig. 1h–j). Strong micronuclear enrichment of cGAS contrasted with weak diffuse cytoplasmic localization in cells without micronuclei, consistent with endogenous relocalization of cGAS to micronuclei. We therefore conclude that cGAS frequently localizes to micronuclei, irrespective of the source of DNA damage initiating their formation. However, given that a nuclear envelope normally encloses micronuclei, it was not clear how cGAS gains access to these structures.

Micronuclear DNA is particularly susceptible to DNA damage, leading to chromothripsis[7,17]. This occurs as a consequence of irreversible nuclear envelope collapse, which arises frequently in micronuclei due to defective nuclear lamina organization[18]. Given that this leads to partial loss of compartmentalization, we postulated that membrane rupture would also result in relocalization of cGAS to micronuclei, to induce a cGAMP-driven proinflammatory response (Fig. 2a). Consistent with this, we observed that micronuclei positive for γH2AX,
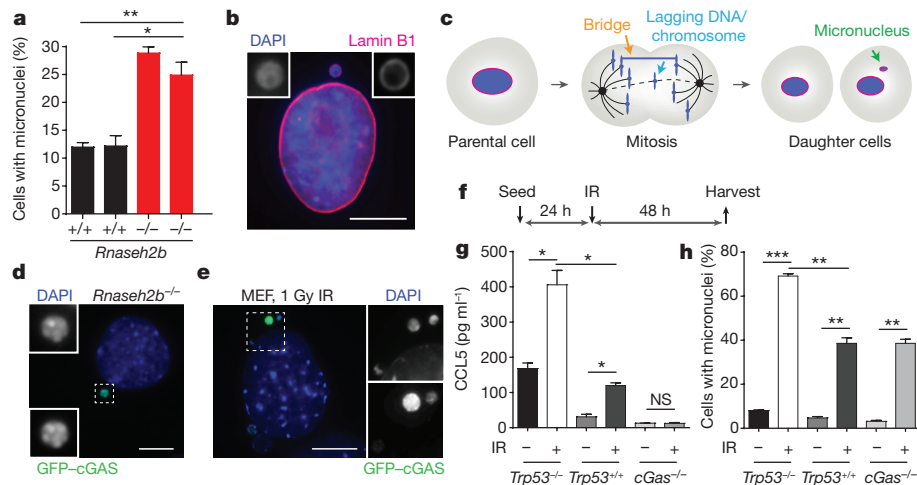
Figure 1 | cGAS localizes to micronuclei resulting from endogenous or exogenous DNA damage. a, Micronuclei form frequently in *Rnaseh2b*[−/−] MEFs and are associated with genome instability. Percentage of cells with micronuclei in two *Rnaseh2b*[+/+] control and two *Rnaseh2b*[−/−] MEF lines. Mean ± s.e.m. of $n = 3$ independent experiments (≥500 cells counted per line). **b**, Micronuclear DNA is surrounded by its own nuclear envelope. Representative image with Lamin B1 (red) staining the nuclear envelope and DAPI (blue) staining DNA. **c**, Micronuclei form after mitosis as a consequence of impaired segregation of DNA during mitosis, originating from chromatin bridges and lagging chromosomes or chromatin fragments. **d**, GFP–cGAS localizes to micronuclei in *Rnaseh2b*[−/−] MEFs. Representative image of GFP–cGAS-expressing *Rnaseh2b*[−/−] MEFs.

e–h, cGAS localizes to micronuclei induced by ionizing radiation and is associated with a cGAS-dependent proinflammatory response. **e**, Representative image of GFP–cGAS-positive micronuclei following 1 Gy irradiation (IR) in *Trp53*[−/−] MEFs. **f**, *Trp53*[−/−], *Trp53*[+/+] and *cGAS*[−/−] (also known as *Mb21d*[−/−]) (cGAS null) MEFs were irradiated (1 Gy), and CCL5 production (**g**) and percentage of cells with micronuclei (**h**) assessed after 48 h. Mean ± s.e.m. of $n = 2$ independent experiments. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, two-tailed *t*-test; NS, not significant. Scale bars, 10 μm. *Rnaseh2b*[+/+] and *Rnaseh2b*[−/−] MEFs in this figure and subsequent figures are on a *Trp53*[−/−] C57BL/6J background (absence of p53 is a prerequisite for generation of *Rnaseh2b*[−/−] MEFs[16]).

a marker of DNA damage, contained cGAS more often than those without γH2AX staining ($P = 0.0169$, *Rnaseh2b*[−/−] MEFs; $P = 0.0005$, U2OS cells; Extended Data Fig. 2a–c). Moreover, co-staining of U2OS cells for cGAS and the nuclear protein retinoblastoma (Rb), a marker for micronuclei with intact nuclear envelopes[18], demonstrated that the majority of cGAS-positive micronuclei had ruptured nuclear envelopes (Fig. 2b, c). Loss of an mCherry-tagged nuclear localization

signal (mCherry–NLS) was also strongly associated with localization of cGAS to micronuclei (Extended Data Fig. 3a, b), further supporting a link between membrane integrity failure and cGAS relocalization. To establish a direct temporal relationship between nuclear envelope rupture and cGAS relocalization, we performed live imaging in U2OS cells expressing mCherry–NLS and GFP–cGAS (Fig. 2d, Extended Data Fig. 3c). cGAS entered the micronuclei rapidly after loss of membrane
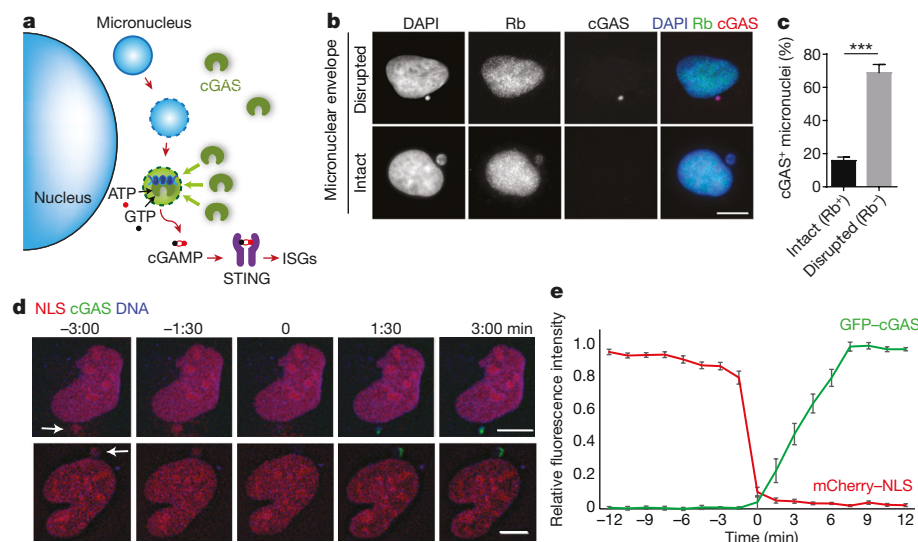
Figure 2 | cGAS localizes to micronuclei upon nuclear envelope rupture. a, Model: micronuclear membrane rupture leads to sensing of DNA by cGAS. Micronuclei are susceptible to nuclear envelope collapse, which permits cytosolic cGAS access to genomic dsDNA, initiating a cGAS–STING-dependent proinflammatory immune response through production of the second messenger cGAMP. **b, c**, Localization of cGAS to micronuclei in U2OS cells inversely correlates with localization of Rb, which is present only in micronuclei with an intact nuclear envelope. **b**, Representative images. **c**, Quantification (mean ± s.e.m. of $n = 3$ independent experiments; ≥250 micronuclei counted per experiment).

cGAS[+], cGAS-stained micronuclei; Rb[+] and Rb[−], micronuclei positive or negative for Rb staining, respectively. ***$P < 0.001$, two-tailed *t*-test. **d**, Representative stills from live imaging of U2OS cells expressing mCherry–NLS and GFP–cGAS. DNA visualized with Hoechst stain. Time (min) relative to loss of mCherry–NLS from micronucleus ($t = 0$, micronuclear membrane rupture). Arrows indicate micronuclei undergoing rupture. **e**, Quantification of cGAS signal accumulating in micronuclei after loss of nuclear envelope integrity. Relative mean fluorescence intensity plotted. Error bars show s.e.m. of $n = 11$ micronuclei. Scale bars, 10 μm.
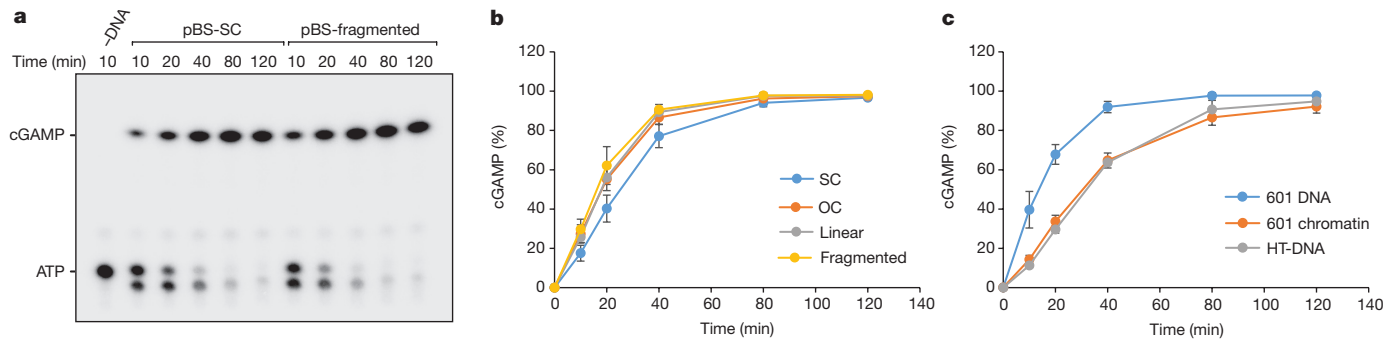
Figure 3 | Continuous and chromatinized DNA activate cGAS.
a, b, Both supercoiled (SC) and fragmented pBluescript (pBS) activate recombinant cGAS to produce cGAMP. **a**, Representative image of thin layer chromatography (TLC) detection of cGAMP. **b**, Quantification of cGAMP measured by TLC over time demonstrates no significant difference in cGAS activation by open circle (OC), linear or fragmented plasmid DNA, with supercoiled DNA showing a slight reduction in cGAMP production. Mean ± s.d., $n = 3$ independent experiments. **c**, Synthetic chromatin activates cGAS at the same level as herring testis (HT) DNA, but slightly less than the corresponding naked 601 DNA. Quantification of cGAMP measured by TLC. Mean ± s.d., $n = 3$ experiments.

integrity, as measured by loss of micronuclear mCherry–NLS (Fig. 2e, Supplementary Video 2). We therefore conclude that nuclear envelope rupture results in exposure of DNA to the cytoplasmic compartment, leading to relocalization of cGAS to micronuclear chromatin.

Activation of cGAS requires dimerization, with each cGAS monomer binding a dsDNA molecule[19]. Combined, these requirements might therefore preclude activation of cGAS that is bound to chromatin. In addition, a published model suggests that there would be steric clashes between bound DNA molecules and predicts that cGAS activation will therefore occur near the ends of dsDNA[20]. However, we reasoned that the flexible nature of DNA could permit cGAS dimerization on continuous DNA and chromatin, given the accessibility of linker DNA between nucleosomes. To test whether cGAS can be activated by DNA that does not contain free ends, we measured production of cGAMP by recombinant cGAS in the presence of plasmid DNA using a chromatography-based assay and found that cGAMP production was similar in the presence of either circular DNA or fragmented plasmid DNA (Fig. 3a, b, Extended Data Fig. 4a, b), establishing that DNA ends are not required for activation of cGAS. Furthermore, supercoiled plasmid DNA induced strong cGAS-dependent CCL5 production in MEFs (Extended Data Fig. 4c). We then prepared synthetic chromatin and found that cGAS can also bind to DNA in the presence of nucleosomes (Extended Data Fig. 5a, b) and that this leads to substantial cGAMP production (Fig. 3c, Extended Data Fig. 5c). Chromatin isolated from cells also activated recombinant cGAS (Extended Data Fig. 5d–g). We therefore conclude that cGAS can bind to and be activated by chromatin.

Small DNA fragments detected by antibodies against dsDNA have been proposed to leak from sites of DNA damage through the interphase nuclear envelope into the cytoplasm and to activate the cGAS–STING pathway[2,21]. In contrast, as micronuclei are generated at mitosis, a prediction of our model is that the immune response will be cell-cycle dependent. To test this prediction, we induced DNA damage in MEFs arrested in G0 by serum starvation. Such cell-cycle-arrested MEFs still displayed functional cGAS signalling, producing similar amounts of cytokine in response to transfected exogenous DNA compared with actively cycling cells (Extended Data Fig. 6a, b). However, cell-cycle-arrested MEFs did not form micronuclei, nor did they exhibit innate immune activation after exposure to equivalent levels of ionizing radiation, despite undergoing equal levels of DNA damage (Fig. 4a–d, Extended Data Fig. 6d, e). Therefore, DNA damage is not sufficient by itself to generate innate immune activation in response to ionizing radiation; with the dependence on cell-cycle progression consistent with mis-segregated DNA at mitosis being the origin of cGAS activation. Quantification of the levels of micronuclear DNA (Extended Data Fig. 7) indicates that it would be sufficient to generate a relevant cytokine response, also supporting micronuclei as an important source of cell-intrinsic immunostimulatory DNA.

We also investigated whether micronuclei initiated by a DNA damage-independent mechanism resulted in similar cGAS relocalization and an associated proinflammatory response. Micronuclei containing whole chromosomes were generated through pharmacological induction of lagging chromosomes by nocodazole treatment[7,18] (Extended Data Fig. 8a, b). This resulted in a substantially increased frequency of micronuclei in both MEFs and U2OS cells (Extended Data Fig. 8c, d). The number of cGAS-positive micronucleated cells
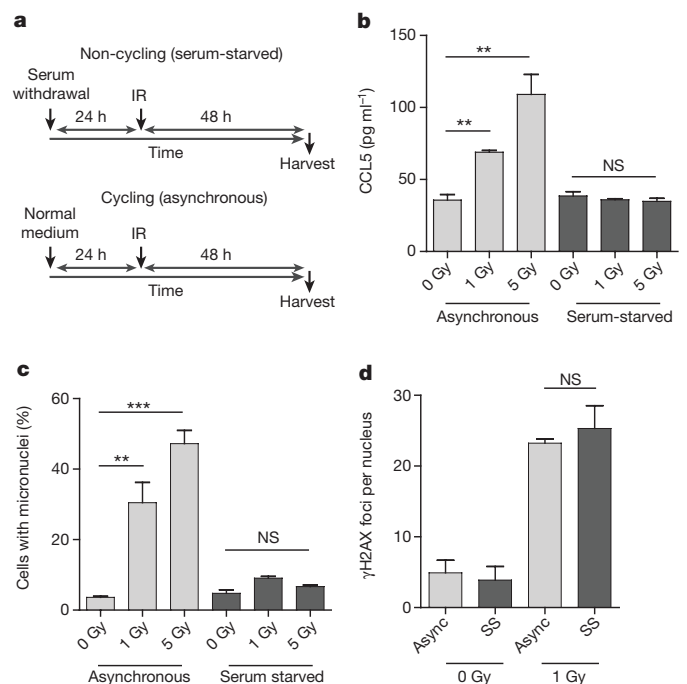

Figure 4 | Innate immune activation after radiation-induced DNA damage is cell-cycle dependent. a, Schematic of experimental protocol. **b**, CCL5 production is significantly increased after irradiation in cycling (asynchronous) cells, but not in cells arrested in G0 after serum starvation. **c**, The percentage of cells containing micronuclei is increased after irradiation in cycling but not G0-arrested cells. Mean ± s.e.m., $n = 3$ independent experiments. **d**, Cycling and G0-arrested cells exhibit the same level of DNA damage, as measured by formation of γH2AX foci. Mean ± s.d., $n = 2$ independent experiments, ≥100 cells analysed per condition per experiment. Only 1 Gy quantified, as γH2AX foci overlapped substantially at 5 Gy; see Extended Data Fig. 6). **P < 0.01, ***P < 0.001, two-tailed t-test; NS, not significant.
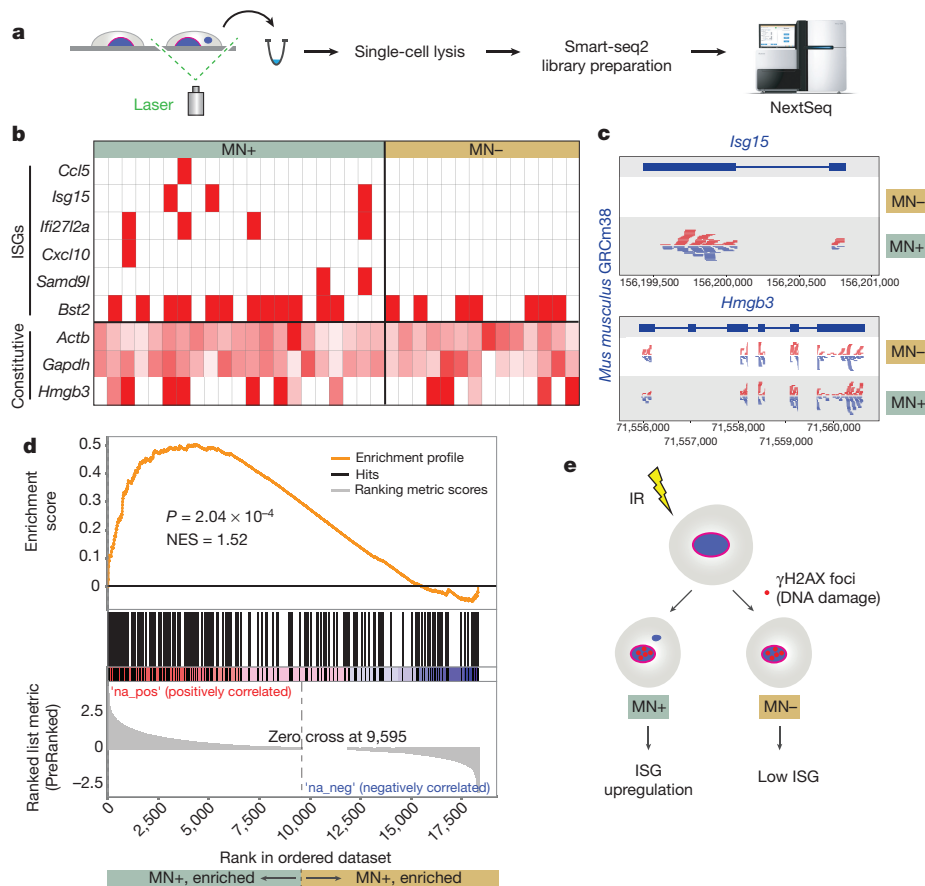
Figure 5 | ISG upregulation occurs specifically in micronucleated cells following DNA damage. a, Experimental outline: 48 h after irradiation (1 Gy) of C57BL/6J MEFs, individual live cells with normal nuclear morphology (MN−) or with micronuclei (MN+) were identified microscopically and excised by laser microdissection. Single-cell transcriptomes were then generated. **b, c**, Transcripts of multiple ISGs were detected only in micronucleated cells. **b**, Heatmap for individual cells, showing ISGs and constitutively expressed control genes (*Actb*, *Gapdh*, *Hmgb3*). Red boxes denote detection of one or more reads per ISG. Control genes shaded in red according to read count relative to the maximum observed for each gene. **c**, Alignment of pooled sequence reads to *Isg15* and *Hmg3b* (Ensembl mouse reference genome version GRCm38. p5). **d**, Transcriptome-wide analysis (GSEA) demonstrates that a set of 336 ISGs is significantly enriched ($P = 2.04 \times 10^{-4}$) in transcriptomes of micronucleated cells. **e**, Interpretation of single-cell data: after induction of DNA damage by irradiation, ISG upregulation occurs preferentially in micronucleated cells rather than those with normal nuclear morphology exposed to identical experimental conditions.

also increased significantly ($P = 0.0083$; Extended Data Fig. 8e), leading us to conclude that nocodazole-induced micronuclei could also be detected by endogenous cGAS. Induction of micronuclei by nocodazole also induced a proinflammatory response with significantly elevated CCL5 cytokine production in MEFs ($P = 0.0008$; Extended Data Fig. 8f). This response, but no increase in DNA damage, was detected from 16 h after nocodazole treatment, consistent with a post-mitotic origin (Extended Data Fig. 8h, j). We therefore conclude that micronuclei arising from mis-segregated chromosomes, as well as genome instability, can induce cGAS signalling.

To confirm a direct relationship between micronuclei and cGAS pathway activation, we assessed at the single-cell level whether the induction of ISGs was specific to micronucleated cells. Forty-eight hours after irradiation (1 Gy), we identified individual micronucleated cells (MN+) and control cells with normal nuclear morphology (MN−) microscopically (Extended Data Fig. 9) and isolated them using laser capture microdissection for subsequent single-cell mRNA sequencing (Fig. 5a). To avoid confounding biases, all cells were collected from the same culture dish and processed in parallel, with library preparation performed in a single 96-well plate with MN+ and MN− cells inter-digitated, and sequencing datasets down-sampled to the same number of reads after mapping. For the thirty-five RNA sequencing libraries that passed quality control (Extended Data Fig. 9a), we first examined a high-confidence list of 11 ISGs that had been previously shown to be induced by endogenous genome instability in MEFs[5]. Six of the

ISGs were represented in the RNA sequencing data, and strikingly we detected transcripts for five of these (*Ccl5*, *Isg15*, *Ifi27l2a*, *Samd9l* and *Cxcl10*) specifically in micronucleated cells ($P = 0.047$, Fisher's exact test, Fig. 5b). Examination of pooled cells down-sampled to identical numbers of sequence reads aligned to these individual genes confirmed differential expression between MN+ and MN− cells (Fig. 5c).

We next assessed the MN+-specific upregulation of a large set of independently defined type I IFN-induced genes ($n = 336$ ISGs) on a transcriptome-wide basis. We performed a gene set enrichment analysis (GSEA) against genes ranked by *z*-scores for differential expression (MN+ versus MN−) calculated by the Single Cell Differential Expression (SCDE) analysis package. GSEA confirmed that expression of these ISGs was significantly enriched in the pool of micronucleated cells over control cells (Fig. 5d; normalized enrichment score 1.52, $P = 2.04 \times 10^{-4}$). These single-cell analyses were therefore consistent with micronucleated cells being the source of DNA damage-induced ISGs (Fig. 5e). In conclusion, while the formal possibility remains that other dsDNA fragments in cells with micronuclei also activate cGAS, this and our other experimental findings strongly implicate micronuclei as a substantial source of cell-intrinsic immunostimulatory DNA.

Micronuclear membrane breakdown provides a mechanism by which dsDNA is exposed to the cytoplasmic sensor cGAS, with spontaneous rupture being frequent and generally irreversible[18]. As chromatin activates cGAS, physiological ruptures in the primary nuclear membrane when cells migrate through tight interstitial spaces

may also cause transient cGAS activation[22,23]. Notably, mitotic chromosomes are temporarily exposed to the cytosol, prompting the important unanswered question of how cGAS–STING pathway activation is prevented during cell division. Although cGAS can localize to nuclear DNA during mitosis (Extended Data Fig. 10a, b), the transient nature of mitosis, hypercompaction of DNA and the peri-chromosomal layer of proteins may mitigate against cGAS binding and activation. In addition, post-translational regulation of cGAS or downstream pathway components in conjunction with transcriptional silencing of mitotic chromosomes could also prevent ISG induction.

Conversely, nuclear membrane breakdown in disease states[24] may be pathologically relevant. Additionally, other aberrant structures generated during cell division may lead to cytosolic DNA exposure. For instance, we infrequently observed cGAS on interphase chromatin bridges in *Rnaseh2b*[−/−] cells and in U2OS cells (Extended Data Fig. 10c, d), so persistent chromatin bridges, such as those arising after telomere crisis[25], could also activate cGAS. Furthermore, while we observed enhanced ISG induction in micronucleated cells, free dsDNA fragments, perhaps released at mitosis, might also activate cGAS in specific contexts. Therefore, while micronuclei provide a substantial source of immunostimulatory DNA, in other pathological contexts additional mechanisms that impair nuclear compartment integrity may play a role.

Micronuclei frequently form in cancer cells[26], and chromosome and genome instability are key drivers of neoplasia[27,28]. Hence, our work predicts that cGAS will often become activated by this route during neoplastic transformation, leading to cGAS- and STING-dependent tumour-suppressive immune responses[29,30]. Consequently, there may be selection pressures during cancer evolution to inactivate cGAS–STING signalling, providing an additional explanation for its frequent inactivation in tumours[31], alongside oncogene-mediated silencing associated with virally induced neoplasia[32]. In conclusion, sensing of ruptured micronuclei by cGAS represents a cell-intrinsic surveillance mechanism that links genome instability to innate immune responses, of relevance to both cancer and autoinflammation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Roers, A., Hiller, B. & Hornung, V. Recognition of endogenous nucleic acids by the innate immune system. *Immunity* **44,** 739–754 (2016).
2. Ahn, J. et al. Inflammation-driven carcinogenesis is mediated through STING. *Nat. Commun.* **5,** 5166 (2014).
3. Gao, D. et al. Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases. *Proc. Natl Acad. Sci. USA* **112,** E5699–E5705 (2015).
4. Härtlova, A. et al. DNA damage primes the type I interferon system via the cytosolic DNA sensor STING to promote anti-microbial innate immunity. *Immunity* **42,** 332–343 (2015).
5. Mackenzie, K. J. et al. Ribonuclease H2 mutations induce a cGAS/STING-dependent innate immune response. *EMBO J.* **35,** 831–844 (2016).
6. Pokatayev, V. et al. RNase H2 catalytic core Aicardi-Goutières syndrome-related mutant invokes cGAS–STING innate immune-sensing pathway in mice. *J. Exp. Med.* **213,** 329–336 (2016).
7. Zhang, C. Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522,** 179–184 (2015).
8. Paludan, S. R. Activation and regulation of DNA-driven immune responses. *Microbiol. Mol. Biol. Rev.* **79,** 225–241 (2015).
9. Gao, P. et al. Cyclic [G(2′,5′)pA(3′,5′)p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153,** 1094–1107 (2013).
10. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339,** 786–791 (2013).
11. Ablasser, A. et al. cGAS produces a 2′-5′-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498,** 380–384 (2013).
12. Ohkuri, T. et al. STING contributes to antiglioma immunity via triggering type I IFN signals in the tumor microenvironment. *Cancer Immunol. Res.* **2,** 1199–1208 (2014).
13. Shen, Y. J. et al. Genome-derived cytosolic DNA mediates type I interferon-dependent rejection of B cell lymphoma cells. *Cell Reports* **11,** 460–473 (2015).
14. Barber, G. N. STING: infection, inflammation and cancer. *Nat. Rev. Immunol.* **15,** 760–770 (2015).
15. Chen, Q., Sun, L. & Chen, Z. J. Regulation and function of the cGAS–STING pathway of cytosolic DNA sensing. *Nat. Immunol.* **17,** 1142–1149 (2016).
16. Reijns, M. A. et al. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **149,** 1008–1022 (2012).
17. Crasta, K. et al. DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482,** 53–58 (2012).
18. Hatch, E. M., Fischer, A. H., Deerinck, T. J. & Hetzer, M. W. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell* **154,** 47–60 (2013).
19. Zhang, X. et al. The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Reports* **6,** 421–430 (2014).
20. Li, X. et al. Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39,** 1019–1031 (2013).
21. Lan, Y. Y., Londoño, D., Bouley, R., Rooney, M. S. & Hacohen, N. Dnase2a deficiency uncovers lysosomal clearance of damaged nuclear DNA via autophagy. *Cell Reports* **9,** 180–192 (2014).
22. Denais, C. M. et al. Nuclear envelope rupture and repair during cancer cell migration. *Science* **352,** 353–358 (2016).
23. Raab, M. et al. ESCRT III repairs nuclear envelope ruptures during cell migration to limit DNA damage and cell death. *Science* **352,** 359–362 (2016).
24. Hatch, E. & Hetzer, M. Breaching the nuclear envelope in development and disease. *J. Cell Biol.* **205,** 133–141 (2014).
25. Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and kataegis induced by telomere crisis. *Cell* **163,** 1641–1654 (2015).
26. Gisselsson, D. et al. Abnormal nuclear shape in solid tumors reflects mitotic instability. *Am. J. Pathol.* **158,** 199–206 (2001).
27. Aguilera, A. & Gómez-González, B. Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.* **9,** 204–217 (2008).
28. Bakhoum, S. F. & Compton, D. A. Chromosomal instability and cancer: a complex relationship with therapeutic potential. *J. Clin. Invest.* **122,** 1138–1143 (2012).
29. Ablasser, A. & Gulen, M. F. The role of cGAS in innate immunity and beyond. *J. Mol. Med. (Berl.)* **94,** 1085–1093 (2016).
30. Woo, S. R. et al. STING-dependent cytosolic DNA sensing mediates innate immune recognition of immunogenic tumors. *Immunity* **41,** 830–842 (2014).
31. Xia, T., Konno, H., Ahn, J. & Barber, G. N. Deregulation of STING signaling in colorectal carcinoma constrains DNA damage responses and correlates with tumorigenesis. *Cell Reports* **14,** 282–297 (2016).
32. Lau, L., Gray, E. E., Brunette, R. L. & Stetson, D. B. DNA tumor virus oncogenes antagonize the cGAS-STING DNA-sensing pathway. *Science* **350,** 568–571 (2015).

METHODS

Mice. The following lines (intercrossed where necessary) were used to provide cells and cell lines for this study: C57BL/6J *Rnaseh2b*[tm1d/+] (referred to as *Rnaseh2b*[+/−])[5], C57BL/6J *Rnaseh2b*[tm2-hgu-A174T] (referred to as *Rnaseh2b*[A174T/A174T])[5], *Trp53*[tm1ty]/J (referred to as *Trp53*[+/−])[33] and C57BL/6NTac-*Mb21d1*[tm1a(EUCOMM)Hmgu]/IcsOrl (referred to as *Mb21d1*[−/−] or *cGAS*[−/−])[34]. The latter was obtained from the Institute Clinique de la Souris via the European Mouse Mutant Archive. All mouse work was performed in accordance with UK Home Office regulations under a UK Home Office project licence (PPL 60/4424).

Cells and cell culture. C57BL/6J *Rnaseh2b*[−/−] *Trp53*[−/−] and *Rnaseh2b*[+/+] *Trp53*[−/−] MEFs were generated from individual embryonic day 10.5 (E10.5) embryos as previously described[5]. C57BL/6NTac *Mb21d1*[−/−] MEFs and C57BL/6J (*Trp53*[+/+]) MEFs were generated from E13.5 embryos. MEFs were maintained in DMEM supplemented with 10% fetal bovine serum (FBS), 50 U ml[−1] penicillin, 50 μg ml[−1] streptomycin and 0.1 mM β-mercaptoethanol at 37 °C in 5% CO_2 and 3% O_2.

U2OS cells were purchased from the European Collection of Authenticated Cell Cultures (ECACC, Cat no. 92022711) and maintained in McCoy's 5A medium with L-glutamine (Gibco) supplemented with 10% FBS, 50 U ml[−1] penicillin and 50 μg ml[−1] streptomycin at 37 °C in 5% CO_2. All cells were mycoplasma-free, with regular checks performed using the Lonza-Mycoalert Mycoplasma Detection Kit.

Generation of stable cell lines. *Rnaseh2b*[+/+] *Trp53*[−/−] and *Rnaseh2b*[−/−] *Trp53*[−/−] MEFs and U2OS cells stably expressing GFP or GFP–cGAS were generated as follows. GFP–cGAS (the result of cloning the mouse or human cGAS coding sequence into a Gateway-compatible version of pEGFP-C1) and GFP were amplified by PCR and cloned into pMSCV. The resulting pMSCV vectors were transfected into Phoenix packaging cells to generate ecotropic or amphotropic viral particles[35], and retroviral supernatant was used to infect the relevant cell lines in the presence of 4 μg ml[−1] polybrene. Cells were selected for stable integration using 2 μg ml[−1] puromycin, and maintained as a heterogeneous pool of stably expressing cells. Stable U2OS GFP–cGAS cells were transfected with pmCherry–NLS (a gift from M. Offterdinger, Addgene plasmid 39319)[36] using lipofectamine 3000 and selected with 500 μg ml[−1] neomycin (G418, Sigma).

Immunofluorescence. Cells were grown on glass coverslips and fixed in 4% paraformaldehyde (PFA) in PBS for 20 min at room temperature. Cells were permeabilized in 0.5% Triton X-100 for 5 min before blocking for 30 min with 1% BSA in PBS. Coverslips were then incubated with primary antibody for 1 h, followed by incubation with a secondary antibody for 45 min. Primary and secondary antibodies were diluted in blocking buffer and all incubations were performed at room temperature. Coverslips were mounted using Vectashield Antifade Mounting Medium with DAPI (Vector Laboratories) and imaged using a Photometrics Coolsnap HQ2 CCD camera and a Zeiss Axioplan II fluorescence microscope with plan-neofluor objectives and images captured with iVision software (BioVision Technologies). All scoring was performed under blinded conditions.

The following antibodies were used for immunofluorescence: cGAS (D1D3G, Cell Signalling, 1:200), phospho-histone H2A.X (Ser139) (2577, Cell Signalling, 1:800), Lamin B1 (ab16048, Abcam, 1:1,000) and Retinoblastoma (554136, BD Biosciences, 1:200). Secondary antibodies, anti-mouse-Alexa488 (A11029), anti-rabbit-Alexa 488 (A11008) and anti-rabbit-Alexa568 (A11036) (all Invitrogen), were used at 1:500 or 1:1,000 dilution.

Determination of micronucleus frequency. After fixation and DAPI staining, the percentage of cells with micronuclei was determined by microscopy under blinded conditions. Micronuclei were defined as discrete DNA aggregates separate from the primary nucleus in cells where interphase primary nuclear morphology was normal. Cells with an apoptotic appearance were excluded.

Erythrocyte micronucleus assay. The presence of micronuclei in erythrocytes was quantified using previously described methods[37]. 50 μl of blood from female *Rnaseh2b*[A174T/A174T] mice and age-matched (7–8 months old) and sex-matched control C57BL/6J mice was collected and expelled into 250 μl heparin solution (500 USP units per ml PBS) on ice. 180 μl of this suspension was expelled into 2 ml of pre-cooled (−80 °C) methanol and aggregates removed by robust tube tapping. The cell suspension was incubated at −80 °C for at least 24 h before staining for flow cytometry.

To stain the cells, 12 ml ice-cold saline solution (0.9% w/v NaCl, 5.3 mM $NaHCO_3$, pH 7.3) was added to the cell suspension and the tube was inverted to mix and placed on ice. Cells were centrifuged at 600*g* for 5 min at 4 °C, supernatant removed and cells resuspended by pipetting. 10 μl of cells per sample was stained with 90 μl of staining solution containing 79 μl saline solution, 10 μl RNase A at 10 mg ml[−1] (Sigma) and 1 μl rat anti-mouse CD71-FITC (Invitrogen RM5301) for 30 min on ice followed by 30 min at room temperature. 1 ml of 1.25 μg ml[−1] propidium iodide (Sigma) was added before data acquisition, using an LSR Fortessa (BD Biosciences). Data were analysed using FlowJo v.7.6.5 software (Tree Star). As the experiment was based on genotypes, no randomization was possible.

Assessment of immune responses. The concentration of CCL5 in supernatants was determined by ELISA (R&D Systems). The cellular response to dsDNA was determined using herring testes DNA (HT-DNA) (Sigma) at a final concentration of 1 μg ml[−1] (Extended Data Fig. 6) or interferon stimulatory DNA (ISD naked, InvivoGen) at a final concentration of 1.33 μg ml[−1] (Extended Data Fig. 8). Cells were incubated overnight and transfected the following day using Lipofectamine 2000/3000 in Opti-MEM reduced serum medium (both Thermo Fisher Scientific). The immune response was assessed at 24 h by ELISA using collected supernatant.

Dose response to HT-DNA. C57BL/6J (*Trp53*[+/+]) MEFs (5×10^4 per well) were seeded in a 12-well plate, incubated overnight and transfected the following day with HT-DNA at the specified concentrations (0–3,000 ng per well; Extended Data Fig. 7) using Lipofectamine 3000 in Opti-MEM reduced serum medium (Thermo Fisher Scientific). After 6 h the medium was replaced; supernatants were collected at 24 h and CCL5 concentrations assessed by ELISA.

RT–qPCR. RNA was extracted from adherent cells using the RNeasy kit (Qiagen) per the manufacturer's instructions and using the included DNase I treatment. cDNA was prepared using Superscript III RT and random oligomer primers (Thermo Fisher Scientific). qRT–PCR was performed using SYBR Select Master Mix (Thermo Fisher Scientific) on an LC480 Real-Time PCR machine (Roche). The expression of target genes was normalized to the housekeeping gene *Hprt* using the formula ($2-\Delta C_t$). Supplementary Table 1 shows the primers used.

X-ray irradiation. 1.3×10^5 cells were seeded per well of a 6-well plate onto glass coverslips (Fig. 1). The following day, cells were irradiated for 1 min at a dose rate of 1 Gy per min using an X-ray irradiation system (Faxitron 43855D, Faxitron X-ray Corporation) operated at 130 kVp. After a further 48 h the culture medium was removed and CCL5 concentration assessed by ELISA (R&D systems). Cells on coverslips were fixed with 4% PFA in PBS for 20 min at room temperature for immunofluorescence imaging.

Live cell imaging. For live cell imaging, *Rnaseh2b*[−/−] *Trp53*[−/−] MEFs transiently expressing mCherry–H2B or U2OS cells stably expressing GFP–cGAS and mCherry–NLS were seeded onto glass-bottomed plates (Greiner Bio One). pmCherry–H2B (gift from P. Vagnarelli, Brunel University, London) was transfected by electroporation with the Neon transfection system (Thermo Fischer Scientific) according to the manufacturer's instructions. Prior to imaging, U2OS cells were treated with 200 ng ml[−1] nocodazole (Sigma) for 12 h and additionally stained with 0.5 μM SiR-DNA (Spirochrome). Cells were maintained at 37 °C in Leibovitz L-15 medium (Gibco). TRITC, FITC and Cy5 image datasets were collected using a 40× plan neofluor (0.75NA) objective on an Axiovert 200 fluorescence microscope (Zeiss) equipped with a Retiga6000 CCD camera (Qimaging) or an Axio-Observer Z1 fluorescence microscope (Zeiss) equipped with an Evolve EMCCD camera (Photometics). Images were recorded every 1.5 min over a 4-h period (U2OS cells) or every 5 min over a 16-h period (MEFs) using Micromanager (https://open-imaging.com/) and subsequently deconvolved using Volocity software (PerkinElmer).

cGAS recruitment following micronuclear envelope disruption was analysed by measuring the mean fluorescence intensity of mCherry–NLS or GFP–cGAS over time. The mean fluorescence intensity for a cytoplasmic background of matched area was subtracted from the micronuclear signal. The background subtracted fluorescence intensity values were then normalized so that the highest fluorescence intensity was 1 and the lowest fluorescence intensity 0. The time of rupture ($t = 0$) was set at the point of maximal decline in the mean mCherry–NLS fluorescence signal. Rupture was observed in $n = 40$ micronuclei during live imaging experiments, in 29 of which detectable levels of cGAS were observed subsequently, and time of entry was then quantified in 11 of these, where movie length was sufficient to follow cGAS accumulation to maximal intensity.

Serum starvation experiments. C57BL/6J (*Trp53*[+/+]) MEFs (1.5×10^5 for asynchronous and 6×10^5 for serum starvation per well) were seeded into a 6-well plate (seeding densities were chosen to achieve similar cell numbers for both conditions at assay endpoint). After 5 h the medium was replaced with low-serum medium (0.25% FBS) for serum starvation, and with fresh medium (10% FBS) for asynchronous/cycling cells. After a further 24 h, the medium was again replaced with 1.6 ml (0.25% FBS for serum starved cells; 10% FBS for control cells), and cells irradiated with 1 or 5 Gy (at 1 Gy per min) (Faxitron 43855D, Faxitron X-ray Corporation). After 48 h the supernatants and cells were harvested, CCL5 measured by ELISA and RT–qPCR performed.

To assess the immune response of serum-starved cells to HT-DNA, C57BL/6J (*Trp53*[+/+]) MEFs (1.5×10^5 per well) were seeded into a 6-well plate and grown in 10% or 0.25% FBS-containing medium for 24 h. Cells were then transfected with 1 μg per ml HT-DNA in Opti-MEM using Lipofectamine 3000 (Thermo Fisher Scientific). After 6 h the medium was replaced with 10% or 0.25% FBS-containing medium, respectively, and 24 h after transfection supernatants were taken for ELISA. To ensure differences in cell number (arising from G0 arrest or irradiation) were not confounders for ELISA results, adherent cells were

counted after trypsinization, with ELISA results corrected for final cell number per well.

To assess the frequency of micronuclei by immunofluorescence, cells were fixed with 4% PFA for 20 min at room temperature and permeabilized with 0.5% Triton X-100 for 5 min. Cover slips were mounted with Vectashield (with DAPI) and the percentage of cells with discrete micronuclei determined under blinded conditions. More than 500 cells per condition per experiment were analysed for micronucleus counts. To assess DNA damage by γH2AX immunofluorescence in serum-starved and asynchronous cells after irradiation, 2×10^5 cells per well of a 6-well plate were seeded on coverslips and incubated in normal or 0.25% FBS medium. Cells were then irradiated with 1 or 5 Gy, and fixed 2 h later. $n > 100$ cells per condition per experiment ($n = 2$).

Chromosome mis-segregation assay. Micronuclei induced by chromosome mis-segregation were generated pharmacologically by nocodazole treatment as described previously[7,18]. MEFs were plated at 8×10^5 cells per 100-mm plate, and 24 h later incubated with 100 ng ml^{-1} nocodazole (Sigma) for 6 h. Mitotic cells were then harvested by shake-off and washed three times with PBS before counting and plating. 1×10^5 mitotic cells were plated per well of a 12-well plate into 800 μl fresh medium. 1.5×10^4 mitotic cells were concurrently plated onto a coverslip in one well of a 12-well plate. After 48 h, culture medium was removed for ELISA and cells on coverslips fixed for imaging to determine micronucleus frequency.

To induce micronuclei in U2OS cells, 7×10^5 cells were seeded per 100-mm plate and incubated for 9 h. Cells were then synchronized with 2 mM thymidine (ACROS Organics) for 27 h and released by washing twice and replacing with fresh medium. After 14 h, U2OS cells were incubated with 100 ng ml^{-1} nocodazole (Sigma) for 6 h and mitotic cells harvested by shake-off. 3×10^4 mitotic cells were plated onto a coverslip in one well of a 12-well plate and incubated for 48 h before processing for microscopic analysis.

Chromosome mis-segregation time-course assay. C57BL/6J $Trp53^{-/-}$ MEFs (8×10^5) were plated in 10-cm plates. The next day, cells were incubated with 100 ng ml^{-1} nocodazole (Sigma) for 6 h and mitotic cells harvested using the shake-off method. At this point supernatant was taken for the $t = 0$ time point. Mitotic cells were then plated into a 12-well plate (1×10^5 per well) or onto coverslips (7×10^4) for immunofluorescence analysis. Asynchronous cells were plated at the same number concurrently. Supernatants were taken and coverslips fixed after 6, 16 and 22 h. Immune response was analysed by ELISA and coverslips were pre-extracted on ice using 0.5% Triton X-100 in PBS for 5 min. Cells were then fixed with 4% PFA for 15 min at room temperature. After blocking for 30 min with 3% BSA at room temperature, γH2AX Ser139 antibody (05-636 Millipore) was added for 2 h at room temperature. Alexa Fluor 568 goat anti-mouse secondary antibody (Life technologies) was then applied and incubated for 1 h at room temperature. Coverslips were mounted using Vectashield antifade mounting medium with DAPI (Vector laboratories) and imaged at room temperature using a Coolsnap HQ CCD camera (Photometrics) and a Zeiss Axioplan II fluorescence microscope with ×40 and ×63 plan-neofluor objectives and acquired using micromanager (http://open-imaging.com/). $n \geq 100$ cells per condition ($n = 1$).

siRNA knockdown. Cells were plated at an optimized density (2.5×10^5 U2OS cells per 6-cm dish) before overnight incubation and transfected the following day with siRNA oligonucleotides targeting cGAS (Dharmacon, M-015607-01-0005, siGENOME 115004) or luciferase (CUUACGCUGAGUACUUCGA, Sigma) at a final concentration of 25 nM. Transfections were performed using Oligofectamine (Thermo Fisher Scientific) per the manufacturer's instructions, in Opti-MEM reduced-serum medium (Thermo Fisher Scientific). Transfection medium was replaced with complete medium after 6 h and protein depletion confirmed 48 h post-transfection by immunoblotting.

