FABRIC GENOMICS™

Application Note

Fabric Genomics Structural Variation Browser

The Fabric Genomics Enterprise (FE) platform provides a Structural Variation browser (SV browser) tool to allow the exploration of patient's structural variants alongside annotations from different sources that are useful for the interpretation of SVs called from whole-genome sequencing data (WGS). Interpreting SVs from WGS data aims to increase next-generation sequencing (NGS) diagnostic yield by replacing diagnosis of microdeletion syndromes by conventional CGH microarrays, further allowing to evaluate SVs at more fine scale.

The FE SV browser is derived from the open source Genoverse browser (http://genoverse.org/) which the DECIPHER project (https://decipher.sanger.ac.uk) uses to display their own patient variant pages. This tool offers interactive visualization of SVs on the reference genome with additional tracks that display affected genes, overlapping variants from DECIPHER patients, overlapping variants from population and other pathogenic variant databases such as NCBI's ClinVar database, ISCA, LSDB, and other resources helpful in the interpretation of SVs in the context of rare genetic diseases.

Requirements

SV browser uses HTML5 technology and will work on all recent internet browsers. It has been tested on recent versions of Google Chrome, Safari, and Firefox 20+.

The samples should have been analyzed by algorithms that identify SVs from WGS NGS data (not provided by Fabric) and filtered to minimize the inclusion of methodological artifacts and common events in populations before interpretation. SVs should have been merged with SNV and small indels calls in a single VCF file and uploaded into Fabric Enterprise for annotation and clinical reporting.

In addition, a clinical test in Fabric enterprise should have been configured with a standard operating procedure (SOP) step that includes interpretation of SVs.

Loading the browser

SV browser is loaded within Fabric Enterprise automatically when clicking on the "VIZ" icon under the variant position column in the "All SVS" SOP tab (see Figure 1). The SV browser will be opened as a separate additional browser tab to allow the user to look back at the SOP steps and variant scoring sheet independently. The view of the browser is automatically centered on the case variant under review, which appears in the "Case SV" track, immediately under the "Genes" track (see Figure 2).

Fabric Genomics | 1611 Telegraph Avenue, Suite 500, Oakland, CA 94612 | 510.595.0800 | info@fabricgenomics.com

© 2019 Fabric Genomics™, Inc. All rights reserved. Not for use in clinical diagnostic.

