

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>

BETA

Database of Genomic Variants
A curated catalogue of human genomic structural variation

BETA

[About the Project](#) [Downloads](#) [Links](#) [Statistics](#) [FAQ](#)
[Genome Browser](#) [Query Tool](#) [Submissions](#) [Contact Us](#) [Tutorial](#)

Keyword, Landmark or Region 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

BETA *Database of Genomic Variants* **BETA**
A curated catalogue of human genomic structural variation

[About the Project](#) [Downloads](#) [Links](#) [Statistics](#) [FAQ](#)
[Genome Browser](#) [Query Tool](#) [Submissions](#) [Contact Us](#) [Tutorial](#)

Keyword, Landmark or Region 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

Search using the Query Tool

Search by landmark or genomic feature

[News: October 19, 2011 The new DGV website has been launched.](#)



Genome Browser

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

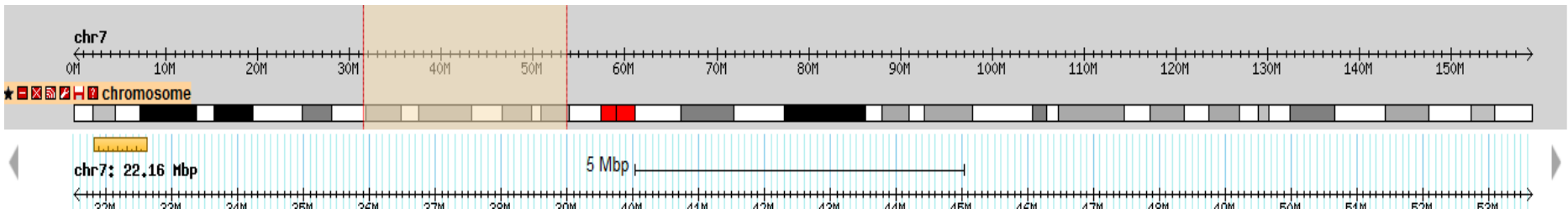
The screenshot shows the top of the DGV genome browser. The header is blue with 'BETA' on the left and right, and 'Database of Genomic Variants' in the center. Below the header is a navigation bar with 'File' and 'Help' menus. The main content area is titled 'Genomic Variants in Human Genome (Build 36: Mar. 2006, hg18): 640 kbp from chr7:104,430,000..105,069,999'. Below this is a 'Browser' tab with sub-tabs for 'Select Tracks', 'Custom Tracks', and 'Preferences'. A 'Search' section contains a 'Landmark or Region' input field with 'chr7:104,430,000..105,069,999' and a 'Search' button. Below the search field are 'Examples' and a 'Data Source' dropdown menu. To the right of the search section are 'Scroll/Zoom' controls with arrows and a 'Show 640 kbp' dropdown, and a 'Flip' checkbox. At the bottom is a 'Filter variants' section with a dropdown menu set to 'ethnicity', an equals sign dropdown, an input field, and plus/minus buttons. 'Filter' and 'Reset' buttons are at the bottom left.

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

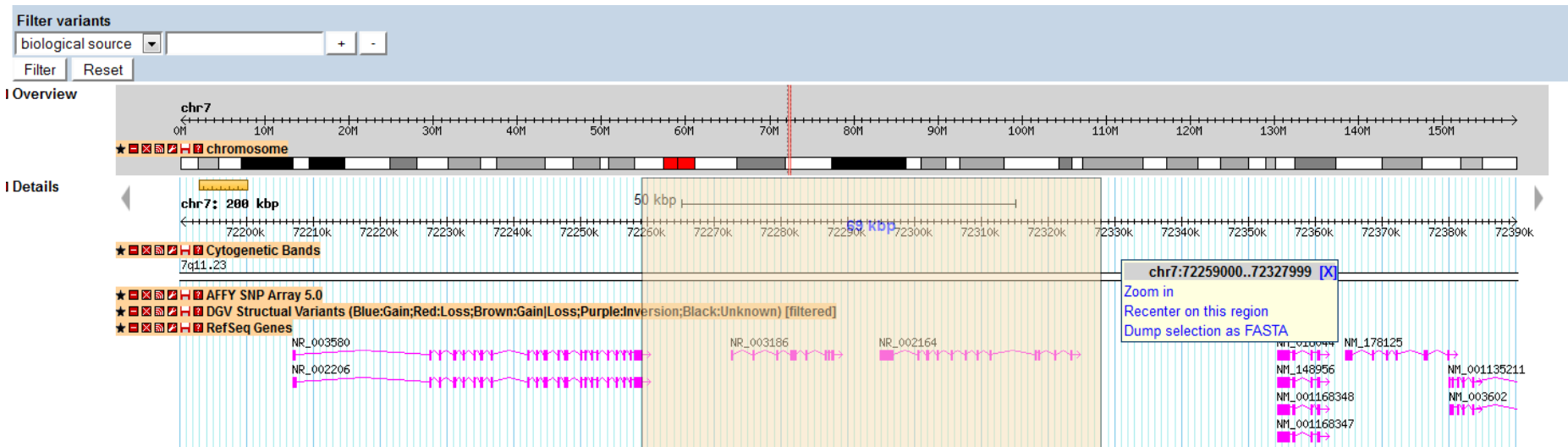
1. To turn on/off various tracks and to modify the display, chose the “Select Tracks” tab.
2. To upload your own custom annotations, or to modify the display, chose the appropriate tabs.
3. To filter the data displayed in the browser, to only show a subset of the variants, use the “Filter Variants” option, 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.

