Making NGS Data Analysis Clinically Practical: Repeatable and Time-Effective Workflows

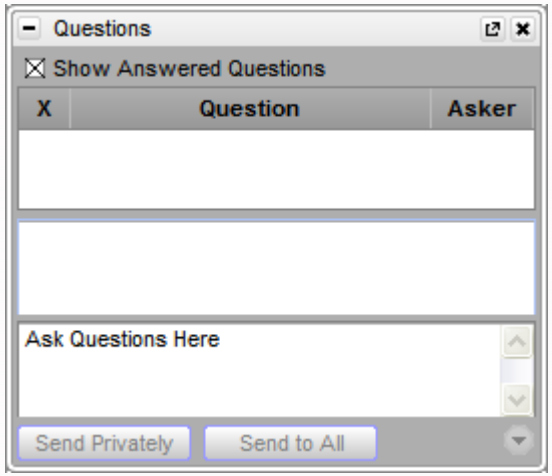
Autumn Laughbaum, Biostatistician

with introduction by Dr. Andreas Scherer, President & CEO

GOLDEN HELIX
Accelerating the Quest for Significance™

Questions during the presentation

Use the Questions pane in your GoToWebinar window



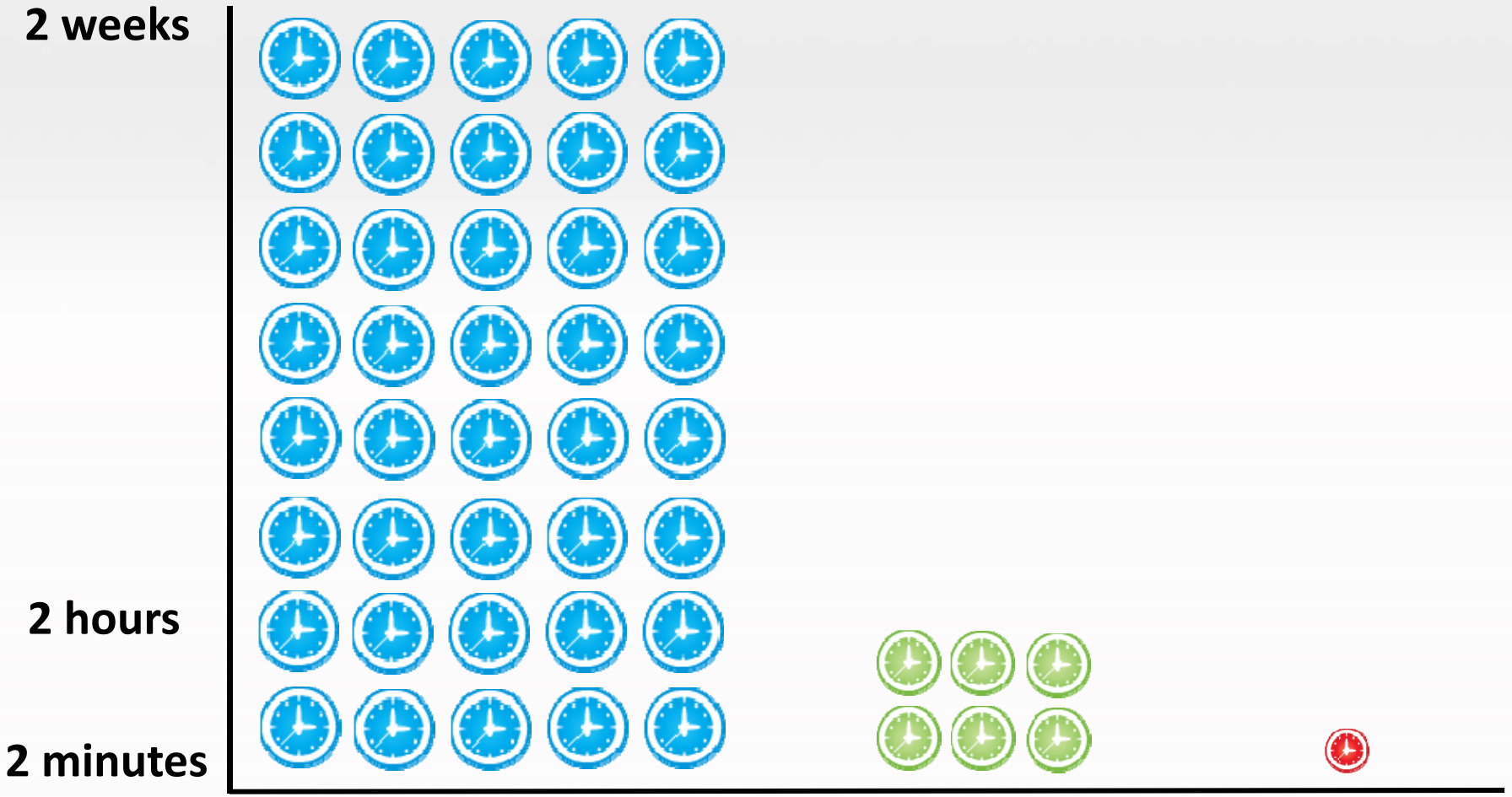
GOLDEN HELIX
Accelerating the Quest for Significance™

Introduction

Researcher

Clinician

Hands-on Time Savings



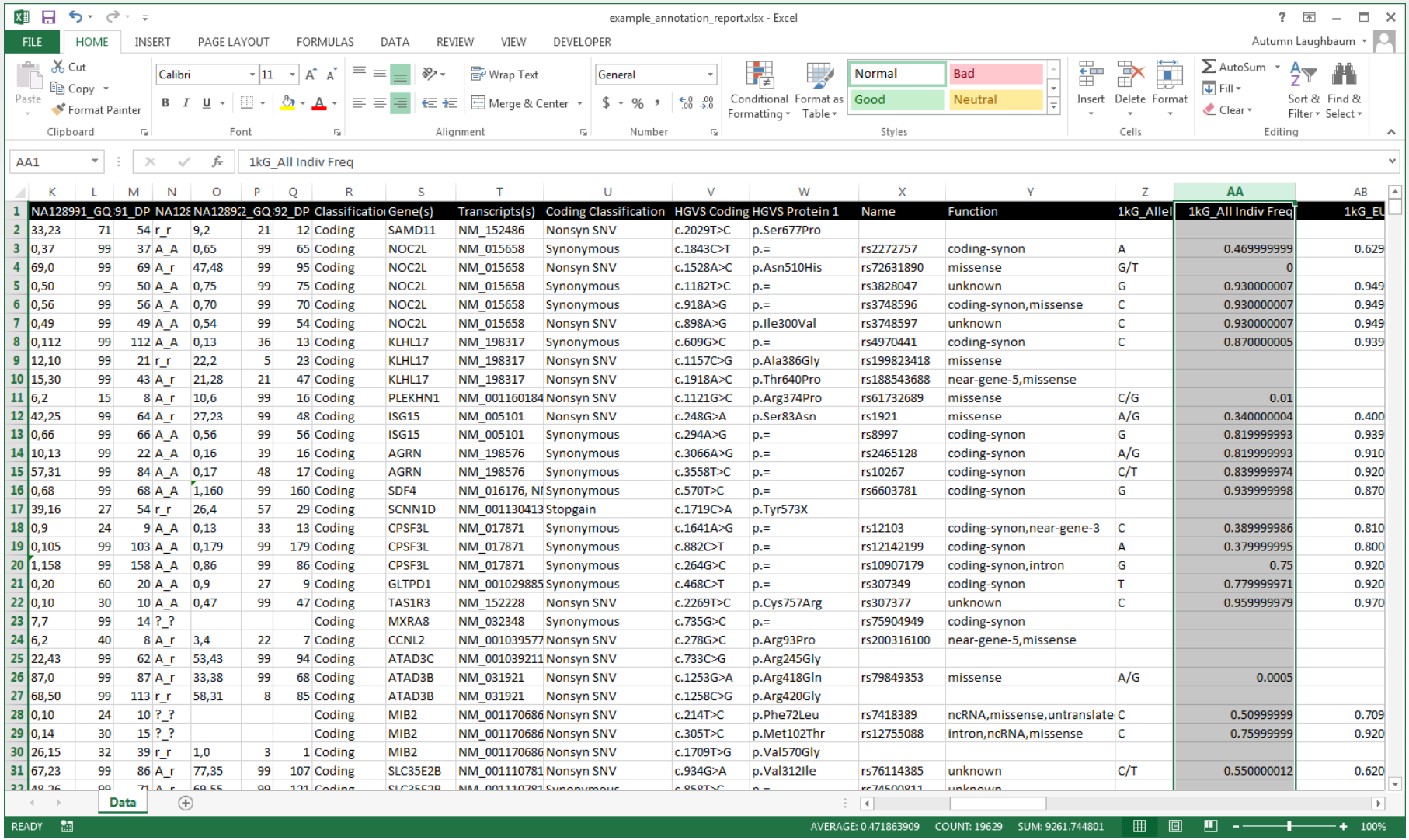
GOLDEN HELIX
Accelerating the Quest for Significance™

Today's Agenda

1. Status Quo
   - Moving from Excel > SVS > Automated Workflows
   - General sequencing workflow
2. Example Workflow: Ogden Syndrome
3. Example Workflow: Trio Analysis
4. Discussion

