

# Bioinformatics - From Sequence Expression and Structure

**Collection Editor:**

Krzysztof Cyran



# Bioinformatics - From Sequence Expression and Structure

**Collection Editor:**

Krzysztof Cyran

**Authors:**

Krzysztof Cyran

Rudy Guerra

Marek Kimmel

**Online:**

< <http://cnx.org/content/col10202/1.24/> >

**C O N N E X I O N S**

**Rice University, Houston, Texas**

This selection and arrangement of content as a collection is copyrighted by Krzysztof Cyran. It is licensed under the Creative Commons Attribution 1.0 license (<http://creativecommons.org/licenses/by/1.0>).

Collection structure revised: May 11, 2004

PDF generated: February 4, 2011

For copyright and attribution information for the modules contained in this collection, see p. 19.

## Table of Contents

<b>Introduction</b>	1
<b>1 Statistical methods for gene expression</b>	
1.1 Gene expression - blotting	4
1.2 DNA Microsatellites Evolution	5
Solutions	7
<b>2 Statistical machine learning</b>	
<b>3 Computational geometry and robotics</b>	
<b>4 Template</b>	13
<b>Glossary</b>	17
<b>Index</b>	18
<b>Attributions</b>	19



# Introduction<sup>1</sup>

## Course roadmap

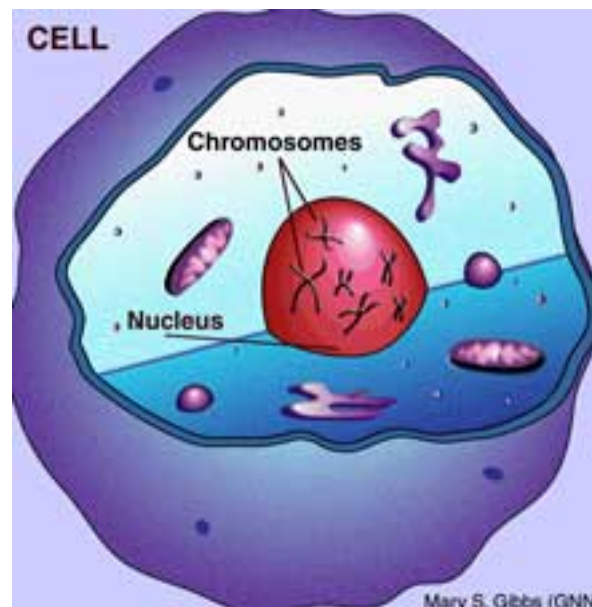
The novel research-oriented course of bioinformatics called *From Sequence to Expression and Structure* is organized into three main parts listed below.

### Course structure

1. Statistical methods for gene expression and gene interaction using DNA-array data as well as other genomic data
2. Statistical machine learning techniques for bioinformatics
3. Application of computational geometry and robotics techniques to the study of biomolecules and receptor-ligand interactions

---

### Biological cell



**Figure 1:** This is a 3 dimensional diagram of the living organism cell

---

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m11106/2.17/>>.

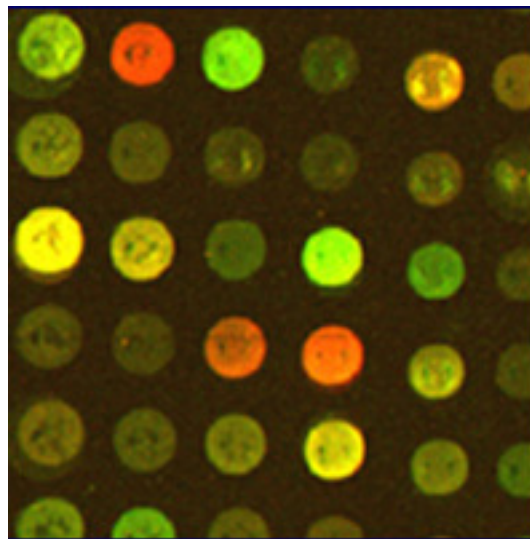
## Statistical methods for gene expression

Multiple gene expression techniques allow simultaneous measurement of expression levels of up to 50,000 genes. These novel methods already are being used to classify human cancers as well as to measure expression changes in experimental conditions. This part of the course introduces the student to the techniques and experiments for obtaining gene expression data, as well as the probabilistic and statistical methods for analysis of such data.

