

Genome-wide Accessible Chromatin Mapping
for Articular Knee Cartilage Reveals
Abnormal Endochondral Ossification in the
Primary Osteoarthritis

November 2018

Liu Ye

Genome-wide Accessible Chromatin Mapping
for Articular Knee Cartilage Reveals
Abnormal Endochondral Ossification in the
Primary Osteoarthritis

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Biological Science
(Doctoral Program in Bioindustrial Sciences)

Liu Ye

Abstract

Osteoarthritis (OA) is a common joint disorder with increasing impact in an aging society. To understand its pathogenesis, molecular profiling such as RNA expression or DNA methylation for diseased tissues has been reported; however, little consensus between expression and DNA methylation has been achieved. On the other hand, chromatin variations in OA are so far unexplored, mainly due to the technical difficulties in applying traditional epigenomic tools on clinical samples.

In this study, I employed an epigenomic method, Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-seq), to characterize the genome-wide chromatin alterations associated with OA. A novel protocol to generate high-quality ATAC-seq data from fresh hard tissue, human cartilage was first established using collagenase II to study human OA disease. Approximately 50,000 chondrocytes nuclei of fresh knee joint tissues were suitable for ATAC-seq library preparation. The quality check qPCR method can be used before sequencing to test the signal to the noise ratio.

In total 109,215 accessible chromatin regions for cartilages were identified, of which 71.1% were annotated as enhancers and 16.9% were annotated as promoters. Integrating the epigenomic data of clinically relevant tissues with the publicly available genetic and transcriptomic data allowed to better understand how the identified loci may contribute to OA pathogenesis. Most of these annotated accessible enhancers were linked to their putative target genes using public datasets. With this enhancer-gene map in chondrocyte, it can better interpret the previously identified OA GWAS (Single Nucleotide Polymorphisms) SNPs or OA differential methylated loci located lie outside of the coding regions. To identify the chromatin signatures relevant to OA, differential accessibility analysis was performed between outer region of lateral tibial plateau (oLT) and inner region of medial tibial plateau (iMT), significantly differentially accessible peaks were determined with false discovery rate (FDR) ≤ 0.05 . Out of the 4,450

differentially accessible peaks, 1,565 are more accessible and 2,885 are less accessible in the damaged tissues compared to the intact tissues.

Further analyses for differential accessible enhancer regions showed bone-related enhancers are more likely to be dysregulated in OA diseased tissue. Consistently, 371 protein-coding genes that are dysregulated both at the transcriptomic (RNA-seq) and epigenomic (ATAC-seq) levels in OA with concordant direction of change. These genes are enriched for pathways regulating chondrogenesis, ossification, and mesenchymal stem cell differentiation.

Moreover, integrating ATAC-seq data with publicly available database also allowed to identify altered cis-regulatory elements as well as transcription factors binding. Taken together, abnormal self-healing in OA knee cartilage was observed in this study, suggesting induced endochondral ossification-like cartilage-to-bone conversion is a characteristic of OA progression.

In conclusion, this study demonstrated genome-wide investigations of accessible chromatin regions is powerful in probing changes of regulatory genomic elements in clinical samples relevant to a disease. These detected regulatory elements can potentially serve as biomarkers or therapeutic targets for OA.

Keywords: Osteoarthritis, epigenetic, cartilage, enhancer, chromatin accessibility, ATAC-seq, RNA-seq, DNA methylation, GWAS

Contents

Abstract.....	I
Contents	III
List of Tables	VI
List of Figures.....	VII
Abbreviations	IX
Chapter 1 General introduction	1
1.1 Osteoarthritis.....	1
1.2 Knee OA.....	2
1.2.1 Current knee OA Statistics	2
1.2.2 Risk factors of knee OA.....	2
1.2.3 Diagnosis of knee OA.....	3
1.2.4 Treatments for knee OA	3
1.3 Cartilage and subchondral bone.....	4
1.4 Epigenetic studies.....	4
1.4.1 Chromatin structure	5
1.4.2 ATAC-seq.....	5
1.5 Research objectives.....	6
Chapter 2 ATAC-seq library preparation using flash frozen and fresh hard tissue.....	13
2.1 Introduction.....	13
2.2 Materials and methods	13
2.2.1 Knee joint tissues	13
2.2.2 Cartilage processing.....	14
2.2.3 Nuclei extraction.....	14

2.2.4 Tagmentation	15
2.2.5 DNA purification	15
2.2.6 PCR amplification.....	15
2.2.7 Size selection and distribution	15
2.2.8 Quality check by qPCR.....	15
2.2.9 KAPA library quantification and Illumine sequencing	16
2.3 Results and discussion	16
2.3.1 ATAC-seq of chondrocyte from human knee cartilage	16
2.3.2 ATAC-seq library quality check	17
2.4 Summary.....	17
Chapter 3 ATAC-seq data analysis	30
3.1 Introduction.....	30
3.2 Materials and methods	31
3.2.1 ATAC-seq data processing, peak calling and quality assessment	31
3.2.2 Defining a set of unified accessible chromatin regions	31
3.2.3 Peak annotation and differentially accessible peak identification	32
3.2.4 Processing of OA associated SNPs.....	32
3.2.5 Definition of cell type-specific enhancers	32
3.2.6 Enrichment analysis for differentially accessible peaks	33
3.2.7 Linking enhancer peaks to their potential target genes.....	33
3.2.8 Integration with publicly available transcriptome data.....	33
3.2.9 Gene ontology analysis	34
3.2.10 Statistics	34
3.3 Results and Discussion.....	34
3.3.1 ATAC-seq data processing, peak calling and quality assessment	34
3.3.2 Accessible chromatin landscape highlights potential enhancers and their target genes relevant to OA.....	34
3.3.3 Identification of differentially accessible enhancers in OA.....	36

3.3.4 Motif enrichment analysis reveals transcription factors relevant to OA	37
3.3.5 Integrative transcriptomics and epigenomics analysis reveals pathways involved in OA.....	38
3.4 Summary.....	40
Chapter 4 Conclusions and future research.....	69
4.1 Conclusions.....	69
4.2 Future research	72
References.....	75
List of publications.....	91
Acknowledgements	92

List of Tables

Table 2-1 Demographic information of patients.	19
Table 2-2 QC RT-qPCR primer sequences (human).	20
Table 2-3 ATAC-seq sequencing statistics.	21
Table 3-1 Differentially accessible enhancers overlapping with OA GWAS SNPs...42	42
Table 3-2 Differentially accessible enhancers overlapping with OA DMLs.	43
Table 3-3 OA associated genes enriched in top 30 GO terms.	44
Table 3-4 Differential genes detected by RNA-seq and ATAC-seq.	45

List of Figures

Figure 1-1 Diagram of cartilage matrix major components.....	9
Figure 1-2 Articular cartilage and subchondral bone in normal and diseased conditions.....	10
Figure 1-3 Interphase of Chromatin.....	11
Figure 1-4 Tn5 tagmentation at open chromatin region.	12
Figure 2-1 Human osteoarthritis (OA) study model system.	23
Figure 2-2 Fresh primary knee OA cartilage section for ATAC-seq.....	24
Figure 2-3 Flow chart of the study design (fresh).....	25
Figure 2-4 Custom tools for frozen joint tissue sectioning and cartilage separation. .	26
Figure 2-5 Bioanalyzer profile overview of ATAC-seq DNA libraries.	27
Figure 2-6 ATAC-seq quality check with qPCR.	28
Figure 2-7 Combined 8 atients ATAC-seq library fragment size distribution.	29
Figure 3-1 The ATAC-seq pipeline schematic.	54
Figure 3-2 Normalized nucleoATAC signal aggregated over all genes shows distinct nucleosome positioning around transcription start sites (TSS).....	55
Figure 3-3 Annotation of identified accessible chromatin regions.	56
Figure 3-4 Identification of accessible chromatin regions with OA susceptible GWAS SNPs and DMLs.	57
Figure 3-5 Principal component analysis on the peak signals across the 16 samples.	58
Figure 3-6 Peaks with smaller FDR (i.e. more confident to be differentially accessible) contain more enhancer peaks.	59
Figure 3-7 Significantly differentially accessible peaks with $FDR \leq 0.05$	60
Figure 3-8 Genome browser views of dysregulated promoters and enhancers.	61
Figure 3-9 Genome browser views of consistency.	62

Figure 3-10 Enrichment of cell type-specific enhancers in differentially accessible peaks.	63
Figure 3-11 DNA motif analysis.....	64
Figure 3-12 A scheme of ATAC-seq and RNA-seq integration analysis.	65
Figure 3-13 Integration of RNA-seq and ATAC-seq.....	66
Figure 3-14 Venn diagram.	67
Figure 3-15 GO enrichment.	68
Figure 4-1 Models for OA initiation proposed to explain pathogenesis of OA.....	73
Figure 4-2 Abnormal self-healing theory in OA pathology.....	74

Abbreviations

AGEs	Advanced Glycation End products
ATAC-seq	Assay for Transposase Accessibility Chromatin sequencing
BED	Browser Extensible Data
CAGE	Cap Analysis Gene Expression
CPM	Read Count Per Million
Ct	Cycle threshold
DAG	Differential ATAC-seq peak associated Gene
DAP	Differential ATAC-seq peak
DEG	Differential Expression Gene
DML	Differentially Methylated Loci
DHS	DNase I Hypersensitive Sites
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
eQTL	expression Quantitative Trait Loci
EO	Endochondral Ossification
FANTOM	Functional Annotation of the Mammalian Genome
FANTOM CAT	FANTOM CAGE associated transcriptom
FDR	False Discovery Rate
GWAS	Genome-wide association studies
GO	Gene Ontology

iLT	Inner region of Lateral Tibial plateau
iMT	Inner region of Medial Tibial plateau
KL grade	Kellgren and Lawrence grade
LN	Liquid Nitrogen
lncRNA	Long Non-Coding RNA
MNase	Micrococcal Nuclease
MRI scan	Magnetic Resonance Imaging scan
MSC	Mesenchymal Stem Cells
NGS	Next-Generation Sequencing
OA	Osteoarthritis
oLT	Outer region of Lateral Tibial plateau
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
RA	Rheumatoid Arthritis
RNA	Ribonucleic Acid
SAM	Sequence Alignment/Map
SNP	Single Nucleotide Polymorphisms
TSS	Transcription Start Site
X-ray	X-radiation

Chapter 1 General introduction

1.1 Osteoarthritis

The literal meaning of Osteoarthritis (OA) is bone arthritis, that infers one type of arthritis and joint inflamed.

Arthritis is a common joint disease. Joints (articulations or articulate surfaces) are the locations where two or more bones meet in a sense, and there are mainly three types of joint judging on moveable, semi-moveable and unmovable, which are respectively to synovial joint, cartilaginous joint and fibrous joint. Commonly what people know and perceive are the synovial joints. Synovial joint is filled with synovial fluid and the most movable type of joint in mammal. Arthritis occurs most often in synovial joint such as the knee, hip, lower back and neck (Jacobson et al., 2008).

Arthritis is not a single disease. It could be more than 100 types (i.e. degenerative arthritis, rheumatoid arthritis, inflammatory arthritis, infectious arthritis, and metabolic arthritis). The common arthritis symptoms are joint pain, swelling, stiffness and decreased ranges of motion. Majority of the estimated arthritis is degenerative arthritis, known as osteoarthritis (Neogi, 2013) and rheumatoid arthritis (Scott et al., 2010).

Osteoarthritis (OA), a degenerative joint chronic disease, is resulting in the degeneration of joint overlying articular cartilage and underlying subchondral bone (Chou et al., 2013; Radin & Rose, 1986), also known as bone-to-bone arthritis or a physiologic imbalance induced “joint failure” (Radin & Rose, 1986). The bone under the cartilage called subchondral bone is smooth and provides support to the articular cartilage (Radin & Rose, 1986), however in the severe OA diseased joint the overlying cartilage has been worn out over time, leading to the underlying subchondral exposure and bone ends rubbing together. Apparently, the OA patients will be suffering it with the excruciating pain and local inflammation; moreover, it could further aggravate the joint degeneration and permanently damage the bone. Therefore, OA is representing one of the most common causes of chronic disability in the world.

In contrast to OA, rheumatoid arthritis (RA) is inflammatory arthritis and autoimmune disorder disease that the joints are eroded by uncontrolled inflammation, and affects the entire body (Beasley, 2012). There is less-inflammation or non-inflammation in most of the cases in OA. The case of inflammation occurred in OA is

owing to degeneration, which is a subsequent consequence of OA. In the early stage of OA, it is a condition of cartilage erosion and the bone is largely spared. It is also showing minor bone spur growth. In the late stage of OA, the breakdown products of cartilage induce a chronic inflammatory response in the synovium and tendon, induced the architecture changed by the further mechanical and inflammatory stress (Grynepas et al., 1991).

1.2 Knee OA

1.2.1 Current knee OA Statistics

OA is highly prevalent worldwide. It has been reported that about 30 percent of US population aged over 60 years has joint problems that could be attributed to primary OA – defined as aging-related, as opposed to injury-related secondary OA. A population-based Framingham Knee Osteoarthritis study suggested that knee OA increases in prevalence throughout the elderly people, and it occurs more common in women than men (Felson, 1990), with most prevalent in the knee (Anderson & Loeser, 2010). Currently, Knee OA is the most challengeable joint disease due to no effective cure to regenerate the original structure and function of the damaged tissue except for knee joint replacement surgery. The knee OA has become a common disease with increasing impact in an aging society.

1.2.2 Risk factors of knee OA

Aging is a risk factor of knee OA. Although OA is not just a disease related to old people and is not an inevitable consequence of growing old (Loeser, 2009; Muller, 2009). However, aging has been considered as a primary risk factor for OA progression, due to changes in the musculoskeletal system, bone structure rarefaction and bone micro damage accumulation (Loeser, 2009; Anderson & Loeser, 2010). Cell replicative senescence is a state of irreversible growth arrest in aging (Campisi, 1997), with phenomena of shortened telomeres (Jiang et al., 2007), DNA damage (Gensler & Bernstein, 1981), chromatin structure changes (i.e. the formation of senescence-associated heterochromatin) (Loeser, 2009; Muller, 2009). These phenomena are mediating aging due to oxidative stress, reactive oxygen species (ROS) or free radicals (Harman, 1956). However, the cell senescence process has also been considered as a mechanism to prevent cells with damaged DNA from being replicated and thus to

protect against tumor formation and cancer development (Collado et al., 2007; Itahana et al., 2004). It has been reported that aging-related changes in the cartilage matrix, such as increased matrix calcification, fatigue failure, advanced glycation end products (AGEs) formation, collagen cleavage, decreased hydration levels and aggrecan size could refer to the development of OA (Chubinskaya et al., 2002; DeGroot et al., 2004; Mouritzen et al., 2003; Nedić et al., 2013; Rosen et al., 1997; Verzijl et al., 2000; Wells et al., 2003; Wilkins et al., 1983). Chondrocytes from older adult exhibit range of changes typically cell senescence, such as decreased growth factor levels, telomere shortening, increased levels of cytokine and MMP production, up-regulated of p21/p53/p16 and SA- β gal accumulation (Dai et al., 2006; Forsyth et al., 2005; Guerne et al., 1995; Hollander et al., 1995; Long et al., 2008; J. A. Martin & Buckwalter, 2001; Martin et al., 2004; Yudoh et al., 2005; Zhou et al., 2004). Taken together, aging is more likely to be inter-related with OA with but not inter-dependent with OA.

Gender is a risk factor of knee OA. The knee OA is more common in elder female, however, there has not yet clear-cut relationship of osteoarthritis with estrogen in female been found (Felson, 1990). The anatomical and kinematic difference between male and female could be potentially involved as the main factor.

Obesity is a risk factor of knee OA. The association of obesity and knee OA has been well known as a risk factor beyond the pathogenic mechanisms, such as the weight induced a direct systemic effect on joints and obesity-associated low-grade inflammation and oxidative stress (Lee & Kean, 2012).

1.2.3 Diagnosis of knee OA

The destruction of the cartilage, joint space narrowing and bony spur formation (bony abnormality) are the main features of OA (Krane et al., 2001). Imaging methods such as X-ray and Magnetic resonance imaging (MRI) scan are commonly used for knee OA diagnosis. And clinical examination included age (> 50 years old), morning stiffness (< 30 minutes), crepitus on active motions, bony enlargement can also be used as a diagnosis of knee osteoarthritis (Altman et al., 1986).

1.2.4 Treatments for knee OA

The pathogenesis of OA is still unclear and lack of treatment to cure the early process of chondral degeneration (Ringdahl & Pandit, 2011). Commonly applied

treatment for knee OA are healthy weight maintains, medication of pain and anti-inflammatory, steroid injections and surgery for repairing the severely damaged joint. Currently, the trials on stem cell research will be a potential treatment for OA aiming to regenerated the degraded cartilage using patient own cells other than alleviating the pain (Pas et al., 2017).

1.3 Cartilage and subchondral bone

Cartilage is a resilient, flexible, smooth and connective elastic tissue, which is composed of unique and specialized cells called chondrocytes, large amounts of type II collagenous extracellular matrix, proteoglycan and elastin fibers (Figure 1-1). The extracellular matrix, proteoglycan and elastin fibers are all produced from chondrocytes. It has been identified that there are mainly three types of cartilage in human, hyaline cartilage, fibrocartilage, and elastic cartilage, based on the relative amounts of type II collagen and proteoglycan.

Particularly, cartilage is an avascular and aneural tissue, which does not contain blood vessels or nerves. A physics process diffusion is applied to assist chondrocytes nutrition from the surrounding fluid flow and underlying subchondral bone (Pan et al., 2009). Compared with other tissue, cartilage has very low self-repair ability.

Subchondral bone refers to the bony components lying distal to calcified cartilage, which is separated into two distinct anatomic entities: subchondral bone plate and subchondral trabecular bone (Goldring & Goldring, 2010; Grynepas et al., 1991; Imhof et al., 1999; Li et al., 2013; Madry et al., 2010; Milz & Putz, 1994). Subchondral bone sclerosis, together with progressive cartilage degradation, has been widely considered as hallmark symptoms of OA (Figure 1-2). Additionally, articular cartilage overlies subchondral bone, by the depth, it encompasses superficial non-calcified cartilage and deeper calcified cartilage. Calcified cartilage is permeable to small molecule transport, and biochemically interacting between non-calcified cartilage and subchondral bone (Li et al., 2013).

1.4 Epigenetic studies

In an organism, all somatic cells share the same DNA sequence copied from the fertilized egg. However, the cell fate and function change in terms of the different gene or group of genes are working at the different time and different cell types. Epigenetics

is a phenomenon that facilitates temporal and spatial regulation of gene expression independent of changes to the underlying DNA sequence and includes DNA methylation, histone modifications, non-coding RNAs (including micro-RNAs) and chromatin accessibility. In contrast to the irreversible genetic variations induced diseases, the reversible epigenetic dysregulation has been attached more and more importance on disease pathogenesis (Haluskova, 2010; Portela & Esteller, 2010). The epigenetic phenomenon is heritable phenotype changing by gene switch on or off, and it is common for normal development such as X-chromosome inactivation (Disteche & Berletch, 2015) and stem cell differentiation (H. Wu & Sun, 2006). The epigenetic regulation can potentially induce aging and diseases.

1.4.1 Chromatin structure

It is widely accepted that the regulation of chromatin structure is a pivotal epigenetic phenomenon (Margueron & Reinberg, 2010). In eukaryotic cells, chromatin can be found, which is the combination of histone protein and genetic complement (DNA and RNA). DNAs is warping with histone proteins and linking RNAs or other regulatory elements can be regulating the structure of each chromatin to form the dynamic chromatin structure. In each organism, all somatic cells also contain the same number of chromatins except for chromatin structure. A nucleosome is the basic unit of chromatin, consisting of 146 DNA nucleotide pairs and 8 histone protein cores. Gene activation is in terms of gene transcription, which demands the regulatory elements access and bind the right loci in the genome to switch on or off the gene transcription. In other words, DNA in open chromatin regions are not tightly packed, and the regulators such as some transcription factors and polymerases can only access the open chromatin regions (Figure 1-3), such as active enhancers and promoters. Open chromatin that is also known as euchromatin, in contrast to heterochromatin, is correlated with active regulatory elements and the dysregulation will be critical in diseases.

1.4.2 ATAC-seq

Chromatin accessibility has been conventionally carried out by using MNase-seq or DNase-seq required millions of cells (Tsompana & Buck, 2014). However, in

the past few years, it has been greatly improved by using prokaryotic Tn5 transposase (Adey et al., 2010).

Assay for Transposase Accessibility Chromatin with high-throughput sequencing (ATAC-seq) is a method for mapping chromatin accessibility genome-wide (Buenrostro et al., 2015). Transposase can catalyze the transposons movement in the genome. The mutated hyperactive Tn5 transposase used in ATAC-seq has high-level activity than the naturally occurring ones allows Tn5 transposons can easily cut the exposed DNA and ligate the specific next-generation sequencing adaptor sequences at each end. In the intact nuclei, only the open chromatin regions can be accessed and cut by Tn5 transposon. The adapted DNA library sequences after PCR amplification can be sequenced by Illumine sequencer (Figure 1-4). In addition, ATAC-seq enables detection of accessible chromatin with fewer cells than other methods with the use of MNase-seq or DNase-seq. By using the bioinformatics, the sequenced DNA can be aligned and mapped to the genome. ATAC-seq method and analysis can profile the closed and open chromatin regions genome-wide in different cells.

1.5 Research objectives

OA is a degenerative joint disease (Findlay & Kuliwaba, 2016; Goldring & Goldring, 2010) that is one of the most common causes of chronic disability in the world (Kwoh, 2012; Martel et al., 2016), of which the knee OA is the most common. Main features of OA include cartilage degradation, subchondral bone thickening, joint space narrowing and osteophytes formation (Krane et al., 2001), resulting in stiffness, swelling, and pain in the joint. Currently available treatments are either pain relief or joint function improvement by strengthening the supporting muscles. However, OA progression ultimately leads to costly total joint replacement surgery, making it a growing global health burden.

Although the causes of OA are not well understood, risk factors such as age, weight, gender, and genetic factors have been identified (Martel et al., 2016). Several models for OA initiation, such as mechanical injury, inflammatory mediators from

synovium, defects in metabolism and endochondral ossification, have been proposed to explain pathogenesis of this disease (Cox et al., 2013; Dreier, 2010; Kapoor, 2015; Kawaguchi, 2008, 2016; Kuyinu et al., 2016; Man & Mologhianu, 2014). To date, GWAS have identified more than 20 loci to be associated with the risk of developing OA (Uhalte et al., 2017). While next-generation sequencing data is being generated to discover rare variants with larger effect size, the identified variants are often located in the non-coding regions of the genome (Corradin & Scacheri, 2014), complicating the identification of the causal genes. Transcriptomic analyses of cartilage in diseased joints of OA patients (taken from the replacement surgeries) provided an opportunity to pinpoint transcriptionally dysregulated genes and pathways relevant to OA (Chou et al., 2013; Dunn et al., 2016). However, such studies have yet to fully reveal the underlying molecular mechanism of how the transcriptions of these genes are dysregulated.

Recently, epigenetic tools have been applied to gain further insight into the pathogenesis of OA. There have been reports of DNA methylation status (den Hollander et al., 2014; Fernández-Tajes et al., 2014; Reynard, 2017; Rushton et al., 2014) in the cartilage of diseased joints. Some of the reports have also revealed the epigenetic marks are potential as mediators of OA genetic risk (den Hollander & Meulenbelt, 2015; den Hollander et al., 2015; Reynard, 2017; Meurs, 2017). However, the change of gene expression is rarely associated with DNA methylation alterations at promoters (Chou et al., 2013, 2015; Zhang et al., 2016). Many of the identified differentially methylated sites, in fact, fall in enhancer regions (den Hollander et al., 2015; Rushton et al., 2014; van Meurs, 2017), which are non-coding regulatory elements, disruption of which may lead to dysregulated transcription, and many are cell type-specific (Bernstein et al., 2012; Herz, 2016). Recent large-scale studies, such as

FANTOM5 (Forrest et al., 2014), Roadmap Epigenomics Project (Bernstein et al., 2010) and GTEx (Lonsdale et al., 2013) have enabled the prediction of regulatory networks between enhancers and their potential target genes (e.g. JEME (Cao et al., 2017)), which could be applied in a clinical context to explore the roles of enhancers in disease pathogenesis.

Here, this study was designed to investigate alterations of enhancers associated with OA by applying ATAC-seq (Buenrostro et al., 2015) on the knee joint cartilages from OA patients, using an optimized protocol for cartilage sample preparation. ATAC-seq maps the accessible chromatin regions, which are often regulatory regions such as promoters and enhancers that play roles in regulation of gene expression. By integrating our ATAC-seq data with the publicly available genetic, transcriptomic and epigenomic data, dysregulated enhancers and their potential target genes were identified. This data highlights a number of OA risk loci and differentially methylated loci (DML) that potentially play roles in cartilage degradation during OA development.

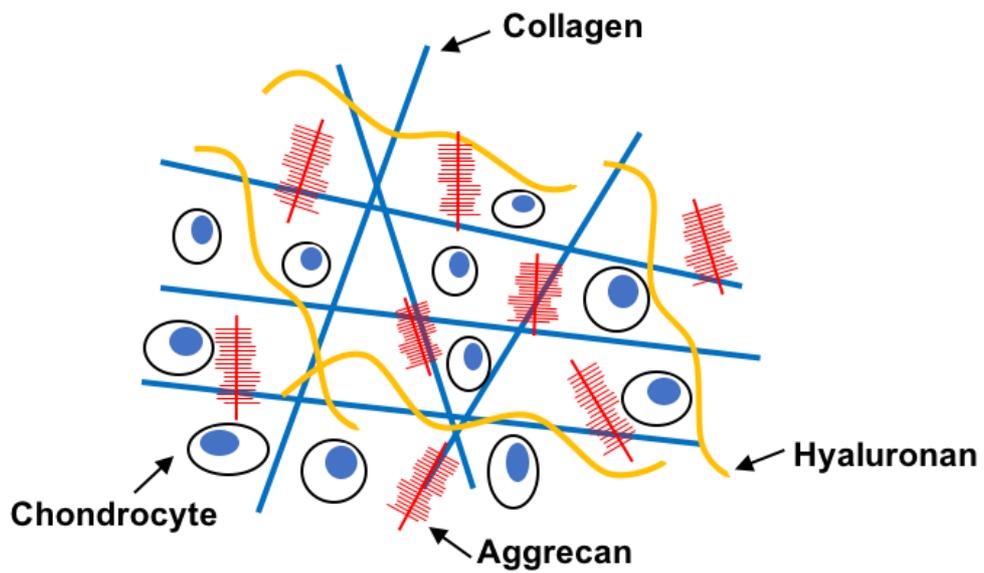


Figure 1-1 Diagram of cartilage matrix major components

Articular cartilage is comprised of chondrocytes surrounding by a matrix of collagens (blue line), proteoglycans (red “bottle brush”) and hyaluronan (yellow line).

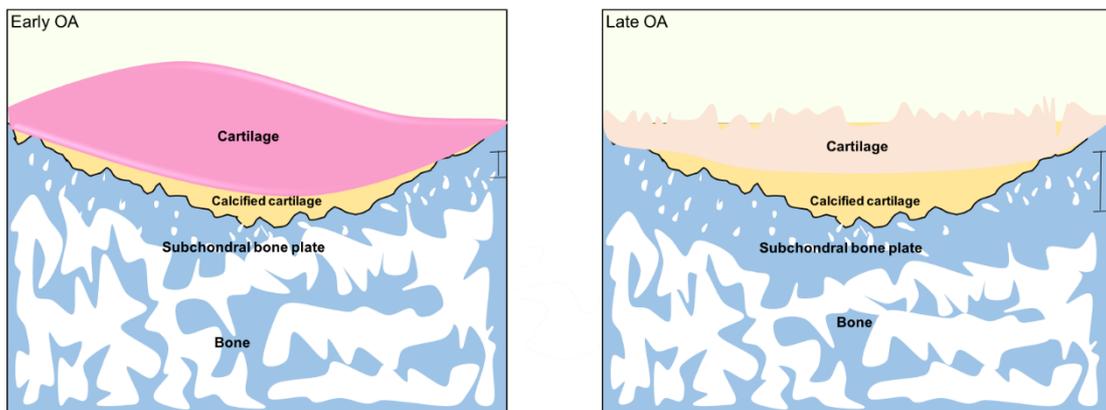


Figure 1-2 Articular cartilage and subchondral bone in normal and diseased conditions.

