CANCER GENOMICS CONSORTIUM

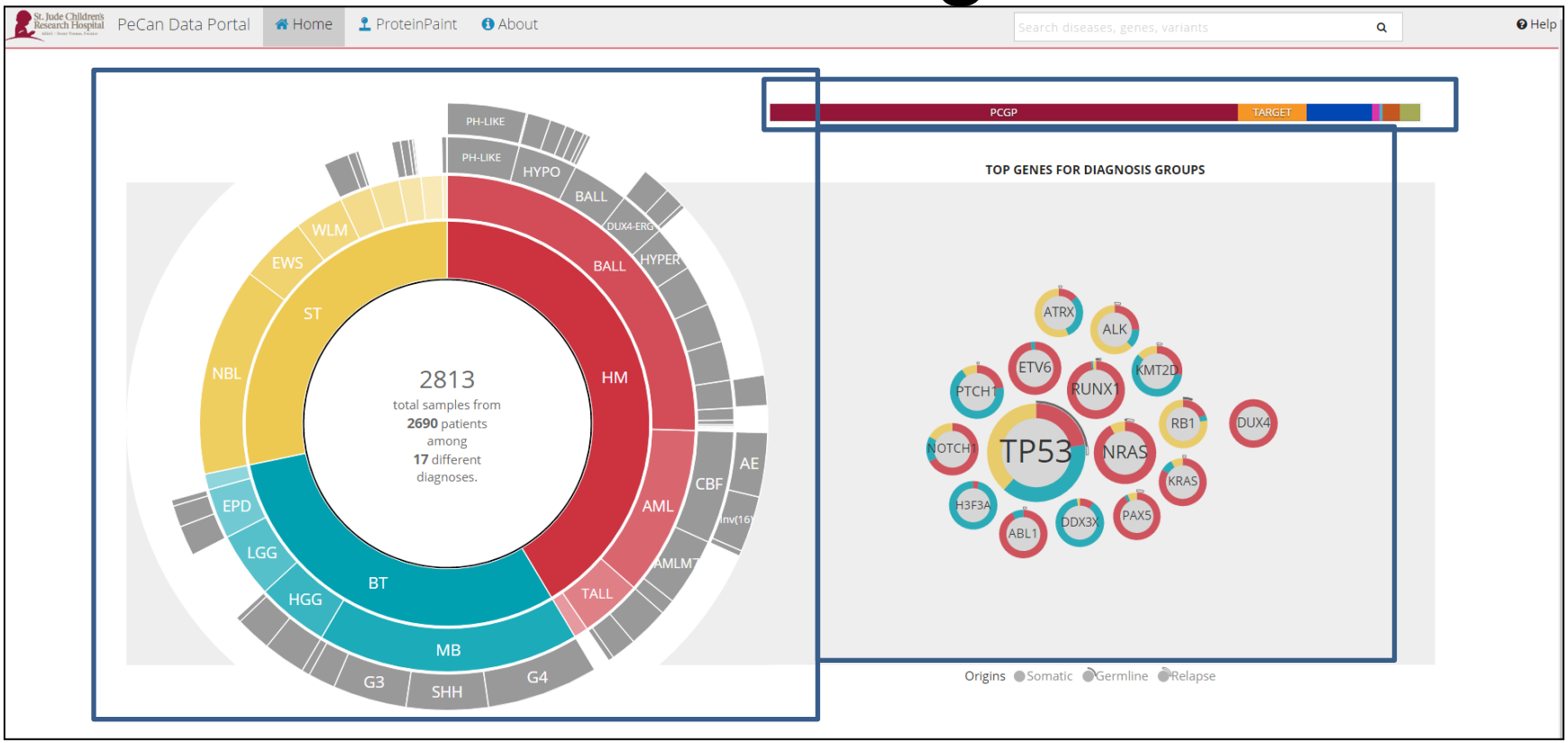
Educating for Best Practices in Clinical Cancer Genomics

PeCan Data Portal

https://pecan.stjude.org/

http://www.nature.com/ng/journal/v48/n1/full/ng.3466.html

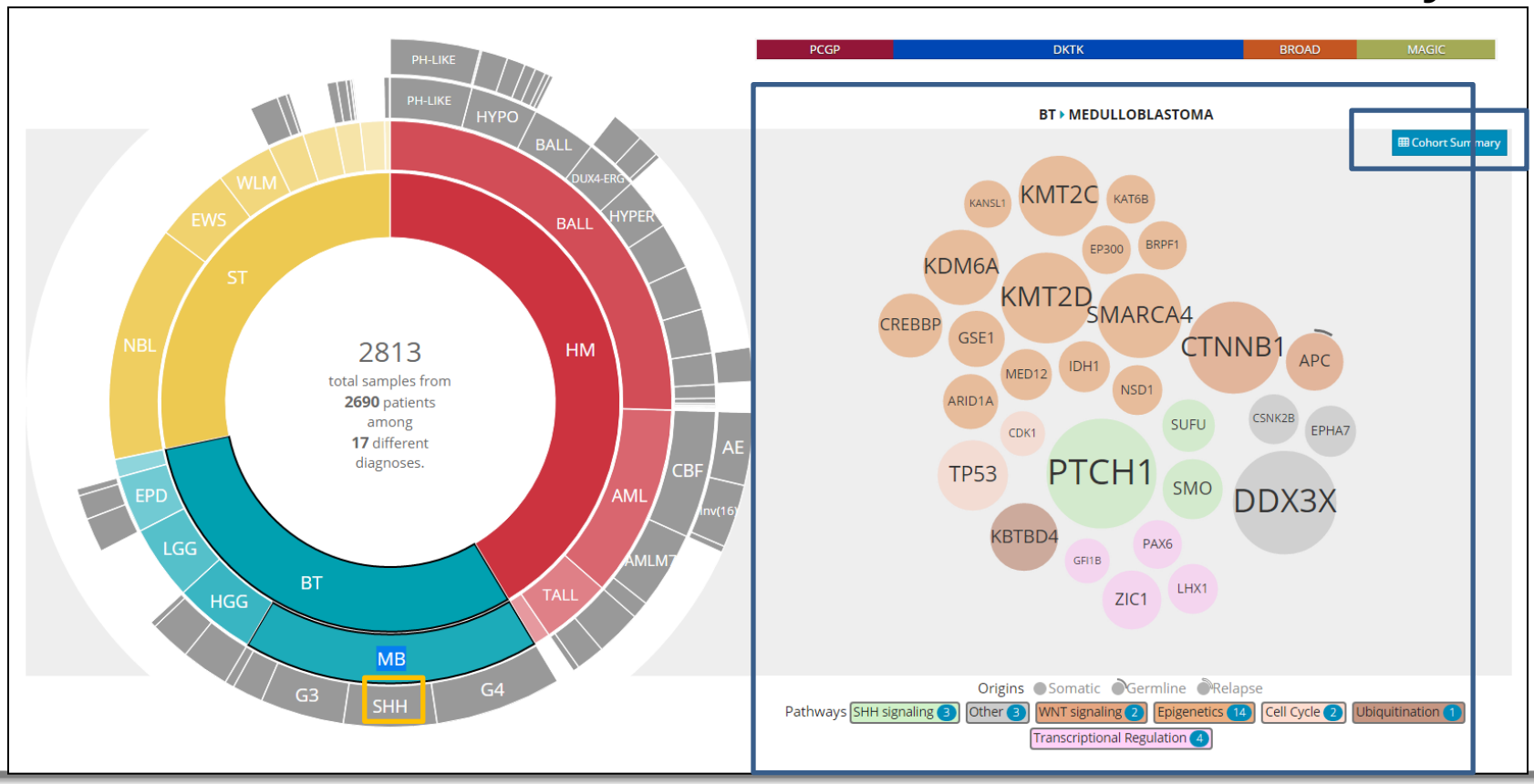
Home Page

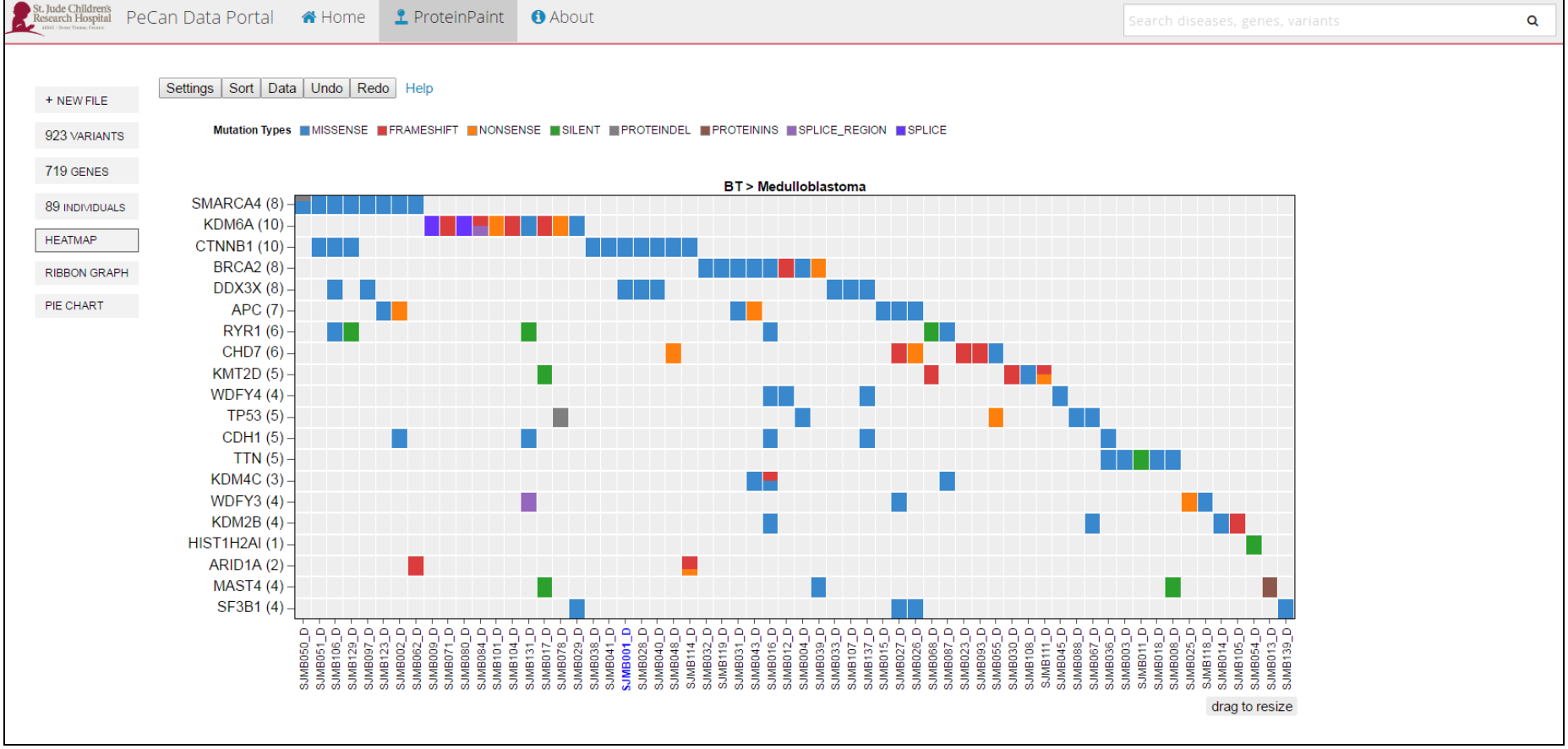


- Pie chart of samples representing cancer types in Data Portal cohorts
- Top genes for diagnosis groups in word cloud
  - If no cohort selected, will show color-coded distribution and mutation origin (not stratified by subtype)
  - If cohort is selected, will show mutation origin
- Bar above top genes shows where studies supplying sample data

Cohort

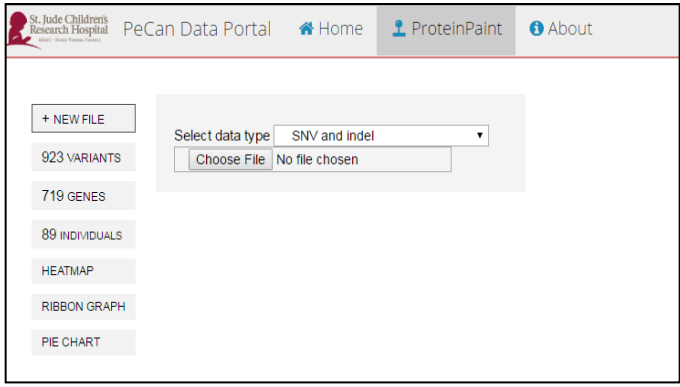
- Cohort selected may show color-coding by pathway in word cloud
- For more details, look at Cohort Summary





