

Wellcome Trust Genome Campus Advanced Course
Genetic Analysis of Population-based Association Studies

Post-GWAS Analysis

Part 2: Functional genomics

Gisela Orozco

Professor of Functional Genomics
University of Manchester, UK

15th September 2023

Session overview

1. Lecture: Functional genomics
 - Limitations of GWAS
 - The use of functional genomics to understand how GWAS SNPs cause disease and overcome some of these limitations
 - Types of data
 - DNase-hypersensitivity
 - ATAC-Seq
 - ChIP-Seq
 - eQTLs
 - Hi-C
 - Large scale functional genomics studies
2. Break
3. Demonstration
 - How can I use publicly available functional genomics data to help me interpret my GWAS results?
 - Exploring online resources

Potential of GWAS for clinical translation

- **Precision medicine**

- Targeting available therapies to groups of patients most likely to respond
- Avoiding therapies in groups of patients likely to develop adverse events.

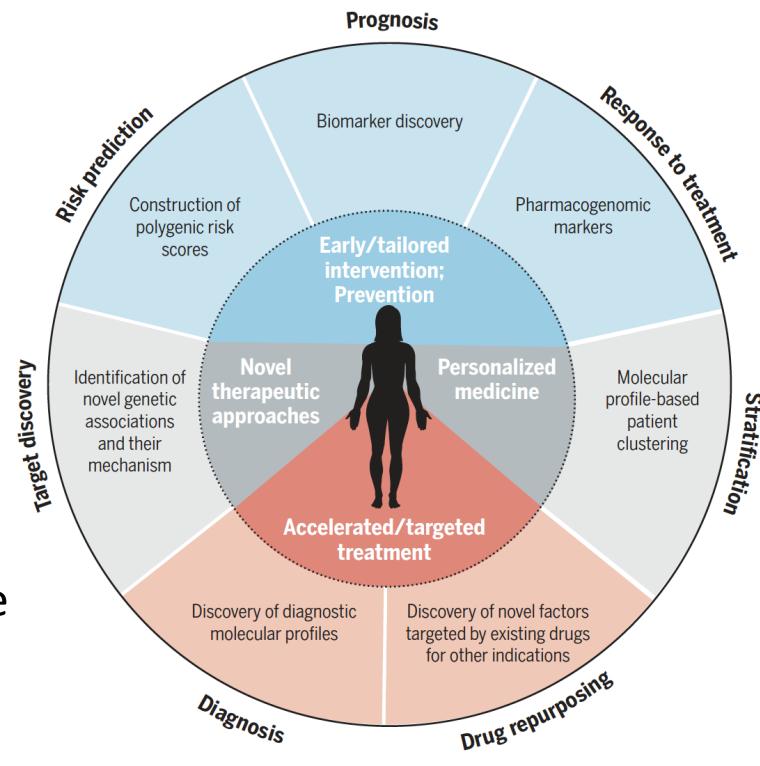
Associations of HLA-C*06:02 with biologic therapy response in psoriasis (Dand, 2018) and *HLA-DRB1* with severity, mortality, and treatment response to biologic drugs in RA (Viatte, 2015).

- **Discovery of novel drug targets**

- Selecting drug targets supported by genetic evidence can double the chance of success in clinical development (Nelson, 2015)
- GWAS signals implicate genes that encode known drug targets, which provide proof of concept of the ability of GWAS to identify potentially new druggable targets.
- GWAS can identify drug-repurposing opportunities, i.e., targets for which there are already approved drugs, for other indications (e.g. IL-23 pathway)

- **Prediction, prevention, and prognosis**

- Genetic risk scores (GRS) can help identify those at highest risk of developing disease, rapid progression or severe manifestations



Zeggini et al.
Science 365, 1409–1413 (2019)

GWAS catalog

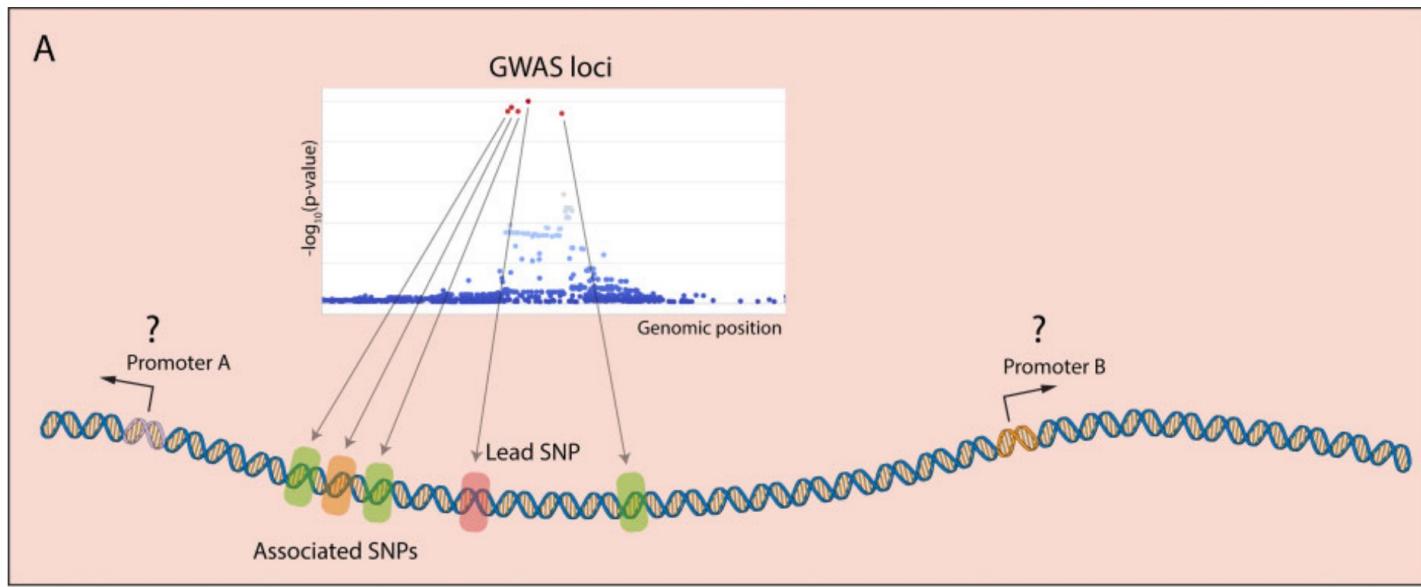
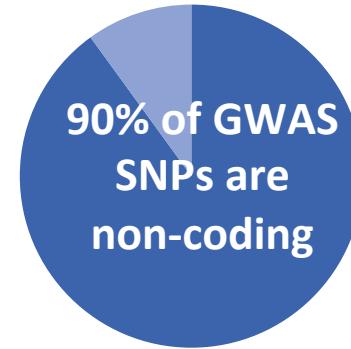
<https://www.ebi.ac.uk/gwas/home>

Accessed 08/09/2023



6,545 publications and 541,660 associations

GWAS have not reached yet their full potential for clinical translation and drug target identification



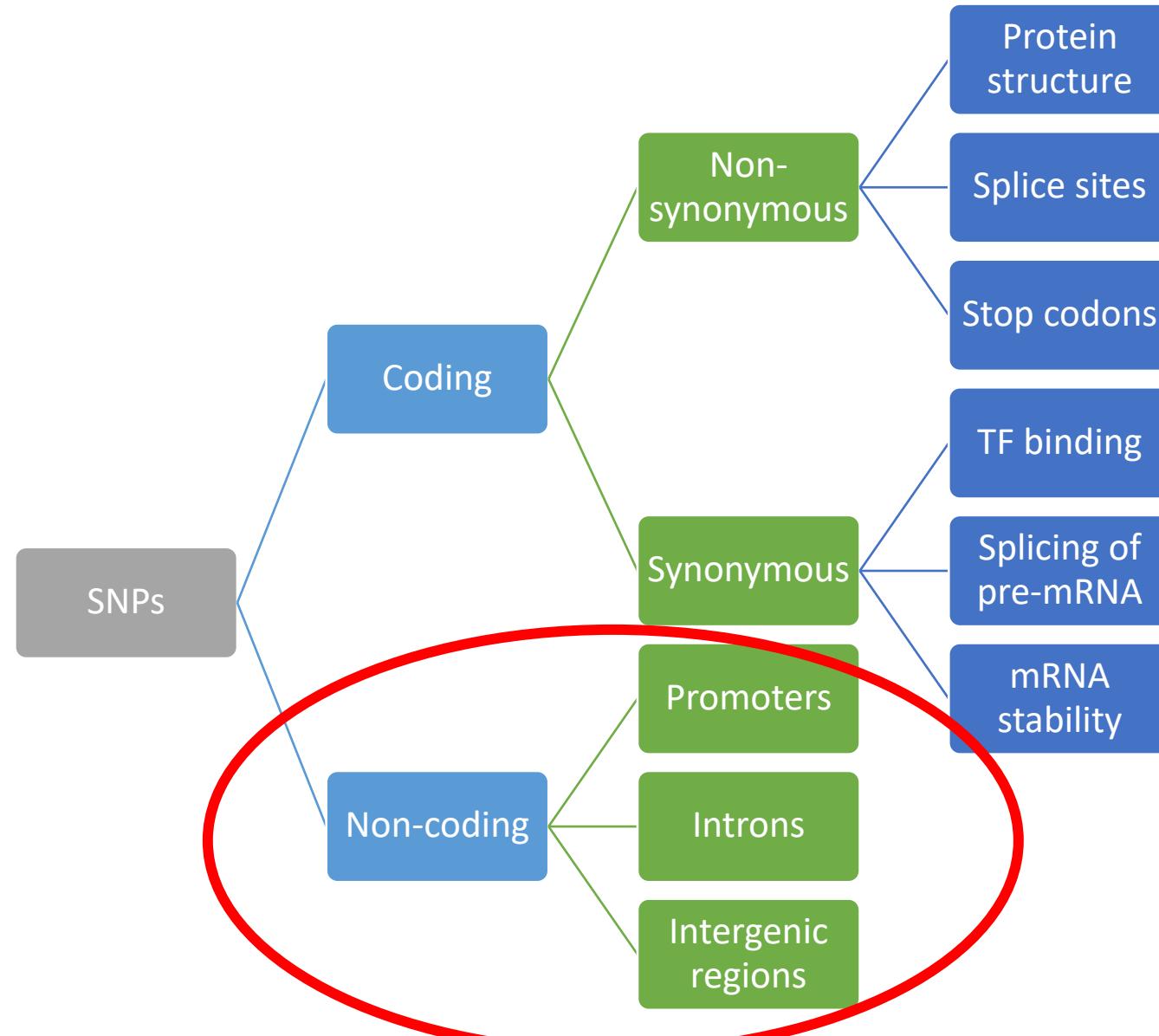
Causal variants?

Function?

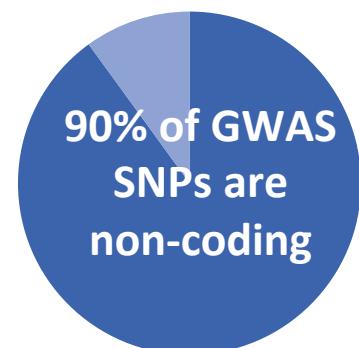
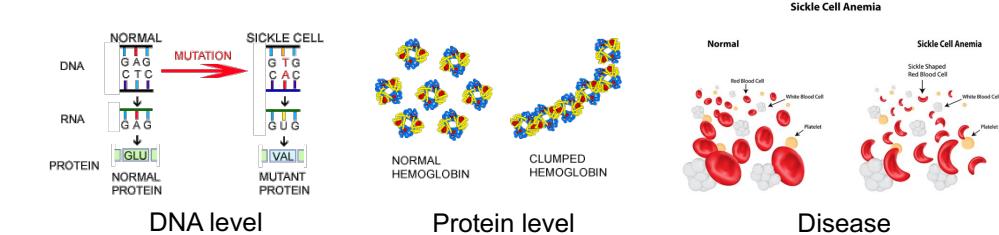
Causal genes?

Causal cell types?

How can SNPs impact biological function?



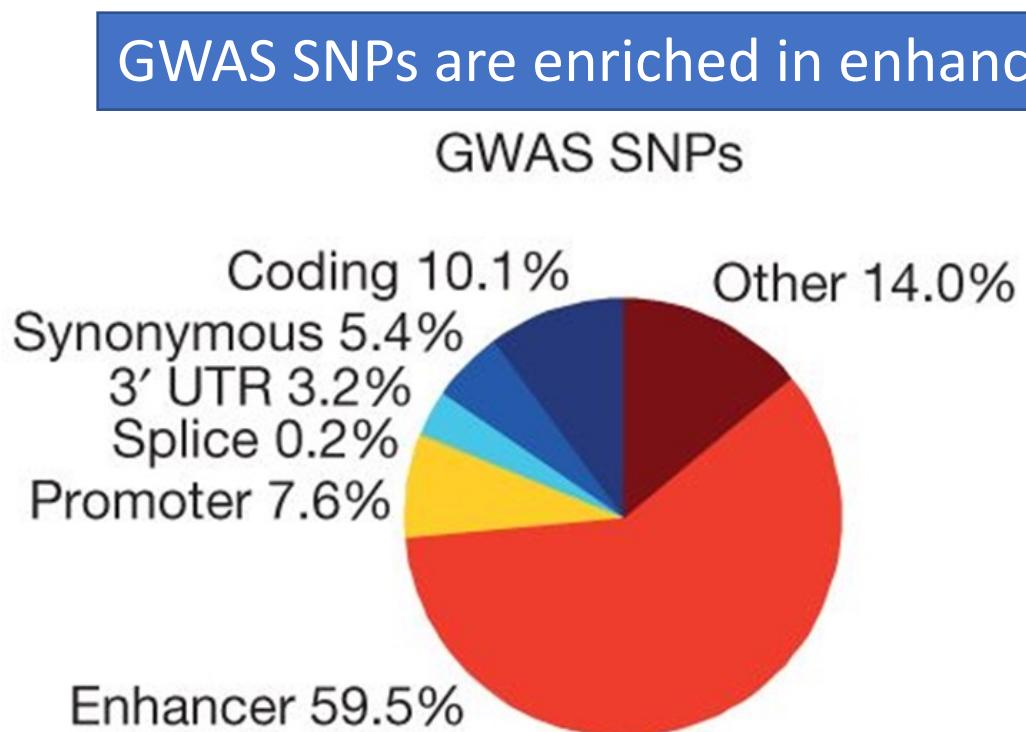
Expected: disease variants result in defective proteins



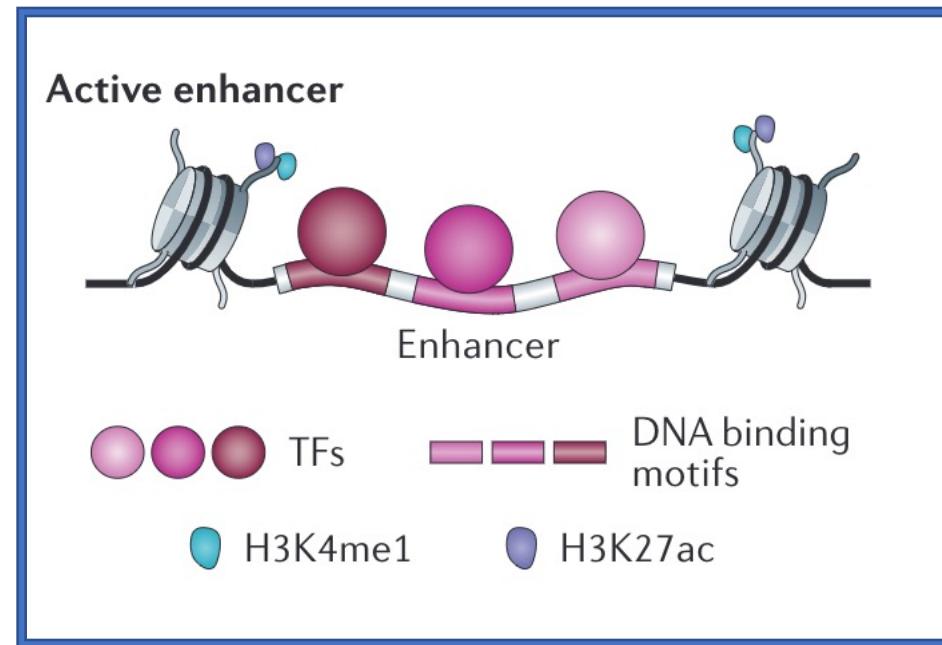
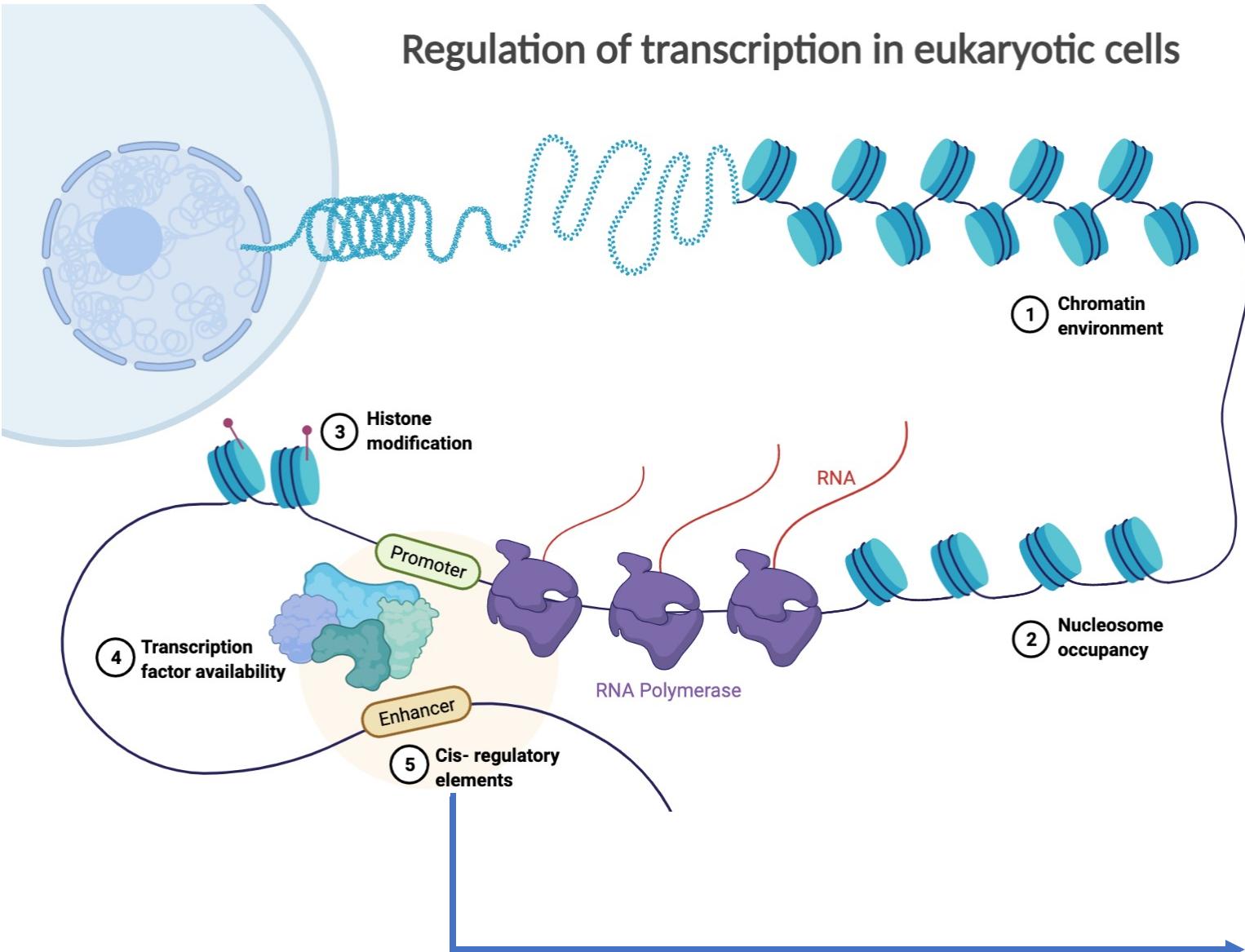
Function???

How can non coding SNPs influence disease?

