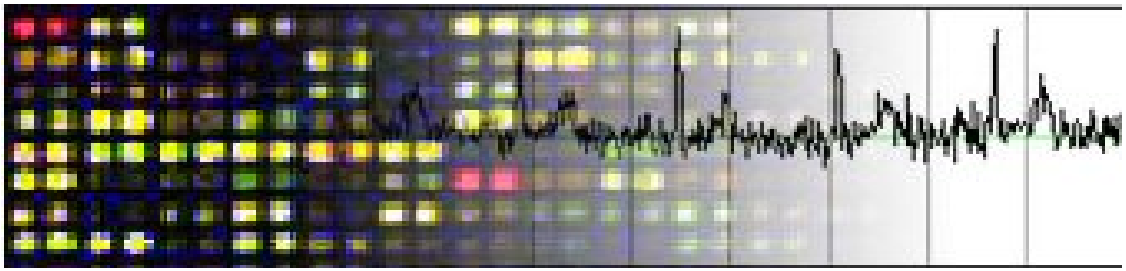
20.453J / 2.771J / HST.958J Biomedical Information Technology
Fall 2008



Biomedical Information Technology

2.771J 20.453J HST.958J SMA5304 Spring 2008

Lecture 9 September 2008, by Hanry Yu

Gaps in Bio-Medical Data and Applications

Scope of Applications (1 week: HY)

*Types and characteristics of biological and medical data **(today)***

Medial Data
Biological Data
Current challenges: fill in the gaps

Examples from liver fibrosis Research (on Thursday)

Biomedical Information Technology

Molecular Networks and Medical Practice

Diagram of myosin molecule structure removed due to copyright restrictions.

Photomicrograph images of cells removed due to copyright restrictions.
See Nikon Small World (http://nikonsmallworld.com)

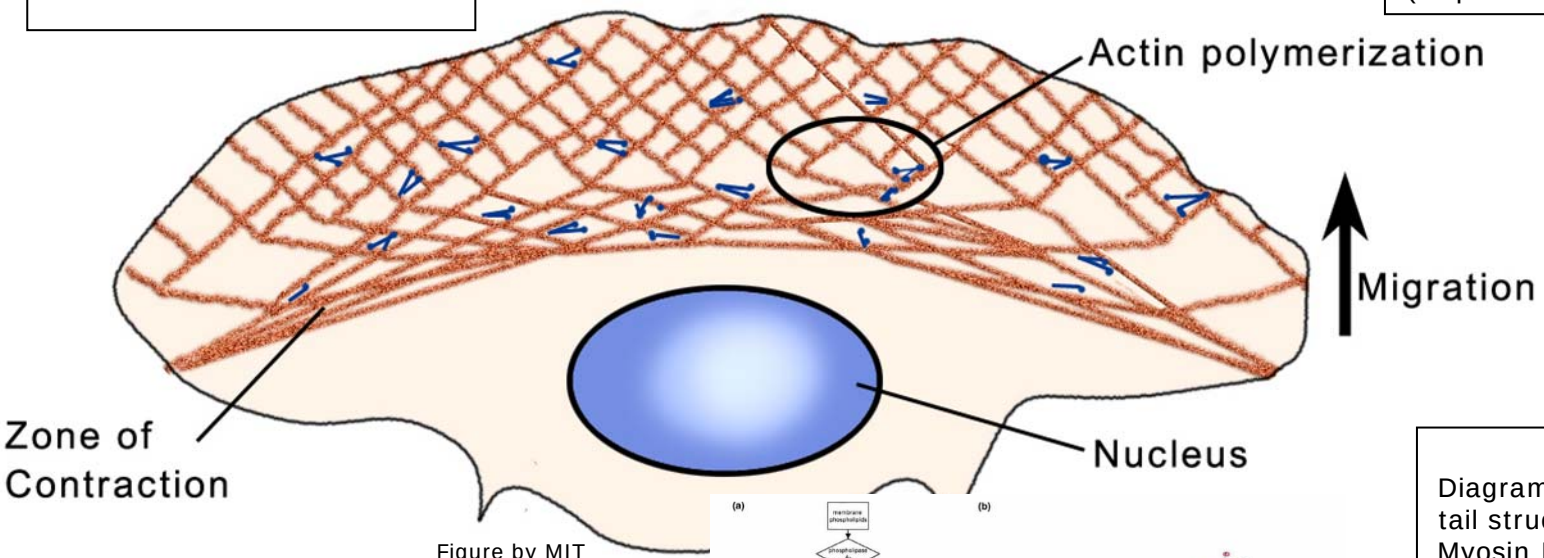
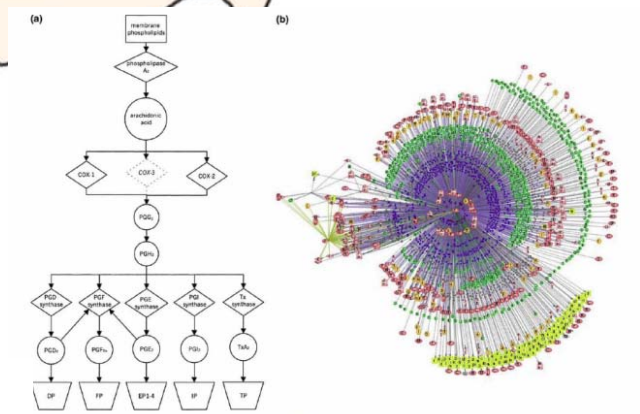


