

## ***In silico* identification and functional validation of allele-dependent AR enhancers**

**Sonia Garritano<sup>1,\*</sup>, Alessandro Romanel<sup>1,\*</sup>, Yari Ciribilli<sup>2,\*</sup>, Alessandra Bisio<sup>2</sup>, Antoneta Gavoci<sup>1</sup>, Alberto Inga<sup>2,#</sup>, Francesca Demichelis<sup>1,3,4,#</sup>**

<sup>1</sup>Laboratory of Computational Oncology, CIBIO, Centre for Integrative Biology, University of Trento, Italy

<sup>2</sup>Laboratory of Transcriptional Networks, CIBIO, Centre for Integrative Biology, University of Trento, Italy

<sup>3</sup>HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College of Cornell University, New York, NY, USA

<sup>4</sup>Institute for Precision Medicine, Weill Medical College of Cornell University and New York Presbyterian Hospital, New York, NY, USA

\*These authors have contributed equally to this work

#These authors share senior authorship

### **Correspondence to:**

Francesca Demichelis, **e-mail:** demichelis@science.unitn.it

**Keywords:** Androgen Receptor (AR), polymorphic regulatory regions, enhancer, allele-specific, Estrogen Receptor (ER)

**Received:** November 19, 2014

**Accepted:** December 30, 2014

**Published:** February 27, 2015

### **ABSTRACT**

**Androgen Receptor (AR) and Estrogen Receptors (ERs) are key nuclear receptors that can cooperate in orchestrating gene expression programs in multiple tissues and diseases, targeting binding elements in promoters and distant enhancers. We report the unbiased identification of enhancer elements bound by AR and ER- $\alpha$  whose activity can be allele-specific depending on the status of nearby Single Nucleotide Polymorphisms (SNP). ENCODE data were computationally mined to nominate genomic loci with: (i) chromatin signature of enhancer activity from activation histone marks, (ii) binding evidence by AR and ER- $\alpha$ , (iii) presence of a SNP. Forty-one loci were identified and two, on 1q21.3 and 13q34, selected for characterization by gene reporter, Chromatin immunoprecipitation (ChIP) and RT-qPCR assays in breast (MCF7) and prostate (PC-3) cancer-derived cell lines. We observed allele-specific enhancer activity, responsiveness to ligand-bound AR, and potentially influence on the transcription of closely located genes (*RAB20*, *ING1*, *ARHGEF7*, *ADAM15*). The 1q21.3 variant, rs2242193, showed impact on AR binding in MCF7 cells that are heterozygous for the SNP. Our unbiased genome-wide search proved to be an efficient methodology to discover new functional polymorphic regulatory regions (PRR) potentially acting as risk modifiers in hormone-driven cancers and overall nominated SNPs in PRR across 136 transcription factors.**

### **INTRODUCTION**

Knowledge of transcriptional and chromatin regulators acting at promoter and enhancer elements has increased considerably in the last decade, highlighting a causative role for gene expression deregulation in different diseases. Cancer, neurological disorders, autoimmunity, and cardiovascular disease can be influenced by mutations in regulatory sequences or in transcription factors (TFs), cofactors, and chromatin regulators [1, 2].

Enhancers contain multiple *cis*-elements able to recruit TFs; the type and concentration of available TFs

at a specific time and location determine the efficiency of initiation complex formation and RNA polymerase II recruitment. *Cis*-elements can be located quite far from TSS, for example within internal introns, intergenic DNA or even in a different chromosome [3]. Promoter and enhancer need to be spatially close, for example through the formation of loops and other chromatin superstructures [4, 5].

Specific histone modifications have been used to map the presence of enhancers or promoters within a particular genomic region. Methylation and Acetylation of Lysine residues in histone tails are key events. There

is strong evidence that the concomitant presence of mono- and tri-methylation of Lysine 4 of histone H3 (H3K4me3) marks transcriptionally active promoters, while mono-methylation, but not tri-methylation, of the same histone residue (H3K4me1) is a marker of enhancers. Moreover, it has been proven that histone H3 Lysine 27 acetylation (H3K27Ac) is able to distinguish active enhancers from inactive or poised enhancer elements containing H3K4me1 alone [6]. These findings are further supported by the enrichment of these epigenetic marks in nucleosome free regions (NFR), p300 binding (a transcriptional co-activator, enzymatically acting as histone acetyl transferase - HAT), and increased nuclease sensitivity [7], all of them markers of open chromatin.

In the last decade, the Encyclopedia of DNA Elements Project (ENCODE) has performed a large number of sequence-based studies to map functional elements across the human genome leading to the biochemical characterization of intronic and intergenic regions [8–10]. The ENCODE results also highlighted cell type specificity of transcriptional regulator binding sites or chromatin states, consistent with the interpretation of noncoding variants relevant to human diseases. Genome-wide association studies (GWAS) have identified more than 150 polymorphic loci associated with increased susceptibility to cancer [11], the majority of which reside outside of known protein-coding sequences potentially influencing the regulation of critical target genes through distal enhancer elements [12–14].

The availability of these annotations opens up to a plethora of *in silico* studies towards the understanding of the role of non-coding inherited in human diseases. So far direct functional implications have been demonstrated only for few of the noncoding SNPs identified through GWAS [15]. To what extent GWAS genetic variants are of clinical or public health importance especially for developing preventive or therapeutic interventions is an open question. The challenge is to demonstrate how single variants or combinations can increment disease susceptibility by perturbing the expression of a transcript, disrupting the function of a protein or affecting regulatory sequences.

We reasoned that ad hoc computational searches of annotated regions combined with genome-wide TF sites from ChIP-Seq experiments and *in vitro* functional assay could identify polymorphisms that likely influence target gene regulation in an allele-dependent manner. As proof of concept, we focused on polymorphisms within regulatory elements bound by two TFs, ER- $\alpha$  and AR, that are key nuclear receptors in common human cancers characterized by a genetic component to their etiology, breast and prostate cancers [16, 17].

Nuclear receptors belong to a large superfamily of evolutionary related TFs that are able to integrate signals coming from outside of the cells and influence gene

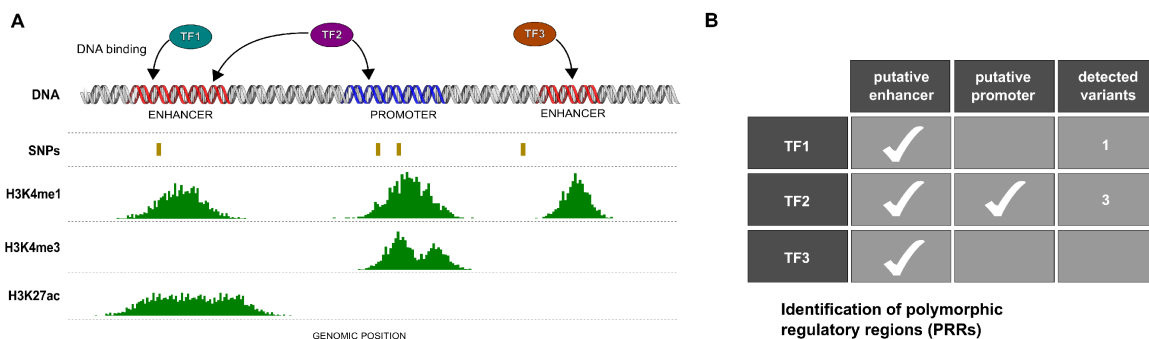
expression. Glucocorticoid Receptor (GR), Androgen Receptor (AR), Estrogen Receptors (ERs) and Retinoic Acid Receptors (RARs) are among the most important and studied members of the family [18]. Ligand binding causes a conformational change that enables dissociation from the inhibitory complex, homo- or hetero-dimerization, nuclear translocation, DNA binding, recruitment of co-activators, thus stimulating transcription of their target genes [19]. Interestingly, it has been discovered that the recruitment of AR can occur more often at gene-distal and intragenic sites rather than at proximal promoter regions. Deregulation of androgen/AR signaling perturbs the normal development of reproductive tract and accounts for a wide range of pathological conditions such as androgen-insensitive syndrome and prostate cancer [20]. Indeed, most prostate cancers express AR, are androgen-dependent for their growth and, as a result of androgen withdrawal, can undergo either cell cycle arrest or even apoptosis. For these reasons, androgen deprivation therapy (ADT) is an effective treatment in prostate cancer, although most patients progress to castration-resistant prostate cancer with an increase of AR expression levels and hypersensitivity to androgen-based therapies. Estrogen Receptor  $\alpha$  and  $\beta$  are sequence-specific TFs that play important roles in development as well as in physiological or pathological conditions in somatic cells, able to influence transcription once activated through the binding to estrogenic compounds ligands. Deregulation of ERs, particularly ER- $\alpha$ , has been extensively studied and associated with cancer development. ER- $\alpha$  induces cell growth and proliferation even if its expression in tumor correlates with a favorable prognosis in endocrine therapy [21, 22].

