Computational prediction of genetic targets for fetal alcohol spectrum disorders

Zané Lombard

Abstract

Fetal alcohol spectrum disorders (FASD) describe the range of disorders that result from in utero alcohol exposure. FASD is a serious global health problem and is observed at exceedingly high frequencies in certain South African communities. Although in utero alcohol exposure is the primary trigger, there is evidence that genetic- and other susceptibility factors contribute towards FASD development. To date, no genome-wide association or linkage studies have been performed for any of the FASD syndromes.

The main objectives of this study were to develop an innovative approach to computationally identify biologically plausible candidate genes for FASD, for a future association study, and to evaluate the appropriateness and validity of this approach. Further, an in silico analysis of known single nucleotide polymorphisms (SNPs) within the top-ranked candidate gene was performed in conjunction with de novo SNP detection, to select a subset of SNPs based on proposed functional impact on gene expression and protein function, for a prospective association study.

A computational binary filtering technique was designed that can be employed to prioritize genes in a candidate list, or could be used to rank all genes in the genome in the absence of such a list. 10174 FASD candidate genes were initially selected from the whole genome using a previously described method. Hereafter the candidates were prioritized using a binary filtering technique. The biological enrichment of the ranked genes was assessed by investigating the protein-protein interactions, functional enrichment and common promoter element binding sites of the top-ranked genes. A group of 87 genes was prioritized as candidates highlighting many strong candidates from the TGF-β, MAPK and Hedgehog signalling pathways, which are all integral to fetal development and potential targets for alcohol's teratogenic effect.

To assess the effectiveness and accuracy of this computational approach, X-linked mental retardation (XLMR) was used as a test disease, considering that XLMR is a set of heterogeneous disorders of which some of the underlying genetics is known. This implementation resulted in a prioritized gene list with a noted enrichment of known XLMR genes among the top-ranked genes. Furthermore, the top-ranked list contained genes that were biologically relevant to XLMR, and could potentially be as yet unknown candidate genes for XLMR. Indeed, many of the top-ranked genes mapped to XLMR candidate regions, confirming their status as good candidates.

Finally, a subset of seven known and novel SNPs was selected within *FGFR1* based on putative functional impact. Data from the HapMap project was used to identify tag SNPs for *FGFR1* to complement the selection made based on function.

The main limitation of the proposed computational approach to candidate gene prediction is that it is primarily based on gene annotation, and that it is therefore biased towards selecting better-annotated genes. However, the results obtained in this study suggest that the described computational method is an effective approach that can identify likely candidates that are biologically relevant to the disease of interest, and therefore appropriate for a candidate-gene association studies. In practice, this technique is an appropriate approach to select a workable set of candidate genes for a complex disease, in a setting where a whole-genome association study is not a viable option.