Androgen Receptor-dependent Enhancers in Prostate Cancer

by

Doğancan Özturan

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Doctor of Philosophy

in

Biomedical Sciences and Engineering

December 2021

Androgen Receptor-dependent Enhancers in Prostate Cancer

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a doctoral dissertation by

Doğancan Özturan

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Assoc. Prof. Nathan A. Lack (Advisor)

_____

Prof. Dr. Tamer T. Önder

_____

Assoc Prof. Umut Şahin

_____

Assist. Prof. Gözde Korkmaz

_____

Prof. Dr. Mesut Muyan

Date: _____

to my beloved parents Emine and Özcan Özturan...

ABSTRACT

Androgen Receptor-dependent Enhancers in Prostate Cancer
Doğancan Özturan
Doctor of Philosophy in Biomedical Sciences and Engineering
December 2021

Prostate cancer is one of the leading causes of cancer-related death in European men. In almost all patients, Androgen Receptor (AR)-mediated transcription is critical for the development and proliferation of cancer. Upon activation, the AR binds to DNA and induces the expression of genes essential for cellular differentiation and tumor growth. AR-mediated transcription is driven by distal regulatory enhancer elements that control gene expression by looping to gene promoters. However, there are exponentially more AR binding sites than differentially expressed genes. It is unknown how these regions work to induce transcription. Delineating the regulatory logic of AR-mediated transcription is critical to understanding how this transcription factor drives prostate cancer growth and progression. To identify these specific AR enhancers, we functionally tested all clinical AR binding sites with a massively multi-parallel enhancer assay to create the first "map" of AR transcriptional activity. From this, we found that only a subset of ARBS had functional enhancer activity AR. Those AR-regulated enhancers act as a regulatory hub that frequently cooperates with other ARBS to drive transcription. Yet intriguingly, we observed that many of the minimal enhancer activity sites are often required for gene transcription, suggesting that ARBS have different functions that are independent of only enhancer activity. To explore this, we systematically characterized the KLK gene locus and found that AR occupancy is dependent on other AR binding sites in the same transcriptional network. Overall, this work provides fundamental insights into AR functions to drive gene activation.

ÖZETÇE

Prostat kanserinde androjen reseptörüne bağımlı enhancer bölgeleri
Doğancan Özturan
Biyomedikal Bilimler ve Mühendislik, Doktora
Aralık 2021

Prostat kanseri, avrupalı erkeklerde kansere bağlı ölümlerin önde gelen nedenlerinden biridir. Hemen hemen tüm hastalarda, Androjen Reseptörü (AR) aracılı transkripsiyon, kanserin gelişimi ve proliferasyonu için kritik öneme sahiptir. Aktivasyon sonrası AR, DNA'ya bağlanır ve hücresel farklılaşma ve tümör büyümesi için gerekli olan genlerin ekspresyonunu indükler. AR aracılı transkripsiyon, gen promotörleri ile etkileşime girerek gen ekspresyonunu kontrol eden distal enhancer elementleri tarafından yönlendirilir. Genomda AR kontrolü ile eksprese edilen genlerden çok daha fazla AR bağlanma bölgesi (ARBB) vardır. Bu bölgelerin transkripsiyonu indüklemek için nasıl çalıştığı bilinmemektedir. AR aracılı transkripsiyonun mantığının tanımlanması, bu transkripsiyon faktörünün prostat kanseri büyümesini ve ilerlemesini nasıl yönlendirdiğini anlamak için kritik öneme sahiptir. Bu spesifik AR enhancer'larını tanımlamak için ve AR transkripsiyonel aktivitesinin ilk "haritasını" oluşturmak için tüm klinik AR bağlanma bölgelerini paralel enhancer deneyi ile işlevsel olarak test ettik. Bundan, ARBB'nin yalnızca bir alt kümesinin, işlevsel olarak enhancer aktivitesine sahip olduğunu bulduk. Bununla birlikte, AR tarafından düzenlenen enhancer'lar, transkripsiyonu yürütmek için sıklıkla diğer ARBB ile işbirliği yapan düzenleyici bir merkez görevi gördüğünü keşfettik. İlginç bir şekilde, minimal aktiviteye sahip enhancer bölgelerinin gen transkripsiyonu için gerekli olduğunu gözlemledik, ve AR'ın sadece enhancer aktivitesinden bağımsız farklı fonksiyonlara sahip olduğunu öne sürdük. Bunu araştırmak için, KLK gen lokusunu sistematik olarak karakterize ettik ve AR'ın bağlanma işlevinin aynı transkripsiyonel ağdaki diğer ARBB'lere bağlı olduğunu saptadık. Genel olarak bu çalışma, ARBB'lerinin AR güdümlü gen aktivasyonunu yönlendirme işlevlerine ilişkin temel bilgiler sağlamıştır.

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my supervisor Dr. Nathan Lack for giving me a scientific path and guiding me through it. With a combination of knowledge and scientific enthusiasm, he is the best advisor I could ever imagine. I will never forget our endless scientific discussions. Specifically, I want to thank him for creating an exciting environment where I felt like day one, every day.

In addition, I would like to thank my thesis observation committee members, Dr. Tamer Önder and Dr. Umut Şahin, for their absolute realism and rigorous scientific guidance. I also wish to express my thanks to Dr. Gözde Korkmaz and Dr. Mesut Muyan for accepting to be in my thesis defense jury. Separately, I would like to thank Dr. Tuğba Bağcı Önder for her support throughout my PhD training.

I want to express my deepest gratitude to Dr. Özgür Gül for being one of a kind scientific role model for me.

I would like to thank past and present Lack Lab members, Tunç Morova, Fırat Uyulur, Hilal Saraç, Fatma Özgün, Ceren Şeref, Betül Ersoy, Bengül Gökbayrak, Flora Huang, Ivan Yu, Shreyas Lingadahalli and Dr. Funda Şar for their friendship and for fruitful scientific discussions.

Special mention goes to the members of Nobelcast: Alişan Kayabölen and Uğur Akcan. Together, we started a fantastic journey that strengthened our friendship with wonderful memories.

My sincere thanks to fellow KUTTAM colleagues, Kenan Sevinç, Gülben Gürhan Sevinç, Nareg Pınarbaşı, Göktuğ Karabıyık, Gizem Nur Şahin, Fidan Şeker for their companionship.

I would like to thank my family Emine and Özcan Özturan, and Duygu, Erhan, and Emir Ayaz, for their unconditional love and support.

Last but not least, my love Gülce, thank you for being (in) my life. I'm glad to say that I look up to you as a person and a scientist. Every day is a joy with you, and with the newest member of our family, Derin, we will create brand new memories together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

3C	Chromatin conformation capture
4C	Circularized chromatin conformation capture
5C	Chromatin conformation capture carbon copy
AR	Androgen receptor
ARBS	Androgen receptor bindings sites
ARE	Androgen response element
ARV	Androgen receptor variant
ARV7	Androgen receptor variant 7
ATACseq	Assay for transposase-accessible chromatin using sequencing
Cas9	CRISPR-associated protein 9
ChIA-PET	Chromatin interaction analysis by paired-end tag sequencing
ChIPseq	Chromatin immunoprecipitation sequencing
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRi	Clustered regularly interspaced short palindromic repeats inhibition
CRPC	Castration-resistant prostate cancer
CTCF	CCCTC-Binding Factor
DBD	DNA binding domain
DHT	Dihydrotestosterone
DNA	Deoxyribonucleic acid
DNAse-seq	DNase I hypersensitive sites sequencing
eGFP	Enhanced green fluorescent protein
ER	Estrogen receptor
FISH	Fluorescence in situ hybridization
GR	Glucocorticoid receptor
GROseq	Global run-on sequencing

H3	Histone 3
HiC	High throughput chromatin conformation capture
HR	Hinge region
HSP	Heat shock protein
K27ac	Acetylation of the lysine residue at N-terminal position 27
kDA	Kilodalton
KLK	Kallikrein
KRAB	Krüppel associated box
LAD	Lamina-associated domain
LBD	Ligand binding domain
MMTV	Mouse mammary tumor virus
MPRA	Massively parallel reporter assays
MR	Mineralocorticoid receptor
PCa	Prostate cancer
PCR	Polymerase chain reaction
PR	Progesterone receptor
PSA	Prostate-specific antigen
PTM	Post-translational modification
qPCR	Quantitative PCR
RNAseq	Ribonucleic acid sequencing
scRNAseq	Single-cell ribonucleic acid sequencing
STARRseq	Self-transcribing active regulatory region sequencing
TAD	Topologically associated domains

Chapter 1

INTRODUCTION

1.1 Fundamentals of the genome

Nuclear DNA is often pictured as a static, passive, and stagnant chemical that simply acts as a template. However, the genome is a dynamic entity that involves numerous complexes from structural proteins that shape the DNA to a particular conformation, nuclear lamins that direct genome organization, and transcription factors (TF) that dictate cell identity. These complex events work in tandem to drive gene transcription in a highly controlled process.

1.1.1 3D genome organization

Chromatin contains histone proteins and genomic DNA compacted hundreds of times to fit into the nucleus. It is organized such that almost 2 meters of DNA is packed into approximately 10μm nuclear space [1–5] (Figure 1.1). Walther Flemming coined the term "chromatin" as it is readily stained by aniline dyes [6]. Early studies with nuclease digestion and electron microscopy showed that the nucleosome complex, the basic unit of the chromatin, consists of core histone proteins (H2A, H2B, H3, and H4) wrapped around a 146bp stretch of DNA [7]. These nucleosomes can form long-distance loops (>Mb) that allow interactions with cis-regulatory DNA elements (CRE). These interactions are mostly restricted to sub-megabase topologically associated domains (TADs). Contact frequency inside TADs is much greater relative to interactions outside of the TAD. As such, TAD boundaries are defined by the frequency of contacts within and outside of a given region [8].

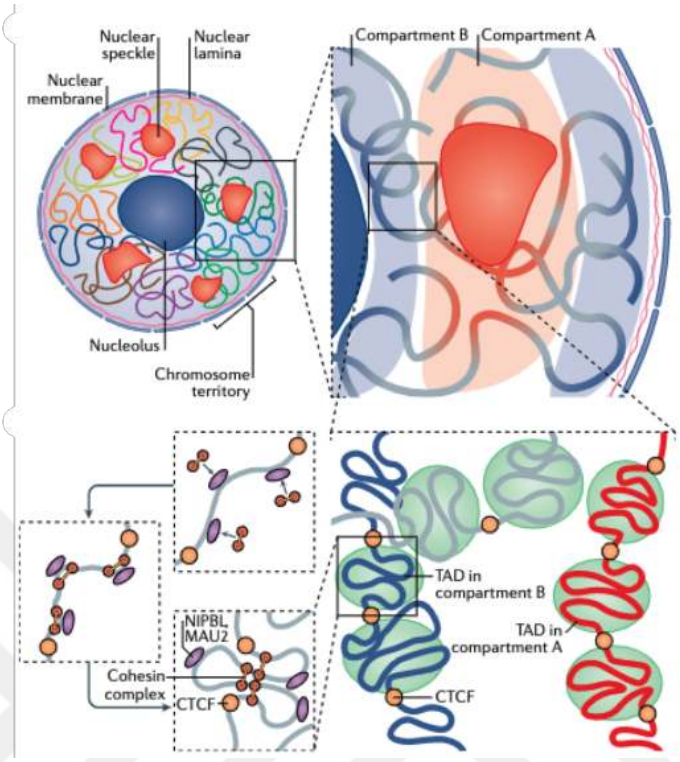


Figure 1.1: Hierarchy of the genomic organization. Nuclear DNA is structured in chromosome territories, compartments, topologically associated domains, and subTAD chromatin loops. Taken from [9]

Although the subTAD interactions can vary across tissues [10, 11], TADs are mostly evolutionarily conserved across different species [8, 12, 13]. TADs are associated with several features, including insulator protein CCCTC-Binding Factor (CTCF) and cohesin which reside on the borders of the TADs [14, 15]. These two proteins work together to form the loop domains via a "loop extrusion" model where cohesin recognizes a CTCF motif and extends the DNA until it reaches a reverse-oriented CTCF binding site [16]. These TADs (or loop domains) are dependent on this interaction and are lost when cohesin is removed [17]. The functional importance of TADs is still controversial as loss of these features does not affect transcription or "compartmentalization". Recent work has suggested that this may be due to the relatively short time these TADS are present [18]

The genome is compartmentalized into two spatial elements: active (Compartment A) and inactive (Compartment B) regions [19]. Active or compartment A regions

are associated with a high number of genes, active transcription, early replication, more accessible DNA, and histone trimethylation in lysine residue 36 (H3K36me3) [20]. Inactive or compartment B regions are associated with lamina-associated domains (LADs) located at the nuclear perimeter that are separated by CTCF binding sites [21]. Compartment B regions generally have decreased transcriptional activity and are associated with late replication [22–24]. As these different compartments are megabases away from each other [25] active regions more frequently contact other active regions [26]. As these different compartments have very different transcriptional activity, there are two main hypotheses about the relationship between gene activation and active/inactive region genomic organization. One theory states that when a gene is activated, its physical location moves towards compartment A [27]. Another claims that translocation happens before the gene is activated, leaving it in a "poised" state [28, 29]. The nucleus is also organized into chromosome territories that contain regions preferentially occupied by particular chromosomes. First conceptualized by Carl Rabl in 1885 [30] and then coined by Theodor Boveri in 1909 [31], this fundamental concept states that each chromosome occurs in a non-random location in the subcellular nuclear space [32]. Although chromosomes were thought to be static and not change their structure or position, there is increasing evidence from single-cell studies that compartmentalization and TAD borders can vary within the same cell lines [33–36].

1.1.2 Enhancers

From early studies in the simian virus 40 (SV40) genome in the late 1970s [37, 38], enhancers have been shown to mediate gene transcription. In essence, enhancers are transcriptional control units required for spatiotemporal gene activation. They are present in mammalians [39], plants [40, 41], yeasts [42], and even in the closest pre-metazoan species, Capsaspora owczarzaki [43]. They are found extensively throughout the human genome [44] and are involved in nearly every cellular function from temporal control of differentiation [45]. Human enhancers are

remarkably cell type-specific [46] to ensure that only specific genes are activated to do specific cellular tasks [47]. Although transcription starts at promoters, enhancers govern gene activation.

Transcription factor (TF) binding to enhancers provides the temporal control that drives a particular phenotype. For example, in Drosophila melanogaster, specific TFs bind to the same enhancer elements at different stages of development [48]. In bacteria, most TFs are present to prevent random RNA polymerase activity since most bacterial promoters are accessible to RNA polymerase. One bacterial TFs with sequence-specific properties is lac Repressor (LacR). It binds nonspecifically to random sequences and then slides via facilitated diffusion through DNA (on average 45bp) until it finds the target. LacR has only 3 binding sites and spends 90% of its time bound to nonspecific regions [49–51].

Enhancer activity generally correlates with several distinct features, including increased chromatin accessibility [52, 53] binding by several protein classes, including transcription factors [54–58] general coactivators such as CEBP/p300 [59], structural proteins [60], and RNA polymerase II (RNAPII) [61]. An extensive list can be found in Table 1.1.

The histone code hypothesis proposes that post-translational modification (PTM) of histone tails will recruit effector proteins that drive function [84, 85]. Therefore, CRE, including enhancers, are often defined based on their specific histone PTM. In support of this, there is extensive evidence that enhancers broadly correlate with mono-methylation of lysine four on the histone protein 3 (H3K4me1) and acetylation of lysine 27 on histone protein 3 (H3K27ac) [68]. RNA is also transcribed from genomic enhancer elements. In 1959, Harris et al. showed that most synthesized RNA doesn't leave the nucleus and originates from noncoding regions [86]. enhancer RNAs (eRNAs) are first discovered in the beta-globin dominant control region, which directs the activation of the beta-globin gene [87]. These transcripts can either be polyadenylated or non-polyadenylated, and transcription from these loci can be directional or bidirectional [88, 89].

Table 1.1: The proteome of enhancer elements

Category	**Examples**
TFs	Sequence-specific TFs [62–65]
	General Transcription Factors [66–68]
Structural Proteins	Modified chromatin histones [69]
	Cohesin [63], Condensin [70], CTCF [71], YY1 [72]
Coregulators	Mediator complex [73, 74] and Integrator complex [75]
	Chromatin remodeling complexes [76] and chromatin modifiers [77]
	Steroid receptor and other TF's coactivators [78–80]
Effector proteins	Histone "readers" [81]
Catalytic Enzymes	RNAPII [68]
	Others (PARP1, Topoisomerase-1, TET2, etc.) [82–84]

1.1.2.1 Enhancer-promoter interactions

Enhancers are not limited to specific chromosomal locations and can act far away from their target genes [90]. As they are orientation-independent, they are commonly located in distal intergenic regions of their target genes or intronic regions of entirely different genes [91]. They mainly function via enhancer-promoter looping, whereby the enhancer comes in close physical proximity to the promoter via chromatin looping [92]. However, although rare, there are non-interacting enhancers such as distal elements of the sonic hedgehog (Shh) gene in the mouse brain [93]. In the chromatin looping model, loops can be formed from scratch or pre-existing. While existing links are suitable for rapid transcriptional activation [94], *de novo* loops are formed by transcription factor-mediated gene regulation [65]. Previous work demonstrated that H3K27Ac and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) or DNase I hypersensitive sites sequencing (DNase-seq) (along with the "generic" HiC) could predict enhancer-promoter pairing, surpassing the "proximity" as a

surrogate of connectivity [95]. However, recent work using a CRISPR-Cas9 mediated system to create systematic deletions of the POU5F1 locus (2Mb), demonstrated that many experimentally validated enhancer regions frequently don't overlap with H3K27Ac (22%) [96].

1.1.2.2 Enhancer evolution

Enhancer biogenesis originates in different ways including exonic artifacts from duplication events [97, 98], accumulation of local point mutations [99], gene loss [100], and transposable element proliferation [101]. Epromoters, a specific subset of gene promoters that show enhancer activities have also been shown to regulate gene expression [102]. Further, duplicated promoters are also reported to show strong enhancer activity [103]. A study on the evolution of enhancers and promoters using H3k27ac and H3K4me3 profiles in 20 mammalian liver tissues demonstrated generally fast enhancer and slow promoter evolution, resulting in little conservation among mammalian species [39]. Enhancers therefore provide a rapid mechanism to adapt to environmental stress.

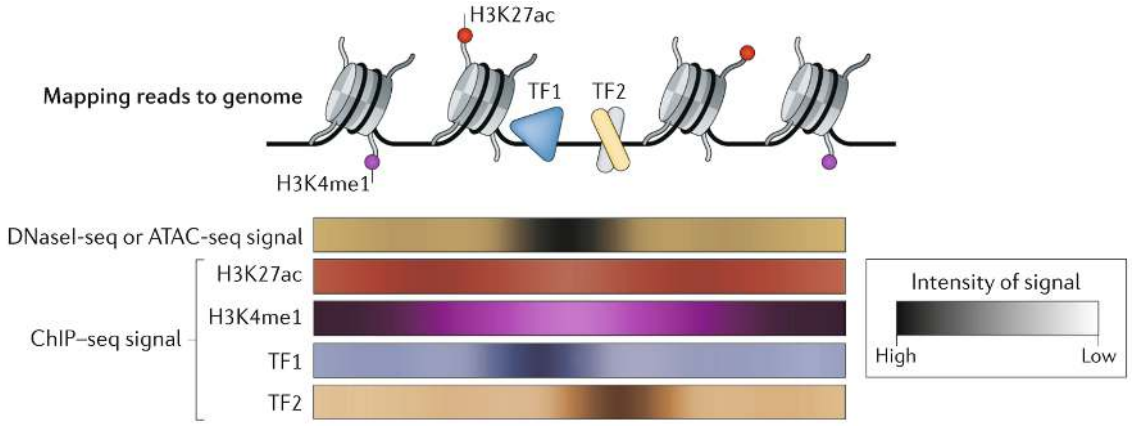


Figure 1.2: Descriptive signatures of enhancer elements. Transcription factor binding in open, H3K27ac, and H3K4me1 marked chromatin is associated with enhancers. Taken from [104]

1.1.2.3 Identification of enhancers

Enhancers are commonly identified by their overlap with several factors such as histone marks (H3K27ac, H3K4me1), open chromatin, or general coactivator binding (Figure 1.2). However, there are many active enhancers in the genome without these features [105]. A recent report demonstrated that H3K27ac loss does not affect chromatin accessibility and gene expression [106]. Further these correlative features are only descriptive and don't provide any information about enhancer strength. There is no evidence that, for example, increased openness or certain marks correlates with increased enhancer activity or vice versa. Numerous experimental methods have therefore been developed to both identify and validate enhancer activity. Enhancer traps have been used to detect endogenous regulatory regions. In this, reporter cassettes are randomly integrated into the genome to detect tissue-specific enhancer activity *ab extra* [107, 108]. Apart from random insertions, integrated reported assays can provide locus-specific enhancer activity [109]. Both enhancer traps and integrated reporter assays quantify enhancer activity but are limited by their low throughput. CRISPRi is an alternative method to test the impact of an enhancer on gene transcription [59, 110]. In this, specific gRNA is used to guide enzymatically dead Cas9 (dCas9) fused to Krüppel associated box (KRAB) to a gene locus where the KRAB effector domain creates heterochromatin upon dCas9 binding to DNA. With this endogenous enhancers can be inhibited to interrogate their impact on gene transcription in an increasingly high-throughput and temporal manner. To improve the repression, alternative repressor domains such as methyl CpG binding protein 2 (MeCP2) [111] and Zinc Finger Imprinted 3 (ZIM3) [112] have also been fused with dCas9. Besides transient epigenetic perturbations, CRISPR/Cas9-based systems can also create permanent mutations or deletions of enhancers [113]. This approach is also scalable, and enhancer-null cells can be generated for further experiments. However, because of the uncertainty in Cas9-mediated non-homologous end joining, extensive validation to characterize the variable genomic scar.

Plasmid-based reporter assays have also been extensively used to functionally test DNA-binding proteins for their transcriptional activity [114]. These assays commonly use the enzyme luciferase and it's naturally occurring bioluminescence substrate luciferin an [115]. In this episomal vector assay, the potential enhancer region is cloned upstream or downstream of a minimal promoter and the luciferase gene. Depending on the enhancer strength, the luciferase gene is transcribed which can be quantified with an enzymatic assay. Although this is a sensitive and inexpensive method, it has very low throughput. Considering the vast amount of putative enhancers in the genome, it can not be applied to test every region.

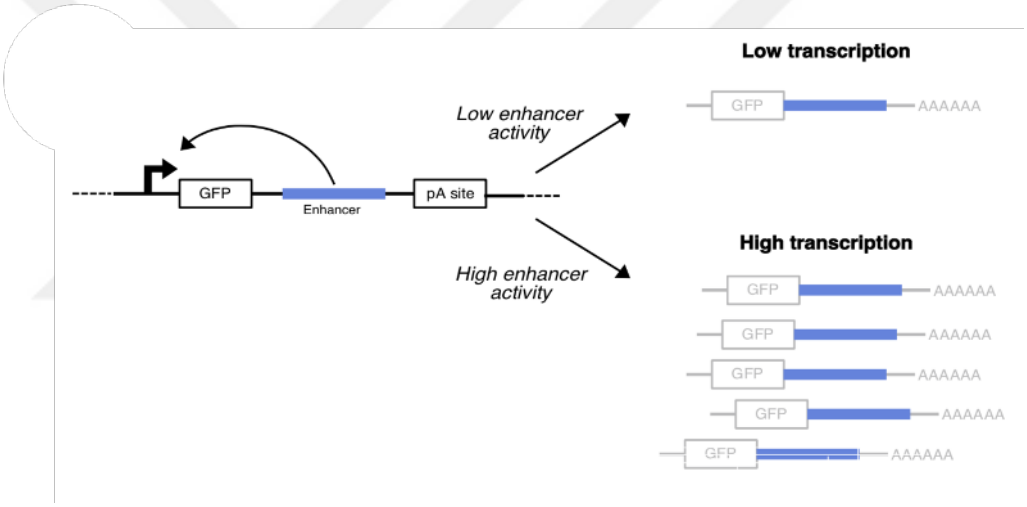


Figure 1.3: Schematics of STARRseq assay. Multiple genomic regions can be cloned into an episomal vector to test the enhancer activity via next-generation sequencing.

To overcome these limitations, massively parallel reporter assays (MPRA) were developed to test thousands of genomic regions for enhancer activity in a single experiment [116, 117]. One commonly used MPRA is Cis-regulatory elements analysis by sequencing (CRE-seq) [118]. This method uses barcoded plasmids and reporter fluorescence proteins to analyze enhancers with fluorescence and RNAseq simultaneously. However, the enhancer-barcode linkage can be lost because of the barcode switching. Another common MPRA is self-transcribing active regulatory region sequencing (STARRseq) [119]. This uses a luciferase assay-like vector containing a minimal promoter, and the regions of interest are cloned downstream

of the promoter. If there is enhancer activity, the region will then drive self-transcription, which can be quantified by next-generation mRNA sequencing. Thus transcript counts from the region of interest act as a reporter for enhancer activity (Figure 1.3). Potential enhancer libraries have been generated by DNA synthesis, which is currently limited to only 150-350bp [120]. Separately, randomly sheared genomic DNA is commonly used to ensure a large insert size (>500bp) [119]. This can either be used directly for genome-wide libraries or enriched by DNA capture [121], ChIP [64], or ATAC [122]. Since STARRseq is an episomal vector-based approach, any orientation can be cloned as it has been reported that enhancers work in a sequence orientation-independent fashion [123]. Most of the initial MPRA (including STARRseq [119]) use pGL3/4 vectors, which have a bacterial origin of replication (ORI), and it has been shown to initiate non-specific transcription rather than the engineered minimal promoter [124]. A second-generation STARRseq plasmid was designed to solve this problem by simply utilizing the ORI as the core promoter [124]. Using this, any human promoter sequence can be integrated into the vectors to test enhancer-promoter specificity [125].

1.1.2.4 Chromosome conformation capture

Enhancers work by dynamically coming in close proximity to promoters through chromatin looping with the help of DNA-binding proteins such as transcription factors [126]. The ENCODE project reported that the average distance of enhancer-promoter (E-P) interactions is approximately 120 kb and that there are nearly four enhancers for each particular active gene [127]. Previously, these specific E-P interactions could only be shown by low-throughput techniques such as fluorescence in situ hybridization (FISH) [128, 129]. However, the introduction of chromosome conformation capture (3C) and related approaches has provided a robust methodology for quantifying physical interaction at different genetic loci [130]. After crosslinking the nuclei, the initial 3C method utilized restriction enzyme digestion and proximity ligation, resulting in a chimeric DNA library.

From this, the strength of all DNA-DNA "interactions" can be interrogated using specific PCR primers. Subsequent approaches were developed to be performed with region-specific PCR using 4C [131–133] or 5C [134] or with oligo pull-down (Capture-C) [135, 136]. The introduction of HiC, a 3C-based method incorporating a streptavidin pull-down followed by massively parallel sequencing, can identify genome-wide interactions to be identified [20]. With this approach, chromosome territories, compartments, and TADs are observed. For example, on a HiC map, compartments A and B show themselves as a Checkerboard pattern [20]. One limitation of all 3C-based methods is that they rely on proximity ligation as a correlate for DNA-DNA interactions. Therefore, if two regions in the genome interact but are not close enough to be crosslinked, that interaction will not be captured during proximity ligation. This is particularly problematic for interactions in specific nuclear bodies. Subsequent methods such as genome architecture mapping (GAM) attempts to solve this limitation by using sections of individual nuclei and detecting the distance between two regions in the genome by sequencing [137]. By doing this to thousands of cryosections, GAM infers the genome-wide chromatin contacts. Split-Pool Recognition of Interactions by Tag Extension (SPRITE) [138] is an approach that also uses crosslinking. However, after that, it applies rounds of tag extension (using 96-well plates) to the ends of the DNA, assuming that regions that are crosslinked together would have the same extended tag [138]. SPRITE reported many contacts at much closer close distances than HiC.

Another limitation of the HiC is the use of restriction enzymes that digests the crosslinked DNA before chimeric DNA ligation. Even with frequent cutters such as DpnI or MseI that recognize and cut every 4 bp, interactions at a resolution window <1-4 kb can not be detected by HiC. The use of micrococcal nuclease to digest naked DNA overcomes this limitation and can map contacts at the single nucleosome level. Termed Micro-C, this method can achieve nearly 200bp resolution [10, 11, 139].

Enhancer-promoter connections are driven mainly by DNA-binding transcription

Table 1.2: Chromosome conformation capture assays

Assay	**Description**
3C	Fundamental assay to test whether two regions in the genome are nearby
4C	Assay to test a given region DNA interactome
Capture-C	Assay to test a given region DNA interactome using the pull-down
5C	Assay to test multiple regions DNA interactome
HiC, GAM, SPRITE	Assay to uncover the whole DNA interactome
ChIA-PET, HiChIP, PLACseq	Assay to uncover the DNA interactions genome-wide centered around a protein of interest

factors (TF). There are numerous methods to capture TF-specific interactions. Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) [140] uses immunoprecipitation followed by ligation of the linker tags to the ends of the DNA to differentiate chimeric reads. HiC can be combined with chromatin immunoprecipitation to make a transcription factor-specific interaction map either in HiChIP [141] or PLACseq [142]). Further, cis-regulatory DNA interactions can be identified from single-cell chromatin accessibility data [143]. Since active regions reside in the open chromatin, co-accessibility profiles for genomic regions strongly correlate with chromatin interactions. The most commonly used chromosome conformation capture assays methods are listed in (Table 1.2).

1.2 Prostate cancer

Prostate cancer (PCa) is one of the most common cancer types in North America and Europe. An estimated one in every nine men will develop prostate cancer in their lifetime [144]. Despite improved detection methods, better therapeutics, and advancements in clinical management, PCa still remains one of the leading causes of cancer death in European men after lung and colorectal cancer. Risk factors

include age, ethnicity, country, and family background [145–148]. The human prostate is a part of the male reproductive system. It is located inferior to symphysis of the pubis, superior to the urogenital diaphragm, and anterior to the rectum and encircles the urethra. The Greek physician Herophlius and Venetian anatomist Niccolò Massa (b.1485, d.1569) first described the prostate [149]. Andreas Vesalius, an anatomist from modern-day Belgium (b.1514, d.1564), created the first detailed drawings of the gland [150]. In modern times the clinical pathologist John E. McNeal reevaluated the structures of the prostate based on histologic characteristics, ductile landscape, and the relationship to the urethra and ejaculatory ducts [151]. From this, he proposed that the prostate organ is comprised of four anatomic regions: a peripheral zone (PZ), central zone (CZ), transition zone (TZ), and anterior fibromuscular zone (stroma). The prostate epithelium is primarily luminal and basal cells with rare neuroendocrine cells [152, 153]. Androgen receptor (AR) signaling is critical in developing and maintaining male sexual organs, including the prostate. Demonstrating the crucial role, AR inactivity results in complete androgen insensitivity syndrome (AIS) in males, which in addition to a female phenotype, prevents prostate development [153].

1.2.1 Prostate Cancer Diagnosis, Progress, and Treatment

Prostate cancer is detected by screening tests, including digital rectal exam (DRE) and prostate-specific antigen (PSA) blood tests. This is then supported by ultrasound, magnetic resonance imaging (MRI), and eventually, a prostate biopsy sample. The prostate cancer stage is characterized by the summation of different parameters: Gleason patterns from the biopsy, prostate-specific antigen (PSA) level, and clinical stage. The Gleason score describes the pathology of the cancerous state by grading the biopsy based on the microscopic appearance and cellularity. After combining these results, a stage (T1-4) of cancer is determined. The stages can be summarized as follows: T1: The tumor isn't detectable with a digital rectal exam (DRE) or imaging but is with tissue biopsy; T2: The tumor is

detectable with a DRE or imaging but is confined to the prostate; T3: Cancer has grown outside the prostate and into the seminal vesicles; T4: Cancer has metastasized into other nearby tissues. If there are suspicions of metastases after the initial diagnosis, imaging tests such as bone scan, computerized tomography (CT), or positron emission tomography (PET) can be used. According to the National Comprehensive Cancer Network (NCCN), risk classification of the prostate cancer stages is as follows [154]: Low risk: T1–T2, Gleason score <=6, and PSA <10 ng/ml; Intermediate risk: T2 or Gleason score 7 or PSA 10–20 ng/ml; High risk: T3 or Gleason score 8–10 or PSA >20 ng/ml. Very high risk for locally advanced prostate cancer: T3–T4 or primary Gleason pattern 5 or >5 cores with Gleason score 8–10; Metastatic risk: N1 (tumor has grown into nearby lymph nodes.) or M1 (tumor has grown into distant areas.) with any T stage.

1.2.2 Castration-Resistant Prostate Cancer

Active surveillance of the rate of increased PSA level in men with less-aggressive types of prostate cancer is associated with a low risk of progression [155–157]. The first line of treatment strategies for PCa includes prostatectomy and radiotherapy. Patients with recurrent or metastatic forms of the disease would be treated with androgen deprivation therapy (ADT). While initially effective, approximately 20-50% of the PCa patients develop resistance to ADT and progress to a more aggressive form of the disease termed castration-resistant prostate cancer (CRPC) [158, 159]. Intriguingly, despite the castrate conditions, AR remains highly active in CRPC [160]. Multiple mechanisms of resistance contribute to CRPC, most commonly involving AR mutations [161], amplifications of either AR gene or enhancer [162–164], and constitutionally variants [163, 164].

1.2.3 Androgen Receptor

The growth of almost all prostate cancers is dependent on AR [165]. First discovered in the cytoplasm of the rat prostate [166], the AR belongs to the steroid hormone group of nuclear receptors that includes estrogen receptor (ER),

glucocorticoid receptor (GR), progesterone receptor (PR), and mineralocorticoid receptor (MR). This receptor family is distinguished from other nuclear receptors such as vitamin D3 and thyroid hormone receptors, as a specific ligand triggers activation. AR is activated by androgens that commonly include testosterone and the more active metabolite dihydrotestosterone (DHT) [46]. Testosterone is produced from Leydig cells within the testes and converted to DHT by 5-alpha reductases once it enters prostate cells [167, 168]. The AR gene resides on the X chromosome q11-12 and comprises eight exons (2750 bp in total). The gene covers 90kb with 10.6kb mRNA and 1.1 kb and 6.9kb 5' and 3' untranslated regions, respectively [169]. When translated, the AR is 919 amino acid long (110 kDa) and consists of four regions: an NH2 transactivation domain (NTD) resides in exon 1, a DNA-binding domain (DBD) that resides in exon 2 and 3, a hinge region that resides in exon 4, and a ligand-binding domain (LBD) that resides in exon 5-8. Figure 1.4).

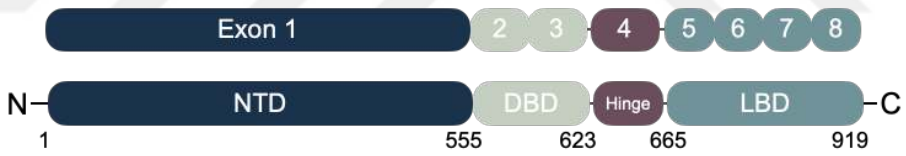


Figure 1.4: Structure of AR. The exonic regions and their corresponding locations are shown in the protein.

The AR-NTD is encoded entirely by exon 1 and comprises more than half of the protein. The NTD structure is yet to be determined as it is intrinsically disordered [170]. Previous work demonstrated that the AR is dependent on NTD for its transcriptional activation via the activation function 1 (AF-1) domain [171]. The NTD interacts with LBD through an N-C intramolecular interaction that is proposed to stabilize the protein and prevent misactivation [172]. The AR-DBD binds to DNA through its three zinc-finger helices [173]. It is highly conserved among steroid hormone receptors [174]. The AR hinge region contains the nuclear localization sequence (NLS), which is exposed upon androgen binding. The AR-LBD is highly conserved among different species but is distinct from other

steroid hormone receptors [175, 176]. Androgens are bound by the AR-LBD at the ligand-binding pocket. Therefore, the deletion of LBD makes AR unresponsive to androgen activation [177]. Structurally, it consists of four -sheets and eleven -helices. The AF2 domain resides in this ligand-binding pocket formed by H12 (Helix 12). The BF3 (binding function 3) surface resides in the domain reported regulating AF2 co-activator interactions [178]. Several drugs (both steroidal and non-steroidal) have been developed to target this site to inhibit AR activity.

1.2.3.1 Post-translational modifications of AR

AR undergoes extensive PTMs such as phosphorylation, acetylation, SUMOylation, and ubiquitination that can influence its binding partners, transcriptional activation and spatiotemporal properties in the cell. Various phosphorylation PTMs have been reported at serine/threonine and tyrosine residues [179, 180]. Specifically, phosphorylation at serine 515 causes AR to degrade by proteasome upon ubiquitination [181]. On the other hand, phosphorylation at serines 16, 81, 256, 308, 424, and 650 g can also increase the transcriptional activity of the nuclear receptor [182]. Acetylations of lysine (630, 632, and 633) residues in the AR hinge region cause decreased translocation of AR upon ligand binding and therefore decreased transcriptional activity. Small ubiquitin-related modifiers (SUMOs) covalently attach to the AR to modify transcription [183] and repression [184]. AR NTD SUMOylation is controlled by SUMO-1 at two lysine residues (386 and 520) and results in transcriptional repression of specific genes [185]. Inhibition of ligand-free AR in the cytosol by histone deacetylases (HDAC) is mediated by SUMOylation [186]. Finally, modifications by SUMO-2/3 cause AR to disengage from the genomic DNA and result in the down-regulation of AR-regulated genes [187]. Interestingly, one of the members of the PIAS family, PIASy, which shows SUMOylation activity [188], represses AR activity independent from this modification [189]. Separately, SENP1 (SUMO Specific Peptidase 1) acts as a co-regulator of AR [190] and positively affects AR-mediated gene regulation through desumoylation of HDAC1 [191].

