

Human COVID19 factors Beaconv2.0: exploring metadata available

July 20, 2020

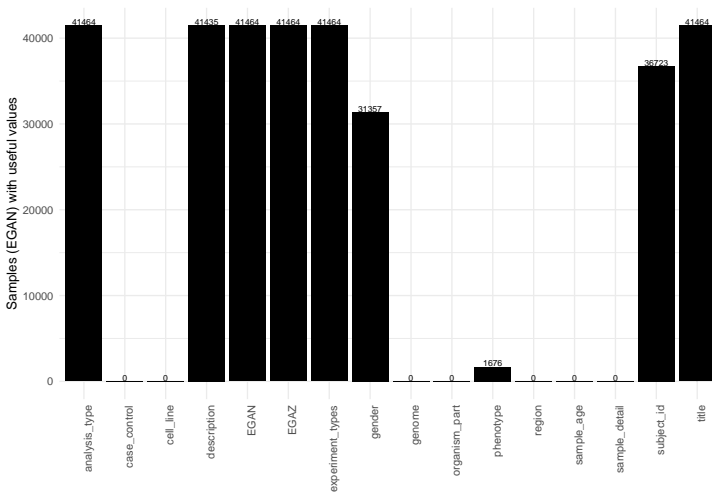
Metadata available for 84K vcf files

Unique values:

- EGAF: 75286
- EGAN: 41464
- subject_id: 35824
- title: 3466
- description: 3388
- phenotype (after rm NA and "N/A", "Not applicable", "not applicable", "Not provided", "None", "unknown", "Proband", "Parent"): 416 unique

"Non cancer", "Cancer breast", "Cancer lung", "Cancer prostate",
"cancer" , "HP:0001263" (411 HP ontology ids)

Metadata available for 84K vcf files



Metadata available for 84K vcf files

Label	Count
Global developmental delay	212
Cancer prostate	108
Cancer lung	98
Cancer breast	96
Non cancer	94
Microcephaly	66
Intellectual disability	54
Seizure	44
Specific learning disability	44
Autism	36

Table: Top most common phenotype values

Metadata available for 84K vcf files

Pending:

- Check custom tags on all XMLs for phenotypes