Utilizing Commercial Real-World Data: Truveta

Barnholtz-Sloan Research Team

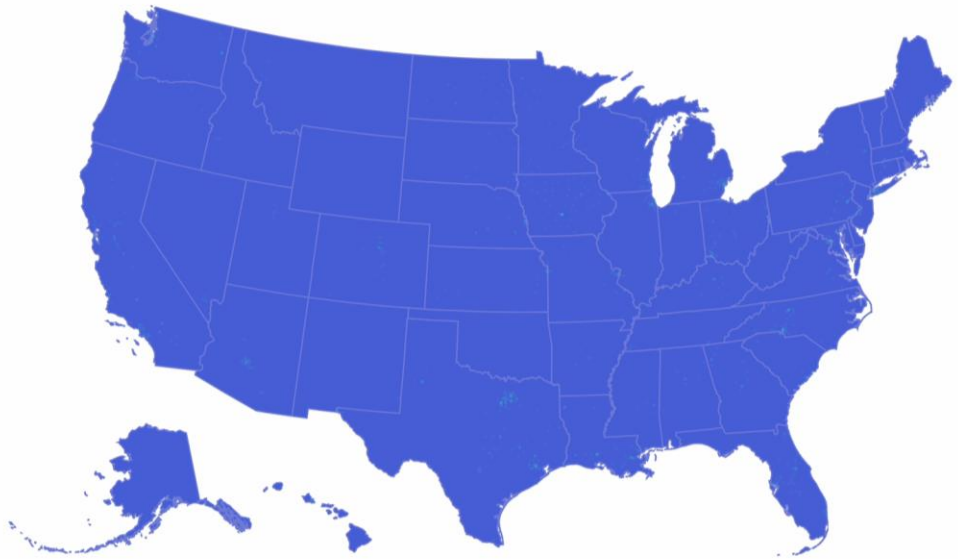
Jill Barnholtz-Sloan, PhD

Kristin Waite, PhD

Gino Cioffi, MPH

Truveta - Data Overview

Truveta healthcare members

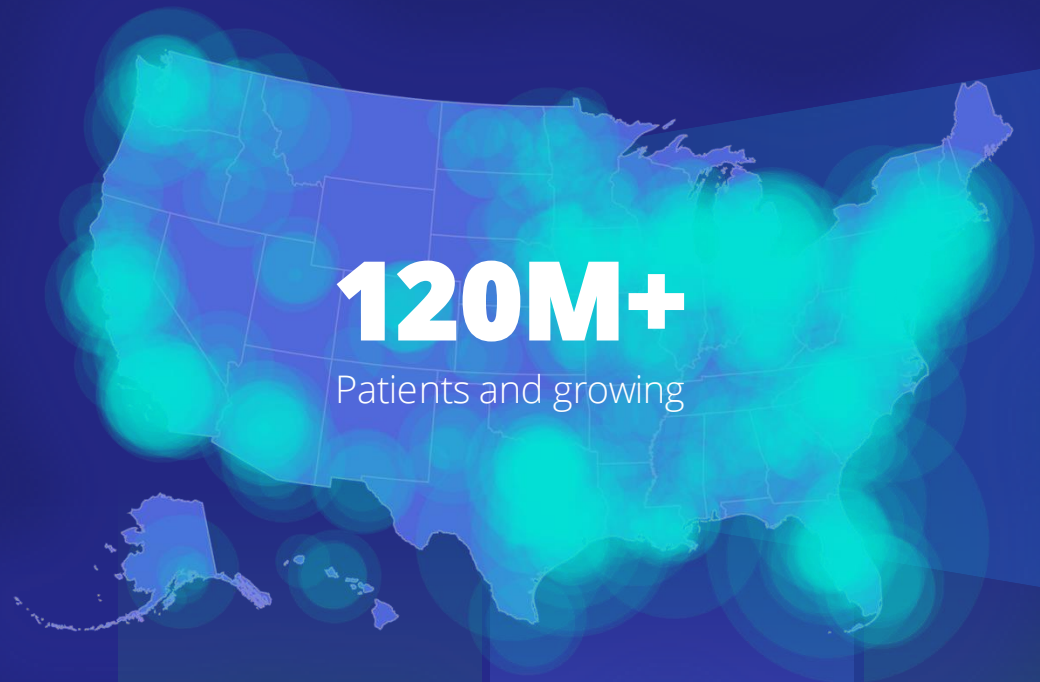


30 of the largest health systems in the US

1 out of 3 Americans represented in Truveta Data

Slide courtesy of Truveta

Truveta Data: Complete, timely, and clean regulatory-grade EHR data



Semi-structured data

- Diagnosis (SNOMED, ICD)
- Procedure (CPT,HCPCS)
- Medication (RXNORM,NDC)
- Labs (LOINC,UCUM)
- Immunizations (CVX)
- Implanted/explanted Device (UDI)
- Inpatient/outpatient care settings
- Pharmacy
- Mother-child link

