Introducing the *NEW* Database of Genomic Variants

Thursday December 8th, 2011

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

The Centre for Applied Genomics (TCAG)
Genome Canada Science & Innovation Centre

1. DNA Sequencing & Synthesis - including NGS (Illumina HiSeq, 5 SOLiD, Roche 454), capillary, IDT portal, modified oligos, HPLC purification
2. Gene Expression & Microarray - Affymetrix, Agilent, Illumina; genotyping, SNP/CNV, expression, tiling
3. Cytogenomics & Genome Resources - karyotyping/SKY, FISH mapping, probe labeling, cDNA and clone libraries
4. Databasing & Biobanking – cell line immortalization and banking, population controls
5. Genetic & Statistical Analysis - project design, statistical genetics, microarray analysis, pathway analysis
6. Informatics – project consultation, sophisticated project design supporting NGS sequencing analysis

More than 1,600 laboratories worldwide have used TCAG services for research in human and animal health, plant/forestry genomics and environment.

Overview

- Introduction to the *New* DGV
- Genome Browser
- Query Tool
- Examples of how to use the data
- Summary/Questions

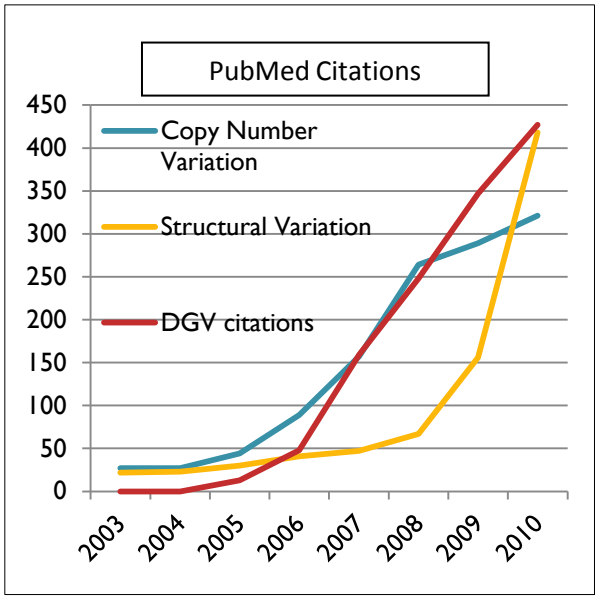
Overview

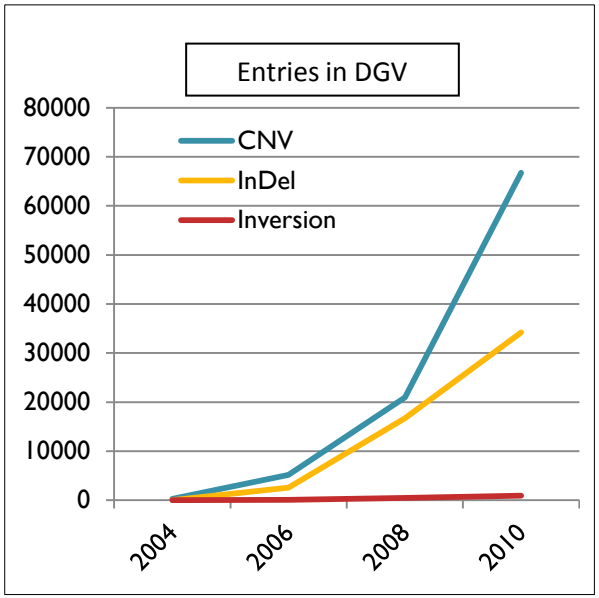
- **Introduction to the *New* DGV**
- Genome Browser
- Query Tool
- Examples of how to use the data
- Summary/Questions

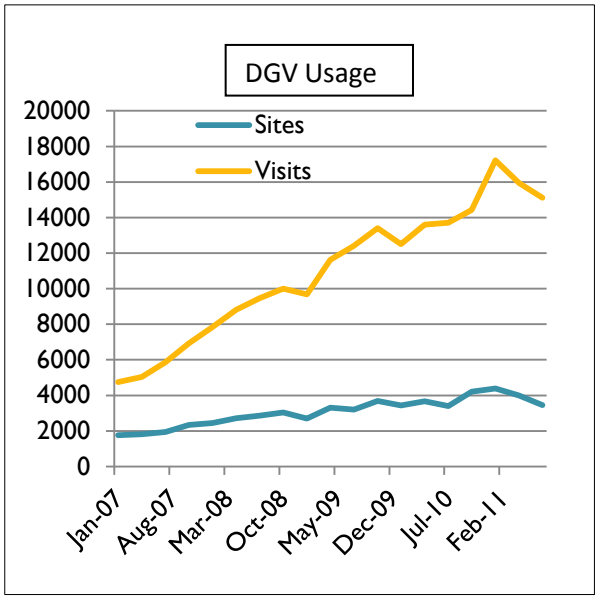
The *New* DGV

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

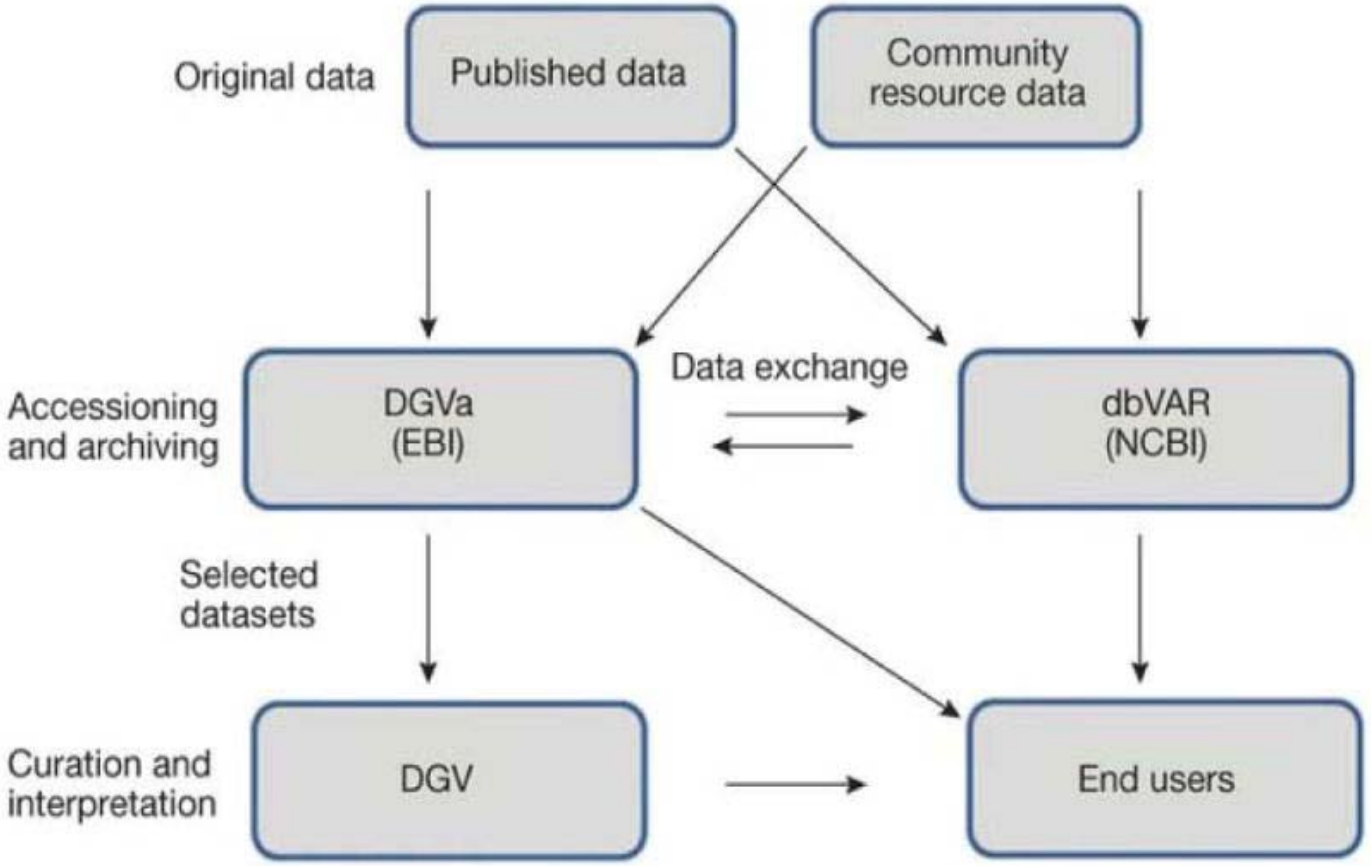
- The Database of Genomic Variants (DGV) has been working in partnership with the new database archives (DGVa and dbVar).
- In 2008, a collaboration between TCAG and EMBL-EBI, was established to collect, organize and curate genome-wide information on copy number variation. The goal was to expand the data curation capabilities to accommodate higher resolution array and sequencing data.





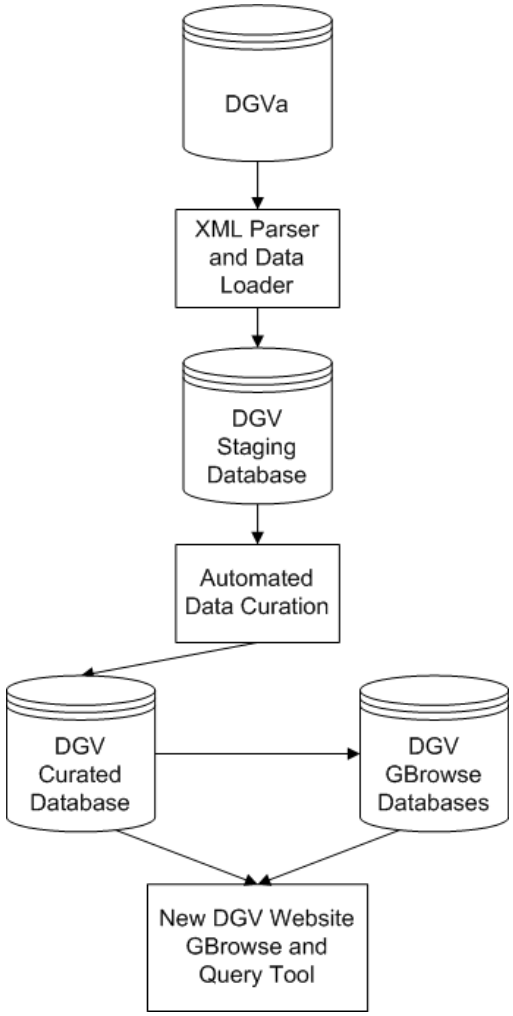


Public data archives for genomic structural variation



Church DM et. al. Nat Genet. 2010 Oct;42(10):813-4

Data Flow Diagram



Data Model/Definitions

Logical model

Project details | ER diagram | Entities | Attributes | Relationships | Domains | Views | Procedures | Sequences

List of entities | Entity details

Analysis
Cohort
Copy_Number
Coverage
Cytoband_Overlap
Dataset
Dataset_Analysis
Dataset_Xref
Feature_Overlap
Filter
Gene_Overlap
Merged_Analysis
Method_Platform_Sample
Method_Study
Platform_Study
Reference
Reference_Assembly
Sample
Sample_Pooled
Sample_Study
Sample_Xref
Study
Study_File_Prefix
Supporting_Merged_Variant
Translocation_Mapping
Variant
Variant_Analysis
Variant_Mapping
Variant_Type
Variant_Xref

Entity: Analysis

Entity details:

Description	This table captures information about each Analysis that is performed in one Study. A Study may perform one or more Analyses within that Study. An Analysis is the comparison of the Study's Dataset to a Reference.
Primary key constraint name	PK_Analysis

Attributes:

Key	Attribute name	Data type	Not null	Description
PK	Analysis_ID	SERIAL	Yes	This is the unique database identifer for one Analysis performed within a Study.
FK	Study_ID	INTEGER	Yes	This is the unique identifier from the 'Study' table and represents the Study in which an instance of an Analysis was employed.
	Analysis_Description	CHARACTER VARYING	No	This field captures a brief description of the Analysis employed in a Study. This information is not always provided in a Study. Note: specific criteria used during the Analysis should be captured in the 'Analysis_Criteria' field and a general description captured in this field.
	Analysis_Type	CHARACTER VARYING	No	If this information is provided, the Analysis Type will be captured in this field; this information is not always provided in a Study. An example of an Analysis Type is "split-read mapping".
	Analysis_Tool	CHARACTER VARYING	No	This field captures the name of the Analysis Tool employed in a given Analysis. Examples of Analysis Tools are "Birdsuite", "CNAG", "Genemapper".
	Analysis_Tool_Version	CHARACTER VARYING	No	This is a version number for the Analysis Tool that was used in a given Study. This information is not always provided for an Analysis Tool.
	Analysis_Tool_Description	CHARACTER VARYING	No	This is a description of the Analysis Tool that was used in a given Study. This information is not always provided for an Analysis Tool.
	Analysis_Tool_Type	CHARACTER VARYING	No	This field captures whether an Analysis Tool is an Algorithm or a Software Suite. This information is not always provided for a given Analysis Tool in a Study; DGVa and dbVar currently do not capture this information.