The data generated by large scale parallel hybridization techniques, such as DNA microarrays, constitute a new generation of data requiring novel methods of statistical analysis.

---

**DNA array**



**Figure 2:** This is an enlarged view on a fragment of cDNA microarray. By calculation of the RED/GREEN ratio for each spot (corresponding to one gene) it is possible to determine the concentration of mRNA (gene expression measure) from test and the reference conditions.

---

You can also refer to the comprehensive presentation of the DNA-Microarray technology<sup>2</sup>

---

<sup>2</sup><http://www.bio.davidson.edu/courses/genomics/chip/chip.html>



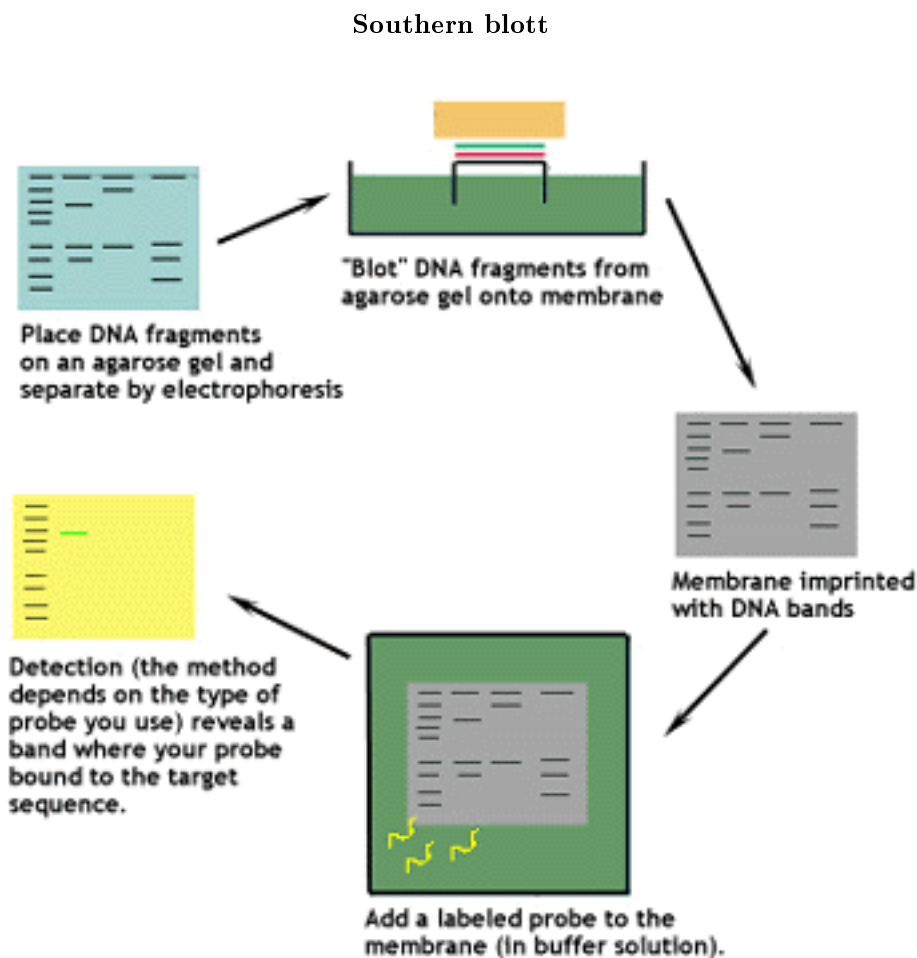


## Chapter 1

# Statistical methods for gene expression

### 1.1 Gene expression - blotting<sup>1</sup>

---



**Figure 1.1:** Basic steps included in a Southern Blot. Courtesy of Mama Ji Molecular Kitchen web page

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m11107/2.6/>>.

Below there is an example of the table in CNXML

head 1	head 2
value1 1	value1 2
value21 1	value2 2

Table 1.1

### Exercise 1.1

This is a question

(Solution on p. 7.)