Unstructured data

- Clinical notes
- Images
- Genomics

Tokenization and integration with third-party datasets

5B+	85M+	1M+	5+	200M	188K+
Clinical notes	Imaging studies	Mother-child pairs	Years longitudinal patient history	Closed Claims Patients	Unique devices

This data is additionally linked with SDOH, mortality, and claims data

Slide courtesy of Truveta

Data from 5B+ clinical notes available for research

Heart Failure - NYHA

Patient Active Problem List Diagnosis • Appendicitis, acute • DM (diabetes mellitus) • Facial cellulitis • Hyponatremia • Essential hypertension • Dilated cardiomyopathy • Pulmonary embolism without acute cor pulmonale • Acute deep vein thrombosis (DVT) of popliteal vein • LBBB (left bundle branch block) • Chronic combined systolic and diastolic CHF, NYHA class 2 and ACA/AHA stage C • Lightheadedness • Suspected sleep apnea • Congestive heart failure • Snoring

History of Present Illness: Information is gathered from patient … with a past medical history of hypertension, hyperlipidemia, cardiomyopathy, history of left lower extremity DVT, diabetes, history of left bundle branch block, pulmonary embolism, combined systolic and diastolic heart failure with EF of 17% at diagnosis

ECHO Complete Result Date:

1. Mildly dilated left ventricle with normal thickness, LVEDD 5.9 cm global hypokinesis with mildly reduced calculated LVEF 42% which on visual comparison is similar to the images
2. The RV is not seen well but the size appears normal with normal lateral annular motion and TAPSE suggestive of normal RV SF
3. Normal biatrial sizes. Intracardiac lead seen intermittently in the right atrium
4. The mitral and aortic valves are seen adequately without echodensities suggestive of vegetation. The tricuspid valve is not seen completely but there are no large echodensities visualized. There is trace tricuspid and mitral valve regurgitations
5. The aortic root and feels normal
6. There is a trace pericardial effusion
7. Normal CVP, normal left-sided filling pressure
8. Sinus 70s ECHO Transesophageal (TEE)

Patient Id	Note Date	Clinical Measure	Value
PT1234	2023-01-22	NYHA Class	2
….	….	…..	….
….	….	….	….

Patient Id	Note Date	Clinical Measure	Value
PT1234	2023-01-22	LVEF	42%
….	….	…..	….
….	….	….	….

Examples of concepts extracted from notes:

- Ejection fraction
- Seizure frequency
- Cancer stage
- Heart failure stage and symptoms
- Cardiac catheterization measures
- Metabolic disease progression

Slide courtesy of Truveta

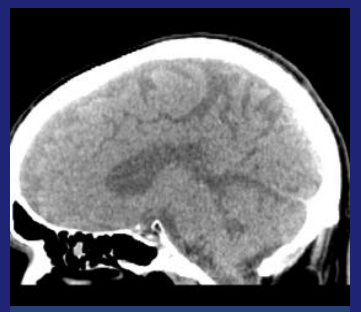
More than 1M matched mother-child pairs

Longitudinal EHR data for more than 1 million mother-child pairs, enabling research from pre-pregnancy through the first 5 years of the child's life

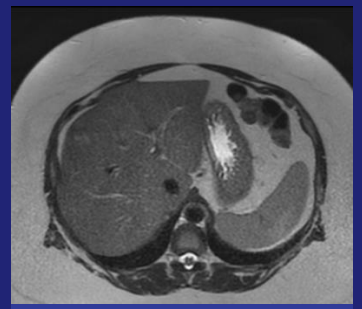
Slide courtesy of Truveta

Images

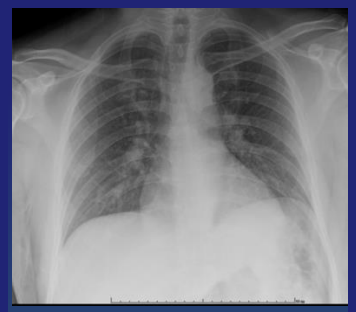
Largest collection of medical images integrated with EHR data