Genomic Variants in Human Genome (Build 36: Mar. 2006, hg18): 1000 kbp from chr7:116,501,604..117,501,603

Browser Select Tracks Custom Tracks Preferences

Show grid Image Width: 600 760 980 1240

Cache tracks

Show tooltips

Highlight feature(s) (feature1 feature2...) [Clear highlighting](#)

Highlight regions (region1:start..end region2:start..end)
chr7:116907253..117095954 [Clear highlighting](#)

Region Size (bp)
200000

[Update Appearance](#)

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

Filter variants: ethnicity = [] + -

Filter Reset

Overview

chr7

0M 10M 20M 30M 40M 50M 60M 70M 80M 90M 100M 110M 120M 130M 140M 150M

chromosome

Details

chr7: 1 Mbp

500 kbp

116.6M 116.7M 116.8M 116.9M 117M 117.1M 117.2M 117.3M 117.4M 117.5M

RefSeq Genes

- ST7|NM_018412
- ST7|NM_021908
- ST7-RS2|NR_002331
- ST70T3|NR_002332
- MNT2|NM_003391
- MNT2|NR_024047
- ASZ1|NM_130768
- ASZ1|NR_023315
- CFTR|NM_000492
- CTTNBP2|NM_033427

Disease Genes (OMIM)

Clone

Cytogenetic Bands

DGV Structural Variants (Blue:Gain;Red:Loss;Brown:Complex;Purple:Inversion;Black:Unknown)

DGV Version 1 Structural Variants (Blue:Gain;Red:Loss;Brown:Complex;Purple:Inversion;Black:Unknown)

Genome Browser Options

Genomic Variants in Human Genome (Build 36: Mar. 2006, hg18): 1000 kbp from chr7:116,501,604..117,501,603

Browser Select Tracks Custom Tracks Preferences

Show grid Image Width
 600 760 980 1240

Cache tracks

Show tooltips

Highlight feature(s) (feature1 feature2...) [Clear highlighting](#)

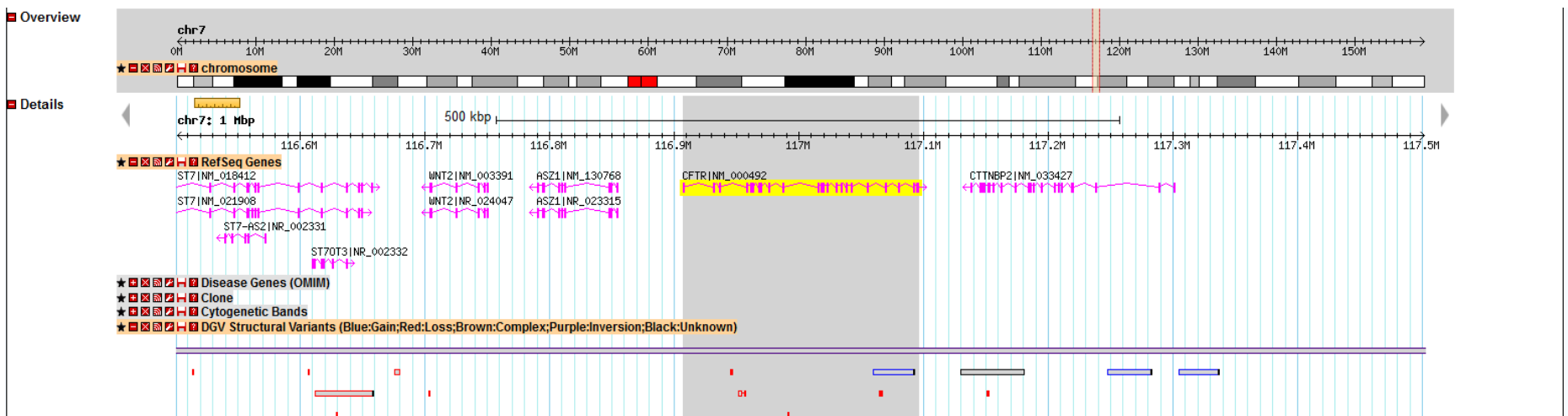
Highlight regions (region1:start..end region2:start..end) [Clear highlighting](#)

chr7:116907253..117095954

Region Size (bp)
200000

[Update Appearance](#)

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

Genomic Variants in Human Genome (Build 36: Mar. 2006, hg18): 50 kbp from chr7:116,976,604..117,026,603

