

# The Database of Genomic Variants

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



FGED Toronto 2012  
Jeff MacDonald

BETA

# Database of Genomic Variants

*A curated catalogue of human genomic structural variation*

BETA

[About the Project](#)  
[Genome Browser](#)

[Downloads](#)  
[Query Tool](#)

[Links](#)  
[Submissions](#)

[Statistics](#)  
[Contact Us](#)

[FAQ](#)  
[Training Resources](#)

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:	179450	326620
Inversions:	459	1330

[Number of Studies:](#) 37

[News: April 2012 Update and Newsletter has been issued](#)

Hosted by The Centre for Applied Genomics  
[Grant support for DGV](#)  
Please read the usage [disclaimer](#)



# Large-Scale Copy Number Polymorphism in the Human Genome

Jonathan Sebat,<sup>1</sup> B. Lakshmi,<sup>1</sup> Jennifer Troge,<sup>1</sup> Joan Alexander,<sup>1</sup> Janet Young,<sup>2</sup> Pär Lundin,<sup>3</sup> Susanne Månér,<sup>3</sup> Hillary Massa,<sup>2</sup> Megan Walker,<sup>2</sup> Maoyen Chi,<sup>1</sup> Nicholas Navin,<sup>1</sup> Robert Lucito,<sup>1</sup> John Healy,<sup>1</sup> James Hicks,<sup>1</sup> Kenny Ye,<sup>4</sup> Andrew Reiner,<sup>1</sup> T. Conrad Gilliam,<sup>5</sup> Barbara Trask,<sup>2</sup> Nick Patterson,<sup>6</sup> Anders Zetterberg,<sup>3</sup> Michael Wigler<sup>1\*</sup>

www.sciencemag.org SCIENCE VOL 305 23 JULY 2004

## BRIEF COMMUNICATIONS

nature  
genetics

### Detection of large-scale variation in the human genome

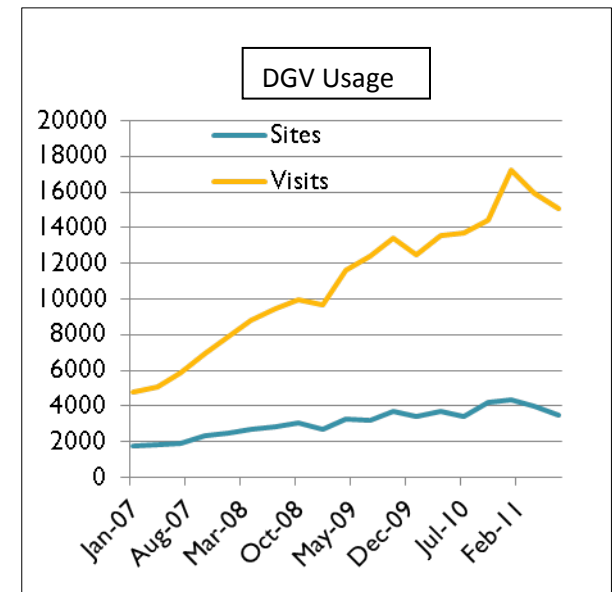
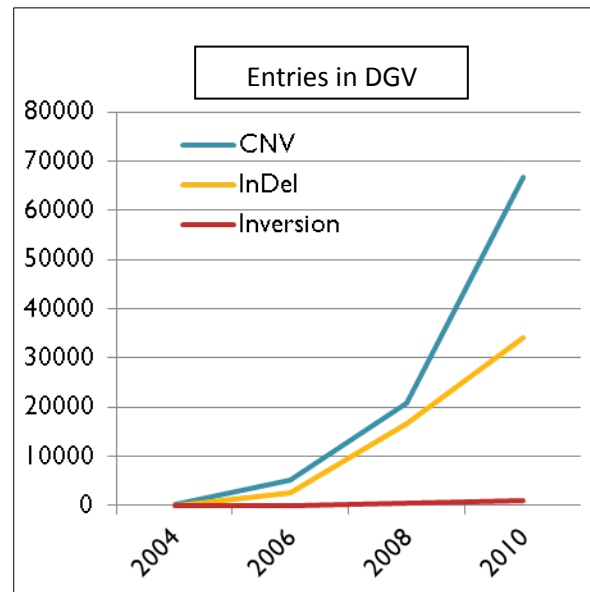
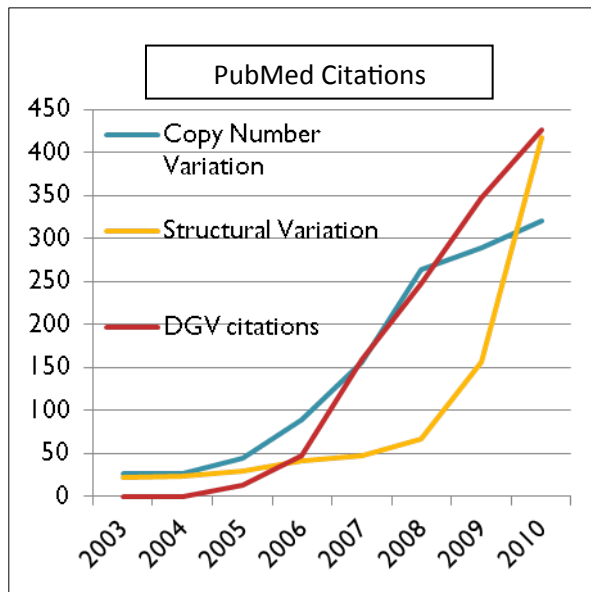
A John Iafrate<sup>1,2</sup>, Lars Feuk<sup>3</sup>, Miguel N Rivera<sup>1,2</sup>, Marc L Listewnik<sup>1</sup>, Patricia K Donahoe<sup>2,4</sup>, Ying Qi<sup>3</sup>, Stephen W Scherer<sup>3,5</sup> & Charles Lee<sup>1,2,5</sup>

