Recent Advances in Genome Biology:
Insights from GWAS to Disease Development

Mehmet Tevfik DORAK, MD PhD

School of Life Sciences, Pharmacy and Chemistry
Kingston University London
London, United Kingdom

MolBiyoKon2017

8-10 September 2017
Boğaziçi University / Istanbul

Outline

What is GWAS / Why GWAS?

Unexpected results

Non-coding region and function

Genetic variants and genome biology

Bioinformatics tools

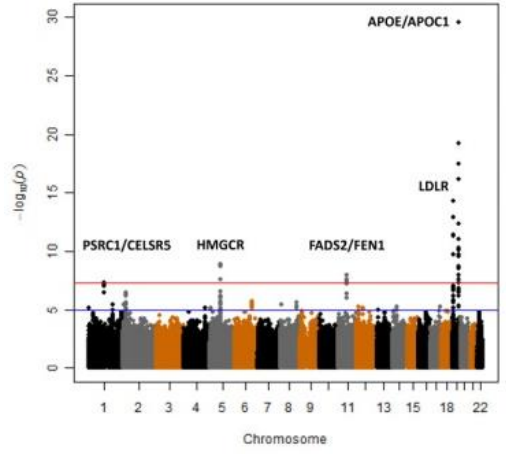
What is GWAS?

Unbiased exploration of correlations between SNPs in the human genome and traits

Up to 5M SNPs can be analyzed and millions more can be imputed

The results are pure statistical correlations

Biological interpretations require further experiments



Kingston University London

What is GWAS?

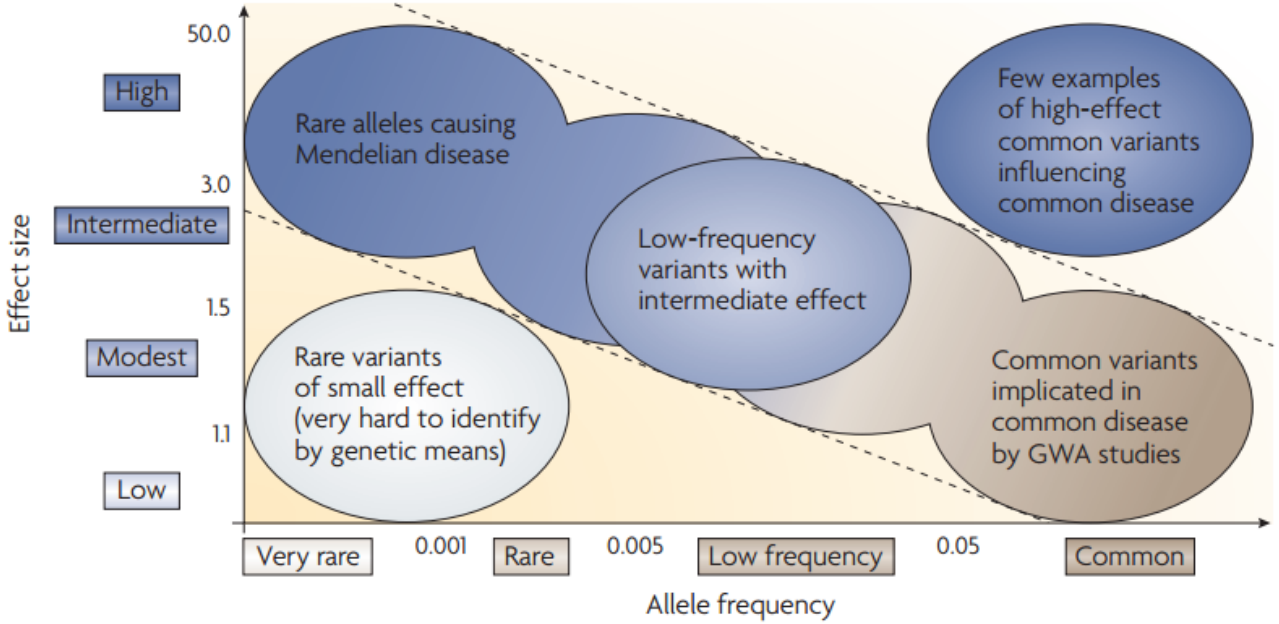


Figure 1 | **Feasibility of identifying genetic variants by risk-allele frequency and strength of genetic effect (odds ratio).** Reproduced, with permission, from *Nature* REF. 10 © (2009) Macmillan Publishers Ltd. All rights reserved. GWA, genome-wide association.

GENOME-WIDE ASSOCIATION STUDIES — VIEWPOINT

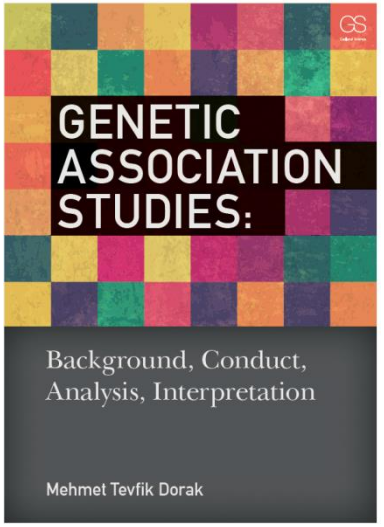
Mendelian disorders and multifactorial traits: the big divide or one for all?

Stylianos E. Antonarakis, Aravinda Chakravarti, Jonathan C. Cohen and John Hardy

Kingston University London

What is GWAS?

Figure 1.1. Intermediate mechanisms mediating causal variant's effect on disease susceptibility. Genetic variants modify disease risk by causing changes in gene expression (most common), splicing process or protein structure.



Kingston University London

Why GWAS?

Genetic association studies have two aims:

To identify genetic markers that can be used for prediction

To unravel disease biology

Unexpected Results

Great majority of results concern the non-coding region

Kingston University London

Unexpected Results

Table 1 | SNPs associated with risk in 8q24.

SNP	Disease	Position	*P*-value	Reference
rs1016343	Prostate cancer	128093297	1×10^{-7}	Eeles et al. (2008)
rs16901979	Prostate cancer	128124916	3×10^{-14}	Gudmundsson et al. (2007)
rs2456449	Chronic lymphocytic leukemia	128192981	8×10^{-10}	Crowther-Swanepoel et al. (2010)
rs16902094	Prostate cancer	128320346	6×10^{-15}	Gudmundsson et al. (2007)
rs378854	Prostate cancer	128323819		Meyer et al. (2011)
rs13281615	Breast cancer	128355618	5×10^{-12}	Easton and Eeles (2008)
rs1562430	Breast cancer, prostate cancer	128387852	6×10^{-7}	Turnbull et al. (2010)
rs10505477	Ovarian cancer	128407443	2×10^{-3}	Ghoussaini et al. (2008), Zanke et al. (2007)
	Colon cancer		3×10^{-11}	
rs10808556	Ovarian cancer	128413147		Ghoussaini et al. (2008)
rs6983267	Ovarian cancer	128413305	9.9×10^{-3}	Yeager et al. (2007), Ghoussaini et al. (2008), Eeles
	Colon cancer		1×10^{-14}	et al. (2008), Thomas et al. (2008), Tomlinson et al.
	Prostate cancer		9×10^{-13}	(2007), Berndt et al. (2008)
rs7837328	Colon cancer	128423127		Berndt et al. (2008)
rs7000448	Prostate cancer	128441170		Ghoussaini et al. (2008)
rs1447295	Prostate cancer, esophageal cancer	128485038	2×10^{-19}	Gudmundsson et al. (2007), Yeager et al. (2007), Lochhead et al. (2011)
rs4242382	Prostate cancer	128517573	3×10^{-19}	Thomas et al. (2008)
rs7017300	Prostate cancer	128525268		Yeager et al. (2007)
rs10090154	Prostate cancer	128532137		Cheng et al. (2008)
rs7837688	Prostate cancer	128539360		Yeager et al. (2007), Berndt et al. (2008)
D8S1128	Type II diabetes	128595148	2×10^{-3}	An et al. (2006)
rs9642880	Bladder cancer	128718068	7×10^{-12}	Ghoussaini et al. (2008), Kiemeney et al. (2008)
rs11993333	End stage renal disease (type I diabetes)	128992487	1.3×10^{-3}	Hanson et al. (2007)
rs2720709	End stage renal disease (type I diabetes)	129058356	2×10^{-5}	Hanson et al. (2007)
rs2648862	End stage renal disease (type I diabetes)	129061785		Hanson et al. (2007)
rs2608053	Hodgkin's lymphoma	129075832	1.16×10^{-7}	Enciso-Mora et al. (2010)
rs1499368	End stage renal disease (type I diabetes)	129094589	6.1×10^{-3}	Hanson et al. (2007)
rs2019960	Hodgkin's lymphoma	129192271	1.26×10^{-13}	Enciso-Mora et al. (2010)
rs1516982	Ovarian cancer	129533646		Goode et al. (2010)
rs10088218	Ovarian cancer	129543949	8×10^{-15}	Goode et al. (2010)
rs10098821	Ovarian cancer	129559228		Goode et al. (2010)

Kingston University London

The 8q24 gene desert: an oasis of non-coding transcriptional activity

*Konrad Huppi *, Jason J. Pitt†, Brady M. Wahlberg† and Natasha J. Caplen*

Unexpected Results

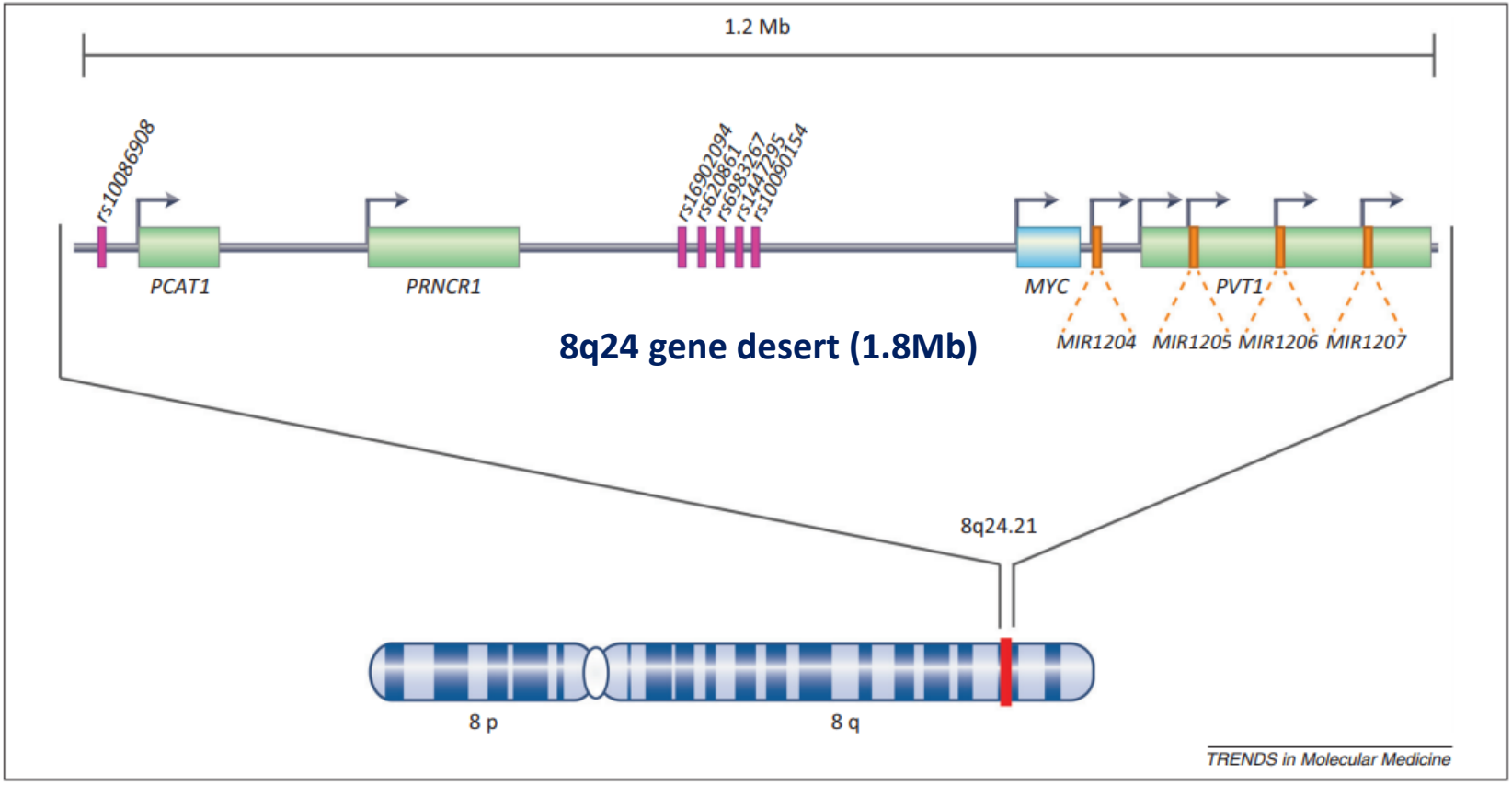


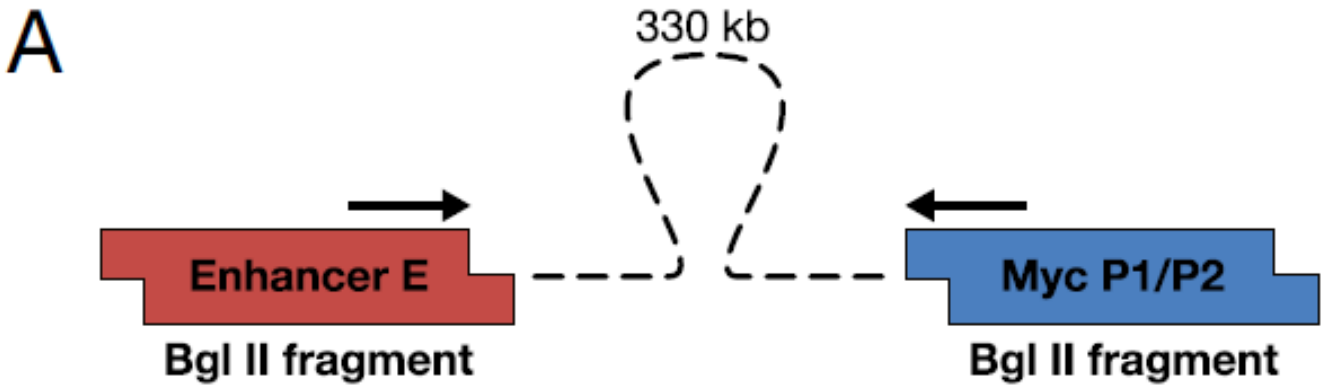
Figure 2. Prostate cancer susceptibility locus is enriched in long noncoding RNAs (lncRNAs). A 1.2 Mb region on chromosome 8q24.21 is a major prostate cancer (PCa)-susceptibility locus. It harbors multiple PCa-associated single-nucleotide polymorphisms (SNPs; shown in pink) and three PCa-associated lncRNAs (prostate cancer associated ncRNA transcript 1 [*PCAT1*], prostate cancer noncoding RNA1 [*PRNCR1*], and Pvt1 oncogene [*PVT1*], all shown in green), and exhibits frequent chromosomal amplification in human cancers. For simplicity, intervening protein-coding genes upstream of the *MYC* oncogene (shown in blue) are not diagrammed. Several microRNA (miRNA) genes colocalize to this region (shown in orange), three of which are housed within the *PVT1* 'host' lncRNA gene. However, no prostate-related functions have yet been ascribed to these miRNAs.

Kingston University London

Long noncoding RNAs and prostate carcinogenesis: the missing 'linc'?

Anna L. Walsh[1,2], Alexandra V. Tuzova[1], Eva M. Bolton[1,2], Thomas H. Lynch[1,2], and Antoinette S. Perry[1]

Unexpected Results



Enhancer E forms close contact with Myc promoter in 3C assay

Kingston University London

Long-range enhancers on 8q24 regulate c-Myc

Jose Sotelo[a,b], Dominic Esposito[b,c], Maria Ana Duhagon[d], Kelley Banfield[a,b], Jennifer Mehalko[b,c], Hongling Liao[a]
Robert M. Stephens[b,e], Timothy J. R. Harris[b], David J. Munroe[b,2], and Xiaolin Wu[a,b]

Non-coding Region and Function

The 8q24 gene desert: an oasis of non-coding transcriptional activity

Konrad Huppi, Jason J. Pitt†, Brady M. Wahlberg† and Natasha J. Caplen*

microRNAs

Long non-coding RNAs

Enhancers

Figure 1.1. Intermediate mechanisms mediating causal variant's effect on disease susceptibility. Genetic variants modify disease risk by causing changes in gene expression (most common), splicing process or protein structure.

Kingston University London

Non-coding Region and Function

ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS

Eric Pasmant[*†,1], **Audrey Sabbagh**[*†], **Michel Vidaud**[*†] **and Ivan Bièche**[*†]

[+] Author Affiliations

[1] Correspondence: UMR745 INSERM, Université Paris Descartes, Faculté des Sciences Pharmaceutiques et Biologiques, 4 avenue de l'Observatoire, 75006, Paris, France. E-mail: eric.pasmant@gmail.com

Abstract

A large noncoding RNA called *ANRIL* (for antisense noncoding RNA in the INK4 locus) has been identified within the *p15/CDKN2B–p16/CDKN2A–p14/ARF* gene cluster. While the exact role of *ANRIL* awaited further elucidation, common disease genomewide association studies (GWAS) have surprisingly identified the *ANRIL* gene as a genetic susceptibility locus shared associated by coronary disease, intracranial aneurysm and also type 2 diabetes. Expression studies have confirmed the coregulation of *p15/CDKN2B*, *p16/CDKN2A*, *p14/ARF*, and *ANRIL*. Among the cluster, *ANRIL* expression showed the strongest association with the multiple phenotypes linked to the 9p21.3 region. More recent GWAS also identified *ANRIL* as a risk locus for gliomas and basal cell carcinomas in accordance with the princeps observation. Moreover, a mouse model has confirmed the pivotal role of *ANRIL* in regulation of *CDKN2A/B* expression through a *cis*-acting mechanism and its implication in proliferation and senescence. The implication of *ANRIL* in cellular aging has provided an attractive unifying hypothesis to explain its association with various susceptibility risk factors. *ANRIL* identification emphasizes the underestimated role of long noncoding RNAs. Many GWAS have identified trait-associated SNPs that felt in noncoding genomic regions. It is conceivable to anticipate that long, noncoding RNAs will map to many of these "gene deserts."—Pasmant, E., Sabbagh, A., Vidaud, M., Bièche, I. *ANRIL*, a long, noncoding RNA, is an unexpected major hotspot in GWAS.

