

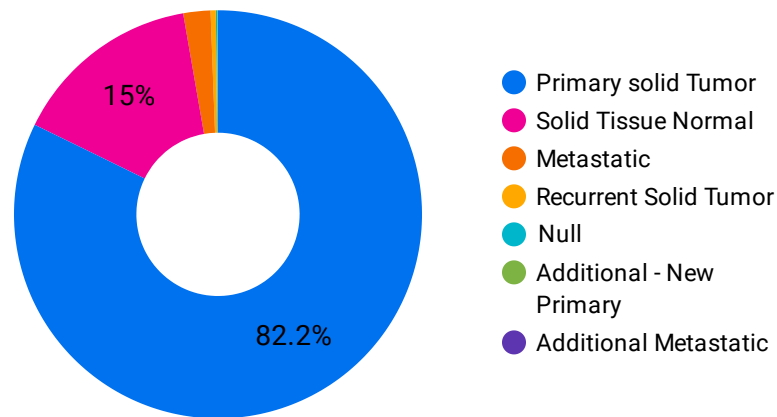
TCGA Dataset Overview

Metadata generated by joining *metadata.TCGA_slide_images*, *TCGA_bioclin_v0.clinical_v1*, *TCGA_bioclin_v0.Biospecimen*

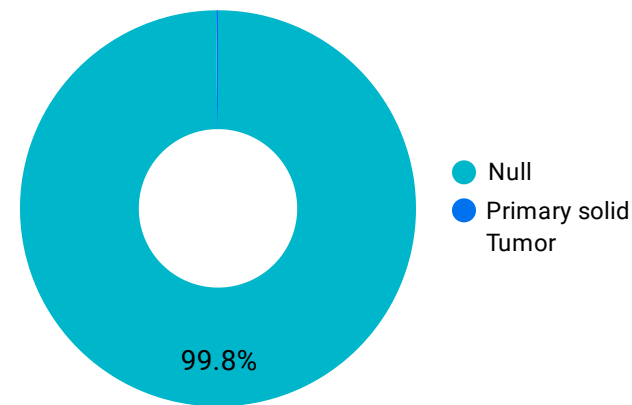
1. Summary and General Stats

# of slides	FFPE/Frozen	# of cohorts	# of sample types
30,081	11,963/18,304	32	6

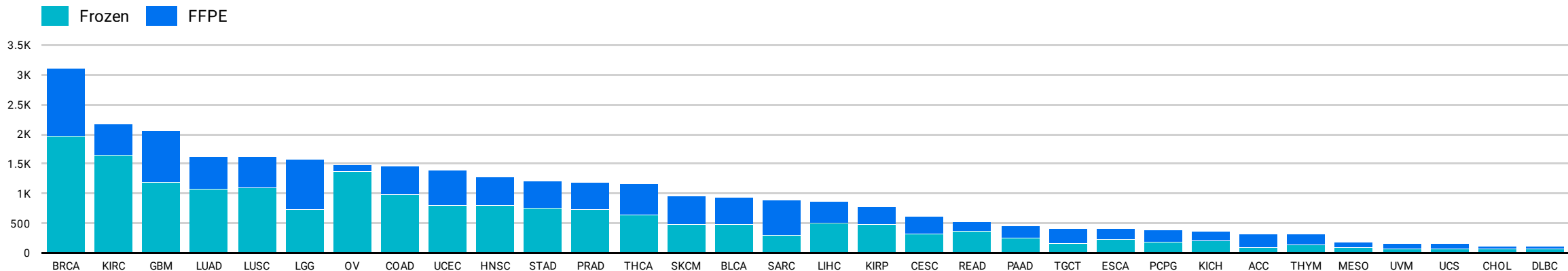
Sample Type Distribution: Frozen



Sample Type Distribution: FFPE



FFPE/Frozen Slides by Cohort



TCGA Dataset Overview

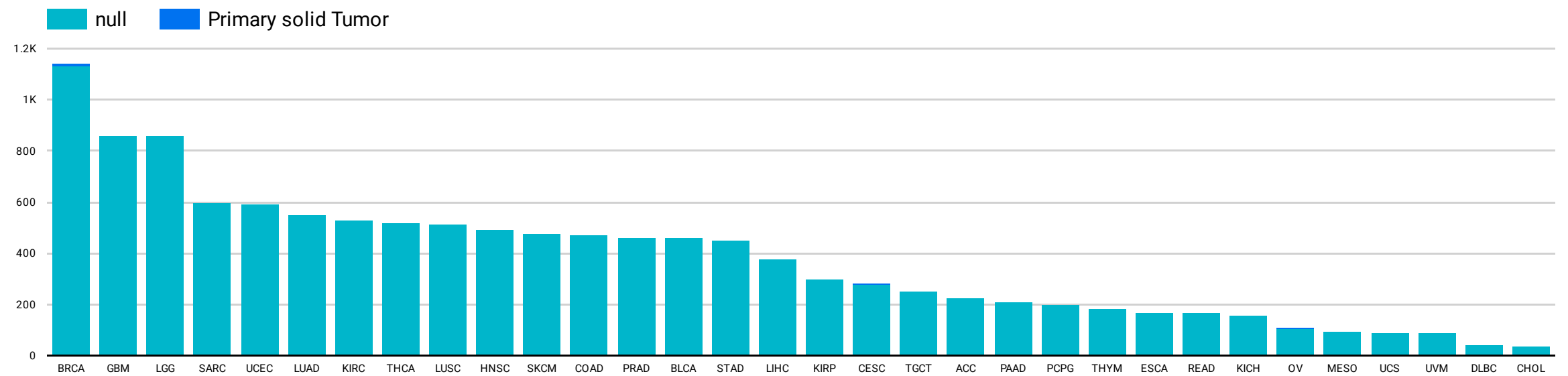
Metadata generated by joining *metadata.TCGA_slide_images*, *TCGA_bioclin_v0.clinical_v1*, *TCGA_bioclin_v0.Biospecimen*