Figure by MIT OpenCourseWare.



Diagrams of head / neck / tail structure for Myosin I, Myosin II, and Myosin V. (from Alberts et al, *Molecular Biology of the Cell*) removed due to copyright restrictions.

Image courtesy of Elsevier, Inc., http://www.sciencedirect.com. Used with permission. Source: Apic, G., et al. FEBS Letters 579 no. 8 (2005): 1872-1877.

What are Medical Data?

- Documentation about individual patient's illness and medical care
- Includes:
  - Medical history
    - All medical events and problems experienced by patient
  - Clinical findings
    - Symptoms: reported by patients (subjective)
    - Signs: detected by physician during physical examination (objective)
    - Laboratory test results
  - Diagnoses
    - Process of identifying disease by its signs, symptoms and laboratory test results

What are Medical Data?

- Therapies
  - Description of application of treatment to effect a cure
  - Treatment includes:
    - Medication, dosage and dosing schedule
    - Care regime, e.g. diet, exercise, etc.
- Prognosis
  - Duration of disease
  - Chances of complications
  - Probable outcomes
  - Prospects for recovery
  - Recovery period
  - Survival rates
  - Death rates

Questions Medical Data Can Answer

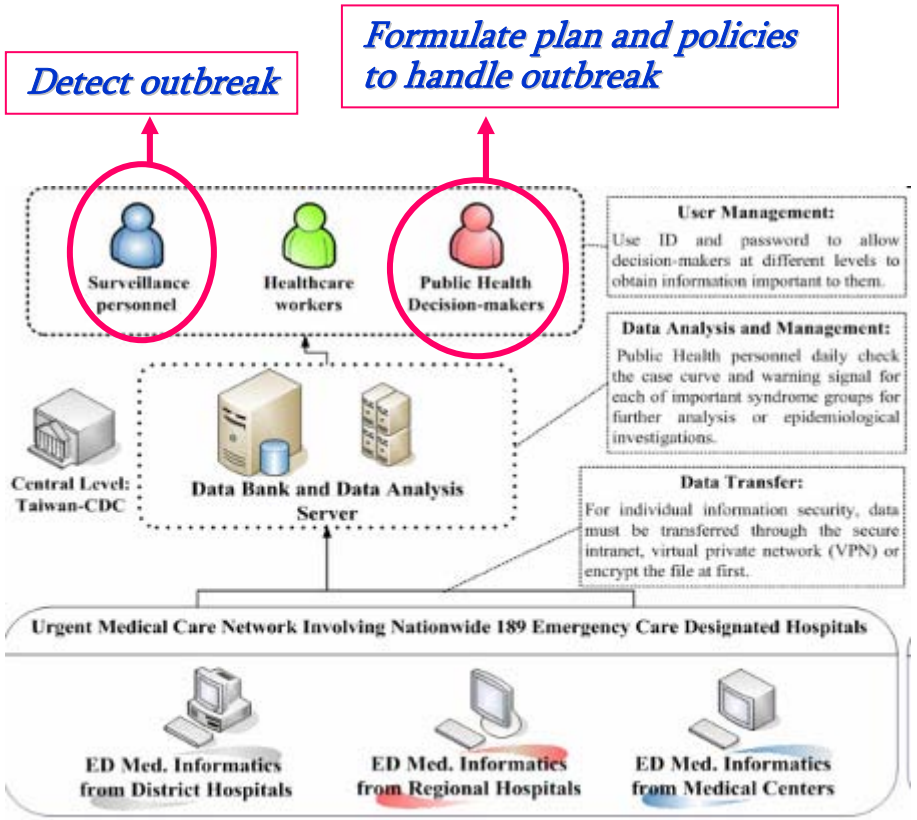
- What is the patient's medical history?
- What are the symptoms?
- What are the examination findings?
- What are the changes in symptoms and signs over time?
  - Progression of chronic diseases
  - Provides information for prescribing treatment
- What are the changes in physiological function over time?
- What were previous treatment given?
- What was the rationale for treatment?

