Today’s Attendance URL and QR code:

https://tinyurl.com/AttendanceFeb27

Reminder: Questions

- Type your questions into the “Q&A” box
- Monitored by committee members
  - Will verbally answer as many as time allows (at the end)
  - May receive a typed answer through the Q&A box from a committee member
  - Unanswered questions will be saved for consideration during the March 12 Q&A webinar

ClinGen CNV Technical Standards Web Series

Resources for Population CNV Data: Database of Genomic Variants (DGV)

http://dgv.tcag.ca

About the Database

- The objective of the Database of Genomic Variants is to provide a comprehensive summary of structural variation in the human genome. We define structural variation as genomic alterations that involve segments of DNA that are larger than 50bp. The content of the database only represents structural variation identified in healthy control samples.
- The Database of Genomic Variants provides a useful catalog of control data for studies aiming to correlate genomic variation with phenotypic data. The database is updated with new data from peer reviewed research studies.

The specific aims of DGV include:

- 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 database has been recently updated and currently contains information from 75 different peer reviewed studies.

There are approximately ~ 980,000 CNV regions and over 4,000 inversions, identified in greater than 47,000 unique individuals.

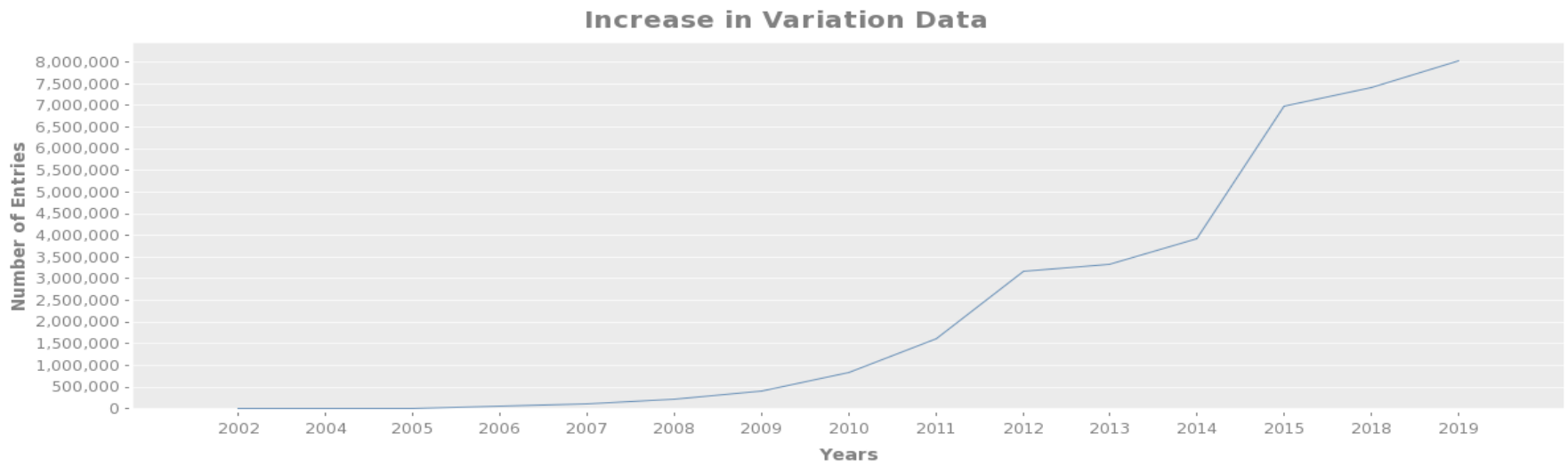
The majority of data sets in DGV are from four different types of studies:

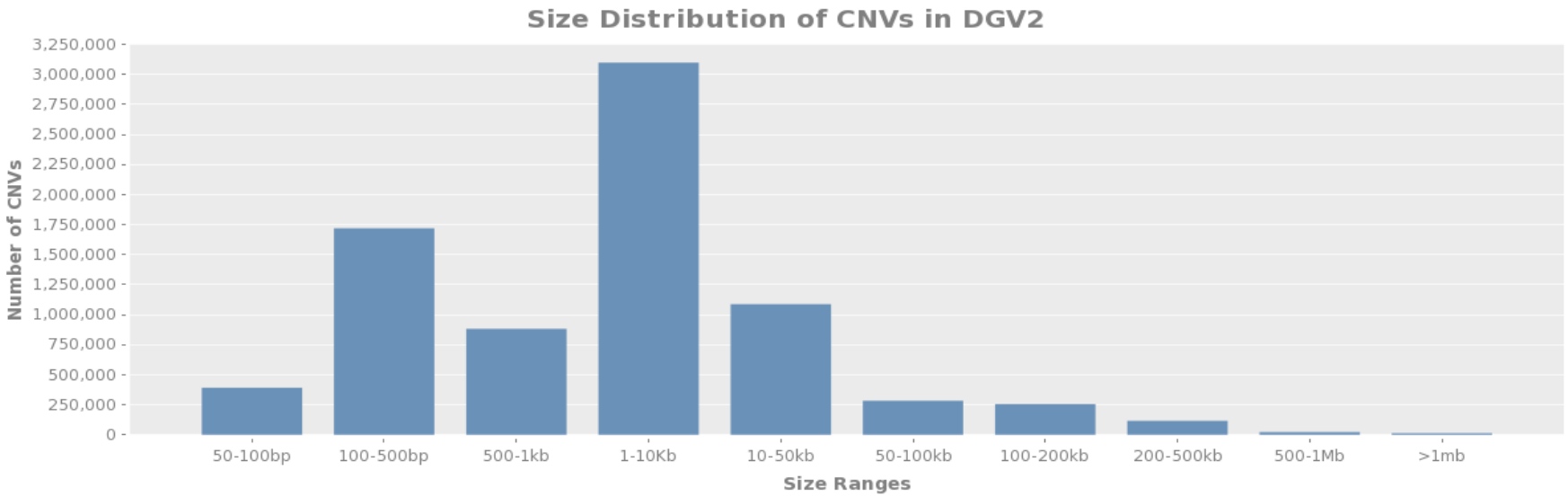
- Array based comparative genomic hybridization (CGH) and comparative intensity analysis (SNP/CNV arrays)
- Whole genome sequencing
- Clone end sequence mapping
- Sequence trace mapping

Of note:

- The database is still limited in content
- Approximately 47,000 individuals are represented, and these are 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

Database Update





Update Summary:

- GRCh38 update for DGV Gold Standard available
- 3 new studies added
  - (nstd166,nstd162,estd217)
- > 400,000 new CNVs
- > 10,000 new unique samples

Database of Genomic Variants (http://dgv.tcag.ca)

Database of Genomic Variants

A curated catalogue of human genomic structural variation

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

Keyword, Landmark or Region Search: Search GRCh37/hg19

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

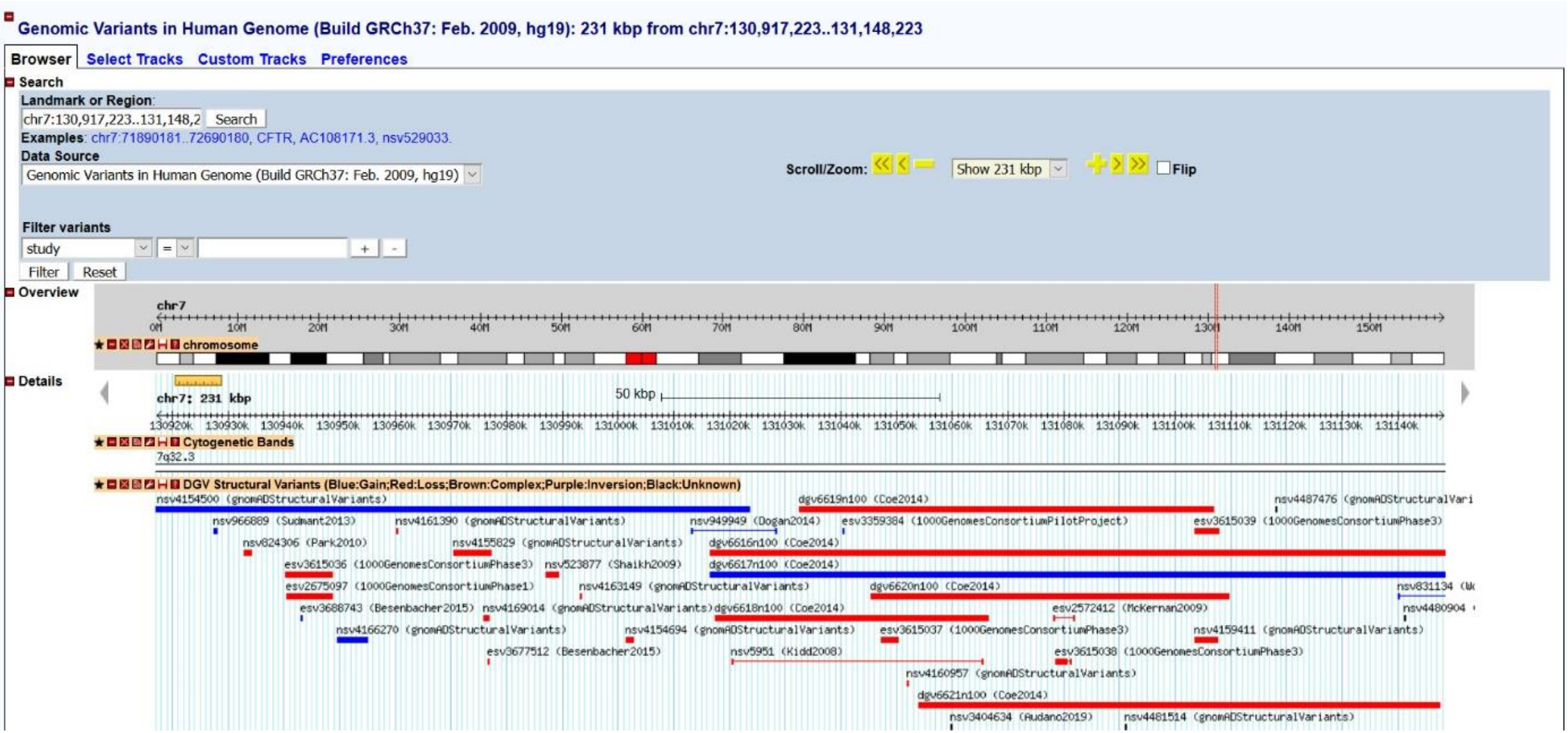
Stat	Merged-level	Sample-level
CNVs:	983845	7021692
Inversions:	4083	32044

