

Description of Centralized IIGB facilities.

The research laboratories of the Institute for Integrative Genome Biology (IIGB) at the UCR are housed in \$53.8 million, 64,000 ft² interdisciplinary, four-story Genomics Building that can accommodate 200 faculty members, graduate students, postdoctoral and staff researchers. The overall Director of IIGB is Dr. Natasha Raikhel. The building unites bioinformaticians and life scientists who use modern genetics, genomics, chemical biology, and bioinformatics to address basic and applied questions in disease, plant, insect, and fungal sciences. The labs and student/researcher office spaces are open to stimulate interaction between different research groups, and the 100-seat auditorium hosts symposiums and trainings in genomics-related research. Adjacent to the Genomics Building is a modern 10,000 ft² instrumentation facility (Keen Hall) that offers expertise and instrumentation in microscopy/imaging, proteomics and genomics (<http://genomics.ucr.edu/>). All facilities utilize a custom on-line password protected customer instrument reservation and automated billing (<http://iigbreservation.ucr.edu/>). Programmable customer-specific electronic locks, permit UCR researchers 24/7 access to all IIGB research instrumentation, enhancing productivity and security. All core facilities offer free training areas including real time PCR, flow cytometry, confocal microscopy, Illumina sample preparation, data analysis utilizing R-based Bioconductor tools and Galaxy, and proteomics sample preparation, among others. We also organize laboratory and lecture workshops on selected topics and provide commonly used reagents such as qPCR mixes at discounted prices to foster research.

I. Genomics Core Facility. The Genomics Core Facility at IIGB provides technical, instrument, and professional development.

a) Next Generation Sequencing. The Genomics Core houses three an Illumina sequencing systems. The HiSeq2500 platform that can generate up to 250bp paired-end sequences. UCR was one of the first institutions on the West Coast to offer HiSeq services which was supported by a NIH high-end MRI award that included essential upgrades to our compute infrastructure including increased CPUs and storage for our compute cluster to process the increased data stream. This is aided by a semi-automated workflow of the HiSeq2500 and the C-bot for cluster generation in high output mode. The current yields of the HiSeq permits multiplexing to reduce the cost per sample cost for researchers at UCR. In addition to the HiSeq, the facility houses an Illumina MiSeq instrument which is capable of read lengths up to 300bp paired-end (600 bases total). The MiSeq is ideal for lower yield applications such as rDNA and other amplicon sequencing as well as exome projects where speed and read length are key considerations. The facility recently acquired a new Illumina NextSeq500 sequencer which expands the NGS offerings. This instrument rapidly runs single sample lanes but at twice the yield (~400 million reads) as a HiSeq lane. The NextSeq adds flexibility in run types and capacity. It also operates very quickly reducing sample processing times for campus researchers. All projects are managed via our secure web portal which includes information about our services and secure uploading of sample information and downloading of data. Beyond instrumentation, a high degree of expertise is required to provide high quality results. At IIGB, in addition to IIGB Research Facilities Director, Glenn Hicks, we employ a full-time Specialist (John Weger) with in-depth knowledge dedicated to Illumina instruments. Other personnel including staff scientists have in-depth knowledge of NSG providing a wealth of experience ensuring the highest quality services. Full bioinformatics support is crucial and is provided by a dedicated programmer and systems administrator as well as the Director of Bioinformatics (Dr. Rakesh Kaundal) who is available for consultation with researchers. The Bioinformatics Core also provides a broad range of free workshops on data analysis that provides a strong training component to successful grants. For custom data analysis services, programming resources are available within the IIGB Bioinformatics Core. Our genomic specialists and programmers work closely as a group to provide free consultation in experimental design and setup. To assist in sample prep, the Genomics Core houses instrumentation for DNA fragmentation