## 1.2 DNA Microsatellites Evolution<sup>2</sup>

Simulated Var vs Variance

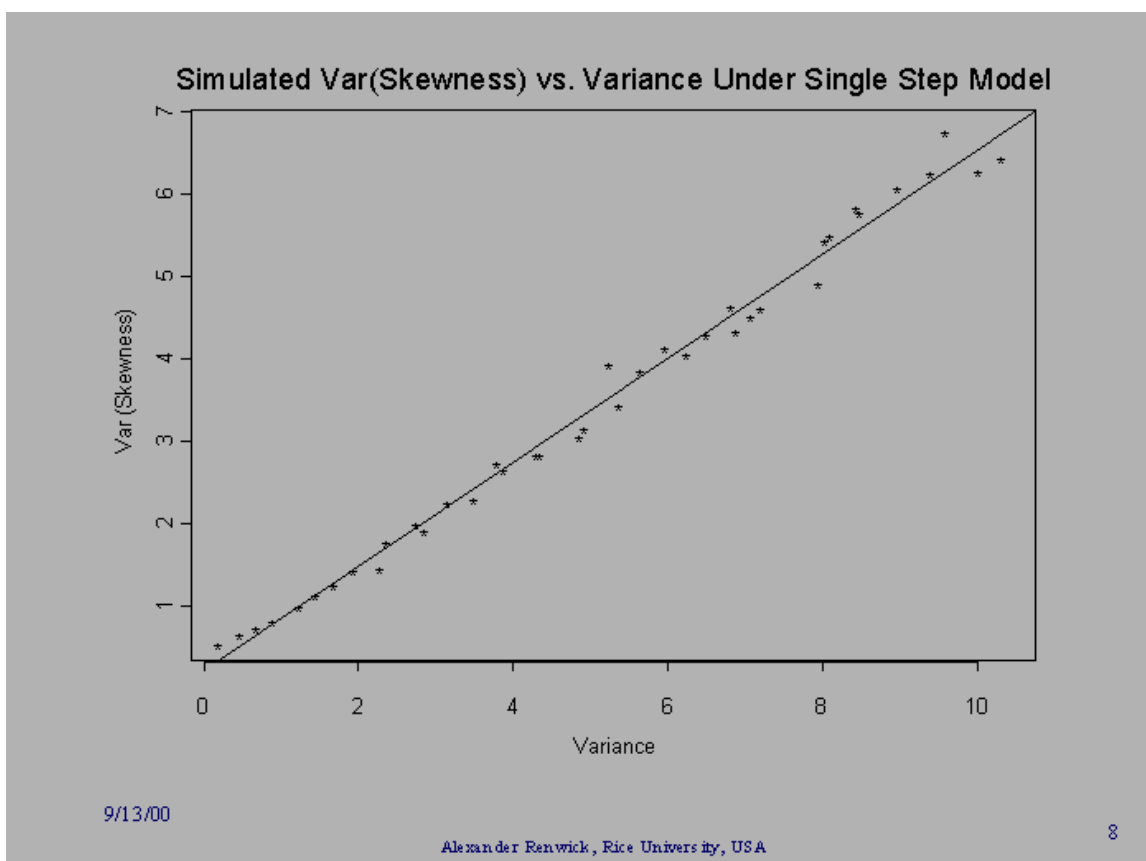


Figure 1.2: Single Step Model

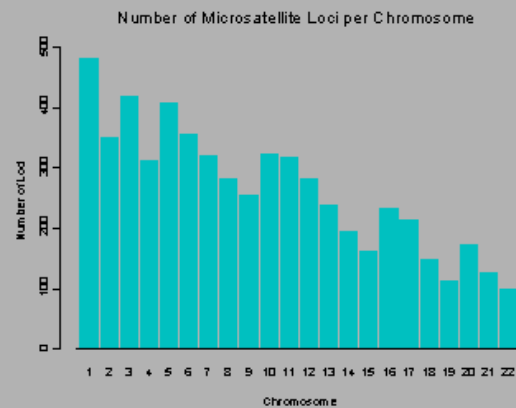
<sup>2</sup>This content is available online at <<http://cnx.org/content/m11110/2.4/>>.

