University *of* Virginia School *of* Medicine

Bioinformatics Core

Centralizing Bioinformatics: Applications, Opportunities, and Challenges with Large-Scale –*Omics* Data

UNIVERSITY of VIRGINIA SCHOOL of MEDICINE

bioinformatics core

bioinformatics.virginia.edu

About the Core

bioinformatics.virginia.edu/about

Mission

The new Bioinformatics Core at UVA serves as a centralized resource for providing expert and timely bioinformatics consulting and data analysis solutions.

The Bioinformatics Core's mission is to build and maintain an infrastructure for service and training that enables the application of strong bioinformatics analysis with a measurable impact on the ability of UVA investigators to both publish their work and obtain new funding.

Stephen D. Turner, Ph.D.
Bioinformatics Core Director
bioinformatics@virginia.edu

Goals

1) Service. The core's primary goal is to identify opportunities and implement solutions for managing, visualizing, analyzing, and interpreting genomic data, including studies of gene expression (RNA-seq and microarrays), pathway analysis, protein-DNA binding (e.g. ChIP-seq), DNA methylation, and DNA variation, using high-throughput platforms in both human and model organisms. We will establish best practices for high-throughput data analysis in a production environment, and develop custom solutions to assist research with data analysis and biological interpretation. See how we can support your research at bioinformatics.virginia.edu/services.

2) Training. It is important that a rapport is established between biomedical researchers and bioinformaticians with whom they collaborate. The Bioinformatics Core will provide outreach and education to the UVA community in the form of seminars, workshops, and training series. Learn more at bioinformatics.virginia.edu/training.

BioConnector

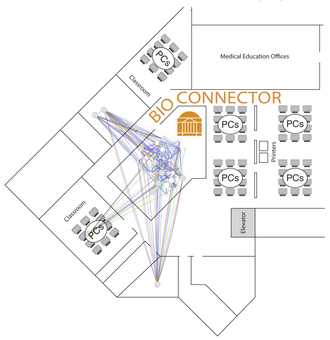
bioinformatics.virginia.edu/bioconnector

The Bioinformatics Core is partnering with the Health Sciences Library to bring together bioinformatic data analysis with information sciences to uniquely position the core/library as an institution-wide nucleus of scientific and biological research support. This partnership, dubbed "BioConnector" will facilitate research support, education and outreach, and collaboration discovery among UVA investigators.

Collaboration discovery: The core and library are outfitting a space with technology to facilitate active sharing of ideas. P.I.s will be encouraged to hold joint lab meetings and other collaborative discussions in this space.

Research support: This collaborative space will be next-door to the Bioinformatics Core director's office. Furthermore, this space will be outfitted with workstations having a research support focus.

Education and Outreach: The core periodically offers training sessions and workshops on various bioinformatics tools and techniques (e.g. Ensembl, UCSC genome browser, OncoMine, RNA-seq).



BioConnector Highlights:
- Collaborative meeting space with multiple display capability.
- Wall-to-wall floor-to-ceiling dry-erase painted walls.
- Next-door to the core director's office and other core staff.

Gene Expression

Microarray

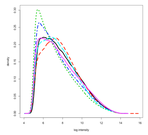
bioinformatics.virginia.edu/microarray

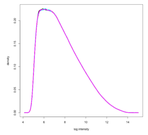
Gene expression is a key determinant of cellular phenotypes. Microarrays have been the workhorse for gene expression studies for over a decade because of their ability to probe the expression of many thousands of transcripts simultaneously. While RNA-seq has many advantages over hybridization-based microarrays, RNA-seq is a young technology undergoing rapid changes in library preparation biochemistry, sequencing platforms, computational pipelines, analysis methods, and statistical treatment. **Services include**:

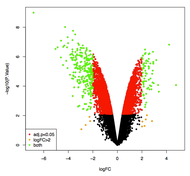
- Accession and analysis of **publicly available data** (e.g. GEO, ArrayExpress).
- Formatting and uploading data to **GEO**.
- **Preprocessing**: background subtraction, summarization, and quantile normalization.
- **Quality assessment**: Visualization of signal intensity distributions of each array using boxplots and density plots. MA plots to visualize signal intensity over average intensity.
- **Principal components analysis** and **clustering** to visualize the overall data (dis)similarity between arrays.
- **Analysis**: Estimation of **fold changes** and standard errors using a linear model with empirical Bayes moderated standard errors.
- Lists of top **differentially expressed genes**, fold changes, statistical significance, **multiple testing correction**.
- Results **visualization**: **Heatmaps, dendrograms**, and **Volcano plots** to visualize statistical significance by fold change.
- **Pathway analysis** – See "Context" section.