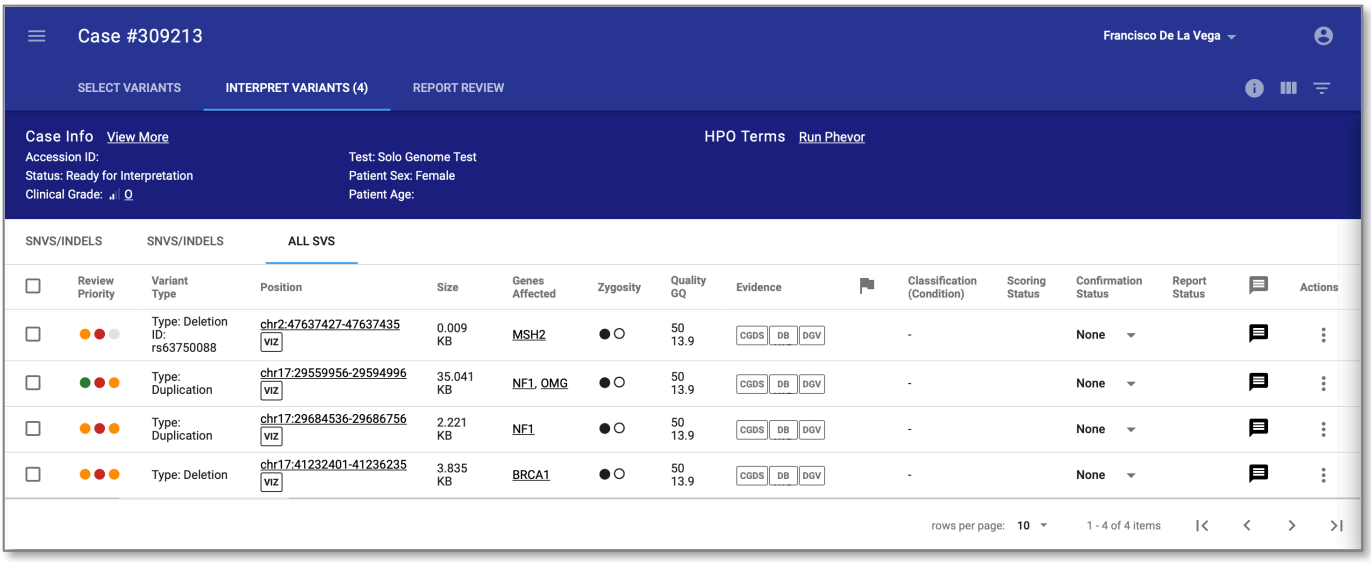

Figure 1. A new "VIZ" icon under variant position in the "ALL SVS" SOP tab launches SV browser in a separate track (Fig. 2), centered on the variant under review.

Navigation

Information is presented horizontally following the coordinates of the genome, in 5' to 3'direction, for the chromosome where the SV of interest resides. Although the browser is already centered in the SV of interest, it is possible to zoom in or out further in the genomic region and navigate to adjacent genomic regions. The navigation menu is on the right of the page, or mouse click-drag action can be also used to control the browser. Clicking on the "i" icon on the right-side panel reveals the tooltips providing help on the buttons, data tracks and general layout of the browser (see Figure 2).

The karyotype representation of the entire chromosome is the first track of the browser. This karyotype has a sliding window that can be moved to quickly visualize variants along the chromosome, resized to increase or decrease selected views.

Often the screen size is not sufficient to see all data tracks and information provided, and it is necessary to scroll down to see additional data tracks. A track height can also be insufficient to show all evidence; in such cases a warning will be shown on the left side of the tack. Track can be resized to show more content by clicking the warning "resize track" or by pulling up or down the small gray handle in the middle bottom of each track.

Fabric Genomics | 1611 Telegraph Avenue, Suite 500, Oakland, CA 94612 | 510.595.0800 | info@fabricgenomics.com

© 2019 Fabric Genomics™, Inc. All rights reserved. Not for use in clinical diagnostic.

