Bioinformatics, 31(14), 2015, 2262–2268
doi: 10.1093/bioinformatics/btv151
Advance Access Publication Date: 18 March 2015
Original Paper

OXFORD

Genome analysis

Family genome browser: visualizing genomes with pedigree information

Liran Juan†, Yongzhuang Liu†, Yongtian Wang, Mingxiang Teng, Tianyi Zang and Yadong Wang*

Center for Bioinformatics, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: John Hancock

Received on December 10, 2014; revised on February 18, 2015; accepted on March 11, 2015

Abstract

Motivation: Families with inherited diseases are widely used in Mendelian/complex disease studies. Owing to the advances in high-throughput sequencing technologies, family genome sequencing becomes more and more prevalent. Visualizing family genomes can greatly facilitate human genetics studies and personalized medicine. However, due to the complex genetic relationships and high similarities among genomes of consanguineous family members, family genomes are difficult to be visualized in traditional genome visualization framework. How to visualize the family genome variants and their functions with integrated pedigree information remains a critical challenge.

Results: We developed the Family Genome Browser (FGB) to provide comprehensive analysis and visualization for family genomes. The FGB can visualize family genomes in both individual level and variant level effectively, through integrating genome data with pedigree information. Family genome analysis, including determination of parental origin of the variants, detection of *de novo* mutations, identification of potential recombination events and identical-by-decent segments, etc., can be performed flexibly. Diverse annotations for the family genome variants, such as dbSNP memberships, linkage disequilibriums, genes, variant effects, potential phenotypes, etc., are illustrated as well. Moreover, the FGB can automatically search *de novo* mutations and compound heterozygous variants for a selected individual, and guide investigators to find high-risk genes with flexible navigation options. These features enable users to investigate and understand family genomes intuitively and systematically.

Availability and implementation: The FGB is available at http://mlg.hit.edu.cn/FGB/.

Contact: ydwang@hit.edu.cn.

1 Introduction

Families with inherited genetic diseases are widely used as cases in Mendelian/complex disease studies. Recently, advances in high-throughput sequencing technologies have led the family genome sequencing to become more and more popular in these studies. Through analysis of family genome sequencing data, such as variant calling, variant prioritizing, etc., researchers are able to better understand the genetic basis of Mendelian/complex diseases in variant level, and create insights for disease prediction, diagnosis and treatment. In family genome sequencing analysis, millions of variants are called from sequencing data. Most of them are rare or have no functional significance. Several methods, such as GEMINI, VariantTools, MendelScan, FARVAT, etc., are developed to identify causal variants in family based studies of disease (Choi *et al.*, 2014;

Koboldt *et al*., 2014; Paila *et al*., 2013; San Lucas *et al*., 2012). However, visualization is always the most intuitive and straightforward approach to compare and validate the remaining candidates of disease causing variants.

Several methods, tools and statistical models have been developed for family based variant detection, annotation and analysis. MendelScan prioritizes candidate variants and searches Mendelian-disease genes by analysing sequencing data and pedigree/phenotype information in family studies of Mendelian diseases (Koboldt *et al*., 2014). GEMINI annotates genetic variants by integrating a diverse and adaptable set of genome annotations into a unified database to facilitate interpretation and data exploration (Paila *et al*., 2013). VariantTools provides a command-line driven toolset for building more sophisticated analytical methods (San Lucas *et al*., 2012). FamAnn automatically selects and annotates variants segregating in each family and shared across families, for facilitating disease variants or genes discovery in family based sequencing studies (Yao *et al*., 2014). VAR-MD analyses the DNA sequence variants produced by human exome sequencing, and generates a ranked list of variants using predicted pathogenicity, Mendelian inheritance models, genotype quality and population variant frequency data (Sincan *et al*., 2012). FARVAT is a family based rare variant association test, and estimates minor allele frequency of each rare variant between affected and unaffected individuals with the best linear unbiased estimators (Choi *et al*., 2014). FamSeq builds on Bayesian networks and the Markov chain Monte Carlo algorithm and provides a confidence measure for variant calls using data from all family members, which can improve the quality of rare variant detection in family based sequencing studies (Peng *et al*., 2013). These tools and methods optimize, simplify and automatize many analysis processes in family based sequencing studies, generating huge amount of genetic variants for downstream validation and visualization.

Visualization provides users intuitive descriptions of raw data. Various visualization tools have been developed for facilitating different types of genomic studies. Comprehensive genome browsers, such as the UCSC genome browser (Kent *et al*., 2002), Ensembl genome browser (Flicek *et al*., 2014), etc., can rapidly and reliably display genomes of many species, together with dozens of integrated genome annotations, including genes and transcripts, variation, comparative genomics, regulatory data, etc. Some standalone genome browsers are available for the interactive exploration of large datasets, especially the high-throughput sequencing datasets, for example, the Integrative Genomics Viewer (Robinson *et al*., 2011) and Savant genome browser (Fiume *et al*., 2012). They can also display a wide variety of genomic annotations to support data analysis. There are many visualization tools focusing on visualizing particular types of data, such as Epigenome Browser (Zhou *et al*., 2011) for visualizing epigenome data, and Hawkeye (Schatz *et al*., 2007) for visualizing large-scale assembly data. GBrowse (Stein *et al*., 2002) and HuRef browser (Axelrod *et al*., 2009) are used to display the genome of a single individual. TASUKE (Kumagai *et al*., 2013) is developed for the visualization of variations, annotations and read depth of multiple genomes. The personal genome browser is designed and developed for visualizing functions of individual genome variants (Juan *et al*., 2014). Although these genome browsers enable researchers to view and interact with genomic annotations and data at any requested portion of genomes, they have no functionality with pedigree information.