The broad coverage of the ENCODE annotations allows for the robust investigation of the impact that both somatic and germ line single nucleotide variants can have on distal *cis*-regulatory sequences [23]. Through a genome wide methodologically unbiased approach applied across multiple cell lines, we identified a set of regulatory elements targeted by one or multiple TF spanning (or in proximity) SNPs and named them polymorphic regulatory regions (PRR). *In vitro* validation experiments on selected loci bound by ER- $\alpha$  and by AR indicate that this approach can detect functionally distinct allelic variants acting as AR-responsive distant enhancers.

## RESULTS

### ***In silico* detection and characterization of polymorphic regulatory regions**

*In silico* analyses identified putative regulatory regions bound by one or more TFs and encompassing polymorphic sites as schematically represented in Figure 1A and 1B, here defined as PRRs. Overall we identified PRRs involving 136 transcription factors and focused on



**Figure 1: Identification of polymorphic regulatory regions (PRRs).** (A) Schematic view of putative PRRs study selection. Based on genomic coordinates regulatory regions (e.g. enhancers and promoters) from ENCODE open chromatin and activation histone marks peaks (e.g. H3K4m1, H3K4me3 and H3K27ac), polymorphic sites (SNPs) and transcription factor (TF) binding regions were combined. (B) Multiple transcription factors may occupy one or more PRRs characterized by different patterns of polymorphic loci.

**Table 1: Number of SNPs (dbSNP138) from the human genome that intersect regulatory regions bound by AR and/or ER- $\alpha$**

	AR	ER- $\alpha$	AR and ER- $\alpha$
H3K4me1	525	500	33
H3K4me1+H3K3me3	257	281	19
H3K27ac	396	410	25
H3K9ac	363	437	33
DNase	184	284	11
FAIRE	76	161	4
Union	591	553	41

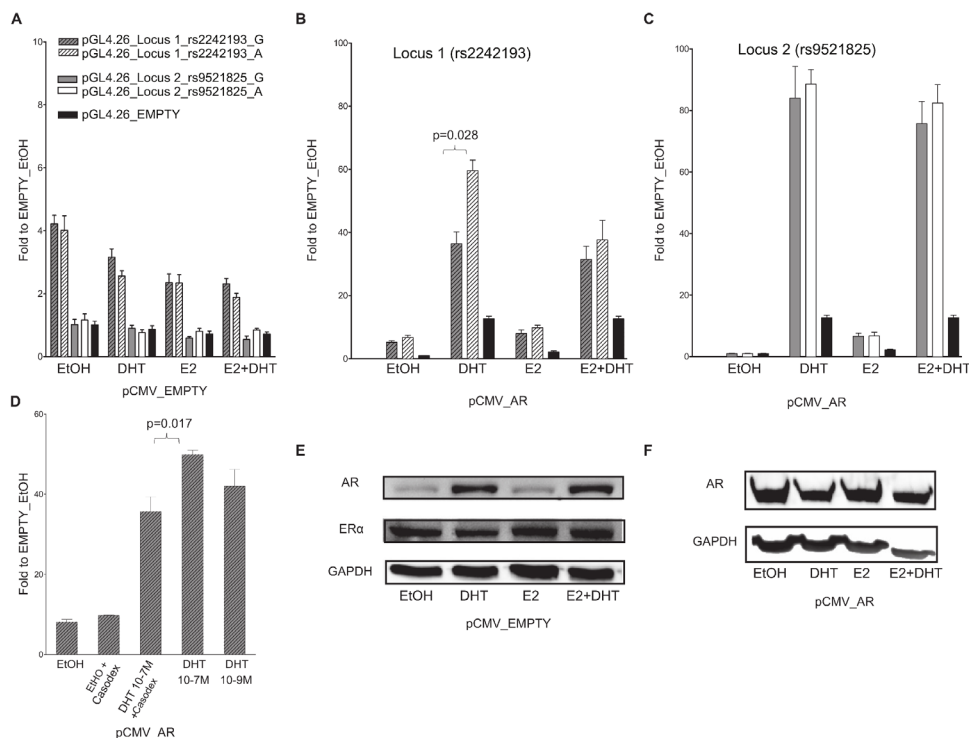
AR and ER- $\alpha$  occupancy data. Table 1 lists the number of SNPs within *consensus regulatory regions* bound by either one or both TFs. A total of 591 (553) SNPs within at least one consensus regulatory region and bound by AR (or ER- $\alpha$ ) were identified. A subset of 41 SNPs were located within consensus regulatory regions bound both by AR and ER- $\alpha$ , of which 14 SNPs were in regions compatible with enhancer activity and 19 with promoter activity, based on histone tail marks. To select PRRs for downstream validation we further considered SNP allele frequencies, availability of genotype-phenotype data (see Materials and Methods) and SNP genotypes of available cell lines. Two PRRs were chosen for experimental characterization; the first located in 1q21.3 (spanning rs2242193), hereafter referred to as Locus 1 and the second in 13q34 (spanning rs9521825), hereafter referred to as Locus 2 (Figure S1). Table S2 contains the SNP characteristics.

Supplementary Table S3A and S3B reports the numbers of SNPs that overlap consensus regulatory regions bound by at least one of the 136 TFs in CEU population and in all populations, respectively. In addition, the complete lists of SNP identifiers within consensus regulatory regions bound by every TF included in the study are available online (<http://demichelislab.unitn.it/PRRTFSNP>).

### The *in silico* selected regions showed enhancer activity

To address whether the *in silico* selected PRRs exhibit enhancer activity modulated by ER and/or AR, we conducted dual luciferase reporter gene assays in MCF7 and PC-3 cells transiently co-transfected with different pGL4.26 reporter constructs, along with pRL-SV40 control vector. No induction of reporter expression was observed after treatment with the DHT and/or E2 compared to the treatment with solvent (ethanol, EtOH). In MCF7 cells a clear induction (luciferase activity relative to the one obtained with the cells transfected with the pGL4-empty vector) was observed with the pGL4-Locus1 sequence in EtOH condition, demonstrating an intrinsic ligand-independent enhancer activity of this sequence, whereas no statistically significant induction was observed in PC-3 cells (Figure 2A, 3A).

In MCF7 cells the increase in luciferase activity was more evident when the pGL4-Locus1 reporter construct was co-transfected along with a pCMV-AR expression vector, particularly after the treatment with 100 nM DHT (Figure 2B). Interestingly, in these latter experimental conditions, the (G > A) SNP alleles within the putative



**Figure 2: The selected locus 1 and locus 2 regulatory regions act as AR-dependent enhancers in the MCF7 cell line. (A)** MCF7 cells were co-transfected with pCMV\_EMPTY vector along with different pGL4.26 reporter constructs containing selected regulatory regions isolated from two putative AR- or ER-target elements (indicated in the legend). Twenty-four hour post-transfection cells were treated for 16 hours with E2, DHT or the combination of the two compounds to stimulate respectively ER- or AR-dependent transcription. **(B – C)** MCF7 cells were co-transfected with pCMV\_AR vector along with pGL4.26\_Locus 1 or \_Locus 2 reporter constructs using the same experimental conditions as in panel A. Wild-type (dashed- or plain-grey bars) and SNPs rs2242193 or rs9521825 (dashed- or plain-white bars) containing constructs were tested. Indicated is the percentage value of statistical relevant differences (Student's *t*-test). **(D)** AR-dependent effect was determined performing luciferase assays by treating MCF7 cells with two different concentrations of DHT and by adding the AR inhibitor Casodex. All the bars presented in the gene reporter assays represent the averages and the standard deviations of at least three biological replicates each performed in triplicate. **(E)** To evaluate the endogenous cellular levels of AR and ER in MCF7 cells untreated and treated with E2, DHT or the combination of the two compounds, a western blot was performed. GAPDH expression was used as loading control. **(F)** A western blot was carried out as in E to verify the amount of AR protein in MCF7 cells over-expressing AR untreated and treated as indicated.

