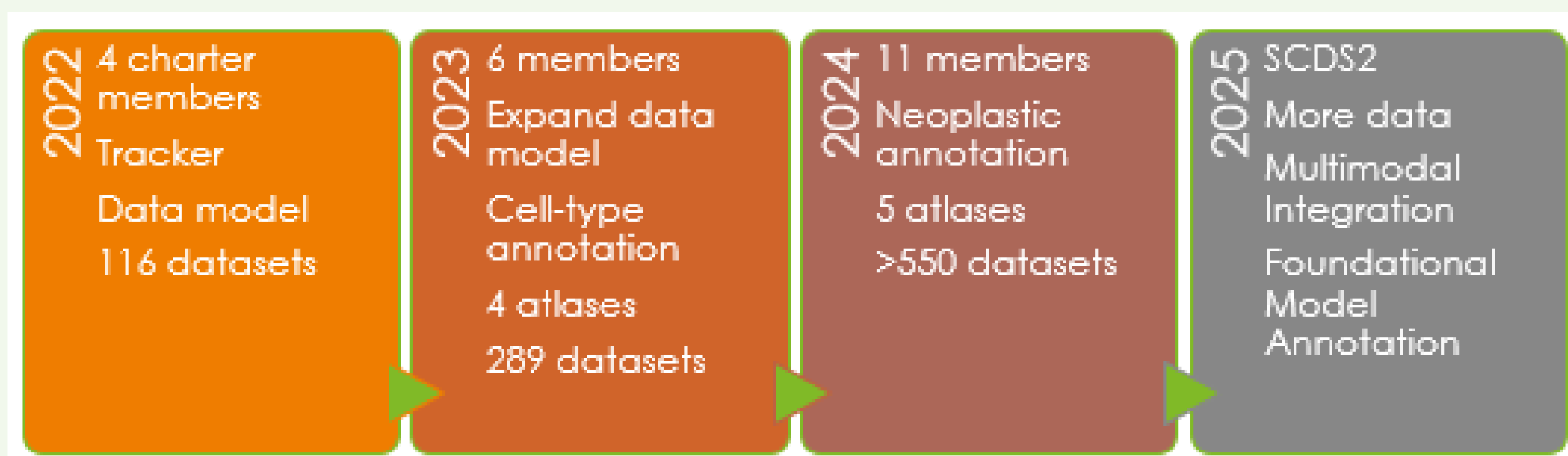
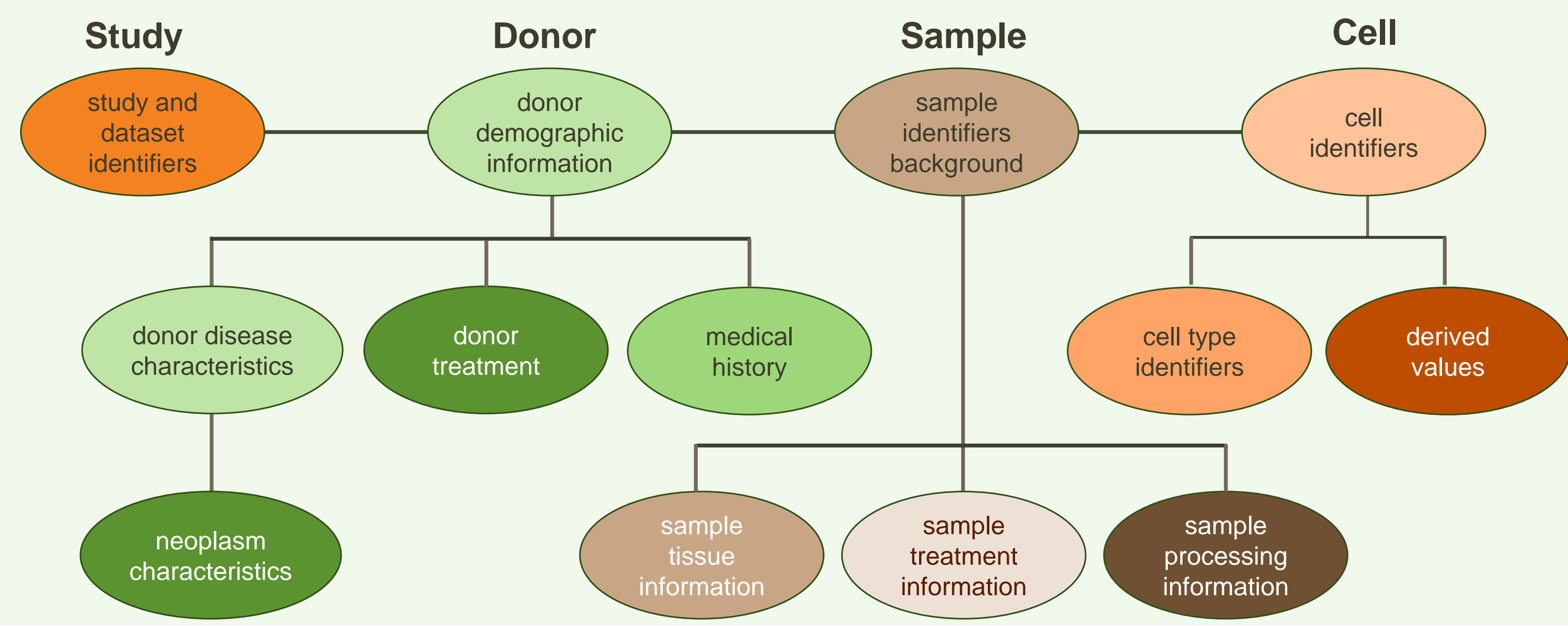


