Application: Rapid exploration, interpretation, and comparison of discrete basis vectors

Loyal Goff - loyalgoff@jhmi.edu

Collaborative Computational Tools for the Human Cell Atlas- FINAL PROPOSAL

Summary

ID: HCA2-0000000053

Status: In Progress

Applicant Details

Completed - Jan 22 2018 04:38 PM (UTC)

Applicant Details

Please complete all requested information. All fields are required.

General Applicant Details

Applicant First Name	Loyal
Applicant Last Name	Goff
Title/Position	Assistant Professor
Address	733 N. Broadway Avenue, Miller Research Building, Rm 449
City	Baltimore
State/Province	MD
Country	United States
Zipcode	21205
Phone Number	732.648.6062
Email	loyalgoff@jhmi.edu

General Organization Details

Organization Name	Johns Hopkins University
Address	733 N. Broadway Avenue, Suite 117
City	Baltimore, MD
State	MD
Country	USA
Zipcode	21205
Tax ID Number	52-0595110

Organizational/Administrative Contact

List the name and contact information of the administrative contact (not the applicant) to discuss additional information needed for the the recommendation of award.

First Name	Kate
Last Name	Kafka
Title/Position	Grants & Contracts Analyst
Address	733 N. Broadway Avenue, Miller Research Building, Rm 443
City	Baltimore
State	MD
Country	United States
Zipcode	21205
Phone Number	(443) 287-0245
Email	kkafka@jhmi.edu

Signing Official

List the name and contact information for the person authorized to sign on behalf of your organization.

First Name	Laura
Last Name	Tyler
Title/Position	Grants Associate
Address	733 N. Broadway Avenue, Miller Research Building, Suite 117
City	Baltimore
State	MD
Country	United States
Zipcode	21205
Phone Number	4105027186
Email	ltyler5@jhmi.edu

Project Details

Completed - Jan 20 2018 02:28 AM (UTC)

Project Details

Please complete all requested information. All fields are required.

PROJECT TITLE

Rapid exploration, interpretation, and comparison of discrete basis vectors contributing to transcriptional signatures of single cells at the scale of the HCA with ProjectoR

PROJECT PURPOSE

Describe the project in one sentence.

Development and benchmarking of rapid transfer learning methods to interrogate biological and technical basis vectors across experiments at the scale of the human cell atlas.

PROJECT SUMMARY

Limit to 250 words maximum

The transcriptional identity of individual cells is derived from the additive effects of discrete basis vectors that describe dependent and independent contributions to the overall transcriptional state. Many methods exist to deconvolve expression matrices into their constitutive patterns. Most do not scale well to large datasets with complex sources of variation, a defining characteristic of single cell analyses. This limitation necessitates the ability to rapidly explore basis vectors, learned on smaller datasets, across larger datasets, and requires the development of statistical and visualization frameworks upon which to evaluate and compare models derived from different computational approaches.

To address this, we have developed ProjectoR, an R package that uses the relationships defined within a high dimensional data set to rapidly interrogate related phenomena in a new data set using Transfer Learning Methods (TLM). TLMs relax many of the constraints of other methods by using feature mappings that connect the samples and relationships, respectively.

We have established a collaborative network of researchers that aim to resolve and characterize discrete and continuous populations of cells through deconstruction and identification of basis vectors using a variety of approaches. As part of this network, we will establish a curated catalog of benchmark training and test data from the developing mouse and human retina, and evaluate and extend ProjectoR as a model visualization and comparison tool to enable rapid transfer learning of basis vectors across datasets at the scale of the HCA.

TOTAL BUDGET AMOUNT

Enter in US dollars and use integers only; include indirect costs in the total amount.

250000

DELIVERABLES

Succinctly describe the deliverables expected from this research. Examples of deliverables include: software libraries / modules, code repositories, algorithms, curated datasets, cells / sequences / images, and protocols. Please also confirm that you can attend at least 2-3 meetings and hackathons, both in person and remote, with the whole group as well as with smaller subgroups of collaborators working on similar projects.

We will provide the R/Bioconductor software package and code repository for ProjectoR to compare single-cell RNA-sequencing data generated through the HCA project, as well as other single-cell data types (e.g. proteomics, chromatin, and metabolomic data).
We will provide definitive benchmarking utilities for analysis at the scale of the HCA.
Benchmarking ProjectoR using existing matched bulk RNA-Seq, two sources of single-cell RNA-Seq data from the developing retina, we will learn and provide basis vectors from these data, in conjunction with our collaborative network of investigators (Fertig, Greene, Patro). Using these vectors, we will estimate the scalability of ProjectoR for use on the entire human cell atlas dataset and demonstrate the utility of this approach for a) identification of technical artifacts, b) classification of major and minor cell-types, c) characterization of continuous biological processes, and d) basis vector reuse. We will classify basis vectors from training sets made using different library preps.
We will develop support functions within ProjectoR to compare gene-wise basis vectors between experimental paradigms allowing us to identify discrete sources of variation not represented in a projection. Finally, we will develop and provide a statistical framework to extend projections beyond directly mappable features by employing a generalized additive model (GAM).
Finally, we will additionally provide benchmark datasets containing bulk and single cell RNA-Seq and ATAC-Seq using a variety of library preparation techniques across the developing mouse retina at high resolution, and where possible, in de-identified human retinal tissue as well.
As part of this project we will attend at least 3 meetings and hackathons in person, and as many as possible remotely with the whole group as well as subgroups of collaborators.

