

## ABSTRACT

The core facility provides bioinformatics expertise for analyses involving **data-intensive experiments** (e.g., next generation sequencing) as well as **traditional bioinformatics analyses**, such as sequence alignments and phylogenetic analyses.

Our support ranges from **standard workflows to customized, in-depth analyses**. We provide **advice and consultation** starting from the initial planning of the experiment to data lifecycle management.

The ABC can also be employed for the development of **websites and customized applications** for managing and visualizing data as well as for the establishment and implementation of **software analysis pipelines**.

In addition, we offer regular seminars, educational workshops and hands-on training.

## OVERVIEW

Our goal is to assist researchers to **address biologically relevant questions** using statistically sound and fully **reproducible computational analyses**.

### MISSION

The core facility's team of bioinformatics scientists enables individual research labs, core facilities, and translational medicine programs at WCM to take advantage of advances in high-throughput experimental data acquisition by employing cutting-edge data analysis tools and methods in areas of biomedical research, both fundamental and translational, ranging from cancer biology to neuroscience, and from diabetes research to drug screening.

### HISTORY

The Applied Bioinformatics Core (ABC) was established in 2013 to address the myriad challenges in storing, analyzing and interpreting large-scale datasets that are the result of continuing advances in high-throughput experimental data acquisition in the biomedical sciences.

### LOCATION

The ABC office is located on the 5<sup>th</sup> floor of the LC building at 1300 York Avenue on the Weill Cornell Medicine campus in New York City.

### OPEN TO ALL

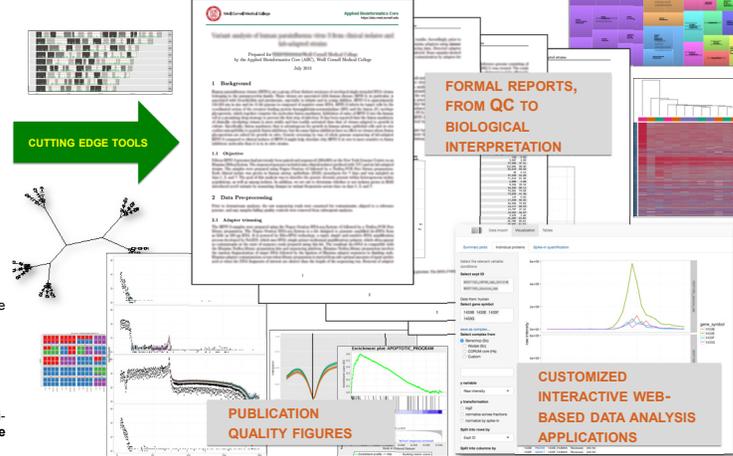
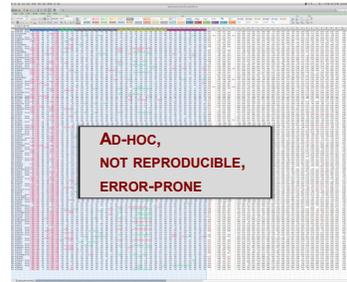
The resources and services of the core facility are open to all investigators at Weill Cornell Medicine, Cornell University and Cornell-affiliated institutions. The facility also provides services to external investigators.

## RESOURCES

Through the Scientific Computing Unit, the ABC has access to High Performance Computing infrastructure with peta-scale data storage and backup and parallel computing (1000s of CPU cores; GPUs).



## WHEN EXCEL IS NOT ENOUGH



The ABC offers support for a wide range of data-intensive and large-scale data analysis projects such as:

- **transcriptomics** (microarrays and RNA-seq)
- **epigenomics** (ChIP-seq, eRRBS, WGBS)
- **whole genome DNA sequencing**
- **genotyping / variant calling**
- **sequence alignment and phylogeny**
- development of **customized software**, including applications with **graphical user interfaces for interactive data exploration** by the researcher

## CONSULTING

### GRANT SUPPORT

The ABC provides **letters of support** for grant applications

- Access to well established infrastructure and dedicated bioinformatics staff are important in addressing reviewers' concerns.
- Addresses the NIH's emerging requirements for the use of rigorous and reproducible data analysis methods.

### SCOPED PROJECTS

The ABC works closely with its clients to understand the **focus of the research**, to determine the scope of the project and how to leverage current best practices to answer their scientific or practical questions. The ABC also advises on the **experimental design** that **best balances cost and statistical power**.