Immunoblotting. Whole-cell extracts were prepared by lysis and sonication of cells in UTB buffer (8 M urea, 50 mM Tris, pH 7.5, 150 mM β-mercaptoethanol, protease inhibitor cocktail (Roche)) and subsequently analysed by SDS–PAGE following standard procedures. In brief, protein samples (20 μg) were resolved on a 4–12% NuPAGE Novex Bis-Tris mini gel (Thermo Fisher Scientific) and transferred onto a nitrocellulose membrane (GE Healthcare Life Sciences). The following antibodies were used for immunoblotting: cGAS (D1D3G, Cell Signalling, 1:1,000) and actin (A2066, Sigma, 1:5,000).

cGAS purification. The coding sequence of the enzymatically active portion of human cGAS (amino acids 157–522) was cloned into pGEX6P1 and protein expressed overnight at 18 °C by induction with 0.3 mM IPTG in Rosetta-2 cells. Cells were lysed by sonication in 25 mM Tris-HCl pH 8, 1 M NaCl, 10% glycerol, 0.5% IGEPAL CA-630, 1 mM DTT. GST–cGAS was affinity purified using glutathione sepharose 4B (GE Healthcare Life Sciences) and cGAS released by PreScission Protease cleavage in 50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 1 mM MgCl$_2$, 1 mM DTT.

In vitro assay for cGAS activity. The cGAS enzyme activity assay was performed as described[11] with minor modifications. Reactions were carried out at 37 °C in 10 μl

with 2 μM recombinant cGAS (amino acids 157–522), 100 ng DNA (or equivalent chromatin), 1 mM GTP and 10 μCi [α-32P]-ATP in 13.75 mM Tris-HCl pH 7.5, 37.5 mM NaCl, 4 mM MgCl$_2$. Reactions were stopped by the addition of EDTA. To separate ATP and cGAMP, TLC was performed on 10×10-cm HPTLC Silica gel 60 F254 glass plates (Merck Millipore). Samples were spotted onto the plates and separation was performed in n-propanol/ammonium hydroxide/water (11:7:2 v/v/v). The plate was air-dried and images collected using phosphorimaging screens and the FLA-5100 imaging system (Fujifilm).

Chromatin synthesis and purification. Soluble chromatin was prepared from NIH3T3 cells as previously described[38,39] but an increased concentration of NP40 detergent (0.2%) was used in buffer NBB and buffer NBR was replaced by buffer NBR2 (buffer NBR modified to contain 1 mM MgCl$_2$ and 1 mM CaCl$_2$). Nuclei were resuspended at 20 A260 in buffer NBR2 and digested with MNase (400 units per ml nuclei; NEB) for 10 min at room temperature. The reaction was stopped by adding EDTA to 10 mM, and nuclei were resuspended in 500 μl TEP20N (10 mM Tris, pH 8; 1 mM EDTA; 20 mM NaCl, 0.5 μM PMSF, 0.05% NP40) and incubated at 4 °C overnight. Soluble chromatin was recovered by centrifugation (5 min, 20,000g) and purified on a 10–50% step gradient in TEP80 (10 mM Tris, pH 8; 1 mM EDTA; 80 mM NaCl, 0.5 μM PMSF) in a SW55 centrifuge tube (Beckman) and centrifuged at 50,000 r.p.m. for 105 min in a SW55 rotor as described[40]. Chromatin was recovered by upward displacement while monitoring the absorbance at 254 nm in 10 samples of 0.5 ml. Aliquots of each fraction were analysed for DNA and protein to check sample integrity. Peak chromatin fractions were dialysed into TEP80 overnight and the concentration determined by measuring the absorbance at 260 nm in 2 M NaCl, 5 M Urea. Naked DNA was prepared from chromatin by proteinase K treatment in 0.25% SDS, 50 mM NaCl, 5 mM EDTA for 2 h at 55 °C, followed by phenol/chloroform extraction and ethanol precipitation.

Synthetic chromatin was prepared using standard approaches adapted from refs 41, 42. Essentially, a 601 DNA template (25×197 bp 601 DNA) was reconstituted with purified chicken core histone octamers at a 1:1 molar ratio in TEP2000 (10 mM Tris, pH 7.5, and 0.2 mM EDTA, 2 M NaCl, 0.5 μM PMSF) and dialysed from 2 M to 400 mM over 6 h and then into 10 mM NaCl overnight, using a Thermofisher microdialysis cap (10,000 MWCO) in a linear gradient maker. Chromatin concentration was measured by measuring absorbance at 254 nm and reconstitution chromatin was analysed by band-shifts, sucrose gradient sedimentation and nuclease digestion.

Single-cell laser capture microdissection. C57Bl/6J MEFs were irradiated (1 Gy) and re-seeded onto 50 mm PEN membrane dishes (Zeiss) 32 h later. Picogreen-containing medium (Quant-iT PicoGreen dsDNA reagent, 4 μl per ml) was added to the cells 48 h after irradiation to stain DNA, and laser capture microdissection (LCM) performed as follows. Micronucleated cells ($n = 32$) and control cells with normal nuclear morphology ($n = 28$) were identified from the same dish using a GFP filter, and LCM of individual cells performed using a Zeiss Palm Microbeam 4 Microscope. Cells were collected into 5 μl 0.2% (v/v) Triton X-100, 2 U μl^{-1} RNasin Ribonuclease Inhibitor (Promega), snap frozen on dry ice and stored at −80 °C until library preparation

cDNA and library preparation. All 60 cells obtained from LCM were processed in two batches (until cDNA amplification), with approximately equal numbers of MN+ and MN− cells in each batch, and cDNA from single cells was obtained using the Smart-seq2 protocol[43] with minor modifications, as described previously[44]. Cells from each batch were processed in LCM-compatible 0.5 ml Eppendorf tubes until cDNA amplification and both batches then transferred to a single 96-well plates for library generation using Illumina Nextera reagents. Libraries were assessed for size distribution on an Agilent Bioanalyser (Agilent Technologies) with the DNA HS Kit, and then quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific) and the Qubit dsDNA HS Assay Kit. Finally, parallel paired end sequencing (2×75 bp) was performed using the NextSeq 500/550 Mid-Output v2 (150 cycle) Kit on the NextSeq 550 platform (Illumina).

Single-cell RNA sequencing data analysis. Reads were mapped against the Ensembl mouse reference genome version GRCm38.p5 with the inclusion of the reference for the spike in controls from the ERCC consortium[45] using a STAR RNA-seq aligner[46]. For quality control and pre-processing, quantification of mapped reads per gene was calculated using Rsubread[47]. Genes with no expression detected in all cells were excluded. The gene counts were loaded into the R package scater and standard quality control metrics were calculated[48]. Quality control exclusion criteria were cells with more than 10% of reads mapping to ERCCs or fewer than 40,000 reads or fewer than 2,000 genes detected (at least one read per gene). For the GSEA analysis, genes were ranked from MN+ to MN− using a z-score calculated with the SCDE R package[49]. The ranked gene list was loaded into GSEA and tested against a list of 336 genes shown to be induced in MEFs in response to type I IFN[50]. For direct comparisons of genes between MN+ and MN− cells in the heatmap and browser shots, mapped reads were imported into SeqMonk (https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/)

and down-sampled to the same number of reads for every cell. Raw reads were counted over genes and genes with at least one read were scored as detected genes. Heatmaps were generated using Matrix2png[51]. The upper part of the heatmap (Fig. 5b) shows detected ISGs, from a high confidence list of genome instability induced ISGs, calculated as the overlap between genes identified in refs 52 and 5. Out of these 11 genes, transcripts for 5 were not detected in any cell. A Fisher's exact test was performed under the null hypothesis that there was no excess of transcribed ISGs in MN+ versus MN− cells. Genome browser shots were taken from pooled reads from all MN+ and MN− cells after pools were down-sampled to the same number of reads.

Statistics. All data are plotted as mean values, with variance as s.e.m. unless stated otherwise. Statistical analysis was performed using Prism (Graphpad Software Inc.). For all quantitative measurements, normal distribution was assumed, with t-tests performed, unpaired and two-sided unless otherwise stated. No statistical methods were used to predetermine sample sizes, which were determined empirically from previous experimental experience with similar assays, and/or from sizes generally employed in the field.

Data availability. Single-cell RNA sequencing data that support the findings of this study have been deposited in GEO with the accession code GSE100771. All other data are available upon reasonable request from the corresponding authors.

33. Jacks, T. *et al.* Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4,** 1–7 (1994).
34. Bridgeman, A. *et al.* Viruses transfer the antiviral second messenger cGAMP between cells. *Science* **349,** 1228–1232 (2015).
35. Swift, S., Lorens, J., Achacoso, P. & Nolan, G. P. in *Current Protocols in Immunology* Chapter 10, Unit 10 17C (Wiley, 2001).
36. Micutkova, L. *et al.* Analysis of the cellular uptake and nuclear delivery of insulin-like growth factor binding protein-3 in human osteosarcoma cells. *Int. J. Cancer* **130,** 1544–1557 (2012).
37. Balmus, G. *et al.* A high-throughput *in vivo* micronucleus assay for genome instability screening in mice. *Nat. Protoc.* **10,** 205–215 (2015).
38. Gilbert, N. *et al.* Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118,** 555–566 (2004).
39. Naughton, C., Sproul, D., Hamilton, C. & Gilbert, N. Analysis of active and inactive X chromosome architecture reveals the independent organization

40. Gilbert, N. *et al.* DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J. Cell Biol.* **177,** 401–411 (2007).
41. Huynh, V. A., Robinson, P. J. & Rhodes, D. A method for the *in vitro* reconstitution of a defined "30 nm" chromatin fibre containing stoichiometric amounts of the linker histone. *J. Mol. Biol.* **345,** 957–968 (2005).
42. Rogge, R. A. *et al.* Assembly of nucleosomal arrays from recombinant core histones and nucleosome positioning DNA. *J. Vis. Exp.* **79,** 50354 (2013).
43. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9,** 171–181 (2014).
44. Kirschner, K. *et al.* Proliferation drives aging-related functional decline in a subpopulation of the hematopoietic stem cell compartment. *Cell Reports* **19,** 1503–1511 (2017).
45. Baker, S. C. *et al.* The External RNA Controls Consortium: a progress report. *Nat. Methods* **2,** 731–734 (2005).
46. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
47. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41,** e108 (2013).
48. McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33,** 1179–1186 (2017).
49. Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11,** 740–742 (2014).
50. Saleiro, D. *et al.* Central role of ULK1 in type I interferon signaling. *Cell Reports* **11,** 605–617 (2015).
51. Pavlidis, P. & Noble, W. S. Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **19,** 295–296 (2003).
52. Schoggins, J. W. *et al.* Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature* **505,** 691–695 (2014).
53. Green, C. M. & Almouzni, G. When repair meets chromatin. First in series on chromatin dynamics. *EMBO Rep.* **3,** 28–33 (2002).
54. Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Mol. Cell* **40,** 179–204 (2010).
55. Staresincic, L. *et al.* Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J.* **28,** 1111–1120 (2009).

of 30 nm and large-scale chromatin structures. *Mol. Cell* **40,** 397–409 (2010).
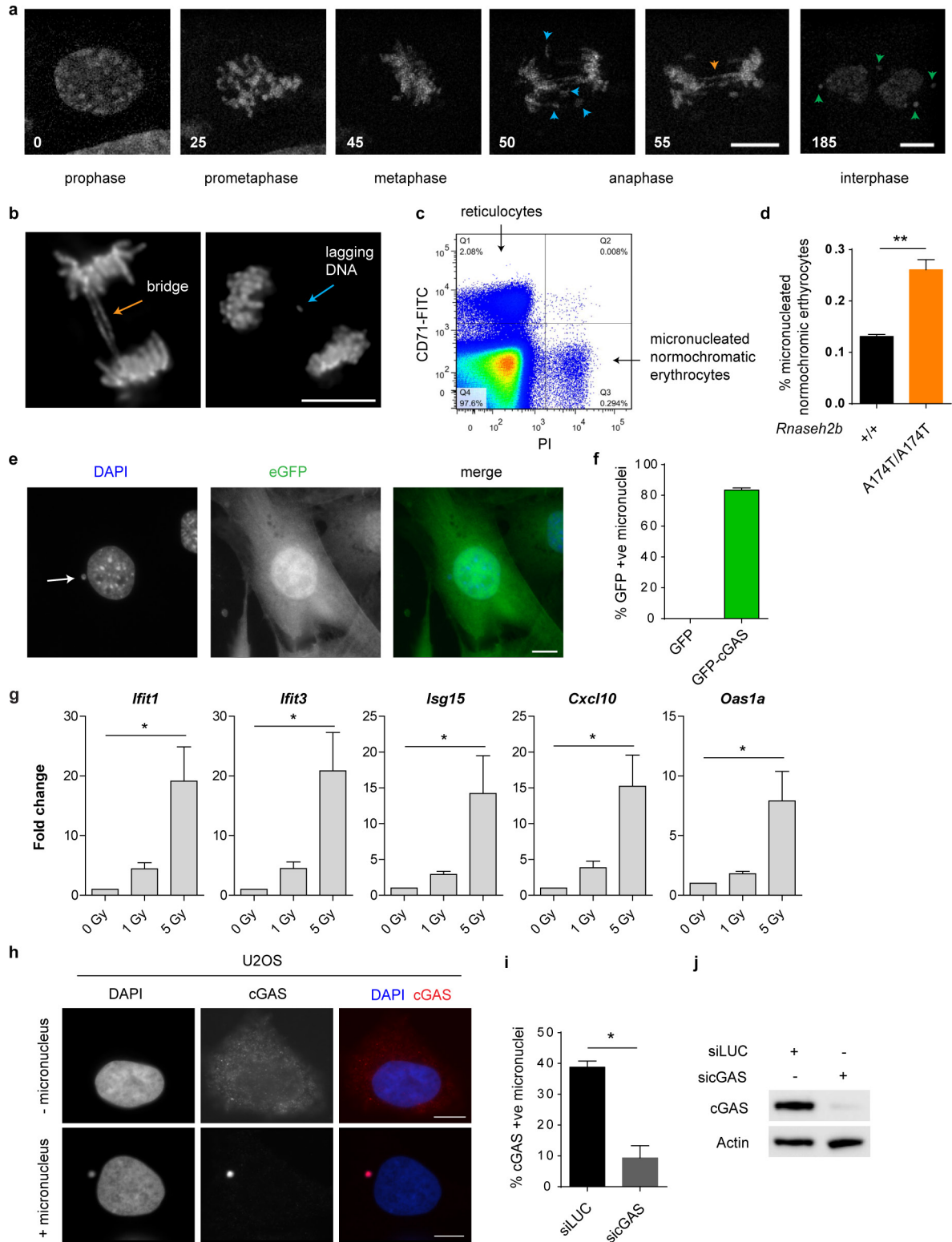
Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Micronuclei form in RNase H2 deficiency, with cGAS localizing to these structures and inducing an ISG response.
**a**, Still images of live imaging in $Rnaseh2b^{-/-}$ MEFs, time in minutes; $t = 0$, prophase. Lagging DNA (blue arrowheads) and DNA bridges (orange arrowhead) at anaphase can result in interphase micronuclei (green arrowheads). **b**, Chromatin bridges and lagging chromosomal DNA (arrows) occur in $Rnaseh2b^{-/-}$ MEFs. Representative fixed cell images. **c, d**, Erythrocyte micronuclei assay[37]. **c**, Representative flow cytometry plot with quadrants containing reticulocytes and micronucleated normochromatic erythrocytes indicated. **d**, $Rnaseh2b^{A174T/A174T}$ mice have a significantly increased frequency of micronucleated erythrocytes. Mean ± s.e.m., $n = 3$ mice per group; two-tailed t-test, $**P < 0.01$. **e, f**, eGFP does not accumulate in micronuclei, whereas the majority of micronuclei show strong accumulation of GFP–cGAS. **e**, Representative image of micronucleus-containing $Rnaseh2b^{-/-}$ MEFs stably expressing eGFP. **f**, Quantification of GFP-positive micronuclei for GFP–cGAS-expressing and GFP-expressing $Rnaseh2b^{-/-}$ MEF lines. Mean ± s.e.m., $n = 4$ experiments (≥500 cells counted per experiment). Scale bars, 10 μm.

g, Increased levels of ISG transcripts (*Ifit1*, *Ifit3*, *Isg15*, *Cxcl10* and *Oas1a*) were detected in C57BL/6J ($Trp53^{+/+}$) MEFs 48 h after irradiation. Transcript levels were normalized to *Hprt*. Mean ± s.e.m., $n = 3$ independent experiments. One-way ANOVA, 2 degrees of freedom, $*P < 0.05$. **h**, Endogenous cytosolic cGAS accumulates in micronuclei in U2OS cells. Representative images of cGAS distribution in cells with or without micronuclei. Images taken using different exposure times (200 vs 700 ms) to visualize weaker cytosolic cGAS signal. **i, j**, Verification of anti-cGAS antibody specificity in human cells. **i**, The percentage of cGAS-positive micronuclei, using anti-cGAS immunofluorescence, was determined microscopically after cGAS or luciferase siRNA knockdown. Mean ± s.e.m., $n = 2$ experiments (500 cells counted per experiment); two-tailed t-test. While several commercial cGAS antibodies were assessed, specific detection of mouse cGAS by immunofluorescence was not possible with these reagents (data not shown). **j**, Immunoblot after siRNA knockdown of cGAS in U2OS cells. siRNA targeting luciferase (siLUC) was used as a negative control. Probing with anti-actin antibody shows equal loading.
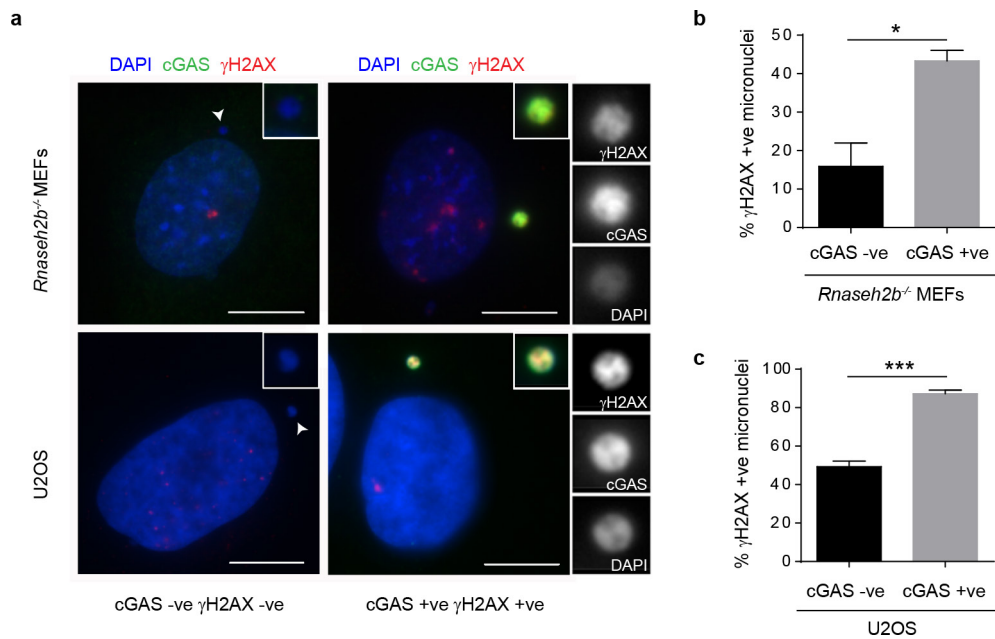
Extended Data Figure 2 | cGAS localization is associated with DNA damage in micronuclei. γH2AX foci in micronuclei correlate with GFP–cGAS localization in *Rnaseh2b*$^{-/-}$ MEFs and endogenous cGAS localization in U2OS cells. **a**, Representative immunofluorescence images: γH2AX, red; cGAS, green. **b**, Percentage of γH2AX-stained micronuclei (γH2AX +ve), either co-stained with cGAS (cGAS +ve), or in which cGAS was not detected (cGAS −ve). *Rnaseh2b*$^{-/-}$ MEFs; ≥500 cells counted per experiment. **c**, Quantification for U2OS cells, ≥250 micronuclei counted per experiment. Mean ± s.e.m., $n = 3$ experiments; *$P < 0.05$, ***$P < 0.001$, two-tailed *t*-test. While our biochemical studies demonstrate that unbroken DNA and chromatin are sufficient to activate cGAS (Fig. 3, Extended Data Figs 4, 5), the increased accessibility of DNA after damage[53] could further assist cGAS binding and activation. Scale bars, 10 μm.

a

b

c
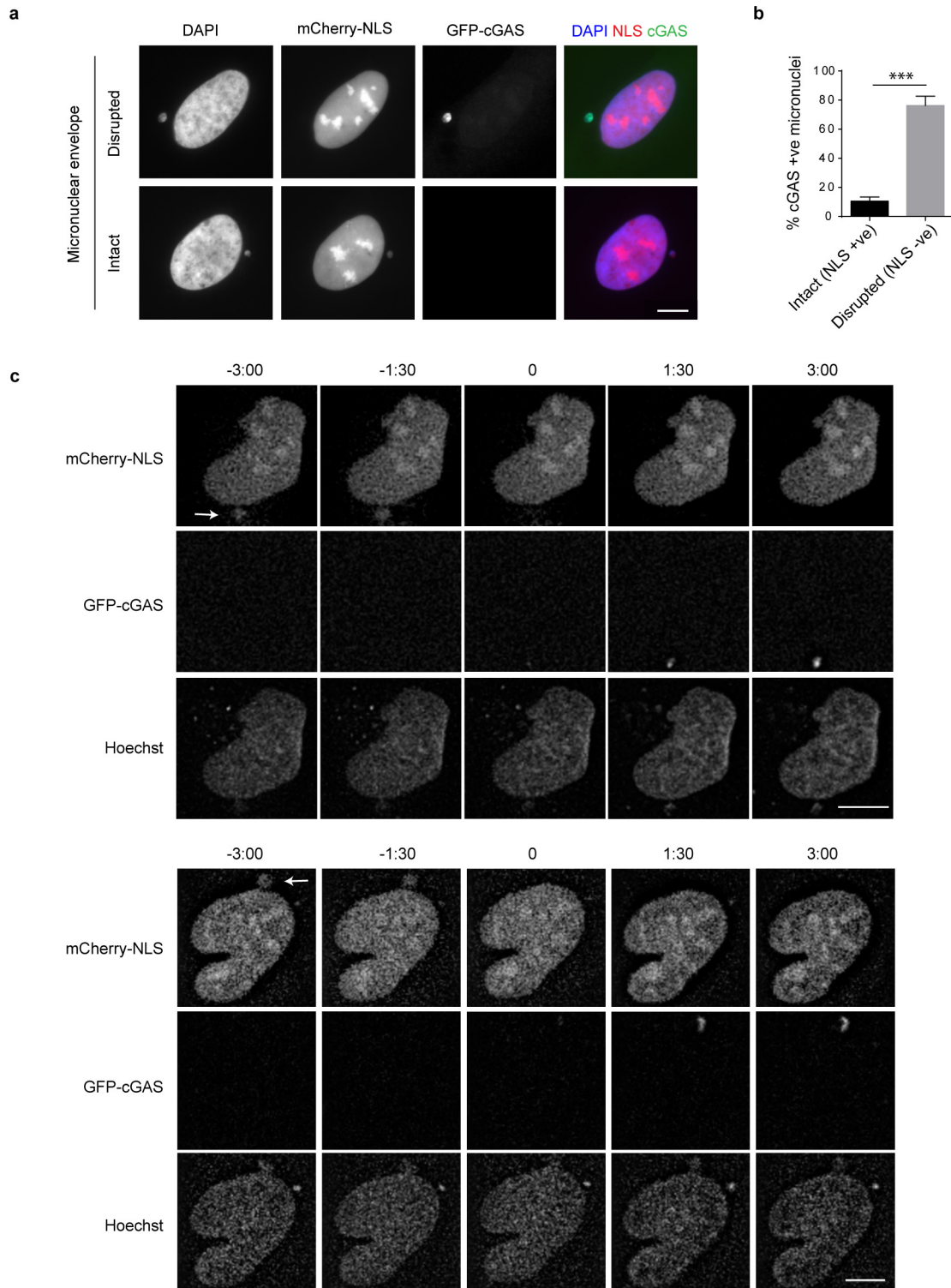
Extended Data Figure 3 | cGAS localizes to micronuclei upon nuclear envelope rupture. a, b, cGAS localization to micronuclei in U2OS cells inversely correlates with localization of mCherry–NLS, which is present only in micronuclei with an intact nuclear envelope. **a,** Representative images of cells containing micronuclei with disrupted or intact nuclear envelopes. **b,** Percentage of intact and disrupted cGAS-positive micronuclei. Mean ± s.e.m., $n = 3$ independent experiments (≥250 micronuclei counted per experiment). NLS +ve and NLS −ve,

mCherry–NLS present in or absent from micronuclei, respectively. cGAS +ve, GFP–cGAS present in micronuclei. ***$P < 0.001$, two-tailed t-test. **c,** Single-channel image for representative stills shown in Fig. 2d from live imaging of U2OS cells expressing mCherry–NLS and GFP–cGAS. DNA visualized with Hoechst stain. Time (min) relative to loss of mCherry–NLS from micronucleus ($t = 0$, micronuclear membrane rupture). Arrows indicate micronuclei undergoing rupture. Scale bars, 10 μm.
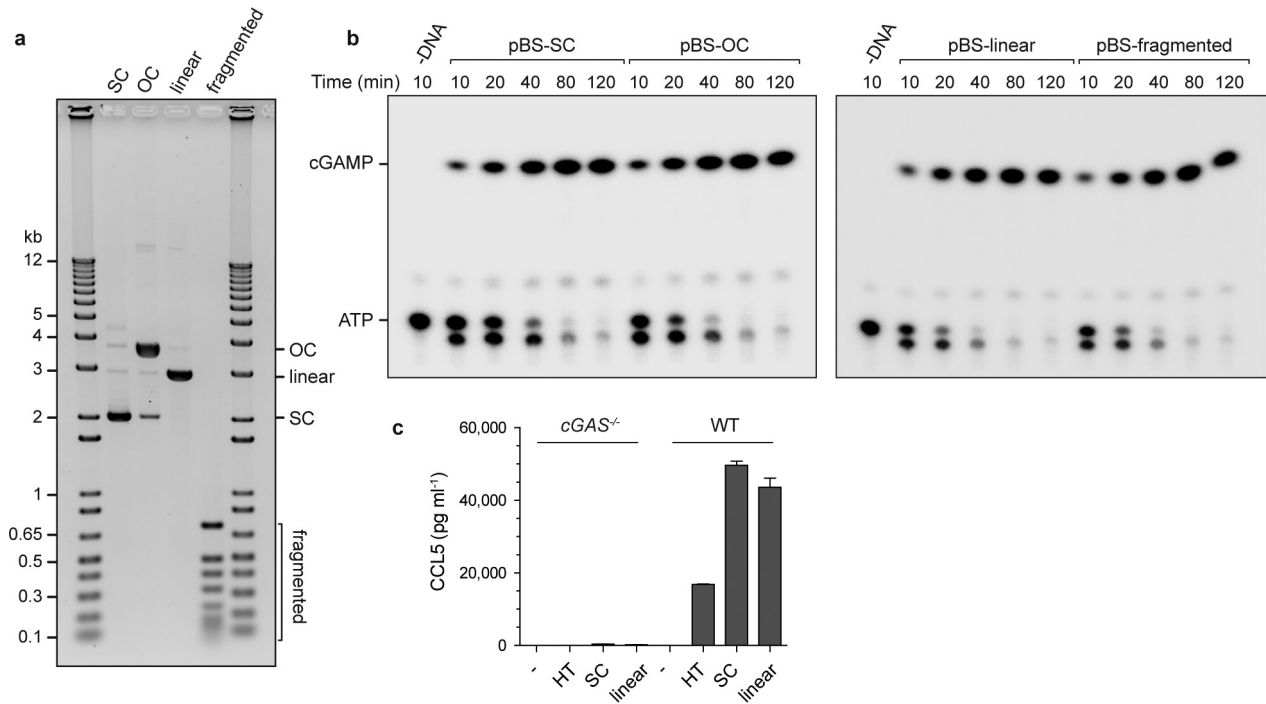
Extended Data Figure 4 | cGAS is activated by circular plasmid DNA.
a, Plasmid DNA (SC, supercoiled; OC, open circle; linear and fragmented) separated by agarose gel electrophoresis. pBluescript II SK(+) supercoiled plasmid DNA was treated with Nt.BspQI nicking endonuclease to generate open circle DNA; with EcoRI to generate a single 3-kb linear fragment; or with HpaII to generate 13 fragments between 710 and 26 bp in size. **b**, Supercoiled, open circle, linear and fragmented pBluescript (pBS) DNA all activate recombinant cGAS to produce cGAMP. Representative images shown. Quantification of $n = 3$ experiments shown in Fig. 3b. **c**, Plasmid DNA induces cGAS-dependent CCL5 production in MEFs. Wild-type and $cGAS^{-/-}$ ($Mb21d1^{-/-}$) MEFs were transfected with 400 ng HT-DNA or supercoiled or linearized pBluescript, and CCL5 production after 24 h measured by ELISA. Mean ± s.e.m., $n = 3$ independent experiments.
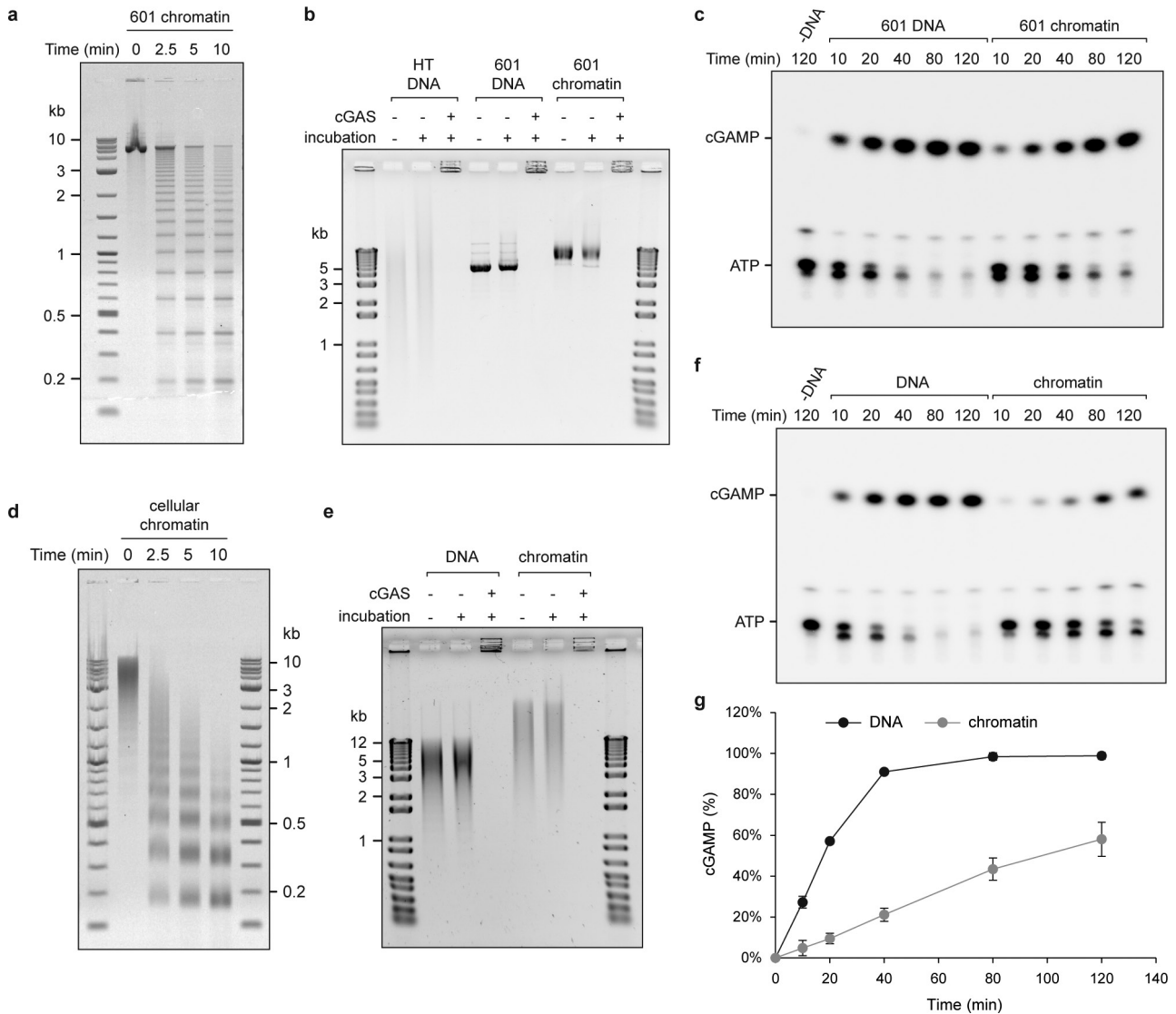
Extended Data Figure 5 | cGAS is activated by chromatin. a, Agarose gel of micrococcal nuclease (MNase)-digested synthetic chromatin assembled onto a 601 DNA template indicates that it has a regular nucleosomal structure. **b,** Chromatin and DNA bind recombinant cGAS; DNA in wells could be the result of near charge neutrality of cGAS–DNA complexes or previously reported cGAS oligomerization. Chromatin is stable under cGAS assay conditions, remaining intact during incubation in cGAS reaction buffer, as evidenced by the bandshift compared to naked DNA. **c,** Representative TLC image demonstrating cGAMP generation by recombinant cGAS in the presence of chromatin. **d,** MNase treatment confirms a nucleosomal ladder pattern for chromatin isolated from mouse NIH3T3 cells. **e,** cGAS binds chromatin, and cellular chromatin is stable under cGAS assay conditions. **f, g,** Cellular chromatin

activates recombinant cGAS, but at a slower rate than the same amount of deproteinized DNA. Representative images shown. Graphs shows quantification from $n = 3$ independent experiments, mean ± s.d. Reduced cGAS activation *in vitro* by chromatin isolated from cells is expected due to the presence of linker histones in addition to the nucleosomal core histones, which has been shown to bind part of the linker DNA, reducing the available sites for cGAS binding, and the use of MNase during the isolation of cellular chromatin. Whereas MNase treatment is needed to fragment the chromatin to allow its purification, it will preferentially cleave accessible non-protein-bound portions, which will further reduce the available sites to which cGAS can bind in the final chromatin preparation. However, such nucleosome-free regions are more likely to allow efficient binding and activation of cGAS *in vivo*.
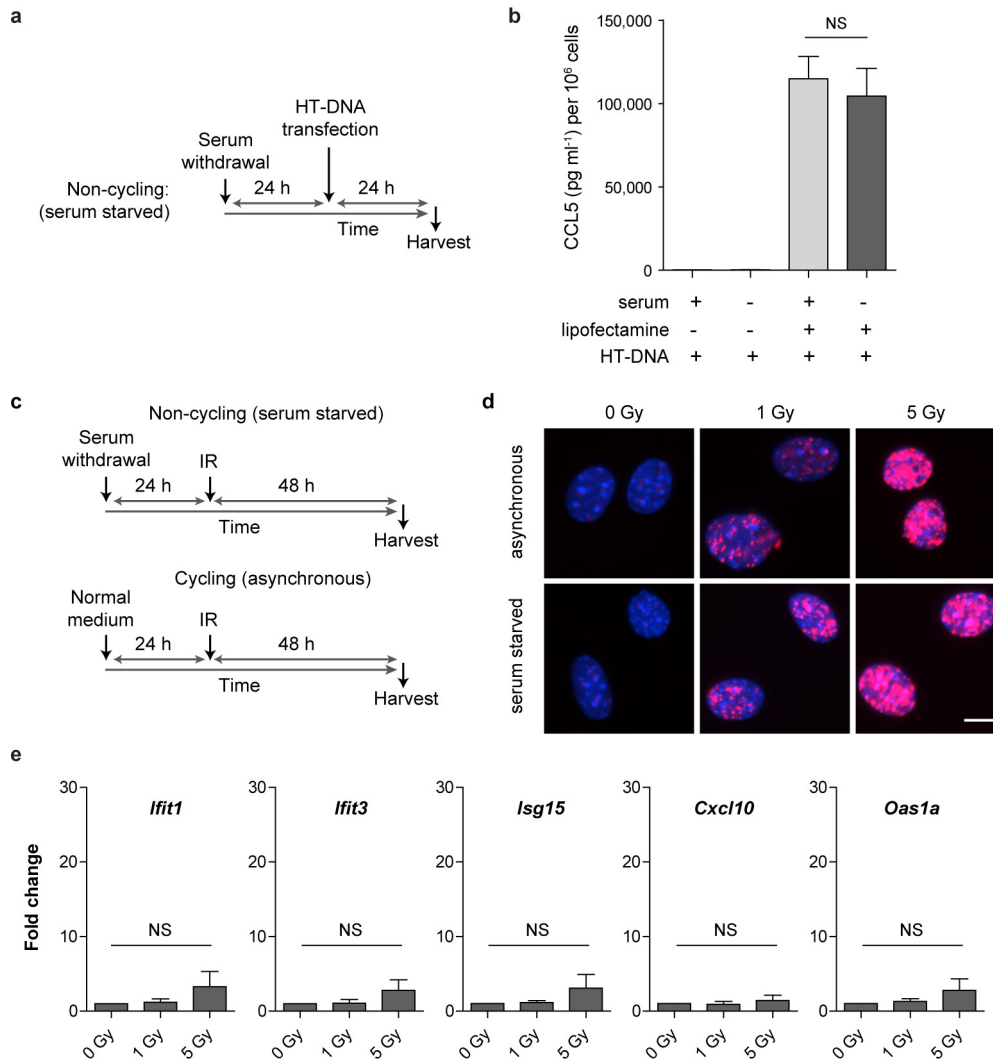
Extended Data Figure 6 | ISG induction by ionizing radiation is abrogated in non-cycling cells. a, Experimental setup: to arrest cells in G0, serum was withdrawn 24 h before transfection with HT-DNA, and supernatant harvested 24 h later. **b**, CCL5 production in response to transfected HT-DNA was equivalent in cycling and serum starved MEFs. Mean ± s.e.m., $n = 2$ independent experiments. **c**, Schematic of experimental protocol. **d**, Cycling and G0-arrested cells exhibit the same level of DNA damage as measured by formation of γH2AX foci.

Representative images; scale bar, 10 μm. Quantifications shown in Fig. 4d. **e**, There is no significant increase in ISG transcripts *Ifit1*, *Ifit3*, *Isg15*, *Cxcl10* and *Oas1a* for cells arrested in G0 after serum starvation (experimental setup as in **c**). Transcript levels were normalized to *Hprt*. Mean ± s.e.m. One-way ANOVA, 2 degrees of freedom, $n = 3$ independent experiments; NS, not significant. Compare to Extended Data Fig. 1g, showing data for matched cycling cells assessed concurrently.
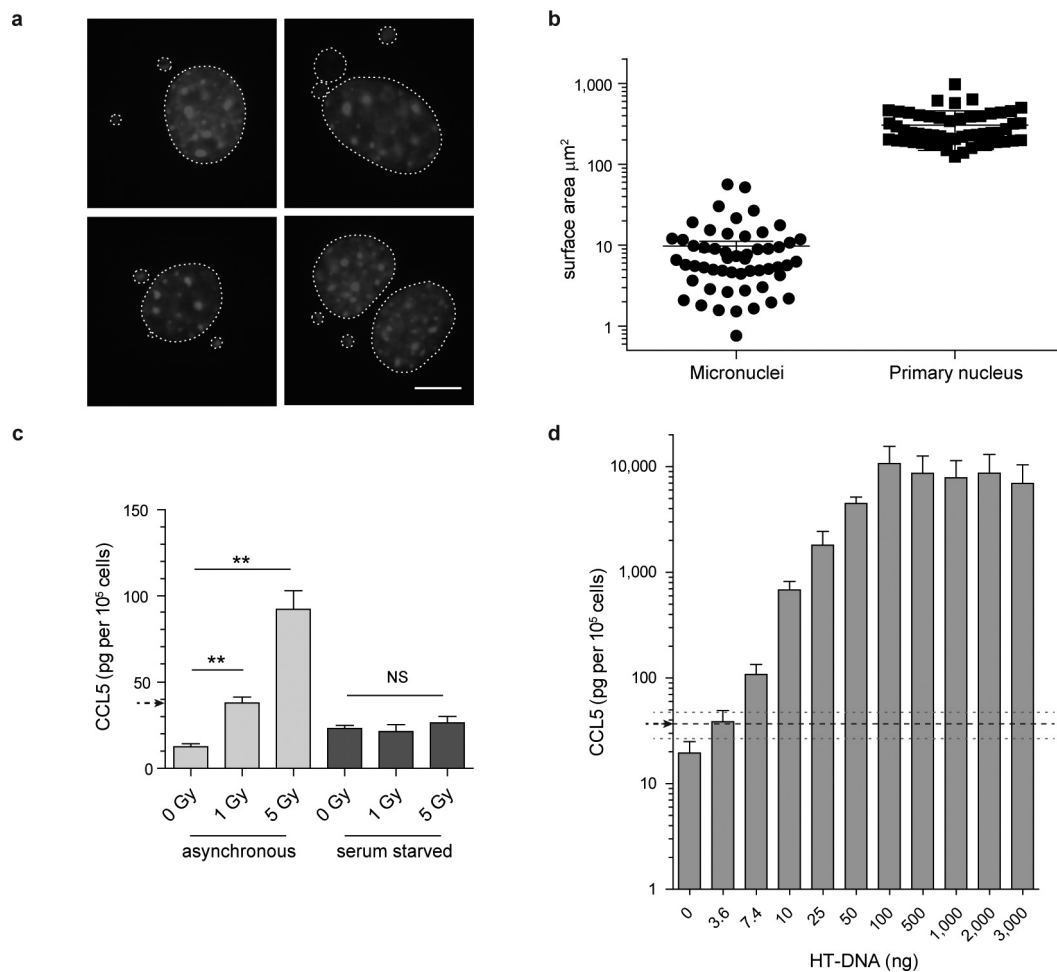
Extended Data Figure 7 | Micronuclear DNA is sufficient to account for the radiation-induced cytokine response. a, b, Measurement of micronuclear DNA content. **a**, Representative images. DAPI-stained primary nuclei and micronuclei surrounded by dotted lines. Scale bar, 10 μm. **b**, Quantification of surface areas of micronuclei and primary nuclei 48 h after 1 Gy irradiation. Micronuclear surface area per cell 9.72 ± 1.46 μm^2, primary nucleus surface area 303 ± 21 μm^2. Horizontal line and error bars: mean ± s.e.m., $n = 54$ cells. Hence, micronuclear content is ~3.2% of the total MEF genome after irradiation, equating to 190 Mbp of DNA. This corresponds to a total of 8.1 ng of micronuclear DNA in 10^5 cells after 1 Gy irradiation (10^5 diploid mouse cells contain a total of 650 ng of genomic DNA, with 39% of cells containing micronuclei, Fig. 1h). **c**, CCL5 response of wild-type C57/BL6 MEFs to ionizing radiation plotted in pg per 10^5 cells. Reanalysis of this dataset (first depicted in Fig. 4b) confirms that the prior statistical analysis is robust to data normalization on the basis of cell counts at assay endpoint. 1 Gy of irradiation in cycling MEFs results in 38 ± 5 pg (mean ± s.d.) of CCL5 per 10^5 cells. **P < 0.01, two-tailed t-test; NS, not significant. **d**, Dose–response curve of secreted CCL5 in wild-type C57BL/6 (*Trp53*$^{+/+}$) MEFs transfected with serial dilutions of transfected HT-DNA.

