GOLDEN HELIX
Enabling Precision Medicine

CNV Annotations: a crucial step in your variant analysis

Darby Kammeraad
Field Application Scientist Manager

CIO Review
20 Most Promising Biotech Technology Providers

Gartner
Hype Cycle for Life sciences

pharma TECH OUTLOOK
Top 10 Analytics Solution Providers

NIH Grant Funding Acknowledgments

GOLDEN HELIX
Enabling Precision Medicine

- Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under:
    - Award Number R43GM128485-01
    - Award Number R43GM128485-02
    - Award Number 2R44 GM125432-01
    - Award Number 2R44 GM125432-02

    - Montana SMIR/STTR Matching Funds Program Grant Agreement Number 19-51-RCSBIR-005

- PI is Dr. Andreas Scherer, CEO Golden Helix.
- The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Who Are We?

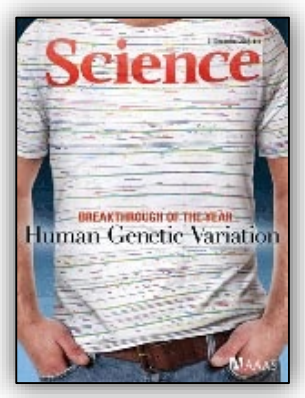
GOLDEN HELIX
Enabling Precision Medicine

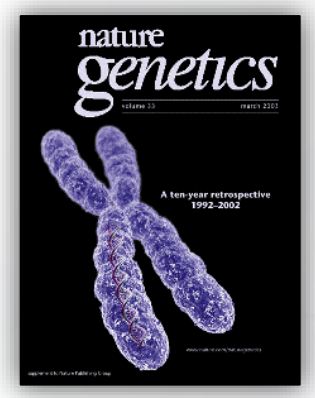
Golden Helix is a global bioinformatics company founded in 1998

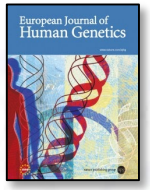
VarSEQ®	VariantWarehouse	SNP & Variation Suite
Filtering and Annotation	Variant Warehouse	CNV Analysis
ACMG Guidelines	Centralized Annotations	GWAS | Genomic Prediction
Clinical Reports	Hosted Reports	Large-N Population Studies
CNV Analysis	Sharing and Integration	RNA-Seq
Pipeline: Run Workflows		Large-N CNV-Analysis

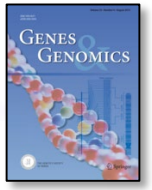
Cited in 1,000s of Peer-Reviewed Publications

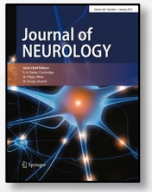
GOLDEN HELIX
Enabling Precision Medicine











Over 400 Customers Globally

GOLDEN HELIX
Enabling Precision Medicine

City of Hope
University of Copenhagen
Huntsman Cancer Institute University of Utah
fulgent
Rutgers The State University of New Jersey
UBC The University of British Columbia
Embrapa
pathology queensland
Northwell Health
the children's hospital at Westmead
Amplexa Genetics
eurofins
Ucla
Hill's
imegen
NSW Government Health Pathology
The University of Iowa
SickKids
Q² Solutions
Genomics for life
Mayo Clinic
Cobb ONE FAMILY. ONE PURPOSE.
National Cancer Institute
cancer research malaysia
Cincinnati Children's
Medizinische Universität Wien
Prevention Genetics DISEASE PREVENTION THROUGH GENETIC TESTING
Fundación Universitaria Sanitas
ALS Therapy Development Institute
UTHealth The University of Texas Health Science Center at Houston
National Eye Institute NATIONAL INSTITUTES OF HEALTH
NHS
MUSC Health Medical University of South Carolina
Shodair Children's Hospital
MEDPACE
zoetis
CPGR
Stanford University
USDA
Nemours
Boston Children's Hospital
Dahl-Chase Diagnostic Services
USP Universidade de São Paulo
tgen An affiliate of City of Hope
NUS National University of Singapore
NIH National Human Genome Research Institute