GOLDEN HELIX
Accelerating the Quest for Significance™

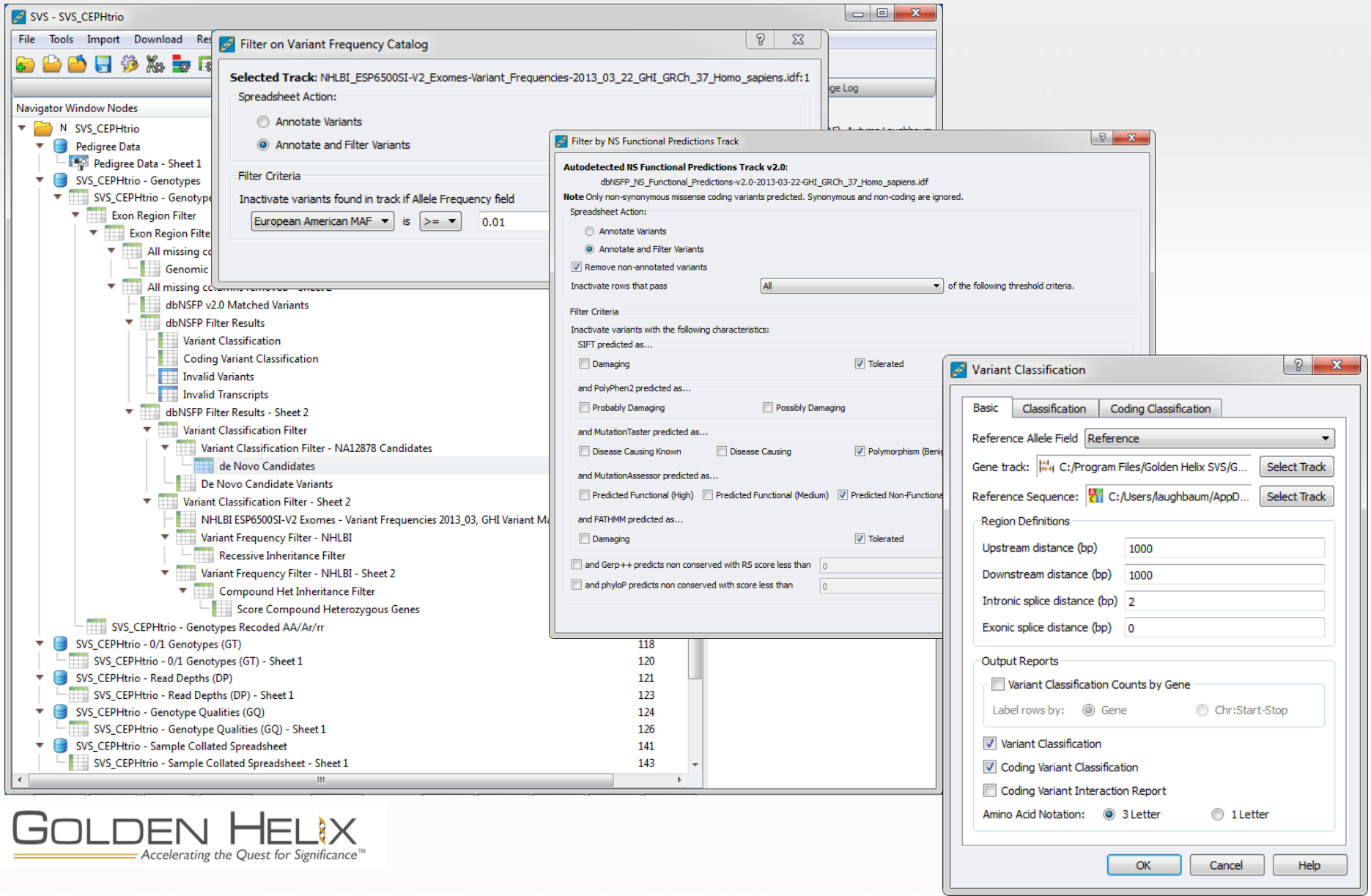
Analyzing Data from a secondary analysis pipeline: Using Excel



K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
NA12891_GQ	91_DP	NA128	NA12892_GQ	92_DP	Classificatio	Gene(s)	Transcripts(s)	Coding Classification	HGVS Coding	HGVS Protein 1	Name	Function	1kG_Allel	1kG_All Indiv Freq	1kG_EU
33,23	71	54	r_r	9,2	21	12	Coding	SAMD11	NM_152486	Nonsyn SNV	c.2029T>C	p.Ser677Pro					
0,37	99	37	A_A	0,65	99	65	Coding	NOC2L	NM_015658	Synonymous	c.1843C>T	p.=	rs2272757	coding-synon	A	0.469999999	0.629
69,0	99	69	A_r	47,48	99	95	Coding	NOC2L	NM_015658	Nonsyn SNV	c.1528A>C	p.Asn510His	rs72631890	missense	G/T	0	
0,50	99	50	A_A	0,75	99	75	Coding	NOC2L	NM_015658	Synonymous	c.1182T>C	p.=	rs3828047	unknown	G	0.930000007	0.949
0,56	99	56	A_A	0,70	99	70	Coding	NOC2L	NM_015658	Synonymous	c.918A>G	p.=	rs3748596	coding-synon,missense	C	0.930000007	0.949
0,49	99	49	A_A	0,54	99	54	Coding	NOC2L	NM_015658	Nonsyn SNV	c.898A>G	p.Ile300Val	rs3748597	unknown	C	0.930000007	0.949
0,112	99	112	A_A	0,13	36	13	Coding	KLHL17	NM_198317	Synonymous	c.609G>C	p.=	rs4970441	coding-synon	C	0.870000005	0.939
12,10	99	21	r_r	22,2	5	23	Coding	KLHL17	NM_198317	Nonsyn SNV	c.1157C>G	p.Ala386Gly	rs199823418	missense			
15,30	99	43	A_r	21,28	21	47	Coding	KLHL17	NM_198317	Nonsyn SNV	c.1918A>C	p.Thr640Pro	rs188543688	near-gene-5,missense			
6,2	15	8	A_r	10,6	99	16	Coding	PLEKHN1	NM_001160184	Nonsyn SNV	c.1121G>C	p.Arg374Pro	rs61732689	missense	C/G	0.01	
42,25	99	64	A_r	27,23	99	48	Coding	ISG15	NM_005101	Nonsyn SNV	c.248G>A	p.Ser83Asn	rs1921	missense	A/G	0.340000004	0.400
0,66	99	66	A_A	0,56	99	56	Coding	ISG15	NM_005101	Synonymous	c.294A>G	p.=	rs8997	coding-synon	G	0.819999993	0.939
10,13	99	22	A_A	0,16	39	16	Coding	AGRN	NM_198576	Synonymous	c.3066A>G	p.=	rs2465128	coding-synon	A/G	0.819999993	0.910
57,31	99	84	A_A	0,17	48	17	Coding	AGRN	NM_198576	Synonymous	c.3558T>C	p.=	rs10267	coding-synon	C/T	0.839999974	0.920
0,68	99	68	A_A	1,160	99	160	Coding	SDF4	NM_016176, NI	Synonymous	c.570T>C	p.=	rs6603781	coding-synon	G	0.939999998	0.870
39,16	27	54	r_r	26,4	57	29	Coding	SCNN1D	NM_001130413	Stopgain	c.1719C>A	p.Tyr573X					
0,9	24	9	A_A	0,13	33	13	Coding	CPSF3L	NM_017871	Synonymous	c.1641A>G	p.=	rs12103	coding-synon,near-gene-3	C	0.389999986	0.810
0,105	99	103	A_A	0,179	99	179	Coding	CPSF3L	NM_017871	Synonymous	c.882C>T	p.=	rs12142199	coding-synon	A	0.379999995	0.800
1,158	99	158	A_A	0,86	99	86	Coding	CPSF3L	NM_017871	Synonymous	c.264G>C	p.=	rs10907179	coding-synon,intron	G	0.75	0.920
0,20	60	20	A_A	0,9	27	9	Coding	GLTPD1	NM_001029885	Synonymous	c.468C>T	p.=	rs307349	coding-synon	T	0.779999971	0.920
0,10	30	10	A_A	0,47	99	47	Coding	TAS1R3	NM_152228	Nonsyn SNV	c.2269T>C	p.Cys757Arg	rs307377	unknown	C	0.959999979	0.970
7,7	99	14	?_?				Coding	MXRA8	NM_032348	Synonymous	c.735G>C	p.=	rs75904949	coding-synon			
6,2	40	8	A_r	3,4	22	7	Coding	CCNL2	NM_001039577	Nonsyn SNV	c.278G>C	p.Arg93Pro	rs200316100	near-gene-5,missense			
22,43	99	62	A_r	53,43	99	94	Coding	ATAD3C	NM_001039211	Nonsyn SNV	c.733C>G	p.Arg245Gly					
87,0	99	87	A_r	33,38	99	68	Coding	ATAD3B	NM_031921	Nonsyn SNV	c.1253G>A	p.Arg418Gln	rs79849353	missense	A/G	0.0005	
68,50	99	113	r_r	58,31	8	85	Coding	ATAD3B	NM_031921	Nonsyn SNV	c.1258C>G	p.Arg420Gly					
0,10	24	10	?_?				Coding	MIB2	NM_001170686	Nonsyn SNV	c.214T>C	p.Phe72Leu	rs7418389	ncRNA,missense,untranslate	C	0.50999999	0.709
0,14	30	15	?_?				Coding	MIB2	NM_001170686	Nonsyn SNV	c.305T>C	p.Met102Thr	rs12755088	intron,ncRNA,missense	C	0.75999999	0.920
26,15	32	39	r_r	1,0	3	1	Coding	MIB2	NM_001170686	Nonsyn SNV	c.1709T>G	p.Val570Gly					
67,23	99	86	A_r	77,35	99	107	Coding	SLC35E2B	NM_001110781	Nonsyn SNV	c.934G>A	p.Val312Ile	rs76114385	unknown	C/T	0.550000012	0.620
48,26	99	71	A_r	69,55	99	121	Coding	SLC35E2B	NM_001110781	Synonymous	c.858T>C	p.=	rs74500811	unknown			