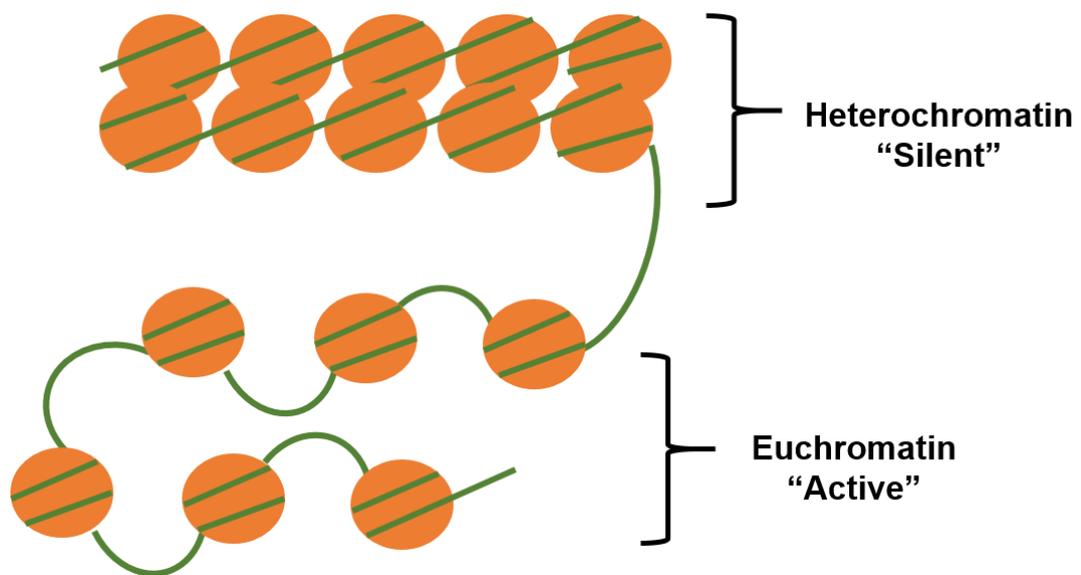


Figure 1-3 Interphase of Chromatin.

Euchromatin region is accessible, considered as active region; heterochromatin region is inaccessible, considered as silent region.

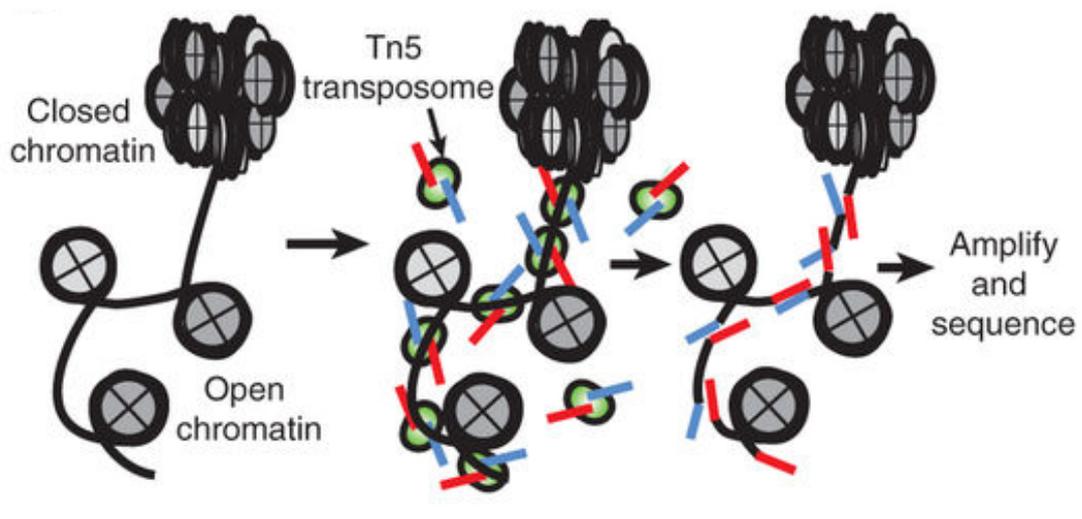


Figure 1-4 Tn5 tagmentation at open chromatin region.

ATAC-seq used the Tn5 transposome can access to the open chromatin and ligate the illumine adaptor sequence, and only the adapted DNA sequence can be further amplified and sequenced (Buenrostro et al., 2015).

Chapter 2 ATAC-seq library preparation using flash frozen and fresh hard tissue

2.1 Introduction

OA has a strong heritable component that is polygenic in nature. However, difficulty in identifying susceptible genes for the disease suggests epigenetic components may be an important factor in the progression of the disease. To date there have been DNA methylation studies on OA (den Hollander & Meulenbelt, 2015), however they are not able to effectively establish the causes for differential expressions in disease samples. Previously developed a unique and proven method was used for extracting high quality material of both normal (or early) and late disease tissues from the same individual donor knee, where accurately isolated chondrocytes from, providing an excellent control for individual-level cofounders (Chou et al., 2013).

To investigate chromatin signatures in articular cartilage associated with OA, ATAC-seq were performed on the chondrocytes isolated from the knee joints of patients. In this study, the ATAC-seq library quality assessed comparing with flash-frozen (n=4) and fresh cartilage (n=8) tissues, according the perversely published tibial plateau section method (Chou et al., 2013), as a model for OA disease progression, including outer region of lateral tibial plateau (oLT) exhibited macroscopically normal cartilage with a visibly smooth cartilage surface, inner region of lateral tibail plateau (iLT) showed intermediate erosion and sufficient cartilage to detect visible fissures on tissue cross-section, and inner region of medial tibial plateau (iMT) had visible most severe erosion of cartilage (Figure 2-1).

2.2 Materials and methods

2.2.1 Knee joint tissues

Human knee joint tibial plates were collected from patients who undertook joint replacement surgery due to severe primary OA at the National Hospital Organization Sagamihara Hospital (Kanagawa, Japan). Demographic information for patients is listed in Table 2-1. The diagnosis of OA was based on the criteria of the American College of Rheumatology (Altman et al., 1986), and all the knees were medially involved in the disease. All the tibial plateaus had medial and lateral compartment

dominant of cartilage erosion. Flash frozen knee joint tissue collected were immediately stored in liquid nitrogen and subsequently stored in -80°C ; fresh knee joint tissue was incubated in 4°C centigrade cultured medium after surgery in less than 6 hours. Informed consent was obtained from each patient enrolled in this study. This study was approved by all the participating institutions (Sagamihara National Hospital, University of Tokyo and RIKEN) with the approval NO. Yokohama H17-23(6).

2.2.2 Cartilage processing

It has previously shown that the oLT region (outer region of the lateral tibial plateau, representing the intact cartilage) is a good control for comparing the iMT region (inner region of medial tibial plateau, representing the damaged cartilage) as a model for OA disease progression (Figure 2-2) (Chou et al., 2013). The previous studies have assessed that these regions could encompass range of histological normal and severity in knee cartilage (Chou et al., 2013; Chou et al., 2015).

The chondrocytes were washed with cold PBS (centrifuge at 500 g for 5 minutes at 4°C). Fresh cartilage was sectioned by using a scalpel (Figure 2-3), and flash frozen knee cartilage was sectioned in liquid nitrogen by using a previous set up custom work platform and followed by thawing with bathing in Dulbecco's Modified Eagle Medium and digestion (Figure 2-4). 100 mg cartilage from each region was sectioned and digested with 0.2% Collagenase Type II (Sigma-Aldrich) at 37°C with rotation for 12 hours to remove matrix debris. No culturing or sorting for purifying chondrocytes, immediately counted approximate 50,000 chondrocytes manually by a cell counting chamber Incyto C-Chip (VWR) and collected by centrifugation at 500 g for 5 minutes at 4°C .

2.2.3 Nuclei extraction

Approximately 50,000 chondrocytes nuclei from both oLT and iMT region of each patient were extracted via homogenizing in 50 μL of pre-cooling and freshly prepared lysis buffer (10 mM Thris.HCl, pH 8.0, 10 mM NaCl, 3 mM MgCl_2 and 0.1% (v/v) Igepal CA-630) on ice and immediately centrifuged at 500 g for 10min at 4°C according to the original protocol (Buenrostro et al., 2015).

2.2.4 Tagmentation

Tn5 transposase (TDE1) and 2× reaction buffer (TD) were from purchased Nextera kit (Illumina). Each tagmentation reaction master mix contains 25 µL TD, 2.5 µL TDE1 and 22.5 µL nuclease free water. 50,000 chondrocyte nuclei were suspended in cold reaction master mix by gently pipetting, avoided making any air bubbles at this step. Reaction was incubated at 37°C for 30 minutes.

2.2.5 DNA purification

The Transposed DNA were purified by using MinElute PCR purification kit (allowed to purify up to 5 ug PCR products between 70 bp to 4 kb in low elution volume; Qiagen). The DNA was eluted in a final 12 µl elution buffer. 1ul loaded into a NanoDrop spectrophotometer (ND-1000) to validate the consistency.

2.2.6 PCR amplification

10 µL transposed DNA was amplified by mixing with 2.5 µl custom PCR forward primer (Ad1.0), 2.5 µL custom PCR reversed barcoded primers (Ad2.n), and 25 µL NEBNext High-Fidelity 2× PCR Master mix (Bao et al., 2015). Thermal cycler condition is 1× (72°C, 5 minutes; 98°C, 30 seconds), (5 + addition cycles)× (98°C, 10 seconds; 63°C, 30 seconds; 72°C, 1 minute). Addition cycles were determined as previously published method by using qPCR (Buenrostro et al., 2015).

2.2.7 Size selection and distribution

After ATAC-seq libraries purified, I used Agencourt AMPure XP (1:1.5 and 1:0.5 sample to beads; Beckman Coulter) for twice size selection, only kept the 100 bp to 1,000 bp length DNA for illumine sequencing. This step would filter out the PCR primers and large DNA fragments which cannot be sequenced in sequencer. The size distribution can be assessed by using bioanalyzer (Agilent Technologies).

2.2.8 Quality check by qPCR

Library quality was assessed before sequencing by qPCR enrichment of a house keeping gene promoter (*GAPDH*) over a gene desert region (Milani et al., 2016). Primer sequences are shown in Table 2-2 (Milani et al., 2016). Briefly, for each PCR reaction, mix 1 µL library DNA (1 ng/µL), 10 µL SYBR Premix ExTaq RnaseH plus

(TAKARA), 0.8 μ L forward primer (10 μ M), 0.8 μ L reverse primer (10 μ M), 0.4 μ L ROX reference dye II (TAKARA), 7.8 μ L nuclease-free water. Thermal cycler condition is 1 \times (95.0 $^{\circ}$ C, 30 seconds) initial denaturation, 40 \times (95 $^{\circ}$ C, 5 seconds; 60 $^{\circ}$ C, 34 seconds) annealing and extension, 1 \times (95 $^{\circ}$ C, 15 seconds; 60 $^{\circ}$ C, 60 seconds; 95 $^{\circ}$ C, 15 seconds; 60 $^{\circ}$ C, 15 seconds) final extension.

2.2.9 KAPA library quantification and Illumine sequencing

Molarity of ATAC-seq libraries were quantified by using KAPA library quantification kit (KAPA Biosystems). Final concentration for sequencing was 5 pM for each lane. ATAC libraries were sequenced on a HiSeq2500 sequencing system (50 bp, PE; Illumina) by internal sequencing service group GeNAS (Genome Network Analysis Support Facility) in RIKEN Yokohama Japan.

2.3 Results and discussion

2.3.1 ATAC-seq of chondrocyte from human knee cartilage

To investigate the variations of chromatin accessibility in cartilage associated with OA, I performed ATAC-seq on human articular cartilage tissues, the tibial plateau of the human knee primary OA selected after replacement surgery (Figure 2-3, 2-4). It has been previously shown that the oLT region (outer region of the lateral tibial plateau, representing the intact cartilage) is a good control for comparing the iMT region (inner region of medial tibial plateau, representing the damaged cartilage) as a model for OA disease progression (Chou et al., 2013), and the transcriptome and methylome of this model have been characterized (Chou et al., 2013; Chou et al., 2015; Zhang et al., 2016; Zhang et al., 2016).

All patients were diagnosed as primary OA, which is induced by aging, none of the patients were younger than 65-year-old when they undertook the joint replacement surgery. The Demographic information of patients were listed in Table 2-1. As expected, female is more suspected to OA, 11 out of 13 patients are female.

Two regions of the cartilage from each patient were selected as oLT (healthy) and iMT (degraded). About 100 mg sectioned cartilage from oLT or iMT region were digested by using collagenase II in 37 $^{\circ}$ C for 12 hours and counted 50,000 live cells for ATAC-seq library. The flash frozen joint tissue was sectioned by using a custom tool shown in Figure 2-4, however, no live cells after digestion could be observed.

2.3.2 ATAC-seq library quality check

Quality of the libraries prepared from flash frozen or fresh cartilage were both assessed. Typically, ATAC-seq library should be generated from live cells and the library length should be distributed between 100 bp to 1000 bp, the nucleosome banding pattern shown in the bioanalyzer profile prior sequencing serves as a quality control for ATAC-seq library (Figure 2-5). However, despite similar bioanalyzer profiles, only chondrocytes isolated from fresh cartilage showed enrichment for accessible loci after sequencing, (Figure 2-6A) and insertion size distribution pooled profile after sequencing (Figure 2-7). This result suggests the chromatin in flash frozen cartilage tissue were randomly fragmented during freezing and thawing. In addition, the expected nucleosome banding patterns were observed in the fragment size distribution for both oLT and iMT libraries (Figure 2-7).

In contrary to previously published ATAC-seq protocol comparing fresh and flash-frozen neurons (Milani et al., 2016), here the bioanalyzer profile was not sufficient to indicate library quality.

Quantitative real-time PCR was performed on the ATAC-seq library to assay enrichment of a known accessible locus (GAPDH promoter) against an intergenic locus (negative control) (Milani et al., 2016). The qPCR enrichment is found to be positively correlated to signal-to-noise ratio (enrichment in DNase hypersensitive sites) listed in the sequencing statistic result (Figure 2-6B; Table 2-3).

Overall, the libraries are of high quality, showing two- to four-fold enrichment in the Roadmap DHS (DHS enrichment score, Methods), with no substantial difference between oLT and iMT libraries.

Taken together, it can be concluded a qPCR assay of known accessible loci is more suitable for ATAC-seq library quality control, and the fresh cartilage tissue is necessary to obtain high quality ATAC-seq data.

2.4 Summary

Conventional epigenomic profiling at the chromatin level, such as chromatin immunoprecipitation sequencing (ChIP-seq) and DNase-seq, is informative in providing insights into the molecular mechanism underlying the regulation of gene expression. However, applying these methods to clinically relevant tissue is less

feasible due to the requirement of large number of cells. Especially with the cartilage tissues, the limited tissue sampling size and the extracellular matrix make collection of sufficient cells difficult. One major advantage of ATAC-seq is that it can be achieved with only thousands of cells, making the direct chromatin profiling of clinical samples feasible.

In this study, ATAC-seq on OA samples was applied to obtain a chromatin accessibility map in articular cartilage, and identified regulatory regions associated with OA. The lack of normal knee tissues due to difficulties associated with collecting them is a limitation for studying OA. However, the previous studies of the pathology and transcriptome showed that the oLT regions are very similar to normal (Chou et al., 2013, 2015). Thus, the oLT regions could serve as a suitable alternative to normal control, which could also reduce the inter-individual variations.

A protocol to generate high quality ATAC-seq data from fresh hard tissue, human cartilage was first established using collagenase II to study human OA disease. Approximately 50,000 chondrocytes nuclei of fresh knee joint tissues were suitable for ATAC-seq library preparation. Quality check qPCR can be used before sequencing to test the signal to the noisy ratio. Thus, it can be concluded that the ATAC-seq libraries had good and indistinguishable quality between oLT and iMT regions.

Table 2-1 Demographic information of patients.

No.	Sample ID	Patient ID	KL-grade	Age	Gender	BMI	Left/right
1	oLT/iMT_01	160908_cartilage_oLT/iMT	4	75	M	29.2	R
2	oLT/iMT_02	160927_cartilage_oLT/iMT	4	78	F	23.3	L
3	oLT/iMT_03	161004_cartilage_oLT/iLT/iMT	3	86	F	25.3	R
4	oLT/iMT_04	161018_cartilage_oLT/iLT/iMT	4	65	F	28.6	L
5	oLT/iMT_05	161027_cartilage_oLT/iLT/iMT	4	70	F	23.5	R
6	oLT/iMT_06	161115_cartilage_oLT/iLT/iMT	4	71	F	22.5	L
7	oLT/iMT_07	161124_cartilage_oLT/iLT/iMT	4	78	F	19.8	R
8	oLT/iMT_08	161212_cartilage_oLT/iLT/iMT	4	82	F	29.9	R
9	Frozen_oLT/ iMT_01	102714YH_oLT/iMT	3	79	F	24.5	R
10	Frozen_oLT/ iMT_02	011916KR_oLT/iMT	4	89	F	24.9	R
11	Frozen_oLT/ iMT_03	012816K_oLT/iMT	4	79	M	23.9	R
12	Frozen_oLT/ iMT_04	120214KL_oLT/iMT	4	80	F	19.5	L

Patient demographic information. Fresh (1-8) and flash frozen (9-12) cartilages were sectioned from oLT and iMT region of each patient.

Table 2-2 QC RT-qPCR primer sequences (human).

Sequence (5' to 3')	Name
CATCTCAGTCGTTCCCAAAGT	GAPDH gene promoter Fw
TTCCCAGGACTGGACTGT	GAPDH gene promoter Rv
AACTGGCTAGTAAGGAGTGAATG	Gene desert region #1 Fw
GGGAATGGAAAGAAGTCCACTAT	Gene desert region #1 Rv
GGAAAGTCCCTCTCTCTAACCT	B2M gene promoter Fw
GCGACGCCTCCACTTATATT	B2M gene promoter Rv
CCCAAACCTCTGAGAGGCTTATT	Gene desert region #2 Fw
GAGCCATCATCTAGACACCTTC	Gene desert region #2 Rv

GAPDH and B2M gene promoter regions are as positive control, and Gene desert region # 1 and # 2 are intergenic regions as negative control.

Table 2-3 ATAC-seq sequencing statistics.

Sample ID	# raw read	# mapped read	# chrM read	# nonduplicated read	# filtered read
oLT_01	129330614	97392138	27968647	89188491	52297657
iMT_01	115602248	92405339	30284689	71819478	44450661
oLT_02	99795730	80640395	21075304	72179831	48677376
iMT_02	99160898	81480868	1844538	90857673	66041550
oLT_03	156914266	121116649	26584322	117399759	74446933
iMT_03	102074864	85253341	22636994	70011859	49032221
oLT_04	123287032	96967328	27774014	86769095	55275069
iMT_04	115122568	92890641	11700019	93796683	65913834
oLT_05	119818700	98934892	46401817	62125879	38103912
iMT_05	100047428	79155363	12299719	81019253	55052706
oLT_06	74390722	61100843	12274153	58131197	41235797
iMT_06	93868388	79771165	16438802	69947345	50727727
oLT_07	127531074	108328720	33782622	83916201	59527656
iMT_07	117432404	90581902	5013315	104949457	71187775
oLT_08	96464546	79464763	17309928	73043617	51451152
iMT_08	100428630	69917383	5558689	86711144	50794381

Sample ID	Fraction of read in DHS	DHS enrichment	Enrichment GAPDH	% mapped	% chrM	% dup	% filter	# called peaks
oLT_01	0.349	2.696	36.08	75.3	28.7	31	53.7	515175
iMT_01	0.391	3.282	18.48	79.9	32.8	37.9	48.1	512610
oLT_02	0.339	2.591	13.85	80.8	26.1	27.7	60.4	539268
iMT_02	0.257	1.717	2.72	82.2	2.3	8.4	81.1	526132
oLT_03	0.303	2.167	19.52	77.2	21.9	25.2	61.5	404851
iMT_03	0.390	3.282	20.22	83.5	26.6	31.4	57.5	480399
oLT_04	0.331	2.503	8.76	78.7	28.6	29.6	57	491146
iMT_04	0.356	2.817	11.91	80.7	12.6	18.5	71	410865
oLT_05	0.450	4.233	31.89	82.6	46.9	48.2	38.5	493423
iMT_05	0.338	2.6	10.11	79.1	15.5	19	69.6	467217
oLT_06	0.354	2.808	14.68	82.1	20.1	21.9	67.5	557619
iMT_06	0.321	2.375	10.92	85	20.6	25.5	63.6	565577
oLT_07	0.370	3.019	12.49	84.9	31.2	34.2	55	432241
iMT_07	0.295	2.088	9.25	77.1	5.5	10.6	78.6	477879
oLT_08	0.354	2.816	14.22	82.4	21.8	24.3	64.7	484188
iMT_08	0.303	2.179	5.27	69.6	8	13.7	72.6	578185

raw read: number of raw ATAC-seq reads from the sequencer.

mapped read: number of reads mapped to the reference genome (hg38).
chrM read: number of reads mapped to the mitochondrial genome.
nonduplicated read: number of the reads after removing PCR duplicates.
filtered read: number of reads after all filtering for MACS2 peak calling.
Fraction of read in DHS: the proportion of filtered reads overlapping Roadmap DHS regions.
DHS enrichment: fold enrichment of reads falling in DHS over randomly shuffled region.
qPCR_enrichment_GAPDH: $2^{\Delta\Delta Ct}$ values generated from RT-qPCR assay of GAPDH promoter (chromatin accessible) against intergenic region (chromatin inaccessible).
% mapped: percentage of reads mapped to the reference genome (hg38).
% chrM: percentage of mapped reads map to the mitochondrial genome.
% dup represents the percentage of duplicated reads in mapped reads.
% filter: percentage of filtered reads for peak calling.
called peaks: number of peaks called by MACS2 for each individual sample.

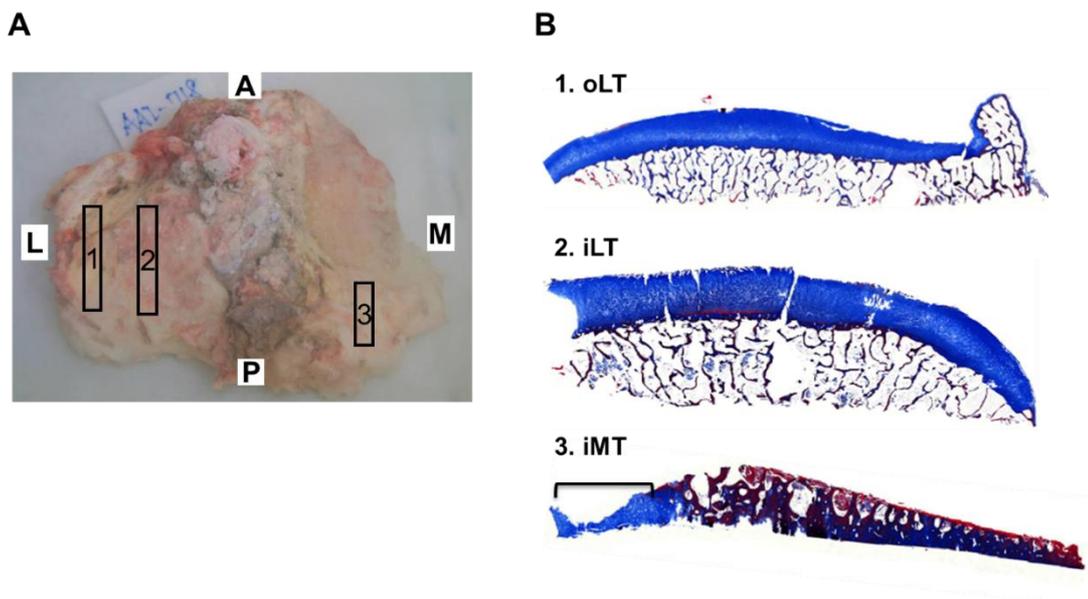


Figure 2-1 Human osteoarthritis (OA) study model system.

(A) Tibial plateau of the human knee primary OA selected after replacement surgery. The black boxes indicate regions for histological analysis and cDNA microarray (Chou et al., 2013), DNA methylation (Zhang et al., 2016) and ATAC-seq. (B) Histological analysis of three regions (Zhang et al., 2016).

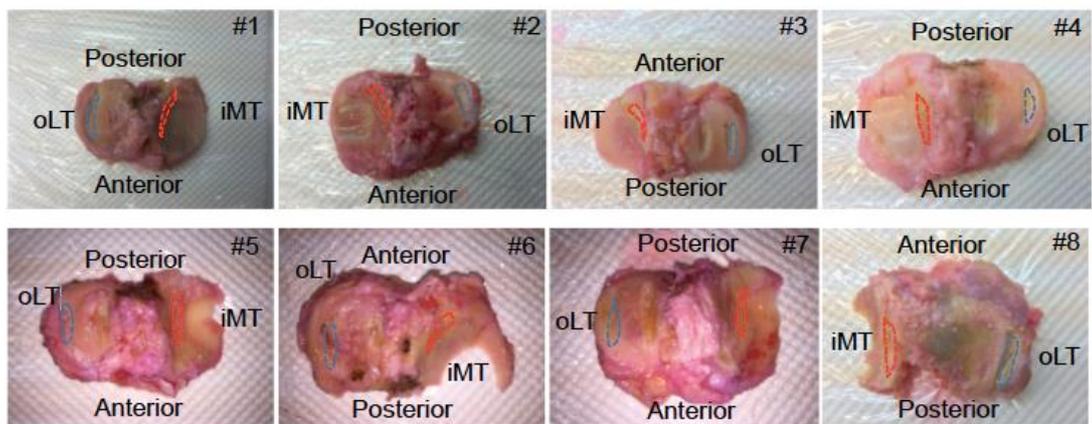


Figure 2-2 Fresh primary knee OA cartilage section for ATAC-seq.

Outer region of lateral tibial plateau (oLT) with a visibly smooth cartilage surface, and inner region of medial tibial plateau (iMT) with visible loss of articular cartilage were collected for isolating chondrocytes.

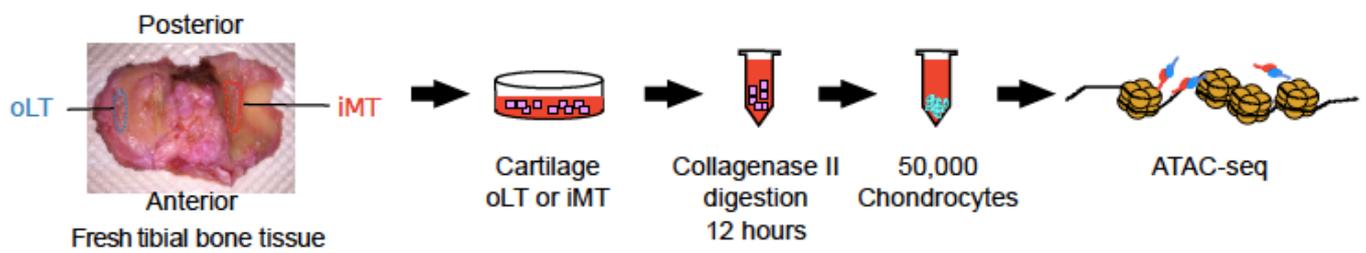


Figure 2-3 Flow chart of the study design (fresh).

Two regions of cartilage from each patient were select as oLT (healthy) and iMT (diseased). About 100 mg sectioned cartilage from each region used for collagenase II digestion in 37°C for 12 hours with rotation, and counted 50,000 live cells for ATAC-seq library preparation.

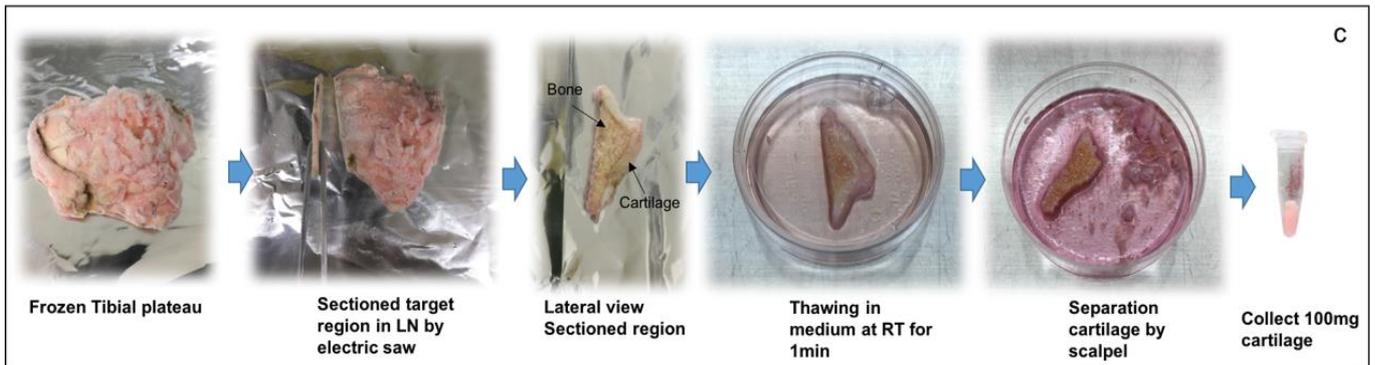
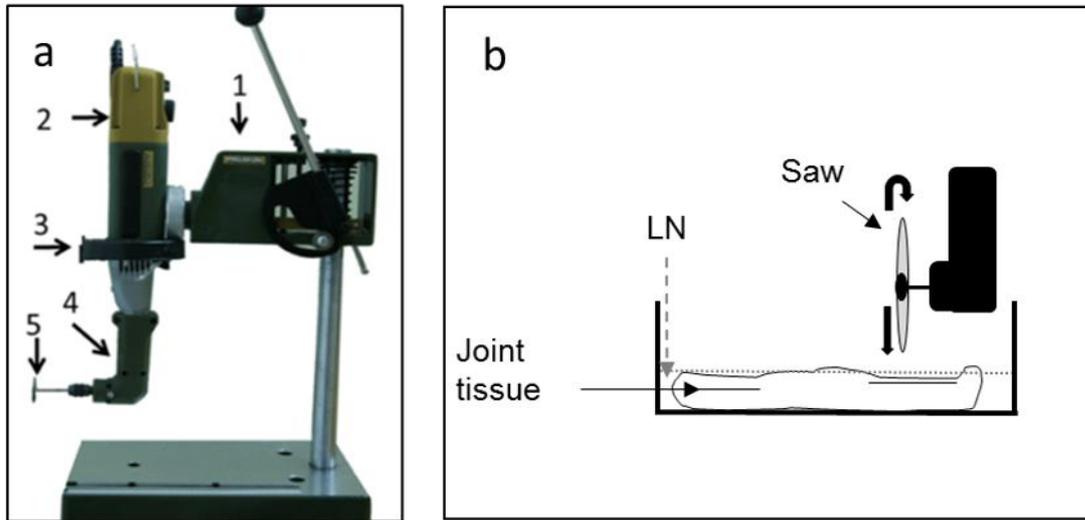


Figure 2-4 Custom tools for frozen joint tissue sectioning and cartilage separation.

