LONG RANGE GENE REGULATION IN HUMAN HEALTH AND DISEASE

Eliza Thulson

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill
in partial fulfillment of the requests for the degree of Doctor of Philosophy
in the Curriculum in Genetics and Molecular Biology.

Chapel Hill
2023

Approved by:

Douglas H. Phanstiel

Robert Duronio

Karen L. Mohlke

J. Mauro Calabrese

Hector L. Franco

ABSTRACT

Eliza Thulson: Long Range Gene Regulation in Human Health and Disease
(Under the direction of Douglas H. Phanstiel)

The human genome is capable of producing a vast number of phenotypically diverse cells, with incredibly unique roles that contribute to tissue- and developmental-specificity. As such, precise transcriptional control during biological processes such as differentiation, development, and response to environmental stimuli is required. A complex variety of regulatory elements are responsible for this regulation, many of which are still being characterized within the non-coding regions of the genome. In this work, I first investigate the function of the transcription factor Activator Protein 1 (AP-1) in loop-based gene regulation in a model of monocyte-to-macrophage differentiation. I utilized genome editing techniques to interrogate the role of AP-1 binding at Interleukin 1 beta (*IL1β*) enhancers, and preliminary results suggest a mechanism in which a DNA loop connects enhancer-bound AP-1 to *IL1β*, influencing gene expression. These data provide new insights into the mechanisms behind transcriptional control and 3D chromatin structure.

I next assay the impact of genetic risk variants on target genes in an *ex vivo* model of osteoarthritis (OA), in which human chondrocytes are treated with fibronectin fragment (FN-f). This model allows for the study of disease-associated variants in the correct cellular and biological context. We integrated hits from OA genome-wide association studies (GWAS), maps of 3D chromatin structure and enhancer activity in chondrocytes, and previously collected RNA-seq data from our OA model. This work revealed a set of putative causal OA variants and their potential target genes, including suppressor of cytokine signaling 2 (*SOCS2*). These results provide unique putative OA risk genes for further research and therapeutic development.

Finally, I describe my generation of high quality transcriptional and genotype data for use in expression quantitative trait locus (eQTL) analyses in an OA phenotype. These data will serve as the basis for QTL studies that assess both gene expression and chromatin accessibility. The overlap with OA

GWAS hits will contribute to the identification of novel putative target genes, risk variants, and their mechanisms.

To my husband, for following me across the country so I could pursue my dreams, and to my son, who came early but also right on time: I couldn't do this without either of you.

ACKNOWLEDGEMENTS

My work was supported by funding from the NIH National Institute of General Medicine Sciences T32 training grant (T32-GM007092) and the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2040435. I would also like to thank Jesse Raab and Thomas Vierbuchen for helpful guidance on CUT & RUN protocols and Samantha Pattenden for use of the Covaris LE220 instrument which was provided by the North Carolina Biotechnology Center Institute Development Program grant 2017-IDG-1005. I would also like to thank the Gift of Hope Organ and Tissue Donor Network, the donor families, and Dr. Susan Chubinskaya for providing normal donor tissue, as well as Dr. Pranav Mishra for donor tissue procurement and Mrs. Arnavaz Hakimiyan for technical assistance.

To the Phanstiel lab: you've all made this journey more than just work. I have looked forward to lab because you aren't just my coworkers, you're my family. Thank you all for every memory – this journey has been the absolute best. Katie, I will forever look up to you. You are the whole reason I rotated here, and I'm so grateful for your constant advice, leadership, compassion, and humor. I could go to you for anything and for that I will always love you. Erika, your talent is amazing, but more so your ability to communicate through art is what I have always admired. Eric, your constant coolness and calm sense of humor have inspired me many a time, and I'm grateful for all the time you've spent explaining code to me. Nicole, I'll never get tired of laughing with you. You always have something witty to say, and the way that you communicate science is something I hope to achieve. Isha, it was an honor to work with you and see you grow throughout your years at UNC. You were relentless in your drive to do science, and you never let any mistakes or obstacles stop you. You did so much for the lab, and I'm so proud to know that you're working toward achieving your dreams in school. Susan, I'm so glad that you ended up working with our lab. I've learned so much about experimental protocols and troubleshooting from you, and I strongly admire your ability to both do lab work and manage the lab. Andrea, I truly felt as if you were one of my mentors during your time in the lab, and I could not have gotten through some of those long protocols without you. CUT & RUN was our scientific brainchild and optimizing it – as tough and confusing as it was

– was made easier with you. Marielle, I'm not sure I've communicated this enough (or at all), but you are truly one of my scientific heroes. You have always had this amazing drive and dedication, and your ability to pick up a new experimental or bioinformatic technique and run with it is extraordinary. I have so many wonderful memories with you – Friday funnies, de-icing the freezer, commiserating over 5-day protocols and lab math. I am so glad you joined the lab, and I know that you will continue the culture because it wouldn't be the Phanstiel lab without you. Yoseli, from the very start of your postdoc, I have felt immensely comfortable with you. I know that I can come to you for lab stuff - to troubleshoot experiments, help me draft an email, read my abstract, listen to me practice a presentation... but I also know that I can come to you when I need an emotional and mental boost, because you are truly the kindest person I know. I have cried to you, vented to you, and complained to you, and you always have my back with some type of wisdom. I hope that I can approach my next adventure the way that you have handled your postdoc – with eagerness, humility, and an open mind.

I would also like to thank the members of my thesis committee: Drs. Robert Duronio, Karen Mohlke, Mauro J. Calabrese, and Hector Franco. I'm grateful to have had a thesis committee so dedicated to the expansion and challenge of my knowledge. Your advice and mentorship over the years has given me the confidence to move forward with faith in myself as a scientist.

To Doug: I don't think I could ever truly express how grateful I am for not only your mentorship, but your kindness and compassion. You've pushed me to become a better scientist in many ways – through the act of research itself, but also as a human being. With your help I was able to receive grants, present at conferences, perform experimental methods I didn't even know existed, and make important connections and friends in the science world. Every time I came to you with a problem or an obstacle in my path, you never made me feel less than or foolish. Instead, you guided me by asking me to look inward, think outside the box, and take a break when needed. I will forever be grateful for your words of wisdom in my first few years: "When science is working, stay late. When it's not, go home." I also can't thank you enough for your support during one of the hardest times in my life. After Noah was born and in the NICU, your flexibility and understanding gave me much-needed space to process and heal. I've always told others that I have a truly unique PI – you truly care about each one of your students and staff, and all future lab members will be incredibly lucky to have you as a mentor. Thank you.

To my other phenomenal professors along the way: Dr. Fenster (I will always fondly refer to you as “Dr. Fenster,” even though you insist I can call you Steve!), you saw me through the very beginning biology courses. I distinctly recall learning about dilutions and “1X” and “10X” factors in BIO113 lab and feeling ever so confused. Yet here I am, thanks to your patience and guidance. I truly admire that I felt you were a combination of a professor and a friend, because I learn best when I feel comfortable and supported. Your classes were so much fun – your enthusiasm for science is one of my biggest inspirations. Katie, I still remember when Bill told me about you, because he knew I’d probably enjoy doing biochemical research with you more than in his chemistry lab. I credit the start of my graduate school career to you, because without your constant fervor for research, I would have never gotten here. You gave me all the skills I needed to apply and interview with grad schools, and you instilled in me a discipline and dedication to everything I do in the lab. I look up to you as a female in STEM, a mentor, and a professor. Doing research with you was the beginning of the start of the rest of my life in science. Thank you for everything.

To the Thulsons, my second family: you all are my home away from home and I love you all so much. You accepted me into your hearts, call me daughter and sister, and look after me as family. I am so very lucky. Jeff, you are the kindest man I know. You are patient and gentle, and your love for your family knows no bounds. Thank you for raising the man of my dreams, and for loving Noah to the moon and back a thousand times. Donny, your straightforwardness and fearlessness give me courage to stand on my own in this world – as a wife, a mom, and a woman. You and Jeff trusted me with your son’s heart, and I am so unbelievably grateful for that. Being your second daughter is an honor like no other. Jetty, thank you for treating me like your sister. You have a level of cool-headedness and calmness that I greatly admire, and your positive outlook on life is something I know I can always rely on. Dillon – SAUCE – your sassiness is out of this world! And thank god for it, because it lightens up every mood. You and Jet are the most fun to hang out with, and I can’t wait to make some more (probably slightly weird) memories together.

To my family: Papa Dick and Grandma Kay, you have always been a supportive rock. From dancing together at family reunions to watching football and cooking out together, I have always had so much fun with you. Thank you for being so wonderfully reliable and encouraging. I love you! Papa Tom,

your constant championing of my educational and career goals has always been my backbone. I love that we can go from discussing the latest science topics to playing Cribbage or Gin Rummy all in one breath. You are my go-to for medical questions and science convos. I love you! Grandma Julie, I daresay you and mom are my best friends. Thank you for always being interested in whatever I'm doing – from science to Instagram to fashion – and for supporting me no matter what. I love your cooking and baking, I love trips to Whole Foods with you, I love our chats. I love YOU. To my aunts, uncles, cousins, and second cousins… I have the biggest, craziest, most wonderful family in the whole wide world! With you all as my support system, I can do anything. Thank you for the laughter, games, and adventures over the years. I love you all!

Mom and dad, where do I begin? From my first breath you have always believed in me. I credit my degrees to you because you never stopped pushing me to do and be my best. There is so much to thank you both for. Mom, you tutored me in geometry and algebra and calculus, helped me with piano and flute, carted me around to all the band events, read and edited my papers, and helped me with my ACTs and applying for college. Dad, you did all of that too, just leave out the math and sub in history or English! I am beyond lucky to have grown up in your household – your relationship taught me about love, your dedication to work taught me to have discipline, and your positive outlook on life taught me how to pick myself up and look forward. I love you both so much.

Gus, you may be my little brother, but I look up to you. You have always been so good at putting your head down and doing the work and having fun while doing it. I love hearing about your work, grabbing a beer with you, playing games (even though you usually kick my butt), and working out together. I'm so proud of everything you've accomplished, and I can't wait to see what comes next for you. I love you, bud!

To all my old and friends: you are why I manage to go on. Alex, growing up with you was the best thing that ever happened to me. We've been through so much together and at times our lives have converged and then taken separate paths, but I can always count on you. You've been my rock during some of the hardest times in my life, and I want you to know how much I appreciate you. Watching you begin a career, become a wife and a mother, start a new career – I'm in awe of you and I always have been. Thank you for being my best friend – I love you. Cassidy, you came into my life at the end of

college and though we both moved far apart, you've remained someone I can lean on, commiserate with, send pictures of Noah to, and discuss TMI things with. Thank you for all the times you've sent me encouraging messages and listened to me vent. Watching you become a mom to Jax has been an inspiration. I love you! Tori, I love that our story starts because of the boys, but continues because of us. You inspire me to stand up for myself, follow my dreams, and be unrelenting in the face of anxiety and fear. I'm so grateful for the strength of our friendship. I love you! To my graduate girlies, Emma and Claire: wow. Grad school would have been SO much harder without you. Emma, I have loved our deep discussions on religion, politics, and relationships. You have such a unique perspective on life, and our conversations always leave me with a new outlook. I admire your work ethic as a scientist, your dedication to your work and your family, and your compassion for others. Thank you for everything these past 6 years – I love you! Claire, your undying support at any time of the day is one of the pillars of my foundation. We've laughed, cried, vented, studied, shopped, partied, shared Wordle and Quordle… we do everything together. Thank you for being my accountability buddy, for talking to me every single day, for loving me and my baby. I love you!

Jon, you are everything I have ever needed in life. How can I thank you for all that you have done and continue to do for me? From day one you have been the most loyal person I know. We've had our rocky moments, for sure, but never once have I doubted the strength of our love. I know that wherever we go, whatever we do – all we need is to be together. You've always supported me, and for that I owe you everything. Thank you for leaving your hometown and following me across the country. Thank you for holding me when I cried and listening to me vent over failed experiments. Thank you for making me coffee and breakfast on days when I didn't know if I had time to eat. Thank you for taking care of the dogs and the cats, for keeping the house clean, for folding and putting away my clothes. Thank you for giving me our child, thank you for being my rock while he was in the NICU, and thank you for being a wonderful dad. I can't wait to see what our future holds, because I know it is going to be so, so good. I love you.

Noah, baby, I had to save you for last because you're first in my heart (don't worry, your dad's a close second). You are the light of my life, the apple of my eye, the icing on my cake. The best part of everything. You came and turned my world upside down, and I'm so grateful because you make me a better person. I never expected the beginning of our journey, but we are stronger for it. With you in it, the

world is brighter. Raising you will be the best thing I ever do. I've already said it, but everything I have ever done and will do is for you. I love you, baby boy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3C	Chromatin conformation capture
3D	Three-dimensional
4C	Chromosome conformation capture-on-Chip
5C	Chromosome conformation capture carbon copy
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
bp	Base pair
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
ChIA-PET	Chromatin interaction analysis by paired-end tag sequencing
ChIP	Chromatin immune precipitation
ChIP-seq	Chromatin immunoprecipitation followed by deep sequencing
dCas9	Catalytically dead Cas9
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DSB	Double-stranded break
eRNA	Enhancer RNA
FISH	Fluorescence *in situ* hybridization
FN-f	Fibronectin fragment
GWAS	Genome-wide association studies
H3K27ac	Histone H3 lysine 27 acetylation
H3K27me3	Histone H3 lysine 27 trimethylation
H3K4me3	Histone H3 lysine trimethylation
H3K9me3	Histone H3 lysine 9 trimethylation
Hi-ChIP	Hi-C followed by chromatin immunoprecipitation
kb	Kilobase
LCR	Locus control region
MS	Mass spectrometry

nCas9	Nickase Cas9
OA	Osteoarthritis
PCA	Principal component analysis
PLAC-seq	Proximity ligation-assisted ChIP-seq
PMA	Phorbol myristate acetate
PTM	Post-translational modification
RNA	Ribonucleic acid
TF	Transcription factor
ZRS	Zone of polarizing activity regulatory sequence

CHAPTER 1: INTRODUCTION

1.1 Gene regulation and cellular diversity

The human genome serves as the blueprint for approximately 200 different types of cells. The functional diversity of these cells relies on the precise control of various complex biological processes. Mis- or dysregulation of these processes can contribute to the onset or progression of diseases, including cancer, auto-immune disorders, and developmental abnormalities (Gregersen and Olsson 2009; Pomerantz and Freedman 2011; Vorstman and Ophoff 2013; The Deciphering Developmental Disorders Study 2015; David *et al.* 2018; Radford and Firth 2019; Bosch and Dalla-Favera 2019).

The question of how one genome could produce such a complex diversity of cellular phenotypes was further complicated by the discovery in the early 2000s that protein-coding genes only make up roughly 2% of genetic material (Carninci *et al.* 2005; ENCODE Project Consortium 2012). The first draft sequence of the human genome in 2001 revealed an estimate of 31,000 genes (Lander *et al.* 2001) and was later revised to 26,588 (Venter *et al.* 2001). This number was further reduced to 24,000 in 2004 following sequencing completion (International Human Genome Sequencing Consortium 2004) and today sits around 20,000 (Salzberg 2018). Of the 3 plus billion base pairs in the human genome, approximately 98% is non-coding sequence. How such a diverse phenotypic array could arise from one genome with such little coding sequence remains a mystery with much to discover.

The concept of phenotypic diversity arising from more than genetic sequence alone has long been demonstrated (King and Wilson 1975). Although originally the non-coding portion of the genome was termed ‘junk DNA,’ it is now widely understood that these regions contain regulatory elements that contribute to the control of gene expression. These elements make up a second layer of information beyond DNA sequence, known as the epigenome. The epigenome is comprised of chemical changes to both DNA bases and histone proteins, controlling chromatin accessibility and affecting gene expression by recruiting or blocking transcription factors (Bannister and Kouzarides 2011).

Post-translational modification (PTM) of histones in the form of acetylation and methylation was discovered in the 1960s (Allfrey *et al.* 1964), and some 30 years later the high-resolution X-ray structure of the nucleosome was solved (Luger *et al.* 1997). Together these breakthroughs gave insight into how PTMs affect chromatin structure and activity (Wolffe 1998). Many histone marks have since been termed canonical, including histone H3 lysine 27 acetylation (H3K27ac, associated with transcriptional activation), H3 lysine 4 trimethylation (H3K4me3, associated with active euchromatin), H3 lysine 27 trimethylation (H3K27me3, associated with transcriptional silencing), and H3 lysine 9 trimethylation (H3K9me3, associated with transcriptional repression), to name a few (Lawrence *et al.* 2016). Histone modifications work in two ways – the first by directly affecting chromatin structure. For example, histone acetylation reduces the affinity between histones and DNA, disrupting these contacts and subsequently unfolding chromatin. This provides transcription factors and other enzymes with access to open DNA (Kouzarides 2007; Bannister and Kouzarides 2011). The second method indirectly affects chromatin structure by modulating the binding of effector molecules such as transcription factors and other chromatin factors. Histone PTM binding proteins utilize various binding sites like plant homeodomains, bromodomains, and chromodomains that function to add or remove additional PTMs or remodel nucleosomes (Sanchez and Zhou 2009, 2011; Yap and Zhou 2011).

The epigenome that mainly exists in non-coding regions of the genome provides significant context into the production of incredible phenotypic diversity from one blueprint. Specific biological conditions are required for these epigenetic factors to do their work, and as such we see distinct histone modification profiles that are correlated with the transcriptional differences in various cell types (Koch *et al.* 2007; Gibney and Nolan 2010). Mapping of the epigenome under specific biological conditions has provided vast annotations for cis-regulatory elements and remains an ongoing effort (Hawkins *et al.* 2010; Wamstad *et al.* 2012; Rada-Iglesias *et al.* 2012; Zhu *et al.* 2013; Gifford *et al.* 2013; Xie *et al.* 2013).

1.2 Discovery of enhancers

The discovery of enhancers in 1981 (Moreau *et al.* 1981; Banerji *et al.* 1981) prompted an entirely new perspective on the capabilities of the human genome and later laid the foundation for the interrogation of the non-coding genome. Prior to this, it had been understood that gene expression was regulated by proteins and that open areas of chromatin sensitive to digestion by deoxyribonuclease

(DNase) I may play a role (Jacob and Monod 1961; Weintraub and Groudine 1976). Tissue-specific regions of DNA sensitive to DNase I significant distances from target genes had also been identified (Stalder *et al.* 1980), but it had yet to be shown that these sequences actually affected gene expression. This was finally demonstrated in 1981 when a non-coding region of the simian virus 40 (SV40) was shown to increase expression of a target gene expressing T-antigen *in vitro*. The key characteristics noted – and these still hold true today – were that the non-coding region affected gene expression independent of the orientation of the sequence and at a remote distance from the promoter (Moreau *et al.* 1981). Close in time to this discovery, the flexibility of the SV40 non-coding region was also shown to influence β-globin gene expression in the same manner (Banerji *et al.* 1981). Soon after, several *in vitro* studies reported the discovery of a tissue-specific, mammalian enhancer at the immunoglobulin heavy chain locus (Banerji *et al.* 1983; Gillies *et al.* 1983; Mercola *et al.* 1983), and finally enhancers were shown to work *in vivo* in a cell-type specific manner by increasing the expression of the large T-antigen exclusively in pancreatic beta cells (Hanahan 1985).

Shortly after the enhancer was discovered, it became clear that mapping these elements would take tremendous effort, as the number of non-coding regions far exceeded the number of genes. Thus, projects such as the NIH Roadmap Epigenome Consortium and the Encyclopedia of DNA Elements (ENCODE) were born, aimed at characterizing the epigenome on a genome-wide scale across many cell types. Today, approximately one million candidate regulatory elements have been catalogued with enhancer-like signatures, spanning ~7.9% of the mappable human genome (ENCODE Project Consortium *et al.* 2020).

We now know that enhancers are critical to the proper execution of many biological processes and contribute to the control of cell-type specificity (Bulger and Groudine 2010). Importantly, it is recognized that many enhancers contain disease-associated variants (Maurano *et al.* 2012), evidence of which is further discussed in section 1.5. The locations of enhancers can be predicted by many experimental techniques, such as the interrogation of transcription factor binding by chromatin immunoprecipitation followed by deep sequencing (ChIP-seq), histone post-translational modifications by ChIP-seq, open chromatin states by assay for transposase-accessible chromatin using sequencing (ATAC-seq), and three-dimensional (3D) chromatin structure by chromatin conformation capture (3C)

assays, to name a few. Together, these methods have identified relatively common enhancer characteristics.

Most enhancers are 100-1000bp in length, and are evolutionarily conserved in both sequence and function (Visel *et al.* 2009). Bound by cohorts of transcription factors, co-regulators, and RNA polymerase II, enhancers are also flanked by histones with specific chromatin marks that demarcate accessibility and activity. These marks commonly include H3K4me and H3K27ac (Heintzman *et al.* 2009) as well as acetylation at various lysines on H3 and H4 (Wang *et al.* 2008; Pradeepa *et al.* 2016). Interestingly, enhancers are actively transcribed, the product of which is termed enhancer RNAs (eRNAs) (De Santa *et al.* 2010; Kim *et al.* 2010). The function of these eRNAs is widely debated, but increasing evidence implicates a role for them in transcriptional regulation (Kim *et al.* 2010; Kaikkonen *et al.* 2013; Arner *et al.* 2015; Cheng *et al.* 2015; Ren *et al.* 2017). Enhancers also contain a wide variety of transcription factor (TF) motifs, binding of which by TFs is context-, cell-, and tissue-specific (Spitz and Furlong 2012).

1.3 Early theories of enhancer-promoter communication

Following the discovery of the enhancer element, the obvious question arose: how do these elements communicate with their target genes? It was well known that enhancers could exert their influence regardless of orientation and over a great distance, even from within the introns of other genes (Moreau *et al.* 1981; Banerji *et al.* 1981, 1983; Gillies *et al.* 1983). Several early theories proposed various models for this phenomenon, the most popular being scanning, linking, and looping (Ptashne 1986; Wang and Giaever 1988; Herendeen *et al.* 1992; Blackwood and Kadonaga 1998; Dorsett 1999; Bulger and Groudine 1999; Engel and Tanimoto 2000; Hatzis and Talianidis 2002).

The scanning model, also known as sliding or tracking, suggested that transcriptional machinery is recruited to enhancers, and subsequently translocates along the DNA until it reaches a promoter, whereupon transcription is initiated (Moreau *et al.* 1981; Blackwood and Kadonaga 1998; Hatzis and Talianidis 2002). Originally proposed in 1981 as a potential mechanism for which the SV40 enhancer could affect transcription (Moreau *et al.* 1981), subsequent studies provided evidence for this model. The bacteriophage T4 late genes were shown to be activated by a remote enhancer, and that transcription of these genes was effectively blocked by the DNA-bound restriction endonuclease EcoRI (Herendeen *et al.*

1992). This model was later exhibited in a eukaryotic system, in which the assembly of transcription factors at the hepatocyte nuclear factor-4α (*HNF-4α*) enhancer was analyzed. Signals representing these factors were found at intervening regions between the enhancer element and the *HNF-4α* gene prior to gene activation (Hatzis and Talianidis 2002).

The linking model – also known as oozing – proposed that the formation of a protein bridge, or scaffold, between the enhancer and promoter allowed for the transmission of transcription signals via RNA polymerase II (Bulger and Groudine 1999). In 2002, Mahmoudi et al. investigated the characteristics and capabilities of the transcription factor GAGA, which contains both DNA binding and protein binding domains. GAGA was shown to form a protein 'link' or bridge between separate DNA fragments as well as promote enhancer function, both *in cis* and *in trans* (Mahmoudi *et al.* 2002).

Neither the scanning nor the linking model explains evidence of physical interaction between an enhancer and a promoter element (Dillon *et al.* 1997; Tahirov *et al.* 2002; Shang *et al.* 2002; Carter *et al.* 2002); however, the looping model does. This model postulates the contact in physical space between an enhancer and a promoter, with intervening DNA looping out. Electron microscopy experiments in 1986 provided some of the first evidence of looping in prokaryotes, visually demonstrating repressor proteins bound at specific locations on DNA, forming a loop (Ptashne 1986). Following presentation of evidence for looping in eukaryotes and *in vitro* (Su *et al.* 1991; Dorsett 1999), *in vivo* support of the looping model was presented in 2002 (Tolhuis *et al.* 2002). This example of looping at the β-globin locus in mice is perhaps one of the most foundational looping studies, serving as the basis for further expansion of the field of 3D chromatin architecture. Utilizing newly developed 3C methods (Dekker *et al.* 2002), the well-known β-globin genes regulatory element, known as the locus control region (LCR), was shown to contact the β-globin genes in physical space in a developmental- and tissue-specific manner. Physical contacts were observed between the LCR and active globin genes in fetal liver cells, with the intervening DNA containing inactive embryonic globin genes looping out. Importantly, loops were not detected in non-expressing murine brain cells (Tolhuis *et al.* 2002).

Forced looping experiments further corroborated the looping model as the mechanism most likely responsible for enhancer-promoter communication (Deng *et al.* 2012, 2014). However, the question of

exactly how loop formation occurs remained. As mentioned earlier, 3C-derived methods have shed significant light on the mechanisms underpinning 3D genomic architecture.

1.4 Chromosome conformation capture (3C) methods

Early studies of the conformation of the genome used cytological technologies, such as fluorescence *in situ* hybridization (FISH), which allows for detection of the location of specific sequences by using fluorescent probes complementary to the region of interest (Schwarzacher *et al.* 1989; Shakoori 2017). FISH has revealed many features of chromatin architecture, such as chromatin territories and chromatin domains (Edelmann *et al.* 2001; Cremer and Cremer 2006; Brown *et al.* 2008; Müller *et al.* 2010). However, FISH is both relatively low-throughput and low resolution, and only allows for visualization at probe-specific regions. Thus, the advent of 3C techniques and its derivates have offered more powerful tools for spatial reconstruction at local and genome-wide views.

3C was developed in 2002 by Dekker et al., and its derivatives soon followed as the need to generate higher and higher-throughput methods increased. All 3C-based methods begin with a similar series of steps, in which chromatin is crosslinked (typically with formaldehyde), digested with restriction endonucleases to fragment the genome, ligated under dilute conditions to favor intra-molecular ligation over inter-molecular ligation, and reverse-crosslinked. The resulting products are chimeric DNA fragments that represent physical 3D contacts within the genome (Dekker *et al.* 2002).

The original 3C method uses semi-quantitative PCR to quantify interactions between a single pair of genomic loci. Because of this, it is often referred to as a ‘one-vs-one’ method. 3C requires prior knowledge of the desired interacting regions so as to design the PCR primers, and can only detect contacts within a few hundred kb (Dekker *et al.* 2002; Simonis *et al.* 2007). Further development of the 3C method resulted in chromosome conformation capture-on-Chip (4C) – known as ‘one vs all.’ 4C combines 3C with microarray or next generation sequencing to assess all of the chromatin interactions at one genomic locus. In this method, 3C DNA templates are cleaved with a second restriction enzyme and re-ligated to create small DNA circles. Interacting fragments are amplified using inverse PCR with bait-specific primers and then evaluated on a microarray or sequencer (Simonis *et al.* 2006; Zhao *et al.* 2006; van de Werken *et al.* 2012). Similarly to 3C, chromosome conformation capture carbon copy (5C) detects interactions within a given region, but in a ‘many vs many’ approach (Dostie *et al.* 2006; Dostie and

Dekker 2007). 5C allows for the detection of millions of interactions via the use of a universal sequence appended to the 5′ ends of the primers, multiplex PCR amplification, and deep sequencing (Dostie *et al.* 2006; Dostie and Dekker 2007).

Hi-C is likely the most commonly utilized 3C-based technique as it provides an 'all vs all' approach for interrogating contacts genome-wide (Lieberman-Aiden *et al.* 2009; van Berkum *et al.* 2010). As in the other 3C methods, the first step is to generate contact segments, but then the ends are filled in with biotin prior to ligation. The fragments are sheared and pulled-down to ensure that only biotinylated junctions are selected for. This results in a sample primarily comprised of hybrid strands that represent 3D physical contacts between two regions of DNA. Without the requirement of a targeted locus or region to focus on, Hi-C provides an unbiased approach for investigating chromatin structure (Lieberman-Aiden *et al.* 2009; van Berkum *et al.* 2010).

Other methods that are based on 3C that investigate chromatin conformation in the context of protein binding include chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (Fullwood *et al.* 2009; Li *et al.* 2010, 2017), Hi-C followed by chromatin immunoprecipitation (Hi-ChIP) (Mumbach *et al.* 2016), and proximity ligation-assisted ChIP-seq (PLAC-seq) (Fang *et al.* 2016). Additionally, sequence capture-based methods have been designed to enrich 3C and Hi-C libraries for specific loci of interest, such as Capture-C (Hughes *et al.* 2014; Davies *et al.* 2016), Capture Hi-C (Mifsud *et al.* 2015), and, more recently, hybrid capture Hi-C (Hi-C^2) (Sanborn *et al.* 2015). There are many other techniques based on the original 3C methods, the development of which allow us to address novel research questions and piece together a deeper understanding of the human genome. These approaches are being used not only to map the organization of the genome, but also to understand the role of its topology in the regulation of gene expression, protein function, and even disease progression.

1.5 Enhancers in disorders and disease

While mutations in protein-coding sequences oftentimes have clear roles in disease, the consequences of misregulation of regulatory elements such as enhancers has been more challenging to elucidate. However, enhancers are increasingly implicated in pathological phenotypes, especially with the advent of genomic techniques such as Hi-C. The non-coding genome offers several regulatory layers, disruption of which may directly or indirectly impact transcription and subsequent protein function.

The sonic hedgehog (SHH) locus is a canonical example in which enhancer mutations contribute to a developmental abnormality. SHH is a key signaling molecule responsible for regulating mammalian embryonic morphogenesis. A regulatory element approximately 770 bp in length and 800-1000 kb away from the *SHH* promoter, termed the zone of polarizing activity regulatory sequence (ZRS), has been shown to regulate expression of *SHH* during limb bud development (Lettice *et al.* 2002, 2003, 2014; Sagai *et al.* 2005). Importantly, point mutations and sequence duplications in the ZRS result in various skeletal defects of the limb and are referred to as "ZRS-associated syndromes." These syndromes include preaxial polydactyly type 2, triphalangeal thumb polysyndactyly, syndactyly type 4, and Werner mesomelic syndrome (Klopocki *et al.* 2008; Furniss *et al.* 2008; Sun *et al.* 2008; Wu *et al.* 2009; Li *et al.* 2009; Wieczorek *et al.* 2010; Anderson *et al.* 2012). Recent studies have identified specific variants within the ZRS that are associated with polydactyly and syndactyly (Xu *et al.* 2020; Kvon *et al.* 2020; Shen *et al.* 2022; Zeng *et al.* 2022). Interestingly, a 2021 study from Ushiki et al. identified CTCF and RAD21 deletions within the ZRS associated with archeiropodia, or congenital limb truncation (Ushiki *et al.* 2021). Taken together, these data show that expression of *SHH* is tightly regulated by its upstream regulatory element, the ZRS, and that various forms of mutations within the ZRS are responsible for a wide variety of developmental abnormalities. Furthermore, these data imply that the ZRS interacts with *SHH* at a long distance, suggesting that 3D chromatin structure plays an important role in *SHH* expression.

Cancer is another state in which loss or gain of enhancer function plays a distinct role in disease onset and progression. This is achieved both by mutations within an enhancer sequence itself or by changes in enhancer activity via epigenetic modifications (Yegnasubramanian *et al.* 2011; Akhtar-Zaidi *et al.* 2012; Stergachis *et al.* 2013; Taberlay *et al.* 2014). Furthermore, fascinating advancements in the field of phase separation have revealed a novel mechanism of action that may contribute to oncogenic transcriptional activation. In this mechanism, fusion proteins – which are comprised of an intrinsically disordered domain from one protein and a DNA-binding domain from another – are proposed to bind to enhancers and gene promoters, recruiting bound loci into phase-separated condensates, prompting aberrant gene expression (Quiroga *et al.* 2022). A remarkable example of one such fusion protein is NHA9, which is comprised of the intrinsically disordered domain from nucleoporin 98 (NUP98) and the DNA binding domain of homeobox A9 (HOXA9). NHA9 is implicated in various forms of leukemia (Gough

et al. 2011), and which was recently shown to induce DNA looping between enhancers and proto-oncogenes via phase separation (Ahn *et al.* 2021). This is similar in concept to enhancer hijacking, in which enhancer gain-of-function can cause transcriptional deregulation correlated with cancer development (Gröschel *et al.* 2014; Lupiáñez *et al.* 2015).

A unique example of enhancer function in disease is within sickle cell disease. Sickle cell disease is an inherited group of disorders in which red blood cells become misshapen and die, decreasing the number of healthy red blood cells and blocking the flow of blood ("What is Sickle Cell Disease?" 2022). The disorder arises from mutations in the adult hemoglobin (also known as β-globin) genes, expression of which predominates over fetal hemoglobin genes from the time of birth (Stamatoyannopoulos 2005; Wang and Thein 2018). Subsequently, the pathological effect of sickle cell disease has been shown to be reduced by elevated levels of fetal hemoglobin, which can be induced by the drug hydroxyurea (Platt 2008; Sankaran and Nathan 2010). Strikingly, forced chromatin looping between the β-globin locus enhancer (also known as the locus control region, or LCR) and the fetal hemoglobin genes triggered transcriptional reactivation of these genes and increased fetal hemoglobin levels in adult sickle cells (Deng *et al.* 2014; Krivega and Dean 2016; Breda *et al.* 2016; Peslak *et al.* 2023). The implications for utilizing 3D chromatin structure to influence gene expression via physical contact with regulatory elements is a unique and tantalizing option for therapeutics. Taken together, these data provide evidence not only for the role of enhancers in causing disease, but also for the manipulation of these regulatory elements in the rescue of disease phenotype.

1.6 Non-coding regulatory elements in osteoarthritis

Investigating the function of regulatory elements in *in vivo* models of disease proves a large challenge: studying disease states in the context of the whole body is complex and requires the use of animals models, which, while necessary for the development and advancement of many scientific endeavors, remain genetically different from humans. *In vitro* models of disease, while more easily manipulable, lack the physiological environment that *in vivo* models provide. *Ex vivo* models combine the benefits of both *in vivo* and *in vitro* systems – the use of cells isolated from human donors for the provision of an appropriate genetic background and the ease of manipulation in culture in a disease-relevant context.

Accordingly, we use an *ex vivo* model of osteoarthritis (OA) in which to study the impact of genetic variants within non-coding regions of the genome on gene expression. This model utilizes the treatment of chondrocytes obtained from human donors with fibronectin fragment (FN-f) (Loeser *et al.* 2012; Reed *et al.* 2021). Chondrocytes are the only cell found in cartilage and have been previously implicated in OA. Furthermore, the breakdown product of fibronectin – a component of the extracellular matrix produced by chondrocytes – is found in high levels of OA tissue (Carnemolla *et al.* 1984; Homandberg *et al.* 1998). This *ex vivo* system was found to sufficiently mimic the OA phenotype (Reed *et al.* 2021).

While several OA genome-wide association studies (GWAS) have identified OA-associated risk variants (Tachmazidou *et al.* 2019; Boer *et al.* 2021), none have integrated DNA looping methods to interrogate the contribution of 3D genomic structure to transcriptional regulation in OA. However, previous studies have shown that 3D chromatin architecture can reveal genes affected by non-coding disease variants (Wang *et al.* 2013; McGovern *et al.* 2016; Shao *et al.* 2019). To that end, we assessed the transcriptional and architectural effects in OA progression in chondrocytes by connecting previously identified OA-associated genetic variants with their putative target genes. This work addresses the need to define the genetic contribution of OA with the use of novel genomic topological and genome editing techniques.

CHAPTER 2: THE ROLE OF AP-1 IN LOOP-BASED GENE REGULATION

2.1 Contributions

While I was the primary lead, this project was initially headed by Kamisha T. Woolery. K.T.W. established culture procedures for THP-1 cells, designed primers and guide RNAs for CRISPR experiments, and developed the first protocols for CRISPR-Cas9 transfection, single cell colony growth, genomic DNA extraction, and validation experiments. Following project handover, I optimized CRISPR and clonal selection protocols, validation experiments, and generated the majority of the data. Robert J. Fisher contributed to the work part-time as an undergraduate research assistant.

2.2 Introduction

The human cellular nucleus is only approximately 6 micrometers in diameter but contains over 3 billion base pairs of DNA measuring two meters in length. How this genetic material is stored in the nucleus has been studied extensively. The first level of packaging comprises DNA wrapped around histone proteins – the nucleosome – which serves as the fundamental structural unit of chromatin (Ozer *et al.* 2015). The 3D structure of the genome is believed to be linked to its function, but the exact relationship between gene expression and genome organization remains an open-ended question, as this interplay appears to be dynamic and strongly dependent on each other (van Steensel and Furlong 2019). Chromatin is compartmentalized in the nucleus based on accessibility: open chromatin is typically found in the interior of the nucleus while closed chromatin is associated with the nuclear periphery (Cremer *et al.* 2006; Bank and Gruenbaum 2011). Open chromatin, otherwise known as euchromatin, contains actively transcribed genes and histone marks such as H3K27ac and H3K4me3, and preferentially associates with other open regions, forming what is considered the ‘A’ compartment. Closed chromatin, or heterochromatin, encompasses inactive genes and histone marks such as H3K9me3 and H3K27me3, and similarly self-associates into the ‘B’ compartment (Rao *et al.* 2014; Hildebrand and Dekker 2020). Recent breakthroughs have revealed that genome organization exists not only in this broad-scale

compartmentalization, but also on a finer scale, as small as thousands of base pairs (Rowley *et al.* 2017; Schwarzer *et al.* 2017).