Project Proposal Part I

Completed - Jan 20 2018 02:31 AM (UTC)

Upload a single PDF document containing the following information. There are no page limits or font/margin requirements.

- Research Plan: You may: use the plan that you submitted previously; alter the scope of the original plan to adjust to a reduced budget; or revise based on discussions with your Science Officer (Jeremy Freeman). Please include literature cited and any figures. Please also mention any specific coordination planned with

your collaborators, either as originally planned or updated based on discussions with your Science Officer as appropriate.

Goff_CZI_Proposal

Filename: Goff_CZI_Proposal.pdf **Size:** 2.9 MB

Project Proposal Part II

Completed - Jan 22 2018 04:46 PM (UTC)

Upload a single PDF containing the following sections. There are no page limits or font/margin requirements.

- **Detailed Budget:** Describe in detail how the budget will be allocated, including personnel, supplies, equipment, travel, subcontracts, other costs, and indirect costs (limited to 15% of direct costs). Names of "to be determined" appointments should be provided, if possible. Detailed budgets should be provided for each subcontractor. Indirect costs may not be assessed on capital equipment or subcontracts, but subcontractors may include up to 15% indirect costs of their direct costs. You may use the budget from your original application, if appropriate.
- **Other Support / Funding Overlap:** Describe the other financial support available to the lab (grant source, number, title, amount). State whether there is overlap with this proposal, and, if so, explain. Please include for all key personnel, including subcontractors.
- **Statement on Sharing / Code and Data Dissemination / Collaboration:** Be specific. We require: a) that code be developed in the open on Github, by creating repositories on Github for each library / module / software component at the start of the project, and coordinating development with collaborators (through Slack, Github Issues, Pull Requests, semantic versioning, etc.), with help and coordination from CZI; b) that full proposals be submitted to a Github repository / wiki that CZI will set up for grantees; c) that any sequence and imaging data collected or curated be deposited in the HCA Data Coordination Platform (details will be provided); d) that any experimental protocols be deposited in a publicly available form (such as Protocols.io); e) that manuscripts be submitted to a preprint server (such as BioRxiv) at the time of submission for publication; and f) that publications cite the grant number or equivalent, as specified in the grant agreement.
- **Statement on Human Tissue / Animal Welfare:** All animal research must follow your institution's policies on ethical treatment of animals. Human data must be fully consented for research and public access without restriction. Provide the name of the institution official that can confirm your statement.

Goff_Project_proposal_part_II_v2

Filename: Goff_Project_proposal_part_II_v2.pdf **Size:** 106.6 kB

Signed Recommendation Administration Form

Completed - Jan 20 2018 02:29 AM (UTC)

Upload this form signed by a person authorized to sign on behalf of your organization, endorsing and verifying your application materials submitted in the grants portal.

Goff108618-signed

Filename: Goff108618-signed.pdf **Size:** 81.0 kB

Summary

Single-cell analysis has demonstrated that population-level gene expression and the 'transcriptional identity' of individual cells, arises from combinations of basis vectors[1]. Reuse and exaptation of co-regulated modules of genes or other cellular features can contribute to diverse phenomena as patterning, tissue organization, cellular physiology, and paralogous functions in disparate tissues. The extent to which basis vectors are shared/reused throughout the human body remains under-explored. Exploring these features at single-cell resolution provides an opportunity to identify and characterize the reuse of co-regulated features.

While many methods exist to deconvolve gene expression into patterns, most methods do not scale to large datasets with complex sources of variation. Further, basis vector identification and evaluation of models is limited to technical metrics with little consideration for the common or disparate biological properties described by each approach. Tools are needed to benchmark the biological activity described by models derived from independent algorithms. Current computational limitations necessitate the ability to rapidly explore basis vectors learned on smaller datasets across larger datasets, and requires the development of statistical and visualization frameworks upon which to evaluate and compare learned models derived from different computational approaches.

Transfer learning methods (TLMs) use previously learned knowledge from one or more sources to improve learning of a new target data. TLMs are able to relax many of the constraints of other methods by using the fact that if two domains are related, there may exist mappings or features that connect the samples [2]. We implemented TLM methodologies to perform integrated analysis of high dimensional multi-omic data in the R package ProjectoR. ProjectoR uses relationships defined within a given data set, to interrogate related biological phenomena in an new data set. Importantly, ProjectoR is agnostic to the source or type of basis vectors (e.g. principal components, metagenes, modules, latent spaces, etc). Instead ProjectoR uses the weights of learned vectors across features from one dataset to establish a feature representation on a target dataset. In this manner, basis vectors corresponding to meaningful biological variation can be compared directly, independent of laboratory of origin or technical artifacts. Projection of artefactual basis vectors, corresponding to technical sources of error in the test dataset, result in little to no information content when projected into the target set. Conversely, biological basis vectors stratify samples consistent with their underlying biological processes. Furthermore, basis vectors learned by independent methods on disparate training sets can be projected into a common test dataset and directly compared. We propose to adapt these TLMs to enable rapid comparisons of multiple data types, bulk and single cell library preparation techniques, developmental time, sex, cell types, and even across species in a well characterized model system that provides an ideal setting to compare. Additionally as part of an open collaborative network, we propose to develop and extend ProjectoR as a statistical framework to evaluate and compare basis vectors learned from disparate algorithms.