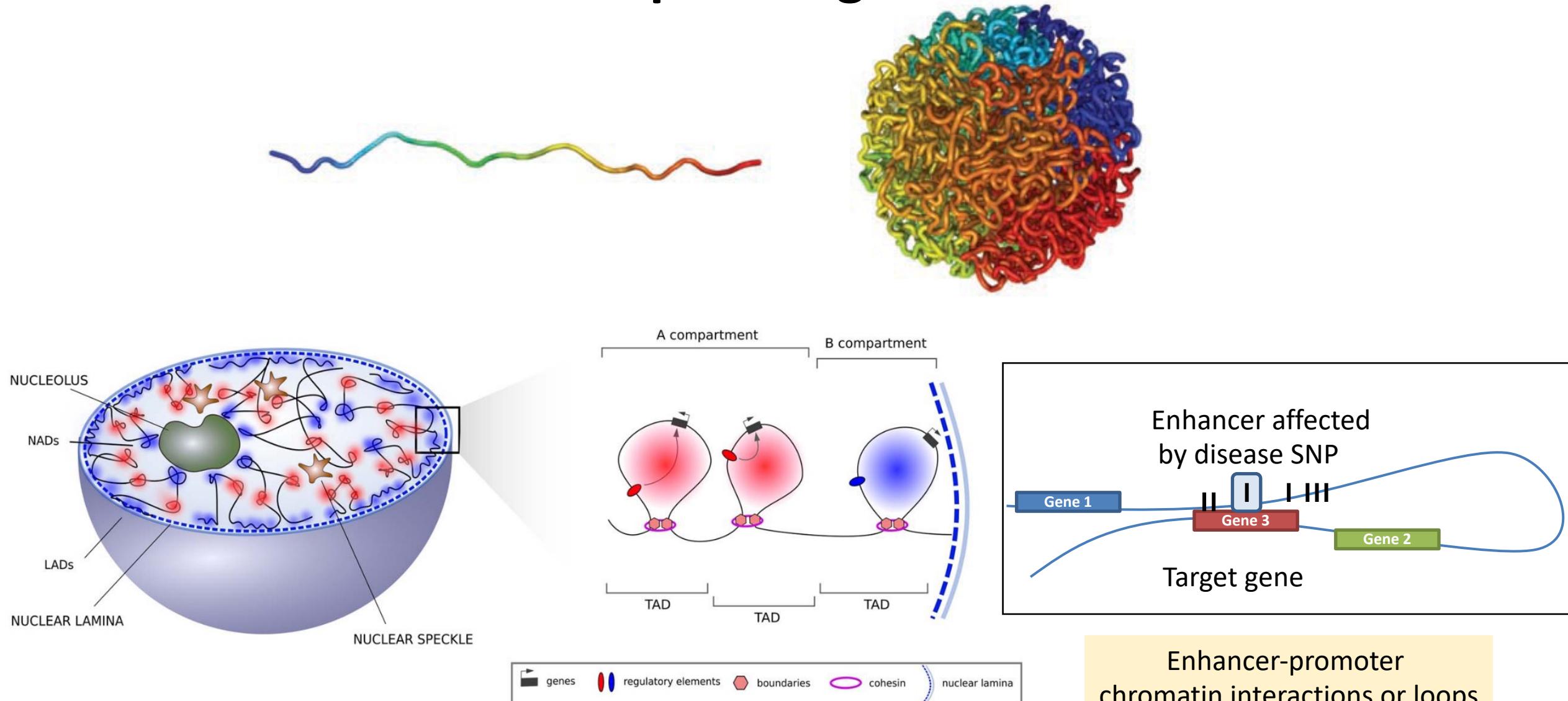
by altering the regulation of gene expression
in disease relevant tissues



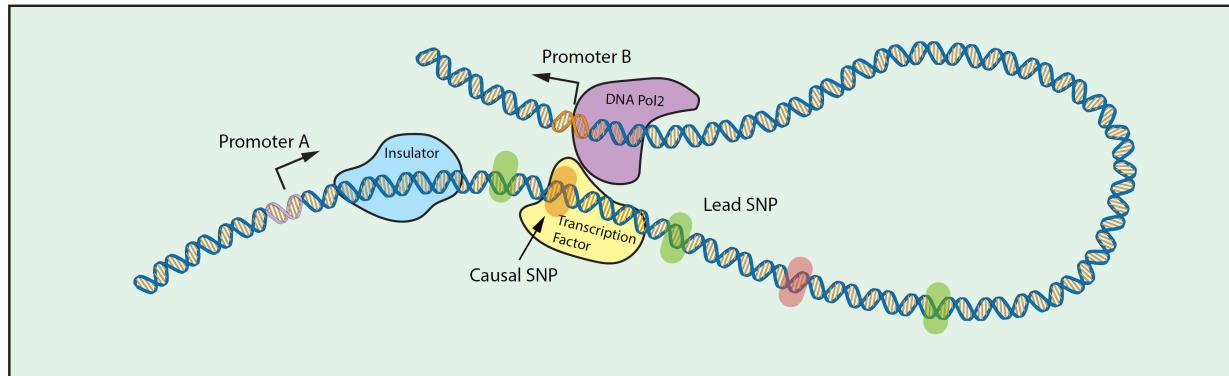
Regulation of transcription in eukaryotic cells



How can enhancers regulate genes that map at long distances?



Functional genomics approaches to understand function of GWAS SNPs



Shi et al. Rheumatology (2020)

A. SNP to function

- Open Chromatin
 - DNase-hypersensitivity
 - ATAC-Seq
- Histone modifications and TF binding
 - ChIP-Seq

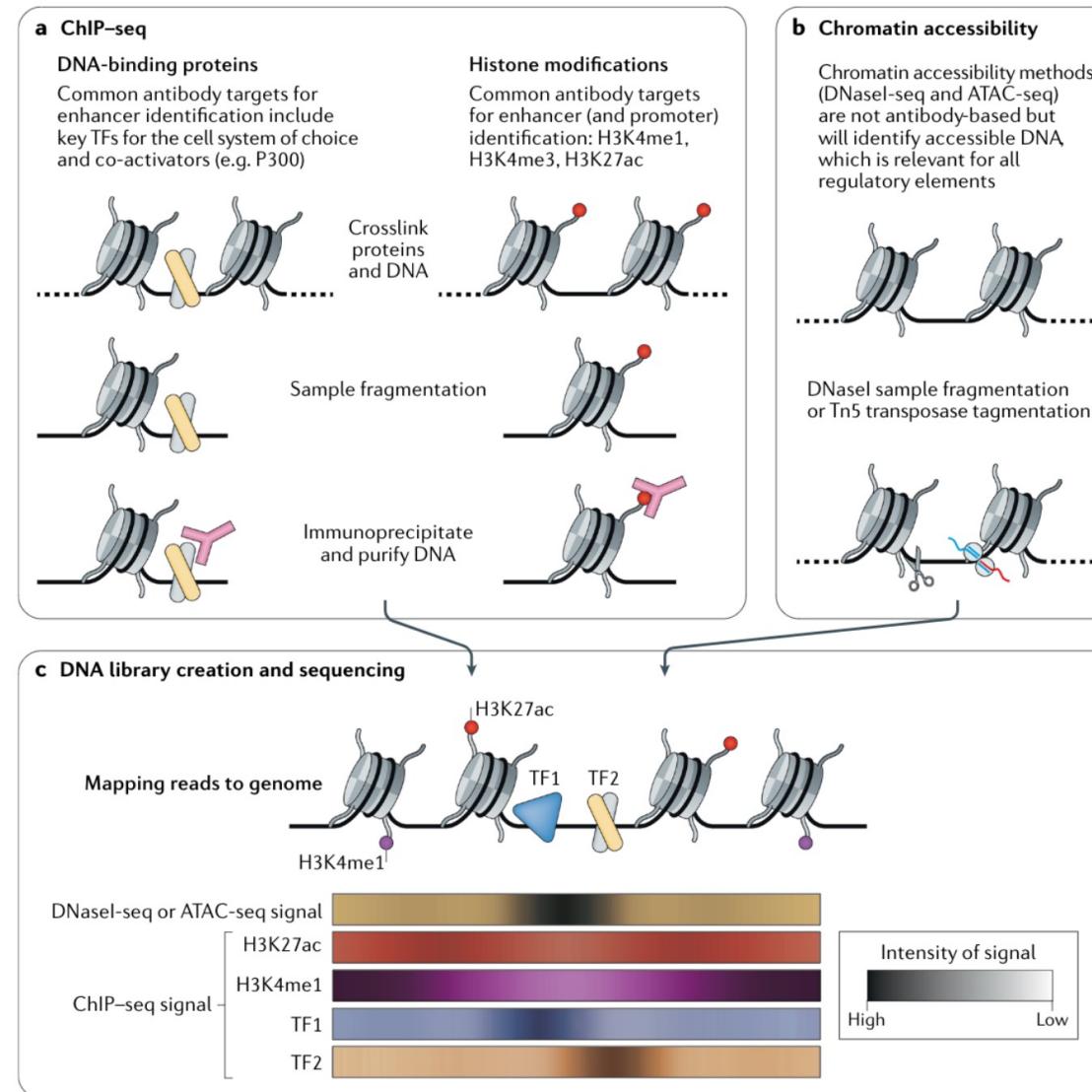
B. SNP to gene

- Gene Expression
 - mRNA sequencing - eQTLs
- Chromatin interactions
 - Hi-C

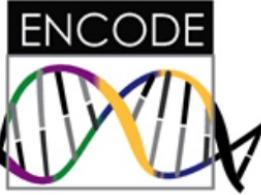
SNP to function

Identification of enhancers by DNA-binding proteins and DNA accessibility

ChIP-seq
Chromatin immuno-precipitation
coupled to sequencing

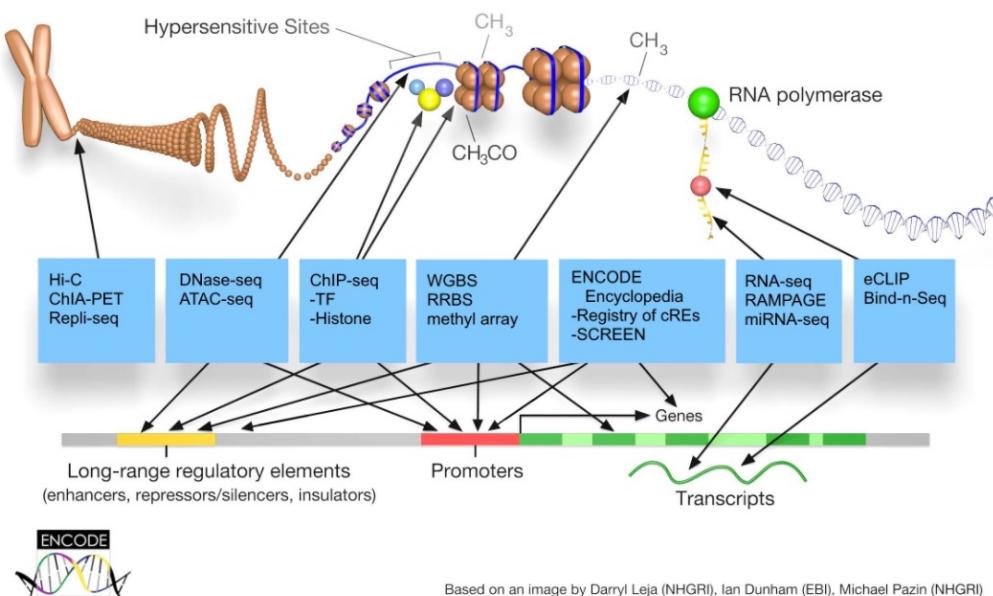


ATAC-seq
Assay for transposase- accessible chromatin

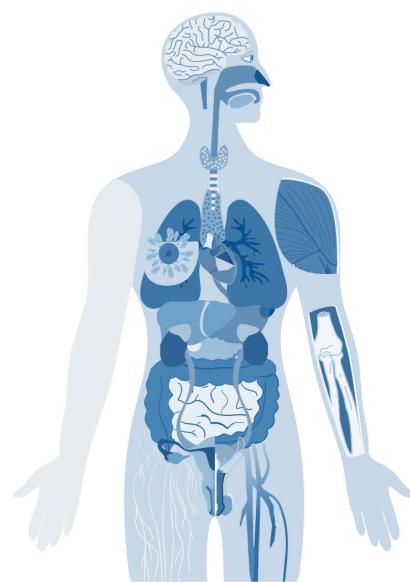


ENCODE: Encyclopaedia of DNA Elements

Experiments



Tissues



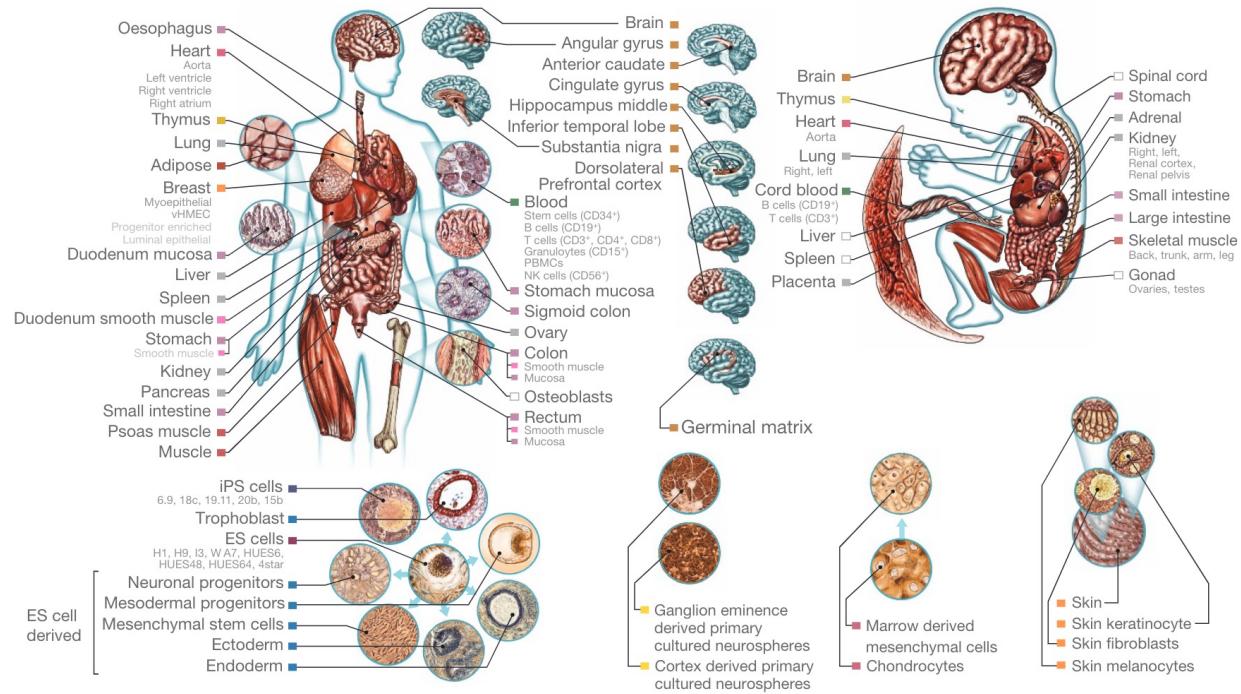
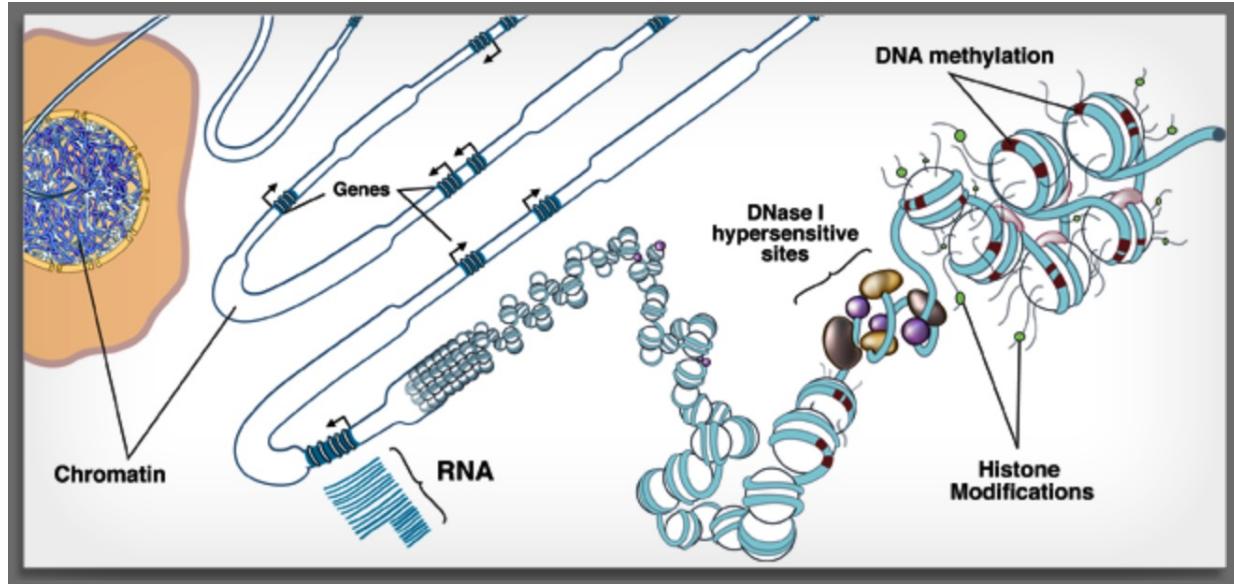
adrenal gland	arterial blood vessel
bone element	brain
breast	bronchus
colon	esophagus
eye	gallbladder
gonad	heart
intestine	kidney
large intestine	limb
liver	lung
mammary gland	mouth
musculature of body	nerve
nose	ovary
pancreas	penis
pericardium	prostate gland
skeleton	skin of body
small intestine	spinal chord
spleen	stomach
testis	thymus
thyroid gland	trachea
tongue	ureter
uterus	urinary bladder
vagina	vein

<https://www.encodeproject.org/>

ENCODE Project Consortium, Jill E. Moore, Michael J. Purcaro, Henry E. Pratt, Charles B. Epstein, Noam Shores, Jessika Adrian, et al. 2020. "Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes." *Nature* 583 (7818): 699–710.

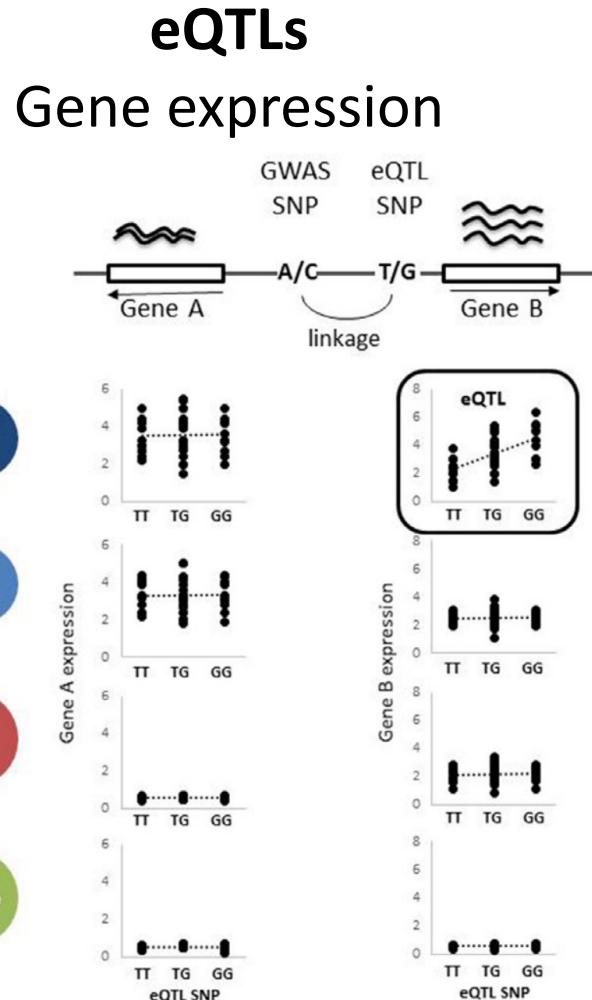
ROADMAP

<http://www.roadmapepigenomics.org/>



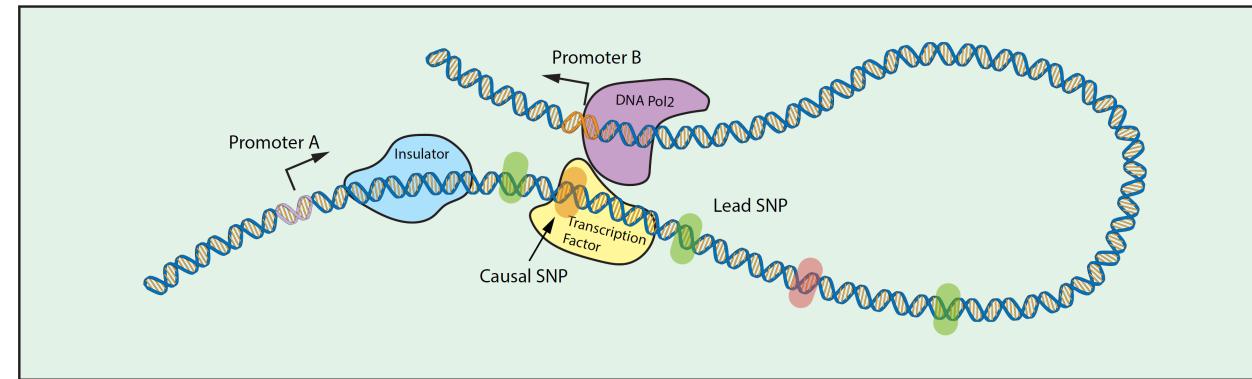
SNP to gene

Identification of the genes impacted by disease SNPs



Hi-C

Chromatin interactions



Shi et al. Rheumatology (2020)

SNP to gene

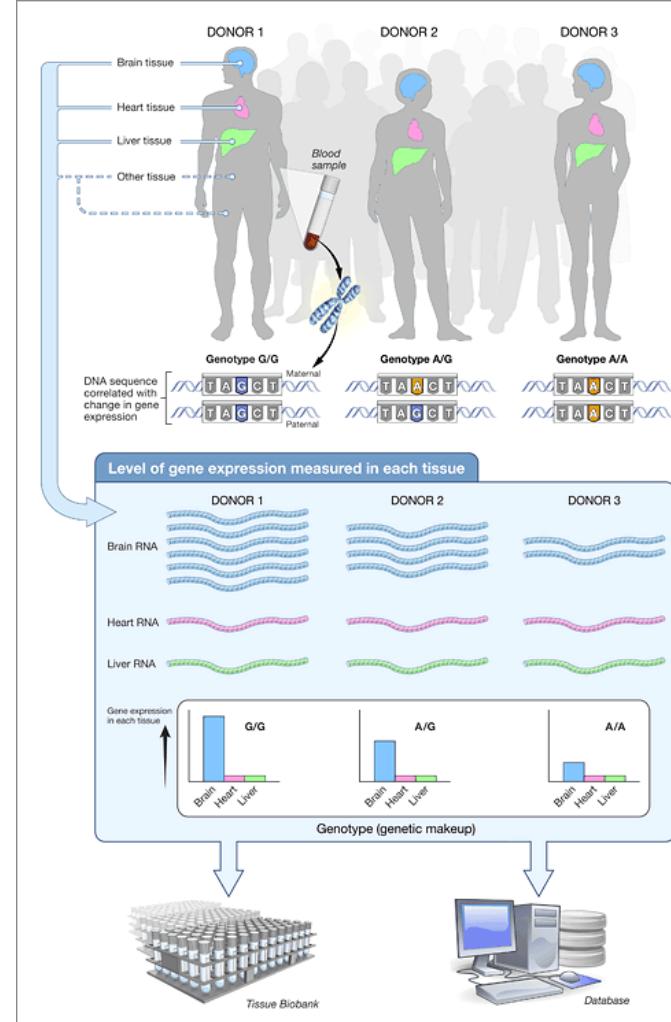
eQTLs: large scale transcriptomics studies