When you choose Golden Helix, you receive more than just the software

GOLDEN HELIX
Enabling Precision Medicine

SOFTWARE IS VETTED

- 20,000+ users at 400+ organizations
- Quality & feedback

DEEPLY ENGRAINED IN SCIENTIFIC COMMUNITY

- Give back to the community
- Contribute content and support

SIMPLE, SUBSCRIPTION-BASED BUSINESS MODEL

- Yearly fee
- Unlimited training & support

INNOVATIVE SOFTWARE SOLUTIONS

- Cited in 1,000s of publications

GENE PANEL | EXOME | GENOME

SEQUENCER

GOLDEN HELIX
Enabling Precision Medicine

PRODUCTS	BIOINFORMATICS PIPELINE	FUNCTION
DNASEQ (Sentieon) TNSEQ (Sentieon) VS-CNV	FASTQ BAM VCF	▸ Single nucleotide variation ▸ Copy number variation & loss of heterozygosity ▸ Chromosomal aberration
Annotations	Annotated VCF	▸ Public & commercial annotations to enrich genomic data sets
VarSeq VSReports VSPipeline	Clinical Report	▸ Annotate & filter ▸ Visually inspect alignments ▸ Variant prioritization ▸ Clinical assessment
VSClinical	Automated ACMG Guidelines	▸ Clinical variant interpretation in coordination with ACMG Guidelines & AMP Guidelines
VSWarehouse	Data Warehousing Web-Enabled Interface + Powerful APQ: JSON, XML, TSV, CSV, SQL, FHIR	▸ Clinical assessment catalog ▸ Advanced data querying ▸ Versioning ▸ Interoperability ▸ Compliance with HIPPA, CLIA & CAP data discovery

VarSeq Suite

GOLDEN HELIX
Enabling Precision Medicine

VARSEQ®

Simple

Flexible

Scalable

Variant annotation
filtering, and interpretation

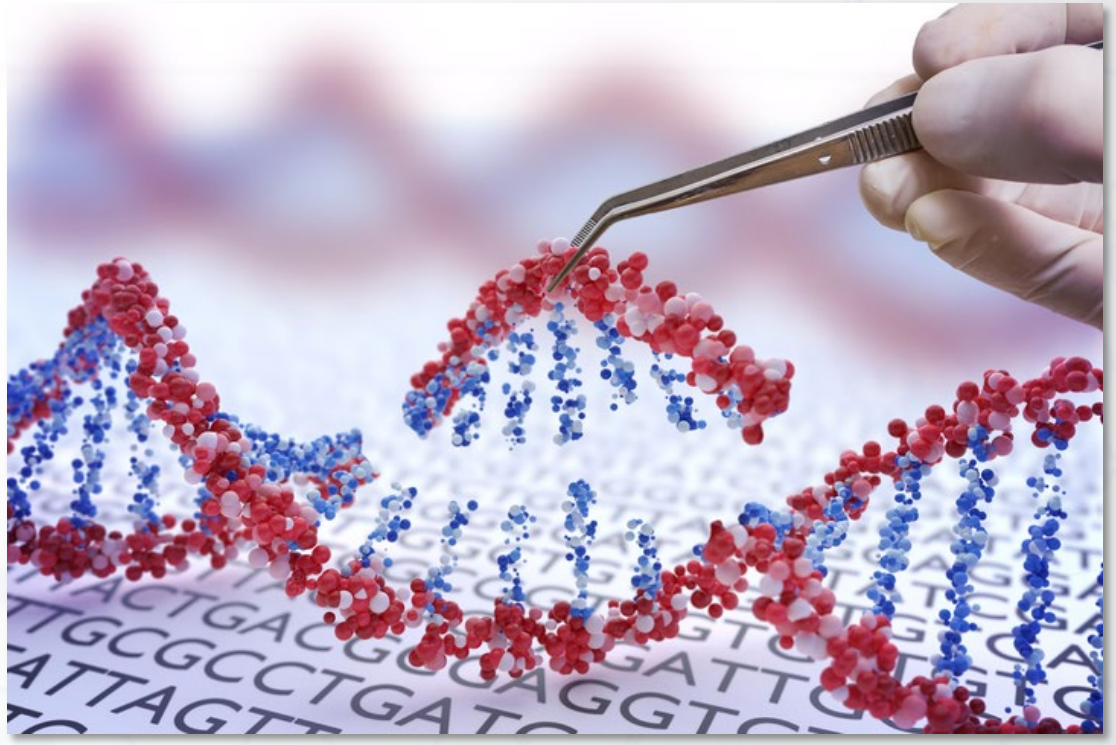
Powerful GUI with
rich visualizations

Repeatable
workflows

CNVs in Clinical Testing

GOLDEN HELIX
Enabling Precision Medicine

- Critical evidence needed for many genetic tests
- Common driver specific cancers, causal hereditary variation
  - EGFR Exon 19 deletion common in lung cancer
  - PIK3CA Amplification in breast cancer
- Large events used heavily in diagnostics
  - Chromosome 13 deletion common in melanoma
  - Autism Spectrum Disorder (ASD)
  - Developmental Delay (DD)
  - Intellectual Delay (ID)