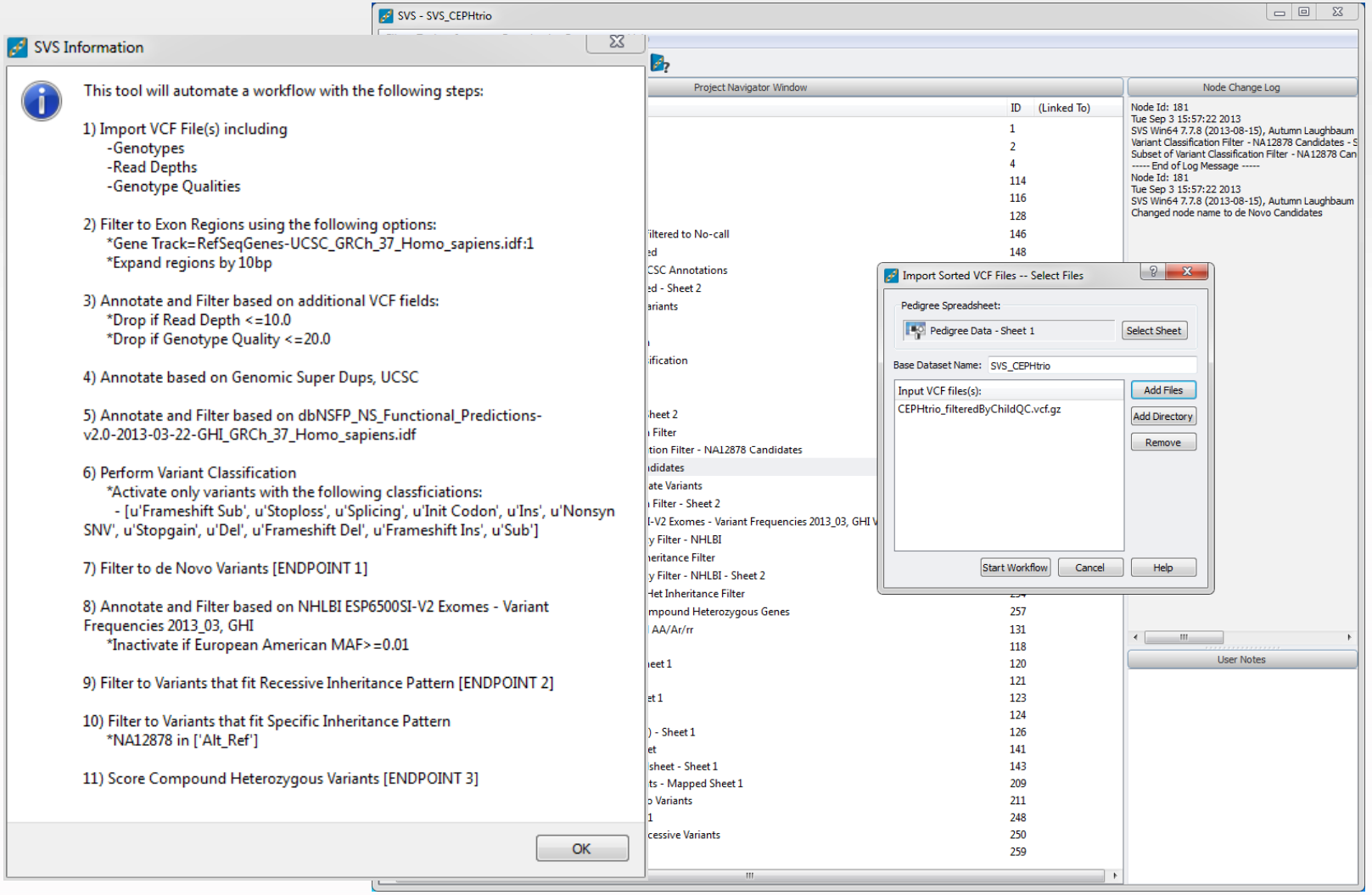
GOLDEN HELIX
Accelerating the Quest for Significance™

Analyzing Data from a secondary analysis pipeline: Using SVS



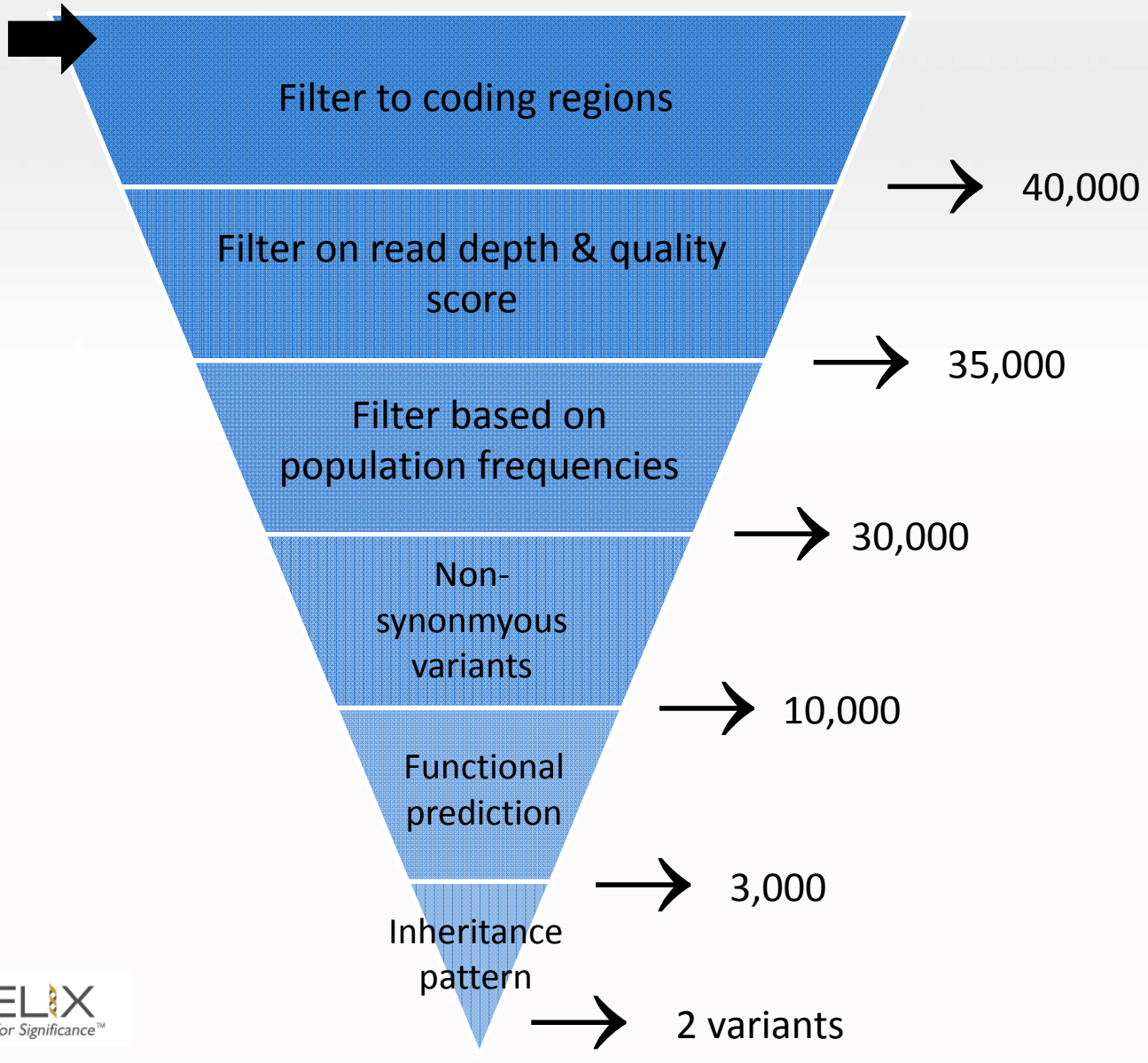
GOLDEN HELIX
Accelerating the Quest for Significance™

