GOLDEN HELIX

Enabling Precision Medicine

Fine-tuning CNV Analysis for the Clinical Analysis of NGS Samples

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

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

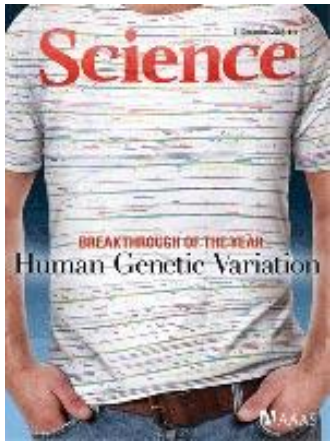
VARIANT WAREHOUSE

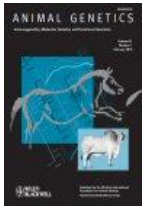
Variant Warehouse
Centralized Annotations
Hosted Reports
Sharing and Integration

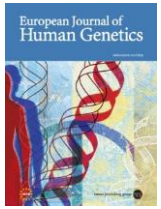
SNP & VARIATION SUITE

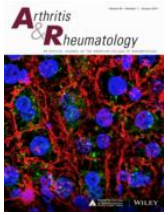
GWAS
Genomic Prediction
Large-N-Population Studies
RNA-Seq
Large-N CNV-Analysis

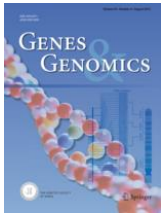
Cited in over 1100 peer-reviewed publications











