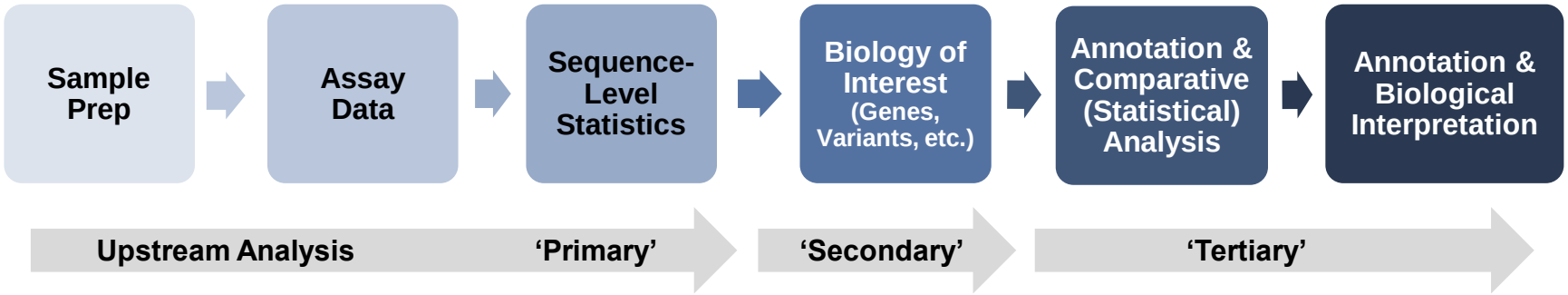
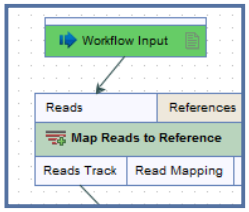
Finding Causal Variants Using Ingenuity Variant Analysis (IVA)

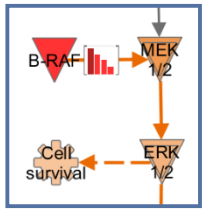
Dev Mistry, Ph.D.
Field Applications Scientist
Devendra.Mistry@qiagen.com

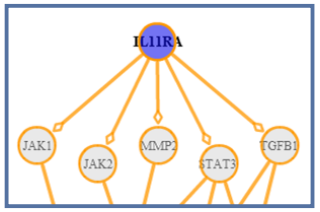
QIAGEN Sample to Insight

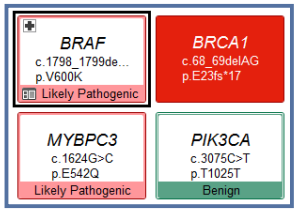


Sample









Insight

QIAGEN

CLC bio A QIAGEN Company

Gx Bx

INGENUITY PATHWAY ANALYSIS

INGENUITY VARIANT ANALYSIS

QIAGEN Clinical Insight

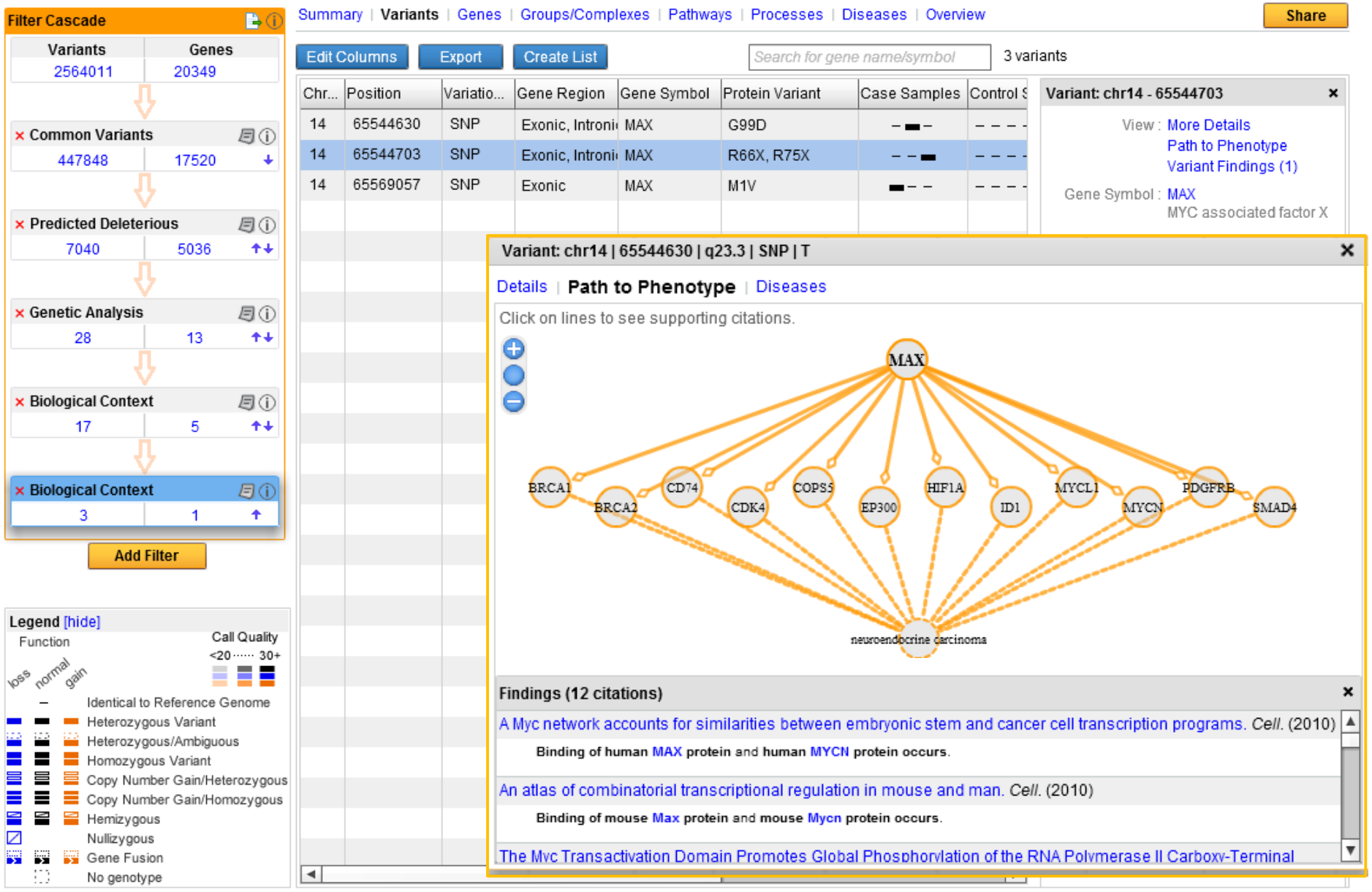
BIOBASE BIOLOGICAL DATABASES

Training Overview

- Introduction
- Data upload and sharing process
- Analysis filter setup
- Results and biological interpretation
- Exporting results
- Summary

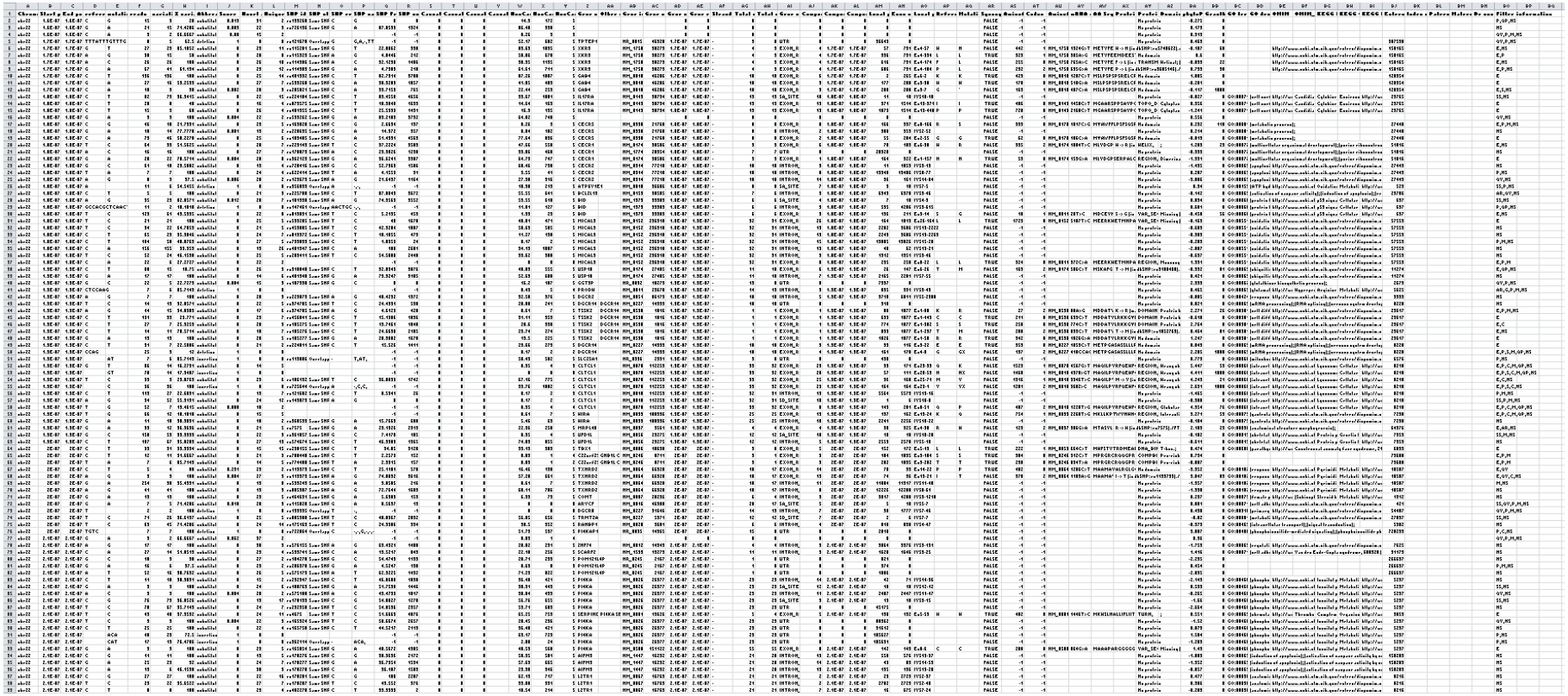
What is Ingenuity Variant Analysis and why use it?

Efficient Cascade to Biologically Meaningful Variants



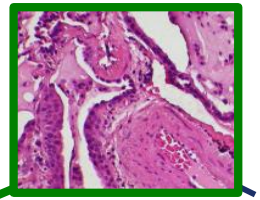
Analysis of Variants the Excel Way

Simply not scalable…



Comprehensive, interactive and visual tools are required to efficiently and correctly interpret an experiment