Analyzing Data from a secondary analysis pipeline: Automated Workflow



Analyzing Sequencing Data

Data from secondary analysis pipeline (VCF) – **2 million variants**

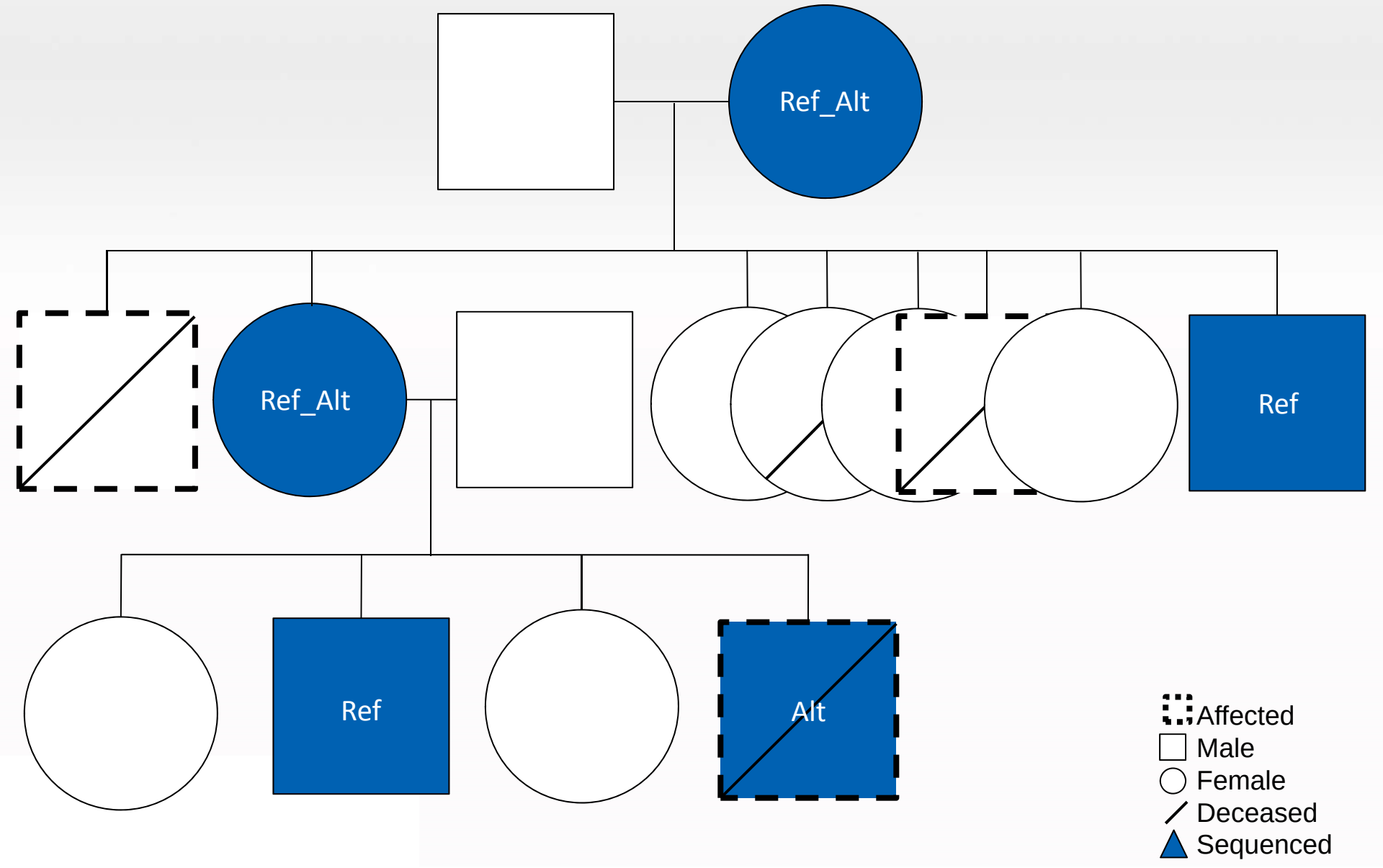


GOLDEN HELIX
Accelerating the Quest for Significance™

Workflow Example – Ogden Syndrome

WORKFLOW EXAMPLE – TRIO ANALYSIS

Pedigree

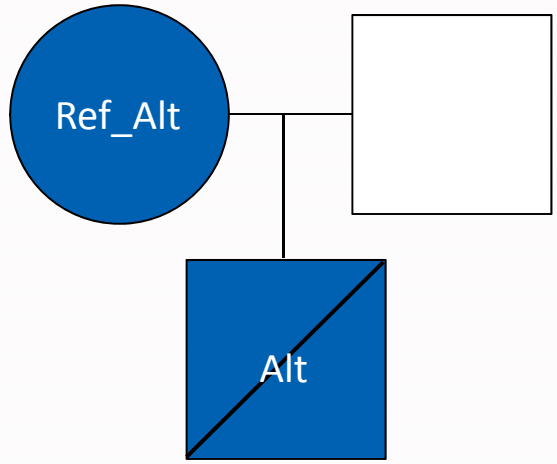

Ref_Alt
Ref_Alt
Ref
Ref
Alt
Affected
Male
Female
Deceased
Sequenced

SVS Demo

Ogden Syndrome

5 Samples
107,000 variants

Filter based on VCF quality metrics → Filter based on population frequencies → Annotate based on functional predictions → Filter based on variant classification → Filter based on inheritance pattern



1 damaging variant

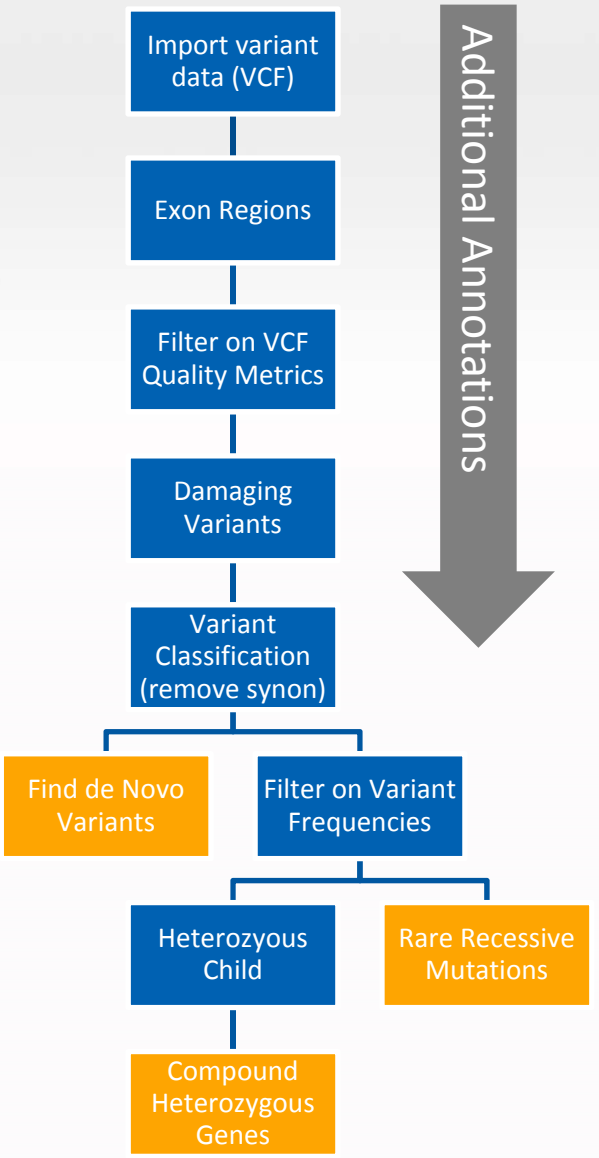
GOLDEN HELIX
Accelerating the Quest for Significance™

Workflow Example – Trio Analysis

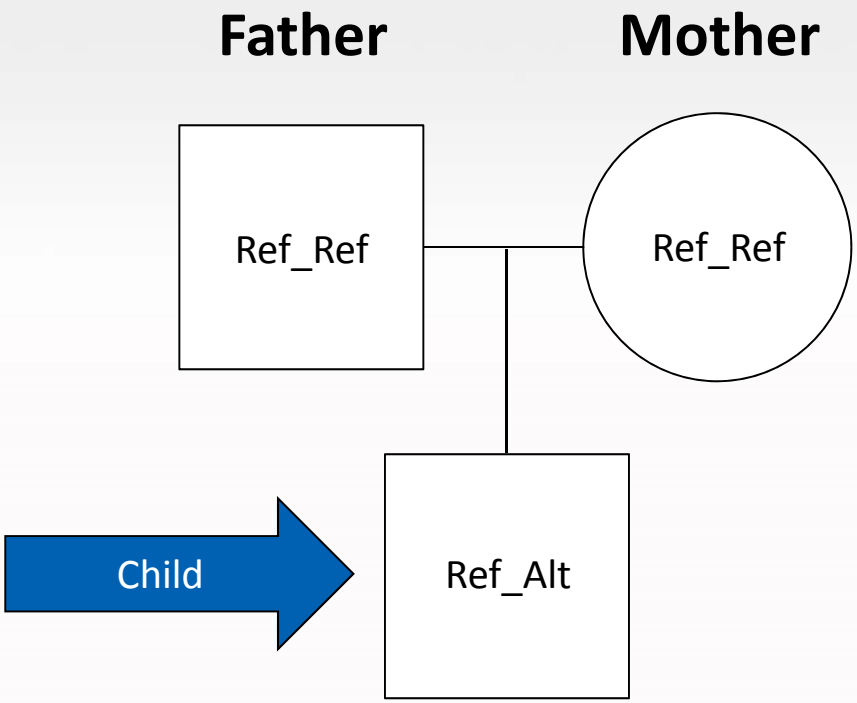
WORKFLOW EXAMPLE – TRIO ANALYSIS

Trio Analysis – 3 Distinct Mutation types

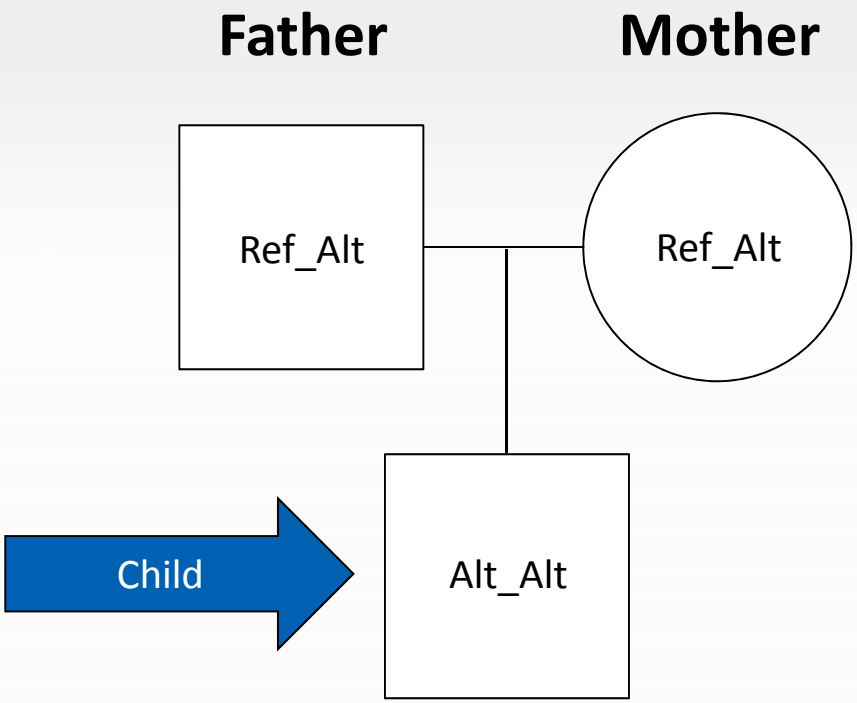
- de Novo Mutations
- Rare Recessive Mutations
- Compound Heterozygous Mutations