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

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

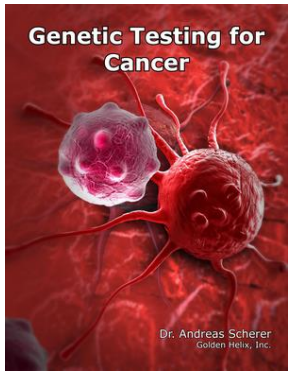
FDA

GOLDEN HELIX
Enabling Precision Medicine

Golden Helix – Who We Are

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

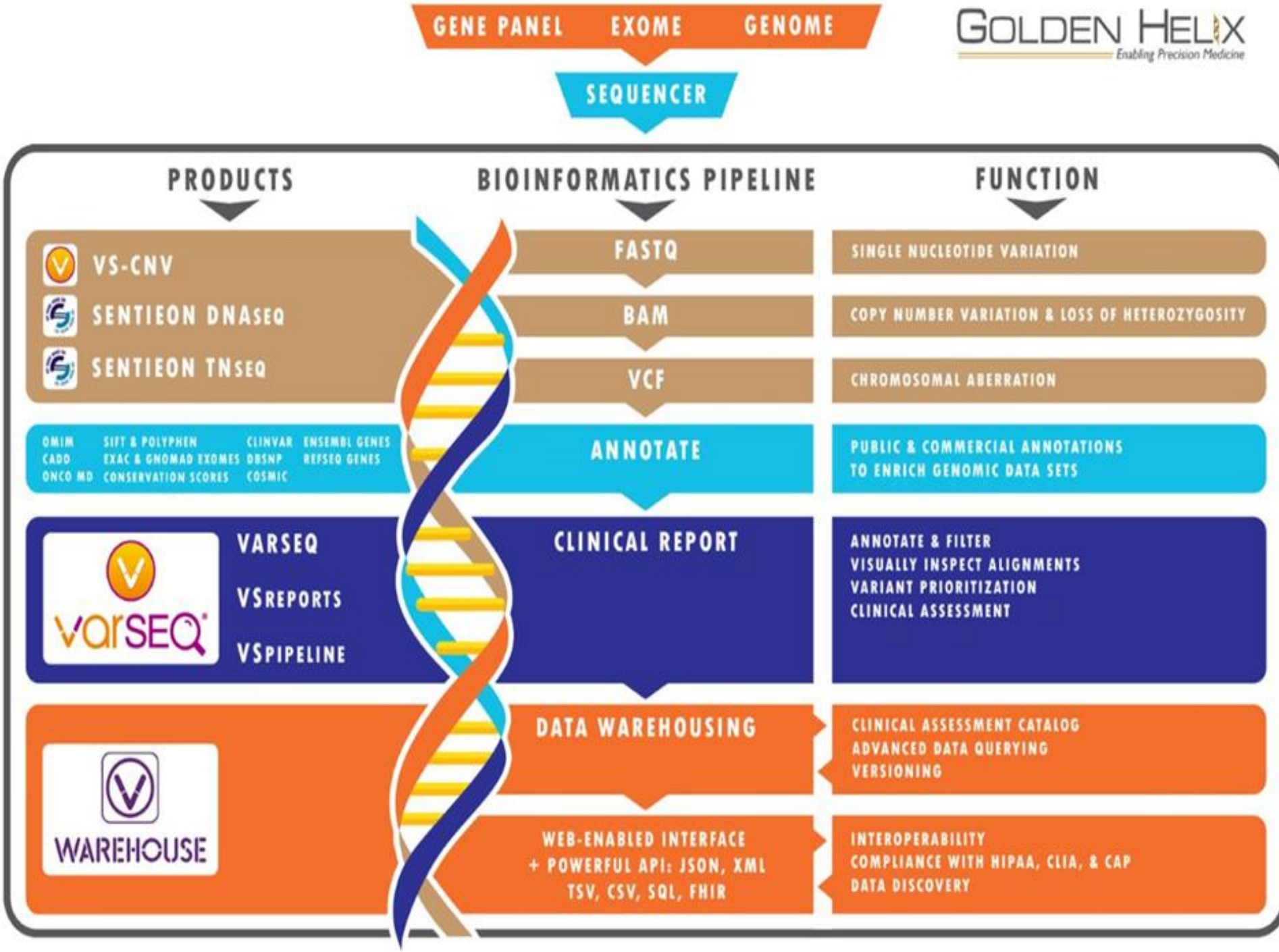
- REPUTATION
- TRUST
- EXPERIENCE



- INDUSTRY FOCUS
- THOUGHT LEADERSHIP
- COMMUNITY

- TRAINING
- SUPPORT
- RESPONSIVENESS

- INNOVATION and SPEED
- CUSTOMIZATIONS

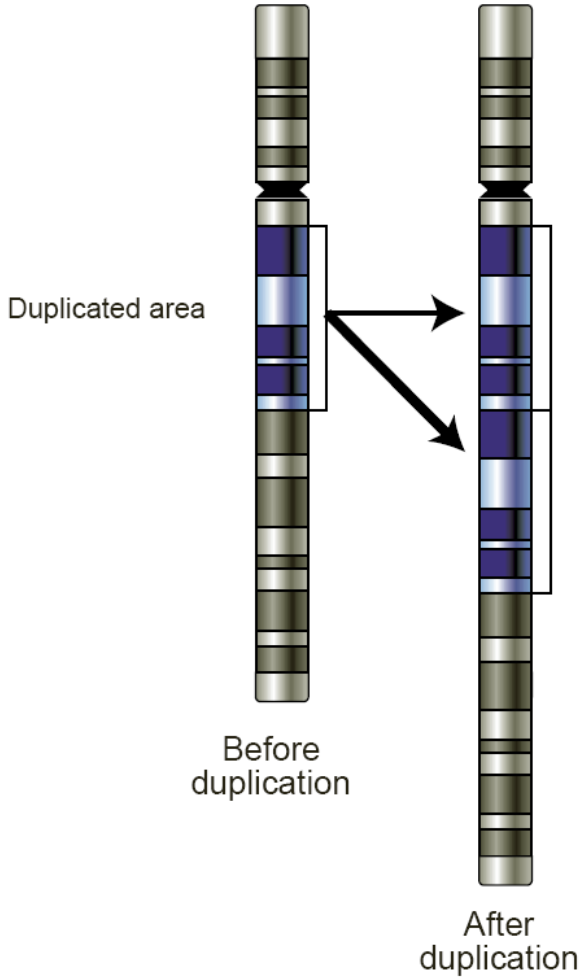

GENE PANEL
EXOME
GENOME
SEQUENCER
GOLDEN HELIX
Enabling Precision Medicine
PRODUCTS
BIOINFORMATICS PIPELINE
FUNCTION
VS-CNV
SENTIEON DNASEQ
SENTIEON TNSEQ
FASTQ
BAM
VCF
SINGLE NUCLEOTIDE VARIATION
COPY NUMBER VARIATION & LOSS OF HETEROZYGOSITY
CHROMOSOMAL ABERRATION
OMIM
CADD
ONCO MD
SIFT & POLYPHEN
EXAC & GNOMAD EXOMES
CONSERVATION SCORES
CLINVAR
DBSNP
COSMIC
ENSEMBL GENES
REFSEQ GENES
ANNOTATE
PUBLIC & COMMERCIAL ANNOTATIONS
TO ENRICH GENOMIC DATA SETS
varSEQ
VARSEQ
VSREPORTS
VSPIPELINE
CLINICAL REPORT
ANNOTATE & FILTER
VISUALLY INSPECT ALIGNMENTS
VARIANT PRIORITIZATION
CLINICAL ASSESSMENT
WAREHOUSE
DATA WAREHOUSING
CLINICAL ASSESSMENT CATALOG
ADVANCED DATA QUERYING
VERSIONING
WEB-ENABLED INTERFACE
+ POWERFUL API: JSON, XML
TSV, CSV, SQL, FHIR
INTEROPERABILITY
COMPLIANCE WITH HIPAA, CLIA, & CAP
DATA DISCOVERY

CNVs in Clinical Testing

- **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)