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 five 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
- Next generation sequencing approaches (NGS)

What is new in the *New* DGV

- Eight new studies added
- New genome browser
- New query tool
- New database schema/backend
- New servers to improve speed/performance

- Why have we launched the new DGV as a BETA site?

New Studies Added

Author/Study	Year	PubmedID	Sample Size	Variant Count	Study Accession
Durbin_et_al	2010	PubMedID=20981092	4719	212777	estd59
Perry et al 2008b	2008b	PubMedID=18775914	62	791	nstd8
Kidd et al 2010b	2010b	PubMedID=21111241	9	1167	nstd47
Teague et al	2010	PubMedID=20534489	4	4205	nstd49
Wilson et al	2010	PubMedID=21212237	1	2642	nstd50
Kidd et al	2010	PubMedID=20440878	9	9825	nstd35
Schuster et a	2010	PubMedID=20164927	1	187	nstd39
Alkan et al	2009	PubMedID=19718026	3	226	nstd31

What is new in the *New* DGV

- Eight new studies added
- New genome browser
- New query tool
- New database schema/backend
- New servers to improve speed/performance

- Why have we launched the new DGV as a BETA site?

DGV Beta Home 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: 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

Download Data

The contents of the Database of Genomic Variants can be downloaded as tab delimited text files. The first line of each file is the column description. Each variant as reported in the original study is given a VariationID and this represents a stable identifier. Since variants may overlap between studies we also cluster overlapping variants into loci. Note that the loci are redefined with each database update and therefore do not represent stable identifiers. Studies where CNV coordinates are based on the NCBI 36 (hg18) assembly are not mapped back to older builds. The Build 35 download file will therefore contain less data than the Build 36 download file.

Release Versions

Release Date	Build 35 (hg17)	Build 36 (hg18)	GRCh 37 (hg19)
2011-10-17	link	link	link
2011-10-14	link	link	link
2011-10-11	link	link	link

Filtered Variants

Release Date	GFF3 File
2011-10-17	link
2011-10-14	link
2011-10-11	link

Studies included in DGV2 and links to the original data

The table below outlines the studies that have been used as sources for the data and the number of regions included from each study. Here you can also download the original dataset for studies with more than ten CNV entries.

Study Summary

Show 50 entries
Showing 1 to 38 of 38 entries (filtered from 76 total entries)

Copy | Print | CSV | Excel | PDF | FAQ

lit. reference	year	regions	pubmed id	assembly	link to archive data	comments
Ahn et al 2009	2009	4091/4298	19470904	NCBI36/hg18	estd19	
Alkan et al 2009	2009	821/226	19718026	NCBI35/hg17	nstd31	A number of variant sub-type descriptions are missing from this study but will be included in the next DGV update.
Conrad et al 2006	2006	935/935	16327808	NCBI34/hg16	nstd17	
Conrad et al 2009	2009	72004/20206	19812545	NCBI36/hg18	estd20	
Cooper et al 2008	2008	368/368	18776910	NCBI35/hg17	nstd14	
De Smith et al 2007	2007	10093/1469	17666407	NCBI35/hg17	estd24	
Durbin_et_al_2010	2010	0/173	20981092		estd59	
Durbin_et_al_2010	2010	170048/212604	20981092	NCBI36/hg18	estd59	
Giglio et al 2002	2002	1/2	12058347	NCBI34/hg16	estd50	
Gusev et al 2009	2009	209/196	18971310	NCBI36/hg18	estd49	
Hinds et al 2006	2006	100/100	16327809	NCBI35/hg17	nstd34	
Iafrate et al 2004	2004	244/255	15286789	GRCh37/hg19	nstd41	This study identified variants in patients and controls. Variants found in control individuals were kept and variants found in patients were only kept if they were also found in the control samples.
Itsara et al 2009	2009	11166/13843	19166990	NCBI35/hg17	nstd27	Only variants detected in samples from the NINDS and HGDP are included in DGV. Variants found in the PARC samples were only kept if they were also found in the NINDS and/or HGDP samples.
Jakobsson et al 2008	2008	4836/1428	18288195	NCBI36/hg18	nstd30	
Kidd et al 2008	2008	18013/7458	18451855	NCBI35/hg17	nstd2	
Kidd et al 2010	2010	20/9825	20440878	NCBI36/hg18	nstd35	
Kidd et al 2010b	2010	1496/1167	21111241	NCBI36/hg18	nstd47	
Korbel et al 2007	2007	1139/1297	17901297	NCBI36/hg18	nstd16	
Levy et al 2007	2007	44/45	17803354	NCBI36/hg18	estd22	
Locke et al 2006	2006	338/388	16826518	NCBI34/hg16	nstd29	

BETA

Database of Genomic Variants

BETA

A curated catalogue of human genomic structural variation

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

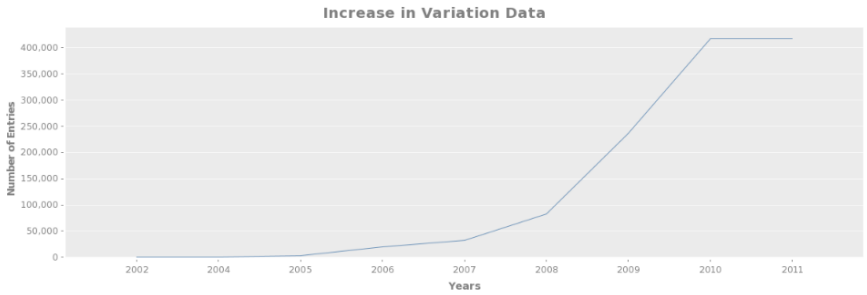
Stat	Merged-level	Sample-level
CNVs:	136804	277913
Inversions:	521	1096

Number of Studies: 37

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

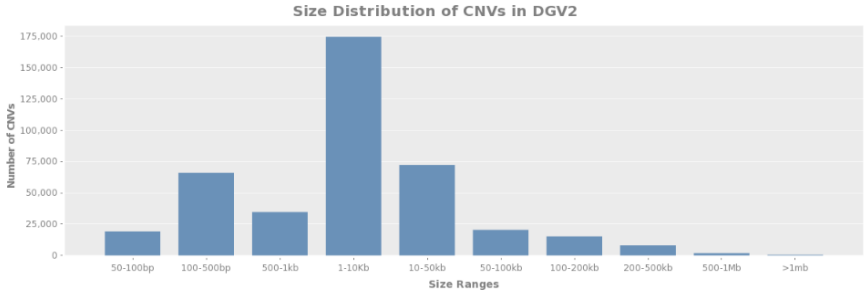
Content Growth

This graph shows the increase in published structural variation data that have been added to the database since its start in 2004; the numbers reflect the year of publication.



Size distribution (CNV)

The graph displays the size distribution of CNVs in the database.



Coverage (CNV):

The table shows the data underlying the graph showing the non-redundant CNV coverage per chromosome. Note that there are several biases going into the underlying data. S
still a lot of smaller CNVs (<30kb) that remain to be identified.

Chr	Nucleotides Covered	Chromosome Length	% Coverage
chr1	121227647	247249719	49.03
chr2	125984828	242951149	51.86
chr3	108996479	199501827	54.63
chr4	100193894	191273063	52.38
chr5	91036041	180857866	50.34
chr6	82312288	170899992	48.16
chr7	101083954	158821424	63.65
chr8	75697497	146274826	51.75
chr9	73156027	140273252	52.15
chr10	76578489	135374737	56.57
chr11	70474623	134452384	52.42
chr12	69013530	132349534	52.14
chr13	52173934	114142980	45.71
chr14	47038666	106368585	44.22
chr15	50921933	100338915	50.75
chr16	51184213	88827254	57.62
chr17	49969317	78774742	63.43
chr18	39134218	76117153	51.41
chr19	37979515	63811651	59.52
chr20	32418199	62435964	51.92
chr21	19108870	46944323	40.71
chr22	24255876	49691432	48.81
chrX	102696418	154913754	66.29
chrY	20698862	57772954	35.83
Total	1623335318	3080419480	52.7

Overlap:

The following table shows the overlaps of CNVs with relevant genomic features. It is important to remember that many CNVs currently have overestimated boundaries, which
category includes genes that are represented in the OMIM Morbid Map.

Feature	# of Features Overlapped by CNVs	# of CNVs Overlapped by Features
Transcripts (39426)	32797 (83.19%)	381752 (45.93%)
Exons (382693)	229150 (59.88%)	185170 (22.28%)
microRNA (1120)	713 (63.66%)	6734 (0.81%)
OMIM (2785)	2399 (86.14%)	62358 (7.5%)
Segmental Duplications (51809)	42108 (81.28%)	167112 (20.11%)

Summary of Filtered Variants

Filtered Summary

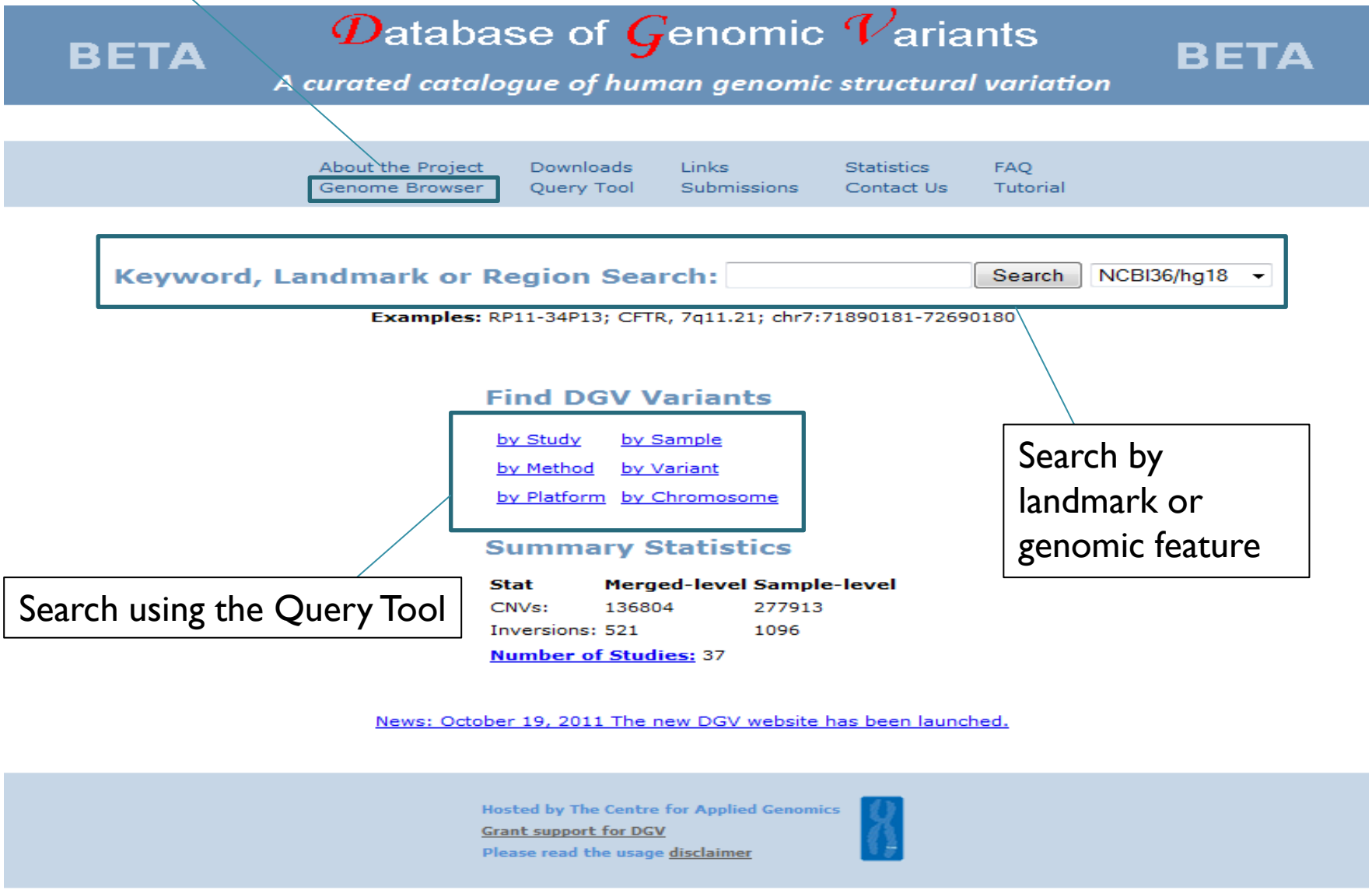
Show 50 entries
Showing 1 to 33 of 33 entries