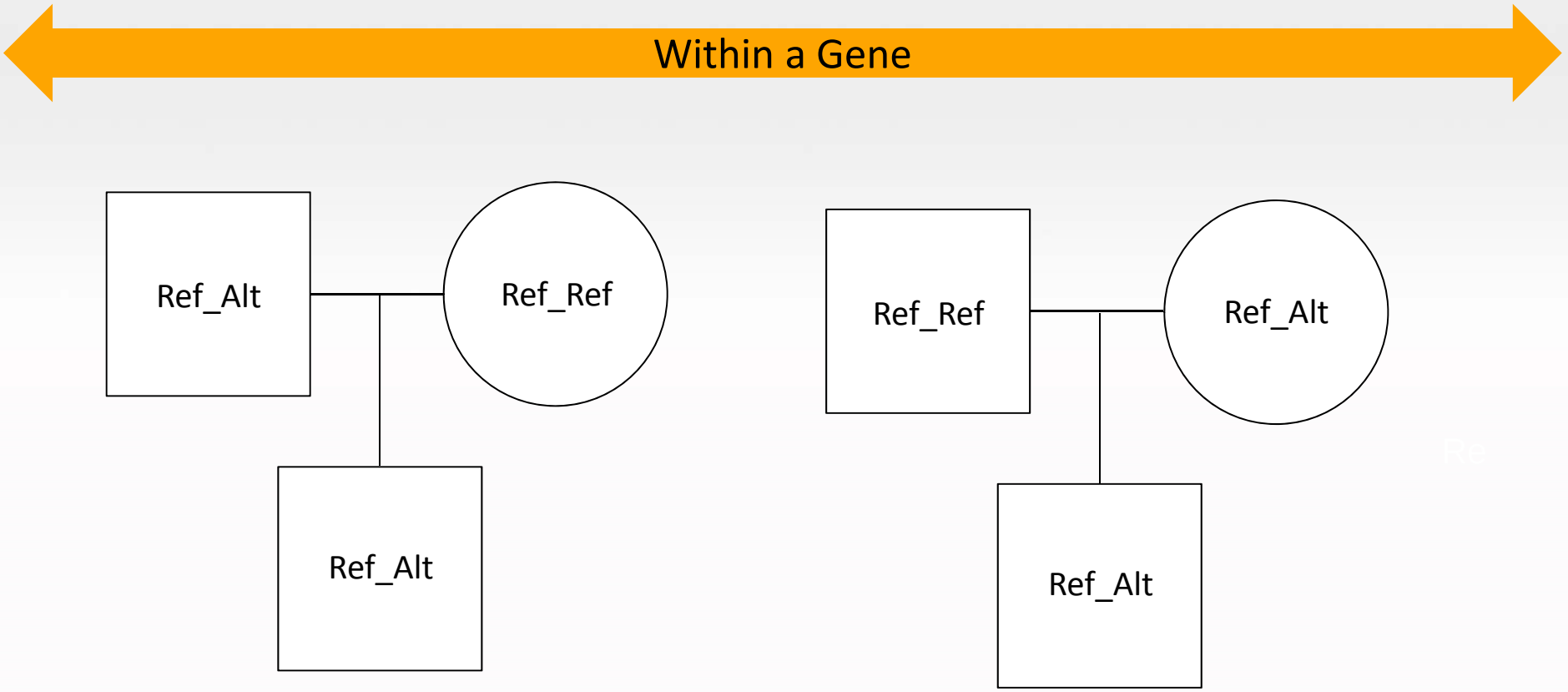
de Novo Mutation



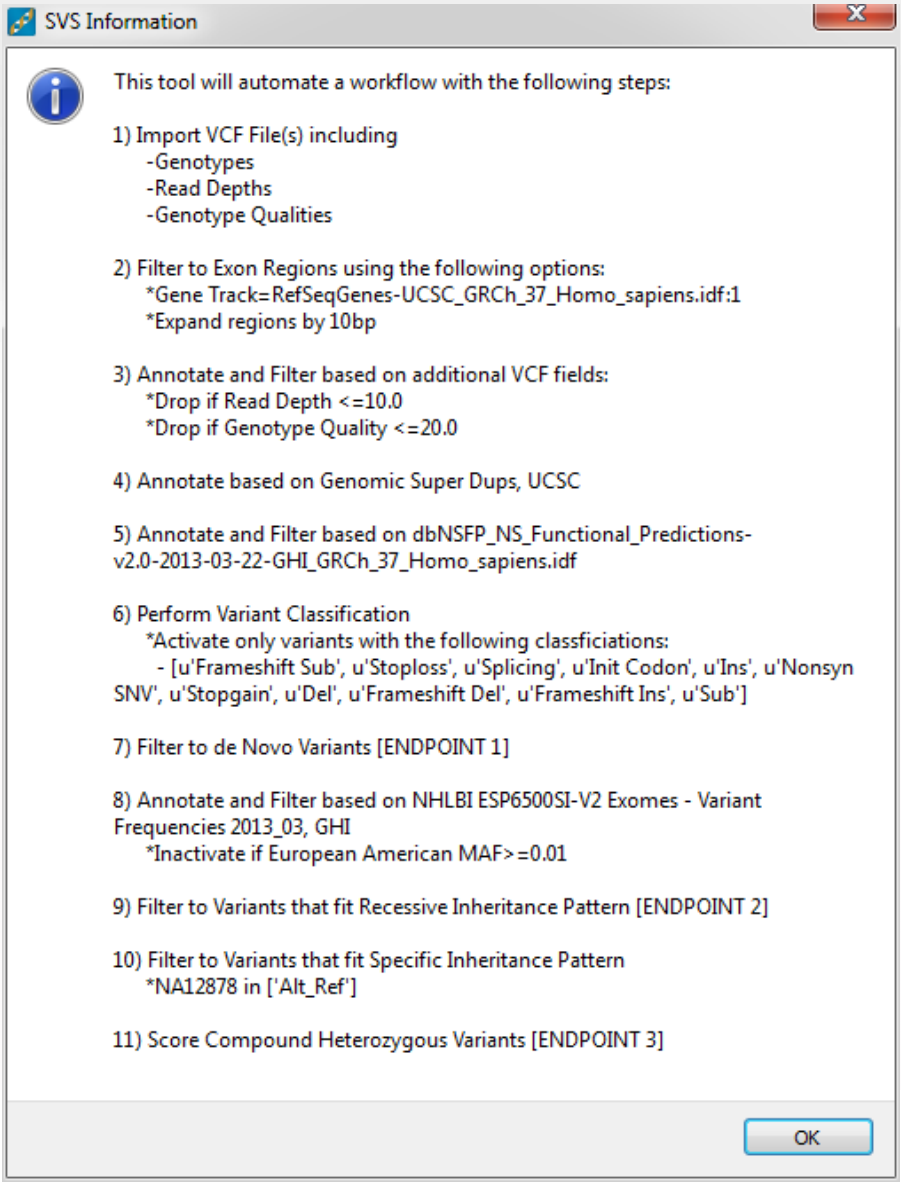
Rare Recessive Mutation



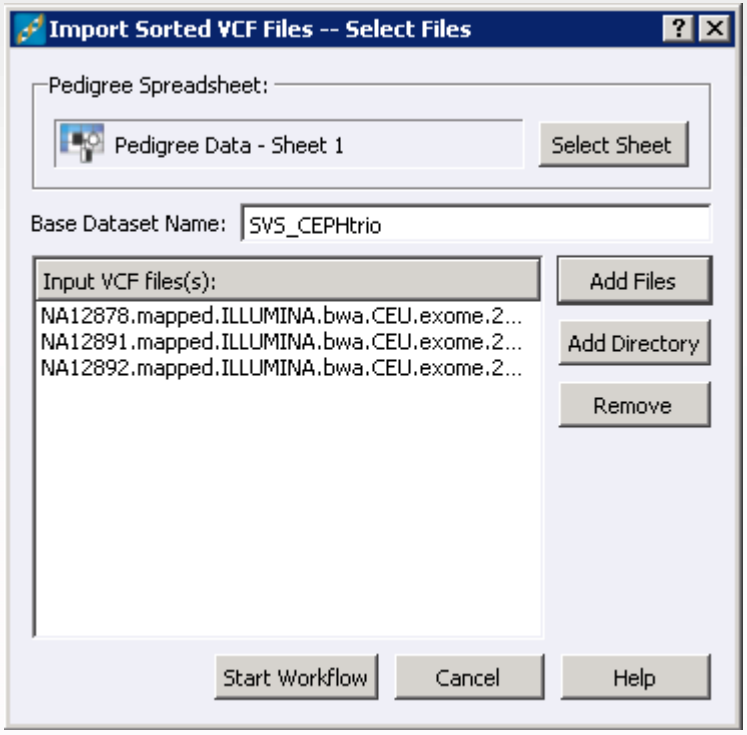
Compound Heterozygous Mutation

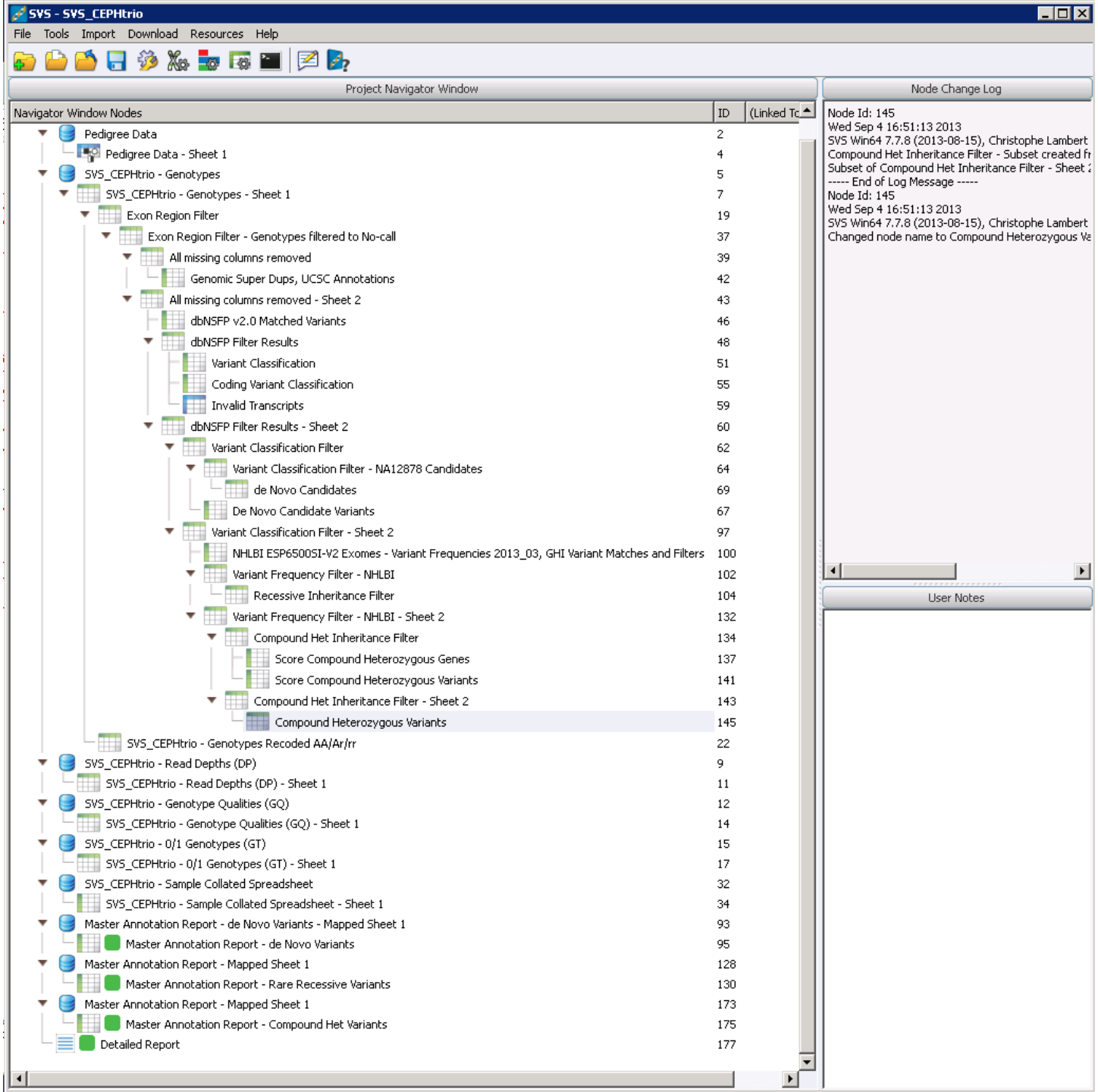


Trio Analysis – Automated Workflow

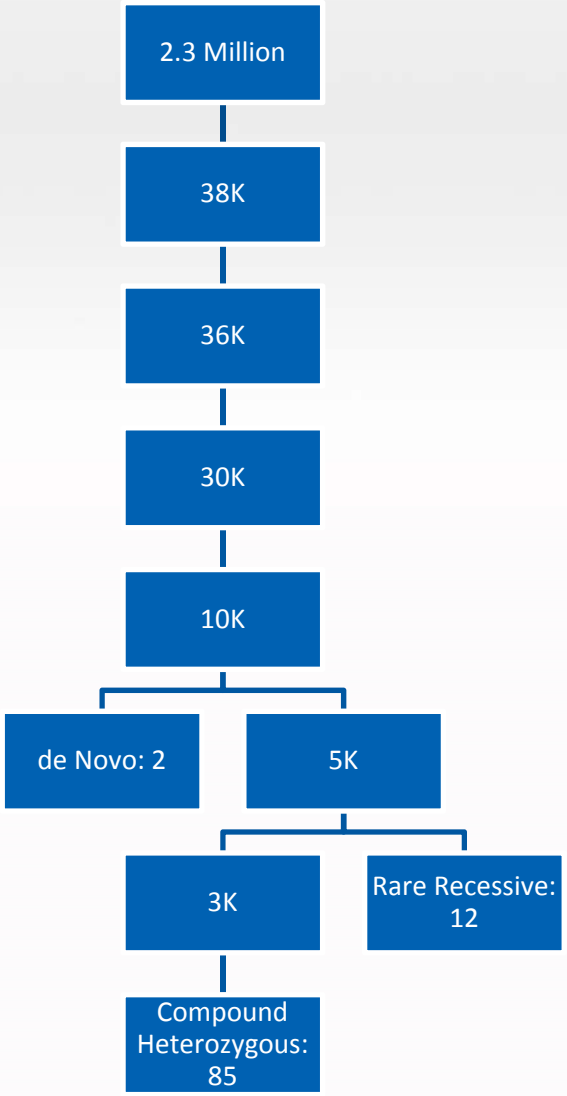


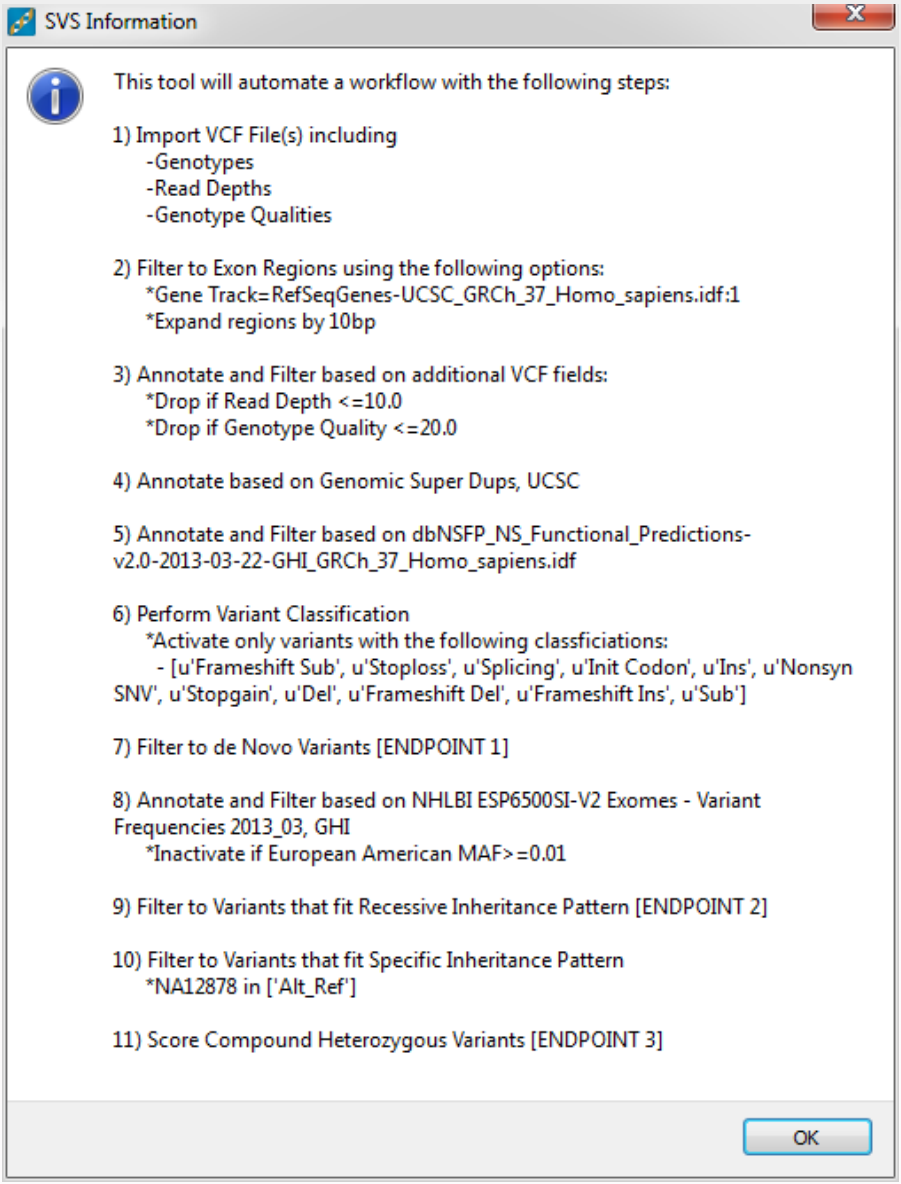
Trio Analysis – Automated Workflow




SVS - SVS_CEPHtrio
File Tools Import Download Resources Help
Project Navigator Window
Navigator Window Nodes
ID
(Linked To
Pedigree Data 2
Pedigree Data - Sheet 1 4
SVS_CEPHtrio - Genotypes 5
SVS_CEPHtrio - Genotypes - Sheet 1 7
Exon Region Filter 19
Exon Region Filter - Genotypes filtered to No-call 37
All missing columns removed 39
Genomic Super Dups, UCSC Annotations 42
All missing columns removed - Sheet 2 43
dbNSFP v2.0 Matched Variants 46
dbNSFP Filter Results 48
Variant Classification 51
Coding Variant Classification 55
Invalid Transcripts 59
dbNSFP Filter Results - Sheet 2 60
Variant Classification Filter 62
Variant Classification Filter - NA12878 Candidates 64
de Novo Candidates 69
De Novo Candidate Variants 67
Variant Classification Filter - Sheet 2 97
NHLBI ESP6500SI-V2 Exomes - Variant Frequencies 2013_03, GHI Variant Matches and Filters 100
Variant Frequency Filter - NHLBI 102
Recessive Inheritance Filter 104