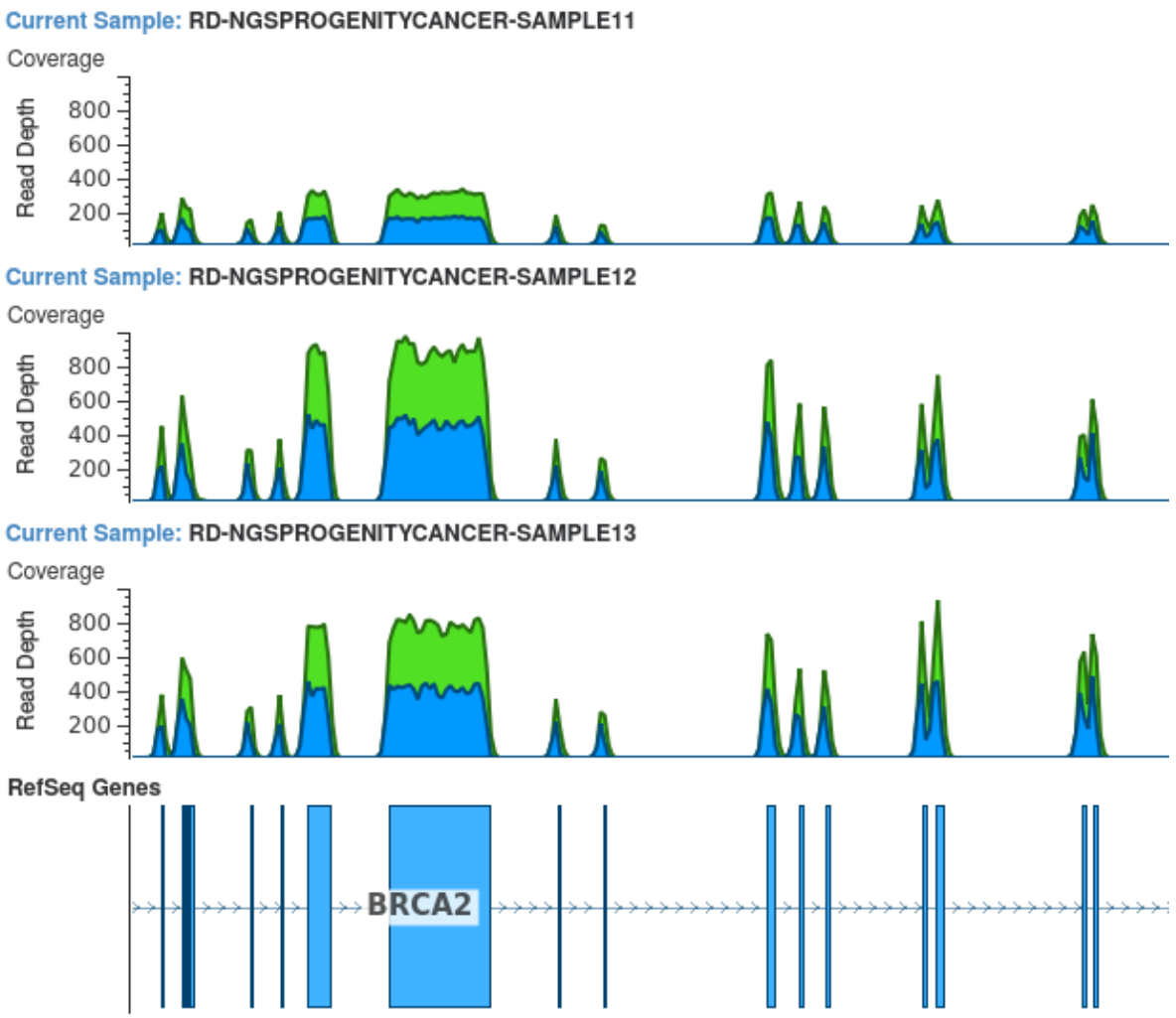
CNV Detection

- **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



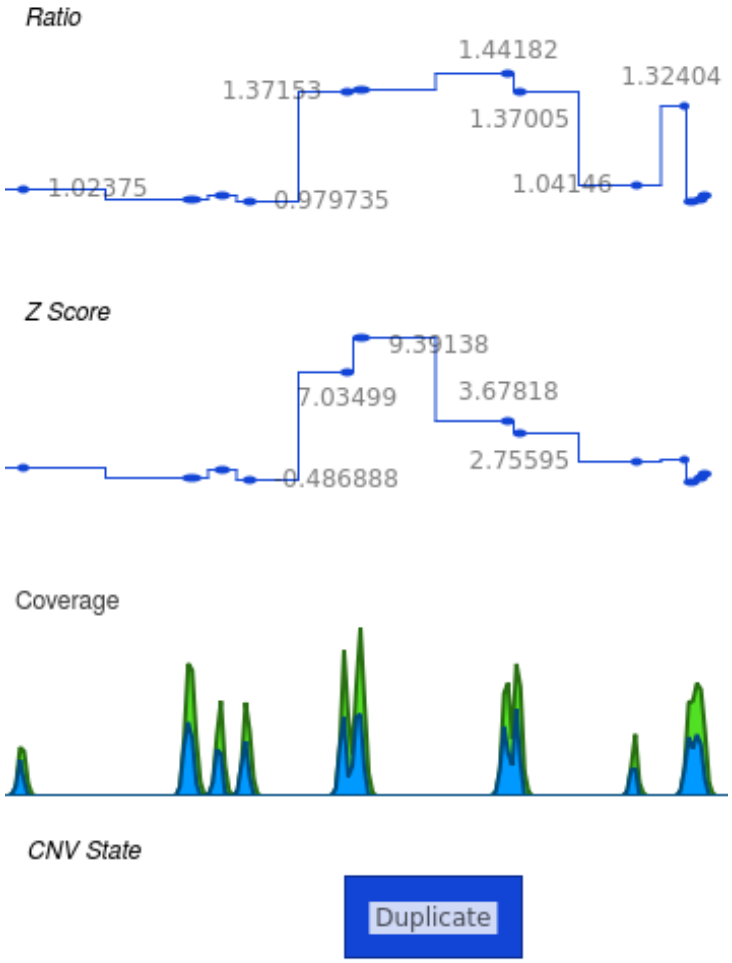
CNV Detection via NGS

- **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



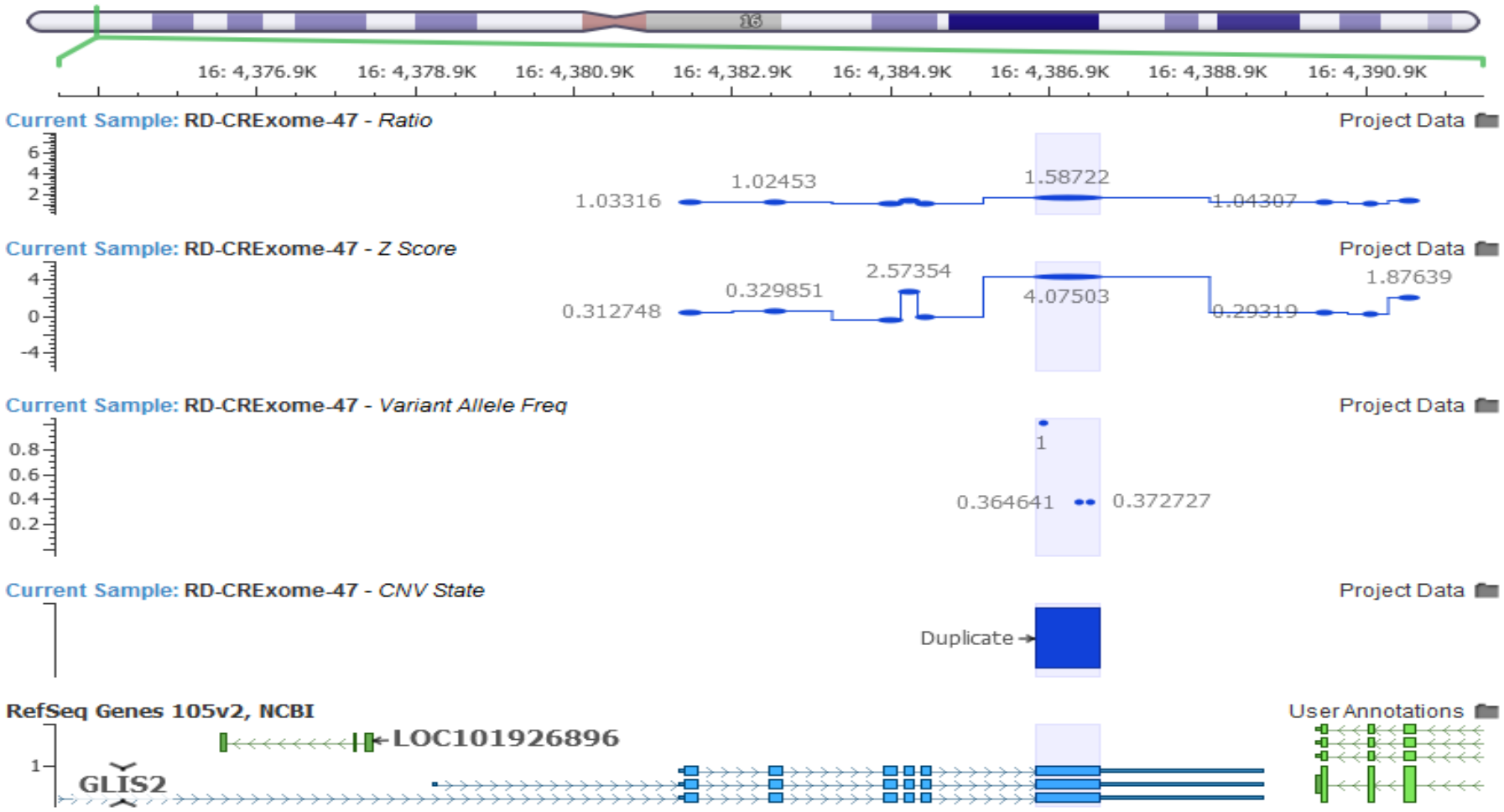
CNV calling in VarSeq

- **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



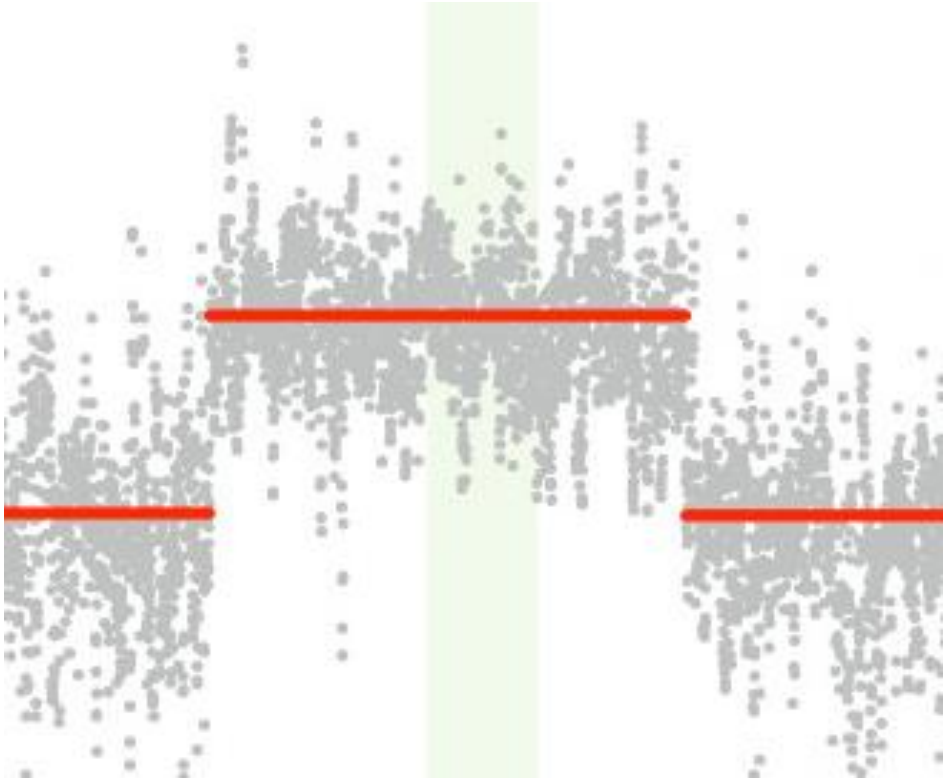