(a) Work platform was constructed from commercial components: (1) rotary tool holder, (2) high speed rotary motor, (3) custom fixation adapter, (4) 90 angle adaptor, (5) 25 mm diameter cutting disc of 0.1 mm thickness saw. (b) Knee joint tissue dipped in liquid nitrogen (LN) ready for section. (c) Steps of cartilage collection from the frozen tibial plateau.

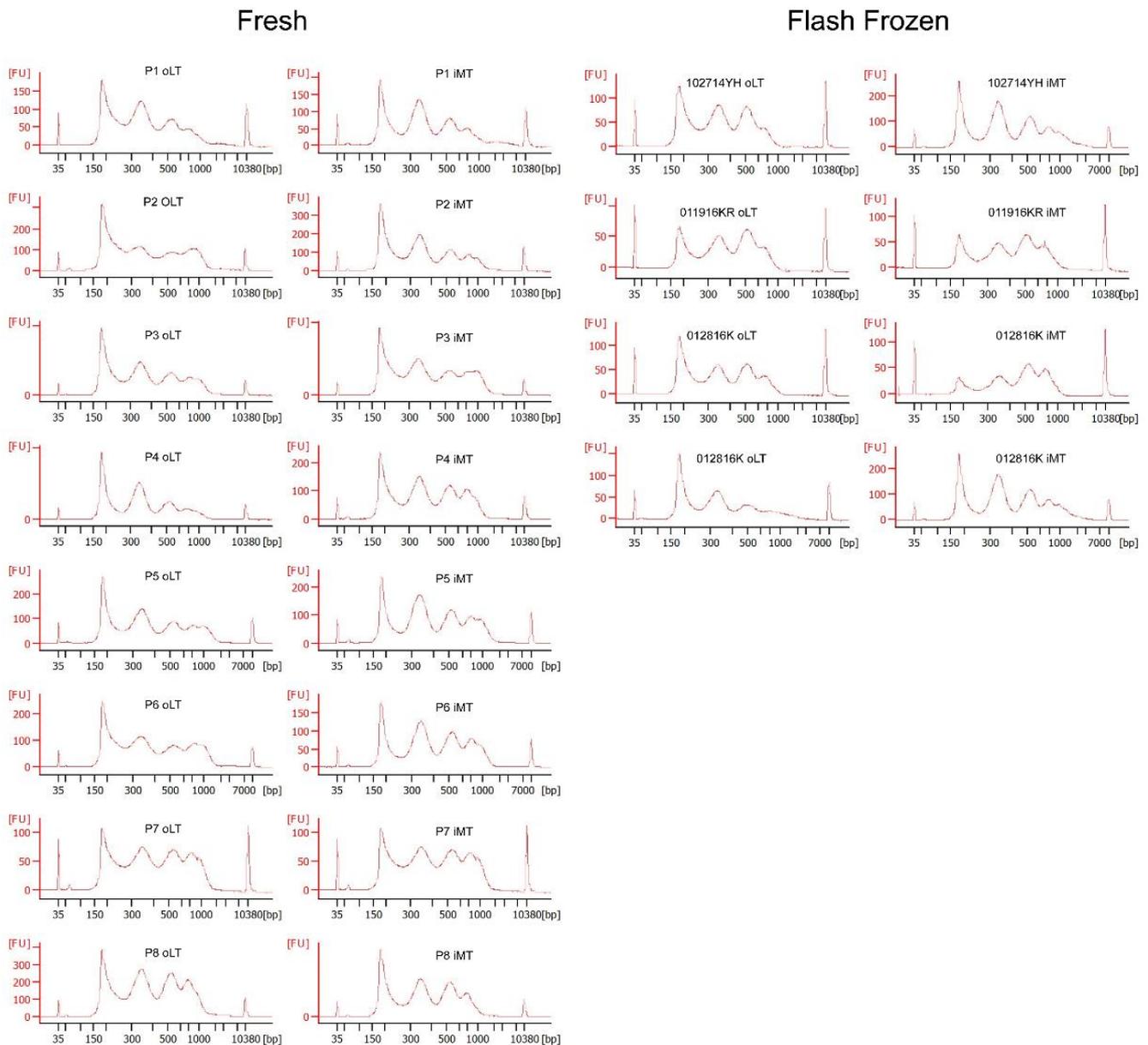


Figure 2-5 Bioanalyzer profile overview of ATAC-seq DNA libraries.

Comparison of fresh and flash frozen samples for ATAC-seq library size distribution (open chromatin DNA length plus sequencing adaptor length), majority of ATAC-seq library length should be distributed between 100 bp to 1000 bp, the peaks in between represents nucleosome free region, mono-nucleosome, de-nucleosome and try-nucleosome from left to right.

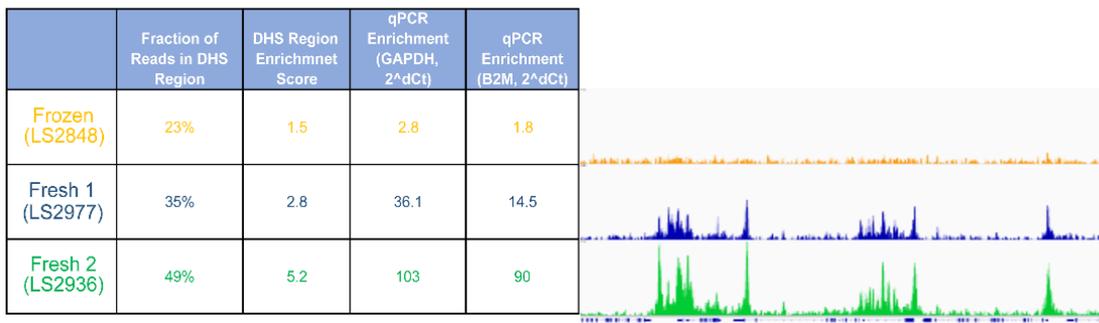
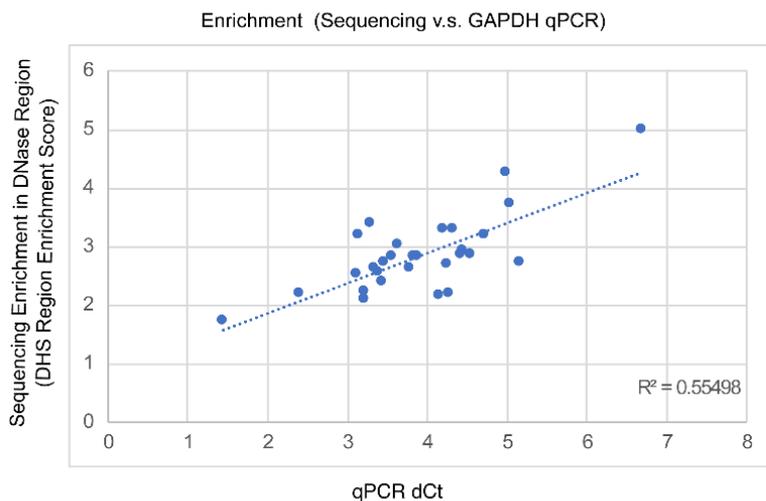
A**B**

Figure 2-6 ATAC-seq quality check with qPCR.

(A) DNase enrichment score and QC-qPCR can potentially be reflecting the ATAC-seq library quality, an essential QC of ATAC-seq library. (B) ATAC-seq library QC-qPCR dCt and ratio of ATAC-seq reads enriched in DNase hypersensitive region is positive linear correlation.

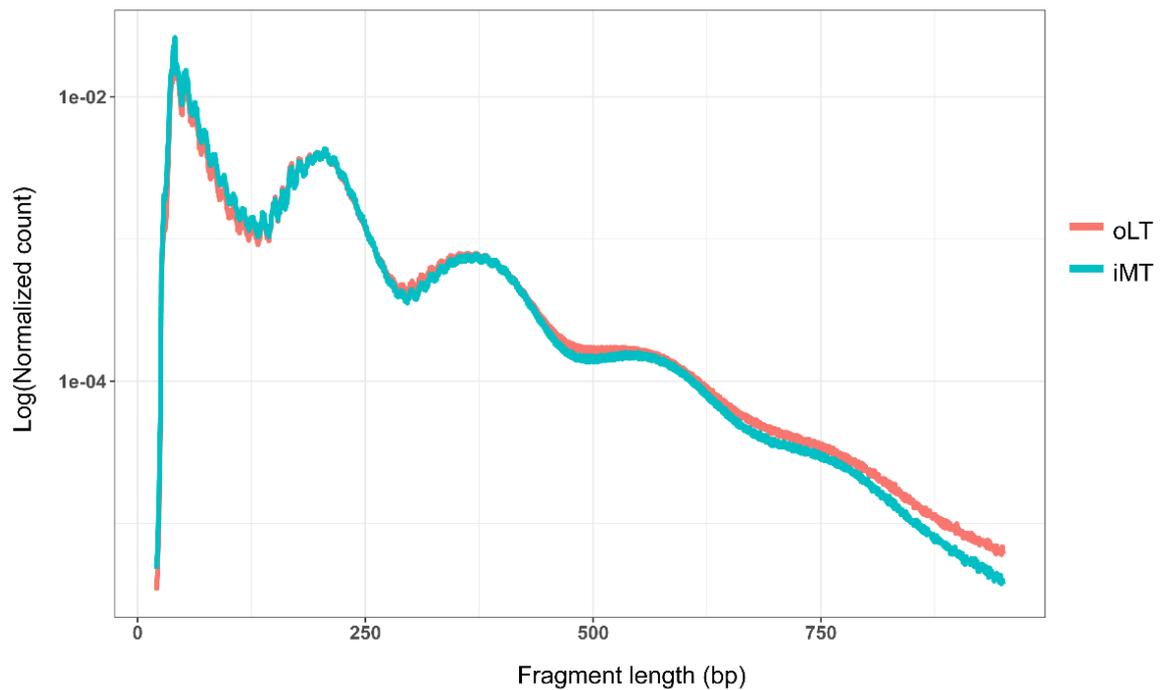


Figure 2-7 Combined 8 atients ATAC-seq library fragment size distribution.

Peaks indicating nucleosome free region, mono-nucleosome, de-nucleosome and tri-nucleosome. Blue curve is pooled 8replicate oLT and red curve is pooled 8replicate iMT. X-axis is open chromatin region DNA length and Y-axis is the log normalized read counts.

Chapter 3 ATAC-seq data analysis

3.1 Introduction

Bioinformatics has become important, with the increasing genomic and epigenomic sequencing techniques developed. The NGS data processing includes pre-processing and data analysis. Similar with other sequencing techniques ATAC-seq data processing pipeline includes sequencing quality check, adapter trimming, alignment, removing PCR duplicates, sorting and peak calling. To date there is no available open source interactive or user-friendly software (without programming) for NGS data processing, and using ATAC-seq data processing pipeline requires bioinformatics skills such as non-interactive command-line application in Linux or Unix environment and programming.

Epigenetic properties of human genome studies such as chromatin structure and DNA methylation, have demonstrated how much epigenetics affect gene translation and expression. ATAC-seq is the assay for Tn5 transposase accessible chromatin followed by next generation sequencing, and the ATAC-seq experiment can provide genome-wide profiles of accessible chromatin and nucleosomes positions through the ATAC-seq data analysis. The working principle of ATAC-seq is by loading the Tn5 transposase and Illumina sequencing primers can be inserted into the open chromatin sites of the whole genome for each cell, the reads after sequencing can be mapped to the reference genome and mark the open chromatin sites.

The genetics of OA is complex. It has very small proportion of genetic heritability in knee OA and currently there is no typical susceptible single gene associated with it (Fernandez-Moreno et al., 2008; Johnson & Hunter, 2014). Therefore, the polygenic nature of OA suggests that epigenetic components may be an important factor in the progression of the disease. However very little epigenetic landscape of OA is known. There have been some reports on the genome-wide DNA methylation studies (Alvarez-Garcia et al., 2016), one of which showed majority (69%) of the differentially methylated regions are hypo-methylated and are found at gene enhancers (Jeffries et al., 2014). Still, relatively few DNA methylation alterations were found in OA diseased tissues, despite many differentially expressed genes. Our previous study by characterizing the genome-wide DNA methylation changes in OA indicated that HOX genes are activated, revealing a self-renewal capacity in diseased tissue (Y. Zhang et al.,

2016). Furthermore, at the chromatin level, there is no genome-wide studies. Here, I sought to investigate genome-wide variations of OA at the chromatin level. This study marks the first application of ATAC-seq on clinically relevant hard tissue and showed how accessible chromatin profiling can provide comprehensive epigenetic information to understand the pathogenesis of a disease.

3.2 Materials and methods

3.2.1 ATAC-seq data processing, peak calling and quality assessment

An ATAC-seq data processing pipeline for read mapping, peak calling, signal track generation, and quality control was implemented (Lee et al., 2016) (Figure 3-1). Briefly, fastq files for all patients were grouped by tissue compartment (oLT or iMT) and input into the pipeline separately, with parameters *-true_rep -no_idr*. Reads were mapped to the hg38 reference genome. Peaks were called by macs2 with default parameters in the pipeline. Basic sequencing information and library quality metrics are listed in Table 2-3.

NucleoATAC was applied to infer genome-wide nucleosome positions and occupancy from the ATAC-seq data (Schep et al., 2015). Briefly, NucleoATAC were run for oLT and iMT separately with default parameters, using bam files merging from all patients as input, and the outputted *nucleoatac_signal.bedgraph* were used for aggregated plot around TSS.

Previously annotated DNase I Hypersensitive Sites (DHS) were used to assess the quality of our libraries. Briefly, a set of DHS defined in Roadmap Epigenomics Project (Boyle et al., 2008) were downloaded (http://egg2.wustl.edu/roadmap/web_portal/DNase_reg.html) and lifted over from hg19 to hg38 using UCSC liftOver tool (Kent et al., 2002) (refer as Roadmap DHS thereafter). For each library, I calculated the ratio of reads that fall within Roadmap DHS versus randomly chosen genomic regions of the same total size (DHS enrichment score). It is noted that the DHS enrichment score is in good correlation with enrichment qPCR for *GAPDH* (Figure 2-6).

3.2.2 Defining a set of unified accessible chromatin regions

Peaks from all 16 libraries were pooled and those that are within 300 bp were merged, resulting in 615,454 merged raw peaks. Reads fall in raw peak regions were

counted for individual samples using bedtools v2.27.0 (Quinlan & Hall, 2010). Read counts were normalized as count-per-million (CPM) based on relative log expression normalization implemented in edgeR v3.18.1 (McCarthy et al., 2012; Robinson et al., 2010). In the end, I defined a set robust peaks (n=109,215) with ≥ 2 CPM in at least 6 oLT or 6 iMT samples for all downstream analyses.

3.2.3 Peak annotation and differentially accessible peak identification

Peaks were annotated as promoters or enhancers based on their intersection with promoter or enhancer as defined in Roadmap DHS (Bernstein et al., 2010). Noted that a peak (n=1,693) was preferentially annotated as a promoter if it intersects both a promoter and an enhancer DHS, in general, those proximal enhancers are epigenetically closed to promoters. Differentially accessible peaks between damaged (i.e. iMT) and intact (i.e. oLT) tissues were identified using edgeR (McCarthy et al., 2012; Robinson et al., 2010). Briefly, reads were normalized by library size and DHS enrichment score of each sample were incorporated as continuous covariates in the design matrix of the generalized linear model implemented in edgeR (McCarthy et al., 2012; Robinson et al., 2010). The Benjamini-Hochberg adjusted p-values (i.e. false discovery rate, FDR) was taken to measure the extent of differential accessibility. A cutoff of $FDR \leq 0.05$ was used to define differentially accessible peaks.

3.2.4 Processing of OA associated SNPs

GWAS lead SNPs of OA were obtained from GWASdb v2 (as of 16 Sep 2017) (M. J. Li et al., 2016), NHGRI-EBI GWAS Catalog (as of 16 Sep 2017) (MacArthur et al., 2017), and a recent study (Zengini et al., 2017). SNPs in linkage disequilibrium with these lead SNPs (i.e. proxy SNPs with $r^2 > 0.5$ and distance limit of 500kb in any three population panels of the 1000 Genomes Project pilot 1 data) were obtained from SNAP database (A. D. Johnson et al., 2008), resulting in 8,973 GWAS SNPs associated with OA. As a negative control, GWAS SNPs associated with Parkinson's disease were obtained the same way from NHGRI-EBI GWAS Catalog. Coordinates for these SNPs in hg38 were obtained from dbSNP Build 150.

3.2.5 Definition of cell type-specific enhancers

Bed files for enhancer clusters with coordinated activity in 127 epigenomes, as

well as the density of clusters per cell type, were obtained from online database (http://egg2.wustl.edu/roadmap/web_portal/DNase_reg.html). DNase I regions selected with $p < 1 \times 10^{-2}$, lifted over from hg19 to hg38 using UCSC liftOver tool. For each cell type, clusters with two-fold density than average (across all cell types) were defined as specific and pooled.

3.2.6 Enrichment analysis for differentially accessible peaks

Fisher's exact test were applied to assess the enrichment for the cell type-specific enhancers. Briefly, differentially accessible peaks ($FDR \leq 0.05$) were compared to all peaks, and same number of randomly selected peaks were used as a control. Permutation of peaks as control were done 25 times.

3.2.7 Linking enhancer peaks to their potential target genes

A promoter peak is assigned to a gene if it intersects with the transcription start sites (TSS) of its transcript as defined in the FANTOM CAGE associated transcriptome (Hon et al., 2017). I noted that a promoter peak might be associated with multiple genes. An enhancer peak is defined as linked to a target gene if it overlaps an expression quantitative trait loci (eQTL) of the corresponding gene (GTEx V7, in any tissues with $p < 1 \times 10^{-5}$) (Consortium, 2015), or is supported by putative enhancer-promoter linkage predicted by JEME method (<http://yiplab.cse.cuhk.edu.hk/jeme/based>) on FANTOM5 and Roadmap Epigenomics Project data (Cao et al., 2017).

3.2.8 Integration with publicly available transcriptome data

An RNA-seq dataset (ArrayExpress E-MTAB-4304) from an independent OA patient cohort was reanalyzed, in which the RNA was extracted from cartilage tissues of both iMT and oLT for 8 patients (Dunn et al., 2016). Read counts on transcripts were estimated by Kallisto v0.43.1 (Bray et al., 2016) using default parameters on FANTOM CAGE associated transcriptome (Hon et al., 2017). Estimated read counts of a gene, defined as the sum of the estimated read counts of its associated transcripts, were used as the input for differential gene expression analysis using edgeR v3.18.1. In total, 3,293 genes were defined as significantly differentially expressed between iMT and oLT ($FDR \leq 0.05$). A gene is defined as “consistently dysregulated both at the epigenomic and transcriptomic levels” when it is upregulated (or downregulated) in RNA-seq with more (or less) accessible promoters or enhancers in ATAC-seq ($n= 371$).

3.2.9 Gene ontology analysis

R (version 3.4.0) was used for performing statistical analyses via an integrated development environment (IDE), Rstudio (version 1.0.143). R/Bioconductor project packages edgeR (version 3.18.1) (Robinson et al., 2010), ChIPseeker (version 1.12.0) (Yu et al., 2015) were utilized. More R packages used in the data processing, ggplot2 (version 2.2.1), dplyr (version 0.5.0), limma (version 3.32.2), data.table (version 1.10.4), magrittr (version 1.5) TxDb.Hsapiens.UCSC.hg38.knownGene (version 3.4.0), rafalib (version 1.0.0), piper (version 0.6.1.3), reshape2 (version 1.4.2).

Linux command line based tools used for data processing, HOMER (version 4.9), bedtools (version 2.26.0), macs2 (version 2.1.1.20160309)

3.2.10 Statistics

In the differential ATAC-seq peak assay, ANOVA (one way analysis of variance) test was performed, with a BH multiple test correction, the adjusted p-value (FDR) ≤ 0.05 was considered significant. HOMER de novo motifs assay, p-value $< 1e-15$ was considered significant. Enrichr GO/pathway analyses, p-value < 0.01 was considered significant. RedeR network assay, p-value < 0.001 was considered significant.

3.3 Results and Discussion

3.3.1 ATAC-seq data processing, peak calling and quality assessment

When applied nucleoATAC to infer genome-wide nucleosome occupancy and positioning from ATAC-seq data 35, it is shown similar aggregated signals around transcription start sites (TSS) for both oLT and iMT libraries, corresponding to -2, -1, +1, +2, +3 nucleosomes as well as nucleosome depletion region at upstream of TSS (Figure 3-2). Thus, it is concluded that the ATAC-seq libraries had good and indistinguishable quality between oLT and iMT regions.

3.3.2 Accessible chromatin landscape highlights potential enhancers and their target genes relevant to OA

Based on the 16 ATAC-seq libraries, I identified a set of unified accessible chromatin regions across all samples (n=109,215 robust peaks, Methods); 77,655 (71.1%) of which were annotated as enhancers and 18,410 (16.9%) as promoters

(Figure 3-3) based on Roadmap DHS annotations (Methods). To assess the relevance of these peaks to OA, these peaks were intersected against OA GWAS (Single Nucleotide Polymorphisms) SNPs and OA DML datasets. 149 peaks (31 promoters and 101 enhancers) were found overlapping with OA-associated GWAS SNPs and 2,508 peaks (595 promoters and 1,715 enhancers) overlapping with OA DML. The fact that majority of these overlapping peaks are enhancers, along with the notion that many disease-associated noncoding variants reside in enhancers (Corradin & Scacheri, 2014; Farh et al., 2015), suggest the dysregulation of enhancer might play a role in OA pathogenesis. To further characterize these potentially OA-relevant enhancers, public resources (Cao et al., 2017; Lonsdale et al., 2013) to predict their target genes (Methods) were utilized. Enhancers that overlap OA GWAS SNPs and OA DML were identified, as well as their predicted target genes, representing candidates of dysregulated transcriptional network during OA pathogenesis. The list includes previously identified OA associated genes, such as *FGFR3* (Tang et al., 2016; S. Zhou et al., 2016) (its enhancer overlaps an OA GWAS SNP, Figure 3-4 up) and *PTEN* (Iwasa et al., 2014) (its predicted enhancer overlaps with an OA DML, Figure 3-4 down), thus validating this approach.

It has been verified 7 OA associated SNPs (rs10851630, rs10851631, rs10851632, rs12905608, rs12910752, rs4238326, and rs35246600) reside in 3 chondrocyte accessible enhancers of the predicted target gene *ALDH1A2*. rs4646626 has previously been identified as a suggestive OA risk locus ($p = 9 \times 10^{-6}$, (Zeggini et al., 2012)), for which it is found two proxy SNPs (rs10851632 and rs12905608), predicted to target the *ALDH1A2* gene, hitting the accessible enhancer region (chr15:57949887-57950974 and chr15:58021340-58022482, respectively) in the samples. Since *ALDH1A2* is inactivated in prechondrogenic mesenchyme during the cartilage development (Hoffman et al., 2006), these SNPs may contribute to OA through disrupting the enhancer of *ALDH1A2* and inappropriately activating a cell differentiation pathway. Consistently, OA genetic risk variants in *ALDH1A2* locus has also been functionally characterized in a recent study (Shepherd et al., 2018).

In addition, I identified several aberrantly methylated enhancers that may be associated with OA. The example is cg09221159 within the enhancer for *PTEN* gene which is hypomethylated in the damaged cartilage (Zhang et al., 2016). *PTEN* is involved in the positive regulation of the apoptotic signaling pathway and its activation is consistent with the notion that chondrocyte apoptosis may contribute to the failure in

appropriately maintaining the cartilage. Thus, my analyses show that the chromatin accessibility map can provide an additional layer of evidence for determining which loci, especially those in the non-coding regions, are associated with OA, which may have been ignored in previous studies.

Therefore, this study provides an accessible chromatin landscape of cartilage tissue for better interpretation of other genetic and epigenomic data relevant to OA and other skeletomuscular disease, although caution should be taken since public data from many populations are involved in the integrative analysis while our ATAC-seq dataset is acquired from the Japanese population.

3.3.3 Identification of differentially accessible enhancers in OA

To compare the accessible chromatin landscape between the intact (oLT) and the damaged (iMT) tissues, principal component analysis was performed on the peak signals across the 16 samples. The first principal component (41.37% of variance) can be attributed to tissue damage variations (i.e. oLT versus iMT), while the second principal component (16.03% of variance) can be attributed to patient-to-patient variations (Figure 3-5). This observation suggests the accessible chromatin landscapes of damaged and intact tissues are readily distinguishable from each other, despite the variations among individual patients.

To identify the chromatin signatures relevant to OA, differential accessibility analysis was performed between oLT and iMT using edgeR (Robinson et al., 2010). The Benjamini-Hochberg adjusted p-values (i.e. false discovery rate, FDR) was taken to measure the extent of differential accessibility with the percentage of the annotation (Figure 3-6), the peaks with smaller FDR (i.e. more confident to be differentially accessible) contain more enhancer peaks, suggesting the enhancers are more likely to be dysregulated in OA cartilage. Significant differentially accessible peaks were determined with $FDR \leq 0.05$. Out of the 4,450 differentially accessible peaks, 1,565 are more accessible and 2,885 are less accessible in the damaged tissues compared to the intact tissues (Figure 3-7, Figure 3-8). The identified differentially accessible peaks are generally consistent among patients (Figure 3-9). Majority (85.3%) of these differentially accessible peaks are enhancers and only 7.6 % are promoters. It is noted that the promoter accessibility of *SOX11* and *RGR* are altered in the OA damaged tissues (more accessible for *SOX11* and less accessible for *RGR*, Figure 3-8), which are

consistent with the previous findings that they are up- and down-regulated in OA damaged tissues, respectively (Chou et al., 2015). Moreover, the differentially accessible enhancers ($FDR \leq 0.05$) overlapping either OA GWAS SNP ($n = 5$; from 4 independent loci) or OA DML ($n = 16$, concordant in direction) are summarized in Table 3-1 and Table 3-2, respectively.

Majority of the differentially accessible peaks are enhancers, which are known to regulate transcription. Cell type-specific enhancers largely drive the transcriptional program required to carry out specific functions for each cell type (Heinz et al., 2015) and their dysregulation may lead to diseases (Kron et al., 2014). To assess the cell type specificity of the differentially accessible enhancers identified in this study, I examined their enrichment of cell type-specific enhancers of 125 cell types defined by the Roadmap Epigenomics Project (Bernstein et al., 2010). These differential accessible enhancers are highly enriched for enhancers were found that are specific to bone-related cell types in damaged tissues (e.g. chondrocyte and osteoblast), as well as mesenchymal stem cells (MSC) and fibroblasts, which have been reported to have similar profiles to MSC (Hematti, 2012) (Figure 3-10).

In summary, the differential accessible chromatin regions identified in this study are enriched in OA-relevant enhancers supported by multiple genetic and epigenomic evidence. These differentially accessible enhancers and their predicted target genes could be used for prioritization of candidate genes to be tested for studying OA disease progression.

3.3.4 Motif enrichment analysis reveals transcription factors relevant to OA

In order to gain more insights into which regulatory pathways may be dysregulated in OA, next the enrichment of transcription factor binding motifs in the differentially accessible regions were examined. Enrichment analyses were performed separately for the regions that are significantly more or less accessible in the damaged tissues, using all accessible regions as background. Transcription factors with significantly enriched motifs are summarized (Figure 3-10A), and their binding prediction in robust peaks are listed (Figure 3-10B). It is noted that most of these transcription factors belong to ETS and bZIP family; many of which are known to regulate genes involved in bone or cartilage development, including AP-1 (Vincenti & Brinckerhoff, 2002), CEBP (Okuma et al., 2015), MafK (Hong et al., 2011), STAT3

(Hayashi et al., 2015), and ERG (Iwamoto et al., 2001; Iwamoto et al., 2007). Moreover, it has been previously proposed that ETS-1 may be involved in OA based on the DML study (Zhang et al., 2016).