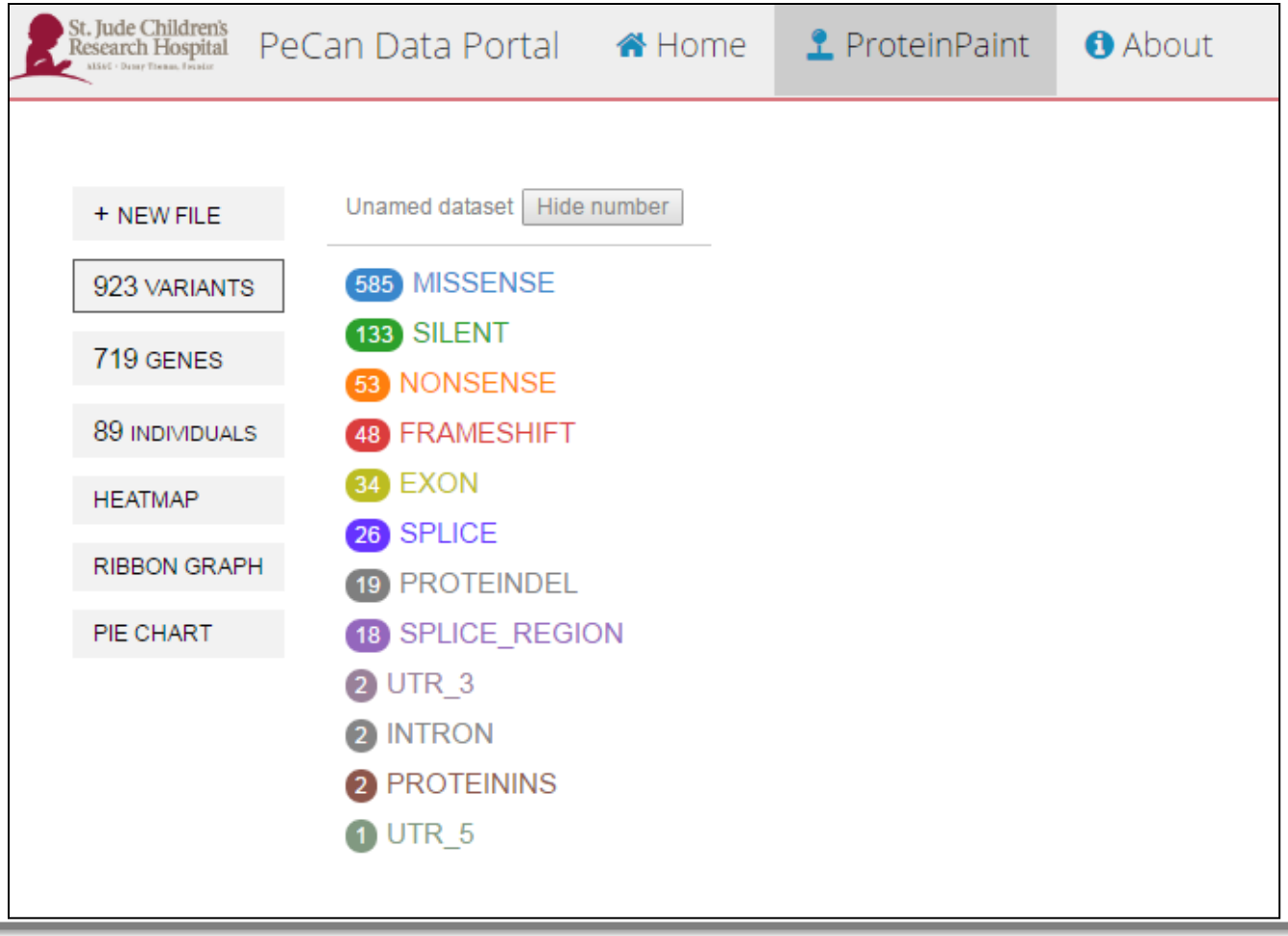
- Default screen in Cohort summary will show heat map
- Menu on left panel
- Can hover over Gene Name on Y-Axis to add or delete gene row or sort selection differently

Add your own data

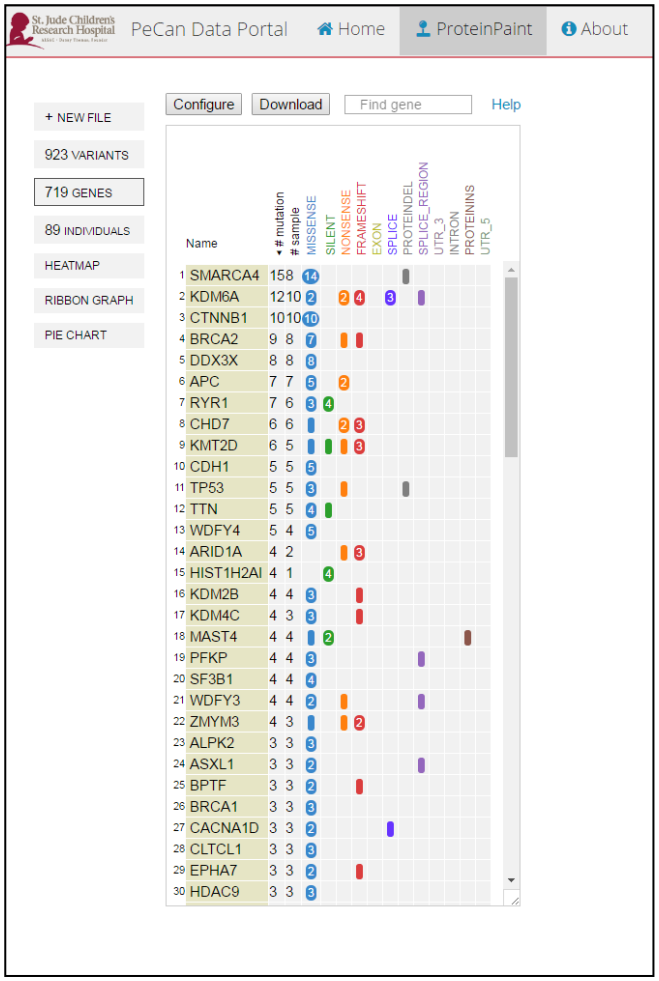


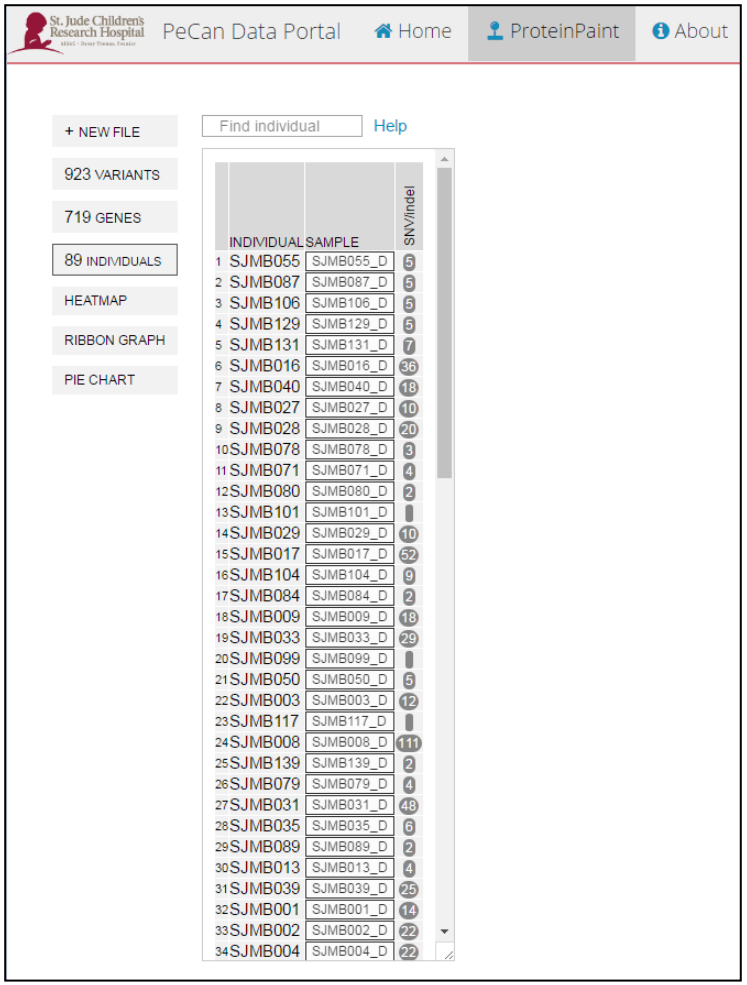
- SNV and indel
  - https://docs.google.com/document/d/1OJ9aXq2_-a3BfIQdKLYCYzrJRTpu4_9i3gephTY-Z38/edit
- SV (tabular format; JSON format to come)
  - https://docs.google.com/document/d/1klDZ0MHVkQTW2-lCu_AvpRE4_FcbhdB-yI17wNdPaOM/edit
- Fusion transcript (tabular format, JSON format to come) – same as SV above
- CNV, gene-level
  - https://docs.google.com/document/d/1WHptqOWNf96V0bYEDpj-EsKZGYnbBNc9aQIrhzdEJaU/edit
- ITD
  - https://docs.google.com/document/d/1Bh9awBsraoHbV8iWXv_3oDeXMsjIAHaOKHr973IJyZc/edit
- Intragenic deletion
  - https://docs.google.com/document/d/1tWbf3rg3BmVIZPGGPk023P0aBkDw_ry5XuZLGyGodyg/edit
- Truncation (N or C terminus loss)
  - https://docs.google.com/document/d/1P1g-Y8r30pSKfan1BhYZcsUtSk7wRb4plaO1S-JCJr4/edit

