

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

Post-GWAS Analysis: Functional genomics and omics integration

Gisela Orozco

Professor of Functional Genomics
University of Manchester, UK

6th September 2023

Session overview

9.30 – 10.30	Functional genomics: Limitations of GWAS How functional genomics can help overcome some of these limitations Types of functional genomics data Large scale functional genomics studies Omics integration Other types of omics data and technologies Validation: CRISPR, MPRAs
10.30 – 11.00	Coffee Break
11.00 – 11.50	Examples of applications to complex diseases SSc OA PsA JIA
(11.50 – 12.00)	Short break
12.00 – 12.45	Online resources and questions
12.45 – 13.15	Wrap up
13.15 – 14.00	Lunch
14.00	Taxies to train station

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.

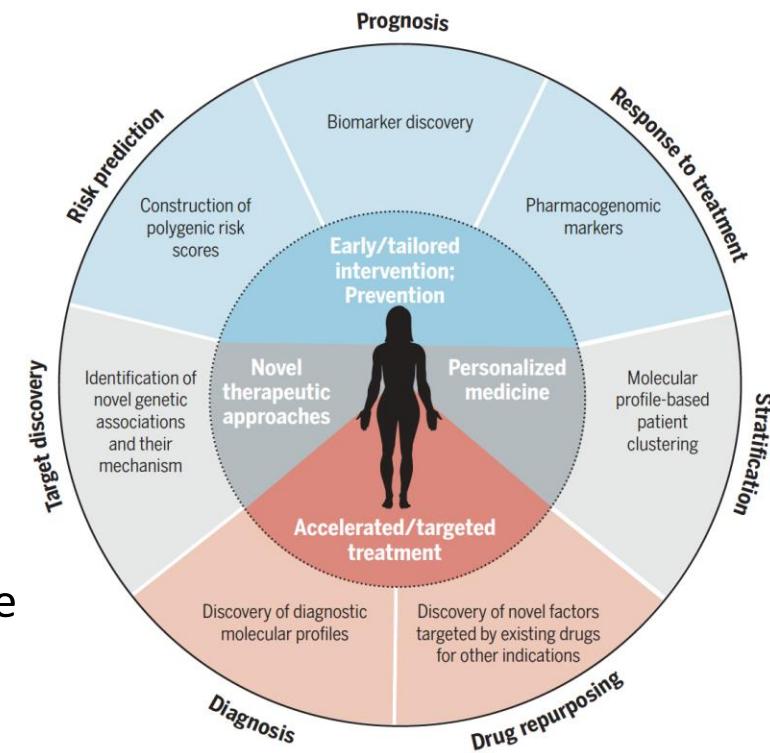
Associations of HLA-C*06:02 with biologic therapy response in psoriasis (Dand, 2018) and *HLA-DRB1* with severity, mortality, and treatment response to biologic drugs in RA (Viatte, 2015).

- **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

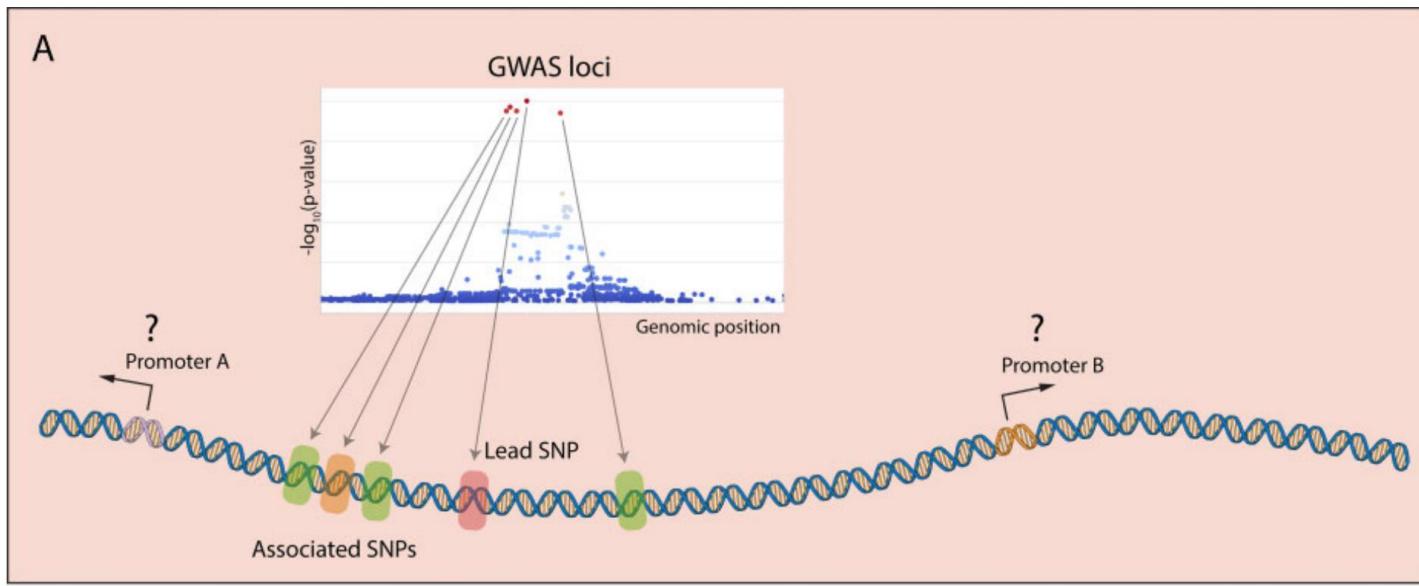
<https://www.ebi.ac.uk/gwas/home>

Accessed 12/08/2024



6,960 publications and 668,514 associations

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



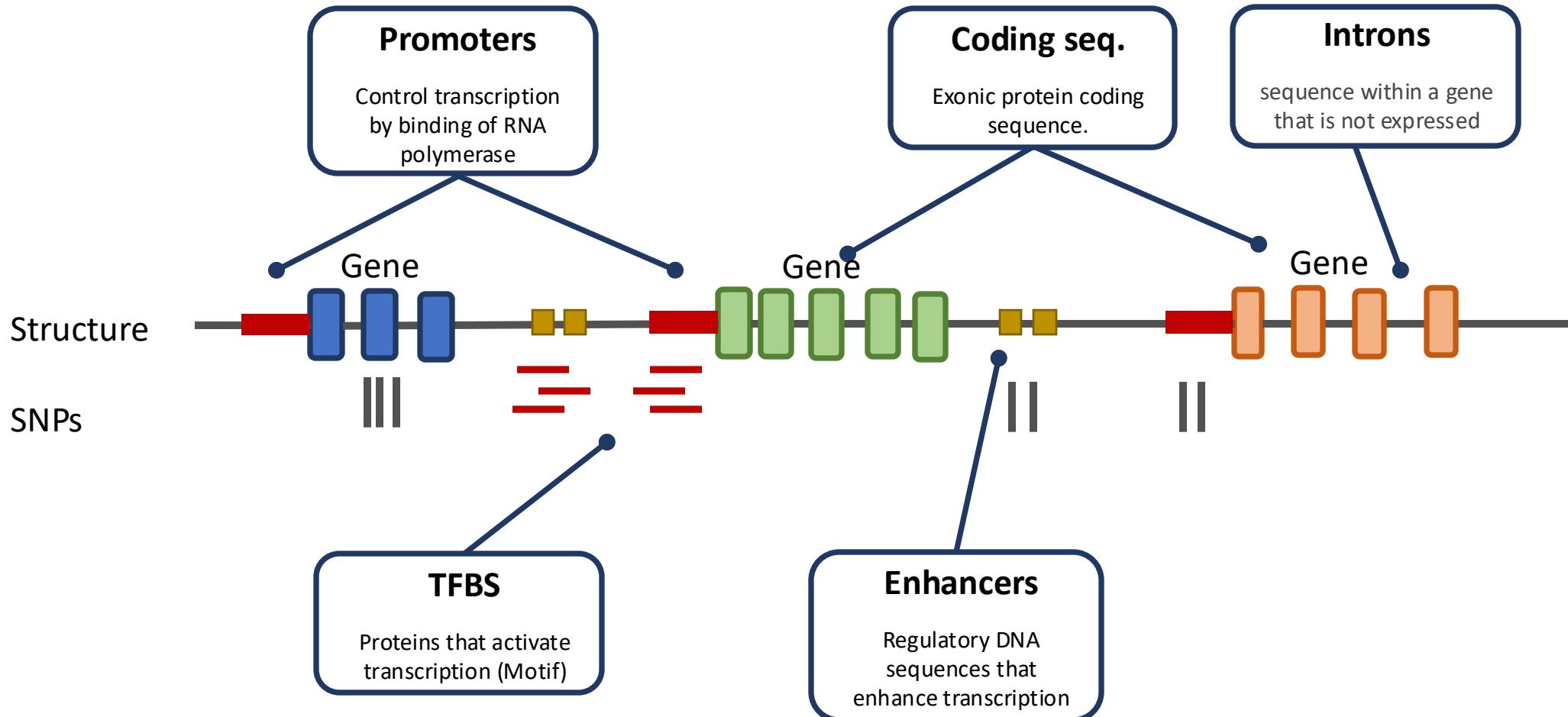
Causal variants?

Function?

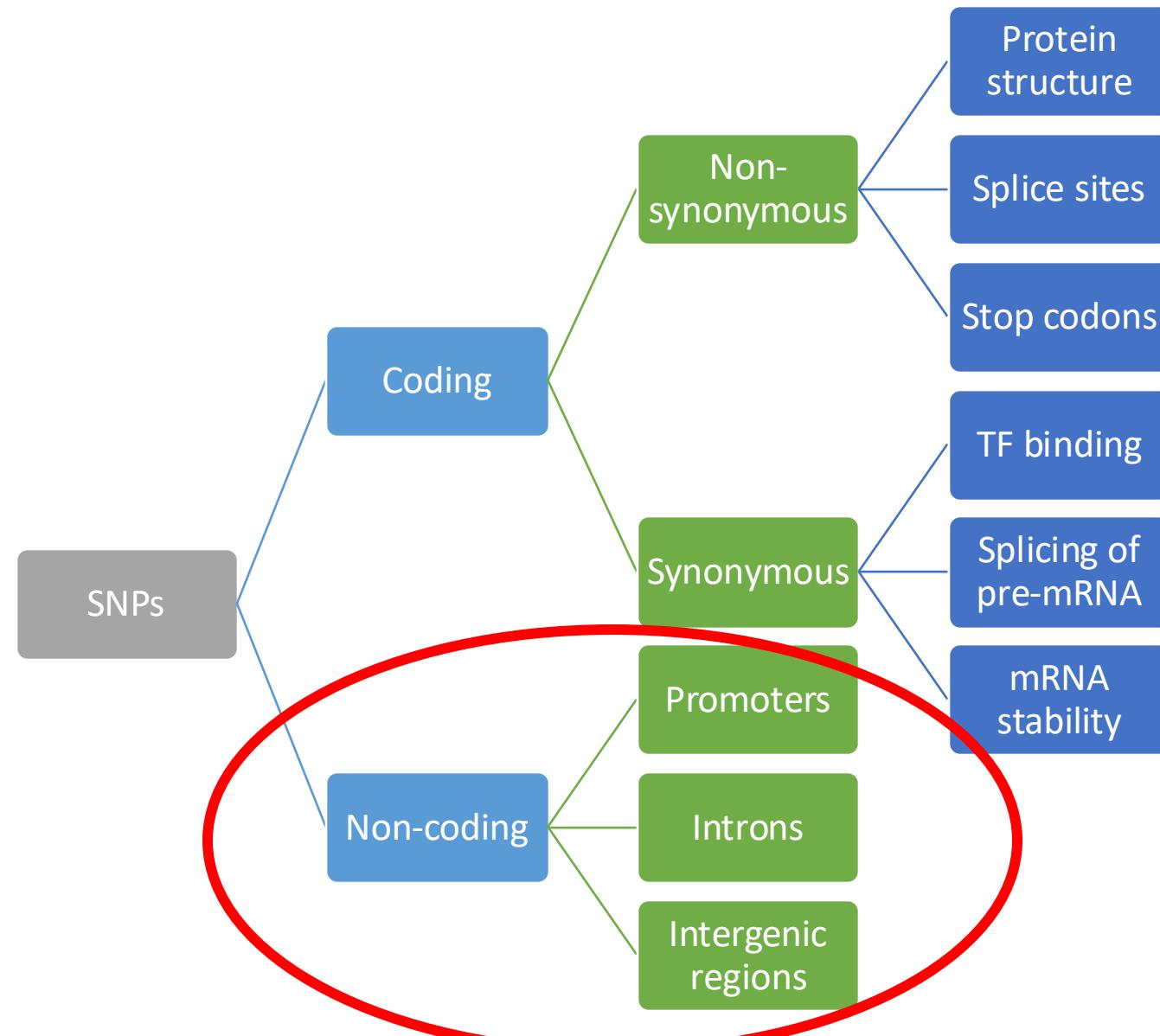
Causal genes?

Causal cell types?

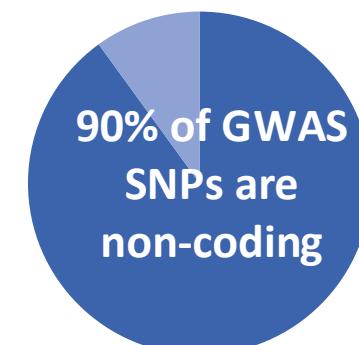
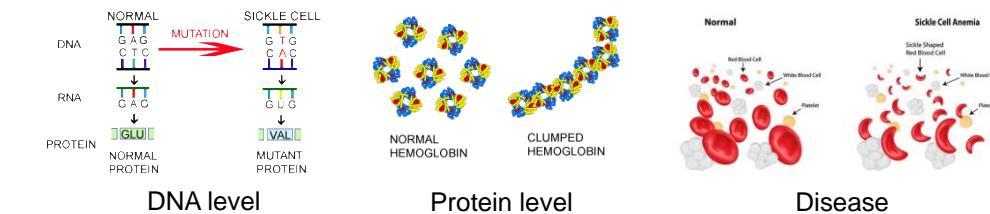
Gene structure and regulatory elements



How can SNPs impact biological function?



Expected: disease variants result in defective proteins

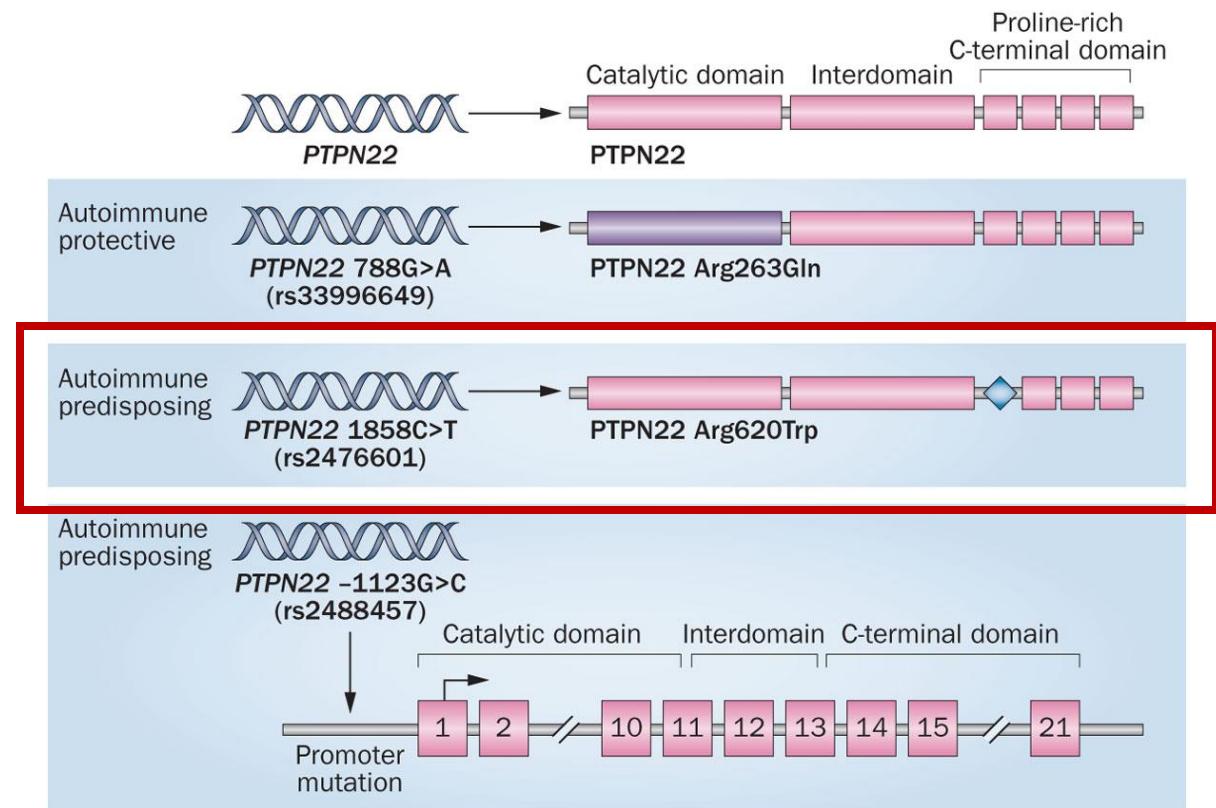


Function???

Coding SNP example

rs2476601

- *PTPN22*: tyrosine phosphatase
- Non-synonymous SNP strongly associated with autoimmunity
- Arginine substituted for tryptophan
- This is mapped to the proline-rich C-terminal domain



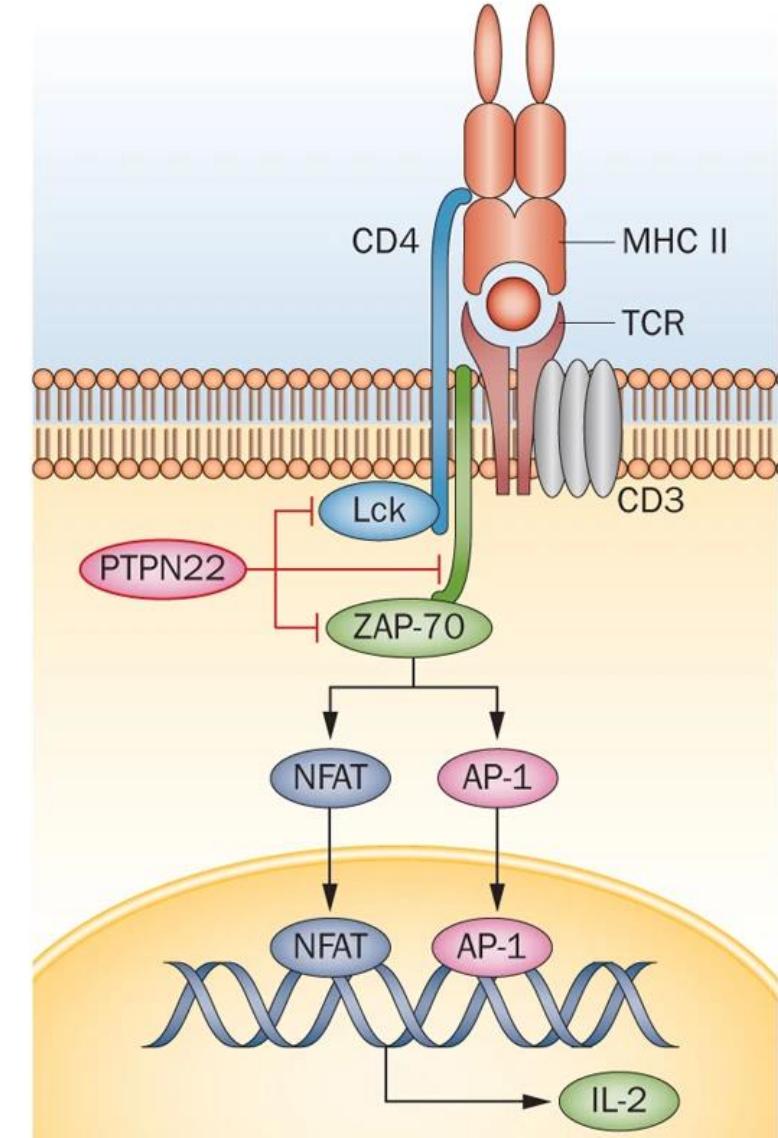
Stanford, S. M. & Bottini, N. (2014) *PTPN22*: the archetypal non-HLA autoimmunity gene
Nat. Rev. Rheumatol. doi:10.1038/nrrheum.2014.109

Function of PTPN22:

- Belongs to a family of protein tyrosine phosphatases (PTPs)
- PTPs regulate cell-signal transduction
- PTPN22 regulates immune cell signalling: e.g. inhibits T-cell activation by restricting signalling downstream of the T-cell receptor (TCR).

Dysregulation due to rs2476601:

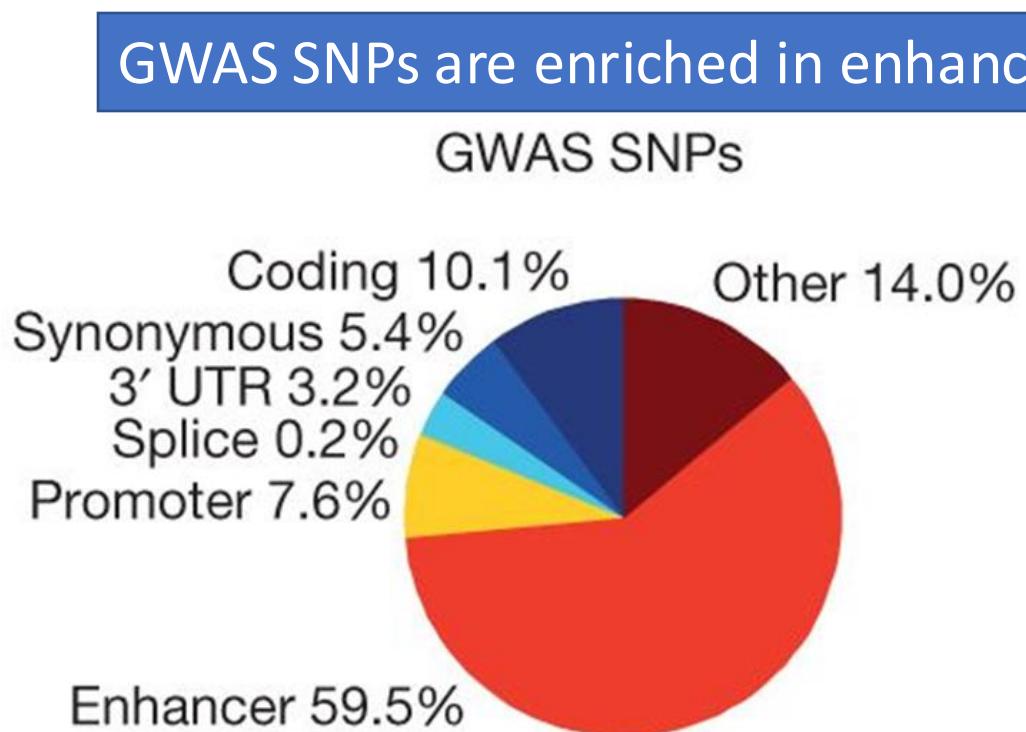
- T allele introduces Arg620Trp
- Proline rich domain
- Altered binding to downstream proteins in the T-cell activation signalling cascade, resulting in an altered immune response.



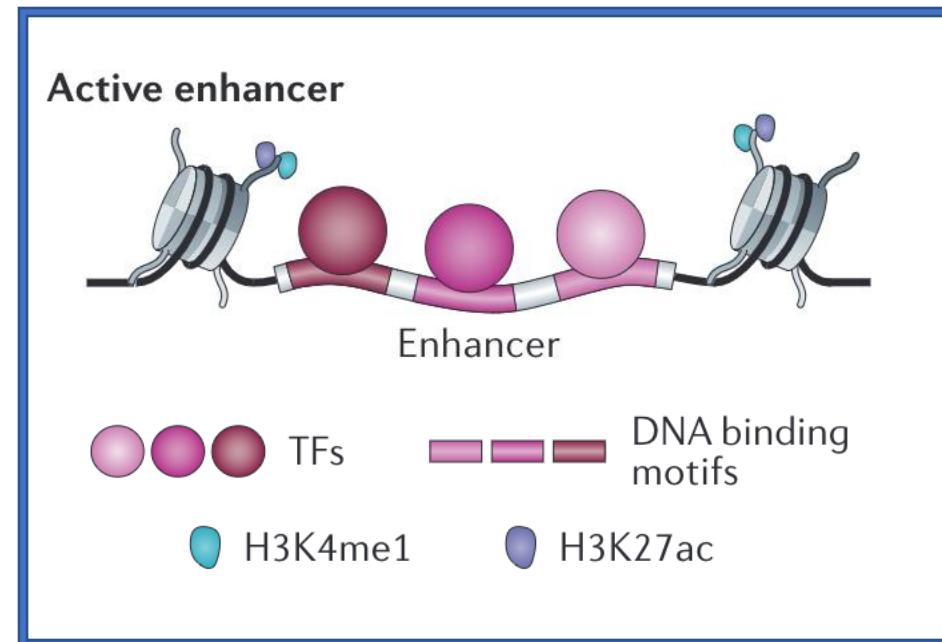
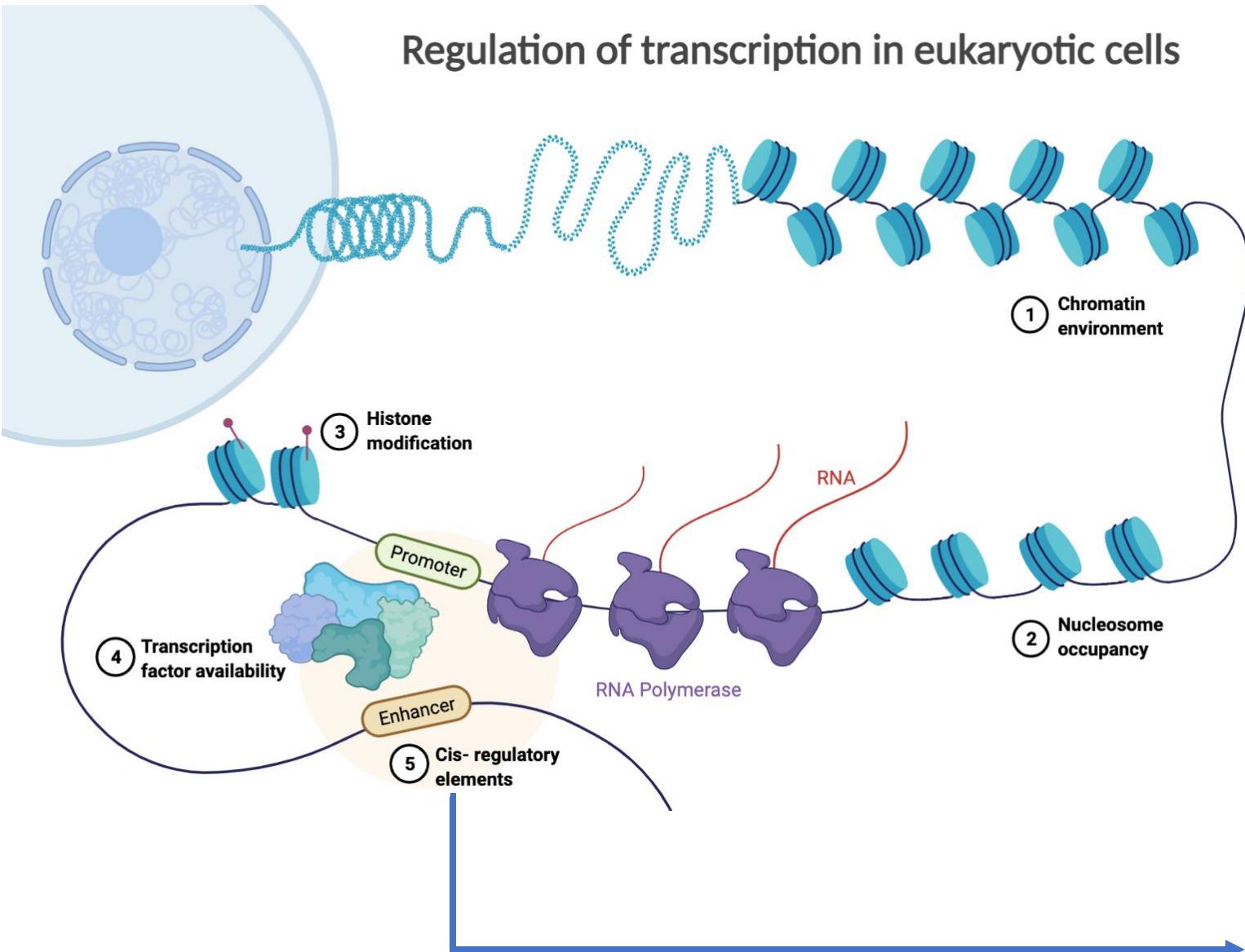
Stanford, S. M. & Bottini, N. (2014) *PTPN22: the archetypal non-HLA autoimmunity gene*
Nat. Rev. Rheumatol. doi:10.1038/nrrheum.2014.109

How can non-coding SNPs influence disease?

