Johns Hopkins University
Department of Biology
Baltimore, MD 21218

July 8, 2017

The European Bioinformatics Institute
Wellcome Genome Campus
Cambridgeshire, CB10 1SD

To Whom It May Concern:

I am applying for the position of HCA Bioinformatician, specifically for building the HCA Data Coordination Platform, advertised on the HCA website. As I finish my postdoctoral research fellowship at Johns Hopkins University, I am looking for new opportunities to contribute to **open-source, reproducible workflow development for promoting effective data sharing and re-use**. I have a passion **for transparent, reproducible, and collaborative science**, especially as it relates to genomics and life sciences research, and I believe that I possess the qualities and experience required to be a productive and integral member of the HCA team.

The opportunity to contribute to the HCA project is especially attractive for me as it aligns with my goal of <u>improving metadata standards and transparent workflow practices</u> in all areas of biological research. As a re-user of published NGS data for the development of Galaxy training materials, I am constantly frustrated by the lack of easily interpretable raw and processed data due to insufficient metadata and computational methods reporting. Oftentimes tool versions and parameters are missing or "custom scripts/code" are referenced, making reproduction of published results nearly impossible. Over the past five years, I have published over 130 NGS libraries to the NCBI's Gene Expression Omnibus in coordination with both primary authorship and collaborative research projects. With each published dataset I aimed to provide sufficient metadata regarding how the data were generated and processed to produce the published results. Further, as part of the Galaxy Training Network I have published Galaxy-based workflows for both mRNA-seq and small RNA-seq data analyses. I am eager to continue collaborating with NGS data generators and users to assess improvements and best practices for big data re-usability.

In addition to experimental techniques, I strongly believe that <u>training in interpretation and assessment of bioinformatic analyses is critical for all scientists</u>. As a postdoctoral research fellow, I have honed my bioinformatics training skills by developing materials for and teaching the *Bioinformatics: Tools for Genome Analysis* course for Johns Hopkins University. As a member of the Galaxy Project, I have led ten workshops teaching bioinformatics tools, workflows, and best practices – in both academic and industry settings – using the Galaxy platform. By leading both small (<10 attendees) and large (>80 attendees) workshops, I have learned how to set up and manage ideal cloud infrastructures (including Amazon Web Services and the US-based XSEDE virtual system) for resource-intensive computing. As a member of the HCA team, I look forward to engaging with scientists on how to interact with the HCA services to enhance their own research through HCA data re-use as well as contribute to the research goals of the scientific community as a whole through HCA data submission.

During all stages of a research project <u>successful collaborations between bioinformaticians and biologists are vital to achieve reproducibility and transparency</u>. Throughout my career, I have had many opportunities to collaborate with both computational and molecular biology/genetics research groups. For example, this May I participated in a three-day hackathon that included bioinformaticians from the Galaxy Project, the Global Organisation for

Bioinformatics Learning, Education and Training (GOBLET), and the European Life Sciences Infrastructure for Biological Information (ELIXIR) initiative. During this event, we curated bioinformatics training materials for the Galaxy Training Network and integrated additional training resources such as ELIXIR's Training e-Support System (TeSS) and the EMBL-EBI's EDAM topics ontology. Further, over the past six years I have worked closely with non-computational researchers by advising in the generation of and analyzing genomics, transcriptomics, proteomics, and metabolomics datasets, which resulted in six non-primary author publications. These scientific achievements emphasize a track record of successful and productive collaborative efforts with a variety of scientists, and I am confident that these skills will be valuable to achieving the HCA mission of delivering simple, open, and direct access to all HCA data, workflows, and analyses.

As a bioinformatician responsible for contributing to the building of the HCA Data Coordination Platform, I believe I can pursue my scientific and career goals as well as contribute to the overall mission of the Human Cell Atlas project. I welcome the opportunity to further discuss my application and the goals of the HCA, and I look forward to hearing from you soon.

Yours sincerely,

Mallory Freeberg, Ph.D.