

# INFORMATICS RESOURCE CENTER (IGS)

## CIBR: Center for Innovative Biomedical Resources

The Informatics Resource Center (IRC) under the direction of Anup Mahurkar provides genomics and bioinformatics services to the UMB campus. The IRC works under the overall guidance of Owen White, PhD, the Director of Bioinformatics for School of Medicine and the Associate Director of the Institute for Genome Sciences. IRC leadership also includes Michelle Gwinn Giglio, PhD, Associate Director for Analysis.

The IRC includes a staff of over 30 scientists, engineers, systems administrators, and analysts that work together to conduct research and development in bioinformatics and provide analysis services. The IRC staff is organized along scientific platforms and functional areas of expertise. Typically, biologists lead scientific platforms and coordinate the engineering and analysis activities needed for individual projects. The major scientific platforms supported by IRC include prokaryotic, eukaryotic, viral, and mammalian genomics, metagenomics, informatics, and systems biology.

### ANALYSIS SERVICES

The IRC has developed and maintains several analysis tools and pipelines that facilitate research at UMSOM. These include:

- Genome assembly and annotation. Pipelines for both prokaryotic and eukaryotic organisms are available. These include both reference-based and reference-independent protocols. Reference-based analysis relies on the transfer of information to the species under study from a closely related species. Reference-independent methods generate assemblies and annotation de novo. The annotated product in all cases includes predicted protein coding genes with functional annotations including protein names, gene symbols, EC numbers and Gene Ontology terms as well as prediction of non-coding RNAs.
- Transcriptome analysis with RNA-Seq. To identify gene and isoform expression patterns, the RNA-Seq alignment pipeline includes the use of the short read aligners BowTie/TopHat, BWA to align the RNA-Seq reads to reference genomes. The alignments are typically visualized using IGV or similar tools.
- Differential expression analysis. The IRC has pipelines to conduct gene and isoform level differential expression analysis using microarrays or RNA-Seq. These pipelines provide differential gene expression that is then used to identify differentially enriched pathways using DAVID, or Ingenuity Pathway Analysis.
- Genome variation analysis. Pipelines for single nucleotide polymorphism (SNP) and copy number variant (CNV) detection along with associated visualization tools.
- Metagenome/Metatranscriptome Analysis. Pipelines are available for analysis of microbiome community composition and functional potential from 16S amplicon and Whole Metagenome Shotgun sequence. RNA-Seq-based metatranscriptome analysis pipelines are available to determine differential expression and functional patterns of community members.
- Custom Programming and Analysis: In addition to using the standard pipelines, IRC staff can develop custom pipelines and analysis tools to meet the needs of individual projects. The IRC staff has expertise in web development, database development, and statistical programming.

### MISSION

To enable basic and translational research by leveraging expertise in genome-scale analysis tools and high-performance computing .

### CORE SERVICES

The IGS Informatics Resource Center engages in for-service (FFS) and collaborative research projects and proposals. For this purpose two cores were created: the Genome Informatics Core (GIC) and the High-Performance Computing Core (HPC). The following are some of the major services available through these IRC cores.

### GENOME ANALYSIS

The GIC has a talented group of researchers and staff with expertise in the following analyses:

- Genome Annotation
- Comparative Genome Analysis
- Metagenome Analysis
- Metatranscriptome Analysis
- Microarray Analysis
- Transcriptome Analysis
- Epigenome Analysis
- Variant Analysis
- Pathway Analysis

### SOFTWARE AND TOOL DEVELOPMENT

GIC software engineers are available to develop custom software solutions that include:

- Website Development
- Custom Programming/Scripting
- Research Data Capture Systems
- Database Design

### RESEARCH COMPUTING

The HPC computational infrastructure is available to researchers to conduct their own analysis. The following are some of the ways researchers can access this infrastructure.

- Cloud Computing
- Pre-packaged Genome Analysis Pipelines
- Command-line tools

### CONTACT

irc-info@som.umaryland.edu  
410-706-5682

# INFORMATICS RESOURCE CENTER (IGS)

## CIBR: Center for Innovative Biomedical Resources

### Visualization Tools

The IRC has developed a number of genome visualization and curation tools that are available to the research community. In addition we have also deployed third-party open source tools. Some of these include:

- SYBIL – A browser for comparative genomics results that provides views for ortholog groups, synteny gradients, genomic regions, and more.
- Circelator – A circular genome visualization tool providing compact figures showing diverse types of information that can be used to compare features of multiple genomes.
- Manatee – A genome annotation query and curation tool that allows one to browse annotations by gene location, function, and biological role. Annotations can be revised by users as well as downloaded in a variety of standard formats.
- Integrative Genome Browser (IGV) – A tool that provides simultaneous visualization of multiple types of genome-associated information including gene models, ortholog data, RNA-Seq alignments, and more.
- JBrowse – A Web-based genome visualization tool.

### Computational Infrastructure

Supporting the informatics at IGS/IRC is a state-of-the-art computational infrastructure that includes a computational grid, an internal 10-gigabit network, clustered database servers, and a hierarchical storage management system.

The grid is built around 5 high-performance high-memory multi-processor machines (512-1024 GB RAM, 4-8 CPU multi-core processors) for memory and compute intensive applications and over 90 high throughput computational nodes (48-256 GB RAM 2 CPU multi-core Intel processor machines) for running distributed applications.

To address the ever-expanding data sets generated by next generation genome sequencing technologies at a reasonable cost we have deployed a hierarchical storage infrastructure consisting of four tiers of random access storage and a fifth tier of serial access tape for archival and data backup. Total storage capacity is nearly 5 petabytes.

To enable genomics-based cancer research, the IRC, on behalf of the Center for Health-related Informatics and Bioimaging (CHIB) and the Greenebaum Cancer Center, has recently deployed a dedicated computational infrastructure with 500 TB of storage and 300 cores, and a set of standardized analysis tools and pipelines. The available pipelines include whole genome analysis, transcriptome analysis, and transcriptome SNP analysis. We have also built graphical interfaces to launch and monitor these pipelines and visualize the results. While this infrastructure is available for all researchers at the campus, cancer center members will get special discounted rate for accessing this infrastructure.

### Outreach and Educational Programs

A vital component of every bioinformatics project at IGS is training and outreach. IGS offers a comprehensive professional development program including multiple workshops each year that provide instruction on genomics, metagenomics, transcriptomics, and programming. More information can be found at <http://www.igs.umaryland.edu/education/workshops.php>.

### LEADERSHIP



Anup Mahurkar, MBA  
Director



Michelle Gwinn Giglio, PhD  
Associate Director for Analysis



Owen White, PhD  
Associate Director, IGS  
Director of Bioinformatics, School of Medicine  
Co-director CHIB

### WEBSITE

<http://www.igs.umaryland.edu/resources/irc/>

### LOCATION

Biopark II, Institute for Genome Sciences  
801 West Baltimore Street  
Baltimore, MD 21201