Proper cellular function requires precise control of gene transcription, which is governed in large part by hundreds of thousands of enhancers. These regulatory elements recruit transcription factors and can influence genes from up to a million base pairs away (Williamson *et al.* 2011). The mechanism by which enhancer interact with genes at such a distance is unclear but is thought to be mediated in part by DNA looping, which brings promoters and enhancers into close 3D proximity. Many long-range interactions between promoters and enhancers have been documented, such as those at the beta globin locus (Carter *et al.* 2002; Tolhuis *et al.* 2002; Palstra *et al.* 2003), sonic hedgehog (SHH) (Amano *et al.* 2009; Williamson *et al.* 2012), PITX1 (Kragesteen *et al.* 2018), and the insulin-like growth factor 2 (IGF2)/H19 locus (Murrell *et al.* 2004), among many others. Importantly, perturbation of enhancer function or long-range contacts between enhancers and promoters has been shown to play role in developmental malformations and disease, such as at limb morphogenesis (Kragesteen *et al.* 2018), sex reversal in flies (Gonen *et al.* 2018), inflammatory bowel disease (Meddens *et al.* 2016), and cancer (Pittman *et al.* 2010; Gröschel *et al.* 2014; Trimarchi *et al.* 2014; Jäger *et al.* 2015).

Several proteins have been well-documented in DNA loop formation, with the key players being CCCTC-binding factor (CTCF), cohesin, Nipped-B-like protein (NIPBL), and Wings apart-like protein homolog (WAPL). The mechanism by which enhancers and promoters are brought into close physical contact is thought to be a process called “loop extrusion.” In this model, NIPBL loads the ring-like cohesin onto chromatin, and DNA is pushed through to form loops (Sanborn *et al.* 2015; Davidson and Peters 2021). CTCF acts as a boundary to cohesin traversing DNA, as evidenced by the association of CTCF at cohesin binding sites (Wendt *et al.* 2008). Cohesin is stalled when it encounters two convergent CTCF motifs, effectively bringing together two linearly distant regions of DNA as “anchors,” with the intervening DNA looping out (Rao *et al.* 2014). Cohesin is then unloaded from DNA by WAPL (Haarhuis *et al.* 2017). Despite the evidence supporting the existence of 3D chromatin structure, the degree to which DNA looping facilitates transcription remains unclear. Studies perturbing loops between enhancers and promoters genome-wide have produced conflicting results. Acute depletion of cohesin in the HCT-116 human colorectal carcinoma cell line revealed loss of loop domains, but with relatively little effect on

transcription (Rao *et al.* 2017). Conversely, NIPBL deletion in mouse liver affected the expression of over one thousand genes (Schwarzer *et al.* 2017). Furthermore, depletion of cohesin in mouse macrophages significantly impacted expression of inflammatory genes in response to a microbial stimulus (Cuartero *et al.* 2018). Taken together, these data indicate a role for DNA looping in regulating transcriptional changes, possibly in a cell-type-specific manner.

While NIPBL, cohesin, CTCF, and WAPL are well-studied as key chromatin architectural proteins in multiple cell types, a recency study has also revealed the enrichment of Activator Protein-1 (AP-1) at novel loops formed during macrophage development (Phanstiel *et al.* 2017). AP-1, a heterodimer made up of varying combinations of FOS, JUN, MAF, ATF, and CREB family proteins, plays many roles in cellular differentiation pathways (Curran and Franza 1988; Angel and Karin 1991; Chinenov and Kerppola 2001). The number of potential transcription factor combinations that make up AP-1 results in varying functional activity for the heterodimer, dependent on tissue- and cell type-specificity (Mechta-Grigoriou *et al.* 2001). Subsequently, the intricate relationships between the AP-1 family proteins have been implicated in cellular proliferation, tumorigenesis, tumor suppression, and apoptosis (Shaulian and Karin 2002; Eferl and Wagner 2003).

Phanstiel et al demonstrated that enhancers containing both CTCF and AP-1 binding sites formed loops with promoters of critical macrophage developmental genes, such as Interleukin 1 beta (*IL1β*) during macrophage differentiation (**Figure 2.1**) (Phanstiel *et al.* 2017). Macrophages, phagocytic cells of the innate immune system, are critical in suppressing inflammation in response to tissue damage (Watanabe *et al.* 2019). Treatment of monocytic precursors using phorbol myristate acetate (PMA) has been well-characterized as a model for studying monocyte-macrophage differentiation (Daigneault *et al.* 2010). *IL1β* is a canonical proinflammatory cytokine produced by activated macrophages that mediates inflammatory responses and is involved in a variety of biological processes, including proliferation, differentiation, and apoptosis (Lopez-Castejon and Brough 2011). Dysregulation of *IL1β* expression has been associated with inflammatory and autoimmune diseases (Lamkanfi and Dixit 2012; Yao *et al.* 2016).

The mechanism by which CTCF and AP-1 might work together to facilitate enhancer-based gene regulation during monocyte-macrophage differentiation remains unknown. We hypothesized a mechanism in which CTCF forms new loops during macrophage development, bringing enhancer-bound

AP-1 into close physical proximity with macrophage-specific promoters such as *IL1β* to activate their transcription (**Figure 2.2**). To address this, we applied genome editing techniques to determine the requirement for CTCF and AP-1 binding in DNA loop formation and *IL1β* transcription. CTCF and AP-1 binding sites at the previously identified upstream *IL1β* enhancers were removed using CRISPR-Cas9 methods in THP-1 monocytic precursor cells (**Figure 2.3**). DNA loops were mapped using hybrid capture Hi-C and *IL1β* expression was determined using qPCR. Quantification of DNA looping and *IL1β* transcription in wild-type and genome-edited clones pointed toward a putative role for AP-1 in loop-based transcriptional activation of *IL1β.*

2.3 Results

2.3.1 Optimization of single cell clonal expansion[1]

CRISPR-Cas9 technology enables the rapid generation of mutations in mammalian cell lines (Doudna and Charpentier 2014). After genomic editing is performed, the bulk population contains a mixture of cells that include non-edited and a heterogenous blend of edited sequences. Due to selective pressure on the heterogenous cell pool, single cells are isolated from the bulk population to grow clonal populations containing the desired edits. Serial or limiting dilution is a common technique utilized to isolate single cells for this purpose.

However, single cell cultures may experience stress and may require additional aid for growth. Conditioned media is a culture media that contains biologically active components obtained from previously cultured cells that have released into the media substances that affect certain cell functions, such as growth. Observations of single cell growth in conditioned media have indicated that culture in conditioned media increases the survival rate during clonal expansion (Rathjen and Geczy 1986; Peters 2014; Yumlu *et al.* 2017).

Wild-type THP-1 populations were initially tested for growth using fresh media. Following serial dilution to 1 cell, 10 cells, 20 cells, and 50 cells in a 96-well plate, plates were incubated at 37ºC, 5% CO_2, for four days, upon which cell counts were taken. It was determined that cells plated in fresh media did not survive.

[1]Optimization of single cell clonal expansion was performed by Kamisha Woolery.

To test the effect of conditioned media on single cell growth, single cell THP-1s were plated in either 100% fresh media, 100% conditioned media, 50% conditioned media/50% fresh media, or 25% conditioned media/75% fresh media. Conditioned media was obtained by spinning down wild-type THP-1 cell culture and transferring the supernatant to a new tube before spinning again. The supernatant was then diluted with fresh media to obtain 25%/ and 50% conditioned media/fresh media. Plates were incubated for 3 weeks at 37ºC, 5% CO_2. Observation of cell growth indicated no growth in wells containing 100% fresh media or 25% conditioned media/75% fresh media and approximately equal growth in 100% conditioned media and 50% conditioned media/50% fresh media.

It was concluded that both 100% and 50% conditioned media/50% fresh media were more optimal for single cell clonal expansion than fresh media. Since both 100% and 50% conditioned media are favorable for cell growth, future experiments would use 75% conditioned media/25% fresh media. Conditioned media would also be vacuum filtered to prevent potential cell contamination.

2.3.2 Optimization of CRISPR-Cas9 editing efficiency in THP-1s

We performed genomic editing of THP-1s using a method of CRISPR-Cas9 delivery termed ribonucleoprotein (RNP) electroporation. In this method, the Cas9 enzyme and guide RNAs (gRNAs) are complexed together and transfected into cells using electroporation (Zuris *et al.* 2015; Wang *et al.* 2017). RNP electroporation offers several advantages over the use of plasmid delivery via cationic lipids: electroporation is faster, eliminates the need for intracellular transcription and translation, reduces off-target effects (Chandrasekaran *et al.* 2018; Lattanzi *et al.* 2019).

Initial genomic editing of THP-1s using RNP electroporation of CRISPR-Cas9 components revealed that editing efficiency was extremely low (**Figure 2.4, panel A**). Various troubleshooting experiments were performed, including editing of very low passage THP-1 cells (**Figure 2.4, panel B**), editing with freshly complexed gRNAs (**Figure 2.4, panel C**), and editing with newly ordered Cas9 enzyme (**Figure 2.4, panel D**). However, gel electrophoresis of the PCR amplified region of interest (in this case, the CTCF binding site at the first *IL1β* enhancer) showed that editing efficiency was either very low or zero in all experiments.

In an attempt to increase editing efficiency, RNP electroporation was performed with the addition of Cas9 Electroporation Enhancer (IDT, cat. no. 1075915). This single-stranded DNA oligonucleotide,

computationally designed to be nonhomologous to human, mouse, and rat genomes, is a Cas9-specific carrier DNA that is optimized to work with the Amaxa Nucleofector device (see Materials and Methods section 2.4.3) ("Start genome editing with CRISPR-Cas9"). During RNP electroporation, the Cas9 Electroporation Enhancer was added to the pooled gRNAs at a final concentration of 2μM prior to combining Nucleofector-resuspended cells with the RNP complex. 'Mock' cells that did not receive RNP complex were included as a control and electroporation was performed as described.

Following DNA extraction and PCR amplification of the region of interest, gel electrophoresis showed that the addition of the Cas9 Electroporation Enhancer appeared to remarkably increase editing efficiency (**Figure 2.5**, 'edited + enh.' lane). 'Mock' cell DNA did not contain any deletions as expected (**Figure 2.5**, 'mock' lane). DNA from edited THP-1 cells without the Cas9 Electroporation Enhancer indicated a very low editing efficiency, represented by an extremely faint band the size of the expected deletion product (**Figure 2.5**, 'edited -enh.' lane).

2.3.3 DNA looping plays a role in the activation of *IL1β* in developing macrophages

Temporally assessing changes during biological events can reveal distinct remodeling of the cellular state that two time points cannot. In order to evaluate the impact of cellular differentiation on chromatin loop-based gene regulation, we collected gene expression, enhancer activation, and DNA looping data at the *IL1β* locus over the course of monocyte differentiation and macrophage activation. Hybrid-capture Hi-C (Hi-C^2) was performed on Hi-C libraries generated from monocytes, macrophages, and LPS/IFγ-activated macrophages at 30, 60, 90, 120, 240, and 360 minutes during treatment. Hi-C^2 uses RNA probes to enrich for target regions within a Hi-C library, allowing for deeper sequencing of a desired region at a reduced cost (Gnirke *et al.* 2009; Sanborn *et al.* 2015).

RNA-sequencing and ChIP-sequencing (H3K27ac) data were also collected from monocytes, macrophages, and LPS/IFγ-activated macrophages at the same timepoints by Kathleen Reed in the Phanstiel lab. Looping data was then overlaid with RNA- and ChIP-seq data (**Figure 2.6**). The data suggests that looping between *IL1β* and its nearest upstream enhancer ("enhancer 1") (**pink line**) precedes enhancer activation (as identified by H3K27ac levels, **blue line**), which precedes *IL1β* expression (**yellow line**) (**Figure 2.6**). This supports a model in which DNA looping plays an important role in the activation of *IL1β* in developing macrophages.

2.3.4 Preliminary data suggests that *IL1β* expression in macrophages is partially dependent on AP-1 binding[1]

THP-1 monocytes were edited with CRISPR-Cas9 to remove the AP-1 binding site at the closest upstream enhancer looped to *IL1β* (see **Figure 2.3, panel A** for a cartoon depiction). CRISPR-edited clones were then differentiated into macrophages with 100nM PMA and RNA was extracted at 0, 0.5, 4, and 24 hours during treatment. RT-qPCR was performed to measure *IL1β* expression over time in wild-type and heterozygous and homozygous deletion clones (**Figure 2.4**). Expression was normalized to actin. While *IL1β* expression increased during treatment in all 3 genotypes, expression was 2-3-fold lower in heterozygous (brown line) and homozygous (gray line) deletion clones when compared to wild-type (yellow line). These results previously determined by the Phanstiel lab suggest that AP-1 enhancer binding is important for proper expression of *IL1β* during macrophage development.

2.4 Discussion

2.4.1 Challenges and limitations

There were several challenges and limitations in this project that resulted in the discontinuance of the project. The following sections describe the main obstacles in this work.

2.4.1.1 Low editing efficiency

Low editing efficiency of THP-1 monocytes was one of the main roadblocks in this project. THP-1s are a human leukemia monocytic cell line that have been used extensively to study monocyte and macrophage functions and immunological pathways such as differentiation and phagocytosis (Chanput *et al.* 2014). This immortal proliferating cell line is often used in research in place of primary monocytes, which do not proliferate in culture (van Furth *et al.* 1979). However, as macrophage precursors, THP-1s are notoriously difficult to transfect due to a mounted immune response against the introduction of foreign nucleic acids (Dokka *et al.* 2000; Keller *et al.* 2018).

Methods of transfection include those biological (via viral transduction), chemical (via polymer or lipid deposition), and physical (via electroporation, microinjection, or biolistic particle delivery). Several protocols have been published that purport efficient transfection of THP-1 cells using a type of

[1]Prelimary data was collected by Kamisha Woolery.

electroporation, termed nucleofection, as opposed to classical methods such as lipofectamine (Martinet *et al.* 2003; Schnoor *et al.* 2009). Nucleofection combines cell-specific electric pulses with chemical reagents to allow the direct transfer of nucleic acid materials into the nucleus. This is a popular method of transfecting eukaryotic primary cells, as viral methods tend to be expensive, time-consuming, and require special laboratory equipment and safety measures. Transfection via nucleofection has been successfully applied to many different cell types, including those considered difficult to transfect (Lakshmipathy *et al.* 2007; "Leukemia and Lymphoma Cell Lines" 2014; "THP-1: Human acute monocytic luekemia" 2016). In comparison to published protocols and methods, THP-1 editing efficiency in our experiments remained low: initial RNP electroporation experiments resulted in little to zero transfection (see **Figure 2.4**). The addition of the Cas9 Electroporation Enhancer did increase transfection efficiency such that deletions were identifiable via PCR amplification and gel electrophoresis in bulk edited populations (see **Figure 2.5**). However, genotyping of clonal populations following single cell expansion revealed, that of the single cells plated that retained viability for growth, only a very small percentage contained heterozygous or homozygous deletions. This low editing efficiency proved to be a major challenge to growing up clonal populations containing the desired CRISPR edits at the AP-1 and CTCF binding sites at *IL1β*s upstream enhancers.

2.4.1.2 Challenges in isolating clonal populations

Genome editing of a bulk population of cells results in heterogenous genotypes throughout the population. Following bulk population editing, clonal populations are established via single cell isolation and expansion, producing a population with a homogenous genotype, also known as a monoclonal population. These monoclonal populations can be further expanded for immediate use in experiments as well as frozen down for future experiments. However, the methods for isolating single cells in suspension cells are often imprecise and unreliable.

Single cell isolation in this project was first attempted using serial dilution (also known as limiting dilution). Serial dilution involves setting up a series of increasing dilutions of the bulk population of cells in order to plate a single cell in wells of a 96-well plate for expansion into a monoclonal population. This method has been used for decades – initially described in antibody production – due to its convenience and low-cost (Goding 1980; Fuller *et al.* 2001). However, because of the statistical nature of serial

dilution, the method is not very efficient. On average, only about one-third of the wells plated will contain a single cell and often-times wells will contain multiple cells, making it nearly impossible to identify whether the expanded population is truly monoclonal. Thus, it may be required that several 96-well plates are prepared via serial dilution. Which wells contain cells have to be confirmed after plating, typically by standard microscopy, which is also imprecise and time-consuming. Confirmation of single cells in a well of a 96-well plate is difficult, even when one well is plated with >1000 cells with which to provide a focus plane for the single cell wells.

Due to low success with serial dilution to obtain clonal populations of CRISPR-edited cells, we also explored a flow cytometry technique known as fluorescence-activated single cell sorting, or FACS. FACS uses a flow cytometer to select for cells with a specified parameter – such as live vs dead and/or fluorescent labels. To that end, we used a fluorescently labeled tracrRNA (tracrRNA-ATTO550; IDT, cat. no 1075927) during CRISPR-Cas9 genome editing, with the intent to isolate single cells that had been successfully transfected. Typically, prior to sorting, cells are washed with PBS containing 1% FBS to minimize non-specific binding. Several control and experimental groups were included: wild-type THP-1s (not edited, electroporated, or washed), non-edited THP-1s (not electroporated but incubated with CRISPR-Cas9 components, with either targeting or non-targeting gRNAs) both washed and not washed, and edited THP-1s (electroporated with either targeting or non-targeting gRNAs) both washed and not washed (**Figure 2.S1**). Control and experimental groups were identifiable as visual peaks within FACS analysis; however, it was not clear what the editing efficiency was in the 'edited THP-1' population due to the presence of a single peak (**Figure 2.S1**, yellow peak) within the FACS analysis. Although this result appears to suggest a transfection efficiency of 100%, that is an unlikely occurrence. Cells from this 'edited THP-1' population (**Figure 2.S1**, yellow peak) were plated as single cells in 96-well plates. Unfortunately, observation of growth revealed that many of the single cells plated did not survive sorting. Of the few cells that did survive and expanded to form clonal populations, genotyping showed that editing efficiency was low, with the majority containing no edits (**Figure 2.S2**). In conclusion, we hypothesized that either the 'edited THP-1' population initially took up the CRISPR-Cas9 components but were not retained long enough for edits to take place, or the ATTO550 fluorophore is sticky such that cells were 'tagged' with the fluorophore and thus appeared as positively transfected in the FACS analysis.

2.4.1.3 Irreproducibility of characterization of edited cell lines

Another major challenge of this work was the difficulty in validating CRISPR-Cas9 deletions. Despite extremely low editing efficiency, occasionally genotyping experiments revealed a putatively successful transfection and CRISPR edit. However, DNA amplification via PCR to visualize these potential deletions was not reproducible. We hypothesized that the long length of the full-length PCR products might be contributing to the struggle in replicating PCRs. Amplicon sizes ranged from 1000bp to over 3000bp for full-length products. We used the Phire Hot Start II DNA Polymerase (Thermo Fisher, cat. no. F122S), which is ideal for amplifying long DNA fragments with short extension times of 10 to 15s/kb. Despite multiple experiments in which annealing temperature, extension time, and PCR additives such as DMSO were adjusted, PCRs remained irreproducible from experiment to experiment.

An additional complication in validating clonal populations was inconclusive Sanger sequencing results. Often, traces came back with no discernible peaks, too much background, and/or abrupt termination of signal. Repeated attempts were made to ensure high quality purified template, sufficient template and primer quantity, and well-designed primers were provided for Sanger sequencing. Chromatogram peaks oftentimes did not provide quality traces that reached the location of the guide RNA sites, making it difficult to identify whether or not a population contained the desired deletion.

Additionally, analysis of traces that were of good quality frequently did not match the gel electrophoresis results. For example, PCR amplification followed by visualization via gel electrophoresis of a clonal population might reveal a homozygous deletion, whereas Sanger sequencing results would indicate a heterozygous deletion. Heterozygous clones were particularly difficult to identify - overlapping Sanger sequencing traces might indicate differential editing on either allele. However, attempts at deconvolution using CRISP-ID (Dehairs *et al.* 2016) often revealed 3 or more overlapping traces. We hypothesized this might be due to inexact single cell isolation, in which two or more cells were initially plated instead of one. This would result in a heterogenous population instead of a monoclonal population. In this scenario, genotyping via PCR and gel electrophoresis might indicate a clonal population with a heterozygous deletion, but Sanger sequencing traces would appear ‘messy,’ and deconvolution would reveal three or more alleles present within the edited cell population due to differential editing of the initially isolated and plated cells. It would be extremely difficult to identify whether the initial cells plated all

had heterozygous deletions or contained a combination of homozygous deletion and wild-type genotypes. This speaks not only to the complexity of decoding Sanger sequencing results, but again to the lack of effectiveness of single cell isolation in these experiments.

2.4.2 Future directions

2.4.2.1 Viral transduction of THP-1 cells as an alternative to RNP electroporation

While viral transduction of cells as a method of genome editing can be time-consuming, expensive, and require special safety measures, it may be a more viable option for editing THP-1 cells than RNP electroporation. For stress-sensitive cells such as THP-1s, the strong electrical current produced by electroporation may result in a high percentage of cell death and/or trigger cellular stress responses. Thus, viral introduction of Cas9 and guide RNAs may provide a more efficient means of editing the desired target region in THP-1s. Several studies have successfully expressed CRISPR-Cas9 components in THP-1 cells using lentiviral transduction (Goetze *et al.* 2017; Xiao *et al.* 2020; Lindner *et al.* 2021). Additionally, several labs at UNC produce stable expression cell lines using lentivirus. The Phanstiel lab has previously been in contact with the lab of Dr. Bernard Weissman for support with lentivirus protocols.

Lentiviral transduction of THP-1s would involve the production of virus by transfecting HEK293FT cells with a vector plasmid containing Cas9 and gRNA sequences as well as an antibiotic resistance gene, viral packaging plasmids, and a transfection efficiency control (a plasmid containing GFP). Transfection efficiency would be assessed via microscopy for GFP. Virus would then be collected and transduction of THP-1s performed. Since THP-1s are a suspension cell, ‘spinfection’ (transducing cells with centrifuging) may be an optimal method of transduction as opposed to simple incubation with the virus (Lindner *et al.* 2021). Successfully transduced cells could be selected for via addition of antibiotics. Alternatively, a Cas9-GFP sequence could be included in the vector plasmid, and successfully transduced cells could be single cell sorted via FACS for subsequent clonal expansion.

2.4.2.2 Gain- and loss-of-function genome-editing at the IL1β locus

Another CRISPR tool involves the use of a catalytically dead Cas9 (dCas9), which can be recruited to and bind specific sequences of DNA but lacks the enzymatic capability to cleave target regions. dCas9 is often used in conjunction with CRISPR inactivation (CRISPRi) or CRISPR activation

(CRISPRa) in order to affect expression at the transcription level. These systems provide more biologically relevant models by affecting genes within their natural environment than traditional techniques such as CRISPR-induced mutation/deletion (also known as CRISPR knockout or CRISPRko) and plasmid gene overexpression. In the interest of investigating the role of AP-1 in looping at the *IL1β* locus specifically, CRISPRko of AP-1 components may not be a viable option due to genome-wide effects. Thus, CRISPRi and CRISPRa serve as alternative options for interrogating the role of looping at *IL1β* on macrophage development. CRISPRi can be achieved by fusing dCas9 with a transcriptional repressor domain, such as Krüppel-associated box, or KRAB (dCas9-KRAB) (Qi *et al.* 2013; Gilbert *et al.* 2014; Kampmann 2018). Similarly, CRISPRa utilizes dCas9 fused to a transcriptional activator domain, such as those derived from the Herpes simplex virion protein 16 (VP16) and 64 (a tetramer of VP16) (Cheng *et al.* 2013; Zalatan *et al.* 2015; Balboa *et al.* 2015).

Our model suggests that the main function of the DNA loop at the *IL1β* locus is to bring AP-1 into close physical proximity of the *IL1β* promoter and that it is this close physical distance between AP-1 and the gene promoter that prompts increase transcription. To test this hypothesis, CRISPRa could be used with dCas9 fused to the transactivating domain of one of major AP-1 subunits, c-Jun (Curran and Franza 1988; Chinenov and Kerppola 2001). By targeting c-Jun to the *IL1β* promoter with dCas9 and a complementary gRNA and subsequently measuring *IL1β* expression via qPCR in monocytes, it could be determined whether or not AP-1 is sufficient for proper *IL1β* transcription. CRISPRi could be used in conjunction to suppress *IL1β* expression in developing macrophages, confirming that loop formation during macrophage development is not affected and is not dependent on *IL1β* expression.

2.4.2.3 Assessing the genome-wide effect of AP-1 on loop-based gene regulation

We hypothesize that loop-based gene regulation by AP-1 is a common phenomenon; therefore, deletion of AP-1 should affect all genes that loop with AP-1-bound enhancers. As an alternative to knocking out core AP-1 subunits, of which there are many (Mechta-Grigoriou *et al.* 2001), a dominant-negative mutant of AP-1 – termed A-Fos or AP-1 DN – has been developed which blocks the function of wild-type AP-1 genome-wide (Olive *et al.* 1997). A-Fos is a variant of the AP-1 subunit Fos with the addition of an acidic amphipathic protein sequence to the N-terminus of the Fos leucine zipper, replacing the normal basic region responsible for DNA binding. A-Fos forms a heterodimeric complex with Jun,

preventing homo- and heterodimeric complex formation with endogenous Jun and Fos proteins and, subsequently, abolishing DNA binding (Olive *et al.* 1997).

A recent study by El-Fattah Ibrahim et al describes an inducible lentiviral protocol for expressing AP-1 DN (El-Fattah Ibrahim *et al.* 2019). Utilizing this technique, human monocytes could be transduced with the lentivirus containing the A-Fos sequence. Following differentiation into macrophages, A-Fos expression can be induced by the addition of doxycycline (El-Fattah Ibrahim *et al.* 2019) and RNA-sequencing would be performed to quantify RNA levels genome-wide. Using our existing maps of macrophage loops and AP-1 binding sites (Phanstiel *et al.* 2017), we would be able to determine how loss of AP-1 affects transcription of genes at the opposite end of loops in macrophages. We would expect to see decreased expression of genes whose promoters either are bound by AP-1 or interact with AP-1-bound enhancers. These data would further shed light on the precise role of AP-1 in chromatin interactions in a cell-type and tissue-specific manner.

2.5 Materials and Methods

2.5.1 THP-1 cell culture

THP-1 human monocyte cells (*ATCC*, cat #: ATCC® TIB-202TM) were cultured in RPMI 1640 with L-glutamine-bicarbonate supplemented with 10% fetal bovine serum and 1% penicillin streptomycin solution. THP-1s were incubated at a density of $4x10^5$ - $1x10^6$ cells/mL at 37°C in a 5% CO_2 atmosphere. Cultures were maintained either by addition of fresh medium or by centrifugation with subsequent resuspension at 2-$4x10^5$ cells/mL.

2.5.2 crispr RNA (crRNA) design

crRNAs were designed using https://chopchop.cbu.uib.no (Montague *et al.* 2014; Labun *et al.* 2016, 2019) and ordered via IDT's Alt-R Custom crRNA Design Tool. See **Table 2.1** for a list of genomic coordinates and sequences.

2.5.3 Ribonucleoprotein (RNP) electroporation

Alt-R CRISPR-Cas9 trans-activating CRISPR RNAs (tracrRNA) were ordered from IDT (cat. no. 1072532). To form the crRNA:tracrRNA duplex, each RNA oligo was resuspended to a final concentration of 200μM in 1X IDTE buffer. The two RNA oligos were then mixed in equimolar concentrations in a sterile

microcentrifuge tube to a final duplex concentration of 100μM, heated at 94°C for two minutes, then allowed to cool to room temperature (stored at -20°C if necessary).

To form the RNP complex, the Amaxa 4D nucleofector solution (Lonza Primary Human Monocytes 4D-Nucleofector® Kit, cat. no. V4XP-3012) was supplemented according to directions and brought to room temperature. 12 well cell culture plates were filled with 1.5mL cell media and stored in the incubator and aliquots of 500μL cell media were prewarmed for post-nucleofection. Enough cell culture was placed into a sterile 15mL conical tube for $1x10^6$ cells per cuvette. For each guide RNA (gRNA), the crRNA:tracrRNA duplex (120pmol final concentration) was mixed with S.p Cas9 Nuclease 3NLS (104pmol final concentration) to a final volume of 5μL and incubated at room temperature for 10-20 minutes. During incubation, the required number of cells were spun down at 300xg for 5 minutes at room temperature, washed with 1X PBS, spun again, and resuspended in 88μL of nucleofector solution. Left and right gRNAs were pooled into one tube and the resuspended cell solution was added along with 100μM Cas9 electroporation enhancer (final concentration 2μM), mixed gently by pipetting, then transferred to a Nucelocuvette. As a control, 'mock' cells were resuspended in 100μL of Nucleofector solution and transferred to a Nucleocuvette as well. Nucleocuvettes were placed in the 4D-Nucleofector X-Unit and the electroporation program for THP-1 monocytes was ran. After run completion, 500μL of pre-warmed cell media was added to each cuvette, mixed gently by pipetting, and transferred to pre-filled culture plates. Cells were then incubated for 48 hours prior to genotyping.

2.5.4 Genotyping CRISPR edited bulk populations

Genomic DNA was extracted from CRISPR-Cas9 edited and mock cells following the Qiagen QIAamp DNA MiniKit (cat. no. 51304) according to manufacturer's instructions for genomic DNA purification from cells (adapted from Qiagen's "QIAamp DNA Mini and Blood Mini Handbook"). Approximately half of the electroporated cell suspension described earlier was used for DNA purification. Following extraction, DNA quantification was performed using a Qubit dsDNA HS Assay (Thermo Fisher Scientific, cat. no. Q32851). PCR amplification was performed following the Thermo Fisher Scientific Phire Hot Start II DNA Polymerase Manual (Phire Hot Start II DNA Polymerase, Thermo Fisher, cat. no. F122S) with 50ng of DNA as an input. The 3-step protocol was used for cycling, with annealing temperatures determined using Thermo Fisher's T_m calculator, and extension time at 10-15s/kb. 25-27

cycles were run. See **Table 2.2** for PCR primer sequences and coordinates. PCR products were then visualized via gel electrophoresis (2% agarose).

2.5.5 Isolating single cells via serial dilution

To prepare conditioned media, THP-1s plated at approximately $4x10^5$ cells/mL were cultured as previously described. 2-3 days after plating, cells were counted to ensure that culture density was between $8x10^5$ cells/mL and $1x10^6$ cells/mL and greater than 90% live. Cells were then transferred to a conical tube and spun down at 300xg for 5 minutes at room temperature. The supernatant was transferred to a 0.22μM vacuum filtration unit and 25% final volume fresh cell media was added (final volume is 75% conditioned media, 25% fresh media). After filtration, 200μL of the final conditioned media was added to each well of a 96-well cell culture plate.

To serially dilute CRISPR edited THP-1 cells in order to isolate single cells, cells were diluted to 500,000 cells/mL in 1mL (dilution 1). 20μL of dilution 1 was added to 980μL of fresh media (dilution 2, ~10,000 cells/mL). 20μL of dilution 2 was then added to 980μL of fresh media (dilution 3, ~200cells/mL). 10μL of dilution 3 was pipetted into the middle wells of the 96-well plate filled with 200μL of conditioned media (~1-2 cells/well). Cells were not added to the wells on the outer edge of the plate due to potential evaporation during incubation. 'Mock' edited cells were diluted in the same way and plated in the bottom row of the 96-well plate. Clonal populations were cultured for approximately 4 weeks before transferring wells with growth to a 12-well plate for further expansion.

2.5.6 Genotyping single cell clonal populations

Genotyping was performed as in section 2.5.4.

2.5.7 PCR cleanup and Sanger sequencing

Prior to Sanger sequencing, the protocol for Enzymatic PCR cleanup using Exonuclease I and Shrimp Alkaline Phosphatase (NEB, cat. no. M0293 and M0371) was followed according to manufacturer's instructions. Sanger sequencing was performed via EtonBio; 5μL of EXO-Sap purified product was provided along with 5μL of 5μM primer.

2.5.8 Macrophage differentiation

Approximately $40x10^6$ cultured THP-1 monocytes were collected, spun down, and resuspended in fresh media in a T125 flask to a final concentration of $1x10^6$ cells/mL. 12-O-tetradecanoylphorbol-l3-

acetate (PMA) (Sigma-Aldrich, cat. no. P1585-1MG) was added to a final concentration of 100μM. Cells were incubated for 72hr at 37°C, 5% CO_2.

2.5.9 Crosslinking of monocytes and differentiated macrophages

Approximately $40x10^6$ cultured THP-1 monocytes were spun down at 3000xg for 5 minutes and supernatant was aspirated prior to resuspension in 1mL of fresh media per 1 million cells. Cells were crosslinked using 1% formaldehyde in RPMI for 10 minutes with gentle shaking. Crosslinking was quenched with 2.0M cold glycine (final concentration 0.2M) for 5 minutes with shaking. Cells were spun down at 3000xg for 5 minutes and supernatant was aspirated prior to washing with 1mL of cold PBS per 5 million cells. Cells were aliquoted in 5 million cell aliquots, spun down at 3000xg for 5 minutes at 4°C. Supernatant was aspirated, and pellets were flash frozen in liquid nitrogen prior to storage at -80°C.

For differentiated macrophages, cells were washed with PBS and trypsinized for 3 minutes. Trypsin was quenched with 2x volume fresh cell media. The cell suspension was spun down at 3000xg for 5 minutes, supernatant was aspirated, and cells were resuspended in fresh media at a concentration of $1x10^6$ cells/mL. Cells were then crosslinked and stored as described in the previous paragraph.

2.5.10 In situ Hi-C library preparation during macrophage differentiation

Two frozen pellets ($5x10^6$ cells each) of both THP-1 monocytes and differentiated macrophages were used to generate separate libraries as technical replicates. Libraries were prepared using the in situ Hi-C protocol as described in Rao et al. 2014 (Rao *et al.* 2014). In brief, crosslinked cells were lysed on ice, nuclei were isolated, and chromatin was digested overnight with the MboI restriction enzyme. Chromatin ends were biotinylated, proximity ligated, and crosslinking was reversed. Samples were sheared on a Covaris LE220 (duty cycle 20, intensity 10, 200 cycles/burst, 90 seconds, 4°C), quantified using Qubit (dsDNA High Sensitivity (HS) assay), and a small sample was run on an agarose gel to ensure proper fragmentation. DNA sized 300-500 bp was selected for using AMPure XP beads, and then eluted. Biotinylated chromatin was then pulled down using streptavidin beads. Following removal of biotin from unligated ends and repair of sheared DNA ends, unique Illumina TruSeq Nano (Set A) indices were ligated onto the samples. Libraries were amplified off of streptavidin beads using 12 PCR cycles based on post-size selection concentrations, quantified again using a Qubit (dsDNA HS assay), and fragment

length was determined using Tapestation (D1000 tape). Libraries were pooled to 4nM, sequenced on the Miniseq sequencer system (Illumina), and analyzed using the Juicer pipeline to assess quality.

2.5.11 Hi-C^2 library preparation during macrophage differentiation

RNA probes were constructed using a unique oligo pool with index sequences. Software was created by Eric Davis (http://phanstiel-lab.med.unc.edu/lure/) to design a pool of oligonucleotides for Hi-C^2. Probe design and the following experimental protocol were adapted from Sanborn et al (Sanborn *et al.* 2015). Probes were ordered via CustomArray.

Region specific sub-pools targeting the *IL1β* region were first amplified from the overall CustomArray oligo pool. Probes for *IL1β* region were designed at the genomic coordinates chr2:113200000-114160000. 160ng of this region-specific oligo pool was PCR amplified with 10μM first (5′-CTGGGACCTCGCCTATCCCAT-3′) and second (5′-CGTGGACAGACCCCGGTAGTG-3′) primers (IDT) and 2X Phusion Master Mix. The following PCR profile was used: 98°C for 30s; cycle 10-18x: 98°C for 10s, 55°C for 30s, 72°C for 30s; 72°C for 7 min, hold at 4°C. The resulting PCR product was re-amplified 3x to create an aliquot stock of the desired region's probe. A 1.8X SPRI clean-up was performed on the PCR product before elution in 10mM Tris-HCl, pH 7.5. The solution was transferred to a new tube and a T7 promoter was added via PCR with the SPRI-cleaned product, 10μM first primer-T7 (5′-GGATTCTAATACGACTCACTATAGGGCCTCGCCTATCCCAT-3′), 10μM primer 2, and 2X Phusion Master Mix. The PCR profile was used: 98°C for 30s; cycle 12-18x: 98°C for 10s, 55°C for 30s, 72°C for 30s; 72°C for 7 min, hold at 4°C. A 1.8X SPRI clean-up was performed on the resulting PCR product before elution in 10mM Tris-HCl, pH 7.5. DNA concentration was quantified with Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, cat. no. Q32851) to verify that a minimum of 1μg of product was produced. The purified product was transcribed using the MAXIscript T7 transcription reaction with the addition of 10mM biotin-16-UTP (Thermo Fisher Scientific, cat. no. AM1320 & 11388908910) by incubating for 90+ minutes at 37°C. 1μL TURBO DNase I was added and incubated at 37°C for 15 min, followed by addition of 1μL 0.5M EDTA. RNA was quantified with Qubit RNA HS Assay kit (Thermo Fisher Scientific, cat. no. Q32852) to verify an expected yield between 5-15μg. RNA was purified with the Zymo Clean and

Concentrator Kit (cat. no. D4060) with elution in 15µL of water. 1U/µL SUPERase-InRNase inhibitor (Thermo Fisher Scientific, cat. no. AM2694) was added and the product was stored at -80°C.

