

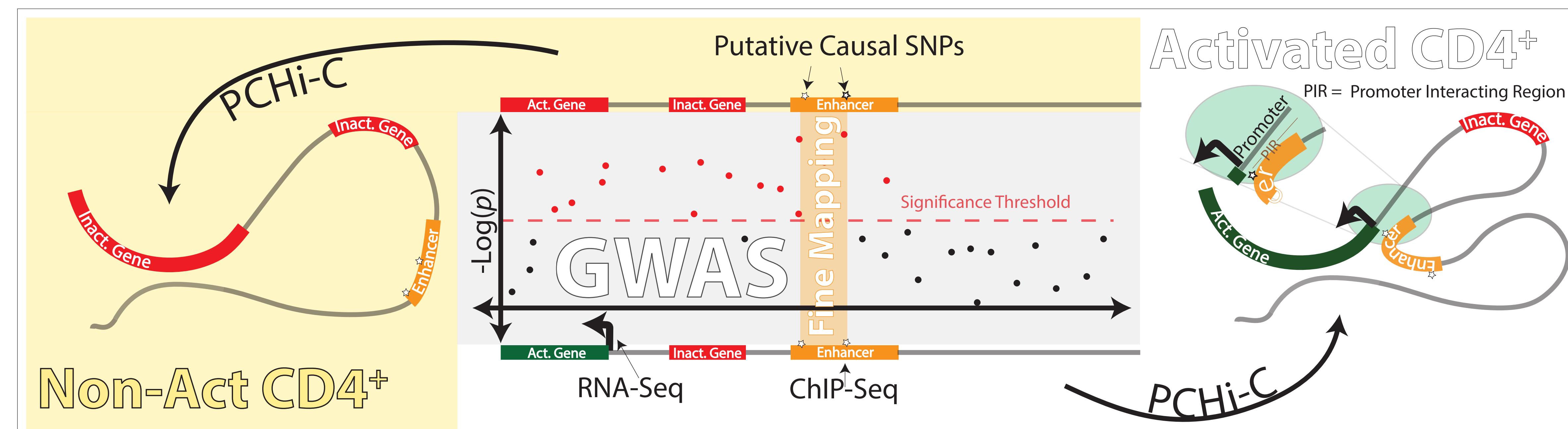
Chromosome contacts in activated T cells identify autoimmune disease-candidate genes.

Oliver S Burren^{1,8}, Arcadio Rubio García¹, Biola-Maria Javierre², Daniel B Rainbow¹, Jonathan Cairns², Nicholas J Cooper¹, John J Lambourne^{3,4}, Ellen Schofield¹, Xaqin Castro Dopico¹, Ricardo C Ferreira¹, Richard Coulson¹, Frances Burden^{3,4}, Sophia P Rowlston^{3,4}, Kate Downes^{3,4}, Steven W Wingett², Mattia Frontini^{3,4,5}, Willem H Ouwehand^{3,4,5,6}, Peter Fraser², Mikhail Spivakov², John A Todd¹, Linda S Wicker¹, Antony J Cutler¹, Chris Wallace^{1,7,8}.

¹JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, University of Cambridge, UK; ²The Babraham Institute, Cambridge, UK; ³Department of Haematology, University of Cambridge, Cambridge, UK; ⁴NHS Blood and Transplant, NIHR Cambridge Biomedical Campus, Cambridge, UK; ⁵British Heart Foundation Centre of Excellence, University of Cambridge, UK; ⁶Wellcome Trust Sanger Institute, Hinxton, UK; ⁷MRC Biostatistics Unit, Cambridge, UK; ⁸Department of Medicine, University of Cambridge, UK. Contact: ob219@cam.ac.uk

