

# OMOPOmics

## Day 2 update

Nick Giangreco, Andrew Clugton, Alex Francette,  
Anish Chakka, Yash Gokhale

# An OMOP extension for patient molecular data

OMOP is a standardized infrastructure for promoting reproducible analysis

Data are encoded as concepts with source values

Data are related with primary and foreign keys

Data tables are encoded in a (SQL) database

Data are extracted with reproducible (SQL) queries

Cohort data are input to custom analysis pipelines

**SPECIMEN\_TABLE**

Column	Name
disease_status_source_value	healthy
treatment_source_value	T cell activation
cell_type_source_value	CD4+ T cells
person_id	P1
disease_status_source_value	healthy

**PERSON\_TABLE**

Column	Name
person_id	P1
gender_concept_id	8507
person_source_value	GSM1484803
gender_source_value	male

**SOURCE\_PROVIDER\_TABLE**

Column	Name
source_id	S1
source_name	GEO
source_value	GSE60682
person_id	P1

**NGS\_TABLE**

Column	Name
fastq	0
bam	0
bed	/pylon5/brz3a1p/codeathon/Chroma_T-Cell/Data/GSE60682/GSM1484803_Donor1_TCA4hrs_Rep2.bed.gz
qc	0
reference_genome_source_value	hg19
sequencing_device_source_value	Illumina HiSeq 2000
read_type_source_value	PAIRED
read_length	0
seq_type_source_value	ATAC-Seq
assay_id	X1

**MOLECULAR\_TEST\_TABLE**

Column	Name
species_name	Homo sapiens
species_ID	40526985
assay_type_source_value	NGS
SOURCE_ID	S1
assay_type_name	ATAC_seq
assay_type_ID	A1
assay_id	X1

# Reproducible queries for cohort data creation

Query:

```
Select file_path  
From file where  
disease == CTCL  
Assay == ATAC-seq
```



Pass file paths into automated  
Analysis script



Generate .txt file w/ list of matching file paths

```
GSM1484807_Donor11_CTCL_Rep1.bed.gz  
GSM1484808_Donor11_CTCL_Rep2.bed.gz  
GSM1484809_Donor11_CTCL_Rep3.bed.gz  
GSM1484810_Donor11_CTCL_Rep4.bed.gz  
GSM1484811_Donor11_CTCL_Rep5.bed.gz  
GSM1484812_Donor11_CTCL_Rep6.bed.gz  
GSM1484813_Donor11_CTCL_Rep7.bed.gz  
GSM1484817_Donor20_CTCL_Rep1.bed.gz  
GSM1484818_Donor20_CTCL_Rep2.bed.gz  
GSM1484819_Donor20_CTCL_Rep3.bed.gz  
GSM1484820_Donor20_CTCL_Rep4.bed.gz  
GSM1484831_Donor5_CTCL_Rep1.bed.gz  
GSM1484832_Donor5_CTCL_Rep2.bed.gz  
GSM1484833_Donor5_CTCL_Rep3.bed.gz  
GSM1484834_Donor5_CTCL_Rep4.bed.gz
```

# Analysis of cohort data by custom pipelines