Custom Family Workflows

varSEQ

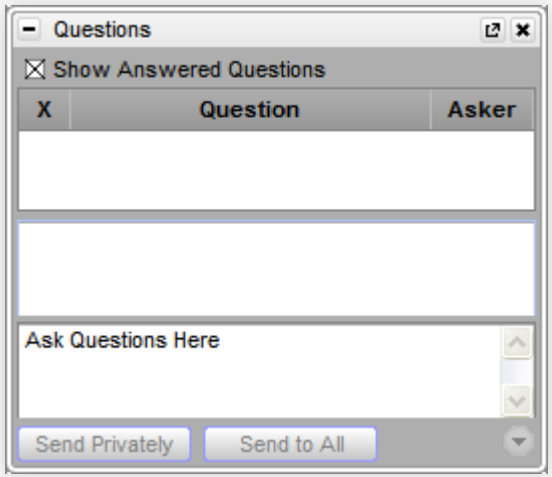
May 11th, 2016

Jami Bartole
Senior Field Application Scientist

GOLDEN HELIX
Accelerating the Quest for Significance™

Questions during the presentation

Use the Questions pane in your GoToWebinar window



Agenda

1. Overview of Golden Helix and VarSeq
2. Quad Workflow
3. Trio plus Unaffected Sibling Workflow
4. Two Affected Siblings Workflow

Golden Helix – Who We Are

Golden Helix is a global bioinformatics company founded in 1998.

gsk GlaxoSmithKline

We are cited in over 1000 peer-reviewed publications

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

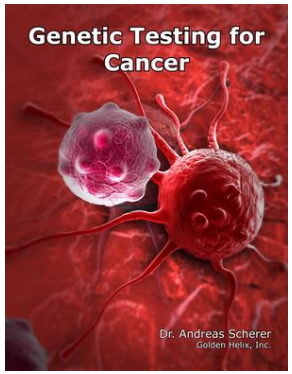
FDA

GOLDEN HELIX
Accelerating the Quest for Significance™

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

VarSeq Suite

- VCF Import
- Variant and Gene Annotations
- Various Algorithms
- Filter Variants

VSReports

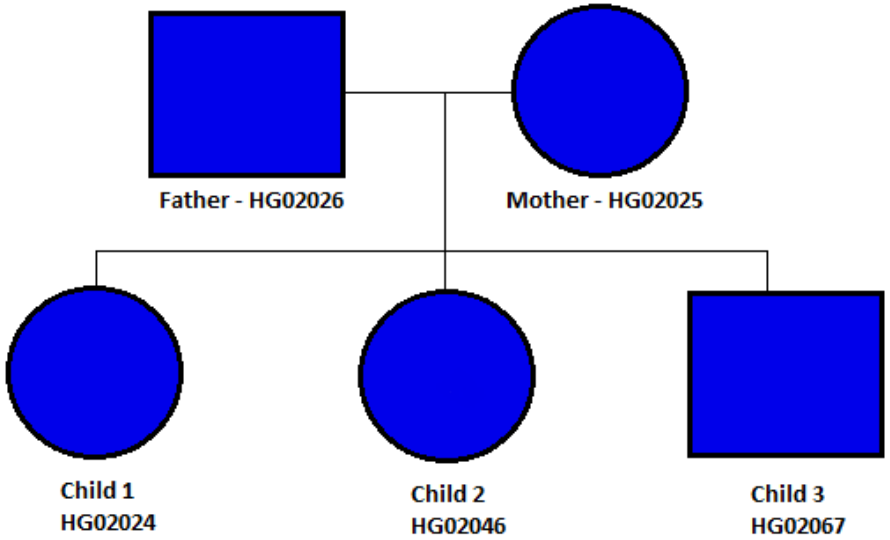
- Flag Variants
- Customized Reports (HTML, PDF, etc.)