Number of Studies: 75

News: May 2016 Update and Newsletter has been issued

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

Database of Genomic Variants: Genome Browser



Database of Genomic Variants: Table Browser

Database of Genomic Variants

A curated catalogue of human genomic structural variation

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

accession = + -

Filter query | Reset

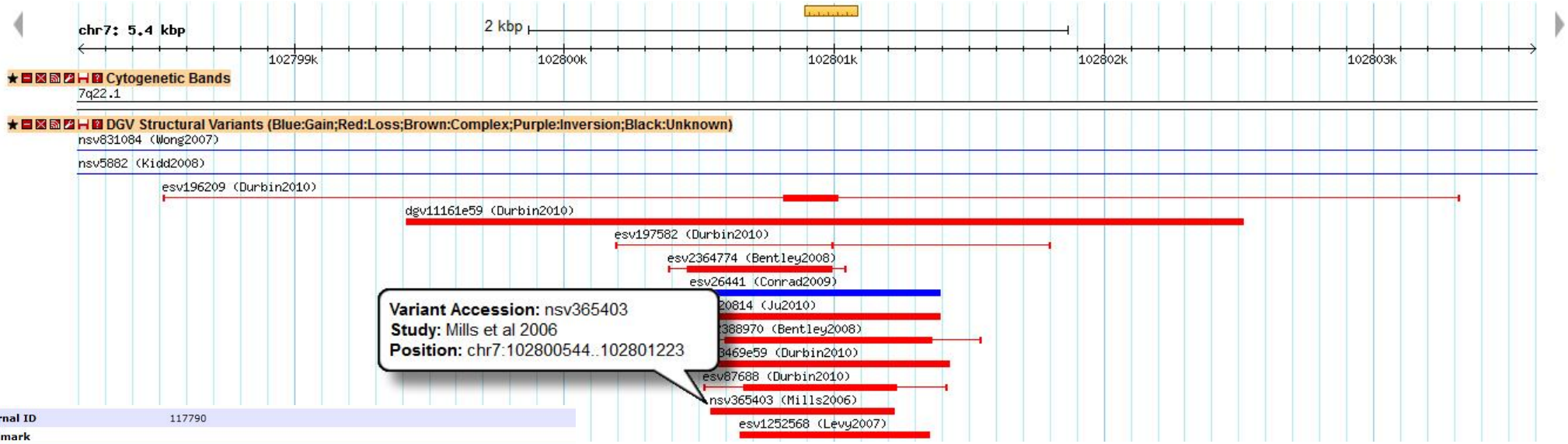
Study | Variants | Samples | Methods | Platforms | Analyses

Show 50 entries
Showing 1 to 50 of 75 entries

Copy | Print | CSV | Excel | PDF | FAQ

accession	study	pubmed id	sample size	variant count
estd1	Redon_et_al_2006	17122850	270	26086
estd176	Banerjee_et_al_2011	21479260	1250	1813
estd180	Pang_et_al_2010	20482838	3	12614
estd19	Ahn_et_al_2009	19470904	1	8590
estd193	Feuk_et_al_2005	16254605	18	6
estd194	Bentley_et_al_2008	18987734	1	11406
estd195	Altshuler_et_al_2010	20811451	1184	159763
estd196	Simon-Sanchez_et_al_2007	17116639	181	638
estd197	McKernan_et_al_2009	19546169	1	13926
estd198	Chia_et_al_2012	23635498	64	780
estd199	1000_Genomes_Consortium_Phase_1	23128226	1151	1225440
estd20	Conrad_et_al_2009	19812545	40	71932
estd201	Wong_et_al_2012b	23290073	96	330655
estd203	Vogler_et_al_2010	21179565	1109	46781
estd209	Pang_et_al_2013b	24192839	1	3470
estd21	Wheeler_et_al_2008	18421352	3	46

DGV Content



Internal ID	117790			
Landmark				
Location Information	Type	Coordinates	Assembly	Other Links
		chr12:109632315..109632415	hg19	UCSC Ensembl
		chr12:108116698..108116798	hg18	UCSC Ensembl
Cytoband	12q24.11			
Allele length	Assembly	Allele length		
	hg19	101		
	hg18	101		
Variant Type	CNV Deletion			
Copy Number				
Allele State	Homozygous			
Allele Origin	Not tested			
Probe Count				
Merged Status	S			
Merged Variants	esv1456600			
Supporting Variants				
Samples	HuRef			
Known Genes	ACACB			
Method	SNP_array			
Analysis	analysis_type=Sequence Alignment			
Comments				
Reference	Levy et al 2007			
Pubmed ID	17803354			
Accession Number(s)	essv3652338			
Frequency	Sample Size	2		
	Observed Gain	0		
	Observed Loss	1		
	Observed Complex	n/a		
	Frequency	n/a		

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

Additional details such as the allele length, number of gains/losses and the allele state (heterozygous/homozygous) have been added (when available).

DGV Tracks

- DGV Structural Variants
  - Contains all the top level/merged variants (dgv or esv/nsv entries, corresponding to those merged by DGV or the non-overlapping nsv/esv variants from the original study.)

- Supporting Variants
  - Contains the underlying supporting or sample level variants (are summarized and represented in the merged or sv variant regions)

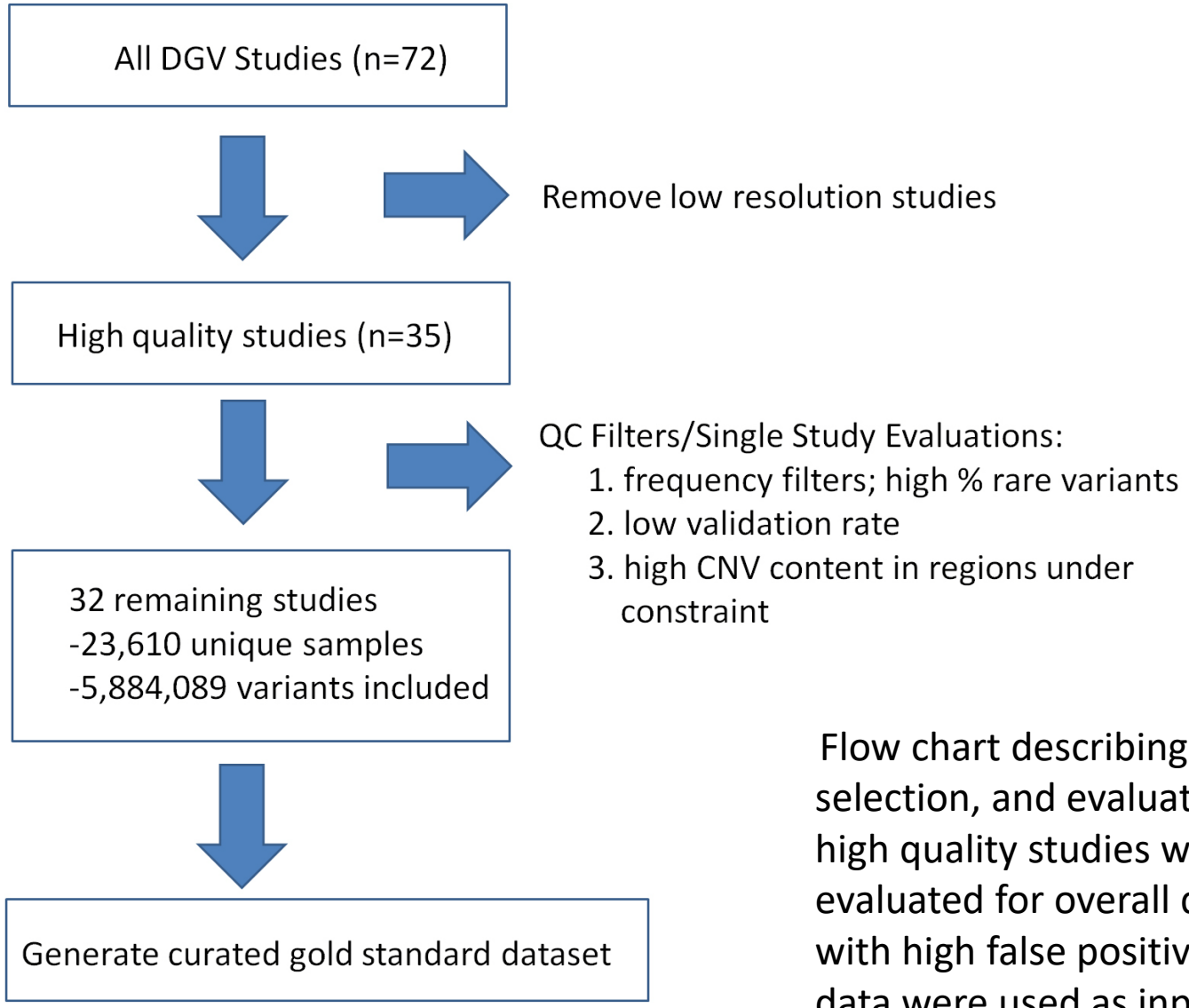
- DGV Gold Standard Variants
  - This track contains a curated set of variants from a select number of studies in DGV. High resolution and high quality studies were evaluated for accuracy and sensitivity and included if they passed strict quality filters. Overlapping sample level variants were merged to create variant regions if they shared at least 50% of their length in a reciprocal manner. CNV regions were kept if there was sufficient support from multiple studies and samples indicating a high quality variant region. The initial criteria required that a variant is found in at least 2 different studies and found in at least two different samples. Information about the underlying samples, ethnicity and population distribution are recorded. This dataset will be dynamic and will continue to be updated to reflect new additions to the database.

DGV Gold Standard

- Increase Sensitivity:
  - leverage large unique cohorts across many studies
  - combine complimentary approaches to better define the underlying variant
  - improve the power to detect variants across diverse genomic regions and size ranges compensating for individual study biases and shortcomings
- Increase Specificity:
  - reduce the false positive rate using support from multiple samples and platforms
  - currently variants in DGV span ~85% of the genome
  - improve resolution and accuracy of variant breakpoints
  - more accurate frequency estimates and broader population representation for each variant region.

A copy number variation map of the human genome. Zarrei M, MacDonald JR, Merico D, Scherer SW. Nat Rev Genet. 2015 Mar;16(3):172-83.PMID:25645873

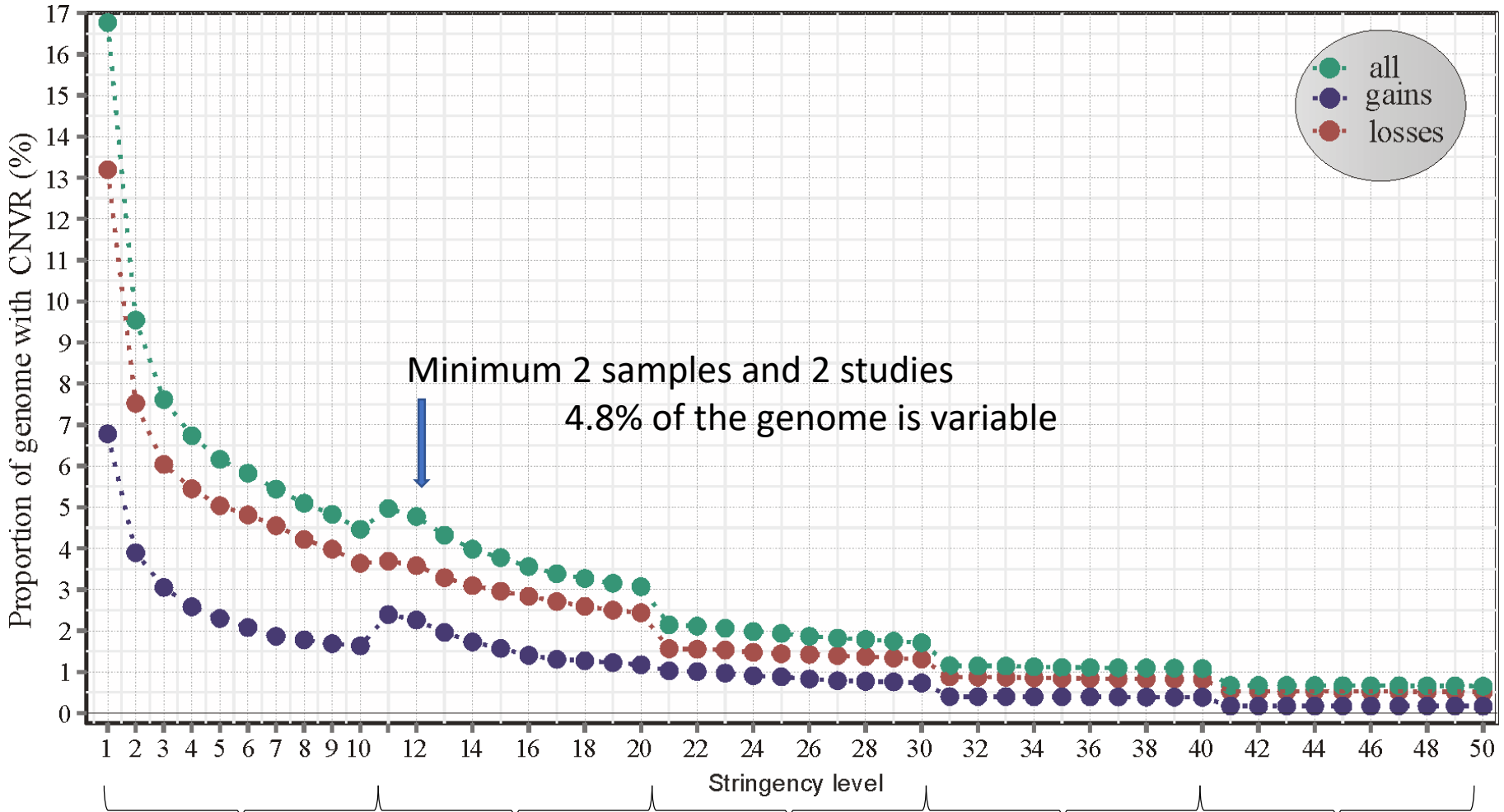
DGV Gold Standard Workflow



Flow chart describing the process for study selection, and evaluation. All high resolution, high quality studies were selected, and evaluated for overall quality, removing those with high false positive rates. The resulting data were used as input to generate the DGV Gold Standard resource.

