

## FACILITIES, EXISTING EQUIPMENT, AND OTHER RESOURCES

The principal investigators have access to extensive computational and experimental facilities available at the University of Chicago and the University of Iowa to carry out the projects outlined.

### University of Chicago facilities, equipment, and other resources: Argonne Leadership Computing Facility (ALCF).

In partnership with Rick Stevens and Argonne National Lab, we will also have access to Argonne's suite of high-performance machines, including ThetaGPU with 28 NVIDIA DGX-3 state-of-the-art AI systems, Polaris, which will be installed in spring, 2021 with 1024 NVIDIA A100 GPUs able to support 7,000 simultaneous training sessions at once for problems that need hyper parameter optimization or for exploring properties of models. Aurora will be installed Summer 2022 and will host 60,000 Intel PVC-XT, performing AI inference at something like 64x faster at over 100 EFs per second, able to train up to 480K models in parallel and some very large models, supporting training models as high as 100 trillion parameters. Human brains have on order 100 trillion synapses ( $10^{11}$  neurons and  $1.8\text{--}3.2 \times 10^{14}$  synapses) and so this machine is, in some sense, a brain scale computer, likely the fastest on the planet when installed, and by far the fastest devoted to AI. Finally, we will have access to the National AI Accelerator Assessment Testbed (N3AT) Facility, designed to assess the usefulness of novel hardware designs from dozens of startups for advancing key research problems in AI. These include computers from Cerebras, SambaNova, Graphcore, Groq and many others. These models have the demonstrative ability to training transformer models, famously large in parameters, much more efficiently than within standard computing arrays (see Data and Computation). We will use Globus ([www.globus.org](http://www.globus.org)) as a platform for data management, which provides managed data transfer solutions for high-speed, reliable, secure file transfer and replication among any variety of storage systems, including on-premise file systems and object stores, desktops and laptops, and cloud storage.

ALCF also houses a range of other high performance computing capabilities spearheaded by its 10-petaflop IBM Blue Gene/Q system, Mira. Mira is comprised of 48 racks, with 1,024 nodes per rack; 768 terabytes of RAM; 35 PB of storage; and 384 I/O nodes.

ALCF also hosts large storage systems. Mira data systems consist of 384 I/O nodes that connect to 16 storage area networks (SANs) that control 8,960 disk drives with a total capacity of 28.8 PB of raw storage and a maximum aggregate transfer speed of 240 GB/s. ALCF computing resources share two 10,000-slot libraries using LTO4 tape technology. The LTO tape drives have built-in hardware compression with compression ratios typically between 1.25:1 and 2:1, depending on the data, giving an effective capacity of 16-24 PB. Two parallel file systems—PVFS and GPFS—are used to access the storage.

ALCF also operates a Globus-based data management and sharing pilot service called Petrel. Petrel provides 1PB of storage that can be used by Argonne researchers and users of Argonne facilities to store and share data with collaborators without the need for local account management.

**Research Computing Center:** The University of Chicago Research Computing Center (RCC) provides high-end research computing resources to researchers at the University of Chicago, which include high-performance computing and visualization resources; high-capacity storage and backup; software; high-speed networking; and hosted data sets. Resources are centrally managed by RCC staff who ensure the accessibility, reliability, and security of the compute and storage systems. A high-throughput network connects the Midway Compute Cluster to the UChicago campus network and the public internet through a number of high-bandwidth uplinks. To support data-driven research RCC hosts a number of large datasets to be accessed within the RCC compute environment.

**Compute Infrastructure.** RCC maintains three pools of servers for distributed high-performance computing. Ideal for tightly coupled parallel calculations, tightly-coupled nodes are linked by a fully non-blocking FDR-10 Infiniband interconnect. Loosely-coupled nodes are similar to the tightly-coupled nodes, but are connected with GigE rather than Infiniband and are best suited for high-throughput jobs. Finally, shared memory nodes contain much larger main memories (up to 1 TB) and are ideal for memory-bound computations. The types of CPU architectures RCC maintains are tabulated in Table 1.

