Facilities and Resources

Laboratory

Dr. Ding's laboratory is located in a state of the art BSL1 laboratory with 500 sq ft of space in room 3101 of the newly built 4515 McKinley Building at Washington University School of Medicine, St. Louis, MO. The laboratory is juxtaposed to all of the other physcian-scientists and translational researchers in the section of Medical Oncology and shares 1000 sq ft of common equipment space with them.

Office

Dr. Ding has 137 sq ft of office space immediately adjacent to lab. She has 500 sq ft of office space for her staff and students. There is also a central office that is available for Dr. Ding's use, which includes printers, fax, copier, office supplies, administrative support, etc.

Administrative assistance

Dr. Ding has full-time assistance from several administrative personnel located in the departmental administrative offices adjacent to the lab. They help with payroll, purchasing, and grants submission in addition to other needs.

The McDonnell Genome Institute, Washington University in St. Louis

Laboratory

The McDonnell Genome Institute is located at 4444 Forest Park Ave., on the northeast corner of the Washington University Medical Center. Other institutions at the Medical Center include Washington University School of Medicine, Barnes-Jewish and St. Louis Children's Hospitals, Central Institute for the Deaf, Mallinckrodt Institute of Radiology and Imaging, the Goldfarb School of Nursing, and St. Louis College of Pharmacy. Currently, The McDonnell Genome Institute occupies 56,660 square feet of space for laboratory and administrative personnel. This space was designed specifically to accommodate production sequencing activities and includes specialized equipment to maintain strict power and temperature requirements. A 900 square-foot technology development laboratory is available for testing prototype equipment and developing new hardware and biochemistry. A large cold room and walk-in freezer are available on the floor.

The CLIA licensed environment (CLE) occupies 2,412 sq ft of space. As a replicate of the production infrastructure necessary to achieve the expected sequencing goals, the CLE maintains all pertinent equipment within this space. This allows for the completion of all exome capture and sequence generation in a controlled environment. The access to the environment is managed by additional security provided by swipe card and keypads on all doors. The space includes laboratory benches, equipment bays, and desk cubicles for laboratory technicians and managers.

Sequencing & Data Production

The McDonnell Genome Institute operates a state-of-the-art high-throughput, high capacity, next-generation sequencing facility. Our library construction core has the capability to generate 1500 dual-indexed small insert Illumina libraries/week. Multiple quality control measures ensure precise control over insert size and library complexity. We utilize a set of 96 unique dual-indexed library adaptors to allow efficient multiplex processing of samples for hybrid capture and/or DNA/RNA sequencing. Our dual-indexed adaptors with matching 6-8 bp index sequences at both ends prevent cross talk during amplification and/or cluster identification on the Illumina instruments and ensure precise matching of DNA sequences with the sample of origin. The sequencing core currently includes 10 HiSeq Xs, 1 HighSeq 4000, 12 Illumina HiSeq 2000s, and 4 HiSeq 2500s instruments. Our Illumina HiSeq capacity is approximately 215 HiSeq flow cells per/month, which is equivalent to ~1720 human genomes with 30X sequence coverage and 475 HiSeq2500 flow cells per/month, which is equivalent to 3,800 human exomes (Nimblegen VCRome; 38 Mb space with 80% of the target space covered at 20X depth with average mean depth of 60X coverage). We also have 4 Illumina MiSeq instruments, 2 Applied Biosystems 3730 XL instruments, two Pacific Biosystems RS2 systems, and an Ion Torrent PGM sequencer. Robotic systems for sample processing include Perkin Elmer Sciclone G3s, Perkin Elmer Janus system, Eppendorf EpMotion 5075, Covaris LE220 DNA Sonicator, and Perkin Elmer LabChip XT. Sample processing is managed by a custom Oracle based laboratory information management system. The McDonnell

Genome Institute has over 100 touchscreen/bar code scanner stations throughout the building which enable data entry and tracking. All sample containers are physically bar code labeled and continuously tracked to monitor progress and ensure sample integrity at every step of the workflow.

Computing Facilities

The McDonnell Genome Institute has a 15,600 sq. ft. state-of-the-art data center that is located across the street from our main building and was completed in 2010 (222 S. Newstead Ave.). The data center contains fully redundant power and cooling systems capable of housing over 100 racks of high-density network, server and storage systems in its 3,100 sq ft raised floor computer room. Electrical power to the facility is supplied by a nearby, double-ended utility power substation with a backup generator. Redundant cooling is supplied by chilled water systems, delivered under the floor. The center has office accommodations, badge secured entry, secuity cameras and a receiving dock. This data center is the first building on the School of Medicine's campus to receive the LEED Gold status by the US Green Building Council. This was a major challenge for architects and engineers who designed the building because of the energy requirements of specialized cooling systems for the computer equipment. In addition to this data center, a legacy 1,200 square-foot server room at the McDonnell Genome Institute's location (4444 Forest Park Ave.) is equipped with raised floors, redundant power and cooling is utilized for equipment with low power and cooling requirements.