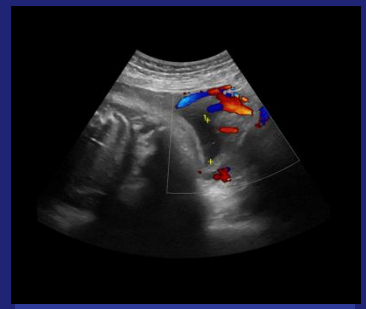
16.5M
Computed tomography



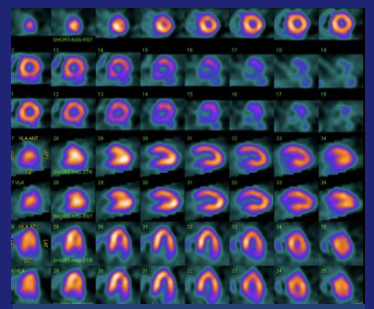
5.2M
Magnetic resonance imaging



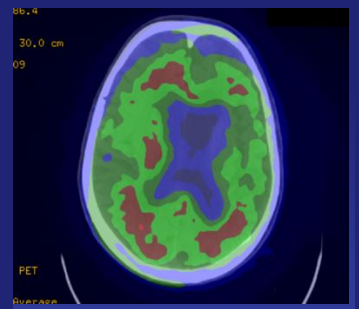
38.7M
Digital x-ray



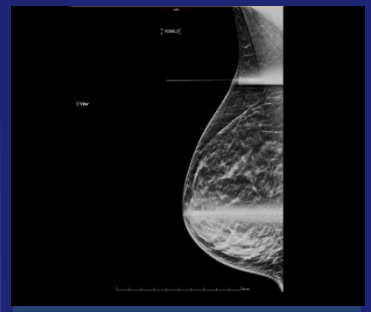
13.9M
Ultrasound



500K
Nuclear medicine



228K
PET imaging



10.1M
Mammography

More than 85 million imaging studies searchable by modality and protocol

Slide courtesy of Truveta

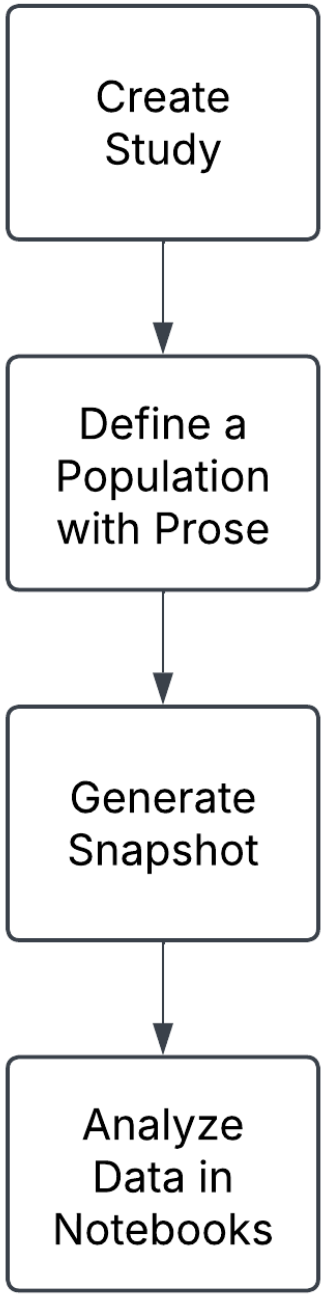
Types of Studies Reported

- Treatment analysis
- Mother - Child studies (pregnancy to 5 yr old)
- Adherence studies
- COVID analysis
- Prescription trends and monitoring
- Testing/screening trends
- Injury analysis
- Office, ER visit analysis
- Monitoring reports