1.2.3.2 AR co-regulators

Co-regulatory interactions strongly influence AR activity. Forkhead box (FOX) protein FOXA1 is a pioneer factor in AR-mediated transcription that can increase local chromatin accessibility at heterochromatin [192, 193]. Histone variants H3.3 and H2A.Z determine the nucleosome stability [194]. Upon DHT binding, FOXA1 displaces the unstable structure and binds there with AR [195]. FOXA1 is commonly mutated in various cancers including prostate [196], breast [174, 197], bladder [198]. Its 3' untranslated region (UTR) [199] and Wing2 domain and R219 residue mutated in prostate cancer resulting in alteration of its pioneer activity [200]. In addition to FOXA1 gene mutations, its DNA binding sites are also commonly mutated in primary prostate tumors [201]. In a large PCa patient cohort, the different FOXA1 aberrations were categorized into three categories based on cancer progression [202]. Homeobox protein HOXB13 is a well-studied AR co-regulator that is upregulated in prostate cancer [203]. It belongs to a family of 39 HOX family members that have a role in the spatial organization of vertebrate development [204]. It has two MEIS1 (Meis Homeobox 1) interaction domains. MEIS1 requires HOXB13 for strong DNA binding [205]. Exogenous expression of FOXA1 and HOXB13 causes a transformation of the AR cistrome in benign prostatic epithelial cells to a more tumor-specific cistrome [206]. TIF2 and SRC1 are hyper-active in the castration state (compared to benign prostate hyperplasia) in prostate cancers, and together they activate AR and promote growth [160]. Intriguingly, the AR both activates and inactivates a similar number of genes. While poorly characterized, gene repression has been proposed to occur via co-repressors, including NCoR (nuclear receptor corepressor 1) and SMRT (silencing mediator of retinoid and thyroid receptors) [207, 208]. These recruit histone deacetylases at androgen receptor binding sites (ARBS) to induce transcriptional repression by deacetylation and chromatin compaction [186]. NCoR is originally a regulator of muscle production in humans and other species such as Caenorhabditis elegans [209]. In PCa, it works by binding to AR to suppress the AR-driven gene expression [207]. Antagonist-bound AR recruits NCoR and SMRT

to the promoter of *PSA/KLK3*, which is a canonical AR-regulated gene [210]. Some studies suggest that bicalutamide-bound AR has increased corepressor recruitment [211].

1.2.3.3 AR binding sites

The unactivated AR is located in the cytosol, where it is maintained in an apo-confirmation by chaperone proteins including HSP27 [212], HSP90 [213], HSP70 [214], and HSP40 [215]. Once activated by androgens, the AR translocates to the nucleus, binds to tens of thousands of ARBS found throughout the genome [216]. ARBS primarily act as cis enhancers of AR-regulated genes. These play crucial roles in the survival and proliferation of normal prostate development, benign prostatic hypertrophy, and tumorigenesis [217]. These cis-acting elements' role in gene activation was first described for the enhancers of the rat probasin gene [218]. AR commonly recognizes the 5'-AGAACA-3'-like motifs [219] and binds as a symmetrical dimer. However, only 25% of the AR binding sites have a full AR motif, while >70% have a half-site motif sequence [220]. These sequences have been shown to initiate transcription with long terminal repeats (LTR) of mouse mammary tumor virus (MMTV) in reporter plasmids [219, 221]. The AR binding sites across genomes correlate with prostate cancer outcomes [222]. AR binding sites have a higher mutational burden exclusively in PCa [223]. In CRPC, there is a dramatic change in the AR cistrome with more than >17,000 new binding sites [224]. Although these new regions are proposed to play a role in the transcriptional activity of genes involved in CPRC, the mechanism of these gained sites is yet to be determined.

1.2.3.4 AR-mediated transcription

Upon activation, AR recruits basal transcriptional machinery to induce gene activation [225]. This complex consists of transcription factors IIs (TFII), including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, and bound to core promoters (including TATA box). CBP/p300 [226] and p160 [227] have also been

directly involved in AR gene activation through histone acetylase activity. Indeed, these complexes are recruited to AR-regulated gene enhancers to induce gene transcription [210] (Figure 1.5).

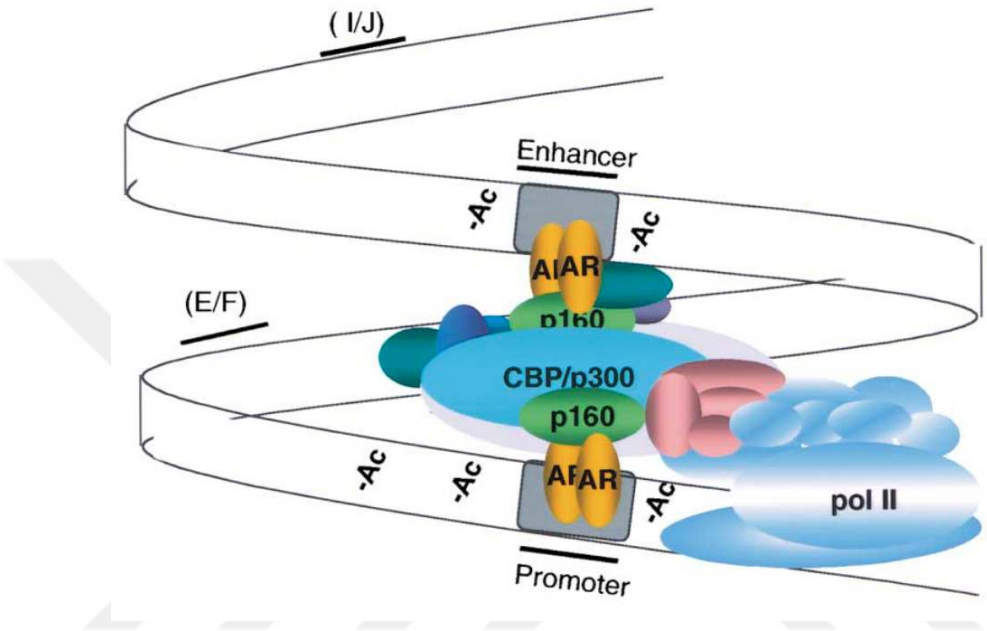


Figure 1.5: The proteome of AR-mediated transcription Coactivators and basal transcriptional machinery are recruited through enhancer-promoter chromatin looping mediated by AR to drive gene activation. The figure is taken from [210].

Although the mechanism of AR-mediated enhancer-promoter interactions is poorly understood, this is believed to occur through MED1 (Mediator Complex Subunit 1) [228]. MED1 is part of the multiprotein Mediator complex, which recruits transcriptional proteins such as RNA Pol II and preinitiation complex upon transcription factor binding. Mediator interacts with the C-terminal of the RNA Pol II at the gene promoters. Upon phosphorylation by CDK7, a component of TFIIH, this interaction disrupts, and RNA Pol II starts to elongate. MED1 binds to the NTD of AR (residues 505 and 537), stabilized by CDK7-mediated phosphorylation of MED1 [229, 230]. MED1 is critical to AR-regulated gene expression [231]. Upon androgen induction, MED1 participates in *PSA/KLK3* gene activation by interacting with both noncoding *PSA/KLK3* enhancer (AREIII) RNA [228] and with AR at *PSA/KLK3* promoter [231]. MED1 is also

reported to be enriched in super-enhancers [232] with BRD4, an interactor of AR [233]. While poorly understood, AR is believed to form liquid-liquid phase separation (LLPS) through AR-MED1 interaction [234]. Separately, AR is also found in LLPS to be tagged for proteasomal degradation via CRL3 (cullin3-RING ubiquitin ligase) regulated by SPOP (speckle-type POZ protein) [235].

1.2.3.5 KLK locus

The Kallikrein (KLK) family of genes has diverse expression profiles throughout human tissue, and they are broadly expressed except in nerve cells. The name "Kallikrein" was given as the initial KLK1 protein was discovered in pancreatic (Greek kallikreas) tissue [236]. Located in chromosome 19q13.3, there are 15 KLK genes found in the KLK locus (Figure 1.6) [237]. Almost all KLK genes consist of 5 exons with variable intronic regions. KLK1 has kininogenase activity and breaks down kininogen, a precursor protein that plays a role in blood coagulation. Unlike KLK1, none of the other genes in the locus have kininogenase activity [238]. These genes are therefore classified as kallikrein-related serine proteases [239]. Other than genes expressed in glands, a plasma kallikrein gene, *KLKB1*, is also reported [240].

Figure 1.6: KLK gene locus. It consists of 15 evolutionary conserved genes related and expressed in diverse tissues in the human body.

Different hormones regulate KLK genes, including androgens, estrogens, and glucocorticoids [241]. *PSA/KLK3* is a canonical AR-activated gene that has been used as a standard model to understand AR-targeted transcriptional action [242–244]. Further, *PSA/KLK3* is a common biomarker used to detect prostate cancer [245]. It encodes a 240 amino acid long glycoprotein with a molecular mass of 33 kDa that degrades proteins involved in the inhibition of semen coagulation [244]. Like most AR-targeted genes, the expression is regulated by a promoter region that works with one or more enhancers [246, 247]. The

proximal promoter regions AREI (-156bp) and AREII (-400bp) are essential for *PSA/KLK3* transcription. The known enhancer region of the *PSA/KLK3* (AREIII) is located approximately 4.5Kb upstream from the transcription start site (TSS) of the gene found in a DNaseI hypersensitive site that is marked with H3K27ac in androgen-treated AR-positive LNCaP cells [248]. In addition to *PSA/KLK3* and *KLK2*, AR expression correlates with *KLK*, *KLK4*, and *KLK15* expression, suggesting an AR-regulated activation mechanism. In AR-negative cell lines such as DU-145, PC-3, RWPE1, and RWPE2, there is no activation of those genes, but interestingly one gene, *KLK14*, is expressed in a hormone-independent mechanism [244]. Recent work demonstrated that a CTCF binding site between *KLK1,2,3,4,15* and *KLK5-14* caused reactivation of inactive genes when disrupted [249].

1.3 Main questions addressed

While we can readily identify ARBS, it is unclear how they contribute to gene transcription. This is further complicated by the vast difference in the number of AR-regulated genes (hundreds) and binding sites (tens of thousands). Delineating this complex mechanism of action is critical to understanding the role of AR in prostate cancer. However, only a fraction of AR binding sites have been characterized. Therefore, this work includes:

1. Identification of AR enhancers by using a massively parallel reporter assay and characterize their role in gene transcription (Chapter 3).

2. Systematic characterization of the role of each ARBS in transcription in KLK locus (Chapter 4).

Chapter 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals and Reagents

All of the material that has been used in this thesis are purchased as follows (unless otherwise stated):

Table 2.1: Material brands

Material	**Brand**
Restriction Endonucleases	New England Biolabs, Thermo Fisher
SYBR™ Green PCR Master Mix	Thermo Fisher
Protein A/G Magnetic Beads	Thermo Fisher
Genomic DNA, Plasmid and RNA Isolation & Gel and PCR cleanup kits	New England Biolabs, Invitrogen, MACHEREY-NAGEL, Qiagen, Zymo Research
Western Blot Reagents	Biorad
Transfection Agents	Mirus Bio
Other Chemicals	Sigma, Merck

2.1.2 Bacterial strains and growth media

Escherichia coli (*E.coli*) competent cells and growth media that have been used in this thesis are listed in Table 2.2 and Table 2.3.

Table 2.2: Competent cells and their genotypes

Strain Name	**Genotype**
Stbl3	F$^-$ $mcrB mrr hsdS20(r_B^-, m_B^-) recA13 supE44 ara-14 galK2 lacY1$ proA2 rpsL20(StrR) xyl-5 $\lambda - leumtl - 1$
DH5α	F$^-$ $endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20$ $\Phi 80 dlacZ\Delta M15 \Delta(lacZYA-argF)U169, hsdR17(r_K^- m_K^+), \lambda$–

Table 2.3: Bacterial broth and agar

Media Name	**Components**
BD Difco™ Dehydrated Culture Media: LB Agar, Miller (Luria-Bertani)	Tryptone 10g/L, Yeast extract 5g/L, Sodium Chloride 10g/L, Agar 15g/L
BD Difco™ Dehydrated Culture Media: LB Broth, Miller (Luria-Bertani)	Tryptone 10g/L, Yeast extract 5g/L, Sodium Chloride 10g/L

2.1.3 Human cell lines

The human cell line stocks used were originally validated via short-tandem repeats (STR) profiling in 2015. All cultures were screened weekly for mycoplasma contamination.

Table 2.4: Human cell lines used in this thesis

Cell Line	**ATCC**
LNCaP	CRL-1740
HEK 293T	CRL-3216
RWPE1	CRL-11609

2.1.4 Plasmids

2.1.4.1 Viral packaging

Packaging vector pCMV-dR8.2dvpr (Addgene plasmid #8455) and pCMV-VSV-G envelope vector (Addgene #8854) were used for packaging retroviruses along with the viral vector of interest. eGFP expressing plenti-CMV-GFP-puro (Addgene plasmid #17448) vectors were used as a positive control.

2.1.4.2 STARRseq, CRISPR/Cas9 and dCas9-KRAB

hSTARR-seq_ORI vector (Addgene plasmid #99296) is used to clone AR binding site inserts with gibson assembly. lentiCRISPRv2 (Addgene plasmid #52961) and lentiguide-puro (Addgene plasmid #52963) plasmid are used in CRISPR/Cas9 mediated knockout and CRISPR/dCas9-KRAB mediated repression, respectively.

2.1.5 gRNA sequences

Table 2.5: CRISPRi gRNAs

Name	**Sequence**
KLK3_inducible	TGAAAACAGACCTACTCTGG
KLK3_promoter	TCCTCACCCTGTCCGTGACG
KLK3_cons	GGCAGATGCTGCGCACACAG
*FKBP5*_promoter	TAAGATGCCCAGTCGGGACT
*FKBP5*_inducible	CACGAGGTTCTGCACCATCT
*FKBP5*_cons	ATCCAGGCGCCACCCAAGCC
*FKBP5*_inactive	AACTCTTTCTCTCCCGCGCT
*STEAP4*_promoter	GATAAGGGACGGGAGCCTTG
*STEAP4*_cons	GGTTGGTTTTGACTTCCAAA
*STEAP4*_inducible	TCTCAACATGGCTTATCAGC
*STEAP4*_inactive	GATACCCCCATTTCTGTGAT
*DBI*_promoter	AAGCACAGGGCAGGACGTCG
*DBI*_inducible	AGAAAAGAACACCGTGACCC

Table 2.6: Non-targeting gRNAs

Name	**Sequence**
Non-targeting_1	TCATGCTTGCTTGGGCAAAA
Non-targeting_2	CTGAAGGTGTCTGGCAGAGC

Table 2.7: Sequences for the creation of KLK3/eGFP cell line

Name	Sequence
KLK3_exon4_1	AGGGGGCAAAAGCACCTGCT
KLK3_exon4_2	CCCTACTCCCAAGATCTTGA
KLK3_exon4_3	TCCCAAGATCTTGAGGGGAA

Table 2.8: KLK CRISPRi screen gRNA targets

Name	Sequence	Name	Sequence
KLK_ARBS1_1	AATAAATAAATCATTATGGA	KLK_ARBS9_1	ACTTTATGACTCTAAATTGT
KLK_ARBS1_2	TAATGATTTATTTATTGGCT	KLK_ARBS9_2	AAGTGGCTCAAAGACAGGAA
KLK_ARBS1_3	ACATGTCCTTCCGGGGCAGG	KLK_ARBS9_3	GTCGCCAATGCAAAGTTAGA
KLK_ARBS1_4	AGATATTATCTTTGTTATGT	KLK_ARBS9_4	CTTTGAAATGGACCCAACAT
KLK_ARBS2_1	GATTGAAAACAGACCTACTC	KLK_ARBS10_1	TGTGTTACAGGCTGAGAAAA
KLK_ARBS2_2	TGAAAACAGACCTACTCTGG	KLK_ARBS10_2	CCAACTCAGTTTACCACTCG
KLK_ARBS2_3	TCTAGGACAGTAAGCAAGCC	KLK_ARBS10_3	ACTGCATCTGCCACCAGAAA
KLK_ARBS2_4	GAGAGAGATATCATCTTGCA	KLK_ARBS10_4	GCCACCAGAAAAGGTGCTGA
KLK_ARBS3_1	GGAGATATTATCTTTATAAT	KLK_ARBS11_1	GATTGAGATAAGGTGTAGGA
KLK_ARBS3_2	GGTTGAAAGCAGACCTACTC	KLK_ARBS11_2	AGCAATTACAGATTGAGATA
KLK_ARBS3_3	TGAAAGCAGACCTACTCTGG	KLK_ARBS11_3	ACCAGCAATAATAATAAACA
KLK_ARBS3_4	GACGTTAACTTTATTATCTA	KLK_ARBS11_4	TCTATGGTGTTTATTACTGT
KLK_ARBS4_1	TCAGTCTGGAGGCATTCCCA	KLK_ARBS12_1	TTATTTCTCACGTTACAAGC
KLK_ARBS4_2	CACATACTAGATCAGTCTGG	KLK_ARBS12_2	CATGTGCCACGCCATCCTTG
KLK_ARBS4_3	GAGGGCCTTGGTCAGCATCT	KLK_ARBS12_3	CGCCCCTTGTGCCAGGACCT
KLK_ARBS4_4	TGGCACCTAGATGCTGACCA	KLK_ARBS12_4	TCCCAGGACCTTACTATTCA
KLK_ARBS5_1	GTTAGCCTGCCACAGTTTTA	KLK_ARBS13_1	CTGAGTGTATCTTACACATT
KLK_ARBS5_2	TGCCAGTATCCATAAAACTG	KLK_ARBS13_2	TGCATGCTGTCCTGTGTGAT
KLK_ARBS5_3	GGATTCTGGATCAAAGGCAA	KLK_ARBS13_3	ATAGGATCTGAGGCCCAGGC
KLK_ARBS5_4	TGATCATTCCTGTGGGACTT	KLK_ARBS13_4	CAGCGCCGGCCGACCCCGCC
KLK_ARBS6_1	CATGATGCAAACCCAAGCAT	KLK_ARBS14_1	ATAGTTTTATAGTCAGGAAA
KLK_ARBS6_2	TGATGCAAACCCAAGCATAG	KLK_ARBS14_2	GGATGGTAGCAGCTGCAATG
KLK_ARBS6_3	ACAAACGAATCCCCTATGCT	KLK_ARBS14_3	TGGAACAACCAAGTTGAAGG
KLK_ARBS6_4	TAAACATCTCAGCAACACTT	KLK_ARBS14_4	GTTCCATAAGTCACTTTTCA
KLK_ARBS7_1	AGAGCCCTCTGTGGGTGGCC	KLK_ARBS15_1	GGACAGTAAGGAAACAGGGC
KLK_ARBS7_2	TGTTGTTCAAGAGCCCTCTG	KLK_ARBS15_2	GGACTATGTCCCACTGTGTT
KLK_ARBS7_3	CTGGCCTGTCTCGTAGAGTC	KLK_ARBS15_3	GGTCAGCAGCGAGCAGGGGA
KLK_ARBS7_4	AATGGATCACGAGCATTAAA	KLK_ARBS15_4	ATGACTTTTTGTTCTACTGA
KLK_ARBS8_1	AAAATATGCTCACACTCATT	KLK_ARBS16_1	AGTTCAGGAGCAAGGGTGGA
KLK_ARBS8_2	AAAATAAGAATAATGAGCAA	KLK_ARBS16_2	AGGCGCAGGGGGCCATGGCT
KLK_ARBS8_3	CAGTACTTCCTTGCATATTT	KLK_ARBS16_3	ACCTGTCCCTGCCGTCATCT
KLK_ARBS8_4	CATATTGACCCAAATATGCA	KLK_ARBS16_4	CTTTCTTGTTTCTCGCTGCC

2.1.6 Antibodies

Table 2.9: Antibodies used in this thesis

Antibody	**Company**	**Catalog No**
Cas9 - Mouse mAB	Cell Signaling Technology	14697
GAPDH - Rabbit pAB	Santa Cruz	25778
PSA - Rabbit pAB	Abcam	ab53774
GFP - Chicken pAB	Abcam	ab13970
α-tubulin - Rabbit mAB	Cell Signalling Technology	21144
Androgen Receptor - Rabbit mAB	Cell Signalling Technology	3202

2.2 Methods

2.2.1 Bacterial growth

2.2.1.1 Competent cell production

(1) Single colony was picked from the previously streaked plate into 2 ml LB broth for overnight culture. (2) 0.5ml LB broth diluted in 100ml new culture and cultured for 3 hours (3) Sample was taken from the culture and optical density (OD)600 value was read using a spectrophotometer (between 0.4 and 0.6). (4) Culture was incubated on ice for 15 minutes and centrifuged for 15 minutes at 3000 rpm. (5) Supernatant was removed and the cell pellet was resuspended in 33ml RF1 buffer (100mM RbCl, 50mM $MnCl_2$-$3H_2O$, 30mM K Acetate, 10mM $CaCl_2$, 15% Glycerol, pH adjusted to 5.8) and incubated for 15 minutes on ice. (6) Supernatant was removed and the cell pellet was resuspended in 8ml RF2 buffer (10mM MOPS, 10mM RbCl, 75mM $CaCl_2$, 15% Glycerol, pH adjusted to 6.8). (7) Mixture was aliquoted and flash froze using liquid nitrogen. (8) Competent cells are stored at 80°C.

2.2.1.2 Bacterial culture conditions

Bacterial cultures have been prepared using LB broth, and appropriate antibiotics (Table 2.10) were added. Cultures were incubated at 37°C at 225 rpm. For each cloned vector, a replica plate has been prepared and glycerol stock has been prepared ($\frac{1}{2}$ 50% glycerol + $\frac{1}{2}$ LB culture).

Table 2.10: Antibiotics and their working concentrations

Antibiotics	**Stock concentration**	**Working concentration**
Ampicillin	100mg/ml	100μg/ml
Carbenicillin	100mg/ml	100μg/ml
Kanamycin	100mg/ml	50μg/ml
Chloramphenicol	50mg/ml	50μg/ml

2.2.2 Molecular biology

ApE (jorgensen.biology.utah.edu/wayned/ape), Snapgene (snapgene.com), and Serial Cloner (serialbasics.free.fr/Serial_Cloner.html) were used for in silico cloning and DNA/protein analysis.

2.2.2.1 Isolation of DNA and RNA

Plasmid DNA, Genomic DNA, and total RNA extraction from the bacterial cultures/human cell lines were performed using the commercial kits of New England Biolabs (Monarch® Plasmid Miniprep - T1010, Monarch® Genomic DNA Purification Kit - T3010, Monarch® Total RNA Miniprep Kit - T2010), Invitrogen (PureLink™ Genomic DNA Mini Kit LSK182001), MACHEREY-NAGEL (NucleoSpin Plasmid Mini - 740588, NucleoBond Xtra Midi - 740410, NucleoBond Xtra Maxi - 740414, NucleoSpin Tissue Mini - 740952, NucleoSpin RNA Mini - 740955), Qiagen (QIAamp DNA Mini - 51304, QIAGEN DNA Midi - 12145, RNeasy Mini Kit - 74104), Zymo Research (Quick-RNA Miniprep - R1054, Zyppy Plasmid Miniprep - D4036, ZR Plasmid Miniprep -

D4015, ZymoPURE II Plasmid Midiprep - D4200). Other than kit-based purification, for the total RNA isolation, TRIzol™ Reagent (Invitrogen 15596026) was used. DNA recovery cleanups from a reaction mixture (e.g. ChIP or 3C), a PCR or an agarose gel were performed using the commercial kits of New England Biolabs (Monarch® DNA Gel Extraction - T1020, Monarch® PCR DNA Cleanup - T1030), MACHEREY-NAGEL (NucleoSpin Gel and PCR Clean-up Mini - 740609), Qiagen (MinElute PCR Purification - 28004, MinElute Reaction Cleanup - 28206).

2.2.2.2 Genomic DNA isolation from 96-well plates

(1) Media was dumped by inverting the plate and excess liquid was removed by blotting on paper towels. (2) Cells are washed with 1X PBS. 150μl PBS was added to each well and then PBS was removed. (3) 50 μl of Bradley Lysis Buffer (10mM Tris-HCl pH 7.5, 10mM EDTA pH 8, 0.5% SDS, 10mM NaCl, H2O) containing Proteinase K (1mg/ml) was added to each well. (4) Lid was replaced and the plate was sealed with parafilm. Plate was put into a humidified chamber overnight at 60ºC. (5) Next day, when plate was cooled to room temperature and 100μl ice-cold EtOH/NaCl (394ml 100% EtOH, 6ml 5M NaCl) added and mixe to precipitate DNA. (6) Plate was incubated at room temperature for 30 min. (7) Plate was placed in a 96-well plate holder centrifuged at 3000 rpm for 20 minutes. (8) Liquid was decant by inverting. Plate was blotted on paper towels to remove excess liquid. (9) 150 μl cold 70% EtOH was added and spun for 10 minutes at 3000 rpm to rinse the pellet. Supernatant was decanted onto paper towels. (10) Washing step was repeated and the DNA pellet was air dried. (11) 30 μl of warmed TE pH 8.0 or elution buffer was added. (12) Plate was covered with wax cover and incubated at 56ºC for approximately 10 min. DNA was quantified.

2.2.2.3 Isolation of nuclei

5x106 cells pellet resuspended in 4.5ml cold permeabilization buffer (10mM TRIS, 10mM NaCl, 0.5% NP-40) + protease inhibitor (PI) cocktail (500μl per 5 ml) and

incubated 30 minutes at 4°C while mixing it gently every 5-10 min or rotate at 4°C.

2.2.2.4 cDNA synthesis

Appropriate amounts of RNA (preferably 1000ng) and water mixture are prepared in 13μl in total. and Mastermix 1 (2.5μl 10mM dNTP (Invitrogen 18427013) and 1μl 50μM random hexamer (Invitrogen N8080127) per reaction) and Mastermix 2 (5μl 5X M-MLV reverse transcriptase buffer (Invitrogen 28025013), 2μl 0.1M DTT (Invitrogen 28025013), 0.5μl RNasin® Ribonuclease Inhibitor (Promega N2111) per reaction) are prepared according to the reaction size. RNA-water mixture and appropriate amount from Mastermix 1 are mixed and they are incubated at 65°C for 5 minutes. An appropriate amount from Mastermix 2 was added to the mixture and it was incubated at room temperature for 10 minutes. 1μl M-MLV Reverse Transcriptase (Invitrogen 28025013) was added to the mixture and it was incubated at 37°C for 1 hour. The mixture then heat-inactivated at 90°C for 10 minutes.

2.2.2.5 Polymerase chain reaction (PCR)

The amplification of DNA from genomic/plasmid DNA was conducted using either a human cell line genomic DNA or plasmid DNA. PCR components are 200 μM dNTP (Invitrogen, 18427013), 0.2 μM forward and reverse primers, (1-100ng) template DNA, 1X GC buffer (NEB, B0519), 1 U Phusion High-Fidelity DNA Polymerase (NEB, M0530), and if the region is GC-rich 1M added to the mixture. For each different primer pair, the annealing temperature would be optimized (55°C, 60°C, 65°C). PCR cycle performed in Veriti^{TM} 96-Well Thermal Cycler is in Figure 2.1.

2.2.2.6 Quantitative polymerase chain reaction (qPCR)

The reaction would be prepared as follows: 2μl cDNA, 10μl 2X SYBR^{TM} Green PCR Master Mix (Thermo Fisher, 4309155), 1μl forward and reverse (0.5μl each), 7μl H2O. qPCR reaction performed using Roche LightCycler® 480 or LightCycler® 96 system. Quantification cycle (Cq) values are calculated using Roche LightCycler® software. Data was analyzed using ΔCq (for 1 replicate) and

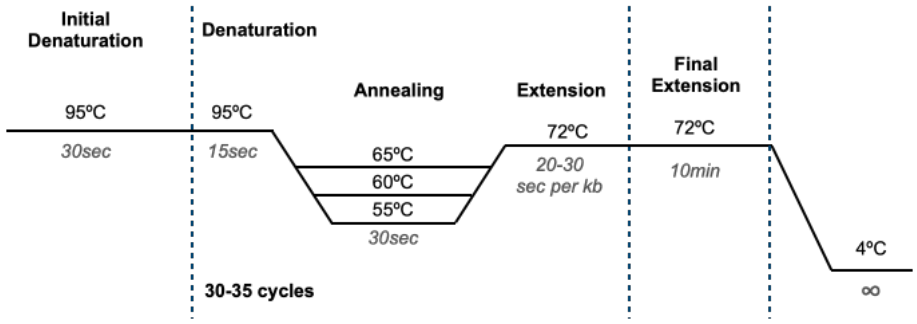


Figure 2.1: Optimization of PCR

ΔΔCq (for >2 biological replicates) methods. qPCR primers are listed in Table 2.11 and Table 2.12.

Table 2.11: qPCR primers

Name	Sequence	Name	Sequence
KLK1_F	GACACCTGGAAGGTGGCAAAGA	KLK10_F	GAGTTGGGTGCTGACGGC
KLK1_R	CATAAGACAGCACTCTGACGGC	KLK10_R	AAGAAGCAGCAGGTGGTCATC
*KLK2*_F	AGCCTGCCAAGATCACAGAT	KLK11_F	TCTGGTGCTCAATGATGGTGA
*KLK2*_R	GCAAGAACTCCTCTGGTTCG	KLK11_R	GGCAGCACGTCCAGC
KLK3_F	GATGCTGTGAAGGTCATGGA	KLK12_F	GCTGTGAGTTACGCCCACACT
KLK3_R	TGGAGGTCCACACACTGAAG	KLK12_R	CATCTTTTTGCTCCTGTGTGTTCTT
KLK4_F	GGCACTGGTCATGGAAAACGA	KLK13_F	GAGACGATGCCATACAGTGTTCTG
KLK4_R	TCAAGACTGTGCAGGCCCAGCC	KLK13_R	GAGACGATGCCATACAGTGTTCTG
KLK5_F	AGATGACACCATGTTCTGCGC	KLK14_F	AGTGTCAGGCTGGGGAACTA
KLK5_R	CACAGGCCCCCCAGAATC	KLK14_R	CCCAGAGTCACCCTGACAAG
KLK6_F	TGGTGCTGAGTCTGATTGCT	*KLK15*_F	TAACTGTGGCGCTTCCCTCATC
KLK6_R	CGCCATGCACCAACTTATT	*KLK15*_R	AGACGTGGTCCGTAGTTGCTCT
KLK7_F	CATCCCCGACTCCAAGAAAA	*FKBP5*_F	TCCCTCGAATGCAACTCTCT
KLK7_R	GCAGGGTACCTCTGCACACC	*FKBP5*_R	GCCACATCTCTGCAGTCAAA
KLK8_F	AAGAGCAGGAACATCCACGTCT	*STEAP4*_F	GCAAAGCATCCAGTGGTCAA
KLK8_R	GAATCAGTAGGTGACCCCGC	*STEAP4*_R	ATGACAGCAAAGCCAAGCAA
KLK9_F	AGTCACCCTGGCAGGAACCT	*DBI*_F	CAGGCGGAGTTTGAGAAAGCT
KLK9_R	AGTCACCCTGGCAGGAACCT	*DBI*_R	TCGCTTGTTTGTAGTGGCTGTAG

Table 2.12: ARBS eRNA qPCR primers

Name	Sequence
ARBS1_eRNA_F	GGGAAGCATCGAGAGCATGA
ARBS1_eRNA_R	TGCTTGACTTTGAAACTCGGA
ARBS2_eRNA_F	AGTCTAGGTGGATGCTGTGC
ARBS2_eRNA_R	TCTCCATGGTTCTGTCACGT
ARBS3_eRNA_F	GGGTTTGTGGCAAAGGTGAG
ARBS3_eRNA_R	GGAAGGCTCTGGGTGAACA
ARBS4_eRNA_F	GTTGCTGTTCATCCTGAGCC
ARBS4_eRNA_R	ACAAGGCGATGGAGAGAACC
ARBS5_eRNA_F	CCTTGGTCCCTCTTCTGTCA
ARBS5_eRNA_R	CTGTGGGTCAAAGCATCGTG

2.2.2.7 Agarose gel electrophoresis

Agarose gels are prepared using Invitrogen UltraPure^{TM} Agarose (16500100) and 1X TAE (40mM Tris, 20mM Acetate, 1mM EDTA, pH 8.6) with 0.5 μg/ml Ethidium UltraPure^{TM} Ethidium Bromide (15585011, 10 mg/mL). For casting 1% agarose gel, 100ml 1X TAE, and 1g agarose were used. An appropriate amount of 6X DNA Gel Loading Dye (R0611) or 6X Purple Gel Loading Dye (B7024S) was mixed with the DNA mixture and loaded into the gel well and electrophoresis was performed using BioRad's horizontal electrophoresis system (https://www.bio-rad.com/en-ca/category/horizontal-electrophoresis-systems).
One of the following DNA ladder was used: 1 kb Plus DNA Ladder (Thermo Scientific, 10787018), 1 kb DNA Ladder (NEB, N3232), 50 bp DNA Ladder (Thermo Scientific, 10416014), 100 bp DNA Ladder (Thermo Scientific, 15628019) to identify the DNA size.

2.2.2.8 T7 Endonuclease Assay

Genomic DNA isolated from wild type and transduced/mutated samples were used to amplify the region of interest. The reaction was PCR cleaned and gel extracted. Re-hybridize the product by increasing the temperature to 95°C and decreasing to 4°C with a ramp rate of 0.3°C/s. Knockout efficiency was determined with an agarose gel.

2.2.2.9 Single and paired gRNA cloning

gRNA cloning into plentiCRISPRv2 (Addgene 52961) or plentiGUIDE-puro (Addgene 52963) vectors has been performed as follows: (1) Plasmid was digested 5ug with 3μl BsmBI (10U/μl) for 3-5 hours at 55°C (2) Product was heat-inactivated at 80°C for 20 minutes, 1μl Antarctic phosphatase (5U/μl) was added (3) Digested plasmid gel purified (4) Each pair of oligos were phosphorylated and annealed: 37°C 30 min and 95°C 5 min and then ramped down to 25°C at 5°C/min. (5) Oligos from Step 4 diluted at a 1:200 dilution into ddH2O. (6) Ligation reaction using Quick Ligase was set up, and it is incubated at room temperature for >1-3 hours (7) Transformation to carbenicillin+ LB agar plates. (8) Validate the positive clones with BsmBI and Sanger sequencing with hu6_f (GGCAAGTTTGTGGAATTGGT).

Paired guide RNA cloning protocol adapted from [250]. Donor plasmids are pDonor_sU6 (Addgene 69351), pDonor_mU6 (Addgene 69350), pDonor_hU6 (Addgene 69312). Two step-paired guide gRNA oligos are listed in Table 2.13: (1)

Table 2.13: Oligos for the paired-gRNA cloning

Name	**Sequence**
Oligo_hU6	CTTGTGGAAAGGACGAAACACCG-[gRNA1-20bp]-GTTTGGGTCTTCGAGAAGACCTCACCG-[gRNA2-20bp]-GTTTTAGAGCTAGAAATAGC
Oligo_s/mU6	CATGCGAGAAAAGCCTTGTTTGG-[gRNA1-20bp]-GTTTGGGTCTTCGAGAAGACCTCACCG-[gRNA2-20bp]-GTTTTAGAGCTAGAAATAGC

For each oligo (paired gRNA), PCR amplified it using Oligo_mu6. (2) pDonor_mU6 is digested with BbsI, loaded to 2% agarose gel, and extracted from the 415bp band. (3) Gibson reaction set up with the PCR amplicon and digested donor and PCR cleanup the reaction. (4) plentiCRISPRv2 digested with BsmBI and Gibson product ligated with digested vector. (5) Transformed using Stbl3 and validated the colonies with NotI/XhoI digestion and Sanger sequencing using hU6_f (GGCAAGTTTGTGGAATTGGT).

2.2.2.10 Restriction digestion cloning

Insert is obtained either from PCR amplification of the target using genomic DNA, a plasmid DNA, and cDNA, or it is from any plasmid target digested with restriction enzymes. Backbone of choice is digested with appropriate enzyme(s). Different insert/backbone ligation ratios are tested and "only backbone" and "only backbone + ligase" control reactions are prepared.

2.2.3 Flow cytometry and cell sorting

Cell line samples are lifted by trypsinization and centrifuged. Cell pellets are resuspended using 5%FBS/PBS solution. In case of cell sorting %1 penicillin-streptomycin added. Control samples are used to gate the right cell population using forward and side scatter height. For each gated sample at least 10,000 cells recorded for further analysis. eGFP is detected at 488 nm and mCherry signal is detected at 550 nm.

2.2.4 Protein methods

2.2.4.1 Protein lysate preparation

Cells collected from a 15cm plate and washed with 1X PBS. 1 protease inhibitor was dissolved in a 5ml Radioimmunoprecipitation assay buffer (RIPA) Lysis buffer (150 mM NaCl, 1% Nonidet P-40, 0.5 % DOC, 0.1% SDS, 50 mM Tris pH 7.4). Cells resuspended in 200μl RIPA Lysis Buffer (protease inhibitor included) and incubate on ice for 30 minutes. Mixture was centrifuged for 20 minutes @4°C 12.000 rpm. Supernatant was transferred to a new tube. Lysates can be stored at -20°C for short term and at -80°C for long term storage.