The pedigree shows the relations and phenotypes of a family. A few tools such as HaploPainter (Thiele and Nurnberg, 2005) and Pedimap (Voorrips *et al*., 2012) are developed for displaying phenotypic and genotypic data for related individuals linked in pedigrees. These tools focus on visualizing pedigrees rather than genomes.

Family genome sequencing data, including both exome sequencing data and whole-genome sequencing data, have been generated by many family based Mendelian and complex disease studies. The visualization and comparison of disease-causing variants of family members can provide intuitive understanding of the genetic basis of Mendelian and complex diseases, thus can support decision making in healthcare. However, currently there is a lack of visualization tools dedicated to analysing variants for family genomes.

We developed the Family Genome Browser (FGB) to support the comprehensive analysis and visualization for genomes of related individuals linked in pedigrees. The FGB enables investigators to upload and view family genome variants with the pedigree information of the family. Based on the relations among family members, users are able to investigate the genetic similarities/differences among consanguineous individuals in variant level. Various kinds of family based analysis, such as determination of parental origin of the variants, recognition of *de novo* mutations, identification of potential recombination events and identical-by-decent (IBD) segments, etc., can be performed by flexibly specifying different family members and analysis modes. Diverse bioinformatics resources, such as dbSNP (Sherry *et al*., 2001), linkage disequilibrium (LD), HGNC genes (Gray *et al*., 2013), RefSeq genes (NCBI Resource Coordinators, 2013), OMIM (Hamosh *et al*., 2005), etc., are integrated in the FGB for comprehensively annotating the genomic variants. For individuals whose parents' variant data are available, the FGB offers a scanning function for automatically finding *de novo* mutations and compound heterozygous variants of the individual. The two kinds of suspicious variants may have the potential to cause dominant and recessive genetic diseases. The scan can be performed on the whole genome, a chromosome, or a cytoband. The FGB provides necessary and convenient genome visualization and analysis features dedicated to family based studies. These features enable users to investigate and understand family genomes intuitively and systematically.

2 Results

The FGB accepts variant call format (VCF) files (Danecek *et al*., 2011) and PED format texts (Purcell *et al*., 2007) as inputs of variant data of multiple genomes and pedigree information (Fig. 1a), illustrates family genomes in a standard pedigree chart (Fig. 1b), and shows genomic variants and their functional consequences in a vertical linear genome display (Fig. 1f).

2.1 Data input

The FGB supports the Bgzip/Tabix (Li, 2011) compressed/indexed VCF files. Investigators can either provide accessible data URLs or upload local data and index files. By inputting the family description in PED format into the 'Pedigree' text area (Fig. 1a), the pedigree chart of the family can be generated and displayed in the 'Pedigree' tab (Fig. 1b).

The drop-down menu can be used to select and display the built-in and uploaded family genome datasets. The uploaded data can be removed by clicking the 'x' button.

We currently hold 100 selected individuals from 1000 genomes project phase 3 (Initial release at June 24) dataset (1000 Genomes Project Consortium, 2012). Each of the 100 selected individuals has at least one relative in the dataset. The pedigree charts of six trios are displayed in the 'Pedigree' tab by default.