2. FFPE Samples

disease_code ▾

avg_percent_tumor_cells ▾

Sample Type Name by Cohort



TCGA Dataset Overview

Metadata generated by joining `metadata.TCGA_slide_images`, `TCGA_bioclin_v0.clinical_v1`, `TCGA_bioclin_v0.Biospecimen`

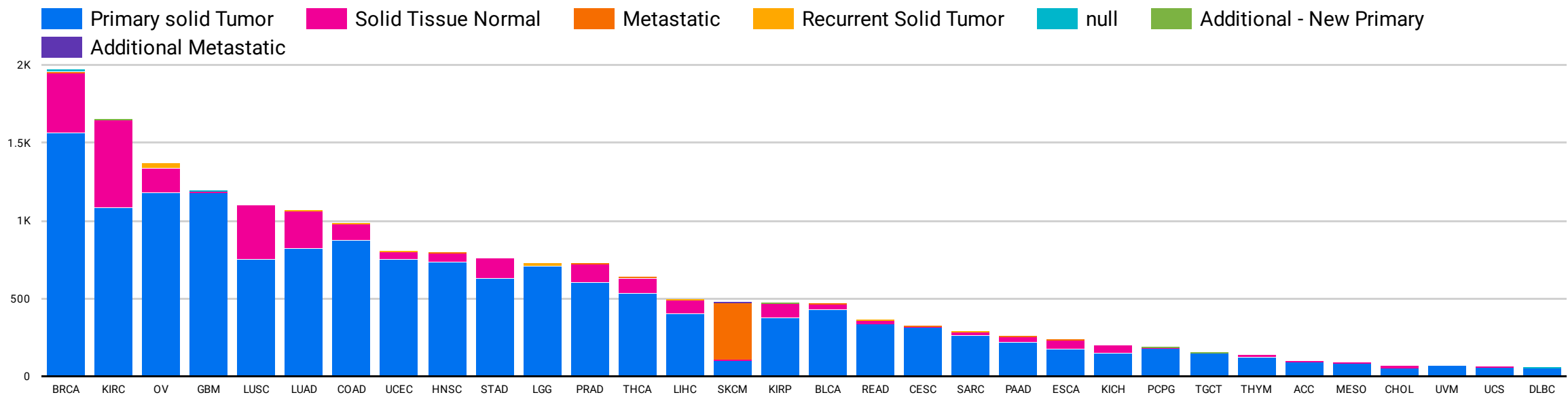
3. Frozen Samples

See cumulative distribution of purity in JupyterLab.
(https://github.com/soltaniehha/Deep-Nexus/blob/master/sandbox/sandbox-olivia/TCGA_Overview.ipynb)