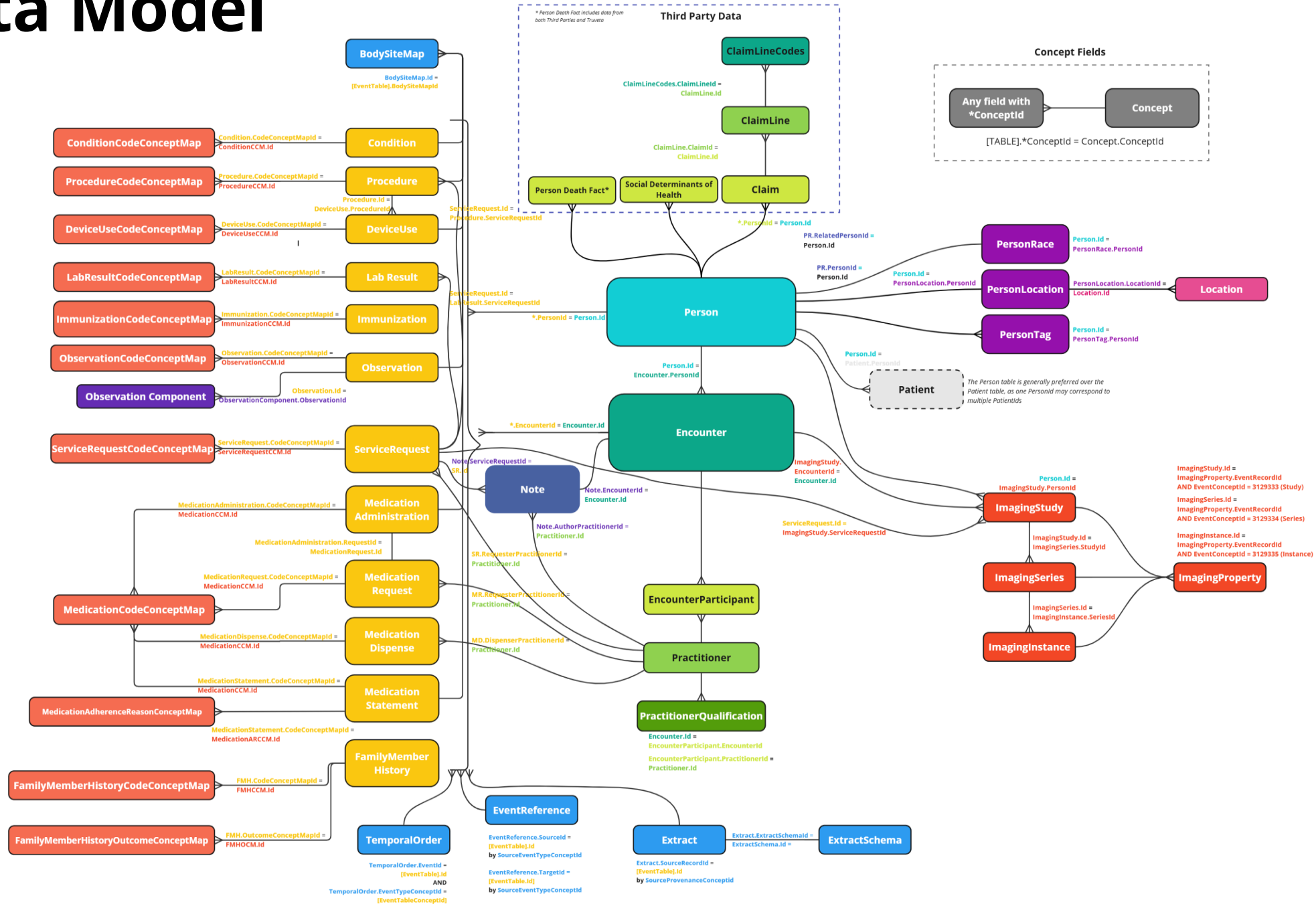
Truveta - User Overview

Truveta Studio

- Web interface for engaging with data
  - All cloud-based, accessed through browser
- “Studies” contain data definitions, snapshots, and notebooks for a given project

