

# Search for and transformation of human cells and cell types with latent space representations

This manuscript ([permalink](#)) was automatically generated from [greenelab/czi-seed-rfa@c97109d](#) on October 11, 2018.

## Authors

---

- **Casey S. Greene**

 [0000-0001-8713-9213](#) ·  [cgreene](#) ·  [greenescientist](#)

Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania

- **Stephanie C. Hicks**

 [0000-0002-7858-0231](#) ·  [stephaniehicks](#) ·  [stephaniechicks](#)

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

- **Jane Roe**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [janeroe](#)

Department of Something, University of Whatever; Department of Whatever, University of Something

# Abstract

---

## Five Key References

---

## Project Team (750 words each)

---

- Elana Fertig
- Loyal Goff
- Casey Greene (Submitter)
- Tom Hampton
- Stephanie Hicks
- Mike Love
- Rob Patro

## Proposal Body (2000 words)

---

- Base enabling technologies:
  - Low dimensional representations (Elana)
  - New methods published by other groups (scVI, etc - other groups)
  - Fast & improved quantification (Rob)
  - Incorporation of uncertainty estimates in low dimensional representations (Rob / Mike / Elana) - note this can be done w/o modification in scCoGAPS

### Aim 1

- Fast search: (in low dimensions, quantifying differences between case and reference maps, twist: shared latent spaces / k-mers)
  - Differences b/w maps (Stephanie)
  - New models for UMI deduplication accounting for transcript-level information (Rob)
    - Parsimony & likelihood based, integrated with gene-level uncertainty
  - Everything FAST! API for search against HCA reference? (Rob)
  - k-mer / quantified latent spaces (Casey / Rob)

### Aim 2

- Eschewing marker genes: Practical exploration of the HCA in latent spaces
  - Search tool for perturbations / signatures in latent-space(s) (Loyal)
    - Differential analysis of latent space usage across contexts

- Latent space transformations for progression? Consider jawns for semi-supervised learning? (Elana)
- Transfer learning of signatures *between/across* tissues (Loyal)
- Reference Cell types
  - Cell-type summarized expression profiles (Mike, Loyal)
    - A 'reference catalog' of reduced dimensional representations
  - Versioning & provenance of cell types / features as the reference dataset changes (Mike)
  - 'Power-user' application of latent-space discovery in novel dataset and projection of HCA into new learned spaces.

### **Aim 3**

- Delivery
  - Training / teaching (scRNAseq, low-dimensional representations, RFA-developed tools) (Tom)
  - Software hardening/testing (Casey - software eng)
  - Bioconductor integration (Stephanie, Mike)

## References

---