APPENDIX C - COMPUTING SUPPORT QUESTIONNAIRE

If you are requesting computing equipment that has a total value of more than $100,000, consultation with Digital Alliance is required. Research and High Performance Computing Services is available to assist you with this process. Please complete the following questionnaire and submit your responses, as well as copies of your Project Summary and Infrastructure sections, to [**Jason Brodeur**,](mailto:brodeujj@mcmaster.ca) Director, Research and High Performance Computing**.**

1. Please provide a summary of your group’s existing IT resources (hardware, operating systems, application programs, staff).

The computational suite requested is new and will be integrated in McMaster’s $8M+ CFI-funded shared high performance computing cluster (HPC). McMaster’s HPC is part of the McMaster Service Lab and Repository which is currently maintained by McMaster. The McMaster HPC is a shared cloud computing environment with 264 CPU cores totaling 614.1 GHz and 3.62 TB of memory. Data storage is provided by 1 PB in high-performance spinning disk storage, 70 TB high speed flash-based storage, and 600 TB of regular spinning disk, including daily incremental backup with 45-60 day retention and with archival solution included. Based on VMWare cloud computing, researchers are able to configure computational instances to fit their needs, both for short term and long term needs, with the ability to reconfigure on demand. This will be essential for the Wilson Pregnancy Lab to set up their specific computational experiments. **The requested equipment is thus an increase upon existing CFI-funded computing capacity.**

1. Will the requested equipment replace or add to any current equipment or system(s)?

All requested equipment is new and will not replace any current equipment. The requested equipment will be added to the existing HPC.

1. Please describe any special or unusual features of the requested IT equipment, and how/why they are required for the proposed research project, if applicable (include any interfaces to specialised equipment).

The requested equipment contains multiple GPU cores and secure large data storage to store meta data, genomic, epigenomic, and gene expression data from human, cell culture, and animal model biospecimans ($83,227.79).

This infrastructure in vital to the Wilson Pregnancy Lab as we are generating large amounts of sequencing and microarray data and storage capacity is a limiting factor. The proposed storage infrastructure will guarantee that we do not run into limited storage for the foreseeable future. The requested storage infrastructure will be integrated into the McMaster CSU data backup and integrity system that is completed for all data within the CSU cluster. The requested GPUs are vital to the Wilson Pregnancy Lab as we need to use substantial running memory to complete our analyses, specifically for our machine learning algorithms. The requested computational suite infrastructure is an incremental expansion on the existing McMaster HPC infrastructure.

The equipment being purchased is to augment the existing cluster to allow the researcher to utilize both the existing equipment and the new purchased equipment to complete the research initiative.  Like many genomics programs, we are consistently limited by storage capacity due to generating large volumes of raw sequencing data and intermediate files on a monthly basis, consistent with McMaster’s "Data, Artificial Intelligence and the Digital Society" research initiative, focusing on generating large-scale data and using artificial intelligence (i.e. machine learning) to tackle important scientific questions.

1. Where will the requested equipment be located? If not in an existing central data centre (e.g. UTS, CSU, RHPCS / SHARCNET) please explain.

The requested equipment (Storage and GPUs) will be hosted in McMaster’s CSU with the existing CFI-funded infrastructure. As this equipment is to add additional resources to the cluster it will require access to all the other equipment to allow the best functionality.

1. Please state to which network(s) the requested equipment will be connected and estimate such connection(s) bandwidth requirements.

The network connections for this equipment require 25GB connections to allow high speed analysis and transfer of data. This requirement is consistent across McMaster’s CSU cluster. Outside of the cluster itself, bandwidth requirements are low beyond initial transfer of data to the cluster.

1. Please identify who will install, commission, operate and maintain the proposed computing systems.

McMaster CSU is now a McMaster Faculty of Health Sciences core facility with installation, commission, operation, and maintenance provided a not cost to investigators.

1. Please explain why any existing centrally-available resources (e.g. Sharcnet, Digital Alliance) are unsuitable to support this research project.

The request is addition of secure storage for patient, genomic, and meta-transcriptomic data from microbiome-derived samples ($$83,227.79):

* The proposed equipment will increase storage and compute to the existing CFI-funded computing environment that fosters wide-spread collaboration among the McMaster community. In addition, the proposed storage will be integrated with the existing McMaster N-ABLE data backup and integrity system for all storage within the CSU cluster, particularly the existing CFI-funded 70 TB high-speed flash-based storage (Pure Storage, Inc.) that enables ground-breaking rapid analysis of genomic, metagenomic, and microbiome data.
* The proposed storage is an incremental expansion of an existing resource at McMaster. No additional purchases are required.
* The existing computing environment contains novel high-performance 70 TB high-speed flash-based storage (Pure Storage, Inc.), enabling high speed I/O beyond traditional spinning disk, critical for rapid outbreak analysis, chemical-genetic screening, and large-scale examination of whole viral and microbial communities in the context of infection and patient genomes. To our knowledge this flash-based storage is not supported by the national platform and any gains would be lost if it had to be accessed from externally hosted infrastructure due to network limitations, compromising research efficiency. In addition, the requested storage would be integrated with CFI-funded high performance, high memory HPE Superdome Flex (144 cores & 3 TB memory) and HPE Apollo (80 CPU + 8 GPU with 384 GB memory) compute critical for the very large multi-omic’s analysis that is part of our research program. This is enabling technology. Without it we would be forced to scale back the size of our experiments. As above, any gains would be lost if it had to be accessed from externally hosted storage.
* Our research is closely linked with clinical teams, often with transfer for RNA or DNA extractions from patients under Material Transfer Agreements or collection of patient genomic data under explicit data usage and sharing agreements. This explicitly applies to the Wilson Pregnancy Lab as their ethics approval does not permit data to leave McMaster Campus, and that cloud computing cannot be used and data cannot be shared outside of the institute.

Signature of Principal Investigator Signature of Director, RHPC

CFI IF 2023 Application Guidance Document – CONFIDENTIAL (for McMaster use only) Page 1