Purposes of Medical Data

- Supports patient care
  - Basis of historical record
  - Communication among care providers
    - Allergic conditions
    - Treatment plan for chronic diseases (e.g. high blood pressure, diabetes)
  - Anticipate future health problems
- Supports medical/clinical research
  - Assists in screening of patient groups for clinical trials
    - Patients' profile can be matched with clinical trial requirements
    - Electronic health record systems can perform match and prompt doctors of suitable candidates during patient's visit – improve recruitment rate
  - Electronic health record systems can systemically generate hypotheses for research based on patients' information
  - Mechanisms and causes of diseases
    - Formulation of possible treatments
  - Assists in studying the efficacy of drugs and medical equipment

Purposes of Medical Data

- Allows surveillance of epidemic and bio-terrorism
  - Epidemic: SARS, avian flu etc
  - Bio-terrorism: anthrax, smallpox etc
  - Electronic medical record systems allows real-time nationwide surveillance.
- Provides legal record



The System Architecture of Nation-Wide Hospital Emergency Department-based Syndromic Surveillance System (ED-SSS) in Taiwan, established in 2003.

Source: Wu, T-S J., et al, 2008. "Establishing a nationwide emergency department-based syndromic surveillance system for better public health responses in Taiwan. *BMC Public Health* 8 (2008): 18. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2249581

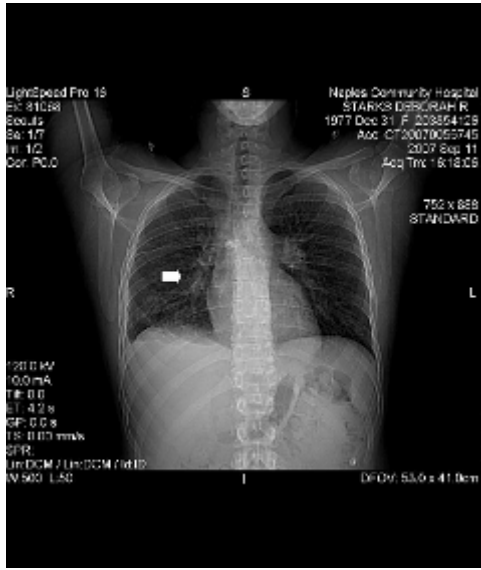
Image courtesy of Wu, T-S J., et al.