Ingenuity Variant Analysis Workflow



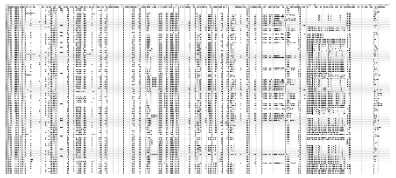
RNA

DNA

Sequencing

Bx

VCF file



Data Analysis

RNA-seq

Variants

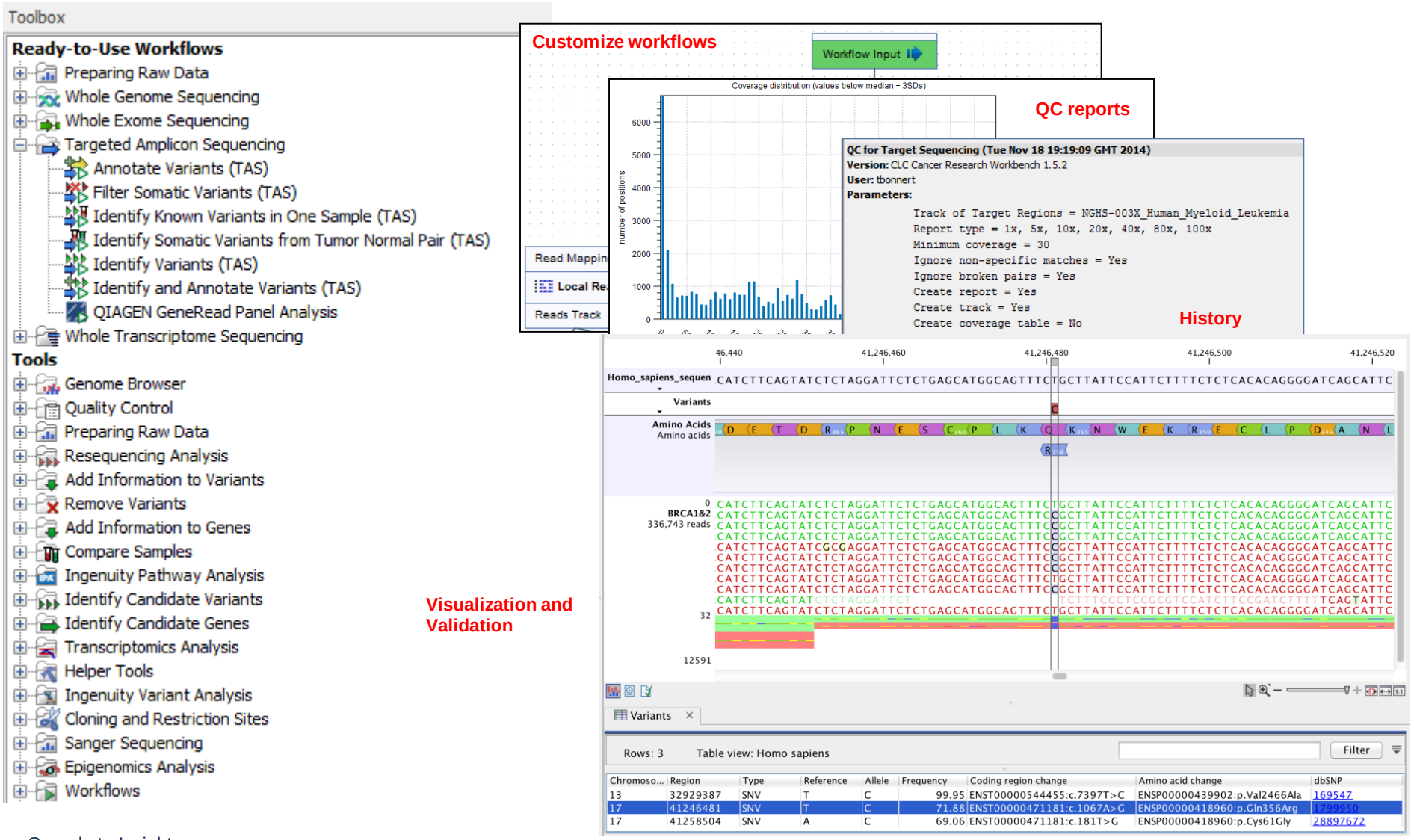
Interpretation

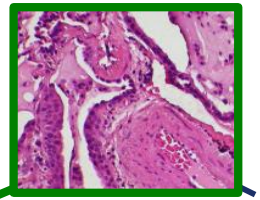
INGENUITY PATHWAY ANALYSIS

INGENUITY VARIANT ANALYSIS

CLC Workbench – Complex Tasks, Simply Done

Streamlined workflows and a rich toolbox to efficiently process data





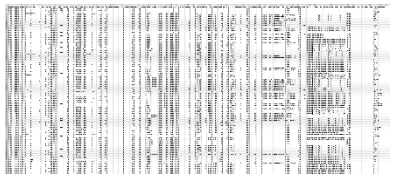
RNA

DNA

Sequencing

Bx

VCF file



Data Analysis

RNA-seq

Variants

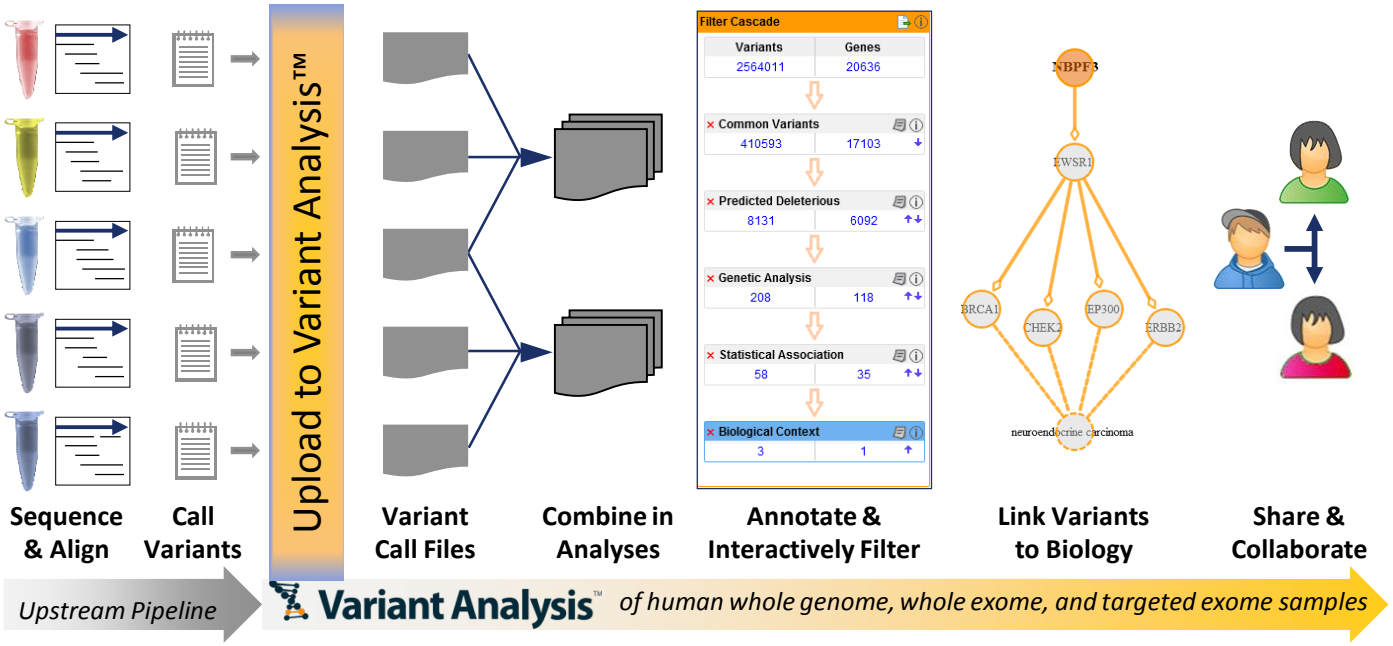
Interpretation

INGENUITY PATHWAY ANALYSIS

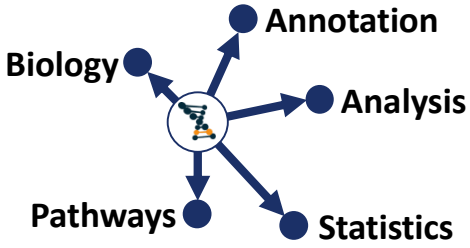
INGENUITY VARIANT ANALYSIS

Identify causal variants from human sequencing data

Ingenuity Variant Analysis



- Stratification Studies ✓
- Large Cancer Studies ✓
- Genetic Disease Cohort ✓
- Trio/Quad Study ✓
- Tumour-Normal Pair ✓
- Personal Genome ✓



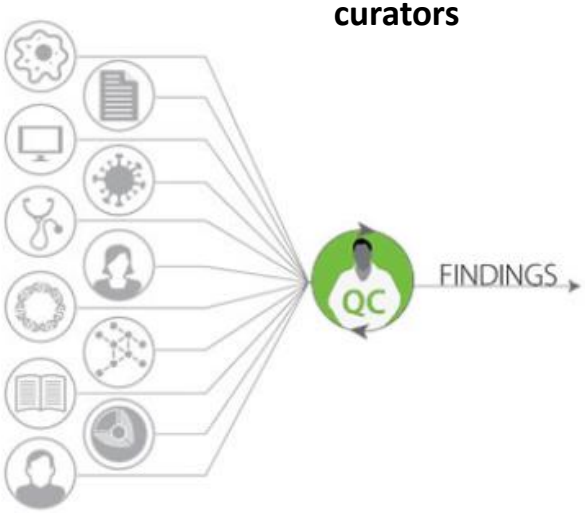
- Pedigree Support
- Disease Identification
- Statistical Burden Testing

A knowledgebase (KB) that's 15 years in the making

Unprecedented Access to Literature Knowledge

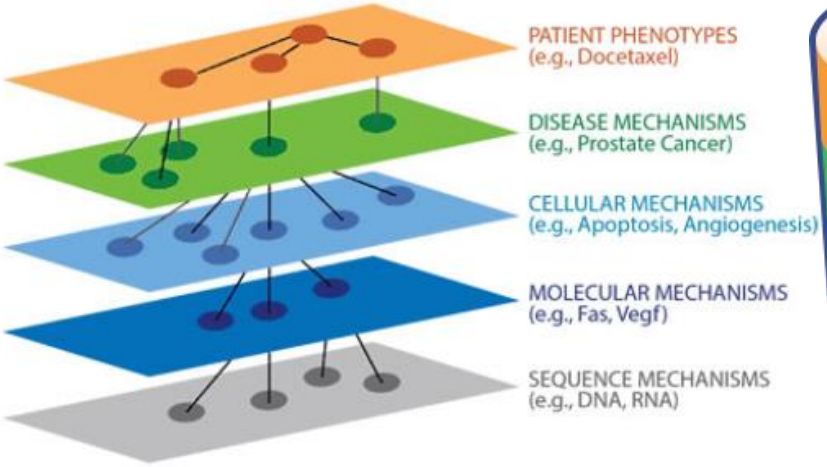
Literature findings

MD/PhD level curators



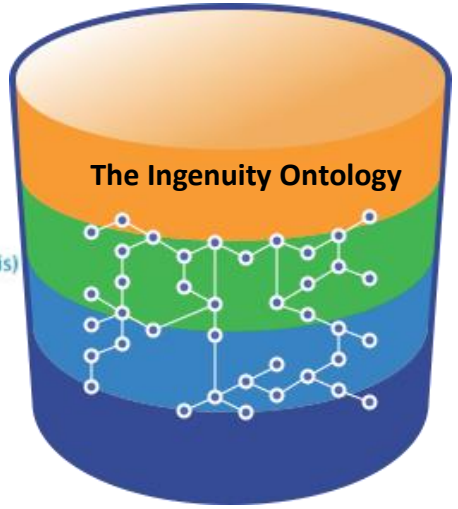
Content Acquisition

Biomedical Ontology



Ingenuity Ontology

The Ingenuity Knowledge Base



Powered By INGENUITY® SYSTEMS

Variant Analysis Content

Quality, Context, Coverage, and Timeliness of Content (ca.1/2014)

COSMIC
Catalogue of somatic mutations in cancer

1000 Genomes
A Deep Catalog of Human Genetic Variation

Mouse Genome Database

The Cancer Genome Atlas
Understanding genomics to improve cancer care

OMIM
Online Mendelian Inheritance in Man
Johns Hopkins University

2014

VISTA

NHLBI Exome Sequencing Project (ESP)
Exome Variant Server

Sorting Intolerant From Tolerant

PolyPhen-2 prediction of functional effects of human nsSNPs

HGMD®
The Human Gene Mutation Database
Cardiff

Additionally

- 349,748+ Ph.D./M.D. expert-curated human phenotype-associated mutation findings
- ~3M+ manual literature findings
- 21,458+ curated disease models
- 185,310+ curated pharmacogenetic (PGx) findings

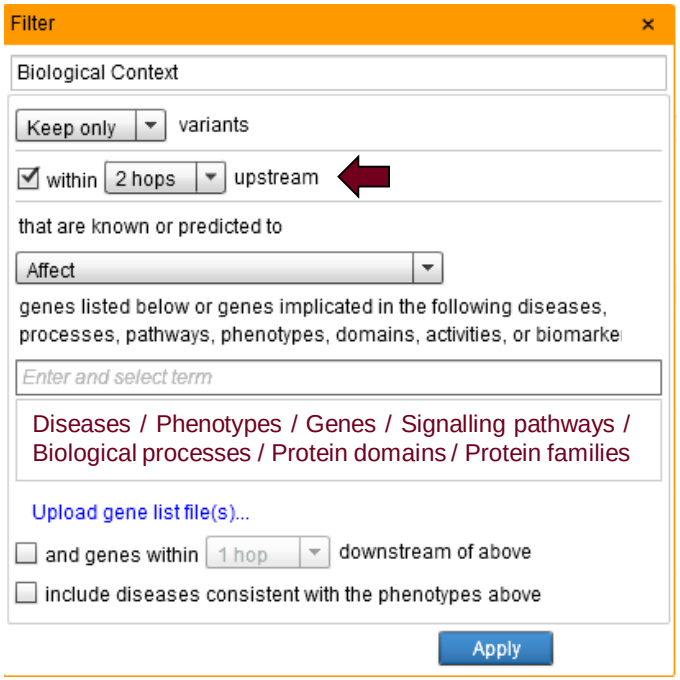
Allele Frequency Community Data

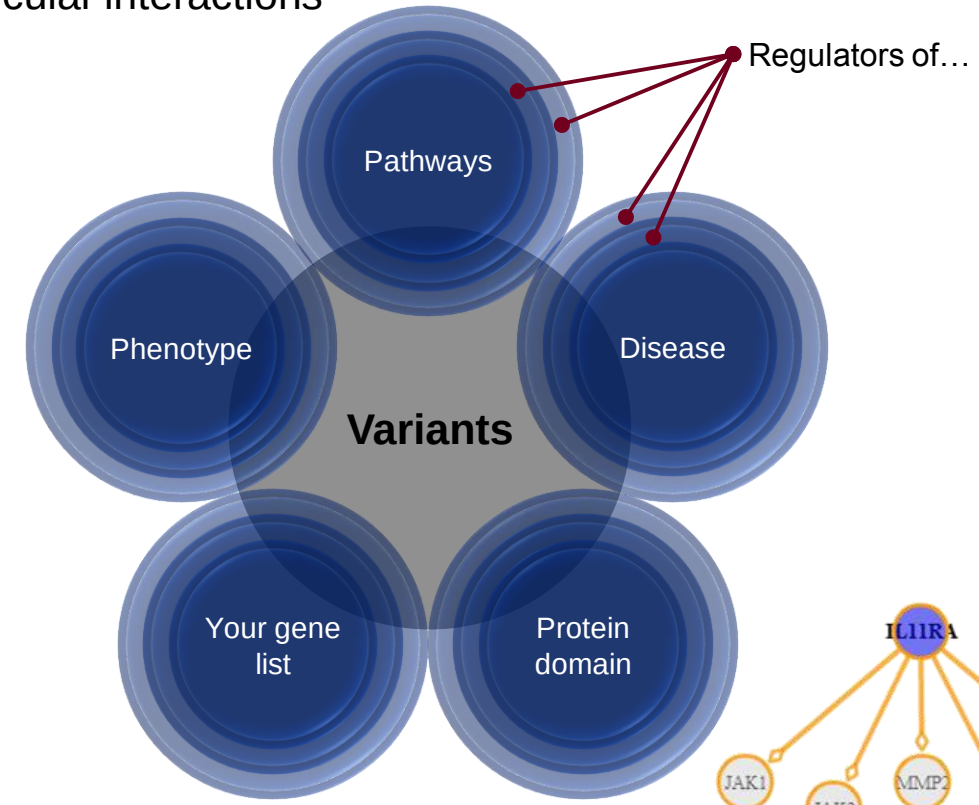
www.allelefrequencycommunity.org

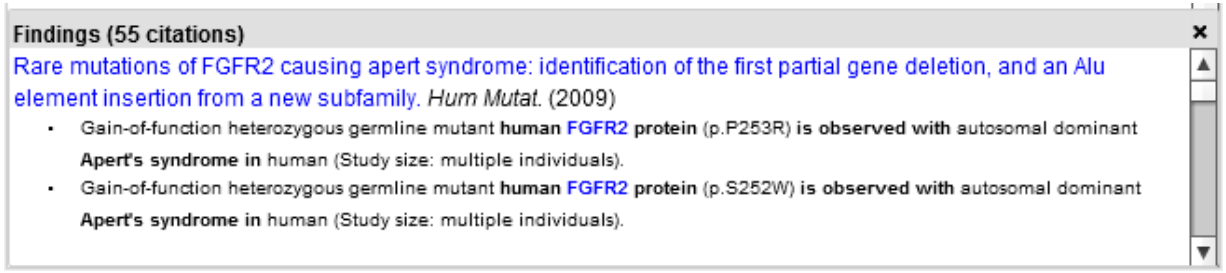
- Leverage the world's largest pool of anonymized allele frequency data
- **Reduces false positives** in analyses by removing variants that are commonly seen in the general population
- Contains Whole Exome AND Whole Genome data
- Better representation of Insertions and Deletions
- **Larger than ANY other public resource**
  - AFC launched with 70,000 samples with >8,000 as whole genomes
  - 12x larger than Exome Variant Server data
- The initial launch version of the database already provides a 43% average false positive rate reduction in a benchmarking set of whole-genome Diagnostic Odyssey cases
- AFC will grow as more people opt-in
  - Launched on 25th February 2015 with 70,000 Samples including 8,000 Whole Genomes
  - Currently at over 100,000 samples, including over 14,000 Whole Genome samples

