DGV Beta User Tutorial

Aims of DGV

- To provide information about **genomic variants identified in control samples**
- To serve as a **resource** to both **clinical and research labs**
- To show **variation in genomic context** in a simple genome browser
- To transparently provide **access to underlying source data** (downloadable)
- To be **publicly available at no cost**

Database Content

- **The majority of data sets in DGV are from four different types of studies:**

- Array based comparative genomic hybridization (CGH) and comparative intensity analysis (SNP/CNV arrays)
- Identification of deletions based on statistical analysis of SNP data
- Clone end sequence mapping
- Sequence trace mapping

What DGV Is Not

- An uncurated repository of primary data

- A substitute for a well-designed control experiment
  - The database is still limited in content
  - About 6,500 individuals represented; not ethnically matched

- Due to biases in studies to date, the content is almost certainly not an accurate representation of structural variation on a population scale.

DGV BetaHome Page

http://dgvbeta.tcag.ca/dgv/app/home

About the Project | Downloads | Links | Statistics | FAQ
Genome Browser | Query Tool | Submissions | Contact Us | Tutorial

Keyword, Landmark or Region Search: Search NCBI36/hg18

Examples: RP11-34P13; CFTR, 7q11.21; chr7:71890181-72690180

Find DGV Variants

by Study | by Sample
by Method | by Variant
by Platform | by Chromosome

Summary Statistics

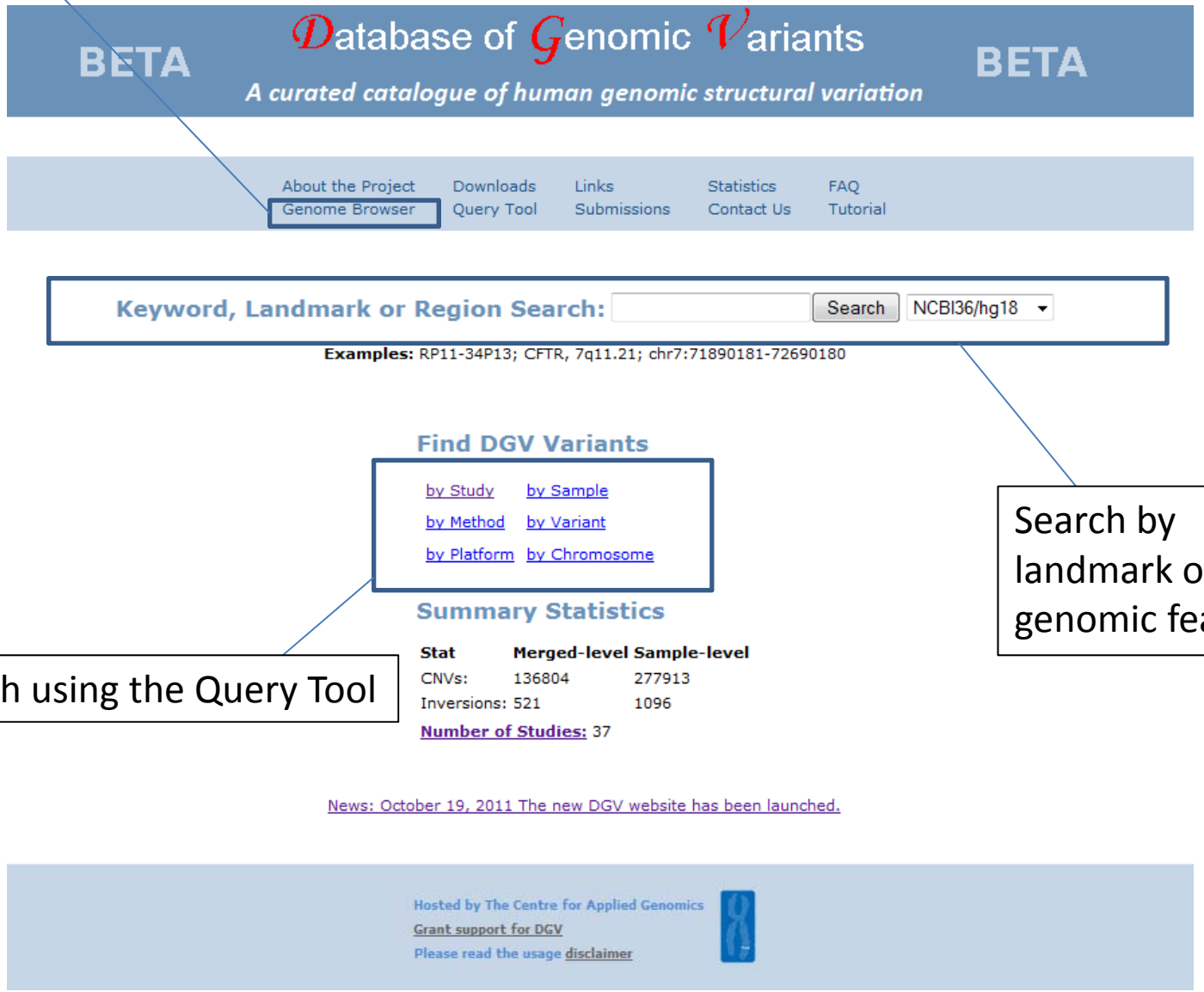
Stat	Merged-level	Sample-level
CNVs:	136804	277913
Inversions:	521	1096
Number of Studies:	37	

News: October 19, 2011 The new DGV website has been launched.

Hosted by The Centre for Applied Genomics
Grant support for DGV
Please read the usage disclaimer

Enter the Genome Browser

Points of Entry into the Database

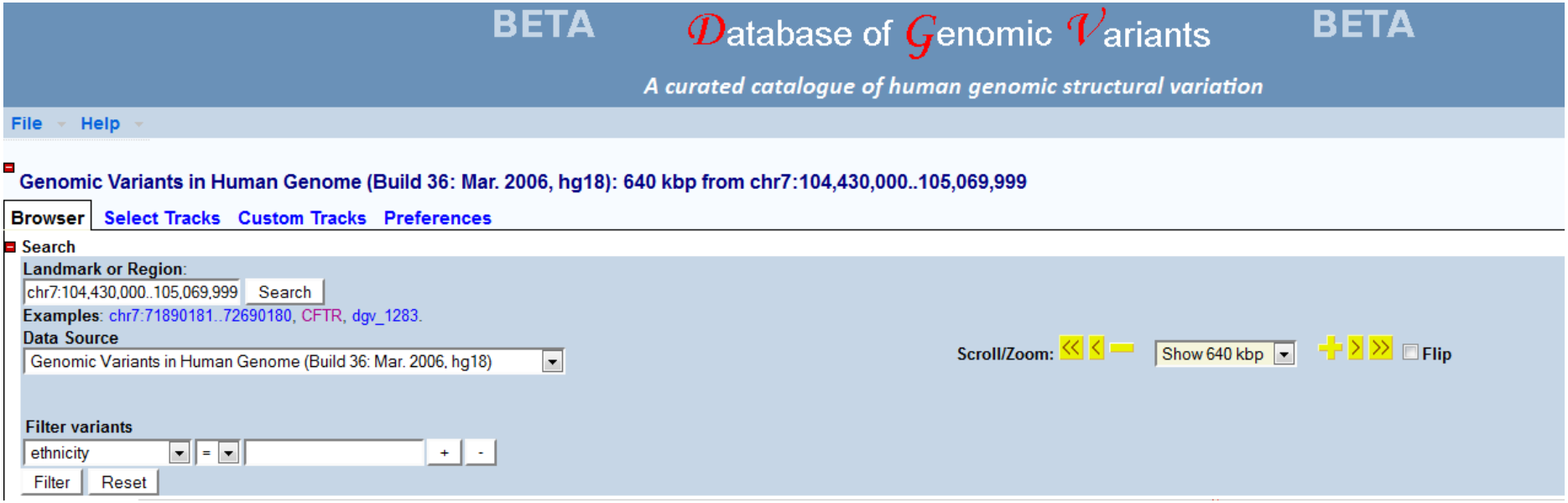


Search using the Query Tool

Search by landmark or genomic feature

Genome Browser

Many features of the genome browser are the same, with some functional differences and some new tools/options.

