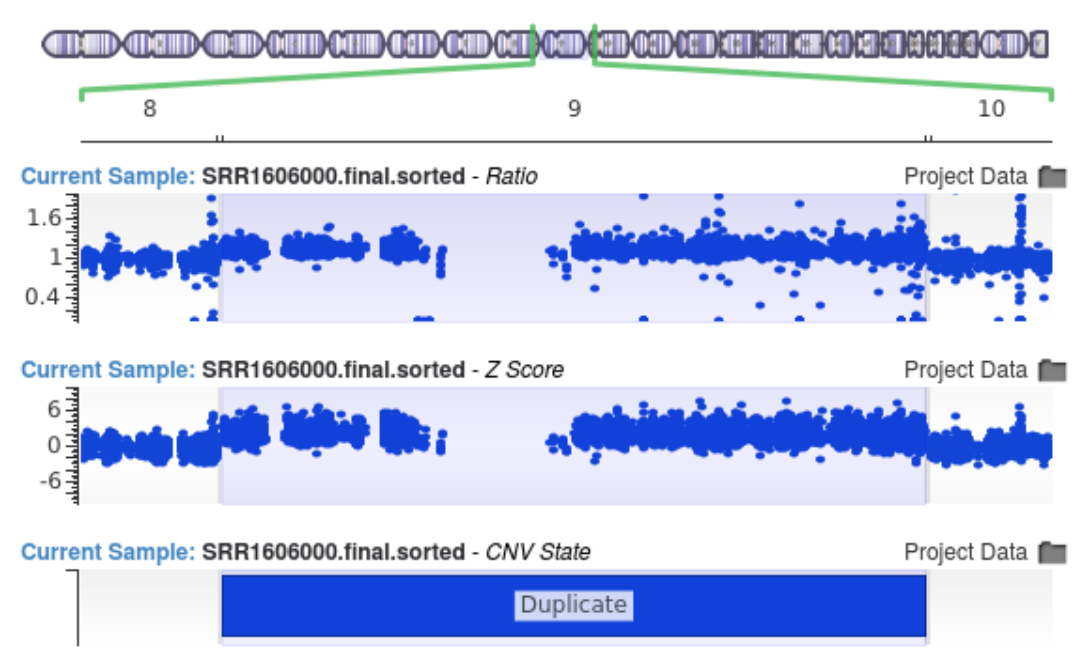
CNV Analysis in VarSeq – A User's Perspective



April 19, 2017

Dr. Nathan Fortier
Senior Software Engineer
& Field Application Scientist

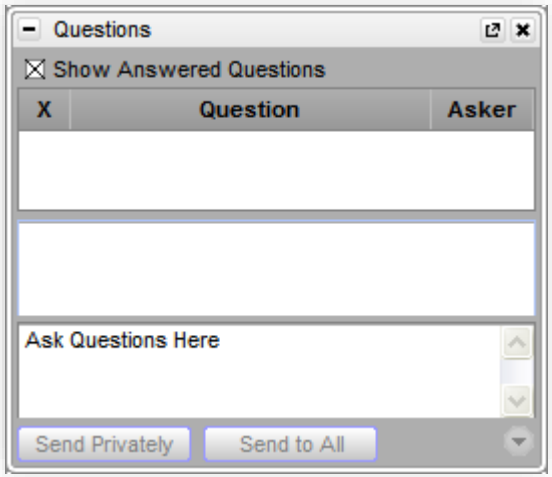
GOLDEN HELIX
Enabling Precision Medicine

Agenda

1. Overview Golden Helix
2. CNV Analysis – A User's Perspective
3. Method and Demo
4. Roadmap

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®

- Filtering and Annotation
- Single Sample CNV-Analysis
- Clinical Reports
- 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

