GOLDEN HELIX

Enabling Precision Medicine

Annotating and Cataloging CNVs in VarSeq

Dr. Nathan Fortier – Director of Research

CIOReview

20 most promising Biotech Technology Providers

pharma TECH OUTLOOK

Top 10 Analytics Solution Providers

Gartner

Hype Cycle for Life sciences

Q & A

Please enter your questions into your GoToWebinar panel

NIH Grant Funding Acknowledgments

- 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
  - Award Number 2R44 GM125432-01
  - Award Number 2R44 GM125432-02
- 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 – Who We Are

Golden Helix is a global bioinformatics company founded in 1998.

GlaxoSmithKline

varSEQ®

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

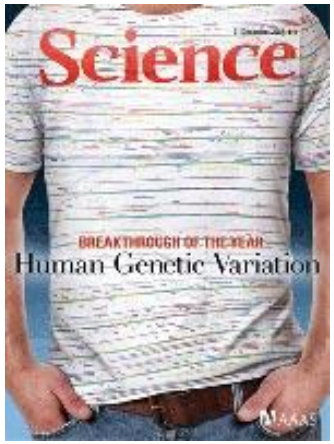
VARIANT WAREHOUSE

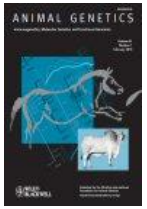
- Variant Warehouse
- Centralized Annotations
- Hosted Reports
- Sharing and Integration

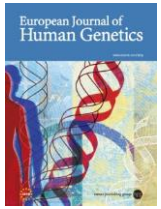
SNP & VARIATION SUITE

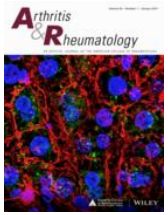
- CNV Analysis
- GWAS
- Genomic Prediction
- Large-N-Population Studies
- RNA-Seq
- Large-N CNV Analysis

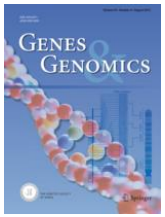
Cited in over 1300 peer-reviewed publications











GOLDEN HELIX
Enabling Precision Medicine

Over 400 customers globally

SickKids THE HOSPITAL FOR SICK CHILDREN

Cincinnati Children's Hospital Medical Center

The Children's Hospital of Philadelphia

Stanford University

UBC

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

USDA Agricultural Research Service

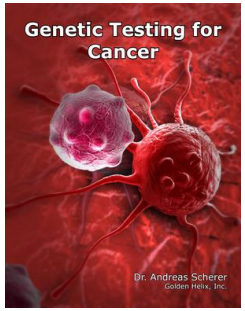
FDA

GOLDEN HELIX Enabling Precision Medicine

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

- INNOVATION and SPEED

GENE PANEL | EXOME | GENOME

SEQUENCER

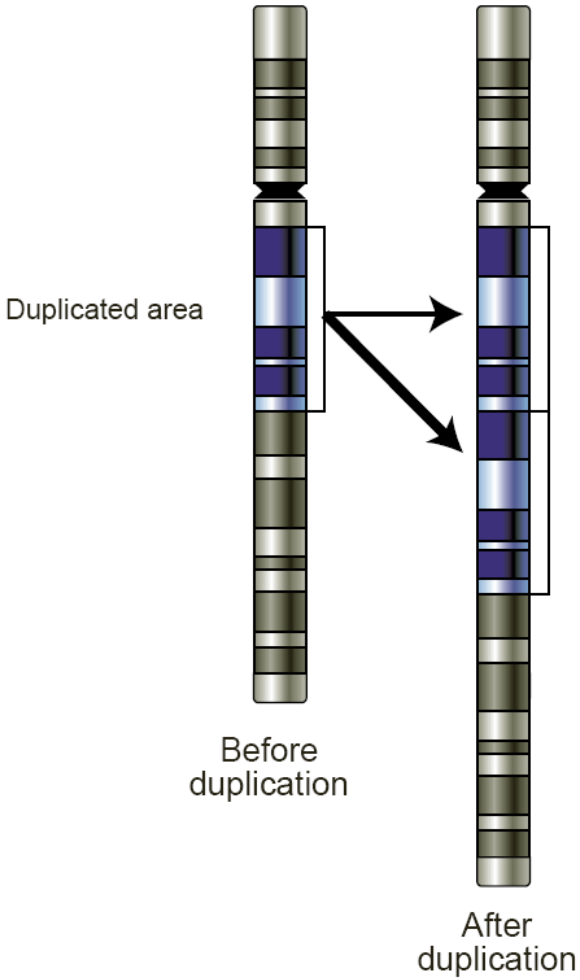
PRODUCTS	BIOINFORMATICS PIPELINE	FUNCTION
DNAseq (Sentieon) TNseq (Sentieon) VS-CNV	FASTQ BAM VCF	▸ Single nucleotide variation ▸ Copy number variation & loss of heterozygosity ▸ Chromosomal aberration
Annotations	Annotated VCF	▸ Public & commercial annotations to enrich genomic data sets
VarSeq VSReports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