Cohort Variant Summary



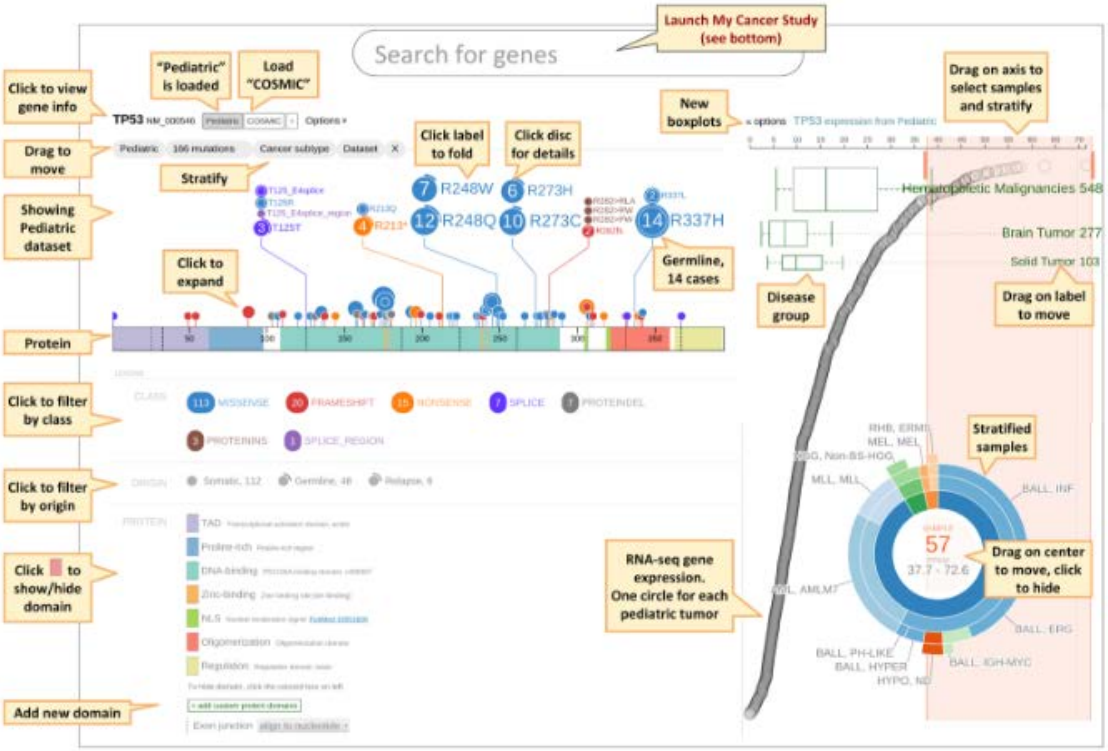
Cohort Gene and Individual Summaries

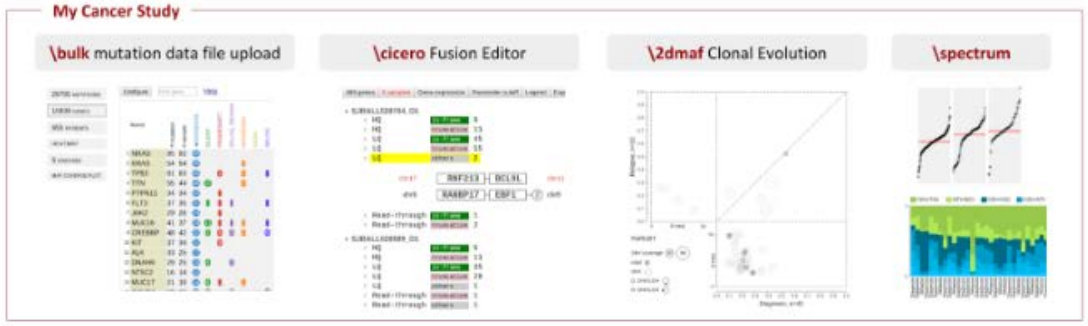




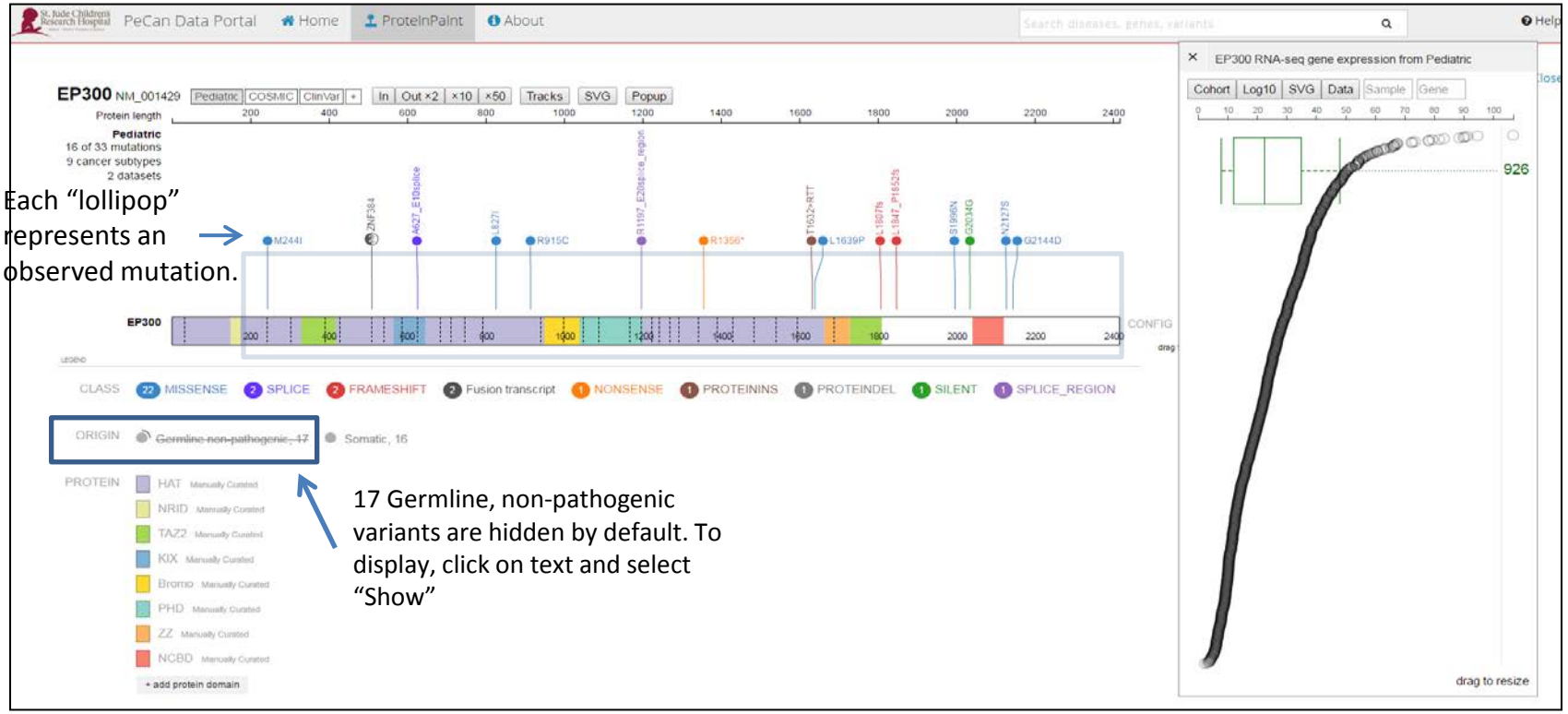
ProteinPaint

St. Jude Children's Research Hospital PeCan Data Portal





ProteinPaint



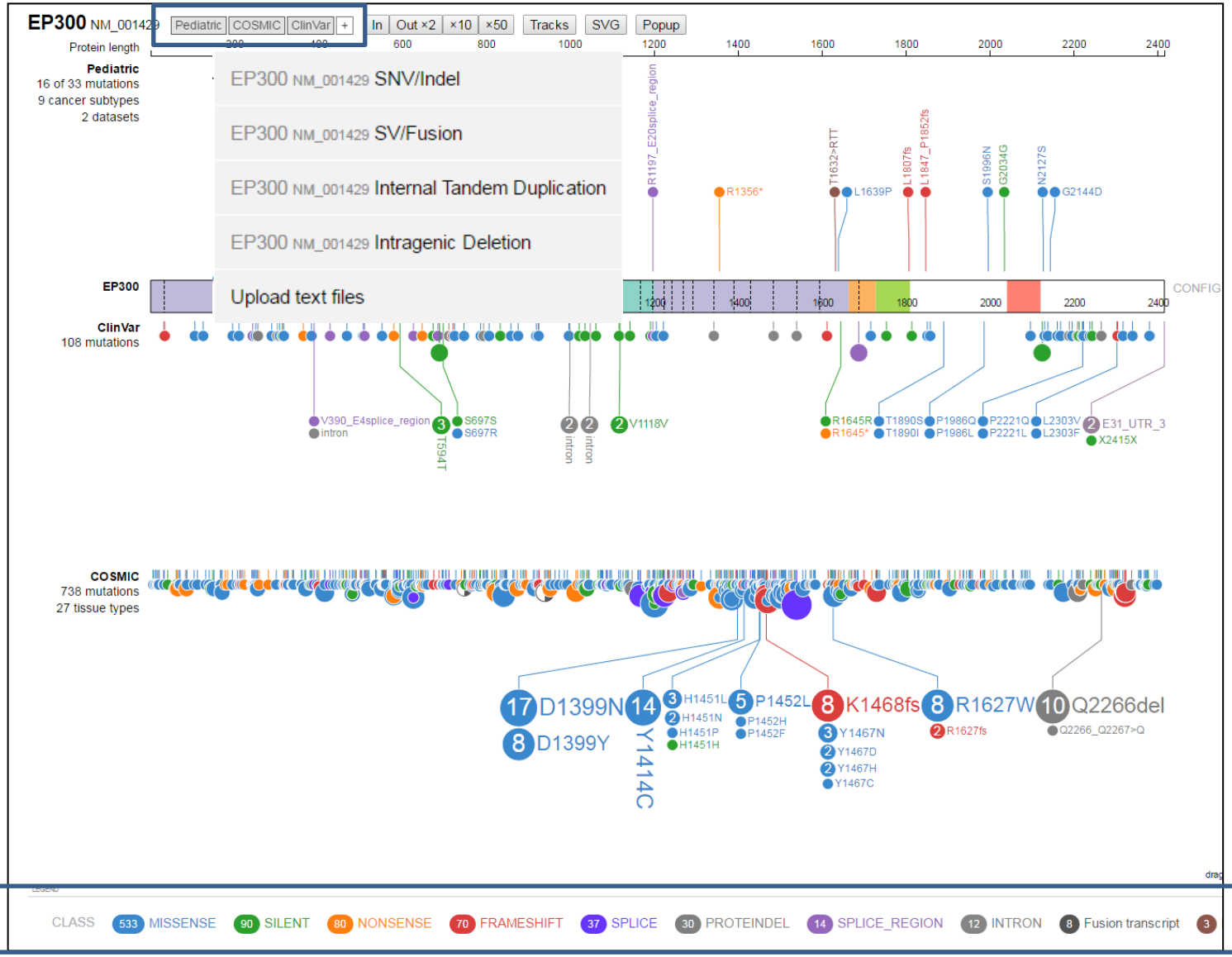
- Choice of transcript – defaults to canonical, but can switch isoforms by clicking on isoform name. NOTE: PeCan shows genes in hg19 genome build
- Multiple Data Sets Available – Pediatric, COSMIC, ClinVar
- Zoom
- Add Custom tracks or adjust view of current tracks
- SVG export or Popup Window of EP300 Protein/Gene View
- RNAseq data

Choice of Transcript and Gene View

- Click on transcript name to make this box pop up.
- Can switch display to genomic, splicing RNA, exon only, protein (default), or an aggregation of all isoforms
- Can also switch viewed isoform

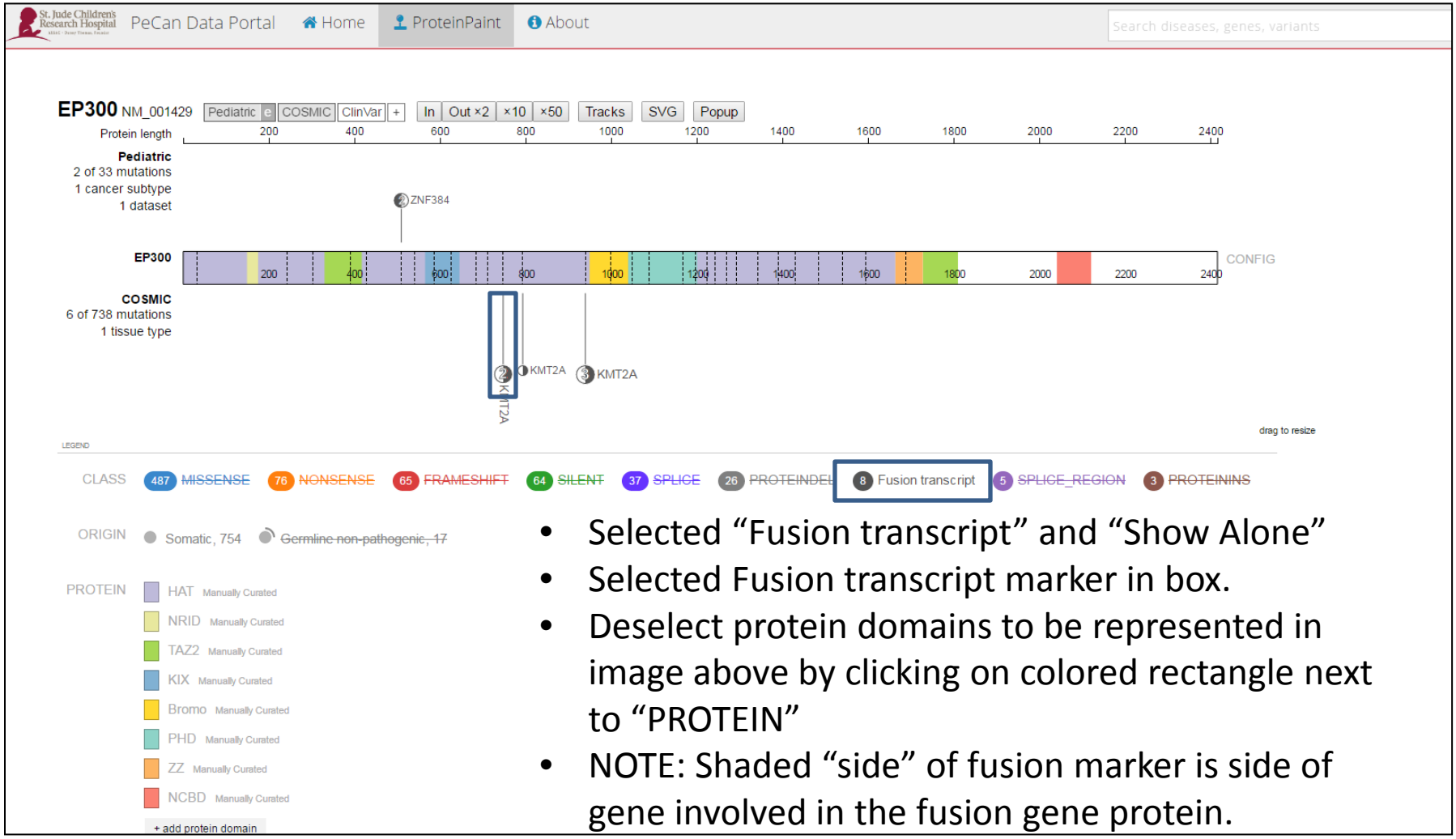
Data Sets

St. Jude Children's Research Hospital PeCan Data Portal



- Can also add your own data by selecting the '+' sign
- Formatting is described within each selection
- To add text file, see examples, format on this page or Slide 5
- Filter selection by class of mutation

