GOLDEN HELIX

Enabling Precision Medicine

Comprehensive Clinical Workflows for Copy Number Variants in VarSeq

Gabe Rudy – VP Product & Engineering

CIOReview
20 most promising Biotech Technology Providers

pharma TECH OUTLOOK
Top 10 Analytics Solution Providers

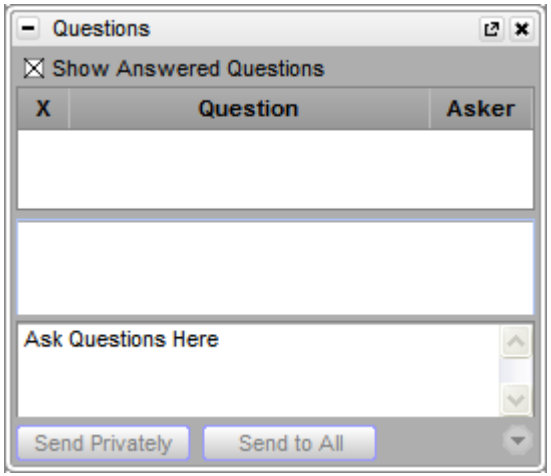
Gartner
Hype Cycle for Life sciences

Agenda

1. Overview Golden Helix
2. CNVs as Part of Clinical Interpretation Workflow
3. VarSeq Demo and Walk Through
4. Up Next & Annoucements

Questions during the presentation

Use the Questions pane in your GoToWebinar window



Golden Helix – Who We Are

Golden Helix is a global bioinformatics company founded in 1998.

GlaxoSmithKline

VarSEQ®

- Variant Calling
- Filtering and Annotation
- Clinical Reports
- CNV Analysis
- Pipeline: Run Workflows

VARIANT WAREHOUSE

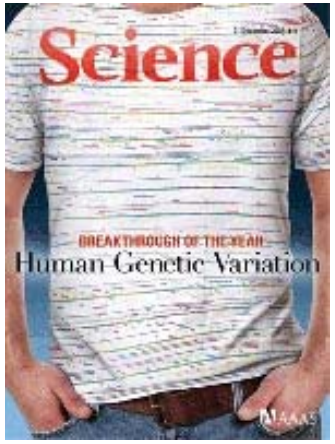
- Variant Warehouse
- Centralized Annotations
- Hosted Reports
- Sharing and Integration

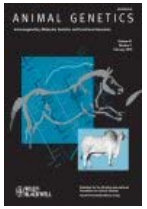
SNP & VARIATION SUITE

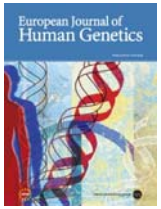
- GWAS
- Genomic Prediction
- Large-N-Population Studies
- RNA-Seq
- Large-N CNV-Analysis

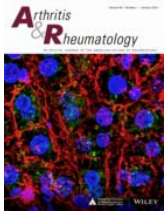
GOLDEN HELIX
Enabling Precision Medicine

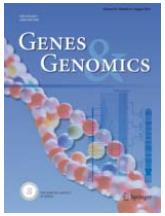
Cited in over 1100 peer-reviewed publications











GOLDEN HELIX
Enabling Precision Medicine

Over 350 customers globally

SickKids
THE HOSPITAL FOR SICK CHILDREN

Cincinnati Children's Hospital Medical Center

The Children's Hospital of Philadelphia

Stanford University

Ucla

Yale

The Feinstein Institute for Medical Research
North Shore-Long Island Jewish Health System