Types of Medical Data

- Images
  - Computed tomography (CT) scan
  - Magnetic resonance imaging (MRI) scan
  - Positron emission tomography (PET) scan
  - Ultrasound imaging etc
- Recorded signals
  - Blood pressure
  - Electrocardiogram (ECG) etc
- Numerical measurements
  - Temperature
  - Laboratory results etc
- Textual description
  - Medical symptoms or signs description
  - Treatment plan
  - Prognosis etc

Image Data - CT scan

- Diagnostic procedure that takes a detailed cross-sectional x-ray picture of a "slice" of the body
  - X-ray machine rotates around the patient and takes pictures from many angles
  - Computer combines the pictures into a very detailed cross-sectional image
- Show the shape, size and the exact location of organs and tissues in any "slice" of the body
- Good spatial resolution
  - Able to distinguish 2 structures which are very near each other as separate entities
- Image size of CT scans depend on:
  - Number of “slices”
  - Resolution of scanner
  - Area scanned
- Other important information:
  - Device type
  - Device settings



CT scan of Fibrosing Mediastinitis

Source: The FM Foundation. http://www.mrsnewengland2007.com/FM-Medical-Information.html

Image Data - MRI

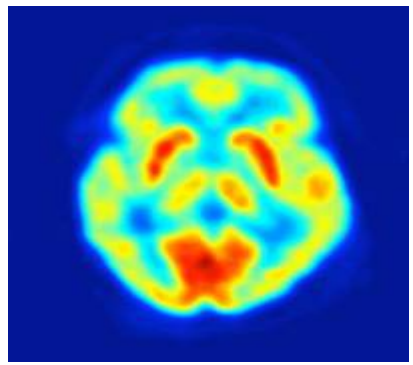
- Diagnostic procedure that uses magnetic/radio waves to affect the body's atoms
  - Radio waves force nuclei (usually hydrogen – body consists mainly of water) into different position
  - When they restore their position, radio waves are emitted
  - Scanner picks up signals and a computer uses the signal to compose the image
  - Image formed is based on location and strength of the signals
- Used to visualize structure and function of the body
- Provides detailed images of body in any plane
- Provides much greater contrast between soft tissues than CT scan
  - Useful for brain, musculoskeletal, cardiovascular and cancer imaging
- Contrast agents sometimes used to enhance appearance of blood vessels, tumors or inflammation.
- Image size of MRI scans depend on:
  - Number of "slices" (for 3D reconstruction)
  - Resolution of scanner
  - Area scanned
- Other important information:
  - Device type
  - Device settings

Image removed due to copyright restrictions.
See http://en.wikipedia.org/wiki/File:Structural.gif.

Animated MRI images of a human head
Source: Dawyne Reed. http://en.wikipedia.org/wiki/Image:Structural.gif

Image Data - PET

- Nuclear medicine imaging technique
  - Patients given injection of very small amount of tracer (e.g. fluorine 18 – radioactive version of glucose. Can be used to image brain where glucose is the main source of energy)
  - PET scanner used to detect emission from the injected tracer
  - 2D and 3D images of the scanned area is created by the computer
- Can be used for:
  - Early detection and monitoring of cancer
    - Can reveal changes in metabolism and how organs and tissues are working
    - Can show if (and where) cancer is spreading to other parts of the body
  - Neurological disease
    - Can provide biochemical function information of the brain
  - Assessment of cardiovascular disease
    - Can be used to assess blood flow to the heart and how the heart is functioning
- Limitations:
  - Can give false results if chemical balances within body are not normal.
  - Resolution of structures of the body may not be as clear compared with CT although information gained from PET scan is much more.



PET image of human brain

Courtesy of Jens Langner. http://en.wikipedia.org/wiki/Image:PET-image.jpg

Image Data – Multi-Modal Imaging

- Combination of PET with CT and MRI
  - CT and MRI scan provides anatomic information
  - PET scan provides metabolic information
- Image size of MRI scans depend on:
  - Number of "slices" (for 3D reconstruction)
  - Resolution of scanner
  - Area scanned
- Other important information:
  - Device type
  - Device settings

CT scan *PET scan* *CT-PET scan*

Image removed due to copyright restrictions.
See http://www.mayoclinic.org/pet/.