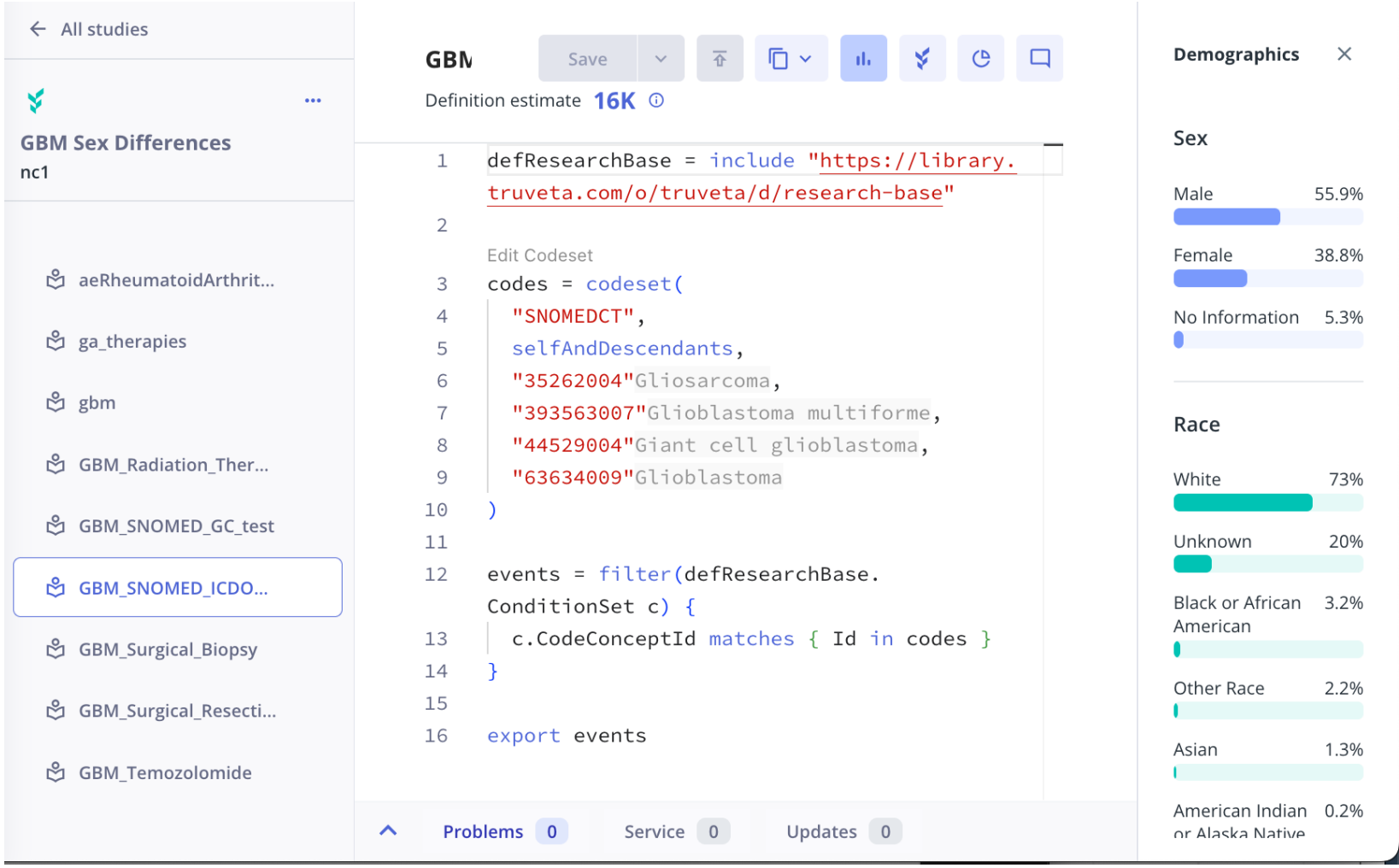


Truveta Data Model



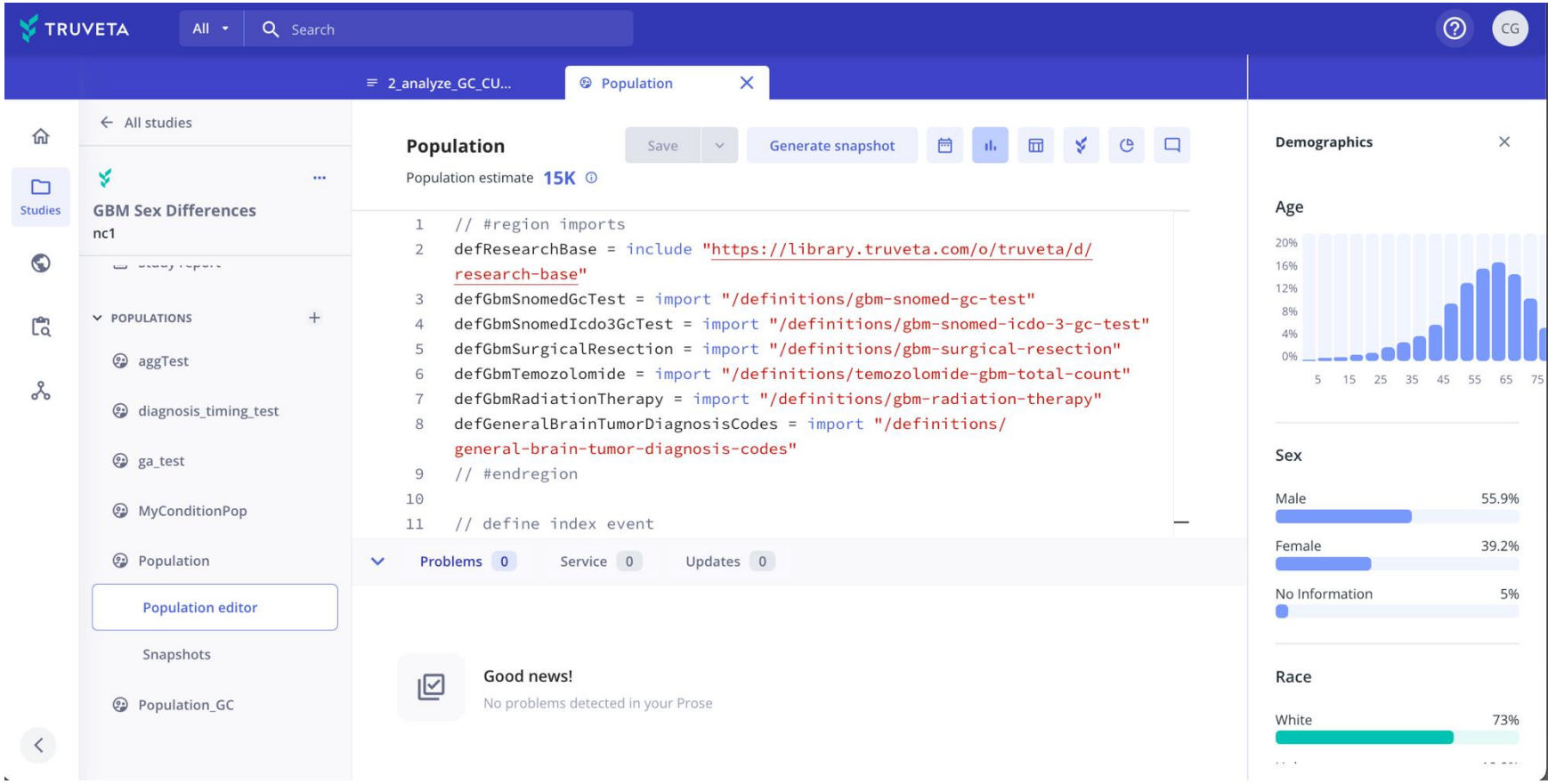
Creating Data Definitions

- Prose
  - Truveta query language
- Event and Codesets
  - Provides broad overview of conditions, procedures, or treatments
- Code identification and mapping tools
  - Truveta Studio Concept Search tool