North Shore LIJ

PREVENTION GENETICS
DISEASE PREVENTION THROUGH GENETIC TESTING

MAYO CLINIC

LINEAGEN

MOFFITT CANCER CENTER

Dartmouth-Hitchcock
NORRIS COTTON CANCER CENTER

BAYER

Lilly

abbvie

TEVA

NATIONAL CANCER INSTITUTE

GOLDEN HELIX
Enabling Precision Medicine

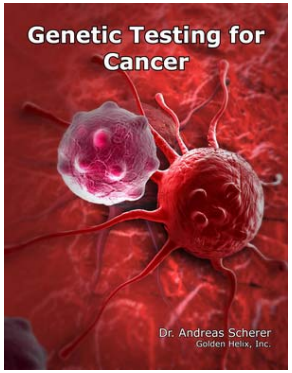
USDA
USDA Agricultural Research Service

FDA

Golden Helix – Who We Are

When you choose a Golden Helix solution, you get more than just software

- REPUTATION
- TRUST
- EXPERIENCE



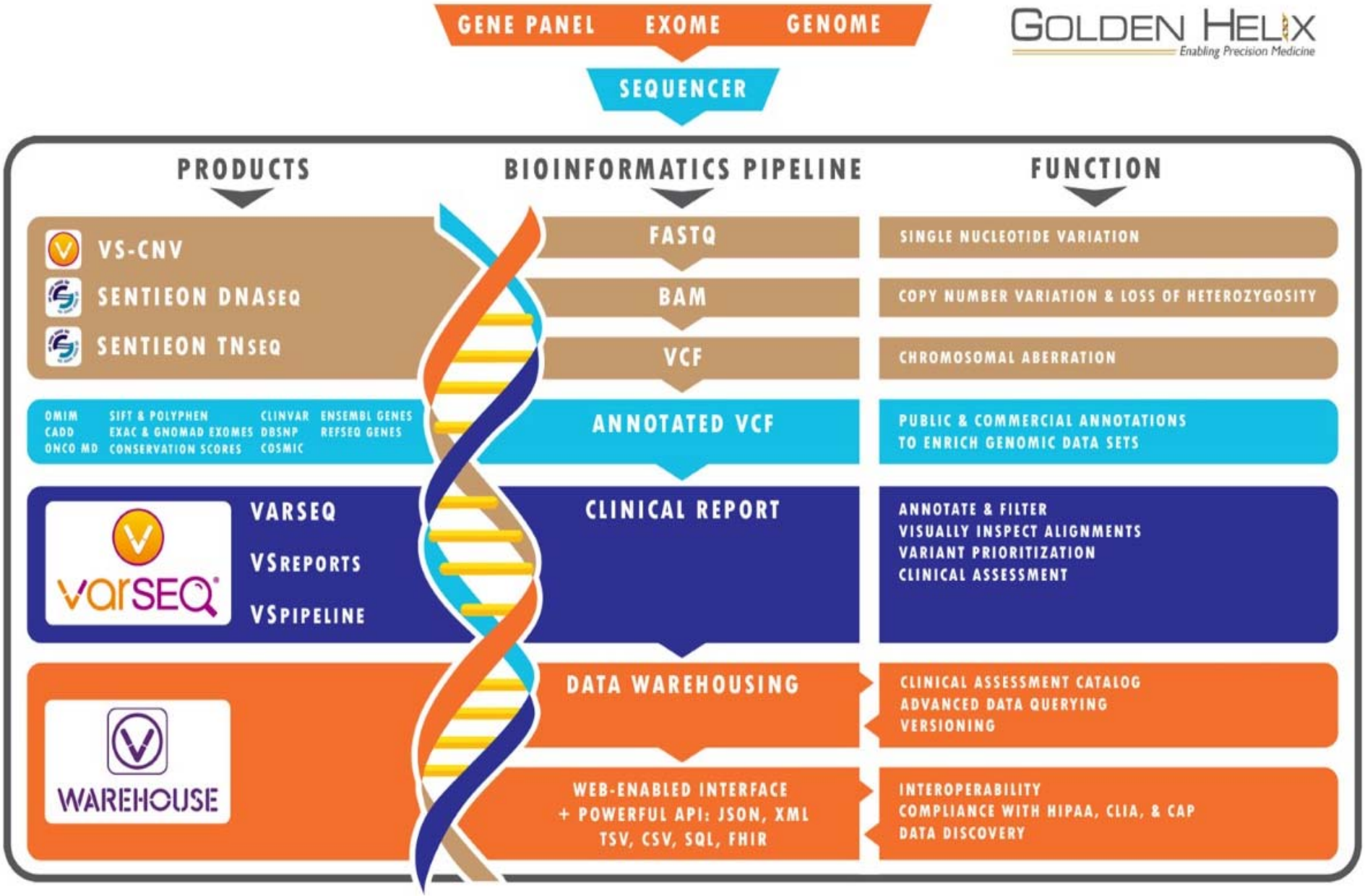
- INDUSTRY FOCUS
- THOUGHT LEADERSHIP
- COMMUNITY

- TRAINING
- SUPPORT
- RESPONSIVENESS

- TRANSPARENCY
- INNOVATION and SPEED
- CUSTOMIZATIONS

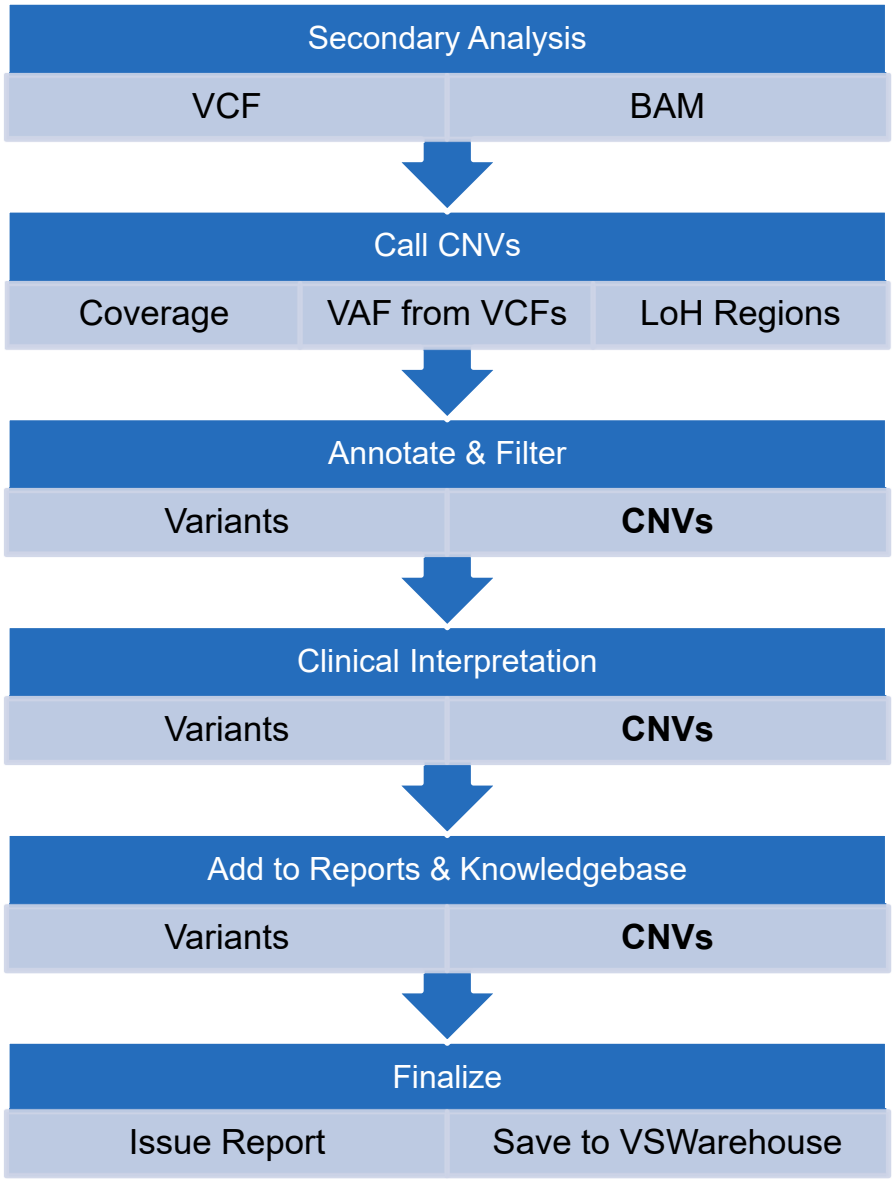
GOLDEN HELIX
Enabling Precision Medicine

VarSeq Clinical Workflows Stack



Clinical Interpretation Workflows (including CNV)