Background

Rancho Biosciences has successfully finished the first round of its 3-year Single-Cell Data Science Consortium (SCDC). This pre-competitive effort grew from 4 charter members to 11 members over the course of **3 years**. The deliverables cover more than 800 publications and 1,000 individual datasets, including recent landmark papers (e.g., Siletti et al, 2022, Tabula Sapiens Consortium et al. 2022 and Tabula Muris Consortium et al., 2020). The ready-to-use deliverables include processed expression data and metadata harmonized across datasets at the study, donor, sample and cell levels. Our dedicated **SCDS data integrity pipeline** combined the best of Rancho accelerators and manual curation by subject matter experts to align metadata to attribute-specific relevant public ontologies, including DOID, UBERON, CL, MeSH, EFO and a custom vocabulary.

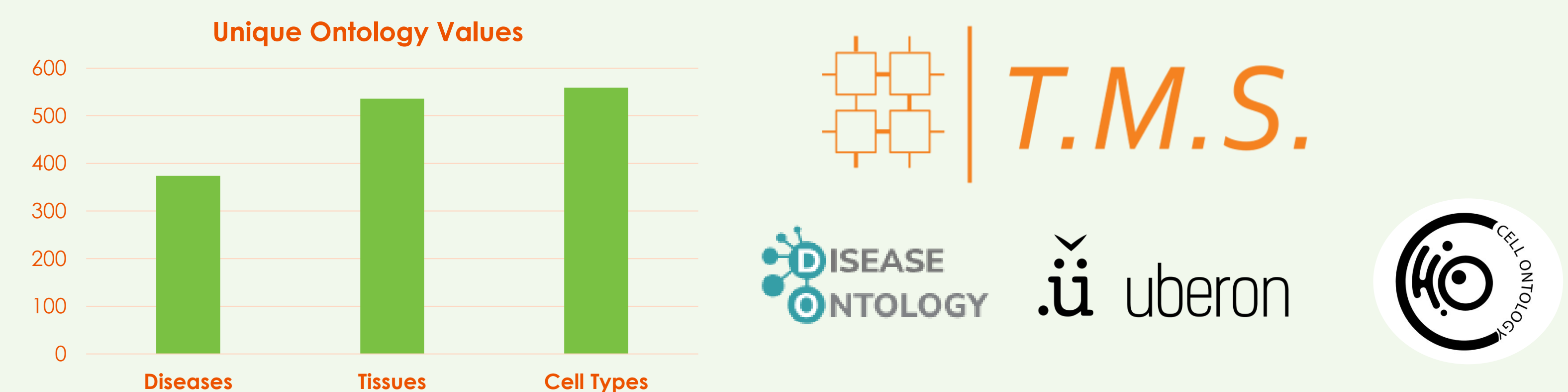


3. SCDS Data Model



We developed a custom 4 entity, 83 attribute data model that is flexible enough to be used for a wide range of experimental purposes

