The recent technological advances have allowed researchers to discover a wide range of variations on the scale of a complete human genome (Redon et al., 2006). The role of single nucleotide polymorphisms (SNPs) and genomic copy number variations (CNVs) in human disease is becoming increasingly evident (Abrahams and Geschwind, 2008; Barkovich et al., 2009; Bucan et al., 2009; Melford et al., 2009), but it is often difficult to move from associations to the identification of causative variations.

The number of genes in CNVs may be large, and non-coding regions may also contribute to the phenotype. On the other hand, the identification of functional SNPs requires predicting what impact a certain polymorphism might have on gene function. This is relatively straightforward with sequence variants leading to amino acid changes in protein coding regions, since these are more likely to disrupt gene function. However, a clear result of genome wide association studies (GWASs) is that functional SNPs with association to disease are as likely to be found in the non-coding portion of the genome (Frazer et al., 2009). Thus data obtained by several types of studies require an informatics approach to organize biological knowledge and prioritize genomic intervals for further analysis.

The current, widely used genome browsers, such as UCSC Browser (Fujita et al., 2011). Ensembl (Flicek et al., 2011) and GBrowse (Donlin, 2009) use individual intervals on a reference genome to display the tracks of functional elements such as genes or RNA. Investigating genomic intervals obtained through CNV or other experiments requires looking at each interval individually, while comparison of functional features in multiple intervals can bring new insights into genome studies.

We describe an interactive online genome analysis tool VISTA RViewer, a new addition to the VISTA suite of tools for comparative genomics (Frazer et al., 2004). RViewer is designed to accept datasets produced by clinical geneticists [either CNVs identified in copy number array comparative genomic hybridization (aCGH) studies or genomic neighborhoods of SNPs of interest from GWASs]. It calculates a set of important parameters for both coding and non-coding regions, and provides investigators with capabilities to compare the intervals and identify those that are likely to be significant in a particular study. This tool also links to mouse phenotype and gene expression data, thus allowing for comparisons and prioritization of significant genes and non-coding regions across a region.

RViewer presents several functionalities not found in currently available tools. First, multiple genomic intervals are displayed simultaneously, allowing for quicker visual inspection. Second, direct links to gene ontology, pathway and mouse phenotype data allow for more rapid prioritization of genes and regions of likely biological significance. And third, the annotation of highly conserved non-coding intervals brings an important component into the analysis, and is likely to be useful for whole genome sequencing studies. To our knowledge, software providing such functions is not currently available.

2 DESCRIPTION

2.1 Software

RViewer allows users to create individual accounts and upload their data in the form of genomic regions from either hg18 or hg19 builds of the human genome in the UCSC BED format, analyze
The tool is implemented as a Java-based web application. The technologies used include Google Web Toolkit, EXT library, Spring Framework and MySQL database server.

In summary, the RViewer is an analytical platform for genomic regions that allows for visualization, comparison and analysis of large-scale genomic data from current biomedical research.

2.2 Case study

To demonstrate the functionality of RViewer, we analyzed nine de novo copy number variants identified by aCGH in patients with infantile spasms (ISS), a form of epilepsy associated with severe developmental outcome. Preliminary analysis suggested involvement of at least two biological networks—ventral forebrain development and synaptic function. The regions of CNVs (hg18) were uploaded into RViewer and analyzed (Fig. 1). Based on the annotations provided by RViewer, 23 genes with brain expression patterns and GO terms similar to known ISS-associated genes were identified. This allowed for identification of the best candidates for follow-up confirmatory studies much faster than would have been possible using currently available tools. Annotation of non-coding intervals also identified highly conserved regions of possible biological significance for further study.

3 DISCUSSION

The high-throughput nature of modern genomic experiments makes the availability of efficient analytical tools critical. The presented novel software RViewer has been utilized in several biomedical test cases, and showed significant improvement in the speed and quality of analysis allowing for fast retrieval and visualization of relevant biological information. The tool is designed in a modular way allowing for customization and expansion such as taking results of other genomic experiments (for example resequencing, whole exome and genome sequencing), adding annotations relevant to a particular disease, analyzing novel variant data and introducing new complex searches. This work is currently in progress.

ACKNOWLEDGEMENTS

We are grateful to Drs Natalia Maltsev, William Dobyns and Elliott Sherr for fruitful ideas and discussions, to Dr Ivan Ovcharenko for providing the HCT data, to Alexander Poliakov and Igor Ratnere for technical support and to Tatyana Smirnova for designing the RViewer Web site.

Funding: National Institutes of Health (R01 HL091495). The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. (DE-AC02-05CH11231).

Conflict of Interest: none declared.

REFERENCES