Therefore, around 4 ng of transfected DNA resulted in a similar level of cytokine production to **c**. Mean and 95% confidence interval indicated by black and grey dashed lines, respectively. Given the similarity of the two estimates, within the same order of magnitude, micronuclear DNA is likely to be sufficient to account for the immune response observed. Conversely, ionizing radiation would not be expected to generate this quantity of small DNA fragments as 1 Gy irradiation generates ~40 double strand breaks (DSBs)[54], and ~1,000 base lesions and single-stranded breaks. DSBs will have an average separation of 150 Mbp, and will therefore be too widely spaced to directly generate small dsDNA fragments. Repair of DNA lesions can generate small single-stranded DNA (ssDNA) fragments through endonuclease activity. The best characterized fragments are those generated by nucleotide excision repair, where endonucleolytic cleavage yields 24–32-nucleotide ssDNA fragments[55]. As such these are not an ideal substrate for cGAS activation, and 5 million such lesions per cell would have to be generated to produce 4 ng of cytosolic DNA in 10^5 cells. Hence, on the basis of our understanding of the current literature, such DNA fragments are likely to be generated at a level that is orders of magnitude lower than that of micronuclear DNA after radiation-induced damage.
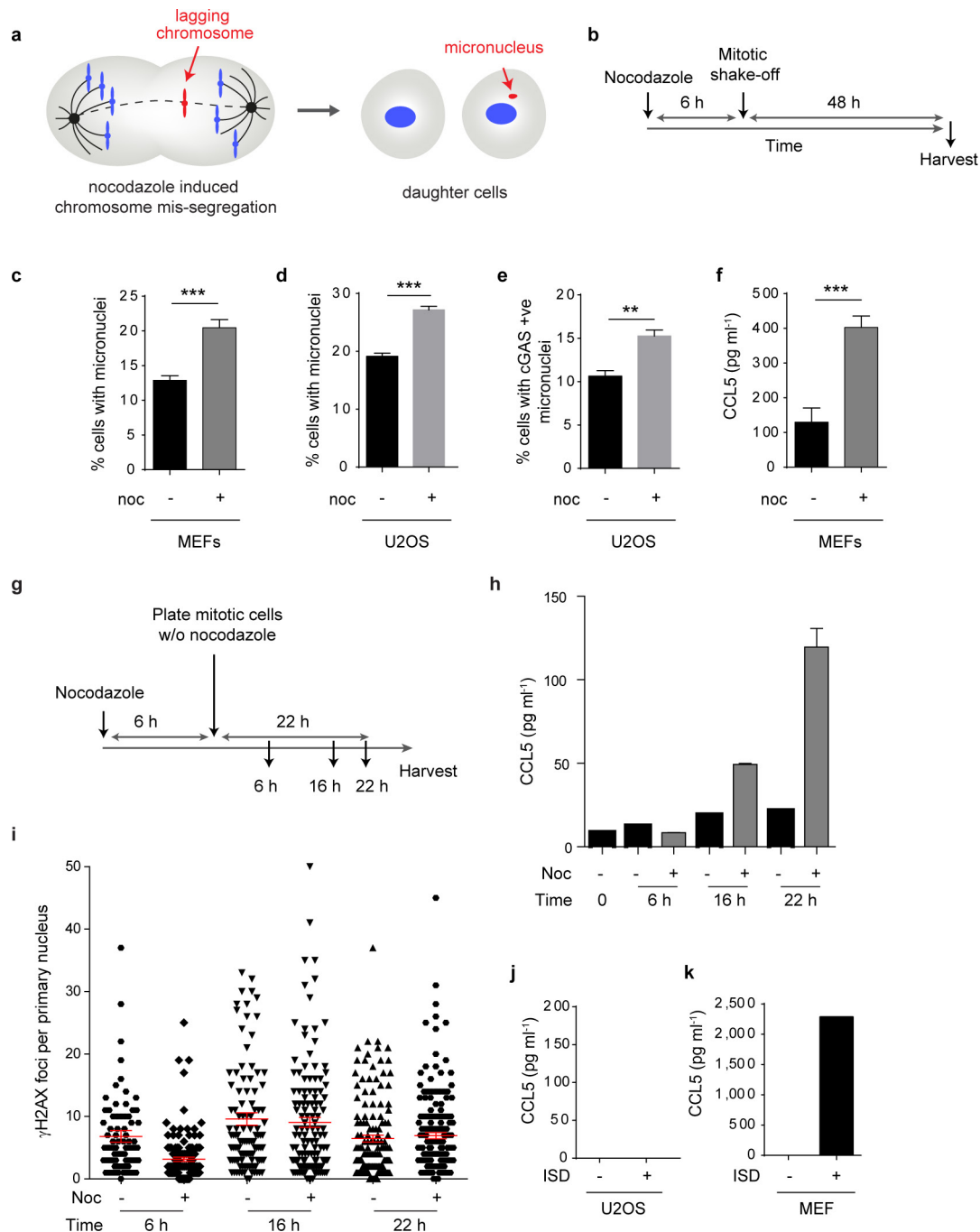
Extended Data Figure 8 | Induction of micronuclei originating from lagging chromosomes leads to a proinflammatory response, but not increased DNA damage in the primary nucleus. a, Model: micronucleus formation after nocodazole treatment. **b**, Schematic of experimental protocol. **c, d**, Percentage of micronucleated cells following nocodazole (noc) treatment of $Trp53^{-/-}$ MEFs (**c**) or U2OS cells (**d**). Mean ± s.e.m., $n = 5$ experiments for $Trp53^{-/-}$ MEFs, $n = 3$ for U2OS cells. **e**, Percentage of U2OS cells with cGAS-positive micronuclei following nocodazole treatment. Mean ± s.e.m., $n = 3$ experiments. **c–e**, ≥500 cells counted per experiment. **f**, CCL5 secretion following nocodazole treatment of $Trp53^{-/-}$ MEFs. Mean ± s.e.m. of $n = 5$ experiments. **P < 0.01, ***P < 0.001, two-tailed t-test. **g–i**, Increased CCL5 production after nocodazole release is observed after 16 h and not associated with increased

DNA damage in the primary nucleus. **g**, Experimental setup: $Trp53^{-/-}$ MEFs were arrested with nocodazole for 6 h and mitotic cells harvested by mitotic shake-off and re-plated in fresh medium with nocodazole omitted. Supernatants and cells were then collected at indicated time points after growth in medium. **h**, Increased CCL5 production was observed from 16 h after release from nocodazole block. Technical duplicate, mean ± s.d. Noc (−), asynchronously grown, plated at the same time as mitotic shake-off Noc (+) cells, arrested with nocodazole. **i**, No increase in the number of γH2AX foci in the primary nucleus was observed after release from nocodazole block. $n ≥ 100$ cells counted per condition. **j, k**, CCL5 response to interferon stimulatory DNA (ISD) is absent in U2OS cells (**j**) but present in MEFs (**k**). CCL5 measured by ELISA 8 h after transfection with ISD. $n = 2$ experiments for U2OS cells, $n = 1$ experiment for MEFs.

a

Total Feature Counts vs ERCC Percentage Counts

b

	Average Total Counts (Library Size)	Average Total Gene Features	Average Percentage ERCC Count
MN + (n = 21)	536,981	6,501	2.04
MN - (n = 14)	846,114	5,937	2.55

c

Cells with normal nuclear morphology (n=14) that passed scRNA-seq QC

d

Micronucleated cells (n=21) that passed scRNA-seq QC
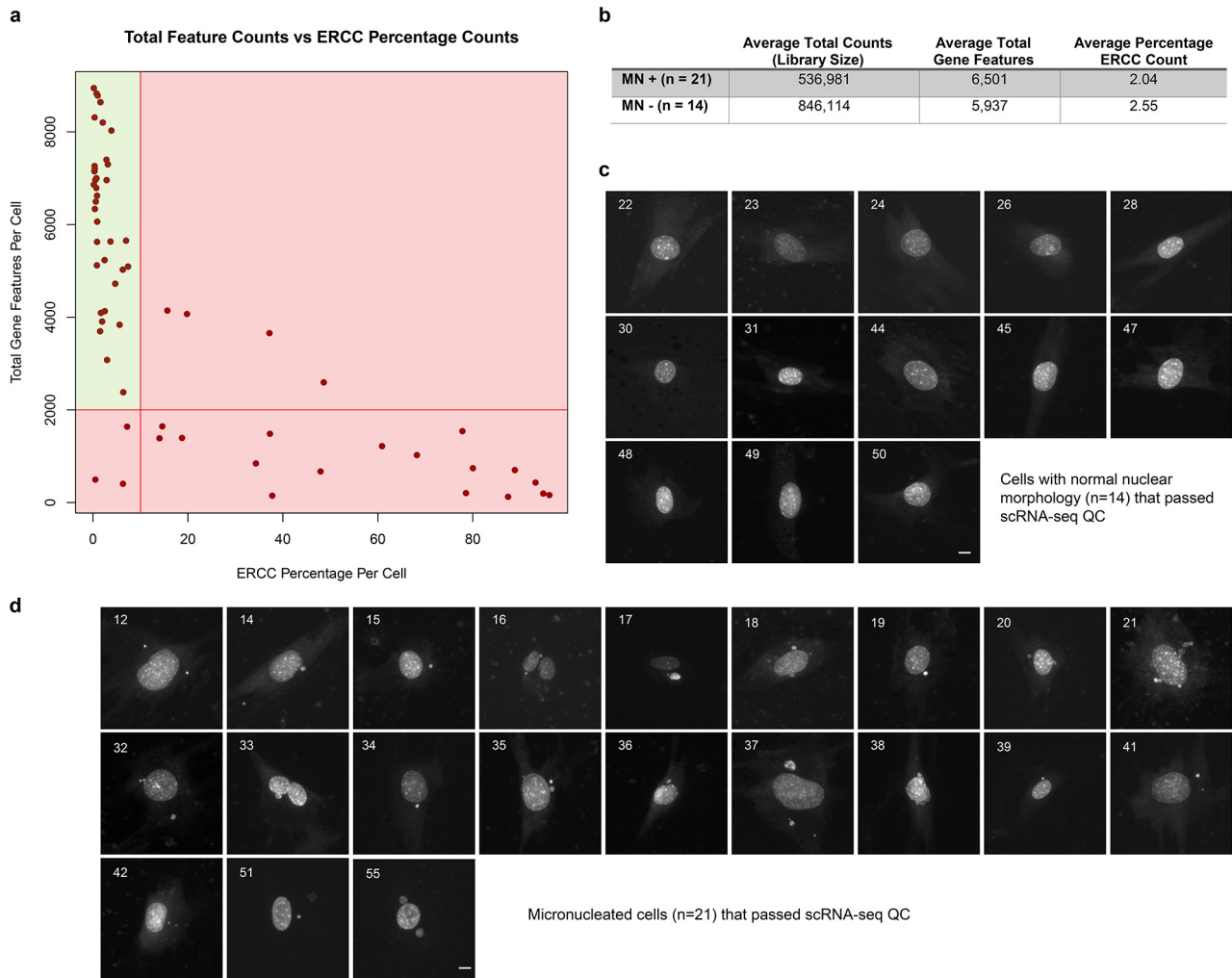
Extended Data Figure 9 | Single-cell RNA sequencing quality control and microscopy images of individual LCM-captured cells. a, Total gene feature counts (reads mapping to a protein coding gene) vs ERCC (RNA spike-in) percentage of total counts per cell. Cells with ERCC percentage counts >10% and/or with feature counts <2,000 were rejected, indicated by red shaded regions. **b**, Summary statistics for 21 micronucleated (MN+) cells and 14 non-micronucleated (MN−) cells that passed quality control. **c**, **d**, Microcopy images of cells captured by LCM that passed quality control after single-cell RNA sequencing. Fourteen live cells without micronuclei (**c**) and 21 live cells with micronuclei (**d**) were isolated from the same culture dish using LCM and used for single-cell mRNA sequencing. DNA was stained with picogreen dsDNA stain. Cells shown are those that passed quality control; numbers indicate the order in which cells were captured. Scale bars, 10 μm.
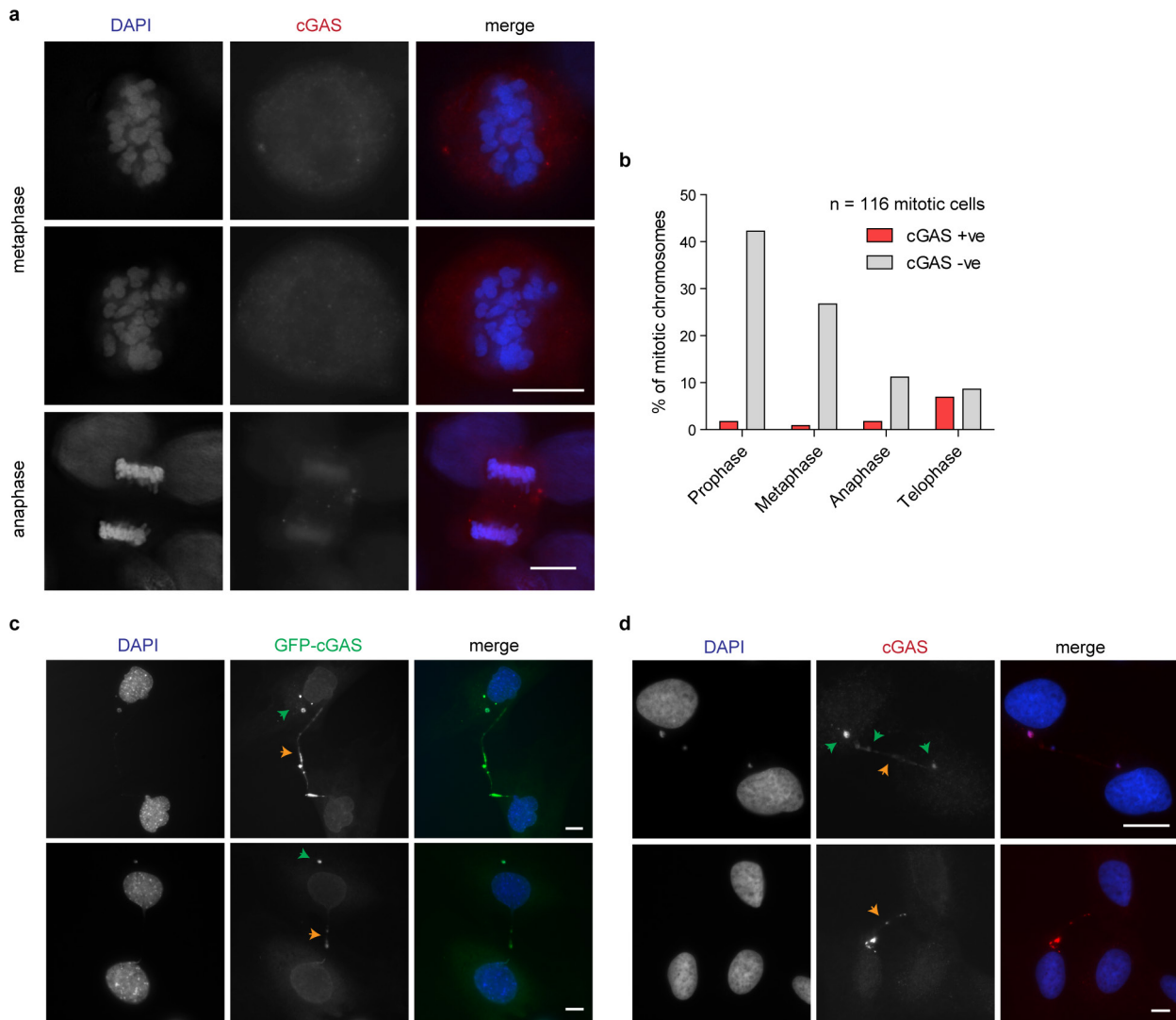
Extended Data Figure 10 | cGAS localizes to telophase chromosomes and DNA bridges. a, Endogenous cGAS was stained by immunofluorescence of U2OS cells in mitosis, showing a diffuse staining pattern without accumulation at the DAPI-stained condensed chromosomes at metaphase. Two representative images shown. During anaphase (and telophase), cGAS staining can be seen on DNA in some cells. Overexpressed GFP–cGAS also localizes more widely to mitotic DNA in U2OS cells and MEFs (data not shown). **b**, Quantification of cGAS staining during mitosis, by stage. **c**, $Rnaseh2b^{-/-} Trp53^{-/-}$ MEFs stably expressing GFP–cGAS show localization of cGAS at DNA bridges (orange arrowheads). **d**, Endogenous cGAS can also be seen to localize to DNA bridges that occasionally occur in U2OS cells. cGAS also localized to micronuclei in the same cells (green arrowheads). Interphase chromatin bridges with cGAS bound in $Rnaseh2b^{-/-} Trp53^{-/-}$ MEFs 0.08% of $n = 1,223$ cells; U2OS cells 0.06% of $n = 1,632$ cells. Scale bars, 10 μm.

Appendix 2: Wilms Tumor 1b defines a wound-specific sheath cell subpopulation associated with notochord repair

Wilms Tumor 1b defines a wound-specific sheath cell subpopulation associated with notochord repair

Juan Carlos Lopez-Baez[1,2], Daniel J Simpson[1†], Laura LLeras Forero[3,4,5†], Zhiqiang Zeng[1,2], Hannah Brunsdon[1,2], Angela Salzano[1], Alessandro Brombin[1,2], Cameron Wyatt[1], Witold Rybski[1,2], Leonie F A Huitema[3], Rodney M Dale[6], Koichi Kawakami[7], Christoph Englert[8,9], Tamir Chandra[1], Stefan Schulte-Merker[3,4,5], Nicholas D Hastie[1*], E Elizabeth Patton[1,2*]

[1]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; [2]CRUK Edinburgh Centre, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; [3]Hubrecht Institute - KNAW & UMC Utrecht, Utrecht, Netherlands; [4]Faculty of Medicine, Institute for Cardiovascular Organogenesis and Regeneration, WWU Münster, Münster, Germany; [5]CiM Cluster of Excellence, Münster, Germany; [6]Department of Biology, Loyola University Chicago, Chicago, United States; [7]Division of Molecular and Developmental Biology, National Institute of Genetics, Mishima, Japan; [8]Department of Molecular Genetics, Leibniz Institute for Age Research-Fritz Lipmann Institute, Jena, Germany; [9]Institute of Biochemistry and Biophysics, Friedrich-Schiller-University, Jena, Germany

*For correspondence:
Nick.Hastie@igmm.ed.ac.uk (NDH);
e.patton@igmm.ed.ac.uk (EEP)

†These authors contributed equally to this work

Abstract Regenerative therapy for degenerative spine disorders requires the identification of cells that can slow down and possibly reverse degenerative processes. Here, we identify an unanticipated wound-specific notochord sheath cell subpopulation that expresses Wilms Tumor (WT) 1b following injury in zebrafish. We show that localized damage leads to Wt1b expression in sheath cells, and that *wt1b*[+] cells migrate into the wound to form a stopper-like structure, likely to maintain structural integrity. *Wt1b*[+] sheath cells are distinct in expressing cartilage and vacuolar genes, and in repressing a Wt1b-p53 transcriptional programme. At the wound, *wt1b*[+] and *entpd5*[+] cells constitute separate, tightly-associated subpopulations. Surprisingly, *wt1b* expression at the site of injury is maintained even into adult stages in developing vertebrae, which form in an untypical manner via a cartilage intermediate. Given that notochord cells are retained in adult intervertebral discs, the identification of novel subpopulations may have important implications for regenerative spine disorder treatments.
DOI: https://doi.org/10.7554/eLife.30657.001

Introduction

Wilms tumour 1 (WT1) is a zinc finger transcription factor that regulates key developmental stages of several mesodermal tissues including the kidneys, gonads and coronary vasculature (*Hastie, 2017*). In the developing kidney, WT1 is required for the maintenance of mesenchymal nephron progenitors (*Kreidberg et al., 1993*; *Motamedi et al., 2014*) as well as differentiation of these progenitors into the epithelial components of the nephron (*Essafi et al., 2011*). In contrast, in the developing heart, WT1 is expressed in the epicardium (mesothelial lining) and required for the production, via an epithelial to mesenchymal transition (EMT), of coronary vascular progenitors (EPDCs) that migrate into

the myocardium (*Martínez-Estrada et al., 2010*). Similarly, WT1-expressing mesothelium is the source of mesenchymal progenitors for specialised cell types within several other developing organs. These include stellate cells within the liver (*Asahina et al., 2011*), interstitial cells of Cajal in the intestine (*Carmona et al., 2013*) and adipocytes within visceral fat depots (*Chau et al., 2014*). WT1 expression is down-regulated in the epicardium postnatally but reactivated in response to tissue damage in both mice (*Smart et al., 2011*) and zebrafish (*Schnabel et al., 2011*). In both organisms, this activation of WT1 in response to damage is associated with new rounds of epicardial EMT, leading to the production of coronary vascular progenitors (*Schnabel et al., 2011*; *Smart et al., 2011*).

Given the reactivation of *Wt1/wt1b* in the damaged epicardium we set out to investigate whether Wt1 programmes are initiated in response to other sources of tissue damage in zebrafish, and uncovered a novel Wt1 response to wounding of the notochord. The notochord is a transient embryonic structure that provides axial support and signalling information (*Stemple, 2005*). The notochord comprises two cell populations, the inner vacuolated cells that provide rigid support to the embryo, and the outer sheath cells, a single cell epithelial layer that surrounds the vacuolated cells and secretes components of the extracellular matrix to provide turgor pressure to the vacuolated cells (*Apschner et al., 2011*; *Ellis et al., 2013*). This rigid axial structure becomes functionally replaced by vertebra of the axial skeleton over time. In zebrafish, a row of metameric mineralized rings, known as chordacentra, forms around the notochord in an anterior to posterior fashion and constitutes the first signs of the definitive vertebral column. The chordacentra delineate the future sites where mature vertebra will form and ossify as the larva grows, while the notochord cells develop into the nucleus pulposus of the adult intervertebral disc, a soft gel-like tissue that provides cushioning and flexibility for the spine (*Parsons, 1977*).

Degeneration of the intervertebral disc leads to extensive back pain, one of the top global causes of years lived with disability (*Lawson and Harfe, 2015*). Treatment primarily consists of managing the pain symptoms, or in more progressed disease includes extensive surgery. One of the major goals of the tissue-engineering field is to identify cells and tissues that will enable novel regenerative therapies to slow down and possibly reverse the degenerative process. Here, we uncover a novel cellular subpopulation in the notochord sheath that emerges at the site of damage and is maintained until formation of a repaired adult vertebra structure. Surprisingly, this subpopulation expresses *wt1b* despite no evidence of *wt1b* expression in physiological notochord development or ossification. Our findings suggest that the zebrafish notochord is protected by a novel wound-specific programme that seals the notochord wound in the embryo and contributes to the subsequent adult vertebra at the injury site.

Results

Wound-specific expression of *wt1b* in the notochord

Given the expression of *wt1b* in the regenerating heart, we wanted to explore the expression of *wt1* in other regenerating tissues, and began with the tail fin regenerative processes. There are two *wt1* paralogues in zebrafish, *wt1a* and *wt1b*, and so we performed tail fin amputations on zebrafish larvae 3 days post fertilization (dpf) using *Tg(wt1a:gfp)* and *Tg(wt1b:gfp)* transgenic lines (*Bollig et al., 2009*; *Perner et al., 2007*) (*Figure 1—figure supplement 1a*). Surprisingly, we discovered that tail fin amputations including partial removal of the notochord triggered a change of cellularity in the notochord, coupled with the specific, de novo upregulation of GFP in a *Tg(wt1b:gfp)* transgenic line. This response was specific to *wt1b* because we did not observe expression of GFP in the notochord of *Tg(wt1a:gfp)* tail fin amputated larvae (*Figure 1—figure supplement 1b–f*).

Next, we developed a needle-based assay to induce localized damage in the developing zebrafish notochord independent of tail fin amputation. Needle injury was induced in 3 dpf *Tg(wt1b:gfp)* that had been crossed with *casper* fish to remove pigmentation and imaged at 72 hr post injury (hpi) (*Figure 1a*). Needle induced wounds triggered a similar, albeit stronger *wt1b:gfp* response compared to the tail fin amputations, that was specifically localised to the site of the wound (*Figure 1b*). Time course imaging showed a progressive expansion of the damaged area over 72 hr, with an increasing expression of GFP signal, concomitant with a change of cellularity in the notochord (*Figure 1c*). Importantly, this was not observed in uninjured zebrafish controls (*Figure 1c*) or in notochord injured *Tg(wt1a:gfp)* transgenic larvae (data not shown). Histological staining of the damaged
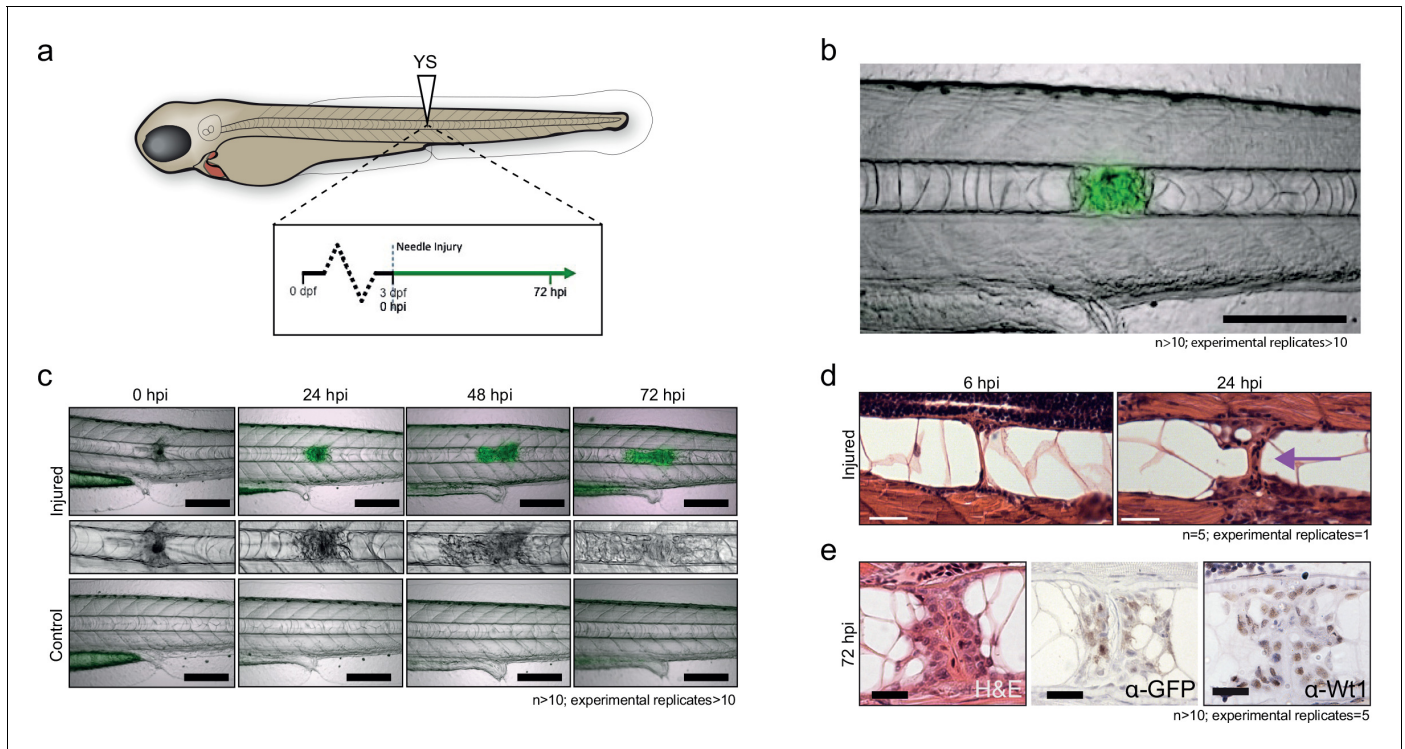
Figure 1. Notochord injury triggers local and sustained *wt1b* expression. (**a**) Schematic of notochord needle-injury protocol. 3 dpf *Tg(wt1b:gfp)*; casper larvae are injured above the yolk sac (YS; at somite 14 or 15) and followed for 72 hr. (**b, c**) Images of *Tg(wt1b:gfp)*; *casper* zebrafish trunk over time following notochord needle injury, and uninjured matched controls. GFP signal is associated with a change of cellularity in the injured notochord (inset). n > 10; experimental replicates >10. Scale bar: 100 µm. (**d**) H and E staining of the injured area at 6 hpi and 24 hpi highlighted the progressive change in cellularity at the site of the injury (arrow). n = 5; experimental replicates = 1. Scale bar: 20 µm. (**e**) Immunohistochemistry of the injured area with α-GFP and α-Wt1 antibodies. n > 10; experimental replicates = 5. Scale bar: 20 µm. dpf = days post fertilization; hpi = hours post injury; H and E = haematoxylin and eosin.

DOI: https://doi.org/10.7554/eLife.30657.002

The following figure supplements are available for figure 1:

Figure supplement 1. *wt1b* expression in tail amputated larvae.
DOI: https://doi.org/10.7554/eLife.30657.003

Figure supplement 2. Wt1 and GFP protein expression in the notochord.
DOI: https://doi.org/10.7554/eLife.30657.004

area revealed the presence of a subpopulation of cells at the site of injury, which contrasted morphologically with the uniform, vacuolated inner cells of the notochord (*Figure 1d*). These cells stained positively for GFP and for endogenous Wt1 protein by immunohistochemistry, validating the faithful expression of the transgene with endogenous *wt1b* expression in this response (*Figure 1e*; *Figure 1—figure supplement 2*). *Tg(wt1b:gfp)* expression was not detected in the notochord outside the wound response by immunohistochemistry for GFP or for Wt1 protein (*Figure 1—figure supplement 2*). Thus, following notochord injury, an unanticipated expression of *wt1b* marks a subpopulation of cells that emerges in the notochord and is associated with the wound.

wt1b expressing cells emerge from the notochord sheath

To determine the origin of the wound-specific *wt1b*+ cells, we examined *wt1b* expression in the vacuolated cells of the notochord, and in notochord sheath cells using two different transgenic lines. The *Tg(SAGFF214A:gfp)* transgenic line labels the cytoplasm of the inner vacuolated cells, and the *Tg(R2col2a1a:mCherry)* transgenic line labels notochord sheath cells. While *col2a1a* is expressed in all notochord cells (*Apschner et al., 2011*), a *Tg(R2col2a1a:mCherry)* line had been generated with a 310 bp conserved regulatory element of the *col2a1a* promoter that is specifically expressed in the

surrounding notochord sheath cells (*Figure 2a*) (*Dale and Topczewski, 2011*; *Yamamoto et al., 2010*).

A needle-induced notochord wound in the *Tg(SAGFF214A:gfp)* transgenic line showed that GFP-expressing cells were lost rapidly upon injury, creating a gap in the row of vacuolated cells. Eventually, this gap was filled with new cells by 144 hpi (*Figure 2—figure supplement 1a,b*). The *SAGFF214A:gfp* response was distinct from the *wt1b⁺* response in time (emerging at 72 hpi compared with 24 hpi), size and number (few and large compared with numerous and small), and in coverage of the wound (visible gaps remaining at the site compared with filling the damage site). These data suggest that *wt1b* expressing cells are distinct from the vacuolated cells at the site of injury.

Next, we explored the role of the notochord sheath cells in this process. We crossed the *Tg(wt1b:gfp)* transgenic line to the *Tg(R2col2a1a:mCherry)* transgenic line. Live confocal and multiphoton imaging revealed *wt1b:gfp* expression in the *R2col2a1a:mCherry* notochord sheath cells following needle induced notochord damage (*Figure 2b–d*; *Video 1*; *Figure 2—figure supplement 1c*), and this was supported by imaging of histological sections (*Figure 2—figure supplement 1d*). *wt1b:gfp* co-expression with *R2col2a1a:mCherry* was visible by 24 hpi in a ring surrounding the notochord vacuolated cells, and by 72 hpi the *wt1b:gfp* subpopulation of sheath cells had migrated into central aspects of the notochord to fill the wound and produce a visible stopper-like seal that was contiguous with the notochord sheath cells, and filled the gap in the notochord caused by the wound (*Figures 1e* and *2d*).

To validate the co-expression of *wt1b:gfp* and *col2a1a:mCherry* in the wounded fish, we FACS sorted cell populations in the injured versus uninjured larvae isolated from the trunk region (*Figure 2e*; 35 larvae pooled per set). Both injured and non-injured larvae contained cells that expressed either GFP⁺ only (presumably *wt1b:gfp* cells of the pronephric duct that were included in the dissected tissue) or mCherry⁺ alone, but the wounded fish had significantly increased numbers of cells that co-expressed *wt1b:gfp* and *col2a1a:mCherry* (GFP⁺mCherry⁺) (*Figure 2—figure supplement 1e*).

Our evidence indicates that the notochord wound triggers a unique *wt1b⁺* subpopulation to emerge in the notochord sheath cells. This *wt1b⁺* sheath cell subpopulation migrates into the wound and generates a stopper-like structure, possibly to prevent further loss of notochord turgor pressure and maintain notochord integrity.

Nystatin mediated disruption of vacuolated cells leads to an increase in *wt1b:gfp* expression

We tested if the *wt1b*-response was specific to wounds that involved rupture of the sheath, or if *wt1b* expressing cells could be induced upon loss of vacuolated cell integrity alone. Mutations in *caveolin* genes lead to collapse of the vacuolated cells, with invasion and replacement from the notochord sheath (*Garcia et al., 2017*). We treated two-day old *Tg(wt1b:gfp; R2col2a1a:mCherry)* zebrafish with nystatin, a small molecule that binds sterols. Nystatin treatment lead to an increase in cellularity of the vacuolated notochord, similar to the phenotype seen in the notochord of *caveolin* mutants (*Figure 2—figure supplement 2*). GFP was expressed in a subpopulation of the mCherry-positive sheath cells at the site of cellularity. Thus, expression of *wt1b* in the sheath does not require a physical breach of the sheath, and *wt1b* expression may be applicable to a wider range of tissue stress and damage situations.

Notochord wound cells express cartilage and mesenchyme genes

To address the molecular process at the site of the wound, we compared the transcriptome of
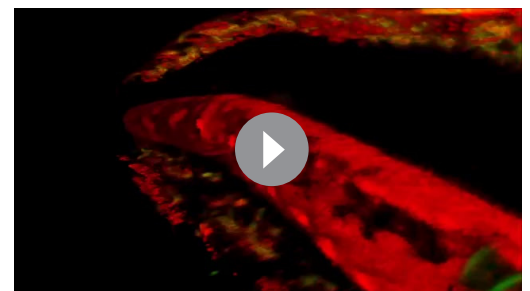
Video 1. Time-lapse imaging of two-photon microscopy of *Tg (wt1b:gfp; R2col2a1a:mCherry)* zebrafish larvae following needle injury over 48 hr. *wt1b:gfp* expression is upregulated in *R2col2a1a:mCherry* expressing notochord sheath cells upon needle injury, leading to the formation of a stopper like structure across the wound
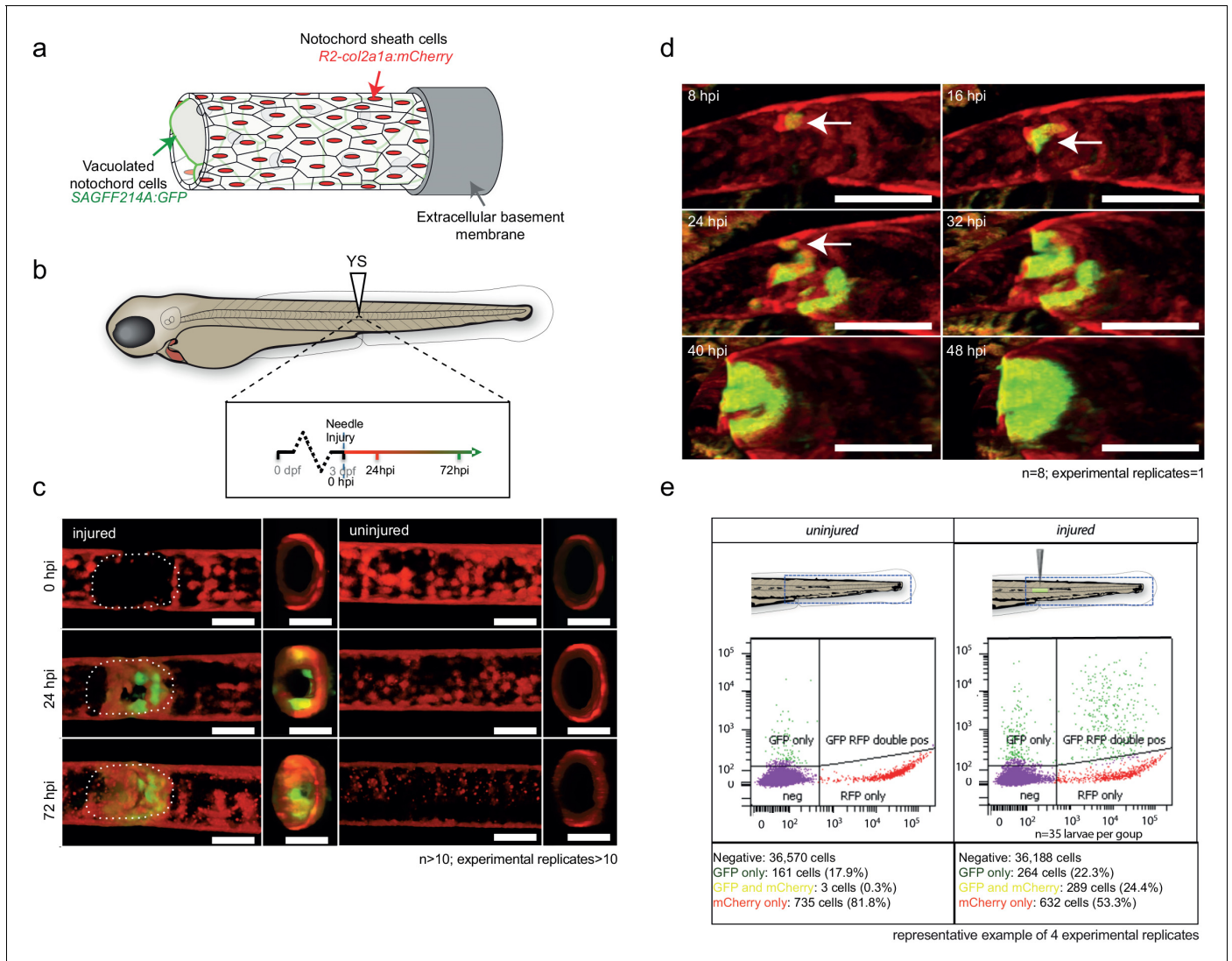
DOI: https://doi.org/10.7554/eLife.30657.009

Figure 2. *wt1b:gfp* expressing notochord sheath cells populate the site of injury in the damaged notochords. (**a**) Schematic diagram of the notochord and transgenic lines used in this study. The notochord is composed of an inner population of highly vacuolated cells (green arrow; *SAGFF214A:gfp*), surrounded by a layer of epithelial-like sheath cells (red arrow; *R2col2a1a:mCherry*), encapsulated by a thick layer of extracellular basement membrane (grey arrow). (**b**) Schematic of experimental design: 3dpf *Tg(wt1b:gfp; R2-col2a1a:mCherry); casper* larvae were needle-injured and imaged at 0, 24 and 72 hpi. (**c**) Needle damage led to the formation of a cell-free gap in the layer of notochord sheath cells (0 hpi – injured; dashed line). GFP expression can be observed in the notochord sheath cells surrounding the area of damage by 24 hpi (inset: cross-sectional view) and these appear to engulf the injured area by 72 hpi (inset). n > 10; experimental replicates >10. Scale bar: 100 μm. (**d**) Multiphoton time-lapse imaging of wound site. Initial upregulation of GFP occurs at eight hpi in the *R2-col2a1a:mCherry* positive cells (arrow) and propagates across the injured area over the next 40 hr to form a seal in the notochord. n = 8; experimental replicates = 1. Scale bar: 100 μm. (**e**) Representative example of FACS analysis of cell populations in injured and non-injured zebrafish trunk tissue. GFP$^+$mCherry$^+$ double positive cells are present in injured *Tg(wt1b:gfp; col2a1a:mCherry)* at 72 hpi. Percentage of fluorescent cells are reported. Note that the dissected tissue can also encompass *wt1b:gfp* expressing cells in the posterior end of the pronephric duct (see also *Figure 1c*). n = 35 larvae per group; experimental replicates = 4. dpf = days post fertilization; hpi = hours post injury.

DOI: https://doi.org/10.7554/eLife.30657.005

The following source data and figure supplements are available for figure 2:

Figure supplement 1. Imaging cell populations at the wound.

DOI: https://doi.org/10.7554/eLife.30657.006

Figure supplement 1—source data 1. Raw data and statistical analyses for *Figure 2—figure supplement 1e*.

DOI: https://doi.org/10.7554/eLife.30657.008

Figure supplement 2. Nystatin treatment leads to upregulation of *wt1b:gfp* expression in notochord sheath cells.

DOI: https://doi.org/10.7554/eLife.30657.007

the trunk region in the injured and uninjured 72 hpi larvae (*Figure 3a,b*; n = 50 larvae per subset). Microarray analysis revealed a highly significant 131-fold increase in expression of *matrix gla protein* (*mgp*), a gene that is known to express in chondrocytic zebrafish tissues (*Gavaia et al., 2006*) and to be involved in the inhibition of hydroxyapatite production during ectopic bone formation (*Schurgers et al., 2013*; *Sweatt et al., 2003*; *Zebboudj et al., 2002*) (*Figure 3c,d*). Other genes included mesenchymal and cell adhesion markers, such as *fn1b*, coagulation factors, such as *f13a1b*, and immune response genes, such as *zgc:92041* and *complement c6* (*Figure 3d*).

The increased expression of *mgp* and *f13a1b* genes implicated the de novo acquisition of chondrogenic features in the injured tissues. Chondrogenic cells in the endochondral tissues of the craniofacial, fin bud and axial skeletons express *mgp* (*Gavaia et al., 2006*) and *FXIIIA* expression is localized to the developing chondrogenic mesenchyme of the pectoral fin bud (*Deasey et al., 2012*). The expression of cartilage genes was unexpected because ossification around the zebrafish notochord occurs via the formation the chordacentra, and does not require the establishment of cartilage anlagen (*Bensimon-Brito et al., 2012*; *Fleming et al., 2004*). To examine the expression of other chondrogenic genes, we analyzed the top 100 significant genes and found an increase in expression of *sox9b*, the master regulator of chondrogenesis, five collagen genes associated with chondrogenic tissues (*col2a1a*, *col2a1b*, *col11a2*, *col9a1* and *col9a2*), the cartilage-specific extracellular structural protein Aggrecan, a microRNA regulator of chondrogenesis microRNA140 and the matrix-cell anchor protein chondroadherin (*chad*) (*Figure 3e*). To validate these findings at the molecular level, we isolated sections of damaged and undamaged tissue, and performed qRT-PCR for *matrix gla protein* (*mgp*) and *sox9b*. We chose these two genes because *mgp* was highly expressed in the microarray analysis and important for bone organization, and because Sox9 is a master cartilage transcription factor. We found *mgp* and *sox9b* to be highly upregulated in the injured tissue compared with the uninjured tissue (*Figure 3f,g*). Our results reveal that notochord wounding leads to the formation of a *wt1b*-positive sheath subpopulation that is characterised by an unexpected increase in genes associated with cartilage.

Single-cell and 10 cell sequencing of *wt1b*-expressing sheath cells

To address the molecular nature of the GFP⁺mCherry⁺ expressing cells, we performed RNA sequencing of single-cells and 10 cell pools of FACS sorted GFP⁺ cells, mCherry⁺ cells and GFP⁺mCherry⁺ cells from injured zebrafish (3dpi) using the SMARTseq2 protocol (*Supplementary file 1*; *Figure 4—figure supplement 1*) (*Kirschner et al., 2017*; *Picelli et al., 2013*). To avoid batch effects, all experimental conditions were sorted onto the same 96 well plate and processed simultaneously (*Baran-Gale et al., 2017*). Sequencing reads were processed using the Scater pipeline (*McCarthy et al., 2017*). Unbiased Single cell consensus clustering (SC3) of the whole transcriptomes revealed that the GFP⁺ cells, mCherry⁺ cells and GFP⁺mCherry⁺ cells clustered into three distinct subpopulations (SC3 cluster 1: GFP⁺, 2: GFP⁺mCherry⁺ and 3: mCherry⁺) (*Figure 4a–c*) (*Kiselev et al., 2017*). Single and 10 cell populations clustering together suggested that sorting conditions led to homogenous 10 cell populations. Expression of *wt1b* was detected in SC3 clusters 1 and 2, and *col2a1a* was expressed in SC3 clusters 2 and 3 (*Figure 4b*). *wt1a* transcripts were not detected in any of the SC3 clusters. Together with the Wt1b antibody immunohistochemistry (*Figure 1e*, *Figure 1—figure supplement 2*), detection of *wt1b* transcripts in GFP⁺mCherry⁺ cells prove endogenous *wt1b* expression in the notochord damage response.