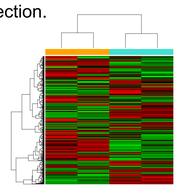
ARRAYEXPRESS

GEO Gene Expression Omnibus







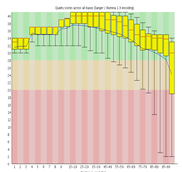


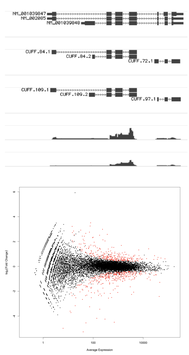
RNA-Seq

bioinformatics.virginia.edu/rna-seq

While microarrays have been the predominant technology for measuring gene expression for years, all hybridization-based technologies are subject to biases and limitations, such as reliance on known gene models and cross-hybridization to probes with similar sequences. Microarrays offer a limited ability to fully catalog and quantify the diverse spectrum of RNA molecules expressed from cellular genomes, and are of little use for detecting previously unknown transcripts, novel exons, or exons with novel splice junctions. RNA-seq does not have these limitations, and furthermore provides a digital readout of gene expression, enabling a higher sensitivity and an essentially unlimited dynamic range. **Services include**:

- Pre- and post-alignment **QC** (per-base quality/GC-content, overrepresented sequences, library complexity, duplicate removal, insert size distribution, coverage estimation).
- **Alignment** to a reference genome (human, mouse, rat, cow, dog, chicken, *Drosophila*, *Arabidopsis*, *C. elegans*, yeast).
- Transcript **assembly**
- **Differential expression testing** (gene level, isoform level, primary transcript, and coding sequence).
- **Differential processing** (differential splicing, transcription start site usage, differential coding output).
- **Visualization** (IGV, Galaxy Track Browser, cummeRbund).





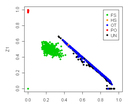
DNA Variation

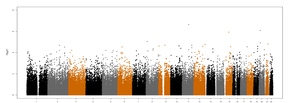
bioinformatics.virginia.edu/dna-variation

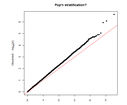
The core currently offers consulting, analysis, management, and interpretation of candidate gene studies, genome-wide high-density DNA **genotyping** data (**GWAS**), and **next-generation sequencing studies**. Services include

- Study design & **power calculations** for SNP genotype-phenotype association studies
- Genotype-phenotype **association studies**: Analysis, interpretation, visualization
- Genotyping: PCA & **Imputation** to a reference population (e.g. HapMap, 1000 Genomes)
- **NGS**: **alignment** to a reference genome, calibration of quality scores, **variant calling**.
- **NGS**: variant **annotation**, SNP **effect prediction**
- Manuscript & Grant preparation
- Accession of publicly available data (**dbGaP**)

dbGaP GENOTYPES and PHENOTYPES





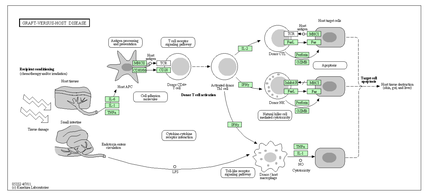


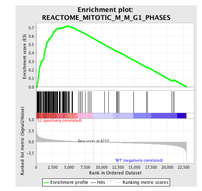
Biological Context

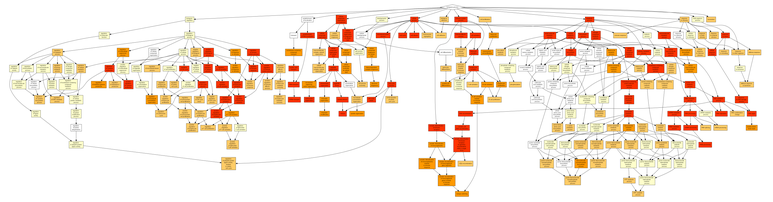
bioinformatics.virginia.edu/context

Many "standard" bioinformatics analyses result in lists of variants, transcripts, or genes and some statistic - e.g., a gene name, fold change, and multiple-testing-corrected P-value for a gene expression study. It is helpful to interpret gene expression data by framing these gene lists in a functional, biological context. This is the basis for "pathway analysis" or functional annotation. The core can assist with these and others:

- **Gene Ontology** overrepresentation analysis (hypergeometric tests)
- **Gene Set Enrichment Analysis** (GSEA, Subramanian et al. PNAS 2005)
- **Ingenuity Pathway Analysis**
- Others (**OncoMine**, **KEGG**, transcription factor networks, etc.)







Other Services

The core is also preparing pipelines for other analyses, including:

- ChIP-seq (alignment, peak calling)
- Small RNA analysis
- DNA methylation

If you would like to become an early adopter and receive a reduced rate for services we are still developing, please contact Stephen Turner at bioinformatics@virginia.edu.

Visit bioinformatics.virginia.edu for more information about any of our services. We can also work with you to develop custom bioinformatics solutions to problems not listed here or online.