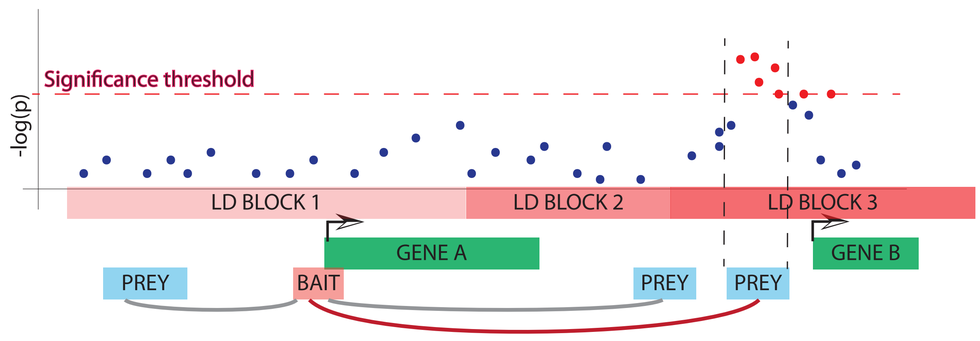
Sailing the Hi-C: Integrating ImmunoChip association summary statistics with promoter capture Hi-C data to prioritise causal genes and tissue contexts across 11 autoimmune diseases.

Oliver S. Burren[1], Ellen Schofield[1], Antony Cutler[1], Arcadio Rubio García[1], Biola-Maria Javierre[2], Jonathan Cairns[2], Tim Carver[1], Premanand Achuthan[1], Steven Hill[3], Sven Sewitz[2], Steven Wilder[4], Daniel Zerbino[4], Mattia Frontini[5], Willem Ouwehand[5,6,7], Linda S. Wicker[1], Peter Fraser[2], John A. Todd[1], Mikhail Spivakov[2] and Chris Wallace[1,3].

[1]JDRF/Wellcome Trust Diabetes and Inflammation Laboratory,University of Cambridge, UK; [2]The Babraham Institute, Cambridge, UK; [3]MRC Biostatistics Unit, Cambridge, UK; [4]EMBL-EBI, Hinxton, UK; [5]Department of Haematology, University of Cambridge, Cambridge, UK; [6]NHS Blood and Transplant, NIHR Cambridge Biomedical Campus, Cambridge, UK; [7]Wellcome Trust Sanger Institute, Hinxton, UK.

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

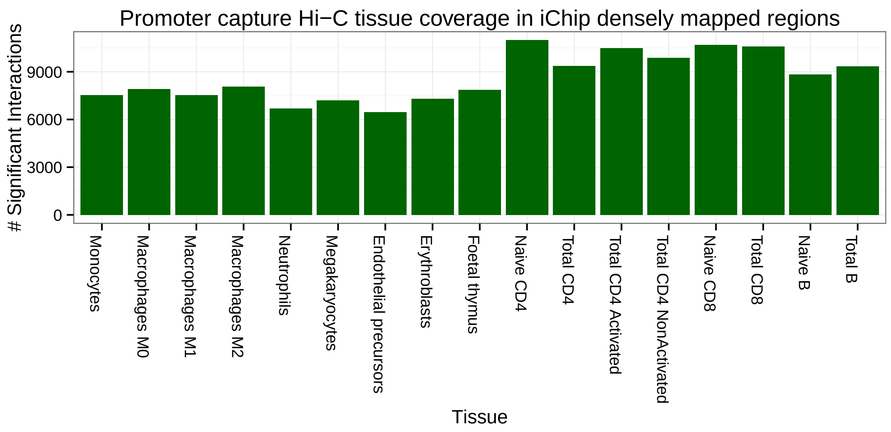
Genome-wide association studies tell us about variation and phenotype. PCHi-C can indicate which genomic regions physically interact with gene promoters in tissue context. By integrating these we can begin to prioritise causal candidate genes and relevant tissue contexts for followup functional studies.



Using traditional LD based techniques we might assign Gene B as a causal candidate gene. However PCHi-C data indicates a physical interaction between LD Block 3 and Gene A.