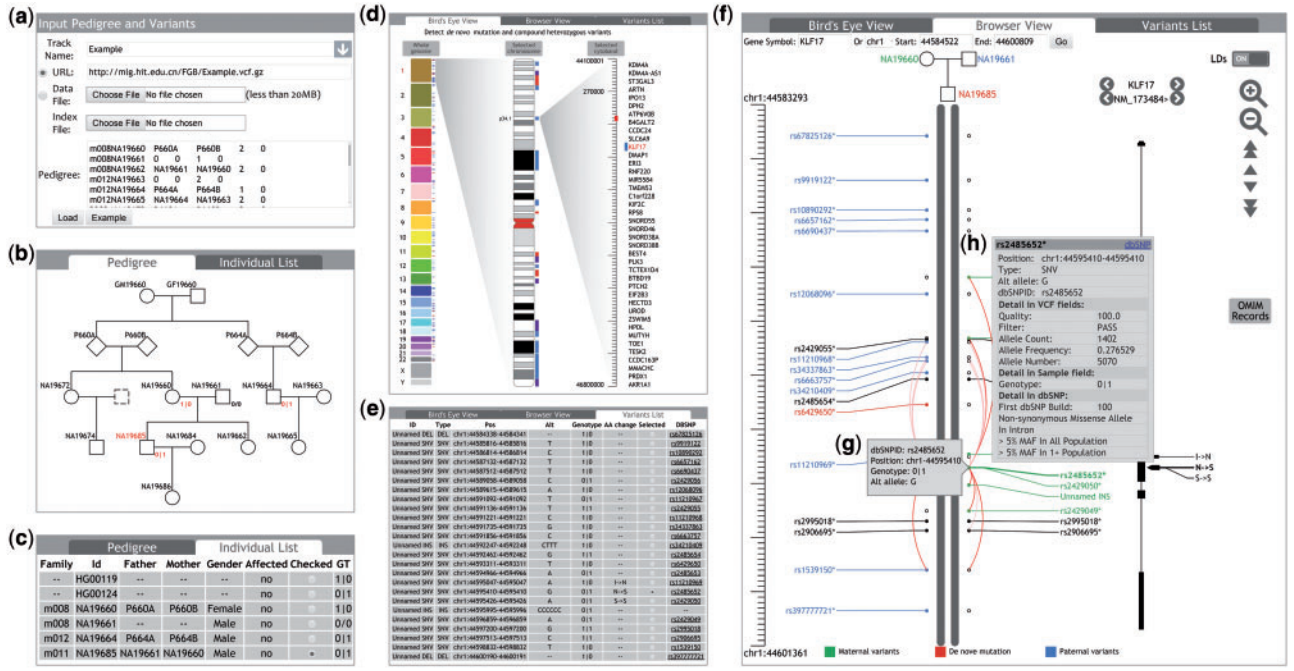

Fig. 1. Overview of the FGB. The interface of the FGB is divided into three panels: Data input panel (top-left of the website), Family panel (bottom-left of the website) and Genome panel (right of the website). (**a**) Data input panel. (**b**) 'Pedigree' tab in Family panel. (**c**) 'Individual List' tab in Family panel. (**d**) 'Bird's Eye View' tab in Genome panel. (**e**) 'Variant List' tab in Genome panel. (**f**) 'Browser View' tab in Genome panel. (**g**) Variant metadata in 'Browser View'. (**h**) 'Detail Window' in 'Browser View'

2.2 Pedigree visualization

Users may choose to input the pedigree information and upload the VCF file at the same time, or to upload the VCF file only. The pedigree information can be updated at any time, with the pedigree chart automatically updated. The individuals in PED text can be not necessarily identical to the samples in VCF data. Because it is very common that genome data of some family members is difficult to be collected by researchers, family studies usually only sequence a subset of members of the whole family. This feature enables users to input and view a more complete pedigree chart than merely the family members whose genomic variants data are available. Selecting a data-available individual in the pedigree chart, genomic variants of the individual can be displayed in the 'Browser View' tab (Fig. 1f).

In some cases, individuals whose data are available may not be in the pedigree. For example, HG00119 and HG00124 are second-order relatives (grandparents, cousins, etc.) as annotated in 1000 genomes project datasets, but the pedigree chart cannot be plotted based on this knowledge. Users can find and select these individuals in the 'Individual List' tab (Fig. 1c). All samples available in the input VCF file are listed in the 'Individual List' tab. If an individual is in the pedigree, his/her basic information described in the input PED text, such as family ID, individual ID, father ID, mother ID, gender, phenotype, etc., are also listed.

2.3 Family genome visualization and analysis

The genomic variants are displayed in the 'Browser View' tab (Fig. 1f) when a user selects an individual in the pedigree chart or individual list. If the genotypes of the genomic variants are phased, alleles displayed in the same flank can be considered to be on the same homologous chromosome. On the contrary, if the genotypes of the genomic variants are unphased, alleles displayed in the same flank cannot be considered to be on the same homologous chromosome.

However, whether the variants are phased or not, the FGB enables users to distinguish the paternal variants, maternal variants and *de novo* mutations by the trio analysis function. Specifying a group of individuals other than the selected individual, denoted by *A*, investigators can view the variants that are shared between the selected individual and *A*, or the variants that are unique in the selected individual genome. Thus, the FGB provides investigators three analysis modes: (i) trio analysis, (ii) highlight shared variants and (iii) highlight unique variants. Users can change the analysis mode on demand by clicking the individual symbol on top of the genome display in the 'Browser View' tab.

By default, for an individual whose parents' data are available, the FGB identifies paternal variants, maternal variants and *de novo* mutations based on the genotype information of the trio, and labels all distinguishable ones in different colours (Fig. 2a). The alternative alleles of maternal variants are labelled in green; the alternative alleles of paternal variants are labelled in blue and the alternative alleles of *de novo* mutations are labelled in red. Alternative alleles of other variants, including homozygous variants and variants whose parental origin cannot be determined, are displayed in black. Regardless of the genotypes of the variants are phased or unphased, this analysis can be always performed. However, if the variants of the trio are all phased by existing phasing tools in advance, such as SHAPEIT2 (Delaneau *et al.*, 2013), MVNCALL (Menelaou and Marchini, 2013), GATK PhaseByTransmission (DePristo *et al.*, 2011), etc., potential recombination event is detected (Kong *et al.*, 2008) and displayed, and an exclamation mark is rendered to indicate the event (Fig. 3a). If haplotype phases of the trio are assigned unambiguously, the displayed event is real recombination event (Auton and McVean, 2012; Roach *et al.*, 2010). For individuals whose parents' data are not available, alternative alleles of all variants are labelled in black.

