



UCLA

Institute for Quantitative & Computational Biosciences

2026

Bruins-In-Genomics (B.I.G.) Summer Program

Summer Undergraduate Research Program

Symposium – August 14, 2026

Director's Message



In its essence, BIG Summer is about providing undergraduate students with the opportunity to experience what it means to be a scientist.

Today's most exciting advances in biomedical and life sciences happen at the intersection of biology, computation, and data sciences. BIG Summer gives students the opportunity to explore this interdisciplinary world through hands-on research, interactive skills development, and close mentorship from UCLA faculty and researchers. Along the way, students develop not only technical skills, but also the ability to ask meaningful questions, work through challenges, communicate their findings, and think like scientists.

BIG Summer is about more than research skills. It is an opportunity for students to discover where their interests and talents can take them. Through mentorship and professional development, students begin to identify their goals and learn how to take the next steps toward graduate school and careers in science.

The abstracts and posters in this booklet represent the culmination of eight weeks of curiosity, hard work, and discovery. Each project reflects a student's contribution to the research community and the beginning of a journey that we hope will continue long beyond this summer.

Congratulations to the 2026 BIG Summer cohort on all that you have accomplished ! And a BIG Thank you to the mentors - the faculty, postdocs and graduate students - whose commitment and generosity make this program possible. Together they reflect the values of BIG Summer and UCLA as a whole.

Alexander Hoffmann

Director, Institute for Quantitative and Computational Biosciences
UCLA

Table of Contents

Student Research Abstract and Poster

1. Avani Athavale	C. Luo lab	Human Genetics
2. Ishya Bahl	N. Harris lab	Neurosurgery
3. Zenya Bian	A. Clark lab	MCDB
4. Kiley Chen	M. Su lab	MIMG
5. Samuel Choi	P. Bradley lab	MIMG
6. Kara Connolly	Q. Cheng lab	Medicine
7. Maggie De Leon	L. Bouchard lab	ChemBiochem
8. Ben Dewey	N. Harris lab	Neurosurgery
9. Ilinca Flacau	S. Lindert lab	ChemBiochem
10. Lipika Goel	A. Meyer lab	Bioengineering
11. Jaiden Ha	K. Kappel lab	ChemBiochem
12. Elsa Harley	B. Knowles lab	EEB
13. Tianpei He	D. Buonomano lab	Neuroscience
14. Jasmine Herrera	A. Hoffmann lab	MIMG
15. Veda Hingarh	W. Hsu lab	Radiological Sciences
16. Ihyun Jeon	K. Kappel lab	ChemBiochem
17. Vyas Koduvayur	A. Hoffmann lab	MIMG
18. Harshita Kukreja	S. Emaminejad lab	Computer Science
19. Derek Lee	T. O'Sullivan lab	MIMG
20. Yi-Ai (Allison) Li	W. Hugo lab	Medicine, Dermatology
21. Emma Liao	C. Luo lab	Human Genetics
22. Annie Linley	V. Savage lab	EEB, CompMed
23. Livia Liu	F. Guo lab	Biological Chemistry
24. Yinuo Liu	X. Xiao lab	IBP
25. Sami Majid	A. Askary lab	MCDB
26. Charissa Mak	V. Savage lab	EEB, CompMed
27. Emma Maliar	J. Chiang lab	Neurosurgery, CompMed
28. Robert Mathews	J. Lusi lab	Medicine/MIMG/Human Genetics
29. Sahel Munshi	A. Hoffmann lab	MIMG
30. Lois Pak	H. Allayee lab	Medicine
31. Jaime Pelayo	A. Hoffmann lab	MIMG
32. Madeline Shono	M. Pellegrini lab	MCDB
33. Chrissy Tetreault	J. Lusi lab	Medicine/MIMG/Human Genetics
34. Nadia Tsai	M. Fox lab	Anthropology, Psychiatry
35. Valerie Vandersall	M. Bouhaddou lab	MIMG
36. Jingyue Wang	N. Harris lab	Neurosurgery
37. Xiangting Wu	W. Liu lab	Applied Mathematics
38. Jessie Xu	A. Covarrubias lab	MIMG
39. Rongying (Flicka) Zhang	X. Yang lab	IDP

Student: Avani Athavale
Faculty Advisor: Chongyuan Luo
Direct Mentor: Cuining Liu
Department: Human Genetics
<https://luogenomics.github.io/>

**Avani
Athavale**

**Pre-
Computational
Biology -
UCLA**

**Expected year
of graduation:
June 2029**

Understanding Cell-type Specific Changes in Alzheimer's Disease Methylome

AVANI ATHAVALE¹, Cuining Liu^{2,4}, Zofia Lok³, Nasser Elhajjaoui², Rosa Nguyen², Min Jen Tsai², Zachary Von Behren², Yi Zhang², Daniel Lee⁵, Xiangdong William Yang⁵, Chongyuan Luo²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Human Genetics, David Geffen School of Medicine, UCLA

³ Computational and Systems Biology, UCLA

⁴ Bioinformatics Interdepartmental Graduate Program, UCLA

⁵ Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, UCLA

Alzheimer's disease (AD) is a progressive neurodegenerative disease affecting millions of older adults. However, the relationship between cell-type specific DNA methylation and regulatory mechanisms implicated in AD is unclear. To better understand this, we used a knock-in mouse model that contains the familial human SAA mutations on the amyloid precursor protein (APP) and analyzed methylomic data generated by single nucleus methylcytosine and transcriptome sequencing (snmCT-seq). We sought to understand cell-type specific Alzheimer's signatures exhibited by cells in the entorhinal cortex, identifying differentially methylated regions (DMRs) that overlap with transcription factor binding motifs and gene bodies for each cell type. Our results implicate previously-found motifs, nominate new ones, and indicate cell-type specificity in the methylation profile of Alzheimer's disease. In the future, by combining this analysis with the transcriptomic profiles of the sampled cells, we will better understand the epigenetics and gene expression landscape of Alzheimer's disease.

Understanding Cell-type Specific Changes in Alzheimer's Disease Methylome

AVANI ATHAVALE¹, Cuining Liu^{2,4}, Zofia Lok³, Nasser Elhajjoui², Rosa Nguyen², Min Jen Tsai², Zachary Von Behren², Yi Zhang², Daniel Lee⁵, Xiangdong William Yang⁵, Chongyuan Luo²

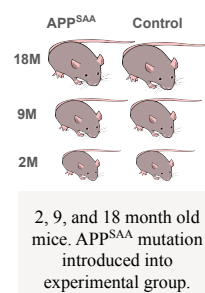
¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA; ² Department of Human Genetics, David Geffen School of Medicine, UCLA; ³ Computational and Systems Biology, UCLA; ⁴ Bioinformatics Interdepartmental Graduate Program, UCLA; ⁵ Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, UCLA

Abstract

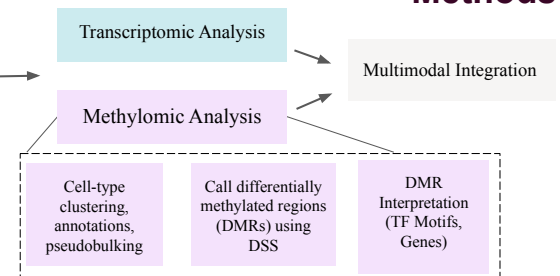
Alzheimer's disease (AD) is a progressive neurodegenerative disease affecting millions of older adults. However, the relationship between cell-type specific DNA methylation and regulatory mechanisms implicated in AD is unclear. To better understand this, we used a knock-in mouse model that contains the familial human SAA mutations on the amyloid precursor protein (APP) and analyzed methylomic data generated by single nucleus methylcytosine and transcriptome sequencing (snmCT-seq). We sought to understand cell-type specific Alzheimer's signatures exhibited by cells in the entorhinal cortex, identifying differentially methylated regions (DMRs) that overlap with transcription factor binding motifs and gene bodies for each cell type. Our results implicate previously-found motifs, nominate new ones, and indicate cell-type specificity in the methylation profile of Alzheimer's disease. In the future, by combining this analysis with the transcriptomic profiles of the sampled cells, we will better understand the epigenetics and gene expression landscape of Alzheimer's disease.

Background

- Alzheimer's disease (AD) affects the brain through accumulation of amyloid beta plaques and neurofibrillary tau tangles
- APP^{SAA} knock-in mouse model introduces a combination of six mutations in the amyloid precursor protein of mice to humanize them
- Entorhinal cortex is where AD pathophysiology begins, but **specific cellular and molecular mechanisms are unclear**
- Our goal is to understand how different cell types in entorhinal cortex exhibit epigenetic alterations in Alzheimer's disease

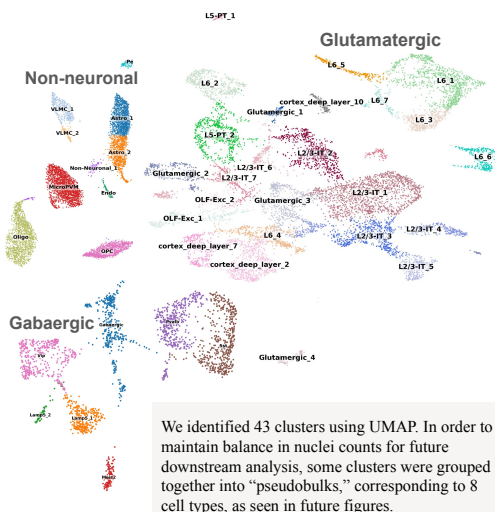


Methods

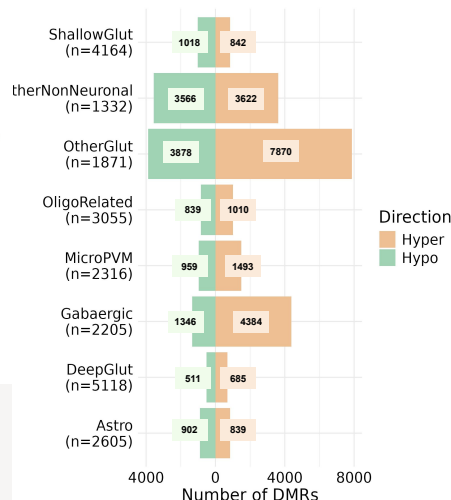


Results

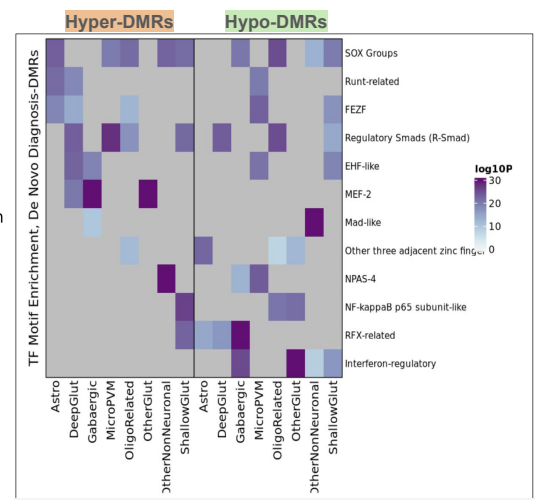
1. UMAP clustering results in expected brain cell types based on canonical gene markers.



2. Each cell type yielded several hundred/thousand differentially methylated regions (DMRs) based on disease status



3. Motif analysis links transcription factor families to DMR regions.



Discussion

- Our findings indicate that different cell types exhibit unique characteristics in their methylome in Alzheimer's disease
 - A lack of correlation between diagnosis based DMRs and nuclei count per cell type indicates that different cell types are presenting distinct signatures
- Differential methylation in binding sites of TFs relevant to neuronal activity/learning/memory suggests that methylation is a valuable insight into how Alzheimer's affects these processes

Future Steps

- Further validate and explore biological significance of TF motifs
- Investigate the overlap of DMRs with genes
- Analyze age-based differentially methylated regions to understand signatures characteristic of normal aging process
- Integrate methylomic and transcriptomic data
- Integrate all age groups to understand how AD pathophysiology intersects with aging process

Acknowledgements: Thank you to collaborators from the Luo lab and Yang lab for guidance on the progress of this project. This work was supported by the National Institute on Aging of the National Institutes of Health under project U01AG076804. Thank you to the BIG Summer Program and the UCLA URC-Sciences Summer Program for supporting this work.



Student: Ishya Bahl
Faculty Advisor: Neil Harris
Direct Mentor: Azad Azargushasb
Department: Neurosurgery
<https://neilharrislab.dgsom.ucla.edu/>

Ishya Bahl

**Computational
and Systems
Biology -
UCLA**

**Expected year
of graduation:
June 2027**

Behavioral Signatures of Repeated Mild Traumatic Brain Injury: Classifying Injured and Sham Mice from Odor-Reversal Reward Learning

ISHYA BAHL¹, BEN DEWEY¹, Azad Azargushasb², Sam Vander Dussen², Neil G. Harris³

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² UCLA Brain Injury Research Center, Department of Neurosurgery, DGSOM, UCLA

³ Intellectual Development and Disabilities Research Center, UCLA

Repeat mild traumatic brain injury (rmTBI) is associated with executive deficits such as impaired inhibitory control and cognitive flexibility. However, the primary behavioral deficit is not fundamentally understood. This project determines whether a mouse is sham or injured using models trained on behavioral inputs during an odor reversal in a probabilistic go/no-go paradigm. 26 mice (13 injured, 13 sham), which previously reached a learned association, were monitored over seven reversal-learning days. Features across specific task periods, like lick rate, velocity, and accuracy, were used to train various classifiers with cross-validation. Using behavioral data alone, most models could not classify sham versus injured mice above 80% accuracy although some models demonstrated a high area under the curve. Analyzing corresponding functional ultrasound imaging (fUSi) data can help identify the compensatory mechanisms in the brain that allow the injured mice to perform nearly the same.



Behavioral Signatures of Repeat Mild Traumatic Brain Injury: Classifying Injured and Sham Mice from Odor-Reversal Reward Learning

Ishya Bahl¹, Ben Dewey¹, Azad Azargushasb², Samuel Vander Dussen², Neil G. Harris³

¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

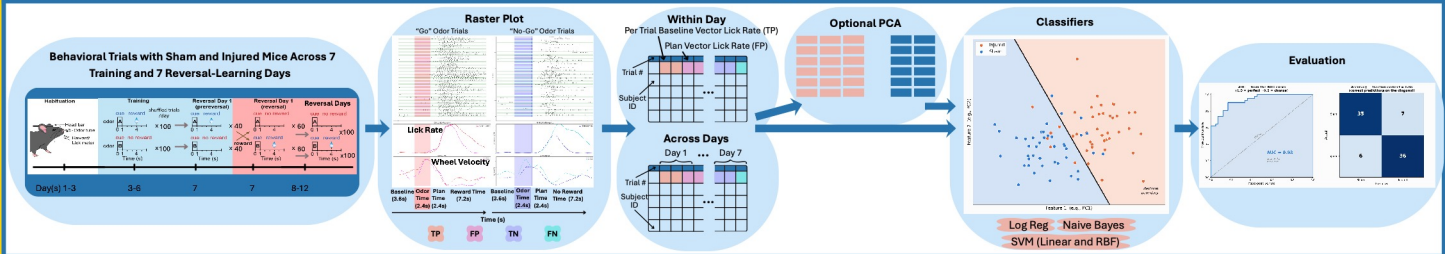
²UCLA Brain Injury Research Center, Department of Neurosurgery, DGSOM, UCLA

³Intellectual Development and Disabilities Research Center, UCLA



*Co-authors

Methodology



Background

- Repeat mild traumatic brain injury (mTBI) produces lasting executive deficits, most consistently in inhibitory control and cognitive flexibility.
- Rodent models confirm these deficits on reversal-learning tasks, but rely on group-level summary measures (percent correct, trials to criterion) that show injured animals differ without identifying the primary deficit.
- Injured mice often reach near-normal performance despite documented pathology, suggesting compensation that averaged behavior cannot capture.
- Trial-level features segmented by task period retain information those summaries discard, so we ask whether a classifier can recover injury status from behavior alone, and use concurrent fUSI to ask what the brain is doing when behavior looks intact.

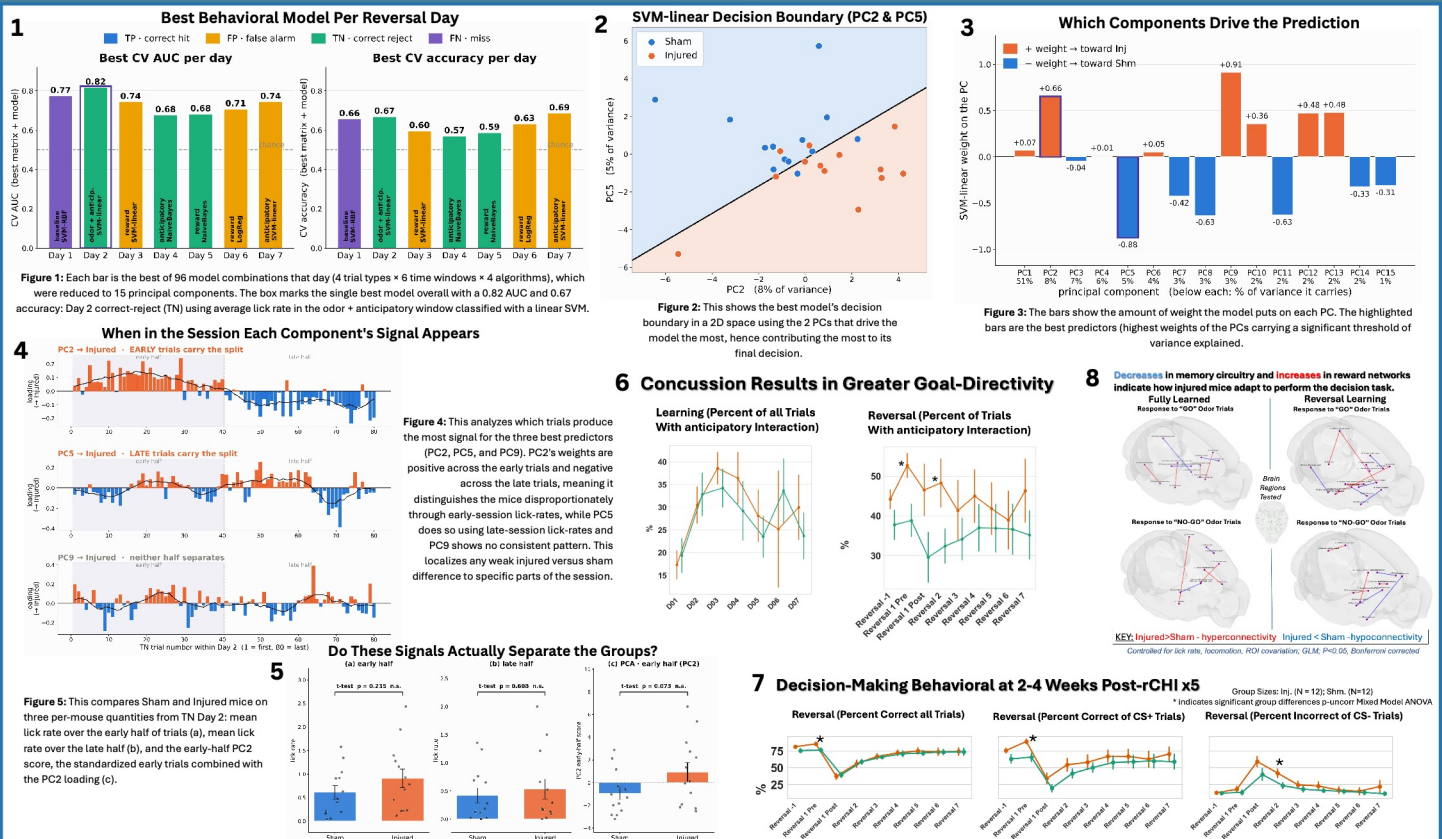
Abstract

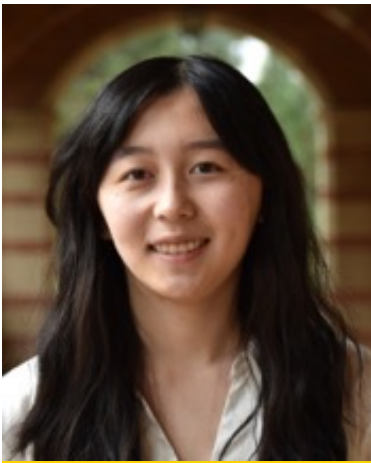
Repeat mild traumatic brain injury (mTBI) is associated with executive deficits such as impaired inhibitory control and cognitive flexibility. However, the primary behavioral deficit is not fundamentally understood. This project determines whether a mouse is sham or injured using models trained on behavioral inputs during an odor reversal in a probabilistic go/no-go paradigm. 26 mice (13 injured, 13 sham), which previously reached a learned association, were monitored over seven reversal-learning days. Features across specific task periods, like lick rate, velocity, and accuracy, were used to train various classifiers with cross-validation. Using behavioral data alone, most models could not classify sham versus injured mice above 80% accuracy although some models demonstrated a high area under the curve. Analyzing corresponding functional ultrasound imaging (fUSI) data can help identify the compensatory mechanisms in the brain that allow the injured mice to perform nearly the same.

Discussion / Future Works

- All classifiers failed to exceed 80% accuracy on behavior alone, indicating that injured mice perform this task closely enough to shams that trial-level behavior carries limited discriminative signal.
- The high AUC in some models despite modest accuracy suggests the separating information is real but weakly expressed, likely concentrated in a subset of animals, task periods, or reversal days rather than distributed across the full dataset.
- Near-normal performance in the face of documented injury points to compensation; analyzing the corresponding fUSI data will test whether injured mice recruit additional or alternative circuitry to sustain it.
- Next steps: identify which features and task periods drive the models that do separate the groups, examine whether classification improves on early reversal days when flexibility demands peak, and evaluate whether adding fUSI features is more capable of delineating between the two groups

Results





Student: Zenya Bian

Faculty Advisor: Amander Clark

Direct Mentor: Mark Larsen

Department: Molecular, Cell, and Developmental Biology (MCDB)

<https://clark.mcdb.ucla.edu/>

Zenya Bian

**Molecular, Cell,
and
Developmental
Biology - UCLA**

**Expected year
of graduation:
June 2027**

Effects of TGF- β /FGF/WNT Signaling on Human Germ Cell Specification

ZENYA BIAN^{1,2}, Mark Larsen²⁻⁵, and Amander Clark²⁻⁵

¹B.I.G. Summer Program, Institute for Quantitative and Computational Biosciences, UCLA.

²Department of Molecular, Cell and Developmental Biology, UCLA.

³Molecular Biology Institute, UCLA.

⁴Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, UCLA.

⁵Center for Reproductive Science, Health and Education, UCLA.

The germline is the only cell lineage capable of giving rise to egg and sperm cells, but mechanisms driving germline specification in the human embryo remain poorly understood due to technical and ethical limitations. Lineage specification of primordial germ cells (PGCs)—the precursor population to eggs and sperm—has been established to be reliant on multiple signaling pathways, including Transforming Growth Factor Beta (TGF- β), Fibroblast Growth Factor (FGF), and WNT signaling. However, the specific interactions between these signaling pathways and their downstream roles in PGC specification remain unresolved. Here, we leveraged a simplified in vitro PGC specification model to robustly specify PGCs and other early embryonic cell lineages. By perturbing the timing and concentration of growth factors, we have demonstrated through single-cell RNA sequencing that specific changes to TGF- β , FGF, and WNT signaling can bias specification towards PGCs, amnion, or mesoderm fates.

Dynamics in TGF- β , FGF, and WNT Signaling Specify Human Germline and Somatic LineagesZENYA BIAN¹, Mark Larsen¹, and Amander Clark¹¹Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, CA.

UCLA QCBio

UCLA College Life Sciences
Biomedical ResearchUCLA Undergraduate Education
Research Centers - SciencesUCLA College Life Sciences
Molecular, Cell &
Developmental Biology

Introduction | The germline is the only cell lineage capable of giving rise to egg and sperm cells, but mechanisms driving germline specification in the human embryo remain poorly understood due to technical and ethical limitations (1). Lineage specification of primordial germ cells (PGCs)—the precursor population to eggs and sperm—has been established to be reliant on multiple signaling pathways, including Transforming Growth Factor Beta (TGF- β), Fibroblast Growth Factor (FGF), and WNT (2, 3). However, the specific interactions between these signaling pathways and their downstream roles in PGC specification remain unresolved. Here, we leveraged a simplified in vitro PGC specification model to robustly specify PGCs and other early embryonic cell lineages. By perturbing the timing and concentration of growth factors, we have demonstrated using single-cell RNA sequencing that specific changes to TGF- β , FGF, and WNT signaling can bias specification towards PGCs or other cell fates. Notably, we have elucidated a requirement for TGF- β 1/2/3 as key drivers of PGC specification and their cooperative behavior with FGF2 signaling.

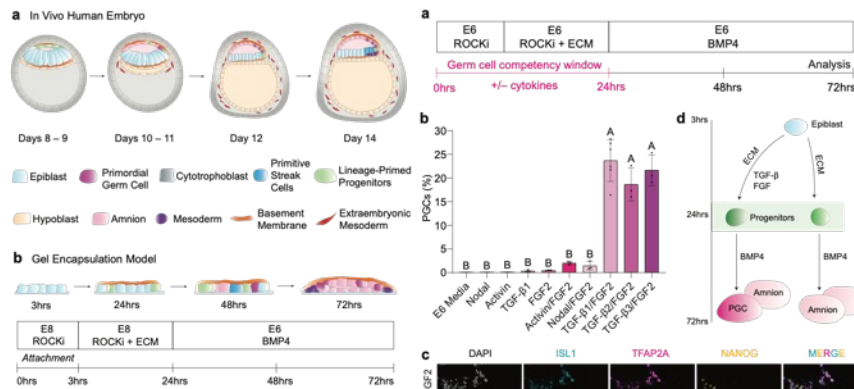


Figure 1: Comparison of PGC Specification In Vivo Versus In Vitro. (a) PGCs are specified from a lineage-primed progenitor and are hypothesized to emerge in the amnion and/or the posterior epiblast around 12 days post-fertilization. (b) PGCs can be robustly specified in vitro using a gel encapsulation model utilizing extracellular matrix and BMP4.

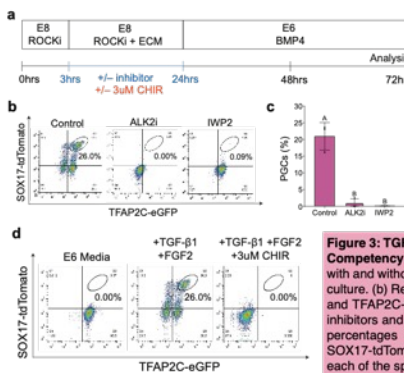


Figure 2: PGC Specification Requires a Combination of TGF- β and FGF2 Signaling. (a) Diagram of culture conditions with and without the addition of inhibitors or CHIR from 3 to 24hrs of culture. (b) Representative flow cytometry plots of SOX17-TdTomato and TFAP2C-eGFP expression as a result of each of the specified inhibitors and control conditions. (c) Quantification of average PGC percentages in (b). (d) Representative flow cytometry plots of SOX17-TdTomato and TFAP2C-eGFP expression as a result of each of the specified culture conditions.

Figure 3: TGF- β , FGF, and WNT Signaling Are Essential for PGC Competency and Specification. (a) Diagram of culture conditions with and without the addition of inhibitors or CHIR from 3 to 24hrs of culture. (b) Representative flow cytometry plots of SOX17-TdTomato and TFAP2C-eGFP expression as a result of each of the specified inhibitors and control conditions. (c) Quantification of average PGC percentages in (b). (d) Representative flow cytometry plots of SOX17-TdTomato and TFAP2C-eGFP expression as a result of each of the specified culture conditions.

Conclusions & Future Directions | By leveraging our in vitro PGC specification model, we have elucidated previously uncharacterized roles for TGF- β 1/2/3 and FGF2 signaling by demonstrating that both pathways are required during specification, and that they interact in a cooperative manner. By activating and inhibiting WNT signaling through targeted perturbation experiments, we show that the specification of PGCs is also highly sensitive to WNT and subject to fate redirection towards amnion or mesoderm cell fates. Our single-cell RNA sequencing data indicate that cells specified under differing signaling conditions during the first 24 hours of differentiation correspond to multiple somatic lineages from an in vivo human embryonic reference dataset. Future analyses will focus on establishing a method of mapping the specified germline.

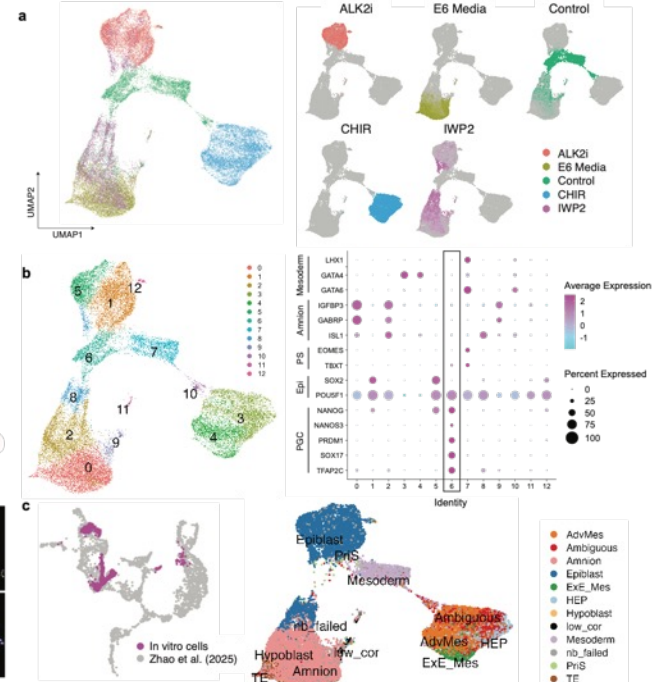


Figure 4: Cells Specified in Different Conditions Map to In Vivo Somatic Embryonic Lineages. (a) UMAP projections created from single-cell RNA sequencing performed on cells cultured in five different signaling environments at 72 hours. (b) An integrated UMAP plot of all conditions and the corresponding dot plot show expression of key PGC marker genes in cluster 6. (c) Using a comprehensive in vivo human embryonic reference (4), all cells were mapped according to reference annotations. Cluster 6 (in vitro germ cells) was excluded due to the lack of PGC annotations from the in vivo reference.

Acknowledgements | Iyngsun Huang generated the SOX17-TdTomato; TFAP2C-eGFP reporter cell line. Vernon Monteiro created the reference mapping scripts. Jeffrey Calm and Ken Yamashita provided technical support from the UCLA BiSCC core facilities. Melody Wu and Gavin Zhao provided library preparation and sequencing services from the UCLA TCGC core facility. This research was supported in part by the Zheng-Hu Family URC-Sciences Research Scholarship and the UCLA Biomedical Research Minor Summer Scholarship. Special thanks to the Clark Lab.

References

- Hughes, P. J., Morris, N. A., & Tam, P. P. L. Technical challenges of studying early human development. *Development* 150, dev201797 (2023).
- Sasaki, K. et al. Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* 17, 178–194 (2015).
- Jin, K. et al. Efficient differentiation of human primordial germ cells through genetic control reveals a key role for Nodal signaling. *eLife* 11, e72811 (2022).
- Zhao, C. et al. A comprehensive human embryo reference tool using single-cell RNA-sequencing data. *Nat. Methods* 22, 193–206 (2025).



Student: Kiley Chen

Faculty Advisor: Maureen Su

Direct Mentor: Christian Bustillos

Department: Microbiology, Immunology,
& Molecular Genetics (MIMG)

<https://maureensulab.wixsite.com/maureensugroup>

Kiley Chen

MIMG - UCLA

**Expected year
of graduation
June 2027**

Single-Cell Reanalysis of Immune Responses to Teplizumab in Type 1 Diabetes

KILEY CHEN^{1, 2}, Maureen Su^{2, 3}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Microbiology, Immunology, and Molecular Genetics, UCLA

³ Department of Pediatrics, UCLA

Teplizumab delays progression from stage 2 to stage 3 type 1 diabetes. However, its long-term effects on immune cell populations remain incompletely understood. We reanalyzed peripheral blood mononuclear single-cell RNA-sequencing data from 14 participants in the TrialNet TN10 study who received teplizumab or placebo. Samples were collected at baseline, 3 months, and 18 months and were integrated at each time point. Cluster identities were assigned by referencing cluster-specific genes with canonical markers. T cells were further subsetted and reintegrated for higher resolution characterization. Comparative analyses across treatment groups and time points are ongoing. This reanalysis aims to refine T-cell subtype identification and characterize immune states associated with the long-term response to teplizumab.

Teplizumab-Associated Transcriptional Changes Across CD4⁺ and CD8⁺ T Cell Populations in Type 1 Diabetes

KILEY CHEN^{1,2}, Maureen Su, M.D.^{2,3}¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA²Department of Microbiology, Immunology, & Molecular Genetics, UCLA³Department of Pediatrics, UCLA

Introduction

Type 1 diabetes (T1D) results from immune-mediated destruction of pancreatic β cells, with autoreactive T cells playing a central role in disease progression.¹ Teplizumab is a non-FcR-binding anti-CD3 monoclonal antibody that can delay progression from stage 2 to stage 3 T1D after a single 14-day treatment course. Previous studies suggest that teplizumab alters T-cell activation and differentiation, promoting features of exhaustion and immune regulation while limiting expansion of autoreactive CD8⁺ T cells.² However, how these effects are distributed across distinct peripheral CD4⁺ and CD8⁺ T-cell states remains incompletely understood. Here, we reanalyzed single-cell RNA-sequencing data from PBMCs collected three months after teplizumab or placebo treatment to examine treatment-associated changes in T-cell composition and transcriptional state.

Methods

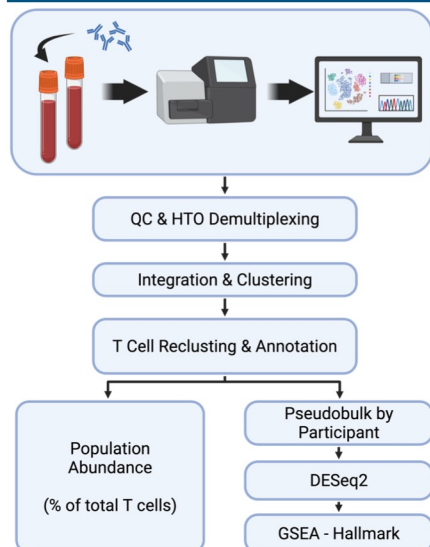


Figure 1: Analysis Workflow. PBMC scRNA-seq data underwent quality control, HTO demultiplexing, integration, clustering, T-cell subsetting, reclustering, and annotation. Treatment-associated differences were evaluated using participant-level population abundance and pseudobulk differential-expression analysis followed by Hallmark GSEA. scRNA-seq data: GSE271063 (Lledó-Delgado et al., 2024)

Results

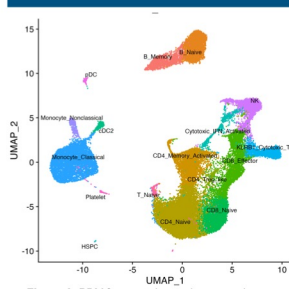


Figure 2: PBMC annotation at three months post-treatment. UMAP of integrated PBMC scRNA-seq data from 14 participants (7 placebo and 7 teplizumab), annotated into major immune cell populations.

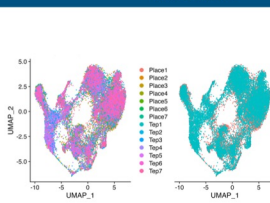


Figure 4: Participant and treatment representation within the T-cell subset. UMAPs of reclustered T cells colored by participant (left) and treatment group (right). Cells from placebo and teplizumab participants were broadly distributed without obvious participant- or treatment-specific bias.

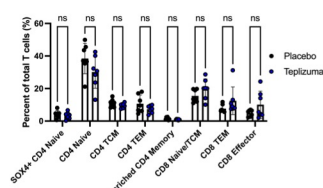


Figure 5: Relative abundance of T-cell populations by treatment group. Each point represents the percentage of total T cells belonging to the indicated population for one participant. No significant differences in the abundance of the analyzed CD4⁺ or CD8⁺ T-cell populations were detected between placebo and teplizumab groups. P values calculated using two-sided Wilcoxon rank-sum tests.

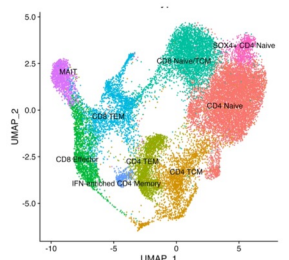


Figure 3: Identification of CD4⁺ and CD8⁺ T-cell states. UMAP of reclustered T cells annotated using canonical marker expression.

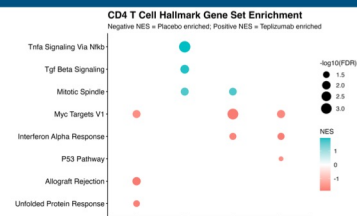


Figure 6: CD4⁺ T-cell Hallmark GSEA. Significant pathways (FDR < 0.05) from participant-level pseudobulk analysis. Placebo favored MYC and IFN- α programs, while teplizumab favored mitotic-spindle and signaling pathways.

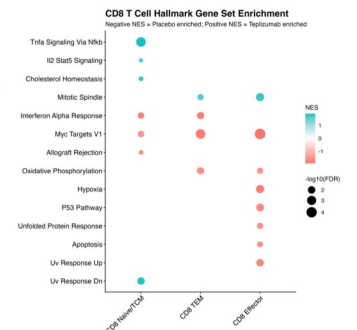


Figure 7: CD8⁺ T-cell Hallmark GSEA. Significant pathways (FDR < 0.05) from participant-level pseudobulk analysis. Placebo favored MYC, IFN- α , and stress/metabolic programs, while teplizumab favored mitotic-spindle-associated pathways.

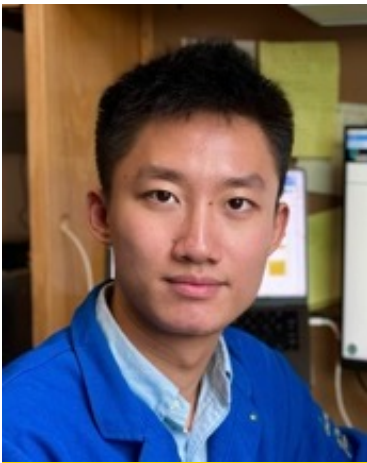
Conclusions

At three months post-treatment, no significant differences were detected in the relative abundance of the major CD4⁺ and CD8⁺ T-cell populations. However, participant-level pseudobulk analysis and GSEA revealed transcriptional differences across multiple T-cell states. Placebo cells showed recurrent enrichment of MYC- and IFN- α -associated programs, with additional stress and metabolic signatures in CD8 effector cells, whereas teplizumab-treated cells showed enrichment of mitotic-spindle-associated and selected signaling pathways. These findings suggest that teplizumab is associated with transcriptional reprogramming without major compositional remodeling, potentially shifting T cells away from a chronic inflammatory/stress-associated state toward alternative signaling transcription signature.

References and Acknowledgements

- Herold KC, Delong T, Perdigo AL, et al. The immunology of type 1 diabetes. *Nat Rev Immunol.* 2024;24:435–451. doi:10.1038/s41577-023-00985-4.
- Lledó-Delgado A, Preston-Hurlburt P, Currie S, et al. Teplizumab induces persistent changes in the antigen-specific repertoire in individuals at risk for type 1 diabetes. *J Clin Invest.* 2024;134(18):e177492. doi:10.1172/JCI177492.

We thank Dr. Ana Lledó-Delgado for providing participant metadata used in this analysis. We thank the UCLA Bruins-in-Genomics (B.I.G.) Summer Program for supporting this research experience and the Maureen Su Lab for mentorship and support. Figures and schematics were created with BioRender.com and GraphPad Prism.



Student: Samuel Choi
Faculty Advisor: Peter Bradley
Direct Mentor: Jim Thach
Department: Microbiology, Immunology,
 & Molecular Genetics (MIMG)
<https://bradleylab.dgsom.ucla.edu/>

Samuel Choi

**Microbiology,
 Immunology,
 & Molecular
 Genetics
 (MIMG) - UCLA**

**Expected year
 of graduation
 June 2027**

Dissecting the interactions and functions of alveolin subcomplexes in the *Toxoplasma* inner membrane complex

SAMUEL J CHOI^{1,2}, Peter J. Bradley^{2,3}

¹ B.I.G. Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

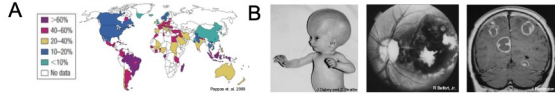
² Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA 90095, USA

³ Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA

The inner membrane complex (IMC) in apicomplexan parasites is a conserved membrane-cytoskeletal system beneath the plasma membrane that modulates parasite replication and motility. In *Toxoplasma gondii*, the IMC includes fourteen intermediate filament-like proteins called alveolins which are characterized by a conserved alveolin domain and play roles in reinforcing the organelle and assisting in daughter cell assembly. Here we identify two alveolin subcomplexes: IMC7 and IMC12 form an interdependent complex exclusive to maternal parasites, while IMC3, IMC6, and IMC10 form a complex enriched in the daughter IMC. Utilizing CRISPR/Cas9 gene knockout and pairwise yeast-two-hybrid (Y2H) assays, we demonstrate direct binding between IMC7 and IMC12. Previous work indicated IMC3 and IMC6 interact via their alveolin domains; our Y2H analysis further reveals an interaction between IMC10 and both IMC3 and IMC6. AlphaFold 3 prediction suggests that the alveolins utilize beta sheets formed via the alveolin domain for interaction and filament formation. Surprisingly however, truncation studies of IMC10 show that the minimal functional region surprisingly lies outside the alveolin domain in the C-terminus of the protein, which includes a conserved valine-rich region that appears to be sufficient for tethering to the IMC and function. In addition, the alveolin domain of IMC10 alone cannot target properly to the IMC. These findings make IMC10 the first alveolin to function without its alveolin domain. While the IMC7-IMC12 complex supports the model of alveolin domain interactions and filament formation, evidence from IMC10's C-terminal binding suggests the presence of uncharacterized protein features required for SPN localization and filament assembly.

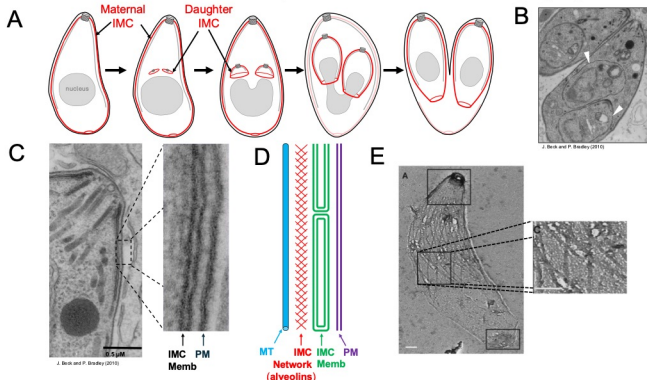
1) B.I.G. Summer Program, Institute for Quantitative and Computational Biosciences, UCLA; 2) Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA 90095; 3) Molecular Biology Institute, University of California, Los Angeles, CA 90095

BACKGROUND

Toxoplasma gondii is an important human pathogen

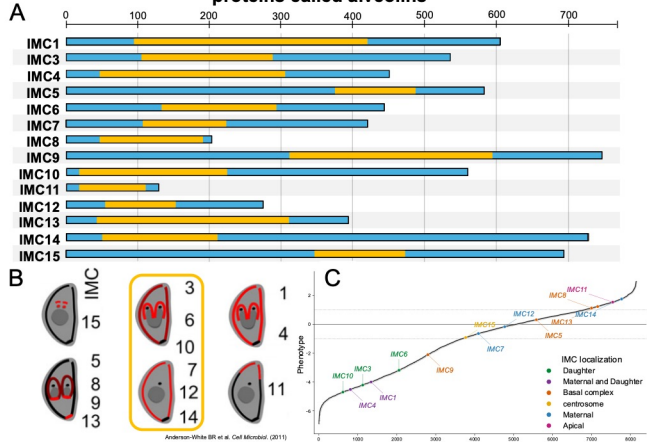
- A) *T. gondii* is a protozoan parasite from the phylum Apicomplexa, which includes other significant pathogens such as *Cryptosporidium* spp. (cryptosporidiosis) and *Plasmodium* spp. (malaria). *Toxoplasma gondii* is distributed worldwide, with some countries having a prevalence > 60%.
- B) Toxoplasmosis is a chronic disease that can cause severe and potentially fatal illness in immunocompromised individuals and neonates. Congenitally-infected infants often suffer from hydrocephaly (left), intellectual disability, and seizures. Immunocompromised individuals, such as AIDS or organ transplant patients, often present with chorioretinitis (center) and encephalitis (right).

The Inner Membrane Complex is a scaffold for endodyogeny



- A) *T. gondii* replicates through endodyogeny and this process is scaffolded by the IMC. Daughter parasites develop inside the maternal parasite with the daughter IMC. Nuclear division takes place after the daughter IMC formation. The maternal IMC is degraded as daughters mature and adopt the maternal plasma membrane.
- B) Transmission EM of replicating *Toxoplasma gondii*, the daughter IMC is highlighted with white arrows.
- C) Transmission EM showing the apical end of *T. gondii*. The periphery of the parasite visible in this image shows the inner membrane complex (IMC) and plasma membrane (PM).
- D) Diagram showing that the IMC consists of two distinct layers, the IMC membrane, which is composed of a series of flattened vesicles, and the IMC network, which is composed of an intermediate filament-like family of proteins called alveolins. A series of 22 microtubules underlies the IMC and provides support for the organelle.
- E) Transmission EM of a detergent-extracted parasite 'ghost' showing the IMC network and microtubules.

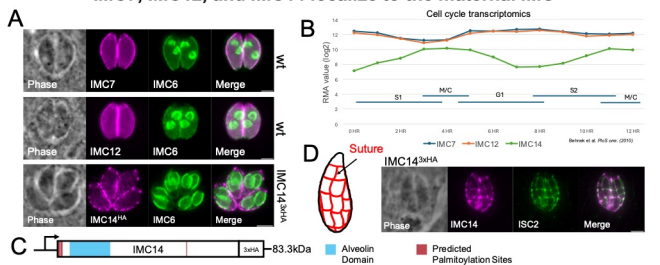
The IMC network is composed of a family of fourteen proteins called alveolins



- A) Diagram depicting the members of the alveolin protein family. Each alveolin has a region of proline-valine repeats called the alveolin domain (yellow). The alveolin domain is thought to facilitate binding between other members of the IMC network as well as to additional peripheral-localizing proteins.
- B) Diagram of the fourteen alveolins showing localization in daughter and maternal parasites. IMC3, IMC6, and IMC10 are enriched in the daughter IMC, whereby IMC7, IMC12, and IMC14 are only present in the maternal IMC.
- C) Phenotype score of alveolin proteins from a genome-wide CRISPR screen, a lower score indicates essential genes.