Project aims, and how they address program goals

Our overall aim for this project is to provide a statistical framework and benchmark datasets to compare methods for identification of basis vectors from single cell data, and determine the biological or artefactual relevance of these vectors using a well-studied mammalian system. We will expand ProjectoR to explore pattern use/reuse at the scale of the HCA with minimal computational effort. Specifically, we will address several of the project goals by 1) developing tools to extract and analyze data organized by genes, cells, or tissues of interest, 2) supporting analytical methods and machine learning approaches, 3) generating curated benchmark datasets, and experimental datasets that directly address computationally-guided questions in quality control, reproducibility, or multimodal integration, and 4) developing new computational approaches comparing and normalizing genomic and imaging data across assays, subjects, and species.

Aim I: *To extend existing benchmark single cell RNA-Seq datasets of the developing retina across library techniques, developmental stages, and species.* Thus, allowing for both discrete cell type identification at

multiple hierarchical levels, as well as continuous properties such as pseudotemporal state, pseudo-spatial state, differentiation state and progenitor competency.

A) We will extend our current catalog of mouse retina single cell data to include sci-RNA-Seq[3] and sci-nucRNA-Seq in mouse developing retina.
B) We will conduct bulk RNA-Seq and single cell sci-RNA-Seq in de-identified human postmortem tissue.
C) We will evaluate the potential for cross-species basis vector comparison using ProjectoR to project developing mouse retina basis vectors into publicly available and newly generated human single-cell RNA-seq and ATAC-seq datasets from retina and other developmentally related tissues.
    a) We will determine whether basis vectors learned in mouse retina, can identify human cells undergoing similar biological processes.

Aim II: *To benchmark ProjectoR using basis vectors (models) from developing mouse and human retina learned from tools across collaborative network*

A) We will evaluate ProjectoR performance on output from various collaborative network models
    a) Benchmark computation speed on projections in exponentially scaled data sets with corresponding increases in dimensionality and complexity.
    b) Using benchmark datasets and *a priori* knowledge we will access accuracy of biological assignments of basis using metrics of sensitivity and specificity of projections to evaluate statistical power.
B) We will identify technical vs biological models and determine methods to QC individual cells via projection of models.

Aim III: *To develop model comparison statistics, pathway enrichment testing, and novel basis vector visualizations in ProjectoR*

A) We will compare collaborative network analytical tools using ProjectoR projections on benchmark datasets to highlight optimal usage for specific biological questions.
    a) We will develop a statistical framework to test for discriminatory power for major cell types or lineages by a given pattern, and develop tools to identify technical artifact patterns.
B) We will develop ProjectoR visualizations to explore shared biological features across benchmark datasets, as well as public and published single cell RNA-Seq datasets to create comprehensive model via projection (e.g. PCA as example).

Prior contributions in this area and preliminary results

Age	Replicate	Sorting	Condition	Library Type	nCells
E11	1	-	Whole Retina	10x Genomics	9,366
E12	2	-	Whole Retina	10x Genomics	26,576
E14	1	-	Whole Retina	10x Genomics	16,154
E14	3	Chx10-GFP(+)	RPCs	Smart-Seq2	244
E14	2	Chx10-GFP(+)	RPCs	Bulk-RNAseq	-
E14	2	Chx10-GFP(-)	Post-mitotic Cells	Bulk-RNAseq	-
E16	1	-	Whole Retina	10x Genomics	5,401
E18	1	-	Whole Retina	10x Genomics	10,574
E18	3	Chx10-GFP(+)	RPCs	Smart-Seq2	286
E18	2	Chx10-GFP(+)	RPCs	Bulk-RNAseq	-
E18	2	Chx10-GFP(-)	Post-mitotic Cells	Bulk-RNAseq	-
P0	1	-	Whole Retina	10x Genomics	9,926
P2	2	-	Whole Retina	10x Genomics	17,968
P2	3	Chx10-GFP(+)	RPCs	Smart-Seq2	280
P2	2	Chx10-GFP(+)	RPCs	Bulk-RNAseq	-
P2	2	Chx10-GFP(-)	Post-mitotic Cells	Bulk-RNAseq	-
P5	1	-	Whole Retina	10x Genomics	6,756
P8	2	-	Whole Retina	10x Genomics	11,729
P14	1	-	Whole Retina	10x Genomics	8,103
				TOTAL	123,363

Table 1: Summary table of preliminary developing mouse retina gene expression data using three distinct library preparations (bulk, Smart-Seq2, and 10x genomics). These data will be used to derive basis vectors using tools developed by members of our collaborative network. We will evaluate the projection of these learned basis vectors across the different data sets to benchmark ProjectoR methodologies and develop novel visualizations and statistical tests for model comparison.