---

### Microsatellites Loci

## Data

- The data, culled from the online Genome Database, consist of,
  - 5800 dinucleotide repeat loci from 22 chromosomes.
  - approximately 40 samples per locus.



9/13/00

Alexander Renwick, Rice University, USA

9

**Figure 1.3:** Number of Microsatellite Loci per chromosome

---

## Solutions to Exercises in Chapter 1

### Solution to Exercise 1.1 (p. 5)

Here is the answer



## Chapter 2

# Statistical machine learning





## Chapter 3

# Computational geometry and robotics



# Chapter 4

## Template<sup>1</sup>

### 4.1 The name of the section

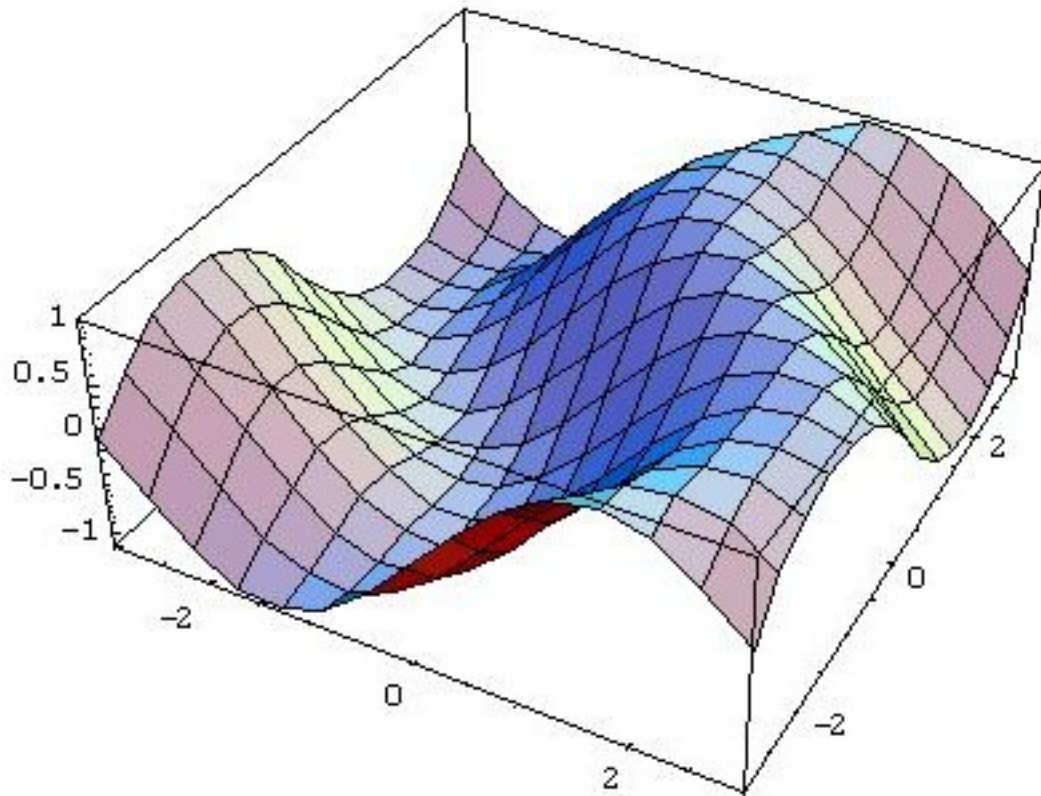
The regular paragraph content should be placed here. This paragraph contains also the example of Connexion-based link to course introductory page (Section ) as well as the example of the Connexions-external link to Microarray presentation page<sup>2</sup>.

---

<sup>1</sup>This content is available online at <<http://cnx.org/content/m11117/2.15/>>.