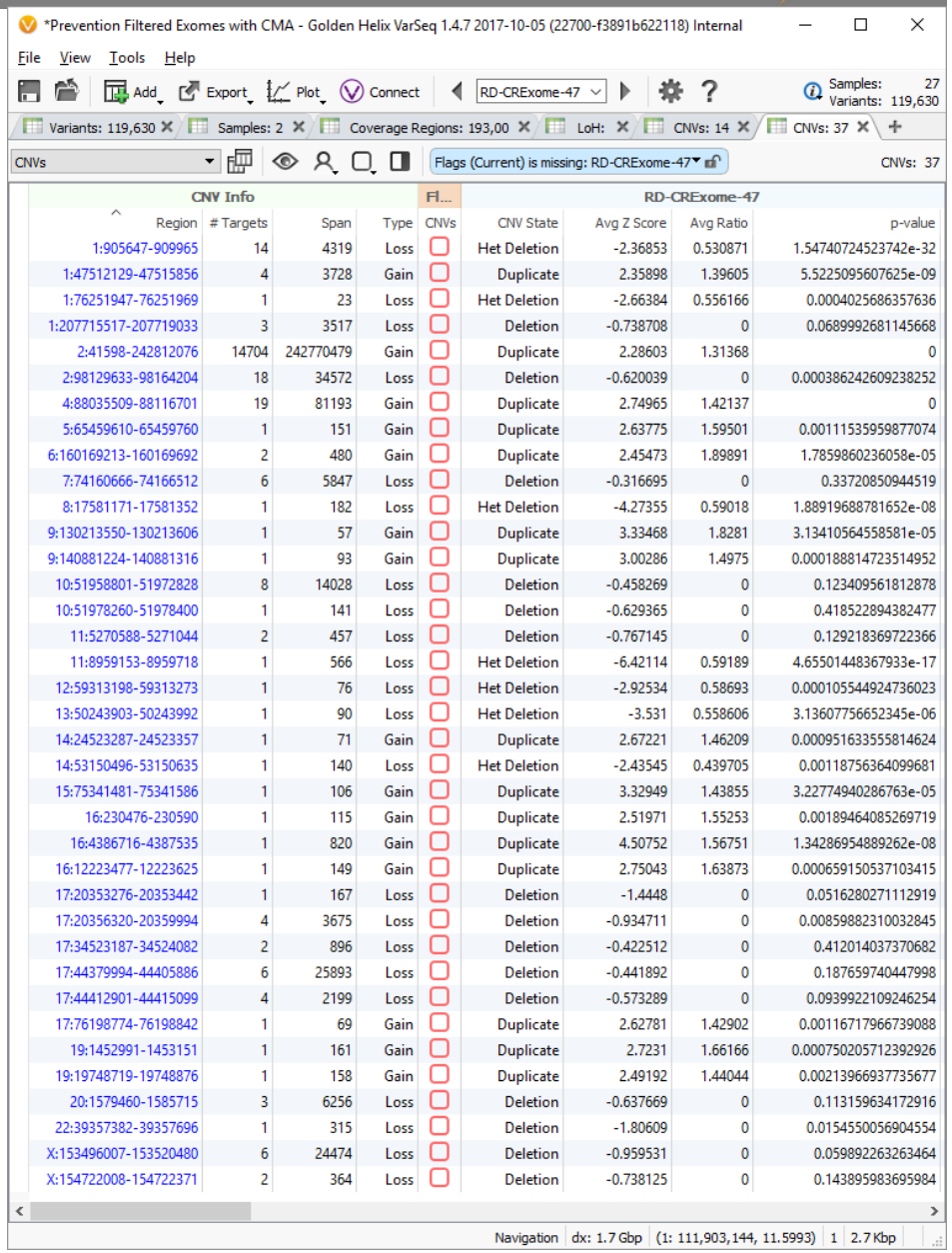
- **Secondary Analysis**
  - Sentieon or Other
  - Small variants called in VCF
  - BAMs provide important coverage data for your targets, also input to CNV
- **VS-CNV Caller 2.0**
  - Can incorporate VCF BAF/VAF
  - Enhanced by LoH regions (exomes)
  - Per-sample CNV calls
    - QC Flags
    - Supporting metrics
    - P-Values
- **After calling:**
  - In some cases we want to treat CNVs like small variants in remaining steps of workflow



GOLDEN HELIX
Accelerating the Quest for Significance™

Why Annotate CNVs?

- **Gene Panels (~50 genes)**
  - No problem looking at one or two CNVs per sample
  - Often have intuition / experience handling handful of atypical genes (PMS2 false-positive prone etc)
  - Often have pre-existing interpretations for most genes. Focused across a single phenotype
- **Large Panels / Exomes**
  - Genes may have multiple phenotype associations
  - Includes regions with high levels of CNVs in population catalogs
  - Clinical interpretations of existing CNVs important
  - Filtering or ranking necessary to reduce manual interpretation to a handful of CNVs