Power of NGS CNV Detection

GOLDEN HELIX
Enabling Precision Medicine

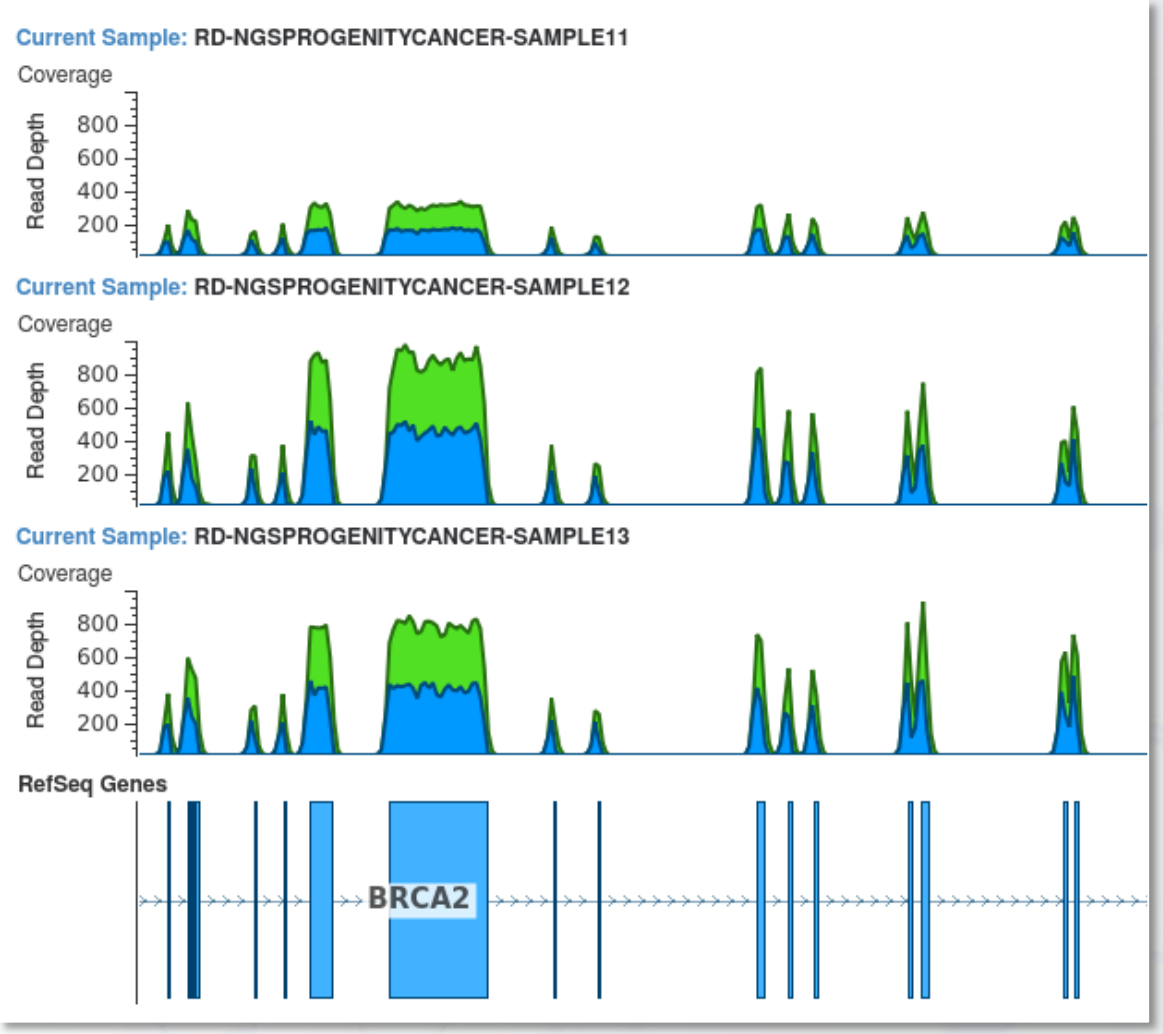
	Detectable events			Supported Data types		
	Small: 150b+	Medium: 1 – 10Kb	Large: 10Kb+	Gene panel	Whole exome	Whole genome
MLPA	✓			✓		
CMA			✓			✓
VS-CNV	✓	✓	✓	✓	✓	✓

- ✓ One single testing paradigm
- ✓ True simplification of clinical workflow
- ✓ Saves time and money – all on site