To avoid confounding factors, for example different ratios of single to 10 cell trancriptomes, when calculating differential expression, we used SC3 on the 10 cell populations only. We found consistent clustering of the different cell populations (GFP⁺, GFP⁺mCherry⁺ and mCherry⁺). Notably, differential marker gene expression in GFP⁺mCherry⁺ cells included the *mgp*, *fn1b* and *f13a1b* genes (*Figure 4c*) that were highly upregulated in the wounded tissue (*Figure 3d*). To validate our findings, we isolated injured notochord tissue from 3dpi and FACS sorted GFP⁺, mCherry⁺ and GFP⁺mCherry⁺ double positive cells, and performed qRT-PCR on sorted cell populations for *mgp*, a SC3 cluster 2 cell marker gene. Expression of *mgp* was selectively enriched in GFP⁺mCherry⁺ double positive cells (*Figure 4d*).

We next calculated differentially expressed genes between GFP⁺mCherry⁺ cells compared with the mCherry⁺ cells using SCDE (*Kharchenko et al., 2014*). Based on the SCDE output genes were ranked and the ranked list was used with the WEB-based gene set analysis toolkit (WebGestalt) to explore the functional nature of the GFP⁺mCherry⁺ cells compared with the mCherry⁺ cells
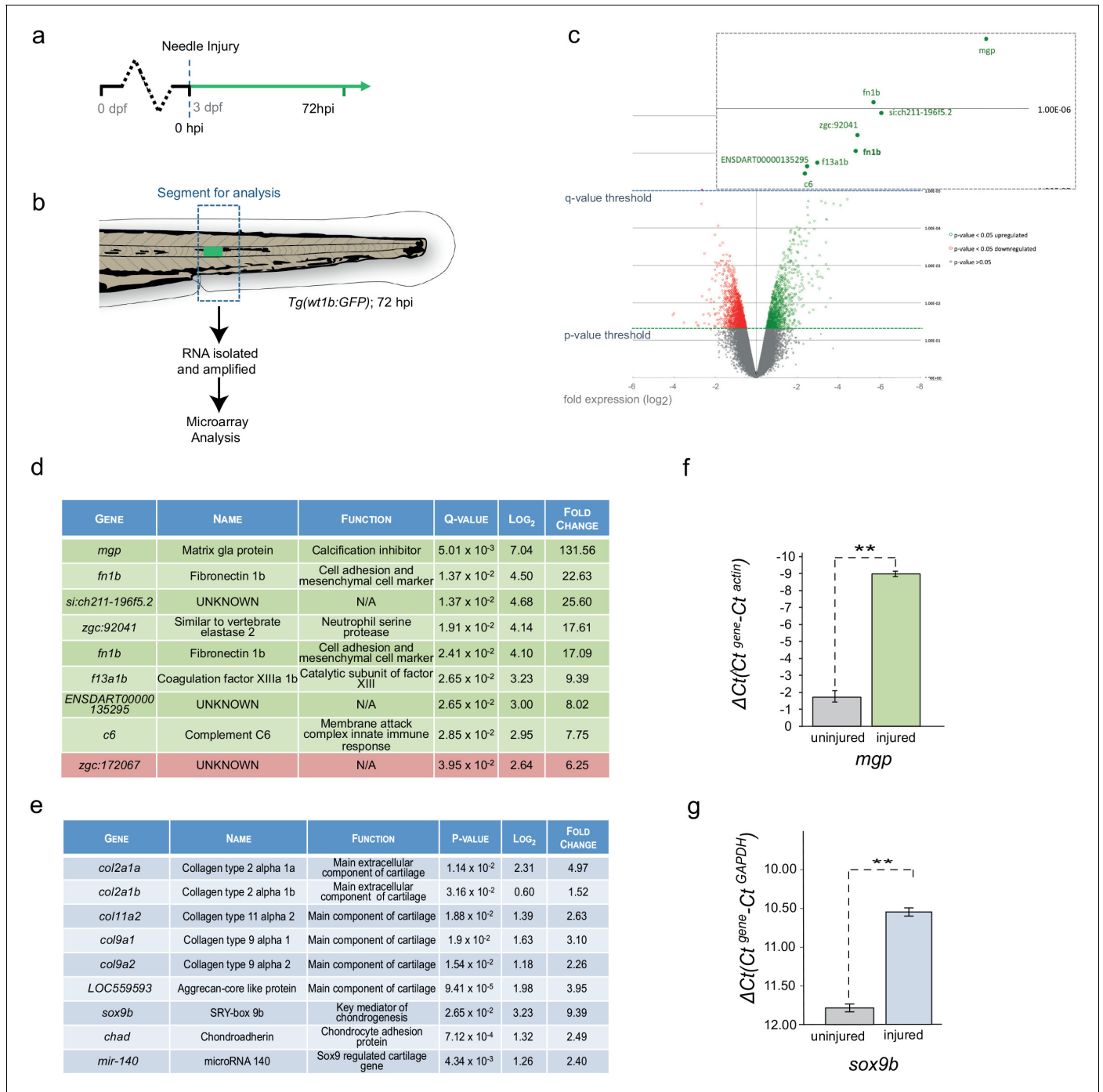
Figure 3. Cartilage genes are expressed in the notochord-injured zebrafish. (**a**) Experimental plan: 3 dpf *Tg(wt1b:gfp)* larvae were needle injured and grown for 72 hr with uninjured age-matched controls (n = 50 larvae per group). (**b**) The area around the *wt1b:gfp* expression was excised at 72 hpi (dotted area) and RNA was extracted and amplified. A similar area was taken from age-matched uninjured controls. (**c**) Volcano plot displaying the differentially expressed genes between injured and non-injured larvae. The y-axis measures the mean expression value of log 10 (p-value) and separates upregulated from downregulated genes. The x-axis represents the log2 fold change of expression. Significantly upregulated genes are shown as green circles or dots and downregulated genes are shown as red circles or dots. Green dotted line represents the p-value threshold (p<0.05) and blue dotted line represents the false discovery rate (FDR) or q-value threshold (q < 0.05). Genes with highest expression change are shown in magnified view. (**d**) Table showing the most significantly differentially expressed genes in injured larvae (q < 0.05). Upregulated genes are shown in green and downregulated genes are shown in red. (**e**) Table showing cartilage-associated genes that were significantly upregulated in the injured larvae (p<0.05). (**f, g**) Results of quantitative real-time PCR (qRT-PCR) of *mgp* and *sox9b*. The y-axis indicates the difference between the cycle threshold (Ct) value of

Figure 3 continued on next page

(*Figure 4e*). Expression of genes in signaling pathways, such as the TGF-ß pathway were reduced, while vacuolar and lysosomal pathway components were highly enriched in the GFP⁺mCherry⁺ cells comparing gene sets from multiple databases. To explore the possibility of lysosome activity in more detail, we performed confocal imaging analysis of the wound site at 7 dpi and observed some GFP⁺mCherry⁺ cells with large inclusions (presumably vacuoles), in the cytoplasm (*Figure 4f*). This suggests that some GFP⁺mCherry⁺ cells may become vacuolated to replace those lost upon injury.

Next, given the expression of cartilage genes by microarray analysis, we performed gene set enrichment analysis (GSEA) with a list of zebrafish cartilage genes curated in AmiGO (*Supplementary file 1b, 1c*). Cartilage genes were significantly enriched in the cell cluster 2 (GFP⁺mCherry⁺ cells) compared with cell cluster 3 (mCherry⁺ cells), suggesting that it is specifically the *wt1b*-expressing sheath cells that express genes involved in cartilage formation (*Figure 4g*).

To explore the role of WT1 in the wound response, we compiled a list of WT1 target genes, and compared it with the rank order list of RNA transcripts expressed in the GFP⁺mCherry⁺ cells by gene set enrichment analysis (GSEA) (*Supplementary file 1b, 1d*) (*Subramanian et al., 2005*). Unexpectedly, we discovered a set of WT1 regulated genes that were specifically repressed in the GFP⁺mCherry⁺ cells (*Figure 4h*). WT1 can function with co-factors to repress or activate gene expression, and this new signature suggests that Wt1b may function as a repressor in the notochord damage response. Next, we performed gene expression analysis for all WT1 co-transcription factors described in (*Toska and Roberts, 2014*), and found *p53* to be most differentially expressed in GFP⁺mCherry⁺ cells compared with mCherry⁺ cells (*Figure 4i,j*). GSEA analysis showed that p53 target genes are enriched overall in the GFP⁺mCherry⁺ cell populations (*Figure 4k*; *Supplementary file 1b, 1e*), however, when we specifically analysed the gene expression for those genes that were present in both the WT1 and p53 target gene list (*Supplementary file 1f*), we found a strong repression of genes that are regulated by both WT1 and p53 (*Figure 4l*). These data uncover an unexpected co-operation between Wt1b and p53 to negatively regulate a select subset of genes in the *wt1b*-expressing sheath cell subpopulation during the wound response.

Vertebra form at the repair site via an unusual cartilage intermediate

The expression of cartilage genes in the wound tissue and in the *wt1b*-expressing sheath cell subpopulation suggests that the notochord wound may induce a previously unknown and alternative bone development process. We stained injured and control animals with alcian blue and alizarin red, which highlight cartilage and bone respectively. Cartilage was clearly visible at the site of injury as soon as three dpi (*Figure 5a*). This staining was significantly stronger and distinct from the highly coordinated segmental cartilage staining that normally occurs during larval development in the region of the future intervertebral discs, which is clearly visible in both injured and non-injured controls by 14 dpi (*Figure 5a*). Similarly, the alizarin red dye identified the anterior to posterior forming chordacentra rings during larval development. However, in injured zebrafish larvae, the normally uniform mineralization pattern was interrupted around the site of damage, leading to delayed formation of the chordacentra at later stages (*Figure 5a*). By 18 dpi, the injured site began to express bone matrix, and was visibly flanked by cartilage expressing segments (*Figure 5b*). This is unusual because during norm-physiological development of the vertebral elements, cartilage and bone stains mark distinct regions of the notochord.
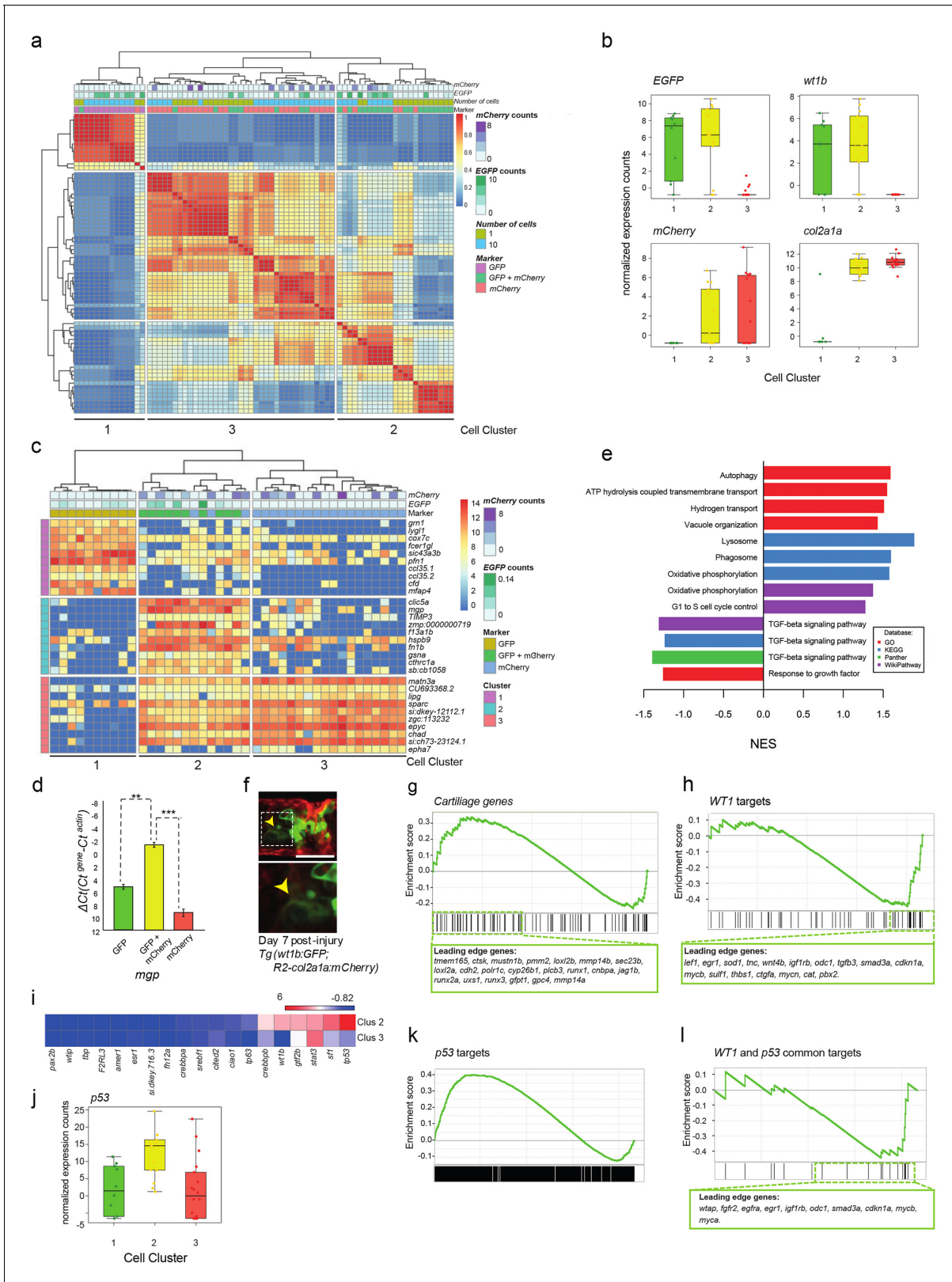
Figure 4. Single-cell and 10 cell sequencing of *wt1b*-shealth cell populations. (**a**) Single-cell and 10 cell SC3 unbiased clustering analysis reveals three distinct cell populations marked by *GFP* (cluster 1), *mCherry* (cluster 3), or *GFP* and *mCherry* (cluster 2). (**b**) *GFP*, *mCherry*, *wt1b* and *col2a1a* expression
Figure 4 continued on next page

To evaluate the outcome of the injury in the ossification process, wild-type larvae were injured and stained using live calcein dye at 21 and 38 dpi (*Du et al., 2001*). The vertebrae that eventually formed were often smaller in a given space interval and appeared supernumerary compared with uninjured age-matched controls (*Figure 5c–e*).

The notochord patterns spine formation via the activation of various signals, and has been proposed to be an essential component of chordacentra formation (*Bensimon-Brito et al., 2012*; *Fleming et al., 2004*). Entpd5a (ectonucleoside triphosphate diphosphohydrolase 5) is an E-type NTPase that is expressed in osteoblasts and is essential for skeletal morphogenesis (*Huitema et al., 2012*). Recent evidence shows that metameric expression of *entpd5a* in notochord sheath cells is an essential requirement for the patterned formation of chordacentra rings (*LL-F and SS-M, personal communication*), with *entpd5a* expression serving as a readout for mineralizing activity (*Huitema et al., 2012*). We crossed the *Tg(wt1b:gfp)* transgenic line to a *Tg(entdp5a:pkRed)* line and followed the wound response. *wt1b* and *entpd5a* expressing cell populations were closely associated at the wound site indicating that mineralizing *entpd5a* cells may directly contribute to *wt1b*⁺-associated chordacentra response (*Figure 6a,b*).

Next, we wanted to explore the relationship between *entpd5a* expression domains and the vertebrae formation at the wound site. By 5 dpf, metameric *entpd5a* expression domains are clearly visible in the anterior notochord. We wounded the notochord in 5 dpf and 7dpf fish either in between two adjacent *entpd5a*-expression domains or aimed at the center of an *entpd5a*-expression domain. Fish that had been wounded between the *entpd5a*-expression domains appeared to have normal vertebrae structures at 25 dpi (n = 6/6). In contrast, damaging the *entpd5a*-expression domain led to a supernumerary vertebra at the wound site (n = 4/4; *Figure 6—figure supplement 1*).

Taken together, these results indicate that wounding alone is not sufficient to alter the vertebrae number, and that *entpd5a* expression domains likely play a role in vertebrae formation following injury. These experiments raise the possibility that the notochord wound assay at 3 dpf disrupts an as of yet unknown precursor cell population. Up-regulation of *entpd5a* at the damage site may be part of a patho-physiological wound repair response that disrupts and/or engages with a precursor cell population (such as the metameric *entpd5a* expression) leading to altered vertebra(e) in the adult.

wt1b⁺ expression perdures into the adult vertebrae

We noticed that the *Tg(wt1b:gfp)* transgene expression was always associated with the site of vertebrae formation in the injured zebrafish that were raised to adulthood. To determine if *wt1b*
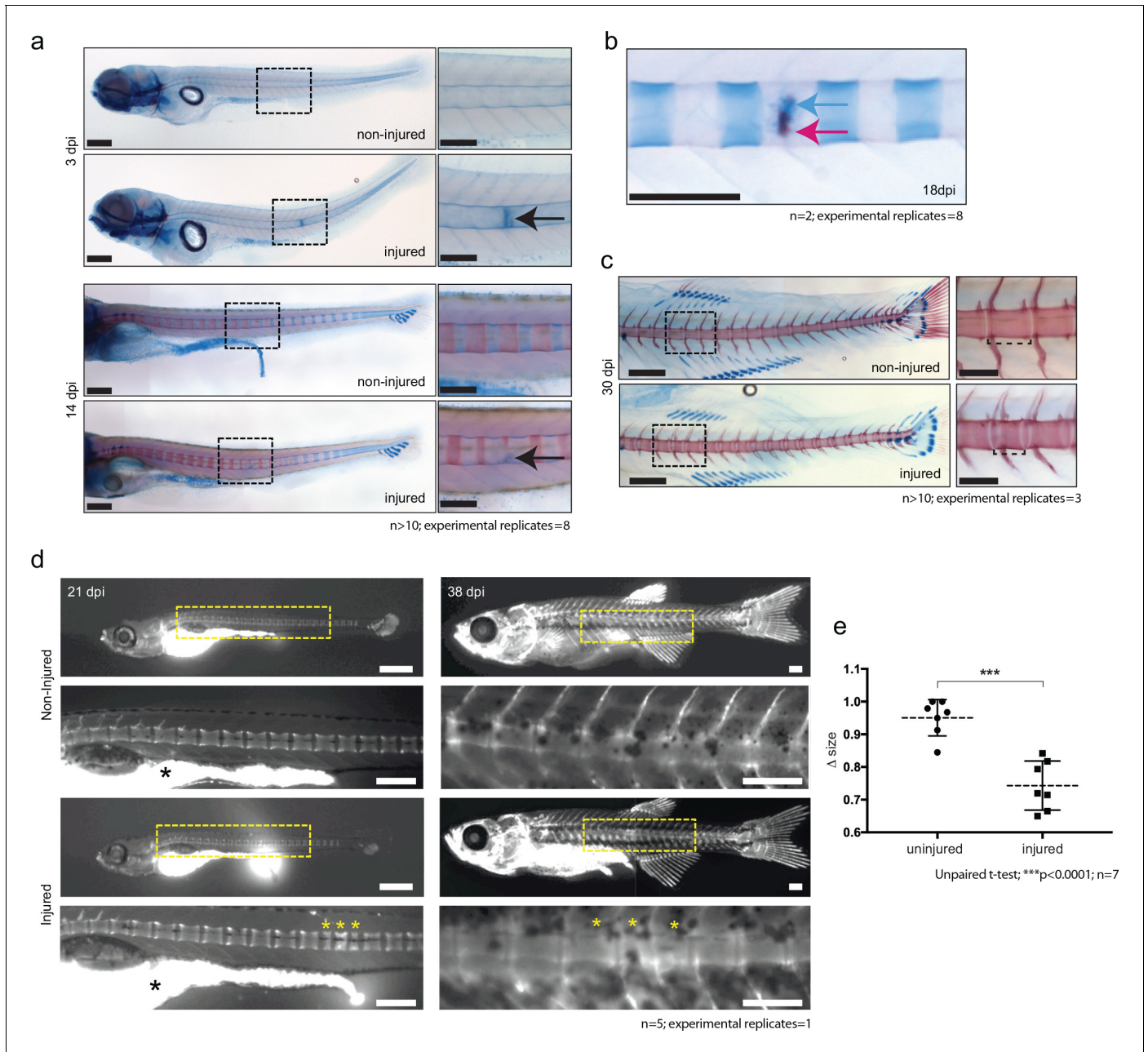
Figure 5. De novo bone formation occurs via a cartilage intermediate at the site of injury. (**a**) Alcian blue and Alizarin red staining at the site of injury in 3 and 14 dpi larvae. Ectopic cartilage deposit is indicated by arrow. n > 10; experimental replicates = 8. Scale bar left panels: 400 µm; scale bar right panels (zoomed images): 200 µm. (**b**) Alcian blue and Alizarin red staining at the site of injury at 18 dpi indicates the presence of bone and cartilage at the repair site (blue arrow = cartilage; red arrow = bone). n = 2; experimental replicates = 8. Scale bar: 200 µm. (**c**) Alcian blue and alizarin red staining of 30 dpi larvae reveals the formation of a smaller vertebra in the damaged area. n > 10; experimental replicates = 3. Scale bar left panels: 400 µm; scale bar right panels (zoomed images): 200 µm. (**d**) Live imaging of calcein stained zebrafish at 21 and 38 dpi in injured and uninjured fish. Vertebrae at damage site are indicated by yellow asterisks. Black asterisk denotes intestinal fluorescence. n = 5; experimental replicates = 1. Scale bar 21 hpf: 200 µm; scale bar 21 hpf zoomed: 100 µm; scale bar 38 hpf: 200 µm; scale bar 38 hpf zoomed: 100 µm. (**e**) The relative vertebra size difference (Δ size) between vertebrae at the site of injury (injured) and vertebrae in non-injured areas (uninjured). Vertebrae at the site of injury were significantly smaller than uninjured vertebrae (Unpaired t-test; ***p<0.0001 two-tailed; mean ±SEM uninjured larvae = 0.9506 + /- 0.02102 n = 7; mean ±SEM injured larvae = 0.7432 + /- 0.0284 n = 7; measurements taken at 30 and 38 dpi).

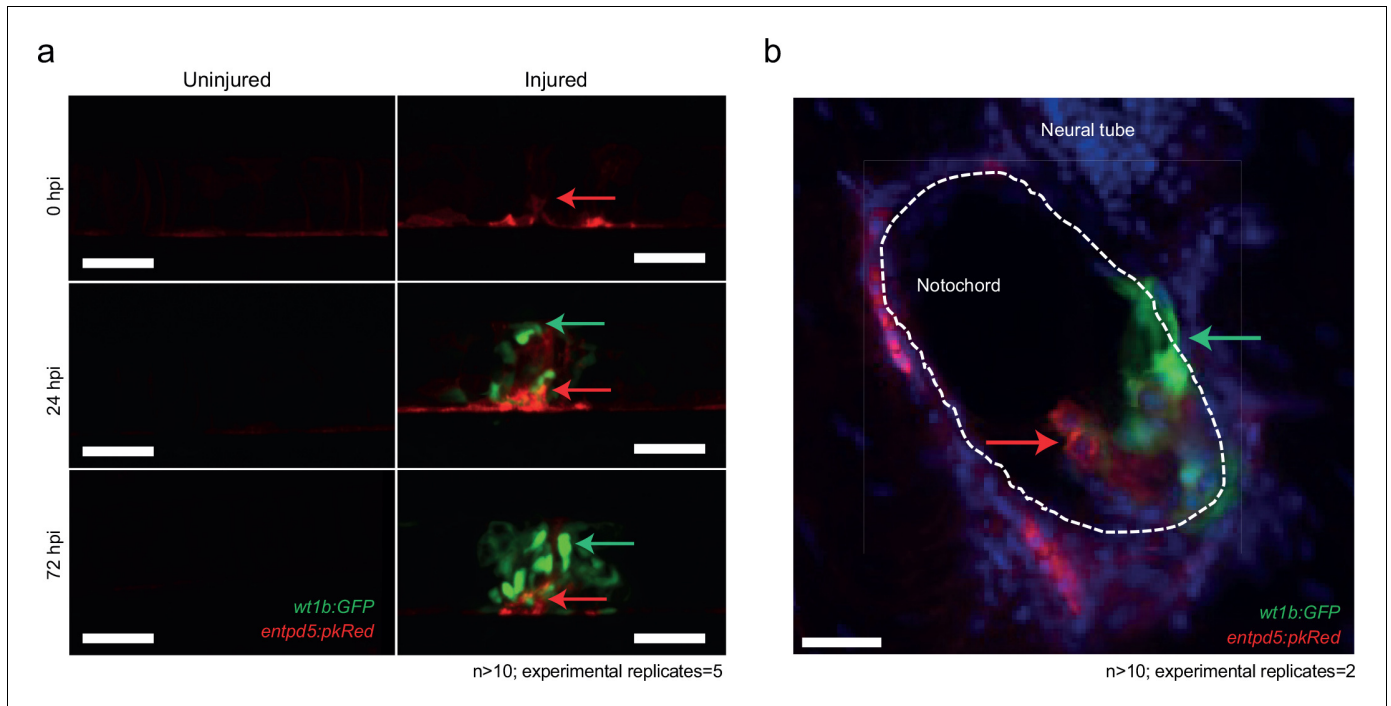DOI: https://doi.org/10.7554/eLife.30657.016

Figure 6. Distinct and closely associated *wt1b* and *entpd5a* subpopulations emerge at the damage site. (a) Live-imaging at the site of notochord injury in *Tg(wt1b:gfp; entpd5a:dkRed)* larvae. Expression of *wt1b:gfp* and *entpd5a:pkRed* at site of damage (green arrows and red arrows respectively) in injured and uninjured fish. n > 10; experimental replicates = 5. Scale bar: 50 μm. (b) Cryo-section of the injured area confirms distinct *wt1b:gfp* and *entpd5a:dkRed* subpopulations at site of damage. n > 10; experimental replicates = 2. Scale bar: 20 μm.

DOI: https://doi.org/10.7554/eLife.30657.017

The following figure supplement is available for figure 6:

Figure supplement 1. Needle damage of the *entpd5a* cell domain leads to supernumerary vertebrae.

DOI: https://doi.org/10.7554/eLife.30657.018

expression was transient at the wound, or sustained throughout the repair process, we raised needle injured *Tg(wt1b:gfp); casper* zebrafish larvae for up to 38 days.

GFP expression was sustained at the wound site, remaining in a small, cellular population at the site of damage, even as chordacentra developed and mineralized around the notochord over time (*Figure 7*). Small GFP expressing cells were further confirmed by α-GFP staining at the site of damage (*Figure 7b*). Strikingly, the *Tg(wt1b:gfp)* transgene maintained expression at this site up to 38 dpi (*Figure 7c,d,g*).

To gain a better understanding of how *wt1b:gfp* expressing cells engage with the newly forming vertebrae, we carried out live, confocal imaging of the area of damage (*Figure 7e–g*). The analysis revealed the presence of both fused and unfused vertebrae at the damaged site, and the sustained and strong expression of *wt1b:gfp* expressing cells associated with the developing vertebra at the repair site area (*Figure 7f*), even in fully formed spine structures (*Figure 7g*).

Taken together these results indicate that *wt1b:gfp* expressing cells both mark a subpopulation of cells that are rapidly activated at the site of the wound and also that these cells persist until adulthood, possibly orchestrating local vertebrae formation with wound repair.

Discussion

We have uncovered wound-specific cellular heterogeneity in the zebrafish notochord that perdures throughout the wound healing process and during adult vertebra formation at the injury site (*Figure 8*). We discover that wounding leads to localized *wt1b* expression in the notochord sheath cells which then invade the site of the injury to form a stopper-like structure, likely to maintain notochord integrity. We show the specific de novo expression of *wt1b* in notochord sheath cells following
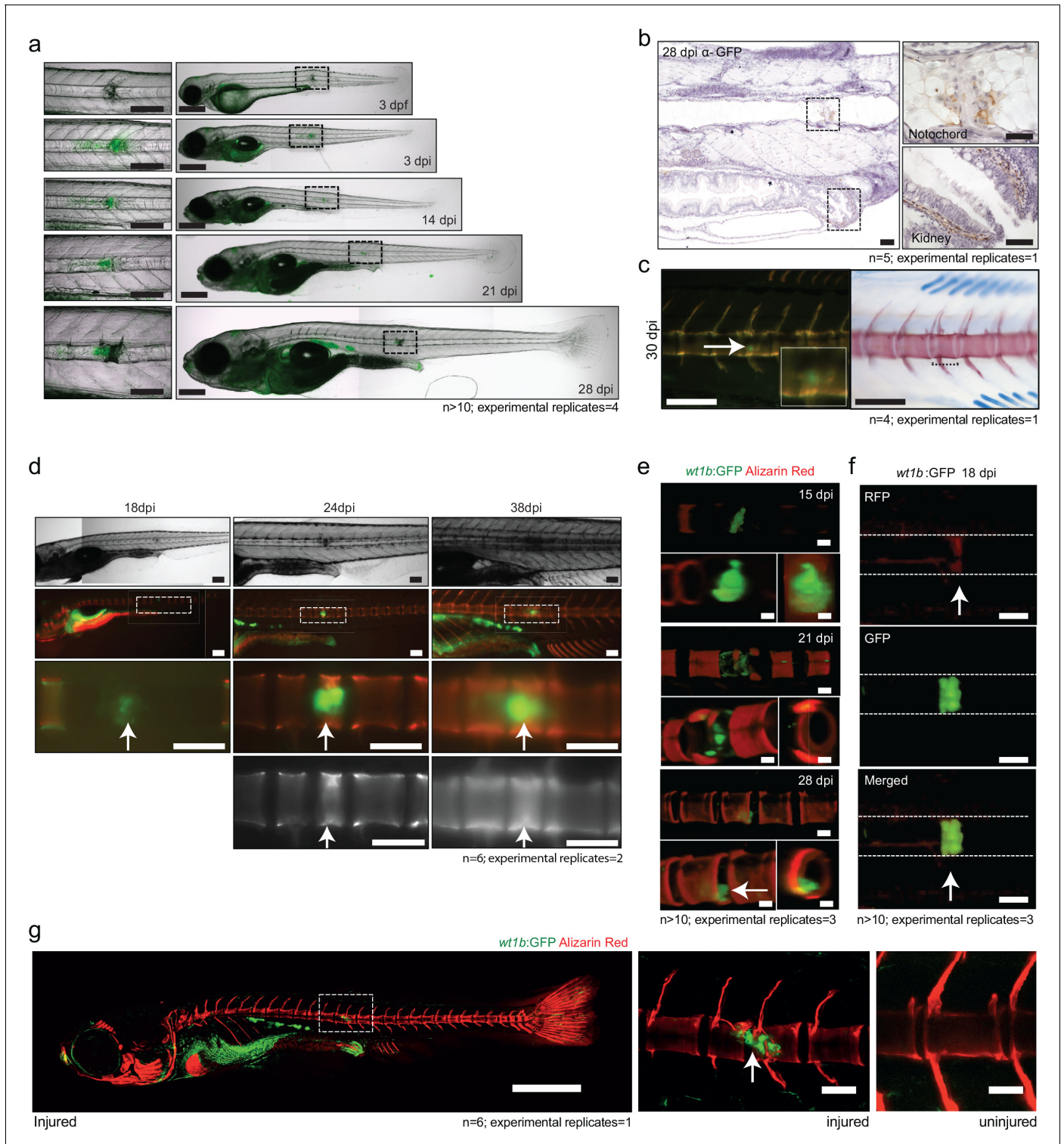
Figure 7. *wt1b* expressing cells are closely associated with vertebral development after injury. (**a**) Images of *Tg(wt1b:gfp)* zebrafish following needle injury at 3 dpf and raised to 28 dpi. n > 10; experimental replicates = 4. Scale bar left panels: 100 µm; scale bar right panels: 200 µm. (**b**) α-GFP staining of 28 dpi larvae at the site of the healing notochord wound and in the kidney. n = 5; experimental replicates = 1. Scale bar left panels: 50 µm. (**c**) Image of fish from *Figure 5a,c*, stained with alizarin red and imaged for *wt1b:gfp* expressing cells. GFP positive cells are found within the ectopic vertebra (white arrow and inset). n = 4; experimental replicates = 1. Scale bar left panels: 100 µm. (**d**) Long-term follow up of alizarin red stained *Tg(wt1b:gfp); casper* larvae shows that chordacentra formation is delayed around the site of injury. GFP cells mark the site of the future vertebra. n = 6; experimental

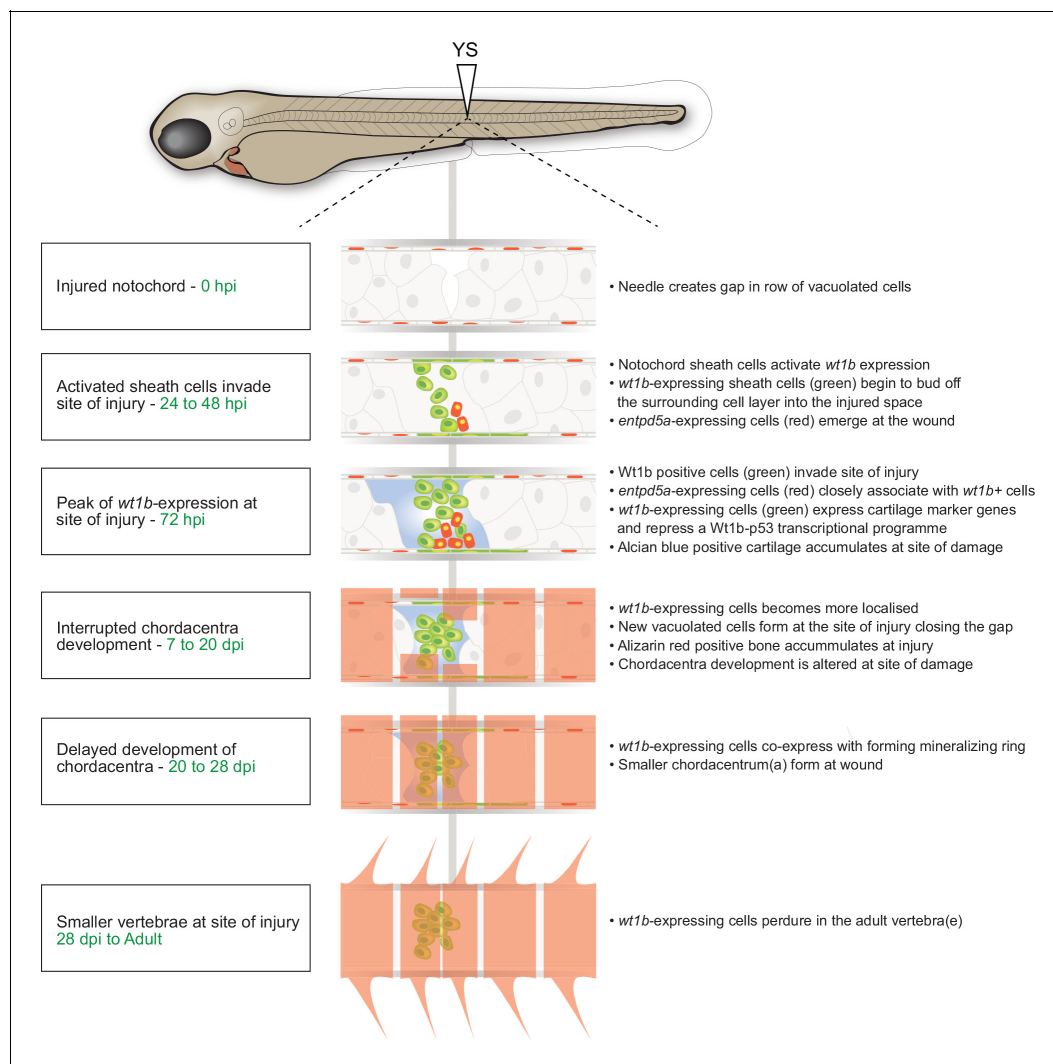*Figure 7 continued on next page*

wounding, despite an absence of *wt1b* expression during notochord development (*Figure 1e*,

Figure 8. Schematic of the notochord wound response.

Figure 1—figure supplement 2).

Very recently, Bagnat and colleagues reported the identification of notochord sheath cells involved in the replacement of vacuolated cells lost due to motion-dependent mechanical damage to the notochord in caveolin mutants (*Garcia et al., 2017*). In this context, sheath cells invade the vacuolated cell layer and differentiate into vacuolated cells to maintain turgor pressure. While we observe that most *wt1b*-expressing cells are tightly associated with a stopper-like (scar-like) structure from embryo to adult (*Figure 7*), we find some *wt1b* expressing cells appear vacuolated at the injury site at later stages (7 days post injury; *Figure 4f*), and that *wt1b*-expressing cells express vacuolar genes (*Figure 4e*). We also detected *entpd5a* expressing cell subpopulations at the wound that are distinct from *wt1b* expressing cells (*Figure 6*). These studies highlight a previously unknown complex and heterogeneous nature of the sheath cell populations, and suggest that the notochord sheath can sense and respond to different types of damage. Motion-dependent shear stress in *caveolin* mutants causes loss of vacuolated cells that are replaced by new vacuolated cells that arise from the sheath (*Garcia et al., 2017*), while acute damage (i.e. needle injury) that encompasses sheath and vacuolated cell damage, leads to sheath cells forming a seal that marks the site of new cartilage and vertebra (*Figure 8*). We show that *wt1b* expression marks a subpopulation of sheath cells in both damage responses (*Figure 1*, *Figure 2—figure supplement 2*), and suggest that additional factors are involved in the ultimate fate of *wt1b*-expressing cells (i.e. vacuolated cells versus scar like structure).

To address the function of Wt1b in the wound response, we generated a CRISPR-Cas9 genetic mutant that removes part of the C-terminal zinc-finger domains that are essential for WT1 function in mammalian systems. We find homozygous *wt1b* mutant zebrafish show no overt difference from wild type fish in the wound response (*Figure 8*; *Figure 8—figure supplement 1*). However, given the dramatic up-regulation of *wt1b* upon wounding, and given the continued expression up to adult stages, we consider it unlikely that Wt1b has no role in the process. Compensatory mechanisms have to be considered, and indeed, we find a small, but significant increase in *wt1a* in *wt1b*$^{\Delta5/\Delta5}$ wounded tissue. Furthermore, compensatory mechanisms downstream the Wt1b-p53 axis could mask a role, and further analysis beyond the scope of this study will be required to fully understand the functional significance of Wt1b in this subpopulation of cells.

By leveraging gene expression profiling, and single-cell and 10 cell sequencing of the wounded tissue, we discovered a mechanism for vertebra formation via a cartilage intermediate at the injury site. This is completely unexpected as in zebrafish, ossification of the chordacentra does not require the establishment of a cartige anlagen, but form via the direct mineralization of the fibrous notochord sheath (*Bensimon-Brito et al., 2012*; *Fleming et al., 2015*). The activation of *wt1b* in sheath cells that migrate towards the center of the notochord is reminiscent of the situation where *wt1b* expression is reactivated in epicardial cells that undergo EMT to produce vascular progenitors and migrate into the heart (*Martínez-Estrada et al., 2010*). This raises the question whether notochord sheath cells may also be mesothelial in nature and if the invading *wt1b* expressing cells are produced via an EMT or, perhaps more accurately, a mesothelial to mesenchyme transition. While *wt1b*-positive cells express some mesenchymal genes (*Figure 3d*), we did not find evidence that these cells express classical gene signatures related to known EMT processes in the damaged tissue. This may be evidence of an as of yet unknown process in the wound response, or possibly because the EMT process was primarily completed by the time of our analysis at 3 days post injury.

Surprisingly, we have uncovered a new Wt1b-p53 gene expression signature that is specifically repressed in *wt1b*$^+$ sheath cells (*Figure 4k*). p53 is a transcription factor that in addition to its well-established role as a tumor suppressor, functions to inhibit premature osteoblast differentiation and bone remodeling (*Liu and Li, 2010*). Several lines of evidence support a direct WT1-p53 interaction, and that p53 can modify activity of WT1 transcriptional activity from an activator to a repressor on select promoters in vitro (*Maheswaran et al., 1995*; *Maheswaran et al., 1993*). However, the in vivo function for the WT1-p53 interaction is not yet understood, and loss of *p53* in *wt1*-null mutant mice does not alter the *wt1*-null phenotype (*Menke et al., 2002*). Here, we identify a Wt1b-p53 axis specifically in the repair of a notochord wound. The Wt1b-p53 gene signature includes repression of genes that regulate osteogenesis in mammals, including *myc* (*a* and *b*), *egr1* and *igfrb* (*Piek et al., 2010*; *Reumann et al., 2011*; *Wang et al., 2015*). We propose that repair-specific transcription factors participate in notochord healing by co-ordinating expression of cartilage genes such as *sox9* and *mgp* (*Schurgers et al., 2013*; *Sweatt et al., 2003*; *Zebboudj et al., 2002*), with a Wt1b-p53

transcriptional axis repressing premature expression of osteogenesis genes in the first few days following wounding. We see *entpd5*[+] notochord sheath cells in the wound area (*Figure 6*), and since *entpd5* is essential for mineralization, it seems likely that these cells, in conjunction with cartilage formation at the site of injury, play a role in centrum formation (*Figure 8*). Eventually, smaller vertebra form at the wound site, and *wt1b:gfp* cells remain tightly associated with this/these vertebra(e) into adulthood. This mode of notochord wound healing and vertebra formation may be a salvage structure to effectively maintain structural integrity of the developing axial skeleton.