identified LCVs may result from gains or losses involving as much as 2 Mb of DNA sequence (1 Mb to each flanking clone). We identified 102 LCVs (41%) that occurred in more than one individual and 24 LCVs that were present in > 10% of the individuals studied. The remaining 153 clones may represent LCVs that occur at lower frequencies. The genomic regions that we identified probably do not represent false positives, because control self-versus-self hybridization experiments

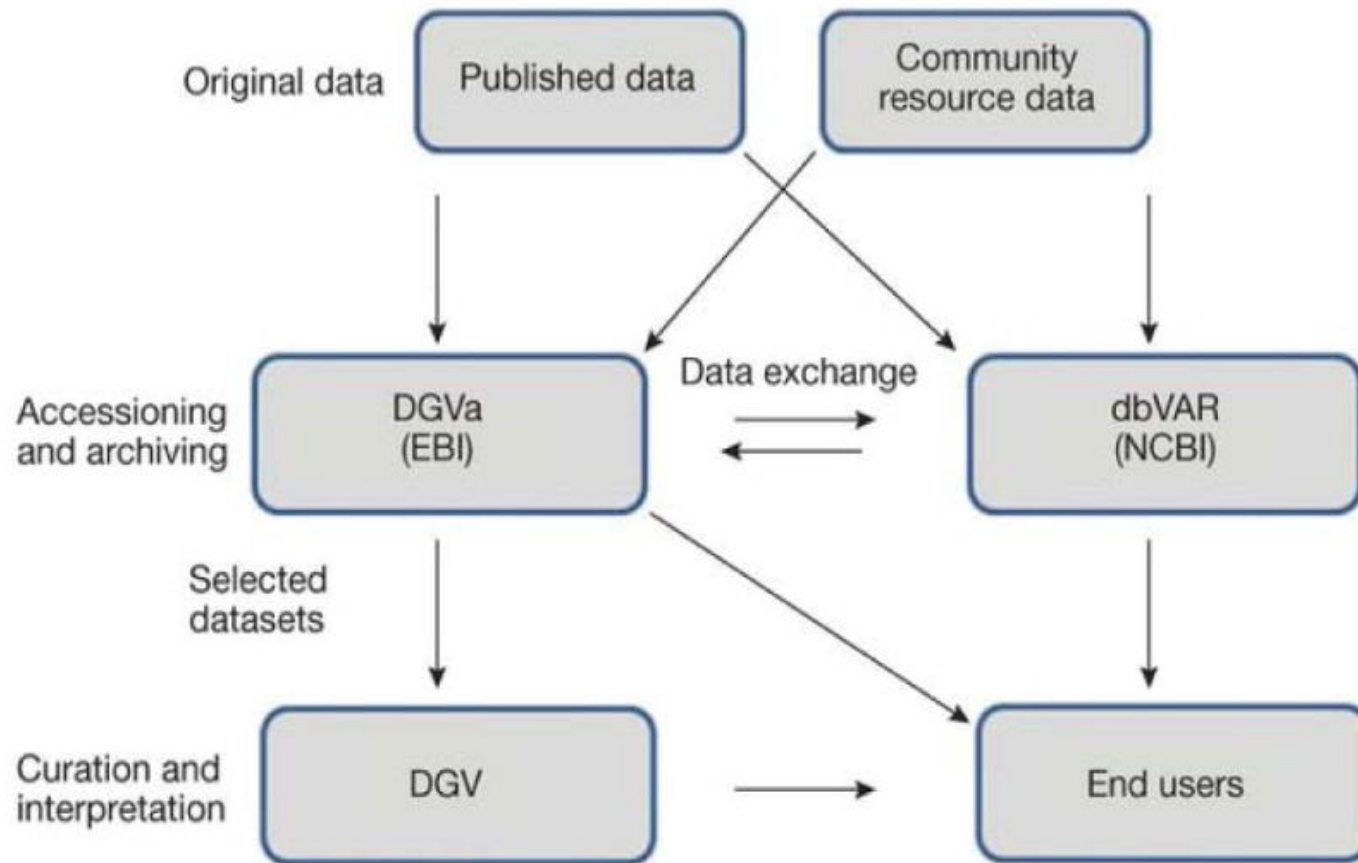
# The Database of Genomic Variants

<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 Management: NCBI JIRA Portal

The screenshot shows the NCBI JIRA Portal dashboard in a Firefox browser window. The browser tabs include 'CAZy - CE1', 'Gmail: Email from Google', and 'System Dashboard - NCBI JIRA'. The address bar shows the URL 'https://ncbijira.ncbi.nlm.nih.gov/secure/Dashboard.jspa'. The page header features the NCBI logo, the user name 'Jeff MacDonald', and a 'Quick Search' field. Below the header is a navigation menu with 'Dashboards', 'Projects', 'Issues', and 'Administration'. The main content area is divided into several sections:

- Projects:** A list of projects with their respective leads. Each project entry includes a project name and a lead name, with icons for details and filters.
- Favorite Filters:** A section indicating 'You have no favorite filters at the moment' with links for 'Create Filter' and 'Manage Filters'.
- Assigned to Me:** A section indicating 'No matching issues found'.
- Issues in progress:** A section indicating 'No matching issues found'.
- Quick Links:** A section with links for 'My Unresolved Reported Issues | Votes | Watches', 'Browse Projects | Search for Issues | Create Issue'.

The Windows taskbar at the bottom shows the system tray with the time '14:55' and date '2012-01-20'.

