

The location of T1 diabetes associated SNPs in regulatory regions

S. Beka¹, I. te Boekhorst¹ ✉, I. Abnizova²

¹School of Computer Science, University of Hertfordshire, Hatfield, United Kingdom

²Wellcome Trust Sanger Institute, Hinxton, United Kingdom

Motivations

Although many association studies on complex diseases focus on variation in coding DNA, recent research shows increasing evidence that the cause of such diseases should be sought in the regulation of gene activity. Rather than studying mutations in genes coding for transcription factors, my work focuses on genetic variants (SNPs) in regulatory modules (TFBS, enhancers, promoters, or other genic locations likely to be involved in gene regulation such as UTR, introns and splice junctions) that are in Type 1 Diabetes (T1D) susceptibility regions of the human genome. The specific research question is: are SNPs associated with T1D more likely to occur in (putative) regulatory regions than other SNPs found in T1D susceptible regions? In addition: Because genes may overlap and/or occurrence in multiple transcripts, one and the same SNP may be associated with more than one genic location and affect more than one functional region (e.g. is a mutation in as well a coding region as a regulatory region). Are SNPs associated with a complex disease such as T1D more likely to be of this kind?

Methods

An extensive search in the databases ENSEMBL and T1Dbase was conducted to collect information all SNPs in T1D susceptible regions [coordinates on genome, type of variant (mutant allele, wildtype allele, transversion or transition, synonymous or non-synonymous), transcriptID, intra-genic region (up/downstream, exon, intron, splice junction, UTR, type of (micro)RNA), type of gene (coding/pseudo/microRNA) and associated diseases] A statistical analysis was performed to assess possible associations between the status of a SNP (associated/not-associated with T1D) on the one hand and features characterising the (intra/inter-) genic position on the other hand.

Results

1) SNPs associated with T1D are more likely to occur in regulatory regions than those that are not associated with T1D 2) SNPs associated with T1D occur more often in multiple genic locations than those that occur in only one genic location. They appear to be relatively over-abundant in transcription factor binding sites, introns and splice-sites.