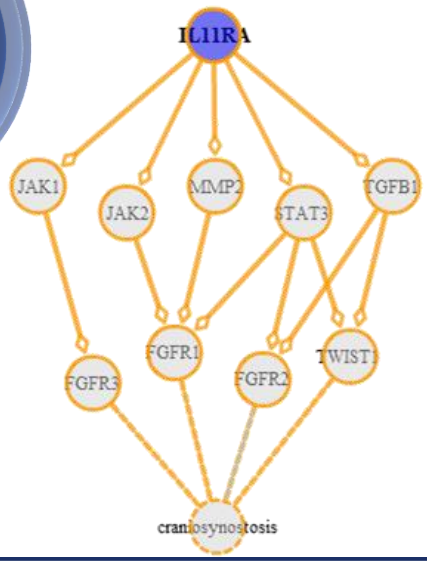
Refine and Dynamically Filter Variants

Use biological associations and molecular interactions









Case Study: Pheochromocytoma (PCC)

Hereditary Pheochromocytoma

Hereditary pheochromocytoma (PCC) is a neuroendocrine tumor of the medulla of the adrenal glands

Whole exome sequencing (Agilent SureSelect) on Illumina Genome Analyzer II

Sequence data obtained from European Nucleotide Archive (ENA)

- http://www.ebi.ac.uk/ena/data/view/ERR031607-ERR031626

Published in Nature Genetics (2011); PMID: 2168591

- Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma

Barcode	Display Name	State	Subject	Variants
ERS025965	924_Hereditary_Pheochromocytoma	case	924	231460
ERS025967	3037_Hereditary_Pheochromocytoma	case	3037	283434
ERS025973	3121_Hereditary_Pheochromocytoma	case	3121	243291
ERS025971	NA11881_Control	control	NA11881	731542
ERS025969	NA12144_Control	control	NA12144	304008
ERS025974	NA12892_Control	control	NA12892	681906
ERS025970	NA12813_Control	control	NA12813	548289
ERS025966	NA12763_Control	control	NA12763	1098602
ERS025968	NA12761_Control	control	NA12761	834283
ERS025972	NA12750_Control	control	NA12750	488217

Independent hereditary pheochromocytoma

Independent HapMap samples

LETTERS

nature genetics

Exome sequencing identifies *MAX* mutations as a cause of hereditary pheochromocytoma

Iñaki Comino-Méndez[1,2,15], Francisco J Gracia-Aznárez[2,3,15], Francesca Schiavi[4,15], Iñigo Landa[1], Luis J Leandro-García[1], Rocío Letón[1], Emiliano Honrado[5], Rocío Ramos-Medina[6], Daniela Caronia[7], Guillermo Pita[7], Álvaro Gómez-Graña[1], Aguirre A de Cubas[1], Lucía Inglada-Pérez[1,2], Agnieszka Maliszewska[1], Elisa Taschin[4], Sara Bobisse[4], Giuseppe Pica[8], Paola Loli[9], Rafael Hernández-Lavado[10], José A Díaz[11], Mercedes Gómez-Morales[12], Anna González-Neira[7], Giovanna Roncador[6], Cristina Rodríguez-Antona[1,2], Javier Benítez[2,3], Massimo Mannelli[13], Giuseppe Opocher[4,14], Mercedes Robledo[1,2] & Alberto Cascón[1,2]

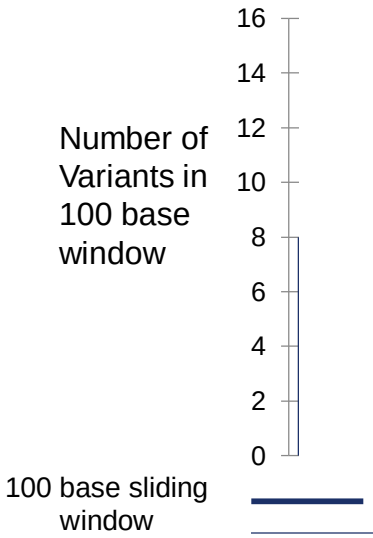
Examples

- Tumor vs Normal
- Trio (Hereditary)
- Stratification (affected vs unaffected)

Live Demo

Keep outside most top *x*% most exonically varible 100 base window

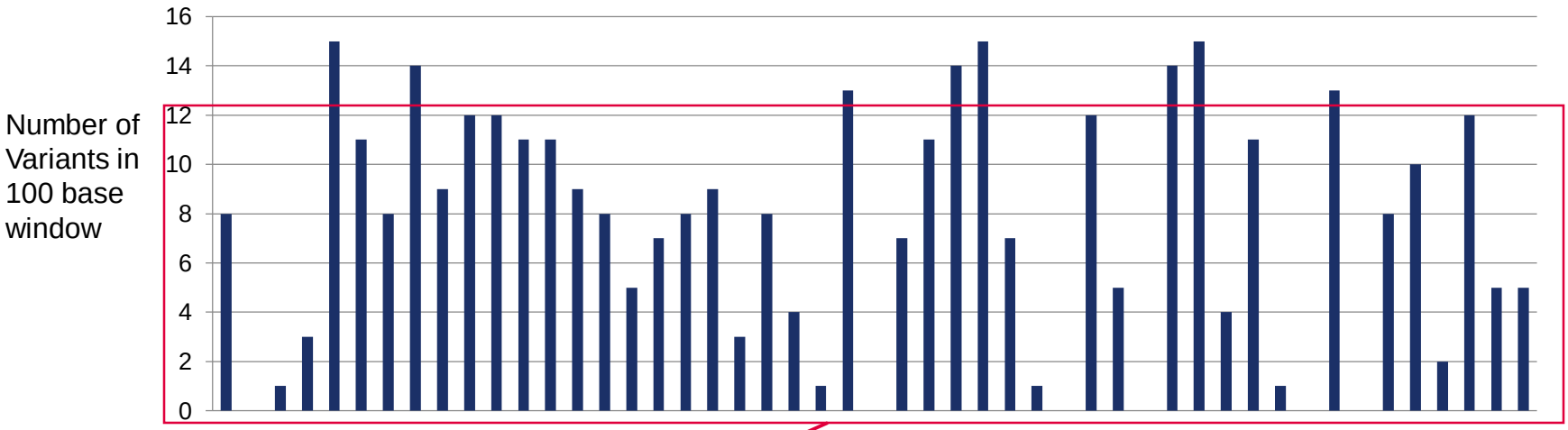
SLIDE EXPLAINING NOISE FILTER



1000 Genomes

Keep outside most top *x*% most exonically variable 100 base window

SLIDE EXPLAINING NOISE FILTER



Will ignore top *x*% most exonically variable 100 base regions

Keep outside top *x*% most exonically variable genes

Identification and Removal of Artifacts

Counts based on single sample, qual 20, allele frequency < 3%

	A	B
1	Gene Symbol	Count
2	MUC4	129
3	PTPRN2	98
4	MUC6	95
5	HLA-DRB1	89
6	MUC16	80
7	PRSS1/PRSS3	63
8	CSF2RA	60
9	PDE4DIP	53
10	BAGE3; BAGE4; BAGE2	53
11	PRIM2	48
12	MAP2K3	46
13	KCNJ12; KCNJ18	45
14	C9orf96	44
15	CDK11A/CDK11B	43
16	MUC12	40
17	AQP7	40
18	TUBGCP2	39
19	CDH4	39
20	HLA-DRB5	39
21	SLC22A2	39
22	C7orf50	36
23	ANKRD36B (includes others)	34
24	PRKAG2	32
25	TTC34	31
26	OR8U1	31
27	EPHA8	29
28	FRG2 (includes others)	29
29	SORCS2	29
30	COL5A1	29

I	J	K	L	M	N	O
Variants per Gene	Count of Genes	Pct of Genes in Analysis	Cumulative Pct of Genes		Cumulative Sum of Variants	Cumulative Pct of Variants
1	3795	56.68%	56.68%		3795	23.40%
2	1307	19.52%	76.21%		6409	39.51%
3	602	8.99%	85.20%		8215	50.64%
4	322	4.81%	90.01%		9503	58.58%
5	194	2.90%	92.91%		10473	64.56%
6	114	1.70%	94.61%		11157	68.78%
7	88	1.31%	95.92%		11773	72.58%
8	48	0.72%	96.64%		12157	74.95%
9	32	0.48%	97.12%		12445	76.72%
10	28	0.42%	97.54%		12725	78.45%
11	32	0.48%	98.01%		13077	80.62%
12	16	0.24%	98.25%		13269	81.80%
13	13	0.19%	98.45%		13438	82.84%
14	15	0.22%	98.67%		13648	84.14%
15	13	0.19%	98.86%		13843	85.34%
16	11	0.16%	99.03%		14019	86.43%
17	7	0.10%	99.13%		14138	87.16%
18	5	0.07%	99.21%		14228	87.71%
19	3	0.04%	99.25%		14285	88.06%
20	6	0.09%	99.34%		14405	88.80%
21	1	0.01%	99.36%		14426	88.93%
22	2	0.03%	99.39%		14470	89.21%
23	2	0.03%	99.42%		14516	89.49%
24	7	0.10%	99.52%		14684	90.52%

For example, if set at 5%:

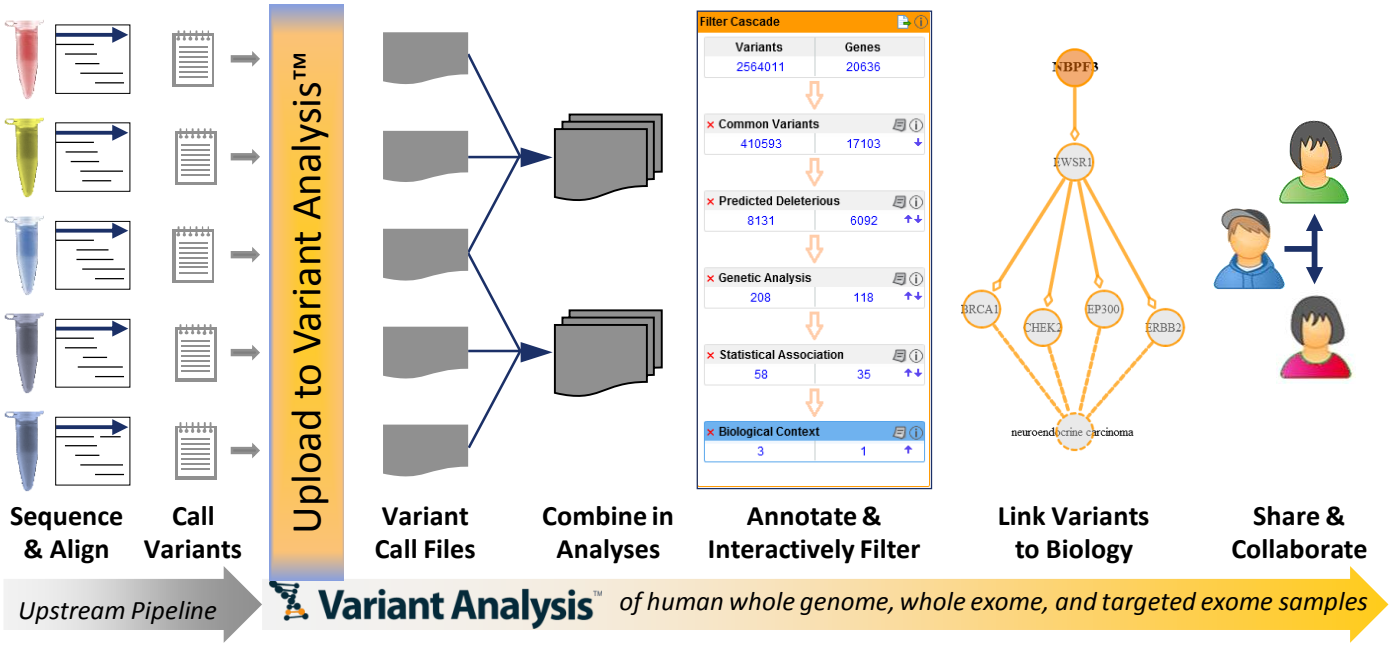
Keep genes that are not in the top 5% most exonically variable

Will ignore the top 5% most exonically variable genes

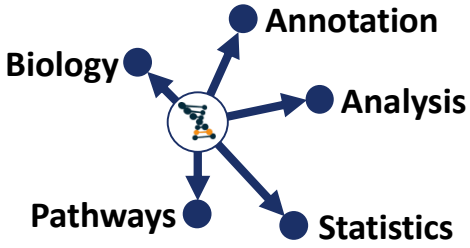
Summary

Identify causal variants from human sequencing data

Ingenuity Variant Analysis

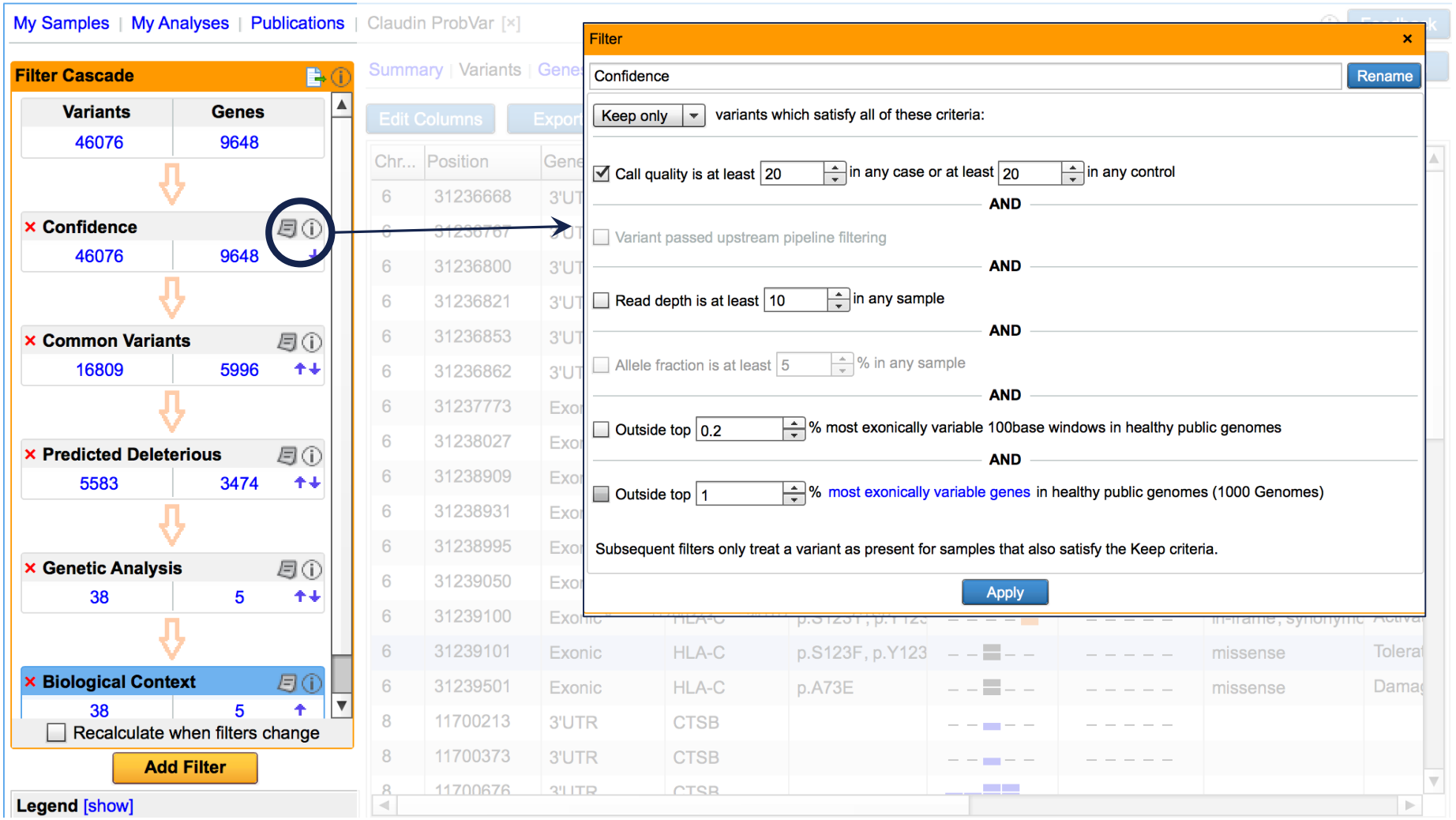


- Stratification Studies ✓
- Large Cancer Studies ✓
- Genetic Disease Cohort ✓
- Trio/Quad Study ✓
- Tumour-Normal Pair ✓
- Personal Genome ✓