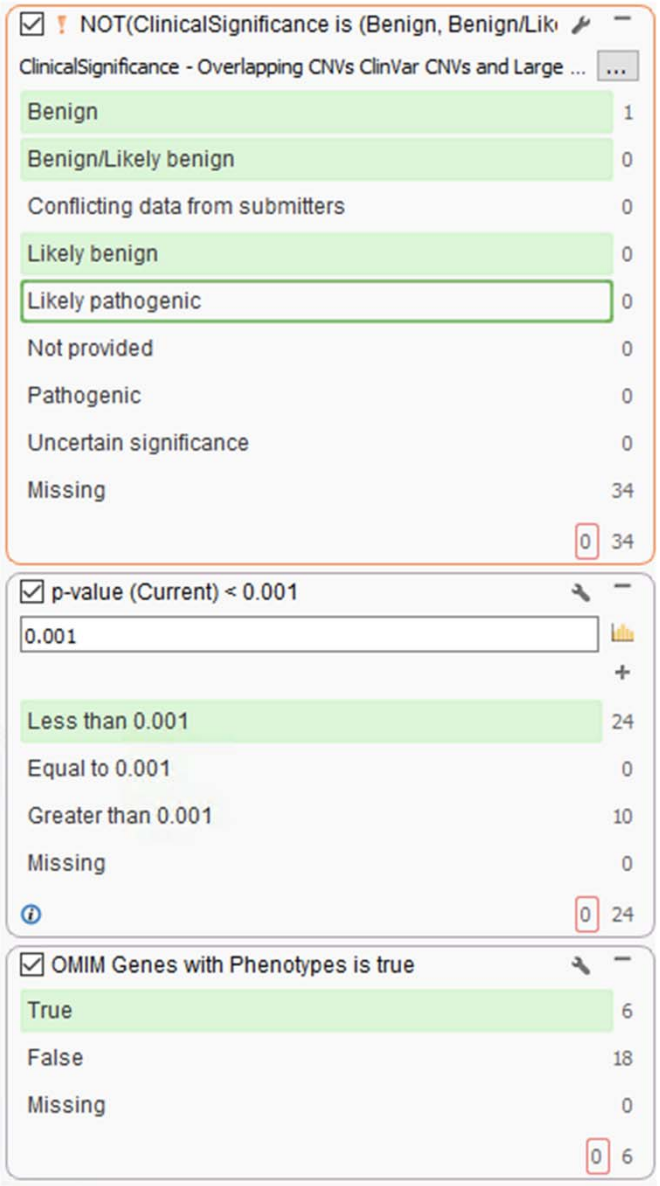


CNV Info				Fl...	RD-CRExome-47			
Region	# Targets	Span	Type	CNVs	CNV State	Avg Z Score	Avg Ratio	p-value
1:905647-909965	14	4319	Loss		Het Deletion	-2.36853	0.530871	1.54740724523742e-32
1:47512129-47515856	4	3728	Gain		Duplicate	2.35898	1.39605	5.5225095607625e-09
1:76251947-76251969	1	23	Loss		Het Deletion	-2.66384	0.556166	0.0004025686357636
1:207715517-207719033	3	3517	Loss		Deletion	-0.738708	0	0.0689992681145668
2:41598-242812076	14704	242770479	Gain		Duplicate	2.28603	1.31368	0
2:98129633-98164204	18	34572	Loss		Deletion	-0.620039	0	0.000386242609238252
4:88035509-88116701	19	81193	Gain		Duplicate	2.74965	1.42137	0
5:65459610-65459760	1	151	Gain		Duplicate	2.63775	1.59501	0.00111535959877074
6:160169213-160169692	2	480	Gain		Duplicate	2.45473	1.89891	1.7859860236058e-05
7:74160666-74166512	6	5847	Loss		Deletion	-0.316695	0	0.33720850944519
8:17581171-17581352	1	182	Loss		Het Deletion	-4.27355	0.59018	1.88919688781652e-08
9:130213550-130213606	1	57	Gain		Duplicate	3.33468	1.8281	3.13410564558581e-05
9:140881224-140881316	1	93	Gain		Duplicate	3.00286	1.4975	0.000188814723514952
10:51958801-51972828	8	14028	Loss		Deletion	-0.458269	0	0.123409561812878
10:51978260-51978400	1	141	Loss		Deletion	-0.629365	0	0.418522894382477
11:5270588-5271044	2	457	Loss		Deletion	-0.767145	0	0.129218369722366
11:8959153-8959718	1	566	Loss		Het Deletion	-6.42114	0.59189	4.65501448367933e-17
12:59313198-59313273	1	76	Loss		Het Deletion	-2.92534	0.58693	0.000105544924736023
13:50243903-50243992	1	90	Loss		Het Deletion	-3.531	0.558606	3.13607756652345e-06
14:24523287-24523357	1	71	Gain		Duplicate	2.67221	1.46209	0.000951633555814624
14:53150496-53150635	1	140	Loss		Het Deletion	-2.43545	0.439705	0.00118756364099681
15:75341481-75341586	1	106	Gain		Duplicate	3.32949	1.43855	3.22774940286763e-05
16:230476-230590	1	115	Gain		Duplicate	2.51971	1.55253	0.00189464085269719
16:4386716-4387535	1	820	Gain		Duplicate	4.50752	1.56751	1.34286954889262e-08
16:12223477-12223625	1	149	Gain		Duplicate	2.75043	1.63873	0.000659150537103415
17:20353276-20353442	1	167	Loss		Deletion	-1.4448	0	0.0516280271112919
17:20356320-20359994	4	3675	Loss		Deletion	-0.934711	0	0.00859882310032845
17:34523187-34524082	2	896	Loss		Deletion	-0.422512	0	0.412014037370682
17:44379994-44405886	6	25893	Loss		Deletion	-0.441892	0	0.187659740447998
17:44412901-44415099	4	2199	Loss		Deletion	-0.573289	0	0.0939922109246254
17:76198774-76198842	1	69	Gain		Duplicate	2.62781	1.42902	0.00116717966739088
19:1452991-1453151	1	161	Gain		Duplicate	2.7231	1.66166	0.000750205712392926
19:19748719-19748876	1	158	Gain		Duplicate	2.49192	1.44044	0.00213966937735677
20:1579460-1585715	3	6256	Loss		Deletion	-0.637669	0	0.113159634172916
22:39357382-39357696	1	315	Loss		Deletion	-1.80609	0	0.0154550056904554
X:153496007-153520480	6	24474	Loss		Deletion	-0.959531	0	0.059892263263464
X:154722008-154722371	2	364	Loss		Deletion	-0.738125	0	0.143895983695984