RESULTS

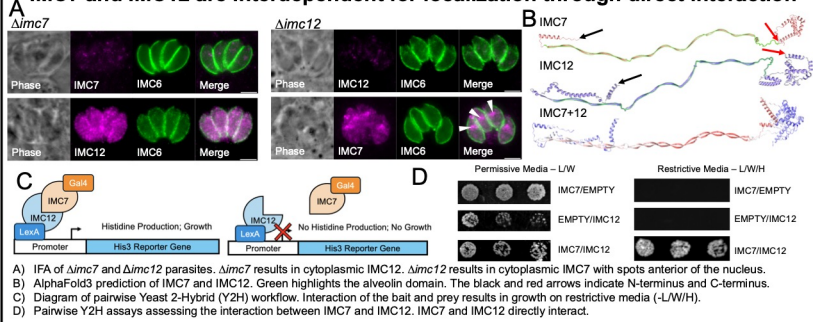
IMC7, IMC12, and IMC14 localize to the maternal IMC



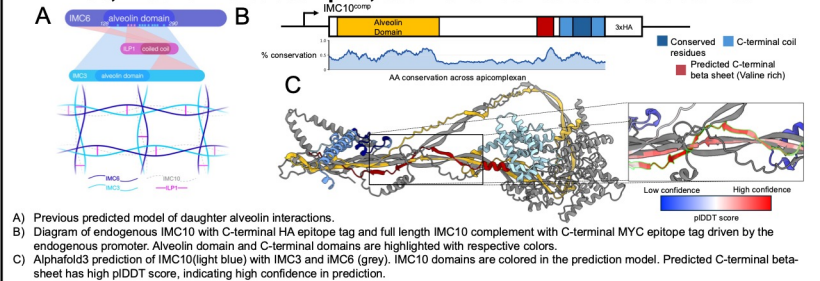
- A) IMC7, IMC12, and IMC14 localization by immunofluorescence (IFA) microscopy. The parasites were stained with anti-IMC7 and anti-IMC12. IMC14 stained with anti-HA after C-terminal integration of a 3xHA epitope tag. Scale bar is 2 μm.
- B) IMC7, IMC12, and IMC14 log2 normalized transcriptomics across the cell cycle.
- C) Diagram of IMC14 with C-terminal 3xHA epitope tag. Palmitoylation sites are predicted by CSS Palm 2.0
- D) Illustration of the IMC sutures of *T. gondii* and IMC14 localization by IFA. The parasites were stained with anti-HA and anti-ISC2. ISC2 is a known marker localized to the IMC suture structure of *T. gondii*.

RESULTS

IMC7 and IMC12 are interdependent for localization through direct interaction

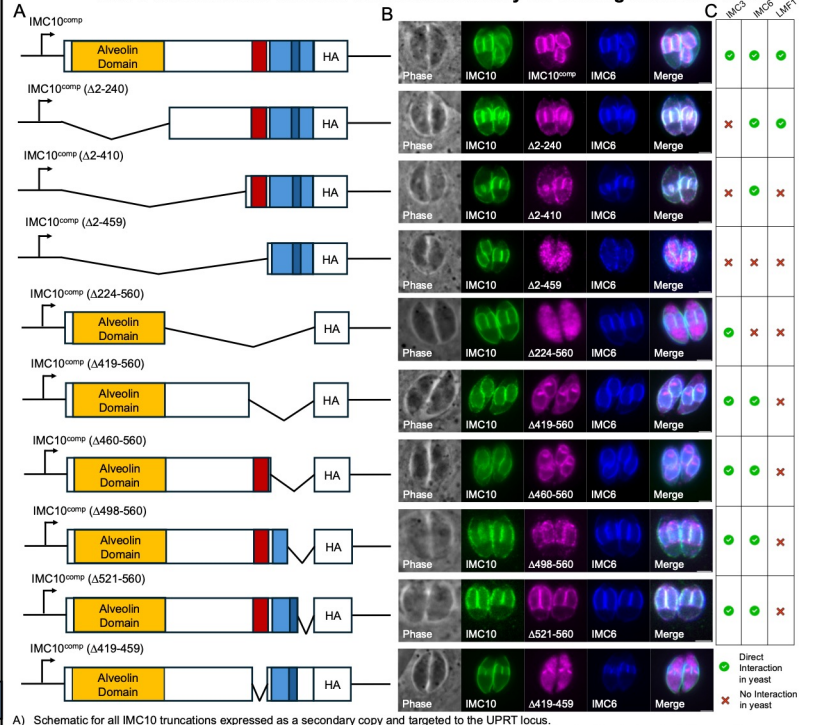


IMC3, IMC6 form a subcomplex, but IMC10 has not been characterized



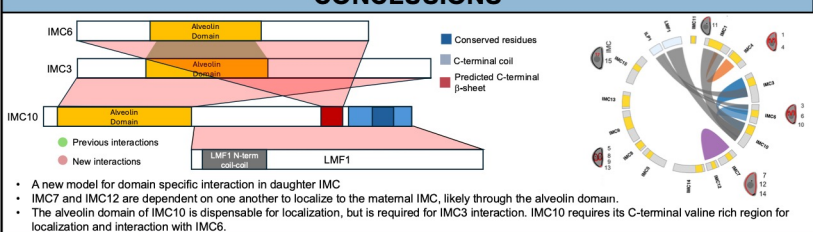
- A) Previous predicted model of daughter alveolin interactions.
- B) Diagram of endogenous IMC10 with C-terminal HA epitope tag and full length IMC10 complement with C-terminal MYC epitope tag driven by the endogenous promoter. Alveolin domain and C-terminal domains are highlighted with respective colors.
- C) AlphaFold3 prediction of IMC10 (light blue) with IMC3 and IMC6 (grey). IMC10 domains are colored in the prediction model. Predicted C-terminal beta-sheet has high pIDOT score, indicating high confidence in prediction.

The C-terminal beta sheet of IMC10 is necessary for binding and localization



- A) Schematic for all IMC10 truncations expressed as a secondary copy and targeted to the UPRT locus.
- B) IFA of the HA-tagged IMC10 deletion series stained with endogenous IMC10 and IMC6 demonstrate that the alveolin domain is dispensable.
- C) Pairwise Y2H assays assessing the interactions between the full length and truncated versions of IMC10 protein with IMC3 and IMC6 and LMF1.

CONCLUSIONS



- A) A new model for domain specific interaction in daughter IMC
- B) IMC7 and IMC12 are dependent on one another to localize to the maternal IMC, likely through the alveolin domain.
- C) The alveolin domain of IMC10 is dispensable for localization, but is required for IMC3 interaction. IMC10 requires its C-terminal valine rich region for localization and interaction with IMC6.

REFERENCES

Anderson-Wilms, B., & Kuo, P. D. (2018). A study of the interactions between the alveolin domain and the alveolin domain of *Toxoplasma gondii*. *Cellular Microbiology*, 18(1), 1-11. <https://doi.org/10.1111/cmi.13014>

Anderson-Wilms, B., & Kuo, P. D. (2018). A study of the interactions between the alveolin domain and the alveolin domain of *Toxoplasma gondii*. *Cellular Microbiology*, 18(1), 1-11. <https://doi.org/10.1111/cmi.13014>

Anderson-Wilms, B., & Kuo, P. D. (2018). A study of the interactions between the alveolin domain and the alveolin domain of *Toxoplasma gondii*. *Cellular Microbiology*, 18(1), 1-11. <https://doi.org/10.1111/cmi.13014>

Anderson-Wilms, B., & Kuo, P. D. (2018). A study of the interactions between the alveolin domain and the alveolin domain of *Toxoplasma gondii*. *Cellular Microbiology*, 18(1), 1-11. <https://doi.org/10.1111/cmi.13014>

Anderson-Wilms, B., & Kuo, P. D. (2018). A study of the interactions between the alveolin domain and the alveolin domain of *Toxoplasma gondii*. *Cellular Microbiology*, 18(1), 1-11. <https://doi.org/10.1111/cmi.13014>



Student: Kara Connolly
Faculty Advisor: Quen Cheng
Direct Mentor: Tanya Ivanov
Department: Medicine
<https://chenglab.dgsom.ucla.edu/>

Kara Connolly

**Biochemistry -
UCLA**

**Expected year
of graduation:
June 2031**

Pinpointing putative transcription factors that contribute to the regulatory logic behind β -glucan-induced innate immune memory

KARA CONNOLLY¹, Tanya Ivanov^{2,3}, Stefanie Brizuela^{2,3}, Quen Cheng³

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Molecular Biology Interdepartmental Graduate Program, UCLA

³ Department of Medicine, Division of Infectious Disease, UCLA

Macrophages, the sentinel immune cell, epigenetically encode memory of previous encounters through coordinated transcription factor (TF) activation, which reprograms the epigenome and dictates responses against secondary stimuli. β -glucan, a fungal derivative, has been shown to program enhanced responses to secondary stimuli in macrophages. However, the TFs regulating the β -Glucan-induced epigenomic state are unknown. To identify candidate TFs, we first found chromatin regions with altered accessibility post- β -glucan stimulation using existing Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) data. We identified 1604 regions with significantly increased accessibility during β -glucan treatment, of which 1264 regions remain differentially accessible 6 days after treatment. We use TF motif analysis to find enriched TF motifs in these regions with sustained reprogramming in order to reveal the putative regulatory logic behind β -Glucan-induced memory. Our findings contribute to the greater understanding of the exact mechanism of epigenetic reprogramming induced by β -glucan treatment.



Pinpointing putative transcription factors that contribute to the regulatory logic behind β -glucan-induced innate immune memory

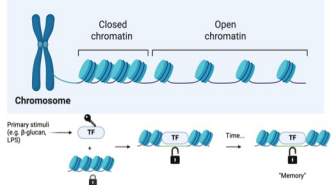
Kara Connolly¹, Tanya Ivanov², Stefanie Brizuela², Quen Cheng³

¹ 1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, ² Molecular Biology Interdepartmental Graduate Program, University of California, Los Angeles, ³ Department of Medicine, Division of Infectious Disease, University of California, Los Angeles



Abstract

Macrophages, the sentinel immune cell, epigenetically encode memory of previous encounters through coordinated transcription factor (TF) activation, which reprograms the epigenome and dictates responses against secondary stimuli. β -glucan, a fungal derivative, has been shown to program enhanced responses to secondary stimuli in macrophages. However, the TFs regulating the β -glucan-induced epigenomic state are unknown. Here we identify enriched TF motifs in regions with sustained reprogramming in order to reveal the putative regulatory logic behind β -glucan-induced memory.



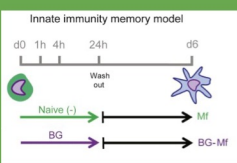
What is Known vs. Unknown

Known:

- Macrophages are multi-functional leukocytes of the innate immune system that encode epigenetic memory of prior stimuli.
- Epigenetic encoding entails the stimulus-prompted coordination of transcription factor activation impacting the accessibility of chromatin (i.e. memory) through histone modification.
- β -glucan training exhibits increased inflammatory defenses against secondary stimuli

Unknown:

- What transcription factor binding sites remain open (inhabited by factors) during the encoding of β -glucan trained memory?
- What transcription factors are involved in memory encoding?



Methods

To identify candidate TFs, we used edgeR to analyze chromatin accessibility from Assay for Transposase Accessible Chromatin (ATAC) sequencing data. By sorting based off of differentially expressed peaks through significant log fold change values ($LFC > 2$), we were then able to use HOMER to identify enriched motifs. All data sets were from Novakovic et al. 2016.

Discussion

- In β -glucan treated samples, there was retained openness of peaks through day 1 and day 6. It can be observed in the clusters that there was minimal overlap in peaks maintained after 6 days across both RPMI and β -glucan treatment.
- The HRE motif was found in activated peaks for both RPMI and β -glucan 1 and 6 days after treatment. This indicates potential contribution to memory formation.
- Transient genes with partial presence show closing after 6 days, whereas memory associated peaks are opened. From this we are able to illustrate the dynamism of peaks associated with memory.

Future Directions

- Our current data reflects chromatin accessibility, but does not show how this accessibility affects gene expression. Future analysis will include pairing this ATAC analysis with a matching RNA-seq dataset.
- To confirm candidate TFs identified through motif analysis, conducting western blot analysis on the identified regulatory, inflammatory, and potential memory-inducing TF factors could confirm their presence and role in memory formation.

Acknowledgements

I would like to acknowledge the QCBio collaborative workshops for their wealth of resources and crucial guidance. Additionally, I'd like to acknowledge my BIG summer peers for their advice and input towards troublesome code.

References

- Glass, C. K. & Natoli, G. Molecular control of activation and priming in macrophages. *Nat Immunol* 17, 26-33 (2015).
- Novakovic, B. et al. β -Glucan Reverses the Epigenetic State of LPS-Induced Immunological Tolerance. *Cell* 167, 1354-1368.e14 (2016).

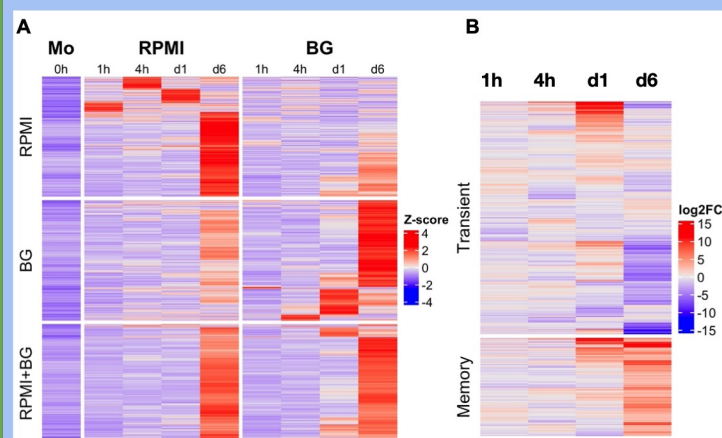


Figure 1. Chromatin accessibility analysis of treatment response peaks. A) Heatmap of 5,235 regions with significantly increased accessibility compared to monocytes (identified with a $\log_{2}FC > 2$). Distinct trends in chromatin accessibility reveal 1,760 peaks unique to β -glucan treatment (BG cluster), 1,748 peaks unique to no treatment (RPMI cluster), and 1,727 peaks shared between both groups. Each cluster is k-means clustered. B) Of the 5,235 regions, 3,682 regions clustered were 'transient', or had a $LFC < 2$ on day 6, while 1,553 regions held 'memory' of β -glucan treatment ($LFC > 2$ on day 6). Regions in the transient cluster show increase in accessibility early time points, but return toward baseline and memory-associated regions exhibit sustained or enhanced accessibility at later time points (d6 most significantly). The increasing accessibility observed in the memory-associated cluster suggests persistent chromatin remodeling following β -glucan exposure.

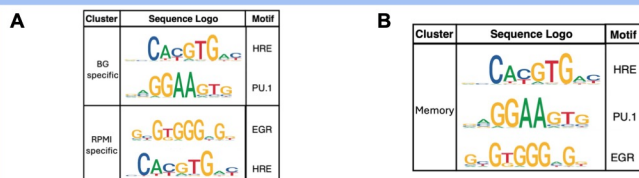


Figure 3. Analysis of enriched TF motifs from β -glucan treatment. BG-specific and RPMI-specific peaks exhibit distinct enriched sequence motifs, suggesting that different transcriptional regulators may contribute to the condition-specific chromatin accessibility patterns. Hypoxia-response element (HRE) was noted as a commonality between both clusters. B) Sequence logos of transcription factor motifs within the 'memory'-associated chromatin accessibility cluster identified in β -glucan ATAC seq. The presence of HRE, PU.1, and EGR motifs suggests that these transcription factors may contribute to the persistent chromatin remodeling associated with β -glucan induced innate immune memory.

Student: Maggie De Leon

Faculty Advisor: Louis Bouchard

Department: Chemistry and Biochemistry

<https://www.chemistry.ucla.edu/directory/bouchard-louis-s/>

Residual Diffusion for Annotation-Free Electrocardiogram Window Denoising

MAGGIE DE LEON³, Abhinav Garg¹, Meicheng Wang², Jeannine Yu⁴, Louis-S. Bouchard⁵

¹ Department of Computer Science, UCLA

² Department of Mathematics, UCLA

³ BIG Summer Program, Institute for Society and Genetics, UCLA

⁴ Department of Molecular, Cell, and Developmental Biology, UCLA

⁵ Departments of Chemistry and Biochemistry and Bioengineering; Jonsson Comprehensive Cancer Center; Molecular Biology Institute, UCLA

Electrocardiogram (ECG) denoising methods must suppress baseline wander and structured noise without distorting waveform morphology. However, classical ECG denoising methods, despite being fast, stable, and easy to deploy, cannot easily adapt to the diverse morphology and interference patterns across records. We present O2, a compact anchor-conditioned residual diffusion denoiser for electrocardiogram windows. O2 learns a residual correction around an infinite impulse response (IIR) filter anchor rather than a full noisy-to-clean transformation. The neural model uses no manual timing annotations, no QRS/P/T labels, no beat-boundary labels, and no record identity as inputs. Our results show that compact anchor-conditioned residual diffusion can improve electrocardiogram window denoising accuracy, harm rate, and tail behavior on the DeepFilter/DeScoD benchmark without requiring annotation-derived inference features.

Maggie De Leon

MIMG - UCLA

Expected year of graduation: June 2028



O2: IIR-Anchored Flow Matching for ECG Denoising

Maggie De Leon^{1†}, Abhinav Garg^{2†}, Meicheng Wang^{3†}, Jeannine Yu^{4†}, Louis-S. Bouchard⁵

¹Institute for Society and Genetics; ²Department of Computer Science; ³Department of Mathematics;

⁴Department of Molecular, Cell, and Developmental Biology; ⁵Departments of Chemistry and Biochemistry, Jonsson Comprehensive Cancer Center, and Molecular Biology Institute, University of California, Los Angeles [†]Equal contribution

UCLA
Chemistry &
Biochemistry

Introduction

Electrocardiogram (ECG) recordings are routinely contaminated by baseline wander, electrode-motion interference, muscle noise, and other structured disturbances. Denoising must therefore reduce error while preserving the P-wave, QRS complex, and T-wave, rather than smoothing them away. We present O2, a compact IIR-anchored residual denoiser trained by conditional flow matching. A fixed IIR filter supplies a stable anchor, while a convolutional velocity network estimates the remaining normalized clean-signal correction.

- No beat detector or annotation-derived model input at inference
- One 345,409-parameter model serves both corruption sets
- Regular-grid overlap-add defines a longer-record extension
- The compact design may facilitate future portable implementation

Methods

I. Denoising pipeline: noisy window to clean recording

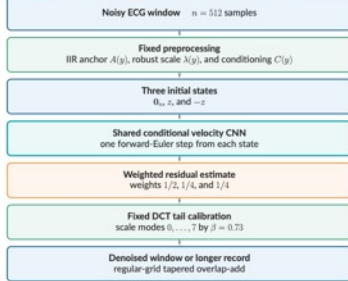


Figure 1. O2 pipeline. Fixed operations are green; the learned velocity network is teal. The model interface requires only noisy samples and deterministic transforms.

II. Random-variable model

On a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $(X, Y) : \Omega \rightarrow \mathbb{R}^n \times \mathbb{R}^n$ be a paired clean-noisy training window with $n = 512$. The variables T and Z satisfy $\mathcal{L}(T) = \text{Unif}[0, 1]$ and $\mathcal{L}(Z) = \mathcal{N}_c(0, \sigma_{T|Z}^2)$ and are mutually independent of (X, Y) . The maps $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\lambda : \mathbb{R}^n \rightarrow (0, \infty)$, and $C : \mathbb{R}^n \rightarrow \mathbb{R}^{4 \times n}$ are respectively the fixed IIR anchor, robust scale, and conditioning map.

$$\begin{aligned} R(\omega) &= X(\omega) - A(Y(\omega)), \\ B_t(\omega) &= (1 - t)Z(\omega) + IR(\omega), \\ \mathcal{J}_{\text{RM}}(\theta) &= \mathbb{E}_n \left[\|v_t(B_t, C(Y), T) - (R - Z)\|_2^2 \right]. \end{aligned}$$

θ parameterizes the CNN $v_t : \mathbb{R}^n \times \mathbb{R}^{4 \times n} \times [0, 1] \rightarrow \mathbb{R}^n$; θ^* is obtained by regularized L-BFGS refinement of 1,905 output-head scalars after AdamW training.

III. Three-state estimator at inference

$$\begin{aligned} z &= Z(\omega), \\ \hat{\mu}_0(y) &= v_0(0, C(y), 0), \\ \hat{\mu}_s(y, \omega) &= z + v_s(z, C(y), \theta), \\ \hat{\mu}_t(y, \omega) &= [\hat{\mu}_0(y) + \hat{\mu}_s(y, \omega) + \hat{\mu}_t(y, \omega)], \\ \hat{X}^{\text{O2}}(y, \omega) &= D^T \Lambda_{0.73} D \{A(y) + \lambda(y) \hat{\mu}_t(y, \omega)\}. \end{aligned}$$

The three $\hat{\mu}$ terms are one-step forward-Euler endpoint estimates. D is the orthonormal DCT-II matrix; $\Lambda_{0.73}$ scales modes $0, \dots, 7$ by 0.73 and leaves all other modes unchanged. AdamW training uses 30,000 updates and 128-window mini-batches.

Results

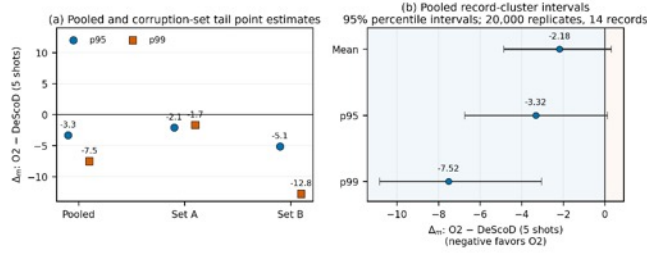


Figure 2. Record-level uncertainty and tail behavior. Paired 95% record-cluster intervals use 20,000 replicates of 14 ECG records. All three point differences favor O2; the p99 interval excludes zero, whereas the mean and p95 intervals include zero. Lower PRDN is better.

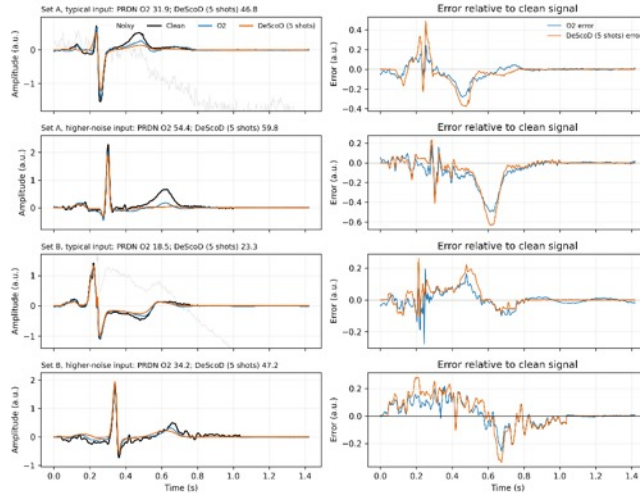


Figure 3. Representative ECG reconstructions. Cases span both corruption sets and distinct records and were selected without access to either denoiser's output or error. Deployed O2 has lower clean-centered PRDN in every displayed case.

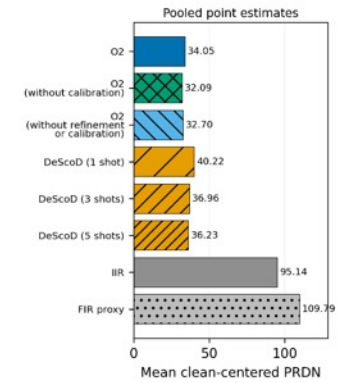


Figure 4. Pooled mean clean-centered PRDN. The plot compares O2 configurations, every public DeScod shot setting, and deterministic filtering over 26,632 matched conditions. Lower is better.

Principal findings

- Against public five-shot DeScod, all nine pooled point estimates favored O2; the same ordering held within both corruption sets
- O2 attained mean/median/p95/p99 PRDN of 34.05/31.02/63.76/85.25 versus 36.23/32.62/67.08/92.76
- The head-refined base model had lower record-mean PRDN than five-shot DeScod on 10 of 14 records (Holm-adjusted $p = 0.0148$) and than each of five additional learned baselines on all 14 records
- One O2 model has 345,409 parameters, 17.9% of one DeScod model, and receives no annotation-derived input at inference

Scope and limitations

- Development-selected DCT calibration lowers p95 and p99 PRDN but raises mean PRDN relative to uncalibrated O2
- Quantitative tests use annotation-derived, 512-sample benchmark windows from 14 ECG records; arbitrary-phase continuous recordings remain to be evaluated
- The reported implementation is buffered and noncausal; real-time throughput and portable-device performance have not been measured

Code

Source code and reproduction instructions:
<https://github.com/lbouchard/O2>

References and acknowledgements

- H. L. G. Dittler, J. Revilla, and A. Li, "DeScod ECG: Deep score-based diffusion model for ECG baseline wander and noise removal," *IEEE J. Biomed. Health Inform.*, 28(9), 5081–5091, 2024.
- Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Li, "Flow matching for generative modeling," *ICML*, 2023.
- Supported by the BIC Summer Program, Institute for Quantitative and Computational Biosciences.



Student: Ben Dewey
Faculty Advisor: Neil Harris
Direct Mentor: Azad Azargushasb
Department: Neurosurgery
<https://neilharrislab.dgsom.ucla.edu/>

Ben Dewey

**Computational
and Systems
Biology -
UCLA**

**Expected year
of graduation:
June 2028**

Behavioral Signatures of Repeated Mild Traumatic Brain Injury: Classifying Injured and Sham Mice from Odor-Reversal Reward Learning

BEN DEWEY¹, ISHYA BAHL¹, Azad Azargushasb², Sam Vander Dussen², Neil G. Harris³

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² UCLA Brain Injury Research Center, Department of Neurosurgery, DGSOM, UCLA

³ Intellectual Development and Disabilities Research Center, UCLA

Repeat mild traumatic brain injury (rmTBI) is associated with executive deficits such as impaired inhibitory control and cognitive flexibility. However, the primary behavioral deficit is not fundamentally understood. This project determines whether a mouse is sham or injured using models trained on behavioral inputs during an odor reversal in a probabilistic go/no-go paradigm. 26 mice (13 injured, 13 sham), which previously reached a learned association, were monitored over seven reversal-learning days. Features across specific task periods, like lick rate, velocity, and accuracy, were used to train various classifiers with cross-validation. Using behavioral data alone, most models could not classify sham versus injured mice above 80% accuracy although some models demonstrated a high area under the curve. Analyzing corresponding functional ultrasound imaging (fUSi) data can help identify the compensatory mechanisms in the brain that allow the injured mice to perform nearly the same.



Behavioral Signatures of Repeat Mild Traumatic Brain Injury: Classifying Injured and Sham Mice from Odor-Reversal Reward Learning

Ishya Bahl¹, Ben Dewey¹, Azad Azargushasb², Samuel Vander Dussen², Neil G. Harris³

¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

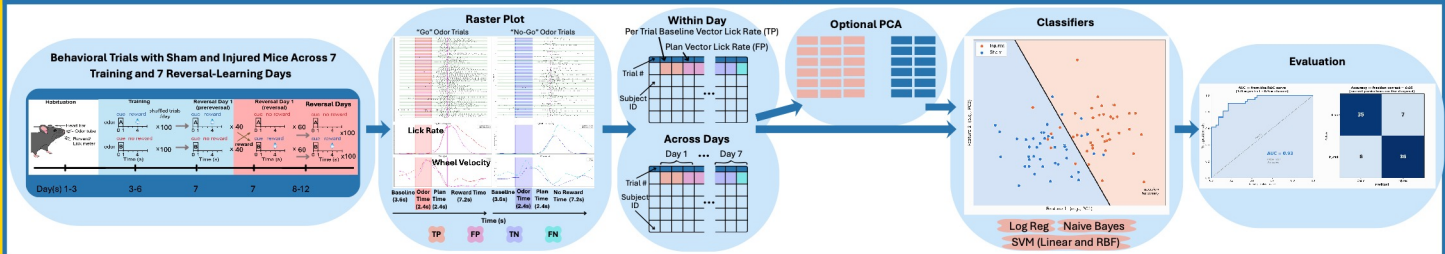
²UCLA Brain Injury Research Center, Department of Neurosurgery, DGSOM, UCLA

³Intellectual Development and Disabilities Research Center, UCLA



*Co-authors

Methodology



Background

- Repeat mild traumatic brain injury (mTBI) produces lasting executive deficits, most consistently in inhibitory control and cognitive flexibility.
- Rodent models confirm these deficits on reversal-learning tasks, but rely on group-level summary measures (percent correct, trials to criterion) that show injured animals differ without identifying the primary deficit.
- Injured mice often reach near-normal performance despite documented pathology, suggesting compensation that averaged behavior cannot capture.
- Trial-level features segmented by task period retain information those summaries discard, so we ask whether a classifier can recover injury status from behavior alone, and use concurrent fUSI to ask what the brain is doing when behavior looks intact.

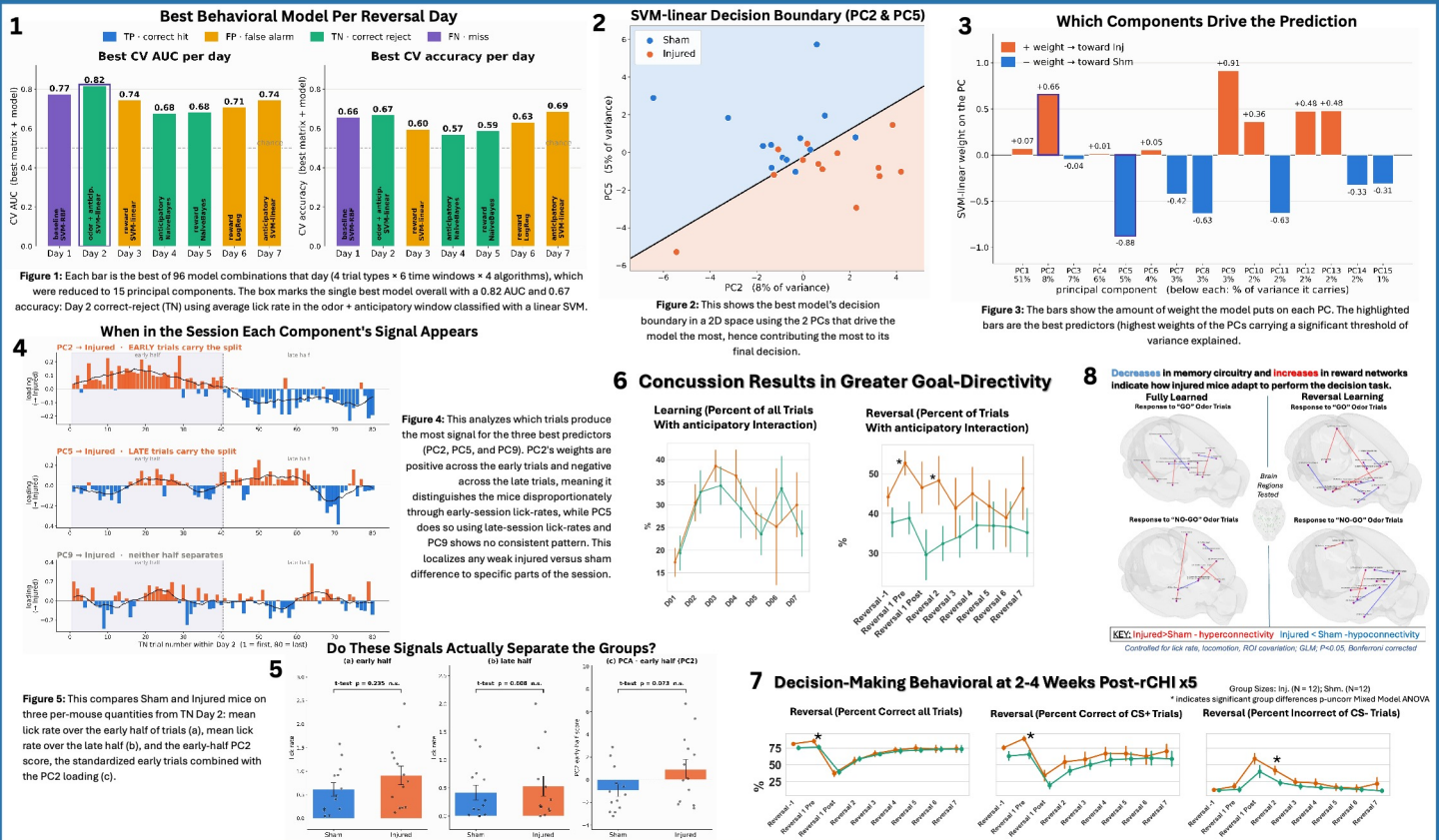
Abstract

Repeat mild traumatic brain injury (mTBI) is associated with executive deficits such as impaired inhibitory control and cognitive flexibility. However, the primary behavioral deficit is not fundamentally understood. This project determines whether a mouse is sham or injured using models trained on behavioral inputs during an odor reversal in a probabilistic go/no-go paradigm. 26 mice (13 injured, 13 sham), which previously reached a learned association, were monitored over seven reversal-learning days. Features across specific task periods, like lick rate, velocity, and accuracy, were used to train various classifiers with cross-validation. Using behavioral data alone, most models could not classify sham versus injured mice above 80% accuracy although some models demonstrated a high area under the curve. Analyzing corresponding functional ultrasound imaging (fUSI) data can help identify the compensatory mechanisms in the brain that allow the injured mice to perform nearly the same.

Discussion / Future Works

- All classifiers failed to exceed 80% accuracy on behavior alone, indicating that injured mice perform this task closely enough to shams that trial-level behavior carries limited discriminative signal.
- The high AUC in some models despite modest accuracy suggests the separating information is real but weakly expressed, likely concentrated in a subset of animals, task periods, or reversal days rather than distributed across the full dataset.
- Near-normal performance in the face of documented injury points to compensation; analyzing the corresponding fUSI data will test whether injured mice recruit additional or alternative circuitry to sustain it.
- Next steps: identify which features and task periods drive the models that do separate the groups, examine whether classification improves on early reversal days when flexibility demands peak, and evaluate whether adding fUSI features is more capable of delineating between the two groups

Results





Student: Ilinca Flacau
Faculty Advisor: Steffen Lindert
Direct Mentor: Zach Drake
Department: Chemistry and Biochemistry
<https://lindertlab.chem.ucla.edu/>

Ilinca Flacau

**Computational
Biology -
UCLA**

**Expected year
of graduation:
June 2027**

Deep-learning protein structure prediction using multi-assay deep mutational scanning

ILINCA FLACAU¹, Zachary Drake², Steffen Lindert²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Chemistry and Biochemistry, UCLA

Recent advances in protein structure prediction achieve high accuracy but rely primarily on evolutionary sequence information. Deep mutational scanning (DMS) provides complementary experimental data by measuring how mutations affect protein function and structure. DMS-Fold showed that incorporating folding stability improves structure prediction, but it is unclear whether this approach generalizes to other types of DMS assays. We investigated whether DMS-Fold remains effective when applied to experimental protein fitness scores from the Human Domainome 1 library and simulated fitness scores generated with ESM1v and ProteinMPNN. We assessed model performance by comparing predicted structures to experimental structures using RMSD and TM-score metrics. DMS-Fold improved structure prediction accuracy over AlphaFold2 for 77% of proteins using experimental fitness data and 74% using simulated fitness data. These findings demonstrate that DMS-Fold successfully leverage fitness-based measurements, expanding the applicability of DMS-guided structure prediction and highlighting the potential of integrating diverse experimental data types into protein modeling.

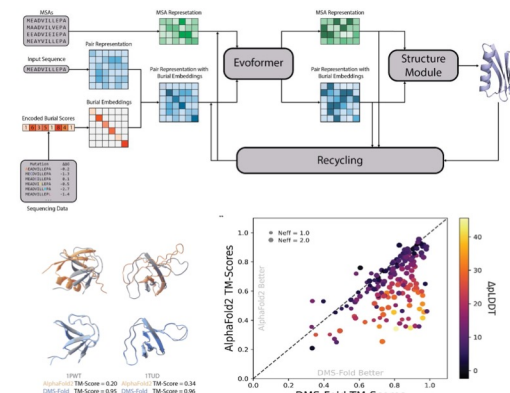
Deep-learning protein structure prediction using multi-assay deep mutational scanning

Ilinca Flacau¹, Zachary Drake², and Steffen Lindert²

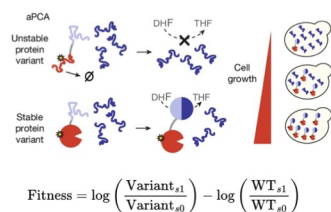
¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA, ²Department of Chemistry and Biochemistry, UCLA,

Background

DMS-Fold integrates $\Delta\Delta G$ deep mutational scanning data into AlphaFold2 to improve structure prediction accuracy

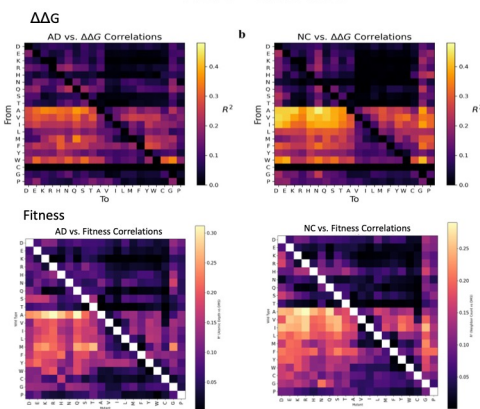


Human Domainome 1 (HD1) quantifies fitness via abundance protein fragment complementation assay deep mutational scanning

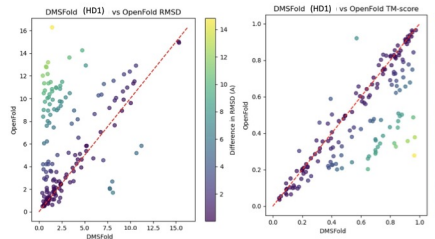


Human Domainome 1 Fitness

DMS-Fold burial metrics are similarly correlated with $\Delta\Delta G$ and HD1 fitness scores

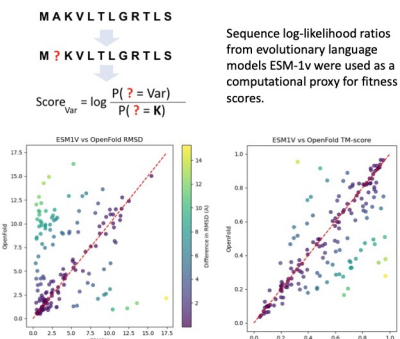


Use of fitness data in DMS-Fold improves protein structure prediction for 77% of proteins



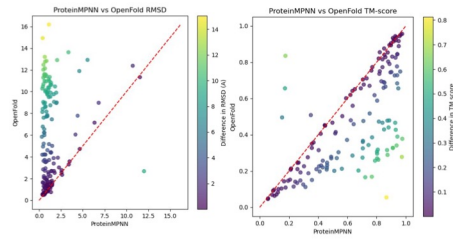
Simulated Fitness

ESM-1V derived fitness score improves protein structure prediction for 65% of proteins



ProteinMPNN derived fitness scores improves protein structure prediction for 87% of proteins

ProteinMPNN amino acid sequence from 3d backbone log-probability differences were used as a computational proxy for fitness scores.
 Fitness Score = $\log P(\text{Mutant} | \text{Structure}) - \log P(\text{WT} | \text{Structure})$





Student: Lipika Goel
Faculty Advisor: Aaron Meyer
Direct Mentor: Joseph Schirle
Department: Bioengineering
<https://asmlab.org/>

Lipika Goel

**Computational
Biology -
UCLA**

**Expected year
of graduation
June 2028**

Building a dynamical systems model to predict cellular response to genetic perturbation

LIPIKA GOEL^{1, 2}, Maxence Larour-Debonneuil³, JJ Schirle⁴, Aaron Meyer⁵

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Computational and Systems Biology Interdepartmental Program, UCLA

³ Chemical & Biomolecular Engineering Department, Samueli School of Engineering, UCLA

⁴ Department of Computational Medicine, UCLA

⁵ Department of Bioengineering, Samueli School of Engineering, UCLA

A fundamental question in systems biology is how a cell responds dynamically to stimuli, warranting the development of virtual cell models that predict cellular state across contexts. In particular, modeling the genetic response to cell perturbation potentially provides insight into disease and therapeutic mechanisms. We adapt Cellbox, an interpretable machine learning model for perturbation biology, for the purpose of building a virtual cell. By inferring regulatory interactions between genes, the model predicts the change in individual genes' expression resulting from a single-gene knockdown. We find that when inferring interactions between genes, the model matches or improves training loss over a baseline model with no gene interactions, highlighting its ability to learn biologically relevant interactions. However, model performance varies greatly when predicting the genetic response to unseen perturbations, indicating room for improvement in the generalizability of the interactions it learns.

Building a dynamical systems model to predict cellular response to genetic perturbation

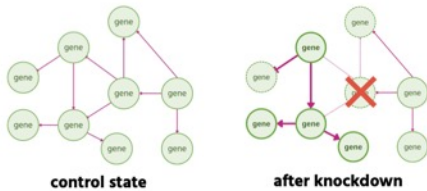
LIPIKA GOEL^{1,2}, Maxence Larour-Debonneuil³, JJ Schirle⁴, Aaron Meyer⁵

¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, ²Computational and Systems Biology Interdepartmental Program, ³Chemical and Biomolecular Engineering Department, ⁴Department of Computational Medicine, ⁵Department of Bioengineering
University of California, Los Angeles

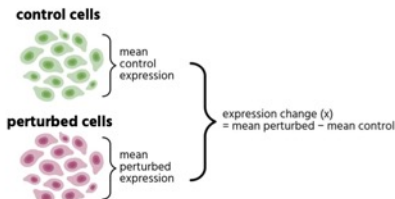
Abstract

A fundamental question in systems biology is how a cell responds dynamically to stimuli, warranting the development of virtual cell models that predict cellular state across contexts. In particular, modeling the genetic response to cell perturbation potentially provides insight into disease and therapeutic mechanisms. We adapt Cellbox, an interpretable machine learning model for perturbation biology, for the purpose of building a virtual cell. By inferring regulatory interactions between genes, the model predicts the change in individual genes' expression resulting from a single-gene knockdown. We find that when inferring interactions between genes, the model matches or improves training loss over a baseline model with no gene interactions, highlighting its ability to learn biologically relevant interactions. However, model performance varies greatly when predicting the genetic response to unseen perturbations, indicating room for improvement in the generalizability of the interactions it learns.

Virtual cells dynamically model cell state, predicting how a cell will respond to certain stimuli or contexts. **We are using a virtual cell model to predict genetic changes in response to gene perturbation.**



The Cellbox model (Yuan et al.) infers intracellular molecular interactions to predict a molecular or phenotypic response to cell perturbation. We adapt this, using gene interactions to predict the change in gene expression following a single-gene knockdown.



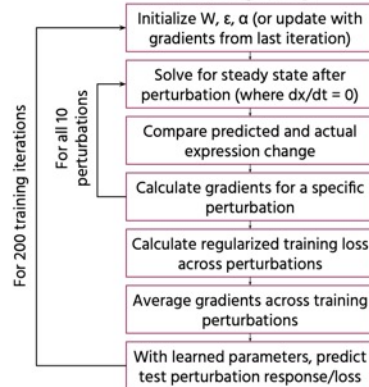
Cell expression values come from a single-cell perturbation-sequencing dataset of human embryonic stem cells. CRISPR-i was used to knock down specific target genes before measurement.

Cellbox Model Equation

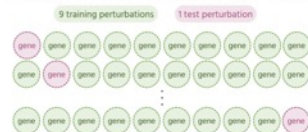
$$\frac{d\vec{x}}{dt} = \vec{\epsilon}\phi(W\vec{x} + \vec{u}) - \alpha\vec{x}$$

change in expression (steady state when $dx/dt = 0$) max expression vector scales perturbation term between -1 and 1 gene interaction matrix expression change caused by perturbation perturbation vector mRNA decay rate

Training Loop



Jackknifing (Leave-One-Out)



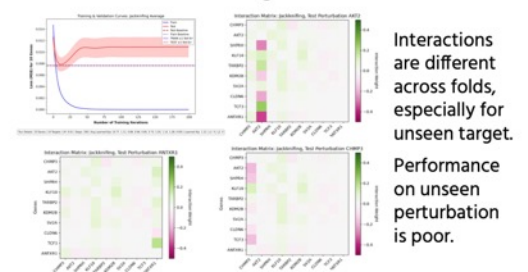
How is the model doing **on average** across this set of genes? How much does each training perturbation affect inferred interactions?

References & Acknowledgements

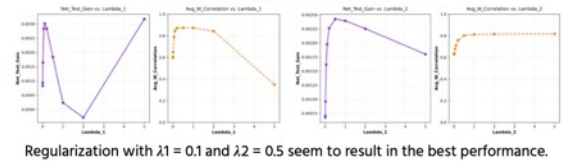
Yuan, B., Shen, C., Luna, A., Korkut, A., Marks, D. S., Ingraham, J., & Sander, C. (2020). CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy. *Cell Systems*, 12(2), 128-140.e4. <https://doi.org/10.1016/j.cels.2020.11.013>

Thank you to JJ and Dr. Meyer for all their guidance, Max for helping me figure out Cellbox, and BIG Summer for helpful workshops and seminars.

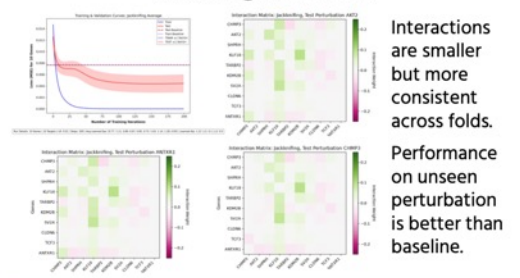
Without Regularization



Regularizing the Model



With Regularization



Discussion

Cellbox can learn meaningful interactions, at least at a small scale.

Future directions: Testing model at bigger scales, using prior information about the gene regulatory network to constrain W .