VAF provides supporting evidence

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



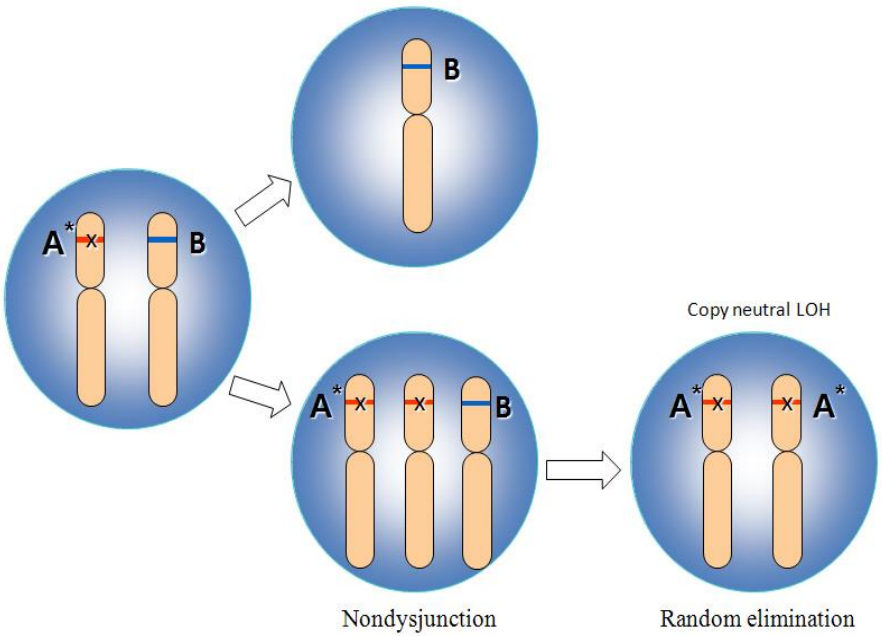
Segmentation

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



LoH Calling

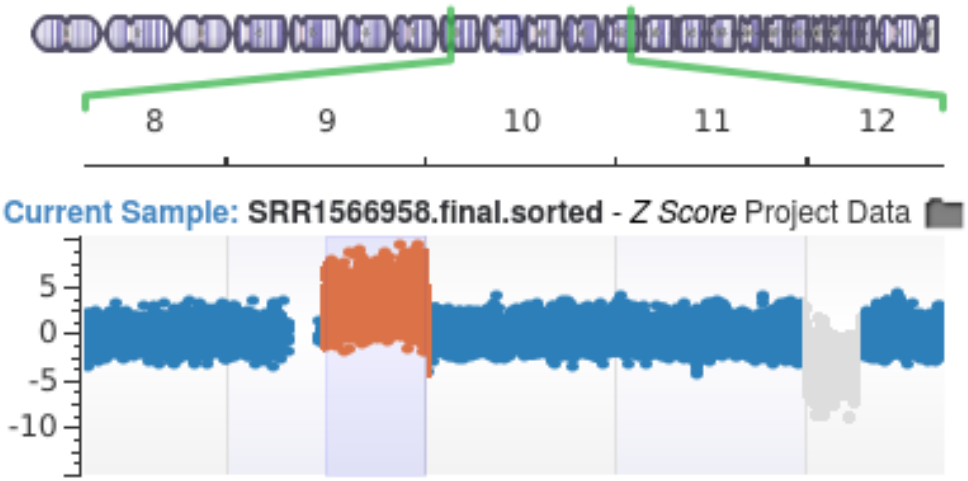
- **Large LoH events need to be interpreted in any gene test that covers large CNVs**
- **New Loss of Heterozygosity(LOH) detection based on H3M2 (Magi *et al.*)**