Copy | Print | CSV | Excel | PDF | FAQ

study	number of filtered variants	reasons for filtering
Ahn et al 2009	8596	Variant was not associated to a reference assembly Inferred variant Variant is too small (< 50 bp) Variant had all its reported mappings filtered out Reported merged variant has had all its supporting variants filtered out
Conrad et al 2006	1870	Inferred variant Reported merged variant has had all its supporting variants filtered out
Conrad et al 2009	94	Overlaps 50% of a decipher region. (hg18) Overlaps 50% of a decipher region. (hg19) Overlaps 50% of a gap region. (hg19) Reported merged variant has had all its supporting variants filtered out Supporting variant had its parent variant filtered.
Cooper et al 2008	736	Inferred variant Reported merged variant has had all its supporting variants filtered out
De Smith et al 2007	388	Overlaps 50% of a decipher region. (hg18) Overlaps 50% of a decipher region. (hg19) Supporting variant had its parent variant mapping filtered. Variant is too small (< 50 bp) Variant had all its reported mappings filtered out Reported merged variant has had all its supporting variants filtered out Supporting variant had its parent variant filtered. Variant is too large (> 3000000 bp)
Durbin_et_al_2010	2270079	Overlaps 50% of a decipher region. (hg18) Overlaps 50% of a decipher region. (hg19) Overlaps 50% of a gap region. (hg18) Inferred variant Supporting variant had its parent variant mapping filtered. Variant is too small (< 50 bp) Overlaps 50% of a gap region. (hg19) Variant had all its reported mappings filtered out Reported merged variant has had all its supporting variants filtered out Supporting variant had its parent variant filtered. Variant is too large (> 3000000 bp)

Enter the Genome Browser

Points of Entry into the Database



Search by landmark or genomic feature

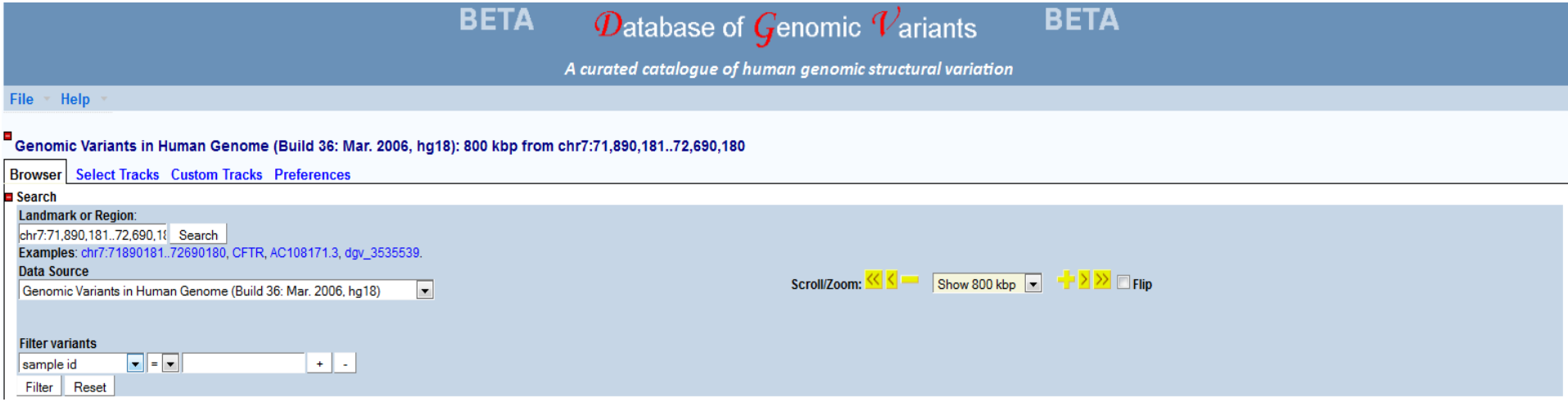
Search using the Query Tool

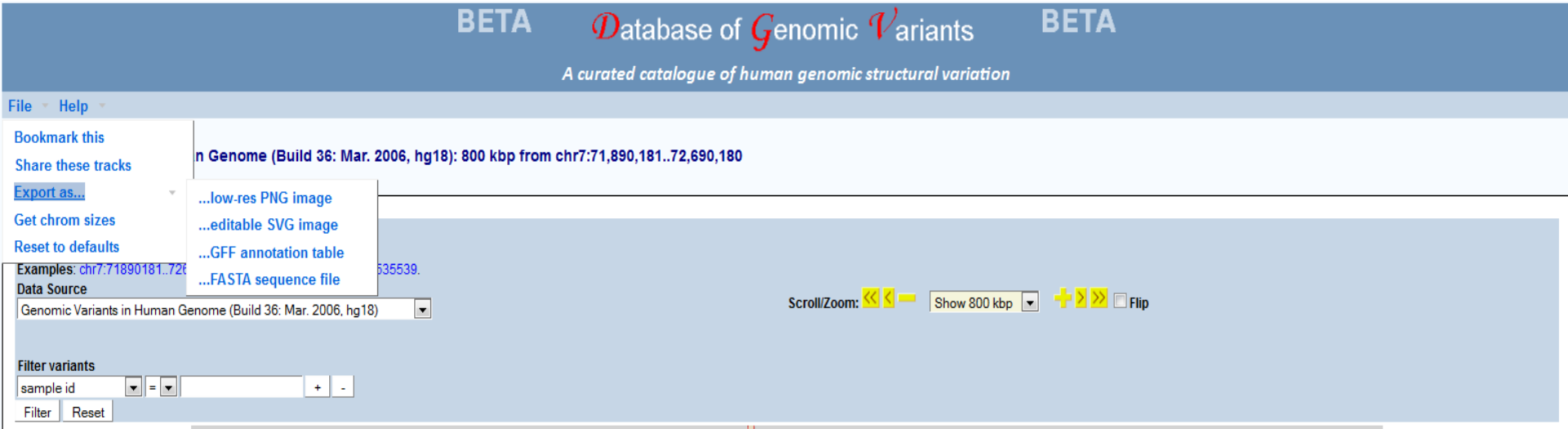
Overview

- Introduction to the *New* DGV
- Genome Browser
- Query Tool
- Examples of how to use the data
- Summary/Questions

Genome Browser

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





Genome Browser Track Options

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

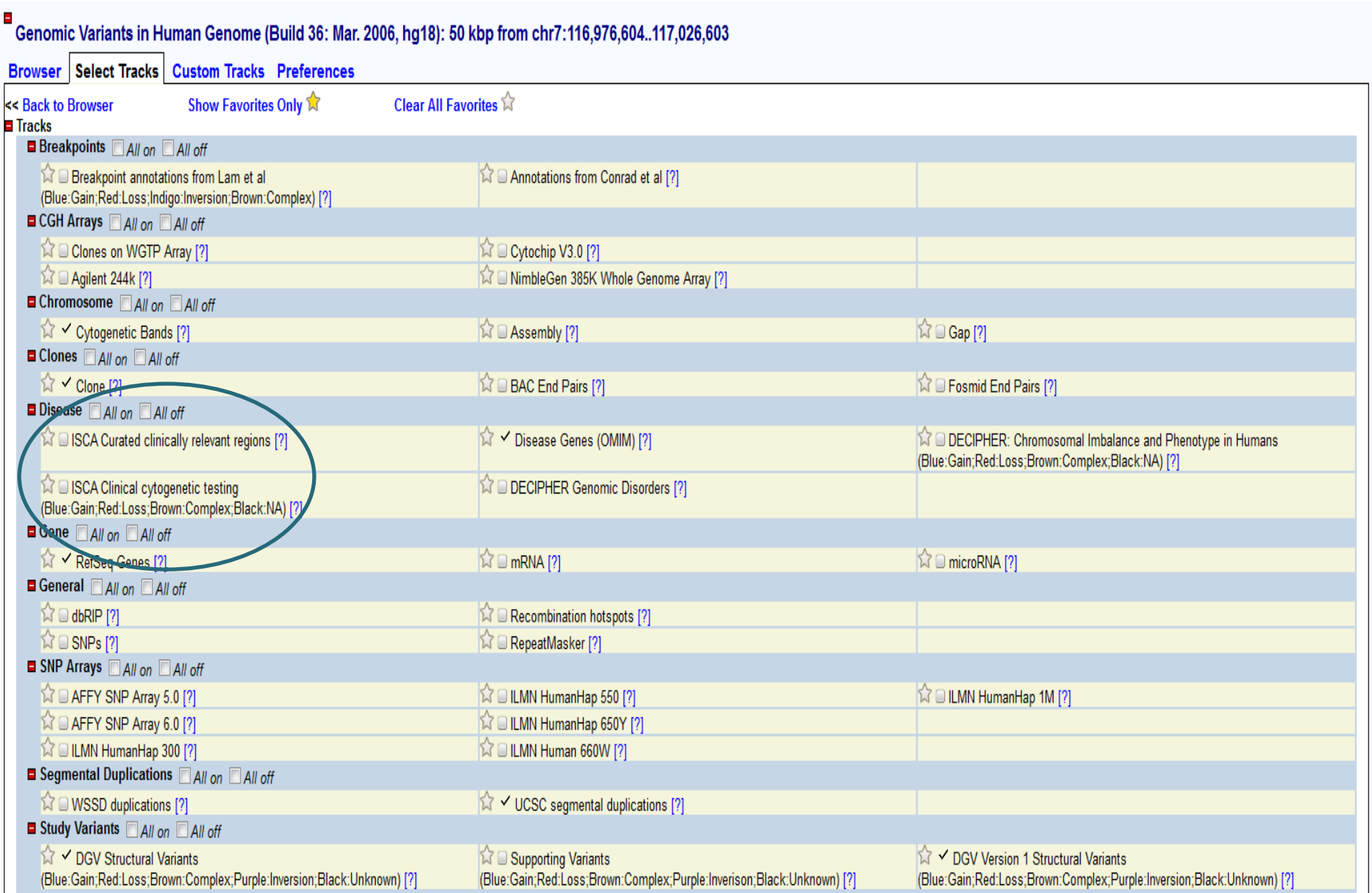
Browser | Select Tracks | Custom Tracks | Preferences

<< Back to Browser Show Favorites Only Clear All Favorites

Tracks

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

Genome Browser Track Options

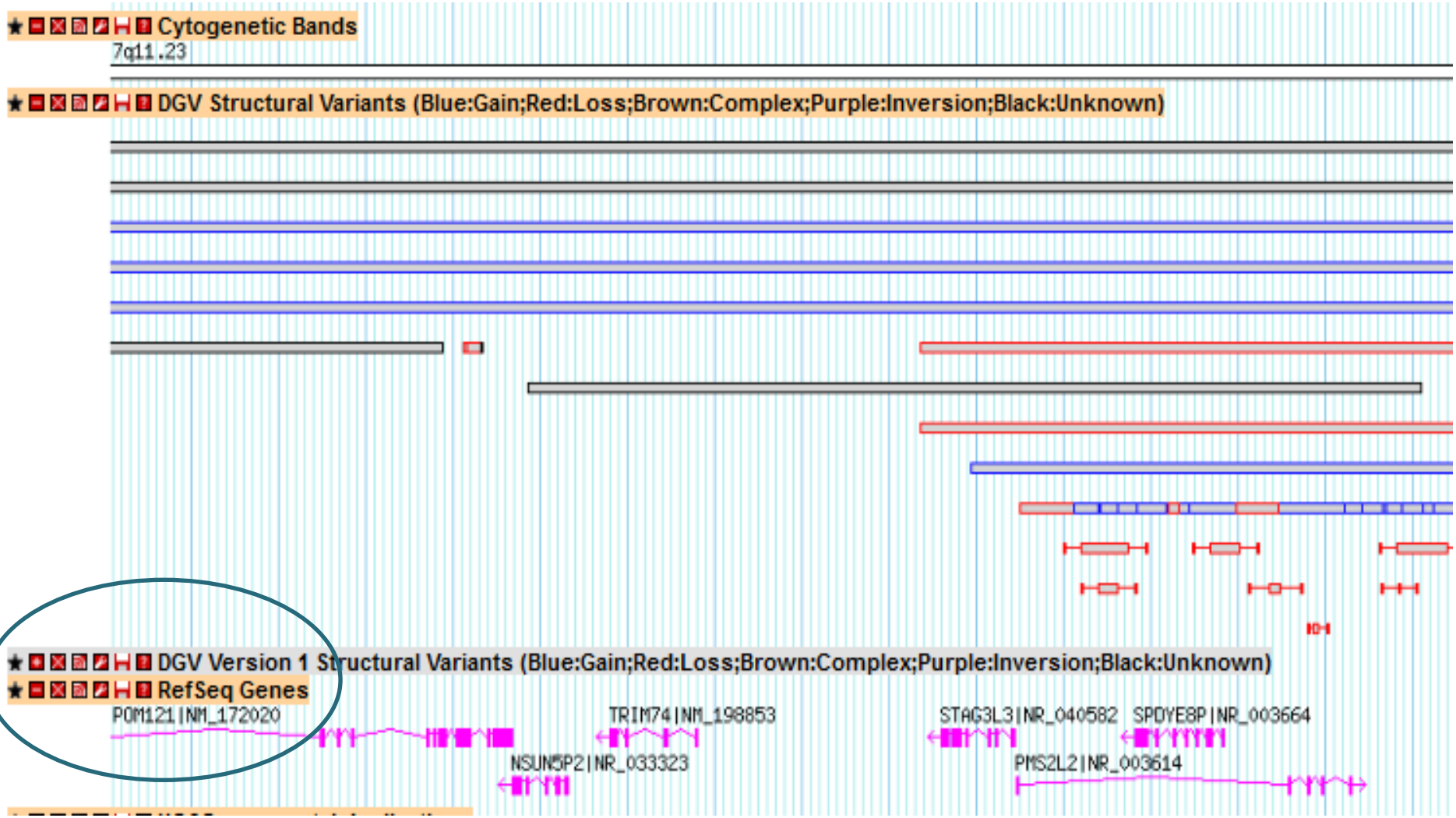


nstd45
Study type: Curated
Submitter: ISCA Consortium
International Standard Cytogenetic Array Consortium Curated Dataset
PubMed: ISCA - Curated 2011
Name: ISCA curated clinically relevant regions for cytogenetic testing
Description:
Annotated clinically relevant regions used in interpretation of cytogenetic testing
Size: 57 (sample data not available)
Subject Phenotype(s): Developmental Delay and additional significant developmental and morphological phenotypes referred for genetic testing