Addressing Issues - CNV Detection via NGS

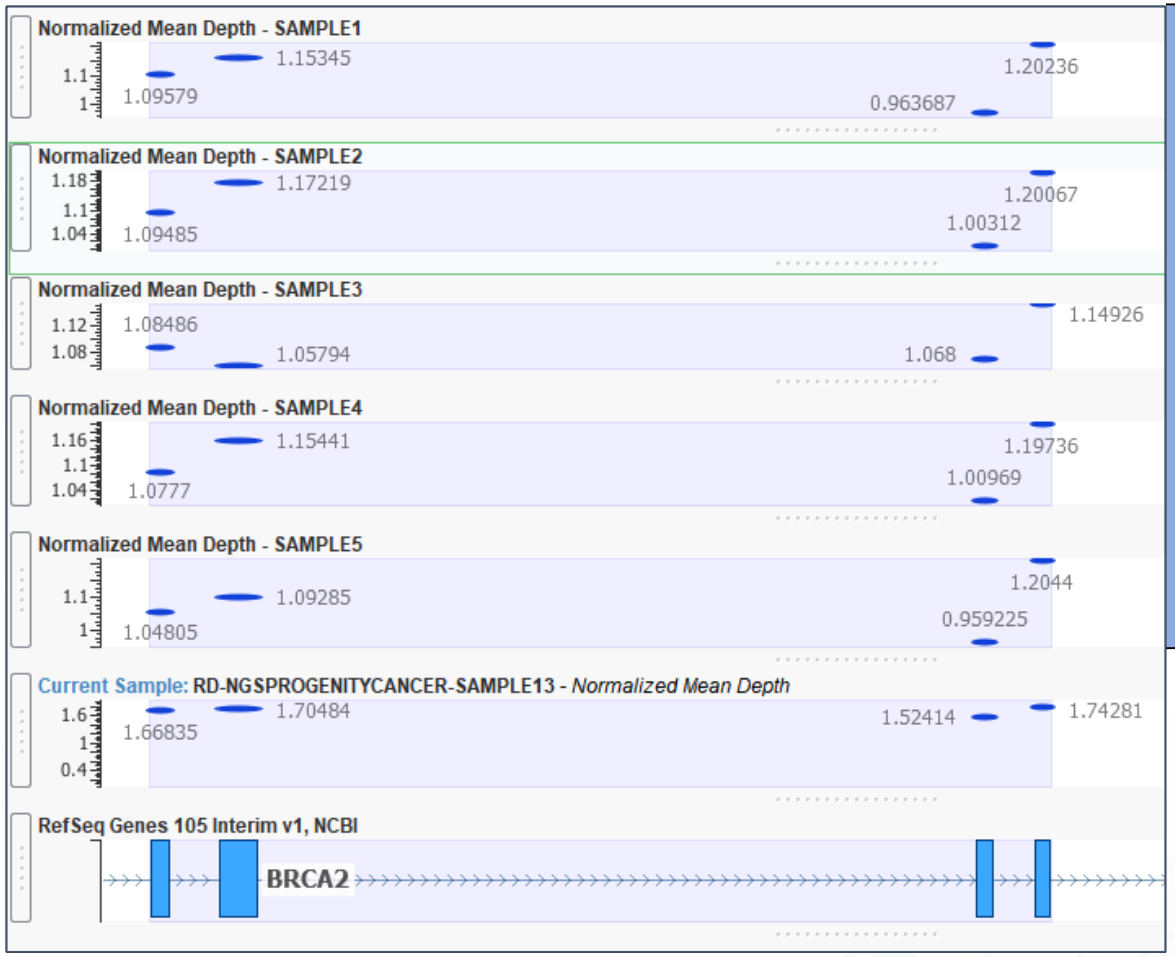
GOLDEN HELIX
Enabling Precision Medicine

- CNVs detected from coverage data in BAM
- Challenges
  - Coverage varies between samples
  - Coverage fluctuates between targets
  - *Systematic biases impact coverage
- Solutions
  - Data Normalization
  - Reference Sample Comparison
  - Algorithm works without case/control data
- <u>Requirements</u>
  - ≥ 30 ref samples
  - From same library prep method
  - Ideally ≥100X coverage



Principle Approach to CNV Calling

GOLDEN HELIX
Enabling Precision Medicine



- Reference samples:
  - Coverage **normalization** and **averaging** represents diploid/normal regions for comparison
  - Reference set is unique for each sample
  - Reference set is selected based on similarity to the sample
  - Non-autosomal regions matched for gender automatically