For hybrid selection, 500ng Hi-C library was mixed and incubated with 2.5µg Cot-1 DNA and 10µg Salmon Sperm DNA at 95°C, 5 min, before holding at 65°C for 5+ min. Hybridization buffer was prepared with 10X SSPE Buffer, 10X Denhardt's Solution, 10mM EDTA, and 0.2% SDS. The RNA probe mixture was prepared at 65°C,2 min with 500ng RNA probes and 20U/µL SUPERase-In RNase Inhibitor. Hybridization buffer, RNA probe mixture, and DNA pond mixture were combined and incubated at 65°C for 24hr.

The Bind-and-Wash buffer (1M NaCl, 10mM Tris-HCl, pH 7.5, 1mM EDTA), the low-stringency wash buffer (1X SSC, 0.1% SDS), and the high stringency wash buffer (0.1X SSC, 0.1% SDS) were prepared. 50µL of streptavidin beads (Thermo Fisher Scientific, cat. no. 65-602) were washed 3x with and resuspended in Bind-and-Wash buffer before incubating with the hybridization mixture for 30 min at room temperature. The beads were reclaimed, washed once with low-stringency buffer, and incubated at room temperature for 15 min. Beads were reclaimed and washed with high-stringency buffer with incubation at 65°C for 10 min for a total of 3x. DNA was eluted off the beads by resuspending and incubating in 0.1M NaOH before reclaiming the beads and transferring the supernatant to a new tube containing 1M Tris-HCl, pH 7.5. A 1X SPRI cleanup was performed using 3X concentrated SPRI beads (Beckman Coulter, cat. no. B23317) with 2 washes in 80% ethanol prior to elution in 10mM Tris-HCl, pH 8.0.
The HiC^2 library was amplified using 2X Phusion Master Mix and Illumina PCR Primer Cocktail with the following PCR profile: 98°C for 30s; cycle 12-18x: 98°C for 10s, 55°C for 30s, 72°C for 30s; 72°C for 7 min, hold at 4°C. Two 0.7X SPRI cleanups were performed, and the final library was eluted in 10mM Tris-HCl, pH 8.0. DNA was quantified using a Qubit dsDNA HS Assay and fragment sizes were obtained using TapeStation (D1000). Libraries were pooled to 4nM, sequenced on the Miniseq sequencer system (Illumina), and analyzed using the Juicer pipeline to assess quality.

2.5.12 Hi-C^2 library preparation of previously collected LPS/IFγ-activated macrophage Hi-C libraries

Hi-C libraries were previously prepared by Kathleen Reed from LPS-IFγ-activated macrophages. Briefly, THP-1 monocytes were differentiated with 25nM PMA for 24hr and then allowed to rest for 72hr in

fresh media. Resting macrophages were then treated with a combination of 10 ng/mL lipopolysaccharide (LPS) and 20 ng/mL interferon gamma (IFNγ) in fresh media. Cells were harvested without treatment, or 0.5, 1, 1.5, 2, 4, 6, or 24 hours after LPS and IFNγ treatment. Crosslinking was performed on cells grown at a density of $2x10^5$ cells/mL using 1% formaldehyde in RPMI for 10 minutes with shaking. Crosslinking was quenched with 0.2M cold glycine for 5 min. The media was removed, and cells were scraped into cold PBS. Each flask was divided into 4 tubes of ~$5x10^6$ cells each. Cells were spun down at 500xg for 5min, resuspended in PBS, and respun to wash residual formaldehyde. Cells were frozen in liquid nitrogen and stored at -80°C for library prep. Hi-C was performed by K.R. as described in section 2.4.10. Hi-C^2 was performed as described in section 2.4.11. Libraries were pooled to 4nM, sequenced on the Miniseq sequencer system (Illumina), and analyzed using the Juicer pipeline to assess quality. For deep sequencing, libraries were pooled to 10nM and paired end 75bp reads were sequenced on a NextSeq (Illumina). Data was processed using the Juicer pipeline.

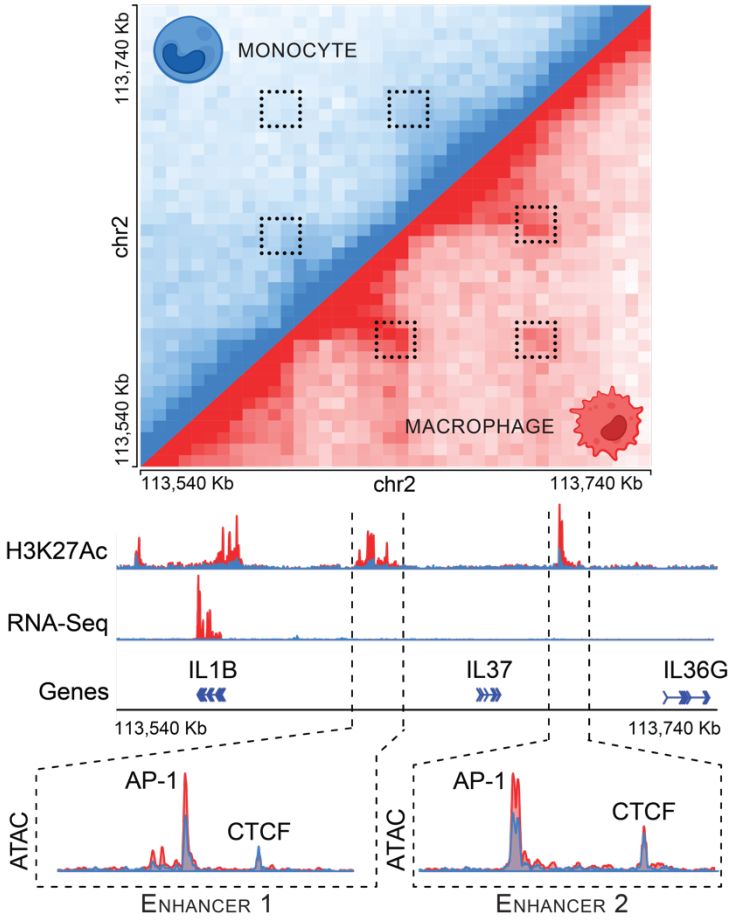

Figure 2.1. Novel loops formed at the *IL1β* locus during macrophage development identify upstream enhancers. Adapted from Phanstiel et al. 2017. Hi-C identified two loops formed at the *IL1β* locus in macrophages (red, heat map) but not in monocytes (blue, heat map). The presence of H3K27ac (as determined by ChIP-seq) at loop anchors (H3K27ac, red tracks), *IL1β* expression (RNA-seq, red tracks), and AP-1 binding motifs (ATAC, red tracks) in macrophages indicates that the *IL1β* promoter is connected via DNA loops to two upstream enhancers, which contribute to *IL1β*s transcription.

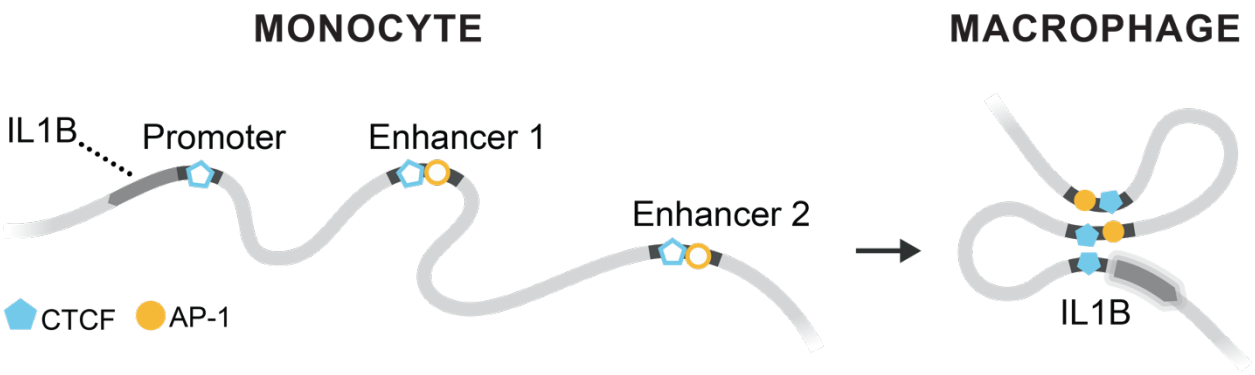

Figure 2.2. Cartoon depicting looping at the *IL1β* locus before and after macrophage differentiation. We hypothesize that DNA loops form at the *IL1β* locus during macrophage development; CTCF binding (blue) promotes loop formation, bringing the enhancer to the promoter, and the resulting proximity of AP-1 (yellow) to the promoter enhances transcription of *IL1β*.

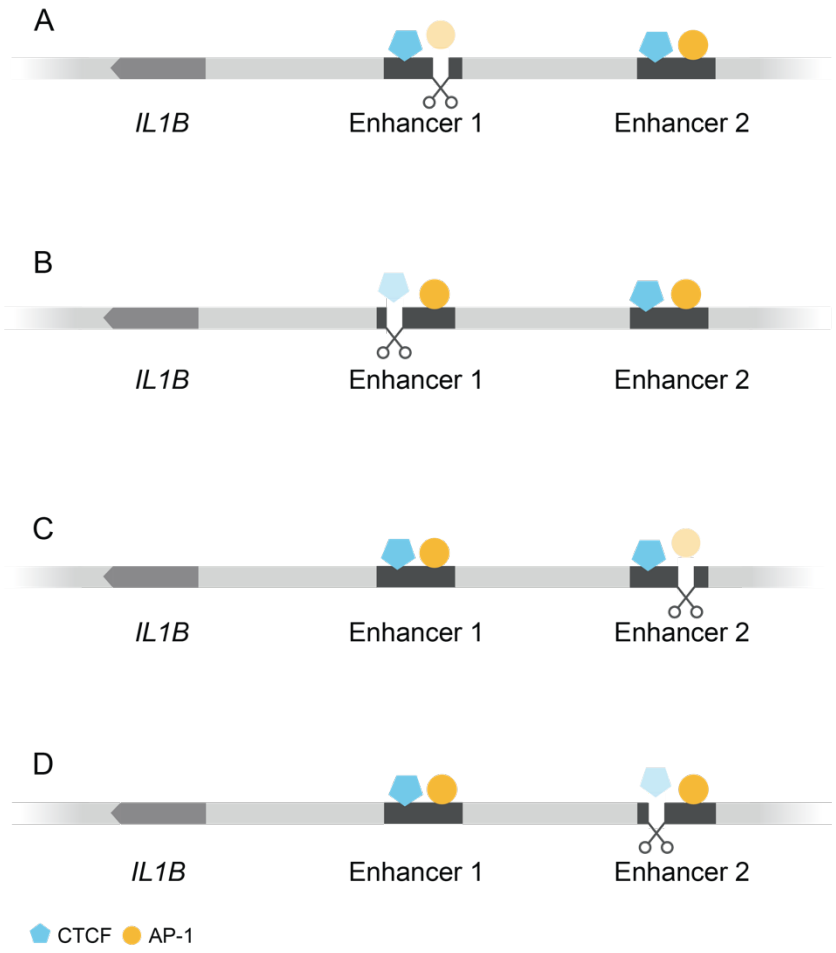

Figure 2.3. Cartoon depicting CRISPR-Cas9 sites at the IL1β upstream enhancers. Guide RNAs were designed to target CTCF (blue) and AP-1 (yellow) binding sites at the *IL1β* enhancer 1 and enhancer 2. The scissors represent Cas9 cleavage sites.

Table 2.1. Sequence and genomic coordinates of crRNAs designed to target AP-1 and CTCF binding sites at the *IL1β* enhancers. crRNAs were designed using https://chopchop.cbu.uib.no (Montague *et al.* 2014; Labun *et al.* 2016, 2019) and ordered via IDT's Alt-R Custom crRNA Design Tool.

Target	Left crRNA		Right crRNA	
	Sequence (5′→3′)	**Genomic coordinates**	**Sequence (5′→3′)**	**Genomic Coordinates**
Enhancer 1 AP-1 binding site	CGTGTGCCCCTTTACCGAGTGGG	chr2:113636616-113636639	CTTACGACATCACTAAGTACTGG	chr2:113639285-113639308
Enhancer 1 CTCF binding cite	CAGGGACCTACGGTTTGTTCGTTTTAGAGCTATGCT	chr2:113641648-113641671	GTGGAAGCAATAAATCGACAGTTTTAGAGCTATGCT	chr2:113642836-113642859
Enhancer 2 AP-1 binding site	TTATAGTGGCTTAGACTGAAAGG	chr2:113693365-113693388	GGCAAAGACTAGGTTAAGTCAGG	chr2:113694387-113694410
Enhancer 2 CTCF binding site	TTTCCACACGGGCTCATAAAAGG	chr2:113698916-113698939	TCCAATGACCTCCCGTTGCCTGG	chr2:113700430-113700453

Table 2.2. Sequence and genomic coordinates of PCR primers surrounding AP-1 and CTCF binding sites at *IL1β*'s upstream enhancers. PCR primers were designed using IDT's PrimerQuest™ Tool (https://www.idtdna.com/pages/tools/primerquest).

Target	Forward primer			Reverse primer		
	Strand	Sequence (5′→3′)	Genomic coordinates	Strand	Sequence (5′→3′)	Genomic Coordinates
Enhancer 1 AP-1 binding site	+	ATGCCACTTGAGTGTCCCTACA	chr2:113635875-113635897	-	AATCACTGTGCCTGAAACAGGG	chr2:113640291-113640313
Enhancer 1 CTCF binding cite	+	CCAATCAGATGGATATTGAGGATAAGGG	chr2:113641538-113641566	-	CCCAACCATCTGAATGGACCC	chr2:113643842-113643863
Enhancer 2 AP-1 binding site	+	AAGGAATAGGAGGGCCTTAGA	chr2:113693199-113693220	-	CATCTCTTCTTCCAGGTGCTC	chr2:113694493-113694514
Enhancer 2 CTCF binding site	+	CACGTCTTCACAAGGAGCTGAAA	chr2:113698377-113698400	-	TCCATGAGCATGTCAGTTGCAC	chr2:113700646-113700668

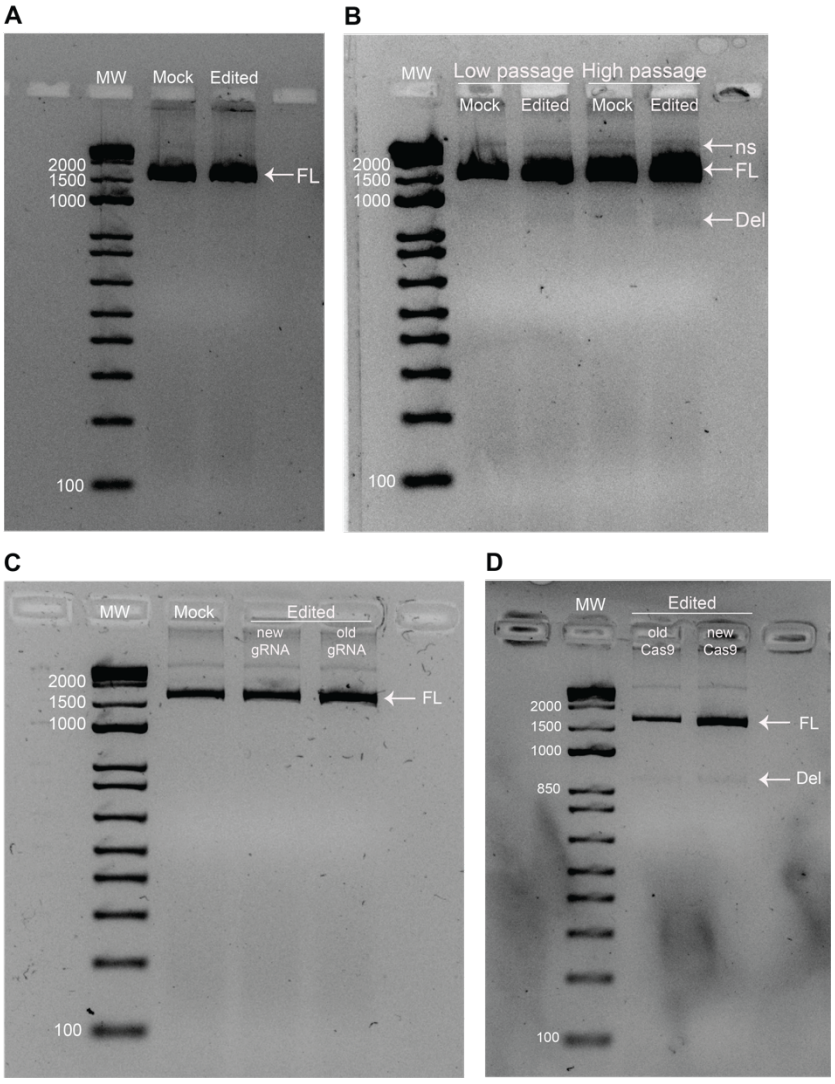

Figure 2.4. Low editing efficiency of THP-1 cells following RNP electroporation. (A) Gel electrophoresis following PCR reveals low editing efficiency. Trouble-shooting editing efficiency included editing with: **(B)** Lower and higher passaged THP-1s, **(C)** newly complexed gRNAs and previously complexed ('old') gRNAs, **(D)** previously used ('old') and fresh ('new') Cas9 enzymes. Following RNP electroporation of THP-1 cells targeting the *IL1β* enhancer 1 CTCF binding site, genomic DNA was extracted, PCR amplified, and visualized on a 2% agarose gel. Mock cells ('mock') were treated as edited cells ('edited') cells were during RNP electroporation but were not mixed with CRISPR-Cas9 components. ns: non-specific band. FL: full-length PCR product. Del: deletion PCR product. MW: molecular weight marker.

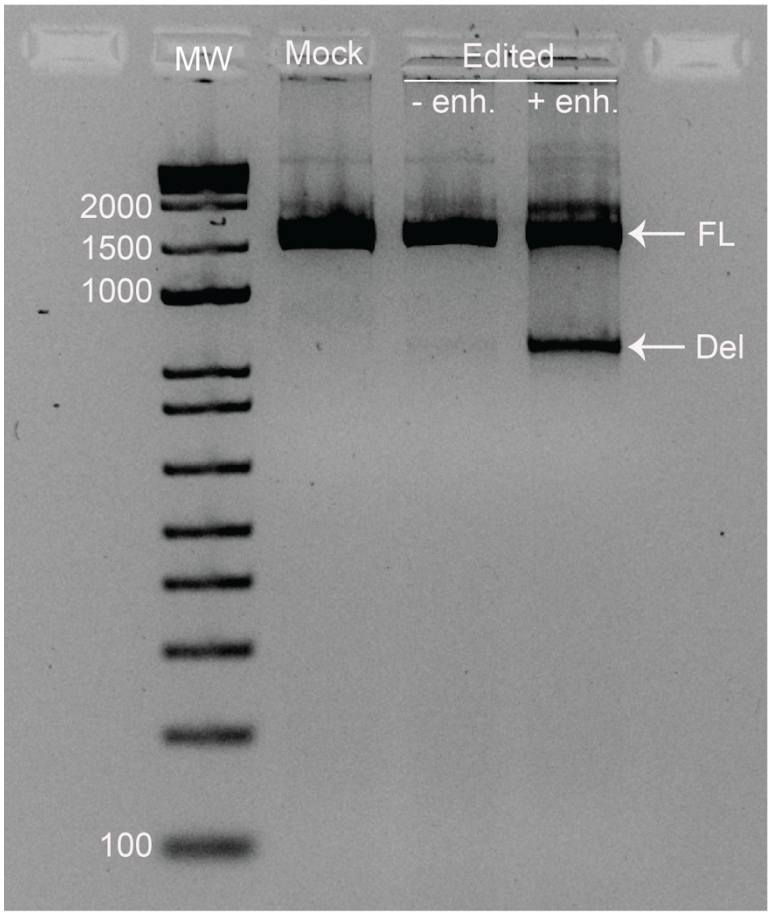

Figure 2.5. The addition of Cas9 Electroporation Enhancer during RNP electroporation increases editing efficiency of THP-1 monocytes. Following RNP electroporation of THP-1 cells targeting the *IL1β* enhancer 1 CTCF binding site, genomic DNA was extracted, PCR amplified, and visualized on a 2% agarose gel. Mock cells ('mock') were treated as edited cells ('edited') cells were during RNP electroporation but were not mixed with CRISPR-Cas9 components. '+ enh.' and '- enh.' indicates whether or not the Cas9 Electroporation Enhancer was included during RNP electroporation. The full-length ('FL') and deletion ('del') PCR products are indicated with white arrows. The darker band from the edited colonies with electroporation enhancer added indicates a higher editing efficiency. MW: molecular weight marker.

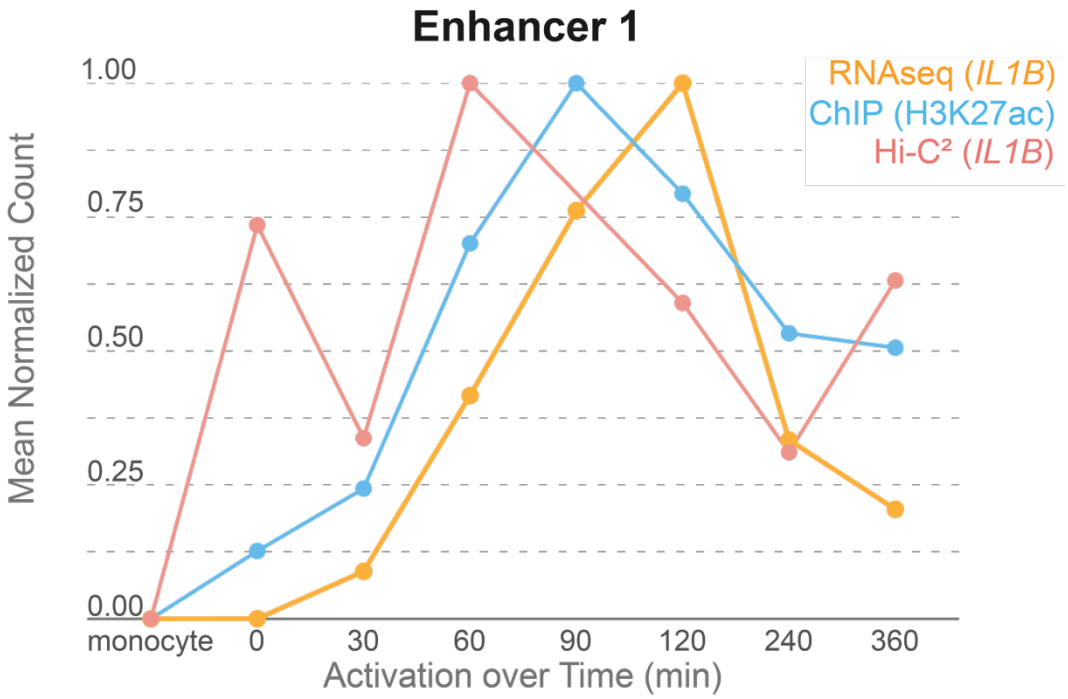

Figure 2.6. Looping at the *IL1β* locus precedes enhancer activation and gene expression. Hi-C^2 (pink), RNA-seq (yellow), and ChIP-seq for H3K27ac (blue) data were collected from monocytes, macrophages, and LPS/IFγ-activated macrophages at 30, 60, 90, 120, 240, and 360 minutes during treatment. Original Hi-C libraries for Hi-C^2 and RNA- and ChIP-seq libraries were generated by Kathleen Reed.

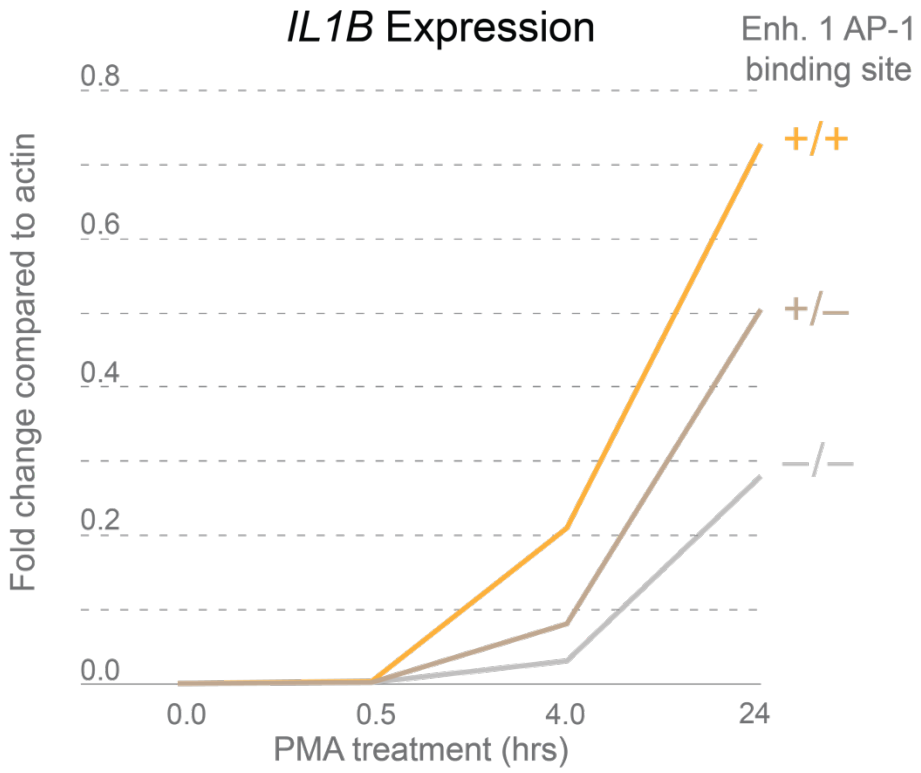

Figure 2.7. *IL1β* expression in macrophages is partially dependent on AP-1 binding. RT-qPCR was used to measure fold-change of *IL1β* expression over 24 hours of PMA treatment in three enhancer 1 AP-1 binding site deletion clones (+/+ = no deletion, +/- = heterozygous deletion, -/- = homozygous deletion). Cells were treated with 100nM PMA for 24 hours and RNA was extracted at 0, 0.5, 4, and 24 hours. n=1. Preliminary data was collected by Kamisha Woolery.

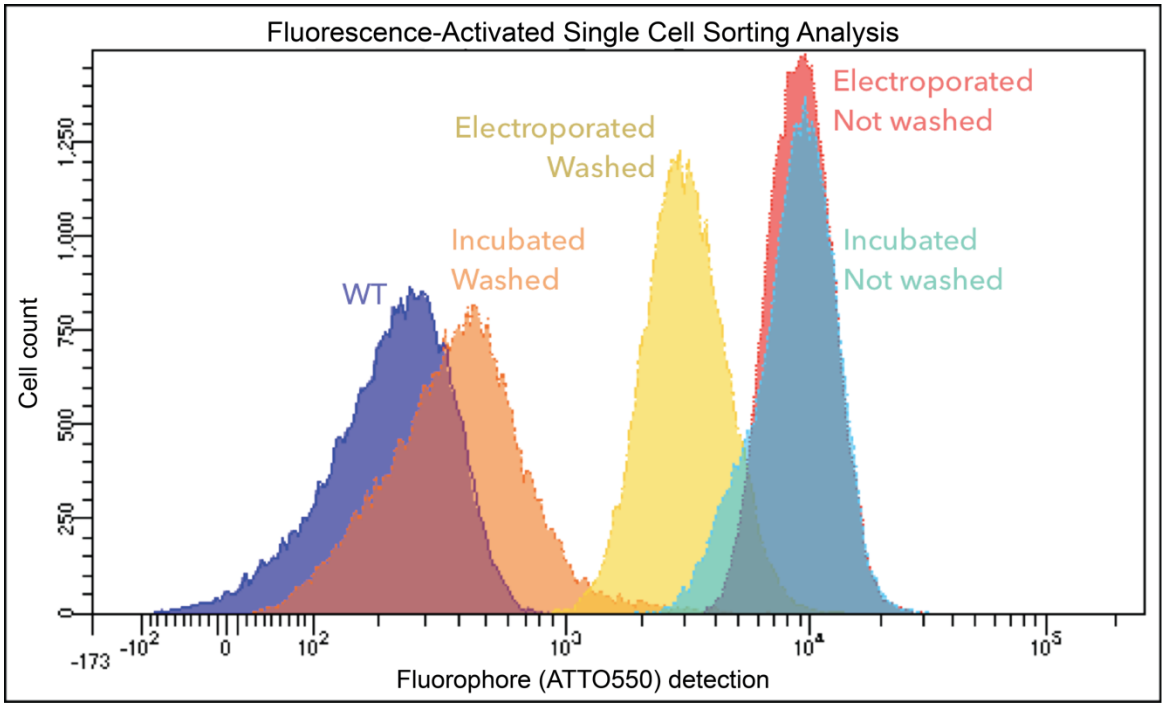

Figure 2.S1 FACS analysis identifies separate control and experimental populations of THP-1 cells. WT (wild-type, purple); WT THP-1s were taken straight from cell culture. Incubated, washed (orange); THP-1 cells were incubated with CRISPR-Cas9 components, but not electroporated, and were washed with PBS containing 1% FBS prior to cell sorting. Electroporated, washed (yellow); THP-1 cells were electroporated with CRISPR-Cas9 components (with either targeting or non-targeting guide RNAs) and were washed with PBS containing 1% FBS prior to cell sorting. Electroporated, not washed (red); THP-1 cells were electroporated with CRISPR-Cas9 components (with either targeting or non-targeting guide RNAs) and were not washed with PBS containing 1% FBS prior to cell sorting. Incubated, not washed (blue); THP-1 cells were incubated with CRISPR-Cas9 components, but not electroporated, and were not washed with PBS containing 1% FBS prior to cell sorting.

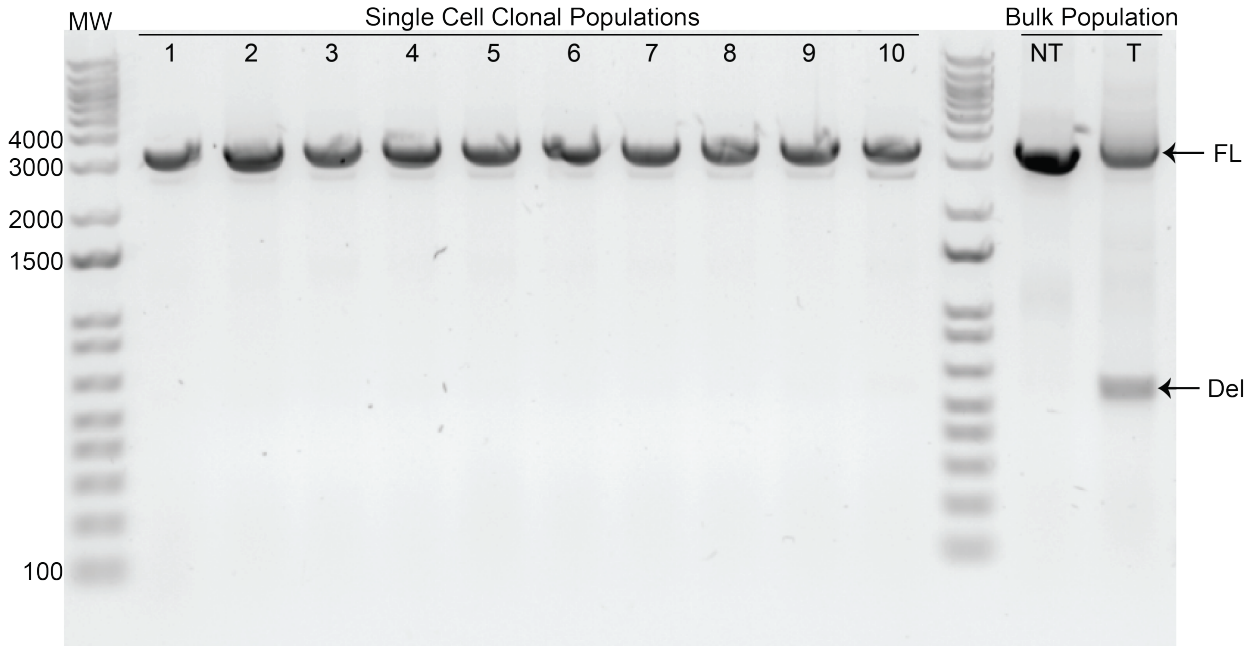

Figure 2.S2. Genotyping following single cell sorting via FACS reveals low editing efficiency. Following RNP electroporation of THP-1 cells targeting the *IL1β* enhancer 1 AP-1 binding site, single cells were isolated using FACS. After growth for approximately 4-6 weeks, genomic DNA was extracted, PCR amplified, and visualized on a 2% agarose gel. For comparison, DNA was also extracted and amplified from bulk populations edited with either non-targeting (NT) or targeting (T) gRNAs. The full-length ('FL') and deletion ('del') PCR products are indicated with black arrows. MW: molecular weight marker.

CHAPTER 3: CHROMATIN STRUCTURE IN CHONDROCYTES IDENTIFIES PUTATIVE OSTEOARTHRITIS RISK GENES[1]

3.1 Contributions

While I was the primary lead, this work is mainly the combined effort of myself and my co-authors, Eric S. Davis and Susan D'Costa. I planned and carried out experimental methods such as Hi-C and Cut and Run. E.S.D. performed Hi-C and Cut and Run data processing and multi-omic data integration. S.D. performed chondrocyte genome editing, transfection, treatment, validation, and Western Blot analysis. Philip R. Coryell performed chondrocyte isolation, culture, and treatment. Nicole E. Kramer contributed to genomic data analysis. Together, I, E.S.D., and S.D. wrote and contributed to the final version of the manuscript. All authors provided feedback and helped shape research and analysis. Douglas H. Phanstiel supervised the project.

3.2 Introduction

Osteoarthritis (OA) affects over 300 million people worldwide, yet treatment options are limited in large part because the mechanisms driving OA are not fully understood (Hunter and Bierma-Zeinstra 2019; Boer *et al.* 2021). Genome-wide association studies (GWAS) have identified over 100 loci associated with OA risk (Reynard and Barter 2020), but translating these broad loci into therapeutic targets has been challenging for several reasons. First, the effects of disease-associated variants are likely cell-type and context specific (Umans *et al.* 2021); therefore, studying these variants in the correct system that mimics the OA phenotype is required. Second, linkage disequilibrium (LD) between nearby variants makes it difficult to identify the causal variant(s) at each locus. Finally, because the majority of OA risk variants occupy non-coding regions of the human genome and can regulate genes up to a million base pairs away, the genes impacted by most OA risk variants are unknown.

Several studies have successfully used genomic and bioinformatic techniques to identify the

[1]The work in this chapter has been previously published. The citation is: Thulson E., E.S. Davis, S. D'Costa, P.R. Coryell, N.E. Kramer, K. L. Mohlke, R. F. Loeser, B. O. Diekman, D. H. Phanstiel. 2022 3D Chromatin structure in chondrocytes identifies putative osteoarthritis risk genes. Genetics 222. https://doi.org/10.1093/genetics/iyac141

genes impacted by gene-distant non-coding GWAS variants for a variety of disease phenotypes (Claussnitzer *et al.* 2015; Won *et al.* 2016; Chesi *et al.* 2019; Laarman *et al.* 2019; Duan *et al.* 2021). Mapping regulatory loci using ChIP-seq, ATAC-seq, or CUT&RUN and intersecting the resulting data with disease-associated variants can identify a short list of putative causal variants. These variants can then be linked to potential target genes by quantifying 3D chromatin contacts using Hi-C or other chromatin conformation capture (3C) techniques. For example, chromatin interaction data was used to determine that an obesity-associated variant located in an intron of the *FTO* gene affects expression of the downstream genes *IRX3* and *IRX5*, which are involved in obesity-related biological processes (Claussnitzer *et al.* 2015). Likewise, Hi-C in human cerebral cortex identified FOXG1 as a distal target of a schizophrenia GWAS variant, supporting its potential role as a schizophrenia risk gene (Won *et al.* 2016).