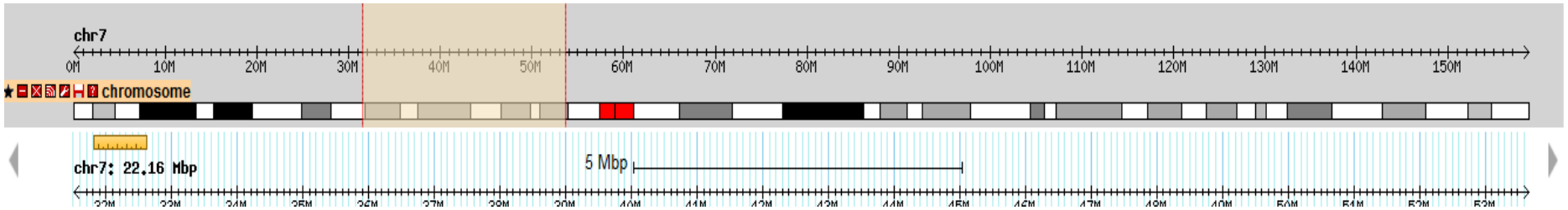


At the top of the browser window, there are options modifying the display in the current browser

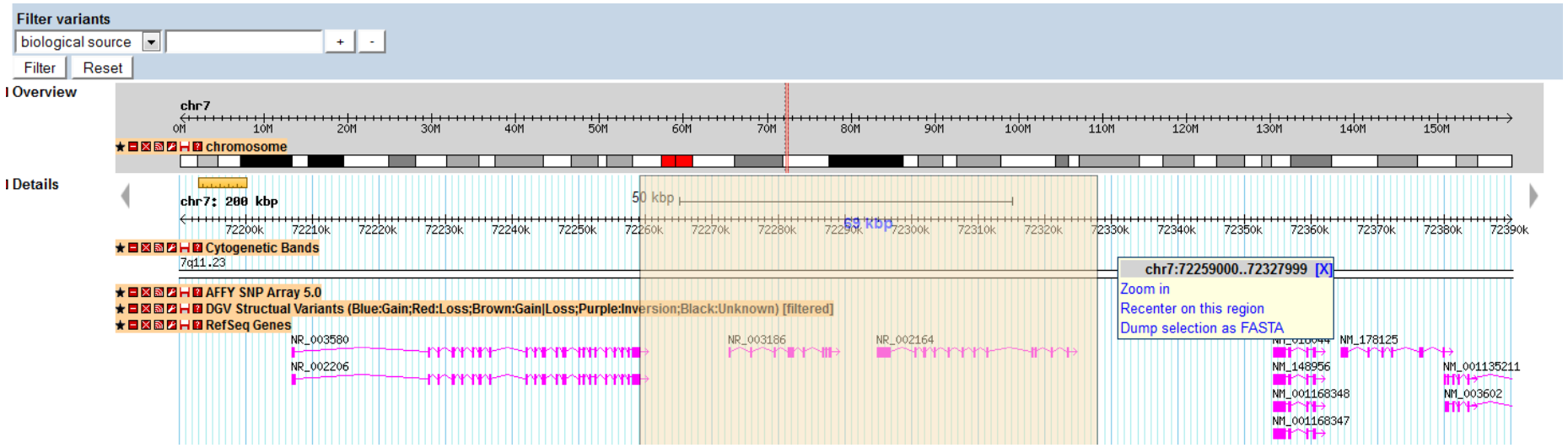
1. To turn on/off various tracks and to modify the display, chose the “Select Tracks” tab.
2. To upload your own custom annotations, or to modify the display, chose the appropriate tabs.
3. To filter the data displayed in the browser, to only show a subset of the variants, use the “Filter Variants” option, select the data type you want to filter by, and add a keyword to the box. To filter by greater than one data type, use the “+” key to add another row.

