Rapid exploration, interpretation, and comparison of discrete basis vectors contributing to transcriptional signatures of single cells at the scale of the HCA wi								
Personnel:	**Role**	**Base**	**Cal. Mos.**	**% Effort**	**Salary Req.**	**Fringe**	**YR1 Total**	
Loyal Goff	PI	$ 130,050	1.80	15%	$19,508	$6,633	*$*	*26,140*
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	12.00	100%	$47,853	$9,236	*$*	*57,089*
TBN	Graduate Student	$ 33,000	12.00	100%	$33,000	$0	*$*	*33,000*
Briana Winer	Research Technologist	$ 34,000	6.00	50%	$17,000	$5,780	*$*	*22,780*
Subtotal					*$117,361*	*$21,648*	**$**	**139,009**
Expenses:								
Supplies & Materials:								
10x V2 library kit (16 samples per kit; 2X $21,000 each)							$	42,000
Chips (6*8 samples/chip, 1x $1,440)							$	1,440
Barcode Index kit (96 samples, 1x $805)							$	805
384 sci-RNA-Seq RT primers							$	6,000
384 sci-RNA-Seq PCR P5 primers							$	6,500
384 sci-RNA-Seq PCR P7 primers							$	2,250
Nextera XT Library Prep Kits							$	6,500
Reagents for enzymatic dissociation of single cells and seq. library preparation							$	5,500
Computational storage space and Rstudio server (Amazon web services)							*$*	*6,500*
Laboratory Consumables (pipette tips, plasticware, etc)							*$*	*5,500*
Adult Human retina tissue samples (4)							*$*	*5,000*
RNAScope probes for model validations							*$*	*3,000*
Services								
Illumina 2500 Sequencing lanes (24*$2,000)							*$*	*48,000*
Computers								
Laptop computer							*$*	*3,500*
Travel to HCA meetings								
Travel expenses							*$*	*2,500*
Subtotal							**$**	**144,995**
Total Direct Costs, Year 1:							**$**	**284,004**
					Facilities & Administration Costs	*15%*	**$**	**42,600.55**
TOTAL COSTS, YEAR 1							**$**	**326,604**

• 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. (Postdoctoral Fellow) Dr. Clark is a postdoctoral fellow 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.

Collaborative Network

Our collaborative network includes several groups responding to this RFA. We shared proposals openly on GitHub and discussed their content via Slack. Selected interactions between groups are summarized in Fig 2 and below. We look forward to these and additional collaborations within this group and the broader HCA consortium.

Model interpretation. Assessing the performance of unsupervised techniques relies critically on interpretation relative to known covariates, gene function, and independent test datasets. Our collaborative network contains several investigators who are proposing new algorithms for unsupervised model learning. By developing ProjectoR as an efficient comparator, we will be able to interpret the function of our inferred transcriptional trajectories relative to gene signatures. Proposed model contributors include **Casey Greene** and Elena Fertig.

Visualization is also critical to such interpretation. Further collaboration with **Lana Garmire**'s group on the development of Granatum will enable automated visualization of unsupervised patterns. We look forward to continued interactions with the consortia to optimize unsupervised model interpretation.

Efficient factorization methods. Several proposals include techniques for efficient factorization methods. We have an existing collaboration with **Elena Fertig** to develop and implement a parallelized version of GW-CoGAPs to identify patterns from single cell data. Much of our preliminary data arises from this fruitful and productive collaboration.

Furthermore, **Rob Patro** proposes these techniques to infer relevant features to quantify transcript abundance. We look forward to collaborating with the consortia to continue to develop state of the art factorization methods and establish a common framework for model validation and comparison.

Deep Learning. Casey Greene's deep learning techniques propose training on arbitrary patches are selected and arbitrary rotations of the data. This ensemble approach maps to the parallelization across subsets of samples in our proposal. We plan to work collaborate to evaluate methods for optimal feature selection across transcripts and samples for efficient and robust pattern inference. Dr. Greene's

Benchmark data. Assessing performance of the proposed algorithm relies on presence of matched, time-course data across bulk and single-cell RNA-sequencing measurement technologies. Benchmark data in the human cell atlas will be critical to optimal algorithm development. We will also benefit from datasets such as the retinal development data in bulk, smart-seq, and 10X platforms provided by **Loyal Goff**. Algorithm development will further benefit from datasets with parallel perturbations in multiple datasets from **Arjun Raj**.

Integration with imaging/spatial single cell data. In collaboration with **Arjun Raj**, we are working to interface their image analysis pipeline for single molecule FISH with a robust statistical test to permit multifactor analysis of these quantifications.