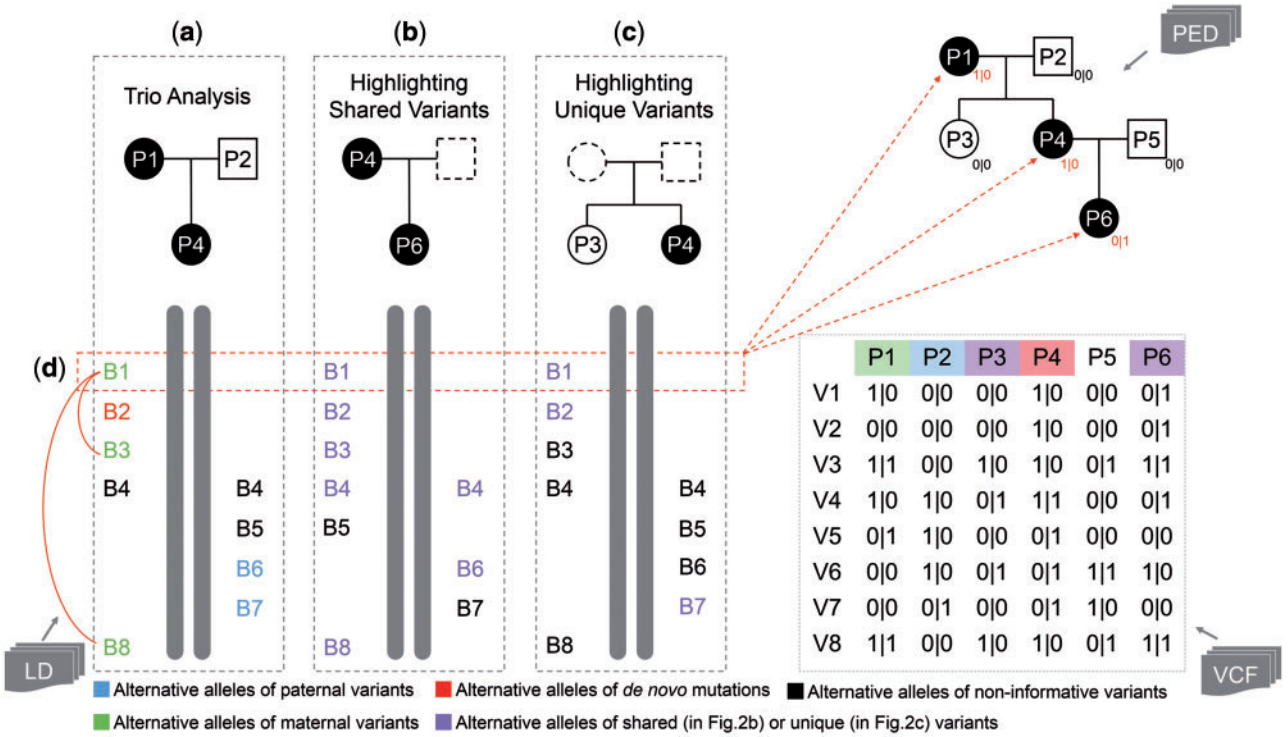

Fig. 2. Family genome visualization and analysis. (**a**) P4-centered trio analysis: P4 versus the parents P1 and P2. (**b**) P4-centered shared variants highlighting: P4 versus P6. (**c**) P4-centered unique variants highlighting: P4 versus P3. (**d**) Selecting a particular variant and illustrating its genotypes in the whole family, as well as LDs with other variants