Over 300 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

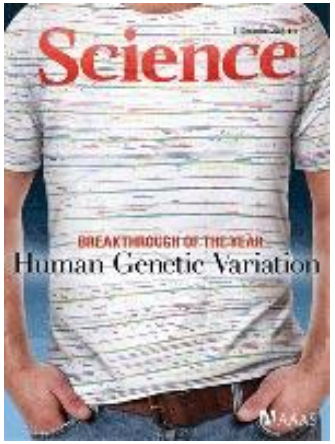
NATIONAL CANCER INSTITUTE

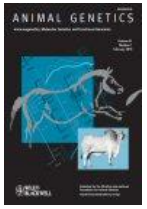
USDA
USDA Agricultural Research Service

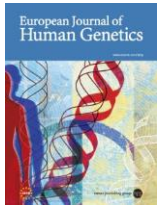
FDA

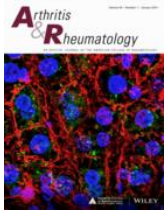
GOLDEN HELIX
Accelerating the Quest for Significance™

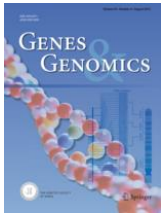
Cited in over 1000 peer-reviewed publications







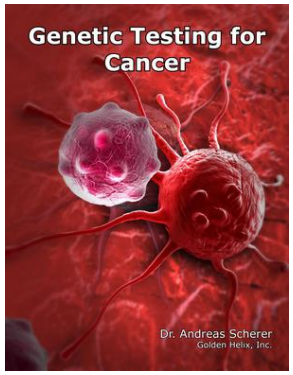




Golden Helix – Who We Are

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

- REPUTATION
- TRUST
- EXPERIENCE



- INDUSTRY FOCUS
- THOUGHT LEADERSHIP
- COMMUNITY

- TRAINING
- SUPPORT
- RESPONSIVENESS

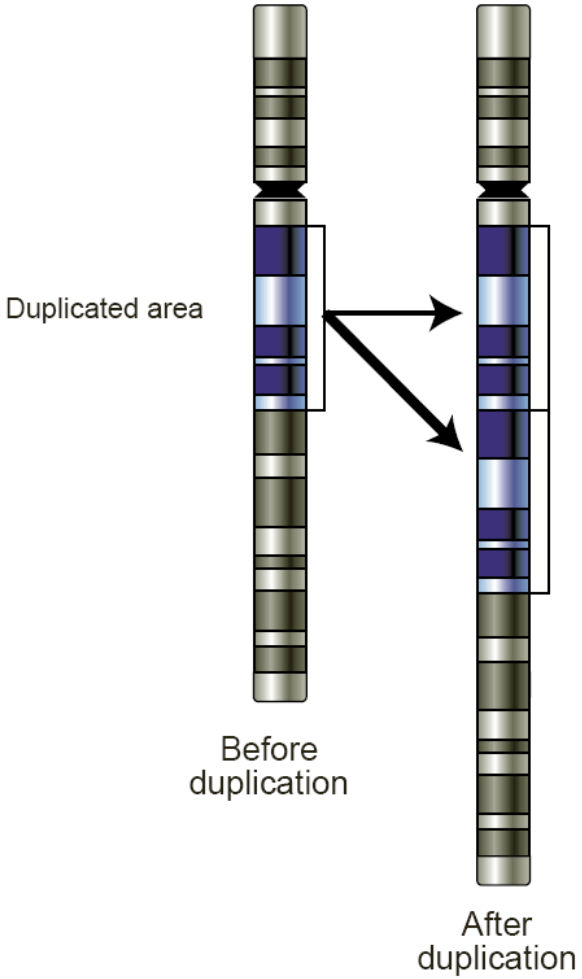
- TRANSPARENCY
- INNOVATION and SPEED
- CUSTOMIZATIONS

CNVs in Clinical Testing

- **Critical evidence needed for many genetic tests**
- **Common driver specific cancers, causal hereditary variation**
  - Chromosome 13 deletion common in melanoma
  - EGFR Exon 19 deletion common in lung cancer
  - PIK3CA Amplification in breast cancer
  - PTEN gross deletion/duplication ASD, *PTEN* hamartoma tumor syndrome (PHTS)
- **Today we will look at CNV calling on Tumor/Normal data**
  - We demonstrate CNVs calling on Melanoma tumor samples
  - Normal controls are used for normalization

