High-Resolution Single-Cell RNA-seq Autoimmune Atlas for Cellular Characterization and Target Discovery in Rheumatoid Arthritis and Systemic Lupus Erythematosus

Dimitris Konstantopoulos[1], Felicia Sabatino[1], Amrita Bhattacharya[1], Sondra Kopyscinski[1], Andy Hope[1], Christofer Karpelian[1], Dan Rozelle[1], Justyna Dopierala[2], Jonny Wray[2]

[1]Rancho BioSciences, LLC

[2]Nucleome Therapeutics Limited

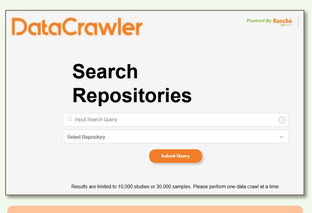
Background

Autoimmune diseases such as Rheumatoid Arthritis (RA) and Systemic Lupus Erythematosus (SLE) exhibit profound cellular heterogeneity across patients and tissues, complicating mechanistic insight and target discovery. In collaboration between **Rancho Biosciences** and **Nucleome Therapeutics**, we built a high-resolution single-cell RNA-seq (scRNA-seq) atlas that unifies curated Single Cell Data Science (**SCDS**) Consortium datasets with newly processed samples to enable consistent, cross-tissue comparisons in RA and SLE.

- **Scale and coverage:** The Autoimmune Atlas integrates **6 datasets** across **RA/SLE pathological states** (including **controls**), totaling **1 million cells** from **218 donors** and **290 samples**, spanning **synovial tissue, skin,** and **peripheral blood**.
- **Source datasets:** Accelerated development combines off-the-shelf analysis-ready datasets with newly re-analyzed datasets under a unified curation and processing strategy.
- **Annotation depth:** Employs a high-resolution cell-annotation procedure to resolve immune and stromal populations, with particular focus in **dendritic cells, memory B cells,** and **regulatory T cells** - cell types central to autoimmune pathogenesis.
- **Biological insight:** Reveals **shared** and **tissue-specific** transcriptional signatures across RA and SLE, supporting mechanism-of-disease hypotheses and prioritization of candidate targets.

1. Data tracking and collection

Query DataCrawler → Crawl Repositories → Annotate Datasets → Select “in scope” Datasets