---

nstd37
Study type: Case-Set
Submitter: ISCA Consortium
Description:
International Standard Cytogenetic Array Consortium
PubMed: ISCA 2011
dbGaP: To gain access to subject level data, see dbGaP. (dbGap release date: June 2011) Name: Clinical_cytogenetic_testing_referral (All)
All patients referred for cytogenetic testing due to clinical phenotypes
Size: 7605 (sample data available via dbGaP)
Subject Phenotype(s): Dev. Delay, or other clinically referred phenotypes

Genome Browser Track Options



Genome Browser Track Options

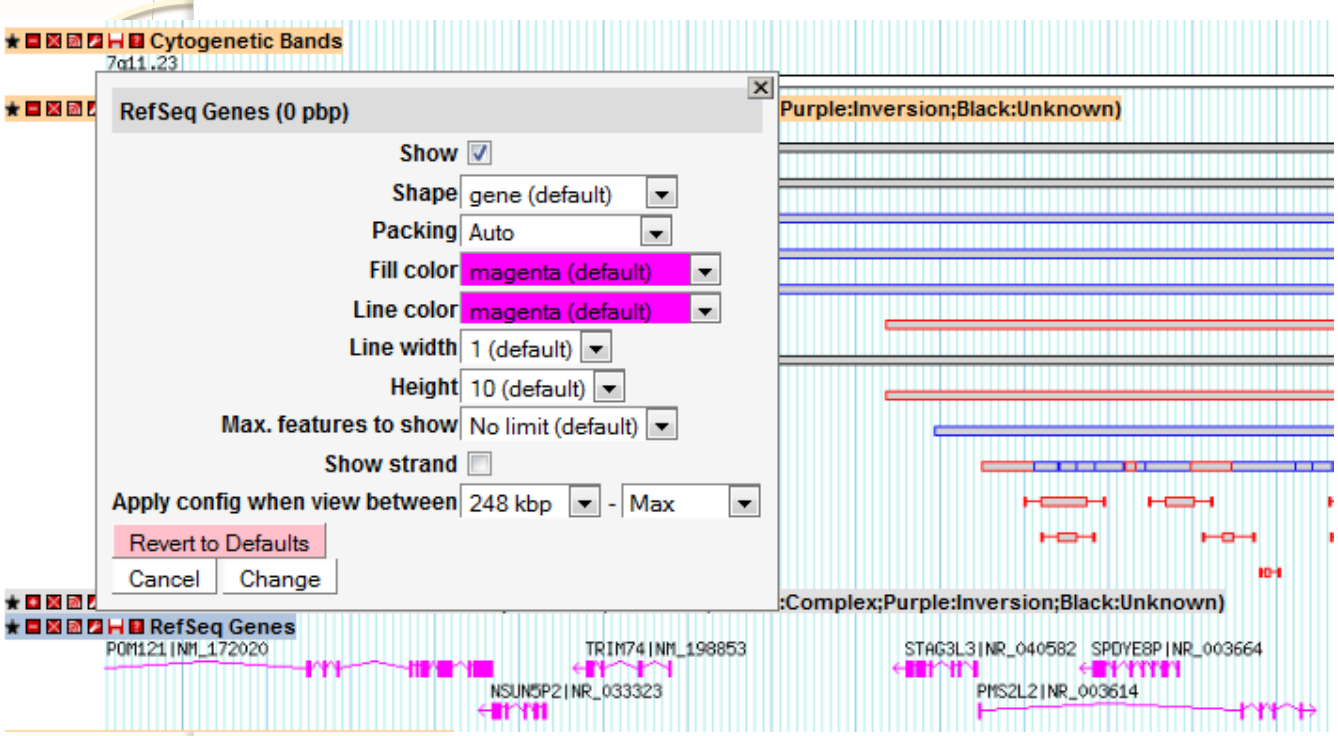
Minimize the track

Displays information about the data
In the track.

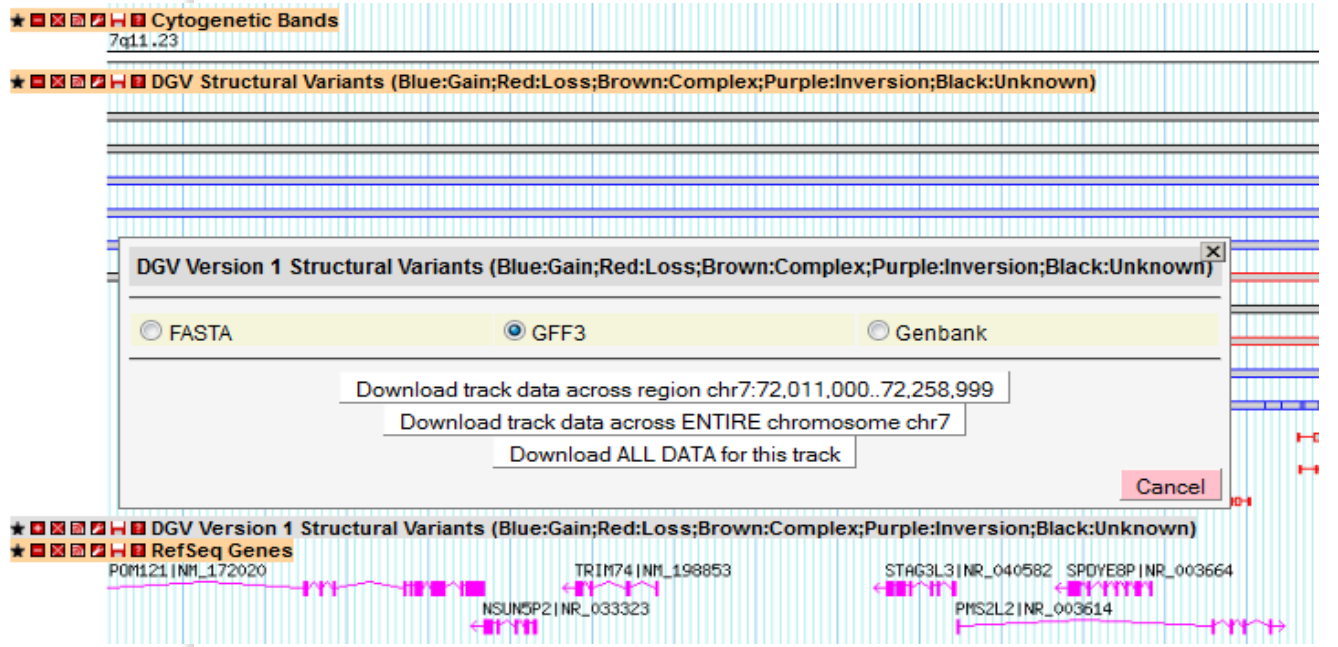
Close the track.

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.

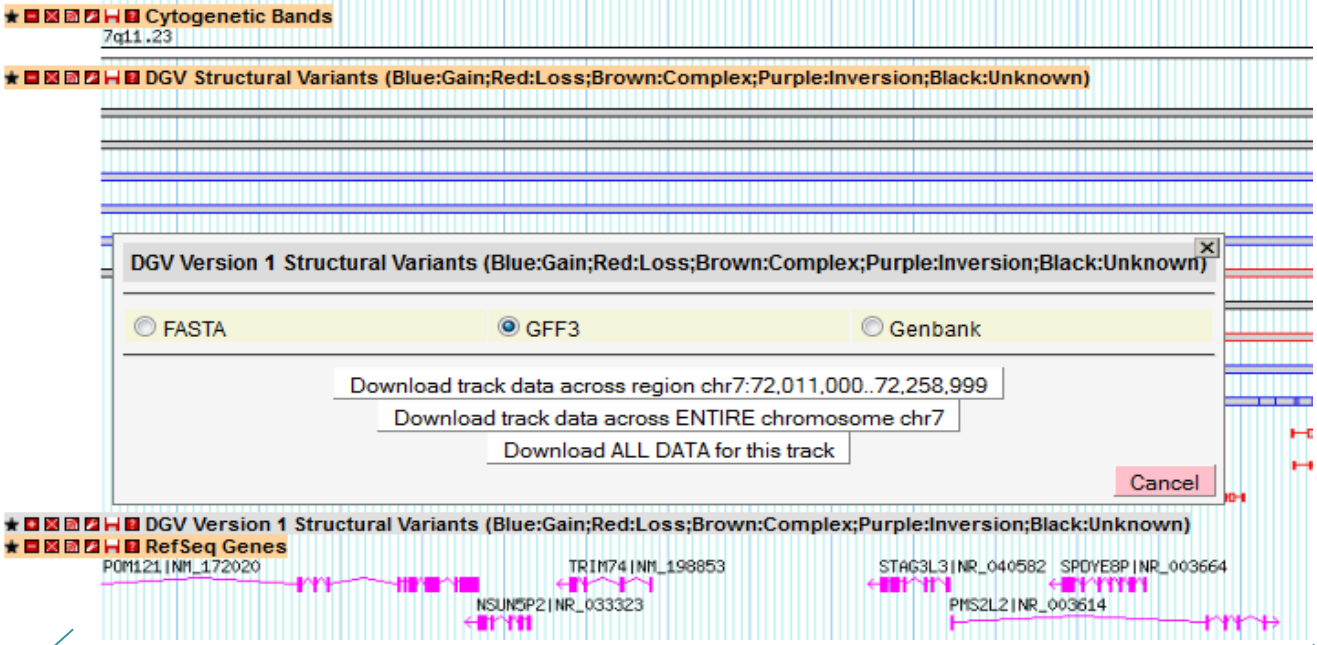
Allows users to change how the data are displayed.