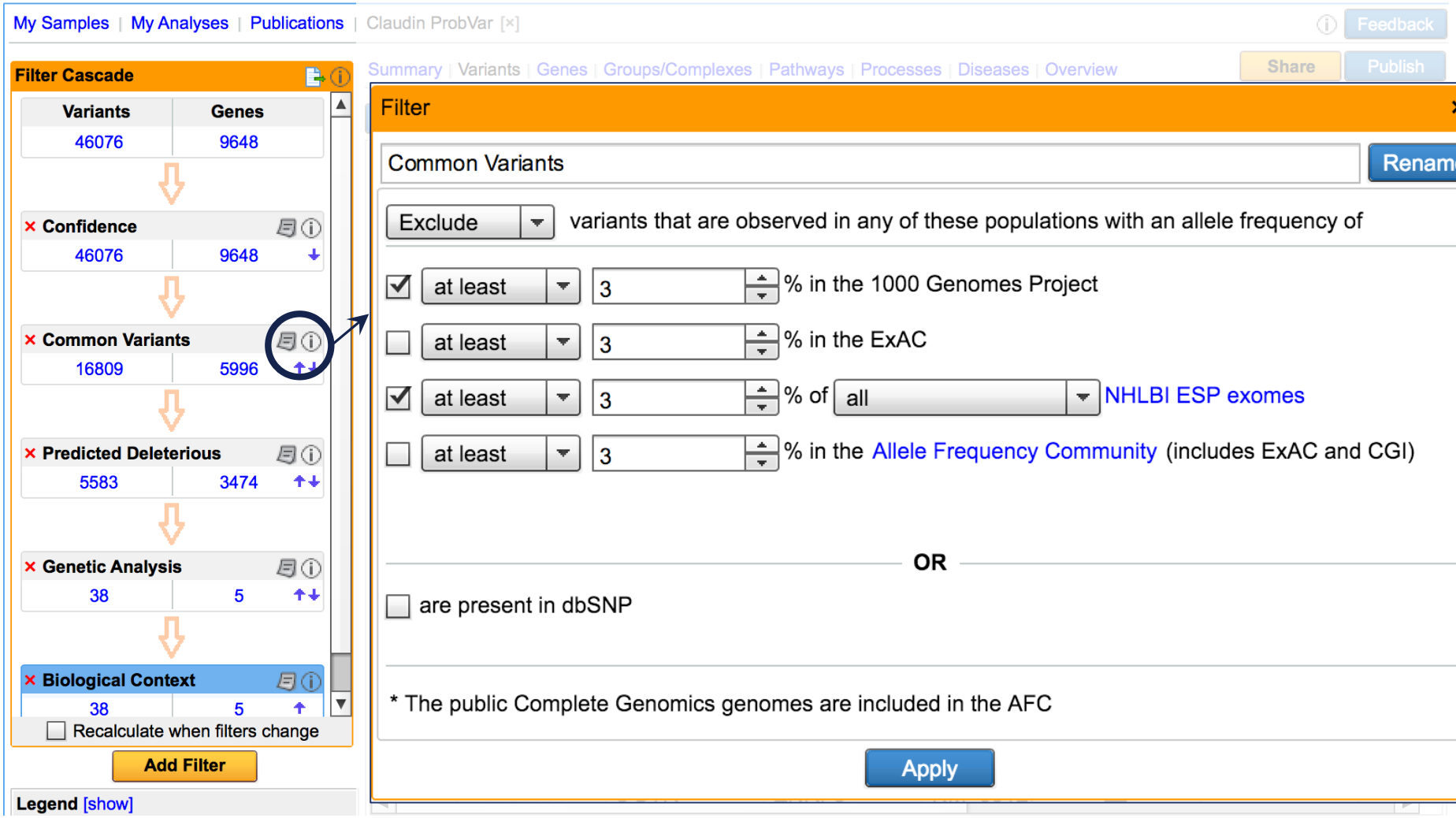


- Pedigree Support
- Disease Identification
- Statistical Burden Testing

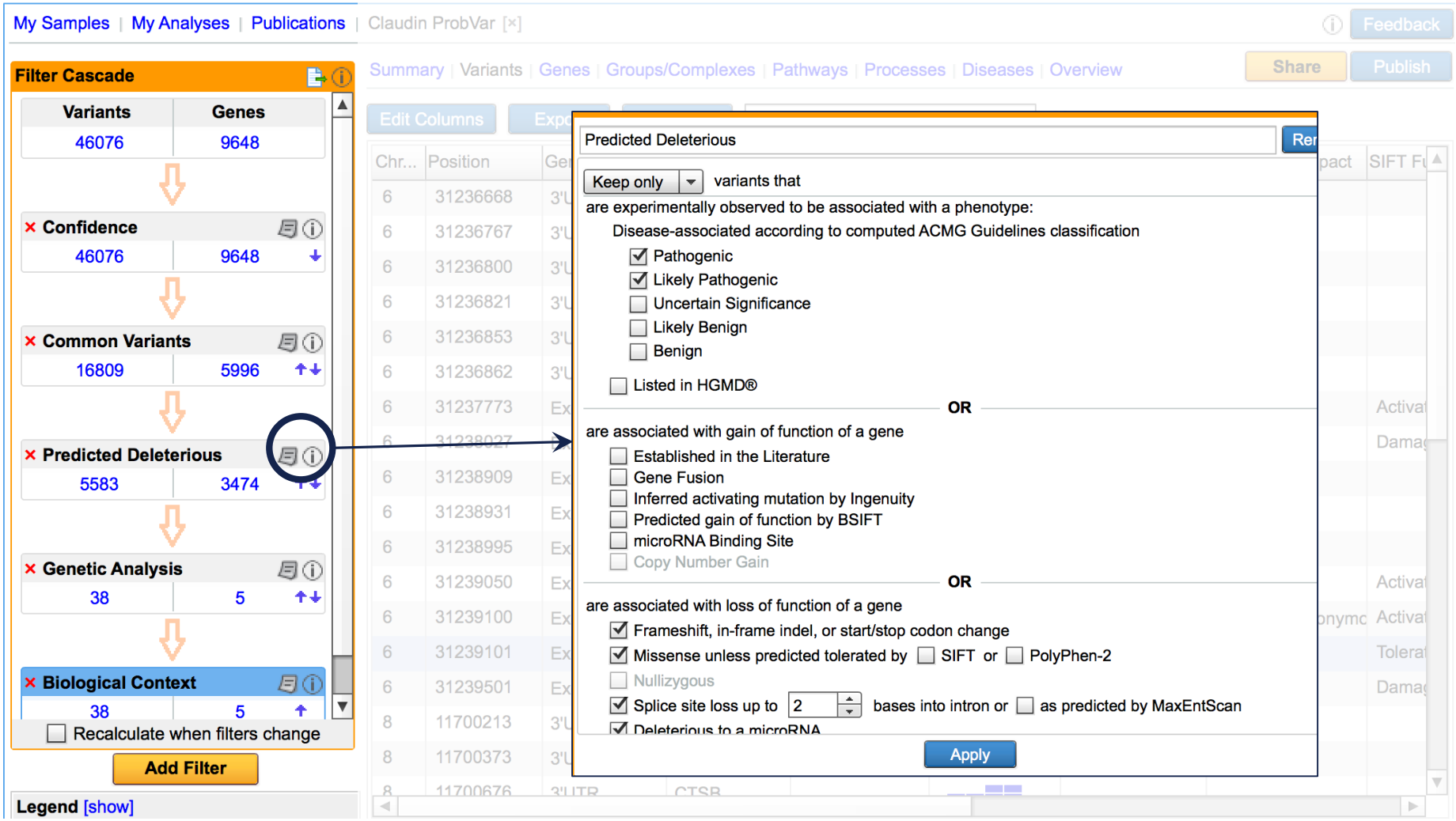
Filter Cascade: Confidence



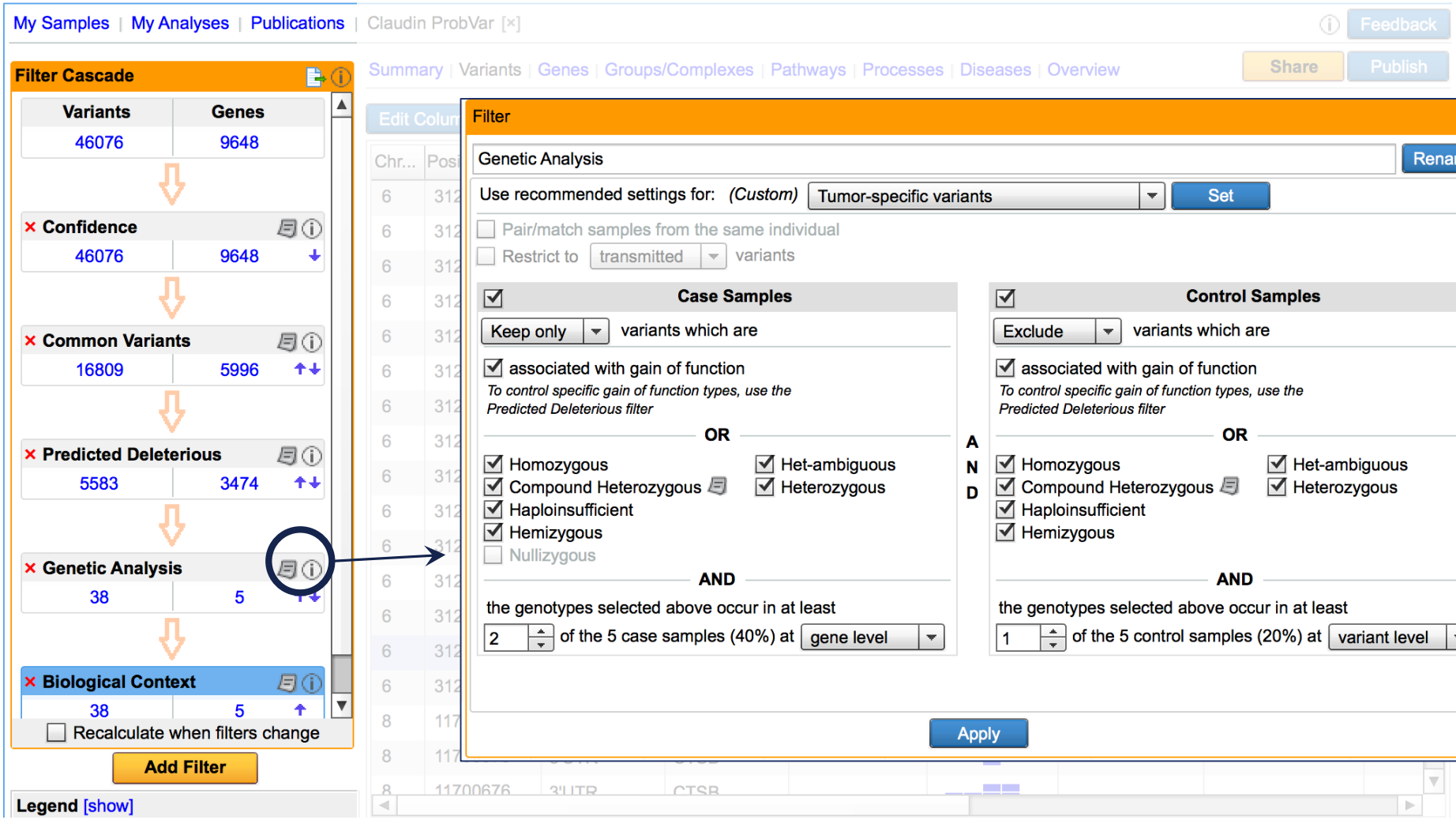
Filter Cascade: Common Variants



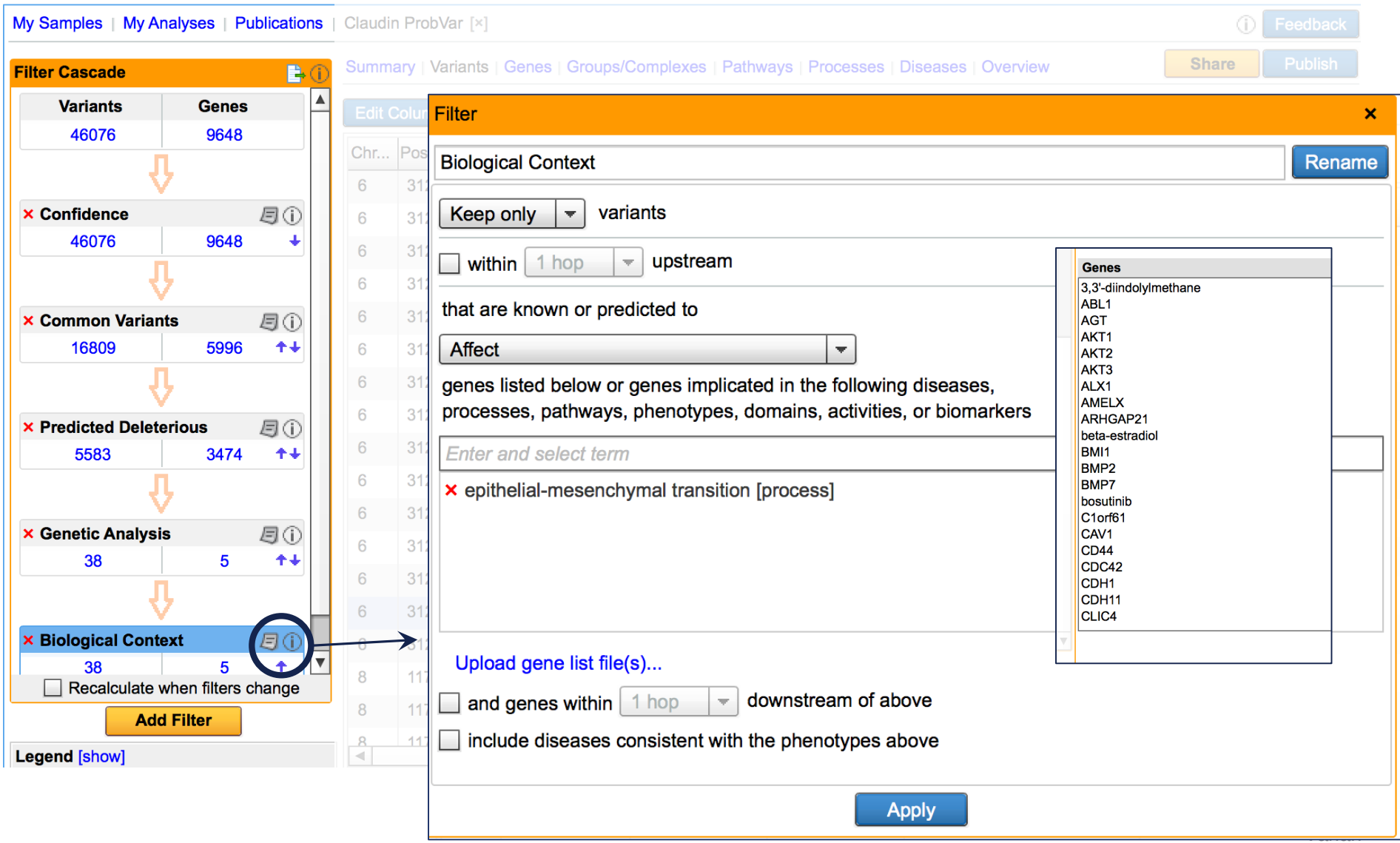
Filter Cascade: Predicted Deleterious



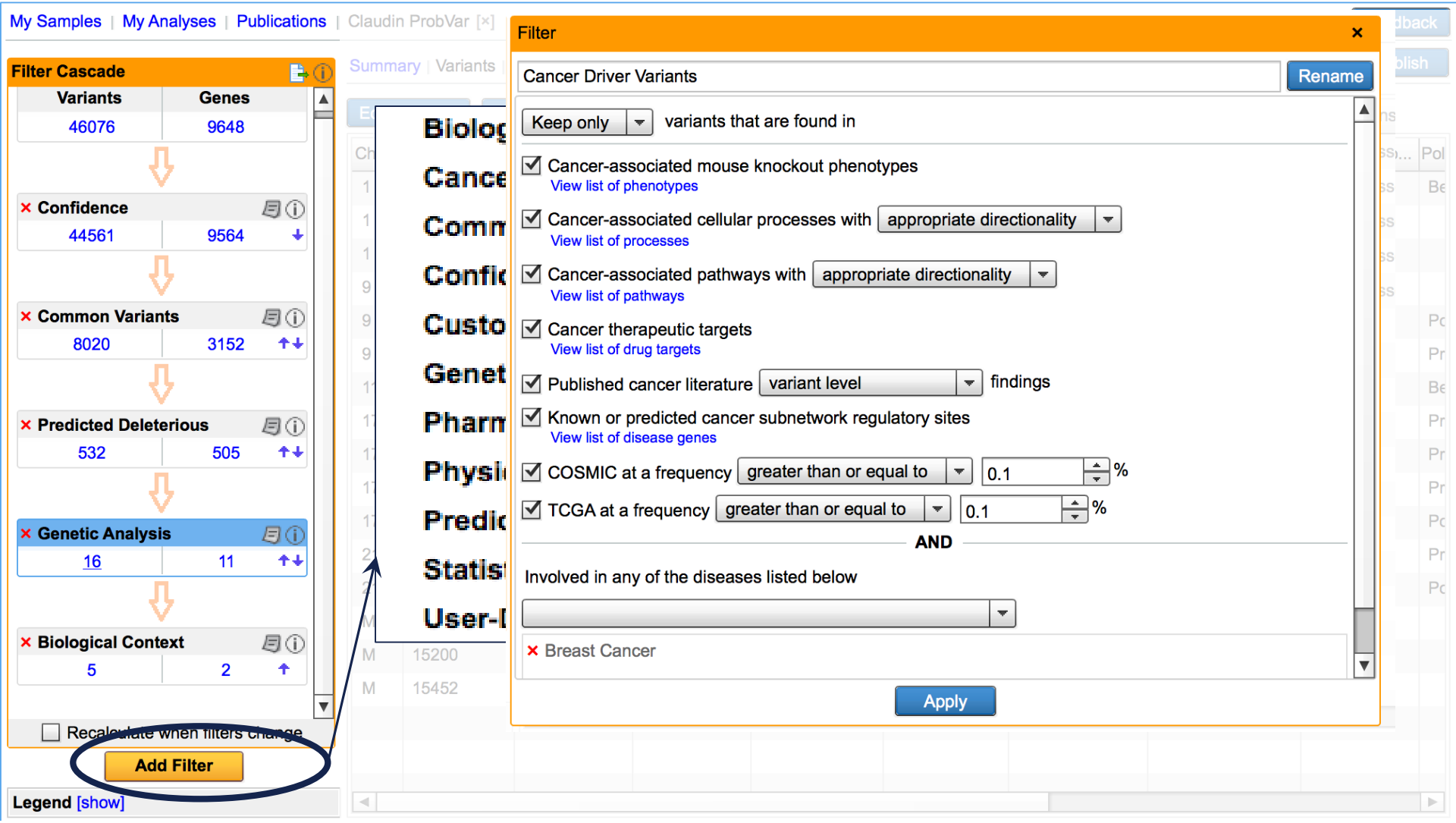
Filter Cascade: Genetic Analysis



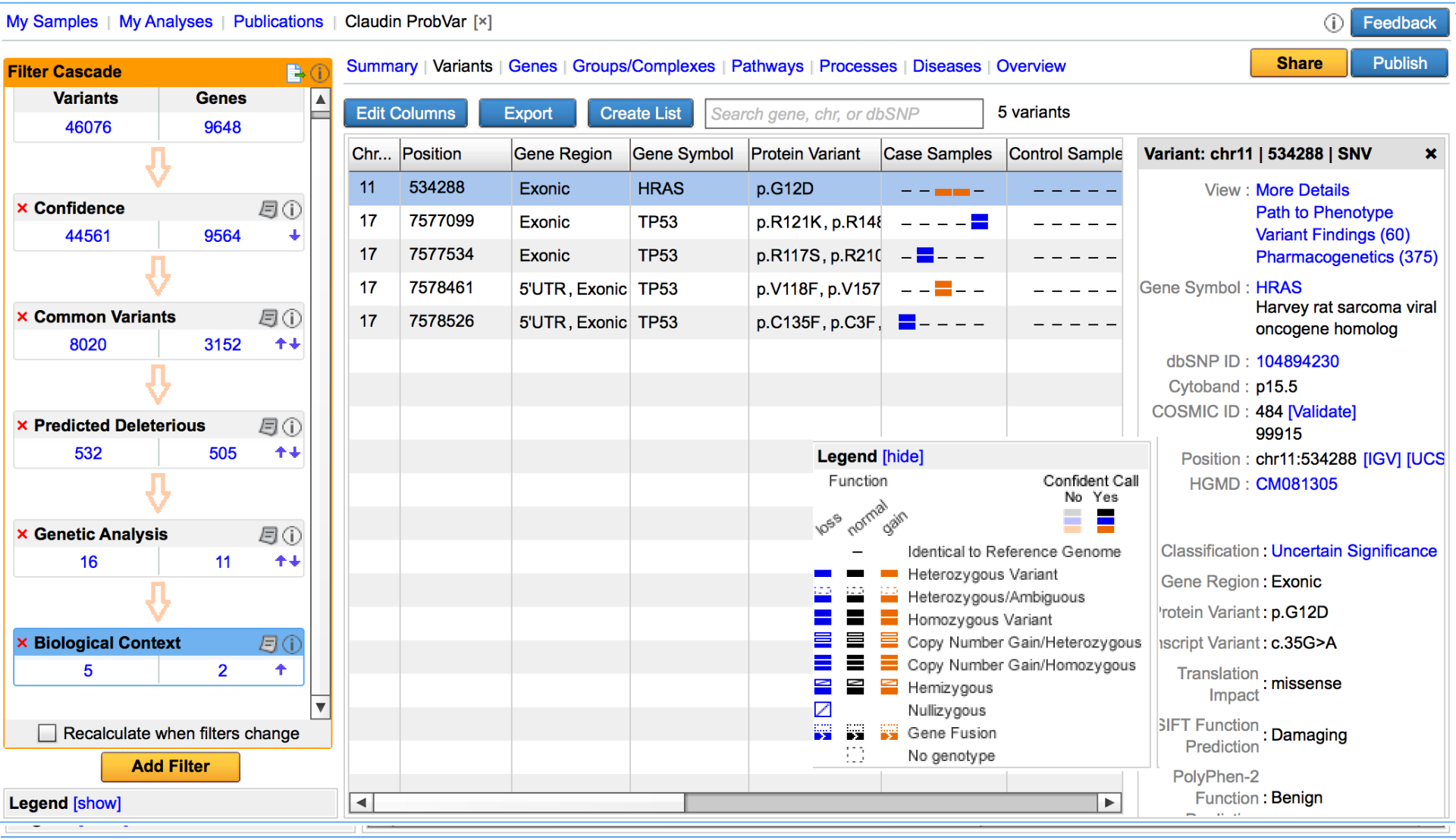
Filter Cascade: Biological Context



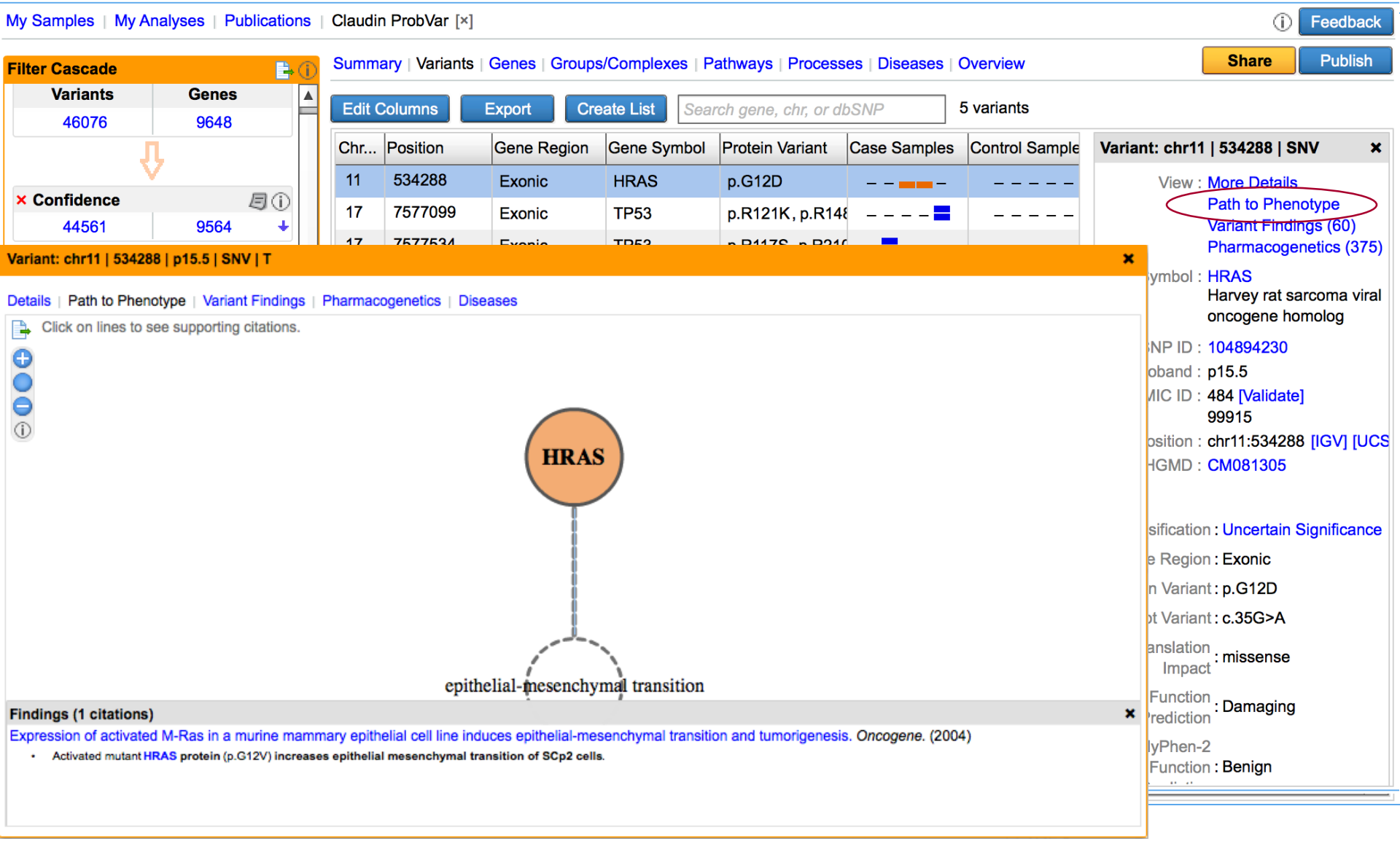
Filter Cascade: Add/Rearrange Filters



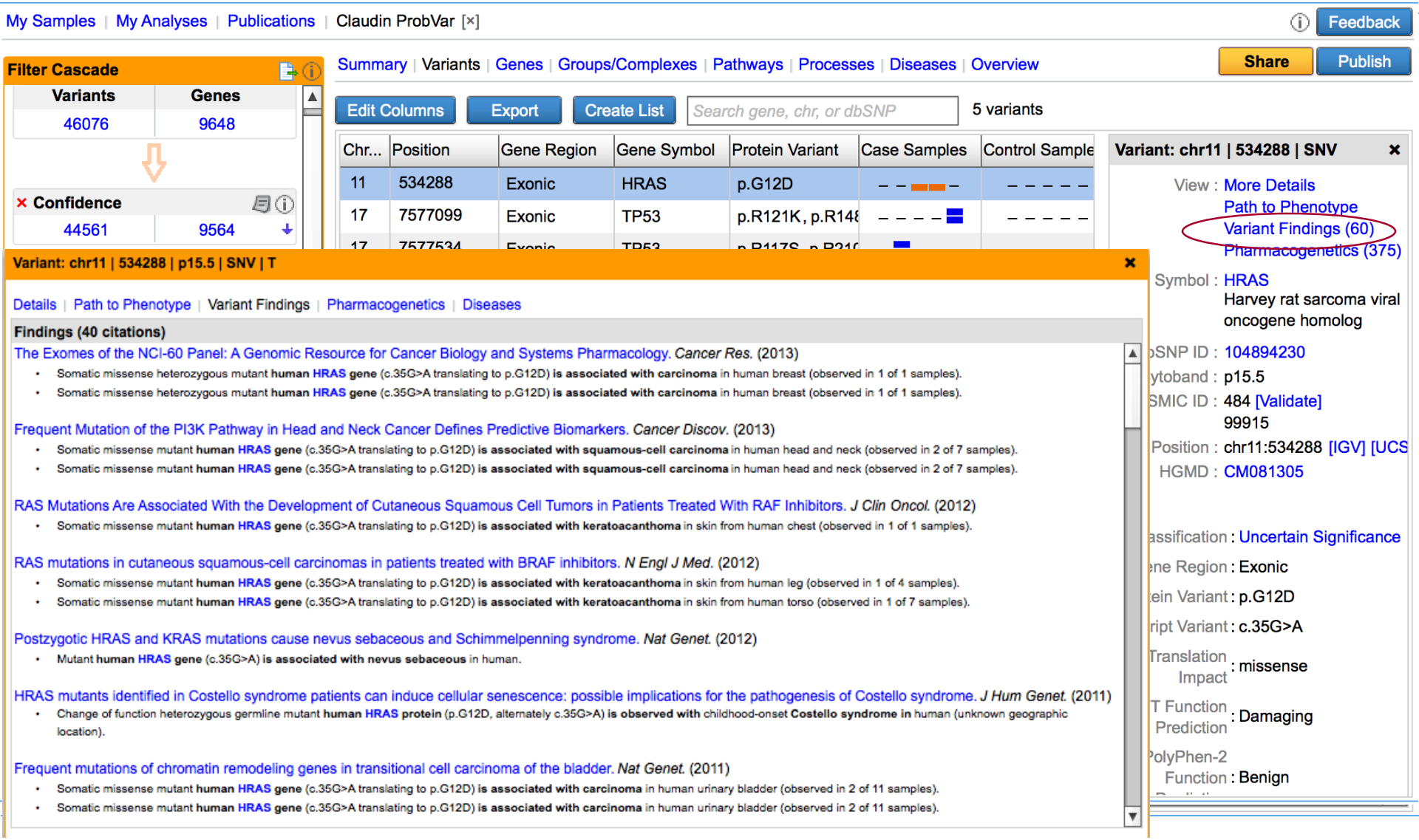