- **Calls LoH events using Hidden Markov Model (HMM)**
  - Observations are variant allele frequencies
  - States are either Homozygous or Non-Homozygous



LoH Calling

P-Values

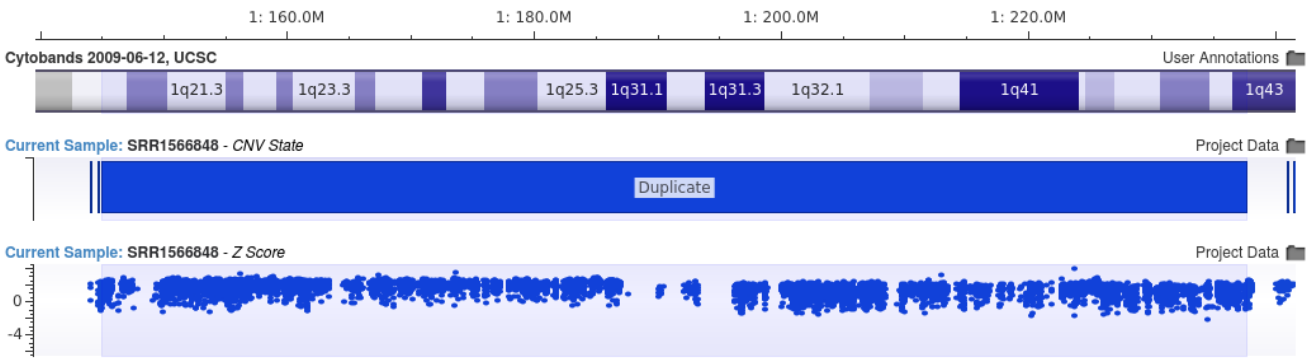
- **P-Values**
  - Probability of z-scores at least as extreme assuming the event targets are diploid
  - Computed using Student's t-test
  - Distribution of event z-scores compared to distribution of diploid targets
- **Quantifies CNV Call Confidence**
  - Values below 0.01 indicate high confidence calls
  - Values above 0.01 indicate lower confidence calls



$$p = 1.4 \cdot 10^{-32}$$

Karyotype Notation

- Karyotype notation provided for large cytogenetic events
- Karyotypes provided at both event and sample level
- Uses common notation
- Specifies chromosome, arm, and band for each mutation



46,XY,dup(1)(q21.1q43)