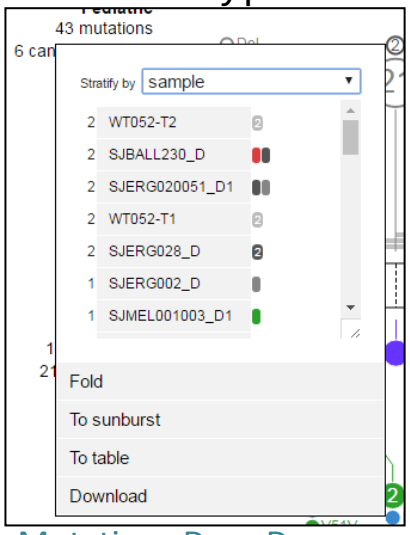
Sort by Mutation Class



- Selected “Fusion transcript” and “Show Alone”
- Selected Fusion transcript marker in box.
- Deselect protein domains to be represented in image above by clicking on colored rectangle next to “PROTEIN”
- NOTE: Shaded “side” of fusion marker is side of gene involved in the fusion gene protein.

Stratifying Data

- Additional labels are underneath the "Pediatric" and "COSMIC" labels if those data types are selected.
  - Mutations
  - Cancer Types
  - Datasets
  - Tissue types



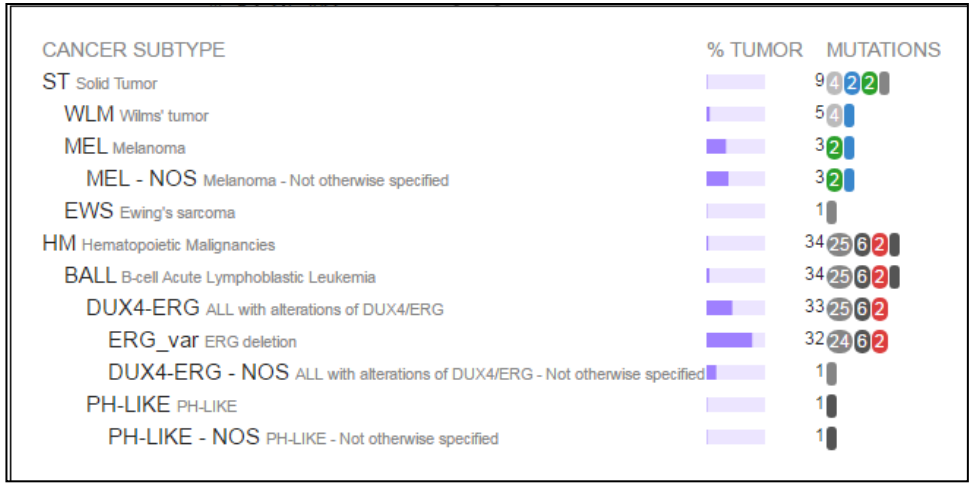
Mutations Drop Down

Sunburst

- Mutations
  - Click on "## mutations" to see stratificaton window (left bottom)
    - Can stratify by
      - Sample – sample ID
      - Specimen – diagnostic?
      - Dataset_label – what dataset is the mutation from?
      - PMID
      - Committee Classification
      - Origin Type
      - LOH
      - In-frame
      - UseNterm
      - +-Strand
    - Will replicate selected information from dataset and visualize below the gene
  - Other options for visualization
    - Sunburst (left)
    - Table
  - Download
    - Information can be copy/pasted into Excel
      - Contains all information on mutations in gene separated by mutation type

Stratifying Data

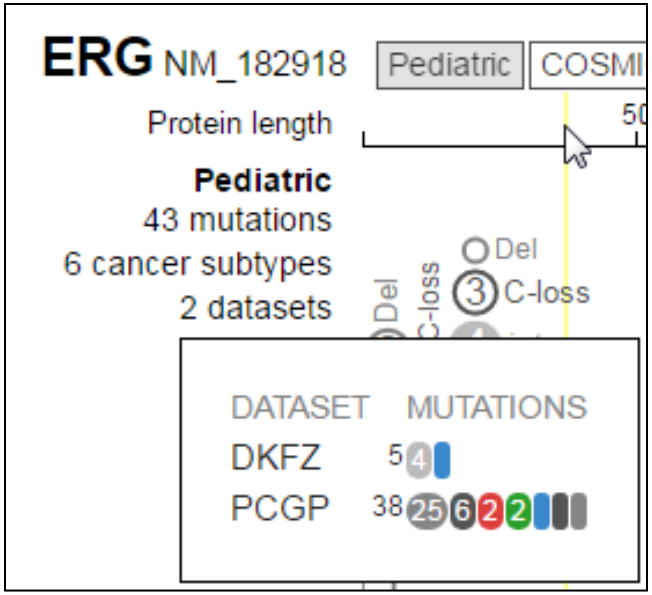
- Additional labels are underneath the "Pediatric" and "COSMIC" labels if those data types are selected.
  - Mutations
  - Cancer Types
  - Datasets
  - Tissue types



- Click on "Cancer Types" label to show breakdown of variants by subtype
- % Tumors within each subtype with a mutation in this gene
- Distribution of mutation classes by subtype
  - Important for inferring driver/passenger events
    - Missense/silent mutations suggest passenger mutations
- Will replicate data from selected cancer subtype to visualize below gene

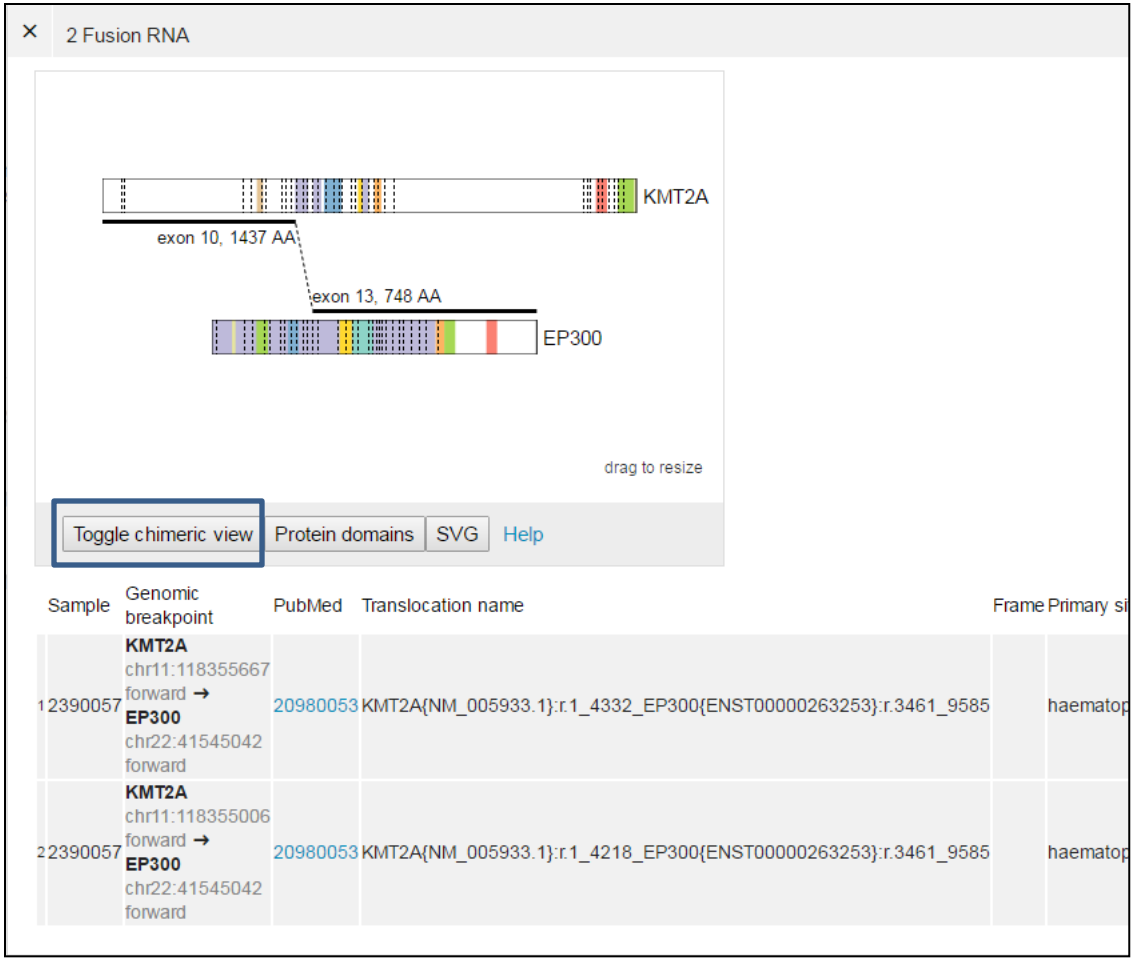
Stratifying Data

- Stratify by Dataset: mirrors datasets highlighted on homepage.
  - PCGP – Pediatric Cancer Genome Project (St. Jude/Wash U)
  - TARGET – Therapeutically Applicable Research to Generate Effective Treatments (NIH)
  - DKFZ – German Cancer Research Center
  - Shanghai Children's Medical Center Pediatric ALL Project
    - UT SW Medical Center Wilms' Tumor Study



- Selecting a dataset will replicate mutations associated with that dataset and visualize below the gene.

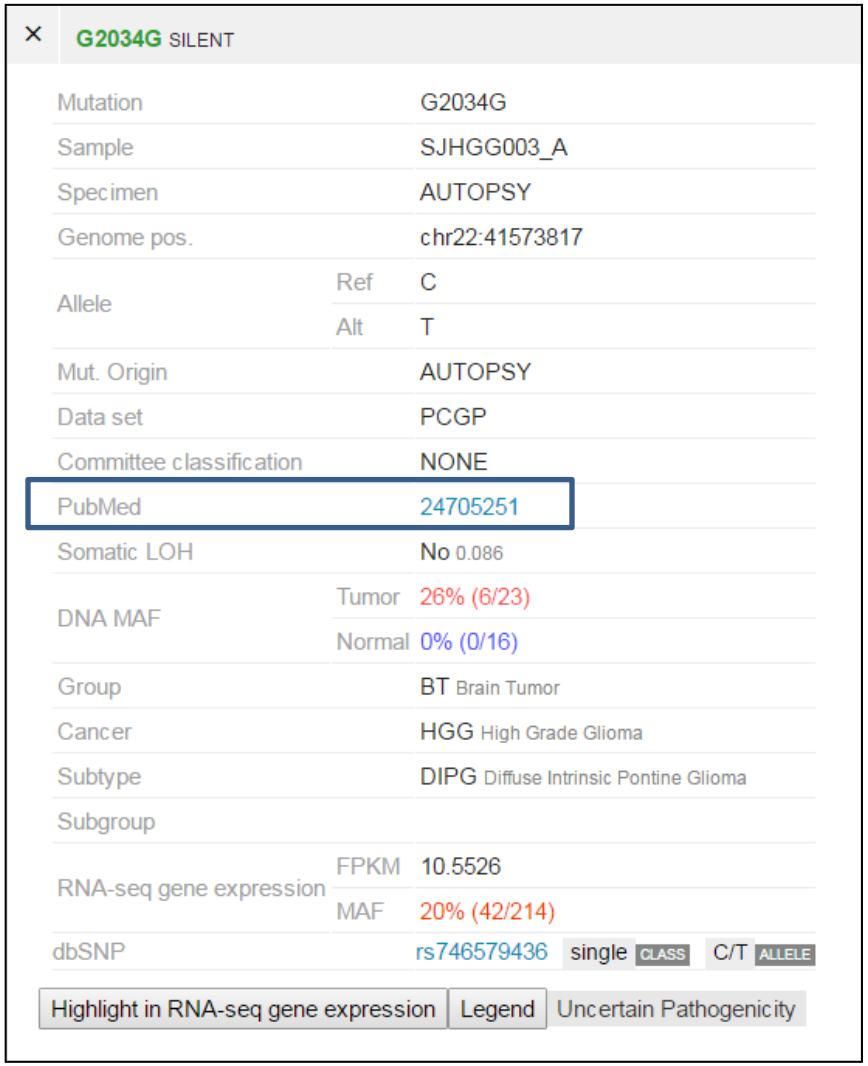
Fusion Visualization



Sample	Genomic breakpoint	PubMed	Translocation name	Frame	Primary si
12390057	**KMT2A** chr11:118355667 forward → **EP300** chr22:41545042 forward	20980053	KMT2A{NM_005933.1}:r.1_4332_EP300{ENST00000263253}:r.3461_9585		haematop
22390057	**KMT2A** chr11:118355006 forward → **EP300** chr22:41545042 forward	20980053	KMT2A{NM_005933.1}:r.1_4218_EP300{ENST00000263253}:r.3461_9585		haematop

- Fusion transcript is annotated and visualized by selecting fusion transcript marker of interest.
- PMID of primary publication in hyperlink in table when applicable along with sample information.
- Toggle chimeric view allows visualization of protein transcript in one bar.