Student: Jaiden Ha

Faculty Advisor: Kalli Kappel

Department: Chemistry and Biochemistry

<https://sites.google.com/view/kappelab/>

Jaiden Ha

**Statistics and
Data Science -
UCLA**

**Expected year
of graduation
June 2027**

Generative Design of Condensate-Forming IDR Sequences via Diffusion Modeling Layered with Oracle-Guided Finetuning

JAIDEN HA^{1,2}, Ihyun Jeon^{1,2}, Mia Lesser², Bryan Zhang², Claire Hsieh^{2,3}, Kalli Kappel²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Chemistry and Biochemistry, UCLA

³ Bioinformatics Interdepartmental PhD Program, UCLA

Biomolecular condensates formed via liquid-liquid phase separation play critical roles in cellular organization, yet designing intrinsically disordered region (IDR) sequences with targeted condensate-forming properties remains challenging due to the poorly understood mapping between sequence and phase behavior. This project explores whether generative models can reliably construct IDR sequences with desired condensate formation and expression characteristics. Our approach uses a neural network fusion classifier, integrating Protein Language Model (PLM) embeddings with biophysical features, as a reward signal guiding two separate fine-tuning extensions of an Order-agnostic Autoregressive Diffusion Model (OADM): Relative Trajectory Balance (RTB) and Direct Reward bAckpropagation with gumbEl Softmax trick (DRAKES). Candidates are filtered by expression intensity, disorder propensity via IUPred3, and ensemble-based predictive uncertainty to prioritize diverse, under-explored sequence space. This framework lays groundwork for computationally guided condensate design, with planned active learning cycles feeding validated candidates back into training data to refine the oracle models.



Generative Design of Condensate-Forming IDR Sequences via Diffusion Modeling Layered with Oracle-Guided Finetuning

JAIDEN HA^{1,2}, Ihyun Jeon^{1,2}, Mia Lesser², Bryan Zhang², Claire Hsieh^{2,3}, Kalli Kappel²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, ² Department of Chemistry and Biochemistry, ³ Bioinformatics Interdepartmental PhD Program, University of California Los Angeles

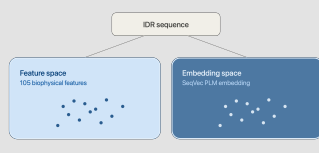
UCLA B.I.G Summer
Bruins in Genomics

Abstract

Biomolecular condensates formed via liquid-liquid phase separation play critical roles in cellular organization, yet designing intrinsically disordered region (IDR) sequences with targeted condensate-forming properties remains challenging due to the poorly understood mapping between sequence and phase behavior. This project explores whether generative models can reliably construct IDR sequences with desired condensate formation and expression characteristics. Our approach uses a neural network fusion classifier, integrating Protein Language Model (PLM) embeddings with biophysical features, as a reward signal guiding two separate fine-tuning extensions of an Order-agnostic Autoregressive Diffusion Model (OADM): Relative Trajectory Balance (RTB) and Direct Reward backpropagation with gumbel Softmax trick (DRAKES). Candidates are filtered by expression intensity, disorder propensity via IUPred3, and ensemble-based predictive uncertainty to prioritize diverse, under-explored sequence space. This framework lays groundwork for computationally guided condensate design, with planned active learning cycles feeding validated candidates back into training data to refine the oracle models.

Materials

The 14,622 CondensSeq IDR sequences are represented in two ways: a 105-feature space and a SeqVec embedding space.



Methodology

Mutual information and Spearman correlation ranked all 105 features (Fig. 1); charge properties (isoelectric point, NCPR, charge ratio) ranked highest, matching CondensSeq's charge-driven condensate finding.

An intermediate-fusion network (Fig. 2) combines feature and embedding branches via cross-attention to classify condensate formation, trained on CondensSeq's binary label; SeqVec achieved 0.877 accuracy.



Figure 2: Feature and embedding branches are fused via cross-attention to predict condensate formation probability.

An 8-model Conformalized Quantile Regression (CQR) ensemble predicts calibrated lower-bound intensity, prioritizing specificity over point-estimate regression to filter candidates unlikely to express well in cells (recall 69.4%, specificity 88.1%).

An order-agnostic diffusion model (OADM), trained via masked-token prediction on 14,622 sequences, generates the baseline distribution guided methods build from. At each step, a random subset of positions is masked and the model minimizes the loss function (shown below) to recover the true residue from context; at generation time, this fill-in-the-blank process repeats, revealing one position at a time until the sequence is complete.

$$\mathcal{L}_{OADM} = -\log P_{\theta}(x_{t+1}^{true} | x_{masked})$$

Three steering strategies compared (Fig. 3): Monte Carlo Best-of-N (oversample, select by score); relative trajectory balance (RTB) fine-tunes OADM toward a distribution-anchored, reward-resampled target; and Direct Reward backpropagation with gumbel Softmax (DRAKES) backpropagates with differentiable discrete sampling, relying on a proxy of the oracle and KL-regularization.

Results

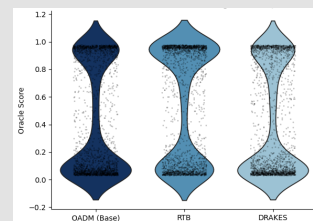


Figure 3: Violin plots of the oracle-predicted condensate-formation probabilities for the OADM set and both fine-tuning model sets, with individual sequence scores overlaid as points.

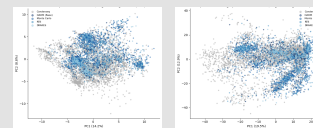


Figure 4: Sequences with experimental or predicted condensate-formation scores above 0.3 were projected via PCA in both feature space and embedding space for the high-scoring subsets (assigned an oracle score > 0.3).

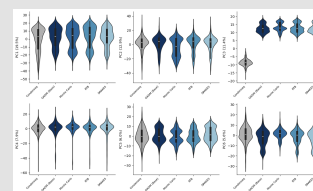


Figure 5: Marginal distribution plots of the principal components explaining more than 5% of variance in the embedding space of the high-scoring subset.

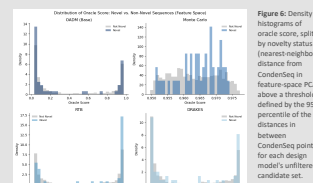


Figure 6: Density histograms of oracle score, split by novelty status (nearest-neighbor distance from CondensSeq in feature-space PCA above a threshold defined by the 95th percentile of the distances in between CondensSeq points), for each design model's unfiltered candidate set.

Conclusion

- Generated sequences from all four design strategies occupy the same region of both feature space and embedding space as real high-scoring CondensSeq sequences (Fig. 4).
- RTB meaningfully optimizes for condensate formation, with the density of sequences within the generated set predicted to form condensates being significantly higher than that of OADM (57.6% vs. 44.3%, a 1.30x increase). The distribution of DRAKES sequences does not improve meaningfully in that regard (Fig. 3).
- RTB is also shown to generate novel sequences that are predicted to be likely to form condensates by the oracle at a much higher rate than that of base OADM and DRAKES (Fig. 7)

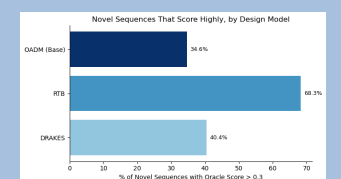


Figure 7: Percentage of feature-space-novel sequences (nearest-neighbor distance to CondensSeq above the 95th percentile threshold) with oracle score exceeding 0.3, shown per design model excluding Monte Carlo Best-of-N.

Future Directions

- The primary intended use of these generated sequences is to feed top-scoring novel candidates into wet-lab experimentation to be iteratively added to the original training datasets and analyze the effects on existing prediction models.
- The immediate next step is to implement a reward signal representing novelty, either by optimizing for further distance from sequences in CondensSeq or by steering towards sequences the oracle model predicts on with higher uncertainty.
- Investigate the specific principal components where generated sequences still diverge from real CondensSeq sequences, even within the high-scoring subset, to determine whether this reflects a real biological gap or a generation-model limitation."

References

- Kappel et al. (2025). Characterizing protein sequence determinants of nuclear condensates by high-throughput pooled imaging with CondensSeq. *Nature Methods*, 22, 1464-1475.
- Hooftboom et al. (2022). Autoregressive diffusion models. *ICLR*. arXiv:2110.02037
- Venkatesan et al. (2024). Assessing interpretable inference in diffusion models for vision, language, and control. *NeurIPS* 37. arXiv:2405.20871
- Wang et al. (2023). Fine-tuning discrete diffusion models via reward optimization with applications to DNA and protein design. *ICLR*. arXiv:2403.12643

Introduction

Building on CondensSeq, a high-throughput dataset of 14,622 experimentally characterized intrinsically disordered region (IDR) sequences, this project develops a generative pipeline, pairing an order-agnostic diffusion model with oracle-guided fine-tuning to design condensate-forming IDR sequences computationally.

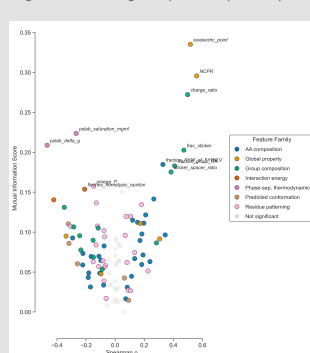


Figure 1: Univariate feature importance to target, represented by mutual information, clustered in feature categories.



Student: Elsa Hartley

Faculty Advisor: Ben Knowles

Department: Ecology and Evolutionary Biology

<https://sites.lifesci.ucla.edu/eeb-knowleslab/>

Elsa Hartley

**Computational
Biology -
UCLA**

**Expected year
of graduation:
June 2027**

Comparison of metagenome-derived bacterial growth rate and bacteriophage lifestyles across environments

ELSA HARTLEY^{1,2} and Ben Knowles^{1,2}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Ecology and Evolutionary Biology, UCLA

Viral infection shapes bacterial communities from the human gut to the open ocean. Viruses can lyse cells after infection or persist chronically. Here, we test whether bacterial growth proxies scale with viral lifestyle across environments by integrating metagenomic assembly of previously published libraries, MAG reconstruction, viral contig lifestyle classification, and growth rate inference using Peak-to-Trough (PTT) and Codon Usage Bias (CUB) methods. These growth proxies use the ratio of origin-adjacent reads to non-adjacent reads (PTT) or of tRNA genes to the codons that require them in the MAG (CUB) as measures of proliferation. Viral lifestyle is predicted through hallmark genes and genomic signatures. We have implemented this pipeline on a preliminary human gut sample set and are extending it to soil and marine datasets. This framework will enable comparison of bacterial proliferation and bacteriophage lifestyle across diverse environments and address bacterial growth as a driver of viral outcomes.



Comparison of metagenome-derived bacterial growth rate and bacteriophage lifestyles across environments

Elsa Hartley^{1,2} and Ben Knowles^{1,2}

¹B.I.G. Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

²Department of Ecology and Evolutionary Biology, UCLA

QCBio

Abstract

Viral infection shapes bacterial communities from the human gut to the open ocean. Viruses can lyse cells after infection or persist chronically. Here, we test whether bacterial growth proxies scale with viral lifestyle across environments by integrating metagenomic assembly of previously published libraries, MAG reconstruction, viral contig lifestyle classification, and growth rate inference using Peak-to-Trough (PTT) and Codon Usage Bias (CUB) methods. These growth proxies use the ratio of origin-adjacent reads to non-adjacent reads (PTT) or of tRNA genes to the codons that require them in the MAG (CUB) as measures of proliferation. Viral lifestyle is predicted through hallmark genes and genomic signatures. We have implemented this pipeline on a preliminary human gut sample set and are extending it to soil and marine datasets. This framework will enable comparison of bacterial proliferation and bacteriophage lifestyle across diverse environments and address bacterial growth as a driver of viral outcomes.

Introduction

- Phage lifestyle (lytic or temperate) can strongly influence bacterial community dynamics.
- Host growth state may affect which viral strategy is favored, but this relationship remains poorly understood across environments.
- Genome-resolved metagenomics enables links between bacterial growth-associated traits and viral lifestyle predictions.
- Question: Does bacterial proliferation correlate with the relative prevalence of lytic versus temperate phages across gut, soil, and marine microbiomes?

Conclusions

- The measures of viral lifestyle, CheckV and Phabox, showed a positive correlation (Figure 2).
- Codon-usage bias (CUB) minimum generation time, after the conversion to a maximum growth potential (1/generation time), showed a negative association with GRiD PTT instantaneous growth (Figure 3).
- As a result of the negative growth metric correlation, observed phage-lifestyle associations were different across growth proxies and were not statistically conclusive with the current sample size (Figures 4 and 5).
- Future work will determine why CUB-based growth potential and PTT-based instantaneous growth estimates diverge, using expanded human, soil, and ocean sample sets and quality-filtered genomes and MAGs.
- Because of divergence of growth metrics, we are unable to make any meaningful conclusions about virus lifestyle at this point.

Acknowledgements

We thank the QCBio Institute at UCLA, especially Alexander Hoffman and Caroline Baron, for their contributions to this project. This work was funded by the Park S. Nobel Scholarship for Plant Biology Research and the Dean's Award for Life Science Research, so a huge thank you to Park S. Nobel and Dr. Tracy Johnson for their support. We'd also like to acknowledge the efforts of 2026 B.I.G. Summer students, and NSF BIO-OCE.

Methods

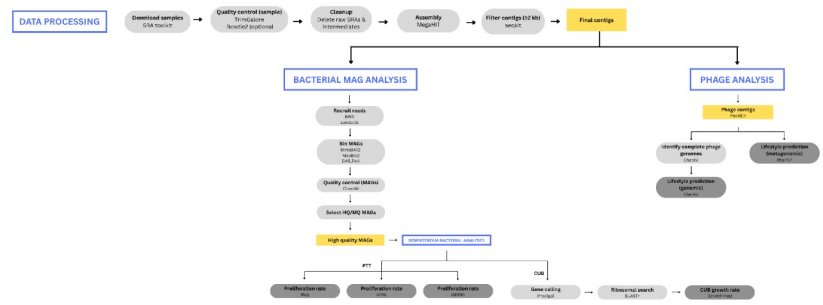


Fig. 1: Metagenomic workflow for bacterial-growth and phage-lifestyle inference

Results

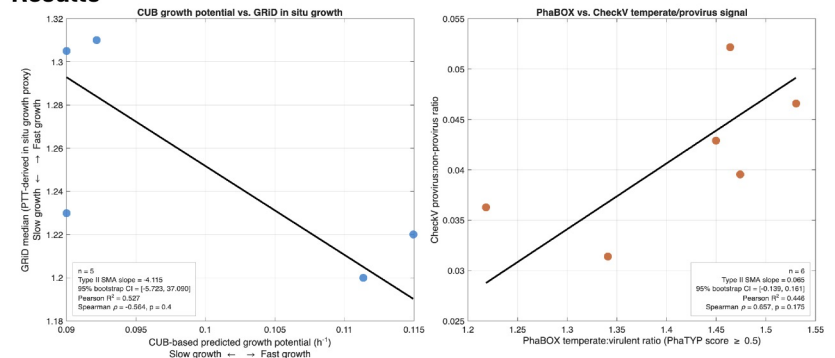


Fig. 2: Phabox and CheckV are positively associated as a measure of viral lifestyle

Fig. 3: CUB-derived growth potential is inversely associated with the GRiD in situ growth proxy

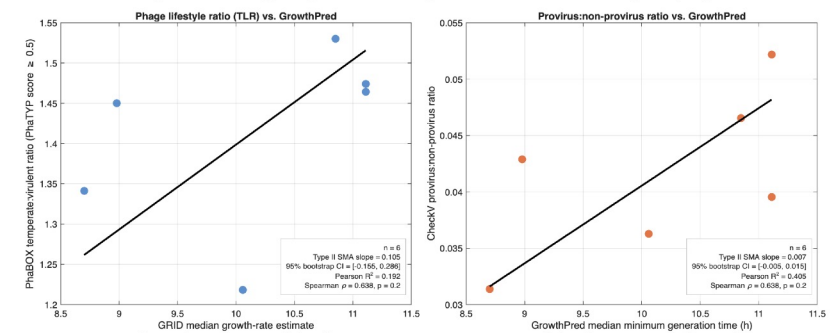


Fig. 4: Associations between bacterial growth proxies and phage lifestyle metrics using CUB GrowthPred

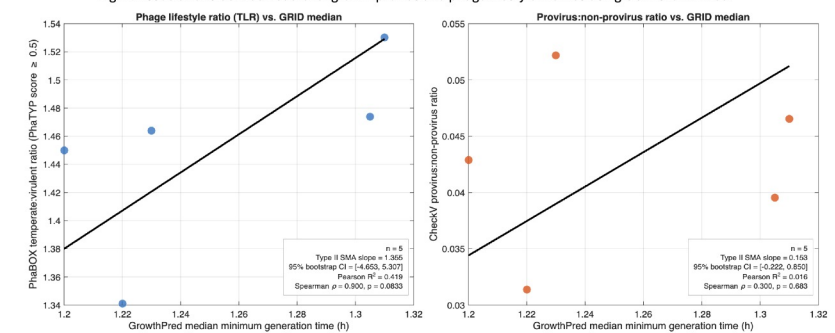
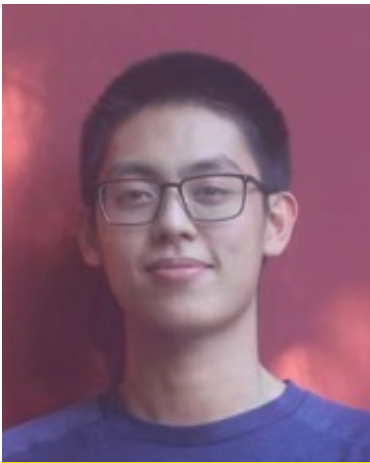


Fig. 5: Associations between bacterial growth proxies and phage lifestyle metrics using PTT GRiD



Student: Tianpei He
Faculty Advisor: Dean Buonomano
Direct Mentor: Andrew Margolis
Department: Neuroscience
<https://www.buonomanolab.com/>

Tianpei He

**Computational
and Systems
Biology &
Applied
Mathematics -
UCLA**

**Expected year
of graduation:
June 2027**

Benchmarking Bayesian Inference of Short-Term Synaptic Plasticity: Protocol-Dependent Parameter Recovery in the Tsodyks-Markram Model

TIANPEI HE¹, Dean Buonomano²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Neurology, David Geffen School of Medicine
 The Tsodyks-Markram model is widely used to quantify short-term synaptic plasticity (STP), describing how synaptic resources deplete and recover. Its parameters are routinely fit to electrophysiology recordings, yet which parameters can be recovered—and under which protocols—remains uncharacterized. Using synthetic traces generated from known parameters, we benchmark parameter recovery, comparing residual-based fitting to a Bayesian method that infers parameters from the probabilistic structure of synaptic responses, including transmission failures. Recovery is protocol-dependent: the facilitation time constant is unidentifiable without a recovery interval, and pooling diverse intervals is most reliable. The Bayesian method helps most for single-protocol data, avoiding the boundary failures of residual fits, recovering release-site number and quantal size, and providing a self-calibrating uncertainty estimate. Applied to the Allen Institute Synaptic Physiology dataset, it shows fast, consistent depression that is uncorrelated with initial synaptic strength—suggesting the two are governed by separable, rather than shared, release parameters.



Benchmarking Bayesian Inference of Short-Term Synaptic Plasticity: Protocol-Dependent Parameter Recovery in the Tsodyks-Markram Model

Tianpei He¹, Dean Buonanno²

¹ BIG Summer Program, Institute for QCBio, UCLA, ² Department of Neurology, David Geffen School of Medicine, UCLA



Abstract

The Tsodyks-Markram (TM) model is widely used to quantify short-term synaptic plasticity (STP), describing how synaptic resources deplete and recover. Its parameters are routinely fit to electrophysiology recordings, yet which parameters can be recovered—and under which protocols—remains uncharacterized. Using synthetic traces generated from known parameters, we benchmark parameter recovery, comparing residual-based fitting to a Bayesian method that infers parameters from the probabilistic structure of synaptic responses, including transmission failures. Recovery is protocol-dependent: the facilitation time constant is unidentifiable without a recovery interval, and pooling diverse intervals is most reliable. The Bayesian method helps most for single-protocol data, avoiding the boundary failures of residual fits, recovering release-site number and quantal size, and providing a self-calibrating uncertainty estimate. Applied to the Allen Institute Synaptic Physiology dataset, it shows fast, consistent depression that is uncorrelated with initial synaptic strength—suggesting the two are governed by separable, rather than shared, release parameters.

Materials and Method

- Simulated 200 synapses with TM and quantal parameters drawn uniformly over physiological ranges, stimulating each with 50 Hz trains followed by a recovery pulse at one of four intervals (250–4000 ms), five trials per protocol.
- Each dataset was fit three ways: least-squares on the mean waveform, least-squares on mean amplitudes, and the exact quantal likelihood marginalized over release-site number.
- Analyzed which ground-truth-free diagnostics actually predict recovery error—data-space R^2 , disagreement between protocols, and the width of the Bayesian posterior.

Figure 1: TM model equations, and experimental protocols

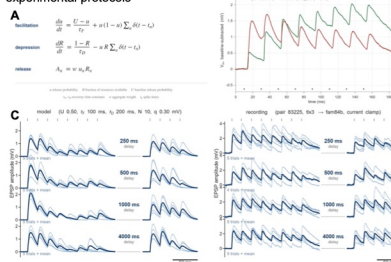
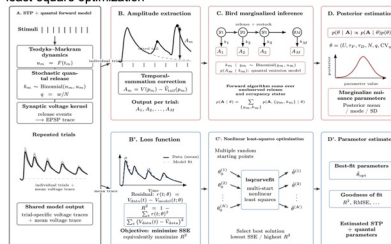


Figure 2: Pipeline for parameter estimation: bayesian inference and least-square optimization



Results—Synthetic data benchmark

Figure 3: Correlation of ground truth parameters to fit/estimate parameters parameter recovery, 200 synapses (500 ms protocol)

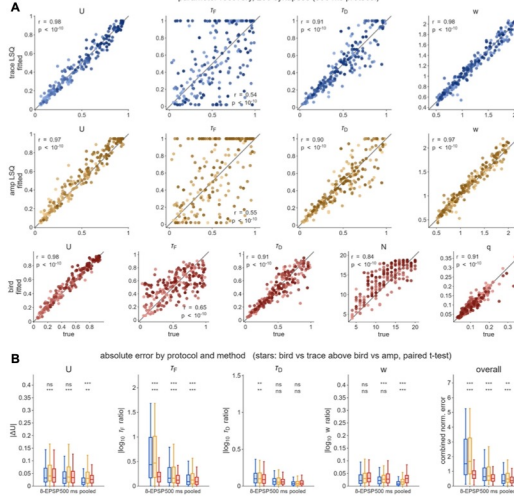
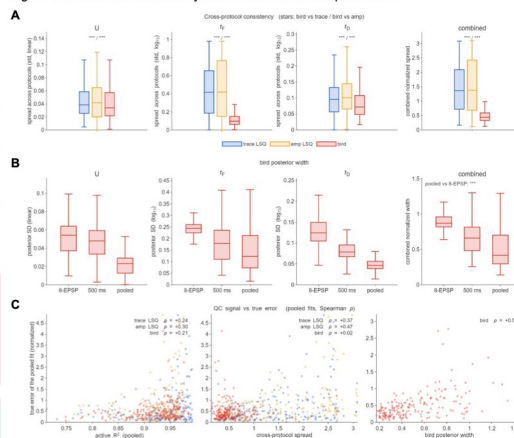


Figure 4: Confidence metrics of synthetic trace fit/estimate parameters



Reference
Tsodyks, M., Pawelzik, K., & Markram, H. (1996). Neural Networks with Dynamic Synapses. *Neural Computation*.
Bird, A. D., Wall, M. J., & Richardson, M. J. E. (2016). Bayesian Inference of Synaptic Quantal Parameters from Correlated Vesicle Release. *Frontiers in Computational Neuroscience*.

Acknowledgements
I thank Professor Dean Buonanno for his mentorship and Graduate student Jamie McDowell and Andrew Margolis for their valuable input and discussions throughout this project.

Results—Parameter estimation on Allen data

Figure 5: Parameter estimation of mouse and human dual patch data using trace lsq optimization and bayesian inference.

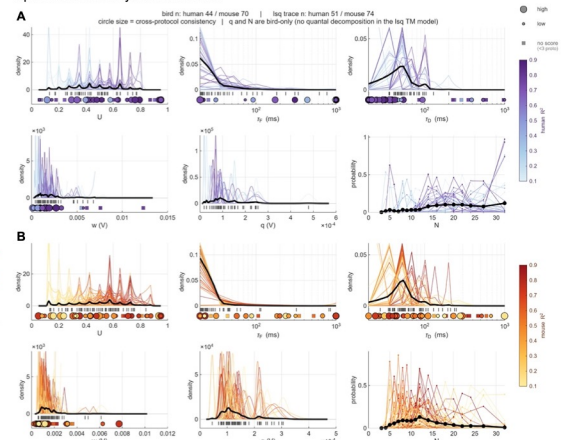
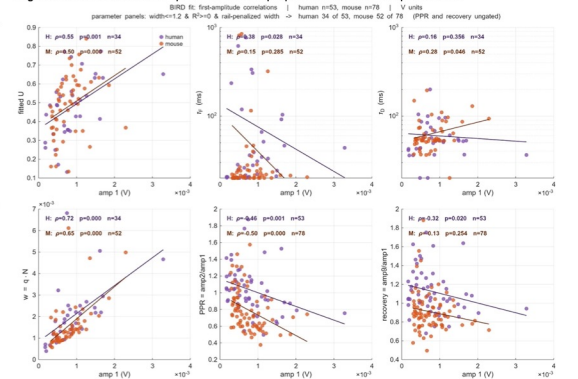


Figure 6: Correlation of STP parameters and properties with initial amplitude size



Discussion and Conclusion

- In synthetic data generated from known ground truth
- Bayesian inference can estimate N and q with relative accuracy while tying lsq in the other three parameters. It's especially better at estimating τ_F , but not as good with U.
 - R^2 -square is anti-predictive of fit accuracy in both methods. Lsq fit accuracy is better predicted by cross-protocol consistency, and bayesian inference is better predicted by combined posterior width.
- In the Allen Institute real data
- Both mouse and human showed similar range of parameters except for N. Both time constants are reported to be low in both methods (possibly still unidentifiable)
 - Higher U prediction is correlated with better R^2 , while posterior width and cross-protocol consistency seems unrelated with parameter values.
 - Initial amplitude strength is positively correlated with U and w, while negatively correlated with PPR and recovery. No strong evidence of correlation with time constants.



Student: Jasmine Herrera

Faculty Advisor: Alexander Hoffmann

Direct Mentor: Xiaolu Guo

Department: Microbiology, Immunology, and
Molecular Genetics (MIMG)

<https://www.signalingsystems.ucla.edu/alexander-hoffmann/>

**Jasmine
Herrera**

**Computational
and Systems
Biology -
UCLA**

**Expected year
of graduation
June 2026**

Two NF κ B-controlled cytokines show remarkably little correlation with NF κ B activity

JASMINE HERRERA¹, Xiaolu Guo², Francisco Lopes³,
Alexander Hoffmann²

¹ BIG Summer Program, Institute for Quantitative and
Computational Biosciences, UCLA

² Signaling Systems Laboratory, Department of MIMG, and
QCBIO, UCLA

³ Federal University of Rio de Janeiro, Brazil

Tumor necrosis factor (TNF) and interferon-beta (IFN β) are critical cytokines produced by macrophages, key coordinators of inflammation and antiviral defense in the innate immune response. While TNF and IFN β are known to be regulated by NF κ B, it remains unknown to what extent NF κ B activity is predictive of cytokine expression. To address this question, we developed an integrated NF κ B/MAPK/IRF model, validated it against experimental data, and simulated cytokine production by heterogeneous macrophages. We found that NF κ B was partially correlated with early steps of LPS-responsive TNF expression, but MAPK exhibited a higher correlation with secreted TNF. For IFN β expression, NF κ B was even less predictive, as IRF3 appeared to be the primary determinant. Nevertheless, TNF and IFN β showed some level of correlation. This work identifies regulators of TNF and IFN β secretion heterogeneity, revealing regulatory principles underlying seemingly noisy cytokine production.

Two NFκB-controlled cytokines show remarkably little correlation with NFκB activity

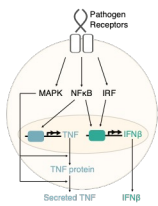
Jasmine Herrera¹, Xiaolu Guo², Francisco Lopes³, Alexander Hoffmann²¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA²Signaling Systems Laboratory, Department of MMG, and QcBio, UCLA³Federal University of Rio de Janeiro, Brazil

UCLA QcBio

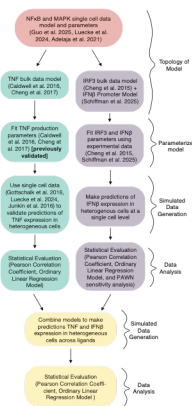
UCLA College Life Sciences
Microbiology, Immunology
& Molecular Genetics

Introduction

- Tumor necrosis factor (TNF) and interferon-beta (IFNβ) are key coordinators of innate immune responses. They are produced by macrophages and regulated by NFκB, MAPK, and IRF signaling pathways.
- While TNF and IFNβ are known to be regulated by NFκB, it remains unknown to what extent NFκB activity is predictive of cytokine expression.
- A previously established mathematical model of MAPK and NFκB signaling can be expanded to include these pathways, validated against experimental datasets, and used to simulate large, heterogeneous single cell populations to determine how these different pathways influence noisy cytokine expression and the correlation between the two expressed cytokines.



Workflow



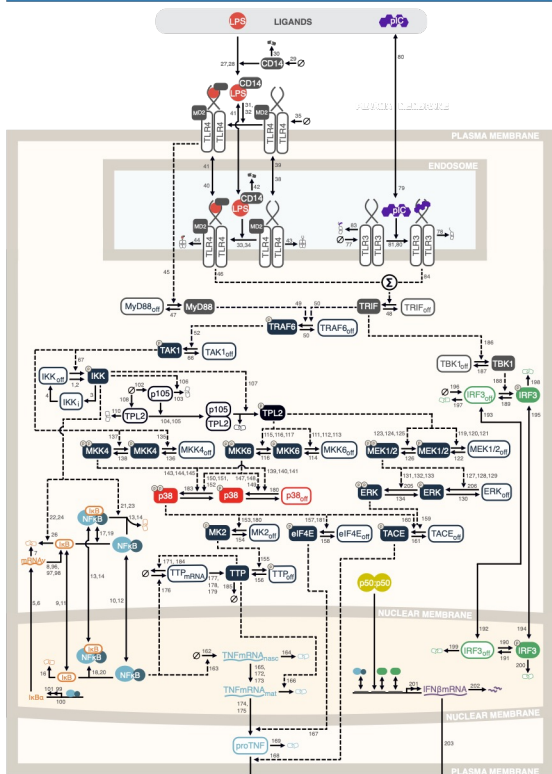
Acknowledgements

We thank the Hu Institute Foundation for funding this research.

We thank the Signaling System Lab for their support, suggestions, and discussions.

This research was supported by the NIH grants to A.H.: R01AI173214, R21AI196850

NFκB-MAPK-IRF-TNF-IFNβ Signaling Network



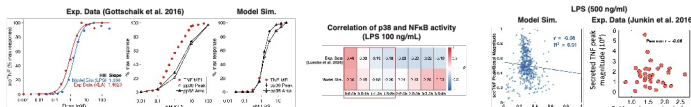
References

- Jurkin M. Cell Reports. 2016; 15(2):411-422. doi:10.1016/j.celrep.2016.03.033.
- Caswell AB. Genes & Development. 2014; 28(17):2120-2133. doi:10.1101/gad.244749.114.
- Cheng CS. Cell Systems. 2017; 4(3):330-343.e5. doi:10.1016/j.cels.2017.01.012.
- Loechle S. Molecular Systems Biology. 2024; 20(8):888-902. doi:10.1038/s44320-024-0047-4.
- Adelstein A. Immunity. 2021; 54(5):916-930. doi:10.1016/j.immuni.2021.04.011.
- Guo X. Nature Communications. 2025; 16(1):5886. doi:10.1038/s41467-025-6991-3.
- Gottschalk RA. Cell Systems. 2016; 2(3):278-290. doi:10.1016/j.cels.2016.04.016.
- Cheng Z. Science Signaling. 2016; 8(385):ra69. doi:10.1126/scisignal.aad208.
- Schiffman A. PNAS. 2025; 122(3):e2025080122. doi:10.1073/pnas.2025080122.

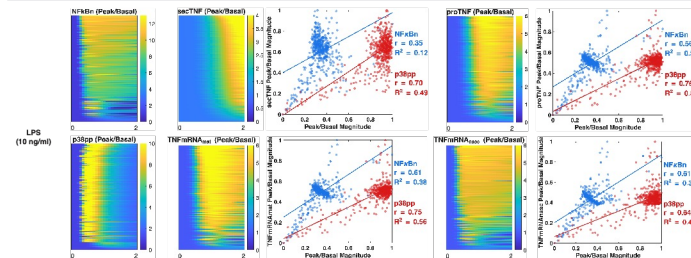
Conclusion

- Expanded previously established framework to include TNF production, IFNβ production, and IRF3 signaling pathways.
- Simulated heterogeneous populations and single-cell trajectories of our models recapitulate both single-cell and bulk experimental data, allowing for justified predictions to be made.
- Correlation analysis and linear regression of the TNF model show that MAPK activity was more strongly correlated with TNF activity than NFκB activity, suggesting TNF production is more strongly predicted by MAPK.
- Correlation, linear regression, and PAVN sensitivity analyses of the IFNβ model indicated that IRF3n was more strongly correlated with, and contributed more highly to, IFNβ activity, whereas NFκB contributed comparably in response to PolyIC but not in response to LPS. This suggests that NFκB contributions may be ligand-specific.
- Correlation analysis and linear regression in the TNF-IFNβ combined model showed that IFNβ is highly correlated with exTNF in response to PolyIC but not in response to LPS. This indicates that NFκB is not the sole strong predictor of TNF and IFNβ expression, and that the loss of correlation in response to LPS may instead be reflect the stronger predictive roles for MAPK and IRF3 pathways.
- Future directions may involve validating single-cell heterogeneous data beyond 2 hours to determine whether the loss of NFκB correlation continues while MAPK correlation increases.

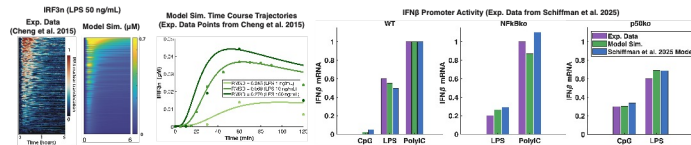
Recapitulating heterogeneous single-cell dose-response distributions and signaling-TNF secretion correlations



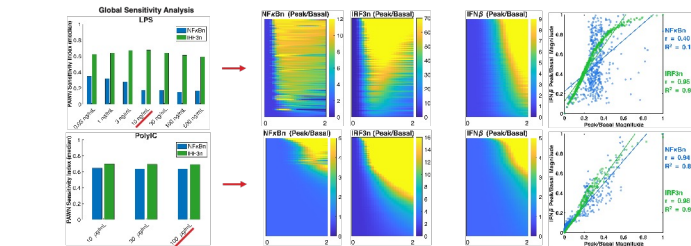
Prediction: MAPK regulates TNF secretion, while NFκB contributes to TNF expression.



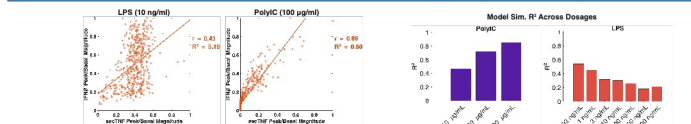
Recapitulating decades of experimental IFNβ promoter activity under diverse perturbations



Prediction: IRF3 dominantly regulates IFNβ, whereas NFκB contributions are ligand-specific



Prediction: TNF shows high correlation between IFNβ in response to PolyIC not LPS





Student: Veda Hingarh
Faculty Advisor: William Hsu
Direct Mentor: Yannan Lin
Department: Radiological Sciences
<https://www.hsu-lab.com/>

Veda Hingarh

**Computational
Biology -
UCLA**

**Expected year
of graduation:
June 2027**

Psychosocial Predictors of Year 1 Survey Retention in the Refer Everyone Advocate for Community Health Cohort

VEDA HINGARH¹, Yannan Lin², Amelia Ferrer², Hannah Milch², Ashley Elizabeth Prosper², William Hsu²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, David Geffen School of Medicine at UCLA, CA, USA

² Department of Radiological Sciences, David Geffen School of Medicine at UCLA, CA, USA

While breast cancer screening is widely adopted, lung cancer screening using low-dose CT remains underutilized nationally despite its proven mortality benefit. The REACH (Refer Everyone Advocate for Community Health) program at UCLA enrolls a diverse cohort through a mobile app capturing demographics, cancer fatalism, consideration of future consequences (CFC), lung and breast cancer knowledge, and longitudinal follow-up over three years, to understand barriers to lung and breast screening participation. We investigated whether baseline fatalism and CFC predict Year 1 survey retention. Analyses were restricted to participants enrolled at least 365 days, reflecting true non-completion rather than insufficient elapsed time. Separate multivariable logistic regression models tested the association between baseline fatalism, and separately CFC, with Year 1 completion, adjusting for age, smoking status, and family lung cancer history. Baseline survey completion was strong (67-76%), but only approximately 20% of eligible participants completed Year 1 follow-up, and neither baseline fatalism (odds ratio [OR] 1.03, 95% confidence interval [CI] 0.92-1.15) nor CFC (OR 0.98, 95% CI 0.94-1.03) independently predicted retention. Completion instead followed an all-or-nothing pattern: participants who returned for Year 1 completed nearly all applicable surveys, suggesting a single engagement event drives retention rather than survey-specific burden or fatigue. An exploratory logistic regression testing a fatalism-by-lung cancer knowledge interaction suggests knowledge may moderate fatalism's effect on retention. These findings suggest attrition in REACH may be driven more by structural or logistical barriers than by psychosocial traits alone, motivating outreach strategies focused on access and engagement via participant reminders rather than psychosocial education.

Psychosocial Predictors of Year 1 Survey Retention in the REACH Cohort

Veda Hingarh¹, Yannan Lin², William Hsu²

¹ B.I.G. Summer Program, Institute for Quantitative and Computational Biosciences, UCLA · ² Medical & Imaging Informatics, Department of Radiological Sciences, UCLA

Background

Low-dose CT screening reduces lung cancer mortality, but it remains badly underutilized. The barrier is not efficacy — it is uptake and follow-through. Because the mortality benefit depends on repeat annual screening, a participant who is scanned once and then disappears captures little of it.

REACH enrolls a diverse cohort through a mobile app built around breast cancer screening, with a subset also lung-eligible. It captures cancer fatalism, consideration of future consequences (CFC), and cancer knowledge at baseline, then follows participants for three years with annual check-ins and a peer referral mechanism.

Question

Do baseline psychosocial traits — cancer fatalism and consideration of future consequences — predict which participants return to complete their Year 1 follow-up survey?

Approach

- Six survey modules across 425 enrolled participants: demographic, fatalism, CFC, breast CAM, lung CAM, breast risk
- True enrollment date computed as the earliest survey start pooled across all modules, recovering participants missed by any single module's baseline
- Retention analyses restricted to participants enrolled ≥ 365 days, so non-completion reflects genuine non-response rather than insufficient elapsed time
- Separate multivariable logistic regressions for fatalism and CFC, each adjusted for age, smoking status, and family lung cancer history
- Exploratory fatalism $\times$ lung cancer knowledge interaction model

Non-completers are long past their window (median 589 days)

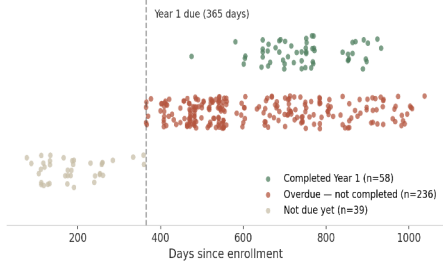


Fig 1. Days since enrollment by Year 1 status. Only 39 participants are not yet due; 236 are overdue and incomplete, at a median of 589 days enrolled (max 1,038). A further 92 participants never started any survey and have no enrollment date to plot.

Results

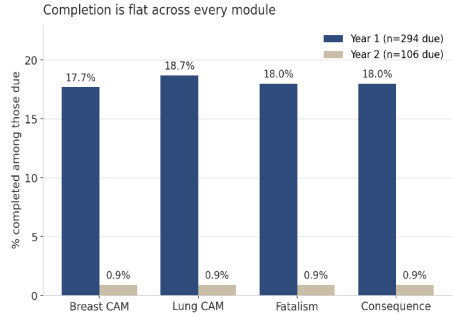


Fig 2. Eligibility-adjusted completion by module. Completion sits within a two-point band across four surveys that differ substantially in length and content — the signature burden effects would leave is absent.

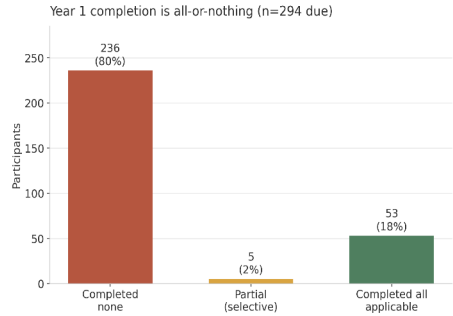


Fig 3. Of 294 participants due for Year 1, 236 completed nothing and 53 completed everything they were eligible for. Only 5 completed some but not all.

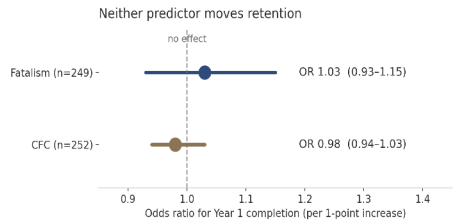


Fig 4. Adjusted odds ratios for Year 1 completion. Both estimates sit essentially on the null rather than clipping it, indicating absence of effect rather than insufficient power.

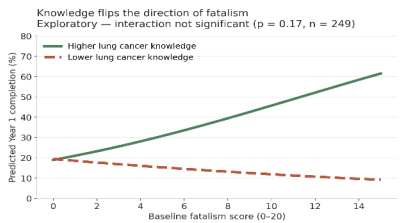


Fig 5. Exploratory interaction. Among higher-knowledge participants, fatalism tracks with higher completion; among lower-knowledge participants, the opposite. Opposing directions may explain the flat main effect in Fig 4.

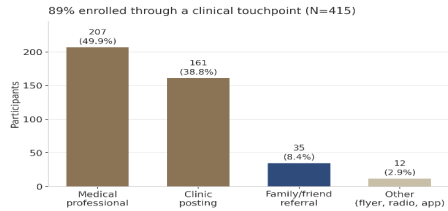


Fig 6. Recruitment source. Peer referral accounts for 35 participants tracing to 22 distinct referrers; 17 referred exactly one person and the longest chain is four.

Key Findings

- Neither fatalism (OR 1.03, 95% CI 0.93–1.15) nor CFC (OR 0.98, 0.94–1.03) predicted Year 1 retention after adjustment.
- Completion was all-or-nothing: 98% of participants due for follow-up either finished every applicable survey or none at all.
- Uniform completion across modules rules out survey-specific burden as the mechanism.
- Nearly 89% of participants enrolled through a clinical touchpoint; the designed peer referral pathway accounts for 8.4%.

Impact

Retention appears to hinge on a single engagement event rather than on participant attitudes or survey fatigue, which relocates the intervention target upstream of survey content entirely. This favors logistics-based approaches — micro check-ins in place of one annual survey, and pairing follow-up with the screening visit itself, the one moment engagement is already established — over psychosocial messaging. Adding a reason-for-non-completion prompt would distinguish notification failure from participant withdrawal, and consistent participant IDs would allow referral source to be linked to clinical outcomes.

Limitations

- REACH measures fatalism with Yes/No items rather than the validated five-point scale, compressing variance and reducing sensitivity
- Cohort is predominantly former and never-smokers entering through breast screening, limiting generalizability to lung screening populations
- Clinical outcomes (LDCT completion and findings) await more mature follow-up and resolution of data linkage gaps

Acknowledgements: UCLA B.I.G. Summer Program · Institute for Quantitative and Computational Biosciences · the REACH study team and participants.

Student: Ihyun Jeon

Faculty Advisor: Kalli Kappel

Department: Chemistry and Biochemistry

<https://sites.google.com/view/kappellab/>

Ihyun Jeon

**Biophysics-
UCLA**

**Expected year
of graduation:
June 2028**

A protein language model's condensate predictions reveal a biophysical blind spot

IHYUN JEON^{1,2}, Jaiden Ha^{1,2}, Bryan Zhang², Mia Lesser², Claire Hsieh^{2,3}, Kalli Kappel²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Chemistry and Biochemistry, UCLA

³ Bioinformatics Interdepartmental PhD Program, UCLA

Protein language models (pLMs) predict biomolecular condensate formation from sequence, but we know little about which sequences they get wrong. We asked whether a finetuned ESM-2 classifier's errors are spread across sequence space or concentrated in a specific biophysical region. We mapped 1,168 held-out predictions onto 106 biophysical features and enumerated feature-bin conjunctions, ranking each by its accuracy deficit. Two three-feature regions dropped to 31% and 37% accuracy against 85% globally. Both were recall collapses with intact specificity: 94 false negatives against 6 false positives. Both held modestly charged, moderately disordered sequences with no significant hydrophobic/aromatic or arginine/aromatic patterning, mostly because they were depleted in these sticker residues. About 75% phase separated anyway, at a mean predicted probability of 0.24. The model treats a muted sticker signature as evidence against phase separation. Both regions are biophysically defined and small enough to test directly by synthesis.

A protein language model's condensate predictions reveal a biophysical blind spot

IHYUN JEON^{1,2}, Jaiden Ha^{1,2}, Bryan Zhang², Mia Lesser², Claire Hsieh^{2,3}, Kalli Kappel²

¹B.I.G. Summer Program, Institute for Quantitative and Computational Biosciences, UCLA · ²Department of Chemistry and Biochemistry, UCLA · ³Bioinformatics Interdepartmental PhD Program, UCLA

Introduction

Biomolecular **condensates** are membraneless compartments that form when intrinsically disordered regions (IDRs) undergo **liquid-liquid phase separation**^{4,5}. Which sequences do this, and why, is still an open question.

Protein language models (pLMs) learn sequence patterns from millions of proteins and can be finetuned to predict phase separation straight from sequence. Our finetuned ESM-2 classifier³ is right **85%** of the time on held-out IDRs.

One accuracy number says nothing about *which* sequences a model gets wrong, and a predictor is only useful if you know where not to trust it^{1,4}. We asked whether the errors scatter across sequence space or concentrate in a specific, biophysically meaningful region.

Approach

Data. 3,895 IDR sequences: 3,687 from the CondensSeq library¹ and 208 experimentally verified IDRs², with condensate phenotype scored by GFP/SNAP imaging. ESM-2 finetuned as a 3-seed ensemble; 2,727 train, 584 validation, 584 test.

Search. Each of **106 biophysical features**^{9–12} was binned, and every two-, three- and four-feature combination of bins was enumerated over the **1,168 held-out** sequences and scored by how far its accuracy falls below the global rate.