Change Display



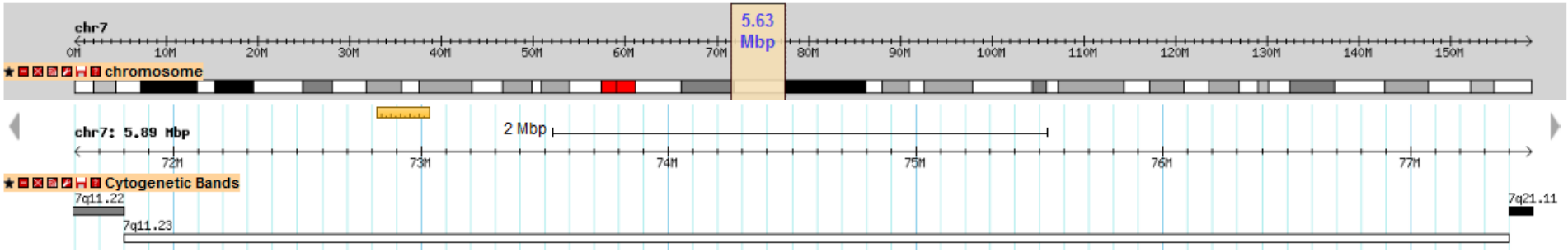
Save the Data



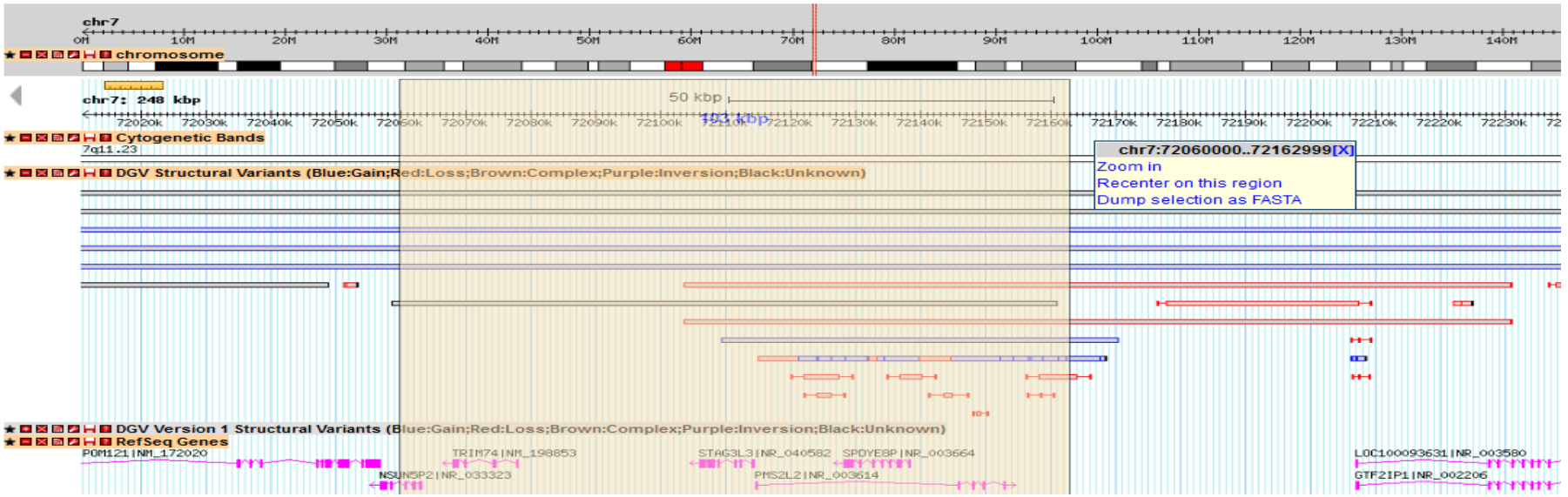

```
##gff-version 3
##date Wed Dec  7 10:24:05 2011
##source gbrowse gbgff gff3 dumper
##sequence-region chr7:73536000..73559999
chr7    DGVA    merged_variant  73547601        73547601        .       .       .       Name=dgv_634129;ID=58156;method_name=Sequencing;variant_
chr7    DGVA    merged_variant  73550100        73550100        .       .       .       Name=dgv_634129;ID=58156;variant_sub_type=Loss;method_na
chr7    DGVA    merged_variant  73548600        73548800        .       .       .       Name=dgv_634129;ID=58156;variant_sub_type=Loss;method_na
chr7    DGVA    merged_variant  73546251        73546251        .       .       .       Name=dgv_634351;ID=58368;method_name=Sequencing;variant_
chr7    DGVA    merged_variant  73550850        73550850        .       .       .       Name=dgv_634351;ID=58368;variant_sub_type=Loss;method_na
chr7    DGVA    merged_variant  73548550        73548551        .       .       .       Name=dgv_634351;ID=58368;variant_sub_type=Loss;method_na
chr7    DGVA    sample_variant  73546851        73546851        .       .       .       Name=dgv_2997458;ID=333879;method_name=Sequencing;variar
chr7    DGVA    sample_variant  73550050        73550050        .       .       .       Name=dgv_2997458;ID=333879;variant_sub_type=Loss;method_
chr7    DGVA    sample_variant  73548050        73548851        .       .       .       Name=dgv_2997458;ID=333879;variant_sub_type=Loss;method_
```

Genome Browser Navigation Options

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

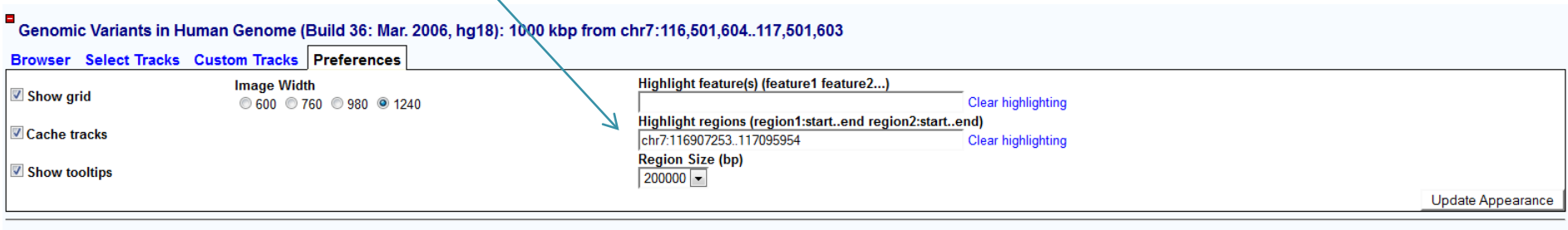


2. Click and drag on the genomic position bar to zoom in or centre on a region of interest.

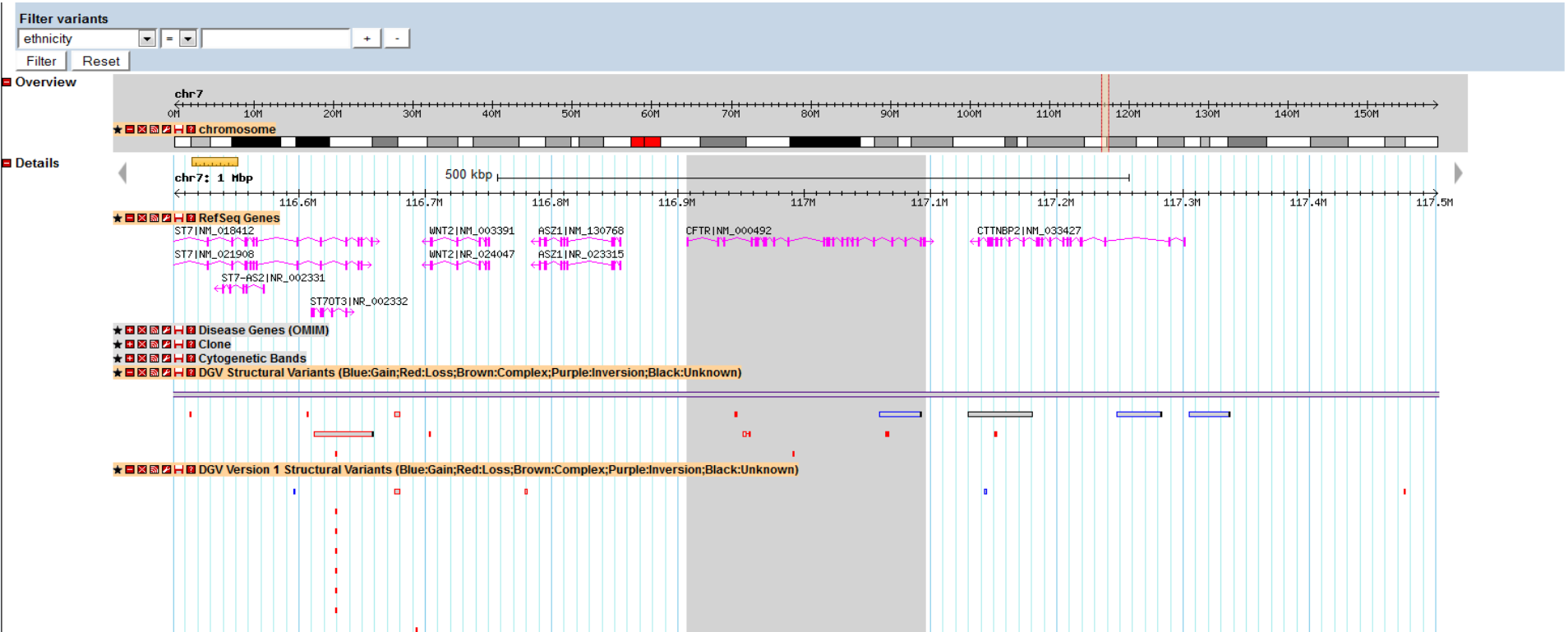


Genome Browser Visualization Options

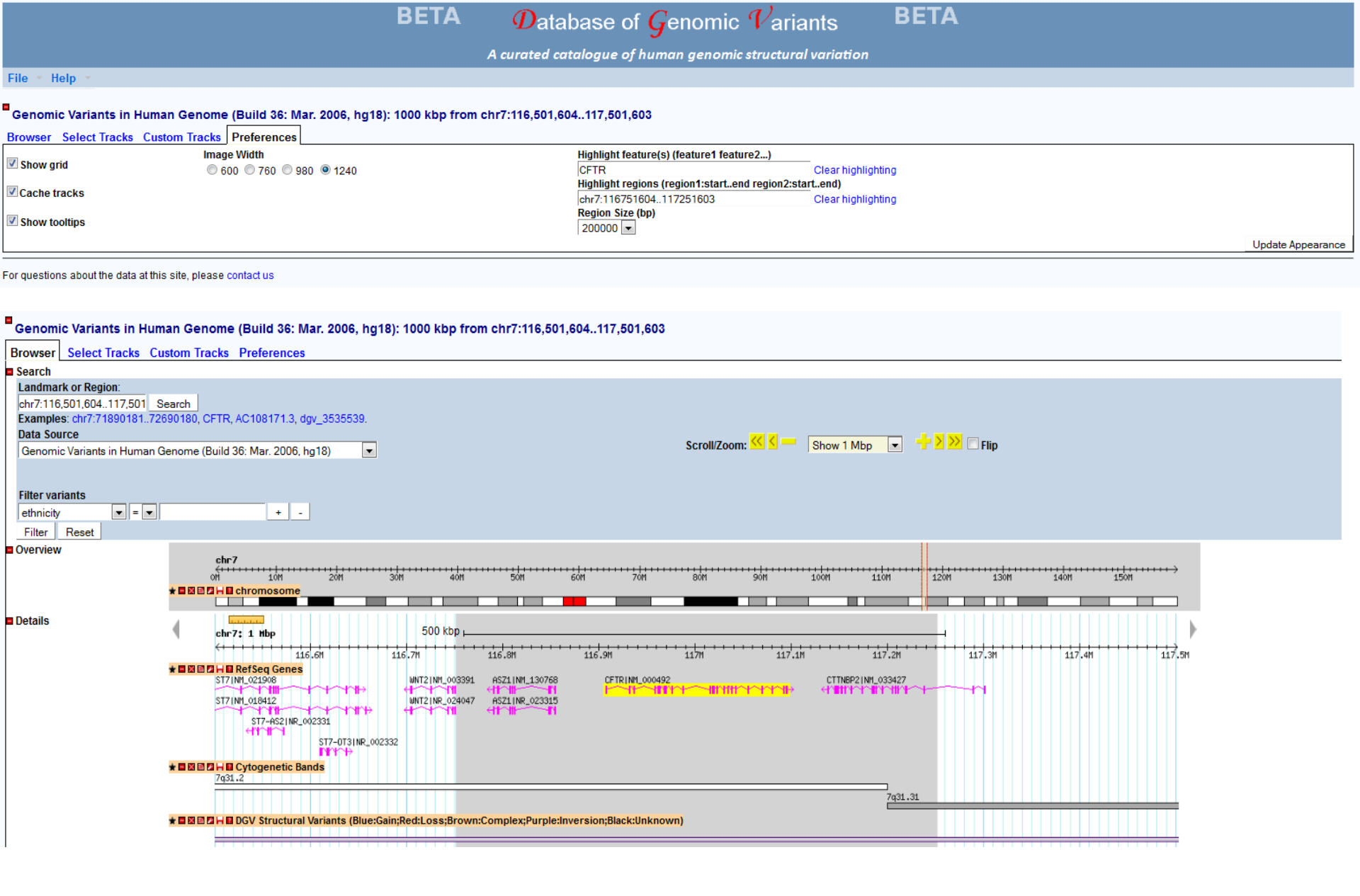
To highlight a region of interest in the browser, select the Preferences tab, and add the target region into this box.