GOLDEN HELIX
Accelerating the Quest for Significance™

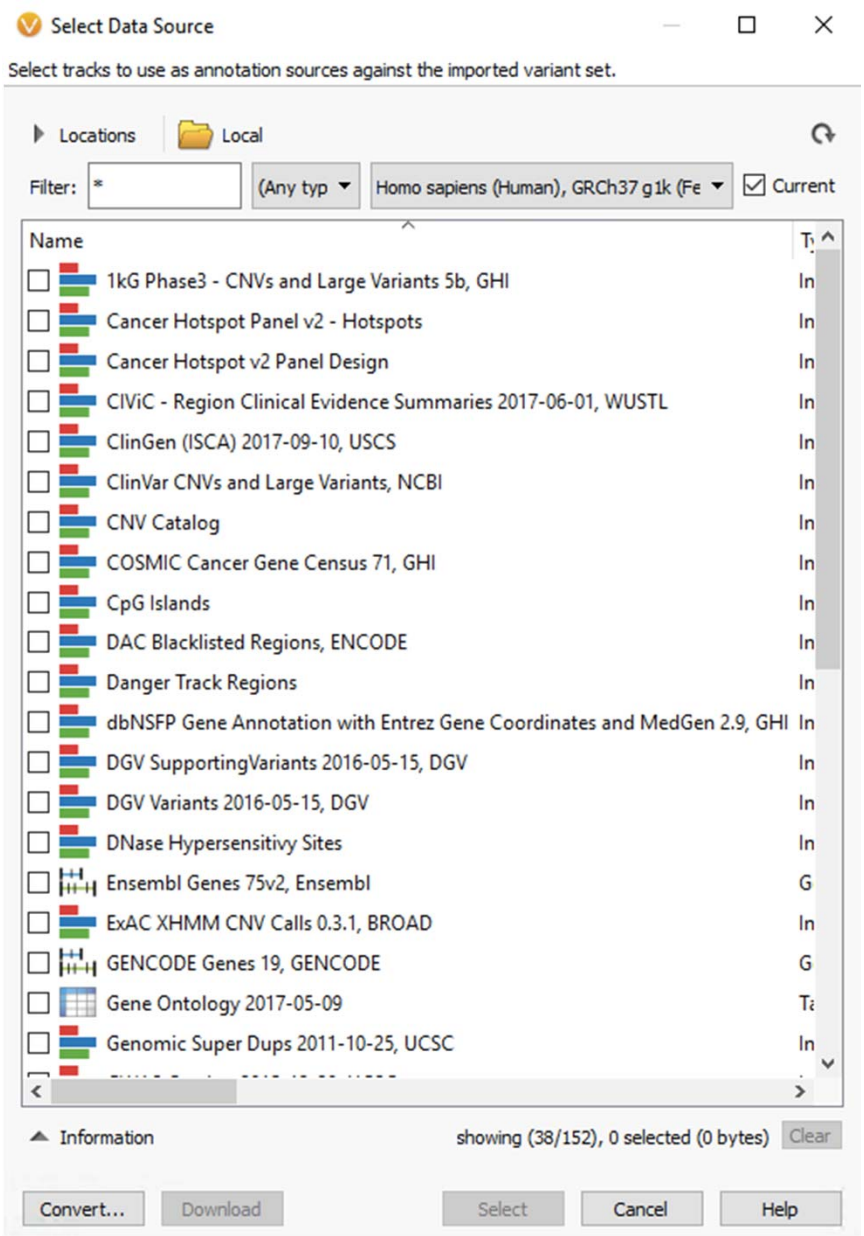
Filtering CNVs

- **Potentially remove "common"**
  - Want to have a match based on type (i.e. is this a common **Gain** or a common **Loss**)
  - Would like to know prevalence in different catalogs
- **Remove likely false-positives or common in internal cohort:**
  - Labs keep track of previously validated / reported CNVs
- **What genes are affected**
  - Including gene-based annotations like phenotype derived gene list
- **Clinical Classification / Disease Association**
  - Prioritize **Pathogenic**, **Likely Pathogenic**
  - Filter out **Benign, Likely Benign**
- **In regions of potential difficulty in genome**



Sources for Annotating CNVs

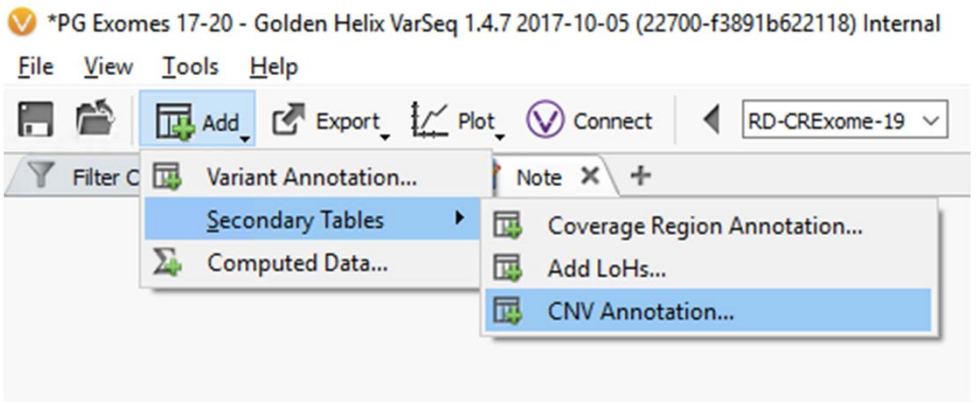
- **CNV calls in Populations:**
  - 1000 Genomes Phase3 Large Variants
  - ExAC per-sample CNV calls
  - DGV large-cohort studies
- **Clinical Interpretations:**
  - ClinVar Large Variants
  - ClinGen (Previously ISCA)
- **Genes**
  - Gene track, which transcripts/exons
  - Special considerations considering large sizes
- **Regions**
  - Genomic Superdups (Large Scale)
  - Low Complexity Regions (Smaller Scale)

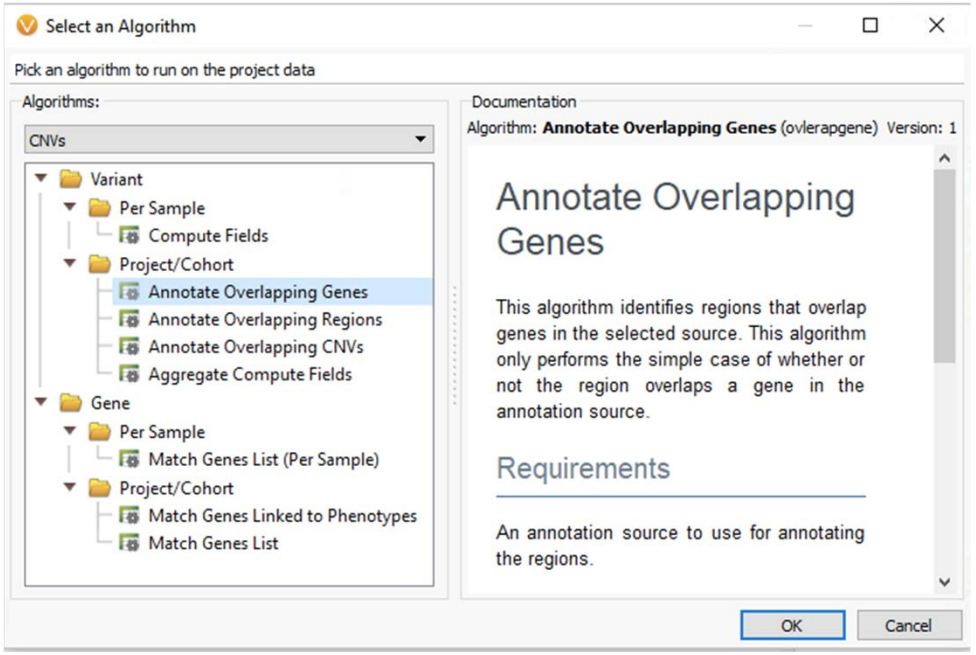


GOLDEN HELIX
Accelerating the Quest for Significance™

Annotation Algorithms: For CNVs / Coverage Regions

- **Need different algorithms than those used for variants**
- **Specialized to annotation data**
- **Annotate Overlapping Genes**
  - Genes and transcripts
- **Annotate Overlapping Regions**
  - Generic intervals
- **Annotate Overlapping CNVs**
  - Catalogs containing CNVs
- **Other algorithms work on "CNVs"**
  - Compute fields (combine / mutate fields)
  - Match Gene Linked to Phenotypes
  - Match Gene List