Genome Browser Options

1. Click and drag on the chromosome to select a particular region

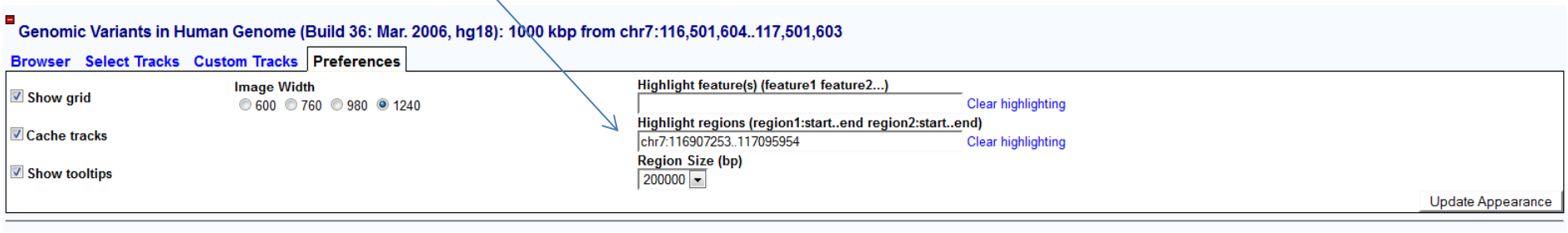


2. Click and drag on the genomic position bar to zoom in or centre on a region of interest.

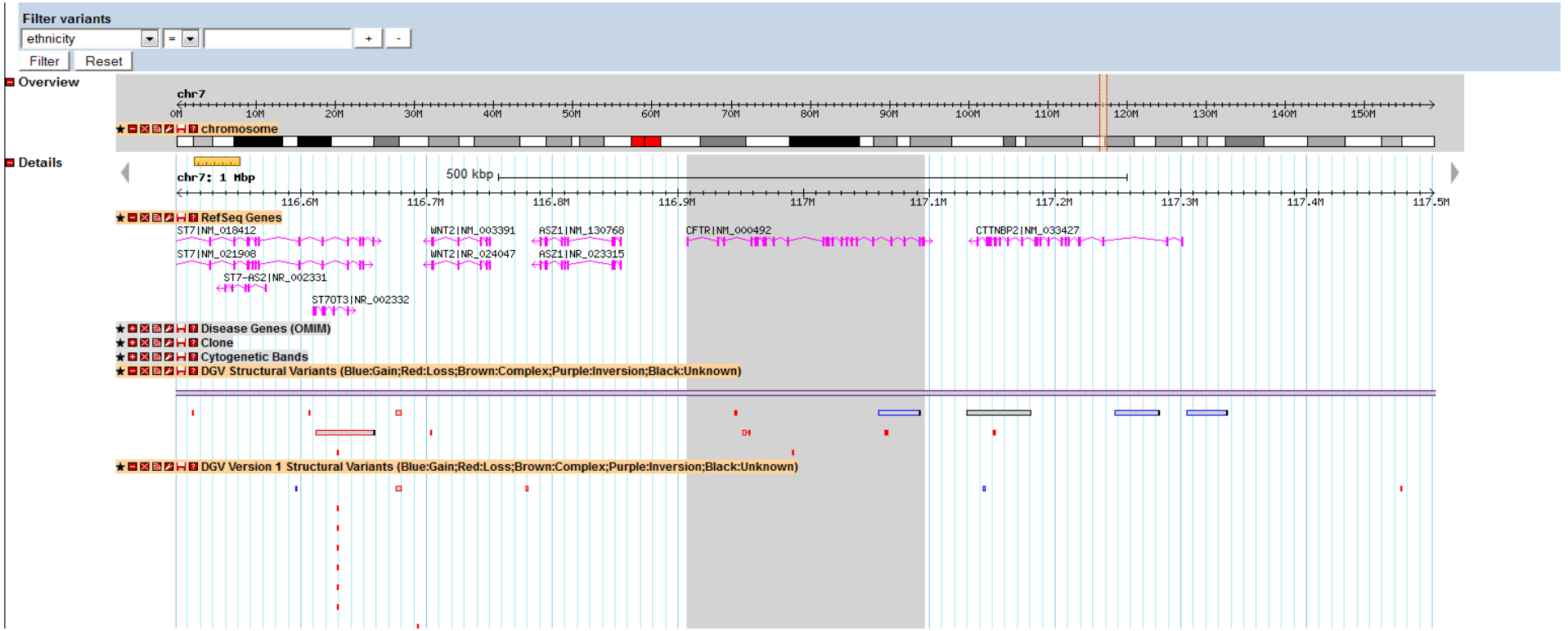


Genome Browser Options

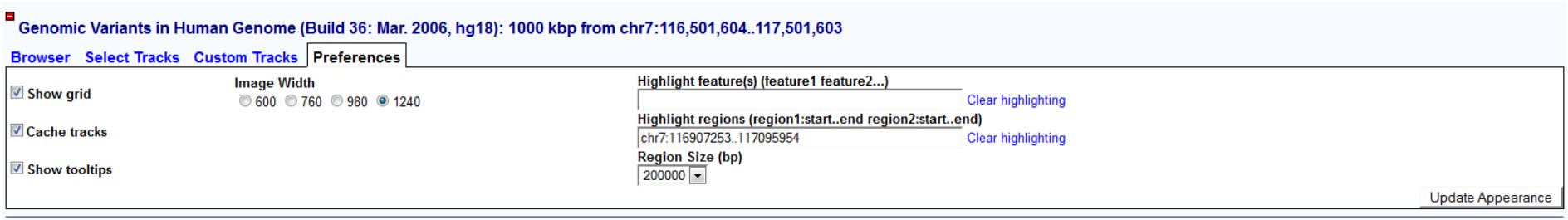
To highlight a region of interest in the browser, select the Preferences tab, and add the target region into this box.



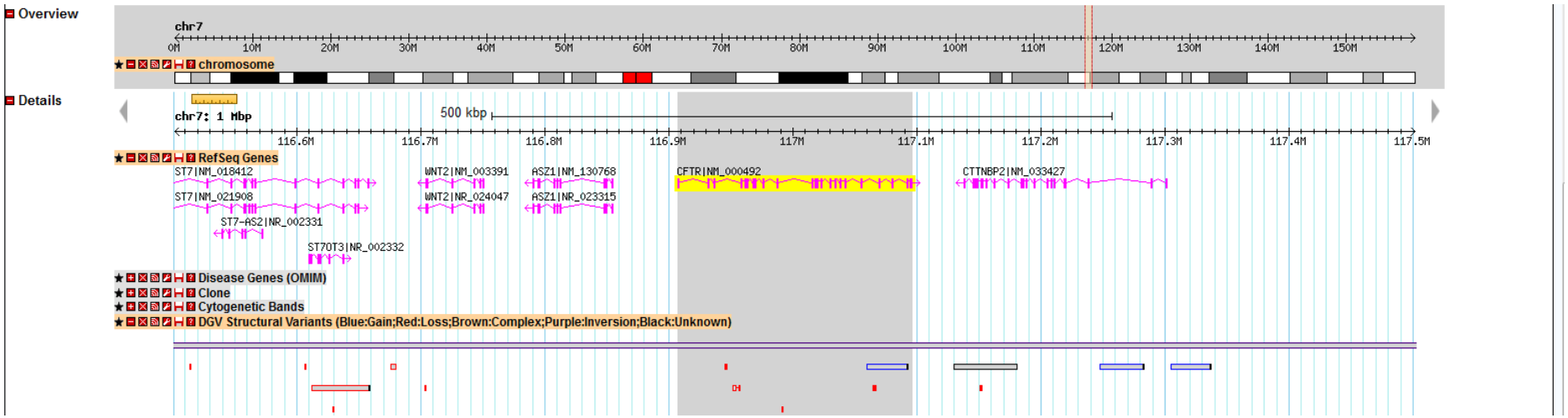
The region will appear as a shaded grey box as seen below.



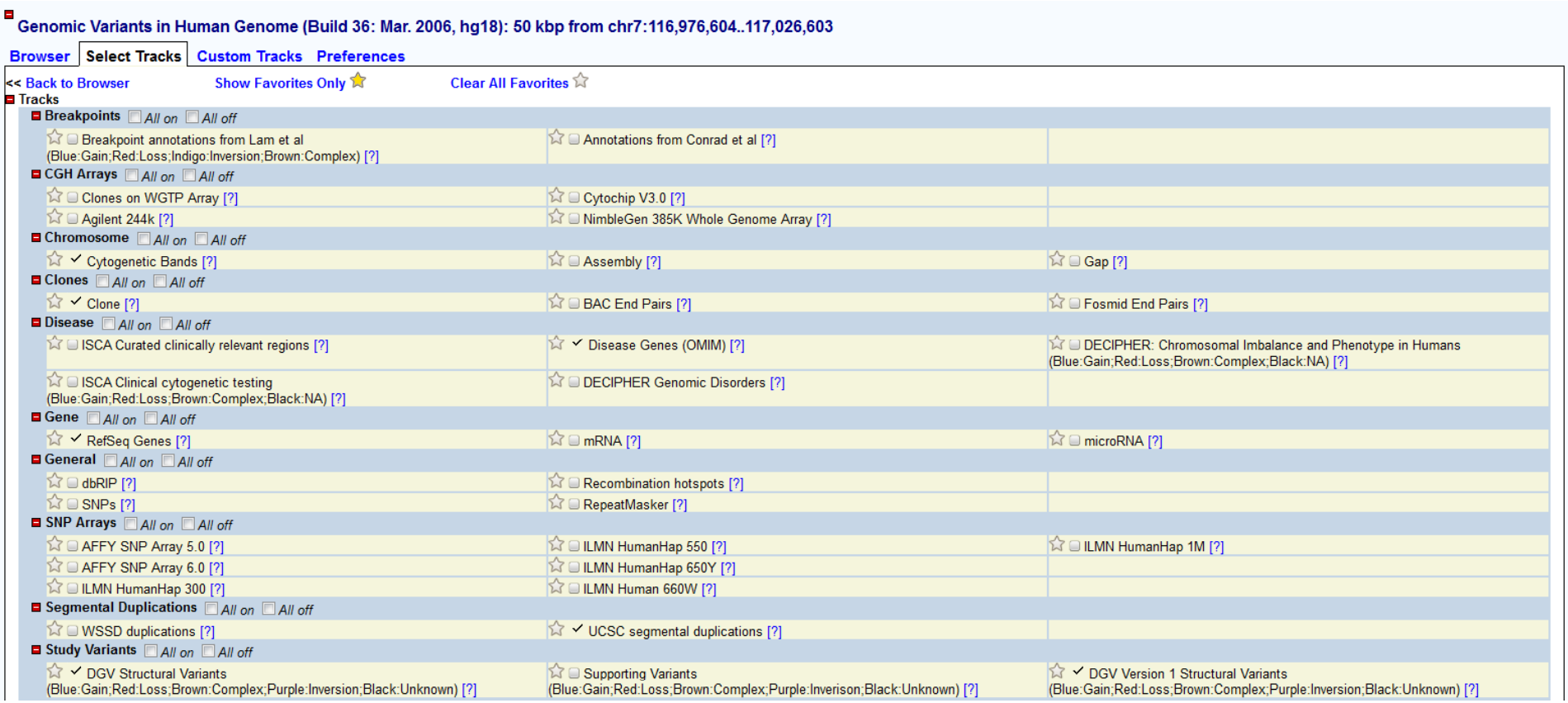
Genome Browser Options



In addition to adding a shaded area to the genome browser, the Preferences tab will allow you to modify the display on your computer by selecting the desired Image Width. You can Also turn the grid pattern on or off, and you may also highlight specific entries in the genome browser that you may be interested in. For example, if you input the term CFTR in the Highlight feature(s) box, you'll get the following display.

