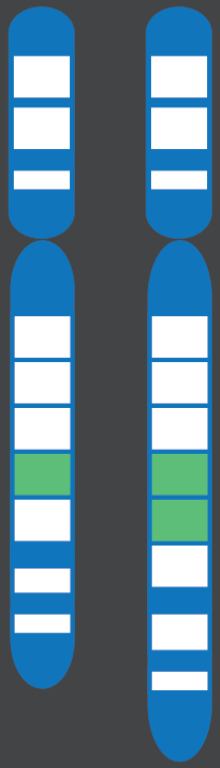
Getting More from your NGS Data: CNV Calling on Target Regions



October 12, 2016

Gabe Rudy
VP Product & Engineering

Dr. Nathan Fortier
Senior Software Engineer
& Field Application Scientist

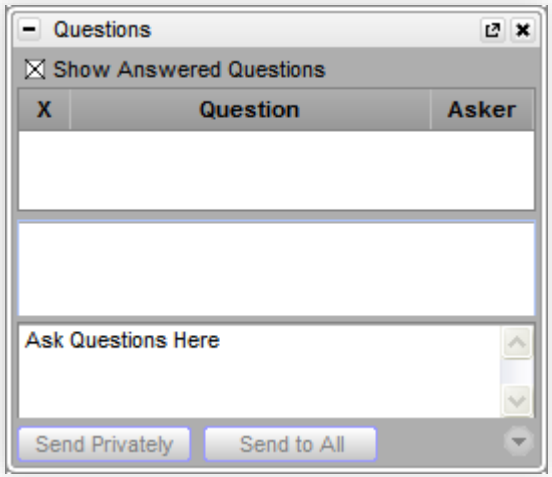
GOLDEN HELIX
Accelerating the Quest for Significance™

Agenda

1. Overview Golden Helix
2. Why Call CNVs in NGS
3. Method and Demo
4. Availability and 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

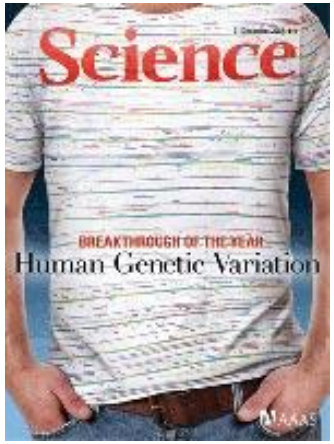
NATIONAL CANCER INSTITUTE

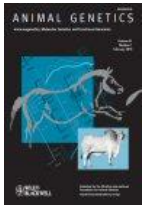
USDA
USDA Agricultural Research Service

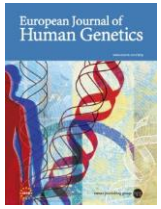
FDA

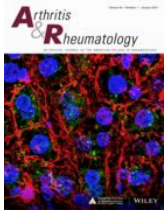
GOLDEN HELIX
Accelerating the Quest for Significance™

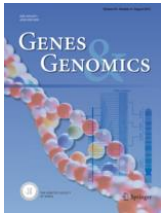
Cited in over 1000 peer-reviewed publications