2.2.4.2 Protein concentration determination

Pierce^{TM} Coomassie Plus (Bradford) Assay Reagent was mixed (23238) by inverting the bottle and equilibrating the necessary volume to room temperature in aluminum covered 50 ml Falcon tube as the Commasie Reagent is light sensitive. 9 different dilutions of Bovine Serum Albumin with 1X PBS were prepared (0, 25,125, 250,

500, 750, 1000, 1500, 2000 μg/ml). 5 μl of the standard solutions and the samples in triplicate was added to each well of a 96 well-plate. 250 μl of Commasie Reagent was added to each well. Plate was shaken for 30 seconds and incubated at room temperature for 10 minutes. Absorbance was read at 595 nm in MicroPlate Reader (Use the Bradford protocol). Analysis: Standard curve was plotted as the average reads of the standards on the y-axis and concentrations on the x-axis. Protein concentration of the lysate determined using the formula of the linear plot (y=ax+b).

2.2.4.3 SDS-PAGE and Western blot

10% separating (0.75mm) Sodium Dodecyl Sulfate Polyacrylamide gel electrophoresis (SDS-PAGE) gel was prepared using following reagents (per gel): 1.9ml distilled H2O, 1.3ml 1.5 M Tris pH 8.8, 1.7ml 30% acrylamide, 50μl 10% SDS, 50μl 10% ammonium persulfate (APS), 2μl tetramethylethylenediamine (TEMED). 5% stacking gel was prepared using following reagents (per gel): 1.4ml distilled H2O, 250μl 1M Tris pH 6.8, 330μl 30% acrylamide, 20μl 10% SDS, 20μl 10% APS, 2μl TEMED (added last). Buffers: 10X Tris Buffered Saline pH 7.4 (200mM Tris Base, 1.5M NaCl), Running Buffer (BioRad, 1X TRIS/Glycine/SDS Buffer), Wash Buffer (TBS-T; 1X TBS, 0.1% Tween-20), Blocking Buffer (1X TBS, 0.1% Tween-20 with 5 % w/v nonfat dry milk), Transfer Buffer (25 mM Tris Base, 192 mM glycine, 10% methanol). 2X loading dye is prepared with 5% ß-mercaptoethanol in the Laemmli sample buffer. 10-30μg cellular lysate mixed with 2x loading buffer and incubated at 95°C for 10 minutes. Sample was run in SDS-PAGE gel at constant 25mAmp until it nearly reached the end of the gel. Transfer stacks were rinsed in the transfer buffer and put into the cassettes. Polyvinylidene fluoride (PVDF) membranes rinsed in the methanol and transfer buffer placed on top of stacks and SDS-PAGE gel was placed on top of the membrane. Whole stock was closed with transfer stacks and proteins were transferred to the membrane with semi-dry blotting (BioRad TransBlot Turbo System). After transfer, the blotted membrane was washed with 1X TBS and blocked with the blocking buffer at room temperature for 2 hours on the shaker.

Appropriate amount of primary antibody is diluted in the blocking buffer (1/3000 for AR, 1/2000 for GAPDH antibody). Membrane was incubated with primary antibody overnight at 4°C on a shaker. Next day, it was washed with TBS-T at least 5-6 times for 5 minutes each on a shaker. At room temperature, the membrane was incubated with the secondary antibody (diluted 1/4000 in blocking buffer) for 1 hour. Membrane was washed with TBS-T at least 5-6 times for 5 minutes each on a shaker. After applying Pierce™ ECL Western Blotting Substrate onto the membrane it was imaged with Odyssey® Fc (LI-COR Biosciences).

2.2.5 Cell culture

LNCaP and all of the LNCaP stable cell lines are cultured in RPMI-1640 (Gibco 12633012) supplemented with 10% fetal bovine serum (Gibco 16140071) and 1% penicillin-streptomycin (Gibco 15140148). HEK293T cells are cultured in Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (Gibco 16140071) and 1% penicillin-streptomycin (Gibco 15140148). All cell lines were grown in 37 C humidified incubator with 5% CO2. They were passaged by trypsinization (0.05% Trypsin, Thermo Fisher, cat no: 25300054). Frozen stocks are prepared by mixing the culture media with minimum 1x106 cells with 10% fetal bovine serum and 10% DMSO. The mixture was transferred to cryovial tubes, and they were placed in Mr. Frosty Freezing Container (Nalgene) at -80°C overnight. They were transferred to the liquid nitrogen tank for long-term storage.

2.2.5.1 Cell culture conditions

2.2.5.2 Viral packaging and PEG precipitation

10cm/6-well plate: $4.4x10^5$/$3x10^5$ 293T cells per well in 6-well plate is seeded. On Day 1, two tubes of a mixture (Fugene mixture: 20μl/4μl Fugene, 180μl/36μl DMEM, DNA mixture: 225ng/50ng VSVG, 2250ng/450ng gag/pol, 2500ng/500ng vector, and DMEM up to 40 μl) was prepared. Two mixes were added together,

and after 30 minutes of incubation at room temperature, it was added to the wells dropwise. On Day 2 (in the morning): Fugene was washed off with PBS. Media was aspirated from each plate and replaced with fresh 8ml/2ml DMEM. On Day 3 (after about 48 hours post-transfection): Supernatant was collected into the appropriate size centrifuge tube. After removing the media, it was replaced with 8ml/2ml DMEM. On Day 4, collection steps were repeated and combined with supernatants from day 3, then they were filtered (0.45μm filters), aliquoted, and made frozen stocks. Packaged viral particles are stored at -80°C. PEG precipitation: 50% PEG solution was added to the 50ml conical tube containing the filtered virus (final concentration: 10%) and mixed by inversion. The mixture was stored overnight at 4°C in the dark. The next day, Precipitated viral particles should be visible in the tube. Tubes spun at 2500 rpm for 20 minutes. The supernatant was removed, and the tube spun again for 5 minutes at 1200 rpm. After removing the excess liquid, 100 uL of PBS was added to the pellet and resuspended the pellet by pipetting. After collecting all the viruses (100 uL), they are aliquoted and stored at –80°C.

2.2.5.3 Viral titration

2x10^5 HEK293T cells seeded in 5 wells of 6 well-plate. Three wells were transduced with different volumes of the virus. (Unconcentrated 10, 100, 1000 μl, Concentrated with PEG 0.1, 1, 10 μl). The media changed the next day. Antibiotic was added according to the infection's antibiotic resistance gene to 4 wells, including one non-transduced well. When antibiotic kills all the cells in non transduced well, the 3 transduced + selected wells and 1 non transduced + not selected well were trypsinized. Cells were counted by hemocytometer. The volume of virus' infection MOI was estimated by using the below formula and table:

$$100 - \left(\frac{\text{transduced and selected cell count}}{\text{nontransduced and not selected cell count}} x100 \right) = \text{virus titer}$$

Starting cell population multiplied with each well's MOI:

$$\text{MOI x } 200.000 = \text{total virus amount}$$

Infected total virus amount and the average volume of virus used was averaged, and the formula was used to estimate virus titer: $\frac{average(\text{total virus amount})}{average(\text{volume of virus})} x1000 = \text{virus titer}$

Table 2.14: MOI calculation.

MOI	%selected cells	MOI	%selected cells
0.1	90.5	0.9	40.6
0.2	81.9	1	36.8
0.3	74.1	2	13.5
0.4	67.0	3	5.0
0.5	60.6	4	1.8
0.6	54.9	5	0.7
0.7	49.6	10	0.0
0.8	45.0		

2.2.5.4 Stable transduction

The day before the experiment, cells are plated in a 6-well plate. Media was warmed upfront, and media was changed before the transduction. 2 μl of protamine sulfate (Stock concentration: 8 μg/ml) was added (which serves as a crosslinker for viral transduction). Once the virus is thawed completely, added dropwise. After 24 hours, the media was changed. If the selection was performed, it started 48 hours after the transduction.

2.2.5.5 Transfection

An appropriate amount of plate and number of cells are plated and incubated overnight. Optimem, Mirus XT2, and growth media warmed upfront. DNA/optimem/Mirus XT2 mixed, respectively. Incubated at room temperature for 30 minutes and added dropwise. eGFP expressing plasmid (Addgene 17448) has been used for the transfection control.

Table 2.15: Parameters for the transfection

Plate	**Cell**	**DNA**	**Mirus XT2**	**Optimem**
12-well	0.1x10^6/well	1μg	3μl	100μl
6-well	0.2x10^6/well	2.5μg	7.5μl	250μl
10 cm	2x10^6	15μg	45μl	1.5ml
15 cm	6x10^6	19μg	57μl	1.9ml

2.2.5.6 Single-cell dilution

100 μl PBS to the edges of 96-well plate added (AHth row and 1st12nd column). 100 μl medium to the remaining wells except 2nd column added. 10,000 cell/ml counted. 200 μl of the cell suspension to the 2nd column added. 100 μl of cell suspension was taken from (n)th column and added to the (n+1)th column; cell suspension was homogenized by doing up-down. (0<n<10). 100 μl of cell suspension taken from the 11th column so that at the end, there was 100 μl of cell suspension in all wells. 100 μl of media was added to all of the wells containing cell suspension. Plate incubated for 2 weeks until the colonies appear and reach enough cell density.

2.2.5.7 Generation of CRISPR/Cas9 eGFP knock-in cell lines

Knock-in cell lines are generated with non-homologous end joining [251]. Donor (Addgene plasmid 83575) plasmid consists of IRES-eGFP cassette flunked by two sequences that can be targeted with a separate sgA plasmid. LNCaP cell lines are transfected with donor, sgA, target gRNA, sgA, and Cas9 expressing plasmid, total amount is 1.6μg (donor 0.6μg + sgA 0.2μg + target gRNA 0.2μg + Cas9 0.6μg) for 12-well plate. After antibiotic selection, cells are collected and plated in a 96-well plate with single-cell dilution protocol. Single cell clones that express eGFP upon DHT treatment (10nM for 6 hours) were selected under the fluorescent microscope. After propagating the single cell clones into 6-well plates, they were separated into three (1) ongoing passaging (2) gDNA extraction (3) flow cytometry. Knock-in was validated with PCR and flow cytometry. Single cell clones are shown in Figure A.2.

2.2.5.8 Generation of CRISPR/Cas9 stable knock-out cell lines

At least 3 gRNAs per perturbation target are used. If the aim was to achieve deletion, 2 gRNAs far apart enough (>100-150bp) were used. Either Cas9/gRNA expressing plasmids or individual Cas9 and gRNA plasmid combinations were used. Cas9/gRNA ratio was >=1. gRNAs either manually designed or CHOPCHOP [252], CRISPR-SURF [253] have been used. Cas-OFFinder [254] has been used to eliminate designed gRNAs with prominent off-target sites. gRNA efficiency was determined with T7 Endonuclease assay. Cells were seeded at $2x10^5$ cells per well in six-well plates. Cells were transfected with 1μg of Cas9/gRNA expressing plasmid using Mirus (1:4 ratio). Approximately 48 h after transfection, cells were treated with appropriate antibiotics. After selection, cells were trypsinized and counted. Starting from the second column of a 96-well plate, cells were seeded at a density of 1000 cells per well and 1:1 diluted in subsequent wells. To prevent evaporation of the medium, perimeter wells of the plate were loaded with 200μl PBS. After 2-4 weeks, individual colonies were isolated and seeded into wells in 48-well plates. As they reached 50-70% confluency, cells were lifted and seeded to 24, 12, and 6-well plates subsequently. In the 6-well state, cells were divided into three (1 for backup, 1 for gDNA extraction, and 1 for continuing the passage). After gDNA extraction, (see the relevant materials and methods section), it was used to amplify the region of interest (gRNA target site). If the aim is to achieve a specific deletion (>100-150bp), after amplification, a comparison with a wild-type cell line was performed to assess the existence of the perturbation. However, Sanger or low-pass amplicon sequencing (50K-100K reads) gave the absolute result for the perturbation validation. Low-pass sequencing analysis has been done using CRISPResso [255]. Preferably, in the 6-well state, a phenotypic assessment can be performed on single-cell clones before genotypic confirmation. PCR oligos used to amplify KLK ARBS are listed in Table 2.16.

Table 2.16: Oligos to PCR amplify KLK ARBS for sequencing.

Name	Sequence
ARBS1_F	GTAGGGAAGCATCGAGAGCA
ARBS1_R	ACACAGTGGTTTGCGTCAAT
ARBS2_F	TTGTGCCACTGGTGAGAAAC
ARBS2_R	TGAGCAAAGACAGCAACACC
ARBS5_F	CTACCTCCCACCCTTCCATC
ARBS5_R	AGAATGTGTCTCCCTCTCGG

2.2.5.9 Generation of paired gRNA reporter efficiency cell lines

plentiCMV-eGFP (Addgene #17448) vector was digested with BamHI (NEB R0136) and using an adapter sequence two g RNA (sgA and sgA2) target sites are ligated along with PacI and XmaI restriction diesgtion sites. loxP-3xPolyA-loxP cassette was PCR amplified from pCAG-loxPSTOPloxP-ZsGreen vector (Addgene #51269) and digested with PacI (NEB R0547) and XmaI (NEB R0180). CMV-sgA-sgA2-eGFP and loxP-3xPolyA-loxP were ligated and the correct colony was validated with Sanger sequencing. Oligos used are listed in Table 2.17. Experiment results are shown in Figure A.5). Transduced the cells with different gRNAs and one paired-guide RNA to create double-stranded break (DSB) in order to delete regions when both of the regions are targeted. Analysis performed with flow cytometry.

Table 2.17: Primers used in paired guide RNA efficiency reporter

Name	Sequence
cutsite_adaptor_F	GATCaaaTTGCGAGTGTTATACGCAC CGTTAGGaaaTTAATTAAaaaaaaCCCG GGaaaTTGCAACGGGTTCTCCCGGCTACAGGaaa
cutsite_adaptor_R	GATCtttCCTGTAGCCGGGAGAACCC GTTGCAAtttCCCGGGttttttTTAATTAA tttCCTAACGGTGCGTATAACACTCGCAAttt
sgA	CGAGTGTTATACGCACCGTT
sgA2	CAACGGGTTCTCCCGGCTAC
loxP-3xPolyA-loxP_F	TCCTTAATTAATAATACGACTCACTATAGGG
loxP-3xPolyA-loxP_R	tttCCCGGGtttGCCGGCCGAATTCGATCT

2.2.6 ChIP

Unless otherwise is stated, all of the steps have been performed on ice. One confluent 15 cm plate is used per cell sample. Required buffers are prepared listed in Table 2.18. (1) Protein A or Protein G beads are prepared one day before preparing the cell samples. 100μl per immunoprecipitation (IP) sample is transferred to 2mL round bottom (Eppendorf LoBind) tubes. With the magnetic rack, beads are separated from their home solution (30 seconds waiting time). The supernatant was removed, and beads were washed with the blocking buffer three times (in The supernatant step, beads were correctly resuspended). Beads were resuspended in 500μl blocking buffer, and 10μg of antibody was added to the mixture. Tubes are incubated at 4^{o}C (cold room) at 10 rpm overnight. (2) Cells are washed with PBS, and 1% formaldehyde is added (10ml in total) to the plate. Plates are incubated at room temperature for 10 minutes. Crosslinking reaction was quenched with 1M glycine (pH 7.5). Formaldehyde was removed from the plate, and the sample was washed with ice-cold PBS with protease inhibitors. After the washing, 500μl ice-cold PBS with protease inhibitors was added, and cells were scraped off the plate and transferred to 1.5ml tubes. Samples were centrifuged at 2000g for 3 minutes at 4^{o}C. The washing step was done twice. (3) 1ml LB1 with protease inhibitors added to the cell pellet and rotated at 4^{o}C for 10 minutes. After centrifuging at 2000g for 5 minutes at 4^{o}C, 1ml LB2 with protease inhibitors was added to the pellet and spun at 4^{o}C for 5 minutes. After centrifuging at 2000g for 5 minutes, the pellet was resuspended in 300μl LB3 with protease inhibitors. (4) Samples were sonicated on high power for 15 minutes (30 seconds on, 30 seconds off). A 10μl aliquot sonicated sample was loaded onto an agarose gel after decrosslinking to validate the fragmentation (Most of the fragments should be below 500bp). 30μl 10% Triton X-100 was added to the sample and centrifuged at 20000g for 10 minutes at 4^{o}C. Antibody-bead mixture washed three times with blocking buffer and finally resuspended in 100μl blocking buffer. 1ml LB3 (with protease inhibitors) added to sonicated and validated lysate. 50μl of this mixture has been set aside for input control. Antibody-bead mixture

was added to the cell lysate. The mixture was rotated at 4^{o}C overnight. (5) Sample was placed on a magnetic rack to collect antibody-DNA bound beads. They are washed with RIPA buffer six times. Finally, they are washed with TE buffer. 200μl elution buffer was added and vortexed. 150μl elution buffer was also added to the input sample. Both ChIP and input samples are incubated at 65^{o}C. (6) Sample was placed on a magnetic rack, and the supernatant was transferred to a new 1.5ml tube. 200μl TE buffer was added to the mixture. 8μl RNase A was added and incubated at 37^{o}C for 10 minutes. Subsequently, 4μl Proteinase K was added and incubated at 55^{o}C for 2 hours. (7) DNA was isolated from the samples using a column-based isolation kit (Qiagen MinElute Reaction Cleanup Kit).

2.2.7 scATACseq

LNCaP cells were androgen deprived for 72 hours and then treated with 100% EtOH or 10 nM DHT for 4 hours. Nuclei from 1x10^6 cells were isolated according to the 10x Genomics protocol (Nuclei Isolation for Single Cell ATAC Sequencing CG000169 Rev D). scATACseq libraries were prepared using 10x Genomics Chromium Next GEM Single Cell ATAC Library Gel Bead Kit v1.1. Libraries were sequenced on a NextSeq500 Illumina sequencer (23,428 unique PE reads/cell DHT and 27,077 unique PE reads/cell EtOH)

2.2.8 CRISPRi

LNCaP dCas9-KRAB cells are generated by transducing LNCaP cells with Lenti-dCas9-KRAB (Addgene 89567) plasmid, and the expression is validated by western blot (antibodies: Cas9 and GAPDH). LNCaP-dCas9-KRAB cells are seeded into 6-well plates (250,000 cells per well). Controls are (1) No Transfection (EtOH) (2) No Transfection (DHT) (3) eGFP (transfection) (4) No Transfection (puromycin) (5) non-targeting gRNA (EtOH) (6) non-targeting gRNA (DHT). (7) 3 gRNAs per condition have been used. gRNAs transfected using Mirus Transit X-T2 (total 1μg DNA). After 48 hours, puromycin was added for the selection. After the selection,

cells were androgen deprived for 24 hours, and after that, 1nM DHT was added for 24 hours. (8) Cells are collected. RNA is extracted and cDNA synthesized using random hexamers.

Table 2.18: Buffers used in ChIP experiments

Buffer Name	**Components**
Lysis Buffer 1 (LB1)	50 mM HEPES–KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10 % glycerol, 0.5 % Igepal CA-630, 0.25 % Triton X-10
Lysis Buffer 2 (LB2)	10 mM Tris–HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGT
Lysis Buffer 3 (LB3)	10 mM Tris–HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % Na-deoxycholate, 0.5 % N-lauroylsarcosine
Blocking Buffer	PBS, 0.5 % bovine serum albumin (BSA)(w/v)
RIPA	50 mM HEPES–KOH, pH 7.5, 500 mM LiCl, 1 mMEDTA, 1 % NP-40 or Igepal CA-630, 0.7 % Na-deoxycholate
TE	10 mM Tris-HCl, pH 8.0, 1 mM EDT
Elution	50 mM Tris–HCl, pH 8.0, 10 mM EDTA, 1% SDS

Table 2.19: qPCR primers used in AR ChIP

Name	**Sequence**
*KLK2*_promoter_qPCR_F	AGCCAGACCCTCTCTCTACA
*KLK2*_promoter_qPCR_R	CCTGTAGCCTTGAACTCCCA
KLK_ARBS1_qPCR_F	ATACAACATGTCCTTCCGGG
KLK_ARBS1_qPCR_R	TGCTTGACTTTGAAACTCGGA
KLK_ARBS2_qPCR_F	TGGATTGAAAACAGACCTACTCT
KLK_ARBS2_qPCR_R	AGGCTTGCTTACTGTCCTAGA
KLK_ARBS3_qPCR_F	GGTTGAAAGCAGACCTACTCT
KLK_ARBS3_qPCR_R	GCCTTGCAAGATGGTATCGC
KLK_ARBS4_qPCR_F	CTCCCCTTCCACAGCTCTG
KLK_ARBS4_qPCR_R	AGCCTTCATTCTCCAGGACC
KLK_ARBS5_qPCR_F	CACATCAGTTCGCCTTGACC
KLK_ARBS5_qPCR_R	CTGTGGGTCAAAGCATCGTG

2.2.9 Chromosome conformation capture

5x10^6 cells per 15ml falcon tube are prepared. Cell pellets are crosslinked with 1% formaldehyde at room temperature for 10 minutes and reaction was quenched with 0.125M glycine. After wash with ice-cold 1X PBS, centrifuged (5 min, 600 g at 4°C) pellet was resuspended in ice-cold permeabilization buffer (10mM TRIS pH 8, 10mM NaCl, 0.5% NP-40, 1 complete protease inhibitor Roche 11245200) and incubated on ice for 30 minutes. Centrifuged pellets (5 min, 600 g at 4°C) resuspended in DpnII (NEB R0543) restriction buffer and 10% SDS and 20% TritonX-100 added. 600U DpnII was added and the reaction was incubated overnight at 37°C while shaking at 900 rpm. Digestion efficiency was determined by agarose gel electrophoresis. Sample was transferred to 50mL Falcon tube and BSA (20mg/ml) and ligase buffer and 10,000U T4 DNA Ligase (NEB M0202S) was added and the reaction was incubated overnight at 16°C. Ligation efficiency was determined by agarose gel electrophoresis. Reaction was reverse crosslinked with Proteinase K (10 mg/ml) and RNase A (10 mg/ml) treatment. DNA was recovered either with phenol-chloroform extraction or isolation kit (Qiagen MinElute Reaction Cleanup Kit). Primers can be found in Table 2.20.

2.2.10 Computational analyses

Unless stated otherwise, all data cleaning and reorganizing performed using tidyverse R packages [256] and all plots were prepared using the ggplot2 R package [257].

2.2.10.1 STARRseq

Short Illumina paired-end reads (75-150 bp) were mapped to the human reference genome (hg19) using BWA [258]. After MAPQ<60 filtration of the mapped reads, coverage was counted with BamCoverage in deepTools [259] and with DESeq2 [260]. Different enhancer classes were classified using this criteria: Inducible enhancers are log2 fold change (LFC) (DHT/EtOH) > 1, inactive enhancers LFC < 1 and constitutively active enhancers have $1 >$ LFC > 1 in both of the plasmid-normalized samples.

Table 2.20: 3C-qPCR primers

Name	Sequence	Coordinates (hg19)
Anchor_AREIII	TCGATTGTCCTTGACAGTAAACA	chr19:51354109-51354131
KLK3_1	TTATCCCTGGTTCCTAGCAC	chr19:51359093-51359112
KLK3_2	CCCTGTTTCTGTTTCATCCT	chr19:51357896-51357915
KLK3_3	CACGTGAGGCTTTGTATGAA	chr19:51357855-51357874
KLK3_4	TCCTGAGTGCTGGTGTCTTA	chr19:51357800-51357819
KLK3_5	GCAATGTATCTGTGGAGCTG	chr19:51357653-51357672
KLK3_6	GATCTGTGGACCACAA	chr19:51357616-51357631
KLK3_7	GCACAGATAGAGTGCACAGG	chr19:51357490-51357509
*KLK2*_A	GGCTCCAGAGGACCATGATT	chr19:51372228-51372247
*KLK2*_B	GGTACCAGCAAACTTGTCCAG	chr19:51372792-51372812
*KLK2*_C	GACGTCATGTAGCTGCGACT	chr19:51373843-51373862
*KLK2*_D	ACCCTTGACCTCCTGGACTT	chr19:51375305-51375324
*KLK2*_E	TGCCCATTGCCTAAAGAAGT	chr19:51378119-51378138
*KLK2*_F	AGATAAGGCCGTGAGCAGAA	chr19:51382312-51382331
*KLK2*_G	CCCCTCGTGGTTAACATCAC	chr19:51385342-51385361
*KLK2*_H	TGGTTCCCACTACATCCACA	chr19:51386600-51386619
*KLK2*_I	GCTGGAGGTGAGTTCTTTGG	chr19:51387189-51387208

2.2.10.2 scATACseq

Raw .fastq files were obtained from 10X genomics platform. Cell barcodes and counts are calculated using 10X Cell Ranger software. Deviation/z-score was calculated with chromVAR [261] and UMAP was calculated in ArchR (v0.9.3) [262]. ATACseq peaks were called using MACS (v2.2.6) [263]. Random genomic regions were generated using bedtools (v2.29.2) [264]. scATACseq co-accessibility scores (Figure 2.2) in EtOH- and DHT-treated LNCaP scores were calculated using Cicero (v1.4.0) [143] with 10X Genomics input. Co-accessible sites overlapping with STARRseq annotated AR enhancers and LNCaP ARBS (GEO:GSM2219854) were calculated with bedtools(v2.29.2) [264] and counted with R (v3.6.1).

2.2.10.3 HiC and ChIA-PET

Publicly available HiC datasets are analyzed with HiC-Pro [[265] .bedpe files are further analyzed and plotted with Sushi [266] and HiTC [267]. VCaP AR

ChIA-PET [268] and LNCaP H3K27ac HiChIP [269] datasets obtained from authors as .bedpe files. Genomic overlaps were calculated either with bedtools [264] or GenomicInteractions R package [270].

2.2.10.4 ChIPseq

Short Illumina paired-end reads (75-150 bp) were mapped to the human reference genome (hg19) using BWA [258]. deepTools [259] was used for reads per kilobase of transcript (RPKM) normalization and for the generation of the .BigWig files.

2.2.10.5 RNAseq

Short Illumina paired-end reads (75-150 bp) were mapped to the human reference genome (hg19) using BWA [258]. Gene count tables were constructed using featureCounts [271]. AR-regulated genes were identified from publicly available RNAseq data of LNCaP cells treated with 100 nM DHT for 6 h (GSE64529) with DESeq2 (v1.26.0) [260]. The distance between androgen-upregulated genes and ARBS was calculated with HOMER (v4.10) [272]. The resulting data was merged in 100-bp bins and the cumulative distribution function was determined with *Ecdf()* in R (v3.6.1). Specific ARBS-promoter interactions were identified from the published AR ChIA-PET performed in the VCaP cell line (GSE54946) by overlapping the loop end of either ARBS or gene's TSS (+/- 5Kb) with GenomicInteractions R package (v1.20.3) [270].

Chapter 3

ANDROGEN RECEPTOR-BOUND ENHANCERS

3.1 Introduction

Androgen receptor (AR) is the critical transcription factor in PCa growth and proliferation. Upon testosterone binding, AR translocates into the nucleus, where it binds to CREs that regulate the expression of target genes. However, while the mechanism of AR activation is well characterized, how these ARBS alter gene expression is poorly understood. While the AR drives transcription through enhancer CREs, this is complicated by the sheer number of ARBS and the vast difference in the number of binding sites (tens of thousands) and AR-regulated genes (hundreds). It is unclear if or how multiple ARBS enhancer elements work together to drive transcription. Understanding this complex process has been limited due to the poor annotation of ARBS enhancer activity. Active enhancers generally correlate with histone marks such as H3K27ac and H4K4me1. They also are found in open chromatin and transcribe enhancer RNA (eRNA). However, these descriptive features are only broadly correlative with active enhancers, and numerous enhancers do not have any of these characteristics. Finally, these correlative features are not quantitative. Functional characterization is critical to identify and understand AR-mediated enhancer activity. Classifying the AR-bound enhancers that directly affect the AR-mediated gene transcription and their action in gene expression is key to revealing the initiation and progression of this disease. Therefore, in this chapter, we aimed to test all of the clinically found ARBS for their enhancer activity with an MPRA. We characterized these different AR enhancers for their impact on gene activation with these annotations. Finally, we demonstrated that inducible AR enhancers act as a regulatory hub that frequently

interacts with other CREs.

3.2 Results

The STARRseq experiment and the analysis were performed by Flora Huang and Tunç Morova in the Vancouver Prostate Centre, University of British Columbia. LNCaP transduction with dCas9-KRAB was done by Bengül Gökbayrak in Koç University. scATACseq libraries were prepared by Funda Şar in Vancouver Prostate Centre, University of British Columbia.

3.2.1 Quantification of ARBS in proximity of AR-regulated genes

While looping can occur between vast genomic distances to bring enhancers near promoters, a recent report demonstrated that the closest gene is a significant predictor of a given enhancer's target [95]. Earlier studies showed that 28% of AR binding sites and 7% of ER binding sites reside in 5kb proximity to a TSS [273, 274]. To quantify this in AR-mediated transcription, upon androgen activation, we determine how many ARBS reside near AR-regulated genes in LNCaP cells using publicly available RNAseq (GSE64529) and AR ChIPseq (GSE83860) datasets. From this, we found that more than 40% of the AR-regulated genes have more than two ARBS within 100kb proximity and more than 60% of the AR-regulated genes have more than two ARBS reside in 200kb proximity (Figure 3.1). This is similar to estrogen receptor (ER), where almost 50% of ER-upregulated genes have two or more ERBS within 100kb proximity [275]. Down-regulated and non-regulated genes less frequently have proximal ARBS. ARBSs which reside near non-regulated genes could simply be enhancer elements for other genes, or some of the ARBS were merely inactive. Moreover, it is noted that there are AR-regulated genes with no ARBS nearby, potentially due to long-distance chromatin looping.

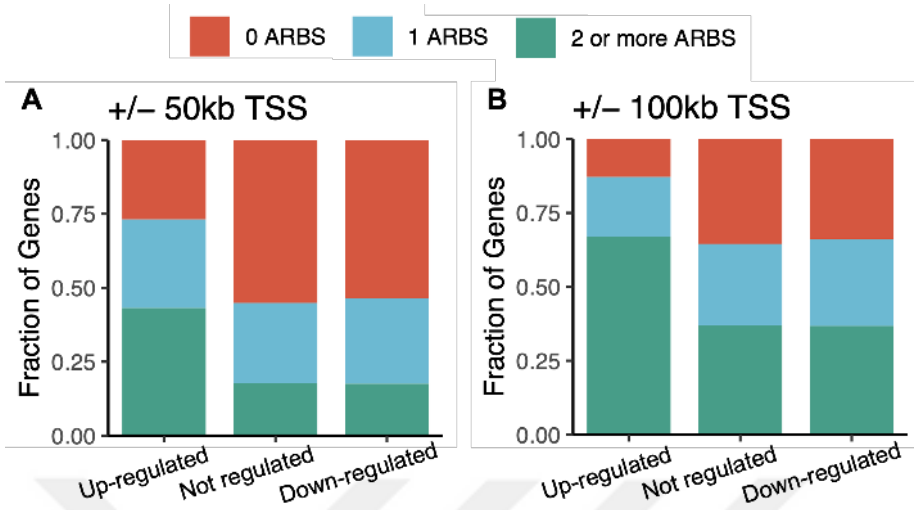


Figure 3.1: ARBSs reside in close proximity to AR-regulated genes in LNCaP cells. AR-activated publicly available RNAseq (GSE64529) and AR ChIPseq (GSE83860) demonstrated the ARBS in proximal AR-regulated genes.

3.2.2 Evaluation of AR-bound enhancers

To better demonstrate the enhancer activity in prostate cancer, we conducted a massively multi-parallel enhancer assay [119] in LNCaP cells to test the activity of clinical ARBS (Figure 3.2). These binding sites were selected from normal prostate (N=7) or prostate cancer tissue samples (N=13) [206]. High confidence AR ChIPseq peaks found in the majority of patients (>90%) were selected, including those from only normal prostate (n=262), those found only in PCa (n=3225), and those found in both (n=652). Known LNCaP enhancers were included as positive controls (n=500). Many regions in the genome contain a canonical ARE motif but do not have AR binding. Due to their similarity in nucleotide composition, we included those regions as a negative control (n=2783). To enrich our target sites, we generated a custom Agilent DNA capture of biotinylated oligonucleotides, which span 700bp per capture region. Using this, we captured these regions from fragmented human male genomic DNA (500-700bp) and cloned them into a second generation STARRseq vector [124]. Next, cloned constructs were electroporated into LNCaP cells (3 biological replicates), and after 72 hours of androgen deprivation, cells were treated either with EtOH or 10nM DHT for 4 hours. Total

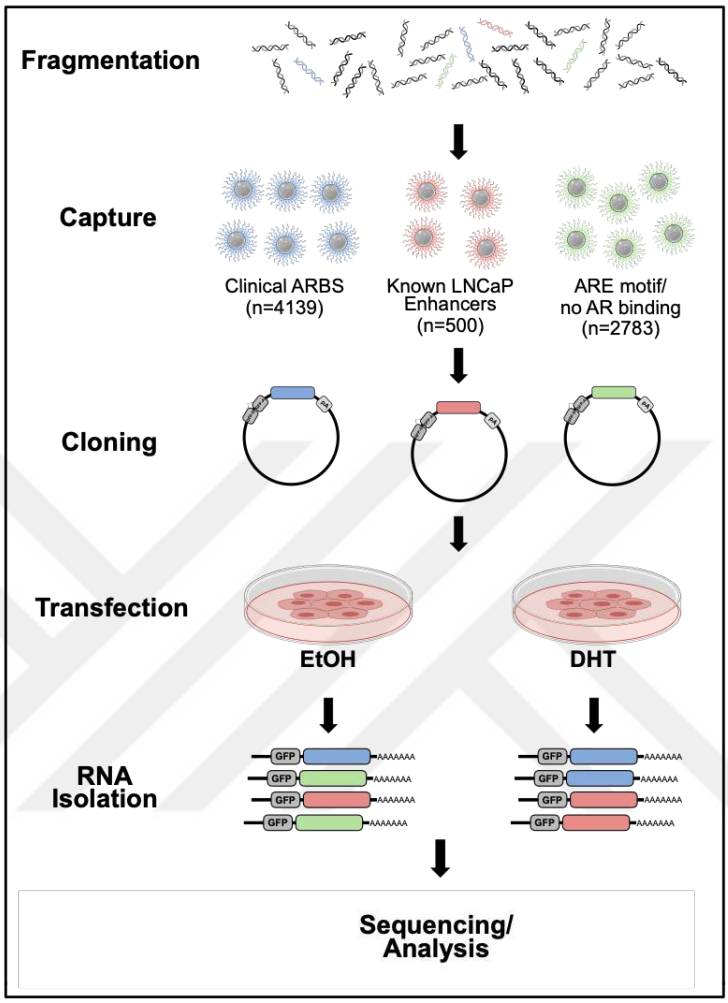


Figure 3.2: Overall scheme of AR STARRseq experiment. Overall scheme of AR STARRseq experiment. Fragmented genomic DNA was captured by oligo beads specific to the clinical ARBS, known LNCaP enhancers, and regions with ARE motif but no AR binding. After cloning those regions to a STARR-seq plasmid which consists of a minimal promoter that drives the expression of the putative enhancer region itself, they are transfected to LNCaP cells. After EtOH/DHT treatment, RNA is isolated and sent for sequencing.

RNA was isolated from these cells and used for library preparation with cassette-specific primers. Samples were sequenced by the Illumina HiSeq4000 (150bp PE-end reads). Comparing the DNA input normalized DHT fold-change over EtOH samples, we observed clear AR-dependent enhancer activity at the known *PSA/KLK3* enhancer AREIII (Figure 3.3).

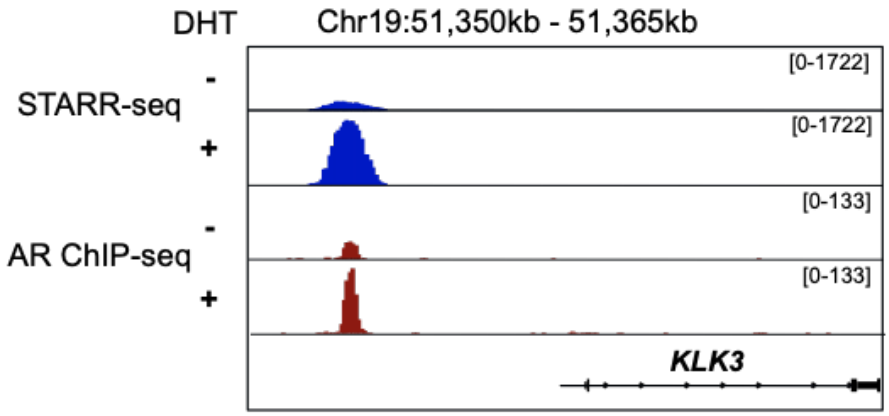


Figure 3.3: KLK3 enhancer AREIII showed AR-inducible STARRseq signal. Genomic track showing the STARRseq signal at the canonical AR enhancer upon DHT treatment.

Intriguingly, we identified three categories of ARBS enhancer elements from this reporter screen (Figure 3.4):

- **Inducible enhancers:** Elements that show increased activity when the AR is present.
- **Constitutively active enhancers:** Elements that are always active regardless of AR binding.
- **Inactive enhancers:** Elements that show no or minimal activity in the presence or absence of AR.

From 4139 clinical ARBS tested, inducible enhancers (n=265) made up only 7% of the regions tested ARBS. Inactive sites (n=3388) with minimal or no enhancer activity made up the bulk of all the tested regions (81%). Finally, constitutively active enhancers (n=465) showed enhancer activity independent from AR at 12% of the ARBS tested. When we compared the RNA Polymerase II binding in those regions, we noticed that inducible enhancers are enriched in RNA PolII binding in the presence of AR. The constitutively active regions are always bound by RNA PolII. Transcriptional activation is a good predictor of active enhancers. Therefore

we compared the bidirectional enhancer RNA transcription from the three enhancers classes (GROseq, GSE83860/GSE84432) and demonstrated that STARRseq signals are correlated with eRNA production. Due to the high frequency of H3K27ac throughout the genome, this histone mark is poorly associated with our STARRseq results. >80% of enhancers called by H3K27ac alone would be false positives. Altogether, these results provided higher resolution characteristics than existing genomic signatures.