RCC also maintains a number of specialty nodes:

- *Large shared memory nodes* - up to 1 TB of memory per node with either 16 or 32 Intel CPU cores. Midway is always expanding, but at time of writing RCC contains a total of 13,500 cores across 792 nodes, and 1.5 PB of storage.
- *Hadoop*: Originally developed at Google, Hadoop is a framework for large-scale data processing.

TABLE 1  
University of Chicago Research Computing Center Capabilities Summary

Cluster	Partition	Compute cores (CPUs)	Memory	Other configuration details
midway1	westmere	12 x Intel X5675 3.07 GHz	24 GB	
	sandyb	16 x Intel E5-2670 2.6GHz	32 GB	
	bigmem	16 x Intel E5-2670 2.6GHz	256 GB	
		32 x Intel E7-8837 2.67GHz	1 TB	
	gpu	16 x Intel E5-2670 2.6GHz	32 GB	2 x Nvidia M2090 or K20 GPU
		20 x Intel E5-2680v2 2.8GHz	64 GB	2 x Nvidia K40 GPU
	mic	16 x Intel E5-2670 2.6GHz	32 GB	2 x Intel Xeon Phi 5100 coprocessor
	amd	64 x AMD Opteron 6386 SE	256 GB	
	ivyb	20 x Intel E5-2680v2 2.8GHz	64 GB	
	midway2	28 x Intel E5-2680v4 2.4GHz	64 GB	
	bigmem2	28 x Intel E5-2680v4 @ 2.4 GHz	512 GB	
	gpu2	28 x Intel E5-2680v4 @ 2.4 GHz	64 GB	4 x Nvidia K80 GPU

- *GPU Computing:* Scientific computing on graphics cards can unlock even greater amounts of parallelism from code. RCC GPU nodes each include two Nvidia Tesla-class accelerator cards and are integrated in the Infiniband network. RCC currently provides access to Fermi-generation M2090 GPU devices and Kepler-generation K20 and K40 devices.
- *Xeon Phi:* The Many Integrated-Core architecture (MIC) is Intel's newest approach to manycore computing. Researchers can experiment with these accelerators by using MIC nodes, each of which have two Xeon Phi cards, and are integrated into the Infiniband network.

**Persistent and High-Capacity Storage.** Storage is accessible from all compute nodes on Midway1 and Midway2 as well as outside of the RCC compute environment through various mechanisms, such as mounting directories as network drives on your personal computer or accessing data as a Globus Online endpoint (at the time of this writing, Globus Online is supported on Midway1). RCC takes snapshots of all home directories (users' private storage space) at regular intervals so that if any data is lost or corrupted, it can easily be recovered. RCC maintains GPFS Filesystem Snapshots for quick and easy data recovery. In the event of catastrophic storage failure, archival tape backups can be used to recover data from persistent storage locations on Midway. Automated snapshots of the home and project directories are available in case of accidental file deletion or other problems. Currently snapshots are available for these time periods: 1) 7 daily snapshots, 2) 4 weekly snapshots.

**Tape Backups.** Backups are performed on a nightly basis to a tape machine located in a different data center than the main storage system. These backups are meant to safeguard against events such as hardware failure or disasters that could result in the complete loss of RCC's primary data center.

**Data Sharing.** All data in RCC's storage environment is accessible through a wide range of tools and protocols. Because RCC provides centralized infrastructure, all resources are accessible by multiple users simultaneously, which makes RCC's storage system ideal for sharing data among your research group members. Additionally, data access and restriction levels can be put in place on an extremely granular level.

**Data Security & Management.** The security of the Research Computing Center's storage infrastructure gives users peace of mind that their data is stored, managed, and protected by HPC professionals. Midway's file management system allows researchers to control access to their data. RCC has the ability to develop data access portals for different labs and groups.

**The Institute for Molecular Engineering at the University of Chicago** house a vibrant research community of multidisciplinary scientists that regularly collaborates to make significant scientific contributions. UChicago also features several translational resources, such as the Human Tissue Research Center, and the Transgenic Animal Center.