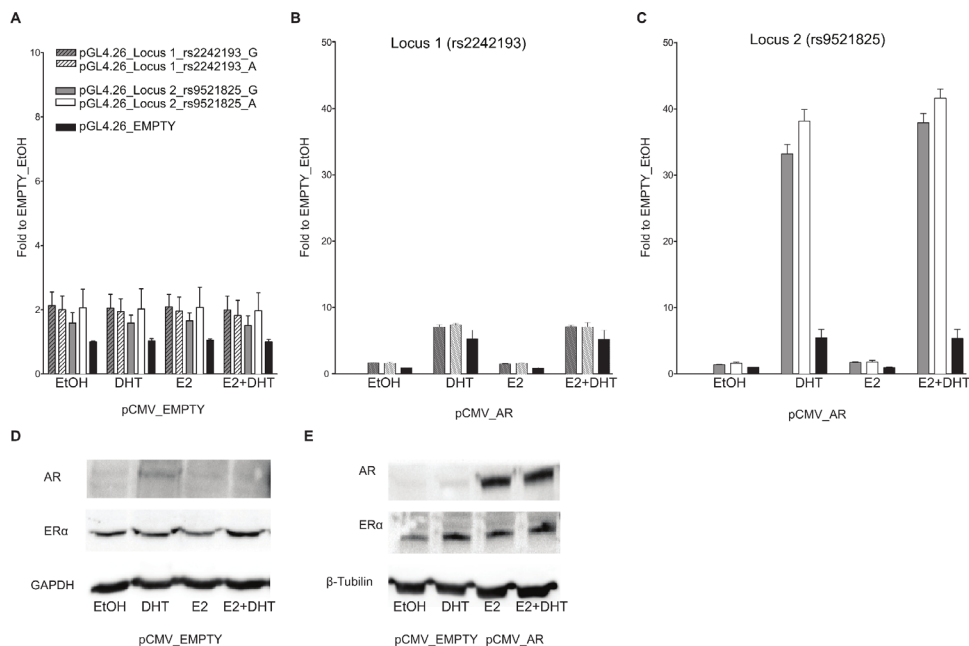
enhancer region exhibited a different responsiveness with significantly higher luciferase activity detected with the pGL4-Locus1\_A construct, ( $p = 0.028$ , determined by Student's *t*-test) (Figure 2B). Surprisingly, this SNP-dependent effect was not significant after combined addition of DHT and E2.

On the contrary, in PC-3 cells the Locus 1 did not elicit any enhancer activity (Figure 3B), suggesting that endogenous tissue-specific cofactors are requested for the transcriptional regulation.

When both cell lines cells were co-transfected with the pGL4-Locus2 reporter along with the pCMV-AR expression vector and supplemented with 100 nM DHT, the induction of the reporter was remarkably enhanced, with equal magnitude for both SNP alleles (Figure 2C, 3C). Furthermore, the treatment with E2 led to a moderate increase in transactivation in MCF7 but not PC-3 cells, and the combination of the two hormones showed neither additive nor antagonistic effects for Locus 2 in both cell lines.

We also demonstrated that DHT treatment elicited the same transcriptional effect also at a lower concentration ( $10^{-9}$ M) ( $p = 0.017$ ; Figure 2D). To further prove the AR-dependent transcriptional activation upon DHT, we treated MCF7 cells over-expressing AR with Casodex ( $10^{-5}$ M, also known as Bicalutamide), a strong non-steroidal anti-androgenic compound. In the presence of Casodex, the increase in pGL4-Locus 1 responsiveness was significantly lower compared to cells treated with DHT alone (Figure 2D). As additional positive control, cells were co-transfected with pGL4.15-PSA reporter and pCMV-AR plasmid. A high induction of reporter expression was shown when cells were treated with DHT compared to pGL4.15 control vector (Figure S2).

As previously demonstrated by Berger and co-workers [24], the treatment of MCF7 cells with 100 mM DHT led to a relevant increase in endogenous AR protein expression when compared to both EtOH or E2 treatment (Figure 2E). On the contrary, the amount of ER- $\alpha$  protein



**Figure 3: Only locus 2 is AR responsive in PC-3 cells.** Experiments were performed with the same experimental setting of Figure 2. Gene reporter assays were conducted in PC-3 cells co-transfected with pCMV\_EMPTY vector (A) or with AR over-expression vector (B) along with different the pGL4.26 reporter constructs (wild-type or containing the SNPs rs2242193 or rs9521825) -pGL4.26\_Locus 1 (panel B), \_Locus 2 (panel C). (D) The amount of ER alpha endogenous protein levels was evaluated through western blot analysis upon treatment with E2, DHT or the combination of the two compounds. (E) The same western blot was performed also in PC-3 cells transiently transfected with an empty vector or an expression vector for AR to demonstrate the increase in protein amount and to test the impact of E2 or DHT treatment on AR or ER expression. GAPDH or β-Tubulin endogenous expression was used as loading control.

was relatively unaffected by the treatments. In PC-3 cells the treatment with DHT and E2 led to the same results as in the MCF7 cells in terms of relative changes in AR and ER-α protein levels (both showing lower endogenous levels compared to MCF7 cells) (Figure 3D and 3E).

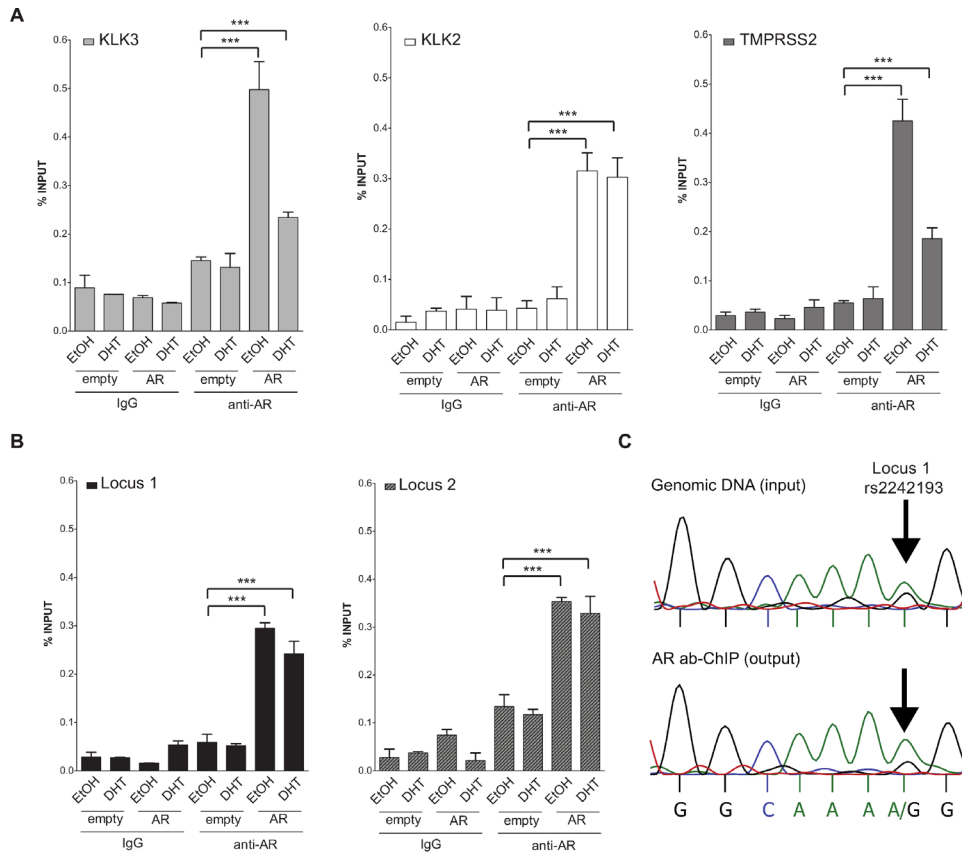
### Potential gene targets of AR bound enhancers

