Gerstein (Bass Central/Main campus)

The Gerstein laboratory is found in two connected buildings. The laboratory consists of 6 rooms and comprises a total of ~1,900 sq. ft. In addition, three conference rooms that have projectors provide venues for interaction. There are 40 gigabit-ready desks, equipped with one or two 23” and 30” LCD screens. The space is properly air conditioned for supporting a large number of computers.

Office:

Mark Gerstein's office space is 178 sq. ft.

Gerstein Lab Computer Infrastructure

Laboratory Network and Storage. The lab’s computing infrastructure is partitioned into a private and a public network. The entire infrastructure is fully gigabit capable and is connected to the Yale backbone via gigabit optic fibre; the network architecture was designed with computing efficiency and network security in mind. The private network consists of individual laptops, desktops and workstations, as well as communal computational servers, dumb terminals, a central fileserver, a consolidated NAS, and printers. There are also servers that provide essential network services such as NIS, NFS, SMB, DHCP, monitoring and backups. The public network consists of numerous production webservers that are either real or virtual machines. The laboratory maintains its own public subnets of 128 public IP addresses and manages many of its own domains (e.g. gersteinlab.org, molmovdb.org, pseudogenes.org, and partslist.org). The lab has a full-time administrator maintaining the network.

The private and public networks obtain gigabit connectivity through four HP Procurve 5300xl switches that are mutually connected via fibre. The private network is behind a Cisco PIX 525, which is concurrently used as an IPSec VPN gateway into the private network. Within the private network are two NetApp storage appliances with 43Tb of raw space, which is configured with 27.5Tb of working space, thirty custom made 4Tb network disks with a total 120Tb capacity, a Dell NAS with a total of 30TB capacity; the NetApp appliances and Dell NAS are used for live user file space, backups of user files and backups of public production webservers. A seven-day incremental backup and a twelve-month incremental backup are currently being implemented in the lab.

Wireless access is available all throughout the lab. Wireless access connects computers directly to the public network.

Available Computers. There are about forty-seven working laptops in the lab, in which eighteen are recent Macbook Pro models.

In total, the lab has 315u of rack space spread over eight racks. Residing in these racks are a dual CPU twelve core Opteron server with 256GB of memory, a dual CPU six core Opteron server with 128GB of memory, a dual CPU four core Opteron server with 64GB of memory, three Intel blade enclosures with 10 dual CPU Intel blades each, fourteen dual cpu 64 bit Xeons servers and six dual cpu 64 bit Opteron servers; these rack servers are in addition to the NetApp storage appliances and the Dell NAS mentioned above. The rack servers have various uses. The dual CPU Opteron servers are for hosting virtual machines, which function as web hosts. In the private network, five rack servers are for essential network services, four are storage head nodes for the Dell SAN and a few are network support or experimental machines. The rest of the rack servers are in the public network acting as webservers. The private network has seven business class color laserjet printers.

Software. A number of open source software, programs created in-house, and proprietary software is used by the lab researchers for their needs. The lab maintains a set of wiki servers for the documentation of internal information and the public dissemination of information. The lab also manages mailman servers for its mailing lists. The compute nodes are mainly used to develop and run Java and Perl code and to perform Matlab and Gromacs calculations. The public webservers are used to deploy Java, Perl, PHP and Python applications.

Individual tasks are coordinated by a web group calendar. Web applications and servers are continually being monitored by a Nagios monitoring system.

Yale Life Sciences Supercomputer. The Gerstein laboratory has priority access to two of the Yale supercomputers, namely Louise and BulldogI, and regular access to six other Yale supercomputers. There are two full-time administrators maintaining the supercomputer.

Louise is a cluster with 112 Dell PowerEdge R610 with (2) quad core E5620 nodes, each with 2.4 Ghz cpu cores and 48 GB RAM. They are interconnected with a Force10 network switch. There is therefore a total of 112*8 cores = 896 cores. Louise has 300 TB (raw) of BlueArc parallel file storage.

BulldogI is a cluster consisting of a head node and 170 Dell PowerEdge 1955 nodes, each containing 2 dual core 3.0 Ghz Xeon 64 bit EM64T Intel cpus, for a total of 680 cores. Each node has 16 GB RAM. The network is Gigabit ethernet. Bulldogi runs a high performance Lustre filesystem. It is managed via PBS. Three 20Tb Dell Power Vault with storage arrays are attached to BulldogI and are dedicated for Gerstein laboratory use.