Variant Frequency Filter - NHLBI - Sheet 2 132
Compound Het Inheritance Filter 134
Score Compound Heterozygous Genes 137
Score Compound Heterozygous Variants 141
Compound Het Inheritance Filter - Sheet 2 143
Compound Heterozygous Variants 145
SVS_CEPHtrio - Genotypes Recoded AA/Ar/rr 22
SVS_CEPHtrio - Read Depths (DP) 9
SVS_CEPHtrio - Read Depths (DP) - Sheet 1 11
SVS_CEPHtrio - Genotype Qualities (GQ) 12
SVS_CEPHtrio - Genotype Qualities (GQ) - Sheet 1 14
SVS_CEPHtrio - 0/1 Genotypes (GT) 15
SVS_CEPHtrio - 0/1 Genotypes (GT) - Sheet 1 17
SVS_CEPHtrio - Sample Collated Spreadsheet 32
SVS_CEPHtrio - Sample Collated Spreadsheet - Sheet 1 34
Master Annotation Report - de Novo Variants - Mapped Sheet 1 93
Master Annotation Report - de Novo Variants 95
Master Annotation Report - Mapped Sheet 1 128
Master Annotation Report - Rare Recessive Variants 130
Master Annotation Report - Mapped Sheet 1 173
Master Annotation Report - Compound Het Variants 175
Detailed Report 177
Node Change Log
Node Id: 145
Wed Sep 4 16:51:13 2013
SVS Win64 7.7.8 (2013-08-15), Christophe Lambert
Compound Het Inheritance Filter - Subset created fr
Subset of Compound Het Inheritance Filter - Sheet 2
----- End of Log Message -----
Node Id: 145
Wed Sep 4 16:51:13 2013
SVS Win64 7.7.8 (2013-08-15), Christophe Lambert
Changed node name to Compound Heterozygous Va
User Notes