The Genotype-Tissue Expression (GTEx) project

V8 Sample Info

V8 Release	# Tissues	# Donors	# Samples
Total	54	948	17382
With Genotype	54	838	15253
Has eQTL Analysis*	49	838	15201

* Number of samples with genotype ≥ 70



 **GTEx** Portal

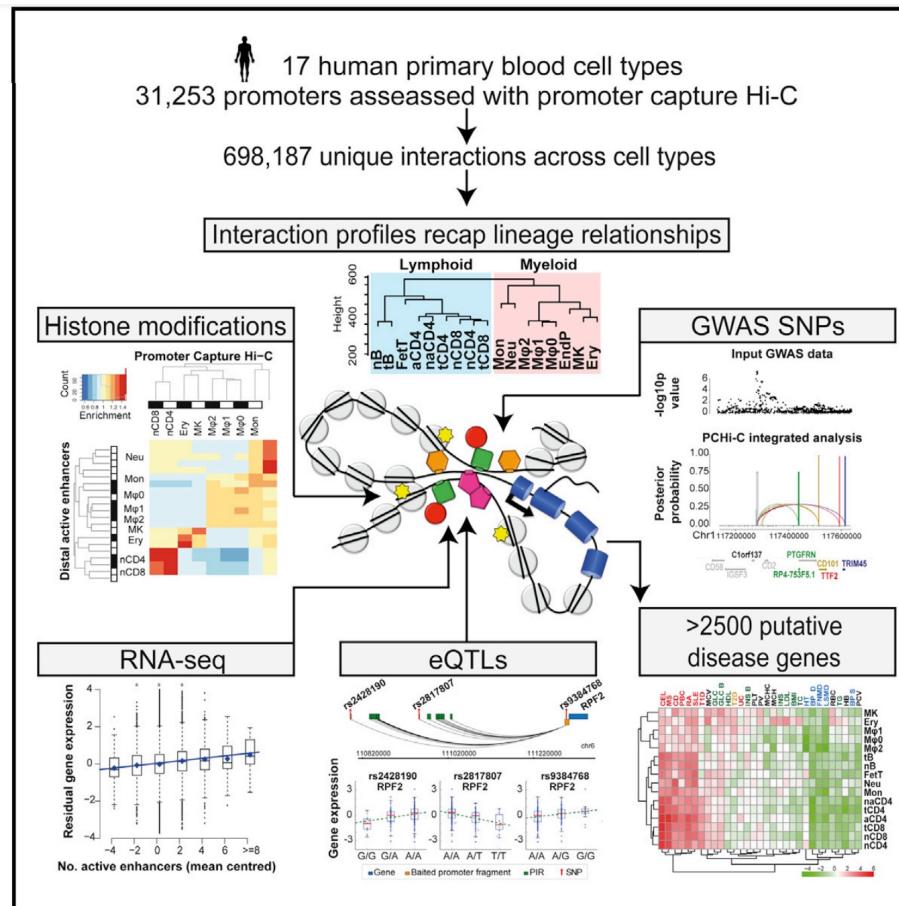
<https://gtexportal.org/home/>

Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters

Biola M. Javierre,^{1,11} Oliver S. Burren,^{2,11} Steven P. Wilder,^{3,11} Roman Kreuzhuber,^{3,4,5,11} Steven M. Hill,^{6,11} Sven Sewitz,¹ Jonathan Cairns,¹ Steven W. Wingett,¹ Csilla Várnai,¹ Michiel J. Thiecke,¹ Frances Burden,^{4,5} Samantha Farrow,^{4,5} Antony J. Cutler,² Karola Rehnström,^{4,5} Kate Downes,^{4,5} Luigi Grassi,^{4,5} Myrto Kostadima,^{3,4,5} Paula Freire-Pritchett,¹ Fan Wang,⁶ The BLUEPRINT Consortium, Hendrik G. Stuennenberg,⁷ John A. Todd,² Daniel R. Zerbino,³ Oliver Stegle,³ Willem H. Ouwehand,^{4,5,8,9} Mattia Frontini,^{4,5,8,*} Chris Wallace,^{2,6,10,*} Mikhail Spivakov,^{1,12,*} and Peter Fraser^{1,*}

SNP to gene: chromatin interactions (Hi-C)

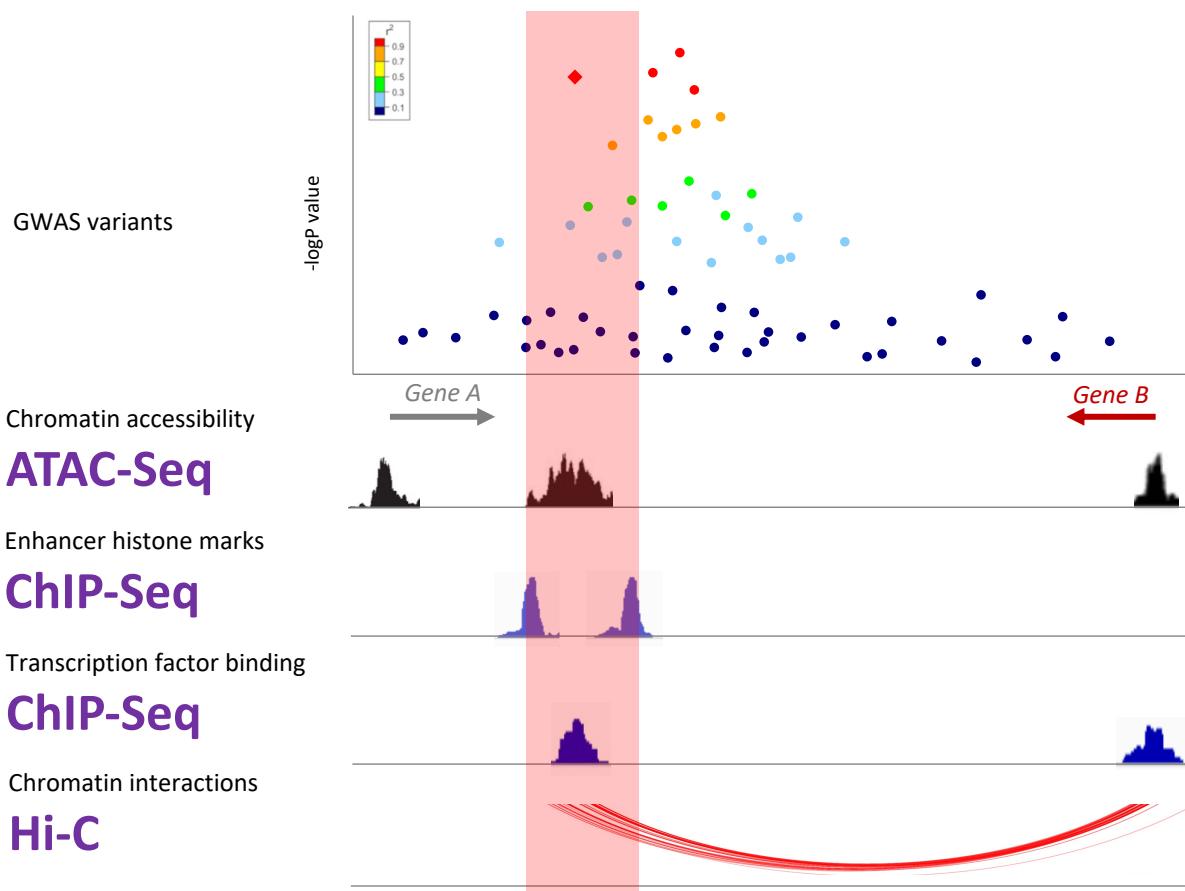
Cell 167, 1369–1384 (2016)



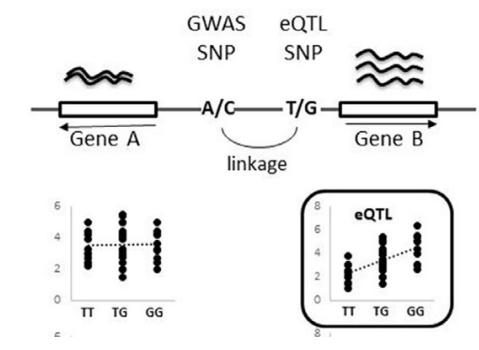
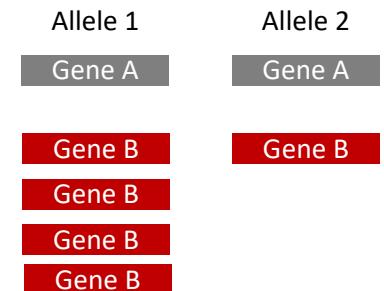
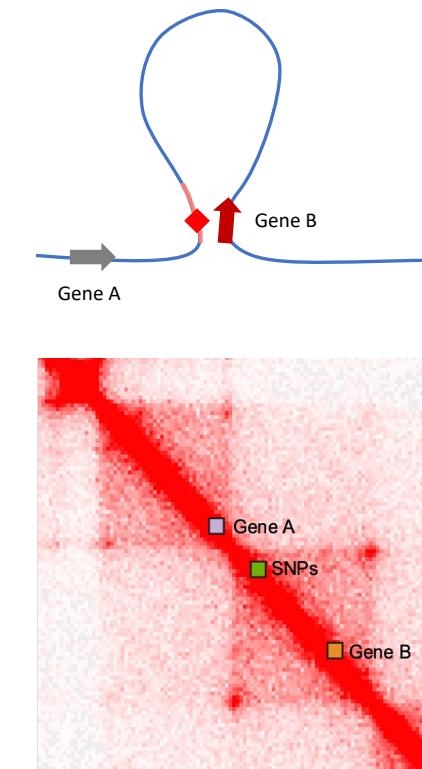
- Chromatin interaction patterns are cell type specific and segregate with the hematopoietic tree
- Promoter-interacting regions are enriched for regulatory chromatin features and eQTLs
- Promoter interactions link non-coding GWAS variants with putative target genes: more than 2,500 putative disease causing genes were identified

Functional annotation of GWAS: Integration with cell-type specific functional genomics datasets

A) From SNP to function

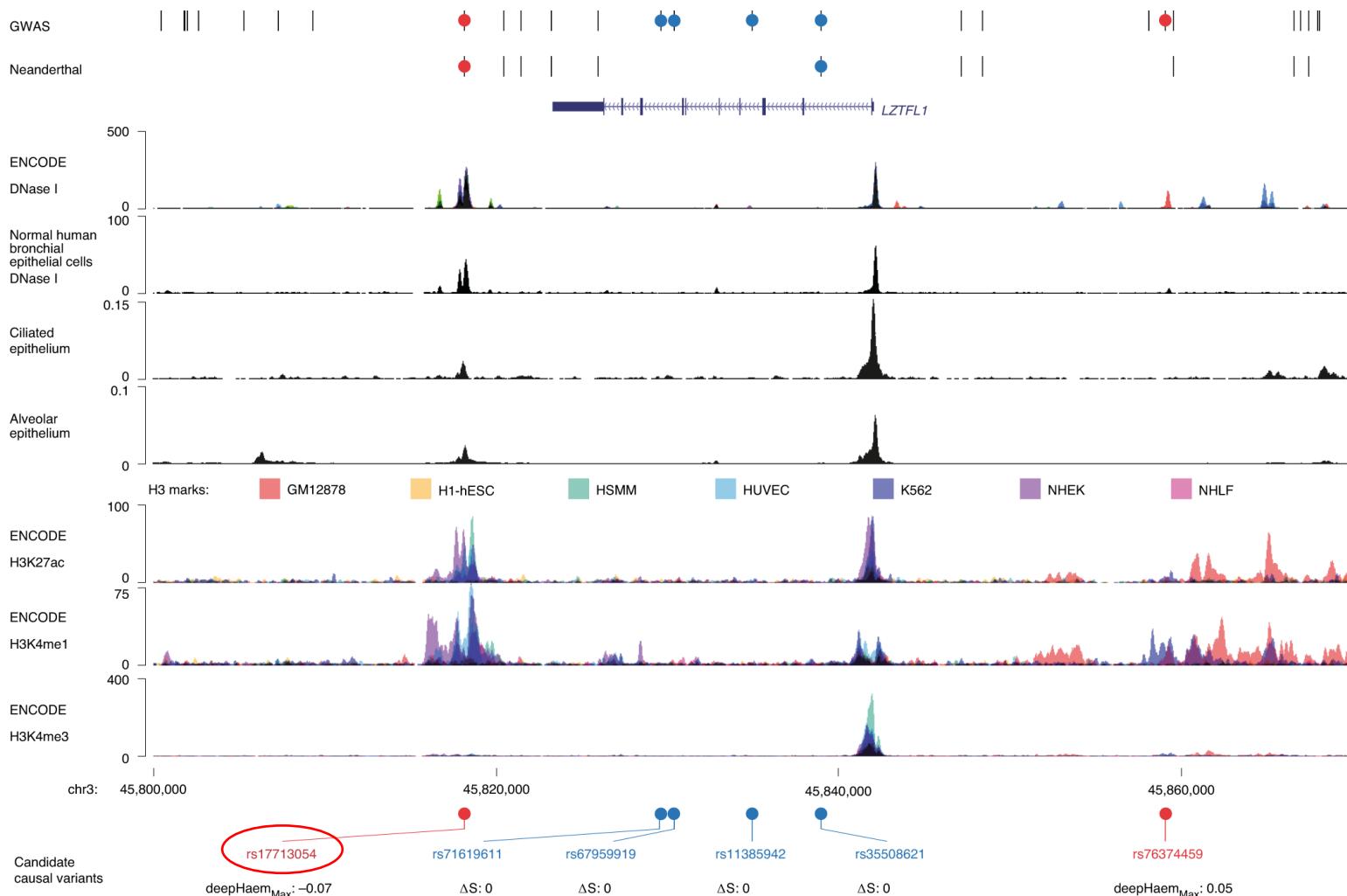


B) From SNP to gene



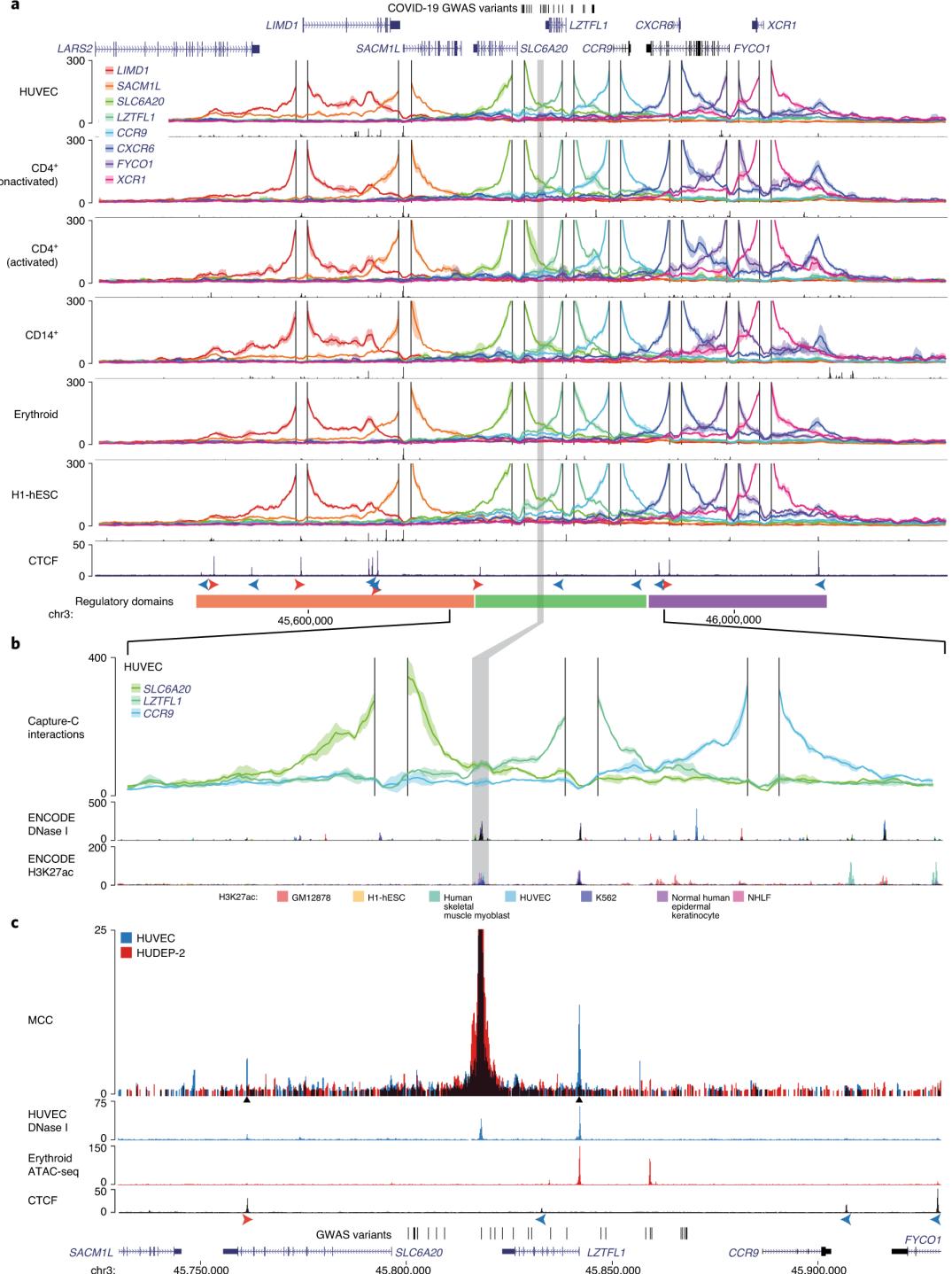
Identification of *LZTFL1* as a candidate effector gene at a COVID-19 risk locus.

Downes et al. Nature Genetics 53, 1606–1615 (2021)



Identification of a potentially causative COVID-19 risk variant.

COVID-19 risk variants from GWAS were assessed for multiple mechanisms. All genome-wide-significant variants and linked variants are shown (GWAS) as are variants present in the Vindija Neanderthal¹² risk haplotype. The circles indicate variants assessed for splicing changes (blue circles, SpliceAI¹⁸: ΔS score (0–1, where 1 is the most damaging)), and presence in *cis*-regulatory elements using open chromatin in 95 ENCODE overlaid DNase I datasets (red circles), normal human bronchial epithelial cells and scATAC-seq from fetal ciliated and alveolar epithelia³⁴. Histone H3 modification tracks show the presence of marks associated with active transcription (H3K27ac) at enhancers (H3K4me1) and promoters (H3K4me3). Variants in open chromatin are given deepHaem damage scores (0–1) with sign indicating increased (–) or decreased (+) accessibility. The region shown is chr3:45,800,000–45,870,000, hg38. HSMM, human skeletal muscle myoblast; NHEK, normal human epidermal keratinocyte.

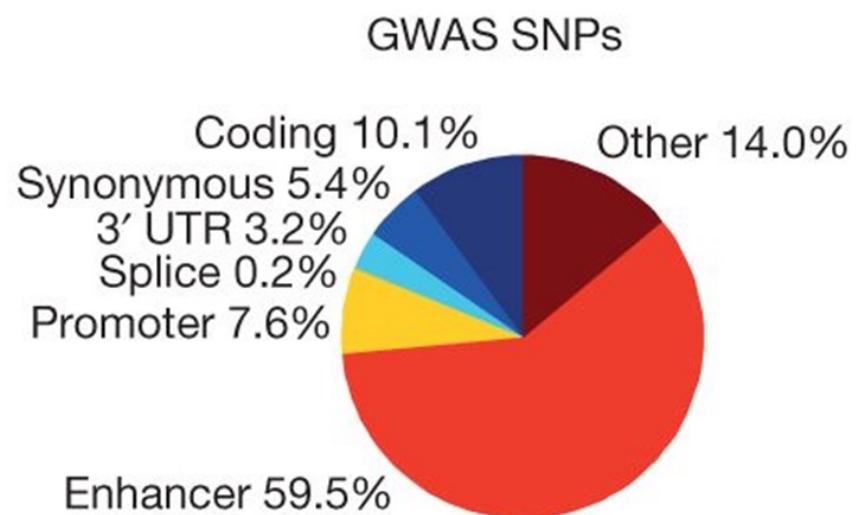
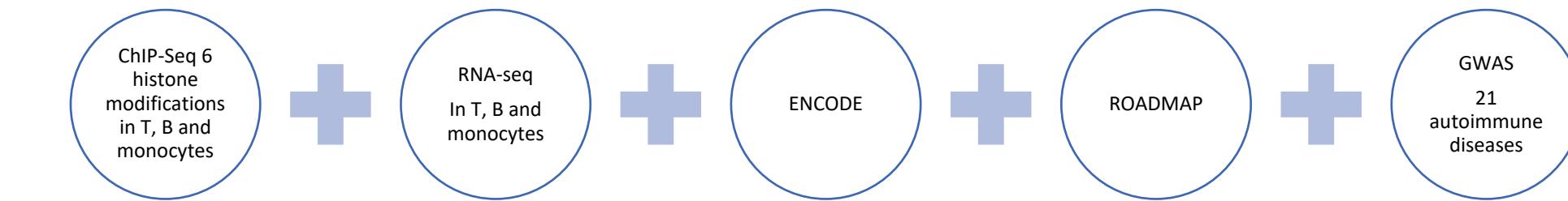


Identification of causal gene:
The rs1773054 enhancer interacts with the *LZTFL1* promoter and the SNP is an eQTL for *LZTFL1*.
This gene is involved in an important viral response pathway.