Because the effects of disease-associated variants are likely limited to particular biological states (Umans *et al.* 2021), studies of their impact must be conducted in the correct cellular and biological context. Several pieces of evidence suggest that chondrocytes—particularly those responding to cartilage matrix damage—are one of the most likely cell types to be affected by OA risk variants. Cartilage breakdown and loss is a primary feature of OA. Chondrocytes are the only cell type found in cartilage and are responsible for maintaining the cartilage matrix. Osteoarthritic cartilage harbors activated chondrocytes that exhibit a proinflammatory phenotype thought to contribute to progressive cartilage degradation, which includes production of bioactive matrix fragments (Loeser 2014; van den Bosch *et al.* 2020). We have developed an *ex vivo* system that simulates the OA chondrocyte phenotype by treating primary human articular chondrocytes with fibronectin fragment (FN-f) (Forsyth *et al.* 2002; Pulai *et al.* 2005; Wood *et al.* 2016; Reed *et al.* 2021). Fibronectin is a ubiquitous extracellular matrix protein, and high levels of FN-f are present in cartilage and synovial fluid of OA joints (Xie *et al.* 1992; Homandberg *et al.* 1998). Subsequently, FN-f has been shown to be an OA mediator that recapitulates gene expression changes associated with OA (Homandberg 1999; Forsyth *et al.* 2002; Pulai *et al.* 2005; Reed *et al.* 2021). We have leveraged this model of OA for use in clonal populations of genome-edited primary human chondrocytes, allowing us to quantify the phenotypic impact of putative target genes of genomic variants in an appropriate disease context.

In this study, we generated CUT&RUN data in primary human chondrocytes and Hi-C data in a human chondrocyte cell line and intersected them with publicly available RNA-seq data from our *ex vivo* OA model, ChIP-seq data from the Roadmap Epigenomics project, and OA GWAS variants from Boer et al. In doing so, we identified 56 putative OA risk genes, including *SOCS2*, whose promoter loops to an OA GWAS variant ~174 Kb away. Deletion of *SOCS2* in primary human chondrocytes using CRISPR-Cas9 led to heightened expression of inflammatory markers in response to treatment with FN-f , providing a possible mechanism for influencing OA risk.

3.3 Results

3.3.1 OA risk variants are enriched in chondrocyte regulatory loci

One of the first steps in decoding GWAS variant mechanisms is to determine the cell types that are likely mediating genetic OA risk. While different risk variants may impact distinct cell types, one approach to help direct research is to determine the cell types which harbor regulatory loci (e.g. enhancers) that are enriched for risk variants. To accomplish this, we performed SNP enrichment analysis using the Genomic Regulatory Elements and Gwas Overlap algoRithm (GREGOR) (Schmidt *et al.* 2015). Publicly available H3K27ac, H3K4me1, and H3K4me3 ChIP-seq peaks from the NIH Roadmap Epigenomics Mapping Consortium (Roadmap) were merged to define regulatory elements for 98 cell types. GREGOR was used to determine each cell type's enrichment for 104 OA GWAS signals recently published in Boer et al. (Boer *et al.* 2021)

The regulatory elements of "Chondrocytes from Bone Marrow Derived Mesenchymal Stem Cell Cultured Cells" (E049) exhibited a strong effect size and p-value of enrichment for OA risk variants (**Fig. 3.1A**), suggesting that many OA risk variants may impact regulatory events in chondrocytes. This is consistent with the known role of chondrocytes in maintaining joint homeostasis. Chondrocytes have been heavily implicated in OA, as activation of chondrocytes by mechanical and inflammatory stimuli triggers downstream inflammatory and catabolic response pathways in diseased tissue (Sandell and Aigner 2001; Pelletier *et al.* 2001; Loeser *et al.* 2012; Caron *et al.* 2015). An example of an OA risk variant that overlaps a chondrocyte-specific regulatory element (H3K27ac peak) is shown in **Figure 3.1B**. For comparison, **Figure 3.1C** shows a non-OA associated variant that overlaps a non-cell type specific, or ubiquitous, enhancer on chromosome 10 that is active in >90% of the 98 cell types evaluated. These

examples underscore the importance of interpreting GWAS risk variants in light of the correct cellular context, as the variant-H3K27ac peak overlap shown in **Figure 3.1B** would not have been detected in any of the other cell types investigated. In addition to E049, IMR90 fetal fibroblasts (E017) and HSMM cell derived Skeletal Muscle Myotubes (E121) were also enriched, suggesting that OA risk variants may also contribute to disease risk through altering the function of fibroblasts and muscle. However, given the strong enrichment in chondrocytes and their documented role in OA biology, we chose to focus our investigation of OA GWAS variants in human chondrocytes.

3.3.2 Multi-omic integration identifies putative variant-gene associations in OA

Due to high LD between variants and the fact the most risk variants reside in non-coding sequences, determining the causal variants and genes they impact remains a major challenge. To address these issues, we generated novel maps of epigenetic features in human chondrocytes and integrated them with GWAS results and publicly available genomic datasets to identify putative variant-gene associations for OA.

First, we identified OA risk variants that are predicted to directly affect protein sequences. We used ENSEMBL's Variant Effect Predictor (VEP) tool to predict the consequences of 1,259 putative OA risk variants that were in high LD ($r^2 > 0.8$) with 104 OA GWAS signals from Boer et al. (Boer *et al.* 2021). VEP identified 29 variants at 19 loci predicted to affect the coding sequence of 24 unique genes (**Fig. 3.2A, top**). 18 of these variants encode a missense mutation impacting 17 genes, while 11 variants encode a synonymous mutation impacting 8 genes (**Fig. 3.2A, top**). Though synonymous variants do not impact the protein sequence directly, differences in transcription efficiency, tRNA availability, and mRNA stability introduced through these variants could contribute to the OA phenotype (Venetianer 2012; Zeng and Bromberg 2019). To identify genes most likely to impact OA risk, we incorporated our previously published RNA-seq FN-f time course data to find genes that change expression in an OA-context. Of the 24 genes identified here, 6 exhibited differential expression in response to FN-f, (**Fig. 3.2A bottom**). Several of the genes identified have been previously implicated in OA, including Interleukin 11 (*IL11*), Solute Carrier Family 39 Member 8 (*SLC39A8*/*ZIP8*), and Serpin Family A Member 1 (*SERPINA1*). IL11 plays a role in bone turnover and is upregulated in subchondral bone and articular cartilage from OA tissue (Tuerlings *et al.* 2021). SLC39A8 is upregulated in OA chondrocytes and suppression of SLC39A8

in a mouse OA model significantly reduces cartilage degradation (Song *et al.* 2013). SERPINA1, a serine protease inhibitor with anti-inflammatory capabilities (Jain *et al.* 2011), is downregulated in OA (Boeuf *et al.* 2008; Wanner *et al.* 2013).

Next, we identified OA risk variants that could impart their phenotypic impact by altering promoter and/or enhancer activity. In order to define the most accurate regulatory regions in chondrocytes we used CUT&RUN to map histone H3K27 acetylation (H3K27ac)—a mark of active enhancers and promoters—in primary human chondrocytes isolated from the knees of 2 cadaveric donors. As expected, Roadmap chondrocyte (E049) H3K27ac peaks showed the highest degree of similarity by jaccard distance to H3K27ac peaks called in our primary human chondrocyte data (**Fig. 3.S1**). Therefore, we merged our primary chondrocyte H3K27ac peaks with all available marks (including active and repressive marks) from the E049 chondrocyte cell line from Roadmap Epigenomics to define a comprehensive set of chondrocyte regulatory elements. Integration of public and newly generated sources of human chondrocyte features allowed us to identify 507 plausible regulatory variants from 1,259 OA risk variants. Intersecting these 507 plausible regulatory variants with gene annotations (UCSC) identified 26 unique variants that overlapped the promoters of 21 genes (**Fig. 3.2B**). Two of these genes were differentially expressed in response to FN-f, both of which have been previously implicated in OA. Growth and Differentiation Factor 5 (GDF5), a member of the TGF-beta family, has roles in skeletal and joint development (Francis-West *et al.* 1999) and has been identified as a major risk locus for OA (Miyamoto *et al.* 2007; Southam *et al.* 2007). Specifically, variants in the GDF5 enhancers *R4* and *GROW1* have been associated with altered anatomical features of the knee and hip, which are thought to confer an increased risk of OA (Capellini *et al.* 2017; Richard *et al.* 2020; Muthuirulan *et al.* 2021). Solute Carrier Family 44 Member 2 (SLC44A2, aka choline transporter-like protein 2) is a mitochondrial choline transporter that has been identified as an expression quantitative trait locus (eQTL) in OA tissue (Steinberg *et al.* 2021) that colocalizes with the OA GWAS signal rs1560707 (Steinberg *et al.* 2020).

In addition to direct regulation of genes by their promoters, long-range regulation of genes also occurs via enhancer-promoter interactions mediated by chromatin loops (Maurano *et al.* 2012). To identify such connections, we conducted deeply sequenced (~2.8 billion reads) in situ Hi-C in C-28/I2 chondrocyte cells and identified 9,271 chromatin loops with Significant Interaction Peak (SIP) caller at 5-

Kb resolution which we expanded to 20-Kb for downstream analysis (Rowley *et al.* 2020). C-28/I2 cells were used because they could be expanded to easily provide the number of cells required for Hi-C analysis. To our knowledge, this is the first Hi-C map in a chondrocyte cell line, enabling us to discover novel OA-associated variant-gene connections. We performed four replicates which exhibited a high degree of reproducibility as measured by stratum adjusted correlation coefficient (SCC > 0.98) with the *HiCRep* package (**Fig. 3.S2**) (Yang *et al.* 2017; Lin *et al.* 2021). Overlapping these data with OA risk variants identified 14 loops connecting 47 variants among 14 loci to 20 unique gene promoters (**Fig. 3.2C**). Four of these genes were differentially expressed in response to FN-f ($p \leq 0.01$, absolute log2 fold-change ≥ 1.25) and are visualized in Figures 2A and S3. Several of these genes have interesting implications for OA, including *FGFR3* (Fibroblast Growth Factor Receptor 3), which plays a role in skeletal development. FGFR3 may have an important function in the maintenance of articular cartilage (Zhou *et al.* 2016; Tang *et al.* 2016; Okura *et al.* 2018), possibly through the Indian hedgehog signaling pathway, which plays a role in regulating chondrocyte hypertrophy and the expression of cartilage matrix-degrading enzymes (Lin *et al.* 2009). FGFR3 is also downregulated in OA tissues, further implicating its potential role in limiting articular cartilage degeneration (Li *et al.* 2012; Shu *et al.* 2016).

Altogether we identified 24 genes impacted by a coding variant, 21 genes with at least one regulatory variant in their promoters, and 20 genes that were connected to a regulatory variant via a chromatin loop. Since genes can fall into multiple categories, the number of total distinct genes identified is 56. All putative variant-gene associations are reported in **Supplementary Table 1** (viewable in the online version of the paper). Boer et al. identified 637 putative effector genes and ranked them by the amount of evidence for association with OA signals (Boer *et al.* 2021). In general, genes with higher tiers of evidence as reported by Boer et al. were more likely to be supported by our analyses (**Fig. 3.S4; Supplementary Table 2**, viewable in the online version of the paper). For example, 67% genes that were supported by six tiers of evidence were also detected in our study, whereas only 2.5% of tier 1 genes were supported by our work. Interestingly, 42% of the genes we identified were not previously implicated by Boer et al. 54% of the genes unique to our study were supported by a chromatin loop compared to only 22% of genes implicated by both studies. This underscores the additional value our study provided by incorporating cell type specific Hi-C data.

3.3.3 Chondrocyte chromatin features identify SOCS2 as a putative regulator of OA

Our multi-omic analysis identified an association between rs7953280 and the promoter of Suppressor Of Cytokine Signaling 2 (*SOCS2).* rs7953280 is located in an intron of the *CRADD* gene, which is expressed at low levels in primary chondrocytes, does not change expression in response to FN-f, and lacks an obvious biological relevance to OA. However, rs7953280 overlaps a putative chondrocyte enhancer (i.e. histone H3K27ac peak), suggesting that it could alter the regulatory capacity of the enhancer and impact the expression of a proximal or distal gene. This enhancer is connected to the promoter of *SOCS2* via a 174 Kb chromatin loop (**Fig. 3.3A**). Unlike *CRADD*, *SOCS2's* expression changes in response to FN-f, peaking at 3 hours (**Fig. 3.2C** and **3.3A orange signal tracks**). Moreover, SOCS2 is known to play a role in resolving inflammatory response through NFKB and is downregulated in knee OA tissues (de Andrés *et al.* 2011; Paul *et al.* 2017) , making it an intriguing candidate as an OA risk gene. No other SNPs from this locus can be assigned to genes using our integrated approach.

To further understand how rs7953280 may confer risk for OA, we examined the sequence surrounding rs7953280 to see if it overlaps and alters any transcription factor (TF) binding motifs. Motif comparison with Tomtom from the MEME suite identified FOS and JUN as matching target motifs (**Fig. 3.3C, Supplementary Table 3**, viewable in the online version of the paper). FOS and JUN are members of the Activator Protein 1 (AP-1) complex, which is upregulated in response to FN-f (Reed *et al.* 2021), and the inhibition of which prevents cartilage degradation in a model of OA (Motomura *et al.* 2018; Fisch *et al.* 2018; Gao *et al.* 2019). We then used SNP effect matrices (SEMs) generated by the SNP effect matrix pipeline (SEMpl) (Nishizaki *et al.* 2020) to assess the predicted consequence of the G (non-risk) to C (risk) variant on binding of JUN or any other of the 211 motifs included with SEMpl (**Fig. 3.3D**). Most TFs are predicted to be unbound at both alleles. However, multiple JUN/AP-1 motifs are predicted to bind to the non-risk, but not the OA-risk sequence (**Fig. 3.3D**) providing further evidence that the G->C mutation in rs7953280 may disrupt JUN/AP-1 binding. Our analysis also showed that STAT-1 was predicted to bind only to the OA-risk sequence, although the SEM score was very close to the cutoff for predicted binding. Nevertheless, since STAT-1 is an important mediator for inflammatory signaling, rs7953280 could influence inflammation during OA progression by modulating STAT-1 binding.

3.3.4 SOCS2 deletion increases proinflammatory gene expression in response to FN-f

To assess the functional role of SOCS2, we used CRISPR-Cas9 to knock out *SOCS2* in primary human chondrocytes isolated from three individual donors. After targeting the *SOCS2* gene with two guide RNAs that flank exon 2 (a constitutive exon that contains the translational start site), we used our previously developed method that employs PCR to screen single-cell-derived colonies (D'Costa *et al.* 2020). The screening primers generated a 1068 bp product if the region was intact and a novel 240 bp amplicon if the two guides successfully deleted the intended 828 bp region (**Fig 3.4A**). We saw efficient deletion in each of the three donors, with 31% of the colonies showing no deletion, 49% of the colonies showing heterozygous deletion, and 20% showing homozygous deletion (**Fig. 3.4B**). Sanger sequencing was used to confirm deletions, while qPCR and western blotting confirmed partial (heterozygous) or complete (homozygous) loss of SOCS2 expression (**Fig. 3.S5**).

Because SOCS2 is a known negative regulator of the inflammatory response in other settings (Paul *et al.* 2017; Monti-Rocha *et al.* 2018), we hypothesized that *SOCS2* deletion would lead to an increased expression of inflammatory cytokines in chondrocytes during the response to FN-f. To test this hypothesis, we treated chondrocytes with defined genotypes (wild type, heterozygous, or homozygous knockout) with either FN-f or PBS for 18 hours and quantified the change in proinflammatory cytokines C-C Motif Chemokine Ligand 2 (*CCL2*) and Interleukin 6 (*IL6*) using qPCR. *IL6* and *CCL2* have previously been shown to exhibit increased expression after 18 hours of FN-f treatment and are also implicated in OA (Wojdasiewicz *et al.* 2014; Wang and He 2018; van den Bosch *et al.* 2020; Reed *et al.* 2021). Deletion of *SOCS2* led to increased expression of both *IL6* and *CCL2* in response to FN-f treatment (**Fig. 3.4C**), and these increases were observed in a dose-dependent fashion, with greater increases observed in the homozygous compared to heterozygous genotypes. These results suggest that the loss of *SOCS2* may promote a heightened inflammatory response to FN-f stimulation, which is consistent with a potential role in OA.

3.4 Discussion

We used a multi-omic approach to identify putative causal SNPs and genes associated with OA risk. The efficacy of this approach was supported by the identification of previously known OA risk genes including *GDF5*, *SLC44A2*, and *IL11*. We generated the first maps of histone H3K27ac in primary human chondrocytes and integrated this dataset with publicly available genomic datasets to reduce thousands of

OA risk GWAS variants to a small list of variants and genes for further study. By generating the first Hi-C contact map of human chondrocytes, we were able to uncover 73 previously unknown connections between OA risk variants and putative target genes. Most looped variant-gene pairs (71 of 73) skipped over the nearest gene, connecting variants to genes as far as 414 Kb away. DNA looping revealed 20 unique genes, 13 of which were not identified by recent fine mapping approaches (Boer *et al.* 2021) and could provide new avenues for therapeutic interventions for OA.

Among the genes identified with Hi-C, four were found to be differentially expressed in our OA model. FGFR3 and SOCS2 have previously been implicated in OA, while Tropomoyosin 1 (TPM1) and Ral Guanine Nucleotide Dissociation Stimulator Like 1 (RGL1) have not. However, TPM1, an actin-binding protein involved in the contractile system of muscle cells and the cytoskeleton of non-muscle cells, has been shown to play roles in an inflammatory response in various cell types, such as human primary coronary artery smooth muscle cells (Li *et al.* 2022) and rod bipolar and horizontal cells in the retina (Gagat *et al.* 2021). RGL1, which functions as a RAS effector protein that activates GTPase by stimulating nucleotide exchange, has also been shown to modulate immune response in both vascular and immune cells (Kirkby *et al.* 2014), and, interestingly, is downregulated in human articular chondrocytes upon treatment with interleukin-1 and oncostatin-M (Yang *et al.* 2021). The functions of TPM1 and RGL1 in inflammatory responses may point to potentially undiscovered roles in osteoarthritis. One especially intriguing gene was *SOCS2*, whose promoter is looped to an OA risk SNP within a histone H3K27ac peak ~170 Kb away. SOCS2 is known to inhibit the JAK/STAT pathway and is induced by various pro-inflammatory cytokines such as interleukin-6, growth hormone, and tumor necrosis factor-alpha (Starr *et al.* 1997; Metcalf *et al.* 2000; Santangelo *et al.* 2005). CRISPR-mediated deletion of *SOCS2* was associated with increased expression of *IL6* and *CCL2* in our *ex vivo* model of OA, suggesting that it may also play a role in mediating inflammation in response to cartilage matrix damage. These findings make SOCS2 a candidate for further studies and the activation of more robust SOCS2 expression could be a goal for future therapeutic development. The regulatory role of SOCS2 in chondrocytes is likely to be subtle, as *Socs2* knockout mice did not show altered OA development (Samvelyan *et al.* 2022). Because that study used a global germline deletion, other members of the

inflammatory cascade may have compensated for *Socs2* loss, and it would be interesting to determine whether the inducible loss of *Socs2* in adult chondrocytes would generate a different result.

Expression quantitative trait locus analysis (GTEx Project v8) provides evidence that variation at rs7953280 is associated with *SOCS2* expression in fibroblasts. In that data set the C allele is associated with increased expression of *SOCS2*. This is contrary to what we predict for chondrocytes but could be explained by differences in cell type or condition (e.g. resting vs stimulated with FN-f). Mapping of QTLs in chondrocytes responding to FN-f could shed light on these differences.

While further work is needed to clarify the role of rs7953280 and SOCS2 in mediating OA risk, our multi-omic analysis suggests the following potential model. In cells harboring the non-risk variant, proinflammatory cytokines such as IL-6 and matrix damage products such as FN-f may activate AP-1 via the JAK/STAT pathway. AP-1 may then bind the enhancer at the rs7953280 locus, increase enhancer activity, and upregulate transcription of *SOCS2* via a chromatin loop between the enhancer and the *SOCS2* promoter. In cells harboring the risk allele, AP-1 binding would be decreased, impeding enhancer activation and proper upregulation of *SOCS2*. As a result, JAK/STAT signaling would remain high, resulting in prolonged or heightened inflammation and further cartilage degradation.

This model, while compelling, will require further experimental investigation and validation. One such experiment would be to use genome editing of noncoding sequences to directly test the effect of rs7953280 on *SOCS2*. While implementing single base changes using CRISPR/Cas9 and homology-directed repair donor oligos in chondrocytes is technically challenging, engineering isogenic chondrocytes with the risk and protective alleles will help validate the association between the variant and *SOCS2* expression and the observed inflammatory response. Moreover, future experiments are required to determine to the degree to which these findings translate from our *ex vivo* model into an *in vivo* system and/or if activation of *SOCS2* could provide therapeutic avenue for OA treatment.

We generated the first maps of histone H3K27ac in primary human chondrocytes, provided the first maps of 3D chromatin contacts in chondrocytes of any type, and identified 56 putative OA risk genes using multi-omic data integration. For each locus, we provide 0, 1 or multiple putative OA risk genes. While these analyses narrow the search space for the genes affected by OA risk variants and allow for the formation of new hypotheses, determining which genes are truly causal will require further

experimental validation similar to the approach described here to investigate SOCS2. We chose to perform functional experiments on SOCS2 because it had the strongest genomic evidence for mediating OA risk of the genes looped to OA risk variants; however, many of the other genes implicated here (via looping or otherwise) may also influence disease risk and warrant further investigation. These putative risk genes and novel epigenetic datasets will provide a foundation for future studies to investigate the genetic variants responsible for OA risk and expedite our search for better prevention and treatment of OA.

3.5 Materials and Methods

3.5.1 Primary chondrocyte isolation and culture

Primary articular chondrocytes were isolated via enzymatic digestion from human talar cartilage obtained from tissue donors, without a history of arthritis, through the Gift of Hope Organ and Tissue Donor Network (Elmhurt, IL) as previously described (Loeser *et al.* 2003; Reed *et al.* 2021)**.** For Cut and Run, two million primary articular chondrocytes from two male donors, ages 39 and 63, were plated onto four 6cm plates in DMEM/F12 media supplemented with 10% fetal bovine serum, 1% penicillin streptomycin solution, 1% amphotericin B, and 0.04% gentamicin. For genome editing, primary chondrocytes from three male donors ages 56, 59, and 64 years were cultured in 6 or 10cm dishes at a density of approximately 70,000 cells/cm^2 in DMEM/F12 media supplemented with 10% FBS and antibiotics.

3.5.2 Fibronectin fragment (FN-f) treatment

After serum starvation, cells were treated with either purified 42 kDa endotoxin-free recombinant FN-f (final concentration 1μM in PBS), prepared as previously described, or PBS as control (Wood *et al.* 2016). Cells were harvested and crosslinked after 90 min or 180 min and immediately subjected to Cut & Run, described below.

3.5.3 Hi-C

C-28/I2 cells were cultured in DMEM/F12 media with 10% fetal bovine serum, 1% penicillin streptomycin solution, 1% amphotericin B, and 0.04% gentamicin. Cells were treated with DMEM/F12 media with 1% ITS-Plus for 48 hours prior to experiments to promote the chondrocyte phenotype. Cells were then washed with 1X PBS and treated with trypsin-EDTA (0.25%) for 3 minutes. Trypsin was

quenched and cells were pelleted at 4°C for 5 minutes at 300*g*. Cells were resuspended in 1mL DMEM/F12 per million cells and crosslinked in 1% formaldehyde for 10 min with rotation before quenching in a final concentration of 0.2M glycine for 5 min with rotation. Cells were pelleted at 300*g* for 5 min at 4°C. Pellets were washed with cold PBS and aliquoted into ~3 million cell aliquots. Pellets were flash frozen in liquid nitrogen and stored at -80°C. In situ Hi-C was performed as previously described (Rao *et al.* 2014). A full description of our methods is provided in the Supplemental Materials.

3.5.4 Hi-C data processing

In situ Hi-C datasets were processed using a modified version of the Juicer Hi-C pipeline (https://github.com/EricSDavis/dietJuicer) with default parameters as previously described (Durand *et al.* 2016). Reads were aligned to the hg19 human reference genome with bwa (v0.7.17) and MboI was used as the restriction enzyme. Four biological replicates were aligned and merged for a total of 2,779,816 Hi-C read pairs in C-28/I2 cells yielding 2,373,892,594 valid Hi-C contacts (85.40%). For visualization, the merged Hi-C contact matrix was normalized with the "KR" matrix balancing algorithm as previously described (Knight and Ruiz 2013) to adjust for regional background differences in chromatin accessibility. Looping interactions were called at 5-Kb resolution with Significant Interaction Peak (SIP) caller (Rowley *et al.* 2020) (v1.6.2) and Juicer tools (v1.14.08) using the replicate-merged, mapq > 30 filtered hic file with the following parameters: "-norm KR -g 2.0 -min 2.0 -max 2.0 -mat 2000 -d 6 -res 5000 -sat 0.01 -t 2000 -nbZero 6 -factor 1 -fdr 0.05 -del true -cpu 1 -isDroso false". Loop anchors were expanded to 20 Kb and loops with overlapping anchors were filtered out (14 loops). This resulted in 9,271 loops after filtering.

3.5.5 Cut and Run

Primary chondrocytes were washed with 1X PBS and treated with trypsin-EDTA (0.25%) for 5 minutes. Trypsin was quenched and cells were pelleted at 4°C for 5 minutes at 1000*g*. Cells were resuspended in 1mL plain DMEM per million cells and crosslinked in 1% formaldehyde for 10 min with rotation before quenching in a final concentration of 125 mM glycine for 5 min with rotation. Cells were pelleted by spinning at 1000*g* for 5 min at 4°C. Each 2 million cell pellet was washed in 1 mL cold PBS prior to flash freezing in liquid nitrogen. We performed Cut and Run following existing protocols (Skene and Henikoff 2017) but modified for crosslinked cells. A full description of our methods is provided in the Supplemental Materials.

3.5.6 Cut and Run data processing and peak calling

Adaptors and low-quality reads were trimmed from paired-end reads using Trim Galore! (v0.4.3). Reads were aligned to the hg19 genome with BWA mem (v0.7.17) and sorted with Samtools (v1.9). Duplicates were removed with PicardTools (v2.10.3) and mitochondrial reads were removed with Samtools idxstats. Samtools was also used to merge donors, and index BAM files. Peaks were called from the merged alignments using MACS2 with the following settings: -f BAM -q 0.01 -g hs --nomodel --shift 100 --extsize 200 --keep-dup all -B --SPMR (v2.1.1.20160309). Peaks were then merged using bedtools (v2.26), and multicov was used to extract counts from each replicate BAM file. Signal tracks were made from alignments using deeptools (v3.0.1).

3.5.7 Genome editing of chondrocytes

3.5.7.1 Preparation of gRNA: Cas9 RNP complex

Two custom SOCS2 Alt-R crRNAs TGACAAGGGCCTATTCCCAC and TTACGCATTCCCAAGGACCC were synthesized by Integrated DNA technologies (IDT). Both sequences are written 5’ to 3’ and do not include PAM sequence. The first crRNA targets the plus strand and the second the minus strand. Ribonucleoprotein (RNP) complexes containing the Cas9 enzyme and sequence-targeting guide RNAs were prepared according to the manufacturer's recommendation. Briefly, Alt-R tracrRNA (1072533, IDT) and crRNA were resuspended in Tris-EDTA buffer to 100 µM concentration and equimolar concentration of crRNA and tracrRNA was combined, heated at 95 °C for 5 min and cooled to room temperature to produce the gRNA. Separate RNP complex for each guide was prepared by combining the gRNA (50 µM) with Alt-R® Cas9 Nuclease (61 µM) (1081058, IDT) and PBS at a ratio of 1:1.1:2 µl at room temperature for 15 min.

3.5.7.2 Transfection of primary human chondrocytes with RNP complex and single cell colony selection

Chondrocytes were trypsinized, washed with PBS and transfected with the RNP complex as previously described with modifications; volumes were scaled up for transfection of more cells in larger cuvettes (D’Costa *et al.* 2020). Two million cells were resuspended in 100 µl of P3 Primary Cell Nucleofector™ solution (V4XP-3024, Lonza). The RNP complex and Alt-R® Cas9 Electroporation Enhancer (1075916, IDT) was added to the cells. The mixture was gently pipetted up and down and transferred to 100 µl Nucleocuvette vessels (V4XP-3024, Lonza) and transfected using program ER-100

on a 4D-Nucleofector™ Core unit (Lonza). Cells were kept at room temperature for 8 minutes and then incubated in prewarmed antibiotic free media containing 20% FBS for recovery. An aliquot of the transfected cells was placed in a 96-well and used for DNA extraction and PCR. Following confirmation of editing, the transfected bulk cells were seeded at low cell density (200 cells per 6 cm^2 dish) for generation of single-cell colonies. Individual colonies were picked under a microscope (EVOS FL, ThermoFisher), the colony was disrupted by pipetting and split into 96- and 24-well plates for genetic analysis and continued expansion, respectively.

3.5.7.3 PCR screening of genome-edited bulk and single-cell derived colonies

DNA was extracted using QuickExtract™ DNA Extraction Solution (Lucigen), depending on confluency 25 to 100 µl of solution was added to the wells containing the cells for 15 minutes at 37C, cell suspension was transferred to tubes and vortexed for a minute. Samples were then heated at 65 °C for 6 minutes, and 98 °C for 2 minutes. The extracted DNA solution was stored at -20 °C. PCR amplification was performed by adding 4 or 5 µl of template DNA, 1 µM forward (SOCS2_F1: accaagtttgtgtgggtgct) and reverse (SOCS2_R1: cttccagcgtgctaagaagc) primers, and EconoTaq PLUS GREEN 2X Master Mix (Lucigen) in a 25 µl reaction. PCR conditions included an initial denaturation at 94 °C for 2 minutes, 35 cycles of denaturation at 94 °C for 30 seconds, annealing at 63 °C for 30 seconds, and extension at 72 °C for 65 seconds, followed by a final extension at 72 °C for 10 minutes. Following amplification of column purified genomic DNA, the PCR product was cleaned up and sequenced using the primers described above and the Bioedit software was used to visualize the chromatograms.

3.5.7.4 Fibronectin fragment (FN-f) treatment and qPCR analysis of genome edited samples

Single cell colonies in 24-well plates were passaged to 6-well plates for expansion. Following genotype confirmation, colonies with similar genotype were combined and seeded at 250,000 cells per well in a 12-well plate. Cultured cells were made serum free and treated with FN-f 1 µM or PBS. Following treatment, media was removed and cells were immediately lysed in the RLT buffer. RNA was isolated with RNeasy Plus columns (Qiagen) and reverse transcribed to cDNA using qScript™ XLT cDNA SuperMix (VWR) or iScript cDNA Synthesis Kit (1708891, Bio-Rad). DNase treatment was used for the second and third donors in order to confirm that detectable *SOCS2* signal in knockout cells was due to the presence of genomic DNA. To evaluate the effect of *SOCS2* editing on inflammatory gene response

quantitative polymerase chain reaction (qPCR) was performed on a QuantStudio™ 6 Flex machine (Applied Biosystems) with TaqMan™ Universal Master Mix and TaqMan Gene Expression Assays for human *CCL2* (Hs00234140_m1), *IL6* (Hs00174131_m1), and housekeeping gene *TBP* (Hs00427620_m1). SOCS2 expression was assessed in pooled colonies with TaqMan Gene Expression Assay Hs00919620_m1.

3.5.7.5 Western Blot analysis

Following genotype identification by PCR, cells from a wildtype, heterozygous and knockout colony were expanded in chondrocyte media supplemented with 5 ng/ml bFGF and 1 ng/ml TGF-β1 (Life technologies) for 11 days. Cells were lysed in standard cell lysis buffer (1X) (Cell signaling technology) containing phenylmethanesulfonyl fluoride (PMSF; Sigma-Aldrich) and phosphatase inhibitor mix. Protein (15 µg) was separated by SDS-PAGE and transferred to nitrocellulose membrane. After blocking in 5% nonfat milk in TBST, the blot was incubated with SOCS2 antibody (PA5-17219; 1:1000; Thermo Fisher) overnight at 4C and secondary antibody solution for 1 hour. The membrane was incubated in Radiance Plus Chemiluminescent Substrate (Azure Biosystems) and signal detected using the Azure c600 gel imaging system. The membrane was striped and incubated with the loading control beta tubulin antibody.

3.5.8 Datasets

3.5.8.1 Osteoarthritis GWAS

Genome-wide association statistics for 11 osteoarthritis phenotypes and lead variants identified in Boer et al. (Boer *et al.* 2021) were obtained from the Musculoskeletal Knowledge Portal (Kiel *et al.* 2020).

3.5.8.2 Epigenome Roadmap Data

Consolidated reference human epigenomes for 98 cell/tissue types were obtained from the NIH Roadmap Epigenomics Project (Bernstein *et al.* 2010) and The Encyclopedia of DNA Elements (ENCODE) project (ENCODE Project Consortium 2012). Processed narrowPeak files for H3K27ac, H3K4me1, and H3K4me3 and BigWig files for H3K27ac were used for each cell/tissue type. Additional narrowPeak files for H3K9ac, H3K9me3, H3K27me3, and H3K36me3 were obtained for mesenchymal stem cell derived chondrocyte cultured cells (E049).

3.5.8.3 RNA-seq time course of fibronectin fragment (FN-f) treatment

RNA-seq data from a prior study of FN-f treated human chondrocytes was obtained from KSM Reed et al. (Reed *et al.* 2021) and vst-normalized, centered, and replicate-combined. The 0-hour FN-f treatment time point was created by combining the 9 PBS-treated replicates. Genes were considered differential with a BH-adjusted p-value of 0.01 and a log2 fold-change threshold > 1.25 across any time point.

3.5.9 Cell type enrichment for OA risk variants

To identify the cell types that likely mediate genetic OA risk, we performed SNP enrichment analysis using GREGOR (Genomic Regulatory Elements and Gwas Overlap algoRithm) (Schmidt *et al.* 2015). Publicly available H3K27ac, H3K4me1, and H3K4me3 ChIP-seq narrowPeaks files from the NIH Roadmap Epigenomics Mapping Consortium were merged and sorted using bedtools (v2.29.2) (Quinlan and Hall 2010) to define regulatory loci for 98 cell types in hg19. GREGOR was used to determine each cell type's enrichment for 104 OA lead SNPs (Boer *et al.* 2021) by comparing the observed overlap between regulatory loci and SNPs with their expected overlap and evaluating significance. Expected overlap is determined using a matched control set of ~500 variants that control for the number of LD proxies, gene proximity and minor allele frequency. Reference data from 1000 Genomes Phase 1 version 2 EUR panel were used with GREGOR to control for LD proxies (1Mb, $r^2 > 0.7$) (1000 Genomes Project Consortium *et al.* 2010). Results were imported into R (v4.1.0) (R Core Team 2021) and visualized with ggplot2 (Wickham 2016) and plotgardener (Kramer *et al.* 2022).

3.5.10 Putative OA risk variants

LD proxies for 104 OA GWAS signals from Boer et al. were identified using the 1000 Genomes European reference panel since the GWAS data primarily analyzed individuals of European ancestry (11 of 13 cohorts are of European descent). r^2 values were calculated with the –ld function in PLINK 1.9 (Purcell *et al.* 2007; 1000 Genomes Project Consortium *et al.* 2010) using a window of 1 Mb for LD calculation. Putative OA risk variants were defined as those in high LD ($r^2 > 0.8$, n = 1,259) with lead variants.

3.5.11 Multi-omic integration for assigning SNPs to putative OA risk genes

We took a multi-omic approach to identify putative SNP-gene pairs implicated in OA. SNPs that 1) were predicted to affect coding regions of genes, 2) overlapped gene promoters, or 3) overlapped a

regulatory peak looped to a gene's promoter were assigned to the "Coding gene", "Gene promoter", or "Loops to gene promoter" categories, respectively. Genes in each category that change in response to FN-f ($p \leq 0.01$ and LFC at any time point ≥ 1.25) were highlighted as putative OA risk genes.

Coding SNP-gene pairs were identified using ENSEMBL's Variant Effect Predictor (VEP) tool. Putative OA risk variants (n = 1,259) were annotated with their predicted consequence on coding sequence using VEP run with the GRCh37.p13 human genome and default parameters. SNPs with a predicted consequence of "missense" or "synonymous" were paired with their affected genes assigned to the "Coding gene" category.

Promoter regions were defined as 2000 bp upstream and 200 bp downstream of the TSS of transcripts obtained with the TxDb.Hsapiens.UCSC.hg19.knownGene Bioconductor package for a total of 8,2960 transcripts. Gene symbols were linked to transcript ranges using the OrganismDbi and Homo.sapiens packages. Transcripts without gene symbols or those not present in the FN-f RNA-seq data were filtered out, leaving a total of 62,590 transcript promoters.