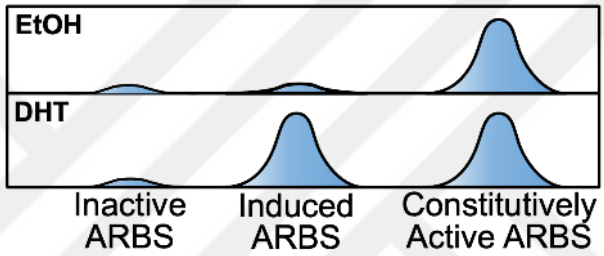


Figure 3.4: AR enhancer classes. STARRseq revealed three distinct classes of enhancer cis-regulatory elements.

3.2.3 Inducible ARBSs

To better understand how AR-mediated enhancer activity impacts gene transcription, we first concentrated on the inducible enhancers which are activated upon androgen treatment. By calculating the distance between an ARBS to its nearest AR-regulated gene and combining these distances into a cumulative distribution function (CDF), we found that inducible ARBS are closer to an AR upregulated gene than the inactive or constitutively active ARBS (Figure 3.5). Subsequently, using the AR-mediated chromatin looping dataset from AR ChIA-PET performed in VCaP, we characterized looping frequency from each ARBS enhancer class to a gene promoter. We observed that more than 15% of the inducible ARBS directly interact with an AR-upregulated gene which is significantly higher than the other two ARBS (Student's t-test, p-value <0.001, Figure 3.6A). No significant enrichment was observed in down-regulated genes

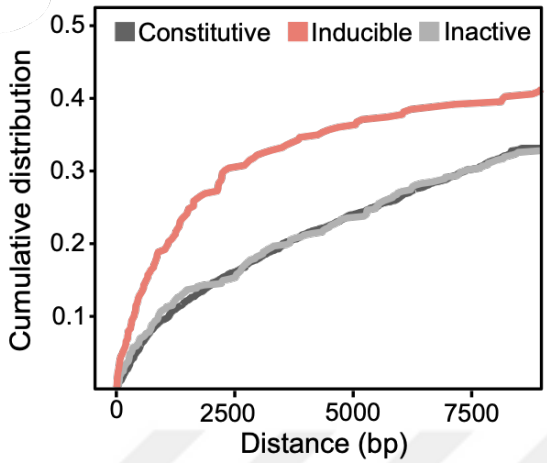


Figure 3.5: Inducible ARBS are more close to an AR upregulated gene. Cumulative distribution function (CDF) shows that inducible enhancer elements tend to be closer to the upregulated genes than the constitutively active or not active enhancers.

(Figure 3.6C). When we expanded looping targets to any genomic loci, we observed that the number of loops from inducible ARBS is significantly greater than inactive ARBS or constitutive enhancers (Student's t-test, p-value <0.001, Figure 3.6B). Breakdown of these AR-regulated chromatin loops showed that all enhancer types have interactions with other genes and other ARBSs (Figure 3.7).

3.2.4 All three enhancer types are required for transcription

After annotating the different ARBS from STARRseq, we then functionally characterize the effect of each ARBS enhancers class on gene transcription. In this, we inactivated each region with a CRISPR-based transcriptional inhibition assay (CRISPRi). This approach fuses an enzymatically inactive dCas9 with The Krüppel associated box (KRAB) domain, a repressor effector protein that induces heterochromatin formation. The dCas9-KRAB can be targeted to a specific enhancer with gRNA, where it would bind DNA and form heterochromatin and inactivate the target non-coding region. We created stably expressing dCas9-KRAB LNCaP cell lines by lentiviral transduction. Cells were selected with blasticidin, and dCas9 expression was validated by western blot (Figure 3.8).

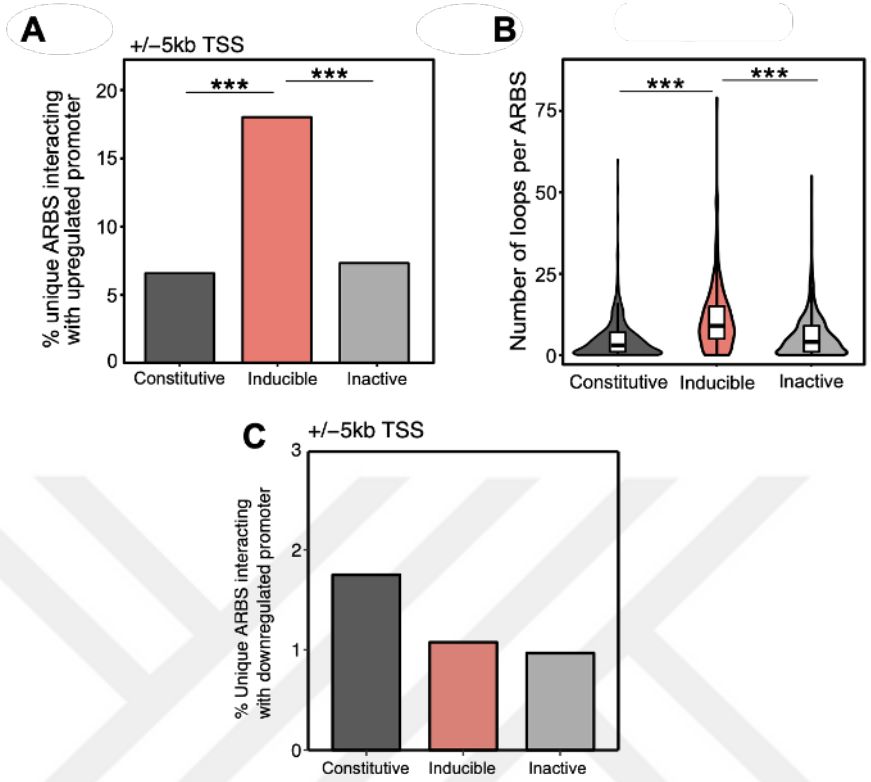


Figure 3.6: Inducible enhancers are more connected. (A) AR ChIA-PET shows a significant enrichment in inducible enhancers, looping more the upregulated genes (+/- 5kb TSS) than the remaining categories. **(B)** Inducible enhancers contain more AR ChIA-PET loops within their region compared to constitutive and inactive enhancers. **(C)** The number of ARBS interacting with a downregulated gene upon AR activation is no more than 2%. (Student's t-test, ***: p <= 0.001)

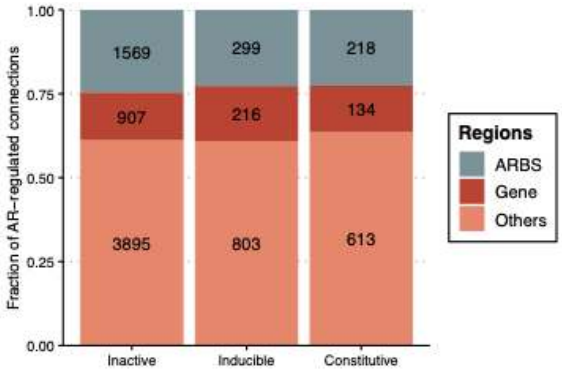


Figure 3.7: AR enhancers chromatin loop landscape. VCaP AR ChIA-PET connections are overlapped with STARRseq annotated ARBS. All of the enhancer classes have similar types of AR-mediated chromatin looping. They interact with other ARBS, genes, and other targets in the genome.

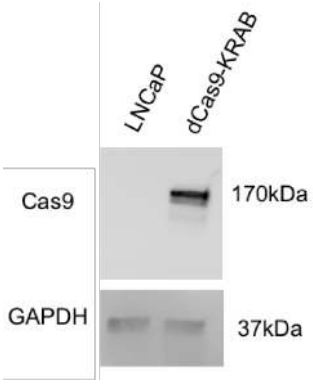


Figure 3.8: Western blot showing the LNCaP-dCas9-KRAB cell line expressing the cassette.

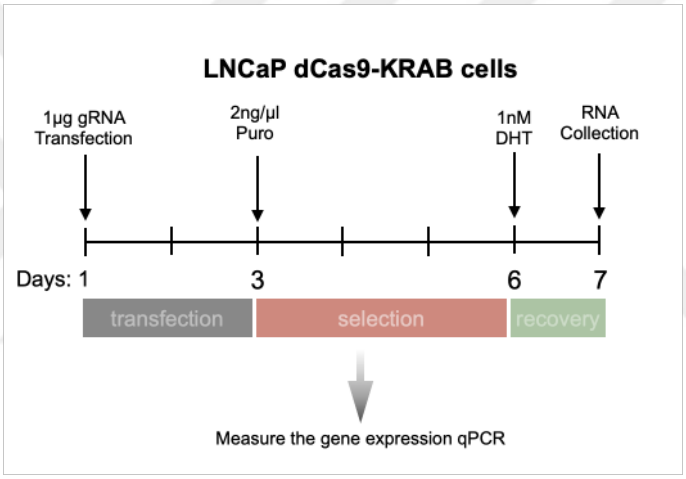


Figure 3.9: Timeline and material schematics of the CRISPRi experiments. 1ug pooled gRNA were transfected into LNCaP cells and incubated for 2 days. After antibiotic (puromycin) selection for 3 days, cells were treated with either EtOH or 1nM DHT for 24 hours before they were collected for RNA extraction.

Following validation, we then optimized the CRISPRi system with a standardized experimental timeline (Figure 3.9). Focusing on AR-mediated transcription, we first designed gRNAs against the *KLK3* promoter and its enhancer, AREIII. gRNA targeting these regions (100-250bp) were designed with CRISPR-SURF [253]. Changes in *TMPRSS2* expression, an AR-regulated gene found on a different chromosome, were assayed to determine non-specific activity. With this, we observed a significant decrease in androgen-regulated *KLK3* expression when CRISPRi was targeted to either the *KLK3* promoter or AREIII

compared to the non-targeting control (p-value < 0.0001). All gRNAs tested did not significantly impact *TMPRSS2* expression. Further, we found no significant change in CRISPRi activity with increasing plasmid DNA amounts (Figure 3.10).

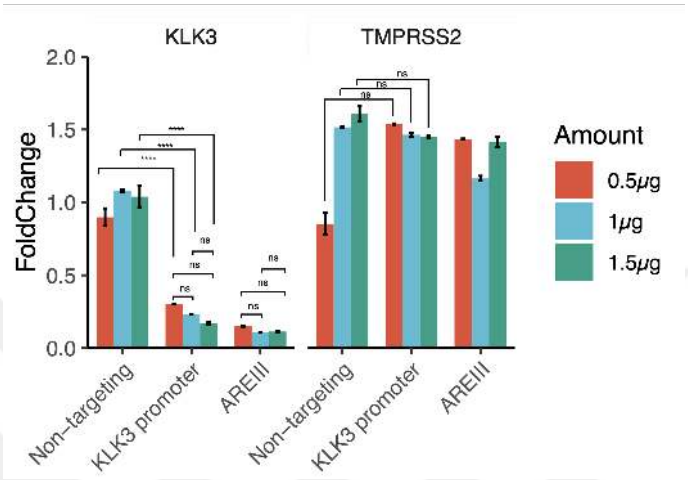


Figure 3.10: Optimization of the CRISPRi experiments. Dose-independent inhibition demonstrated only in target *KLK3* but not in *TMPRSS2*. (3 biological replicates, Two-way ANOVA, ns: $p > 0.05$, ****: $p <= 0.0001$)

Having established the CRISPRi methodology in LNCaP cells, we identified several genes that possess different ARBS enhancer types (Figure 3.11). The simplest of these was *DBI*, with only one inducible enhancer looped to its promoter. Next, the canonical *KLK3* enhancer was found to have a strong inducible enhancer (AREIII) and an upstream constitutive enhancer that loops *KLK3* and *KLK2* genes. Finally, *FKBP5* and *STEAP4* contain a complex mixture of inactive, constitutive, and induced AR enhancers. Of particular note, the induced AR enhancer of *STEAP4* is >2Mb from the TSS of the gene. To determine if any of these targeted genes were essential for cell growth, we compared these results to published genome-wide screens in LNCaP [276] and detected that *DBI* is required for cell proliferation (Figure A.1).

We then systematically inhibited all annotated ARBS using CRISPRi to better understand how the different AR enhancer classes drive transcription. Initially focusing on *KLK3*, we used multiple gRNA (n=3) to inhibit either the upstream

Genes	Loop from			Induced within +/- 50kb	Essentiality	AR Regulated
	Induced	Cons. Active	Not Active			
KLK3	✓	✓	—	✓	—	✓
FKBP5	✓	✓	✓	✓	—	✓
STEAP4	✓	✓	✓	—	—	✓
DBI	✓	—	—	✓	✓	✓

Figure 3.11: Genes selected for CRISPRi screen. 4 AR-regulated genes are selected for various genomic properties.

constitutive enhancer, the inducible AREIII enhancer, and also the gene promoter (positive control). Similar to our optimization, inhibition of AREIII showed a reduction in *KLK3* gene expression. Surprisingly, the constitutive ARBS inhibition also resulted in *KLK3* expression reduction. This indicates a co-operative action of two different enhancers for one gene (Figure 3.12A). Further, we investigated an essential AR-regulated gene. *DBI* has a proximal ARBS where only one inducible enhancer is thought to be responsible for its expression. When targeted, gene expression is significantly reduced (Figure 3.12B).

Next, we targeted those ARBS that loop to the *STEAP4* promoter. This includes an intronic constitutive, inactive and inducible AR-enhancer located a considerable distance from the *STEAP4* gene at the primary sequence (inactive=1.4Mb, inducible ARBS=2.0Mb) through chromosomal loop are brought in proximity with the gene promoter. All ARBS enhancer classes, including the inducible, impaired the *STEAP4* gene expression (Figure 3.12C).

Finally, for *FKBP5*, we targeted 3 ARBS, including an intronic constitutively active and inducible AR enhancer and an inactive ARBS near the TSS. (Figure 3.12D). *FKBP5* gene expression is significantly reduced when both the inducible and inactive were targeted. Although constitutive enhancer results are not significant, there is evidence of co-operative enhancer action of inducible and inactive enhancers. Overall these CRISPRi results suggest that all three enhancer classes are required for gene expression.

To test this potential role of ARBSs, we analyzed the inducible enhancers and

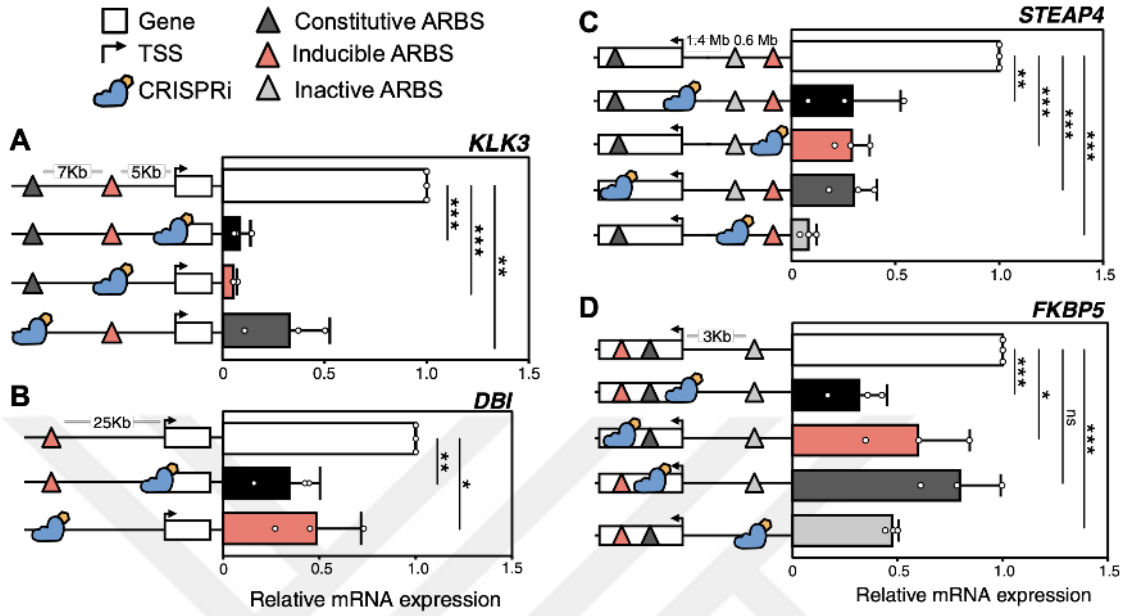


Figure 3.12: CRISPR-mediated enhancer activity inhibition. Showing that all three enhancer types are required for gene expression. (3 biological replicates +/− SD; Student's t-test, ns: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001)

their interaction partners and how they are working together to gene expression. To do that, we compared the expression profile (upon AR activation) of two subsets of AR-regulated genes: (1) genes that interact only with inducible enhancers (inducible only), and (2) genes that are in contact with both inducible enhancer and a constitutively active and/or inactive enhancer. We found that inducible enhancers that are in contact with other types of ARBS enhancers result in more transcriptional output than inducible enhancers alone (Figure 3.13).

3.2.5 ARBS chromatin accessibility associated with AR-activated cells

Having shown that AR enhancers work together to drive gene expression, we then investigated how each AR enhancer class contributes to the chromatin profile networks. Given that there is extensive evidence demonstrating that the AR binding to DNA induces increased chromatin accessibility, we, therefore, performed

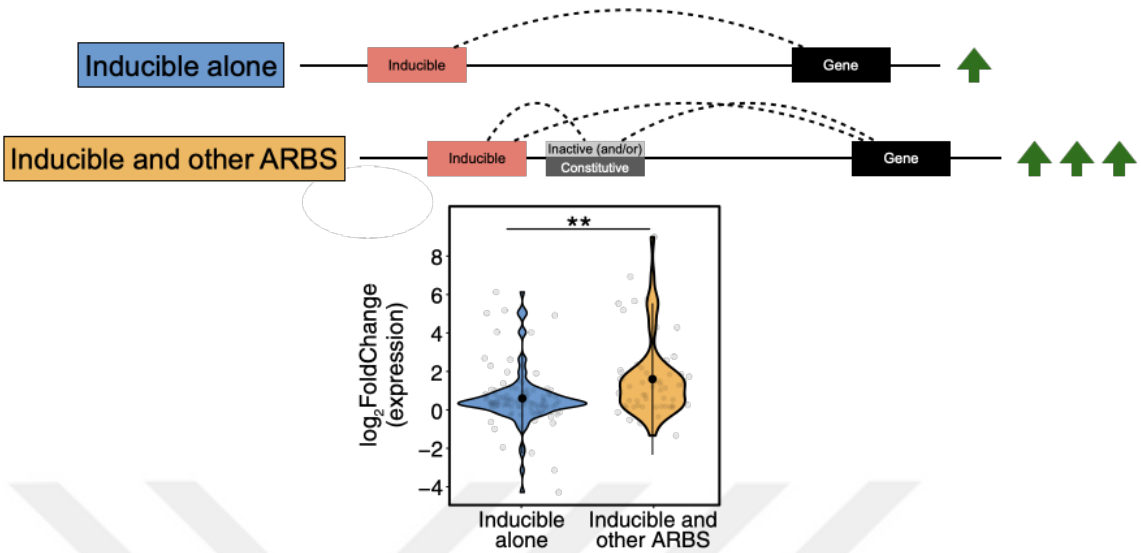


Figure 3.13: Expressions of different ARBS-interacting genes. AR-regulated genes that are in contact with only inducible enhancers which are in contact with other ARBS (n=58) have more androgen-induced gene expression than genes regulated by only inducible enhancers (n=102). (Student's t-test, **: p <= 0.01)

single-cell ATACseq (scATACseq) using LNCaP cells in either AR-inactivated (EtOH) or AR-activated (DHT) cell populations (Figure 3.14).

Sample	Number of cells	Total number of read pairs	Read pairs per cell	Read pairs mapped confidently to genome	Estimated duplicates
LNCaP DHT 2000	6,978	241,839,790	~34,657	89.4%	32.3%
LNCaP EtOH 2000	8,335	251,730,862	~30,201	89.3%	37.7%

Figure 3.14: scATACseq descriptive statistics. More than 7000 cells were processed for ATACseq library preparation, and mapping efficiency was more than 89%.

With this scATACseq data, we created co-accessibility profiles for each PCa cell using Cicero [143]. This analysis involves using accessibility information of all pairs of open chromatin regions (Figure 3.15). By calculating the frequency of pair accessibility, we can define co-accessible regions upon AR activation. Importantly, these co-accessibility networks have been shown to strongly correlate with chromatin looping [143].

When comparing these two treatment arms, we observed a clear difference between the AR-inactive and AR-active single cell populations in a Uniform Manifold Approximation and Projection (UMAP) (Figure 3.16A).

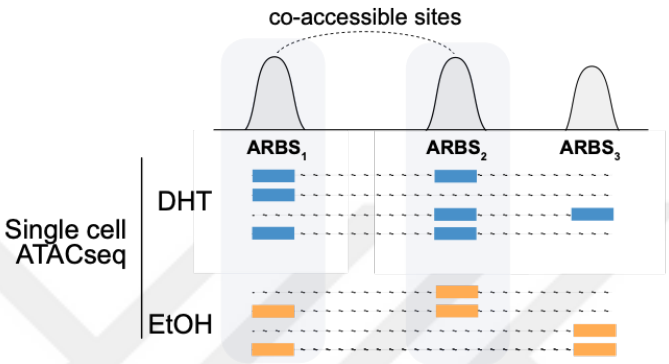


Figure 3.15: Co-accessibility analysis. The profile is calculated with the frequency of pairs of open chromatin regions in each single cell.

Using these projections, we compared the chromatin accessibility profiles of all ARBS classes. To calculate the enrichment of each peak set per single cell, we calculated the deviation scores by chromVAR [261]. This analysis calculates a z-score based on the difference between individual single cells and the average expected bulk chromatin accessibility for a given sample. By calculating these scores in projected cell populations, we showed that all three enhancer types' chromatin accessibility profiles are associated with AR-activated cells (Figure 3.16B). Supporting these results, we observed that ARBS had increased accessibility when bound by AR [277]. Using a pseudobulk ATACseq generated from the aggregated single cell populations, we showed that all three enhancer types had increased accessibility upon AR activation (Figure 3.17).

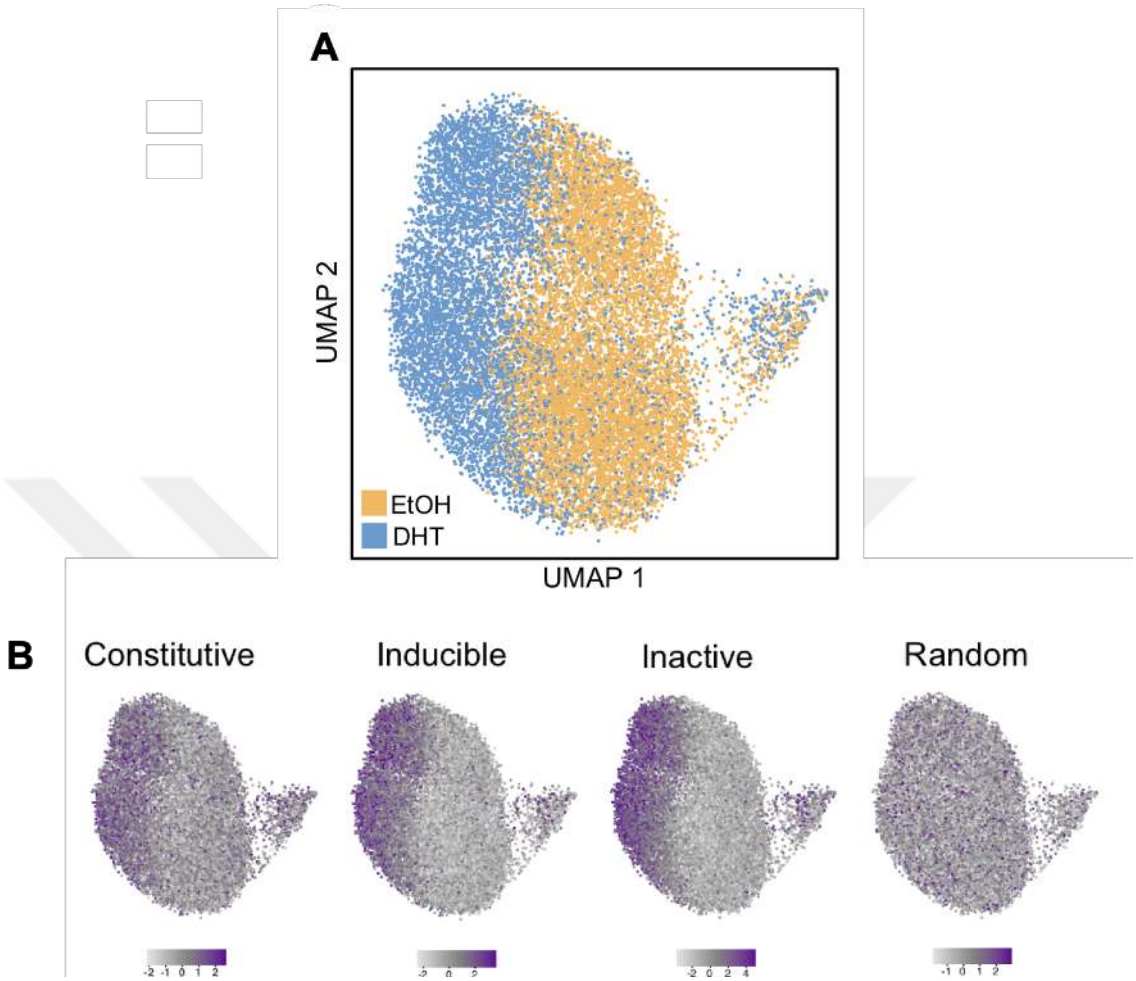


Figure 3.16: Single cell projections of AR-active and AR-inactive samples. **(A)** UMAP of the EtOH and DHT treated cells. **(B)** Deviation scores of ARBS enhancers and random regions in the genome. This score calculates the likelihood of convergence of the chromatin profiles of given regions to projected cell populations.

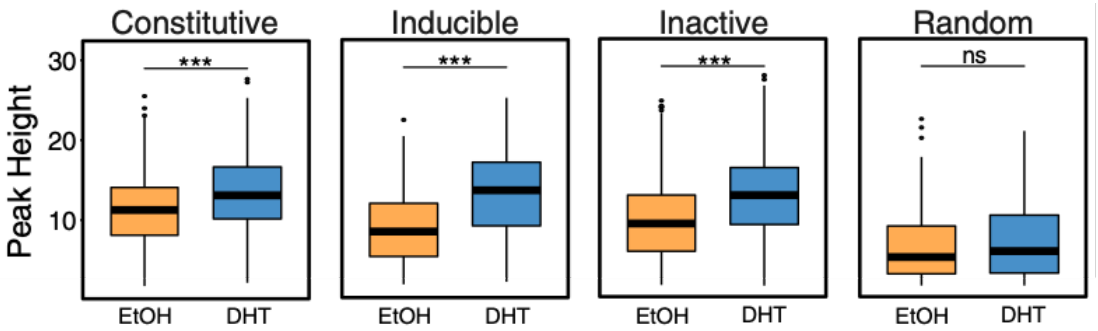


Figure 3.17: All three types of enhancers are opening upon AR activation. The different ARBS peak height of the aggregated ATACseq shows the AR-dependent opening of the chromatin. (Student's t-test, ns: $p > 0.05$, ***: $p <= 0.001$)

3.2.6 All enhancer types are co-accessible with other ARBS upon AR activation

With this scATACseq data, we then calculated the co-accessible ARBS with annotated and non-tested ARBSs. In this, we found that all annotated ARBS had increased co-accessibility with other ARBS upon AR activation (Figure 3.18).

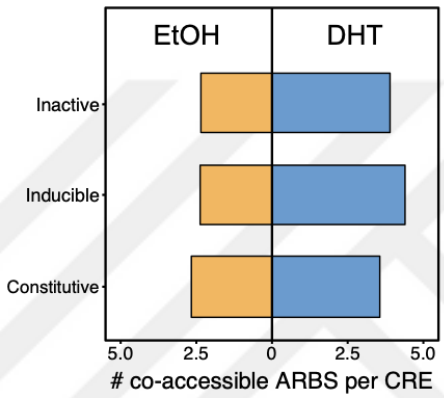


Figure 3.18: All three types of enhancers are co-accessible with other ARBS Co-accessibility profiles of all 3 types of enhancers with other ARBSs.

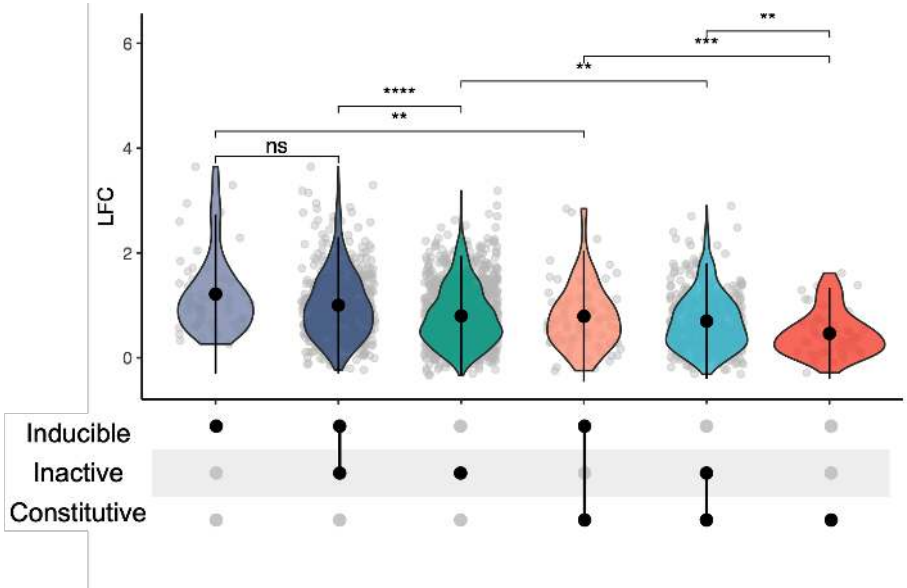


Figure 3.19: Chromatin accessibility profiles of connected ARBSs. Co-accessible ARBS enhancers that are connected via chromatin looping are calculated. (Student's t-test, ns: p > 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001)

Combining pseudobulk chromatin accessibility profile with AR-mediated chromatin looping data, we observed that AR-inducible regions frequently interact and open together. These open regions interact with themselves higher than other types of enhancers. Interestingly, the second-most interacting regions are the inducible-inactive pair. These results suggest that inducible regions may act as a regulatory hub of AR-mediated transcription while other types of enhancers are also required for gene transcription (Figure 3.19).

3.3 Discussion

Upon activation by androgens, AR translocates into the nucleus and binds to DNA, where it induces genes that are essential for PCa differentiation and growth. Most of these genes are in proximity (+/- 100kb) with multiple ARBS. Although these binding sites are reported in the genome mainly using ChIP assays, how these regions act as enhancers and activate the gene expression is poorly understood.

In this chapter, we systematically tested high-confidence clinical ARBS for their enhancer activity using STARRseq. We identified three different classes of enhancers (inducible, constitutive, inactive). When overlaying our enhancer annotation with chromosomal looping data, we found that up-regulated genes are significantly more likely to contact an induced than an inactive or constitutively active AR enhancer. We did not observe any enrichment with downregulated genes suggesting that the AR-mediated gene repression occurs through an indirect mechanism. These findings are similar to what is observed with ER suggesting this may be a common mechanism for nuclear receptor-mediated transcription [275]. However, while inducible enhancers may be critical to AR-mediated transcription, this is not to say that inactive or constitutively enhancers have no role. When we systematically inactivated different AR enhancer classes with CRISPRi, we observed that inactive and constitutively active enhancers could be critical for transcription. Indeed, all three enhancer types were evolutionary conserved, suggesting functional importance [57]. These results demonstrate an additional

function at ARBS that is not directly related to canonical AR enhancer activity but is still required for transcription. How this occurs is poorly understood. Recent work using CRISPR paired gRNA deletions demonstrated that the multiple binding sites of the ER can form both hierarchical and synergistic models to induce enhancer-activated gene expression [278]. In brief, the hierarchical model suggests a need for a primary TF motif-contained enhancer for stable gene expression. If there is a secondary weak motif enhancer, this site will only activate when the primary enhancer is active. In contrast, the synergistic model of ER activity suggests a cooperative action of those multiple sites. Our AR enhancer data cannot separate between these models but indicates that the ARBS enhancer strength is an important feature that should be considered more than motif strength. For example, the results with *KLK3* showed a co-operative regulation with different classes of enhancers where an AR-independent enhancer loops to an AR-inducible enhancer. Both enhancers are required for *KLK3* gene expression, although AR activates only one. In addition, the FKBP5 enhancers showed other characteristics where an inactive enhancer contributed to gene expression as significantly as an AR-inducible enhancer. This suggests a co-operative enhancer mechanism where the non-inducible ARBS can play a crucial role in directing the gene expression. These results indicate that inactive and constitutive enhancers may be bound by additional but essential TF's that are independent of AR. These only contribute to transcription when AR binds DNA, and inducible enhancers act as regulatory hubs. The minimally active ARBS are still vital since there is a significant reduction in the gene expression without AR binding.

Overall our data suggest that multiple enhancer sites are required for stable gene expression but surprisingly, not all those regions have to be inducible ARBS.

Chapter 4

KLK LOCUS ENHANCERS

4.1 Introduction

One unexpected finding from our functional characterization of several AR-regulated genes (Chapter 3) was the critical role of both inactive and constitutively active ARBS in gene transcription. As AR binding does not impact the enhancer activity in an in vitro enhancer assay, this suggests that different types of ARBS must work in concert to drive gene transcription.

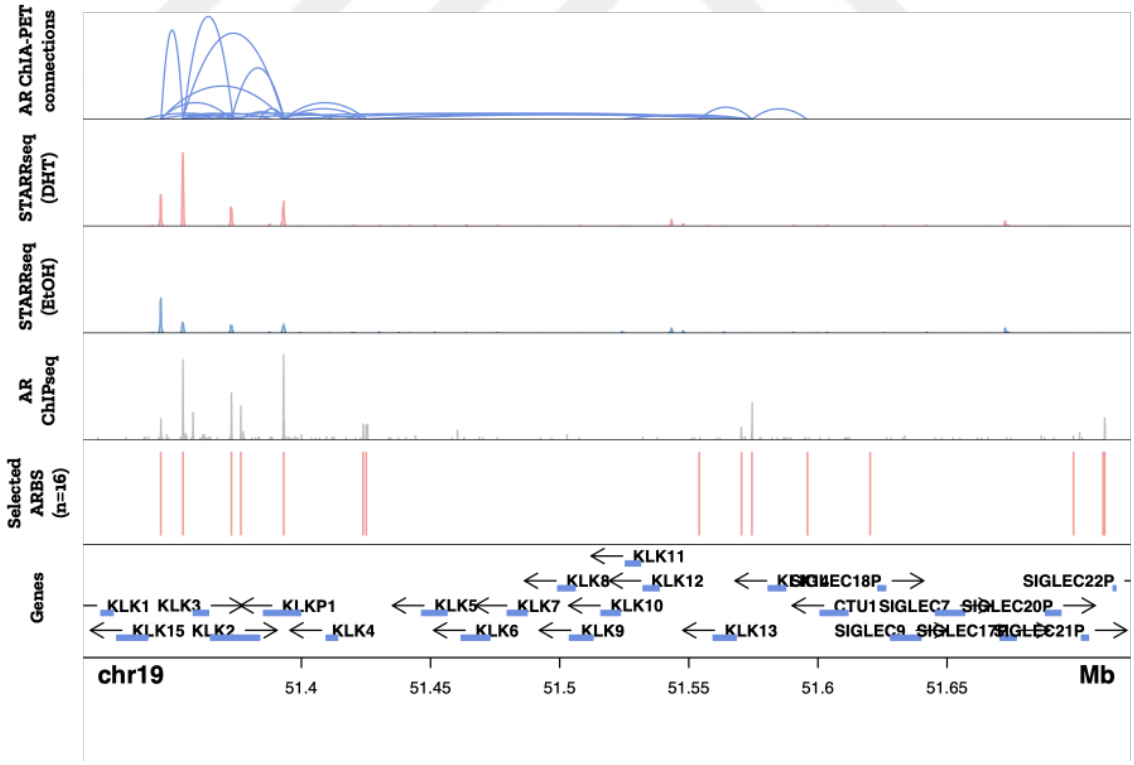


Figure 4.1: The KLK locus. It consists of 15 KLK genes with 16 ARBS within the AR-mediated chromatin loop domain.

To further explore the complex transcription control of AR-mediated gene expression, we focused on a single locus with multiple AR binding sites. We selected the KLK locus as a model as there are 16 ARBS and 15 KLK genes (Figure 4.1). This includes the well-characterized *PSA/KLK3* enhancer AREIII, which we termed ARBS2. Importantly, none of these were essential for LNCaP growth in published DepMap data (Figure 4.2). Therefore any disruption to KLK gene function will have no impact on LNCaP proliferation.