We developed the ProjectoR package to enable rapid transfer learning of basis vectors from one biological dataset into another. On top of the generic function, the base projectoR function is currently coded for regression, PCA, CoGAPS, and clustering-derived basis vectors. The ProjectoR package currently contains statistics to evaluate cluster/set overlap (intersectoR), methods to transform learned vectors to more accurately reflect biology (rotatoR), evaluate correlation (correlateR), and standardize gene composition/feature mapping across datasets (geneMatchR).

We have in parallel established a compendium of RNA-Seq libraries (bulk and

single-cell) from dissociated mouse retina at various time points using different library preparations (Table 1). Leveraging our existing catalog of bulk and single-cell RNA-Seq data from the developing mouse retina (Table 1; described below), we identified 18 basis vectors corresponding to both biologically meaningful patterns of gene expression, batch-specific effects, and sources of technical variation (Figure 1). Gene-weights were projected into 10x single-cell RNA-Seq. Projection of >120,000 cells took 55.5 seconds. Using these projections, we identified lineage-specific patterns of gene expression (Figure 1B-C), as well as patterns both correlated and anti-correlated with developmental age (Figure 1D-E). Importantly, modules corresponding to technical artifacts demonstrated no enrichment across cells.

Proposed work and deliverables

We will develop and extend ProjectoR to compare single-cell RNA-sequencing data generated through the HCA project, as well as other single-cell data types (e.g. proteomics, chromatin, and metabolomic data). We will provide definitive benchmarking utilities for analysis at the scale of the HCA.

Benchmarking ProjectoR using existing matched bulk RNA-Seq, two sources of single-cell RNA-Seq data from the developing retina, we will learn basis vectors from these data, in conjunction with our collaborative network of investigators (Fertig, Greene, Patro) [4, 5, 6]. Using these vectors, we will estimate the scalability of ProjectoR for use on the entire human cell atlas dataset and demonstrate the utility of this approach for a) identification of technical artifacts, b) classification of major and minor cell-types, c) characterization of continuous biological processes, and d) basis vector reuse. We will classify basis vectors from training sets made using different library preps to:

- Identify vectors corresponding to technique-specific artefacts
- Identify vectors corresponding to quality of input data
    - Doublets -> incompatible vectors in same cell
    - Low quality vectors (correlated with metrics of poor scRNA-Seq performance)

We will then use and develop support functions within ProjectoR to compare gene-wise basis vectors between experimental paradigms allowing us to identify discrete sources of variation not represented. Finally, we will develop a statistical approach to extend projections *beyond directly mappable features* by employing a generalized additive model (GAM) in which pattern weights, projected using only a subset of mapped features, are used as explanatory variables to identify other features whose expression may be dependent.

Proposal for evaluation and dissemination of methods, resources, or results.

All primary and benchmark data will be deposited with the HCA data coordination platform, as well as freely accessible online, and archived on the NCBI short read archive along with all available metadata. Learned basis vectors, workflows, and analyses will be publicly available on a custom web site, archived in conjunction with published manuscripts, and posted on Biorxiv. Stable software builds will be released in the ProjectoR package.

Statement of commitment to share proposals, methods, data, and code

We commit to making all scripts, code, and software available via github under the permissive MIT license consistent with the standard operating procedures of the Goff Lab. The proposal has been developed in the open in conjunction with the collaborative network and is publicly shared at https://github.com/gofflab/czi-rfa-2017/. We acknowledge that the proposal will be made public. Proposed work in benchmark data annotation, model identification and characterization, and analysis will be conducted on a publicly available github repository as they are generated.