**The University of Chicago Comprehensive Cancer Center (UCCCC):** One of only two NCI-designated Comprehensive Cancer Centers in Illinois, the UCCCC has a reputation for excellence and innovation and a commitment to address cancer through clinical and basic science cancer research and training, clinical

cancer care, and expertise in population research. UCCCC researchers have access to a comprehensive set of shared technologies with the University of Chicago Biological Sciences Division (BSD), including 13 Core facilities. The UCCCC offers a wealth of intellectual, technological, and financial resources to pursue a comprehensive, collaborative research program involving more than 215 renowned scientists and clinicians.

**Translational and collaborative research:** The University of Chicago's strong physical sciences division, including my home department of Chemistry, is located in direct proximity to the medical school and hospital system. Indeed, I chose to start my independent career here at UChicago specifically so that I could develop a group whose work could impact human health. I have now witnessed firsthand the benefits of this proximity and have developed several strategic collaborations with clinicians and clinical researchers to develop new technologies. The University devotes substantial resources to translational research, which provides a clear path to move from the bench to the clinic. This includes the University of Chicago Innovation Exchange (<https://innovation.uchicago.edu>), which provides seed money and expertise to translate basic science discoveries into commercial ventures and to foster collaborations between the basic science divisions, medical school, and national labs. Therefore, the University of Chicago is an exceptional location to pioneer paradigm-shifting biomedical technologies.

**University of Chicago Core Facilities** The University of Chicago offers extensive access to cutting-edge research technologies and expertise needed for my work via a well-funded and expertly staffed set of core facilities:

- The Microscopy Core – Maintains an extensive array of microscopes that our group can access, including a variety of confocal and 2-photon fluorescence scopes. The primary microscope we currently use is a Nikon Confocal featuring multiple laser excitation sources, a 2-photon light source, an automated stage, and a heated chamber. This will be very useful for the time course, live cell imaging experiments associated with the proposed research when our group's microscope is insufficient.
- The Biophysics Core - Provides access and training to state-of-the-art biophysics equipment, such as a Biacore, Plate Readers, and a ProteOn XPR36 protein interaction array system.
- Transgenic/ES Cell Technology Mouse Core Facility - Provides genetically manipulated mice through transgenic technology or embryonic stem (ES) cell manipulation. The facility provides a comprehensive set of technical services and a fully operational construction and gene targeting service.

**University of Iowa, Facilities and Other Resources: Laboratory Space.** Dr. Manicassamy's Laboratory is housed in the Department of Microbiology & Immunology on the second floor of the Bowen Science Building (BSB; Core 400) at The University of Iowa. This state of the art facility is comprised of 1,700 square feet of newly renovated laboratory space. In addition to the infrastructure available in the Department of Microbiology & Immunology, Dr. Manicassamy's research is supported by excellent core facilities at the University of Iowa, including Next Generation Sequencing, Bioinformatics Core, Flow Cytometry, Central Microscopy Research Facility, Small Animal Imaging, DNA sequencing. Dr. Manicassamy has full-time administrative support through the Department of Microbiology & Immunology.

**Office Space.** Dr. Manicassamy has 200 square feet of separate office space adjunct to the laboratory in BSB. PC and Mac computers, computational network, laser printers, color printers, and scanners are available. Manuscripts and desktop publishing of papers can be prepared in several offices available to scientists working in BSB.

**Scientific Environment.** The University of Iowa has a highly collaborative research community with several leading Virologists and Immunologists, including Drs. Stanley Perlman (Coronaviruses pathogenesis/host responses), Mark Stinski (Herpes Virus), John Harty (T cell responses to Infection), Kevin Legge (dendritic cell-T cell responses to Influenza virus), Steven Varga (Host responses to RSV), Gail Bishop (Viral Immunology), Wendy Maury (Filovirus entry), Richard Roller (Molecular Herpes virology), Jack Stapleton (HIV/HCV pathogenesis), and Hillel Haim (HIV evolution/pathogenesis). Weekly Microbiology seminar series, and the Virology and Immunology journal clubs provide excellent opportunities for students and postdocs for scientific interactions. In addition, research program in pulmonary biology provide an excellent scientific forum for discussion and collaboration. The Levitt Center for Viral Pathogenesis provides funding for seminar speakers and student travel and is another venue for interactions with research interests related to the project. Moreover, the students and postdocs are supported by several NIH T32 training grants.