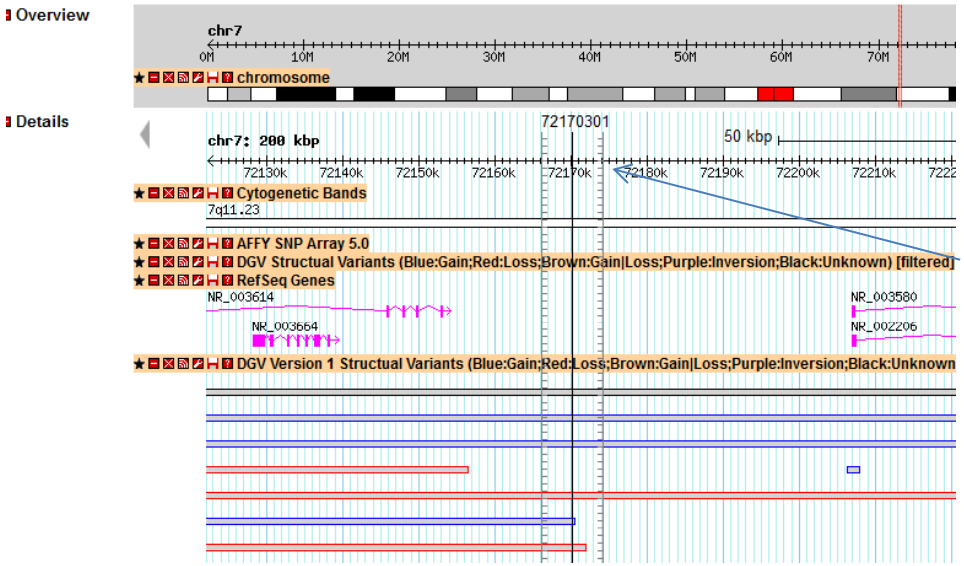


Genome Browser Options

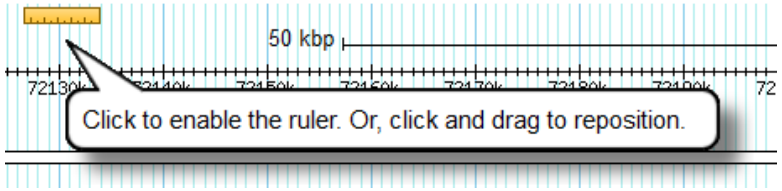


To select the tracks you want shown in the browser, navigate to the “Select Tracks” tab. Check the box beside the track name to turn the track on or off. To turn all the tracks on/off within a group (i.e. Disease), use the “All on” or “All off” options. The “?” beside the tracks will provide information regarding the content of the data displayed in this track.

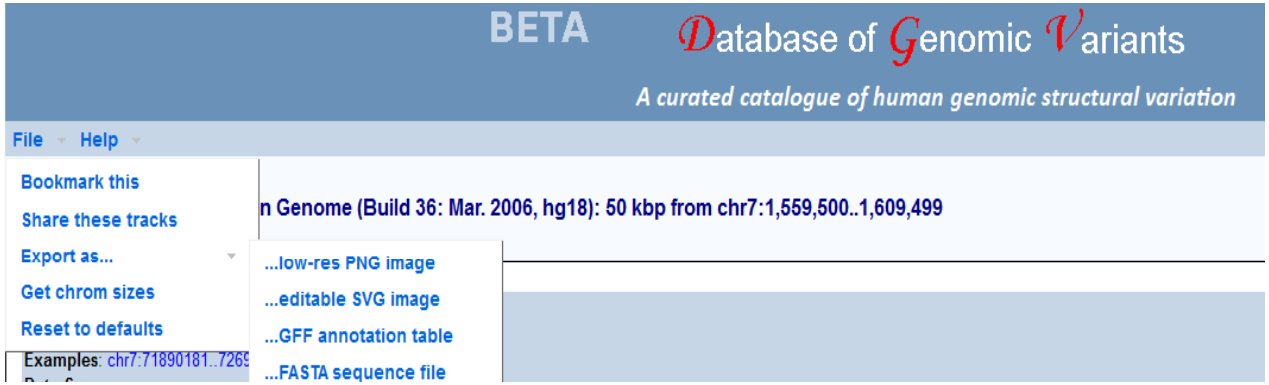
Genome Browser Options



Click on the ruler to add a vertical marker to align features.



File option allows users to export graphics for Figures, and can access GFF tables and FASTA sequences.



Genome Browser Options
Track Options

Minimize the track

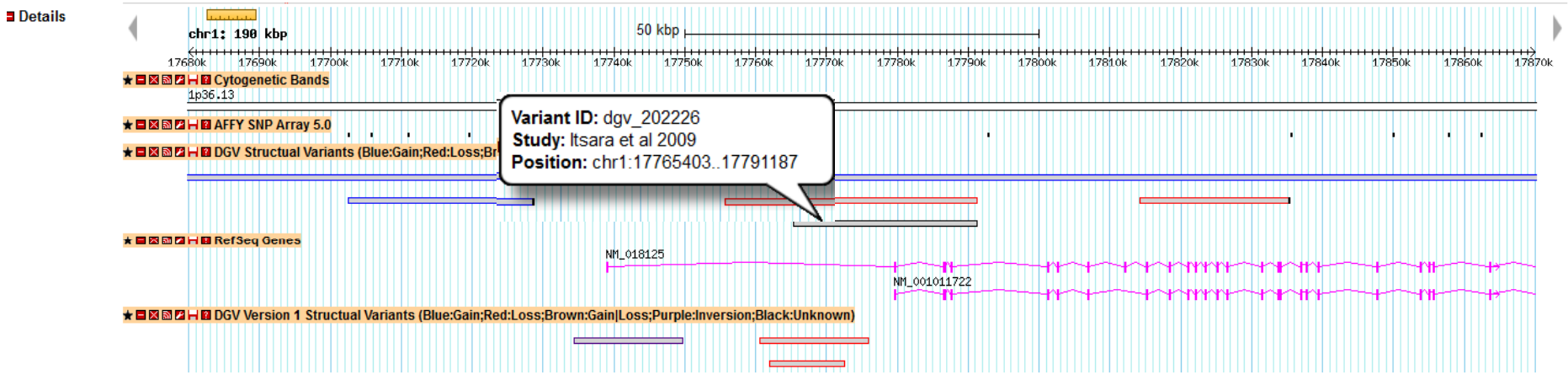
Displays information about the data
In the track.

Close the track.

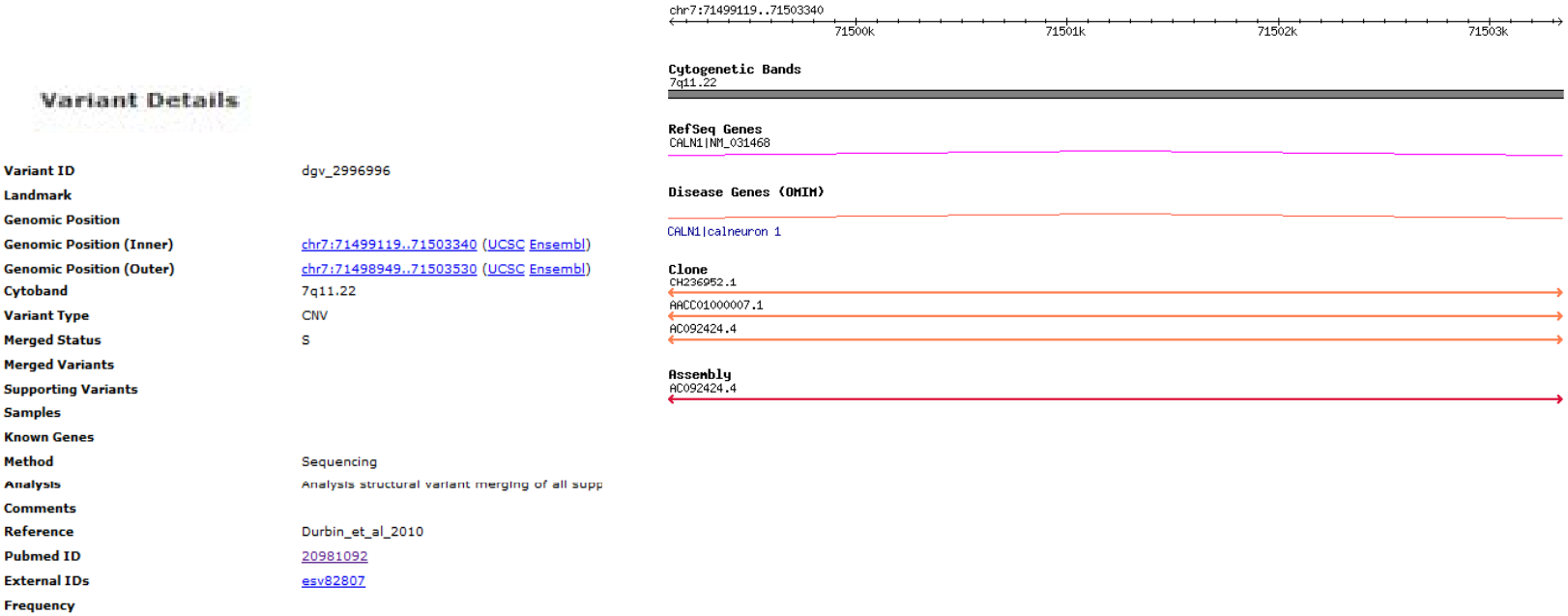
Save the data from this track
This allows the user to download or save the underlying data as FASTA, GFF or GenBank format.
User may select the region displayed in the window, the whole chromosome or genome wide.