The transcription factor Jun-AP1, which is enriched in the more accessible regions of the damaged tissues, is known to regulate chondrocyte hypertrophic morphology, which contributes to longitudinal bone growth, consistent with the notion of endochondral ossification (He et al., 2016). A recent study showed injection of adipose-derived stromal cells overexpressing an AP-1 family transcription factor Fra-1 can inhibit OA progression in mice (Schwabe et al., 2016). Consistently, in this ATAC-seq analysis, other members of AP-1 family (e.g. BATF, FOSL2) also showed enrichment in differentially accessible regions.

Taken together, the motif enrichment analysis of the differentially accessible regions in the damaged tissues is consistent with the hypothesis that the transcriptional program for chondrocyte differentiation may be disrupted during OA progression, and suggests that cell type-specific enhancers may be dysregulated through the ETS and bZIP family transcription factors.

3.3.5 Integrative transcriptomics and epigenomics analysis reveals pathways involved in OA

To evaluate the effects of the dysregulated regulatory regions on gene expression in OA, the RNA-seq dataset from a different cohort that used the same disease model (i.e. oLT vs. iMT) (Dunn et al., 2016) were reanalyzed and integrated it with the differentially accessible regions identified in this study (Figure 3-11). Firstly, the dysregulated regulatory regions have detectable effects on the gene expression were verified; the genes with more accessible promoters or enhancers have significantly higher expression fold-changes between oLT and iMT ($p < 0.001$, Student's *t*-test) than the ones with less accessible promoters or enhancers (Figure 3-12). To further investigate the congruence between these transcriptomic and epigenomic changes, the significant differentially expressed genes ($n=3,293$, $FDR \leq 0.05$) were overlapped onto the genes with differentially accessible promoters ($n=255$) or enhancers ($n=2,406$) (Figure 3-13). The overlaps are statistically significant ($p < 0.001$ in both promoters and enhancers, Fisher's exact test), and further demonstrate the chromatin accessibility dataset from this study is generally consistent with the transcriptomic dataset. As a

result, 371 genes were identified that are consistently dysregulated both at the epigenomic and transcriptomic levels, representing a shortlist of OA-related candidate genes supported by multiple lines of evidence (Figure 3-13, Table 3-3 and Table 3-4).

It was found that *BMPR1B* (bone morphogenetic protein receptor type 1B) is upregulated and both its promoter and enhancer are more accessible in the damaged tissue. Its activation is consistent with the ossification pathway activation, since it encodes a transmembrane serine kinase that binds to BMP ligands that positively regulate endochondral ossification and abnormal chondrogenesis (J. Li & Dong, 2016; Yoon et al., 2005). Consistently, an osteoblast marker gene *MSX2* (Msh Homeobox 2) involved in promoting osteoblast differentiation (Cheng et al., 2003; Matsubara et al., 2008), is upregulated and both its promoter and enhancer are more accessible in the damaged tissue, suggesting the osteoblast differentiation may be activated in OA. Furthermore, it was found *ROR2* (receptor tyrosine kinase like orphan receptor 2) is down regulated and its enhancers are less accessible in the damaged samples. Since it is required for cartilage development (Dickinson & Hollander, 2017; Schille et al., 2016), it suggests that the normal chondrocyte development and cartilage formation may be compromised in OA. Consistently, I have also determined *FGFR2* and *STAT1*, which are known to inhibit chondrocyte proliferation (Karuppaiah et al., 2016), are upregulated.

To elucidate the biological pathways dysregulated in OA, the enrichment of gene ontology (GO) terms of the 371 OA-related candidate genes using enrichr (Kuleshov et al., 2016) (Figure 3-14, top 30 terms listed) was examined. Overall, I observed the enrichment of GO terms related to cell fate and differentiation, including MSC differentiation, ossification and bone development (Figure 3-14). In the ‘positive regulation of ossification pathway’, two genes were identified to be more accessible and upregulated: *SOX11* (only the promoter is more accessible) and *WNT5A* (both the promoter and the enhancer are more accessible). It has been shown that *WNT5A* protein can induce matrix metalloproteinase production and cartilage destruction (Huang et al., 2017), and its upregulation is consistent with ossification being an important process and signature of OA progression. Other susceptible genes and pathways that support the ossification during OA from this analysis include *LRP5*, *FGFR2* and *BMPR1B* in the endochondral bone morphogenesis pathway, which have been reported as OA associated genes previously (den Hollander et al., 2014; Ellman et al., 2013; Fernández-Tajes et al., 2014; Papathanasiou et al., 2010; Rushton et al., 2014; Wu et al., 2016). In

conclusion, the integrative analysis of ATAC-seq and publicly available RNA-seq datasets indicates that dysregulated chondrocyte differentiation and endochondral ossification are associated with OA progression.

3.4 Summary

The strategy of integrating the epigenomic data of clinically relevant tissues with the publicly available genetic and transcriptomic data allowed better understanding how the identified loci may contribute to OA pathogenesis. Most of these accessible chromatin regions are enhancers and the large scale public datasets enable like them to their putative target genes. Previous studies have suggested the potential role of OA-associated epigenetic changes within enhancers in disease pathogenesis (den Hollander & Meulenbelt, 2015; van Meurs, 2017). With this enhancer-gene map in chondrocyte, it can now better interpret the previously identified OA GWAS SNPs or OA differential methylated loci located lie outside of the coding regions.

In general, the differential enhancer analysis shows MSC, chondrocyte and osteoblast-specific enhancers are dysregulated in the damaged tissues. Furthermore, motif enrichment analysis of differentially accessible loci has identified many dysregulated transcription factors, the functions of which are known to be in chondrocyte development regulation.

In the integrative analysis of ATAC-seq and RNA-seq, many dysregulated genes related to lineage differentiation of MSC pathways were observed.

The pathogenesis for OA is not yet fully understood, despite multiple genes and pathways that have been characterized to be dysregulated (Dunn et al., 2016; Shen et al., 2017; Steinberg et al., 2017; Uhalte et al., 2017; Usami et al., 2016). In this study, by integrating clinically relevant epigenomic data with genetic and transcriptomic data, it provides multiple lines of evidence supporting a number OA candidate genes and

pathways that may be crucial to OA pathogenesis, which could potentially be used for clinical diagnostic or as therapeutic targets.

Table 3-1 Differentially accessible enhancers overlapping with OA GWAS SNPs.

Enhancer peak ID	Proxy SNP ID	Lead SNP ID	r ²	GWAS P value	Phenotype	Population	GWAS literature	log ₂ (fold change)	FDR	Predicted target gene(s)
chr7:32611048-32612186	rs10807862	rs7805536	0.8	4.4×10 ⁻⁸	OA	EUR	(Zeggini et al., 2012)	-0.96	0.01	<i>ZNRF2P1</i> ; <i>DPY19L1P1</i> ; <i>AVL9</i> ; <i>LSM5</i>
chr7:32616344-32617105	rs10951345	rs7805536	0.8	4.4×10 ⁻⁸	OA	EUR	(Zeggini et al., 2012)	-0.77	0.03	<i>DPY19L1P1</i> ; <i>AVL9</i>
chr10:119370132-119371726	rs17098787	rs11198893	0.9	9.0×10 ⁻⁶	OA (knee and hip)	EUR	(Panoutsopoulou et al., 2011)	-0.59	0.02	<i>RP11-215A21.2</i>
chr7:44238166-44238952	rs67391165	rs3757837	1.0	8.0×10 ⁻¹⁰	OA (hip)	EUR	(Evangelou et al., 2014)	0.62	0.03	<i>MYL7</i> ; <i>YKT6</i>
chr8:125503356-125504061	rs7832357	rs4512391	1.0	1.1×10 ⁻⁶	OA (knee and hip)	EUR	(Panoutsopoulou et al., 2011)	-0.54	0.04	NA

Table 3-2 Differentially accessible enhancers overlapping with OA DMLs.

Enhancer peak ID	DML	Delta beta*	Log ₂ (fold change)	FDR	Predicted target gene(s)
chr1:19398124-19398841	cg06360604	-0.177	1.194	0.003	<i>AKR7A2</i> ; <i>AKR7A3</i> ; <i>CAPZB</i> ; <i>PQLC2</i>
chr16:66511371-66512170	cg02934719	-0.168	0.827	0.021	<i>BEAN1</i> ; <i>TK2</i>
chr11:74025533-74026295	cg01117339	-0.164	0.623	0.037	<i>COA4</i> ; <i>RAB6A</i> ; <i>UCP2</i>
chr1:85606965-85607684	cg00418071	-0.157	1.214	0.001	<i>CYR61</i> ; <i>DDAH1</i> ; <i>ZNHIT6</i> ; <i>SYDE2</i>
chr4:158809913-158811140	cg11637968	-0.218	0.878	0.032	<i>ETFDH</i> ; <i>FAM198B</i> ; <i>FNIP2</i> ; <i>PPID</i> ; <i>C4orf46</i>
chr5:32767956-32768626	cg26647771	-0.198	0.924	0.010	<i>GOLPH3</i> ; <i>NPR3</i> ; <i>TARS</i>
chr1:226708432-226710048	cg04503570	-0.166	0.500	0.046	<i>PSEN2</i>
chr2:1720779-1722443	cg08216099	-0.193	0.732	0.016	<i>PXDN</i>
chr10:119523073-119524950	cg18591136	-0.152	0.508	0.040	<i>RGS10</i>
chr6:168412123-168412714	cg26379705	-0.174	0.736	0.009	<i>SMOC2</i>
chr6:146850220-146851275	cg18291422	-0.184	0.960	0.012	<i>STXBPS</i>
chr8:81688679-81689538	cg04491064	0.164	-0.959	0.046	<i>CHMP4C</i>
chr12:124700740-124701544	cg04658679	0.160	-0.514	0.030	<i>FAM101A</i> ; <i>UBC</i>
chr4:26412447-26413348	cg22992279	0.159	-0.737	0.028	<i>RBPJ</i>
chr11:35344856-35345647	cg13971030	0.157	-0.656	0.038	<i>SLC1A2</i>
chr12:132846887-132848225	cg21232015	0.151	-0.615	0.038	<i>ZNF10</i> ; <i>ZNF268</i> ; <i>ZNF605</i> ; <i>ZNF84</i> ; <i>GOLGA3</i> ; <i>ANKLE2</i> ; <i>CHFR</i>

* Positive or negative delta beta indicates hypermethylation or hypomethylation in iMT, respectively.

Table 3-3 OA associated genes enriched in top 30 GO terms.

Term	GO term ID	Overlap	P.value	% enrichment	Differential gene enriched in terms		
					pro	enh	pro&enh
positive regulation of TRAIL-activated apoptotic signaling pathway	GO:1903984	2/6	0.005	33.3	PMAIP1	THBS1	
positive regulation of mesenchymal stem cell proliferation	GO:1902462	2/6	0.005	33.3	HMGA2;SOX11		
negative regulation of collateral sprouting of intact axon in response to injury	GO:0048685	3/9	0.000	33.3	DCC	PTPRS;RTN4RL1	
mesenchymal cell fate commitment	GO:0014030	2/6	0.005	33.3	HMGA2	FGFR2	
mesenchymal cell differentiation involved in salivary gland development	GO:0060692	2/6	0.005	33.3	HMGA2	FGFR2	
mesenchymal cell differentiation	GO:0048762	2/6	0.005	33.3	HMGA2	FGFR2	
embryonic ectodermal digestive tract morphogenesis	GO:0048613	3/9	0.000	33.3	SOX11	FOXF1;FGFR2	
collateral sprouting of intact axon in response to injury	GO:0048673	2/6	0.005	33.3		RTN4RL1;BDNF	
choline catabolic process	GO:0042426	2/6	0.005	33.3		SLC44A1;CHDH	
cardiac endothelial to mesenchymal transition	GO:0140074	2/6	0.005	33.3	HMGA2	FGFR2	
positive regulation of metanephric cap mesenchymal cell proliferation	GO:0090096	3/10	0.001	30.0	CHRD	LRP5;FGFR2	
positive regulation of mesenchymal cell proliferation involved in lung development	GO:2000792	3/10	0.001	30.0	CHRD	LRP5;FGFR2	
post-embryonic limb morphogenesis	GO:0035127	2/7	0.007	28.6		TBX4	BMPR1B
negative regulation of optical nerve axon regeneration	GO:1905592	2/7	0.007	28.6		PTPRS;RTN4RL1	
lens fiber cell development	GO:0070307	2/7	0.007	28.6	WNT5B	MAF	
positive regulation of mesenchymal cell proliferation involved in ureter development	GO:2000729	3/11	0.001	27.3	CHRD	LRP5;FGFR2	
positive regulation of ossification	GO:0045778	2/8	0.009	25.0	SOX11		WNT5A
positive regulation of cilium movement	GO:0003353	2/8	0.009	25.0		RPS19;ETS1	
positive regulation of actin filament-based movement	GO:1903116	2/8	0.009	25.0		RPS19;ETS1	
hindlimb morphogenesis	GO:0035137	2/8	0.009	25.0		TBX4	BMPR1B
hexose transmembrane transport	GO:0035428	2/8	0.009	25.0		SLC2A12;SLC2A5	
fat body development	GO:0007503	2/8	0.009	25.0		ZNF516;LRP5	
endochondral bone morphogenesis	GO:0060350	3/13	0.002	23.1		LRP5;FGFR2	BMPR1B
post-embryonic camera-type eye morphogenesis	GO:0048597	2/9	0.011	22.2	TENM3	IFT122	
positive regulation of extrathymic T cell differentiation	GO:0033090	2/9	0.011	22.2	ZMIZ1;PRDM1		
positive regulation of cellular component movement	GO:0051272	2/9	0.011	22.2		RPS19;ETS1	
negative regulation of dendritic spine development	GO:0061000	2/9	0.011	22.2	DCC	PTPRS	
fat pad development	GO:0060613	2/9	0.011	22.2		ZNF516;LRP5	
establishment of epithelial cell apical/basal polarity involved in camera-type eye morphogenesis	GO:0003412	2/9	0.011	22.2	TENM3	IFT122	
adipose tissue development	GO:0060612	2/9	0.011	22.2		ZNF516;LRP5	

Table 3-4 Differential genes detected by RNA-seq and ATAC-seq.

geneName	geneID	geneType	logFC.rna	logCPM.rna	FDR.rna	venn_diagram
ZNF775	ENSG00000196456	protein_coding	-0.418	3.309	0.009	rna __na enhancer
ZNF740	ENSG00000139651	protein_coding	-0.337	4.624	0.016	rna __na enhancer
ZNF662	ENSG00000182983	protein_coding	-0.367	3.310	0.042	rna __na enhancer
ZNF648	ENSG00000179930	protein_coding	-1.141	0.848	0.016	rna __na enhancer
ZNF592	ENSG00000166716	protein_coding	-0.283	4.926	0.031	rna __na enhancer
ZNF516	ENSG00000101493	protein_coding	-0.695	2.375	0.004	rna __na enhancer
ZNF273	ENSG00000198039	protein_coding	-0.473	2.705	0.016	rna __na enhancer
ZNF107	ENSG00000196247	protein_coding	-0.456	3.826	0.011	rna __na enhancer
ZMIZ1	ENSG00000108175	protein_coding	-0.732	7.696	0.007	rna promoter __na
ZIC4	ENSG00000174963	protein_coding	1.260	0.236	0.019	rna promoter __na
ZIC1	ENSG00000152977	protein_coding	0.810	2.676	0.025	rna promoter __na
ZDHHC2	ENSG00000104219	protein_coding	0.641	4.820	0.008	rna __na enhancer
YWHAZ	ENSG00000164924	protein_coding	0.455	7.892	0.004	rna __na enhancer
YKT6	ENSG00000106636	protein_coding	0.244	6.147	0.024	rna __na enhancer
YEATS4	ENSG00000127337	protein_coding	0.442	2.921	0.014	rna __na enhancer
WWP2	ENSG00000198373	protein_coding	-0.761	9.389	0.005	rna __na enhancer
WNT9A	ENSG00000143816	protein_coding	0.800	2.292	0.029	rna promoter __na
WNT5B	ENSG00000111186	protein_coding	0.791	3.189	0.025	rna promoter __na
WNT5A	ENSG00000114251	protein_coding	1.247	0.766	0.033	rna promoter enhancer
WNK2	ENSG00000165238	protein_coding	-1.553	3.500	0.005	rna __na enhancer
VWC2	ENSG00000188730	protein_coding	1.203	2.751	0.005	rna promoter __na
VRK3	ENSG00000105053	protein_coding	-0.437	2.917	0.009	rna __na enhancer
VPS53	ENSG00000141252	protein_coding	-0.245	3.957	0.037	rna __na enhancer
VIT	ENSG00000205221	protein_coding	-1.830	6.609	0.014	rna __na enhancer
USP2	ENSG00000036672	protein_coding	0.616	1.241	0.039	rna __na enhancer
UROD	ENSG00000126088	protein_coding	0.267	4.831	0.030	rna __na enhancer
UBLCP1	ENSG00000164332	protein_coding	0.534	4.288	0.005	rna __na enhancer
UBL3	ENSG00000122042	protein_coding	0.473	5.783	0.004	rna __na enhancer
UBAP1L	ENSG00000246922	protein_coding	-0.890	1.785	0.037	rna __na enhancer
UBAC2	ENSG00000134882	protein_coding	0.648	5.645	0.003	rna __na enhancer
TXNDC5	ENSG00000239264	protein_coding	0.259	7.772	0.018	rna __na enhancer
TUBA1A	ENSG00000167552	protein_coding	0.615	6.885	0.040	rna __na enhancer
TTC28	ENSG00000100154	protein_coding	-0.434	4.524	0.011	rna __na enhancer
TSPYL5	ENSG00000180543	protein_coding	-0.286	5.176	0.017	rna __na enhancer
TSEN15	ENSG00000198860	protein_coding	0.369	4.644	0.014	rna __na enhancer
TSC22D1	ENSG00000102804	protein_coding	-0.502	9.713	0.019	rna __na enhancer
TRIM36	ENSG00000152503	protein_coding	2.298	1.545	0.008	rna promoter __na
TRANK1	ENSG00000168016	protein_coding	-0.588	3.316	0.019	rna __na enhancer
TRAF1	ENSG00000056558	protein_coding	-0.747	3.936	0.004	rna __na enhancer
TOM1L1	ENSG00000141198	protein_coding	1.610	2.373	0.005	rna promoter __na

TNS3	ENSG00000136205	protein_coding	-0.567	8.527	0.005	rna __na enhancer
TMEM8B	ENSG00000137103	protein_coding	-0.603	4.156	0.003	rna __na enhancer
TMEM59L	ENSG00000105696	protein_coding	2.220	1.995	0.012	rna promoter __na
TMEM37	ENSG00000171227	protein_coding	-0.844	1.189	0.042	rna __na enhancer
TMEM14B	ENSG00000137210	protein_coding	0.312	5.106	0.046	rna __na enhancer
TMCO3	ENSG00000150403	protein_coding	0.355	7.058	0.013	rna __na enhancer
THBS1	ENSG00000137801	protein_coding	-0.577	9.426	0.037	rna __na enhancer
TGFBI	ENSG00000120708	protein_coding	1.023	9.654	0.035	rna __na enhancer
TEX30	ENSG00000151287	protein_coding	0.581	1.704	0.039	rna __na enhancer
TENM3	ENSG00000218336	protein_coding	1.584	2.780	0.008	rna promoter __na
TEN1	ENSG00000257949	protein_coding	0.404	2.287	0.024	rna __na enhancer
TBX4	ENSG00000121075	protein_coding	-0.810	4.995	0.010	rna __na enhancer
TBPL1	ENSG0000028839	protein_coding	0.439	3.503	0.014	rna __na enhancer
TBL1X	ENSG00000101849	protein_coding	-0.383	3.556	0.027	rna __na enhancer
TATDN3	ENSG00000203705	protein_coding	0.352	2.778	0.021	rna __na enhancer
TATDN1	ENSG00000147687	protein_coding	0.315	3.777	0.041	rna __na enhancer
TARS	ENSG00000113407	protein_coding	0.304	5.463	0.033	rna __na enhancer
STK32A	ENSG00000169302	protein_coding	-0.718	4.777	0.034	rna __na enhancer
ST6GALNAC5	ENSG00000117069	protein_coding	1.545	4.163	0.004	rna promoter __na
ST6GAL2	ENSG00000144057	protein_coding	1.465	0.758	0.048	rna promoter __na
ST3GAL6	ENSG00000064225	protein_coding	-0.733	2.740	0.020	rna __na enhancer
SSR1	ENSG00000124783	protein_coding	0.219	7.768	0.047	rna __na enhancer
SPTLC3	ENSG00000172296	protein_coding	0.417	2.646	0.043	rna __na enhancer
SPRED2	ENSG00000198369	protein_coding	-0.372	4.693	0.024	rna __na enhancer
SPECC1	ENSG00000128487	protein_coding	0.758	5.856	0.009	rna __na enhancer
SPATA6	ENSG00000132122	protein_coding	-0.412	2.789	0.047	rna __na enhancer
SOX11	ENSG00000176887	protein_coding	2.471	2.233	0.022	rna promoter __na
SOCS5	ENSG00000171150	protein_coding	0.346	5.023	0.035	rna __na enhancer
SNRNP25	ENSG00000161981	protein_coding	0.309	3.221	0.042	rna __na enhancer
SMIM5	ENSG00000204323	protein_coding	0.881	2.869	0.007	rna __na enhancer
SLC7A2	ENSG00000003989	protein_coding	1.216	7.892	0.011	rna __na enhancer
SLC44A1	ENSG00000070214	protein_coding	0.290	5.270	0.021	rna __na enhancer
SLC43A3	ENSG00000134802	protein_coding	-0.668	6.480	0.003	rna __na enhancer
SLC43A1	ENSG00000149150	protein_coding	-0.400	2.345	0.016	rna __na enhancer
SLC38A5	ENSG00000017483	protein_coding	1.088	2.946	0.024	rna __na enhancer
SLC35B1	ENSG00000121073	protein_coding	0.360	4.834	0.012	rna __na enhancer
SLC2A5	ENSG00000142583	protein_coding	1.656	3.203	0.009	rna __na enhancer
SLC2A12	ENSG00000146411	protein_coding	0.854	4.881	0.019	rna __na enhancer
SLC24A3	ENSG00000185052	protein_coding	1.125	0.991	0.029	rna promoter enhancer
SLC15A4	ENSG00000139370	protein_coding	0.474	4.498	0.008	rna __na enhancer
SIDT2	ENSG00000149577	protein_coding	-0.527	5.426	0.003	rna __na enhancer
SH3RF2	ENSG00000156463	protein_coding	-0.872	1.651	0.016	rna __na enhancer
SH3BP5	ENSG00000131370	protein_coding	0.473	5.773	0.011	rna __na enhancer

SGPL1	ENSG00000166224	protein_coding	0.340	3.414	0.024	rna __na enhancer
SGK1	ENSG00000118515	protein_coding	1.478	6.384	0.002	rna __na enhancer
SETBP1	ENSG00000152217	protein_coding	-0.724	5.050	0.005	rna __na enhancer
SERTAD4	ENSG00000082497	protein_coding	-0.453	7.270	0.021	rna __na enhancer
SERPINE2	ENSG00000135919	protein_coding	1.462	8.253	0.006	rna __na enhancer
SERPINA3	ENSG00000196136	protein_coding	-0.695	11.820	0.007	rna __na enhancer
SEMA6C	ENSG00000143434	protein_coding	-0.441	2.208	0.031	rna __na enhancer
SEMA5A	ENSG00000112902	protein_coding	1.319	2.745	0.023	rna __na enhancer
SEC22A	ENSG00000121542	protein_coding	0.292	3.290	0.043	rna __na enhancer
SEC11C	ENSG00000166562	protein_coding	0.325	4.297	0.037	rna __na enhancer
SDC3	ENSG00000162512	protein_coding	-1.094	7.039	0.002	rna __na enhancer
SCUBE2	ENSG00000175356	protein_coding	-0.996	3.976	0.003	rna __na enhancer
SCN4A	ENSG00000007314	protein_coding	-1.370	-0.192	0.006	rna __na enhancer
S100B	ENSG00000160307	protein_coding	-0.500	7.558	0.018	rna __na enhancer
RTN4RL1	ENSG00000185924	protein_coding	-0.755	4.608	0.003	rna __na enhancer
RTKN2	ENSG00000182010	protein_coding	1.844	-1.662	0.004	rna promoter __na
RRM2	ENSG00000171848	protein_coding	1.323	0.713	0.023	rna __na enhancer
RPS8	ENSG00000142937	protein_coding	0.367	9.224	0.031	rna __na enhancer
RPS3A	ENSG00000145425	protein_coding	0.395	10.041	0.034	rna __na enhancer
RPS19	ENSG00000105372	protein_coding	0.361	8.087	0.041	rna __na enhancer
RP11-182J1.16	ENSG00000259511	protein_coding	-1.361	2.230	0.007	rna __na enhancer
RP11-111M22.2	ENSG00000179240	protein_coding	-0.586	1.627	0.028	rna __na enhancer
RORA	ENSG00000069667	protein_coding	0.554	6.170	0.005	rna __na enhancer
ROR2	ENSG00000169071	protein_coding	-0.770	4.836	0.005	rna __na enhancer
ROR1	ENSG00000185483	protein_coding	0.809	2.676	0.041	rna promoter enhancer
RNF167	ENSG00000108523	protein_coding	-0.269	5.495	0.040	rna __na enhancer
RNF152	ENSG00000176641	protein_coding	0.929	4.860	0.010	rna __na enhancer
RNF150	ENSG00000170153	protein_coding	-0.982	1.364	0.046	rna promoter __na
RNF139	ENSG00000170881	protein_coding	0.531	4.985	0.003	rna __na enhancer
RIT1	ENSG00000143622	protein_coding	0.482	3.384	0.009	rna __na enhancer
RHPN2	ENSG00000131941	protein_coding	0.896	0.896	0.041	rna promoter __na
RHOBTB3	ENSG00000164292	protein_coding	-0.292	6.193	0.046	rna __na enhancer
RGAG4	ENSG00000242732	protein_coding	-0.539	3.249	0.006	rna __na enhancer
REEP3	ENSG00000165476	protein_coding	0.298	6.148	0.024	rna __na enhancer
RDH14	ENSG00000240857	protein_coding	0.442	3.161	0.008	rna __na enhancer
RBBP8	ENSG00000101773	protein_coding	0.526	3.032	0.013	rna __na enhancer
RASL11B	ENSG00000128045	protein_coding	1.044	-0.670	0.018	rna __na enhancer
RARRES2	ENSG00000106538	protein_coding	-1.381	4.048	0.029	rna __na enhancer
RAD51AP2	ENSG00000214842	protein_coding	1.346	0.218	0.007	rna promoter __na
RAD51AP1	ENSG00000111247	protein_coding	1.037	-0.071	0.013	rna __na enhancer
RAD51	ENSG00000051180	protein_coding	1.499	-0.842	0.017	rna __na enhancer
RAB36	ENSG00000100228	protein_coding	-0.479	1.528	0.037	rna __na enhancer
PTPRS	ENSG00000105426	protein_coding	-0.823	5.880	0.005	rna __na enhancer