4. Metadata Harmonization



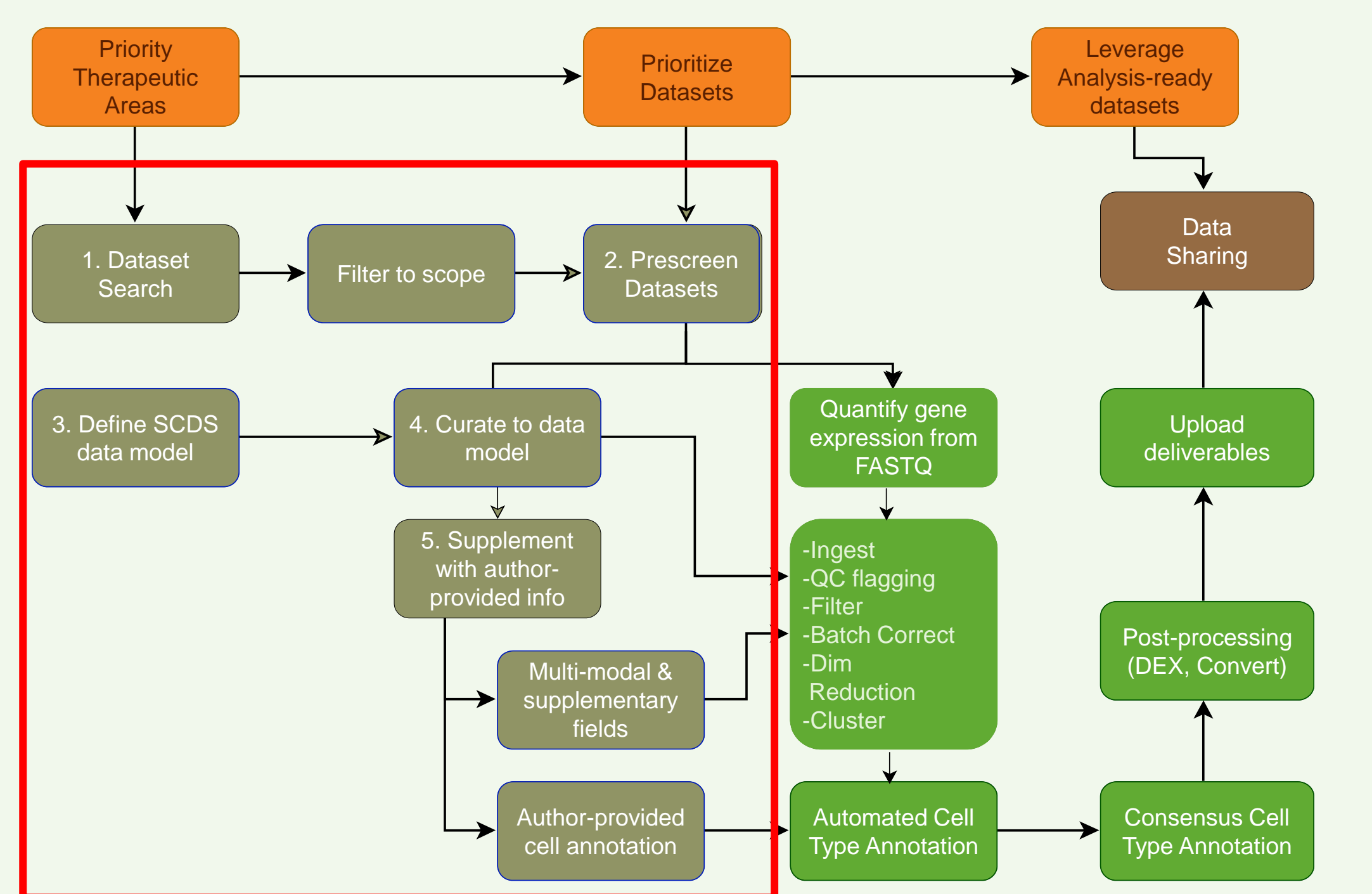
- Metadata is aligned to attribute-specific relevant public ontologies, including DOID, UBERON, CL, MeSH, EFO and a custom vocabulary using Rancho accelerator, Terminology Management Solution
- In total, 559 unique cell types, 374 unique diseases and 536 unique tissue values were harmonized to allow for intra- and inter-dataset comparisons
- Strict adherence to the data model and accompanying dictionaries reduced differences across batches and curators.