**Animal Care Unit (ACU)** at The University of Iowa is in full compliance with all NIH guidelines and regulations pertaining to the care and use of experimental animals (PHS Assurance no. A3021-01). The

ACU has enjoyed accreditation from the American Association of Accreditation of Laboratory Animal Care since 1994. The ACU maintains centralized animal housing facilities, which are staffed by highly trained individuals to provide husbandry and research support services. In addition to providing daily animal care, all ordering and receipt of animals, quarantine and health monitoring is performed by the ACU. The ACU also provides research support services, such as anesthesia and surgical support, rodent breeding assistance, diagnostic laboratory services, and investigator training. The ACU veterinarians are faculty members of The University of Iowa. They are available to assist investigators and their staff with all aspects of their animal research activities.

**Biosafety Facility.** The select agent containment laboratories are housed on the 5<sup>th</sup> floor of the Carver Biomedical Research Building (CBRB) and on the 4<sup>th</sup> floor of the Medical Laboratories (ML) Building, University of Iowa, Iowa City, IA. BSL3 facilities are CDC certified and is managed by Ms. Dana Reis (Director). Dr. Manicassamy has one BSL3 suite with one class IIb biosafety cabinet and one Animal BSL3 suite available for work on highly pathogenic influenza viruses and coronaviruses.

#### **Major Equipment in Dr. Manicassamy's laboratory::**

- 1) **Biosafety cabinets:** Sterile environment for handling infectious materials, safeguarding researcher and samples.
- 2) **Tissue culture incubators:** Optimal temperature, humidity, and gas conditions for cell and tissue growth.
- 3) **PCR machines and real-time PCR machines:** Amplify specific DNA sequences, monitor amplification process in real-time.
- 4) **Table top centrifuges:** Separate liquid sample components based on density.
- 5) **Microscopes:** Visualize microscopic organisms, cells, and minute structures.
- 6) **-80 freezers and -20 freezers:** Long-term storage and preservation of biological samples.
- 7) **Bacterial incubators:** Optimal conditions for bacterial growth and development.
- 8) **Bacterial shakers:** Facilitate aeration and mixing of bacterial cultures.
- 9) **Thermal cyclers:** Precise temperature control for DNA amplification and molecular biology applications.
- 10) **UV-Spectrophotometers:** Measure light absorbance, determine concentration and purity of biomolecules.
- 11) **Fluorescence microscope:** Visualize and analyze samples using fluorescence, study cellular structures and biomolecular interactions.
- 12) **Gel doc unit:** Capture and document images of DNA, RNA, and protein samples in gel electrophoresis experiments.

#### **UI Department of Microbiology and Immunology Resources:**

- 1) **3 flow cytometers:** Measure and analyze physical and chemical properties of cells or particles in fluid.
- 2) **Fluorescence plate readers:** Measure fluorescence intensity in microplate format, high-throughput analysis of cellular and biochemical events.
- 3) **Zeiss inverted fluorescent and confocal microscopes:** High-resolution imaging of biological samples, visualize living cells and tissues.
- 4) **qRT-PCR thermocyclers:** Quantify RNA transcripts in real-time, provide insights into gene expression.
- 5) **Fuji CCD imaging system:** Capture high-resolution images of fluorescent and chemiluminescent samples, sensitive and accurate detection of biomolecules.
- 6) **High-speed and ultracentrifuges with varied rotors:** Rapid separation of samples based on size, shape, and density, accommodate various rotor types.
- 7) **Typhoon imaging system:** Detect, quantify, and analyze proteins, nucleic acids, and biomolecules in gels, membranes, and microplates.
- 8) **ELISA plate readers:** Measure absorbance of enzyme-linked immunosorbent assays (ELISA), quantify proteins, peptides, and hormones.
- 9) **TopCount:** High-throughput quantification of radioactivity in samples, study biochemical and cellular processes.
- 10) **LiCor imaging system:** Sensitive and accurate detection of fluorescent and chemiluminescent samples in various formats.
- 11) **Darkroom facilities:** Controlled processing of light-sensitive materials, such as photographic films and imaging plates.

- 12) **Cold and warm rooms:** Temperature-controlled spaces for storage and experiments requiring specific temperature conditions.