Allows users to change how the data are displayed.

DGV data

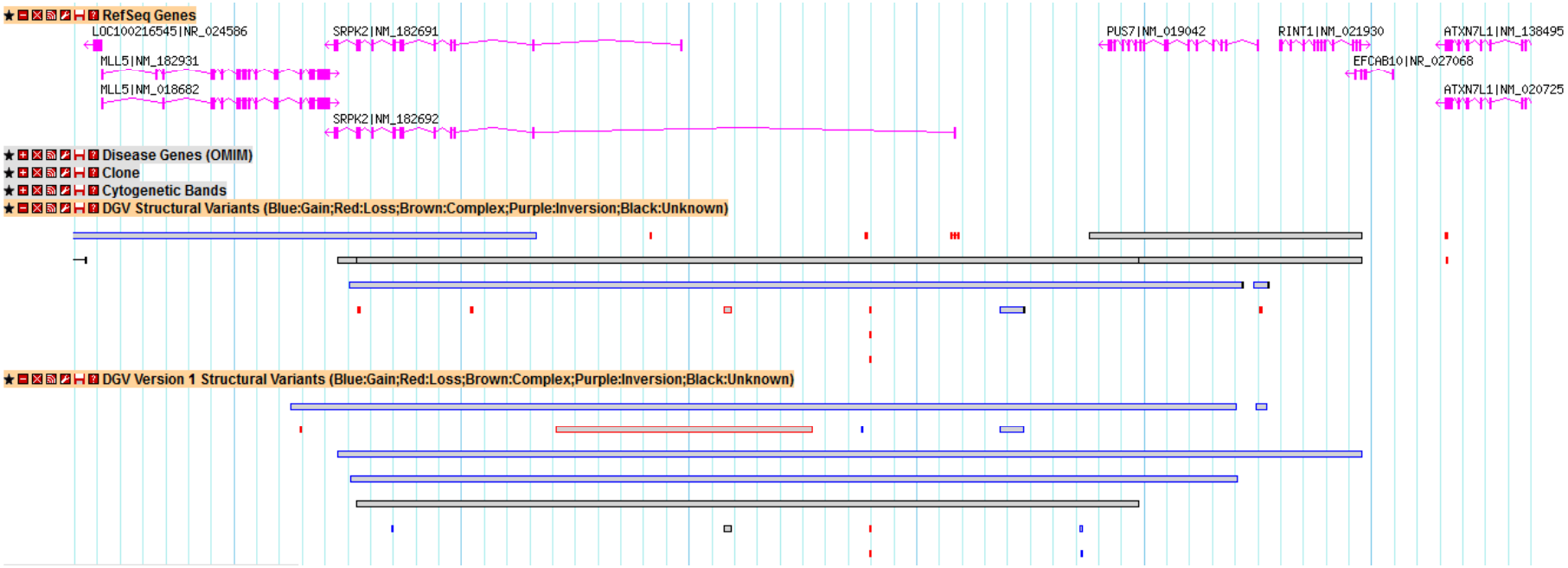


Users can mouse over the variations to obtain summary information, or can click on the image to go to the variants detail page (below).



DGV data

The original (DGV Version 1) and new DGV structural variation data are displayed for comparison.



DGV data

There are a few different types if images used to display structural variants in the genome browser. The different types are shown below with a corresponding description.

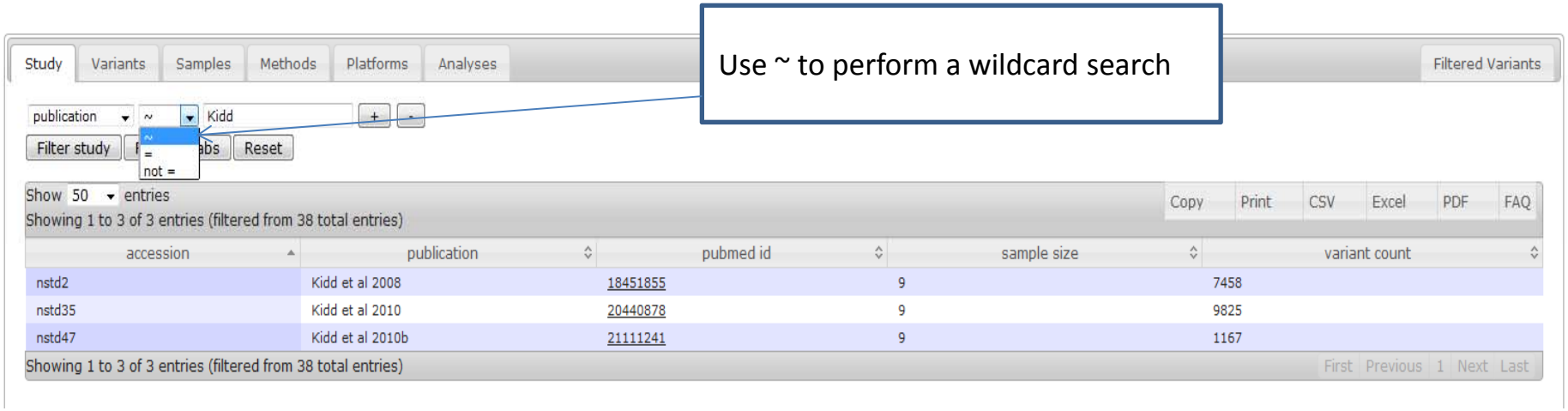
The first type is a solid bar (purple). This variant has a single start, and single end position defined in the study. The second type (red), has been described using a two coordinate System. The bar in the middle represents the inner most start and stop positions of the variant, and the thin bars extending out, represent the outer coordinates reported by the authors. In some studies, where there is breakpoint ambiguity, a confidence interval is established, and the actual breakpoint would likely reside somewhere between the inner and outer coordinates shown. By clicking on the variant, you will be taken to the variant details summary page, and the coordinates used are displayed.