To identify potential target genes of the selected enhancers, we queried a human tissue dataset for gene transcripts in their surrounding genomic regions and selected four genes around Locus 1 (*PBX1P*, *SHC1*, *DCST2*, *ADAM15*) and three genes around Locus 2 (*RAB20*, *ING1*, *ARHGEF7*). Although the luciferase assays showed high responsiveness for both enhancer regions, we were not able to detect any increase in the expression of the selected genes in MCF7, with the exception of *ADAM15* in the presence of AR over-expression and upon DHT treatment (see Figure S3A and S3B). The treatment of PC-3 cells with 100 nM DHT for 24 hours led to a considerable increase in endogenous *RAB20*, *ARHGEF7*, and *ING1* expression when compared to EtOH treated cells (Figure S3D). The expression level of *ING1* increased also when the cells were treated with DHT for 16 or 36 hours. Consistent with the luciferase assays, in PC-3 cells genes located near the Locus 1 were only slightly inducible by

DHT (*PBX1P* and *ADAM15*) (Figure S3C). As expected, MCF7 and PC-3 cells showed *KLK3* up-regulation upon DHT treatment (Figure S3A–S3C, inserts). *TFE1* resulted in up-regulation upon E2 treatment consistent with ER-α dependent transactivation in MCF7 cells, where no up-regulation of endogenous *TFE1* was detected in PC-3 cells (mainly expressing ER-β) (Figure S3B–S3D, inserts).

### AR binds Locus 1 and Locus 2 PRRs and rs2242193 impacts on AR recruitment

As both loci were confirmed as transcriptionally responsive to DHT by luciferase assay, we next opted for ChIP assays with AR antibody (or with normal IgG as a control) using MCF7 cells that are heterozygous at rs2242193 in Locus 1, but homozygous for the reference allele at Locus 2. Using quantitative PCR (qPCR), we were able to detect AR binding to both selected loci in MCF7 cells transiently over-expressing AR (Figure 4A, 4B). Occupancy levels at *KLK3*, *KLK2* and *TMPPRS2* were measured as positive controls. Moreover, to assess whether AR showed allele-specific DNA binding at rs2242193, we amplified AR-enriched Locus 1 region by standard PCR followed by double-strand direct DNA sequencing analysis. Quantification of the electropherograms showed that the A allele was significantly enriched ( $p < 10^{-22}$ , Fisher test)



**Figure 4: Both locus 1 and locus 2 are directly bound by AR.** (A) A series of ChIP-qPCRs were performed in MCF7 cells (heterozygous for SNP rs2242193 within Locus 2) to determine AR chromatin binding at positive control enhancers -KLK3 (light grey bars), KLK2 (white bars) and TMPRSS2 (dark grey bars)- and (B) at Locus 1 and Locus 2 regions (presented respectively as black and grey-patterned bars, respectively). Mean  $\pm$  s.d. of three technical replicates were plotted. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.005$ , Student's *t*-test. (C) ChIP analysis from panel (B) was followed by standard PCR to amplify Locus 1 region and direct-sequencing was performed to quantify AR recruitment. The specific peaks involving the SNP rs2242193 are highlighted with arrows. Electropherograms showed that AR was preferentially recruited to the A allele of the SNP rs2242193 ( $p < 10^{-22}$ ). Input samples from ChIP assay were used as a control.

in chromatin fragments immunoprecipitated with antibody against AR compared to input genomic DNA (Figure 4C and Figure S4A for the biological replicate). Allele-specific PCR confirmed the higher relative occupancy of AR to the A allele (Figure 4C and Figure S4B for qPCR profiles). Collectively, these results showed that AR was preferentially recruited to the A allele of rs2242193. These observations were consistent with the significant increase in luciferase activity obtained with the reporter construct containing the A allele in MCF7 cells (Figure 2B).

## DISCUSSION

Inherited variants and somatic mutations located in intronic or intergenic genomic regions far from any oncogene or tumor suppressor gene may alter cancer susceptibility, influencing distal enhancer elements that regulate the expression of critical target genes [23, 25]. Indeed, the majority of all noncoding GWAS SNPs either lies within a DNase I Hypersensitive Sites or is in

complete linkage disequilibrium with SNPs in a nearby DNase I Hypersensitive Sites [26]. Few of them were demonstrated to modulate transcription factor binding, leading to allele specific transcriptional activity. The SNP rs6983267 located within an enhancer modulates the expression of the *MYC* oncogene [13] and affects a binding site for the Wnt-regulated transcription factor TCF4 [14]; the rs4590952 SNP that resides in a functional p53-binding site regulates the transcription of the *KITLG* gene [27], and the rs758550 SNP influences *CYP2D6* expression [28]. A recent study showed that androgen receptor binding and transactivation within an androgen response element upstream of the *TMPRSS2* gene is affected by a variant, rs8134378, [29]. In the context of tumour genomes, noncoding driver mutations in the *TERT* promoter were reported in multiple tumor types [30, 31] and nearly additional hundred noncoding driver candidates across the genome were nominated through an integrative study exploiting variants annotations from more than one thousand individuals [23].

In this study we applied a computational approach to select new putative polymorphic regulatory regions (PRRs), defined as regulatory elements spanned by SNPs that may influence the binding of transcription factors. Forty-one SNPs spanning regulatory regions were identified and two PRRs were selected for *in vitro* characterization (on 1q21.3 and 13q34). Towards their functional characterization, we cloned the two PRRs and tested them in gene reporter and ChIP assays separately examining the two SNP alleles. Both PRRs demonstrated enhancer activity and exhibited androgen-responsiveness in at least one cell line. The rs2242193 on 1q21.3 (Locus 1) exhibited allele-specific differences in MCF7 cells, where no enhancer activity was elicited in PC-3 cells. The 13q34 region (Locus 2) showed a modest trend for negative impact of the SNP allele in PC-3 cells, while for highly responsive MCF7 cells no effect was appreciated. These cell specific differences can be related to the expression of endogenous tissue-specific cofactors that are needed for the transcriptional regulation. Specifically, ligand-bound AR translocates to the nucleus, binds to androgen responsive elements (AREs) and modulates gene expression through the induction of chromatin reorganization, epigenetic histone modifications at target genomic loci, and through the recruitment of multiple co-regulator complexes. Proteins that interact with the AR can be divided into three general classes: (i) components of the general transcriptional machinery (e.g. TFIIB and TFIIF [32]); (ii) functionally different proteins with AR co-activating or co-repressing properties (e.g. histone acetyl transferases (HATs), co-activators such as NCOA1 (SRC1, AIB1), NCOA2 (TIF2, SRC2), NCOA3 (SRC3), and co-repressors like SIRT1 and NCOR1 [33–36] and (iii) specific transcription factors that differ from general transcription factors. Some of them interact directly with AR (DAX-1 with the AR Ligand Binding Domain -LBD-) [37] and affect its ability to be recruited at ARE sites without binding directly to the DNA. Other factors such as AP-1 can compete with AR for co-regulators that are present in limited amount within the cell [38]. Alternatively, some transcription factors (e.g., Foxa2 [39]) might bind to DNA sequences allowing cooperation and transcriptional co-regulation of target genes. The dependency of AR on its co-regulators to form a productive transcriptional complex could explain the tissue-selective androgen-dependent gene expression.