  - ATHENA



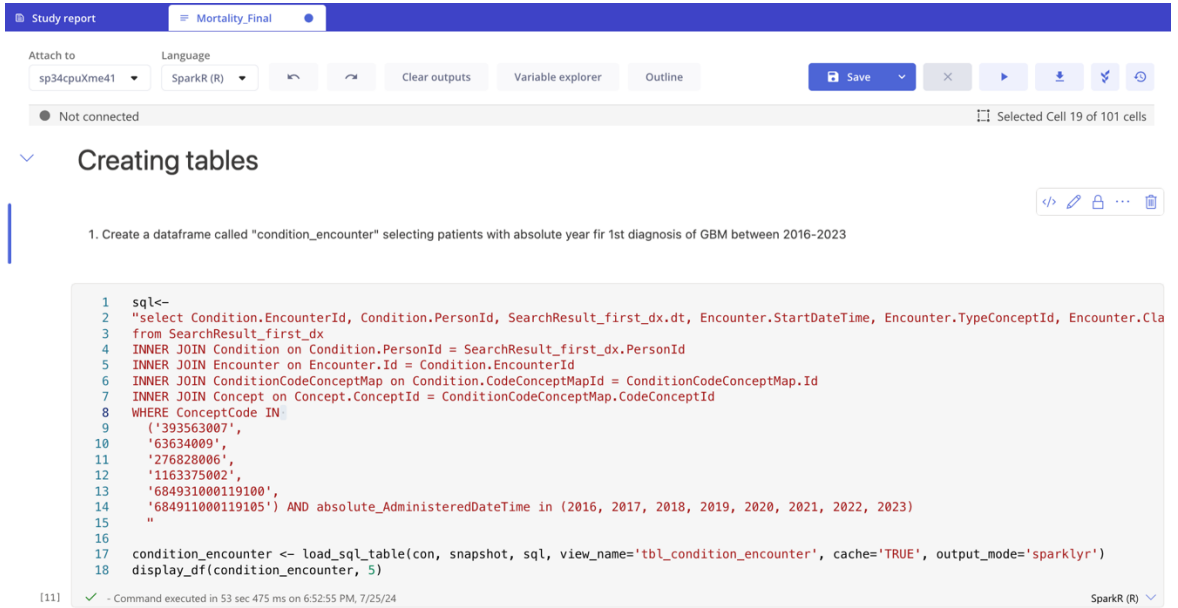
Defining a Study Population

- Use Prose Definitions to define population
- Generate “Snapshot”
  - Point-in-time collection of records
  - To be analyzed in notebook environment
- Select Tables to be included
  - Remove unnecessary tables to reduce cloud costs