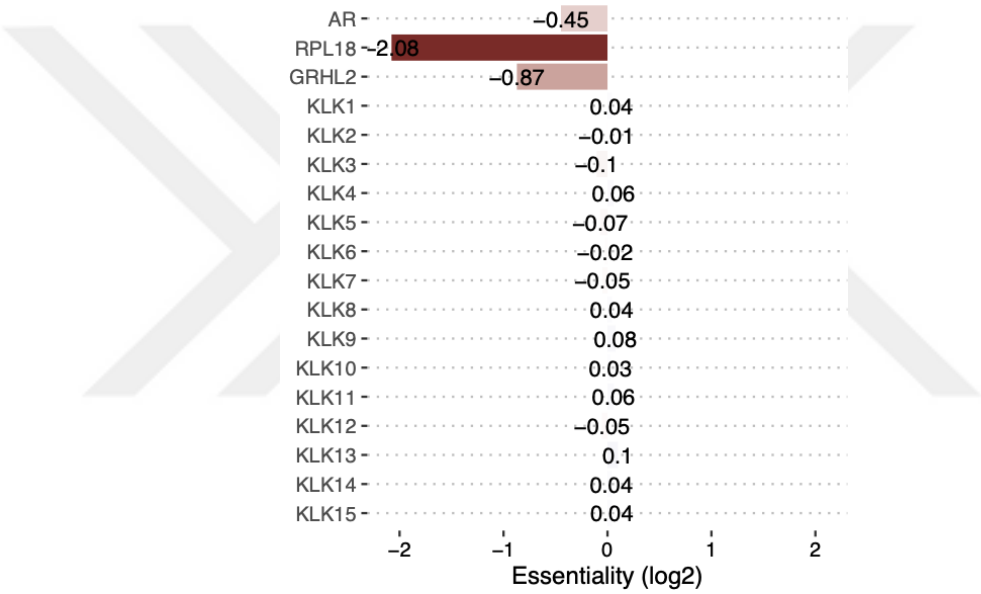


Figure 4.2: LNCaP essentiality of KLK genes. In genome-wide CRISPR screens, AR, RPL18 ribosomal protein (positive control), inducing, and AR-coregulator GRHL2 are essential for LNCaP proliferation. No KLK genes were required for growth. Produced from DepMap 21Q4 Public [279]

4.2 Results

To better understand the genomic characteristics of how ARBS enhancers impact KLK expression and, we integrated co-accessibility (scATACseq), AR-mediated chromatin looping (ChIA-PET), and H3K27ac HiChIP to characterize the chromatin looping between ARBS and the KLK genes (Figure 4.4). From this, we observed two distinct groups of genes that are affected by chromatin openness,

histone acetylation, and AR-mediated chromatin looping: *KLK1*, *KLK2*, *PSA/KLK3*, *KLK4*, *KLK15* (subset 1: S1), and *KLK5-14* (subset 2: S2) gene sets are separated by genomic locus. S1 regions of genes are also segregated from the rest of the KLK gene locus (S2) with a CTCF binding site [249].

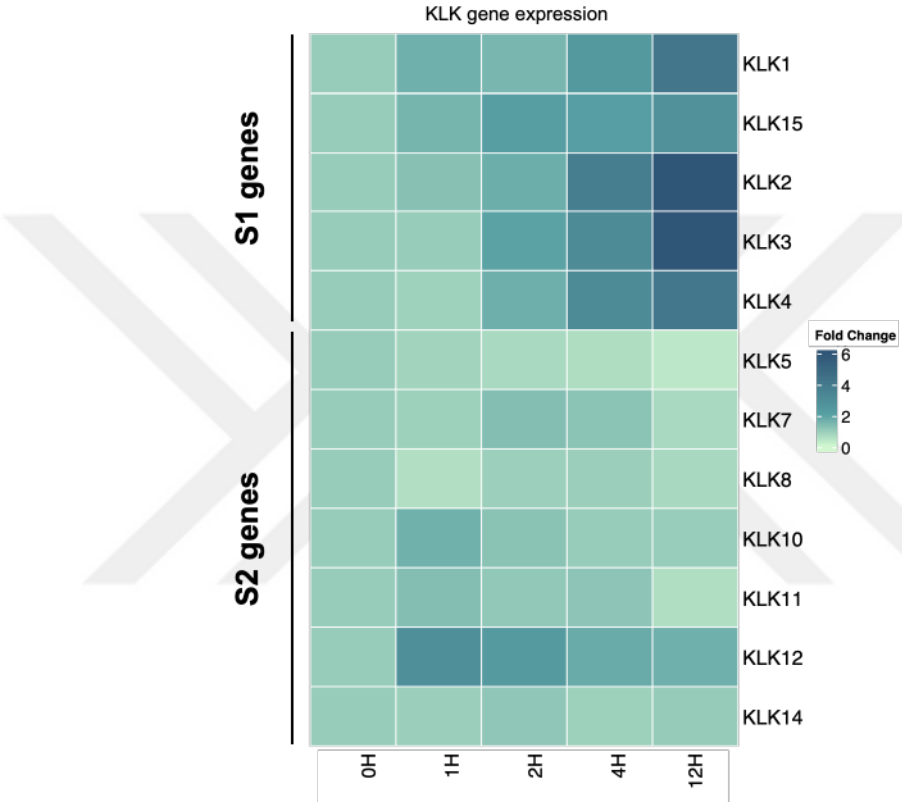


Figure 4.3: Time course gene expression results of the entire KLK locus. 200.000 (seeded in 6-well plate) LNCaP cells are androgen deprived for 72 hours and treated either with EtOH or 10nM DHT for 6 hours (3 biological replicates). *KLK1-4* and *KLK15* (S1) and KLK5-14 (S2) show distinct differential expression profiles upon AR activation.

Although co-accessibility profiles of those two subsets are similar, S1 genes were more frequently co-accessible with ARBS1, ARBS2/AREIII, and ARBS5, which is higher than other ARBSs. Further, AR-mediated looping shows the enrichment in the same S1 genes and ARBSs. Finally, H3K27ac marked chromatin looping demonstrates that, broadly, only S1 genes are participating in this looping event. This is supported by transcriptional data as only S1 genes are also upregulated by their androgen treatment (Figure 4.3). Supporting these groupings, S1 genes have

progressive androgen-dependent gene expression while S2 genes are not transcribed in LNCaP cells.

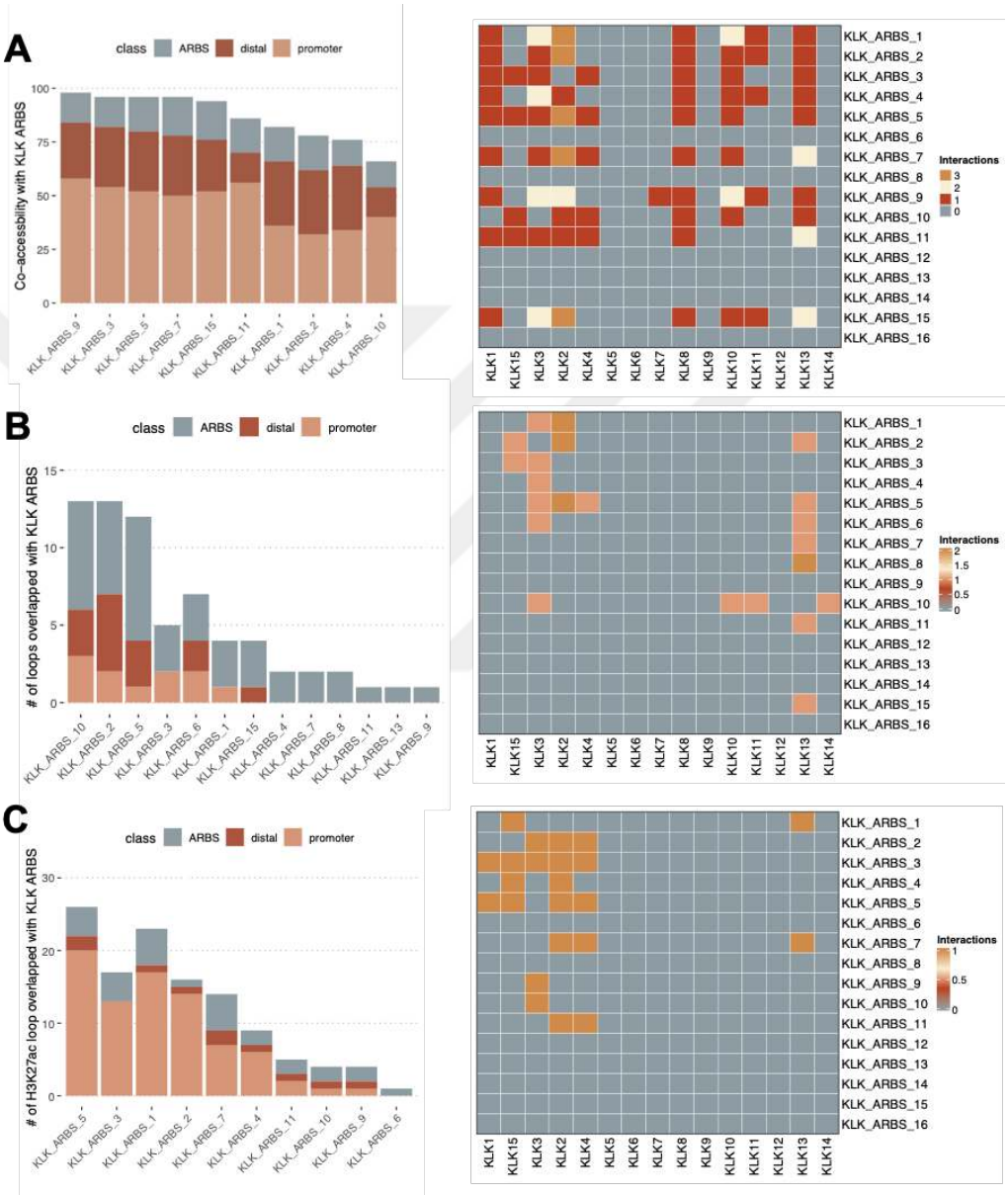


Figure 4.4: Genomic landscape of the KLK locus and ARBS. (A) Co-accessibility profiles of KLK genes and ARBS. Number of ARBS, promoter, and other genomic sites found in open chromatin with KLK ARBS. Distal regions are the regions other than promoters or ARBS. **(B)** AR-mediated chromatin loops in KLK genes and ARBS. Number of ARBS, promoter, and other genomic sites connected with KLK ARBS through AR-mediated chromatin loops. **(C)** H3K27ac chromatin loops in KLK genes and ARBS. Number of ARBS, promoter and other genomic sites connected with KLK ARBS through AR-mediated chromatin loops.

4.2.1 Locus-wide ARBS perturbation with CRISPRi

To further investigate ARBS enhancers, we applied CRISPRi to each ARBS (n=16) in the KLK locus and quantified how they impacted gene expression (n=15) across numerous biological replicas (n=6) (Figure 4.5).

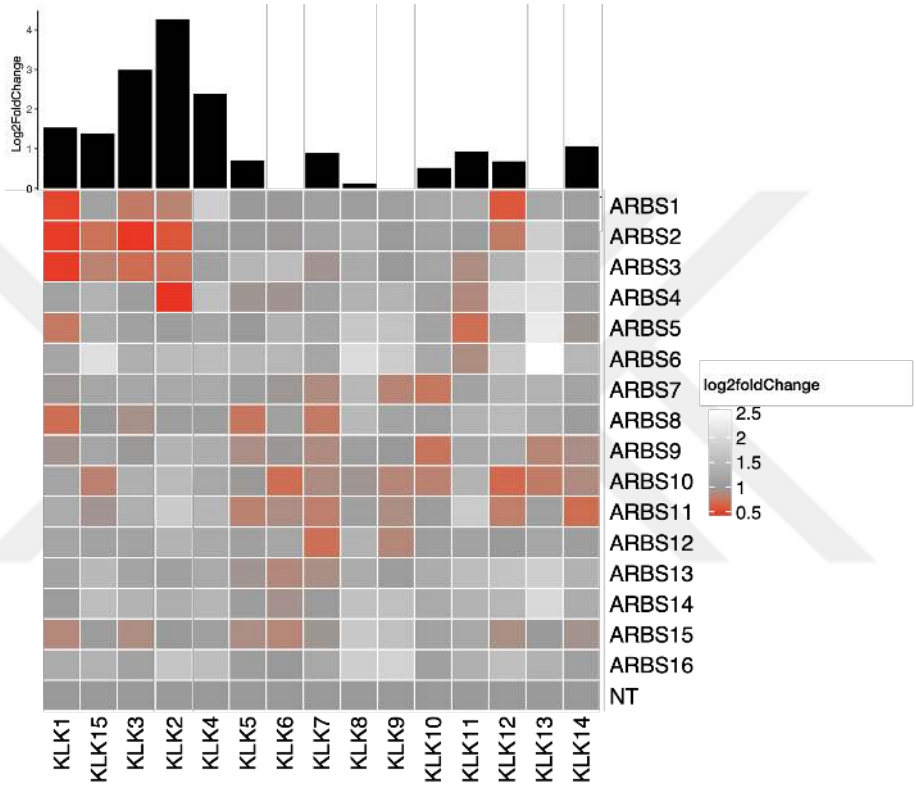


Figure 4.5: Heatmap showing the CRISPRi for the KLK locus. 6 biological replicates were targeting each ARBS in KLK locus with dCas9-KRAB in LNCaP cells. (normalized to non-targeting gRNA). Top: KLK gene expression in LNCaP cells upon 10nM DHT treatment for 4 hours.

CRISPRi inhibition of each ARBS revealed a significant cluster of genes affected by a group of ARBSs that interact via chromatin loops and reside in close proximity (Figure 4.6). This cluster consists of ARBS1 to ARBS5 (ARBS4 is the *KLK2* promoter – not included), and the genes affected by it are *KLK*, *KLK2*, *PSA/KLK3*, *KLK15*. To note, KLK4 was not affected by any individual ARBS.

From this, we observed that inducible enhancer ARBS2/AREIII had a more significant impact on *PSA/KLK3* and *KLK2* gene expressions than either

constitutive (ARBS1) or inactive enhancers (ARBS3). Nevertheless, when inhibited, the inactive enhancer ARBS3 also affected the gene expression of *KLK*, *KLK2*, *PSA/KLK3*, and *KLK15*. The decrease in *PSA/KLK3* and *KLK2* gene expression, in the absence of ARBS2/AREIII and ARBS3 (*KLK2* AREII), was previously published using mouse models [280].

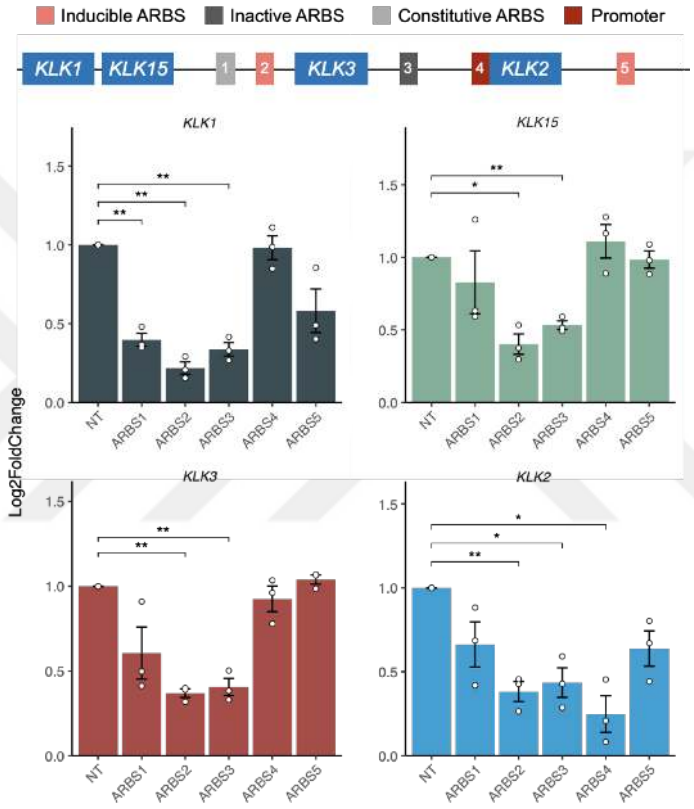


Figure 4.6: CRISPRi targeting ARBS and selected KLK gene expression. *KLK*, *KLK15*, *PSA/KLK3*, and *KLK2* gene expression decreased significantly after proximal ARBS perturbation. NT: non-targeting gRNA (3 biological replicates, Student's t-test, *: $p < 0.05$, **: $p <= 0.01$).

4.2.2 CTCF loss and KLK gene expression

The S1 region in the genome comprises 70kb and is protected by two flanking CTCF binding sites. Given their role in both E-P and TAD looping, we, therefore, interrogated whether these CTCF binding sites affect the expression of genes that are either inside (*PSA/KLK3* and *KLK2*) or outside (*KLK* and *KLK15*) of these

two CTCF binding sites. After targeting gRNAs (n=3) to the upstream (5') and downstream (3') CTCF binding sites, we found no significant impact of these on the gene expression (Figure 4.7). Several studies demonstrated CTCF-loss has mixed effects on gene expression [281, 282]. A recent report revealed that the S2 genes are activated when the CTCF binding site between S1 and S2 KLK gene locus is disrupted [249]. This may suggest that CTCF binding site functionality is specific to transcriptional networks. Further investigation of the AR occupancy upon the CTCF boundary mentioned above and the effect of the other CTCF binding sites are required to delineate the CTCF and AR binding with KLK gene activation.

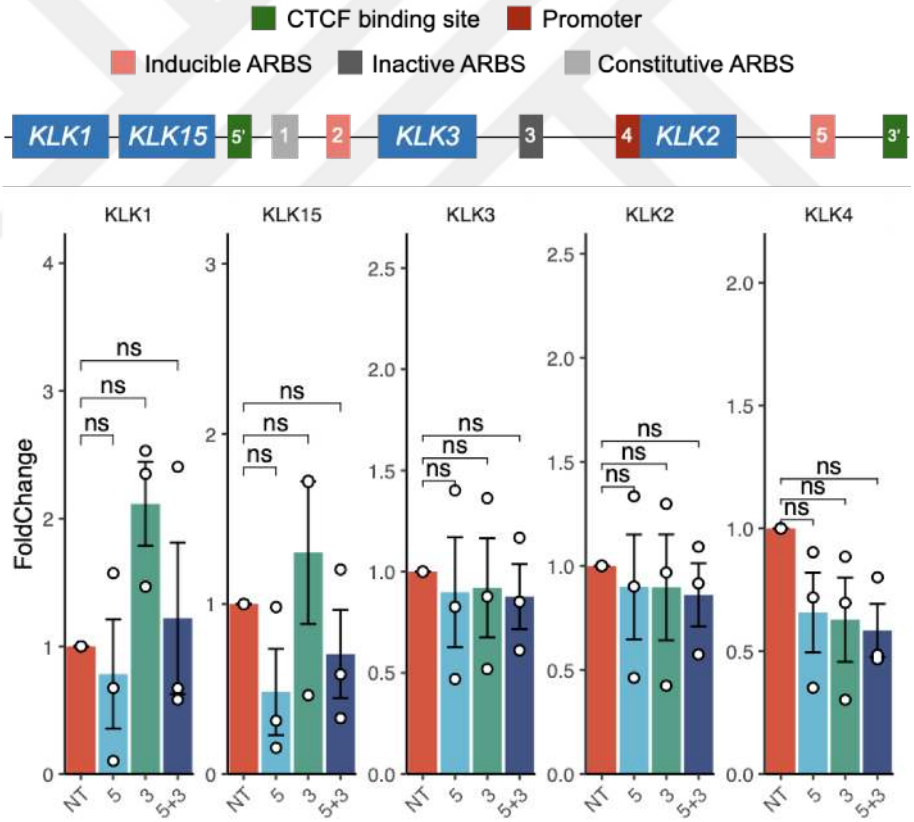


Figure 4.7: CRISPRi targeting CTCF binding sites in KLK locus. NT: non-targeting gRNA, 5: upstream CTCF binding site, 3: downstream CTCF binding site, 5+3: both sites are targeted with CRISPRi. (3 biological replicates +/− SD; Student's t-test, ns: $p > 0.05$)

4.2.3 AR-mediated chromatin looping in PSA/KLK3 and KLK2

As ARBS2/AREIII was found to control the gene expression of both *PSA/KLK3* and *KLK2*, we quantified the interactions with this enhancer and gene promoters with 3C. To do this, we conducted a 3C assay from ARBS2/AREIII to *PSA/KLK3* and *KLK2* gene promoters (Figure 4.8). Similar to previous work, we observed that the ARBS2-*KLK2* chromatin loop had increased frequency when treated with AR [228]. Similarly, *PSA/KLK3* gene promoters have an increased interaction frequency upon AR activation, demonstrating the AREIII-KLK3 chromatin looping.

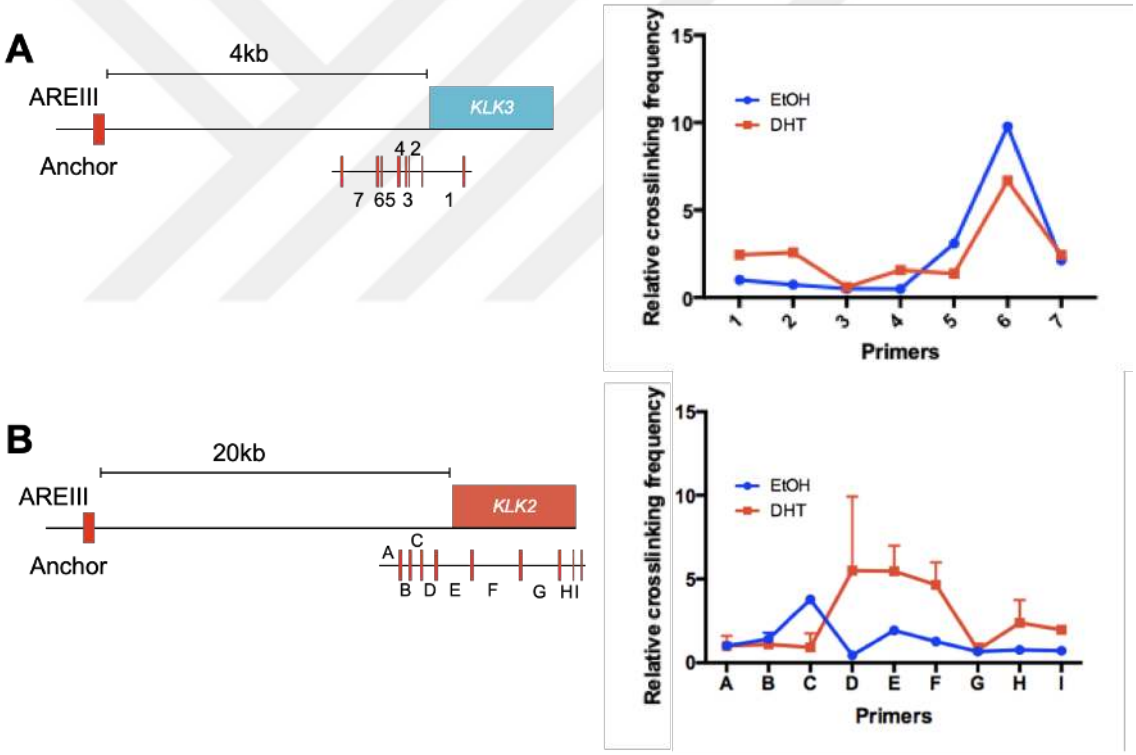


Figure 4.8: ARBS loops to KLK genes in LNCaP cells. LNCaP cells are androgen-deprived for 3(A) AREIII-KLK3 and (B) AREIII-*KLK2* interactions. 3C-qPCR. Reproduced from [228].

4.2.4 S1 gene locus and ARBS

In the previous chapter, we reported that inducible enhancers are significantly more connected to any genomic target than other enhancer classes. Supporting

this, in the KLK locus, the inducible enhancers ARBS2/AREIII and ARBS5 had more AR-mediated chromatin loops to cis-regulatory elements, including and gene promoters, than ARBS1, ARBS3, and ARBS4 (Figure 4.9).

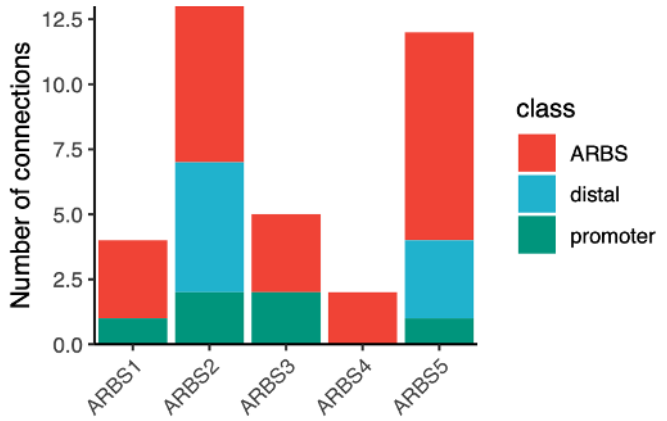


Figure 4.9: AR-mediated chromatin loops of selected KLK ARBS. All ARBS are connected to other ARBS and other KLK genes via AR-mediated chromatin looping.

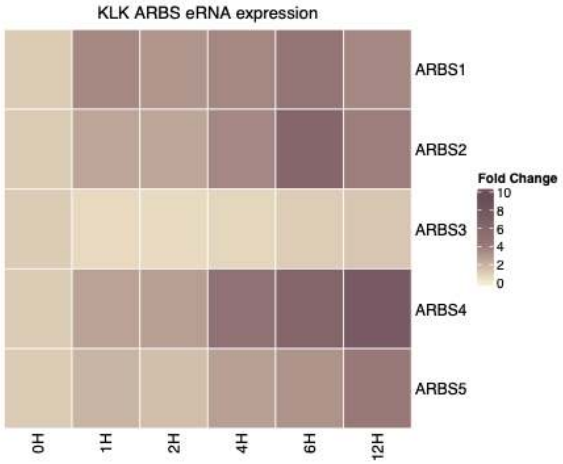


Figure 4.10: ARBSs express eRNA. LNCaP cells are androgen starved for 72 hours, and upon 10nM DHT treatment for different time points, eRNA was quantified with qPCR.

Almost all enhancer elements transcribe either polyadenylated or non-polyadenylated eRNA. We, therefore, quantified the eRNA expression from ARBS1-5 by qPCR. Other than ARBS3 (inactive), all ARBS were found to

produce RNA following AR binding (Figure 4.10).

Using our selected ARBS, we demonstrated that, although there is a similarity in AR binding site motifs at promoters of *PSA/KLK3* and *KLK2* genes, there were no apparent differences between enhancer types (Figure 4.11). This supports both previous reports and our general observation that more than >50% of ARBS do not contain ARE motif [283].

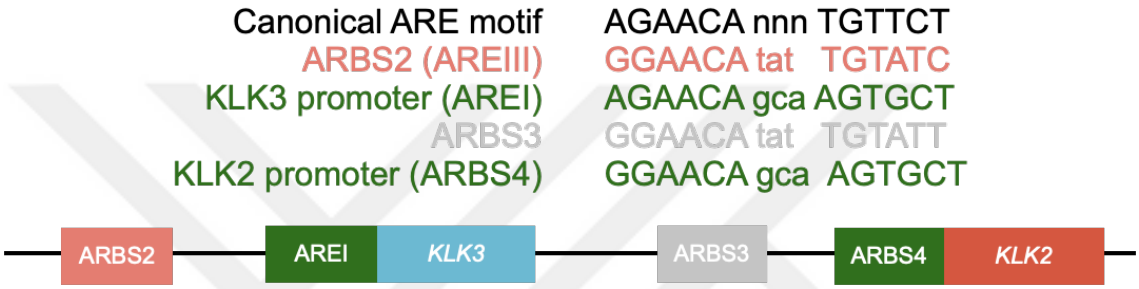


Figure 4.11: *PSA/KLK3* and *KLK2* gene promoters, enhancers, and core motif sequences. Although two enhancers (ARBS2/AREIII and ARBS3) are bound by AR and motifs are similar, they are of different enhancers (inducible and inactive, respectively).

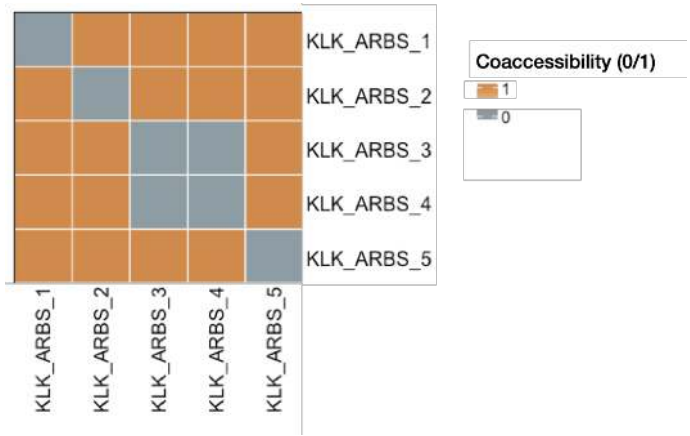


Figure 4.12: Co-accessibility profile of KLK ARBS1-5. scATACseq revealed that S1 KLK ARBS are opening upon AR activation.

Active enhancers reside in open chromatin regions. To show the collective activation of the ARBS1-5, we used scATACseq co-accessibility profiles. Indeed, except, ARBS3 and ARBS4, all other ARBS became co-accessible upon activation (Figure 4.12).

4.2.5 Enhancer rewiring

Previous work has demonstrated that ARBS2/AREIII has two gene targets: *PSA/KLK3* and *KLK2* [228]. This targeting is through enhancer-promoter chromatin looping potentially mediated by eRNA transcription. Supporting these results, we observed that when ARBS2/AREIII was inactivated with CRISPRi, both *PSA/KLK3* and *KLK2* showed a significant decrease in AR-mediated transcription (Figure 4.10). Interestingly, when we targeted the *PSA/KLK3* promoter with CRISPRi, we observed a significant decrease of *PSA/KLK3* and an unexpected increase in *KLK2* transcription. This may be enhancer rewiring, a recently described phenomenon that occurs when a target promoter is disrupted, leading the enhancer to have increased activity with a different promoter [284] (Figure 4.13).

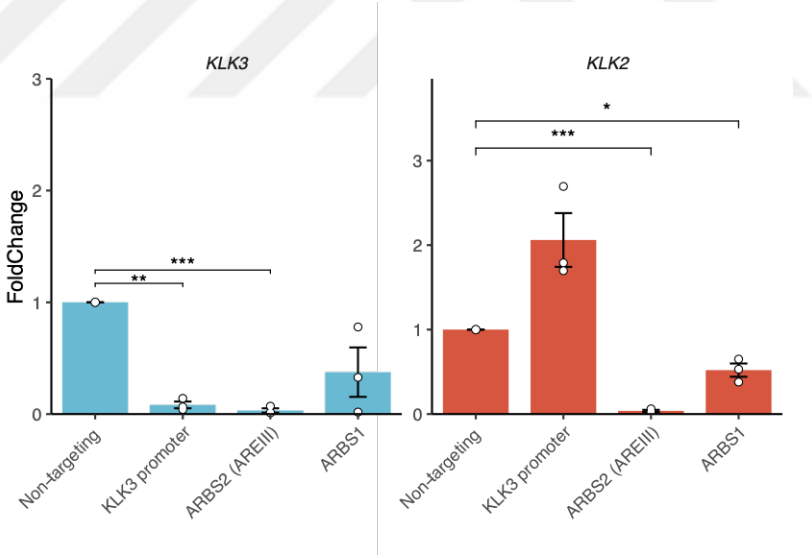


Figure 4.13: Enhancer rewiring in KLK locus. *KLK2* and *PSA/KLK3* gene expression upon the inhibition of their enhancers. (3 biological replicates, Student's t-test, *: $p < 0.05$, **: $p <= 0.01$, ***: $p <= 0.001$)

4.2.6 Enhancer-null cells phenocopied the decrease in enhancer-linked gene expression

In our previous work, we transiently transfected gRNAs into LNCaP-dCas9 expressing cells to capture the effect of the ARBS. To ensure this was not an

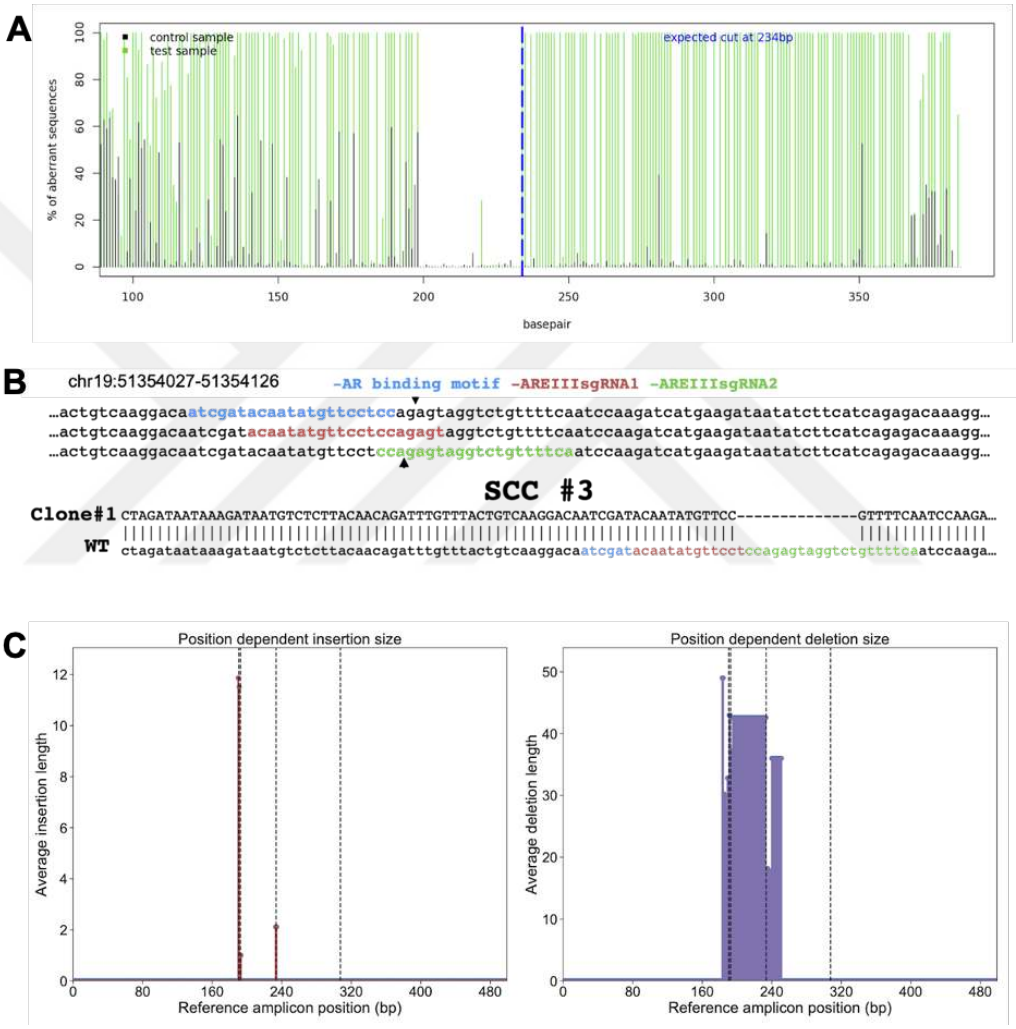


Figure 4.14: Indel validation of the ARBS-null cells. **(A)** ARBS1 indel in LNCaP KO cells. TIDE algorithm [285] was used to analyze Sanger sequencing data. It revealed the 40bp indel in ARBS1. **(B)** ARBS2/AREIII indel in LNCaP KO cells. Sanger sequencing revealed that single-cell clone 3 has 14bp deletion around the AR motif. **(C)** ARBS5 indel in LNCaP KO cells. CHOPCHOP [252] was used to analyze PCR amplicon sequencing data. It revealed the 60bp homozygous indel in ARBS5.

assay-specific artifact and allow detailed characterization, we created single-cell knock-out (KO) clones for ARBS1, ARBS2/AREIII, and ARBS5 with

CRISPR/Cas9. We chose these ARBSs because they contain both inducible and constitutive enhancers and profoundly impact KLK gene expression (Figure 4.6). After gRNA and Cas9 transfection to LNCaP cells, we grew single-cell clones from the antibiotic selected population using limiting dilution. As CRISPR-mediated

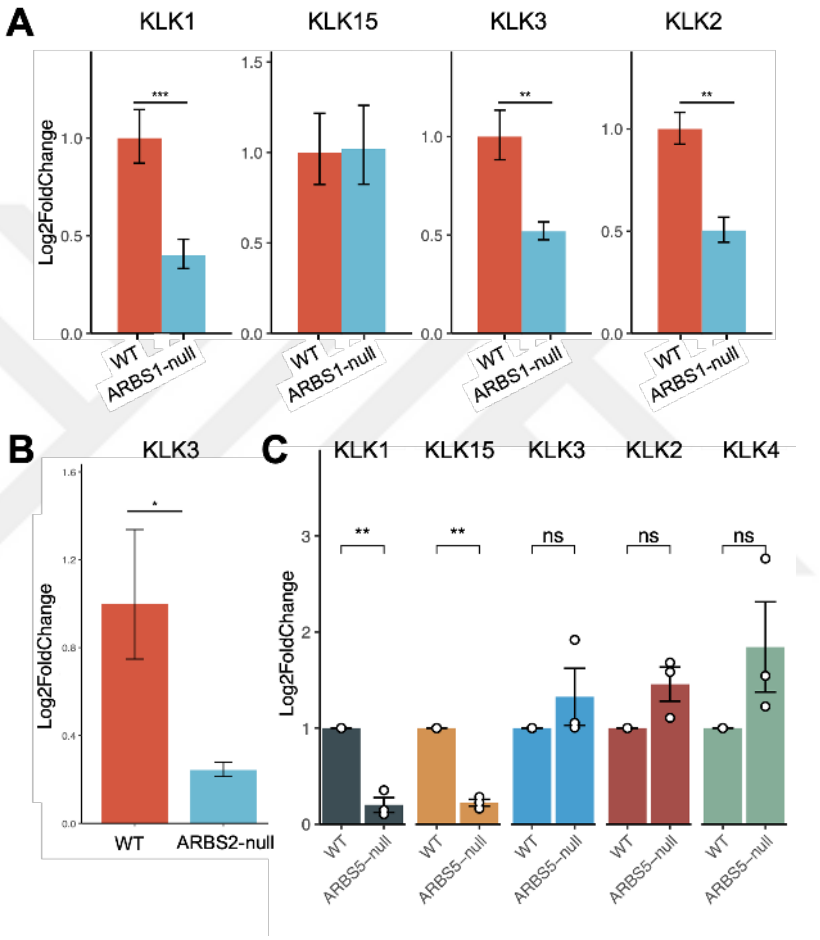


Figure 4.15: ARBS-null cells showing the decreased gene expression. **(A)** ARBS1-null cells KLK gene expression compared to the WT LNCaP **(B)** ARBS2-null single-cell clones cells KLK gene expression compared to the WT LNCaP **(C)** ARBS5-null cells KLK gene expression compared to the WT LNCaP. (3 biological replicates +/− SD; Student's t-test, ns: p > 0.05, **: p <= 0.01, ***: p <= 0.001)

indel formation is dependent on gRNA design, gene target, and cell karyotype, we tested an average of 20 single-cell clones per genomic to ensure the cells contained a homozygous deletion. These clones were then validated by Sanger and/or high-throughput sequencing to confirm the presence of an indel at the target

ARBS (Figure 4.14). From this, we generated homozygous deletions of all targeted ARBS in LNCaP. Using gRNAs targeting ARBS1 (n=3), we generated ARBS1-null cells, which contain a 40bp deletion within the target site (Figure 4.13A). We next generated ARBS2/AREIII-null cells by targeting the AR-motif multiple gRNAs (n=2). Out of four isolated ARBS2/AREIII-null cells, single-cell clone 3 (SCC3) was selected after Sanger sequencing revealed a 4bp indel around those two target sites (Figure 4.13B). ARBS5-null cell line were validated with high-throughput PCR amplicon sequencing and found to contain a 60bp homozygous indel (Figure 4.13C).