Chondrocyte regulatory regions were defined by combining Roadmap Epigenomics data with data from primary human articular chondrocytes. Specifically, H3K4me1, H3K4me3, H3K9ac, H3K9me3, H3K27ac, H3K27me3, and H3K36me3 peaks from mesenchymal stem cell derived chondrocyte cultured cells (E049) obtained through AnnotationHub (v3.1.7, snapshot date 2021-10-20) (Morgan *et al.* 2017) were combined with Donor-merged H3K27ac peaks from primary human articular chondrocytes. OA SNPs were overlapped with chondrocyte regulatory regions resulting in 507 SNPs.

SNPs overlapping chondrocyte regulatory regions that also overlapped a promoter region were assigned to their affected gene and added to the "Gene promoter" category. SNPs overlapping chondrocyte regulatory regions were intersected with loop calls from Hi-C in the C-28/I2 chondrocyte cell line (see Methods on Hi-C processing and loop calling). The linkOverlaps function from the InteractionSet package was used to identify chondrocyte regulatory SNPs that are connected to promoters by loops. These SNP-gene pairs were assigned to the "Loops to gene promoter" category.

3.5.12 Motif analysis

Tomtom (v5.4.1; release date: Sat Aug 21 19:23:23 2021 -0700) from the MEME suite was used to identify motif matches for sequences surrounding the rs7953280 variant (Gupta *et al.* 2007). All 7-mers

surrounding rs7953280 ("GGCTTTG", "GCTTTGA", "CTTTGAG", "TTTGAGG", "TTGAGGC", "TGAGGCA", "GAGGCAT") and the entire 13 bp sequence ("GGCTTTGAGGCAT") were used to identify motif matches. Sequences were input into the online motif comparison tool and queried against the JASPAR2022_CORE_vertebrates_non-redundant_v2 and HOCOMOCOv11_core_HUMAN_mono_meme_format motif databases. Pearson correlation coefficient was used as the motif column comparison function and the significance threshold was set to an E-value < 10; no q-value threshold was set and reverse complementing of motifs was permitted. The following command summarizes the parameters used: “tomtom -no-ssc -oc . -verbosity 1 -min-overlap 5 -mi 1 -dist pearson -evalue -thresh 10.0 -time 300 query_motifs motif_databases."

3.5.13 Transcription factor (TF) motif binding propensity

We used SNP Effect Matrix scores (SEMs) to predict the TF binding propensity between risk and non-risk SNPs in OA. Pre-calculated SEMs for 211 TF motifs were obtained from SEMpl (https://github.com/Boyle-Lab/SEMpl) and used for scoring risk and non-risk SNP sequences (Nishizaki *et al.* 2020). Binding propensity scores were determined by generating frame-shifted K-mers covering each TF motif position for both risk and non-risk sequences. K-mers were scored against 211 TF SEMs using position-weight matrix (PWM) scoring functions from the Biostrings Bioconductor package (Pagès *et al.* 2021). The best scoring K-mer frame for each TF motif was used to select the binding score for risk and non-risk sequences. Scores were normalized by applying inverse-log transformation, subtracting the scrambled baseline provided with each SEM, and dividing the result by the absolute value of that baseline. TFs with positive scores are predicted to be bound while negative scores are predicted to be unbound.

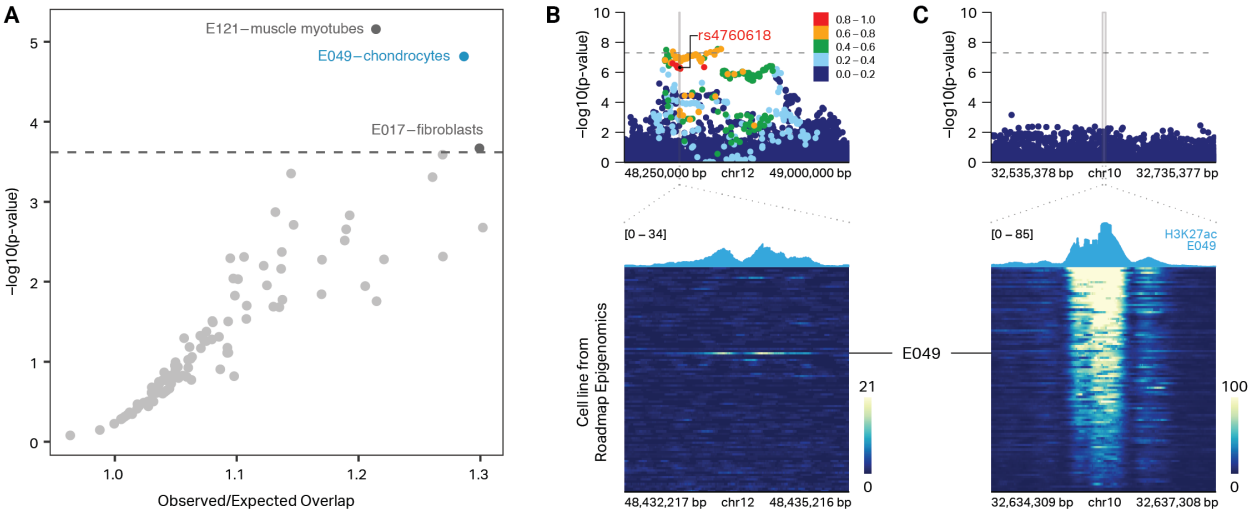

Figure 3.1. OA risk variants are enriched in chondrocyte regulatory elements. (A) Enrichment analysis of 98 cell types from the NIH Roadmap Epigenomics Mapping Consortium reveals that OA GWAS variants are enriched in the regulatory regions (H3K27ac, H3K4me1, or H3K4me3 ChIP-seq peaks) of chondrocytes, skeletal muscle myotubes, and fibroblasts. **(B)** Heatmap of H3K27ac signal from 98 cell types (bottom) highlights a chondrocyte-specific enhancer that overlaps Knee/Hip osteoarthritis risk variant (Boer *et al.* 2021) (rs4760618, circled) that is in high LD ($r^2 > 0.8$) with the lead variant (rs7967762, red diamond) for this locus (top). **(C)** Heatmap of H3K27ac signal from 98 cell types (bottom) highlights a ubiquitous enhancer (active in >90% of cell types) that does not overlap an OA GWAS variant (top).

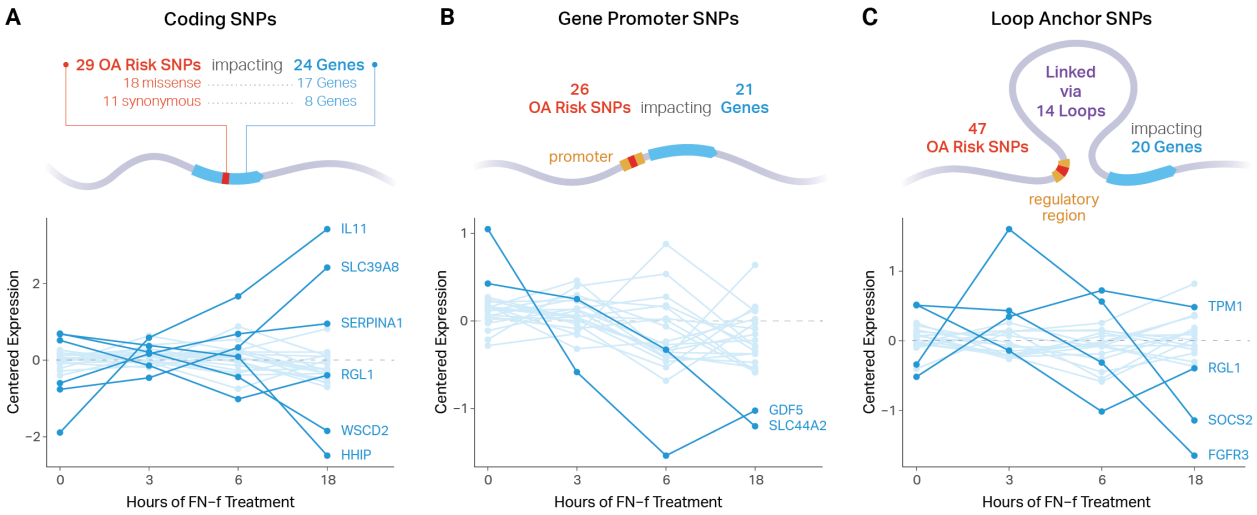

Figure 3.2. Multi-omic integration for assigning SNPs to putative OA risk genes. (A) ENSEMBL's Variant Effect Predictor tool identified 29 unique OA risk SNPs (18 missense and 11 synonymous) overlapping coding regions of 24 unique genes (17 missense, 8 synonymous). **(B)** 26 unique OA risk SNPs overlapped both a chondrocyte regulatory region (H3K27ac, H3K4me1, or H3K4me3 ChIP-seq peaks) and a gene promoter for 21 unique genes. **(C)** 47 unique SNPs overlapped chondrocyte regulatory regions connected to 20 unique gene promoters via 14 C-28/I2 chromatin loops. RNA-seq data from our *ex-vivo* OA model depicts how putative OA risk genes change in response to FN-f. Normalized expression of genes are shown below each category over an 18 hour time-course of fibronectin fragment (FN-f) treatment. Differential genes ($p \leq 0.01$, absolute $\log_2$ fold-change ≥ 1.25) are colored and labeled.

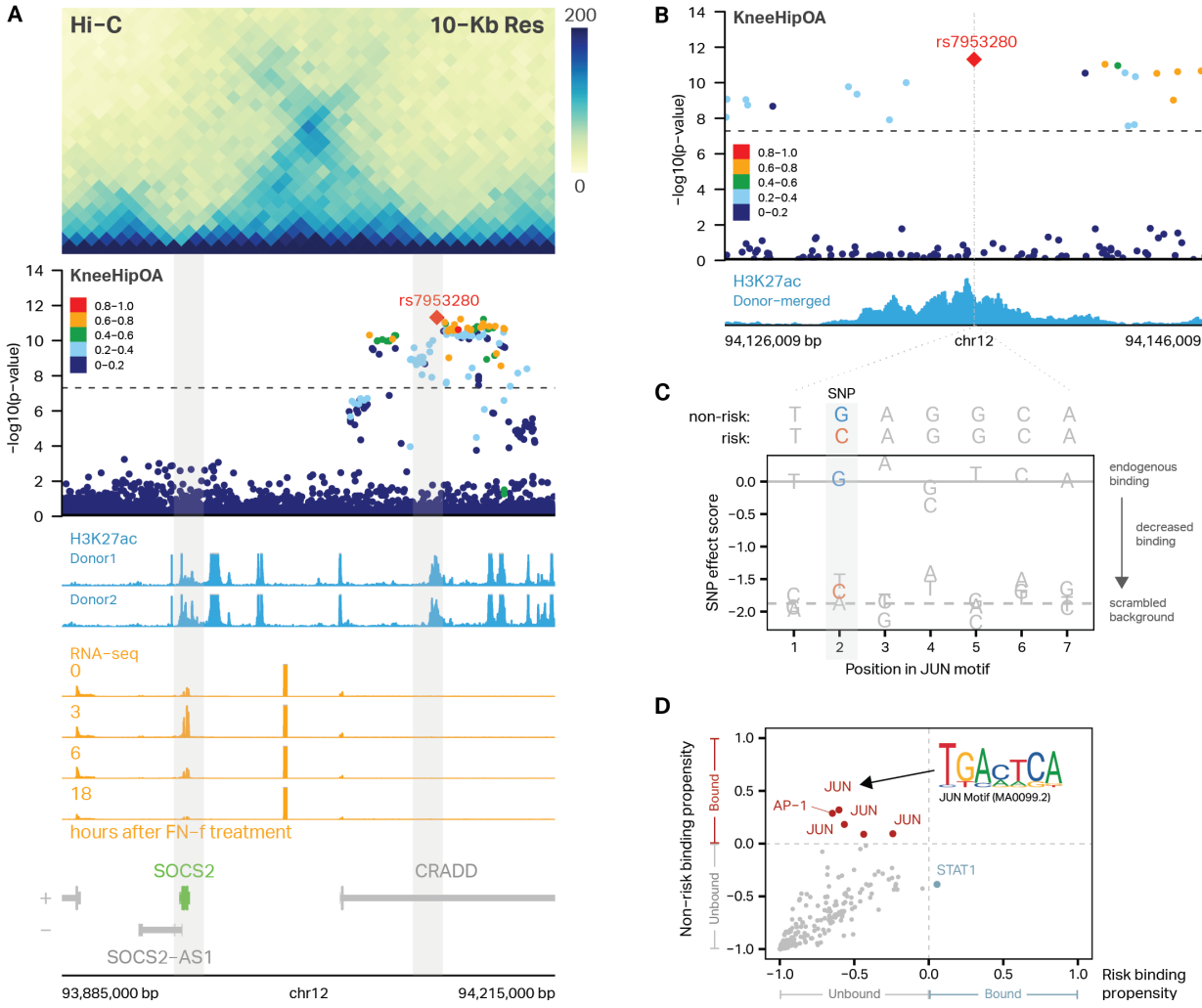

Figure 3.3. 3D chromatin interactions identify SOCS2 as a putative regulator of OA. (A) Hi-C performed in C-28/I2 cells reveals a chromatin loop connecting OA risk variant rs7953280 (right gray bar) to the promoter of *SOCS2* (left gray bar). rs7953280 is located in an intronic region of *CRADD* and overlaps an H3K27ac peak in primary articular human chondrocytes from two donors (blue signal tracks). *SOCS2* is differentially expressed in response to treatment with FN-f. Gene tracks are shown below with +/- indicating gene strand. **(B)** Zoom-in on rs7953280 shows that the SNP is located within an H3K27ac peak in primary articular human chondrocytes. **(C)** Motif analysis identifies a JUN binding site at rs7953280. SNP effect matrix (SEM) data predicts decreased binding at the JUN motif (JASPAR ID: MA0099.2) with a G to C polymorphism in the second position. **(D)** Motif analysis from 211 precomputed SEMs from SEMpl predicts that JUN/AP-1 motifs (red, upper left quadrant) bind to the non-risk but not the risk allele.

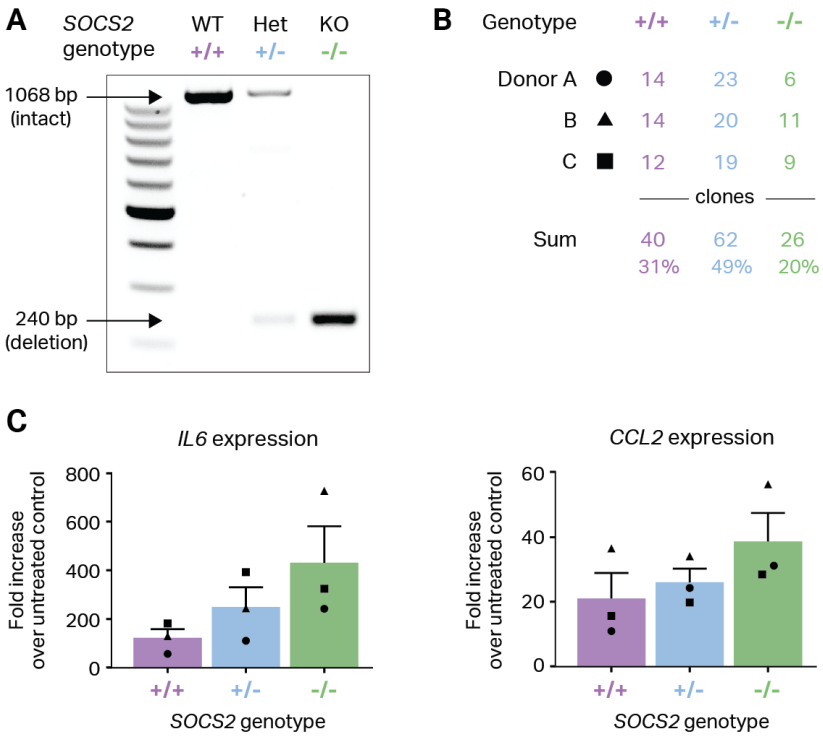

Figure 3.4. SOCS2 deletion increases proinflammatory gene expression in response to FN-f. (A) PCR primers surrounding the intended *SOCS2* deletion were used to screen single-cell-derived colonies from three independent donors. WT - wildtype (+/+. purple); het - heterozygous deletion (+/-, blue); KO - homozygous knockout deletion (-/-, green). **(B)** Efficiency of deleting the intended *SOCS2* deletion in primary human chondrocytes from three independent donors. **(C)** qPCR at 18 hours after FN-f treatment revealed increased expression of the proinflammatory genes *IL6* and *CCL2* in *SOCS2* deletion colonies from three independent donors.

3.6 Supplemental Methods

3.6.1 Hi-C

In situ Hi-C was performed as previously described (Rao *et al.* 2014). Pellets were lysed in ice-cold Hi-C lysis buffer (10mM Tris-HCl pH 8.0, 10mM NaCl, 0.2% IGEPAL CA630) with 50µL of protease inhibitors for 15 min on ice. Cells were pelleted and washed using the same buffer. Pellets were resuspended in 50µL 0.5% SDS and incubated at 62°C for 7 min. Reactions were quenched with 145µL water and 25µL 10% Triton X-100 at 37°C for 15 min. Chromatin was digested overnight with 25µL 10X NEBuffer2 and 100U MboI at 37°C with rotation.

Reactions were incubated at 62°C for 20 min then cooled to RT. Fragment overhangs were repaired by adding 37.5µL 0.4mM biotin-14-dATP; 1.5µL each 10mM dCTP, dGTP, dTTP; 8µL 5U/µL DNA Polymerase I, Large (Klenow) Fragment and incubating at 37°C for 1.5 h with rotation. Ligation was performed by adding 673µL water, 120µL 10X NEB T4 DNA ligase buffer, 100µL 10% Triton X-100, 6µL 20mg/mL BSA, and 1µL 2000U/µL T4 DNA ligase and incubating at RT for 4 h with slow rotation. Samples were pelleted at 2500g, resuspended in 432µL water, 18µL 20mg/mL proteinase K, 50µL 10% SDS, and 46µL 5M NaCl, incubated at 55°C for 30 min, and then transferred to 68°C overnight. Samples were cooled to RT and 1.6x volumes of pure ethanol and 0.1x volumes of 3M sodium acetate pH 5.2 were added to each sample, prior to incubation at -80°C for over 4-6 h. Samples were spun at max speed at 2°C for 15 min and washed twice with 70% ethanol. Pellets were dissolved in 130µL 10mM Tris-HCl pH 8.0 and incubated at 37°C for 1-2 h. Samples were stored at 4°C overnight.

DNA was sheared using the Covaris LE220 (Covaris, Woburn, MA) to a fragment size of 300-500bp in a Covaris microTUBE. DNA was transferred to a fresh tube and the Covaris microTUBE was rinsed with 70µL of water and added to the sample. A 1:5 dilution of DNA was run on a 2% agarose gel to verify successful shearing.

Sheared DNA was size selected using AMPure XP beads. 0.55x volumes of 2X concentrated AMPure XP beads were added to each reaction and incubated at RT for 5 min. Beads were reclaimed on a magnet and the supernatant was transferred to a fresh tube. 30µL 2X concentrated AMPure XP beads were added and incubated for 5 min at RT. Beads were reclaimed on a magnet and washed with fresh

70% ethanol. Beads were dried for 5 min at RT prior to DNA elution in 300µL 10mM Tris-HCl pH 8. Undiluted DNA was run on a 2% agarose gel to verify successful size selection between 300-500 bp. 150µL 10mg/mL Dynabeads MyOne Streptavidin T1 beads were washed with 400µL 1X Tween washing buffer (TWB; 250µL Tris-HCl pH 7.5, 50µL 0.5M EDTA, 10mL 5M NaCl, 25µL Tween 20, 39.675µL water). Beads were then resuspended in 300µL 2X Binding Buffer (500µL Tris-HCl pH 7.5, 100µL 0.5M EDTA, 20mL 5M NaCl, 29.4mL water), added to the DNA sample, and incubated at RT for 15 min with rotation. DNA-bound beads were then washed twice with 600µL 1X TWB at 55°C for 2 min with shaking. Beads were resuspended in 100µL 1X NEBuffer T4 DNA ligase buffer, transferred to a new tube, and reclaimed.

Sheared ends were repaired by resuspending the beads in 88µL 1X NEB T4 DNA Ligase Buffer with 1mM ATP, 2µL 25mM dNTP mix, 5µL 10U/uL NEB T4 PNK, 4uL 3U/uL NEB T4 DNA polymerase I, and 1uL 5U/uL NEB DNA polymerase 1, large (Klenow) fragment and incubating at RT for 30 min. Beads were washed two more times with 1X TWB for 2 min at 55°C with shaking. Beads were washed once with 100uL 1X NEBuffer 2, transferred to a new tube, and resuspended in 90uL 1X NEBuffer 2, 5uL 10mM dATP, and 5uL NEB Klenow exo minus, and incubated at 37°C for 30 min. Beads were washed two more times with 1X TWB for 2 min at 55°C with shaking. Beads were washed in 100uL 1X Quick Ligation Reaction Buffer, transferred to a new tube, reclaimed, and resuspended in 50uL 1X NEB Quick Ligation Reaction Buffer. 2uL NEB DNA Quick Ligase and 3uL of an appropriate Illumina indexed adapter (TruSeq nano) were added to each sample before incubating at RT for 15 minutes. Beads were reclaimed and washed twice with 1X TWB for 2 min at 55°C. Beads were washed in 100uL 10mM Tris-HCl pH 8, transferred to a new tube, reclaimed, and resuspended in 50uL 10mM Tris-HCl pH 8.
Hi-C libraries were amplified directly off T1 beads with 8 cycles in 5uL PCR primer cocktail, 20uL Enhanced PCR mix, and 25uL of DNA on beads. The PCR settings were as follows: 3 min at 95°C followed by 4-12 cycles of 20s 98°C, 15s at 60°C, and 30s at 72°C. Samples were held at 72°C for 5 min before holding at 4°C. Amplified samples were transferred to a new tube and brought to 250uL in 10mM Tris-HCl pH 8.

Beads were reclaimed and the supernatant containing the amplified library was transferred to a new tube. Beads were resuspended in 25uL 10mM Tris-HCl pH 8 and stored at -20°C. 0.7x volumes of

warmed AMPure XP beads were added to the supernatant sample and incubated at RT for 5 min. Beads were reclaimed and washed with 70% ethanol without mixing. Ethanol was aspirated. Beads were resuspended in 100uL 10mM Tris-HCl pH 8, 70uL of fresh AMPure XP beads were added, and the solution was incubated for 5 min at RT. Beads were reclaimed and washed twice with 70% ethanol without mixing. Beads were left to dry and DNA was eluted in 25uL 10mM Tris-HCl pH 8. The resulting libraries were then quantified by Qubit and Tapestation. A low depth sequence was performed first using the Miniseq sequencer system (Illumina) and analyzed using the Juicer pipeline to assess quality. The resulting libraries underwent paired-end 2x150bp sequencing on an Illumina NovaSeq sequencer. Each replicate was sequenced to an approximate depth of 750 million reads. The full sequencing depth was 2.8 billion reads.

3.6.2 Cut and Run

Following flash freezing, thawed pellets were resuspended in 1mL ice-cold nuclei isolation buffer (NE1 buffer; 20mM HEPES pH 7.5, 10mM KCl, 1mM MgCl2, 1mM DTT, 0.1% Triton X-100, 1X CPI added fresh) and incubated for 10 min at 4°C with rotation. Nuclear pellet was collected by centrifugation at 1000g for 5 min at 4°C, then resuspended in 1mL of ice-cold wash buffer (WB; 20mM HEPES pH 7.5, 0.2% Tween-20, 150mM NaCl, 150mM BSA, 0.5mM Spermidine, 10mM Na-Butyrate, 1X CPI added fresh). 10uL concanavalin A lectin beads washed and resuspended in binding buffer (BB; 20mM HEPES pH 7.5, 10mM KCl, 1mM CaCl2, 1mM MnCl2) were add- ed to each sample and incubated for 10 min at RT. Beads were reclaimed and resuspended in 50uL antibody buffer (AbB; WB supplemented with 0.1% Triton X-100 and 2mM EDTA). 0.01ug/uL H3K27ac antibody in AbB was added to each sample and samples were incubated overnight at 4°C with mixing at 1000xRPM.

Beads were reclaimed and washed with 1mL triton wash buffer (TwB; WB supplemented with 0.1% Triton X-100) without mixing. Beads were reclaimed and resuspended in 50uL AbB. 2.5uL CUTANA pAG-MNase (Epicypher, #15-1016) was added and samples were incubated for 1 h at 4°C on a metal block. Beads were reclaimed and washed twice with 1mL TwB before resuspension in 100uL TwB. To digest chromatin, 2uL 100mM CaCl2 was added and samples were incubated for 30 min at 4°C on a metal block. Digestion was halted by the addition of 100uL 2X STOP buffer (340mM NaCl, 20mM EDTA, 4mM EGTA, 0.1% Triton X-100, 50ug/mL RNAse A). Samples were incubated for 20 min at 37°C to

release pA-MNase cleaved fragments from nuclei). Beads were placed on a magnet and the supernatant containing DNA fragments was transferred to a new tube. To reverse crosslinks, 2uL 10% SDS and 2uL 20mg/mL proteinase K were added to each sample and incubated for 1 h at 65°C. DNA was purified using the Zymo DNA Clean & Concentrator Kit according to manufacturer's protocols using 5 volumes of DNA binding buffer. DNA was eluted in 55uL water.

Sequencing libraries were prepared from CUT&RUN fragments using KAPA HyperPrep with library amplification kit (no. KK8504) following the manufacturer's instructions. Post-ligation bead cleanup was performed with two rounds of 1.2X volumes of beads and DNA was eluted in a final volume of 25uL 10mM Tris-HCl pH 8. Library amplification was performed with 20uL of the adapter ligated DNA with 12 PCR cycles. One round post amplification cleanup was performed with 1.2X volumes of beads. The resulting libraries were then quantified by Qubit and Tapestation. A low depth sequence was performed first using the Miniseq sequencer system (Illumina) and analyzed using the Juicer pipeline to assess quality control. The resulting libraries underwent paired-end 2x150bp sequencing on an Illumina NextSeq sequencer.

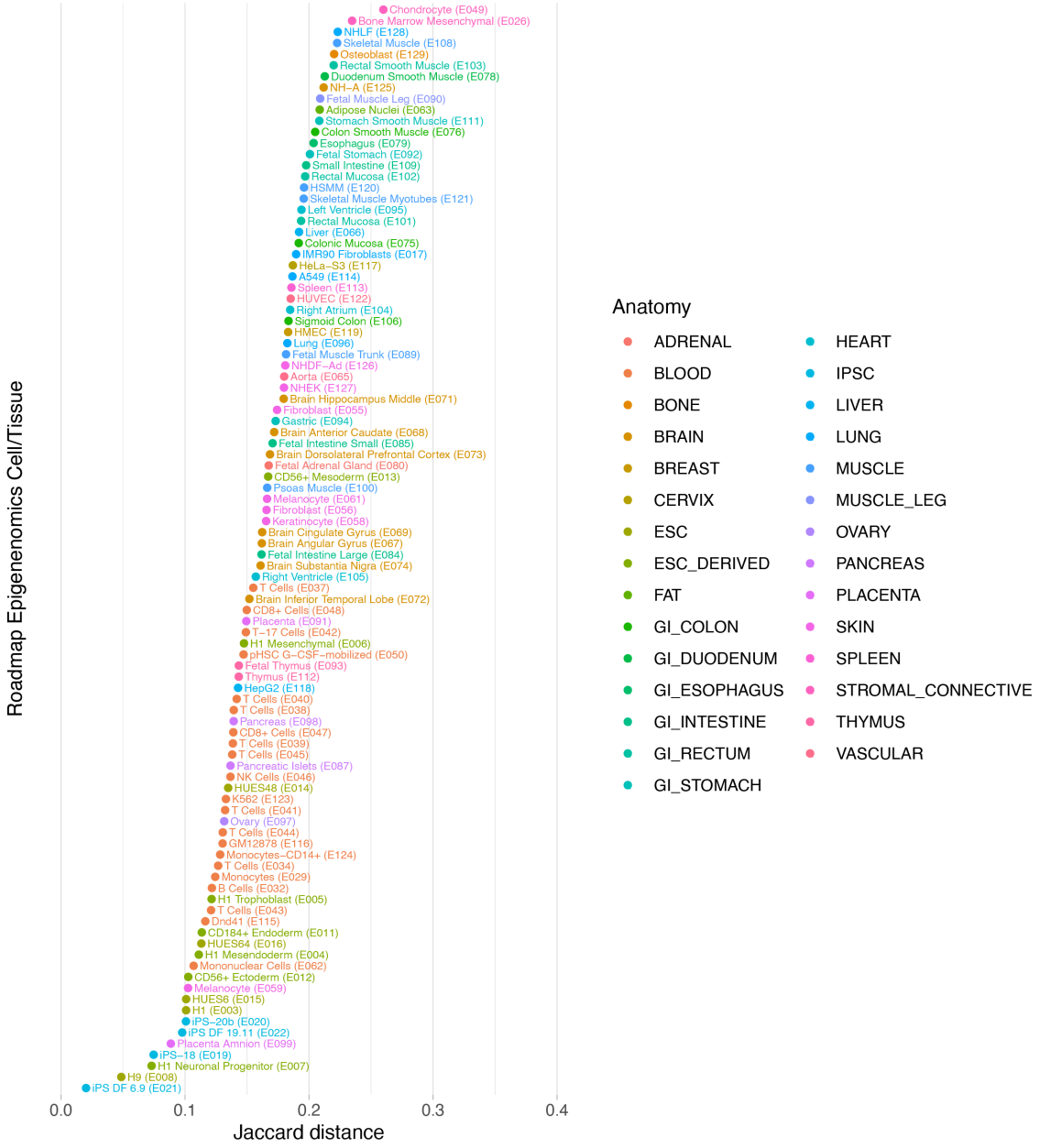

Figure 3.S1. Jaccard distance (similarity) between primary human chondrocytes and each cell type from the Roadmap Epigenomics Project. The cells with the highest similarity to primary human chondrocytes are “E049 - Mesenchymal Stem Cell Derived Chondrocyte Cultured Cells”. The Jaccard distance was calculated as intersection over union between each set of H3K27ac ChIP-seq peaks.

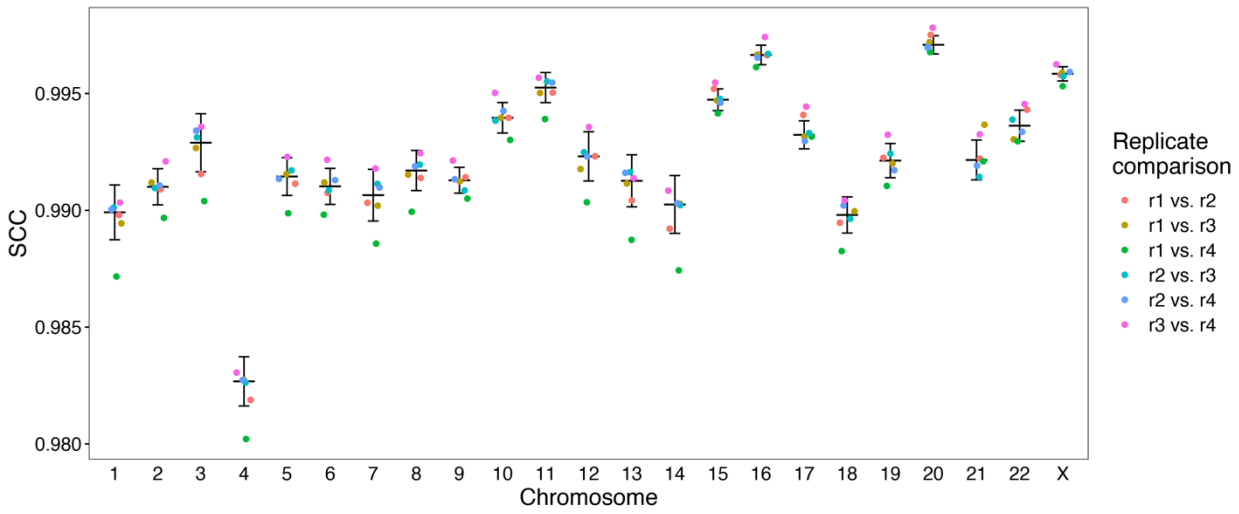

Figure 3.S2. Quantifying similarity of Hi-C replicates. Stratum adjusted correlation coefficient (SCC) comparing similarity between C-28/I2 Hi-C replicates for each chromosome (Yang *et al.* 2017). The python implementation of HiCRep was used to calculate SCC values at 10-kb resolution with a smoothing parameter of --h 20 and maximum distance of --dBPMax 5000000 (Lin *et al.* 2021).

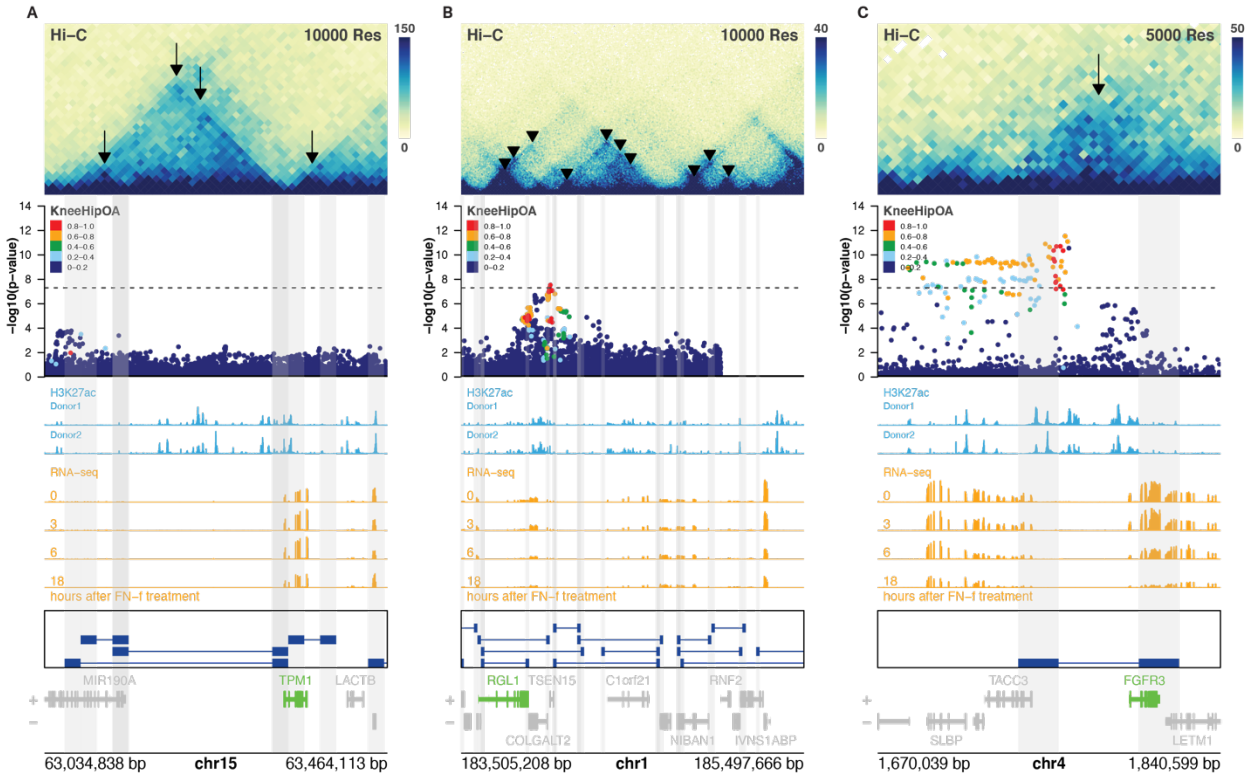

Figure 3.S3. Loci of looped variant-gene pairs identified as differentially expressed in response to FN-f. Hi-C in C-28/I2 cells (Hi-C heatmap) shows chromatin loop anchors (gray vertical highlights and blue bars) connecting OA GWAS variants (Manhattan plot) to promoters (bottom gene track) of differentially expressed genes in response to a timecourse of FN-f treatment (orange signal tracks). Blue signal tracks show CUT&RUN data for H3K27ac from two donors of primary human chondrocytes. Loci of variant-gene pairs shown are (a) rs746239049-TPM1, (b) rs10797938-RGL1, and (c) rs4535386/rs2896518-FGFR3.