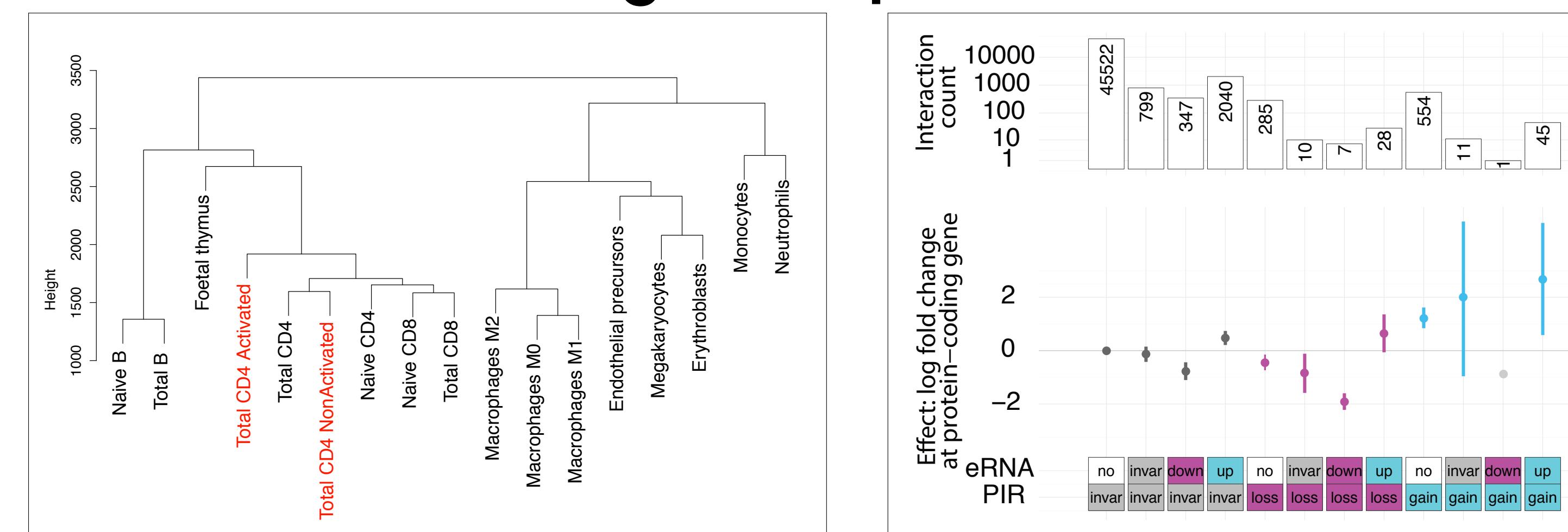
Promoter capture Hi-C (PCHi-C) can link putative causal variants to genes

Genome-wide association studies tell us about variation and phenotype. PCHi-C can indicate which genomic regions physically interact with gene promoters in specific tissue contexts. By integrating these we can develop a data-driven approach to prioritise causal candidate genes and tissue contexts for follow up functional studies.



As part of a larger study encompassing 17 primary human haematopoietic tissues(1), we used a combination of PCHi-C, total RNA-Seq and ChIP-Seq to examine the effect of activation on human CD4⁺ T cells at a 4 hour time point. We prioritised genes and tissue contexts for functional validation by integrating these data with summary GWAS statistics for 11 autoimmune traits.