With 343,000 three-feature candidates, the worst will look bad by chance. A region therefore counts only if it fails independently on validation and on test, and survives Benjamini-Hochberg FDR at 10%. We carry **two** forward. Region B clears every guard. Region A fails on both splits by the same margin but holds 28 test sequences against our floor of 30, so it is reported as a named region, not as a search hit.

Errors spread through the whole feature space

The errors do not form a visible cluster in the top two components

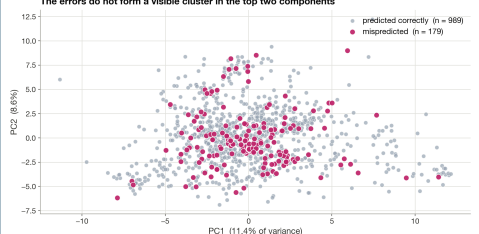


Figure 1. All 1,168 held-out sequences projected onto the first two principal components of the biophysical feature space. The 179 mispredicted sequences spread through the cloud rather than pooling anywhere, and a classifier trained on these two coordinates alone separates errors from correct calls at **AUC 0.52**, essentially chance. The structure is real, but it lies outside the directions of greatest variance, so we searched feature combinations instead.

One feature at a time only goes so far

Alone, the worst feature only pulls accuracy down to 63%

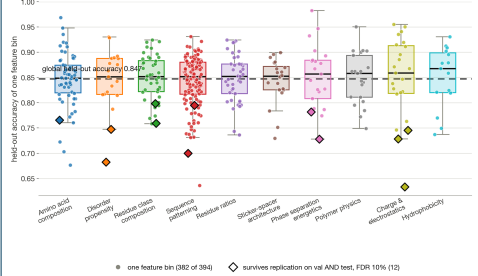


Figure 2. Held-out accuracy of each of the 394 single-feature bins, grouped by the kind of biophysical property measured. Every category straddles the global rate and no category is broadly worse than any other. Only **12 of 394 bins** survive replication on validation and test plus FDR control (outlined diamonds), and the worst of them still sits at **0.63 over 191 sequences**. Figure 3 pairs two of these axes and reaches 0.46 over 94.

Two features together open a hole

Together, two features carve out a weak hole at just 46%



Figure 3. Held-out accuracy across the plane of net charge per residue and hydrophobic moment. Neither feature alone is remarkable, reaching 0.74 and 0.75 in their worst bins. The cell at **NCPH +0.01 to +0.05**, mildly cationic, and **hydrophobic moment 0.22 to 0.26** collapses to **0.46** over 84 held-out sequences, while its immediate neighbours run 0.73 to 0.88. Region A's two-feature parent is outlined in red.

A third feature and recall collapses

Adding a third axis to the weak two-feature holes above narrows each one into a region where recall all but disappears, **0.32 over 68** held-out sequences for A and **0.37 over 83** for B.

Region A mildly cationic, NCPH +0.01 to +0.05 · hydrophobic moment 0.22 to 0.26 · ILMV vs FWY unclustered, z -2 to 0

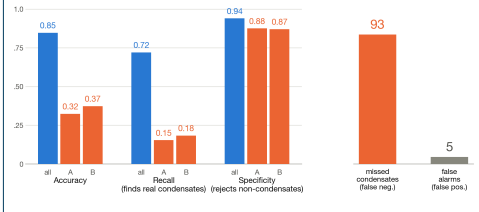
Region B disorder-promoting 0.70 to 0.775 · basic, pI 8 to 10 · RJK vs FWY unclustered, z -2 to 0

That third axis says the same thing in both. The drivers are present but **evenly dispersed, not clustered**. Held-out performance, pooled validation and test.

	n	accuracy	recall	specificity
all held-out	1,168	0.85	0.72	0.94
region A	68	0.32	0.15	0.88
region B	83	0.37	0.18	0.87

Inside both regions the model predicts a mean probability of **0.24**, while about **75%** of those sequences phase separate.

Specificity barely moves. Recall falls off a cliff.



Errors run one way



Figure 4. In both regions recall collapses while specificity holds: the model still rejects non-condensates but almost stops finding real ones. Pooled there are **93 missed condensates against 5 false alarms**, so this is not a class-balance artifact.

These sequences really do form condensates

Every sequence on this row phase separated in the assay. The model got the left half wrong.

Model says NO

Model says YES

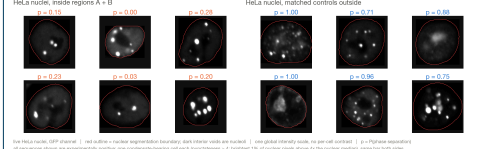


Figure 5. CondensSeq microscopy rules out an imaging or labelling artifact¹³. Panels are live **HLA nuclei**: an NLS on the library makes the outlined object the nucleus, not the whole cell. Labels here are **cheaper** than the library average (nucleolar 0.013 against 0.056), so nucleolar mis-scoring runs the wrong way. Region sequences do reach high-contrast puncta less often: **0.8% against 14.2%** of cells, p = 0.03 weighting sequences equally; 90 held-out sequences against 90 matched controls, but the median cell looks the same in both sets.

Conclusions and future directions

- pLM condensate errors concentrate in **definable biophysical regions**, and only **conjunctions of features** reveal them. Scanned one feature at a time, the model looks uniformly competent.
- The model learned that stickers and cation- π contacts drive phase separation^{4–8}, and it has no rule for their absence. Both regions are depleted in aromatic and hydrophobic residues, and about 7% of them phase separate anyway.
- Specificity holds at 0.88 and 0.87 inside the two regions, so a positive prediction stays reliable. The blind spot is confined to negatives.
- Region B was **memorised from 46 training sequences** with zero training false negatives and generalises to nothing, a data gap that more training sequences would close.
- Next: synthesise the 101-sequence union**, prioritising those the model calls negative. A predicted-negative that comes back positive tells us the most.

Acknowledgements

Thanks to **Dr. Kalli Kappel** and the whole Kappel Lab for hosting this work, and to the **B.I.G. Summer Program** and the Institute for Quantitative and Computational Biosciences for support and funding. CondensSeq imaging data courtesy of the Kappel Lab. Correspondence: ihyunjeon@ucla.edu

References

1. Kappel, K. et al. *Nat. Methods* **22**, 1464–1475 (2025).
2. Maiti, S. et al. *Sci. Adv.* **11**, eady1420 (2025).
3. Lin, Z. et al. *Science* **379**, 1123 (2023). ESM-2.
4. Banani, S. F. et al. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
5. Holehouse, A. S. & Kragelund, B. B. *Nat. Rev. Mol. Cell Biol.* **7**, Das, S. et al. *PNAS* **117**, 28795–28805 (2020).
6. Pappu, R. V. et al. *Chem. Rev.* **123**, 8945–8987 (2023).
7. von Bülow, S. et al. *PNAS* (2025).
8. Cohan, M. C. et al. *J. Mol. Biol.* **434**, 167373 (2022).
9. NARDINI.
10. Ginell, G. M. et al. *Science* **388**, eadq8381 (2025).
11. FINCHES.
12. Emenecker, R. J., Griffith, D. & Holehouse, A. S. *Biophys. J.* (2021). Metapredict.
13. Alberti, S., Gladfelter, A. & Mittag, T. *Cell* **176**, 419–434 (2019).
14. Prabhakaran, R. & Bromberg, Y. *Nat. Methods* (2026).



Student: Vyas Koduvayur

Faculty Advisor: Alexander Hoffmann

Direct Mentor: Aleksander Gorin

Department: Microbiology, Immunology, and
Molecular Genetics (MIMG)

<https://www.signalingsystems.ucla.edu/alexander-hoffmann/>

**Vyas
Koduvayur**

**Computational
& Systems
Biology- UCLA**

**Expected year
of graduation
June 2027**

Clinical biopsies reveal macrophages with distinct histories of cytokine exposure

VYAS KODUVAYUR^{1,2}, Alexander Hoffmann², and Aleksandr Gorin³

¹ BIG Summer Program, Institute for Quantitative and
Computational Biosciences, UCLA

² Department of Microbiology, Immunology, and Molecular
Genetics, UCLA

³ Department of Medicine, Division of Infectious Diseases, UCLA
Macrophages are highly plastic innate immune cells that respond to environmental cues and regulate tissue homeostasis and inflammation. Macrophage activation is traditionally described using the M0/M1/M2 framework, which classifies cells by cytokine exposure. *In vivo*, cytokine stimulation can occur before, during, or after monocyte-to-macrophage differentiation. Consequently, the same cytokine may induce distinct transcriptional programs depending on maturation state, which current methods don't reveal. Here, we generated maturation- and cytokine-specific RNAseq signatures from monocytes and macrophages exposed to IFN γ or IL4 during MCSF differentiation. We developed a hierarchical scoring framework that classifies cells by maturation stage before assigning cytokine-polarized states, enabling context-aware annotation of macrophages in single-cell datasets from patient biopsies. Assigned cell states retained defining transcriptional signatures, and trajectory analyses recapitulated expected differentiation pathways. Our framework demonstrates that *in vitro* datasets can be leveraged to infer both current programming state and prior history of macrophages in clinical samples.



Clinical biopsies reveal macrophages with distinct histories of cytokine exposure

Vyas Koduvayur^{1,2}, Alexander Hoffmann², and Aleksandr Gorin³



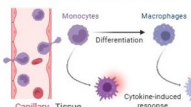
UCLA

¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, ²Department of Microbiology, Immunology, and Molecular Genetics, ³Department of Medicine, Division of Infectious Diseases, University of California, Los Angeles

Abstract

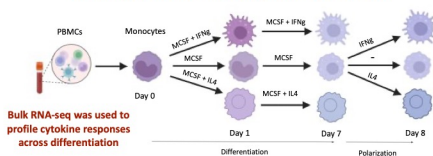
- Macrophages are highly plastic innate immune cells that respond to environmental cues and regulate tissue homeostasis and inflammation. Macrophage activation is traditionally described using the M0/M1/M2 framework, which classifies cells by cytokine exposure.
- In vivo*, cytokine stimulation can occur before, during, or after monocyte-to-macrophage differentiation. Consequently, the same cytokine may induce distinct transcriptional programs depending on maturation state, which current methods don't reveal.
- Here, we generated maturation- and cytokine-specific RNAseq signatures from monocytes and macrophages exposed to IFN γ or IL4 during MCSF differentiation. We developed a hierarchical scoring framework that classifies cells by maturation stage before assigning cytokine-polarized states, enabling context-aware annotation of macrophages in single-cell datasets from patient biopsies.
- Assigned cell states retained defining transcriptional signatures, and trajectory analyses recapitulated expected differentiation pathways. Our framework demonstrates that *in vitro* datasets can be leveraged to infer both current programming state and prior history of macrophages in clinical samples.

In vivo, cytokines can act before, during, or after monocyte-to-macrophage differentiation



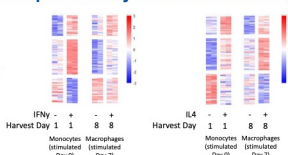
- Do cytokines induce distinct cellular states in monocytes compared to macrophages?
- Can we identify signatures of these states that can be translated to clinical data?

In vitro model of cytokine exposure during and after monocyte-to-macrophage differentiation

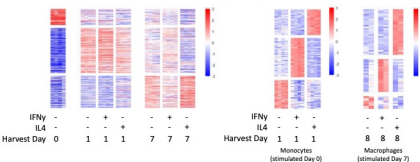


Cytokine-driven responses vary with maturation stage

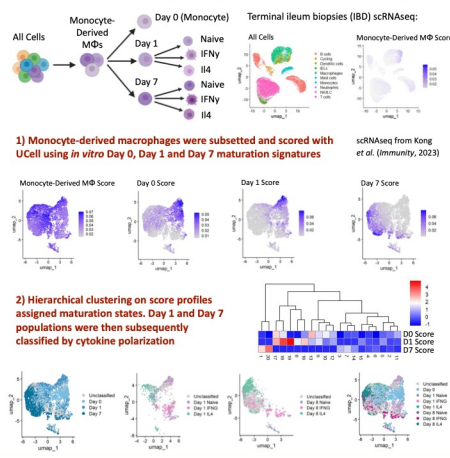
A 1-day cytokine exposure elicits distinct transcriptional responses in monocytes versus macrophages



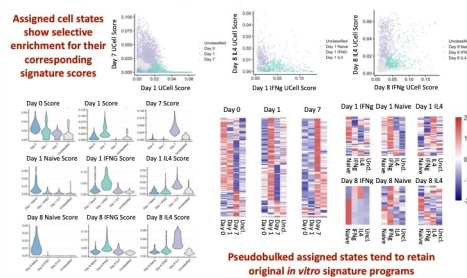
Maturation stage signatures define each timepoint, revealing distinct cytokine programs within each stage



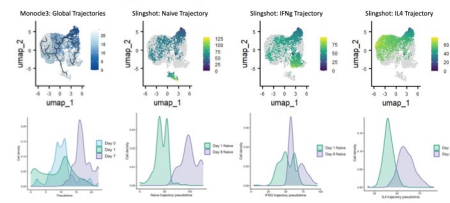
Hierarchical scoring identifies maturation stages and stage-specific cytokine polarization states *in vivo*



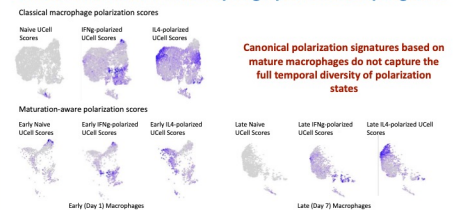
Validation of transcriptionally defined maturation states



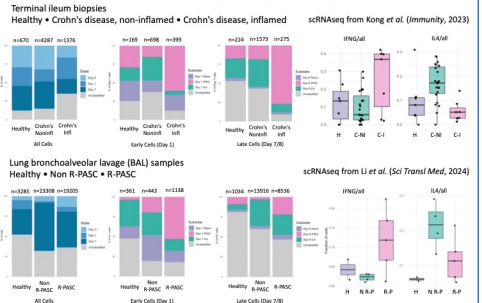
Trajectory inference recapitulates macrophage maturation and cytokine-specific differentiation



Maturation state-aware signatures contextualize classical M0/M1/M2 macrophage polarization programs



Macrophage maturation landscapes differ across tissues and disease states



Discussion

- In vitro* cytokine perturbations define transcriptional signatures of macrophage maturation and inflammatory exposure.
- These signatures can be used to interpret transcriptional states of macrophages in clinical samples and characterize their cytokine exposure history.
- This methodology provides a framework for translating controlled *in vitro* perturbation experiments into biological insights from *in vivo* clinical samples.
- Applying this approach across tissues and diseases could reveal how macrophage maturation, polarization, and inflammatory history vary across disease contexts.
- Next steps
 - Refine maturation/exposure signatures and improve classification of unassigned macrophages.
 - Apply the classifier to various organs and clinical datasets to identify disease-associated macrophage states and exposure histories.

Acknowledgements

This project was made possible by funding from the Hu Institute Foundation Scholarship.

Special thanks to the members of the Signaling Systems Lab for their guidance and support.

We also thank the QCBio community and the BIG Summer Program for providing us with their resources and the opportunity to present this work.



Student: Harshita Kukreja
Faculty Advisor: Sam Emaminejad
Direct Mentor: Alireza Mohseni
Department: Computer Science
<https://www.ee.ucla.edu/sam-emaminejad/>

**Harshita
Kukreja**

**Computer
Science and
Engineering -
UCLA**

**Expected year
of graduation
June 2028**

Can Routine Laboratory Tests Detect Chronic Heart Failure?

HARSHITA KUKREJA^{1,2,3}, Alireza Mohseni⁴, Michael Makhoul^{4,5}

¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department Computer Science, Henry Samueli School of Engineering and Applied Science, UCLA

³ Department of Disability Studies, UCLA

⁴ Interconnected and Integrated Bioelectronics Lab (I2BL), Department of Electrical and Computer Engineering, Henry Samueli School of Engineering and Applied Science, UCLA

⁵ Department of Bioengineering, Henry Samueli School of Engineering and Applied Science, UCLA

NT-proBNP, a biomarker released in response to cardiac wall stress, supports CHF diagnosis alongside other techniques. However, it is rarely ordered in hospitalized patients (0.96% of admissions in dataset). We tested whether routine labs can identify documented CHF (ICD-9/10) using MIMIC-IV data (123,626 unique patients; 14,388 [11.6%] with CHF). Panel selected empirically from all biomarker combinations by cross-validated AUC-to-sample-size tradeoff. XGBoost achieved AUC 0.769 (95% CI 0.765–0.772; sensitivity 0.78, specificity 0.62), modestly outperforming logistic regression (AUC 0.762; difference 0.006, 95% CI 0.004–0.009). SHAP identified creatinine as second largest contributor to model output (next to age). Predicted probabilities were poorly calibrated to true CHF prevalence. Restricting to patients without documented renal disease reduced CHF prevalence to 4.7% (cohort composition change) and AUC to 0.712. External validation beyond this single-center cohort is needed.

Can Routine Lab Tests Detect Chronic Heart Failure?

UCLA Samueli School of Engineering

HARSHITA KUKREJA^{1,2,3}, Alireza Mohseni⁴, Michael Makhoul^{4,5}

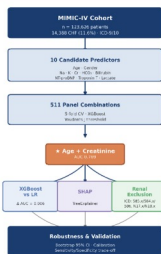
UCLA QCBio

¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA ²Department Computer Science, Henry Samueli School of Engineering and Applied Science, UCLA ³Department of Disability Studies, UCLA ⁴Interconnected and Integrated Bioelectronics Lab (I2BL), Department of Electrical and Computer Engineering, Henry Samueli School of Engineering and Applied Science, UCLA ⁵Department of Bioengineering, Henry Samueli School of Engineering and Applied Science, UCLA

ABSTRACT

NT-proBNP, a biomarker released in response to cardiac wall stress, supports CHF diagnosis alongside other techniques. However, it is rarely ordered in hospitalized patients (0.96% of admissions in the dataset). We tested whether routine labs can identify documented CHF (ICD-9/10) using MIMIC-IV data (123,626 unique patients; 14,388 [11.6%] with CHF). Panel selected from a cross-validated search of 511 age/biomarker combinations; Age + Creatinine achieved within 0.004 AUC of the top-scoring panel while preserving a larger complete-case cohort. XGBoost achieved AUC 0.769 (95% CI 0.765–0.772; sensitivity 0.78, specificity 0.62), modestly outperforming logistic regression (AUC 0.762; difference 0.006, 95% CI 0.004–0.009). SHAP analysis of the broader biomarker panel identified age and creatinine as the dominant predictors, motivating their selection as the minimal reported panel. Predicted probabilities were poorly calibrated to true CHF prevalence. Restricting to patients without documented renal disease reduced CHF prevalence to 4.7% (cohort composition change) and AUC to 0.712. External validation beyond this single-center cohort is needed.

METHODS



Cohort definition, combinatorial panel selection (511 combinations, 5-fold cross-validated XGBoost), and downstream validation steps.

CONCLUSION

A two-lab panel (age + creatinine) approaches the discrimination ceiling achievable from routine labs alone (AUC 0.769, 95% CI 0.765–0.772, bootstrap-validated on cross-validated predictions), modestly but reliably outperforming logistic regression. Excluding documented renal disease reduced both AUC and creatinine's relative importance consistent with, though not conclusive proof of, a renal contribution to creatinine's signal. Routine labs may support CHF risk flagging ahead of a formal workup, but are not a substitute for NT-proBNP or clinical diagnosis. External validation in an independent cohort is needed before any clinical use.

RESULTS

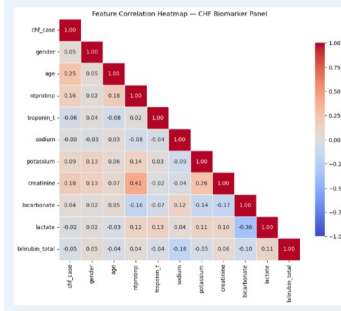


Figure 1. Pairwise correlation matrix for candidate biomarkers, patient demographics, and CHF diagnosis (pairwise-complete observations, n varies by pair). Age ($r = 0.25$) and creatinine ($r = 0.17$) show the strongest univariate correlation with CHF diagnosis; no single lab value is strongly correlated with case status on its own, motivating a multivariable model over any single threshold-based rule.

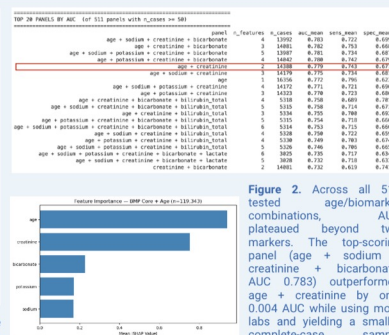


Figure 3. SHAP feature importance for the 5-marker BMP-core + Age panel ($n=13,987$; AUC 0.781). Age and creatinine are the dominant predictors; sodium, potassium, and bicarbonate contribute comparatively little, supporting Age + Creatinine as a near-lossless minimal panel.

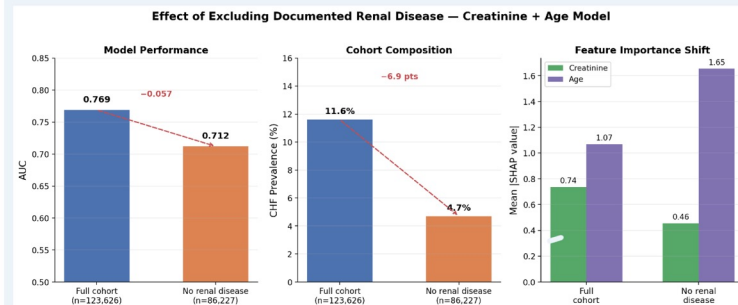


Figure 4. Effect of excluding patients with documented renal disease (ICD-coded CKD/AKI/renal failure) on the Age + Creatinine model. AUC declines from 0.769 to 0.712 as CHF prevalence drops from 11.6% to 4.7%. Creatinine's SHAP importance falls by 38% (0.74→0.46) while age's rises (1.07→1.65), indicating creatinine's predictive signal is substantially, though not entirely, attributable to comorbid renal impairment rather than heart failure specifically.

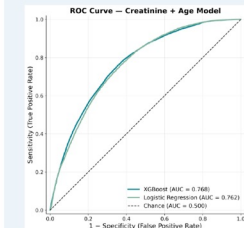


Figure 5. ROC curves for XGBoost vs. logistic regression on the Age + Creatinine panel. XGBoost modestly outperformed logistic regression (AUC 0.769 vs. 0.762; Δ 0.006, 95% CI 0.004–0.009).

REFERENCES

O'Kane M, Porter D, McCann M, Jülicher P, Christenson R, Oellerich M, Price CP, St John A. A value proposition for natriuretic peptide measurement in the assessment of patients with suspected acute heart failure. *Clin Chim Acta*. 2020 Jan;500:98–103. doi: 10.1016/j.cca.2019.09.023. Epub 2019 Oct 23. PMID: 31655057.

DISCUSSION

1. The top-scoring panel among all 511 tested combinations exceeded Age + Creatinine by only 0.004 AUC at the cost of two additional labs and a smaller sample → supporting Age + Creatinine as a near-lossless minimal panel.
2. Bootstrap resampling (1,000 iterations) of cross-validated, out-of-fold predictions gave XGBoost vs. logistic regression a 95% CI of 0.004–0.009 for the AUC difference, excluding zero, so the modest advantage is real rather than sampling noise.
3. Excluding patients with documented renal disease reduced both AUC (0.769→0.712) and creatinine's relative SHAP importance, consistent with comorbid renal disease contributing to creatinine's predictive value, though cohort size and CHF prevalence changed simultaneously with this exclusion and were not held constant, so this contribution cannot be fully isolated from the present analysis.

LIMITATIONS

Predicted probabilities were poorly calibrated to true CHF prevalence, its raw percentages should not be read as literal probabilities without recalibration.

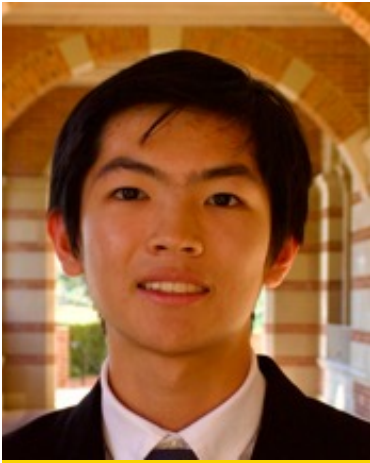
Analysis used complete-case observations rather than imputation for missing labs, which is conservative but discards patients with partial data.

CHF ascertainment relied on ICD-coded diagnoses. Class imbalance (11.6% prevalence) was addressed via SMOTE oversampling within training folds, which yielded higher AUC and sensitivity than untreated or weighted alternatives at a modest cost to specificity (Δ Sens +0.026/+0.048, Δ Spec -0.018/-0.028 vs. no handling/scale_pos_weight).

Findings are from a single academic medical center and require external validation.

ACKNOWLEDGEMENTS

We thank the MIT Laboratory for Computational Physiology and Beth Israel Deaconess Medical Center for providing the MIMIC-IV dataset (Johnson, A.E.W., Bulgarelli, L., Shen, L. et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data* 10, 1 (2023); PhysioNet, Goldberger et al. 2000)



Student: Derek Lee

19

Faculty Advisor: Timothy O'Sullivan

Direct Mentor: Wesley Armstrong

Department: Microbiology, Immunology, and
Molecular Genetics (MIMG)

<https://osullivanlab.com/>

Derek Lee

**Molecular, Cell,
and
Developmental
Biology - UCLA**

**Expected year
of graduation
June 2028**

Validating Putative Cell-Cell Interactions in Human Lung Adenocarcinoma using Spatial Transcriptomics

DEREK LEE¹, Tim O'Sullivan²

¹ BIG Summer Program, Institute for Quantitative and
Computational Biosciences, UCLA

² Department of Microbiology, Immunology, and Molecular
Genetics, UCLA

The tumor microenvironment consists of a complex web of cell-cell interactions that impact tumor progression and clinical outcomes. While these interactions are unique to each cancer type and may serve as promising therapeutic targets, the clinical significance of individual cell-cell communication networks remains unclear. To address this, we developed SKyNet, a clinically informed framework integrating matched transcriptomic and clinical data from the Cancer Genome Atlas to identify ligand-receptor interactions significantly associated with patient survival. Here, we analyze publicly available tumor and adjacent tissue from lung adenocarcinoma patients using 10x Visium spatial transcriptomics to evaluate spatial colocalization between predicted ligand-receptor interactions. Using Jaccard similarity, we quantified ligand-receptor colocalization in both tumor and adjacent tissue and identified three distinct categories of interactions, suggesting that differential mechanisms of interaction between these categories may drive tumor progression. Our spatial transcriptomic analysis provides a validation step to refine *in silico* ligand-receptor communication networks.

Validating Putative Cell-Cell Interactions in Human Lung Adenocarcinoma using Spatial Transcriptomics

DEREK LEE¹, Timothy E. O'Sullivan²¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA² Department of Microbiology, Immunology, and Molecular Genetics, UCLA

Background

The Tumor Microenvironment Milieu

The tumor microenvironment (TME) contains a complex web of non-cancerous cells (e.g., immune cells and fibroblasts) that communicate to create a tumor permissive environment.

Single-cell based Kaplan-Meier Analysis Network (SKyNet)

Incorporates matched transcript and clinical data from the Cancer Genome Atlas for prioritizing ligand-receptor interactions associated with survival outcomes in cancer patients.

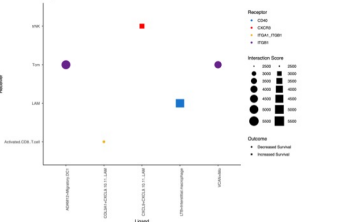


Figure 1: Identifying ligand-receptor interactions associated with clinical outcomes using SKyNet. Map of ligand-receptor interactions correlated with significant survival outcomes in breast cancer patients. Interaction scores were calculated from the mean survival difference and p-values generated from Kaplan-Meier survival analyses.

Project Aim

Communication networks like SKyNet predict ligand-receptor interactions based on **transcript enrichment** and/or **clinical outcomes** but DO NOT address tumor heterogeneity (i.e., ligand and receptor may be restricted to distinct locations within the tumor). We used 10x Visium spatial transcriptomics samples from lung adenocarcinoma (LUAD) patients to address this limitation by evaluating colocalization of predicted ligand-receptor interactions.

Methods

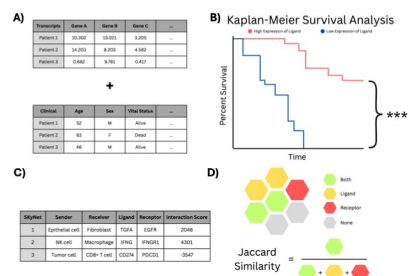


Figure 2: Spatial Transcriptomic Analysis Workflow. (A) Paired bulk RNA-seq and clinical data spanning 11 cancers were downloaded from the Cancer Genome Atlas. (B) Ligand-receptor interactions from the scDiffCom database were evaluated through Kaplan-Meier analysis and permutation testing. (C) SKyNet outputs a spreadsheet of ligand-receptor interactions significantly associated with survival outcomes based on Kaplan-Meier analyses, along with their sender and receiver cell types. (D) Ligand-receptor colocalization were mapped on five matched tumor and adjacent 10x Visium spatial transcriptomics samples from LUAD patients and quantified using Jaccard similarity.

Figure 3

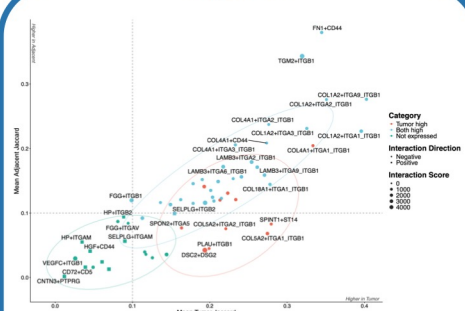


Figure 3: Quantifying predicted ligand-receptor colocalization in LUAD spatial transcriptomic samples using Jaccard Index. Calculated Jaccard similarity in tumor and adjacent tissue across five patients for 58 unique ligand-receptor interactions predicted by SKyNet in LUAD patients. The minimum cutoff for ligand-receptor colocalization was set at a Jaccard index of 0.1 (10%). Normal confidence ellipses generated with a confidence level of 0.90.

Figure 4

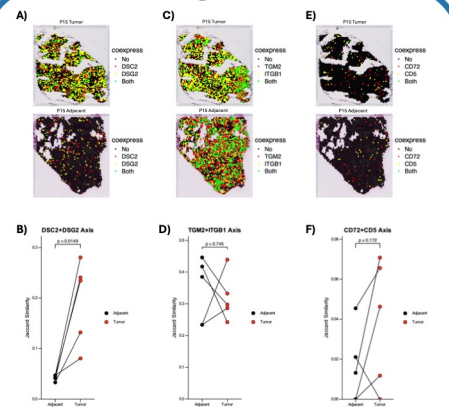


Figure 4: Identification of three categories of ligand-receptor interactions through Jaccard Index. (A-B) Tumor high: Jaccard index above 10% in tumor tissue and significantly higher than adjacent tissue ($p < 0.05$). (A) Jaccard similarity in tumor and adjacent tissue for DISC2/D5G2 axis. (B) Representative spatial transcriptomic sample for colocalization of DISC2/D5G2 axis. (C-D) Both high: Jaccard index above 10% in tumor tissue and not significantly higher than adjacent tissue ($p > 0.05$). (C) Jaccard similarity in tumor and adjacent tissue for TGM2/ITGB1 axis. (D) Representative spatial transcriptomic sample for colocalization of TGM2/ITGB1 axis. (E-F) Not expressed: Jaccard index below 10% in both tumor and adjacent tissue. (E) Jaccard similarity in tumor and adjacent tissue for CD72/CD5 axis. (F) Representative spatial transcriptomic sample for colocalization of CD72/CD5 axis. P-values for (B), (D), and (F) were calculated using paired t-tests.

Figure 5

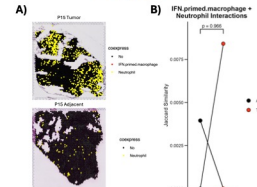


Figure 5: 10x Visium spatial transcriptomics lacks the resolution to evaluate spatial colocalization of cell-cell interactions. (A) Jaccard similarity in tumor and adjacent tissue for interferon-primed macrophage and neutrophil interactions. (B) Representative spatial transcriptomic sample for colocalization of interferon-primed macrophage and neutrophil interactions. Cell type presence was determined by calculating module scores based on lineage genes derived from scRNA-seq and taking the top three module scores in each Visium spot. P-value for (B) was calculated using paired t-test.

Discussion

Three Categories of Ligand-Receptor Interactions

- Tumor High:** Presence of interaction associated with survival outcomes → interaction drives survival outcomes
- Both High:** Presence of interaction does not affect prognosis → tumor microenvironment modulates interaction's impact on tumor progression
- Not Expressed:** possible limitations of 10x Visium sequencing depth

Future Directions

- The power of our analysis was limited sample size ($n=5$); a larger cohort would increase confidence
- Draw association between spatial colocalization and clinical outcomes
- 10x Xenium achieves single cell resolution at the expense of sequencing depth (maximum of ~5k genes)

Acknowledgments

This work used publicly available data originally published by De Ziani et al., 2024, which is accessible online through [E-MTAB-13530](https://doi.org/10.1101/2024.03.13.583330).

I would like to thank Dr. O'Sullivan, my mentor Wesley Armstrong, and Bruins-In-Genomics (BIG) Summer Program for guiding and feedback throughout the project. This work was made possible by generous support from the Hu Institute Foundation Scholarship. I would also like to thank Jade Liang and Varchas Bharadwaj for developing the SKyNet communication algorithm.

References

- De Ziani, M. et al. Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. *Nat. Commun.* **15**, 4388 (2024).
- Hoadley, K. A. et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
- Laguer, C. et al. scDiffCom: a tool for differential analysis of cell-cell interactions provides a mouse atlas of aging changes in intercellular communication. *Nat. Aging* **3**, 1446-1461 (2023).
- Liu, Z., Sun, D. & Wang, C. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. *Genome Biol.* **23**, 218 (2022).
- Mayer, S. et al. The tumor microenvironment shows a hierarchy of cell-cell interactions dominated by fibroblasts. *Nat. Commun.* **14**, 5810 (2023).

Student: Yi-Ai (Allison) Li
Faculty Advisor: Willy Hugo
Department: Department of Medicine, Dermatology
<https://www.uclahealth.org/cancer/members/willy-hugo>

**Yi-Ai (Allison)
Li**

**Molecular, Cell,
and
Developmental
Biology - UCLA**

**Expected year
of graduation
June 2027**

Characterization of Intratumoral Immune Cell Populations in Glioblastoma Patients with Extended Survival

ALLISON LI¹, Christopher Wong², Willy Hugo²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Medicine, David Geffen School of Medicine, UCLA

Glioblastoma (GBM) is a highly aggressive primary brain tumor characterized by extensive intratumoral heterogeneity. Autologous tumor lysate-pulsed dendritic cell (ATL-DC) vaccination has been shown to improve survival in some patients, with a subset achieving long-term survival of over 10 years without recurrence. However, the immune mechanisms underlying the distinct outcomes of long-term responders (LTRs) and short-term responders (STRs) remain poorly understood. We analyzed pre-therapy 10x Genomics Visium spatial transcriptomics data with from GBM tumors to characterize the spatial organization of the tumor immune microenvironment. Distinct immune cell populations were enriched in either the tumor microenvironment of LTRs or STRs, suggesting differences in cellular composition that may influence therapeutic response and survival outcomes. These findings provide insights into the spatial immune landscape of GBM and may guide future optimization of ATL-DC vaccination and the development of more effective immunotherapeutic strategies.

Characterization of Intratumoral Immune Cell Populations in Glioblastoma Patients with Extended Survival

Allison Li¹, Christopher Wong², Willy Hugo²¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA, ² Department of Medicine, David Geffen School of Medicine, UCLA

ABSTRACT

Inflammatory cellular communities (CNs) featuring high immune infiltration and mesenchymal-like (MES-like) tumor cells are consistently enriched in short-term responders (STRs). Conversely, distinct CNs with patient-specific immune cell enrichment characterize long-term responders (LTRs). Overall, these insights into GBMs spatial immune landscape can guide future optimization of ATL-DC vaccination and targeted immunotherapies.

INTRODUCTION

Glioblastoma (GBM) is one of the most aggressive primary brain tumors, characterized by extensive intratumoral heterogeneity and an immunosuppressive tumor microenvironment that contributes to therapeutic resistance and poor survival. Autologous tumor lysate-pulsed dendritic cell (ATL-DC) vaccination aims to induce tumor-specific anti-tumor immunity and has demonstrated survival benefits in patients with methylated MGMT (mMGMT) GBM [3]. While most patients with GBM survive less than 30 months despite treatment, a rare subset of ATL-DC-treated patients achieved long-term survival of over 10 years without recurrence.

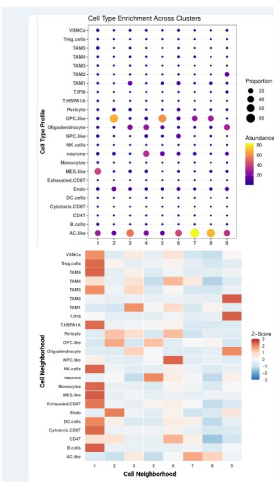
However, the cellular and spatial mechanisms distinguishing LTRs from STRs remain poorly understood. Here, we analyzed pre-treatment GBM samples from four LTRs and four STRs using spatial transcriptomics to characterize differences in cellular composition and spatial tumor-immune organization that may contribute to long-term response.

KEY FINDINGS:

- STRs were enriched with inflammatory, MES-like/TAM-rich neighborhoods.
- TAM-MES-like tumor co-enrichment may contribute to STR-associated niches and maladaptive inflammation.

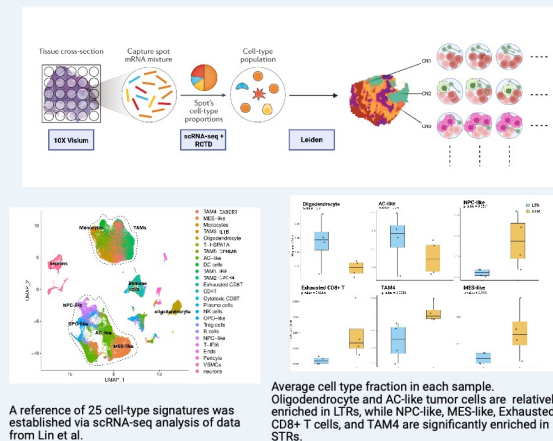
METHODS

To analyze the spatial tumor-immune microenvironment, 10x Genomics Visium spatial transcriptomics data were evaluated. Single-cell RNA sequencing (scRNA-seq) data from Lin et al. [1] were processed using Seurat (v4.3.0), incorporating Harmony (v2.0.5) batch correction and Louvain clustering, to establish reference signatures across 25 distinct cell types [1,4,5]. The cell-type composition of each Visium spot was subsequently estimated via Robust Cell Type Decomposition (RCTD; spacexr package v2.2.1) using maximum-likelihood estimation in Full Mode. Based on these inferred cellular compositions, spatial networks were constructed and partitioned into cellular neighborhoods (CNs) using Leiden clustering (v0.4.3.1). Finally, differences in cell-type proportions within CNs and overall CN abundance across samples were statistically evaluated using Mann-Whitney U tests in R.



CN1 is enriched with MES-like, TAMs, and immune cells; CN6 is enriched with AC-like, NPC-like, and TAMs, and CD4. Other CNs are enriched with distinct cell type combinations, revealing diverse microenvironments.

RESULTS



A reference of 25 cell-type signatures was established via scRNA-seq analysis of data from Lin et al.

Average cell type fraction in each sample. Oligodendrocyte and AC-like tumor cells are relatively enriched in LTRs, while NPC-like, MES-like, Exhausted CD8+ T cells, and TAM4 are significantly enriched in STRs.

DISCUSSION

The enrichment of CN1 (MES-like tumor cells, TAMs, CD8+ T cells) and CN6 (NPC-like tumor cells, CD4+ T cells, TAMs) in STRs reveals a highly inflammatory spatial niche. Consistent with previous studies, MES-like tumor cells co-localized with myeloid and immune cells, particularly TAM5. This spatial alignment aligns with known macrophage-tumor interactions, where TAM-derived signals promote MES-like GBM transitions [2,4] and shape the TAM/MES-rich microenvironments preferential to STRs.

Unexpectedly, this immune-dense environment did not translate into a favorable therapeutic response. Although MES-like states are linked to heightened cytotoxic CD8+ T-cell infiltration [2], our findings demonstrate that elevated inflammatory infiltration alone is insufficient to drive effective anti-tumor immunity or enhance clinical response to ATL-DC vaccination.

FUTURE DIRECTION

Cellular Interactions: Map single-cell ligand-receptor interactions within specific CNs and expand to unmethylated MGMT cohorts.

Survival Mechanisms & Strategy: Uncover shared immune features in >10-year GBM survivors to optimize ATL-DC vaccines and future immunotherapies.

ACKNOWLEDGEMENTS

This work is supported by the UCLA Brain SPORE Developmental Research Program (P50 CA211015). We thank the Prins and Liao labs for identifying patient cohorts, sample quality control, collection, and insightful guidance, alongside the UCLA BIG Summer Program for project support.

REFERENCES

- Greenwald et al. *Cell* 2024
- Hara et al. *Cancer Cell* 2021
- Liao et al. *JAMA Oncology* 2023
- Lin et al. *Nature Neuroscience* 2026
- Neffel et al. *Cell* 2019



Emma Liao

**Computational
Biology - UCLA**

**Expected year
of graduation
June 2027**

Student: Emma Liao
Faculty Advisor: Chongyuan Luo
Direct Mentor: Cuining Liu
Department: Human Genetics
<https://luogenomics.github.io/>

Lineage Tracing via Natural Mitochondrial Variants and CNV Calling in Single Cell Data

EMMA K. LIAO^{1,2}, Emma Tu², Hannah Chao², Cuining Liu^{3,5}, Yu Sun, Terence W. Li^{3,4,5}, Justin Langerman⁶, Zach Von Behren³, Min Jen Tsai³, Yi Zhang³, Kathrin Plath⁶, Chongyuan Luo³

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Computational Biology, College of Letters of Science, UCLA

³ Department of Human Genetics, David Geffen School of Medicine, UCLA

⁴ Department of Computational Medicine, David Geffen School of Medicine, UCLA

⁵ Bioinformatics Interdepartmental Program, UCLA

⁶ Department of Biological Chemistry, David Geffen School of Medicine, UCLA

Methods such as pseudotime have been used to examine trajectories of cell development in single-cell data. However, pseudotime cannot map the clonal origin of individual cells, which is important for understanding variable starting populations. While there are wet lab methods to track the division of cells, they are expensive. This raises the question of whether lineage tracing can use natural variants to track the pathway of cells along their respective trajectories. This project compares lineage tracing methods against pseudotime to examine clonal variation, using fibroblast reprogramming into induced pluripotent stem cells. We investigated packages used for lineage tracing in single-cell data using natural mitochondrial variants and developed new CNV-calling methods and applied them to fibroblast reprogramming. Our pipeline demonstrates how lineage tracing can be applied to single-cell data to uncover clonal relationships underlying cell fate decisions, offering a generalizable framework for studying individual variability in reprogramming systems.

Lineage Tracing via Natural Mitochondrial Variants and CNV Calling in Single Cell Data

EMMA K. LIAO, Emma Tu, Hannah Chao, Cuining Liu, Terence W. Li, Justin Langerman, Yu Sun, Zach Von Behren, Min Jen Tsai, Yi Zhang, Kathrin Plath, Chongyuan Luo

ABSTRACT

Methods such as pseudotime have been used to examine trajectories of cell development in single-cell data. However, pseudotime cannot map the clonal origin of individual cells, which is important for understanding variable starting populations. While there are wet lab methods to track the division of cells, they are expensive. This raises the question of whether lineage tracing can use natural variants to track the pathway of cells along their respective trajectories. This project compares lineage tracing methods against pseudotime to examine clonal variation, using fibroblast reprogramming into induced pluripotent stem cells. We investigated packages used for lineage tracing using natural mitochondrial variants and developed new CNV-calling methods and applied them to fibroblast reprogramming single-cell data. Our pipeline demonstrates how lineage tracing can be applied to single-cell data to uncover clonal relationships underlying cell fate decisions, offering a generalizable framework for studying individual variability in reprogramming systems.

BACKGROUND

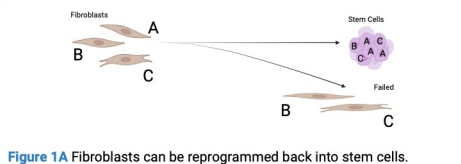


Figure 1A Fibroblasts can be reprogrammed back into stem cells.

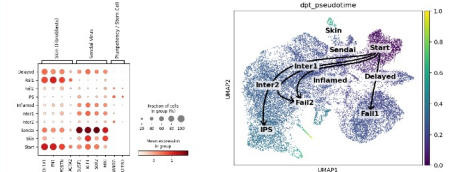


Figure 1B Marker genes used to annotate cell states. Figure 1C Trajectories of reprogramming fibroblasts back into stem cells.

METHODS

1 Mitochondrial Variant Method

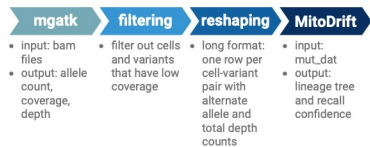


Figure 2A Lineage tracing pipeline based on mitochondrial variants. Using existing packages: mgatk and MitoDrift.

2 CNV Method

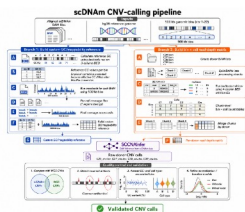


Figure 2B New CNV-calling method.

RESULTS

1 Lineage trees can be built from mitochondrial variants, but have poor precision.

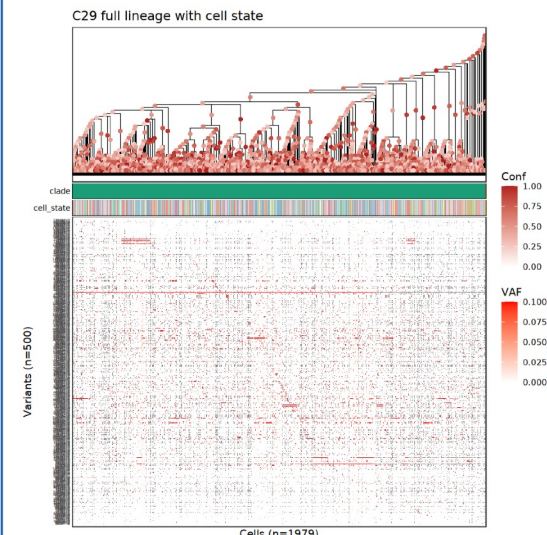


Figure 3A Lineage tree built from the top 500 variants from mgatk to MitoDrift. Almost all clones merge into one clade.