Results: The Short List



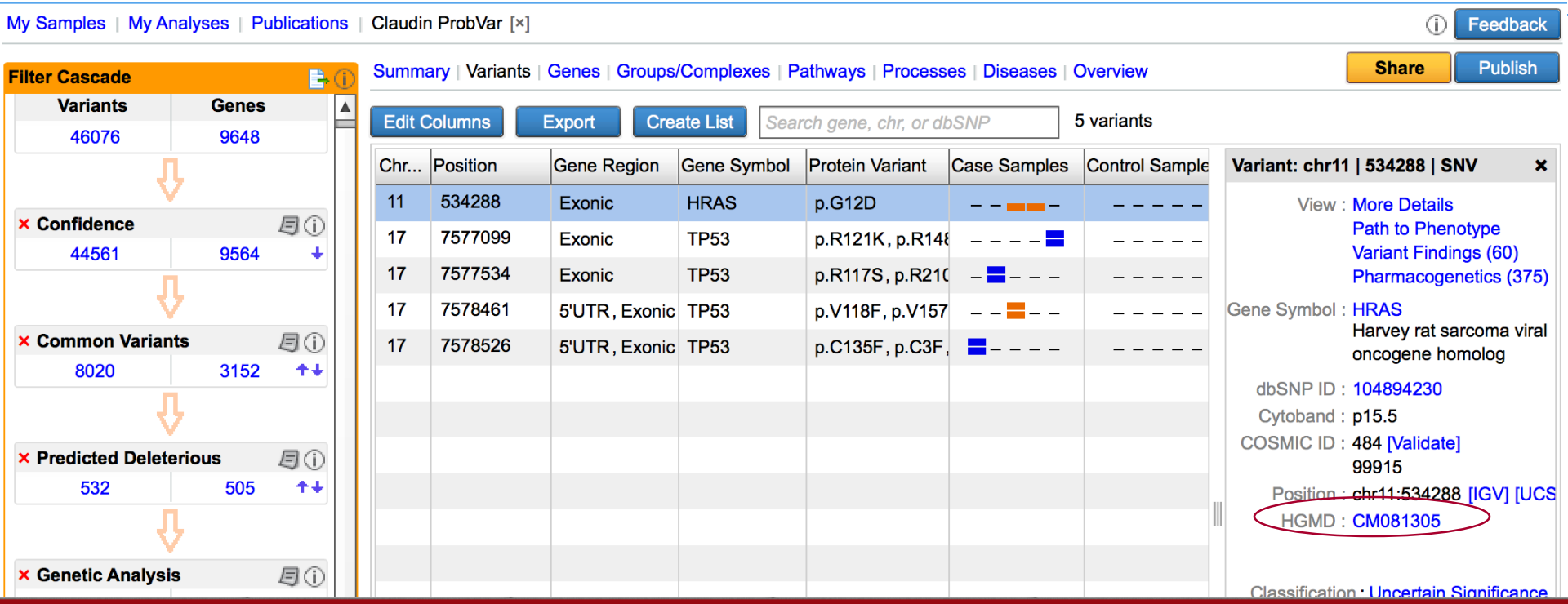
Results: The Short List



Results: The Short List



Results: The Short List



HGMD accession	Disease/phenotype	Gene symbol	Codon change	Amino acid change	Codon number	Feedback
CM081305	Costello syndrome	HRAS	GGC-GAC	Gly-Asp	12	Feedback

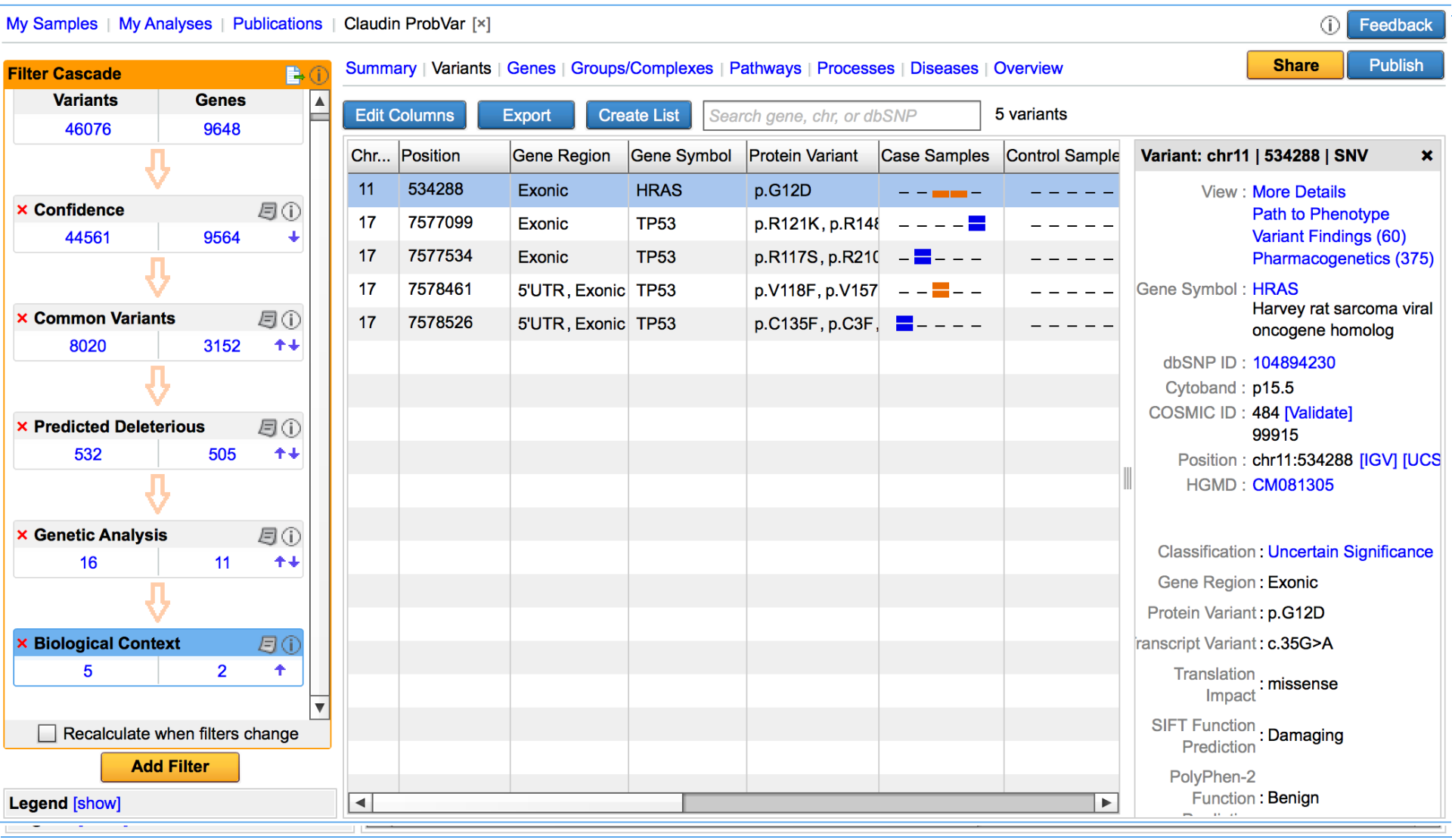
The G12D substitution exhibits a shift in polarity from non-polar to negatively charged and displays a decrease in Kyte-Doolittle hydrophobicity from -0.4 to -3.5. Approximately 1.95% of missense mutations in HGMD are Gly-Asp. The mutation occurs 178 amino acids from the end of the protein.