Lung cancer

Scans showing lung cancer (bright spot in the chest).
Left: CT scan; Center: PET scan; Right: combined CT-PET scan
Source: Mayo Clinic. http://www.mayoclinic.org/pet/

Recorded Signals - ECG

- Records the electrical activity of the heart over time
- Measured by an array of electrodes placed on the body surface
- American Heart Association requires the ECG signal to consist of 3 individual leads, each recording 10 bits per sample, and 500 samples per second
- Some ECG signals, may require 12 leads, 11 bits per second, 1000 samples per second, and last 24 hours
  - This ECG record requires 1.36 gigabytes of storage when converted to digital format

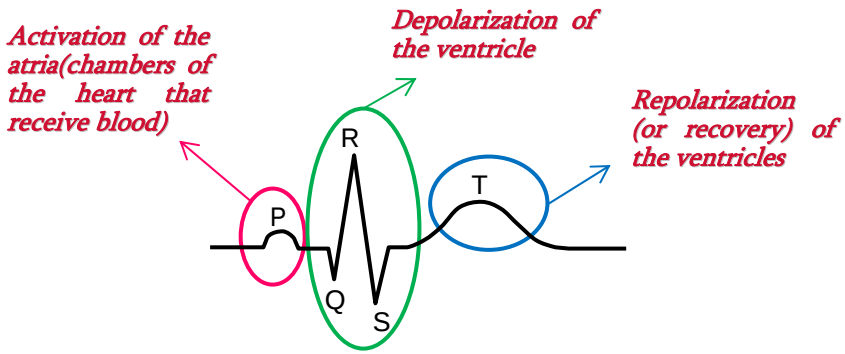
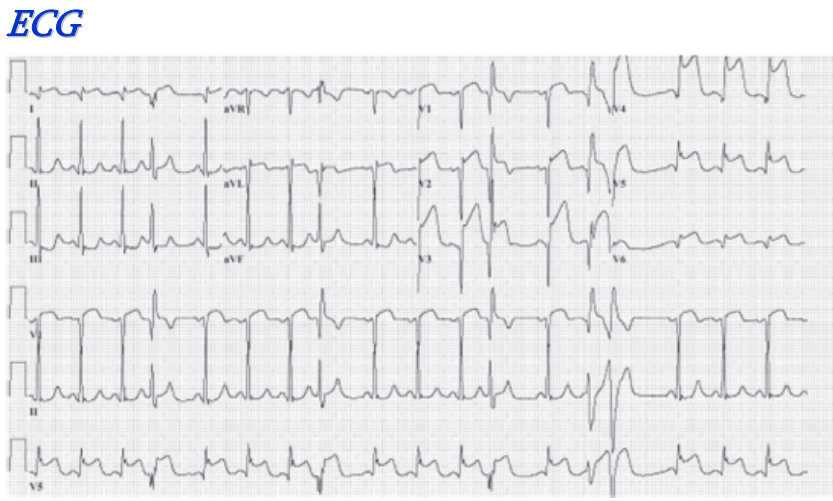


Figure by MIT OpenCourseWare, after Lynch, 1985.



Source: Clifford, G. F. Azuaje, and P. McSharry. "The Physiological Basis of the Electrocardiogram." *Advanced Methods and Tools for ECG Data Analysis*. Norwood, MA: Artech House, 2006. Courtesy of Artech House. Used with permission. © Artech House, 2006.

Characteristics of Medical Data

- Medical data = single instance of observation (signs, symptoms, clinical findings) or description of diagnosis, therapy or prognosis

- Defined by:
  - Patient
  - Parameter being observed
  - Description of parameter
    - Numerical value
    - Textual description
    - Graphical image
  - Device used (when applicable)
  - Device setting (when applicable)
  - Time of observation