DataCrawler

GEO, SRA, Synapse, dbGAP repositories

Single-cell and single-nuclei datasets

Catalog of 6,600 datasets

Client examines and picks datasets

id	shortname	title	authors	journal	pmid	abstract	organism	tissue	disease	number_cell	technique	repo_identifi	processed_d	in_scope
1	Absinta_2021_Nat	A lymphocyte	Martina Absi	Nature	34497421	Multiple sclei	Human	brain	Nervous Syst	66432	Chromium	GSE180759	TXT;CSV	in scope
2	Fernandes_2020_C	Single-Cell Tr	Hugo J.R. Fer	Cell Rep	33053338	The advent of	Human	mature B cell		15325	Chromium	E-MTAB-9154	ZIP;TXT	in scope
3	Pappalardo_2020_	Transcriptom	Jenna L. Pap	Sci Immunol	32948672	T cells provid	Human	brain;spine;t	Nervous Syst	6000	Chromium	GSE157277	CSV	in scope
4	Agarwal_2020_Nat	A single-cell	Devika Agarw	Nat Commun	32826893	We describe	Human	brain		18120	Chromium	GSE140231	TAR	in scope
5	Schafflick_2020_N	Integrated sir	David Schaffl	Nat Commun	31937773	Cerebrospina	Human	anatomical p		65326	Chromium	GSE138266	TSV;MTX	in scope
6	Schirmer_2019_Na	Neuronal vul	Lucas Schirm	Nature	31316211	Multiple sclei	Human	brain	Nervous Syst	48919	Chromium	PRJNA544731		in scope
7	Hong_2021_Front_I	Single-Cell Rl	Xiaoping Hor	Front Immun	33603736	Objective: Pri	Human	blood	Musculoskel	57288	Chromium	GSE157278	TSV;MTX	in scope
8	Caron_2020_Sci_Re	Single-cell ar	Maxime Caro	Sci Rep	32415257	Childhood ac	Human	bone marrow		39375	Chromium	GSE132509	MTX;TSV	in scope
9	Alivernini_2020_Na	Distinct syno	Stefano Alive	Nat Med	32601335	Immune-regu	Human	type B synovi	Musculoskel		Chromium	E-MTAB-8322	ZIP	in scope
10	Kim_2020_Nat_Mec	Targeted ther	Doyoung Kim	Nat Med	31959990	Drug-induce	Human	blood;skin	Mental Disor	54476	Chromium	GSE132802	H5	in scope
11	Aganezov_2021_Ge	Comprehensi	S. Aganezov,	Genome Res	32887686	Improved ide	Human		Neoplasms		Chromium	PRJNA476239		in scope
12	Bondoc_2021_Com	Identificatior	Alexander Bo	Commun Bio	34497364	Hepatoblastc	Human	liver;tumor	Neoplasms	26886	Chromium	GSE180665	XLSX;TAR;H5/	in scope
13	Sahu_2021_Oncog	Human cell tr	Biswajyoti Sa	Oncogene	34302118	Cancer is the	Human	mature B cell	Neoplasms	3500	Chromium	PRJEB31262		in scope
14	Jansky_2021_Nat_C	Single-cell tra	Selina Jansky	Nat Genet	33767450	Neuroblastor	Human	adrenal glan	Neoplasms		Chromium	GSE163431	TSV	in scope
15	Mercatelli_2021_Bi	Single-Cell Gr	Daniele Merc	Biomolecules	33525507	Neuroblastor	Human	mature B cell	Neoplasms	2067	Chromium	PRJNA668226		in scope
16	Luoma_2020_Cell	Molecular Pa	Adrienne M. I	Cell	32603654	Checkpoint b	Human	colon	Digestive Sys	51652	Chromium	GSE144469	TAR;CSV	in scope
17	Ji_2020_Cell	Multimodal A	Andrew L. Ji,	Cell	32579974	To define the	Human	integument	Neoplasms	50009	Chromium	GSE144240	JPG;TSV	in scope
18	Lu_2020_Cell	Complement	Yiwen Lu, Qi	Cell	32142650	Understandir	Human	regulation of	Neoplasms	998	Chromium	GSE135710	MTX;TSV	in scope
19	Laugney_2020_Nat	Regenerative	Ashley M. Lal	Nat Med	32042191	Development	Human	lung;tumor;n	Neoplasms	40505	Chromium	GSE123904	CSV	in scope
4850	Gideon_2022_Imm	Multimodal p	Gideon,Hann	Immunity	35483355	Mycobacteri	Monkey	granulocyte	Infections		Seq-Well	GSE200151		in scope
20	Rao_2020_Cold_Sp	Comparative	Rao M, Oh K,	Cold Spring H	32054662	Genomic ana	Human	breast	Neoplasms		Chromium	GSE140312		in scope
21	Wang_2020_Stem_(	Adult Human	Rong Wang, F	Stem Cell Rep	32004492	Radial glia (F	Human	brain;tumor	Neoplasms	12367	Chromium	GSE139448	CSV	in scope
22	Rendeiro_2020_Na	Chromatin m	André F. Ren	Nat Commun	31996669	The Bruton ty	Human		Neoplasms	43000	Chromium	GSE111015	TAR;MTX;TSV	in scope
4865	Hay_2023_Cancer_t	Low TCR Bind	Hay Zachary	Cancer Immu	36787375	T-cell recepto	Mouse	lymphocyte	Neoplasms		Chromium	GSE212980		in scope
23	Cillo_2020_Immun	Immune Lanc	Anthony R. Ci	Immunity	31924475	Head and nec	Human	neck	Neoplasms	131224	Chromium	GSE139324	MTX;TSV	in scope
24	Yost_2019_Nat_Me	Clonal replac	Kathryn E. Yo	Nat Med	31359002	Immunother	Human	regulation of	Neoplasms	79046	Chromium	GSE123813	TXT	in scope
4875	Hsu_2023_Inflamm	Responsive n	Hsu Paul, Ch	Inflamm Bow	37235748	Background:	Human	blood	Digestive Sys		CITE-Seq	GSE228165		in scope
25	Merino_2019_Nat_(	Barcoding re	D. Merino, T.	Nat Commun	30770823	Primary triple	Human		Neoplasms	25000	Chromium	GSE123926	MTX;TSV	in scope
26	Ben-David_2018_N	Genetic and t	Uri Ben-Davic	Nature	30089904	Human cance	Human	mature B cell	Neoplasms	26465	Chromium	GSE114462	CSV	in scope
27	Azizi_2018_Cell	Single-Cell M	Elham Azizi, /	Cell	29961579	Knowledge o	Human	regulation of	Neoplasms	72000	Chromium	GSE114727	CSV	in scope
28	Rodriguez_2022_Sc	Preneoplastic	Sara Rodrigu	Sci Adv	35044826	Normal cell c	Human	bone marrow	Neoplasms,	40169	Chromium	GSE185158	RDATA	in scope
29	Li_2022_J_Clin_Inve	Cardiomyocy	Shen Li, Tom	J Clin Invest	34813507	Various popu	Mouse	fibroblast of	Cardiovascul		Chromium	GSE185060	CSV	in scope
30	Guo_2022_JCI_Insig	Resolving the	Wei Guo, Cui	JCI Insight	34793335	Colorectal ca	Mouse	colon	Neoplasms	27927	Chromium	GSE188711	MTX;TSV	in scope