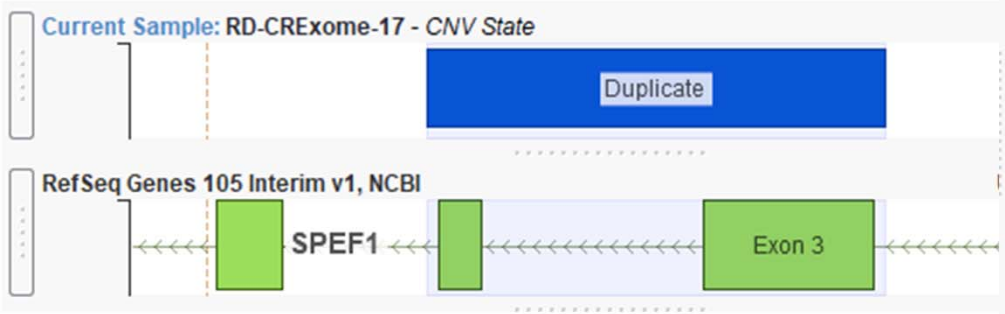




GOLDEN HELIX
Accelerating the Quest for Significance™

Annotation Algorithms: Overlapping Genes

Overlapping a few exons:



20:3759588-3760009 (422 bp)

Avg Z Score: 2.4057

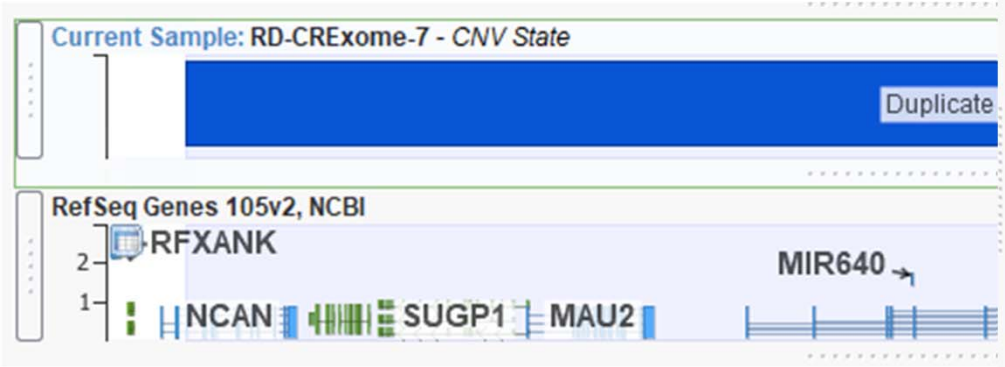
Overlapping Genes RefSeq Genes 105 Interim v1, NCBI

Gene Names	SPEF1
# Genes	1

Overlapping Transcripts RefSeq Genes 105 Interim v1, NCBI

	1
Transcript Name	NM_015417.4
Gene Name	SPEF1
% Covered	10.6781
Overlapping Exons	3-4

Overlapping many genes:



19:19329714-19779785 (450.1 Kbp)

Span: 450072

Overlapping Genes RefSeq Genes 105 Interim v1, NCBI

Gene Names	ATP13A1, CILP2, GATAD2A, GMIP, HAPLN4, LPAR2, MAU2, NCAN, NDUFA13, PBX4, SUGP1, TM6SF2, TSSK6, YJEFN3, ZNF101
# Genes	15

Overlapping Transcripts RefSeq Genes 105 Interim v1, NCBI

	1	2	3
Transcript Name	NM_004386.2	NM_001300949.1	NM_033204.3
Gene Name	NCAN	ZNF101	ZNF101
% Covered	82.7905	5.36668	1.27726
Overlapping Exons	3-15	1	1