The McDonnell Genome Institute maintains a 1 Gigabit external network link protected by a modern firewall and a 10 Gigabit link on the Internet 2 research network, which is protected by a central Washington University router with a whitelist of collaborating institutions such as CGHub and NCBI sequence data repositories. Within our network, the McDonnell Genome Institute has a highly scalable storage system consisting of over 16 petabytes of raw data storage spread across 23 disk controllers organized into 6 clusters based on usage patterns on our 16 Gigabit SAN network. In addition, MGI has a high performance and highly expandable tape robot managing a tape library of 5 petabytes, which allows the shuttling of data from live disk to much less expensive tape and back again on demand. The McDonnell Genome Institute's computational cluster has 517 servers, 5,926 cores, and 1.1 petabytes of RAM, and runs on average 2.5 million individual computational jobs per month equal to 131 years wall clock time. The newest computational servers have dual 10 Gigabit network links, 40 cores with hyper-threading, 384 gigabyte RAM and 3.2 terabyte local SSD storage per node, which are networked to a redundant 40 Gigabit Ethernet backplane to each storage node. These computational servers and 162 additional operational servers within our computing facilities are managed with automation tools such as 1) PXE+Kickstart for image distribution and boot 2) Puppet+Git+mcollective for server build configuration, change control, and deployment 3) Jenkins+Git for continues integration. The MGI also manages large database instances of Oracle, PostgreSql, and MySQL and utilize many monitor systems such as Zenoss, Nagios, Graphite, Logstash, RTM (LSF), Netflow Collector, OSSEC, Piwik Web Analytics, DBTuna and Google Analytics for maintaining stability, troubleshooting and tuning our systems. To insure continuity of services in the case of a disaster, we have defined service level agreements and nightly backups of critical data, which are stored monthly and retained for one year at an off-site location.

Washington University School of Medicine

The Washington University School of Medicine, consistently ranked in the top 5 medical schools in the United States by U.S. News & World Report and by funding from the National Institutes of Health, has a rich, 122-year scientific history in basic, clinical, and translational research. The Medical School is organized into 20 Departments, 14 clinical Departments and 6 basic science Departments, and includes a total of 1,874 faculty and 1,349 students. Since its founding in 1891, it has trained nearly 8,000 physicians and has contributed groundbreaking discoveries in many areas of medical research. The Medical School also has robust clinical translational infrastructure through its 30 program project or center grants funded by the National Institutes of Health. The School's faculty members are the staff physicians at Barnes-Jewish Hospital and St. Louis Children's Hospital that form the academic hub for the 5,252-bed BJC HealthCare System, the Medical School's hospital partner. The School of Medicine and these fine hospitals, which are perennially recognized for excellence in patient care by U.S. News & World Report and also provide a superb atmosphere for collaborative translational research and for training students, residents, and fellows, are the principal components of the Washington University Medical Center. The compact nature of this 230-acre academic medical center in 12 city blocks enhances the collaborative opportunities for translational research.

Resources - Brown University - Center for Computational Molecular Biology and the Department of Computer Science

Facilities, Equipment, and Other Resources
Dr. Raphael has academic appointments in the Department of Computer Science and the Center for Computational Molecular Biology (CCMB) at Brown University. Both entities reside in the same building. Dr. Raphael has an office in the Computer Science Department and computer lab space in the CCMB. The CCMB has 2,098 square feet of office space. These offices are equipped with 22 workstations.

Center for Computational Molecular Biology (CCMB):

OSCAR Core Compute Cluster: The Brown University Center for Computing and Visualization (CCV) maintains a high-performance computing system for campus use. The nucleus of the system is a 166-node Linux cluster purchased from IBM in the Fall of 2009. Each node has two Intel Xeon 5540 (2.53 GHz) quad-core Nehalem processors, 24 Gigabytes of DDR-3 memory (1333 GHz) and a 40 Gb/s Quad-Data-Rate (QDR) Infiniband messaging interface
Large Memory Cluster: In addition to the dual-processor Nehalem nodes, there are 7 quad-processor AMD Opteron nodes with Shanghai-series processors clocked at 2.6 GHz. Six of these nodes have 64 Gigabytes of memory. The seventh has 128 Gigabytes. All have QDR Infiniband interfaces.