PTGES	ENSG00000148344	protein_coding	1.382	5.799	0.022	rna __na enhancer
PSMD6	ENSG00000163636	protein_coding	0.440	4.922	0.008	rna __na enhancer
PSMD1	ENSG00000173692	protein_coding	0.346	6.124	0.012	rna __na enhancer
PRUNE	ENSG00000143363	protein_coding	-0.225	4.204	0.043	rna __na enhancer
PREX1	ENSG00000124126	protein_coding	-0.481	6.731	0.007	rna __na enhancer
PRDM11	ENSG0000019485	protein_coding	-0.433	3.395	0.013	rna __na enhancer
PRDM1	ENSG00000057657	protein_coding	1.993	0.552	0.012	rna promoter __na
PPP1R32	ENSG00000162148	protein_coding	-0.841	0.333	0.031	rna __na enhancer
PPHLN1	ENSG00000134283	protein_coding	0.208	4.873	0.032	rna __na enhancer
POMGNT2	ENSG00000144647	protein_coding	-0.270	3.373	0.031	rna __na enhancer
POM121C	ENSG00000135213	protein_coding	-0.331	4.500	0.013	rna __na enhancer
PODXL2	ENSG00000114631	protein_coding	0.593	2.728	0.031	rna promoter __na
PODXL	ENSG00000128567	protein_coding	0.711	3.452	0.044	rna promoter __na
PMAIP1	ENSG00000141682	protein_coding	1.645	0.408	0.005	rna promoter __na
PLEKHG3	ENSG00000126822	protein_coding	-0.312	4.554	0.031	rna __na enhancer
PLEKHA6	ENSG00000143850	protein_coding	-0.794	0.559	0.025	rna __na enhancer
PLCG1	ENSG00000124181	protein_coding	-0.491	5.668	0.004	rna __na enhancer
PIK3R1	ENSG00000145675	protein_coding	-0.728	7.355	0.012	rna __na enhancer
PIGN	ENSG00000197563	protein_coding	0.365	4.455	0.047	rna __na enhancer
PHF10	ENSG00000130024	protein_coding	0.250	6.229	0.032	rna __na enhancer
PEG3	ENSG00000198300	protein_coding	-0.676	4.470	0.037	rna __na enhancer
PDE4D	ENSG00000113448	protein_coding	0.406	5.320	0.035	rna __na enhancer
PAX1	ENSG00000125813	protein_coding	2.901	0.192	0.011	rna promoter __na
PAK1IP1	ENSG00000111845	protein_coding	0.312	3.876	0.022	rna __na enhancer
P4HA3	ENSG00000149380	protein_coding	1.179	5.227	0.026	rna promoter __na
OSBPL3	ENSG00000070882	protein_coding	0.923	3.404	0.023	rna __na enhancer
ODC1	ENSG00000115758	protein_coding	0.430	4.999	0.012	rna __na enhancer
OBSCN	ENSG00000154358	protein_coding	-1.109	4.559	0.004	rna __na enhancer
NUDCD1	ENSG00000120526	protein_coding	0.307	3.438	0.038	rna __na enhancer
NTRK3	ENSG00000140538	protein_coding	-2.662	1.025	0.012	rna __na enhancer
NT5E	ENSG00000135318	protein_coding	1.030	8.599	0.007	rna __na enhancer
NRIP3	ENSG00000175352	protein_coding	1.058	0.488	0.040	rna promoter __na
NPR3	ENSG00000113389	protein_coding	2.465	2.818	0.015	rna __na enhancer
NPAS2	ENSG00000170485	protein_coding	-0.341	3.710	0.041	rna __na enhancer
NOA1	ENSG00000084092	protein_coding	0.279	3.812	0.031	rna __na enhancer
NKX3-2	ENSG00000109705	protein_coding	-0.622	4.697	0.009	rna __na enhancer
NGF	ENSG00000134259	protein_coding	1.333	3.034	0.016	rna __na enhancer
NFIL3	ENSG00000165030	protein_coding	1.019	5.284	0.049	rna __na enhancer
NFIA	ENSG00000162599	protein_coding	-0.461	5.533	0.030	rna __na enhancer
NFATC2	ENSG00000101096	protein_coding	0.581	7.142	0.047	rna __na enhancer
NEDD9	ENSG00000111859	protein_coding	1.074	3.556	0.010	rna __na enhancer
NDUFB9	ENSG00000147684	protein_coding	0.315	5.190	0.015	rna __na enhancer
NCOA6	ENSG00000198646	protein_coding	-0.350	4.579	0.044	rna __na enhancer

NALCN	ENSG00000102452	protein_coding	1.559	-1.609	0.032	rna __na enhancer
NAE1	ENSG00000159593	protein_coding	0.294	4.681	0.035	rna __na enhancer
MYO6	ENSG00000196586	protein_coding	0.509	4.479	0.006	rna __na enhancer
MYO5A	ENSG00000197535	protein_coding	0.425	4.907	0.012	rna __na enhancer
MYO1B	ENSG00000128641	protein_coding	0.771	6.002	0.002	rna __na enhancer
MYH14	ENSG00000105357	protein_coding	-1.436	3.634	0.003	rna __na enhancer
MUC20	ENSG00000176945	protein_coding	-0.909	3.817	0.004	rna __na enhancer
MTSS1L	ENSG00000132613	protein_coding	-0.649	5.099	0.004	rna __na enhancer
MTSS1	ENSG00000170873	protein_coding	1.419	2.252	0.037	rna promoter enhancer
MSX2	ENSG00000120149	protein_coding	2.936	2.620	0.003	rna promoter enhancer
MRPS10	ENSG00000048544	protein_coding	0.409	3.721	0.009	rna __na enhancer
MPPED2	ENSG00000066382	protein_coding	-1.089	3.342	0.004	rna __na enhancer
MOK	ENSG00000080823	protein_coding	-0.360	2.886	0.015	rna __na enhancer
MITF	ENSG00000187098	protein_coding	0.466	2.798	0.031	rna __na enhancer
MICAL2	ENSG00000133816	protein_coding	0.696	5.435	0.048	rna __na enhancer
MFSD11	ENSG00000092931	protein_coding	0.366	4.359	0.005	rna __na enhancer
MFHAS1	ENSG00000147324	protein_coding	-0.506	3.654	0.016	rna __na enhancer
MERTK	ENSG00000153208	protein_coding	-0.545	1.992	0.008	rna __na enhancer
MED13	ENSG00000108510	protein_coding	0.397	5.277	0.018	rna __na enhancer
MARK2	ENSG00000072518	protein_coding	-0.338	3.796	0.029	rna __na enhancer
MAP3K2	ENSG00000169967	protein_coding	0.336	5.784	0.029	rna __na enhancer
MAP1B	ENSG00000131711	protein_coding	1.137	3.415	0.020	rna promoter __na
MAOB	ENSG00000069535	protein_coding	-0.746	7.116	0.004	rna __na enhancer
MAF	ENSG00000178573	protein_coding	-0.853	5.480	0.005	rna __na enhancer
LZTS1	ENSG00000061337	protein_coding	-0.582	3.299	0.010	rna __na enhancer
LRR8C	ENSG00000171488	protein_coding	1.387	3.940	0.004	rna __na enhancer
LRR8B	ENSG00000197147	protein_coding	0.932	0.948	0.018	rna __na enhancer
LRR1	ENSG00000137269	protein_coding	0.417	3.645	0.038	rna __na enhancer
LRP5L	ENSG00000100068	protein_coding	-0.736	0.903	0.011	rna __na enhancer
LRP5	ENSG00000162337	protein_coding	-0.637	4.529	0.004	rna __na enhancer
LRP11	ENSG00000120256	protein_coding	0.227	5.707	0.037	rna __na enhancer
LRIG1	ENSG00000144749	protein_coding	-0.707	6.054	0.005	rna __na enhancer
LIMCH1	ENSG00000064042	protein_coding	-0.420	7.594	0.012	rna __na enhancer
LGALS3	ENSG00000131981	protein_coding	0.494	8.066	0.005	rna __na enhancer
LEMD2	ENSG00000161904	protein_coding	-0.194	4.783	0.039	rna __na enhancer
LAMA1	ENSG00000101680	protein_coding	-0.673	-0.077	0.035	rna __na enhancer
KRBOX1	ENSG00000273291	protein_coding	-1.406	0.616	0.014	rna __na enhancer
KIAA1161	ENSG00000164976	protein_coding	-0.716	3.857	0.010	rna __na enhancer
KIAA0196	ENSG00000164961	protein_coding	0.220	4.952	0.043	rna __na enhancer
KCNS3	ENSG00000170745	protein_coding	1.539	1.810	0.009	rna promoter enhancer
KCNIP3	ENSG00000115041	protein_coding	-0.743	1.775	0.025	rna __na enhancer
KCNH1	ENSG00000143473	protein_coding	-1.441	-1.122	0.021	rna __na enhancer
ITPR1	ENSG00000150995	protein_coding	0.610	4.126	0.022	rna __na enhancer

ITIH5	ENSG00000123243	protein_coding	-0.432	6.401	0.038	rna __na enhancer
ITGA4	ENSG00000115232	protein_coding	1.315	0.163	0.005	rna promoter __na
INHBA	ENSG00000122641	protein_coding	0.853	7.680	0.013	rna __na enhancer
IL17RB	ENSG00000056736	protein_coding	-1.051	4.866	0.010	rna __na enhancer
IL11	ENSG00000095752	protein_coding	4.101	3.706	0.025	rna promoter __na
IGFBP7	ENSG00000163453	protein_coding	1.279	8.044	0.012	rna promoter enhancer
IFT122	ENSG00000163913	protein_coding	-0.336	4.155	0.025	rna __na enhancer
IAH1	ENSG00000134330	protein_coding	0.376	4.600	0.008	rna __na enhancer
HUNK	ENSG00000142149	protein_coding	1.600	1.033	0.006	rna promoter __na
HSD3B7	ENSG00000099377	protein_coding	0.755	4.794	0.022	rna __na enhancer
HOXD8	ENSG00000175879	protein_coding	0.368	4.581	0.024	rna __na enhancer
HOXB2	ENSG00000173917	protein_coding	1.598	2.415	0.003	rna promoter __na
HMGA2	ENSG00000149948	protein_coding	2.064	0.218	0.021	rna promoter __na
HES6	ENSG00000144485	protein_coding	2.010	-0.460	0.008	rna promoter __na
HDDC2	ENSG00000111906	protein_coding	0.240	5.493	0.043	rna __na enhancer
HAT1	ENSG00000128708	protein_coding	0.321	5.105	0.014	rna __na enhancer
HAAO	ENSG00000162882	protein_coding	-0.518	2.814	0.016	rna __na enhancer
H6PD	ENSG00000049239	protein_coding	-0.308	6.176	0.027	rna __na enhancer
GUCY1A3	ENSG00000164116	protein_coding	-0.987	3.738	0.021	rna __na enhancer
GSTO2	ENSG00000065621	protein_coding	-0.498	3.029	0.011	rna __na enhancer
GRIN2C	ENSG00000161509	protein_coding	-1.120	2.442	0.012	rna __na enhancer
GRIA2	ENSG00000120251	protein_coding	1.587	3.646	0.011	rna promoter __na
GRAMD1A	ENSG00000089351	protein_coding	-0.392	4.548	0.023	rna __na enhancer
GPR176	ENSG00000166073	protein_coding	0.765	2.567	0.017	rna promoter __na
GPC4	ENSG00000076716	protein_coding	1.304	2.329	0.005	rna promoter __na
GMDS	ENSG00000112699	protein_coding	-0.517	5.421	0.003	rna __na enhancer
GLYATL2	ENSG00000156689	protein_coding	1.620	0.501	0.011	rna promoter __na
GGT7	ENSG00000131067	protein_coding	-0.349	4.511	0.014	rna __na enhancer
GDF7	ENSG00000143869	protein_coding	-1.549	1.799	0.012	rna promoter __na
GDF10	ENSG00000107623	protein_coding	-1.052	9.343	0.007	rna __na enhancer
GALNT13	ENSG00000144278	protein_coding	1.361	1.879	0.041	rna promoter __na
FZD10	ENSG00000111432	protein_coding	1.037	2.677	0.029	rna __na enhancer
FXYD3	ENSG00000089356	protein_coding	-0.877	1.288	0.018	rna __na enhancer
FSTL3	ENSG00000070404	protein_coding	0.936	4.046	0.029	rna promoter enhancer
FOXF1	ENSG00000103241	protein_coding	2.301	0.558	0.003	rna __na enhancer
FNIP2	ENSG00000052795	protein_coding	0.880	5.953	0.007	rna __na enhancer
FHL2	ENSG00000115641	protein_coding	1.012	3.541	0.007	rna promoter __na
FGFR2	ENSG00000066468	protein_coding	-0.780	7.732	0.008	rna __na enhancer
FGF9	ENSG00000102678	protein_coding	1.811	0.395	0.012	rna promoter __na
FGD6	ENSG00000180263	protein_coding	-0.696	2.469	0.028	rna __na enhancer
FDX1	ENSG00000137714	protein_coding	0.645	2.569	0.025	rna __na enhancer
FBRSL1	ENSG00000112787	protein_coding	-0.392	3.527	0.032	rna __na enhancer
FBLN7	ENSG00000144152	protein_coding	-1.049	7.333	0.003	rna __na enhancer

FBLN2	ENSG00000163520	protein_coding	1.001	2.610	0.038	rna __na enhancer
FAM78B	ENSG00000188859	protein_coding	-1.262	1.303	0.010	rna __na enhancer
FAM63A	ENSG00000143409	protein_coding	-0.436	3.875	0.005	rna __na enhancer
FAM198A	ENSG00000144649	protein_coding	-1.292	3.383	0.006	rna __na enhancer
FAM177B	ENSG00000197520	protein_coding	1.283	-1.178	0.024	rna __na enhancer
FAM101A	ENSG00000178882	protein_coding	-1.343	3.618	0.005	rna __na enhancer
F3	ENSG00000117525	protein_coding	1.053	1.364	0.011	rna __na enhancer
ETV4	ENSG00000175832	protein_coding	0.661	1.514	0.045	rna __na enhancer
ETS1	ENSG00000134954	protein_coding	-0.541	6.914	0.008	rna __na enhancer
ERG	ENSG00000157554	protein_coding	-0.446	7.201	0.027	rna __na enhancer
EPB41L1	ENSG00000088367	protein_coding	-0.371	5.463	0.005	rna __na enhancer
EMP2	ENSG00000213853	protein_coding	-0.289	7.159	0.043	rna __na enhancer
EML1	ENSG00000066629	protein_coding	0.429	5.921	0.022	rna __na enhancer
ELK4	ENSG00000158711	protein_coding	0.480	4.755	0.010	rna promoter __na
EHD3	ENSG00000013016	protein_coding	-0.479	5.434	0.020	rna __na enhancer
EGR2	ENSG00000122877	protein_coding	1.401	3.775	0.030	rna __na enhancer
EEPD1	ENSG00000122547	protein_coding	-0.795	1.067	0.027	rna __na enhancer
ECH1	ENSG00000104823	protein_coding	-0.310	4.932	0.017	rna __na enhancer
EBF3	ENSG00000108001	protein_coding	2.084	0.823	0.007	rna promoter __na
DYSF	ENSG00000135636	protein_coding	0.853	5.391	0.022	rna __na enhancer
DYNC111	ENSG00000158560	protein_coding	-0.781	5.033	0.003	rna __na enhancer
DUSP4	ENSG00000120875	protein_coding	1.594	1.733	0.014	rna promoter enhancer
DSG2	ENSG00000046604	protein_coding	1.047	5.995	0.006	rna __na enhancer
DNAJC22	ENSG00000178401	protein_coding	1.021	2.997	0.004	rna promoter __na
DLX4	ENSG00000108813	protein_coding	0.759	3.289	0.039	rna promoter __na
DLGAP1	ENSG00000170579	protein_coding	-0.993	0.174	0.032	rna __na enhancer
DKK3	ENSG00000050165	protein_coding	1.224	8.156	0.010	rna promoter __na
DIRAS1	ENSG00000176490	protein_coding	1.113	3.654	0.004	rna promoter enhancer
DHX32	ENSG00000089876	protein_coding	0.222	3.886	0.040	rna __na enhancer
DENND2D	ENSG00000162777	protein_coding	-0.474	3.539	0.019	rna __na enhancer
DEGS1	ENSG00000143753	protein_coding	0.329	6.165	0.027	rna __na enhancer
DCUN1D3	ENSG00000188215	protein_coding	0.346	2.764	0.047	rna __na enhancer
DCC	ENSG00000187323	protein_coding	-2.581	3.705	0.002	rna promoter __na
DAAM2	ENSG00000146122	protein_coding	-0.491	6.506	0.006	rna __na enhancer
CYP27B1	ENSG00000111012	protein_coding	0.614	1.216	0.032	rna promoter __na
CUL4A	ENSG00000139842	protein_coding	0.229	5.059	0.041	rna __na enhancer
CTNNB1	ENSG00000168036	protein_coding	0.358	7.674	0.016	rna __na enhancer
CTGF	ENSG00000118523	protein_coding	0.585	12.225	0.019	rna __na enhancer
CSDC2	ENSG00000172346	protein_coding	1.010	1.204	0.032	rna __na enhancer
CSAD	ENSG00000139631	protein_coding	-0.436	3.639	0.042	rna __na enhancer
CRYL1	ENSG00000165475	protein_coding	-0.313	4.471	0.044	rna __na enhancer
CRTC3	ENSG00000140577	protein_coding	-0.203	4.521	0.049	rna __na enhancer
CRLF1	ENSG00000006016	protein_coding	2.165	7.398	0.010	rna promoter __na

CREB1	ENSG00000118260	protein_coding	0.228	4.633	0.041	rna __na enhancer
COMMD2	ENSG00000114744	protein_coding	0.343	4.606	0.029	rna __na enhancer
COLCA2	ENSG00000214290	protein_coding	-1.087	-0.929	0.017	rna __na enhancer
COL9A3	ENSG00000092758	protein_coding	-0.820	10.899	0.009	rna __na enhancer
COL8A1	ENSG00000144810	protein_coding	2.144	4.563	0.015	rna __na enhancer
COL18A1	ENSG00000182871	protein_coding	1.128	5.301	0.008	rna promoter __na
CNOT7	ENSG00000198791	protein_coding	0.223	5.540	0.025	rna __na enhancer
CMTM4	ENSG00000183723	protein_coding	0.333	4.265	0.027	rna __na enhancer
CLDN23	ENSG00000253958	protein_coding	-0.753	1.770	0.029	rna __na enhancer
CITED4	ENSG00000179862	protein_coding	1.306	1.866	0.015	rna __na enhancer
CHRD	ENSG00000090539	protein_coding	0.603	3.273	0.023	rna promoter __na
CHDH	ENSG00000016391	protein_coding	-0.682	4.440	0.014	rna __na enhancer
CFDP1	ENSG00000153774	protein_coding	-0.355	5.283	0.030	rna __na enhancer
CENPK	ENSG00000123219	protein_coding	1.003	1.011	0.017	rna __na enhancer
CDYL2	ENSG00000166446	protein_coding	0.994	1.505	0.029	rna promoter __na
CDC42EP3	ENSG00000163171	protein_coding	-0.925	5.279	0.008	rna __na enhancer
CCRN4L	ENSG00000151014	protein_coding	0.887	0.826	0.021	rna __na enhancer
CCNYL1	ENSG00000163249	protein_coding	0.598	3.472	0.007	rna __na enhancer
CCNDBP1	ENSG00000166946	protein_coding	0.227	4.633	0.026	rna __na enhancer
CCND1	ENSG00000110092	protein_coding	0.897	5.222	0.014	rna promoter __na
CCDC91	ENSG00000123106	protein_coding	0.428	3.801	0.012	rna __na enhancer
CCDC85A	ENSG00000055813	protein_coding	-0.555	5.237	0.009	rna __na enhancer
CCDC25	ENSG00000147419	protein_coding	0.320	5.247	0.010	rna __na enhancer
CAMTA2	ENSG00000108509	protein_coding	-0.396	4.188	0.022	rna __na enhancer
CALU	ENSG00000128595	protein_coding	0.357	8.873	0.019	rna __na enhancer
CAAP1	ENSG00000120159	protein_coding	0.506	3.046	0.009	rna __na enhancer
C7orf73	ENSG00000243317	protein_coding	0.341	6.199	0.048	rna __na enhancer
C6orf132	ENSG00000188112	protein_coding	1.123	0.820	0.021	rna promoter enhancer
C4orf46	ENSG00000205208	protein_coding	0.658	1.770	0.015	rna __na enhancer
C4BPA	ENSG00000123838	protein_coding	-0.960	3.526	0.005	rna __na enhancer
C2orf69	ENSG00000178074	protein_coding	0.621	3.563	0.003	rna __na enhancer
C1orf56	ENSG00000143443	protein_coding	-0.553	1.083	0.043	rna __na enhancer
C17orf72	ENSG00000224383	protein_coding	-1.244	0.803	0.005	rna __na enhancer
C16orf72	ENSG00000182831	protein_coding	0.307	5.365	0.017	rna __na enhancer
C12orf60	ENSG00000182993	protein_coding	0.385	1.474	0.049	rna __na enhancer
BRD4	ENSG00000141867	protein_coding	-0.360	4.734	0.036	rna __na enhancer
BOC	ENSG00000144857	protein_coding	-0.820	7.744	0.003	rna __na enhancer
BMPR1B	ENSG00000138696	protein_coding	2.309	0.063	0.009	rna promoter enhancer
BEGAIN	ENSG00000183092	protein_coding	-0.472	2.133	0.011	rna __na enhancer
BDNF	ENSG00000176697	protein_coding	1.622	-0.641	0.011	rna __na enhancer
BCAR3	ENSG00000137936	protein_coding	-0.300	3.508	0.016	rna __na enhancer
BAALC	ENSG00000164929	protein_coding	0.958	5.777	0.004	rna __na enhancer

ATP5H	ENSG00000167863	protein_coding	0.415	5.820	0.005	rna __na enhancer
ATP5A1	ENSG00000152234	protein_coding	0.376	7.280	0.036	rna __na enhancer
ATP13A2	ENSG00000159363	protein_coding	-0.395	2.971	0.022	rna __na enhancer
ASAP2	ENSG00000151693	protein_coding	0.594	6.664	0.008	rna __na enhancer
ARSI	ENSG00000183876	protein_coding	0.812	2.258	0.018	rna __na enhancer
ARPP19	ENSG00000128989	protein_coding	0.245	6.229	0.037	rna __na enhancer
ARMC10	ENSG00000170632	protein_coding	0.311	4.615	0.018	rna __na enhancer
ARHGEF4	ENSG00000136002	protein_coding	-0.518	2.289	0.043	rna __na enhancer
ARHGAP44	ENSG00000006740	protein_coding	1.159	1.745	0.009	rna __na enhancer
ARHGAP29	ENSG00000137962	protein_coding	-0.327	5.135	0.050	rna __na enhancer
ARFGAP3	ENSG00000242247	protein_coding	0.379	6.190	0.048	rna __na enhancer
AP2S1	ENSG00000042753	protein_coding	0.448	4.097	0.045	rna __na enhancer
ANXA4	ENSG00000196975	protein_coding	0.416	6.738	0.011	rna __na enhancer
AMPD3	ENSG00000133805	protein_coding	-0.299	3.017	0.033	rna __na enhancer
ALS2CR11	ENSG00000155754	protein_coding	0.601	0.577	0.021	rna __na enhancer
ALS2CL	ENSG00000178038	protein_coding	0.893	3.598	0.013	rna __na enhancer
AKAP8L	ENSG00000011243	protein_coding	-0.233	4.589	0.035	rna __na enhancer
AKAP11	ENSG00000023516	protein_coding	0.313	5.444	0.027	rna __na enhancer
AIDA	ENSG00000186063	protein_coding	0.342	5.349	0.014	rna __na enhancer
AGPS	ENSG00000018510	protein_coding	0.602	4.567	0.003	rna __na enhancer
ADTRP	ENSG00000111863	protein_coding	1.106	3.629	0.008	rna __na enhancer
ADPRHL1	ENSG00000153531	protein_coding	0.830	3.146	0.024	rna __na enhancer
ADO	ENSG00000181915	protein_coding	0.298	3.953	0.024	rna __na enhancer
ADCY3	ENSG00000138031	protein_coding	-0.397	5.069	0.017	rna __na enhancer
ADAMTS6	ENSG00000049192	protein_coding	1.207	5.473	0.006	rna __na enhancer
ADAMTS14	ENSG00000138316	protein_coding	1.713	1.638	0.034	rna __na enhancer
ACCS	ENSG00000110455	protein_coding	-0.431	2.829	0.038	rna __na enhancer
AC104532.2	ENSG00000267314	protein_coding	0.723	2.902	0.025	rna __na enhancer
ABHD2	ENSG00000140526	protein_coding	0.533	7.313	0.018	rna __na enhancer
AAAS	ENSG00000094914	protein_coding	-0.357	4.555	0.016	rna __na enhancer

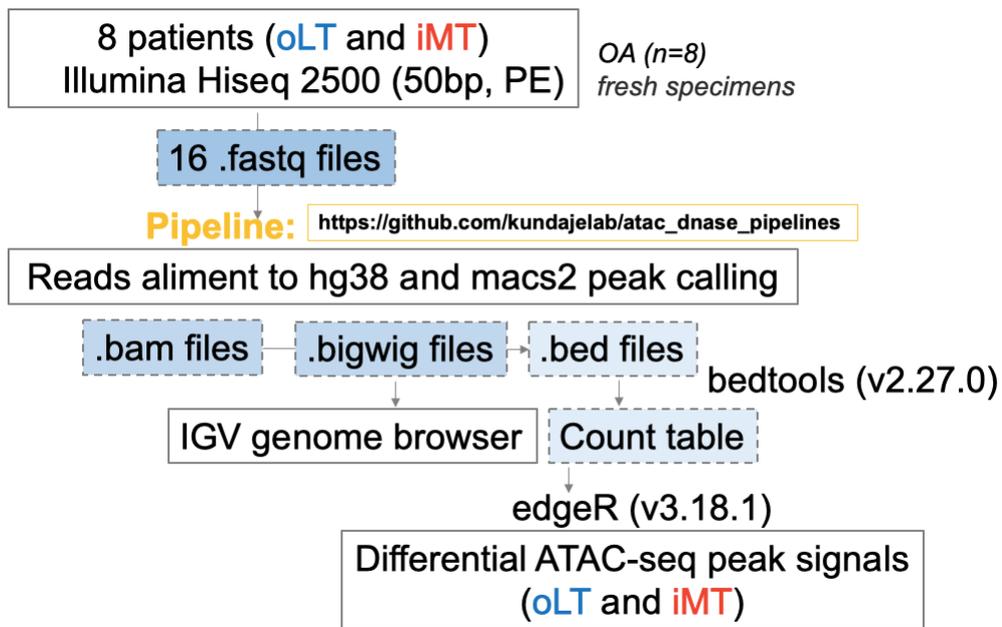


Figure 3-1 The ATAC-seq pipeline schematic.

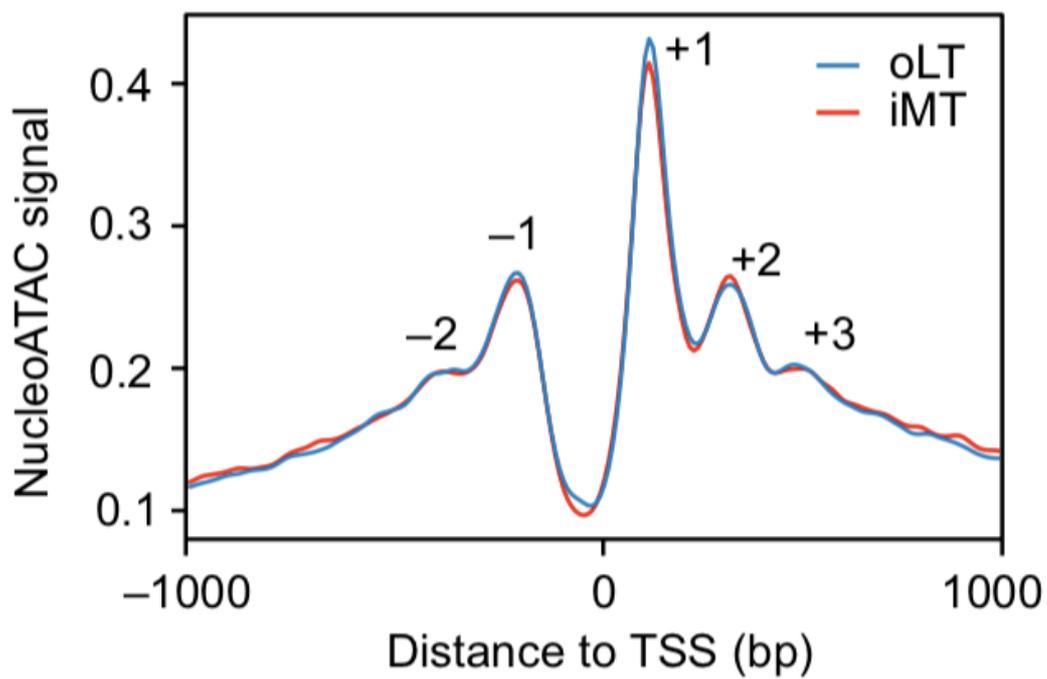


Figure 3-2 Normalized nucleoATAC signal aggregated over all genes shows distinct nucleosome positioning around transcription start sites (TSS).

Positive and negative numbers (+1, +2, +3, -1, -2) indicate the TSS upstream and downstream nucleosomes.

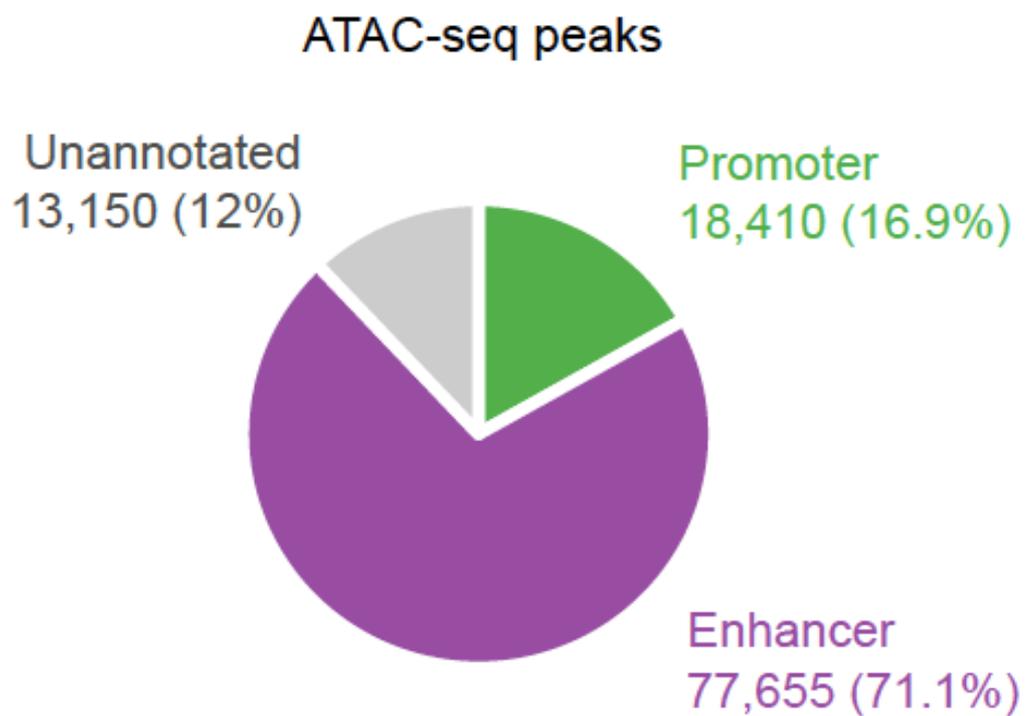


Figure 3-3 Annotation of identified accessible chromatin regions.