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



Genome Browser Visualization Options

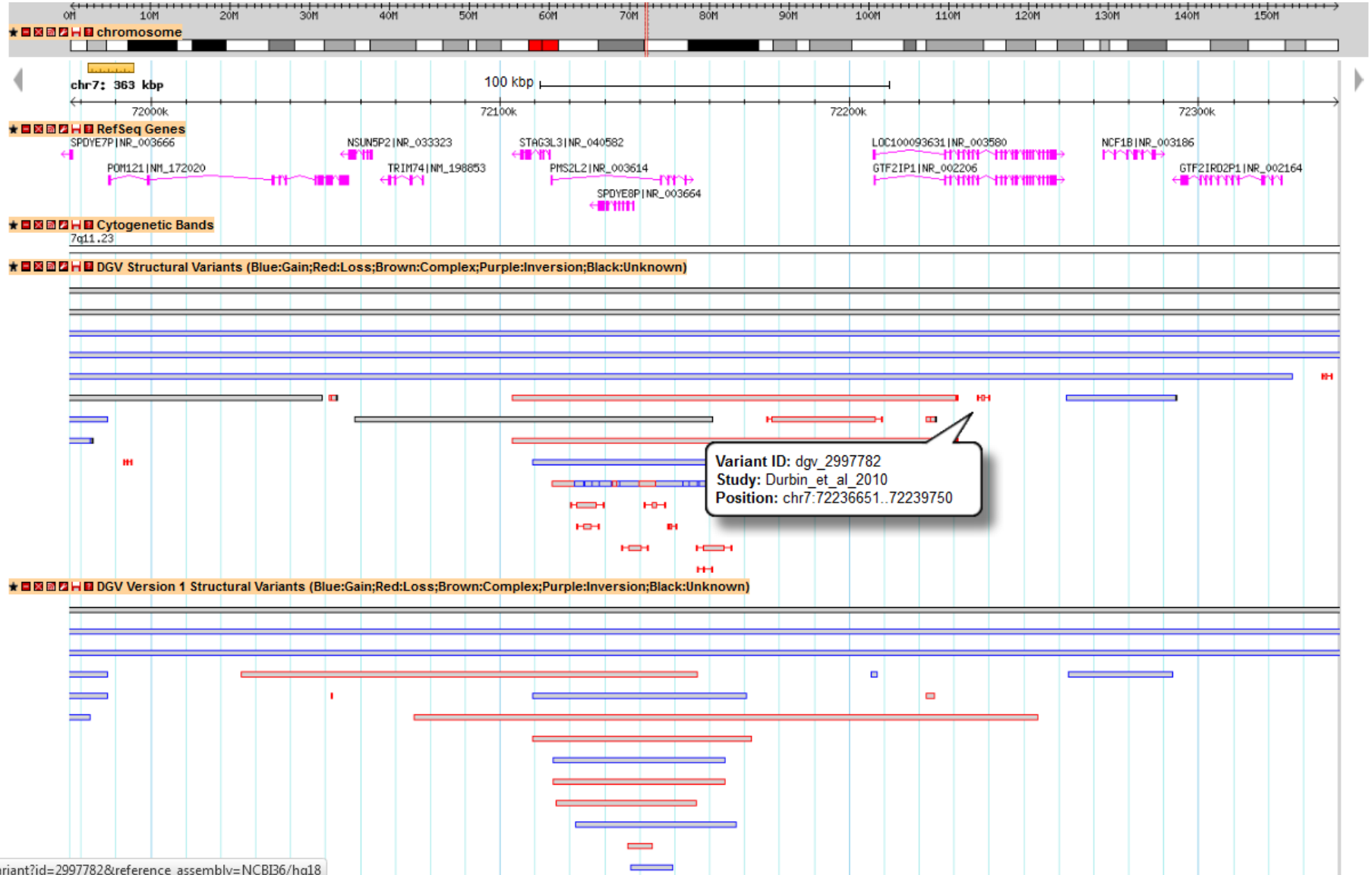


DGV Genomic Variant Data

- We have a total of 275,536 sample level variants and 137,493 merged variants in DGV from 37 different studies from a total of 5,709 samples.
- We have both the original and new data displayed as separate tracks for completeness and to allow users to compare/contrast the entries in both. We have maintained active links to the old DGV as well..
- The underlying supporting variants are also displayed in a separate track so users can visualize the contributing variants which comprise the merged calls.
- Variant types now include Inversions, CNV and Other/Complex. Variant subtypes are used to describe gain/loss/insertion/deletion.
- DGV now contains variants that are 50bp and larger and no longer differentiate InDels and CNVs.

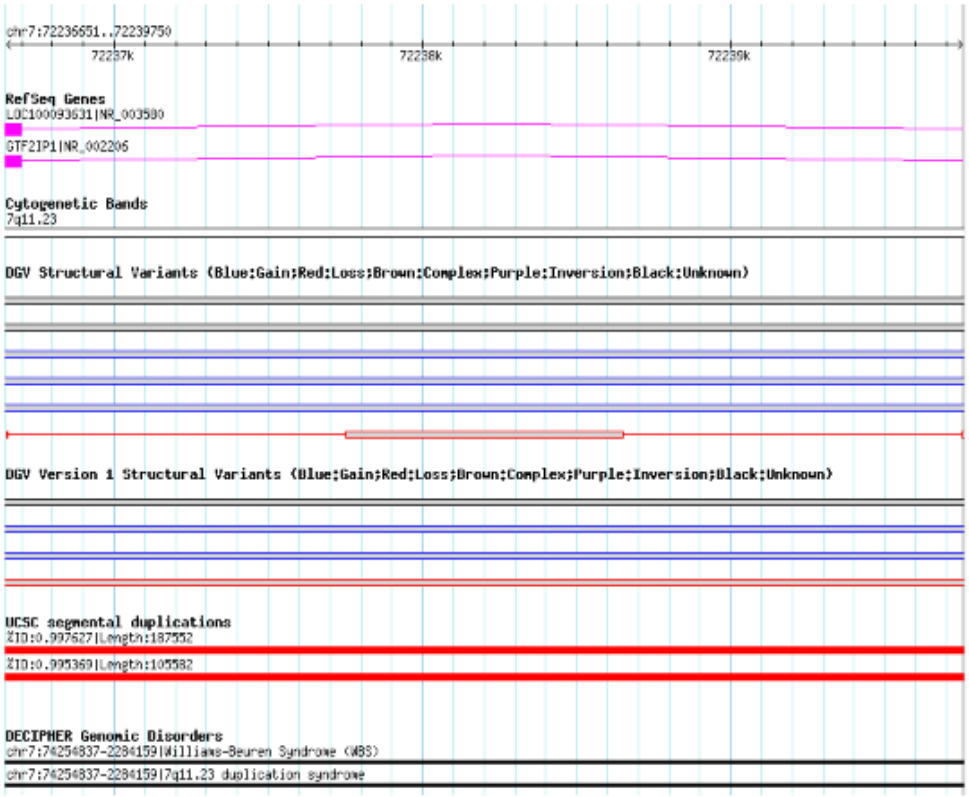
DGV Genomic Variant Data

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



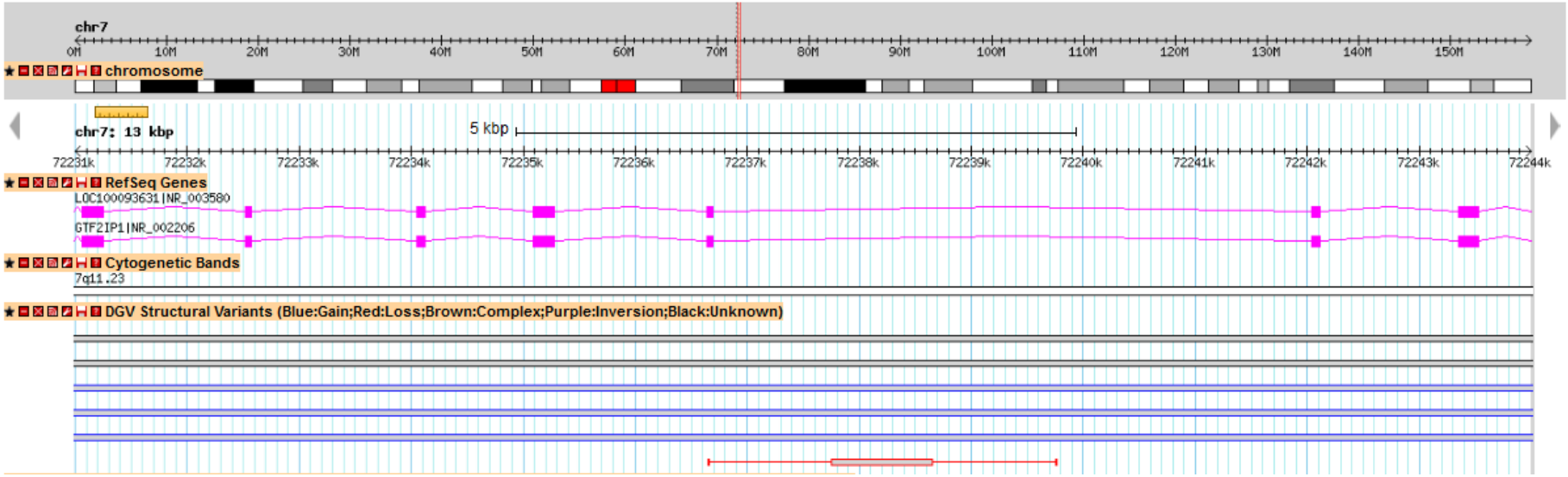
Variant Details

Variant: dgv_2997782



Variant ID	dgv_2997782
Landmark	
Genomic Position	
Genomic Position (Inner)	chr7:72238651..72237750 (UCSC Ensembl)
Genomic Position (Outer)	chr7:72236651..72239750 (UCSC Ensembl)
Cytoband	7q11.23
Variant Type	CNV
Merged Status	S
Merged Variants	
Supporting Variants	
Samples	
Known Genes	
Method	Sequencing
Analysis	Analysis structural variant merging of all supporting structural variant calls generated by type of computational approach: read depth analysis.
Comments	
Reference	Durbin_et_al_2010
Pubmed ID	20981092
External IDs	esv82353
Frequency	

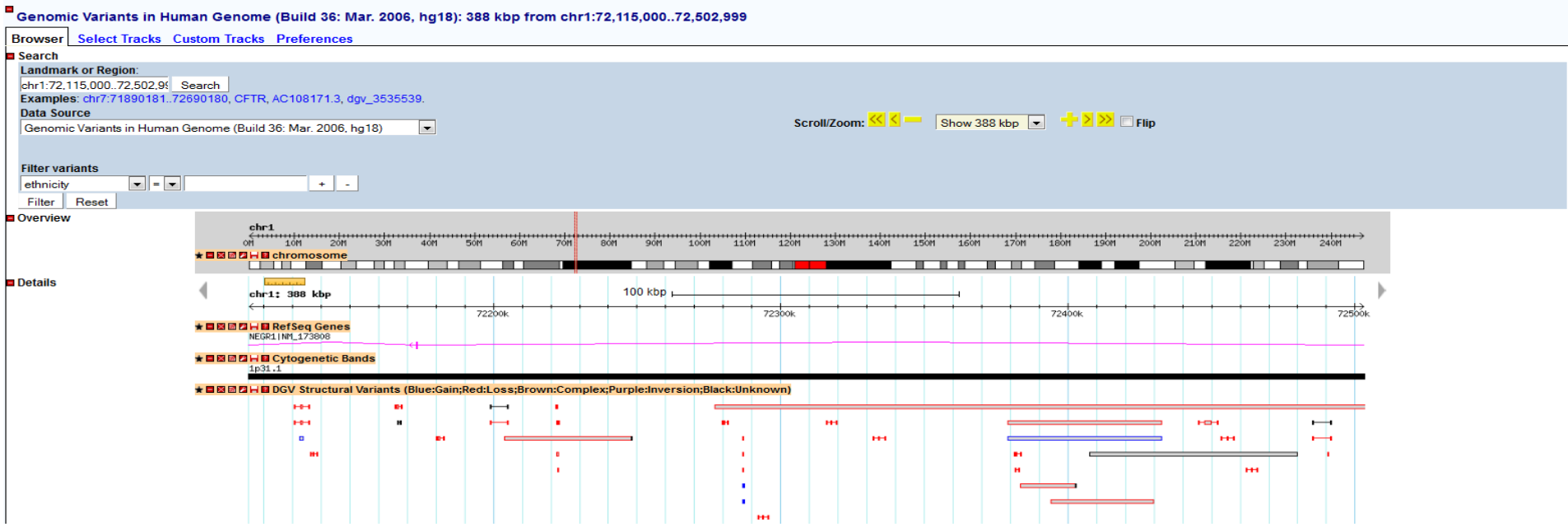
There are a few different types if images and different colours used to display structural variants in the genome browser.