Evaluating Thresholds

% of genome that is variable at various stringency levels



Zarrei M, MacDonald JR, Merico D, Scherer SW. Nat Rev Genet. 2015 Mar;16(3):172-83.PMID:25645873

Summary Statistics DGV Gold Standard

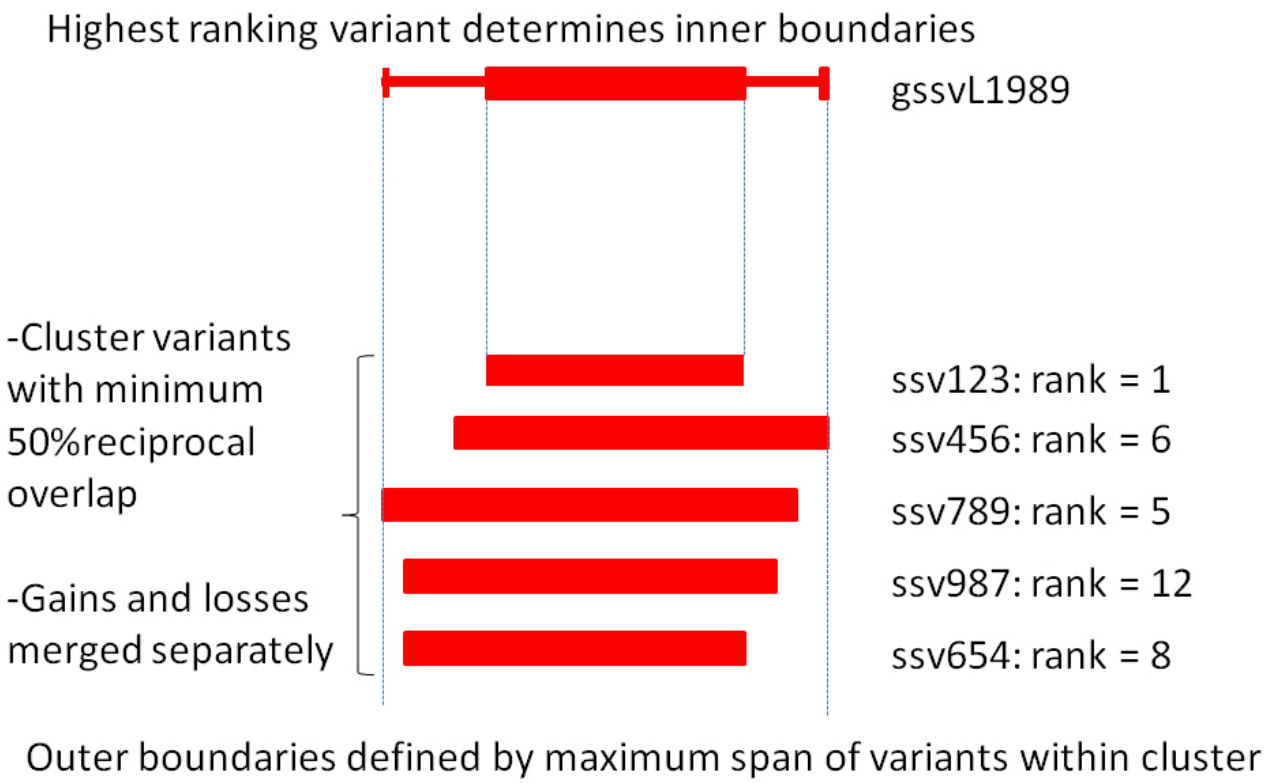
Gold Standard Statistics	# Entries
Studies	32
Variants	5,844,089
Gold Standard Regions	38,185
Gains	6,209
Losses	31,976
Unique Samples	23,610
% Genome/CNV	17.04%

Study Name	Study Accession	Platform/Method
1000 Genomes Pilot Phase	etd59	Illumina_GA
1000GenomesPhase1	estd199	Illumina_II_IIX_HiSeq
1000GenomesPhase3	estd214	Multiple_NGS_Sequencing
Ahn2009	estd199	Illumina_GA
Alkan2009	nstd31	Illumina_GA+454
Altshuler2010	estd195	Affymetrix6.0_Illumina1M
Coe2014	nstd100	Affymetrix_6.0
Arlt2010	nstd50	Illumina_GA + IlluminaOmni1-Quad
Bentley2008	estd194	Illumina_GA
Boomsma2014	estd215	Illumina_HiSeq
Campbell2011	nstd46	AgilentCustom180K_v2.1+v3.0
Conrad2009	estd20	NimbleGen42M
Cooper2008	nstd14	Illumina1M
Cooper2011	nstd54	Custom_Illumina_1.2M
Dogan2014	nstd73	Illumina_HiSeq
Ju2010	nstd65	Illumina_GA
Kidd2010b	nstd47	SangerSequencing
Kim2009	nstd43	Agilent24M
Levy2007	estd22	SangerSequencing
McCarroll2008	nstd22	Affymetrix6.0
McKernan2009	estd197	SOLiD
Mills2006	nstd6	SangerSequencing
Mokhtar2014	estd213	Affymetrix6.0
Pang2010	estd180	Affymetrix6.0 + NimbleGen42M + Agilent24M + Illumina1M
Park2010	nstd67	Agilent24M
Perry2008	nstd4	AgilentCustom_015685+015686+244K
Schuster2010	nstd39	Illumina_GA+454
Sudmant2013	nstd82	Illumina_HiSeq
Uddin2014	estd212	CytoScanHD_2.7M
Vogler2010	estd203	Affymetrix6.0
Wang2008	estd3	Illumina_GA
Wong2012b	estd201	Illumina_GA

DGV Gold Standard Input

- All DGV datasets were evaluated using a number of criteria to identify those with a low false positive rate (see Zarrei et al, Nature Reviews Genetics 2015, PMID: 25645873 for details)
- To improve breakpoint accuracy, datasets were selected based on the approach used to identify SV's.
  - Sequencing based (NGS or sanger)
  - Array based with platforms that had a minimum of 1M probes
  - Custom, focused arrays with local, high density coverage)

DGV Gold Standard Clustering and Breakpoint Assignment



Overview of the process for clustering variants together to generate a gold standard variant record. Each combination of platform and algorithm was assigned a rank to measure breakpoint capture accuracy. The highest ranking variant in the cluster is used to define the inner start and stop coordinates. Outer coordinates are defined by the maximum span of the remaining entries.

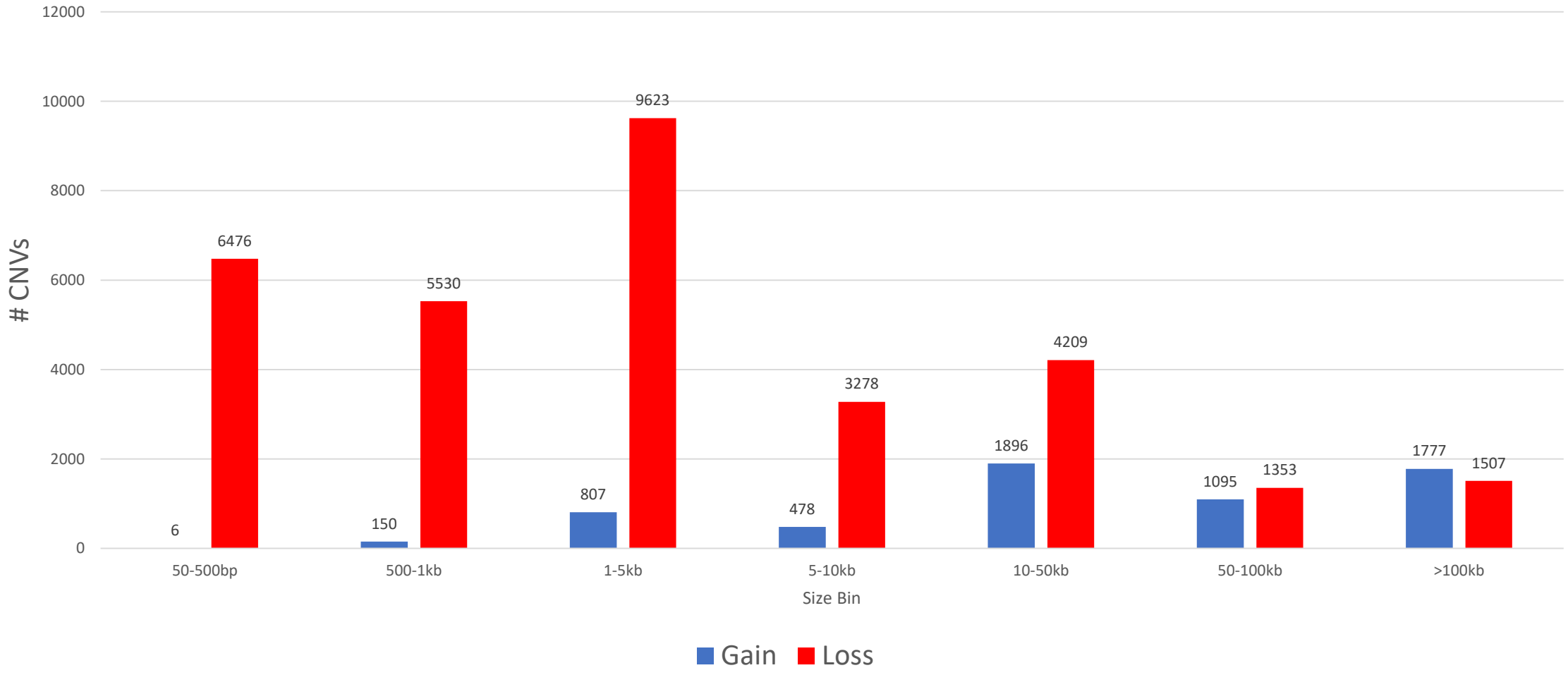
Study Accession	Study Name	Platform	Algorithm	Rank	SampleCount	SamplePopulations	VariantCount
estd20	Conrad2009	NimbleGen 42M	GADA	7	40	African European	51600
estd59	1000GenomesPilot	Illumina II	Mobile Insertion Mosaik (PEM) analysis_num=	9	144	African Asian European	7531
estd59	1000GenomesPilot	454 GS FLX	Mobile Insertion Mosaik (SplitRead) analysis_	4	-	-	-
estd59	1000GenomesPilot	Illumina GA II + 454 GS FLX	Mobile Insertion Mosaik (PEM) analysis_num=	9	-	-	-
estd59	1000GenomesPilot	Illumina II	TandemDuplication Spanner (PEM) analysis_n	9	-	-	-
estd195	Altshuler_et_al_2010	Affymetrix SNP 6.0 + Illumina Human1Mv1	pooled raw data; QuantiSNP Birdseye	15	1184	African Asian European Mexican Native American	157507
nstd82	Sudmant2013	Illumina_HiSeq	ReadDepth	8	10	African Asian European Oceania South American	121639
estd203	Vogler2010	Affymetrix6.0	PennCNV+Birdseye	16	1109	African European	42342
estd199	1000GenomesPhase 1	Illumina II, IIX, HiSeq	PEM; Breakdancer analysis_num=1	9	1055	African Asian European Mexican South American	1202915
estd199	1000GenomesPhase 1		ReadDepth; CNVnator analysis_num=2	8	-	-	-
estd199	1000GenomesPhase 1		PEM; DELLY analysis_num=3	9	-	-	-
estd199	1000GenomesPhase 1		ReadDepth+PairedEnd; GenomeStrip analysis_	8	-	-	-
estd199	1000GenomesPhase 1		SplitRead; Pindel analysis_num=5	5	-	-	-
nstd100	Coe2014(ARIC)	Affymetrix_6.0	Custom_HMM	16	8,733	European	213345
nstd100	Coe2014(WTCCC2)	Affymetrix_6.0	Custom_HMM	16	1500	European	62448
nstd54	Cooper2011	Custom_Illumina_1.2M	Custom_HMM	14	1500	European	405414
estd212	Uddin2014	CytoScanHD_2.7M	ChAS	13	873	African Asian European Admixed	71178
estd215	Boomsma2014	Illumina_HiSeq	GATK+BreakDancer+Pindel+GenomeSTRiP+CN	8	767	European	28083
estd214	1000GenomesPhase3	Multiple_NGS_Sequencing	RD+PEM	8	2504	African Asian European Mexican Native American North American South American	2872046

