FACILITIES & OTHER RESOURCES

Updated September 1, 2020

Specific Fields Relevant for Emory Integrated Computational Core (EICC) Users

**Other**:

**EMORY INTEGRATED COMPUTATIONAL CORE (EICC)**

The **Emory Integrated Computational Core (EICC)**, one of the **Emory Integrated Core Facilities (EICF)**, offers comprehensive computational services and bioinformatics pipelines for the analysis of -omics data. The EICC has 1000 sq ft of dedicated office space on the 7th floor of the Woodruff Memorial Research Building that provides for meeting customers, weekly meetings with the members of the Emory Integrated Genomics Core (EIGC), and for monthly meetings of computational service providers from other cores within the EICF.

The EICC operates a small HPC system for short computational jobs. The cluster serves multiple functions related to core projects, including running NGS analysis pipelines, high-performance and parallel computing for disciplines such as proteomics, metabolomics, and imaging. The cluster is composed of 1 head node, 5 high-memory compute nodes, and one GPU node with 4x Tesla V100 GPUs. The cluster has a 2PB local storage array and offers access to Emory Isilon storage (1PB research-grade storage, 500TB of HIPAA compliant storage). A 10 Gbps ethernet switch provides a high-speed Storage Area Network (SAN) fabric. All storage arrays and compute nodes utilize the SAN for data transfer and are configured to connect via the 10 Gbps high-speed network. The cluster is connected to the Internet2 high-speed network for large data transfers to and from external systems. The cluster runs Scientific Linux 7 64-bit operating system on all nodes and utilizes Slurm for job submission and management. Configuration: One head node: 2 x 3.3 GHz 8-core CPUs, 64 GB RAM. Five compute nodes: 4 x 2.2 GHz 16-core CPUs, 512 GB RAM, 10 Gbps ethernet. One GPU node: 2 x 2.3 GHz 16-core CPUs, 384 GB RAM, 4 x nVidia Tesla V-100 GPU with 32 GB RAM.

Amazon Web Services (AWS) is an on-demand delivery of IT resources in the cloud with pay-as-you-go pricing. The AWS infrastructure is highly durable, available, elastic and scalable. The Emory AWS environment is an AWS environment that is established according to the Emory business, security and compliance practices. Access to the Emory AWS Console must be authenticated with Emory Single sign-on. The virtual private cloud within the environment is protected by the Emory firewall. Secured connection (SSH or RDP) to an EC2 instance must be made from a workstation already located on the Emory Core network or via a VPN tunnel that is authenticated by 2-factor authentication. All AWS services have been reviewed by the Emory Security Team, and specific guidelines about utilizing these services for HIPAA or identifiable health information will be published. Emory University provides access to the Emory AWS environment to researchers as part of the overall IT support for research. AWS computing and data storage expenses, however, are not covered by the university, and must be budgeted for in grant applications. The EICC works with LITS to provide guidance on AWS usage and optimization.

The EICC also provides comprehensive computational services. We divide computational services into two main categories. The first enables expert users to access existing pipelines or develop their own custom analyses. The second category provides investigators the ability to have analyses performed by an EICC computational/bioinformatics expert for a set fee per project. Galaxy provides a wide variety of bioinformatic tools that allow the analysis, manipulation and visualization of large genome-wide datasets from a wide variety of platforms, including microarrays and next-generation sequencing instruments. The EICC also supports an enterprise HIPAA compliant LabKey server for Emory investigators. Collaborators outside Emory can also access this LabKey server infrastructure when collaborating with Emory investigators.

Standard analysis pipelines using other open-source software packages are implemented for DNA/RNA-seq/ChIP-seq/16S microbiome sequencing projects for human, animal, and microbial genomes. We have implemented the QIIME 2 pipeline for microbiome data analyses. Custom tools or other pipelines (such as mothur) are also available. For the analysis of RNA-seq data, we have implemented the Star and HTseq-count pipeline. Custom tools and pipelines can be developed for specialized projects such as fusion transcript detection. For targeted sequencing, exome sequencing, and whole genome sequencing, we use a custom PEMapper and PECaller pipeline. For variant annotation, we use the bystro.io software package. We have also implemented and analyze data sets with other mapping and variant identification pipelines (BWA, GATK). Substantial capacity exists for these integrated computing resources to support computational/bioinformatic analyses for EICC users.As part of our comprehensive bioinformatics services, we facilitate collaborations with existing biostatistics/bioinformatics service centers at Emory. These include the Winship Biostatistics and Bioinformatics Shared Resource headed by Dr. Jeffrey Switchenko, the Bioinformatics and Systems Biology Shared Resource headed by Dr. Manoj Bhasin, and the Biostatistics Consulting Core (BCC) headed by Dr. Renee Moore.