Sample Type Name by Cohort

disease_code

avg_percent_tumor_cells

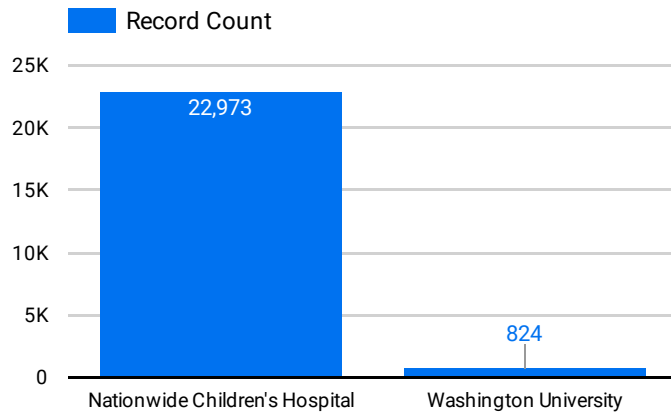


TCGA Dataset Overview

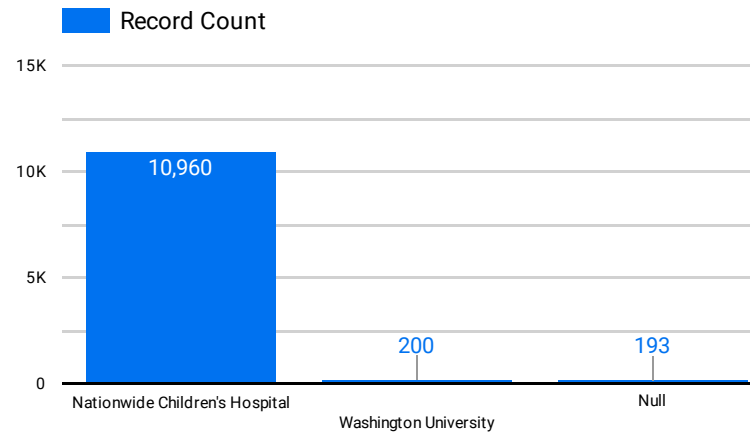
Metadata generated by joining *metadata.TCGA_slide_images*, *TCGA_bioclin_v0.clinical_v1*, *TCGA_bioclin_v0.Biospecimen*

4. Data Source

Biospecimen



clinical_v1

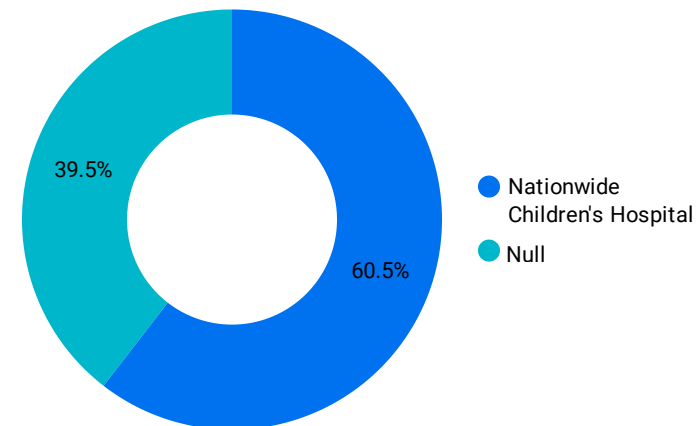
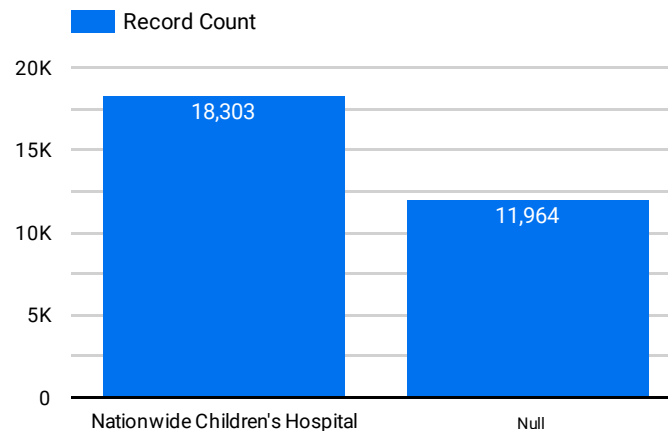


TCGA_slide_images

NO INFO
 $\circ(\pi \rightsquigarrow \pi)\circ$

↓ join

TCGA_slide_metadata



TCGA Dataset Overview

Metadata generated by joining `metadata.TCGA_slide_images`, `TCGA_bioclin_v0.clinical_v1`, `TCGA_bioclin_v0.Biospecimen`

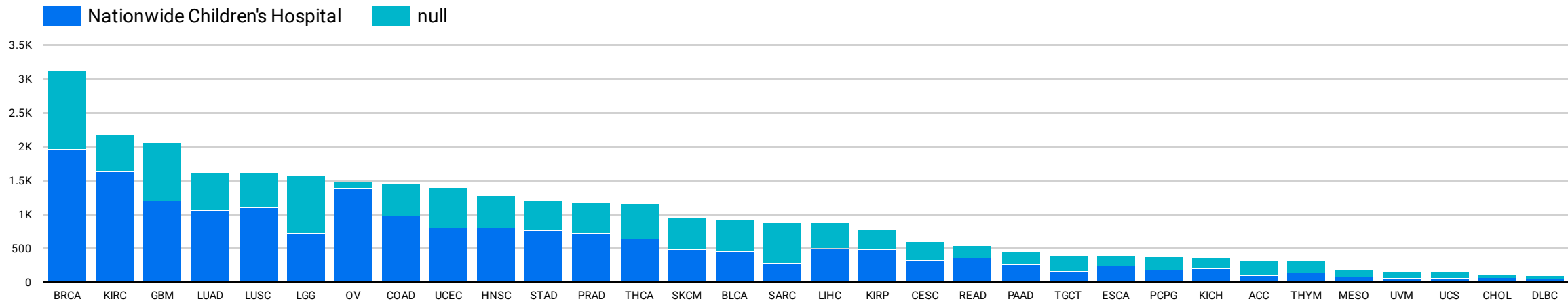
4. Data Source (cont'd)

disease_code

Data Resource by Cohort

total # of slides

30,267



Source: Washington University, Nationwide Children's Hospital

Sample Type by Cohort, with Known Source (NCH)

total # of slides

18,303