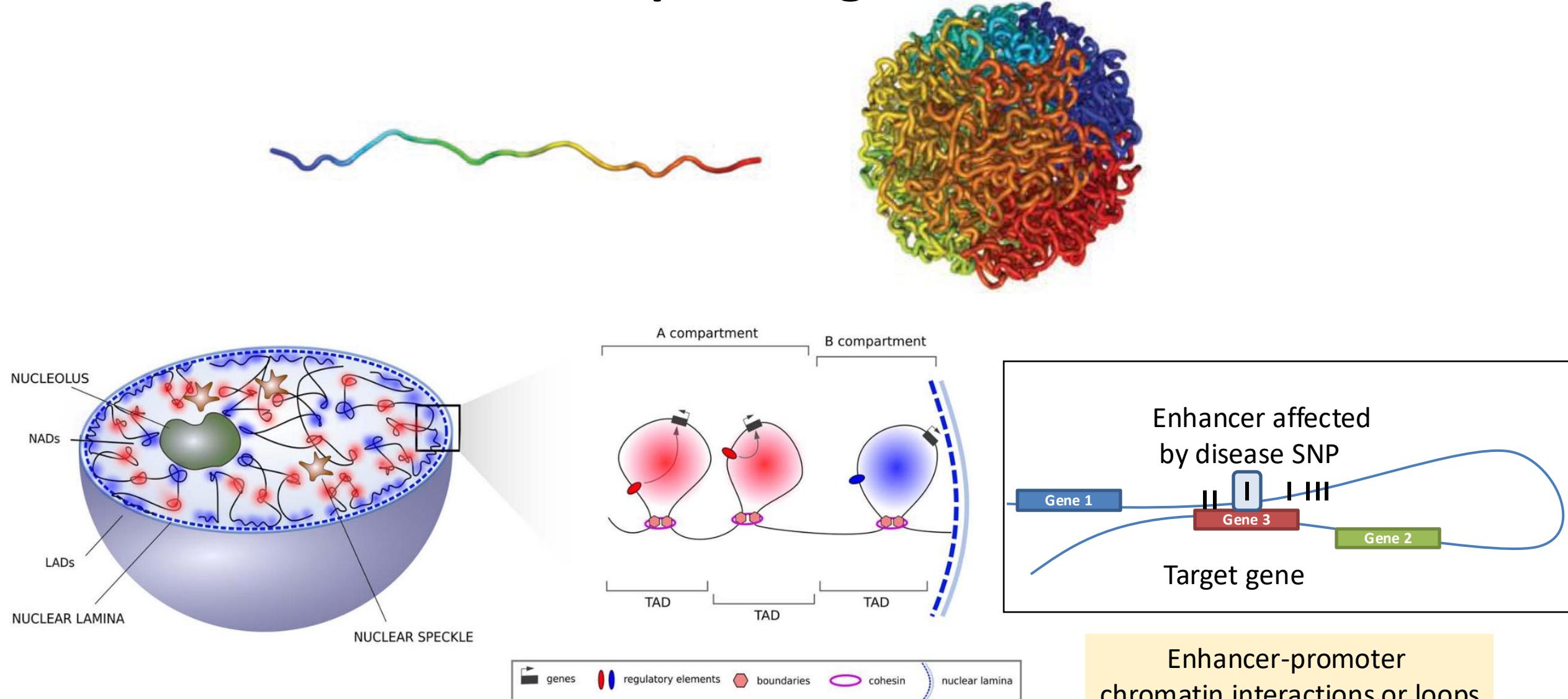
by altering the regulation of gene expression
in disease relevant tissues



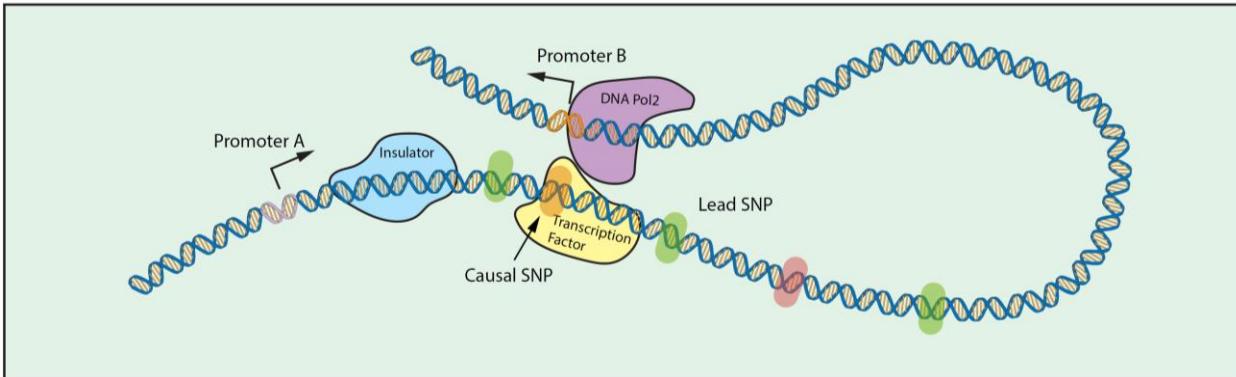
Regulation of transcription in eukaryotic cells



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 (CUT&Tag, CUT&RUN)

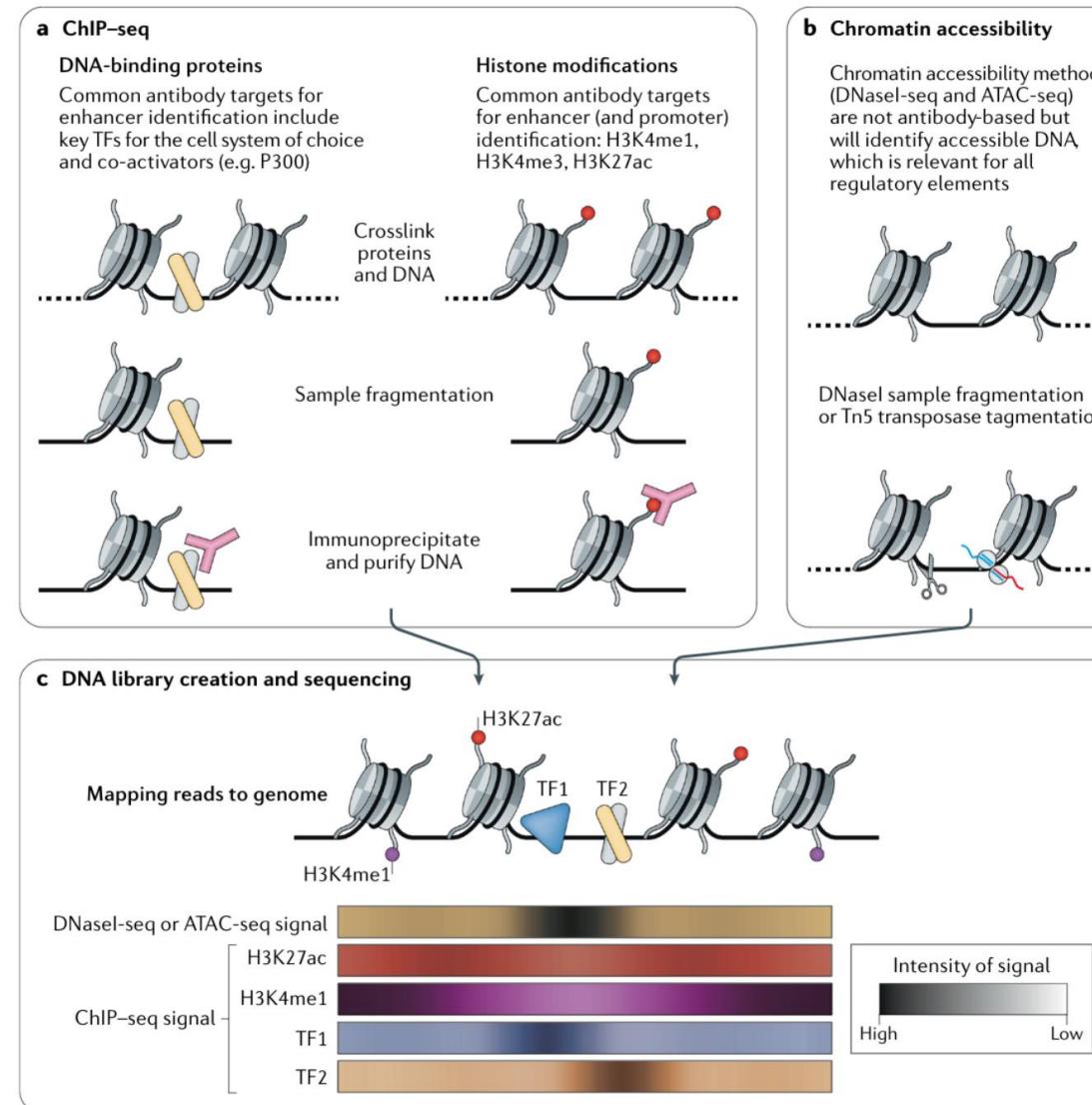
B. SNP to gene

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

SNP to function

Identification of regulatory elements/enhancers by DNA-binding proteins and DNA accessibility

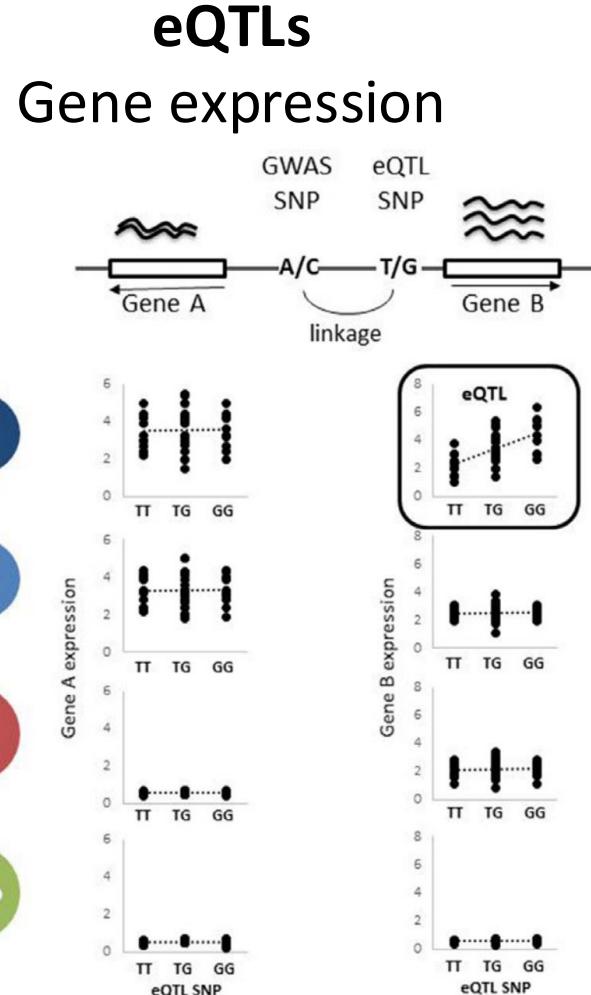
ChIP-seq
Chromatin immuno-precipitation coupled to sequencing



ATAC-seq
Assay for transposase- accessible chromatin

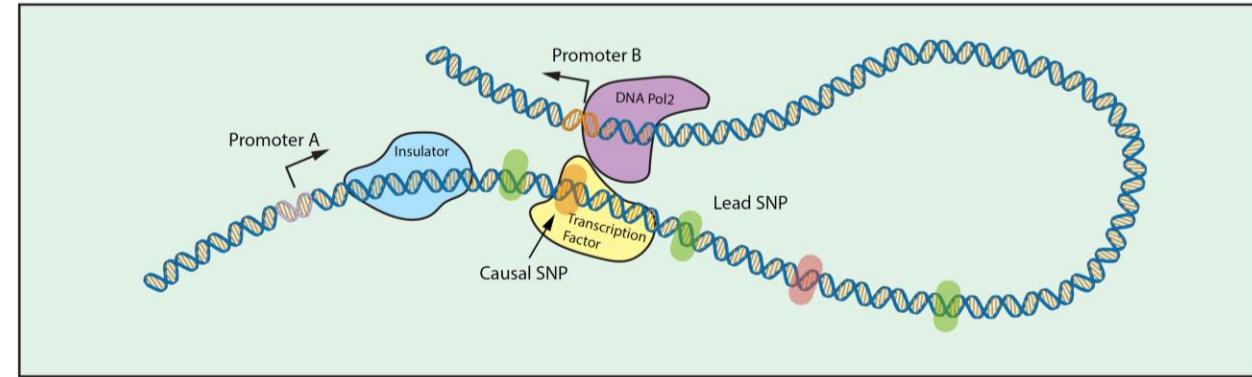
SNP to gene

Identification of the genes impacted by disease SNPs



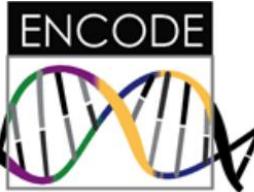
Hi-C

Chromatin interactions



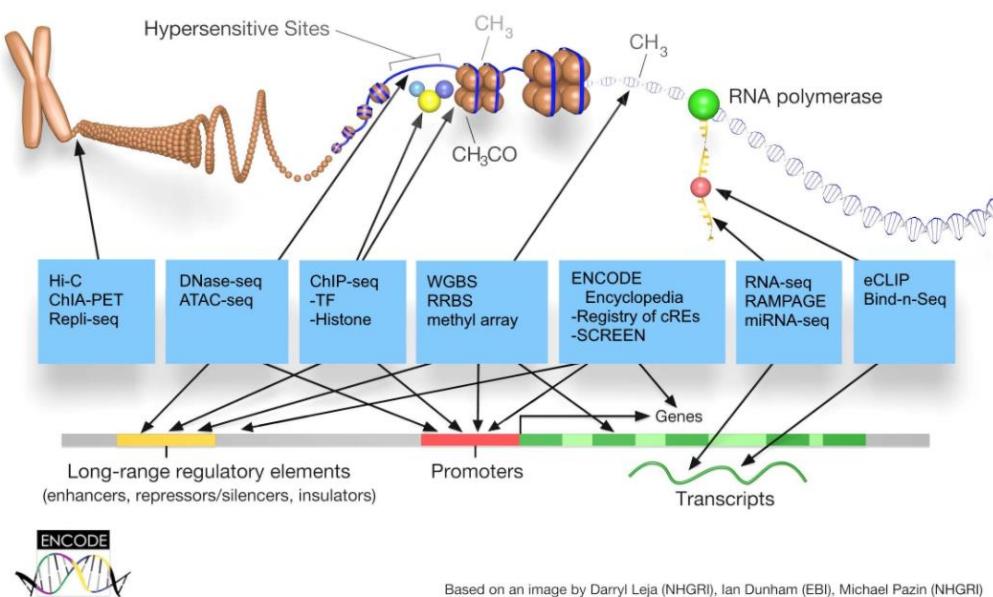
Shi et al. Rheumatology (2020)

Large-scale studies

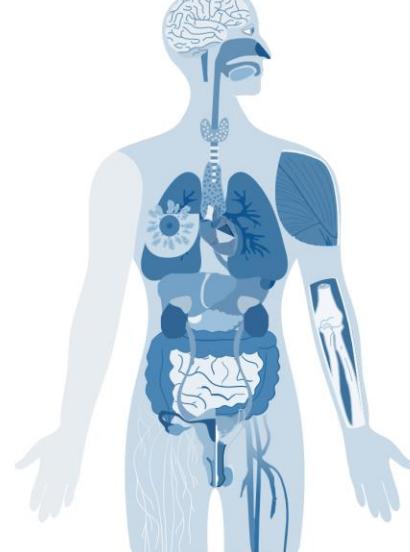


ENCODE: Encyclopaedia of DNA Elements

Experiments



Tissues



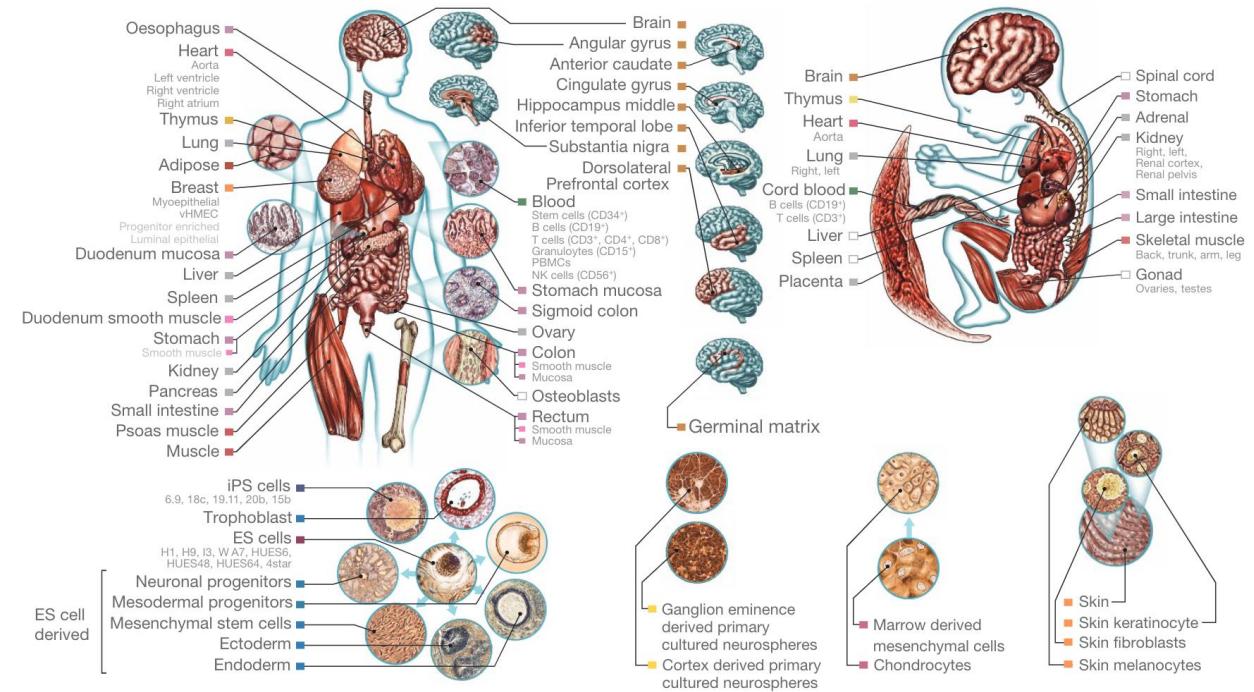
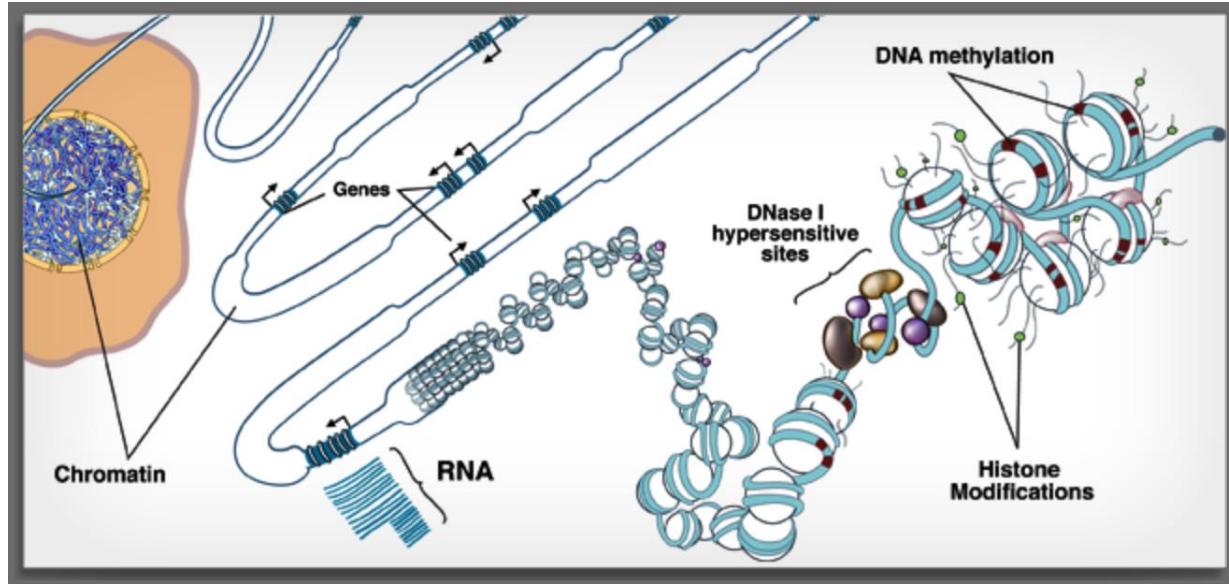
adrenal gland	arterial blood vessel
bone element	brain
breast	bronchus
colon	esophagus
eye	gallbladder
gonad	heart
intestine	kidney
large intestine	limb
liver	lung
mammary gland	mouth
musculature of body	nerve
nose	ovary
pancreas	penis
pericardium	prostate gland
skeleton	skin of body
small intestine	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.roadmapepigonomics.org/>



SNP to gene

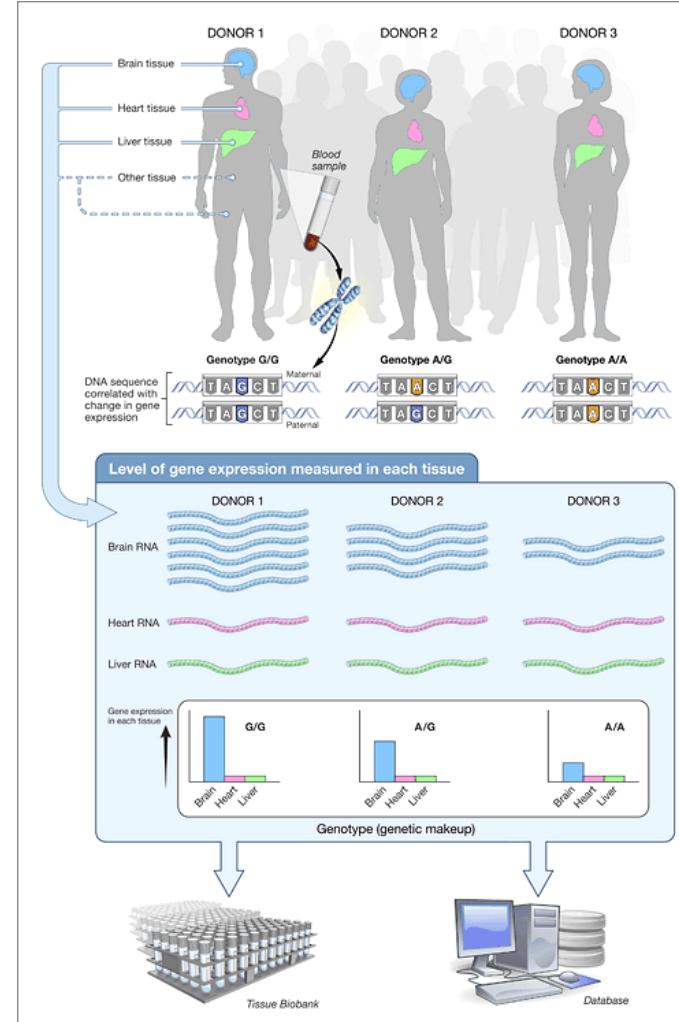
eQTLs: large scale transcriptomics studies

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

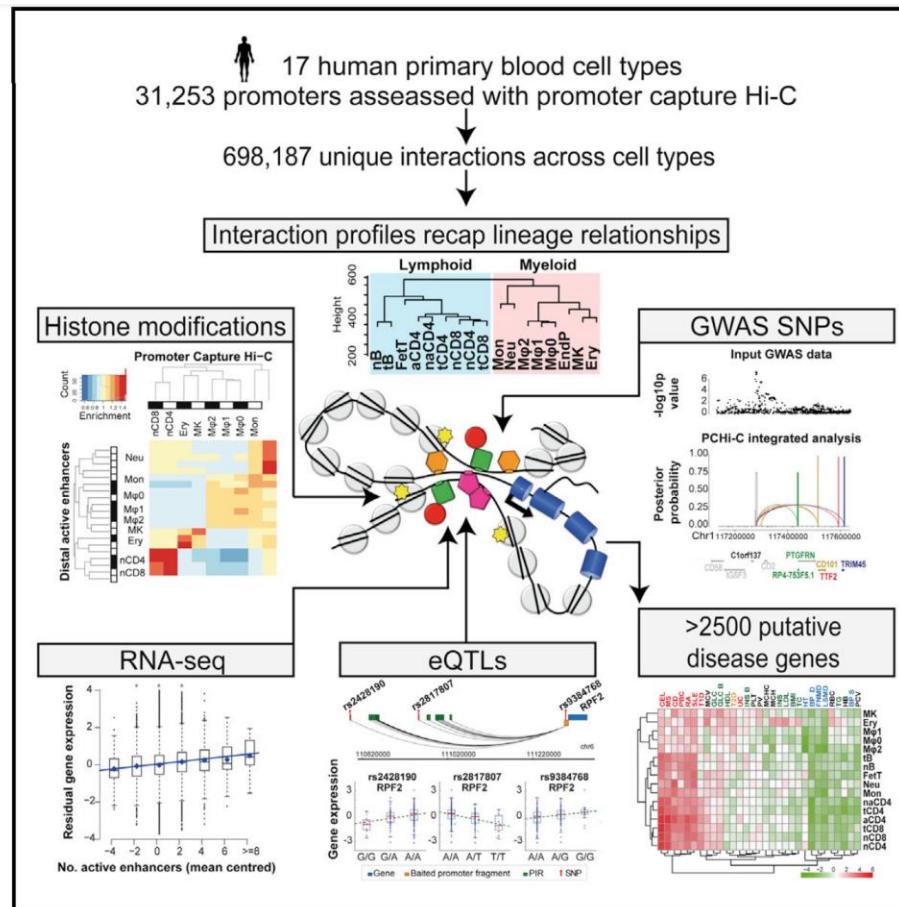
<https://gtexportal.org/home/>

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,*}

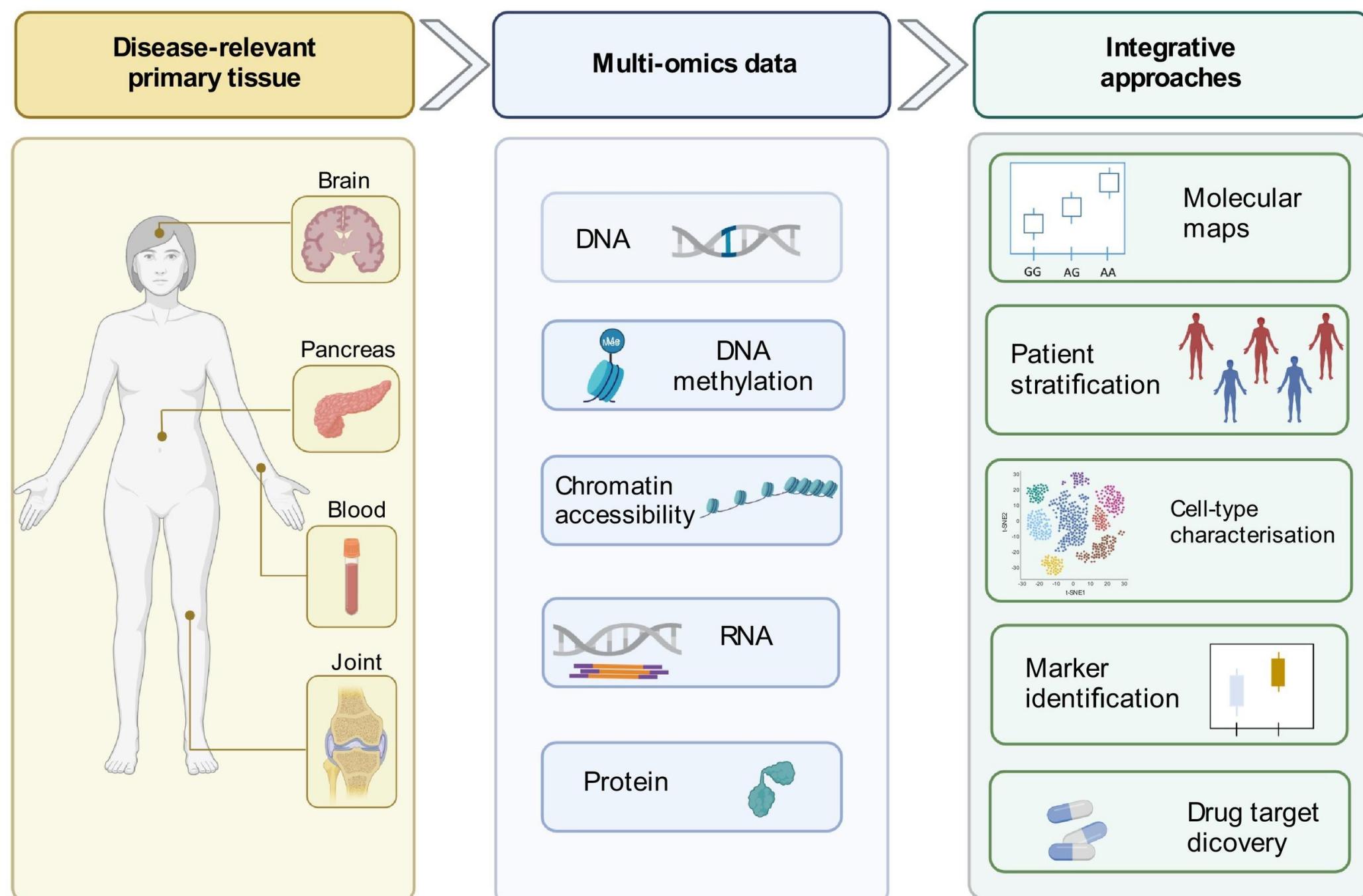
SNP to gene: chromatin interactions (Hi-C)

Cell 167, 1369–1384 (2016)



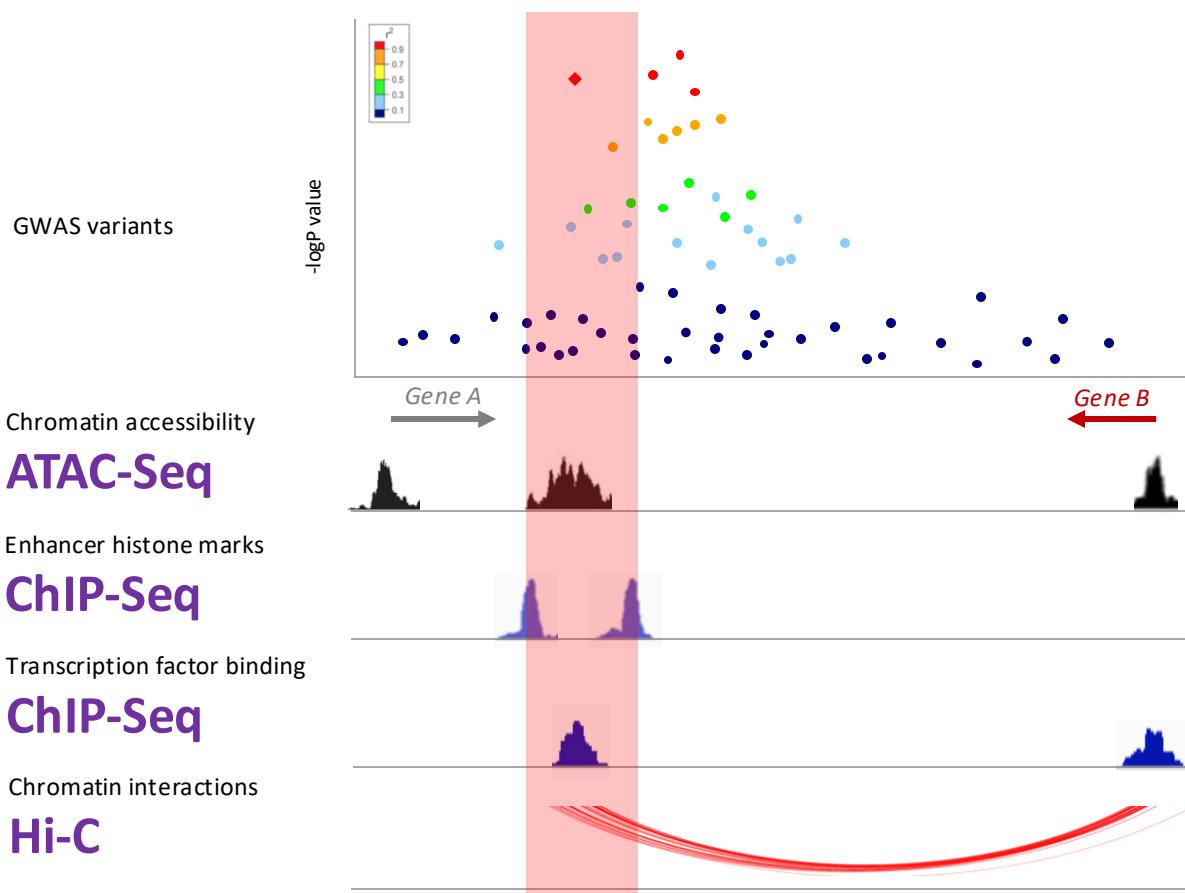
- Chromatin interaction patterns are cell type specific and segregate with the hematopoietic tree
- Promoter-interacting regions are enriched for regulatory chromatin features and eQTLs
- Promoter interactions link non-coding GWAS variants with putative target genes: more than 2,500 putative disease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