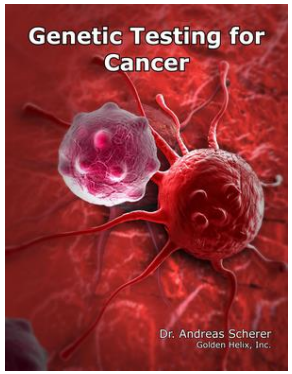




Golden Helix – Who We Are

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

- REPUTATION
- TRUST
- EXPERIENCE



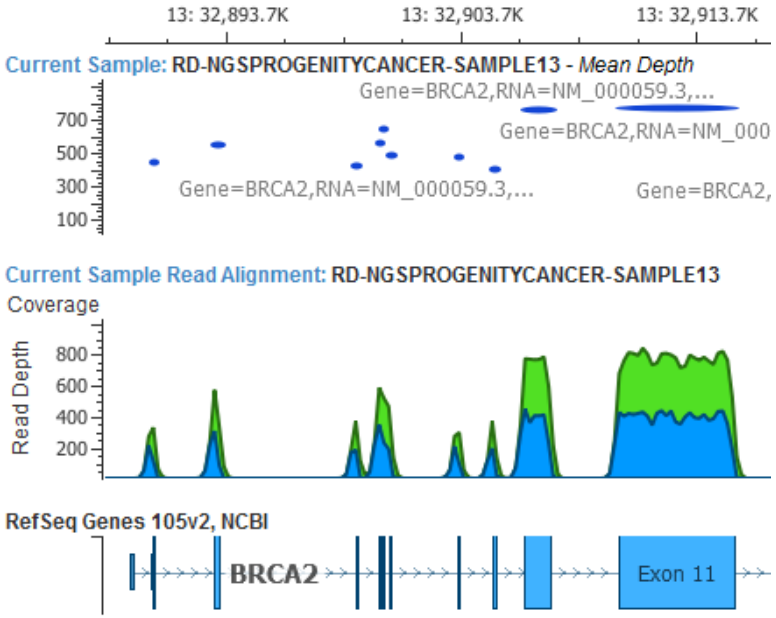
- INDUSTRY FOCUS
- THOUGHT LEADERSHIP
- COMMUNITY

- TRAINING
- SUPPORT
- RESPONSIVENESS

- TRANSPARENCY
- INNOVATION and SPEED
- CUSTOMIZATIONS

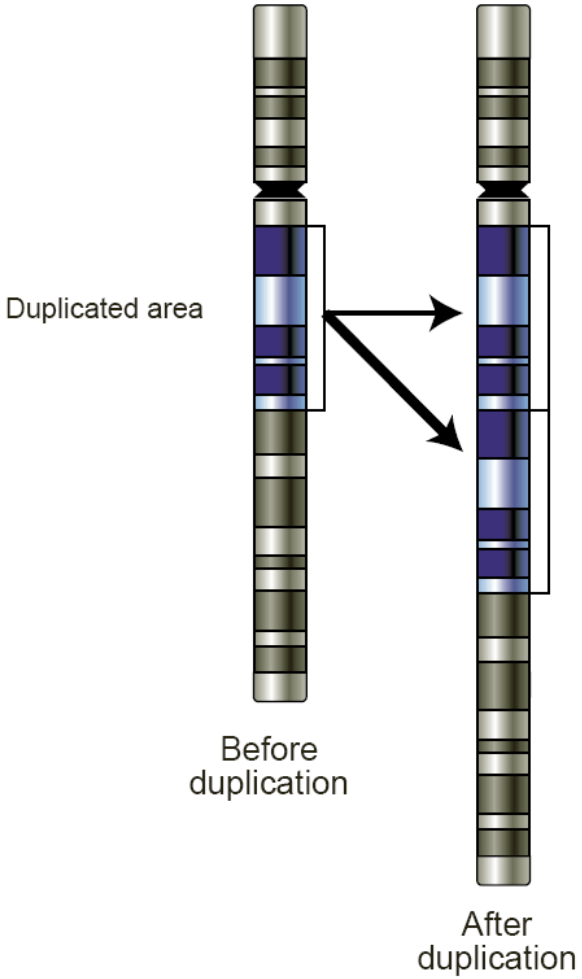
Why Call CNVs on your Gene Panels?

- **Gene Panels in Clinical Use**
  - Add value to existing workflow
- **Precision**
  - NGS provides the precision to call single-exon events that can be missed by whole-genome micro-arrays (CMAs)
- **But Challenging (Coverage Variable)**
  - Variation is systematic to sequencing kit
  - Can be overcome with repo of reference samples
  - Even with false-positives, very few to review



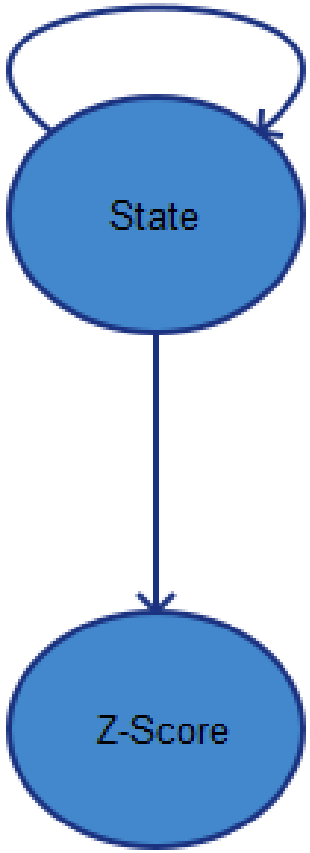
CNV Detection

- **CNV detection involves:**
  - Normalization
  - Data Correction
  - Classification
- **Several approaches exist for each step**
- **All approaches rely on coverage information**