Selection of SNV

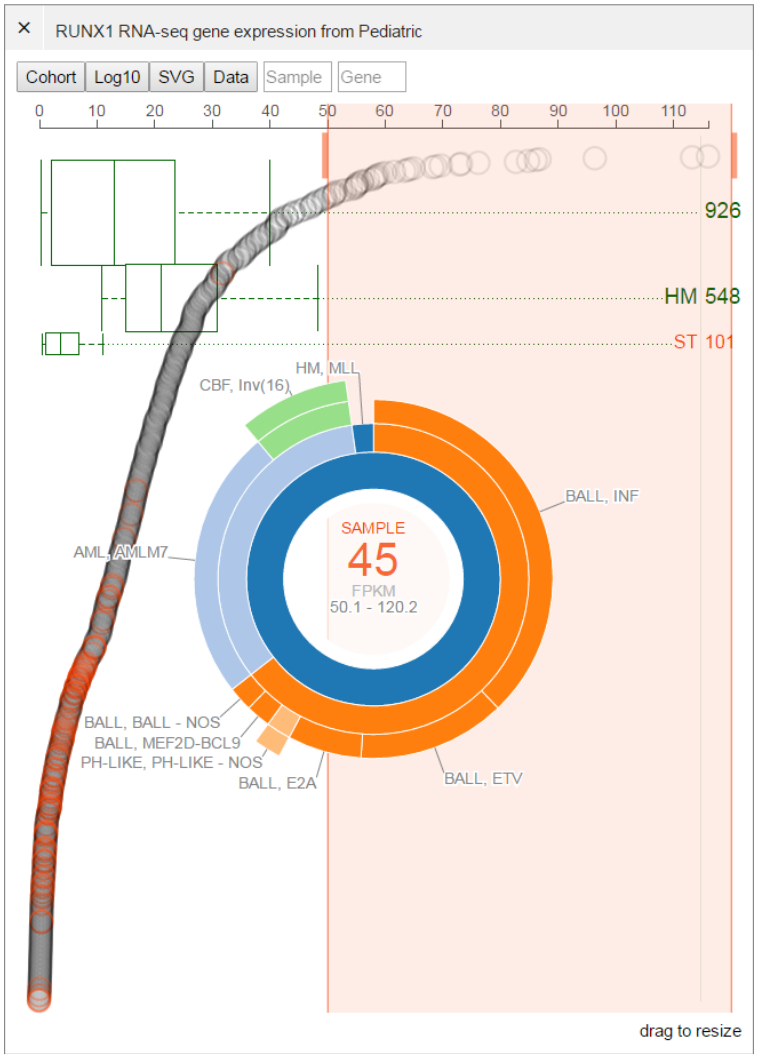


× **G2034G** SILENT

Mutation		G2034G
Sample		SJHGG003_A
Specimen		AUTOPSY
Genome pos.		chr22:41573817
Allele	Ref	C
	Alt	T
Mut. Origin		AUTOPSY
Data set		PCGP
Committee classification		NONE
PubMed		24705251
Somatic LOH		No 0.086
DNA MAF	Tumor	26% (6/23)
	Normal	0% (0/16)
Group		BT Brain Tumor
Cancer		HGG High Grade Glioma
Subtype		DIPG Diffuse Intrinsic Pontine Glioma
Subgroup		
RNA-seq gene expression	FPKM	10.5526
	MAF	20% (42/214)
dbSNP		rs746579436 single CLASS C/T ALLELE

Highlight in RNA-seq gene expression | Legend | Uncertain Pathogenicity

RNA-Seq Data



- FPKM = fragments per kilobase of transcript per million mapped reads.
- Select Cohort to make green box plots for different cancer types.
- Can toggle between normal scale or log 2 for FPKM (x-axis)
- Make a selection on x-axis and drag to make pie chart appear. That data is stratified by disease type.
- When collapsed, you can recover RNA-Seq Data by clicking “e” next to the “Pediatric” button above the gene.

Hg18 Data or Mouse mm9 Data?

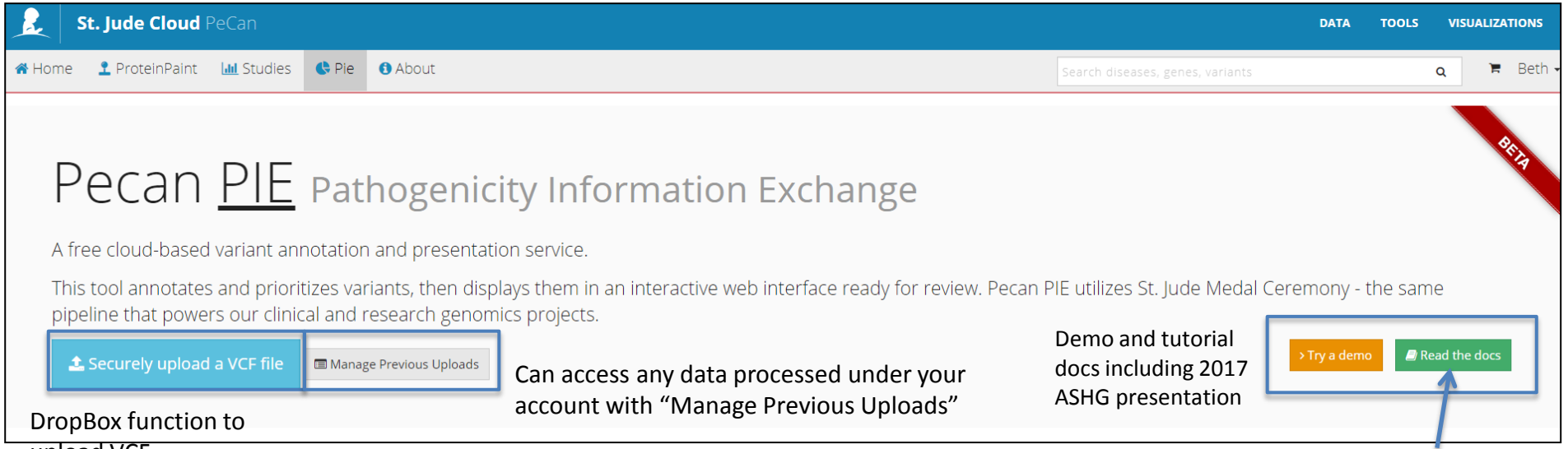
- To use genes found in hg18 or mm9 mouse genome (and other features hidden on PeCan site), use this link: https://pecan.stjude.org/pp
  - Select genome from the drop down menu first, then type in gene name
  - Also on this version of the website
    - API: under Help→Embed in your website
    - Documentation on URL parameters, organizing data into a study.
    - Several other features under "Apps"

PeCan PIE (Pathogenicity Information Exchange)

- As of 10/23/17 – Beta version
- Cloud-based cancer annotation and presentation service. (DNA Nexus application)
- Will annotate and rank variants
- Display on one interactive web page for your review
- Can select ACMG criteria met and add custom interpretations of data.
- Requires VCF file as input

PeCan PIE

- Click on PIE button on home screen.
- You may need to set up a DNA Nexus account.



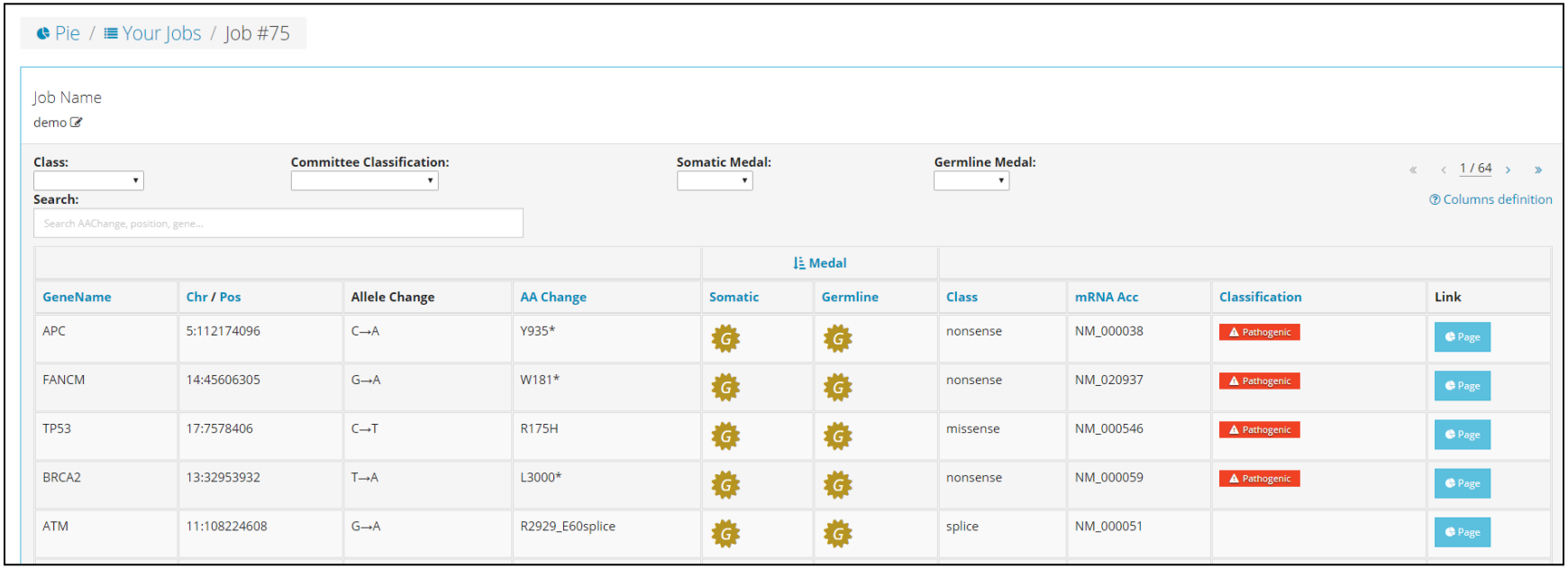
DropBox function to upload VCF

Can access any data processed under your account with "Manage Previous Uploads"

Demo and tutorial docs including 2017 ASHG presentation

Great tutorial docs! Use these and the ASHG presentation!

PeCan PIE – Job Page



Pie / Your Jobs / Job #75

Job Name
demo

Class: | Committee Classification: | Somatic Medal: | Germline Medal:

Search:
Search AAChange, position, gene...

