

Welcome to BIOS 567

Mikhail Dozmorov
Fall 2017

Course logistics

Instructor: Mikhail G. Dozmorov
mikhail.dozmorov@vcuhealth.org, Ph.D.

Schedule: Monday, Wednesday 9:00am – 10:20am

Classroom: One Capital Square (OCS) 5009

Office hours: Mon, Wed 10:30am – 12:00pm, OCS 730

Course documents: <https://blackboard.vcu.edu>,
<https://mdozmorov.github.io/bios567.2017/>

Course prerequisites

Prerequisites: BIOS 524 Biostatistical Computing, BIOS 553 Linear Regression, and BIOS 554 ANOVA

Required text: Sorin Draghici “Statistics and Data Analysis for Microarrays Using R and Bioconductor”, 2nd Ed., Chapman & Hall/CRC Press, 2012. ISBN-978-1-4398-0975-4.

- Supplemental reading and course materials will be provided in-class

Software: The R programming environment

3/13

Computational Genomics

Genomics/Bioinformatics (general)

- Generation, organization, analysis and interpretation of biological data (initially genomic data)

4/13

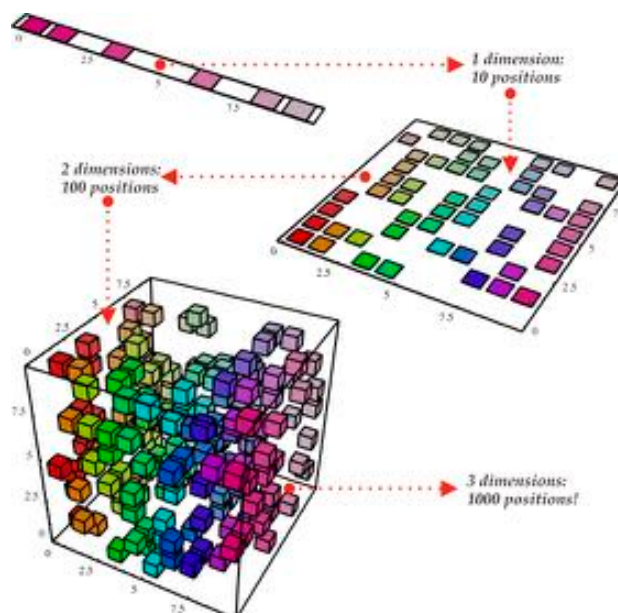
Biostatistics

Biostatistics

- Class of statistical methods for dealing with large biological data sets
- Goal: statistically identify significant changes in biological processes to answer relevant biological questions.
- High-throughput studies; get data matrix; mine the matrix for information

5/13

Meet the curse of dimensionality



6/13

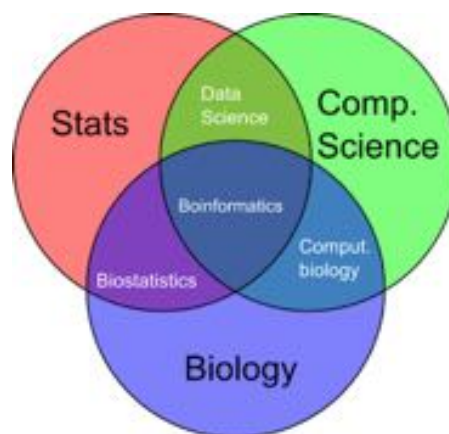
Distributions to know

- Normal (Gaussian) distribution
- t distribution
- Chi-square distribution
- F distribution
- Binomial distribution
- Poisson distribution
- Gamma distribution
- Negative binomial distribution
- Beta distribution
- Beta Binomial distribution
- Multinomial distribution