Accessing Data

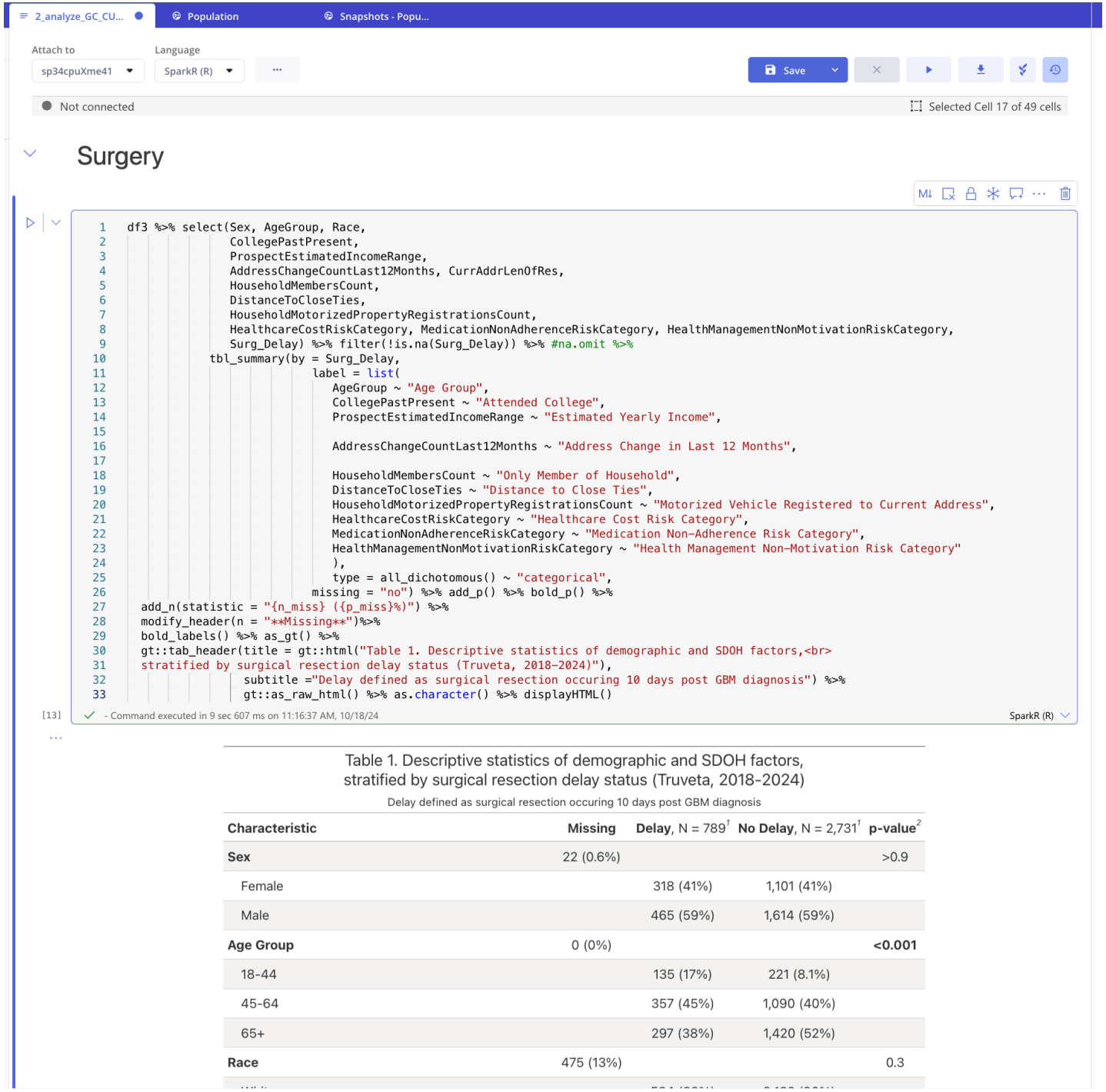
- Spark-based integrated notebook environment
  - PySpark (Python)
  - SparkR (R)
- Data Wrangling
  - SQL
  - Koalas (Python)
  - Sparklyr (R)