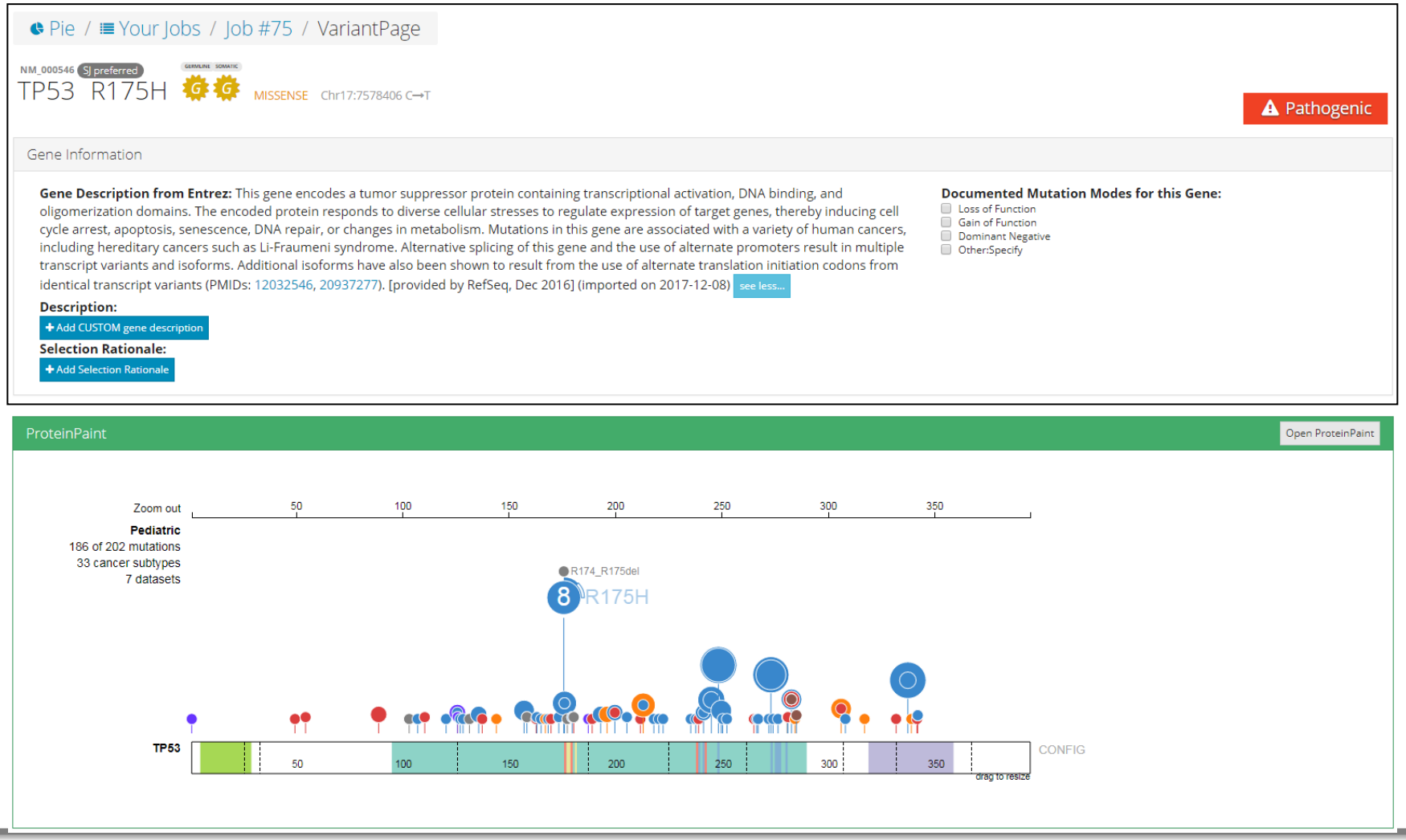
1 / 64

Columns definition

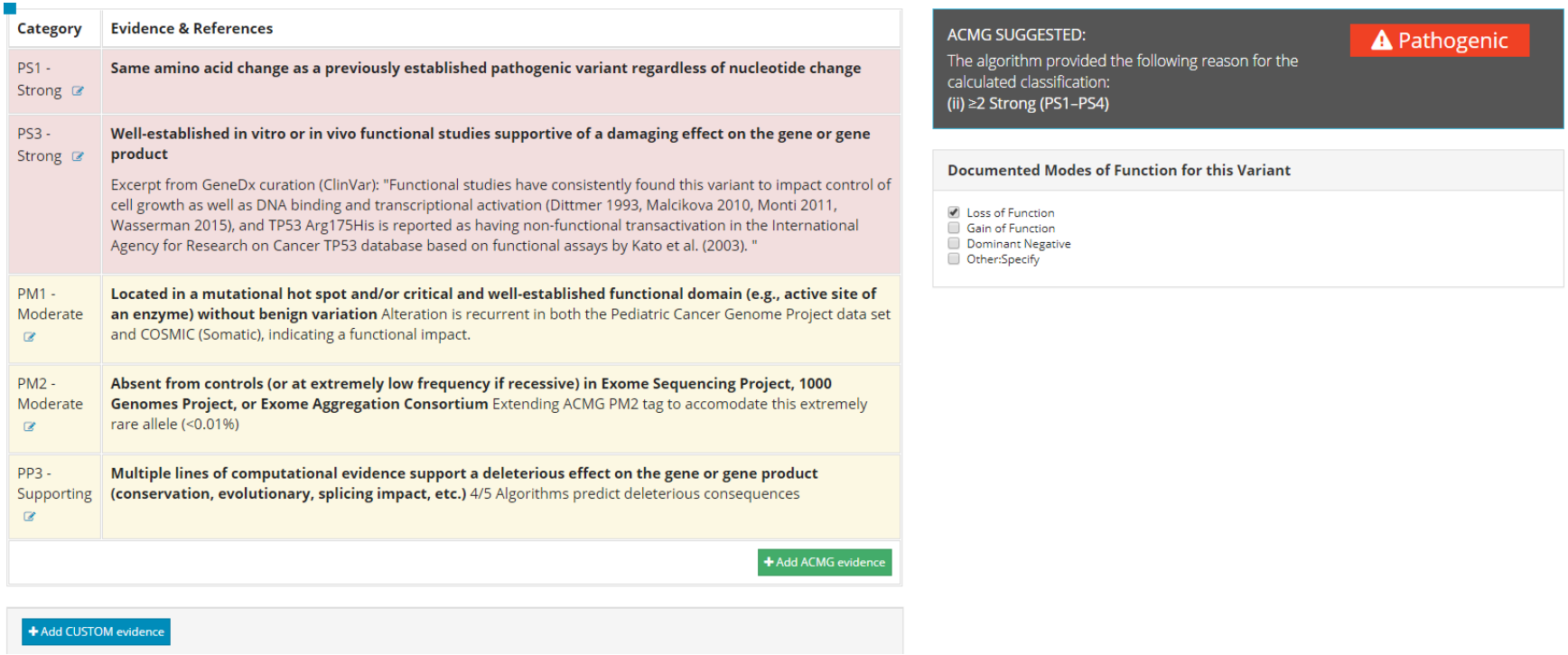
				Medal					
GeneName	Chr / Pos	Allele Change	AA Change	Somatic	Germline	Class	mRNA Acc	Classification	Link
APC	5:112174096	C→A	Y935*			nonsense	NM_000038	Pathogenic	Page
FANCM	14:45606305	G→A	W181*			nonsense	NM_020937	Pathogenic	Page
TP53	17:7578406	C→T	R175H			missense	NM_000546	Pathogenic	Page
BRCA2	13:32953932	T→A	L3000*			nonsense	NM_000059	Pathogenic	Page
ATM	11:108224608	G→A	R2929_E60splice			splice	NM_000051		Page

- Ranked in order of relevance based on somatic and germline medal classifications.
- Link to Variant page on right column

PeCan PIE Variant Page



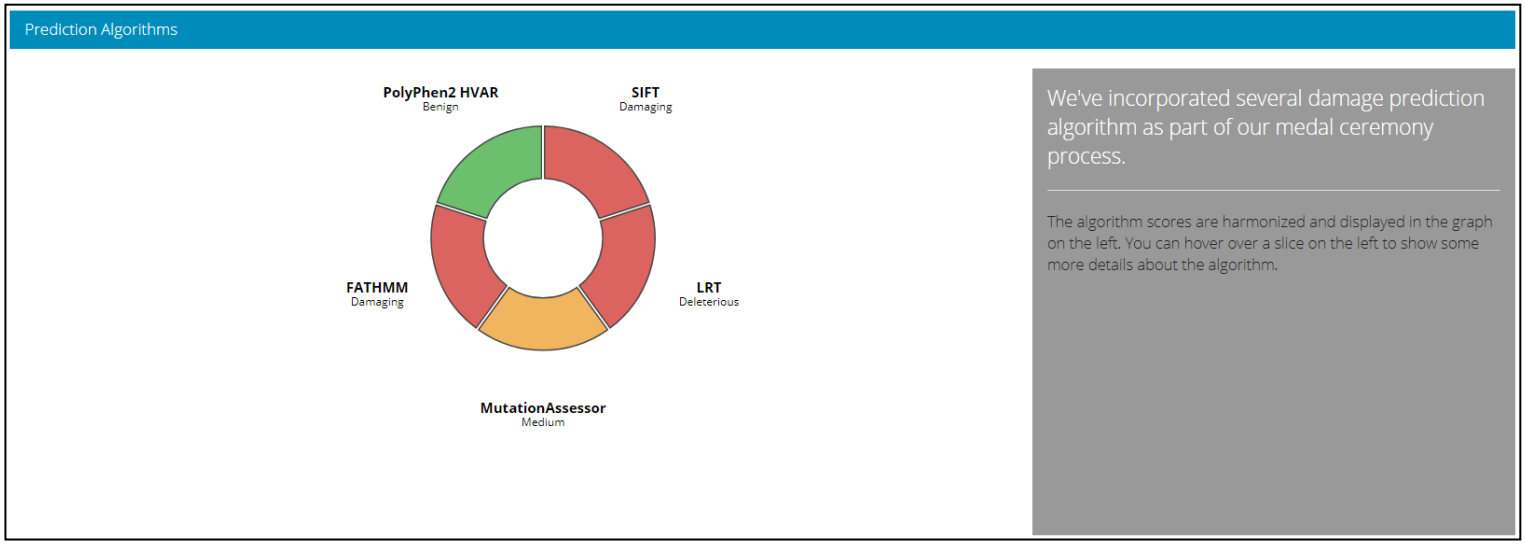
PeCan PIE Variants Page

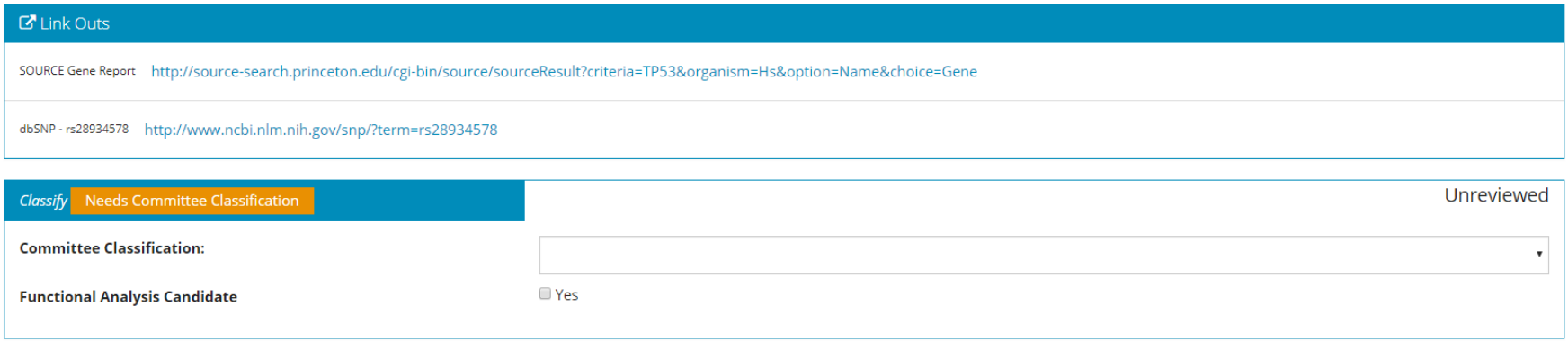


Category	Evidence & References
PS1 - Strong	**Same amino acid change as a previously established pathogenic variant regardless of nucleotide change**
PS3 - Strong	**Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product** Excerpt from GeneDx curation (ClinVar): "Functional studies have consistently found this variant to impact control of cell growth as well as DNA binding and transcriptional activation (Dittmer 1993, Malcikova 2010, Monti 2011, Wasserman 2015), and TP53 Arg175His is reported as having non-functional transactivation in the International Agency for Research on Cancer TP53 database based on functional assays by Kato et al. (2003). "
PM1 - Moderate	**Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation** Alteration is recurrent in both the Pediatric Cancer Genome Project data set and COSMIC (Somatic), indicating a functional impact.
PM2 - Moderate	**Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium** Extending ACMG PM2 tag to accomodate this extremely rare allele (<0.01%)
PP3 - Supporting	**Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)** 4/5 Algorithms predict deleterious consequences

* Also includes ClinVar, Allele frequency information, Medal ceremony determination fields, links to gene information on dbSNP and SOURCE if available.

