May 6, 2019

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

To Whom It May Concern:

I am applying for the position of Project Lead with the European Genome-phenome Archive (EBI01399) to coordinate archiving and distributing UK Biobank whole genome and exome data through the EGA. I am excited for this opportunity to take on a leadership and coordination role, and I believe that I possess the qualities and experience required to be a productive and effective Project Lead with the EGA.

The described role is especially attractive to me as it aligns with my belief in the potential for big data to have a profoundly positive impact on the quality of human life: the UK Biobank aims to integrate genetic and environmental data to investigate the development of human disease, and the EGA aims to archive and share this type of data while supporting data access regulations. The potential of big data can only be realized, however, if data are Findable, Accessible, Interoperable, and Re-usable (FAIR). In my graduate and post-doctoral research, I re-used published next-generation sequencing data to achieve my research aims. These datasets were often accessible in standard archives and interoperable in standard formats, but, frustratingly, they were not always findable or re-usable without significant effort. A growing desire to improve the FAIR-ness of biological data led me to join the Human Cell Atlas (HCA) Data Coordination Platform (DCP) project to define metadata standards for cellular resolution data that adhere to FAIR data principles. As a Project Lead for the EGA, I hope to continue working towards maximizing the positive impact of big data on the scientific community and, ultimately, on human life.

While part of the Galaxy Project and the HCA DCP, I coordinated work among members of multiple teams to achieve project goals. I worked with other trainers from the Galaxy Training Network to identify what bioinformatic analysis topics our users wanted training on and to improve the quality of the training materials by incorporating feedback from workshop attendees. With the HCA DCP, I have multiple roles that require organization and flexibility to changing priorities. For example, my role as senior data wrangler involves planning biweekly sprints to coordinate work with the five other data wranglers from the EBI and the University of California at Santa Cruz (UCSC). Early communication struggles between the data wranglers led me to implement and chair weekly data wrangler meetings where we report on, plan, and prioritize our work. Planning this work requires translating DCP-wide priorities decided by the project leads into the specific data wrangling tasks that need to be accomplished. As part of the HCA metadata team, I chair monthly meetings to coordinate our work and to discuss metadata requirements from other DCP teams at the Broad Institute and the Chan Zuckerberg Initiative. I often deputize for our project coordinator and project lead to chair project coordination calls or attend project management meetings when they are unavailable. The project coordination skills I have learned from these experiences will be valuable for collaborating with both the EGA and UK Biobank teams to translate the project's goals into high-quality and timely deliverables.

The Galaxy Project and the HCA DCP both aim to provide a resource for scientists that supports FAIR data. Their success relies on collaborations with external scientists and organizations, and I strongly believe that all big data projects must be able to integrate with external resources to take advantage of others' expertise. Recently, I had the opportunity to coordinate work between the EBI's HCA DCP team, the Unified Submission Interface team, and EBI archive teams (ENA, BioSamples, BioStudies) to establish an automated pipeline for accessioning HCA datasets in EBI archives. Through prioritization of features, assessment of solutions to eliminate blockers, and facilitation of communication between relevant developers, I was able to ensure delivery of a workable pipeline and successfully archive the first two HCA datasets. This work required having technical discussions with developers and priority discussions with project managers, skills that are highly relevant for a Project Lead role with the EGA.

In addition to communicating with external scientists, I believe that it's important to engage in discussions about big data with the public to foster a sense of shared stewardship and build trust. To this end, I have spoken about the HCA DCP with a wide range of audiences at public engagement events. For example, last summer I spoke to members of the Cambridge Society for the Application of Research about the HCA, answering their questions about data privacy and describing how the HCA is contributing to improving human health. These examples of my interactions with external groups emphasize a track record of successful and productive collaborative efforts with a variety of teams and stakeholders These skills will be valuable to achieving the EGA's mission of making UK Biobank genome and exome sequencing data FAIR for the scientific community.

I believe I can pursue my scientific and career goals as well as contribute to the overall missions of both the EGA and the UK Biobank through the role of Project Lead for the EGA. I welcome the opportunity to further discuss my application, and I look forward to hearing from you soon.

Yours sincerely,

Mallory Freeberg, Ph.D.