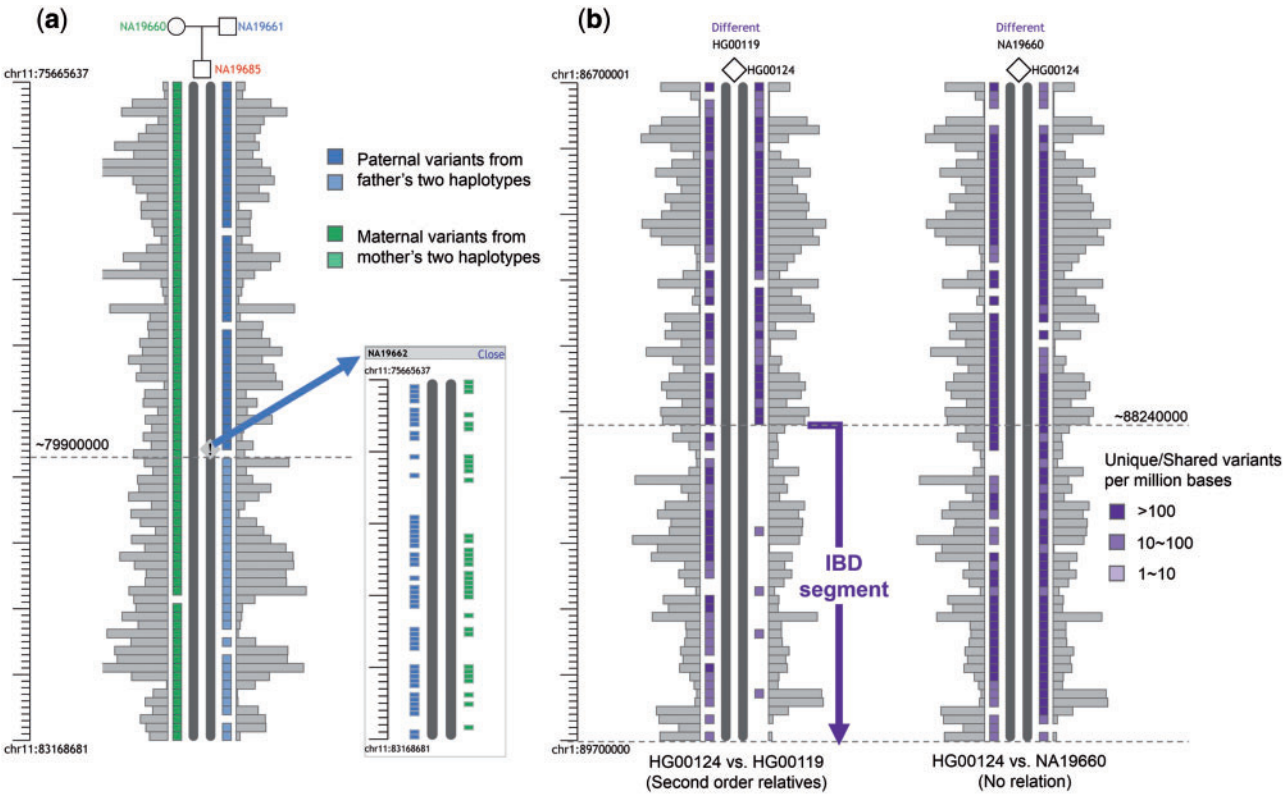

Fig. 3. Identification of potential recombination events and IBD segments. (**a**) A potential recombination event at ~79 900 000, 11p15.3 of NA19685 trio. (**b**) An IBD segment from ~88 240 000, 1p22.3 for second order relatives HG00124 and HG00119

Table 1. Functional roles and effects of coding region variants

Functional role of variant	Examples of variant effects displayed alongside genes	Variant type
Synonymous SNV	Q->Q, Stop->Stop . . .	SNV
Non-synonymous SNV	Q->R, . . .	SNV
Stop gain	Stop gain	SNV/INDEL
Stop loss	Stop loss	SNV/INDEL
Start codon loss	Initiator loss	SNV/INDEL
Frame shift	Frame shifting	INDEL%3!=0[a]
Insertion	_->R, QR->QPR, . . .	INS %3==0[a]
Deletion	Q->_, QPR->QQ, . . .	DEL %3==0[a]
ASS/DSS loss	ASS/DSS loss	SNV/INDEL

[a]Percentage refers to the MOD function.

Users may specify a group of individuals and investigate the shared variants among these individuals and the selected individual. The alternative alleles of shared variants are labelled in purple. As shown in Figure 2b, this feature is developed to find disease causal variants that are shared among disease-affected family members.

Users may also specify a group of individuals and investigate the unique variants that only present on the selected individual. In this mode, the alternative alleles of unique variants are also labelled in purple. As shown in Figure 2c, by comparing the selected individual which is affected by a phenotype, and the individual group that is not affected by the phenotype, this feature can help investigators to eliminate non-causal variants. It can also be used to identify IBD segments. For example, Figure 3b clearly illustrates an IBD segment between the second order relatives HG00119 and HG00124.

2.4 Comprehensive annotation to genomic variants

For facilitating users to analyse the family genome data, the FGB can annotate the input genomic variants by integrating with several knowledge bases. The annotation is performed in three layers: (i) genetic variants, (ii) transcriptional and translational consequences caused by the variants and (iii) potential phenotypes.

For genetic variants, dbSNP records and LDs are annotated to the variants. The FGB retrieves dbSNP records based on chromosomal positions of the variants, and provides dbSNP IDs and corresponding links to dbSNP website for the variants in the 'Variant List' tab. If a variant is recorded in dbSNP but unnamed in the input VCF file, the queried dbSNP ID is shown in the 'Browser View' tab as the name of the variant. In this case, we append a '*' to the dbSNP ID to distinguish the unnamed variants from the variants that already have dbSNP IDs in the input VCF file. For VCF files that have been annotated with an earlier version of dbSNP, users can quickly focus on variants that are submitted to the later versions of dbSNP recently.