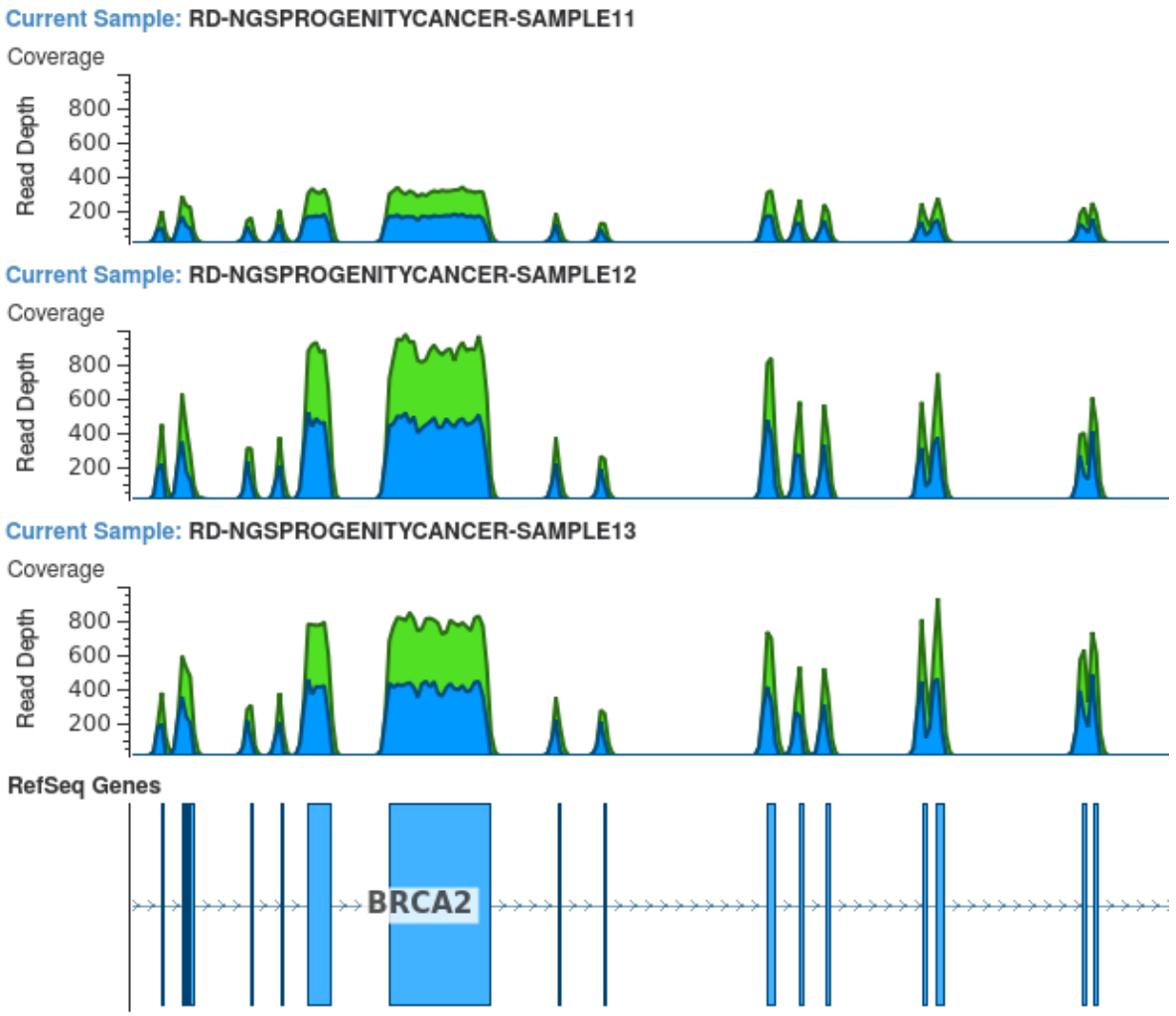
CNV Detection

- **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



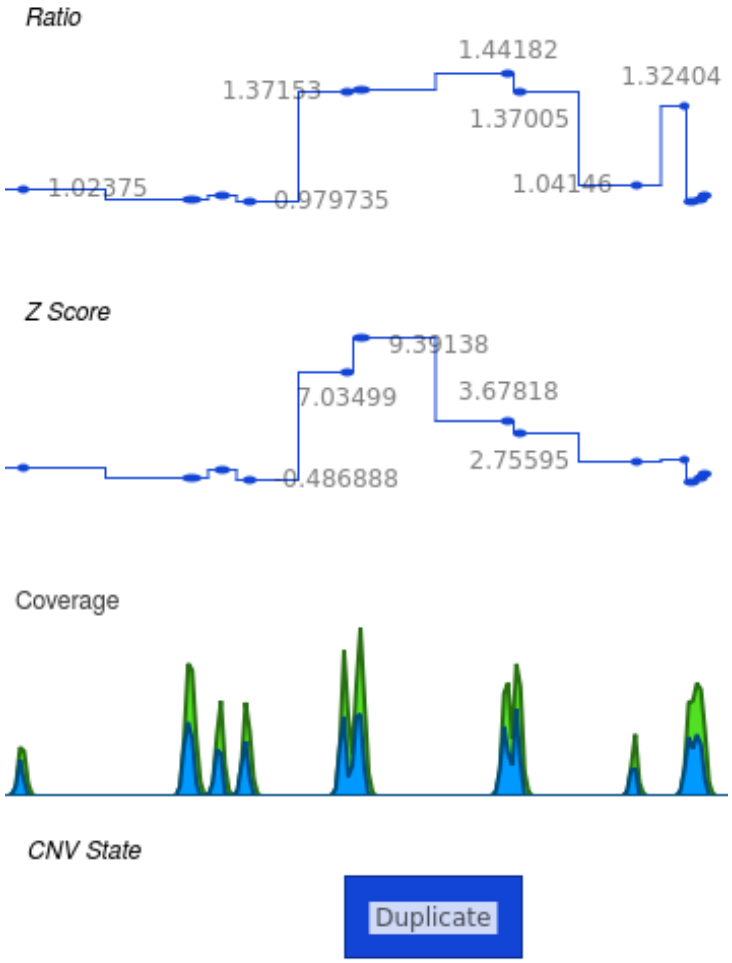
CNV Detection via NGS

- **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



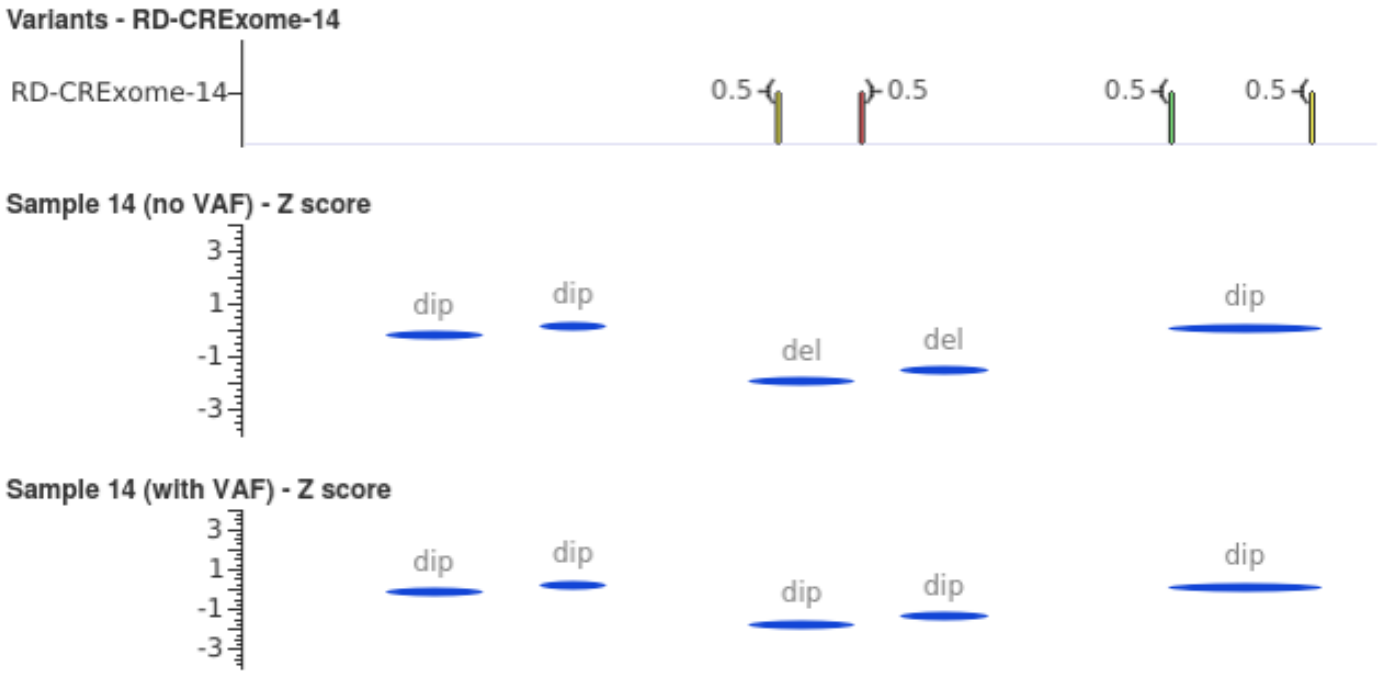
CNV calling in VarSeq

- **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