# Study Tracking: NCBI JIRA Portal

The screenshot shows the NCBI JIRA Portal for the 'Structural Variation Load' project. The browser tabs include 'CAZy - CE1', 'Gmail: Email from Google', and 'Structural Variation Load - NCBI JIRA'. The URL is 'https://ncbijira.ncbi.nlm.nih.gov/browse/VLOAD'. The page header shows the NCBI logo, navigation menus (Dashboards, Projects, Issues, Administration), and a user profile for 'Jeff MacDonald'. The main content area is titled 'Structural Variation Load' and includes a 'Create Issue' button and options for 'Task' and 'Discussion'. A left sidebar contains a navigation menu with 'Summary' selected, along with 'Issues', 'Road Map', 'Change Log', 'Popular Issues', 'Versions', and 'Components'. The 'Summary' section includes a 'Description' (This project will be used by dbVar and DGVA to jointly manage loading and exchange tasks.), 'Lead: Church, Deanna', and 'Key: VLOAD'. Below this is an 'Issues: Due' section listing three issues: VLOAD-76 (Submitter curated dataset), VLOAD-148 (Hou et al 2011), and VLOAD-149 (Hou et al 2011b). An 'Issues: 30 Day Summary' section features a bar chart showing the number of issues due over a 30-day period, with a peak of 3 issues on the final day. The right side of the page displays 'Versions: Due' with three items: 'Data Cleanup ready to load DGVA', 'Aug 2011 Public Data Release' (Release Date: Aug/25/11), and 'Submission Template Updates' (Release Date: Sep/01/11). An 'Activity Stream' section shows comments from John Lopez and Tim Hefferon on issue VLOAD-158, dated January 18, 2012.

Firefox CAZy - CE1 Gmail: Email from Google Structural Variation Load - NCBI JIRA

https://ncbijira.ncbi.nlm.nih.gov/browse/VLOAD

NCBI Jeff MacDonald Quick Search

Dashboards Projects Issues Administration Create Issue

Structural Variation Load Create: Task Discussion

Summary Issues Road Map Change Log Popular Issues Versions Components

**Summary**

Description

This project will be used by dbVar and DGVA to jointly manage loading and exchange tasks.

Lead: Church, Deanna

Key: VLOAD

Issues: Due

- VLOAD-76 Submitter curated dataset
- VLOAD-148 Hou et al 2011
- VLOAD-149 Hou et al 2011b

Issues: 30 Day Summary

3 Public Data Release

2

1

Versions: Due

- Data Cleanup ready to load DGVA
- Aug 2011 Public Data Release Release Date: Aug/25/11
- Submission Template Updates Release Date: Sep/01/11

Activity Stream

January 18 - 1:00 PM

- John Lopez commented on VLOAD-158 (Lopez-Herrera et al 2012) saying: I reloaded the files. The corrected placements can be seen here: [http://dev.ncbi.nlm.nih.gov/dbvarvariants/nsv209111/?db=DBVAR\\_LOAD:dbVarDev#tab-2](http://dev.ncbi.nlm.nih.gov/dbvarvariants/nsv209111/?db=DBVAR_LOAD:dbVarDev#tab-2)
- Tim Hefferon commented on VLOAD-158 (Lopez-Herrera et al 2012) saying: Hi John L,  
Thanks - I contacted the submitter, who corrected the coordinates to start="152111739", stop="152222852".  
The submission has been updated, and XML regenerated.  
Tim

January 18 - 12:00 PM

- John Lopez commented on VLOAD-158 (Lopez-Herrera et al 2012) saying: Hi Tim,  
Can you please confirm the start and stop coordinates?  
I think something is wrong.

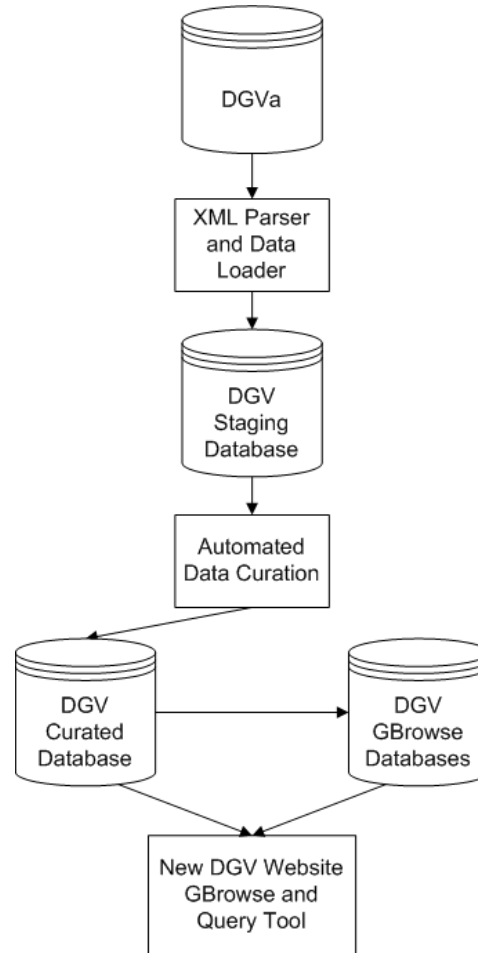
14:57 2012-01-20

# Automation, Standardization and Integration

- Developed tools and pipelines to import, format and load studies into DGV
- Shared data structure, vocabulary and terminology between all three sites
- Interactive links between each database (DGV, DGVa and dbVAR)