Kingston University London

Genetic Variants and Genome Biology

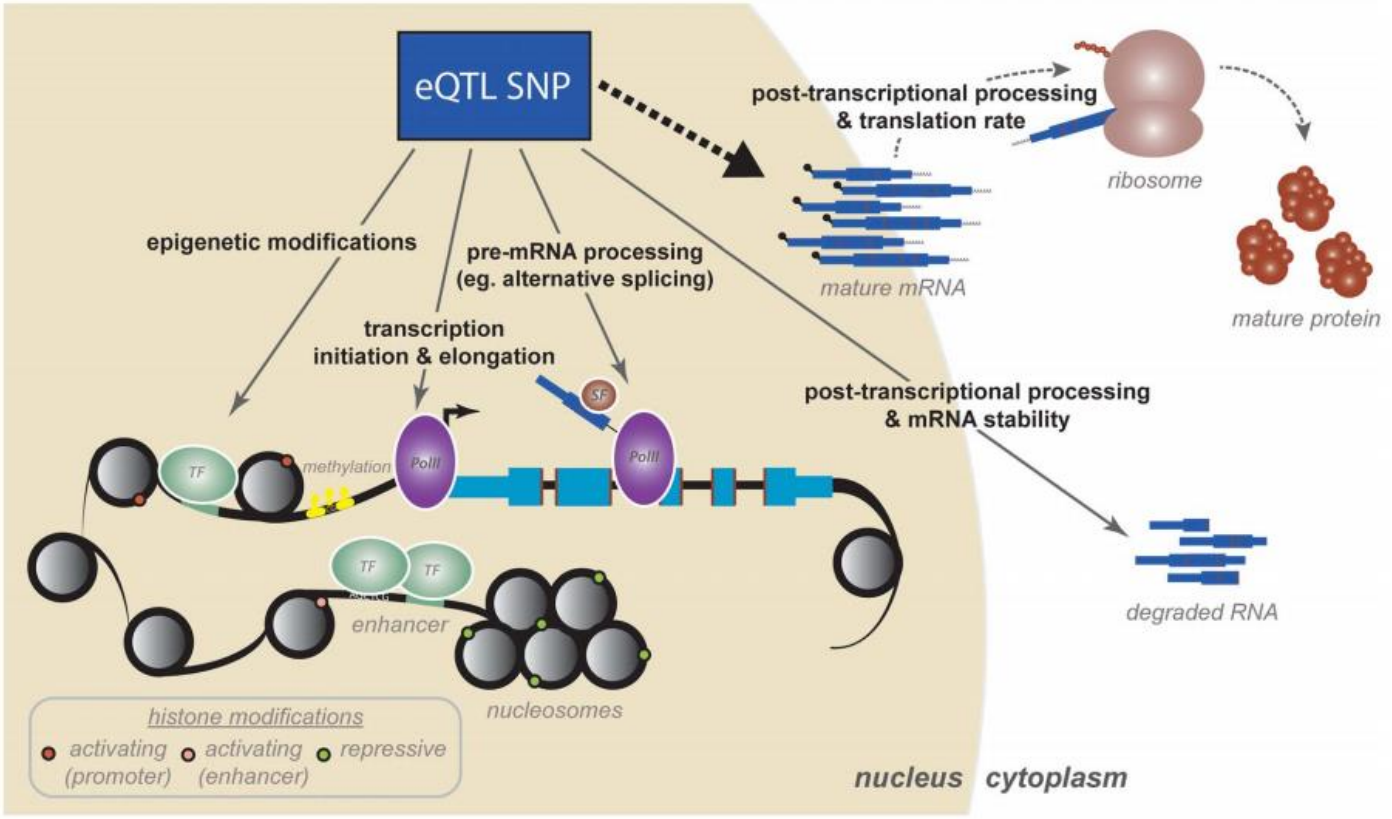


Fig. 1. A cascade of regulatory mechanisms by which an eQTL SNP can affect gene expression. Studies mapping regulatory QTLs have identified a variety of mechanisms, many of which are coordinated, by which eQTLs might act to affect variation in mature mRNA levels. First, eQTL SNPs can impact epigenetic modifications and transcription initiation. These include regulatory processes such as transcription factor binding, histone modifications, enhancer activity (perhaps mediated by chromatin architecture and conformation), and DNA methylation. Transcriptional mechanisms, and specifically transcription factor binding, are likely the strongest contributors to variation in steady-state mRNA levels. Second, recent work has increased appreciation for transcriptional and cotranscriptional processes as major contributors to variation in gene expression levels and mRNA isoform diversity. These include mechanisms such as transcriptional elongation (by PolII traveling rates), cotranscriptional splicing, and mRNA processing and modification. Third, eQTL SNPs both within and outside the transcript have been shown to influence posttranscriptional mRNA processing, which includes mechanisms such as general mRNA degradation, defects in polyadenylation, and targeting by miRNAs. Finally, preliminary studies have shown that we do not yet fully appreciate the extent to which variation in mRNA expression might impact or even correlate to variation in downstream protein products, the synthesis of which are additionally regulated by a set of posttranscriptional and translational mechanisms.
doi:10.1371/journal.pgen.1004857.g001

Genetic Variants and Genome Biology

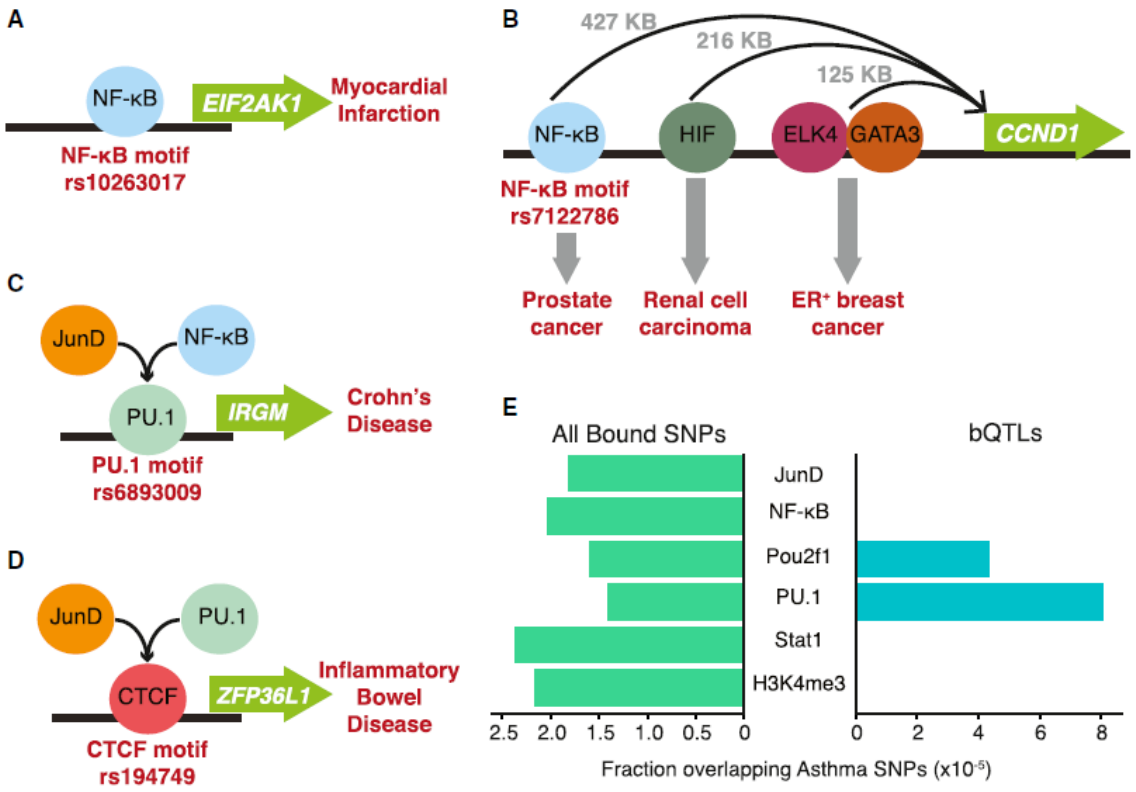


Figure 6. Effects of bQTLs on Disease Risk

Each example involves a SNP in a TF binding motif that is also a bQTL and is associated with disease risk.

(A) bQTL for NF-κB (highlighted in Figure 1B) in an NF-κB binding motif that is also an eQTL for *EIF2AK1* and associated with myocardial infarction.

(B) bQTL for NF-κB in an NF-κB binding motif that is also an eQTL for *CCND1* and is associated with prostate cancer. See also Figure S5.

(C) bQTL for PU.1, NF-κB, and JunD in a PU.1 binding motif that is also associated with Crohn's disease.

(D) bQTL for JunD and PU.1 in a CTCF binding motif that is also associated with inflammatory bowel disease.

(E) Comparing enrichments for asthma-associated SNPs: all bound SNPs (left) and bQTLs (right).

Kingston University London

Pooled ChIP-Seq Links Variation in Transcription Factor Binding to Complex Disease Risk

Ashley K. Tehranchi,[1] Marsha Myrthil,[2] Trevor Martin,[1] Brian L. Hie,[3] David Golan,[4,5] and Hunter B. Fraser[1,*]

Genetic Variants and Genome Biology

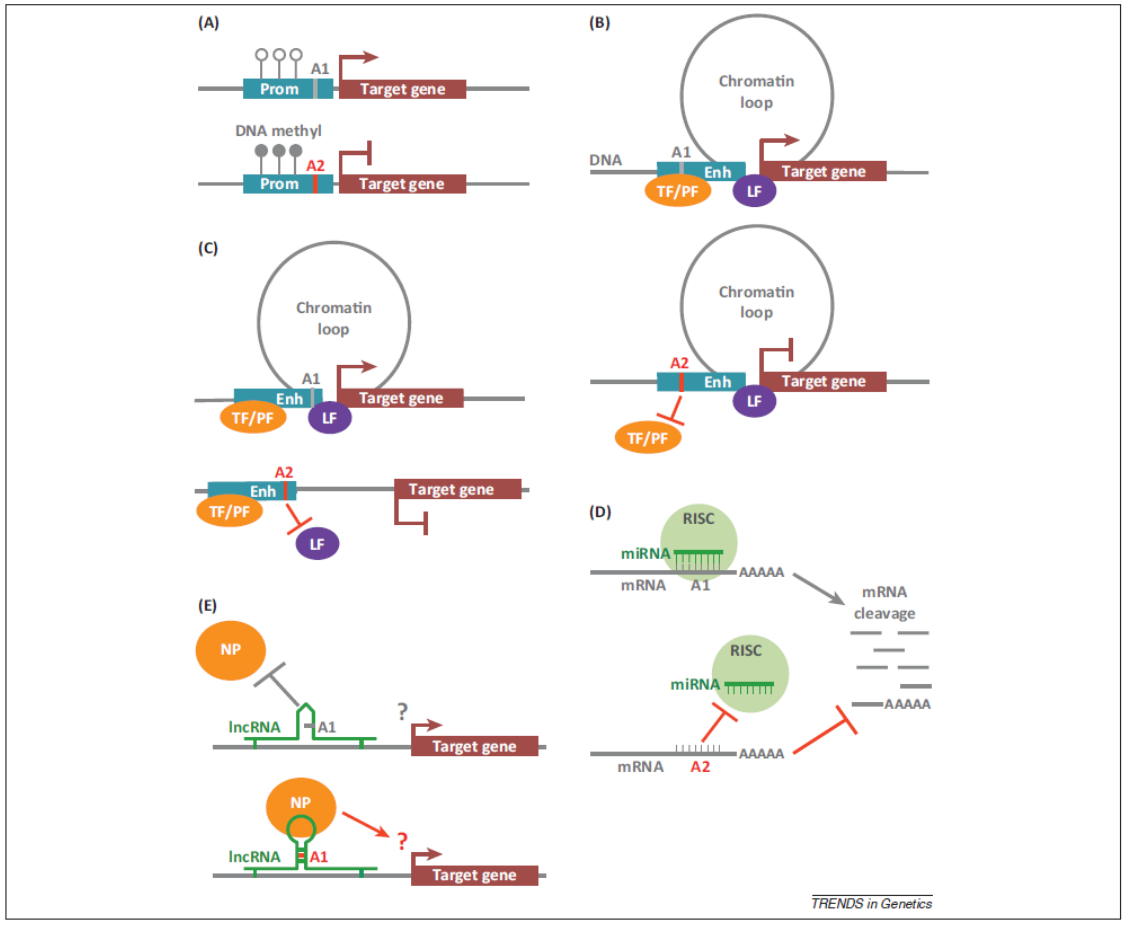


Figure 2. Non-coding genetic risk variants. **(A)** Genetic risk variants influence DNA methylation level at promoter regions. **(B)** Genetic risk variants modulate transcription factor binding to chromatin. **(C)** Genetic risk variants alter chromatin loop formation bridging enhancers and promoters. **(D)** Genetic risk variants influence the repression effect of miRNAs. **(E)** Genetic risk variants influence the interaction of lncRNAs with target proteins. Abbreviations: Enh, enhancer; LF, chromatin looping factor; lncRNA, long non-coding RNA; miRNA, microRNA; NP, nuclear protein; PF, pioneer factor; Prom, promoter; RISC: RNA-induced silencing complex; TF, transcription factor.

Genetic risk variants can alter chromatin loop formation bridging enhancers and promoters

The human genome is organized in a 3D architecture which is thought to regulate a diverse set of DNA-templated processes [87–91]. This allows regulatory elements, such as enhancers and promoters, to interact physically through long-range chromatin interactions, or chromatin loops, to regulate gene expression [39,42]. The human pigmentation-associated SNP, rs12913832, imposes allele-specific chromatin loop formation [92]. The rs12913832 SNP resides in an enhancer 21 kb upstream of the *OCA2* (oculocutaneous albinism II) pigment gene [92]. The T allele of this SNP favors chromatin loops to the *OCA2* gene compared to the C allele and is associated with a darker pigmentation in melanocytes [92]. Specific DNA binding proteins, including the cohesin and mediator complex as well as the insulator protein CTCF (CCCTC-binding factor), promote chromatin loop formation [93–95]. Although the rs12913832 SNP is the only genetic risk variant known to modulate chromatin loop formation, variants altering the DNA affinity for looping factors will likely also result in allele-specific chromatin loop formation (Figure 2C).

Kingston University London

Laying a solid foundation for Manhattan – 'setting the functional basis for the post-GWAS era'

Xiaoyang Zhang[1*,†], Swneke D. Bailey[2,3,†], and Mathieu Lupien[2,3,4]

3D Genome: Chromatin Interactions

The eukaryotic nucleus is a complex 3D environment in which genome function depends not only on the linear arrangement of regulatory sequence elements, but also on their spatial organization for effective control of gene expression.

Analysis of the role of chromatin 3D organization in gene expression is progressing rapidly, largely due to the development of chromosome conformation capture methods such as Hi-C.

Sequences within "Topologically Associated Domains" (TADs) interact more frequently with sites inside than outside the domain. TADs with a median size of 880 kb have been found in mammals.

Kingston University London

Breaking TADs: How Alterations of Chromatin Domains Result in Disease

Darío G. Lupiáñez,[1,2,3,@] Malte Spielmann,[1,2,3] and Stefan Mundlos[1,2,3,*]

3D Genome: Chromatin Interactions

Hi-C Contact Map

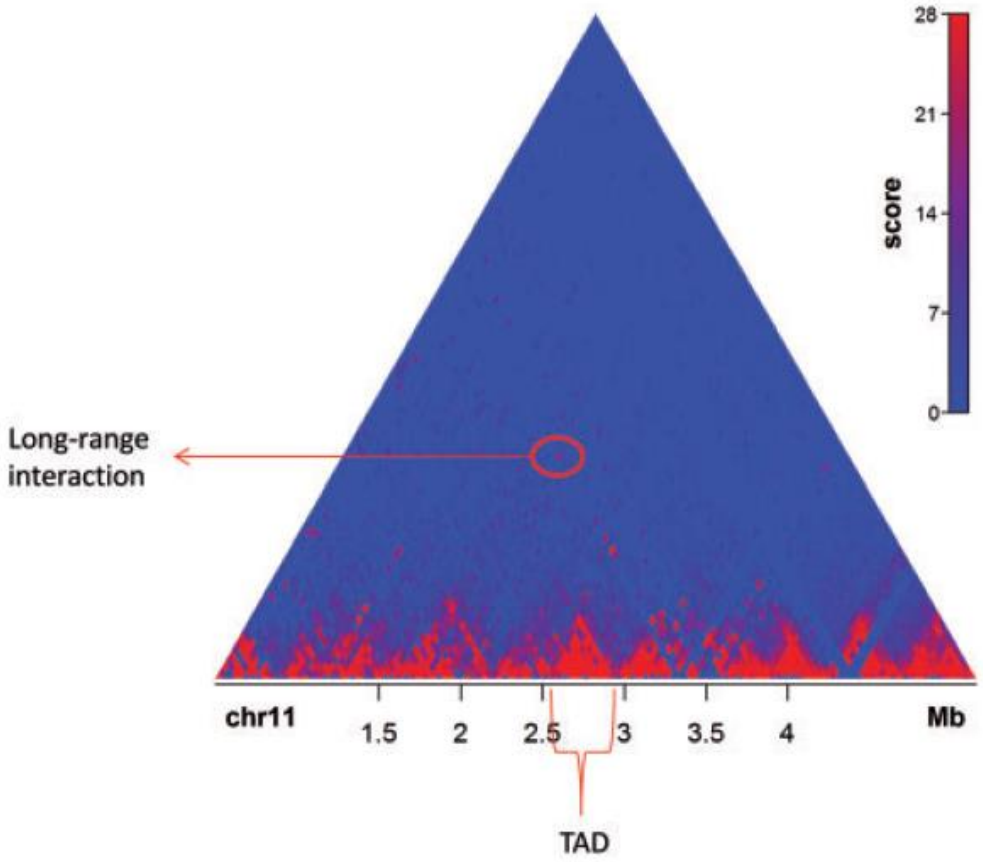


Figure 2. An example of a Hi-C contact map. Hi-C contact map of a segment of mouse chromosome 11, generated using Sushi [90] from Dixon *et al.* [85] data. A TAD and a long-range interaction between two loci are annotated. A colour version of this figure is available online at BIB online: https://academic.oup.com/bib.

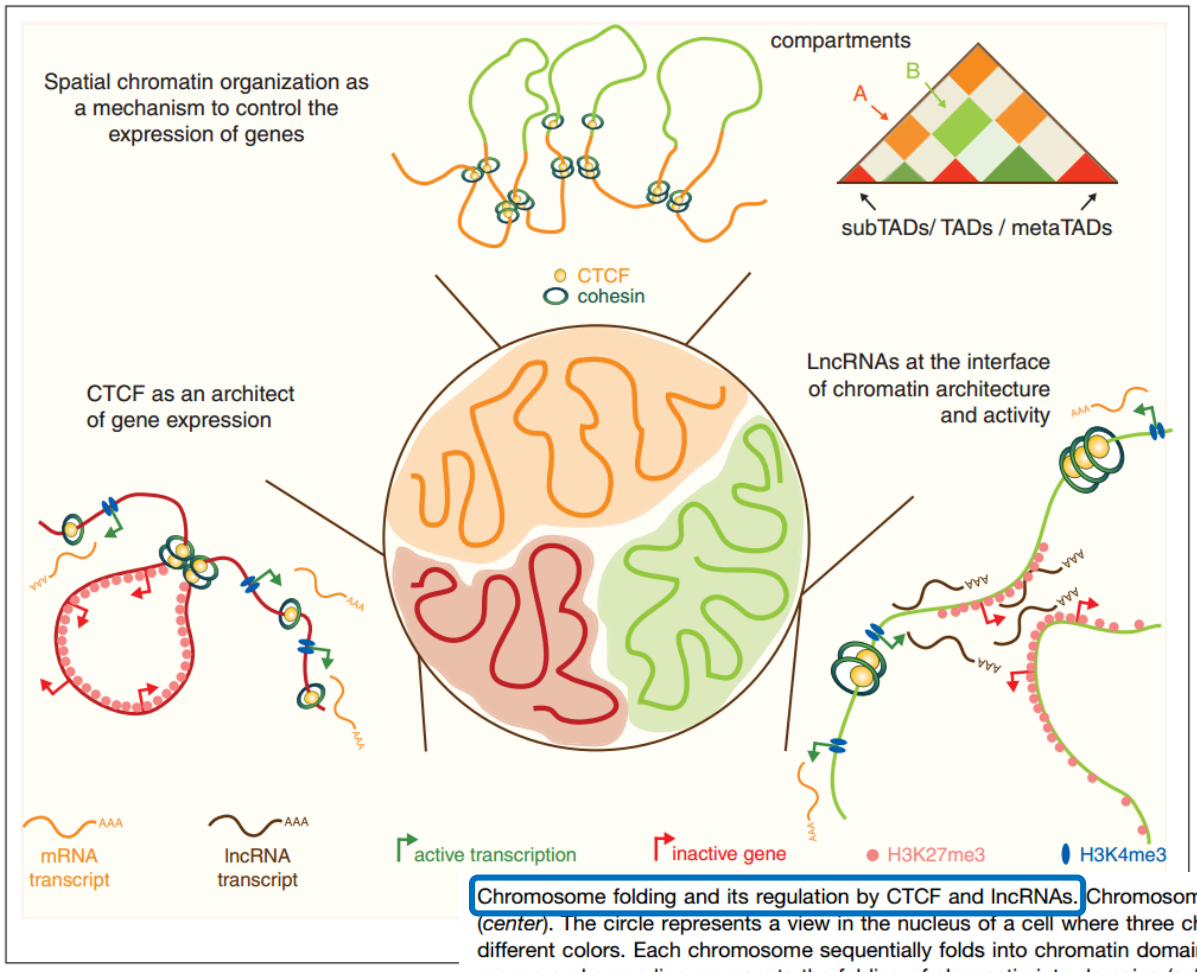
Briefings in Bioinformatics, 17(6), 2016, 980–995

doi: 10.1093/bib/bbv097
Advance Access Publication Date: 19 November 2015
Software Review

In the loop: promoter–enhancer interactions and bioinformatics

Antonio Mora, Geir Kjetil Sandve, Odd Stokke Gabrielsen and Ragnhild Eskeland

3D Genome: Chromatin Interactions



Chromosome folding and its regulation in health and disease

Xue Qing David Wang and Josée Dostie

Chromosome folding and its regulation by CTCF and lncRNAs. Chromosomes occupy distinct territories in the nucleus of mammalian cells (*center*). The circle represents a view in the nucleus of a cell where three chromosomes (lines) and their territories (highlights) are shown in different colors. Each chromosome sequentially folds into chromatin domains across size scales to guide proper gene expression (*top*). The orange and green line represents the folding of chromatin into domains (subTAD/TAD/metaTAD) delimited by the cohesin complex and architectural protein CTCF. Areas where the line is orange indicate a domain where transcription is active whereas inactive domains are shown with green lines. Active and inactive domains preferentially interact with each other respectively, and they tend to spatially colocalize in either active (A, orange) or inactive (B, green) compartments. Chromatin organization into domains or compartments as measured with 3C technologies is shown in heatmap form (*top right*). CTCF and cohesin regulate transcription by various mechanisms including the formation of loop structures that restrict the spread of activating histone marks (*left*). In the example shown, CTCF binding and dimerization forms a loop structure that contains a cluster of inactive genes (red arrows) silenced by H3K27me3 (pinks dots). Formation of this subTAD also serves to prevent further spreading of silencing marks into neighboring regions where genes are active (green arrows). lncRNAs bind histone-modifying complexes to regulate chromatin activity and structure (*right*). In this example, transcription of a lncRNA gene yields transcripts that recruit chromatin-modifying complexes at nearby genes to silence them while CTCF provides boundaries to limit the scope of this effect. This regulation can occur within chromosomes (*cis*) or between them (*trans*).

