RESOURCES

Follow the 398 application instructions in Part I, 4.7 Resources.

The GRCF occupies about approximately 4,000 sf of laboratory and office space on 10th floor of the Blalock building, 6,000 sf at an off-site location (Lighthouse Point), 19,000 sf shared space with the Center for Inherited Disease Research in the Triad bldg. at the Bayview campus, and ~2,000 sf of LN2 repository space at the Fayette St. loading dock.

JHU Cell Center:
The tissue culture laboratory is a CAP accredited facility on the Johns Hopkins Hospital campus. The facility includes 3 biological safety cabinets and 6 ThermoFisher humidified incubators in 2 negative air pressure rooms. Cells are quantified and quality control monitored by way of 2 Vi-Cell XRs. The Rees Environmental Monitoring System continually monitors the cell culture conditions. Single cell genomics is accomplished using Fluidigm's C1 Single Cell Auto Prep. Single cell isolates are then analyzed through JHU genotyping laboratory and JHU DNA sequencing laboratory.

JHU Biorepository:
The Biorepository is a CAP accredited facility that houses 2 Taylor Warton LABS 80K LN2 vapor phase freezers, 6 Taylor Warton LABS 40K LN2 vapor phase freezers, 2 MVE High Efficiency LN2 vapor phase freezers and 2 ThermoFisher 20CUFT chest freezers. For cryogenic transport of biospecimens, the Biorepository utilizes 2 portable LabRep Co. Liquid Nitrogen Cryocarts. Freezer security and integrity is maintained through the Rees Environmental Monitoring System, with an automated Centron monitor. Repository inventories are maintained through the Freezerworks database.

JHU genotyping laboratory:
The JHU lab houses a variety of major equipment used for sample receipt, sample processing, sample pretesting, SNP GWAS, SNP linkage, SNP custom genotyping, epigenetic methylation assays and NGS. Equipment for sample intake includes 1 Nanodrop ND-1000 and 1 Nanodrop 8000 Spectrophotometer, 1 BioMicroLab XL-20 and 1 BioMicroLab VolumeCheck instrument. Genotyping equipment includes 3 Illumina iScan scanners and 2 Illumina HiScan scanners with 3 autoloaders (capable of loading BeadChips on 2 scanners) for 24/7 continuous BeadChip scanning. Two Illumina BeadXpress readers are used for sample pretesting and small custom projects because this platform has lower costs per genotype for small numbers of SNPs per sample.
The Fragment Analysis Facility (FAF), a laboratory within the Genetic Resources Core Facility (GRCF), consists of approximately 6000 sq. ft of space at the Lighthouse Point facility located approximately 1.5 miles from the Johns Hopkins Medical school campus. The FAF houses equipment for both STR and SNP genotyping, nucleic acid extraction, and mycoplasma detection services. Equipment includes a Taqman 7900HT, four Veriti (Applied Biosystems) thermocyclers, a Qiagen Autopure DNA extractor, a QIAcube that provides automated processing of Qiagen spin columns, a Hamilton Microlab STAR liquid handling workstation, Integra Biosciences VIAFLO 96 and a CheckScanner for reading mycoplasma MycoDtect microarrays. The lab also has -80 and -20 freezer storage, various electrophoresis and centrifugation equipment, a Nanodrop 2000 spectrophotometer, a Spectramax Gemini XS UV plate reader, a Fragment Analyzer (Advanced Analytical Technologies) for nucleic acid quality, quantity and sizing analysis, and a Millipore water purification system. The FAF also has access to the DNA Analysis Facility's 3730 XL DNA Analyzers for fluorescent fragment analysis.