Materials and methods

Key resources table

Reagent type or resource	Designation	Source or reference	Identifiers	Additional information
Gene (Danio Rerio)	*sagff214a*	NA	ZFIN ID: ZDB-ALT-110315–2	
Gene (Danio Rerio)	*wt1a*	NA	ZFIN ID: ZDB-GENE-980526–558	
Gene (Danio Rerio)	*col2a1a*	NA	ZFIN ID: ZDB-GENE-980526–192	
Gene (Danio Rerio)	*entpd5a*	NA	ZFIN ID: ZDB-GENE-100419–1	
Gene (Danio Rerio)	*sox9b*	NA	ZFIN ID: ZDB-GENE-001103–2	
Gene (Danio Rerio)	*wt1b*	NA	ZFIN ID: ZDB-GENE-050420–319	
Genetic reagent (Danio Rerio)	Tg(entpd5:kaede)	*Geurtzen et al., 2014* doi: 10.1242/dev.105817	ZFIN ID: ZDB-ALT-150223–1: hu6867	Same BAC used as *Huitema et al. (2012)* (DOI: 10.1073/pnas.1214231110) with kaede insertion at first translated ATG
Genetic reagent (Danio Rerio)	Tg(entpd5:pkRed)	This paper	ZFIN ID: hu7478	Same BAC used as *Huitema et al. (2012)* (DOI: 10.1073/pnas.1214231110) with pkRed insertion at first translated ATG
Genetic reagent (Danio Rerio)	Tg(SAGFF214a;UAS:gfp)	*Yamamoto et al. (2010)* DOI: 10.1242/dev.051011	ZFIN ID: ZDB-FISH-150901–18089	
Genetic reagent (Danio Rerio)	Tg(wt1b:GFP,R2col 2a1a:mCherry)	This paper	ZFIN ID: ZDB-ALT-180105–1; zfin.org:ue401Tg	
Genetic reagent (Danio Rerio)	Tg(wt1a:GFP)	*Bollig et al. (2009)* DOI: 10.1242/dev.031773	ZFIN ID: ZDB-FISH-150901–2540	
Genetic reagent (Danio Rerio)	Tg(wt1b:GFP)	*Perner et al. (2007)* DOI: 10.1016/j.ydbio.2007.06.022	ZFIN ID: ZDB-FISH-150901–1774	
Genetic reagent (Danio Rerio)	casper	*White et al. (2008)* DOI: 10.1016/j.stem.2007.11.002	ZFIN ID: ZDB-ALT-990423–22	
Genetic reagent (Danio Rerio)	zebrafish codon optimised cas9 mRNA	*Jao et al. (2013)* DOI: 10.1073/pnas.1308335110		
Genetic reagent (Danio Rerio)	Wt1b p.F319fsX321	this paper	ZFIN ID: ZDB-ALT-180 105–2; zfin.org:ue402	zebrafish wt1b mutant line, mutation is in the exon coding the zinc finger 2
Genetic reagent	Tol2 transposase	*Kawakami, 2007* DOI: 10.1186/gb-2007–8 s1-s7		

Continued on next page

Continued

Reagent type or resource	Designation	Source or reference	Identifiers	Additional information
Antibody	anti-WT1 (rabbit polyclonal)	This paper, Cambridge Research Biochemicals antibody production services		(1:25000); anti-WT1 was designed using the TARGET antibody production protocol from Cambridge Research Biochemicals using a conserved protein sequence from the C-terminal of the zebrafish Wt1a and Wt1b proteins.
Antibody	AlexaFluor 488 antibody (rabbit polyclonal)	Invitrogen	Donkey anti-Rabbit IgG (H + L) Secondary Antibody, Alexa Fluor 488: R37602; RRID:AB_221544	(1:800)
Antibody	anti-GFP (rabbit polyclonal)	Cell Signaling Technology	Cell Signaling Technology: GFP Antibody (Rabbit): 2555S; RRID:AB_10692764	(1:1500)
Recombinant DNA reagent (plasmid)	R2-col2a1a:mCherry	*Dale and Topczewski (2011)* DOI: 10.1016/j.ydbio.2011.06.020		
Sequence-based reagent	wt1b mutant sgRNA	this paper		GGTCAGACCTGGAGAAGCGG
Commercial assay or kit	Dako REAL EnVision Detection System kit	Dako	Dako REAL EnVision Detection System, Peroxidase /DAB+, Rabbit/Mouse: Code K5007	
Commercial assay or kit	Low Input Quick Amp Labelling Kit	Agilent Technologies	Low Input Quick Amp Labeling Kit, one-color: 5190–2305	
Commercial assay or kit	Nextera XT DNA Library Preparation Kit (96 samples),	Illumina	Nextera XT DNA Library Preparation Kit (96 samples),: Cat: FC-131–1096	
Commercial assay or kit	4 × 44K Whole Zebrafish (V3) Genome Oligo Microarray	Agilent Technologies		
Chemical compound, drug	DPX Mountant for histology	Sigma-Aldrich	DPX Mountant for histology: 06522–100 ML	
Chemical compound, drug	ProLong Gold Antifade Mountant with DAPI	Invitrogen	ProLong Gold Antifade Mountant with DAPI: P36931	
Chemical compound, drug	Trizol	Invitrogen	TRIzol Reagent: 15596026	
Chemical compound, drug	FACSmax cell disassociation solution	Genlantis	FACSmax Cell Dissociation Solution: AMS.T200110	
Chemical compound, drug	OCT compound Tissue-Tek	Sifam Instruments LTD	OCT COMPOUND TISSUE-TEK: SIFAAGR1180	
Chemical compound, drug	Nystatin	Sigma-Aldrich	Nystatin powder, BioReagent, suitable for cell culture: N6261-500KU	
Software, algorithm	Color Inspector 3D	ImageJ 1.51 n plugin	RRID:SCR_002285	
Software, algorithm	Fiji	ImageJ 1.51 n	RRID:SCR_002285	
Software, algorithm	Feature Extraction Software	Agilent Technologies	RRID:SCR_014963	
Software, algorithm	Rsubread package	R-3.3.3; *Liao et al. (2013)*. DOI: 10.1093/nar/gkt214	RRID:SCR_009803	
Software, algorithm	SCDE	*Kharchenko et al. (2014)* DOI: 10.1038/nmeth.2967	RRID:SCR_015952	

Continued on next page

Continued

Reagent type or resource	Designation	Source or reference	Identifiers	Additional information
Software, algorithm	SC3 package	*Kiselev et al. (2017)* DOI: 0.1038/nmeth.4236	RRID:SCR_015953	
Software, algorithm	Scater package	*McCarthy et al., 2017* DOI: 10.1093/bioinformatics/btw777	RRID:SCR_015954	
Software, algorithm	STAR RNA-seq aligner	*Dobin et al. (2013)* DOI: 10.1093/bioinformatics/bts635	RRID:SCR_015899	
Software, algorithm	FACSDiva software	Version 6.1.3; BD Biosciences	RRID:SCR_001456	
Software, algorithm	Webgestalt	*Wang et al. (2013)* DOI: 10.1093/nar/gkt439	RRID:SCR_006786	
Software, algorithm	Rosetta Resolver gene expression data analysis system	Rosetta Biosoftware	RRID:SCR_008587	
Other	Alizarin Red	Fisher Scientific	Alizarin Red S Sodium Salt25G:11329707	
Other	Alcian Blue	Sigma	Alcian Blue 8Gx: A5268-10G	

Zebrafish lines

All experimental procedures were approved by the University of Edinburgh Ethics Committee and were in accordance with the UK Animals (Scientific Procedures) Act 1986. Transgenic lines for this study include: *Tg(entpd5a:pkRed)* (*Huitema et al., 2012*), *Tg(SAGFF214A:GalFF;UAS:gfp)* (*Yamamoto et al., 2010*), *Tg(wt1a:gfp)* (*Bollig et al., 2009*), *Tg(wt1b:gfp)* (*Bollig et al., 2009*; *Perner et al., 2007*). Many of the studies were performed in a transparent background created by crossing homozygous *Tg(wt1b:gfp)* fish to homozygous pigment-free transparent *casper* fish (*White et al., 2008*). The *Tg(wt1b:gfp;R2col2a1a:mCherry)* line was created by injecting the *R2col2a1a:mCherry* construct (*Dale and Topczewski, 2011*) with a Tol2 transposase (*Kawakami, 2007*) into *Tg(wt1b:gfp;casper)* zebrafish embryos, generating *Tg(R2col2a1a:mCherry)*[ue401Tg].

Notochord needle injury and tail amputation assays

For notochord wounds on day 3, larvae were anaesthetised in tricaine, placed sagittally on a petri dish and either inserted gently with an electrolysis-sharpened tungsten wire or tail amputated at different levels. Injured larvae were transferred to fresh water to recover and observe. Non-injured age-matched larvae were grown as non-injured controls. For injuries on day 5 and 7 pf larvae, the notochord wounds were generated using stainless steel insect pins (0.10 mm), under fluorescence light in a Leica (Germany) M165FC with a 1.0X plan Apo objective. All pictures (brightfield, Kaede and alizarin red stains) where taken using an Olympus (Japan) szx16 with a 1.5X Plan Apo objective with a Leica DFC 450C camera.

Whole-mount microscopy

Live and fixed whole-mount time-course and time-lapse experiments were performed using an AZ100 upright macroscope (Nikon; Japan) using a x2 and x5 lens with a Retiga Exi camera (Qimaging) or Coolsnap HQ2 camera (Photometrics; Tucson, Arizona) or a Leica MZFLIII fluorescence stereo microscope fitted with a Qimaging Retiga Exi camera. Images were analyzed and processed using the IPLab Spectrum and Micro-Manager software. Live and fixed whole-mount confocal imaging was performed using an A1R confocal system (Nikon) using x10 and a x20 lens over a Z-plane range of 80–100 μm (approximate width of the notochord) using a 480 nm laser (GFP), a 520 nm (RFP) and/or a 561 nm laser (alizarin red). Images were captured and analysed using Nis-Elements C software (Nikon). Images of the nystatin-treated larvae were acquired by using a 20x lens on the Imaging Platform Dragonfly (Andor Technologies, Belfast UK) with 488 nm (GFP) and 561 nm (RFP) lasers built on a Nikon TiE microscope body with a Perfect focus system (Nikon Instruments). Z stacks through the notochord were collected in Spinning Disk 25 μm pinhole mode on the Zyla 4.2 camera using a Bin of 1 and frame averaging of 1 using Fusion v1.4 software. Data were visualised using Fiji, and

histograms generated using its Color Inspector 3D plugin. Multiphoton confocal time-lapse imaging was performed using an SP5 confocal microscope (Leica) equipped with a Ti:Sapphire multiphoton laser (Spectra Physics; Santa Clara, California) and a three axis motorised stage. For confocal imaging and time-lapse experiments, anaesthetised injured and non-injured larvae were embedded sagittally in a drop of 1% low-melting point agarose prior to imaging, in a specially designed glass insert, which was covered in a mixture of E3 medium and anaesthetic. All time-lapse imaging was done at 30 or 60- min intervals over 48 hr using an incubation chamber (Solent Scientific; UK) under a constant temperature of 28°C and larvae were terminated in an overdose of tricaine at the end of each experiment.

Histology

Zebrafish larvae younger than 20 dpf were culled and fixed overnight in 4% PFA/PBS at 4°C. The fixed larvae were washed in PBS, dehydrated in rising methanol/PBS concentrations and cleared in xylene before being paraffin embedded for sectioning. Older zebrafish were culled and fixed in 4% PFA/PBS at 4°C for 3 days with an abdominal incision to ensure tissue penetrance of the fixative (*Walker and Kimmel, 2007*; *Wojciechowska et al., 2016*). Fish were decalcified using 0.5M EDTA (pH 7.5) for 5 days in a rocker at 4°C and dehydrated in 70% ethanol at 4°C. Fish were embedded in paraffin using a Miles Scientific Tissue TEK VIP automated processor. Embedded larvae and older zebrafish were sectioned using a Leica RM2235 rotary microtome to a width of 5 µm. Sections were haematoxylin and eosin (H and E) stained and mounted using DPX mountant for histology (Sigma-Aldrich; St. Louis, Missouri). For cryosections, zebrafish larvae were embedded in OCT compound Tissue Tek (Sifam Instruments LTD; UK) and cut to 8 µm following protocols available at www.zfin.org.

Wt1 zebrafish antibody

The Wt1 antibody was synthesised by Cambridge Research Biochemicals (CRB; UK) antibody production services (http://www.crbdiscovery.com/home). The antibody was created using the CRB TARGET antibody production protocol (https://www.crbdiscovery.com/antibodies/target-antibodies/), which used a HPLC-purified peptide made from the third zinc finger domain of zebrafish Wt1 (CQRKFSRSDHLKTHTRT) to immunise two rabbits. This epitope is found in both Wt1a and Wt1b, and the antibody is expected to detect both zebrafish Wt1a and Wt1b. The serum from each rabbit was collected at multiple time points and tested for the presence of Wt1 antibodies using an electrophoretic mobility shift assay (EMSA). The purified polyclonal antibody was extracted from the rabbit serum on the final collection day. Western blot analysis of lysates from zebrafish (24 hpf) revealed a strong band at approximately 45 kDa, consistent with the size of zebrafish Wt1a/b protein. Immunofluorescence on paraffin-embedded sections with Wt1 antibody (diluted 1:33,000) revealed cell-specific staining in the kidney and notochord wound site that was depleted by co-incubation of the Wt1 antibody with the Wt1 epitope peptide.

Immunohistochemistry

Slides were de-waxed in xylene and rehydrated through decreasing ethanol washes, before being incubated in a bleach solution to remove pigment. Antigen-unmasking was performed as previously described (*Patton et al., 2005*). with the Dako REAL EnVision Detection System kit (Dako; UK) following manufacturer's instructions. Slides were incubated overnight at 4°C with the following antibodies: anti-rabbit α-GFP (1:1,500; Cell Signaling Technology) and anti-rabbit α-WT1 (1:25,000; Cambridge Research Biochemicals; UK). An Axioplan II fluorescence microscope (Zeiss; Germany) with a Plan Apochromat objective was used for brightfield imaging of tissue sections. Images were captured using a Qimaging Micropublisher 3.3mp cooled CCD camera and analysed using the IPLab Spectrum software.

Immunofluorescence

Slides were processed as described above and blocked in 10% heat inactivated donkey serum for 2 hr. Slides were incubated overnight at 4°C with α-WT1 (1:33,000) antibody diluted in 1% heat inactivated donkey serum in TBSTw. Slides were incubated for 1 hr in a secondary anti-rabbit AlexaFluor 488 antibody (1:800) (Invitrogen; Carlsbad, California) in 1% heat inactivated donkey serum and

mounted in ProLong Gold Antifade Mountant with DAPI (Invitrogen) overnight before being imaged in a fluorescent stereomicroscope.

Tissue staining

Live bone staining was performed using 0.2% (w/v) calcein or using 50 µg/ml alizarin red (Fisher Scientific; UK) as previously described (*Kimmel et al., 2010*). For cartilage and bone staining, we used alcian blue and alizarin red following the protocol outlined in (*Walker and Kimmel, 2007*) with modifications from protocols on www.zfin.org. Bone and cartilage staining in fixed larvae was performed on PFA fixed and then methanol dehydrated specimens, treated overnight at 4°C with 0.02% (weight to volume) alizarin red in 70% ethanol. Specimens were bleached (H_2O_2) and cleared before storing in glycerol for imaging.

RNA extraction and microarray analysis

Tg(wt1b:gfp) zebrafish larvae were needle injured and grown to 72 hpi with age-matched non-injured controls. The area around the site of injury was dissected and transferred into 1 ml of chilled RNA-later. The samples were centrifuged into a pellet at 4°C and macerated in 500 µl of Trizol (Invitrogen) using a 25G $^{5/8}$1 ml syringe. RNA was extracted following Trizol manufacturer's instructions and eluted into 15 µl of distilled H_2O. Extracted RNA was sent to Myltenyi Biotec (Germany) who conducted the microarray analysis. Injured and non-injured samples were sent in triplicates and the RNA was amplified and Cy3-labelled using a Low Input Quick Amp Labelling Kit (Agilent Technologies; UK) following manufacturer's instructions. The labelled cRNA was hybridised against a 4 × 44K Whole Zebrafish (V3) Genome Oligo Microarray (Agilent Technologies). The microarray images were processed using the Feature Extraction Software (FES – Agilent Technologies) and differential gene expression was determined using the Rosetta Resolver gene expression data analysis system (Rosetta Biosoftware).

Fluorescence-Activated cell sorting

The trunk region of fifty *Tg(wt1b:gfp; R2col2a1a:mCherry)* injured larvae and non-injured 72 hpi larvae were dissected and collected separately in cold PBS + 2% fetal calf serum (FCS). Tissue disassociation was adapted from a previously described protocol (*Manoli and Driever, 2012*) and centrifuged cells were collected in FACSmax cell disassociation solution (Genlantis; San Diego, California). The samples were passed twice through a 40 µm cell strainer, collected in an agar-coated petri dish on ice and transferred into an eppendorf tube to be sorted by a FACSAria2 SORP instrument (BD Biosciences; UK) equipped with a 405 nm, a 488 nm and a 561 nm laser. Green fluorescence was detected using GFP filters 525/50 BP and 488 nm laser, red fluorescence was detected using 585/15 BP filter and 561 nm laser. Data were analysed using FACSDiva software (BD Biosciences) Version 6.1.3. For single cell sequencing, single cell or 10 cells were sorted into 96-well plates; for quantitive realtime PCR (qPCR) analysis, cells were collected by centrifuging at 6000 rpm for 5 min.

wt1b mutant line

The *wt1b* genetic mutant line was generated by CRISPR/Cas9 with a guide RNA target GGTCA-GACCTGGAGAAGCGG (on the reverse strand) in the exon of *wt1b* that encodes zinc finger 2. Crispr/Cas9 genome editing was carried out following the Joung lab protocol (*Hwang et al., 2013*), with injection of a zebrafish codon optimized Cas9 mRNA (*Jao et al., 2013*). Two founders carrying germline mutations at the target site were identified: one mutation is a deletion of 12 bp (leading to an in-frame deletion) and the other is a deletion of 5 bp (*wt1b*$^{\Delta 5)ue402}$, and was used in this study.

qRT-PCR analysis

RNA extraction, in vitro synthesis and PCR amplification of cDNA were performed using the Smart Seq2 protocol (*Picelli et al., 2014*). Amplified cDNA was quantified using a bioanalyzer, and directly used for qRT-PCR without further dilution of the cDNA template. qPCR was performed in a Roche LightCycler480 using a SYBR green protocol. ΔCt (the difference between the cycle threshold (Ct) value of the gene of interest and the Ct value of *ß-actin* or *gapdh*) was used to compare the

expression level of genes. Statistics (St Dev and paired T-test) were performed using Matlab (Natick, Massachussets). Primers are listed in *Supplementary file 2*.

Single-cell and 10 cell sequencing

RNA extraction, in vitro synthesis and PCR amplification of cDNA, and construction of a sequencing library using the Nextera XT DNA Library Preparation Kit (Illumina; San Diego California) according to the Smart Seq2 protocol with minor modifications as described before (*Picelli et al., 2014*). Libraries were sequenced on a NextSeq Illumina sequencer. Reads were mapped against the Ensembl *Danio rerio* reference genome version GRCz10.90 (*Ensembl, 2017*) with the inclusion of the reference for the spike in controls from the ERCC consortium, as well as the coding sequence for EGFP and mCherry, using STAR RNA-seq aligner (*Dobin et al., 2013*). For quality control and preprocessing, quantification of mapped reads per gene was calculated using the Rsubread package in R-3.3.3 (*Liao et al., 2013*). Genes that were not expressed in any cells were excluded. The gene counts were loaded as a scater object in R-3.3.3 (using the scater package) and standard quality control metrics were calculated (*McCarthy et al., 2017*). Quality control exclusion criteria were cells with more than 25% of reads mapping to ERCCs or fewer than 100,000 reads or fewer than 1000 genes detected (at least one read per gene) were rejected (see *Figure 4—figure supplement 1* and *Supplementary file 1a*).

Consensus clustering set to three clusters was conducted on the single and 10 cells using the SC3 package (*Kiselev et al., 2017*). The 10 cell group was isolated and SC3 consensus clustering set to three clusters was conducted on these cells alone. Differential expression between cluster 2 and cluster 3 of the SC3 10 cell analysis was conducted using SCDE (*Kharchenko et al., 2014*). A differential expression list, ranked from cluster 2 to cluster three according to z-score was used for the GSEA analysis (*Mootha et al., 2003*; *Subramanian et al., 2005*). The differential expression list was tested against gene lists compiled from online resources (*Supplementary file 1b*). Functional analysis between the ranked 10 cell list and online gene lists for gene ontology (biological processes, nonredundant) and pathways (KEGG, Panther, Reactome and WikiPathway databases were used) using the online tool WebGestalt and gene set enrichment function (*Wang et al., 2013*).

Vertebrae size measurements and statistical analysis

The vertebrae size difference in injured zebrafish larvae (age range 30 dpi to 38 dpi) were compared between vertebrae at the site of injury (injured) and vertebrae outside of the site of injury (uninjured). Injured vertebrae and uninjured vertebrae were measured and the average length was recorded for each group. The average lengths were then compared and the relative size difference was calculated. The relative size difference between each group (injured:uninjured vs. uninjured:uninjured) was compared using an unpaired t-test.

Author contributions

Juan Carlos Lopez-Baez, Conceptualization, Resources, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing—original draft, Writing—review and editing; Daniel J Simpson, Conceptualization, Data curation, Software, Formal analysis; Laura LLeras Forero, Conceptualization, Formal analysis, Validation, Investigation; Zhiqiang Zeng, Resources, Generating transgenic lines and molecular biology tools; Hannah Brunsdon, Conceptualization, Formal analysis, Investigation; Angela Salzano, Investigation, Methodology; Alessandro Brombin, Visualization, Writing—review and editing; Cameron Wyatt, Formal analysis, Validation, Investigation; Witold Rybski, Resources, Maintaining genetic lines and genetic crosses; Leonie F A Huitema, Resources, Generation of transgenic line; Rodney M Dale, Conceptualization, Resources, Writing—review and editing, Result interpretation, and sharing of reagents and expertise; Koichi Kawakami, Christoph Englert, Resources, Writing—review and editing, Result interpretation, and sharing of reagents and expertise; Tamir Chandra, Conceptualization, Software, Formal analysis, Supervision, Funding acquisition; Stefan Schulte-Merker, Nicholas D Hastie, Conceptualization, Resources, Supervision, Funding acquisition, Writing—original draft, Writing—review and editing; E Elizabeth Patton, Conceptualization, Resources, Data curation, Formal analysis, Supervision, Funding acquisition, Investigation, Methodology, Writing—original draft, Project administration, Writing—review and editing

Author ORCIDs

Alessandro Brombin http://orcid.org/0000-0001-8262-9248
Witold Rybski http://orcid.org/0000-0002-6025-2918
Rodney M Dale https://orcid.org/0000-0003-4255-4741
Koichi Kawakami http://orcid.org/0000-0001-9993-1435
Christoph Englert http://orcid.org/0000-0002-5931-3189
Stefan Schulte-Merker http://orcid.org/0000-0003-3617-8807
E Elizabeth Patton http://orcid.org/0000-0002-2570-0834

Ethics

Animal experimentation: All work presented in this study has been performed in accordance with the UK legal requirements for the protection of animals used for experimental or other scientific research under the Animal (Scientific Procedures) Act 1986. All experiments were approved by the University of Edinburgh Ethics Committee, and performed under the Home Office Project License 70/8000 to EEP. Zebrafish welfare and husbandry were closely monitored by the MRC Human Genetics Unit Zebrafish Facility staff.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.30657.027
Author response https://doi.org/10.7554/eLife.30657.028

Additional files

Supplementary files

• Supplementary file 1. (a) Single-cell differential expression list. (b) Gene List Sources. (c) Zebrafish cartilage genes. (d) WT1 gene targets. (e) p53 gene targets. (f) WT1 and p53 shared gene targets
DOI: https://doi.org/10.7554/eLife.30657.023

• Supplementary file 2. List of primers used for qRT-PCR and genotyping.
DOI: https://doi.org/10.7554/eLife.30657.024

• Transparent reporting form
DOI: https://doi.org/10.7554/eLife.30657.025

References

Apschner A, Schulte-Merker S, Witten PE. 2011. Not all bones are created equal - using zebrafish and other teleost species in osteogenesis research. *Methods in Cell Biology* **105**:239–255. DOI: https://doi.org/10.1016/B978-0-12-381320-6.00010-2, PMID: 21951533

Asahina K, Zhou B, Pu WT, Tsukamoto H. 2011. Septum transversum-derived mesothelium gives rise to hepatic stellate cells and perivascular mesenchymal cells in developing mouse liver. *Hepatology* **53**:983–995. DOI: https://doi.org/10.1002/hep.24119, PMID: 21294146

Baran-Gale J, Chandra T, Kirschner K. 2017. Experimental design for single-cell RNA sequencing. *Briefings in Functional Genomics*. DOI: https://doi.org/10.1093/bfgp/elx035, PMID: 29126257

Bensimon-Brito A, Cardeira J, Cancela ML, Huysseune A, Witten PE. 2012. Distinct patterns of notochord mineralization in zebrafish coincide with the localization of Osteocalcin isoform 1 during early vertebral centra formation. *BMC Developmental Biology* **12**:28. DOI: https://doi.org/10.1186/1471-213X-12-28, PMID: 23043290

Bollig F, Perner B, Besenbeck B, Köthe S, Ebert C, Taudien S, Englert C. 2009. A highly conserved retinoic acid responsive element controls wt1a expression in the zebrafish pronephros. *Development* **136**:2883–2892. DOI: https://doi.org/10.1242/dev.031773, PMID: 19666820

Carmona R, Cano E, Mattiotti A, Gaztambide J, Muñoz-Chápuli R. 2013. Cells derived from the coelomic epithelium contribute to multiple gastrointestinal tissues in mouse embryos. *PLoS One* **8**:e55890. DOI: https://doi.org/10.1371/journal.pone.0055890, PMID: 23418471

Chau YY, Bandiera R, Serrels A, Martínez-Estrada OM, Qing W, Lee M, Slight J, Thornburn A, Berry R, McHaffie S, Stimson RH, Walker BR, Chapuli RM, Schedl A, Hastie N. 2014. Visceral and subcutaneous fat have different origins and evidence supports a mesothelial source. *Nature Cell Biology* **16**:367–375. DOI: https://doi.org/10.1038/ncb2922, PMID: 24609269

Dale RM, Topczewski J. 2011. Identification of an evolutionarily conserved regulatory element of the zebrafish col2a1a gene. *Developmental Biology* **357**:518–531. DOI: https://doi.org/10.1016/j.ydbio.2011.06.020, PMID: 21723274

Deasey S, Grichenko O, Du S, Nurminskaya M. 2012. Characterization of the transglutaminase gene family in zebrafish and in vivo analysis of transglutaminase-dependent bone mineralization. *Amino Acids* **42**:1065–1075. DOI: https://doi.org/10.1007/s00726-011-1021-0, PMID: 21809079

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. DOI: https://doi.org/10.1093/bioinformatics/bts635, PMID: 23104886

Du SJ, Frenkel V, Kindschi G, Zohar Y. 2001. Visualizing normal and defective bone development in zebrafish embryos using the fluorescent chromophore calcein. *Developmental Biology* **238**:239–246. DOI: https://doi.org/10.1006/dbio.2001.0390, PMID: 11784007

Ellis K, Hoffman BD, Bagnat M. 2013. The vacuole within: how cellular organization dictates notochord function. *Bioarchitecture* **3**:64–68. DOI: https://doi.org/10.4161/bioa.25503, PMID: 23887209

Ensembl. 2017. Zebrafish (GRCz10). http://www.ensembl.org/Danio_rerio/Info/Index [Accessed 6 November, 2017].

Essafi A, Webb A, Berry RL, Slight J, Burn SF, Spraggon L, Velecela V, Martinez-Estrada OM, Wiltshire JH, Roberts SG, Brownstein D, Davies JA, Hastie ND, Hohenstein P. 2011. A wt1-controlled chromatin switching mechanism underpins tissue-specific wnt4 activation and repression. *Developmental Cell* **21**:559–574. DOI: https://doi.org/10.1016/j.devcel.2011.07.014, PMID: 21871842

Fleming A, Keynes R, Tannahill D. 2004. A central role for the notochord in vertebral patterning. *Development* **131**:873–880. DOI: https://doi.org/10.1242/dev.00952, PMID: 14736741

Fleming A, Kishida MG, Kimmel CB, Keynes RJ. 2015. Building the backbone: the development and evolution of vertebral patterning. *Development* **142**:1733–1744. DOI: https://doi.org/10.1242/dev.118950, PMID: 25968309

Garcia J, Bagwell J, Njaine B, Norman J, Levic DS, Wopat S, Miller SE, Liu X, Locasale JW, Stainier DYR, Bagnat M. 2017. Sheath cell invasion and trans-differentiation repair mechanical damage caused by loss of caveolae in the zebrafish notochord. *Current Biology* **27**:1982–1989. DOI: https://doi.org/10.1016/j.cub.2017.05.035, PMID: 28648824

Gavaia PJ, Simes DC, Ortiz-Delgado JB, Viegas CS, Pinto JP, Kelsh RN, Sarasquete MC, Cancela ML. 2006. Osteocalcin and matrix Gla protein in zebrafish (Danio rerio) and Senegal sole (Solea senegalensis): comparative gene and protein expression during larval development through adulthood. *Gene Expression Patterns* **6**:637–652. DOI: https://doi.org/10.1016/j.modgep.2005.11.010, PMID: 16458082

Geurtzen K, Knopf F, Wehner D, Huitema LF, Schulte-Merker S, Weidinger G. 2014. Mature osteoblasts dedifferentiate in response to traumatic bone injury in the zebrafish fin and skull. *Development* **141**:2225–2234. DOI: https://doi.org/10.1242/dev.105817, PMID: 24821985

Hastie ND. 2017. Wilms' tumour 1 (WT1) in development, homeostasis and disease. *Development* **144**:2862–2872. DOI: https://doi.org/10.1242/dev.153163, PMID: 28811308

Huitema LF, Apschner A, Logister I, Spoorendonk KM, Bussmann J, Hammond CL, Schulte-Merker S. 2012. Entpd5 is essential for skeletal mineralization and regulates phosphate homeostasis in zebrafish. *PNAS* **109**: 21372–21377. DOI: https://doi.org/10.1073/pnas.1214231110, PMID: 23236130

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* **31**:227–229. DOI: https://doi.org/10.1038/nbt.2501, PMID: 23360964

Jao LE, Wente SR, Chen W. 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *PNAS* **110**:13904–13909. DOI: https://doi.org/10.1073/pnas.1308335110, PMID: 23918387

Kawakami K. 2007. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biology* **8**:S7. DOI: https://doi.org/10.1186/gb-2007-8-s1-s7, PMID: 18047699

Kharchenko PV, Silberstein L, Scadden DT. 2014. Bayesian approach to single-cell differential expression analysis. *Nature Methods* **11**:740–742. DOI: https://doi.org/10.1038/nmeth.2967, PMID: 24836921

Kimmel CB, DeLaurier A, Ullmann B, Dowd J, McFadden M. 2010. Modes of developmental outgrowth and shaping of a craniofacial bone in zebrafish. *PLoS One* **5**:e9475. DOI: https://doi.org/10.1371/journal.pone.0009475, PMID: 20221441

Kirschner K, Chandra T, Kiselev V, Flores-Santa Cruz D, Macaulay IC, Park HJ, Li J, Kent DG, Kumar R, Pask DC, Hamilton TL, Hemberg M, Reik W, Green AR. 2017. Proliferation drives aging-related functional decline in a subpopulation of the hematopoietic stem cell compartment. *Cell Reports* **19**:1503–1511. DOI: https://doi.org/10.1016/j.celrep.2017.04.074, PMID: 28538171

Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, Natarajan KN, Reik W, Barahona M, Green AR, Hemberg M. 2017. SC3: consensus clustering of single-cell RNA-seq data. *Nature Methods* **14**:483–486. DOI: https://doi.org/10.1038/nmeth.4236, PMID: 28346451

Kreidberg JA, Sariola H, Loring JM, Maeda M, Pelletier J, Housman D, Jaenisch R. 1993. WT-1 is required for early kidney development. *Cell* **74**:679–691. DOI: https://doi.org/10.1016/0092-8674(93)90515-R, PMID: 8395349

Lawson L, Harfe BD. 2015. Notochord to nucleus pulposus transition. *Current Osteoporosis Reports* **13**:336–341. DOI: https://doi.org/10.1007/s11914-015-0284-x, PMID: 26231139

Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research* **41**:e108. DOI: https://doi.org/10.1093/nar/gkt214, PMID: 23558742

Liu H, Li B. 2010. p53 control of bone remodeling. *Journal of Cellular Biochemistry* **111**:529–534. DOI: https://doi.org/10.1002/jcb.22749, PMID: 20589754

Maheswaran S, Englert C, Bennett P, Heinrich G, Haber DA. 1995. The WT1 gene product stabilizes p53 and inhibits p53-mediated apoptosis. *Genes & Development* **9**:2143–2156. DOI: https://doi.org/10.1101/gad.9.17.2143, PMID: 7657166

Maheswaran S, Park S, Bernard A, Morris JF, Rauscher FJ, Hill DE, Haber DA. 1993. Physical and functional interaction between WT1 and p53 proteins. *PNAS* **90**:5100–5104. DOI: https://doi.org/10.1073/pnas.90.11.5100, PMID: 8389468

Manoli M, Driever W. 2012. Fluorescence-activated cell sorting (FACS) of fluorescently tagged cells from zebrafish larvae for RNA isolation. *Cold Spring Harbor Protocols* **2012**:pdb.prot069633. DOI: https://doi.org/10.1101/pdb.prot069633, PMID: 22854565

Martínez-Estrada OM, Lettice LA, Essafi A, Guadix JA, Slight J, Velecela V, Hall E, Reichmann J, Devenney PS, Hohenstein P, Hosen N, Hill RE, Muñoz-Chapuli R, Hastie ND. 2010. Wt1 is required for cardiovascular

progenitor cell formation through transcriptional control of Snail and E-cadherin. *Nature Genetics* **42**:89–93. DOI: https://doi.org/10.1038/ng.494, PMID: 20023660

McCarthy DJ, Campbell KR, Lun ATL, Wills QF. 2017. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **247**:btw777. DOI: https://doi.org/10.1093/bioinformatics/btw777

Menke AL, Clarke AR, Leitch A, Ijpenberg A, Williamson KA, Spraggon L, Harrison DJ, Hastie ND. 2002. Genetic interactions between the Wilms' tumor 1 gene and the p53 gene. *Cancer Research* **62**:6615–6620. PMID: 12438257

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, et al. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**:267–273. DOI: https://doi.org/10.1038/ng1180, PMID: 12808457

Motamedi FJ, Badro DA, Clarkson M, Lecca MR, Bradford ST, Buske FA, Saar K, Hübner N, Brändli AW, Schedl A. 2014. WT1 controls antagonistic FGF and BMP-pSMAD pathways in early renal progenitors. *Nature Communications* **5**:4444. DOI: https://doi.org/10.1038/ncomms5444, PMID: 25031030

Parsons R. 1977. *The Vertebrate Body*. fifth edition. Philadelphia: W.B.Saunders Company.

Patton EE, Widlund HR, Kutok JL, Kopani KR, Amatruda JF, Murphey RD, Berghmans S, Mayhall EA, Traver D, Fletcher CD, Aster JC, Granter SR, Look AT, Lee C, Fisher DE, Zon LI. 2005. BRAF mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Current Biology* **15**:249–254. DOI: https://doi.org/10.1016/j.cub.2005.01.031, PMID: 15694309

Perner B, Englert C, Bollig F. 2007. The Wilms tumor genes wt1a and wt1b control different steps during formation of the zebrafish pronephros. *Developmental Biology* **309**:87–96. DOI: https://doi.org/10.1016/j.ydbio.2007.06.022, PMID: 17651719

Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods* **10**:1096–1098. DOI: https://doi.org/10.1038/nmeth.2639, PMID: 24056875

Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols* **9**:171–181. DOI: https://doi.org/10.1038/nprot.2014.006, PMID: 24385147

Piek E, Sleumer LS, van Someren EP, Heuver L, de Haan JR, de Grijs I, Gilissen C, Hendriks JM, van Ravestein-van Os RI, Bauerschmidt S, Dechering KJ, van Zoelen EJ. 2010. Osteo-transcriptomics of human mesenchymal stem cells: accelerated gene expression and osteoblast differentiation induced by vitamin D reveals c-MYC as an enhancer of BMP2-induced osteogenesis. *Bone* **46**:613–627. DOI: https://doi.org/10.1016/j.bone.2009.10.024, PMID: 19857615

Reumann MK, Strachna O, Yagerman S, Torrecilla D, Kim J, Doty SB, Lukashova L, Boskey AL, Mayer-Kuckuk P. 2011. Loss of transcription factor early growth response gene 1 results in impaired endochondral bone repair. *Bone* **49**:743–752. DOI: https://doi.org/10.1016/j.bone.2011.06.023, PMID: 21726677

Schnabel K, Wu CC, Kurth T, Weidinger G. 2011. Regeneration of cryoinjury induced necrotic heart lesions in zebrafish is associated with epicardial activation and cardiomyocyte proliferation. *PLoS One* **6**:e18503. DOI: https://doi.org/10.1371/journal.pone.0018503, PMID: 21533269

Schurgers LJ, Uitto J, Reutelingsperger CP. 2013. Vitamin K-dependent carboxylation of matrix Gla-protein: a crucial switch to control ectopic mineralization. *Trends in Molecular Medicine* **19**:217–226. DOI: https://doi.org/10.1016/j.molmed.2012.12.008, PMID: 23375872

Smart N, Bollini S, Dubé KN, Vieira JM, Zhou B, Davidson S, Yellon D, Riegler J, Price AN, Lythgoe MF, Pu WT, Riley PR. 2011. De novo cardiomyocytes from within the activated adult heart after injury. *Nature* **474**:640–644. DOI: https://doi.org/10.1038/nature10188, PMID: 21654746

Stemple DL. 2005. Structure and function of the notochord: an essential organ for chordate development. *Development* **132**:2503–2512. DOI: https://doi.org/10.1242/dev.01812, PMID: 15890825

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* **102**:15545–15550. DOI: https://doi.org/10.1073/pnas.0506580102, PMID: 16199517

Sweatt A, Sane DC, Hutson SM, Wallin R. 2003. Matrix Gla protein (MGP) and bone morphogenetic protein-2 in aortic calcified lesions of aging rats. *Journal of Thrombosis and Haemostasis* **1**:178–185. DOI: https://doi.org/10.1046/j.1538-7836.2003.00023.x, PMID: 12871556

Toska E, Roberts SG. 2014. Mechanisms of transcriptional regulation by WT1 (Wilms' tumour 1). *Biochemical Journal* **461**:15–32. DOI: https://doi.org/10.1042/BJ20131587, PMID: 24927120

Walker MB, Kimmel CB. 2007. A two-color acid-free cartilage and bone stain for zebrafish larvae. *Biotechnic & Histochemistry* **82**:23–28. DOI: https://doi.org/10.1080/10520290701333558, PMID: 17510811

Wang J, Duncan D, Shi Z, Zhang B. 2013. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Research* **41**:W77–W83. DOI: https://doi.org/10.1093/nar/gkt439, PMID: 23703215

Wang T, Wang Y, Menendez A, Fong C, Babey M, Tahimic CG, Cheng Z, Li A, Chang W, Bikle DD. 2015. Osteoblast-specific loss of IGF1R signaling results in impaired endochondral bone formation during fracture healing. *Journal of Bone and Mineral Research* **30**:1572–1584. DOI: https://doi.org/10.1002/jbmr.2510, PMID: 25801198

White RM, Sessa A, Burke C, Bowman T, LeBlanc J, Ceol C, Bourque C, Dovey M, Goessling W, Burns CE, Zon LI. 2008. Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell* **2**:183–189. DOI: https://doi.org/10.1016/j.stem.2007.11.002, PMID: 18371439

Wojciechowska S, Zeng Z, Lister JA, Ceol CJ, Patton EE. 2016. Melanoma Regression and Recurrence in Zebrafish. *Methods in Molecular Biology* **1451**:143–153. DOI: https://doi.org/10.1007/978-1-4939-3771-4_10, PMID: 27464806

Yamamoto M, Morita R, Mizoguchi T, Matsuo H, Isoda M, Ishitani T, Chitnis AB, Matsumoto K, Crump JG, Hozumi K, Yonemura S, Kawakami K, Itoh M. 2010. Mib-Jag1-Notch signalling regulates patterning and structural roles of the notochord by controlling cell-fate decisions. *Development* **137**:2527–2537. DOI: https://doi.org/10.1242/dev.051011, PMID: 20573700

Zebboudj AF, Imura M, Boström K. 2002. Matrix GLA protein, a regulatory protein for bone morphogenetic protein-2. *Journal of Biological Chemistry* **277**:4388–4394. DOI: https://doi.org/10.1074/jbc.M109683200, PMID: 11741887

Appendix 3: Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity

WILEY | Aging Cell

Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity

Nelly Olova[1],[*] | Daniel J. Simpson[1],[*] | Riccardo E. Marioni[2] | Tamir Chandra[1]

[1]MRC Human Genetics Unit, MRC, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

[2]Centre for Cognitive Ageing and Cognitive Epidemiology, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

Correspondence
Tamir Chandra, MRC Human Genetics Unit, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, UK.
Email: tamir.chandra@igmm.ed.ac.uk

Abstract

Induced pluripotent stem cells (IPSCs), with their unlimited regenerative capacity, carry the promise for tissue replacement to counter age-related decline. However, attempts to realize in vivo iPSC have invariably resulted in the formation of teratomas. Partial reprogramming in prematurely aged mice has shown promising results in alleviating age-related symptoms without teratoma formation. Does partial reprogramming lead to rejuvenation (i.e., "younger" cells), rather than dedifferentiation, which bears the risk of cancer? Here, we analyse the dynamics of cellular age during human iPSC reprogramming and find that partial reprogramming leads to a reduction in the epigenetic age of cells. We also find that the loss of somatic gene expression and epigenetic age follows different kinetics, suggesting that they can be uncoupled and there could be a safe window where rejuvenation can be achieved with a minimized risk of cancer.

KEYWORDS
aging, aging clock, epigenetic age, iPSC, partial reprogramming, rejuvenation

1 | INTRODUCTION, RESULTS AND DISCUSSION

The human aging process is accompanied by multiple degenerative diseases. Our understanding of such aging related disorders is, nevertheless, fragmented, and the existence and nature of a general underlying cause are still much debated (Faragher, 2015; Gladyshev & Gladyshev, 2016). The generation of induced pluripotent stem cells (iPSCs) allows the reprogramming of somatic cells back to an embryonic stem cell (ESC)-like state with an unlimited regenerative capacity. This has led to multiple strategies for tissue replacement in degenerative diseases (Takahashi et al., 2007). Clinical application of

iPSCs, however, is at its infancy (Singh, Kalsan, Kumar, Saini, & Chandra, 2015; Soria-Valles et al., 2015; Takahashi & Yamanaka, 2016), and the potency of iPSCs bears risks, not least cancer induction. For example, in vivo experiments with iPSCs have shown that continuous expression of Yamanaka factors (Oct4, Sox2, Klf4 and c-Myc, thus OSKM) in adult mice invariably leads to cancer (Abad et al., 2013; Ohnishi et al., 2014).

To avoid this risk, a parallel concept of epigenetic rejuvenation has been proposed: the aging process in cells can be reversed whilst avoiding dedifferentiation (Manukyan & Singh, 2012; Singh & Zacouto, 2010). In other words, an old dysfunctional heart cell could be rejuvenated without the need for it to be passed through an embryonic/iPSC state. The concept of epigenetic rejuvenation

*These authors contributed equally to this work.

requires that rejuvenation and dedifferentiation each follow a distinct pathway. Nevertheless, it is not well understood whether rejuvenation and dedifferentiation are invariably intertwined, or instead whether it is possible to manipulate age without risking dedifferentiation.

The epigenetic rejuvenation potential of partial reprogramming with OSKM factors was previously shown by the forced expression of OSKM+LIN28 in senescent human fibroblasts, which led to recovering the high mobility of histone protein 1β by day 9, a feature characteristic for young fibroblasts (Manukyan & Singh, 2014). Ocampo et al. further demonstrated that partial reprogramming by transient cyclic induction of OSKM ameliorates signs of aging and extends lifespan in progeroid mice, with no resulting teratoma formation (Ocampo et al., 2016). This established partial reprogramming as a promising candidate intervention for age-related disease. Estimating epigenetic age, which is a promising molecular proxy for biological age (Jylhävä, Pedersen, & Hägg, 2017; Wagner, 2017), was, however, not possible to measure in mice at the time of the Ocampo study. This has left the nature (i.e., dedifferentiation/rejuvenation) of the described cellular changes unexplored:

1. Does the epigenetic remodelling seen truly reflect rejuvenation (i.e., a reduction in cellular/tissue age)? If so, can we observe a decrease in epigenetic age in partially reprogrammed human cells?
2. What is the extent of rejuvenation upon reaching a partially reprogrammed state (e.g., years of epigenetic age decrease)?
3. What are the dynamics of dedifferentiation in early reprogramming?

A major obstacle in understanding the relation between differentiation and aging has been our inability to accurately measure cellular age with a high correlation to the chronological age of the organism. However, over the last five years, a number of age predictors have been developed, the most accurate of which utilize DNA methylation (known as epigenetic clocks) (Hannum et al., 2013; Horvath, 2013; Horvath et al., 2018; Levine et al., 2018; Weidner et al., 2014), with the first Horvath multitissue age predictor being the most widely applicable and used ($r = 0.96$). This "Horvath clock" shows the highest correlation to chronological age, predicting the age (or epigenetic age, eAge) of multiple tissues with a median error of 3.6 years (Horvath, 2013). eAge is distinct from and poorly correlated with other age-related biomarkers, such as senescence and telomere length, which have been shown to correlate independently with the process of aging (Lowe, Horvath, & Raj, 2016; Marioni et al., 2016). Moreover, an acceleration of epigenetic age as measured by the "Horvath clock" is associated with a higher risk of all-cause mortality (Christiansen et al., 2016; Marioni et al., 2015; Perna et al., 2016), premature aging syndromes (Down and Werner) (Horvath et al., 2015; Maierhofer et al., 2017), frailty and menopause (Breitling et al., 2016; Levine et al., 2016). All of these studies suggest that eAge may capture a degree of biological aging.