# Data Flow Diagram

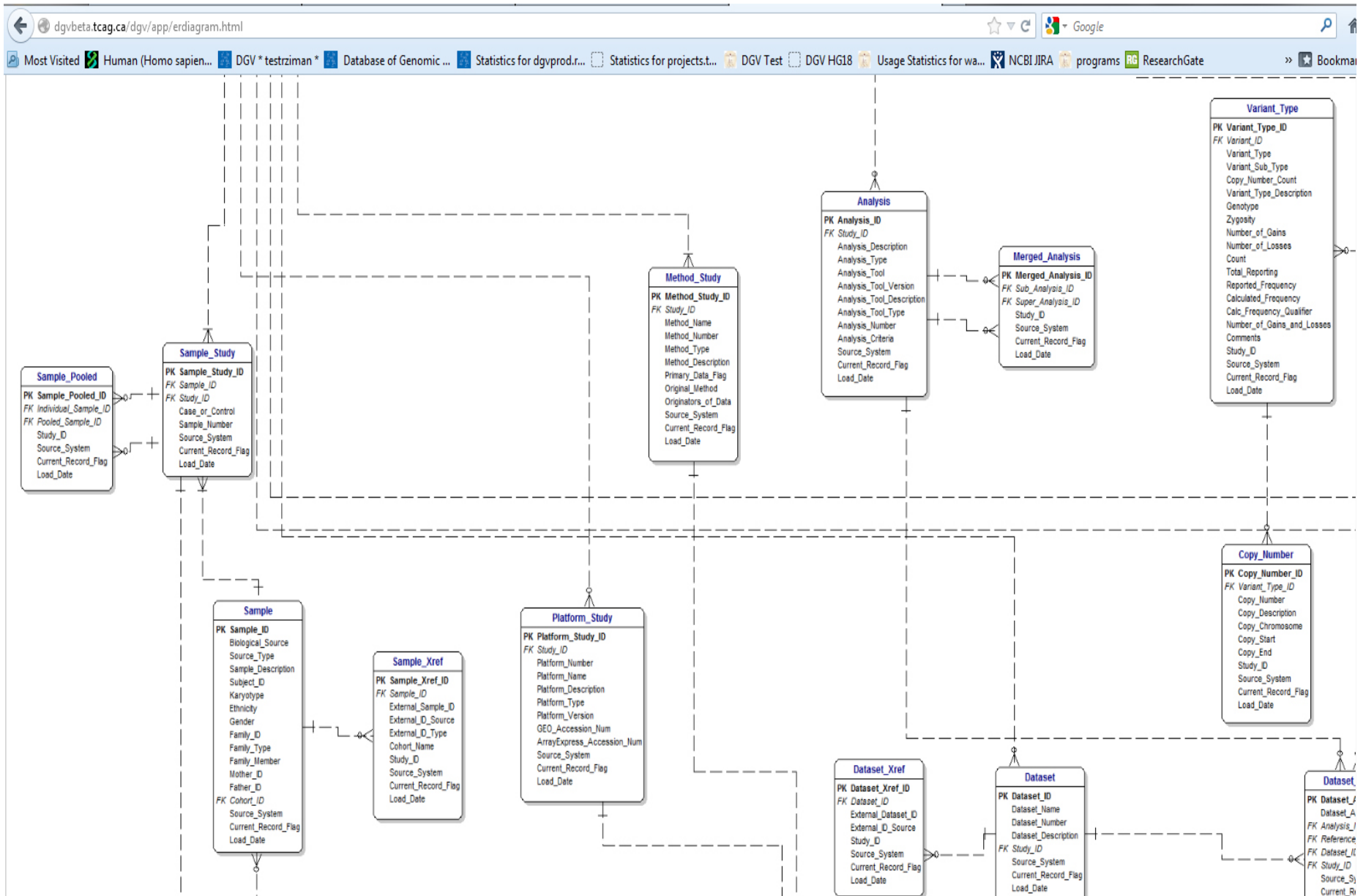


# Standardized Submission Template

	A	B	C	D
1	<b>STUDY</b>			
2	<b>FIELD</b>	<b>TO BE COMPLETED BY SUBMITTER</b>	<b>NOTES</b>	
3	<b>study_id</b>		A study identifier in the format <b>AuthorYear</b> where <b>Author</b> is the last name of the first author and <b>Year</b> is the year (or anticipated year) of publication, e.g. <b>Smith2012</b> .	
4	<b>study_description</b>		A description (<700 words) of the study. If the study has been published, you may use the abstract.	
5	<b>study_alias</b>		A comma-delimited list of alternative names by which the study will be indexed, e.g. 1000 Genomes, Thousand Genomes.	
6	<b>study_type</b>		Choose a term from the list: Case Control = comparison between healthy and diseased samples, Case Set = disease-associated variation, Control Set = variation within normal samples, Curated = variation compiled from the literature or online resources, Tumour vs. Matched Normal = variation between tumour and healthy samples from the same individual.	
7	<b>taxonomy_id</b>		The NCBI taxonomy identifier(s), as a comma-delimited list, for the species studied. To find a taxonomy ID, go to <a href="http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy">http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy</a> , click the link for the species and look for "Taxonomy ID=".	
8	<b>project_id</b>		The European Bioinformatics Institute project identifier.	
9	<b>pubmed_id</b>		** REQUIRED if the study has been published. The NCBI Pubmed ID(s), as a comma-delimited list.	
10	<b>study_url</b>		If the study has a website, provide the URL.	
11	<b>EGA_id</b>		If the study has been submitted to the European Genome-phenome Archive (EGA), provide the EGA study accession number.	
12	<b>hold_date</b>		If you wish to delay public release of the study data, enter a date for public release, in the format YYYY-MM-DD.	
13				
14				
15				
16				
17				



# DGV Data Model



# Data Access; download and archive

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](#) [Training Resources](#)

## 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 non-stable identifier which will change each time the database is updated. Each variant is also provided with an accession which is a stable identifier and will remain constant. 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)
2012-03-29	<a href="#">link</a>	<a href="#">link</a>	<a href="#">link</a>

## Filtered Variants

Release Date	GFF3 File