Sample of Interest

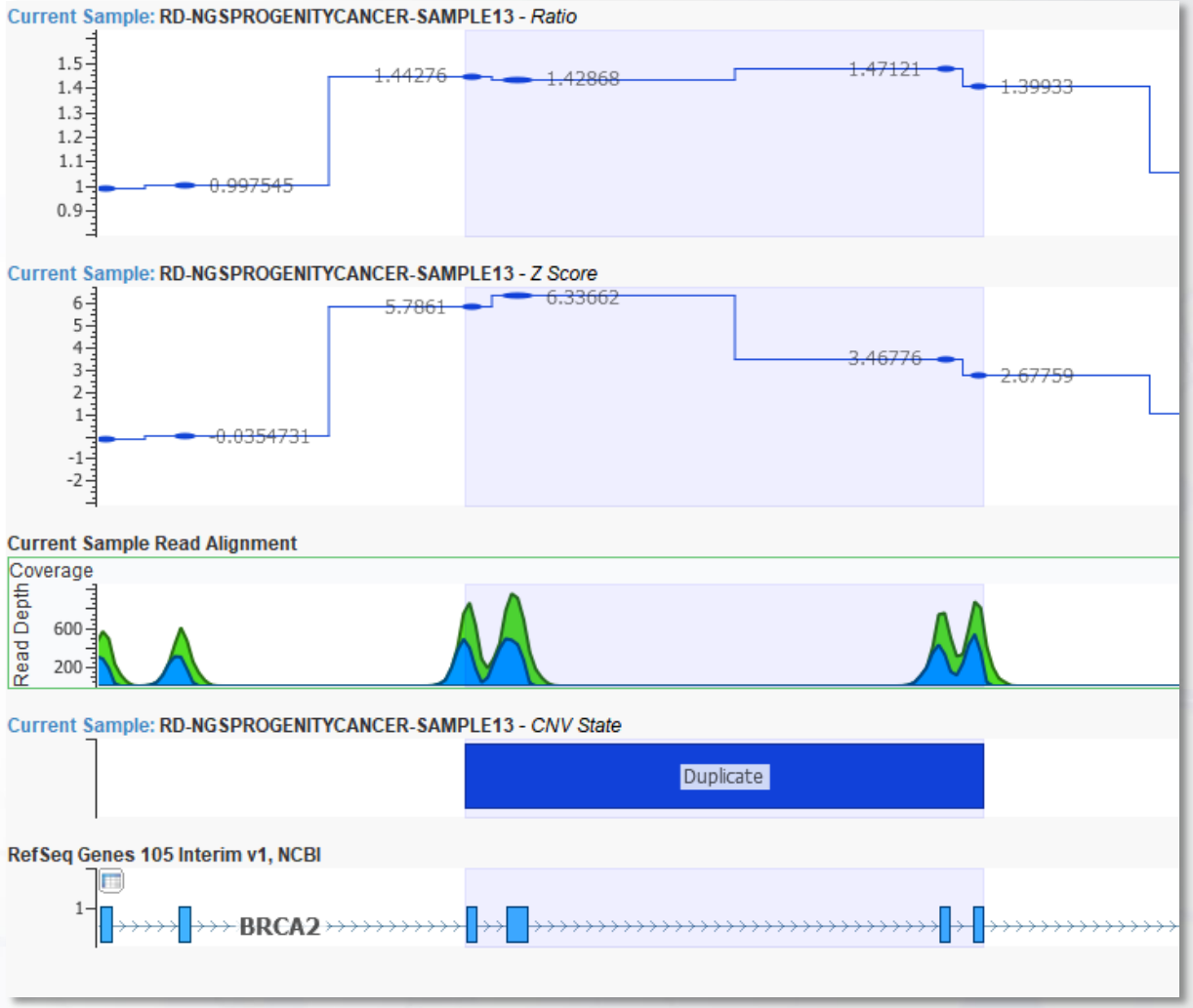
Sample 13 normalized coverage

Reference set normalized coverage

Coverage Region I...	Overlapp...	RD-NGSP...	Target Copy Number State for RD-NGSPROGENITYCANCER-SAMPLE13		
Region	Gene Names	Mean Depth	CNV State	Normalized Mean Depth	Avg. Normalized Control Depth
13:32936641-32936850	BRCA2	810.281	Duplicate	1.66835	1.17383
13:32937297-32937690	BRCA2	828.008	Duplicate	1.70484	1.20154
13:32944520-32944714	BRCA2	740.241	Duplicate	1.52414	1.02256
13:32945074-32945257	BRCA2	846.446	Duplicate	1.74281	1.22956

CNV Detection: Ratio, Z-score, and VAF

- Metrics
  - Ratio: sample coverage divided by reference sample mean
  - Z-score: standard deviations from reference sample mean
  - VAF: Variant Allele Frequency
- For Gene Panels and Exomes
  - Probabilistic model used to call CNVs
  - Segmentation identifies large cytogenetic events
- For Whole Genome Data
  - Targets segmented using Z-scores
  - Events called based on Z-score and Ratio thresholds