DGV Gold Standard Frequency Calculations

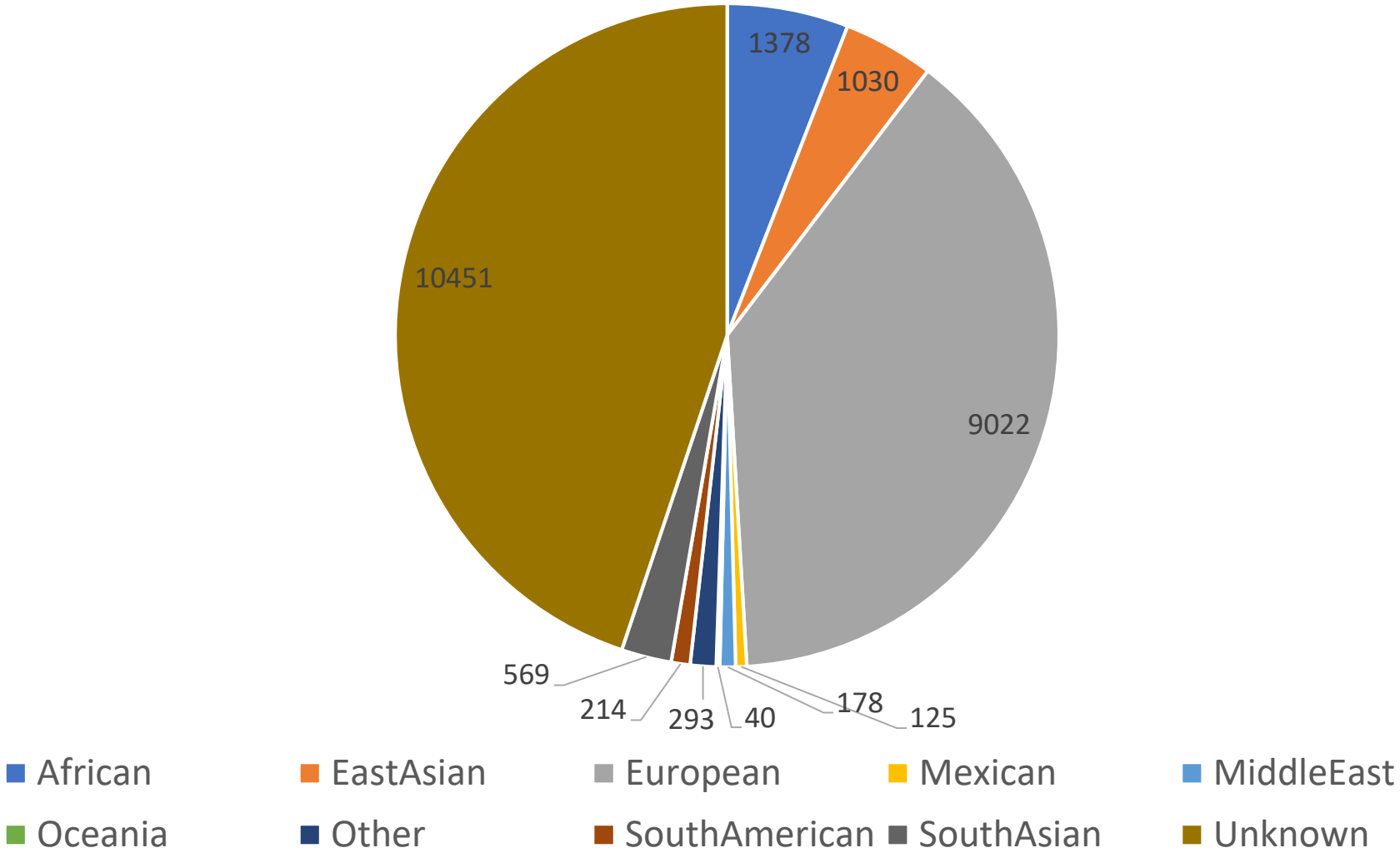
Frequency:

The overall frequency of variants in the region was determined across all studies. To calculate the frequency, we identified which studies had the power to detect variants at each region (accounting for differences in coverage and resolution of the various approaches). The number of unique samples reported to contain a variant as a fraction of the number of unique samples that were tested at each loci was used to determine the frequency. Redundancy in common cohorts and samples across the various studies was addressed.