In concordance with the dCas9-KRAB inhibition, ARBS1 disruption decreased *KLK1*, *KLK2*, *PSA/KLK3* gene expression while *KLK15* remained unaffected (Figure 4.15A). ARBS2/AREIII-null LNCaP cells showed decreased *PSA/KLK3* gene expression (Figure 4.15B). ARBS5-null cell KLK gene expression profile aligns with the KLK locus-wide CRISPRi screen where *KLK1* and *KLK15* are less expressed while other remains not significantly changed (Figure 4.15C).

4.2.7 Loss of ARBS affects AR binding in both target genes and nearby ARBS

Based on our earlier findings, we hypothesized that *KLK1*, *KLK15*, *PSA/KLK3*, and *KLK2* transcriptional regulation is tightly connected via chromatin looping of common enhancers. To determine how AR binding in this locus is impacted following ARBS disruption, we conducted AR ChIP-qPCR in each ARBS-null cell line. Using constitutive ARBS1-null and inducible ARBS2-null cells, we compared the AR binding upon DHT treatment with wild-type (WT) LNCaP cells at both ARBS2/AREIII and the *KLK2* promoter.

Disruption of the ARBS1 and ARBS2/AREIII resulted in a significant decrease in AR occupancy in the *KLK2* promoter. Surprisingly, we observed that AR binding around the ARBS1 region was also reduced upon ARBS2/AREIII disruption (Figure 4.16A). To investigate this phenomenon, we tested ARBS5-null cells to characterize the AR binding throughout the selected locus. Indeed, ARBS5

disruption resulted in a significant decrease in AR binding in all other ARBS (Figure 4.16B). Although there is a significant decrease in AR binding through ARBS1-5, only *KLK1* and *KLK15* gene expressions are affected. This suggests that this closed (S1) transcriptional network may have a compensation mechanism where ARBSs are interchangeable throughout the S1 genes. More insights on the 3D conformation of this S1 locus are required to explain the complex enhancer-promoter arrangements.

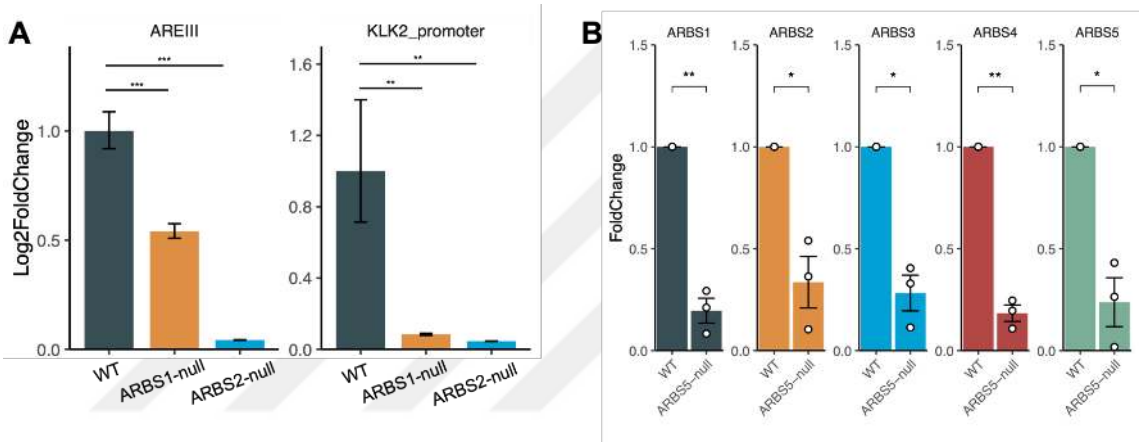


Figure 4.16: AR binding depends on other ARBS. (A) ARBS and *KLK2* gene promoter AR occupancy upon ARBS1 disruption (AREIII=ARBS2). **(B)** ARBS AR occupancy upon ARBS5 disruption. (3 biological replicates SD; (A) ANOVA and (B) Student's t-test, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001)

Overall, this phenotype suggests that different enhancers cooperatively work together to drive gene expression, and enhancer-binding requires mutual AR occupancy in the same transcriptional network.

4.3 Discussion

In this chapter, we systematically inhibited all ARBS in the KLK locus using CRISPRi. From this, we focused on a subset of AR-activated 5 KLK genes (*KLK1-3* and *KLK15*) within the S1 cluster. Since the KLK locus most likely arose through gene duplications, it is plausible that the expression of each KLK gene is regulated separately by conserved regulatory elements. However, because of

the very same reason, they might have shared regulatory elements that form a unique transcriptional complex. The close proximity of the *KLK1-4* and *KLK15* supports this hypothesis that duplications occur exclusively on the gene side, and enhancers rapidly evolve to control those duplicated genes. We showed that *PSA/KLK3* and *KLK2* share the same ARBS1 and ARBS2/AREIII enhancers via chromatin looping. Upon inactivation of *PSA/KLK3*, we observed enhancer rewiring that directs ARBS2/AREIII from *PSA/KLK3* promoter to *KLK2*, increasing the *KLK2* transcriptional output. Using ARBS-null stable cell lines, we measured the AR occupancy in other KLK ARBS. ARBS1 perturbation causes a significant decrease in nearby ARBS2/AREIII AR occupancy. As a side note, like ARBS2, the effect of transcribed ARBS enhancer RNAs on KLK gene expression remains to be interrogated. Similarly, ARBS5 inhibition resulted in a decrease in AR occupancy in all ARBS found in the same transcriptional complex. However, gene expression of 5 KLK genes upon ARBS5 knockout is more complex. Both *KLK1* and *KLK15* gene expression levels strongly correlate with the initial CRISPRi screen. However, although we see a decrease in AR binding in ARBS4 (which is the promoter of *KLK2*) and ARBS1, ARBS2/AREIII (enhancers that control the expression *PSA/KLK3* and *KLK2*), we did not see a significant decrease in their gene expression. This result suggests that a secondary compensation mechanism can sustain the gene expression of these specific genes. There is a possibility that upon disruption of the ARBSs, this transcriptional complex that consists of 4 genes and 5 ARBSs could be disrupted, and other KLK ARBS can re-target themselves to recover the activation of the genes. Combining CRISPRi and CRISPR/Cas9 mediated perturbations, we report the KLK genes that are regulated by proximal and looped ARBSs (Table 4.1).

A recent study demonstrated that S1 KLK genes (*KLK1-4* and *KLK15*) are expressed upon AR activation while S2 (*KLK5-14*) genes are inactive in LNCaP cells [249]. When the boundary CTCF is deleted, S2 genes become active while the whole KLK locus is inactive upon AR activation. We used the same reasoning and targeted the CTCF sites that cover the ARBS1-5 and checked the gene expression

Table 4.1: ARBS and their KLK S1 gene targets.

	ARBS1	**ARBS2**	**ARBS3**	**ARBS4**	**ARBS5**
KLK1	+	+	+	–	+
KLK15	–	+	+	–	+
PSA/KLK3	+	+	+	–	–
KLK2	+	+	+	+	–

of the genes located inside (*PSA/KLK3*, *KLK2*) and outside (*KLK1*, *KLK15*) of the two CTCF binding sites and. There was no significant expression change of the S1 genes, suggesting that not all CTCF sites are functional throughout the locus. This chapter demonstrates that using the KLK locus as a model, genes share ARBS enhancer elements, and AR binding is dependent on other proximal or interacting ARBS. The observed enhancer re-targeting, AR-mediated chromatin loops, multiple AR enhancers for each gene overall suggest a co-operative action of AR-mediated gene transcription.

Chapter 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

In this thesis, we interrogated how AR drives gene expression through enhancer activity at ARBS. Previous work demonstrated that AR-mediated transcription is induced by distal regulatory enhancer elements that loop to gene promoters to control gene expression by [286, 287]. Given that the majority of AR up-regulated genes are in close proximity with multiple ARBS, it is unclear how each binding site contributes to transcription.

To characterize the activity of each ARBS we functionally tested all clinical AR binding sites with a massively multi-parallel enhancer assay to create the first map of AR transcriptional activity. From this we identified three different classes of AR enhancers: inducible AR-driven enhancers, constitutive enhancers that are not impacted by AR binding, and inactive minimally active enhancers sites. AR-inducible enhancers were commonly found near AR-regulated genes and frequently interacted with gene promoters via AR-mediated chromatin looping. We observed that upon DHT treatment, inducible enhancers increase looping to other CREs significantly more than other ARBS. From these results we hypothesized that AR-inducible enhancers act as a regulatory hub that cooperates with other ARBS and CREs to drive transcription.

Yet despite the critical role of inducible AR enhancers, further characterization revealed that all classes of ARBS were required for gene transcription. While the exact role of inactive and constitutive enhancers is an active area of research, our results provide insights into potential mechanisms. Conceivably, these ARBS may simply regulate the expression of alternative gene targets [288]. However, this

seems unlikely as we demonstrated that both constitutively and inactive ARBS are required for AR-driven gene expression. This suggests a secondary role of these enhancers. In transcriptional networks, the DNA binding of a given transcription factor might depend on a second transcription factor to stabilize the interaction. For instance, multiple transcription factors could be bound to a constitutive enhancer in the absence of AR but only selectively activate AR-mediated gene transcription upon binding through induced looping. Alternatively, inactive ARBSs identified by ChIPseq could be shadow peaks. To identify TF binding, ChIP relies on proximal crosslinking [289]. If an ARBS is in physical contact through chromatin looping with a non-AR bound CRE it could potentially crosslink the AR to both the true-positive (ARBS) and false-positive (CRE) binding site. Alternatively, given that STARRseq is an episomal reporter assay that only tests a single region, AR potentially may not bind to these plasmids because of the absence of a chromosomal looping or the local concentration of a specific transcription factor Supporting this, we observed in the KLK locus that AR binding at inactive enhancers was dependent on looped AR inducible enhancers. Further studies are needed to delineate the role of inactive or constitutively active AR enhancers in PCa.

Having identified these different AR enhancer classes, we focused on the KLK gene locus to better understand how AR binding impacts transcription. We systematically inhibited all ARBS and found that most binding sites do not impact KLK gene transcription. However, we identified several ARBS, including 2 inducible, 1 inactive, and 1 constitutive enhancer, that specifically impact the gene expression of the S1 cluster of KLK genes. In these, ARBS2/AREIII regulates KLK3 and KLK2 expression by chromatin looping. We showed that in the absence of the target KLK3 promoter, the ARBS2/AREIII enhancer rewires to increase KLK2 transcriptional output. Further when we inactivated the inducible enhancers ARBS5 there was a significant decrease in AR binding at sites within the regulatory network. Shared regulatory control is a commonly observed phenomenon in genetics, as many genes are activated in such a way [290]. Selective

pressure may lead to diverging control mechanisms where there are multiple regulatory regions that establish stable gene activation. In support of this, we show that AR binding can depend on other ARBS in the same transcriptional network to either stabilize or initiate the DNA interaction. Our data suggests that KLK genes must share ARBS enhancers to synergistically induce transcription.
Altogether, this work provides a comprehensive enhancer annotation of all clinical ARBS and provides fundamental insights of how AR-mediated enhancers drive transcription.

5.2 Future directions

This work has raised numerous questions for further studies. Our CRISPRi experiments demonstrated that all three types of enhancers are required for gene transcription. However, as this has only been studied on a handful of genomic regions, it is unclear if this is unique to the specific genetic loci or a universal feature of nuclear receptor signaling. Large-scale data is needed to understand this critical signaling pathway. Recently, single cell technologies have grown from its infancy and become a much more reliable set of technologies. To be able to interrogate a certain modality in a single cell while doing the perturbation and maintaining the information, it is possible to understand AR and its action of gene activation using different cell line models. Unlike the low-throughput methods such as individual CRISPR inhibition of the specific genes and ARBS, high throughput CRISPR/dCas9-KRAB assays combined with single-cell RNA-seq could be used to identify key ARBSs that affect the most important AR-regulated genes. Furthermore, a novel method can be implemented which involves using STARRseq in single-cell resolution. Coupling with CRISPR perturbations, enhancer activation can be inhibited and the whole transcriptome can be measured upon that change. By doing that for selected AR binding sites in the genome, it is possible to assess enhancer-promoter connectivity and direct effect of an enhancer to a particular gene expression. Furthermore, chromatin accessibility difference can be measured in the genome upon a particular enhancer being inhibited.

H3K27ac histone mark and chromatin looping are considered important markers of enhancers. Other than functional testing, using KO cell lines, 4C-seq to provide the landscape of the looping and H3K27ac ChIP-seq of all ARBS investigated in KLK locus could be performed. With this, it is possible to test whether enhancer action requires certain histone marks or chromatin looping. using a recently reported CTCF binding site that affects the expression of S2 genes [248] and combining this with our STARRseq annotated enhancers, using a CTCF binding site-null LNCaP cell line, AR occupancy alterations could be determined. With that, S2 gene reactivation via AR binding could be investigated. Finally, to expand the KLK gene locus example, other AR-regulated genes (such as *STEAP4*) could be used to investigate enhancer rewiring and change in the AR occupancy.

APPENDIX

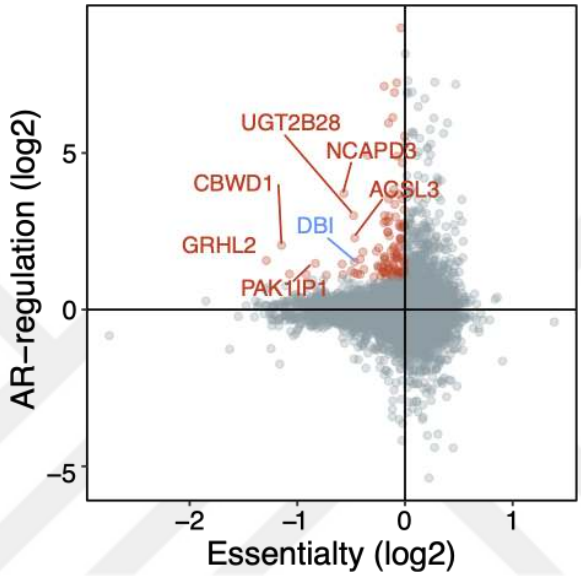


Figure A.1: AR upregulated DBI is essential in LNCaP growth.

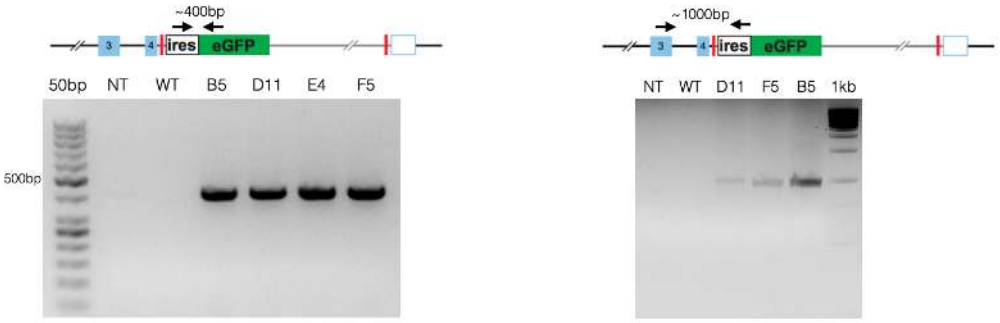


Figure A.2: eGFP knock-in PCR validation. PCRs showing the successful eGFP integration in exon 4 of the *PSA/KLK3* gene. B5, D11, E4 and F5 are the single clones that selected for validation.

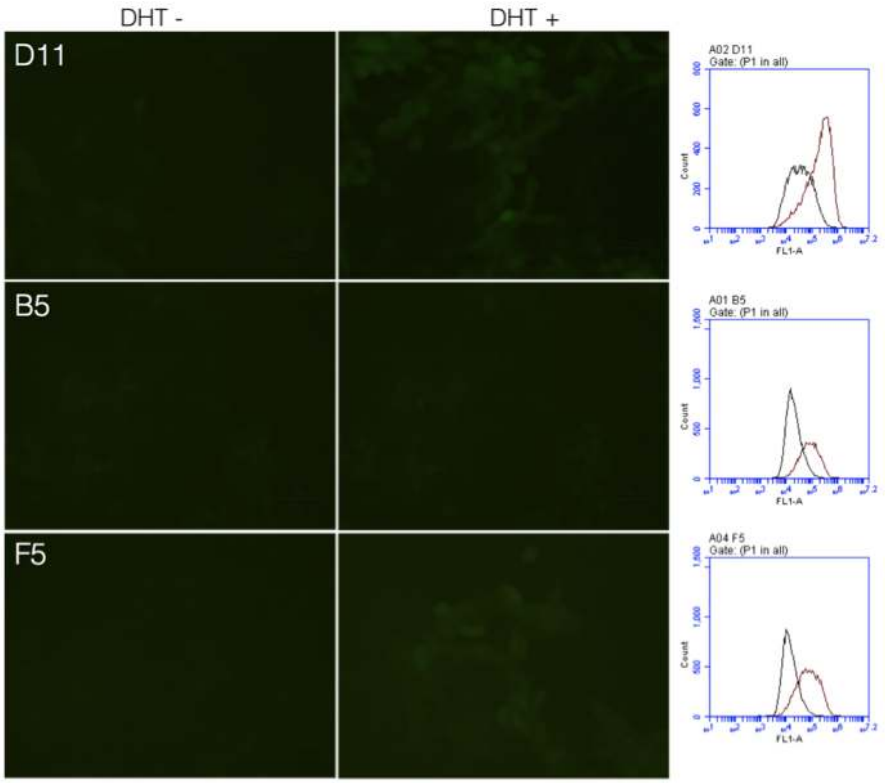


Figure A.3: eGFP signal upon DHT treatment. D11, B5 and F5 single clones are androgen starved for 72 hours and induced with 10nM DHT for 6 hours and then processed for imaging and flow cytometry.

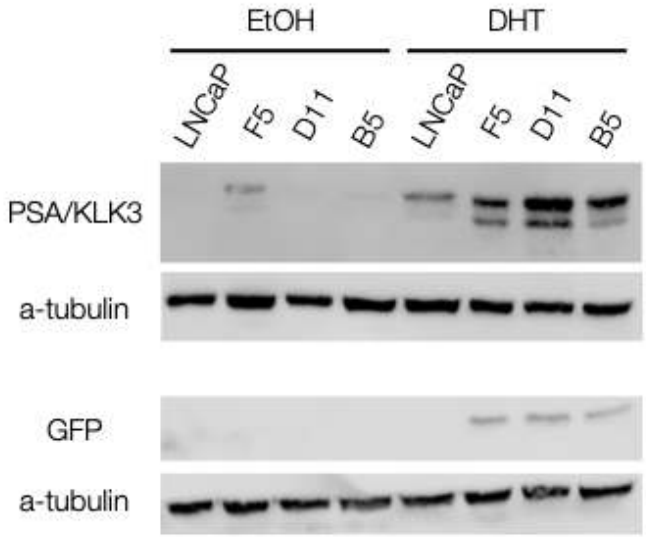


Figure A.4: eGFP protein expression upon DHT treatment. D11, B5 and F5 single clones are androgen starved for 72 hours and induced with 10nM DHT for 6 hours and then processed for GFP and *PSA/KLK3* western blot.

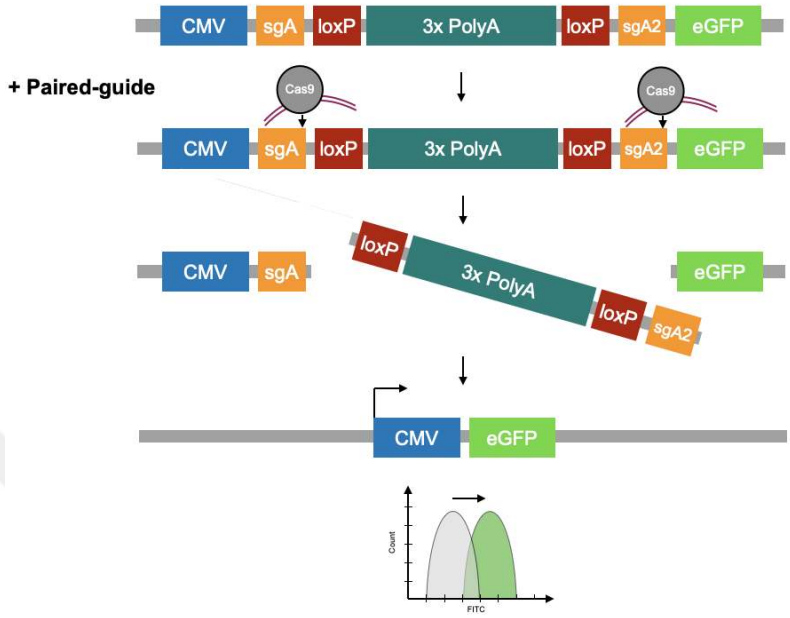


Figure A.5: Schematics of the paired gRNA efficiency reporter assay

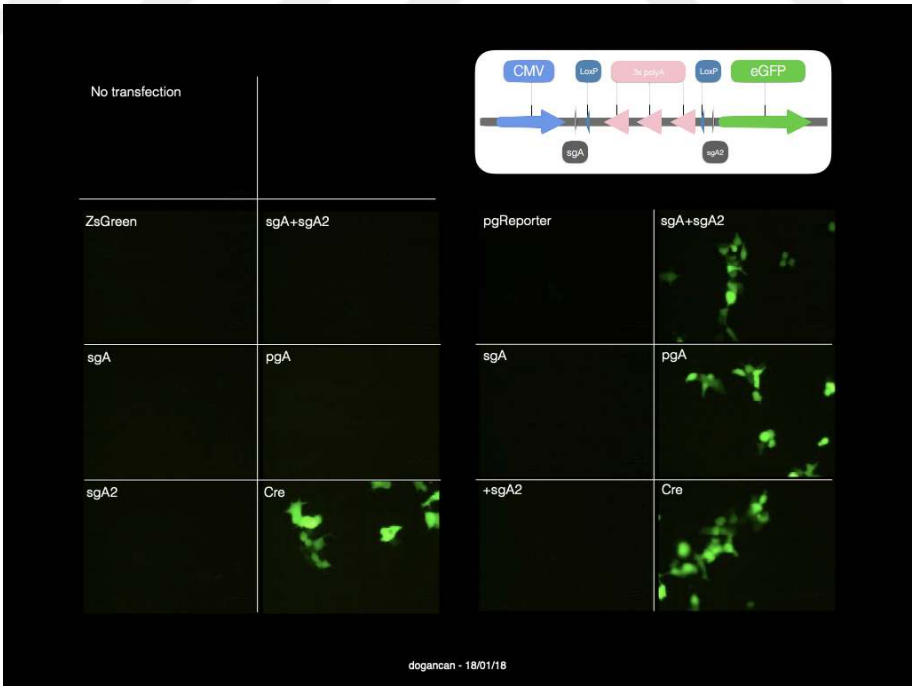
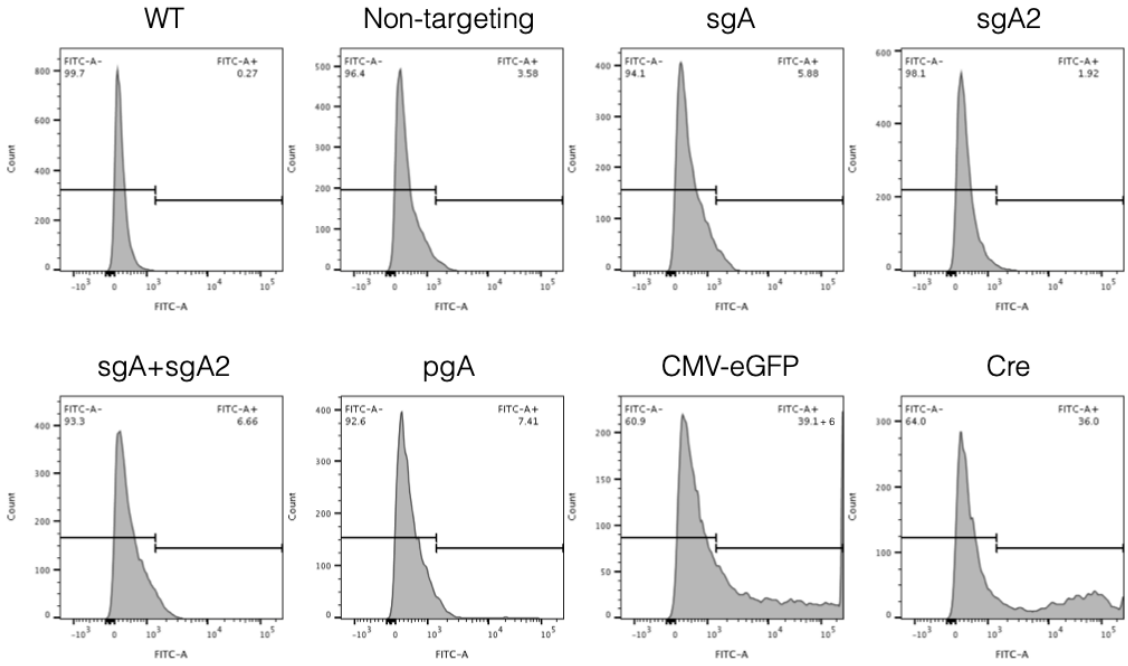
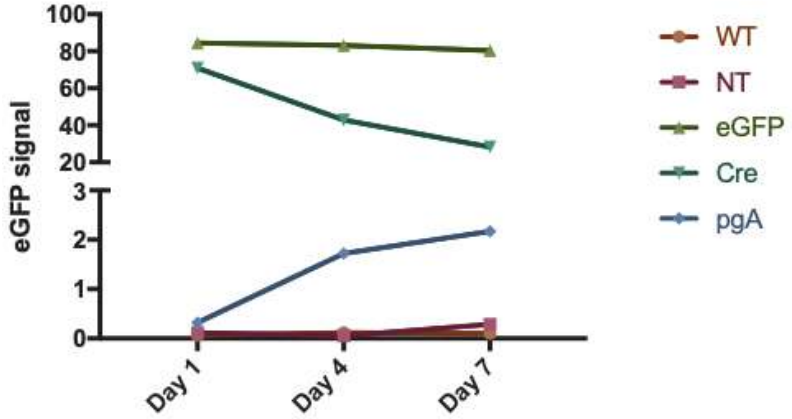


Figure A.6: Paired gRNA efficiency reporter assay. 293T cells are transfected with single gRNAs (sgA, sgA2), paired gRNA (pgA) targeting the same location of the paired-reported cassette. Cre was used as a positive control in both Zsgreen plasmid and paired-guide reporter plasmids which has LoxP-stop-LoxP cassette.



(a)



(b)

Figure A.7: Flow cytometry results for paired gRNA efficiency stable cell line LNCaP cells were transduced with eGFP reporter cassette in low MOI and selected with puromycin. **(a)** Flow cytometry of eGFP reporter cassette upon targeting with both single or paired gRNA. Cre was used for positive control. **(b)** Quantification of eGFP from flow cytometry in different time points upon transfection with shown plasmids.

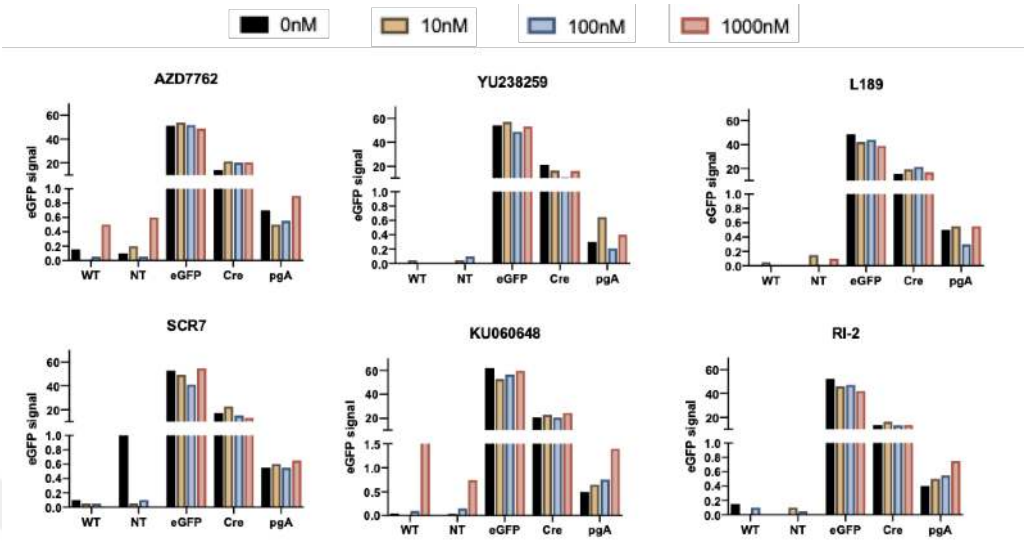


Figure A.8: Increasing deletion efficiency with DNA-repair inhibitors. DNA inhibitors are used to inhibit rapid DNA repair to increase the chance of simultaneous double-stranded break.

Table A.1: DNA-repair pathway inhibitors

Pathway	**Compound**	**Target**
NHEJ	SCR7	Ligase IV
	Ku-0060648	DNA-PK & PI3K
	L189	DNA ligases I, III, and IV
HDR	YU238259	HDR
	AZD 7762	CHK1 & CHK2
	RI-2	RAD51

BIBLIOGRAPHY

[1] H B Sun, J Shen, and H Yokota. Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys. J.*, 79(1):184–190, July 2000.

[2] Luis A Parada, Sotiria Sotiriou, and Tom Misteli. Spatial genome organization. *Exp. Cell Res.*, 296(1):64–70, May 2004.

[3] D W Ross. The human genome: information content and structure. *Hosp. Pract.*, 34(6):49–54, 56–60, 65, June 1999.

[4] Craig Venter and Daniel Cohen. The century of biology. *New Perspect. Q.*, 21(4):73–77, September 2004.

[5] Allison Piovesan, Maria Chiara Pelleri, Francesca Antonaros, Pierluigi Strippoli, Maria Caracausi, and Lorenza Vitale. On the length, weight and GC content of the human genome. *BMC Res. Notes*, 12(1):106, February 2019.

[6] Edward Albert Schäfer. Abstract of three lectures on secretion. *Br. Med. J.*, 1(1278):1287–1289, June 1885.

[7] T J Richmond, J T Finch, B Rushton, D Rhodes, and A Klug. Structure of the nucleosome core particle at 7 a resolution. *Nature*, 311(5986):532–537, 1984.

[8] Jesse R Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, April 2012.

[9] Hui Zheng and Wei Xie. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.*, 20(9):535–550, September 2019.

[10] Tsung-Han S Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S Hansen, Oliver J Rando, Robert Tjian, and Xavier Darzacq. Resolving the 3D landscape of Transcription-Linked mammalian chromatin folding. *Mol. Cell*, 78(3):539–553.e8, May 2020.

[11] Nils Krietenstein, Sameer Abraham, Sergey V Venev, Nezar Abdennur, Johan Gibcus, Tsung-Han S Hsieh, Krishna Mohan Parsi, Liyan Yang, René Maehr, Leonid A Mirny, Job Dekker, and Oliver J Rando. Ultrastructural details of mammalian chromosome architecture. *Mol. Cell*, 78(3):554–565.e7, May 2020.

[12] Elphège P Nora, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L van Berkum, Johannes Meisig, John Sedat, Joost Gribnau, Emmanuel Barillot, Nils Blüthgen, Job Dekker, and Edith Heard. Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485(7398):381–385, April 2012.

[13] Jan Krefting, Miguel A Andrade-Navarro, and Jonas Ibn-Salem. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol.*, 16(1):87, August 2018.

[14] M Jordan Rowley and Victor G Corces. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.*, 19(12):789–800, 2018.

[15] Orsolya Symmons, Veli Vural Uslu, Taro Tsujimura, Sandra Ruf, Sonya Nassari, Wibke Schwarzer, Laurence Ettwiller, and François Spitz. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.*, 24(3):390–400, March 2014.

[16] Johannes Nuebler, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid A Mirny. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U. S. A.*, 115(29):E6697–E6706, July 2018.

[17] Suhas S P Rao, Su Chen Huang, Brian Glenn St Hilaire, Jesse M Engreitz, Elizabeth M Perez, Kyong Rim Kieffer-Kwon, Adrian L Sanborn, Sarah E Johnstone, Gavin D Bascom, Ivan D Bochkov, Xingfan Huang, Muhammad S Shamim, Jaeweon Shin, Douglass Turner, Ziyi Ye, Arina D Omer, James T Robinson, Tamar Schlick, Bradley E Bernstein, Rafael Casellas, Eric S Lander, and Erez Lieberman Aiden. Cohesin loss eliminates all loop domains. *Cell*, 171(2):305–320.e24, October 2017.

[18] Michele Gabriele, Hugo B Brandão, Simon Grosse-Holz, Asmita Jha, Gina M Dailey, Claudia Cattoglio, Tsung-Han S Hsieh, Leonid Mirny, Christoph Zechner, and Anders Sejr Hansen. Dynamics of CTCF and cohesin mediated chromatin looping revealed by live-cell imaging. December 2021.

[19] Britta A M Bouwman and Wouter de Laat. Getting the genome in shape: the formation of loops, domains and compartments. *Genome Biol.*, 16:154, August 2015.

[20] Erez Lieberman-Aiden, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, Richard Sandstrom, Bradley Bernstein, M A Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A Mirny, Eric S Lander, and Job Dekker. Comprehensive mapping

of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, October 2009.

[21] Lars Guelen, Ludo Pagie, Emilie Brasset, Wouter Meuleman, Marius B Faza, Wendy Talhout, Bert H Eussen, Annelies de Klein, Lodewyk Wessels, Wouter de Laat, and Bas van Steensel. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453(7197):948–951, June 2008.

[22] Tyrone Ryba, Ichiro Hiratani, Junjie Lu, Mari Itoh, Michael Kulik, Jinfeng Zhang, Thomas C Schulz, Allan J Robins, Stephen Dalton, and David M Gilbert. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.*, 20(6):761–770, June 2010.

[23] Eitan Yaffe, Shlomit Farkash-Amar, Andreas Polten, Zohar Yakhini, Amos Tanay, and Itamar Simon. Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS Genet.*, 6(7):e1001011, July 2010.

[24] Benjamin D Pope, Tamir Chandra, Quinton Buckley, Matthew Hoare, Tyrone Ryba, Frances K Wiseman, Anna Kuta, Michael D Wilson, Duncan T Odom, and David M Gilbert. Replication-timing boundaries facilitate cell-type and species-specific regulation of a rearranged human chromosome in mouse. *Hum. Mol. Genet.*, 21(19):4162–4170, October 2012.

[25] Jop Kind, Ludo Pagie, Sandra S de Vries, Leila Nahidiazar, Siddharth S Dey, Magda Bienko, Ye Zhan, Bryan Lajoie, Carolyn A de Graaf, Mario Amendola, Geoffrey Fudenberg, Maxim Imakaev, Leonid A Mirny, Kees Jalink, Job Dekker, Alexander van Oudenaarden, and Bas van Steensel. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell*, 163(1):134–147, September 2015.

[26] Irina Solovei, Katharina Thanisch, and Yana Feodorova. How to rule the nucleus: divide et impera. *Curr. Opin. Cell Biol.*, 40:47–59, June 2016.

[27] Chien-Hui Chuang and Andrew S Belmont. Moving chromatin within the interphase nucleus-controlled transitions? *Semin. Cell Dev. Biol.*, 18(5):698–706, October 2007.

[28] Pierre Therizols, Robert S Illingworth, Celine Courilleau, Shelagh Boyle, Andrew J Wood, and Wendy A Bickmore. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science*, December 2014.

[29] C Francastel, D Schübeler, D I Martin, and M Groudine. Nuclear compartmentalization and gene activity. *Nat. Rev. Mol. Cell Biol.*, 1(2):137–143, November 2000.

