International Journal of Medical Sciences & Pharmaceutical Research, Volume1, Issue1, Jan-Jun-2017, pp 25-39,

BIOSTATISTICAL ASSESSMENT OF GENIC POSITIONS AND PROFILES OF SIGNIFICANT TRANSCRIPTION FACTOR BINDING SITES IN SINGLE NUCLEOTIDE POLYMORPHISM SENSITIVITY

BEKA, SYLVIA [1] . RENE TE BOEKHORST [2]. IRINA ABNIZOVA [3] BEKA, NATHAN[4]

[1,2] Department of Computer Science, University of Hertfordshire, Hatfield, Hertfordshire

[3] The Wellcome Trust Sanger Institute, Hinxton, Cambridge

[4] Biocomputation Group, Science and Technology Research Institute, University of Hertfordshire, Hatfield, Hertfordshire

ABSTRACT

Human complex diseases, like Diabetes and Cancer, affect many people worldwide today. Despite existing knowledge, many of these diseases are still not preventable. Complex diseases are known to be caused by a combination of genetic factors, as well as environmental and life style factors. The scope of this investigation covered the genomics of Type 1 Diabetes (T1D). There are 49 human genomic regions that are known to carry markers (disease-associated single nucleotide mutations) for T1D, and these were extensively studied in this work. The aim was to find out in how far this disease may be caused by problems in gene regulation rather than in gene coding. For this, the genetic factors associated with T1D, including the single point mutations and susceptibility regions, were characterised on the basis of their genomic attributes. Furthermore, mutations that occur in binding sites for transcription factors were analysed for change in the conspicuousness of their binding region, caused by allele substitution. This is called SNP (Single nucleotide polymorphism) sensitivity. From this study, it was found that the markers for T1D are mostly non-coding SNPs that occur in introns and non-coding gene transcripts, these are structures known to be involved in gene regulatory activity. It was also discovered that the T1D susceptibility regions contain an abundance of intronic, non-coding transcript and regulatory nucleotides, and that they can be split into three distinct groups on the basis of their structural and functional genomic contents. Finally, using an algorithm designed for this study, thirty-seven SNPs that change the representation of their surrounding region were identified. These regulatory mutations are non-associated T1D-SNPs that are mostly characterised by Cytosine to Thymine (C-T) transition mutations. They were found to be closer in average distance to the disease-associated SNPs than other SNPs in binding sites,

and also to occur frequently in the binding motifs for the USF (Upstream stimulatory factor) protein family which is linked to problems in Type 2 diabetes.