Figures

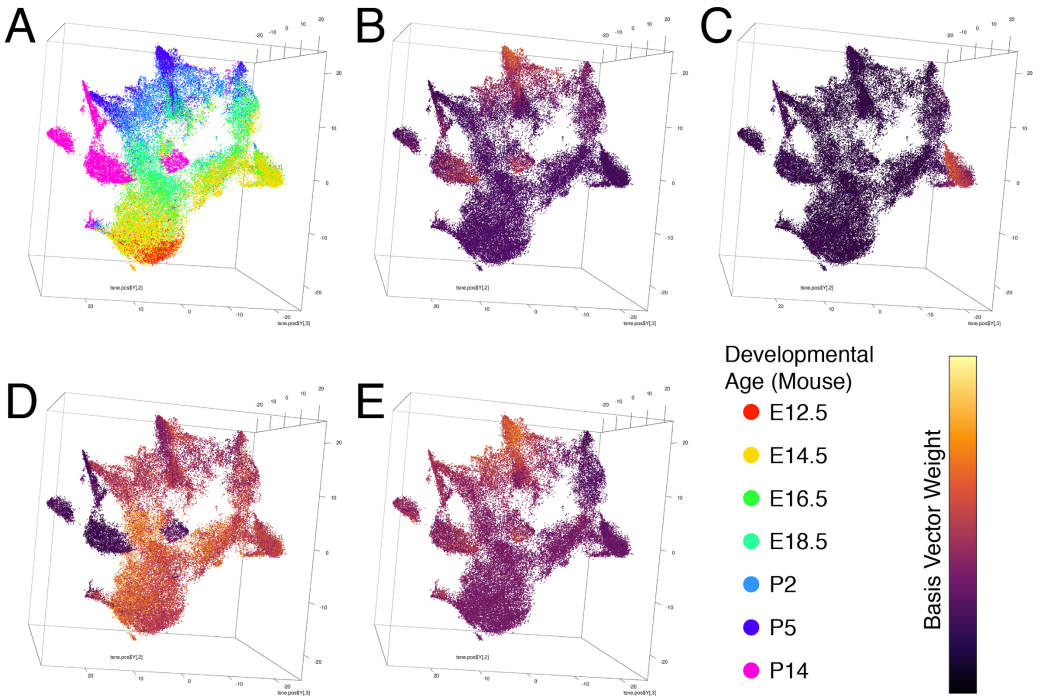

Figure 1: 3D t-stochastic neighbor embedding (tSNE) plot of 54,463 dissociated cells from select timepoints from developing mouse retina showing lineage commitment and maturation of retinal progenitor cells. (A) Cells are colored by developmental age at collection. B-E) cells are colored based on ProjectoR projections of select basis vectors learned from an independent bulk RNA-Seq study of mouse retina development. ProjectoR is able to identify cells with lineage-specific basis vector usage (B & C) as well as basis vectors describing both decreasing usage (D) and increasing usage (E) over developmental time, consistent with

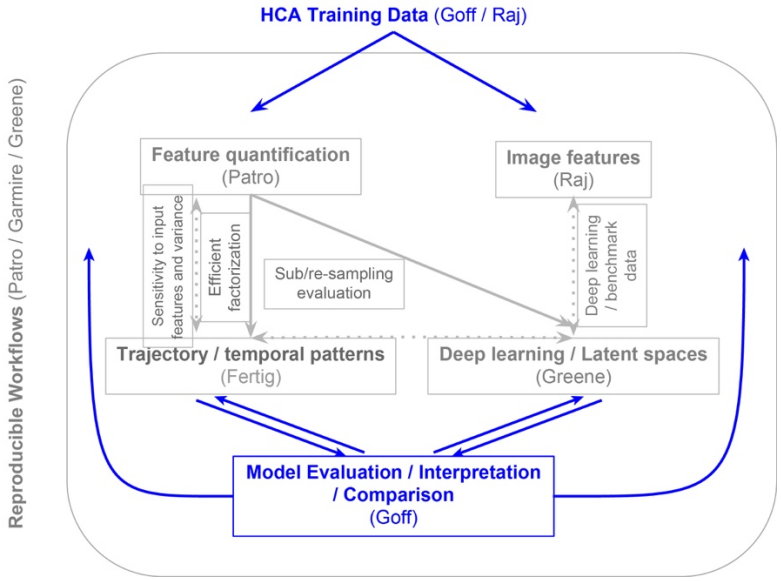

Figure 2: Proposed collaborative network. Contributions from this proposal are highlighted in blue.

References cited

1. Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34,** 1145–1160 (2016).
2. Pan, D. An integrative framework for continuous knowledge discovery. *Journal of Convergence Information Technology* (2010).
3. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357,** 661–667 (2017).
4. Fertig, E. J., Ding, J., Favorov, A. V., Parmigiani, G. & Ochs, M. F. CoGAPS: an R/C++ package to identify patterns and biological process activity in transcriptomic data. *Bioinformatics* **26,** 2792–2793 (2010).
5. Fertig, E. J. *et al.* Gene expression signatures modulated by epidermal growth factor receptor activation and their relationship to cetuximab resistance in head and neck squamous cell carcinoma. *BMC Genomics* **13,** 160 (2012).
6. Stein-O'Brien, G. *et al.* PatternMarkers and Genome-Wide CoGAPS Analysis in Parallel Sets (GWCoGAPS) for data-driven detection of novel biomarkers via whole transcriptome Non-negative matrix factorization (NMF). *bioRxiv* doi:10.1101–083717 (2016). doi:10.1101/083717