Size Distribution of CNV in DGV Gold Standard

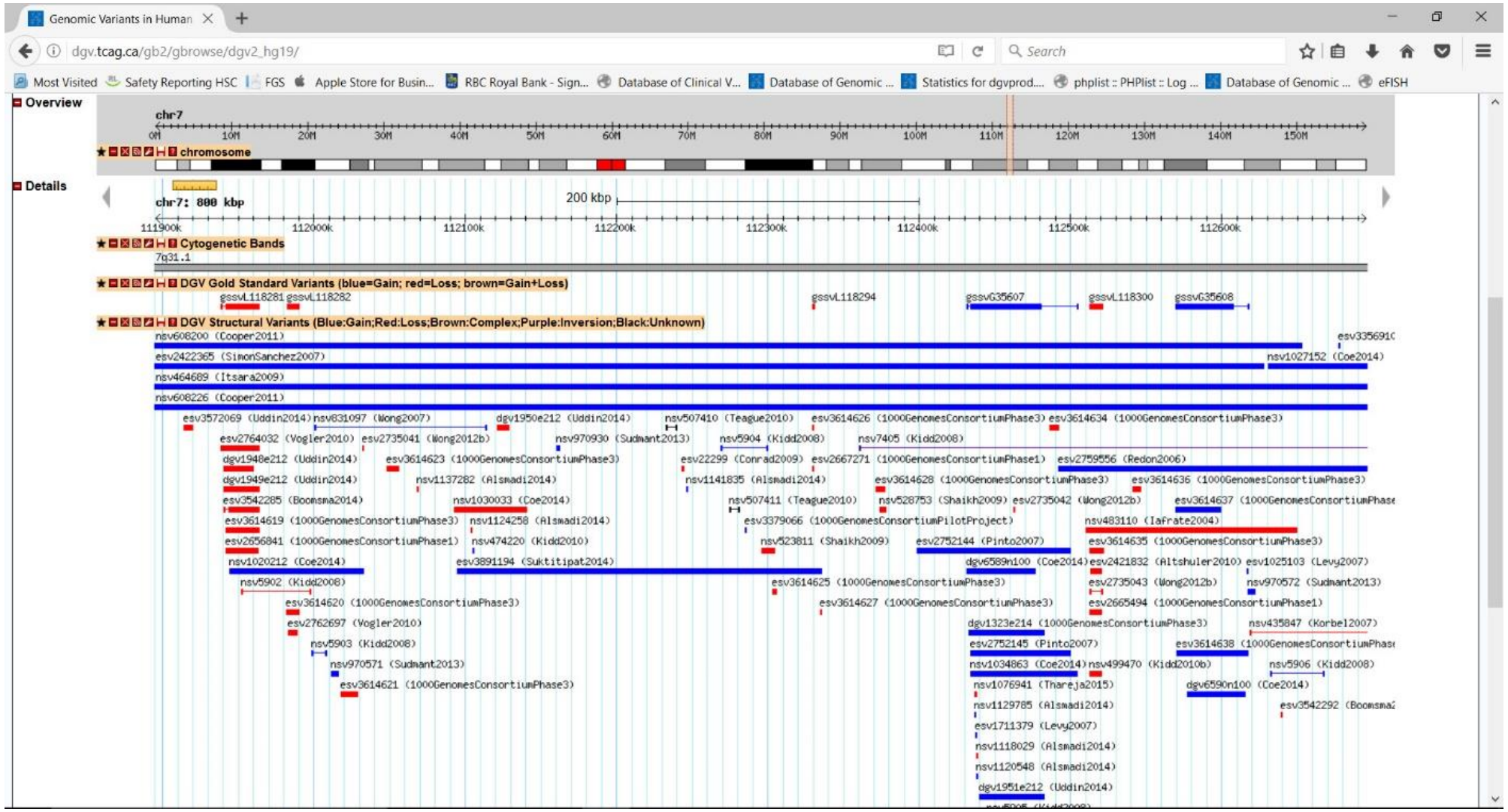


Population Summary of DGV Gold Standard

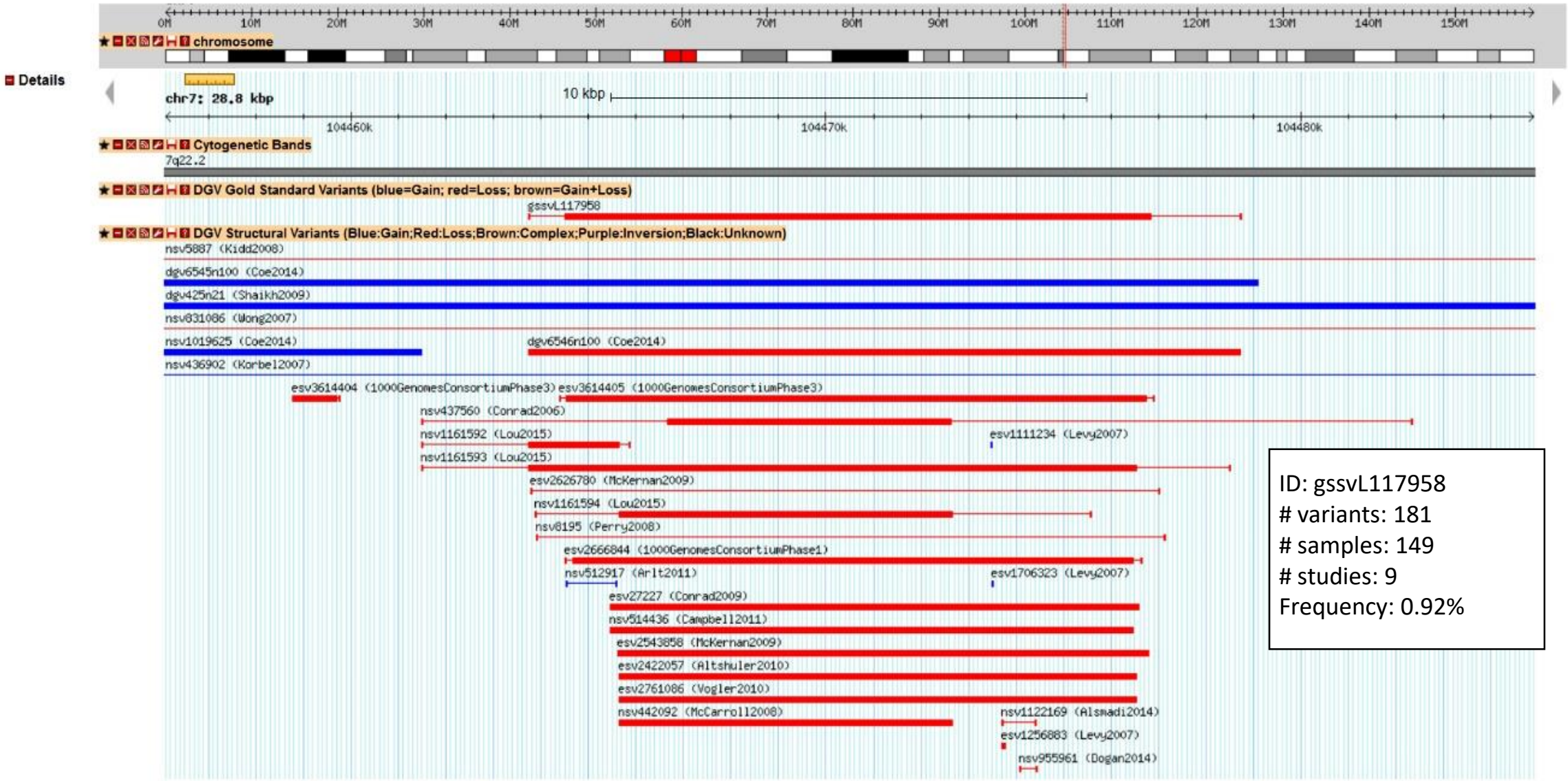


DGV vs. DGV Gold Standard

The DGV structural variation data and DGV Gold Standard data are displayed for comparison.
DGV gold standard CNV regions contain summary information for CNV regions with support across multiple studies and samples.



DGV Gold Standard Example
gssvL117958



ID: gssvL117958
variants: 181
samples: 149
studies: 9
Frequency: 0.92%

GBrowse Details: id:33629

dgv.tcag.ca/gb2/gbrowse_details/dgv2_hg19?ref=chr7;start=104463733;end=104478718;na

exac consortium

Most Visited | my Sun Life | Sun Life ... | Database of Genomic ... | Strand Home | Safety Reporting HSC | Go Transit Guarantee | Office 365 | Rogers: Wireless, TV, I... | Apple Store for Busine... | RB

gssvL117958 Details

Name:	**gssvL117958**
Type:	copy_number_variation_region
Description:	
Source:	CNV
Position:	chr7:104463733..104478718
Length:	14986
Frequency:	0.92%
Platforms:	Affymetrix_6.0
	Affymetrix6.0_Illumina1M
	SOLiD
	Affymetrix6.0
	NimbleGen42M
	Illumina_II_IIX_HiSeq
	AgilentCustom_015685+015686+244K
	AgilentCustom180K_v2.1+v3.0
	Multiple_NGS_Sequencing
PopulationSummary:	African 74:Asian 0:European 0:Mexican 2:MiddleEast 0:NativeAmerican 0:NorthAmerican 0:Oceania 0:SouthAmerican 1:Turkish 0:Admixed 0:Unknown 72
Studies:	Vogler2010
	Campbell2011
	Perry2008
	Conrad2009
	Altshuler2010
	Coe2014
	McKernan2009
	1000GenomesPhase1
	1000GenomesPhase3

Summary Details:

Clicking on the DGV Gold Standard variant will provide a summary information page with details on the CNV region.

The source of the underlying calls, sample identifiers, population summary, frequency and size are reported.

GBrowse Details: id:33629

dgv.tcag.ca/gb2/gbrowse_details/dgv2_hg19?ref=chr7;start=104463733;end=104478718;na

exac consortium

Most Visited | my Sun Life | Sun Life ... | Database of Genomic ... | Strand Home | Safety Reporting HSC | Go Transit Guarantee | Office 365 | Rogers: Wireless, TV, I... | Apple Store for Busine... | RB

algorithms:	PennCNV+Birdseye
	PEM+ReadDepth+SplitRead
	Custom_HMM
	QuantiSNP+BirdsEye
	GADA
	PEM
	ReadDepth
	RD+PEM
	ADM2
inner_end:	104476834
inner_rank:	7
inner_start:	104464519
load_id:	gssvL117958
num_platforms:	9
num_samples:	149
num_studies:	9
num_unique_samples_tested:	16207
num_variants:	181
number_of_algorithms:	9
outer_end:	104478718
outer_start:	104463733
samples:	RW_0258
	HG02449
	RW_0268

Summary Details (continued):

- The number of studies, and the number of samples are indicated, and the number of unique samples tested is reported.

- The frequency is calculated as
(# samples/ # of unique samples tested) x 100

- The inner rank is also displayed to give a general idea on the quality of the breakpoint assignments (1 = best, 16 = lowest)

DGV Struct Var Track Settings

genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=802695427_6n2ModOaZOTP827KYDMxB

Most Visited | my Sun Life | Sun Life ... | Database of Genomic ... | Strand Home | Safety Reporting HSC | Go Transit Guarantee | Office 365 | Rogers: Wireless, TV, I... | Apple Store for Busine... | RBC Royal Bank - Sign...

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Projects | Help | About Us

DGV Struct Var Track Settings

Database of Genomic Variants: Structural Variation (CNV, Inversion, In/del) (▲All Variation tracks)

Display mode: hide Submit Cancel Reset to defaults

List subtracks: ○ only selected/visible ◉ all

☑ hide	DGV Struct Var	Database of Genomic Variants: Structural Var Regions (CNV, Inversion, In/del)	schema
☑ hide	DGV Supp Var	Database of Genomic Variants: Supporting Structural Var (CNV, Inversion, In/del)	schema
☑ hide	DGV Gold Standard	Database of Genomic Variants: Gold Standard Variants	schema

Data version: 2016-05-15

Description

This track displays copy number variants (CNVs), insertions/deletions (InDels), inversions and inversion breakpoints annotated by the Database of Genomic Variants (DGV), which contains genomic variations observed in healthy individuals. DGV focuses on structural variation, defined as genomic alterations that involve segments of DNA that are larger than 1000 bp. Insertions/deletions of 50 bp or larger are also included.

Display Conventions

This track contains three subtracks:

- Structural Variant Regions: annotations that have been generated from one or more reported structural variants at the same location.
- Supporting Structural Variants: the sample-level reported structural variants.
- Gold Standard Variants: curated variants from a selected number of studies in DGV.

Color is used in these subtracks to indicate the type of variation:

- **Inversions** and **inversion breakpoints** are purple.

DGV | Highlight All | Match Case | Match Diacritics | Whole Words | 1 of 1 match | Reached end of page, continued from top

Type here to search

3:28 PM
2020-02-21

Human hg19 chr7:104448747-1

genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeEx

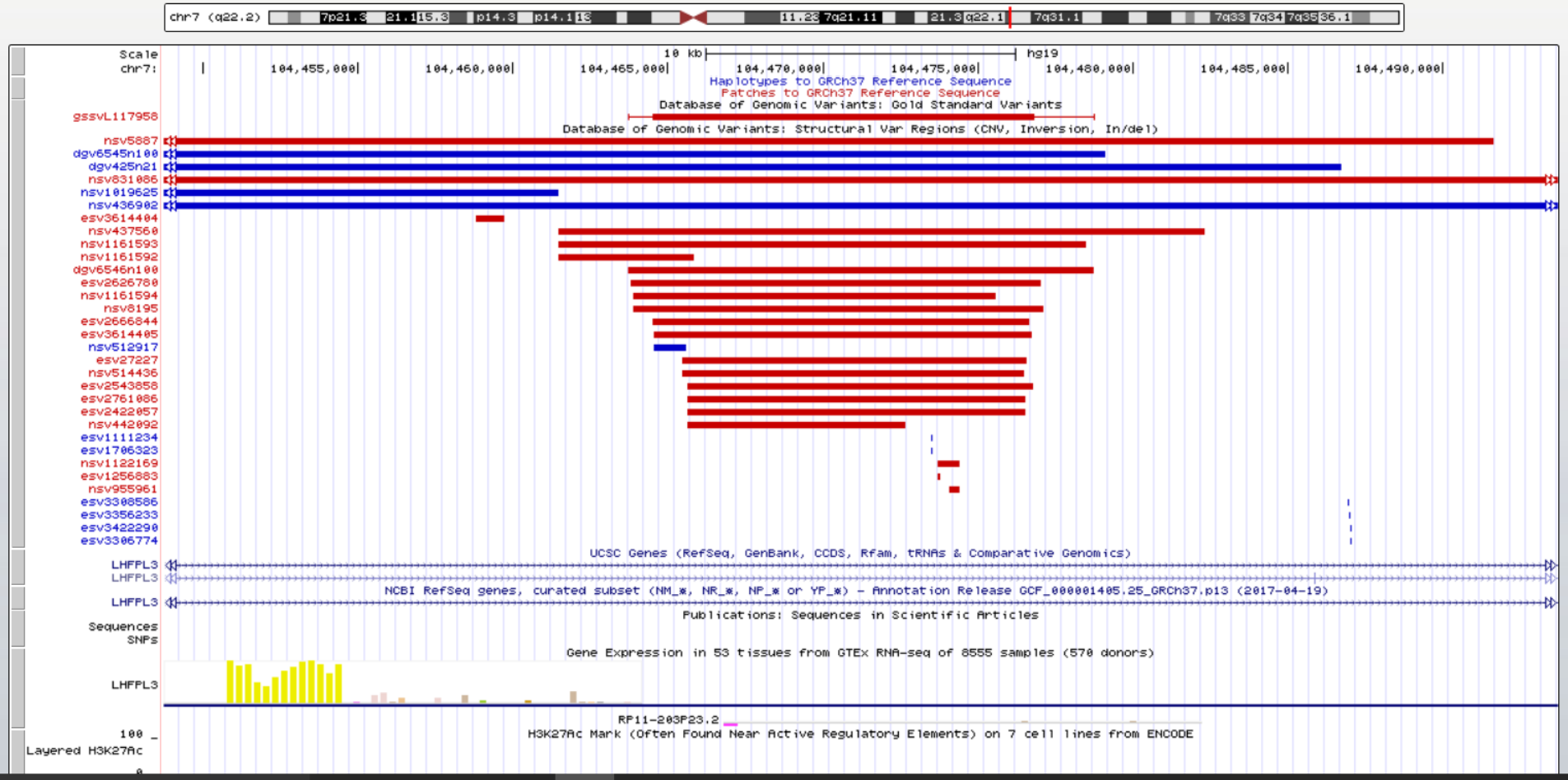
Most Visited | my Sun Life | Sun Life ... | Database of Genomic ... | Strand Home | Safety Reporting HSC | Go Transit Guarantee | Office 365 | Rogers: Wireless, TV, I... | Apple Store for Busine... | RBC Royal Bank - Sign...