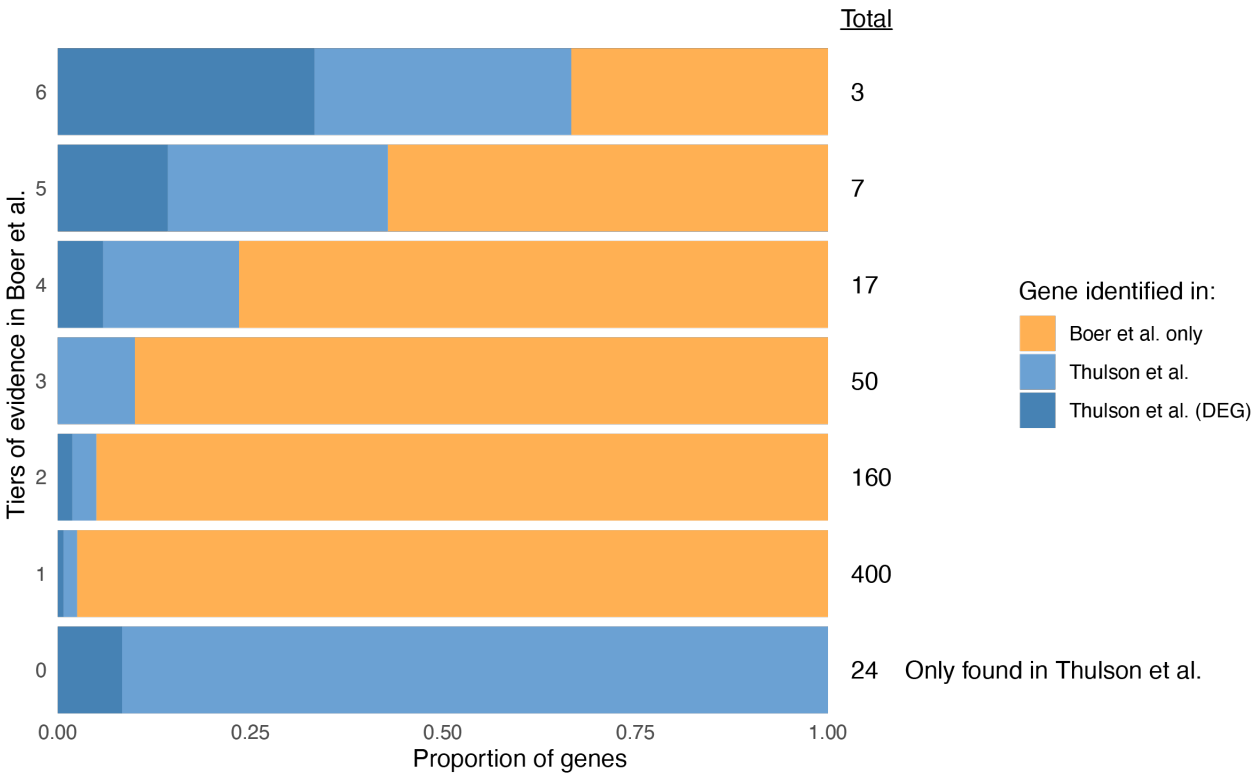

Figure 3.S4. Effector gene comparison between Boer et al. and Thulson et al. Proportion of genes identified in Thulson et al. by evidence score in Boer et al. Differential gene in Thulson et al. indicates the proportion of genes that changed significantly in response to FN-f from Reed et al. (see methods). Evidence level 0 indicate novel genes that were found in Thulson et al. but not in Boer et al.

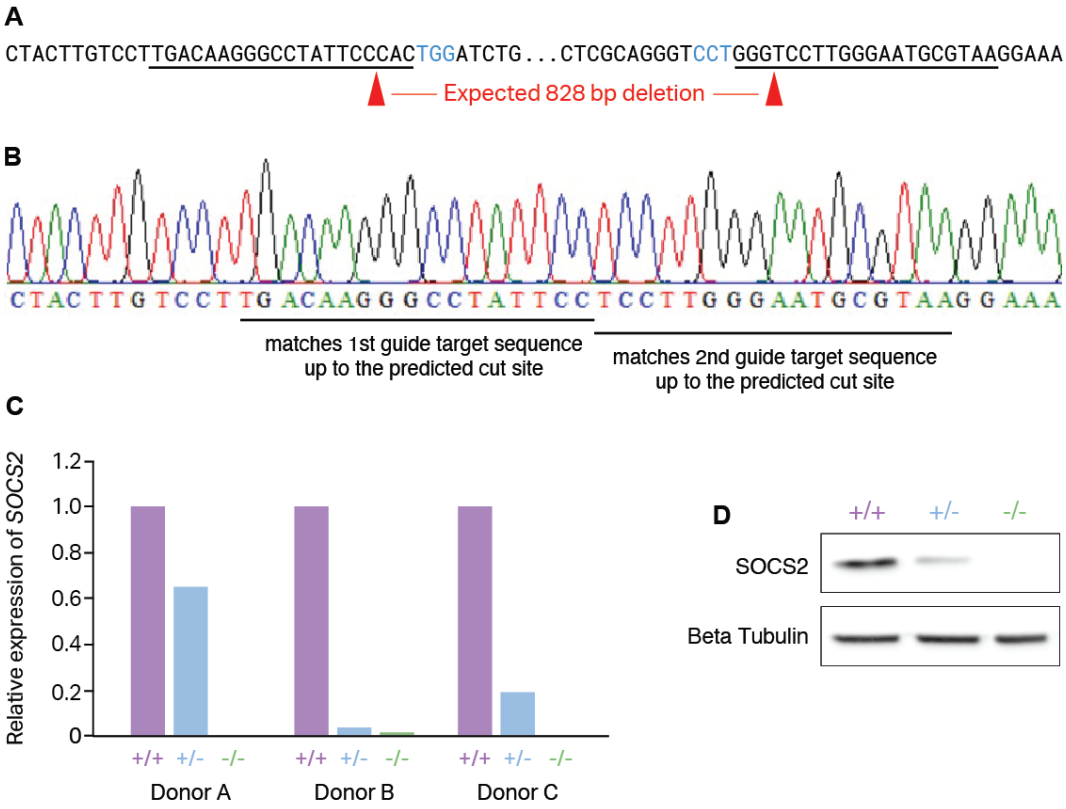

Figure 3.S5. Validation of *SOCS2* knockout. (A) Region of SOCS2 targeted by editing, PAM sites in blue, guide RNA target sites underlined, expected cut sites depicted by red triangles. **(B)** Sanger sequencing of knockout colony confirmed deletion. **(C)** SOCS2 expression was analyzed in pooled colonies by qPCR, RNA from colonies of donors A and C were treated with DNase before qPCR to eliminate any background signal of SOCS2 from remaining genomic DNA. Expression was normalized to housekeeping gene TBP and relative expression calculated. **(D)** Western blot analysis confirmed decreased expression in heterozygous colony and loss of SOCS2 in knockout colony.

CHAPTER 4: GENERATING GENOTYPE AND TRANSCRIPTIONAL DATA FOR USE IN QTL ANALYSES

4.1 Contributions

This work was the combined effort of myself and Philip R. Coryell, Susan D'Costa, and Nicole E. Kramer. P.R.C. cultured and treated human primary articular chondrocytes and collected cells and RNA for downstream application. I extracted DNA and RNA, performed QC, and shipped the samples to UNC's Mammalian Genotyping Core for genotyping and to the New York Genome Center for RNA-seq library preparation and sequencing. S.D. contributed to DNA and RNA extraction and QC as well. N.E.K. performed data analysis.

4.2 Introduction

Osteoarthritis (OA) has a strong genetic component (MacGregor and Spector 1999; Aubourg *et al.* 2022), which many studies have attempted to characterize by aiming to identify putative causal genes and the mechanisms underlying disease progression. Recent genome-wide association studies (GWAS) have identified genetic variants, or single nucleotide polymorphisms (SNPs), that are associated with the disease (Tachmazidou *et al.* 2019; Boer *et al.* 2021). However, determining the causal SNPs, the molecular mechanisms, and the genes affected remains difficult because many SNPs do not directly affect coding regions and may instead influence gene expression by altering regulatory elements. Furthermore, many neighboring variants are in high linkage disequilibrium (LD), a phenomenon in which these variants are inherited together due to co-segregation during meiotic recombination (Slatkin 2008; Cano-Gamez and Trynka 2020). Because of this, LD makes it difficult to discern the causal variant. Expression quantitative trait loci (eQTL) are genomic regions that contain one or more variants that influence the expression level of a gene. eQTL mapping is used as a statistical tool to connect genetic variation with changes in gene expression so as to elucidate potential mechanisms behind gene regulation. The integration of GWAS with eQTL mapping can help discover previously unidentified disease loci in the correct cell, condition, and tissue type (Hormozdiari *et al.* 2016; Umans *et al.* 2021).

eQTLs are identified first by collecting genotype and gene expression data from a population of genetically diverse individuals. Genotypes are compared to expression levels using association; individuals are grouped according to the allele they carry and if expression levels are higher in one group than the other, it can be postulated that the variant being tested influences expression of the gene (Albert and Kruglyak 2015). eQTLs have been mapped in a large number of tissues, in large part by the Genotype-Tissue Expression Project (GTEx) (GTEx Consortium 2013), in an effort to interpret GWAS findings for translational research by providing a resource database and tissue bank. However, studies that focus on the effects of genetic variation on gene expression in a context-specific fashion – such as in response to a disease-relevant stimulus – are still needed in a majority of tissues and diseases (Umans *et al.* 2021). Importantly, the GTEx database has yet to include genetic variation studies in chondrocytes, which would provide significant data for future research on mechanisms underlying OA.

In following with the concept of studying genetic variation on gene expression in a context-specific fashion, we utilized an *ex vivo* model for simulating an OA phenotype. In this model, chondrocytes are treated with fibronectin fragment (FN-f). Fibronectin is an extracellular protein present in cartilage, and stimulation with its breakdown product has been shown to recapitulate OA features (Xie *et al.* 1992; Homandberg *et al.* 1998; Homandberg 1999; Collins *et al.* 2019). One of our previous studies characterized the response of chondrocytes isolated from healthy tissue to FN-f treatment, validating that the transcriptional changes seen in this model mimic those seen in OA (Reed *et al.* 2021).

In this study, 105 samples of chondrocytes were isolated from healthy donor articular cartilage and treated with either PBS as a control or FN-f for 18 hours. DNA from non-treated cells was extracted and genotyped using the Mammalian Genotyping Core at the University of North Carolina at Chapel Hill. RNA from control and FN-f treated cells was extracted and sent to the New York Genome Center for RNA-seq library preparation and sequencing. Preliminary analyses indicate that our in vitro model of OA robustly mimics the OA phenotype. My generation of high quality data provides the basis for further QTL analysis and overlap with OA GWAS studies in order to identify novel putative target genes, risk variants, and mechanisms underpinning OA progression.

4.3 Results

4.3.1 Donor characteristics

We currently have collected DNA and RNA samples from a total of 105 donors. Of those 105 donors, ~84% are male (88) and ~16% are female (17) (**Figure 4.2A**). 65% (68) of the donors are Caucasian, 10% (10) are Hispanic, 19% (20) are Black, 3% (3) are Asian, 1% (1) are Arab, and 3% (3) were not disclosed (**Figure 4.2B**). The youngest donor was 34, and the oldest 84, with the average age being 61 years. 63 was the most common age (7 donors), followed by 50, 57, and 72 years (6 donors each) (**Figure 4.2C**).

In order to confirm the ancestry of our donor samples, we performed a principal component analysis (PCA) and combined our genotyping data with the 1000 Genomes Reference (**Figure 4.2D)**. By mapping our population's data (black dots) on top of the 1000 Genomes Reference Data (colored dots), we can confirm that the self-reported ancestries of our donor samples are correct. Ancestry data is important as it can help explain trait variance related to genetic variation and subsequently evaluate the impact of linkage disequilibrium.

4.3.2 Data quality

4.3.2.1 RNA samples

Prior to submitting RNA samples to the New York Genome Center (NYGC) for RNA-sequencing, quality control was performed to assess the integrity of the samples. The quality of an RNA sample is especially important because the RNA molecule is less stable and degrades much more quickly than DNA for two main reasons: RNA is more reactive due to its 2′-hydroxyl group on the pentose ring and enzymes that degrade RNA, ribonucleases, are highly abundant in nature. Thus, a measure of quality, termed the 'RNA integrity number,' or RIN, is calculated after extraction and before use of the sample for further experimentation. The RIN was developed by Agilent Technologies in 2006 (Schroeder *et al.* 2006), and utilizes a bioanalyzer to perform a combination of microfluidics, microcapillary electrophoresis, and fluorescence detection of an RNA sample. There are many detailed features of a sample that are recorded by the bioanalyzer, which uses a software algorithm to integrate the entirety of a RNA's electrophoretic trace. This culminates in a degradation score, with 10 being the least degraded and 1 being the most (Schroeder *et al.* 2006).

The NYGC requires samples to have a RIN of at least 7 to proceed with sample preparation and sequencing, as higher quality RNA samples require less amplification and thus can decrease the number of amplification errors. Subsequently, we used the Agilent TapeStation to collect RINs for all RNA samples. The average RIN was 9.8, with the lowest being 8.8 and the highest 10. The majority of the samples had a RIN over 9.4, with the largest number of samples with a RIN of 10 (**Figure 4.3A**). With these high quality scores, we were able to confidently submit the RNA samples to NYGC for sample prep and sequencing.

Following RNA sequencing, sequencing depths were analyzed as a measure of post-sequencing sample quality. Sequencing depth, or read depth, describes the number of times that a given nucleotide is read during sequencing. Sequencing depth varies by instrument – from a few million reads on a Miniseq to tens or hundreds of millions on a deep sequencer such as a Novaseq – and the deeper the sequencing, the more confidence in base identity. Subsequently, we requested a sequencing depth of 80 million reads per sample. Our samples had an average sequencing depth of 103,856,738 reads, with the lowest at 70,647,361 reads and the highest at 238,293,701 reads (**Figure 4.3B**). Such high sequencing depth results in an increased number of informational reads and thus the statistical power to detect differential gene expression.

4.3.2.2 DNA samples

Quality of DNA samples was assessed prior to submitting to the UNC-CH Mammalian Genotyping Core. Purity was evaluated using a Nanodrop to quantify the ratio of sample absorbance at 260 (A260) and 280nm (A280). DNA absorbs wavelengths at 260 and 280nm and the ratio of A260 to A280 indicates potential contamination of a sample. For DNA, a ratio of ~1.7 to ~2.0 is considered pure. Lower than 1.7 may indicate contamination with residual phenol or other reagents associated with the extraction process. Higher than 2 may indicate the presence of RNA in a DNA sample. Our samples had an average A260/A280 ratio of 1.97, with the lowest being 1.87 and the highest 2.14 (**Figure 4.4**). Although 2.14 is slightly high, it is within the acceptable range of ratios.

Following genotyping, computational quality control was performed to evaluate the number of SNPs that passed (to retain SNPs with missing genotype rate less than 10%, deviations from Hardy-Weinberg equilibrium at a p-value greater than 1 x 10e-6, minor allele frequency greater than 1%, and

sufficient imputation quality (R2 > 0.3); further details can be found in the dissertation of Nicole Kramer). The final dataset contained 10,419,216 autosomal variants for 79 donor samples (although we now have 105 donors collected, at the time of data analysis, only 79 donor samples had data availability).

4.3.3 Differential RNA analyses confirm FN-f treatment drives major transcriptional changes

In order to confirm that any transcriptional changes occurring in chondrocyte samples treated with FN-f were due to FN-f treatment, several differential RNA analyses were performed. A principal component analysis (PCA) of global gene expression revealed two separate sample clusters based on treatment with FN-f (**Figure 4.5A, pink**) vs PBS (**Figure 4.5A, blue**), validating that gene expression changes seen were condition-specific. In line with this, the gene expression profile of transcriptome-wide gene expression changes after FN-f treatment displayed that donor-specific factors such as age, sex, and race did not contribute to differential gene expression (**Figure 4.5B**). Furthermore, the gene expression profile revealed the up- and down-regulation of several genes implicated in OA, such as *WNT10B*, *MMP13*, *CXCL2*, *GDF5*, and *COLA2* (**Figure 4.5B**), upholding our treatment of chondrocytes with FN-f as an appropriate model of OA.

To identify the phenotypic functions and pathways in which the FN-f-responsive genes are enriched in, we also performed gene ontology (GO) term (Alexa and Rahnenfuhrer 2021) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis (Tenenbaum and Maintainer 2021) (**Figure 4.5C**). Differentially expressed genes were enriched in processes related to stimulus response and signaling, inflammatory response and cytokine production, and cell morphogenesis as well as in TNF, IL-17, and NF-kappa B pathways. These pathways were highlighted by Reed et al. (2021), in which our *in vitro* model of OA was initially described. These results further corroborate our use of FN-f treatment of chondrocytes in order to simulate transcriptional changes as seen in OA.

4.4 Discussion

eQTL analyses are a way of annotating genetic variants that affect gene expression and are associated with disease. Studying these eQTLs in response to disease-appropriate stimuli can provide deeper insights into the genetic variation that drives disease progression. Therefore, we used a cell model of OA in which we induce an inflammatory response in chondrocytes with fibronectin fragment (FN-f) (Reed *et al.* 2021) with the intent of capturing cell-type- and disease-state-specific eQTLs.

This project was a joint effort between myself, my lab mate Nicole Kramer, and our collaborators in the lab of Dr. Richard Loeser. My work on this project focused specifically on the collection and generation of high quality genotype and transcriptional data for the use in eQTL analyses. The Loeser lab isolated primary articular chondrocytes from human talar cartilage obtained from tissue donors and treated with FN-f or PBS as a control. From these 105 donors, I collected both DNA and RNA and performed quality control analyses to confirm the quality of the samples. Genotyping was performed by the UNC-CH Mammalian Genotyping Core and RNA-sequencing was performed by the New York Genome Center. Following collection of data, further eQTL analyses were performed by Nicole Kramer in the Phanstiel lab.

With the generation of high quality data, we were able to confirm that the transcriptional changes occurring in FN-f treated chondrocytes were due to stimulation with the fragment alone, that FN-f treatment induced a robust, distinctive transcriptional change compared to PBS as a control, and that differentially expressed genes were enriched in processes related to stimulus response and signaling, inflammatory response and cytokine production, and cell morphogenesis as well as in TNF, IL-17, and NF-kappa B pathways. These results validate the use of the chondrocyte FN-f model system to recapitulate the OA phenotype.

The results of the genotyping and transcriptional data will provide the basis for eQTL mapping. The genotyping data allows for detection of genetic variants within a population, such as our collection of donors. When combined with the RNA-sequencing data obtained from FN-f- and PBS-treated samples, we can putatively describe which genetic variants are significantly associated with OA. These findings within our controlled *in vitro* model of OA will identify novel gene candidates and point to possible mechanisms by which disease-associated variants influence disease-specific gene expression and subsequently disease progression. We hope that these results will aid in the discovery of novel OA therapeutic targets.

4.5 Materials and Methods

4.5.1 Primary chondrocyte isolation and culture

Human primary articular chondrocytes were isolated via enzymatic digestion from human talar cartilage obtained from tissue donors, without a history of arthritis, through the Gift of Hope Organ and

Tissue Donor Network (Elmhurt, IL) as previously described (Loeser *et al.* 2003; Reed *et al.* 2021). After isolation, $1x10^6$ chondrocytes were plated in 6-well plates in 1:1 DMEM/F12 supplemented with 10% FBS and incubated for four days prior to treatment.

4.5.2 Fibronectin fragment (FN-f) treatment[1]

Chondrocytes were washed twice with PBS and replaced with 1mL serum-free 1:1 DMEM/F12, then incubated for two hours. Serum-free media was then transferred to centrifuge tubes and either treated with 12.5µL PBS or 12.5µL FN-f (final concentration 1µM) prior to transferring back into respective wells. Cells were incubated for 18 hours before proceeding.

4.5.3 RNA lysis collection before extraction and QC[1]

After FN-f treatment, the work area was first cleaned with RNase inhibitor RNaseZap (Sigma Aldrich, cat. no. R2020-250ML). The supernatant was aspirated from the plates and cells were washed once with PBS. 350µL RNA buffer RLT from the Qiagen RNeasy Mini Kit (cat. no. 74104) was added to each well and plates were swirled to ensure complete coverage. RNA lysate was transferred to microcentrifuge tubes, vortexed, and stored at -80°C.

4.5.4 Cell collection before DNA extraction and genotyping

After FN-f treatment, cells were washed twice with PBS prior to trypsinization and quenching with 1:1 DMEM/F12 supplemented with 10% FBS. Media with cells was transferred to microcentrifuge tubes and centrifuged at 300xg for 10 minutes. Supernatant was aspirated and cells were flash frozen in liquid nitrogen prior to storage at -80°C.

4.5.5 RNA extraction, QC, and transfer to New York Genome Center for RNA-Seq

Frozen lysates were thawed on the bench at room temperature and 1 volume of freshly made 70% ethanol was added. RNA was extracted using the RNeasy Mini Kit (Qiagen, cat. no. 74104) following the Quick Protocol for RNeasy Mini Kit, part 1 for cells. On-column DNase digestion was performed following steps 1-4 in the Quick Protocol for RNeasy Mini Kit, part 2. RNA was eluted in 30µL RNase-free water.

QC was performed using the Qubit RNA HS Assay kit (Thermo Fisher Scientific, cat. no.

[1]These methods were performed by Philip R. Coryell.

Q32852). The RNA integrity numbers were confirmed to be greater than or equal to 7 using a TapeStation High Sensitivity RNA ScreenTape. At least 2µg in 15-30µL volume of samples were sent to the New York Genome Center for prep (KAPA Stranded RNA-seq with RiboErase (HMR)) and RNA-sequencing (Novaseq 6000 2x100bp, 80 million reads per sample).

4.5.6 Genomic DNA extraction, QC, and genotyping

Frozen cells were resuspended in PBS to a final volume of 200µL. Genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen, cat. no. 51304) according to the manufacturer's instructions for genomic DNA purification from cells (adapted from Qiagen's "QIAamp DNA Mini and Blood Mini Handbook"). Final samples were eluted in 30µL nuclease-free water and stored at -20°C prior to QC. DNA quantification was performed both using the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, cat. no. Q32851) and a nanodrop. Absorbance was recorded at 260, 280, 230, and 270. Samples were transferred to the University of North Carolina at Chapel Hill's Mammalian Genotyping Core and genotyped using the Infinium Global Diversity Array-8 v.10 Kit (Illumina, 20031669).

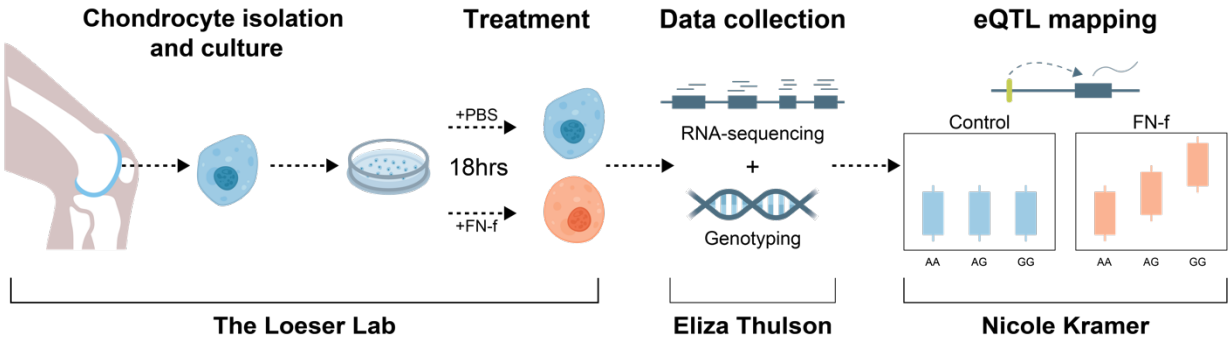

Figure 4.1. eQTL study design. This project was performed in collaboration with the lab of Dr. Richard Loeser and Nicole Kramer of the Phanstiel lab. The Loeser lab isolated human chondrocytes from donors of healthy human articular talar cartilage, cultured cells, and treated with either PBS as a control or fibronectin fragment (FN-f) for 18 hours to simulate a cell-type-relevant osteoarthritis phenotype. I collected data for RNA-sequencing and genotyping by extracting RNA and DNA and performing quality control analysis. Nicole Kramer performed eQTL mapping with the intent of identifying disease-specific eQTLs.

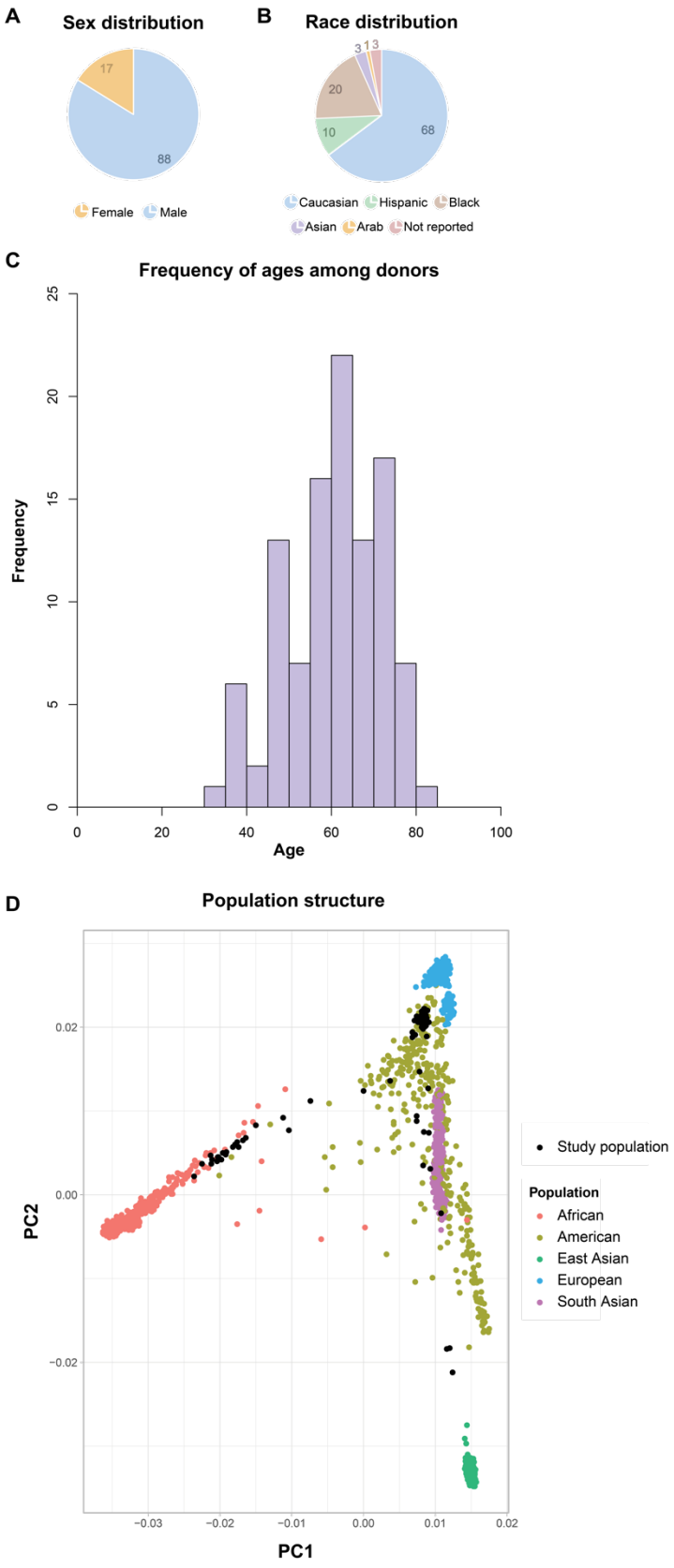
A
Sex distribution
17
88
Female
Male
B
Race distribution
3
1
3
20
10
68
Caucasian
Hispanic
Black
Asian
Arab
Not reported
C
Frequency of ages among donors
Frequency
0
5
10
15
20
25
0
20
40
60
80
100
Age
D
Population structure
PC2
0.02
0.00
−0.02
−0.03
−0.02
−0.01
0.00
0.01
0.02
PC1
Study population
Population
African
American
East Asian
European
South Asian

Figure 4.2. Clinical characteristics of human talar cartilage tissue donors. Human primary articular chondrocytes were isolated via enzymatic digestion from healthy human talar cartilage obtained from tissue donors through the Gift of Hope Organ and Tissue Donor Network (Elmhurt, IL) as previously described (Loeser *et al.* 2003; Reed *et al.* 2021). **(A)** Sex distribution. Of the 105 donors collected for samples, approximately 16% were female and 84% were male. **(B)** Race distribution. The majority of donors were Caucasian (~65%), while 19% were Black, 10% were Hispanic, 3% were Asian, 1% were Arab, and 3% were not disclosed. Races are described as reported in the pathologist's report. **(C)** Frequency of donor ages. The average donor age was 56, with the youngest being 34 and the oldest being 84. **(D)** Principal component analysis of sample genotypes (for 79 donors collected at the time) overlaid with 1000 Genomes Data as assessed by Nicole Kramer. Data from 1000 Genomes are colored by superpopulation and samples from our study are colored in black.

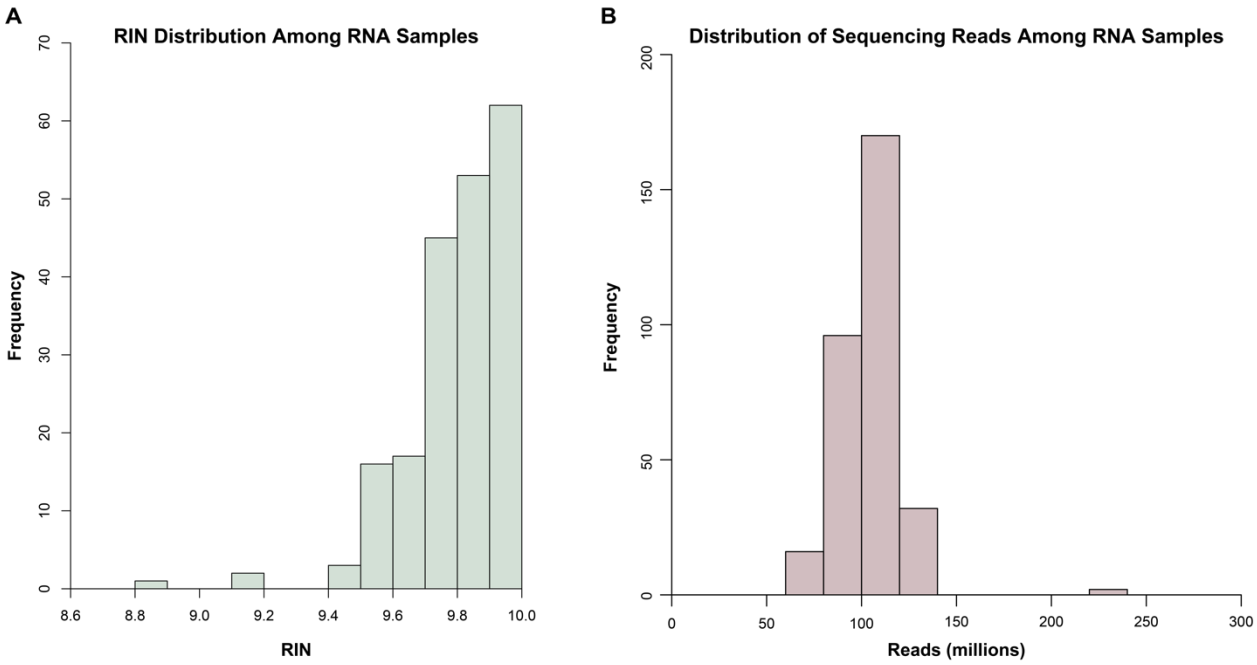

Figure 4.3. RNA-sample and sequencing quality control results. Chondrocytes from donors were treated with either PBS (control or "CTL") or FN-f ("FNF") for 18 hours and RNA was extracted. **(A)** Quality control (QC) was performed to obtain the RNA integrity numbers (RINs) via TapeStation. The frequency distribution as shown by the histogram of RINs indicates most samples had RINs of 9.6 and higher. **(B)** The number of sequencing reads were obtained following sequencing at the New York Genome Center. The frequency distribution as shown by the histogram of sequencing reads indicates most samples were sequenced to a depth of 100 million reads.

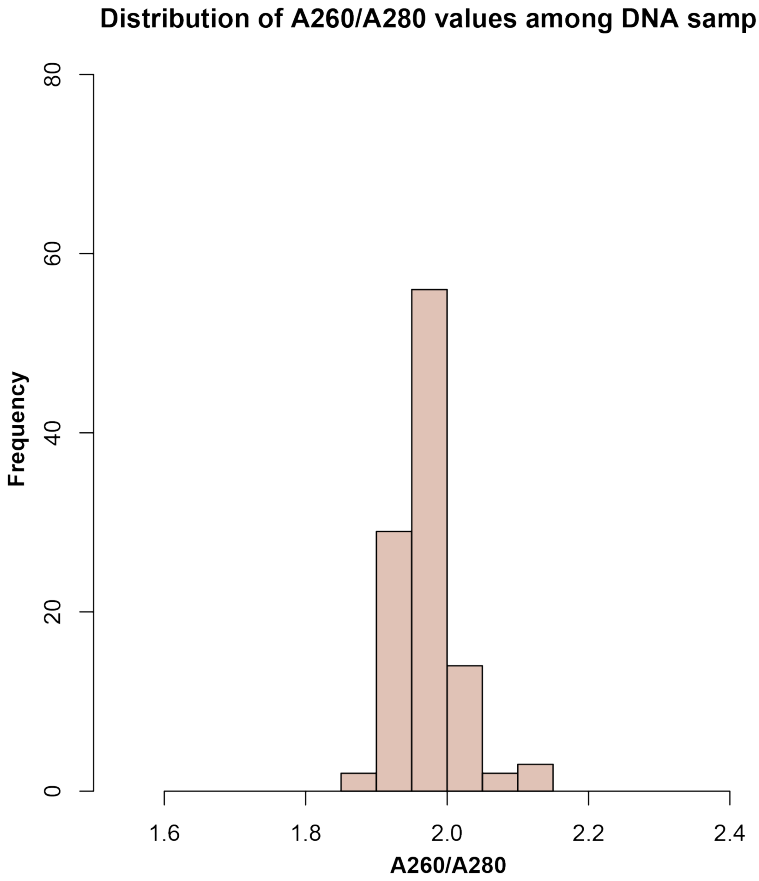

Figure 4.4. DNA sample quality control results. DNA was extracted from chondrocytes harvested from donors. Quality control (QC) was performed via Nanodrop to assess the A260/A280 ratio as a measure of DNA purity. The ratio of DNA absorbance at 260 and 280nm indicates potential contamination of a sample - a ratio of ~1.7 to ~2.0 is considered pure. Lower than 1.7 may indicate contamination with residual phenol or other reagents associated with the extraction process. Higher than 2 may indicate the presence of RNA in a DNA sample. Our samples had an average A260/A280 ratio of 1.97, with the lowest being 1.87 and the highest 2.14.

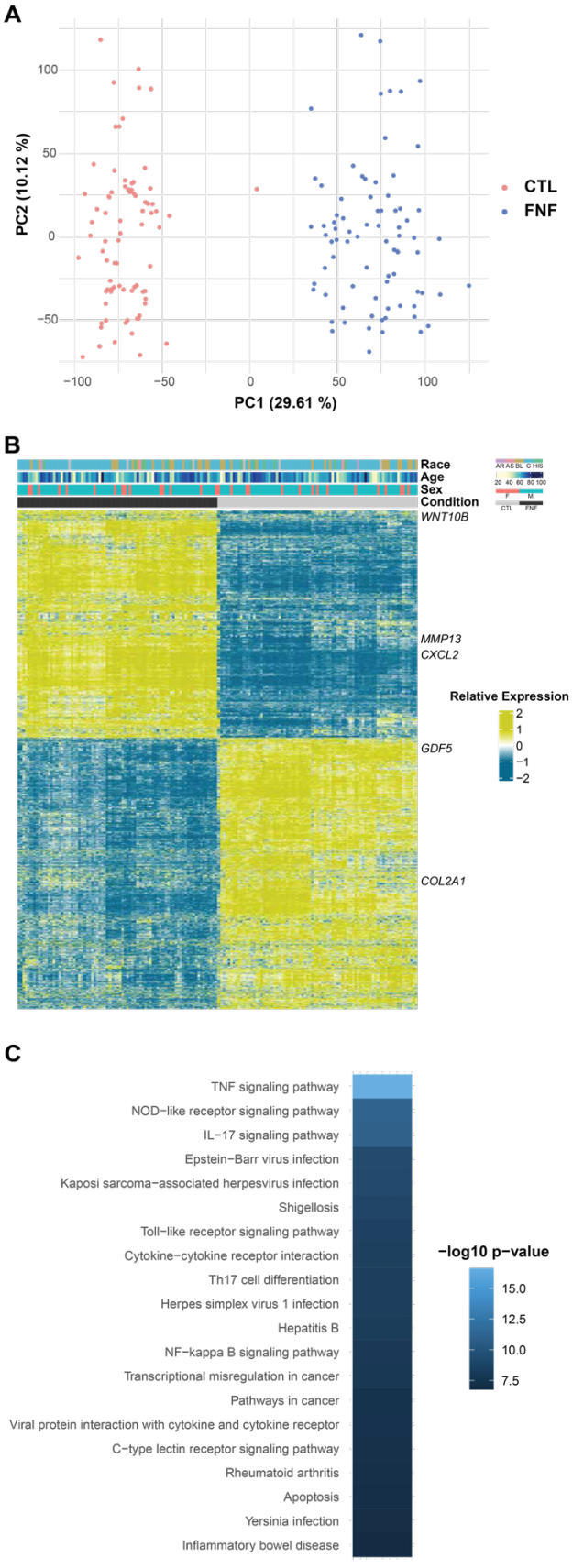
A
100
50
0
−50
PC2 (10.12 %)
−100
−50
0
50
100
PC1 (29.61 %)
CTL
FNF
B
Race
Age
Sex
Condition
WNT10B
MMP13
CXCL2
GDF5
COL2A1
Relative Expression
2
1
0
−1
−2
C
TNF signaling pathway
NOD−like receptor signaling pathway
IL−17 signaling pathway
Epstein−Barr virus infection
Kaposi sarcoma−associated herpesvirus infection
Shigellosis
Toll−like receptor signaling pathway
Cytokine−cytokine receptor interaction
Th17 cell differentiation
Herpes simplex virus 1 infection
Hepatitis B
NF−kappa B signaling pathway
Transcriptional misregulation in cancer
Pathways in cancer
Viral protein interaction with cytokine and cytokine receptor
C−type lectin receptor signaling pathway
Rheumatoid arthritis
Apoptosis
Yersinia infection
Inflammatory bowel disease
−log10 p−value
15.0
12.5
10.0
7.5

Figure 4.5. Differential RNA analyses confirm FN-f treatment drives transcriptional changes[1]. (A) Principal component analysis (PCA) of control (CTL, blue) and FN-f (pink) treatment conditions indicates clustering occurs due to condition. **(B)** Gene expression profile of transcriptome-wide gene expression changes following FN-f treatment. Genes are colored by their relative expression (yellow to blue). Previously implicated OA genes are indicated (Reed *et al.* 2021). Samples are labeled for self-reported race, age, and sex as well as condition ("CTL" or "FNF"). **(C)** Top 20 KEGG pathways enriched in response to FN-f treatment. The heatmap color represents -log10 p-value enrichment for the labeled pathway.