Variant ID	dgv_2993496
Landmark	
Genomic Position	
Genomic Position (Inner)	chr7:116381201..116380950 (UCSC Ensembl)
Genomic Position (Outer)	chr7:116379801..116382650 (UCSC Ensembl)
Cytoband	7q31.2
Variant Type	CNV
Merged Status	S

Variant ID	dgv_380175
Landmark	
Genomic Position	
Genomic Position (Inner)	
Genomic Position (Outer)	chr7:112147851..121027925 (UCSC Ensembl)
Cytoband	7q31.1
Variant Type	OTHER
Merged Status	S

Query Tool

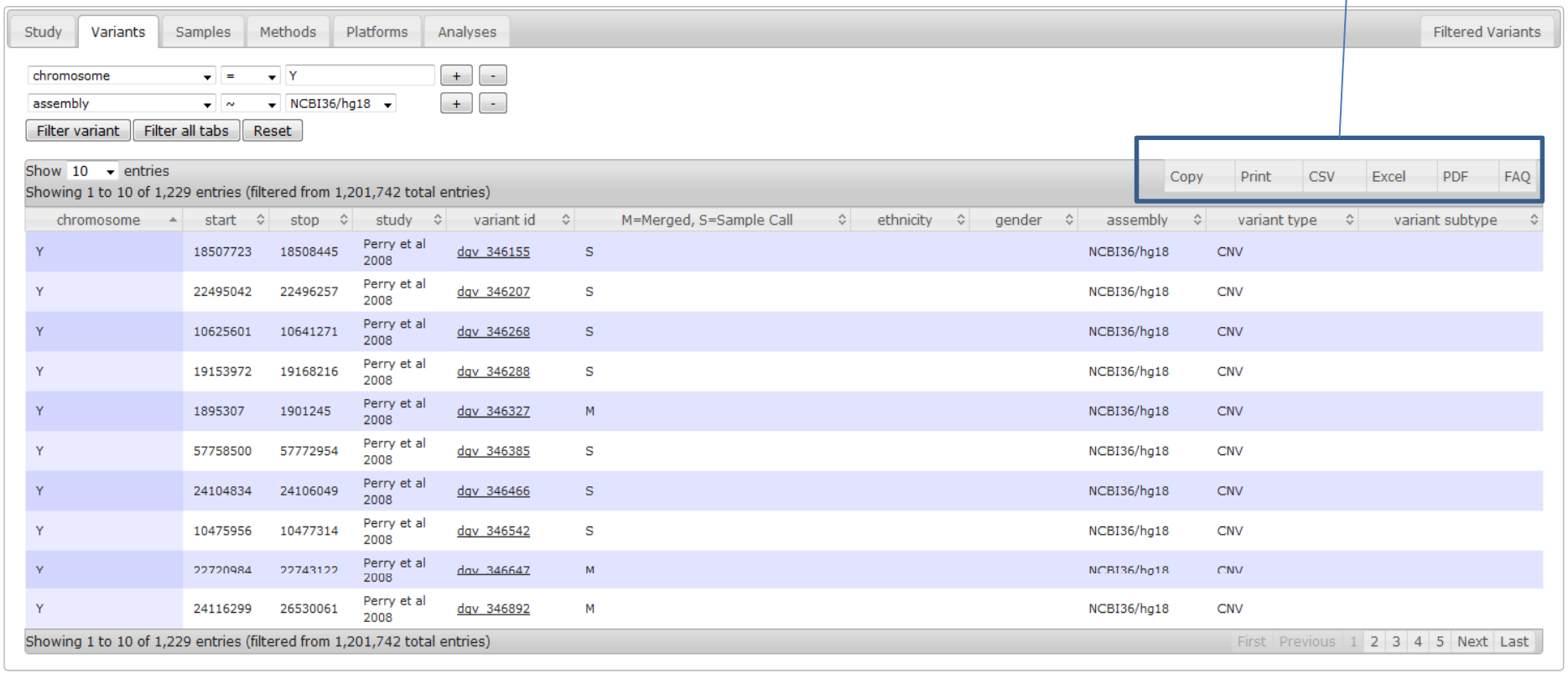
Users can filter data based on the options presented in the tabs above.
On the right side, users can access variants published as part of the original study, but which have been removed from inclusion in DGV for various reasons (Filtered).



Example, where a user has selected to filter the database to obtain studies where the primary author is “Kidd”.

Query Tool

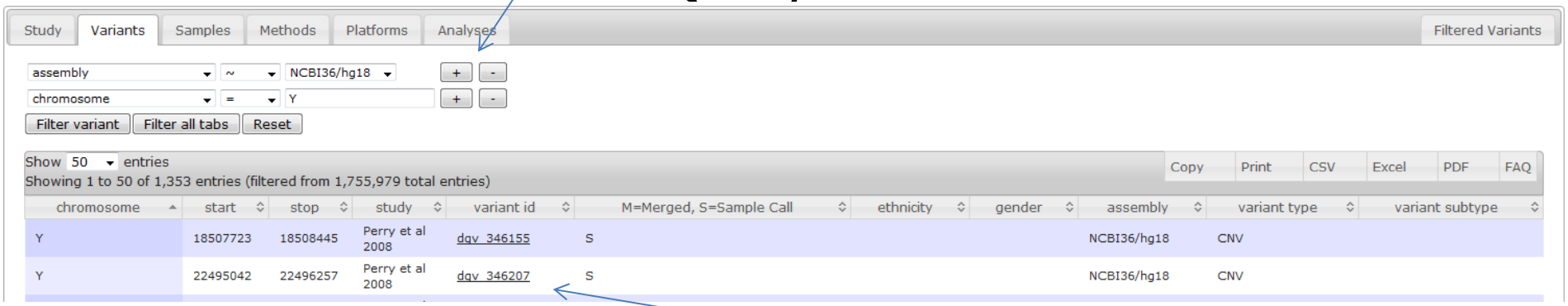
Users can save, copy or print the output using these options.



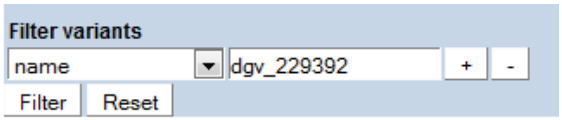
Another example where the user has selected variants on chromosome Y, mapped to assembly version hg18. To filter across all tables, select the “Filter all” button.

Use the “+” button to add additional terms
Use the “-” button to remove search terms

Query Tool



Follow link to variant summary page and genome browser:
-note, any filtering done in the query tool will be carried forward to subsequent displays. Select the reset button to display all data.



It is important to note that when you select the “Filter all” button, this function is applied once to the database. If you perform another filter function, the previous search is reset and a new query is performed.

Interpreting DGV Data
Case Study #1

- Complex region that contains a known disease/disorder locus

- Flanking segmental duplications mediate many short CNVs that only partially overlap deletion/duplication region.

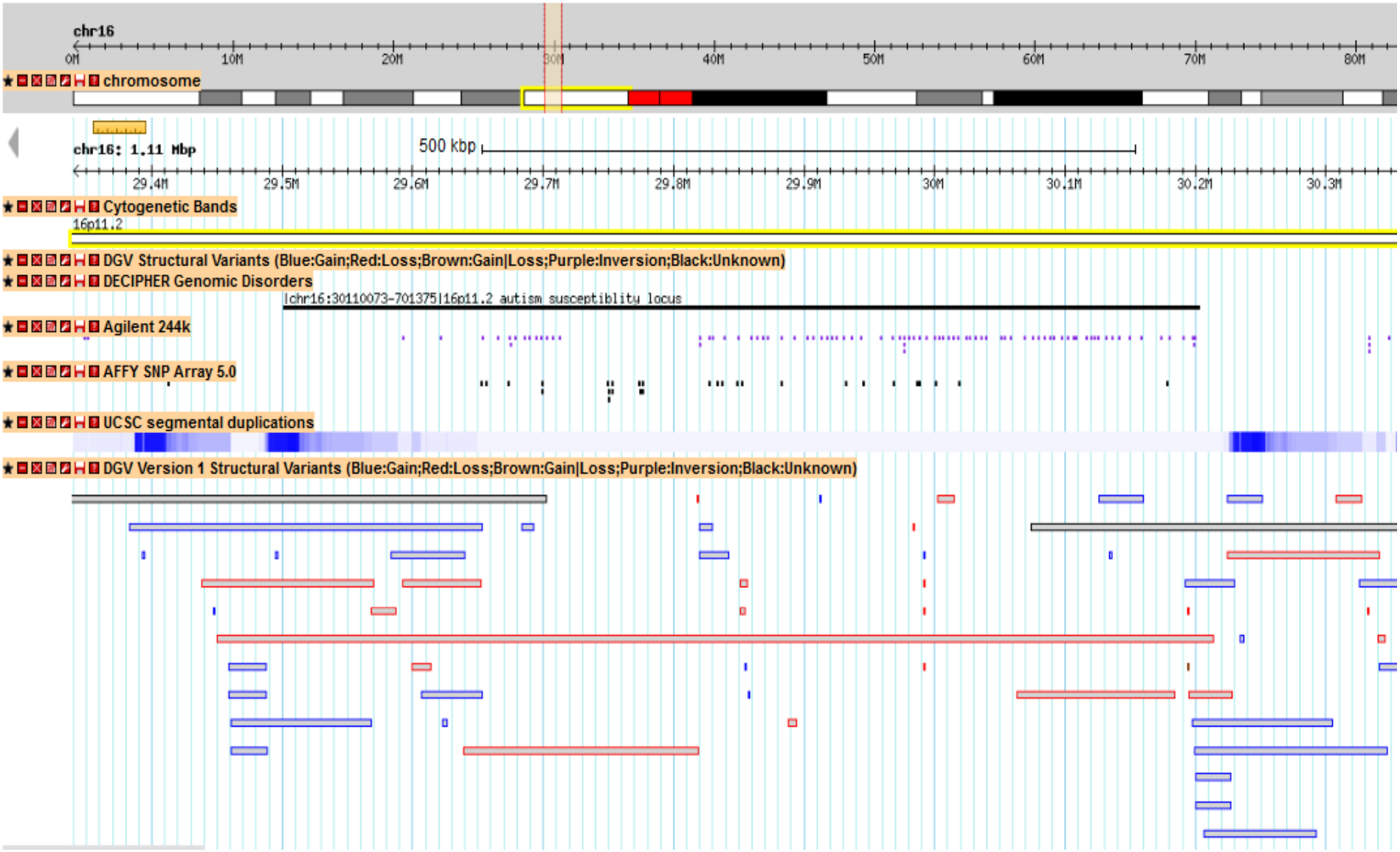
- Very different coverage in some regions on SNP vs. CGH arrays