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | View | Help | About Us

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

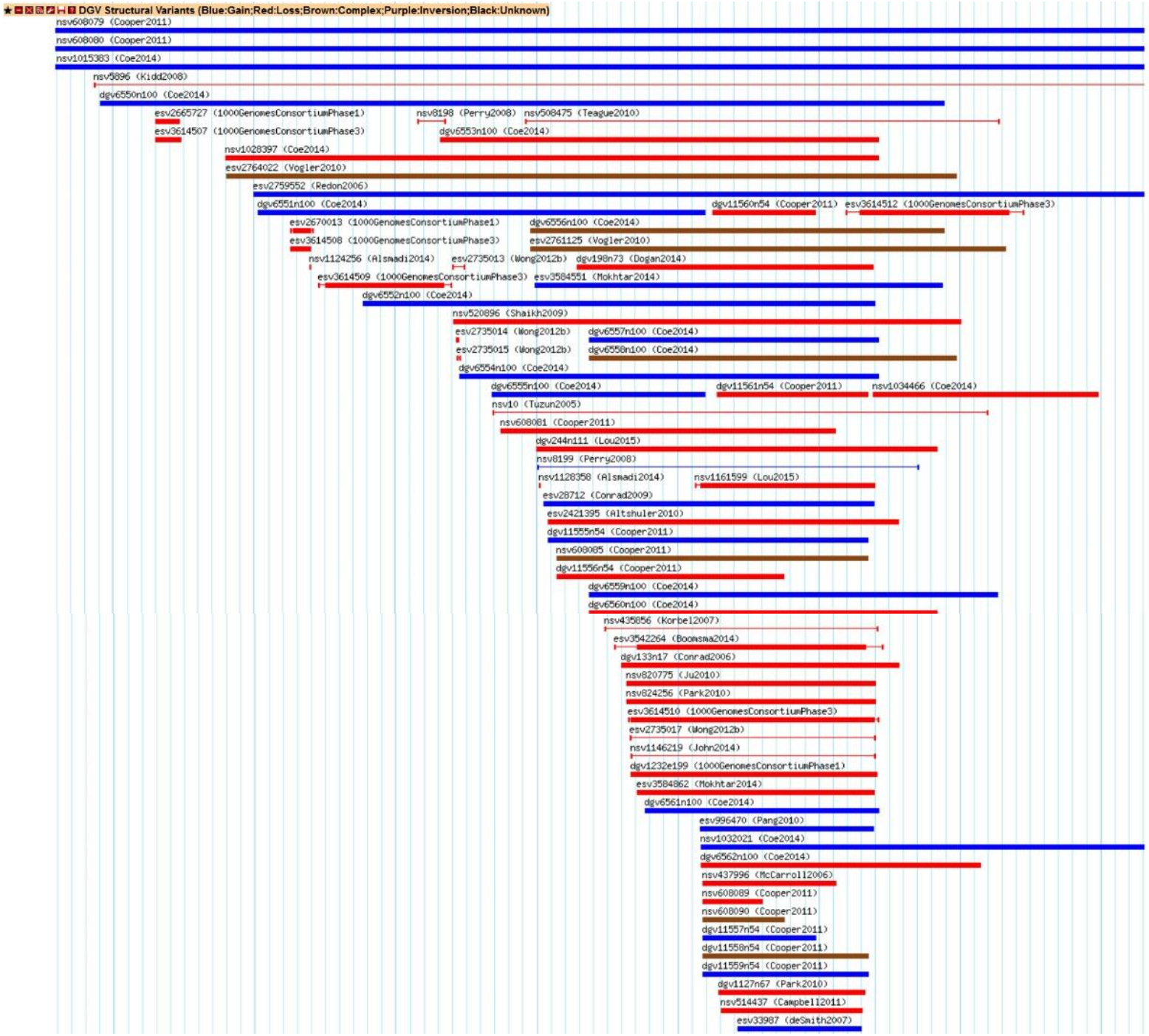
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr7:104,448,747-104,493,704 44,958 bp. enter position, gene symbol, HGVS or search terms go



Type here to search

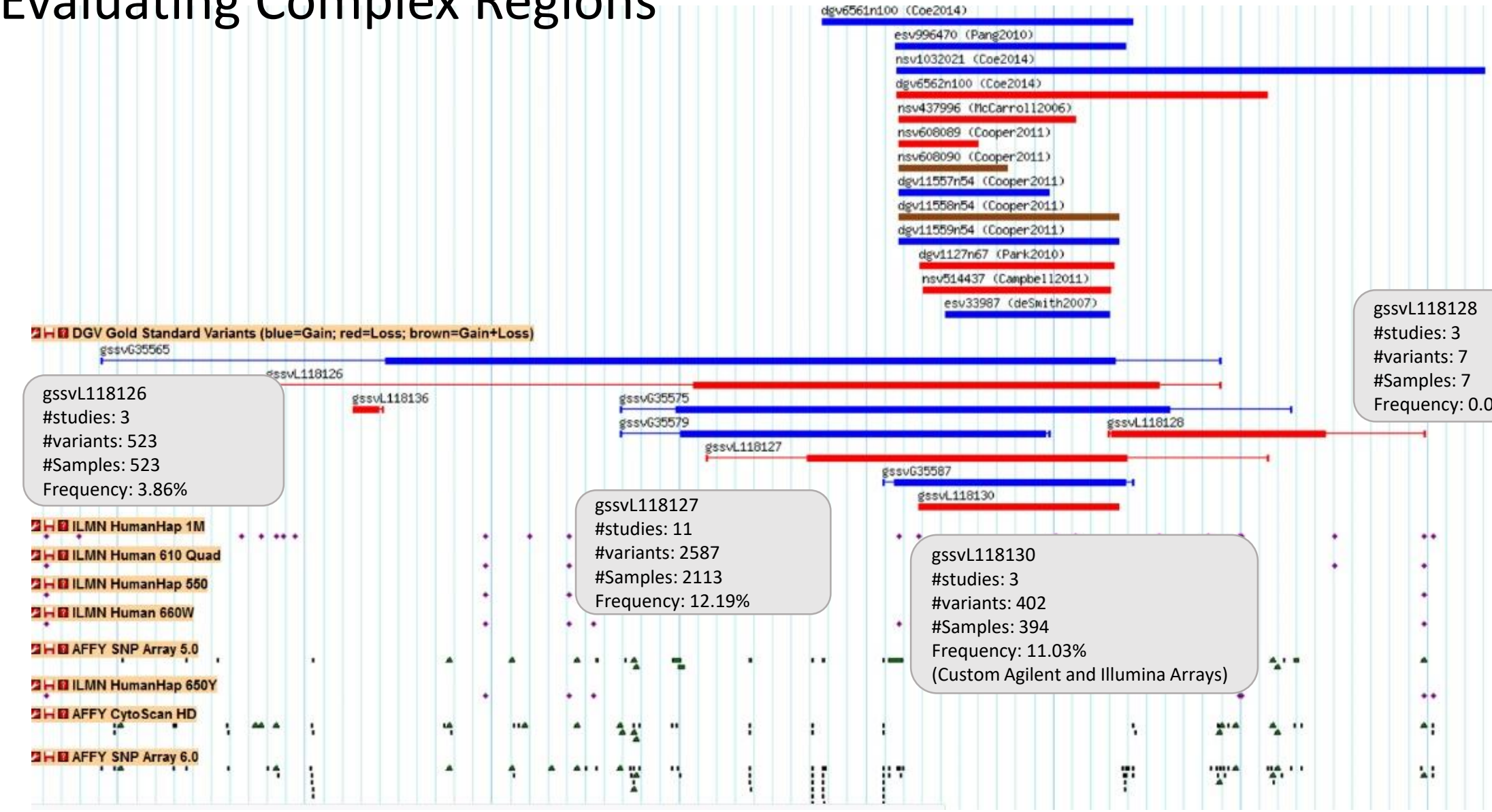
3:33 PM
2020-02-21



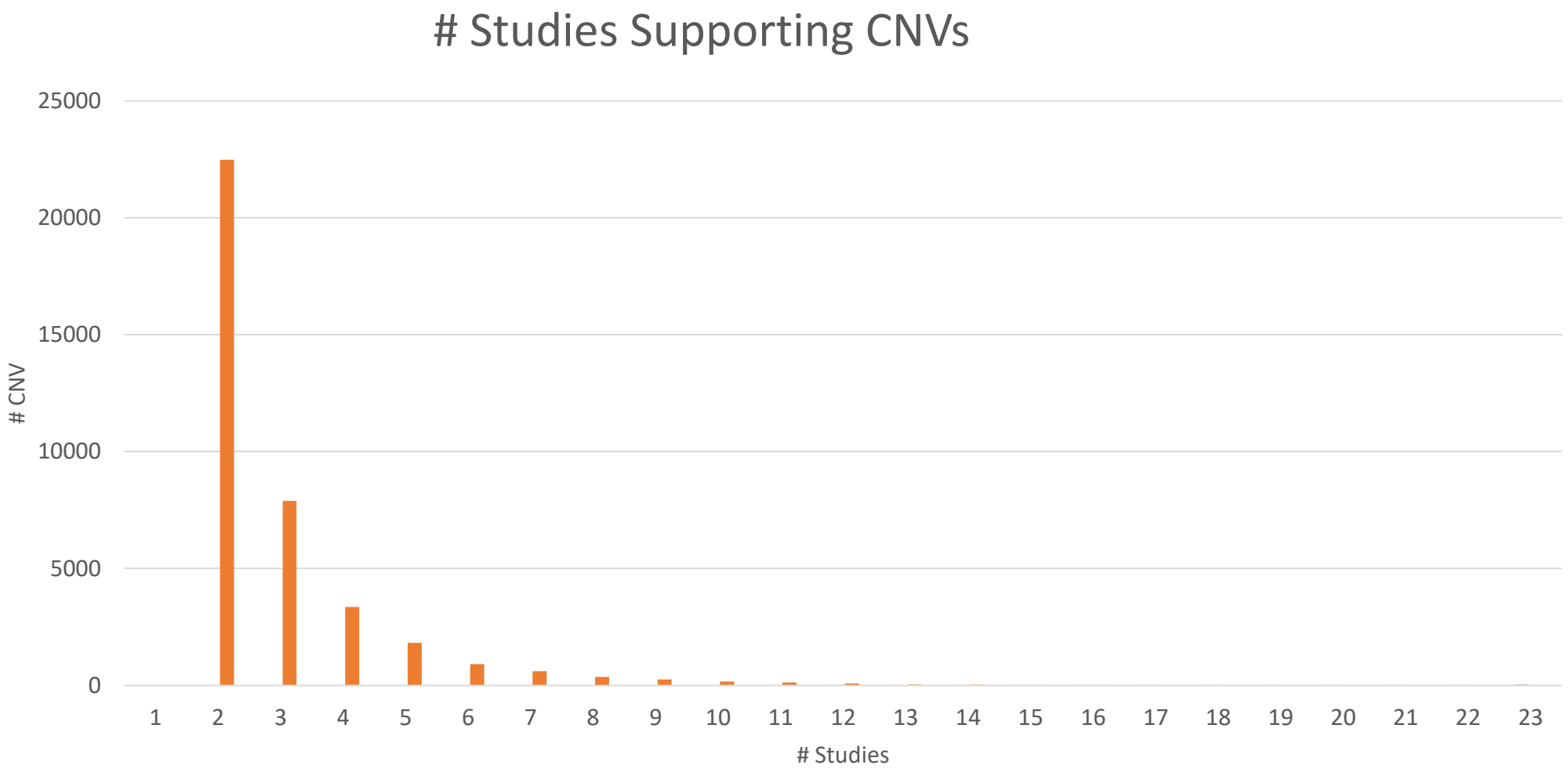
Evaluating Complex Regions

- Some regions may contain a large number of overlapping CNVs which may result in multiple overlapping DGV Gold Standard variants
- This may result from the underlying genomic complexity of the region (segmental duplications, repeat rich regions etc).
- Overlapping entries in the Gold Standard my also result from an inability to resolve variants derived from different platforms/approaches.