The CCMB has purchased dedicated access to 144 cores on the OSCAR computer cluster, under a "condominium" model.

In addition, a web server consisting of 4 dual core AMD Opteron processors and 16 GB memory hosts web applications developed in the CCMB for public distribution.

The Brown University Department of Computer Science:
The department provides leading-edge computing technology to all its Faculty and students. We have over 500 desktop systems running Linux, OSX or Windows. Most of these are custom-built machines configured and assembled by the department's technical staff. Components include quad-core processors with up to 16GB of memory and dual 19" or single 24" flat-panel monitors. These systems are connected to the department's 1Gb/s switched Ethernet network with access to both Internet1 and Internet2 via the University's fiber-optic backbone. An 802.11n (300Mb/s) wireless network is accessible throughout the department.

The department has three electronic classrooms. The first, a banked auditorium, holds seventy-three systems running Linux. This room serves as the primary computer facility for undergraduate computer science students. The second contains twenty-five seats, each with a Microsoft Windows and Linux system. The layout of this space makes it an ideal room for sections, seminars, and interactive learning. The third contains an advanced audiovisual system that supports recording and streaming of lectures and talks. In its default configuration, it accommodates 50 people, but the back wall opens into an atrium, equipped with large screen TVs and speakers providing overflow seating for around 200. Five research labs further enrich the environment with specialized hardware and advanced workstations from a variety of vendors.

Desktop and research systems are supported by a data center with fully redundant servers that offer a wide range of services. Central file storage is built upon IBM's General Parallel File System (GPFS). This approach provides a scalable, high performance, cost effective solution based on IBM hardware and currently hosts approximately 343TB of RAID-6 storage. A Sun Grid Engine cluster of 181 computational servers running Linux provides 1828 cores and over 4TB of combined memory. The most powerful of these are quad processor systems each with a total of 64 cores and 256GB of memory.

Resources - New York University School of Medicine

The School of Medicine has a proud history dating back to its founding in 1841. Over our history, NYU scientists have produced groundbreaking discoveries, some of which have led to Nobel Prizes, and all of which have helped advance the diagnosis and treatment of disease. Our researchers have been at the forefront of

devising creative and effective solutions to some of the world's most complex healthcare issues. In the last three decades our researchers have registered more than 1,450 inventions and been issued more than 650 patents, more than half of which have been licensed. We have ranked first among U.S. universities in income from technology licensing over the past five years. Our extraordinary record puts us in a leadership position in what has become an increasingly important role for academic medical centers: the transformation of laboratory research into new diagnostics, medicines, and devices. NIH funded more than seventy percent of the $263 million in research grants awarded to NYU Langone Medical Center in the fiscal year of 2010. The same year, baseline NIH funding grew from $126.4 million to $139.8 million – a ten percent increase from the previous year and a five-fold rate increase compared to other medical schools around the country.

NYU Langone Medical Center

NYU Langone Medical Center is one of the nation's premier centers of excellence in healthcare, biomedical research, and medical education. Located in Manhattan, NYU Langone consists of three hospitals—Tisch Hospital, a 705-bed acute-care tertiary facility; Rusk Institute of Rehabilitation Medicine, the first rehabilitation hospital in the world, with extensive inpatient and outpatient rehabilitation programs; and the 190-bed Hospital for Joint Diseases, one of only five hospitals in the world dedicated to orthopaedics and rheumatology—plus the NYU School of Medicine, one of the nation's preeminent academic institutions and one of 18 schools and colleges of New York University. In addition, NYU Langone Medical Center offers ambulatory-care services in various Manhattan neighborhoods, the outer boroughs, Long Island, New Jersey and Westchester County, bringing services directly to where our patients live and work. NYU Langone's medical students, residents, and faculty also provide patient care at Bellevue Hospital Center, the nation's oldest public hospital, and the Medical Center is affiliated with Woodhull Hospital in Brooklyn, Gouverneur Healthcare Services in Manhattan, and the New York Harbor Veterans Affairs Medical Center. The Medical Center's trifold mission to serve, teach, and discover is achieved on a daily basis through the seamless integration of an academic culture devoted to excellence in patient care, education and research.

NYU Langone Medical Center (consisting of New York University School of Medicine and NYU Hospitals Center) and the broader New York University community offer a physical, organizational and intellectual infrastructure that creates a unique and rich environment to foster and support research, education, and clinical care. NYU Langone has over 50 centers, more than 20 core facilities and shared technology resources, 29 academic departments in the clinical and basic sciences, and 533,000 square feet of research space as of fall 2011. During the 2010-2011 academic year, the School of Medicine faculty consisted of 1,360 full-time and 2,175 part-time members, including nearly 100 endowed professors. The student and trainee population consisted of 736 MD candidates, 68 MD/PhD candidates, 248 PhD candidates, 1,125 residents and fellows, 381 postdoctoral fellows, and nearly 5,000 registrants in postgraduate continuing medical education courses.