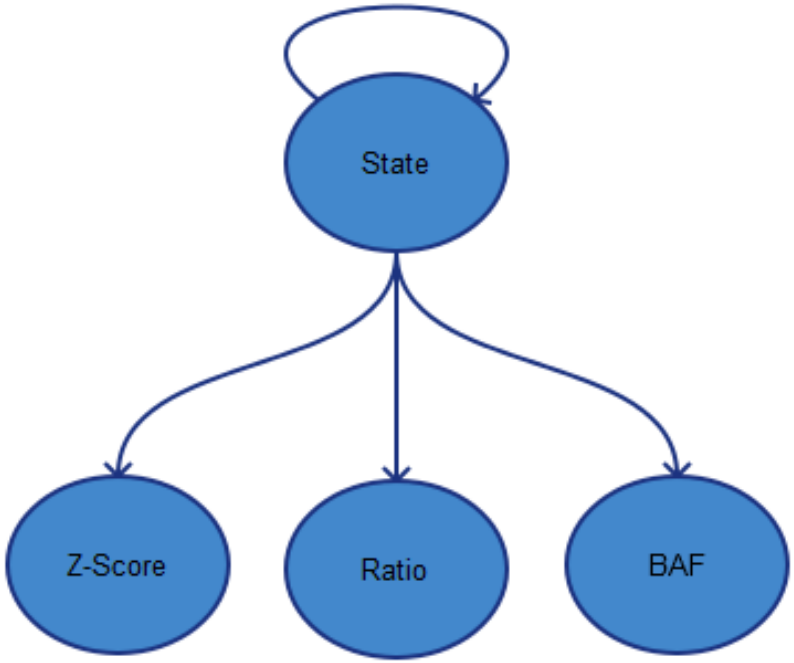
Classification

- **Thresholding**
  - Combine *multiple measures* using logical operators to determine state
- **Hidden Markov Models**
  - Markov process with hidden state
  - Use a *single metric* (such as coverage “Z-Score”) as evidence for hidden CNV state
  - Call CNVs by finding most probable assignment to hidden state