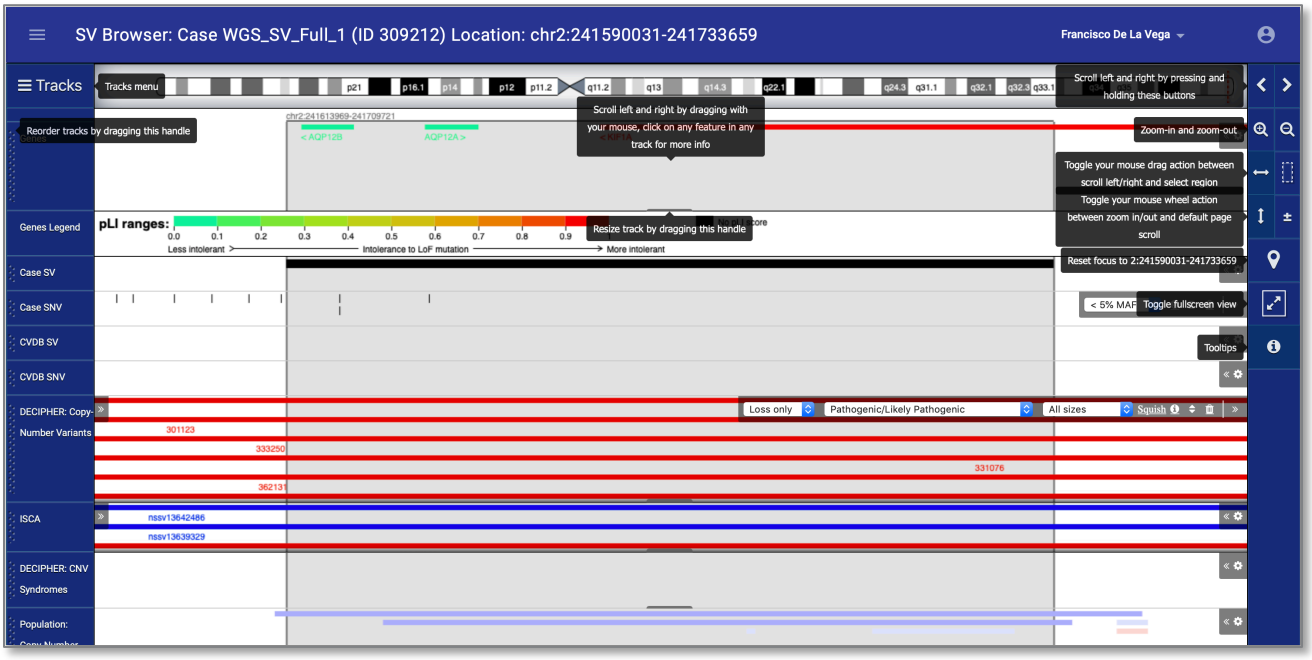

Figure 2. SV browser main screen with tooltips activated (by clicking on the “i” icon), showing the functions of the navigation tools and other functionality.

Browser Tracks

The browser has many data tracks available including genes, the case SV and SNVs, the SV and SNVs present in the workspace Classified Variant Database (CVDB), and other relevant data tracks sources directly from DECIPHER (Table 1). The two boundaries of the tracks are used to provide information regarding the tracks. The left side of the tracks is used for informational messages such as “Reference sequence not displayed at this zoom level” etc. These messages may be dismissed by clicking on the “<<” on these messages. The right-side of the track (<<) provides additional filters, help and a toggle for removing the track altogether.

By default, all protein coding genes are shown colored by their gnomAD *pLI* scores in the genes track, but these may be filtered to show only those genes that have LSDB or are known development disorder genes. The view may be changed to “squish” to compact the display of features in the track and reduce default width:

Fabric Genomics | 1611 Telegraph Avenue, Suite 500, Oakland, CA 94612 | 510.595.0800 | info@fabricgenomics.com

© 2019 Fabric Genomics™, Inc. All rights reserved. Not for use in clinical diagnostic.

Other tracks have filters that are relevant to the data type, such as filtering by minor allele frequency for case SNVs, deletions (loss) or duplications (gains) in DECIPHER CNV tracks, size range of events, and consequence of ClinVar variants. The “i” icon and the trashcan are used for showing help and turning the track off, correspondingly.

Moving/Enabling/Disabling Tracks

By default, certain tracks are opened and displayed in a default order (see Table 1). However, the browser tracks may be repositioned to create a custom view. To move tracks up or down, click and drag the tracks on the left side. Further, additional tracks are available and can be selected for display by clicking on the “Tracks” drawer on the left of the karyotype to open the track management panel and clicking on the “+” icon (see Figure 3). Likewise, tracks can be removed from this panel by clicking on the “x“ icon.

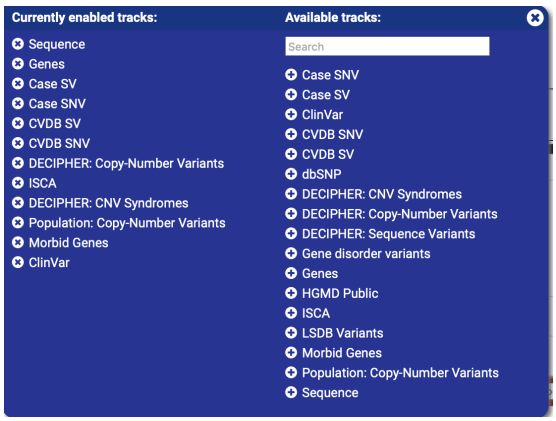

Figure 3. Track management panel.

Repositioning tracks allows groups of relevant data to be viewed together.

Tracks may be turned off from the track-specific menus on the right of each track (<<) or from the “Tracks” button on the top right. The trashcan on the listed tracks (left) will turn the track off, the “+” on the available tracks (right) will add the track to the browser. Tracks may be enabled/disabled/duplicated in this manner.