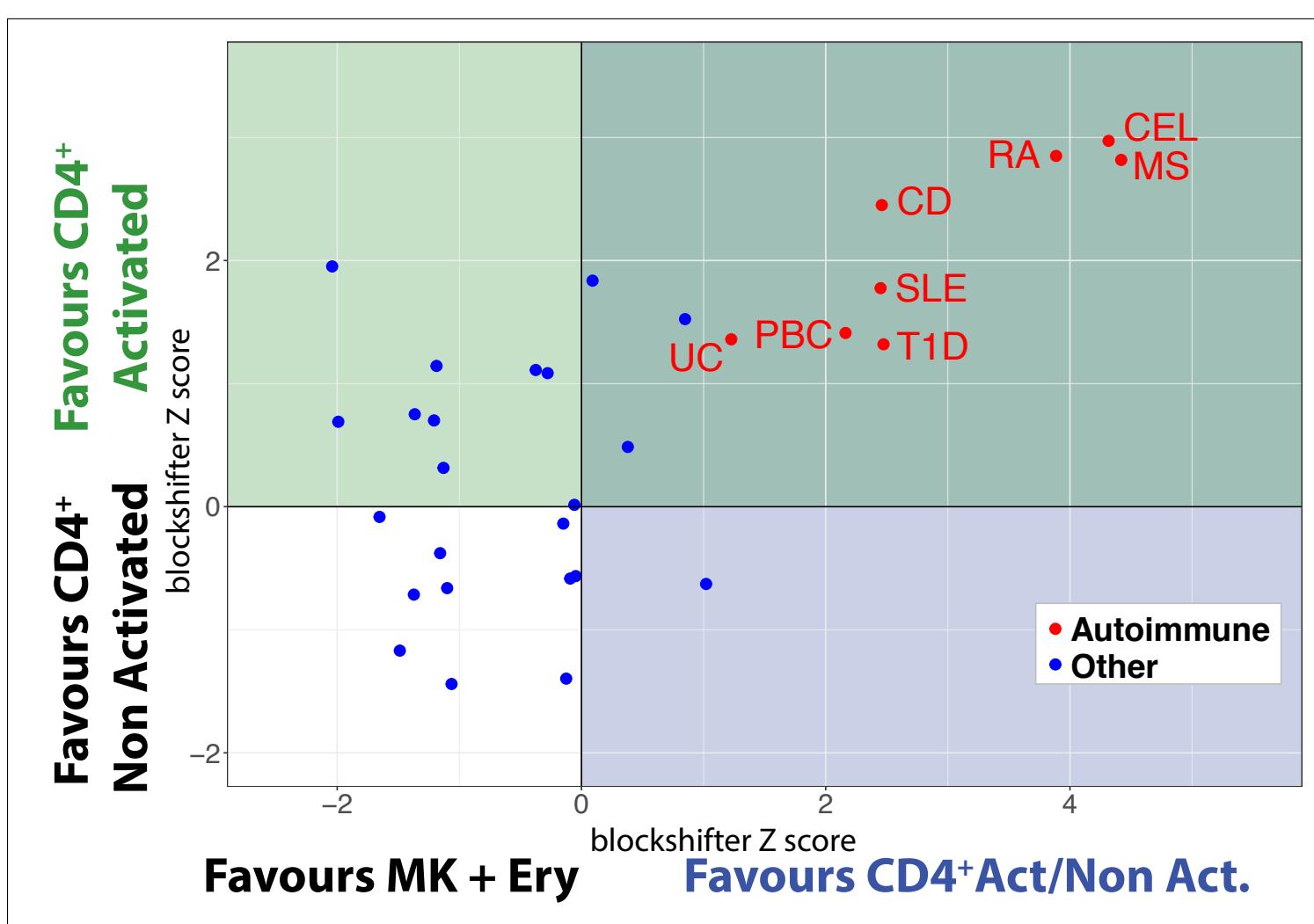
Chromosome interactions influence lineage fate and gene expression



Clustering of chromatin interactions for 17 primary human tissues identified by PCHi-C recapitulates the haematopoietic lineage tree.

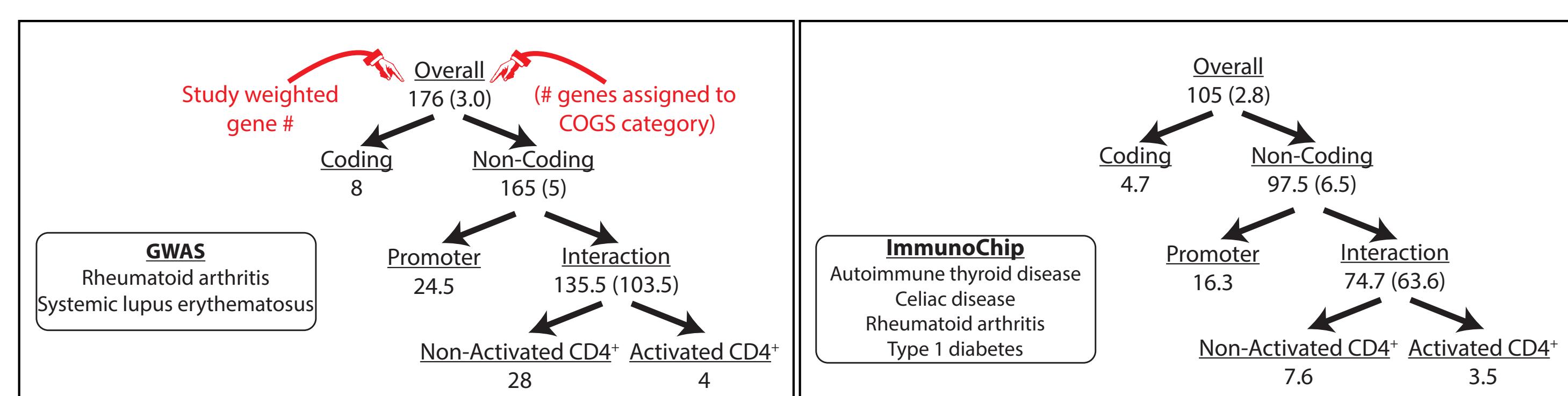
PCHi-C target gene expression correlates synergistically with promoter interaction region (PIR) gain and enhancer RNA (eRNA) expression on activation.