Literature citation	Citation type	Notes
1. Lo (2008) *J Med Genet* **45:** 167 PubMed: 18039947	Primary literature report	None
2. Niihori (2011) *J Hum Genet* **56:** 707 PubMed: 21850009	Functional characterisation	None
3. Cheng (2012) *PLoS Comput Biol* **8:** e1002738 PubMed: 23093928	Functional characterisation	Systemic impact quantified as a combination of protein stability change and pathway perturbation.

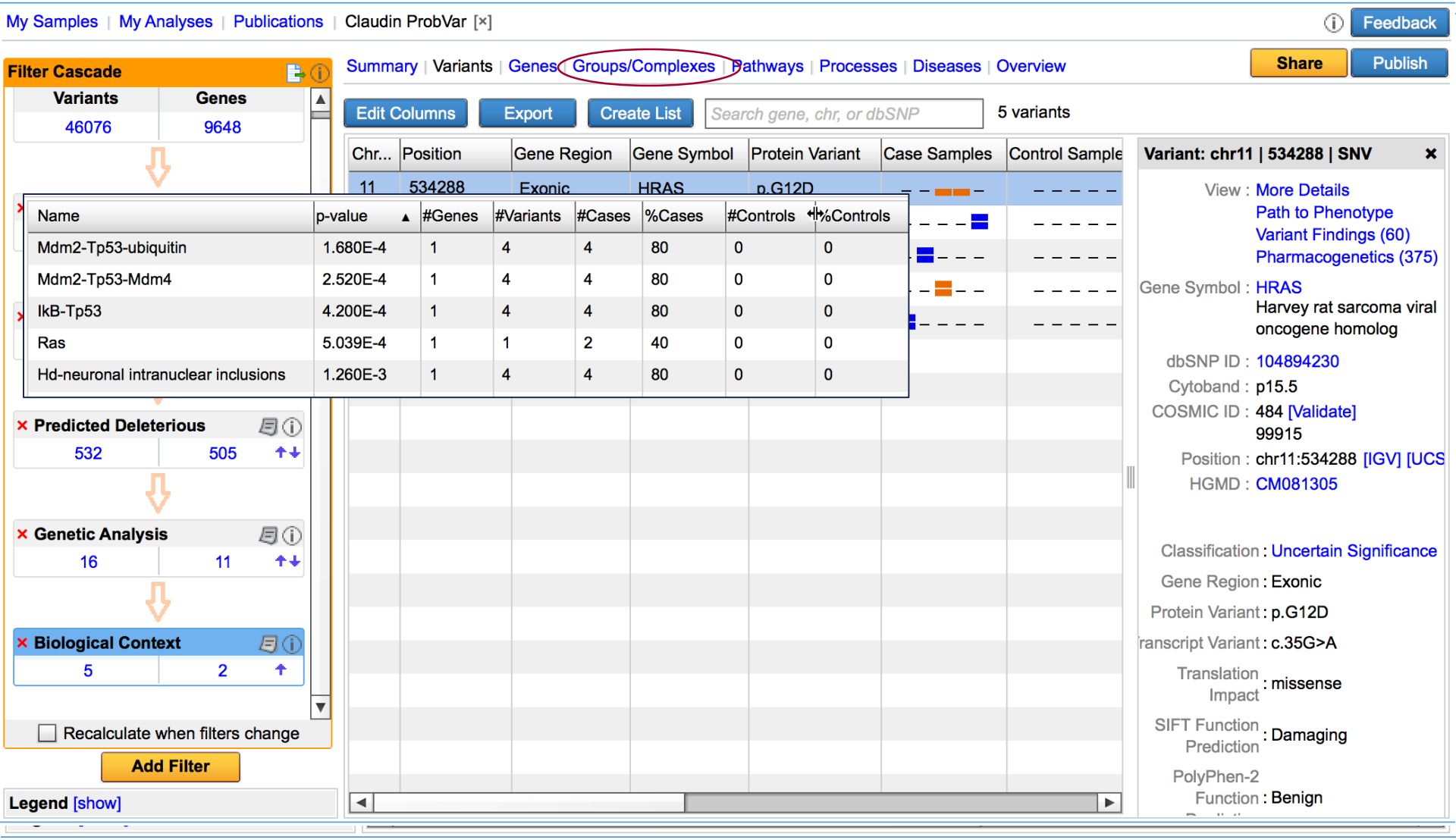
Extra information

Coding strand genomic sequence (GRCh37.3) CGGAATATAAGCTGGTGGTGGTGGGCGCCG(G/A)CGGTGTGGGCAAGAGTGCGCTGACCATCCA