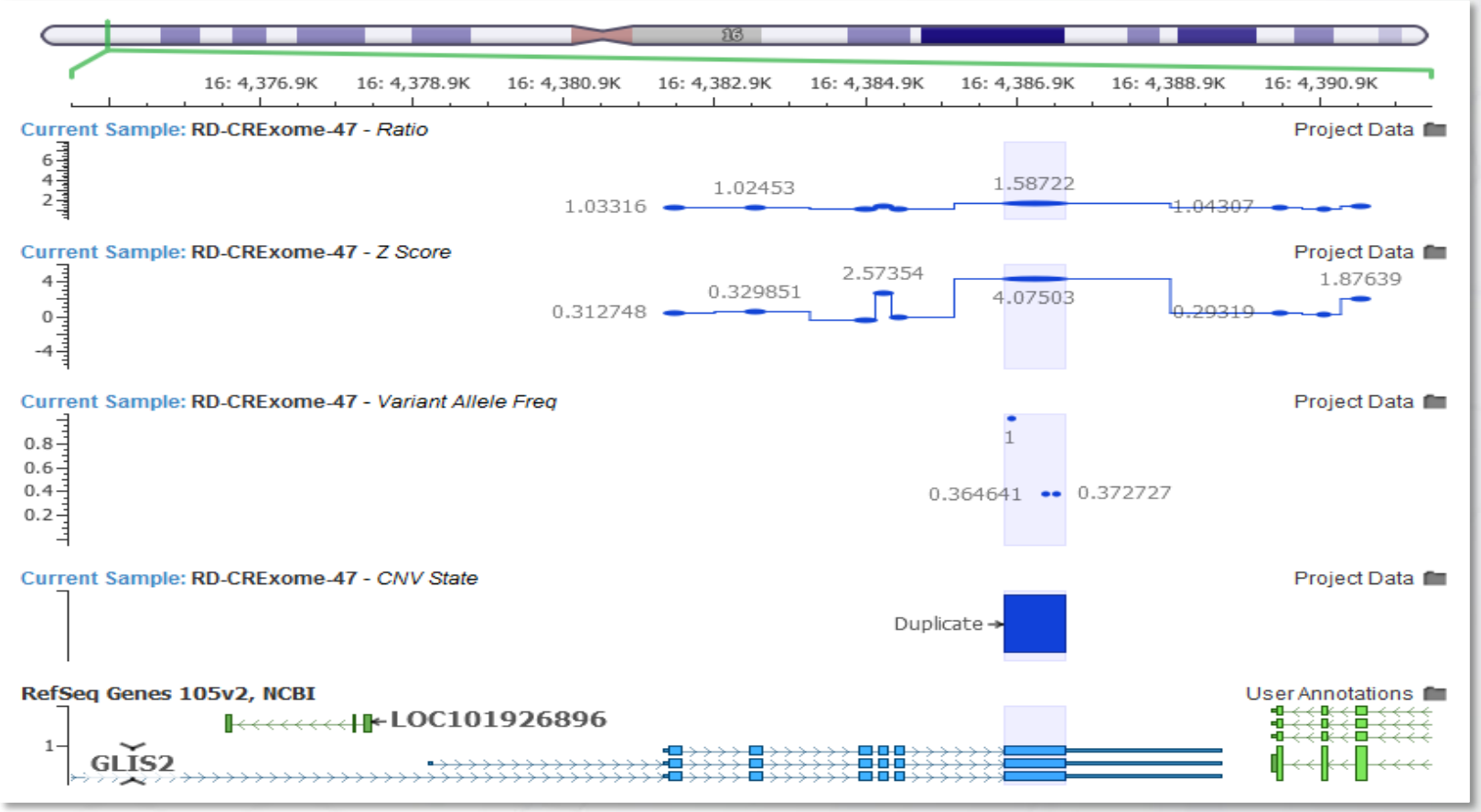


Optimizing CNV Detection - Segmentation

- Metrics are noisy over large regions
- Outliers cause large events to be called as many small events
- Solved using segmentation:
  - Regions containing many events are segmented
  - Small events sharing a segmented region are merged