VSPipeline

- Automate Workflow
- High-Throughput Analysis

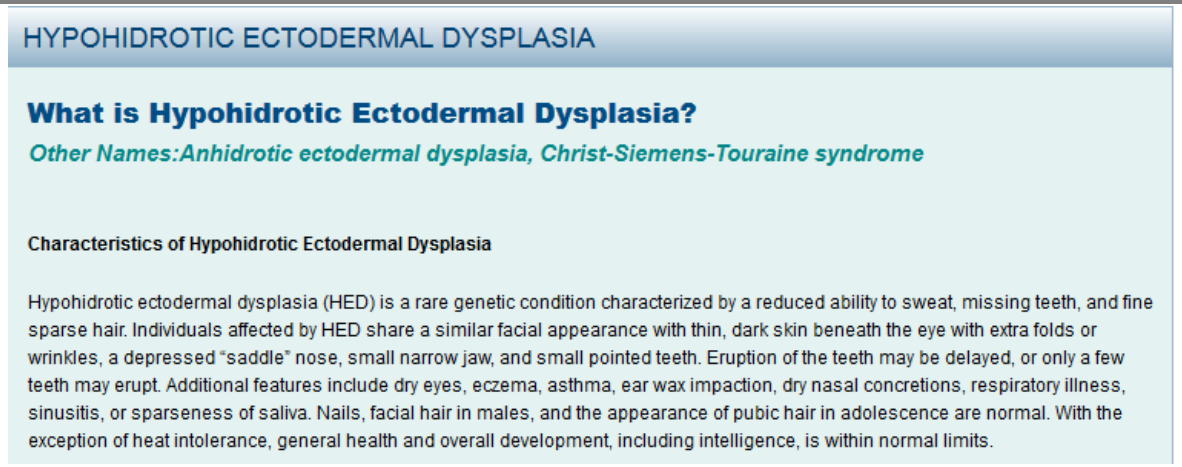
- Organize Samples into Projects
- Projects as Variant Frequency Annotations
- Centralized VSReports Hosting
- Scalable Technology
- Multiple Interfaces

Sample Data for our Custom Family



- **Family was simulated from the 1000 Genomes Vietnamese (KHV) related samples using the VN049 trio and the VN056 sibling pair.**
- **Family formed with two affected daughters and one unaffected son.**
- **BAM files and VCF files available for all five samples**
  - Downloaded Illumina Exome Alignment BAM files from the 1000 Genomes Phase 3 Project
  - GATK was used for variant calling

Phenotype for the Affected Samples



HYPOHIDROTIC ECTODERMAL DYSPLASIA

What is Hypohidrotic Ectodermal Dysplasia?

Other Names:Anhidrotic ectodermal dysplasia, Christ-Siemens-Touraine syndrome

Characteristics of Hypohidrotic Ectodermal Dysplasia

Hypohidrotic ectodermal dysplasia (HED) is a rare genetic condition characterized by a reduced ability to sweat, missing teeth, and fine sparse hair. Individuals affected by HED share a similar facial appearance with thin, dark skin beneath the eye with extra folds or wrinkles, a depressed "saddle" nose, small narrow jaw, and small pointed teeth. Eruption of the teeth may be delayed, or only a few teeth may erupt. Additional features include dry eyes, eczema, asthma, ear wax impaction, dry nasal concretions, respiratory illness, sinusitis, or sparseness of saliva. Nails, facial hair in males, and the appearance of pubic hair in adolescence are normal. With the exception of heat intolerance, general health and overall development, including intelligence, is within normal limits.

- HED may be inherited in one of 3 patterns: X-linked recessive (95%), autosomal recessive or autosomal dominant (5%)
- Changes or mutations in the EDA, EDAR, EDARADD, and WNT10A genes are most commonly associated with HED. These genes tell the body to make proteins that are needed early in life (before birth and shortly after) for the normal development of sweat glands, teeth, hair, skin, and other mucous glands.
- These four genes account for 90% of hypohidrotic/anhidrotic ectodermal dysplasia cases

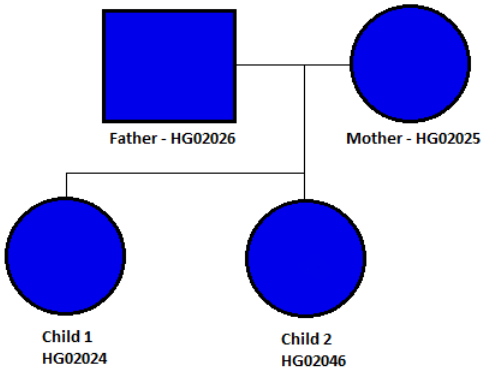
References:

- http://nfed.org/index.php/about_ed/hypohidrotic-ectodermal-dysplasia
- http://rarediseases.org/rare-diseases/hypohidrotic-ectodermal-dysplasia/#affected-populations
- http://www.ncbi.nlm.nih.gov/pubmed/17354266