(Covaris S220, Bioruptor), size selection (eGel system, Blue Pipin), quantification (qPCR, Qubit), and the quality control of libraries (Agilent Bioanalyzer and recently acquired Advanced Analytical Sciences Fragment Analyzer for higher throughput QC). We also provide guidance in all aspects of NGS and sponsor hands-on laboratory workshops to provide training in sample prep. The Genomics Core also provides sample prep services for novice labs or those who do not want to devote resources to preparing samples in their labs. IIGB has experience with RNASeq small RNA, genomic, ChIPseq, bisulfite sequencing, and custom library applications. We recently acquired the latest Illumina Neoprep sample prep robot (<http://www.illumina.com/systems/neoprep-library-system.html> <http://www.illumina.com/systems/neoprep-library-system.html>) which is in service for DNA sample prep and will enter service for RNASeq in February 2016. The small, reagent cassette-based device can rapidly prep up to 16 DNA and RNASeq samples at a time in an automated manner including quantification for sample pooling at lower cost. For sample information upload, data download and information about services, we maintain an Illumina web portal with secure access to UCR Illumina customers (<http://illumina.ucr.edu>). We provide additional workshops, starter grants, and the hosting of campus-wide seminars in NGS to promote methods sharing and broader campus-wide research. Overall, our facility for NGS has maintained cutting edge technology focusing on appropriate technologies that match the needs of UCR researchers at competitive re-charge rates.

b) Other gene expression services. For routine Sanger sequencing, the facility utilizes a 96-capillary sequencer (ABI 3730XL), whereas for genotyping, amplified microsatellites can be detected using a 16-capillary instrument (ABI 3130XL) and software (ABI GeneMapper). For gene expression, microarray services have not been replaced by NGS for all applications. For commercial Affymetrix arrays, hybridization (Hybridization Oven 640), staining and washing (Fluidics Station 450), and quality control of RNAs (Agilent Bioanalyzer, Fragment Analyzer) are offered. For array data analysis Molecular Devices Acuity and Stratagene ArrayAssist are available. For nucleic acid and protein blots, a Typhoon 9410 scanner (GE) with a resolution of 10-500 μ can quantify images from gels, blots and other formats utilizing fluorescence, chemiluminescence or phosphoimaging. The core has a gel doc (Biorad Easy Doc and imaging software) with UV, white light and other filters. For array scanning the core offers a two-color GenePix 4000B (Molecular Devices) and a five color ScannArray Express (Perkin Elmer). Gene expression can be validated by quantitative PCR using one of six Biorad real time qPCR instruments (CFX96, CFX Connect x 2, iQ5, MYiQ).

c) Flow Cytometry. For fluorescence activated cell sorting (FACS) applications for whole cells, the Genomics Core houses a BD FACSAria which is a high-speed sorter with fixed-alignment cuvette flow cell and three air-cooled lasers (488nm, 633nm, and 407nm). This is operated by a dedicated specialist with significant experience in FACS (Holly Eckelhoefer) who also provides training and workshops.

II. Bioinformatics Core. UC Riverside's research compute infrastructure is provided by a high performance computing facility located in the Institute for Integrative Genome Biology (IIGB) and managed with the Genomics, Proteomics, and Microscopy & Imaging Cores. The research compute component is managed by Director of Bioinformatics Dr. Rakesh Kaundal and is located in the interdepartmental Genomics Building centrally located on the UC Riverside campus. The space dedicated to the computing facility consists of a server room and several offices for systems administrators, the director of the Facility and up to six programmers, students and postdocs. A multifunctional lecture hall, available in the same building, is frequently used for hands-on data analysis workshops with up to 75 participants. All compute systems of the facility, including a Linux cluster, web/database servers and data storage systems, are located in its state-of-the-art server room with 600 sqft of compute rack space. The server room has a raised floor cooling system with 2x 25 tons of redundant AC cooling and uninterrupted power is provided by a large 75KVA UPS and backup power generator. Its network is provided by a 10 Gbps connection to central campus and the

CalREN High Performance Research Network, which is part of the Internet2, along with a 4 Gbps connection to commodity internet. A secondary, co-located server room is utilized as part of the central campus data center to provide a geographically separated location for backups of production systems and data. Former Bioinformatics Core Director Thomas Girke was recently awarded NIH S10 and NSF MIR grants to expand and update the IIGB compute cluster ensuring its continued position in the future as the largest and most up-to-date biological computing center on the UCR campus.

