Yale Center for Genome Analysis- Resources

The Yale Center for Genome Analysis (YCGA) is a state-of-the-art DNA Sequencing Center launched in 2010 on Yale's West Campus to provide a centralized facility for services, equipment, and expertise required for large-scale sequence analysis studies. Yale has allocated the entire building to YCGA with over 9000 sq. ft. of custom-designed laboratory and office space with all modern amenities. Essential infrastructure such as uninterrupted and emergency power supply critical for the 24/7 operation of the sequencers, high-performance clusters (HPC), data storage file systems, and a high-speed ethernet connection required for efficient data transfer is already in place for our current sequencing platforms. Since the laboratory temperature is critical for the optimum performance of Illumina Genome Analyzers, Yale West Campus central facility staff constantly monitors and maintains a temperature close to 21 ^{0} C. The YCGA is very well supported by Yale Information Technology Services (ITS). Approximately 7000 sq. ft. of the area is occupied by the laboratory that houses both sequencing and other associated instrumentation, a walk-in 40C cold room, and a dedicated area for library preparation with additional unoccupied space that will be used for future expansion. The office and the administrative regions occupy approximately 2,500 sq. ft. of space. A glass wall completely separates the laboratory and office spaces for safety reasons. The laboratory can be accessed from the office area through two glass doors situated at each end of the laboratory. YCGA has 12 offices that Ph.D.-level and senior staff mostly use. The administrative area is used mainly by the laboratory technicians to carry out computer work and to hold LIMS and other high-end computers used for high throughput analysis. One conference room that has a projector provides venues for interaction.

The list of the major sequencing and microarray instruments available at YCGA is as follows:

1. Two Illumina NovaSeq6000 sequencing systems
2. One Illumina NovaSeq X plus sequencing system
3. One Illumina NeXTs 2000 sequencing system
4. One Element AVITI sequencing system
5. One Pacific Bioscience Sequel II sequencing system
6. One Pacific Bioscience Ravio sequencing system
7. One Illumina MiSeq
8. One Ion torrent PGM system
9. Three 10 x genomics Single Cell Genomics System
10. One Mission Bio single-cell genomics platform
11. MERFISH spatial multiomics system
12. 10X Xenium Spatial Transcriptomics System
13. Nanostring CosMx Spatial Transcriptomics System
14. CurioSeeker technology for Spatial transcriptomics
15. Altasxomics technology for spatial ATAC-Seq
16. Lenovo Performance Cluster, with a total of 4,624 cores/CPUs
17. 4.5 Petabytes of Panasas and Hitachi Data systems storage
18. Two Covaris E210 and R230 focus acoustic energy shearing systems
19. One Hamilton NGS STAR system for library prep
20. PerkinElmer Sciclone G3 NGSx system for library prep
21. PerkinElmer Janus liquid handling system
22. Two Integra VIOFLOW pipetting systems
23. One Formulatrix Mantis liquid dispensing system
24. Caliper LabChip GX system
25. Two Agilent TapeStation 4200 systems
26. Two Roche LightCycler 480 ll qPCR systems
27. One Fisher QuantStudio Absolute Q Digital PCR System
28. One BioRad CFX Opus 96 RT-PCR system
29. 16 BioRad 96-well Thermo Cyclers
30. Three Agilent Bioanalyzers for Q/C of RNA, cDNA, cRNA, and fragmented cRNA
31. Fragment analyzer capillary electrophoresis system
32. Eppendorf epMotion 5075LH automated liquid handler
33. Fisher scientific Kingfisher automation for nucleic acid extraction

34. Molecular Devices SpectraMax II 96/384-well spectrophotometer
35. Illumina throughput microarray system with one Tecan robotics and HiScans with an autoloader.
36. Affymetrix GeneChip Model 3000 Scanners with autoloaders
37. Three Affymetrix GeneChip Model 450 fluidics stations
38. Eco Revolve fluorescent microscope
39. MERFISH technology
40. More than 50 computer workstations and laptops.

Laboratory Information Management System at YCGA: All aspects of data generation, including sample collection, library preparation, sequencing, and billing, are tracked using the YCGA LIMS system. YCGA LIMS system consists of ilab and SLIMs modules from Agilent customized based on the workflow needs of YCGA. YCGA LIMS system is an online, flexible, scalable, and password-protected database integrated with Yale's CAS (central authentication system) and central billing system. ilab system tracks samples, projects, and billing and can create reports on billing, resource utilization, and projects per need. This LIMS system provides sample status information to users and YCGA staff throughout the sample's passage in the library preparation and sequencing workflow, including primary analysis of the sequence data and its delivery to the end user. Before receiving the DNA samples, the investigator or laboratory staff enters the metadata associated with each sample, including analysis parameters. The user-set parameters of the reference genome and analysis type are selected from drop-down menus when the samples are submitted to YCGA. All subsequent critical aspects of sample tracking, namely the sample to flowcell lane assignments and associated results files, are entered and monitored via this SLIMS system by the Library prep, sequencing, and Bioinformatics teams of YCGA.
Additionally, the machine runs can be monitored remotely through the LIMS system. Once the primary analysis is completed and assessed for quality, an automated email is sent to the end user and the Principal investigator. The YCGA LIMS is expandable to include any range of data outputs and multiple sequencing platforms to handle rapidly evolving second and third-generation sequencing technologies and new genomics-based workflows.