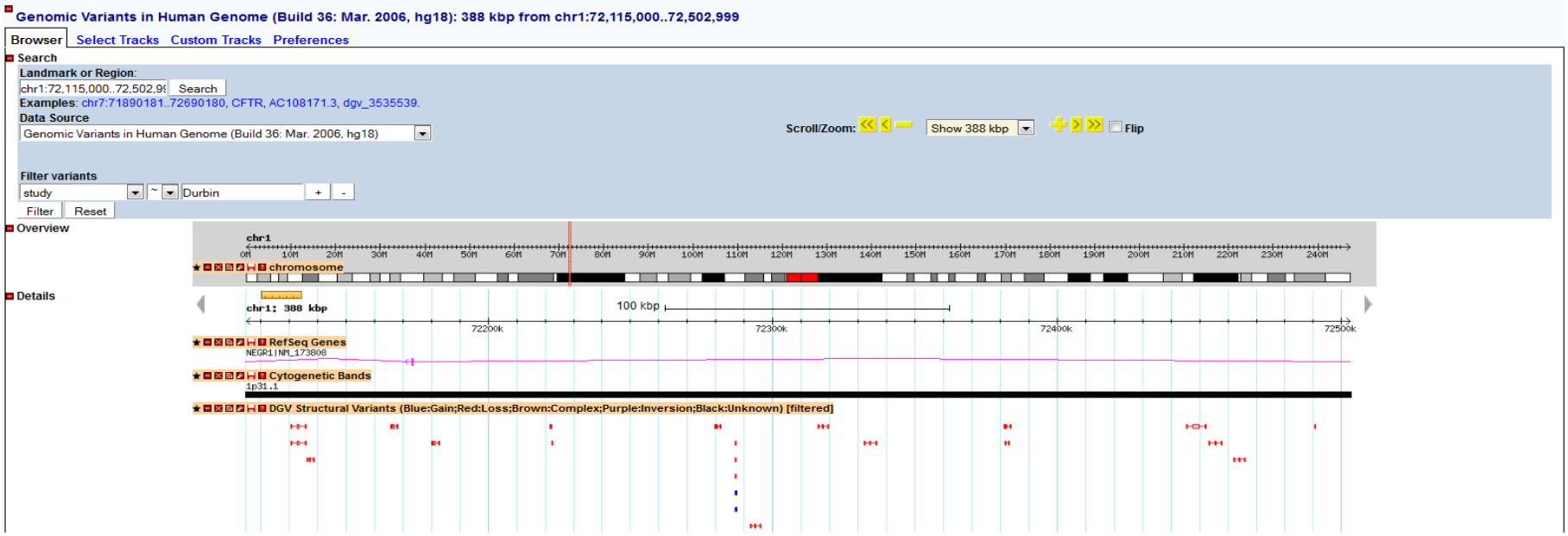


Variant ID	dgv_2993496
Landmark	
Genomic Position	
Genomic Position (Inner)	chr7:116381201..116380950 (UCSC Ensembl)
Genomic Position (Outer)	chr7:116379801..116382650 (UCSC Ensembl)
Cytoband	7q31.2
Variant Type	CNV
Merged Status	S

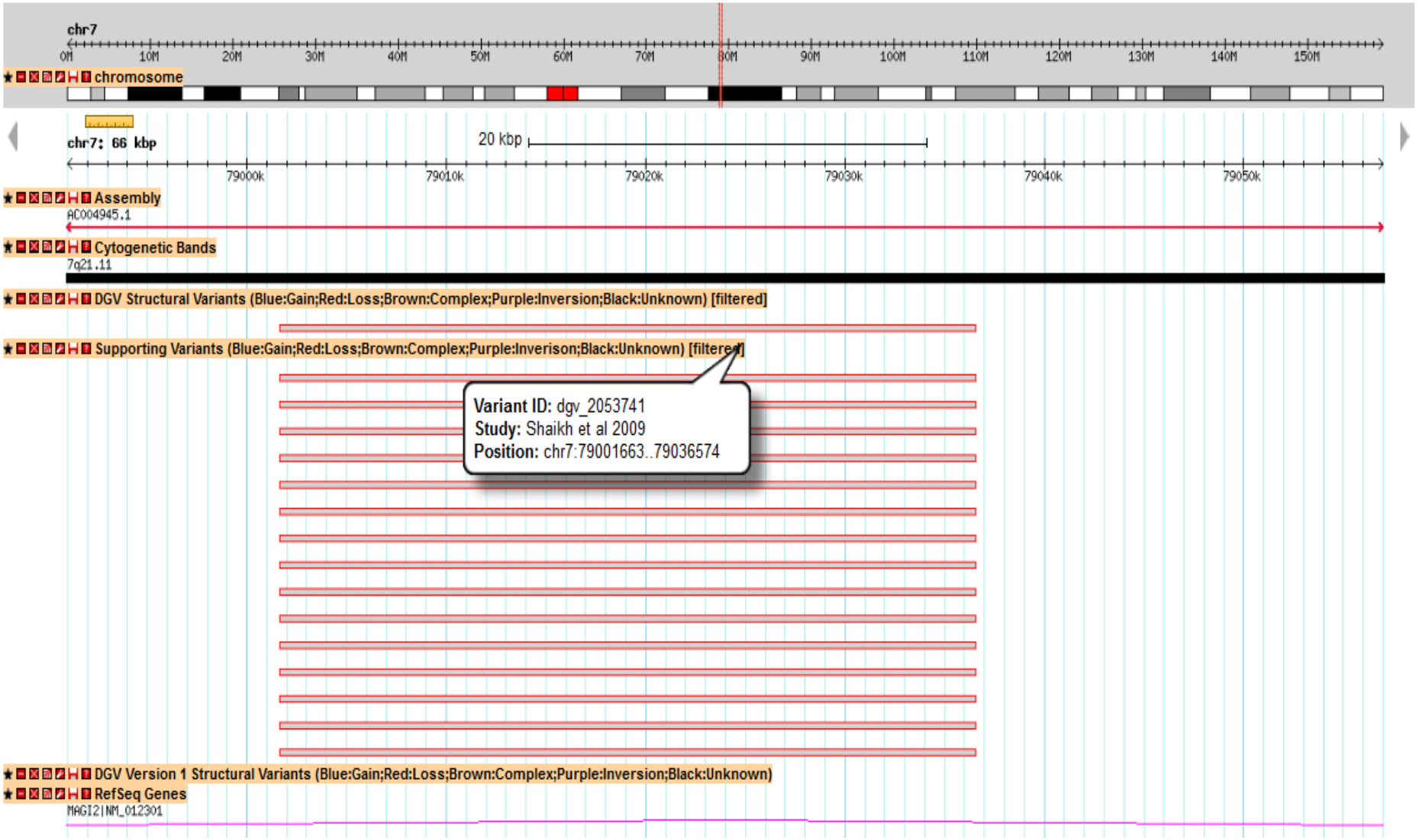
Variant ID	dgv_380175
Landmark	
Genomic Position	
Genomic Position (Inner)	
Genomic Position (Outer)	chr7:112147851..121027925 (UCSC Ensembl)
Cytoband	7q31.1
Variant Type	OTHER
Merged Status	S

Applying Filters to DGV Variants





DGV Supporting Variant Data



Overview

- Introduction to the *New* DGV
- Genome Browser
- Query Tool
- Examples of how to use the data
- Summary/Questions

The Query Tool

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

assembly = NCBI36/hg18 + -

Filter query | Reset

Study | Variants | Samples | Methods | Platforms | Analyses | Filtered Variants

Show 50 entries | Copy | Print | CSV | Excel | PDF | FAQ

Showing 1 to 35 of 35 entries (filtered from 37 total entries)

accession	study	pubmed id	sample size	variant count
estd1	Redon et al 2006	17122850	271	3103
estd19	Ahn et al 2009	19470904	1	4091
estd20	Conrad et al 2009	19812545	451	72004
estd21	Wheeler et al 2008	18421352	3	23
estd22	Levy et al 2007	17803354	2	44
estd24	De Smith et al 2007	17666407	51	10093
estd3	Wang et al 2008	18987735	1	2663
estd48	Stefansson et al 2005	15654335	1	1
estd49	Gusev et al 2009	18971310	270	209
estd50	Giglio et al 2002	12058347	0	1
estd55	Pinto et al 2007	17911159	745	974
estd59	Durbin_et_al_2010	20981092	186	170048
nstd1	Tuzun et al 2005	15895083	1	296
nstd14	Cooper et al 2008	18776910	9	368
nstd16	Korbel et al 2007	17901297	2	1139
nstd17	Conrad et al 2006	16327808	60	935
nstd2	Kidd et al 2008	18451855	9	18013
nstd20	McCarroll et al 2006	16468122	269	4444
nstd21	Shaikh et al 2009	19592680	2026	65018
nstd22	McCarroll et al 2008	18776908	270	1319

Introduction to the Query Tool

assembly = NCBI36/hg18 + -

Filter query | Reset

Study | Variants | Samples | Methods | Platforms | Analyses | Filtered Variants

Show 50 entries | Copy | Print | CSV | Excel | PDF | FAQ

Showing 1 to 35 of 35 entries (filtered from 37 total entries)

accession	study	pubmed id	sample size	variant count
estd1	Redon et al 2006	17122850	271	3103
estd19	Ahn et al 2009	19470904	1	4091

assembly = NCBI36/hg18 + -

Filter query | Reset

Study | Variants | Samples | Methods | Platforms | Analyses | Filtered Variants

Show 50 entries | Copy | Print | CSV | Excel | PDF | FAQ

Showing 1 to 50 of 415,566 entries (filtered from 1,983,157 total entries)

chromosome	start	stop	study	variant id	M=Merged, S=Sample Call	ethnicity	gender	assembly	variant type	variant subtype
1	27337400	27338705	Conrad et al 2009	dgv 2691686	S	YRI	Female	NCBI36/hg18	CNV	Loss
1	61855378	61856336	Durbin_et_al_2010	dgv 3014505	M			NCBI36/hg18	CNV	Loss

assembly = NCBI36/hg18 + -

Filter query | Reset

Study | Variants | Samples | Methods | Platforms | Analyses | Filtered Variants

Show 50 entries | Copy | Print | CSV | Excel | PDF | FAQ

Showing 1 to 50 of 2,830 entries (filtered from 26,060 total entries)

study	external sample id	family id	source	sample description	ethnicity	gender	cohort name
Ahn et al 2009	SJK		DNA		KOREAN	Male	
Conrad et al 2006	NA19094		DNA		YORUBA	Female	30 Yoruba trios
Conrad et al 2006	NA19208		DNA		YORUBA	Male	30 Yoruba trios

Introduction to the Query Tool (2)

assembly = NCBI36/hg18 + -

Filter query | Reset

Study | Variants | Samples | **Methods** | Platforms | Analyses | Filtered Variants

Show 50 entries — Copy | Print | CSV | Excel | PDF | FAQ

Showing 1 to 43 of 43 entries (filtered from 496 total entries)

method name	study	method type	method description
BAC_aCGH	Locke et al 2006	Discovery	Array hybridizations were performed as described by Snijders et al., 2001 with use of the segmental duplication array. The segmental duplication array consists of 2,007 BACs, spotted in triplicate, that were targeted to 130 complex regions of the genome and flanked by intrachromosomal segmental duplications. All 269 individuals were hybridized, with dye-swap replicate experiments, to the segmental duplication array with use of a single reference individual (Coriell ID GM15724) for comparison.
BAC_aCGH	Redon et al 2006	Discovery	Comparative hybridization using a Whole Genome TilePath (WGTP) array. A total of 26,678 large insert clones were selected from the published Golden Path to cover the human genome in tiling path resolution. Clones were screened for T1 phage and Pseudomonas contamination and verified by finger printing and end sequencing.

assembly = NCBI36/hg18 + -

Filter query | Reset

Study | Variants | Samples | Methods | **Platforms** | Analyses | Filtered Variants

Show 50 entries — Copy | Print | CSV | Excel | PDF | FAQ

Showing 1 to 43 of 43 entries (filtered from 504 total entries)

study	platform name	platform type	platform version	geo accession	arrayexpress accession
Ahn et al 2009	Illumina Genome Analyzer				
Alkan et al 2009	Illumina Genome Analyzer and Roche/454 sequencer.				
Conrad et al 2006	Not Provided				
Conrad et al 2009	Sanger H. Sapiens NimbleGen 42M CGH Array				E-MTAB-40
Cooper et al 2008	Illumina Human1Mv1 DNA Analysis BeadChip (Human1Mv1_C)			GPL6983	

study ~ Alkan + -

Filter query | Reset

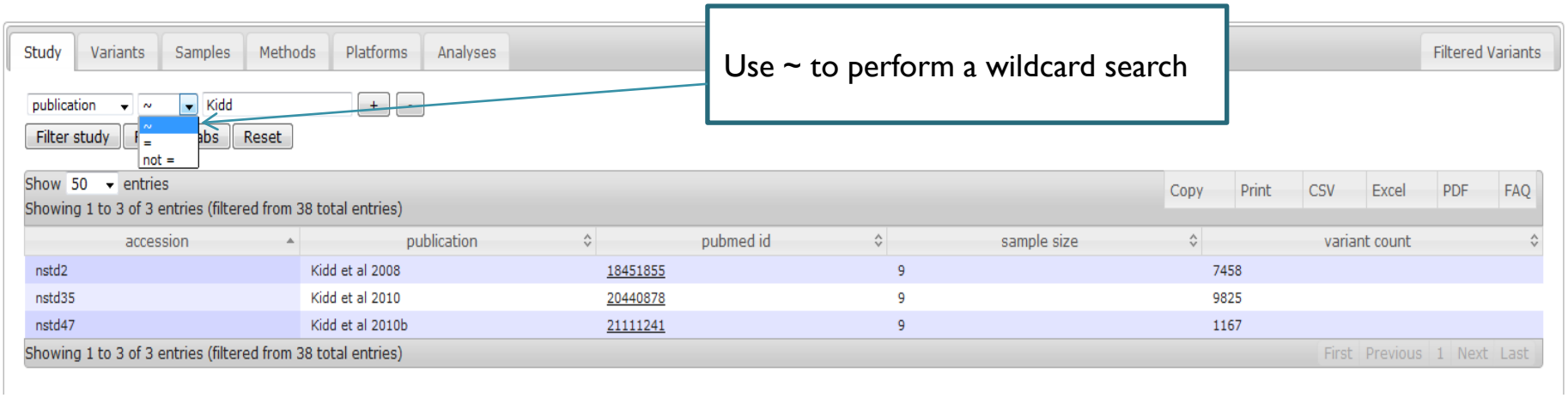
Study | Variants | Samples | Methods | Platforms | **Analyses** | Filtered Variants

