Oral presentation

Prioritisation of Disease Gene Candidates: A Systems Biology Approach
Frances S Turner*

Address: Medical Research Council Human Genetics Unit, Western General Hospital, Edinburgh University, UK.

Email: Frances S Turner* - fturner@hgu.mrc.ac.uk

* Corresponding author

Although much work has been done in the linkage mapping of many common, genetically complex diseases studies have typically identified disease susceptibility regions of tens of megabases, rather than individual genes [1]. Typically genes are prioritized for further study, by looking for those whose characteristics (function/expression etc) fit with what is already known about the disease. However, by definition such approaches would fail to identify novel disease genes that do not fit with prior knowledge of the disease or where little is known about the underlying aetiology of the disease.

It is intuitive that genes predisposing to the same phenotype may encode proteins with similar functions, or be present in the same pathways or complexes, which may be reflected by shared annotation, co-expression or protein interaction. Therefor where multiple susceptibility regions have been mapped, good candidate genes may potentially be selected according to their similarity to genes in the other susceptibility regions. However, looking for potential disease related pathways or functions from long lists of candidate genes leads to numerous false positives due to the large number of non-disease genes compared to disease genes.

Here we present POCUS (Prioritisation Of Candidate genes Using Statistics) [2]. POCUS searches for expression profiles or functional annotation including Gene Ontology (GO) terms and InterPro domains shared between genes within different susceptibility regions for the same disease. Each gene is scored according to the features it shares with genes in other regions. The scores reflect the probability of the observed similarity being seen by chance, so the false positive rate is controlled. We demonstrate that POCUS can successfully identify genes underlying polygenic diseases, and genes in common human pathways using pathway data from the Reactome database [3]. This method could also be applied to model organisms where mapping data is often of higher resolution.

References
1. Risch NJ: **Searching for genetic determinants in the new millennium.** *Nature* 2000, **405**:847-856.
2. Turner FS, Clutterbuck DR, Semple CA: **POCUS: mining genomic sequence annotation to predict disease genes.** *Genome Biol* 2003, **4**:R75.
3. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L: **Reactome: a knowledgebase of biological pathways.** *Nucleic Acids Res* 2005, **33**:D428-32.