YCGA High-Performance Computation (HPC): DNA sequence data generated by YCGA's Sequencers are directly transferred to YCGA's HPC for further analysis. The YCGA HPC consists of 400 nodes with 4614 cores/CPUs, sixteen GPU's, and approximately five Petabytes of high-performance parallel storage (Panasas Inc.; Hitachi Data Systems); it runs a Linux operating system. All machines are connected via gigabit Ethernet. Hardware and software support for the HPC is provided by ITS and is housed in a secure, card-key controlled, state-of-the-art machine room at Yale's West Campus facility. The components are installed in an APC modular pod that provides power, targeted in-line cooling, and fire suppression. ITS is installing a tape archival service to use the high-performance storage efficiently while ensuring all data is adequately backed up. The cluster has redundant login nodes (two nodes with failover) and uses virtual machines as login hosts, all designed to prevent access to the cluster by rogue user processes. Industry-standard cluster management techniques and software packages are used to maintain and control the cluster. The nodes run CentOS, an open-source version of Linux. Node operating systems are provisioned using Rocks. Individual users are scheduled onto nodes using Maui Torque, a batch queuing system. Nagios and Ganglia are used to monitor the state of the cluster 24x7. The sequencers at YCGA communicate with the HPC cluster via a dedicated fiber link between a Force 10 switch on the cluster and an HP Proliant switch in the sequencing center building. This effectively links the two facilities into a single high-speed network (1 Gigabit/sec, with 10 GB links to the Panasas storage units). The entire facility sits behind a firewall that only allows SSH-authenticated access by users. Two Ph.D. computer scientists and four system administrator staff from Yale ITS oversee the operation management of HPC. Several useful packages are installed on both clusters, including LAM mpi, mpich, R, java, Intel, and PGI compilers.

YCGA staff: The YCGA has 34 staff members highly trained in high-throughput sequencing, HPC, and data analysis. This includes 13 Ph.D. and six MS-level staff. The day-to-day operation is overseen by Dr. Bony Kumar, who has more than 15 years of experience in molecular genetics technologies and analysis. He leads nine scientists involved in sequence production. Four systems administrators maintain the IT infrastructure. It is overseen by Rob Bjornson, Ph.D., and Jason Ignatius, M.S., who have extensive parallel computing experience. Eight Ph.D. scientists provide the bioinformatics support.

Bioinformatics: The bioinformatics support for analyzing sequence data generated at YCGA is provided by eight Ph.D. scientists. The bioinformatics staff has extensive experience developing new programs to study data generated by next-generation sequencing platforms. It conducts high-level analyses and develops new algorithms and data analysis tools. They work closely with the HPC team to manage data storage and develop analysis pipelines. They also guide other bioinformatics staff working in individual laboratories.

Services provided by YCGA: YCGA supports diverse sequencing applications, including Whole-exome (WES), whole-genome (WGS), targeted sequencing, mRNA-Seq, ChIP-Seq, methyl-Seq, microRNA-Seq, and de novo genome sequencing. YCGA also provides library preparation and sequencing services for CUT & Tag, HiC, and ATAC-Seq. YCGA supports SNP validation and genotyping services using droplet PCR and NGS platforms. It provides microarray genotyping, expression, and methylation analysis using Affymetrix and Illumina microarray platforms. YCGA also offers spatial transcriptomics through 10X Xenium, Nanostring CosMX, and CuriSeeker(Curio Biosciences). Spatial ATAC-Seq is offered using the AtlasXomics.Since 2010, YCGA has carried out sequence analyses (library prep, sequencing, and analysis) for 397 Yale and 419 non-Yale investigators from 158 national and 80 international institutions, including the Broad, Harvard, UCLA, Columbia, Duke, MIT, Memorial Sloan Catering Cancer Center, UCLA, and John Hopkins. YCGA has sequenced >130,000 exomes and several thousands of samples for other applications. It has been approved for CLIA and CAP certification since 2012. In addition to providing sequencing analyses, YCGA staff actively participate in developing new technologies, work collaboratively with the Yale investigators, and provide much-needed grant-writing support to its users.

Scientific accomplishment: More than 1200 publications (that we know of) have resulted from using the resources provided by YCGA. More than 70 publications have resulted in high-profile journals, including Science, Nature, Cell, New England Journal of Medicine, and Nature Genetics. The facility has shown scientific leadership by publishing its first significant publication after the completion of the human genome project that discovered the association of complement factor H with age-related macular degeneration in 2005. (Science, 2005). Similarly, YCGA was the first to perform clinical diagnosis using exome sequencing in 2009.

Financial management/Grant support: YCGA operates on a cost-recovery basis. Since its inception, YCGA's budget has been balanced. Yale University and the School of Medicine support YCGA's mission and have generously provided millions of dollars in funds to acquire new technologies and maintain cutting-edge genomics. YCGA users have received over 250 million dollars in grant funding from multiple extramural funding agencies. Dr. Mane, the director of YCGA, has received over $35 million in funding as a principal investigator from NIH and other funding agencies. YCGA is the home of NHGRI/Yale Center for Mendelian Genomics(YCMG), one of the four national centers established in November 2011 and then renewed in 2016 (http://www.mendelian.org/). It is supported by a $22 million grant from NIH/NHGRI. The YCMG applies NGS and computational approaches to discover the genes and variants that underlie Mendelian conditions, including skin diseases, intellectual disability, intracranial aneurysm, migraine, musculoskeletal malformations, Dent's disease, electrolyte homeostasis, polycystic kidney disease, cardiovascular disorders, Kidney and inherited eye diseases and more. Over the past eight years, the Center has analyzed over 35,000 samples for 110 NIH-funded investigators from 80 institutions from 13 countries, and these efforts have led to several high-profile publications, including Nature, Science, and Cell. In addition to YCMG, YCGA is closely associated with and serves as a gnomic core for the Yale O'Brien Kidney Center, Yale Cancer Center, Yale Center for Clinical Investigation, and Yale/Gilead Cancer Research Collaboration.