Browser | **Select Tracks** | Custom Tracks | Preferences

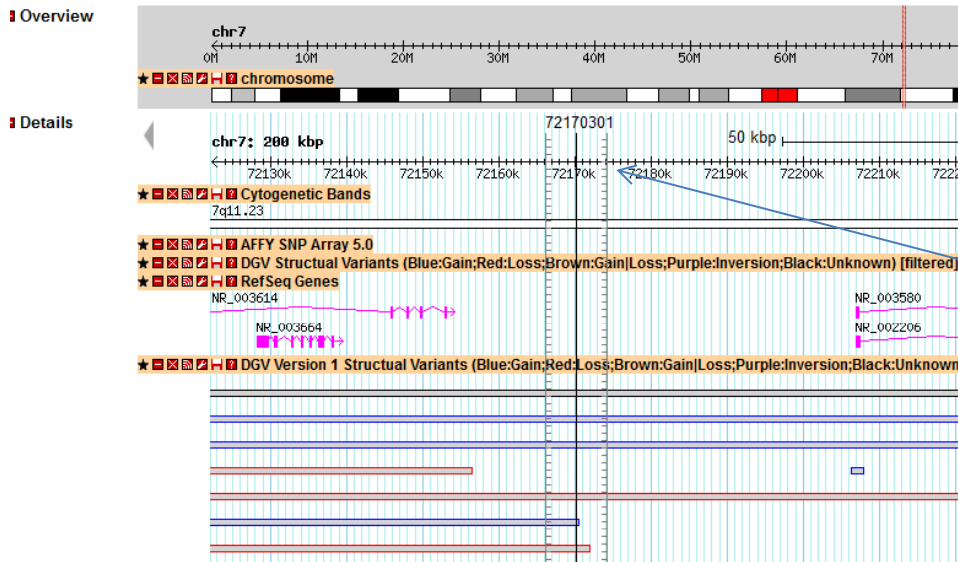
<< Back to Browser | Show Favorites Only ★ | Clear All Favorites ☆

Tracks

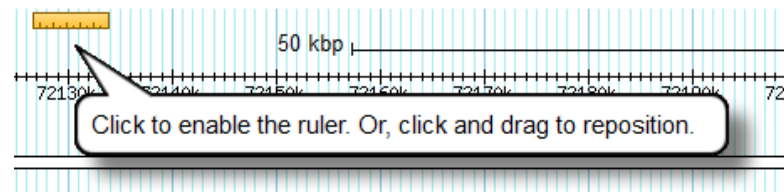
Breakpoints <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input type="checkbox"/> Breakpoint annotations from Lam et al (Blue:Gain;Red:Loss;Indigo:Inversion;Brown:Complex) [?]	☆ <input type="checkbox"/> Annotations from Conrad et al [?]	
CGH Arrays <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input type="checkbox"/> Clones on WGTP Array [?]	☆ <input type="checkbox"/> Cytochip V3.0 [?]	
☆ <input type="checkbox"/> Agilent 244k [?]	☆ <input type="checkbox"/> NimbleGen 385K Whole Genome Array [?]	
Chromosome <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input checked="" type="checkbox"/> Cytogenetic Bands [?]	☆ <input type="checkbox"/> Assembly [?]	☆ <input type="checkbox"/> Gap [?]
Clones <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input checked="" type="checkbox"/> Clone [?]	☆ <input type="checkbox"/> BAC End Pairs [?]	☆ <input type="checkbox"/> Fosmid End Pairs [?]
Disease <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input type="checkbox"/> ISCA Curated clinically relevant regions [?]	☆ <input checked="" type="checkbox"/> Disease Genes (OMIM) [?]	☆ <input type="checkbox"/> DECIPHER: Chromosomal Imbalance and Phenotype in Humans (Blue:Gain;Red:Loss;Brown:Complex;Black:NA) [?]
☆ <input type="checkbox"/> ISCA Clinical cytogenetic testing (Blue:Gain;Red:Loss;Brown:Complex;Black:NA) [?]	☆ <input type="checkbox"/> DECIPHER Genomic Disorders [?]	
Gene <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input checked="" type="checkbox"/> RefSeq Genes [?]	☆ <input type="checkbox"/> mRNA [?]	☆ <input type="checkbox"/> microRNA [?]
General <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input type="checkbox"/> dbRIP [?]	☆ <input type="checkbox"/> Recombination hotspots [?]	
☆ <input type="checkbox"/> SNPs [?]	☆ <input type="checkbox"/> RepeatMasker [?]	
SNP Arrays <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input type="checkbox"/> AFFY SNP Array 5.0 [?]	☆ <input type="checkbox"/> ILMN HumanHap 550 [?]	☆ <input type="checkbox"/> ILMN HumanHap 1M [?]
☆ <input type="checkbox"/> AFFY SNP Array 6.0 [?]	☆ <input type="checkbox"/> ILMN HumanHap 650Y [?]	
☆ <input type="checkbox"/> ILMN HumanHap 300 [?]	☆ <input type="checkbox"/> ILMN Human 660W [?]	
Segmental Duplications <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input type="checkbox"/> WSSD duplications [?]	☆ <input checked="" type="checkbox"/> UCSC segmental duplications [?]	
Study Variants <input type="checkbox"/> All on <input type="checkbox"/> All off		
☆ <input checked="" type="checkbox"/> DGV Structural Variants (Blue:Gain;Red:Loss;Brown:Complex;Purple:Inversion;Black:Unknown) [?]	☆ <input type="checkbox"/> Supporting Variants (Blue:Gain;Red:Loss;Brown:Complex;Purple:Inversion;Black:Unknown) [?]	☆ <input checked="" type="checkbox"/> DGV Version 1 Structural Variants (Blue:Gain;Red:Loss;Brown:Complex;Purple:Inversion;Black:Unknown) [?]