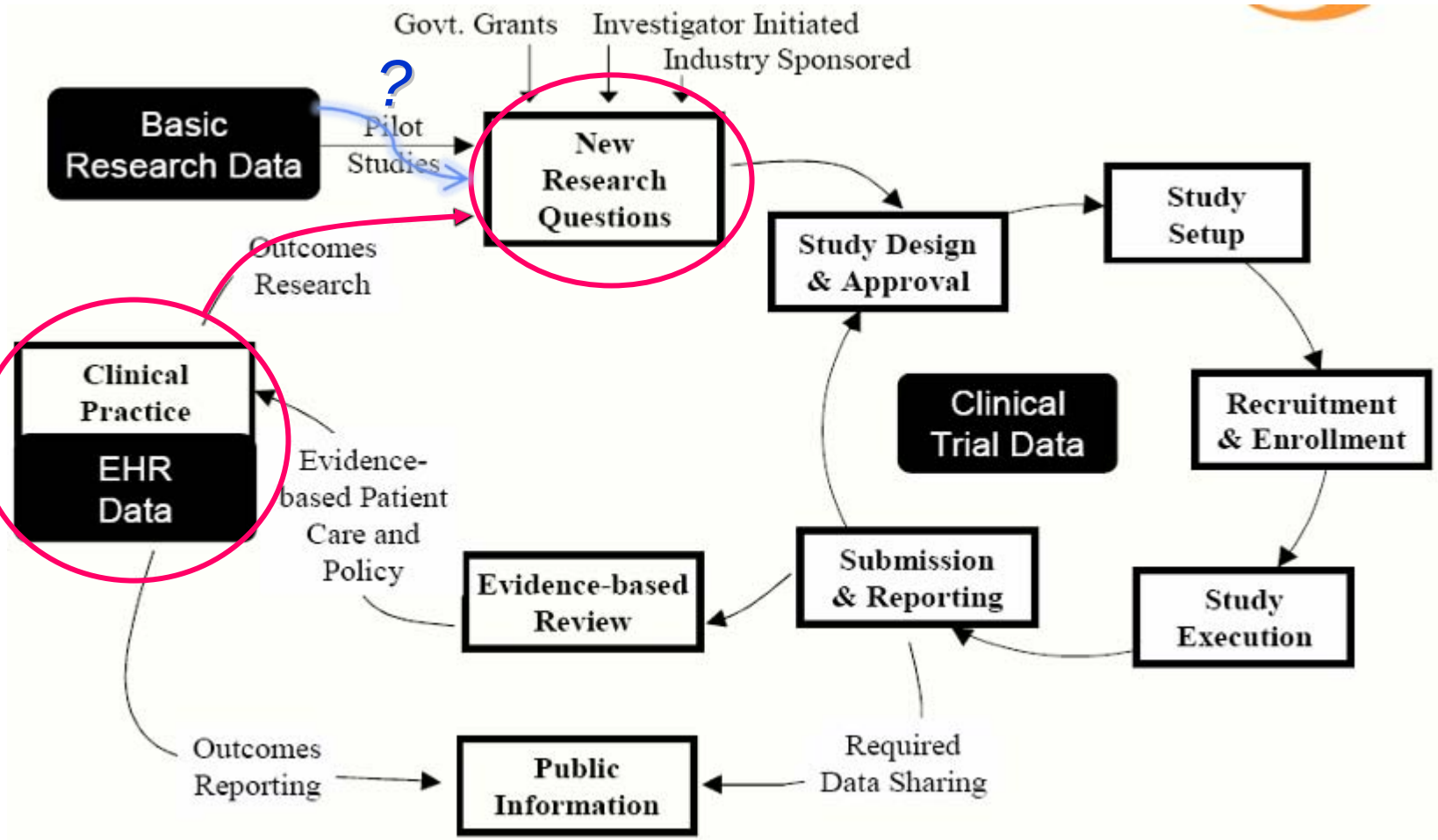
Example:

Peter Tan has a fever of 38.1°C on 20th May 2008 at 5.30pm

Patient: Peter Tan
Parameter: Body temperature
Description: 38.1°C
Time: 20th May 2008, 5.30pm

- Medical data = collection of medical data

Lifecycle of Current Clinical Research



Source: Kahn, M.G. Integrating Electronic Health Records and Clinical Trials – An Examination of Pragmatic Issues. http://www.esi-bethesda.com/ncrrworkshops/clinicalResearch/pdf/MichaelKahnPaper.pdf

Courtesy of Michael G. Kahn MD, PhD. Used with permission.

What are Biological Data?

- Chemical Molecules (DNA, RNA, Carbohydrates, Proteins, Lipids): Identities & Interactions
- Pathways and networks
  - ~25000 genes, 100-200 pathways and a fraction relevant to diseases per cell type (Bauch A. et al., Immune Review 2006, 210: 187-207)
- Phenotypes (*in vitro* and *in vivo*)
- Meso-cale mechanistic studies

Biological Data

Image removed due to copyright restrictions.
See Fig. 5 in: Bauch, A., and G. Superti-Furga. "Charting Protein Complexes, Signaling Pathways, and Networks in the Immune System." *Immunological Reviews* 210, no. 1 (2006): 187-207.