1 parameter; 2 parameters; 3 parameters

7/13

Bioinformatics



Data Scientist = statistician + programmer + coach + storyteller + artist

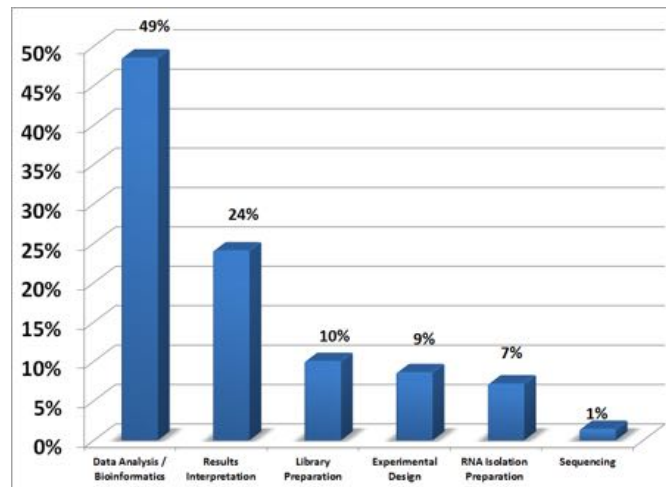
- Shlomo Aragon

<https://genomejigsaw.wordpress.com/2015/09/27/faq>

8/13

Data analysis

Which step of an RNA-Seq project do you find most intimidating?



<http://www.rna-seqblog.com/rna-seq-blog-poll-results-25/>

9/13

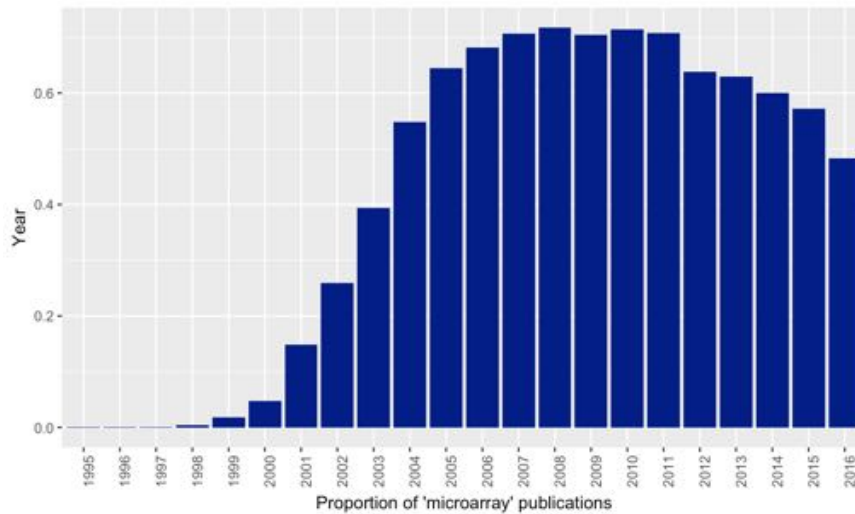
Still microarray?

- Microarray is still widely used because of lower costs, easier experimental procedure and more established analysis methods.
- Still around 50% of all publications annually.
- Similar problems are presented in newer technologies such as RNA-seq, and similar statistical techniques can be borrowed.

10/13

Microarrays

Counts of papers per year in PubMed having the word microarray



11/13

Microarrays today

Researchers find that modern microarrays can outperform RNA-Seq in terms of reproducibility and cost

Posted by: RNA-Seq Blog | In Information, Publications | June 13, 2017 | 2,660 Views

RNA sequencing (RNA-seq) and microarrays are two transcriptomics techniques aimed at the quantification of transcribed genes and their isoforms. Here researchers from the [Luxembourg Institute of Health](#) compare the latest Affymetrix HTA 2.0 microarray with Illumina 2000 RNA-seq for the analysis of patient samples – normal lung epithelium tissue and squamous cell carcinoma lung tumours. Protein coding mRNAs and long non-coding RNAs (lncRNAs) were included in the study.

<http://www.rna-seqblog.com/researchers-find-that-modern-microarrays-can-outperform-rna-seq-in-terms-of-reproducibility-and-cost/>

12/13

Expectations

- Attend every class
- Read every assigned reading
- Complete all homework in time
- Be active in class. Ask questions
- Work hard - it will pay off