LD describes the non-random association of alleles in population genetics, which is very useful in family based studies. The FGB visualizes LDs by the curves connecting the alternative alleles of variants. Colour gradients of the red curves represent the R^2 values of LDs between the alleles. The curves are not displayed unless the R^2 values are >0.2. The LD curves can be optionally displayed or hidden by clicking the 'LDs' switch on the top-right corner of the 'Browser View' tab (Fig. 1f).

Transcripts of genes are displayed on the right of genetic variants. By switching the option buttons left and right (Fig. 1f), users can selectively view a single transcript, all transcripts of a gene, or all transcripts of all genes in the browsing region. Users can also investigate the details of genes and transcripts in the HGNC and RefSeq websites by clicking the names of the genes and the transcripts.

The effects of variants disrupting protein coding, i.e., coding region variations, are displayed alongside the genes. A variety of coding region variations, such as amino acid changes, splicing event changes, etc., play important roles in molecular mechanisms of genetic diseases. The visualization of the functional effects of genetic variants can greatly help family based genetic disease studies. The major coding region variations are summarized in Table 1.

For the phenotype analysis, OMIM records that are associated to genes in the browsing region can be queried and displayed by clicking the 'OMIM record' button, if applicable. Similar to the dbSNP, HGNC and RefSeq annotations, the FGB provides links to the corresponding OMIM webpage for details of OMIM entries.

2.5 Detail information

The position, alleles, dbSNP ID and genotype of the variant are displayed when users hover over an individual variant (Fig. 1g). To further investigate the genotypes of a variant in the whole family, users can view the variant genotypes of all samples in the input VCF file by selecting a variant (Fig. 2d). The genotypes are displayed in the right-bottom corner of the symbols of family members in the pedigree chart, and in the eighth column of the individual list. If the LDs option is on, LD curves irrelevant to the selected variant are hidden for highlighting the ones connected to it. If the selected variant locates in a protein-coding region, its effect, such as amino acid change, is also highlighted using a bold typeface. Moreover, more detailed variant annotation data from the VCF fields is displayed in a 'Detail Window' (Fig. 1h).

Users may also browse the displayed variants in the 'Variant List' tab (Fig. 1e). The detail information of the variants, such as ID, type, chromosomal position, alternates, genotypes, effects in protein-coding region, the corresponding dbSNP ID and link to dbSNP website, etc., are listed in a table. In the table, the variants are sorted by their chromosomal positions. Users can also select any variant in the table to view its genotypes in the whole family.

2.6 Automatic detection of *de novo* mutations, compound heterozygous variants

Bird's Eye View is a three-column genome view (Fig. 1d). All chromosomes are displayed as rectangle buttons in the left. The heights of rectangles are proportionate to the lengths of the corresponding chromosomes. When a chromosome is selected, the cytobands of the chromosome are displayed in a chromosome shape in the middle. When a cytoband is selected, the genes located in the selected cytoband are listed in the right area. Clicking on the gene symbols, users can view the corresponding region in the 'Browser View' tab.

In the 'Bird's Eye View' tab, the FGB can scan an individual genome to detect *de novo* mutations and compound heterozygous variants in coding regions of a selected individual. *De novo* mutations do not present on parents. Coding region *de novo* mutations, especially the non-synonymous ones, may cause dominant genetic diseases that do not affect the parents. Compound heterozygous variants, inherited from the father and the mother, respectively, are both non-synonymous and locate in the same gene. The parents may appear normal as they still have the reference allele, but the offspring can be affected by the recessive disease, for he/she inherits both deleterious variants.

After scanning, potential high risk genes containing the two kinds of suspicious variants are highlighted with red, blue or purple

marks. Red marks indicate the genes containing coding region *de novo* mutations. Blue marks indicate the genes containing compound heterozygous variants. Purple marks indicate the genes containing both of them. Similarly, red, blue or purple marks are labelled to the cytobands and chromosomes containing colour-marked genes. The detection of *de novo* mutations and compound heterozygous variants can be performed on the whole genome or a selected chromosome/cytoband. The variants data of the selected individual's parents should be available in the input VCF files. This feature offers investigators an automatic method to find candidates of disease-related genes and disease-causing variants for rare genetic disease patients whose parents appear normal.

2.7 Navigation

In the 'Browser View' tab, the axis and the range of current browsing region are displayed in the left. Moving the cursor over the axis, a horizontal line and the corresponding coordinate are shown to help users positioning the variants, the genes and coding region variations. By clicking and dragging the cursor on the axis, users can navigate to an interested browsing region.

Besides, the FGB provides flexible navigation ways to enable users to view a desired genomic region by specifying the genomic coordinates or gene symbols, as well as panning and zooming the browsing region. Users can also navigate to the corresponding region of a selected gene in the 'Bird's Eye View' tab.

