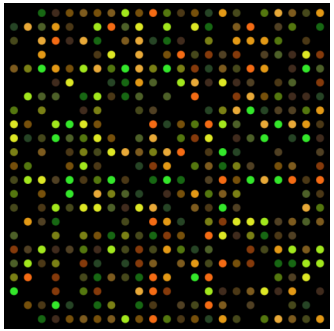
Introduction to Systems Biology
Lecture 7 Part A-1 Iyengar

Gathering and Analyzing Large Data Sets

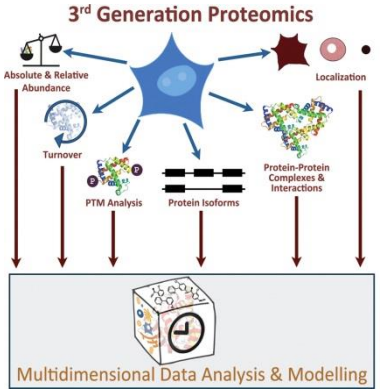
Technologies developed over the past two decades allow for the measurement of many cellular components such as mRNA and proteins simultaneously

Measurement of mRNAs by DNA Microarrays



DNA microarray Wikimedia Commons Guillaume Paumier

Measurement of proteins by Mass-spectrometry



Lamond AI et al. Mol Cell Proteomics 2012;11:O112.017731

These types of experiments can provide information about how the cellular landscape changes upon perturbation such as -

during signaling

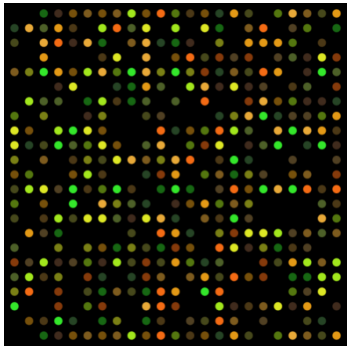
during change in cell state - differentiation

Onset of disease state

Introduction to Systems Biology
Lecture 7 Part A-2 Iyengar

Genomics

Started with development of **Microarrays**



Wikimedia Commons
Guillaume Paumier

Based on Southern blots (developed by E. Southern) –
DNA-DNA probe hybridization to identify specific DNA sequences

Used to quantify and identify mRNA expression profiles
under changing conditions - Transcriptomics

Extract mRNA → Convert to cDNA → Couple to dyes →hybridize →visualize → Computationally analyze
by reverse transcriptase to separate signal from noise

A typical micro array is shown. Each spot can represent a gene

Classic paper Iyer VR et.al. Science 283: 83 (1999) >8500 genes analyzed .
In response to serum stimulation of fibroblasts expression of over 517 genes changed
Serum induces proliferation …a broad picture of the cellular program moving it to a proliferative state

Introduction to Systems Biology
Lecture 7 Part A-3 Iyengar

Measurements in Genomics - I

Sequencing the whole genome --
Massively parallel DNA sequencing technologies -- enable whole genome sequencing in a week
Deep Sequencing - Repeated sequencing of a DNA fragment – region of interest in a chromosome
Substantial increase in sensitivity and accuracy

Tells you everything about the genome in terms of sequence -
Need extensive computational analysis to separate signal from noise

SNPs -- Single nucleotide polymorphisms - single base pair variations in the genome that occur with and relatively high frequency

CNVs -- copy number variation - alterations in DNA structure such a region of the chromosome is abnormally duplicated or deleted

Exome Sequencing -- Sequencing the expressed genome
Separate the part of the whole genome that codes for proteins (the exons) and then sequence
Often used to study genetic variations involved in disease states

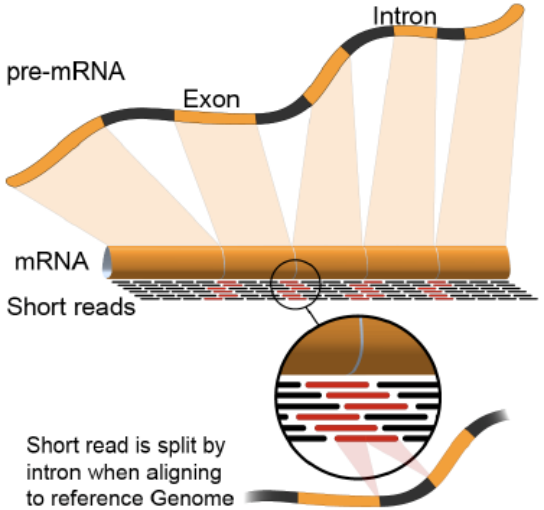
CHiP- Seq Sequencing transcription factor bound DNA
CHiP - chromatin – immunoprecipitation - using an antibody against a transcription factor of interest
Precipitate the DNA crosslinked to TF-- Sequence
To identify DNA binding sites for a TF on a genome-wide basis