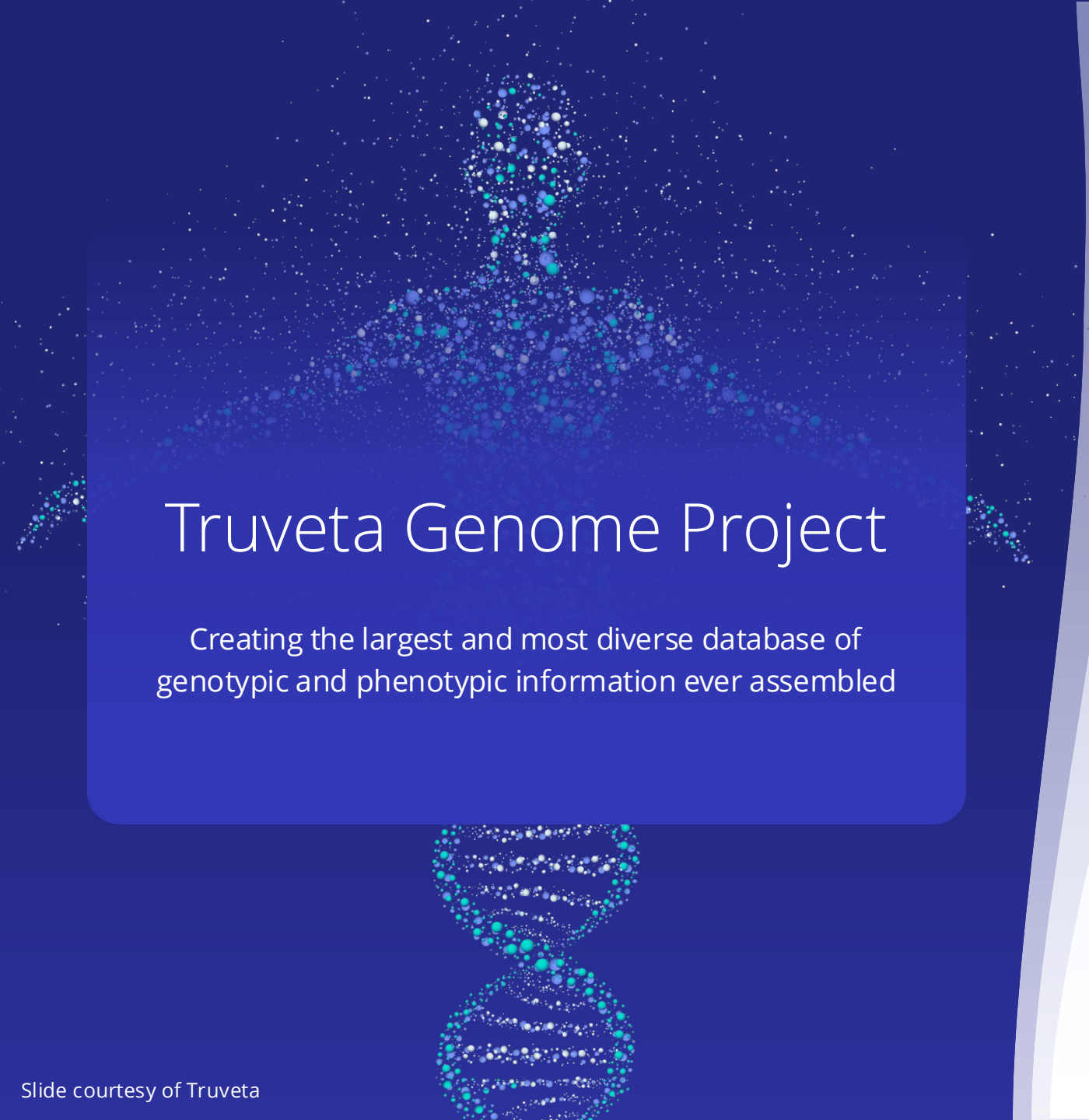
Analyzing Data & Exporting Results

- Analytical dataset can be analyzed using regular R/Python syntax

- Exporting Results
  - Can export figures (png, jpeg, etc.) and tables (html)
  - Can **not** export data files



Future Applications



Truveta Genome Project

Creating the largest and most diverse database of genotypic and phenotypic information ever assembled

- Delivering genetic data on 10s of millions, over 10x the scale of previous endeavors
- Creating the world's largest genetic biobank to support future multi-omics
- Enabling drug discovery and optimized clinical trials
- Transforming how diseases are prevented, diagnosed, and cured

Slide courtesy of Truveta

Truveta Genome Project: Learning the biology of disease

Healthcare member

Patient consents to leftover biospecimens being used for research

Leftover biospecimens after lab test shipped to RGC

Regeneron Genetics Center

RGC

Sequences DNA and banks leftover for future ‘omics

Truveta

Enables research by healthcare, public health, and life science

Slide courtesy of Truveta

Questions

- For feasibility questions, please reach out to Kasie Bailey at Truveta
  - Email: kasieb@truveta.com