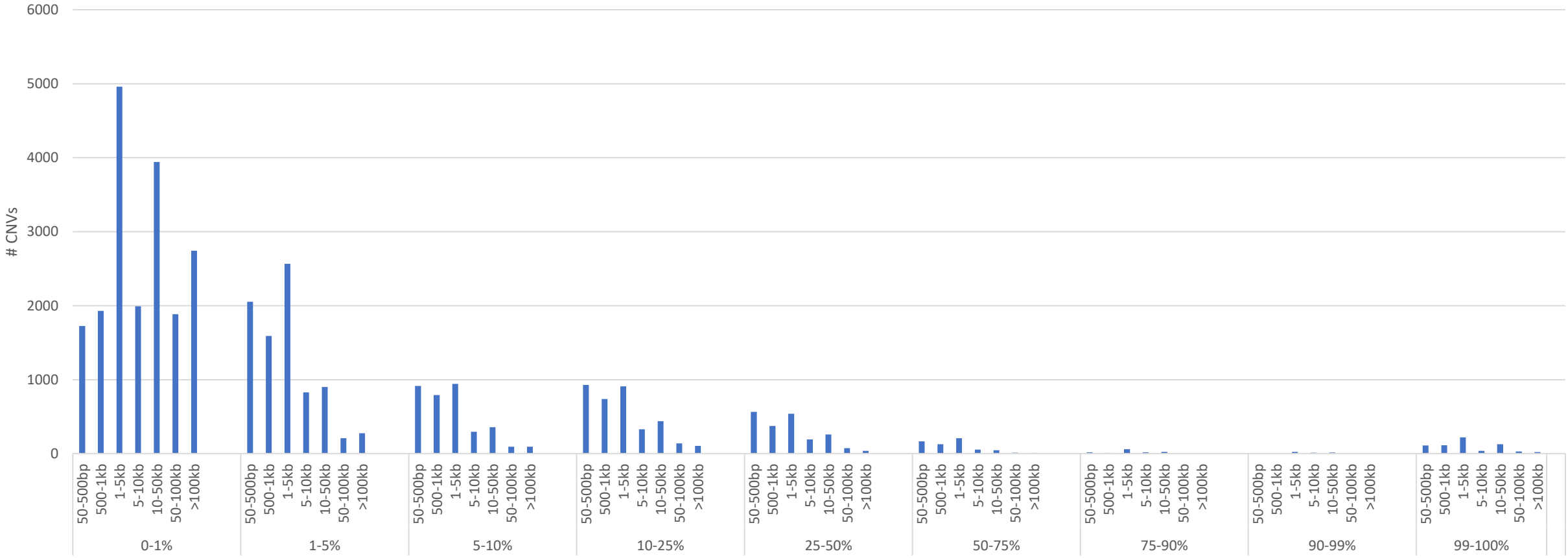
Evaluating Complex Regions



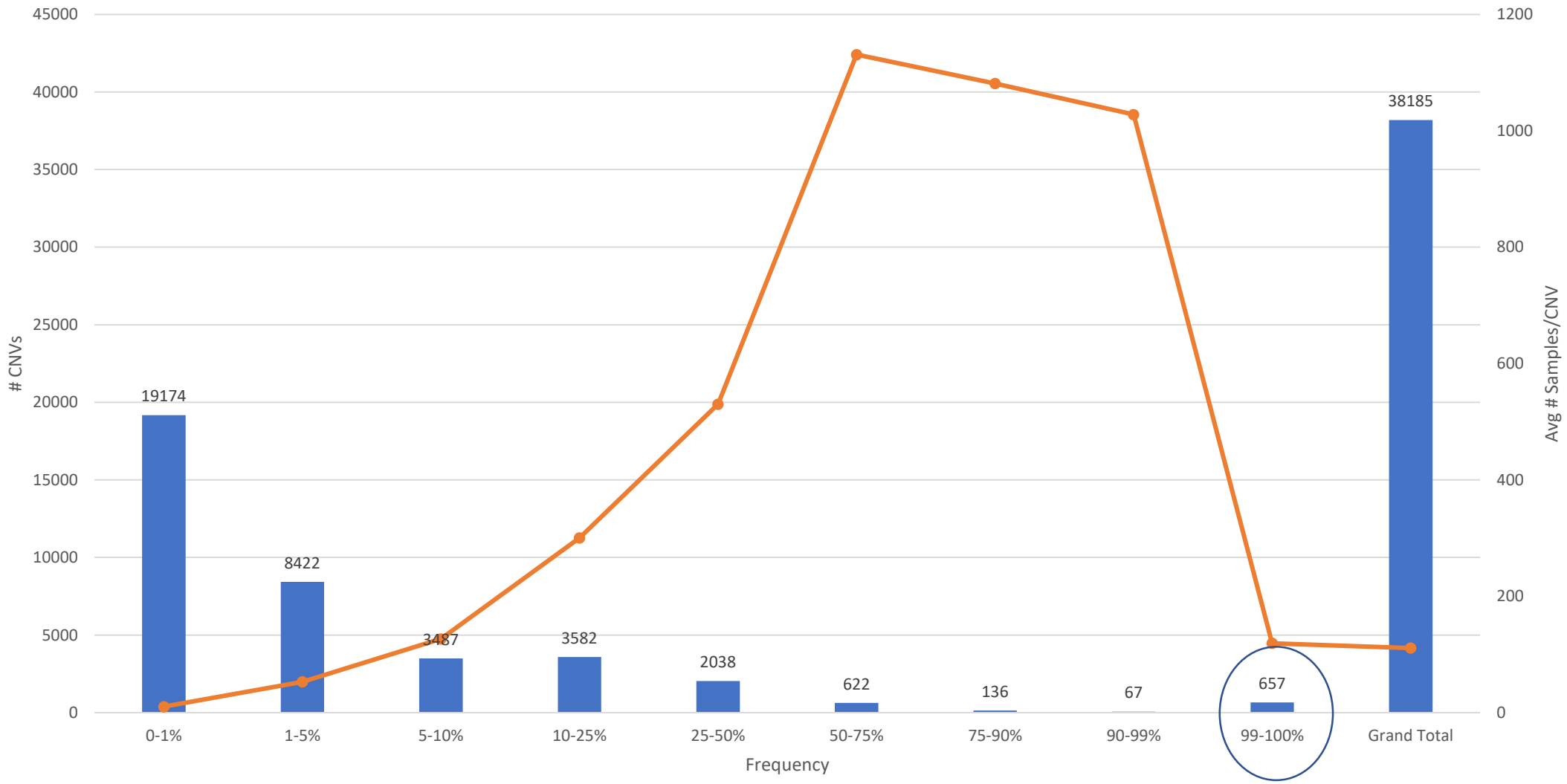
DGV Gold Standard Aggregate Data



DGV Gold Standard CNV Distribution by Size and Frequency



DGV Gold Standard Frequency Metrics



Things to Consider

The category that appears to be of the most concern are the extremely common calls (> 99%), where there is a possibility that these may in fact be reference artifacts (false calls).

When looking at this category (> 99% frequency) in more detail to determine if there are any specific patterns that emerge we see:

1. The majority of the variants in this group are small (<10kb)
2. Most are also only found in 2 subjects and 2 studies (minimum inclusion criteria (400/657)
3. Many of these reside in repeat rich regions (> 75% of the CNV length is annotated as complex)
    - -Complex regions include simple repeats, repeat masker, segmental duplications, satellite repeats, and assembly gaps
    - -CNVs in this category which overlap these complex regions (> 75% of their length) should be evaluated with caution.

Evaluating the DGV Gold Standard Variants with ClinGen Dosage Sensitivity Map

(https://dosage.clinicalgenome.org/)

(A) Haploinsufficiency and DGV Gold Standard Deletions

60 Genes with evidence for haploinsufficiency overlap a deletion.

-remove deletions where > 75% of the region is covered by repeats or segmental duplications, n = 53 genes

-remove deletions where the frequency of the variant is less than 1% or greater than 99%, n = 7 genes

-remove deletions where there is only support in 2 samples and 2 studies (minimum criteria), n = 4

Gene	CNV	Frequency	Evidence	Phenotype
GP6	gssvL59631	2.69	AR	BLEEDING DISORDER, PLATELET-TYPE, 11; BDPLT11
MCEE	gssvL63781	2.87	AR	METHYLMALONYL-CoA EPIMERASE DEFICIENCY
PSPH	gssvL115680	1.04	AR	PHOSPHOSERINE PHOSPHATASE DEFICIENCY; PSPHD
TSEN2	gssvL79955	2.69	AR	PONTOCEREBELLAR HYPOPLASIA, TYPE 2B; PCH2B

-if we require > 2 studies, and > 2 samples, we do not find any CNVs which overlap.

(B) Triplosensitivity and DGV Gold Standard Duplications

1 Gene with evidence for triplosensitivity overlaps a duplication (found at 0.02% frequency)

Gene	CNV	Frequency	Evidence	Phenotype
PAK2	gssvG26773	0.02	Little evidence for dosage pathogenicity	In a study comparing a large cohort of children with ID and/or DD to unaffected adult controls, the incidence of observed gains of PAK2 was 7/15,767 cases vs 0/8,329 controls. These duplications were likely not focal and included surrounding genes.

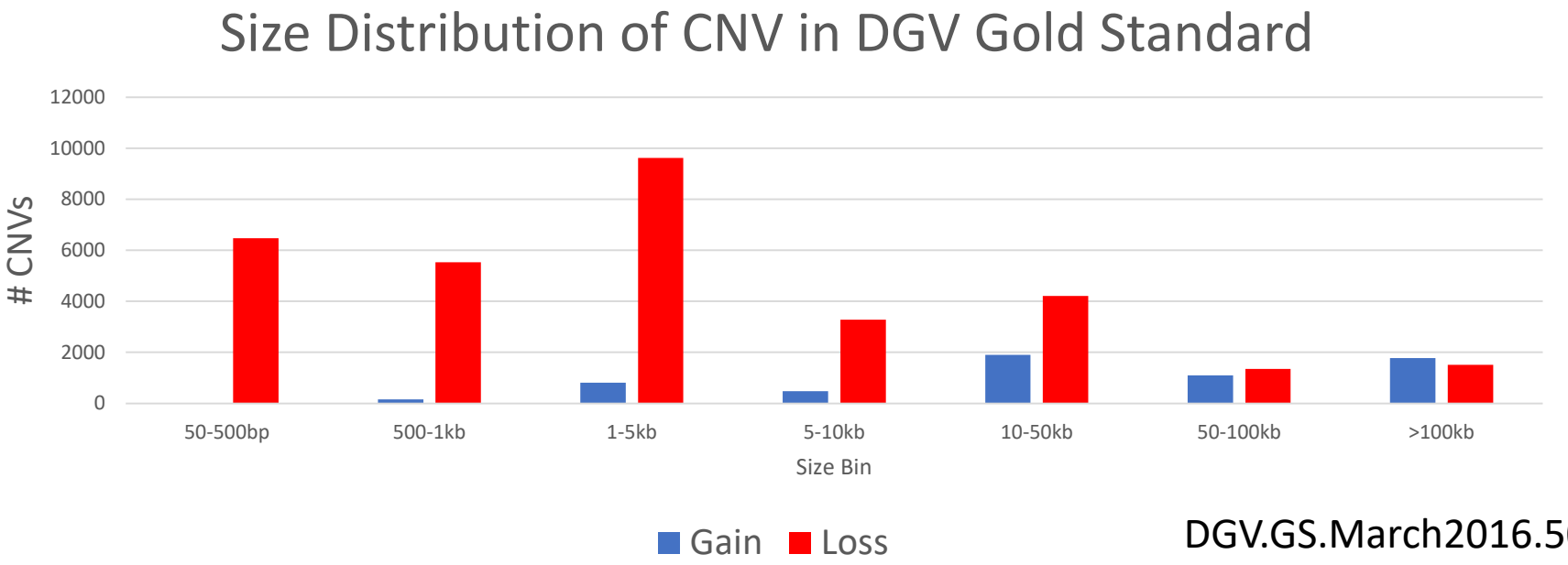
How does the DGV Gold Standard compare to other population CNV data?