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.

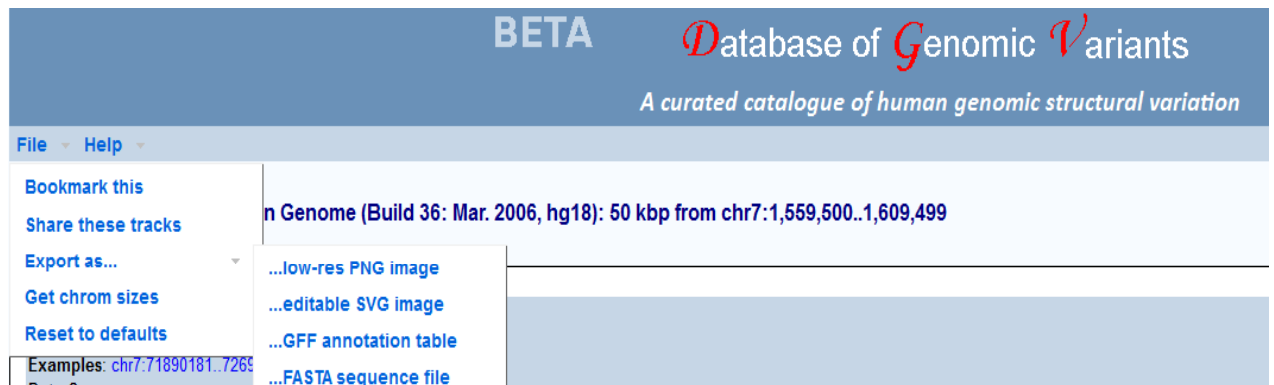
Genome Browser Options



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.



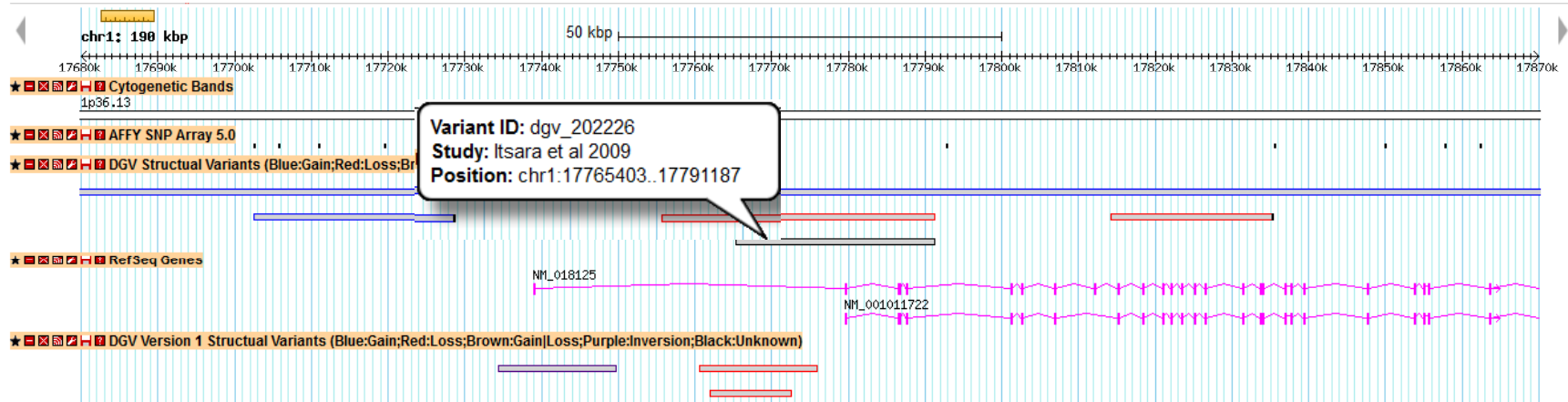
Close the track.