Omics integration

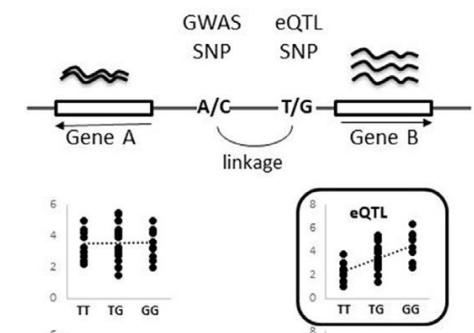
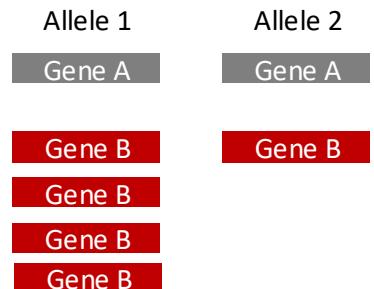
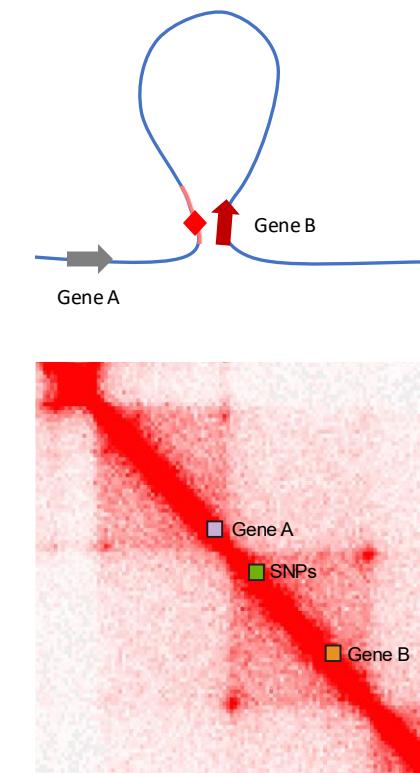


Functional annotation of GWAS: Integration with cell-type specific functional genomics datasets

A) From SNP to function



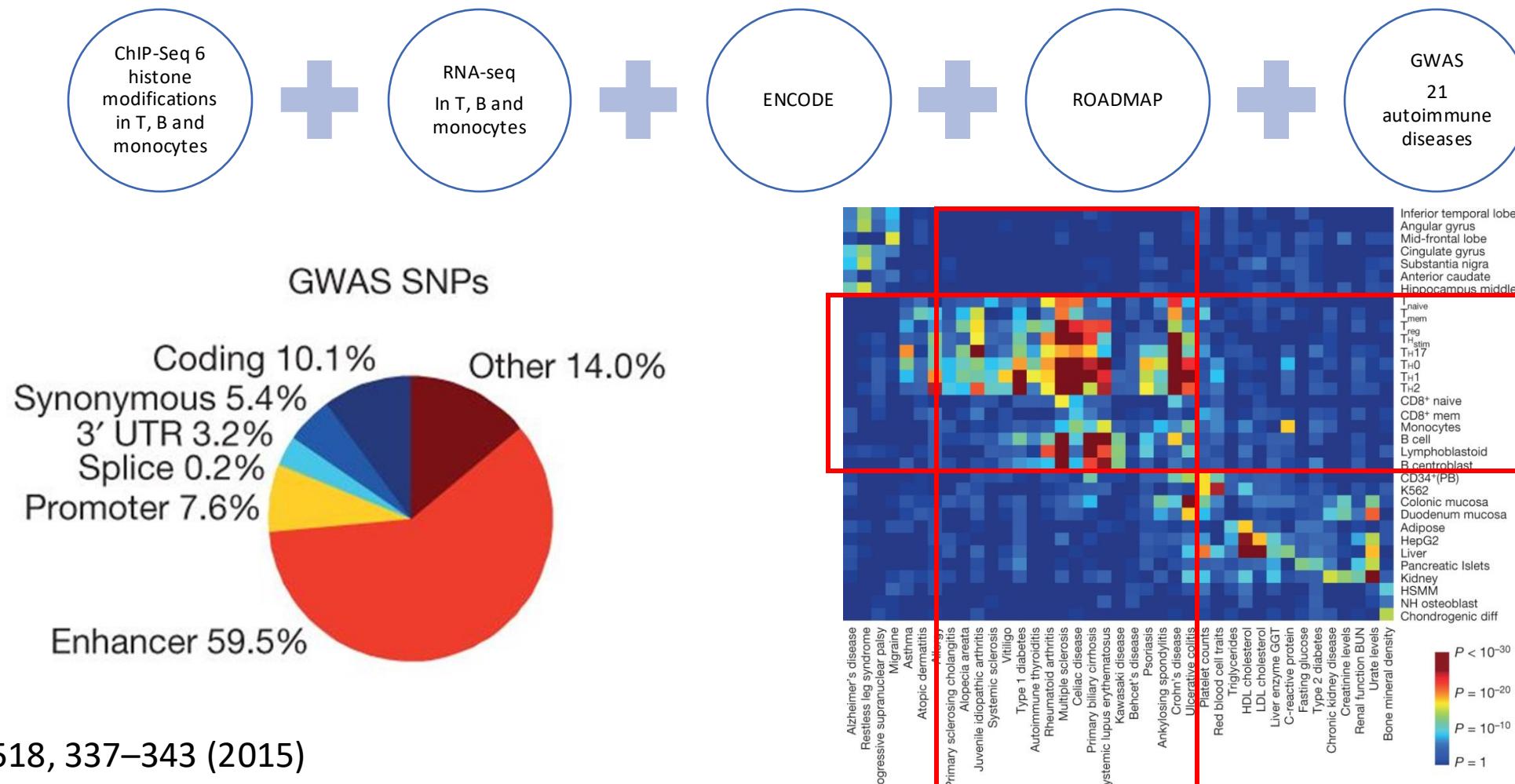
B) From SNP to gene



Genetic and epigenetic fine mapping of causal autoimmune disease variants

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 Shores¹, 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§}

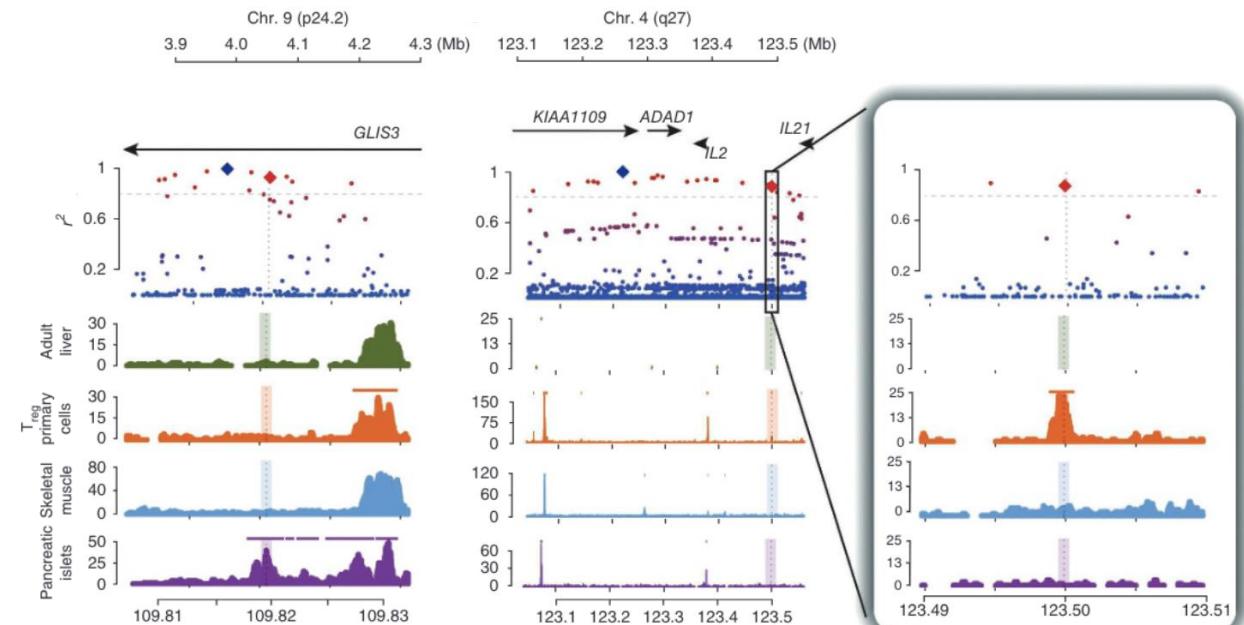
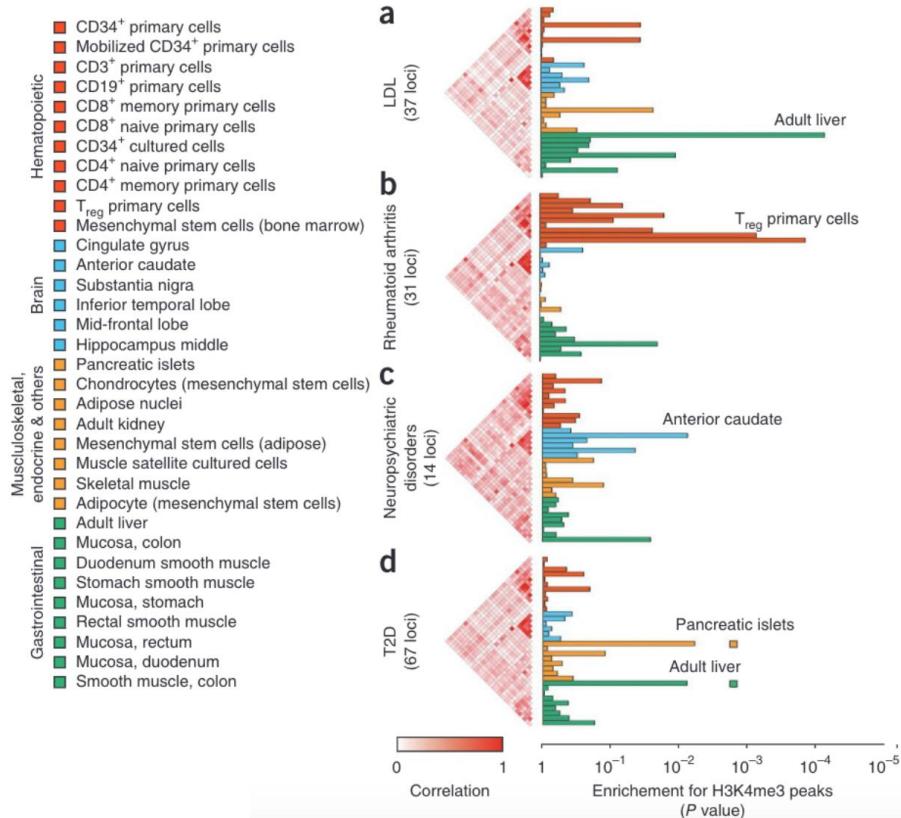
- Functional annotation
- Fine mapping
- Identification of disease relevant cells



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}

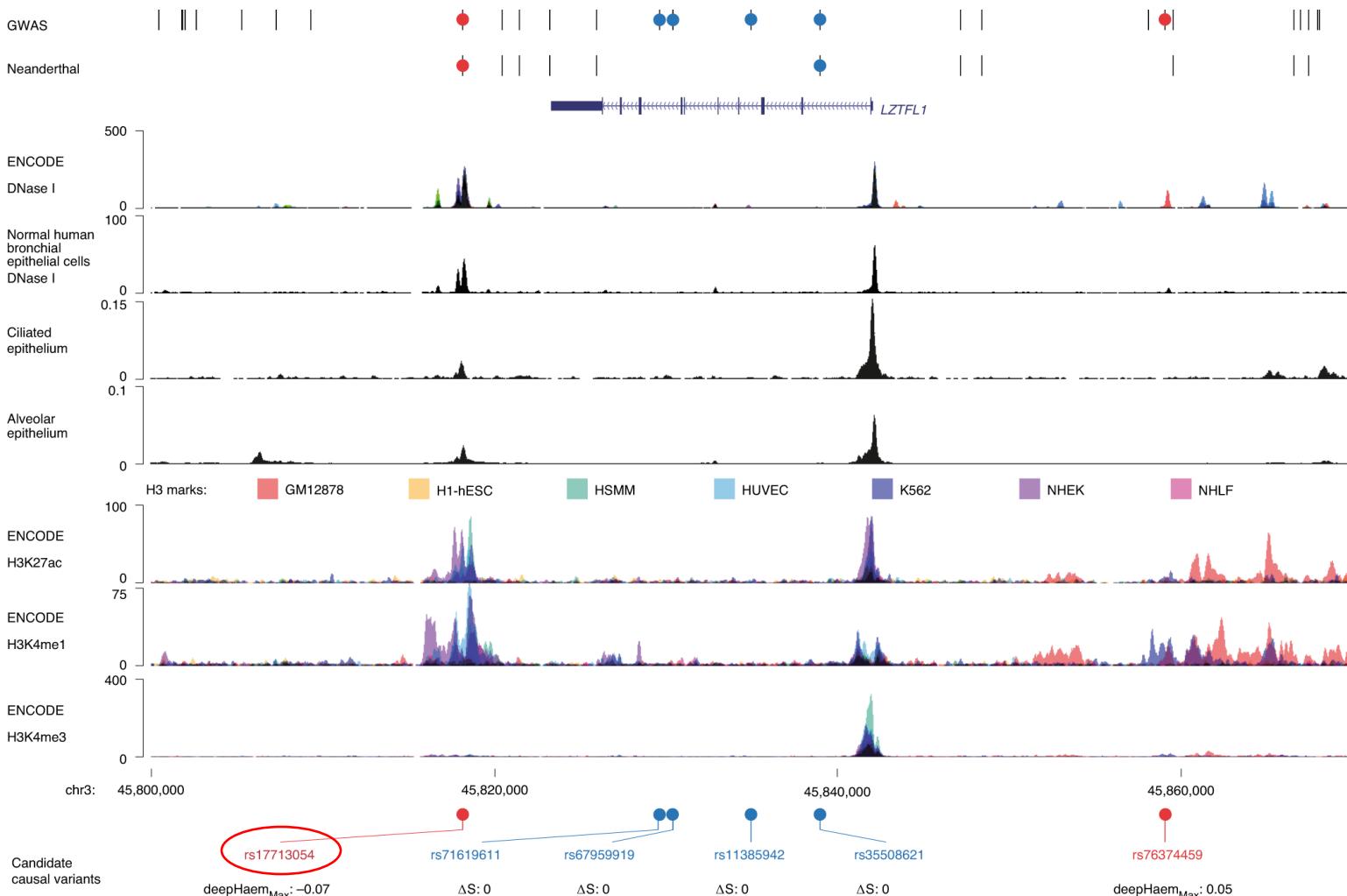


GLIS3 locus – T2D

IL2-IL21 locus: RA

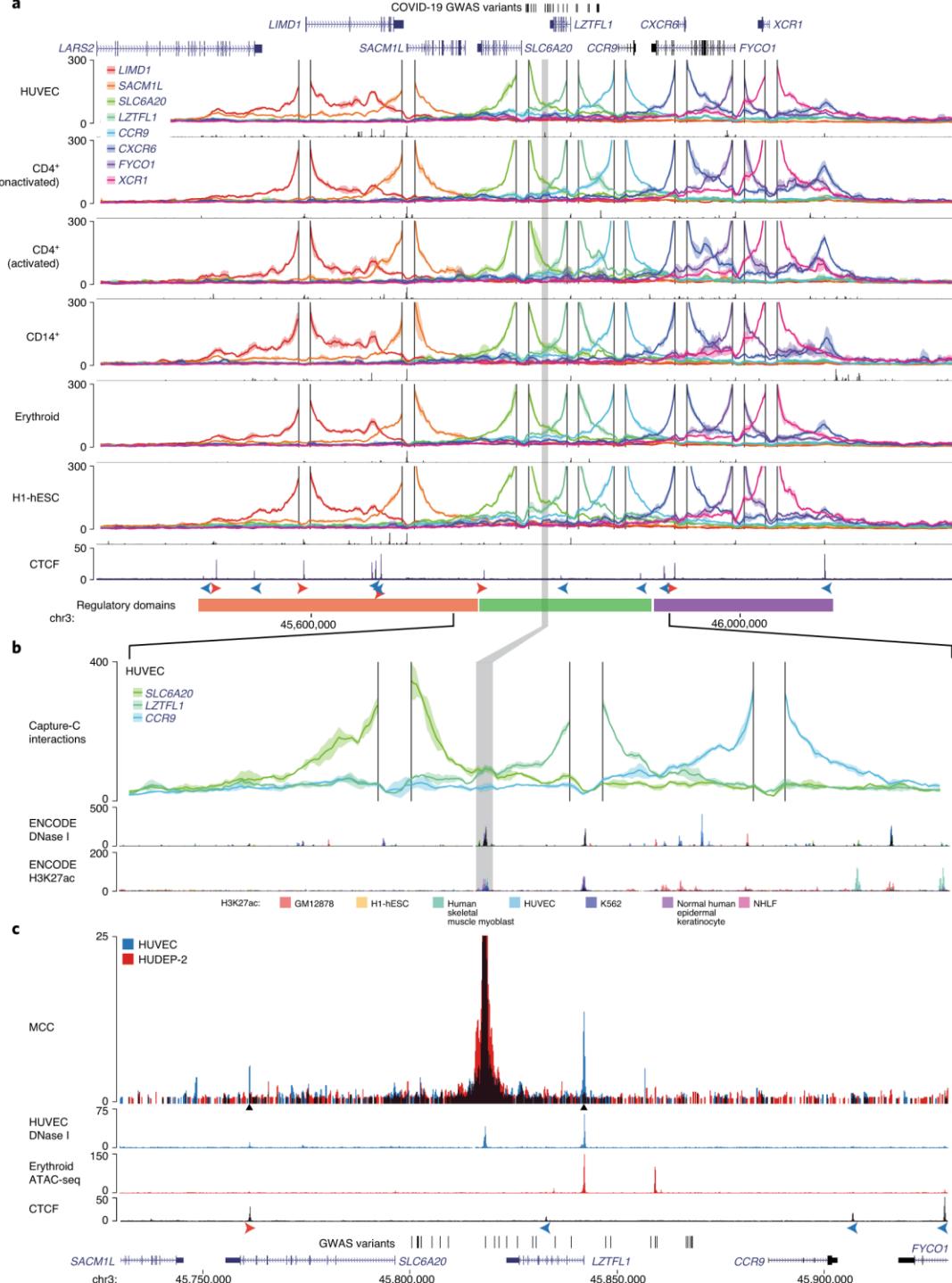
Identification of *LZTFL1* as a candidate effector gene at a COVID-19 risk locus.

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.

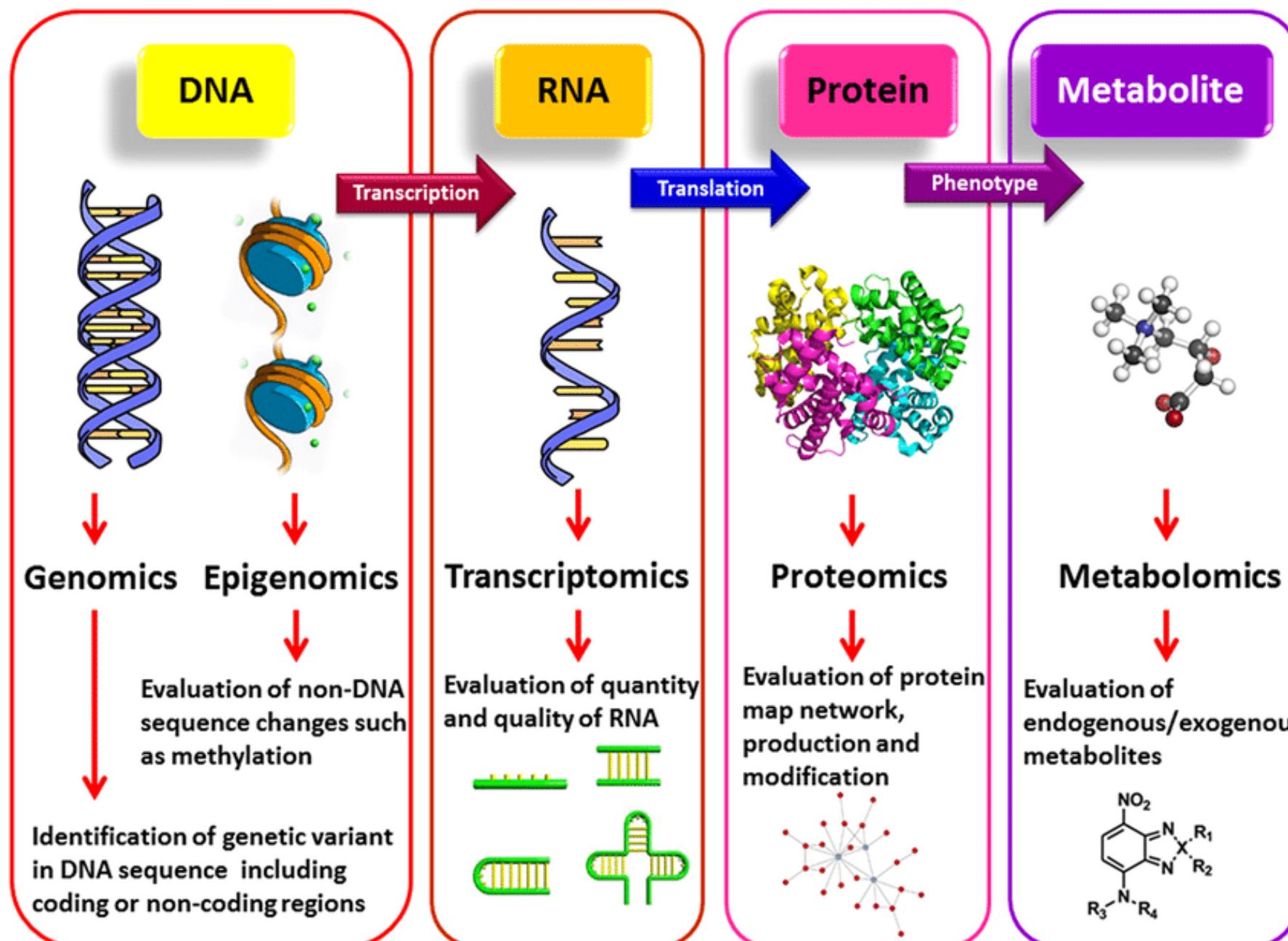


Identification of causal gene:

The rs1773054 enhancer interacts with the *LZTFL1* promoter and the SNP is an eQTL for *LZTFL1*.

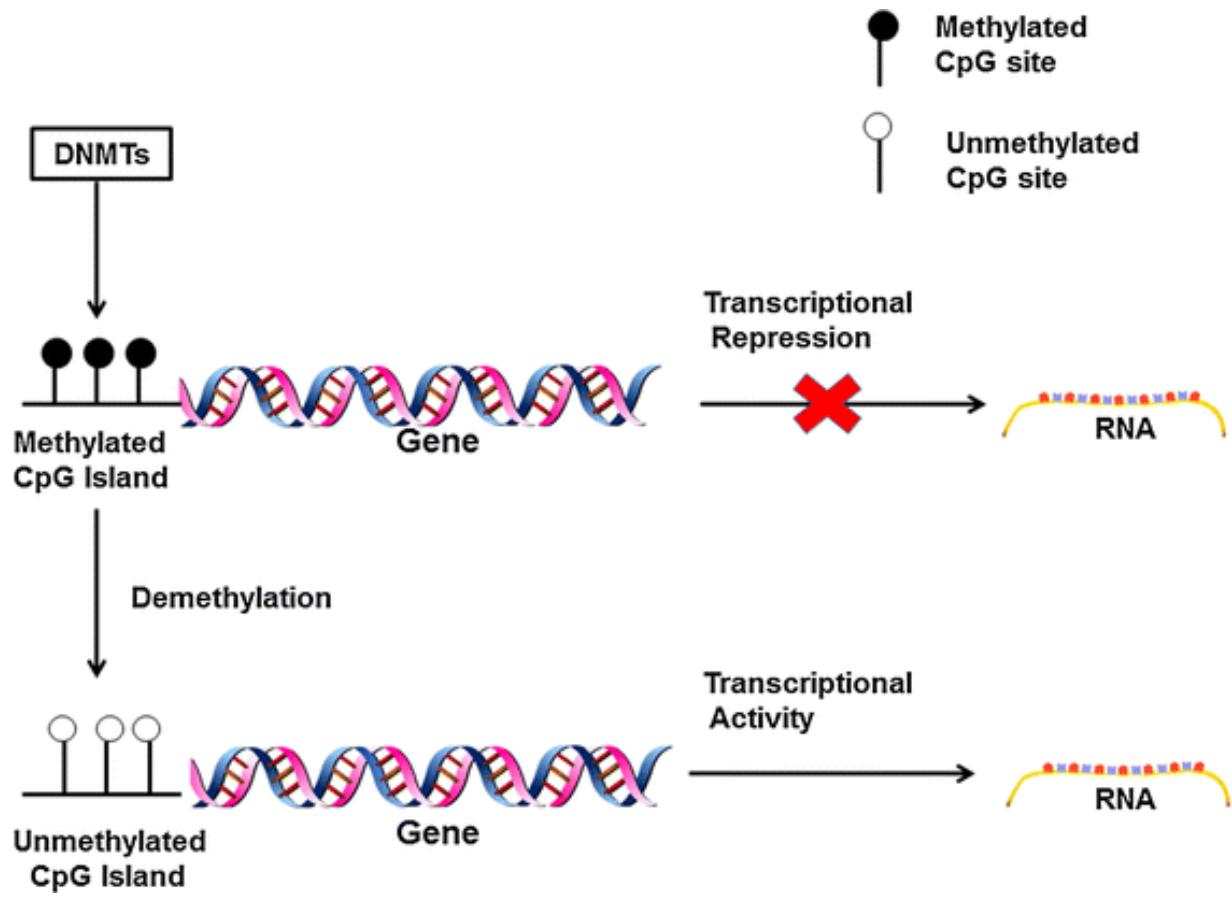
This gene is involved in an important viral response pathway.

Other omics and technologies



Methylation

- Methylation of DNA (not to be confused with histone methylation) is a common epigenetic marker.
- A methyl (CH_3) group is added to the C5 position of a cytosine residue, forming 5-methylcytosine (5mC)
- The majority of DNA methylation occurs on cytosines that precede a guanine nucleotide, or CpG sites.
- Its role is not completely understood, but it is involved in the regulation of gene expression
 - Repression, in general: “blocks” binding of transcriptional regulators e.g. TFs
- Easily measurable as a biomarker using microarrays



DNMTs: DNA methyltransferases

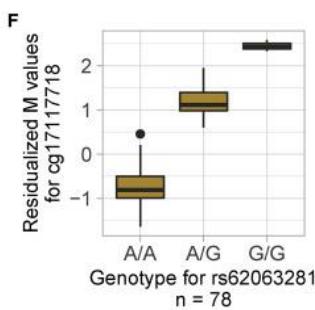
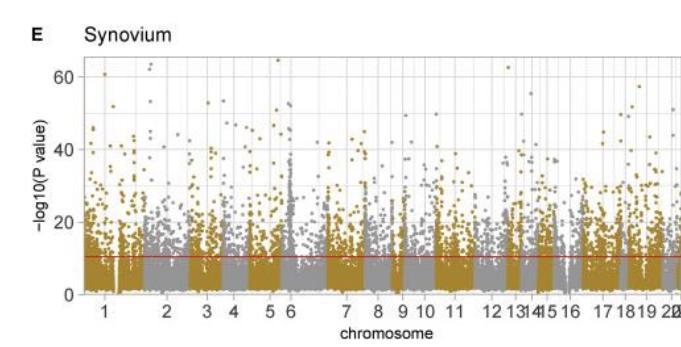
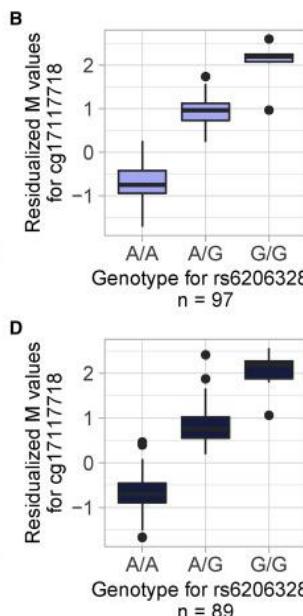
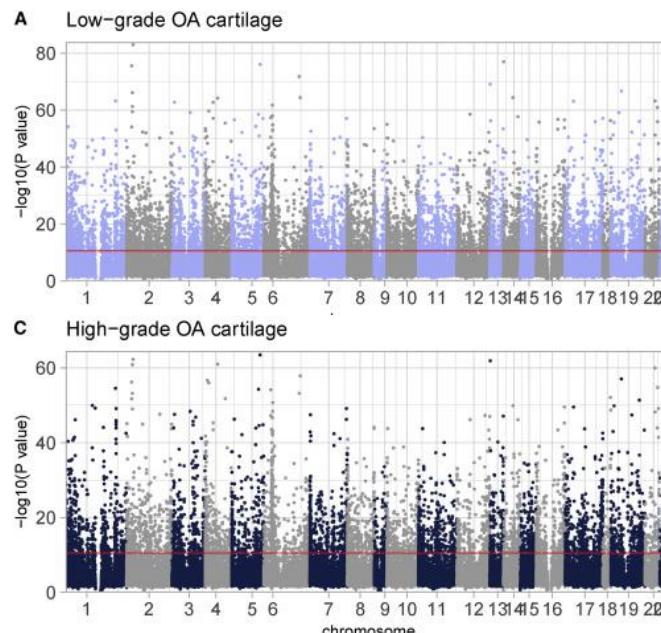
Example:

An epigenome-wide view of osteoarthritis in primary tissues

Peter Kreitmaier,^{1,2} Matthew Suderman,³ Lorraine Southam,¹ Rodrigo Coutinho de Almeida,⁴ Konstantinos Hatzikotoulas,¹ Ingrid Meulenbelt,⁴ Julia Steinberg,^{1,5} Caroline L. Relton,³ J. Mark Wilkinson,^{6,8,*} and Eleftheria Zeggini^{1,7,8,*}

The American Journal of Human Genetics 109, 1255–1271, July 7, 2022

- Genotype and methylation profiles of knee intact and degraded cartilage and synoviocytes from 98 OA patients
- Identification of DNA methylation markers of cartilage degeneration using Epigenome-wide association study (EWAS): 15,328 differentially methylated sites (DMSs)
- Methylation QTL (meQTL) maps



Genetic variants that are significantly associated with methylation levels of proximal methylation sites (<1 Mb; *cis*-mQTLs). Tissue specific.