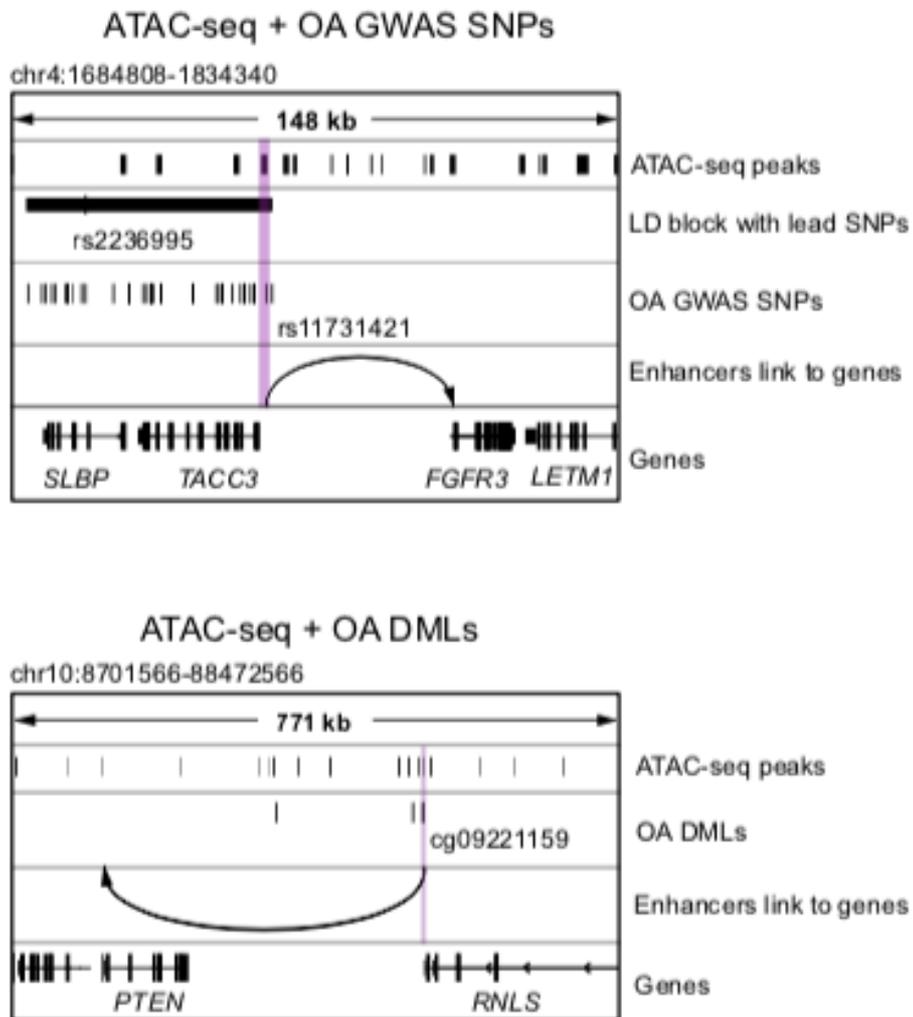


Figure 3-4 Identification of accessible chromatin regions with OA susceptible GWAS SNPs and DMLs.

An example of an OA GWAS SNP rs11731421 (up) and differentially methylated locus cg09221159 (down) overlapping with an open enhancer in cartilages from OA patients. The arrow points to the predicted target genes.

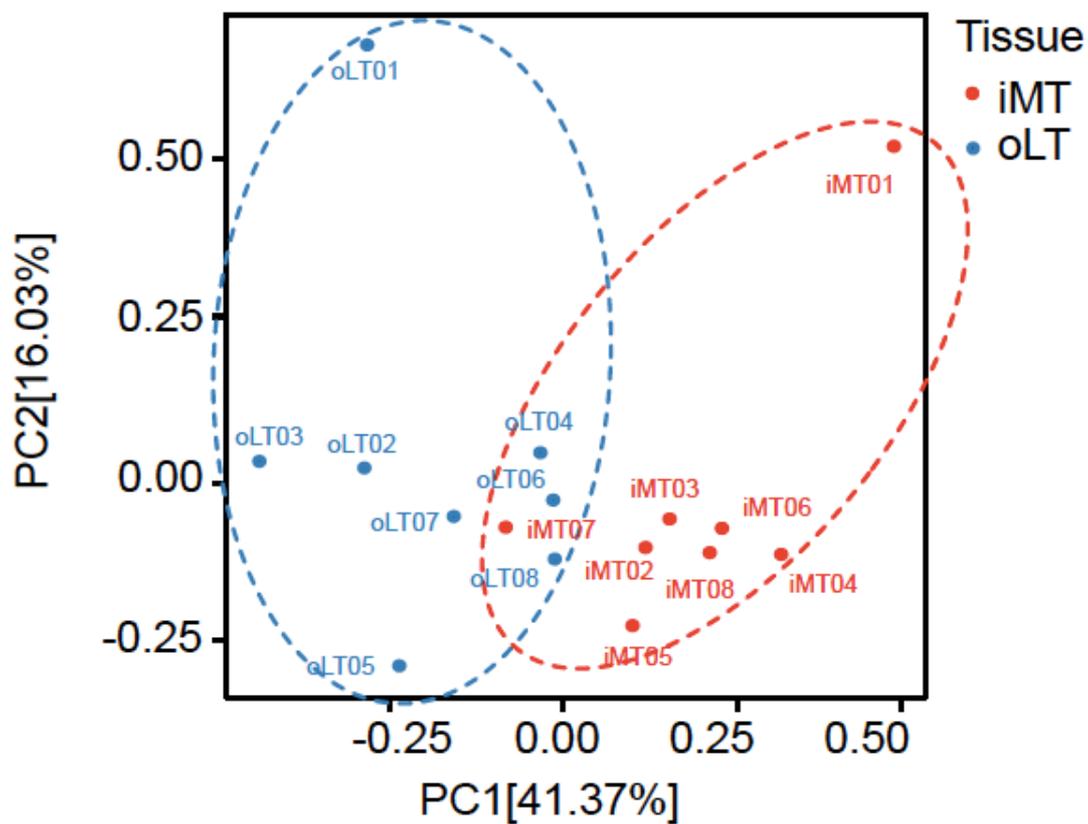


Figure 3-5 Principal component analysis on the peak signals across the 16 samples. PC1: separation between oLT and iMT; PC2: separation between each patient.

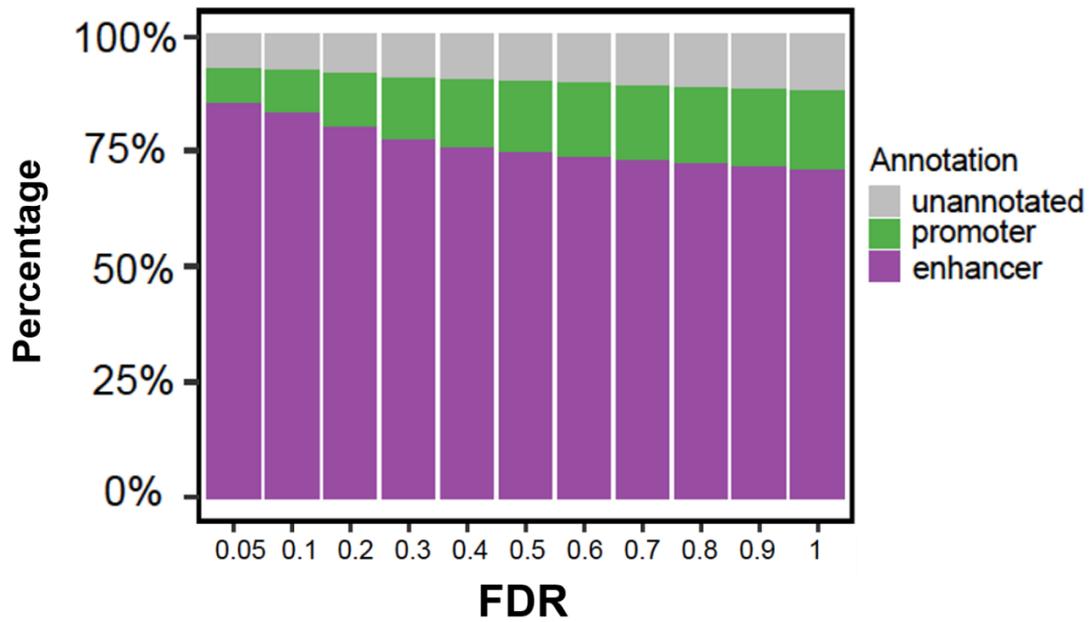


Figure 3-6 Peaks with smaller FDR (i.e. more confident to be differentially accessible) contain more enhancer peaks.

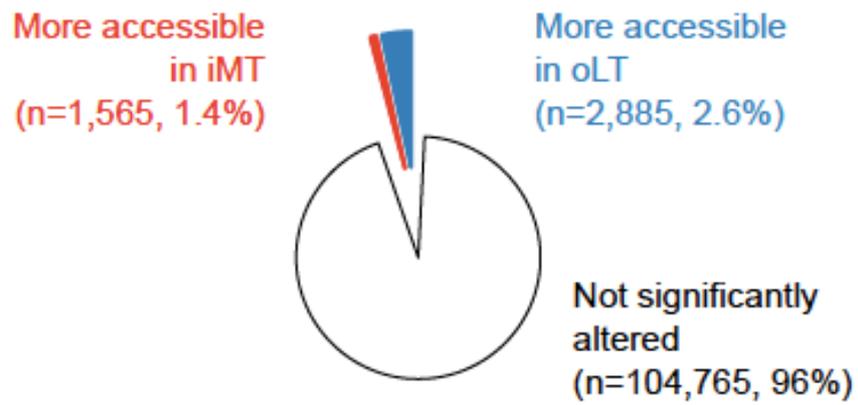


Figure 3-7 Significantly differentially accessible peaks with $FDR \leq 0.05$.

Pie chart shows proportions of the ATAC-seq peaks that are more accessible in iMT (red) and oLT (blue).

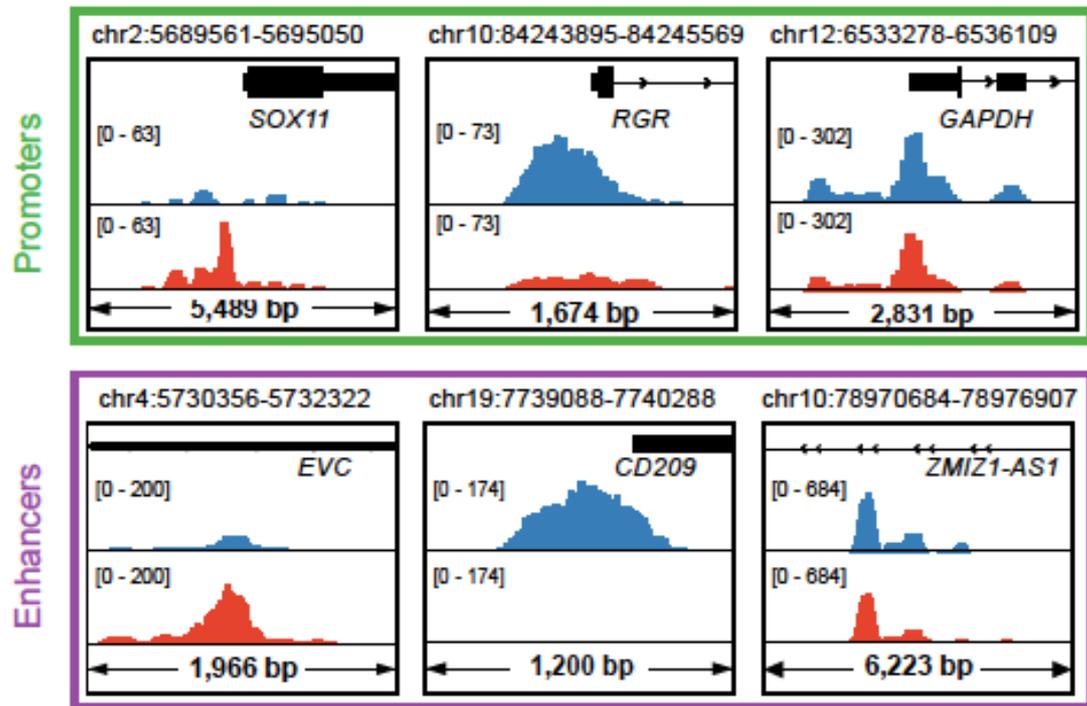


Figure 3-8 Genome browser views of dysregulated promoters and enhancers. Example loci (all patients pooled) showing differential accessible regions at promoters (top) and enhancers (bottom) with more accessible in iMT (left), more accessible in oLT (middle), and not significantly altered between oLT and iMT (right).

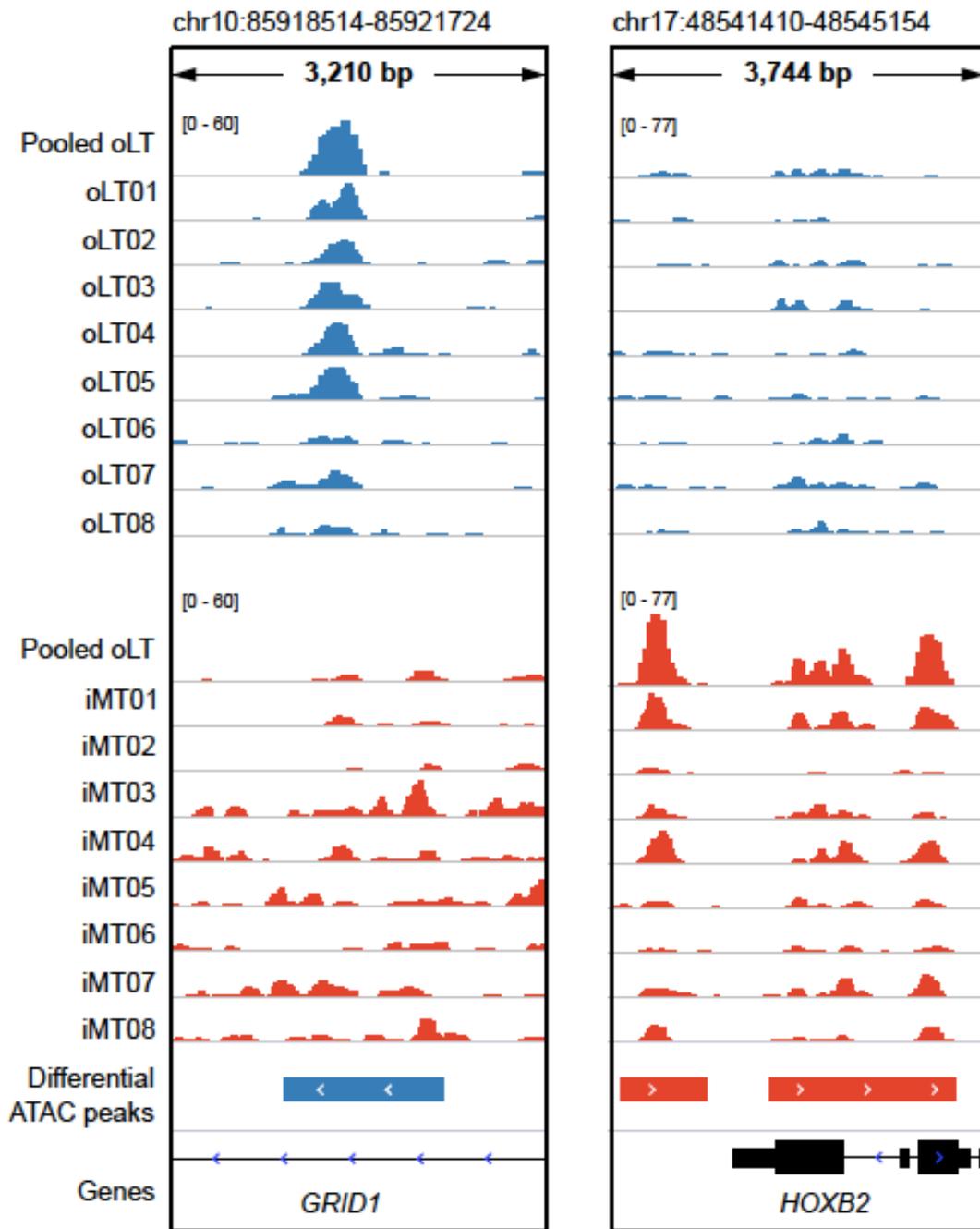


Figure 3-9 Genome browser views of consistency.

Genome browser views showing consistency of ATAC-seq signals across patients at examples loci.

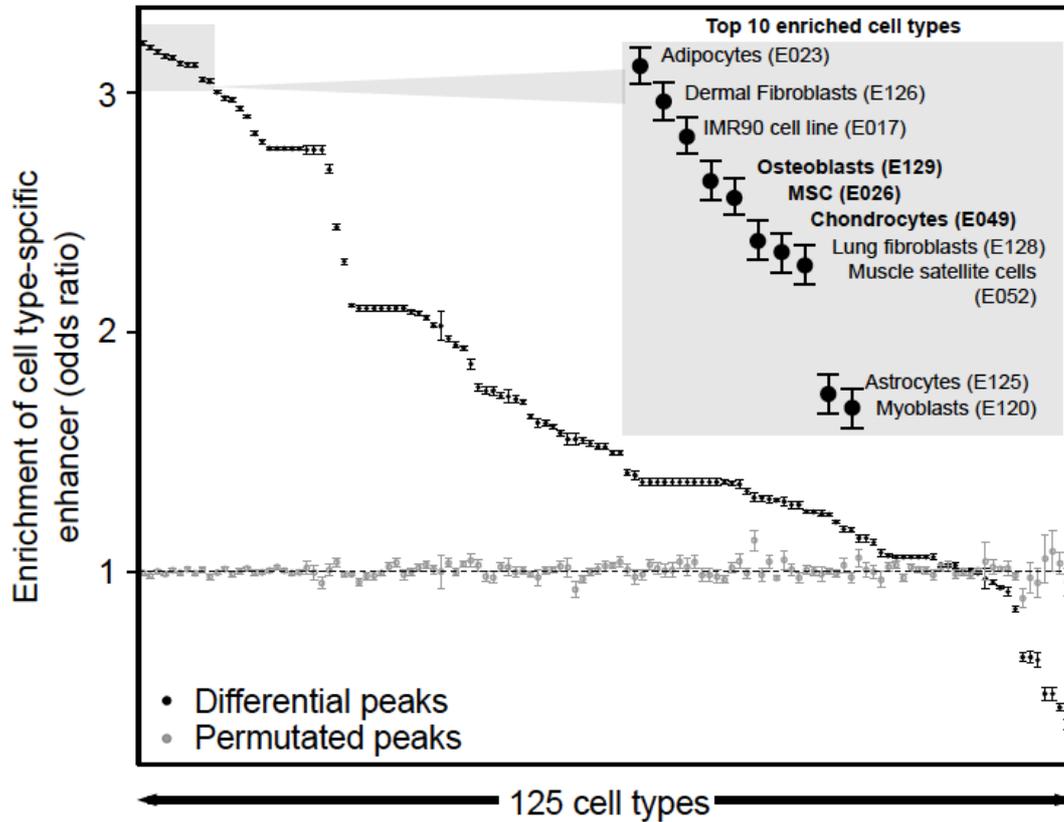


Figure 3-10 Enrichment of cell type-specific enhancers in differentially accessible peaks.

Top 10 are listed (inset), which includes bone- and chondrocyte-related cell types (bold). Error bars represent 95% confidence interval.

Top de novo Motif in oLT
more accessible region

(P value: 1e-22)



CEBPE/MA0837.1/Jaspar(0.936)

Top de novo Motif in iMT
more accessible region

(P value: 1e-83)



Jun-AP1 (bZIP)/ Homer (0.977)

More accessible
region in oLT

More accessible
region in iMT

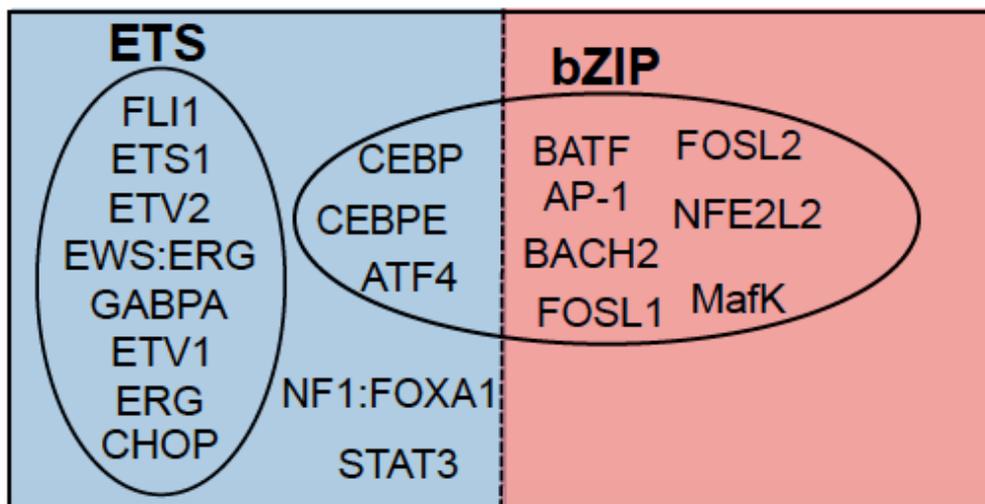


Figure 3-11 DNA motif analysis.

Top de novo motif (top) and top predicted known transcription factors (bottom) enriched in differentially accessible regions.

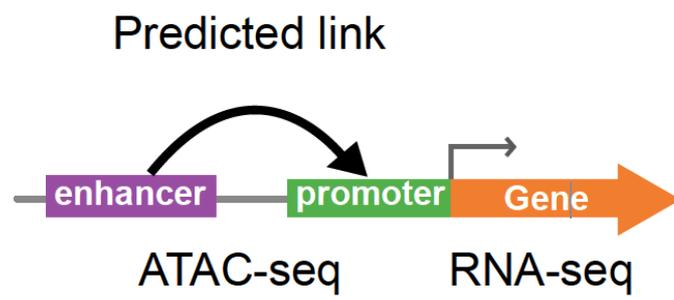


Figure 3-12 A scheme of ATAC-seq and RNA-seq integration analysis.

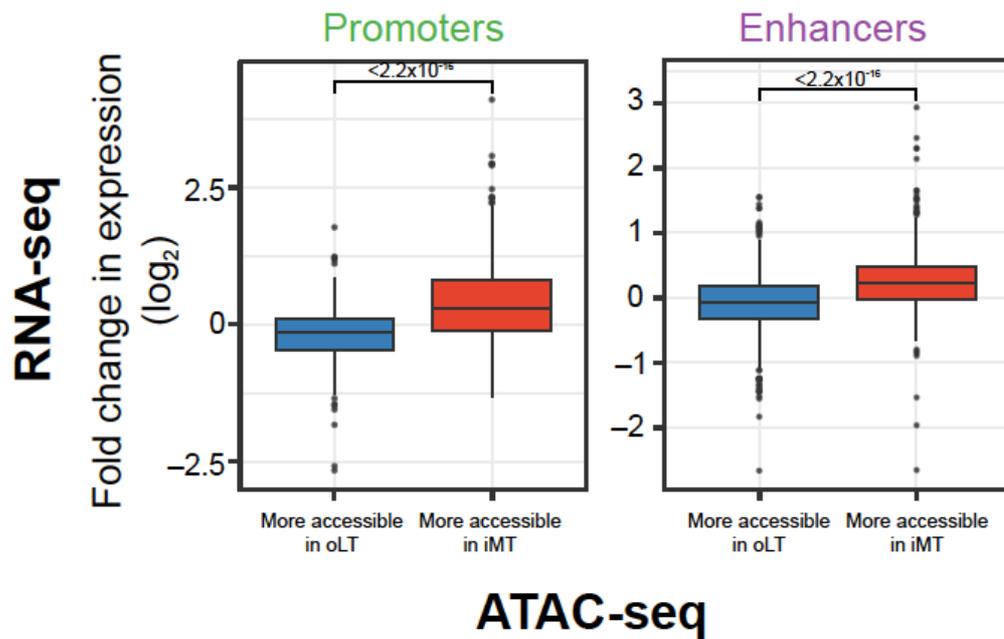


Figure 3-13 Integration of RNA-seq and ATAC-seq.

Differential chromatin accessibility (ATAC-seq) at both promoter (left) and enhancer (right) in OA is generally consistent with differential expression (RNA-seq). Fold change between iMT and oLT is plotted. Box plots show the median, quartiles, and Tukey whiskers.

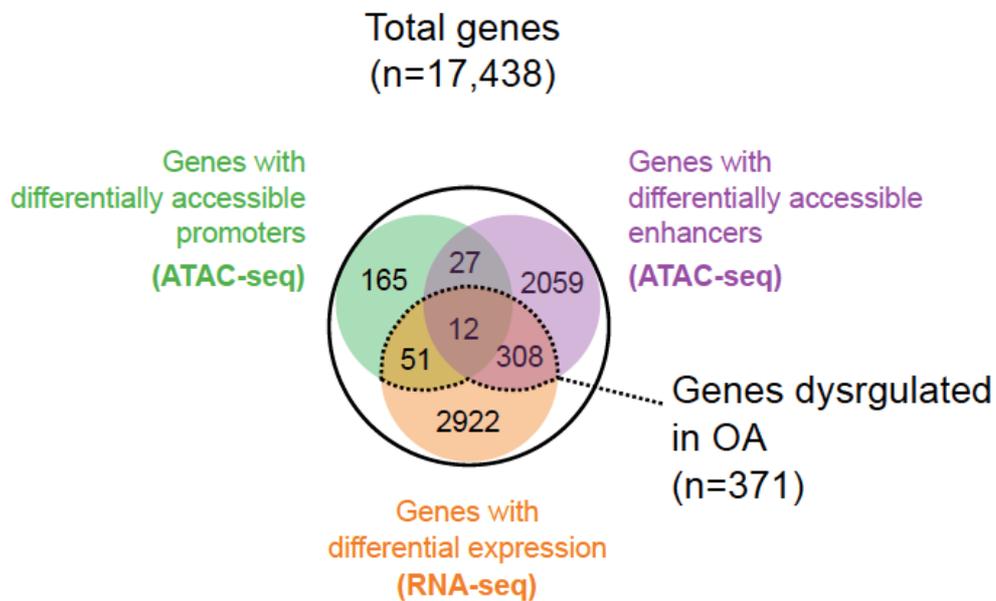


Figure 3-14 Venn diagram.

Venn diagram summarizing protein-coding genes that are dysregulated at both the transcriptomic (RNA-seq) and epigenomic (ATAC-seq) levels in OA with concordant direction of change.