JHU DNA High Throughput sequencing:
Sequencing equipment includes 1 Covaris E210, 1 Covaris S2, 1 Caliper LabChip GX, 2 Agilent BioAnalyzers, 3 Illumina HiSeq 2500 instruments, 2 Illumina cBot instruments and 2 Illumina MiSeqs that are used for validation and optimization experiments as well as to support CLIA services (see below). The lab also has one IonProton being evaluated for WES. Liquid handlers include 9 modified Tecan Genesis/ Evo/EvoII/200 instruments, 2 Perkin Elmer Multiprobes, 1 Perkin Elmer Janus, 1 Agilent Bravo, 1 Beckman Biomek FX, 1 Caliper Sciclone, 2 96-well Robbins Hydras and 3 Biomek 2000 instruments. A Veriteq alarm system monitors the freezers and laboratory environment and alerts laboratory staff 24/7 of freezer failures, power failures or the presence of water on the floors. All freezers containing DNA samples or key reagents

are alarmed using this system. The entire building is protected by back-up emergency generator power. Additional resources and equipment for DNA extraction, bisulfite conversion, real time PCR, Sanger sequencing and pyrosequencing are available at a separate location.

Informatics Infrastructure: The lab has been continuously expanding and improving our informatics infrastructure, including our data movement and storage capacity, computational and network capacity as well as server room space and power. The increase in services and explosion in data volume has led to rapid proliferation of servers and storage equipment. To accommodate this growth, we expanded our original server room into adjoining office space in 2008, nearly doubling the size to 315 sf. A third 40kVA UPS and new power panel is currently being added to this server room, providing a total of 115kVA to address power limitations. Additional space is also now being renovated, which will add another 681 sf. The racks, UPS units and A/C systems needed to complete these renovations are budgeted for Year One of Task One. Available building power was recently doubled and sufficient emergency generator capacity was added to power all building systems (including HVAC) in the event of a city power outage. The new generators operated successfully through multiple outages during the summer thunderstorm season of 2011.
IT Systems: The primary high-performance sequence analysis platform is a cluster of 6 Dell C6100 systems running Sun Grid Engine connected to Isilon NAS via 10-gigabit ethernet. Each C6100 comprises 4 servers for a total of 24 cluster nodes, 576 cores, 2.3TBs of RAM, and 43TB of local disk. This allows us to complete analysis for sequencing runs in 24-36 hours. Additional capacity can be added in 4-server increments to meet increased demand. Two bladesystems provide high-density computing platforms with shared power and network backplane. An HP C-class bladesystem currently has 9 (expandable to 16) multi-core blades, each of which is a slim but fully-functional server. These are primarily set up as powerful “remote workstations” that are especially useful to run large and/or long-running applications without tying up users’ desktop PCs – for instance, processing the largest Illumina Infinium projects. In addition, a previous-generation HP Proliant bladeserver contains 10 dual-processor blades used as dedicated servers for various purposes. As these are migrated to the new blades or VMs or decommissioned, this bladeserver will be decommissioned. Virtualization is now a key lab IT strategy because it saves space, money, power and minimizes the management overhead associated with server proliferation. VMware ESX is increasingly the first choice for Linux and Windows servers due to ease of setup and quick availability. VMware runs on three Sun 16-core X4450 hosts for redundancy and to minimize downtime by live migration of VMs between hosts using vMotion. OS virtualization (e.g., OpenVZ) is also used where it makes good sense.
All raw and analytical data is stored on a robust, high-performance Isilon network-attached storage (NAS) system, which currently consists of 720TB total raw capacity, an accelerator node and a backup node. The entire system has a total bandwidth approaching 250Gb/s. Automated balancing of network connections among ports prevents overloading of individual nodes and maximizes network throughput. Storage space and throughput is increased by adding new nodes with no service interruption. With current protection level settings, 1 disk and 1 entire node can be lost with no data loss and thus far we have never lost any data using this system. Critical data is further protected by “snapshot” backups to disk, and the backup node enables NDMP dumps at very high speed over optical fiber directly to tape on the i500 system. Server disk is provisioned from a new Xiotech Emprise 7000 storage area network system (SANS) with 53TB of fiber channel disk and multiple controllers providing redundant data access paths. Xiotech uses patented self-healing “DataPacs” of 10 disks and 2 spares, which are guaranteed not to fail for 5 years. In 2011 this storage capacity was almost doubled, and currently has the ability to nearly double again. Windows systems include 2 Dell servers running multi-threaded AutoCall genotype calling pipeline, 3 large shared fileservers, 3 directory/domain servers, 2 LIMS servers, 2 backup servers, a terminal server for thin clients used by research techs, 2 Oracle application servers for SQL*LIMS and several utility servers. Two large-memory HP workstations are available for computation of large genotyping projects and other demanding applications. There are approximately 75 other PCs, including 50 user desktop machines, all of which are centrally managed. Other IT equipment includes 4 Dell/Ocarina appliances that use proprietary algorithms to continuously compress static data with very high efficiency to reduce disk space usage, 2 Sun X4150 Linux servers as our Oracle 11g RAC platform, a 16-core, 256GB RAM Sun X4600 and 2 smaller Sun general purpose Linux servers, two Sun X4500 48TB storage servers, a HP Alphaserver cluster comprised of HP ES-47 and ES40 systems (database and compute servers), a Linux virtualization server (OpenVZ) and several Linux workstations. CIDR has two high-capacity network printers, several individual desktop printers and dedicated barcode/label printers in the labs. Hand scanners for barcode labels are available at all workstations.