Example:

An epigenome-wide view of osteoarthritis in primary tissues

Peter Kreitmaier,^{1,2} Matthew Suderman,³ Lorraine Southam,¹ Rodrigo Coutinho de Almeida,⁴ Konstantinos Hatzikotoulas,¹ Ingrid Meulenbelt,⁴ Julia Steinberg,^{1,5} Caroline L. Relton,³ J. Mark Wilkinson,^{6,8,*} and Eleftheria Zeggini^{1,7,8,*}

The American Journal of Human Genetics 109, 1255–1271, July 7, 2022

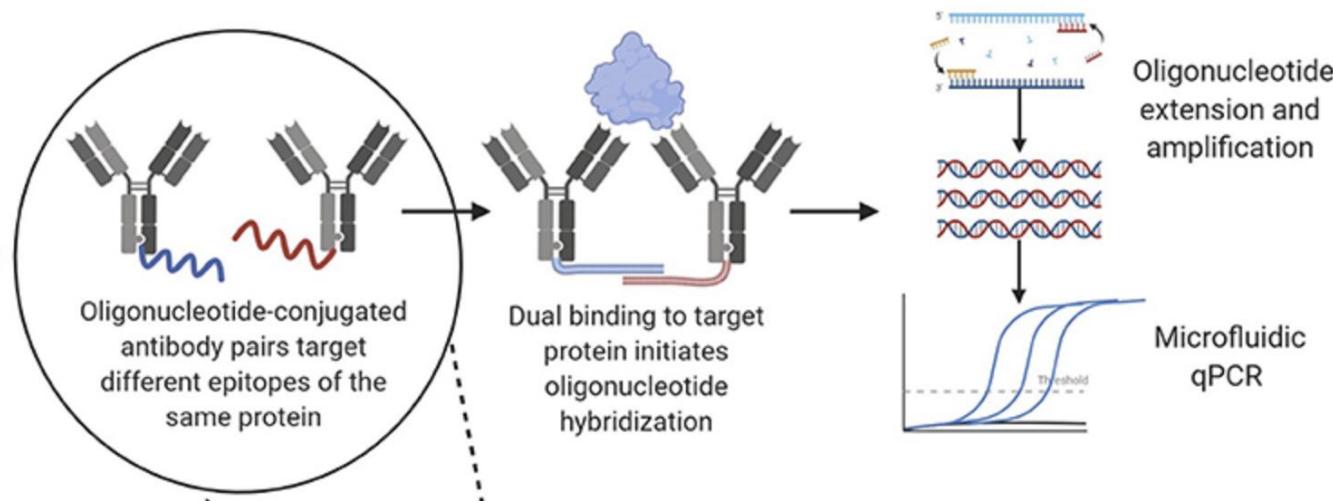
- Colocalization with GWAS signals, for the identification of osteoarthritis-linked genetic risk variants that exert their effect through the regulation of nearby methylation sites.
- They found osteoarthritis-linked genetic variants that colocalize with a methylation site and additionally show an eQTL effect at nominal significance on the gene annotated to the respective methylation site in the same tissue.

GWAS Trait	GWAS lead variant	RA	mQTL Tissue	Msite	Msite Location (Gene)	RA effect on meth	Coloc PP	Gene	RA effect on exp	p-value (RA effect on exp)	Effect Meth on exp	p-value (Effect Meth on exp)
alloOA	rs798726	C	H-G	cg15763121	3'UTR	-	0.81	FAM53A	-	0.002	-	-
alloOA	rs2061027	A	H-G	cg26672287	Body	+	0.87	LTBP1	-	0.008	-	0.046
alloOA	rs798726	C	L-G	cg00626702	Body	-	0.82	FAM53A	-	3e-05	-	-
alloOA	rs35206230	T	L-G	cg05627522	1776_tss	+	0.94	RPP25	-	0.022	-	0.004
alloOA	rs798726	C	L-G	cg07428655	3'UTR	-	0.91	FAM53A	-	3e-05	-	-
alloOA	rs798726	C	L-G	cg07929082	5'UTR	+	0.94	FAM53A	-	3e-05	-	-
alloOA	rs1126464	G	L-G	cg08723459	3'UTR	+	0.88	CHMP1A	-	0.007	-	-
alloOA	rs798726	C	L-G	cg15763121	3'UTR	-	0.91	FAM53A	-	3e-05	-	-
alloOA	rs35206230	T	L-G	cg18612461	1928_tss	+	0.94	RPP25	-	0.022	-	0.005
alloOA	rs798726	C	L-G	cg21231739	3'UTR	-	0.91	FAM53A	-	3e-05	+	0.037
alloOA	rs3771501	A	L-G	cg27223728	3415_tss	-	0.96	TGFA	-	0.002	-	-
kneeOA	rs4775006	A	L-G	cg02900766	Body	-	0.99	ALDH1A2	+	0.003	-	0.023
kneeOA	rs4775006	A	L-G	cg12382153	Body	-	0.91	ALDH1A2	+	0.003	-	-
TKR	rs34958088	G	L-G	cg02900766	Body	-	0.99	ALDH1A2	+	7e-04	-	0.023
TKR	rs34958088	G	L-G	cg12382153	Body	-	0.93	ALDH1A2	+	7e-04	-	-
alloOA	rs2171126	T	Synov	cg19571390	Body	+	0.86	CRADD	+	0.003	-	-

Proteomics

- Proteomics is the large-scale study of proteomes.
 - A proteome is a set of proteins produced in an organism, system, or biological context
- Recent technological advances enable high-throughput, multiplex protein biomarker analysis

Olink platform: Proximity Extension Assay (PEA) technology uses oligonucleotide-conjugated antibody pairs for targeting different epitopes of the same protein. Only when antibody pairs bind the same protein are the oligonucleotides close enough to hybridize. Annealed sequences are then extended, amplified, and measured by microfluidic qPCR.



Example:

Article

Plasma proteomic associations with genetics and health in the UK Biobank

Nature | Vol 622 | 12 October 2023

<https://doi.org/10.1038/s41586-023-06592-6>

Received: 17 June 2022

Accepted: 31 August 2023

Published online: 4 October 2023

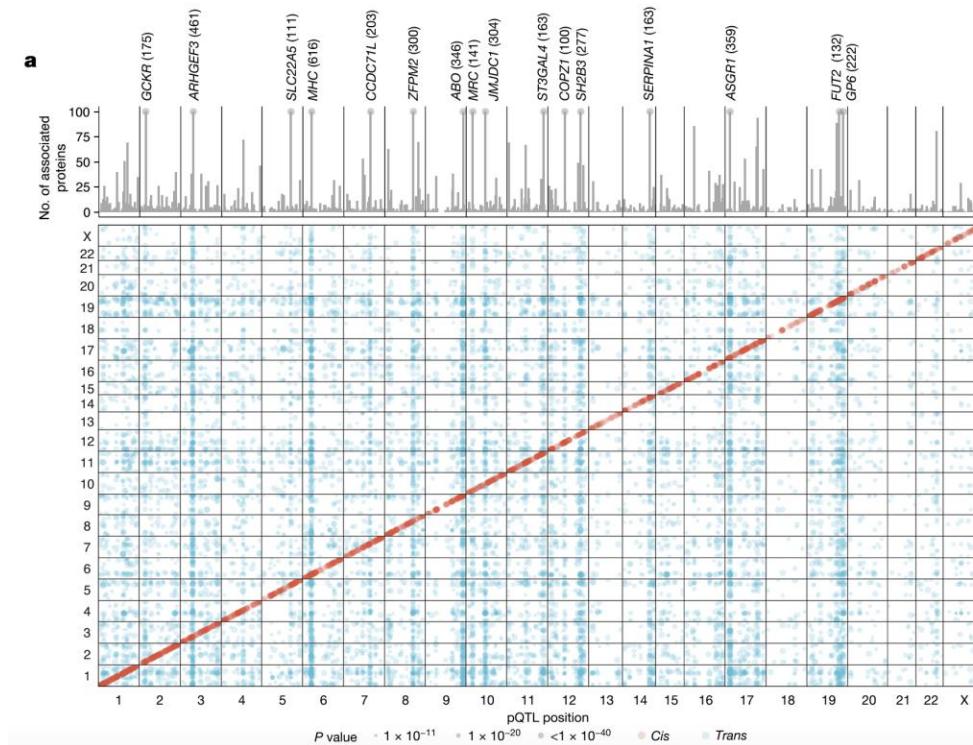
Open access

 Check for updates

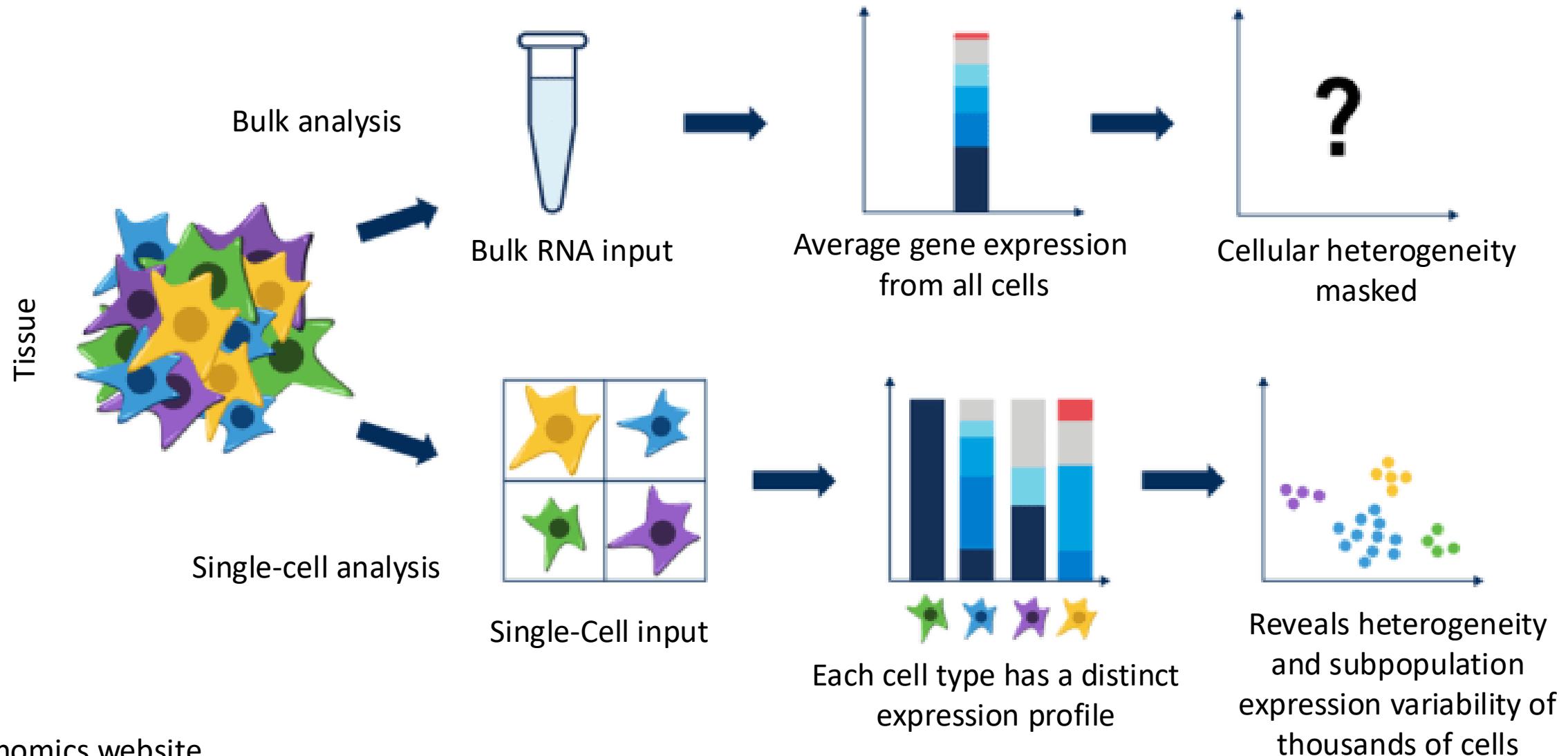
Benjamin B. Sun^{1,2*}, Joshua Chiou^{2,26}, Matthew Traylor^{3,26}, Christian Benner^{4,26}, Yi-Hsiang Hsu^{5,26}, Tom G. Richardson^{3,6,26}, Praveen Surendran^{6,26}, Anubha Mahajan^{4,26}, Chloe Robins^{7,26}, Steven G. Vasquez-Grinnell^{8,26}, Liping Hou^{1,26}, Erika M. Kvistad^{3,26}, Oliver S. Burren¹⁰, Jonathan Davitte⁷, Kyle L. Ferber¹¹, Christopher E. Gillies¹², Åsa K. Hedman¹³, Sile Hu³, Tinchi Lin¹⁴, Rajesh Mikkilineni¹⁵, Rion K. Pendergrass⁴, Corran Pickering¹⁶, Bram Prins¹⁰, Denis Baird¹, Chia-Yen Chen¹, Lucas D. Ward¹⁷, Aimee M. Deaton¹⁷, Samantha Welsh¹⁶, Carissa M. Willis¹⁷, Nick Lehner¹⁸, Matthias Arnold^{18,19}, Maria A. Wörheide¹⁸, Karsten Suhré²⁰, Gabi Kastenmüller¹⁶, Anurag Sethi²¹, Madeleine Cule²¹, Anil Raj²¹, Alnylam Human Genetics^{*}, AstraZeneca Genomics Initiative^{*}, Biogen Biobank Team^{*}, Bristol Myers Squibb^{*}, Genentech Human Genetics^{*}, GlaxoSmithKline Genomic Sciences^{*}, Pfizer Integrative Biology^{*}, Population Analytics of Janssen Data Sciences^{*}, Regeneron Genetics Center^{*}, Lucy Burkitt-Gray¹⁶, Eugene Melamud²¹, Mary Helen Black⁹, Eric B. Fauman², Joanna M. M. Howson³, Hyun Min Kang³, Mark I. McCarthy⁴, Paul Nioi¹⁷, Slavé Petrovski^{10,22}, Robert A. Scott⁴, Erin N. Smith²³, Sándor Szalma²³, Dawn M. Waterworth²⁴, Lyndon J. Mitnaul¹², Joseph D. Szustakowski^{3,27}, Bradford W. Gibson^{5,27}, Melissa R. Miller^{2,27} & Christopher D. Whelan^{1,26,27}

- Landmark study - UK Biobank Pharma Proteomics Project (UKB-PPP): largest and most comprehensive study (at the time of publication) on the effects of common genetic variation on proteins circulating in the blood and how these associations can contribute to disease
- Researchers measured the abundance of nearly 3,000 circulating proteins, many of which were previously difficult to capture, from over 54,000 participants in the UK Biobank – which has been collecting data and tracking the health of 500,000 volunteer participants enrolled between 2006 and 2010

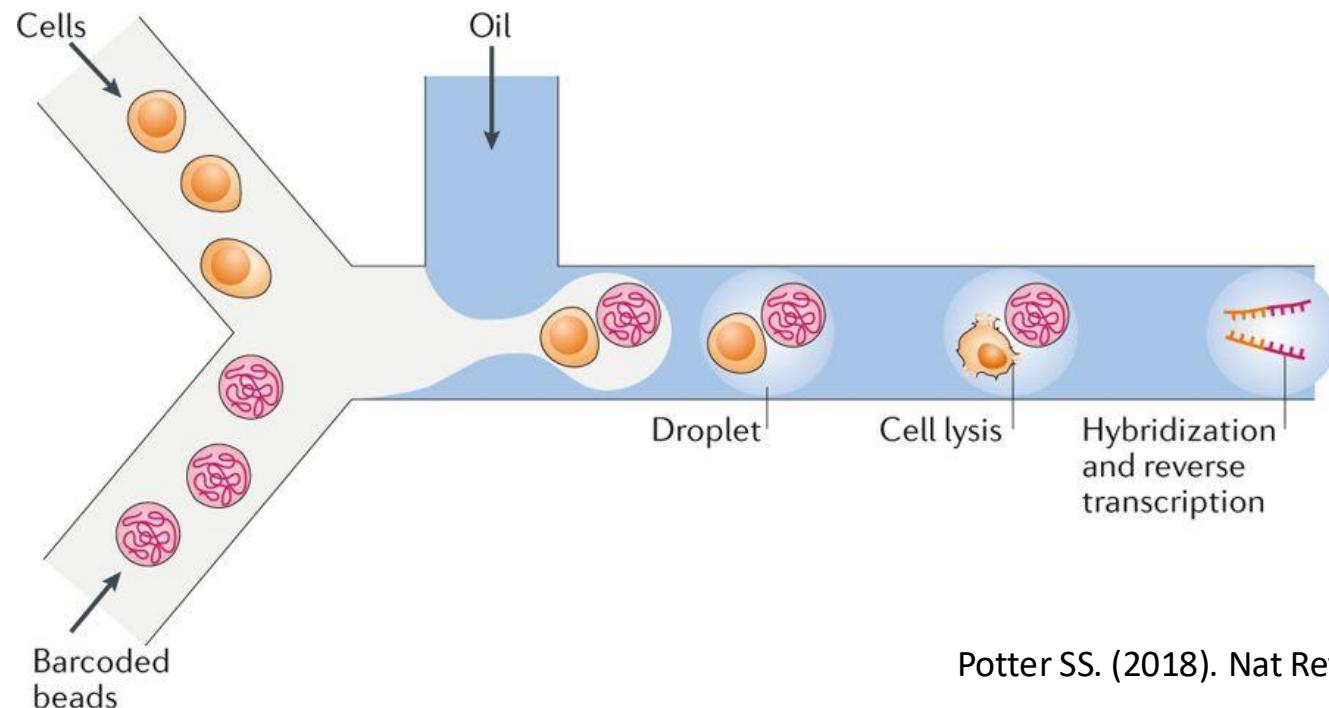
- Profiling of blood protein levels was performed across the top 20 most common health conditions in UK Biobank. This revealed that, for example, inflammatory proteins, long thought to contribute towards mental health conditions, are significantly higher in patients with depression.
- The study presents comprehensive **protein quantitative trait locus (pQTL)** mapping of 2,923 proteins that identifies 14,287 primary genetic associations, of which 81% are previously undescribed
 - accessible at <http://ukb-ppp.gwas.eu>



Single cell technologies



- A microfluidics system is used to make oil microdroplets; each droplet contains only one cell mixed with specific beads.
- Each bead has oligonucleotides that are uniquely barcoded for that bead and are in a solution that contains a mild detergent, which lyses the cells after mixing.
- The RNAs from the lysed cell anneal to the bead oligonucleotides, and subsequent reverse transcription incorporates the bead-specific barcode into the cDNA, thereby allowing the sequences of those cDNAs to be assigned to a specific cell.



Example:

Article

<https://doi.org/10.1038/s41588-024-01682-1>

Tissue-specific enhancer–gene maps from multimodal single-cell data identify causal disease alleles

Received: 7 March 2023

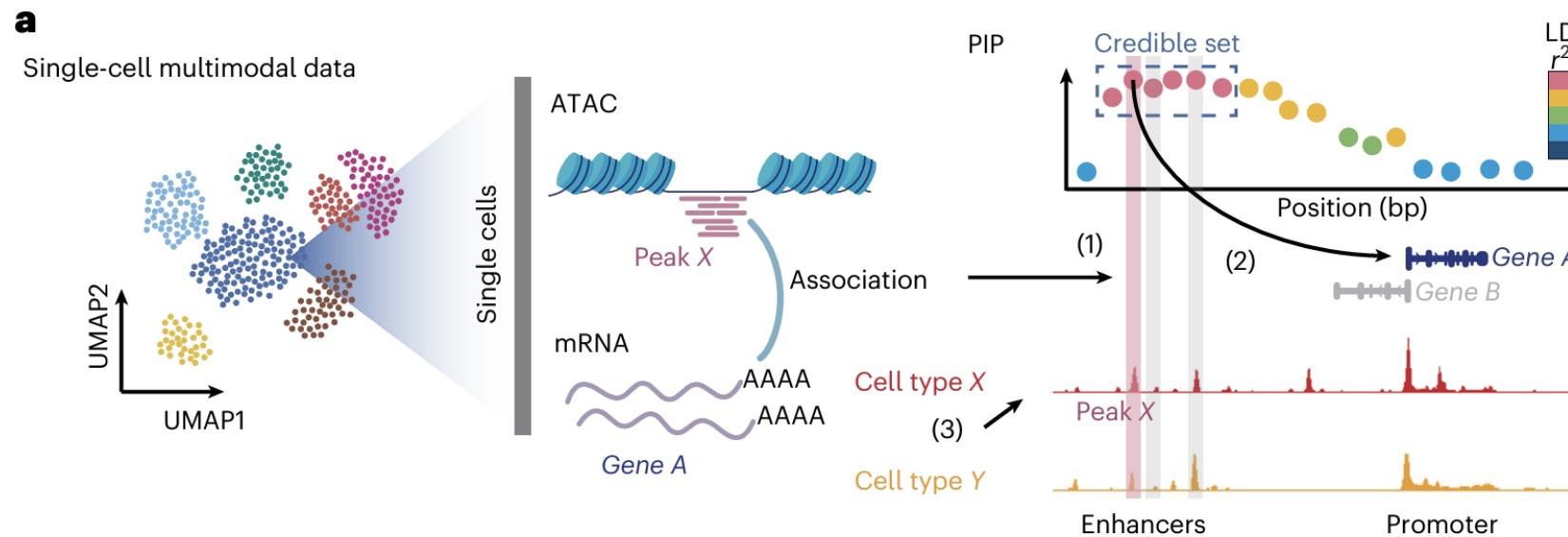
Accepted: 7 February 2024

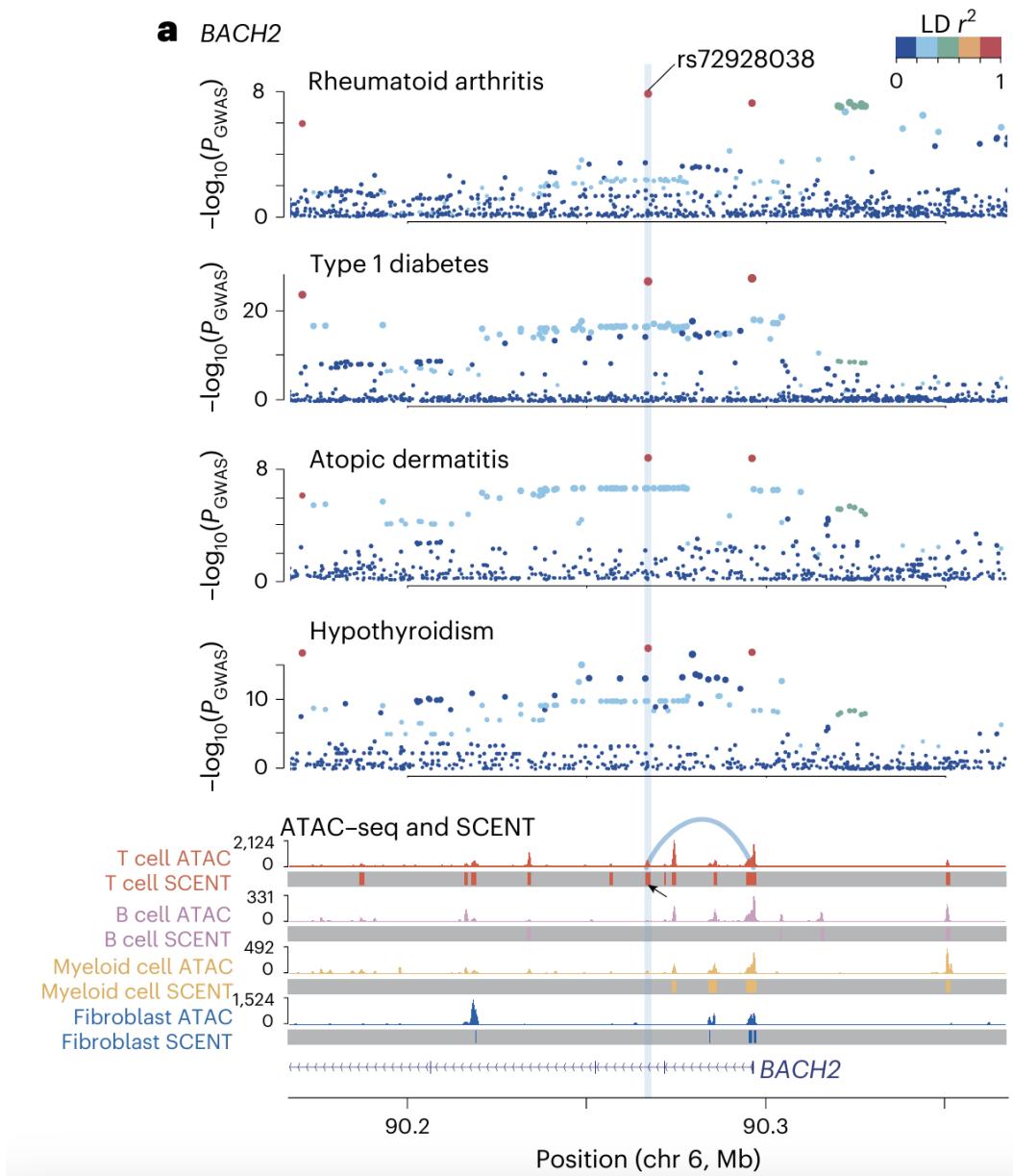
Published online: 9 April 2024

Check for updates

Saori Sakaue 1,2,3, Kathryn Weinand 1,2,3,4, Shakson Isaac 1,2,3,4, Kushal K. Dey 3,5, Karthik Jagadeesh 3,5, Masahiro Kanai 3,6,7,8, Gerald F. M. Watts⁹, Zhu Zhu⁹, Accelerating Medicines Partnership® RA/SLE Program and Network*, Michael B. Brenner 9, Andrew McDavid¹⁰, Laura T. Donlin^{11,12}, Kevin Wei⁹, Alkes L. Price 3,5,13 & Soumya Raychaudhuri 1,2,3,4

- Developed a statistical method, SCENT (single-cell enhancer target gene mapping), that models association between enhancer chromatin accessibility and gene expression in single-cell multimodal RNA sequencing and ATAC sequencing data.
- SCENT can identify (1) active cis-regulatory regions and (2) their target genes in (3) a specific cell type.
- These SCENT results can be used to define likely causal variants, genes and cell types for GWAS loci.





rs72928038 was prioritized as the causal variant at *BACH2* locus by T-cell-specific SCENT enhancer-gene map for rheumatoid arthritis (RA), type 1 diabetes (T1D), atopic dermatitis and hypothyroidism.

Validation: genome editing, MPRAs

Genome editing

- Genome editing is the deliberate alteration of a selected DNA sequence in a living cell.
 - A strand of DNA is cut at a specific point and naturally existing cellular repair mechanisms fix the broken DNA strands.
 - The way they are repaired can affect gene function and new DNA sequences can be delivered when the DNA is cut and act as templates for generating an altered sequence.
 - Genome editing techniques can be used to delete sections of DNA or alter how a gene functions: for example, by changing a variant that may give rise to disease to one that functions normally.

CRISPR-Cas9 genome editing

Gene editing within a living cell

A technique called CRISPR-Cas9 is one method of what is called genome editing, allowing scientists to precisely repair or modify genes in the quest for new ways to treat or prevent inherited diseases.

Guide RNA

Scientists customize RNA molecules to guide an enzyme named Cas9 to the correct spot.

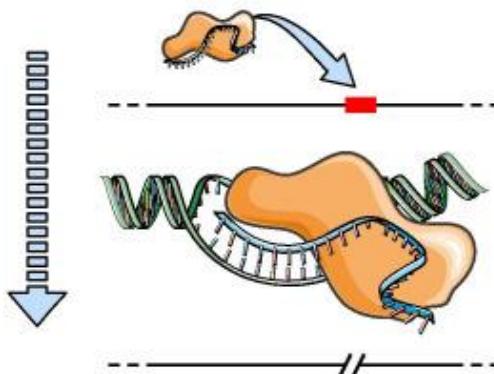
Cas9

The enzyme precisely cuts DNA like scissors. The cell repairs itself or a new piece of DNA can be inserted.