To assemble a catalog of RA and SLE studies, we surveyed expression repositories, including GEO, SRA, Synapse, and dbGaP, for single-cell or single-nuclei datasets. Our catalog includes almost 8,000 publications from the last 10 years and is annotated for species, disease, tissue, and assay types.
Rancho have processed > 1000 datasets through Single Cell Data Science Consortium (SCDS), a member-led initiative to create resources which advance therapeutic discovery. After 1 year, SCDS datasets are commercialized and available as a high-quality product that enhances our client's research.

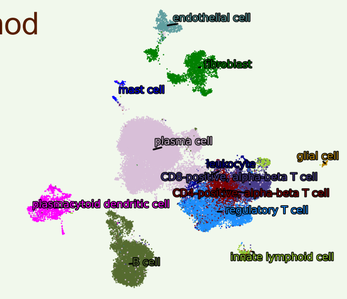
2. QC, data ingestion, and curated metadata import

- Rancho's **QC** and **ingestion pipeline** was applied to ensure high-quality data while preserving the underlying biological signal.
- Our data integrity team collaborated with study authors to resolve discrepancies, link donor information with sample identifiers, and enrich metadata for downstream use.
- Any non-SCDS datasets were processed using the same principles and standardized procedures, ensuring consistency across the entire atlas.

3. Automated cell-type annotation workflow

CZ CELL×GENE · scGPT · Azimuth · CellTypist · scArches

For Cell-type annotation we used a comprehensive, multi-method strategy leveraging CELLxGENE Census reference data.



- Transfer/label-transfer approaches (scArches, Azimuth)
- Supervised classification (CellTypist)
- Representation learning (scGPT)
- Reference-backed annotation (CELLxGENE Census)

Predictions were reconciled into **consensus labels** and validated against canonical marker-gene signatures with targeted manual review. Cell-type categories were further organized into parent classes to enable intuitive browsing and cross-dataset comparisons.

4. Data Integration and atlas creation

Integration & Feature Selection

- Integration model: scVI (variational autoencoder; scvi-tools) used as the batch-correction/latent-representation method of choice.
- Tissue-aware markers: Curated canonical cell-type markers for synovium, skin, and peripheral blood from CellMarker 2.0.
- Variable features: Aggregated union of highly variable genes (HVGs) across datasets (a gene counted if HVG in ≥1 dataset).
- Final integration genes: Intersection of the marker panel and the HVG union → ~2,000 genes used to train scVI.