Trio Analysis – Automated Workflow





GOLDEN HELIX
Accelerating the Quest for Significance™

Golden Helix SVS Workflow Automation

Results from automated workflow "NGS Trio Complete Workflow" in project SVS_CEPHtrio at 2013-09-04 15:51:57

Final deNovo Variant Count: 2

*4:144801662-SNV at 4:144801662, Ref/Alt=C/G,T,T, gene=GYPE, transcript=NM_002102, NM_198682, p.Gly13Ala
*17:18647625-SNV at 17:18647625, Ref/Alt=T/A, gene=FBXW10, transcript=NM_031456, p.Ile23Asn

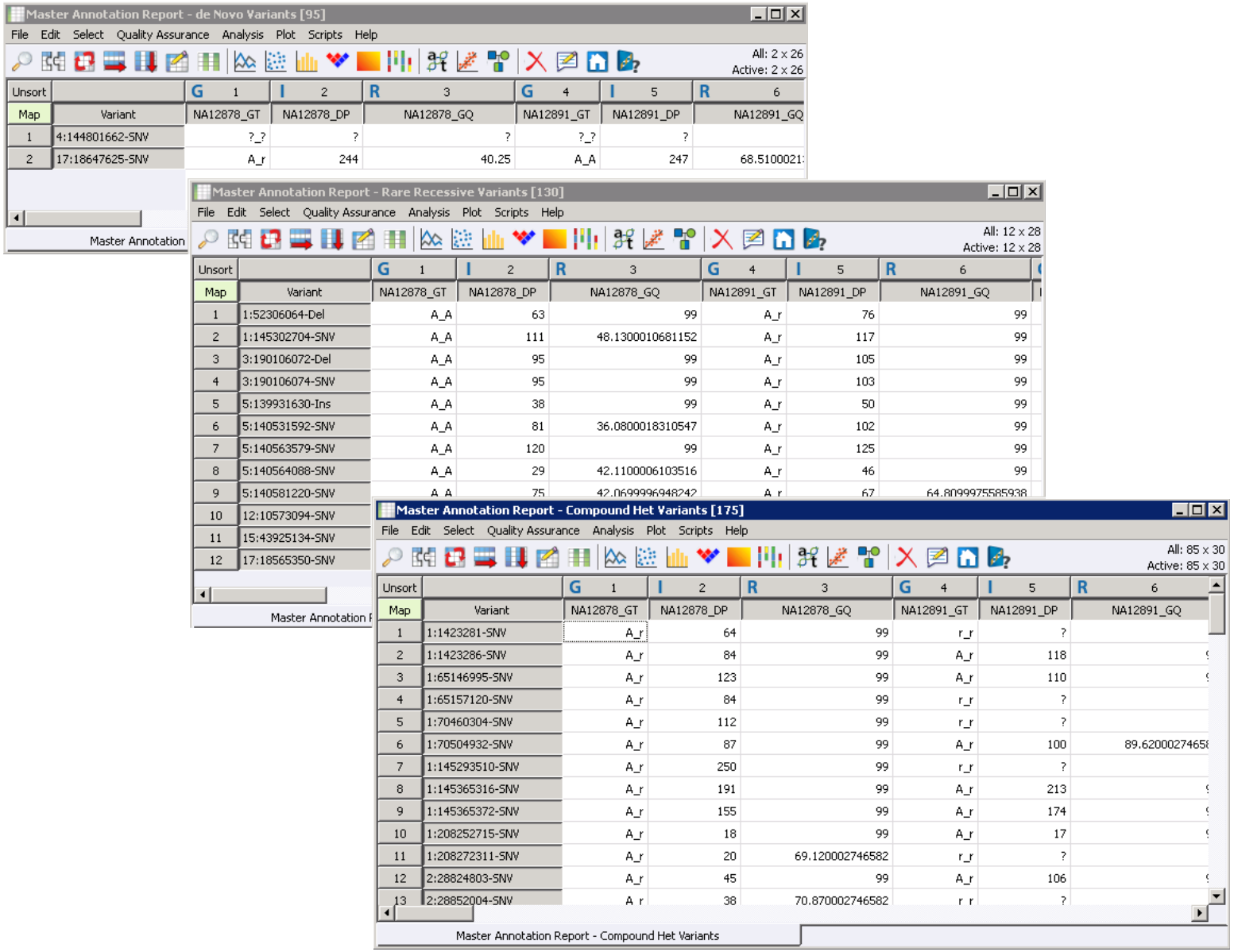
Final Rare Recessive Variant Count: 12

*1:52306064-Del at 1:52306064, Ref/Alt=TCT/-, gene=NRD1, transcript=NM_001101662, NM_001242361, NM_002525, p.Glu22del
*1:145302704-SNV at 1:145302704, Ref/Alt=A/G, gene=NBPF10, transcript=NM_001039703, p.Lys381Arg
*3:190106072-Del at 3:190106072, Ref/Alt=G/-, gene=CLDN16, transcript=NM_006580, p.Ala56fs
*3:190106074-SNV at 3:190106074, Ref/Alt=G/C, gene=CLDN16, transcript=NM_006580, p.Ala56Pro
*5:139931630-Ins at 5:139931630, Ref/Alt=-/G, gene=SRA1, transcript=NM_001035235, p.Val110fs
*5:140531592-SNV at 5:140531592, Ref/Alt=C/T, gene=PCDHB6, transcript=NM_018939, p.Thr585Ile
*5:140563579-SNV at 5:140563579, Ref/Alt=C/T, gene=PCDHB16, transcript=NM_020957, p.Thr482Ile
*5:140564088-SNV at 5:140564088, Ref/Alt=C/T, gene=PCDHB16, transcript=NM_020957, p.Arg652Cys
*5:140581220-SNV at 5:140581220, Ref/Alt=C/A, gene=PCDHB11, transcript=NM_018931, p.Arg625Ser
*12:10573094-SNV at 12:10573094, Ref/Alt=C/G, gene=KLRC3, transcript=NM_002261, NM_007333, p.Trp19Ser
*15:43925134-SNV at 15:43925134, Ref/Alt=T/A, gene=CATSPER2, transcript=NM_172095, ?
*17:18565350-SNV at 17:18565350, Ref/Alt=G/C, gene=ZNF286B, transcript=NM_001145045, p.Thr490Ser

Final Compound Het Variant Count: 85

*1:1423281-SNV at 1:1423281, Ref/Alt=G/A, gene=ATAD3B, transcript=NM_031921, p.Arg418Gln
*1:1423286-SNV at 1:1423286, Ref/Alt=C/G, gene=ATAD3B, transcript=NM_031921, p.Arg420Gly
*1:65146995-SNV at 1:65146995, Ref/Alt=A/C, gene=CACHD1, transcript=NM_020925, p.Asp1103Ala
*1:65157120-SNV at 1:65157120, Ref/Alt=C/G, gene=CACHD1, transcript=NM_020925, p.Pro1183Arg
*1:70460304-SNV at 1:70460304, Ref/Alt=A/G, gene=LRRC7, transcript=NM_020794, p.Asn293Ser
*1:70504932-SNV at 1:70504932, Ref/Alt=C/A, gene=LRRC7, transcript=NM_020794, p.Ala1104Glu

Automated Workflow Results



Master Annotation Report - de Novo Variants [95]

File Edit Select Quality Assurance Analysis Plot Scripts Help

All: 2 x 26
Active: 2 x 26

Unsort		G 1	I 2	R 3	G 4	I 5	R 6
Map	Variant	NA12878_GT	NA12878_DP	NA12878_GQ	NA12891_GT	NA12891_DP	NA12891_GQ
1	4:144801662-SNV	?_?	?	?	?_?	?	
2	17:18647625-SNV	A_r	244	40.25	A_A	247	68.5100021:

Master Annotation

Master Annotation Report - Rare Recessive Variants [130]

File Edit Select Quality Assurance Analysis Plot Scripts Help

All: 12 x 28
Active: 12 x 28

Unsort		G 1	I 2	R 3	G 4	I 5	R 6
Map	Variant	NA12878_GT	NA12878_DP	NA12878_GQ	NA12891_GT	NA12891_DP	NA12891_GQ
1	1:52306064-Del	A_A	63	99	A_r	76	99
2	1:145302704-SNV	A_A	111	48.1300010681152	A_r	117	99
3	3:190106072-Del	A_A	95	99	A_r	105	99
4	3:190106074-SNV	A_A	95	99	A_r	103	99
5	5:139931630-Ins	A_A	38	99	A_r	50	99
6	5:140531592-SNV	A_A	81	36.0800018310547	A_r	102	99
7	5:140563579-SNV	A_A	120	99	A_r	125	99
8	5:140564088-SNV	A_A	29	42.1100006103516	A_r	46	99
9	5:140581220-SNV	A_A	75	42.0699996948242	A_r	67	64.8099975585938
10	12:10573094-SNV						
11	15:43925134-SNV						
12	17:18565350-SNV						

Master Annotation R

Master Annotation Report - Compound Het Variants [175]