Agenda

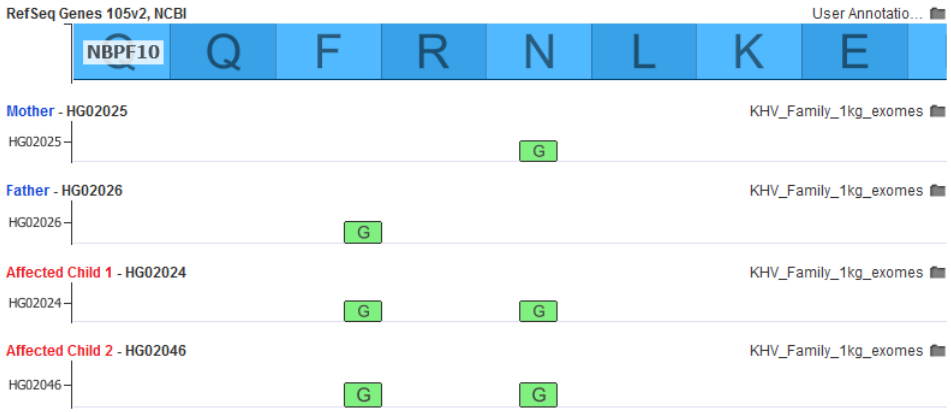
1. Overview of Golden Helix and VarSeq
2. Quad Workflow
3. Trio plus Unaffected Sibling Workflow
4. Two Affected Siblings Workflow

Quad Analysis Workflow



Mother, Father and Two Affected Siblings

- Using Variant Sets to look at the combined set of de Novo candidate variants between the two daughters.



- Using a custom Filter Chain to identify common variants between the two affected daughters.
  - de Novo Candidates
  - Compound Heterozygous Candidates

HPO Terms For Variant Sites

- HP:0007607: hypohidrotic ectodermal dysplasia

Group by Genes	HED PhoRank		
Gene Names	HED Gene Rank	HED Gene Score	HED Path
A1BG	0.343972	3.07775e-006	A1BG, GO:0008150 (biological_process), GO:0050789 (regulati...
A1CF	0.397163	3.82478e-006	A1CF, GO:0005737 (cytoplasm), EDARADD, HP:0007607 (Hyp...
A2M	0.241135	1.91239e-006	A2M, GO:0048863 (stem cell differentiation), GO:0030154 (cell...
A2ML1	0.765957	0.00086194	A2ML1, HP:0011354 (Generalized abnormality of skin), HP:000...

- Prioritize variants using PhoRank to be included in a custom Clinical Report!

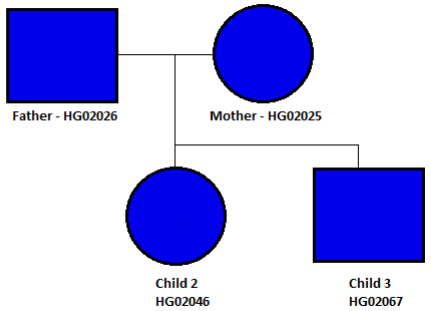
VarSeq Demonstration

Agenda

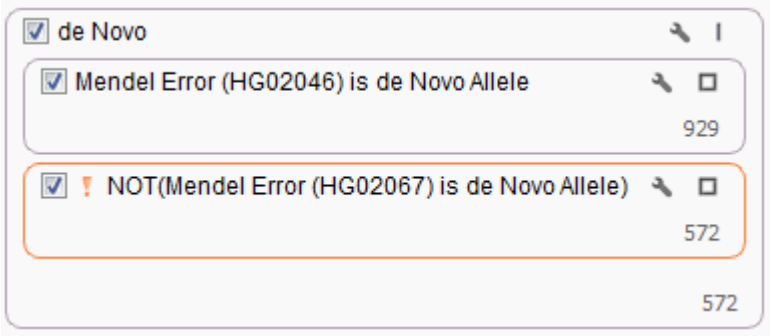
1. Overview of Golden Helix and VarSeq
2. Quad Workflow
3. Trio plus Unaffected Sibling Workflow
4. Two Affected Siblings Workflow