> *Duplicated tracks are useful to view the same subset of data using different filters. For example, duplicating the genes tracks will allow developmental disorder genes and genes*

Fabric Genomics | 1611 Telegraph Avenue, Suite 500, Oakland, CA 94612 | 510.595.0800 | info@fabricgenomics.com

© 2019 Fabric Genomics™, Inc. All rights reserved. Not for use in clinical diagnostic.

colored by pLI to be seen together. Similarly, CNV track could be duplicated into gain and loss track and seen separately rather than together in a single track or one at a time.

Context information on track items

Each item displayed on a track has an associated context information box that is used to provide more information about the item clicked, as well as provide links to other relevant external resources, when available. Following these links will create a new tab in the external website describing in more details the data feature, e.g. a DECIPHER case. Close these menus by clicking on the “x” icon on the top right.

Quick focus on feature

The context menu contains a new “Focus here” item which can be used to quickly focus on the relevant feature. Click on focus here and the browser immediately zooms in or out to show the feature in question as relevant resolution. The “focus here” function is available for every item shown on the genome browser allowing quick navigation and analysis of region of interest. At any time, the “reset focus” button may be used to go back to initial region of interest.

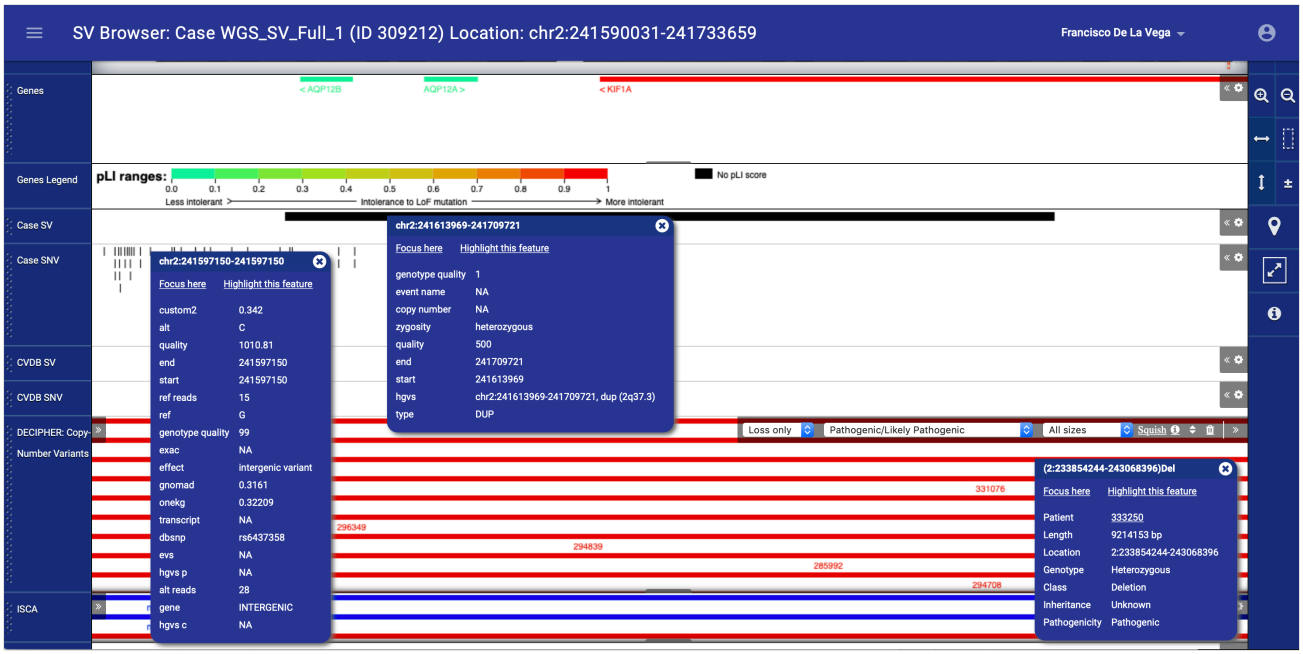

Figure 4. Context information panels can be displayed by clicking on a given feature in a track. Content varies base on the feature selected. Some hyperlinks to the original source area available. Focus here and highlight functions can be invoked by clicking on the corresponding text.