Results: The Short List



Results: The Short List



Results: The Short List

My Samples | My Analyses | Publications | Claudin ProbVar [×]

Feedback

Filter Cascade

Variants	Genes
46076	9648

Confidence: 44561 | 9564

Common Variants: 8020

Predicted Deleteri: 532

Genetic Analysis: 16

Biological Context: 5

Recalculate whe

Add Fi

Legend [show]

Summary | Variants | Genes | Groups/Complexes | Pathways | Processes | Diseases | Overview

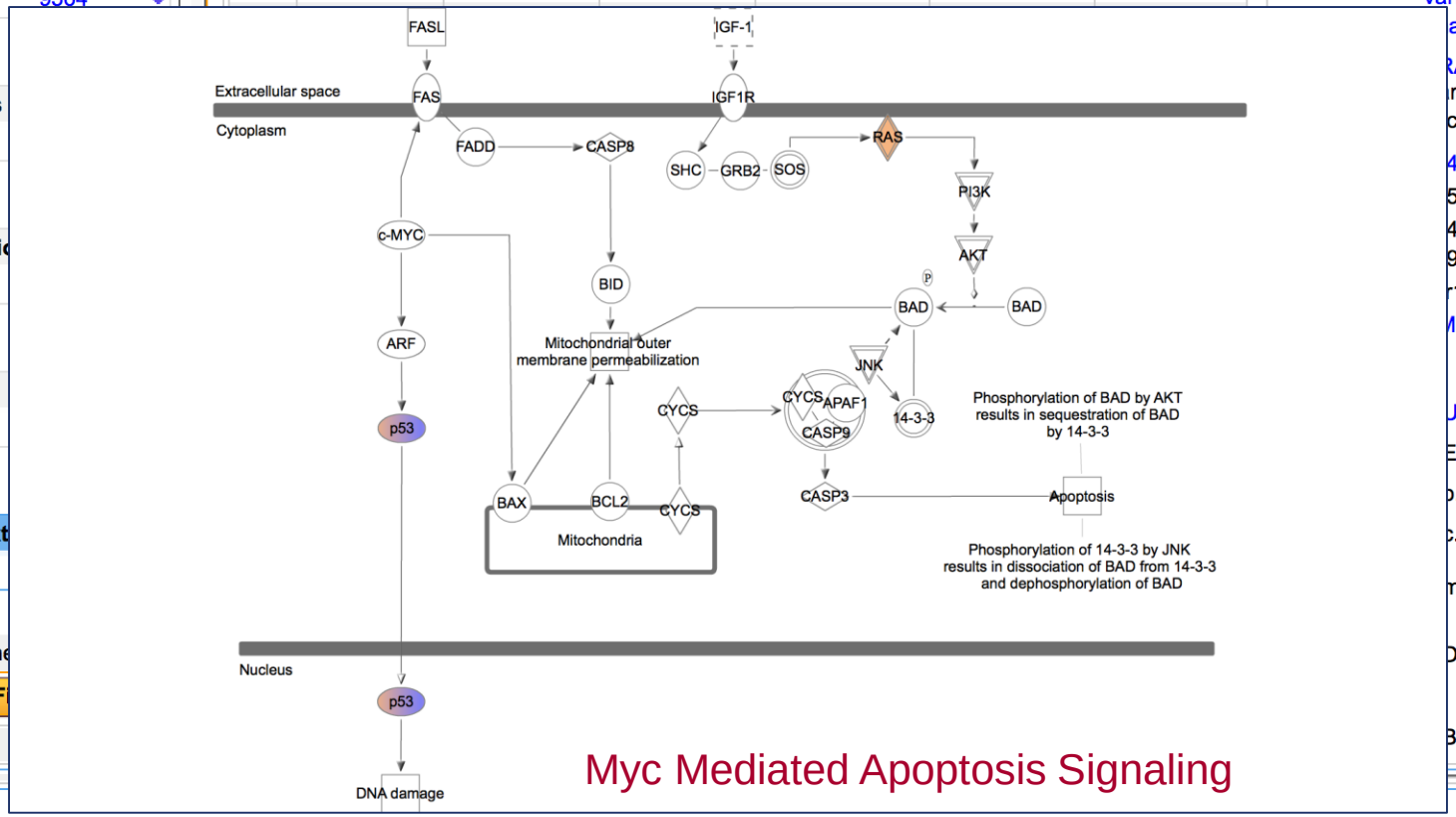
Share | Publish

Edit Columns | Export | Create List | Search gene, chr, or dbSNP | 5 variants

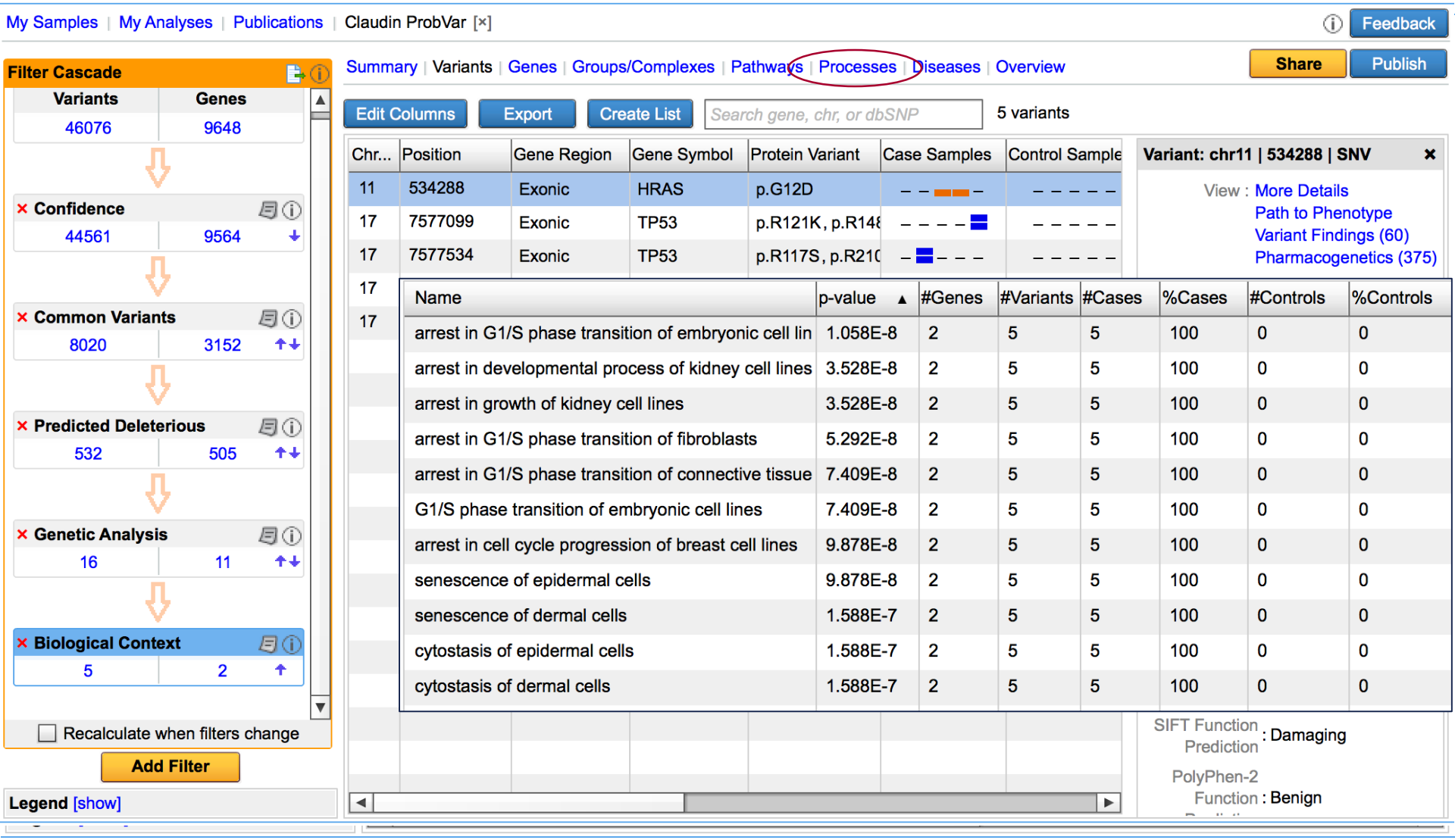
Chr...	Position	Gene Region	Gene Symbol	Protein Variant	Case Samples	Control Sample
11	534288	Exonic	HRAS	p.G12D		
17	7577099	Exonic	TP53	p.R121K, p.R148		

Variant: chr11 | 534288 | SNV

View : More Details
Path to Phenotype
Variant Findings (60)
armacogenetics (375)
RAS
rvey rat sarcoma viral
cogene homolog
4894230
5.5
4 [Validate]
915
r11:534288 [IGV] [UCS
M081305
Uncertain Significance
Exonic
.G12D
.35G>A
missense
Damaging
Benign



Results: The Short List

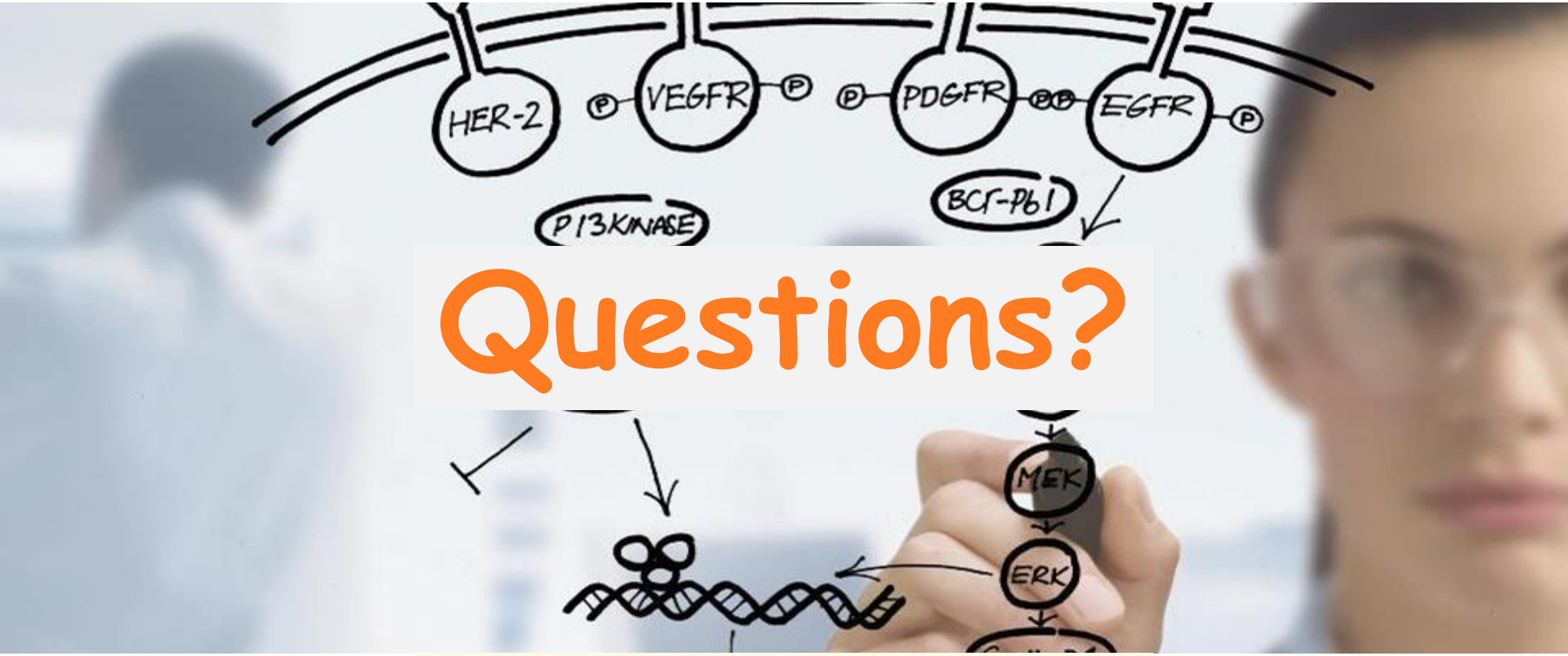


Benefits

Ingenuity Variant Analysis

- Easy to use, allows scientific iteration
  - Filter variants, real-time, based on both standard and proprietary criteria
- Provides variant filtering using public and proprietary databases
  - Allele Frequency Community
- Powerful tools for genetic and statistical tests
- Relate variants to literature and biology
  - Narrow-down, discover, and prioritize variants based on literature findings, pathway and biological associations
- Integrate variants with GWAS, gene expression, or other gene or positional data using BED files