[1]RNA analyses were performed by Nicole Kramer of the Phanstiel lab.

CHAPTER 5: DISCUSSION

5.1 Summary

My thesis work focuses on long range gene regulation in human health and disease, combining several techniques for interrogating enhancer activity, 3D chromatin structure, and the relationship between genotype and phenotype in the context of disease-relevant cell types and states. The findings and data collection presented here address the critical question of how 3D chromatin interactions contribute to gene expression in specific biological contexts, such as cellular activation and stimulation. I achieved this both by exploring a single genomic locus during a specific cellular developmental state as well as by providing the foundation for identifying and studying variation at many genomic loci and their target genes in the broader context of a disease phenotype. The results ultimately address the transcriptional control that takes place in all nucleated human cells and offer insight into how extensive cellular diversity can be achieved from one genome.

5.2 Characterization of the role of Activator Protein 1 (AP-1) in loop-based gene regulation

5.2.1 AP-1 loops to and activates *IL1β* during monocyte-to-macrophage differentiation

The role of Activator Protein 1 (AP-1), a heterodimeric transcription factor made up of varying combinations of FOS, JUN, MAF, ATF and CREB family proteins, has been well-documented in the differentiation of hematopoietic cell lineages (Liebermann *et al.* 1998; Valledor *et al.* 1998). While AP-1 has been shown to bind at the anchors of chromatin loops (Chavanas *et al.* 2008; Qiao *et al.* 2015), its participation in long-range gene regulation via DNA looping in macrophages has only recently been described (Phanstiel *et al.* 2017). Here we provide further insights into the mechanisms behind AP-1-mediated gene regulation through 3D chromatin structure at the Interlekin 1 beta (*IL1β*) locus, a key macrophage regulatory gene.

IL1β is a proinflammatory cytokine secreted by a variety of cell types, including macrophages, during immune response to infection and injury (Dinarello 1996). Phanstiel et. al (2017) demonstrated the formation of a novel loop at the *IL1β* locus during monocyte-to-macrophage differentiation, with enriched

binding of AP-1 alongside the canonical loop factor CTCF. These results imply a role for AP-1 in dynamic promoter-distal gene regulation during macrophage development. I propose a model in which CTCF forms new loops during monocyte-to-macrophage differentiation, bringing enhancer-bound AP-1 into close physical proximity with macrophage-specific genes, such as *IL1β*, to activate transcription.

5.2.2 Interrogating the mechanism by which AP-1 activates *IL1β*

The exact mechanism by which AP-1 participates in this form of loop-based gene regulation at the *IL1β* locus remains an open question. While our preliminary results indicate a role for the binding of AP-1 at the first upstream *IL1β* enhancer, it is still unclear if and how the second upstream enhancer contributes to *IL1β* expression. To help elucidate whether *IL1β*'s enhancers have separate, redundant, and/or additive functions, *IL1β* expression in THP-1 mutants with AP-1 binding site deletions at the first (E1 AP-1 del), second (E2 AP-1 del), and first and second enhancers (E1+E2 AP-1 del) could be compared during monocyte-to-macrophage differentiation. *IL1β* expression in E1 AP-1 deletion mutants should recapitulate what our preliminary results show – that removal of AP-1 binding sites on one and then both alleles leads to decreased *IL1β* expression during macrophage differentiation. I would expect to see a similar decline in *IL1β* expression in E2 AP-1 deletion mutants, with an even more severe decrease in transcriptional output in E1+E2 AP-1 deletion mutants. Hi-C^2 could be performed alongside qPCR for gene expression in these mutants to confirm that the DNA loop connecting *IL1β* with its enhancers is not affected by deletion of the AP-1 binding sites.

Additionally, the CTCF binding sites at the *IL1β* enhancers could be deleted, and qPCR and Hi-C performed in these mutants during macrophage development. Execution of these experiments in CTCF binding site mutants alongside the AP-1 binding site mutants would further corroborate our hypothesis of a CTCF-driven looping event culminating in *IL1β* activation by subsequent AP-1 binding. Furthermore, given the expected decrease in *IL1β* expression in these CTCF and AP-1 binding site mutants, canonical inflammatory macrophage pathways should be evaluated so as to gauge the impact of these mutations on macrophage development. Such pathways might include genes such as V-maf musculoaponeurotic fibrosarcoma oncogene homolog B (*MAFB*), an AP-1 subunit that plays a role in macrophage homeostasis (Hamada *et al.* 2020), Toll-like receptors (*TLRs*), cell-surface membrane receptors that

functions in detection of foreign pathogens in the innate immune response (Medzhitov 2001), and interferon-regulatory factors (*IRFs*), intracellular proteins that regulate immune cell maturation and polarization (Tamura *et al.* 2008; Günthner and Anders 2013).

5.2.3 Exploring the composition of AP-1 family members at the *IL1β* locus

Given its many functions in diverse biological processes (Shaulian and Karin 2002; Eferl and Wagner 2003), the specific composition of the AP-1 complex may contribute to its interactions on DNA in a cell-type and tissue-specific manner. However, many of the AP-1 subunits share consensus binding motifs and thus it has been challenging to differentiate between family member specificity (Mechta-Grigoriou *et al.* 2001; Isakova *et al.* 2017). Subsequently, it may prove difficult to perform chromatin immunoprecipitation (ChIP) for a specific AP-1 family member. Enzyme-catalyzed proximity-labeling technologies, such as BioID, TurboID, and APEX, have served as tools for studying protein-protein interactions. These technologies work by utilizing promiscuous labeling enzymes fused to a specific protein to initiate tagging of endogenous proteins in close proximity to the labeling enzyme. These tagged interacting proteins can then be pulled down and identified via mass spectrometry (MS) (Roux *et al.* 2012, 2013, 2018; Kim *et al.* 2014; Kim and Roux 2016; Branon *et al.* 2018; Trinkle-Mulcahy 2019). However, these technologies do not identify protein-protein interactions at specific genomic locations as ChIP does. A 2016 study by Schmidtmann et al. developed a proximity-labeling approach to identify chromatin interacting proteins at specific DNA sequences, termed CasID (Schmidtmann *et al.* 2016). By combining the RNA-programmable DNA binding capabilities of dCas9 with the proximity-dependent biotin identification function of the promiscuous biotin ligase enzyme, the CasID approach can be used to biotinylate and identify proteins at specific genomic loci (Schmidtmann *et al.* 2016). This provides an alternative method to ChIP, which requires suitable antibodies and provides global rather than locus-specific data.

The CasID tool could be used to identify which AP-1 family members are binding at the *IL1β* locus – and potentially other macrophage-specific genes – and functioning during macrophage differentiation. This would shed significant light into the precise proteins involved in AP-1-specific looping events. By identifying non-redundant biological roles for distinct AP-1 subunits which bind a common

recognition motif, we can further elucidate the mechanisms underlying the collaborative interactions of AP-1 contributing to its transcriptional specificity.

5.3 Interrogating the non-coding genome for the identification of osteoarthritis (OA) risk genes

5.3.1 *SOCS2* and other genes putatively contribute to OA progression

While over 100 loci have been identified that are associated with OA, the majority of OA risk variants are non-coding, making it difficult to identify the impacted genes. To identify causal OA risk variants and their target genes, we leveraged an *ex vivo* model of OA whereby primary human chondrocytes are stimulated with fibronectin fragment (FN-f) to identify and functionally characterize potential OA risk genes. By combining OA GWAS, H3K27ac ChIP-seq, RNAseq, and 3D chromatin structure data in our model of OA, we were able to construct a refined set of putative causal OA variants and their putative target genes.

One of these genes was Suppressor of Cytokine Signaling 2 (*SOCS2*), which we discovered was connected by a 174 Kb loop to the OA risk variant rs7953280. *SOCS2* has been shown to have a role in the resolution of inflammatory response and is downregulated in knee OA tissue (de Andrés *et al.* 2011; Paul *et al.* 2017). Importantly, rs7953280 is located within a peak of H3K27ac in an intron of the *CRADD* gene. While *CRADD* is not expressed in chondrocytes during FN-f treatment, *SOCS2* changes in expression over the course of stimulation. Furthermore, rs7953280 changes the sequence of an AP-1 binding motif, and our binding predictions indicate that Jun, an AP-1 subunit, binds to the non-risk but not the risk allele. Taken together, these data imply that rs7953280 may be a putative causal variant that affects *SOCS2* expression in an OA environment. The negative impact of *SOCS2* deletion on inflammatory gene response further corroborates our hypothesized model.

Other genes we identified via Hi-C that looped to OA risk variants include Fibroblast Growth Receptor 3 (*FGFR3*), Tropomysin 1 (*TPM1*) and Ral Guanine Nucleotide Dissociation Stimulator Like 1 (*RGL1*). *FGFR3* has been previously implicated in OA, as it has been found to be downregulated in OA tissues (Li *et al.* 2012; Shu *et al.* 2016). The association of *TPM1* and *RGL1* with OA specifically is novel; however, these genes have been shown to modulate inflammatory responses in various cell types (Kirkby *et al.* 2014; Gagat *et al.* 2021; Li *et al.* 2022). It would be interesting to further investigate these genes in the context of our *ex vivo* OA model. In a similar fashion to our experiments in *SOCS2* knockout

chondrocytes, the impact of *FGFR3, TPM1, and RGL1* knockouts on inflammatory gene response could be explored. Their potential for influencing disease risk may point to these genes and their pathways as additional therapeutic targets for future studies.

5.3.2 Validation of putative causal variants and their target genes in our *ex vivo* model of OA

While we have preliminary evidence for our model of decreased AP-1 binding at the risk allele of rs7953280 influencing *SOCS2* expression, further work is needed to clarify the role of this risk SNP and *SOCS2* in mediating OA risk. To that end, future experiments should focus on the genome editing of non-coding sequences in order to directly examine the effect of risk variants on their putative target genes – such as in the case of rs7953280 and *SOCS2*. However, using CRISPR-Cas9 technology to create point mutations is technically challenging due to the mechanisms of the intracellular repair systems homology-directed repair (HDR) and nonhomologous end joining (NHEJ). HDR and NHEJ, triggered by the double-stranded breaks (DSBs) induced by Cas9, often result in unintended insertions and deletions (Yeh *et al.* 2019). While this may or may not prove an issue for the deletion of large genomic regions, editing single base pairs requires both precision and accuracy.

Recent advances in CRISPR have explored genome editing without inducing DSBs and subsequently triggering HDR and NHEJ (Komor *et al.* 2016; Gaudelli *et al.* 2017; Anzalone *et al.* 2019). The first of these approaches utilizes a base editor fused to either a catalytically dead Cas9 (dCas9) or a nickase Cas9 (nCas9) targeted to a specific location via a guide RNA. These base editors typically modify C → T and G → A (cytidine deaminase) (Komor *et al.* 2016) and A → G and T → C (Gaudelli *et al.* 2017). A 2021 study from Chen et al. also describes a C → G and G → C system which takes advantage of cytidine deaminase in conjunction with other base excision repair proteins (Chen *et al.* 2021). This would prove useful for our interest in rs7953280 (a C in the risk allele and a G in the non-risk allele).
Another approach for inducing single base substitutions is prime editing. Instead of a deaminase, this method involves a prime editing extended guide RNA (pegRNA)-guided reverse transcriptase fused to nCas9 (Anzalone *et al.* 2019). The pegRNA comprises three specific sequences: one that guides nCas9 to the target DNA region, a template containing the desired edit, and a primer binding sequence. The complex binds to the target DNA and nCas9 nicks one strand, generating a DNA flap. The pegRNA's primer binding sequence binds to this flap, and then the template containing the desired edit on the

pegRNA is reverse transcribed. This edited sequence is incorporated into the DNA at the end of the nicked flap, and the target DNA is repaired with the newly reverse transcribed DNA. This leaves a mismatch, which is ultimately repaired to match the newly edited sequence (Anzalone *et al.* 2019; Scholefield and Harrison 2021). Recent studies have validated the functionality of prime editing for creating specific point mutations; for example, in the gene responsible for Duchenne Muscular Dystrophy (Happi Mbakam *et al.* 2022) and in other various disease models such as liver cancer and Wilson disease (Schene *et al.* 2020).

These newer approaches to CRISPR that forgo the creation of DSBs would require protocol optimization. However, they provide promising alternative options to canonical CRISPR-Cas9 techniques for editing single bases. Taking advantages of base editing and prime editing CRISPR systems by engineering edited chondrocytes with both the risk and protective allele would allow us to confidently investigate the association between rs7953280 and *SOCS2*, as well as other gene-variant pairs identified.

5.4 QTLs for investigating non-coding disease-associated SNPs

5.4.1 Powering eQTLs with an increased sample size

In general, the power to detect significant disease- or trait-associated loci in a population is increased with a larger sample size, both within GWAS and eQTL studies (Spencer *et al.* 2009; Huang *et al.* 2018; Wang and Xu 2019; Chitre *et al.* 2023). Subsequently, continuing to collect samples from human talar cartilage tissue donors remains a high priority. As of this time, we have collected both genotyping and RNA-seq data from 105 unique donors. We expect that as we continue to collect data from a larger number of samples, the number of loci significantly associated with OA discovered will increase. Ideally, we would obtain at least 150 donors for genotyping and RNA-sequencing.

5.4.2 Colocalization of eQTLs, caQTLs, and GWAS loci to identify disease-causing SNPs and genes

Chromatin accessibility QTLs (caQTLs) are similar in concept to eQTLs, but instead of affecting gene expression, these genetic variants/loci affect nucleosome packing, position, and overall chromatin accessibility (Degner *et al.* 2012; Kumasaka *et al.* 2016). In order to more confidently identify candidate causal SNPs and genes, colocalization may be utilized, in which we see traits (eg gene expression and chromatin accessibility) sharing a putative causal variant (Giambartolomei *et al.* 2014; Hormozdiari *et al.* 2016). This integration of GWAS hits with both caQTLs and eQTLs would provide deeper insights into the

regulatory mechanisms underlying the risk for developing OA whereby noncoding risk variants affect transcription factor binding to regulatory elements such as enhancers, affecting chromatin accessibility (caQTLs) and also affect transcriptional output (eQTLs).

Subsequently, identification of caQTLs would require performing assay for transposase-accessible chromatin using sequencing (ATAC-seq) in our *ex vivo* model of OA. To that end, the Phanstiel lab at this time has collected both FN-f and PBS-treated samples from 9 donors, with an ultimate goal of collecting samples from 20 donors. This number is significantly lower than that of the donors required for eQTL analysis for several reasons: the first being that ATAC is a much more intensive protocol and data collection is time-consuming and more expensive, and the second being that caQTLs are easier to detect from ATAC data due to fewer factors affecting the binding of transcription factors within an ATAC peak compared to the numerous factors that contribute to gene expression. Following caQTL mapping in our OA model, we can utilize transcription factor motif analysis in conjunction with our colocalization analysis to identify risk variants that alter regulatory function as well as gene expression.

5.4.3 Validation of predicted causal variants and genes using Hi-C and genome editing

After identification of putative causal SNPs via colocalization of caQTLs, eQTLs, and GWAS data, the use of Hi-C could be employed to connect these variants with potential target genes, in a manner similar to how our discovery of *SOCS2* looped to the OA risk SNP rs7953280. The role of these target genes can be assessed by generating populations of heterozygous and homozygous knockout cell line and quantifying changes in inflammatory response pathways and other OA-related biological processes. To model the impacts of the risk variants, genome editing would also be performed to engineer clonal populations with the risk and protective alleles in the form of single base substitutions. We would expect to see changes in target gene expression (as measured by qPCR) due to these modifications. These experiments would result in the further characterization of new OA risk variants and the genes they regulate, paving the way for continued study in OA pathogenesis.

REFERENCES

1000 Genomes Project Consortium, G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, *et al.*, 2010 A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.

Ahn J. H., E. S. Davis, T. A. Daugird, S. Zhao, I. Y. Quiroga, *et al.*, 2021 Phase separation drives aberrant chromatin looping and cancer development. Nature 595: 591–595.

Akhtar-Zaidi B., R. Cowper-Sal-lari, O. Corradin, A. Saiakhova, C. F. Bartels, *et al.*, 2012 Epigenomic enhancer profiling defines a signature of colon cancer. Science 336: 736–739.

Albert F. W., and L. Kruglyak, 2015 The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16: 197–212.

Alexa A., and J. Rahnenfuhrer, 2021 topGO: Enrichment Analysis for Gene Ontology

Allfrey V. G., R. Faulkner, and A. E. Mirsky, 1964 Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl. Acad. Sci. U. S. A. 51: 786–794.

Amano T., T. Sagai, H. Tanabe, Y. Mizushina, H. Nakazawa, *et al.*, 2009 Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev. Cell 16: 47–57.

Anderson E., S. Peluso, L. A. Lettice, and R. E. Hill, 2012 Human limb abnormalities caused by disruption of hedgehog signaling. Trends Genet. 28: 364–373.

Andrés M. C. de, K. Imagawa, K. Hashimoto, A. Gonzalez, M. B. Goldring, *et al.*, 2011 Suppressors of cytokine signalling (SOCS) are reduced in osteoarthritis. Biochem. Biophys. Res. Commun. 407: 54–59.

Angel P., and M. Karin, 1991 The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. Biochim. Biophys. Acta Rev. Cancer 1072: 129–157.

Anzalone A. V., P. B. Randolph, J. R. Davis, A. A. Sousa, L. W. Koblan, *et al.*, 2019 Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576: 149–157.

Arner E., C. O. Daub, K. Vitting-Seerup, R. Andersson, B. Lilje, *et al.*, 2015 Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science 347: 1010–1014.

Aubourg G., S. J. Rice, P. Bruce-Wootton, and J. Loughlin, 2022 Genetics of osteoarthritis. Osteoarthritis Cartilage 30: 636–649.

Balboa D., J. Weltner, S. Eurola, R. Trokovic, K. Wartiovaara, *et al.*, 2015 Conditionally stabilized dCas9 activator for controlling gene expression in human cell reprogramming and differentiation. Stem Cell Reports 5: 448–459.

Banerji J., S. Rusconi, and W. Schaffner, 1981 Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27: 299–308.

Banerji J., L. Olson, and W. Schaffner, 1983 A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33: 729–740.

Bank E. M., and Y. Gruenbaum, 2011 The nuclear lamina and heterochromatin: a complex relationship. Biochem. Soc. Trans. 39: 1705–1709.

Bannister A. J., and T. Kouzarides, 2011 Regulation of chromatin by histone modifications. Cell Res. 21: 381–395.

Berkum N. L. van, E. Lieberman-Aiden, L. Williams, M. Imakaev, A. Gnirke, *et al.*, 2010 Hi-C: a method to study the three-dimensional architecture of genomes. J. Vis. Exp. https://doi.org/10.3791/1869

Bernstein B. E., J. A. Stamatoyannopoulos, J. F. Costello, B. Ren, A. Milosavljevic, *et al.*, 2010 The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. 28: 1045–1048.

Blackwood E. M., and J. T. Kadonaga, 1998 Going the distance: a current view of enhancer action. Science 281: 60–63.

Boer C. G., K. Hatzikotoulas, L. Southam, L. Stefánsdóttir, Y. Zhang, *et al.*, 2021 Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell 184: 6003–6005.

Boeuf S., E. Steck, K. Pelttari, T. Hennig, A. Buneb, *et al.*, 2008 Subtractive gene expression profiling of articular cartilage and mesenchymal stem cells: serpins as cartilage-relevant differentiation markers. Osteoarthritis Cartilage 16: 48–60.

Bosch F., and R. Dalla-Favera, 2019 Chronic lymphocytic leukaemia: from genetics to treatment. Nat. Rev. Clin. Oncol. 16: 684–701.

Bosch M. H. J. van den, P. L. E. M. van Lent, and P. M. van der Kraan, 2020 Identifying effector molecules, cells, and cytokines of innate immunity in OA. Osteoarthritis Cartilage 28: 532–543.

Branon T. C., J. A. Bosch, A. D. Sanchez, N. D. Udeshi, T. Svinkina, *et al.*, 2018 Efficient proximity labeling in living cells and organisms with TurboID. Nat. Biotechnol. 36: 880–887.

Breda L., I. Motta, S. Lourenco, C. Gemmo, W. Deng, *et al.*, 2016 Forced chromatin looping raises fetal hemoglobin in adult sickle cells to higher levels than pharmacologic inducers. Blood 128: 1139–1143.

Brown J. M., J. Green, R. P. das Neves, H. A. C. Wallace, A. J. H. Smith, *et al.*, 2008 Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. J. Cell Biol. 182: 1083–1097.

Bulger M., and M. Groudine, 1999 Looping versus linking: toward a model for long-distance gene activation. Genes Dev. 13: 2465–2477.

Bulger M., and M. Groudine, 2010 Enhancers: the abundance and function of regulatory sequences beyond promoters. Dev. Biol. 339: 250–257.

Cano-Gamez E., and G. Trynka, 2020 From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. Front. Genet. 11: 424.

Capellini T. D., H. Chen, J. Cao, A. C. Doxey, A. M. Kiapour, *et al.*, 2017 Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. Nat. Genet. 49: 1202–1210.

Carnemolla B., M. Cutolo, P. Castellani, E. Balza, S. Raffanti, *et al.*, 1984 Characterization of synovial fluid fibronectin from patients with rheumatic inflammatory diseases and healthy subjects. Arthritis Rheum. 27: 913–921.

Carninci P., T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, *et al.*, 2005 The transcriptional landscape of the mammalian genome. Science 309: 1559–1563.

Caron M. M. J., P. J. Emans, D. A. M. Surtel, P. M. van der Kraan, L. W. van Rhijn, *et al.*, 2015 BAPX-1/NKX-3.2 Acts as a Chondrocyte Hypertrophy Molecular Switch in Osteoarthritis. Arthritis & Rheumatology 67: 2944–2956.

Carter D., L. Chakalova, C. S. Osborne, Y.-F. Dai, and P. Fraser, 2002 Long-range chromatin regulatory interactions in vivo. Nat. Genet. 32: 623–626.

Chandrasekaran A. P., M. Song, K.-S. Kim, and S. Ramakrishna, 2018 Different methods of delivering CRISPR/Cas9 into cells. Prog. Mol. Biol. Transl. Sci. 159: 157–176.

Chanput W., J. J. Mes, and H. J. Wichers, 2014 THP-1 cell line: an in vitro cell model for immune modulation approach. Int. Immunopharmacol. 23: 37–45.

Chavanas S., V. Adoue, M.-C. Méchin, S. Ying, S. Dong, *et al.*, 2008 Long-range enhancer associated with chromatin looping allows AP-1 regulation of the peptidylarginine deiminase 3 gene in differentiated keratinocyte. PLoS One 3: e3408.

Chen L., J. E. Park, P. Paa, P. D. Rajakumar, H.-T. Prekop, *et al.*, 2021 Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. Nat. Commun. 12: 1384.

Cheng A. W., H. Wang, H. Yang, L. Shi, Y. Katz, *et al.*, 2013 Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. 23: 1163–1171.

Cheng J.-H., D. Z.-C. Pan, Z. T.-Y. Tsai, and H.-K. Tsai, 2015 Genome-wide analysis of enhancer RNA in gene regulation across 12 mouse tissues. Sci. Rep. 5: 12648.

Chesi A., Y. Wagley, M. E. Johnson, E. Manduchi, C. Su, *et al.*, 2019 Genome-scale Capture C promoter interactions implicate effector genes at GWAS loci for bone mineral density. Nat. Commun. 10: 1260.

Chinenov Y., and T. K. Kerppola, 2001 Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. Oncogene 20: 2438–2452.

Chitre A. S., O. Polesskaya, D. Munro, R. Cheng, P. Mohammadi, *et al.*, 2023 Exponential increase in QTL detection with increased sample size. bioRxiv.

Claussnitzer M., S. N. Dankel, K.-H. Kim, G. Quon, W. Meuleman, *et al.*, 2015 FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N. Engl. J. Med. 373: 895–907.

Collins J. A., L. Arbeeva, S. Chubinskaya, and R. F. Loeser, 2019 Articular chondrocytes isolated from the knee and ankle joints of human tissue donors demonstrate similar redox-regulated MAP kinase and Akt signaling. Osteoarthritis Cartilage 27: 703–711.

Cremer T., M. Cremer, S. Dietzel, S. Müller, I. Solovei, *et al.*, 2006 Chromosome territories--a functional nuclear landscape. Curr. Opin. Cell Biol. 18: 307–316.

Cremer T., and C. Cremer, 2006 Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories. Eur. J. Histochem. 50: 161–176.

Cuartero S., F. D. Weiss, G. Dharmalingam, Y. Guo, E. Ing-Simmons, *et al.*, 2018 Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. Nat. Immunol. 19: 932–941.

Curran T., and B. R. Franza Jr, 1988 Fos and jun: The AP-1 connection. Cell 55: 395–397.

Daigneault M., J. A. Preston, H. M. Marriott, M. K. B. Whyte, and D. H. Dockrell, 2010 The identification of markers of macrophage differentiation in PMA-stimulated THP-1 cells and monocyte-derived macrophages. PLoS One 5: e8668.

David T., S. F. Ling, and A. Barton, 2018 Genetics of immune-mediated inflammatory diseases. Clin. Exp. Immunol. 193: 3–12.

Davidson I. F., and J.-M. Peters, 2021 Genome folding through loop extrusion by SMC complexes. Nat. Rev. Mol. Cell Biol. 22: 445–464.

Davies J. O. J., J. M. Telenius, S. J. McGowan, N. A. Roberts, S. Taylor, *et al.*, 2016 Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat. Methods 13: 74–80.

D'Costa S., M. J. Rich, and B. O. Diekman, 2020 Engineered Cartilage from Human Chondrocytes with Homozygous Knockout of Cell Cycle Inhibitor p21. Tissue Eng. Part A 26: 441–449.

De Santa F., I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, *et al.*, 2010 A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. 8: e1000384.

Degner J. F., A. A. Pai, R. Pique-Regi, J.-B. Veyrieras, D. J. Gaffney, *et al.*, 2012 DNase I sensitivity QTLs are a major determinant of human expression variation. Nature 482: 390–394.

Dehairs J., A. Talebi, Y. Cherifi, and J. V. Swinnen, 2016 CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. Sci. Rep. 6. https://doi.org/10.1038/srep28973

Dekker J., K. Rippe, M. Dekker, and N. Kleckner, 2002 Capturing chromosome conformation. Science 295: 1306–1311.

Deng W., J. Lee, H. Wang, J. Miller, A. Reik, *et al.*, 2012 Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell 149: 1233–1244.

Deng W., J. W. Rupon, I. Krivega, L. Breda, I. Motta, *et al.*, 2014 Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell 158: 849–860.

Dillon N., T. Trimborn, J. Strouboulis, P. Fraser, and F. Grosveld, 1997 The effect of distance on long-range chromatin interactions. Mol. Cell 1: 131–139.

Dinarello C. A., 1996 Biologic basis for interleukin-1 in disease. Blood 87: 2095–2147.

Dokka S., D. Toledo, X. Shi, J. Ye, and Y. Rojanasakul, 2000 High-efficiency gene transfection of macrophages by lipoplexes. Int. J. Pharm. 206: 97–104.

Dorsett D., 1999 Distant liaisons: long-range enhancer-promoter interactions in Drosophila. Curr. Opin. Genet. Dev. 9: 505–514.

Dostie J., T. A. Richmond, R. A. Arnaout, R. R. Selzer, W. L. Lee, *et al.*, 2006 Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. 16: 1299–1309.

Dostie J., and J. Dekker, 2007 Mapping networks of physical interactions between genomic elements using 5C technology. Nat. Protoc. 2: 988–1002.

Doudna J. A., and E. Charpentier, 2014 Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346: 1258096.

Duan A., H. Wang, Y. Zhu, Q. Wang, J. Zhang, *et al.*, 2021 Chromatin architecture reveals cell type-specific target genes for kidney disease risk variants. BMC Biol. 19: 38.

Durand N. C., M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, *et al.*, 2016 Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3: 95–98.

Edelmann P., H. Bornfleth, D. Zink, T. Cremer, and C. Cremer, 2001 Morphology and dynamics of chromosome territories in living cells. Biochim. Biophys. Acta Rev. Cancer 1551: M29–M39.

Eferl R., and E. F. Wagner, 2003 AP-1: a double-edged sword in tumorigenesis. Nat. Rev. Cancer 3: 859–868.

El-Fattah Ibrahim S. A., A. Abudu, E. Johnson, N. Aftab, S. Conrad, *et al.*, 2019 Correction: The role of AP-1 in self-sufficient proliferation and migration of cancer cells and its potential impact on an autocrine/paracrine loop. Oncotarget 10: 799.

ENCODE Project Consortium, 2012 An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74.

ENCODE Project Consortium, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, *et al.*, 2020 Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583: 699–710.

Engel J. D., and K. Tanimoto, 2000 Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. Cell 100: 499–502.

Fang R., M. Yu, G. Li, S. Chee, T. Liu, *et al.*, 2016 Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. Cell Res. 26: 1345–1348.

Fisch K. M., R. Gamini, O. Alvarez-Garcia, R. Akagi, M. Saito, *et al.*, 2018 Identification of transcription factors responsible for dysregulated networks in human osteoarthritis cartilage by global gene expression analysis. Osteoarthritis Cartilage 26: 1531–1538.

Forsyth C. B., J. Pulai, and R. F. Loeser, 2002 Fibronectin fragments and blocking antibodies to alpha2beta1 and alpha5beta1 integrins stimulate mitogen-activated protein kinase signaling and increase collagenase 3 (matrix metalloproteinase 13) production by human articular chondrocytes. Arthritis Rheum. 46: 2368–2376.

Francis-West P. H., J. Parish, K. Lee, and C. W. Archer, 1999 BMP/GDF-signalling interactions during synovial joint development. Cell Tissue Res. 296: 111–119.

Fuller S. A., M. Takahashi, and J. G. Hurrell, 2001 Cloning of hybridoma cell lines by limiting dilution. Curr. Protoc. Mol. Biol. Chapter 11: Unit11.8.

Fullwood M. J., M. H. Liu, Y. F. Pan, J. Liu, H. Xu, *et al.*, 2009 An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 462: 58–64.

Furniss D., L. A. Lettice, I. B. Taylor, P. S. Critchley, H. Giele, *et al.*, 2008 A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. Hum. Mol. Genet. 17: 2417–2423.

Furth R. van, J. A. Raeburn, and T. L. van Zwet, 1979 Characteristics of human mononuclear phagocytes. Blood 54: 485–500.

Gagat M., W. Zielińska, K. Mikołajczyk, J. Zabrzyński, A. Krajewski, *et al.*, 2021 CRISPR-Based Activation of Endogenous Expression of TPM1 Inhibits Inflammatory Response of Primary Human Coronary Artery Endothelial and Smooth Muscle Cells Induced by Recombinant Human Tumor Necrosis Factor α. Frontiers in Cell and Developmental Biology 9.

Gao X., Y. Sun, and X. Li, 2019 Identification of key gene modules and transcription factors for human osteoarthritis by weighted gene co-expression network analysis. Exp. Ther. Med. 18: 2479–2490.

Gaudelli N. M., A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, *et al.*, 2017 Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551: 464–471.

Giambartolomei C., D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, *et al.*, 2014 Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10: e1004383.

Gibney E. R., and C. M. Nolan, 2010 Epigenetics and gene expression. Heredity (Edinb.) 105: 4–13.

Gifford C. A., M. J. Ziller, H. Gu, C. Trapnell, J. Donaghey, *et al.*, 2013 Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell 153: 1149–1163.

Gilbert L. A., M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, *et al.*, 2014 Genome-scale CRISPR-mediated control of gene repression and activation. Cell 159: 647–661.

Gillies S. D., S. L. Morrison, V. T. Oi, and S. Tonegawa, 1983 A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33: 717–728.

Gnirke A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, *et al.*, 2009 Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. 27: 182–189.

Goding J. W., 1980 Antibody production by hybridomas. J. Immunol. Methods 39: 285–308.

Goetze R. W., D.-H. Kim, R. F. Schinazi, and B. Kim, 2017 A CRISPR/Cas9 approach reveals that the polymerase activity of DNA polymerase β is dispensable for HIV-1 infection in dividing and nondividing cells. J. Biol. Chem. 292: 14016–14025.

Gonen N., C. R. Futtner, S. Wood, S. A. Garcia-Moreno, I. M. Salamone, *et al.*, 2018 Sex reversal following deletion of a single distal enhancer of Sox9. Science 360: 1469–1473.

Gough S. M., C. I. Slape, and P. D. Aplan, 2011 NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. Blood 118: 6247–6257.

Gregersen P. K., and L. M. Olsson, 2009 Recent advances in the genetics of autoimmune disease. Annu. Rev. Immunol. 27: 363–391.

Gröschel S., M. A. Sanders, R. Hoogenboezem, E. de Wit, B. A. M. Bouwman, *et al.*, 2014 A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 157: 369–381.

GTEx Consortium, 2013 The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45: 580–585.

Günthner R., and H.-J. Anders, 2013 Interferon-regulatory factors determine macrophage phenotype polarization. Mediators Inflamm. 2013: 731023.

Gupta S., J. A. Stamatoyannopoulos, T. L. Bailey, and W. S. Noble, 2007 Quantifying similarity between motifs. Genome Biol. 8: R24.

Haarhuis J. H. I., R. H. van der Weide, V. A. Blomen, J. O. Yáñez-Cuna, M. Amendola, *et al.*, 2017 The cohesin release factor WAPL restricts chromatin loop extension. Cell 169: 693-707.e14.

Hamada M., Y. Tsunakawa, H. Jeon, M. K. Yadav, and S. Takahashi, 2020 Role of MafB in macrophages. Exp. Anim. 69: 1–10.

Hanahan D., 1985 Heritable formation of pancreatic beta-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes. Nature 315: 115–122.

Happi Mbakam C., J. Rousseau, G. Tremblay, P. Yameogo, and J. P. Tremblay, 2022 Prime editing permits the introduction of specific mutations in the gene responsible for Duchenne Muscular Dystrophy. Int. J. Mol. Sci. 23: 6160.

Hatzis P., and I. Talianidis, 2002 Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol. Cell 10: 1467–1477.

Hawkins R. D., G. C. Hon, L. K. Lee, Q. Ngo, R. Lister, *et al.*, 2010 Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 6: 479–491.

Heintzman N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, *et al.*, 2009 Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459: 108–112.

Herendeen D. R., G. A. Kassavetis, and E. P. Geiduschek, 1992 A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. Science 256: 1298–1303.

Hildebrand E. M., and J. Dekker, 2020 Mechanisms and functions of chromosome compartmentalization. Trends Biochem. Sci. 45: 385–396.

Homandberg G. A., C. Wen, and F. Hui, 1998 Cartilage damaging activities of fibronectin fragments derived from cartilage and synovial fluid. Osteoarthritis Cartilage 6: 231–244.

Homandberg G. A., 1999 Potential regulation of cartilage metabolism in osteoarthritis by fibronectin fragments. Front. Biosci. 4: D713-30.

Hormozdiari F., M. van de Bunt, A. V. Segrè, X. Li, J. W. J. Joo, *et al.*, 2016 Colocalization of GWAS and eQTL signals detects target genes. Am. J. Hum. Genet. 99: 1245–1260.

Huang Q. Q., S. C. Ritchie, M. Brozynska, and M. Inouye, 2018 Power, false discovery rate and Winner's Curse in eQTL studies. Nucleic Acids Res. 46: e133.

Hughes J. R., N. Roberts, S. McGowan, D. Hay, E. Giannoulatou, *et al.*, 2014 Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet. 46: 205–212.