Introduction to Systems Biology
Lecture 7 Part A-4 Iyengar

Measurements in Genomics-II

RNA Seq - Sequencing the expressed mRNA

Extract and fragment RNA
Convert fragments to cDNA
Sequence DNA fragments and map on to reference genome



Wikipedia RNA-Seq

Advantage: very quantitative and can be very precise
For genes that are moderate and expressed at low levels
Many reads are required for accurate measurement of levels
In analysis of the fly genome ---
even after 50 million mapped reads , all transcripts had not been found
(Malone & Oliver BMC Biology 9:34)
Currently mRNA-Seq is several times costlier than microarrays

Introduction to Systems Biology
Lecture 7 Part A-5 Iyengar

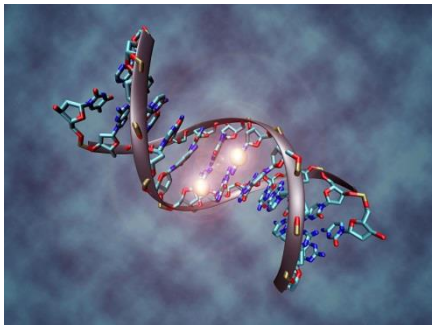
Measurements in Genomics - III

DNA Methylation

Addition of methyl groups to C in DNA in mammals

Typically 5' position of C in CpG dinucleotides are methylated leading to inhibition of gene expression

Detection by genome wide bisulfite sequencing
Bisulfite converts C→ U but not Me-C
Has some limitations but useful to know about silencing of genes



The crystal structure of a short DNA helix with sequence "accgcCGgcgcc", which is methylated on both strands at the center cytosine
Christoph Bock - Wiki Media Commons

MicroRNAs miRNAs

Small (21-25 nucleotide) RNA - regulates gene expression
Can be sequenced using RNA-Seq starting with size selected RNAs
Around 1100 human mirs mir-001, or mir-123 or mir-500

Each mir can target multiple mRNA and each mRNA can be regulated by several mirs --- Some combinatorial complexity
Aberrant mir levels associated with disease—cancers, heart disease

Why are these genomic measurements essential for a systems-level understanding ?

- All changes in cell state both normal and disease related are at least part due to changes in gene expression – hence surveying mRNA expression patterns is informative

- Both levels and activities of proteins are regulated by genomic and epigenomic characteristics and hence these characteristics are often involved in disease initiation and progression

  mutated genes produce continuously active proteins e.g. mutant Ras: lung, colon

  changes in levels of proteins by methylation status or mirs also regulate both normal and disease process

Introduction to Systems Biology
Lecture 7 Part A-7 Iyengar

Proteomics

Measuring the levels of many proteins simultaneously by mass-spectrometry

By measuring levels

Identify proteins within protein complexes

Identify proteins in an organelle

Determine the protein composition in a cell at a certain state

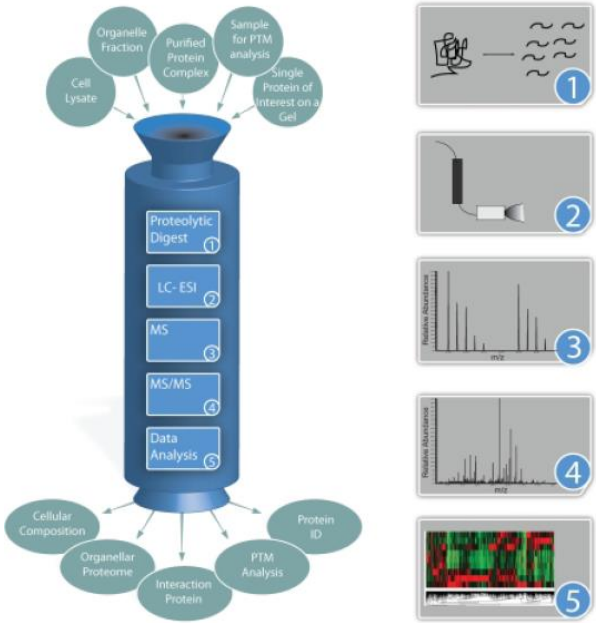
By measuring protein state

Dynamics of cellular activities

Phosphoproteomics measuring phosphorylated peptides

Ser, Thr or Tyr

Acetylation – Lys - functional complexes



Walther TC & Mann M (2010)
JCB 190: 491- 500

Introduction to Systems Biology
Lecture 7 Part A-8 Iyengar

Why do we need proteomics ?