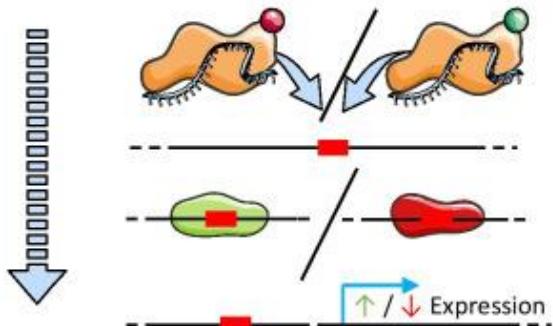


Studying the function of disease associated variants using CRISPR-Cas9 methods

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



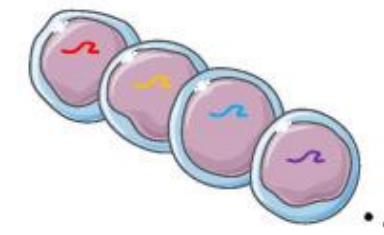
ii. Tiled gRNAs



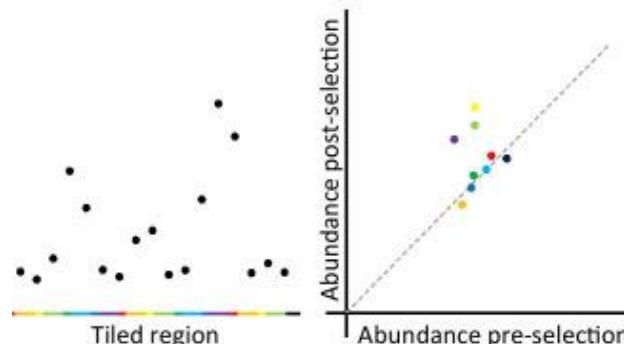
iii. Lentiviral particles



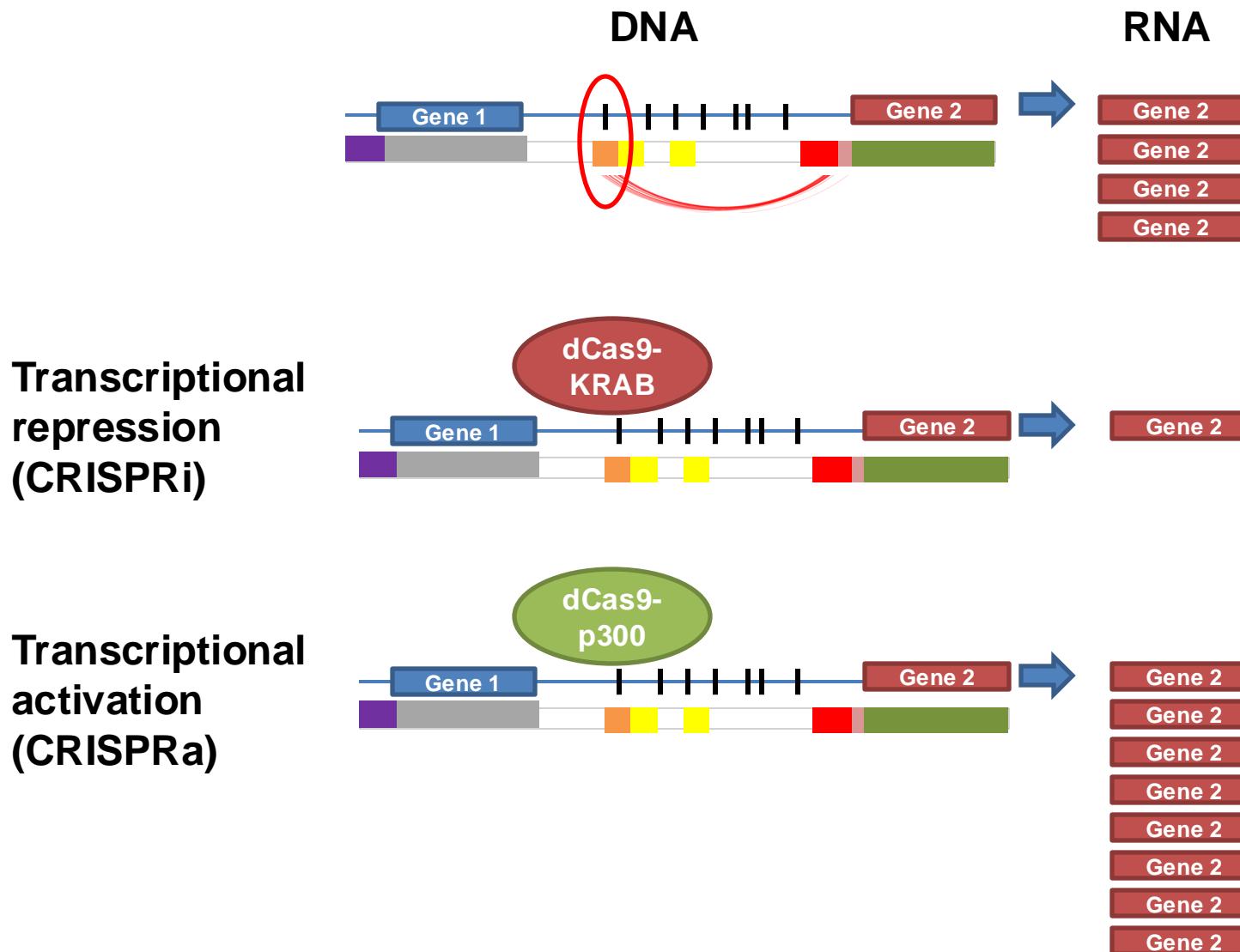
iv. Transduced cells



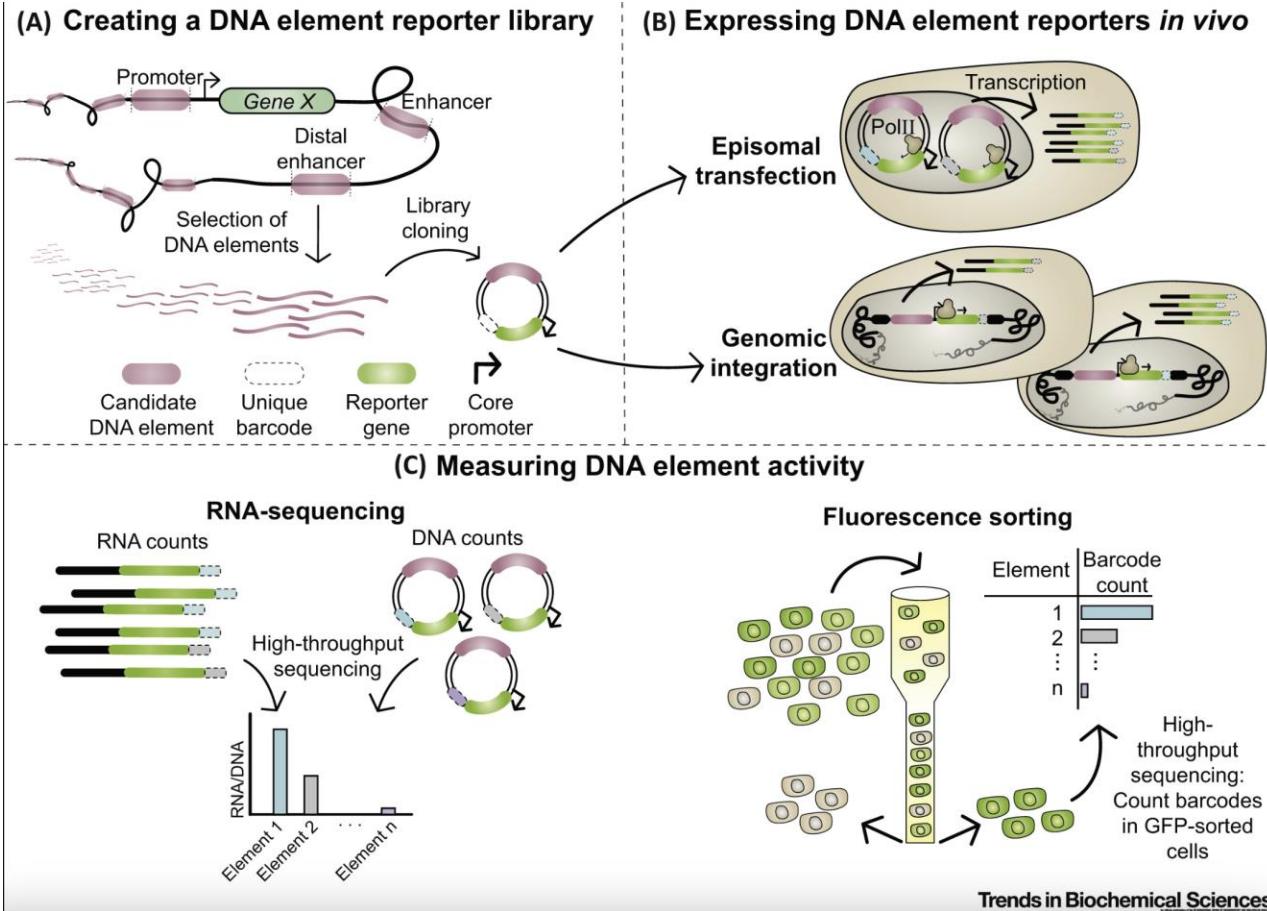
v. scRNAseq/selection and DNA sequencing



Functional validation: dCas9-mediated transcriptional modulation



Massively parallel reporter assays (MPRAs)



- MPRAs can measure the transcriptional activity of up to millions of DNA elements, such as putative enhancers containing GWAS variants.
- These elements are cloned in a reporter vector, with each element typically linked to a unique DNA barcode that is linked to a reporter gene (A).
- The activity of the elements is inferred from the abundance of the reporter gene in mRNA after *in vivo* expression (B).
- Alternatively, the reporter does not carry barcodes but encodes a fluorescent protein; cells are then sorted by fluorescence levels and the activity of the query elements is inferred from their abundance in the sorted cell pools (C).

Examples of application to complex diseases

Identification of Mechanisms by Which Genetic Susceptibility Loci Influence Systemic Sclerosis Risk Using Functional Genomics in Primary T Cells and Monocytes

David González-Serna,¹  Chenfu Shi,² Martin Kerick,¹ Jenny Hankinson,³ James Ding,² Amanda McGovern,² Mauro Tutino,³ Gonzalo Villanueva-Martin,¹  Norberto Ortego-Centeno,⁴ José Luis Callejas,⁴ Javier Martin,¹ and Gisela Orozco⁵



Chenfu Shi



Javier Martin

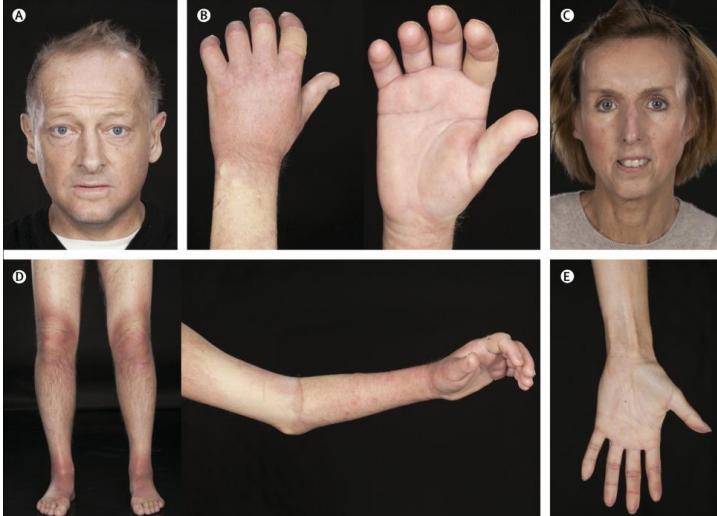


David Gonzalez Serna

ipbIn

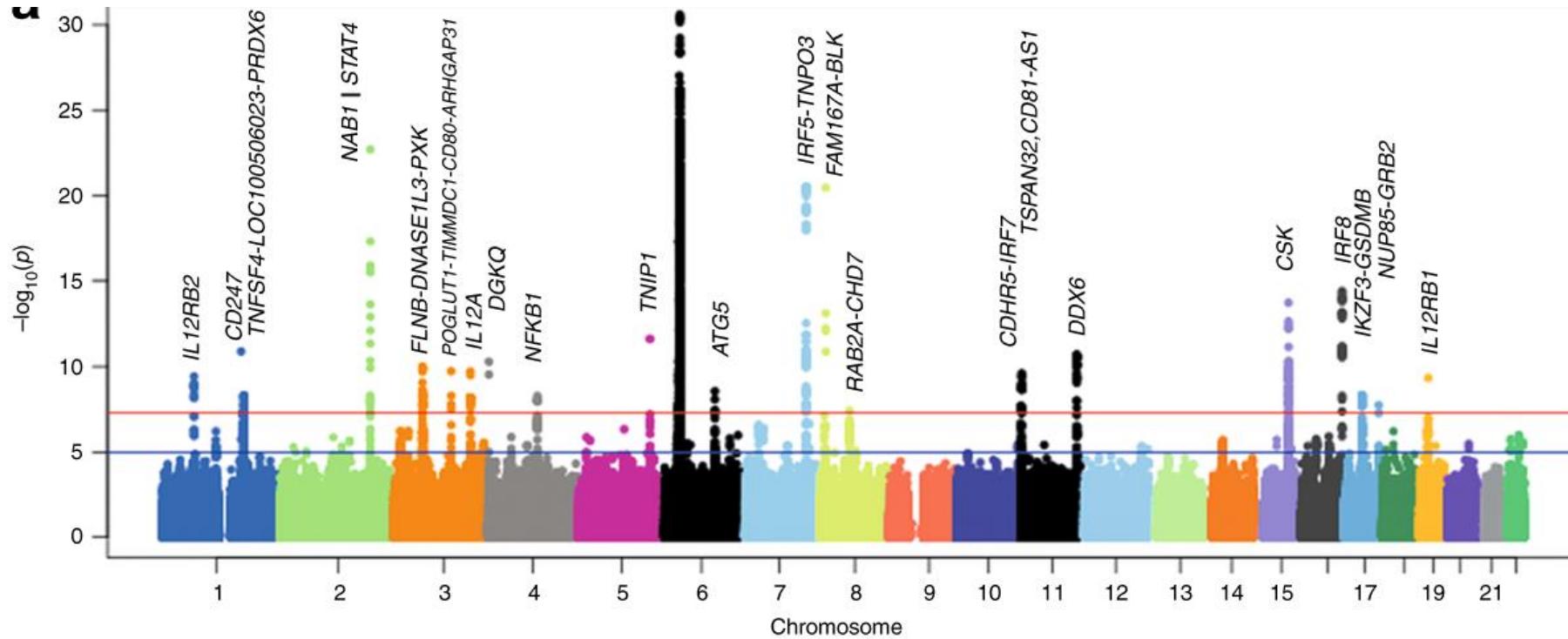
 **CSIC**
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Systemic Sclerosis (SSc) or scleroderma



- Complex, chronic, autoimmune disease
- Affects the connective tissue, characterized by an immune imbalance, vascular alterations, and an excessive collagen deposition leading to fibrosis
- It can affect the skin as well as internal organs
- Etiology is poorly understood
- There are no specific treatments available

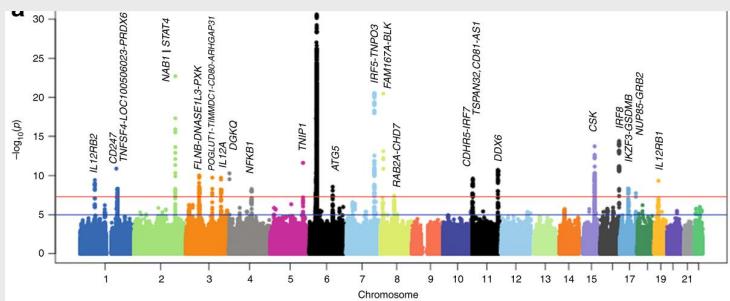
GWAS have identified 27 risk loci for SSc, but causal genes are unknown



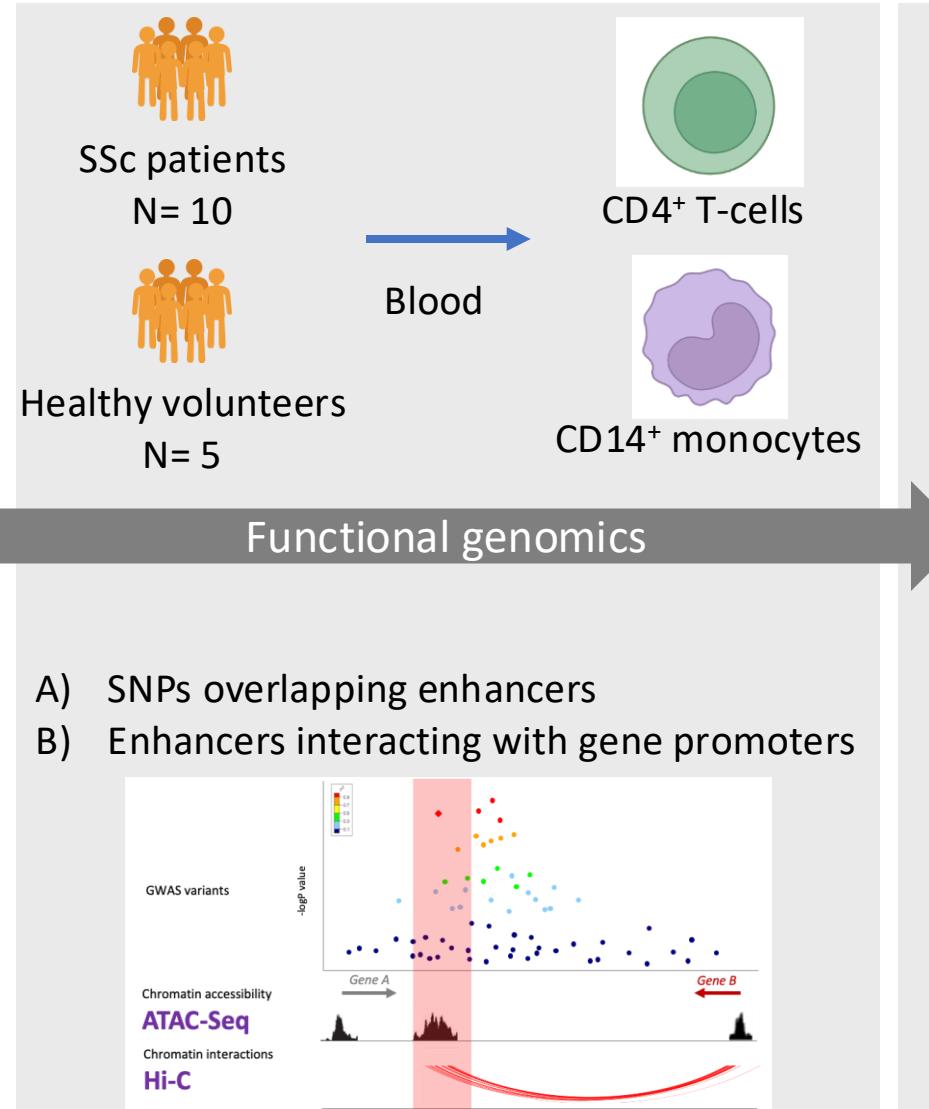
Elena Lopez-Isac, Javier Martin et al.
Nature Communications 2019

Identification of SSc causal genes: study design

GWAS



Lopez-Isac et al.
Nature Communications 2019



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

Shi, Serna et al.
Arthritis and Rheumatology, 2022

Linking differential expression and differential interactions in CD4⁺ T cells vs CD14⁺ monocytes

- Chromatin conformation is highly cell type specific
 - CD4⁺ T-cells vs CD14⁺ monocytes:
 - 71,213 significant differential interactions
- Differentially expressed genes are significantly enriched in differentially interacting genes, indicating how chromatin conformation is highly linked to gene expression
 - CD4⁺ T-cells vs CD14⁺ monocytes:
 - 2,257 strongly differentially expressed genes
- Our results demonstrate the importance of using different cell types to define promoter interactions, and how they are linked with the expression of important genes for those cell types.

Differential interactions and expression between SSc patients and healthy controls

- In contrast, we found that 3D chromatin structure is largely preserved between SSc patients vs Controls
 - CD4+ T-cells:
 - 4,858 modest significant differential interactions
 - 62 differentially expressed genes
 - CD14+ monocytes:
 - No differential interactions
 - 63 differentially expressed genes
- Subtle differences in chromatin conformation can lead to big changes in gene expression.
 - Subtle differences are difficult to detect: power, too much variability between samples and technical limitations of the methods

46 target genes for the 23 systemic sclerosis risk GWAS loci in CD4⁺ T cells and CD14⁺ monocytes: 39 new candidate genes and 7 confirmed genes

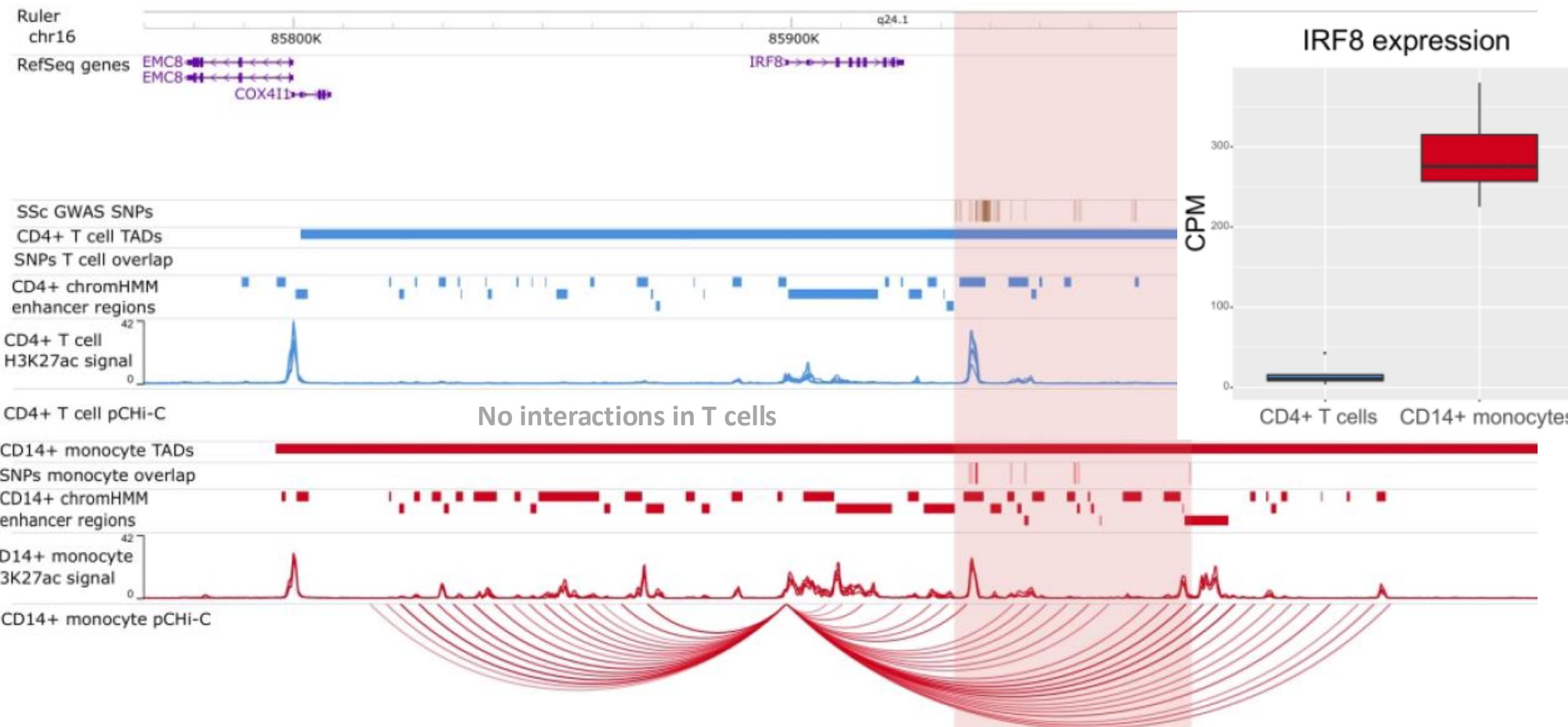
Chr	Bp (start - end) ¹	GWAS Locus ²	pCHi-C target genes ³		CD4 ⁺ vs CD14 ⁺	
			CD4 ⁺ T cells	CD14 ⁺ monocytes	Differential interactions	Differential expression
1	67326053 - 6744804	<i>IL12RB2</i>				
1	167445635 - 167465040	<i>CD247</i>	<i>CD247, CREG1</i>		<i>CD247, CREG1</i>	<i>CD247, CREG1</i>
1	173337507 - 173391947	<i>TNFSF4- LOC100506023 - PRDX6</i>				
2	190642047 - 190698201	<i>NAB1</i>	<i>MFSD6, NEMP2</i>	<i>MFSD6, NEMP2, HIBCH:INPP1</i>	<i>MFSD6, NEMP2, HIBCH:INPP1</i>	<i>MFSD6, NEMP2, HIBCH, INPP1</i>
2	191035723 - 191108308	<i>STAT4</i>	<i>STAT4, NABP1</i>		<i>STAT4, NABP1</i>	<i>STAT4, NABP1</i>
3	58084620 - 58482701	<i>FLNB -DNASE1L3-PXK</i>	<i>RPP14, KCTD6</i>	<i>RPP14, KCTD6</i>	<i>RPP14, KCTD6</i>	<i>KCTD6</i>
3	119384733 - 119546340	<i>POGLUTI-TIMMDC1-CD80- ARHGAP31</i>	<i>TMEM39A;POGLUT1</i>		<i>TMEM39A;POGLUT1</i>	<i>TMEM39A, POGLUT1</i>
3	160002484 - 160030580	<i>IL12A</i>		<i>SMC4;IFT80</i>	<i>SMC4;IFT80</i>	<i>SMC4, IFT80</i>
4	960523 - 990021	<i>DGKQ</i>	<i>GAK;TMEM175, FGFR1</i>	<i>GAK, TMEM175, FGFR1</i>	<i>FGFR1</i>	<i>GAK, TMEM175 FGFR1</i>
4	102477892 - 102615256	<i>NFKB1</i>	<i>SLC39A8, NFKB1, UBE2D3;CISD2, SLC9B1, BDH2</i>	<i>SLC39A8, UBE2D3;CISD2, BDH2</i>	<i>SLC39A8, NFKB1, UBE2D3;CISD2, SLC9B1, BDH2</i>	<i>SLC39A8, UBE2D3, CISD2, BDH2</i>
5	151064651 - 151080486	<i>TNIP1</i>				
6	106181815 - 106339294	<i>ATG5</i>				
7	128933913 - 129095960	<i>IRF5-TNPO3</i>				
8	11474517 - 11544554	<i>FAM167A-BLK</i>				
8	60638547 - 60664239	<i>RAB2A-CHD7</i>	<i>ASPH, SDCBP, CHD7</i>		<i>ASPH, SDCBP</i>	<i>ASPH, SDCBP, CHD7</i>
11	554659 - 619789	<i>CDHR5 -IRF7</i>				
11	2311894 - 2363262	<i>TSPAN32,CD81-AS1</i>	<i>TSSC4</i>		<i>TSSC4</i>	<i>TSSC4</i>
11	118704617 - 118875175	<i>DDX6</i>	<i>CXCR5, UPK2, DDX6, IFT46;ARCNI</i>	<i>CXCR5</i>	<i>CXCR5, UPK2, DDX6</i>	<i>CXCR5, DDX6, ARCNI</i>
15	74739180 - 75148328	<i>CSK</i>	<i>CSK, CLK3, ULK3, SCAMP2, MPI, FAM219B, COX5A</i>	<i>CSK, CLK3, ULK3, SCAMP2, MPI, FAM219B, COX5A, C15orf39</i>	<i>CSK, ULK3, SCAMP2, MPI, FAM219B, COX5A, C15orf39</i>	<i>CSK, CLK3, ULK3, SCAMP2, MPI, FAM219B, COX5A, C15orf39</i>
16	85932852 - 85979945	<i>IRF8</i>		<i>IRF8</i>	<i>IRF8</i>	<i>IRF8</i>
17	39747478 - 39933464	<i>IKZF3-GSDMB</i>	<i>IKZF3, ERBB2, PSMD3</i>	<i>ERBB2</i>	<i>IKZF3, ERBB2, PSMD3</i>	<i>IKZF3, ERBB2, PSMD3</i>
17	75193533 - 75279345	<i>NUP85-GRB2</i>				
19	18068862 - 18093031	<i>IL12RB1</i>	<i>PIK3R2, RAB3A</i>	<i>RAB3A</i>	<i>PIK3R2, RAB3A</i>	

Genes classically associated with systemic sclerosis through proximity to GWAS loci are highlighted in bold.