a) Compute Hardware CPU and Memory Resources. The main compute cluster provides access to 512 CPU cores, 1.2TB of total RAM and the latest Infiniband interconnect. For memory intensive tasks, four dedicated high-memory nodes (additional 176 CPU cores total) are available with 64GB, 256GB and 2x 512GB of usable RAM. The latest release of the Debian Linux distribution is used as OS on all systems. The queuing system, scheduler and resource manager on the cluster are based on Maui and Torque. To manage the individual compute nodes, we make heavy use of Puppet for configuration management, Graphite for graphing trends in the operational environment, Jenkins for business process automation and Icinga for alerting and monitoring.

b) Central Data Storage System. The current data storage system is based on a storage area network (SAN) from LSI with a total of 164TB of raw disk storage space organized across two tiers, both being fully backed up to our secondary facility. Of the 164TB of raw disk space, 24TB is fast Tier 1 storage and 140TB is slower Tier 2 storage. Our Tier 1 storage consists of 300GB 15k RPM FC drives and our Tier 2 storage consists of 2TB and 3TB 7200 RPM SATA drives. Each Tier is served by a dedicated NFS server running OpenIndiana and making heavy use of ZFS. The data from all production systems are replicated via nightly differential backup protocols to a 196TB backup server system located in the geographically separated secondary server room. Two recent awards from both NIH and NSF are supporting the expansion of high performance computing and storage capacity.

c) Software Resources. The facility is strongly committed to maintaining a comprehensive software infrastructure for research applications in applied and basic biosciences. Currently, it maintains over 500 open source bioinformatics software packages for NGS analysis, comparative genomics, data mining, statistics, molecular modeling, cheminformatics, evolutionary biology, and all common programming environments. In areas dependent on commercial software applications, the facility provides access to a smaller number of commercial software tools. The primary focus on freely available and open source software offers access to the widest spectrum of software tools with the most advanced algorithms, and it maximizes freedom to operate in a highly diverse and multidisciplinary academic research environment.

III. Proteomics Core. The W. M. Keck Proteomics Core has a series of instruments for protein separation, sample preparation, and mass spectrometry analyses and houses a LTQ-Orbitrap Fusion, Waters Q-TOF nano-ESI MS/MS, an ABI Q-STAR XL oMALDI MS/MS, 2D-nanoUPLC, capillary LC equipment, protein digestion, 2-dimensional gel electrophoresis equipment, tissue homogenizer, and a GelFree 8000 protein separation system. The facility is managed by Academic Coordinator Dr. Sonqin Pan and has an upgraded Q-TOF Premier technologies including continuous expression scanning, ultra-performance liquid chromatography (UPLC) with non-splitting direct nano-flow, and more advanced data processing software, which will enable researchers to perform large-scale quantitative proteomics analyses with increased sensitivity and sequence coverage. With the support of an NIH S10 instrument grant (PI Yinsheng Wang), a highly-advanced Orbitrap Fusion MS system has been added to the facility to elevate its analytical capability and sensitivity to an unprecedented level. The Orbitrap Fusion features three fragmentation technologies including ETD, HCD, and CID. It is highly suitable for studies of protein post-translational modification (PTM) such as phosphorylation and glycosylation. Its ultra-high resolution also allows characterization of intact proteins through top-down approaches. Low abundance proteins from limited biological samples can be readily detected with its sub-femtomole level detection sensitivity. Combined with a two-

dimension nanoUPLC system, the Orbitrap Fusion can routinely perform MudPIT analysis for highly-complex biological samples both qualitatively and quantitatively.