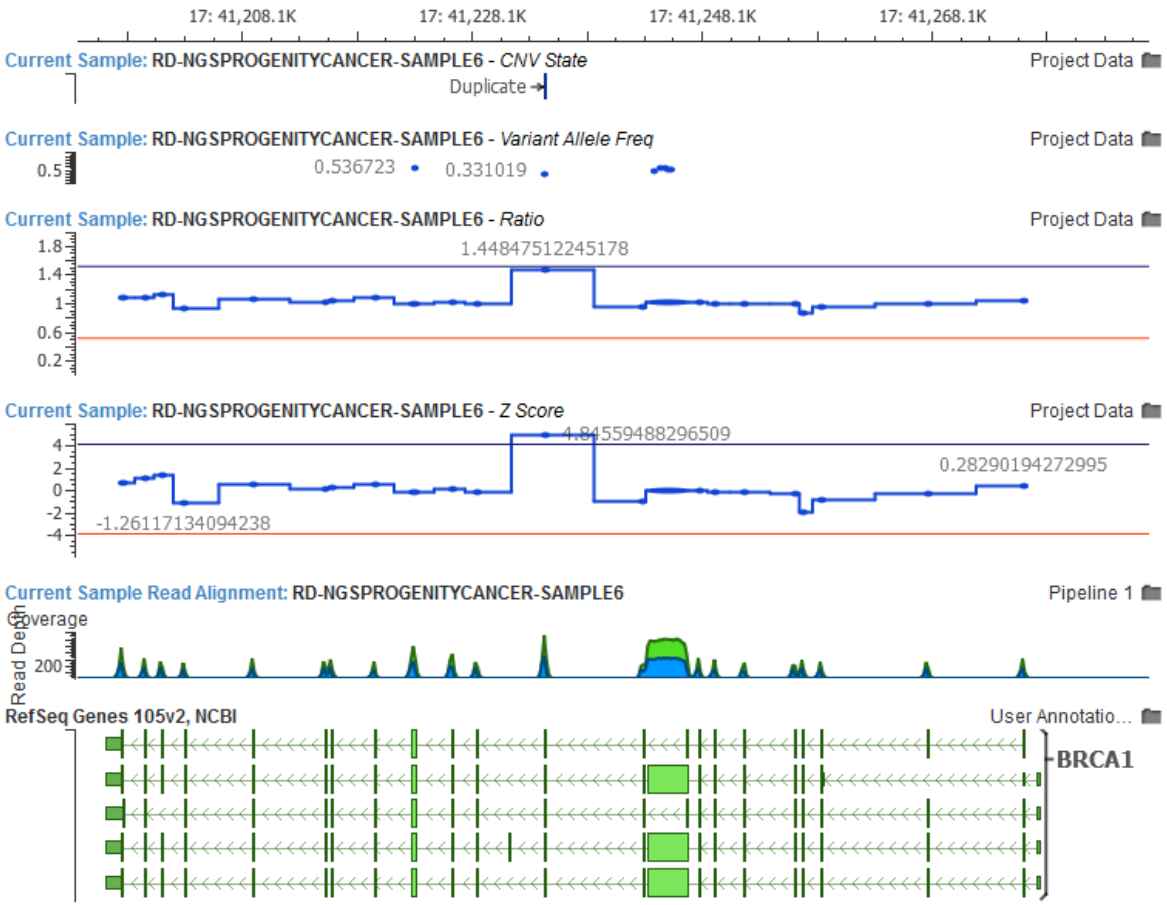
Dynamic Bayesian Networks

- **Combine the best of both worlds**
  - HMMs cannot easily incorporate multiple evidence metrics
  - Dynamic Bayesian Networks overcome this limitation
- **Dynamic Bayesian Networks**
  - Directed graph in which:
    - Nodes represent random variable
    - Edges represent conditional dependencies
  - Each Node encodes a probability distribution conditioned on its parents