[30] RABL and C. Uber zelltheilung. *Morphol. Jahrb.*, 10:214–330, 1885.

[31] Th Boveri. Die blastomerenkerne von ascaris megalocephala. *Archiv für Zellforschung*, 3:181, 1909.

[32] Thomas Cremer and Marion Cremer. Chromosome territories. *Cold Spring Harb. Perspect. Biol.*, 2(3):a003889, March 2010.

[33] Leslie J Mateo, Sedona E Murphy, Antonina Hafner, Isaac S Cinquini, Carly A Walker, and Alistair N Boettiger. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, 568(7750):49–54, 2019.

[34] Elizabeth H Finn, Gianluca Pegoraro, Hugo B Brandão, Anne-Laure Valton, Marlies E Oomen, Job Dekker, Leonid Mirny, and Tom Misteli. Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell*, 176(6):1502–1515.e10, March 2019.

[35] Bogdan Bintu, Leslie J Mateo, Jun-Han Su, Nicholas A Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N Boettiger, and Xiaowei Zhuang. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, October 2018.

[36] Jun-Han Su, Pu Zheng, Seon S Kinrot, Bogdan Bintu, and Xiaowei Zhuang. Genome-Scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell*, 182(6):1641–1659.e26, September 2020.

[37] Christophe Benoist and Pierre Chambon. In vivo sequence requirements of the SV40 early promoter region, 1981.

[38] P Gruss, R Dhar, and G Khoury. Simian virus 40 tandem repeated sequences as an element of the early promoter. *Proc. Natl. Acad. Sci. U. S. A.*, 78(2):943–947, February 1981.

[39] Diego Villar, Camille Berthelot, Sarah Aldridge, Tim F Rayner, Margus Lukk, Miguel Pignatelli, Thomas J Park, Robert Deaville, Jonathan T Erichsen, Anna J Jasinska, James M A Turner, Mads F Bertelsen, Elizabeth P Murchison, Paul Flicek, and Duncan T Odom. Enhancer evolution across 20 mammalian species. *Cell*, 160(3):554–566, January 2015.

[40] Tobias Jores, Jackson Tonnies, Michael W Dorrity, Josh T Cuperus, Stanley Fields, and Christine Queitsch. Identification of plant enhancers and their

constituent elements by STARR-seq in tobacco leaves. *Plant Cell*, 32(7):2120–2131, July 2020.

[41] Rurika Oka, Johan Zicola, Blaise Weber, Sarah N Anderson, Charlie Hodgman, Jonathan I Gent, Jan-Jaap Wesselink, Nathan M Springer, Huub C J Hoefsloot, Franziska Turck, and Maike Stam. Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.*, 18(1):137, July 2017.

[42] Alexandre A Boudreault, Dominique Cronier, William Selleck, Nicolas Lacoste, Rhea T Utley, Stéphane Allard, Julie Savard, William S Lane, Song Tan, and Jacques Côté. Yeast enhancer of polycomb defines global esa1-dependent acetylation of chromatin. *Genes Dev.*, 17(11):1415–1428, June 2003.

[43] Arnau Sebé-Pedrós, Cecilia Ballaré, Helena Parra-Acero, Cristina Chiva, Juan J Tena, Eduard Sabidó, José Luis Gómez-Skarmeta, Luciano Di Croce, and Iñaki Ruiz-Trillo. The dynamic regulatory genome of capsaspora and the origin of animal multicellularity. *Cell*, 165(5):1224–1237, May 2016.

[44] ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, September 2012.

[45] Martin S C Larke, Ron Schwessinger, Takayuki Nojima, Jelena Telenius, Robert A Beagrie, Damien J Downes, A Marieke Oudelaar, Julia Truch, Bryony Graham, M A Bender, Nicholas J Proudfoot, Douglas R Higgs, and Jim R Hughes. Enhancers predominantly regulate gene expression during differentiation via transcription initiation. *Mol. Cell*, 81(5):983–997.e7, March 2021.

[46] Nathaniel D Heintzman, Gary C Hon, R David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F Harp, Zhen Ye, Leonard K Lee, Rhona K Stuart, Christina W Ching, Keith A Ching, Jessica E Antosiewicz-Bourget, Hui Liu, Xinmin Zhang, Roland D Green, Victor V Lobanenkov, Ron Stewart, James A Thomson, Gregory E Crawford, Manolis Kellis, and Bing Ren. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243):108–112, May 2009.

[47] Thomas Sandmann, Charles Girardot, Marc Brehme, Waraporn Tongprasit, Viktor Stolc, and Eileen E M Furlong. A core transcriptional network for early mesoderm development in drosophila melanogaster. *Genes Dev.*, 21(4):436–449, February 2007.

[48] Benjamin P Berman, Barret D Pfeiffer, Todd R Laverty, Steven L Salzberg, Gerald M Rubin, Michael B Eisen, and Susan E Celniker. Computational identification of developmental enhancers: conservation and function of

transcription factor binding-site clusters in drosophila melanogaster and drosophila pseudoobscura. *Genome Biol.*, 5(9):R61, August 2004.

[49] Petter Hammar, Prune Leroy, Anel Mahmutovic, Erik G Marklund, Otto G Berg, and Johan Elf. The lac repressor displays facilitated diffusion in living cells. *Science*, 336(6088):1595–1598, June 2012.

[50] Johan Elf, Gene-Wei Li, and X Sunney Xie. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316(5828):1191–1194, May 2007.

[51] Robert B Winter and Peter H Von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. the escherichia coli lac repressor-operator interaction: equilibrium measurements. *Biochemistry*, 20(24):6948–6960, November 1981.

[52] Salvatore Spicuglia and Laurent Vanhille. Chromatin signatures of active enhancers. *Nucleus*, 3(2):126–131, March 2012.

[53] Alan P Boyle, Sean Davis, Hennady P Shulha, Paul Meltzer, Elliott H Margulies, Zhiping Weng, Terrence S Furey, and Gregory E Crawford. High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 132(2):311–322, January 2008.

[54] Iros Barozzi, Marta Simonatto, Silvia Bonifacio, Lin Yang, Remo Rohs, Serena Ghisletti, and Gioacchino Natoli. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol. Cell*, 54(5):844–857, June 2014.

[55] Outi Hallikas, Kimmo Palin, Natalia Sinjushina, Reetta Rautiainen, Juha Partanen, Esko Ukkonen, and Jussi Taipale. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell*, 124(1):47–59, January 2006.

[56] Nasun Hah, Shino Murakami, Anusha Nagari, Charles G Danko, and W Lee Kraus. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.*, 23(8):1210–1223, 2013.

[57] Chia-Chi Flora Huang, Shreyas Lingadahalli, Tunc Morova, Dogancan Ozturan, Eugene Hu, Ivan Pak Lok Yu, Simon Linder, Marlous Hoogstraat, Suzan Stelloo, Funda Sar, Henk van der Poel, Umut Berkay Altintas, Mohammadali Saffarzadeh, Stephane Le Bihan, Brian McConeghy, Bengul Gokbayrak, Felix Y Feng, Martin E Gleave, Andries M Bergman, Colin Collins, Faraz Hach, Wilbert Zwart, Eldon Emberly, and Nathan A Lack. Functional mapping of androgen receptor enhancer activity. *Genome Biol.*, 22(1):149, May 2021.

[58] Ian C McDowell, Alejandro Barrera, Anthony M D'Ippolito, Christopher M Vockley, Linda K Hong, Sarah M Leichter, Luke C Bartelt, William H Majoros, Lingyun Song, Alexias Safi, D Dewran Koçak, Charles A Gersbach, Alexander J Hartemink, Gregory E Crawford, Barbara E Engelhardt, and Timothy E Reddy. Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding. *Genome Res.*, 28(9):1272–1284, September 2018.

[59] Shiqi Xie, Jialei Duan, Boxun Li, Pei Zhou, and Gary C Hon. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol. Cell*, 66(2):285–299.e5, April 2017.

[60] Yoon Hee Jung, Hsiao-Lin V Wang, Daniel Ruiz, Fiorella C Grandi, Brianna J Bixler, Hannah Linsenbaum, Jian-Feng Xiang, Samantha Forestier, Andrew M Shafik, Peng Jin, and Others. Recruitment of CTCF to an fto enhancer is responsible for transgenerational inheritance of obesity. *bioRxiv*, pages 2020–2011, 2020.

[61] François Spitz and Eileen E M Furlong. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, 13(9):613–626, September 2012.

[62] Jian Yan, Martin Enge, Thomas Whitington, Kashyap Dave, Jianping Liu, Inderpreet Sur, Bernhard Schmierer, Arttu Jolma, Teemu Kivioja, Minna Taipale, and Jussi Taipale. Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell*, 154(4):801–813, August 2013.

[63] Chia-Lin Wei, Qiang Wu, Vinsensius B Vega, Kuo Ping Chiu, Patrick Ng, Tao Zhang, Atif Shahab, How Choong Yong, Yutao Fu, Zhiping Weng, Jianjun Liu, Xiao Dong Zhao, Joon-Lin Chew, Yen Ling Lee, Vladimir A Kuznetsov, Wing-Kin Sung, Lance D Miller, Bing Lim, Edison T Liu, Qiang Yu, Huck-Hui Ng, and Yijun Ruan. A global map of p53 transcription-factor binding sites in the human genome. *Cell*, 124(1):207–219, January 2006.

[64] Christopher M Vockley, Anthony M D'Ippolito, Ian C McDowell, William H Majoros, Alexias Safi, Lingyun Song, Gregory E Crawford, and Timothy E Reddy. Direct GR binding sites potentiate clusters of TF binding across the human genome. *Cell*, 166(5):1269–1281.e19, 2016.

[65] Douglas Vernimmen, Marco De Gobbi, Jacqueline A Sloane-Stanley, William G Wood, and Douglas R Higgs. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.*, 26(8):2041–2051, April 2007.

[66] Frederic Koch, Romain Fenouil, Marta Gut, Pierre Cauchy, Thomas K Albert, Joaquin Zacarias-Cabeza, Salvatore Spicuglia, Albane Lamy de la Chapelle, Martin Heidemann, Corinna Hintermair, Dirk Eick, Ivo Gut, Pierre Ferrier, and Jean-Christophe Andrau. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.*, 18(8):956–963, July 2011.

[67] Steven Hahn. Structure and mechanism of the RNA polymerase II transcription machinery, 2004.

[68] Eliezer Calo and Joanna Wysocka. Modification of enhancer chromatin: what, how, and why? *Mol. Cell*, 49(5):825–837, March 2013.

[69] Wenbo Li, Yiren Hu, Soohwan Oh, Qi Ma, Daria Merkurjev, Xiaoyuan Song, Xiang Zhou, Zhijie Liu, Bogdan Tanasa, Xin He, Aaron Yun Chen, Kenny Ohgi, Jie Zhang, Wen Liu, and Michael G Rosenfeld. Condensin I and II complexes license full estrogen receptor α-Dependent enhancer activation. *Mol. Cell*, 59(2):188–202, July 2015.

[70] Gang Ren, Wenfei Jin, Kairong Cui, Joseph Rodrigez, Gangqing Hu, Zhiying Zhang, Daniel R Larson, and Keji Zhao. CTCF-Mediated Enhancer-Promoter interaction is a critical regulator of Cell-to-Cell variation of gene expression. *Mol. Cell*, 67(6):1049–1058.e6, September 2017.

[71] Abraham S Weintraub, Charles H Li, Alicia V Zamudio, Alla A Sigova, Nancy M Hannett, Daniel S Day, Brian J Abraham, Malkiel A Cohen, Behnam Nabet, Dennis L Buckley, Yang Eric Guo, Denes Hnisz, Rudolf Jaenisch, James E Bradner, Nathanael S Gray, and Richard A Young. YY1 is a structural regulator of Enhancer-Promoter loops. *Cell*, 171(7):1573–1588.e28, December 2017.

[72] Benjamin L Allen and Dylan J Taatjes. The mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.*, 16(3):155–166, March 2015.

[73] Célia Jeronimo and François Robert. The mediator complex: At the nexus of RNA polymerase II transcription. *Trends Cell Biol.*, 27(10):765–783, October 2017.

[74] Fan Lai, Alessandro Gardini, Anda Zhang, and Ramin Shiekhattar. Integrator mediates the biogenesis of enhancer RNAs. *Nature*, 525(7569):399–403, September 2015.

[75] Burak H Alver, Kimberly H Kim, Ping Lu, Xiaofeng Wang, Haley E Manchester, Weishan Wang, Jeffrey R Haswell, Peter J Park, and Charles

W M Roberts. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.*, 8:14648, March 2017.

[76] M Yu Mazina and N E Vorobyeva. Chromatin modifiers in transcriptional regulation: New findings and prospects. *Acta Naturae*, 13(1):16–30, January 2021.

[77] Shino Murakami, Anusha Nagari, and W Lee Kraus. Dynamic assembly and activation of estrogen receptor α enhancers through coregulator switching. *Genes Dev.*, 31(15):1535–1548, August 2017.

[78] Song Liu, Sangeeta Kumari, Qiang Hu, Dhirodatta Senapati, Varadha Balaji Venkadakrishnan, Dan Wang, Adam D DePriest, Simon E Schlanger, Salma Ben-Salem, Malyn May Valenzuela, Belinda Willard, Shaila Mudambi, Wendy M Swetzig, Gokul M Das, Mojgan Shourideh, Shahriah Koochekpour, Sara Moscovita Falzarano, Cristina Magi-Galluzzi, Neelu Yadav, Xiwei Chen, Changshi Lao, Jianmin Wang, Jean-Noel Billaud, and Hannelore V Heemers. A comprehensive analysis of coregulator recruitment, androgen receptor function and gene expression in prostate cancer. *Elife*, 6:e28482, August 2017.

[79] Charles E Foulds, Qin Feng, Chen Ding, Suzanna Bailey, Tamra L Hunsaker, Anna Malovannaya, Ross A Hamilton, Leah A Gates, Zheng Zhang, Chunshu Li, Doug Chan, Amol Bajaj, Celetta G Callaway, Dean P Edwards, David M Lonard, Sophia Y Tsai, Ming-Jer Tsai, Jun Qin, and Bert W O'Malley. Proteomic analysis of coregulators bound to ERα on DNA and nucleosomes reveals coregulator dynamics. *Mol. Cell*, 51(2):185–199, July 2013.

[80] Sean D Taverna, Haitao Li, Alexander J Ruthenburg, C David Allis, and Dinshaw J Patel. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.*, 14(11):1025–1040, November 2007.

[81] Evangelos Pefanis, Jiguang Wang, Gerson Rothschild, Junghyun Lim, David Kazadi, Jianbo Sun, Alexander Federation, Jaime Chao, Oliver Elliott, Zhi-Ping Liu, Aris N Economides, James E Bradner, Raul Rabadan, and Uttiya Basu. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell*, 161(4):774–789, May 2015.

[82] W Lee Kraus. Transcriptional control by PARP-1: chromatin modulation, enhancer-binding, coregulation, and insulation. *Curr. Opin. Cell Biol.*, 20(3):294–302, June 2008.

[83] Janusz Puc, Piotr Kozbial, Wenbo Li, Yuliang Tan, Zhijie Liu, Tom Suter, Kenneth A Ohgi, Jie Zhang, Aneel K Aggarwal, and Michael G Rosenfeld.

Ligand-dependent enhancer activation regulated by topoisomerase-i activity. *Cell*, 160(3):367–380, January 2015.

[84] B D Strahl and C D Allis. The language of covalent histone modifications. *Nature*, 403(6765):41–45, January 2000.

[85] B M Turner. Histone acetylation and an epigenetic code. *Bioessays*, 22(9):836–845, September 2000.

[86] H Harris. Turnover of nuclear and cytoplasmic ribonucleic acid in two types of animal cell, with some further observations on the nucleolus. *Biochem. J*, 73:362–369, October 1959.

[87] P Collis, M Antoniou, and F Grosveld. Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression, 1990.

[88] Gioacchino Natoli and Jean-Christophe Andrau. Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.*, 46:1–19, August 2012.

[89] Robin Andersson, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, Xiaobei Zhao, Christian Schmidl, Takahiro Suzuki, Evgenia Ntini, Erik Arner, Eivind Valen, Kang Li, Lucia Schwarzfischer, Dagmar Glatz, Johanna Raithel, Berit Lilje, Nicolas Rapin, Frederik Otzen Bagger, Mette Jørgensen, Peter Refsing Andersen, Nicolas Bertin, Owen Rackham, A Maxwell Burroughs, J Kenneth Baillie, Yuri Ishizu, Yuri Shimizu, Erina Furuhata, Shiori Maeda, Yutaka Negishi, Christopher J Mungall, Terrence F Meehan, Timo Lassmann, Masayoshi Itoh, Hideya Kawaji, Naoto Kondo, Jun Kawai, Andreas Lennartsson, Carsten O Daub, Peter Heutink, David A Hume, Torben Heick Jensen, Harukazu Suzuki, Yoshihide Hayashizaki, Ferenc Müller, Alistair R R Forrest, Piero Carninci, Michael Rehli, and Albin Sandelin. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493):455–461, March 2014.

[90] Laura A Lettice, Simon J H Heaney, Lorna A Purdie, Li Li, Philippe de Beer, Ben A Oostra, Debbie Goode, Greg Elgar, Robert E Hill, and Esther de Graaff. A long-range shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.*, 12(14):1725–1735, July 2003.

[91] S Kong, D Bohl, C Li, and D Tuan. Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol. Cell. Biol.*, 17(7):3955–3965, July 1997.

[92] Michael Petrascheck, Dominik Escher, Tokameh Mahmoudi, C Peter Verrijzer, Walter Schaffner, and Alcide Barberis. DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Res.*, 33(12):3743–3750, July 2005.

[93] Nezha S Benabdallah, Iain Williamson, Robert S Illingworth, Shelagh Boyle, Graeme R Grimes, Pierre Therizols, and Wendy A Bickmore. PARP mediated chromatin unfolding is coupled to long-range enhancer activation. June 2017.

[94] Guillaume Andrey, Thomas Montavon, Bénédicte Mascrez, Federico Gonzalez, Daan Noordermeer, Marion Leleu, Didier Trono, François Spitz, and Denis Duboule. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science*, 340(6137):1234167, June 2013.

[95] Charles P Fulco, Joseph Nasser, Thouis R Jones, Glen Munson, Drew T Bergman, Vidya Subramanian, Sharon R Grossman, Rockwell Anyoha, Benjamin R Doughty, Tejal A Patwardhan, Tung H Nguyen, Michael Kane, Elizabeth M Perez, Neva C Durand, Caleb A Lareau, Elena K Stamenova, Erez Lieberman Aiden, Eric S Lander, and Jesse M Engreitz. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations, 2019.

[96] Yarui Diao, Rongxin Fang, Bin Li, Zhipeng Meng, Juntao Yu, Yunjiang Qiu, Kimberly C Lin, Hui Huang, Tristin Liu, Ryan J Marina, Inkyung Jung, Yin Shen, Kun-Liang Guan, and Bing Ren. A tiling-deletion-based genetic screen for cis -regulatory element identification in mammalian cells. *Nat. Methods*, 14(6):629–635, April 2017.

[97] Xianjun Dong, Pavla Navratilova, David Fredman, Øyvind Drivenes, Thomas S Becker, and Boris Lenhard. Exonic remnants of whole-genome duplication reveal cis-regulatory function of coding exons. *Nucleic Acids Res.*, 38(4):1071–1085, March 2010.

[98] Andrej A Arsovski, Julian Pradinuk, Xu Qiu Guo, Sishuo Wang, and Keith L Adams. Evolution of Cis-Regulatory elements and regulatory networks in duplicated genes of arabidopsis. *Plant Physiol.*, 169(4):2982–2991, December 2015.

[99] J R Stone and G A Wray. Rapid evolution of cis-regulatory sequences via local point mutations. *Mol. Biol. Evol.*, 18(9):1764–1770, September 2001.

[100] Michael P Eichenlaub and Laurence Ettwiller. De novo genesis of enhancers in vertebrates. *PLoS Biol.*, 9(11):e1001188, November 2011.

[101] Zefu Lu, Alexandre P Marand, William A Ricci, Christina L Ethridge, Xiaoyu Zhang, and Robert J Schmitz. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants*, 5(12):1250–1259, December 2019.

[102] David Santiago-Algarra, Charbel Souaid, Himanshu Singh, Lan T M Dao, Saadat Hussain, Alejandra Medina-Rivera, Lucia Ramirez-Navarro, Jaime A Castro-Mondragon, Nori Sadouni, Guillaume Charbonnier, and Salvatore Spicuglia. Epromoters function as a hub to recruit key transcription factors required for the inflammatory response. *Nat. Commun.*, 12(1):6660, November 2021.

[103] R Kay, A Chan, M Daly, and J McPherson. Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. *Science*, 236(4806):1299–1302, June 1987.

[104] Robin Andersson and Albin Sandelin. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.*, 21(2):71–87, February 2020.

[105] Marc A J Morgan and Ali Shilatifard. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nat. Genet.*, 52(12):1271–1281, December 2020.

[106] Tiantian Zhang, Zhuqiang Zhang, Qiang Dong, Jun Xiong, and Bing Zhu. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol.*, 21(1):45, February 2020.

[107] Marcela Rojas-Pierce and Patricia S Springer. Gene and enhancer traps for gene discovery. *Methods Mol. Biol.*, 236:221–240, 2003.

[108] L Manseau, A Baradaran, D Brower, A Budhu, F Elefant, H Phan, A V Philp, M Yang, D Glover, K Kaiser, K Palter, and S Selleck. GAL4 enhancer traps expressed in the embryo, larval brain, imaginal discs, and ovary of drosophila. *Dev. Dyn.*, 209(3):310–322, July 1997.

[109] Jennifer Hammelman, Konstantin Krismer, Budhaditya Banerjee, David K Gifford, and Richard I Sherwood. Identification of determinants of differential chromatin accessibility through a massively parallel genome-integrated reporter assay. *Genome Res.*, 30(10):1468–1480, October 2020.

[110] Kailong Li, Yuxuan Liu, Hui Cao, Yuannyu Zhang, Zhimin Gu, Xin Liu, Andy Yu, Pranita Kaphle, Kathryn E Dickerson, Min Ni, and Jian Xu. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat. Commun.*, 11(1):1–16, 2020.

[111] Nan Cher Yeo, Alejandro Chavez, Alissa Lance-Byrne, Yingleong Chan, David Menn, Denitsa Milanova, Chih-Chung Kuo, Xiaoge Guo, Sumana Sharma, Angela Tung, Ryan J Cecchi, Marcelle Tuttle, Swechchha Pradhan, Elaine T Lim, Noah Davidsohn, Mo R Ebrahimkhani, James J Collins, Nathan E Lewis, Samira Kiani, and George M Church. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods*, 15(8):611–616, August 2018.

[112] Nader Alerasool, Dmitri Segal, Hunsang Lee, and Mikko Taipale. An efficient KRAB domain for CRISPRi applications in human cells. *Nat. Methods*, 17(11):1093–1096, 2020.

[113] Thomas J Cunningham, Joseph J Lancman, Marie Berenguer, P Duc Si Dong, and Gregg Duester. Genomic knockout of two presumed forelimb tbx5 enhancers reveals they are nonessential for limb development. *Cell Rep.*, 23(11):3146–3151, June 2018.

[114] Stephen T Smale. Luciferase assay. *Cold Spring Harb. Protoc.*, 2010(5):db.prot5421, May 2010.

[115] W D McElroy. A history of luminescence from the earliest times until 1900 . vol. 44 of memoirs of the american philosophical society. e. newton harvey. american philosophical society, philadelphia, pa., 1957. xiii + 692 pp, plates. $6. *Science*, 126(3278):847–848, October 1957.

[116] Rupali P Patwardhan, Choli Lee, Oren Litvin, David L Young, Dana Pe'er, and Jay Shendure. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.*, 27(12):1173–1175, December 2009.

[117] Rupali P Patwardhan, Joseph B Hiatt, Daniela M Witten, Mee J Kim, Robin P Smith, Dalit May, Choli Lee, Jennifer M Andrie, Su-In Lee, Gregory M Cooper, Nadav Ahituv, Len A Pennacchio, and Jay Shendure. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.*, 30(3):265–270, February 2012.

[118] Jamie C Kwasnieski, Ilaria Mogno, Connie A Myers, Joseph C Corbo, and Barak A Cohen. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl. Acad. Sci. U. S. A.*, 109(47):19498–19503, November 2012.

[119] Cosmas D Arnold, Daniel Gerlach, Christoph Stelzer, Łukasz M Boryń, Martina Rath, and Alexander Stark. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, 339(6123):1074–1077, March 2013.

[120] Hansol Choi, Yeongjae Choi, Jaewon Choi, Amos Chungwon Lee, Huiran Yeom, Jinwoo Hyun, Taehoon Ryu, and Sunghoon Kwon. Purification of multiplex oligonucleotide libraries by synthesis and selection. *Nat. Biotechnol.*, July 2021.

[121] Laurent Vanhille, Aurélien Griffon, Muhammad Ahmad Maqbool, Joaquin Zacarias-Cabeza, Lan T M Dao, Nicolas Fernandez, Benoit Ballester, Jean Christophe Andrau, and Salvatore Spicuglia. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat. Commun.*, 6:6905, April 2015.

[122] Xinchen Wang, Liang He, Sarah M Goggin, Alham Saadat, Li Wang, Nasa Sinnott-Armstrong, Melina Claussnitzer, and Manolis Kellis. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat. Commun.*, 9(1):1–15, 2018.

[123] Jason C Klein, Vikram Agarwal, Fumitaka Inoue, Aidan Keith, Beth Martin, Martin Kircher, Nadav Ahituv, and Jay Shendure. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat. Methods*, 17(11):1083–1091, November 2020.

[124] Felix Muerdter, Łukasz M Boryń, Ashley R Woodfin, Christoph Neumayr, Martina Rath, Muhammad A Zabidi, Michaela Pagani, Vanja Haberle, Tomáš Kazmar, Rui R Catarino, Katharina Schernhuber, Cosmas D Arnold, and Alexander Stark. Resolving systematic errors in widely used enhancer activity assays in human cells, 2018.

[125] Drew T Bergman, Thouis R Jones, Vincent Liu, Layla Siraj, Helen Y Kang, Joseph Nasser, Michael Kane, Tung H Nguyen, Sharon R Grossman, Charles P Fulco, Eric S Lander, and Jesse M Engreitz. Compatibility logic of human enhancer and promoter sequences. October 2021.

[126] Ilias K Nolis, Daniel J McKay, Eva Mantouvalou, Stavros Lomvardas, Menie Merika, and Dimitris Thanos. Transcription factors mediate long-range enhancer–promoter interactions. *Proc. Natl. Acad. Sci. U. S. A.*, 106(48):20222–20227, December 2009.

[127] Amartya Sanyal, Bryan R Lajoie, Gaurav Jain, and Job Dekker. The long-range interaction landscape of gene promoters. *Nature*, 489(7414):109–113, September 2012.

[128] J G Gall and M L Pardue. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc. Natl. Acad. Sci. U. S. A.*, 63(2):378–383, June 1969.

[129] Jeffrey M Levsky and Robert H Singer. Fluorescence in situ hybridization: past, present and future. *J. Cell Sci.*, 116(Pt 14):2833–2838, July 2003.

[130] Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–1311, February 2002.

[131] Peter H L Krijger, Geert Geeven, Valerio Bianchi, Catharina R E Hilvering, and Wouter de Laat. 4c-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods*, 170:17–32, January 2020.

[132] Marieke Simonis, Petra Klous, Erik Splinter, Yuri Moshkin, Rob Willemsen, Elzo De Wit, Bas Van Steensel, and Wouter De Laat. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4c). *Nat. Genet.*, 38(11):1348–1354, November 2006.

[133] Zhihu Zhao, Gholamreza Tavoosidana, Mikael Sjölinder, Anita Göndör, Piero Mariano, Sha Wang, Chandrasekhar Kanduri, Magda Lezcano, Kuljeet Singh Sandhu, Umashankar Singh, Vinod Pant, Vijay Tiwari, Sreenivasulu Kurukuti, and Rolf Ohlsson. Circular chromosome conformation capture (4c) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.*, 38(11):1341–1347, November 2006.

[134] Maria A Ferraiuolo, Amartya Sanyal, Natalia Naumova, Job Dekker, and Josée Dostie. From cells to chromatin: Capturing snapshots of genome organization with 5C technology. *Methods*, 58(3):255–267, 2012.

[135] Borbala Mifsud, Filipe Tavares-Cadete, Alice N Young, Robert Sugar, Stefan Schoenfelder, Lauren Ferreira, Steven W Wingett, Simon Andrews, William Grey, Philip A Ewels, Bram Herman, Scott Happe, Andy Higgs, Emily Leproust, George A Follows, Peter Fraser, Nicholas M Luscombe, and Cameron S Osborne. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.*, 47(6):598–606, June 2015.

[136] Pelin Sahlén, Ilgar Abdullayev, Daniel Ramsköld, Liudmila Matskova, Nemanja Rilakovic, Britta Lötstedt, Thomas J Albert, Joakim Lundeberg, and Rickard Sandberg. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol.*, 16:156, August 2015.

[137] Robert A Beagrie, Antonio Scialdone, Markus Schueler, Dorothee C A Kraemer, Mita Chotalia, Sheila Q Xie, Mariano Barbieri, Inês de Santiago, Liron-Mark Lavitas, Miguel R Branco, James Fraser, Josée Dostie, Laurence Game, Niall Dillon, Paul A W Edwards, Mario Nicodemi, and Ana Pombo. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, 543(7646):519–524, March 2017.

[138] Sofia A Quinodoz, Noah Ollikainen, Barbara Tabak, Ali Palla, Jan Marten Schmidt, Elizabeth Detmar, Mason M Lai, Alexander A Shishkin, Prashant Bhat, Yodai Takei, Vickie Trinh, Erik Aznauryan, Pamela Russell, Christine Cheng, Marko Jovanovic, Amy Chow, Long Cai, Patrick McDonel, Manuel Garber, and Mitchell Guttman. Higher-Order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell*, 174(3):744–757.e24, July 2018.

[139] Tsung-Han S Hsieh, Assaf Weiner, Bryan Lajoie, Job Dekker, Nir Friedman, and Oliver J Rando. Mapping nucleosome resolution chromosome folding in yeast by Micro-C. *Cell*, 162(1):108–119, July 2015.

[140] Guoliang Li, Melissa J Fullwood, Han Xu, Fabianus Hendriyan Mulawadi, Stoyan Velkov, Vinsensius Vega, Pramila Nuwantha Ariyaratne, Yusoff Bin Mohamed, Hong-Sain Ooi, Chandana Tennakoon, Chia-Lin Wei, Yijun Ruan, and Wing-Kin Sung. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.*, 11(2):R22, February 2010.

[141] Maxwell R Mumbach, Adam J Rubin, Ryan A Flynn, Chao Dai, Paul A Khavari, William J Greenleaf, and Howard Y Chang. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods*, 13(11):919–922, November 2016.

[142] Rongxin Fang, Miao Yu, Guoqiang Li, Sora Chee, Tristin Liu, Anthony D Schmitt, and Bing Ren. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.*, 26(12):1345–1348, December 2016.

[143] Hannah A Pliner, Jonathan S Packer, José L McFaline-Figueroa, Darren A Cusanovich, Riza M Daza, Delasa Aghamirzaie, Sanjay Srivatsan, Xiaojie Qiu, Dana Jackson, Anna Minkina, Andrew C Adey, Frank J Steemers, Jay Shendure, and Cole Trapnell. Cicero predicts cis-regulatory DNA interactions from Single-Cell chromatin accessibility data. *Mol. Cell*, 71(5):858–871.e8, 2018.

[144] Rn Ms Cns Anp, Pamela Hallquist Viale, Pamela Hallquist Viale, RN, MS, CNS, and ANP. The american cancer society's facts & figures: 2020 edition, 2020.

[145] J W Moul, R R Connelly, R M Mooneyhan, W Zhang, I A Sesterhenn, F K Mostofi, and D G McLeod. Racial differences in tumor volume and prostate specific antigen among radical prostatectomy patients. *J. Urol.*, 162(2):394–397, August 1999.

[146] P N Brawn, E H Johnson, D L Kuhl, M W Riggs, V O Speights, C F Johnson, 3rd, P P Pandya, M L Lind, and N F Bell. Stage at presentation and survival of white and black patients with prostate carcinoma. *Cancer*, 71(8):2569–2573, April 1993.

[147] Wael A Sakr, David J Grignon, Gabriel P Haas, Lance K Heilbrun, J Edson Pontes, and John D Crissmana. Age and racial distribution of prostatic intraepithelial neoplasia, 1996.

[148] J E Fowler, Jr, S A Bigler, C Lynch, S S Wilson, and P B Farabaugh. Prospective study of correlations between biopsy-detected high grade prostatic intraepithelial neoplasia, serum prostate specific antigen concentration, and race. *Cancer*, 91(7):1291–1296, April 2001.

[149] A Christofides and J Goddard. 42 the evolution in our understanding of prostatic anatomy and its surgical management. *Eur. Urol. Suppl.*, 12(1):e42, March 2013.

[150] Andreas Vesalius, John Bertrand de Cusance Morant Saunders, and Charles Donald O'Malley. *The Illustrations from the Works of Andreas Vesalius of Brussels: With Annotations and Translations, a Discussion of the Plates and Their Background, Authorship and Influence, and a Biographical Sketch of Vesalius*. Courier Corporation, 1950.

[151] Steven H Selman. The McNeal prostate: A review. *Urology*, 78(6):1224–1228, December 2011.

[152] R J Cohen, G Glezerson, L F Taylor, H A Grundle, and J H Naudé. The neuroendocrine cell population of the human prostate gland. *J. Urol.*, 150(2 Pt 1):365–368, August 1993.

[153] Susan Kasper. Exploring the origins of the normal prostate and prostate cancer stem cell. *Stem Cell Rev.*, 4(3):193–201, September 2008.

[154] Ding-Wei Ye, Hua Xu, Yao Zhu, and Bo Dai. National comprehensive cancer network (NCCN) risk classification in predicting biochemical recurrence after radical prostatectomy: a retrospective cohort study in chinese prostate cancer patients, 2018.

[155] Christopher Warlick, Bruce J Trock, Patricia Landis, Jonathan I Epstein, and H Ballentine Carter. Delayed versus immediate surgical intervention and prostate cancer outcome. *J. Natl. Cancer Inst.*, 98(5):355–357, March 2006.

[156] H Ballentine Carter, Anna Kettermann, Christopher Warlick, E Jeffrey Metter, Patricia Landis, Patrick C Walsh, and Jonathan I Epstein. Expectant

management of prostate cancer with curative intent: an update of the johns hopkins experience. *J. Urol.*, 178(6):2359–64; discussion 2364–5, December 2007.

[157] Ryan K Berglund, Timothy A Masterson, Kinjal C Vora, Scott E Eggener, James A Eastham, and Bertrand D Guillonneau. Pathological upgrading and up staging with immediate repeat biopsy in patients eligible for active surveillance. *J. Urol.*, 180(5):1964–7; discussion 1967–8, November 2008.

[158] M Kirby, C Hirst, and E D Crawford. Characterising the castration-resistant prostate cancer population: a systematic review. *Int. J. Clin. Pract.*, 65(11):1180–1192, November 2011.

[159] Huggins Ch and C Hodges. The effect of castration, of estro gen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. *Cancer Res.*, 1941.

[160] C W Gregory, B He, R T Johnson, O H Ford, J L Mohler, F S French, and E M Wilson. A mechanism for androgen receptor-mediated prostate cancer recurrence after androgen deprivation therapy. *Cancer Res.*, 61(11):4315–4319, June 2001.

[161] Nada Lallous, Stanislav V Volik, Shannon Awrey, Eric Leblanc, Ronnie Tse, Josef Murillo, Kriti Singh, Arun A Azad, Alexander W Wyatt, Stephane LeBihan, Kim N Chi, Martin E Gleave, Paul S Rennie, Colin C Collins, and Artem Cherkasov. Functional analysis of androgen receptor mutations that confer anti-androgen resistance identified in circulating cell-free DNA from prostate cancer patients. *Genome Biol.*, 17:10, January 2016.

[162] Kati K Waltering, Alfonso Urbanucci, and Tapio Visakorpi. Androgen receptor (AR) aberrations in castration-resistant prostate cancer. *Mol. Cell. Endocrinol.*, 360(1-2):38–43, September 2012.

[163] David Y Takeda, Sándor Spisák, Ji-Heui Seo, Connor Bell, Edward O'Connor, Keegan Korthauer, Dezső Ribli, István Csabai, Norbert Solymosi, Zoltán Szállási, David R Stillman, Paloma Cejas, Xintao Qiu, Henry W Long, Viktória Tisza, Pier Vitale Nuzzo, Mersedeh Rohanizadegan, Mark M Pomerantz, William C Hahn, and Matthew L Freedman. A somatically acquired enhancer of the androgen receptor is a noncoding driver in advanced prostate cancer. *Cell*, 174(2):422–432.e13, July 2018.

[164] Srinivas R Viswanathan, Gavin Ha, Andreas M Hoff, Jeremiah A Wala, Jian Carrot-Zhang, Christopher W Whelan, Nicholas J Haradhvala, Samuel S Freeman, Sarah C Reed, Justin Rhoades, Paz Polak, Michelle Cipicchio, Stephanie A Wankowicz, Alicia Wong, Tushar Kamath, Zhenwei Zhang,

Gregory J Gydush, Denisse Rotem, PCF/SU2C International Prostate Cancer Dream Team, J Christopher Love, Gad Getz, Stacey Gabriel, Cheng-Zhong Zhang, Scott M Dehm, Peter S Nelson, Eliezer M Van Allen, Atish D Choudhury, Viktor A Adalsteinsson, Rameen Beroukhim, Mary-Ellen Taplin, and Matthew Meyerson. Structural alterations driving Castration-Resistant prostate cancer revealed by Linked-Read genome sequencing. *Cell*, 174(2):433–447.e19, July 2018.