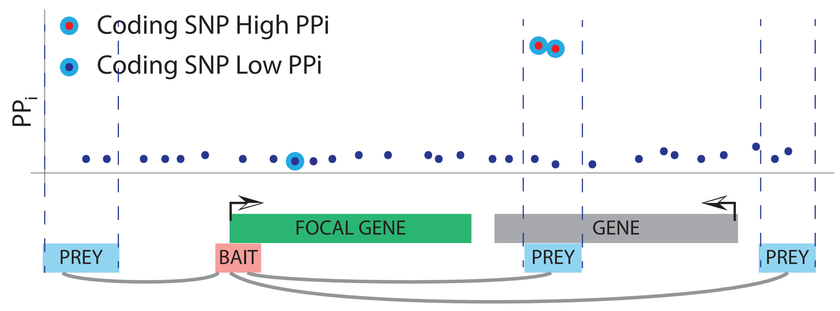
Densely genotyped ImmunoChip regions overlap PCHi-C interactions

`http://www.immunobase.org` has ImmunoChip summary statistics for 11 diseases, including 177 distinct densely genotyped regions. PCHi-C datasets are available for 17 human tissues, with a focus on Haematopoiesis. Significant interactions were called using CHiCAGO `http://regulatorygenomicsgroup.org/chicago`.

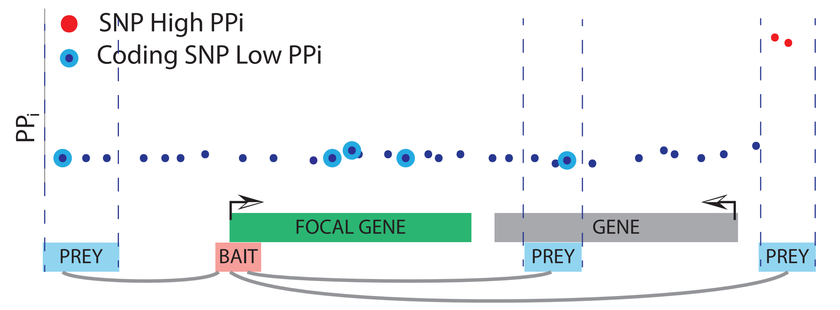


Integrating PCHi-C with summary statistics

We can use posterior probabilities (see top right box) to compare specific hypothesis using Bayes factors. Below is an example where a coding SNP hypothesis is supported.



The following figure shows an example where a causal variant overlapping an interaction is the most likely hypothesis across tissues.



To narrow down the tissue context we split interactions into binary tissue groups based on hierarchical clustering of all interactions, using a Bayes factor to select the most likely group based on the data. We continue this process of selecting groups of tissue interactions until we are unable to choose between groups or a single tissue hypothesis is most likely.

References

Wakefield 2009 *Genetic Epidemiology* 33:79-86, **Maller et. al 2012** *Nat. Genet.* 44:1294-301, **Pickrell 2014** *Am. J. Hum. Genet.* 95:559-573, **IMSGC et. al 2013** *Nat. Genet.* 5:1353-60

Using posterior probabilities to define a gene score

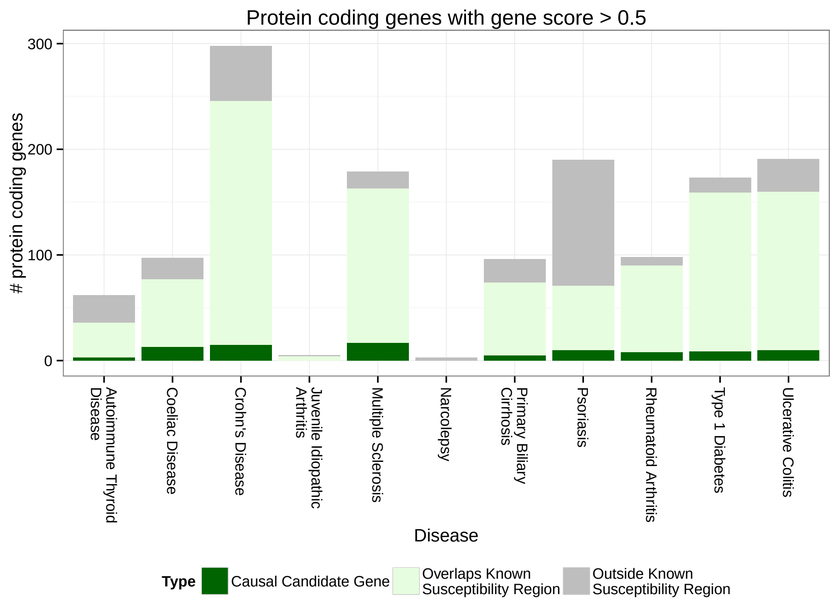
Using methods described by Wakefield (2009) and Maller et. al (2012) we convert summary p-values to posterior probabilities.