To understand the dynamics of eAge during reprogramming, we applied Horvath's multitissue age predictor over a previously published reprogramming time course on human dermal fibroblasts (HDFs) (Horvath, 2013; Ohnuki et al., 2014). After OSKM transfection, successfully transformed subpopulations were isolated and analysed at regular time points during 49 days for gene expression and DNA methylation (detailed schematic shown in Supporting Information Figure S1). Epigenetic rejuvenation, that is, decrease in eAge, commenced between days 3 and 7 after OSKM transduction in the partially reprogrammed TRA-1-60 (+) cells (characterized in Tanabe, Nakamura, Narita, Takahashi, & Yamanaka, 2013) and continued steadily until day 20, when eAge was stably reset to zero (Figure 1a). A broken stick model (comprising two linear regressions joined at a break point) showed a good fit to the observed data starting from day 3 and measured a steady decrease with 3.8 years per day until day 20 (SE 0.27, $p = 3.8 \times 10^{-7}$) (Figure 1a). The TRA-1-60 (+) cell populations at days 7 and 11 have been previously characterized as "partially reprogrammed" for their high expression of pluripotency markers but also high reversion rates towards somatic state (Tanabe et al., 2013). Therefore, the observed eAge decline at days 7 and 11 suggests that partial reprogramming can indeed be considered a rejuvenation mechanism in human cells.

Horvath's multitissue age predictor is the most accurate and widely used for various cell types and tissues (Wagner, 2017). Nevertheless, we calculated eAge from alternative DNA methylation-based age predictors: four tissue-specific clocks (Hannum et al., 2013; Horvath et al., 2018; Weidner et al., 2014), one that incorporates clinical measures, called PhenoAge (Levine et al., 2018), and individual CpGs previously correlated with age (Garagnani et al., 2012). All clocks consistently reached the point of reset to their iPSC eAge at day 20, despite the cells not being fully reprogrammed before day 28 (Ohnuki et al., 2014) (Supporting Information Figure S2). Again, eAge showed a steady decline from day 3 to day 20 in the skin and blood and Weidner 99 CpG clocks, PhenoAge declined from day 7 to day 20, whilst the Hannum and Weidner 3 CpG clocks did not produce informative trajectories. Overall, eAge values and "years" of decrease varied between the clocks (actual chronological age of HDF donors is not available for reference) (Supporting Information Figure S2). The highest age associated individual CpG (ELOVL2's cg16867657) showed a similar trajectory to the Horvath eAge decline; however, the remaining CpGs produced inconsistent trajectories (Supporting Information Figure S2). The observed differences are not surprising, given that the alternative clocks were validated for blood (Hannum et al., 2013; Weidner et al., 2014), forensic applications (Horvath et al., 2018), whole organisms (Levine et al., 2018) or various tissues as for the individual CpGs (Garagnani et al., 2012).

In Ocampo et al. partial reprogramming was achieved after just two days of OKSM induction in mice carrying an inducible OSKM transgene (Ocampo et al., 2016). However, such "secondary" systems for direct reprogramming are known to have up to 50-fold higher efficiency and accelerated kinetics in comparison with virally transduced in vitro systems (Wernig et al., 2008). To facilitate comparison
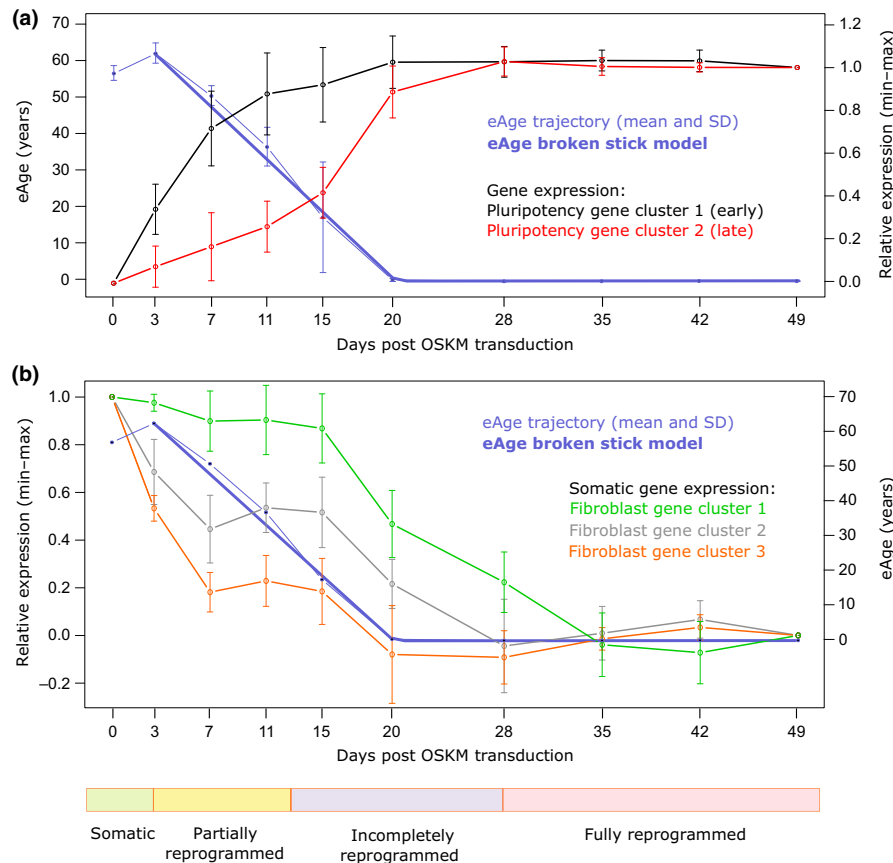
FIGURE 1 Dynamics of eAge and gene expression in a 49-day HDF reprogramming time course. (a) Left Y axis: eAge trajectory of Horvath's multitissue age predictor calculated from DNA methylation arrays from the following cell populations: day 0 (HDFs), day 3 (OSKM-expressing EGFP (+) HDFs), day 7, 11, 15, 20 and 28 (human pluripotency marker TRA-1–60 (+) cells at intermediate stages of reprogramming), and fully reprogrammed iPSCs from days 35, 42 and 49. Data were fit with a broken stick model composed of two linear sections. Error bars represent *SD*. Measured rate (years per day) of eAge decrease [day 3 – day 20] = −3.8, *SE* 0.27, *p* = 3.8×10^{-7}. Right Y axis: Composite gene expression trajectories of key pluripotency markers statistically clustered as per Genolini, Alacoque, and Marianne Sentenac (2015). Microarray expression data were obtained for the same time points and cell subpopulations as for eAge. Relative expression values were log2-transformed and presented as arbitrary units starting from "0" for "day 0" to "1" for "day 49." Error bars represent *SD*. (b) Left Y axis: Composite gene expression trajectories of key fibroblast markers generated as described for the pluripotency markers in (a). Relative expression values were presented as arbitrary units starting from "1" for "day 0" to "0" for "day 49." Right Y axis: eAge as in (a, left Y axis), without *SD*

to other systems and associate eAge with intermediate states in the reprogramming trajectory, we compared it to gene expression measured in the same samples. We analysed corresponding microarray expression data for 19 well-established pluripotency marker genes (Table 1 and Supporting Information Figure S3) as a proxy for reaching a mature pluripotent state (Boyer et al., 2005; Cai et al., 2006; Galan et al., 2013; Ginis et al., 2004; Mallon et al., 2013). We statistically clustered the expression patterns of those genes (Genolini et al. 2015), which resulted in two composite trajectories. These followed previously described expression dynamics of early (cluster 1) and late (cluster 2) activated pluripotency genes (Figure 1a) (Buganim et al., 2012; Chung et al., 2014; Takahashi & Yamanaka, 2016; Tanabe et al., 2013). Pluripotency gene cluster 1 included *NANOG*, *SALL4*, *ZFP42*, *TRA-1-60*, *UTF1*, *DPPA4* and *LEFTY2*, and their expression increased dramatically within the first 10 days and then established stable pluripotency expression levels by day 20. In contrast, pluripotency gene cluster 2 (containing late expressing genes such as

LIN28, *ZIC3* and *DNMT3B*) elevated expression more slowly and reached stable pluripotency levels by day 28 (Chung et al., 2014; Tanabe et al., 2013). Interestingly, eAge resets to zero at the same time that the genes in cluster 1 reached their pluripotent state levels, which temporally precedes full pluripotency. This also coincided with a peak in expression of a number of embryonic developmental genes between days 15 and 20, and might suggest that the reset marks a point where the cells reach an embryonic-like state but are not yet fully pluripotent (Table 1 and Supporting Information Figure S4). In summary, eAge decline is observed well within the first wave of pluripotency gene expression.

Therapeutic partial reprogramming will depend on rejuvenation with minimal dedifferentiation, which carries the risk of malignancies. We studied the dynamics of fibroblast gene downregulation as a proxy for the loss of somatic cell identity. The individual trajectories of 19 commonly used fibroblast marker genes (Chang, Li, & Guo, 2014; Goodpaster et al., 2008; Janmaat et al., 2015; Kalluri &

TABLE 1 List of pluripotency and fibroblast marker genes used in gene expression clusters

Marker	Gene	Protein name	Accession	Cluster
Pluripotency	NANOG	Nanog homeobox	A_23_P204640	1 (early)
Pluripotency	REX1 (ZFP42)	Zinc Finger Protein 42	A_23_P395582	1 (early)
Pluripotency	TRA−1−60/81 (PODXL)	Podocalyxin	A_23_P215060	1 (early)
Pluripotency	UTF1	Undifferentiated embryonic cell transcription factor 1	A_33_P3294217	1 (early)
Pluripotency	DPPA4	Developmental pluripotency associated 4	A_23_P380526	1 (early)
Pluripotency	TDGF1 (CRIPTO)	Teratocarcinoma-derived growth factor 1	A_23_P366376	1 (early)
Pluripotency	SALL4	Spalt-like transcription factor 4	A_23_P109072	1 (early)
Pluripotency	LEFTY1	Left–right determination factor 1	A_23_P160336	1 (early)
Pluripotency	LEFTY2	Left–right determination factor 2	A_23_P137573	1 (early)
Pluripotency	DNMT3A	DNA methyl-transferase 3A	A_23_P154500	1 (early)
Pluripotency	TFCP2L1	Transcription factor CP2-like 1	A_23_P5301	1 (early)
Pluripotency	TERF1	Telomeric repeat binding factor (NIMA-interacting) 1	A_23_P216149	2 (late)
Pluripotency	DPPA5	Developmental pluripotency associated 5	A_32_P233950	2 (late)
Pluripotency	TERT	Telomerase reverse transcriptase	A_23_P110851	2 (late)
Pluripotency	ZIC3	Zic family member 3	A_23_P327910	2 (late)
Pluripotency	LIN28a	LIN28 homolog A	A_23_P74895	2 (late)
Pluripotency	LIN28b	LIN28 homolog B	A_33_P3220615	2 (late)
Pluripotency	LECT1	Leukocyte cell derived chemotaxin 1	A_23_P25587	2 (late)
Pluripotency	DNMT3B	DNA methyl-transferase 3B	A_23_P28953	2 (late)
Fibroblast	COL3A1	Pro-collagen a2(III)	A_24_P935491	1
Fibroblast	FSP-1	Fibroblast surface protein	A_23_P94800	1
Fibroblast	TGFB3	Transforming growth factor beta 3	A_23_P88404	1
Fibroblast	TGFB2	Transforming growth factor beta 2	A_24_P402438	1
Fibroblast	COL1A2	Pro-collagen a2(I)	A_24_P277934	2
Fibroblast	ITGA1	Integrin a1b1 (VLA-1)	A_33_P3353791	2
Fibroblast	DDR2	Discoidin-domain-receptor-2	A_23_P452	2
Fibroblast	P4HA3	Prolyl 4-hydroxylase	A_24_P290286	2
Fibroblast	THY1	Thy-1 cell surface antigen; CD90	A_33_P3280845	2
Fibroblast	FAP	Fibroblast activation protein	A_23_P56746	2
Fibroblast	CD248	Endosialin, TEM1	A_33_P3337485	2
Fibroblast	VIM	Vimentin	A_23_P161190	2
Fibroblast	COL1A1	Pro-collagen a1(I)	A_33_P3304668	3
Fibroblast	ITGA5	Integrin a5b1	A_23_P36562	3
Fibroblast	P4HA1	Prolyl 4-hydroxylase	A_33_P3214481	3
Fibroblast	P4HA2	Prolyl 4-hydroxylase	A_33_P3394933	3
Fibroblast	TGFB1	Transforming growth factor beta 1	A_24_P79054	3
Fibroblast	HSP47	Serpin family H member 1, SERPINH1	A_33_P3269203	–
Fibroblast	CD34	Hematopoietic progenitor cell antigen	A_23_P23829	–

Note. Key pluripotent marker genes were selected from Ginis et al. (2004); Cai et al. (2006); Mallon et al. (2013); Galan et al. (2013); Boyer et al. (2005). Fibroblast marker genes were selected from Kalluri and Zeisberg (2006); Zhou et al. (2016); Janmaat et al. (2015); Pilling et al. (2009); Chang et al. (2014); Goodpaster et al. (2008); MacFadyen et al. (2005).

Zeisberg, 2006; MacFadyen et al., 2005; Pilling, Fan, Huang, Kaul, & Gomer, 2009; Zhou, Yang, Randall Wickett, & Zhang, 2016) (Table 1 and Supporting Information Figure S5) clustered into three composite expression patterns, two of which (clusters 2 and 3) went into an immediate decline after OSKM induction (Figure 1b). However, one fibroblast-specific cluster (cluster 1) remained stable in its expression

for the first 15 days. Interestingly, after day 7, fibroblast-specific gene expression in clusters 2 and 3 stopped declining and plateaued until day 15, coinciding with a peak in expression of senescence markers between days 11 and 15 (Supporting Information Figure S6). Vimentin (*VIM*), for example, remained at 60% of maximal expression until day 15 of reprogramming, similarly to *FAP*, *CD248* and *COL1A2*

in cluster 2 (Supporting Information Figure S5). After day 15, fibroblast gene expression declined rapidly in all three clusters, and only by day 35 had all reached ESC expression levels, marking a complete loss of somatic identity (Figure 1b). Cluster 1, which contains the well-described indicators of fibroblast identity *FSP1*, *COL3A1* and *TGFB2/3* (Kalluri & Zeisberg, 2006), showed the slowest decline and was also the last to reach ESC expression levels. In summary, we found that a number of fibroblast-specific genes maintained high expression levels until day 15, by which time a substantial drop in eAge has been observed.

Epigenetic rejuvenation or the reversal of cellular age is a promising concept as it could avoid the oncogenic risks associated with dedifferentiation. Here, we analysed a reprogramming time-course on HDFs and show that eAge declines in partially reprogrammed cells before their somatic identity is entirely lost.

It is well established that partial reprogramming happens within an early, reversible phase during the iPSC reprogramming time-course, which involves the stochastic activation of pluripotency genes. It is followed by a more deterministic maturation phase with predictable order of gene expression changes, where cell fate is firmly bound towards pluripotency (Smith, Sindhu, & Meissner, 2016; Takahashi & Yamanaka, 2016). Indeed, it has been shown that mouse fibroblasts fail to become iPSC and revert to their original somatic state if OSKM expression is discontinued during the initial stochastic phase (Brambrink et al., 2008; Stadtfeld, Maherali, Breault, & Hochedlinger, 2008). Previously, Tanabe et al. showed that TRA-1-60 (+) cells at reprogramming days 7 and 11 have not yet reached maturation and are partially reprogrammed (Tanabe et al., 2013) but our analysis already shows a decrease in their eAge according to multiple age predictors (Figure 1a and Supporting Information Figure S2). We have also shown that a large proportion of fibroblast marker genes maintain relatively high levels of expression until day 15 (Figure 1b and Supporting Information Figure S5). Nearly, unchanged levels of expression on day 15 were previously also shown for a large proportion of somatic genes (Tanabe et al., 2013). Together with increased senescence gene expression between days 11 and 15 (Supporting Information Figure S6), this likely contributes to the high propensity of partially reprogrammed TRA-1-60 (+) cells to revert back to somatic phenotype before day 15 in the time-course (Tanabe et al., 2013). Interestingly, the stepwise decline of fibroblast gene expression coinciding with a peak in expression of senescence genes seems to delay the loss of somatic identity but not the expression of pluripotency genes. Taken together, the different dynamics between the stepwise fibroblast expression and the linear decline in eAge further indicate that dedifferentiation and epigenetic rejuvenation can be uncoupled.

Our data suggest a window of opportunity within the uncommitted reprogramming phase, where a decline of eAge happens alongside partial maintenance of fibroblast gene expression. A deeper understanding of the kinetics of rejuvenation will be required to master therapeutic partial reprogramming, since any progress of dedifferentiation, even in a small subpopulation, carries the risk of malignancies. Our bulk expression analysis does not allow for a precise definition of the safe rejuvenation boundaries, and further experiments on a single cell level and in in vivo conditions are needed to determine a safe epigenetic rejuvenation window in different reprogramming systems. Upon defining safe boundaries, consideration should also be given to the steep decline of eAge, which resets to zero well ahead of the establishment of a pluripotent state, according to a number of age predictors (Supporting Information Figure S2). Most likely, this marks the point of reaching prenatal or embryonic stage, as suggested by the peak in expression of key developmental genes (Supporting Information Figure S4).

The extent of epigenetic rejuvenation in years (human) or months (mouse), which can be achieved through partial reprogramming, also needs further attention and will most likely differ with the different reprogramming systems. The "Horvath clock" shows up to 10 years of rejuvenation in Ohnuki et al.'s system by day 7 and another 10 + years by day 11. However, the intrinsic median estimation error of 3.6 years in this age predictor, the varying eAge rejuvenation values between the different age predictors and the intrareplicate biological variation seen from the large error bars highlight the need for more experiments and repetitions before this is established with a higher certainty.

Despite the obvious differences in reprogramming kinetics, our results also suggest that the improvements observed by Ocampo et al. in their OSKM-inducible secondary reprogramming system might be due to epigenetic rejuvenation. It remains to be shown how stable in time the rejuvenated phenotype is in either of the systems. Further analysis is also needed regarding the effect of partial reprogramming on adult stem cells or premalignant cells, which have already shown a higher propensity of transforming to malignancy (Abad et al., 2013; Ohnishi et al., 2014). It is possible that a premalignant phenotype could be attenuated or amplified by partial reprogramming. In summary, our findings reveal exciting possibilities but also open a number of questions and highlight areas that need further attention.

CONFLICT OF INTEREST

The authors of this paper have no conflict of interests to declare.

REFERENCES

Abad, M., Mosteiro, L., Pantoja, C., Cañamero, M., Rayon, T., Ors, I., … Serrano, M. (2013). Reprogramming in vivo produces teratomas and iPS cells with totipotency features. *Nature*, *502*, 340–345. https://doi.org/10.1038/nature12586

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4*. *67*.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., … Young, R. A. (2005). Core transcriptional regulatory circuitry

in human embryonic stem cells. *Young*, 122, 947–956. https://doi.org/10.1016/j.cell.2005.08.020

Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H., & Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell*, 2, 151–159. https://doi.org/10.1016/j.stem.2008.01.004

Breitling, L. P., Saum, K.-U., Perna, L., Schöttker, B., Holleczek, B., & Brenner, H. (2016). Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clin. Epigenetics*, 8, 21. https://doi.org/10.1186/s13148-016-0186-5

Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., … Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, 150, 1209–1222. https://doi.org/10.1016/j.cell.2012.08.023

Cai, J., Chen, J., Liu, Y., Miura, T., Luo, Y., Loring, J. F., … Zeng, X. (2006). Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells*, 24, 516–530. https://doi.org/10.1634/stemcells.2005-0143

Chang, Y., Li, H., & Guo, Z. (2014). Mesenchymal stem cell-like properties in fibroblasts. *Cellular Physiology and Biochemistry*, 34, 703–714. https://doi.org/10.1159/000363035

Christiansen, L., Lenart, A., Tan, Q., Vaupel, J. W., Aviv, A., McGue, M., & Christensen, K. (2016). DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell*, 15, 149–154. https://doi.org/10.1111/acel.12421

Chung, K. M., Kolling, F. W., Gajdosik, M. D., Burger, S., Russell, A. C., & Nelson, C. E. (2014). Single cell analysis reveals the stochastic phase of reprogramming to pluripotency is an ordered probabilistic process. *PLoS ONE*, 9, e95304. https://doi.org/10.1371/journal.pone.0095304

Faragher, R. G. A. (2015). Should we treat aging as a disease? The consequences and dangers of miscategorisation. *Frontiers in Genetics*, 6, 1–7.

Galan, A., Diaz-Gimeno, P., Poo, M. E., Valbuena, D., Sanchez, E., Ruiz, V., … Simon, C. (2013). Defining the genomic signature of totipotency and pluripotency during early human development. *PLoS ONE*, 8, 20–23. https://doi.org/10.1371/journal.pone.0062135

Garagnani, P., Bacalini, M. G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., … Franceschi, C. (2012). Methylation of ELOVL2 gene as a new epigenetic marker of age Aging Cell. *Aging Cell*, 11, 1132–1134.

Genolini, C., Alacoque, X., & Marianne Sentenac, C. A. (2015). kml and kml3d: R packages to cluster longitudinal data. *Journal of Statistical Software*, 65, 1–34.

Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., … Rao, M. S. (2004). Differences between human and mouse embryonic stem cells. *Developmental Biology*, 269, 360–380.

Gladyshev, T. V., & Gladyshev, V. N. (2016). A Disease or Not a Disease? Aging As a Pathology. *Trends in Molecular Medicine*, 22, 995–996.

Goodpaster, T., Legesse-Miller, A., Hameed, M. R., Aisner, S. C., Randolph-Habecker, J., & Coller, H. A. (2008). An Immunohistochemical method for identifying fibroblasts in formalin-fixed, paraffin-embedded tissue. *Journal of Histochemistry and Cytochemistry*, 56, 347–358. https://doi.org/10.1369/jhc.7A7287.2007

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., … Friend, S. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 49, 359–367.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, 14, R115. https://doi.org/10.1186/gb-2013-14-10-r115

Horvath, S., Garagnani, P., Bacalini, M. G., Pirazzini, C., Salvioli, S., Gentilini, D., … Franceschi, C. (2015). Accelerated epigenetic aging in Down syndrome. *Aging Cell*, 14, 491–495. https://doi.org/10.1111/acel.12325

Horvath, S., Oshima, J., Martin, G. M., Lu, A. T., Quach, A., Cohen, H., … Raj, K. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany. NY)*, 10, 1758–1775.

Janmaat, C. J., De Rooij, K. E., Locher, H., De Groot, S. C., De Groot, J. C. M. J., Frijns, J. H. M., & Huisman, M. A. (2015). Human dermal fibroblasts demonstrate positive immunostaining for neuron- and glia-specific proteins. *PLoS ONE*, 10, 1–14. https://doi.org/10.1371/journal.pone.0145235

Jylhävä, J., Pedersen, N. L., & Hägg, S. (2017). Biological age predictors. *EBioMedicine*, 21, 29–36. https://doi.org/10.1016/j.ebiom.2017.03.046

Kalluri, R., & Zeisberg, M. (2006). Fibroblasts in cancer. *Nature Reviews Cancer*, 6, 392–401.

Kuznetsova, A., Brockhoff, P. B., & Bojesen Christensen, R. H.(2016). *lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-33*. Retrieved from https://cran.r-project.org/web/packages/lmerTest/index.html

Levine, M. E., Lu, A. T., Chen, B. H., Hernandez, D. G., Singleton, A. B., Ferrucci, L., … Horvath, S. (2016). Menopause accelerates biological aging. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 9327–9332.

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Hou, L., … Ferrucci, L. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, 10, 573–591. https://doi.org/10.18632/aging.101414

Lowe, D., Horvath, S., & Raj, K. (2016). Epigenetic clock analyses of cellular senescence and ageing. *Oncotarget*, 7, 8524–8531. https://doi.org/10.18632/oncotarget.7383

MacFadyen, J. R., Haworth, O., Roberston, D., Hardie, D., Webster, M. T., Morris, H. R., … Isacke, C. M. (2005). Endosialin (TEM1, CD248) is a marker of stromal fibroblasts and is not selectively expressed on tumour endothelium. *FEBS Letters*, 579, 2569–2575. https://doi.org/10.1016/j.febslet.2005.03.071

Maierhofer, A., Flunkert, J., Oshima, J., Martin, G. M., Haaf, T., & Horvath, S. (2017) Accelerated epigenetic aging in Werner syndrome. *Aging (Albany NY)*, 9, 1143–1152.

Mallon, B. S., Chenoweth, J. G., Johnson, K. R., Hamilton, R. S., Tesar, P. J., Yavatkar, A. S., … McKay, R. D. G. (2013). StemCellDB: The human pluripotent stem cell database at the National Institutes of Health. *Stem Cell Res.*, 10, 57–66. https://doi.org/10.1016/j.scr.2012.09.002

Manukyan, M., & Singh, P. B. (2012). Epigenetic rejuvenation. *Genes to Cells*, 17, 337–343. https://doi.org/10.1111/j.1365-2443.2012.01595.x

Manukyan, M., & Singh, P. B. (2014). Epigenome rejuvenation: HP1β mobility as a measure of pluripotent and senescent chromatin ground states. *Scientific Reports*, 4, 1–8.

Marioni, R. E., Harris, S. E., Shah, S., McRae, A. F., von Zglinicki, T., Martin-Ruiz, C., … Deary, I. J. (2016). The epigenetic clock and telomere length are independently associated with chronological age and mortality. *International Journal of Epidemiology*, 45, 424–432.

Marioni, R. E., Shah, S., McRae, A. F., Chen, B. H., Colicino, E., Harris, S. E., … Deary, I. J. (2015). DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biology*, 16, 25. https://doi.org/10.1186/s13059-015-0584-6

Ocampo, A., Reddy, P., Martinez-Redondo, P., Platero-Luengo, A., Hatanaka, F., Hishida, T., … Izpisua Belmonte, J. C. (2016). In vivo amelioration of age-associated hallmarks by partial reprogramming. *Cell*, 167, 1719–1733.e12.

Ohnishi, K., Semi, K., Yamamoto, T., Shimizu, M., Tanaka, A., Mitsunaga, K., … Yamada, Y. (2014). Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. *Cell*, 156, 663–677.

Ohnuki, M., Tanabe, K., Sutou, K., Teramoto, I., Sawamura, Y., Narita, M., … Takahashi, K. (2014). Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation

potential. *Proceedings of the National Academy of Sciences*, *111*, 12426–12431.

Perna, L., Zhang, Y., Mons, U., Holleczek, B., Saum, K.-U., & Brenner, H. (2016). Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clinical Epigenetics*, *8*, 64. https://doi.org/10.1186/s13148-016-0228-z

Pilling, D., Fan, T., Huang, D., Kaul, B., & Gomer, R. H. (2009). Identification of markers that distinguish monocyte-derived fibrocytes from monocytes, macrophages, and fibroblasts. *PLoS ONE*, *4*, 31–33. https://doi.org/10.1371/journal.pone.0007475

Singh, V. K., Kalsan, M., Kumar, N., Saini, A., & Chandra, R. (2015). Induced pluripotent stem cells: Applications in regenerative medicine, disease modeling, and drug discovery. *Frontiers in Cell and Developmental Biology*, *3*, 1–18. https://doi.org/10.3389/fcell.2015.00002

Singh, P. B., & Zacouto, F. (2010). Nuclear reprogramming and epigenetic rejuvenation. *Journal of Biosciences*, *35*, 315–319. https://doi.org/10.1007/s12038-010-0034-2

Smith, Z. D., Sindhu, C., & Meissner, A. (2016). Molecular features of cellular reprogramming and development. *Nature Reviews Molecular Cell Biology*, *17*, 139–154.

Soria-Valles, C., Osorio, F. G., Gutiérrez-Fernández, A., De Los, A. A., Bueno, C., Menéndez, P., … López-Otín, C. (2015). NF-κB activation impairs somatic cell reprogramming in ageing. *Nature Cell Biology*, *17*, 1004–1013.

Stadtfeld, M., Maherali, N., Breault, D. T., & Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell*, *2*, 230–240. https://doi.org/10.1016/j.stem.2008.02.001

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, *131*, 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Takahashi, K., & Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. *Nature Reviews Molecular Cell Biology*, *17*, 183–193.

Tanabe, K., Nakamura, M., Narita, M., Takahashi, K., & Yamanaka, S. (2013). Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 12172–12179.

Wagner, W. (2017). Epigenetic aging clocks in mice and men. *Genome Biology*, *18*, 107. https://doi.org/10.1186/s13059-017-1245-8

Weidner, C., Lin, Q., Koch, C., Eisele, L., Beier, F., Ziegler, P., … Wagner, W. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology*, *15*, R24. https://doi.org/10.1186/gb-2014-15-2-r24

Wernig, M., Lengner, C. J., Hanna, J., Lodato, M. A., Steine, E., Foreman, R., … Jaenisch, R. (2008). A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotechnology*, *26*, 916–924.

Zhou, L., Yang, K., Randall Wickett, R., & Zhang, Y. (2016). Dermal fibroblasts induce cell cycle arrest and block epithelial–mesenchymal transition to inhibit the early stage melanoma development. *Cancer Medicine*, *5*, 1566–1579. https://doi.org/10.1002/cam4.707

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Olova N, Simpson DJ, Marioni RE, Chandra T. Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity. *Aging Cell*. 2019;18:e12877. https://doi.org/10.1111/acel.12877

Appendix 4: Kidney Single-Cell Atlas Reveals Myeloid Heterogeneity in Progression and Regression of Kidney Disease

Kidney Single-Cell Atlas Reveals Myeloid Heterogeneity in Progression and Regression of Kidney Disease

Bryan R. Conway,[1] Eoin D. O'Sullivan [ID],[2] Carolynn Cairns,[1] James O'Sullivan,[1]
Daniel J. Simpson [ID],[3] Angela Salzano,[3] Katie Connor,[1,2] Peng Ding,[2] Duncan Humphries,[2]
Kevin Stewart,[1] Oliver Teenan,[1] Riinu Pius,[4] Neil C. Henderson,[2] Cécile Bénézech,[1]
Prakash Ramachandran [ID],[2] David Ferenbach [ID],[2] Jeremy Hughes,[2] Tamir Chandra,[3] and
Laura Denby [ID] [1]

[1]Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom
[2]Medical Research Council Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom
[3]Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom
[4]Centre for Medical Informatics, University of Edinburgh, Edinburgh, United Kingdom

ABSTRACT

Background Little is known about the roles of myeloid cell subsets in kidney injury and in the limited ability of the organ to repair itself. Characterizing these cells based only on surface markers using flow cytometry might not provide a full phenotypic picture. Defining these cells at the single-cell, transcriptomic level could reveal myeloid heterogeneity in the progression and regression of kidney disease.

Methods Integrated droplet– and plate-based single-cell RNA sequencing were used in the murine, reversible, unilateral ureteric obstruction model to dissect the transcriptomic landscape at the single-cell level during renal injury and the resolution of fibrosis. Paired blood exchange tracked the fate of monocytes recruited to the injured kidney.

Results A single-cell atlas of the kidney generated using transcriptomics revealed marked changes in the proportion and gene expression of renal cell types during injury and repair. Conventional flow cytometry markers would not have identified the 12 myeloid cell subsets. Monocytes recruited to the kidney early after injury rapidly adopt a proinflammatory, profibrotic phenotype that expresses *Arg1*, before transitioning to become *Ccr2*[+] macrophages that accumulate in late injury. Conversely, a novel *Mmp12*[+] macrophage subset acts during repair.

Conclusions Complementary technologies identified novel myeloid subtypes, based on transcriptomics in single cells, that represent therapeutic targets to inhibit progression or promote regression of kidney disease.

CKD affects approximately 10% of the global population[1] and is a major risk factor for ESKD and cardiovascular disease.[2–4] It is now recognized that CKD is not always progressive, but that regression of albuminuria and improvement in renal function can occur if the injurious stimulus is removed.[5–7] Furthermore, regression of established fibrosis, the best histologic predictor of outcome,[8] has been observed after prolonged normalization of blood glucose levels after successful pancreas transplantation.[9,10] However, the cellular and molecular

pathways mediating injury regression are poorly understood, partly because renal biopsies are rarely performed in patients who are clinically improving.

The innate immune system has been implicated in both progression and regression of fibrosis in multiple organs, including the kidney.[11–15] Recruitment of proinflammatory monocytes[16,17] to the injured kidney *via* CCL1-CCR2 signaling[18,19] may exacerbate tissue damage through the release of proinflammatory factors and by activating myofibroblasts. Tissue macrophages are heterogeneous and inherently plastic, and may adopt different phenotypes in response to environmental cues. Hence, they may be injurious, but, in addition, they may mediate repair by scavenging cell debris, degrading excess extracellular matrix (ECM), and by secreting factors that may promote regeneration of injured tissue.[20–23]

Most studies have used panels of cell surface markers to characterize myeloid cell subsets by flow cytometry; however, this approach is inherently biased and is unlikely to capture the full phenotypic spectrum. Recent advances in transcriptomics, including single-cell RNA sequencing (scRNA-seq), have facilitated detailed analysis of myeloid cells in the healthy kidney, and after AKI,[24–26] and in other organs.[27–31] However, macrophage heterogeneity during regression of fibrosis in the kidney remains uncertain. Hence, in this study, we employ scRNA-seq to characterize myeloid cell subsets in the reversible, unilateral-ureteric-obstruction (R-UUO) model, in which we[32] and others[33] have demonstrated regression of established tubulointerstitial fibrosis after reversal of obstruction. We identified myeloid cell subsets that were indistinguishable using standard flow cytometry markers, with the relative proportions of the subsets changing dynamically during injury and repair. Pseudotime analysis and paired blood exchange (PBE) support dynamic changes in monocyte and macrophage phenotype in response to induction and removal of injury. Similar myeloid cell phenotypes are observed in the human kidney, suggesting they may represent specific targets to slow progression of CKD or promote renal repair.

METHODS

Animal Models
All protocols and surgical procedures were approved by the Animal Ethics Committee, University of Edinburgh. Animal experiments were conducted in accordance with the United Kingdom Animals Scientific Procedures Act 1986, under Home Office project licenses 70/8093 and 70/8867.

R-UUO Model
The R-UUO model was performed as previously described.[32] Briefly, 8-week-old male C57BL/6JOlaHsd mice (Enviago) underwent laparotomy, and the left ureter was isolated and the distal portion was ligated twice with 6/0 black-braided silk

Significance Statement

The innate immune system is central to injury and repair in the kidney, but the heterogeneity of myeloid cell subsets behind these processes is unknown. Complementary technologies—including bulk tissue transcriptomics, integrated droplet– and plate-based single-cell RNA sequencing, and paired blood exchange—resolved myeloid cell heterogeneity in a murine model of reversible unilateral ureteric obstruction, creating a single-cell atlas. The identified novel myeloid subsets could be targeted to ameliorate injury or enhance repair, including an *Arg1*⁺ monocyte subset present during injury and *Mmp12*⁺ macrophages present during repair. Standard flow cytometry to detect cell surface markers would have missed these subsets. Complementary techniques capture the complexity and dynamics of monocyte, dendritic cell, and macrophage phenotypes in the injured and repairing kidney.

sutures close to the bladder. In mice that required reversible ureteric obstruction, a silastic tube was placed around the ureter immediately proximal to the ligature to prevent excessive dilation. After 7 days of obstruction, the ureter was reanastomosed into the bladder, and the peritoneum and skin were sutured closed. Mice were euthanized by carbon dioxide narcosis and dislocation of the neck at day 2 (UUO-2) or day 7 (UUO-7) after UUO, or 7 (R-UUO 1 week), 14 (R-UUO 2 weeks), or 28 days (R-UUO 4 weeks) after ureteric reanastomosis after 7 days of obstruction. SMART-seq2 studies used MacGreen mice,[34] in which the *Csf1r* promoter drives enhanced green fluorescent protein (EGFP) as a reporter specific to myeloid cells.

PBE
Male C57BL/6NCrl mice, which are homozygous for CD45.2, were paired with male Ly5.1 mice (Charles River), which are CD45.1/CD45.2 heterozygotes. Four pairs of mice were used. The PBE was performed as previously described.[35] Briefly, all animals had a right jugular venous catheter inserted before the UUO surgery. At 1 day post-UUO, 15×150-μl aliquots of whole blood were exchanged between each animal in the pair over a 20-minute period. Two pairs were euthanized at both 2 and 7 days post-UUO, and whole blood and kidney tissue (UUO and contralateral kidney) was harvested.

Immunohistochemistry
Kidney tissue was fixed and formalin-fixed, paraffin-embedded, 4-μm tissue sections were prepared. Sections were rehydrated and staining was performed using the Sequenza system (Thermo Scientific, Waltham, MA). Sections were incubated with the avidin/biotin blocking kit (SP2001; Vector Laboratories) and blocked with serum-free protein block (X0909; Dako). Tissue sections were incubated with primary antibody (Supplemental Table 1) diluted in antibody diluent (S202230; Dako UK Ltd.), overnight at 4°C, before incubation with biotinylated secondary antibody (Supplemental Table 1) for 30 minutes at room temperature. Vectastain RTU ABC Reagent (PK7100; Vector Laboratories)

was then applied, followed by incubation with the DAB+ Substrate Chromogen System (K3468; Dako), and then counterstaining with hematoxylin before dehydration and mounting with Pertex mounting medium (3808707E; Histolab Products AB). The stained section was scanned with a Zeiss Axio Scan.Z1 Slide Scanner (Carl Zeiss Microscopy). The percentage of DAB staining per section ($n=6$–8 per group) was determined by ImageJ.

Immunofluorescence

Slides were de-waxed in xylene (2×5 minutes), rehydrated, and antigen retrieved (7 minutes at 60% microwave power). Slides were allowed to cool at room temperature, mounted in Sequenza (Thermo Fisher) racks, rinsed twice in PBS, and blocked for 45 minutes in Gentex block (120 μl) at room temperature. Primary antibodies were incubated at the concentrations in Supplemental Table 1 in antibody diluent (Abcam) and incubated overnight at 4°C. Slides were then washed twice with PBS, and secondary antibodies were added at a concentration of 1:200, diluted in antibody diluent, and incubated at room temperature for 30 minutes. Slides were again washed twice with PBS and then blocked with Gentex block for 45 minutes at room temperature. The second primary antibody was added to the slides, and incubated at 4°C overnight. Slides were washed and secondary antibodies applied. For dual immunofluorescence, the slides were boiled in 10 mM citrate and blocked with serum-free protein block for 1 hour before incubation of the second primary antibody. The antibodies were visualized by incubation with tyramide red or green (Perkin Elmer) for 10 minutes. After washing, slides were mounted with 4′,6-diamidino-2-phenylindole (DAPI) Fluoromount-G (Southern Biotech) and a coverslip was applied before visualization.

Bone Marrow–Derived Macrophage Culture

The hind legs of C57BL/6JOlaHsd mice were removed before the skin and underlying muscle were excised with sterile scissors and forceps to isolate the femur. The bone marrow was then flushed out in DMEM (Gibco) containing 10% L929 conditioned medium, 10% FCS, and 1% penicillin-streptomycin. Cell suspensions were cultured for 1 week in 60-ml, sterile Teflon pots at 37°C with 5% carbon dioxide. Macrophages were then plated into six-well plates, incubated with 10 μg FITC-conjugated collagen (D12052), and left overnight. Cells were collected and run through flow cytometry to quantify the FITC signal.

RNA Extraction, Gene Expression, and Bulk RNA-seq

Total RNA from cortical kidney tissue was isolated using the RNeasy kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. For quantitative PCR analysis of targeted gene expression, cDNA was synthesized from 1 μg of template RNA, using the QuantiTect Reverse Transcription Kit (Qiagen, Venlo, Netherlands). Quantitative PCR was performed using the PerfeCTa FastMix II Probe Master

(VWR, Lutterworth, United Kingdom) and TaqMan Gene Expression Assay–specific primers (Life Technologies; Supplemental Table 2) and normalized to hypoxanthine-guanine phosphoribosyltransferase.

Before RNA-seq, RNA integrity was checked using Agilent Nano Chips, and only samples with an RNA integrity number greater than seven were used in subsequent analysis. Four mice per group underwent RNA-seq, with the animals selected on the basis that their *Havcr1* gene expression, as determined by quantitative RT-PCR, was closest to the mean of that group. A poly(A) library was constructed and run on a HiSeq2500, using 2×100-bp, paired-end (PE) sequencing. FastQC was used for initial quality control (QC), reads were mapped to the mm10 transcriptome using RSEM and Bowtie2, and DESeq2 was used for differential gene expression analysis. Data were deposited in the National Center for Biotechnology Information Gene Expression Omnibus database (accession number GSE145053).

The shinyNGS R package was used to generate gene clusters in Figure 1C. Genes with an average fragments per kilobase of transcript per million mapped reads of less than one across all groups were excluded from analysis. Genes that were not significantly differentially expressed (DE; adjusted $P<0.05$, determined using DESeq2 R package) in any group (compared with sham) were excluded from analysis. Using the feature-based clustering module of the shinyNGS package, 7810 genes were assigned to one of six clusters, based on expression change between each of the groups.

Kidney Digestion for Flow Cytometry and Single-Cell Sequencing

Immediately after euthanasia, mice were perfused with 10 ml PBS. Kidneys were excised, decapsulated, and placed in ice-cold PBS. Equal portions of renal cortex from each mouse were finely minced in digest buffer (4.25 mg/ml Collagenase V [Sigma-Aldrich, St. Louis, MO], 6.25 mg/ml Collagenase D [Roche, Basel, Switzerland], 10 mg/ml Dispase [Thermo Scientific], and 300 μg/ml DNase [Roche] in RPMI 1640 [10% FCS, 1% penicillin/streptomycin/L-glutamine]) before homogenization in gentleMACS C Tubes, using the gentleMACS Dissociator (Miltenyi Biotec, Auburn, CA). Samples were incubated at 37°C with shaking to maximize digestion. The kidney suspension was then subjected to a second gentleMACS homogenization and digestion, neutralized with an equal volume of FACS buffer (PBS, 2mM EDTA, and 2% FCS). Kidney cell suspensions were then passed sequentially through 100-, 70-, and 40-μm sieves. Any residual red blood cells were lysed by Red Blood Cell Lysis Buffer (Sigma-Aldrich). Cells were resuspended in ice-cold FACS buffer, ready for use.

