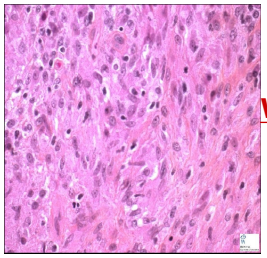
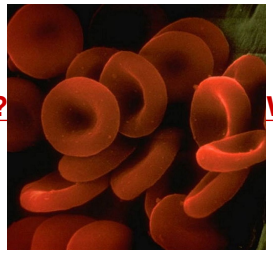
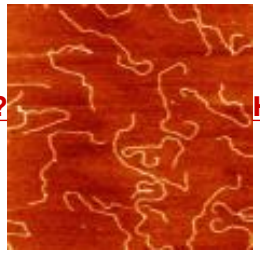
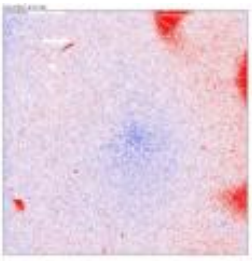
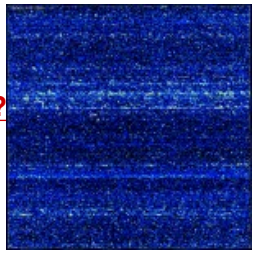

Who?

What?

How?

Why?


STAT 466* Statistics for Genomics
For grad students offered under
STAT 857* Statistics for Life Sciences
Room change: now in 101 - Jeffery Hall
Instructor: Julia Brettschneider, PhD

Tutorials with Elsa Hansen:
Monday 4:30, Tuesday 9:30
(choose one)

Instructor's office hours:
Wednesdays and Fridays
after class in 101 Jeffery

Background material (collection of links): gene detective's blog
Course material (in reverse chronological order):

- **Clustering talk from Bioconductor workshop**
- Peng Wu (grad student): Normalization in a cDNA study on prostate cancer
- Lectures Nov 28: Multiple testing (false discovery rate)
- Lectures Nov 24: Multiple testing (family wise error rate)
- Homework 5
- Jie Zheng and Shiqin Helen Guan (grad students): QA in a short oligo fly embryo time course experiment
- Lectures Nov 21: Finding differentially expressed genes
- Lectures Nov 14 and 17: Overview of cDNA microarray normalization methods
- Jing Jin (grad student), Nov 17: Connectivity Map
- Mark Rogers (grad student), Nov 15: Genomic-scale population divergence in flies & cDNA microarray image analysis
- Computer lab 2: cDNA microarray data analysis
- File with functions for cDNA microarray data analysis "smaFunctions.txt"
- Guidance for computer lab 2 "sampleAnalysis.txt"
- ApoAIdata_R.txt ApoAIdata_G.txt ApoAIdata_Rb.txt ApoAIdata_Gb.txt
- Introduction to R for biologists (by Natalie Roberts, WEHI, Melbourne)
- R manuals under link "Manuals" (left column)
- Lecture November 8: basics about cDNA microarrays 2
- Chi square example (as .doc file)
- Chi square example (as .rtf file)
- Homework 4
- Lecture November 3: basics about cDNA microarrays I
- Guest lecture October 31: basics about microarray experiments
- Copy of my email with instructions regarding homework 3 (due Nov.3)
- Undergrad projects - additional info
- Lecture October 13: R help for Computer lab 1
- Lecture October 13 supplement: R interactive session
- Computer lab 1 with E.coli data:
    - 123nt
    - cacttt_gap
    - gcatgc_gap
    - gtattg_20kb_count
    - gtattg_5kb_count
    - purine100

- purine1000
- purine10000
- tataat_20kb_count
- tataat_5kb_count
- tataat_gap

Projects for undergrads
Homework 2
Course info part II
Homework 1
Example 2 (micorarrays) from 19.9.06 lecture (about a minute to download this pdf)
Course Info part I (includes info on textbook and other literature)

Who is this class being offered for? And who offers it? There are several kinds of students that may take this class for a number of different reasons. For example: A **math/stats student** who is interested in genomics or who is generally curious to see how statistics and probability theory are used in scientific research and about the role of statisticians in collaborative work. A **biology or biochemistry student** who would like to get ready for the statistical challenges that inevitably come with evaluating genomic experiments. An **epidemiology student** who wants to explore new biochemical assay technologies and their quantification and meaning in the context of study design. A **health sciences student** interested in the opportunities and limitations of new genomic technologies in biomedical research and screening/diagnostics.