The laboratory also has priority access to a SGI F1240 system. This system has 12 Xeon E5345 Quad-Core 2.33GHz CPUs (for a total of 48 processor cores), with 2 x 4M L2 cache per CPU, a 1333MHz front side bus, 96GB of memory, and 6 Raptor 150GB, 10K rpm SATA drives. It runs SUSE Linux Enterprise Server 10 as a system single image. That is, all 48 cores are managed by a single process scheduler, and the 96 GB memory is, in principle, addressable by a single process. In practice, system caches and buffers reduce the maximum amount of memory available to any given process to about 70 GB. In many ways then, the system can be thought of as an SMP, but in terms of hardware architecture it is closer to an infiniband-connected cluster.

Core Lab. The Gerstein Lab is adjacent to the Yale Center for Structural Biology (CSB) Core laboratory. The Core laboratory resources are available to members of the Gerstein lab. The Core laboratory supports the work of all the people associated with the CSB, in total about 200 users and >200 computers. These computers include a number of high-performance graphics workstations for visualizing macromolecular structures and complex data sets. The CSB Core staff of 2 FTE provides support to the associated CSB laboratories as well as the Core computers.

Oracle Server. Yale University has an institutional site license for the Oracle database management system. As a result, many major administrative computing systems at Yale are being developed using Oracle, and Yale’s ITS staff has extensive Oracle experience. Yale ITS maintains and operates several Oracle database systems at the School of Medicine, and provides access to these machines to many different projects. There are several advantages to using institutional servers. The ITS staff backs up each database on a regular schedule, typically with full backups weekly and partial backups several times a day. The ITS staff maintains the hardware of the database machine, the system software, and the Oracle software. They perform periodic upgrades when new versions of the software become available. They also handle any systems problems that occur, and are available to help troubleshoot any application problems that arise.

FACILITIES AND OTHER RESOURCES:

CLINICAL: The Yale New Haven Hospital is the 6th largest hospital in the United States and currently has capacity for over 1500 patients at any given time. The Yale Cancer Center and Smilow Cancer Hospital is an NCI-designated Comprehensive Cancer Center and is part of the National Comprehensive Cancer Network. The Genitourinary Cancer Program has an active clinical and translational research team. The hospital is ideally set up adjacent to research laboratories for translational research studies. Patients are consented for research in the clinic and procurement of blood is obtained in the blood draw station or prior to surgery. These specimens are immediately transported to the laboratory. The operating rooms are adjacent to the pathology station to allow procurement and processing of fresh tissue. This allows rapid acquisition of patient material for research purposes.

OFFICE: Dr. Shuch also has a 120 sq. ft. office located within the Department of Urology research space adjacent to the Brady Memorial Urology laboratory. Additionally he shares a Department administrator for academic support who is located adjacent to his office.

PERSONAL COMPUTING: In Dr. Shuch’s office there is a brand new high-powered Macintosh desktop with 32 GB 1600 MHz with a 3.4 GHz Intel Core. There are also two 27-inch thunderbolt displays. This three-screen set up is useful for running various genomic programs. The machine also has Mac Parallels installed to run Windows for specific genomics programs. Additionally he has a Macbook Pro Laptop 8 GB 1600MHz with a 3 GHz Core.

LABORATORY: Dr. Shuch’s laboratory is in a shared Urology Laboratory consisting of resources for molecular biology, cellular biology, and genetics research. His laboratory is approximately 700 sq. ft. of space as is located in the Brady Memorial Laboratory. There are two laminar flow culture hoods, two -80˚C freezers, two -20˚liquid nitrogen tanks, a cold room, shared dark room, several inverted fluorescent/brightfield scopes, and common equipment space. The laboratory is equipped with typical tools of molecular biology and genetics research including BioRad Systems for western blots transfer, an Immunohistochemistry station, a 96-well PCR thermal cyclers, and an Applied Biosystems qRT-PCR analyzer. There is a PerkinElmer Fluorescence reader for 12-96 well microplates. There is an apparatus for both agarose and acrylamide gel electrophoresis.

YALE TISSUE PROCUREMENT SERVICES (YTPS): This shared resource is run by the Department of Pathology. The research team ensures that specimens removed from the operating room are procured within 30 minutes. Using standard operating procedures for each approved protocol; specimens can be obtained based on investigator preferences. For our heterogeneity assessment the research team obtains three distinct tumor regions that are snap frozen in liquid nitrogen for later use. Other aspects of this service are the ability to retrieve well-annotated formalin-fixed pathology specimens. The Yale Pathology Archives dates back to the 1950s’ and there are over 2000 nephrectomy specimens available.