Overlapping Tx Aux Fields RefSeq Genes 105 Interim v1, NCBI

copy

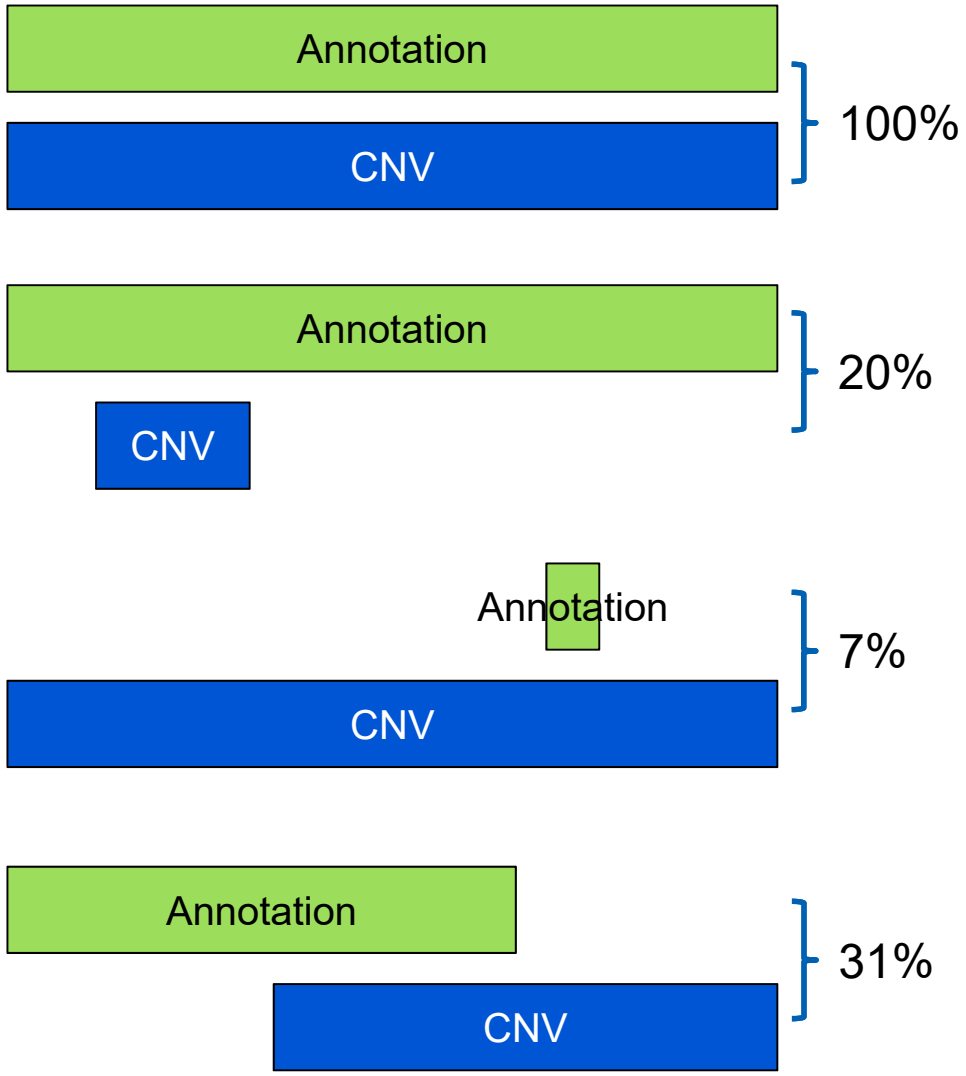
GOLDEN HELIX
Accelerating the Quest for Significance™

Annotation Algorithms: Overlapping Regions

- **Not expect exact matches**
- **Percent overlap not correct metric**
- **Need metric of "sameness"**
- **Jaccard index:**
  - "similarity coefficient"

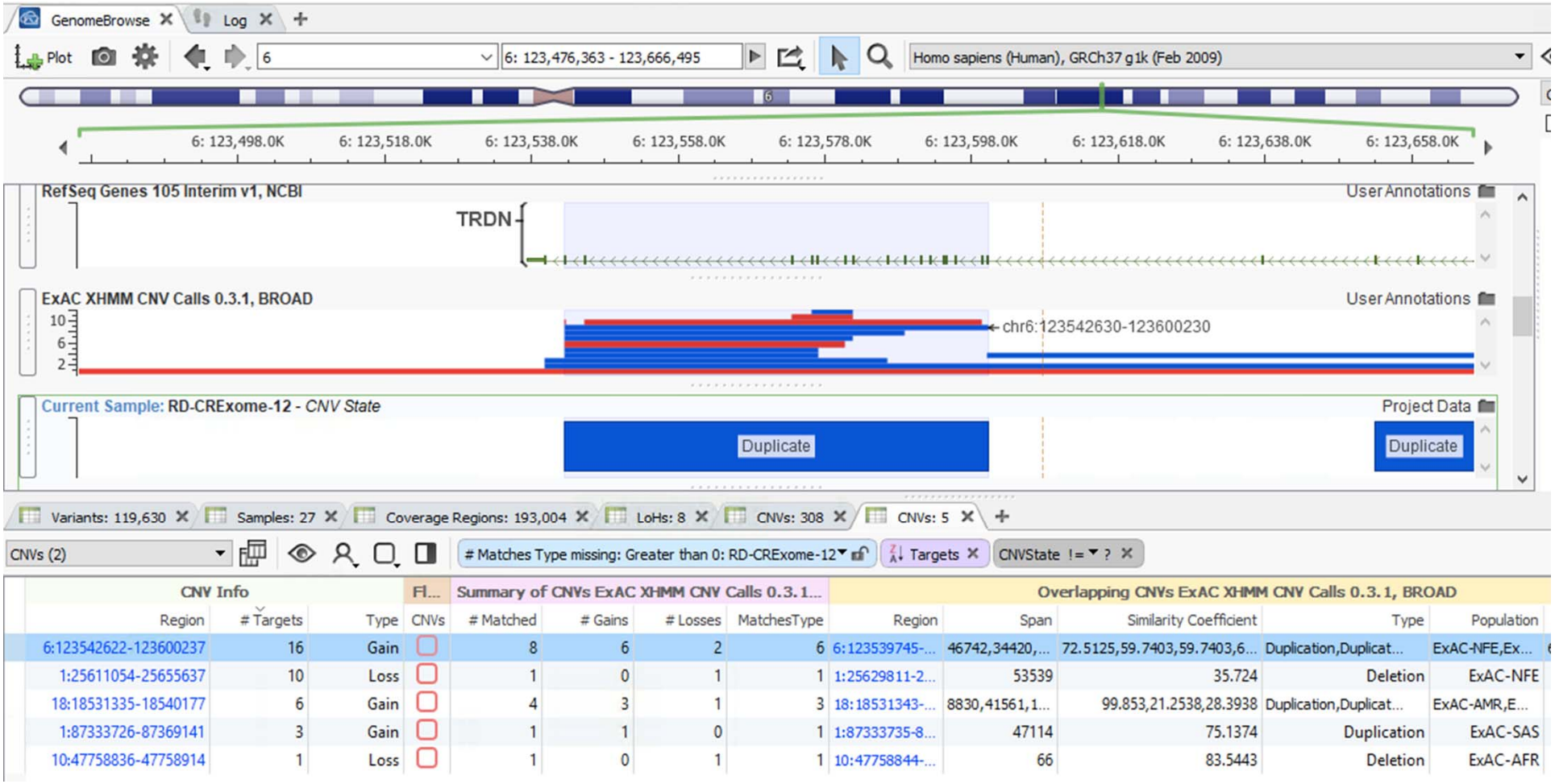
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

  - For fully overlapped regions, the percent overlap of the smaller to the larger
  - Default value of 20% for annotations
  - If set to 0%, then any overlap matches
  - If set to 100%, then exact matches