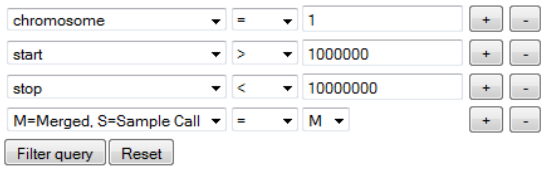
Show 50 entries — Copy | Print | CSV | Excel | PDF | FAQ

Showing 1 to 3 of 3 entries (filtered from 1,557 total entries)

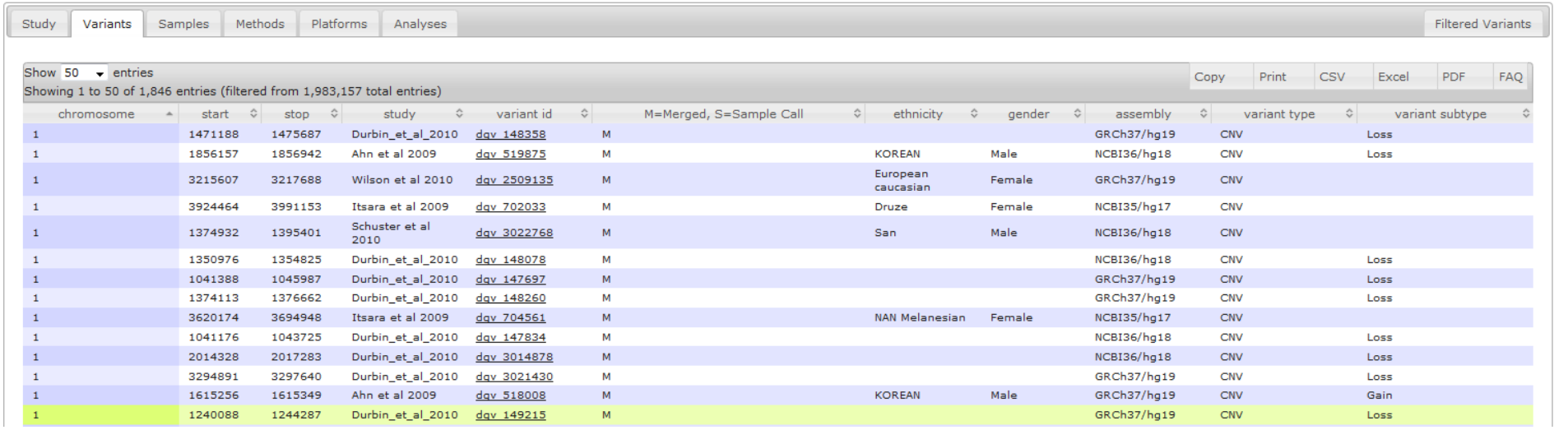
study	analysis type	analysis tool	analysis description	reference type	reference	reference description
Alkan et al 2009	Detection	mrFAST	We constructed duplication maps for each of the three genomes and estimated the absolute copy number of each duplication interval larger than 20 kb in length. We considered a given segment to be duplicated within an individual if the median estimated copy number for that individual was >2.5.	Ref_sequence	NCBI35/hg17	

Query Tool Functionality



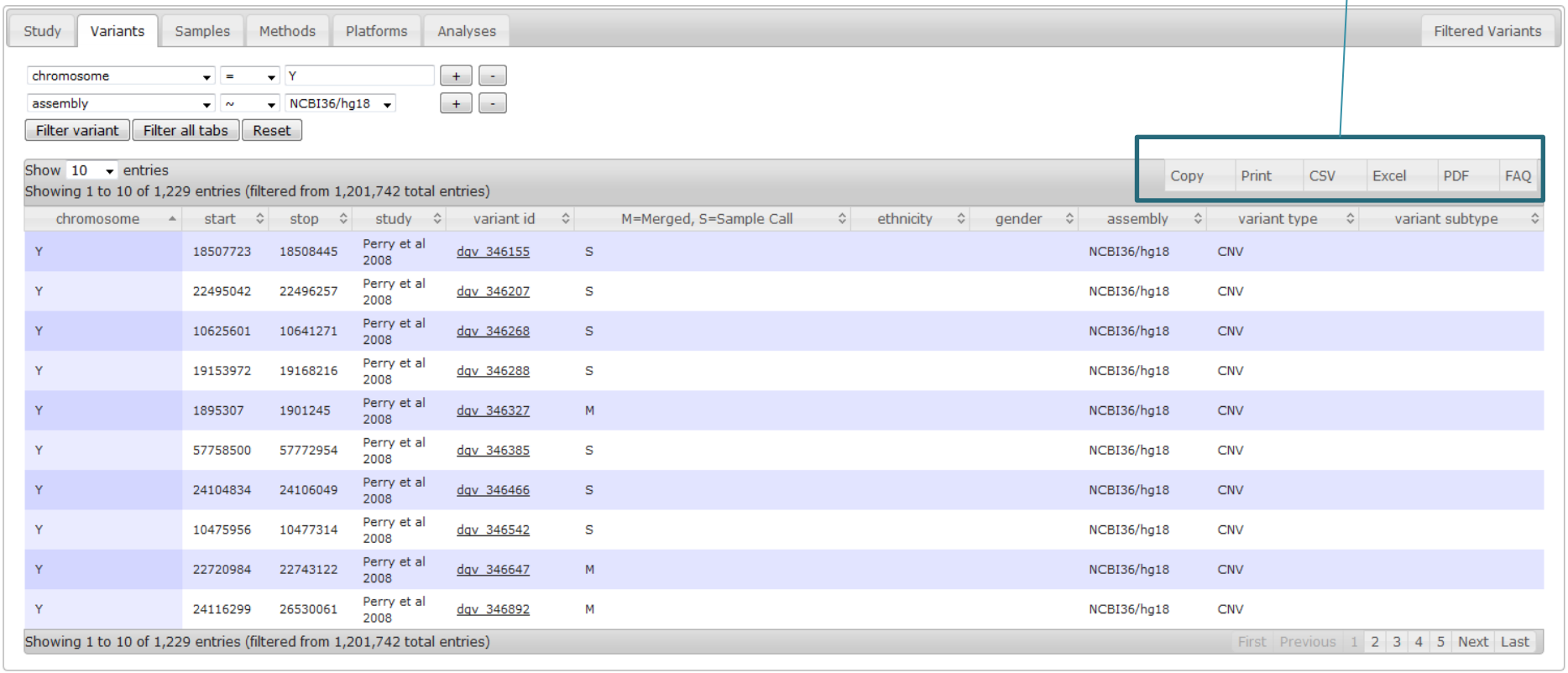


Use the "+" button to add additional terms
Use the "-" button to remove search terms



Query Tool

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



Overview

- Introduction to the *New* DGV
- Genome Browser
- Query Tool
- Examples of how to use the data
- Summary/Questions

Examples of Common Searches

•There are a few common types of searches that individuals may use to query or search the database.

- Extract only variants detected in a specific population (CEU)
- Identifying which studies have detected variants in a specific individual (NA15510)
- Find variants located in a specific region/locus of interest.
- Locate which studies and which variants are derived from HapMap samples.
- Search for variants that were detected using a sequencing based approach.

Example One
Search by sample name: NA15510

external sample id | = | NA15510 | + | -

Filter query | Reset

Study | Variants | Samples | Methods | Platforms | Analyses | Filtered Variants

Show 50 entries | Copy | Print | CSV | Excel | PDF | FAQ

Showing 1 to 8 of 8 entries (filtered from 26,060 total entries)

study	external sample id	family id	source	sample description	ethnicity	gender	cohort name
Conrad et al 2009	NA15510		DNA				
Cooper et al 2008	NA15510		DNA				
Kidd et al 2008	NA15510		DNA				
Kidd et al 2010	NA15510		Cell-culture				Fosmid9
Kidd et al 2010b	NA15510		Cell-culture				Fosmid9
Korbel et al 2007	NA15510		DNA			Female	
Mills et al 2006	NA15510		DNA				TSC
Pinto et al 2007	NA15510		DNA			Male	HapMap

Showing 1 to 8 of 8 entries (filtered from 26,060 total entries) | First | Previous | 1 | Next | Last

Select the common reference sample NA15510 using external sample ID and filter.

-Results returned include a list of all the studies that have used this sample

-Select the variants tab to get a list of all variants detected in NA15510

Example One (continued)

external sample id = NA15510 + -

Filter query Reset

Study | Variants | Samples | Methods | Platforms | Analyses | Filtered Variants

Show 50 entries

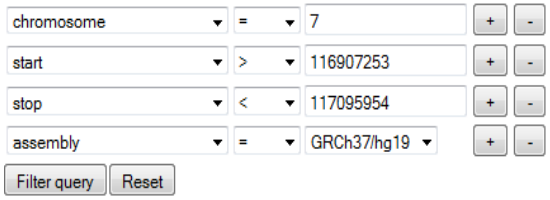
Showing 1 to 50 of 4,558 entries (filtered from 1,983,157 total entries)

Copy Print CSV Excel PDF FAQ

chromosome	start	stop	study	variant id	M=Merged, S=Sample Call	ethnicity	gender	assembly	variant type	variant subtype
1	10499	92013	Conrad et al 2009	dgv 2673064	S			GRCh37/hg19	CNV	Gain
1	444338	449236	Conrad et al 2009	dgv 2672840	S			GRCh37/hg19	CNV	Gain
1	449286	450511	Conrad et al 2009	dgv 2673356	S			GRCh37/hg19	CNV	Loss
1	869415	870347	Conrad et al 2009	dgv 2672675	S			GRCh37/hg19	CNV	Gain
1	1074379	1076019	Conrad et al 2009	dgv 2673401	S			GRCh37/hg19	CNV	Gain
1	1223594	1225599	Conrad et al 2009	dgv 2672664	S			GRCh37/hg19	CNV	Loss
1	1285400	1286900	Conrad et al 2009	dgv 2672237	S			GRCh37/hg19	CNV	Gain
1	1910377	1911917	Conrad et al 2009	dgv 2672632	S			GRCh37/hg19	CNV	Loss
1	2409767	2450161	Kidd et al 2008	dgv 372002	S			GRCh37/hg19	OTHER	Inversion

We now have a list of variants for NA15510 detected in multiple studies.

Example Two:
Search for variants in a defined region



Study | Variants | Samples | Methods | Platforms | Analyses | Filtered Variants

Show 50 entries — Copy | Print | CSV | Excel | PDF | FAQ

Showing 1 to 8 of 8 entries (filtered from 1,983,157 total entries)

chromosome	start	stop	study	variant id	M=Merged, S=Sample Call	ethnicity	gender	assembly	variant type	variant subtype
7	117018506	117063855	Kidd et al 2008	dgv 380508	S			GRCh37/hg19	CNV	
7	117035303	117035613	Mills et al 2006	dgv 888512	S			GRCh37/hg19	CNV	
7	117018506	117063855	Kidd et al 2008	dgv 367649	S	UTAH/MORMON	Female	GRCh37/hg19	CNV	Loss
7	117081446	117086131	Shaikh et al 2009	dgv 1997989	S			GRCh37/hg19	CNV	Loss
7	117081446	117086131	Shaikh et al 2009	dgv 2003874	S			GRCh37/hg19	CNV	Loss
7	116915618	116915869	Durbin_et_al_2010	dgv 2994296	S			GRCh37/hg19	CNV	Loss
7	117035303	117035613	Mills et al 2006	dgv 1110850	S			GRCh37/hg19	CNV	Loss
7	117081446	117086131	Shaikh et al 2009	dgv 2051043	M			GRCh37/hg19	CNV	

Showing 1 to 8 of 8 entries (filtered from 1,983,157 total entries) — First | Previous | 1 | Next | Last

Compare a list of variants in DGV to a region defined in your study or in a clinical array report.

-Can obtain a list of all variants across studies that fall within a defined region

-Can use this for classifying potential rare/de-novo variants, or as evidence to suggest your variant of interest may or may not be pathogenic.

(Note: it is very important to only use DGV data as a guide. The content is still limited and should not replace proper experimental controls.)

Overview

- Introduction to the *New* DGV
- Genome Browser
- Query Tool
- Examples of how to use the data
- Summary/Questions

Summary/Questions

- Launched a fully integrated, accessioned and archived database with several new studies including the 1,000 Genomes.
- Introduced a new genome browser and implemented an interactive query tool for improved access to the data.
- Improved the performance and function of the database, while maintaining the user-friendly interface and community database design.

DGV Inter-Operates With Other Genome Databases

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

DECIPHER

Ensembl

UCSC

HapMap

GeneCards

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 participate in the webinar today, your support and feedback is greatly appreciated.

Sincerely,
The DGV Team

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