By ChIP assay experiments in MCF7 cells, we found that both loci were enriched in chromatin fragments immunoprecipitated with AR antibody and, importantly, detected AR preferentially recruited to the A allele of the SNP rs2242193 ( $p$ -value < 0.05). This result suggested that the genetic variant rs2242193 might have an impact on the AR recruitment to the chromatin by changing a single base of ARE sequence within the enhancer. In order to verify if the SNPs of interest fall within AR- and/or ER-DNA binding sites we aligned [40] short sequences surrounding the SNPs within Locus 1 and Locus 2 against the ARE (half-site ARE: RGNACR) [41] and the ERE

(RGGTCANNNTGASCY) [42] consensus sequences (Figure S5). This comparison indicates that rs2242193 maps to an important AR-DNA contact site (although the site has two additional nonconsensus bases) whereas rs9521825 maps to a less conserved position of a site that is overall a better match to the ARE consensus. Overall this agrees with the luciferase assay results, where the rs2242193 variant A strongly affected the androgen-responsiveness and with the ChIP assay where AR was preferentially recruited to the A allele. The ERE consensus is more conserved compared to the ARE consensus (Figure S5). It has been demonstrated that two nucleotide changes, one in each arm of the palindrome, are sufficient to inhibit ER- $\alpha$  binding [43] and different studies report that cell-specific factors regulate ER transcriptional activation [44, 45]. Moreover, PC-3 cells expresses mainly ER- $\beta$  rather than ER- $\alpha$  [46]. All together, these observations may explain the luciferase assay results where the E2 treatment led to increased transactivation in MCF7 but not in PC-3 cells and the induction of TFF1 endogenous expression upon E2 treatment in MCF7 cells only.

We posit that AR occupancy at Locus 1 and Locus 2 distal enhancers potentially influenced by the SNP genotype would possibly contribute to AR-dependent modulation of the expression of putative target genes and performed qPCR analysis to evaluate the endogenous expression levels of genes located in the proximity of the PRRs and their relative abundance in response to ectopic AR and/or DHT treatment. Based on RNA-seq data, we selected four (*PBX1P*, *SHC1*, *DCST2*, *ADAM15*) and three (*RAB20*, *ING1*, *ARHGEF7*) genes in the surroundings of Locus 1 and Locus 2, respectively. The treatment of PC-3 over-expressing AR with 100 nM DHT led to a relevant increase in endogenous *RAB20*, *ARHGEF7*, and *ING1* expression. Consistent with the luciferase assays, no gene located near the Locus 1 was induced by DHT. Albeit indirectly, these results suggest that the transcriptional activity of the aforementioned genes are under the control of this enhancer. Recent studies suggested a role for a number of RAB proteins in human cancers both as activators and inhibitors in tumor progression. In particular, *RAB20* expression is increased in exocrine pancreatic adenocarcinomas [47] and in breast cancer [48]. Moreover, *RAB20* amplification was also correlated with high-grade dysplastic colorectal adenomas [49]. Over-expression of *ING1* can induce cell cycle arrest and apoptosis [50], enhance expression of the *BAX* gene and was reported to alter mitochondrial membrane potential in a p53-dependent manner [51]. INGs also function in histone acetylation [52, 53] acting as stoichiometric partner for HAT and HDAC complexes [54]. In MCF7 cells where AR was over-expressed, only *ADAM15* gene expression increased significantly upon treatment with DHT (Figure S3). These results underline the strong variability of AR-induced responsiveness in different cell lines previously reported also among cell lines of

the same tissue derivation; for instance, only 11% of the androgen-responsive genes reported in HPr-1AR cells [41] were consistently activated by AR in LNCaP cells [55, 56]. The metalloproteinase ADAM15 is a multi-domain disintegrin protease that maps to a region of documented amplification associated with the metastatic progression of human cancers, including prostate, breast, ovarian, colon, and melanoma [57–59]. *ADAM15* mRNA and protein levels are increased in prostate cancer and its expression is significantly increased during metastatic progression. *ADAM15* may down-regulate adhesion of tumor cells to the extracellular matrix, reduce cell-cell adhesion, and promote metastasis through the activity of its disintegrin and metalloproteinase domains [60]. Given its diverse functions, ADAM15 may represent a pivotal regulatory component of tumor progression and an important target for therapeutic intervention.

Whereas our approach to the study of PRRs can in principle be pursued for any transcription factor of interest, we reasoned that AR and ER in the context of prostate and breast cancer cells would provide an excellent proof of principle as both diseases are characterized by a prominent hereditary component and as alternative roles for AR and ER have been suggested for both diseases. A significant number of poorly differentiated breast carcinomas are ER-negative but AR-positive suggesting AR as a useful marker for the further refinement of breast cancer subtype classification and as an independent prognostic factor and therapeutic target for the triple-negative breast cancers [61, 62]. On the other hand, although the AR remains the major target for prostate cancer prevention and treatment, ER is also involved in prostate cancer development and tumour progression. ER- $\alpha$  signalling potentiates the carcinogenic effects of androgens on the prostatic epithelium [22] and levels of E2 that increase with age may contribute to prostate cancer risk [63]. Epidemiologic studies suggested that African Americans, at the highest risk of prostate cancer, show the highest levels of estrogens [64]. Moreover, there is evidence that the treatment of the high-grade prostatic intraepithelial neoplasia with the antiestrogen toremifene leads to decreased cumulative risk of prostate cancer diagnosis [65]. Where the functional significance of the potential co-occupancy by AR and ER at enhancer sites (see Table S3) and the impact of androgen and estrogen relative levels remain to be established, we hypothesize that AR/ER bound PRRs can play a cell specific role in AR/ER-dependent modulation of target genes based upon the individual's genotype and overall potentially contribute to carcinogenesis mechanisms.

Several studies have noted that GWAS associated risk SNPs map to enhancers at higher than random rates. Altogether, our results show that the unbiased genome-wide search for PRRs is an efficient methodology to discover new functional *cis*-elements relevant to hormone driven diseases and beyond by providing experimental

evidence for selected variants mapping to regulatory regions. Further studies will extend on the understanding of specific disease mechanisms and on the design of strategies for individuals' risk assessment and treatment to eventually improve the processes of drug selection and dosing. The broad collection of more than 130 TFs analyzed in this study identified PRRs of potential interest for the research community towards the understanding of rare and common variants in *cis*-regulatory sequences.

## MATERIALS AND METHODS

### Selection of responsive regulatory regions

ChIP-Seq ENCODE data were queried for 17 cell-lines selected based on H3K4m1 and H3K4m3 data availability. The set includes the following; GM12878, H1-hESC, Hela-S3, HepG2, HMEC, HSMM, HSMMtube, HUVEC, K562, Monocytes-CD14+, NHA, NHDF-Ad, NHEK, NHLF, Osteobl, Dnd41, LNCaP. *Consensus regulatory regions* were determined for enhancer/promoter pattern (H3K4me1/H3K4me1+H3K4me3) and for additional activation and open chromatin markers (H3K9ac, H3K27ac, DNase-seq and FAIRE-seq). Specifically, for each marker the consensus was generated as the merge of all the peak regions that were detected in at least two cell lines as follow: (i) retain a peak if it overlaps for at least 50% of its length with a peak in second cell line; (ii) concatenate all the peaks obtained from step i; (iii) sort and merge retained peak regions. ENCODE data as per January 2013.

For each ENCODE TF ( $N = 134$ ) and for ER- $\alpha$  [66] and AR [67] ChIP-Seq data, a list of regions of interest was compiled based on 3,568,312 SNPs from dbSNP138 (merged list across populations) upon selection of binding peak regions across cell lines (as described above). To exploit paired genotype/phenotype data from sequencing experiments and high density oligonucleotide data [68, 69] SNPs represented on the Affymetrix SNP 6.0 ( $N \text{ SNP} = 924,395$ ) were tracked through the study and prioritized for functional validation.

### Cell lines and compounds

Breast adenocarcinoma-derived MCF7 cell line (positive both for Estrogen and Androgen Receptors) and Prostate Cancer-derived PC-3 cell line (positive for Estrogen Receptor alpha and negative for Androgen Receptor) were purchased from ATCC (American Type Culture Collection, LGC Standards, Milan, Italy).