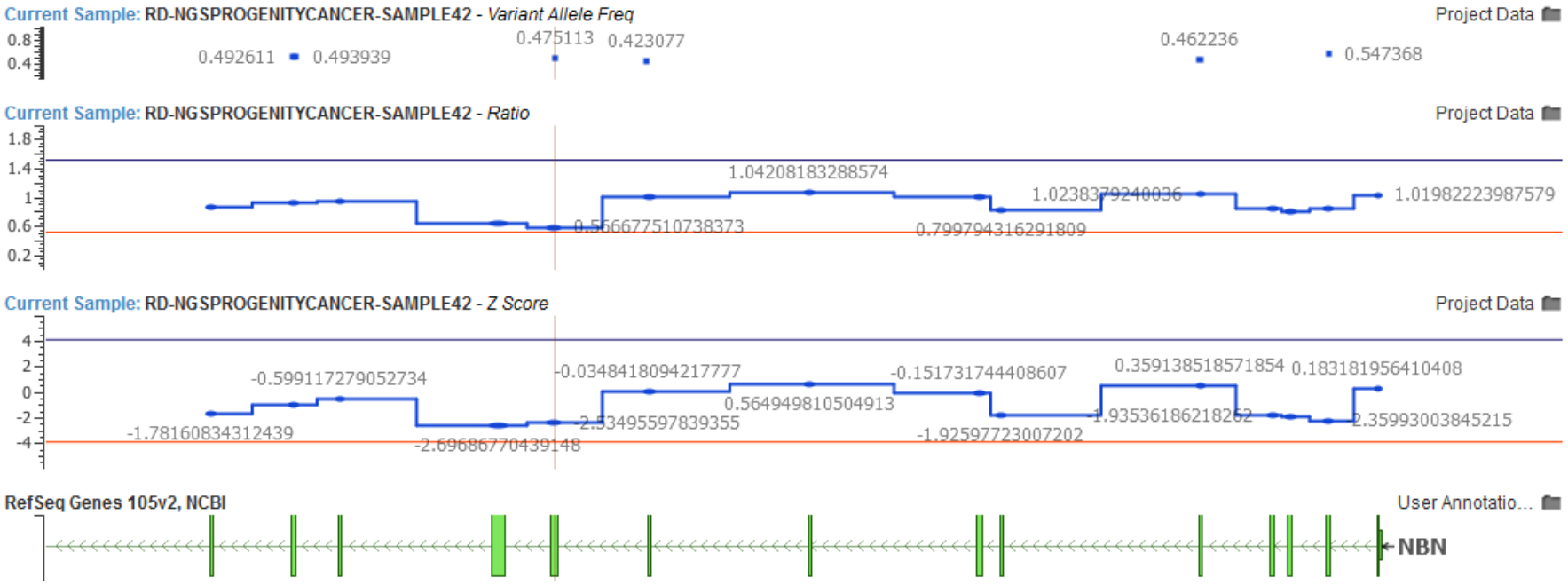
B Allele Frequency

- **Already in VarSeq:**
  - VarSeq maps target regions to containing variants
  - The Variant Allele Frequency field is automatically computed on import
- **Provides supporting evidence:**
  - For duplication
    - ratios like 1/3, 2/3 etc
  - Against deletions
    - Any non 0 or 1 ratio
    - Reduce FP deletion calls



This single-target duplication is called with higher probabilities due to the 0.33 BAF of containing variant

BAF Help Reduce False-Positives



Potential deletion event not called because variant in target has ~0.5 Variant Allele Frequency

CNV Calling in VarSeq

Coverage Computed on Targets in BED File

Add Samples to Reference Set

Run CNV for each Sample

- Pick 30 Matched Controls
- Compute Ratio, Z-Score, BAF
- Classify Targets with DBN
- Event Segmentation with DBN

Outputs

- Target Region Table
- CNV Event Table
- Sample Summary Table

Reference Samples

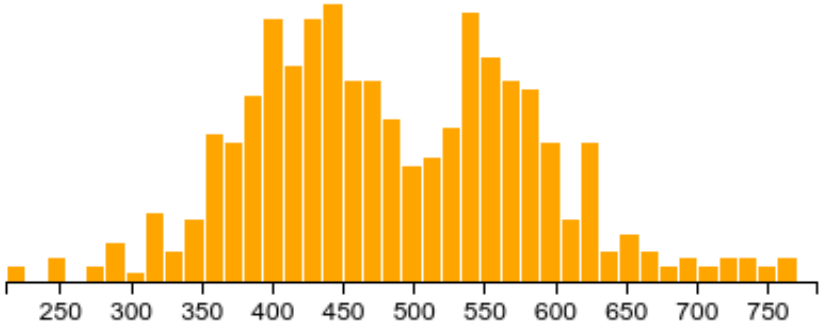
- Global
- Matched to Panel
- Sample Aware

Benchmarking Data

- **144 cancer panel samples provided by PreventionGenetics**
  - Samples cover 549 target regions in 31 cancer genes
  - 91 of these regions fall within CNV events, including heterozygous deletions and duplications
  - Algorithm was run over 48 of these samples
  - All 144 samples were used as controls
- **Experiments illustrate effect of the CNV probability on**
  - Sensitivity (true positive rate)
  - Precision (positive predictive value)