Chemical Molecules

nucleic acids, proteins, carbohydrates, lipids

Images removed due to copyright restrictions.

"Anchoring of integral proteins to the plasma membrane…"
Figure 3-36 in Lodish, H., et al. Molecular Cell Biology. 4th edition. New York, NY: W. H. Freeman, 2000.
Viewable at the NCBI Bookshelf
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mcb.figgrp.618

Diagram of phospholipid bilayer, containing retinal pigment.

Molecule structure image, highlighting small domain and large domain.

Conformations and Interactions

Molecular Signals and Metabolism via molecular interactions (nucleic acids, proteins, carbohydrates, lipids)

Image removed due to copyright restrictions.

Figure 6-10 in Silverthorn, D. *Human Physiology*. 2nd ed. Prentice-Hall, 2000.
http://cwx.prenhall.com/bookbind/pubbooks/silverthorn2/medialib/Image_Bank/CH06/FG06_10.jpg

Phosphorylation

Image removed due to copyright restrictions.

"The two branches of the inositol phospholipid pathway."
Figure 15-36 in Alberts, B., et al. *Molecular Biology of the Cell*. 4th edition. New York, NY: Garland Science, 2002.

Viewable at the NCBI Bookshelf
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mboc4.figgrp.2812

Pathway: e.g. Glycolysis Regulation

Image removed due to copyright restrictions.
See http://www.biocarta.com/pathfiles/h_GLYCOLYSISPATHWAY.asp.

Pathway Interactions

Image removed due to copyright restrictions.

"Five parallel intracellular signaling pathways…"
Figure 15-61 in Alberts, B., et al. *Molecular Biology of the Cell*. 4th edition. New York, NY: Garland Science, 2002.

Viewable at the NCBI Bookshelf
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mboc4.figgrp.2866

Network Formation

Image removed due to copyright restrictions.

"Chart of the major signaling pathways relevant to cancer in human cells…"
Figure 23-31 in Alberts, B., et al. *Molecular Biology of the Cell*. 4th edition. New York, NY: Garland Science, 2002.

Viewable at the NCBI Bookshelf
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?highlight=23-31&rid=mboc4.figgrp.4327

Pathways and Networks

Image removed due to copyright restrictions.

"Glycolysis and the citric acid cycle provide the precursors needed to synthesize many important biological molecules."
Figure 2-87 in Alberts, B., et al. *Molecular Biology of the Cell*. 4th edition. New York, NY: Garland Science, 2002.

Viewable at the NCBI Bookshelf
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?highlight=2-87&rid=mboc4.figgrp.318

Image removed due to copyright restrictions.

"Glycolysis and the citric acid cycle are at the center of metabolism."
Figure 2-88 in Alberts, B., et al. *Molecular Biology of the Cell*. 4th edition. New York, NY: Garland Science, 2002.

Viewable at the NCBI Bookshelf
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?highlight=2-88&rid=mboc4.figgrp.320

Network Dimensions

Image removed due to copyright restrictions.

"A representation of all of the known metabolic reactions involving small molecules in a yeast cell." Figure 2-89 in Alberts, B., et al. *Molecular Biology of the Cell*. 4th edition. New York, NY: Garland Science, 2002. [From Jeong, H., et al. *Nature* 411 (2001): 41-42]

Viewable at the NCBI Bookshelf
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?highlight=2-89&rid=mboc4.figgrp.321