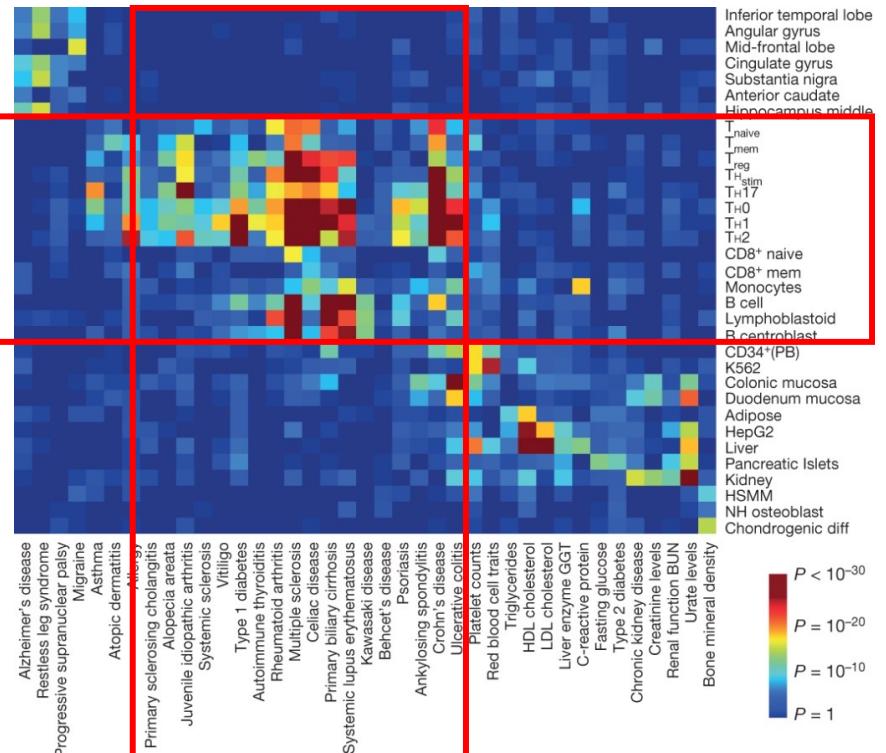
Genetic and epigenetic fine mapping of causal autoimmune disease variants

- Functional annotation
- Fine mapping
- Identification of disease relevant cells

Kyle Kai-How Farh^{1,2*}, Alexander Marson^{3*}, Jiang Zhu^{1,4,5,6}, Markus Kleinewietfeld^{1,7†}, William J. Housley⁷, Samantha Beik¹, Noam Shoresh¹, Holly Whitton¹, Russell J. H. Ryan^{1,5}, Alexander A. Shishkin^{1,8}, Meital Hatan¹, Marlene J. Carrasco-Alfonso⁹, Dita Mayer⁹, C. John Luckey⁹, Nikolaos A. Patsopoulos^{1,10,11}, Philip L. De Jager^{1,10,11}, Vijay K. Kuchroo¹², Charles B. Epstein¹, Mark J. Daly^{1,2}, David A. Hafler^{1,7§} & Bradley E. Bernstein^{1,4,5,6§}



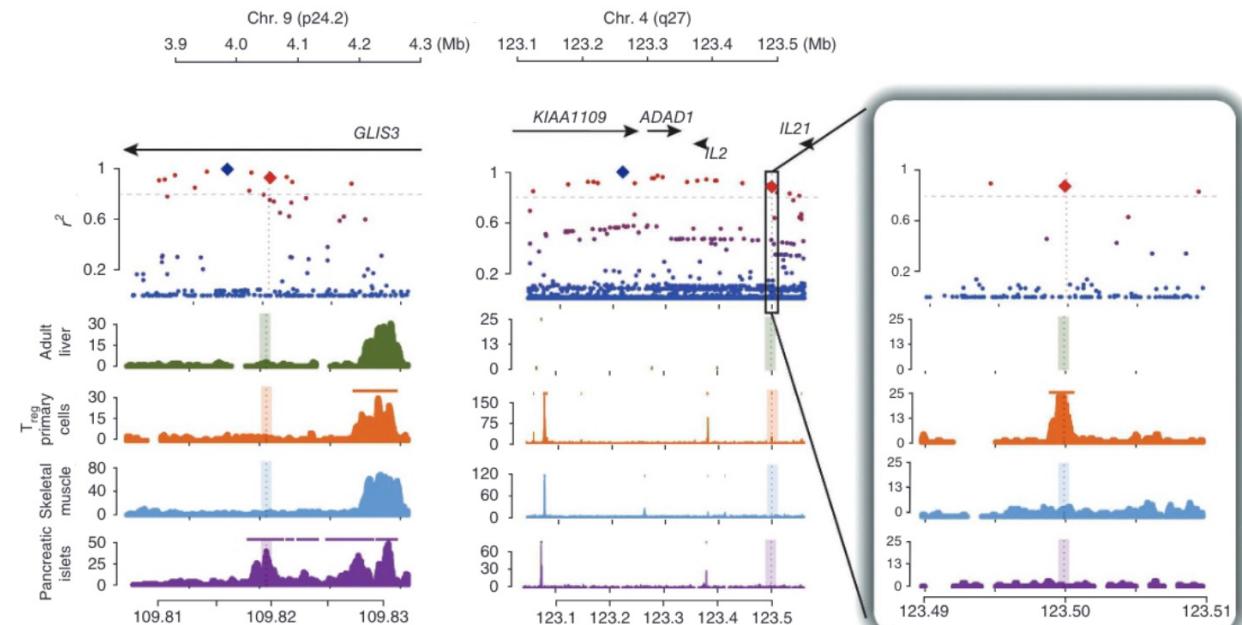
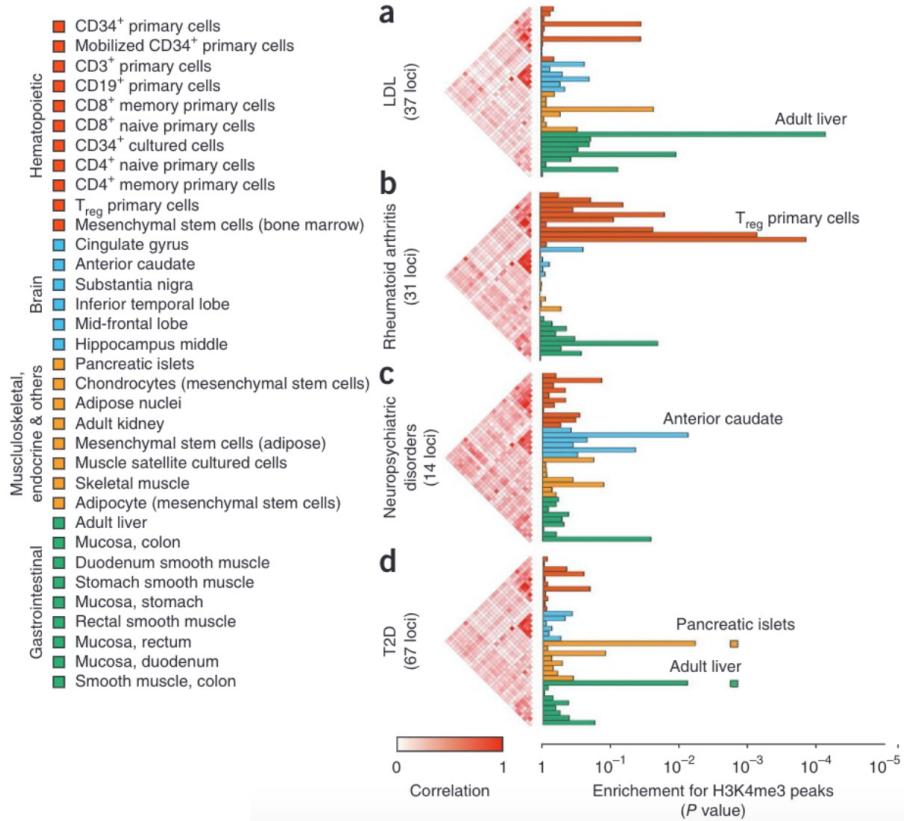
Nature 518, 337–343 (2015)



Chromatin marks identify critical cell types for fine mapping complex trait variants

Nature Genetics 45, 124–130 (2013)

Gosia Trynka^{1–4,8}, Cynthia Sandor^{1–4,8}, Buhm Han^{1–4}, Han Xu⁵, Barbara E Stranger^{1,4,7}, X Shirley Liu⁵ & Soumya Raychaudhuri^{1–4,6}

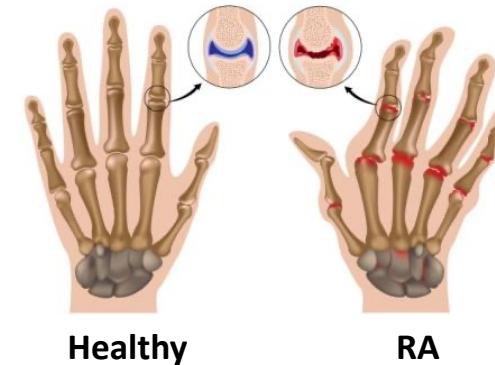


GLIS3 locus – T2D

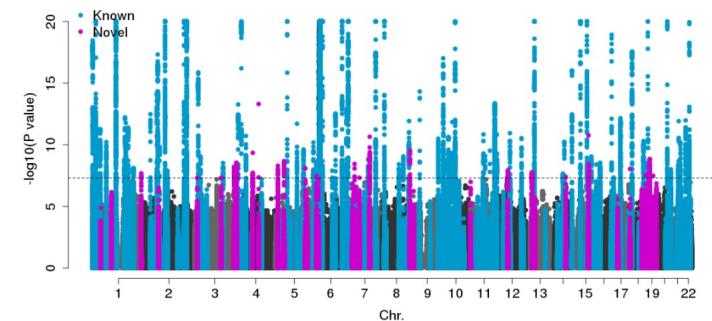
IL2-IL21 locus: RA

Rheumatoid arthritis (RA)

- Autoimmune rheumatic disease characterized by inflammation and destruction of the joints
- Common, affect millions worldwide
- Chronic diseases with NO CURE
- Correlated with:
 - Increased morbidity
 - Early mortality
- Disability
- Many patients do not respond to available treatments



124 loci



Linking RA SNPs to potential causal genes



ARTICLE

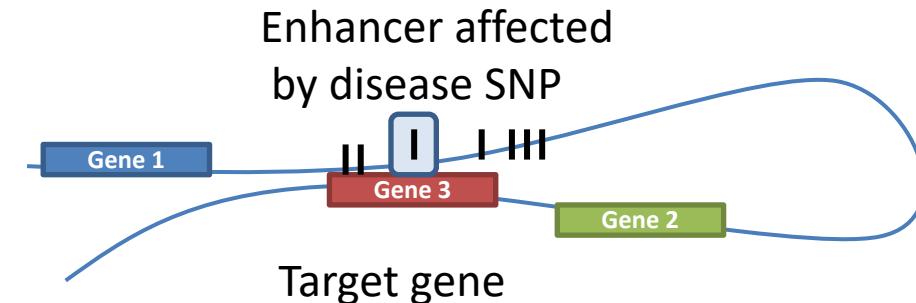
Received 15 Jun 2015 | Accepted 28 Oct 2015 | Published 30 Nov 2015

DOI: 10.1038/ncomms10069 OPEN

Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci

Paul Martin^{1,*}, Amanda McGovern^{1,*}, Gisela Orozco^{1,*}, Kate Duffus¹, Annie Yarwood¹, Stefan Schoenfelder², Nicholas J. Cooper³, Anne Barton^{1,4}, Chris Wallace^{3,5}, Peter Fraser², Jane Worthington^{1,4} & Steve Eyre¹

* These authors contributed equally to this work.



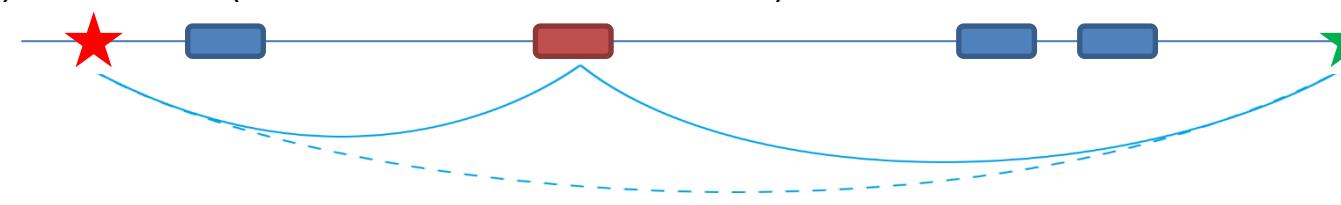
Capture Hi-C: method to map chromatin interactions at specific regions of the genome in a targeted manner

- We targeted all known risk loci (211) for RA, JIA, PsA and T1D to identify potential target genes
- T cell line (Jurkat) and B cell line (GM12878)
- **850 potential causal genes**

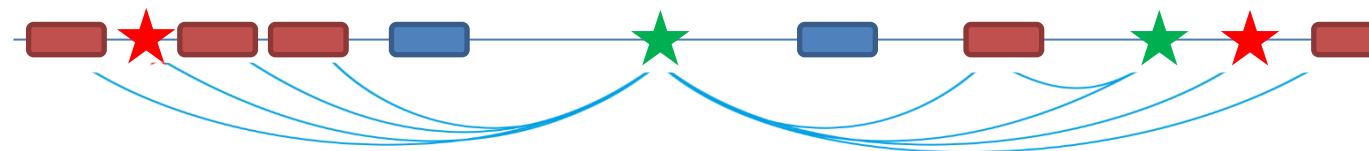
1) Disease-associated SNPs often do not interact with the nearest gene but with genes that map at long distances. Impacts how GWAS SNPs are annotated: GWAS SNPs are traditionally annotated just to the nearest gene . Examples: *FOXO1* (*COG6*), *AZI2* (*EOMES*).



2) SNPs associated with different diseases can interact with each other and the same gene, so different genetic associations may be modulating the same immune pathways. Examples: *PTPRC* (PsA: *DENND1B*; RA: *PTPRC*) , *DEXI* (JIA: *RMI2*; RA: *ZC3H7A*), *ZFP36L1* (RA: *RAD51B*; JIA: *ZFP36L1*)



3) Regions often show a complex pattern of interactions. Examples: *IL20RA* (*TNFAIP3*)



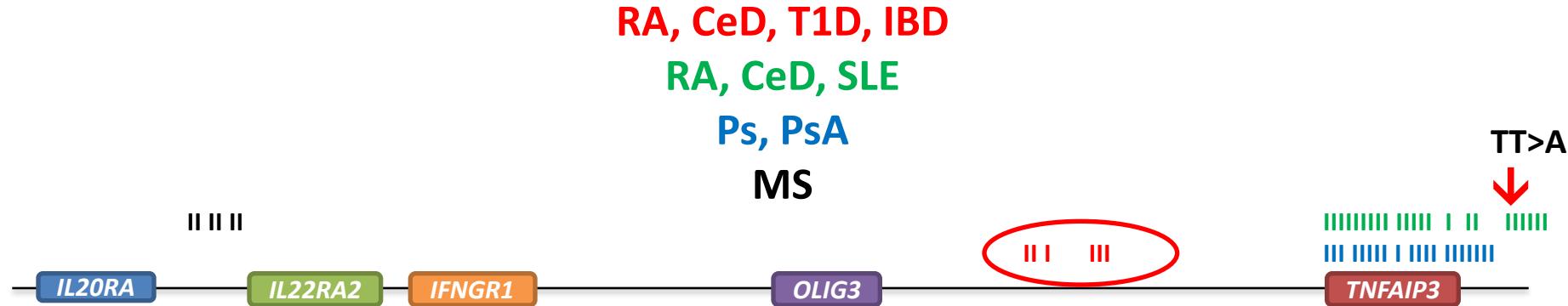
Disease-causing gene

Non disease-causing gene

SNP associated with disease A

SNP associated with disease B

The 6q23 genetic locus in autoimmunity



TT>A alters *TNFAIP3* protein expression
(Musone et al, *Nat Genet* 2008; Adrianto et al, *Nat Genet* 2011; Wang et al, *Plos Genet*, 2013)

RA intergenic region: FUNCTION UNKNOWN

RESEARCH

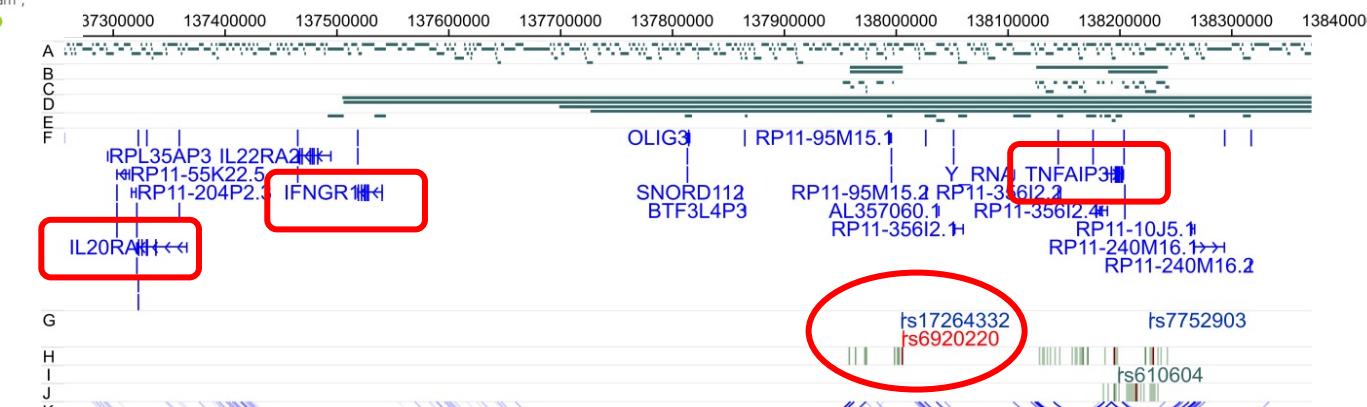
Open Access



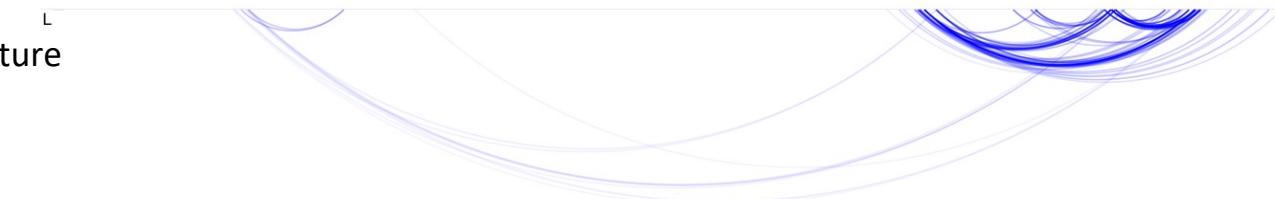
Capture Hi-C identifies a novel causal gene, *IL20RA*, in the pan-autoimmune genetic susceptibility region 6q23

Amanda McGovern¹, Stefan Schoenfelder², Paul Martin¹, Jonathan Massey¹, Kate Duffus¹, Darren Plant^{1,3}, Annie Yarwood¹, Arthur G. Pratt⁴, Amy E. Anderson⁴, John D. Isaacs⁴, Julie Diboll⁴, Nishanthi Thalayasingam⁴, Caroline Ospelt⁵, Anne Barton^{1,3}, Jane Worthington^{1,3}, Peter Fraser², Stephen Eyre¹ and Gisela Orozco¹

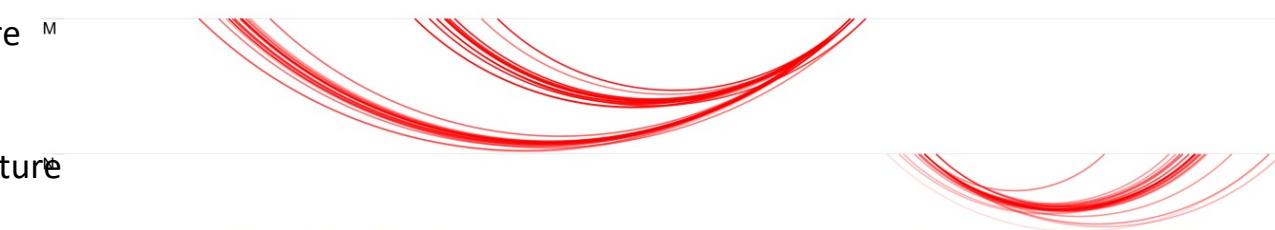
Region capture
B cell line



Promoter capture
B cell line



Region capture
T cell line



Promoter capture
T cell line

Intergenic RA SNPs interact not only with *TNFAIP3*, but also with *IL20RA* and *IFNGR1*