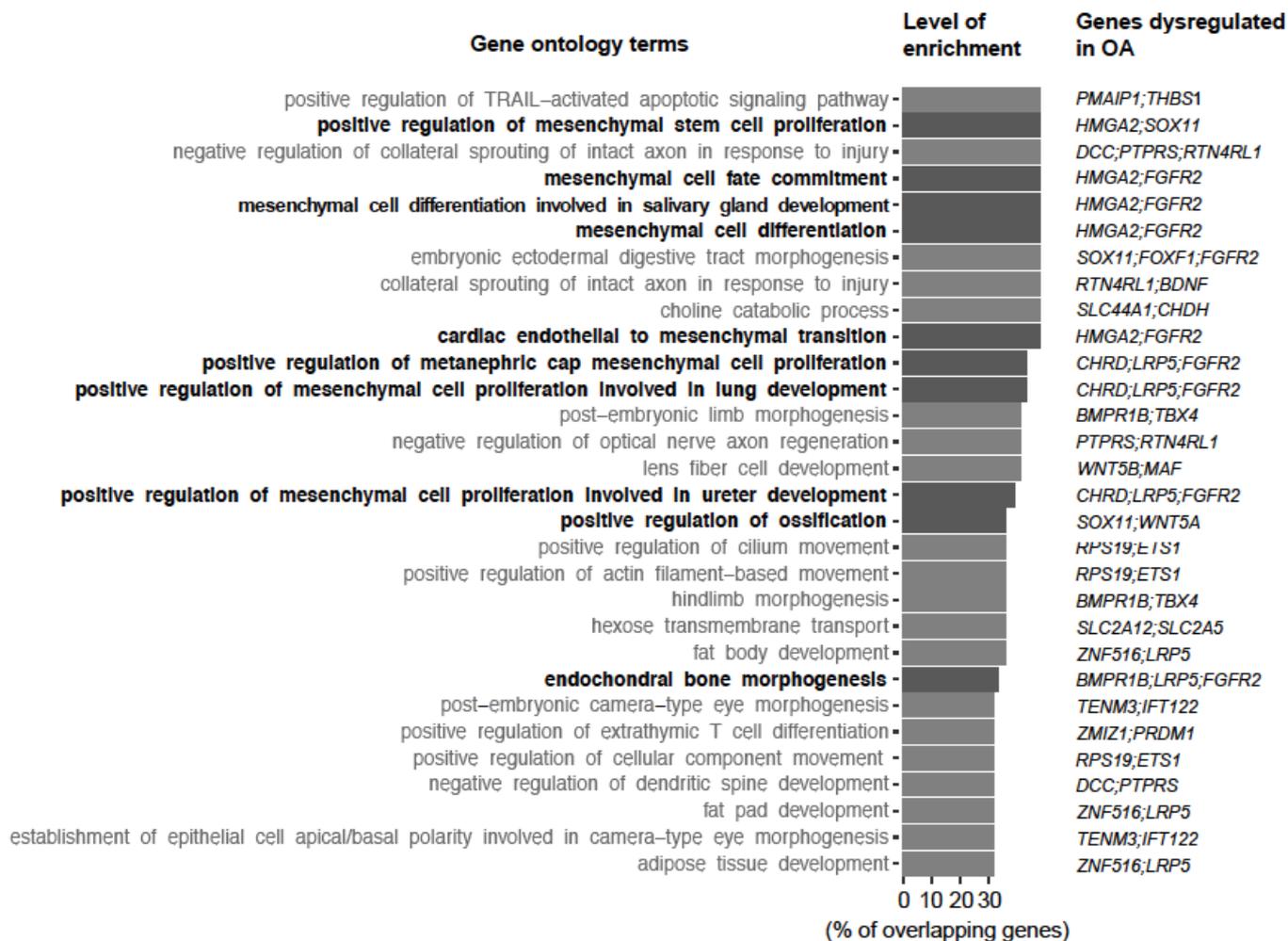


Figure 3-15 GO enrichment.

GO enrichment analysis of 371 dysregulated protein coding genes observed by both ATAC-seq and RNA-seq analysis. Top 30 terms in GO biological process ranked by the level of enrichment and overlapping dysregulated genes are listed. Terms related to MSC, bones, and chondrocytes are highlight in bold.

Chapter 4 Conclusions and future research

4.1 Conclusions

OA is a degenerative joint disease with increasing impact in an aging society and that is the most common causes of Chronic disability in the world. The knee OA is the most common. Normally used method for OA diagnosis is X-ray imaging. Hallmarks of OA include joint space narrowing, cartilage degradation, osteophytes formation and subchondral bone thickening. Risk factors such as aging, obesity, gender and genetic factors have been identified previously, however, the pathogenesis remains incompletely understood. The joint tissue from replacement surgery has been used as a model for studying OA disease progression by comparing the cartilage from preserved joint space region (oLT; containing intact cartilage) and less joint space region (iMT; containing damaged cartilage). The transcriptome assay of this model has been characterized previously. It has been identified that the chondrocytes from the intact cartilage region of OA patient have similar gene expression level with non-OA cartilage, indicating that the intact region cartilage is able to be served as a positive control. Transcriptomic analyses of cartilage provided an opportunity to pinpoint transcriptionally dysregulated genes and pathways relevant to OA. However, such studies have yet to fully reveal the underlying molecular mechanism of how the transcription of these genes are dysregulated. The epigenetic regulations can gain more insight to the disease pathogenesis. Epigenetic modification can regulate gene expression without changing the DNA sequence including DNA methylation, chromatin accessibility, DNA binding proteins, histone modifications and non-coding RNAs. It is known that gene regulatory regions such as promoter and enhancers that play roles in regulation of gene expression. The active promoters and enhancers are chromatin accessible regions which allow the binding proteins (i.e. transcription factors) to access into. ATAC-seq is a method to profile the chromatin accessible regions globally by using Nextera Tn5 transposase. This study was designed to investigate alterations of enhancers or promoters associated with OA by applying ATAC-seq on the knee joint cartilages from OA patients. Fresh human cartilage has been identified to be more suitable than the flash frozen human cartilage to generate high-quality ATAC-seq data using collagenase II.

Multi-omics profiling of tissues can provide insight into the molecular pathogenesis of a disease. While earlier works have been focused on detecting causal genetic variant for disorders, recent technical advances have brought the application of transcriptome and DNA methylome analysis to disease-relevant clinical samples. However, epigenomic analysis at the chromatin level for diseased tissue can pose a larger obstacle due to the requirement of millions of cells. In this study, I took advantage of ATAC-seq that it requires much less material than, for example, DNase-seq or FAIRE-seq, successfully applied on OA articular knee cartilage, and identified genome-wide open chromatin variations potentially associated with OA pathogenesis.

This study was presented as an application of ATAC-seq in OA in a clinically relevant setting. The chromatin accessibility map in cartilage will be a resource for future GWAS and DNA methylation studies in OA and other musculoskeletal diseases. This study identified altered promoters and enhancers of genes that might be involved in the pathogenesis of OA. The analyses suggest aberrant enhancer usage associated with MSC differentiation and chondrogenesis in OA. Understanding these molecular bases of OA is necessary for future therapeutic intervention.

At the individual gene level, this result observed chromatin differences corresponding to expression changes in diseased tissue for known OA related genes. Both of HOXB2 and SOX11 are genes functioning in embryonic-fetal development (Casaca et al., 2016; Wang & Neiva, 2008). It has been reported that SOX11 can induce cartilage growth plate formation in mouse embryos through promoting Wnt signaling (Kato et al., 2015) and contribute joint maintenance through regulating GDF5 (Growth/differentiation factor 5) (Kan et al., 2013). HSD11B1 codes a microsomal enzyme that can lead and catalyze the conversion of stress hormone (active glucocorticoid) cortisol to an inactive metabolite cortisone, in other words, HSD11B1 reduces cortisone to the active hormone cortisol that activates glucocorticoid receptors. Too much stress hormone cortisol can induce central obesity and metabolic syndrome (Abraham et al., 2013). The higher cortisol level is also correlated with higher chronic osteoarthritis pain subscale score (Carlesso et al., 2016; Khoromi et al., 2006).

Moreover, the traditional researches for diseases focus on causal coding genes, noncoding genes and regulatory elements which are proved to have critical implication in recent years (Esteller, 2011). Among them, long non-coding RNAs (lncRNAs) are more likely to be cell-type-specific and proposed to play a regulatory role for gene expression. In many cases lncRNAs have been shown to recruit regulatory complexes

through RNA-protein interaction or nucleosome remodeling (Geisler & Coller, 2013). However, they are usually low abundant and difficult to detect at the RNA level. Here, this study utilized ATAC-seq to probe genome-wide chromatin accessibility variation including those non-coding RNAs.

The novel findings in this work have potential to serve as diagnostic or therapeutic targets. A validation crossed three different platforms, patients and races (European and Japanese) suggest the genes that involved in the endochondral ossification signals pathway should be critical for OA development and progression. Furthermore, the related regulators can be the clinical therapeutic targets of OA progression.

In addition, mesenchymal stem cell transplantation therapy for curing OA is promising but falls stagnant since the way of mesenchymal stem cell mediating healing is unclear and could induce abnormal repair (Gupta et al., 2012; Manieri et al., 2015). Models for OA initiation (Figure 4-1) proposed to explain pathogenesis such as mechanical injury, inflammatory mediators from synovium, defects in metabolism and endochondral ossification (Cox et al., 2013; Dreier, 2010; Kapoor, 2015; Kawaguchi, 2008, 2016; Kuyinu et al., 2016; Man & Mologhianu, 2014). The results from this work are supporting the endochondral theory. In the fetus or child, the bone is formed through cartilage turnover processes (i.e. endochondral ossification) including mesenchymal stem cell to chondrocyte differentiation, hypertrophic chondrocyte differentiation, and osteoblast differentiation. In the adult, cartilage is not capable of regenerating or healing, since the mesenchymal stem cell to chondrocyte differentiation is more likely to be blocked. However, in OA, with the loss of cartilage and disorder in the joint, the mesenchymal stem cell can be forced to differentiate to chondrocyte or osteoblast as a self-healing mechanism, similar to endochondral ossification processes that result of cartilage turnover to the bone. How to force the chondrocyte to regenerate cartilage will be a critical and essential cure for OA patients (Figure 4-2).

Altogether, I believe the findings are of very high interest to the field, as this extensive analysis on ATAC-seq datasets on clinical samples of OA patients could expound the possible pathogenesis mechanism, leading to the identification of potential targets of biomarkers or therapies.

4.2 Future research

The osteoblast-like signal or ossification signal is from a chondrocyte undergoing terminal differentiation, or from MSC direct differentiation cannot be distinguished from bulk tissue analysis, but may be revealed by further experiments such as single-cell transcriptomics or single-cell ATAC-seq (Haque et al., 2017). Single cell RNA-seq and ATAC-seq analysis enable look for candidates signals that are responsible for initiating EO differentiation, ligand/receptor that are specific to the ‘damaged’ tissues. If stopping these signals at the early onset of OA, it may minimize the pain the patient may otherwise experience by stop the abnormal healing. Animal models can be used for candidate gene functional validation.

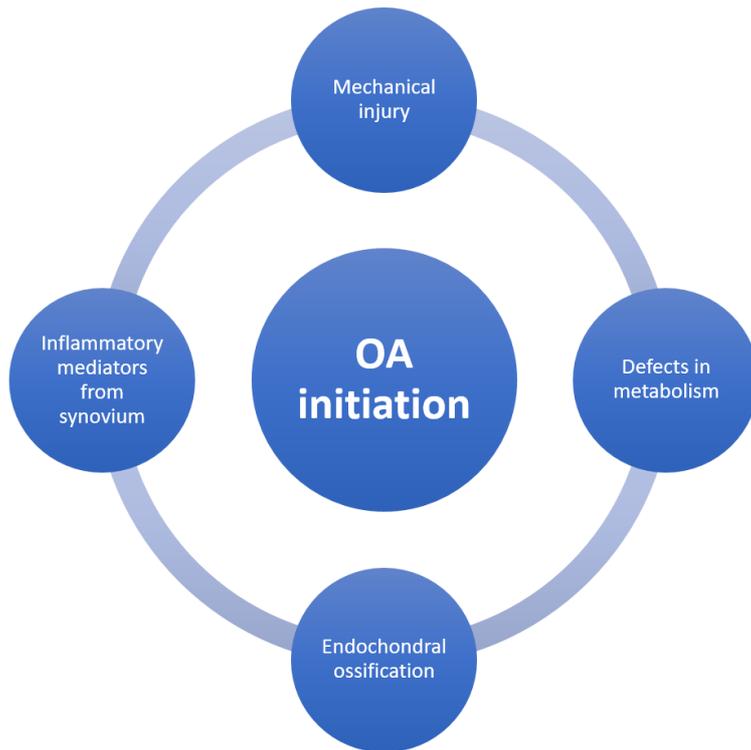


Figure 4-1 Models for OA initiation proposed to explain pathogenesis of OA.

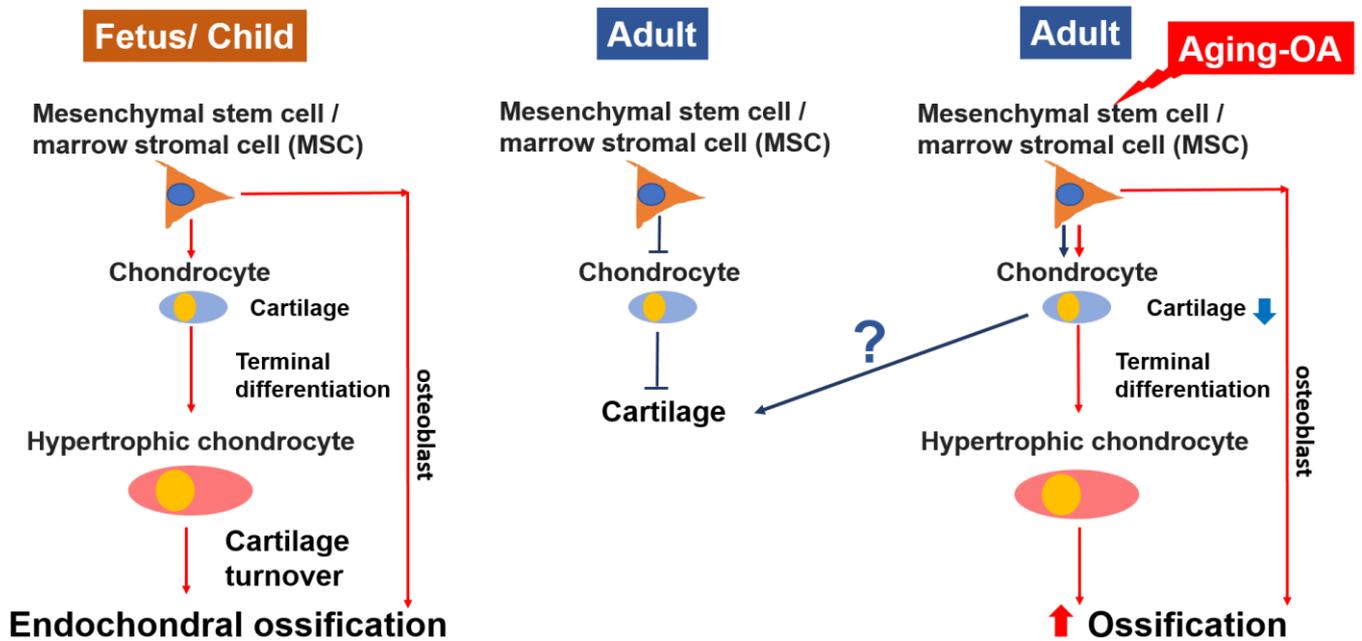


Figure 4-2 Abnormal self-healing theory in OA pathology.

Mesenchymal stem cells are launched in the OA degenerative joint, as a capacitive for self-healing, the positive regulation of osteoblast differentiation in disease inducing the bone formation, which is similar as endochondral ossification, the main skeletal formation process during fetal growing.

References

- Abraham, S. B., Rubino, D., Sinaii, N., Ramsey, S., & Nieman, L. K. (2013). Cortisol, obesity, and the metabolic syndrome: a cross-sectional study of obese subjects and review of the literature. *Obesity (Silver Spring, Md.)*, 21(1), E105-17.
- Adey, A., Morrison, H. G., Asan, Xun, X., Kitzman, J. O., Turner, E. H., Stackhouse, B., MacKenzie, A. P., Caruccio, N. C., Zhang, X., & Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology*, 11(12), R119. <https://doi.org/10.1186/gb-2010-11-12-r119>
- Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., Christy, W., Cooke, TD., Greenwald, R., Wolfe, F. (1986). Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis & Rheumatism*, 29(8), 1039–1049. <https://doi.org/10.1002/art.1780290816>
- Alvarez-Garcia, O., Fisch, K. M., Wineinger, N. E., Akagi, R., Saito, M., Sasho, T., Su, A. I., & Lotz, M. K. (2016). Increased DNA methylation and reduced expression of transcription factors in human osteoarthritis cartilage. *Arthritis & Rheumatology*, 68(8), 1876–1886.
- Bao, X., Rubin, AJ., Qu, K., Zhang, J., Giresi, PG., Chang, HY., Khavari, PA. (2015). A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. *Genome Biology*, 16(1), 284. <https://doi.org/10.1186/s13059-015-0840-9>
- Beasley, J. (2012). Osteoarthritis and rheumatoid arthritis: Conservative therapeutic management. *Journal of Hand Therapy*. <https://doi.org/10.1016/j.jht.2011.11.001>
- Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., & Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57–74. <https://doi.org/10.1038/nature11247>
- Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J., Hirst, M., Lander, E. S., Mikkelsen, T. S., & Thomson, J. A. (2010). The NIH roadmap epigenomics mapping consortium. *Nature Biotechnology*. 28(10), 1045-8.
- Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey,

- T. S., & Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 132(2), 311–322.
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5), 525–527.
- Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 2015, 21.29.1-21.29.9.
- Campisi, J. (1997). The biology of replicative senescence. *European Journal of Cancer Part A*. [https://doi.org/10.1016/S0959-8049\(96\)00058-5](https://doi.org/10.1016/S0959-8049(96)00058-5)
- Cao, Q., Anyansi, C., Hu, X., Xu, L., Xiong, L., Tang, W., Mok, M. T. S., Cheng, C., Fan, X., Gerstein, M., Cheng, A. S. L., & Yip, K. Y. (2017). Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nature Genetics*, 49(10), 1428–1436.
- Carlesso, L. C., Sturgeon, J. A., & Zautra, A. J. (2016). Exploring the relationship between disease-related pain and cortisol levels in women with osteoarthritis. *Osteoarthritis and Cartilage*, 24(12), 2048–2054.
- Casaca, A., Novoa, A., & Mallo, M. (2016). Hoxb6 can interfere with somitogenesis in the posterior embryo through a mechanism independent of its rib-promoting activity. *Development*, 143(3), 437–448.
- Cheng, S. L., Shao, J. S., Charlton-Kachigian, N., Loewy, A. P., & Towler, D. A. (2003). Msx2 promotes osteogenesis and suppresses adipogenic differentiation of multipotent mesenchymal progenitors. *Journal of Biological Chemistry*, 278(46), 45969–45977. <https://doi.org/10.1074/jbc.M306972200>
- Chou, C.-H., Wu, C.-C., Song, I.-W., Chuang, H.-P., Lu, L.-S., Chang, J.-H., Kuo, S.-Y., Lee, C.-H., Wu, J.-Y., Chen, Y.-T., Kraus, V. B., & Lee, M. T. M. (2013). Genome-wide expression profiles of subchondral bone in osteoarthritis. *Arthritis Research & Therapy*, 15(6), R190.
- Chou, C. H., Lee, C. H., Lu, L. S., Song, I. W., Chuang, H. P., Kuo, S. Y., Wu, J. Y., Chen, Y. T., Kraus, V. B., Wu, C. C., & Lee, M. T. M. (2013). Direct assessment of articular cartilage and underlying subchondral bone reveals a progressive gene expression change in human osteoarthritic knees. *Osteoarthritis and Cartilage*, 21(3), 450–461.
- Chou, C. H., Lee, M. T. M., Song, I. W., Lu, L. S., Shen, H. C., Lee, C. H., Wu, J. Y., Chen, Y. T., Kraus, V. B., & Wu, C. C. (2015). Insights into osteoarthritis

- progression revealed by analyses of both knee tibiofemoral compartments. *Osteoarthritis and Cartilage*, 23(4), 571–580.
- Chubinskaya, S., Kumar, B., Merrihew, C., Heretis, K., Rueger, D. C., & Kuettner, K. E. (2002). Age-related changes in cartilage endogenous osteogenic protein-1 (OP-1). *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1588(2), 126–134.
- Collado, M., Blasco, M. A., & Serrano, M. (2007). Cellular Senescence in Cancer and Aging. *Cell*, 130(2), 223–33.
- Consortium, G. Te. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, 348(6235), 648–660.
- Corradin, O., & Scacheri, P. C. (2014). Enhancer variants: Evaluating functions in common disease. *Genome Medicine*, 6(10), 85. <https://doi.org/10.1186/s13073-014-0085-3>
- Cox, L. G. E., Van Donkelaar, C. C., van Rietbergen, B., Emans, P. J., & Ito, K. (2013). Alterations to the subchondral bone architecture during osteoarthritis: Bone adaptation vs endochondral bone formation. *Osteoarthritis and Cartilage*, 21(2), 331–338.
- Dai, S. M., Shan, Z. Z., Nakamura, H., Masuko-Hongo, K., Kato, T., Nishioka, K., & Yudoh, K. (2006). Catabolic stress induces features of chondrocyte senescence through overexpression of caveolin 1: Possible involvement of caveolin 1-induced down-regulation of articular chondrocytes in the pathogenesis of osteoarthritis. *Arthritis and Rheumatism*, 54(3), 818–831.
- DeGroot, J., Verzijl, N., Wenting-Van Wijk, M. J. G., Jacobs, K. M. G., Van El, B., Van Roermund, P. M., Bank, R. A., Bijlsma, J. W. J., TeKoppele, J. M., & Lafeber, F. P. J. G. (2004). Accumulation of advanced glycation end products as a molecular mechanism for aging as a risk factor in osteoarthritis. *Arthritis and Rheumatism*, 50(4), 1207–1215.
- den Hollander, W., & Meulenbelt, I. (2015). DNA Methylation in Osteoarthritis. *Current Genomics*, 16(6), 419–426. <https://doi.org/10.2174/1389202916666150817212711>
- den Hollander, W., Ramos, Y. F. M., Bomer, N., Elzinga, S., Van Der Breggen, R., Lakenberg, N., De Dijcker, W. J., Suchiman, H. E. D., Duijnsveld, B. J., Houwing-Duistermaat, J. J., Slagboom, P. E., Bos, S. D., Nelissen, R. G. H. H., &

- Meulenbelt, I. (2015). Transcriptional associations of osteoarthritis-mediated loss of epigenetic control in articular cartilage. *Arthritis and Rheumatology*.
- den Hollander, W., Ramos, Y. F. M., Bos, S. D., Bomer, N., van Der Breggen, R., Lakenberg, N., De Dijcker, W. J., Duijnisveld, B. J., Slagboom, P. E., Nelissen, R. G. H. H., & Meulenbelt, I. (2014). Knee and hip articular cartilage have distinct epigenomic landscapes: Implications for future cartilage regeneration approaches. *Annals of the Rheumatic Diseases*, 73(12), 2208-2212.
- Dickinson, S., & Hollander, A. (2017). The wnt5a receptor, receptor tyrosine kinase-like orphan receptor 2, is a predictive cell surface marker of human mesenchymal stem cells with an enhanced capacity for chondrogenic differentiation. *Stem Cells*, 26(9), 2399–2407.
- Dsteche, CM., Berietch, JB. (2015). X-chromosome inactivation and escape. *Journal of Genetics*, 94(4), 591–599.
- Dreier, R. (2010). Hypertrophic differentiation of chondrocytes in osteoarthritis: the developmental aspect of degenerative joint disorders. *Arthritis Research & Therapy*, 12(5), 216.
- Dunn, S. L., Soul, J., Anand, S., Schwartz, J. M., Boot-Handford, R. P., & Hardingham, T. E. (2016). Gene expression changes in damaged osteoarthritic cartilage identify a signature of non-chondrogenic and mechanical responses. *Osteoarthritis and Cartilage*, 24(8), 1431–1440.
- Ellman, M. B., Yan, D., Ahmadinia, K., Chen, D., An, H. S., & Im, H. J. (2013). Fibroblast growth factor control of cartilage homeostasis. *Journal of Cellular Biochemistry*. 114(4), 735-42. <https://doi.org/10.1002/jcb.24418>
- Esteller, M. (2011). Non-coding RNAs in human disease. *Nature Reviews Genetics*, 12(12), 861–874.
- Evangelou, E., Kerkhof, HJ., Styrkarsdottir, U., Ntzani, E.E., Bos, S.D., Esko, T., Evans, D.S., Metrustry, S., Panoutsopoulou, K., Ramos, Y.F., Thorleifsson, G., Tsilidis, K.K.; arcOGEN Consortium, Arden, N., Aslam, N., Bellamy, N., Birrell, F., Blanco, F.J., Carr, A., Chapman, K., Day-Williams, A.G., Deloukas, P., Doherty, M., Engström, G., Helgadottir, H.T., Hofman, A., Ingvarsson, T., Jonsson, H., Keis, A., Keurentjes, J.C., Kloppenburg, M., Lind, P.A., McCaskie, A., Martin, N.G., Milani, L., Montgomery, G.W., Nelissen, R.G., Nevitt, M.C., Nilsson, P.M., Ollier, W.E., Parimi, N., Rai, A., Ralston, S.H., Reed, M.R., Riancho, J.A., Rivadeneira, F., Rodriguez-Fontenla, C., Southam, L., Thorsteinsdottir, U.,

- Tsezou, A., Wallis, G.A., Wilkinson, J.M., Gonzalez, A., Lane, N.E., Lohmander, L.S., Loughlin, J., Metspalu, A., Uitterlinden, A.G., Jonsdottir, I., Stefansson, K., Slagboom, P.E., Zeggini, E., Meulenbelt, I., Ioannidis, J.P., Spector, T.D., van Meurs, J.B., & Valdes, A.M. (2014). A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Annals of the Rheumatic Diseases*, 73(12), 2130-2136. <https://doi.org/10.1136/annrheumdis-2012-203114>
- Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shores, N., Whitton, H., Ryan, R.J., Shishkin, A.A., Hatan, M., Carrasco-Alfonso, M.J., Mayer, D., Luckey, C.J., Patsopoulos, N.A., De Jager, P.L., Kuchroo, V.K., Epstein, C.B., Daly, M.J., Hafler, D.A., & Bernstein, B.E. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539), 337–343.
- Felson, D. T. (1990). The epidemiology of knee osteoarthritis: Results from the framingham osteoarthritis study. *Seminars in Arthritis and Rheumatism*, 20(3 SUPPL. 1), 42–50. [https://doi.org/10.1016/0049-0172\(90\)90046-I](https://doi.org/10.1016/0049-0172(90)90046-I)
- Fernandez-Moreno, M., Rego, I., Carreira-Garcia, V., & Blanco, F. (2008). Genetics in Osteoarthritis. *Current Genomics*, 9(8), 542–547. <https://doi.org/10.2174/138920208786847953>
- Fernández-Santiago, R., Carballo-Carbajal, I., Castellano, G., Torrent, R., Richaud, Y., Sánchez-Danés, A., Vilarrasa-Blasi, R., Sánchez-Pla, A., Mosquera, J.L., Soriano, J., López-Barneo, J., Canals, J.M., Alberch, J., Raya, Á., Vila, M., Consiglio, A., Martín-Subero, J.I., Ezquerra, M., & Tolosa, E. (2015). Aberrant epigenome in iPSC-derived dopaminergic neurons from Parkinson’s disease patients. *EMBO Molecular Medicine*, 7(12), 1529–1546. <https://doi.org/10.15252/emmm.201505439>
- Fernández-Tajes, J., Soto-Hermida, A., Vázquez-Mosquera, M. E., Cortés-Pereira, E., Mosquera, A., Fernández-Moreno, M., Oreiro, N., Fernández-López, C., Fernández, J. L., Rego-Pérez, I., & Blanco, F. J. (2014). Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients. *Annals of the Rheumatic Diseases*, 73(4), 668-677.
- Findlay, D. M., & Kuliwaba, J. S. (2016). Bone–cartilage crosstalk: a conversation for understanding osteoarthritis. *Bone Research*, 4, 16028.
- Forrest, A. R. R., Kawaji, H., Rehli, M., Baillie, J. K., De Hoon, M. J. L., Haberle, V.,

- FANTOM Consortium, Hayashizaki, Y. (2014). A promoter-level mammalian expression atlas. *Nature*, 507(7493), 462–470. <https://doi.org/10.1038/nature13182>
- Forsyth, C. B., Cole, A., Murphy, G., Bienias, J. L., Im, H.-J., & Loeser, R. F. (2005). Increased matrix metalloproteinase-13 production with aging by human articular chondrocytes in response to catabolic stimuli. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 60(9), 1118–1124.
- Geisler, S., & Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature Reviews. Molecular Cell Biology*, 14(11), 699–712.
- Gensler, H. L., & Bernstein, H. (1981). DNA damage as the primary cause of aging. *The Quarterly Review of Biology*, 56(3), 279-303. <https://doi.org/10.1086/412317>
- Goldring, M. B., & Goldring, S. R. (2010). Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis. In *Annals of the New York Academy of Sciences* (Vol. 1192, pp. 230–237).
- Grynpas, M. D., Alpert, B., Katz, I., Lieberman, I., & Pritzker, K. P. H. (1991). Subchondral bone in osteoarthritis. *Calcified Tissue International*, 49(1), 20–26.
- Guerne, P. A, Blanco, F., Kaelin, A., Desgeorges, A., & Lotz, M. (1995). Growth factor responsiveness of human articular chondrocytes in aging and development. *Arthritis and Rheumatism*, 38(7), 960–968.
- Gupta, P. K., Das, A. K., Chullikana, A., & Majumdar, A. S. (2012). Mesenchymal stem cells for cartilage repair in osteoarthritis. *Stem Cell Research & Therapy*, 3(4), 25.
- Haluskova, J. (2010). Epigenetic studies in human diseases. *Folia Biol (Praha)*, 56(3):83-96.
- Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine*, 9(1), 75.
- Harman, D. (1956). Aging: a theory based on free radical and radiation chemistry. *Journal of Gerontology*, 11(3), 298–300.
- Hayashi, S., Fujishiro, T., Hashimoto, S., Kanzaki, N., Chinzei, N., Kihara, S., Takayama, K., Matsumoto, T., Nishida, K., Kurosaka, M., & Kuroda, R. (2015). p21 deficiency is susceptible to osteoarthritis through STAT3 phosphorylation. *Arthritis Research & Therapy*, 17, 314.