$$P(SNPi\ causal \mid Data) = PP_i = \frac{BF_i\pi_i}{1 + \sum_{i=1} BF_i\pi i} \quad (1)$$

It improves our ability to resolve causal SNPs at the cost of making the assumption of one causal variant in a region. Within a region we can sum PP_i for groups of SNPs that share interactions and combine across regions, R, by assuming statistical independence between LD blocks.

$$Gene\ score = 1 - \prod_j \left(1 - \left(\sum_{i \in R_j} 1 - PP_i\right)\right) \quad (2)$$

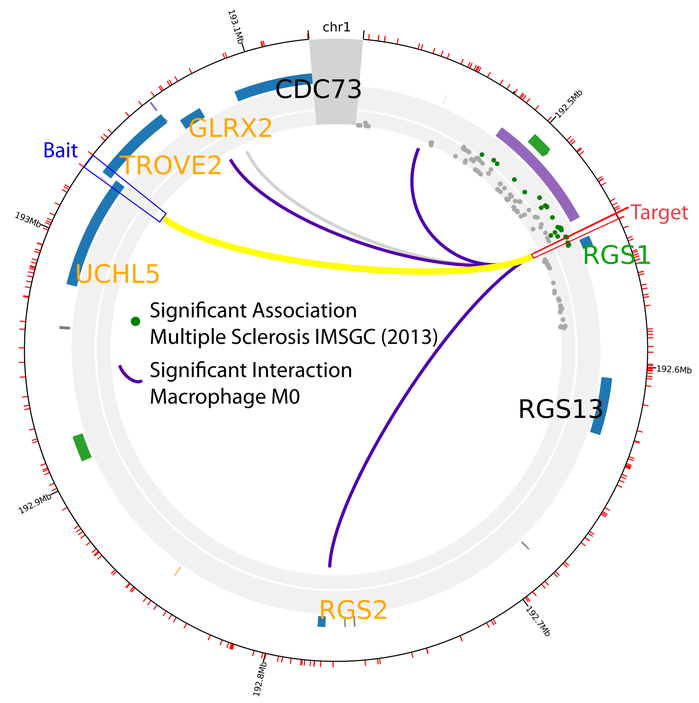
Current causal candidate gene catalogues are inadequate



These results highlight the incomplete nature of published catalogues of causal candidate genes. They also demonstrate that there is evidence for the causal candidacy of genes outside of known susceptibility regions.

CHiCP: a tool to visualise prioritised genes, SNPs & interactions

In the following example for Multiple Sclerosis the causal candidate gene is reported as *RGS1*. However, gene prioritisation using PCHi-C as described also highlights *TROVE2*, *UCLH5*, *GLRX2* and *RGS2* as additional candidates through interactions in Macrophage M0 tissue.



Visualisation created in CHiCP, an interactive beta version is available by visiting `http://chicp.immunobase.org`.

Further Work

- Incorporation of genomic annotations using fgwas Pickrell (2014) to improve prioritisation.
- Characterisation of gene overlap between diseases and identification of novel pathways.
- Development of gene set enrichment analysis using gene scores.
- Further development of CHiCP visualisation tool `http://chicp.immunobase.org`.