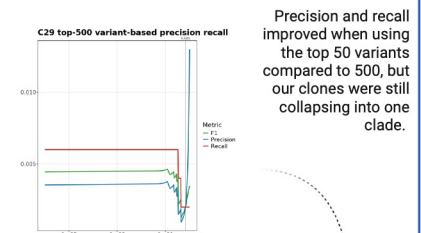


Figure 3B Very low precision and recall.

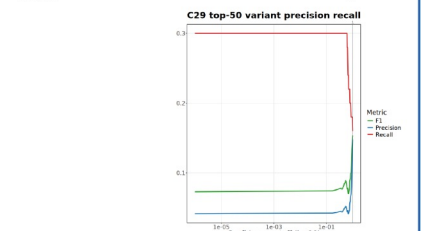


Figure 3C Improved precision and recall.

2 The CNV method gave us more balanced clades, but ~90% of cells still merged in one clade.

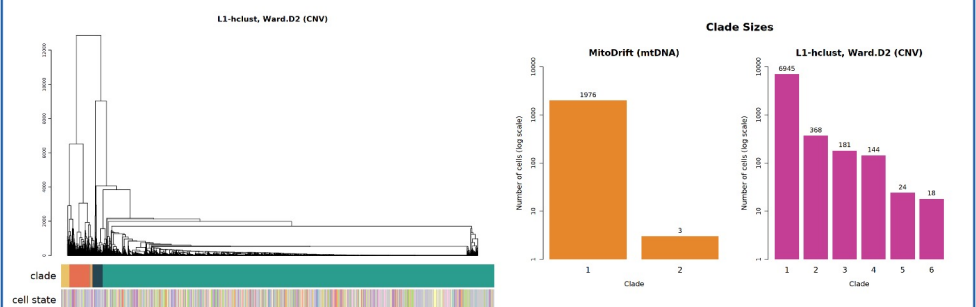


Figure 3D Lineage tree built from the CNVs using Ward.D2 clustering.

Figure 3E The lineage tree built from the CNVs with Ward.D2 clustering had more balanced clade distribution compared to that built from MitoDrift.

DISCUSSION

- Building trustworthy single-cell lineage trees was harder than expected
- mtDNA variants are too sparse/low-depth to produce reliable trees
- CNV clustering (Ward.D2) revealed real substructure that held up under QC checks (not driven by outlier bins or missing data)

FUTURE DIRECTIONS

- Continue to develop CNV-calling method
- More QC on CNV lineage tree method (cluster stability, strand-concordance filtering)
- Test additional donors and clustering approaches
- Study individual variation in fibroblast reprogramming

ACKNOWLEDGEMENTS

A special thank you to Cuining Liu, Dr. Chongyuan Luo, and the UCLA BIG Summer program for the mentorship and opportunities, Terence Li for sequencing and alignment support, Emma Tu for the CNVs, and Hannah Chao for the pseudotime background.



Student: Annie Linley
Faculty Advisor: Van Savage
Direct Mentor: Kai Akamatsu
Department: EEB, Computational Medicine
<https://savagelab.eeb.ucla.edu/>

Annie Linley

**Physics,
Mathematics -
UCLA**

**Expected year
of graduation
June 2027**

Modeling 1D Blood Flow Through Branching Vascular Networks to Describe Growth and Drug Delivery in Both Healthy Tissue and Tumors

ANNIE LINLEY^{1'2'4}, Kai Akamatsu², Van M. Savage^{2'3}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Departments of Computational Medicine and Ecology & Evolutionary Biology, UCLA

³ Santa Fe Institute

⁴ Departments of Physics and Mathematics, UVA

Medical imaging reveals that tumor vasculature differs substantially from that of healthy tissue: tumor vessels are less hierarchical, more tortuous, and less evenly distributed. Tumor vasculature is thus difficult to mathematically model, since it likely cannot be faithfully approximated by a self-similar straight-vessel bifurcating tree in the same way as whole-organism vasculature. My project makes progress on the goal of modeling tumor vasculature by combining one-dimensional Navier-Stokes fluid flow equations with power loss minimization techniques to create a more physically accurate fluid flow description of a healthy mammalian vascular system. I created an algorithm informed by this framework which, given the geometric and elastic parameters of the relevant vessels from either simulated or empirical networks, can describe the fluid flow comprehensively. This algorithm could be used to confirm the characteristics and volume of blood flow into and out of a tumor, potentially informing growth predictions, drug delivery, and treatment options.



Background and Introduction

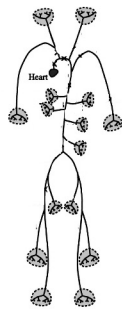


Figure 1. Schematic showing our branching model. The large vessels with a few major bifurcations and the smaller-vessel self-similar bifurcation trees [1].

Blood flow in healthy mammalian vascular systems is approximated well by a **one-dimensional mathematical model based on axisymmetric Navier-Stokes equations for tapering elastic vessels** [1]. However, these equations do not address evolutionary optimization: **energy loss from dissipation in small vessels and pulsatile wave reflections in large vessels is minimized** in real vasculature [2]. To create the most physically accurate description of healthy mammalian vasculature, we united these conditions using papers by Olufsen [1] and Savage [2, 3].

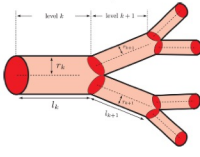


Figure 2. Schematic of vessel branching describing radius and length changes [2]. The vessel extends in the positive-x direction. The length of a vessel at level k is l_k , and the radius at a given x and t is r_k . The total number of vessels at level k is $N_k = 2^k$, since we consider only bifurcations.

Baseline equations

1D Navier-Stokes fluid flow averaged across vessel cross-sectional area is governed by three equations: the **continuity equation** ensuring blood is conserved at a bifurcation, the **momentum equation** describing the forces on the fluid, and, if needed, the **state equation** relating pressure to cross-sectional area. For large vessels, we have [1]:

$$\text{Continuity equation} \rightarrow \frac{\partial q}{\partial x} + \frac{\partial A}{\partial t} = 0$$

where q is the flow rate (vol/time) and $A = \pi r_k^2$ is the cross-sectional area.

$$\text{Momentum equation} \rightarrow \frac{\partial q}{\partial t} + \frac{\partial}{\partial x} \left(\frac{q^2}{A} + B \right) = -\frac{2\pi r_k \mu}{\delta A} \frac{\partial q}{\partial x} + \frac{\partial B}{\partial x} \frac{dr_k}{dx}$$

where $B(p(x,t), p_0(x,t))$ is the integral of the area over the pressure gradient, δ is the width of the boundary layer of blood, μ is the ratio of the fluid viscosity μ to its density ρ , and r_k is the equilibrium radius including vessel tapering.

$$\text{State equation} \rightarrow p(x,t) - p_0 = \frac{4Eh}{r_k} \left(1 - \sqrt{\frac{A}{A_0}} \right)$$

where p is the pressure, p_0 the equilibrium pressure, E is the Young's modulus of a vessel wall with thickness h , and $A_0 = \pi r_k^2$ is the cross-sectional area when $p = p_0$.

For small vessels, we have [1]:

$$\text{Continuity equation} \rightarrow i\omega C P + \frac{\partial Q}{\partial x} = 0$$

where C is the compliance of the vessel, and $Q = q e^{i\omega t}$ and $P(x) = p(x,t) e^{i\omega t}$ are the complex-valued flow rate and pressure, respectively.

$$\text{Momentum equation} \rightarrow i\omega U_x + \frac{\partial P}{\partial x} = \frac{\mu}{r_k} \frac{\partial}{\partial x} \left(\frac{\partial U_x}{\partial x} \right)$$

where $U_x(r,x) = u_x(r,x,t) e^{i\omega t}$ is the complex-valued fluid velocity.

The small vessel equations can be manipulated to give a **one-dimensional wave equation** for each of Q and P : $\frac{\partial^2 Q}{\partial x^2} + \frac{\partial^2 P}{\partial x^2} = 0$ or $\frac{\partial^2 P}{\partial x^2} = 0$, with wave velocity $c = \sqrt{\frac{A_0(1-F_2)}{\rho C}}$, where F_2 is a ratio of Bessel functions which emerges in the solutions.

Impedance

The large-vessel equations require a numerical technique to solve, but we can get Q and P analytically for the small vessels by solving their continuity and wave equations. We can thus calculate **impedance**, $Z(x,\omega) = \frac{P(x,\omega)}{Q(x,\omega)}$ for each vessel (think resistance to fluid flow) [1]:

$$Z(x,\omega) = \frac{i(b \cos(\omega x/c) - a \sin(\omega x/c))}{cC(a \cos(\omega x/c) + b \sin(\omega x/c))}$$

We can also calculate the **root impedance**, $Z(0,\omega)$, at the junction of the large and small vessels.

Power loss minimization

We use the method of Lagrange multipliers to minimize the function $\mathcal{P}$ describing power loss due to dissipation [3]:

$$\mathcal{P} = Q^2 Z_{\text{out}} + \lambda V_{\text{blood}} + \lambda_M M + \sum_{k=0}^N \lambda_k 2^k l_k^3$$

Here, Z_{out} is the total impedance of the network, $V_{\text{blood}} = \sum_{k=0}^N N_k \pi r_k^2 l_k$, M is the body mass, and the λ_k s are the Lagrange multipliers corresponding to each quantity. Once power loss is minimized, we conclude that the **most efficient scaling ratio for small vessels** is $\frac{l_{k+1}}{l_k} = 2^{-1/3}$ [3].

To minimize power loss due to reflection, we need $\frac{1}{Z_k} = \frac{2}{Z_{k+1}}$ to hold across all junctions in the large vessels. For a simplified Z_k , $\frac{Z_{k+1}}{Z_k}$ will give the **optimal branching ratio for large vessels** (Savage's group calculated $2^{-1/2}$) [3].

Boundary conditions

Now that we understand the physics of the blood flow in our chosen vascular architecture, we can impose the boundary conditions that let us unite the large- and small-vessel regimes:

	Inflow	Bifurcation	Outflow
Large	empirical value	$q_k = 2q_{k+1}$ $p_k = p_{k+1}$	$p(x_L, t)$
Small	large vessels	$\frac{1}{Z_k} = \frac{2}{Z_{k+1}}$	$Z_{\text{terminal}} = 0$

Table 1. The boundary conditions for each vessel size. Note conservation of flow and pressure continuity for large vessels and impedance matching and zero terminal resistance for small vessels. $p(x_L, t)$ is a convolution integral of q and a over time [1].

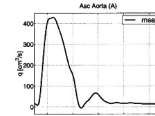


Figure 3. Flow into the ascending aorta captured with MRI. From [1].

The empirical inflow for large vessels would look something like Figure 3. The function represented can be inserted into the numerical solution for the large-vessel flow.

What does this look like at the transition point?

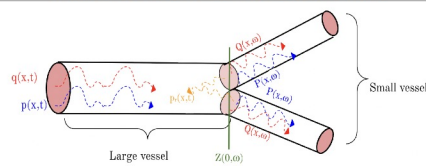


Figure 4. Schematic of calculated quantities at the transition from large to small vessel regime.

In Figure 4, we give a visual representation of the measurable/calculable quantities at the transition junction from large to small vessels. Of note:

- For large vessels, the non-linear pressure-area relationship leads to pulsatile waveforms expressed in the time domain, while for small vessels, a dominant viscous force term gives sinusoidal waveforms expressed in the frequency domain. We can go between the two domains using Fourier series [1].
- The $q(x,t)$ and $p(x,t)$ waveforms were sketched from those collected with magnetic resonance [1], like in Figure 3.
- $Q(x,\omega)$ and $P(x,\omega)$ are more regular sinusoids, and their amplitude is damped by the increased friction present in small vessels.
- We can match impedances at every level within the large artery structure to eliminate reflection [3], except at the transition to small vessels. Thus, the reflected wave $p_r(x,t)$ is shown at the junction.

How can we turn this into an algorithm?

We used all our equations to create a **simulated vascular system with physically realistic flow**. The physical parameters of this system can be estimated from literature (simulated), or, in the future, taken from empirical networks. A flowchart of the basic equation-to-function pipeline is given in Figure 4.

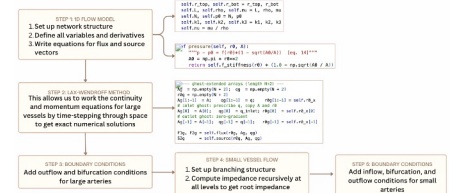


Figure 5. Flowchart of theory to algorithm.

Algorithm progress

We currently have a roughly accurate simulated single vessel. The steps for this code:

1. Define a vessel class with all physical parameters as variables.
2. Use variables to write equations like pressure and wave speed [1].
3. Establish a simple Poiseuille terminal resistance for the outlet boundary condition.
4. Solve the momentum and continuity equations using the **two-step Lax-Wendroff method** [1].
5. Check by placing probes down the vessel to measure flow and pressure at each point.

So, for the aortic arch ($r_0 = 1.14\text{cm}$, $l = 1.8\text{cm}$), we produced Figure 5:

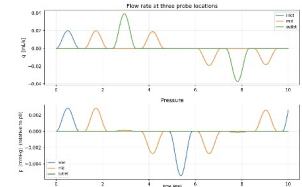


Figure 6. Simulated vessel results show the flow and pressure waves propagate in the correct direction. The outflow pressure is correctly nearly zero, and we even see the predicted reflection [1].

We also calculated the simulated wave speed $c = 7.384\text{ m/s}$, while the analytical wave speed was $c = 7.393\text{ m/s}$. The .124% relative error is largely because the Lax-Wendroff method produces only second-order accuracy in flow solutions.



Figure 7. Empirical network of lung arteries from a CT scan of a patient with non-small cell lung cancer.

The goal is to be able to extend this algorithm to model full systems like that shown in Figure 7. The vascular system of the tumor, shown in red, is less hierarchical, more tortuous, and less evenly distributed than in the lung. The healthy model provides insight into how these characteristics may affect blood flow in the tumor. If we can accurately extract characteristics of flow and pressure waves into the tumor, we could potentially predict patterns of growth or drug delivery.

References

- [1] M. S. Olufsen, C. S. P. de Vries, E. M. Pedersen, A. Nordin, and J. Larsen. Numerical simulation and experimental validation of blood flow in arteries with structured tree outflow conditions. *Annals of Biomedical Engineering*. 2013;41:1293–1299. DOI: 10.1007/s10439-013-0800-1. URL: <https://doi.org/10.1007/s10439-013-0800-1>.
- [2] Van M. Savage, Eric J. Olufsen, and Walter Fontana. Setting up abnormic scaling theory. *Fluid Computational Biology*. 4(1):0100017, 2009. doi: 10.1071/journal.pcbi.1000017. URL: <https://doi.org/10.1071/journal.pcbi.1000017>.
- [3] Van M. Savage, Eric J. Olufsen, and Walter Fontana. The SI: Energy loss in vascular networks. Supplementary Material for "Setting Up Abnormic Scaling Theory". *Fluid Computational Biology*. 4(1):0100017, 2009. URL: <https://doi.org/10.1071/journal.pcbi.1000017>.



Student: Livia Liu
Faculty Advisor: Feng Guo
Department: Biological Chemistry
<https://guolab.mbi.ucla.edu/>

Livia Liu

**Computational
Biology - UCLA**

**Expected year
of graduation
June 2027**

Designing antisense oligonucleotides for favoring specific RNA switch conformations

LIVIA LIU¹, Feng Guo²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Biological Chemistry, David Geffen School of Medicine, UCLA

Antisense oligonucleotides are short, synthetic DNA molecules that bind to target mRNA and allow for the modification of this mRNA through various means. RNA switches are RNA based regulatory mechanisms that adopt two alternative conformations with different effects on gene regulation. This project extends upon the work of Goodarzi et al, which only analyzed the effects of antisense oligonucleotides on the conformations on one RNA switch, the RORC RNA switch. Using the RNA sequence data for the other RNA switches that were discovered, we used ShapeRNA to visualize secondary structures for 10 RNA switches and attempted to design ASOs that would favor certain conformations of these RNA switches. We are also developing a computer based program to automate the generation of potential ASOs for RNA switches based on secondary structure, which can help contribute to future studies focusing on ASOs and targeting specific RNA switches, in terms of convenience.

Designing antisense oligonucleotides for favoring specific RNA switch conformations

LIVIA LIU, Feng Guo

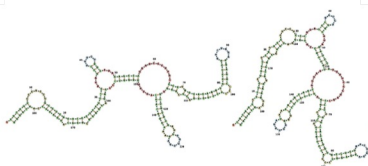
Guo Lab, Department of Biological Chemistry, University of California, Los Angeles

Abstract

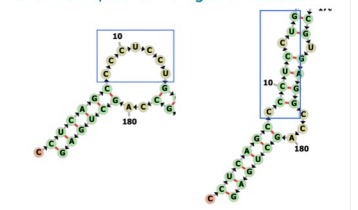
Antisense oligonucleotides are short, synthetic DNA molecules that bind to target mRNA and allow for the modification of this mRNA through various means. RNA switches are RNA based regulatory mechanisms that adopt two alternative conformations with different effects on gene regulation. This project extends upon the work of Goodarzi et al, which only analyzed the effects of antisense oligonucleotides on the conformations on one RNA switch, the RORC RNA switch. Using the RNA sequence data for the other RNA switches that were discovered, we used ShapeRNA to visualize secondary structures for 10 RNA switches and attempted to design ASOs that would favor certain conformations of these RNA switches.

Methodology

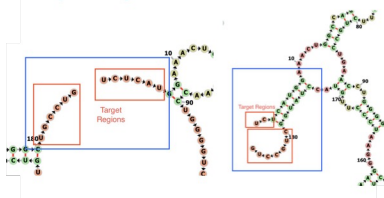
All of the testing in the study was performed in silico using ShapeRNA. Predictions of the conformations were obtained using ShapeRNA's secondary structure ensemble prediction.



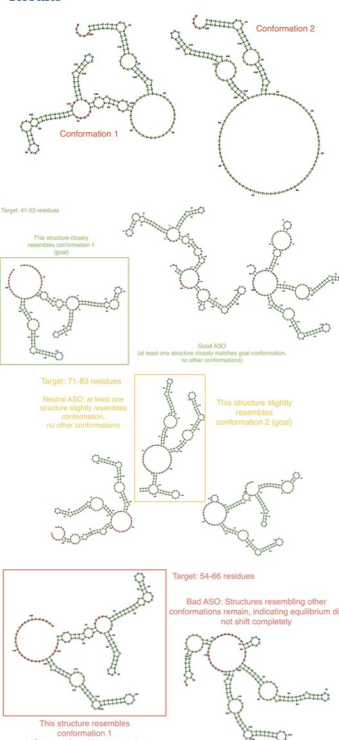
Strategy: finding areas in the RNA where the residues were bound in a hairpin stem in the other conformations, but not in the goal conformation.



Alternative strategy: Stabilize hairpin structures unique to the goal conformation.



Results

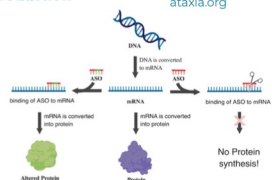


ASO List - Example

Target	Complementary / ASO	Position	#	Switch
CACAGAUUGGAA	GTGTCTAGACCTT	41-53	1	1
UCACAGAUUGGAA	AGTGTCTAGACCTT	40-53	1	1
CAUUUUUUUUCUC	GTAATAAAAGAG	29-41	1	1
CAUUUUUUUUCUCA	GUAAAUAAGAGU	29-42	1	1
GCUCAGGUCCAA	CGACGUCCAGGUU	54-66	2	2
GCUCAGGUCCAAC	CGACGUCCAGGUUG	54-67	2	2
GAGGGUGGGUGG	CUCCCCACCCACC	71-83	2	2
GAGGGUGGGUGGG	CUCCCCACCCACC	71-84	2	2

Introduction

Image Source: ataxia.org



- RNA switches change conformation in response to binding of specific metabolites.
- ASOs can alter the expression of mRNA through various means, such as inducing inclusion/skipping of exons or degradation of mRNA.
- ASOs can be used to "shift" the equilibrium between the conformations of an RNA structural switch.

Discussion / Future Goals

- Finding a "good" ASO with good or perfect results is actually difficult. Most ASOs turned out to have "neutral" or "bad" results.
- As an effective program for comparing secondary structures of RNA sequences could not be found, most of our analysis/results were based on objective and observation based judgement.
- One future goal is to use the ASO design process practiced manually in this study to create a computational program/process to facilitate the generation of ASOs for a switch based on secondary structure.
- We would also like to test the responses of the RNA switch sequences to ASO binding in vitro, using similar methods to those outlined in Goodarzi et al.

Acknowledgements

I would like to thank Professor Guo for offering me advice and guidance throughout the duration of this study, as well as caring about my progress and regularly checking in and discussing with me when there was time.



Student: Yinuo Liu

Faculty Advisor: Grace Xinshu Xiao

Direct Mentor: Anna Fraser

Department: Integrative Biology and Physiology

<https://xiaolab.ibp.ucla.edu/>

Yinuo Liu

**Computational
and Systems
Biology - UCLA**

**Expected year
of graduation
June 2027**

Characterizing Bidirectional Transcription and its Link to Neuroinflammation in Parkinson's Disease

YINUO LIU^{1,2}, Anna Fraser³, Xinshu Grace Xiao⁴

¹ BIG summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Computational and Systems Biology Interdepartmental Program, UCLA

³ Molecular Biology Interdepartmental Program, UCLA

⁴ Integrative Biology and Physiology, UCLA

Parkinson's Disease (PD) is a neurodegenerative disease with motor and cognitive symptoms marked by Lewy Bodies (LB) that trigger neuroinflammation and eventually lead to dopaminergic cell death. Previous studies have shown that intra-molecular double-stranded RNA (dsRNA) play a role in cancer, autoimmune, and neurodegenerative diseases. However, the contribution of intermolecular dsRNA from bidirectional transcription is undercharacterized. In this study, we characterized bidirectional transcript configuration and expression from PD-affecting brain regions in patient and control samples to investigate immune-related functions. Our results suggest bidirectional transcripts that overlap transposable elements are enriched for innate immune-related genes and could serve as a potential source of dsRNA that contribute to neuroinflammation in PD. These findings could uncover a novel pathway linking the role of non-coding regions sourcing dsRNA with innate immunity and major PD pathogenesis.

Characterizing Bidirectional Transcription and its Link to Neuroinflammation in Parkinson's Disease

Yinuo Liu,¹ Anna M.J. Fraser,² Xinshu Grace Xiao³

1. Computational and Systems Biology Interdepartmental Program, UCLA

2. Molecular Biology Interdepartmental Program, UCLA

3. Integrative Biology and Physiology, UCLA



Abstract

Parkinson's Disease (PD) is a neurodegenerative disease with motor and cognitive symptoms marked by Lewy Bodies (LB) that trigger neuroinflammation and eventually lead to dopaminergic cell death. Previous studies have shown that intra-molecular double-stranded RNA (dsRNA) play a role in cancer, autoimmune, and neurodegenerative diseases. However, the contribution of intermolecular dsRNA from bidirectional transcription is undercharacterized. In this study, we characterized bidirectional transcript configuration and expression from PD-affecting brain regions in patient and control samples to investigate immune-related functions. Our results suggest bidirectional transcripts that overlap transposable elements are enriched for innate immune-related genes and could serve as a potential source of dsRNA that contribute to neuroinflammation in PD. These findings could uncover a novel pathway linking the role of non-coding regions sourcing dsRNA with innate immunity and major PD pathogenesis.

Background

- Exogenous dsRNA from virus is known for triggering antiviral immune pathways.²
- Endogenous dsRNA is "marked" by ADAR1 editing to prevent autoimmune responses³. However, this is also a vulnerable mechanism that could be exploited by cancer, autoimmune, and neurodegenerative diseases.⁵
- dsRNA is recognized by sensor proteins and leads to the production of Type I interferon and proinflammatory cytokines.^{2,3}
- LB is a hallmark of PD and could trigger neuroinflammation that perpetuates neurodegeneration.⁴
- Caudate Nucleus and Putamen are two brain regions connected to the nigrostriatal pathway, making them vulnerable to dopamine deficiency due to loss of neuromelanin. Lewy Body is tagged by ubiquitination for autophagy.

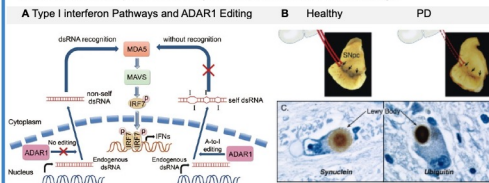


Fig. 1A Reprinted from [3] dsRNA is identified by sensors like MDA5 where it triggers the interferon pathway producing proinflammatory cytokines. ADAR1 editing distinguishes exogenous dsRNA with endogenous "self" dsRNA where edited dsRNA is no longer sensed by MDA5. **Fig. 1B, 1C Adapted from [4]** PD SNpc cells are depigmented due to loss of neuromelanin. Lewy Body is tagged by ubiquitination for autophagy.

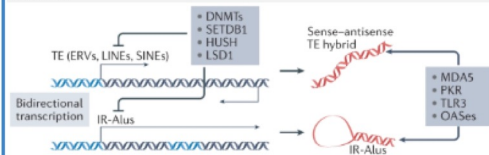


Fig. 1D Adapted from [2] Endogenous dsRNA can be intra-molecular (folding of the same strand due to IR-Alu) or inter-molecular (bidirectional transcription).

Method

Pipeline filter for uniquely-mapped bidirectional regions as inputs for downstream analysis:
Strand Correlation, Transcript Configuration, Transposable Element Overlap

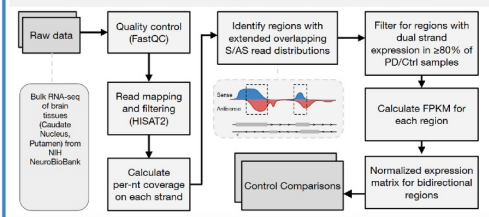
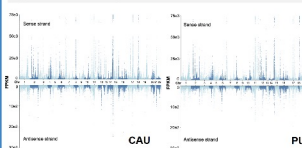
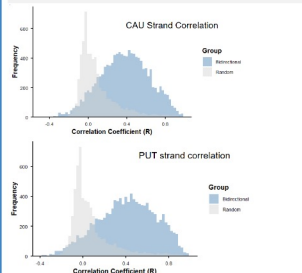


Fig. 2 Bidirectional Regions in Caudate Nucleus and Putamen



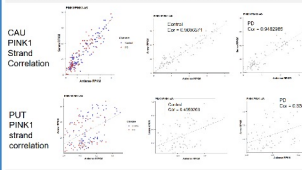
Symmetric relative expression on both S/AS strands show bidirectional regions found across the genome for CAU and PUT

Fig. 3 Strand Correlation of Bidirectional Regions



Bidirectional regions showed significant strand correlation compared to random for both brain regions as anticipated

Fig. 4 Strand Correlation Case study of PINK1/PINK1-AS



PINK1 is a PD-relevant gene where PINK1-AS regulates PINK1 expression. CAU has significantly better strand correlation for PINK1. Control samples of PUT have greater correlation than PD.

Results

Fig. 5 Transcript Configuration and Gene Annotation

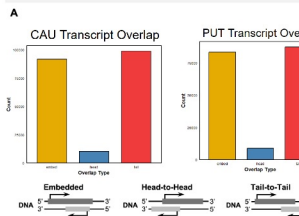


Fig. 7

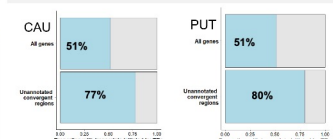


Fig. 7 Unannotated convergent regions contains greater proportion of transcripts starting in a TE than a random region in the genome

Fig. 8 Gene Ontology Analysis

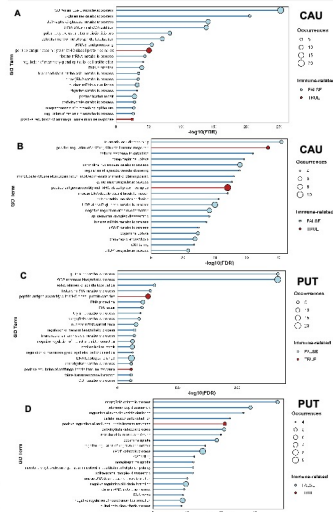


Fig. 8B, 8D (antisense unannotated) show slightly more significant enrichment in innate immune pathways than Fig. 8A, 8C (both anno.)

Fig. 5A Most transcripts are tail or embedded configuration.

Fig. 5B Most transcripts overlap exons and 3'UTR.

Fig. 6 Transcription Units Initiate in Transposable Elements

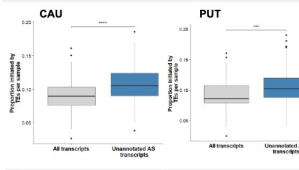


Fig. 6 As shown in the boxplots for both brain regions, there are significantly more transcripts from Unannotated Antisense regions that initiate in transposable elements than a random transcript.

Conclusions

- 1,587 bidirectional regions were found for CAU, 5525 for PUT, among which 30% of all regions have previously unannotated antisense strand.
- Strong strand correlation, indicating coregulated transcription that supports the formation of dsRNA and the stabilization of the sense gene by the binding of antisense strand.
- Most bidirectional regions overlap a transcription unit that initiates in a TE, and the proportion remains high for transcripts in unannotated convergent regions.
- Unannotated Antisense regions show enrichment for innate immune genes.
- Uncovered a set of previously unannotated antisense transcripts that may contribute to the activation of innate immunity through dsRNA formation.

Future Directions

- ISG Correlation of bidirectional region expression in control and patient samples
- Analyze proportion of ADAR1 editing to gain insight on whether these bidirectional regions are marked as "self" in control and patient samples
- Explore energetic favorable folding to see how well the pipeline captured intermolecular dsRNA sources as opposed to intramolecular dsRNA.

References

1. Irmady K, et al. Blood transcriptomic signatures associated with molecular changes in the brain and clinical outcomes in Parkinson's disease. *Nat Commun*. 2023;14(1):3956. doi:10.1038/s41467-023-39562-6
2. Chen YG, Hur S. Cellular origins of dsRNA, their recognition and consequences. *Nat Rev Mol Cell Biol*. 2022;23(4):296-301. doi:10.1038/s41586-021-00430-1
3. Zhou Y, Chen T, Cao X. RNA editing by ADAR1 marks dsRNA as "self". *Cell Res*. 2015;25(12):1283-1284. doi:10.1038/cr.2015.135
4. Dauer W, Przedborski S. Parkinson's disease: mechanisms and models. *Neuron*. 2003 Sep 11;39(6):889-909. doi:10.1016/S0969-6473(03)00568-3
5. Xie X, Weng Y, Gu J, Wang J, Liu J, Zhao X, Jin Z. Cellular endogenous double-stranded RNA: From pathophysiology to therapeutic implications. *Genes Dis*. 2024; In Press. doi:10.1016/j.gendis.2024.101345

Acknowledgements

- Thanks to Xiao Laboratory members for their support, guidance, and valuable contributions to this work.
- Thanks to BIG summer program faculty and students for their encouragement and support to this work.





Student: Sami Majid

Faculty Advisor: Amjad Askary

Direct Mentor: Mykel Barrett

Department: Molecular, Cell, and Developmental Biology (MCDB)

<https://amjad.mcdb.ucla.edu/>

Sami Majid

Biology - UCLA

**Expected year
of graduation
June 2029**

Genome-Wide Characterization of Ligand-Specific Transcriptional Responses to BMP2, BMP10, and BMP2+GDF5 in C2C12 Myoblasts

SAMI MAJID^{1,2}, Amir Shami², Amjad Askary^{2,3,4}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Molecular, Cell, and Developmental Biology, UCLA

³ Institute for Quantitative and Computational Biosciences, UCLA

⁴ Brain Institute Research, UCLA

BMP2, BMP10, and BMP2+GDF5 have been reported to elicit distinct phenotypic outcomes in C2C12 myoblasts, but the genome-wide transcriptional changes underlying these differences remain uncharacterized, such as the specific genes and timepoints at which they arise. We treated C2C12 myoblasts with BMP2, BMP10, or BMP2+GDF5 (combined), alongside untreated controls, over a one-, two-, and three-day time course and performed transcriptome-wide RNA sequencing on all 12 samples. Differential expression analysis is currently being done to identify genes that respond similarly across ligands versus those that are specific to the BMP ligands we focused on, with attention to osteogenic and myogenic markers. Defining these shared and ligand-specific transcriptional markers will help provide a foundation for future studies into how distinct BMP ligands generate specificity in downstream signaling and cell fate outcomes.

Genome-Wide Characterization of Ligand-Specific Transcriptional Responses to BMP2, BMP10, and BMP2+GDF5 in C2C12 Myoblasts

SAMI MAJID^{1,2}, Amir Shami², Amjad Askary^{2,3,4}

¹BIG Summer Program, IQCB, UCLA ²Dept. of Molecular, Cell, and Developmental Biology, UCLA ³Institute for Quantitative and Computational Biosciences, UCLA ⁴Brain Institute Research, UCLA



1. BACKGROUND + HYPOTHESIS

Background

- BMP signaling can redirect C2C12 myoblasts from muscle toward bone fate.
- Different BMP ligands may activate different transcriptional programs, but the genome-wide differences were previously uncharacterized.

Hypothesis

- All three ligands share a core BMP response, driven by overlapping SMAD signaling.
- BMP2 most strongly induces osteogenic genes, based on prior literature.

BMP-SMAD Signaling

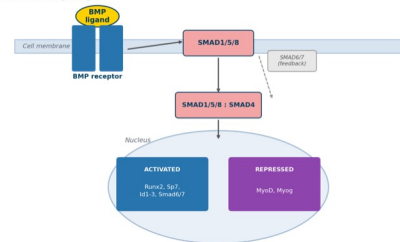


Figure 1. BMP ligand binding activates SMAD1/5/8, which complexes with SMAD4 to activate osteogenic genes and repress myogenic genes.

Experimental Design

C2C12 myoblasts treated across 3 time points with 4 conditions

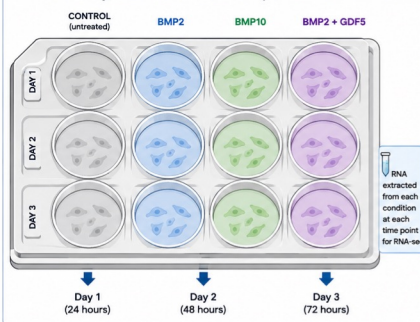


Figure 2. C2C12 myoblasts treated with Control, BMP2, BMP10, or BMP2+GDF5 across a multi-day time course; RNA sequenced at each timepoint.

4

conditions

13,212

genes

9

libraries

Why C2C12 + BMP?

- C2C12 is a well-established myoblast line that reliably differentiates into muscle, making lineage switches easy to detect.
- BMP-driven conversion to bone-forming cells is one of the clearest examples of signal-driven cell fate change in developmental biology, and a strong system for testing ligand specificity.

2. METHODS

Sequencing (PacBio/illumina)

- FastQC, fastp trimming, STAR alignment, UMI deduplication, featureCounts.
- This Analysis
- Combined 3 batches with median-of-ratios normalization; filtered to 13,212 of 78,409 genes.
- Moderated linear model (log2 CPM) with empirical Bayes shrinkage, blocking on day.

Analysis Pipeline

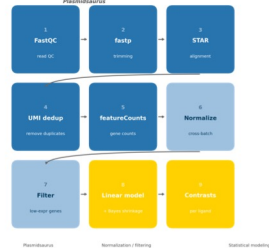


Figure 3. Steps performed by PacBio/illumina versus this analysis.

3. RNA-SEQ DATA + QC

Batch Effect Check

- One sequencing batch was sequenced far shallower than the other two; that batch, not treatment, is the largest source of variation (PC1).

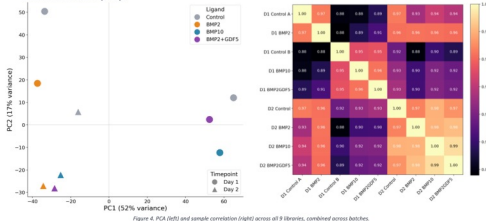


Figure 4. PCA (left) and sample correlation (right) across all 9 libraries, combined across batches.

Top Candidate Genes

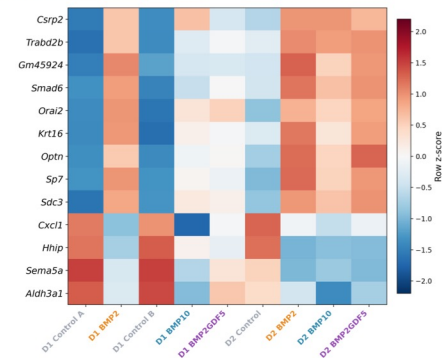


Figure 5. Expression (row z-score) of top candidate genes across all 9 libraries.

4. KEY FINDINGS

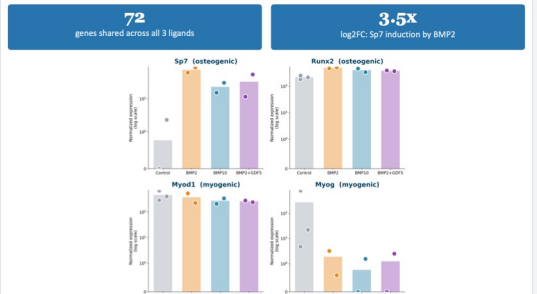


Figure 6. Sp7 and Runx2 (osteogenic) rise with treatment; MyoD1 and Myog (myogenic) fall, most strongly under BMP2.

5. RESULTS + DISCUSSION

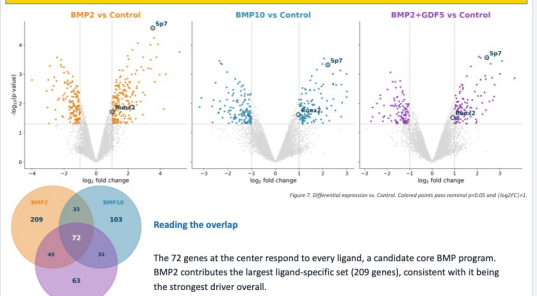


Figure 7. Differential expression vs. Control. Colored points pass nominal $p < 0.05$ and $|\log2FC| > 1$.

Figure 8. Overlap of candidate genes altered by each ligand relative to control.

Discussion

- All three ligands engage the same core BMP-SMAD pathway at different strengths; BMP2 is the most potent driver of the osteogenic switch.

6. FUTURE DIRECTIONS

From Candidate Genes to Mechanism



Why it matters

Mapping ligand-specific BMP programs helps explain how the same signaling family drives different outcomes in bone and muscle, informing future regenerative and precision-signaling strategies.

Figure 9. Day 3 sequencing (left) at our provider; next steps validate candidates and extend to more ligands.

Student: Charissa Mak
Faculty Advisor: Van Savage
Direct Mentor: Kai Akamatsu
Department: EEB, Computational Medicine
<https://savagelab.eeb.ucla.edu/>



Charissa Mak

**Molecular, Cell,
and
Developmental
Biology - UCLA**

**Expected year
of graduation
June 2027**

Quantitative PET-CT Radiomics Characterization of Tumor Metabolism and Vascular Architecture in Non-Small Cell Lung Cancer

CHARISSA MAK¹, Kai Akamatsu², Van M. Savage^{2, 3, 4}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Computational Medicine, David Geffen School of Medicine, UCLA

³ Department of Ecology and Evolutionary Biology, UCLA

⁴ Santa Fe Institute

Non-small cell lung cancer (NSCLC) is characterized by abnormalities in tumor metabolism and vascular structure that contribute to differences in tumor phenotype, progression, and clinical outcomes. We investigated whether quantitative PET-CT-derived biomarkers could differentiate the two most common NSCLC histological subtypes—adenocarcinoma and squamous cell carcinoma—and classify clinical stage. We developed a multimodal computational pipeline that extracted metabolic, volumetric, and vascular branching and scaling features from co-registered PET-CT images and integrated them into supervised machine learning models. Machine learning analyses demonstrated that quantitative PET-CT biomarkers more accurately classified clinical stage than histological subtype. Histological classification relied primarily on metabolic intensity features, whereas clinical stage classification benefited from integrating structural vascular metrics with volumetric metabolic features, improving cross-validated model performance. These findings demonstrate that complementary metabolic and vascular biomarkers capture distinct aspects of NSCLC biology and support multimodal PET-CT analysis as a source of robust, non-invasive biomarkers for precision oncology.

Quantitative PET-CT Radiomics Characterization of Tumor Metabolism and Vascular Architecture
in Non-Small Cell Lung CancerCharissa Mak¹ Kai Akamatsu² Van M. Savage^{2,3,4}¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA²Department of Computational Medicine, David Geffen School of Medicine, UCLA³Department of Ecology and Evolutionary Biology, UCLA⁴Santa Fe Institute

Abstract

Non-small cell lung cancer (NSCLC) is characterized by abnormalities in tumor metabolism and vascular structure that contribute to differences in tumor phenotype, progression, and clinical outcomes. We investigated whether **quantitative PET-CT-derived biomarkers** could differentiate the two most common NSCLC histological subtypes—adenocarcinoma and squamous cell carcinoma—and classify clinical stage. We developed a multimodal computational pipeline that extracted metabolic, volumetric, and vascular branching and scaling features from co-registered PET-CT images and integrated them into supervised machine learning models. **Machine learning analyses** demonstrated that quantitative PET-CT biomarkers more accurately classified clinical stage than histological subtype. **Histological classification** relied primarily on metabolic intensity features, whereas **clinical stage classification** benefited from integrating structural vascular metrics with volumetric metabolic features, improving cross-validated model performance. These findings demonstrate that complementary metabolic and vascular biomarkers capture distinct aspects of NSCLC biology and support multimodal PET-CT analysis as a source of robust, non-invasive biomarkers for precision oncology.

Introduction

- Clinical Need:**
- NSCLC-dominated by **Adenocarcinoma (ADC)** and **Squamous Cell Carcinoma (SCC)**—requires non-invasive, objective tools for precise histological subtyping and clinical staging [1, 2].
- Biological Mechanism & Vascular Scaling parameters:**
- Accelerated tumor expansion triggers both **pathological angiogenesis** (yielding altered vascular branching geometry on CT) and **elevated glycolysis** (causing metabolic heterogeneity on PET) [3].
 - Vascular architecture is modeled across branch levels $k \rightarrow k+1$ using fundamental scaling parameters (Fig. 1):

Radius scaling ratio (β): Quantifies vessel tapering across bifurcations:

$$\beta = \frac{r_{k+1}}{r_k}$$

Length scaling ratio (γ): Quantifies segment length attenuation:

$$\gamma = \frac{l_{k+1}}{l_k}$$

Branching Asymmetry ($\Delta\beta, \Delta\gamma, \Delta\beta, \Delta\gamma$): Quantifies structural heterogeneity between daughter vessels (1, 2) at a bifurcation node ($\Delta\beta = |\beta_1 - \beta_2|, \Delta\gamma = |\gamma_1 - \gamma_2|$) and averaged across the full tumor vascular network ($\bar{\Delta\beta}, \bar{\Delta\gamma}$).

- Hypothesis & Objective:**
- Integrating CT-derived structural and vascular metrics with PET metabolic features improves **machine learning classification** of clinical stage and histological subtype.

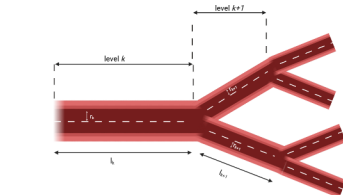


Figure 1. Vessel Bifurcation Geometry & Scaling Parameters. Schematic of a single vascular bifurcation showing parent vessel radius r_k branching into daughter vessels r_{k+1} , defining the radius ratio $\beta = r_{k+1}/r_k$ and length ratio $\gamma = l_{k+1}/l_k$ extracted via Angiart [3, 4].

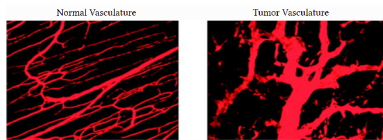


Figure 2. Structural Comparison of Normal vs. Tumor Vasculature. Microscopic view highlighting the contrast between the organized, hierarchical branching geometry of normal microvasculature (left) and the chaotic, disorganized, and hypervascular branching characteristic of tumor angiogenesis (right) [5].

Dataset, Cohort & Radiomics Pipeline

Dataset & Analytic Cohort:

Data Source: Co-registered PET-CT scans from **TCIA Lung-PET-CT-Dx dataset** [6].

Histology Cohort ($n = 99$): 75 **Adenocarcinoma (ADC)** vs. 24 **Squamous Cell Carcinoma (SCC)**.

Clinical Stage Cohort ($n = 102$): 68 **Early Stage (I–II)** vs. 34 **Advanced Stage (III–IV)**.

Inclusion Criteria: Complete primary tumor 3D PET/CT volume, annotated DICOM **metadata**, and pathologically confirmed staging.

Multimodal Feature Extraction Pipeline:

CT Vessel Skeletonization: Tumors segmented via **3D U-Net (lungtumorsnet)** [7] and clipped to annotations. Lung CT vessels enhanced via **Hessian** filtering and skeletonized via **Angiart** [4] to extract topological metrics (r_k, β, γ , vessel counts).

PET SUV Quantification: 3D CT tumor masks mapped to PET voxel space via DICOM physical transformation matrices [6]; voxel intensities converted to **Body-Weight Standardized Uptake Values (SUV)**.

Feature Fusion & Preprocessing: Extracted metabolic (SUV_{max}, SUV_{avg} , MTV, TLG, Tumor:Background Ratio) and vascular metrics. Missing values median-imputed across cohort prior to bootstrap-LASSO feature selection.

Pipeline Architecture

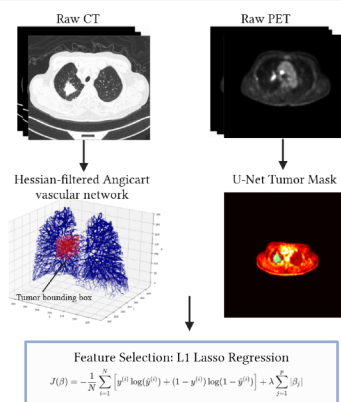


Figure 3. End-to-End Multimodal Workflow. Automated U-Net tumor segmentation and PET SUV conversion run in parallel with Hessian vessel filtering and Angiart skeletonization for feature fusion and classification.

Pre-Machine Learning Statistical Analyses

Mann-Whitney U testing and Cohen's d effect sizes were evaluated across candidate features ($n = 99$ histology; $n = 102$ clinical stage). High feature redundancy was observed across raw candidate metrics prior to stability selection.

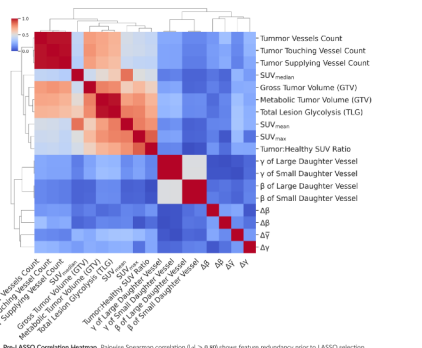


Figure 4. Pre-LASSO Correlation Heatmap. Pairwise Spearman correlation (ρ) at $n = 98$ shows feature redundancy prior to LASSO selection.