- **Probabilistic model used to call CNVs**



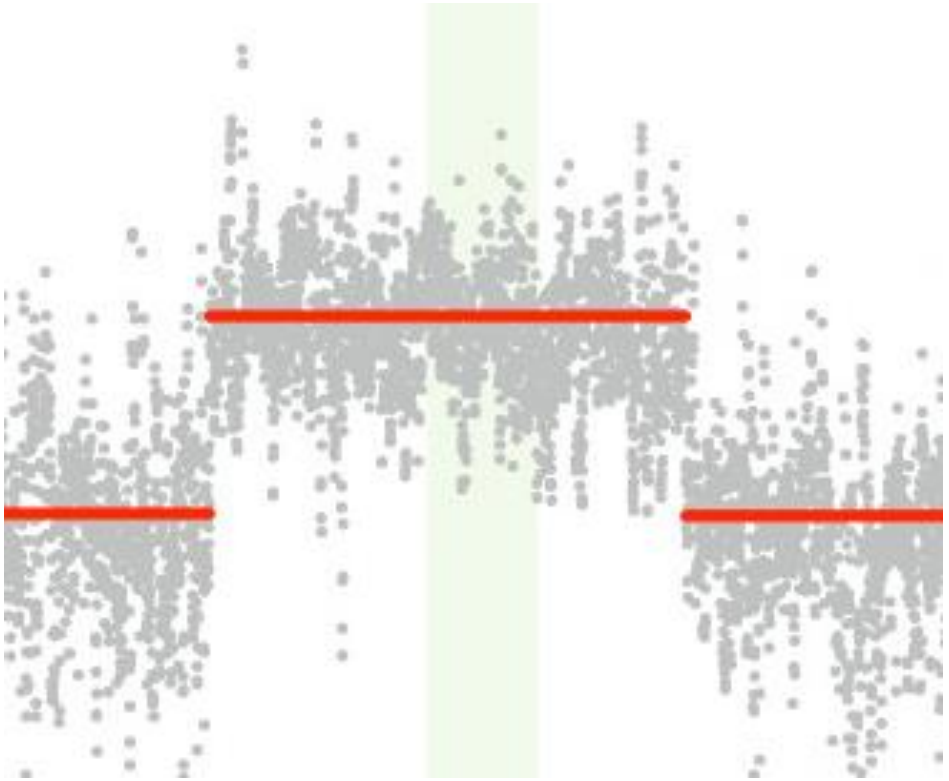
VAF

- **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



Segmentation

- Metrics are noisy over large regions
- Outliers cause large events to be called as many small events
- Addressed using CNAM optimal segmentation



Segmentation

- **First Pass**
  - Group targets into large megabase sized bins
  - Segment these bins to obtain large cytogenetic events
  - Perform fine tuning on the edges of the events
- **Second Pass**
  - Identify regions containing many small events
  - Perform segmentation these regions
  - Merge events that share a segmented region