Kingston University London

3D Genome: Chromatin Interactions

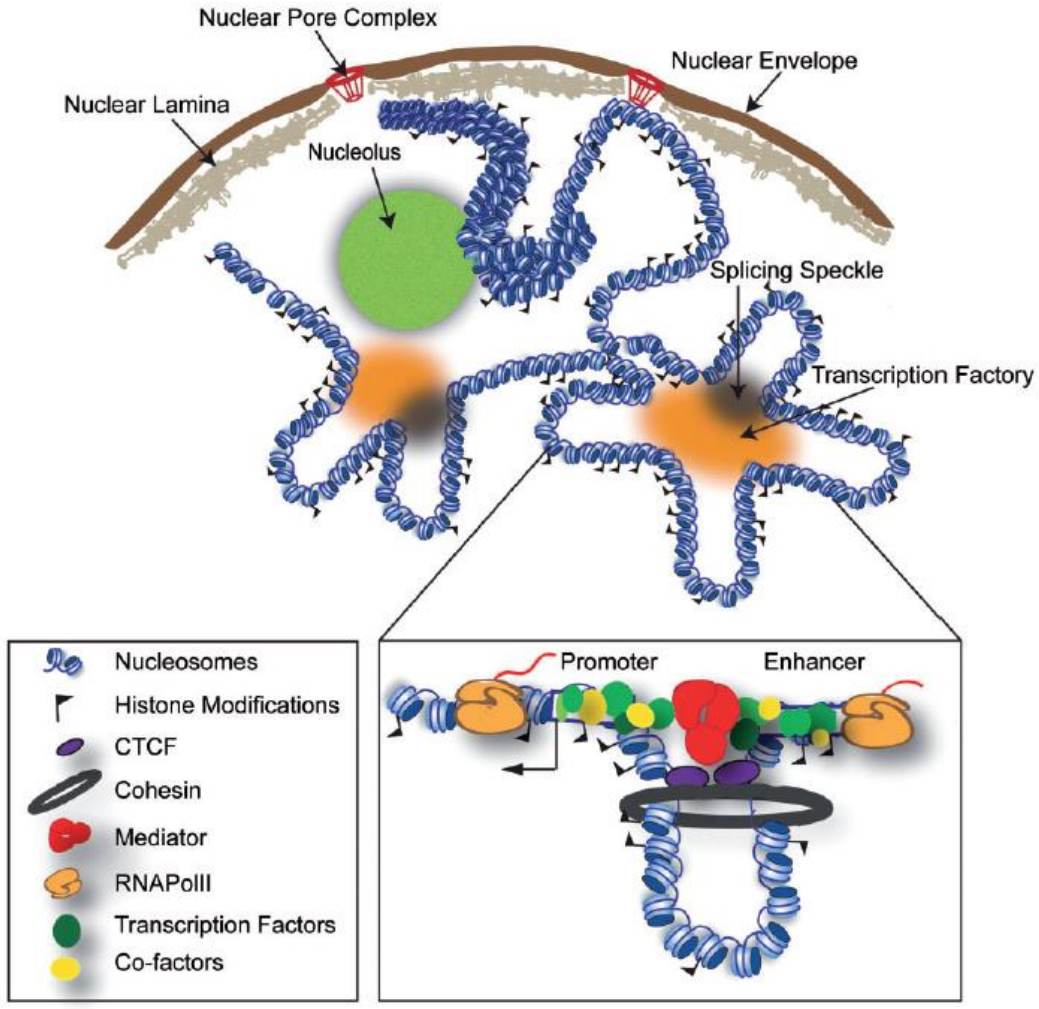


Figure 1. Models of chromatin organization. A diagram of different models of chromatin organization in the nuclear space. Interphase chromatin that interacts with the nuclear lamina (grey), nucleolus (green), nuclear pores (red), transcription factories (orange) and splicing speckles (black) are depicted here. Generally, lamin- and nucleolar-associated domains are transcriptionally repressed and have a more condensed chromatin, whereas chromatin that loops to the nuclear pore, transcription factories and splicing speckles are transcriptionally active and therefore have a more open chromatin structure (here, depicted as 10 nm chromatin fibre). Enhancers can activate gene expression over a distance and contain binding sites for TFs that recruit co-factors (activators or repressors). A promoter–enhancer looping mechanism mediated by cohesin (brown), CTCF (purple) and the mediator complex (red) that brings the enhancer into close proximity to its target promoter are presented in the enlarged box. The enhancer and promoter are marked with white boxes, and the transcription start site of the transcribed target gene is annotated with an arrow. TFs (green) and co-factors (yellow) bind the enhancer and are brought close to the basal transcription machinery at the promoter. RNAPolII (orange) transcribes pre-mRNA from the target gene and eRNA from the enhancer. Some of these models may co-exist for different PEIs; however, there are also other models that we could not show. A colour version of this figure is available online at BIB online: https://academic.oup.com/bib.

OXFORD

Briefings in Bioinformatics, 17(6), 2016, 980–995

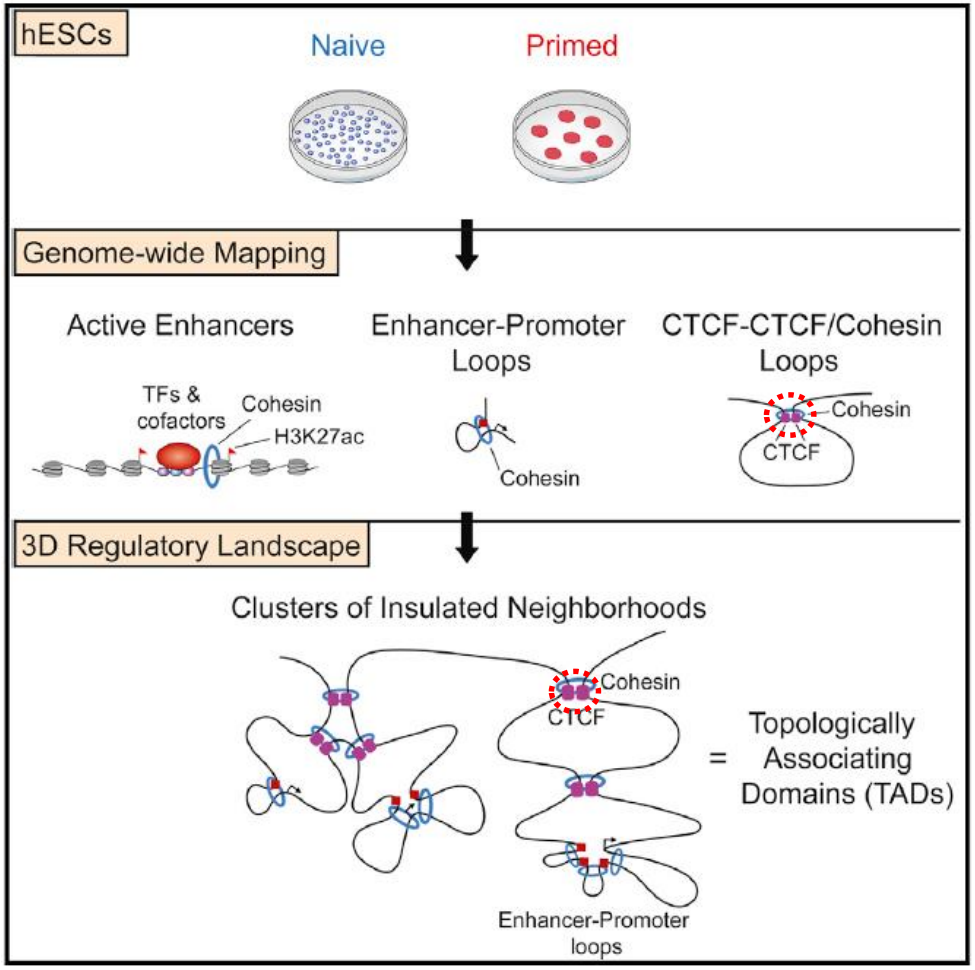
doi: 10.1093/bib/bbv097
Advance Access Publication Date: 19 November 2015
Software Review

In the loop: promoter–enhancer interactions and bioinformatics

Antonio Mora, Geir Kjetil Sandve, Odd Stokke Gabrielsen and Ragnhild Eskeland

3D Genome: Chromatin Interactions

Ji et al. map the chromosome organizational structures that underlie gene regulation in human naive and primed pluripotent cells. Their framework of cohesin-associated CTCF loops, and the cohesin-associated enhancer-promoter loops within them, provides a reference map for future interrogation of regulatory interactions.



- ChIA-PET analysis maps enhancers and insulators into looped domains
- Cohesin-associated loops organize topologically associating domains (TADs)
- Regulatory changes during cell state transitions take place within TADs
- The conserved anchors of CTCF-CTCF loops are frequently mutated in cancer

Kingston University London

3D Chromosome Regulatory Landscape of Human Pluripotent Cells

Xiong Ji,[1,6] Daniel B. Dadon,[1,2,6] Benjamin E. Powell,[1,6] Zi Peng Fan,[1,3,6] Diego Borges-Rivera,[1,2,6] Sigal Shachar,[4] Abraham S. Weintraub,[1,2] Denes Hnisz,[1] Gianluca Pegoraro,[5] Tong Ihn Lee,[1] Tom Misteli,[4] Rudolf Jaenisch,[1,2,*] and Richard A. Young[1,2,*]

3D Genome: Chromatin Interactions

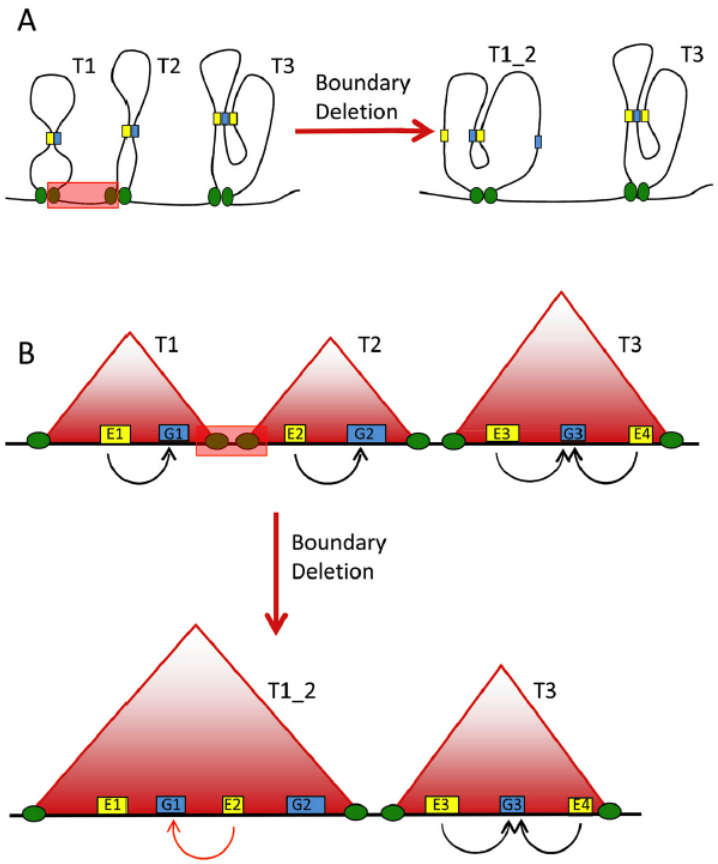


Figure 1. **Schematic diagram of regulatory re-wiring following the deletion of a domain boundary.** (**A**) Interactions between enhancers and their target genes occur within chromatin domains. The deletion of a boundary region leads to novel gene-enhancer interactions between previously insulated elements; this process may lead to the spatial or temporal mis-expression of genes. (**B**) The same scenario as in (**A**) is drawn as represented by a high-throughput chromosome conformation capture (Hi-C) interaction map. Red triangles: topologically associating domains; yellow boxes: regulatory elements; blue boxes: target genes; green circles: insulator elements. Further examples of pathogenic genomic rearrangements, including insulator-spanning tandem duplications, are illustrated in 31.

Kingston University London

F1000Research F1000Research 2017, 6(F1000 Faculty Rev):314 Last updated: 24 MAR 2017

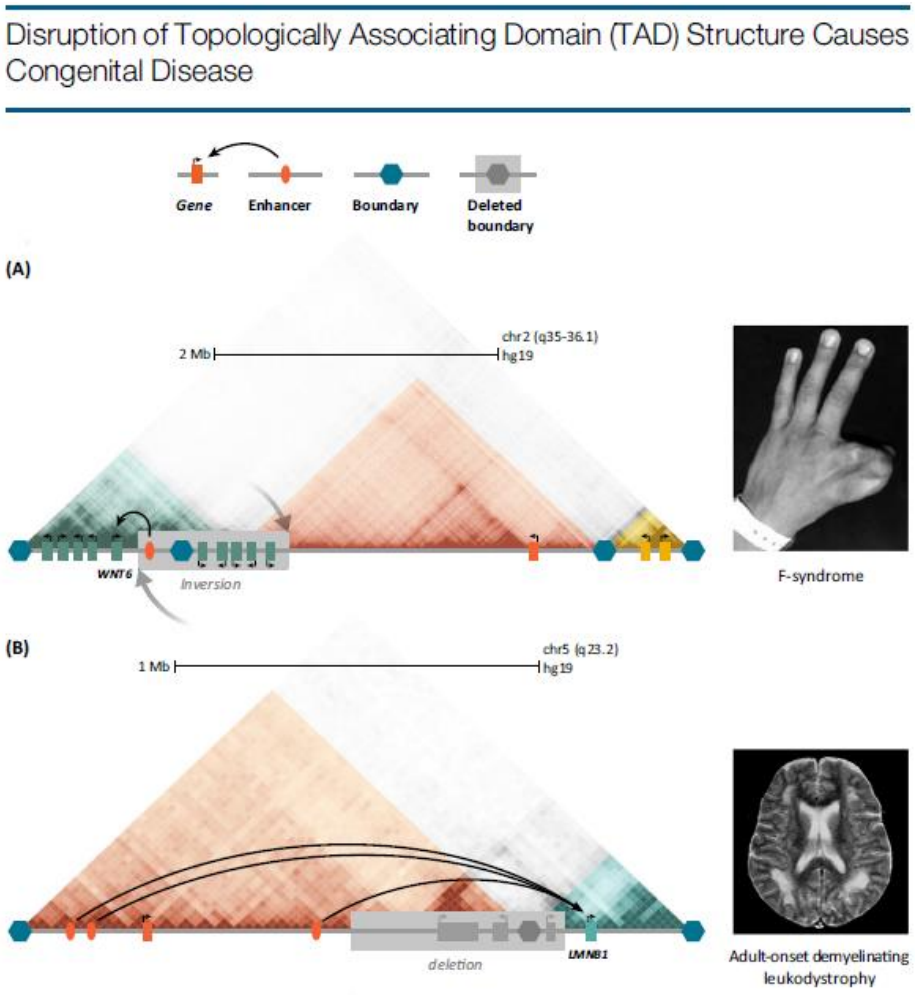
Check for updates

REVIEW

When TADs go bad: chromatin structure and nuclear organisation in human disease [version 1; referees: 2 approved]

Vera B Kaiser, Colin A Semple

3D Genome: Chromatin Interactions



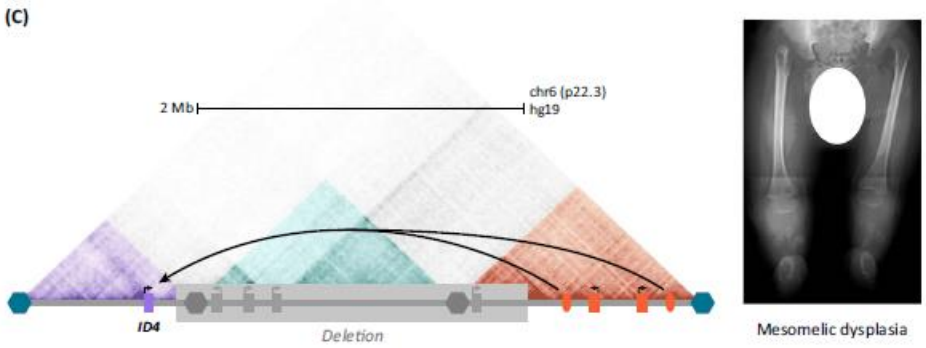


Figure 3. (A) F syndrome is a rare dominantly inherited skeletal disorder characterized by syndactyly of the first and second fingers. An inversion leaves the TAD boundary intact but places a cluster of limb enhancers from a neighboring TAD in front of WNT6 causing misexpression in digit 1 and 2. (B) Autosomal-dominant adult-onset demyelinating leukodystrophy (ADLD) is a rare neurological disorder characterized by progressive central nervous system demyelination due to overexpression of LMNB1. A 600-kb deletion including a TAD boundary was shown to result in pathological interactions between three strong forebrain enhancer elements and the LMNB1 promoter resulting in cerebral lamin B1 overexpression and myelin degeneration. (C) Structural variations can alter the TAD architecture of the genome by deleting, duplicating, or inverting TADs and their boundaries, thereby allowing enhancers from neighboring domains to ectopically activate genes causing misexpression and disease. Deletions on chromosome 6p22.3 have been shown to cause mesomelic dysplasia featuring hypoplastic tibiae and fibulae. The deletions span three TADs and remove two TAD boundaries. This brings several potential limb enhancers into close proximity with ID4, presumably resulting in misexpression in the developing limb bud.

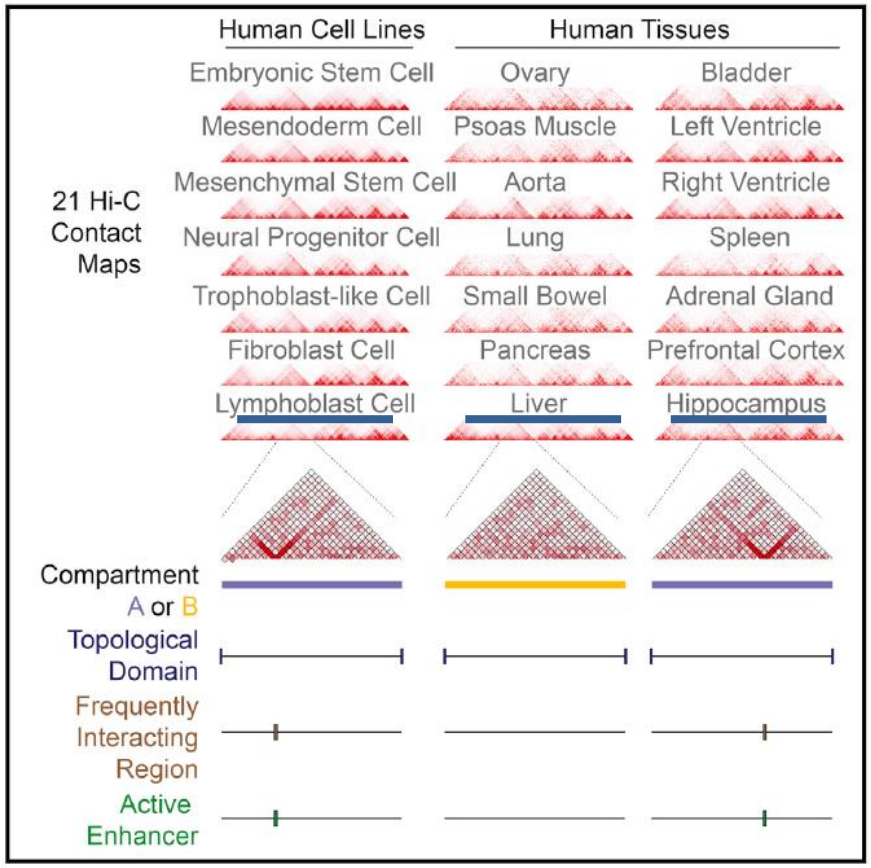
Breaking TADs: How Alterations of Chromatin Domains Result in Disease

Darío G. Lupiáñez,[1,2,3,@] Malte Spielmann,[1,2,3] and Stefan Mundlos[1,2,3,*]

Kingston University London

3D Genome: Chromatin Interactions

Tissue Specificity



In Brief

Schmitt et al. analyze Hi-C maps in 21 human cell lines and primary tissues and uncover a class of genome organizational features termed FIREs. FIREs are local interaction hotspots, highly tissue-specific, and correspond to active enhancers. We discuss the implications of our findings for the study of gene regulation and disease. Explore the Cell Press IHEC web portal at http://www.cell.com/consortium/IHEC.