IV. Microscopy/Imaging Core. The Microscopy Core provides a comprehensive suite of confocal microscopes and supporting peripheral equipment. Under the supervision of a full-time imaging expert and Academic Coordinator, Dr. David Carter, the core has high throughput imaging capability and a range of optical imaging applications for use mammalian, plant and other systems. The Microscopy Core has five confocal microscopes: A high resolution UV spectral system with 9 laser lines (Leica TCS SP2 UV); a high resolution visible system with five laser lines (Zeiss LSM-510) and built-in laser ablation capability (Photonic Microsystems MicroPoint); one fully automated white-light confocal imaging workstation (BD Pathway HT); and a Yokogawa spinning disc system with EMCCD camera (PhotonMax 512B) for low light imaging. The Yokogawa system has a Terabyte of storage memory and includes an Optical Insights DualView eyepiece for simultaneous imaging of CFP/YFP fluorescent proteins for FRET analysis. We also operate a Leica SP5 inverted confocal microscope which brings enhanced sensitivity, software and flexibility in imaging. Imaris Bitplane and Mercury Amira software are used for 3D image rendering and analysis. Other microscopy core equipment include a Hitachi TM-1000 tabletop SEM; a Molecular Devices Arcturus XT for laser capture; one ultramicrotome (Leica Ultracut T) with cryosectioning system for EM immunolocalization studies (Leica EM FCS); two fluorescence dissection stereo microscopes (Leica MZIII), one having attachments for fluorescence imaging at sub-cellular resolution (Fluoro Combi/SPOT RT210-3), and the other having a Spot Pursuit 4Mpixel camera for fluorescence and color imaging; a cryostat for rapid sectioning of frozen tissue (Microm HM 500 OMV); a microtome for sectioning wax or plastic embedded samples (Hacker 5030 Microtome); an oscillating tissue slicer (EMS5000); a particle gun for inserting DNA into living tissue (BioRad PDS1000/HE Hepta); an automated microscope with premium image analysis system (GE MCID Elite) and a fluorescence microplate reader (Wallac Victor 2).

V. Chemical screening facilities. IIGB and CEPCEB maintain diverse chemical libraries, robotics and cheminformatics that are operated across the Genomics, Microscopy and Bioinformatics Cores to support screening for bioactive molecules. Current collections consist of about 50,000 diverse compounds mostly from commercial sources that are maintained in dedicated freezers with autodial alarm systems. Collections are described at <http://chemminedb.ucr.edu/intro/data-sources/> and include commercial and some custom sets. Access is based on a modest rate per compound. For high throughput screens, we have two small BioTek Precision 2000 liquid-handling robots in laminar flow hoods for clean work, and a Beckman Coulter Biomek FX double bridge fluid handling robot with Cytomat hotel, for library management and distribution. The Biomek FX has eight independent Teflon tipped dispensers for cherry picking, a 96 tip head for bulk distribution, and two pin tool fixtures for dispensing sub-microliter volumes. This robot is maintained by the Microscopy Core and is available for use by students and other researchers free of charge. Additional equipment include a Percival Arabidopsis growth chamber and a Stanford Photonics Onyx luminescence imager with mouse handling capability (warm pad and anaesthetic manifold). Additional equipment in support of screening includes plate readers (Promega Glomax Multi+; Wallac Victor 2) which provide 96 and 384 well screening by luminescence, UV fluorescence or absorbance with heating/shaking if needed. For microscopy-based screening the Microscopy Core houses a fully automated white-light confocal imaging workstation (BD Pathway HT). This microscope is capable of automated imaging from 96 well or 384 well plates including autofocus and image capture. The Pathway is also capable of liquid handling and basic image analysis. The combined imaging resources permit phenotyping at the macro and micro levels. All equipment is available on a reasonable user fee basis, and training and consultation is provided free of charge. The Bioinformatics Core also developed a comprehensive screening database known as Chemmine (<http://chemminedb.ucr.edu/>) to maintain compound data

sources as well as links to Chemmine Tools. ChemMine Web Tools (<http://chemmine.ucr.edu/>) is an online service for analyzing and clustering small molecules by structural similarities, physicochemical properties or custom data types. Compounds structures can be imported into the workbench by simple copy/paste, local files or from a PubChem search. The site is free and open to all users with no login requirement.

VI. Metabolomics Core. In late 2016, a new Metabolomics Core will be developed as part of the IIGB infrastructure. This is a major addition that will enhance UCR's competitiveness in this new and expanding field of biology. This latest addition will be initiated under the direction of new incoming IIGB Director, Katie Dehesh.