PREVENTION GENETICS

DISEASE PREVENTION THROUGH GENETIC TESTING



Average Read Depth of Samples

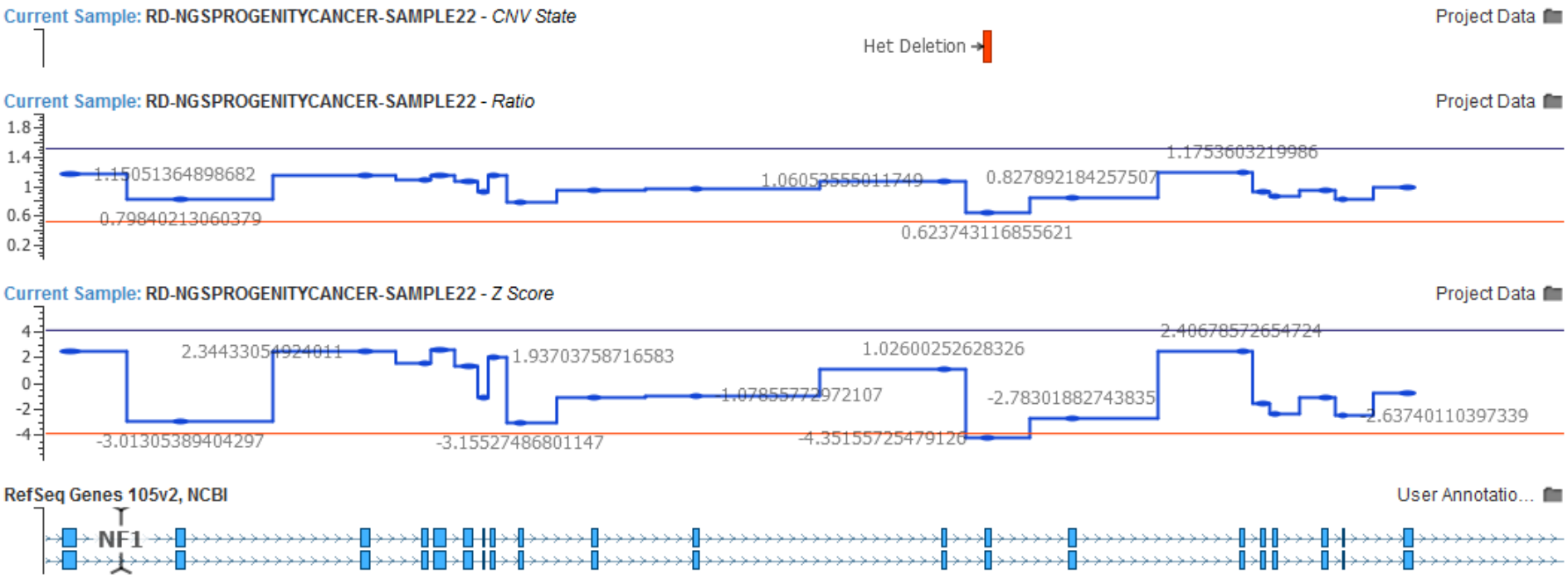
Results

- **Performance as CNV probability is adjusted**

P(CNV)	TP	FP	TN	FN	Sens.	Prec.
1E-12	84	31	26345	7	92.3%	73.0%
1E-10	87	43	26348	4	95.6%	66.9%
1E-8	89	51	26350	2	97.8%	63.6%
1E-6	91	63	26352	0	100%	59.1%

Flagging Events

- **Low quality events can be flagged if**
  - *Low Controls Depth*: The mean of the matched controls read depth is exceptionally low
  - *High Controls Variation*: The variation of the matched controls read depth was high
  - *Within Regional IQR*: The cannot be differentiated from the noise of local region
- **Filtering flagged events improves precision**



Results

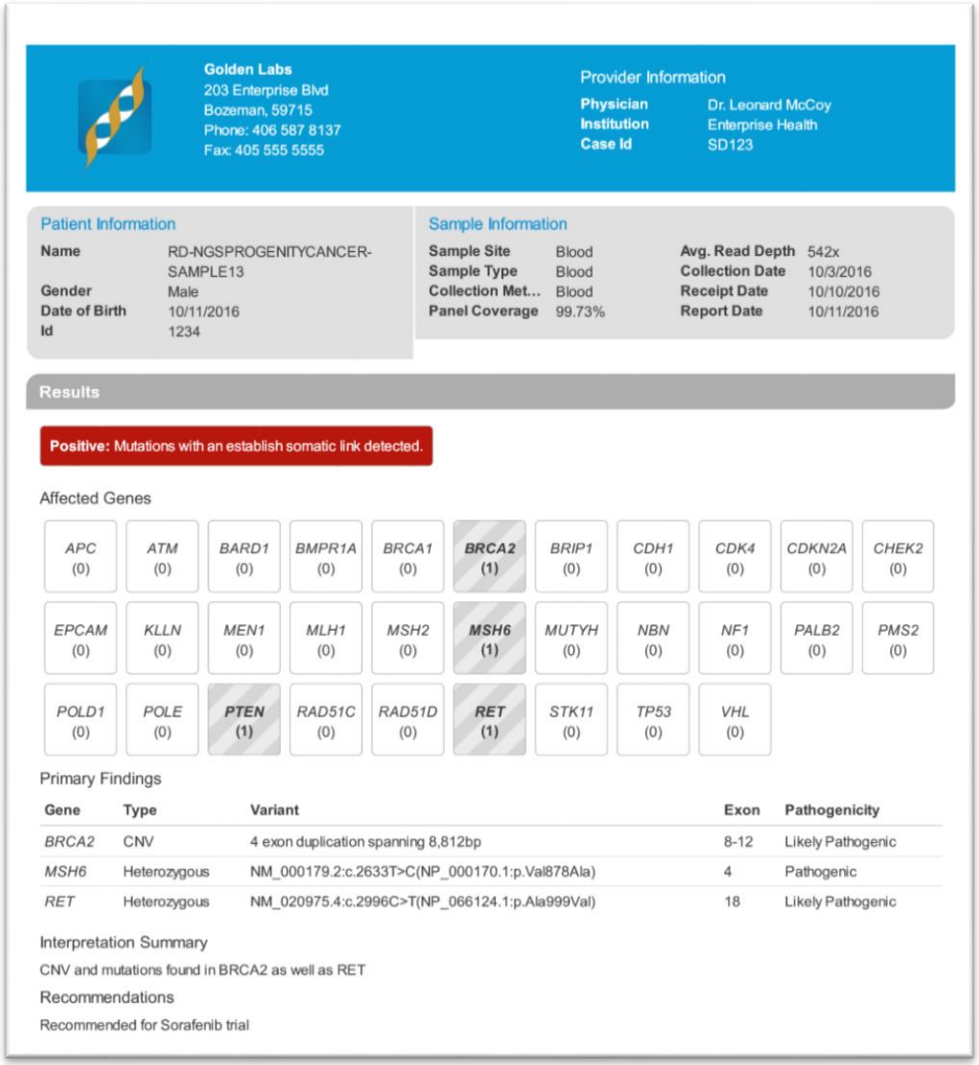
- **Performance excluding flagged events**