Highlights

- Integrative analysis of chromatin architecture in a broad set of human tissues
- FIREs are an architectural feature of chromatin organization
- FIREs are enriched for super-enhancers and show tissue-specific chromatin interactions
- FIRE formation is partially dependent on CTCF and the Cohesin complex

Kingston University London

A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]

3D Genome: Chromatin Interactions

Distance

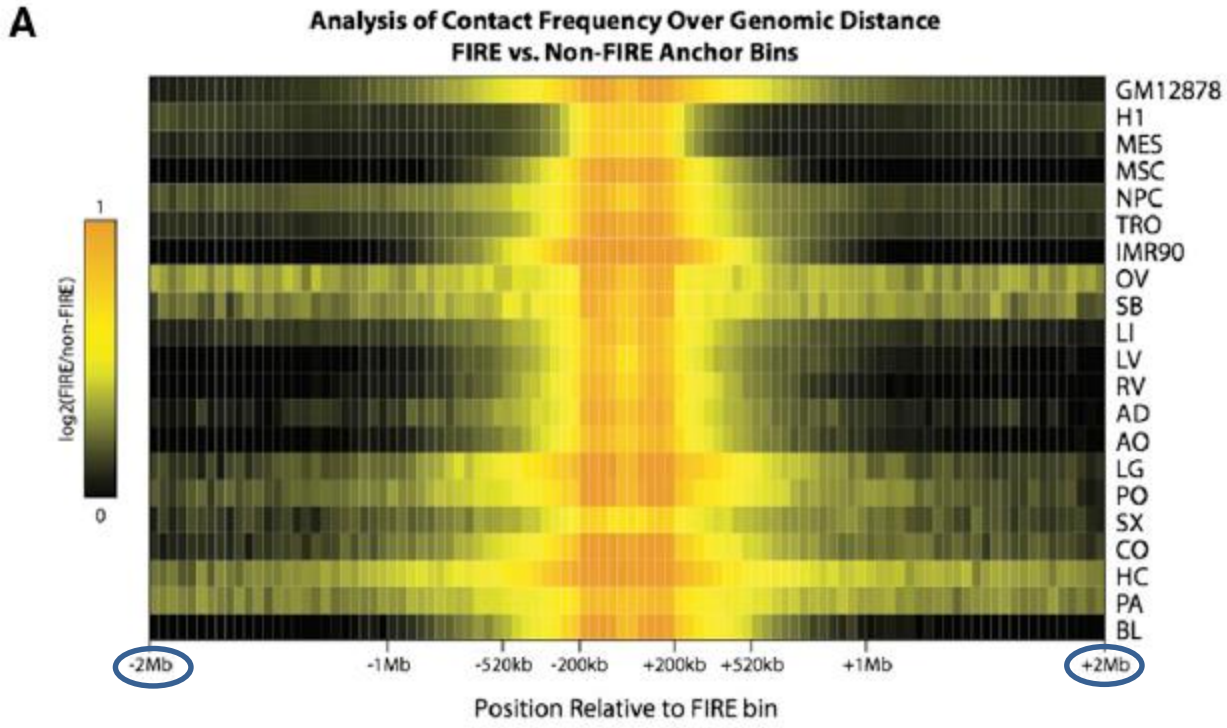


Figure 7. FIREs Have Several Targets and Are Self-Interactive

(A) Heat map showing the relationship between the mean observed contact frequencies at FIREs compared to the mean observed contact frequency at non-FIREs. Enrichment is shown as the ratio between the two contact observed mean contact frequencies (FIRE:non-FIRE) per unit genomic distance, from ± 40 kb to ± 2 Mb, centered on FIRE bins. Each row represents the analysis of a different sample, and the color intensity corresponds to the enrichment value.

Kingston University London

A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]

3D Genome: Chromatin Interactions

Tissue Specificity

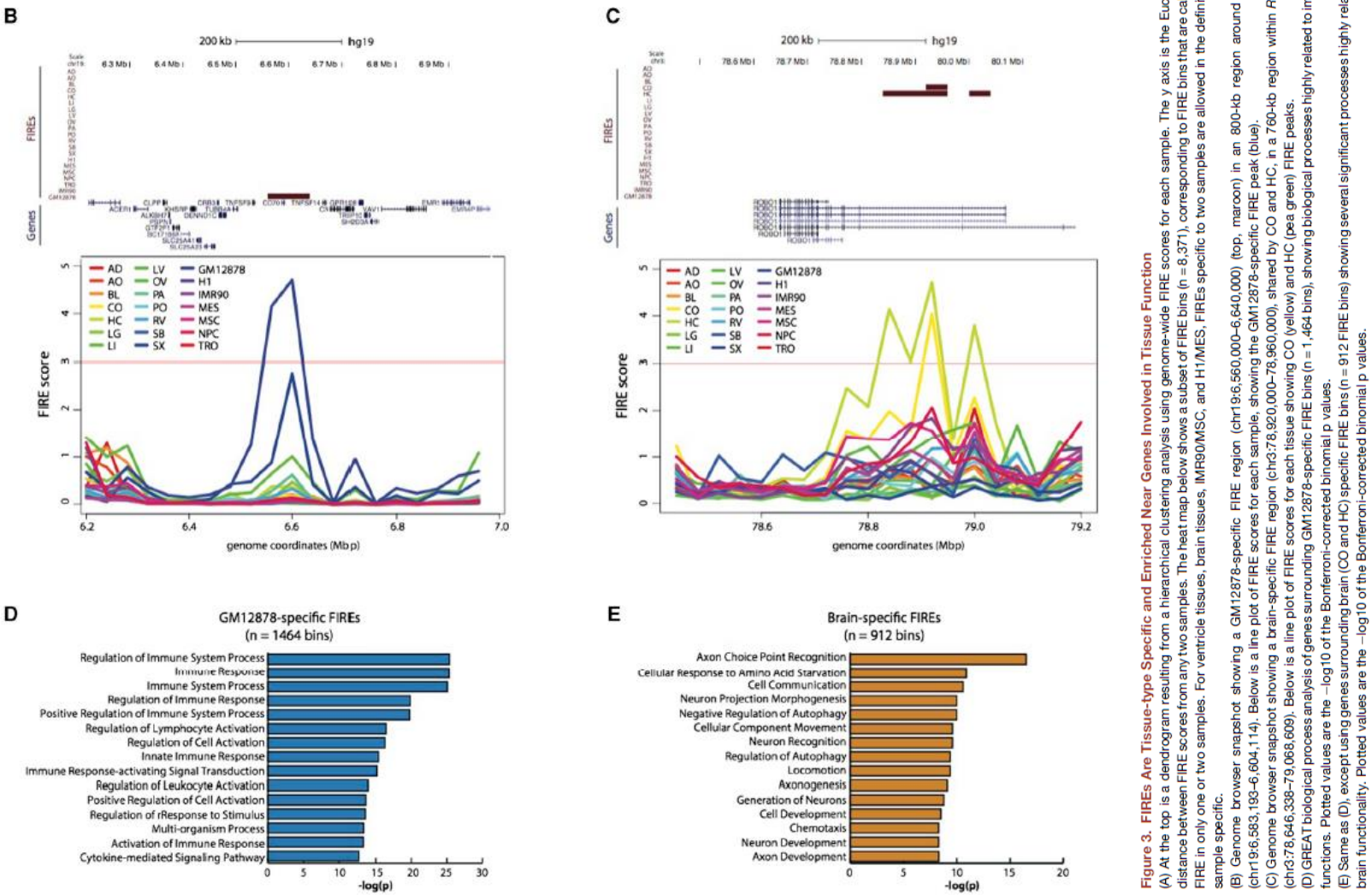


Figure 3. FIREs Are Tissue-type Specific and Enriched Near Genes Involved in Tissue Function

(A) At the top is a dendrogram resulting from a hierarchical clustering analysis using genome-wide FIRE scores for each sample. The y axis is the Euclidean distance between FIRE scores from any two samples. The heat map below shows a subset of FIRE bins (n = 8,371), corresponding to FIRE bins that are called as FIRE in only one or two samples. For ventricle tissues, brain tissues, IMR90/MSC, and H1/MES, FIREs specific to two samples are allowed in the definition of sample specific.

(B) Genome browser snapshot showing a GM12878-specific FIRE region (chr19:6,560,000–6,640,000) (top, maroon) in an 800-kb region around *CD70* (chr19:6,583,193–6,604,114). Below is a line plot of FIRE scores for each sample, showing the GM12878-specific FIRE peak (blue).

(C) Genome browser snapshot showing a brain-specific FIRE region (chr3:78,920,000–78,960,000), shared by CO and HC, in a 760-kb region within *ROBO1* (chr3:78,646,338–79,068,609). Below is a line plot of FIRE scores for each tissue showing CO (yellow) and HC (pea green) FIRE peaks.

(D) GREAT biological process analysis of genes surrounding GM12878-specific FIRE bins (n = 1,464 bins), showing biological processes highly related to immune functions. Plotted values are the −log10 of the Bonferroni-corrected binomial p values.

(E) Same as (D), except using genes surrounding brain (CO and HC) specific FIRE bins (n = 912 FIRE bins) showing several significant processes highly related to brain functionality. Plotted values are the −log10 of the Bonferroni-corrected binomial p values.

A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]

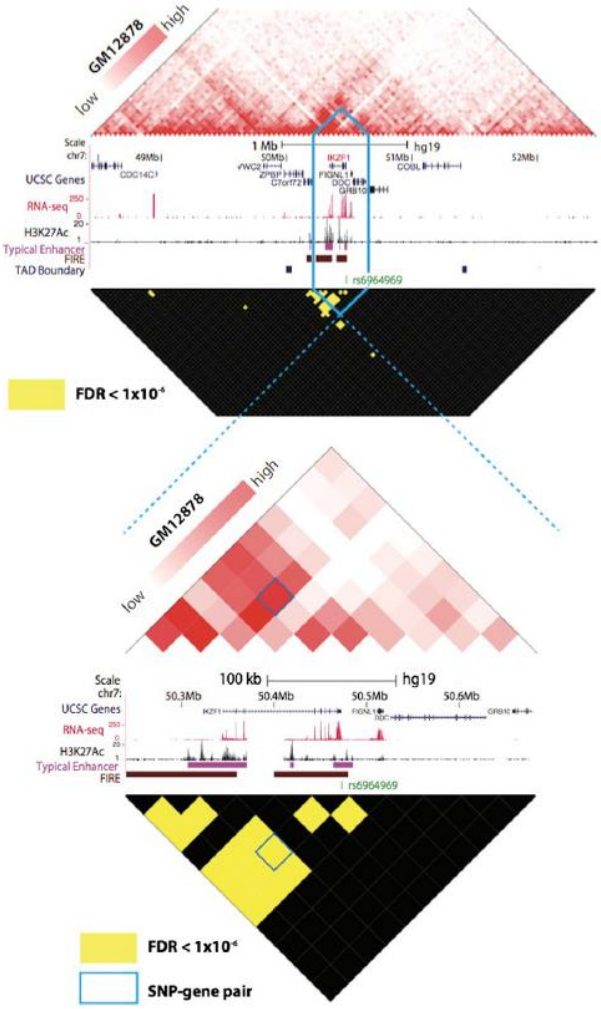
Kingston University London

3D Genome: Chromatin Interactions

Disease Associations

FIREs Are Enriched for Disease-Associated SNPs

Our analyses have indicated that FIREs are enriched for active enhancers and super-enhancers (Figures 4A–4D; Figures S3B, S3C, S3F, S3G, S3N, and S3O). Because typical and super-enhancers contain a significant proportion of disease-associated SNPs (Hnisz et al., 2013), we further investigated the overlap between FIREs and disease-associated SNPs. First, we mapped 4,327 previously annotated disease-associated non-coding SNPs to FIREs defined in each cell line and tissue (see Supplemental Experimental Procedures) (Hnisz et al., 2013). Consistent with previous results (Hnisz et al., 2013), we observed 7.06 and 3.76 SNPs per megabase, and among 354 GM12878 FIREs overlapped with super-enhancers and 2,800 GM12878 FIREs overlapped with typical enhancers, respectively (Figure S5A). Surprisingly, among 1,615 GM12878 FIREs that do *not* overlap an annotated enhancer, we also observed 3.33 SNPs per megabase, which is ~2.3-fold higher than the genome-wide SNP density (1.42 SNPs per megabase) (Figure S5A). Importantly, these SNPs would not be captured by directly overlapping super-enhancers or typical enhancers with disease-associated SNPs (Hnisz et al., 2013).



Kingston University London

A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]

3D Genome: Chromatin Interactions

IKZF1 rs6964969
Childhood A.L.L.

chr7:50472800-50473400	Flanking Active TSS	GM12878-XiMat	blood

Figure 6. FIREs Are Enriched with Disease-Associated GWAS SNPs

(A) Heat map showing the enrichment of disease-associated GWAS SNPs (see Supplemental Experimental Procedures) in FIRE bins for each cell line or tissue (columns). Rows represent the enrichment of disease-associated SNPs for one disease, and all rows in the presented heat map are sorted from high to low based on enrichment score in GM12878 (lymphoblast cell line). Only diseases with >15 SNPs are shown. Noted to the right are the top 15 diseases for which disease-associated SNPs are most enriched in GM12878 FIREs, showing the high enrichment of several diseases (all except mean corpuscular volume) with previously noted immune-mediated pathology (Jostins et al., 2012).

(B) Normalized Hi-C contact matrix of a 2.16-Mb locus (chr1:65,120,000–67,280,000) in GM12878 cells. The tracks below depict the presence of two SNPs associated with acute lymphoblastic leukemia (rs546784 and rs6683977) located within a FIRE bin (brown, chr1:66,760,000–66,800,000), ~30 kb outside of a GM12878-specific super-enhancer (red) and also within the *PDE4B* gene sequence. To the right of the Hi-C contact matrix is the FIRE score.

(C) Bar plots showing the enrichment of Parkinson's disease-associated SNPs across 14 primary adult tissue FIRE annotations, also highlighting the highest enrichment in FIREs from both brain tissues (CO and HC).

(D) Bar plots showing the enrichment of SNPs associated with the quantitative triglycerides trait across 14 primary adult tissue FIRE annotations, also highlighting the highest enrichment in liver FIREs.

(E) Normalized Hi-C contact matrix (top) in GM12878 for a 4.04-Mb locus (chr7:48,440,000–52,480,000) centered on *IKZF1* (red text). The Hi-C color scale ranges from the 15th to 99th percentile normalized contact frequencies within this locus. The reflected matrix shows the statistically significant (FDR < 1e−6) bin-pairs within 2-Mb genomic distance across the locus. Only bin pairs with FDR < 1e−6 are yellow; the rest are black. Between the matrices are a UCSC gene annotations (blue, top), RNA-seq data (red), H3K27Ac data (black), typical enhancer annotations (Hnisz et al., 2013) (purple), FIRE annotations (brown), TAD boundary calls (blue), and an SNP that is statistically linked to the IKZF1 TSS (green). The blue lines outline the 440-kb locus (chr7:50,240,000–50,680,000) that is shown in (F).

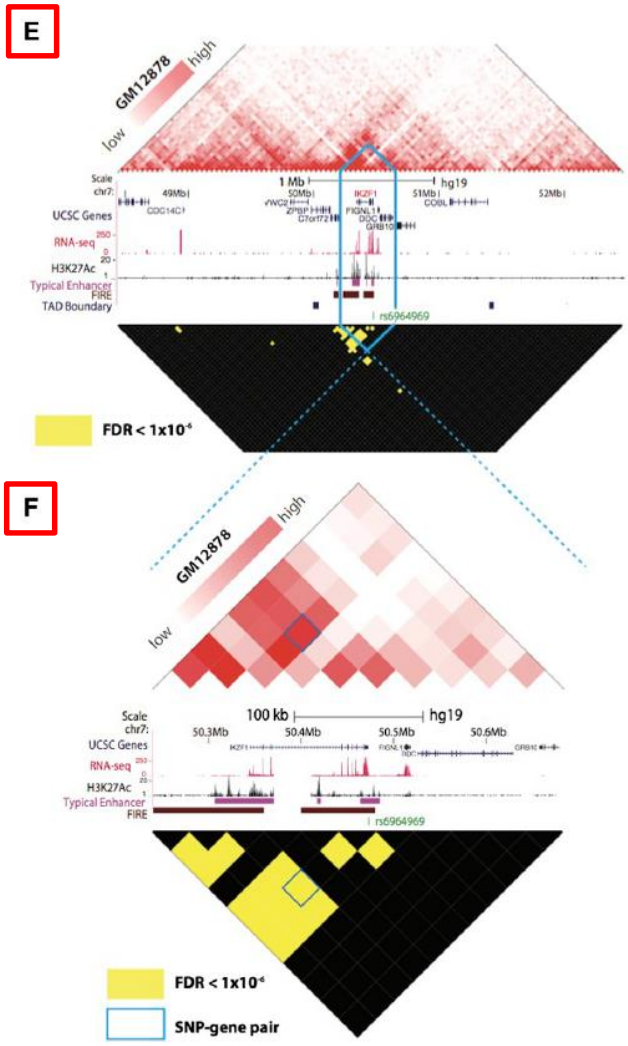
(F) Same as (E), except a zoomed-in snapshot of a 440-kb locus (chr7:50,240,000–50,680,000) centered on a SNP-bearing FIRE bin (chr7:50,440,000–50,480,000) containing the 3′ UTR of *IKZF1* and the SNP rs6964969. The blue box outlines the bin pair that is the significant interaction between previously known SNP-gene pairs.

(G) Bar plots showing the enrichment of liver GTEx eQTLs in FIRE peak bin pairs as a function of the subset of top liver FIRE peaks (based on the lowest false discovery rate) determined by Fit-Hi-C.

(H) Same as (G), except using aorta GTEx eQTLs, FIREs, and FIRE peaks.

(I) Same as (G), except using left ventricle GTEx eQTLs, FIREs, and FIRE peaks.

(J) Same as (G), except using cortex GTEx eQTLs, FIREs, and FIRE peaks.



Kingston University London

A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

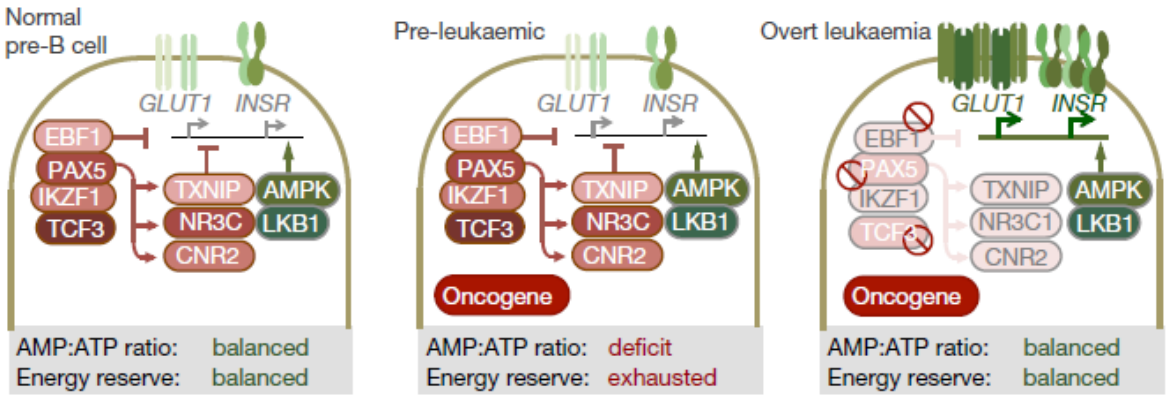
Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]

IKZF1

Metabolic gatekeeper function of B-lymphoid transcription factors

Lai N. Chan[1,2], Zhengshan Chen[1,2], Daniel Braas[3], Jae-Woong Lee[1,2], Gang Xiao[1,2], Huimin Geng[4], Kadriye Nehir Cosgun[1,2], Christian Hurtz[4], Seyedmehdi Shojaee[4], Valeria Cazzaniga[4], Hilde Schjerven[4], Thomas Ernst[5], Andreas Hochhaus[5], Steven M. Kornblau[6], Marina Konopleva[6], Miles A. Pufall[7], Giovanni Cazzaniga[8], Grace J. Liu[9], Thomas A. Milne[10], H. Phillip Koeffler[11,12], Theodora S. Ross[13], Isidro Sánchez-García[14], Arndt Borkhardt[15], Keith R. Yamamoto[4], Ross A. Dickins[9], Thomas G. Graeber[3] & Markus Müschen[1,2]

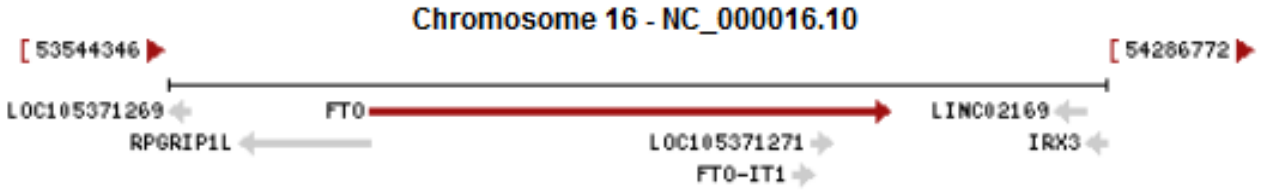
B-lymphoid transcription factors function as metabolic gatekeepers by limiting the amount of cellular ATP to levels that are insufficient for malignant transformation.



Kingston University London

3D Genome: Chromatin Interactions

FTO rs1421085



Chr16:53703963..54114467

Claussnitzer and colleagues recently showed that the *FTO* allele, which shows the strongest genome wide association signal for obesity, acts a gain of function. Using CRISPR/Cas9 genome editing, they showed that the disease-associated single-nucleotide variant rs1421085 T to C disrupts a conserved motif for the *ARID5B* repressor, which unleashes a pre-adipocyte enhancer, leading to a doubling of *IRX3* and *IRX5* expression during early adipocyte differentiation.