In different zooming levels, the FGB displays the variants in different ways to avoid requesting over-sufficient data beyond the display limit. For browsing region <50 bases, reference genome sequences are shown by colours and letters instead of a simple axis. For browsing region >50 bases and <425 bases, reference genome sequences are shown by colours only. For the displayed variants number <55, the FGB displays dots to represent alternative alleles of the variants at their chromosomal positions, and labels the variants' names to the dots. For variants number >55 and <270, the FGB displays dots only. For variants number >270, the FGB divides the browsing region into dozens of bins and displays histogram bars to show the relative variants number in each bin (Fig. 3). In this resolution, the black-coloured dots are not displayed as the dots are illegible. If applicable, colourful dots are replaced by dozens of coloured mosaics, for helping users to recognize *de novo* mutations, recombination events, IBD segments, etc. The FGB is able to visualize family genomes from a single nucleotide to a whole chromosome in the 'Browser View' panel.

More details about the interfaces and usages of the FGB are available at http://mlg.hit.edu.cn/FGB/tutorial.html.

3 Implementation

The FGB is a browser/server architecture-based web application. The back end of the FGB is implemented in JAVA. Apache Tomcat is used to provide web services. The genomic data processing results are packed into XML objects for transferring and displaying. In the front end of the FGB, the Asynchronous JavaScript and XML (AJAX) technique is adopted for exchanging data asynchronously between the browser and the server to avoid full page reloads. The scalable vector graphics (SVG) is used as the graphic engine to plot visual elements and to interact with users. The advances of the SVG enable the FGB to be freely zoom in and out while remaining the high definition of the visualization, and thus can be used on display devices in any resolution.

Comprehensive resources are integrated in the FGB to annotate family genomes, including hg19, dbSNP142, LD, HGNC, RefSeq, OMIM, etc. Most of the genomic annotation data were downloaded from the UCSC genome browser database (Karolchik *et al*., 2014). The LD R^2 values used in FGB are calculated based on 1000 genomes phase 3 dataset (1000 Genomes Project Consortium, 2012) by PLINK (Purcell *et al*., 2007). Related individuals are excluded in the LD calculation. Tabix API (Li, 2011) is embedded in FGB for parsing the Bgzip/Tabix compressed/indexed VCF files.

To access the public site for more information, please visit http://mlg.hit.edu.cn/FGB/. For general questions regarding the FGB, please contact user support via email at pgbrowser@gmail.com. The FGB is best accessed using Google Chrome and works smoothly as well with other web-browsers, including Mozilla Firefox, Safari, Microsoft Internet Explorer (Version 10 or later), Opera, etc. A tutorial of the FGB is available at http://mlg.hit.edu.cn/FGB/tutorial.html. Users may freely obtain a copy (http://mlg.hit.edu.cn/FGB/FGB-1.0.tar.gz) of the FGB to install it locally. The source codes of the FGB are also publicly available at GitHub (https://github.com/lrjuan/FGB/).

4 Discussion

The FGB provides investigators accurate, comprehensive and global visualization for family genomes. The visualization of family genomes is an important and special case of multiple genome visualization. First, the relations and phenotypes of family members are essential for family based studies, and should be described and integrated in variant-level genome visualization. Second, the similarity between the genomes of consanguineous family members is much higher than the similarity between unrelated individuals. Thus the redundancy of family genomes should be well-handled in visualization. The FGB is designed and developed to achieve these special requirements, and provides users accurate and appropriate visualization for family genomes.

Though the VCF files support the storage of variants of multiple genomes, the pedigree information, e.g., genders, relations and phenotypes of the multiple genomes, can hardly be described by the data format. The FGB accepts the *de facto* standard format, PED format, as the input format of pedigree information, integrates the pedigree information with the VCF data, then analyses and visualizes the genomic variants in the three versatile analysis modes. By using a single or combination of the three analysis modes, users are able to properly investigate family genomes in variant-level, such as distinguishing paternal variants, maternal variants and *de novo* mutations, identifying potential recombination events and IBD segments, finding or eliminating disease-causing variants and genes, etc.

As the solution to reduce the redundancy in family genomes, the FGB displays an individual-centered genome view, and coloured variants to show their presence/absence states on genomes of other family members. Comparing genomes is the most basic purpose of the multiple genomes visualization. Comparative genomics visualization tools focus on visualizing large-scale differences among genomes of different species. Human multiple genomes visualization tools are committed to visualize 3 million differences, not 3 billion bases. For family genome visualization, differences are far <3 million. Thus the FGB chooses the individual-centered genome view as the basis of the comparison, and enables users to concentrate on differences/similarities between the central individual and a group of specified family members. This feature avoids showing dozens of similar genomes aligned together, which is not only hard to capture the key information, but also costly in performance. By selecting the

switchable central individual, the comparing group of individuals, and the analysis mode, users may fully customize the analysis method and the family members to be analysed, then view the analysis result. To further investigate a particular variant, users can select the variant and view its genotypes in the whole family.

The FGB provides comprehensive annotations for family genomes. dbSNP memberships, LDs, genes and variants functional effects, potential associated phenotypes, etc., are efficiently annotated to the displayed genetic variants on the fly. This feature may help users to understand the details and the potential functional effects of the variants comprehensively, and thus to quickly locate the suspicious variants.