Data Ingestion Workflow

Members
 • Define Priorities

Data Integrity
 • Maintain catalog
 • Prescreen datasets
 • Harmonize sample-level metadata

Bioinformatics
 • Provide analysis-ready datasets

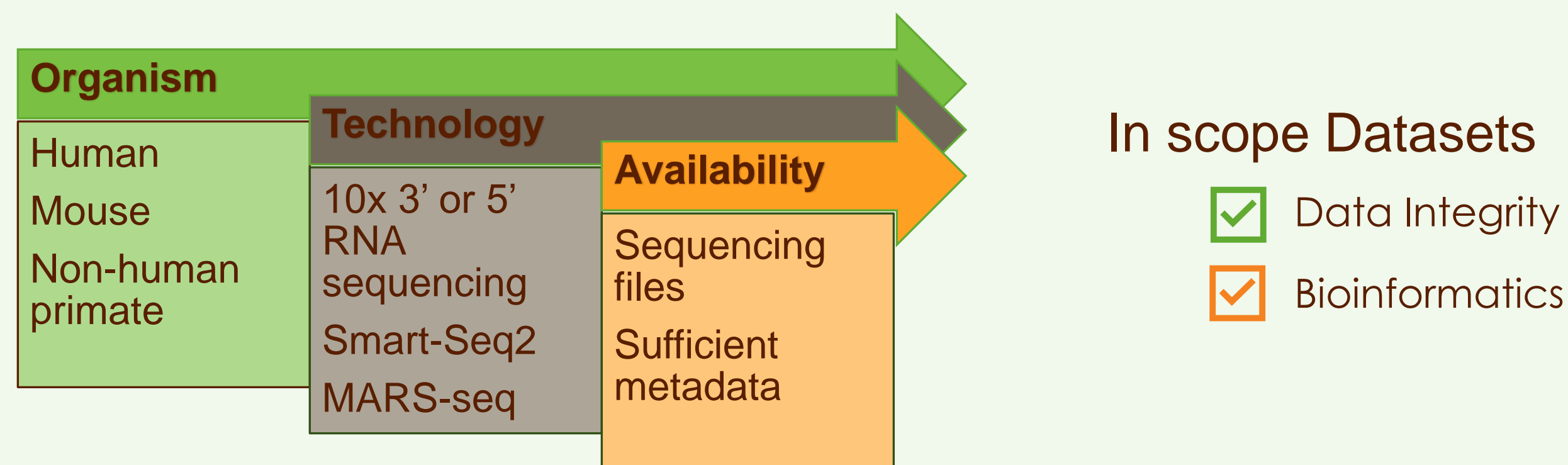


1. Building a Catalog

id	shortname	title	authors	journal	pmid	abstract	organism	tissue	disease	number	cell	technique	repo	ident	processed	d	in_scope
1	Absarin_2021_Naiv	Single-Cell Transcriptomics of Human T Cells	Martina Absarin	Nature	34497421	Multiple scRNA-seq datasets from human T cells provide insights into the diversity of T cell subsets and their developmental trajectories.	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
2	Fernandes_2020_C	Single-Cell Transcriptomics of Human T Cells	Luca R. Fernandes	Cell	33053338	The advent of single-cell transcriptomics has revealed the heterogeneity of T cells and their developmental trajectories.	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
3	Pappalardo_2020	Transcriptomic Profiling of Human T Cells	Jenna L. Pappalardo	Immunity	32948672	T cells provide human immunity through their ability to kill infected cells and tumor cells.	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
4	Agarwal_2020_Naiv	Single-Cell Transcriptomics of Human T Cells	Divyanshu Agarwal	Nat Commun	32260993	We describe the transcriptomic heterogeneity of naive T cells and their developmental trajectories.	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
5	Schifflick_2020_N	Integrated Single-Cell Transcriptomics of Human T Cells	David Schifflick	Nat Commun	31957773	Neoplasms	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
6	Schirmer_2019_Na	Neuronal and Immune Cell Transcriptomics	Lucas Schirmer	Nature	31316211	Multiple scRNA-seq datasets from human T cells provide insights into the diversity of T cell subsets and their developmental trajectories.	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
7	Hong_2021_Front	Single-Cell Transcriptomics of Human T Cells	Front Immunol	32603736	Objective: To describe the transcriptomic heterogeneity of naive T cells and their developmental trajectories.	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT					in scope
8	Caron_2020_Sci	Single-Cell Transcriptomics of Human T Cells	Maxime Caron	Sci Rep	32415257	Childhood acute lymphoblastic leukemia (ALL) is a common childhood cancer.	Human	bone marrow	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
9	Allaverdi_2020_Na	Distinct Single-Cell Transcriptomics of Human T Cells	Stefano Allaverdi	Nat Med	32611335	Immune-regulation in human T cells is controlled by a complex network of transcription factors.	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
10	Kim_2020_Nat	Metabolic Regulation of Human T Cell Development	Doyoung Kim	Nat Med	31959590	Drug-induced T cell exhaustion is a major cause of T cell dysfunction in cancer and chronic infections.	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
11	Agarwal_2021_Ce	Comprehensive Single-Cell Transcriptomics of Human T Cells	Divyanshu Agarwal	Genome Res	32687696	Improved understanding of T cell development and function through single-cell transcriptomics.	Human	blood	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