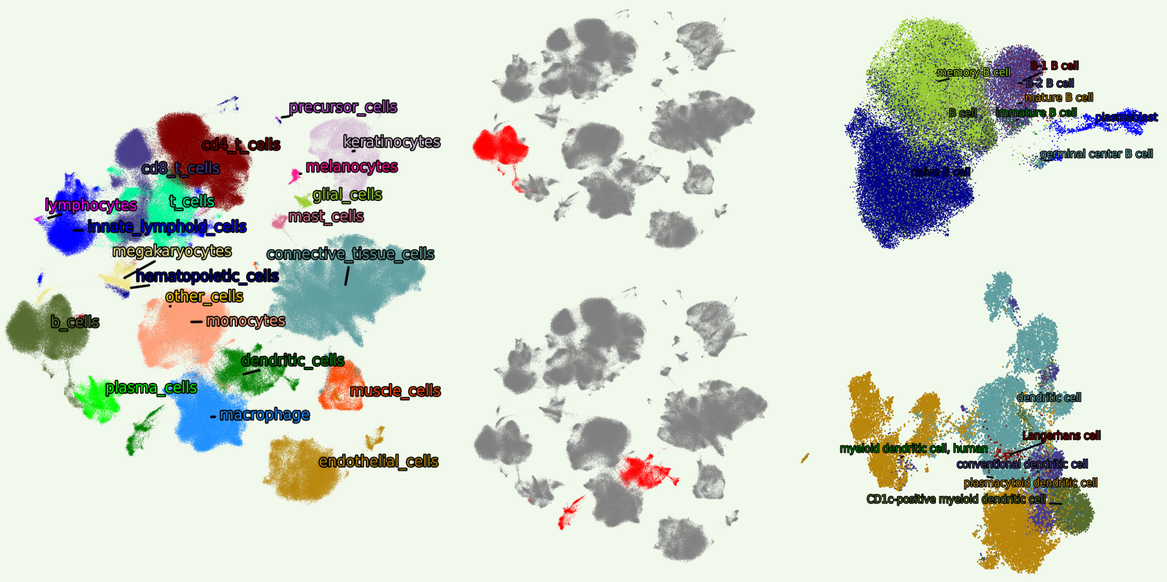
Latent-Space QC & Refinement

- Neighborhood-concordance filtering: In the scVI latent space, a cell is retained if ≥55% of its neighbors share the same parent cell class (reduces ambiguous/outlier assignments).
- Minimum category size: Parent cell classes with <40 cells (total) were removed to stabilize downstream analyses.

Clustering & Visualization

- Graph-based clustering and UMAP dimensionality reduction were performed on the scVI-corrected embedding.

5. Atlas-level Queries: Zooming into Biological Signal



- Our robust annotation workflow enables high-confidence cell identities across tissues and studies, supporting hierarchical exploration from major lineages to fine-grained subsets.
- This makes the atlas query-ready: users can “zoom” from refined parent classes to disease-relevant components (B cells, dendritic cells) across RA/SLE and tissues.
- Downstream analyses can resolve intra-cell-type diversity (e.g., memory B-cell states, dendritic-cell subtypes), providing a principled path to interrogate underlying biology and generate hypotheses.

Conclusions

The Nucleome-Rancho Autuimmune atlas uncovers both shared and tissue-specific cellular signatures in RA and SLE, offering a powerful foundation for understanding disease pathogenesis. The atlas scale and resolution creates a unique opportunity to query disease-relevant cell states, identify candidate biomarkers, and guide target discovery in future analyses. Beyond this, the atlas establishes a scalable and reproducible framework for disease-focused single-cell analysis, enabling the construction of reference datasets that support hypothesis generation and biomedical research.

Nucleome THERAPEUTICS · Rancho biosciences