CNVs in Clinical Testing

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

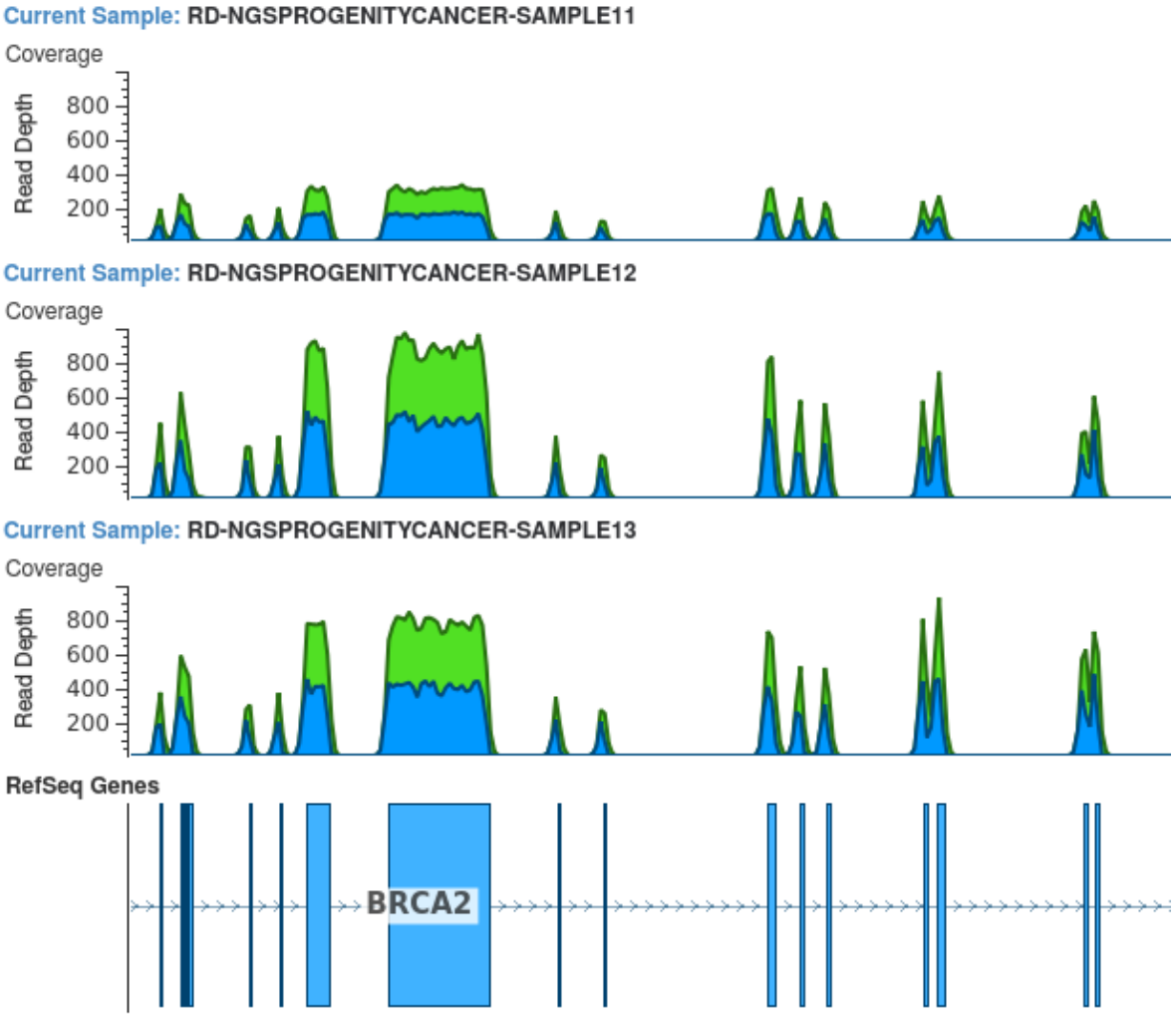
CNV Detection

- **Chromosomal microarray**
  - Current best practice
  - Slow
  - Additional expense
  - Only detects large events
- **CNV calling from NGS data**
  - Calls from existing coverage data
  - Detects small single-exon events
  - Provides faster results, simplified clinical workflow



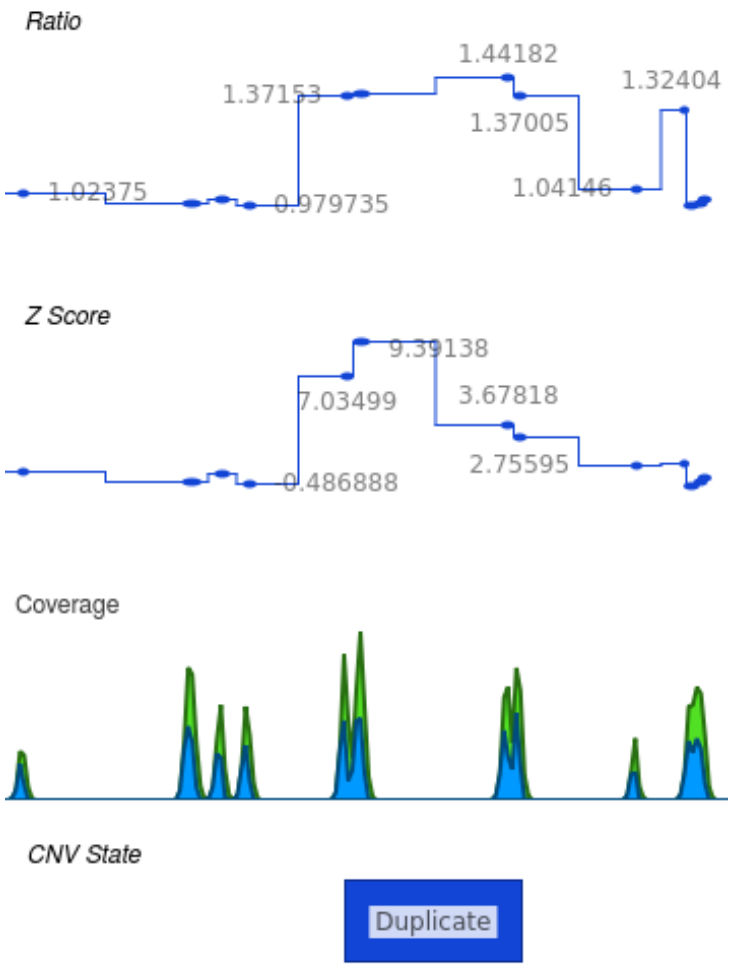
CNV Detection via NGS

- **CNVs are called from coverage data**
- **Challenges**
  - Coverage varies between samples
  - Coverage fluctuates between targets
  - Systematic biases impact coverage
- **Solutions**
  - Data Normalization
  - Reference Sample Comparison



CNV calling in VarSeq

- **Reference samples used for normalization**
- **Metrics**
  - Z-score: number of standard deviations from reference sample mean
  - Ratio: sample coverage divided by 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