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

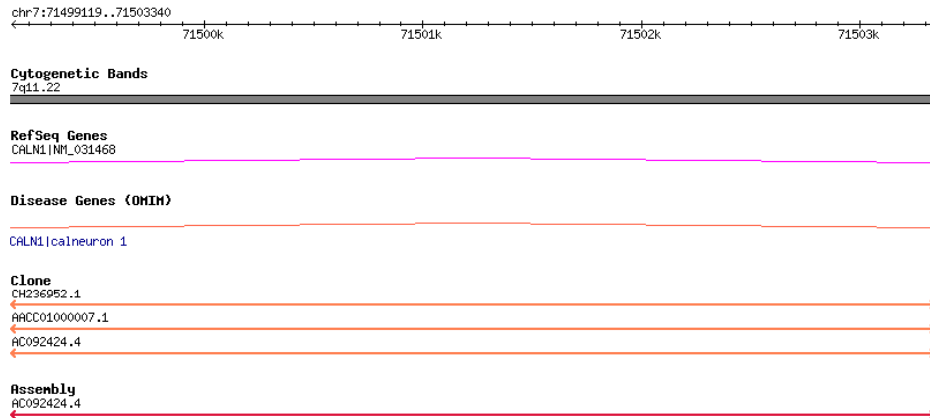
Details



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

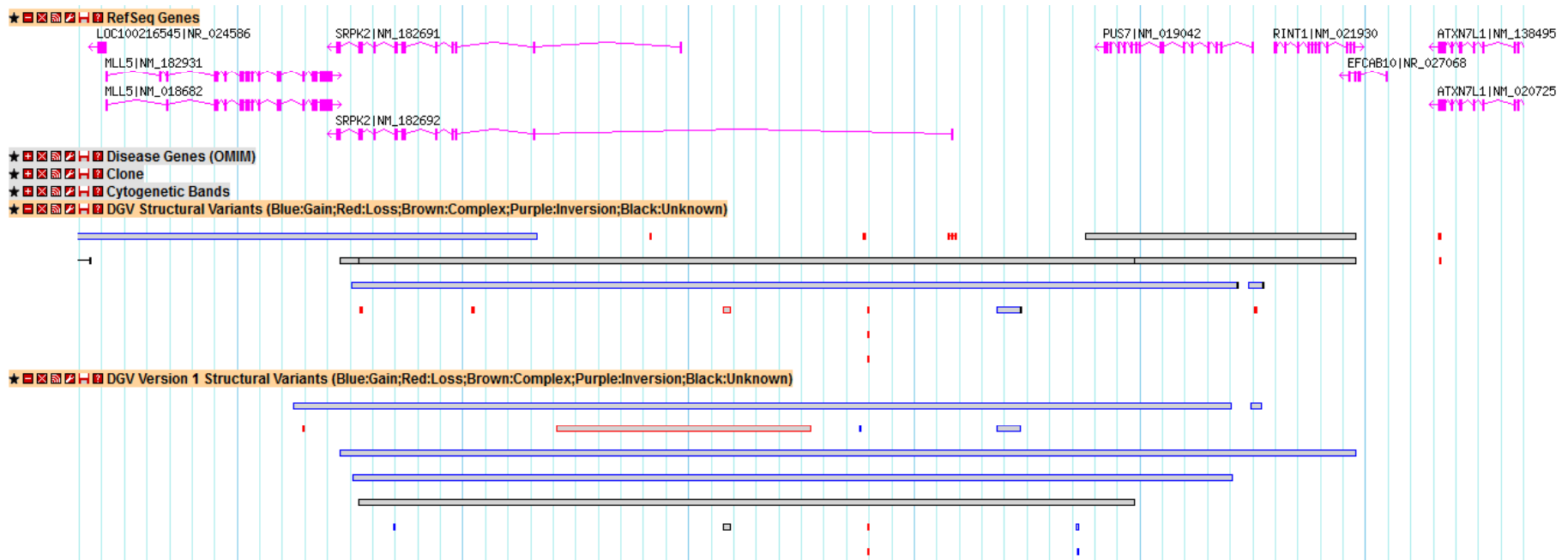
Variant Details

Variant ID	dgv_2996996
Landmark	
Genomic Position	
Genomic Position (Inner)	chr7:71499119..71503340 (UCSC Ensembl)
Genomic Position (Outer)	chr7:71498949..71503530 (UCSC Ensembl)
Cytoband	7q11.22
Variant Type	CNV
Merged Status	S
Merged Variants	
Supporting Variants	
Samples	
Known Genes	
Method	Sequencing
Analysis	Analysis structural variant merging of all supp
Comments	
Reference	Durbin_et_al_2010
Pubmed ID	20981092
External IDs	esv82807
Frequency	



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 of 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).

The screenshot shows the DGV Query Tool interface. A search filter is applied to the "publication" field with the value "Kidd". A callout box points to the search input field with the text "Use ~ to perform a wildcard search". The search results are displayed in a table with columns: accession, publication, pubmed id, sample size, and variant count. The table shows 3 entries filtered from 38 total entries.

accession	publication	pubmed id	sample size	variant count
nstd2	Kidd et al 2008	18451855	9	7458
nstd35	Kidd et al 2010	20440878	9	9825
nstd47	Kidd et al 2010b	21111241	9	1167

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.