	DGV Gold Standard	gnomAD SV
# Datasets/Studies	32	53
# Samples	23,610	10,847
# Variants Calls	5,844,089	313,581
# Variant Regions	38,185	304,733
Data Type	Microarray, multiple platforms	Sequencing (NGS) High - 30x coverage
	Sequencing (NGS) Low - 15x coverage	
	Sequencing (Sanger)	

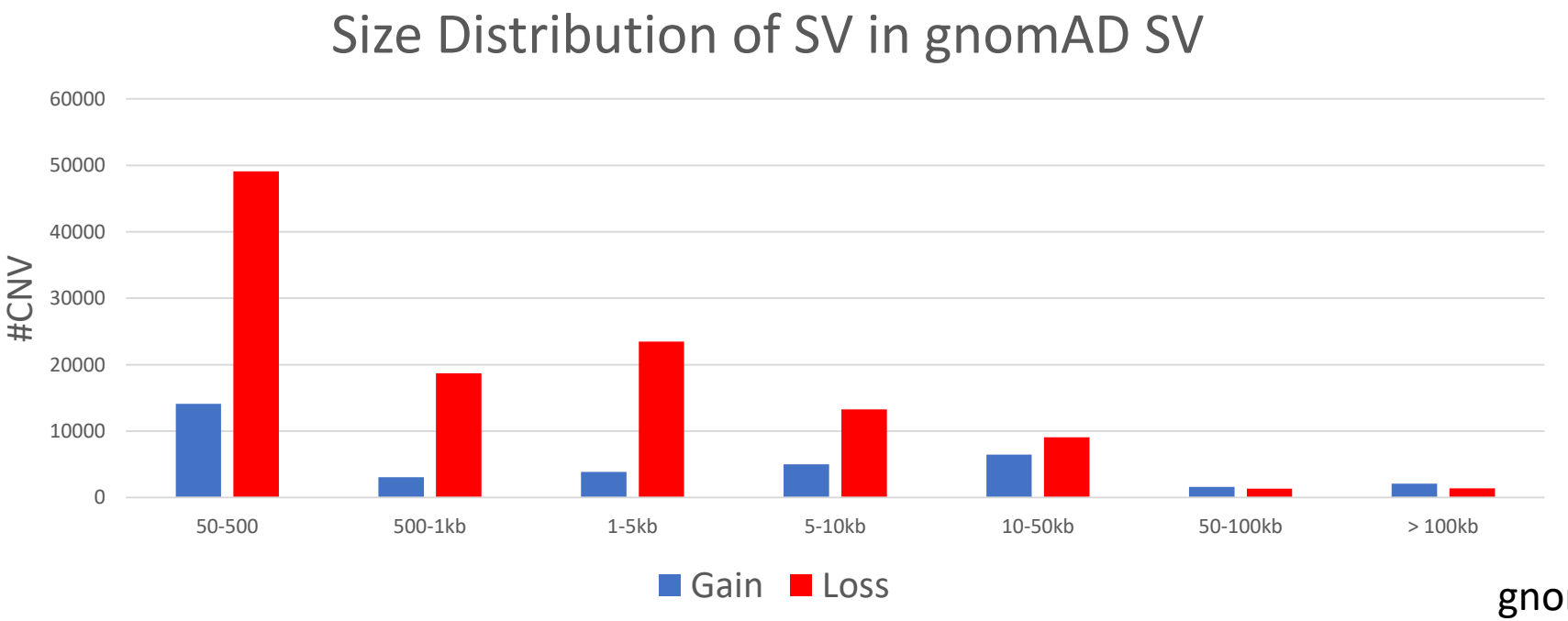
Note: The samples included in each database are unique, and do not overlap

	gnomAD SV	DGV Gold Standard	DGV
Classes of SV	Deletion	Deletion	Gain
	Duplication	Duplication	Loss
	Multiallelic CNV		Duplication
	Insertion		Deletion
	Inversion		Insertion
	Translocation		Tandem Duplication
	Complex SV		Inversion
	Breakends		Complex
			Sequence Alteration
			Novel Sequence Insertion
			Copy Number Variation

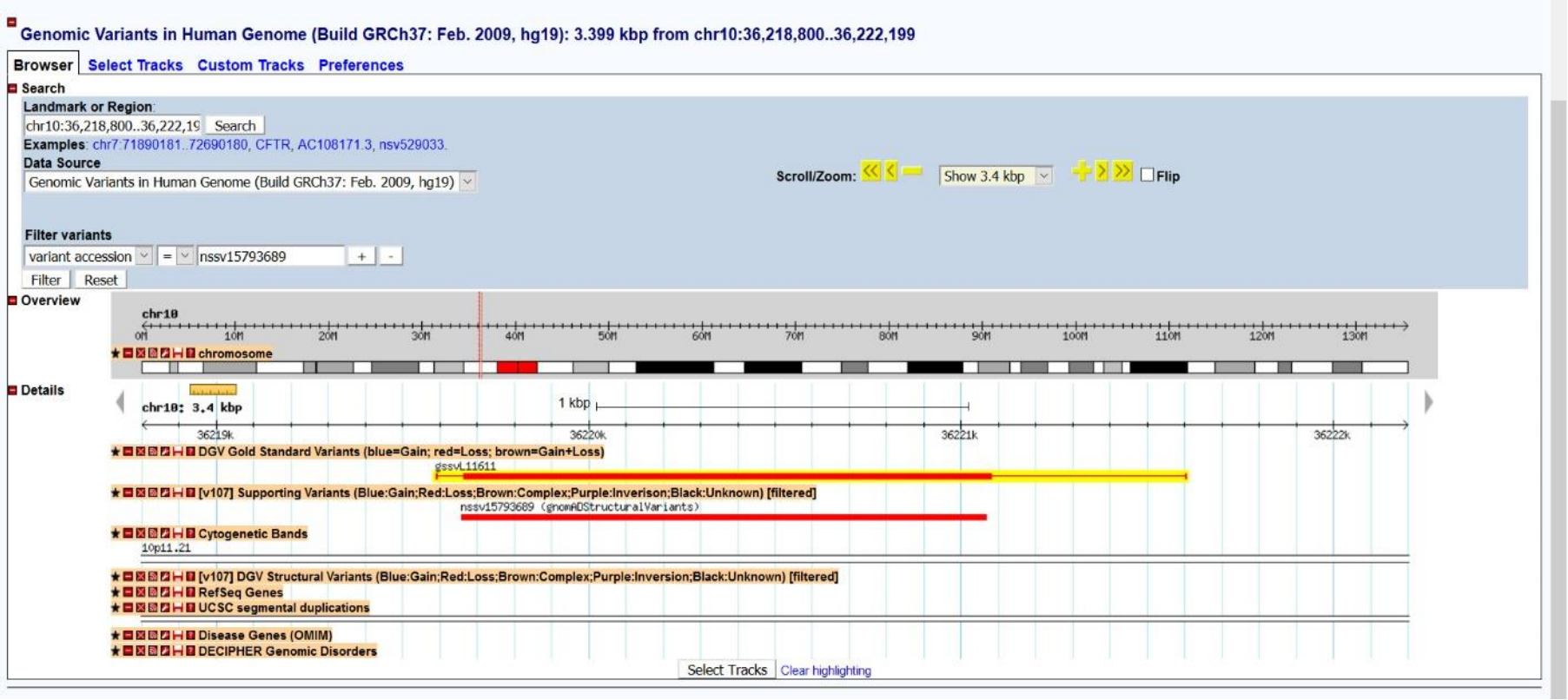
(https://www.ncbi.nlm.nih.gov/dbvar/studies/nstd166/ AND https://gnomad.broadinstitute.org/about).



DGV.GS.March2016.50percent.GainLossSep.Final.hg19.gff3



gnomad_v2.1_sv.controls_only.sites.bed



Comparing The Reported Frequency of Common Variants

DGV Gold Standard	gnomAD SV	Comparison
gssvL11611	nssv15793689 (DEL_10_109480)	# Homozygous = 33
num_samples: 467	allele_count=1242	# samples with CNV = 1209
num_studies:7	num_studies=n/a	
num_unique_samples_tested: 3460	allele_number=21692	# samples tested = 10846
num_variants: 743	num_variants=n/a	
Frequency: 13.5%	allele_frequency=0.057256	Frequency = 11.14%
https://gnomad.broadinstitute.org/variant/DEL_10_109480?dataset=gnomad_sv_r2_1		

More info on the database and the DGV Gold standard:

Nucleic Acids Research Advance Access published October 29, 2013

Nucleic Acids Research, 2013, 1–7
doi:10.1093/nar/gkt958

The database of genomic variants: a curated collection of structural variation in the human genome

Jeffrey R. MacDonald[1], Robert Ziman[1], Ryan K. C. Yuen[1], Lars Feuk[2,*] and Stephen W. Scherer[1,3,*]

[1]The Centre for Applied Genomics, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, 686 Bay Street, Toronto, Ontario M5G 0A4, Canada, [2]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala SE-751 08, Sweden and [3]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Received August 16, 2013; Revised September 25, 2013; Accepted September 27, 2013

Nucleic Acids Res. 2014 Jan;42. **The Database of Genomic Variants: a curated collection of structural variation in the human genome.**
MacDonald JR[1], Ziman R, Yuen RK, Feuk L, Scherer SW.

s | AOP, published online 3 February 2015; doi:10.1038/nrg3871

ANALYSIS

Downloaded fro

A copy number variation map of the human genome

Mehdi Zarrei[1], Jeffrey R. MacDonald[1], Daniele Merico[1] and Stephen W. Scherer[1,2]

Abstract | A major contribution to the genome variability among individuals comes from deletions and duplications — collectively termed copy number variations (CNVs) — which alter the diploid status of DNA. These alterations may have no phenotypic effect, account for adaptive traits or can underlie disease. We have compiled published high-quality data on healthy individuals of various ethnicities to construct an updated CNV map of the human genome. Depending on the level of stringency of the map, we estimated that 4.8–9.5% of the genome contributes to CNV and found approximately 100 genes that can be completely deleted without producing apparent phenotypic consequences. This map will aid the interpretation of new CNV findings for both clinical and research applications.

Nat Rev Genet. 2015 Mar;16(3):172-83. doi: 10.1038/nrg3871.
A copy number variation map of the human genome.
Zarrei M[1], MacDonald JR[1], Merico D[1], Scherer SW[2].

The Hospital for Sick Children and TCAG

Dr. Stephen Scherer
Dr. Mehdi Zarrei
Dr. James Stavropoulos
Joseph Whitney
Bhooma Thiruvahindrapuram
Dr. Daniele Merico
Dr. Christian Marshall
Robert Ziman
Dr. Giovanna Pellecchia

web: http://dgv.tcag.ca/
email: dgv-contact@sickkids.ca

External Collaborators/DGV Advisory Board

Dr. Lars Feuk
Dr. Ilkka Lappalainen
Dr. Deanna Church
Dr. Tim Hefferon
Dr. Paul Flicek
Dr. David Ledbetter
Angie Hinrichs

Funding for DGV provided by:

TCAG
SickKids
GenomeCanada
CIHR IRSC
McLaughlin Centre
UNIVERSITY OF TORONTO