Fabric Genomics | 1611 Telegraph Avenue, Suite 500, Oakland, CA 94612 | 510.595.0800 | info@fabricgenomics.com

© 2019 Fabric Genomics™, Inc. All rights reserved. Not for use in clinical diagnostic.

Sequence-level view

Sequence variants and associated population and pathogenic tracks are best viewed at high zoom levels where the browser automatically switches to show sequence variants on the reference genome sequence. To zoom in to a variant, use the navigation “+” from the right side, or shift-click-drag to select a short region of the view and choose “zoom here”. After a certain zoom level, the genomic reference sequence will begin to show and the DECIPHER patient sequence variant will be shown on the reference sequence (see Figure 5).

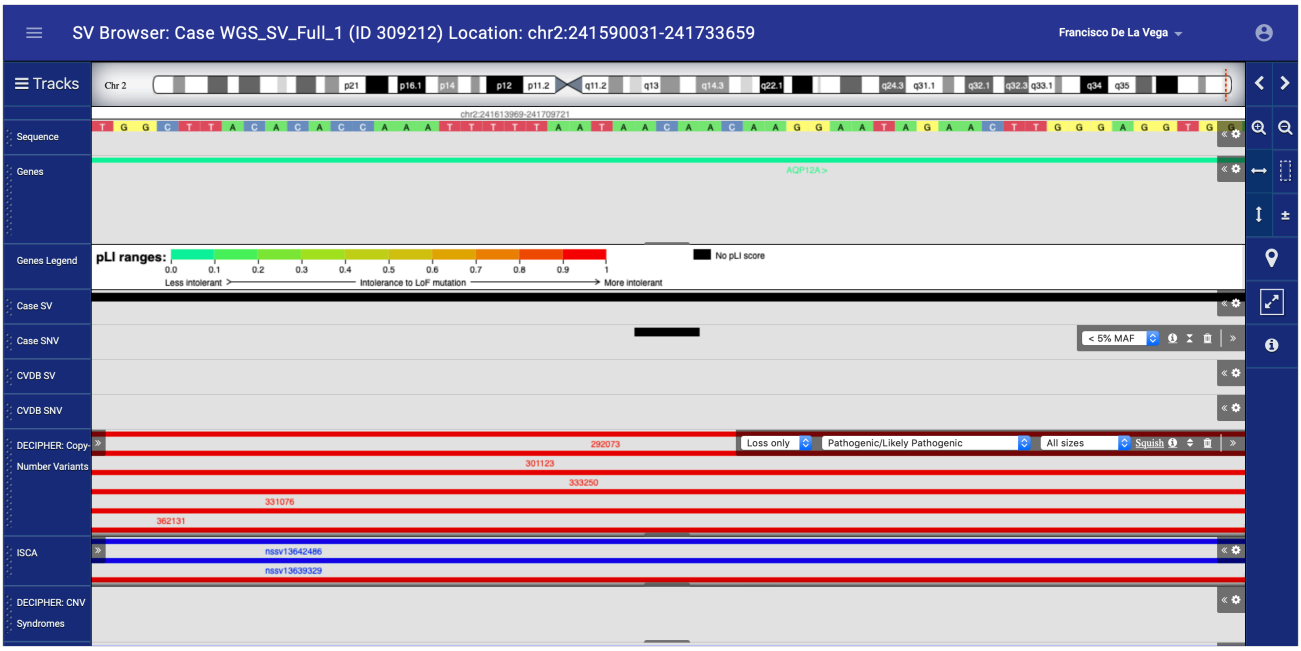

Figure 5. Sequence level view can be shown when zooming to a very small feature such as a SNV or by zooming in.

External Data Tiers

Beside the case SV and SNVs, and the SV and SNVs from the workspace CVDB, all other data tracks are provisioned directly from DECIPHER and mirror what is available in their browser (https://decipher.sanger.ac.uk/browser). Data is updated by DECIPHER approximately every two months.

> *Please note that Fabric Genomics does not express any warranty with respect to the accuracy of the data coming from the external data sources such as DECIPHER. The DECIPHER consortium provides these data in good faith as a research tool, but without verifying the accuracy, clinical validity or utility of the data. The DECIPHER consortium,*

Fabric Genomics | 1611 Telegraph Avenue, Suite 500, Oakland, CA 94612 | 510.595.0800 | info@fabricgenomics.com

© 2019 Fabric Genomics™, Inc. All rights reserved. Not for use in clinical diagnostic.

makes no warranty, express or implied, nor assumes any legal liability or responsibility for any purpose for which the data are used.