P(CNV)	TP	FP	TN	FN	Sens.	Prec.
1E-12	82 (2)	1 (30)	26343	9	90.1%	98.8%
1E-10	85 (2)	3 (40)	26346	6	93.4%	96.6%
1E-8	87 (2)	5 (46)	26348	4	95.6%	94.6%
1E-6	88 (3)	9 (54)	26539	3	96.7%	90.7%

VarSeq Demonstration

Availability and Roadmap

- **Early Access Now**
  - Part of VarSeq Clinical w/ VSReports
  - Being validated in clinical workflows
  - Will be in upcoming VarSeq 1.4.2
- **Validate Your Workflows**
  - Need reference samples
  - Validate against known CNVs or orthogonal assay
- **Exomes Upcoming**
  - Say strategy will scale to exomes
  - Exomes have extreme regional effects
  - May require additional normalization, regional knowledge and QC flags



Golden Labs
203 Enterprise Blvd
Bozeman, 59715
Phone: 406 587 8137
Fax: 405 555 5555

Provider Information

Physician	Dr. Leonard McCoy
Institution	Enterprise Health
Case Id	SD123

Patient Information

Name	RD-NGSPROGENITYCANCER-SAMPLE13
Gender	Male
Date of Birth	10/11/2016
Id	1234

Sample Information

Sample Site	Blood	Avg. Read Depth	542x
Sample Type	Blood	Collection Date	10/3/2016
Collection Met...	Blood	Receipt Date	10/10/2016
Panel Coverage	99.73%	Report Date	10/11/2016

Results

Positive: Mutations with an establish somatic link detected.

Affected Genes

APC (0)	ATM (0)	BARD1 (0)	BMPR1A (0)	BRCA1 (0)	**BRCA2 (1)**	BRIP1 (0)	CDH1 (0)	CDK4 (0)	CDKN2A (0)	CHEK2 (0)
EPCAM (0)	KLLN (0)	MEN1 (0)	MLH1 (0)	MSH2 (0)	**MSH6 (1)**	MUTYH (0)	NBN (0)	NF1 (0)	PALB2 (0)	PMS2 (0)
POLD1 (0)	POLE (0)	**PTEN (1)**	RAD51C (0)	RAD51D (0)	**RET (1)**	STK11 (0)	TP53 (0)	VHL (0)		

Primary Findings

Gene	Type	Variant	Exon	Pathogenicity
BRCA2	CNV	4 exon duplication spanning 8,812bp	8-12	Likely Pathogenic
MSH6	Heterozygous	NM_000179.2:c.2633T>C(NP_000170.1:p.Val878Ala)	4	Pathogenic
RET	Heterozygous	NM_020975.4:c.2996C>T(NP_066124.1:p.Ala999Val)	18	Likely Pathogenic

Interpretation Summary

CNV and mutations found in BRCA2 as well as RET

Recommendations

Recommended for Sorafenib trial

Questions or more info:

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