VAF Provides Supporting Evidence

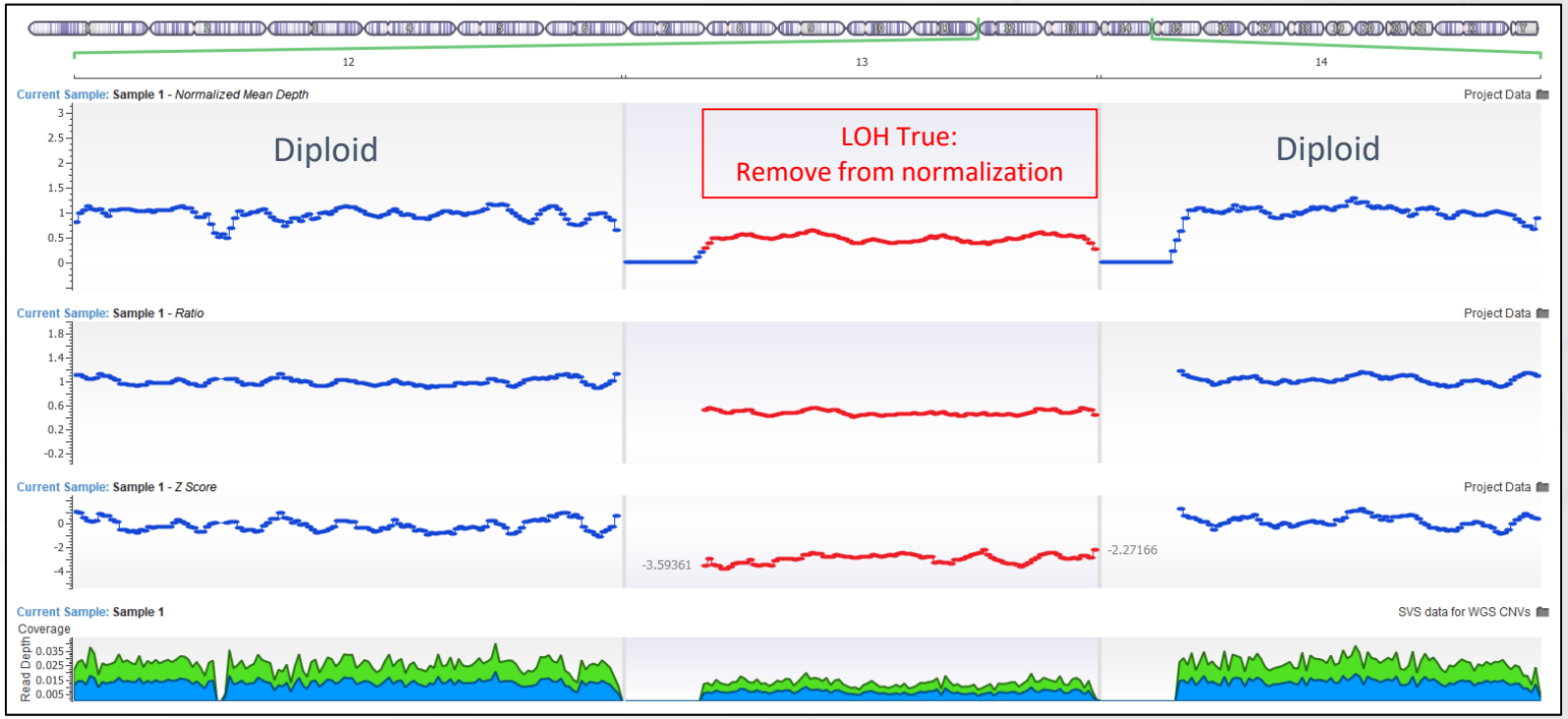
- Values other than 0 or 1 are evidence against het deletions
- Values of 2/3 and 1/3 are evidence for duplications



Advanced Optimization for CNV Detection - LOH

GOLDEN HELIX
Enabling Precision Medicine

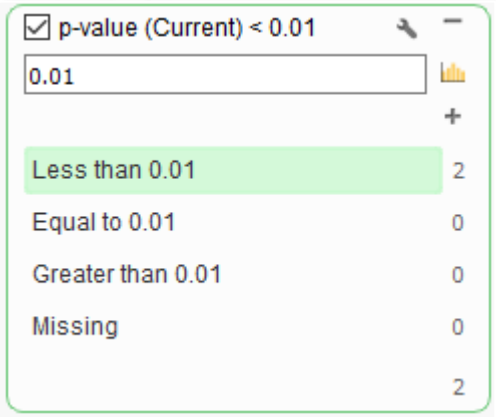
- **Issue** - Large chromosomal deletions and duplications can skew the mean coverage of a sample
  - The previous approaches don't account for this
- **Solution** - Detection of LOH areas to optimize definition of normal regions
  - Prior to running CNV caller
  - Probabilistic model based on VAF (Whole Exome/Genome)
  - Identifies and excludes non-diploid regions from normalization
  - Quality control step to improve CNV caller