- Same region occurs, at least rarely, in “control” individuals

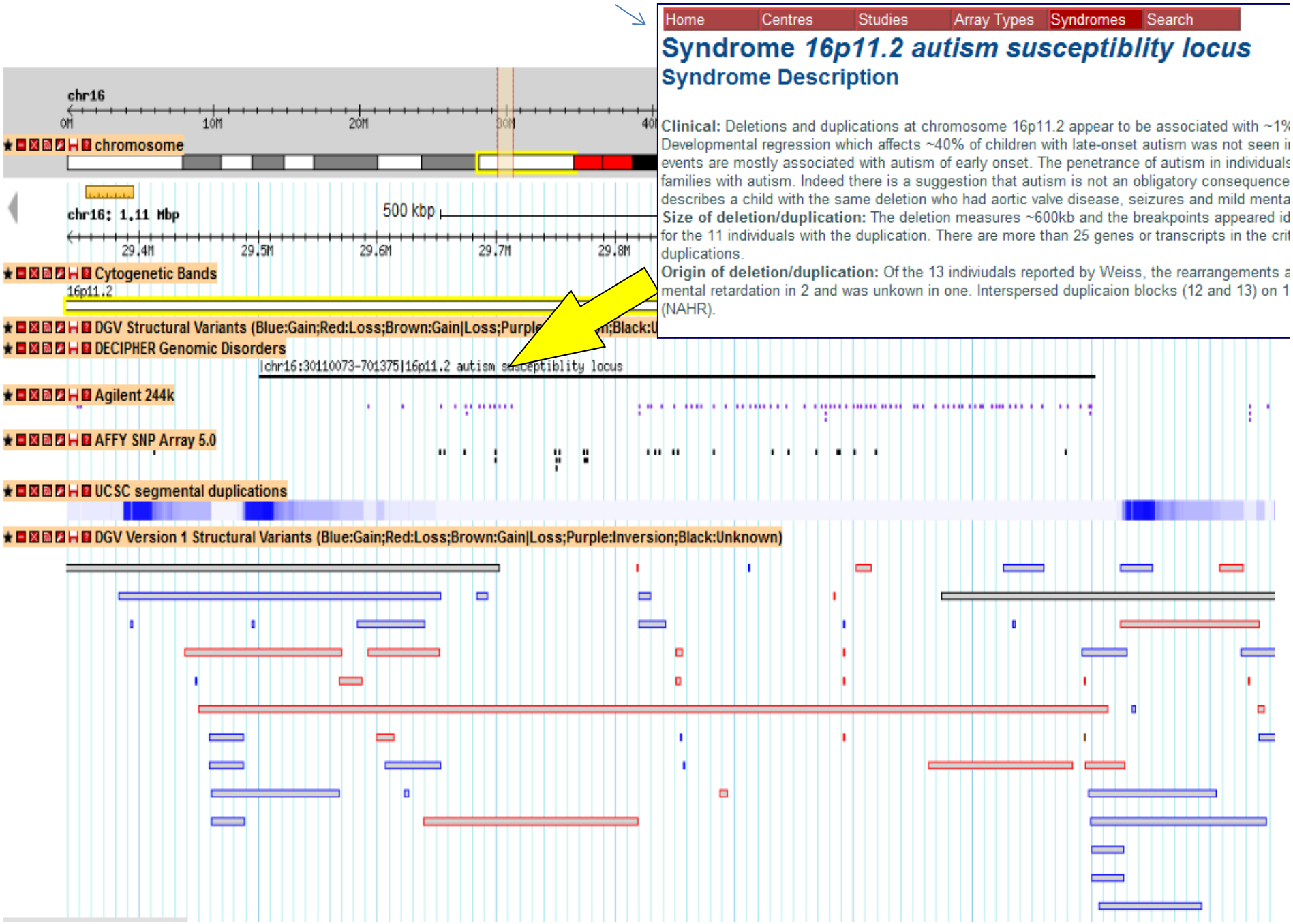
– not surprising, even for rare CNVs that occur in relatively common and potentially under-diagnosed disorders (e.g. autism)

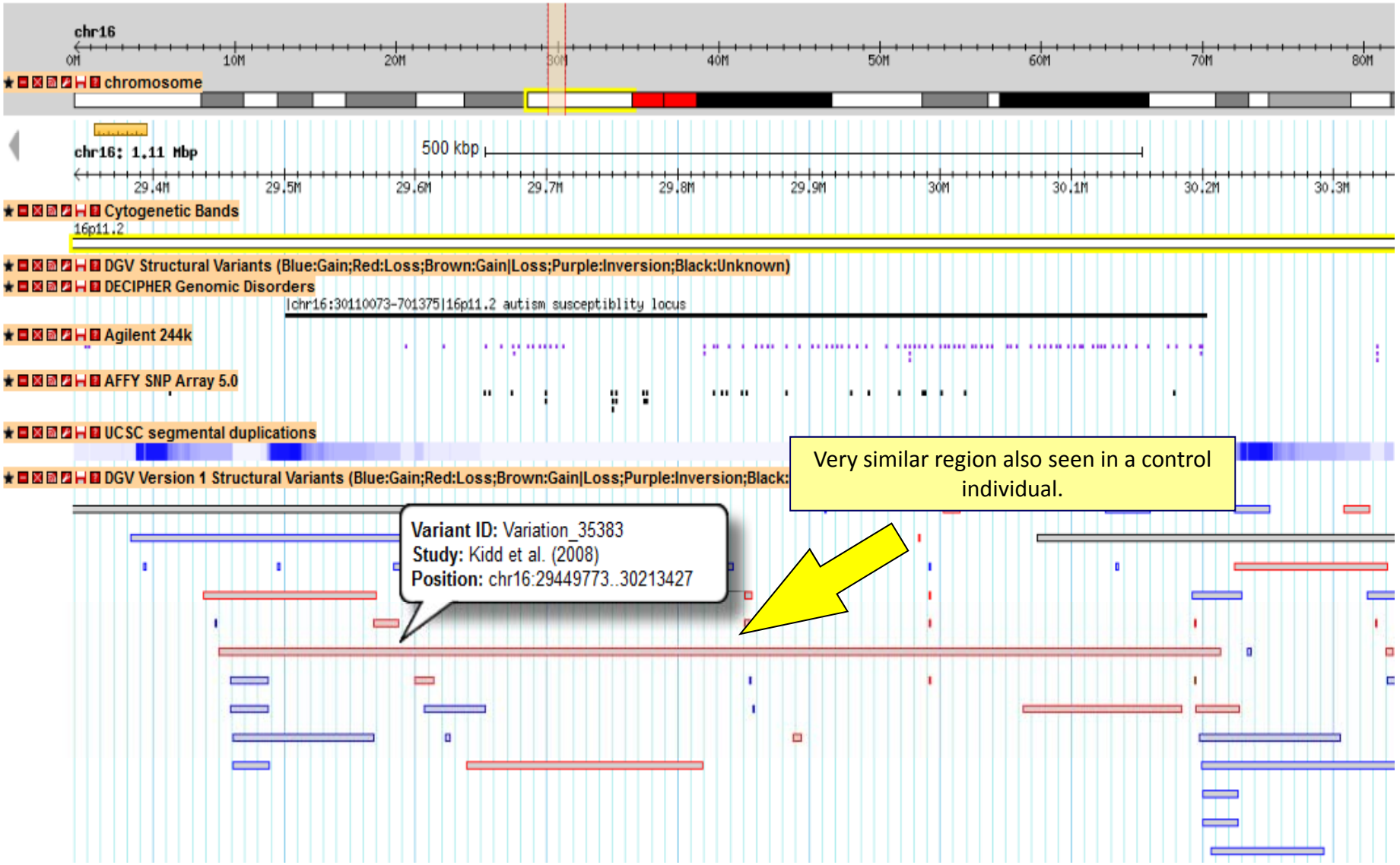
– susceptibility CNVs may act like SNPs; common variants vs. rare variants