Kingston University London

3D Genome: Chromatin Interactions

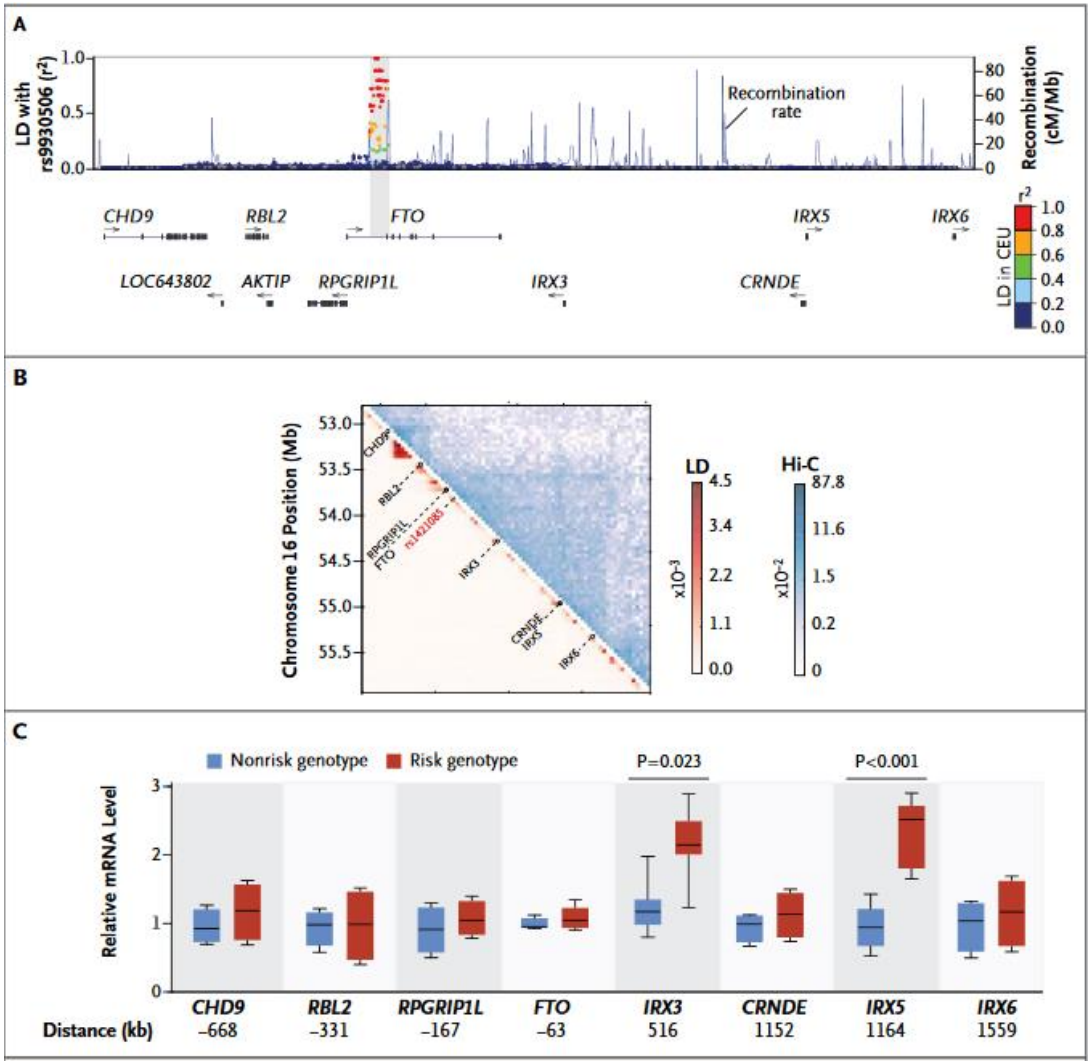


Figure 2. Activation of *IRX3* and *IRX5* Expression in Human Adipocyte Progenitors by the *FTO* Obesity Risk Genotype.
Panel A shows gene annotations and LD with array tag variant rs9930506 in a 2.5-Mb window; LD is expressed as r2 values in the CEU population. Arrows indicate the direction of transcription of annotated genes in the locus. **Panel B** shows chromosome conformation capture (Hi-C) interactions contact probabilities in human IMR90 myofibroblasts, revealing a 2-Mb topologically associating domain, and LD mean r2 statistics for all SNV pairs at 40-kb resolution. **Panel C** shows box plots for expression levels, after 2 days of differentiation, in human adipose progenitors isolated from 20 risk-allele carriers and 18 non-risk-allele carriers, evaluated by means of a quantitative PCR analysis for all genes in the 2.5-Mb locus. The horizontal line within each box represents the median, the top and bottom of each box indicate the 75th and 25th percentile, and I bars indicate the range.

Kingston University London

FTO Obesity Variant Circuitry and Adipocyte Browning in Humans

Melina Claussnitzer, Ph.D., Simon N. Dankel, Ph.D., Kyoung-Han Kim, Ph.D., Gerald Quon, Ph.D., Wouter Meuleman, Ph.D., Christine Haugen, M.Sc., Viktoria Glunk, M.Sc., Isabel S. Sousa, M.Sc., Jacqueline L. Beaudry, Ph.D., Vijitha Puviindran, B.Sc., Nezar A. Abdennur, M.Sc., Jannel Liu, B.Sc., Per-Arne Svensson, Ph.D., Yi-Hsiang Hsu, Ph.D., Daniel J. Drucker, M.D., Gunnar Mellgren, M.D., Ph.D., Chi-Chung Hui, Ph.D., Hans Hauner, M.D., and Manolis Kellis, Ph.D.

3D Genome: Chromatin Interactions

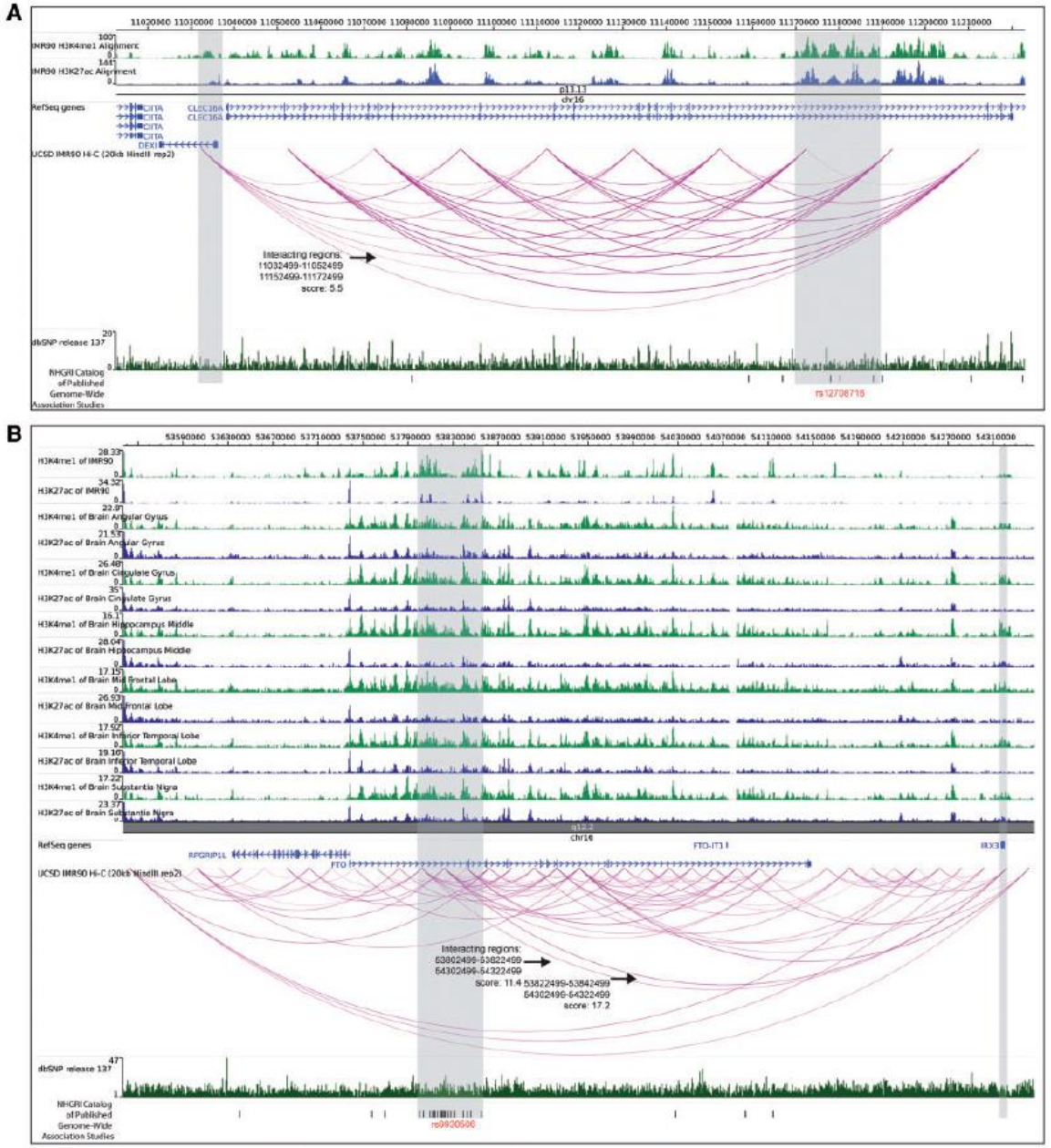


Figure 3. Long-range interactions functionally connect disease-associated SNPs with disease candidate genes. (**A**) Physical proximity between *DEXI* gene locus and autoimmune disease-associated SNPs in the intron of *CLEC16A* from human foetal lung (IMR-90) cells (from the Ren lab, [85]) show interactions between *CLEC16A* intron 19 and the *DEXI* locus. The enhancer marks in IMR-90 cells for H3K4me1 and H3K27ac are shown in green and blue, respectively, and the filter threshold for the Hi-C data was set to 5. SNPs in the region are in black, and the eQTL SNP rs12708716 is marked in red. The arc (pink) for interacting regions (grey) is highlighted with an arrow. (**B**) Long-range interactions links obesity-associated variants in *FTO* with the *IRX3* locus. Hi-C data in human foetal lung (IMR-90) cells show interactions between the first intron of *FTO* with *IRX3*. The tracks for H3K4me1 and H3K27ac are shown in green and blue from IMR-90 cells and different human brain tissues from the NIH Roadmap Epigenomics Mapping Consortium. The filter threshold for the Hi-C data was set to 10. SNPs in the region are in black, and the BMI-associated SNP rs9930506 is marked in red. Arcs (pink) for interacting regions (grey) are highlighted with arrows. These public data sets are available and visualized with the WashU EpiGenome Browser (http://epigenomegateway.wustl.edu/browser/). dbSNP release 137 is shown in dark green, and the The National Human Genome Research Institute (NHGRI) Catalogues of GWAS are visualized in UCSC browser (http://genome-euro.ucsc.edu) [166]. A colour version of this figure is available online at BIB online: https://academic.oup.com/bib.

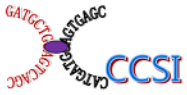
In the loop: promoter–enhancer interactions and bioinformatics

Antonio Mora, Geir Kjetil Sandve, Odd Stokke Gabrielsen and Ragnhild Eskeland

Kingston University London

3D Genome: Chromatin Interactions

FTO rs1421085



Chromatin Chromatin Space Interaction

Home | Method | Search | Download | Help | Update

Method: All Species: hg38 Cell Type: All

Search by Ensembl ID or gene name

Input Ensembl ID or gene name: FTO Example: ENSG00000230368 or FAM41C

Search by chromatin fragment

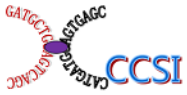
Input chromatin fragment: Example: chr1:1-1000

Submit

Kingston University London

3D Genome: Chromatin Interactions

FTO rs1421085



Chromatin Chromatin Space Interaction

Home | Method | Search | Download | Help | Update

Fragment1	Fragment2	Contact	FDR	P-Value	Method	Species	Cell Type	Enhancer	Antibody	Resolution	Geo
chr16:53431869-53437682	chr16:53701668-53718194	24	0.000500354	NA	CHi-C	hg38	GM12878	NA	NA	7.3kb	NA
chr16:53501368-53503674	chr16:53701668-53718194	24	0.001130303	NA	CHi-C	hg38	GM12878	NA	NA	7.3kb	NA
chr16:53701668-53718194	chr16:53683060-53692052	62	0.000414481	NA	CHi-C	hg38	GM12878	FANTOM5	NA	5.2kb	NA
chr16:53701668-53718194	chr16:53654234-53663350	41	0.000511939	NA	CHi-C	hg38	GM12878	FANTOM5	NA	5.2kb	NA
chr16:53701668-53718194	chr16:53678989-53682648	32	0.000515715	NA	CHi-C	hg38	GM12878	FANTOM5	NA	5.2kb	NA
chr16:53701668-53718194	chr16:53735718-53740760	43	0.000534632	NA	CHi-C	hg38	GM12878	FANTOM5	NA	5.2kb	NA
chr16:53701668-53718194	chr16:53671359-53675086	40	0.000544885	NA	CHi-C	hg38	GM12878	FANTOM5	NA	5.2kb	NA
chr16:53701668-53718194	chr16:53753814-53756390	34	0.000560596	NA	CHi-C	hg38	GM12878	FANTOM5	NA	5.2kb	NA
chr16:53701668-53718194	chr16:53725659-53735717	39	0.0006141	NA	CHi-C	hg38	GM12878	FANTOM5	NA	5.2kb	NA
chr16:53701668-53718194	chr16:53764083-53773653	25	0.000629421	NA	CHi-C	hg38	GM12878	FANTOM5	NA	5.2kb	NA

Showing 1 to 10 of 43 entries

Previous 1 2 3 4 5 Next

Kingston University London

3D Genome: Chromatin Interactions

SNCA rs356168

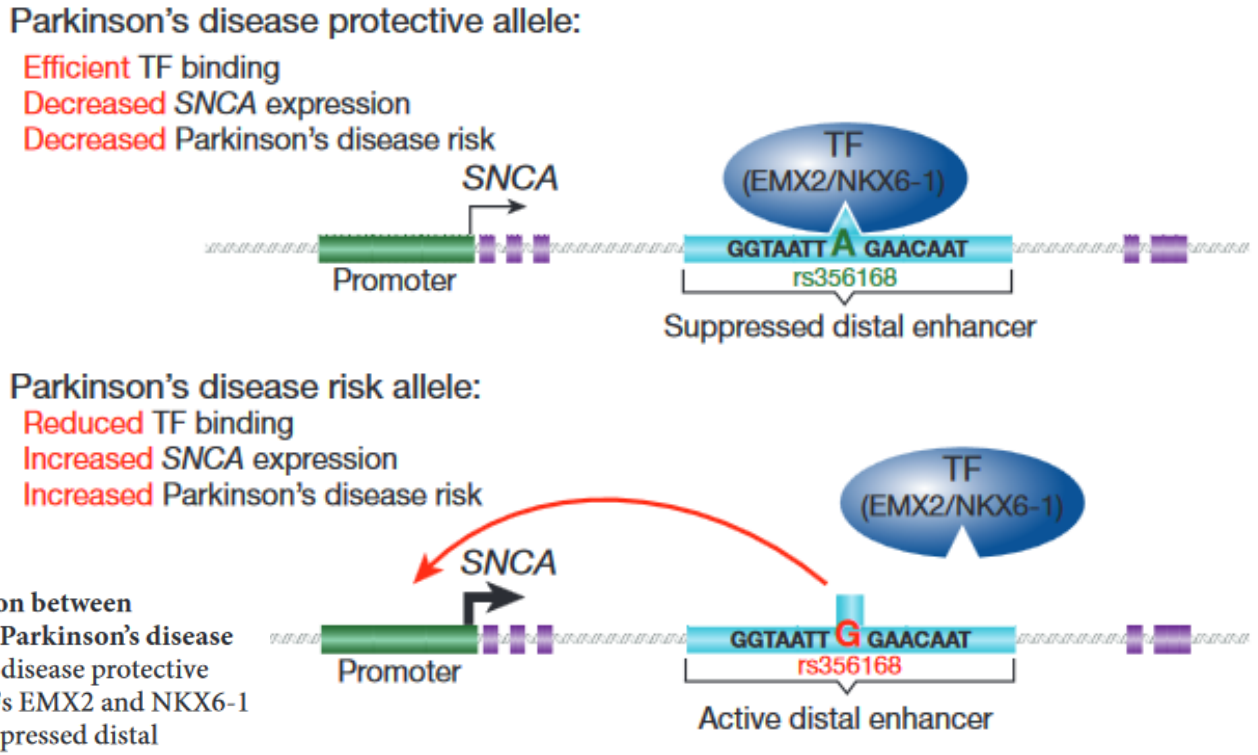


Figure 4 | Proposed model describing the correlation between SNP-dependent TF binding, *SNCA* expression and Parkinson's disease risk. Carriers of the A allele at rs356168 (Parkinson's disease protective allele) show efficient binding of the brain-specific TFs EMX2 and NKX6-1 at the distal intron-4 enhancer, which results in a suppressed distal enhancer and consequently lower expression of *SNCA* associated with a reduced risk to develop Parkinson's disease. In contrast, carriers of the G allele at rs356168 (Parkinson's disease risk allele) show reduced TF binding, which results in an active distal enhancer leading to increased expression of *SNCA* and increased risk of developing Parkinson's disease.

Kingston University London

Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression

Frank Soldner[1], Yonatan Stelzer[1], Chikdu S. Shivalila[1,2], Brian J. Abraham[1], Jeanne C. Latourelle[3], M. Inmaculada Barrasa[1], Johanna Goldmann[1], Richard H. Myers[3], Richard A. Young[1,2] & Rudolf Jaenisch[1,2]