The FGB offers a function to automatically detect *de novo* mutations and compound heterozygous variants by scanning the genomes of the selected individual and his/her parents. Based on two common models of disease inheritance, dominant model and recessive model, the FGB searches rare variants or rare combinations of variants that may cause genetic diseases. This feature guides users to create global insight of high risk genes on the selected individual.

Advanced bioinformatics technologies, such as asynchronous data transferring, SVG, flexible navigation options, legibility-based data detail limiting, etc., have been adopted to improve the FGB performance, such as reducing the bandwidth requirements, balancing the server load and improving the user experiences. With increasingly producing family genomes, the FGB can be widely used to display, interpret and analyse family genomes and greatly benefit academic researchers and clinical physicians.

Acknowledgements

We thank Prof. Qingyuan Zhang M.D., Dean of Harbin Medical University Cancer Hospital, Dr Yunlong Liu, Associate Professor of Indiana University School of Medicine and Dr Shijia Zhu, Postdoctoral Fellow of Mount Sinai Hospital, for their valuable comments and suggestions to the system design. We thank Dr Guohua Wang, Dr Jian Liu, Dr Yang Hu, Chengwu Yan, Ling Wang, Yang Bai, Jiajie Peng and Yanshuo Chu for extensive FGB testing.

Funding

This study is supported by grants 2012AA020404, 2012AA02A602, 2012AA02A604, and 2015AA020101, sponsored by National High-Tech Research and Development Program (863) of China.

Conflict of Interest: none declared.

References

Auton,A. and McVean,G. (2012) Estimating recombination rates from genetic variation in humans. *Methods Mol. Biol.*, **856**, 217–237.

Axelrod,N. *et al.* (2009) The HuRef Browser: a web resource for individual human genomics. *Nucleic Acids Res.*, **37**, D1018–D1024.

Choi,S. *et al.* (2014) FARVAT: a family-based rare variant association test. *Bioinformatics*, **30**, 3197–3205.

NCBI Resource Coordinators, (2013) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **41**, D8–D20.

Danecek,P. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Delaneau,O. *et al.* (2013) Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods*, **10**, 5–6.

DePristo,M.A. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet*, **43**, 491–498.

Fiume,M. *et al.* (2012) Savant Genome Browser 2: visualization and analysis for population-scale genomics. *Nucleic Acids Res.*, **40**, W615–W621.

Flicek,P. *et al.* (2014) Ensembl 2014. *Nucleic Acids Res.*, **42**, D749–D755.

1000 Genomes Project Consortium, (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.

Gray,K.A. *et al.* (2013) Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res.*, **41**, D545–D552.

Hamosh,A. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.

Juan,L. *et al.* (2014) The personal genome browser: visualizing functions of genetic variants. *Nucleic Acids Res.*, **42**, W192–W197.

Karolchik,D. *et al.* (2014) The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.*, **42**, D764–D770.

Kent,W.J. *et al.* (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.

Koboldt,D.C. *et al.* (2014) Exome-based mapping and variant prioritization for inherited Mendelian disorders. *Am. J. Hum. Genet.*, **94**, 373–384.

Kong,A. *et al.* (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.*, **40**, 1068–1075.

Kumagai,M. *et al.* (2013) TASUKE: a web-based visualization program for large-scale resequencing data. *Bioinformatics*, **29**, 1806–1808.

Li,H. (2011) Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics*, **27**, 718–719.

Menelaou,A., Marchini,J. (2013) Genotype calling and phasing using next-generation sequencing reads and a haplotype scaffold. *Bioinformatics*, **29**, 84–91.

Paila,U. *et al.* (2013) GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS. Comput. Biol.*, **9**, e1003153.

Peng,G. *et al.* (2013) Rare variant detection using family-based sequencing analysis. *Proc. Natl. Acad. Sci. USA.*, **110**, 3985–3990.

Purcell,S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.

Roach,J.C. *et al.* (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*, **328**, 636–639.

Robinson,J.T. *et al.* (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

San Lucas,F.A. *et al.* (2012) Integrated annotation and analysis of genetic variants from next-generation sequencing studies with variant tools. *Bioinformatics*, **28**, 421–422.

Schatz,M.C. *et al.* (2007) Hawkeye: an interactive visual analytics tool for genome assemblies. *Genome Biol.*, **8**, R34.

Sherry,S.T. *et al.* (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.

Sincan,M. *et al.* (2012) VAR-MD: a tool to analyze whole exome-genome variants in small human pedigrees with mendelian inheritance. *Hum. Mutat.*, **33**, 593–598.

Stein,L.D. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.

Thiele,H. and Nurnberg,P. (2005) HaploPainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics*, **21**, 1730–1732.

Voorrips,R.E. *et al.* (2012) Pedimap: software for the visualization of genetic and phenotypic data in pedigrees. *J. Hered.*, **103**, 903–907.

Yao,J. *et al.* (2014) FamAnn: an automated variant annotation pipeline to facilitate target discovery for family-based sequencing studies. *Bioinformatics*, **30**, 1175–1176.

Zhou,X. *et al.* (2011) The Human Epigenome Browser at Washington University. *Nat. Methods*, **8**, 989–990.