Flow Cytometry

Single-cell suspensions were incubated with Fc Block (BD Biosciences, San Jose, CA) and then incubated with preconjugated antibodies (Supplemental Table 3) in round-bottomed plates. Controls were set up, including unstained
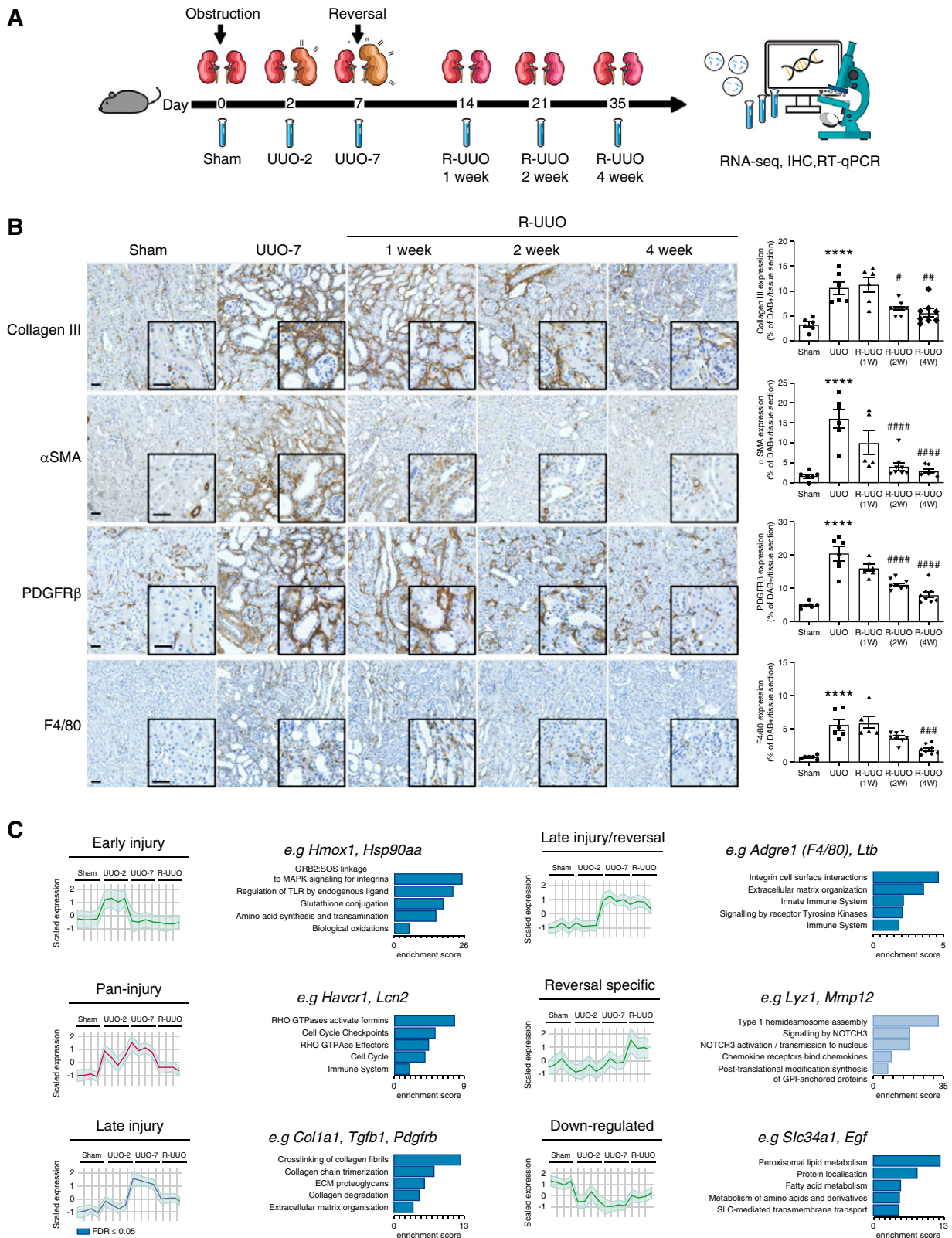
Figure 1. Phases of progression and regression of fibrosis in the R-UUO model are associated with dynamic changes in the renal transcriptome. (A) Male 6- to 8-week-old C57BL/6J mice underwent either UUO or sham surgery, and were either euthanized 2 days later, or left obstructed for 7 days and then euthanized, or had their ureter reimplanted to reverse obstruction before euthanasia at 1, 2, or 4 weeks post UUO (n=6–8 per group). (B) Representative images and quantification of fibrosis (collagen III) and fibroblast (PDGFR-β), myofibroblast (α-smooth muscle actin [α-SMA]), or macrophage (F4/80) accumulation during the R-UUO model. Scale bar, 50 μM.

cells, beads with single stains of each antibody, and fluorophore minus one controls. DAPI was used to distinguish live and dead cells. For cellular composition analysis, the following antibodies were used: CD45, CD31, LTL, PDGFRβ, and F4/80. For SMART-seq2 experiments, cells were incubated with the antibodies CD45, CD11b, CD11c, F4/80, MHCII, CD206, CD64, and CD24; with a lineage dump gate including TCRβ (T cells), CD19 (B cells), Siglec-F (eosinophils), and Ly6G (neutrophils); all conjugated to BV421 and run on the BD FACS ARIA II. For PBE, blood and tissue were analyzed on the BD 6L LSRFortessa using the following antibody panels: for blood, CD45.1, CD45.2, CD3, F4/80, GR-1, CD11b, and CD19; for kidney tissue, CD45.1, CD45.2, F4/80, Ly6C, MHCII, CD11b, CD206, CD24, and CD64; with a dump gate including TCRβ, CD19, Siglec-F, and Ly6G. All files were exported in FCS format and analyzed with FlowJo software version 10.

Single-Cell Droplet Library Preparation

For scRNA-seq analysis on the 10$\times$ Genomics platform, single-cell suspensions from renal cortex were prepared from pools of three animals from each group, as outlined by the 10$\times$ Genomics Single Cell 3$'$ Reagent Kit User Guide version 2. A total of 50,000 live (DAPI$-$) cells were sorted on the BD FACS ARIA II. Samples were washed twice in PBS (Sigma), followed by centrifugation at 500 $\times$ g for 5 minutes at 4°C. Sample viability was assessed using trypan blue (Sigma) with an automated cell counter (Bio-Rad), and the appropriate volume for each sample was calculated. The chip was loaded with 10,700 cells per lane.

After droplet generation, samples were transferred onto a prechilled 96-well plate, heat sealed, and reverse transcription was performed using a C1000 Touch Thermal Cycler (Bio-Rad). After reverse transcription, cDNA was recovered using the 10$\times$ Genomics Recovery Agent, and a Silane Dyna-Bead (Thermo Fisher) cleanup was performed. Purified cDNA was amplified and cleaned using SPRIselect beads (Beckman). Samples were diluted at 4:1 (elution buffer [Qiagen]/cDNA) and an initial concentration check was performed on a Qubit fluorometer (Invitrogen) to ensure adequate cDNA concentration. The final cDNA concentration was checked on a bioanalyzer (Invitrogen).

SMART-seq2 Library Preparation

For the SMART-seq2 experiment, we performed flow cytometry for $n=3$ animals per group, and sequenced one animal per group. The flow cytometry patterns within each group were broadly similar, mitigating against the selected animal being unrepresentative of the group. A single, live (DAPI$+$), EGFP$+$ (Csf1r$+$) cell was sorted into each well of a 96-well plate, and all fluorochrome information was recorded using the index-sort capability of the BD FACS ARIA II. Equal numbers of cells from each time point were sorted into each plate to reduce batch effect, with 192 cells per time point included in total. Single cells were processed as previously described.[36] Briefly, cells were lysed immediately in lysis buffer containing 5% RNase inhibitor and 0.025% Triton X-100. Oligo(dT) primers were added and reverse transcription was performed. SMART-seq2 libraries were prepared according to the previously described protocol,[36] with a few modifications[37]: at step 5, 0.1 μl of the External RNA Controls Consortium (ERCC) spike-in mix (10:5 diluted, 4456740; Life Technologies) was added with 0.1 μl of 100 μM oligo(dT) primer and 1 μl of dNTP mix and 0.8 μl of water, yielding the same concentrations of primer and oligo as originally reported. Fluidigm protocol (PN 100-7168 M1) was used for tagmentation library generation. The final cDNA concentration was checked on a bioanalyzer (Agilent).

Sequencing

The 10$\times$ libraries were pooled and normalized by molarity before being sequenced across four lanes on a single Illumina flow cell. Sequencing was performed on an Illumina HiSeq platform, with a target of approximately 350 million PE reads per lane, giving approximately 525 million PE reads per sample, comprising 2$\times$150-bp PE configuration and 8-bp index reads. The SMART-seq2 libraries were sequenced as 8$\times$ 96-well plates, which were pooled and sequenced on an Illumina HiSeq 4000 (50-bp, single-end reads). Data were deposited in the National Center for Biotechnology Information Gene Expression Omnibus database (accession number GSE140023).

scRNA-seq Analysis

For the droplet-based dataset, the cellranger mkfastq wrapper (Cell Ranger Single Cell Software suite 2.1.0, http://10xgenomics.com) de-multiplexed the Illumina output BCL files to library-specific FASTQ files. Subsequently, alignment was performed using the cellranger count function, using STAR aligner 2.5.1b[38] against the Ensembl mouse reference genome version GRCm38.68. Correction and filtering of cell bar code and unique molecular identifiers followed, and the retained bar codes were quantified and used to generate a gene

---

****$P<0.0001$ versus sham, #$P<0.05$ versus UUO, ##$P<0.01$ versus UUO, ###$P<0.001$, ####$P<0.0001$ versus UUO. (C) Unbiased clustering analysis of bulk RNA sequencing data from the renal cortex of mice ($n=4$ per group) during the R-UUO time course identified six discrete temporal patterns of gene expression. Representative genes and enriched pathways are provided for each cluster. The number of genes included in each cluster is as follows: Down-regulated 1562; Early injury, 779; Pan-injury, 1479; Late injury, 1619; Late injury/reversal, 1708; and Reversal specific, 663. Shaded error range is the SD of the mean scaled gene expression for each animal. Dark and light blue pathways are those demonstrating gene enrichment at a false discovery rate of <0.05 and >0.05, respectively. FDR, false discovery rate; IHC, immunohistochemistry; RT-qPCR, quantitative RT-PCR.

expression matrix. Summary sequencing statistics are provided in Supplemental Table 4. The SMART-seq2 raw reads were similarly mapped against the Ensembl mouse reference genome version GRCm38.68, using the STAR RNA-seq aligner,[38] with the additional inclusion of the sequences for the ERCC spike-ins.

For our droplet-based dataset, a standard sequence of filtering, highly variable gene selection, dimensionality reduction, and clustering were performed using the scRNA-seq analysis R package Seurat (version 2.3.4).[39] After alignment and initial preprocessing, we began our R workflow with 15,046 genes across 7073 cells in our sham group, 16,450 genes across 5088 cells in our UUO-2 group, 17,368 genes across 7124 cells in our UUO-7 group, and 17,227 genes across 6096 cells in our R-UUO group. To exclude low-quality cells in both single-cell experiments, we then filtered cells that expressed <300 genes and less than 500 unique molecular identifiers, and to exclude probable doublets, cells with >10,000 unique molecular identifiers and <3000 genes were removed. This would have removed the majority of injured and apoptotic cells.[40] We used a mitochondrial filter to remove cells in which >50% of genes were mitochondrial, consistent with other renal-specific scRNA-seq projects.[41,42] This is a higher filter than has been used in nonrenal single-cell analysis, but reflects the high mitochondrial content in renal tubular epithelial cells. Any gene not expressed in at least three cells was removed.

For the SMART-seq2 data, identical metrics were used as above, but the mitochondrial filter threshold was lowered to 25%, and cells with >25% of reads mapping to the ERCCs were additionally excluded. After filtering and QC, 15,046 genes across 4540 samples of sham mice, 16,450 genes across 3101 samples of the UUO-2 mice, 17,368 genes across 5563 samples of UUO-7 mice, 17,227 genes across 4308 samples of the R-UUO mice, and 13,517 genes across 362 samples in the SMART-seq2 data were taken forward for analysis, resulting in 92 cells in the sham group, 102 cells in UUO-2, 103 cells in UUO-7, and 65 cells from the R-UUO group.

Normalization was performed using the Seurat package to reduce biases introduced by technical variation, sequencing depth, and capture efficiency. We used the default global-scaling normalization method "logNormalize," which normalized gene expression per cell by the total expression and multiplied the result by a scaling factor before log transformation. We then scaled the data and regressed out variation between cells due to the number of unique molecular identifiers and the percentage of mitochondrial genes.

The expression matrix subsequently underwent dimensionality reduction, using principal component analysis of the highly variable genes within the dataset. Using Seurat's FindVariableGenes function (and computed using the LogVMR argument) we used log-mean expression values between 0.0125 and three, and a dispersion cutoff of 0.5 to select genes. Principal component analysis was performed using these selected genes, and 20 principal components were identified for subsequent analysis in each dataset, selected both visually using the elbow point on the elbow plot and *via* the jackstraw method.

Further cluster-based QC was performed in the droplet data using the density-based spatial clustering algorithm DBscan, which was used to identify cells on a t-distributed stochastic neighbor embedding (tSNE) map. We initially set an eps value of 0.5 and removed clusters with fewer than ten cells. The remaining cells were then clustered again with an eps value of one, followed by removing the clusters with <20 cells. Of note, this allowed identification of a cluster characterized by high expression of heat-shock genes, including *Fos*, *Jun*, and *Atf3*. This cluster was removed because it was considered to be an artifact of cell stress due to the experimental protocol, as recently described.[43] This procedure removed 158 (3.4%) cells from a total of 4540 cells in sham mice, 137 (4.4%) of 3101 cells in UUO-2 mice, 167 (3%) of 5563 cells in UUO-7 mice, and 83 (1.9%) of 4308 cells in R-UUO mice.

Clusters were then identified using Seurat's FindClusters function, built using the first ten principal components and a resolution parameter of 1.5. The original Louvain modularity optimization algorithm was used. tSNE (using the Rtsne package Barnes–Hut implementation) was then used for further dimensionality reduction and visualization, which was run on a reduced dimensional space of the first five to ten dimensions, using perplexity values of 15–50.

For all single-cell differential expression tests, we used the Wilcoxon rank-sum test to identify a unique expression profile for each cluster, with differential expression tested between each cluster and all other clusters combined. The FindAllMarkers test, as implemented in Seurat, returns an "adj_pval" (Bonferroni-adjusted *P* values) and an "avg_logFC" (average log fold change) for each gene. Genes were ranked in order of average log fold change and visualized using heatmaps.

The process of acquiring the myeloid subsets required a combination of cluster-based cell pruning and a gene-based cell filter. The clusters were annotated by cell type, using the generated DE marker genes, and recognized markers from known biology and from the single-cell literature to date. All myeloid clusters were then isolated, renormalized, and rescaled as described above, and were processed again through the pipeline described above, although without the DBscan QC step. Clustering resolution was lowered for each downstream implementation of the FindClusters function, and again clusters were identified based on DE genes. Any nonmyeloid cell clusters were removed, and the data were, once again, reprocessed and reclustered. This entire process was repeated three times in total, with successively higher resolutions used to generate greater numbers of clusters, to coerce any nonmyeloid cells into forming distinct clusters that could be removed as the data became cleaner. The expression of 52 key nonmyeloid genes was then assessed, and any cell expressing such genes was removed. Finally, tSNE graphs were visually inspected, and any unusual clusters were manually selected and DE genes inspected—any stray unwanted cluster was manually pruned and the data reprocessed as before.

Differential Proportion Analysis

Differential proportion analysis was developed to detect changes in population proportions across groups within single-cell experiments. The algorithm is described in detail, along with source code, by Farbehi *et al.*[44] Briefly, this approach uses a novel permutation-based statistical test to analyze whether observed changes in proportions of cell populations were greater than expected by chance. This approach attempts to consider sources of technical variation within the experimental technique, such as differing absolute cell numbers within the experiment, cell-type capture bias (a known feature of current single-cell workflows[45]), and variation due to *in silico* analysis (cluster assignment accuracy, for example). A proportion table of clusters per phenotype/group is created from the count table, and the difference in cluster proportion is compared with a null distribution. This distribution is constructed by random permutations of random subsamples of cluster labels across a random proportion of total cells. A new proportion table is generated from these data, and the process is repeated multiple times, with the resulting difference in cluster proportions across the data forming the null distribution. The observed distribution is then compared with this null distribution, and final *P* values are calculated based on the minimum *P* values of any observed increases or decreases in proportion. As per the original paper,[44] we used a w parameter of 0.1, where lower values will trend toward a stricter test (fewer significant hits), and higher values trend toward higher numbers of significant hits.

Assignment to Myeloid Cell Phenotype on the Immunological Genome Project Consortium

Cluster Identity Predictor (version 2[46]) was used to generate Spearman correlation values for each cluster within our own data, as compared with cluster gene signatures with Immunological Genome Project (ImmGen) mouse immune cell datasets based on the entire gene expression dataset. The algorithm first subsets genes common to both datasets before using a one-to-many (cluster-to-references) calculation of correlation coefficients. A single correlation coefficient is calculated for each reference cell type and for each cluster, allowing each cluster in our experiment to be analyzed against each known cell type in the reference file and scored for its overall similarity. To further validate these assignments, we used SingleR[27] to assign myeloid cell classification, using the murine ImmGen dataset, and create a consensus matrix with our classifications. Briefly, this pipeline is based on correlating reference bulk-transcriptomic datasets of pure cell types with single-cell gene expression. Similarly to the Cluster Identity Predictor, a Spearman coefficient was calculated for single-cell gene expression, with each of the samples in the reference dataset based only on the variable genes in the reference dataset. This is performed iteratively until a classification is reached. One myeloid cluster mapped to a mixture of cells and expressed cell cycle genes, such as *Mki67* and *Top2a*, consistent with proliferation and was therefore assigned to the proliferation cluster.

Assignment to Myeloid Phenotype in Recovery from Ischemia-Reperfusion Injury

To compare the transcriptome from our myeloid cell clusters with that of myeloid cells in the kidney during recovery from renal ischemia-reperfusion injury (IRI), we used the datasets generated by Lever *et al.*[47] and deposited in the Gene Expression Omnibus (GSE121410). We used SingleR software, as previously described, to align our macrophage clusters to embryonic and adult resident renal macrophages from healthy kidneys, and macrophages and infiltrating monocytes in kidneys from mice 6 days after IRI.

Ligand-Pair Interactions

Heatmaps and dotplots of number of ligand-pair interactions were generated using the CellphoneDB tool (https://www.cellphonedb.org/) developed by the Teichmann Lab (Wellcome Sanger Institute, Cambridge, United Kingdom).[48] The lower cut-off for expression proportion of any ligand or receptor in a given cell type was set to 10%, and the number of permutations was set to 1000. The clusters were not subsampled.

Platform Integration

Because our SMART-seq2 library included an index sort, the transformed FACS data corresponding to each cell was then exported as .fcs files and analyzed using FlowJo software. The index-sort data were then extracted using the index FlowJo plugin available from FlowJo exchange. These data were then matched to the cell bar code and imported as both a separate "protein" assay and as metadata into the Seurat object to allow for visualization.

Our droplet-based and SMART-seq2 datasets were then integrated using the "anchoring" approach introduced in Seurat version 3.[49] Here, we created an integrated reference dataset and transferred the cell type labels onto our SMART-seq2 data. Briefly, this approach requires identification of "anchors" between the datasets, which represent shared biologic states. This involves jointly reducing the dimensionality of both datasets using diagonalized canonic correlational analysis before searching for mutual nearest neighbors in the new shared space. The paired cells are treated as anchors that represented shared biology across the datasets. Anchors were identified using the default parameters of the FindIntegrationAnchors function, with the argument dims=1:20.

To map the monocyte, dendritic cell (DC), and macrophage clusters onto conventional flow cytometry myeloid cell gates, we used UUO-2, UUO-7, and R-UUO, respectively, because not all clusters were represented at all time points in the SMART-seq2 dataset.

Pseudotime Analysis

Lineage reconstruction and pseudotime inference was performed using the slingshot package.[50] This method works by learning cluster relationships in an unsupervised manner and constructing smooth curves representing the lineage across
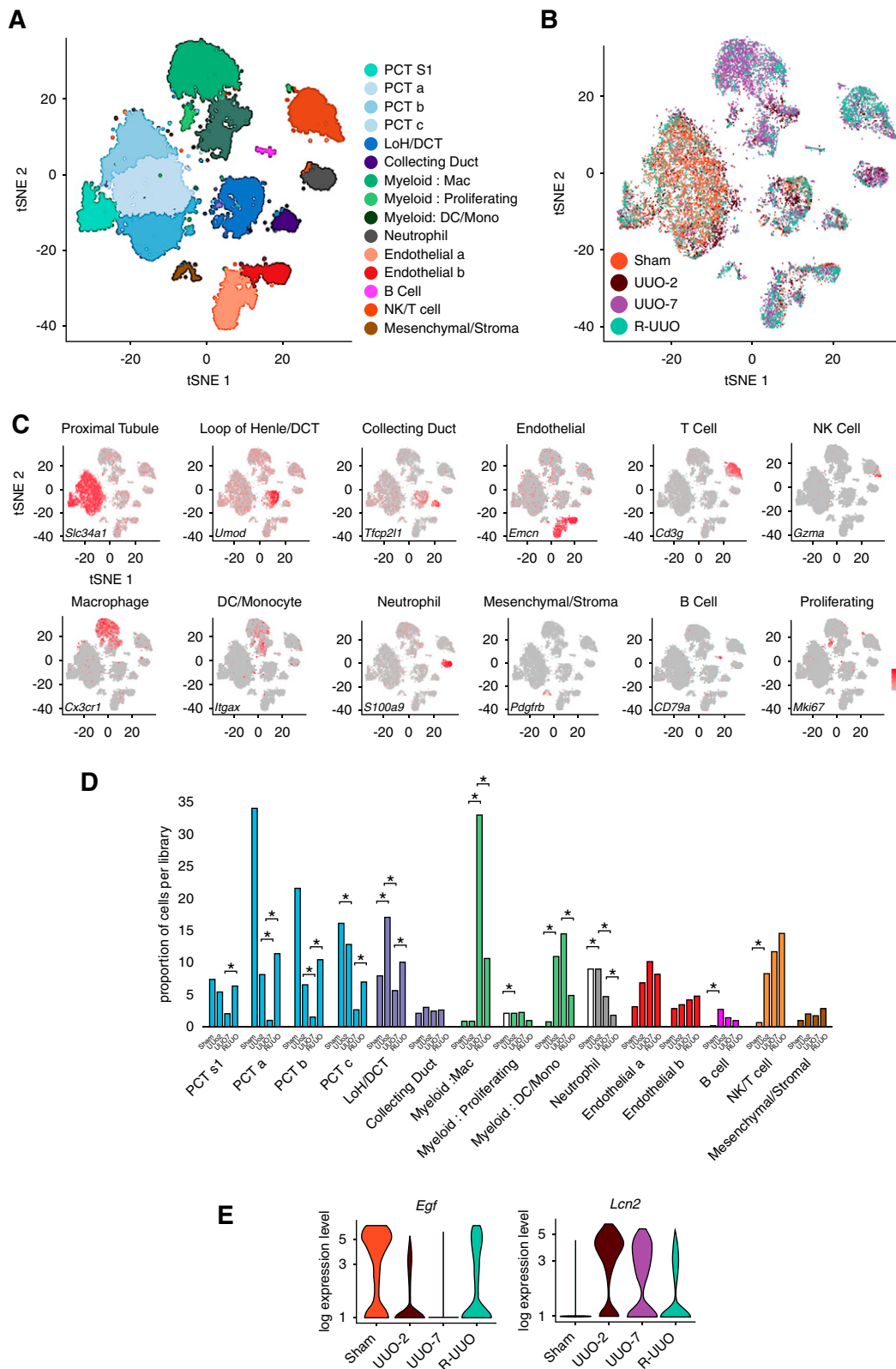
Figure 2. scRNA-seq analysis identifies discrete renal cell types, with dynamic changes in the proportion and transcriptome of each cell type observed across the R-UUO model. tSNE plots of 17,136 cells from libraries pooled from mice that underwent sham, UUO-2, UUO-7, or R-UUO (2 weeks) (*n*=3 per time point) classified by (A) cell cluster and (B) time point. (C) Expression of selected marker genes for each cell classification projected onto tSNE plot. Color scale is log$_{10}$ expression levels of genes. (D) Relative proportions of cells assigned to each cluster by time point. Statistical significance derived using differential proportional analysis, with a mean error of 0.1

two visualized pseudotime dimensions. Briefly, this involved first creating the raw expression matrix of the subsetted cells that were classified as "*Arg1+* monocyte," "*Ly6c2+* monocyte," and "*Ccr2+* macrophage" by Seurat. This was followed by filtering genes not expressed in any cluster with less than ten cells, and having at least three reads within that cluster. Full quantile normalization was then performed before dimensionality reduction using diffusion maps *via* the destiny package.[51] Next, cells were clustered to allow slingshot to infer a global structure of the lineage, using a Gaussian mixture modeling implemented in Mclust,[52] revealing eight underling clusters in the data. Slingshot then constructed the cluster-based minimum spanning tree and fitted the principal curve.

Pathway Analysis

Gene set enrichment analysis (GSEA) and over-representation analysis (ORA) were used to identify enriched pathways based on the DE genes, using WebGestalt (http://www.webgestalt.org/). For GSEA, we generated a rank for each gene in the list of DE genes using the formula rank=(average log fold change)×(−log [adjusted P value]). To perform ORA of bulk RNA-seq data, significantly DE genes were selected for the algorithm based on a minimum of two-fold upregulation (or 50% of baseline for the downregulated genes) against the appropriate comparator (*e.g.*, UUO-2 versus sham when considering early injury). Enrichment categories were discarded if they contained less than five or >2000 genes. These thresholds were calculated by WebGestalt based on the number of overlapping genes between the annotated genes in the category and the reference gene list for the ORA method. For the GSEA method, categories were discarded if they contained <15 genes or >500 genes. The Benjamini–Hochberg method was used to correct for multiple testing during ORA, and the top ten enriched categories—as ranked by false discovery rate—were selected. The reference gene list used was the Illumina MouseRef-8. We used pathway gene sets from the protein analysis through evolutionary relationships, PANTHER (http://www.pantherdb.org), Reactome, and Kyoto Encyclopedia of Genes and Genomes (https://www.genome.jp/kegg/) as our reference gene lists.

Statistical Analyses

Animal group size was determined from previous pilot experiments. Comparisons between two unpaired, non-normally distributed data points were carried out *via* Mann–Whitney test. Comparisons between two unpaired, normally distributed data points were carried out *via* t test. Comparisons between multiple groups were performed with one-way ANOVA

with the Tukey multiple comparison test. All statistical analysis was performed using GraphPad Prism version 10.

RESULTS

Degradation of Excess ECM after Reversal of Ureteric Obstruction Is Associated with Persistence of Immune Cells

To determine the pathways that mediate renal injury and repair, we used the murine R-UUO model (Figure 1A). Within 7 days of ureteric obstruction (UUO-7), there was expansion of interstitial PDGF-β^+ cells (Figure 1B), activation to α-smooth muscle actin$^+$ myofibroblasts (Figure 1B, Supplemental Figure 1A), and collagen deposition (Figure 1B), with induction of tubular injury markers such as *Havcr1* (encodes kidney injury molecule-1; Figure 1C). After reimplantation of the ureter, there was a decline in *Havcr1* expression, a significant reduction in interstitial PDGFR-β^+ cells with loss of myofibroblastic phenotype (α-smooth muscle actin$^-$) (Supplemental Figure 1A), and a gradual regression of collagen deposition over 4 weeks (Figure 1B), as has been observed previously.[33] Macrophages (F4/80$^+$) accumulated in the kidney during UUO, and persisted through the early stages of R-UUO, before trending toward baseline levels by 4 weeks after reversal (Figure 1B), consistent with previous findings in this model.[33]

We first performed bulk RNA-seq of the renal cortex in the R-UUO model, which revealed six discrete temporal patterns of gene expression (Figure 1C, Supplemental Figure 1, B–D, Supplemental Table 5). Three of the clusters were characterized by gene upregulation predominantly during the injury phase: "early injury" genes (UUO-2 only) were enriched for damage-associated molecular pattern–Toll-like receptor (TLR) signaling, MAPK signaling, and oxidative stress pathways (Figure 1C); "pan-injury" genes included cell cycle genes and markers of kidney injury (*Havcr1* and *Lcn2*); and "late injury" genes (UUO-7) were enriched for ECM components, ECM crosslinkers, and inhibitors of ECM degradation.

Two clusters were characterized by gene activation predominantly during R-UUO and were enriched for genes implicated in innate and adaptive immunity (Figure 1C). Remarkably, five of the top ten genes induced specifically after R-UUO (*Lyz1*, *Mmp12*, *Gpnmb*, *Ccl8*, and *Retnla*; Supplemental Figure 1, D and E) were also induced in macrophages in our model of resolution of liver fibrosis.[23] Multiple podocyte-specific genes were also included in the cluster upregulated during reversal, consistent with the loss of tubular mass and relative glomerular preservation in this model of predominantly tubular injury.

over 100,000 iterations. *$P<0.05$. (E) Violin plots of *Egf* and *Lcn2* (encodes neutrophil gelatinase-associated lipocalin) gene expression in the loop of Henle/distal convoluted cell cluster. The *y* axis shows the log-scale normalized read count. a–c, PCT subclusters colored by shared nearest neighbor; DCT, distal convoluted tubule; LoH, loop of Henle; Mac, macrophage; Mono, monocyte; NK, natural killer cell; PCT, proximal convoluted tubule; S1, S1 segment.
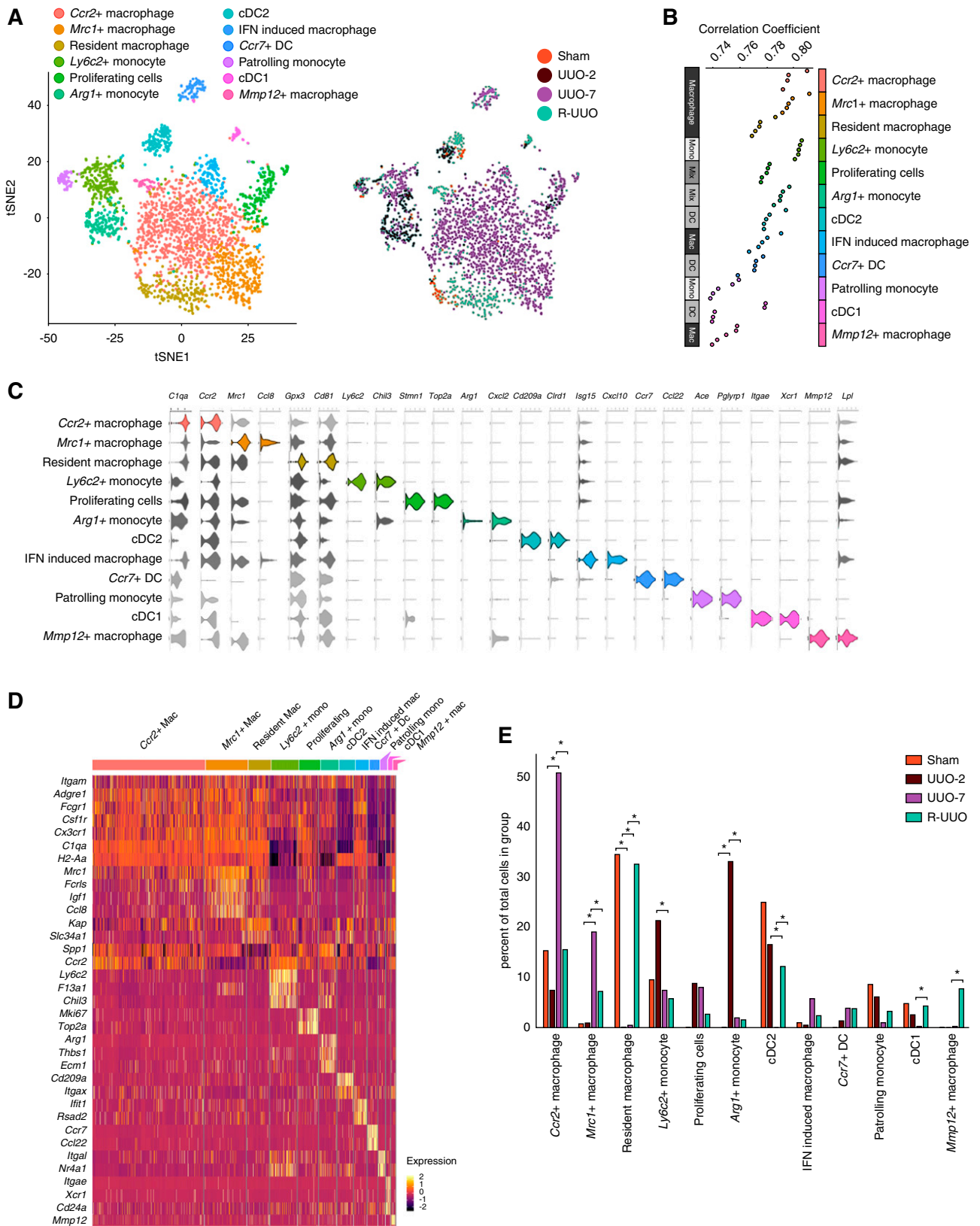
Figure 3. scRNA-seq analysis identifies 12 discrete myeloid cell clusters, with dynamic changes in the proportion of cells assigned to each cluster across the R-UUO model. (A) tSNE plot of 2956 pooled myeloid cells from each time point, annotated by cell type and time point. (B) Top five ImmGen reference immune cell types correlating with each cluster, ranked by Spearman correlation coefficients after a cluster-to-references analysis using Cluster Identify Predictor version 2. (C) Violin plots showing the expression levels of selected

The patterns of gene expression in the R-UUO model were consistent with those observed in a rat model of reversible diabetes and hypertension (Supplemental Figure 1, F and G),[53] suggesting common mechanisms of injury and repair across multiple models and organs,[23] with regression of fibrosis being characterized by the presence of a specific macrophage phenotype.

scRNA-seq Demonstrates Dynamic Changes in the Proportion of Intrinsic and Immune Cells during the R-UUO Model

To characterize the heterogeneity of cells during injury and repair and to ascribe the bulk-transcriptomic changes to specific cell types, we performed scRNA-seq on single-cell suspensions from the renal cortex of animals at four time points: baseline, UUO-2, UUO-7, and 2 weeks after R-UUO. Unsupervised clustering of the aggregated data from approximately 17,500 individual transcriptomes identified 15 discrete clusters (Figure 2, A and B, Supplemental Figure 2, A–D, Supplemental Table 6), which were classified using established cell-specific markers in murine kidneys (Figure 2C, Supplemental Figure 2B).[41] Because this is the first scRNA-seq dataset during renal injury and reversal of fibrosis, we have created an interactive tool for data exploration at http://www.ruuo-kidney-gene-atlas.com/.

Differential proportional analysis[44] determined that, after UUO, there was a marked reduction in the proportion of cells derived from the proximal tubule, which partly reversed after R-UUO (Figure 2D). Additionally, there were dynamic changes in the tubular transcriptome after UUO, with induction of injury markers such as *Lcn2* (adjusted $P=1.45 \times 10^{-293}$) and reduced expression of *Egf*, a biomarker of tubular cell health (adjusted $P=5.67 \times 10^{-05}$) (Figure 2E).[54,55] There was early recruitment of neutrophils and natural killer cells to the obstructed kidney, followed by expansion of macrophages and T cells, which persisted beyond the reversal of obstruction, as observed previously in the repair phase in a model of IRI (Figure 2D).[26] The changes in cell proportions were replicated on flow cytometry using markers of key cell types: proximal tubule cells (LTL), endothelial cells (CD31), fibroblasts/pericytes (PDGFR-β), immune cells (CD45), and F4/80^Hi and F4/80^Lo macrophages (Supplemental Figure 2E). Of note, although we detected a small number of cells expressing podocyte genes, these were too few to constitute a discrete cluster, as has been reported previously.[42] This likely reflects the difficulty in isolating single glomerular cells from whole kidney cortex, whereas extraction of single nuclei is more efficacious, with small-nuclear RNA-seq detecting 20-fold more podocytes.[42] Similarly, we

may not have captured other glomerular cells, such as endothelial or mesangial cells and monocytes or macrophages, that are unique to the glomerulus.

scRNA-seq Reveals Myeloid Heterogeneity during Injury and Repair

Our previous data in the kidney and liver suggest a pivotal role for the plasticity of myeloid cells in injury and repair.[23,53] To further characterize the myeloid cell heterogeneity and phenotype, we repeated the SNN clustering specifically on myeloid cells, partitioning these cells into 12 clusters (Figure 3A). We first assigned clusters as monocytes, macrophages, or DCs by generating Spearman correlation values for each cluster as compared with gene signatures of mouse immune cells obtained from ImmGen (Figure 3B, Supplemental Figure 3, A and C). We then refined this classification using cluster-defining DE genes (Supplemental Figure 3B, Supplemental Table 7) and the following genes encoding cell surface protein markers that define specific myeloid cell subsets on flow cytometry: *Itgam* (CD11b), *Adgre1* (F4/80), *Fcgr1* (CD64), *Itgax* (CD11c), and *H2-Aa* (MHCII) (Figure 3, C and D).

Early Accumulation of *Ly6c2*⁺ and *Arg1*⁺ Monocytes after Ureteric Obstruction

We identified the first of three monocyte clusters as patrolling monocytes because they expressed *Nr4a1* and *Itgal*, which are implicated in survival[56,57] and adhesion[56,58] of CX3CR1⁺/Ly6C⁻ patrolling monocytes. We classified the second cluster as inflammatory Ly6C⁺ monocytes because they expressed *Ly6c2*, *Ccr2*, *F13a1*, and *Chil3* genes.[24] Cells in this cluster increased at UUO-2 (Figures 3, A and E and 4A), indicating early recruitment of Ly6C⁺ inflammatory monocytes to the kidney during injury.

A third cluster uniquely expressed *Arg1* (Figure 3, C and D) and additionally markers of Ly6C⁺ inflammatory monocytes, including *Ccr2*, *Chil3*, and *F13a1*, although not *Ly6c2* (Figures 3D and 4B). DE genes in this cluster included early response genes (*Ier3*, *Fos*, *Jun*), hypoxia genes (*Hif1a*, *Vegfa*), proinflammatory genes (*Thbs1*, *Spp1*), profibrotic genes (*Tgfb1*, *Tgfbi*), and genes encoding ECM components (*Fn1*, *Ecm1*) or ECM crosslinkers (*Tgm2*; Figure 4B). These *Arg1*⁺ cells were exclusively present at UUO-2 (Figure 3, A and E), and the profibrotic gene expression suggested they may initiate fibrosis by interacting with mesenchymal cells. Accordingly, we determined expression of ligand-receptor pairs between each monocyte subset and mesenchymal cells (Figure 4C). More monocyte ligand–mesenchymal receptor pairs were expressed in *Ly6c2*⁺ and *Arg1*⁺ monocytes than patrolling monocytes. In addition, compared with the *Ly6c2*⁺ cells, the *Arg1*⁺

marker genes in each cluster. The x axis shows the log-scale normalized read count. (D) Heatmap of selected marker gene expression in each cluster, calculated using Wilcoxon signed-rank test. The color scheme is based on z-score distribution. (E) Relative proportions of each cell type at each time point. Statistical significance tested using differential proportional analysis with a mean error of 0.1 over 100,000 iterations. *$P<0.05$. Mac, macrophage; mono, monocyte; IFN, interferon.
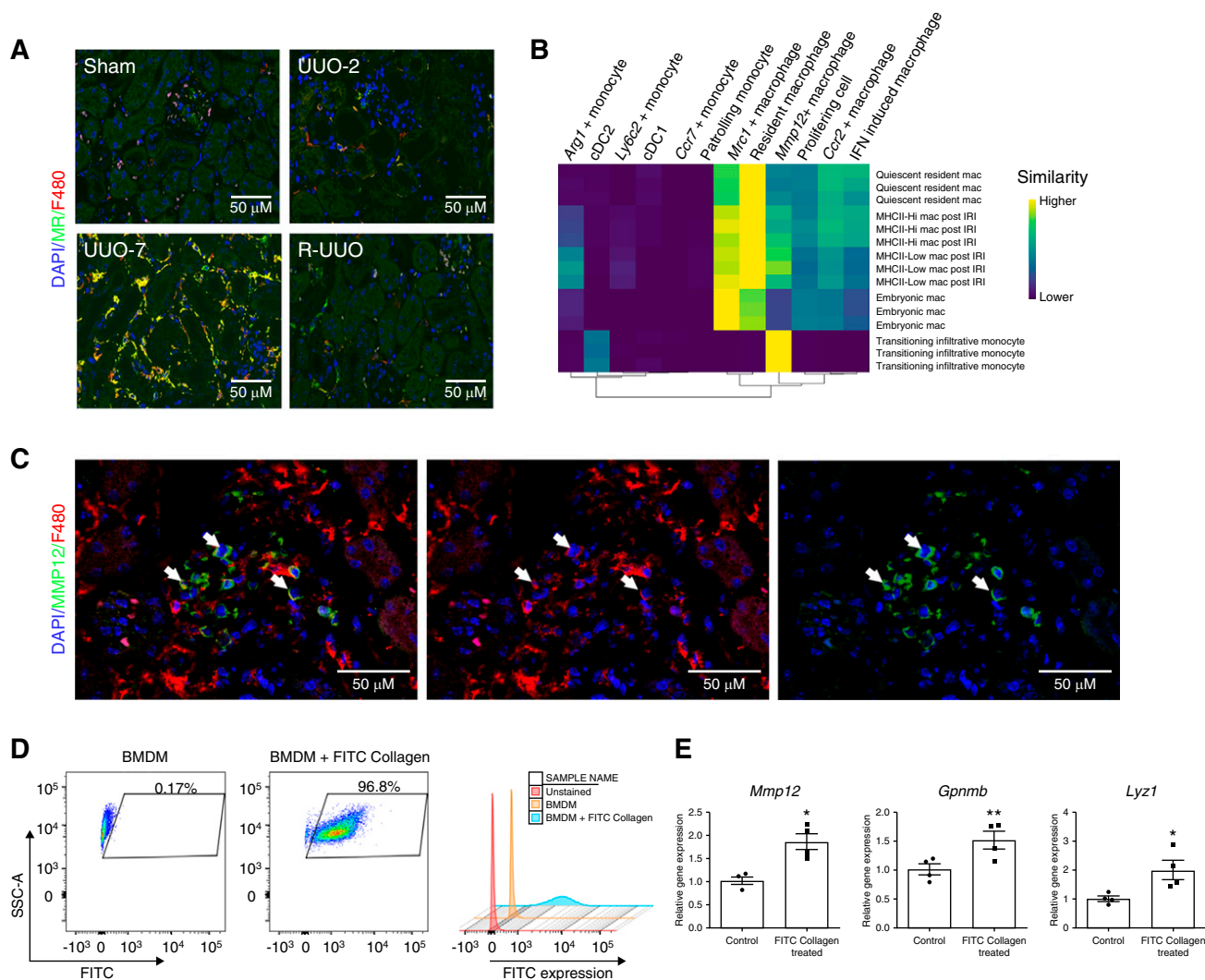
Figure 4. Monocytes recruited to the kidney at UUO-2 exhibit a pro-fibrotic phenotype that may activate mesenchymal cells. (A) Representative immunofluorescence images from each time point in the R-UUO model (sham, UUO-2, UUO-7, and R-UUO [2 weeks]) for Chil3 (marker of Ly6C$^+$ monocytes, red) and the pan-macrophage marker F4/80 (green). (B) Violin plots showing log$_{10}$ expression levels of selected genes in the three monocyte clusters. The *y* axis shows the log-scale normalized read count. (C) Significant monocyte ligand–mesenchymal receptor pairs across the three monocyte subclusters. Color of dot is proportional to mean expression values for all of the interacting partners, and size is inversely proportion to *P* value.

cells demonstrated greater potential for *Fn1-integrin*, *Pdgfa-Pdgfrß*, and *Tnf-Tnfsfr1* signaling to mesenchymal cells. Taken together, these data suggest *Arg1$^+$* cells may be derived from recruited Ly6C$^+$ monocytes that become activated acutely in the hypoxic and inflammatory milieu of the injured kidney toward a profibrotic phenotype.

Macrophages Adopt Differing Phenotypes during Injury and Resolution Phases

We identified five clusters (Figure 3B) that expressed genes consistent with macrophage identity, including those encoding CSF1 receptor (*Csf1r*), MHCII (*H2-Aa*), and *Cd81*, a conserved marker of renal macrophages (Figure 3, C and D).[24] We define a first cluster as quiescent resident macrophages,

detected exclusively in kidneys that underwent sham surgery or were de-obstructed (Figure 3, A and E). In contrast to other macrophage clusters, these cells did not express *Spp1*, which encodes osteopontin, a marker of activated macrophages[59] that promotes renal injury after UUO.[60] Although they were tagged by proximal tubular genes, the expression was much lower than in tubular cells (Figure 2C); therefore, this likely represents ambient tubular RNA, which is most prevalent in sham or R-UUO mice due to the higher proportion of proximal tubular cells (Figure 2D).