Rapid exploration, interpretation, and comparison of discrete basis vectors contributing to transcriptional signatures of single cells at the scale of the HCA with ProjectoR								
Personnel:		**Role**	**Base**	**Cal. Mos.**	**% Effort**	**Salary Req.**	**Fringe**	**YR1 Total**
Loyal Goff		PI	$ 132,651	1.80	15%	$19,898	$6,765	*$ 26,663*
Seth Blackshaw		Collaborator	$ -	0.00	0%	$0	$0	*$ -*
Brian Clark		Research Associate	$ -	0.00	0%	$0	$0	*$ -*
Genevieve Stein-O'Brien		Postdoctoral Fellow	$ 47,853	6.00	50%	$23,927	$4,618	*$ 28,544*
TBN		Graduate Student	$ 31,936	9.00	75%	$23,952	$4,400	*$ 28,352*
Briana Winer		Research Technologist	$ 34,000	6.00	50%	$17,000	$5,780	*$ 22,780*
Subtotal						*$84,776*	*$21,563*	**$ 106,339**
Expenses:								
Supplies & Materials:								
10x V2 library kit (16 samples per kit; 2X $21,000 each)								$ 42,000
Chips (6*8 samples/chip, 2x $1,440)								$ 2,880
Barcode Index kit (96 samples, 1x $805)								$ 805
Nextera XT Library Prep Kits								$ 2,000
Reagents for enzymatic dissociation of single cells and seq. library preparation								$ 3,000
Computational storage space and Rstudio server (Amazon web services)								*$ 6,000*
Laboratory Consumables (pipette tips, plasticware, etc)								*$ 4,867*
RNAScope probes for model validations								*$ 4,000*
Services								
Illumina 2500 Sequencing lanes (20*$2,000)								*$ 40,000*
Computers								
Laptop computer								*$ 3,000*
Travel to HCA meetings								
Travel expenses								*$ 2,500*
Subtotal								**$ 111,052**
Total Direct Costs, Year 1:								**$ 217,391**
						Facilities & Administration Costs	*15%*	**$ 32,608.67**
TOTAL COSTS, YEAR 1								**$ 250,000**

• Loyal A. Goff, Ph.D. (PI) will be responsible for overall project design and coordination, direction on ProjectoR development and collaborative efforts on single cell collection and biological interpretation (Goff, Blackshaw).

• Seth Blackshaw, Ph.D. (Collaborator) Dr. Blackshaw is a renowned expert in the biology of mammalian retinal development. Drs. Goff, Blackshaw, and collaborative network member Dr. Fertig have an existing collaboration built around a detailed characterization of retinal cell developmental biology that has contributed much of the preliminary data for this proposed project. Dr. Blackshaw will continue to provide biological interpretations to learned basis vectors and will provide necessary resources for validations.

• Brian Clark, Ph.D. (Research Associate) Dr. Clark is a research associate (K99) jointly mentored in the labs of the PI (Goff) and collaborator (Blackshaw). He has years of experience in studying mouse retina development and has extensive experience working with RNA. He will be responsible for tissue acquisition, processing, and will continue to generate the bulk and single cell libraries in mouse and human retina. Dr. Clark will also be primarily responsible for validation of learned basis vectors through in situ fluorescence hybridization analysis.

• Genevieve Stein-O'Brien, Ph.D. (Postdoctoral Fellow) is the chief developer and maintainer of the ProjectoR package and will be responsible for implementing the transfer learning statistics into the software package. Dr. Stein-O'Brien is a postdoctoral fellow co-supervised by Dr. Goff and collaborative network member Dr. Fertig. She is listed in both proposals. In this proposal, Dr. Stein-O'Brien is listed as responsible for the algorithm development and analyses proposed in this award. Her work will be completed in collaboration with all key personnel on this proposal and co-supervised by Dr. Goff and Dr. Fertig. If both awards are funded, a TBN postdoc will be hired to collaborate with Dr. Stein-O'Brien on these efforts.

• TBN (Graduate Student) in conjunction with the research technologist will perform the required single cell library preparations and metadata aggregation and will be primarily responsible for generating and processing the human retinal single cell benchmark data.

• Briana Winer (Research Technologist). In conjunction with the TBN graduate student, Briana will be responsible for performing the additional 10x genomics and sci-RNA-Seq datasets under the direct supervision of the PI, as well as aggregation and organization of all associated metadata.

Other Support

GOFF, L.A.

ACTIVE:

IOS-1665692 (Brown/Goff) 03/01/2017 – 02/28/2021 1.20 Calendar
National Science Foundation $225,000
Cell type specific gene expression differences induced by experience-dependent plasticity

2016-MSCRFI-2805 (Goff/Bjornsson) 06/01/2016 – 05/31/2019 2.40 Calendar
Maryland Stem Cell Research Commission $200,000
Single cell analysis of hippocampal neurogenesis defects in Kabuki Syndrome 1

SLI (Goff/Brown) 06/01/2016 – 05/30/2018 1.20 Calendar
Johns Hopkins Science of Learning Institute $100,000
Cell-type specific heterogeneity in experience-induced gene expression

Target ALS (Goff – Co-PI) 05/01/2017 – 04/30/2018 1.80 Calendar
TargetALS Foundation $120,000
Cellular Mechanisms of Cortical Hyperexcitability
This project will explore the cell-type-specific effects of familial ALS mutations on hyperexcitability of cortical neurons, and the common and distinct gene expression changes that evoke this phenotype in ALS mouse models.

Synergy Award (Goff) 07/01/2017 – 06/31/2018 1.20 Calendar
Johns Hopkins School of Medicine Discovery Fund $100,000
Systematic characterization of transcriptional variation in retinal development at single cell resolution

PENDING:

SPARC (Chakravarti/Goff) 1/1/2018 - 12/31/2020 2.40 Calendar
NIH Common Fund (SPARC OT2) $310,000
Comprehensive mapping and characterization of the intrinsic and extrinsic connection matrix of the enteric Nervous system.