12	Bondoc_2021_Co	Comprehensive Single-Cell Transcriptomics of Human T Cells	Alexander Bondoc	Commun Bio	34497364	Hepatoblastoma is a rare childhood liver cancer.	Human	liver	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
13	Sahu_2021_Oncog	Single-Cell Transcriptomics of Human T Cells	Biswajit Sahu	Oncogene	34302118	Cancer is the leading cause of death worldwide.	Human	colon	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
14	Jansky_2021_Naiv	Single-Cell Transcriptomics of Human T Cells	Jens Jansky	Nat Genet	32176450	Neuroblastoma is a common childhood cancer.	Human	adrenal gland	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
15	Mercantelli_2021	Single-Cell Transcriptomics of Human T Cells	Giuseppe Mercantelli	Commun Bio	33252507	Neuroblastoma is a common childhood cancer.	Human	adrenal gland	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
16	Luoma_2020_Cell	Molecular Profiling of Human T Cells	Adrienne M. Luoma	Cell	32603604	Checkpoints in human T cell development and function.	Human	colon	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
17	Ji_2020_Cell	Multi-modal Profiling of Human T Cells	Andrew L. Ji	Cell	32579074	To define the human T cell transcriptome and its relationship to the genome.	Human	intestine	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
18	Lu_2020_Cell	Comprehensive Single-Cell Transcriptomics of Human T Cells	Yiwei Lu	Qiy Cell	32142650	Understanding human T cell development and function through single-cell transcriptomics.	Human	regulation of	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
19	Launay_2020_Nat	Regenerative Medicine and Human T Cell Development	Ashley M. Launay	Nat Med	32042191	Development of human T cells is controlled by a complex network of transcription factors.	Human	lung	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
4850	Gleason_2022_Imm	Multi-modal Profiling of Human T Cells	Immunology	35483355	Myoblasts and myotubes are essential for muscle development and function.	Human	granulocyte	Infections	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
20	Rao_2020_Cell	Comparative Single-Cell Transcriptomics of Human T Cells	Rao M. Rao	Cell	32044662	Genomic and transcriptomic heterogeneity in human T cells.	Human	breast	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
21	Wang_2020_Stem	Adult Human Radial Glia	Rong Wang	Stem Cell Rep	32004492	Radial glia in the human brain is a diverse population of cells.	Human	brain	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
22	Rendtorff_2020_Na	Chromatin Profiling of Human T Cells	Andre F. Rendtorff	Nat Commun	31996569	The Bruton tyrosine kinase is a key component of the B cell receptor.	Human	lymphocyte	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
4856	Hay_2023_Cancer	Low-Throughput Single-Cell Transcriptomics of Human T Cells	Benjamin Hay	Br J Haematol	36787375	T cell receptor signaling is essential for T cell development and function.	Human	head and neck	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
23	Cillo_2020_Imm	Immune Response in Human T Cells	Anthony R. Cillo	Immunity	31504475	Head and neck cancer is a common cause of death.	Human	head and neck	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