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

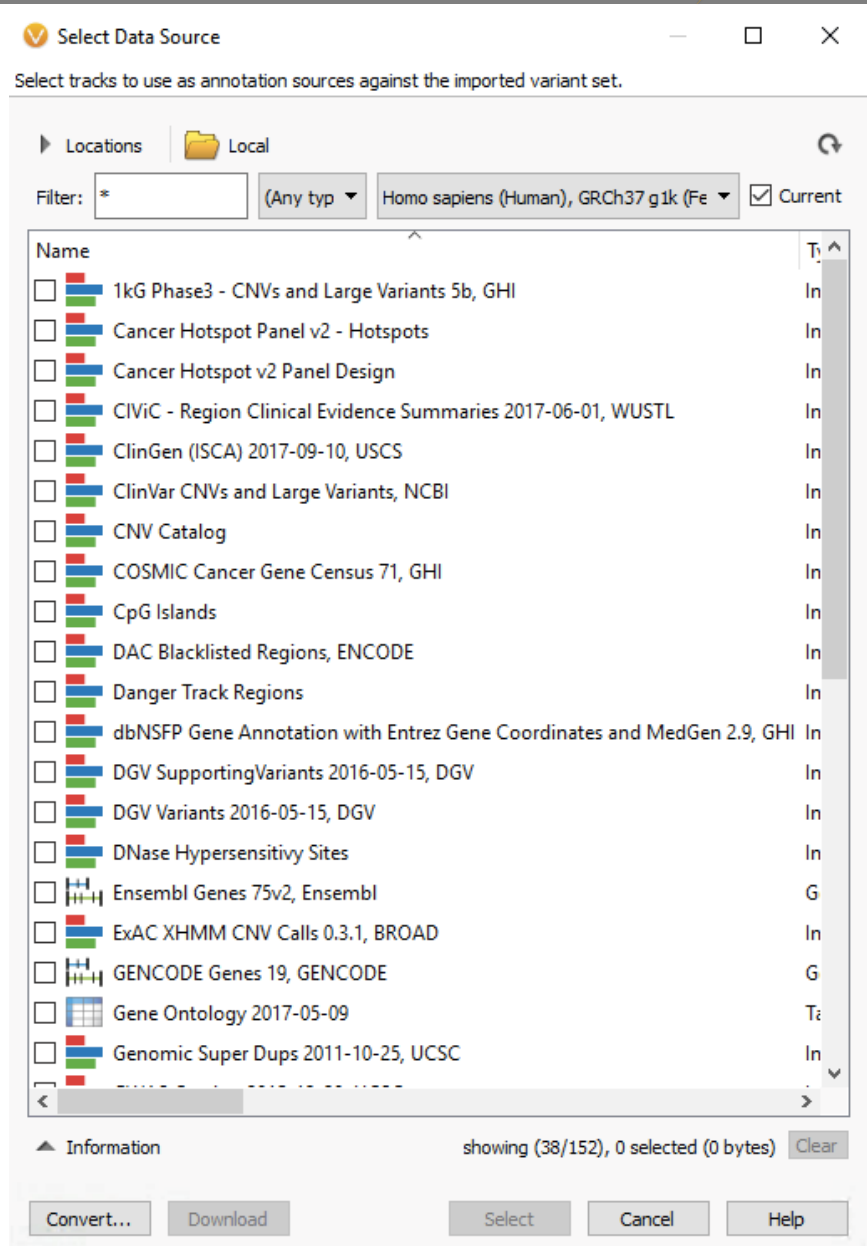
- **100x Coverage**
- **Reference samples**
  - Recommend at least 30 references
  - Minimum of 10
  - From same platform and library preparation
  - Gender matched references required for non-autosomal calls

Non-Autosomal Normalization

- **Sex is inferred from coverage data**
  - Sample is inferred female if
    - Y chromosome coverage is low
    - X chromosome coverage matches the autosome
  - Otherwise the sample is inferred to be male
- **Samples are matched on inferred sex**
- **Same-sex samples are used for normalization of non-autosomal chromosomes**

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)