<sup>2</sup><http://www.bio.davidson.edu/courses/genomics/chip/chip.html>

The name of the figure



**Figure 4.1:** The caption of the figure should be here

---

The name of the list

- item 1
- item 2
- item 3

```
//the programming code
for (i=0;i<100;i++){
  for (j=i;j<100;j++){
    S[i,j]+=i*j;
  }
}
```

**Definition 4.1: defined term**  
here is the meaning of the defined term

**Example 4.1**  
The example illustrating the definition. Instead of paragraph a figure can be used as an example

**The integral of some function**

$$\int_{-1}^1 \sqrt[3]{\frac{x + \sin(x)}{a^2 + \cos(x)}} dx$$

(4.1)

**Exercise 4.1**  
This is a question to be answered by a student

(Solution on p. 16.)

**Law 4.1:**  
Here is the content of the presented law. Can be also expressed by equation

**Proof:**  
The proof of the law. The proof can be also more sophisticated structure than just the paragraph.

---

**The name of the table**

header 1	header 2	header 3
value 11	value 12	value 13
value 21	value 22	value 23

**Figure 4.2:** The table can have not only the general title above, but also the capture, which is just here

---

## Solutions to Exercises in Chapter 4

### **Solution to Exercise 4.1 (p. 15)**

And a solution to the question

## Glossary

**D** defined term

here is the meaning of the defined term

## Index of Keywords and Terms

**Keywords** are listed by the section with that keyword (page numbers are in parentheses). Keywords do not necessarily appear in the text of the page. They are merely associated with that section. *Ex.* apples, § 1.1 (1) **Terms** are referenced by the page they appear on. *Ex.* apples, 1

**B** bioinformatics, DNA microarrays, gene expression, classification, machine learning, robotics, genetic networks, § (1)

**C** CNXML, § 4(13)

**D** defined term, 15

**G** gene expression, northern blotting, southern blotting, § 1.1(4)

**M** microsatellites, mutation behavior, probabilistic models, § 1.2(5)

**T** tags, § 4(13)  
templates, § 4(13)



## Attributions

Collection: *Bioinformatics - From Sequence Expression and Structure*

Edited by: Krzysztof Cyran

URL: <http://cnx.org/content/col10202/1.24/>

License: <http://creativecommons.org/licenses/by/1.0>

Module: "Introduction"

By: Krzysztof Cyran, Marek Kimmel, Rudy Guerra

URL: <http://cnx.org/content/m11106/2.17/>

Pages: 1-2

Copyright: Krzysztof Cyran, Marek Kimmel, Rudy Guerra

License: <http://creativecommons.org/licenses/by/1.0>

Module: "Gene expression - blotting"

By: Krzysztof Cyran

URL: <http://cnx.org/content/m11107/2.6/>

Pages: 4-5

Copyright: Krzysztof Cyran

License: <http://creativecommons.org/licenses/by/1.0>

Module: "DNA Microsatellites Evolution"

By: Krzysztof Cyran

URL: <http://cnx.org/content/m11110/2.4/>

Pages: 5-6

Copyright: Krzysztof Cyran

License: <http://creativecommons.org/licenses/by/1.0>

Module: "Template"

By: Krzysztof Cyran

URL: <http://cnx.org/content/m11117/2.15/>

Pages: 13-16

Copyright: Krzysztof Cyran

License: <http://creativecommons.org/licenses/by/1.0>

## **Bioinformatics - From Sequence Expression and Structure**

NOTE: THIS COURSE IS IN PRAPARATION STAGE. AS LONG AS THIS NOTE IS VISIBLE, DO NOT RELY ON THE CONTENT!!!!!! This is an advanced research-oriented Bioinformatics course: From Sequence to Expressions and Structure

## **About Connexions**

Since 1999, Connexions has been pioneering a global system where anyone can create course materials and make them fully accessible and easily reusable free of charge. We are a Web-based authoring, teaching and learning environment open to anyone interested in education, including students, teachers, professors and lifelong learners. We connect ideas and facilitate educational communities.

Connexions's modular, interactive courses are in use worldwide by universities, community colleges, K-12 schools, distance learners, and lifelong learners. Connexions materials are in many languages, including English, Spanish, Chinese, Japanese, Italian, Vietnamese, French, Portuguese, and Thai. Connexions is part of an exciting new information distribution system that allows for **Print on Demand Books**. Connexions has partnered with innovative on-demand publisher QOOP to accelerate the delivery of printed course materials and textbooks into classrooms worldwide at lower prices than traditional academic publishers.