3D Genome: Chromatin Interactions

8q24

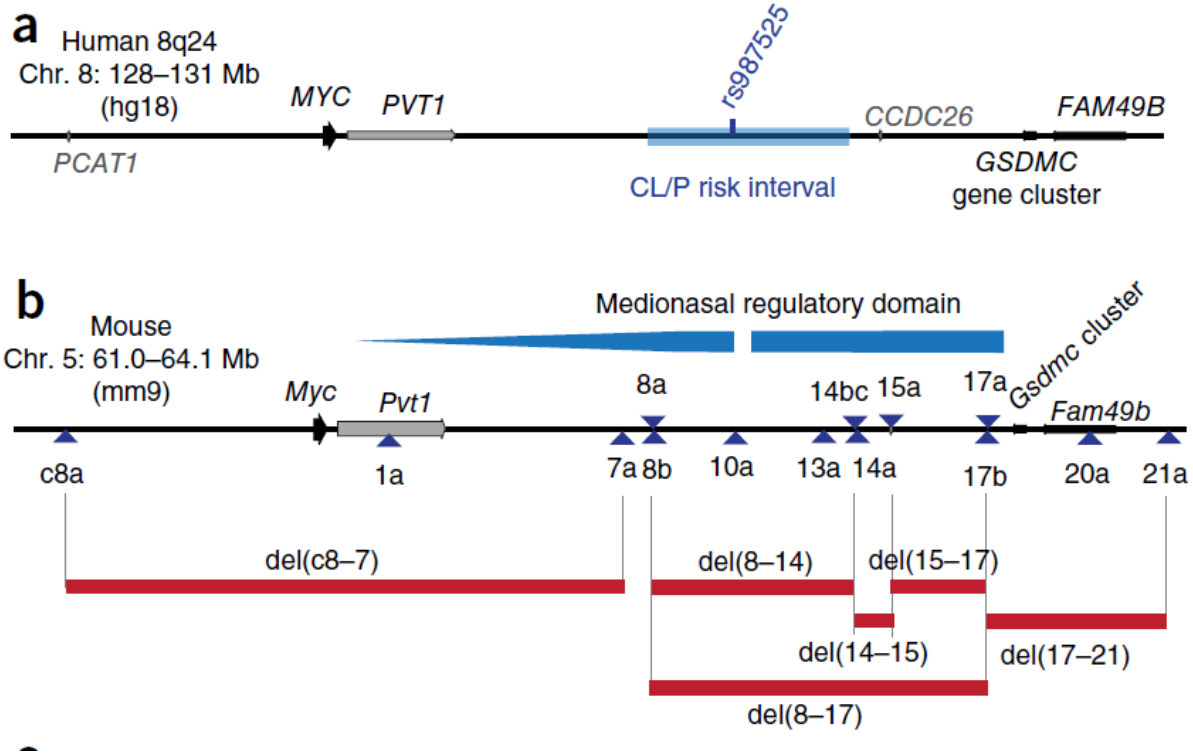


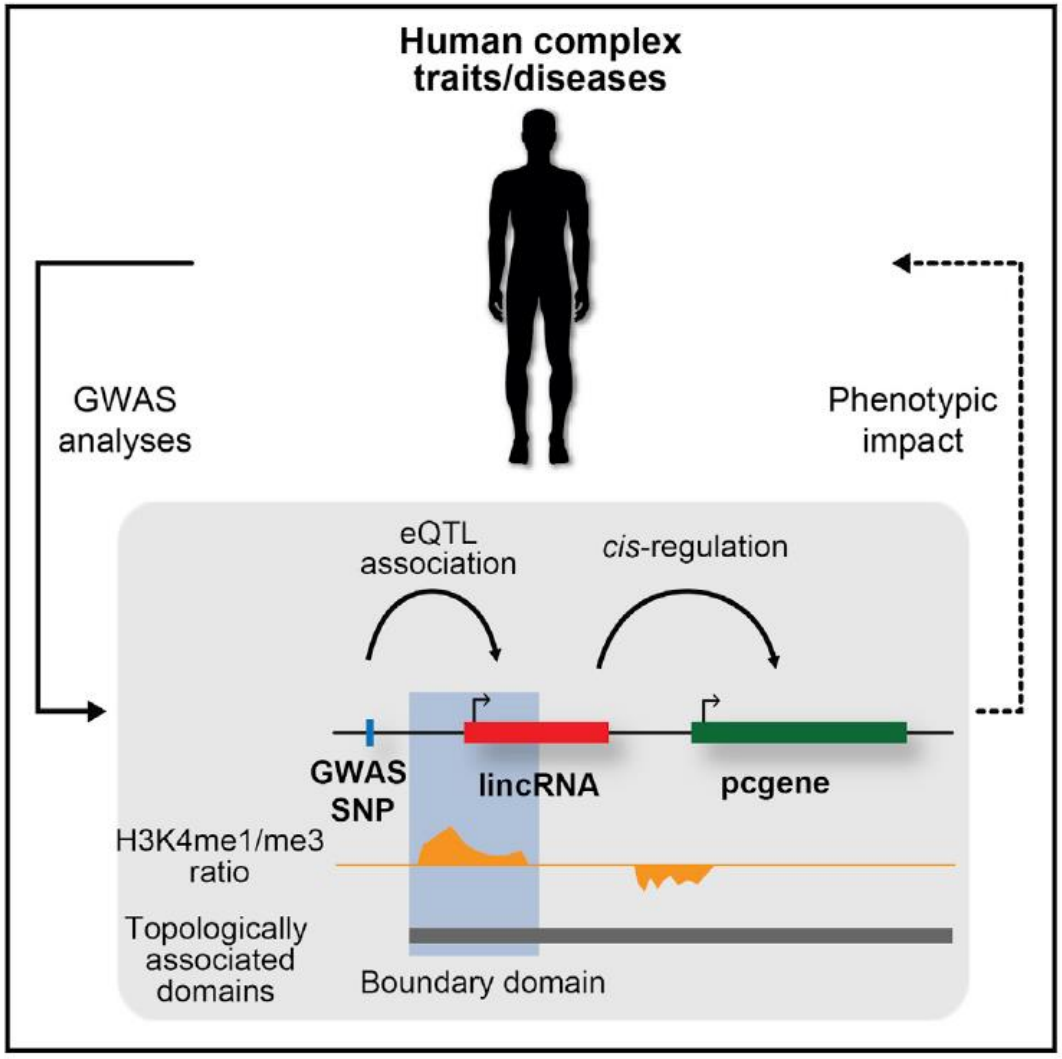
Figure 1 Functional characterization of the 8q24 CL/P regulatory landscape. (**a**) The human 8q24 interval associated with CL/P risk. Genes are shown as plain arrows (black, protein-coding genes; gray, annotated noncoding transcripts). The CL/P interval[3] and the most significantly associated SNP (rs987525) are indicated in blue. (**b**) Syntenic organization of the mouse locus, depicting transposon insertions (blue triangles) and deletions (red bars) used in this study. An expanded list of insertions and alleles is given in **Supplementary Table 1**. The expression patterns of adjacent insertions (shown in **c**) define a broad 'medionasal regulatory domain' indicated by a blue bar whose width represents relative LacZ expression levels.

Kingston University London

Long-range enhancers regulating *Myc* expression are required for normal facial morphogenesis

Veli Vural Uslu[1], Massimo Petretich[1], Sandra Ruf[1], Katja Langenfeld[1], Nuno A Fonseca[2], John C Marioni[2] François Spitz[1]

3D Genome: Chromatin Interactions



In Brief

Tan et al. identify and characterize 69 human complex trait/disease-associated lincRNAs in LCLs. They show that these loci are often associated with *cis*-regulation of gene expression and tend to be localized at TAD boundaries, suggesting that these lincRNAs may influence chromosomal architecture.

- We identify 69 lincRNAs associated with human complex traits (TR-lincRNAs)
- TR-lincRNAs are conserved in humans and interact with other disease-relevant loci
- TR-lincRNAs often associate with *cis*-regulation of proximal protein-coding gene expression
- TR-lincRNAs are enriched at TAD boundaries and may modulate chromatin architecture

Kingston University London

cis-Acting Complex-Trait-Associated lincRNA Expression Correlates with Modulation of Chromosomal Architecture

Jennifer Yihong Tan,[1,2,*] Adam Alexander Thil Smith,[1,2] Maria Ferreira da Silva,[1,2] Cyril Matthey-Doret,[1,2] Rico Rueedi,[2,3] Reyhan Sönmez,[2,3] David Ding,[4] Zoltán Kutalik,[3,5] Sven Bergmann,[2,3] and Ana Claudia Marques[1,2,6,*]

3D Genome: Chromatin Interactions

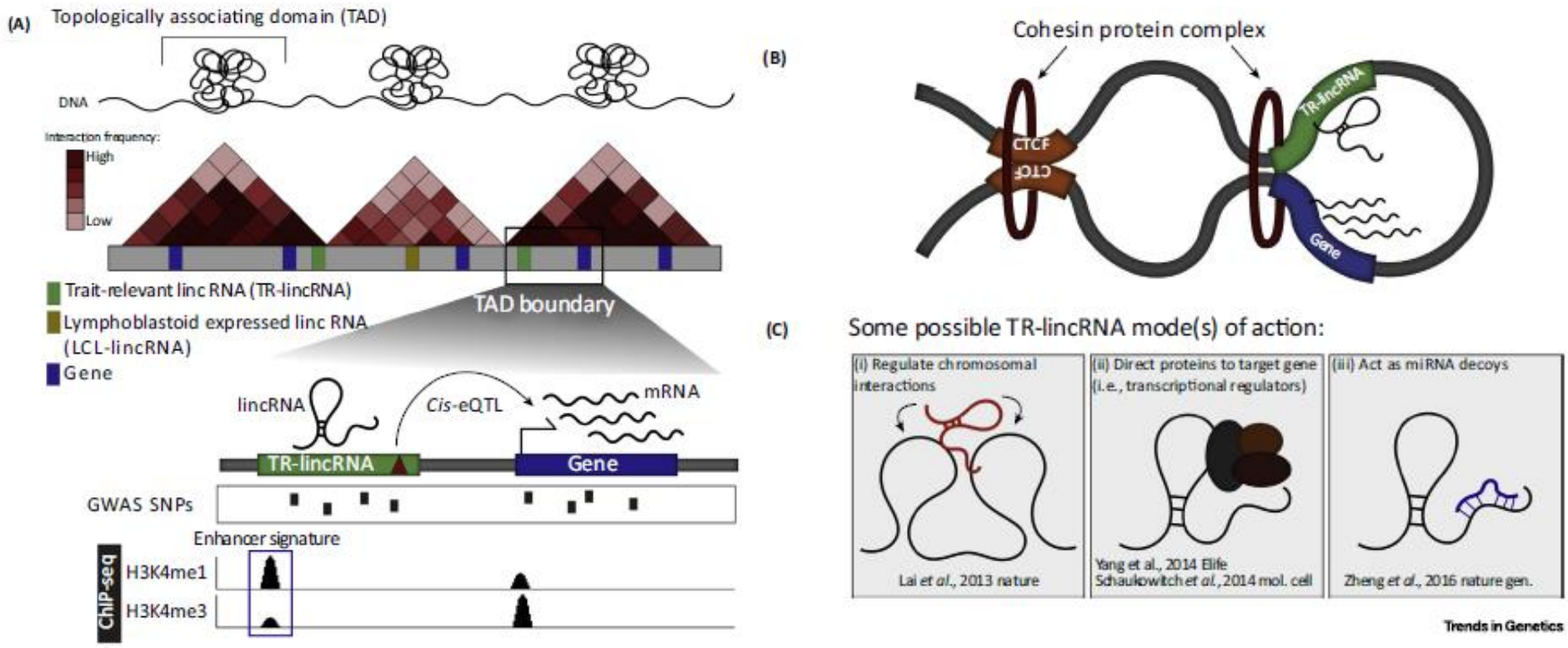


Figure 1. *Cis*-Acting TR-lincRNAs May Influence Target Gene Expression by Altering Intra-TAD Chromosomal Looping. (A) By using GWAS *cis*-eQTL analysis, Tan *et al.* [8] identified regulatory interactions between *cis*-acting TR-lncRNAs and target genes. TADs containing TR-lincRNAs display a higher density of chromosomal contacts. TR-lincRNAs are often located near the boundaries of TADs and arise from enhancer regions. (B) TR-lincRNAs were shown to be occupied by cohesin, but not by CTCF. (C) Possible TR-lincRNA mode (s) of action. TR-lincRNAs may (i) regulate chromosomal looping directly; (ii) use chromosomal looping to bring protein binding partners, such as transcriptional regulators or chromatin remodelers, in close proximity to target genes; or (iii) act as miRNA decoys. Abbreviations: ChIP-seq; chromatin immunoprecipitation with deep sequencing; CTCF, CCCTC-binding factor; eQTL, expression quantitative trait locus; GWAS, genome-wide association study; HSK4me1/3, histone H2 lysine 4 mono/trimethylation; lincRNA, long intergenic non-coding RNA

Spotlight

Enhancer-Derived lncRNAs Regulate Genome Architecture: Fact or Fiction?

Stephanie Fanucchi[1,2,*] and Musa M. Mhlanga[1,2,*]

How does the non-coding portion of the genome contribute to the regulation of genome architecture? A recent paper by Tan *et al.* focuses on the relationship between *cis*-acting complex-trait-associated lincRNAs and the formation of chromosomal contacts in topologically associating domains (TADs).

Kingston University London

3D Genome: Chromatin Interactions

An intronic polymorphism of IRF4 gene influences gene transcription *in vitro* and shows a risk association with childhood acute lymphoblastic leukemia in males

Thuy N. Do, Esma Ucisik-Akkaya, Charronne F. Davis, Brittany A. Morrison, M. Tevfik Dorak *

Genomic Immunoepidemiology Laboratory, HUMIGEN LLC, The Institute for Genetic Immunology, 2439 Kuser Road, Hamilton, NJ 08690-3303, USA

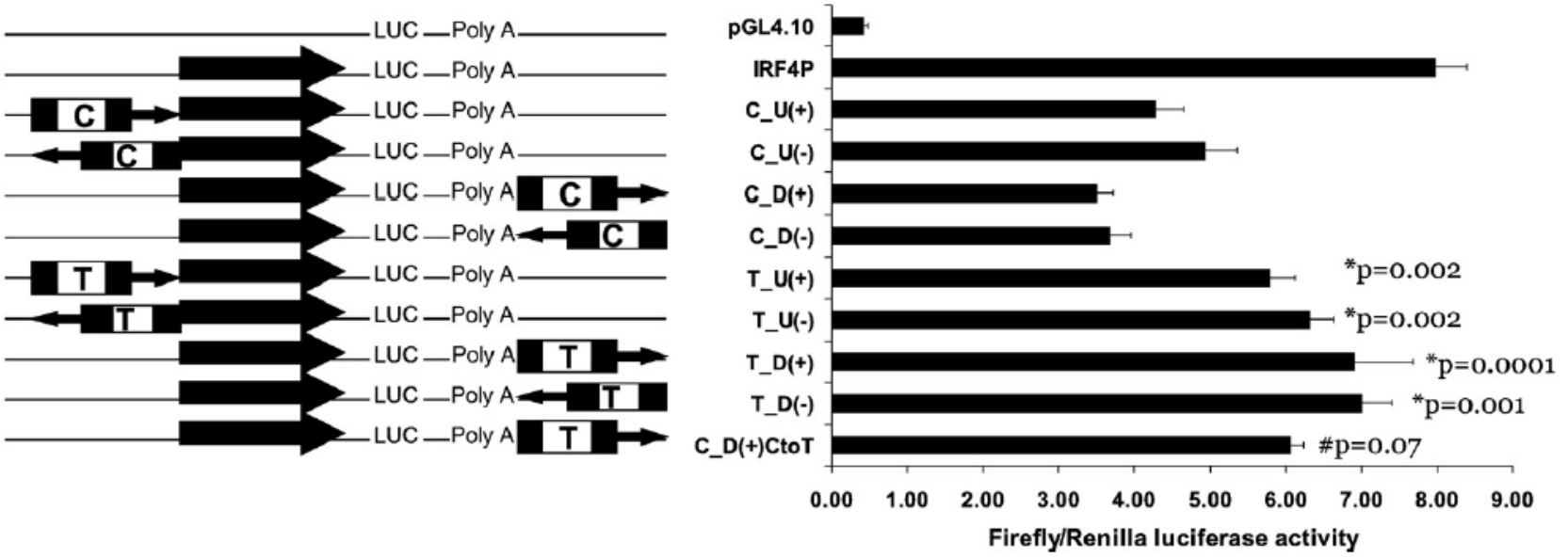


Fig. 1. IRF4 intron 4 with wild type allele C at SNP rs12203592 represses IRF4 promoter activity while IRF4 intron 4 with variant allele T significantly alleviates this repressive effect. Both work in an orientation- and position-independent manner. The full 1.2-kb fragment of intron 4 of the human IRF4 gene (contains either a wild type C or variant allele T at SNP rs12203592) was subcloned into the luciferase-reporter plasmid driven by a 2.4-kb IRF4 promoter (the big black arrow right before luciferase gene (LUC)). In all of the constructs, the LUC is used as a reporter gene whose mRNA is stabilized by a polyadenylation/splice signal from the simian virus 40 (Poly A). Raji cells were co-nucleofected with these constructs and with the internal control plasmid pGL4.13[hRenilla/SV40] and then assayed for both firefly and *Renilla* luciferases after 24 h. To adjust for differences in transfection efficiencies, firefly luciferase values were standardized to *Renilla* luciferase values. The results are from three independent experiments. The error bars represent standard errors. *Comparison between intron 4 with the variant allele T and intron 4 with the wild type allele C; #comparison between CD(+)CtoT with TD(+).

Kingston University London

3D Genome: Chromatin Interactions

OXFORD

Human Molecular Genetics, 2015, Vol. 24, No. 9 2649–2661

doi: 10.1093/hmg/ddv029
Advance Access Publication Date: 27 January 2015
Original Article

ORIGINAL ARTICLE

Allele-specific transcriptional regulation of *IRF4* in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the *IRF4* promoter

Mijke Visser, Robert-Jan Palstra† and Manfred Kayser*

Department of Forensic Molecular Biology, Erasmus MC University Medical Centre Rotterdam, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

The rs12203592 enhancer physically interacts with the *IRF4* promoter through an allele-dependent chromatin loop.

Kingston University London

3D Genome: Chromatin Interactions

The rs12203592 enhancer physically interacts with the *IRF4* promoter through an allele-dependent chromatin loop.

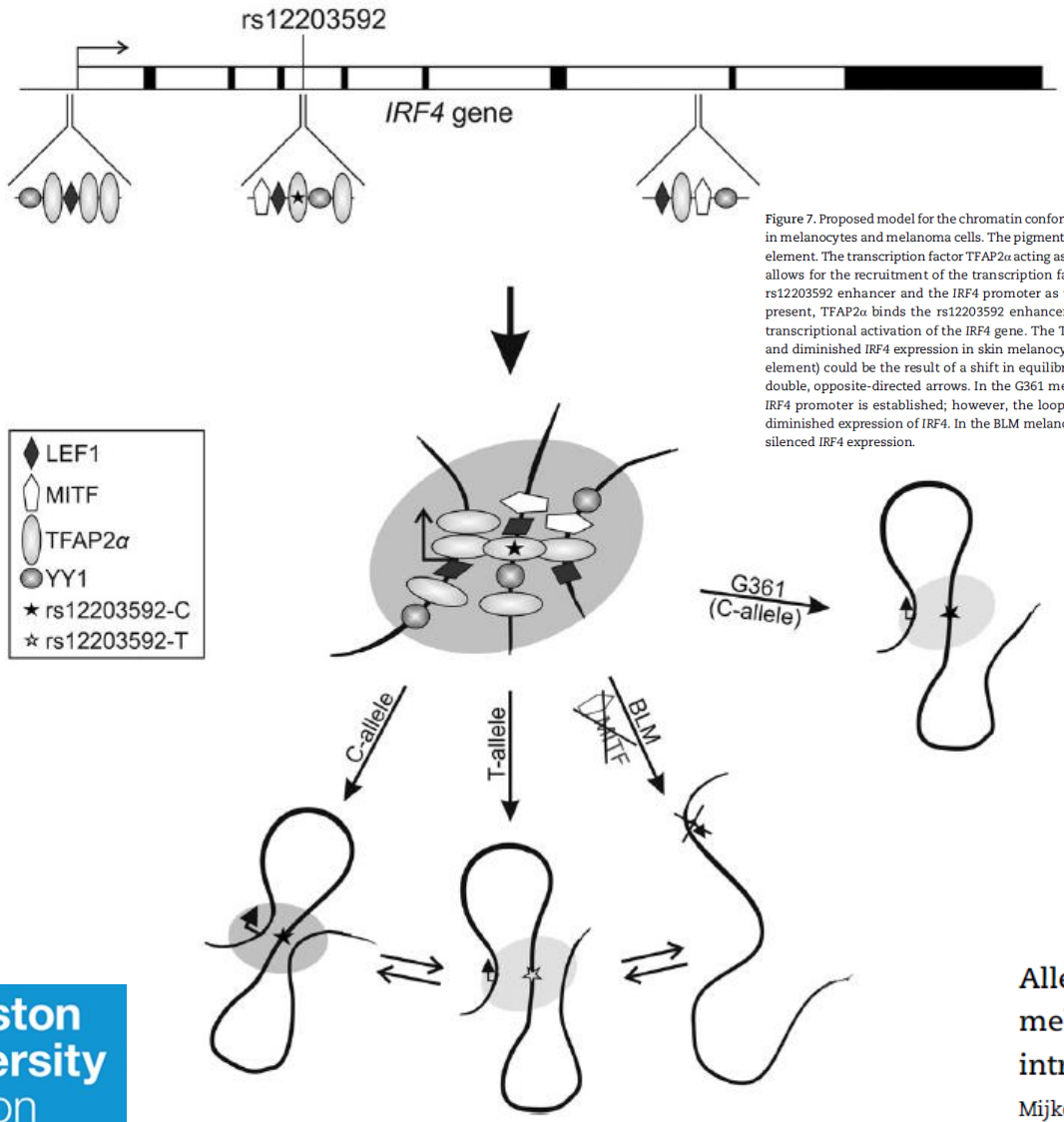


Figure 7. Proposed model for the chromatin conformation of the *IRF4* locus and subsequent transcriptional regulation of *IRF4*, depending on the allelic status of rs12203592 in melanocytes and melanoma cells. The pigmentation-associated SNP rs12203592 is located in intron 4 of *IRF4* and the region around this SNP functions as an enhancer element. The transcription factor TFAP2α acting as sequence specific DNA binding factor recognizes the rs12203592 enhancer in an allele-dependent manner, which then allows for the recruitment of the transcription factors MITF, YY1 and potential additional transcription factors like LEF1. Chromatin loops are formed between the rs12203592 enhancer and the *IRF4* promoter as well as the intron-7 YY1-element, both interactions depend on the allelic status of rs12203592. With the C-allele present, TFAP2α binds the rs12203592 enhancer, followed by recruitment of additional factors like MITF, YY1 and potentially LEF1, loop formation and proper transcriptional activation of the *IRF4* gene. The T-allele is unable to bind TFAP2α, which leads to reduced recruitment of additional factors, reduced loop formation and diminished *IRF4* expression in skin melanocytes. The different interactions between the rs12203592 enhancer and the *IRF4* promoter (as well as the intron-7 YY1-element) could be the result of a shift in equilibrium between the unfolded and interacting state that depends on the rs12203592 alleles, which is indicated by the double, opposite-directed arrows. In the G361 melanoma cell line, the TFAP2α factor presumably still binds the rs12203592 enhancer and loop formation toward the *IRF4* promoter is established; however, the loop toward the intron-7 YY1 element is disrupted, resulting in a less stable chromatin structure and consequently, diminished expression of *IRF4*. In the BLM melanoma cell line, MITF is absent and no chromatin loops are formed, resulting into a linear chromatin conformation and silenced *IRF4* expression.