PeCan PIE – Prediction Algorithms and Final Classification





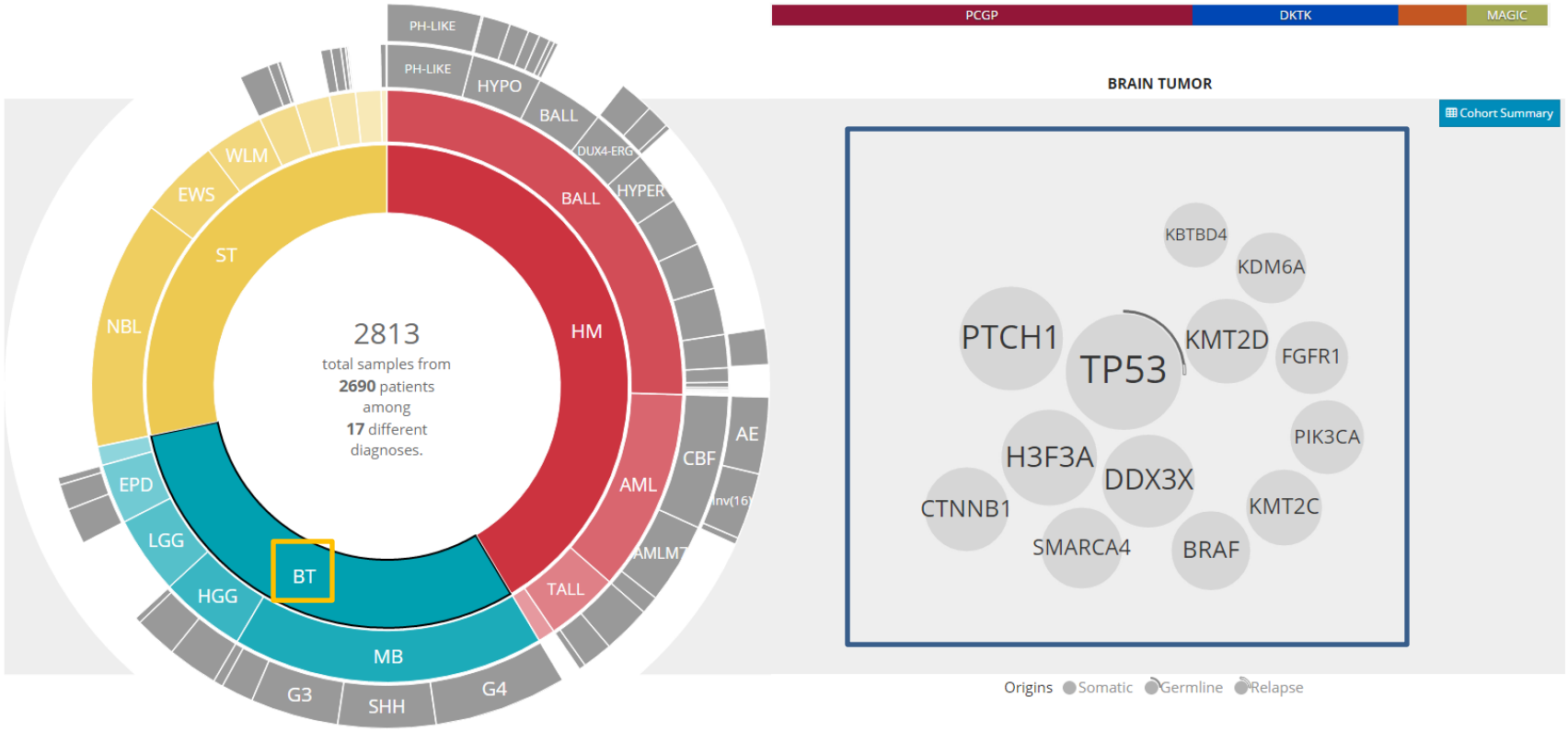
Other Pipeline Services (in Beta)

- Rapid RNA-Seq
  - FASTQ/BAM input to find gene fusions and ITDs in dataset
- ChIP-Seq Peak Calling
  - Paired ChIP-Seq FASTQ files
- WARDEN Differential Expression
  - Multiple paired RNA-Seq FASTQs, sample list file
  - Performs alignment, coverage analysis, gene counts, and differential expression.
- HLA Typing and Neoepitope Prediction
  - Single-end/paired-end WGS FASTQ files or an aligned WGS BAM file
  - Identifies HLA alleles and predicted epitope affinities of peptides.

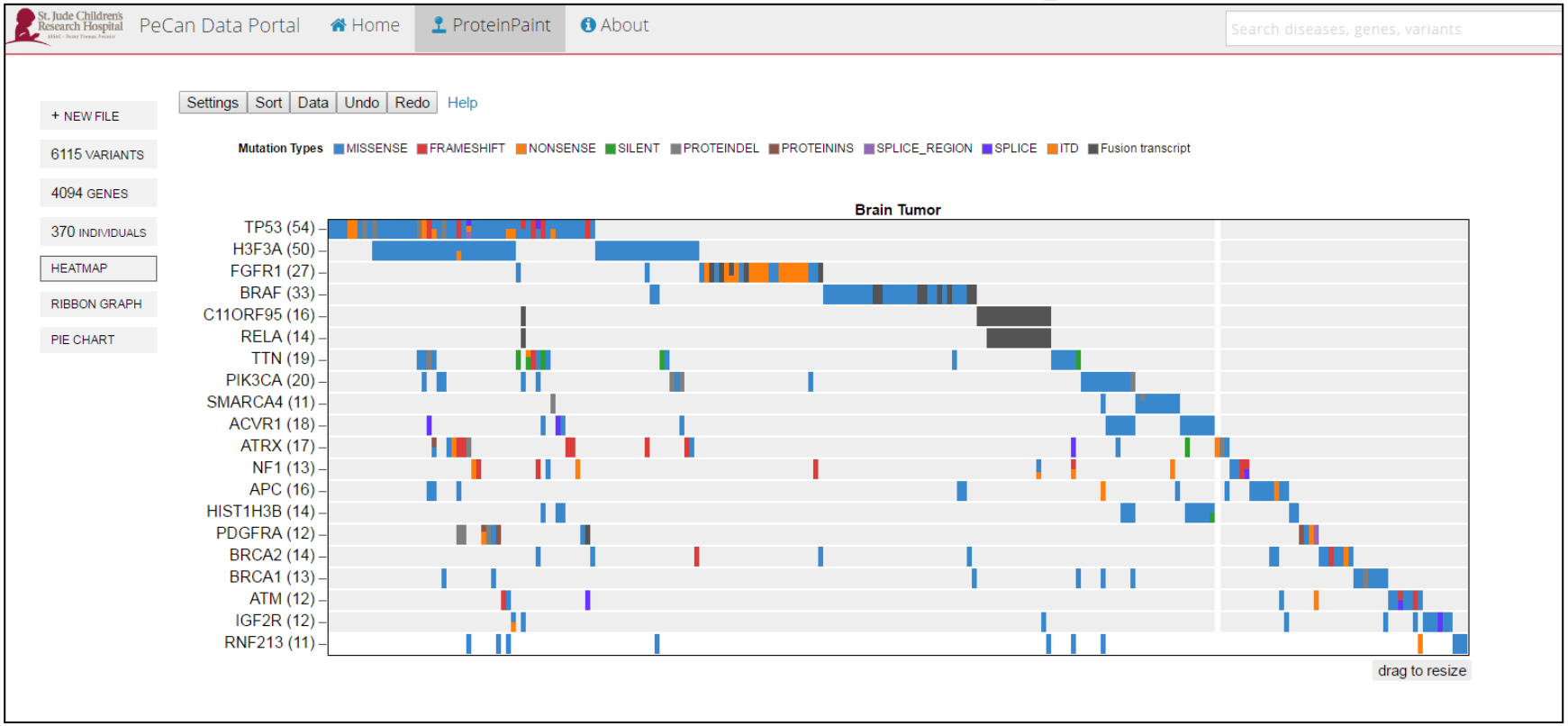
Scenario #1

- Your lab wants to make an NGS panel for pediatric brain tumors. What genes should you look into using for your panel?
  - Gene Cloud
  - Cohort Summary

Select 'BT' from Home Page or type in Search bar and see Gene Cloud



Go Into Cohort Summary and Observe Heat Map

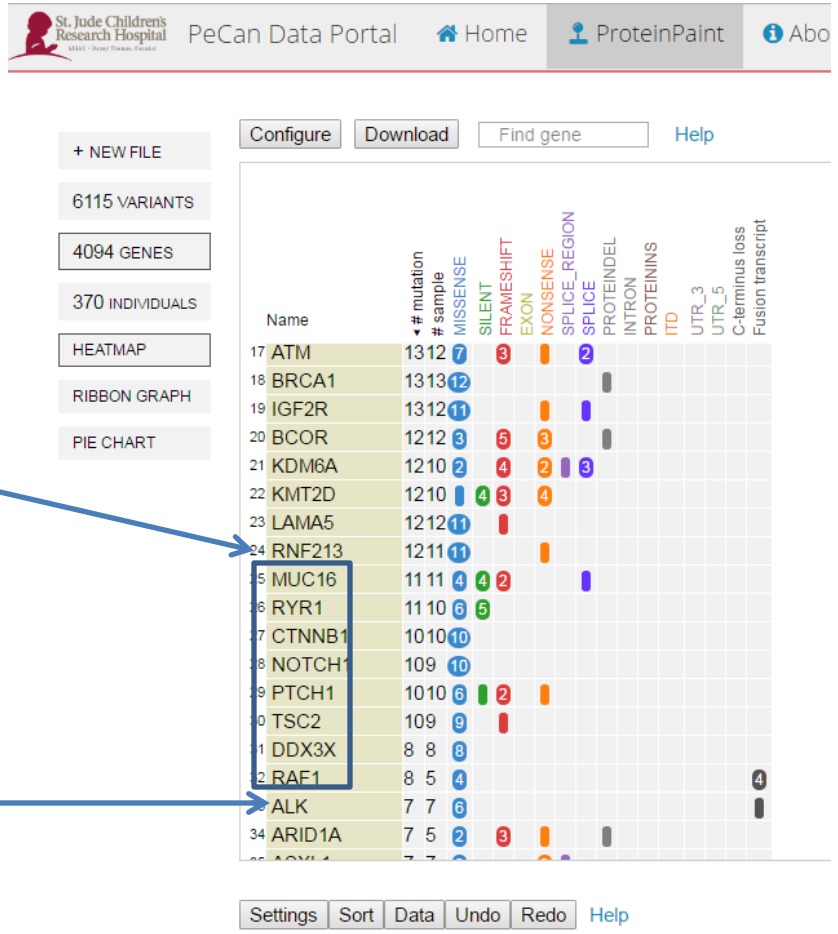


- If you would like to see more genes see next slide.
- If you would like to narrow NGS regions down to regions where mutations have been previously observed, look at ProteinPaint for each gene (choose appropriate data sets) and narrow regions from there.