MCF7 and PC-3 cell lines were maintained respectively in Dulbecco's Modified Eagle Medium (DMEM) or Roswell Park Memorial Institute medium (RPMI) (Gibco, Life Technologies, Milan, Italy) that were supplemented with 10% Fetal Bovine Serum (FBS), 100 units/ml penicillin, 100  $\mu\text{g/ml}$  streptomycin, and 2 mM



L-Glutamine. Cells were grown in humidified atmosphere at 37°C with 5% CO<sub>2</sub> in a cell culture incubator.

Sex hormone depletion (androgens and estrogens), prior to dihydrotestosterone (DHT) (Sigma-Aldrich, Milan, Italy) and 17β-estradiol (E2, Sigma-Aldrich, Milan, Italy) treatments, was achieved by growing the cells in medium without Phenol Red (Euroclone, Celbio, Milan, Italy), supplemented with 10% charcoal/dextran treated FBS (Hyclone, Celbio, Milan, Italy) for 48 hours.

### Plasmids and dual luciferase assay

The genomic sequence of the selected responsive regulatory regions was amplified using 5PRIME MasterTaq kit (5PRIME, Milan, Italy) in order to get an adequate replication fidelity on genomic DNA available obtained from HUVEC cells (Human Umbilical Vein Endothelial Cells from a male donor). Primers were selected with the Primer-BLAST web tool (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) (Table S1). PCR products for the two selected loci of interest (fragment sizes of 1009 bp (Locus 1) and 1205 bp (Locus 2)) were cloned upstream the luciferase cDNA into pGL4.26 vector (Promega, Milan, Italy) using KpnI and XhoI restriction endonucleases and T4 DNA Ligase (New England Biolabs, Euroclone). Constructs harboring the alternative allele for rs2242193 and rs9521825 were created with the GeneArt®eSite-Directed Mutagenesis System (Invitrogen, Life Technologies) according to the manufacturer's instructions. (Table S1). The correct insertion of the desired genomic sequence and SNP alleles in pGL4.26 was checked by direct DNA cycle sequencing (BMR Genomics, Padua, Italy). All plasmids were purified from XL1-Blue *E. coli* bacterial cells using the PureYield™ Plasmid Midiprep system protocol (Promega). The day before transfection, MCF7 and PC-3 cells (6 or 8 x 10<sup>4</sup> cells, respectively) were seeded in 24 well plates. Cells were co-transfected using the FugeneHD reagent (Promega) or TransIT-LT1 reagent (Mirus, TemaRicerca, Bologna, Italy) with pGL4.26-derived reporter vectors (350 ng) and pCMV-AR24Q expression vector (100 ng, to over-express AR). The total DNA amount was kept constant by the addition of the empty vector pCMV-NeoBam. All transfections were normalized for efficiency using 50 ng of the pRL-SV40 vector (Promega). Twenty-four hours after transfection, cells were treated with either 10<sup>-7</sup> M DHT, or 10<sup>-7</sup> M E2 or the combination of the two compounds for at least 16 h, diluted in hormone- and serum-free medium. To determine whether DHT elicited the same transcriptional effect also at a lower concentration, MCF7 cells were treated with 10<sup>-9</sup> M DHT. Moreover, to verify the AR-dependent transactivation of the selected regulatory regions, we used the non-steroidal anti-androgen compound Bicalutamide (10<sup>-5</sup> M, also known as Casodex, Sigma). Forth-eight hours after transfection, cells were lysed using Passive Lysis Buffer

1 × (Promega) and firefly and Renilla luciferase activities were measured as previously described [70] with Dual-Luciferase Reporter Assay (Promega) using the Infinite M200 multi-plate reader (Tecan, Milan, Italy).

### Real-time qPCR

Total RNA was extracted from MCF7 (seeded as 1.5–3 x 10<sup>5</sup> cells in a 6 well plate) and PC-3 (2.5–3.5 x 10<sup>5</sup>) cells transfected with AR or empty vectors (2.5 μg per well) and treated with DHT, E2 or the combination of the two compounds using the RNeasy kit (Qiagen, Milan, Italy) according to the manufacturer's instructions.

Two μg of total RNA was converted in cDNA using the RevertAid First Strand cDNA Synthesis Kit and the M-MuLV reverse transcriptase enzyme (ThermoFisher Scientific, Milan, Italy). Then, quantitative PCR reactions in real-time were performed as previously described [71] using KAPA SYBR® FAST Universal 2 × qPCR Master Mix (Kapa Biosystems, Resnova, Ancona, Italy) using the CFX384 or CFX96 Detection Systems (BioRad, Milan, Italy). Primer specificity and efficiency was tested with standard procedures. Analysis of relative mRNA expression was performed using the ΔΔCt method with GAPDH (Glyceraldehyde 3-phosphate dehydrogenase) and B2M (beta-2 microglobulin) as reference genes. Canonical targets for ER and AR, TFF1 and KLK3 respectively, were used as positive controls. qPCR analysis was also performed to evaluate the endogenous expression levels of genes located in the proximity of the region of interest (Table S1).

### Western blot

Endogenous as well as ectopically expressed soluble proteins were extracted from 6 well plates with 100 μl of ice-cold RIPA (Radio Imuno-Precipitation Assay) lysis buffer supplemented with protease inhibitor cocktail (Roche, Milan, Italy). Equal amounts of proteins (50 μg) were resolved by 12% SDS-PAGE and transferred to nitrocellulose membranes using the semi-dry iBlot Transfer System (Invitrogen, Life Technologies). Membranes were blocked with 5% non-fat skim milk dissolved in PBS-T and incubated at 4°C overnight with primary anti-AR (clone #: D6F11, Cell Signaling, Milan, Italy) (1:1000) and -ER-α (clone #: E115, Millipore, Milan, Italy) (1:500) antibodies diluted in 1% milk in PBS-T. Detection of GAPDH levels (clone #: 6C5, Santa Cruz Biotechnology, Milan, Italy) or β-Tubulin (clone #: 3F3-G2, Santa Cruz) served as loading control. Membranes were then incubated with secondary goat anti-mouse (1:10000) or goat anti-rabbit antibodies (Sigma-Aldrich) (1:12000) for 1 h at room temperature. Detection was achieved using the ECL Select detection reagent (Amersham, GE Health Care, Milan, Italy) with the ChemiDoc XRS+ System (BioRad).

## Chromatin immunoprecipitation (ChIP) assay

MCF7 cells were maintained into 150 mm Petri dishes in RPMI medium without Phenol Red, supplemented with 10% charcoal/dextran treated FBS. Two days after, cells were transfected with pCMV-AR24Q expression vector or the pCMV-NeoBam empty vector. Then, cells were treated with DHT ( $10^{-7}$ M) or EtOH. After 16 hours of treatment, cells were harvested and ChIP assays were performed as previously described [72, 73]. Briefly, protein-DNA complexes were cross-linked by the addition of 1% formaldehyde that was quenched with 0.125 M Glycine. Chromatin was sheared with cycles of sonication using a Misonix S-4000 sonicator (Misonix, Newtown, CT, USA) to generate fragments with an average size of 150–400 bp. Small aliquots of sample (10%) were used as input DNA. Chromatin immunoprecipitation was performed with anti-AR antibody (ChIPAb+ androgen receptor Assay Kit, Millipore) using the Magna ChIP G kit (Upstate, Millipore) according to the manufacturer's instructions. We next performed real-time quantitative PCR with Sybr Green as described above, followed by cycle sequencing analysis (BMR Genomics). Known AR target genes (*KLK3*, *KLK2*, *TMPRSS2*) were selected as positive control. Amplification of the selected regulatory regions was also performed (Table S1). AR specific recruitment was calculated as % of input signals respect to the EtOH treated cells according to the  $\Delta$ Ct method. ChIP analysis was followed by PCR to amplify and direct-sequencing Locus 1 region (both forward and reverse strand). The area under the peak corresponding to A and G allele of the SNP was integrated and quantified using the ImageJ software. The area of A allele (A) was normalized to the area of allele G (G) using the following formula:  $(A/(A + G)) * 100$ . The difference between the input and output DNA was compared using the Fisher test. Allelic specific PCR was performed to further support the impact of rs2242193 on AR occupancy, starting from DNA obtained by ChIP experiments in MCF7 cells that are heterozygous for this SNP (Table S1).