Further investigating the function of the intergenic SNPs in transcriptional regulation

rs6920220 is strongly correlated with 8 other SNPs



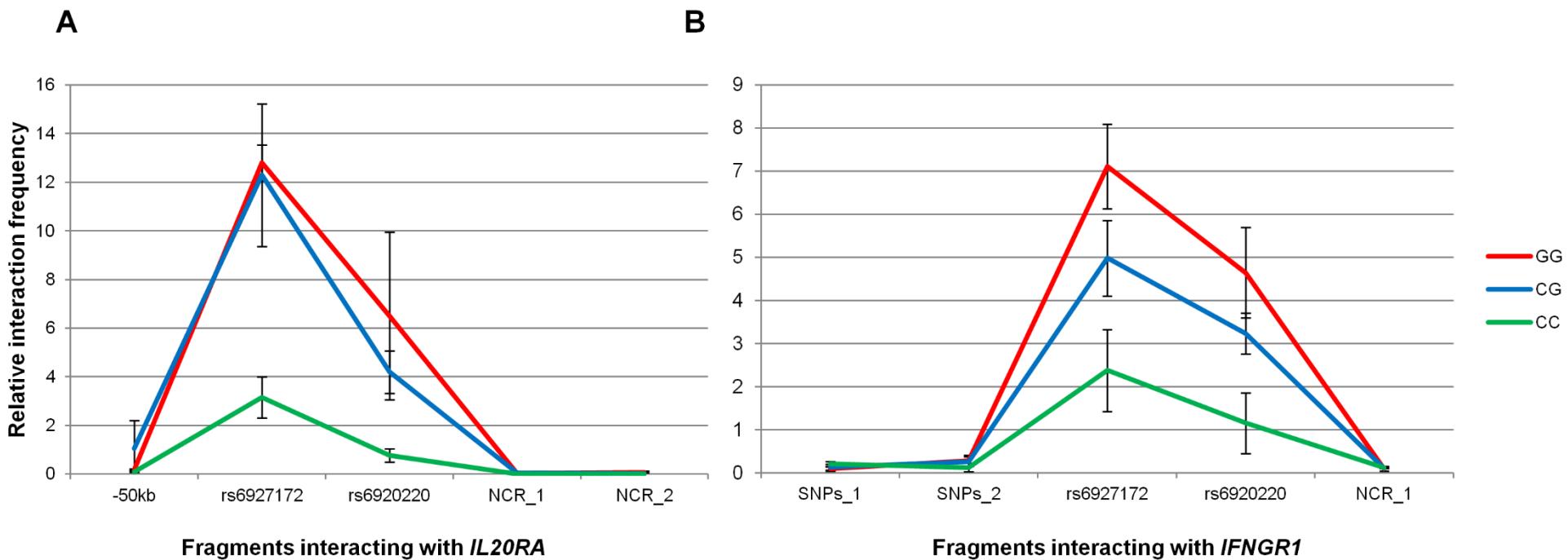
Prioritization of SNPs for functional studies: Functional annotation

**rs6927172 showed
more evidence of regulatory activity**

- Conserved
- Enhancer histone marks in 15 cell types
- DNAse hypersensitivity sites in 14 cell types
- Binding of regulatory proteins in 13 cell types
- Changes binding sites for 8 transcription factors

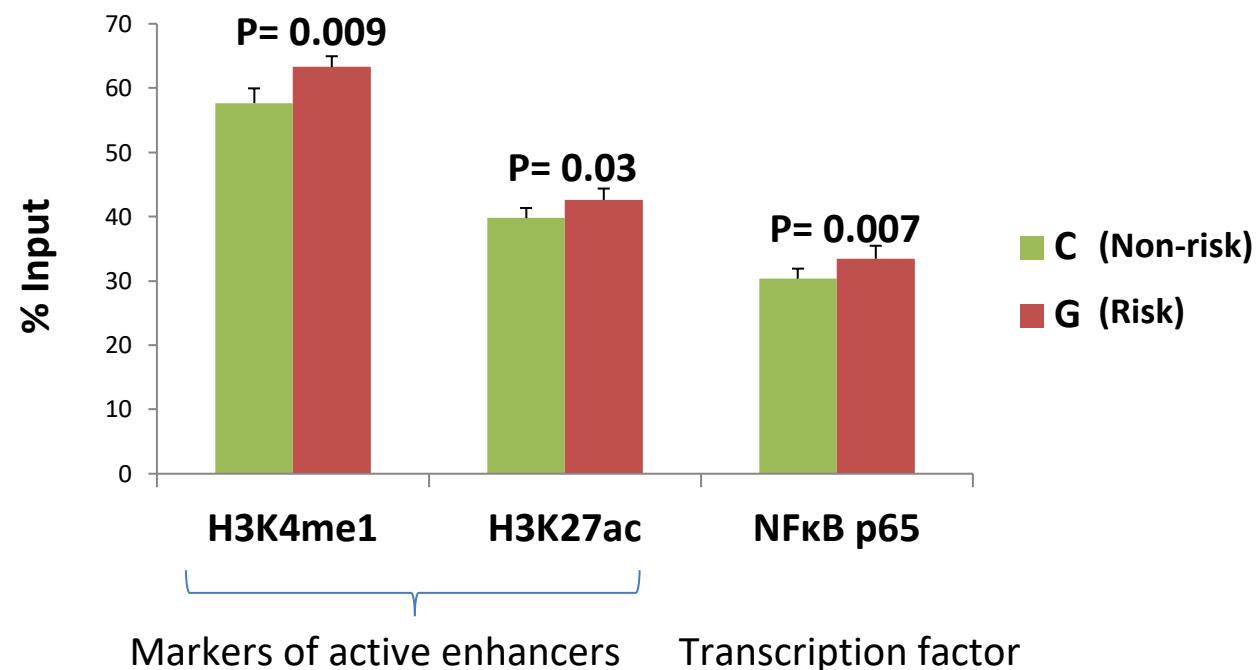
rs6927172 risk allele shows higher frequency of interactions with *IL20RA* and *IFNGR1*

Genotype-specific 3C



rs6927172 risk allele shows a modest but significant increase in binding of regulatory proteins in T-cells

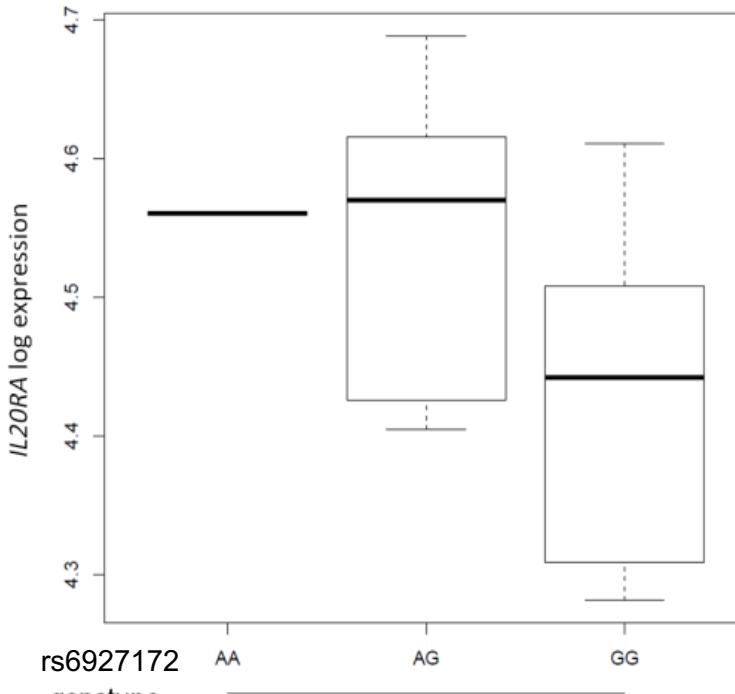
Allele-specific ChIP-qPCR in Jurkat T cells



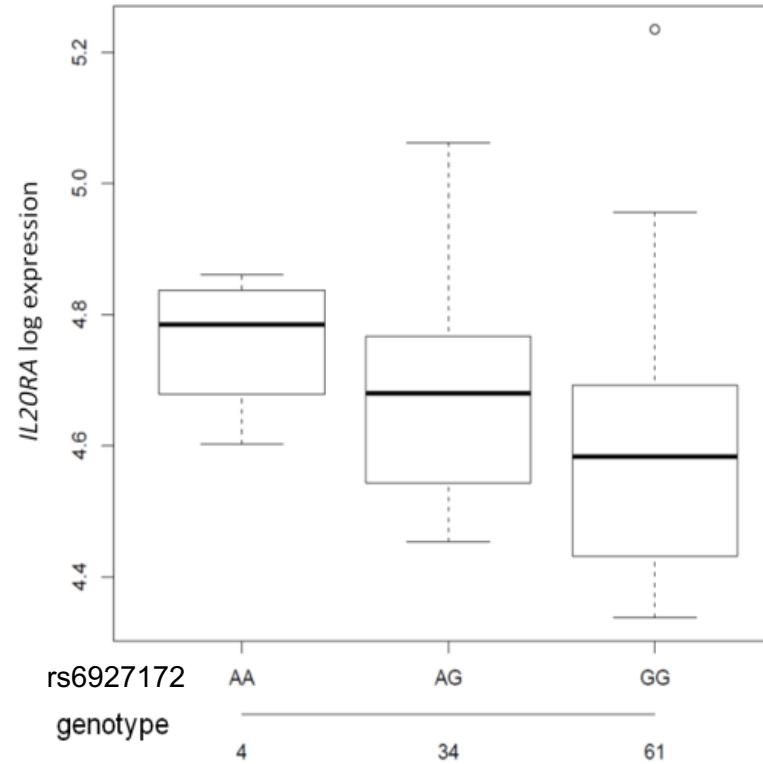
The risk allele may be correlated with an increased regulatory activity

rs6927172 risk allele is correlated with increased expression of *IL20RA* in primary CD4+ T cells

eQTL in T cells
from RA patients

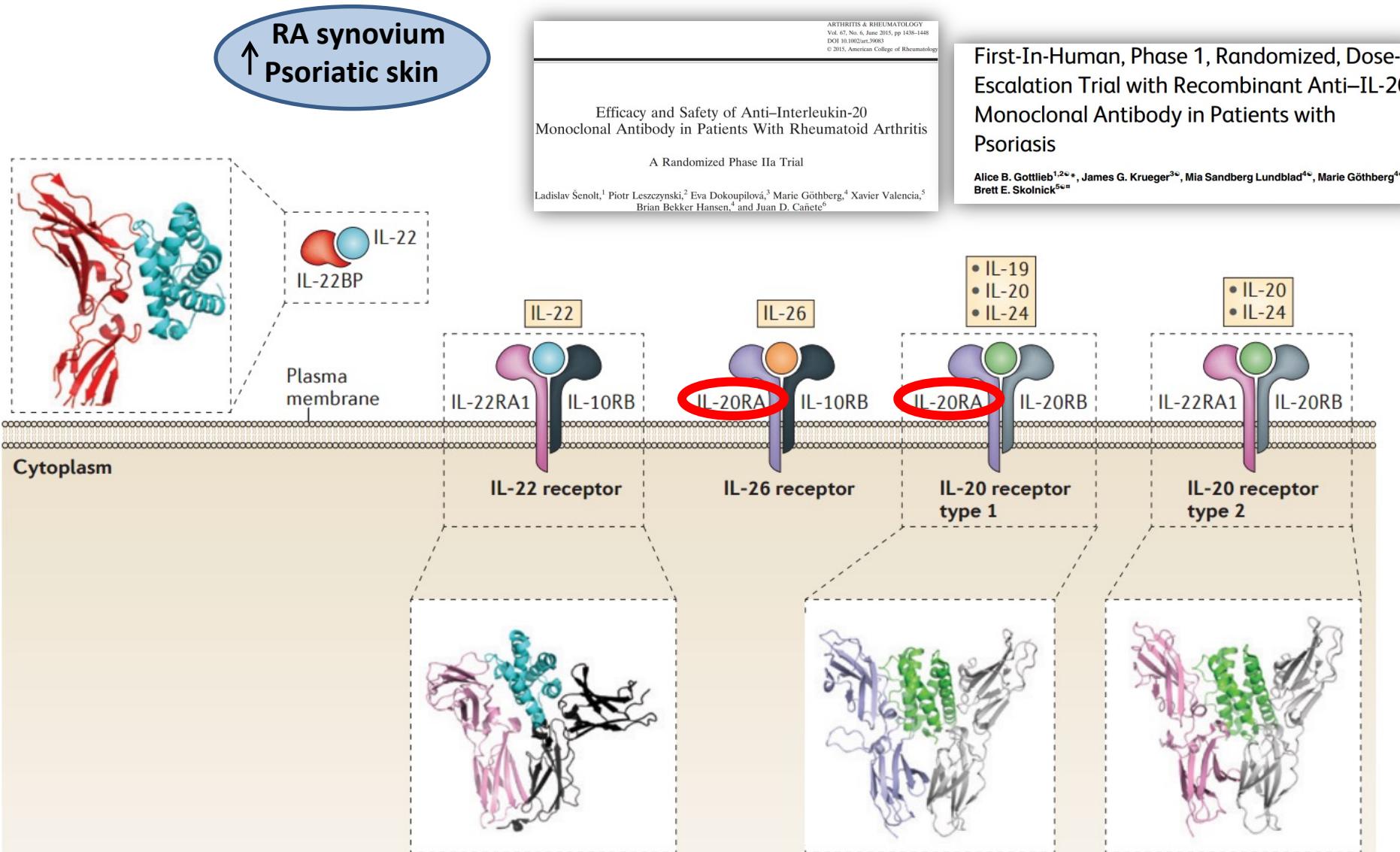


21 Healthy individuals
P= 0.02



99 RA patients
P= 0.006

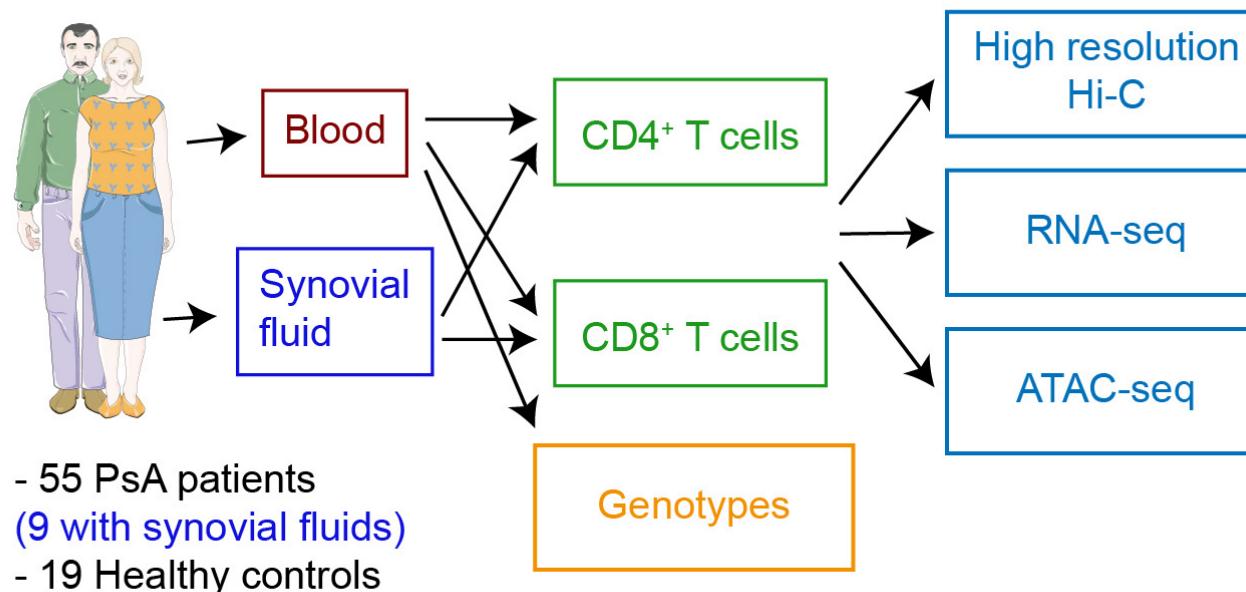
IL-20 is the target of an existing drug

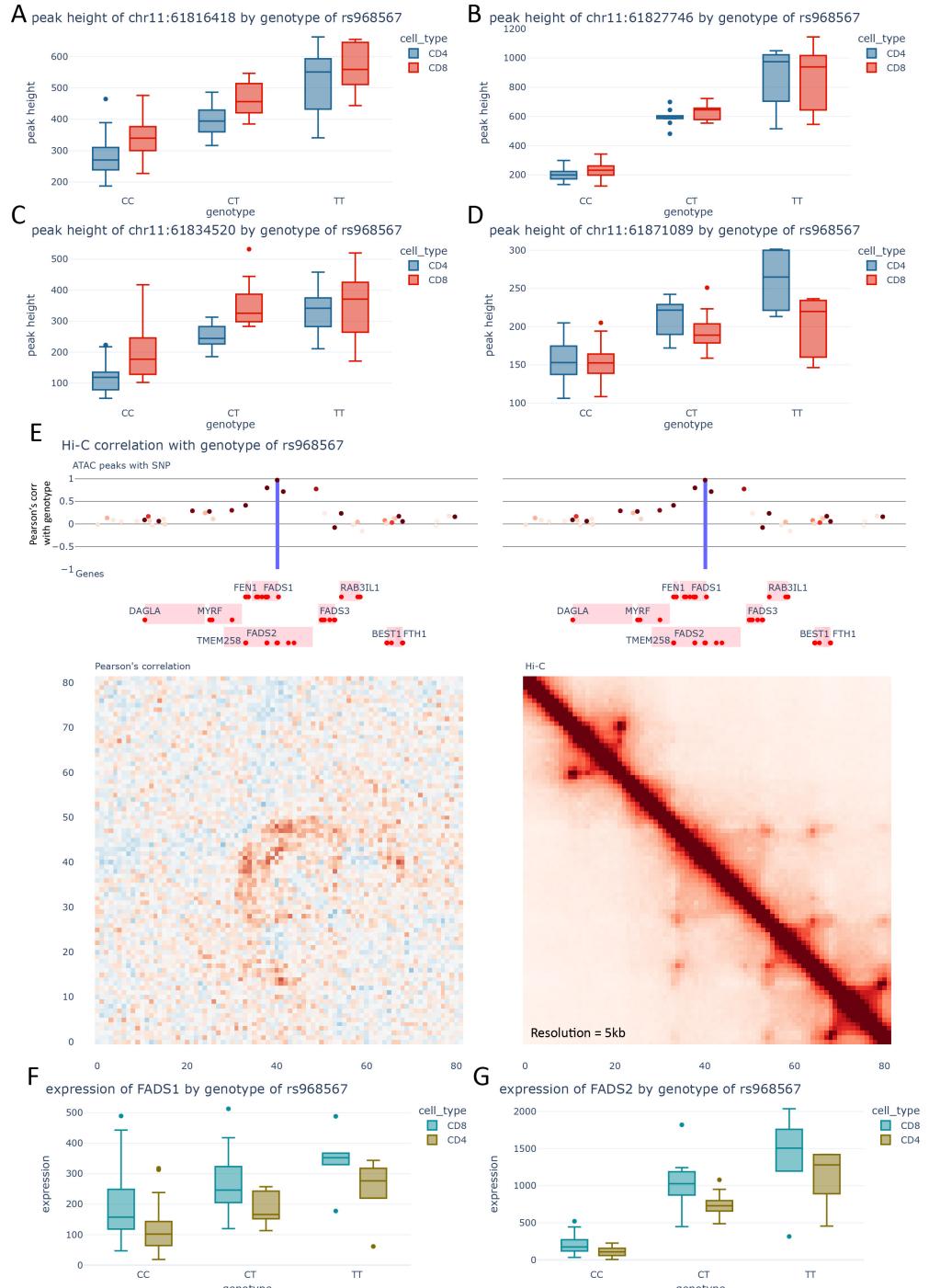


Multi-omics analysis in primary T cells elucidates mechanisms behind disease associated genetic loci

Shi et al, medRxiv 2023.07.19.23292550

The most extensive dataset of chromatin conformation data with matching gene expression and chromatin accessibility from primary T cells to date

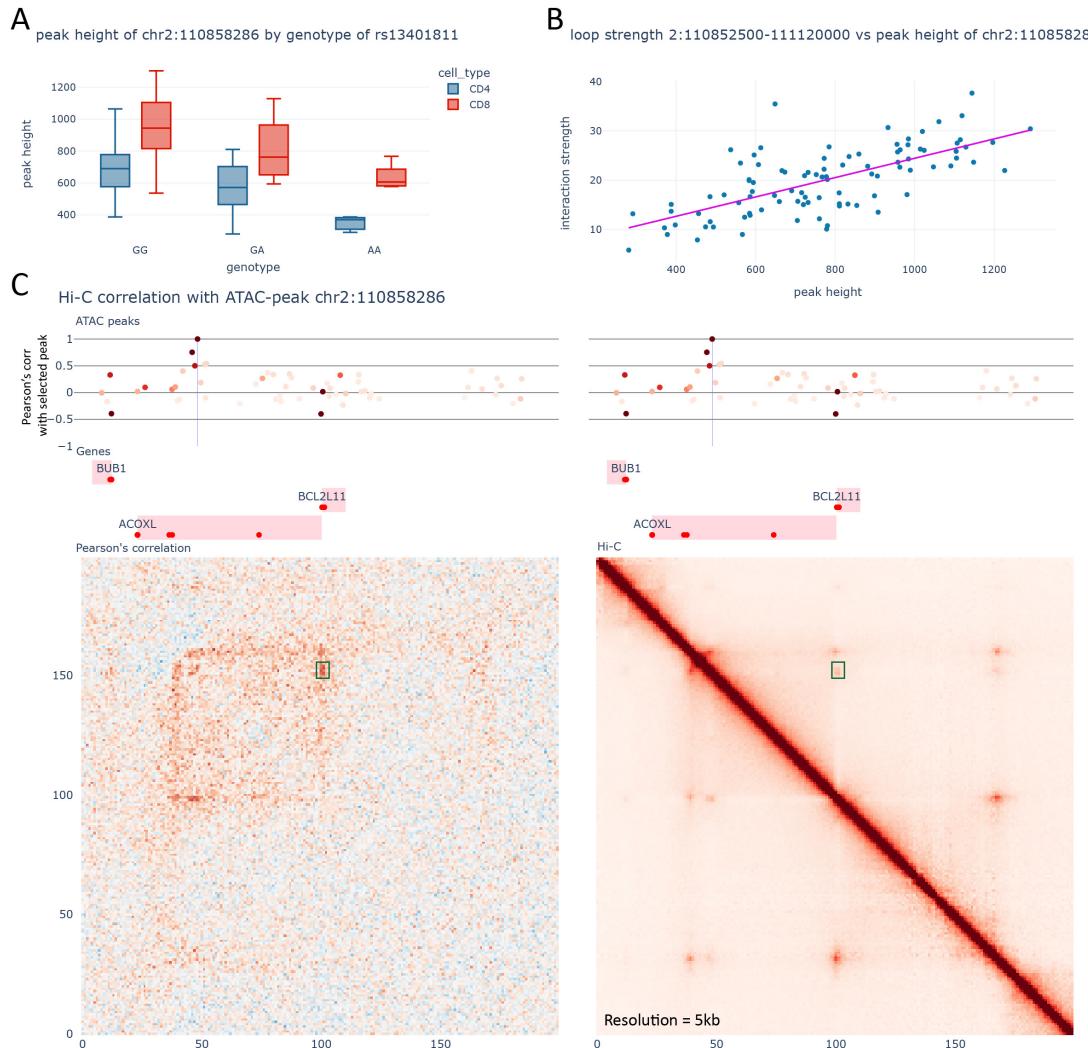




Mechanism: ALLELE SPECIFIC EFFECTS

RA locus on chr 11 tagged by lead SNP rs7943728:

- rs968567 is the only SNP in the LD block that overlaps chromatin accessibility regions (enhancers)
- Shows strong allelic imbalance accompanied by an 8.2-fold increase in chromatin accessibility for the protective (ALT) allele.
- This SNP is a strong caQTL for 4 separate ATAC-peaks, indicating that the protective allele of rs968567 causes an increased activity of other regulatory elements in the locus as well (A-D).
- This allele also results in increased chromatin interactions in, particularly those bringing the genes *FADS2*, *FADS1*, *FADS3* and *FTH1* in closer contact with the enhancer elements affected by rs968567 (E).
- Moreover, this allele is correlated with an increase in the expression of *FADS2* and *FADS1* (F-G).
- These genes are crucial for the pathogenesis of RA as they play a key role in the biosynthesis of longchain polyunsaturated fatty acids, like omega-3 and omega-6, known to influence inflammation and immune response mechanisms pivotal in the disease's pathogenesis.