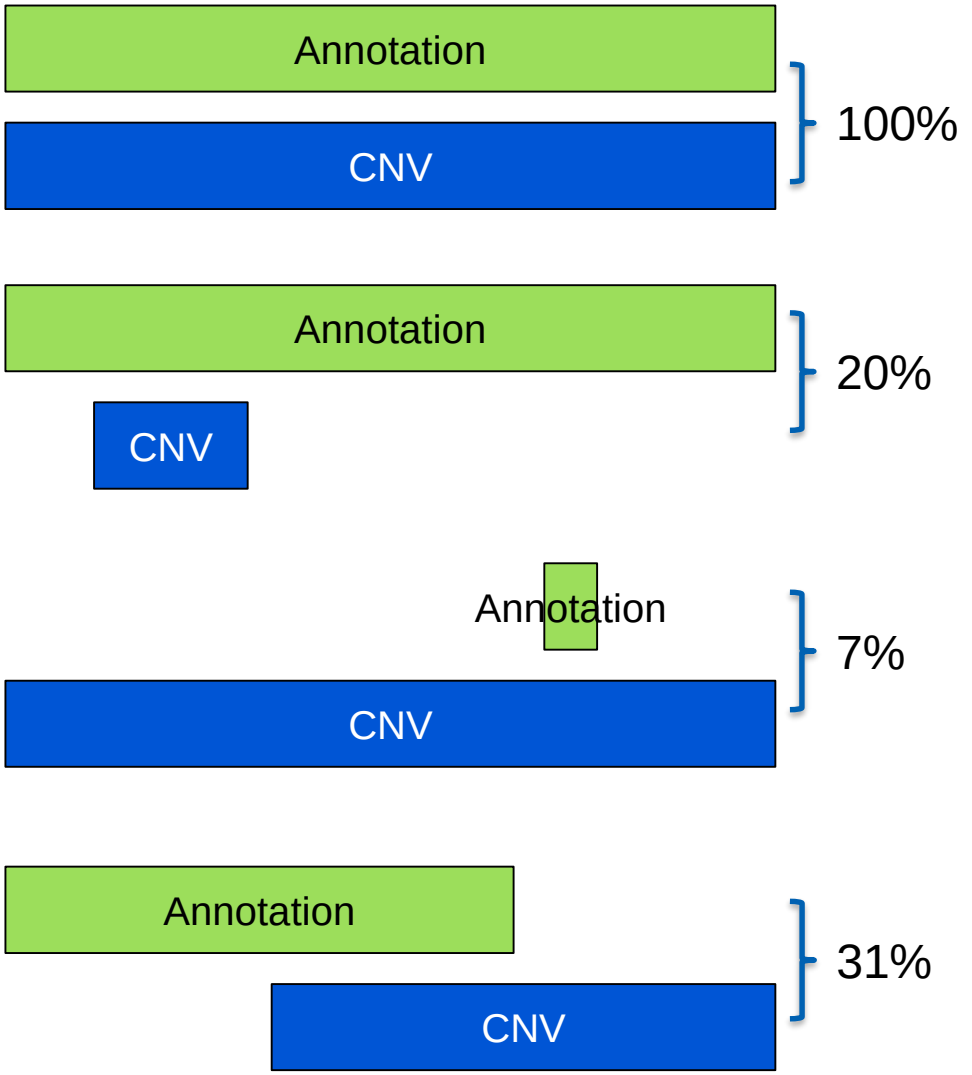


Annotation Algorithms: Overlapping Regions

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

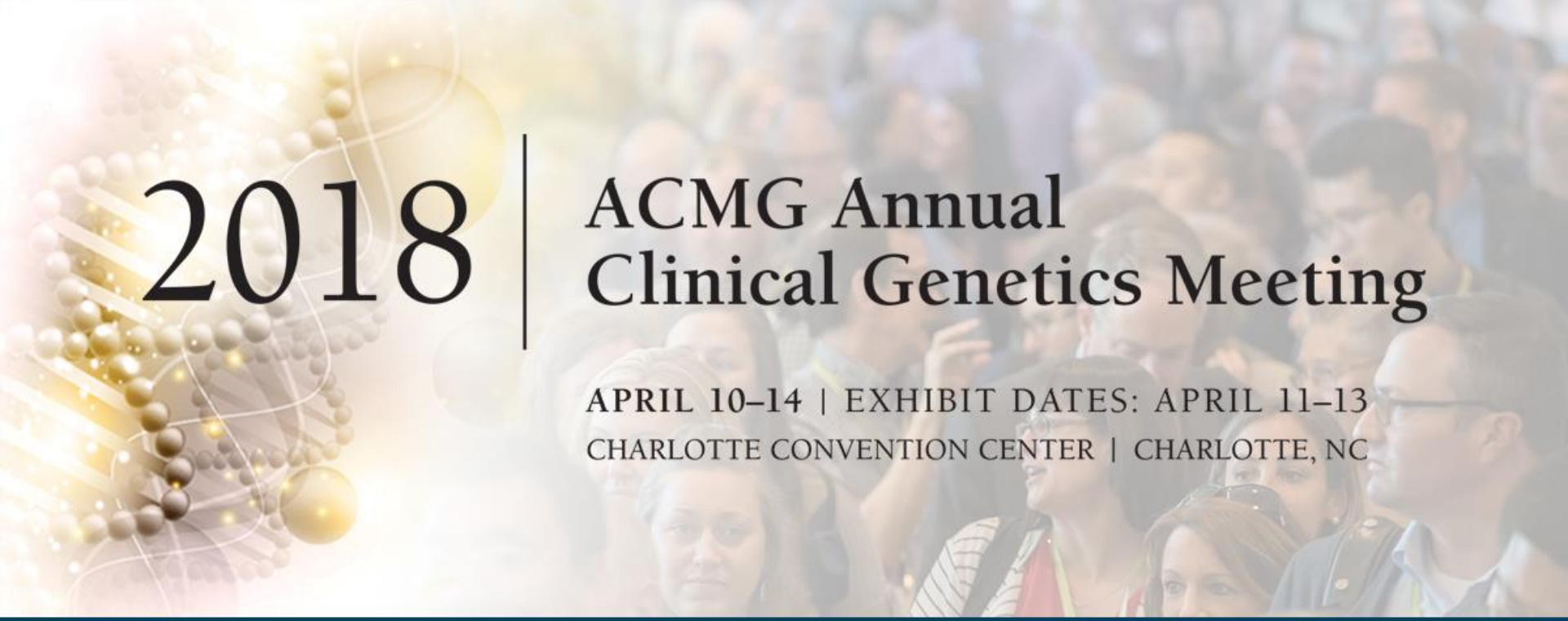
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

  - 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



Q & A

Please enter your questions into your GoToWebinar Panel



We're Exhibiting at ACMG 2018

Find us (and the t-shirts) in booth 1306!

Q & A

Please enter your questions into your GoToWebinar Panel