Bioinformatics Core (http://facility.bioinformatics.ucr.edu/): UC Riverside’s research compute infrastructure is provided by a central Research Compute Facility located in the Institute for Integrative Genome Biology (IIGB), which houses three additional technology cores including Genomics, Microscopy and Proteomics. The bioinformatics facility provides access to high-performance computing resources, data analysis and programming expertise that serve scientists at UC Riverside to master the informatics needs of their research in a proficient and cost-effective manner. The facility also hosts a broad range of free workshops on data analysis that provides a strong training component to successful grants. The Research Compute component is located in a brand new interdepartmental Genomics Building centrally located on the UC Riverside campus. The space dedicated to the Research Compute 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 90 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 80 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. In 2014, the Bioinformatics Core was awarded a substantial NIH S10 grant 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.

**Compute Hardware**

(i) CPU and Memory Resources. The Research Compute Facility’s main compute cluster provides access to ~2400 CPU cores, 16TB of total RAM and the latest Infiniband interconnect. For memory intensive tasks, seven dedicated high-memory nodes (additional 256 CPU cores total) are available with 2x 128GB, 3x 512GB and 2x 1TB of usable RAM. For GPU-based computations, two nodes are available containing each two NVIDIA Tesla K80 GPUs with a total of 20,000 cuda cores. The latest release of the CentOS 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, Cacti for graphing trends in the operational environment, Jenkins for business process automation and Icinga for alerting and monitoring.

(ii) Central Data Storage System. The facility’s current data storage system is based on GPFS, a high-performance clustered file system deployed in distributed parallel modes with a total of ~1.3PB of storage capacity (175TB in home, 1.1PB in bigdata). An additional data storage system based on a storage area network (SAN) from LSI is also available 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 OmniOS and making heavy use of ZFS. There is an additional NAS-based 150TB of usable disk storage available for the users. **Back-up:** The data from all production systems are replicated via nightly differential backup protocols to the 780TB GPFS based backup servers, and a 254TB backup LSI/NAS-based server; both located in the geographically separated secondary server room. Two recent awards from both
NIH and NSF have supported the expansion of high-performance computing in terms of compute and storage capacity.

**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 depending on commercial software applications, the facility also 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.