Clinical Stage ($n = 102$)					Histological Subtype ($n = 99$)				
Feature	Early ($n=34$)	Adv. ($n=68$)	p -val	Cohen's d	Feature	ADC ($n=75$)	SCC ($n=24$)	p -value	Cohen's d
Tumor Healthy Ratio	9.81 ± 3.44	20.09 ± 19.63	7.4×10^{-4}	0.89	SUV_{max}	6.30 ± 11.53	17.73 ± 11.87	1.2×10^{-4}	1.01
SUV_{max}	1.54 ± 1.19	3.06 ± 1.63	1.2×10^{-4}	1.19	SUV_{avg}	1.94 ± 1.89	2.14 ± 1.66	0.0140	0.50
MTV	7.08 ± 10.05	37.22 ± 10.22	2.8×10^{-4}	1.17	TLG	13.42 ± 75.94	55.63 ± 96.48	0.0049	0.06
GTV	22.04 ± 66.88	107.7 ± 219.2	1.3×10^{-4}	0.86					
Tumor vessels	21.00 ± 20.50	41.00 ± 31.00	0.0016	0.84					
γ	0.71 ± 0.32	0.67 ± 0.20	0.0005	-0.42					
Length asymmetry	0.73 ± 0.01	0.71 ± 0.01	0.0258	-0.48					

Table 1. Clinical Stage Feature Comparison. Mann-Whitney U test p -values and Cohen's d effect sizes comparing Early (I–II) vs. Advanced (III–IV) NSCLC cohorts.

Histological Subtype Classification (ADC vs. SCC)

Performance across 5-fold cross-validation on the full median-imputed cohort ($n = 99$), reported as mean $\pm$ SD across folds. Top-performing metric per column is highlighted in **bold**.

Model	Sens.	Spec.	Prec.	F1	AUC	Acc.
L1 Logistic	0.55±0.20	0.65±0.13	0.34±0.12	0.41±0.13	0.69±0.10	0.62±0.10
Random Forest	0.30±0.21	0.88±0.10	0.45±0.32	0.33±0.21	0.76±0.09	0.74±0.07
SVM (RBF)	0.57±0.24	0.77±0.12	0.45±0.17	0.49±0.16	0.74±0.10	0.72±0.09

Table 3. Full-cohort CV metrics (mean $\pm$ SD) for histological subtyping (ADC vs. SCC).

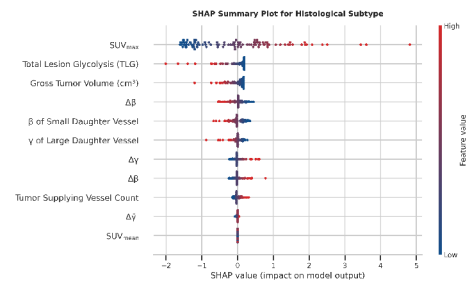


Figure 5. SHAP Feature Importance (Histology Subtype). Feature contributions demonstrate primary reliance on PET metabolic intensity (SUV_{max}).

Clinical Stage Classification (Early vs. Advanced)

Performance across 5-fold cross-validation on the full median-imputed cohort ($n = 102$), reported as mean $\pm$ SD across folds. Top-performing metric per column is highlighted in **bold**.

Model	Sens.	Spec.	Prec.	F1	AUC	Acc.
L1 Logistic	0.62±0.19	0.72±0.12	0.54±0.14	0.57±0.14	0.76±0.11	0.69±0.10
Random Forest	0.56±0.22	0.80±0.11	0.59±0.18	0.55±0.17	0.82±0.09	0.72±0.09
SVM (RBF)	0.62±0.21	0.67±0.13	0.49±0.13	0.53±0.14	0.73±0.11	0.65±0.10

Table 4. Full-cohort CV metrics (mean $\pm$ SD) for clinical stage classification (Early I–II vs. Advanced III–IV).

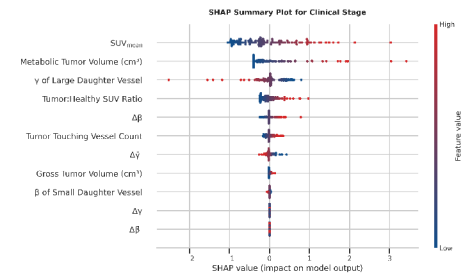


Figure 6. SHAP Feature Importance (Clinical Stage). SHAP summary values demonstrate joint contributions from metabolic burden and vascular metrics.

Discussion & Future Directions

Conclusions & Biomarker Drivers:

- **Histology:** Subtyping (ADC vs. SCC) is driven primarily by metabolic intensity.
- **Clinical Stage:** Advanced-stage tumors exhibit significantly higher tumor-touching vessel counts and metabolic intensity.

Translational Implications:

- **Non-Invasive Diagnostics:** Multimodal radiomic signatures provide a scalable alternative to physical biopsies [1].
- **Clinical Trials:** Enables oncologists and drug developers to evaluate anti-angiogenic and metabolism-targeting therapies.
- **Early-Stage Risk Stratification:** Scoring models can identify "high-risk" Stage I–II patients carrying aggressive microenvironmental features despite small primary sizes, prioritizing them for intensive treatment.

Methodological Limitations & Future Work:

- **Vascular Scaling:** Implement Horton-Strahler for length scaling (γ) to preserve true anatomical lengths and normalize fractal dimension (D) predictions.
- **Microenvironments & Segmentation:** Incorporate Hessian-based structural micro-environment detection.
- **Data Imputation:** Replace current median imputation with deep learning methods for missing values.

Acknowledgements & References

Acknowledgements: Many thanks to all members of the Savage Lab for their guidance, feedback, and support. We gratefully acknowledge The Cancer Imaging Archive (TCIA) for providing access to the Lung PET-CT data dataset. Supported by the Brauer-In-Germania (BIG) Summer Program and QCBio at UCLA.

References:

1. Jemal A, Murray T, Samuels A, et al. Cancer statistics, 2019. *CA Cancer Clin Oncol*. 2019;69(1):20–34.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of the worldwide incidence of and mortality from major cancer sites. *CA Cancer Clin Oncol*. 2018;68(6):694–715.
3. Brunner AB, Savage VM, Cline R, et al. A multi-scale approach to quantifying tumor heterogeneity using a quantitative radiomics approach. *Nat Commun*. 2018;9:4005.
4. Neufeldt M, Driesch M, Kopp M. Scaling Properties of Biological Vascular Tree-like Arterioles. *PLoS One*. 2015;10(12):e0140455.
5. Gao S, Durr DC, Hill M, et al. The Cancer Imaging Archive (TCIA): maintaining and archiving a public information repository. *J Digit Imaging*. 2013;26(5):505–527.
6. Cline R, Neufeldt M, Kopp M, et al. The Cancer Imaging Archive (TCIA): maintaining and archiving a public information repository. *J Digit Imaging*. 2013;26(5):505–527.
7. Frokwin V, Sivas DCM, Anderson A, Lange T, Kitz G, Lindner P. Teacher student approach for lung tumor segmentation from mixed supervised datasets. *PLoS One*. 2021;16(6):e0245147.



Student: Emma Maliar
Faculty Advisor: Jeffrey Chiang
Department: Neurosurgery/Computational
Medicine
<https://chianglab.healthsciences.ucla.edu/>

Emma Maliar

**Computational
and Systems
Biology - UCLA**

**Expected year
of graduation
June 2027**

Interpretable Sparse Autoencoder Features for TBI Prognosis from Multimodal MRI

EMMA MALIAR¹, Jeffrey Chiang²

¹ BIG Summer Program, QCBio, University of California, Los Angeles

² Department of Computational Medicine, University of California, Los Angeles

Traumatic brain injury (TBI) is the leading cause of death and disability in people under 40. Current prognostic standards like CRASH rely solely on clinical variables (Glasgow Coma Scale and pupillary reactivity), often reporting high false-positive rates for unfavorable prognoses, increasing the risk for premature withdrawal of life-sustaining therapy. Multimodal MRI captures detailed information about hemorrhaging, microstructural damage, and diffuse axonal injury that clinical variables miss entirely. Prior work has shown that foundation models can extract features from T1, T2, SWI, and ADC MRI and classify six-month functional outcomes. However, a key gap for clinical adoption remains: it is unclear which brain regions drive model predictions. Therefore this project aims to yield interpretable, spatially localized features by applying sparse autoencoder methods to foundation model embeddings of multimodal TBI MRI data.

Interpretable Sparse Autoencoder Features for TBI Prognosis from Multimodal MRI

Emma Maliar¹, Jeffrey N. Chiang²

1 BIG Summer Program, QCBio, UCLA 2 Department of Computational Medicine and Neurosurgery, UCLA

Abstract

Traumatic brain injury (TBI) is the leading cause of death and disability in people under 40. Current prognostic standards like CRASH rely solely on clinical variables (Glasgow Coma Scale and pupillary reactivity), often reporting high false-positive rates for unfavorable prognoses, increasing the risk for premature withdrawal of life-sustaining therapy.

Multimodal MRI captures detailed information about hemorrhaging, microstructural damage, and diffuse axonal injury that clinical variables miss entirely. Prior work has shown that foundation models can extract features from T1, T2, SWI, and ADC MRI and classify six-month functional outcomes.

However, a key gap for clinical adoption remains: it is unclear which multimodal signals drive model predictions. Therefore this project aims to yield interpretable, spatially localized features by applying sparse autoencoder methods to foundation model embeddings of multimodal TBI MRI data.

Introduction

What is Traumatic brain injury or TBI?

- characterized by a disruption of normal brain function due to an external force to the head
- common sources of TBI are sports, military, motor vehicle accident related
- 5.3 million people in the United States living with TBI-related disability (Ashkhan et al., 2025)

Current standard for clinical diagnosis of TBI is the Glasgow Coma Scale

- diagnoses TBI severity based on the patient's eye opening, verbal, and motor responses (Maas et al., 2022)
- pupillary reactivity is also a key diagnostic tool, as it is a sign of neurological health reflecting proper brainstem function
- age and demographic factors are also included in this scale as well as CT (Computed Tomography) scans that check for hemorrhaging

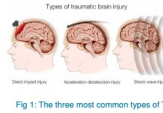


Fig 1: The three most common types of TBI

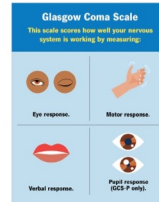


Fig 2: The main components of the Glasgow Coma Scale

CRASH and IMPACT are prognosis models that are frequently used to predict how a TBI will develop over time

- CRASH (Corticoid Randomization after Significant Head Injury)
- IMPACT (International Mission for Prognosis and Analysis of Clinical Trials in TBI)
- both developed to predict 6-month outcomes and rely only on clinical variables
- have limited predictive power as TBI is a fairly heterogeneous condition
- over-diagnosis of unfavorable outcomes can be detrimental to patient mental states and can lead to over monitoring of patients who do not require intensive care

One potential power source of information regarding TBI prognosis is patient brain MRI or magnetic resonance imaging scans

- MRI is a non-invasive method which has the capacity to capture multi-planar imaging, which provides a comprehensive anatomical view of the brain
- models including multimodal MRI features for other conditions such as multiple sclerosis and Alzheimer's significantly outperformed clinical covariates alone (Sriram et al., 2025)

Methods

Research Goal:

We apply sparse autoencoders to decompose multimodal MRI embeddings into spatially localized, interpretable features, identifying which signals drive outcome predictions and whether they align with known neuroradiological biomarkers of injury severity.

Dataset & Preprocessing

- 84 severe TBI patients, binary outcome (GOSe ≥ 5 favorable / < 5 unfavorable)
- 4 MRI modalities per patient (T1, T2, SWI, DWI/ADC), normalized and resized

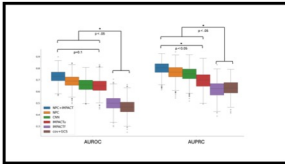


Fig 3: Comparison of AUROC (under the receiver operating characteristic) and AUPRC (precision-recall curve) across different models showing comparable performance to established TBI prognosis models (IMPACT) and IMPACT for non-parametric combination with IMPACT

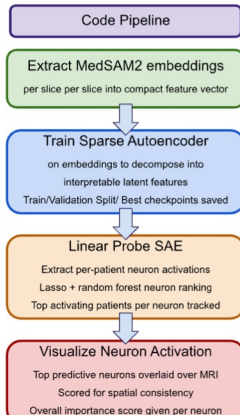


Fig 4: Code pipeline for sparse auto encoder training and embedding generation and utilization

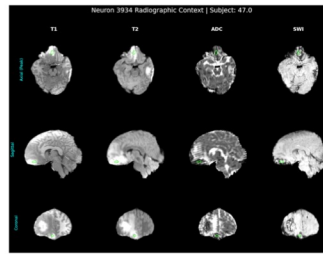


Fig 5: Image shows axial, coronal, sagittal views for 4 different modalities from the dataset with the area with peak activation of the neuron circled in green. One single patient is shown with the same slice in each image.

Neuron Importance Scoring

- Spatial consistency: mean pairwise voxel distance between a neuron's peak activation across top-5 subjects
- Importance score = (normalized l1weight $\times$ spatial consistency) weighted sum

Optimized prompt (diagnosis criteria) ... as an expert neuroradiologist, tasked with differentiating brain MRI scans as either Normal / Mild TBI or Moderate-to-Severe TBI / Pathological Injury Trauma ... The definitive diagnostic criterion for Moderate-to-Severe TBI is ... structural brain parenchymal disruption, extra-axial hemorrhage, or microvascular damage extending beyond superficial soft tissue ... Assess the following features, weighted by importance:	
1. Parenchymal Damage & Diffuse Axonal Injury (DAI): Is there ... evidence of punctate microhemorrhages, particularly at the gray-white matter junction, corpus callosum, or brainstem on SWI/T2* sequences? ... clear and consistent foci of restricted diffusion (DWI/ADC hyperintensity/hypointensity) indicating acute cytotoxic edema or axonal tearing ...	
2. Extra-Axial Hemorrhage & Fluid Collections: ... Is the subarachnoid, subdural, or epidural space clear? Be impartial - minor, isolated benign findings or minor artifacts can occur ... Structural Disruption: Presence of crescentic (subdural) or biconvex (epidural) hyperintensities, subarachnoid blood filling sulci, or significant extra-axial mass effect.	
3. Ventricles & Midline Shift: ... Are the lateral ventricles symmetric and age-appropriate in size, or is there compression? ... Normal Ventricles: Symmetric, well-defined, no effacement. Atypical/Pathological: Asymmetric compression, sulcal effacement, or midline shift due to edema or hemorrhage.	
4. Edema & Signal Intensity: ... Is the tissue signal uniform on T2/FLAIR, or is there surrounding vasogenic/cytotoxic edema (FLAIR hyperintensity)? Keep in mind that non-traumatic white matter hyperintensities (e.g., small vessel ischemic disease) can be present in older adults. Normal Parenchyma: Intact gray-white differentiation. Traumatic Edema: Geographic FLAIR hyperintensity adjacent to contusions or shear injuries.	
5. Brainstem & Herniation Signals: Are the basal cisterns patent, or is there evidence of uncal, tonsillar, or subfalcine herniation?	
Features Suggestive of Normal / Unremarkable Scans:	
• Preserved gray-white matter differentiation and intact structural architecture	• Symmetric, patent basal cisterns and ventricular system without midline shift.
• Absence of foci of susceptibility artifact (microhemorrhages) on SWI/GRE	• Absence of extra-axial blood collections or acute restricted diffusion

Fig 6: Prompt used with a LLM to predict TBI prognosis based on MRI embeddings, prompting the model to think as a neurologist and describing which features are identified and then use those features to make a prediction

LLM-Based Diagnostic Reasoning

- Structured prompt directs an LLM to evaluate scans/embeddings against defined radiological criteria (DAI, extra-axial hemorrhage, midline shift, edema, herniation) and predict prognosis
- Adapted from paper distinguishing normal tissue from invasive carcinoma with similar framework (Hong et al., 2026)

Results

51 Predictive Neurons Identified

- L1-logistic regression identified 51 SAE neurons with non-zero predictive weight for TBI outcome
- Each neuron mapped to its top-5 activating patients and visualized across orthogonal views
- 2 neurons discarded after visual review - activation regions were anatomically nonsensical (showed activation on negative space)

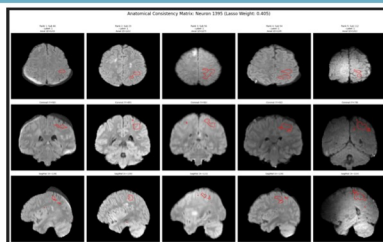


Figure 7: Most Predictive & Spatially Consistent Neuron. Neuron 1395 ranked highest overall (Lasso weight = 0.405, spatial consistency = 0.876). Consistent activation was localized to the same anatomical region across all 5 top-activating subjects. The neuron was visualized via anatomical consistency matrix (axial/coronal/sagittal overlays per subject).

Aggregate Importance Heatmap

- Top 10 neurons' peak activation locations were averaged and overlaid over an MRI
- Produced 8 slices per orthogonal view, showing where model-relevant features concentrate across the brain
- Reveals a consolidated, non-random spatial pattern of predictive activity rather than scattered noise

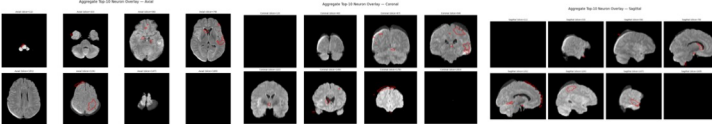


Figure 8: Three orthogonal views (axial, coronal, and sagittal) of MRI overlaid with the average voxel area of the top 10 neurons ranked by spatial similarity and importance

Discussion

Does this abstract, mathematically-derived SAE neuron correspond to something anatomically real and consistent?

Consistent localization across independent subjects suggests it is tracking a genuine anatomical/pathological structure rather than dataset noise (Fig 7)

- Circled areas correspond to a localized bruises, smashes, or bleeding on the brain tissue
- LLM identified focal traumatic parenchymal injury

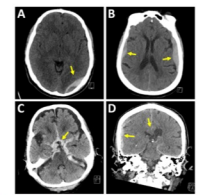


Figure 9: Examples of common lesions seen on MRI after TBI identified by radiologists (Douglas et al., 2018)

Conclusion

Predictive signal is not arbitrary

- a number of SAE-derived features are spatially consistent across patients and localize to regions flagged as pathological by radiological criteria
- suggests foundation-model-derived imaging features can serve as an interpretable, complement to black-box prediction
- necessary for clinical adoption, where opacity (not accuracy) is often the primary adoption barrier

Future Directions: Clinical Pipeline

- Proposed pipeline: generate embeddings $\rightarrow$ highlight top-activating features directly on MRI slices $\rightarrow$ feed annotated images into an LLM prompted to act as a radiologist
- LLM output: prognosis prediction, a confidence score, and a feature-by-feature breakdown explaining which highlighted regions contributed to the diagnosis
- Long-term goal: merge this interpretable, MRI-driven scoring approach into a new prognostic scale (similar in purpose to CRASH and IMPACT) but grounded in imaging-derived features rather than clinical variables alone

References

- Ashkhan, S. A. (2025). Linking the gap between traumatic brain injuries & cognitive impairments through the creation of machine learning-based diagnostic & prognostic clinical tools (Doctoral dissertation prospectus). University of California, Los Angeles.
- Douglas, D. B., Ro, T., Tabbal, T., Krawchuk, B., Malmgren, J., Gullu, J., Dalberger, A., Anderson, A. E., Douglas, P. K., & Wintermark, M. (2018). Neuroimaging of Traumatic Brain Injury. *Medical Sciences*, 7(1), 2.
- Hong, Y., Kuo, K.-C., Edwards, L. et al. Adaptive diagnostic reasoning framework for pathology with multimodal large language models. *Commun Med*, 6, 296 (2026). <https://doi.org/10.1038/s43864-026-01481-z>
- Maas, A. I. B., Alonso, D. K., Bontje, G. T., Arora, M., Alkandari, C., Jendek, N., Arns, M., Bultmann, T., Bell, M. J., Bodien, Y. G., Brett, B. L., Buki, A., Chonka, R. M., Cizero, G., Clark, D., Chesny, B., Cooper, D. J., Coles, E., Czoszynska, M., Dams-O'Connor, K., ... IITBI Participants and Investigators (2022). Traumatic brain injury: progress and challenges in prevention, clinical care, and research. *The Lancet*, *Neurology*, 21(1), 1054-1065. [https://doi.org/10.1016/S1473-3099\(20\)30200-9](https://doi.org/10.1016/S1473-3099(20)30200-9)
- Sriram, S., Nuvirthi, V., Anu Kaarthy, T. P., Archita, S., & Murugan, T. (2025). Advanced MRI based Alzheimer's diagnosis through ensemble learning techniques. *Scientific reports*, 15(1), 33840

Acknowledgements

We thank the UCLA Health Brain Injury Research Center (BIRC) for providing the BIRC cohort dataset.

Limitations

- Across subjects a given voxel coordinate may not correspond to the same anatomical location so spatial consistency quantifications can only be used as an estimate of structural similarity
- The current method does not account for left-right symmetry: the same functional feature could appear on opposite hemispheres across subjects and be scored as "inconsistent" when it may reflect the same underlying structure
- Peak activations for different neurons/subjects can occur at different axial slice depths, making it difficult to consolidate this information visually

Student: Robert Matthews

Faculty Advisor: Jake Lusic

Direct Mentor: Dylan Sarver

Department: Medicine/MIMG/Human Genetics

<https://lusic.genetics.ucla.edu/>

Under Pressure: Mapping the Remodeled Temporal Structure in Beige Adipose Reveals Diverging Responses to Nutritional Timing Restrictions

ROBERT MATHEWS¹, Dylan Sarver^{2,3,4}, Aldons J Lusic^{2,3,4}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Medicine, Division of Cardiology, UCLA

³ Department of Microbiology, Immunology, and Molecular Genetics, UCLA

⁴ Department of Human Genetics, UCLA

Nutritional timing is a potent regulator of circadian metabolism, yet how it reorganizes temporal transcription in adipose tissue remains poorly understood. Previous work demonstrated that restricting food access between two active-phase feeding bouts induces white adipose tissue beiging; increasing sympathetic signaling, mitochondrial biogenesis, and UCP1 expression. We sought to define how daily transcriptional programs are remodeled during this transition. Using circadian RNA sequencing from white adipose collected under ad-libitum and twice-a-night feeding, we developed an analytical pipeline which captures complex temporal expression patterns beyond conventional sinusoidal rhythm detection. We characterized temporally organized mitochondrial-associated transcripts, transcription factors, and secreted signaling molecules; clustering genes with shared expression dynamics and identifying coordinated pathways. This systems-level approach reveals remodeling of transcriptional networks accompanying adipose beiging, increasing our understanding of how nutritional timing reshapes rhythmic adipose physiology. This strategy offers a generalizable platform for extracting biologically meaningful temporal programs from time-series transcriptomic datasets.

**Robert
Matthews**

**Molecular,
Cellular, and
Developmental
Biology - UCLA**

**Expected year
of graduation
June 2026**

Under Pressure: Mapping the Remodeled Temporal Structure in Beige Adipose Reveals Diverging Responses to Nutritional Timing

UCLA

Robert Mathews¹, Dylan Sarver^{2,3,4}, Aldons J Lusis^{2,3,4}¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
²Department of Microbiology, Immunology, and Molecular Genetics, UCLA³Department of Medicine, Division of Cardiology, UCLA
⁴Department of Human Genetics, UCLA

Abstract

Nutritional timing is a potent regulator of circadian metabolism, yet how it reorganizes temporal transcription in adipose tissue remains poorly understood¹. Previous work demonstrated that restricting food access between two active-phase feeding bouts induces white adipose tissue beige; increasing sympathetic signaling, mitochondrial biogenesis, and UCP1 expression². We sought to define how daily transcriptional programs are remodeled during this transition. Using circadian-sampled bulk RNA sequencing from white adipose collected under ad-libitum (AL) and twice-a-night (TAN) feeding, we developed an analytical pipeline which captured complex temporal expression patterns beyond conventional sinusoidal rhythm detection. We characterized temporally organized mitochondrial-associated transcripts³, transcription factors, and secreted signaling molecules; clustering genes with shared expression dynamics and identifying coordinated pathways. This systems-level approach reveals remodeling of transcriptional networks accompanying adipose beigeing, increasing our understanding of how nutritional timing reshapes rhythmic adipose physiology. This strategy offers a generalizable platform for extracting biologically meaningful temporal programs from time-series transcriptomic datasets.

¹Hatori et al., 2012. doi.org/10.1016/j.cmet.2012.04.019
²Mattar et al., 2024. doi.org/10.1016/j.celrep.2024.114390
³Rath et al., 2021. doi: 10.1093/nar/gkaa1011

Fig. 1: Filtered Clusters Reveal Restructured Temporal Architecture

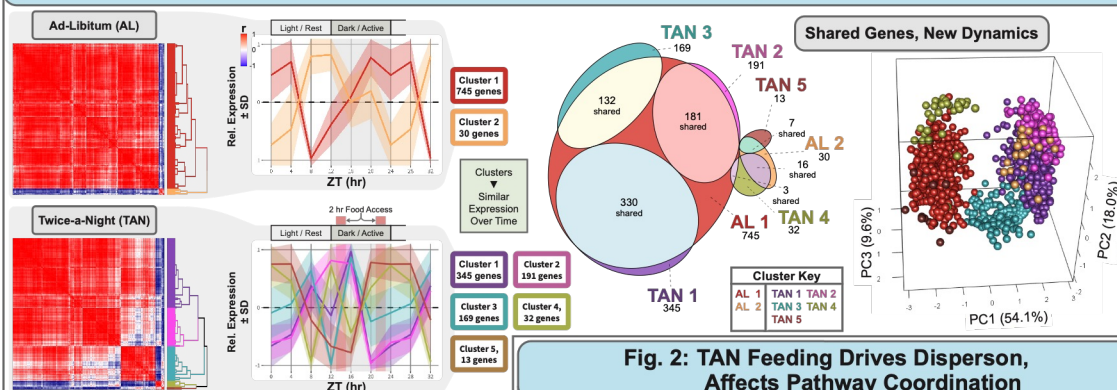


Fig. 2: TAN Feeding Drives Dispersion, Affects Pathway Coordination

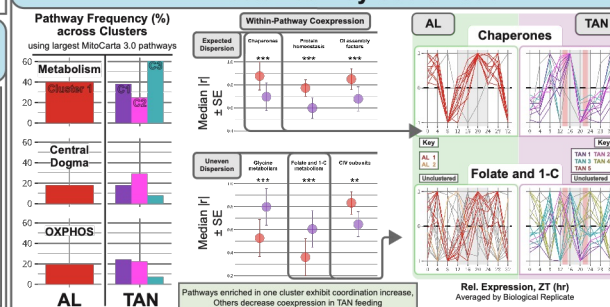
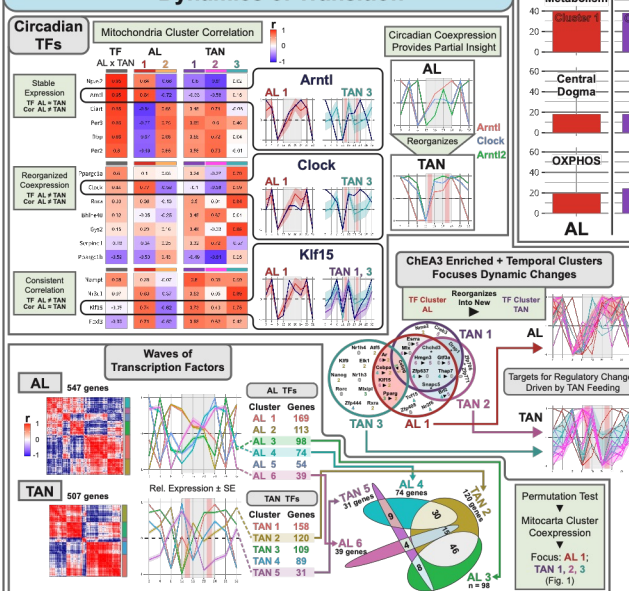


Fig. 3: Transcription Factors Expand Dynamics of Transition

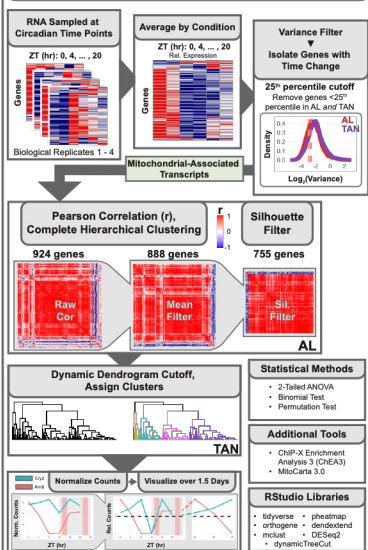


Conclusions

- Adipose of TAN fed mice exhibits diverging reactions to nutritional signal
- Reorganized adipose appears through shared gene clusters
- New cluster organization presents different temporal architecture
- MitoCarta pathways enriched in one cluster have higher coordination across time
- Circadian genes present some reorganization, but do not fully capture the TAN driven changes
- Notably, the core interactions of Clock-Arntl and Clock-Arntl2 are distinct between AL and TAN groups
- Modeling clusters of transcription factors adds insights into core signaling changes
- ChEA3 results incorporate known enrichment with our temporal changes

The integration of temporal coexpression, transcriptional waves, and gene expression provides a nuanced view into the TAN adipose and enables further investigation into the mechanisms underpinning the response to nutritional timing. The use of this strategy and developed tools may enable further investigation into other perturbation studies, the goal of which is the accumulation of temporal motifs to build a more comprehensive understanding of nutritional timing, cellular responses, and organism adaptation.

Methodology of Approach



Student: Sahel Munshi

Faculty Advisor: Alexander Hoffmann

Direct Mentor: Vyas Koduvayur

Department: Microbiology, Immunology, and
Molecular Genetics (MIMG)

<https://www.signalingsystems.ucla.edu/alexander-hoffmann/>

Sahel Munshi

**High School
Student**

Systematic Transcriptional Profiling Reveals Stimulus-Specific Programs in Human Macrophage Training

SAHEL MUNSHI¹, Vyas Koduvayur^{1,2}, Noah Yan³, and Alexander Hoffmann²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Microbiology, Immunology, and Molecular Genetics, UCLA

³ Department of Medicine, Division of Digestive Diseases, UCLA
Macrophages are innate immune cells that adapt to different tissue contexts by a process termed innate immune training, thereby affecting their subsequent immune responses. Although some macrophage states have been extensively studied, comparative transcriptional profiling across a broad panel of training stimuli has not been systematically performed. Here, we used bulk RNA-seq to profile primary human macrophages from seven donors exposed to eleven immunological stimuli to characterize stimulus-specific transcriptional programs. Following quality control, TMM normalization, and filtering, principal component analysis identified stimulus type as the primary source of transcriptional variation, with interferon-type stimuli (IFN γ , IFN β , PolyIC) separating from other conditions along PC1. Relative to naive macrophages, IFN γ and IFN β induced the largest numbers of differentially expressed genes. Clustering revealed distinct, stimulus-specific expression patterns, and Gene Ontology analysis identified the biological processes underlying these stimulus-specific differences. This dataset provides a transcriptional resource for understanding human macrophage training across diverse immune stimuli.



Systematic transcriptional profiling reveals stimulus-specific programs in human macrophage training

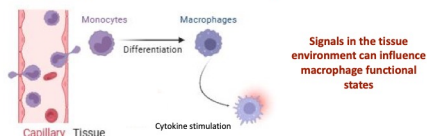
UCLA

SAHEL MUNSHI¹, Vyas Koduvayur^{1, 2}, Noah Yan³, and Alexander Hoffmann²¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA, ²Department of Microbiology, Immunology, and Molecular Genetics, UCLA, ³Department of Medicine, Division of Digestive Diseases, UCLA

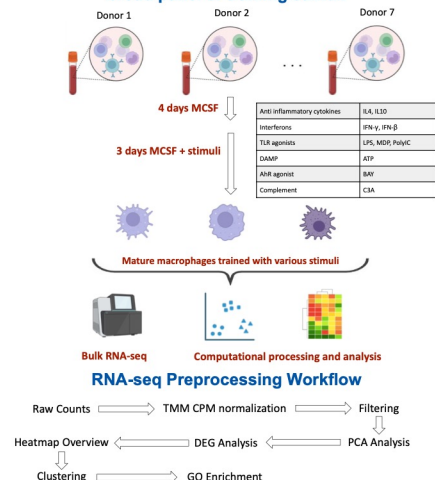
Abstract

- Macrophages are innate immune cells that adapt to different tissue contexts by a process termed innate immune training, thereby affecting their subsequent immune responses. Although some macrophage states have been extensively studied, comparative transcriptional profiling across a broad panel of training stimuli has not been systematically performed.
- Here, we used bulk RNA-seq to profile primary human macrophages from seven donors exposed to eleven immunological stimuli to characterize stimulus-specific transcriptional programs.
- Following quality control, TMM normalization, and filtering, principal component analysis identified stimulus type as the primary source of transcriptional variation, with interferon-type stimuli (IFN- γ , IFN- β , PolyIC) separating from other conditions along PC1.
- Relative to naive macrophages, IFN- γ and IFN- β induced the largest numbers of differentially expressed genes. Clustering revealed distinct, stimulus-specific expression patterns, and Gene Ontology analysis identified the biological processes underlying these stimulus-specific differences.
- This dataset provides a transcriptional resource for understanding human macrophage training across diverse immune stimuli.

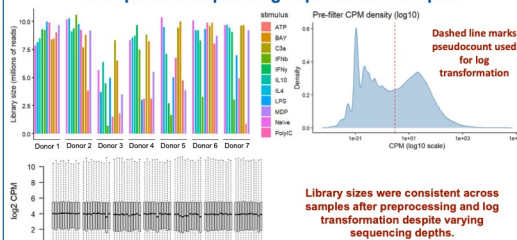
Macrophages can be trained in response to stimulus



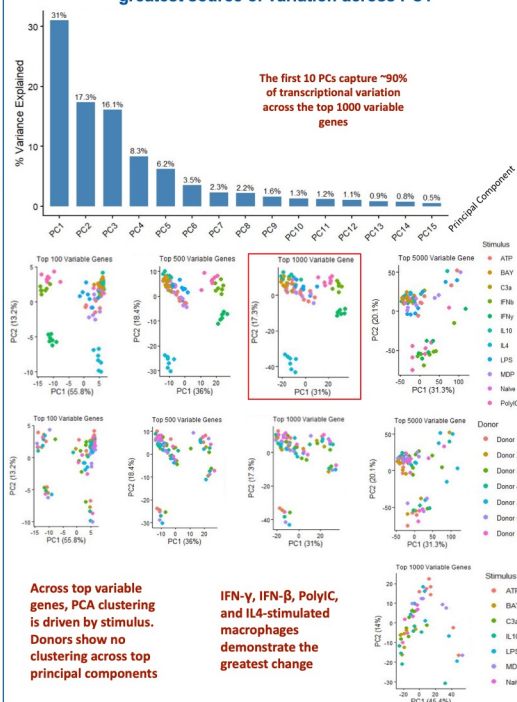
Monocyte-derived macrophages were profiled across a broad panel of training stimuli



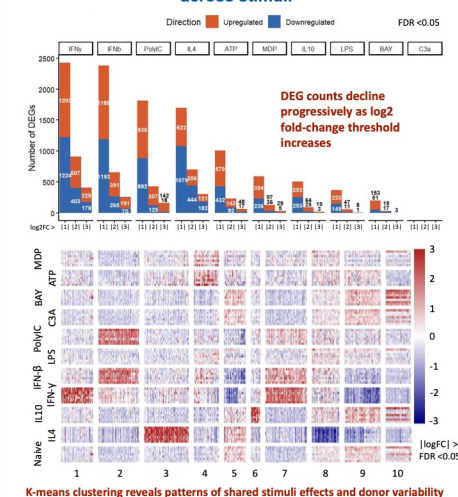
Quality control and normalization ensure comparable sequencing depth across samples



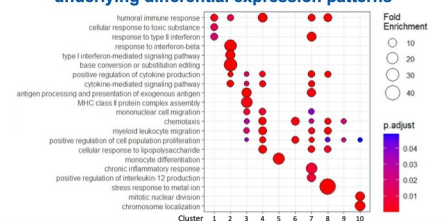
Principal component analysis identifies stimulus as the greatest source of variation across PC1



Distinct patterns of differential gene expression emerge across stimuli



Gene Ontology reveals enriched biological pathways underlying differential expression patterns

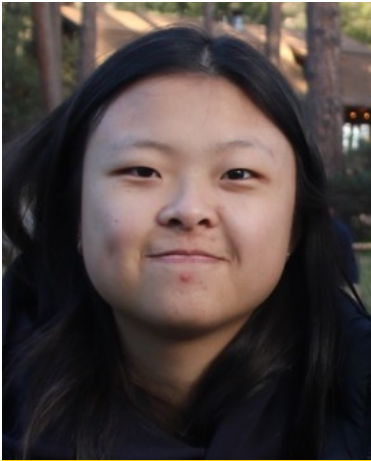


Conclusions

- Our *in vitro* system models macrophage exposure to diverse training stimuli encountered *in vivo*, allowing for analysis of stimulus-specific transcriptional programs.
- PCA identifies stimulus, not donor, as the dominant axis of variation; interferon-associated stimuli (IFN- γ , IFN- β , PolyIC) and IL-4 show greatest changes.
- K-means clustering reveals distinct gene modules shared across stimuli; GO enrichment links each module specific biological processes underlying stimulus-specific response.
- This dataset serves as a resource to match specific tissue/immune contexts to transcriptional responses.
- Next steps:
 - Validation of key DEGs to explore roles in macrophage training phenotypes
 - Further analysis into specific gene clusters to identify pathway drivers and key regulators

Acknowledgements

We acknowledge the members of the Signaling Systems Lab for their guidance and support. We also thank the QCBio community and the BIG Summer Program for providing us with their resources and the opportunity to present this work.



Student: Lois Pak
Faculty Advisor: Hooman Allayee
Department: Medicine
<https://profiles.ucla.edu/hooman.allayee>

Lois Pak

**Computational
and Systems
Biology - UCLA**

**Expected year
of graduation
June 2027**

Genetic determinants of circulating metabolites associated with cardiometabolic diseases

LOIS PAK¹, Rajat Gupta¹, James R. Hilser¹, Zeneng Wang², Joseph A. DiDonato², Jaana A. Hartiala¹, Aldons J. Lusis¹, Stanley L. Hazen², Hooman Allayee

¹ Department of Medicine, David Geffen School of Medicine of UCLA, Los Angeles, CA 90095;

² Department of Heart, Blood, and Kidney Research, Cleveland Clinic, Cleveland, OH 44195

Coronary artery disease (CAD), myocardial infarction (MI), and type 2 diabetes (T2D), account for significant morbidity and mortality. Each has a multifactorial etiology largely influenced by heritable factors in the context of environmental factors. Despite this recognition, the hundreds of susceptibility loci identified for CAD, MI, and T2D collectively explain only a fraction of their heritabilities. Previous studies have also identified several metabolites associated with cardiometabolic diseases, including indolepropionate (IPA) and phenylacetylglutamine (PAG), which are derived from gut bacteria-mediated metabolism of dietary tryptophan and phenylalanine, respectively. For example, higher and lower levels of IPA and PAG are associated with decreased and increased risk of T2D and CAD/MI, respectively. However, the mechanisms underlying these associations are not understood. Here, we performed a large-scale, multi-cohort genome-wide association study (GWAS) meta-analysis to identify novel loci associated with levels of each metabolite, followed by functional validation of positional candidate genes in mouse models.

Genetic determinants of circulating metabolites associated with cardiometabolic diseases

LOIS PAK^{1,2}, Rajat Gupta², James R. Hilser², Zeneng Wang³, Joseph A. DiDonato³, Jaana A. Hartiala², Aldons J. Lusis², Stanley L. Hazen³, Hooman Allayee²

¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA; ²Department of Medicine, David Geffen School of Medicine of UCLA, Los Angeles, CA 90095
³Department of Heart, Blood, and Kidney Research, Cleveland Clinic, Cleveland, OH 44195

Abstract

Coronary artery disease (CAD), myocardial infarction (MI), and type 2 diabetes (T2D) account for significant morbidity and mortality. Each has a multifactorial etiology largely influenced by heritable factors in the context of environmental factors. Despite this recognition, the hundreds of susceptibility loci identified for CAD, MI, and T2D collectively explain only a fraction of their heritabilities. Previous studies have also identified several metabolites associated with cardiometabolic diseases, including indolepropionate (IPA) and phenylacetylglutamine (PAG), which are derived from gut bacteria-mediated metabolism of dietary tryptophan and phenylalanine, respectively. For example, higher and lower levels of IPA and PAG are associated with decreased and increased risk of T2D and CAD/MI, respectively. However, the mechanisms underlying these associations are not understood. Here, we performed a large-scale, multi-cohort genome-wide association study (GWAS) meta-analysis to identify novel loci associated with levels of each metabolite, followed by functional validation of positional candidate genes in mouse models.

Methods

GWAS summary statistics were collected for **IPA (14 cohorts)**, **PAG (13 cohorts)**, and **PAA (13 cohorts)**, with each metabolite analyzed separately.

Variant IDs, positions, alleles, association statistics, sample sizes, and genome builds were standardized, with duplicate and invalid variants removed.

Separate Z-score meta-analyses were performed using **METAL**, with genome-wide significance defined as $P < 5 \times 10^{-8}$.

Lead variants were identified and independent signals were evaluated using genomic distance and LD ($r^2 < 0.001$).

Lead SNPs for **IPA and PAA** were compared across both GWAS to determine whether associations were shared or metabolite-specific.

Regional plots were generated around selected lead variants to evaluate LD patterns and prioritize positional candidate genes.

Selected candidate genes were evaluated in mouse models for their effects on circulating metabolite levels.

Manhattan plots

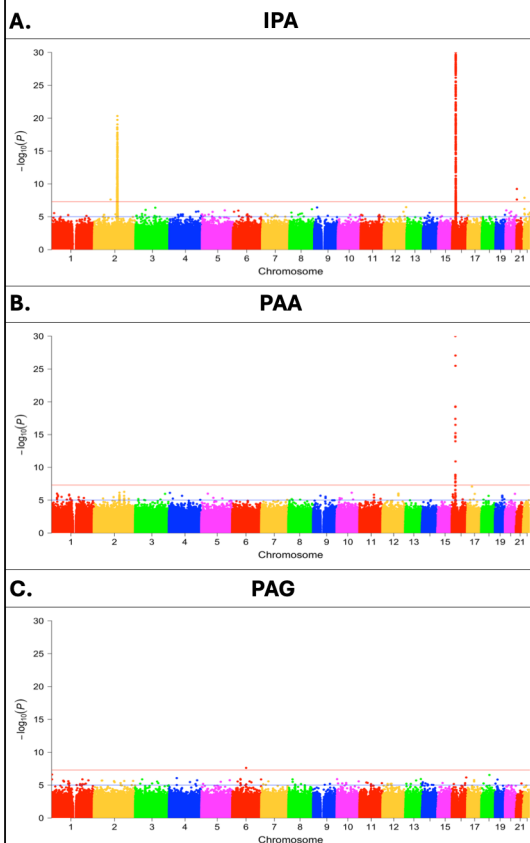


Figure 1. Manhattan plot of results from genome-wide association study meta-analyses for IPA, PAA, and PAG. (A) One novel locus on chromosome 16 was significantly associated with IPA where the lead SNP was rs4603548 ($p=6.16 \times 10^{-9}$). (B) The same novel locus on chromosome 16 was also significantly associated with PAA but with a different lead SNP (rs139233971; $p=5.72 \times 10^{-8}$). (C) No significantly associated locus was identified for PAG. Genome-wide thresholds for significant ($p \leq 5 \times 10^{-8}$) and suggestive ($p \leq 5 \times 10^{-5}$) associations are indicated by the horizontal orange and blue lines, respectively. P-values are truncated at $-\log_{10}(p) = 30$.

Regional plots for chr 16

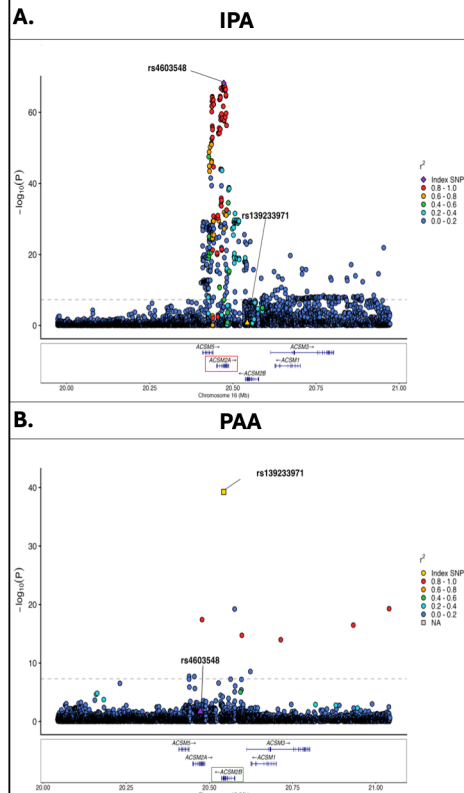


Figure 2. Regional association plots of the chromosome 16 locus associated with IPA and PAA. (A) The lead SNP identified for IPA on chromosome 16 is rs4603548 (highlighted in purple) and localizes to intron 6 of *ACSM2A* (boxed in red). (B) The lead SNP for PAA on chromosome 16 is different (rs139233971, highlighted in yellow) and maps to intron 11 of *ACSM2B* (boxed in green). Variants are plotted according to their genomic location at the chromosome 16 locus and the strength of their association with each respective metabolite, represented as $-\log_{10}(P)$. The degree of linkage disequilibrium between each lead SNP and nearby variants is represented by r^2 values according to the color-coded legend. Nearby genes within each locus are displayed in the lower panel to provide genomic context and highlight candidate genes located near the associated signals.