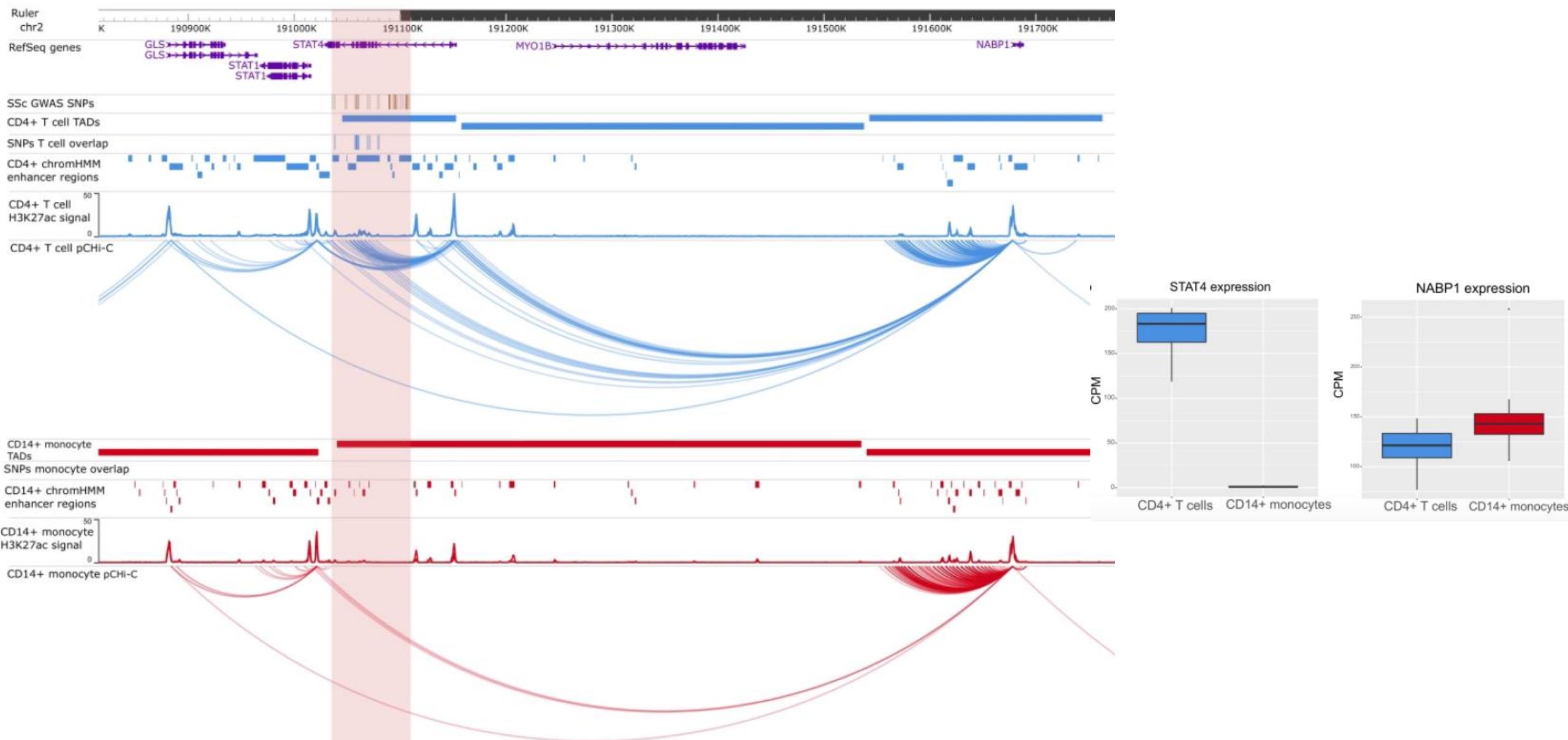
¹Bp in GRCh38 (hg 38) assembly

²Locus as defined by López-Isac et al.(9)

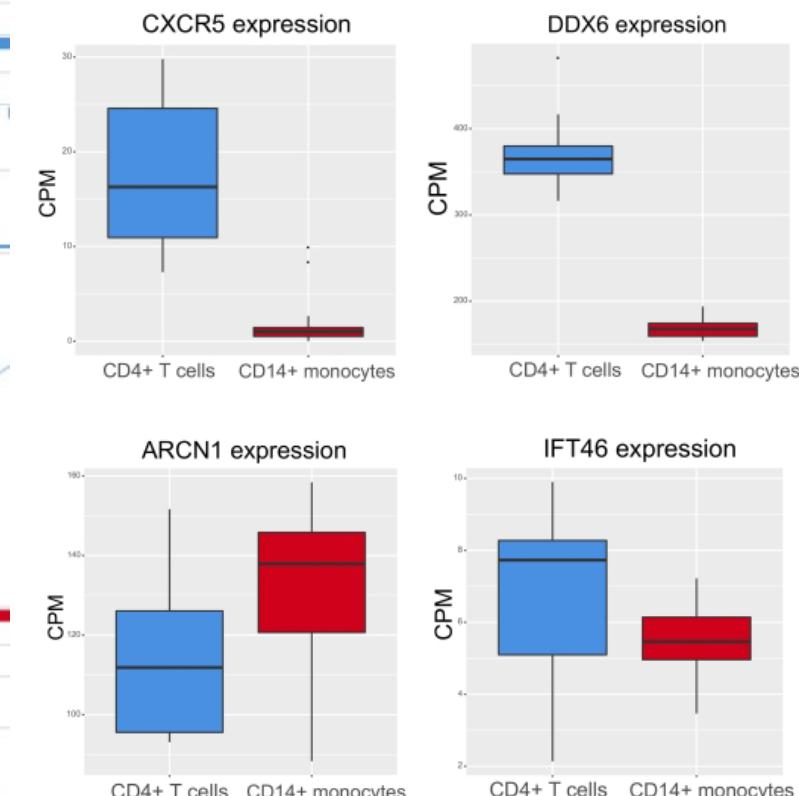
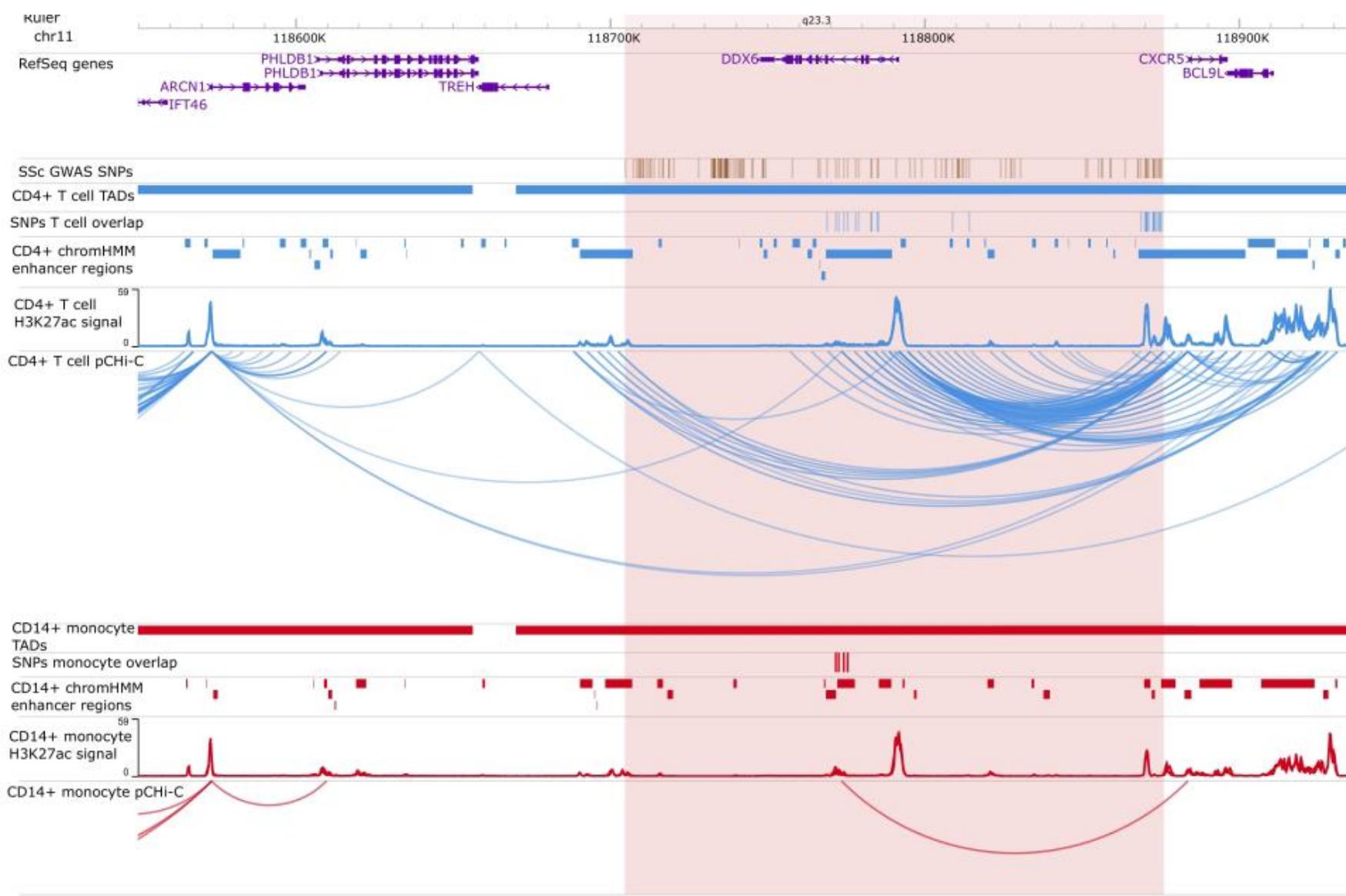
IRF8 is a monocyte specific SSc gene



STAT4 is a T-cell specific SSc gene



Novel SSc candidate genes: *DDX6*, *CXCR5*, *UPK2*, *IFT46/ARCN1*



Genomics driven drug repurposing

- Genetically validated drug targets are more likely to lead to approved drugs
 - Nelson et al, Nature Genetics, 2015
 - King et al, PLoS Genetics, 2019
- Disease genes identified using GWAS are correlated with targets of known drugs
 - RA: TNF inhibitors, JAK inhibitors (Okada et al, Nature, 2014)
- GWAS findings have sparked the successful repositioning of drugs
 - IL-12/IL-23 in PsA and psoriasis

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

<https://platform.opentargets.org>
<https://www.drugbank.com>

[§]Only related immune-mediated diseases were included. All clinical trials at least in completed phase III.

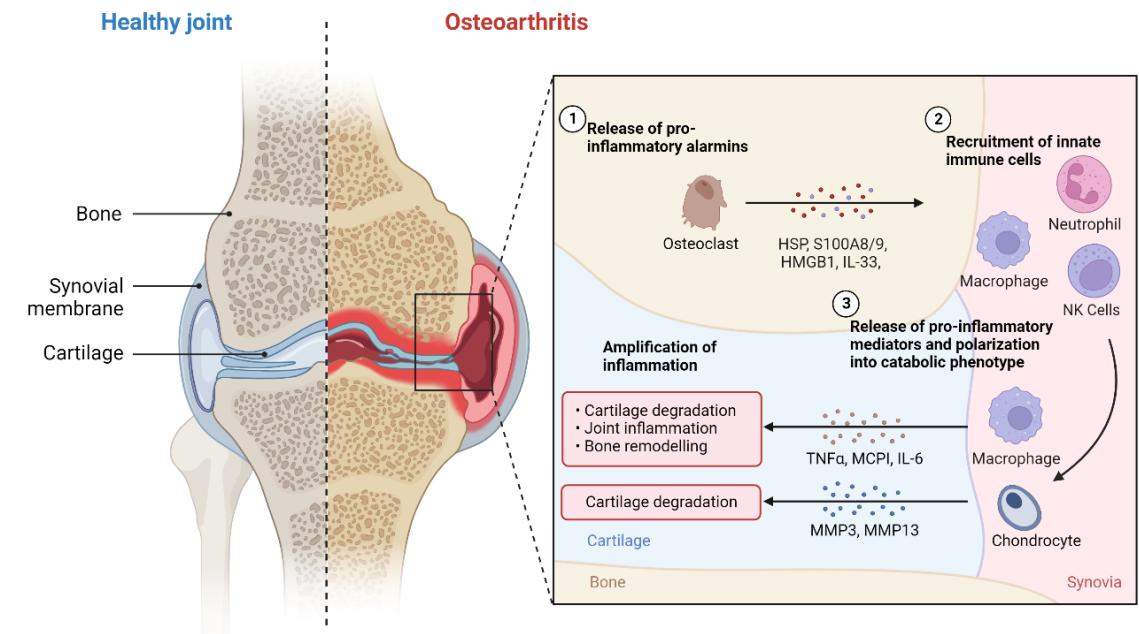
*These drugs present phase III or lower clinical trials in systemic sclerosis.

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 [§]
<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>		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	<i>KEAP1</i>	Dimethyl Fumarate	Psoriasis, Multiple sclerosis, Disseminated sclerosis
			<i>IMPDH1</i>	Mycophenolic acid*	Systemic lupus erythematosus, Immunosuppression
			<i>TUBB4B</i>	Colchicine	Osteoarthritis, Advanced fibrosis
<i>RAB2A-CHD7</i>	<i>SDCBP</i>	CD4+ T cells	<i>PPARG</i>	Mesalamine	Crohn's disease, Ulcerative colitis
	<i>CHD7</i>	CD4+ T cells	<i>SIPR3</i>	Fingolimod	Multiple sclerosis, Disseminated sclerosis
<i>CSK</i>	<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>		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>IL12RB1</i>	<i>PIK3R2</i>	CD4+ T cells	<i>ADRA1B</i>	Epinephrine	Crohn's disease
			<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>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>RAB3A</i>		CD4+ T cells, CD14+ monocytes	<i>HSPA8</i>	Forigerimod	Systemic lupus erythematosus

Osteoarthritis (OA)

- Most common MSK disorder:
 - 300M people worldwide
 - 40% of people over the age of 70 affected
- Characterized by structural damage and inflammation of the joints
- Leads to pain and disability/decreased mobility
- Current treatment approaches focus on pain relief, physiotherapy and joint replacement for end- stage disease
 - No targeted treatments available





OPEN ACCESS

TRANSLATIONAL SCIENCE

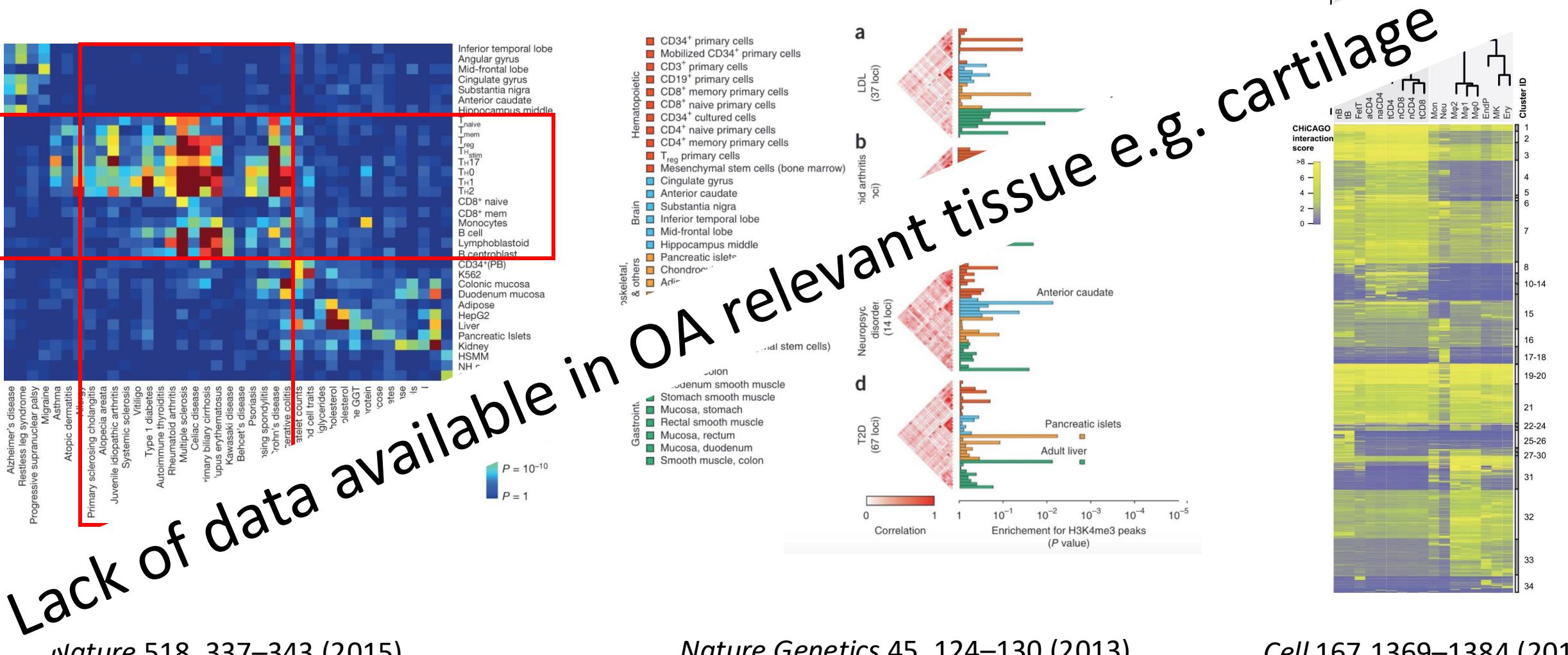
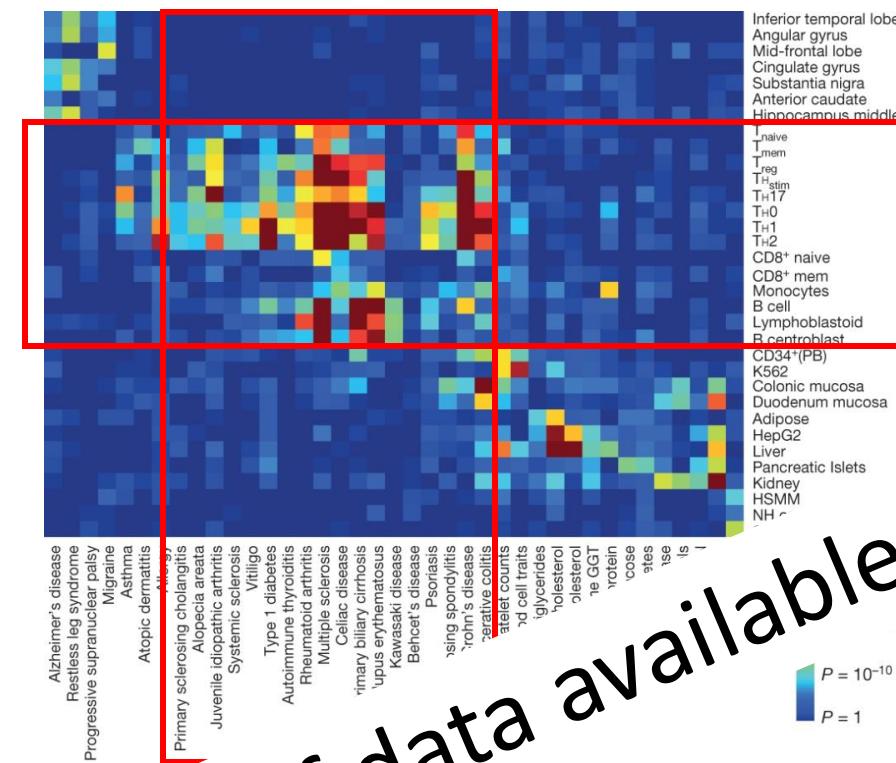
Primary osteoarthritis chondrocyte map of chromatin conformation reveals novel candidate effector genes

Norbert Bittner ,¹ Chenfu Shi,² Danyun Zhao,² James Ding,² Lorraine Southam,¹ Diane Swift,³ Peter Kreitmaier,^{1,4,5} Mauro Tutino,¹ Odysseas Stergiou,¹ Jackson T S Cheung,⁶ Georgia Katsoula,^{1,4,5} Jenny Hankinson,¹ Jeremy Mark Wilkinson,⁷ Gisela Orozco,^{2,8} Eleftheria Zeggini  ^{1,5}

Bittner N, et al. *Ann Rheum Dis* 2024;0:1–12. doi:10.1136/ard-2023-224945

- First whole genome chromosome conformation analysis (Hi-C) map of primary OA chondrocytes
- Aim: To identify novel candidate effector genes for the disease

The need for data in primary OA tissue



- 9 OA samples
- Arima Hi-C kit
 - 1M chondrocytes/sample
- 450 million reads per library

Analysis workflow

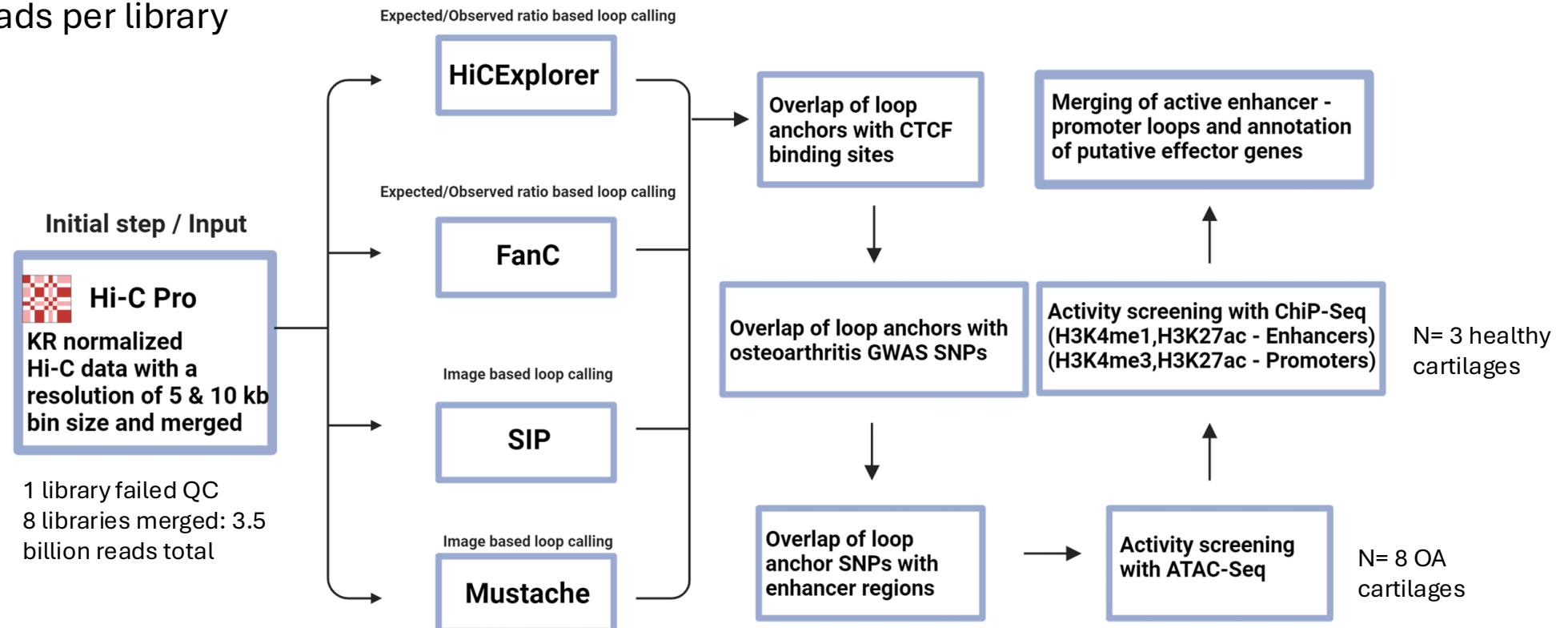


Figure 1 Workflow for the analysis of putative effector genes. Hi-C data from primary tissue chondrocytes with a 5 kb and 10 kb bin size were Knight-Ruiz (KR) normalised, merged and analysed separately with each of the four applied algorithms. Loop anchors were overlapped with CTCF binding sites as a quality control measure (table 1). Osteoarthritis genome-wide association studies (GWAS) single-nucleotide polymorphisms (SNPs)³ were overlapped with loop anchors and enhancer regions from ENCODE SCREEN³⁰ (V.3). Activity screening for SNP-containing enhancers was performed with overlap of public ATAC-seq¹¹ and ChIP-seq data¹². Active promoters on loop anchors contacting active SNP-enhancer regions were identified with the same epigenetic datasets. Overlapping enhancer-promoter loops were merged and putative effector genes were annotated. Figure was created with BioRender.com.

Heritability enrichment

- Are the identified loops from chondrocytes informative for the functional annotation of the osteoarthritis GWAS signals?
 - We applied stratified LD score regression (S-LDSC) to all 11 OA phenotypes, as reported in the largest OA GWAS meta-analysis to date (Boer et al, Cell, 2021)
 - No enrichment for all ENCODE enhancers (expected)
 - Strong significant enrichment in active enhancers overlapping loop anchors in chondrocytes



Using the promoter-enhancer contact matrix to functionally annotate the GWAS signals

- 345 variants within 472 loop-anchor regions associated with 77 GWAS signals
- 14 variants are found within active enhancers of 41 loop anchors
- 10 variants were located within 17 active enhancer- promoter loops
- 3 variants were present in an enhancer-enhancer loops
- 9 loop anchors which do not overlap with any cCRE element but show active epigenetic marks
 - This may indicate novel, not yet identified enhancer elements specific to chondrocytes.

OA associated variants in active enhancers interacting with active promoters in chondrocytes

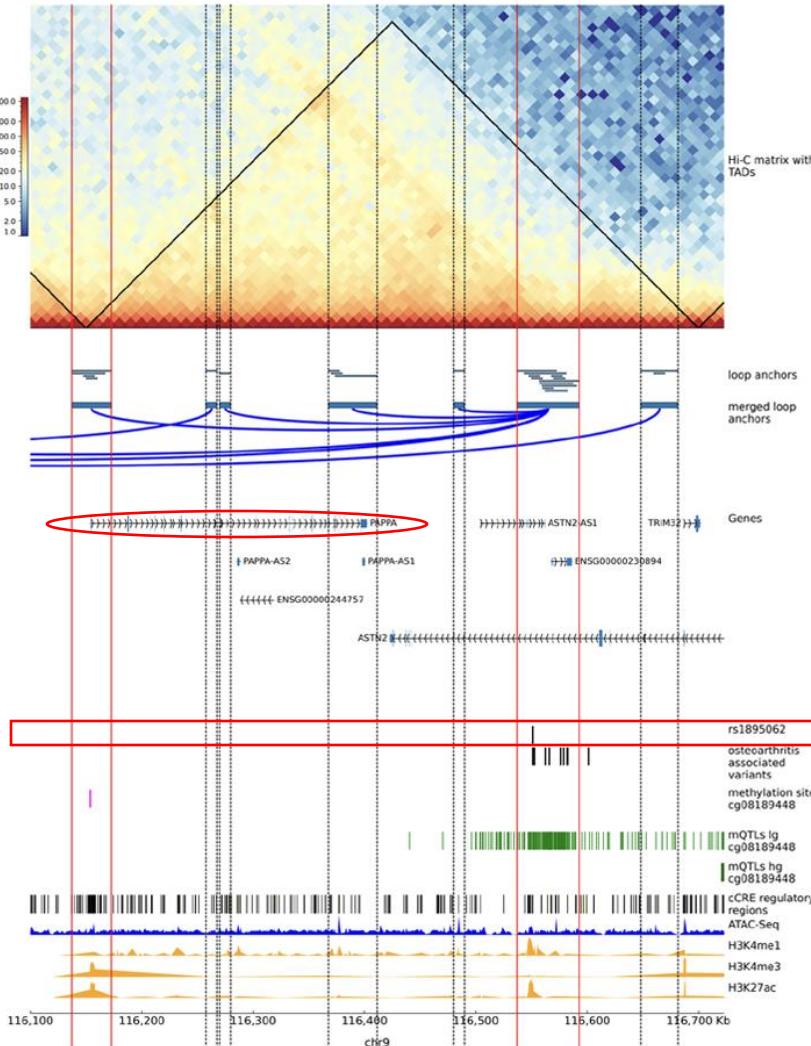
Lead variant (rsID)	rs2276749	rs8112559	rs143083812	rs12908498	rs10405617	rs10062749	rs1321917
Phenotype	THR	Hand OA Finger OA	THR/Hip OA	Hip OA/THR/TJR Knee-Hip OA/All OA	All OA/Knee OA	Hand OA/Thumb OA Knee OA	THR/Hip OA/TJR
Variant position	3:11601991	19:45887197	7:129203569	15:67074150	19:10642292	5:142425523	9:116562650
Credible set variants	rs2276749 rs6799718	rs8112559	rs143083812	rs1498506 rs1498507	rs10984	rs28538668 rs6861056	rs1895062
Annotated effector gene	<i>VGLL4</i>	<i>APOE3</i>	-	<i>SMAD3</i>	<i>ILF3</i> <i>SMARCA4</i>	<i>NR3C1</i>	-
Putative effector gene	-	-	-	-	<i>SLC44A2*</i>	<i>SPRY4</i>	<i>PAPPA</i>
Methylation and gene targets (QTLs)	-	-	-	cg09501821 eQTL (SMAD3)	cg01654627, cg17710535 eQTL (SLC44A2)	cg19514721	cg08189448
Transcription factor affected	-	IRF, Stat		Hic1, ATF3, ATF6 E2F, HEY1, Pax-4	AP-1, ATF3, E2F INSM1, Jundm2	Foxd1	AP-1, GATA, Pou2

Risk-associated phenotypes and lead variants were identified in the study by Boer *et al.*³ Lead variants are defined as the most significantly associated SNV for each of the credible set variants. Credible set variants are variants residing in the 95% credible sets, with a posterior probability of causality >3%. Annotated effector genes were genes identified to be associated with the lead variant in the study by Boer *et al.* Putative effector genes were identified in this study. Methylation targets and its QTLs were identified in a study by Kreitmaier *et al.*⁴¹ and gene targets and their eQTLs in a study by Steinberg *et al.*³⁹ Transcription factors affected by the credible set variant were identified with HaploReg V.4.2.^{43 44}

* Colocalisation has been previously observed in the study by Boer *et al* in Genotype-Tissue Expression data, here we add additional support for SLC44A2.

All OA, osteoarthritis at any joint site; eQTL, expression QTL; Finger OA, finger osteoarthritis; Hand OA, hand osteoarthritis; Hip OA, hip osteoarthritis; Knee-Hip OA, knee and/or hip osteoarthritis; Knee OA, knee osteoarthritis; QTL, quantitative trait loci; SNV, single nucleotide variant; THR, total hip replacement; TJR, total joint replacement.

Chr9, lead variant rs1321917 : rs1895062 – *PAPPA*

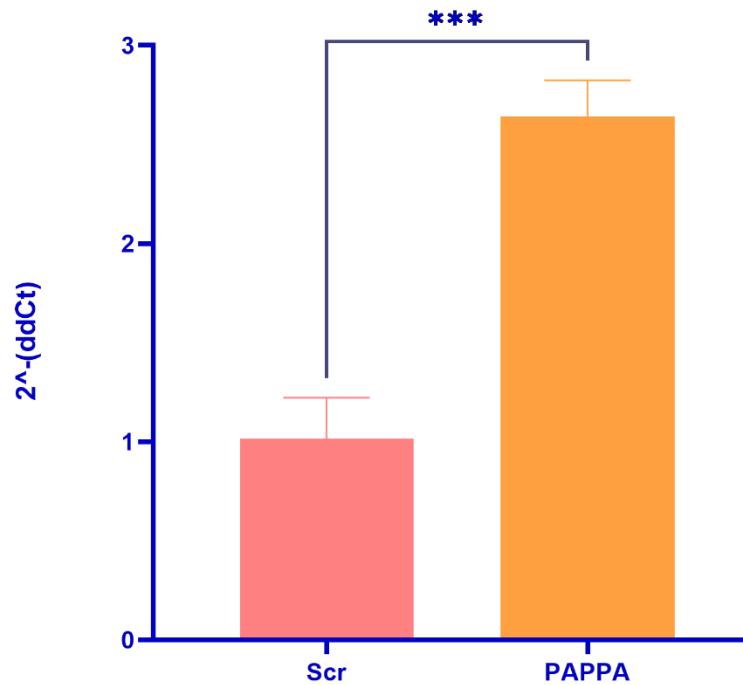


- No high confidence effector genes had been identified.
- One SNP from the credible set, rs1895062, is in an active enhancer, interacting with a single gene:
 - *PAPPA* (pregnancy-associated plasma protein A):
 - *PAPPA* is upregulated in osteoarthritis affected cartilage
 - Cleaves the insulin growth factor 1 transporter proteins IGFBP4 and IGFBP5 and releases IGF- 1 to its receptor
 - high levels of PAPPA result in increased cleavage of IGFBP5
 - Increased levels of intact IGFBP5 result in an improvement in joint architecture during development of osteoarthritis in dogs
 - In PAPPA- null mice, there was marked increase rate of bone formation
 - rs1895062 is a methQTL in low- grade cartilage for cg08189448 and cg08189448: GWAS signal and methQTL signal colocalise
 - rs1895062 is predicted to disrupt motifs of three different transcription factors.