A novel causal gene, *BCL2L11*, in the RA locus *ACOXL*:

- Only one SNP, rs13401811, shows strong allelic imbalance (FDR 1.86E-96), with a reduction of 40% in chromatin accessibility associated with the protective (ALT) allele.
- This SNP is also a strong caQTL for the overlapping ATAC-seq peak (A).
- This locus has been previously linked to *ACOXL* because the SNP in LD rs1554005, is a missense variant for *ACOXL*. However, the function of this gene is not especially intriguing for RA pathogenesis and is not expressed in immune cells.
- Instead, chromatin conformation data reveals a loop connecting the enhancer affected by rs13401811 to the promoter of *BCL2L11* a gene located more than 300kb downstream of the GWAS SNPs.
- The activity of this enhancer appears to be highly correlated with the strength of the loop (B-C).
- BCL2L11* has a critical role within the immune system, acting as a pro-apoptotic stimulator and modulating thymic negative selection.

All datasets are available:

- All pre-processed data, including precomputed chromatin conformation maps, correlations with gene expression, chromatin accessibility, genotype, QTL datasets and code to replicate our analysis and annotate further GWAS results at

<http://bartzabel.ls.manchester.ac.uk/orozcolab/SNP2Mechanism/>

Genetics and genomics can identify drug targets

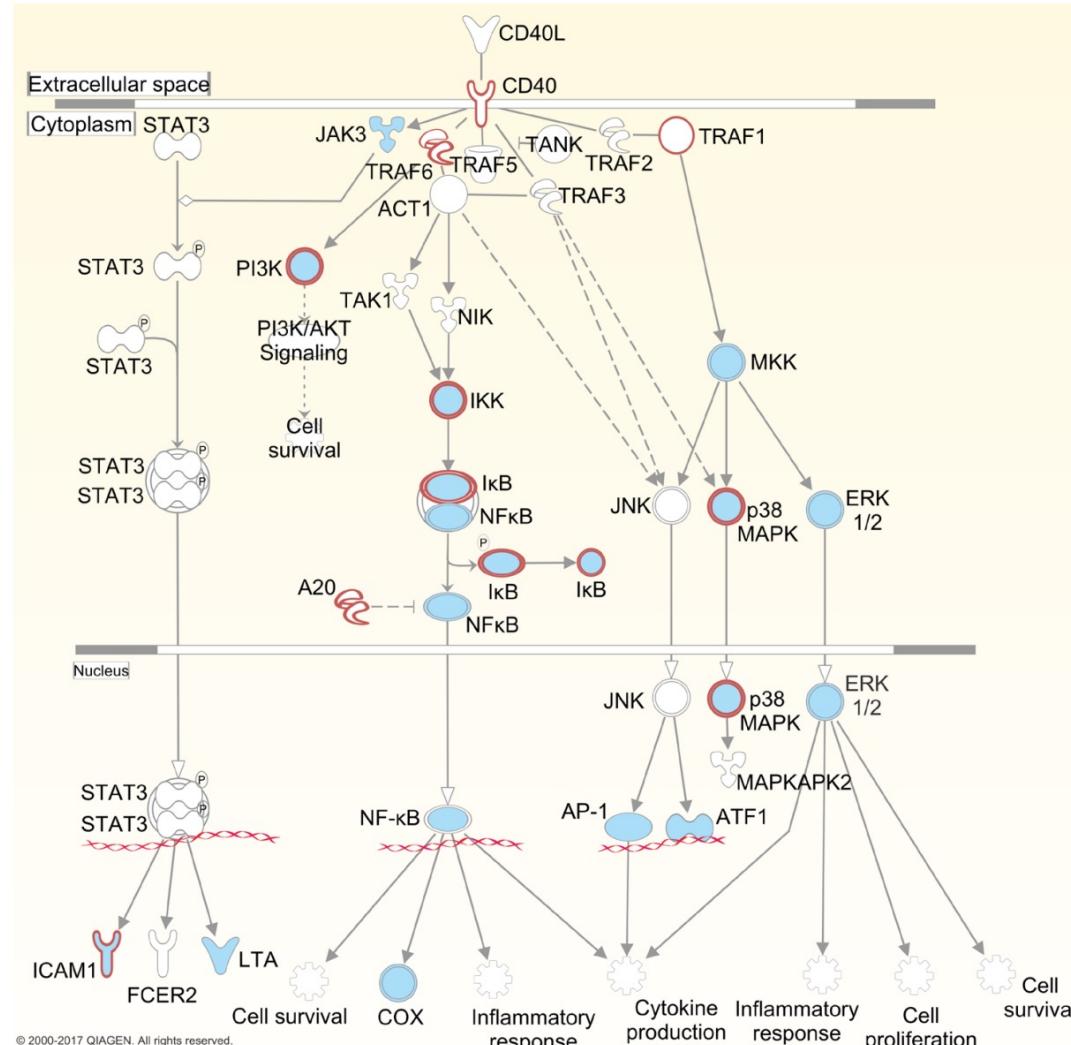
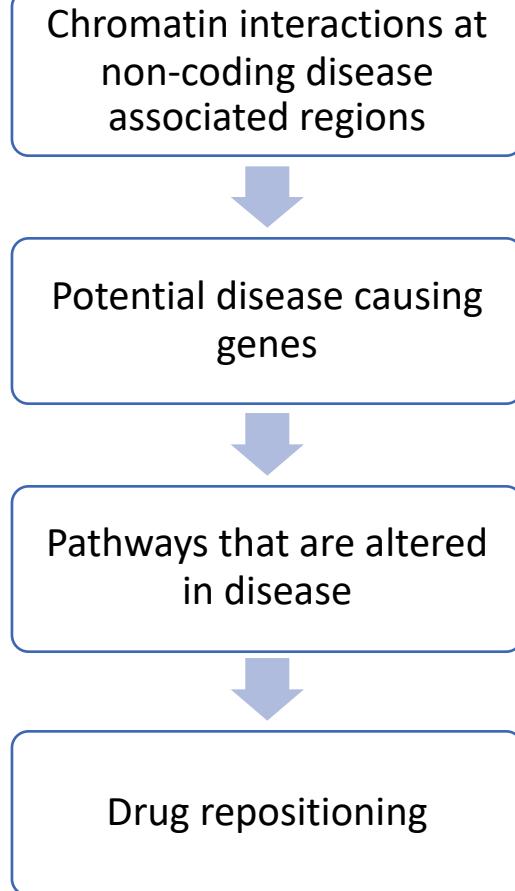


Figure 3 CD40 signalling pathway. Capture Hi-C-identified genes are outlined in red and existing drug targets are shaded in blue.

Genetics and genomics can identify drug targets

Chromatin interactions reveal novel gene targets for drug repositioning in rheumatic diseases

Paul Martin,^{1,2} James Ding,² Kate Duffus,^{1,2} Vasanthi Priyadarshini Gaddi,² Amanda McGovern,² Helen Ray-Jones,^{2,3} Annie Yarwood,^{2,3} Jane Worthington,² Anne Barton,^{2,3} Gisela Orozco²

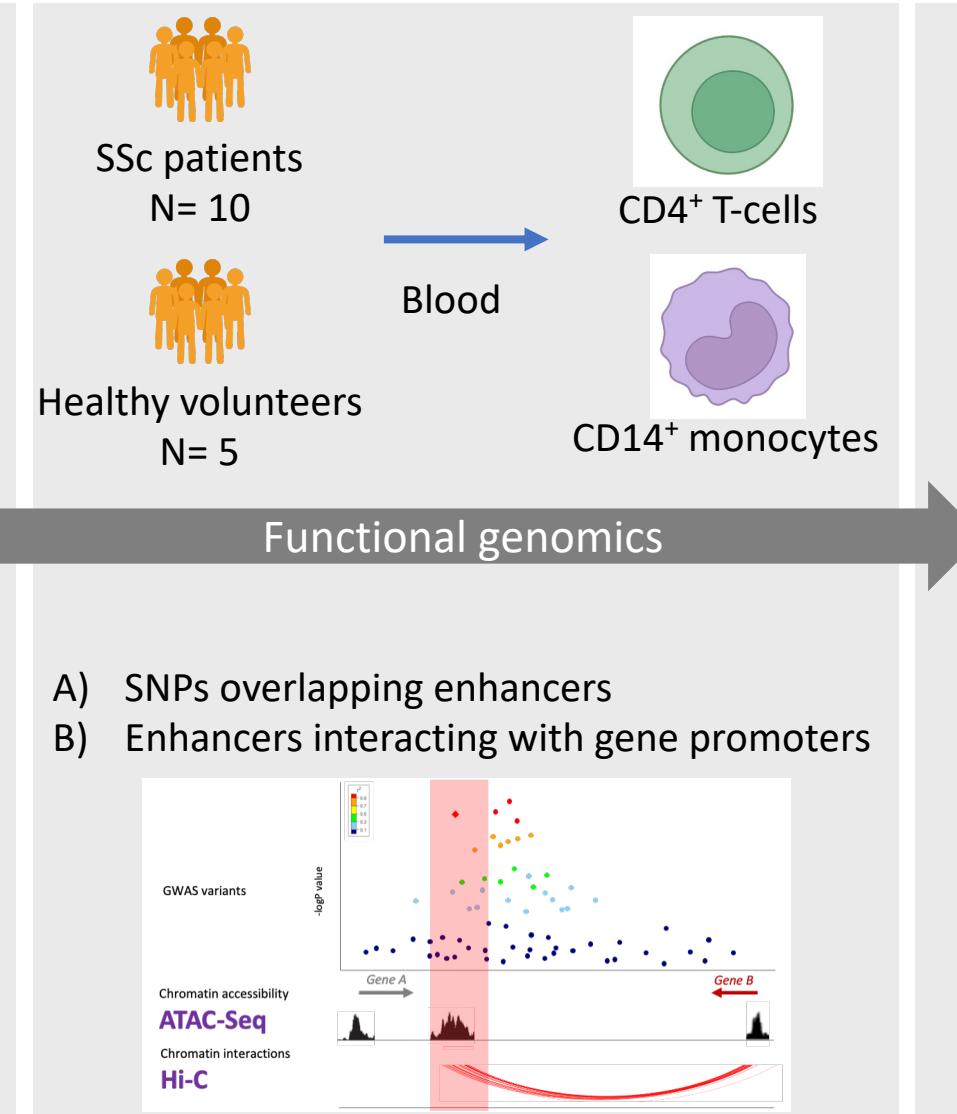
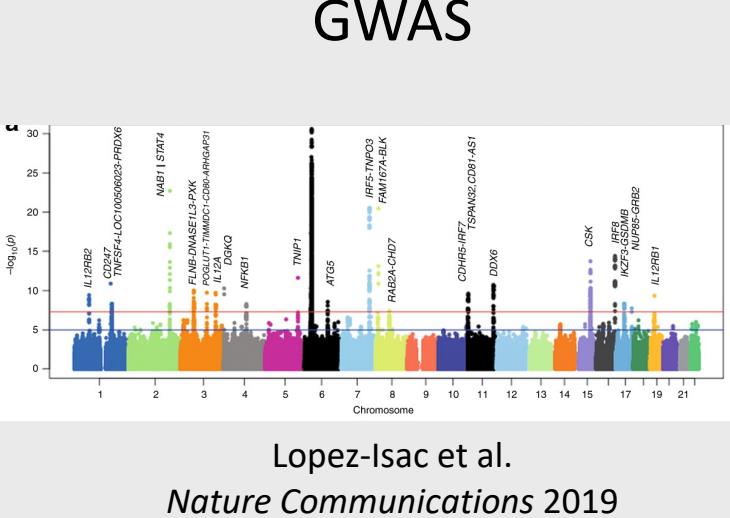
Annals of Rheumatic Diseases, 2019

Disease	Genes identified by CHi-C (n)	Genes which are existing drug targets (n)	Drugs identified (n)	Drugs currently used (n)	Drugs for potential repositioning (n)	Potential pathway targets (n)	Potential pathway drugs (n)
RA	50	13	38	8	30	283	398
PsA	9	2	2	0	2	47	87
JIA	10	2	4	0	4	205	325
All	59	14	39	8	31	307	412

CHi-C, Capture Hi-C; JIA, juvenile idiopathic arthritis; PsA, psoriatic arthritis; RA, rheumatoid arthritis.

- Drugs currently in use:
 - Biologic therapies: adalimumab, etanercept and rituximab (*FCGR2A*), sarilumab and tocilizumab (*IL6R*) and tofacitinib (*TYK2*) in RA.
- Drugs with potential for repositioning: alemtuzumab (leukaemia and multiple sclerosis), natalizumab (MS) and daclizumab (MS).

Identification of causal genes in SSc



23 loci linked to
39 new candidate
genes
and 7 previously
identified genes

Shi et al.
Arthritis and Rheumatology, 2022

Drug repurposing in SSc

46 candidate genes identified with CHi-C

21 drugs that target 13 SSc genes:

- Tocilizumab and nintedanib are already approved by FDA for its use in SSc-associated interstitial lung disease
- Tofacitinib, bosentan, methylprednisolone and mycophenolic acid: advanced clinical trials in SSc
- 15 drugs for potential repurposing, eg metformin or dimethyl fumarate

^sOnly related immune-mediated diseases were included. All clinical trials at least in completed phase III.

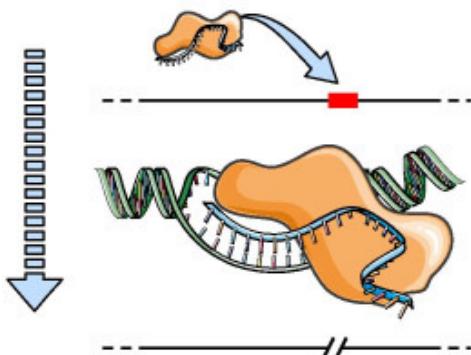
*These drugs present phase III or lower clinical trials in systemic sclerosis.

GWAS genome-wide association studies, pCHi-C promoter capture Hi-C, PPI protein-protein interaction.

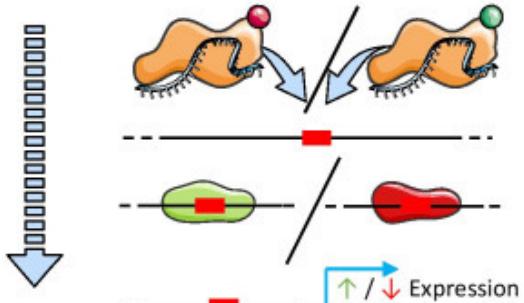
GWAS locus	pCHi-C interacting genes	Cell type with interactions	Genes in strong PPI	Targeted drug	Disease indication ^s		
<i>CD247</i>	<i>CREG1</i>	CD4+ T cells	<i>TUBB4B</i>	Colchicine	Osteoarthritis, Advanced fibrosis		
<i>FLNB-DNASE1L3-PXK</i>	<i>RPP14</i>	CD4+ T cells, CD14+ monocytes	<i>KEAP1</i>	Dimethyl Fumarate	Psoriasis, Multiple sclerosis, Disseminated sclerosis		
			<i>AGTR1</i>	Candesartan	Type 1 Diabetes		
			<i>HSPA8</i>	Forigerimod	Systemic lupus erythematosus		
			<i>IL12B</i>	Ustekinumab	Psoriasis, Crohn's disease, Ulcerative colitis		
			<i>IL1R1</i>	Anakinra	Rheumatoid arthritis		
			<i>IL23A</i>	Tildrakizumab	Psoriasis		
<i>NFKB1</i>	<i>NFKB1</i>	CD4+ T cells	<i>JAK2</i>	Tofacitinib	Systemic sclerosis, Rheumatoid arthritis, Ulcerative colitis, Interstitial lung disease, Takayasu Arteritis		
			<i>NR3C1</i>	Methylprednisolone*	Rheumatoid arthritis, Crohn's disease, Psoriatic arthritis, Ulcerative colitis, Behcet's syndrom		
			<i>UBE2D3</i>	CD4+ T cells, CD14+ monocytes	Psoriasis, Multiple sclerosis, Disseminated sclerosis		
<i>RAB2A-CHD7</i>	<i>SDCBP</i>	CD4+ T cells	<i>IMPDH1</i>	Mycophenolic acid*	Systemic lupus erythematosus, Immunosuppression		
			<i>TUBB4B</i>	Colchicine	Osteoarthritis, Advanced fibrosis		
			<i>CHD7</i>	<i>PPARG</i>	Crohn's disease, Ulcerative colitis		
<i>CSK</i>	<i>DDX6</i>	CD4+ T cells, CD14+ monocytes	<i>SIPR3</i>	Fingolimod	Multiple sclerosis, Disseminated sclerosis		
			<i>CSK</i>	CD4+ T cells, CD14+ monocytes	<i>FLT4</i>	Nintedanib	Systemic sclerosis, Idiopathic pulmonary fibrosis, Interstitial lung disease
			<i>COX5A</i>	CD4+ T cells, CD14+ monocytes	<i>NDUFB10</i>	Metformin	Type 1 Diabetes, Type 2 Diabetes
<i>IKZF3-GSDMB</i>	<i>IKZF3</i>	CD4+ T cells	<i>JAK1</i>	Baricitinib	Rheumatoid arthritis		
			<i>JAK3</i>	Upadacitinib	Rheumatoid arthritis		
			<i>IL2RA</i>	Basiliximab	Type 1 Diabetes		
<i>ERBB2</i>	<i>IL12RB1</i>	CD4+ T cells, CD14+ monocytes	<i>IL6R</i>	Tocilizumab	Systemic sclerosis, Rheumatoid arthritis, Juvenile idiopathic arthritis, Giant cell arteritis		
			<i>JAK</i> kinases	Tofacitinib	Systemic sclerosis, Rheumatoid arthritis, Ulcerative colitis, Interstitial lung disease, Takayasu Arteritis		
			<i>ADRA1B</i>	Epinephrine	Crohn's disease		
<i>PIK3R2</i>	<i>PIK3R2</i>	CD4+ T cells	<i>AGTR1</i>	Candesartan	Type 1 Diabetes		
			<i>EDNRA</i>	Bosentan	Systemic sclerosis, Idiopathic pulmonary fibrosis, Pulmonary artery hypertension		
			<i>JAK1</i>	Baricitinib	Rheumatoid arthritis		
<i>RAB3A</i>	<i>IL12RB1</i>	CD4+ T cells, CD14+ monocytes	<i>JAK</i> kinases	Tofacitinib	Systemic sclerosis, Rheumatoid arthritis, Ulcerative colitis, Interstitial lung disease, Takayasu Arteritis		
			<i>PDGFRB</i>	Nintedanib	Systemic sclerosis, Idiopathic pulmonary fibrosis, Interstitial lung disease		
			<i>HSPA8</i>	Forigerimod	Systemic lupus erythematosus		