QC Flags

- **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

Reference Samples

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

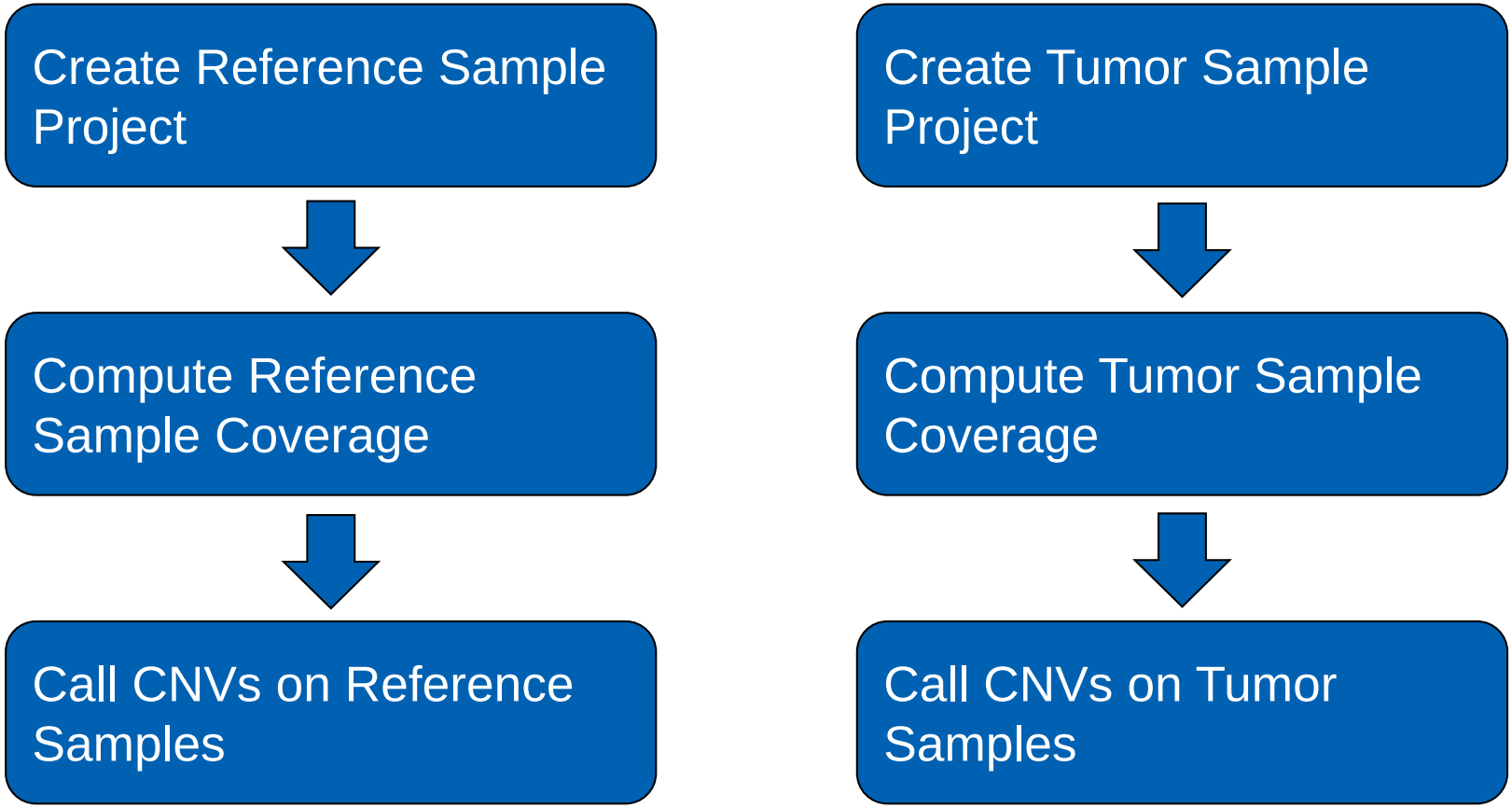
Requirements

- **100x Coverage**
- **Reference samples**
  - Recommend at least 30 references
  - Minimum of 10
  - From same platform and library preparation
  - Gender matched references required for Non-autosomal calls