## ACKNOWLEDGMENTS

This work was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC, IG 13562 to F.D. and S.G.). We are thankful to Dr. Maria Pennuto (CIBIO, University of Trento) for providing the pCMV-AR24Q expression vector and to Dr. Dimple Chakravarty for valuable input on the study.

## REFERENCES

1. Heinz S, Romanoski CE, Benner C, Allison KA, Kaikkonen MU, Orozco LD, Glass CK. Effect of natural genetic variation on enhancer selection and function. *Nature*. 2013; 503:487–492.

2. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, Romano-Adesman A, Bjornson RD, Breitbart RE, Brown KK, Carriero NJ, Cheung YH, Deanfield J, DePalma S, Fakhro KA, Glessner J, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*. 2013; 498:220–223.
3. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet*. 2005; 76:8–32.
4. Krivega I, Dean A. Enhancer and promoter interactions—long distance calls. *Curr Opin Genet Dev*. 2012; 22:79–85.
5. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012; 489:109–113.
6. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:21931–21936.
7. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*. 2007; 39:311–318.
8. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74.
9. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447:799–816.
10. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, Dunham I, Elnitski LL, Farnham PJ, Feingold EA, Gerstein M, Giddings MC, et al. Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:6131–6138.
11. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research*. 2014; 42:D1001–1006.
12. Li Q, Seo JH, Stranger B, McKenna A, Pe'er I, Laframboise T, Brown M, Tyekuceva S, Freedman ML. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell*. 2013; 152:633–641.
13. Pomerantz MM, Ahmadyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A,

- Davis M, Yao K, Kehoe SM, Lenz HJ, Haiman CA, Yan C, Henderson BE, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nature genetics*. 2009; 41:882–884.
14. Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Bjorklund M, Wei G, Yan J, Niittymaki I, Mecklin JP, Jarvinen H, Ristimaki A, Di-Bernardo M, East P, Carvajal-Carmona L, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nature genetics*. 2009; 41:885–890.
  15. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:9362–9367.
  16. Demichelis F, Stanford JL. Genetic predisposition to prostate cancer: Update and future perspectives. *Urol Oncol*. 2014 Jul 1. pii: S1078–1439(14)00163-X. doi: 10.1016/j.urolonc.2014.04.021.
  17. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *The New England journal of medicine*. 2000; 343:78–85.
  18. Robinson-Rechavi M, Carpentier AS, Duffraisse M, Laudet V. How many nuclear hormone receptors are there in the human genome? *Trends in genetics : TIG*. 2001; 17:554–556.
  19. Sonoda J, Pei L, Evans RM. Nuclear receptors: decoding metabolic disease. *FEBS letters*. 2008; 582:2–9.
  20. Green SM, Mostaghel EA, Nelson PS. Androgen action and metabolism in prostate cancer. *Molecular and cellular endocrinology*. 2012; 360:3–13.
  21. Bieche I, Parfait B, Laurendeau I, Girault I, Vidaud M, Lidereau R. Quantification of estrogen receptor alpha and beta expression in sporadic breast cancer. *Oncogene*. 2001; 20:8109–8115.
  22. Bonkhoff H, Berges R. The evolving role of oestrogens and their receptors in the development and progression of prostate cancer. *Eur Urol*. 2009; 55:533–542.
  23. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, Sboner A, Lochovsky L, Chen J, Harmanci A, Das J, Abyzov A, Balasubramanian S, Beal K, Chakravarty D, Challis D, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science*. 2013; 342:1235587.
  24. Berger R, Lin DI, Nieto M, Sicinska E, Garraway LA, Adams H, Signoretti S, Hahn WC, Loda M. Androgen-dependent regulation of Her-2/neu in prostate cancer cells. *Cancer Res*. 2006; 66:5723–5728.
  25. Chen CY, Chang IS, Hsiung CA, Wasserman WW. On the identification of potential regulatory variants within genome wide association candidate SNP sets. *BMC medical genomics*. 2014; 7:34.
  26. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutayavin T, Stehling-Sun S, Johnson AK, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012; 337:1190–1195.
  27. Zeron-Medina J, Wang X, Repapi E, Campbell MR, Su D, Castro-Giner F, Davies B, Peterse EF, Sacilotto N, Walker GJ, Terzian T, Tomlinson IP, Box NF, Meinshausen N, De Val S, Bell DA, et al. A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell*. 2013; 155:410–422.
  28. Wang D, Papp AC, Sun X. Functional characterization of CYP2D6 enhancer polymorphisms. *Hum Mol Genet*. 2014 Nov 7. pii: ddu566.
  29. Clinckemalie L, Spans L, Dubois V, Laurent M, Helsen C, Joniau S, Claessens F. Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element. *Mol Endocrinol*. 2013; 27:2028–2040.
  30. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. *Science*. 2013; 339:957–959.
  31. Killela PJ, Reitman ZJ, Jiao Y, Bettegowda C, Agrawal N, Diaz LA Jr., Friedman AH, Friedman H, Gallia GL, Giovannella BC, Grollman AP, He TC, He Y, Hruban RH, Jallo GI, Mandahl N, et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:6021–6026.
  32. Lavery DN, McEwan IJ. Functional characterization of the native NH2-terminal transactivation domain of the human androgen receptor: binding kinetics for interactions with TFIIF and SRC-1a. *Biochemistry*. 2008; 47:3352–3359.
  33. Endler A, Chen L, Shibasaki F. Coactivator recruitment of AhR/ARNT1. *International journal of molecular sciences*. 2014; 15:11100–11110.
  34. McInerney EM, Rose DW, Flynn SE, Westin S, Mullen TM, Kronen A, Inostroza J, Torchia J, Nolte RT, Assa-Munt N, Milburn MV, Glass CK, Rosenfeld MG. Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. *Genes & development*. 1998; 12:3357–3368.
  35. Bevan CL, Hoare S, Claessens F, Heery DM, Parker MG. The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1. *Molecular and cellular biology*. 1999; 19:8383–8392.
  36. Needham M, Raines S, McPheat J, Stacey C, Ellston J, Hoare S, Parker M. Differential interaction of steroid hormone receptors with LXXLL motifs in SRC-1a depends on residues flanking the motif. *The Journal of steroid biochemistry and molecular biology*. 2000; 72:35–46.