- The ABC prepares a **detailed proposal** that outlines the planned work and elucidates the analyses to be performed. The proposal contains a **fixed cost** for the deliverables.

- At the end of each project, the ABC delivers a **comprehensive report** that contains all analysis details, results, and a discussion section that puts the results into the context of the specific research project.

- In addition, the ABC provides all result files of interest, including **publication-ready images** as well as **all software tools and scripts** that were used to generate the results.

### AUTOMATED PIPELINES

The ABC **hardens software** for computational labs that aim to **streamline their analysis and ensure reproducible results**. Pipelines can be integrated into existing workflows and alternative analyses with optional parameters can be subsequently ordered.

### AGILE PROJECTS

ABC staff can be engaged on an **hourly basis for short-term or open-ended projects** where the preparation of a formal scope may not be practical or desirable.

## EXAMPLE PROJECTS

- RNA virus genome assembly and variant analysis
- Assessment of transcriptome and methylome differences in different cancer cell lines
- Identification and annotation of genomic variants in different parasite mutants
- Investigating the effects of DNA methyltransferase inhibitors on the transcriptome and methylome of different cancer cell types
- Identification of cancer-induced transcriptome changes in blood cell samples
- Identification of gene candidates for enzymes involved in glycosylation pathways using phylogenetic profiling
- R shiny app for capturing, organizing, normalizing and visualization of high-throughput bacterial survival assays
- R shiny app for visualization of proteomics data
- Website for the Core for Smooth Muscle Analysis (<https://csma.med.cornell.edu>)
- Pipelines for RNA-seq and DNA-seq alignment
- Pipeline for the identification of genomic variants in tumor patient samples

## RECENT PUBLICATIONS

Ramirez F, Ryan DP, Grünig B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T.  
*deepTools2: a next generation web server for deep-sequencing data analysis.*  
Nucleic Acids Res. 2016 Apr 13. doi: 10.1093/nar/gkw257

Palermo LM, Uppel M, Skrabanek L, Zumbo P, Gerner S, Toussaint NC, Rima BK, Huey D, Niewiesk S, Porotto M, Moscona A.  
*Features of Circulating Parainfluenza Virus Required for Growth in Human Airway.*  
MBio. 2016 Mar 15;7(2). doi: 10.1128/mBio.00235-16.

Webster AF, Zumbo P, Fostel J, Gandara J, Hester SD, Rocio L, Williams A, Wood CE, Yaik CL, Mason CE.  
*Mining the Archives: A Cross-Platform Analysis of Gene Expression Profiles in Archived Formalin-Fixed Paraffin-Embedded Tissues.*  
Toxicol Sci. 2015 Dec;148(2):460-72. doi: 10.1093/toxsci/kfv195.

Li S, Labaj PP, Zumbo P, Sykaeck P, Shi W, Shi L, Phan J, Wu PY, Wang M, Wang C, Thierry-Mieg D, Thierry-Mieg J, Kneil DP, Mason CE.  
*Detecting and correcting systematic variation in large-scale RNA sequencing data.*  
Nat Biotechnol. 2014 Sep;32(9):888-95. doi: 10.1038/nbt.3000.

## TRAINING

### ONE-ON-ONE AND GROUP TRAINING

We offer individualized training for software analysis tools and pipelines. We also give dedicated workshops to small groups.

### HANDS-ON WORKSHOPS

The ABC hosts a series of workshops that promote the effective use of bioinformatics and computational methods in scientific research. Topics include:

- Introduction to Unix
- Introduction to R
- RNA-seq analysis

See <http://www.triio.org/courses/> for more information

### BIOINFORMATICS WALK-IN CLINICS

Researchers can get immediate help with any kind of bioinformatics-related question

- Thursdays, 12-1:30 pm, LC504 at 1300 York Avenue

### DATA ANALYSIS & BIOINFORMATICS USER GROUP (D:BUG)

The ABC hosts an informal platform for fostering lore exchange between WCM personnel that code on a regular basis

- select Thursdays, LC504 at 1300 York Avenue

See <https://github.com/abcbug/dbug>

## CONTACT

### Applied Bioinformatics Core

Luca Skrabanek, Ph.D. | Paul Zumbo | Friederike Dündar, Ph.D.  
ls2017@med.cornell.edu | paz2005@med.cornell.edu | fd2007@med.cornell.edu

<https://abc.med.cornell.edu>