[165] Edward P Gelmann. Androgen receptor biology in prostate cancer, 2010.

[166] W I Mainwaring. A soluble androgen receptor in the cytoplasm of rat prostate. *J. Endocrinol.*, 45(4):531–541, December 1969.

[167] N Bruchovsky and J D Wilson. The conversion of testosterone to 5-alpha-androstan-17-beta-ol-3-one by rat prostate in vivo and in vitro. *J. Biol. Chem.*, 243(8):2012–2021, April 1968.

[168] K M Anderson and S Liao. Selective retention of dihydrotestosterone by prostatic nuclei. *Nature*, 219(5151):277–279, July 1968.

[169] G G Kuiper, P W Faber, H C van Rooij, J A van der Korput, C Ris-Stalpers, P Klaassen, J Trapman, and A O Brinkmann. Structural organization of the human androgen receptor gene. *J. Mol. Endocrinol.*, 2(3):R1–4, May 1989.

[170] Scott M Dehm and Donald J Tindall. Androgen receptor structural and functional elements: role and regulation in prostate cancer. *Mol. Endocrinol.*, 21(12):2855–2863, December 2007.

[171] J A Simental, M Sar, M V Lane, F S French, and E M Wilson. Transcriptional activation and nuclear targeting signals of the human androgen receptor. *J. Biol. Chem.*, 266(1):510–518, January 1991.

[172] P Doesburg, C W Kuil, C A Berrevoets, K Steketee, P W Faber, E Mulder, A O Brinkmann, and J Trapman. Functional in vivo interaction between the amino-terminal, transactivation domain and the ligand binding domain of the androgen receptor. *Biochemistry*, 36(5):1052–1064, February 1997.

[173] P L Shaffer, A Jivan, D E Dollins, F Claessens, and D T Gewirth. Crystal structure of androgen receptor DNA-binding domain bound to a direct repeat response element, June 2004.

[174] Giovanni Ciriello, Michael L Gatza, Andrew H Beck, Matthew D Wilkerson, Suhn K Rhie, Alessandro Pastore, Hailei Zhang, Michael McLellan, Christina Yau, Cyriac Kandoth, Reanne Bowlby, Hui Shen, Sikander Hayat, Robert Fieldhouse, Susan C Lester, Gary M K Tse, Rachel E Factor, Laura C

Collins, Kimberly H Allison, Yunn-Yi Chen, Kristin Jensen, Nicole B Johnson, Steffi Oesterreich, Gordon B Mills, Andrew D Cherniack, Gordon Robertson, Christopher Benz, Chris Sander, Peter W Laird, Katherine A Hoadley, Tari A King, TCGA Research Network, and Charles M Perou. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell*, 163(2):506–519, October 2015.

[175] Xiao Hu and John W Funder. The evolution of mineralocorticoid receptors. *Mol. Endocrinol.*, 20(7):1471–1478, July 2006.

[176] C S Chang, J Kokontis, and S T Liao. Molecular cloning of human and rat complementary DNA encoding androgen receptors. *Science*, 240(4850):324–326, April 1988.

[177] G Jenster, H A van der Korput, C van Vroonhoven, T H van der Kwast, J Trapman, and A O Brinkmann. Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. *Mol. Endocrinol.*, 5(10):1396–1404, October 1991.

[178] Eva Estébanez-Perpiñá, Leggy A Arnold, Phuong Nguyen, Edson Delgado Rodrigues, Ellena Mar, Raynard Bateman, Peter Pallai, Kevan M Shokat, John D Baxter, R Kiplin Guy, Paul Webb, and Robert J Fletterick. A surface on the androgen receptor that allosterically regulates coactivator binding. *Proc. Natl. Acad. Sci. U. S. A.*, 104(41):16074–16079, October 2007.

[179] L J Blok, P E de Ruiter, and A O Brinkmann. Androgen receptor phosphorylation. *Endocr. Res.*, 22(3):197–219, August 1996.

[180] Zhiyong Guo, Bojie Dai, Tianyun Jiang, Kexin Xu, Yingqiu Xie, Oekyung Kim, Issa Nesheiwat, Xiangtian Kong, Jonathan Melamed, Venkatesh D Handratta, Vincent C O Njar, Angela M H Brodie, Li-Rong Yu, Timothy D Veenstra, Hegang Chen, and Yun Qiu. Regulation of androgen receptor activity by tyrosine phosphorylation. *Cancer Cell*, 10(4):309–319, October 2006.

[181] Pierre Chymkowitch, Nicolas Le May, Pierre Charneau, Emmanuel Compe, and Jean-Marc Egly. The phosphorylation of the androgen receptor by TFIIH directs the ubiquitin/proteasome process. *EMBO J.*, 30(3):468–479, February 2011.

[182] Daniel Gioeli, Scott B Ficarro, Jesse J Kwiek, David Aaronson, Mathew Hancock, Andrew D Catling, Forest M White, Robert E Christian, Robert E Settlage, Jeffrey Shabanowitz, Donald F Hunt, and Michael J Weber. Androgen receptor phosphorylation. regulation and identification of the phosphorylation sites. *J. Biol. Chem.*, 277(32):29304–29314, August 2002.

[183] Jaclyn R Gareau and Christopher D Lima. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat. Rev. Mol. Cell Biol.*, 11(12):861–871, December 2010.

[184] Eckardt Treuter and Nicolas Venteclef. Transcriptional control of metabolic and inflammatory pathways by nuclear receptor SUMOylation. *Biochim. Biophys. Acta*, 1812(8):909–918, August 2011.

[185] Päivi Sutinen, Marjo Malinen, Sami Heikkinen, and Jorma J Palvimo. SUMOylation modulates the transcriptional activity of androgen receptor in a target gene and pathway selective manner. *Nucleic Acids Res.*, 42(13):8310–8319, July 2014.

[186] Y Yang, A K-W Tse, P Li, Q Ma, S Xiang, S V Nicosia, E Seto, X Zhang, and W Bai. Inhibition of androgen receptor activity by histone deacetylase 4 through receptor SUMOylation. *Oncogene*, 30(19):2207–2218, May 2011.

[187] Miia Rytinki, Sanna Kaikkonen, Päivi Sutinen, Ville Paakinaho, Vesa Rahkama, and Jorma J Palvimo. Dynamic SUMOylation is linked to the activity cycles of androgen receptor in the cell nucleus. *Mol. Cell. Biol.*, 32(20):4195–4205, October 2012.

[188] Noora Kotaja, Ulla Karvonen, Olli A Jänne, and Jorma J Palvimo. PIAS proteins modulate transcription factors by functioning as SUMO-1 ligases. *Mol. Cell. Biol.*, 22(14):5222–5234, July 2002.

[189] Mitchell Gross, Randy Yang, Irina Top, Christina Gasper, and Ke Shuai. PIASy-mediated repression of the androgen receptor is independent of sumoylation. *Oncogene*, 23(17):3059–3066, April 2004.

[190] Sanna Kaikkonen, Tiina Jääskeläinen, Ulla Karvonen, Miia M Rytinki, Harri Makkonen, Daniel Gioeli, Bryce M Paschal, and Jorma J Palvimo. SUMO-specific protease 1 (SENP1) reverses the hormone-augmented SUMOylation of androgen receptor and modulates gene responses in prostate cancer cells. *Mol. Endocrinol.*, 23(3):292–307, March 2009.

[191] Jinke Cheng, Dachun Wang, Zhengxin Wang, and Edward T H Yeh. SENP1 enhances androgen receptor-dependent transcription through desumoylation of histone deacetylase 1. *Mol. Cell. Biol.*, 24(13):6021–6028, July 2004.

[192] Nan Gao, Jianfeng Zhang, Mira A Rao, Thomas C Case, Janni Mirosevich, Yongqing Wang, Renjie Jin, Aparna Gupta, Paul S Rennie, and Robert J Matusik. The role of hepatocyte nuclear factor-3 alpha (forkhead box a1) and androgen receptor in transcriptional regulation of prostatic genes. *Mol. Endocrinol.*, 17(8):1484–1507, August 2003.

[193] Yu Zhao, Donald J Tindall, and Haojie Huang. Modulation of androgen receptor by FOXA1 and FOXO1 factors in prostate cancer. *Int. J. Biol. Sci.*, 10(6):614–619, June 2014.

[194] Chunyuan Jin and Gary Felsenfeld. Nucleosome stability mediated by histone variants h3.3 and H2A.Z. *Genes Dev.*, 21(12):1519–1529, June 2007.

[195] Mathieu Lupien, Jérôme Eeckhoute, Clifford A Meyer, Qianben Wang, Yong Zhang, Wei Li, Jason S Carroll, X Shirley Liu, and Myles Brown. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*, 132(6):958–970, March 2008.

[196] Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell*, 163(4):1011–1025, November 2015.

[197] Martin G Dalin, Alexis Desrichard, Nora Katabi, Vladimir Makarov, Logan A Walsh, Ken-Wing Lee, Qingguo Wang, Joshua Armenia, Lyndsay West, Snjezana Dogan, Lu Wang, Deepa Ramaswami, Alan L Ho, Ian Ganly, David B Solit, Michael F Berger, Nikolaus D Schultz, Jorge S Reis-Filho, Timothy A Chan, and Luc G T Morris. Comprehensive molecular characterization of salivary duct carcinoma reveals actionable targets and similarity to apocrine breast cancer. *Clin. Cancer Res.*, 22(18):4623–4633, September 2016.

[198] J N Weinstein, R Akbani, B M Broom, W Wang, R G Verhaak, D McConkey, S Lerner, M Morgan, C J Creighton, C Smith, and Others. Cancer genome atlas research network. comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*, 507(7492):315–322, 2014.

[199] Matti Annala, Sinja Taavitsainen, Gillian Vandekerkhove, Jack V W Bacon, Kevin Beja, Kim N Chi, Matti Nykter, and Alexander W Wyatt. Frequent mutation of the FOXA1 untranslated region in prostate cancer, 2018.

[200] Elizabeth J Adams, Wouter R Karthaus, Elizabeth Hoover, Deli Liu, Antoine Gruet, Zeda Zhang, Hyunwoo Cho, Rose DiLoreto, Sagar Chhangawala, Yang Liu, Philip A Watson, Elai Davicioni, Andrea Sboner, Christopher E Barbieri, Rohit Bose, Christina S Leslie, and Charles L Sawyers. FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes. *Nature*, 571(7765):408–412, July 2019.

[201] Stanley Zhou, James R Hawley, Fraser Soares, Giacomo Grillo, Mona Teng, Seyed Ali Madani Tonekaboni, Junjie Tony Hua, Ken J Kron, Parisa Mazrooei, Musaddeque Ahmed, Christopher Arlidge, Hwa Young Yun, Julie Livingstone, Vincent Huang, Takafumi N Yamaguchi, Shadrielle M G Espiritu, Yanyun Zhu, Tesa M Severson, Alex Murison, Sarina Cameron, Wilbert Zwart, Theodorus van der Kwast, Trevor J Pugh, Michael Fraser, Paul C Boutros,

Robert G Bristow, Housheng Hansen He, and Mathieu Lupien. Noncoding mutations target cis-regulatory elements of the FOXA1 plexus in prostate cancer. *Nat. Commun.*, 11(1):441, January 2020.

[202] Abhijit Parolia, Marcin Cieslik, Shih-Chun Chu, Lanbo Xiao, Takahiro Ouchi, Yuping Zhang, Xiaoju Wang, Pankaj Vats, Xuhong Cao, Sethuramasundaram Pitchiaya, Fengyun Su, Rui Wang, Felix Y Feng, Yi-Mi Wu, Robert J Lonigro, Dan R Robinson, and Arul M Chinnaiyan. Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. *Nature*, 571(7765):413–418, July 2019.

[203] S Edwards, C Campbell, P Flohr, J Shipley, I Giddings, R Te-Poele, A Dodson, C Foster, J Clark, S Jhavar, G Kovacs, and C S Cooper. Expression analysis onto microarrays of randomly selected cDNA clones highlights HOXB13 as a marker of human prostate cancer. *Br. J. Cancer*, 92(2):376–381, January 2005.

[204] L Zeltser, C Desplan, and N Heintz. Hoxb-13: a new hox gene in a distant region of the HOXB cluster maintains colinearity. *Development*, 122(8):2475–2484, August 1996.

[205] W F Shen, J C Montgomery, S Rozenfeld, J J Moskow, H J Lawrence, A M Buchberg, and C Largman. AbdB-like hox proteins stabilize DNA binding by the meis1 homeodomain proteins. *Mol. Cell. Biol.*, 17(11):6448–6458, November 1997.

[206] Mark M Pomerantz, Fugen Li, David Y Takeda, Romina Lenci, Apurva Chonkar, Matthew Chabot, Paloma Cejas, Francisca Vazquez, Jennifer Cook, Ramesh A Shivdasani, Michaela Bowden, Rosina Lis, William C Hahn, Philip W Kantoff, Myles Brown, Massimo Loda, Henry W Long, and Matthew L Freedman. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.*, 47(11):1346–1351, November 2015.

[207] Shinta Cheng, Sabrina Brzostek, Suzanne R Lee, Anthony N Hollenberg, and Steven P Balk. Inhibition of the dihydrotestosterone-activated androgen receptor by nuclear receptor corepressor. *Mol. Endocrinol.*, 16(7):1492–1501, July 2002.

[208] Myles C Hodgson, Inna Astapova, Shinta Cheng, Larissa J Lee, Manon C Verhoeven, Eunis Choi, Steven P Balk, and Anthony N Hollenberg. The androgen receptor recruits nuclear receptor CoRepressor (N-CoR) in the presence of mifepristone via its N and C termini revealing a novel molecular mechanism for androgen receptor antagonists. *J. Biol. Chem.*, 280(8):6511–6519, 2005.

[209] Hiroyasu Yamamoto, Evan G Williams, Laurent Mouchiroud, Carles Cantó, Weiwei Fan, Michael Downes, Christophe Héligon, Grant D Barish, Béatrice Desvergne, Ronald M Evans, Kristina Schoonjans, and Johan Auwerx. NCoR1 is a conserved physiological modulator of muscle mass and oxidative function. *Cell*, 147(4):827–839, November 2011.

[210] Yongfeng Shang, Molly Myers, and Myles Brown. Formation of the androgen receptor transcription complex. *Mol. Cell*, 9(3):601–610, March 2002.

[211] X Yuan, C Cai, S Chen, S Chen, Z Yu, and S P Balk. Androgen receptor functions in castration-resistant prostate cancer and mechanisms of resistance to new agents targeting the androgen axis. *Oncogene*, 33(22):2815–2825, May 2014.

[212] Amina Zoubeidi, Anousheh Zardan, Eliana Beraldi, Ladan Fazli, Richard Sowery, Paul Rennie, Colleen Nelson, and Martin Gleave. Cooperative interactions between androgen receptor (AR) and heat-shock protein 27 facilitate AR transcriptional activity. *Cancer Res.*, 67(21):10455–10465, November 2007.

[213] Virginie Georget, Béatrice Térouanne, Jean-Claude Nicolas, and Charles Sultan. Mechanism of antiandrogen action: key role of hsp90 in conformational change and transcriptional activity of the androgen receptor. *Biochemistry*, 41(39):11824–11831, October 2002.

[214] Jun Dong, Zeyu Wu, Dan Wang, Laura E Pascal, Joel B Nelson, Peter Wipf, and Zhou Wang. Hsp70 binds to the androgen receptor n-terminal domain and modulates the receptor function in prostate cancer cells. *Mol. Cancer Ther.*, 18(1):39–50, January 2019.

[215] Avrom J Caplan, Elizabeth Langley, Elizabeth M Wilson, and Johanna Vidal. Hormone-dependent transactivation by the human androgen receptor is regulated by a dnaj protein (). *J. Biol. Chem.*, 270(10):5251–5257, March 1995.

[216] Mark L Cutress, Hayley C Whitaker, Ian G Mills, Murray Stewart, and David E Neal. Structural basis for the nuclear import of the human androgen receptor. *J. Cell Sci.*, 121(Pt 7):957–968, April 2008.

[217] Mari Kaarbø, Tove I Klokk, and Fahri Saatcioglu. Androgen signaling and its interactions with other signaling pathways in prostate cancer. *Bioessays*, 29(12):1227–1238, December 2007.

[218] P S Rennie, N Bruchovsky, K J Leco, P C Sheppard, S A McQueen, H Cheng, R Snoek, A Hamel, M E Bock, and B S MacDonald. Characterization of two

cis-acting DNA elements involved in the androgen regulation of the probasin gene. *Mol. Endocrinol.*, 7(1):23–36, January 1993.

[219] J Ham, A Thomson, M Needham, P Webb, and M Parker. Characterization of response elements for androgens, glucocorticoids and progestins in mouse mammary tumour virus. *Nucleic Acids Res.*, 16(12):5263–5276, June 1988.

[220] Qianben Wang, Wei Li, X Shirley Liu, Jason S Carroll, Olli A Jänne, Erika Krasnickas Keeton, Arul M Chinnaiyan, Kenneth J Pienta, and Myles Brown. A hierarchical network of transcription factors governs androgen Receptor-Dependent prostate cancer growth. *Mol. Cell*, 27(3):380–392, 2007.

[221] F Claessens, L Celis, B Peeters, W Heyns, G Verhoeven, and W Rombauts. Functional characterization of an androgen response element in the first intron of the c3(1) gene of prostatic binding protein. *Biochem. Biophys. Res. Commun.*, 164(2):833–840, October 1989.

[222] Suzan Stelloo, Ekaterina Nevedomskaya, Henk G van der Poel, Jeroen de Jong, Geert J L H van Leenders, Guido Jenster, Lodewyk F A Wessels, Andries M Bergman, and Wilbert Zwart. Androgen receptor profiling predicts prostate cancer outcome. *EMBO Mol. Med.*, 7(11):1450–1464, November 2015.

[223] Tunç Morova, Daniel R McNeill, Nada Lallous, Mehmet Gönen, Kush Dalal, David M Wilson, 3rd, Attila Gürsoy, Özlem Keskin, and Nathan A Lack. Androgen receptor-binding sites are highly mutated in prostate cancer. *Nat. Commun.*, 11(1):832, February 2020.

[224] Mark M Pomerantz, Xintao Qiu, Yanyun Zhu, David Y Takeda, Wenting Pan, Sylvan C Baca, Alexander Gusev, Keegan D Korthauer, Tesa M Severson, Gavin Ha, Srinivas R Viswanathan, Ji Heui Seo, Holly M Nguyen, Baohui Zhang, Bogdan Pasaniuc, Claudia Giambartolomei, Sarah A Alaiwi, Connor A Bell, Edward P O'Connor, Matthew S Chabot, David R Stillman, Rosina Lis, Alba Font-Tello, Lewyn Li, Paloma Cejas, Andries M Bergman, Joyce Sanders, Henk G van der Poel, Simon A Gayther, Kate Lawrenson, Marcos A S Fonseca, Jessica Reddy, Rosario I Corona, Gleb Martovetsky, Brian Egan, Toni Choueiri, Leigh Ellis, Isla P Garraway, Gwo Shu Mary Lee, Eva Corey, Henry W Long, Wilbert Zwart, and Matthew L Freedman. *Prostate cancer reactivates developmental epigenomic programs during metastatic progression*, volume 52. Springer US, 2020.

[225] M Beato and A Sánchez-Pacheco. Interaction of steroid hormone receptors with the transcription initiation complex. *Endocr. Rev.*, 17(6):587–609, December 1996.

[226] Andrew J Bannister and Tony Kouzarides. The CBP co-activator is a histone acetyltransferase, 1996.

[227] W Feng, M P Nguyen, D Chen, and others. Estrogen receptor activation function 1 works by binding p160 coactivator proteins. *Molecular*, 1998.

[228] Chen Lin Hsieh, Teng Fei, Yiwen Chen, Tiantian Li, Yanfei Gao, Xiaodong Wang, Tong Sun, Christopher J Sweeney, Gwo Shu Mary Lee, Shaoyong Chen, Steven P Balk, Xiaole Shirley Liu, Myles Brown, and Philip W Kantoff. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc. Natl. Acad. Sci. U. S. A.*, 111(20):7319–7324, 2014.

[229] Reyaz ur Rasool, Ramakrishnan Natesan, Qu Deng, Shweta Aras, Priti Lal, Samuel Sander Effron, Erick Mitchell-Velasquez, Jessica M Posimo, Shannon Carskadon, Sylvan C Baca, Mark M Pomerantz, Javed Siddiqui, Lauren E Schwartz, Daniel J Lee, Nallasivam Palanisamy, Goutham Narla, Robert B Den, Matthew L Freedman, Donita C Brady, and Irfan A Asangani. CDK7 inhibition suppresses Castration-Resistant prostate cancer through MED1 inactivation, 2019.

[230] Madesh Belakavadi, Pradeep K Pandey, Ravi Vijayvargia, and Joseph D Fondell. MED1 phosphorylation promotes its association with mediator: implications for nuclear receptor signaling. *Mol. Cell. Biol.*, 28(12):3932–3942, June 2008.

[231] Feng Jin, Frank Claessens, and Joseph D Fondell. Regulation of androgen receptor-dependent transcription by coactivator MED1 is mediated through a newly discovered noncanonical binding motif. *J. Biol. Chem.*, 287(2):858–870, January 2012.

[232] Deborah Hay, Jim R Hughes, Christian Babbs, James O J Davies, Bryony J Graham, Lars Hanssen, Mira T Kassouf, A Marieke Marieke Oudelaar, Jacqueline A Sharpe, Maria C Suciu, Jelena Telenius, Ruth Williams, Christina Rode, Pik-Shan Li, Len A Pennacchio, Jacqueline A Sloane-Stanley, Helena Ayyub, Sue Butler, Tatjana Sauka-Spengler, Richard J Gibbons, Andrew J H Smith, William G Wood, and Douglas R Higgs. Genetic dissection of the α-globin super-enhancer in vivo. *Nat. Genet.*, 48(8):895–903, August 2016.

[233] Irfan A Asangani, Vijaya L Dommeti, Xiaoju Wang, Rohit Malik, Marcin Cieslik, Rendong Yang, June Escara-Wilke, Kari Wilder-Romans, Sudheer Dhanireddy, Carl Engelke, Mathew K Iyer, Xiaojun Jing, Yi-Mi Wu, Xuhong Cao, Zhaohui S Qin, Shaomeng Wang, Felix Y Feng, and Arul M Chinnaiyan. Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature*, 510(7504):278–282, June 2014.

[234] Fan Zhang, Samantha Wong, Joseph Lee, Shreyas Lingadahalli, Christopher Wells, Neetu Saxena, Christophe Sanchez, Bei Sun, Ana Karla Parra-Nuñez,

Novia Chan, Jennifer M Bui, Yuzhuo Wang, Paul S Rennie, Nathan Lack, Artem Cherkasov, Martin Gleave, Jörg Gsponer, and Nada Lallous. Dynamic phase separation of the androgen receptor and its coactivators to regulate gene expression. March 2021.

[235] Jill J Bouchard, Joel H Otero, Daniel C Scott, Elzbieta Szulc, Erik W Martin, Nafiseh Sabri, Daniele Granata, Melissa R Marzahn, Kresten Lindorff-Larsen, Xavier Salvatella, Brenda A Schulman, and Tanja Mittag. Cancer mutations of the tumor suppressor SPOP disrupt the formation of active, Phase-Separated compartments. *Mol. Cell*, 72(1):19–36.e8, October 2018.

[236] H Kraut, E K Frey, and E Werle. Der nachweis eines kreislaufhormons in der pankreasdrüse. (IV. mitteilung über dieses kreislaufhormon.), 1930.

[237] George M Yousef, Albert Chang, Andreas Scorilas, and Eleftherios P Diamandis. Genomic organization of the human kallikrein gene family on chromosome 19q13.3–q13.4. *Biochem. Biophys. Res. Commun.*, 276(1):125–133, September 2000.

[238] Carla A Borgoño, Julie-Ann Gavigan, Juliano Alves, Ben Bowles, Jennifer L Harris, Georgia Sotiropoulou, and Eleftherios P Diamandis. Defining the extended substrate specificity of kallikrein 1-related peptidases. *Biol. Chem.*, 388(11):1215–1225, November 2007.

[239] Ioannis Prassas, Azza Eissa, Gennadiy Poda, and Eleftherios P Diamandis. Unleashing the therapeutic potential of human kallikrein-related serine proteases. *Nat. Rev. Drug Discov.*, 14(3):183–202, March 2015.

[240] H Yu, P J Anderson, B I Freedman, S S Rich, and D W Bowden. Genomic structure of the human plasma prekallikrein gene, identification of allelic variants, and analysis in end-stage renal disease. *Genomics*, 69(2):225–234, October 2000.

[241] Miltiadis Paliouras and Eleftherios P Diamandis. Intracellular signaling pathways regulate hormone-dependent kallikrein gene expression. *Tumour Biol.*, 29(2):63–75, June 2008.

[242] J A Clements. The glandular kallikrein family of enzymes: tissue-specific expression and hormonal regulation. *Endocr. Rev.*, 10(4):393–419, November 1989.

[243] J A Clements, B A Matheson, and J E Funder. Tissue-specific regulation of the expression of rat kallikrein gene family members by thyroid hormone. *Biochem. J*, 267(3):745–750, May 1990.

[244] Mitchell G Lawrence, John Lai, and Judith A Clements. Kallikreins on steroids: structure, function, and hormonal regulation of prostate-specific antigen and the extended kallikrein locus. *Endocr. Rev.*, 31(4):407–446, August 2010.

[245] W J Catalona, D S Smith, T L Ratliff, K M Dodds, D E Coplen, J J Yuan, J A Petros, and G L Andriole. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.*, 324(17):1156–1161, April 1991.

[246] K B Cleutjens, C C van Eekelen, H A van der Korput, A O Brinkmann, and J Trapman. Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specific antigen promoter. *J. Biol. Chem.*, 271(11):6379–6388, March 1996.

[247] Kitty B J M Cleutjens, Kitty B J, Hetty A G M van der Korput, Conny C E M van Eekelen, Henri C J van Rooij, Peter W Faber, and Jan Trapman. An androgen response element in a far upstream enhancer region is essential for high, Androgen-Regulated activity of the Prostate-Specific antigen promoter, 1997.

[248] Dennis J Hazelett, Suhn Kyong Rhie, Malaina Gaddis, Chunli Yan, Daniel L Lakeland, Simon G Coetzee, Brian E Henderson, Houtan Noushmehr, Wendy Cozen, Zsofia Kote-Jarai, Rosalind A Eeles, Douglas F Easton, Christopher A Haiman, Wange Lu, Peggy J Farnham, and Gerhard A Coetzee. Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS Genet.*, 10(1):e1004102, January 2014.

[249] Amanda Khoury, Joanna Achinger-Kawecka, Saul A Bert, Grady C Smith, Hugh J French, Phuc-Loi Luu, Timothy J Peters, Qian Du, Aled J Parry, Fatima Valdes-Mora, Phillippa C Taberlay, Clare Stirzaker, Aaron L Statham, and Susan J Clark. Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat. Commun.*, 11(1):54, January 2020.

[250] Joana A Vidigal and Andrea Ventura. Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. *Nat. Commun.*, 6:8083, August 2015.

[251] Xiangjun He, Chunlai Tan, Feng Wang, Yaofeng Wang, Rui Zhou, Dexuan Cui, Wenxing You, Hui Zhao, Jianwei Ren, and Bo Feng. Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.*, 44(9):e85, May 2016.

[252] Kornel Labun, Tessa G Montague, Maximilian Krause, Yamila N Torres Cleuren, Håkon Tjeldnes, and Eivind Valen. CHOPCHOP v3:

expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.*, 47(W1):W171–W174, July 2019.

[253] Jonathan Y Hsu, Charles P Fulco, Mitchel A Cole, Matthew C Canver, Danilo Pellin, Falak Sher, Rick Farouni, Kendell Clement, Jimmy A Guo, Luca Biasco, Stuart H Orkin, Jesse M Engreitz, Eric S Lander, J Keith Joung, Daniel E Bauer, and Luca Pinello. CRISPR-SURF: discovering regulatory elements by deconvolution of CRISPR tiling screen data. *Nat. Methods*, 15(12):992–993, December 2018.

[254] Sangsu Bae, Jeongbin Park, and Jin Soo Kim. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of cas9 RNA-guided endonucleases. *Bioinformatics*, 30(10):1473–1475, 2014.

[255] Luca Pinello, Matthew C Canver, Megan D Hoban, Stuart H Orkin, Donald B Kohn, Daniel E Bauer, and Guo-Cheng Yuan. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.*, 34(7):695–697, July 2016.

[256] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Pedersen, Evan Miller, Stephan Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *J. Open Source Softw.*, 4(43):1686, November 2019.

[257] Hadley Wickham. Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.*, 3(2):180–185, March 2011.

[258] H Li and R Durbin. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 2009.

[259] Fidel Ramírez, Friederike Dündar, Sarah Diehl, Björn A Grüning, and Thomas Manke. deeptools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, 42(Web Server issue):W187–91, July 2014.

[260] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15(12):550, 2014.

[261] Alicia N Schep, Beijing Wu, Jason D Buenrostro, and William J Greenleaf. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods*, 14(10):975–978, October 2017.

[262] Jeffrey M Granja, M Ryan Corces, Sarah E Pierce, S Tansu Bagdatli, Hani Choudhry, Howard Y Chang, and William J Greenleaf. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.*, 53(3):403–411, March 2021.

[263] Yong Zhang, Tao Liu, Clifford A Meyer, Jérôme Eeckhoute, David S Johnson, Bradley E Bernstein, Chad Nussbaum, Richard M Myers, Myles Brown, Wei Li, and X Shirley Shirley. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, 9(9), 2008.

[264] Aaron R Quinlan and Ira M Hall. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, March 2010.

[265] Nicolas Servant, Nelle Varoquaux, Bryan R Lajoie, Eric Viara, Chong-Jian Chen, Jean-Philippe Vert, Edith Heard, Job Dekker, and Emmanuel Barillot. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.*, 16:259, December 2015.

[266] Douglas H Phanstiel, Alan P Boyle, Carlos L Araya, and Michael P Snyder. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics*, 30(19):2808–2810, October 2014.

[267] Nicolas Servant, Bryan R Lajoie, Elphège P Nora, Luca Giorgetti, Chong-Jian Chen, Edith Heard, Job Dekker, and Emmanuel Barillot. HiTC: exploration of high-throughput 'c' experiments, 2012.

[268] Zhizhuo Zhang, Kern Rei Chng, Shreyas Lingadahalli, Zikai Chen, Mei Hui Liu, Huy Hoang Do, Shaojiang Cai, Nicola Rinaldi, Huay Mei Poh, Guoliang Li, Ying Ying Sung, Charlie L Heng, Leighton J Core, Si Kee Tan, Xiaoan Ruan, John T Lis, Manolis Kellis, Yijun Ruan, Wing-Kin Sung, and Edwin Cheung. An AR-ERG transcriptional signature defined by long-range chromatin interactomes in prostate cancer cells. *Genome Res.*, 29(2):223–235, February 2019.

[269] Claudia Giambartolomei, Ji-Heui Seo, Tommer Schwarz, Malika Kumar Freund, Ruth Dolly Johnson, Sandor Spisak, Sylvan C Baca, Alexander Gusev, Nicholas Mancuso, Bogdan Pasaniuc, and Matthew L Freedman. H3K27ac HiChIP in prostate cell lines identifies risk genes for prostate cancer susceptibility. *Am. J. Hum. Genet.*, 108(12):2284–2300, December 2021.

[270] Nathan Harmston, Elizabeth Ing-Simmons, Malcolm Perry, Anja Barešić, and Boris Lenhard. GenomicInteractions: An R/Bioconductor package for manipulating and investigating chromatin interaction data. *BMC Genomics*, 16:963, November 2015.

[271] Yang Liao, Gordon K Smyth, and Wei Shi. featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, April 2014.

[272] Sven Heinz, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C Lin, Peter Laslo, Jason X Cheng, Cornelis Murre, Harinder Singh, and Christopher K Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, 38(4):576–589, May 2010.

[273] Eric C Bolton, Alex Y So, Christina Chaivorapol, Christopher M Haqq, Hao Li, and Keith R Yamamoto. Cell- and gene-specific regulation of primary target genes by the androgen receptor. *Genes and Development*, 21(16):2005–2017, August 2007.

[274] Chin-Yo Lin, Vinsensius B Vega, Jane S Thomsen, Tao Zhang, Say Li Kong, Min Xie, Kuo Ping Chiu, Leonard Lipovich, Daniel H Barnett, Fabio Stossi, Ailing Yeo, Joshy George, Vladimir A Kuznetsov, Yew Kok Lee, Tze Howe Charn, Nallasivam Palanisamy, Lance D Miller, Edwin Cheung, Benita S Katzenellenbogen, Yijun Ruan, Guillaume Bourque, Chia-Lin Wei, and Edison T Liu. Whole-Genome cartography of estrogen receptor α binding sites. *PLoS Genet.*, 3(6):e87, June 2007.

[275] Julia B Carleton, Kristofer C Berrett, and Jason Gertz. Multiplex enhancer interference reveals collaborative control of gene regulation by estrogen receptor α-Bound enhancers. *Cell Systems*, 5(4):333–344.e5, 2017.

[276] Benedikt Rauscher, Florian Heigwer, Marco Breinig, Jan Winter, and Michael Boutros. GenomeCRISPR - a database for high-throughput CRISPR/Cas9 screens. *Nucleic Acids Res.*, 45(D1):D679–D686, January 2017.

[277] Alok K Tewari, Galip Gürkan Yardimci, Yoichiro Shibata, Nathan C Sheffield, Lingyun Song, Barry S Taylor, Stoyan G Georgiev, Gerhard A Coetzee, Uwe Ohler, Terrence S Furey, Gregory E Crawford, and Phillip G Febbo. Chromatin accessibility reveals insights into androgen receptor activation and transcriptional specificity. *Genome Biol.*, 13(10):R88, October 2012.

[278] Julia B Carleton, Matthew Ginley-Hidinger, Kristofer C Berrett, Ryan M Layer, Aaron R Quinlan, and Jason Gertz. Regulatory sharing between estrogen receptor α bound enhancers. *Nucleic Acids Res.*, 48(12):6597–6610, July 2020.

[279] DepMap: The cancer dependency map project at broad institute. `http://depmap.org`. Accessed: 2021-12-17.

[280] C Wei, B P Callahan, M J Turner, R A Willis, E M Lord, R K Barth, and J G Frelinger. Regulation of human prostate-specific antigen gene expression in transgenic mice: evidence for an enhancer between the PSA and human glandular kallikrein-1 genes, 1998.

[281] Elphège P Nora, Anton Goloborodko, Anne-Laure Valton, Johan H Gibcus, Alec Uebersohn, Nezar Abdennur, Job Dekker, Leonid A Mirny, and Benoit G Bruneau. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, 169(5):930–944.e22, May 2017.

[282] Haoyue Zhang, Jessica Lam, Di Zhang, Yemin Lan, Marit W Vermunt, Cheryl A Keller, Belinda Giardine, Ross C Hardison, and Gerd A Blobel. CTCF and transcription influence chromatin structure re-configuration after mitosis. *Nat. Commun.*, 12(1):5157, August 2021.

[283] Charles E Massie, Boris Adryan, Nuno L Barbosa-Morais, Andy G Lynch, Maxine G Tran, David E Neal, and Ian G Mills. New androgen receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Rep.*, 8(9):871–878, September 2007.

[284] Soohwan Oh, Jiaofang Shao, Joydeep Mitra, Feng Xiong, Matteo D'Antonio, Ruoyu Wang, Ivan Garcia-Bassets, Qi Ma, Xiaoyu Zhu, Joo-Hyung Lee, Sreejith J Nair, Feng Yang, Kenneth Ohgi, Kelly A Frazer, Zhengdong D Zhang, Wenbo Li, and Michael G Rosenfeld. Enhancer release and retargeting activates disease-susceptibility genes. *Nature*, pages 1–6, May 2021.

[285] Eva K Brinkman, Tao Chen, Mario Amendola, and Bas van Steensel. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.*, 42(22):e168, December 2014.

[286] Xiaoyang Zhang, Richard Cowper-Sal-lari, Swneke D Bailey, Jason H Moore, and Mathieu Lupien. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res.*, 22(8):1437–1446, August 2012.

[287] Joseph Nasser, Drew T Bergman, Charles P Fulco, Philine Guckelberger, Benjamin R Doughty, Tejal A Patwardhan, Thouis R Jones, Tung H Nguyen, Jacob C Ulirsch, Heini M Natri, Elle M Weeks, Glen Munson, Michael Kane, Helen Y Kang, Ang Cui, John P Ray, Tom M Eisenhaure, Kristy Mualim, Ryan L Collins, Kushal Dey, Alkes L Price, Charles B Epstein, Anshul Kundaje, Ramnik J Xavier, Mark J Daly, Hailiang Huang, Hilary K Finucane, Nir Hacohen, Eric S Lander, and Jesse M Engreitz. Genome-wide maps of enhancer regulation connect risk variants to disease genes. *bioRxiv*, page 2020.09.01.278093, 2020.

[288] Robert-Jan Palstra and Frank Grosveld. Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. *Front. Genet.*, 3:195, September 2012.

[289] Terrence S Furey. ChIP-seq and beyond: New and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.*, 13(12):840–852, 2012.

[290] R J van Rooijen, M J Gasson, and W M de Vos. Characterization of the lactococcus lactis lactose operon promoter: contribution of flanking sequences and LacR repressor to promoter activity. *J. Bacteriol.*, 174(7):2273–2280, April 1992.