Annotation Algorithms: Overlapping CNVs

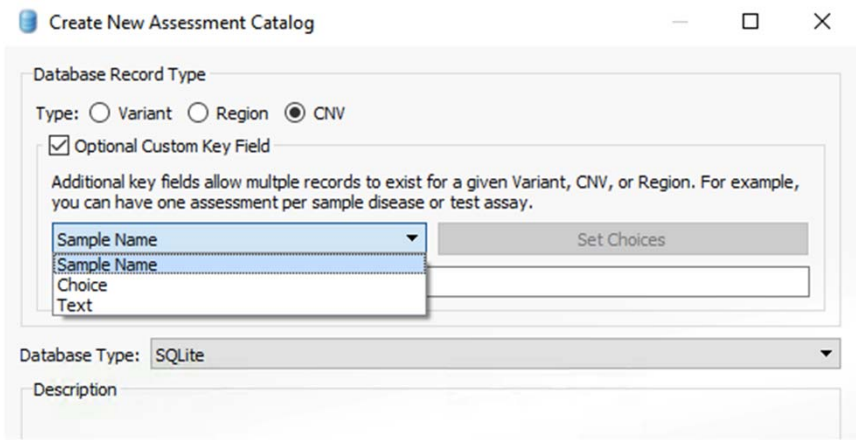
- **First Match based on Overlap Regions**
  - Use 20% similarity coefficient
- **Then Match the CNV Type to the CNV Type of the source:**

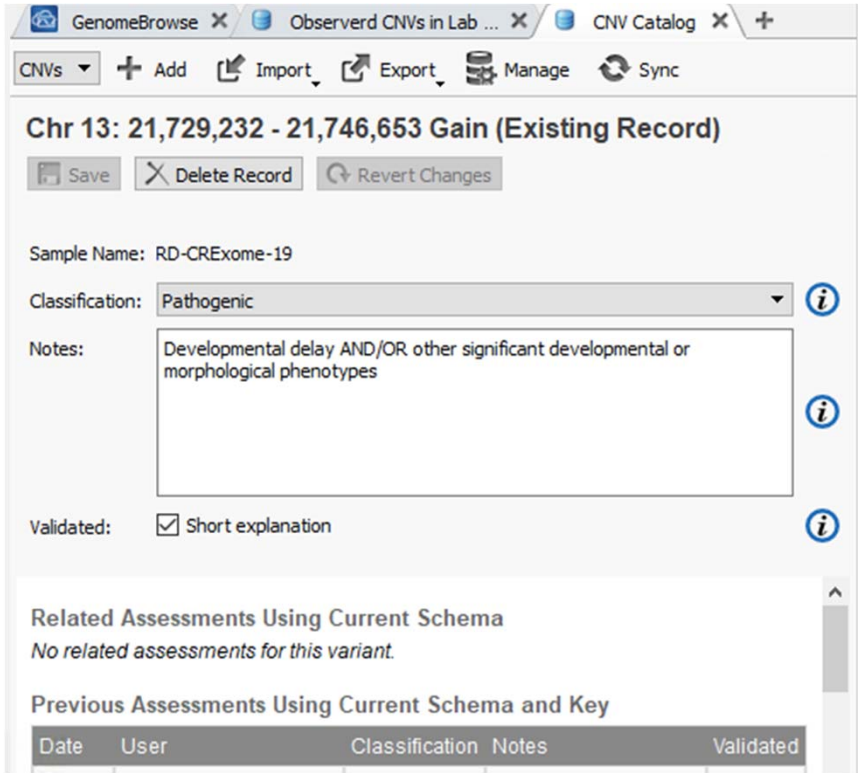


CNV Info			Fl...	Summary of CNVs ExAC XHMM CNV Calls 0.3.1...				Overlapping CNVs ExAC XHMM CNV Calls 0.3.1, BROAD				
Region	# Targets	Type	CNVs	# Matched	# Gains	# Losses	MatchesType	Region	Span	Similarity Coefficient	Type	Population
6:123542622-123600237	16	Gain		8	6	2	6	6:123539745-...	46742,34420,...	72.5125,59.7403,59.7403,6...	Duplication,Duplicat...	ExAC-NFE,Ex...
1:25611054-25655637	10	Loss		1	0	1	1	1:25629811-2...	53539	35.724	Deletion	ExAC-NFE
18:18531335-18540177	6	Gain		4	3	1	3	18:18531343-...	8830,41561,1...	99.853,21.2538,28.3938	Duplication,Duplicat...	ExAC-AMR,E...
1:87333726-87369141	3	Gain		1	1	0	1	1:87333735-8...	47114	75.1374	Duplication	ExAC-SAS
10:47758836-47758914	1	Loss		1	0	1	1	10:47758844-...	66	83.5443	Deletion	ExAC-AFR

Knowledge Capture for CNVs

- **Assessment Catalogs**
  - Capture knowledge
  - Custom field schema
  - Act as annotation/plot sources
- **Use Cases**
  - Clinical assessments
  - All QC'd CNVs catalog
  - Gene / Region interpretations
- **New "Custom Key"**
  - Can be auto-set to sample
  - Can set to phenotype
- **CNV Catalogs**
  - Automatically capture the CNV "Type" (Gain vs Loss)
  - Centralize on VSWarehouse





GOLDEN HELIX
Accelerating the Quest for Significance™

VarSeq Demonstration

GOLDEN HELIX
Enabling Precision Medicine

Looking Forward

- **Release Schedule**
  - VarSeq 1.4.7 (early October)
  - VSWarehouse 1.4 (late October)
- **Up Next for VarSeq:**
  - Improved clinical interpretation of variants with emphasis on accurate and up-to-date per-transcript in-silico predictions
  - Per-transcript splice site prediction algorithms in VarSeq
    - SpliceSiteFinder-like, MaxEntScan, GeneSplicer, HumanSplicingFinder, NNSplice
  - Pre-transcript functional prediction in VarSeq
    - SIFT, PolyPhen2
  - Per-base conservation scores and multi-species alignment

Announcements

- **Next Webcast: October 11th!**
  - CNV Annotations: User Experience
  - Building some of these workflows from scratch
  - More examples of the interpretation process
  - VSWarehouse Integration
- **Golden Helix at ASHG 2017!**
  - Booth 902
  - Come see demos and ask us questions!
- **See Gabe's talk at pre-ASHG meeting**
  - *"Rethinking the 5 splice site algorithms used in clinical genomics"*
  - October 17, HGVS meeting @ Hilton

Questions or more info:

- Email info@goldenhelix.com
- Request an evaluation of the software at www.goldenhelix.com