A second macrophage cluster expressed high levels of *Mrc1* (encodes mannose receptor), and cells from this cluster were most commonly observed at UUO-7 and, to a lesser extent, at R-UUO (Figures 3, A and E and 5A). The *Mrc1$^+$* cells

Figure 5. Mannose receptor (MR)$^+$ F4/80^{Hi} macrophages and MMP12$^+$F4/80^{Lo} cells are observed in late injury or specifically following reversal of obstruction, respectively. (A) Representative immunofluorescence images across the R-UUO time course (sham, UUO-2, UUO-7, and R-UUO [2 weeks]) for mannose receptor (MR, marker of *Mrc1*$^+$ macrophages, green) and F4/80 (red). (B) Single R analysis comparing the transcriptome of the myeloid clusters with those of embryonic macrophages, adult macrophages before and after renal IRI, and infiltrating monocytes during repair of renal IRI. (C) Immunofluorescence images showing colocalisation of MMP12 (green) and F4/80 (red) in renal macrophages. (D) Flow cytometry plots of bone marrow–derived macrophages (BMDMs) demonstrating fluorescence after phagocytosis of FITC-collagen. (E) Expression of reparative macrophage genes measured by quantitative RT-PCR in BMDMs after phagocytosis of FITC-collagen versus BMDMs cultured in medium alone (control). *n*=4 replicates. *$P<0.05$, **$P<0.01$. Mac, macrophage.

expressed multiple scavenger receptors (*Mrc1*, *Fcrls*, *Stab1*), suggesting a role in scavenging debris/excess ECM (Figure 3D). In addition, they expressed *Igf1*, which is upregulated in reparative macrophages in the liver[23] and promotes regression of cirrhosis and liver regeneration,[61] and *Apoe*, which dampens inflammation[62] and promotes regeneration.[63] The *Mrc1*$^+$ cells exhibited lower H2-Aa expression, indicating downregulation of MHCII. This is of interest because resident renal macrophages downregulate MHCII and adopt a phenotype consistent with embryonic macrophages during repair from acute ischaemia-reperfusion injury (IRI).[47] To assess this further, we compared the transcriptome of our macrophage clusters with those of embryonic kidney macrophages, resident renal macrophages before and after IRI, and infiltrative monocytes.[47] The transcriptome of the *Mrc1*$^+$ cells most closely aligned to embryonic macrophages, consistent with reprogramming toward a developmental phenotype to facilitate renal repair (Figure 5B).

A third macrophage cluster predominantly comprised cells from UUO-7 kidneys. Cells in this cluster expressed *Ccr2* (Figure 3D), raising the possibility they may be derived from Ly6C$^+$/CCR2$^+$ monocytes that are recruited to the injured kidney before transitioning to adopt a transcriptome highly similar to resident macrophages (Figures 3, C and D and 5B).
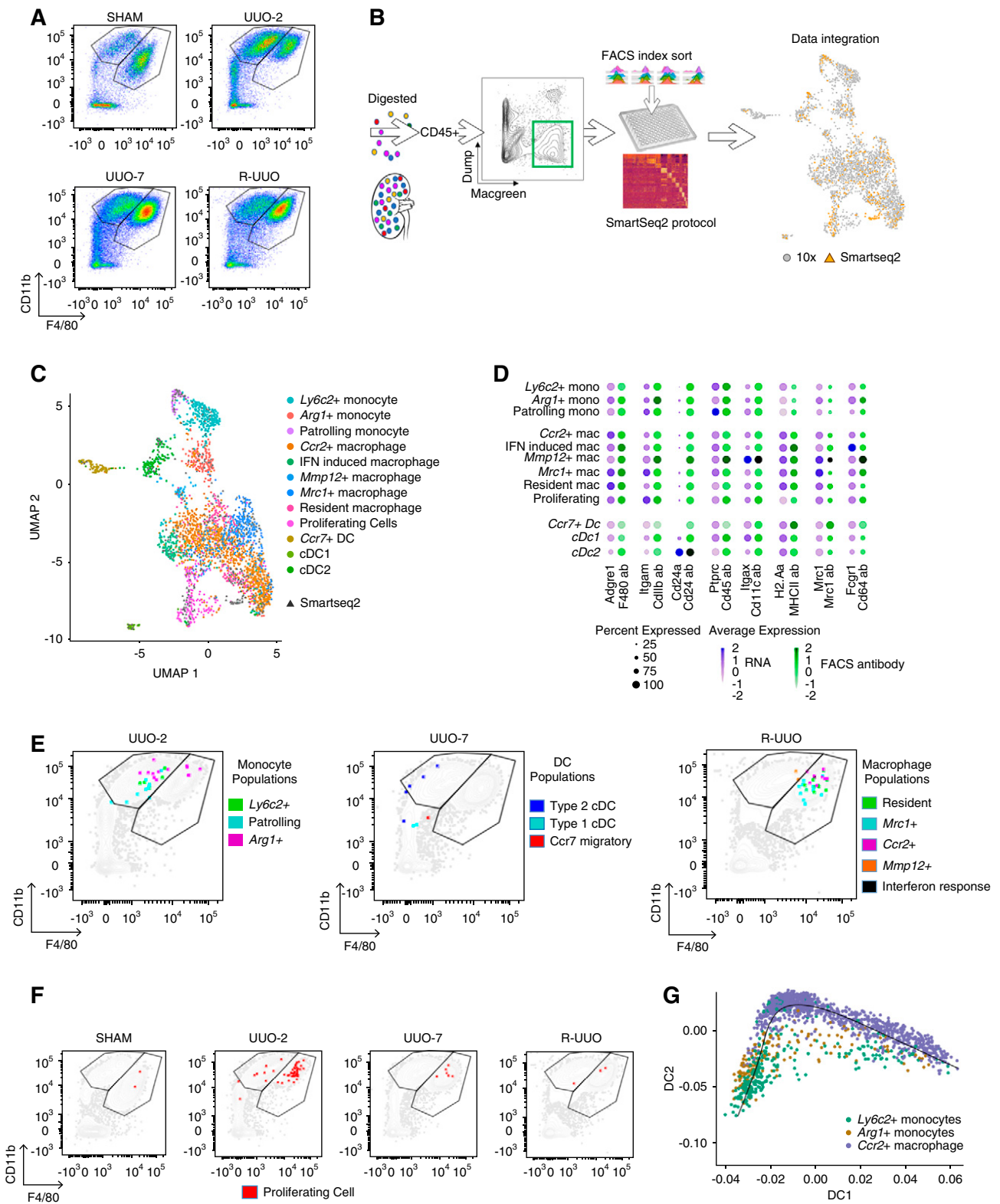
Figure 6. Integration of droplet– and plate-based scRNA-seq datasets determines that clusters identified by scRNA-seq would not be readily discriminated on conventional flow cytometry. (A) Representative flow cytometry plots from kidney cell suspensions from each time point after gating on CD45$^+$MacGreen$^+$TCRβ^-CD19$^-$Ly6G$^-$Siglec-F$^-$ myeloid cells. Cells segregated into CD11b$^+$F4/80^{Lo} monocyte and CD11b$^+$F4/80^{Hi} macrophage gates. (B) Strategy to integrate gene and cell surface protein expression at the single-cell level. Kidneys were digested into single-cell suspensions and single CD45$^+$MacGreen$^+$TCR1β^-CD19$^-$Ly6G$^-$Siglec-F$^-$ myeloid cells were sorted into individual wells after index linkage to cell surface marker expression. They underwent scRNA-seq using the SMART-

A fourth cluster was uniquely characterized by expression of IFN-stimulated genes (Figure 3, C and D). Although their function remains unknown, similar cells have been observed after injury in other organs, including the heart.[28]

Intriguingly, a final macrophage cluster was observed solely in kidneys that had undergone R-UUO. These cells were characterized by expression of *Mmp12*, a macrophage-specific metalloproteinase, and scavenger receptors (*Mrc1*, *Fcrls*), suggesting they may be involved in matrix remodeling (Figures 3, C and D and 5C). In addition, the *Mmp12*+ cluster–expressed genes implicated in efferocytosis and lipid transport, suggesting they may be involved in clearance of apoptotic cells (Supplemental Figure 3D). This is in keeping with our previous work, where *Mmp12*+ expression defines a reparative macrophage phenotype that mediates regression of liver fibrosis.[23] In response to phagocytosis of collagen (Figure 5D), bone marrow–derived macrophages upregulated the degradative enzymes expressed in reparative macrophages including *Mmp12*, *Gpnmb*, and *Lyz1*[23] (Figure 5E), suggesting they may switch to a matrix-degrading phenotype on encountering scarred matrix. Although the *Mmp12*+ cells mapped to macrophages on the ImmGen databases (Figure 3B), they morphologically resembled monocytes and expressed low levels of F4/80 (Figure 5C) but high levels of *Ccr2* (Figure 3C). Furthermore, their transcriptome most closely aligned to monocytes infiltrating the kidney during recovery from IRI (Figure 5B). Taken together, these data suggest they may be derived from monocytes that adopt a unique *Mmp12*+ phenotype on encountering a kidney during the reparative phase after injury.

DCs Adopt a Migratory Phenotype during Late-Stage Injury and Resolution

We assigned three clusters as DCs, expressing MHC genes (*H2-Aa*) but not macrophage markers (*Cd81*, *C1q*; Figure 3D). One cluster expressed *Itgae* (encodes CD103), whereas another expressed *Cd209a*, consistent with type 1 and type 2 conventional DCs (cDC1, cDC2), respectively.[24] cDC1s and cDC2s were proportionally fewer in UUO-7 before returning during R-UUO (Figure 3, A and E). In contrast, the third DC cluster, which expressed *Ccr7*, was not detected in sham animals, but appeared by UUO-7 and persisted through R-UUO (Figure 3, A and E). This cluster mapped specifically to lymph node DCs in the ImmGen database (Supplemental Figure 3A). Taken together, the data suggest that, after kidney injury, resident DCs upregulate *Ccr7* and this may

promote migration to draining lymph nodes by binding to CCR19/CCR21.[64]

Conventional Flow Cytometry Does Not Capture the Full Heterogeneity of Myeloid Cells

To assess how our scRNA-seq–derived clusters corresponded to conventional cytometry, we performed scRNA-seq using plate-based SMART-seq2 technology, which enabled linkage of each transcriptome to an abundance of cell surface markers (FACS intensity) using index sorting.[65] We repeated the R-UUO model using MacGreen mice (express EGFP under the Csf1r promoter) and performed flow cytometry, gating on CD45+MacGreen+TCRβ−CD19−Ly6G−Siglec-F− myeloid cells. There was expansion of the CD11b+F4/80Lo population at UUO-2, consistent with early recruitment of monocytes in response to injury (Figure 6A). By UUO-7, monocyte recruitment had diminished, but there was an increase in CD11b+F4/80Hi macrophages, which persisted through 2 weeks after R-UUO (Figure 6A).

We captured 192 individual CD45+MacGreen+TCR1β− CD19−Ly6G−Siglec-F− myeloid cells from each time point for scRNA-seq on the SMART-seq2 platform (Figure 6B). This dataset was then integrated with the original droplet dataset, with SMART-seq2 cells distributed through every cluster (Figure 6, B and C). Index linkage demonstrated moderate correlation between the gene and corresponding surface protein expression in each cluster (Figure 6D). Next, we mapped the cells from each myeloid cluster onto the monocyte and macrophage gates on flow cytometry (Figure 6E). Cells from the *Ly6c2*+ and patrolling monocyte clusters mapped to the CD11b+F4/80Lo monocyte gate, as expected. Conversely, resident, *Mrc1*+, *Ccr2*+, and IFN-response macrophages all mapped appropriately to the CD11b+F4/80Hi macrophage gate. Cells from the *Arg1*+ cluster straddled the monocyte and macrophage gates (Figure 6E), with a proportion of the *Arg1*+ cells colocating with *Ccr2*+ macrophages in the CD11bHiF4/80Hi region, suggesting they may be transitioning to *Ccr2*+ macrophages. Furthermore, trajectory analysis supported transition of *Ly6c2*+ monocytes to *Ccr2*+ macrophages, with *Arg1*+ monocytes representing an intermediate transitional state (Figure 6G). Cells from the *Mmp12*+ cluster also straddled the monocyte and macrophage gates (Figure 6E), consistent with their intermediate F4/80 expression on immunofluorescence (Figure 5B). Cells from the cDC1 and *Ccr7*+ clusters were CD11b−F4/80−, whereas the cDC2s mapped to the CD11b+F4/80Lo monocyte gate

---

seq2 protocol before integration with the 10× dataset. (C) Uniform Manifold Approximation and Projection (UMAP) of the combined 10× and SMART-seq2 dataset. (D) Dotplot of cell surface protein and corresponding gene expression in each cluster. The size of the dot denotes the percentage of cells in each cluster expressing the relevant gene/protein; the intensity of color represents mean gene/protein expression. (E) Representative flow cytometry plots from UUO-2, UUO-7, and R-UUO (2 weeks) illustrate mapping of cells from each myeloid cluster onto the CD11b+F4/80Lo monocyte and CD11b+F4/80Hi macrophage gates. (F) Mapping of proliferating cells (red) at each time point onto the flow cytometry plots. (G) Pseudotime analysis of the transcriptomes of the *Ly6c2*+, *Arg1*+, and *Ccr2*+ clusters. Mac, macrophage; mono, monocyte.
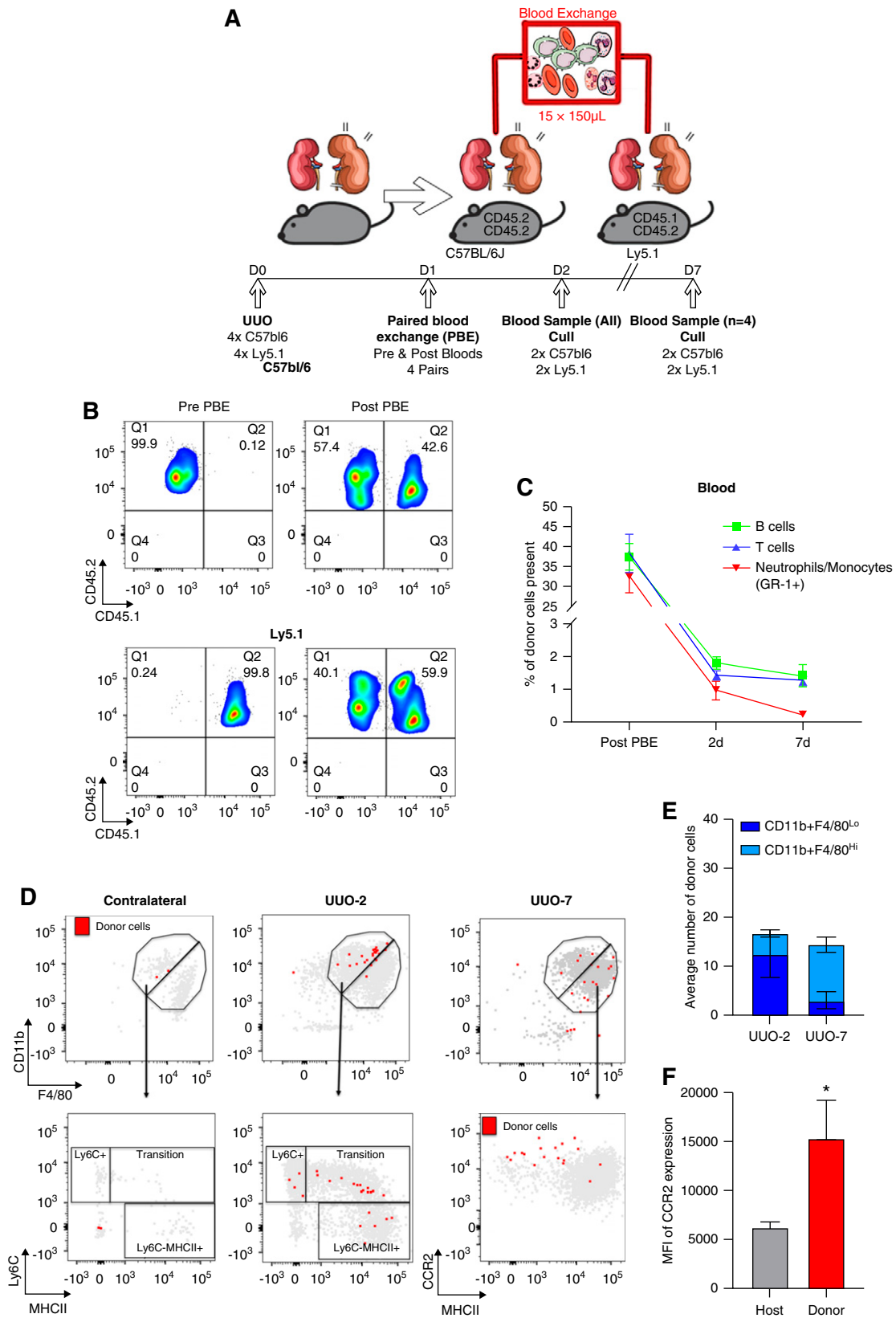
Figure 7. Paired blood exchange (PBE) demonstrates that CCR2$^+$ macrophages observed at UUO-7 are derived from donor monocytes. (A) Schemata of experimental strategy for PBE to track fate of immune cells recruited to the kidney. One day after UUO, whole blood exchange was performed between pairs of C57BL/6 and Ly5.1 mice ($n=4$) and pairs were euthanized either 2 or 7 days after UUO. (B) Representative flow cytometry plots of circulating CD45$^+$ cells from pairs of mice pre- and immediately post-blood exchange,

(Figure 6E). Cells from the proliferating myeloid cluster were observed predominantly at UUO-2 and mapped to the monocyte and, more particularly, the macrophage gates (Figure 6F). Consistent with this, the scRNA-seq cell cluster designated as proliferating cells (express *Mki67*) predominantly expressed macrophage markers, with representation of all macrophage clusters.

Monocytes Recruited Early after UUO Transition to a Macrophage Phenotype

Pseudotime trajectory analysis and flow cytometry data suggested that Ly6c$^+$ monocytes recruited to the obstructed kidney transitioned to *Ccr2*$^+$ macrophages by UUO-7. To test this hypothesis, we performed PBE between Ly5.1 mice (CD45.1/CD45.2 heterozygous) and C57BL6/J mice (CD45.2 homozygous) (Figure 7A). Immediately after the PBE, approximately 40% of total CD45$^+$ circulating cells were derived from the donor (Figure 7B). The proportion of donor-derived circulating monocytes/neutrophils fell rapidly to approximately 1% within 24 hours, with negligible numbers persisting through to UUO-7 (Figure 7C, full gating strategy in Supplemental Figure 4A). There was a similarly rapid reduction in the proportion of donor-derived circulating T and B lymphocytes; however, a small number of donor lymphocytes persisted in the circulation through UUO-7 (Figure 7C).

To determine the fate of donor monocytes recruited to the obstructed kidney, we performed flow cytometry on kidney cell suspensions, gating on CD45$^+$CD64$^+$TCRβ^-CD19$^-$Ly6G$^-$Siglec-F$^-$ myeloid cells (Supplemental Figure 4B). At UUO-2, donor cells were recruited preferentially to the obstructed kidney (Figure 7D) and mapped almost exclusively to the CD11b$^+$F4/80^{Lo} monocyte gate (Figure 7E). Indeed, they spanned the monocyte "waterfall," suggesting transition from Ly6C^{Hi}/MHC^{Lo} monocytes toward a Ly6C^{Lo}MHCII^{Hi} macrophage-like phenotype Figure 7D). By UUO-7, the majority of donor cells were located in the CD11b$^+$F4/80^{Hi} macrophage gate, and they expressed high levels of CCR2 compared with the global macrophage population (Figure 7F). In combination with the trajectory analysis and flow cytometry data, these data suggest donor monocytes are recruited selectively to the obstructed kidney by UUO-2, transition to a CCR2^{Hi} macrophage by UUO-7, and, hence, are the likely source of the cells in the *Ccr2*$^+$ macrophage cluster observed at UUO-7 in the scRNA-seq dataset (Figure 3A).

Myeloid Cell Subsets Correlate with Fibrosis in Human Kidney Disease

We next assessed whether similar myeloid cell phenotypes were observed in the human kidney using the Human Protein Atlas (Figure 8A). Cells that stained with F13A1 (marker of *Ly6c2*$^+$ monocytes) and DOK2 (*Arg1*$^+$ monocytes) were located specifically in focal areas of injury/inflammation, whereas ITGAL (patrolling monocytes) was largely restricted to cells within the circulation. CD68, a pan-macrophage marker (including a marker of resident macrophages), was widely distributed in the healthy kidney, whereas mannose receptor (*Mrc1*$^+$ macrophages) and CCR2 (*Ccr2*$^+$ macrophages) localized to areas of tissue injury. IRF8 and CD209, markers of types 1 and 2 conventional DCs, respectively, localized to areas of renal injury, with CCR7 (*Ccr7*$^+$ migratory DCs) staining a cluster of cells that resembled a tertiary lymphoid follicle. Furthermore, in biopsy specimens of healthy donors and patients with diabetic nephropathy or FSGS (https://www.nephroseq.org/resource/login.html), expression of each myeloid marker correlated with expression of *Col1a1* (encodes collagen I; Figure 8B). Expression of *Mmp12* was not detected in the healthy human kidney or in patients with CKD, which is consistent with the fact that cells in the *Mmp12*$^+$ cluster were specific to the resolution phase of kidney injury.

DISCUSSION

Our scRNA-seq studies, the first detailed characterization of myeloid cell heterogeneity in the kidney during progression and regression of fibrosis, have identified novel monocyte and macrophage subsets not previously observed in the kidney.

Acute injury induces a novel population of cells that are transcriptomically aligned to monocytes but that uniquely express *Arg1*. Although *Arg1* has traditionally been thought of as a marker of alternative macrophage activation,[67] the *Arg1*$^+$ cells do not express other markers of alternative activation, such as *Mrc1* or MHCII-encoding genes, suggesting that *in vitro* immune activation assays do not reflect the complex *in vivo* milieu. Indeed, the *Arg1*$^+$ cells express proinflammatory and profibrotic genes, and future work should determine whether specific depletion of these cells could reduce disease severity. Intriguingly, a novel *Mmp12*$^+$ macrophage subset emerged specifically during the resolution phase. We have
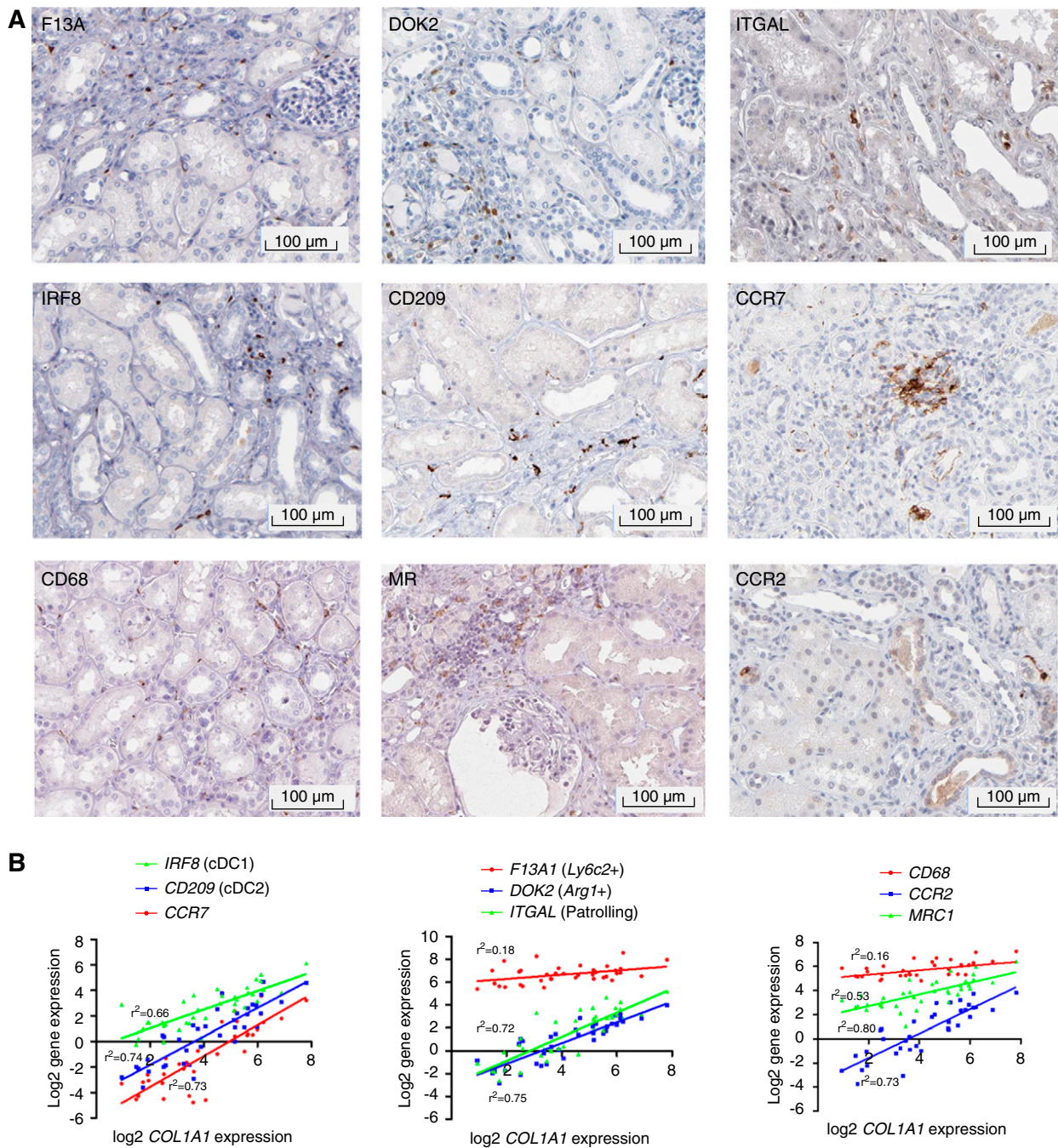
Figure 8. Markers of murine myeloid subsets are also observed in human kidney disease. (A) Immunostaining for myeloid cluster–defining markers in human kidney tissue obtained from the Human Protein Atlas.[66] (B) Gene expression of cluster-defining markers for DCs (left), monocytes (center), and macrophages (right) against COL1A1 gene expression in the tubulointerstitium of kidneys from healthy controls ($n=9$), patients with diabetic nephropathy ($n=10$), and patients with FSGS ($n=18$). Data obtained from www.Nephroseq.org.

previously reported a similar macrophage phenotype during the resolution of liver disease, suggesting common reparative mechanisms across organs.[23] Our *in vitro* studies suggest that ingestion of excess ECM or cell debris may be a stimulus for induction of this phenotype; however, their cellular origin and role in resolution requires further study because strategies to induce this phenotype may enhance scar degradation in the

diseased kidney and other organs. To definitively conclude that the $Mmp12^+$ macrophages we identified are essential to repair, as our studies infer, functional studies in $MMP12^{-/-}$ mice demonstrating a lack of repair would be required.

Uniquely, we have integrated plate-based and droplet scRNA-seq with index linkage to map our myeloid subsets onto monocyte and macrophage gates on flow cytometry.

Table 1. Cell surface markers for assignment of myeloid clusters by flow cytometry

Cluster	F4/80 (Adgre1)	CD11b (Itgam)	CD64 (Fcgr1)	MR (Mrc1)	CD40 (Cd40)	CCR2 (Ccr2)	Ly6C (Ly6c2)	CCR1 (Ccr1)	CD11a (Itgal)	CD203 (Itgae)	CD209 (Cd209a)	CCR7 (Ccr7)
Resident macs	+++	++	++	+	–	–	–	–	–	–	–	–
Mrc1+ macs	+++	+++	++	+++		–	–	+	–	–	–	–
IFN-response macs	+++	++	+++	+	++	–	–	–	–	–	–	–
Ccr2+ macs	++	++	++	+		++	–	–	–	–	–	–
Mmp12+ macs	+	+	–	++		+	–	–	–	–	–	–
Ly6c2+ monos	–	++	–	–		+++	+++	+	+	–	–	–
Arg1+ monos	–	+++	+	–		+	–	+++	–	–	–	–
Patrolling monos	–	+	–	–		–	–	–	+++	–	–	–
cDC1	–	–	–	+		–	–	–	–	+++	–	–
cDC2	–	++	–	–		–	–	+	–	–	+++	–
Ccr7+ DCs	–	–	–	–		–	–	–	–	–	–	+++

Gene expression (italicised in brackets) of corresponding cell surface protein markers that could be used to assign cells to myeloid clusters on flow cytometry. Macs, macrophages; monos, monocytes; IFN, interferon.

Importantly, the novel myeloid subsets would not have been identified by conventional markers on flow cytometry, highlighting the utility of scRNA-seq to characterize myeloid heterogeneity in an unbiased and more detailed way. Based on the DE genes in each subset, we suggest a panel of cell surface markers could be used in combination to differentiate the cells from each myeloid cluster on flow cytometry (Table 1), with the caveat that there is only a modest correlation between gene expression and cell surface marker expression of the corresponding protein.

To track the fate of circulating immune cells recruited to the kidney, we used PBE, which has previously been used to assess the effects of donor serum[35] but not, to our knowledge, to track immune cells. By combining PBE, flow cytometry, and pseudotime analyses, we demonstrate early recruitment of monocytes specifically to the obstructed kidney and that these subsequently adopt a macrophage phenotype, but continue to express CCR2. These results suggest circulating monocytes are the source of the large Ccr2+ macrophage cluster observed at UUO-7 in the scRNA-seq dataset, which is consistent with lineage tracing and parabiosis studies after myocardial infarction.[28] Remarkably, although they continue to express Ccr2, their transcriptome is otherwise almost identical to resident macrophages. Genetic or pharmacologic inactivation of CCR2 after renal IRI reduces the expansion of F4/80+ macrophages in the kidney and the severity of renal fibrosis, suggesting CCR2+ cells may be detrimental,[19] although it is unclear whether they remain detrimental once they transition to a macrophage phenotype. One advantage of PBE over parabiosis or bone marrow transfer is that the donor cells persist at large numbers in the circulation for a relatively short time, and, therefore, enable the tracking of cells at multiple discrete time points after injury or during resolution of disease. Future studies using PBE may help determine whether cells in the Mmp12+ cluster are indeed derived from infiltrating monocytes entering a repairing kidney. The short circulating time of donor cells may also be a limitation of the technique, in that only a small proportion of recruited cells are derived from the donor, therefore, a limited number of cells are available for downstream analysis. Further refinements including performing several consecutive PBEs may increase the yield.

In summary, by combining multiple complementary technologies, our studies have identified novel subsets of myeloid cells, which may also be present in human kidney disease and, hence, may represent therapeutic targets to inhibit progression and enhance resolution of kidney disease.

DISCLOSURES

All authors have nothing to disclose.

FUNDING

SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2020060806/-/DCSupplemental.

Supplemental Figure 1. Immunofluorescence and bulk RNA-seq analysis.

Supplemental Figure 2. Whole library analysis of single cell RNA-seq and cell proportions determined by flow cytometry.

Supplemental Figure 3. Analysis of myeloid transcriptomes.

Supplemental Figure 4. Flow cytometry gating strategy.

Supplemental Table 1. Antibodies utilized in immunohistochemistry/immunofluorescence studies.

Supplemental Table 2. Taqman probes used for qRT-PCR in studies.

Supplemental Table 3. Antibodies utilized in flow cytometry.

Supplemental Table 4. Summary of $10\times$ sequencing statistics

Supplemental Table 5. Bulk RNA-seq matrix of differentially expressed genes in each gene cluster, expressed in FKPM.

Supplemental Table 6. Single cell RNA-seq table of differentially expressed genes in the global data as described in Figure 2, expressed per cluster.

Supplemental Table 7. Single cell RNA-seq table of differentially expressed genes in the myeloid only data as described in Figure 3, expressed per cluster.

REFERENCES

1. Mills KT, Xu Y, Zhang W, Bundy JD, Chen CS, Kelly TN, et al.: A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010. *Kidney Int* 88: 950–957, 2015

2. Astor BC, Matsushita K, Gansevoort RT, van der Velde M, Woodward M, Levey AS, et al.: Chronic Kidney Disease Prognosis Consortium: Lower estimated glomerular filtration rate and higher albuminuria are associated with mortality and end-stage renal disease. A collaborative meta-analysis of kidney disease population cohorts. *Kidney Int* 79: 1331–1340, 2011

3. Chen TK, Knicely DH, Grams ME: Chronic kidney disease diagnosis and management: A review. *JAMA* 322: 1294–1304, 2019

4. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY: Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 351: 1296–1305, 2004

5. Hu B, Gadegbeku C, Lipkowitz MS, Rostand S, Lewis J, Wright JT, et al.: African-American Study of Kidney Disease and Hypertension Group: Kidney function can improve in patients with hypertensive CKD. *J Am Soc Nephrol* 23: 706–713, 2012

6. Perkins BA, Ficociello LH, Silva KH, Finkelstein DM, Warram JH, Krolewski AS: Regression of microalbuminuria in type 1 diabetes. *N Engl J Med* 348: 2285–2293, 2003

7. Ruggenenti P, Perticucci E, Cravedi P, Gambara V, Costantini M, Sharma SK, et al.: Role of remission clinics in the longitudinal treatment of CKD. *J Am Soc Nephrol* 19: 1213–1224, 2008

8. Mariani LH, Martini S, Barisoni L, Canetta PA, Troost JP, Hodgin JB, et al.: Interstitial fibrosis scored on whole-slide digital imaging of kidney biopsies is a predictor of outcome in proteinuric glomerulopathies. *Nephrol Dial Transplant* 33: 310–318, 2018

9. Fioretto P, Steffes MW, Sutherland DE, Goetz FC, Mauer M: Reversal of lesions of diabetic nephropathy after pancreas transplantation. *N Engl J Med* 339: 69–75, 1998

10. Fioretto P, Sutherland DE, Najafian B, Mauer M: Remodeling of renal interstitial and tubular lesions in pancreas transplant recipients. *Kidney Int* 69: 907–912, 2006

11. Duffield JS: Macrophages in kidney repair and regeneration. *J Am Soc Nephrol* 22: 199–201, 2011

12. Duffield JS, Tipping PG, Kipari T, Cailhier JF, Clay S, Lang R, et al.: Conditional ablation of macrophages halts progression of crescentic glomerulonephritis. *Am J Pathol* 167: 1207–1219, 2005

13. Lee S, Huen S, Nishio H, Nishio S, Lee HK, Choi BS, et al.: Distinct macrophage phenotypes contribute to kidney injury and repair. *J Am Soc Nephrol* 22: 317–326, 2011

14. Nelson PJ, Rees AJ, Griffin MD, Hughes J, Kurts C, Duffield J: The renal mononuclear phagocytic system. *J Am Soc Nephrol* 23: 194–203, 2012

15. Ricardo SD, van Goor H, Eddy AA: Macrophage diversity in renal injury and repair. *J Clin Invest* 118: 3522–3530, 2008

16. Lin SL, Castaño AP, Nowlin BT, Lupher ML Jr., Duffield JS: Bone marrow Ly6Chigh monocytes are selectively recruited to injured kidney and differentiate into functionally distinct populations. *J Immunol* 183: 6733–6743, 2009

17. Montgomery TA, Xu L, Mason S, Chinnadurai A, Lee CG, Elias JA, et al.: Breast regression protein-39/chitinase 3-like 1 promotes renal fibrosis after kidney injury via activation of myofibroblasts. *J Am Soc Nephrol* 28: 3218–3226, 2017

18. Chow FY, Nikolic-Paterson DJ, Ozols E, Atkins RC, Rollin BJ, Tesch GH: Monocyte chemoattractant protein-1 promotes the development of diabetic renal injury in streptozotocin-treated mice. *Kidney Int* 69: 73–80, 2006

19. Xu L, Sharkey D, Cantley LG: Tubular GM-CSF promotes late MCP-1/CCR2-mediated fibrosis and inflammation after ischemia/reperfusion injury. *J Am Soc Nephrol* 30: 1825–1840, 2019

20. Lin SL, Li B, Rao S, Yeo EJ, Hudson TE, Nowlin BT, et al.: Macrophage Wnt7b is critical for kidney repair and regeneration. *Proc Natl Acad Sci U S A* 107: 4194–4199, 2010

21. Wang Y, Chang J, Yao B, Niu A, Kelly E, Breeggemann MC, et al.: Proximal tubule-derived colony stimulating factor-1 mediates polarization of renal macrophages and dendritic cells, and recovery in acute kidney injury. *Kidney Int* 88: 1274–1282, 2015

22. Zhang MZ, Yao B, Yang S, Jiang L, Wang S, Fan X, et al.: CSF-1 signaling mediates recovery from acute kidney injury. *J Clin Invest* 122: 4519–4532, 2012

23. Ramachandran P, Pellicoro A, Vernon MA, Boulter L, Aucott RL, Ali A, et al.: Differential Ly-6C expression identifies the recruited macrophage phenotype, which orchestrates the regression of murine liver fibrosis. *Proc Natl Acad Sci U S A* 109: E3186–E3195, 2012

24. Zimmerman KA, Bentley MR, Lever JM, Li Z, Crossman DK, Song CJ, et al.: Single-cell RNA sequencing identifies candidate renal resident macrophage gene expression signatures across species. *J Am Soc Nephrol* 30: 767–781, 2019

25. Stewart BJ, Ferdinand JR, Young MD, Mitchell TJ, Loudon KW, Riding AM, et al.: Spatiotemporal immune zonation of the human kidney. *Science* 365: 1461–1466, 2019

26. do Valle Duraes F, Lafont A, Beibel M, Martin K, Darribat K, Cuttat R, et al.: Immune cell landscaping reveals a protective role for regulatory T cells during kidney injury and fibrosis. *JCI Insight* 5: e130651, 2020

27. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al.: Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 20: 163–172, 2019

28. Dick SA, Macklin JA, Nejat S, Momen A, Clemente-Casares X, Althagafi MG, et al.: Self-renewing resident cardiac macrophages limit adverse remodeling following myocardial infarction. *Nat Immunol* 20: 29–39, 2019

29. Fernandez DM, Rahman AH, Fernandez NF, Chudnovskiy A, Amir ED, Amadori L, et al.: Single-cell immune landscape of human atherosclerotic plaques. *Nat Med* 25: 1576–1588, 2019

30. Masuda T, Sankowski R, Staszewski O, Böttcher C, Amann L, Sagar , et al.: Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* 566: 388–392, 2019

31. Ramachandran P, Dobie R, Wilson-Kanamori JR, Dora EF, Henderson BEP, Luu NT, et al.: Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* 575: 512–518, 2019

32. Hesketh EE, Vernon MA, Ding P, Clay S, Borthwick G, Conway B, et al.: A murine model of irreversible and reversible unilateral ureteric obstruction. *J Vis Exp* [94]: 52559, 2014

33. Cochrane AL, Kett MM, Samuel CS, Campanale NV, Anderson WP, Hume DA, et al.: Renal structural and functional repair in a mouse model of reversal of ureteral obstruction. *J Am Soc Nephrol* 16: 3623–3630, 2005

34. Sasmono RT, Williams E: Generation and characterization of MacGreen mice, the Cfs1r-EGFP transgenic mice. *Methods Mol Biol* 844: 157–176, 2012

35. Rebo J, Mehdipour M, Gathwala R, Causey K, Liu Y, Conboy MJ, et al.: A single heterochronic blood exchange reveals rapid inhibition of multiple tissues by old blood. *Nat Commun* 7: 13363, 2016

36. Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R: Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9: 171–181, 2014

37. Kirschner K, Chandra T, Kiselev V, Flores-Santa Cruz D, Macaulay IC, Park HJ, et al.: Proliferation drives aging-related functional decline in a subpopulation of the hematopoietic stem cell compartment. *Cell Rep* 19: 1503–1511, 2017

38. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al.: STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21, 2013

39. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R: Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36: 411–420, 2018

40. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al.: Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* 17: 29, 2016

41. Park J, Shrestha R, Qiu C, Kondo A, Huang S, Werth M, et al.: Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* 360: 758–763, 2018

42. Wu H, Kirita Y, Donnelly EL, Humphreys BD: Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: Rare cell types and novel cell states revealed in fibrosis. *J Am Soc Nephrol* 30: 23–32, 2019

43. O'Flanagan CH, Campbell KR, Zhang AW, Kabeer F, Lim JLP, Biele J, et al.: CRUK IMAXT Grand Challenge Team: Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biol* 20: 210, 2019

44. Farbehi N, Patrick R, Dorison A, Xaymardan M, Janbandhu V, Wystub-Lis K, et al.: Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. *eLife* 8: e43882, 2019

45. O'Sullivan ED, Mylonas KJ, Hughes J, Ferenbach DA: Complementary roles for single-nucleus and single-cell RNA sequencing in kidney disease research. *J Am Soc Nephrol* 30: 712–713, 2019

46. Ekiz HA, Conley CJ, Stephens WZ, O'Connell RM: CIPR: a web-based R/shiny app and R package to annotate cell clusters in single cell RNA sequencing experiments. *BMC Bioinformatics* 21: 191, 2020 10.1186/s12859-020-3538-2

47. Lever JM, Hull TD, Boddu R, Pepin ME, Black LM, Adedoyin OO, et al.: Resident macrophages reprogram toward a developmental state after acute kidney injury. *JCI Insight* 4: 125503, 2019

48. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R: CellPhoneDB: Inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* 15: 1484–1506, 2020

49. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al.: Comprehensive integration of single-cell data. *Cell* 177: 1888–1902.e21, 2019

50. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al.: Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19: 477, 2018

51. Angerer P, Haghverdi L, Büttner M, Theis FJ, Marr C, Buettner F: Destiny: Diffusion maps for large-scale single-cell data in R. *Bioinformatics* 32: 1241–1243, 2016

52. Scrucca L, Fop M, Murphy TB, Raftery AE: Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R J* 8: 289–317, 2016

53. Conway BR, Betz B, Sheldrake TA, Manning JR, Dunbar DR, Dobyns A, et al.: Tight blood glycaemic and blood pressure control in experimental diabetic nephropathy reduces extracellular matrix production without regression of fibrosis. *Nephrology (Carlton)* 19: 802–813, 2014

54. Betz BB, Jenks SJ, Cronshaw AD, Lamont DJ, Cairns C, Manning JR, et al.: Urinary peptidomics in a rodent model of diabetic nephropathy highlights epidermal growth factor as a biomarker for renal deterioration in patients with type 2 diabetes. *Kidney Int* 89: 1125–1135, 2016

55. Ju W, Nair V, Smith S, Zhu L, Shedden K, Song PXK, et al.: ERCB, C-PROBE, NEPTUNE, and PKU-IgAN Consortium: Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci Transl Med* 7: 316ra193, 2015

56. Carlin LM, Stamatiades EG, Auffray C, Hanna RN, Glover L, Vizcay-Barrena G, et al.: Nr4a1-dependent Ly6C(low) monocytes monitor endothelial cells and orchestrate their disposal. *Cell* 153: 362–375, 2013

57. Hanna RN, Carlin LM, Hubbeling HG, Nackiewicz D, Green AM, Punt JA, et al.: The transcription factor NR4A1 (Nur77) controls bone marrow differentiation and the survival of Ly6C- monocytes. *Nat Immunol* 12: 778–785, 2011

58. Auffray C, Fogg D, Garfa M, Elain G, Join-Lambert O, Kayal S, et al.: Monitoring of blood vessels and tissues by a population of monocytes with patrolling behavior. *Science* 317: 666–670, 2007

59. Guo H, Cai CQ, Schroeder RA, Kuo PC: Osteopontin is a negative feedback regulator of nitric oxide synthesis in murine macrophages. *J Immunol* 166: 1079–1086, 2001

60. Ophascharoensuk V, Giachelli CM, Gordon K, Hughes J, Pichler R, Brown P, et al.: Obstructive uropathy in the mouse: Role of osteopontin in interstitial fibrosis and apoptosis. *Kidney Int* 56: 571–580, 1999

61. Sobrevals L, Rodriguez C, Romero-Trevejo JL, Gondi G, Monreal I, Pañeda A, et al.: Insulin-like growth factor I gene transfer to cirrhotic liver induces fibrolysis and reduces fibrogenesis leading to cirrhosis reversion in rats. *Hepatology* 51: 912–921, 2010

62. Yin C, Ackermann S, Ma Z, Mohanta SK, Zhang C, Li Y, et al.: ApoE attenuates unresolvable inflammation by complex formation with activated C1q. *Nat Med* 25: 496–506, 2019

63. Arnold L, Perrin H, de Chanville CB, Saclier M, Hermand P, Poupel L, et al.: CX3CR1 deficiency promotes muscle repair and regeneration by enhancing macrophage ApoE production. *Nat Commun* 6: 8972, 2015

64. Menetrier-Caux C, Montmain G, Dieu MC, Bain C, Favrot MC, Caux C, et al.: Inhibition of the differentiation of dendritic cells from CD34(+)

progenitors by tumor cells: Role of interleukin-6 and macrophage colony-stimulating factor. *Blood* 92: 4778–4791, 1998

65. See P, Lum J, Chen J, Ginhoux F: A single-cell sequencing guide for immunologists. *Front Immunol* 9: 2425, 2018

66. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al.: Proteomics. Tissue-based map of the human proteome. *Science* 347: 1260419, 2015

67. Gordon S: Alternative activation of macrophages. *Nat Rev Immunol* 3: 23–35, 2003