- He, X., Ohba, S., Hojo, H., & McMahon, A. P. (2016). AP-1 family members act with Sox9 to promote chondrocyte hypertrophy. *Development (Cambridge, England)*, 1, dev.134502.
- Heinz, S., Romanoski, C. E., Benner, C., & Glass, C. K. (2015). The selection and function of cell type-specific enhancers. *Nature Reviews Molecular Cell Biology*, 16(3), 144–154. <https://doi.org/10.1038/nrm3949>
- Hematti, P. (2012). Mesenchymal stromal cells and fibroblasts: a case of mistaken identity? *Cytotherapy*. 14(5), 516-521.
- Herz, H. M. (2016). Enhancer deregulation in cancer and other diseases. *BioEssays*.
- Hoffman, L. M., Garcha, K., Karamboulas, K., Cowan, M. F., Drysdale, L. M., Horton, W. A., & Underhill, T. M. (2006). BMP action in skeletogenesis involves attenuation of retinoid signaling. *Journal of Cell Biology*, 174(1), 101–113.
- Hollander, A. P., Pidoux, I., Reiner, A., Rorabeck, C., Bourne, R., & Poole, A. R. (1995). Damage to type II collagen in aging and osteoarthritis starts at the articular surface, originates around chondrocytes, and extends into the cartilage with progressive degeneration. *Journal of Clinical Investigation*, 96(6), 2859–2869.
- Hon, C., Ramilowski, J., Harshbarger, J., Bertin, N., Rackham, O., Gough, J., FANTOM Consortium., Forrest, A. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, 543(7644), 199–204.
- Hong, E., Di Cesare, P. E., & Haudenschild, D. R. (2011). Role of c-Maf in chondrocyte differentiation. *CARTILAGE*, 2(1), 27–35. <https://doi.org/10.1177/1947603510377464>
- Huang, G., Chubinskaya, S., Liao, W., & Loeser, R. F. (2017). Wnt5a induces catabolic signaling and matrix metalloproteinase production in human articular chondrocytes. *Osteoarthritis and Cartilage*, 25(9), 1505–1515. <https://doi.org/10.1016/j.joca.2017.05.018>
- Imhof, H., Breitenseher, M., Kainberger, F., Rand, T., & Trattnig, S. (1999). Importance of subchondral bone to articular cartilage in health and disease. *Top Magn Reson Imaging*. 10(3), 180-92.
- Itahana, K., Campisi, J., & Dimri, G. P. (2004). Mechanisms of cellular senescence in human and mouse cells. *Biogerontology*, 5(1), 1–10.
- Iwamoto, M., Higuchi, Y., Enomoto-Iwamoto, M., Kurisu, K., Koyama, E., Yeh, H., Rosenbloom, J., & Pacifici, M. (2001). The role of ERG (ets related gene) in cartilage development. *Osteoarthritis and Cartilage*, 9, S41–S47.

<https://doi.org/10.1053/JOCA.2001.0443>

- Iwamoto, M., Tamamura, Y., Koyama, E., Komori, T., Takeshita, N., Williams, J. A., Nakamura, T., Enomoto-Iwamoto, M., & Pacifici, M. (2007). Transcription factor ERG and joint and articular cartilage formation during mouse limb and spine skeletogenesis. *Developmental Biology*, 305(1), 40–51.
- Iwasa, K., Hayashi, S., Fujishiro, T., Kanzaki, N., Hashimoto, S., Sakata, S., Chinzei, N., Nishiyama, T., Kuroda, R., & Kurosaka, M. (2014). PTEN regulates matrix synthesis in adult human chondrocytes under oxidative stress. *Journal of Orthopaedic Research*, 32(2), 231–237.
- Jacobson, J. A., Girish, G., Jiang, Y., & Sabb, B. J. (2008). Radiographic Evaluation of Arthritis: Degenerative Joint Disease and Variations. *Radiology*, 248(3), 737–747.
- Jeffries, M. A., Donica, M., Baker, L. W., Stevenson, M. E., Annan, A. C., Humphrey, M. B., James, J. A., & Sawalha, A. H. (2014). Genome-Wide DNA methylation study identifies significant epigenomic changes in osteoarthritic cartilage. *Arthritis and Rheumatology*, 66(10), 2804–2815.
- Jiang, H., Ju, Z., & Rudolph, K. L. (2007). Telomere shortening and ageing. *Zeitschrift Fur Gerontologie Und Geriatrie*. 40(5):314-324. <https://doi.org/10.1007/s00391-007-0480-0>
- Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'Donnell, C. J., & De Bakker, P. I. W. (2008). SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, 24(24), 2938–2939.
- Johnson, V. L., & Hunter, D. J. (2014). The epidemiology of osteoarthritis. *Best Practice and Research: Clinical Rheumatology*. 28(1), 5-15. <https://doi.org/10.1016/j.berh.2014.01.004>
- Kan, A., Ikeda, T., Fukai, A., Nakagawa, T., Nakamura, K., Chung, U., Kawaguchi, H., & Tabin, C. J. (2013). SOX11 contributes to the regulation of GDF5 in joint maintenance. *BMC Developmental Biology*, 13(1), 4.
- Kapoor, M. (2015). Pathogenesis of Osteoarthritis. In *Osteoarthritis* (pp. 1–28). Cham: Springer International Publishing.
- Karuppaiah, K., Yu, K., Lim, J., Chen, J., Smith, C., Long, F., & Ornitz, D. M. (2016). FGF signaling in the osteoprogenitor lineage non-autonomously regulates postnatal chondrocyte proliferation and skeletal growth. *Development*, 143(10), 1811–1822. <https://doi.org/10.1242/dev.131722>
- Kato, K., Bhattaram, P., Penzo-Mendez, A., Gadi, A., & Lefebvre, V. (2015). SOXC

- transcription factors induce cartilage growth plate formation in mouse embryos by promoting noncanonical WNT signaling. *Journal of Bone and Mineral Research : The Official Journal of the American Society for Bone and Mineral Research*, 30(9), 1560–1571.
- Kawaguchi, H. (2008). Endochondral ossification signals in cartilage degradation during osteoarthritis progression in experimental mouse models. *Molecules and Cells*, 25(1), 1–6.
- Kawaguchi, H. (2016). The canonical Wnt signal paradoxically regulates osteoarthritis development through the endochondral ossification process. *Integrative Molecular Medicine*, 3(3), 672–674.
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, 12(6), 996–1006. <https://doi.org/10.1101/gr.229102>
- Khoromi, S., Muniyappa, R., Nackers, L., Gray, N., Baldwin, H., Wong, K. A., Matheny, L. A., Moquin, B., Rainer, A., Hill, S., Remaley, A., Johnson, L. L., Max, M. B., & Blackman, M. R. (2006). Effects of chronic osteoarthritis pain on neuroendocrine function in men. *The Journal of Clinical Endocrinology & Metabolism*, 91(11), 4313–4318.
- Krane, S. M., Quigley, J., & Krane, S. (2001). Petulant cellular acts: destroying the ECM rather than creating it. *The Journal of Clinical Investigation*, 107(1), 31–32.
- Kron, K. J., Bailey, S. D., & Lupien, M. (2014). Enhancer alterations in cancer: a source for a cell identity crisis. *Genome Med*, 6(9), 77.
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1), W90–W97.
- Kuyinu, E. L., Narayanan, G., Nair, L. S., & Laurencin, C. T. (2016). Animal models of osteoarthritis: Classification, update, and measurement of outcomes. 1, 19. *Journal of Orthopaedic Surgery and Research*.
- Kwoh, C. K. (2012). Epidemiology of osteoarthritis. In *The Epidemiology of Aging* (pp. 523–536).
- Lee, J., Christoforo, G., Christoforo, G., Foo, C., Probert, C., Kundaje, A., Boley, N., Kohpangwei, Kim, D., & Dacre, M. (2016). Kundajelab/Atac_Dnase_Pipelines:

0.3.0.

- Lee, R., & Kean, W. F. (2012). Obesity and knee osteoarthritis. *Inflammopharmacology*, 20(2), 53-8. <https://doi.org/10.1007/s10787-011-0118-0>
- Li, G., Yin, J., Gao, J., Cheng, T. S., Pavlos, N. J., Zhang, C., & Zheng, M. H. (2013). Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes. *Arthritis Research & Therapy*, 15(6), 223. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24321104>
- Li, J., & Dong, S. (2016). The signaling pathways involved in chondrocyte differentiation and hypertrophic differentiation. 2016, 2470351. *Stem Cells International*.
- Li, M. J., Wang, P., Liu, X., Lim, E. L., Wang, Z., Yeager, M., Wong, M. P., Sham, P. C., Chanock, S. J., & Wang, J. (2016). GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Research*, 44(D1), 869–876.
- Loeser, R. F. (2009). Aging and osteoarthritis: the role of chondrocyte senescence and aging changes in the cartilage matrix. *Osteoarthritis and Cartilage*, 17(8), 971–979.
- Long, D., Blake, S., Song, X.-Y., Lark, M., & Loeser, R. F. (2008). Human articular chondrocytes produce IL-7 and respond to IL-7 with increased production of matrix metalloproteinase-13. *Arthritis Research & Therapy*, 10(1), R23.
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., GTEx Consortium, & Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*. 45(6):580-585.
- MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., MayPendlington, Z., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., & Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, 45(D1), D896–D901.
- Madry, H., van Dijk, C. N., & Mueller-Gerbl, M. (2010). The basic science of the subchondral bone. *Knee Surgery, Sports Traumatology, Arthroscopy*, 18(4), 419–433. <https://doi.org/10.1007/s00167-010-1054-z>
- Man, G. S., & Mologhianu, G. (2014). Osteoarthritis pathogenesis - a complex process that involves the entire joint. *Journal of Medicine and Life*, 7(1), 37–41.
- Manieri, N. A., Mack, M. R., Himmelrich, M. D., Worthley, D. L., Hanson, E. M., Eckmann, L., Wang, T. C., & Stappenbeck, T. S. (2015). Mucosally transplanted

- mesenchymal stem cells stimulate intestinal healing by promoting angiogenesis. *Journal of Clinical Investigation*, 125(9), 3606–3618.
- Margueron, R., & Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. *Nature Reviews Genetics*, 11(4), 285-96. <https://doi.org/10.1038/nrg2752>
- Martel-pelletier, J., Barr, A. J., Cicuttini, F. M., Conaghan, P. G., Cooper, C., Goldring, M. B., Goldring, S. R., Jones, G., & Pelletier, J. (2016). Osteoarthritis. *Nature Reviews Disease Primers*, 2, 16072.
- Martin, J. A., & Buckwalter, J. A. (2001). Telomere erosion and senescence in human articular cartilage chondrocytes. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 56(4), B172-B179. <https://doi.org/10.1093/gerona/56.4.B172>
- Martin, J. a, Klingelhutz, A. J., Moussavi-Harami, F., & Buckwalter, J. a. (2004). Effects of oxidative damage and telomerase activity on human articular cartilage chondrocyte senescence. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 59(4), 324–337.
- Matsubara, T., Kida, K., Yamaguchi, A., Hata, K., Ichida, F., Meguro, H., Aburatani, H., Nishimura, R., & Yoneda, T. (2008). BMP2 regulates osterix through Msx2 and Runx2 during osteoblast differentiation. *Journal of Biological Chemistry*, 283(43), 29119–29125. <https://doi.org/10.1074/jbc.M801774200>
- McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10), 4288–4297.
- Milani, P., Escalante-Chong, R., Shelley, B. C., Patel-Murray, N. L., Xin, X., Adam, M., Mandefro, B., Sareen, D., Svendsen, C. N., & Fraenkel, E. (2016). Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells. *Scientific Reports*, 6, 25474.
- Milz, S., & Putz, R. (1994). Quantitative morphology of the subchondral plate of the tibial plateau. *Journal of Anatomy*, 185(Pt 1), 103–110.
- Mouritzen, U., Christgau, S., Lehmann, H.-J., Tankó, L. B., & Christiansen, C. (2003). Cartilage turnover assessed with a newly developed assay measuring collagen type II degradation products: influence of age, sex, menopause, hormone replacement therapy, and body mass index. *Annals of the Rheumatic Diseases*, 62(February 2005), 332–336.

- Muller, M. (2009). Cellular senescence: molecular mechanisms, in vivo significance, and redox considerations. *Antioxidants & Redox Signaling*, 11(1), 59–98.
- Nedić, O., Rattan, S. I. S., Grune, T., & Trougakos, I. P. (2013). Molecular effects of advanced glycation end products on cell signalling pathways, ageing and pathophysiology. *Free Radical Research*, 47 Suppl 1(August), 28–38.
- Neogi, T. (2013). The epidemiology and impact of pain in osteoarthritis. *Osteoarthritis and Cartilage*, 21(9), 1145–1153.
- Okuma, T., Hirata, M., Yano, F., Mori, D., Kawaguchi, H., Chung, U. I., Tanaka, S., & Saito, T. (2015). Regulation of mouse chondrocyte differentiation by CCAAT/enhancer-binding proteins. *Biomed Res*, 36(1), 21–29. <https://doi.org/10.2220/biomedres.36.21>
- Pan, J., Zhou, X., Li, W., Novotny, J. E., Doty, S. B., & Wang, L. (2009). In situ measurement of transport between subchondral bone and articular cartilage. *Journal of Orthopaedic Research*, 27(10), 1347–1352.
- Panoutsopoulou, K., Southam, L., Elliott, K.S., Wrayner, N., Zhai, G., Beazley, C., Thorleifsson, G., Arden, N.K., Carr, A., Chapman, K., Deloukas, P., Doherty, M., McCaskie, A., Ollier, W.E., Ralston, S.H., Spector, T.D., Valdes, A.M., Wallis, G.A., Wilkinson, J.M., Arden, E., Battley, K., Blackburn, H., Blanco, F.J., Bumpstead, S., Cupples, L.A., Day-Williams, A.G., Dixon, K., Doherty, S.A., Esko, T., Evangelou, E., Felson, D., Gomez-Reino, J.J., Gonzalez, A., Gordon, A., Gwilliam, R., Halldorsson, B.V., Hauksson, V.B., Hofman, A., Hunt, S.E., Ioannidis, J.P., Ingvarsson, T., Jonsdottir, I., Jonsson, H., Keen, R., Kerkhof, H.J., Kloppenburg, M.G., Koller, N., Lakenberg, N., Lane, N.E., Lee, A.T., Metspalu, A., Meulenbelt, I., Nevitt, M.C., O'Neill, F., Parimi, N., Potter, S.C., Rego-Perez, I., Riancho, J.A., Sherburn, K., Slagboom, P.E., Stefansson, K., Styrkarsdottir, U., Sumillera, M., Swift, D., Thorsteinsdottir, U., Tsezou, A., Uitterlinden, A.G., van Meurs, J.B., Watkins, B., Wheeler, M., Mitchell, S., Zhu, Y., Zmuda, J.M.; arcOGEN Consortium, Zeggini, E., & Loughlin, J. (2011). Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Annals of the Rheumatic Diseases*. 70(5), 864-867
- Papathanasiou, I., Malizos, K. N., & Tsezou, A. (2010). Low-density lipoprotein receptor-related protein 5 (LRP5) expression in human osteoarthritic chondrocytes. *Journal of Orthopaedic Research*, 28(3), 348–353. <https://doi.org/10.1002/jor.20993>

- Pas, H. I., Winters, M., Haisma, H. J., Koenis, M. J., Tol, J. L., & Moen, M. H. (2017). Stem cell injections in knee osteoarthritis: a systematic review of the literature. *British Journal of Sports Medicine*, 51(15):1125-1133. <https://doi.org/10.1136/bjsports-2016-096793>
- Portela, A., & Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology*, 28(10), 1057–1068. Retrieved from <http://www.nature.com/doifinder/10.1038/nbt.1685>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842.
- Radin, E. L., & Rose, R. M. (1986). Role of subchondral bone in the initiation and progression of cartilage damage. *Clinical Orthopaedics and Related Research*, (213), 34-40. <https://doi.org/10.1097/00003086-198612000-00005>
- Reynard, L. N. (2017). Analysis of genetics and DNA methylation in osteoarthritis: What have we learnt about the disease? *Seminars in Cell and Developmental Biology*, 62:57-66.
- Ringdahl, E., & Pandit, S. (2011). Treatment of knee osteoarthritis. *American Family Physician*, 83(11), 1287–1292.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140.
- Rosen, F., McCabe, G., Quach, J., Solan, J., Terkeltaub, R., Seegmiller, J. E., & Lotz, M. (1997). Differential effects of aging on human chondrocyte responses to transforming growth factor beta: increased pyrophosphate production and decreased cell proliferation. *Arthritis Rheum*, 40(7), 1275–1281.
- Rushton, M. D., Reynard, L. N., Barter, M. J., Refaie, R., Rankin, K. S., Young, D. A., & Loughlin, J. (2014). Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. *Arthritis and Rheumatology*, 66(9), 2450-2460.
- Schep, A. N., Buenrostro, J. D., Denny, S. K., Schwartz, K., Sherlock, G., & Greenleaf, W. J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Research*, 25(11), 1757–1770.
- Schille, C., Bayerlová, M., Bleckmann, A., & Schambony, A. (2016). Ror2 signaling is required for local upregulation of GDF6 and activation of BMP signaling at the neural plate border. *Development*, 143(17), 3182–3194.

- Schwabe, K., Garcia, M., Ubieta, K., Hannemann, N., Herbort, B., Luther, J., Noël, D., Jorgensen, C., Casteilla, L., David, J. P., Stock, M., Herrmann, M., Schett, G., & Bozec, A. (2016). Inhibition of Osteoarthritis by Adipose-Derived Stromal Cells Overexpressing Fra-1 in Mice. *Arthritis and Rheumatology*, 68(1), 138–151.
- Scott, D. L., Wolfe, F., & Huizinga, T. W. J. (2010). Rheumatoid arthritis. *The Lancet*, 376(9746), 1094–1108. [https://doi.org/10.1016/S0140-6736\(10\)60826-4](https://doi.org/10.1016/S0140-6736(10)60826-4)
- Shane Anderson, A., & Loeser, R. F. (2010). Why is osteoarthritis an age-related disease? *Best Practice and Research: Clinical Rheumatology*. 24(1), 15.
- Shen, J., Abu-Amer, Y., O’Keefe, R. J., & McAlinden, A. (2017). Inflammation and epigenetic regulation in osteoarthritis. *Connective Tissue Research*. 58(1), 49–63.
- Shepherd, C., Zhu, D., Skelton, A. J., Combe, J., Threadgold, H., Zhu, L., Vincent, T. L., Stuart, P., Reynard, L. N., & Loughlin, J. (2018). Functional characterisation of the osteoarthritis genetic risk residing at ALDH1A2 identifies rs12915901 as a key target variant. *Arthritis & Rheumatology*. 70(10):1577-1587.
- Steinberg, J., Ritchie, G. R. S., Roumeliotis, T. I., Jayasuriya, R. L., Clark, M. J., Brooks, R. A., Binch, A. L. A., Shah, K. M., Coyle, R., Pardo, M., Le Maitre, C. L., Ramos, Y. F. M., Nelissen, R. G. H. H., Meulenbelt, I., McCaskie, A. W., Choudhary, J. S., Wilkinson, J. M., & Zeggini, E. (2017). Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. *Scientific Reports*, 7(1), 8935.
- Tang, J., Su, N., Zhou, S., Xie, Y., Huang, J., Wen, X., Wang, Z., Wang, Q., Xu, W., Du, X., Chen, H., & Chen, L. (2016). Fibroblast growth factor receptor 3 inhibits osteoarthritis progression in the knee joints of adult mice. *Arthritis & Rheumatology*, 68(10), 2432–2443.
- Tsompana, M., & Buck, M. J. (2014). Chromatin accessibility: A window into the genome. *Epigenetics and Chromatin*. 7(1), 33. <https://doi.org/10.1186/1756-8935-7-33>
- Uhalte, E. C., Wilkinson, J. M., Southam, L., & Zeggini, E. (2017). Pathways to understanding the genomic aetiology of osteoarthritis. *Human Molecular Genetics*. 26(R2), R193-R201
- Usami, Y., Gunawardena, A. T., Iwamoto, M., & Enomoto-Iwamoto, M. (2016). Wnt signaling in cartilage development and diseases: Lessons from animal studies. *Laboratory Investigation*, 96(2), 186–196.
- van Meurs, J. B. J. (2017). Osteoarthritis year in review 2016: genetics, genomics and

- epigenetics. *Osteoarthritis and Cartilage*. 25(2), 181-189.
<https://doi.org/10.1016/j.joca.2016.11.011>
- Verzijl, N., DeGroot, J., Thorpe, S. R., Bank, R. A., Shaw, J. N., Lyons, T. J., Bijlsma, J. W. J., Lafeber, F. P. J. G., Baynes, J. W., & TeKoppele, J. M. (2000). Effect of collagen turnover on the accumulation of advanced glycation end products. *Journal of Biological Chemistry*, 275(50), 39027–39031.
- Vincenti, M. P., & Brinckerhoff, C. E. (2002). Transcriptional regulation of collagenase (MMP-1, MMP-13) genes in arthritis: integration of complex signaling pathways for the recruitment of gene-specific transcription factors. *Arthritis Research*, 4(3), 157. <https://doi.org/10.1186/ar401>
- Wang, H. L., & Neiva, R. F. (2008). Socket Augmentation: Rationale and Technique. In *Fundamentals of Esthetic Implant Dentistry* (pp. 140–224).
- Wells, T., Davidson, C., Mörgelin, M., Bird, J. L. E., Bayliss, M. T., & Dudhia, J. (2003). Age-related changes in the composition, the molecular stoichiometry and the stability of proteoglycan aggregates extracted from human articular cartilage. *The Biochemical Journal*, 370, 69–79.
- Wilkins, E., Dieppe, P., Maddison, P., & Evison, G. (1983). Osteoarthritis and articular chondrocalcinosis in the elderly. *Annals of the Rheumatic Diseases*, 42(3), 280–284.
- Wu, H., & Sun, Y. E. (2006). Epigenetic Regulation of Stem Cell Differentiation. *Pediatric Research*, 59, 21R–25R.
<https://doi.org/10.1203/01.pdr.0000203565.76028.2a>
- Wu, M., Chen, G., & Li, Y.-P. (2016). TGF- β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Research*, 4, 16009.
- Yoon, B. S., Ovchinnikov, D. A., Yoshii, I., Mishina, Y., Behringer, R. R., & Lyons, K. M. (2005). *Bmpr1a* and *Bmpr1b* have overlapping functions and are essential for chondrogenesis in vivo. *Proceedings of the National Academy of Sciences*, 102(14), 5062–5067.
- Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14), 2382–2383.
- Yudoh, K., Nguyen, van T., Nakamura, H., Hongo-Masuko, K., Kato, T., & Nishioka, K. (2005). Potential involvement of oxidative stress in cartilage senescence and

development of osteoarthritis: oxidative stress induces chondrocyte telomere instability and downregulation of chondrocyte function. *Arthritis Research & Therapy*, 7(2), R380-91.

Zeggini, E., Panoutsopoulou, K., Southam, L., Rayner, N. W., Day-Williams, A. G., Lopes, M. C., arcOGEN Consortium, & Loughlin, J. (2012). Identification of new susceptibility loci for osteoarthritis (arcOGEN), a genome-wide association study. *Lancet*, 380(9844), 815–823.

Zengini, E., Hatzikotoulas, K., Tachmazidou, I., Steinberg, J., Hartwig, F.P., Southam, L., Hackinger, S., Boer, C.G., Styrkarsdottir, U., Gilly, A., Suveges, D., Killian, B., Ingvarsson, T., Jonsson, H., Babis, G.C., McCaskie, A., Uitterlinden, A.G., van Meurs, J.B.J., Thorsteinsdottir, U., Stefansson, K., Davey Smith, G., Wilkinson, J.M., & Zeggini, E. (2018). Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nature Genetics*. 50(4):549-558.

Zhang, Y., Fukui, N., Yahata, M., Katsuragawa, Y., Tashiro, T., Ikegawa, S., & Lee, M. T. M. (2016). Identification of DNA methylation changes associated with disease progression in subchondral bone with site-matched cartilage in knee osteoarthritis. *Scientific Reports*, 6, 34460.

Zhang, Y., Fukui, N., Yahata, M., Katsuragawa, Y., Tashiro, T., Ikegawa, S., & Michael Lee, M. T. (2016). Genome-wide DNA methylation profile implicates potential cartilage regeneration at the late stage of knee osteoarthritis. *Osteoarthritis and Cartilage*, 24(5), 835–843.

Zhou, H. W., Lou, S. Q., & Zhang, K. (2004). Recovery of function in osteoarthritis chondrocytes induced by p16INK4a-specific siRNA in vitro. *Rheumatology*, 43(5), 555–568.

Zhou, S., Xie, Y., Li, W., Huang, J., Wang, Z., Tang, J., Xu, W., Sun, X., Tan, Q., Huang, S., Luo, F., Xu, M., Wang, J., Wu, T., Chen, L., Chen, H., Su, N., Du, X., Shen, Y., & Chen, L. (2016). Conditional Deletion of Fgfr3 in Chondrocytes leads to Osteoarthritis-like Defects in Temporomandibular Joint of Adult Mice. *Scientific Reports*, 6:24039.

List of publications

1. Ye Liu, Jen-Chien Chang, Chung-Chau Hon, Naoshi Fukui, Nobuho Tanaka, Zhenya Zhang, Ming Ta Michael Lee, and Aki Minoda. "Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis". *Scientific Reports*. 2018; 8(1), 15499.
2. Priyanshu Bhargava, Vidhi Malik, Ye Liu, Jihoon Ryu, Sunil C. Kaul, Durai Sundar, and Renu Wadhwa. "Molecular insights to Withaferin-A induced senescence: bioinformatics and experimental evidence to the role of NF-kappa B and CARF". *Journal of Gerontology: Medical Sciences*. 2018; Series A, gly107.
3. Satoshi Kojo, Hirokazu Tanaka, Takaho Endo, Sawako Muroi, Ye Liu, Wooseok Seo, Mari Tenno, Kiyokazu Kakugawa, Yoshinori Naoe, Krutula Nair, Kazuyo Moro, Yoshinori Katsuragi, Akinori Kanai, Toshiya Inaba, Takeshi Egawa, Byrappa Venkatesh, Aki Minoda, Ryo Kominami, and Ichiro Taniuchi. "Priming of lineage-specifying genes by Bcl11b is required for lineage choice in post-selection thymocytes." *Nature Communications*. 2017; 8(1), 702.
4. Kejuan Li, Yue Yu, Shuang Sun, Ye Liu, Sukant Garg, Sunil C. Kaul, Zhongfang Lei, Ran Gao, Renu Wadhwa, and Zhenya Zhang. "Functional characterisation of anticancer activity in the aqueous extract of *Helicteres angustifolia* L. roots". *PLoS One*. 2016; 11(3), e0152017.
5. Jihoon Ryu, Zeenia Kaul, A. Rum Yoon, Ye Liu, Tomoko Yaguchi, Youjin Na, Hyo Mi Ahn, Ran Gao, Il Kyu Choi, Chae Ok Yun, Sunil C. Kaul, and Renu Wadhwa. "Identification and functional characterization of nuclear mortalin in human carcinogenesis". *Journal of Biological Chemistry*. 2014; 289(36), 24832-24844.
6. Nishant Saxena, Shashank P. Katiyar, Ye Liu, Abhinav Grover, Ran Gao, D. Sundar, Sunil C. Kaul and Renu Wadhwa. "Molecular Interactions of Bcl-2 and Bcl-Xl with Mortalin: Identification and Functional Characterization." *Bioscience Reports*. 2013; 33(5), e00073.

Acknowledgements

It is my pleasure to thank all the people who have supported and helped me. Nothing would have happened without their support and help.

First and foremost, I would like to thank my supervisors /advisors, Prof. Zhenya Zhang, Dr. Ming Ta Michael Lee and Dr. Aki Minoda for their enthusiasm and continuous support for my Ph.D. study and research. I would also like to thank other academic advisors, Prof. Renu Wadhwa, Dr. Sunil Kaul and Piero Carninci for their motivation, encouragement and support during my studies at National Institute of Advanced Industrial Science and Technology (AIST) and RIKEN.

I would like to express my appreciation to my thesis committee, Prof. Yingnan Yang, Associate Prof. Utsumi Motoo for their valuable and insightful comments in my thesis defenses. I also appreciate the help from Sumiko Nishigawa, Xianglan Quan and Xi Yang for dealing with all my graduation documents patiently.

I would further thank Dr. Shiro Ikegawa (RIKEN & Tokyo University) for building up this human knee osteoarthritis research study, and Dr. Naoshi Fukui (Sagamihara National Hospital & Tokyo University) for supporting all samples.

I also would like to thank the fellowship supported by RIKEN Junior Research Associate (JRA) Program. Special thanks to my mentors Dr. Jihoon Ryu and Dr. Ran Gao for training my experiment skills and supported and introduced me to work in their laboratory; Dr. Jen-Chien Chang, Dr. Chung-Chau Hon and Dr. Jay W. Shin especially helping on bioinformatics. I also thank all my friends and colleges in the Minoda lab, Zhang lab, Lee lab and Renu lab.

Lastly, no words are measurable enough to appreciate and express my love to my families.