The screenshot shows the Query Tool interface with the following components:

- Navigation tabs: Study, Variants, Samples, Methods, Platforms, Analyses. A 'Filtered Variants' indicator is on the right.
- Filter controls: chromosome (dropdown), = (operator), Y (value), assembly (dropdown), ~ (operator), NCBI36/hg18 (value). Buttons for '+', '-' are next to the values.
- Action buttons: Filter variant, Filter all tabs, Reset.
- Table controls: Show 10 entries (dropdown), Showing 1 to 10 of 1,229 entries (filtered from 1,201,742 total entries).
- Table with columns: chromosome, start, stop, study, variant id, M=Merged, S=Sample Call, ethnicity, gender, assembly, variant type, variant subtype.
- Table content (10 rows):

chromosome	start	stop	study	variant id	M=Merged, S=Sample Call	ethnicity	gender	assembly	variant type	variant subtype
Y	18507723	18508445	Perry et al 2008	dgv_346155	S			NCBI36/hg18	CNV	
Y	22495042	22496257	Perry et al 2008	dgv_346207	S			NCBI36/hg18	CNV	
Y	10625601	10641271	Perry et al 2008	dgv_346268	S			NCBI36/hg18	CNV	
Y	19153972	19168216	Perry et al 2008	dgv_346288	S			NCBI36/hg18	CNV	
Y	1895307	1901245	Perry et al 2008	dgv_346327	M			NCBI36/hg18	CNV	
Y	57758500	57772954	Perry et al 2008	dgv_346385	S			NCBI36/hg18	CNV	
Y	24104834	24106049	Perry et al 2008	dgv_346466	S			NCBI36/hg18	CNV	
Y	10475956	10477314	Perry et al 2008	dgv_346542	S			NCBI36/hg18	CNV	
Y	22720984	22743122	Perry et al 2008	dgv_346647	M			NCBI36/hg18	CNV	
Y	24116299	26530061	Perry et al 2008	dgv_346892	M			NCBI36/hg18	CNV	

Showing 1 to 10 of 1,229 entries (filtered from 1,201,742 total entries)

Callout box options: Copy, Print, CSV, Excel, PDF, FAQ.

Page navigation: First, Previous, 1, 2, 3, 4, 5, Next, Last.

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

The screenshot shows the Query Tool interface with the following components:

- Navigation tabs: Study, Variants, Samples, Methods, Platforms, Analyses
- Search filters: assembly (NCBI36/hg18), chromosome (Y)
- Buttons: Filter variant, Filter all tabs, Reset
- Table header: chromosome, start, stop, study, variant id, M=Merged, S=Sample Call, ethnicity, gender, assembly, variant type, variant subtype
- Table body (showing 1 to 50 of 1,353 entries):

chromosome	start	stop	study	variant id	M=Merged, S=Sample Call	ethnicity	gender	assembly	variant type	variant subtype
Y	18507723	18508445	Perry et al 2008	dgv_346155	S			NCBI36/hg18	CNV	
Y	22495042	22496257	Perry et al 2008	dgv_346207	S			NCBI36/hg18	CNV	

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.

★ 🗨️ 📄 📄 📄 📄 📄 📄 DGV Structural Variants (Blue:Gain;Red:Loss;Brown:Gain|Loss;Purple:Inversion;Black:Unknown) [filtered]

Filter variants

name + -

Filter Reset

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



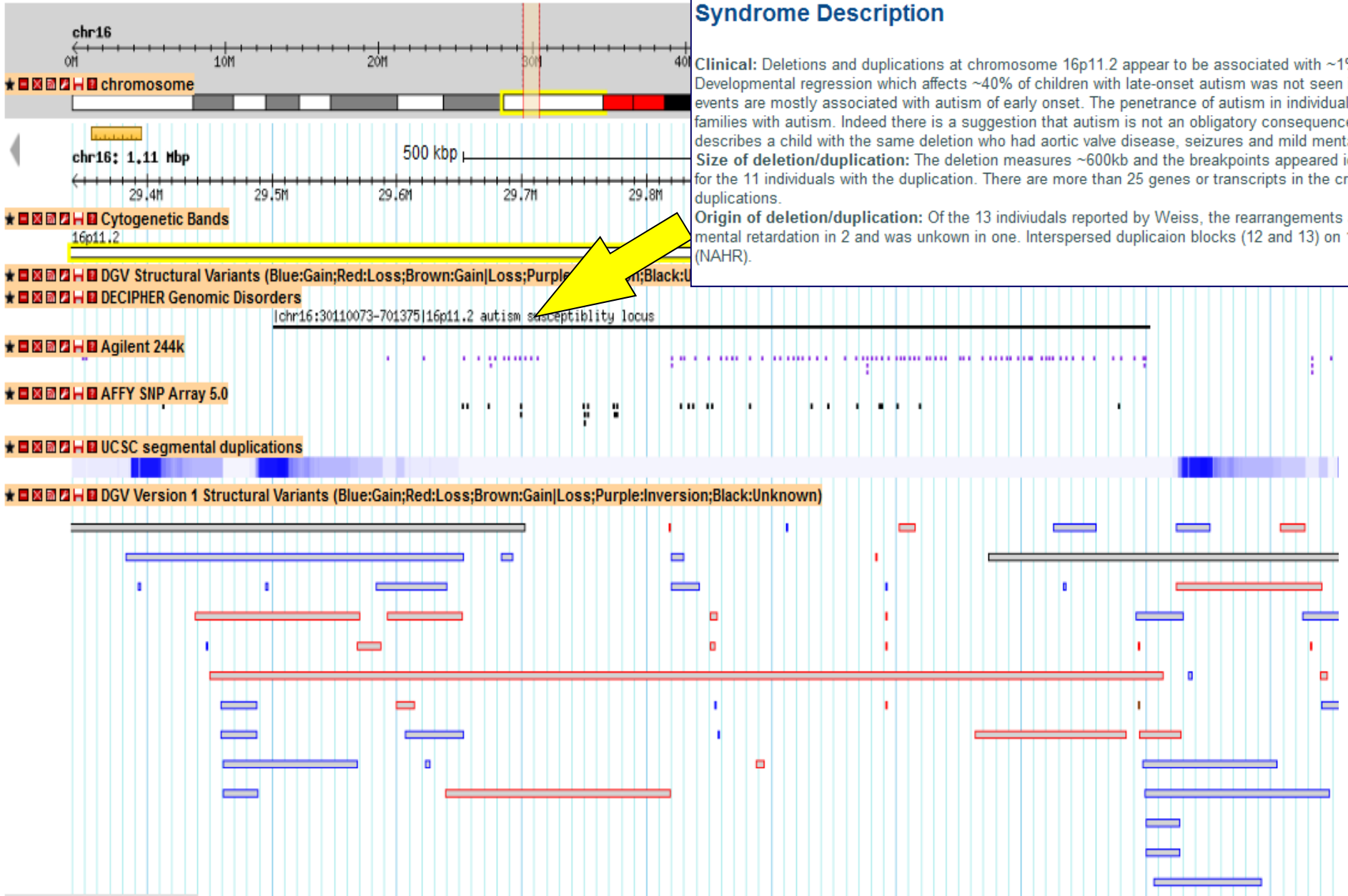
Summary available from link to DECIPHER database