NSF (Fertig) 07/01/2018-06/30/2022 1.20 Calendar
National Science Foundation $941,508
Scalable Methods for Smooth-spare non-negative Matrix Factorization in Genomics
This project is to develop efficient algorithms for pattern detection in genomics with smooth-sparse matrix factorization.

1R21AI139358-01 (Potter) 07/01/2018-06/30/2020 .6 Calendar
NIH/NIAID $275,000
Identification and characterization of mosquito sensory neurons detecting human-related cues
Mosquitoes utilize a diverse array of senses to navigate and identify human hosts for biting, and a better understanding of these sensory systems could lead to new effective methods to control mosquito populations.

OVERLAP: None

Statement on Sharing / Code and Data Dissemination / Collaboration

We commit to making all scripts, code, and software available via github under the permissive MIT license consistent with the standard operating procedures of the Goff Lab. Each software package and analysis will have an independent, open-source repository for coordinating development with collaborators. Proposed work in benchmark data annotation, model identification and characterization, and analysis will also be conducted on a publicly available github repository as they are generated. Additional communication/coordination will be through Slack, Github Issues, and Pull Requests as required. The proposal has been developed in the open in conjunction with the proposed collaborative network and is currently publicly shared at https://github.com/gofflab/czi-rfa-2017/. Additional collaboration in the development of this joint proposal was conducted on Slack. We acknowledge that the final proposal will be made public and submitted to a specific Github repository / wiki established by CZI. All primary and benchmark data will be deposited with the HCA data coordination platform, as well as freely accessible online, and archived on the NCBI short read archive along with all available metadata. Learned basis vectors, workflows, and analyses will be publicly available on a custom web site, archived in conjunction with published manuscripts, and posted to Biorxiv. Stable software builds will be released in the ProjectoR package through the Bioconductor project. Any experimental protocols developed will be depositied with Protocols.io. Preprints derived from this work will be made available through BioRxiv at the time of submission for publication and any published works will cite the CZI grant number for this proposed project as requested.

Human Tissue / Vertebrate Animals:

Use of human tissue:

All human tissue used will be obtained from sources that have broadly consented donors for use in this proposed study and were collected for uses other than this specific study. All proposed activities meet the criteria for human subjects exemption, as defined by the National Institutes of Health, as tissue will be de-identified prior to acquisition and all investigators will be blinded to the donors and will not be able to ascertain the identity of the donors. Investigators will not have access to personally identifiable information and will not generate additional identifiable genomic data.

Detailed description of the proposed use of animals:

Both male and female C57BL/6J or CD-1 IGS wildtype mice, ages E11 to P30, will be used for the proposed experiments to provide retinal tissue for the single cell analysis. These studies will require approximately 20 pregnant females and 30 male and female mice (newborn to adult). Experiments will be carried out on excised, and enzymatically dissociated retinal tissue. The tissue will be isolated and immediately processed as required for each specific assay. Prior to sacrifice, animals will be housed in the Johns Hopkins American Association of Laboratory Animal Care (AALAC) accredited animal facility.

Justification for the use of animals, choice of species and numbers of animals to be used: The major goal of this research is to describe the developmental organization of the mammalian retina across time and space using single cell measurements. Because of fundamental gaps in our knowledge of retina development, no effective computational models have been developed that could replace the experimentation with animals proposed here. Nor can the precise and complex interplay of activity in intact tissue organization be mimicked in cell culture or other *in vitro* systems. Because of the increasing sophistication of the genetic tools available in the mouse, mice have become a standard model organism for studying the function of neural circuits. Therefore, all the proposed experiments will be performed in wild type mice. The use of experimental animals has been carefully considered in the design of this proposal. We are restricting our investigations *in vivo* to questions where an *in vivo* investigation is critical. The number of animals assigned to each experiments represents the minimum that we expect to make possible statistical analysis and thus interpretation of results. This number takes into account that some animals may be needed for unanticipated validation experiments. Based on experience, we computed approximately 10% more mice than those necessary for statistical analysis. Every effort will be made to minimize the number of animals used for the described experiments.

Veterinary care: Mice will be housed in a barrier facility at Johns Hopkins (BRB), an AALAC accredited facility. All procedures will be performed in the BRB according to governmental and NIH requirements. Veterinary care is provided by the Research Animal Resources (RAR) at Johns Hopkins. JHU maintains the expectation that all veterinary care for JHU research animals will be provided by RAR veterinarians or with RAR veterinary guidance to ensure quality and contemporary standards of care. Veterinary care is available 24/7 via routine rounds and a rotating on-call schedule. Veterinary care is provided by 4 faculty veterinarians, and 6 veterinary fellows. Surgical and technical support is provided by RAR rodent technical specialists. Daily husbandry procedures are provided by trained animal technicians. The veterinarians work directly with the animal care staff on programs designed to reduce the prevalence of infectious disease, the monitoring of animal health, and the diagnosis and treatment of illness and disease. Veterinary staff reserves the right to intervene in all cases in which animals are experiencing unalleviated pain or distress that has not been justified in the protocol as necessary to accomplish scientific objectives and for which provisions for palliative care have not been provided. An animal protocol, which is reviewed annually, has been approved for these studies by the Institutional Animal Care and Use Committee at Johns Hopkins

Procedure to minimize pain and discomfort: Every effort will be made to minimize any discomfort or distress to the animals being used. The surgical procedures are relatively straightforward and will be performed after extensive training and under supervision by experienced laboratory personnel and RAR veterinary staff. Mice will be carefully monitored for signs of distress as they recover from anesthesia and in the days following the surgery. If there is any sign of distress, animals will be euthanized immediately.