QC Events

- **Low quality events can be flagged if**
  - Event targets have low coverage
  - There is high variation between samples at event targets
  - Event cannot be differentiated from noise at a region

- **Samples can be flagged if**
  - The sample does not match the references
  - The sample has extremely low coverage
  - There is high variance across the target regions

- **Filtering flagged events improves precision**

Reference Samples

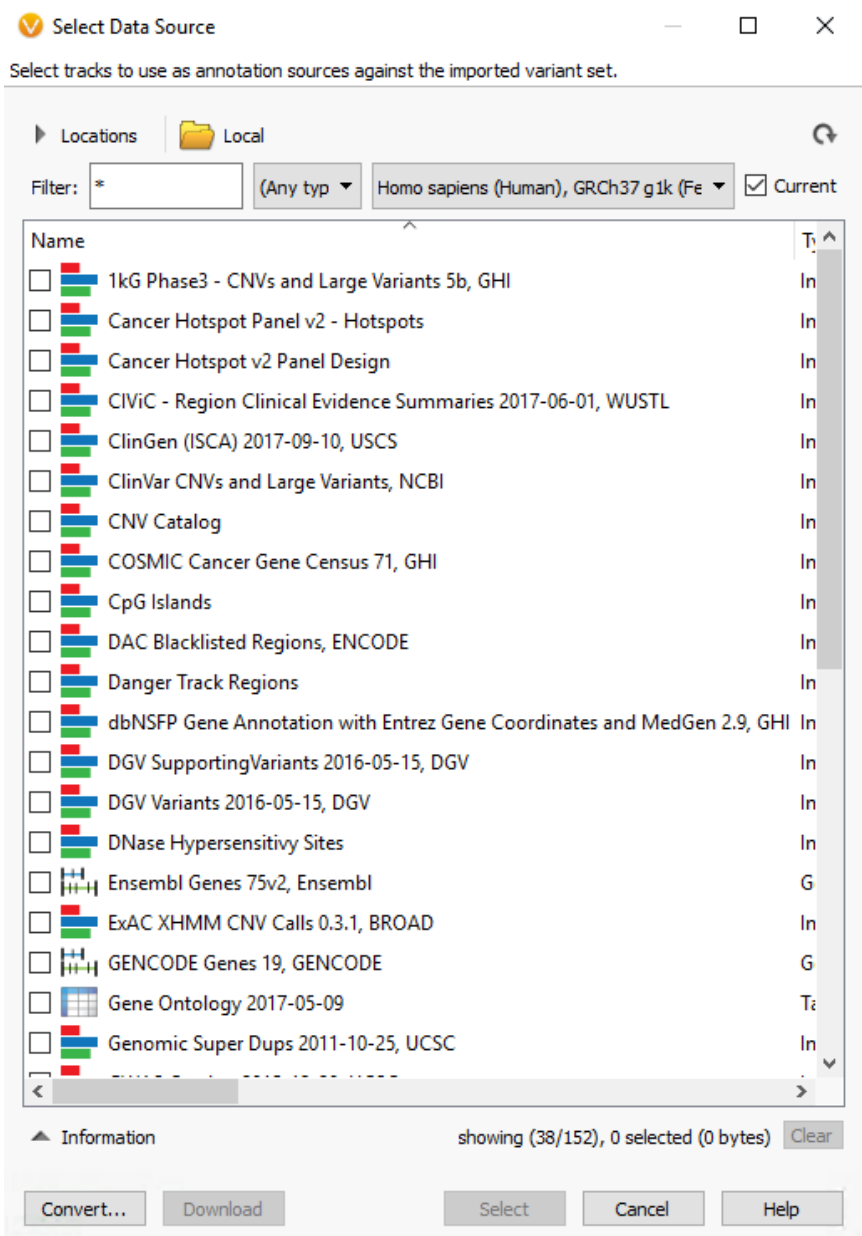
- Match references are chosen for each sample
- Samples with lowest percent difference chosen
- Performance affected if controls don’t have matching coverage profile
- Samples are flagged if the average percent difference is above 20%

Requirements

- **100x Coverage**
- **Reference samples**
  - Recommend at least 30 references
  - From same platform and library preparation
  - Automatic gender matched references for non-autosomal calls