Validation: CRISPR-Cas9 genome editing

a) Small deletions are the predominant mutagenic consequence of WT Cas9 targeted using a single gRNA



b) dCas9 fusion proteins can be used to activate / repress target chromatin regions



c) Both approaches can be scaled up for screens

i. GWAS locus



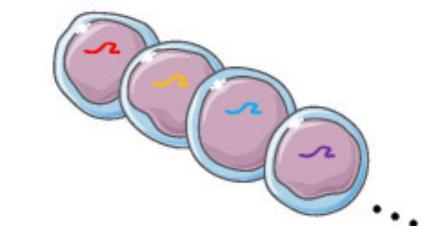
ii. Tiled gRNAs



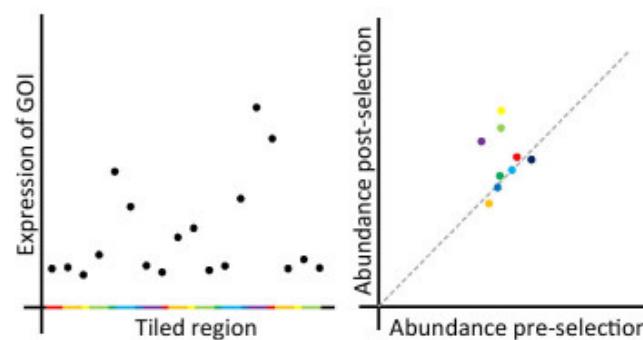
iii. Lentiviral particles



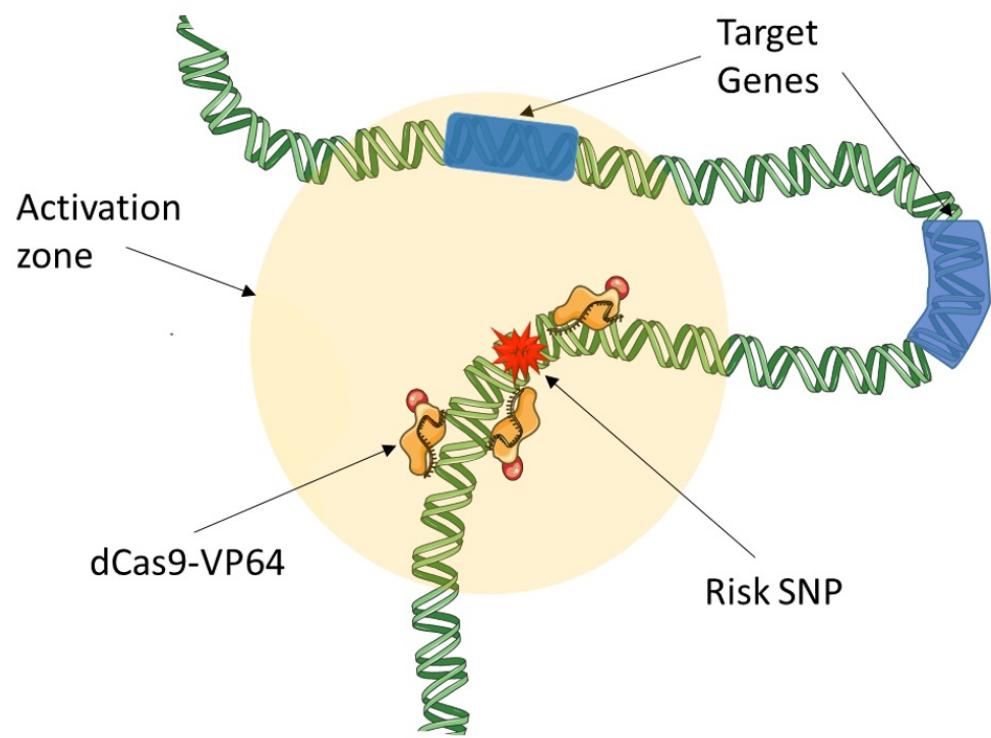
iv. Transduced cells



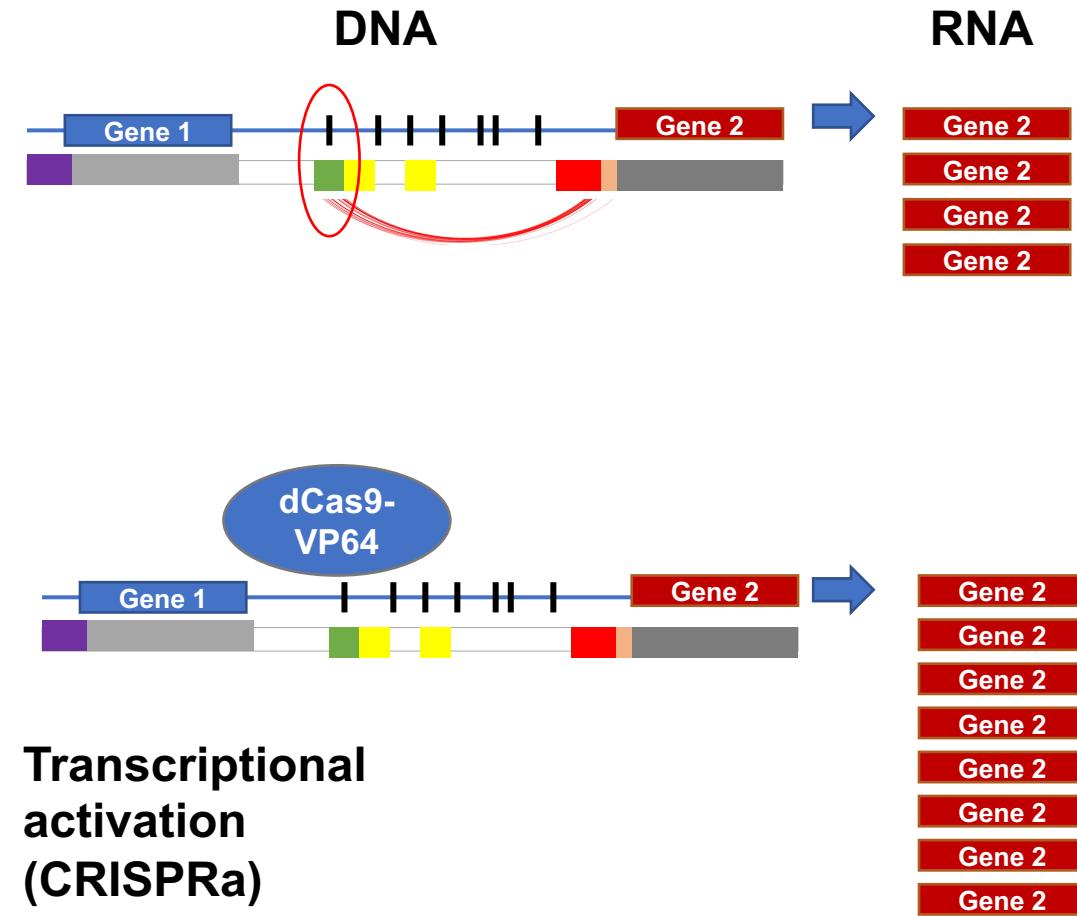
v. scRNAseq/selection and DNA sequencing



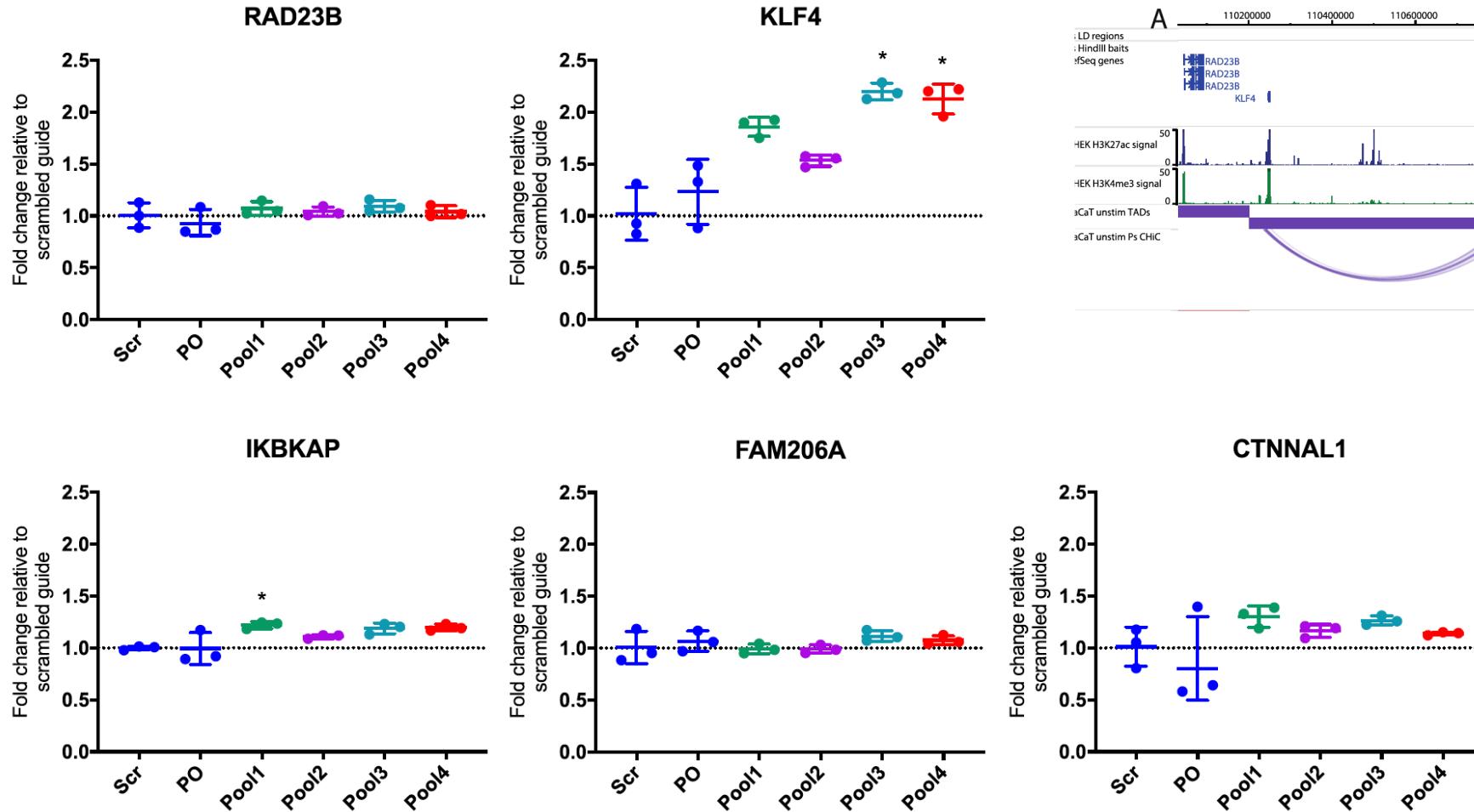
dCas9-mediated CRISPR activation (CRISPRa)



Antonios Frantzeskos

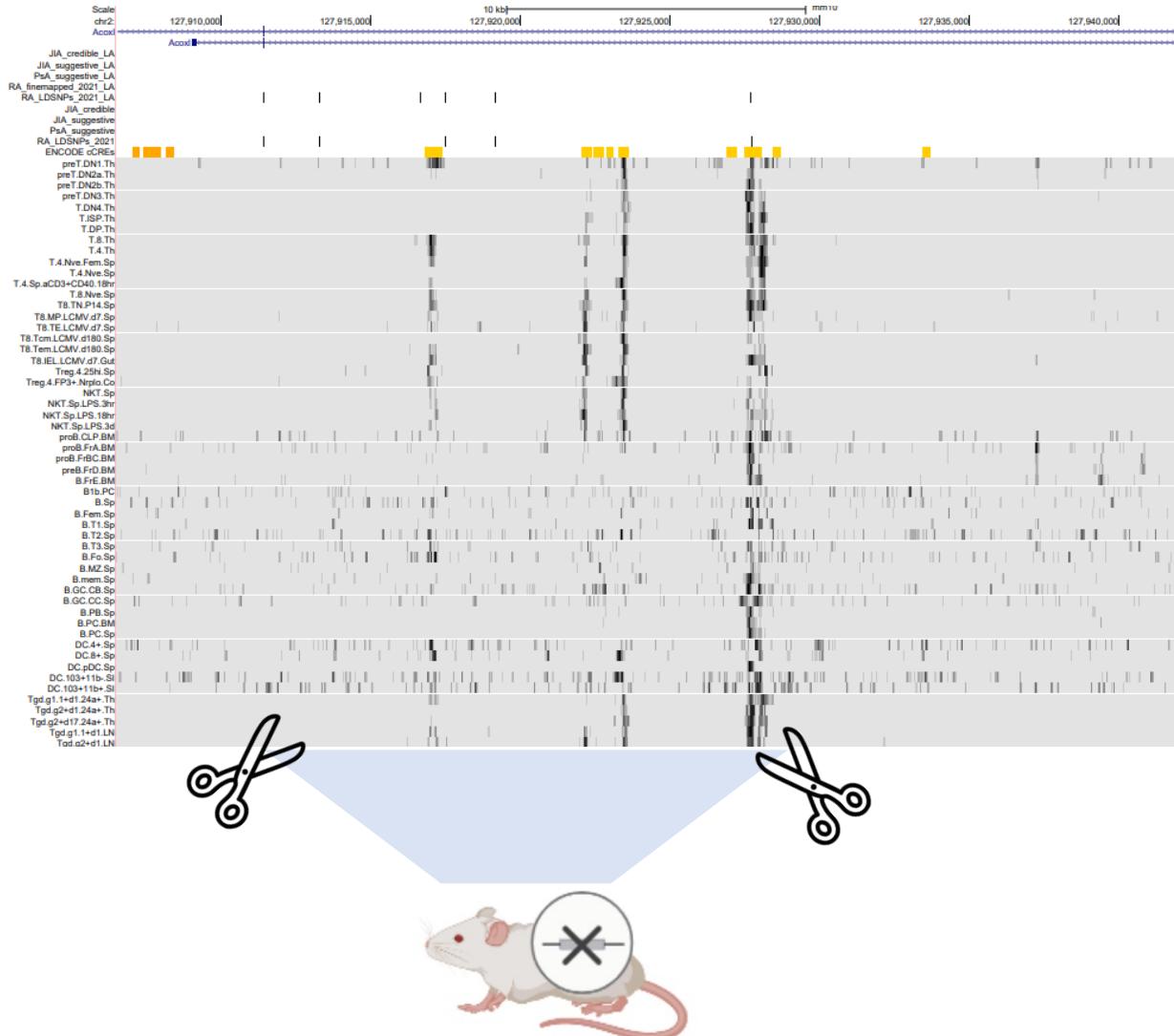


CRISPRa shows that a psoriasis-associated enhancer in 9q31 regulates *KLF4* expression but not other genes in the vicinity



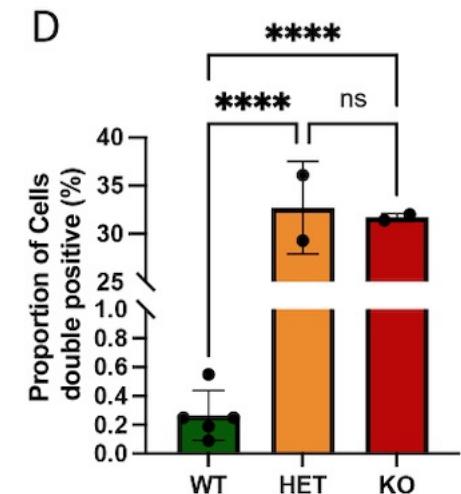
Expression of genes within the 9q31.2 locus in HaCaT cells expressing dCas9-P300

CRISPR Knock out mouse for *ACOXL* enhancer harboring RA risk variants: effect on the distant *BCL2L11* gene



RA associated SNPs
Human enhancers

Mouse enhancers



Marked increase in the presence of a population of CD4/CD8 double-positive T cells in the spleen. This indicates a possible evasion of central tolerance mechanisms in T lymphocytes, which suggests that **apoptosis is impaired (*BCL2L11* is involved in regulation of apoptosis)**.

Summary

- GWAS have not reached their full potential for translation to patient benefit because most disease SNPs are non-coding and their biological impact is challenging to assess
- Most disease SNPs are thought to affect expression of **disease causing genes** through dysregulation of **enhancers** in specific disease **cell types**
- Functional genomics techniques can be used to identify disease **enhancers** (DNase-seq, ATAC-seq, ChIP-Seq), **causal genes** (eQTLs, Hi-C) and disease **cell types** (co-localization and data integration)
- Most of these datasets are publicly available and can be used to identify biological disease mechanism and identify drug targets

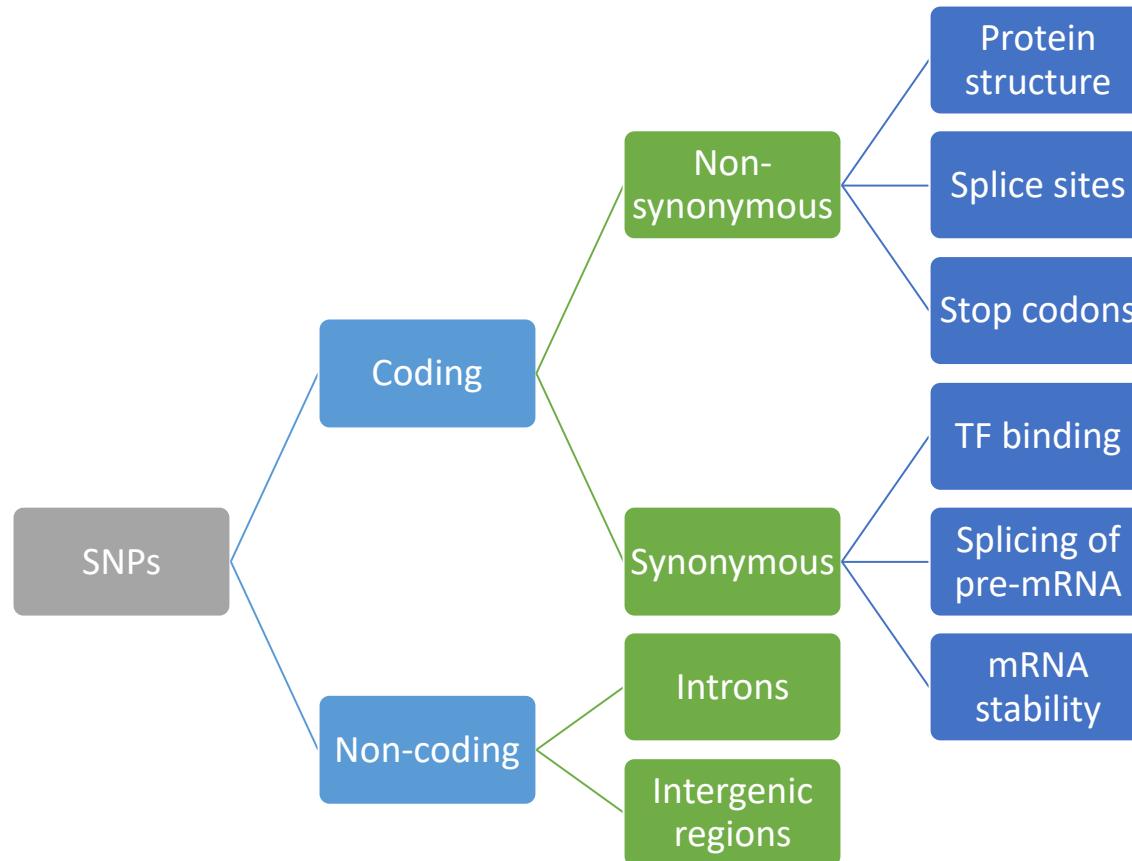
Wellcome Trust Genome Campus Advanced Course
Genetic Analysis of Population-based Association Studies

Session: Post-GWAS Analysis - Part 2: Functional genomics

Online resources to functionally annotate GWAS variants

How can I use publicly available functional genomics data to help me
interpret my GWAS results?

Types of genetic variants: The Ensembl Variant Effect Predictor (VEP)



Good for prediction of
effect of protein CODING variants

<https://www.ensembl.org/Tools/VEP>

 [Login/Register](#)

[BLAST/BLAT](#) | [VEP](#) | [Tools](#) | [BioMart](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

 [Search all species...](#) 

VEP ▾

Web Tools

- Web Tools
- BLAST/BLAT
- Variant Effect Predictor**
- Linkage Disequilibrium Calculator
- Variant Recoder
- File Chameleon
- Assembly Converter
- ID History Converter
- VCF to PED Converter
- Data Slicer
- Post-GWAS

Variant Effect Predictor ?

New job Clear form | Close

Species:  Homo_sapiens X

Assembly: GRCh38.p13
[Add/remove species](#)

If you are looking for VEP for Human GRCh37, please go to [GRCh37 website](#).