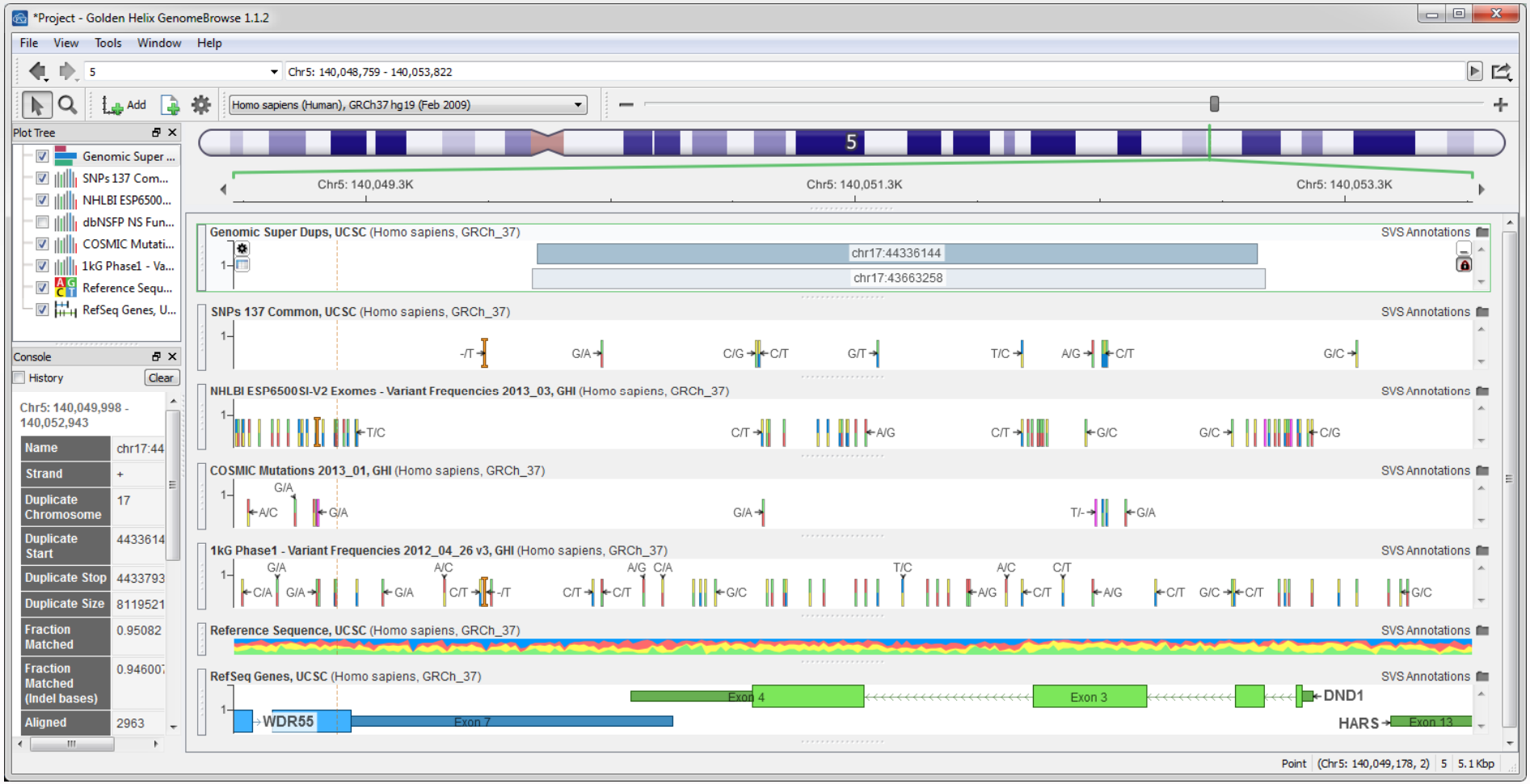
File Edit Select Quality Assurance Analysis Plot Scripts Help

All: 85 x 30
Active: 85 x 30

Unsort		G 1	I 2	R 3	G 4	I 5	R 6
Map	Variant	NA12878_GT	NA12878_DP	NA12878_GQ	NA12891_GT	NA12891_DP	NA12891_GQ
1	1:1423281-SNV	A_r	64	99	r_r	?	
2	1:1423286-SNV	A_r	84	99	A_r	118	9
3	1:65146995-SNV	A_r	123	99	A_r	110	9
4	1:65157120-SNV	A_r	84	99	r_r	?	
5	1:70460304-SNV	A_r	112	99	r_r	?	
6	1:70504932-SNV	A_r	87	99	A_r	100	89.6200027465
7	1:145293510-SNV	A_r	250	99	r_r	?	
8	1:145365316-SNV	A_r	191	99	A_r	213	9
9	1:145365372-SNV	A_r	155	99	A_r	174	9
10	1:208252715-SNV	A_r	18	99	A_r	17	9
11	1:208272311-SNV	A_r	20	69.120002746582	r_r	?	
12	2:28824803-SNV	A_r	45	99	A_r	106	9
13	2:28852004-SNV	A_r	38	70.870002746582	r_r	?	

Master Annotation Report - Compound Het Variants

Additional Annotations



GOLDEN HELIX
Accelerating the Quest for Significance™

Example in GenomeBrowse

WORKFLOW EXAMPLE – TRIO ANALYSIS

Discussion

WORKFLOW EXAMPLE – TRIO ANALYSIS

Clinicians vs. Researchers

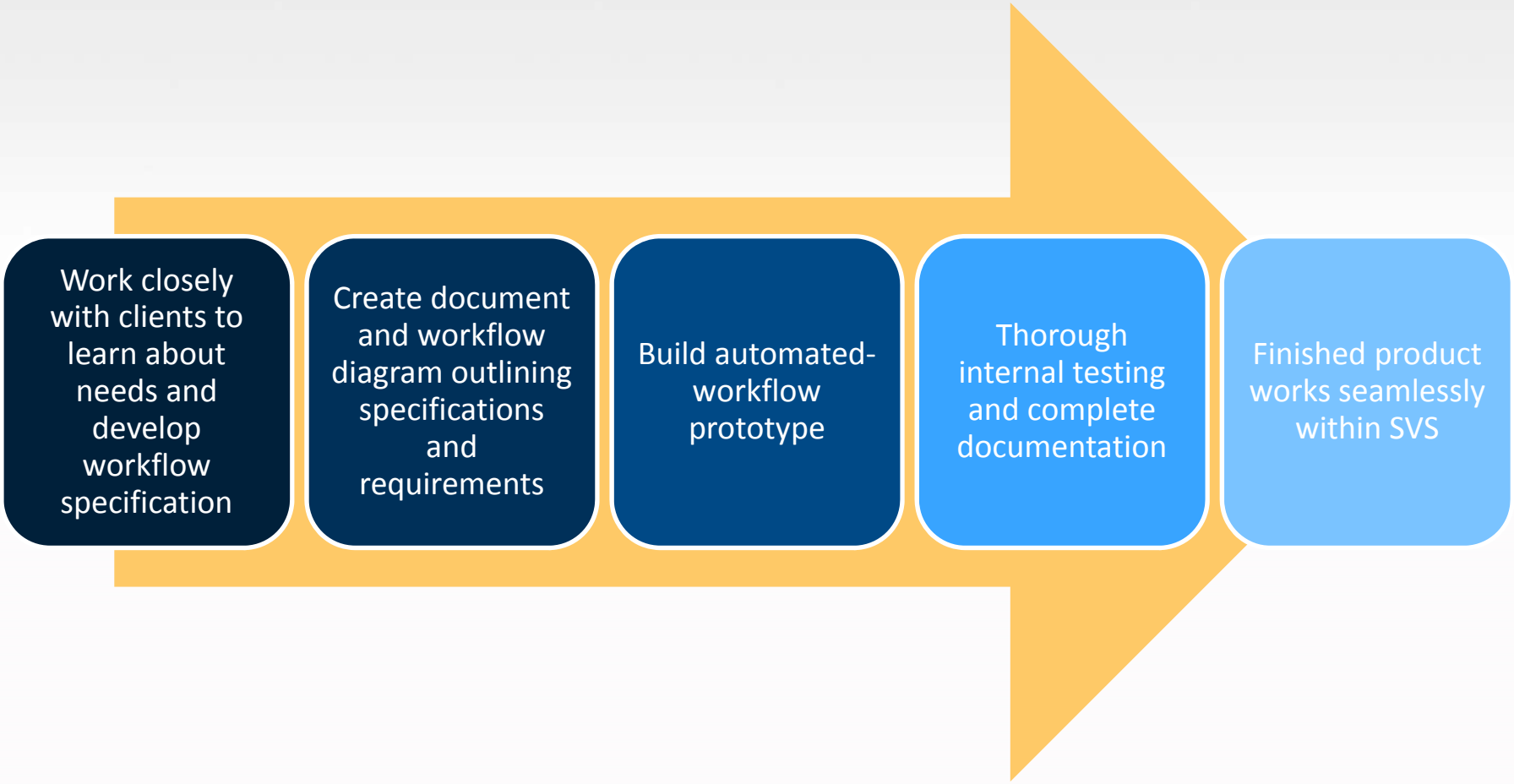
Clinicians

- Running well-defined workflow on additional samples
- Minimum user-interface knowledge
- Small learning curve
- Limited hands-on time

Researchers

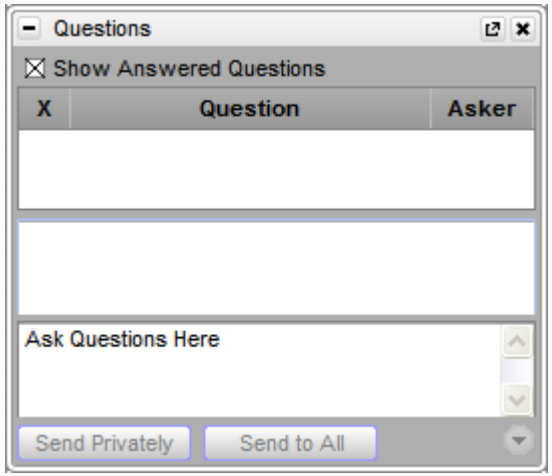
- Building and testing workflows
- More complex but intuitive interface
- Larger learning curve
- Power to investigate and manipulate data

Conclusion – The GHI Approach



Questions during the presentation

Use the Questions pane in your GoToWebinar window



GOLDEN HELIX
Accelerating the Quest for Significance™