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)



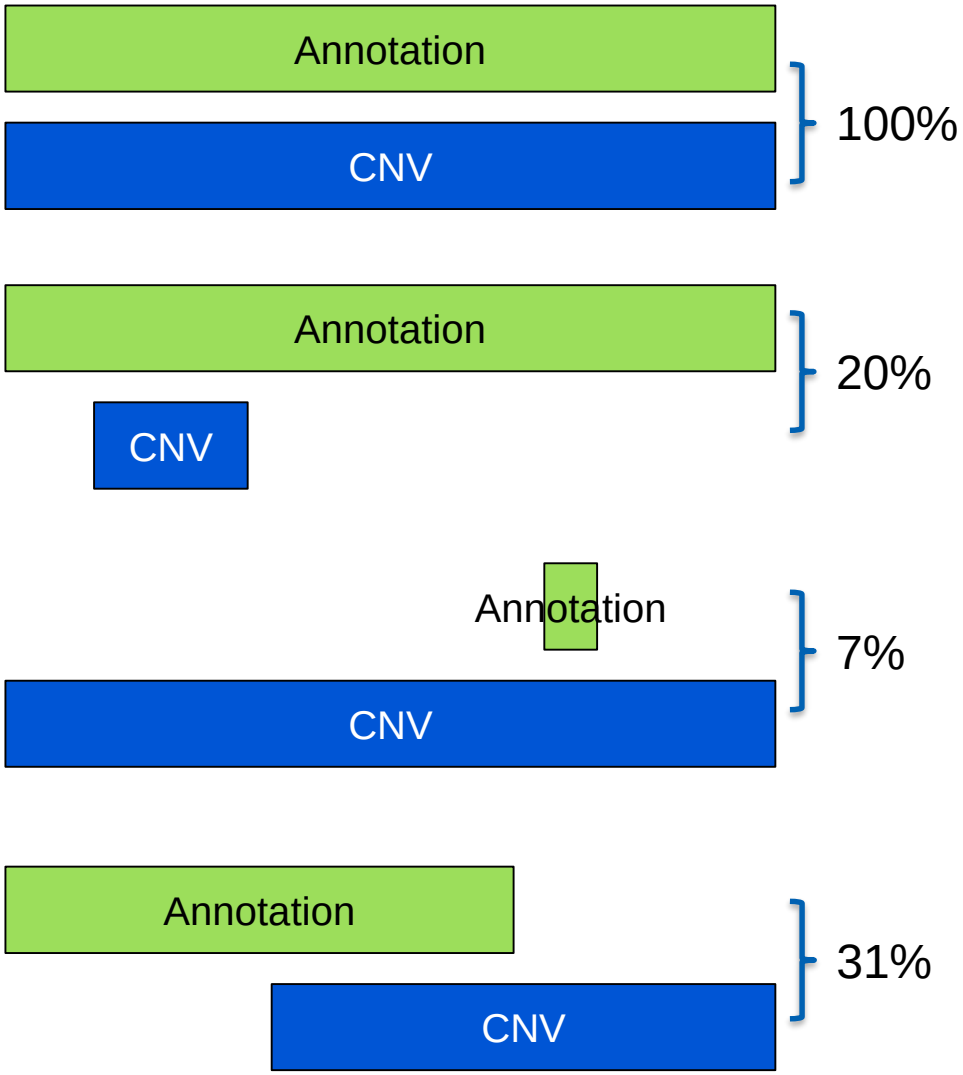
GOLDEN HELIX
Enabling Precision Medicine

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



VarSeq Demonstration

NIH Grant Funding Acknowledgments

- 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
  - Award Number 2R44 GM125432-01
  - Award Number 2R44 GM125432-02
- 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.

Q & A

See us at AMP 2018 in Booth #1801

November 1-3, 2018 | San Antonio, TX

Booth 1801

5-minute demos at the Golden Helix booth

Free t-shirts & chance to win an iPad Pro

GOLDEN HELIX
Enabling Precision Medicine

Q & A