YALE GENTIOURINARY BIOSPECIMEN REPOSITORY: This Urology Laboratory houses the Genitourinary Biospecimen Repository. Patient blood and urine samples are procured in clinic or prior to surgery. Blood samples are spun down for component separation with an Eppendorf 5804

Refrigeration Centrifuge unit. Cancer tissue obtained from with the YTPS is processed in the laboratory. DNA from whole blood or tumor tissue is extracted using an automated system, the Promega, Maxwell 16 research instrument. A Thermo Scientific Nanodrop unit is available for determination of DNA concentration. Specimens are labeled with a Zebra barcoding system and tracking monitored with an OnCore Biospecimen Management system. All biospecimen materials are barcoded, catalogued, and stored in a dedicated -86˚C Thermo Scientific Ultra-Low Temperature Freezer with storage space for over 86,000 biospecimens.

YALE CENTER FOR GENOME ANALYSIS (YCGA) has multiple next-generation sequencing systems including Illumina HiSeq 2000 (8 machines), HiSeq 2500 (2 machines), and MiSeq (1 machine). The core has a Agena MassArray MALDI-TOF system for multiplex PCR analysis of a panel of gene regions. The core facility also has expertise in library preparation for whole genome sequencing, exome sequencing, mRNA sequencing, and ChIP sequencing. They have technical expertise to capture exomes from formalin fixed paraffin embedded tissues and capture mRNA libraries from <10,000 cells. Additionally they have specific expertise in running experiments involving Molecular Inversion Probe (MIP) sequencing.

Other available equipment includes:

- One Pacific Bioscience third generation sequencer
- One Ion Torrent sequencer
- Dell High Performance Cluster; 110 nodes with a total of approximately 1000 cores/CPUs
- Caliper LabChip GX system
- Caliper Sciclone liquid hander for library preparation
- 3 Agilent Bioanalyzer for Q/C of RNA, cDNA, cRNA and fragmented cRNA
- Eppendorf epMotion 5075LH automated liquid handler
- Molecular Devices SpectraMax II 96/384-well spectrophotometer
- Illumina high throughput BeadArray system with two Tecan robotics and HiScan with autoloader.
- Affymetrix GeneChip Model 3000 Scanners with autoloaders
- 4 Affymetrix GeneChip Model 450 fluidics stations
- Nimblegen array platform with three Maui Hybridization instruments

YALE BIOINFORMATICS RESOURCES: Four Ph.D. scientists provide technical assistance in the analysis of data generated at YCGA. This includes the analysis of data generated by next gene sequencing platforms, microarray analyses, and MIP sequencing. Additionally there is support from Keck's Sequencing center for specific bioinformatics and biostatistics needs. Other faculty members of Yale's Computational Biology and Biology have extensive experience developing new data analysis tools.

KECK DNA SEQUENCING FACILITY: For Sanger sequencing needs, the Keck Shared Biotechnology Resource has served as the center for high-end instrumentation at Yale School of Medicine for nearly a quarter century. The Keck Facility performs custom sequencing on templates from over 1,000 users from within Yale and abroad. This facility provides competitive and timely DNA sequencing in a cost effective manner. The Keck Facility has extensive experience designing molecular inversion probes for sequencing. Sequencing is carried out on Applied Biosystems 3730xL DNA Analyzers, along with Applied Biosystems Big Dye chemistries. MIP primers will be synthesized using Dr. Oligo 192 and Applied Biosystems 394 synthesizers.

HIGH PERFORMANCE COMPUTING (HPC) CENTER: The HPC center runs 2 linux-based clusters that are free for Yale research faculty. BulldogN is a cluster with 170 computer nodes in size from 8 to 64 computer cores with varying amounts of memory per node. All nodes are connected to several petabytes of high performance parallel storage. BulldogN is dedicated for the production, storage and analysis of genomic data generated from the YCGA. Louise is the main Biomedical and Biological Sciences High-performance computing (HPC) resource. This cluster contains 321 nodes with 3680 computer cores. Louise has a BlueArc file system with over 500 TB of usable storage. Louise is dedicated for the storage and analysis of genomic data generated from external data sources.

YALE CANCER CENTER CORE LABS: The Smilow Cancer Center has shared core facilities for members to use at no charge. This equipment is located in 2 common rooms on the medical campus. Equipment includes:

- Applied Biosystems ViiA7 Real Time PCR (Two)
- Zeiss Axiocam microscope camcer
- Zeiss Axio Observer A1- Inverted Microscope (Two)
- Sorvall Thermoregulatable RC6+ High Speed Centrifuge (Two)
- BioRad Chemidoc MP (Two)
- Paradigm Fluorescence Plate Reader
- Gentle Max tissue prep and magnetic bead sorter
- Cytospin
- Leica RM2235 rotary microtome
- Leica CM1950 cryostat
- GE ÄKTA pure chromatography system
- Lonza Nucleofactor/Electroporator
- Guava easyCyte Flow Cytometer
- The Mitocell S200 micro respirometry system
- Virtis Advantage Plus bench top freeze dryer