Cohort Demographics

Study	N	Ancestry
Canadian Longitudinal Study on Aging (CLSA)	60	African
Canadian Longitudinal Study on Aging (CLSA)	104	East Asian
Canadian Longitudinal Study on Aging (CLSA)	108	South Asian
Canadian Longitudinal Study on Aging (CLSA)	8,299	European
Interval-Epic	19,994	European
Gene Bank	3,188	European
TOPMed	10,142	Mixed Ancestry
Knight-Alzheimer Disease Research Center (ADRC)	417	African American
Knight-Alzheimer Disease Research Center (ADRC)	2,392	European
Metabolic Syndrome in Men (METSIM)	6,076	European
Schlosser	4,960	European
Hispanic Community Health Study/ Study of Latinos (SOL)	3,926	Hispanic
TwinsUK/KORA	7,358	European
Tohoku Medical Megabank Japan	1,008	Japanese
Total	68,032	

Functional Validation of *Acsm2* in mice

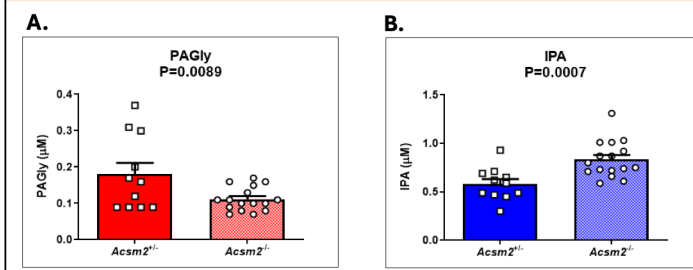


Figure 3. Plasma levels of IPA and PAG in *Acsm2* deficient mice. IPA and PAG levels were measured in plasma by mass spectrometry in heterozygous ($n=11$) and homozygous ($n=16$) knockout mice.

References

1. Chen, Y. et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nature Genetics* 55, 44–50 (2023).
2. Schlesselman, P. et al. Genetic studies of paired metabolites reveal enzymatic and transport processes at the interface of plasma and urine. *Nature Genetics* 55, 995–1008 (2023).
3. Yang, G. et al. European and African ancestry-specific plasma protein-QTL and metabolite-QTL analyses identify ancestry-specific T2D effector proteins and metabolites. *Nature Communications* 16, 7412 (2025).
4. Sureshbabu, P. et al. Rare and common genetic determinants of metabolic individuality and their effects on human health. *Nature Medicine* 28, 2321–2332 (2022).
5. Fedotova, E. V. et al. A Genome-wide Association Study Discovers 46 Loci of the Human Metabolome in the Hispanic Community Health Study/Study of Latinos. *American Journal of Human Genetics* 107, 849–863 (2020). [EOL]
6. Yin, X. et al. Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. *Nature Communications* 13, 1644 (2022).
7. Akiyama, S. et al. The Tohoku Medical Megabank Project: Design and Mission. *Journal of Epidemiology* 28, 493–511 (2018).
8. Wang, N. et al. Genetic architecture and analysis practices of circulating metabolites in the NHLBI Trans-Omics for Precision Medicine Program. *American Journal of Human Genetics* 112, 2720–2738 (2023).
9. Shin, S.-K. et al. An atlas of genetic influences on human blood metabolites. *Nature Genetics* 48, 543–550 (2016).

Results/Conclusions

1. Large-scale genetic analysis identifies the same locus on chromosome 16 harboring multiple members of the *ACSM* family of enzymes as being significantly associated with circulating levels of both IPA and PAA.
2. No locus was identified for PAG.
3. The lead variant for IPA pointed to *ACSM2A* as the causal gene whereas the lead variant for PAA pointed to *ACSM2B*.
4. *ACSM2* was functionally validated to affect levels of both metabolites since mice only have one isoform of the enzyme.
5. The dual role of *ACSM2* in catabolizing IPA and generating phenylacetyl CoA from PAA has diverged in the human lineage to *ACSM2A* (IPA) and *ACSM2B* (PAA).

Student: Jaime Pelayo

Faculty Advisor: Alexander Hoffmann

Direct Mentor: Jennifer Chia

Department: Microbiology, Immunology, and
Molecular Genetics (MIMG)

<https://www.signalingsystems.ucla.edu/alexander-hoffmann/>

Capturing hematopoietic development using trajectory analysis of immunophenotypic data.

JAIME PELAYO^{1, 2}, Alexander Hoffmann^{1, 3, 4, 5}, Jennifer Chia^{3, 4, 5, 6}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Integrative Biology and Physiology, UCLA

³ Department of Microbiology, Immunology, and Molecular Genetics, UCLA

⁴ Jonsson Comprehensive Cancer Center, UCLA

⁵ Broad Stem Cell Research Center, UCLA

⁶ Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA

Hematopoietic development has been modeled as either discrete, continuous, or as a continuum containing specific stable intermediate states. Although cell immunophenotyping is the gold standard for identifying distinct hematopoietic populations, conventional approaches poorly capture continuous development. Advances in high-dimensional immunophenotyping now allow us to test whether surface-protein data can resolve this continuum. Here, we calculate Euclidean distances between cell clusters to identify optimal marker sets for resolving populations, then use pseudotime trajectory analysis to order hematopoietic cells along the differentiation continuum. We quantify monocyte progenitor cell stability using silhouette analysis to ask whether discrete stable states exist within the developmental continuum. We found that immunophenotypic data can capture hematopoietic trajectories at a continuous resolution without full-transcriptome measurements. Further, pseudotime resolution depends more on marker specificity to cell types than on the total number of parameters. Understanding hematopoiesis as a discretized continuum may improve detection of abnormal developmental states in disease.

Jaime Pelayo

**Computational
and Systems
Biology,
Physiological
Science -
UCLA**

**Expected year
of graduation
June 2027**

Capturing hematopoietic development using trajectory analysis of surface immunophenotypes

Jaime Pelayo^{1, 2}, Alexander Hoffmann^{1, 3, 4, 5}, Jennifer Chia^{4, 5, 6, 7}

¹Institute for Quantitative and Computational Biosciences, UCLA

²Department of Integrative Biology and Physiology, UCLA³Department of Microbiology, Immunology, and Molecular Genetics, UCLA⁴Jonsson Comprehensive Cancer Center, UCLA

⁵Broad Stem Cell Research Center, UCLA

⁶Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA

⁷Molecular Biology Institute, UCLA



BACKGROUND & SIGNIFICANCE

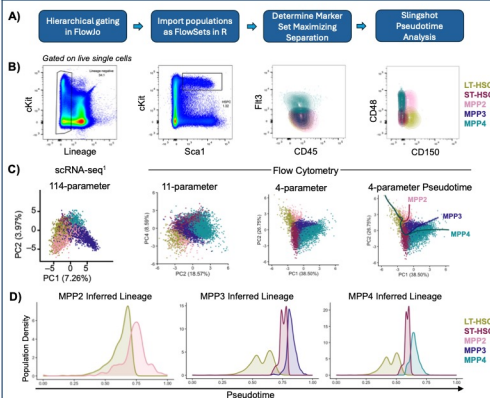
- Hematopoiesis begins with rare HSCs to produce all blood cells.¹
 - Hematopoiesis has been modeled as **either a discrete or continuous process**.²
- The diagram illustrates two models of hematopoiesis. The **Discrete Model** shows a hierarchical lineage starting from a single **LTHSC** (Lymphoid-Tissue Hematopoietic Stem Cell), which differentiates into various progenitor populations: **MPC** (Myeloid Progenitor Cell), **CMP** (Common Myeloid Progenitor Cell), **CLP** (Common Lymphoid Progenitor Cell), **GMP** (Granulocyte-Macrophage Progenitor Cell), **MEP** (Myeloid Erythroid Progenitor Cell), **CDP** (Common Dendritic Progenitor Cell), and **GEP** (Granulocyte Erythroid Progenitor Cell). These progenitors further differentiate into mature blood cells: **Erythrocyte**, **Megakaryocyte**, **Platelet**, **Neutrophil**, **Monocyte**, **Macrophage**, **T cell**, **B cell**, **DC** (Dendritic Cell), and **NK cell** (Natural Killer cell). The **Continuous Model** shows a single population of **Hematopoietic Progenitor Cells** that directly differentiate into the same mature blood cell types.
- Transcriptomic data has been used for trajectory analysis of hematopoietic development, but it **does not well capture the cell-surface phenotype**.
- The diagram illustrates the limitations of transcriptomic data for capturing cell-surface phenotype. It shows a single cell with a transcriptome (represented by a cloud of dots) and a cell-surface phenotype (represented by a cloud of dots). The transcriptome is shown to be insufficient for capturing the full range of cell-surface phenotypes.
- Cell Immunophenotyping is the gold standard** for identifying hematopoietic populations; however, it has been limited by the number of markers and thus has **not been effective at capturing the hematopoietic continuum**.
 - Cellular Indexing of Transcriptome and Epitopes by Sequencing (CITE-seq) and spectral flow cytometry have increased the dimensionality of immunophenotypic data, **enabling pseudotime trajectory inference to capture continuous development at a single-cell resolution**.³

RESEARCH QUESTIONS

Can immunophenotype-based trajectory analysis capture the continuum of hematopoietic development?

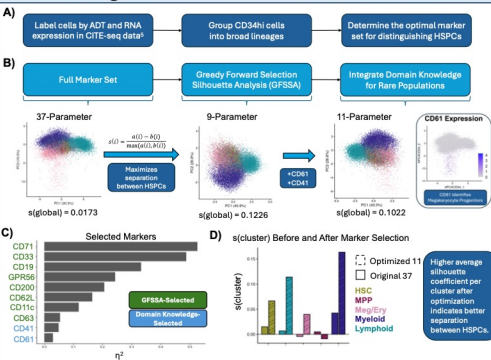
- Can this approach resolve relatively homogeneous stem and multipotent progenitor populations by their lineage biases?
- Can we resolve heterogeneous intermediates along the hierarchy from stem cells to mature cells?

Figure 1: Few markers can resolve highly similar multipotent cell states with lineage biases



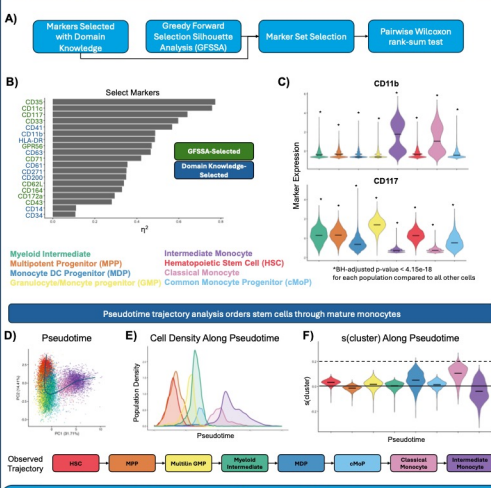
A) Workflow; pseudotime performed with Slingshot.³ **B**) Gating strategy for murine stem cell and multipotent progenitor populations (HSPCs).¹ **C**) Principal component analysis (PCA) of HSPCs using different data types and marker sets. Gold standard separation with scRNA-seq data (left); new analysis of flow cytometry data (right), with distinct trajectory branches for megakaryocyte/erythroid-biased MPP2s, myeloid-biased MPP3s, and lymphoid-biased MPP4s. **D**) Pseudotime for cells with >80% probability of residing along each branch.

Figure 2: An unbiased approach aids marker selection in higher dimensional human HSPC data



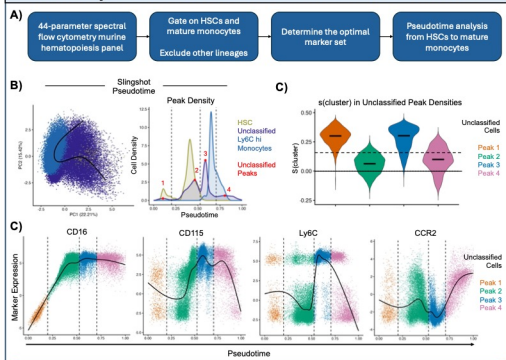
A Workflow schematic using pre-clustered human CITE-seq data. **B** Silhouette Analysis used to measure how well a cell fits into its assigned cluster, and is distinct from nearby clusters. **C** Greedy Forward Selection Analysis (GFSSA) performed to select markers that cumulatively increase the global average silhouette coefficient, but can miss rare populations. Addition of domain knowledge allows retention of these cell types. **D** η^2 values from Analysis of Variance (ANOVA) to measure how much variance between clusters is explained by each marker. **E** Mean silhouette coefficient for each HSPC population, compared between the full 37-parameter set and optimized 10-parameter markers.

Figure 3: Trajectory analysis of human data can resolve hematopoiesis from HSCs through mature monocytes



A) Workflow schematic. **B)** ANOVA η^2 for markers for PCA selected by GFSSA and domain knowledge. **C)** Expression of representative marker across all populations; statistical calculation with Wilcoxon rank-sum test between a population and all other cells. **D)** Singhopt pseudotime; start cluster set as HSCs and end cluster set as intermediate monocytes. **E)** Population density across pseudotime; observed order of intermediates matches that expected from the literature.⁶ **F)** Silhouette analysis of populations annotated by a combination of CITE-seq and scRNA-seq⁴ shows that many intermediate progenitors are not distinct from one another.

Figure 4: Immunophenotypic trajectories resolve hematopoietic intermediates without classification



A) Workflow schematic. HSCs Lineage-*c/kit*-*Sca1*-*Flt3*-*CD40-CD151*o, megakaryocytes (Lineage-*CD11b*-*CD115*-*Ly6c*-*VSSC*)-, unclassified intermediate cells are negative for lymphoid, erythroid, megakaryocyte, granulocyte, and DC lineage markers. **B)** HSCs are set as the root and monocytes as the end cluster for pseudotime analysis. Peak densities within the unclassified cells may represent distinct progenitor populations. **C)** Silhouette analysis of peak densities in unclassified cells shows that many intermediate progenitors are not distinct from one another. **D)** Evaluation of marker expression in unclassified peaks to compare with known progenitor phenotypes.

CONCLUSIONS

- Immunophenotypic trajectory analysis captures continuous hematopoietic development in stem and multipotent progenitor populations.
 - Even closely related HSPC states can be resolved using an optimized marker set containing informative parameters.
 - Resolution depends more on marker specificity than on panel dimensionality.
 - Combining computational and domain knowledge improves marker selection.
- Immunophenotypic data can resolve the full monocyte lineage from HSCs through mature cell with continuous development and few abundant phenotypic states.
 - Silhouette analysis suggests gated progenitor populations are relatively indistinct.
 - Murine data with unlabeled progenitor populations contains densities along pseudotime that may correspond to previously defined progenitors.

FUTURE DIRECTIONS

- Future work will resolve intermediate cell state further, and extend to other lineages
- This will allow us to create a map of the hematopoietic continuum from stem cells through mature cells based on surface immunophenotypes.
- We will use this map to profile the bone marrow of murine inflammatory models and humans with hematopoietic disease to determine how specific inflammatory networks impact hematopoiesis in health and disease.

ACKNOWLEDGEMENTS

We acknowledge support from the **Whitcome Summer Undergraduate Research Fellowship** and the **Program for Excellence in Education and Research in the Sciences** (PEERS). For support and helpful discussions, thank you to Noah Yan, Kai Silkwood, Vyas Koduvayur, Jansen Samaniego, and members of the **Chia Laboratory and Signaling Systems Laboratory**. Thank you to Caroline Baron for organizing BIG Summer.

REFERENCES

- ¹Pietras et al. (2015). *Cell Stem Cell*, 17(1), 35–46.
²Zhang et al. (2024). *Nature Immunology*, 25, 703–715.
³Velten et al. (2017). *Nature Cell Biology*, 19, 271–281.
⁴Guilliams et al. (2018). *Immunity*, 49(4), 595–613.
⁵Melsen et al. (2020). *Immunology*, 205(3), 864–871.
⁶Singh et al. (2025). *Blood*, 145(12), 1293–1308.
⁷Chia et al. (2025). *Cell Reports*, 44(9), 116193.

CONTACTS

JChia@mednet.ucla.edu jdpelayo@g.ucla.edu

Student: Madeline Shono

Faculty Advisor: Matteo Pellegrini

Department: Molecular, Cell, and Developmental Biology (MCDB)

<https://www.pellegrini.mcdb.ucla.edu/>

Age and Weight Associated Changes in the Canine Oral Microbiome

MADELINE SHONO¹, Zhiyue Jiang², Aisling Archdeacon³, Matteo Pellegrini³

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Computational and Systems Biology Interdepartmental Program, UCLA

³ Department of Molecular, Cell and Developmental Biology, UCLA

Periodontal disease (PD) is one of the most common oral health concerns in dogs, associated with shifts in the microbiome community and an increase in disease-correlated species. PD often goes untreated until it has progressed significantly and symptoms are visible. Age and weight are known to contribute to the risk of periodontal disease, however, large-scale characterizations of how these host factors relate to the canine oral microbiome remain limited. We ask if and how the canine oral microbiome changes with age and weight, and if there are health implications. Buccal swabs from 1,125 dogs were whole genome shotgun sequenced, enabling taxonomic profiling using MetaPhlAn. Findings showed an age-associated restructuring in the oral microbiome with a progression towards anaerobic, disease-associated taxa. As weight increased, disease-related taxa decreased and commensal taxa increased, but associations were weaker compared to age. This work could improve early detection and diagnosis of periodontal disease.

**Madeline
Shono**

**Computational
and Systems
Biology - UCLA**

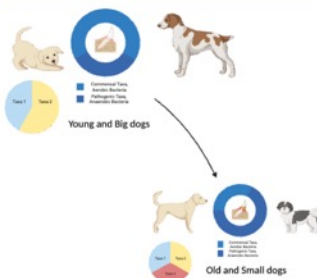
**Expected year
of graduation
June 2026**

Age and Weight Associated Changes in the Canine Oral Microbiome

MADLINE SHONO¹, Zhiyue Jiang², Aisling Archdeacon³, Matteo Pellegrini³

BIG Summer Program, Institute for Quantitative and Computational Biosciences¹, UCLA; Computational and Systems Biology Interdepartmental Program, UCLA²; Department of Molecular, Cell and Developmental Biology, UCLA³

Abstract



Overall, younger dogs and larger dogs are at less risk for oral diseases, such as periodontal disease, based on their oral microbiome. The shift towards an oral microbiome dominated with anaerobic bacteria is correlated with an increased risk of disease.

Introduction

Oral Microbiome: The oral microbiome changes with age and weight (body size) – specifically in the composition and abundance of different bacteria.

Bacteria: Bacteria can be classified by oxygen preference (aerobic, anaerobic or facultative) or by correlation with disease (disease correlated or commensal). Specific bacteria such as *Conchiformibius* (under the order Neisseriales) and *Porphyromonas* (in the order Bacteroidales) have been widely classified as commensal and oral disease correlated, respectively.

Periodontal Disease (PD): PD is a prevalent oral disease in dogs, in which age and weight/body size are contributing factors. Periodontal disease is associated with changes in the canine/dog oral microbiome. Specifically, healthy dogs have oral microbiomes dominated by aerobic taxa and an increase in anaerobic taxa in dogs with periodontal disease.

Question: How does the dog oral microbiome change with age? Based on these changes, can we make any inferences about a dog's health?



Results

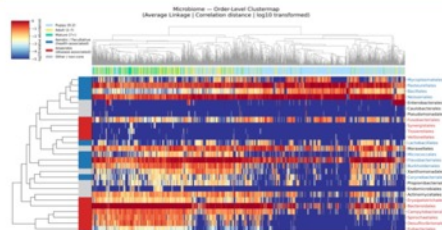


Figure 1: Order level cluster map of log10 transformed relative abundances. Samples (dogs) are columns ($n = 1,125$) and rows are order level abundances. Order names are colored based on the bacteria's oxygen tolerance (aerobic, anaerobic or other.) The samples are annotated based on age group. Clustering was done using average linkage (correlation distance.)

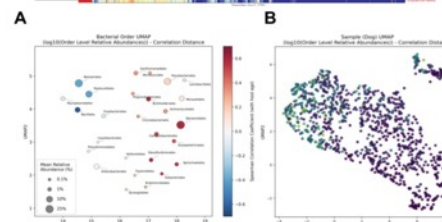


Figure 2A: UMAP of bacterial orders. Sized by log10 transformed mean relative abundances and colored by Spearman correlation with dog age. **Figure 2B: UMAP of samples' (dogs') log10 transformed order level relative abundance profiles, colored by age.** Both UMAPs use correlation distance metric.

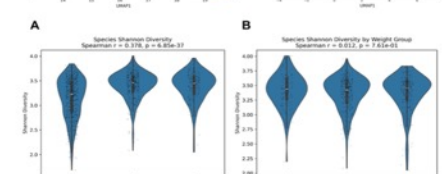


Figure 3A: Species level Shannon diversity values of oral microbiomes grouped by age with corresponding Spearman r value. Ages were grouped according to the following: 0-2, 2-7 and 7+ years old. **Figure 3B:**

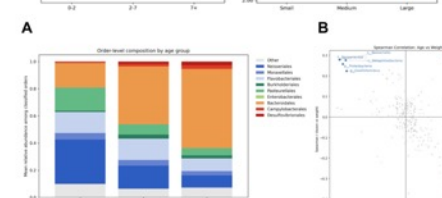


Figure 3B: Species level Shannon diversity values of oral microbiomes, grouped by weight with corresponding Spearman correlation value. The weights were grouped according to the following: 0-20, 20-60 and 60+ lbs.



Figure 4A: Stacked bar plot of mean relative order level abundances across various age groups (0-2, 2-7 and 7+ years old.) Bars are colored according to oxygen preference of the bacteria (green = facultative, red/orange = anaerobic and blue = aerobic.)

Figure 4B: Spearman correlation values of relative abundance of taxa at all levels with dog age (x-axis) and weight (y-axis.)

Methods

Sample Collection: Randomly collected buccal swabs ($n = 1,125$). We used whole genome shotgun sequencing on the samples. Ages and weights of dogs were self reported by the owners.

Data Processing: We ran each sample through the MetaPhlAn software to conduct taxonomic profiling for each dog's oral microbiome. In other words, abundances of bacteria at different taxonomic levels are calculated from the DNA sequences.

Data Cleaning: We normalized each bacteria's abundance to be a fraction of the total bacteria abundance (1.) We then log10 transformed the data when reducing the effect of outliers. Orders that were classified by MetaPhlAn but not formally named were removed from the analysis.

Diversity Analysis: Spearman correlation coefficient was used to calculate correlation between diversity and age. Shannon Diversity (H) was calculated using the following formula:

$$H = - \sum_{i=1}^S (p_i \ln p_i)$$

Conclusion

The dog oral microbiome changes with age. Younger dogs tend to be dominated with aerobic, commensal taxa. Older dogs tend to be dominated with anaerobic, disease correlated taxa. Changes in the oral microbiome with weight are weaker. However, there was a handful of taxa across all taxonomic levels that were correlated positively with age and negatively with weight and vice versa. UMAP analysis revealed orders tended to divide based on their correlation with age.

Diversity shows a significant species level shift as dogs age. Younger dogs usually have a wider variety in diversity while older dogs tend to have more diverse oral microbiomes. The correlation between diversity and dog weight is insignificant.

Future directions include getting clinical data – such as number of teeth – and collecting breed data. These new pieces of information could help us make health based inferences based on our dataset and answer if specific breeds oral microbiome changes are associated with breed. Additionally, we are currently working on a functional metabolic profiling pipeline for the same samples.

Acknowledgements

I would like to acknowledge the principal investigator, Dr. Matteo Pellegrini. I also would like to acknowledge fellow undergraduates that have helped me with this project including Zhiyue (Jennifer) Jiang and Aisling Archdeacon. Diagrams created in <https://BioRender.com>.

Student: Chrissy Tetreault

Faculty Advisor: Jake Lusic

Direct Mentor: Dylan Sarver

Department: Medicine/MIMG/Human Genetics

<https://lusic.genetics.ucla.edu/>

Establishing a Pipeline for Estimating Time-Course Protein Concentrations in Murine Liver

CHRISSY TETREULT^{1,2}, Dylan Sarver^{2,3,4}, Robert Mathews^{1,5}, Jake Lusic^{2,3,4}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Microbiology, Immunology, and Molecular Genetics, UCLA

³ Department of Medicine, Division of Cardiology, UCLA

⁴ Department of Human Genetics, University of California

⁵ Department of Molecular, Cellular, and Developmental Biology, UCLA

Determining absolute protein abundance is more time and resource intensive than relative quantification, but crucial for accurate pharmacokinetic modeling. We reasoned that relative-to-absolute scaling factors could be calculated if both measurements were present for circadian proteins, which could be applied to all relatively quantified proteins. Utilizing RStudio, we identified high-confidence peptides via ambiguous and phosphorylated peptide exclusion and reproducibility and mass spectrometry (MS) quality metrics. Scaling factors were calculated in a timepoint-specific manner using both relatively and absolutely quantified peptides and validated against matching circadian protein concentrations quantified in independent experiments. The same scaling factors were applied to the remaining relatively quantified proteins. This scalable pipeline enables researchers to estimate absolute protein abundance from relative measurements paired with minimal absolute data. We hope additional scaling factor quantification across a broad range of peptides and MS instrumentation will enable improved pharmacokinetic modeling of optimal drug timing and dosage for chronotherapeutic interventions.

**Chrissy
Tetreault**

**Microbiology,
Immunology,
and Molecular
Genetics -
UCLA**

**Expected year
of graduation
June 2026**



3-Dimensional Integration Framework for Atlas Generation: Absolute Circadian Proteome Abundance Estimation in Murine Liver

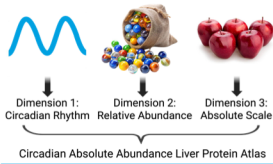
Chrissy Tetreault and Dylan Sarver

Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles

Contact: christet@g.ucla.edu



Graphical Abstract



Background

- Molecular reaction kinetics depend on reactant and product concentrations
- Measuring protein concentrations directly is resource-intensive
- Relative protein abundances (fold-changes) are more accessible
- Gene Regulatory Networks currently rely on relative data →
 - Limited utility - context dependent
- Time-specific absolute predictions are essential for chronopharmacology and personalized medicine development

Hypothesis

SINCE matched absolute and relative abundances can be used to calculate a scaling factor,
IF high-quality circadian absolute anchors are paired to their corresponding relative abundances, **THEN** the scaling factor estimates absolute abundances for all proteins quantified in the same relative dataset.

References

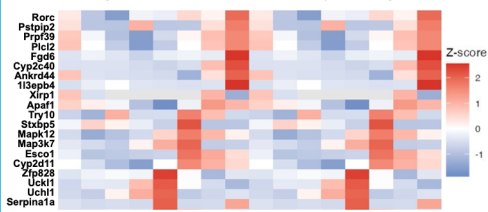
- Duong, Hai A., Kenichi Baba, Jason P. DeBruyne, et al. 2024. "Environmental Circadian Disruption Re-Writes Liver Circadian Proteomes." *Nature Communications* 15 (1): 9537. <https://doi.org/10.1038/s41467-024-49852-3>.
- Narumi, Ryohel, Yoshihiro Shimizu, Maki Ueki-Tadenuma, et al. 2016. "Mass Spectrometry-Based Absolute Quantification Reveals Rhythmic Variation of Mouse Circadian Clock Proteins." *Proceedings of the National Academy of Sciences* 113 (24): E3461-67. <https://doi.org/10.1073/pnas.1603799113>.
- Wang, Yunchi, Lai-Song Mingwei, Liu, et al. 2018. "A Proteomics Landscape of Circadian Clock in Mouse Liver." *Nature Communications* 9: 1-16. <https://doi.org/10.1038/s41467-018-03888-2>.
- Wei, Jing, Chen Ding, Jinsong Zhang, et al. 2014. "High-Throughput Absolute Quantification of Proteins Using an Improved Two-Dimensional Reversed-Phase Separation and Quantification Concatemer (QconCAT) Approach." *Analytical and Bioanalytical Chemistry* 406: 4183-93. <https://doi.org/10.1007/s00216-014-7784-x>.

Acknowledgements

Dr. Aldons Jake Lusis, PhD.
Robert Mathews

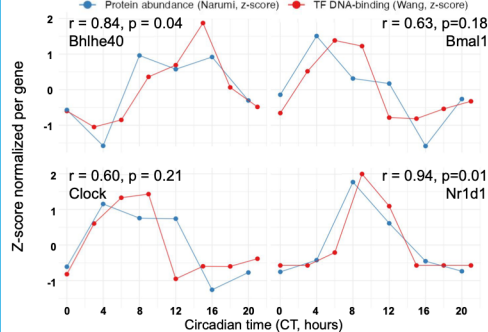
Methods and Results

Fig. 1: **Duong Dimension 1 - Top 20 rhythmic genes**



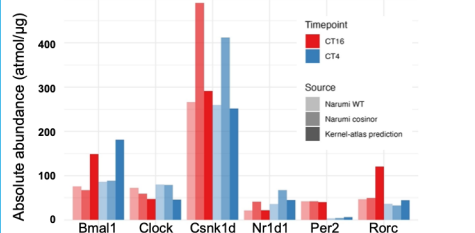
- Label Free Quantification Mass Spectrometry, Z-score normalized
- Duong-defined circadian rhythms, Dimension 1, 6501 genes

Fig. 3: **Transcription Factor abundance vs DNA-binding activity**



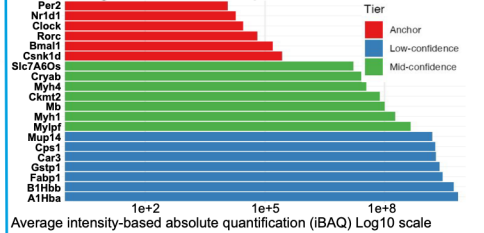
- Protein abundance lags slightly behind TF binding activity
- Not replicated for all TFs - these had most significant Pearson correlation
- Increases confidence in Narumi-defined absolute abundances, Dimension 3

Fig. 5: **Wildtype Validation: Narumi WT vs Gaussian (h) model**
CT4 and CT16 only (independent experiment 2-timepoint validation)



- Anchor weighting method for SF calculations incurred trade-offs between TF predictive accuracy against Narumi WT and QconCAT predictive accuracy

Fig. 2: **Wang Dimension 2 - Spread of iBAQ values**



- Average intensity-based absolute quantification (iBAQ) Log10 scale
- Semi-absolute, not truly absolute quantification, Dimension 2
- Wang-defined protein-to-protein average relative abundance, 6780 genes

Fig. 4a: **Global vs Local SF methods - high confidence genes**
n = 13 high-confidence QconCAT genes per method (same genes)
p-values: paired Wilcoxon local weighted method vs. global scale factor (SF)

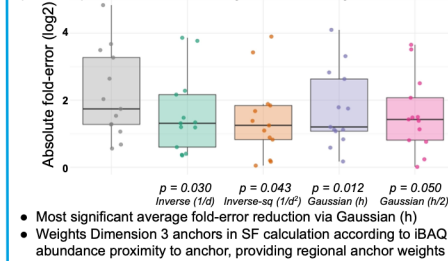
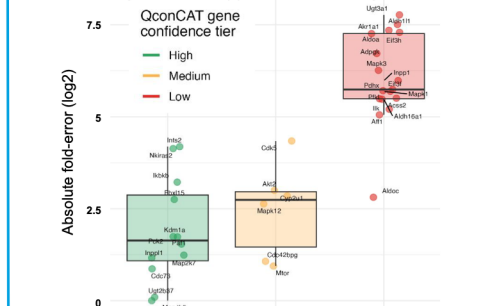


Fig. 4b: **Abundance determines confidence of atlas estimates**
n = 35 QconCAT validation genes across 7 orders of magnitude (log10 iBAQ)
Distance-weighted Gaussian (h) kernel SF instead of global constant SF



- n = 2061 or 32%, n = 2289 or 35%, and n = 2150 or 33% of atlas gene estimates fall in high, medium, and low confidence tiers, respectively.

Conclusions

- Proof-of-concept framework for estimating protein concentrations from relative abundance, rhythms, and high-confidence absolute anchors
- Accuracy increases closer to anchor calibration range
- High-confidence estimate error magnitude remains high - 3x on average, up to 17x
- TF rhythms anticipate DNA binding activity
- Estimates maintain rhythm
- Adjusting anchor parameters to increase predictive accuracy of one validation dataset decreases predictive accuracy of another
- Gaussian (h) kernel SF best predicts local abundance trends

Limitations

- Method proof-of-concept due to uncertainty
- Atlas **currently** invalid for predictions of GRN models or chronopharmacology
- Wang iBAQ underestimates low abundance and overestimates high abundance proteins
- Even local SFs assume ionization efficiency, which depends on multiple peptide properties, is entirely abundance dependent
- Anchor peptide set has unique structural characteristics

Future Directions

- Improve predictive accuracy with better and more dimension 2 and 3 training data
- Identify medium and high abundance protein anchors
- Normalize ionization efficiency variation
- Diversify anchor protein sample set to include metabolic and ribosomal enzymes, chromatin regulators, kinases, etc.
- Extend to nuclear fraction

Student: Nadia Tsai

Faculty Advisor: Molly Fox

Direct Mentor: Jennifer Bramen

Department: Anthropology, Psychiatry

<https://motherhoodlabucla.mystrikingly.com/>

Regulatory T Cells and the Old Friends Hypothesis: Investigating the Effects of Urbanization on Immune Regulation

NADIA TSAI¹, S. Madigan Durham²

Urbanization has been associated with a rising incidence of autoimmune and chronic inflammatory diseases, prompting interest in how environmental change influences immune regulation. The Old Friends Hypothesis proposes that reduced exposure to evolutionarily familiar microorganisms disrupts immunoregulatory mechanisms, particularly those mediated by regulatory T (Treg) cells. While previous work has established the evolutionary basis of this hypothesis and characterized the molecular pathways governing Treg development and function, few studies have integrated these perspectives across human populations experiencing different degrees of urbanization. We investigated whether transcriptomic patterns of Treg-associated genes in Orang Asli communities in Malaysia, which represent a gradient of traditional and urbanized lifestyles, are consistent with immunoregulatory changes predicted by the Old Friends Hypothesis. Candidate genes involved in Treg identity, suppressive function, and immune regulation were identified through a literature review and selected for future RNA-seq analyses, including differential gene expression testing, principal component analysis, hierarchical clustering, and pathway enrichment. This work will establish a framework for examining how environmental modernization may influence immune homeostasis and inform future studies integrating environmental exposures with immunological mechanisms underlying autoimmune and chronic inflammatory disease.

Nadia Tsai

**Neuroscience -
UCLA**

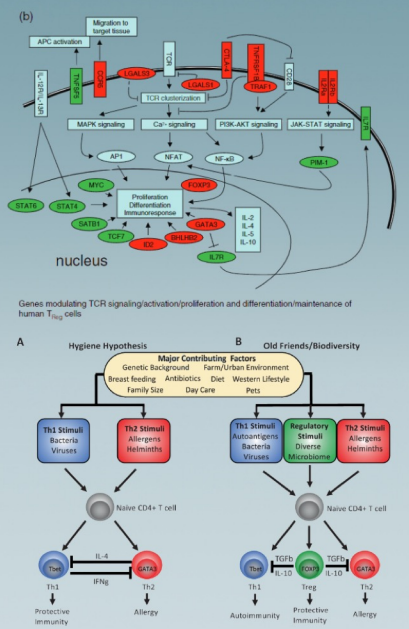
**Expected year
of graduation
June 2027**

Regulatory T Cells and the Old Friends Hypothesis: Investigating the Effects of Urbanization on Immune Regulation

Nadia Tsai¹, S. Madigan Durham, Dr. Molly Fox
Biological Anthropology of Motherhood Lab, Department of Anthropology, University of California - Los Angeles, Los Angeles, CA, USA



Background

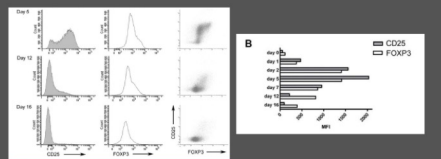
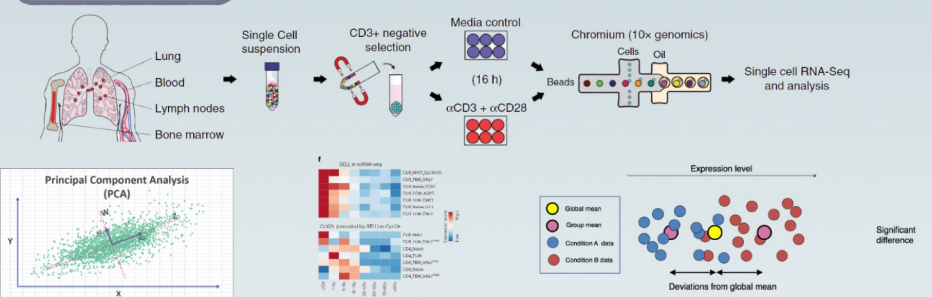


Results

Literature review identified a panel of Treg-associated genes implicated in lineage specification, suppressive function, and immune homeostasis based on the Pfoertner et al. paper.

Gene symbol	Gene name	Annotation
IL2	Interleukin 2	
ACTN1	Actinin, alpha 1	SLE, CHA
IL21	Interleukin 21	
IL22	Interleukin 22	UC, MC, COPD, T2D, AS, IAD, RA, ALPS, SLE
STAT1	TGF-beta induced transcription factor 1	
STAT3	TGF-beta induced transcription factor 1	
IL21R	Interleukin 21 receptor	
IL22R1	Interleukin 22 receptor, alpha	JA, RA, BD
IL22R2	Interleukin 22 receptor, beta	RA, JA
IL22R3	Interleukin 22 receptor, gamma	T1D, SLE, SLE, BD
IL23	Interleukin 23	
IL24	Interleukin 24	
IL25	Interleukin 25	
IL26	Interleukin 26	
IL27	Interleukin 27	
IL28	Interleukin 28	
IL29	Interleukin 29	
IL30	Interleukin 30	
IL31	Interleukin 31	
IL32	Interleukin 32	
IL33	Interleukin 33	
IL34	Interleukin 34	
IL35	Interleukin 35	
IL36	Interleukin 36	
IL37	Interleukin 37	
IL38	Interleukin 38	
IL39	Interleukin 39	
IL40	Interleukin 40	
IL41	Interleukin 41	
IL42	Interleukin 42	
IL43	Interleukin 43	
IL44	Interleukin 44	
IL45	Interleukin 45	
IL46	Interleukin 46	
IL47	Interleukin 47	
IL48	Interleukin 48	
IL49	Interleukin 49	
IL50	Interleukin 50	
IL51	Interleukin 51	
IL52	Interleukin 52	
IL53	Interleukin 53	
IL54	Interleukin 54	
IL55	Interleukin 55	
IL56	Interleukin 56	
IL57	Interleukin 57	
IL58	Interleukin 58	
IL59	Interleukin 59	
IL60	Interleukin 60	
IL61	Interleukin 61	
IL62	Interleukin 62	
IL63	Interleukin 63	
IL64	Interleukin 64	
IL65	Interleukin 65	
IL66	Interleukin 66	
IL67	Interleukin 67	
IL68	Interleukin 68	
IL69	Interleukin 69	
IL70	Interleukin 70	
IL71	Interleukin 71	
IL72	Interleukin 72	
IL73	Interleukin 73	
IL74	Interleukin 74	
IL75	Interleukin 75	
IL76	Interleukin 76	
IL77	Interleukin 77	
IL78	Interleukin 78	
IL79	Interleukin 79	
IL80	Interleukin 80	
IL81	Interleukin 81	
IL82	Interleukin 82	
IL83	Interleukin 83	
IL84	Interleukin 84	
IL85	Interleukin 85	
IL86	Interleukin 86	
IL87	Interleukin 87	
IL88	Interleukin 88	
IL89	Interleukin 89	
IL90	Interleukin 90	
IL91	Interleukin 91	
IL92	Interleukin 92	
IL93	Interleukin 93	
IL94	Interleukin 94	
IL95	Interleukin 95	
IL96	Interleukin 96	
IL97	Interleukin 97	
IL98	Interleukin 98	
IL99	Interleukin 99	
IL100	Interleukin 100	

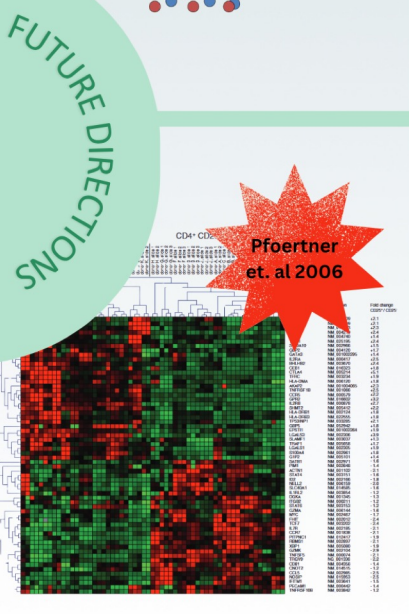
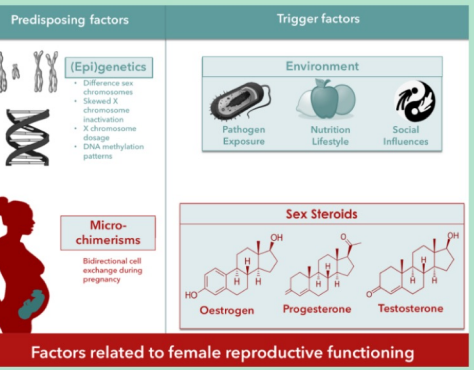
Methodology



Studies consistently recommend peripheral blood mononuclear cells (PBMCs) as an appropriate sample source for evaluating systemic Treg-associated gene expression, making them well suited for comparisons across human populations.

Conclusions

- A literature review identified a targeted panel of regulatory T cell (Treg)-associated genes, including FOXP3, IL2RA (CD25), CTLA4, IL10, ENTPD1 (CD39), TIGIT, and IKZF2 (Helios), that collectively represent Treg identity, suppressive function, and maintenance of immune tolerance.
- These candidate genes will serve as the foundation for downstream analyses comparing immune regulatory profiles across Orang Asli communities representing different degrees of urbanization.
- Future analyses will include RNA-seq quality control, normalization, principal component analysis (PCA), differential gene expression analysis (DESeq2), hierarchical clustering, and pathway enrichment to identify patterns associated with environmental modernization.
- By integrating gene expression with environmental and lifestyle data, this study aims to evaluate whether immunoregulatory signatures are consistent with predictions of the Old Friends Hypothesis.
- This work will provide a foundation for understanding how environmental change may influence immune homeostasis and inform future research on the mechanisms underlying autoimmune and chronic inflammatory diseases



References

Huang, T., et al. (2025). Integrating single-cell RNA and T cell/B cell receptor sequencing with mass cytometry reveals dynamic trajectories of human peripheral immune cells from birth to old age.

Pfoertner, S., et al. (2006). Signatures of human regulatory T cells: an encounter with old friends and new players.

Selechnik, D., et al. (2019). Immune and environment-driven gene expression during invasion: An eco-immunological application of RNA-Seq.

Szabo, P. A., et al. (2019). Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease.

Baecher-Allan, C., et al. (2006). Human regulatory T cells and their role in autoimmune disease.

Orang Asli Health and Lifeways Project. <https://www.orangaslihealth.org/>



Student: Valerie Vandersall
Faculty Advisor: Mehdi Bouhaddou
Direct Mentor: Jennifer Bramen
Department: Microbiology, Immunology, and
Molecular Genetics (MIMG)
<https://www.bouhaddoulab.org/>

**Valerie
Vandersall**

**Biochemistry -
UCLA**

**Expected year
of graduation
June 2027**

Conserved phosphorylation sites influence alphavirus replication and suggest host kinase regulation of viral fitness

VALERIE VANDERSALL ^{1,2,3,5,6}, Declan M. Winters ^{2,4,6}, Evan Wynn ^{2,3,6}, Erin Kim ^{3,5}, Kareem Alba ^{2,3,6}, Justin Selby ^{2,3,6}, Melody MH Li ^{2,3}, Mehdi Bouhaddou ^{2,3,5}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Microbiology, Immunology, & Molecular Genetics, UCLA

³ Molecular Biology Institute, UCLA

⁴ Department of Human Genetics, UCLA

⁵ Department of Chemistry and Biochemistry, UCLA

⁶ Institute for Quantitative and Computational Biosciences, UCLA

Alphaviruses are enveloped, single-stranded, positive-sense RNA viruses primarily transmitted by mosquitoes. They represent an increasing global health concern, yet no approved antiviral treatments are currently available. Although phosphorylation is a well-established regulator of viral protein function across different virus families, whether and how phosphorylation regulates alphavirus replication remains poorly understood. Using phosphoproteomics of SINV-infected HEK293T cells across three independent experiments, we identified 49 reproducibly detected phosphorylation sites that correspond to conserved residues across 11 major human-infecting alphaviruses. Each site was mutated to make either phosphomimetic or phosphodead variants, and infectious particle production was assayed by plaque assays to determine sites important for viral fitness. We next leveraged Kinase Library motif prediction scores to investigate whether host kinase activity may underlie these events. We observed that sites with higher predicted kinase motif scores were more frequently associated with replication defects upon mutation. This may indicate that phosphorylation of viral proteins by host kinases is an important regulator of the alphavirus life cycle.

Conserved phosphorylation sites influence alphavirus replication and suggest host kinase regulation of viral fitness

VALERIE VANDERSALL^{1,2,3,4}, Declan M. Winters^{2,4}, Evan Wynn^{2,3,4}, Erin Kim^{2,3}, Kareem Alba^{2,3,4}, Justin Selby^{2,3,4}, Melody MH Li^{3,4} & Mehdi Bouhaddou^{2,3,4}

¹ Bruins in Genomics (BIG) Summer Program, Institute for Quantitative and Computational Biosciences, University of California, Los Angeles

² Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles

³ Molecular Biology Institute, University of California, Los Angeles

⁴ Institute for Quantitative and Computational Biosciences, University of California, Los Angeles



Introduction

- Alphaviruses are mosquito-borne, enveloped, positive-sense single-stranded RNA viruses that cause a range of disease from chronic arthritis to encephalitis
- Outbreaks pose a growing health concern, yet there are no approved antiviral treatments
- Phosphorylation regulates viral protein function across diverse viral families, but its contribution to alphavirus replication is poorly understood¹
- Defining functional phosphorylation sites may reveal how alphaviruses interact with and respond to host signaling pathways²

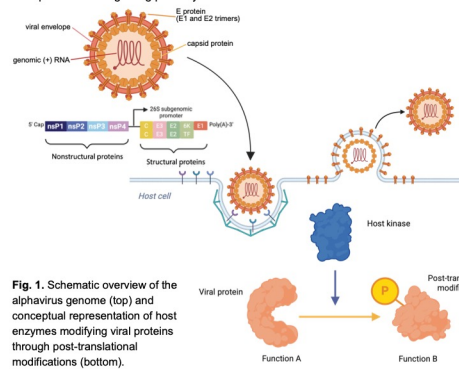


Fig. 1. Schematic overview of the alphavirus genome (top) and conceptual representation of host enzymes modifying viral proteins through post-translational modifications (bottom).

Experimental Workflow

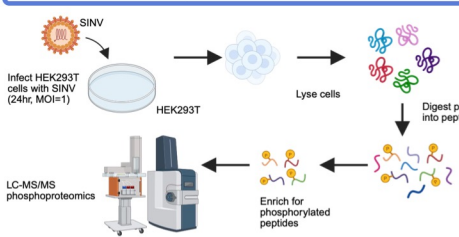


Fig. 2. LC-MS/MS phosphoproteomics to identify phosphorylation sites on viral proteins.