Performance

- **Performance on Gene Panels:**
  - Sensitivity: 98.8 %
  - Specificity: 99.9 %
  - Precision: 99.6 %
- **We are currently evaluating performance on Tumor/Normal exomes**

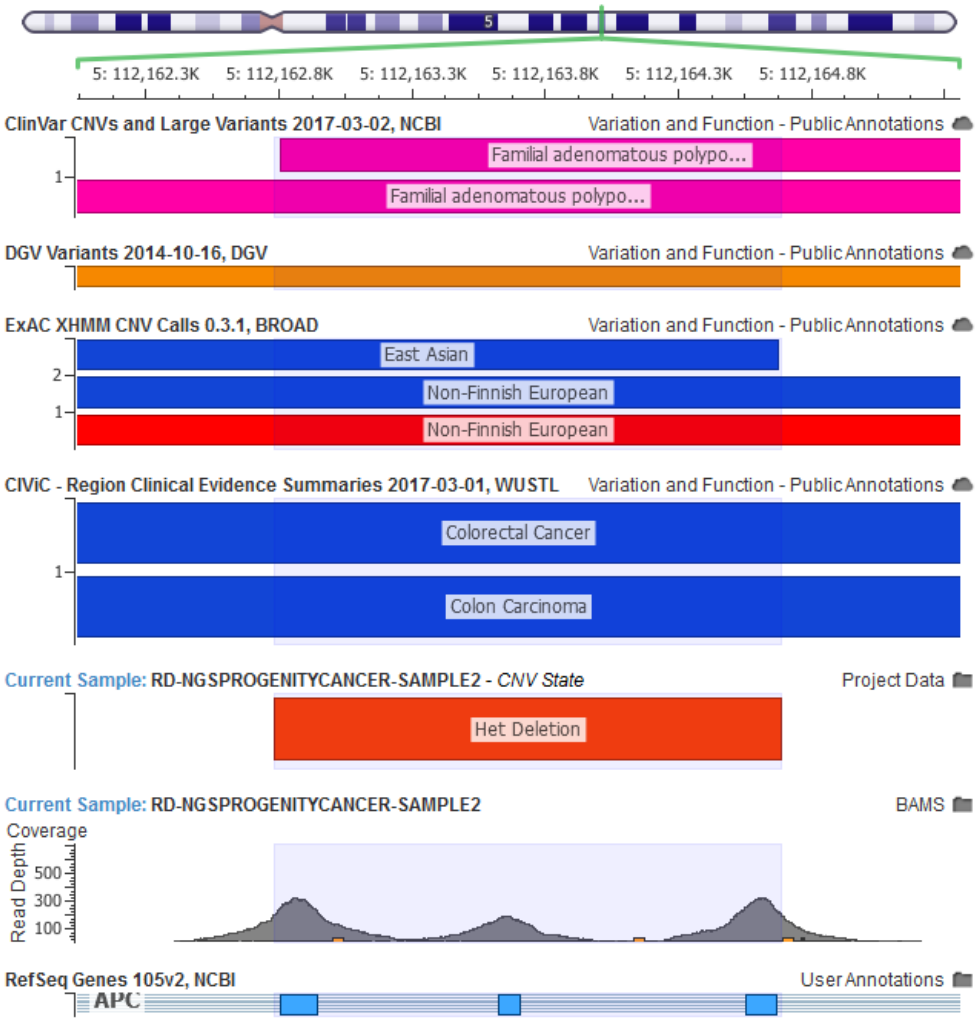
Tumor Normal Workflow



VarSeq Demonstration

Roadmap

- **Whole Genome Analysis**
- **CNV Reporting**
  - Able to add to CNVs to VSReport
- **CNV Annotations**
  - CNV annotations currently available
  - Next release integrate regional/overlap annotation of CNVs
  - Allow for more advanced filtering and interpretation workflows



Questions or more info:

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