Autoimmune GWAS signals are enriched in activated CD4⁺ T Cell PIRs



Inspired by GOSHIFTER(2), we developed *blockshifter*, a method to examine the enrichment of GWAS summary statistics between tissue specific PIRs in the presence of correlation. We examined summary statistics for 31 GWAS traits and found autoimmune traits are most strongly enriched in activated CD4⁺ T cell PIRs

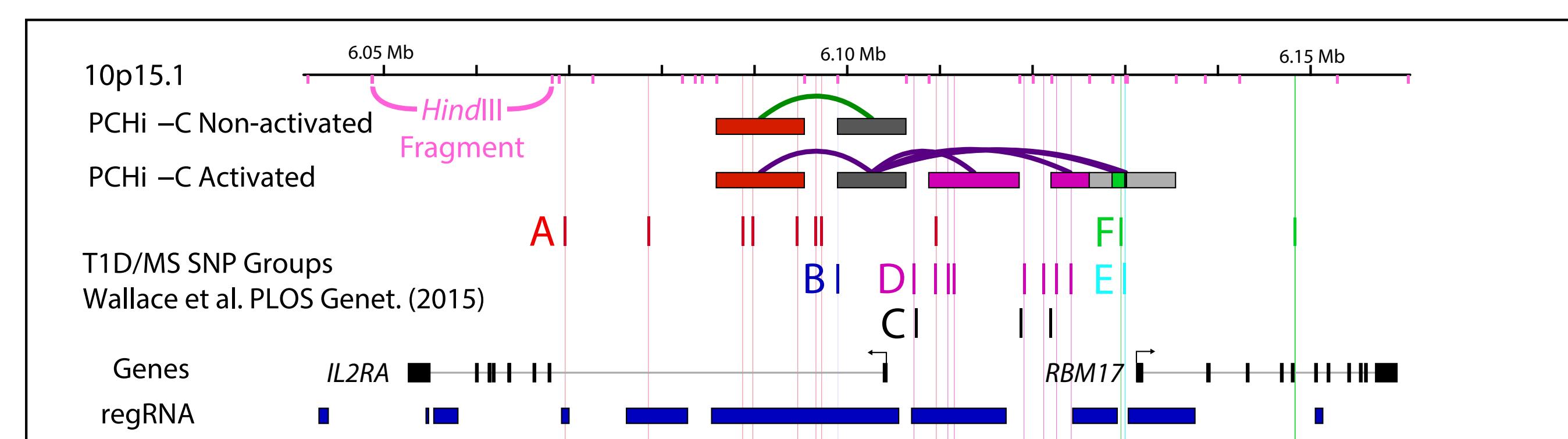
PCHi-C assisted gene prioritisation across 6 autoimmune traits



Functional gene prioritisation across 11 autoimmune diseases using genome wide (GWAS) or targeted genotyping array (ImmunoChip) data for autoimmune disease loci curated in <http://www.immunobase.org>

Interaction-mediated regulation of *IL2RA* expression

COGS analysis prioritised *IL2RA* in multiple autoimmune diseases. We previously fine mapped five signals (A-F) in type 1 diabetes (T1D) and multiple sclerosis (MS)(2). PCHi-C data shows constitutive interactions between the promoter of *IL2RA* and SNP set 'A' regardless of tissue state. On activation other interactions between downstream regions overlapping SNP sets 'D-F' form.