The table below shows the current list of available data tracks, whether they appear by default and their order, and version of the external data source, when available, from DECIPHER.

Tier	Default?	Source and version
Genes	YES	ENSEMBL Genes, colored by gnomAD pLI scores (gnomAD vr2.01).
Case SV	YES	The SV (deletion, duplication, inversion) of the case being interpreted.
Case SNV	YES	SNVs and small indels from the case in the region of zoom.
CVDB SV	YES	SVs previously reported in other cases present in the workspace Classified Variant Database.
CVDB SNVs	YES	SNVs and small indels previously reported in other cases present in the workspace Classified Variant Database.
DECIPHER CNVs	YES	Copy-number variants observed in DECIPHER patients. Deletions are represented in red and duplications in blue. The CNVs are also color-coded by user-defined pathogenicity values. You may filter these CNVs for gains, for losses, or by pathogenicity (v9.29).
ISCA	YES	Copy-number variants from the International Standards for Cytogenomic Arrays (ISCA) Consortium. You may filter these CNVs for gains, for losses, or by pathogenicity. Updated every two months.
DECIPHER Syndromes	YES	Micro deletion/duplication syndromes (curated by the DECIPHER consortium, v9.29).
Population: CNVs	YES	Common copy-number variants observed in the general population (curated by the DECIPHER consortium, v9.29).
Morbid Genes	YES	Genes that are associated with disease phenotypes (curated and maintained by OMIM). Updated every two months.
DECIPHER sequence variants	NO	Sequence variants observed in DECIPHER patients (v9.29).
ClinVar	NO	Variant data from the ClinVar database maintained by the NCBI. Updated every two months.
Gene Disorder Variants	NO	Variants associated with known disorders as annotated by GeneReviews. Updated every two months.
HGMD Public	NO	Sequence variants reported in the public version of the Human Gene Mutation Database (HGMD). Updated every two months.
dbSNP	NO	All sequence variants from the database of Single Nucleotide Polymorphisms (dbSNP). Updated every two months.
LSDB variants	NO	Variants from the Locus-specific databases (LSDB) maintained by the Leiden Open Variation Database (LOVD). Updated every two months.

Fabric Genomics | 1611 Telegraph Avenue, Suite 500, Oakland, CA 94612 | 510.595.0800 | info@fabricgenomics.com

© 2019 Fabric Genomics™, Inc. All rights reserved. Not for use in clinical diagnostic.

Known Limitations

There are a number of known limitations for the current release of the FE SV Browser:

- SV Browser is currently available only on the hGRCb37 assembly.
- While it is not disabled in WES and Panel data, the annotation tracks are useful for the analysis of SVs from patients called from WGS data and for the diagnosis of rare genetic diseases. These data and visualizations are not very useful for the analysis of CNVs called from targeted resequencing data and for other use cases.
- SV browser is independent of the FE SOP application panel. Any annotations referenced in the browser need to be brought manually to variant interpretation notes if desired or necessary by copy-paste.
- Track customization and display are not saved between sessions or variants.

Conclusions

The Fabric Enterprise Structural Variation browser allows users to review evidence from multiple sources relevant for the interpretation of SVs in rare genetic disorders and in an interactive form that can be used to interpret a case SV with respect to the patient phenotype. The visual format allows users to better appreciate if overlaps with such evidence is meaningful, and the ability to filter by even type, variant consequence and size, gene intolerance to changes, previously reported pathogenicity by DECIPHER, and curated microdeletion syndromes by ISCA and DECIPHER, thus enabling appropriate classification of such variants and increasing the diagnostic yield of clinical WGS.

Fabric Genomics | 1611 Telegraph Avenue, Suite 500, Oakland, CA 94612 | 510.595.0800 | info@fabricgenomics.com

© 2019 Fabric Genomics™, Inc. All rights reserved. Not for use in clinical diagnostic.