Name for this job (optional):

Input data: **Either paste data:**

```
rs1156485833
rs1258750482
rs867704559
```

Run instant VEP for current line >

Examples: [Ensembl default](#), [VCF](#), [Variant identifiers](#), [HGVS notations](#), [SPDI](#)

Or upload file: Choose file No file chosen

Functional annotation of non-coding variants: Haploreg

Ward and Kellis. Nucleic Acid Research (2011)

- HaploReg is a tool for exploring annotations of the noncoding genome at variants on haplotype blocks.
- Integrated datasets:
 - LD information from the 1000 Genomes Project
 - Chromatin states (enhancers, promoters etc) and protein binding (TFs etc)
 - Roadmap Epigenomics
 - ENCODE
 - Sequence conservation across mammals
 - Effect of SNPs on regulatory motifs
 - Effect of SNPs on expression from eQTL studies
 - GTEx analysis V6, the GEUVADIS analysis, and 10 other studies

<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>

RegulomeDB

Boyle et al. Genome Research (2012)

- RegulomeDB is a database that annotates SNPs with known and predicted regulatory elements in the intergenic regions
 - Transcription factor binding sites
 - Position-Weight Matrix for TF binding (PWM)
 - DNase Footprinting
 - Open Chromatin
 - Chromatin States
 - eQTLs
 - Validated functional SNPs

<https://www.regulomedb.org/regulome-search>

RegulomeDB

Boyle et al. Genome Research (2012)

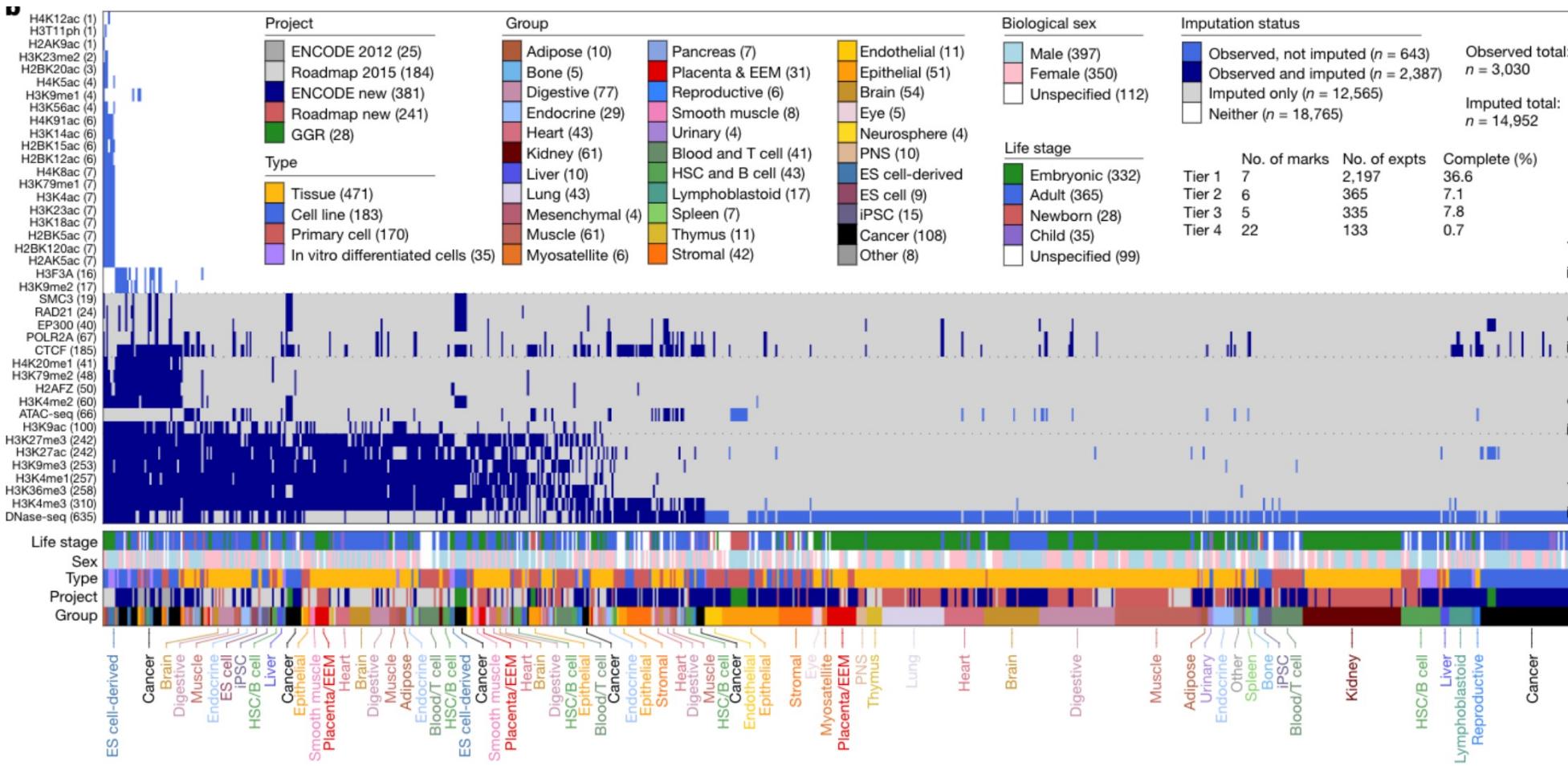
- RegulomeDB ranks SNPs with a scoring system that predicts how likely they are to be functional

Score	Supporting data
1a	eQTL + TF binding + matched TF motif + matched DNase Footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase Footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding / DNase peak
2a	TF binding + matched TF motif + matched DNase Footprint + DNase peak
2b	TF binding + any motif + DNase Footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Motif hit
7	Other

EpiMap

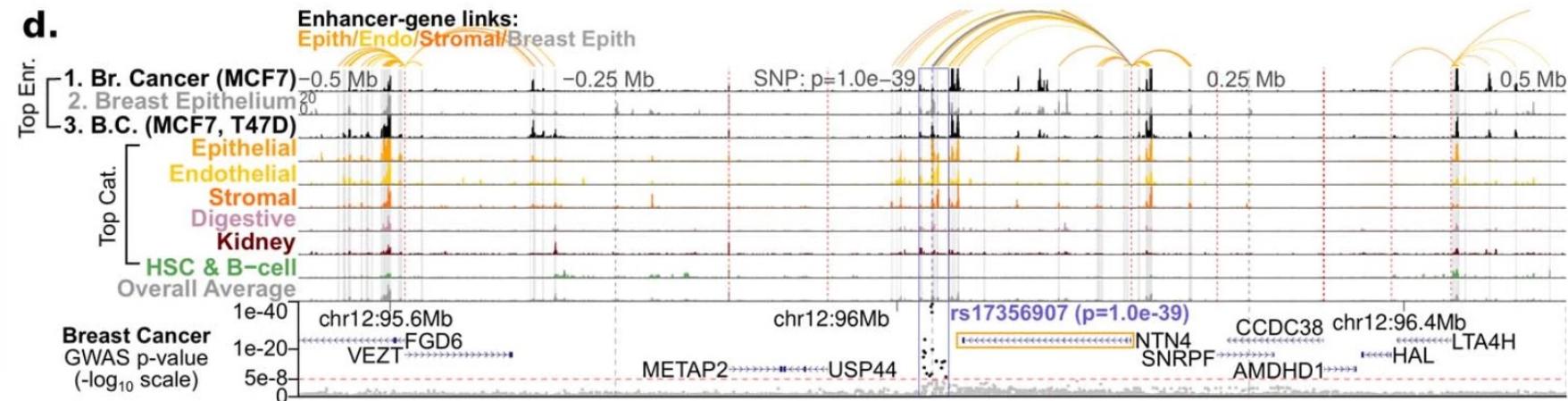
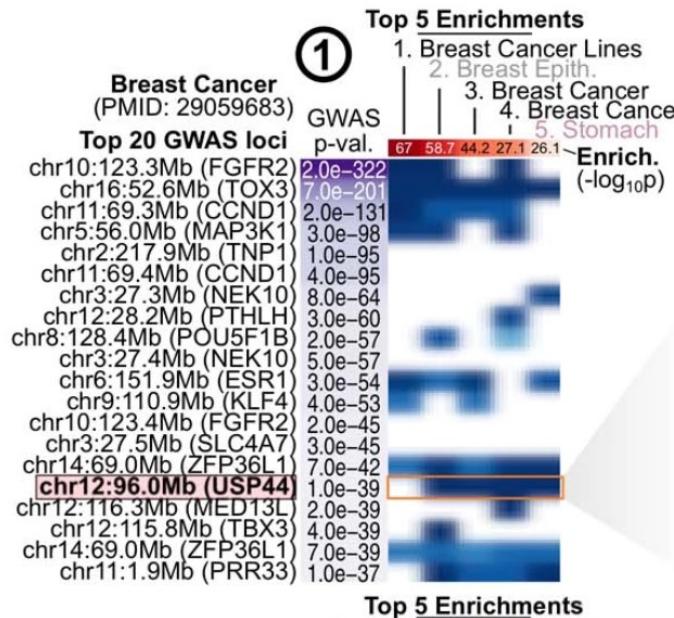
Boix et al. Nature Genetics (2021)

- Compendium of 10,000 epigenomic maps including 800 samples
- These datasets were used to define chromatin states, high-resolution enhancers, enhancer modules, upstream regulators and downstream target genes.
- Annotation of 30,000 genetic loci that are associated with 540 traits from the GWAS catalog, predicting trait-relevant tissues, putative causal nucleotide variants in enriched tissue enhancers and candidate tissue-specific target genes for each.



EpiMap

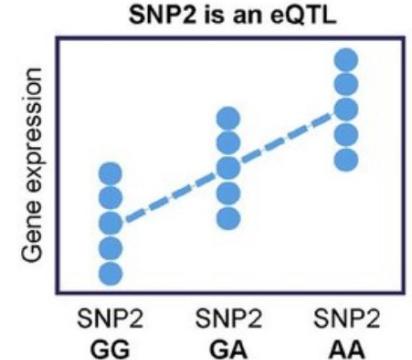
<http://comppbio.mit.edu/epimap>



<https://epilogos.altius.org/>

Boix et al. Nature Genetics (2021)

eQTLs: GTEx



- The Genotype-Tissue Expression (GTEx) project is an ongoing effort to build a comprehensive public resource to study tissue-specific gene expression and regulation
- RNA-Seq data available from 54 non-diseased tissue sites across nearly 1000 individuals
- The GTEx Portal provides open access to data including gene expression, QTLs, and histology images

<https://gtexportal.org/home/>



eQTL Catalogue

Expression and splicing QTLs recomputed from public datasets

<https://www.ebi.ac.uk/eqtl>

Kerimov et al. Nature Genetics (2021)

- There are many eQTL studies that have published their summary statistics, but technical differences between datasets are a barrier to their widespread use.
- The eQTL Catalogue aims to provide uniformly processed gene expression and splicing QTLs from all available public studies on human.

RNA-seq studies

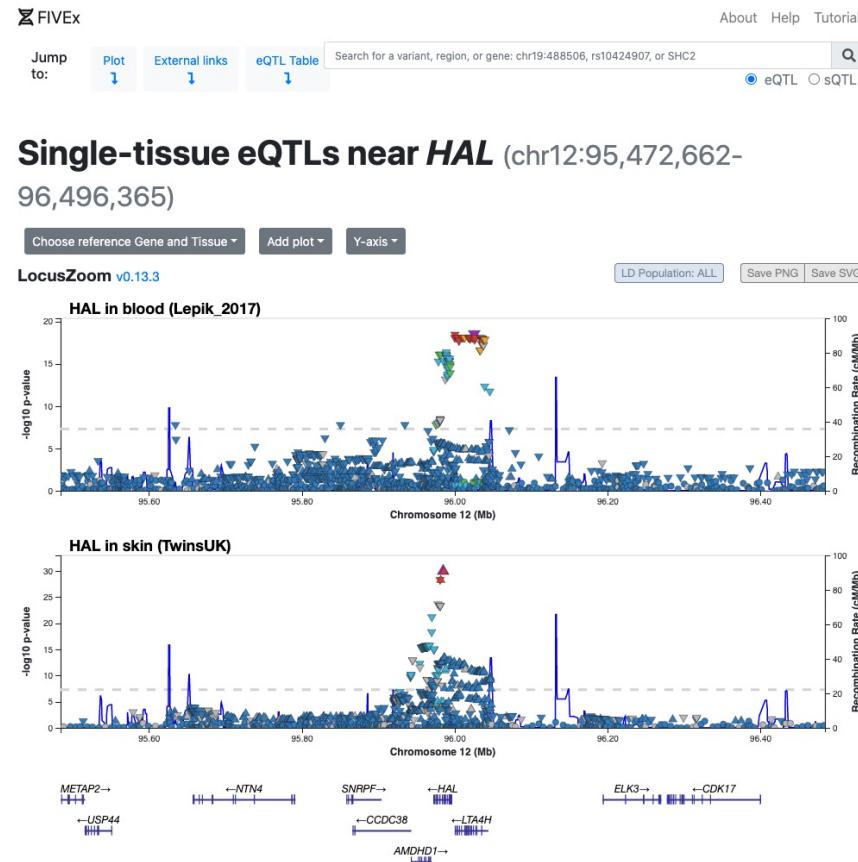
Study	Cell types or tissues	Conditions	Sample s	Donors
Alasoo_2018	macrophages	IFNg, Salmonella, IFNg + Salmonella	336	84
BLUEPRINT	monocytes, neutrophils, CD4+ T cells		554	197
GENCORD	LCLs ¹ , fibroblasts, T cells		560	195
GEUVADIS	LCLs ¹		445	445
HipSci	iPSCs ²		322	322
Nedelec_2016	macrophages	Listeria, Salmonella	493	168
Quach_2016	monocytes	LPS, Pam3CSK4, R848, IAV	969	200
Schwartzentruber_2018	sensory neurons		98	98
TwinsUK	adipose, LCLs ¹ , skin, blood		1364	433
van_de_Bunt_2015	pancreatic islets		117	117
Schmiedel_2018	15 immune cell types	α CD3+ α CD28 (4h)	1331	91
BrainSeq	brain (DLPFC ³)		484	484
ROSMAP	brain (DLPFC ³)		576	576
Lepik_2017	blood		491	491
FUSION	adipose, muscle		559	302
GTEX(v8)	49 tissues		15178	838
CAP	LCLs ¹	statin	296	148
Peng_2018	placenta		149	149
PhLiPS	iPSCs ² , hepatocytes		168	87
iPScore	iPSCs ²		107	107
CommonMind	brain (DLPFC ³)		590	590
Braineac2	brain (putamen, substantia nigra)		167	110
Steinberg_2020	synovium, cartilage		210	73
Young_2019	microglia		104	104

→ DICE study

Microarray studies

Study	Cell types or tissues	Condition s	Samples	Donors
CEDAR	CD4+ and CD8+ T cells, monocytes, neutrophils, platelet, B cells, ileum, rectum, transverse colon		2388	322
Fairfax_2012	B cells		282	282
Fairfax_2014	monocytes	IFN24, LPS2, LPS24	1372	424
Kasela_2017	CD4+ and CD8+ T cells		553	297
Naranhai_2015	neutrophils		93	93

- The eQTL Catalogue gene expression and splicing QTLs can be visualised with the FIVE_x eQTL browser: <https://fivex.sph.umich.edu/>

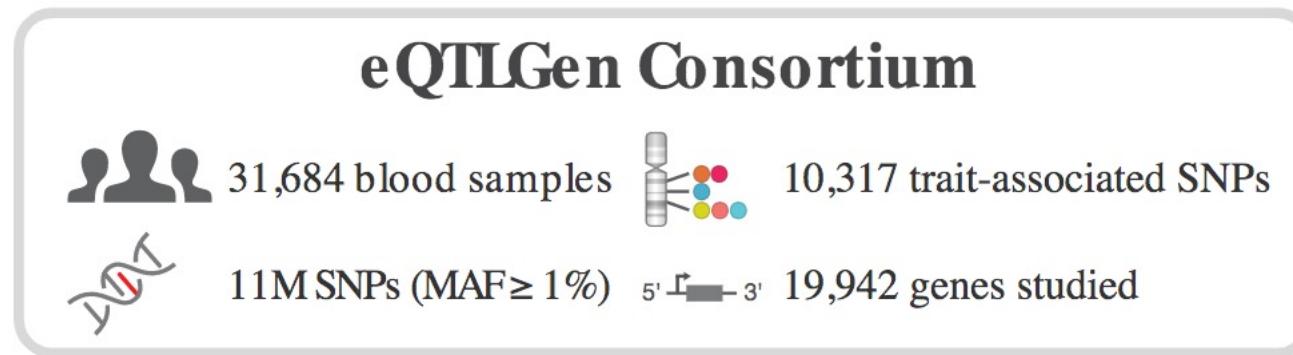


- Data can be downloaded from their FTP site <ftp://ftp.ebi.ac.uk/pub/databases/spot/eQTL>

eQTLGen

Vosa et al. Nature Genetics (2021)

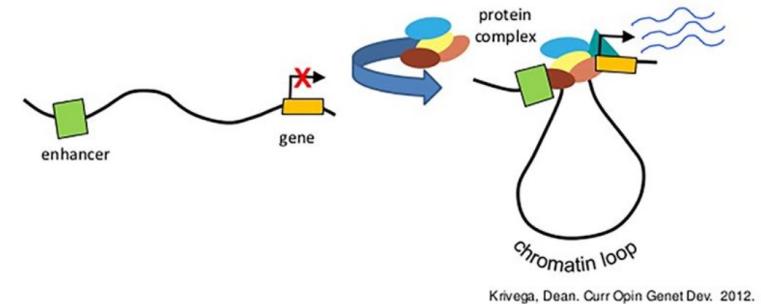
- Includes cis- and trans-expression quantitative trait locus (eQTL) using blood-derived expression from 31,684 individuals



- All data available to download at <https://www.eqtlgen.org>

Capture Hi-C plotter (CHiCP)

Schofield et al. Bioinformatics (2016)

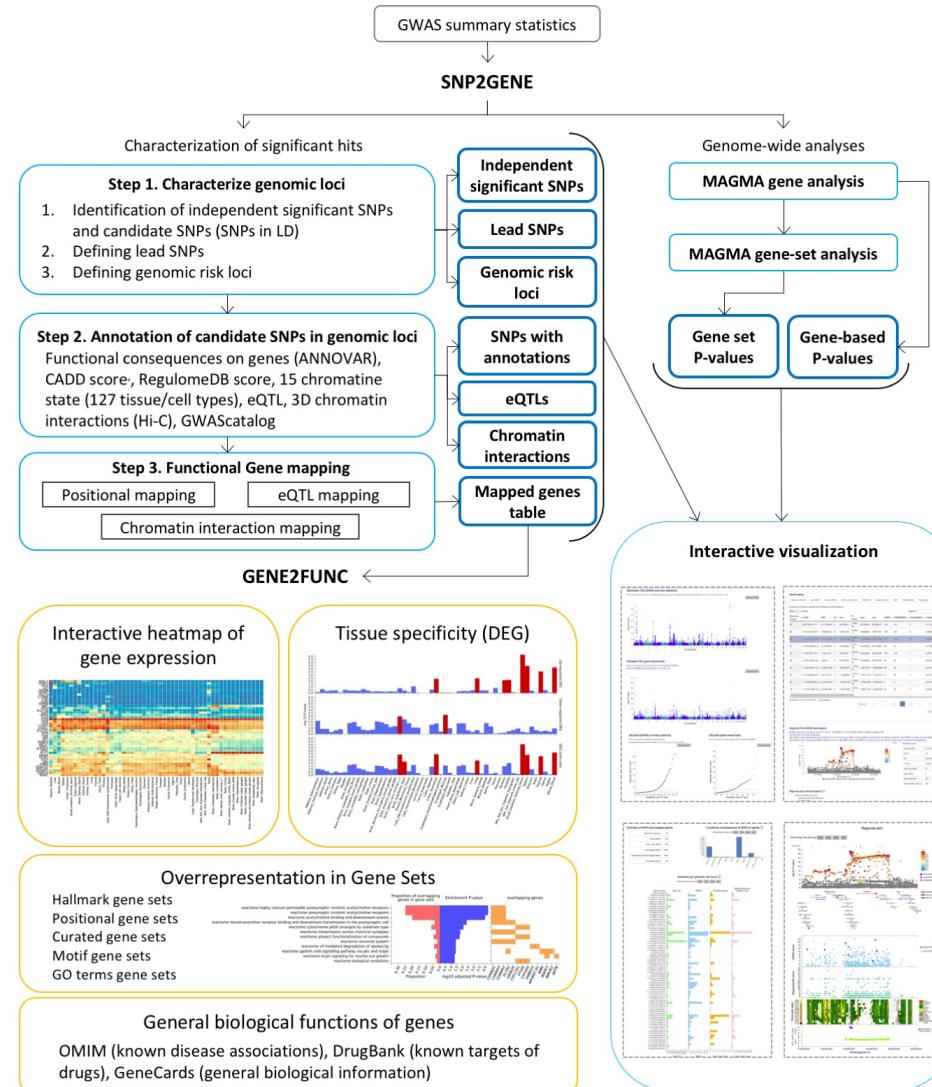


- CHiCP is a web application for visually integrating GWAS data with promoter capture Hi-C data
- It supports the analysis described in:
 - Javierre et al (2016) Cell - 17 human primary hematopoietic cell types
 - Mifsud et al (2015) Nat Genet – CD34, GM12878
 - Miguel-Escalada et al (2015) Nat Commun – Pancreatic islets
 - Choy et al (2018) Nat Commun - hESC Derived Cardiomyocytes
- Incorporates autoimmune focused population genetic data (GWAS and ImmunoChip) from Immunobase
 - There is the option of adding your own data

<https://www.chicp.org/>

FUMA GWAS

Functional Mapping and Annotation of Genome-Wide Association Studies



Watanabe et al. *Nat. Commun.* 8:1826. (2017).
<https://www.nature.com/articles/s41467-017-01261-5>

<https://fuma.ctglab.nl/>

<https://www.ensembl.org/Tools/VEP>

<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>

<https://www.regulomedb.org/regulome-search>

<http://compbio.mit.edu/epimap>

<https://genetics.opentargets.org/>

<https://www.encodeproject.org/>

<https://gtexportal.org/home/>

<https://www.ebi.ac.uk/eqtl/>

<https://www.eqtlgen.org>

<https://dice-database.org/>

<https://www.chicp.org/>

<https://fuma.ctglab.nl/>