Follow up – *PAPPA* locus

CRISPR validation *PAPPA* locus

Significantly increased *PAPPA* gene expression following CRISPR Activation targeting SNP rs1895062 in SW1353 chondrocyte cell line



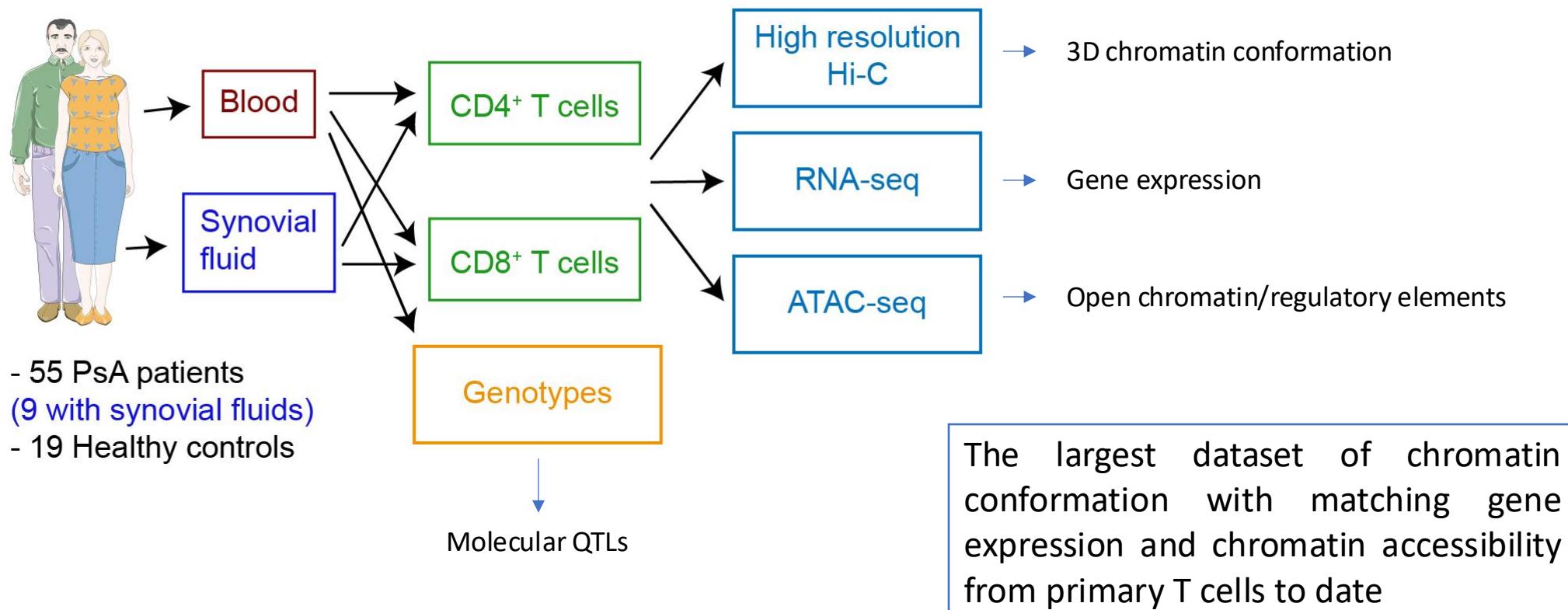
Psoriatic arthritis (PsA)

- Autoimmune disease:
 - Inflammation and destruction of the joints
 - Skin involvement: Psoriasis
- Cause not completely understood:
 - No cure
 - Lack of targeted treatments
 - Many patients do not respond to available therapies

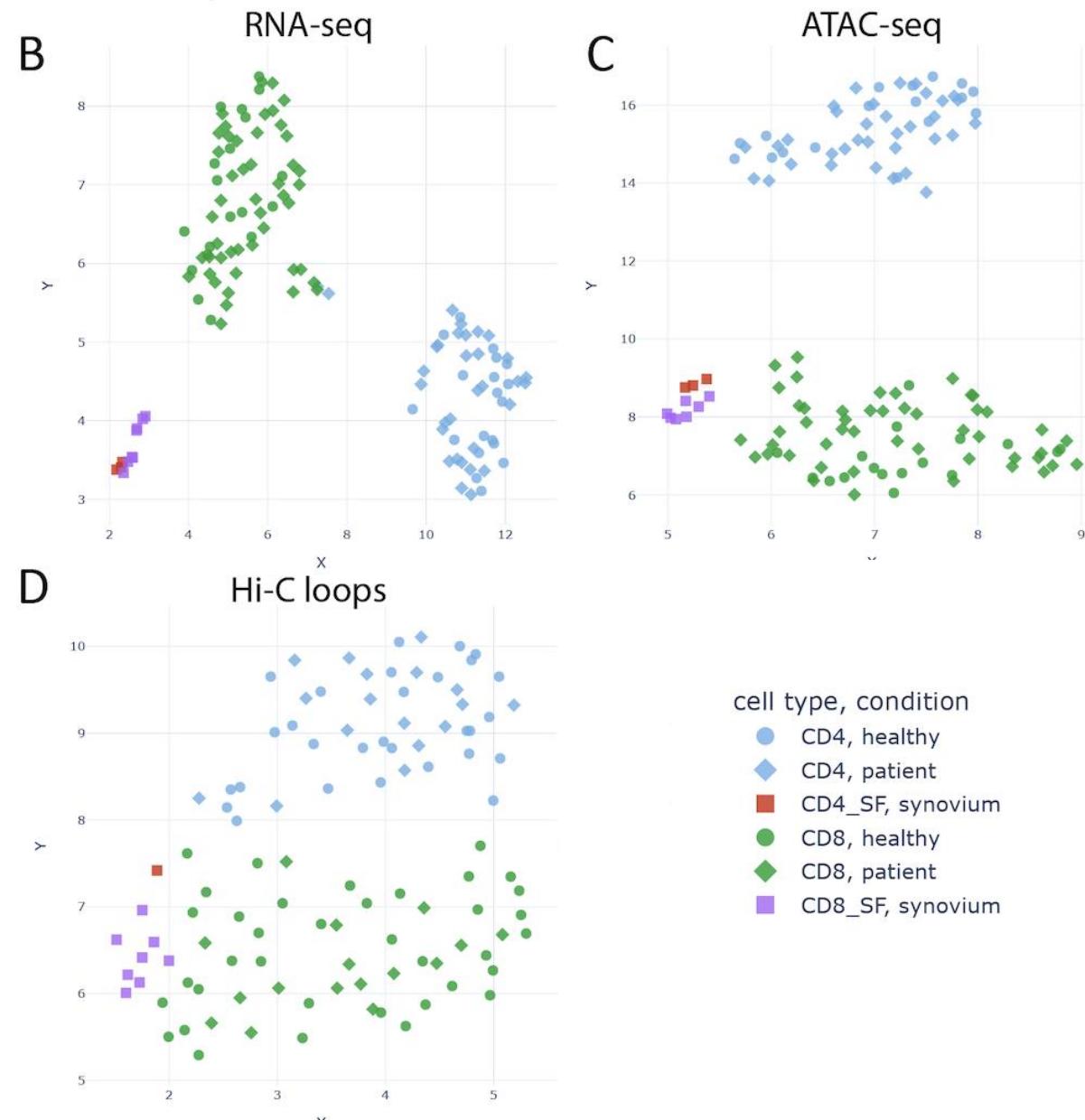


Multi-omics analysis in primary T cells elucidates mechanisms behind disease associated genetic loci

Shi et al, medRxiv 2023.07.19.23292550



A compendium of functional genomics data in primary T cells



For all data types, samples separate by cell type and location

Differentially expressed genes and differentially bound ATAC-seq peaks

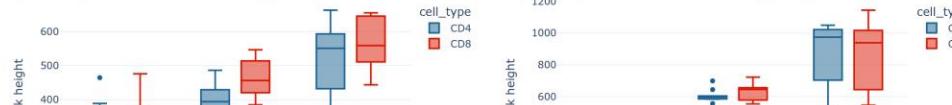
Comparison	Differentially expressed genes	Differentially bound ATAC-seq peaks
Active vs non-active disease CD8 ⁺ T cells	38	3
Active vs non-active disease CD4 ⁺ T cells	166	1463
Active disease vs controls CD8 ⁺ T cells	437	127
Active disease vs controls CD4 ⁺ T cells	110	116
CD8 ⁺ vs CD4 ⁺ T cells	10109	43648
CD8 ⁺ SF vs CD8 ⁺ Active disease Blood	12709	30079
CD4 ⁺ SF vs CD4 ⁺ Active disease Blood	11536	37411

No significant differential looping between groups

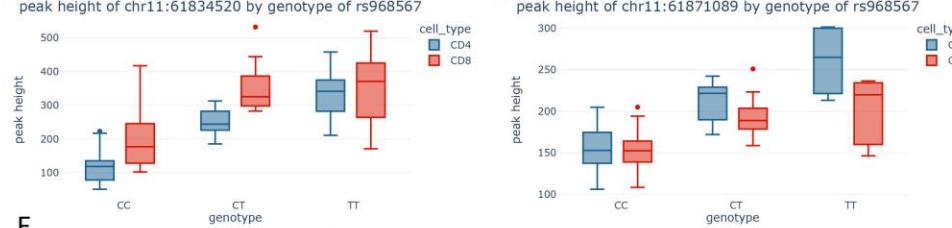
Genetic variation is strongly associated with alterations in gene expression, chromatin accessibility and chromatin conformation

QTL methodology	Number of significant QTLs (phenotypes)
CD8 ⁺ T cells eQTL	3610
CD4 ⁺ T cells eQTL	3150
CD8 ⁺ T cells caQTL	21,033
CD4 ⁺ T cells caQTL	19,342
CD8 ⁺ T cells insQTL	26,021
CD4 ⁺ T cells insQTL	24,807
CD8 ⁺ T cells loopQTL	12,307
CD4 ⁺ T cells loopQTL	11,595

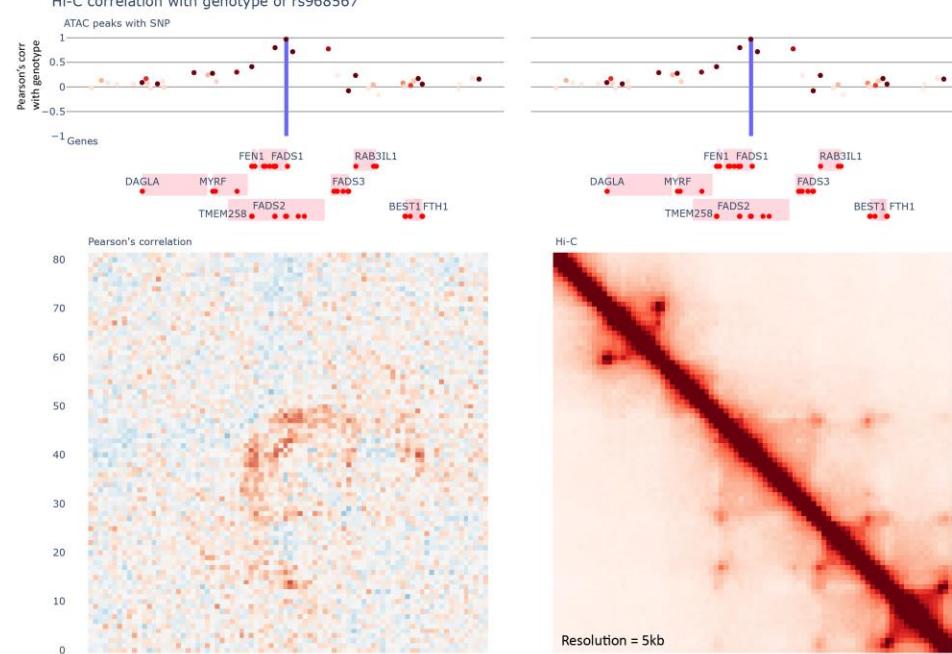
A peak height of chr11:61816418 by genotype of rs968567



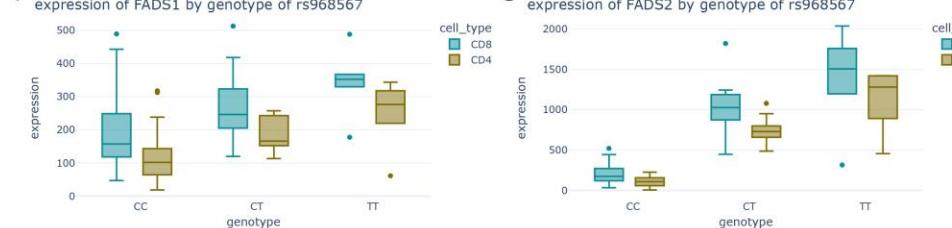
C peak height of chr11:61834520 by genotype of rs968567



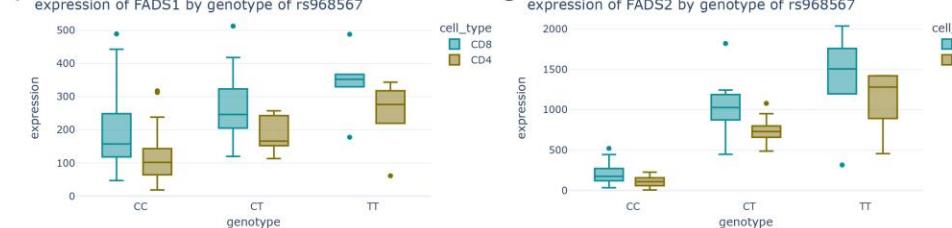
E Hi-C correlation with genotype of rs968567



F expression of FADS1 by genotype of rs968567



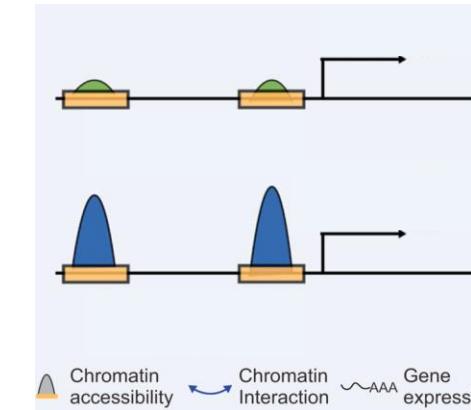
G expression of FADS2 by genotype of rs968567



Mechanism: ALLELE SPECIFIC EFFECTS

RA locus on chr 11 tagged by lead SNP rs7943728: Intergenic, 4 SNPs in LD block

- rs968567 is the only SNP in the LD block that overlaps a chromatin accessibility region (enhancer): **identification of putative causal variant**
- Protective allele leads to **INCREASED chromatin accessibility** → increased enhancer activity (4 ATAC peaks)
- The same allele is correlated with **INCREASED chromatin interactions** with *FADS2*, *FADS1*, *FADS3* and *FTH1*
- The same allele is correlated with **HIGHER expression** of *FADS2* and *FADS1*



Risk

Protection

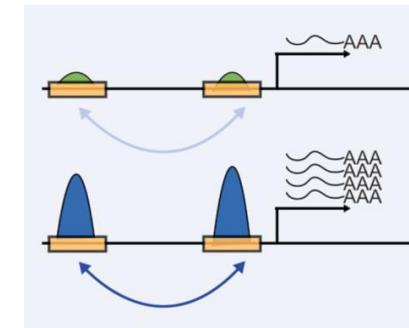
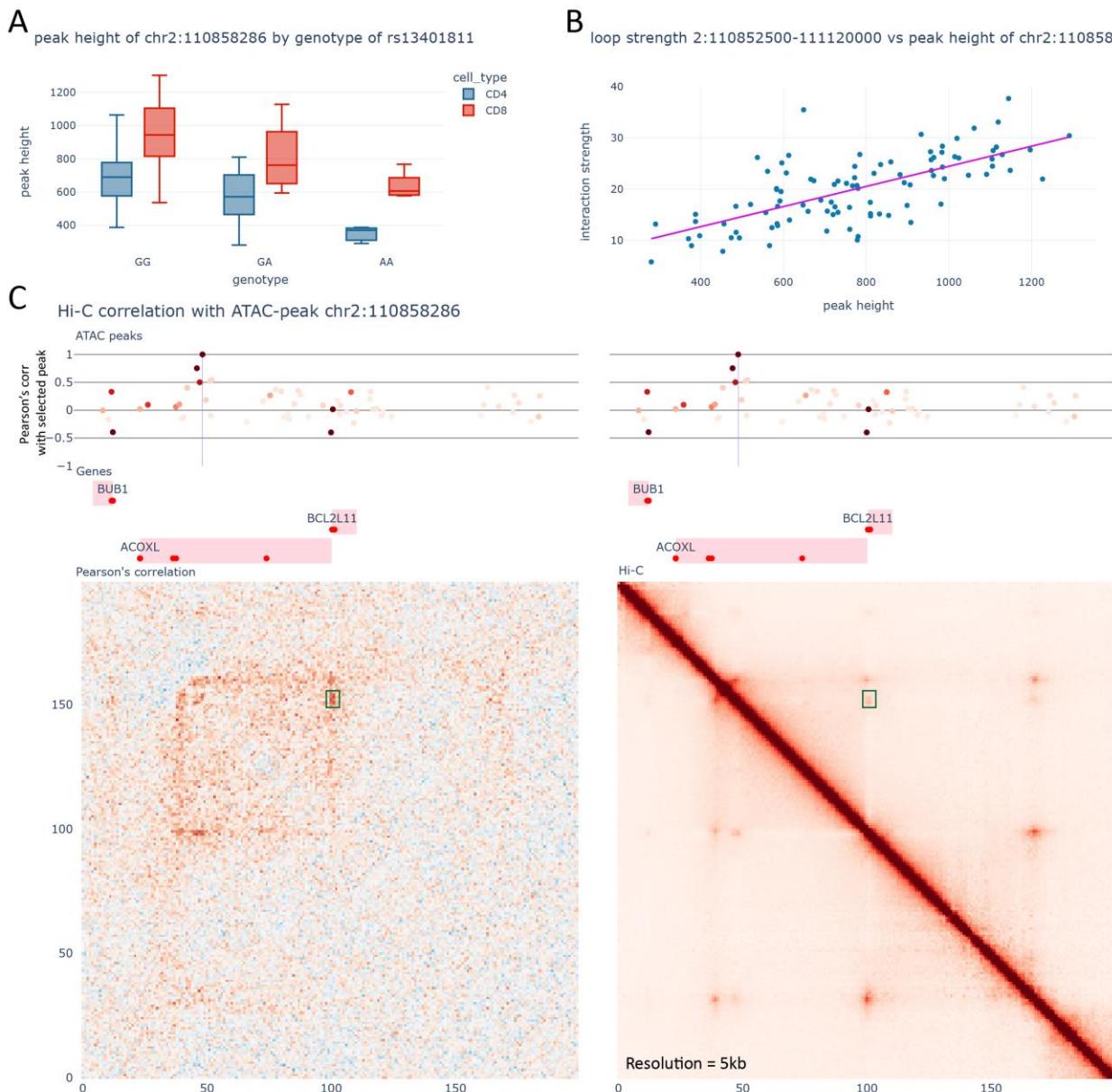
- These genes are involved in the synthesis of omega-3 and omega-6, known to influence inflammation and immune response mechanisms pivotal in the disease's pathogenesis.
 - Potential drug targets

Novel causal gene: *BCL2L11*

RA locus on chr 2 tagged by lead SNP rs13396472:

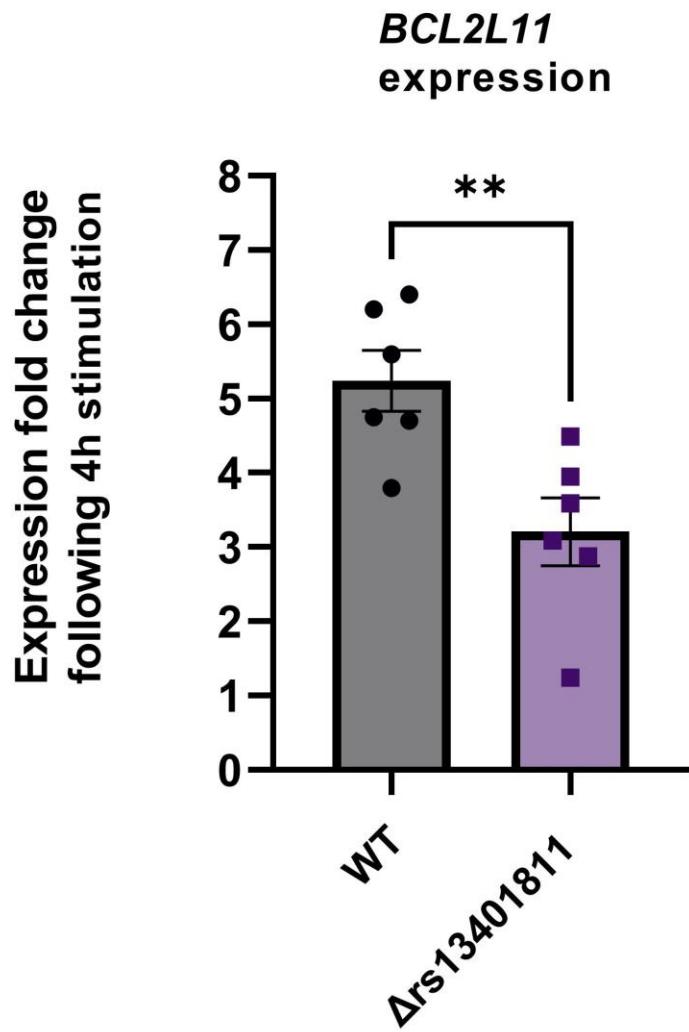
8 SNPs in LD block, overlap *ACOXL* gene

- Intronic rs13401811 is the only SNP in the LD block that overlaps a chromatin accessibility region (enhancer): **identification of putative causal variant**
- rs13401811 interacts with *BCL2L11*, located more than 300Kb away: **identification of novel causal gene**
- Risk allele leads to 40% increase in chromatin accessibility, and increased chromatin interactions with the promoter of *BCL2L11*



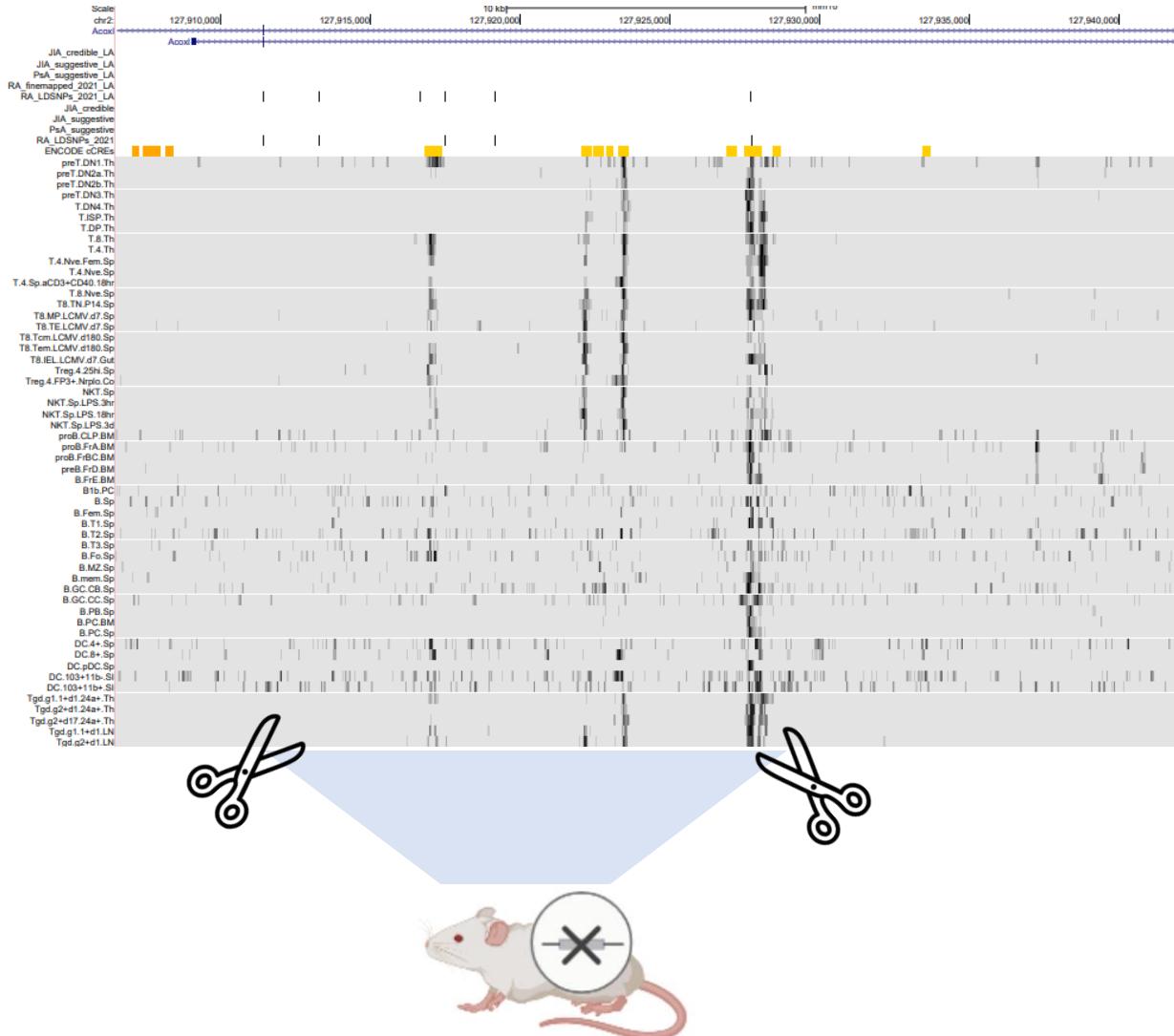
- *BCL2L11* has a critical role within the immune system, acting as a pro-apoptotic stimulator and modulating thymic negative selection

Validation using CRISPR/Cas9 in primary human CD4+ T cells



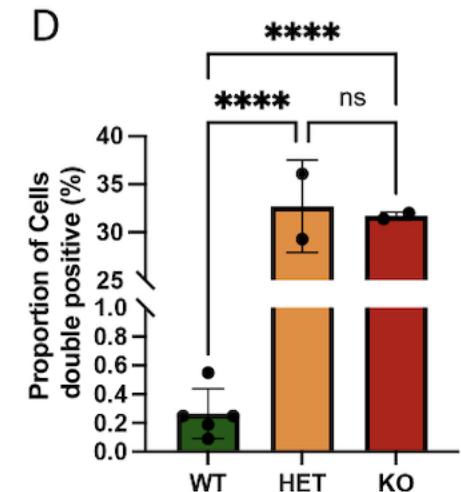
- Deletion of ~200bp encompassing the rs13401811 ATAC-seq peak.
- We examined the change in expression of *BCL2L11* following a 4-hour stimulation with CD3/CD28 beads.
- We observed a marked decrease in the upregulation of *BCL2L11* in the CD4+ T cells following stimulation (*p* value 0.008) in the enhancer knock out T-cells compared to wild type T-cells.
- This result supports the role of the rs13401811 enhancer region in regulating *BCL2L11* expression CD4+ T cells.
- Knockout mice for *BCL2L11* display progressive autoimmune disease. But what happens if we KO the enhancer in mice?

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)**.

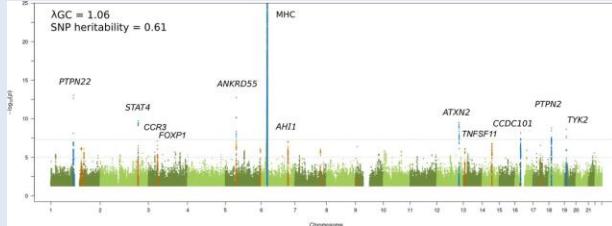
Juvenile idiopathic arthritis (JIA)



- JIA has been understudied:
 - Rare
 - Availability of biological materials
- Recent advances in functional genomics methods allow low input samples
 - Mikhail Spivakov – Imperial College London

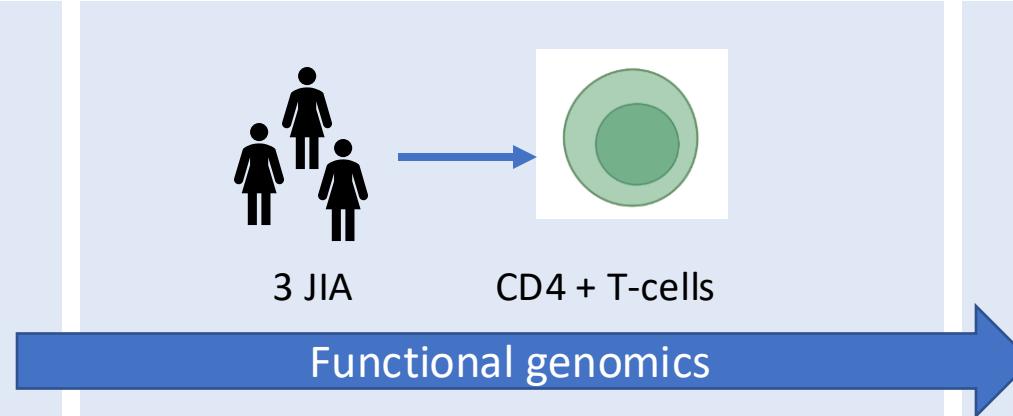
Functional genomics in JIA: linking SNPs to potential causal genes

GWAS



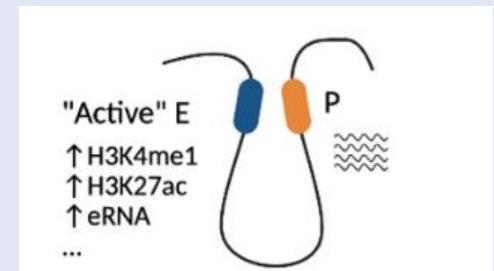
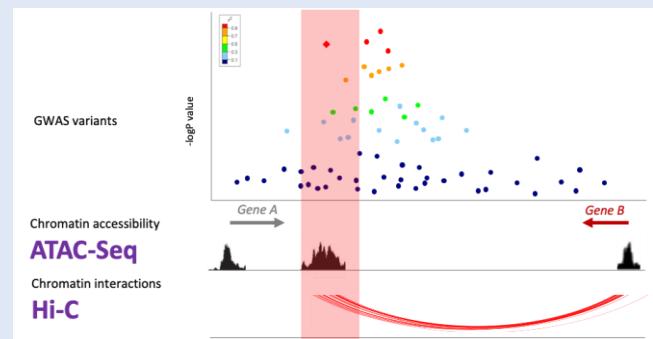
López-Isac et al. *Ann Rheum Dis* 2021

44 risk loci ($p < 10^{-5}$)



19 loci linked to
61 target genes

- A) SNPs overlapping T-cell enhancers
 - ATAC-seq
- B) Enhancers interacting with gene promoters
 - Low input pCHi-C



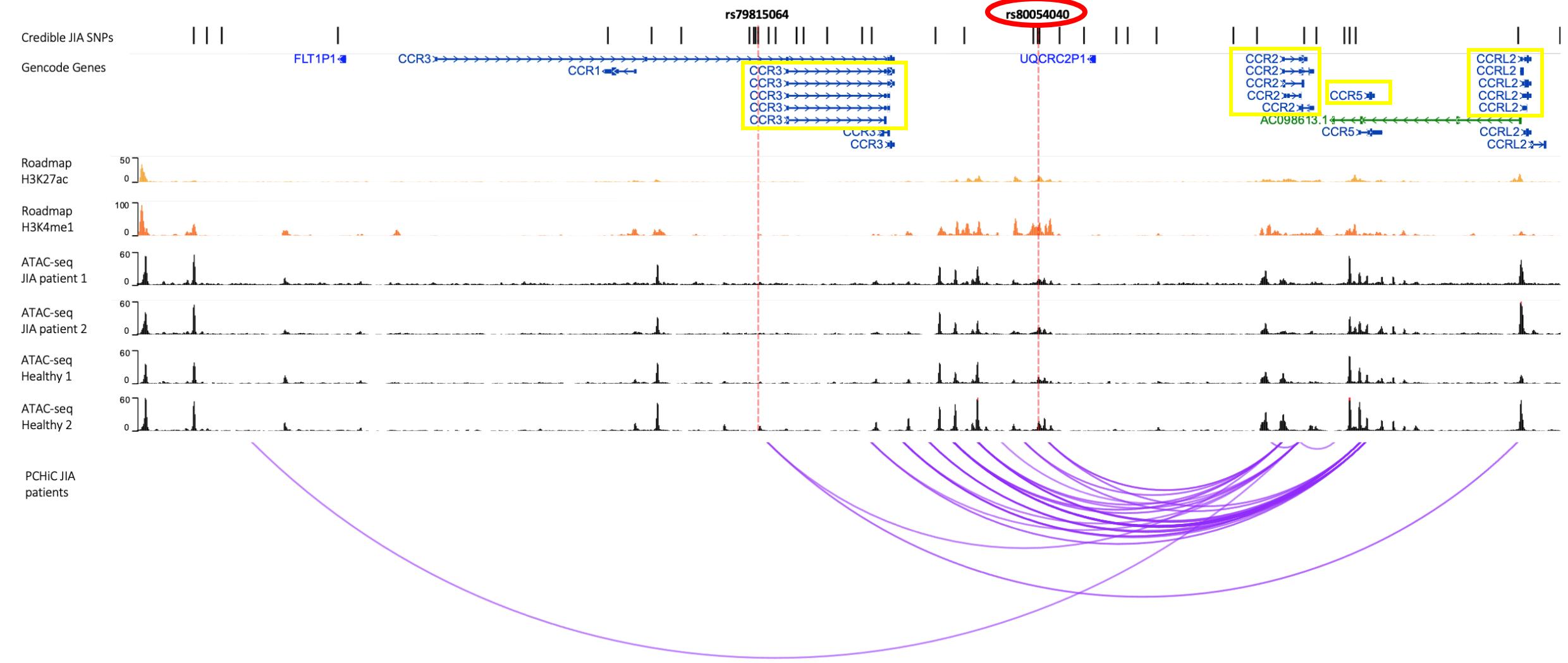
Frantzeskos A et al.
Manuscript in preparation