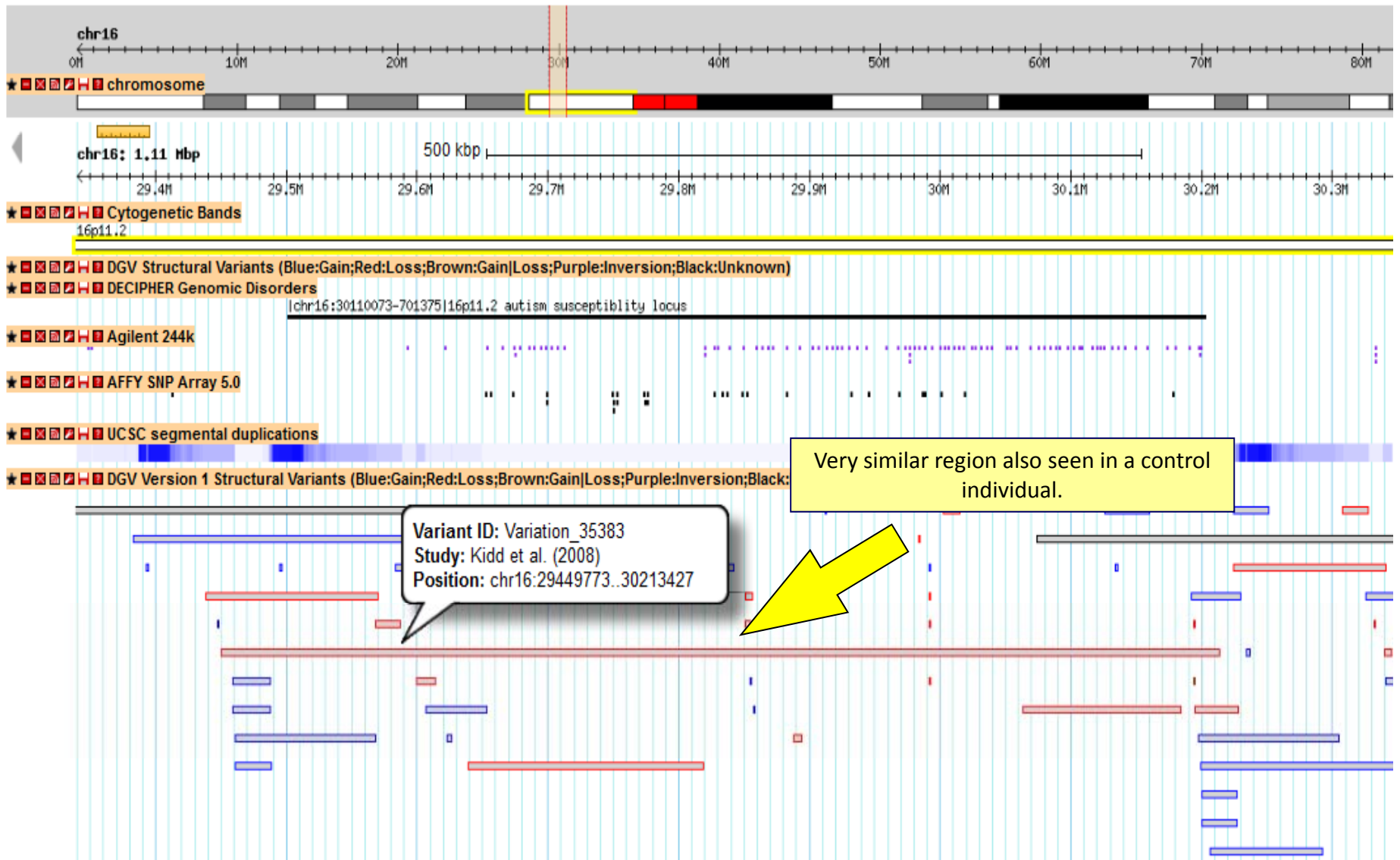
Home Centres Studies Array Types Syndromes Search

