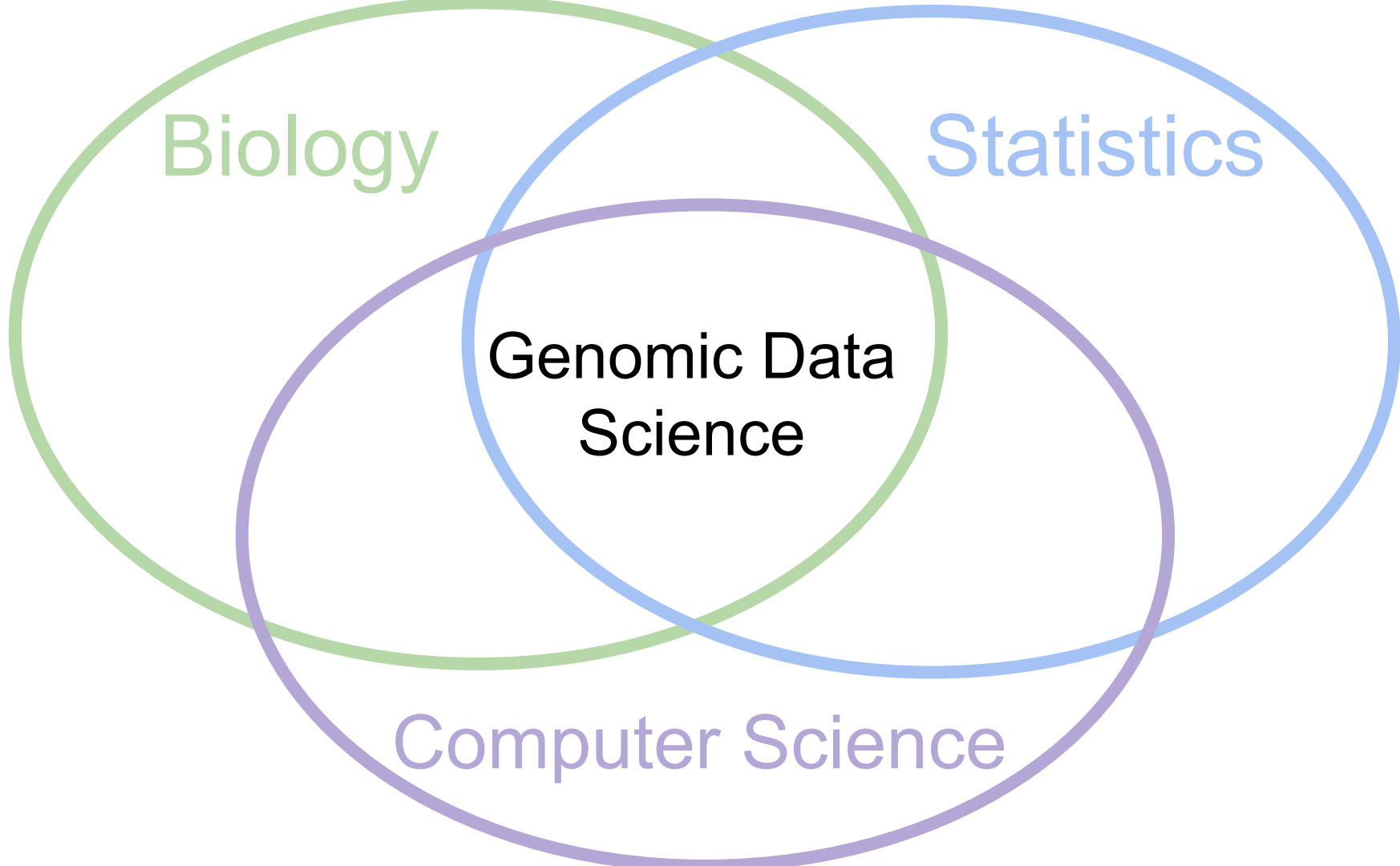


Statistics for Genomic Data Science Wrap Up

Jeff Leek

@jtleek

www.jtleek.com



Biology

Statistics

Genomic Data
Science

Computer Science

Abstract

Formula display: **MathJax** 

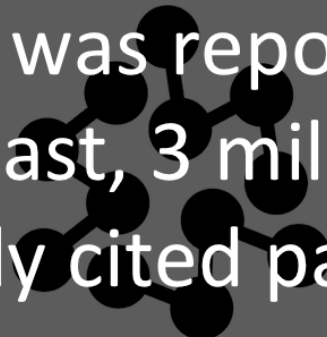
Background

Many groups, including our own, have proposed the use of DNA methylation profiles as biomarkers for various disease states. While much research has been done identifying DNA methylation signatures in cancer vs. normal etc., we still lack sufficient knowledge of the role that differential methylation plays during normal cellular differentiation and tissue specification. We also need thorough, genome level studies to determine the meaning of methylation of individual CpG dinucleotides in terms of gene expression.

Results

In this study, we have used (insert statistical method here) to compile unique DNA methylation signatures from normal human heart, lung, and kidney using the Illumina Infinium 27 K methylation arrays and compared those to gene expression by RNA sequencing. We have identified unique signatures of global DNA methylation for human heart, kidney and liver, and showed that DNA methylation data can be used to correctly classify various tissues. It indicates that DNA methylation reflects tissue specificity and may play an important role in tissue differentiation. The integrative analysis of methylation and RNA-Seq data showed that gene methylation and its transcriptional levels were comprehensively correlated. The location of methylation markers in terms of distance to transcription start site and CpG island showed no effects on the regulation of gene expression by DNA methylation in normal tissues.