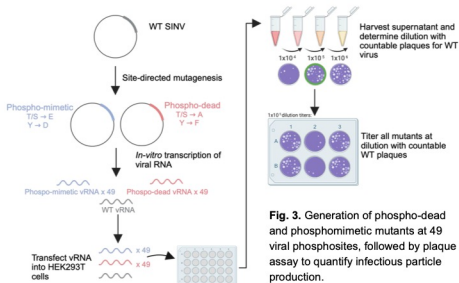


Fig. 3. Generation of phospho-dead and phosphomimetic mutants at 49 viral phosphosites, followed by plaque assay to quantify infectious particle production.

Experimental Results

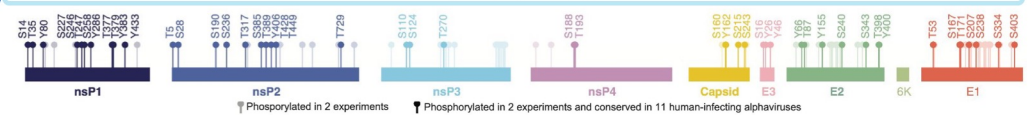


Fig. 4. Schematic representation of highly conserved phosphorylation sites detected across the SINV proteome in 2/3 independent infections.

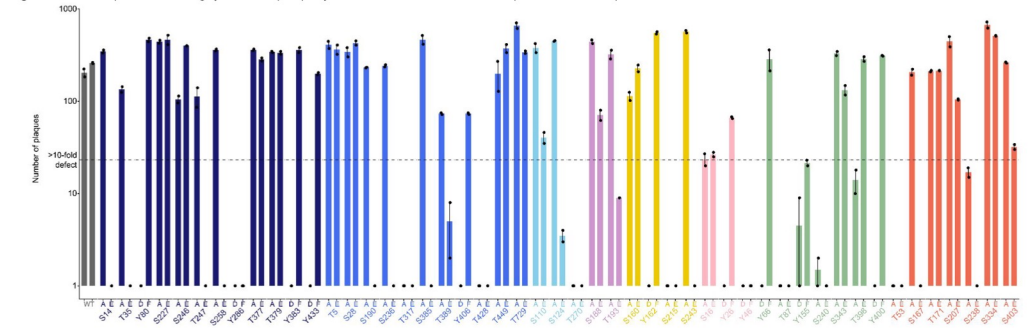
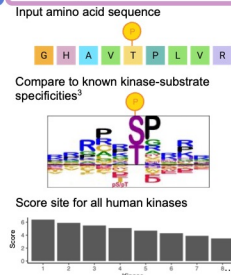


Fig. 5. Plaque counts at a 10⁻⁵ dilution for 49 phospho-dead and phosphomimetic mutants, with WT controls shown on the left. Dashed line represents >10-fold defect compared to WT.

Computational Workflow



Conclusions

- Phosphosite mutations reduce SINV infectious particle production
- Strongly defective sites are enriched for distinct kinase families
- Defective nsP1 sites have higher predicted MAPK scores than non-defective sites
- nsP1 phosphorylation may couple SINV replication to host MAPK signaling

References

- Bouhaddou et al. 2020. *Cell*.
- Alba et al. 2025. *bioRxiv*.
- Johnson et al. 2023. *Nature*.

Computational Analysis

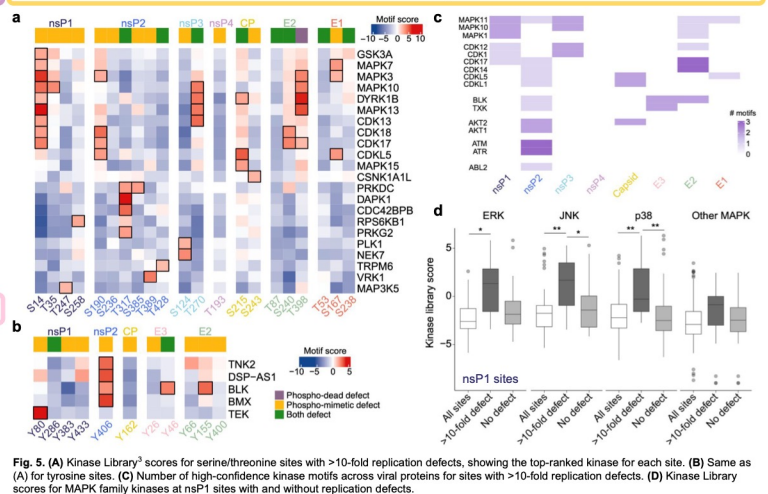


Fig. 5. (A) Kinase Library³ scores for serine/threonine sites with >10-fold replication defects, showing the top-ranked kinase for each site. (B) Same as (A) for tyrosine sites. (C) Number of high-confidence kinase motifs across viral proteins for sites with >10-fold replication defects. (D) Kinase Library scores for MAPK family kinases at nsP1 sites with and without replication defects.

Acknowledgements

This work was supported by the UCLA AIDS Institute, the James B. Pendleton Charitable Trust, and the McCarthy Family Foundation. We would like to acknowledge funding from the National Institutes of Health (NIH-NIGMS R35GM160071 to MB), the Cellular and Molecular Biology T32 at UCLA (NIH-NIGMS T32GM145388 to DMW), and the Microbial Pathogenesis Training Grant T32 at UCLA (T32AI007323 to DMW).



Student: Jingyue Wang
Faculty Advisor: Neil Harris
Direct Mentor: Azad Azargushasb
Department: Neurosurgery
<https://neilharrislab.dgsom.ucla.edu/>

Jingyue Wang

**Neuroscience -
UCLA**

**Expected year
of graduation
June 2028**

Amide Proton Transfer (APT) MRI in Revealing Cellularity Shifts in Traumatic Brain Injury

JINGYUE WANG¹, Gregory Smith², Azad Azargushasb², Samuel Vander Dussen², Neil Harris²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² David Geffen School of Medicine, UCLA

Standard neuroimaging captures structural traumatic brain injury (TBI) damage, but fails to detect cellularity changes in the subacute to chronic phase. Therefore, this study aims to determine whether Amide Proton Transfer (APT) MRI can complement structural imaging by revealing biological mechanisms across TBI progression. APT MRI was integrated alongside diffusion-weighted imaging, restricted diffusion imaging (RDI), and Jacobian determinants in a controlled cortical impact rat model at days 3 (subacute) and 30 (chronic), utilizing principal component analysis and voxel-wise analysis. Analysis revealed that subacute lesions spanned a continuous spectrum of heterogeneity, whereas chronic injuries clustered into two phenotypes. Furthermore, early APT value elevation is aligned with RDI and edema, whereas chronic APT value decreases correlate with atrophy. The findings reveal the meaning of APT changes in the context of TBI, as it shows neuroinflammation and altered cellularity; therefore, it is shown to be an additional non-invasive tool for tracking injury progression.

Amide Proton Transfer (APT) MRI in Revealing Cellularity Shifts in Traumatic Brain Injury

JINGYUE WANG¹, Gregory Smith², Azad Azargushasb², Samuel Vander Dussen², Neil Harris²

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² UCLA Brain Injury Research Center, Department of Neurosurgery, David Geffen School of Medicine, UCLA

Abstract

Standard neuroimaging captures structural traumatic brain injury (TBI) damage, but fails to detect cellularity changes in the subacute to chronic phase. Therefore, this study aims to determine whether Amide Proton Transfer (APT) MRI can complement structural imaging by revealing biological mechanisms across TBI progression.

APT MRI was integrated alongside diffusion-weighted imaging, restricted diffusion imaging (RDI), and Jacobian determinants in a controlled cortical impact rat model at days 3 (subacute) and 30 (chronic), utilizing voxel-wise analysis.

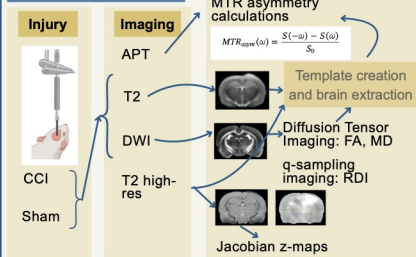
Analysis revealed that early APT value elevation is aligned with RDI and edema, whereas chronic APT value decreases correlate with atrophy.

The findings reveal the significance of APT changes in the context of TBI, as they show neuroinflammation and altered cellularity; therefore, they are shown to be an additional non-invasive tool for tracking injury progression.

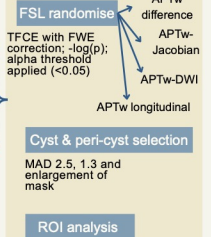
Background

- **Amide Proton Transfer (APT) MRI**
Non-invasively map mobile cellular proteins, peptides, and tissue pH, measured in MTR_{asym}, and referred to as % CEH
- **Controlled Cortical Impact (CCI)**
Experimental model to study traumatic brain injury (TBI) in mice or rats where a metal pit hits the brain tissue
- **Diffusion-weighted imaging (DWI)**
MRI scan that measures the movement of water molecules
 - Fractional anisotropy (FA): scale of isotropic (0) to anisotropic (1) used for white-matter integrity
 - Mean diffusivity (MD): average magnitude of water molecule diffusion within tissue
 - Restricted diffusion imaging (RDI): density of restricted water molecules for cellularity and inflammation

Methodology



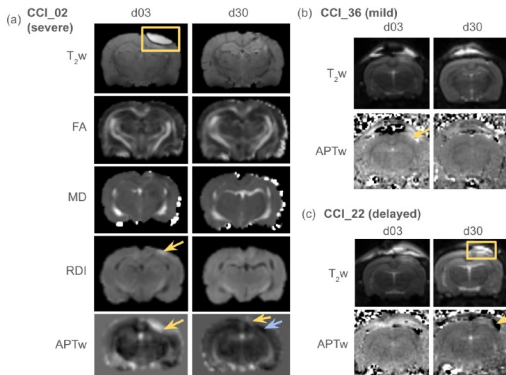
Analysis



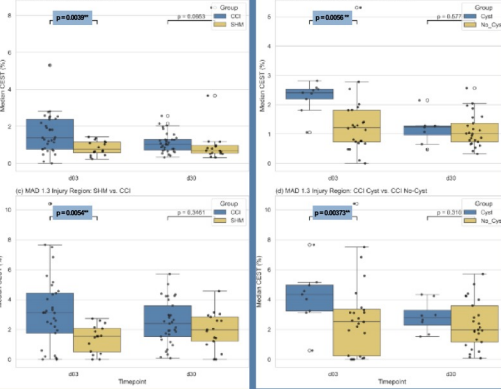
Results

1. APTw in relations to T2w imaging and cyst formation

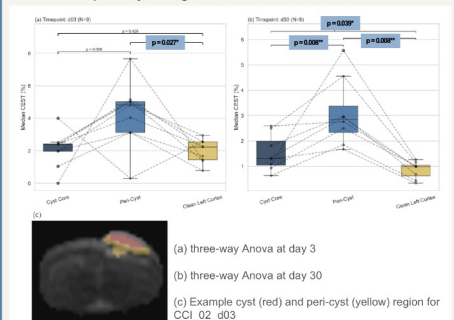
(i) Example MR images of CCI rats



(ii) APTw comparison of subjects with and without cysts

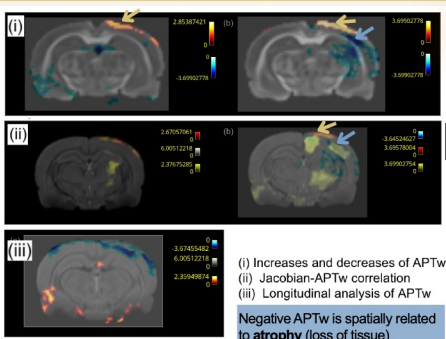


(iii) Within subject comparison of APTw at cyst region, peri-cyst region, and the left cortex

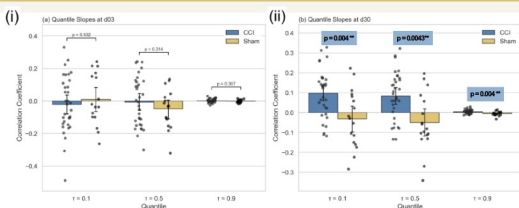


(ii) Results suggest that subjects with cysts at day 3 have a higher local (injury site) and global (left cortex) median CEST% (APT_w), meaning that APT_w increases are directly related to inflammation and swelling.

2.FSL randomise voxel-wise analysis of APTw, APTw-Jacobian, and longitudinal APTw changes



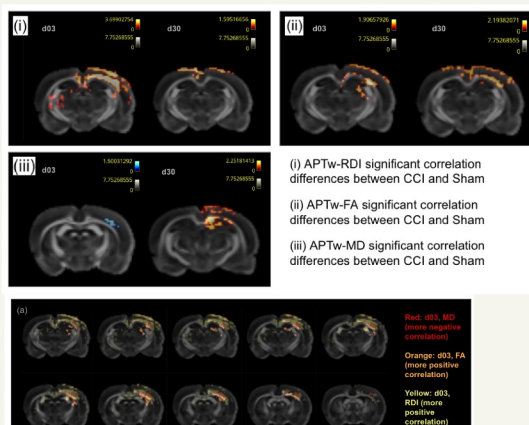
3. APTw-Jacobian correlations at various quartiles



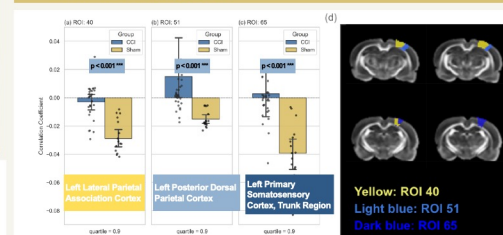
(i) No significant differences in the correlations at day 3

(ii) Significant differences in correlation at both the high and low values of APTw at day 30, with positive correlations for CCI compared to Sham

4. APTw-DWI correlation and overlap analysis



5. Subject-wise APTw-RDI analysis specific to ROIs near the injury site



Discussion

Significance:

- **Significance:**
 - APTW increase, especially at day 3, is related to damaged tissue (T2w hyperintensity) and is related to inflammation, but the increase is not limited to the CSF cyst itself
 - APTW decrease is mainly observed chronically at day 30 at the sensory cortex above the hippocampus, and it is correlated with the overlapping d03 of APTW-DWI
 - APTW-RDI has the largest changes in correlation between CCI and Sham, with healthy tissue correlation being negative and damaged tissue correlation, especially at high RDI regions, being positive, further supporting the inflammation hypothesis
 - Jacobian negative at day 30 is spatially correlated with APTW decrease, suggesting that APTW decrease is related to loss of cellularity and tissue
 - Longitudinally, there is a general depression of APTW near the injury site from day 3 to day 30

Future directions:

- Finalize PCA & GMM categorization of APTw
- Use biomarkers such as GFAP, tau, NFL, and Aldolase-C to evaluate changes in cellularly and match with APTw
- Compare area under the curve (AUC) prediction accuracy of behavioral outcome using APTw, DWI, and T2W
- Use APTw to evaluate the usefulness of the surgeon's severity labeling (High and Low)

Conclusion

Results showed APTw changes in the context of TBI as neuroinflammation and loss of cellularity longitudinally, and they have correlations with DWI metrics

Acknowledgements

Jinyuan Zhou group at JHU for providing code for APTw analysis



Student: Xiangting Wu
Faculty Advisor: Wentai Liu
Direct Mentor: Ryan Kochis
Department: Applied Mathematics
<https://brl.seas.ucla.edu/>

Xiangting Wu

**Applied
Mathematics -
UCLA**

**Expected year
of graduation
June 2027**

Multiscale iEEG Biomarkers and Patient-Specific Digital Brain Twins for Alzheimer's Disease Risk and Neuromodulation

XIANGTING WU^{1,2,3,4}, Siarra Huang^{3,4}, Sana Meher^{3,4}, Ryan Kochis^{3,4}, Sandeep Vardhan^{3,4}, Wentai Liu^{3,4}

¹ BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

² Department of Mathematics, UCLA

³ Department of Bioengineering, UCLA Samueli School of Engineering

⁴ California NanoSystems Institute, UCLA

High-frequency oscillations (HFOs), interictal epileptiform discharges (IEDs), and low-frequency dynamics provide complementary signatures of network dysfunction relevant to epilepsy and Alzheimer's disease (AD). We are developing a multiscale human iEEG framework integrating automated biomarker detection, AD-risk evaluation, and patient-specific digital brain modeling. Using CRCNS recordings, expert-annotated IED data, and UCLA clinical iEEG, we developed frequency-informed HFO detection with contextual calibration and are integrating HFO/fast-ripple, IED, and low-frequency features into individualized electrophysiological profiles. Our HFO classifier achieved 0.876 recall and 0.839 precision, while temporal event localization remained a major limitation. These biomarkers are now being incorporated into patient-specific neural-mass models to construct virtual brain twins that reproduce individual network dynamics and support in-silico optimization of neuromodulation targets, timing, and stimulation parameters before clinical validation.

Multiscale iEEG Biomarkers and Patient-Specific Digital Brain Twins for Alzheimer's Disease Risk and Neuromodulation

XIANGTING WU*, Sana Meher, Sierra Huang, Ryan Kochis, Sandeep Vardhan, Wentai Liu
Department of Bioengineering, UCLA California NanoSystems Institute, University of California, Los Angeles

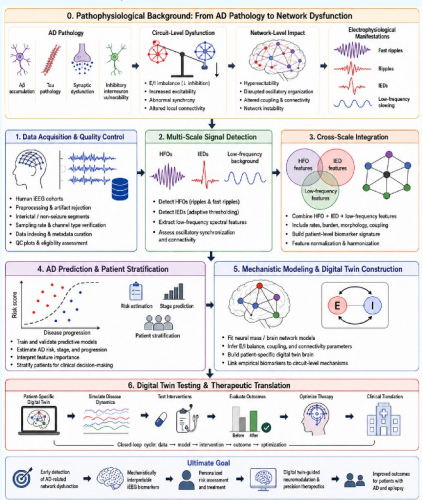
Abstract

High-frequency oscillations (HFOs), interictal epileptiform discharges (IEDs), and low-frequency dynamics provide complementary signatures of network dysfunction relevant to epilepsy and Alzheimer's disease (AD). We are developing a multiscale human iEEG framework integrating automated biomarker detection, AD-risk evaluation, and patient-specific digital brain modeling. Using CRNS recordings, expert-annotated IED data, and UCLA clinical iEEG, we developed frequency-informed HFO detection with contextual calibration and are integrating HFO/fast-ripple, IED, and low-frequency features into individualized electrophysiological profiles. Our HFO classifier achieved 0.876 recall and 0.839 precision, while temporal event localization remained a major limitation. These biomarkers are now being incorporated into patient-specific neural-mass models to construct virtual brain twins that reproduce individual network dynamics and support in-silico optimization of neuromodulation targets, timing, and stimulation parameters before clinical validation.

Background

Pathophysiological Rationale

AD-related molecular and circuit dysfunction may produce detectable changes in neural excitability before overt structural pathology. We therefore integrate focal epileptiform biomarkers with broader network dynamics to characterize early, patient-specific electrophysiological dysfunction. Longitudinal HAP-120 data support this rationale, showing E/I imbalance and pathological HFO activity preceding detectable Aβ plaque accumulation.



Pipeline Objective

The framework progresses from human iEEG acquisition to HFO, IED, and low-frequency network-state characterization, then integrates these signals into patient-level biomarkers for AD prediction. Combining event morphology with epileptiform and background context is intended to improve pathological specificity—particularly because ripples may be physiological or pathological and can differ in their coupling to IEDs, slow waves, and sleep spindles. The resulting biomarker profile can subsequently constrain mechanistic neural-mass and digital-twin models for patient-specific simulation and intervention testing.

Fast-Ripple Detection and Artifact Discrimination

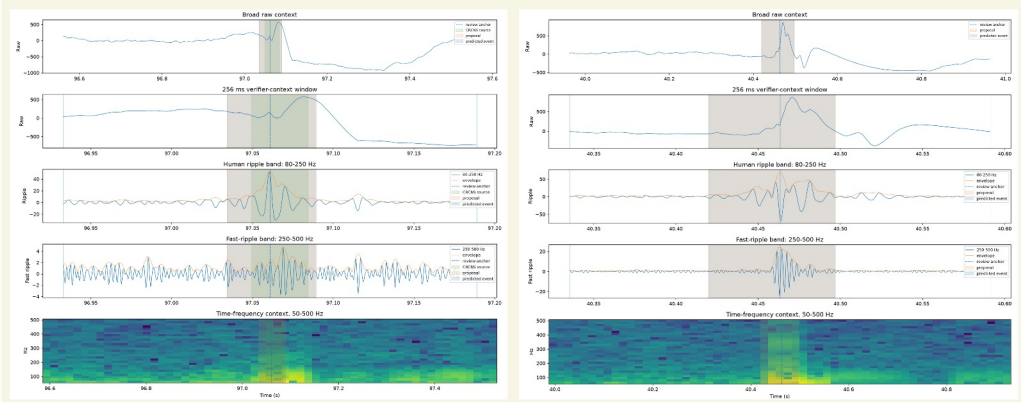


Figure. Representative fast-ripple detection and artifact discrimination.

A true Fast Ripple (left) shows a compact 250-500 Hz oscillatory event with localized spectral energy, while a sharp transient (right) generates filter-induced high-frequency activity resembling a Fast Ripple. Comparison across raw signal, frequency-specific traces, and time-frequency representations helps separate genuine HFOs from transient-related artifacts.

Cross-Patient Generalization of HFO Spatial Burden and Fast Ripple Classification

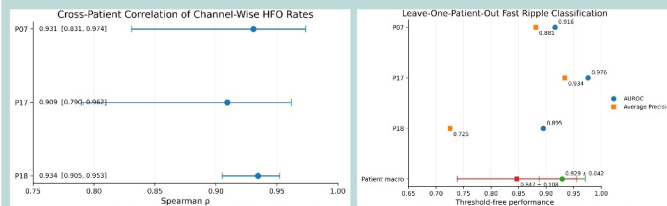


Figure. Predicted channel-wise HFO burden strongly preserved the CRNS spatial ranking across all three patients (Spearman $\rho=0.931, 0.909, 0.934$), supporting relative spatial burden as a robust downstream biomarker. Correlation evaluated at channel level rather than by absolute event count, emphasizing preservation of spatial disease burden.

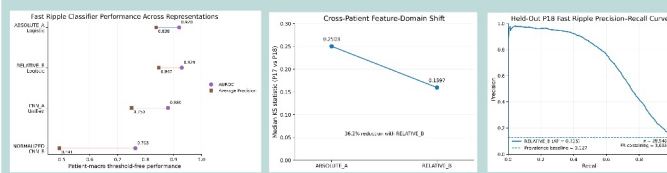


Figure. Among the four FR classifiers tested, Figure. Predicted channel-wise HFO rates strongly Figure. Completely held-out P18 data, preserved the CRNS reference ranking across RELATIVE_B achieved AP = 0.725 with a macro performance (AP = 0.847; AUROC = 0.929), three patients (Spearman $\rho=0.931, 0.909, 0.934$), positive-event prevalence of 0.127 supporting patient-relative spectral features as the best-performing cross-patient representation in this comparison. Spearman correlation evaluates channel-wise rank (3,633/28,548). Substantial precision-recall ordering rather than agreement in absolute event performance above the prevalence baseline. counts, supporting relative spatial HFO burden as a candidate downstream feature.

Model for Prediction and Intervention

$$\dot{\mathbf{x}}_i = \mathbf{f}_i(\mathbf{x}_i, \mathbf{b}_i) + \sum_j \mathbf{S}_{ij} \mathbf{h}_j(\mathbf{x}_j) + \mathbf{p}_i(\mathbf{x}_i, \mathbf{b}_i, \mathbf{w})$$

Established local dynamics + Established network coupling + Proposed pathology modulation

- Local dynamics (Jansen-Rit)**

$$\dot{\mathbf{x}}_i = \mathbf{f}_i(\mathbf{x}_i, \mathbf{b}_i) + \sum_j \mathbf{S}_{ij} \mathbf{h}_j(\mathbf{x}_j) + \mathbf{p}_i(\mathbf{x}_i, \mathbf{b}_i, \mathbf{w})$$

$$\mathbf{f}_i = \begin{bmatrix} \dot{V}_i \\ \dot{V}_i^{\text{exc}} \\ \dot{V}_i^{\text{inh}} \end{bmatrix} = \begin{bmatrix} -V_i + V_i^{\text{exc}} - V_i^{\text{inh}} \\ V_i^{\text{exc}} - V_i^{\text{inh}} \\ V_i^{\text{inh}} - V_i^{\text{exc}} \end{bmatrix}$$

$$\mathbf{h}_i = \begin{bmatrix} V_i^{\text{exc}} \\ V_i^{\text{inh}} \end{bmatrix}$$

$$\mathbf{p}_i = \begin{bmatrix} V_i^{\text{exc}} \\ V_i^{\text{inh}} \end{bmatrix}$$
- Whole-brain coupling**

$$\mathbf{S}_{ij} = \mathbf{C}_{ij} \mathbf{S}_{ij}(\mathbf{x}_i, \mathbf{x}_j) \quad \mathbf{S}_{ij} = \text{structural connectivity} \quad \mathbf{C}_{ij} = \text{global coupling}$$
- Proposed pathology-informed modulation**

$$\mathbf{p}_i = \mathbf{J}_i \mathbf{f}_i \mathbf{b}_i \quad \mathbf{J}_i = \begin{bmatrix} \mathbf{J}_i^{\text{exc}} \\ \mathbf{J}_i^{\text{inh}} \end{bmatrix} \quad \Delta \mathbf{b}_i = \mathbf{W}_i \mathbf{b}_i$$

$$\mathbf{f}_i(\mathbf{x}_i, \mathbf{b}_i) = \mathbf{f}_i(\mathbf{x}_i, \mathbf{b}_i) + \mathbf{J}_i \mathbf{f}_i \mathbf{b}_i$$

Mechanistic modeling framework for patient-specific electrophysiology. Established Jansen-Rit neural-mass dynamics and whole-brain network coupling are extended with a proposed pathology-informed term derived from HFO, IED, spectral, and connectivity biomarkers. The model will be fit to estimate latent circuit parameters related to E/I balance and coupling, generate synthetic iEEG, and support future AD-risk modeling and patient-specific intervention simulation.

References

- Liao KY, Han X, Chen WY, Hsin YL, Liu W. Temporal Dynamics of High-Frequency Oscillations in Alzheimer's Disease: A Longitudinal Study in HAP-120 Mice. *bioRxiv*. 2025.
- Ye H, Chen C, Weiss SA, Wang S. Pathological and Physiological High-Frequency Oscillations on Electroencephalography in Patients with Epilepsy. *Neurosci Bull*. 2024;40:609-620.
- Vossel KA, et al. Seizures and Epileptiform Activity in the Early Stages of Alzheimer Disease. *JAMA Neurol*. 2013;70:1158-1166.
- Zijlmans M, Worrell GA, Dümpelmann M, et al. How to record high-frequency oscillations in epilepsy: A practical guideline. *Epilepsia*. 2017;58:1305-1315.
- Frauscher S, Bartolomei F, Kobayashi K, et al. High-frequency oscillations: The state of clinical research. *Epilepsia*. 2017;58:1316-1329.
- Jansen BH, Rit VG. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol Cybern*. 1995;73:357-366.
- Fedelet T, Burnos S, Boran E, et al. Resection of high frequency oscillations predicts seizure outcome in the individual patient. *Sci Rep*. 2017;7:13836. CRNS leq-1 dataset, doi:10.6080/K06QVQV5.



Student: Jessie Xu

Faculty Advisor: Anthony Covarrubias

Direct Mentor: Utkarsh Tripathi

Department: Microbiology, Immunology, and
Molecular Genetics (MIMG)

<https://www.covarrubiaslab.com/>

Jessie Xu

**Molecular, Cell
and
Developmental
Biology - UCLA**

**Expected year
of graduation
June 2027**

Loss of the Y Chromosome Reshapes Senescence-Associated Transcriptional Programs in Mouse Macrophages

JESSIE XU¹, Utkarsh Tripathi¹, Anthony J. Covarrubias¹

¹ Dept. of Microbiology, Immunology & Molecular Genetics (MIMG), DGSOM, UCLA

Mosaic loss of the Y chromosome (LOY) is a common age-associated chromosomal alteration in men linked to age-related disease, yet its effects on macrophage senescence remain unclear. We asked whether LOY alters macrophage transcriptional responses to senescence. Bone marrow-derived macrophages from XY and XO mice were exposed to 10 Gy irradiation and analyzed 10 days later by bulk RNA-seq (n=3/group), followed by DESeq2, genotype-by-senescence interaction, and gene-set enrichment analyses. Senescence dominated transcriptional variation (PC1, 85.4%), while genotype separated samples along PC2 (5.8%). Senescence altered 3,102 genes in XY and 4,010 in XO macrophages ($p_{adj} < 0.05$, $|\log_2 FC| > 0.5$). Interaction analysis identified 959 genes with significant genotype-by-senescence interactions and revealed stronger suppression of ribosomal, DNA-replication, cell-cycle, and oxidative-phosphorylation programs in XO macrophages, with differential lipid/steroid and lysosomal remodeling. These findings suggest that Y-chromosome status shapes macrophage senescence programs relevant to aging.

Loss of the Y Chromosome Reshapes Senescence-Associated Transcriptional Programs in Mouse Macrophages

JESSIE XU¹, Utkarsh Tripathi¹, Anthony J. Covarrubias¹¹ Dept. of Microbiology, Immunology & Molecular Genetics (MIMG), David Geffen School of Medicine, UCLA

Abstract

Mosaic loss of the Y chromosome (LOY) is a common age-associated chromosomal alteration in men linked to age-related disease, yet its effects on macrophage senescence remain unclear. We asked whether LOY alters macrophage transcriptional responses to senescence. Bone marrow-derived macrophages from XY and XO mice were exposed to 10 Gy irradiation and analyzed 10 days later by bulk RNA-seq (n=3/group), followed by DESeq2, genotype-by-senescence interaction, and gene-set enrichment analyses. Senescence dominated transcriptional variation (PC1, 85.4%), while genotype separated samples along PC2 (5.8%). Senescence altered 3,102 genes in XY and 4,010 in XO macrophages (padj<0.05, |log2FC|>0.5). Interaction analysis identified 959 genes with significant genotype-by-senescence interactions and revealed stronger suppression of ribosomal, DNA-replication, cell-cycle, and oxidative-phosphorylation programs in XO macrophages, with differential lipid/steroid and lysosomal remodeling. These findings suggest that Y-chromosome status shapes macrophage senescence programs relevant to aging.

Introduction

Background:

- LOY is a common somatic chromosomal alteration that increases in frequency with age.
- LOY is associated with increased risk of age-related disease.
- Prior Covarrubias Lab work showed that partial CRISPR-mediated Y-chromosome loss increased β -gal expression and cell-cycle arrest, providing a rationale to investigate senescence-associated transcriptional programs.

Knowledge gap: How Y-chromosome loss alters the macrophage transcriptional response to senescence remains unclear.

Research question: Does Y-chromosome status modify the macrophage transcriptional response to senescence?

Working hypothesis: Y-chromosome loss modifies the transcriptional programs engaged during macrophage senescence.

Methods

Experimental design: 2x2 factorial

NS-XY n=3	SEN-XY n=3
NS-XO n=3	SEN-XO n=3

- Bone marrow-derived macrophages (BMDMs) were derived from XY and XO mice, with XO macrophages used as a model of Y-chromosome loss.
- Senescence was induced by 10 Gy irradiation; cells were analyzed 10 days post-irradiation
- 12 biological samples underwent paired-end 150-bp bulk RNA-seq

$(XO_SEN-XO_NS) - (XY_SEN-XY_NS)$
Tests whether the RESPONSE differs by genotype.

- Baseline genotype effect:** NS-XO vs NS-XY | What differs before senescence?
- Senescence response:** SEN vs NS within each genotype | How does each genotype respond?
- Genotype $\times$ senescence interaction:** Tests whether the senescence response differs by genotype.

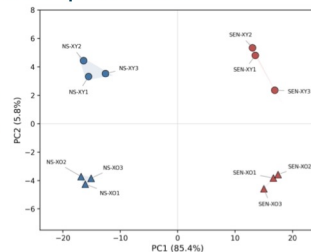
RNA-seq Analysis Workflow



Reference: mm10/GRCm38; annotation: GENCODE M25 | Direct DEG threshold: padj<0.05, |log2FC|>0.5

Results 1

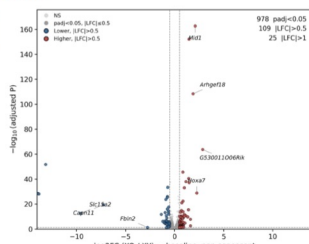
Senescence dominates global transcriptional variation



PC1 (85.4%) separates NS from SEN; PC2 (5.8%) separates XY from XO. Senescence is the dominant source of transcriptional variation, with genotype contributing a smaller secondary pattern.

Results 2

LOY alters the baseline transcriptional state



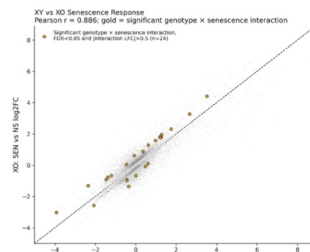
Before senescence, XO macrophages already exhibit a distinct transcriptional state. These differences extend beyond the expected loss of Y-linked transcripts.

XO-higher	XY-higher
Ribosome	Cholesterol / steroid
Cell cycle	Fatty-acid metabolism
DNA replication	Sphingolipid metabolism
ECM / focal adhesion	PPAR / AMPK signaling

LOY-associated differences precede senescence induction.

Results 3

The senescence response is largely shared



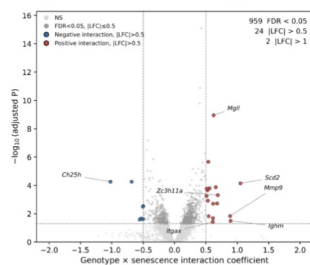
Despite baseline differences, senescence-associated transcriptional changes are highly concordant between genotypes. Genotype primarily modifies response magnitude within a broadly shared program.

Shared DEGs: 1,391 up | 1,136 down | 0 opposite-direction

Where does LOY change the trajectory?

Results 4

Genotype-dependent effects are widespread but modest in magnitude



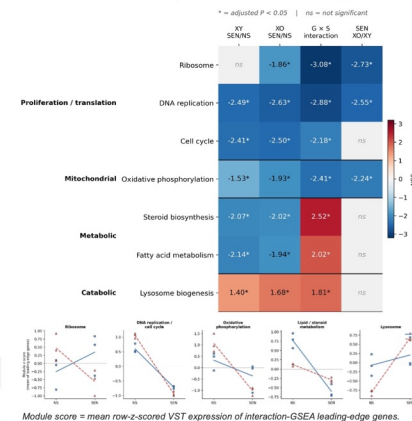
- Negative: XO changes more negatively.
- Positive: XO changes more positively.

Positive interaction does not automatically mean pathway activation.

Most effects are individually modest, motivating pathway-level analysis.

Results 5

LOY reshapes the trajectory of key pathways



XO shows stronger suppression of translational, DNA-replication, and OXPHOS programs, together with relative preservation of lipid/steroid programs and stronger lysosomal induction.

Conclusion

Y-chromosome status alters the macrophage baseline state and quantitatively reshapes a broadly shared senescence-associated transcriptional response.

- XO macrophages enter senescence from a distinct baseline transcriptional state while retaining a highly concordant global response with XY macrophages.
- Genotype-dependent differences are concentrated in the magnitude of pathway remodeling, particularly translation, DNA replication, oxidative phosphorylation, lipid/steroid metabolism, and lysosomal programs.

Next steps: Validate key mitochondrial, lipid-metabolic, and lysosomal phenotypes, and identify and test candidate molecular regulators underlying the genotype-dependent transcriptional response.

Acknowledgements

Acknowledgements: I sincerely thank Dr. Utkarsh Tripathi and Dr. Anthony J. Covarrubias for their mentorship, guidance, and support throughout this project. I am also grateful to Dr. Alexander Hoffmann for hosting the BIG Summer Program and to Caroline for her coordination throughout.

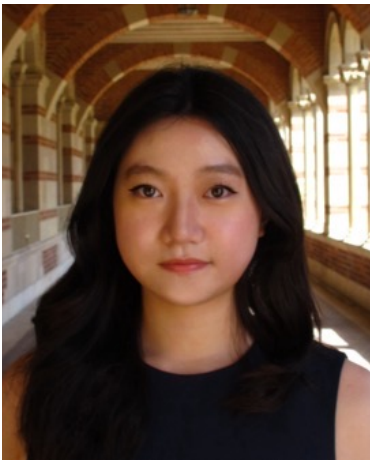
Student: Rongying (Flicka) Zhang

Faculty Advisor: Xia Yang

Direct Mentor: Monty Blencowe

Department: Integrative Biology and Physiology

<https://yanglab.ibp.ucla.edu/>



**Rongying
(Flicka) Zhang**

**Computational
Biology - UCLA**

**Expected year
of graduation
June 2027**

Cross-Context Multiomics Network Analysis Identifies Risk Modules and Key Drivers in Colorectal Cancer

RONGYING (FLICKA) ZHANG^{1,2,3}, Montgomery Blencowe¹, Yang Xia¹

¹ Department of Integrative Biology and Physiology, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

² Department of Computational and Systems Biology, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

³ BIG Summer Program, Institute for Quantitative and Computational Biosciences, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

Colorectal cancer (CRC) involves diverse genetic and epigenetic risks, but whether they converge or diverge on regulatory networks remains unclear. We aim to use a multiomics informed network framework to identify regulatory pathways, networks, and driver genes of CRC. We used the Mergeomics method to integrate genome-wide association studies (GWAS; common variants), whole exome studies (WES; rare variants), and epigenome-wide association studies (EWAS) datasets with pathway, gene coexpression and gene regulatory networks across non-CRC colon, normal adjacent tissues (NAT), and tumors. We found that CRC-associated genetic and epigenetic signals converged on stage-specific programs, with non-CRC and NAT tissues informed immune, epithelial, stromal-vascular, cytoskeletal, and vesicle-trafficking pathways whereas tumors mainly informed cell cycle and metabolic reprogramming. We identified network drivers, many of which were known CRC somatic-mutation genes, linking susceptibility networks to tumor-acquired genomic disruption. Together, the results provide mechanistic insights into the interconnected networks driving CRC progression.



Cross-Context Multi-Omics Network Analysis Identifies Risk Modules and Key Drivers in Colorectal Cancer

RONGYING (FLICKA) ZHANG², Montgomery Blencowe¹, Yang Xia¹

Department of Integrative Biology and Physiology, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

2 BIG Summer Program, Institute for Quantitative and Computational Biosciences, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

UCLA College | Life Sciences
Computational & Systems Biology



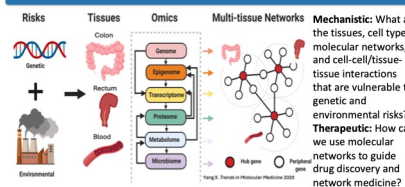
YANG LAB



ABSTRACT

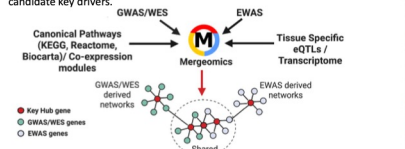
- Colorectal cancer (CRC) involves diverse genetic and epigenetic risks, but whether they converge or diverge on regulatory networks remains unclear. We aim to use a multilomics informed network framework to identify regulatory pathways, networks and driver genes of CRC.
- We used a MergeRank method to integrate genome-wide association studies (GWAS; common variants), whole exome studies (WES; rare variants), and epigenome-wide association studies (EWAS) datasets with pathway, gene coexpression and gene regulatory networks across non-CRC colon, normal adjacent tissues (NAT), and tumors. We found that CRC-associated genetic and epigenetic signals converged on stage-specific programs, with the WES, NAT and EWAS datasets identifying distinct programs: stromal-muscular, cytoskeletal, and vesicle-trafficking pathways whereas tumors mainly informed cell cycle and metabolic reprogramming. We identified network drivers, many of which were known CRC somatic-mutation genes, linking susceptibility to tumor progression. We also identified novel mechanistic insights into the interconnected networks driving CRC progression.

BACKGROUND



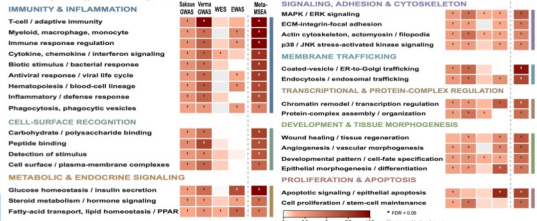
CRC is shaped by genetic, epigenetic, transcriptional, and environmental factors, so no single molecular layer fully explains disease risk or progression. CRC involves progressive tissue remodeling from non-cancer colon to normal-adjacent tissue and tumor.

Single-cell analyses can miss coordinated regulation, because disease signals often act through interconnected pathways and network neighborhoods. A systems-level framework integrates multiomics with coexpression and regulatory networks to identify convergent modules, context-specific rewiring, candidate key drivers.

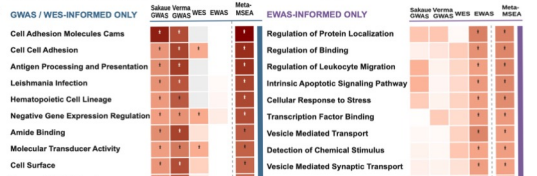


MSEA Results

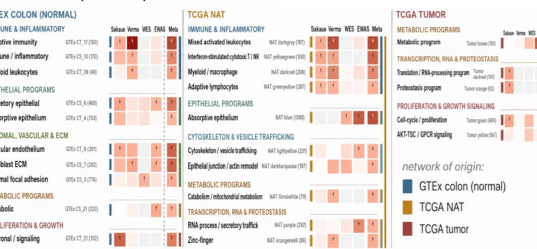
MSEA Top Biological Pathways Enriched



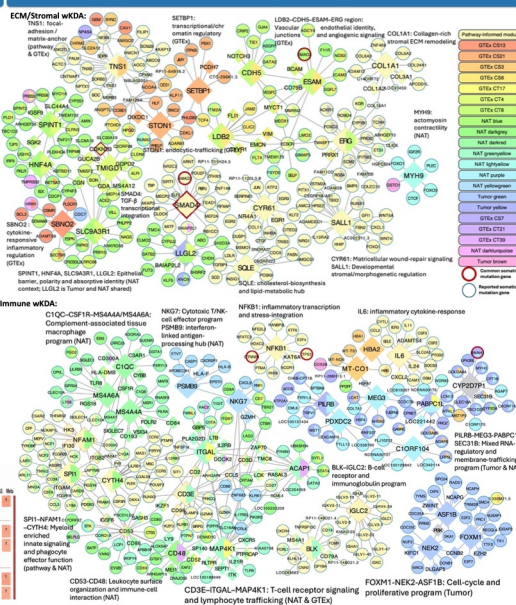
Pathways Showing Bias for Omics Type



MSEA Context-specific Co-expression Network Modules



Context-specific WGCNA and pathway level subnetworks



CONCLUSION AND FUTURE STEPS

- **CRC-associated genetic and epigenetic risks converge on coordinated biological programs rather than isolated genes.** Many omics MSEA consistently highlighted immune, inflammatory epithelial, stromal-vascular, trafficking, metabolic, and growth-related pathways. **These risk-associated programs are strongly context dependent across healthy colon, NAT, and tumor.** Healthy colon and NAT retain broad immune, epithelial, stromal, cytoskeletal, and vesicle-trafficking programs, whereas tumors show a more restricted shift toward proliferation, RNA/proteostasis, and metabolic reprogramming.
- **Different omics layers contribute complementary information to CRC biology.** Genetic signals from GWAS/WES and epigenetic signals from EWAS overlap in some pathways but also preferentially inform distinct biological processes, supporting the value of integrating multiple data types to reveal evidence of biological mechanisms.
- **Network integration identifies candidate key drivers that may coordinate CRC-associated programs.** WkODA identifies hubs linking immune, inflammatory, epithelial, and stromal/ECM neighborhoods, providing mechanistic candidates for how inherited or acquired risk signals may propagate through tissue-specific regulatory networks.

Future Steps:
Leverage UCLA ATLAS genetic data linked to longitudinal EMRs to connect patient genotypes with treatment history and clinical outcomes.
Test drug–variant / pharmacogenomic associations to identify genetic variants associated with differential drug response or toxicity.
Construct polygenic risk scores (PRS) and evaluate whether inherited CRC risk modifies treatment response, disease progression, or patient stratification.
Integrate PharmOmics/network-based drug candidates with EMR validation across UCLA and external cohorts to prioritize clinically actionable therapies.

ACKNOWLEDGEMENTS: I wish to acknowledge the Hu Scholarship, the BIG Summer Program Organizers, the UCLA Computational & Systems Biology department, the Yang Xia Lab, and Rainy Liu for supporting this project.

MATERIALS AND METHODS

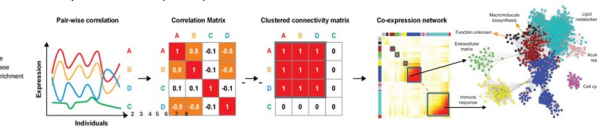
Bayesian Network (BN):

- Constructed context-specific BNs separately for GTEx healthy colon, TCGA NAT, and TCGA tumor using their corresponding bulk transcriptomic expression profiles.
- Gene-gene dependencies were inferred from the expression data to define the network structure.
- Edge direction was informed using external regulatory evidence from eQTL relationships and the ChEA Transcription Factor Targets dataset.
- The three networks were subsequently integrated into a combined cross-context network while retaining edge provenance.

wKDA (Weighted Key Driver Analysis):

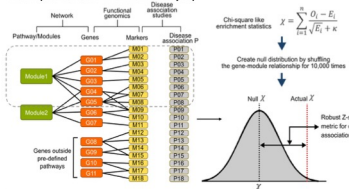
- Project CRC-associated pathways' and WGCNA modules' gene sets onto the context specific Bayesian networks.
- Identify highly connected candidate hub genes within the network. Test each hub's local neighborhood for enrichment of disease-module genes, accounting for network topology and edge weights.
- Prioritize significant hubs as key drivers that may coordinate broader CRC-related programs.

Gene Co-Expression Networks (WGCNA):



Key data resources:
Verma 2024 GWAS: 446,766 European; 120,570 African ancestry; 59,419 Hispanic/Latino.
Sakaue 2021 GWAS: 407,002 European; 167,691 East Asian.
UK Biobank WES: 1,290 CRC cases and 72,940 controls.
EWAS Open Platform CRC study: 279 samples from Germany.
eQTL resources: GTEx v8 colon sigmoid/transverse and TCGA COAD/READ.
Network contexts: GTEx healthy colon, TCGA NAT, and TCGA tumor for Bayesian Networks.

MSEA (Marker Set Enrichment Analysis)

 PharmOmics UCLA Atlas Biobank
paired w/ EMR