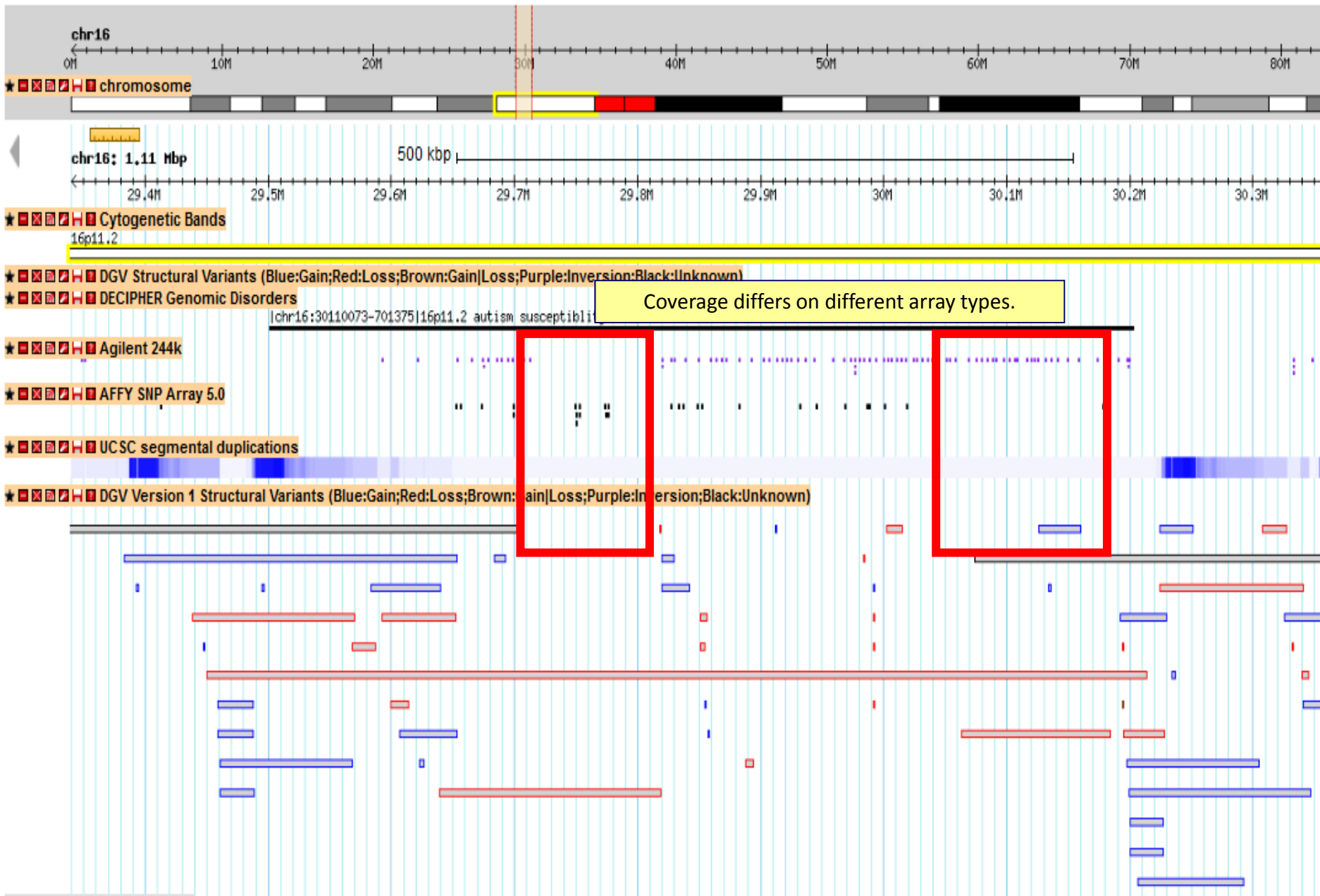
Syndrome 16p11.2 autism susceptibility locus Syndrome Description

Clinical: Deletions and duplications at chromosome 16p11.2 appear to be associated with ~1% Developmental regression which affects ~40% of children with late-onset autism was not seen in events are mostly associated with autism of early onset. The penetrance of autism in individuals families with autism. Indeed there is a suggestion that autism is not an obligatory consequence describes a child with the same deletion who had aortic valve disease, seizures and mild mental retardation. **Size of deletion/duplication:** The deletion measures ~600kb and the breakpoints appeared identical for the 11 individuals with the duplication. There are more than 25 genes or transcripts in the critical region.

Origin of deletion/duplication: Of the 13 individuals reported by Weiss, the rearrangements a mental retardation in 2 and was unknown in one. Interspersed duplication blocks (12 and 13) on 1 (NAHR).







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