Allele-specific transcriptional regulation of *IRF4* in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the *IRF4* promoter

Mijke Visser, Robert-Jan Palstra† and Manfred Kayser*

Kingston University London

Bioinformatics

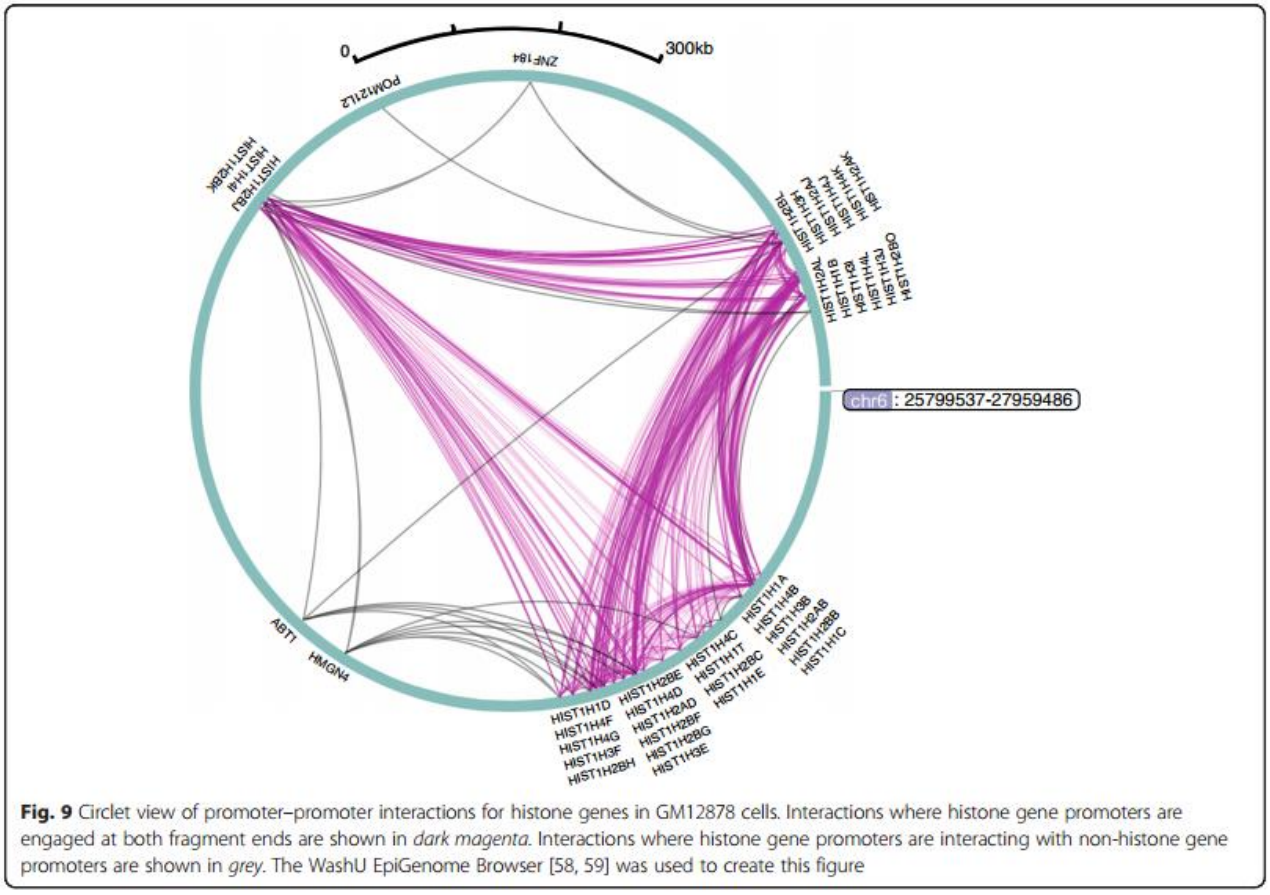


Fig. 9 Circlet view of promoter–promoter interactions for histone genes in GM12878 cells. Interactions where histone gene promoters are engaged at both fragment ends are shown in *dark magenta*. Interactions where histone gene promoters are interacting with non-histone gene promoters are shown in *grey*. The WashU EpiGenome Browser [58, 59] was used to create this figure

Cairns *et al. Genome Biology* (2016) 17:127
DOI 10.1186/s13059-016-0992-2

Genome Biology

METHOD Open Access

CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data

Jonathan Cairns[1†], Paula Freire-Pritchett[1†], Steven W. Wingett[1,2], Csilla Várnai[1], Andrew Dimond[1], Vincent Plagnol[3], Daniel Zerbino[4], Stefan Schoenfelder[1], Biola-Maria Javierre[1], Cameron Osborne[5], Peter Fraser[1] and Mikhail Spivakov[1*]

Kingston University London

Bioinformatics Tools

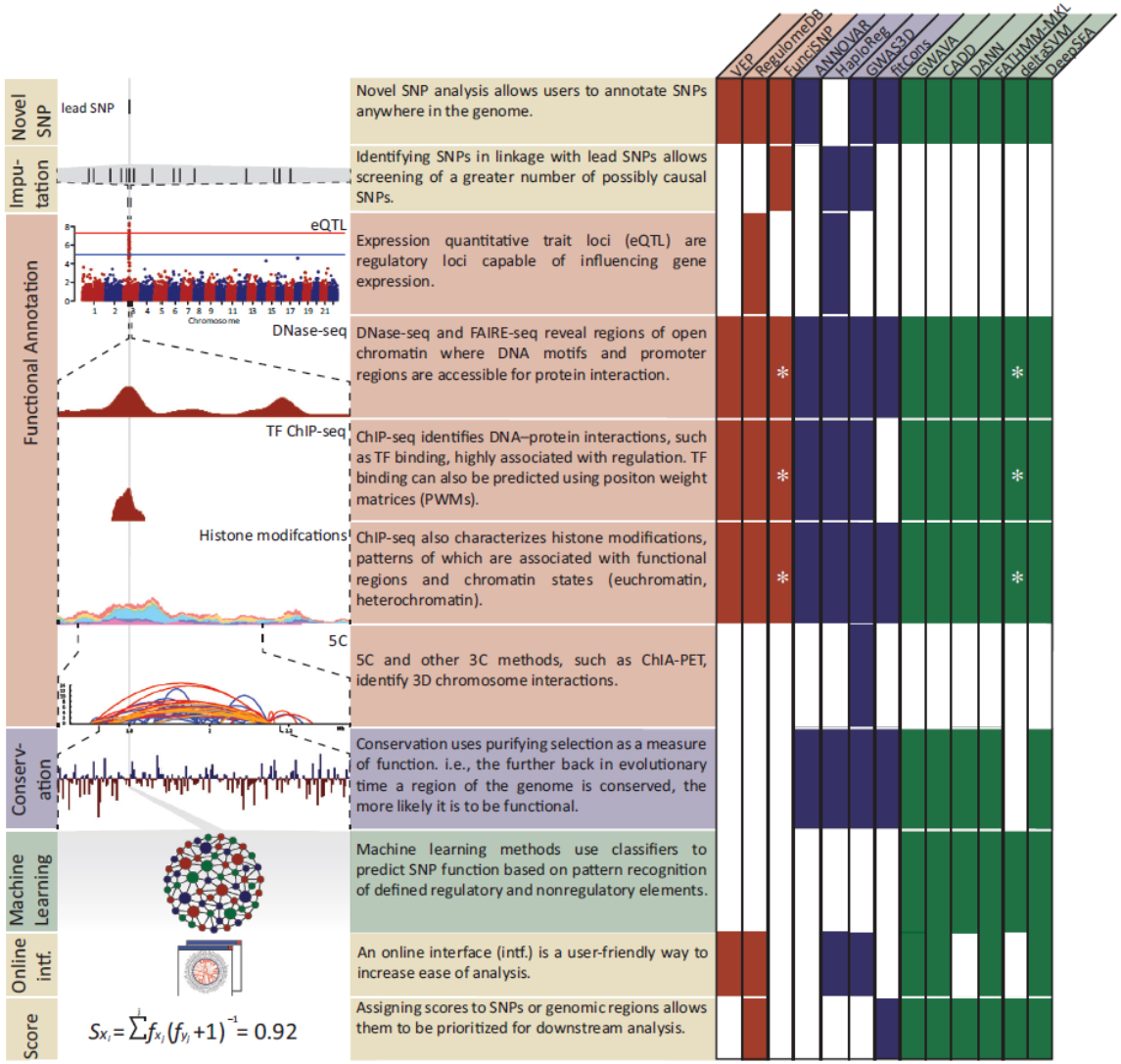


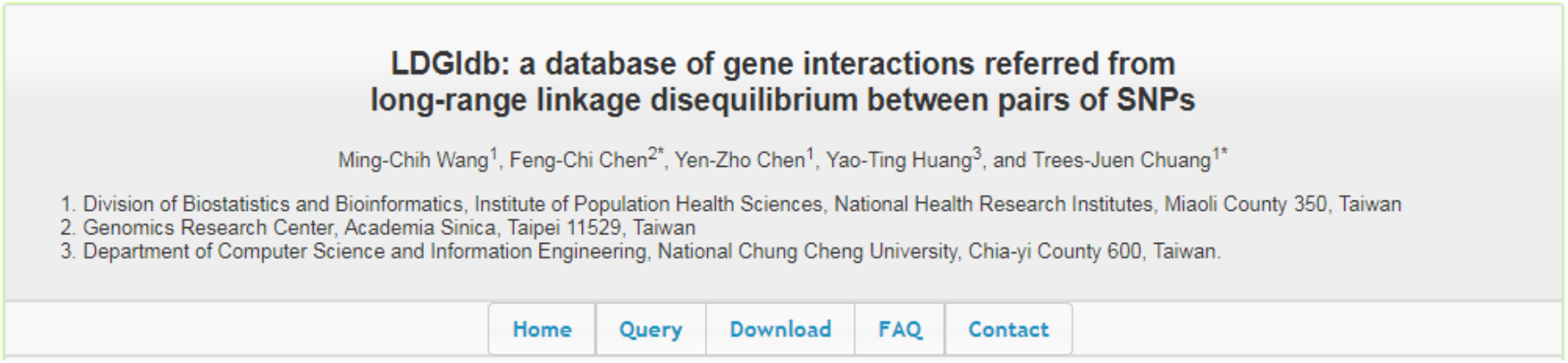
Figure 1. Data and Tools Used to Analyze Noncoding Variants. Single nucleotide polymorphism (SNP) aligned with functional (red) and conservation (blue) data, machine learning methods (green), and tool features (yellow). Each tool discussed in this perspective is labeled with annotation types used in its noncoding variant analysis platform. * represents optional input data sets supplied by the user. Abbreviations: 3C, chromosome conformation capture; 5C, chromosome conformation capture carbon copy; CADD, combined annotation-dependent depletion; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; DANN, deleterious annotation of genetic variants using neural networks; DNase-seq, DNase I hypersensitive sites sequencing; eQTL, expression quantitative trait loci; FAIRE, formaldehyde-assisted isolation of regulatory elements; FunciSNP, Functional Identification of SNPs; GWAVA, genome-wide annotation of variants; TF, transcription factor; VEP, Variant Effect Predictor.

Mining the Unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms

Sierra S. Nishizaki[1] and Alan P. Boyle[1,2,*]

Kingston University London

Bioinformatics Tools



LDGIdb: a database of gene interactions referred from long-range linkage disequilibrium between pairs of SNPs

Ming-Chih Wang[1], Feng-Chi Chen[2*], Yen-Zho Chen[1], Yao-Ting Huang[3], and Trees-Juen Chuang[1*]

1. Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Miaoli County 350, Taiwan
2. Genomics Research Center, Academia Sinica, Taipei 11529, Taiwan
3. Department of Computer Science and Information Engineering, National Chung Cheng University, Chia-yi County 600, Taiwan.

Home | Query | Download | FAQ | Contact

LDGIdb: a database of gene interactions inferred from long-range strong linkage disequilibrium between pairs of SNPs

Ming-Chih Wang[1], Feng-Chi Chen[2,3,4*], Yen-Zho Chen[1], Yao-Ting Huang[5] and Trees-Juen Chuang[1*]

Kingston University London

Bioinformatics Tools

DATABASE

Database, 2017, 1–17
doi: 10.1093/database/bax028
Original article

Original article

GeneHancer: genome-wide integration of enhancers and target genes in GeneCards

Simon Fishilevich[1,†], Ron Nudel[1,†], Noa Rappaport[1], Rotem Hadar[1], Inbar Plaschkes[1], Tsippi Iny Stein[1], Naomi Rosen[1], Asher Kohn[2], Michal Twik[1], Marilyn Safran[1], Doron Lancet[1,*] and Dana Cohen[1,*]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel and [2]LifeMap Sciences Inc, Marshfield, MA 02050, USA

Genomics for IRF4 Gene

Products: Regulatory Element

Regulatory Elements for IRF4 Gene

Enhancers for IRF4 Gene IMPROVED!

Filter: (40 results) See all 40 »

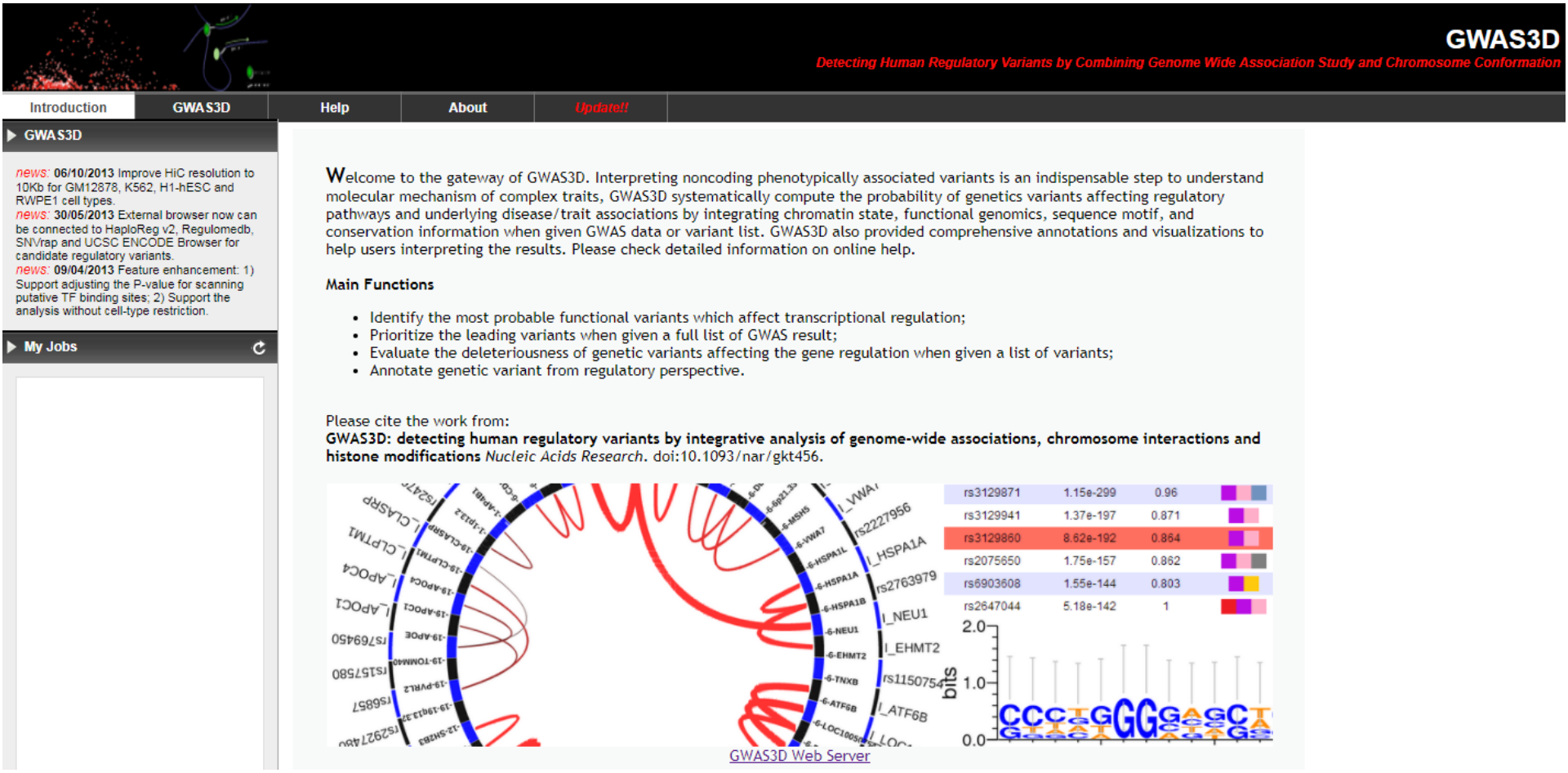
	GeneHancer Identifier	Enhancer Score	Enhancer Sources	Gene-Enhancer Score	Total Score	TSS distance (kb)	Number of Genes Away	Size (kb)	Transcription Factor Binding Sites within enhancer	Gene Targets for Enhancer
⊕	GH06G000390	1.6★	FANTOM5, ENCODE, dbSUPER	26.1★	41.76	+0.9	0	3.6	65 TFs HDGF PKNOX ⊕	5 genes IRF4 ⊕
⊕	GH06G000219	1.8★	FANTOM5, Ensembl, ENCODE, dbSUPER	12.8★	23.04	-170.8	19	3.0	93 TFs HNRNPUL1 … ⊕	5 genes IRF4 ⊕
⊕	GH06G000225	1.7★	FANTOM5, Ensembl, ENCODE, dbSUPER	13.5★	22.95	-166.1	19	2.6	39 TFs HDGF TBP ⊕	6 genes IRF4 ⊕
⊕	GH06G000316	1.5★	FANTOM5, ENCODE, dbSUPER	14.3★	21.45	-74.3	12	2.2	54 TFs HDGF TBP ⊕	5 genes IRF4 ⊕
⊕	GH06G000290	2★	FANTOM5, Ensembl, ENCODE, dbSUPER	9.5	19	-96.7	13	10.0	206 TFs HNRNPUL1 … ⊕	5 genes EXOC2 ⊕

★ - Elite enhancer and/or Elite enhancer-gene association

Download GeneHancer data dump

Kingston University London

Bioinformatics Tools



Kingston University London

GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications

Mulin Jun Li[1,2], Lily Yan Wang[1,2], Zhengyuan Xia[2,3], Pak Chung Sham[2,4,5,6] and Junwen Wang[1,2,4,*]

Bioinformatics Tools

Chromatin Chromatin Space Interaction

Home | Method | Search | Download | Help | Update

Welcome to CCSI database

Here, **CCSI (Chromatin Chromatin Space Interaction)** database presents 3,017,962 chromatin interaction pairs with annotation of genes, enhancers and SNPs in many cell lines of human, mouse and yeast. These data were obtained by means of **3C, 4C, 5C, ChIA-PET and Hi-C** technology in a cell's natural state, nearly all of which detected the three-dimensional architecture of chromosome by coupling ligation in close spatial proximity followed by high-throughput sequencing. So transcriptional regulatory mechanism in disease pathogenesis associated with spatial interactions among genes, enhancers and SNPs could be explored on the base of it.

Kingston University London

DATABASE

Database, 2016, 1–7
doi: 10.1093/database/bav124
Original article

Original article

CCSI: a database providing chromatin–chromatin spatial interaction information

Xiaowei Xie[1], Wenbin Ma[1], Zhou Songyang[1], Zhenhua Luo[1,2], Junfeng Huang[1], Zhiming Dai[3,*] and Yuanyan Xiong[1,4,*]

Bioinformatics Tools

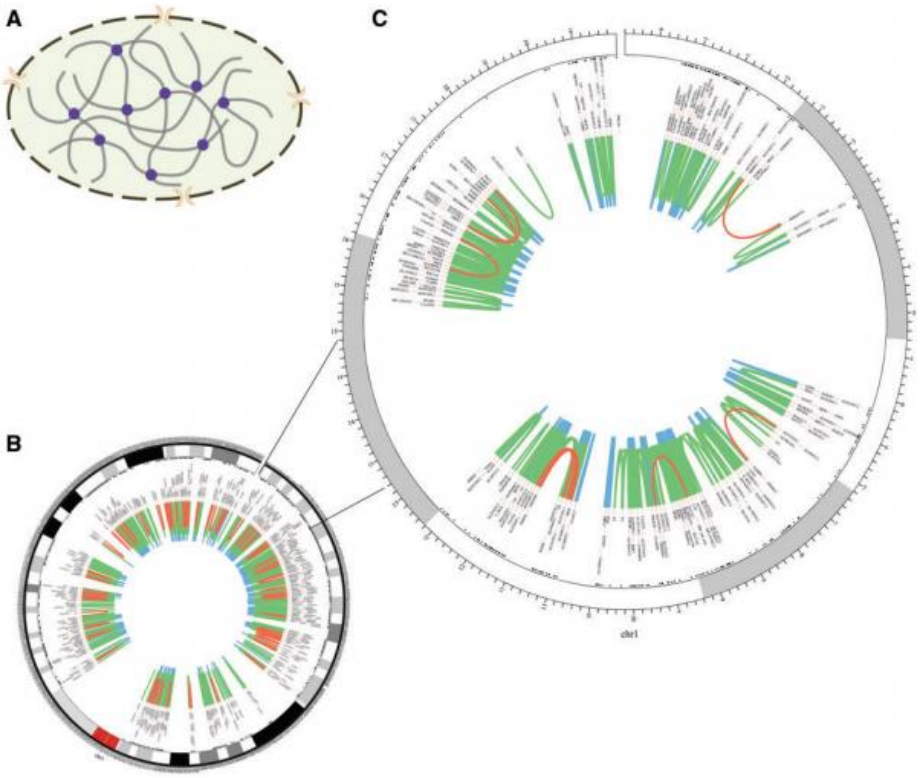


Figure 2. Chromosomal 3D structure and promoter–promoter interactions of Chr1 in IMR90 cell line based on a set of Hi-C data. (**A**) Chromosomal 3D structure. The dashed circle with two orange crescents that stand for nuclear pore complex is the nucleus membrane. The thick grey lines are chromatin and the purple circles stand for proteins that link chromatin together. (**B**) Promoter–promoter interactions of Chr1. (**C**) Promoter–promoter interactions of Chr1:1-20000000, zooming into the interactions. The red lines stand for long-range interactions (distance between interaction pair > 500 kb), while the blue lines for short-range (distance < 50 kb) and the green lines for middle-range (distance spanning 50–500 kb). The black texts are the gene names of corresponding loci.