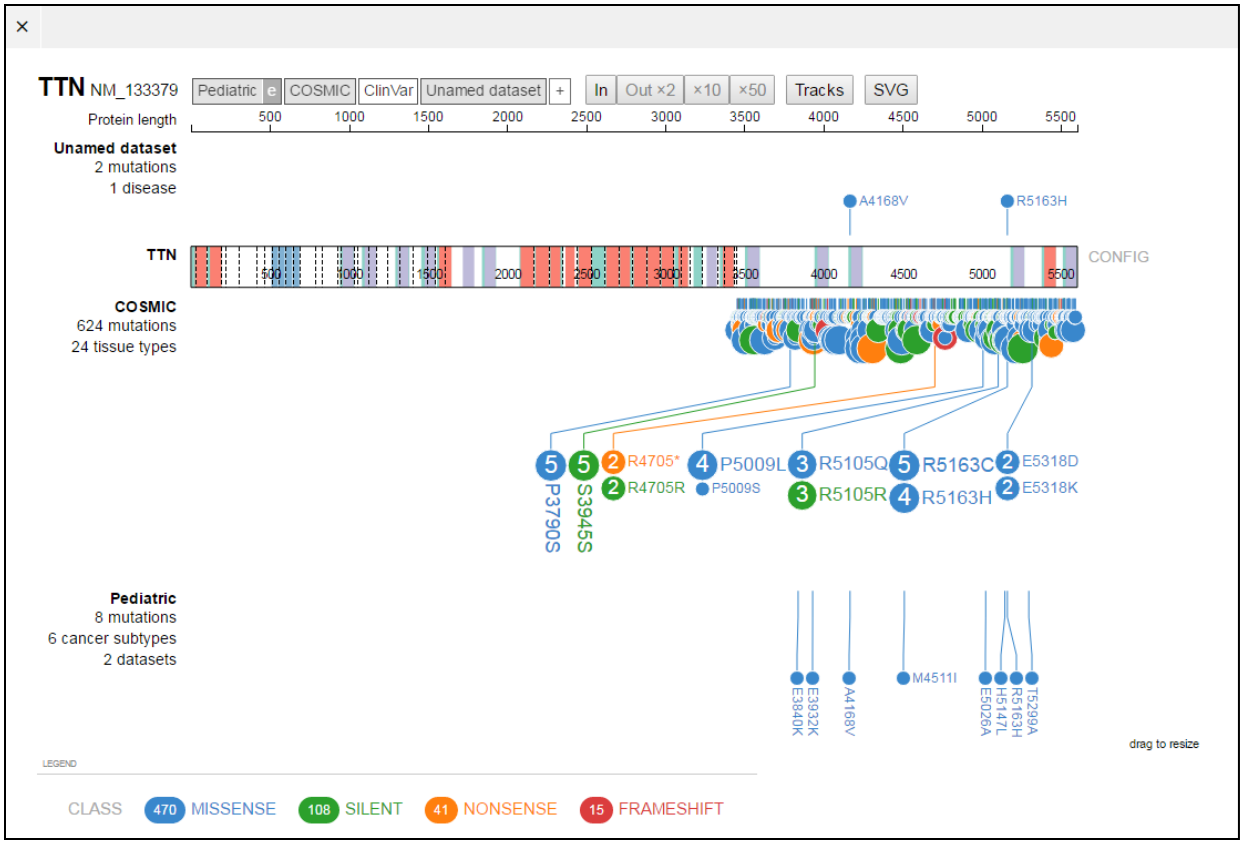
Want to Investigate more genes for your panel?

RNF213 = Last gene on heat map

If you set desired cutoff at 8 mutations observed, you may want to look into these additional 8 genes.



Whole genes or critical regions to interrogate for NGS panel?



- Example gene: TTN – all observed mutations in COSMIC and Pediatric data sets fall on exon 46.
- Is it worth covering the whole gene in your NGS panel?

Scenario #2

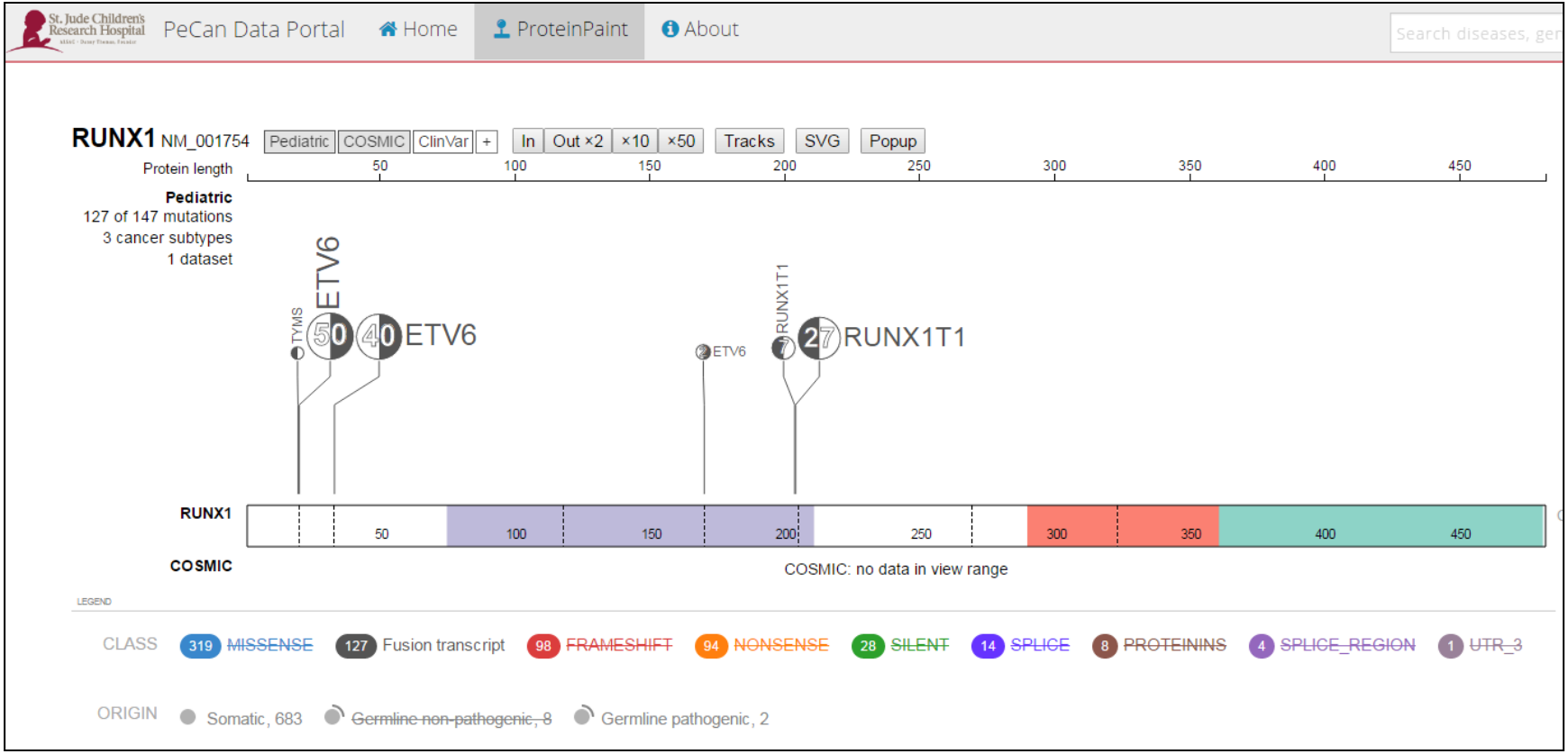
- You want to design PCR primers overlapping DNA binding domains (Zinc Finger domains) in IKZF1 – which exons should you target?
  - ProteinPaint

IKZF1

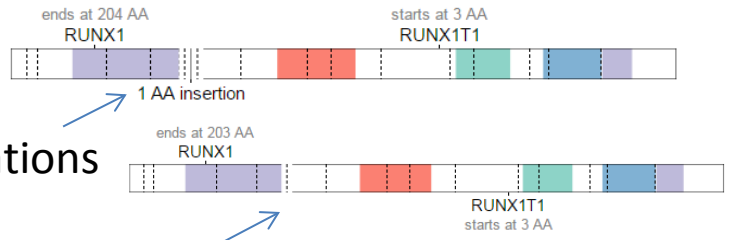
- Search for gene name IKZF1 in search bar in ProteinPaint or Home Page
- Manually curated ZnF domains in exons 4, 5, 6, and 8.
  - Hover over Protein to see exons – boundaries annotated by dashed vertical line.

Scenario #3

- You would like to design a D-FISH strategy targeting RUNX1-RUNX1T1 gene fusions
  - ProteinPaint

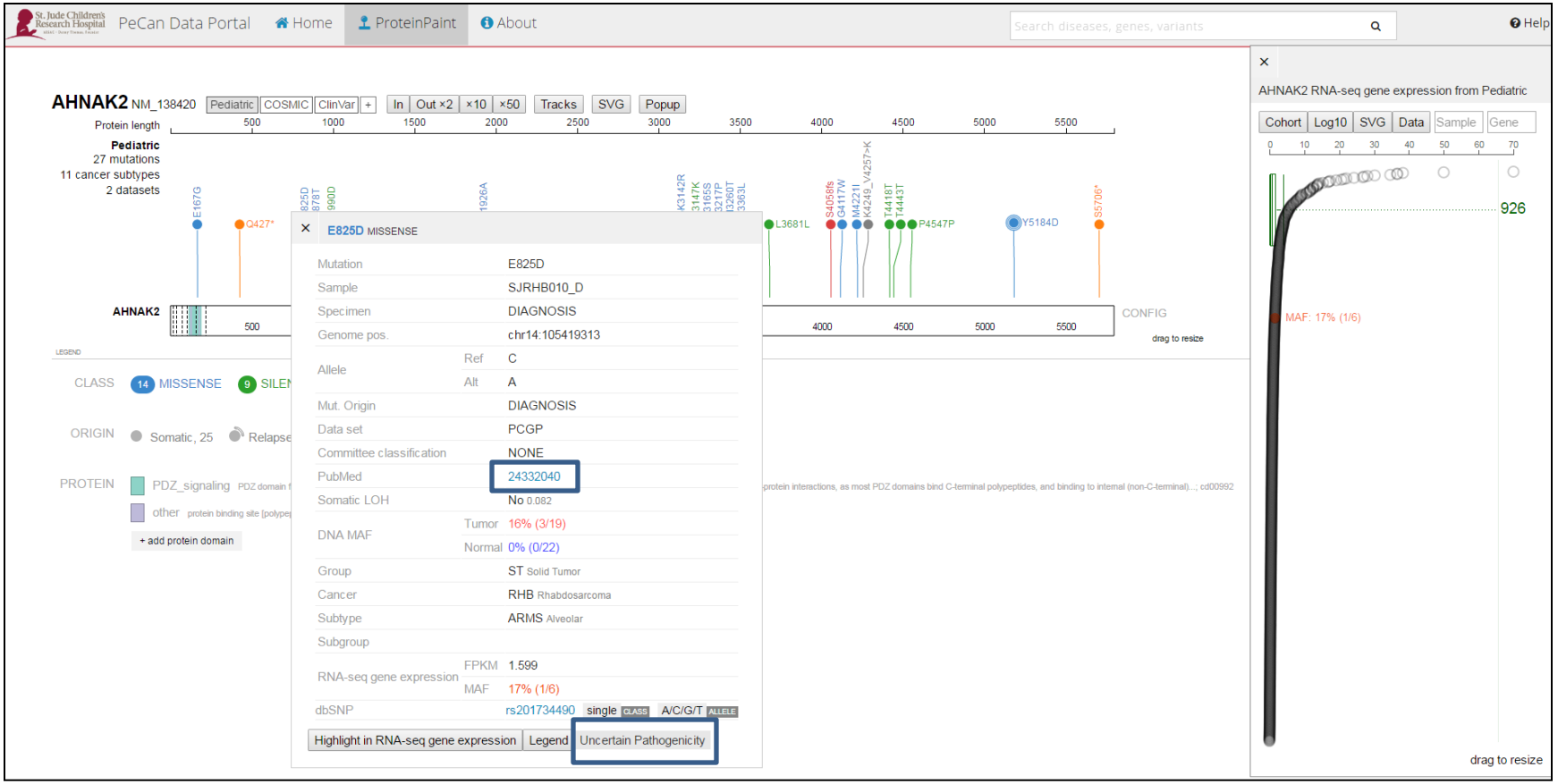


- Use Search bar to search for RUNX1
- Find most common RUNX1T1
- Filter by fusion transcript
- Select fusion transcript marker representing 27 mutations
- Within Fusion view, toggle to chimeric view
- Repeat for fusion transcript marker representing 7 mutations
- Use transcript information to guide probe placement



Scenario #4

- You are a laboratory consultant trying to interpret a pediatric neuroblastoma case with just one mutation identified, which your lab has not seen before.
  - AHNAK2 – E825D (missense)
    - ProteinPaint



- Potential clues:
  - PMID connected with variant
  - Manual curation interpretation
  - RNA-Seq data

Online Tutorials

- Tutorial Document: https://docs.google.com/document/d/1JWKq3ScW62GlSFGuJvAajXchcRenZ3HAvpaxlLeGaw0/edit
- Video: https://www.youtube.com/watch?v=UDSoIs-2ZfU

Contacts

- Xin Zhou: Xin.Zhou@STJUDE.org
- Jinghui Zhang: Jinghui.Zhang@STJUDE.org

  – https://pecan.stjude.org/pp

User community on Google+ to learn about latest development and provide feedback: https://pecan.stjude.org/community