Hunter D. J., and S. Bierma-Zeinstra, 2019 Osteoarthritis. Lancet 393: 1745–1759.

International Human Genome Sequencing Consortium, 2004 Finishing the euchromatic sequence of the human genome. Nature 431: 931–945.

Isakova A., R. Groux, M. Imbeault, P. Rainer, D. Alpern, *et al.*, 2017 SMiLE-seq identifies binding motifs of single and dimeric transcription factors. Nat. Methods 14: 316–322.

Jacob F., and J. Monod, 1961 Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3: 318–356.

Jäger R., G. Migliorini, M. Henrion, R. Kandaswamy, H. E. Speedy, *et al.*, 2015 Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. Nat. Commun. 6. https://doi.org/10.1038/ncomms7178

Jain S., V. Gautam, and S. Naseem, 2011 Acute-phase proteins: As diagnostic tool. J. Pharm. Bioallied Sci. 3: 118–127.

Kaikkonen M. U., N. J. Spann, S. Heinz, C. E. Romanoski, K. A. Allison, *et al.*, 2013 Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol. Cell 51: 310–325.

Kampmann M., 2018 CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine. ACS Chem. Biol. 13: 406–416.

Keller A.-A., M. B. Maeß, M. Schnoor, B. Scheiding, and S. Lorkowski, 2018 Transfecting macrophages. Methods Mol. Biol. 1784: 187–195.

Kiel D. P., J. P. Kemp, F. Rivadeneira, J. J. Westendorf, D. Karasik, *et al.*, 2020 The Musculoskeletal Knowledge Portal: Making Omics Data Useful to the Broader Scientific Community. J. Bone Miner. Res. 35: 1626–1633.

Kim T.-K., M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, *et al.*, 2010 Widespread transcription at neuronal activity-regulated enhancers. Nature 465: 182–187.

Kim D. I., K. C. Birendra, W. Zhu, K. Motamedchaboki, V. Doye, *et al.*, 2014 Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc. Natl. Acad. Sci. U. S. A. 111: E2453-61.

Kim D. I., and K. J. Roux, 2016 Filling the void: Proximity-based labeling of proteins in living cells. Trends Cell Biol. 26: 804–817.

King M. C., and A. C. Wilson, 1975 Evolution at two levels in humans and chimpanzees. Science 188: 107–116.

Kirkby N. S., M. H. Lundberg, W. R. Wright, T. D. Warner, M. J. Paul-Clark, *et al.*, 2014 COX-2 protects against atherosclerosis independently of local vascular prostacyclin: identification of COX-2 associated pathways implicate Rgl1 and lymphocyte networks. PLoS One 9: e98165.

Klopocki E., C.-E. Ott, N. Benatar, R. Ullmann, S. Mundlos, *et al.*, 2008 A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. J. Med. Genet. 45: 370–375.

Knight P. A., and D. Ruiz, 2013 A fast algorithm for matrix balancing. IMA J. Numer. Anal. 33: 1029–1047.

Koch C. M., R. M. Andrews, P. Flicek, S. C. Dillon, U. Karaöz, *et al.*, 2007 The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res. 17: 691–707.

Komor A. C., Y. B. Kim, M. S. Packer, J. A. Zuris, and D. R. Liu, 2016 Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533: 420–424.

Kouzarides T., 2007 Chromatin modifications and their function. Cell 128: 693–705.

Kragesteen B. K., M. Spielmann, C. Paliou, V. Heinrich, R. Schöpflin, *et al.*, 2018 Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat. Genet. 50: 1463–1473.

Kramer N. E., E. S. Davis, C. D. Wenger, E. M. Deoudes, S. M. Parker, *et al.*, 2022 Plotgardener: Cultivating precise multi-panel figures in R. Bioinformatics. https://doi.org/10.1093/bioinformatics/btac057

Krivega I., and A. Dean, 2016 Chromatin looping as a target for altering erythroid gene expression. Ann. N. Y. Acad. Sci. 1368: 31–39.

Kumasaka N., A. J. Knights, and D. J. Gaffney, 2016 Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat. Genet. 48: 206–213.

Kvon E. Z., Y. Zhu, G. Kelman, C. S. Novak, I. Plajzer-Frick, *et al.*, 2020 Comprehensive in vivo interrogation reveals phenotypic impact of human enhancer variants. Cell 180: 1262-1271.e15.

Laarman M. D., G. Geeven, P. Barnett, N. Null, G. J. E. Rinkel, *et al.*, 2019 Chromatin Conformation Links Putative Enhancers in Intracranial Aneurysm–Associated Regions to Potential Candidate Genes. J. Am. Heart Assoc. 8: e011201.

Labun K., T. G. Montague, J. A. Gagnon, S. B. Thyme, and E. Valen, 2016 CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res. 44: W272-6.

Labun K., T. G. Montague, M. Krause, Y. N. Torres Cleuren, H. Tjeldnes, *et al.*, 2019 CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. 47: W171–W174.

Lakshmipathy U., S. Buckley, and C. Verfaillie, 2007 Gene transfer via nucleofection into adult and embryonic stem cells. Methods Mol. Biol. 407: 115–126.

Lamkanfi M., and V. M. Dixit, 2012 Inflammasomes and their roles in health and disease. Annu. Rev. Cell Dev. Biol. 28: 137–161.

Lander E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, *et al.*, 2001 Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Lattanzi A., V. Meneghini, G. Pavani, F. Amor, S. Ramadier, *et al.*, 2019 Optimization of CRISPR/Cas9 delivery to human hematopoietic stem and progenitor cells for therapeutic genomic rearrangements. Mol. Ther. 27: 137–150.

Lawrence M., S. Daujat, and R. Schneider, 2016 Lateral thinking: How histone modifications regulate gene expression. Trends Genet. 32: 42–56.

Lettice L. A., T. Horikoshi, S. J. H. Heaney, M. J. van Baren, H. C. van der Linde, *et al.*, 2002 Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. U. S. A. 99: 7548–7553.

Lettice L. A., S. J. H. Heaney, L. A. Purdie, L. Li, P. de Beer, *et al.*, 2003 A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. 12: 1725–1735.

Lettice L. A., I. Williamson, P. S. Devenney, F. Kilanowski, J. Dorin, *et al.*, 2014 Development of five digits is controlled by a bipartite long-range cis-regulator. Development 141: 1715–1725.

Unaltered Differentiation Potential of Leukemia and Lymphoma Cell Lines HL-60, THP-1 and U-937 after Nucleofection™, 2014 Lonza.

Li H., C.-Y. Wang, J.-X. Wang, G.-S. Wu, P. Yu, *et al.*, 2009 Mutation analysis of a large Chinese pedigree with congenital preaxial polydactyly. Eur. J. Hum. Genet. 17: 604–610.

Li G., M. J. Fullwood, H. Xu, F. H. Mulawadi, S. Velkov, *et al.*, 2010 ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol. 11: R22.

Li X., M. B. Ellman, J. S. Kroin, D. Chen, D. Yan, *et al.*, 2012 Species-specific biological effects of FGF-2 in articular cartilage: implication for distinct roles within the FGF receptor family. J. Cell. Biochem. 113: 2532–2542.

Li X., O. J. Luo, P. Wang, M. Zheng, D. Wang, *et al.*, 2017 Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. Nat. Protoc. 12: 899–915.

Li R., Y. Liang, and B. Lin, 2022 Accumulation of systematic TPM1 mediates inflammation and neuronal remodeling by phosphorylating PKA and regulating the FABP5/NF-κB signaling pathway in the retina of aged mice. Aging Cell 21: e13566.

Lieberman-Aiden E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, *et al.*, 2009 Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326: 289–293.

Liebermann D. A., B. Gregory, and B. Hoffman, 1998 AP-1 (Fos/Jun) transcription factors in hematopoietic differentiation and apoptosis. Int. J. Oncol. 12: 685–700.

Lin A. C., B. L. Seeto, J. M. Bartoszko, M. A. Khoury, H. Whetstone, *et al.*, 2009 Modulating hedgehog signaling can attenuate the severity of osteoarthritis. Nature Medicine 15: 1421–1425.

Lin D., J. Sanders, and W. S. Noble, 2021 HiCRep.py : Fast comparison of Hi-C contact matrices in Python. Bioinformatics. https://doi.org/10.1093/bioinformatics/btab097

Lindner B., E. Martin, M. Steininger, A. Bundalo, M. Lenter, *et al.*, 2021 A genome-wide CRISPR/Cas9 screen to identify phagocytosis modulators in monocytic THP-1 cells. Sci. Rep. 11: 12973.

Loeser R. F., C. A. Pacione, and S. Chubinskaya, 2003 The combination of insulin-like growth factor 1 and osteogenic protein 1 promotes increased survival of and matrix synthesis by normal and osteoarthritic human articular chondrocytes. Arthritis Rheum. 48: 2188–2196.

Loeser R. F., S. R. Goldring, C. R. Scanzello, and M. B. Goldring, 2012 Osteoarthritis: a disease of the joint as an organ. Arthritis Rheum. 64: 1697–1707.

Loeser R. F., 2014 Integrins and chondrocyte-matrix interactions in articular cartilage. Matrix Biol. 39: 11–16.

Lopez-Castejon G., and D. Brough, 2011 Understanding the mechanism of IL-1β secretion. Cytokine Growth Factor Rev. 22: 189–195.

Luger K., A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, 1997 Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389: 251–260.

Lupiáñez D. G., K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, *et al.*, 2015 Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161: 1012–1025.

MacGregor A. J., and T. D. Spector, 1999 Twins and the genetic architecture of osteoarthritis. Rheumatology (Oxford) 38: 583–588.

Mahmoudi T., K. R. Katsani, and C. P. Verrijzer, 2002 GAGA can mediate enhancer function in trans by linking two separate DNA molecules. EMBO J. 21: 1775–1781.

Martinet W., D. M. Schrijvers, and M. M. Kockx, 2003 Nucleofection as an efficient nonviral transfection method for human monocytic cells. Biotechnol. Lett. 25: 1025–1029.

Maurano M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, *et al.*, 2012 Systematic localization of common disease-associated variation in regulatory DNA. Science 337: 1190–1195.

McGovern A., S. Schoenfelder, P. Martin, J. Massey, K. Duffus, *et al.*, 2016 Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23. Genome Biol. 17: 212.

Mechta-Grigoriou F., D. Gerald, and M. Yaniv, 2001 The mammalian Jun proteins: redundancy and specificity. Oncogene 20: 2378–2389.

Meddens C. A., M. Harakalova, N. A. M. van den Dungen, H. Foroughi Asl, H. J. Hijma, *et al.*, 2016 Systematic analysis of chromatin interactions at disease associated loci links novel candidate genes to inflammatory bowel disease. Genome Biol. 17: 247.

Medzhitov R., 2001 Toll-like receptors and innate immunity. Nat. Rev. Immunol. 1: 135–145.

Mercola M., X. F. Wang, J. Olsen, and K. Calame, 1983 Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. Science 221: 663–665.

Metcalf D., C. J. Greenhalgh, E. Viney, T. A. Willson, R. Starr, *et al.*, 2000 Gigantism in mice lacking suppressor of cytokine signalling-2. Nature 405: 1069–1073.

Mifsud B., F. Tavares-Cadete, A. N. Young, R. Sugar, S. Schoenfelder, *et al.*, 2015 Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat. Genet. 47: 598–606.

Miyamoto Y., A. Mabuchi, D. Shi, T. Kubo, Y. Takatori, *et al.*, 2007 A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. Nat. Genet. 39: 529–533.

Montague T. G., J. M. Cruz, J. A. Gagnon, G. M. Church, and E. Valen, 2014 CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. 42: W401-7.

Monti-Rocha R., A. Cramer, P. Gaio Leite, M. M. Antunes, R. V. S. Pereira, *et al.*, 2018 SOCS2 Is Critical for the Balancing of Immune Response and Oxidate Stress Protecting Against Acetaminophen-Induced Acute Liver Injury. Front. Immunol. 9: 3134.

Moreau P., R. Hen, B. Wasylyk, R. Everett, M. P. Gaub, *et al.*, 1981 The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. 9: 6047–6068.

Morgan M., M. Carlson, D. Tenenbaum, and S. Arora, 2017 AnnotationHub: Client to access AnnotationHub resources. R package version 2.

Motomura H., S. Seki, S. Shiozawa, Y. Aikawa, M. Nogami, *et al.*, 2018 A selective c-Fos/AP-1 inhibitor prevents cartilage destruction and subsequent osteophyte formation. Biochem. Biophys. Res. Commun. 497: 756–761.

Müller I., S. Boyle, R. H. Singer, W. A. Bickmore, and J. R. Chubb, 2010 Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. PLoS One 5: e11560.

Mumbach M. R., A. J. Rubin, R. A. Flynn, C. Dai, P. A. Khavari, *et al.*, 2016 HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods 13: 919–922.

Murrell A., S. Heeson, and W. Reik, 2004 Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat. Genet. 36: 889–893.

Muthuirulan P., D. Zhao, M. Young, D. Richard, Z. Liu, *et al.*, 2021 Joint disease-specificity at the regulatory base-pair level. Nat. Commun. 12: 4161.

Nishizaki S. S., N. Ng, S. Dong, R. S. Porter, C. Morterud, *et al.*, 2020 Predicting the effects of SNPs on transcription factor binding affinity. Bioinformatics 36: 364–372.

Okura T., M. Matsushita, K. Mishima, R. Esaki, T. Seki, *et al.*, 2018 Activated FGFR3 prevents subchondral bone sclerosis during the development of osteoarthritis in transgenic mice with achondroplasia. J. Orthop. Res. 36: 300–308.

Olive M., D. Krylov, D. R. Echlin, K. Gardner, E. Taparowsky, *et al.*, 1997 A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. J. Biol. Chem. 272: 18586–18594.

Ozer G., A. Luque, and T. Schlick, 2015 The chromatin fiber: multiscale problems and approaches. Curr. Opin. Struct. Biol. 31: 124–139.

Pagès H., P. Aboyoun, R. Gentleman, and S. DebRoy, 2021 Biostrings: Efficient manipulation of biological strings

Palstra R.-J., B. Tolhuis, E. Splinter, R. Nijmeijer, F. Grosveld, *et al.*, 2003 The beta-globin nuclear compartment in development and erythroid differentiation. Nat. Genet. 35: 190–194.

Paul I., T. S. Batth, D. Iglesias-Gato, A. Al-Araimi, I. Al-Haddabi, *et al.*, 2017 The ubiquitin ligase Cullin5SOCS2 regulates NDR1/STK38 stability and NF-κB transactivation. Sci. Rep. 7: 42800.

Pelletier J. P., J. Martel-Pelletier, and S. B. Abramson, 2001 Osteoarthritis, an inflammatory disease: potential implication for the selection of new therapeutic targets. Arthritis Rheum. 44: 1237–1247.

Peslak S. A., S. Demirci, V. Chandra, B. Ryu, S. K. Bhardwaj, *et al.*, 2023 Forced enhancer-promoter rewiring to alter gene expression in animal models. Mol. Ther. Nucleic Acids 31: 452–465.

Peters D., 2014 Genome editing in human pluripotent stem cells. Stembook. https://doi.org/10.3824/stembook.1.94.1

Phanstiel D. H., K. Van Bortle, D. Spacek, G. T. Hess, M. S. Shamim, *et al.*, 2017 Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. Mol. Cell 67: 1037-1048.e6.

Pittman A. M., S. Naranjo, S. E. Jalava, P. Twiss, Y. Ma, *et al.*, 2010 Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. PLoS Genet. 6: e1001126.

Platt O. S., 2008 Hydroxyurea for the treatment of sickle cell anemia. N. Engl. J. Med. 358: 1362–1369.

Pomerantz M. M., and M. L. Freedman, 2011 The genetics of cancer risk. Cancer J. 17: 416–422.

Pradeepa M. M., G. R. Grimes, Y. Kumar, G. Olley, G. C. A. Taylor, *et al.*, 2016 Histone H3 globular domain acetylation identifies a new class of enhancers. Nat. Genet. 48: 681–686.

Ptashne M., 1986 Gene regulation by proteins acting nearby and at a distance. Nature 322: 697–701.

Pulai J. I., H. Chen, H.-J. Im, S. Kumar, C. Hanning, *et al.*, 2005 NF-κB Mediates the Stimulation of Cytokine and Chemokine Expression by Human Articular Chondrocytes in Response to Fibronectin Fragments. The Journal of Immunology 174: 5781–5788.

Purcell S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559–575.

Qi L. S., M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, *et al.*, 2013 Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152: 1173–1183.

Qiao Y., C.-N. Shiue, J. Zhu, T. Zhuang, P. Jonsson, *et al.*, 2015 AP-1-mediated chromatin looping regulates ZEB2 transcription: new insights into TNFα-induced epithelial-mesenchymal transition in triple-negative breast cancer. Oncotarget 6: 7804–7814.

Quinlan A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Quiroga I. Y., J. H. Ahn, G. G. Wang, and D. Phanstiel, 2022 Oncogenic fusion proteins and their role in three-dimensional chromatin structure, phase separation, and cancer. Curr. Opin. Genet. Dev. 74: 101901.

R Core Team, 2021 R: A Language and Environment for Statistical Computing

Rada-Iglesias A., R. Bajpai, S. Prescott, S. A. Brugmann, T. Swigut, *et al.*, 2012 Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell Stem Cell 11: 633–648.

Radford E. J., and H. V. Firth, 2019 The genetics of developmental disorders. Paediatr. Child Health (Oxford) 29: 422–430.

Rao S. S. P., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, *et al.*, 2014 A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159: 1665–1680.

Rao S. S. P., S.-C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, *et al.*, 2017 Cohesin loss eliminates all loop domains. Cell 171: 305-320.e24.

Rathjen D. A., and C. L. Geczy, 1986 Conditioned medium from macrophage cell lines supports the single-cell growth of hybridomas. Hybridoma 5: 255–261.

Reed K. S. M., V. Ulici, C. Kim, S. Chubinskaya, R. F. Loeser, *et al.*, 2021 Transcriptional response of human articular chondrocytes treated with fibronectin fragments: an in vitro model of the osteoarthritis phenotype. Osteoarthritis Cartilage 29: 235–247.

Ren C., F. Liu, Z. Ouyang, G. An, C. Zhao, *et al.*, 2017 Functional annotation of structural ncRNAs within enhancer RNAs in the human genome: implications for human disease. Sci. Rep. 7. https://doi.org/10.1038/s41598-017-15822-7

Reynard L. N., and M. J. Barter, 2020 Osteoarthritis year in review 2019: genetics, genomics and epigenetics. Osteoarthritis Cartilage 28: 275–284.

Richard D., Z. Liu, J. Cao, A. M. Kiapour, J. Willen, *et al.*, 2020 Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. Cell 181: 362-381.e28.

Roux K. J., D. I. Kim, M. Raida, and B. Burke, 2012 A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. 196: 801–810.

Roux K. J., D. I. Kim, and B. Burke, 2013 BioID: a screen for protein-protein interactions. Curr. Protoc. Protein Sci. 74: 19.23.1-19.23.14.

Roux K. J., D. I. Kim, B. Burke, and D. G. May, 2018 BioID: A screen for protein-protein interactions. Curr. Protoc. Protein Sci. 91: 19.23.1-19.23.15.

Rowley M. J., M. H. Nichols, X. Lyu, M. Ando-Kuri, I. S. M. Rivera, *et al.*, 2017 Evolutionarily conserved principles predict 3D chromatin organization. Mol. Cell 67: 837-852.e7.

Rowley M. J., A. Poulet, M. H. Nichols, B. J. Bixler, A. L. Sanborn, *et al.*, 2020 Analysis of Hi-C data using SIP effectively identifies loops in organisms from C. elegans to mammals. Genome Res. 30: 447–458.

Sagai T., M. Hosoya, Y. Mizushina, M. Tamura, and T. Shiroishi, 2005 Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development 132: 797–803.

Salzberg S. L., 2018 Open questions: How many genes do we have? BMC Biol. 16: 94.

Samvelyan H. J., C. Huesa, L. Cui, C. Farquharson, and K. A. Staines, 2022 The role of accelerated growth plate fusion in the absence of SOCS2 on osteoarthritis vulnerability. Bone Joint Res. 11: 162–170.

Sanborn A. L., S. S. P. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, *et al.*, 2015 Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. U. S. A. 112: E6456-65.

Sanchez R., and M.-M. Zhou, 2009 The role of human bromodomains in chromatin biology and gene transcription. Curr. Opin. Drug Discov. Devel. 12: 659–665.

Sanchez R., and M.-M. Zhou, 2011 The PHD finger: a versatile epigenome reader. Trends Biochem. Sci. 36: 364–372.

Sandell L. J., and T. Aigner, 2001 Articular cartilage and changes in Arthritis: Cell biology of osteoarthritis. Arthritis Res. Ther. 3: 107.

Sankaran V. G., and D. G. Nathan, 2010 Thalassemia: an overview of 50 years of clinical research. Hematol. Oncol. Clin. North Am. 24: 1005–1020.

Santangelo C., A. Scipioni, L. Marselli, P. Marchetti, and F. Dotta, 2005 Suppressor of cytokine signaling gene expression in human pancreatic islets: modulation by cytokines. Eur. J. Endocrinol. 152: 485–489.

Schene I. F., I. P. Joore, R. Oka, M. Mokry, A. H. M. van Vugt, *et al.*, 2020 Prime editing for functional repair in patient-derived disease models. Nat. Commun. 11: 5352.

Schmidt E. M., J. Zhang, W. Zhou, J. Chen, K. L. Mohlke, *et al.*, 2015 GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. Bioinformatics 31: 2601–2606.

Schmidtmann E., T. Anton, P. Rombaut, F. Herzog, and H. Leonhardt, 2016 Determination of local chromatin composition by CasID. Nucleus 7: 476–484.

Schnoor M., I. Buers, A. Sietmann, M. F. Brodde, O. Hofnagel, *et al.*, 2009 Efficient non-viral transfection of THP-1 cells. J. Immunol. Methods 344: 109–115.

Scholefield J., and P. T. Harrison, 2021 Prime editing - an update on the field. Gene Ther. 28: 396–401.

Schwarzacher T., A. R. Leitch, M. D. Bennett, and J. S. Heslop-Harrison, 1989 In situ localization of parental genomes in a wide hybrid. Ann. Bot. 64: 315–324.

Schwarzer W., N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, *et al.*, 2017 Two independent modes of chromatin organization revealed by cohesin removal. Nature 551: 51–56.

Shakoori A. R., 2017 Fluorescence in situ hybridization (FISH) and its applications, pp. 343–367 in *Chromosome Structure and Aberrations*, Springer India, New Delhi.

Shang Y., M. Myers, and M. Brown, 2002 Formation of the androgen receptor transcription complex. Mol. Cell 9: 601–610.

Shao L., X. Zuo, Y. Yang, Y. Zhang, N. Yang, *et al.*, 2019 The inherited variations of a p53-responsive enhancer in 13q12.12 confer lung cancer risk by attenuating TNFRSF19 expression. Genome Biol. 20: 103.

Shaulian E., and M. Karin, 2002 AP-1 as a regulator of cell life and death. Nat. Cell Biol. 4: E131-6.

Shen X., S. Zhang, X. Zhang, T. Zhou, and Y. Rui, 2022 Two nonsense GLI3 variants are associated with polydactyly and syndactyly in two families by affecting the sonic hedgehog signaling pathway. Mol. Genet. Genomic Med. 10: e1895.

Shu C. C., M. T. Jackson, M. M. Smith, S. M. Smith, S. Penm, *et al.*, 2016 Ablation of Perlecan Domain 1 Heparan Sulfate Reduces Progressive Cartilage Degradation, Synovitis, and Osteophyte Size in a Preclinical Model of Posttraumatic Osteoarthritis. Arthritis & Rheumatology 68: 868–879.

Simonis M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, *et al.*, 2006 Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. 38: 1348–1354.

Simonis M., J. Kooren, and W. de Laat, 2007 An evaluation of 3C-based methods to capture DNA interactions. Nat. Methods 4: 895–901.

Skene P. J., and S. Henikoff, 2017 An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife 6. https://doi.org/10.7554/eLife.21856

Slatkin M., 2008 Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9: 477–485.

Song J., D. Kim, C. H. Lee, M. S. Lee, C.-H. Chun, *et al.*, 2013 MicroRNA-488 regulates zinc transporter SLC39A8/ZIP8 during pathogenesis of osteoarthritis. J. Biomed. Sci. 20: 31.

Southam L., J. Rodriguez-Lopez, J. M. Wilkins, M. Pombo-Suarez, S. Snelling, *et al.*, 2007 An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. Hum. Mol. Genet. 16: 2226–2232.

Spencer C. C. A., Z. Su, P. Donnelly, and J. Marchini, 2009 Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 5: e1000477.

Spitz F., and E. E. M. Furlong, 2012 Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. 13: 613–626.

Stalder J., A. Larsen, J. D. Engel, M. Dolan, M. Groudine, *et al.*, 1980 Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell 20: 451–460.

Stamatoyannopoulos G., 2005 Control of globin gene expression during development and erythroid differentiation. Exp. Hematol. 33: 259–271.

Starr R., T. A. Willson, E. M. Viney, L. J. Murray, J. R. Rayner, *et al.*, 1997 A family of cytokine-inducible inhibitors of signalling. Nature 387: 917–921.

Start genome editing with CRISPR-Cas9, Integrated DNA Technologies.

Steensel B. van, and E. E. M. Furlong, 2019 The role of transcription in shaping the spatial organization of the genome. Nat. Rev. Mol. Cell Biol. 20: 327–337.

Steinberg J., L. Southam, N. C. Butterfield, T. I. Roumeliotis, A. Fontalis, *et al.*, 2020 Decoding the genomic basis of osteoarthritis. bioRxiv 835850.

Steinberg J., L. Southam, T. I. Roumeliotis, M. J. Clark, R. L. Jayasuriya, *et al.*, 2021 A molecular quantitative trait locus map for osteoarthritis. Nat. Commun. 12: 1309.

Stergachis A. B., S. Neph, A. Reynolds, R. Humbert, B. Miller, *et al.*, 2013 Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell 154: 888–903.

Su W., S. Jackson, R. Tjian, and H. Echols, 1991 DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes Dev. 5: 820–826.

Sun M., F. Ma, X. Zeng, Q. Liu, X.-L. Zhao, *et al.*, 2008 Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. J. Med. Genet. 45: 589–595.

Taberlay P. C., A. L. Statham, T. K. Kelly, S. J. Clark, and P. A. Jones, 2014 Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Res. 24: 1421–1432.

Tachmazidou I., K. Hatzikotoulas, L. Southam, J. Esparza-Gordillo, V. Haberland, *et al.*, 2019 Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat. Genet. 51: 230–236.

Tahirov T. H., K. Sato, E. Ichikawa-Iwata, M. Sasaki, T. Inoue-Bungo, *et al.*, 2002 Mechanism of c-myb–C/EBPβ cooperation from separated sites on a promoter. Cell 108: 57–70.

Tamura T., H. Yanai, D. Savitsky, and T. Taniguchi, 2008 The IRF family transcription factors in immunity and oncogenesis. Annu. Rev. Immunol. 26: 535–584.

Tang J., N. Su, S. Zhou, Y. Xie, J. Huang, *et al.*, 2016 Fibroblast Growth Factor Receptor 3 Inhibits Osteoarthritis Progression in the Knee Joints of Adult Mice. Arthritis Rheumatol 68: 2432–2443.

Tenenbaum D., and B. P. Maintainer, 2021 KEGGREST: Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG)

The Deciphering Developmental Disorders Study, 2015 Large-scale discovery of novel genetic causes of developmental disorders. Nature 519: 223–228.

THP-1: Human acute monocytic luekemia, 2016 Lonza.

Tolhuis B., R.-J. Palstra, E. Splinter, F. Grosveld, and W. de Laat, 2002 Looping and interaction between hypersensitive sites in the active β-globin locus. Mol. Cell 10: 1453–1465.

Trimarchi T., E. Bilal, P. Ntziachristos, G. Fabbri, R. Dalla-Favera, *et al.*, 2014 Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell 158: 593–606.

Trinkle-Mulcahy L., 2019 Recent advances in proximity-based labeling methods for interactome mapping. F1000Res. 8: 135.

Tuerlings M., M. Hoolwerff, E. Houtman, E. H. E. Suchiman, N. Lakenberg, *et al.*, 2021 RNA Sequencing Reveals Interacting Key Determinants of Osteoarthritis Acting in Subchondral Bone and Articular Cartilage: Identification of *IL11* and *CHADL* as Attractive Treatment Targets. Arthritis & Rheumatology 73: 789–799.

Umans B. D., A. Battle, and Y. Gilad, 2021 Where Are the Disease-Associated eQTLs? Trends Genet. 37: 109–124.

Ushiki A., Y. Zhang, C. Xiong, J. Zhao, I. Georgakopoulos-Soares, *et al.*, 2021 Deletion of CTCF sites in the SHH locus alters enhancer-promoter interactions and leads to acheiropodia. Nat. Commun. 12: 2282.

Valledor A. F., F. E. Borràs, M. Cullell-Young, and A. Celada, 1998 Transcription factors that regulate monocyte/macrophage differentiation. J. Leukoc. Biol. 63: 405–417.

Venetianer P., 2012 Are synonymous codons indeed synonymous? Biomol. Concepts 3: 21–28.

Venter J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, *et al.*, 2001 The sequence of the human genome. Science 291: 1304–1351.

Visel A., E. M. Rubin, and L. A. Pennacchio, 2009 Genomic views of distant-acting enhancers. Nature 461: 199–205.

Vorstman J. A. S., and R. A. Ophoff, 2013 Genetic causes of developmental disorders. Curr. Opin. Neurol. 26: 128–136.

Wamstad J. A., J. M. Alexander, R. M. Truty, A. Shrikumar, F. Li, *et al.*, 2012 Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell 151: 206–220.

Wang J. C., and G. N. Giaever, 1988 Action at a distance along a DNA. Science 240: 300–304.

Wang Z., C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, *et al.*, 2008 Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. 40: 897–903.

Wang S., F. Wen, G. B. Wiley, M. T. Kinter, and P. M. Gaffney, 2013 An enhancer element harboring variants associated with systemic lupus erythematosus engages the TNFAIP3 promoter to influence A20 expression. PLoS Genet. 9: e1003750.

Wang M., Z. A. Glass, and Q. Xu, 2017 Non-viral delivery of genome-editing nucleases for gene therapy. Gene Ther. 24: 144–150.

Wang X., and S. L. Thein, 2018 Switching from fetal to adult hemoglobin. Nat. Genet. 50: 478–480.

Wang T., and C. He, 2018 Pro-inflammatory cytokines: The link between obesity and osteoarthritis. Cytokine Growth Factor Rev. 44: 38–50.

Wang M., and S. Xu, 2019 Statistical power in genome-wide association studies and quantitative trait locus mapping. Heredity (Edinb.) 123: 287–306.

Wanner J., R. Subbaiah, Y. Skomorovska-Prokvolit, Y. Shishani, E. Boilard, *et al.*, 2013 Proteomic profiling and functional characterization of early and late shoulder osteoarthritis. Arthritis Res. Ther. 15: R180.

Watanabe S., M. Alexander, A. V. Misharin, and G. R. S. Budinger, 2019 The role of macrophages in the resolution of inflammation. J. Clin. Invest. 129: 2619–2628.

Weintraub H., and M. Groudine, 1976 Chromosomal subunits in active genes have an altered conformation. Science 193: 848–856.

Wendt K. S., K. Yoshida, T. Itoh, M. Bando, B. Koch, *et al.*, 2008 Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451: 796–801.

Werken H. J. G. van de, G. Landan, S. J. B. Holwerda, M. Hoichman, P. Klous, *et al.*, 2012 Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat. Methods 9: 969–972.

What is Sickle Cell Disease?, 2022 National Heart, Lung, and Blood Institute.

Wickham H., 2016 ggplot2: Elegant Graphics for Data Analysis

Wieczorek D., B. Pawlik, Y. Li, N. A. Akarsu, A. Caliebe, *et al.*, 2010 A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes Werner mesomelic syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. Hum. Mutat. 31: 81–89.

Williamson I., R. E. Hill, and W. A. Bickmore, 2011 Enhancers: from developmental genetics to the genetics of common human disease. Dev. Cell 21: 17–19.

Williamson I., R. Eskeland, L. A. Lettice, A. E. Hill, S. Boyle, *et al.*, 2012 Anterior-posterior differences in HoxD chromatin topology in limb development. Development 139: 3157–3167.

Wojdasiewicz P., Ł. A. Poniatowski, and D. Szukiewicz, 2014 The role of inflammatory and anti-inflammatory cytokines in the pathogenesis of osteoarthritis. Mediators Inflamm. 2014: 561459.

Wolffe A. P., 1998 Packaging principle: How DNA methylation and histone acetylation control the transcriptional activity of chromatin. J. Exp. Zool. 282: 239–244.

Won H., L. de la Torre-Ubieta, J. L. Stein, N. N. Parikshak, J. Huang, *et al.*, 2016 Chromosome conformation elucidates regulatory relationships in developing human brain. Nature 538: 523–527.

Wood S. T., D. L. Long, J. A. Reisz, R. R. Yammani, E. A. Burke, *et al.*, 2016 Cysteine-mediated redox regulation of cell signaling in chondrocytes stimulated with fibronectin fragments. Arthritis rheumatol. 68: 117–126.

Wu L., D. Liang, N. Niikawa, F. Ma, M. Sun, *et al.*, 2009 A ZRS duplication causes syndactyly type IV with tibial hypoplasia. Am. J. Med. Genet. A 149A: 816–818.

Xiao J., J. Deng, Q. Zhang, P. Ma, L. Lv, *et al.*, 2020 Targeting human cytomegalovirus IE genes by CRISPR/Cas9 nuclease effectively inhibits viral replication and reactivation. Arch. Virol. 165: 1827–1835.

Xie D. L., R. Meyers, and G. A. Homandberg, 1992 Fibronectin fragments in osteoarthritic synovial fluid. J. Rheumatol. 19: 1448–1452.

Xie W., M. D. Schultz, R. Lister, Z. Hou, N. Rajagopal, *et al.*, 2013 Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell 153: 1134–1148.

Xu C., X. Yang, H. Zhou, Y. Li, C. Xing, *et al.*, 2020 A novel ZRS variant causes preaxial polydactyly type I by increased sonic hedgehog expression in the developing limb bud. Genet. Med. 22: 189–198.

Yang T., F. Zhang, G. G. Yardımcı, F. Song, R. C. Hardison, *et al.*, 2017 HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. Genome Res. 27: 1939–1949.

Yang Y., P. Shen, T. Yao, J. Ma, Z. Chen, *et al.*, 2021 Novel role of circRSU1 in the progression of osteoarthritis by adjusting oxidative stress. Theranostics 11: 1877–1900.

Yao Y., J. B. Wang, M. M. Xin, H. Li, B. Liu, *et al.*, 2016 Balance between inflammatory and regulatory cytokines in systemic lupus erythematosus. Genet. Mol. Res. 15. https://doi.org/10.4238/gmr.15027626

Yap K. L., and M.-M. Zhou, 2011 Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. Biochemistry 50: 1966–1980.

Yegnasubramanian S., Z. Wu, M. C. Haffner, D. Esopi, M. J. Aryee, *et al.*, 2011 Chromosome-wide mapping of DNA methylation patterns in normal and malignant prostate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. BMC Genomics 12: 313.

Yeh C. D., C. D. Richardson, and J. E. Corn, 2019 Advances in genome editing through control of DNA repair pathways. Nat. Cell Biol. 21: 1468–1478.

Yumlu S., J. Stumm, S. Bashir, A.-K. Dreyer, P. Lisowski, *et al.*, 2017 Gene editing and clonal isolation of human induced pluripotent stem cells using CRISPR/Cas9. Methods 121–122: 29–44.

Zalatan J. G., M. E. Lee, R. Almeida, L. A. Gilbert, E. H. Whitehead, *et al.*, 2015 Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell 160: 339–350.

Zeng Z., and Y. Bromberg, 2019 Predicting Functional Effects of Synonymous Variants: A Systematic Review and Perspectives. Frontiers in Genetics 10.

Zeng L., J.-Y. Jin, F.-M. Luo, Y. Sheng, P.-F. Wu, *et al.*, 2022 ZPA regulatory sequence variants in Chinese patients with preaxial polydactyly: Genetic and clinical characteristics. Front. Pediatr. 10: 797978.

Zhao Z., G. Tavoosidana, M. Sjölinder, A. Göndör, P. Mariano, *et al.*, 2006 Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. 38: 1341–1347.

Zhou S., Y. Xie, W. Li, J. Huang, Z. Wang, *et al.*, 2016 Conditional Deletion of Fgfr3 in Chondrocytes leads to Osteoarthritis-like Defects in Temporomandibular Joint of Adult Mice. Scientific Reports 6.

Zhu J., M. Adli, J. Y. Zou, G. Verstappen, M. Coyne, *et al.*, 2013 Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell 152: 642–654.

Zuris J. A., D. B. Thompson, Y. Shu, J. P. Guilinger, J. L. Bessen, *et al.*, 2015 Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat. Biotechnol. 33: 73–80.