All production computer systems and critical desktops/workstations are routinely backed up to tape on daily, weekly, and/or monthly schedules using CommVault Sympana 9 enterprise backup software. Several CommVault media servers are directly connected via fiberchannel to a Quantum i500 robotic tape system containing 10 ultra-highspeed high-capacity LTO-5 tape drives and 125 tape library slots for continuous operation. Periodic backups are rotated weekly to a secure off-site storage facility in Frederick, MD, while local copies of backup tapes (and other media, such as extras copies of projects released on portable hard drives) are kept on-site in a large fireproof safe.

Software Systems: The CIDR contract has allowed us to develop a robust IT capability and custom software to manage both genotyping and sequencing studies. The principal pipeline used for NGS is CIDRSeqSuite which was developed to automate a variety of bioinformatic workflows and analyses for sequencing and will be used in the prospective phase of this study. It is designed to run on 64-bit Linux and is written in Java SE 6 and Perl 5. The primary analysis workflow combines custom tools to numerous third-party secondary and tertiary analysis software including ANNOVAR9 , SAMtools 10, Picard 11, BWA 12, the Genome Analysis Toolkit (GATK)11, VCFtools and Tabix 13. Beginning with the qseq.txt files produced from binary basecall files (BCL) files by the Illumina BCL converter, CIDRSeqSuite demultiplexes samples based on their indexes, then combines and converts them into FASTQ files. From there, independent analyses by sample are launched. This analysis includes aligning the FASTQ files with BWA to a reference genome for a paired-end sequencing run, sorting of the BAM with Picard, local realignment around Indels using GATK, synchronizing mate-pair information and flagging of molecular duplicates with Picard, recalibration of base call quality scores with GATK, variant calling with GATK UnifiedGenotyper and annotation of variants with ANNOVAR. To manage the 1,000's of files generated at each step in the pipeline, CIDRSeqSuite automatically renames and moves files by project into a central location. CIDRSeqSuite parallelizes these independent analyses across a Sun Grid Engine-managed cluster.

Data Security: Sequence received from Illumina will arrive as encrypted hard drives. At JHU, the systems and network are protected by local and NIH/CIT-managed firewalls, and are exhaustively and continuously monitored for performance, problems and potential security compromise, with automated notification systems that contact responsible staff by telephone, email and/or text message. We are in compliance with both NIH and university data security standards.

JHU DNA Services

This laboratory within the GRCF is located in ~400 sf the 10th floor of the Blalock building on the Johns Hopkins Medical Campus. This laboratory houses equipment for qPCR, digital PCR and medium throughput genotyping (2 QuantStudio 12K Flex machines, one Taqman 7900HT and one QuantStudio 3D), Sanger sequencing (2 3730XL machines), and Pyrosequencing (Pyromark Q24).