Euthanasia: Animals are inspected daily by the veterinary care support staff. Any moribund animals are noted and treated as appropriate (hydrogel addition, saline injection) and monitored closely over the next 24 hours. Those animals not responding to treatment will be euthanized. Euthanasia in all instances will be terminal inhalation of carbon dioxide or isoflurane vapor, followed by secondary euthanasia through cervical dislocation or decapitation. Methods of euthanasia are consistent with the American Veterinary Medical Association (AVMA) Guidelines on Euthanasia, and in accordance with protocols approved by the Animal Care and Use Committee at Johns Hopkins University.

RECOMMENDATION ADMINISTRATION FORM

Please provide the following information using the Chan Zuckerberg Initiative online grants administration portal (http://apply.chanzuckerberg.com/). Please also have this form signed by an official authorized to sign on your organization's behalf, endorsing this application and verifying that the information entered/uploaded into the grants portal is accurate. This signed form must be uploaded into the grant portal for your final application to be considered complete. Please contact Nina Cardoza, Science Grants Manager, with any administrative questions (nina@chanzuckerberg.com) or Jeremy Freeman, Director of Computational Biology, with any scientific questions (jeremy@chanzuckerberg.com).

Please enter the following information where indicated in the grants portal:

Applicant Details:

- **General**: Applicant name, title, address, phone number, email

Organization Details:

- **General:** Organization name, address, tax ID number
- **Organizational/Administrative Contact:** Name, title, address, phone, email for administrative contact (not the applicant) to discuss additional information needed for the the recommendation of award
- **Signing Official:** Name, title, address, phone, email for the person authorized to sign on behalf of your organization and who will be signing this form

Project Details:

- **Project Title**
- **Project Purpose** (1 sentence)
- **Project Summary** (250 words maximum): You may use the abstract from your original application, if appropriate.
- **Total budget amount** in US dollars
- **Deliverables:** Succinctly describe the deliverables expected from this research. Examples of deliverables include: software libraries / modules, code repositories, algorithms, curated datasets, cells / sequences / images, and protocols. Please also confirm that you can attend at least 2-3 meetings and hackathons, both in person and remote, with the whole group as well as with smaller subgroups of collaborators working on similar projects.

Please upload the following documents where indicated in the grants portal:

Project Proposal Part I: Please upload a single PDF document containing the following information. There are no page limits or font/margin requirements.

- **Research Plan:** You may: use the plan that you submitted previously; alter the scope of the original plan to adjust to a reduced budget; or revise based on discussions with your Science Officer (Jeremy Freeman). Please include literature cited and any figures. Please also mention any specific coordination planned with your collaborators, either as originally planned or updated based on discussions with your Science Officer as appropriate.

Project Proposal Part II: Please upload a single PDF containing the following sections. There are no page limits or font/margin requirements.

- **Detailed Budget:** Describe in detail how the budget will be allocated, including personnel, supplies, equipment, travel, subcontracts, other costs, and indirect costs (limited to 15% of direct costs). Names of "to be determined" appointments should be provided, if possible. Detailed budgets should be provided for each subcontractor. Indirect costs may not be assessed on capital equipment or subcontracts, but subcontractors may include up to 15% indirect costs of their direct costs. You may use the budget from your original application, if appropriate.

- **Other Support / Funding Overlap:** Describe the other financial support available to the lab (grant source, number, title, amount). State whether there is overlap with this proposal, and, if so, explain. Please include for all key personnel, including subcontractors.

- **Statement on Sharing / Code and Data Dissemination / Collaboration:** Be specific. We require: a) that code be developed in the open on Github, by creating repositories on Github for each library / module / software component at the start of the project, and coordinating development with collaborators (through Slack, Github Issues, Pull Requests, semantic versioning, etc.), with help and coordination from CZI; b) that full proposals be submitted to a Github repository / wiki that CZI will set up for grantees; c) that any sequence and imaging data collected or curated be deposited in the HCA Data Coordination Platform (details will be provided); d) that any experimental protocols be deposited in a publicly available form (such as Protocols.io); e) that manuscripts be submitted to a preprint server (such as BioRxiv) at the time of submission for publication; and f) that publications cite the grant number or equivalent, as specified in the grant agreement.

- **Statement on Human Tissue / Animal Welfare:** All animal research must follow your institution's policies on ethical treatment of animals. Human data must be fully consented for research and public access without restriction. Provide the name of the institution official that can confirm your statement.

Signed Recommendation Administration Form: Upload this form signed by a person authorized to sign on behalf of your organization, endorsing and verifying your application materials submitted in the grants portal.

Endorsement and verification of application materials: Executed by a person authorized to sign on behalf of your organization

Name: Laura Tyler

Title: Grants Associate

Signature:  Date 1/19/2018