DATABASE

Database, 2016, 1–7
doi: 10.1093/database/bav124
Original article

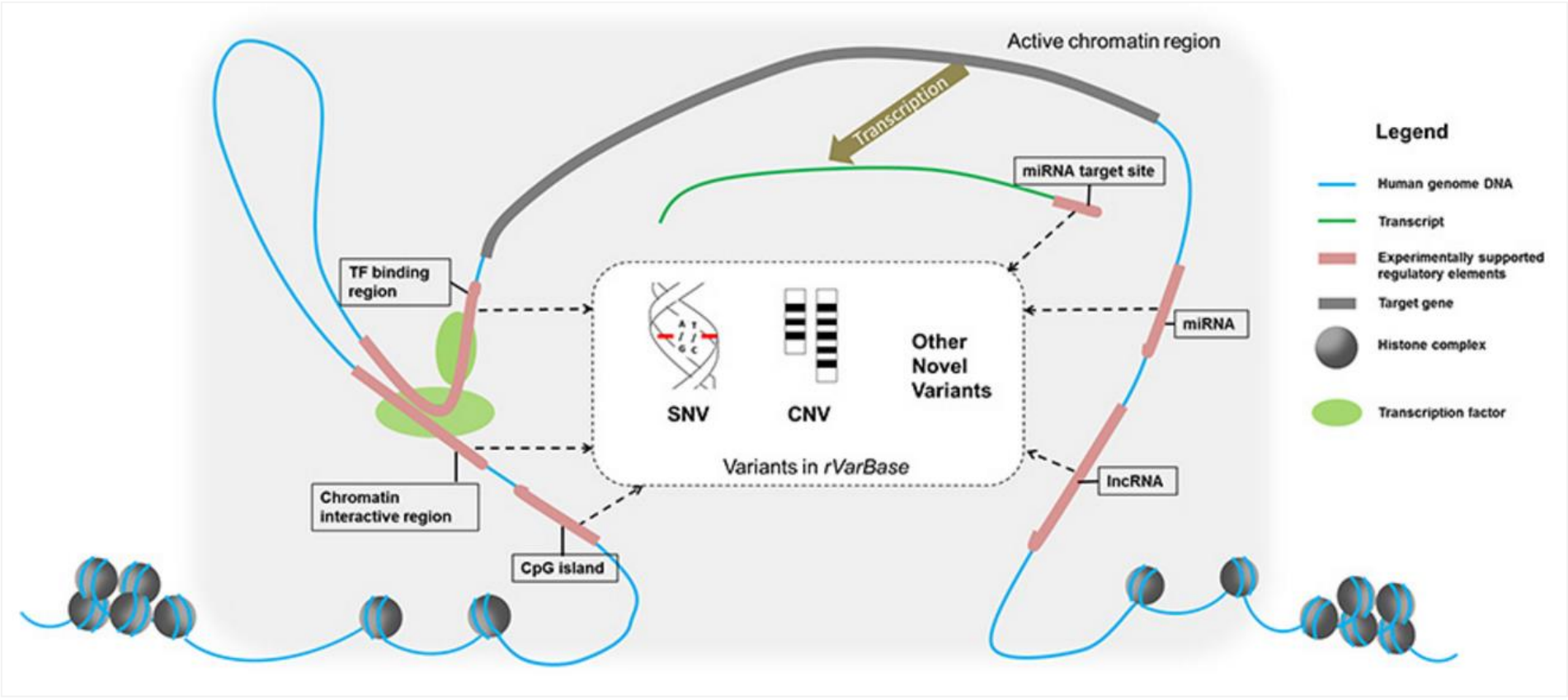
Original article

CCSI: a database providing chromatin–chromatin spatial interaction information

Xiaowei Xie[1], Wenbin Ma[1], Zhou Songyang[1], Zhenhua Luo[1,2], Junfeng Huang[1], Zhiming Dai[3,*] and Yuanyan Xiong[1,4,*]

Kingston University London

Bioinformatics Tools

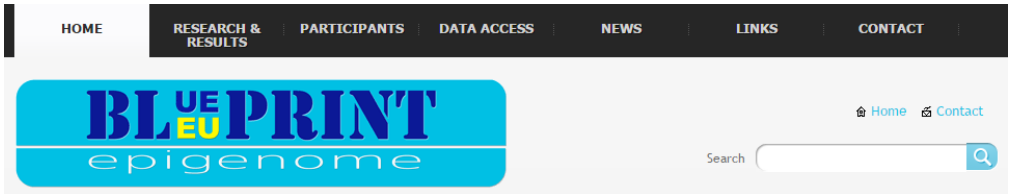


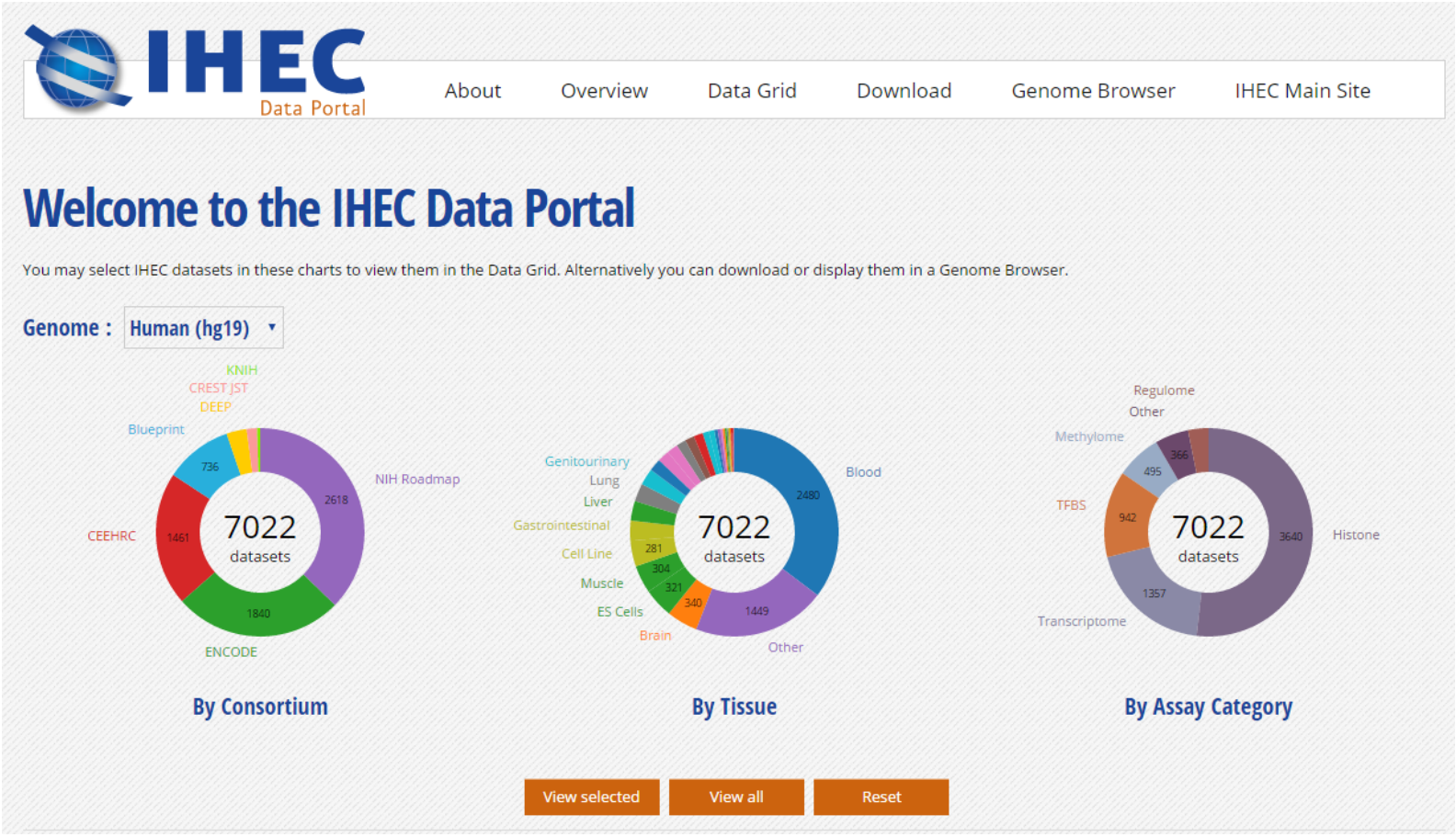
Kingston University London

rVarBase:an updated database for regulatory features of human variants

Chromatin states, regulatory elements and target genes

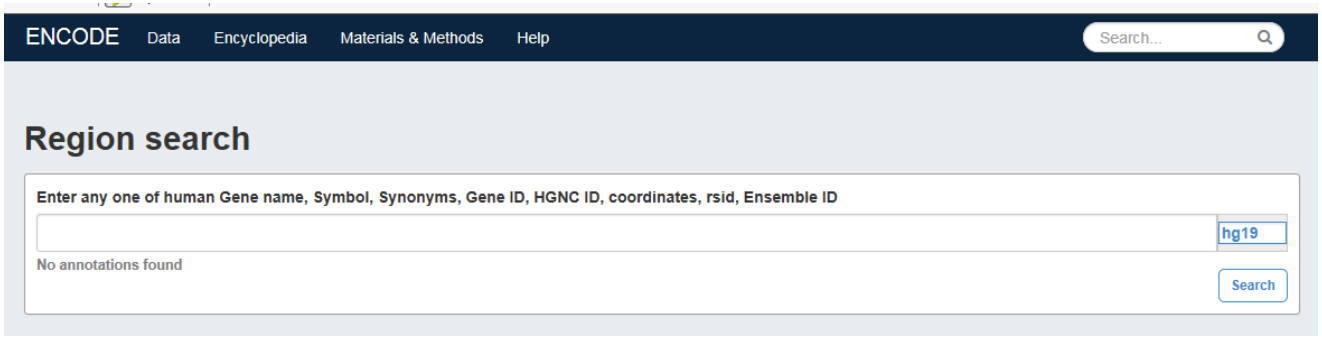
Bioinformatics Tools

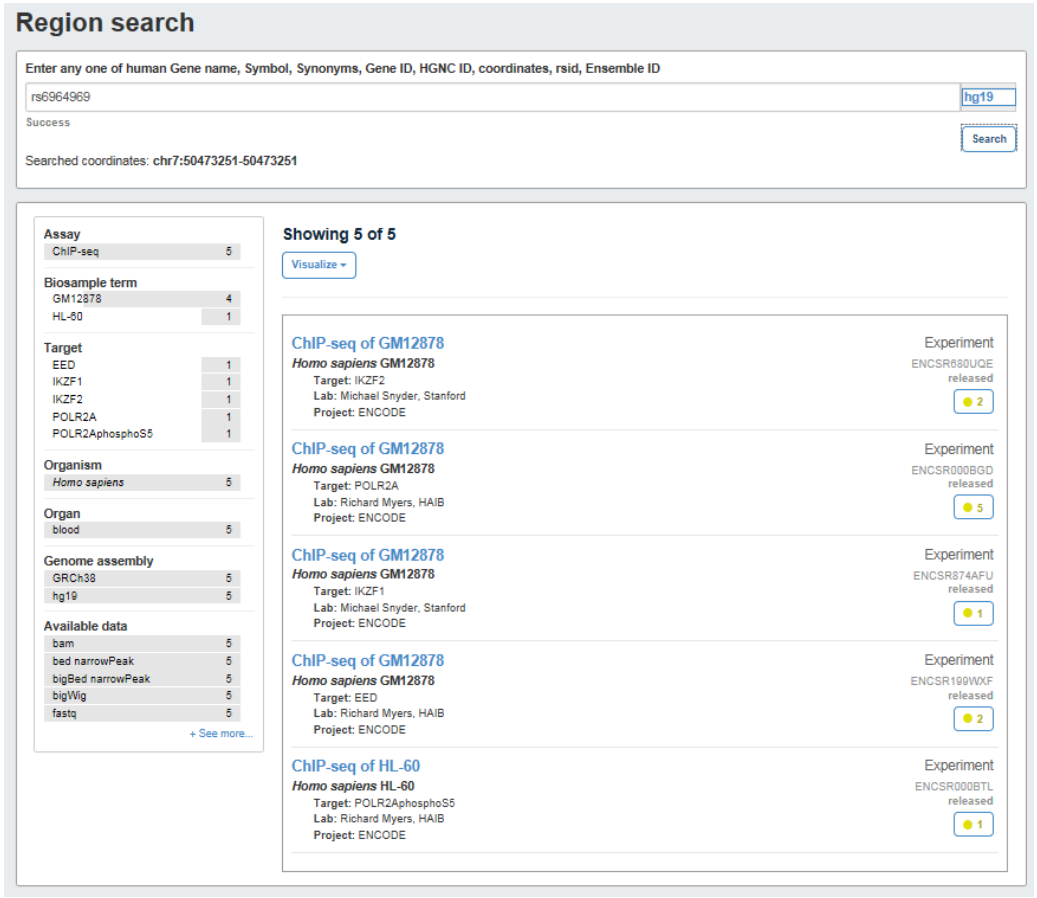




Kingston University London

Bioinformatics Tools





Kingston University London

Bioinformatics Tools

HaploReg v4

HaploReg is a tool for exploring annotations of the noncoding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci. Using LD information from the 1000 Genomes Project, linked SNPs and small indels can be visualized along with chromatin state and protein binding annotation from the Roadmap Epigenomics and ENCODE projects, sequence conservation across mammals, the effect of SNPs on regulatory motifs, and the effect of SNPs on expression from eQTL studies. HaploReg is designed for researchers developing mechanistic hypotheses of the impact of non-coding variants on clinical phenotypes and normal variation.

Update 2015.09.15: Version 4 now includes many recent eQTL results including the GTEx pilot, four different options for defining enhancers using Roadmap Epigenomics data, and a complete set of source files for download and local analysis. Older versions available: v3, v2, v1.

Build Query | Set Options | Documentation

Use one of the three methods below to enter a set of variants. If an r^2 threshold is specified (see the Set Options tab), results for each variant will be shown in a separate table along with other variants in LD. If r^2 is set to NA, only queried variants will be shown, together in one table.

Query (comma-delimited list of rsIDs OR a single region as chrN:start-end): rs6964969

or, upload a text file (one refSNP ID per line): Browse...

or, select a GWAS:

Submit Query

Query SNP: rs6964969 and variants with r^2 >= 0.8

chr	pos (hg38)	LD (r^2)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	eQTL results	Motifs changed	GENCODE genes	dbSNP func annot
7	50398132	0.96	0.99	rs62447205	A	G	0.23	0.22	0.13	0.31			BLD	4 tissues			4 altered motifs	IKZF1	intronic
7	50398606	0.96	0.99	rs11978267	A	G	0.22	0.22	0.13	0.31		LNG	BLD			7 eQTL results	Foxp1,Pax-6	IKZF1	intronic
7	50398997	0.96	0.99	rs6973210	G	A	0.23	0.22	0.13	0.31			BLD	BLD,PANC		7 eQTL results	4 altered motifs	IKZF1	intronic
7	50399099	0.97	0.99	rs6960400	A	G	0.23	0.22	0.13	0.31			BLD	BLD,BLD			ATF3,Maf,NF-E2	IKZF1	intronic
7	50401254	0.98	0.99	rs10278451	G	T	0.23	0.22	0.13	0.31						7 eQTL results	Sox	IKZF1	3'-UTR
7	50401853	0.98	0.99	rs11552047	C	T	0.22	0.22	0.12	0.31			BLD				Nr2e3,Zec	IKZF1	3'-UTR
7	50402283	0.98	0.99	rs11980379	T	C	0.23	0.22	0.12	0.31			BLD	THYM		8 eQTL results	NF-E2	IKZF1	3'-UTR
7	50402678	0.98	0.99	rs200338223	C	CT	0.23	0.22	0.12	0.31			BLD				12 altered motifs	IKZF1	3'-UTR
7	50402680	0.96	0.99	rs33999320	T	TC	0.23	0.22	0.12	0.31			BLD				10 altered motifs	IKZF1	3'-UTR
7	50402906	0.98	0.99	rs4132601	T	G	0.23	0.22	0.12	0.31			BLD			8 eQTL results		IKZF1	3'-UTR
7	50403915	0.99	1	rs11980407	G	A	0.22	0.22	0.12	0.30			BLD			7 eQTL results	Pou5f1	IKZF1	3'-UTR
7	50404626	1	1	rs62445866	G	A	0.22	0.22	0.12	0.30			BLD	12 tissues	CTCF,RAD21		NRSF	IKZF1	3'-UTR
7	50405144	1	1	rs58923657	C	T	0.22	0.22	0.12	0.30		BLD	BLD, THYM, SPLN	BLD,BLD			CHOP::CEBPalpha	42bp 3' of IKZF1	
7	50405553	1	1	rs6964969	A	G	0.23	0.22	0.12	0.30		BLD	4 tissues	BLD,BLD,BLD	NFKB	7 eQTL results	Pdx1,ZBTB33	451bp 3' of IKZF1	
7	50405592	0.97	1	rs6956014	T	C	0.22	0.22	0.12	0.30		BLD	4 tissues	BLD	NFKB		Myc,RFX5,SREBP	490bp 3' of IKZF1	
7	50406172	1	1	rs28462675	A	G	0.22	0.22	0.12	0.30			4 tissues	BLD,BLD				1.1kb 3' of IKZF1	
7	50407229	0.98	1	rs150935798	CA	C	0.11	0.21	0.12	0.30			4 tissues	PLCNT,THYM			BDP1,SREBP	2.1kb 3' of IKZF1	
7	50407232	0.98	1	rs200342481	GC	G	0.11	0.21	0.12	0.30			4 tissues	PLCNT,THYM			HDAC2,NF-I	2.1kb 3' of IKZF1	
7	50407623	1	1	rs10264390	T	C	0.22	0.22	0.12	0.30				BLD,THYM,BLD		7 eQTL results		2.5kb 3' of IKZF1	
7	50408133	0.99	1	rs28696237	C	G	0.18	0.22	0.12	0.30		SPLN	BLD, THYM					3kb 3' of IKZF1	
7	50409446	0.97	1	rs10230978	G	A	0.18	0.22	0.13	0.30							4 altered motifs	4.3kb 3' of IKZF1	
7	50409515	0.98	1	rs10272724	T	C	0.18	0.22	0.13	0.30				THYM,PANC		7 eQTL results	9 altered motifs	4.4kb 3' of IKZF1	
7	50409816	0.97	0.99	rs17133805	T	G	0.18	0.22	0.13	0.30		BLD	6 tissues	BLD,BLD,BLD			HNF4	4.7kb 3' of IKZF1	
7	50409913	0.98	0.99	rs62445869	G	A	0.12	0.21	0.13	0.30		BLD	6 tissues	5 tissues	TCF12		6 altered motifs	4.8kb 3' of IKZF1	
7	50409989	0.97	0.99	rs17133807	G	A	0.19	0.22	0.13	0.30		BLD	6 tissues	6 tissues	4 bound proteins	7 eQTL results	CTCF,HEN1	4.9kb 3' of IKZF1	
7	50410929	0.89	0.97	rs1110701	A	G	0.27	0.25	0.13	0.32			5 tissues	4 tissues		7 eQTL results	4 altered motifs	5.8kb 3' of IKZF1	

Kingston University London

Conclusions

GWAS has fulfilled the aim of unraveling disease biology more than providing markers for disease prediction

Unexpected results in gene deserts and other intergenic regions shed light on the function of these non-coding regions

What used to be called "junk DNA" codes for non-coding RNA and involved in chromatin interactions

Non-coding genetic variants modify the function of regulatory elements as crucially as coding variants alter protein structure

Projects like ENCODE, NIH Epigenomics Road Map and BluePrint provide sufficient data to examine effects of genetic variation

Conclusions

The data indicate that the target gene of a genetic variation may be far away from it (even on a different chromosome); the nearest gene is unlikely to be the target gene

The target genes are usually tissue-specific

The candidate genes involved in a disease process may not have a variation of itself implicated in the disease development

5th International Congress of the Molecular Biology Association of Turkey

Many thanks to the MolBiyoKon17 Organisation Committee, in particular to

Prof Alaattin Şen, Dr Umut Şahin, Dr İbrahim Yaman

Moleküler Biyoloji Derneği

as well as

the Sponsors

Moleküler Biyoloji Derneği

EMBO excellence in life sciences

United Nations Educational, Scientific and Cultural Organization · Turkish National Commission for UNESCO

ANT TEKNİK

COMPECTA The HPC Company

jove Journal of Visualized Experiments

aka Biyoteknoloji

MedSanTek

BO&GA

ARDI medikal

prizma

Nescafé Dolce Gusto

GEN Plaza Biyoteknoloji Merkezi San. ve Tic. Ltd. Şti.

FEBS Journal

PHI Tech Bioinformatics

Gen-Era

CLEANROOM

SACEM Hayat Teknolojileri A.Ş

bio medya

sentromer