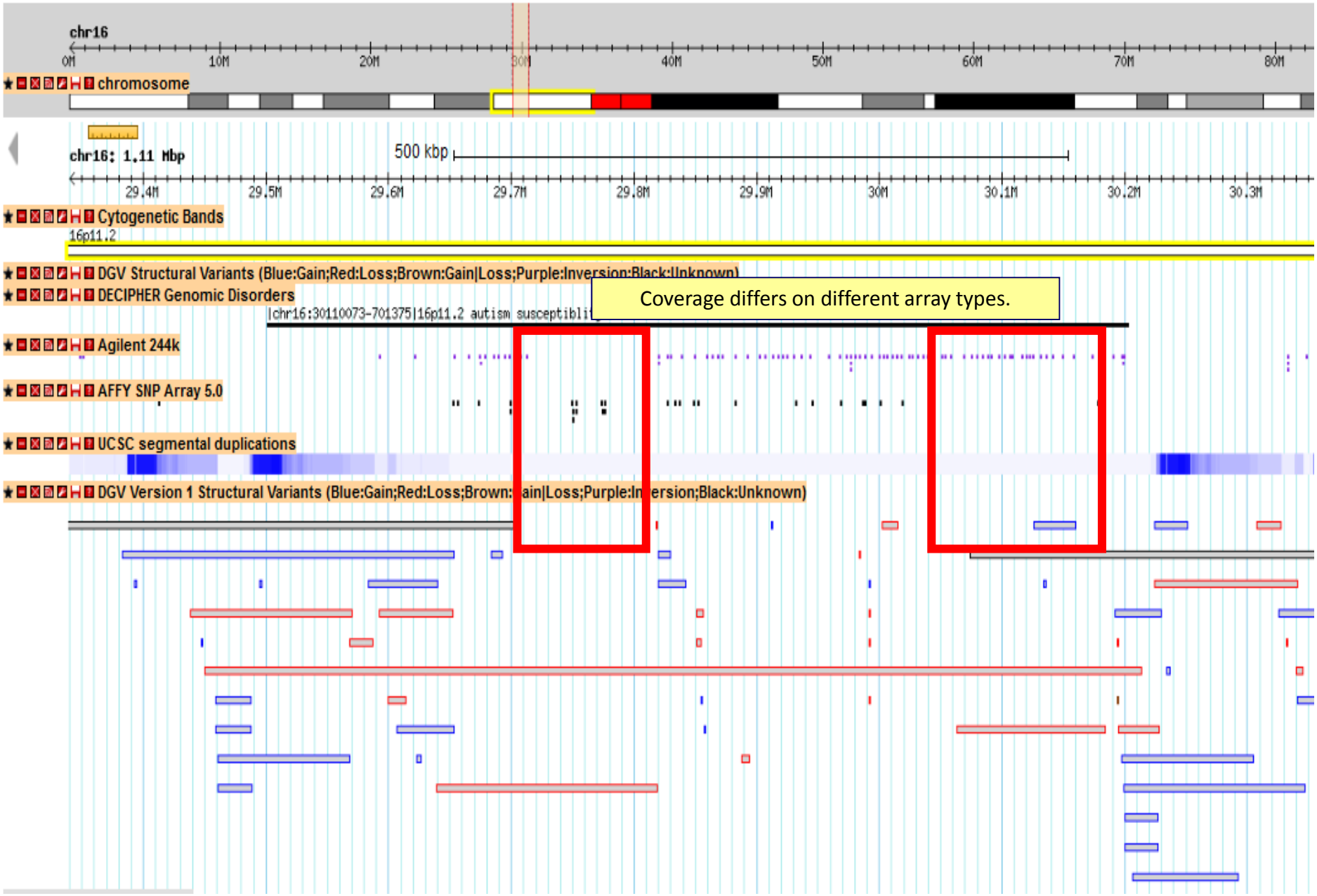
Complex Region 16p11.2



Summary available from link to DECIPHER database




chr16
0M
10M
20M
30M
40M
50M
60M
70M
80M
chromosome
chr16: 1.11 Mbp
500 kbp
29.4M
29.5M
29.6M
29.7M
29.8M
29.9M
30M
30.1M
30.2M
30.3M
Cytogenetic Bands
16p11.2
DGV Structural Variants (Blue:Gain;Red:Loss;Brown:Gain|Loss;Purple:Inversion;Black:Unknown)
DECIPHER Genomic Disorders
|chr16:30110073-701375|16p11.2 autism susceptiblity locus
Agilent 244k
AFFY SNP Array 5.0
UCSC segmental duplications
DGV Version 1 Structural Variants (Blue:Gain;Red:Loss;Brown:Gain|Loss;Purple:Inversion;Black:
Variant ID: Variation_35383
Study: Kidd et al. (2008)
Position: chr16:29449773..30213427
Very similar region also seen in a control individual.


chr16
0M
10M
20M
30M
40M
50M
60M
70M
80M
chromosome
chr16: 1.11 Mbp
500 kbp
29.4M
29.5M
29.6M
29.7M
29.8M
29.9M
30M
30.1M
30.2M
30.3M
Cytogenetic Bands
16p11.2
DGV Structural Variants (Blue:Gain;Red:Loss;Brown:Gain|Loss;Purple:Inversion;Black:Unknown)
DECIPHER Genomic Disorders
|chr16:30110073-701375|16p11.2 autism susceptibli
Coverage differs on different array types.
Agilent 244k
AFFY SNP Array 5.0
UCSC segmental duplications
DGV Version 1 Structural Variants (Blue:Gain;Red:Loss;Brown:
ain|Loss;Purple:In
ersion;Black:Unknown)

DGV Inter-Operates With Other Genome Databases

- **Other sites displaying data from DGV:**

DECIPHER

Ensembl

UCSC

HapMap

GeneCards

A Few Related Resources

- DGV questions and support: – **dgv-contact@sickkids.ca**
- DGV Mailing List: Fill out form on the “Contacts” page.

- DGVa: http://www.ebi.ac.uk/dgva

- dbVar: http://www.ncbi.nlm.nih.gov/dbvar/

- GMOD / GBROWSE – a comprehensive Wiki describing the underlying tools
- **http://gmod.org/wiki/GBrowse**

Help and Support

http://dgvbeta.tcag.ca/dgv/app/contacts

If you have any questions while using the DGV beta site, or if you notice any errors or bugs in the database, please contact the DGV Team at your earliest convenience.

Email: dgv-contact@sickkids.ca.

If you would like to receive updates and notifications about DGV, please sign up for our newsletter

Thank you for taking the time to test and use the database, your assistance and feedback is greatly appreciated.

Sincerely,
The DGV Team