2012-03-29	<a href="#">link</a>

## 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.

variantid	chr	start	end	varianttype	variantsubtype	reference	pubmedid	method	platform	mergeid	frequency	samplesize	cohortdescription	genes	
4322	20	206901	209000	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several populations that		
4323	20	60781151	60784750	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4324	20	1500351	1505850	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II	6764		6	Samples from several populations that		
4325	20	54607001	54608950	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4326	20	51764151	51767850	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4327	20	48979501	48982600	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4328	20	56643601	56645700	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4329	20	5736451	5740500	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II	6659		6	Samples from several populations that		
4330	20	15088051	15091650	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4331	20	25988951	25993150	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4332	20	54288401	54291450	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4333	20	47617276	47619400	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4334	20	40562751	40566750	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4335	20	26205451	26219350	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			2699971	6	Samples from several	
4336	20	22370651	22373550	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4337	20	62194501	62198300	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4338	20	53239451	53242950	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4339	20	11422701	11424750	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		
4340	20	637701	640400	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several populations that		
4341	20	59846651	59849650	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			2700003	6	Samples from several	
4342	20	25737951	25774950	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6615	6	Samples from several	
4343	20	26012151	26020750	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			2699966	6	Samples from several	
4344	20	1693801	1696000	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several populations that		
4345	20	19024251	19026550	CNV	Loss	Durbin_et_al_2010	20981092	Sequencing	Illumina Genome Analyzer II			6	Samples from several		