If he was cited every time a p-value was reported
get a bad rap - but we
sometimes find them useful.
his paper would have, at the very least, 3 million
citations* - making it the most highly cited paper
of all time.



The **p-value** is the most widely-known statistic. P-values are reported in a large majority of scientific publications that measure and report data. **R.A. Fisher** is widely credited with inventing the p-value. If he was cited every time a p-value was reported his paper would have, at the very least, 3 **million** citations* - making it the **most highly cited paper** of all time.

However, the p-value has a large number of very vocal critics. The criticisms of p-values, and hypothesis testing more generally, range from philosophical to practical. There are even **entire websites** dedicated to “debunking” p-values! One



RECENT POSTS

[The statistics department Moneyball opportunity](#)

[The Mozilla Fellowship for Science](#)

There is more to learn!



Advanced Statistics for the Life Sciences

We dive into the best methods for analyzing complex data in the life sciences with a focus on genomics.



Starts Self-Paced

[Enroll Now](#)

- I would like to receive email from Harvard University and learn about its other programs.

About this course

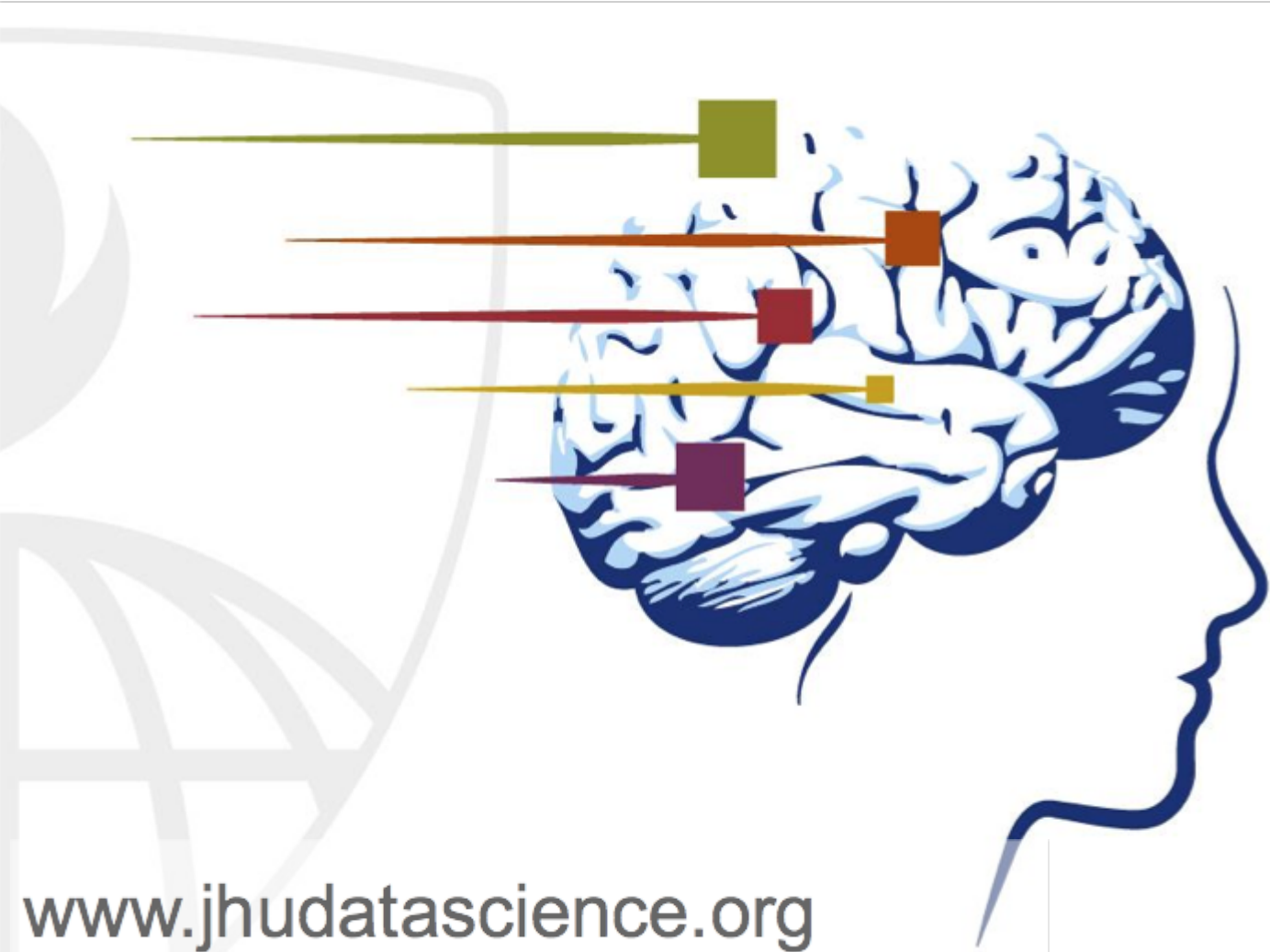
2 Reviews 4/5 ★★★★★

Here we will cover the advanced techniques being used by data analysis experts in the life sciences. These methods are required to analyze some of the more complex datasets, such as those found in genomics. We will cover several topics including statistical modeling, multiple test correction, clustering, prediction methods, factor analysis and empirical Bayes methods. We will also elaborate on the use of R markdown to

[See more](#)

What you'll learn

Level:	Intermediate
Length:	4 weeks
Effort:	6 hours / week
Subject:	Biology & Life Sciences
Institution:	HarvardX
Languages:	English



www.jhudatascience.org

Thank you!