A recent study by

Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z. (2012)

Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells.

Mol Syst Biol. 2012;8:608

Aneuploidy –extra chromosome copies in a cell - often can result in serious disease such as Down's syndrome (trisomy of chromosome 21)

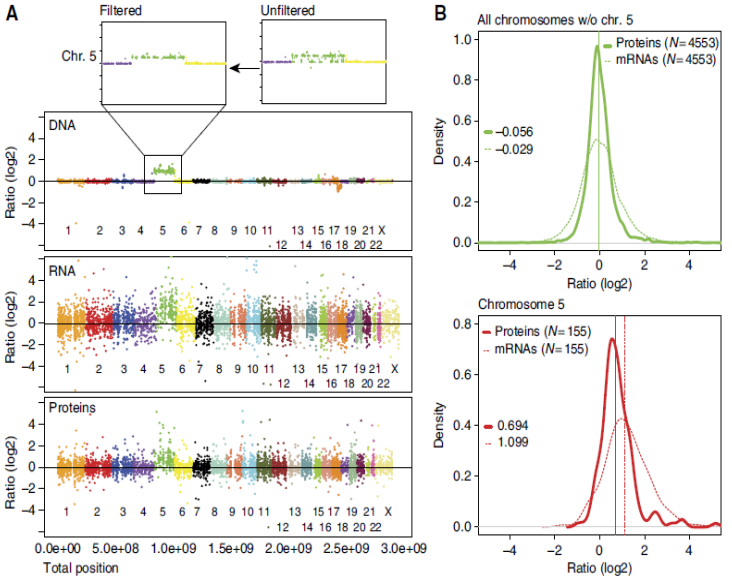
Study checked what happened in cells with extra chromosomes

mRNA levels - by microarray…protein levels by mass spectrometry

Cells try to compensate for protein levels…In fact amount of protein does not proportionally increase with amount of increased DNA or mRNA

Free subunits of complexes are degraded

Autophagy is stimulated



Stingele A et al (2012) MSB 8: 608

Introduction to Systems Biology
Lecture 7 Part A-9 Iyengar

Metabolomics

Metabolome

The full set of metabolites found in a cell, tissue organism

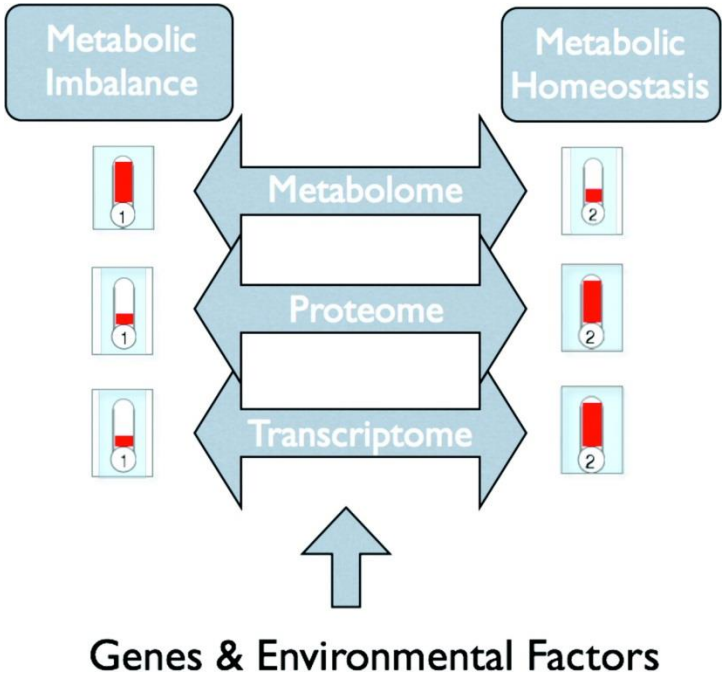
Useful in understanding how phenotypic changes occur or not

Example 1 - small change in DNA – point mutation leads to small change in level of mRNA or protein, but big change in activity

Example 2 - Stress induced big changes in mRNA and protein but no big change in metabolites or function

Mass spectrometry coupled to chromatography (Gas or Liquid)

NMR – generally not high throughput



Mayr M Circ Cardiovasc Genet 2008;1:58-65