# Interactive Access to Data: Query Tool

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](#)

assembly = NCBI36/hg18 + -

Study Variants Samples Methods Platforms Analyses Filtered Variants

Show 50 entries  
Showing 1 to 35 of 35 entries (filtered from 37 total entries)

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

# Data Extract: Query Tool

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

The screenshot shows a web-based query tool interface. At the top, there are tabs for 'Study', 'Variants', 'Samples', 'Methods', 'Platforms', and 'Analyses'. Below the tabs are filter fields: 'chromosome' set to 'Y', 'assembly' set to 'NCBI36/hg18', and buttons for '+', '-', 'Filter variant', 'Filter all tabs', and 'Reset'. A 'Filtered Variants' button is in the top right. Below the filters, it says 'Show 10 entries' and 'Showing 1 to 10 of 1,229 entries (filtered from 1,201,742 total entries)'. A menu with options 'Copy', 'Print', 'CSV', 'Excel', 'PDF', and 'FAQ' is highlighted with a blue box. Below the menu is a table of variant data. At the bottom, it says 'Showing 1 to 10 of 1,229 entries (filtered from 1,201,742 total entries)' and has pagination controls: 'First', 'Previous', '1', '2', '3', '4', '5', 'Next', 'Last'.

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	<a href="#">dqv_346155</a>	S			NCBI36/hg18	CNV	
Y	22495042	22496257	Perry et al 2008	<a href="#">dqv_346207</a>	S			NCBI36/hg18	CNV	
Y	10625601	10641271	Perry et al 2008	<a href="#">dqv_346268</a>	S			NCBI36/hg18	CNV	
Y	19153972	19168216	Perry et al 2008	<a href="#">dqv_346288</a>	S			NCBI36/hg18	CNV	
Y	1895307	1901245	Perry et al 2008	<a href="#">dqv_346327</a>	M			NCBI36/hg18	CNV	
Y	57758500	57772954	Perry et al 2008	<a href="#">dqv_346385</a>	S			NCBI36/hg18	CNV	
Y	24104834	24106049	Perry et al 2008	<a href="#">dqv_346466</a>	S			NCBI36/hg18	CNV	
Y	10475956	10477314	Perry et al 2008	<a href="#">dqv_346542</a>	S			NCBI36/hg18	CNV	
Y	22720984	22743122	Perry et al 2008	<a href="#">dqv_346647</a>	M			NCBI36/hg18	CNV	
Y	24116299	26530061	Perry et al 2008	<a href="#">dqv_346892</a>	M			NCBI36/hg18	CNV	

# 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

<b>Breakpoints</b> <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 [?]	
<b>CGH Arrays</b> <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 [?]	
<b>Chromosome</b> <input type="checkbox"/> All on <input type="checkbox"/> All off		
<input checked="" type="checkbox"/> Cytogenetic Bands [?]	<input type="checkbox"/> Assembly [?]	<input type="checkbox"/> Gap [?]
<b>Clones</b> <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 [?]
<b>Disease</b> <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 [?]	
<b>Gene</b> <input type="checkbox"/> All on <input type="checkbox"/> All off		
<input checked="" type="checkbox"/> RefSeq Genes [?]	<input type="checkbox"/> mRNA [?]	<input type="checkbox"/> microRNA [?]
<b>General</b> <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 [?]	
<b>SNP Arrays</b> <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 [?]	
<b>Segmental Duplications</b> <input type="checkbox"/> All on <input type="checkbox"/> All off		
<input type="checkbox"/> WSSD duplications [?]	<input checked="" type="checkbox"/> UCSC segmental duplications [?]	
<b>Study Variants</b> <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) [?]



# Future Directions and Challenges

- How to improve submission process to increase number of entries
  - work/interact with journals to push for mandatory submission
- Should we expand and include cases/patient data in DGV
  - privacy/confidentiality issues
- How to handle/display a large number of future personal genomes data.
  - what data do we want to include, which are better suited for dbSNP(?)
  - need to optimize data acquisition and processing pipelines
- Generating a DGV curated Gold Standard Set
  - based on feedback from community, each individual may have different requirements; is there a one size fits all solution?