Trio + Unaffected Sibling Workflow

Mother, Father, Affected Daughter and Unaffected Son

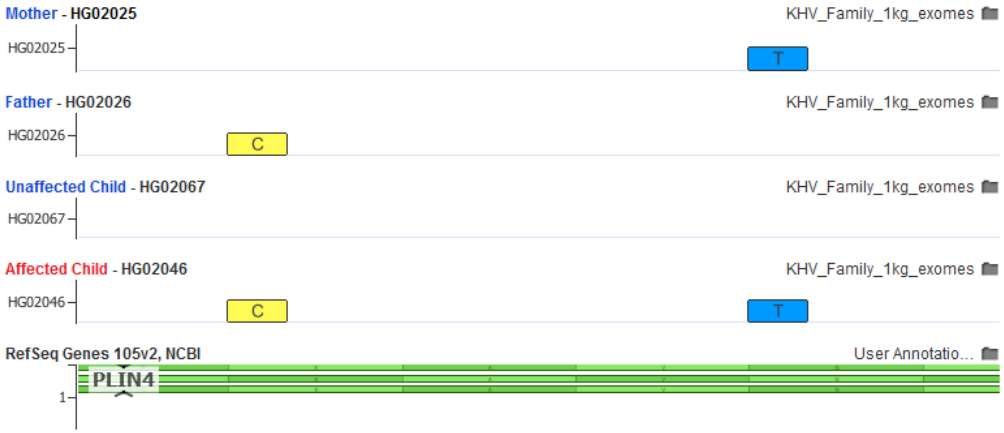


- Custom template creating separate sample specific filter cards for each child



- Inverting filter cards to remove variants present in unaffected sibling

- Using Count Allele algorithm to restrict the number of heterozygous calls for each variant

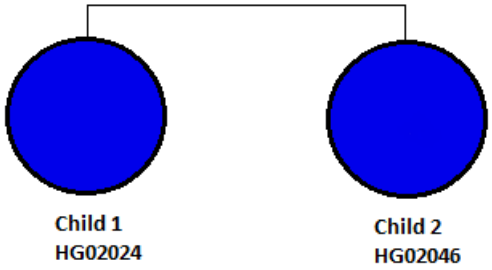


VarSeq Demonstration

Agenda

1. Overview of Golden Helix and VarSeq
2. Quad Workflow
3. Trio plus Unaffected Sibling Workflow
4. Two Affected Siblings Workflow

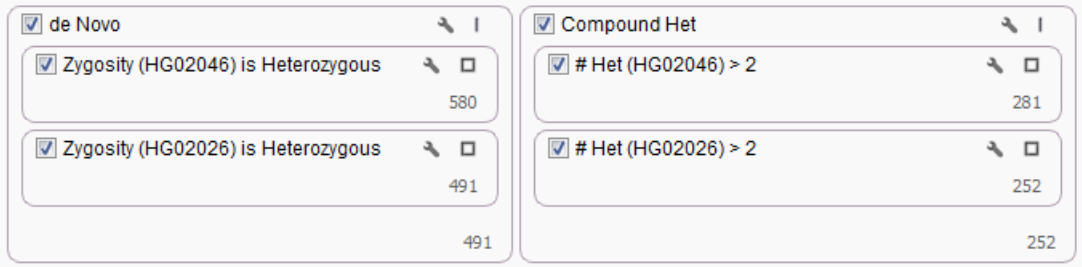
Affected Sibling Pair Workflow



Two Affected Daughters

- Using Custom template to set sample specific filters

- Using Genotype Zygosity Algorithm to find shared “de Novo” candidate variants
- Using Count Alleles by Gene Algorithm to find Shared “Compound Het” candidate regions



Variant Sites		CADD Scores 1.3			
Chr:Pos	Ref/Alt	Ref/Alt	Raw Score	PHRED Score	Estimated?
1:1431165	C/T	C/T	-0.195452	1.122	False
1:12785295	C/T	C/T	3.66079	23.2	False
1:12854090	T/A	T/A	7.00174	33	False
1:12854105	A/G	A/G	2.03281	16.42	False
1:12854162	C/T	C/T	-2.06069	0.001	False
1:12854188	T/C	T/C	0.760341	9.216	False
1:12854414	G/A	G/A	-0.751712	0.055	False
1:12856010	C/G	C/G	-1.43082	0.003	False
1:12856111	C/T	C/T	0.35293	6.194	False
1:12907358	T/C	T/C	-1.92604	0.001	False

- Prioritize variants using CADD annotation (Coming SOON to a VarSeq near you!)
  - CADD provides pre-computed scores for all possible 8.6 billion single-letter substitutions, as well as 20 million previously observed indels. For novel indels, the score is estimated using scores from flanking or deleted bases

VarSeq Demonstration

Summary

- Quad Workflow
  - Used Variant Sets to create a multi-sample VCF file, which includes de Novo Candidate variants present in either affected sample
  - Setting sample specific filter cards created a custom filter chain for de Novo and Compound Heterozygous Gene candidates in common between the two affected samples
  - Prioritized variants using PhoRank to be added to customized Clinical Report.

- Trio plus Unaffected Sibling Workflow
  - Used the ability to invert filter cards to exclude de Novo variants in the unaffected sample
  - Used Count Alleles algorithm to restrict the number of heterozygous calls for each variant position, which excluded hets present in the unaffected sample

- Two Affected Siblings Workflow
  - Used Genotype Zygosity to identify possible de Novo variants
  - Used Count Alleles by Gene to identify possible Compound Het genes
  - Looked at prioritizing variants based on new CADD annotation.

Questions or more info:

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