37. Holter E, Kotaja N, Makela S, Strauss L, Kietz S, Janne OA, Gustafsson JA, Palvimo JJ, Treuter E. Inhibition of androgen receptor (AR) function by the reproductive orphan nuclear receptor DAX-1. *Mol Endocrinol.* 2002; 16:515–528.
38. Aarnisalo P, Palvimo JJ, Janne OA. CREB-binding protein in androgen receptor-mediated signaling. *Proceedings of the National Academy of Sciences of the United States of America.* 1998; 95:2122–2127.
39. Yu X, Suzuki K, Wang Y, Gupta A, Jin R, Orgebin-Crist MC, Matusik R. The role of forkhead box A2 to restrict androgen-regulated gene expression of lipocalin 5 in the mouse epididymis. *Mol Endocrinol.* 2006; 20:2418–2431.
40. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23:2947–2948.
41. Bolton EC, So AY, Chaivorapol C, Haqq CM, Li H, Yamamoto KR. Cell- and gene-specific regulation of primary target genes by the androgen receptor. *Genes & development.* 2007; 21:2005–2017.
42. Shu FJ, Sidell N, Yang D, Kallen CB. The tri-nucleotide spacer sequence between estrogen response element half-sites is conserved and modulates ERalpha-mediated transcriptional responses. *The Journal of steroid biochemistry and molecular biology.* 2010; 120:172–179.
43. Klinge CM. Estrogen receptor interaction with estrogen response elements. *Nucleic acids research.* 2001; 29:2905–2919.
44. Eeckhoutte J, Carroll JS, Geistlinger TR, Torres-Arzayus MI, Brown M. A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. *Genes & development.* 2006; 20:2513–2526.
45. Jones PS, Parrott E, White IN. Activation of transcription by estrogen receptor alpha and beta is cell type- and promoter-dependent. *J Biol Chem.* 1999; 274:32008–32014.
46. Lau KM, LaSpina M, Long J, Ho SM. Expression of estrogen receptor (ER)-alpha and ER-beta in normal and malignant prostatic epithelial cells: regulation by methylation and involvement in growth regulation. *Cancer Res.* 2000; 60:3175–3182.
47. Amillet JM, Ferbus D, Real FX, Antony C, Muleris M, Gress TM, Goubin G. Characterization of human Rab20 overexpressed in exocrine pancreatic carcinoma. *Human pathology.* 2006; 37:256–263.
48. Turner N, Lambros MB, Horlings HM, Pearson A, Sharpe R, Natrajan R, Geyer FC, van Kouwenhove M, Kreike B, Mackay A, Ashworth A, van de Vijver MJ, Reis-Filho JS. Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets. *Oncogene.* 2010; 29:2013–2023.
49. Habermann JK, Brucker CA, Freitag-Wolf S, Heselmeyer-Haddad K, Kruger S, Barenboim L, Downing T, Bruch HP, Auer G, Roblick UJ, Ried T. Genomic instability and oncogene amplifications in colorectal adenomas predict recurrence and synchronous carcinoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2011; 24:542–555.
50. Maher SK, Helbing CC. Modulators of inhibitor of growth (ING) family expression in development and disease. *Current drug targets.* 2009; 10:392–405.
51. Bose P, Thakur S, Thalappilly S, Ahn BY, Satpathy S, Feng X, Suzuki K, Kim SW, Riabowol K. ING1 induces apoptosis through direct effects at the mitochondria. *Cell death & disease.* 2013; 4:e788.
52. Loewith R, Meijer M, Lees-Miller SP, Riabowol K, Young D. Three yeast proteins related to the human candidate tumor suppressor p33(ING1) are associated with histone acetyltransferase activities. *Molecular and cellular biology.* 2000; 20:3807–3816.
53. Vieyra D, Toyama T, Hara Y, Boland D, Johnston R, Riabowol K. ING1 isoforms differentially affect apoptosis in a cell age-dependent manner. *Cancer Res.* 2002; 62:4445–4452.
54. Doyon Y, Cayrou C, Ullah M, Landry AJ, Cote V, Selleck W, Lane WS, Tan S, Yang XJ, Cote J. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Molecular cell.* 2006; 21:51–64.
55. DePrimo SE, Diehn M, Nelson JB, Reiter RE, Matese J, Fero M, Tibshirani R, Brown PO, Brooks JD. Transcriptional programs activated by exposure of human prostate cancer cells to androgen. *Genome biology.* 2002; 3:RESEARCH002.
56. Nelson PS, Clegg N, Arnold H, Ferguson C, Bonham M, White J, Hood L, Lin B. The program of androgen-responsive genes in neoplastic prostate epithelium. *Proceedings of the National Academy of Sciences of the United States of America.* 2002; 99:11890–11895.
57. Alers JC, Rochat J, Krijtenburg PJ, Hop WC, Kranse R, Rosenberg C, Tanke HJ, Schroder FH, van Dekken H. Identification of genetic markers for prostatic cancer progression. *Laboratory investigation; a journal of technical methods and pathology.* 2000; 80:931–942.
58. Balazs M, Adam Z, Treszl A, Begany A, Hunyadi J, Adany R. Chromosomal imbalances in primary and metastatic melanomas revealed by comparative genomic hybridization. *Cytometry.* 2001; 46:222–232.
59. Glinsky GV, Krones-Herzig A, Glinskii AB. Malignancy-associated regions of transcriptional activation: gene expression profiling identifies common chromosomal regions of a recurrent transcriptional activation in human prostate, breast, ovarian, and colon cancers. *Neoplasia.* 2003; 5:218–228.

60. Lucas N, Day ML. The role of the disintegrin metalloproteinase ADAM15 in prostate cancer progression. *Journal of cellular biochemistry*. 2009; 106:967–974.
61. Park S, Koo J, Park HS, Kim JH, Choi SY, Lee JH, Park BW, Lee KS. Expression of androgen receptors in primary breast cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2010; 21:488–492.
62. Moinfar F, Okcu M, Tsybrovskyy O, Regitnig P, Lax SF, Weybora W, Ratschek M, Tavassoli FA, Denk H. Androgen receptors frequently are expressed in breast carcinomas: potential relevance to new therapeutic strategies. *Cancer*. 2003; 98:703–711.
63. Kaufman JM, Vermeulen A. The decline of androgen levels in elderly men and its clinical and therapeutic implications. *Endocrine reviews*. 2005; 26:833–876.
64. Rohrmann S, Nelson WG, Rifai N, Brown TR, Dobs A, Kanarek N, Yager JD, Platz EA. Serum estrogen, but not testosterone, levels differ between black and white men in a nationally representative sample of Americans. *The Journal of clinical endocrinology and metabolism*. 2007; 92:2519–2525.
65. Price D, Stein B, Sieber P, Tutrone R, Bailen J, Goluboff E, Burzon D, Bostwick D, Steiner M. Toremifene for the prevention of prostate cancer in men with high grade prostatic intraepithelial neoplasia: results of a double-blind, placebo controlled, phase IIB clinical trial. *J Urol*. 2006; 176:965–970; discussion 970–961.
66. Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li R, Hennig S, Mosquera JM, Pauwels J, Park K, Kossai M, MacDonald TY, Fontugne J, Erho N, Vergara IA, Ghadessi M, Davicioni E, et al. The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nature communications*. 2014; 5:5383.
67. Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, Wang X, Wu L, Li J, Hu M, Gong Y, Cheng H, Laxman B, Vellaichamy A, Shankar S, Li Y, Dhanasekaran SM, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell*. 2010; 17:443–454.
68. Banerjee S, Oldridge D, Poptsova M, Hussain W, Chakravarty D, Demichelis F. A Computational Framework Discovers New Copy Number Variants with Functional Importance. *PloS one*. 2011; 6:E17539.
69. Demichelis F, Setlur SR, Banerjee S, Chakravarty D, Chen JY, Chen CX, Huang J, Beltran H, Oldridge DA, Kitabayashi N, Stenzel B, Schaefer G, Horninger W, Bektic J, Chinnaiyan AM, Goldenberg S, et al. Identification of functionally active, low frequency copy number variants at 15q21.3 and 12q21.31 associated with prostate cancer risk. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:6686–6691.
70. Ciribilli Y, Andreotti V, Menendez D, Langen JS, Schoenfelder G, Resnick MA, Inga A. The coordinated p53 and estrogen receptor cis-regulation at an FLT1 promoter SNP is specific to genotoxic stress and estrogenic compound. *PloS one*. 2010; 5:e10236.
71. Ciribilli Y, Monti P, Bisio A, Nguyen HT, Ethayathulla AS, Ramos A, Foggetti G, Menichini P, Menendez D, Resnick MA, Viadiu H, Fronza G, Inga A. Transactivation specificity is conserved among p53 family proteins and depends on a response element sequence code. *Nucleic acids research*. 2013; 41:8637–8653.
72. Bisio A, De Sanctis V, Del Vescovo V, Denti MA, Jegga AG, Inga A, Ciribilli Y. Identification of new p53 target microRNAs by bioinformatics and functional analysis. *BMC cancer*. 2013; 13:552.
73. Lion M, Bisio A, Tebaldi T, De Sanctis V, Menendez D, Resnick MA, Ciribilli Y, Inga A. Interaction between p53 and estradiol pathways in transcriptional responses to chemotherapeutics. *Cell Cycle*. 2013; 12:1211–1224.