http://www.ingenuity.com/products/variant-analysis



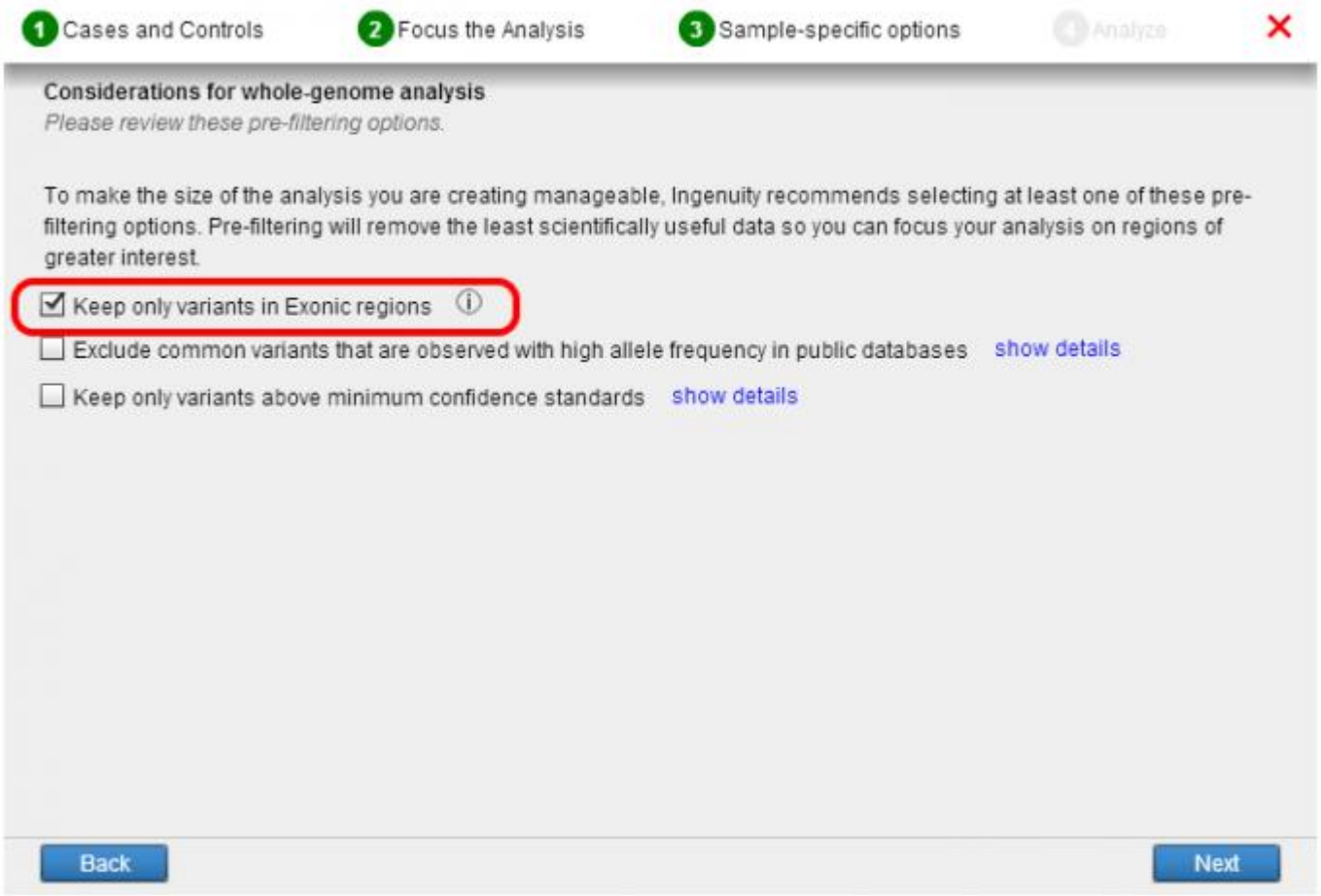
Questions?

Thank You!!

Dev Mistry, Ph.D.
Field Applications Scientist
Devendra.Mistry@qiagen.com

OTHER SLIDES

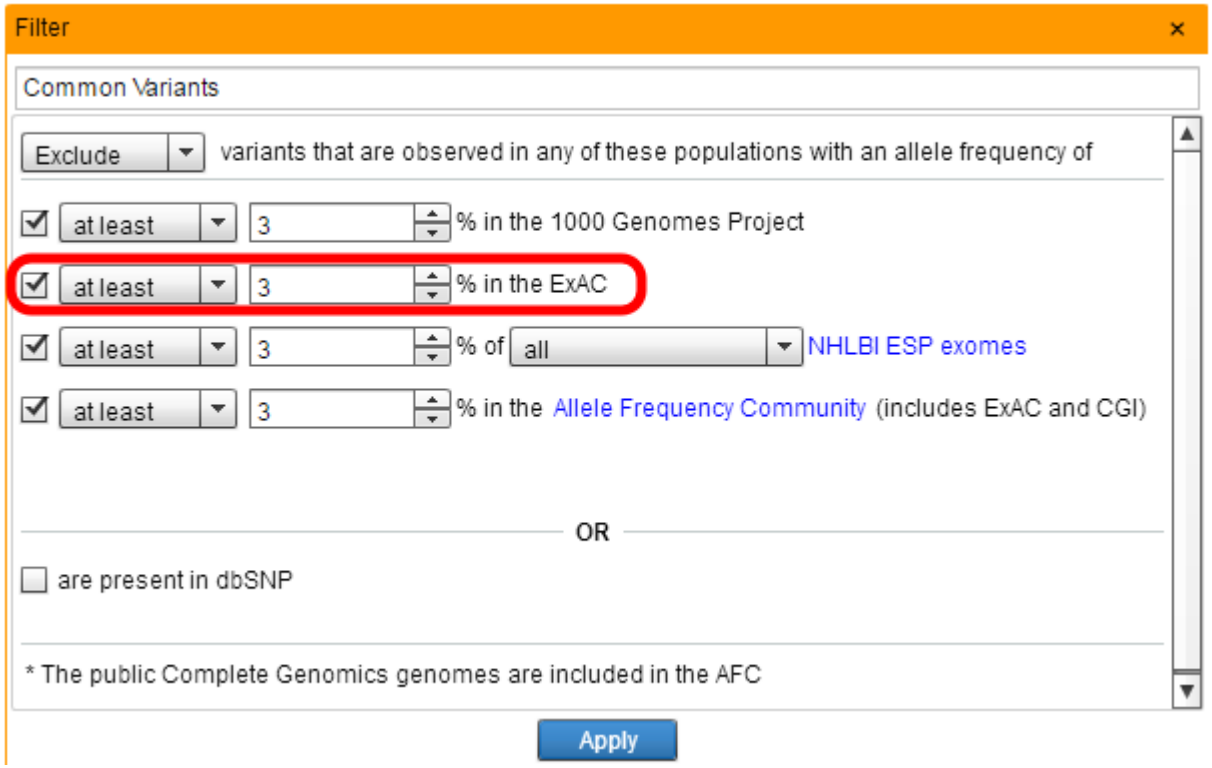
Increase application performance by pre-filtering large samples sets



Remove unnecessary low quality or low value variants upfront to improve analysis speed

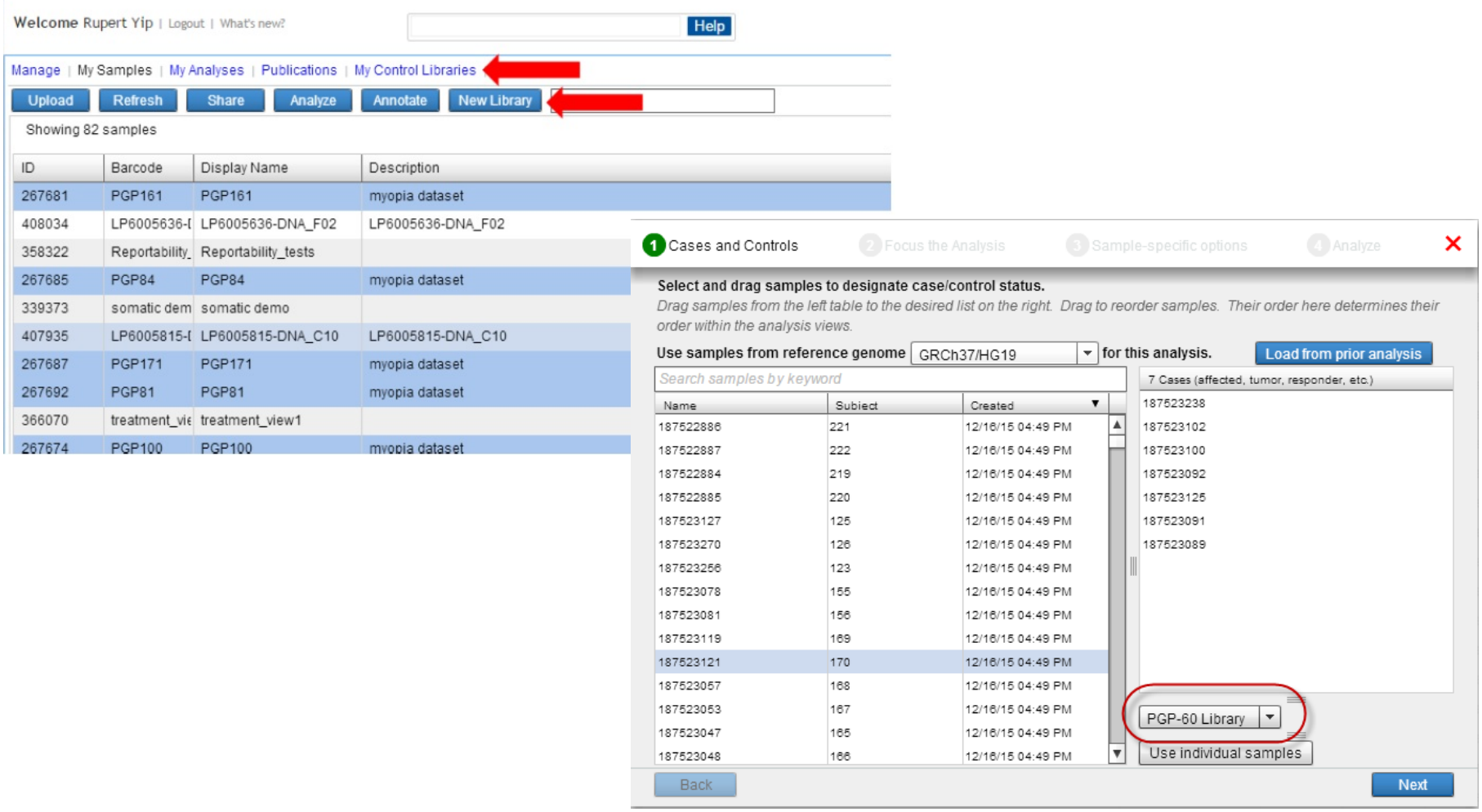
Recent Features

Reduce false positives with inclusion of the ExAC database for filtering out common variants

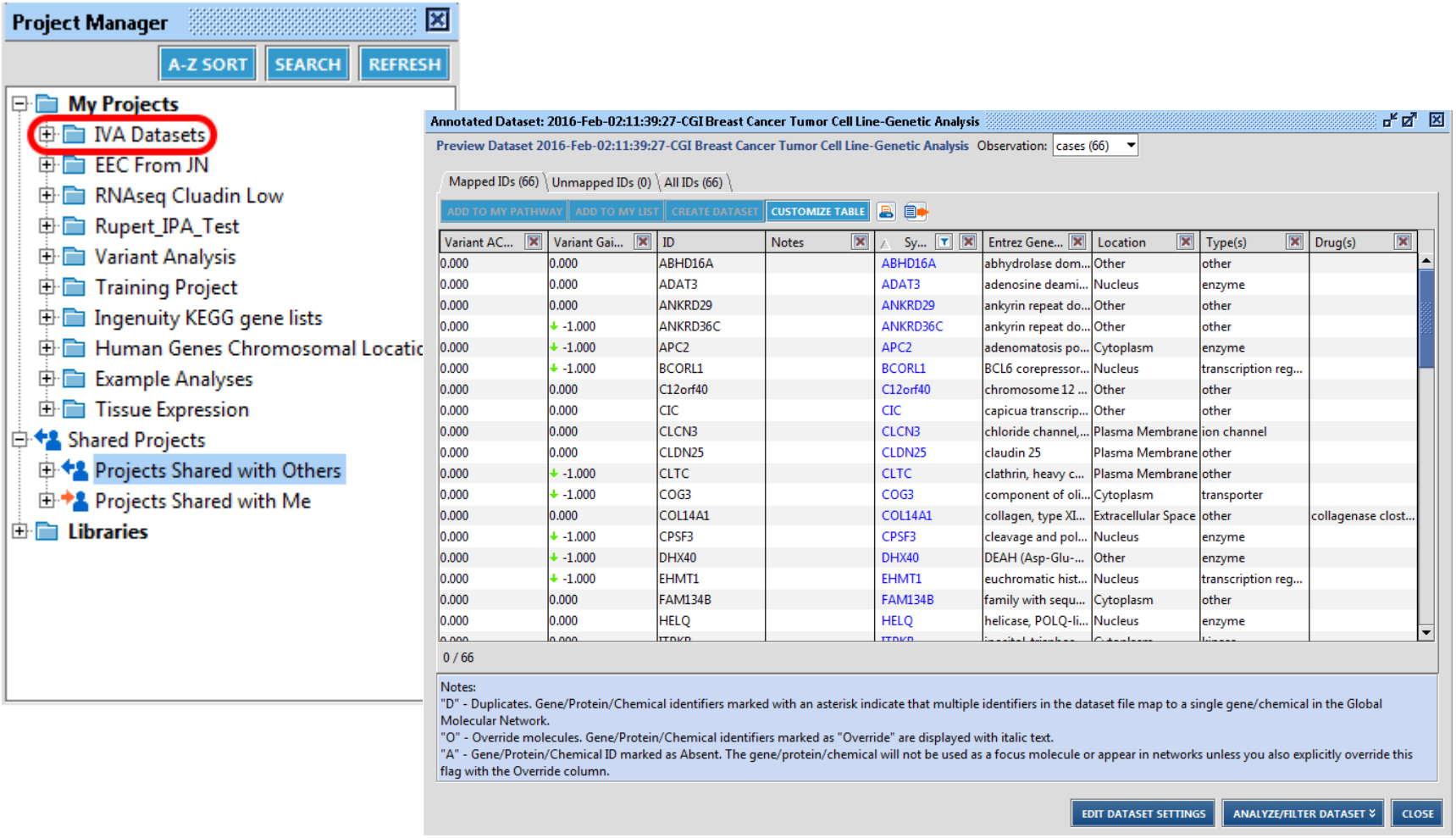


- The Exome Aggregation Consortium (ExAC)
- Exome sequencing data from a wide variety of large-scale sequencing projects
- Spans 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies.

Faster Variant Analysis setup by grouping control samples into defined libraries



Export directly to IPA with inferred gain/loss-of-function values



Allele Frequency Community Data

www.allelefrequencycommunity.org

- Leverage the world's largest pool of anonymized allele frequency data
- **Reduces false positives** in analyses by removing variants that are commonly seen in the general population
- Contains Whole Exome AND Whole Genome data
- Better representation of Insertions and Deletions
- **Larger than ANY other public resource**
  - AFC launched with 70,000 samples with >8,000 as whole genomes
  - 12x larger than Exome Variant Server data
- The initial launch version of the database already provides a 43% average false positive rate reduction in a benchmarking set of whole-genome Diagnostic Odyssey cases
- AFC will grow as more people opt-in
  - Launched on 25th February 2015 with 70,000 Samples including 8,000 Whole Genomes
  - Currently at over 100,000 samples, including over 14,000 Whole Genome samples

Benefits

Ingenuity Variant Analysis

- Easy to use, allows scientific iteration
  - Filter variants, real-time, based on both standard and proprietary criteria
- Provides variant filtering using public and proprietary databases
  - Allele Frequency Community
- Powerful tools for genetic and statistical tests
- Relate variants to literature and biology
  - Narrow-down, discover, and prioritize variants based on literature findings, pathway and biological associations
- Integrate variants with GWAS, gene expression, or other gene or positional data using BED files

http://www.ingenuity.com/products/variant-analysis