24	Yost_2019_Nat	Clonal Replication of Human T Cells	Kathryn E. Yost	Nat Med	31359002	Regulation of human T cell development and function through single-cell transcriptomics.	Human	regulation of	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
4879	Hsu_2023_Inflam	Response to Inflammation in Human T Cells	Hsu Paul	Chin J Inflamm Res	37225148	Background: Human T cells are essential for immune responses.	Human	background	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
25	Merino_2020_Naiv	Barcoding of Human T Cells	Merino T.	Nat Commun	30770823	Primary triple-negative breast cancer is a rare and aggressive form of breast cancer.	Human	breast	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
26	Ben-David_2018_N	Genetic and Transcriptional Profiling of Human T Cells	Ben-David N.	Nature	30089904	Human cancer genomes are highly heterogeneous.	Human	mature B cell	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
27	Asai_2018_Cell	Single-Cell Transcriptomics of Human T Cells	Shih-Wei Asai	Cell	22961379	Knowledge of human T cell development and function through single-cell transcriptomics.	Human	regulation of	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
28	Rodriguez_2022_Sc	Preneoplastic Human T Cells	Sara Rodriguez	Sci Adv	35044826	Normal cell cycle regulation in human T cells.	Human	bone marrow	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
11	Li_2022_J_Clin	Investigating the Role of Human T Cells in Cancer	Shen Li	Tom J Clin Invest	34813507	Various types of cancer are caused by genetic mutations.	Human	fibroblast of	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope
30	Guo_2022_JCI	Insights into the Role of Human T Cells in Cancer	Wei Guo	JCI Insight	34793335	Colorectal cancer is a common cause of death.	Human	colon	Neoplasms	15325	Chromium	GSE180789	TCV-TXT				in scope

To assemble a catalog of relevant datasets, we surveyed several repositories, including GEO, SRA, AE, and dbGAP for single cell or single nuclei datasets. Our catalog includes over 6,600 publications or datasets published since 2013 and is annotated for disease, tissue, and assay.

2. Prescreening



Data Integrity team members check each dataset selected by consortium members to ensure they are compatible with the bioinformatics pipeline and have sufficient metadata to provide downstream value.

5. QC and Supplemental Information



- Stringent **QC tools** and scripts were applied to ensure the highest quality content.
- Data integrity team members sought cell-level metadata such as cell type annotations or cluster information on alternative websites and contacted authors for additional information.
- We worked with authors to resolve discrepancies, link the donor information and sample identifiers, and otherwise enrich the metadata.

Conclusions

Curated datasets and atlases delivered as part of this consortium are valuable resources that increase the accuracy and efficiency of downstream analyses, accelerate reproducible science and facilitate joint analysis of public data.

SCDS2: Year 4 and Beyond

SCDS2 kicked off in February 2025 with 6 charter members

Download Poster