The prerequisites ("some exposure to statistics or probability") are few. We expect students to do guided reading and be curious to learn about a field that is new to them and to communicate with fellow students from other disciplines. The instructor Dr. Julia Brettschneider is an **applied statistician and assistant professor in Mathematics/Statistics and Community Health/Epidemiology.** She has been collaborating with molecular biologists and biomedical researchers on projects involving innovative high-throughput genomic technologies for the past 5 years.

What is this about? In highschool you learned that genes are the blueprint of organisms. All the cells in your body have essentially the same set of genes. How is it that your liver and your brain are different*?

The biochemical activity of a gene, also called gene expression, depends on the cell type and the cell's condition, and it varies substantially from gene to gene. In turn, measuring how a gene's expression changes in response to altered circumstances can shed light on its functions. For example, a gene that turns out to be differentially expressed in tumour versus control tissue may be directly or indirectly involved in the genesis or the consequences of the cancer. New biochemical assay technologies such as **microarrays**, enable scientists to measure the expression of thousands of genes simultaneously.

The major challenge in analysing such data is to correctly distinguish biological variation in the data from technological noise and bias in tens of thousands of simultaneous measurements, typically with a very small number of replicates. More specifically, this leads to statistical issues such as non-linear normalization, reliability and accuracy of simultaneous measurements, and massive multiple testing.

For more information see Genome Canada's genomics tutorial and NCBI's introductions to microarrays. Terry Speed's Micorarray homepage portraits the statistical challenges related to microarray data. The Nature article describes the perspective of biologists facing heaps of noisy genomic data including their urgent need for better methods and computationally and statistically skilled support. The richest source of freely available packages for genomic data analysis is Bioconductor.

How is this class going to be taught? The focus of this course will be on building a bridge between science and statistics, and on how scientific context affects choice and application of statistical techniques. Background in genomics, statistics, and use of basic computational packages, will be developed as needed. As to statistical topics, we will cover descriptive statistics, regression, non-linear regression, basic probability concepts, basic sampling, hypothesis testing, multiple hypothesis testing, basic experimental design, data quality assessment and control, clustering and classification. See Syllabus for more details. The course will be similar to my former course Statistics for Bioinformatics at UC Berkeley. There will be some amount of traditional lecture style, and there will be the usual program of homework assignments, quizzes, and a final exam (but no midterm). Additionally, we will:
- solve problems in class by discussions in groups
- work in computer labs
- have short oral presentations by students
- organize a conference style poster session
- mix students in interdisciplinary groups so that they can support each other and can learn how to commicate across disciplines

Tentative plan for determining the grades: final exam (42%), poster (20%), oral presentation (10%), two quizzes (5% each), 3 computer labs and 5 homework assignments (count 6 best ones with 3% each).

Why are we offering this course? And why in this form? Our first objective is to **teach statistical methodology and computational tools needed for scientific research based on innovative genomic technologies.** Our second objective is to teach **how science and statistics go hand in hand, and stimulate each other.** This includes probabilistic modeling and statistical data analysis for answering scientific questions. It also includes formulating scientific questions in a way that makes them answerable by experiments. For example, an interest expressed as "What are genes doing in cancer cells?" could be phrased more usefully in a question like "Which genes are differentially expressed in kidney tumour samples compared to healthy tissue samples?". Computer labs are part of the class, because we believe that the hard work of computational implementation is most rewarding in the context of a question that needs to be answered.

We offer this class for an **interdisciplinary audience** because we believe that statisticians and scientists need to get comfortable working together as early as possible in their careers - an important skill in both the academic and the commercial sector. This includes learning each other's language. It means explaining one's own field to an outsider in a sufficiently accurate but still simple enough way. It also means to get used to such explanations from the other field. It further includes getting an understanding of the other side's questions and priorities. Sir R. A. Fisher's famous quote describes what happens if the two sides don't communicate well: "To call in a statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of".

Questions? For further questions email instructor ac.usneeuq.tsam@ailuj (read backwards). You can also email to get on a list for updates regarding this course.

*If that doesn't apply you don't really need this class...

- A_01.cel
- A_02.cel
- A_03.cel
- A_04.cel
- A_05.cel
- A_06.cel
- A_07.cel
- A_08.cel
- A_09.cel
- A_10.cel
- A_11.cel

A_12.cel
B_01.cel
B_02.cel
B_03.cel
B_04.cel
B_05.cel
B_06.cel
B_07.cel
B_08.cel
B_09.cel
B_10.cel
B_11.cel
B_12.cel
C_01.cel
C_02.cel
C_03.cel
C_04.cel
C_05.cel
C_06.cel
C_07.cel
C_08.cel
C_09.cel
C_10.cel
C_11.cel
C_12.cel