Gene and SNP prioritization for validation

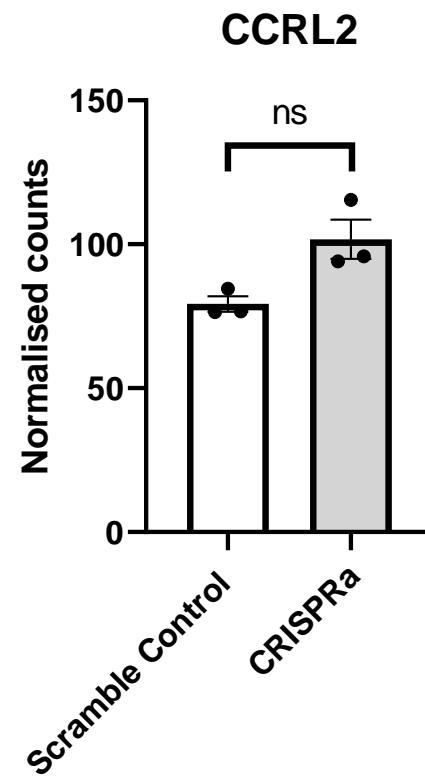
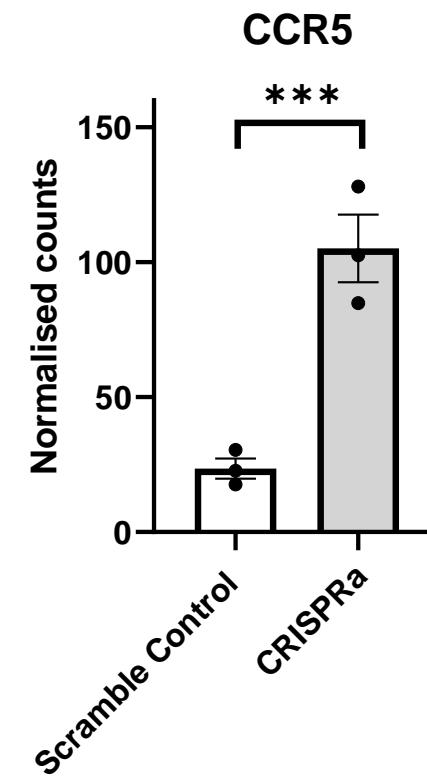
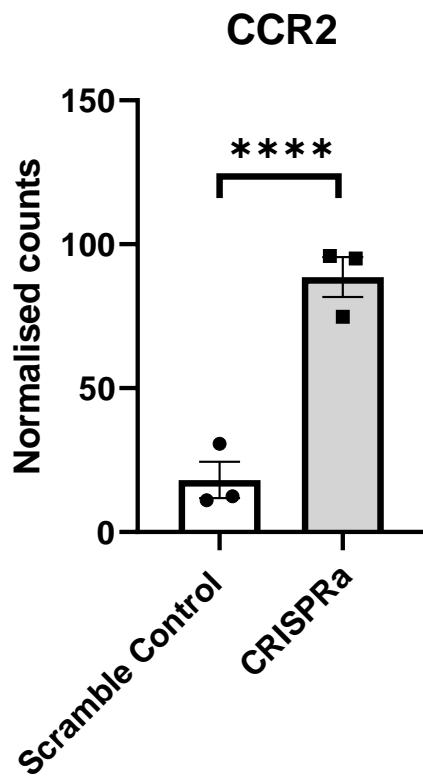
- Gene prioritization
 - COGS (Capture Hi-C Omnibus Gene Score): Bayesian method that integrates chromatin interactions, GWAS fine mapping and epigenetic marks to provide a score that indicates how likely a gene is to be causal for disease
 - 7 genes from 5 loci
- SNP prioritization
 - overlap EpiMap active enhancers in CD4+ T-cells
 - Overlap ATAC-seq peaks CD4+ T-cells from JIA children

RANKING	CHR	LEAD SNP	GWAS ASSIGNED GENE (BASED ON PROXIMITY)	COGS PRIORITISED GENE	COGS SCORE	Putative causal SNPs
1	5	rs12654812	<i>RGS14</i>	<i>RGS14</i>	0.88	-
2	5	rs4869314	<i>ERAP2</i>	<i>ERAP2</i>	0.82	-
3	1	rs6679677	<i>RSBN1</i>	<i>HIPK1</i>	0.69	rs6679677
4	3	rs34173901	<i>GLB1</i>	<i>CCR4</i>	0.65	4 SNPs
5	3	rs79815064	<i>CCR3</i>	<i>CCRL2</i>	0.64	rs80054040
6	3	rs79815064	<i>CCR3</i>	<i>CCR2</i>	0.57	rs80054040
7	3	rs79815064	<i>CCR3</i>	<i>CCR3</i>	0.51	rs80054040

The *CCR3* locus: prioritization of the most likely causal SNPs

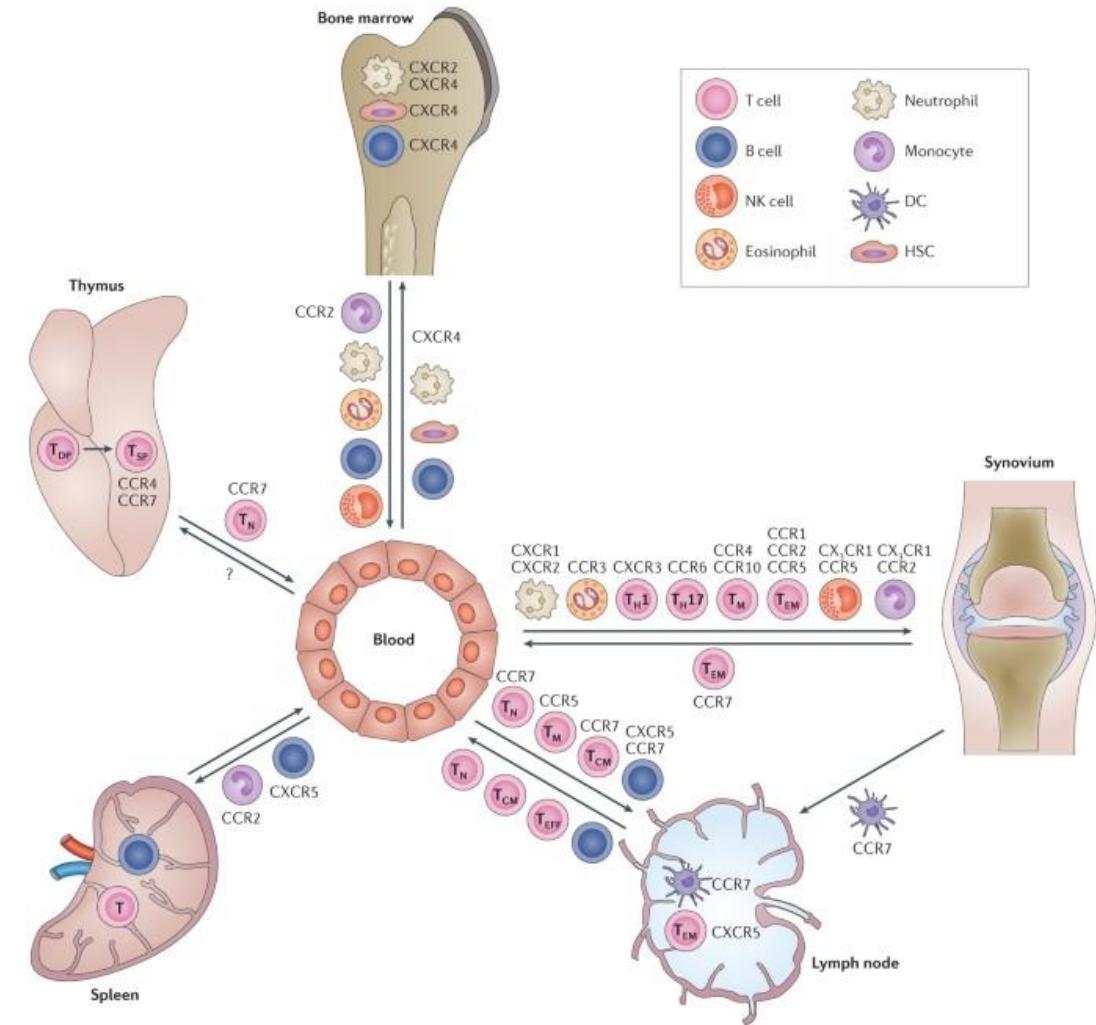


CRISPRa/i in Jurkat confirms enhancer containing rs80054040 regulates expression of interacting genes *CCR2* and *CCR5*



CCR2 and CCR5 in JIA: potential for drug repurposing

- CCR2 and CCR5 encode chemokine receptors (CCRs) that control migration of immune cells and therefore have a key role in the immune response. They have been implicated in a wide range of inflammatory diseases, including arthritis. E.g. they are upregulated in JIA patients.
- Antagonists of CCR5 were originally developed for the treatment of HIV, are already on the market and are well tolerated by patients
- Dual CCR5/CCR2 targeting is emerging as a more efficacious strategy than targeting either receptor alone in the treatment of complex human disorders
- These antagonists could have great potential for the treatment of JIA



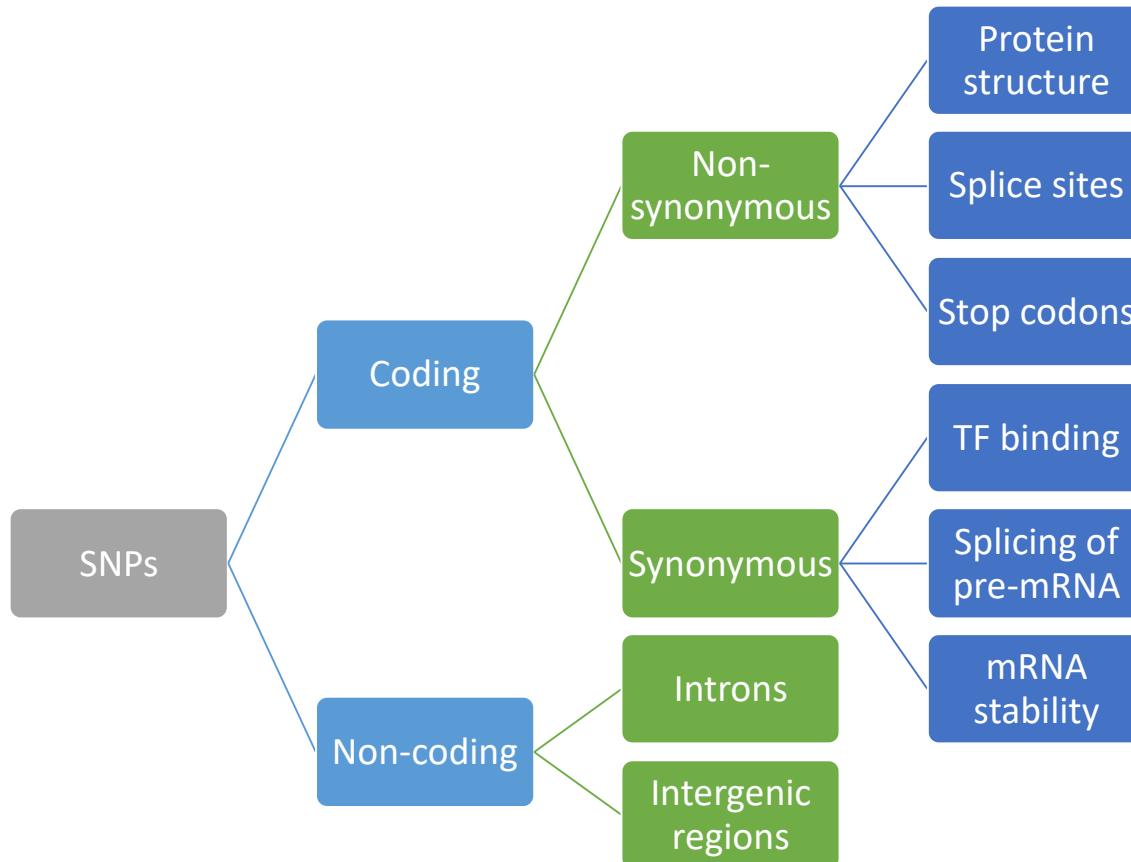
Summary

- 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**
- “Multi-omics” data 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)
- These data helps us identify key biological pathways that are affected in disease, which can in turn suggest novel therapeutic targets.

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

[Configure this page](#)

[Custom tracks](#)

[Export data](#)

[Share this page](#)

[Bookmark this page](#)

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)

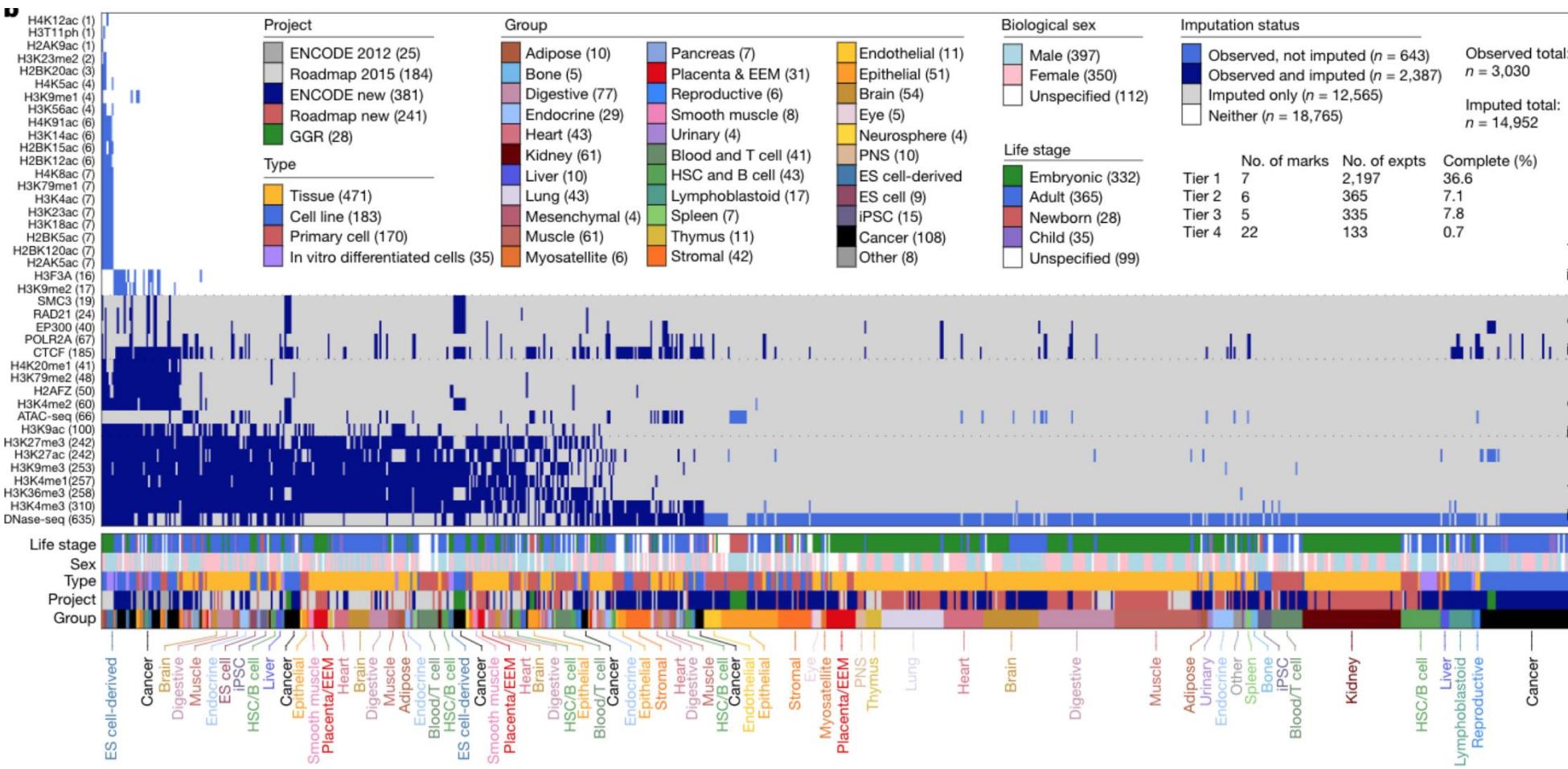
- RegulomeDB ranks SNPs with a scoring system that predicts how likely they are to be functional

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

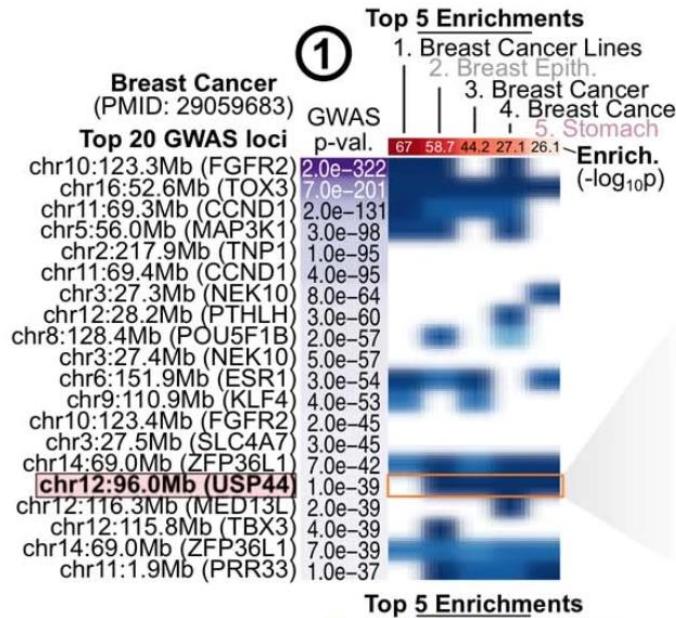
Boix et al. Nature Genetics (2021)

- 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.

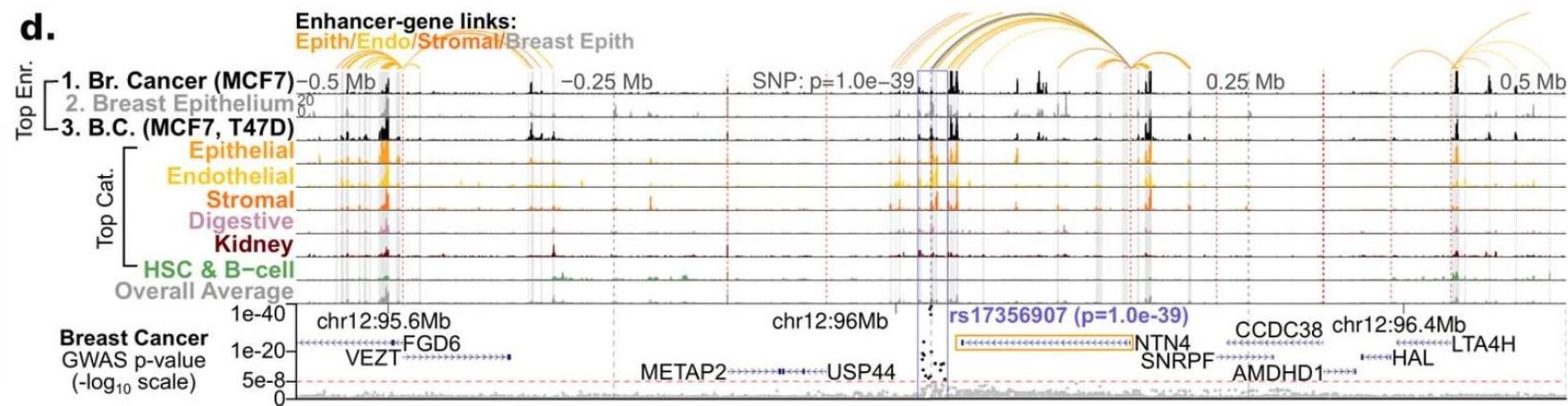


EpiMap

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



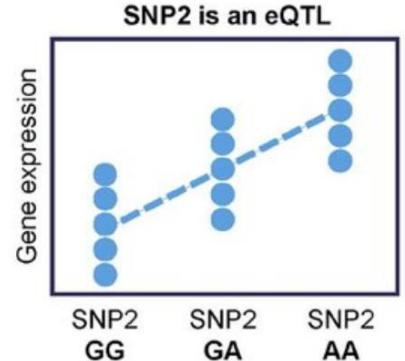
d.



<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

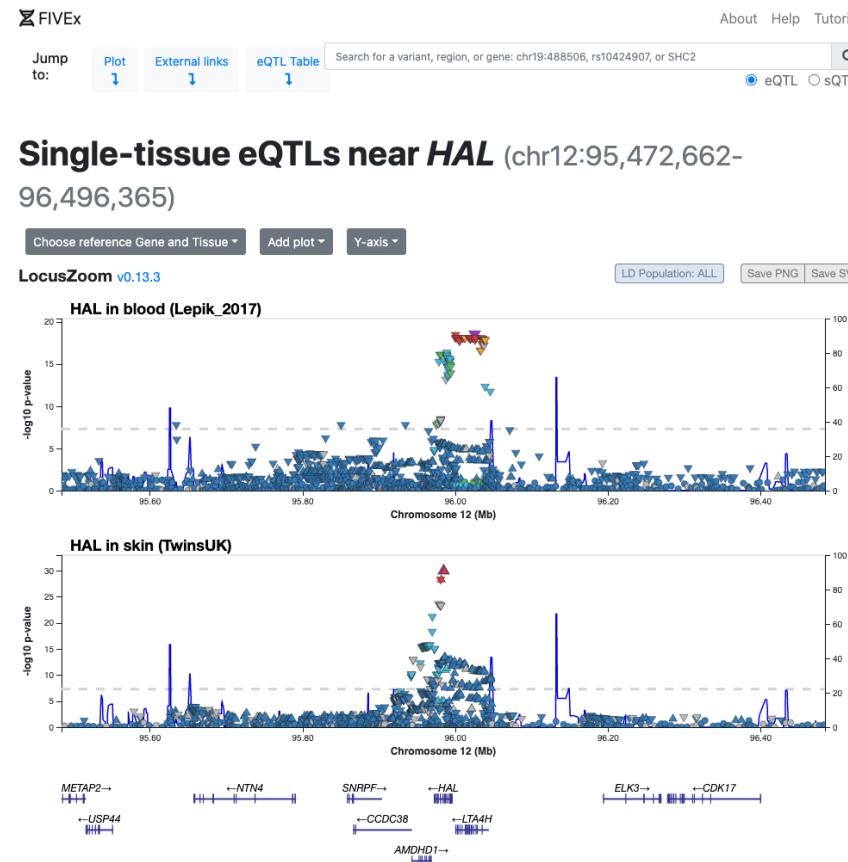
Study	Cell types or tissues	Conditions	Samples	Donors
Alasoo_2018	macrophages	IFNg, Salmonella, IFNg + Salmonella	336	84
BLUEPRINT	monocytes, neutrophils, CD4+ T cells		554	197
GENCORD	LCLs ¹ , fibroblasts, T cells		560	195
GEUVADIS	LCLs ¹		445	445
HipSci	iPSCs ²		322	322
Nedelec_2016	macrophages	Listeria, 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	Conditions	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 FIVEEx eQTL browser: <https://fivex.sph.umich.edu/>

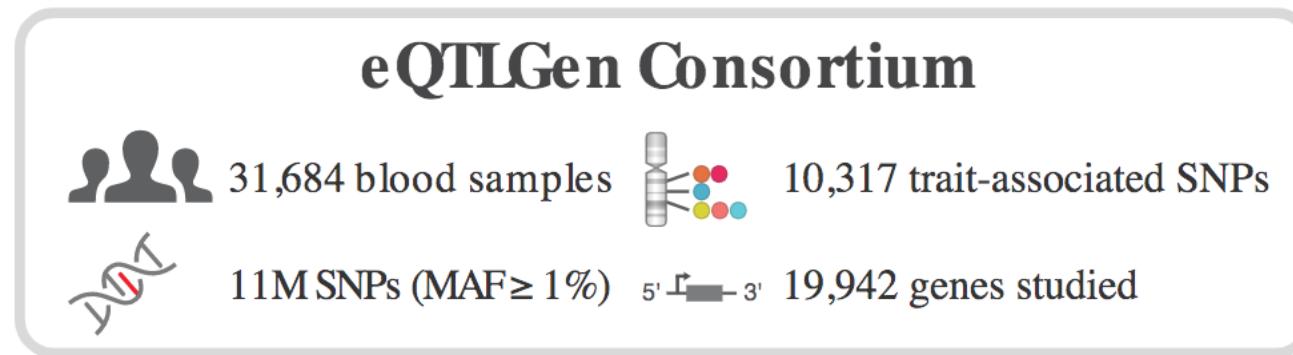


- Data can be downloaded from their FTP site <ftp://ftp.ebi.ac.uk/pub/databases/spot/eQTL>

eQTLGen

Vosa et al. Nature Genetics (2021)

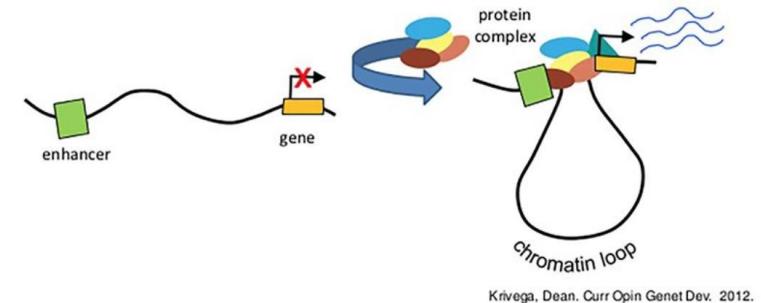
- Includes cis- and trans-expression quantitative trait locus (eQTL) using blood-derived expression from 31,684 individuals



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

Capture Hi-C plotter (CHiCP)

Schofield et al. Bioinformatics (2016)

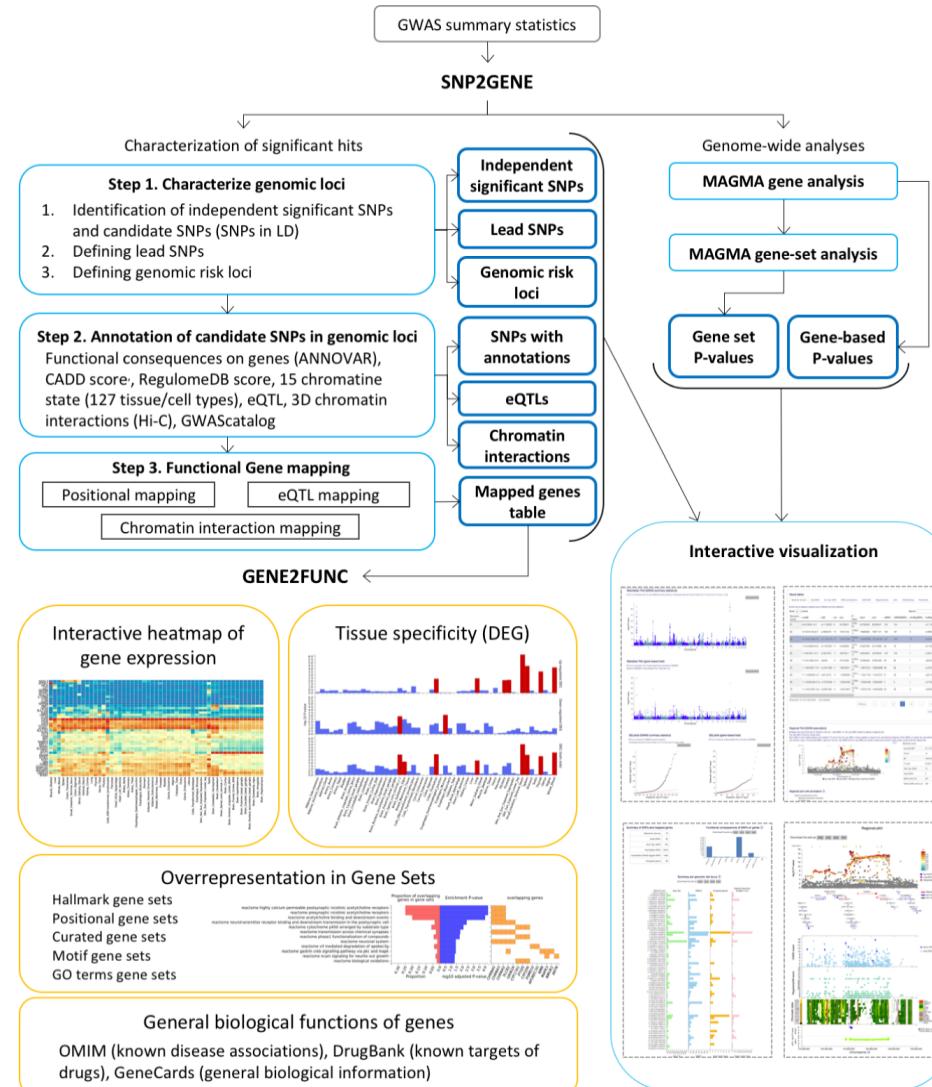


- CHiCP is a web application for visually integrating GWAS data with promoter capture Hi-C data
- It supports the analysis described in:
 - Javierre et al (2016) Cell - 17 human primary hematopoietic cell types
 - Mifsud et al (2015) Nat Genet – CD34, GM12878
 - Miguel-Escalada et al (2015) Nat Commun – 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/>

<https://www.drugbank.com>