Bioinformatics Resources

Space:

The Center for Health Informatics and Bioinformatics (CHIBI) and the associated Graduate Training Program in Biomedical Informatics is located in the newly renovated space at 227 E. 30th Street on the 7th Floor. CHIBI has a total of 20 private, windowed offices (approx 12x13 feet) for faculty and staff, 30 cubicles (approx 8X6 feet) of which 26 are available for the exclusive use of pre and postdoctoral trainees and an additional 8 cubicles (approx 5x4 feet) for trainee and staff use. The entire floor is exposed to natural light owing to the innovative architectural design, which allows light entering windows from faculty offices along the floor's perimeter to flow into the internal space. In addition, there are two large conference rooms on the floor that accommodate more than 30 people each comfortably. These rooms are used for meeting, teaching and presentation purposes. Both rooms are fully equipped with state-of-the-art audiovisual equipment and include videoconferencing and teleconferencing capabilities and built-in computers and projectors for presentation purposes. On the first floor of 227 E. 30th Street is a large auditorium that will be used for lectures, grand rounds, and various talks that draw large audiences. This auditorium will similarly be equipped with state-of-the-art equipment for conferences and presentations. The location is electronically secured. CHIBI provides 2 shared printers, a fax machine, copier/scanner. The location is wired and connected to Internet through NYULMC secure network.

High Performance Computing:

CHIBI owns/manages a state of the art HPC facility currently includes:

Data Storage: **1 PB (petabytes) of scalable high-bandwidth Isilon storage system**. This system is critical for the operation of next-gen research such as projects utilizing high throughput sequencing, RNAi screen assays, microarrays, proteomics, and other high-throughput assays.

Computing Power:

The HPC cluster consists of the following:

- 2 head nodes,
- 64 compute nodes (with 16 of the latest intel cores and 128GB of RAM)
- 5 additional compute nodes with the latest NVIDIA Graphics Processing Unit (GPU) accelerators
- A high memory node with 1TeraByte of RAM
- Network Attach Storage (NAS) system with 80 TB of disk storage accessible by the cluster nodes only
- All cluster nodes are interconnected by a private 10Gbit network and have dedicated, high-bandwidth 10Gbit access to the primary data storage.

Physical Location: The data storage and computer cluster of the HPC are located in a secure data center protected by both the medical center enterprise firewall and HPC-specific security using two-factor authentication.

Affiliated Personnel:

i. HPC architecture & management: 2 dedicated FTEs
ii. Storage, network, power/cooling, cybersecurity and backup services provided by MCIT department through multiple shared FTEs
iii. Informatics faculty with operational and scientific roles: 5 faculty
iv. Informatics faculty with parallel algorithm development expertise for high-dimensional complex data analytics: 2 faculty
v. Informatics scientific and administrative staff (shared): 2 administrative and 4 analyst/programmers

Software: A large variety of bioinformatics, data analysis and other software licenses and codes are supported.

Best Practices Integrative Informatics Core (BPIC):

BPIC provides a one-stop consultation service to address the expanding needs for high-level bioinformatics, high dimensionality data mining, database management, and health informatics services required by NYULMC research efforts. BPIC is designed to integrate pre-existing bioinformatics resources of NYU and in the future establish strong working relationships with NYU academic programs in mathematics, computer science and biology. BPIC Objectives:

- Identify and prioritize specific informatics research support needs across the medical center.
- Provide researchers with access to selected set of state-of-the-art bioinformatics tools and the resources necessary for data storage and computation to apply these tools to basic science, translational and clinical projects.
- Benchmark internally- and externally-developed methods, study and synthesize related literature to develop/disseminate existing best practice guidelines.
- Execute analyses or software on behalf of researchers, construct custom data analysis pipelines, or point to external resources capable of so doing.
- Provide expert bioinformatics and health informatics consultation and training to researchers for all stages of research from design to execution, to publication.
- Provide a dissemination channel and evaluation test bed for novel informatics methods developed by NYU researchers (e.g., in Computational Biology, Mathematics, Biostatistics, and Computer Science).
- Identify needs for new methods development and communicate these needs to the method developers closing the loop between methods development conception and deployment.
- Compile a bioinformatics reference resource with answers/solutions to common questions and research needs.