CNV Confidence: P-Values

GOLDEN HELIX
Enabling Precision Medicine

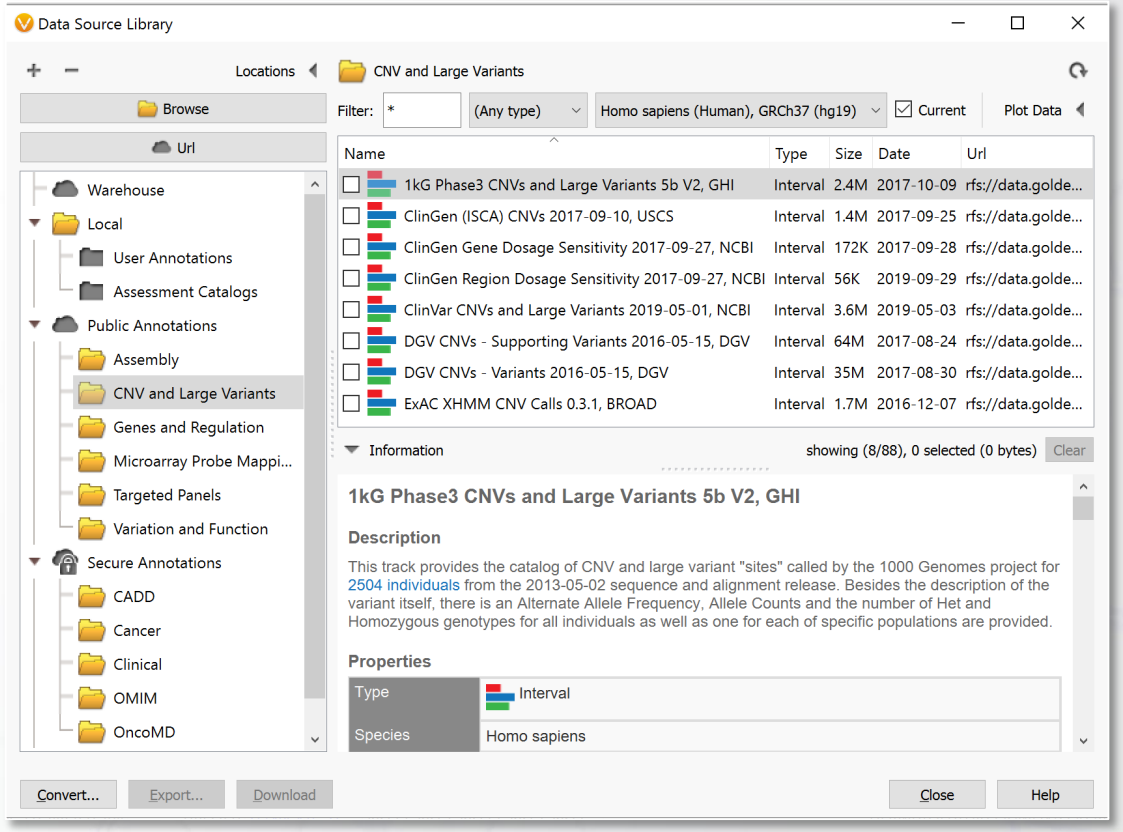
- P-Values
  - Introspective ability to define confidence in the CNV event
  - Confidence threshold definable in your workflow/protocol
  - You can modify the threshold at any point
  - Lower the p-value/probability the higher the confidence the event is real
- Typical p-values
  - <0.05*
  - <0.01**
  - <0.001***



CNV Info						
Region	Type	# Targets	# Samples	Span	CNV State	p-value
11:108160310...	Loss	21	1	39676	Het Deletion	4.68818626586653e-07
11:108128189...	Gain	9	1	22167	Duplicate	8.93228134373203e-05
17:29683459-...	Gain	6	1	4283	Duplicate	0.00131093873642385
8:90965453-9...	Gain	5	1	17353	Duplicate	0.0050437287427485
3:37053483-3...	Gain	5	1	14036	Duplicate	0.00561221409589052
17:56769986-...	Gain	4	1	10725	Duplicate	0.00574351288378239
11:64573088-...	Gain	4	1	1624	Duplicate	0.0100980764254928
17:41219606-...	Gain	4	1	9046	Duplicate	0.0126859014853835
17:41201106-...	Gain	4	1	14305	Duplicate	0.0156013108789921
11:108151711...	Loss	4	1	6752	Het Deletion	0.0184763930737972
11:108121409...	Gain	3	1	2251	Duplicate	0.0187129583209753
3:37045873-3...	Loss	4	1	7501	Het Deletion	0.0200324188917875

CNV Workflow and Annotations

- Workflow Issue: How to screen through many CNV events?
  - Especially relevant with WES
  - What steps are necessary to find events relevant to the patient?
- First step
  - Prioritize CNVs specific to sample
  - High quality
  - High confidence
- Second step
  - DGV CNVs - Exclude CNVs in known healthy individuals
  - Genomic Sup Dups - Exclude CNVs in known duplication regions
  - ExAC + 1kG Phase3 - Eliminate common CNVs
  - ClinGen + ClinVar - Eliminate known benign CNVs
- Third step
  - Prioritize individual sample phenotype + gene list



GOLDEN HELIX
Enabling Precision Medicine

Project Demonstration

NIH Grant Funding Acknowledgments

GOLDEN HELIX
Enabling Precision Medicine

- Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under:
    - Award Number R43GM128485-01
    - Award Number R43GM128485-02
    - Award Number 2R44 GM125432-01
    - Award Number 2R44 GM125432-02

    - Montana SMIR/STTR Matching Funds Program Grant Agreement Number 19-51-RCSBIR-005

- PI is Dr. Andreas Scherer, CEO Golden Helix.
- The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

GOLDEN HELIX
Enabling Precision Medicine

We're headed to ESHG 2019 in June!