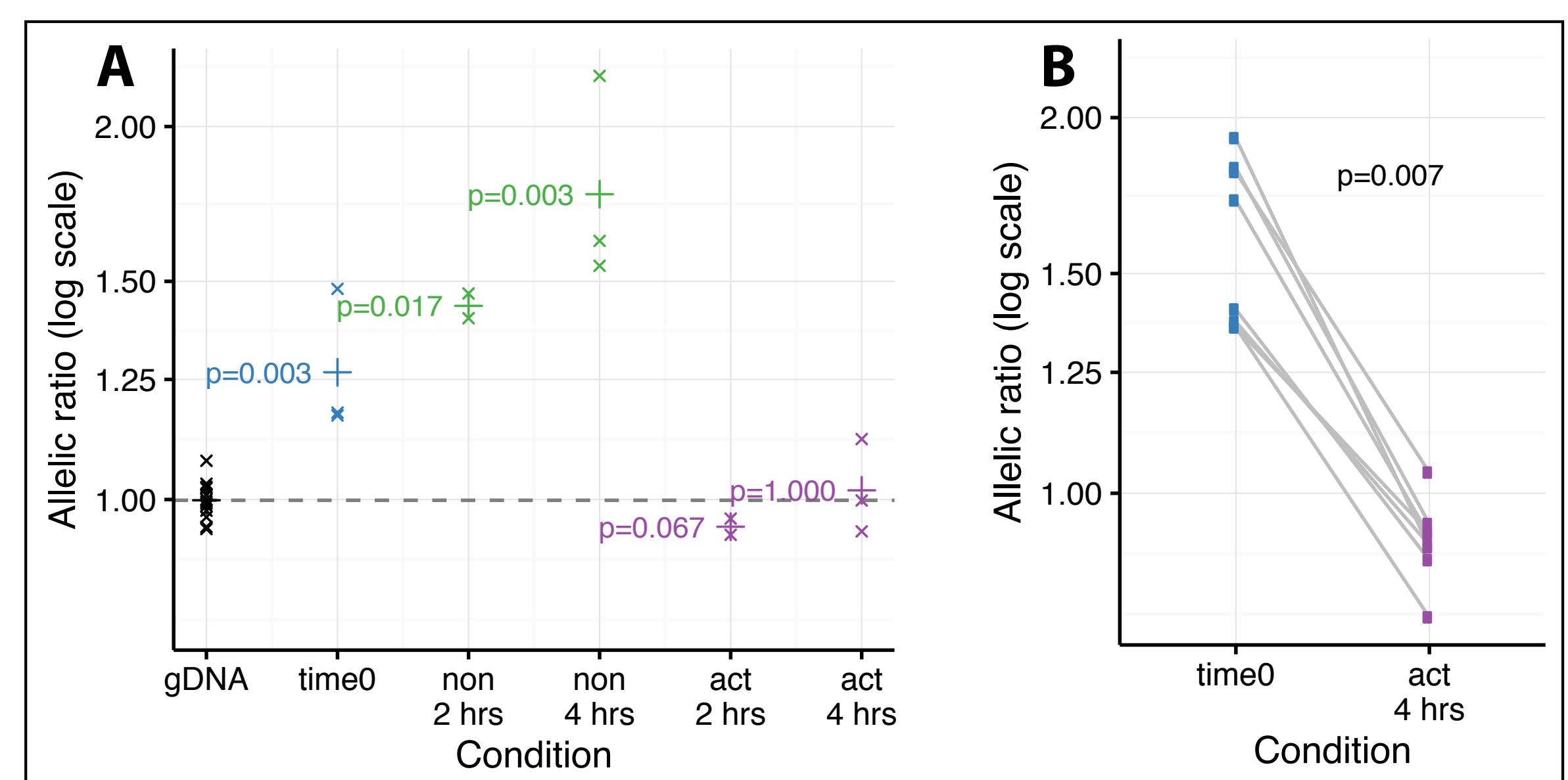


Allelically imbalanced transcription of *IL2RA* within CD4⁺ T cells of individuals heterozygous for PCHi- C linked autoimmune-associated variants is specific to non-activated cells.

Integrating PCHi-C with GWAS summary statistics

Capture Hi-C Omnibus Gene Score (COGS) is a Bayesian method for integrating PCHi-C data and GWAS to prioritise gene and tissue contexts according to SNP overlap with:

1. Coding Regions.
2. "Promoter" regions/short range interactions, refractive to PCHi-C interrogation.
3. Constitutive and tissue specific PIRs.



Allelic imbalance for individuals heterozygous for group 'A' SNPs. A) Allelic ratio is balanced (1:1) in genomic DNA (black), and activated cells at 2 and 4 hours (purple), but imbalanced in freshly sorted (blue) or non-activated (green) cells. Allelic ratio is defined as the ratio of counts from the two alleles, shown on a log scale. Each point represents the mean over 2-3 replicates for each of 4-5 individuals. B) Allelic ratio in memory CD4⁺ T cells falls upon activation in cells from eight individuals.

References

- (1) Javierre et al. (2016) *Cell*, (2) Trynka et al. (2015) *Am. J. Hum. Genet.*, (3) Wallace et al. (2015) *PLOS Genet.*

Data freely available for interactive exploration at <http://www.chicp.org>

Pre-print available at <http://biorkxiv.org/content/early/2017/01/17/100958>