Importance of Biological Pathways

- Association with phenotypes and diseases
- Drug discovery

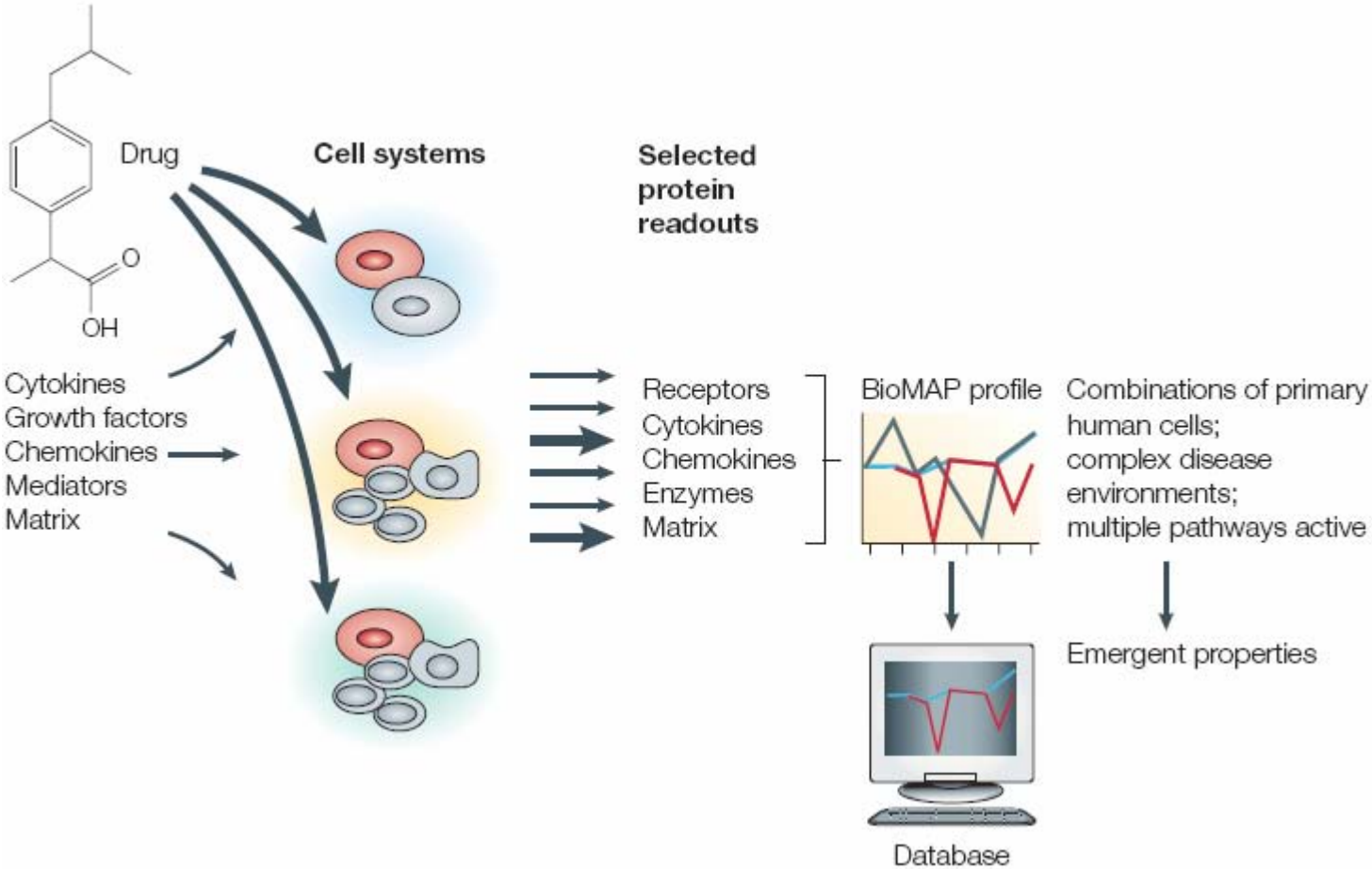
Image removed due to copyright restrictions.
See Table 1 in: Bauch, A., and G. Superti-Furga. "Charting Protein Complexes, Signaling Pathways, and Networks in the Immune System." *Immunological Reviews* 210, no. 1 (2006): 187-207.

Bauch, A., and G. Superti-Furga. "Charting protein complexes, signaling pathways, and networks in the immune system." Immunological Reviews 210, no. 1 (2006): 187-207.

Drug Discovery Approaches

Image removed due to copyright restrictions.
See Fig. 1 in: Butcher, E. C. "Can Cell Systems Biology Rescue Drug Discovery?" *Nature Reviews Drug Discovery* 4, (June 2005): 461-467.

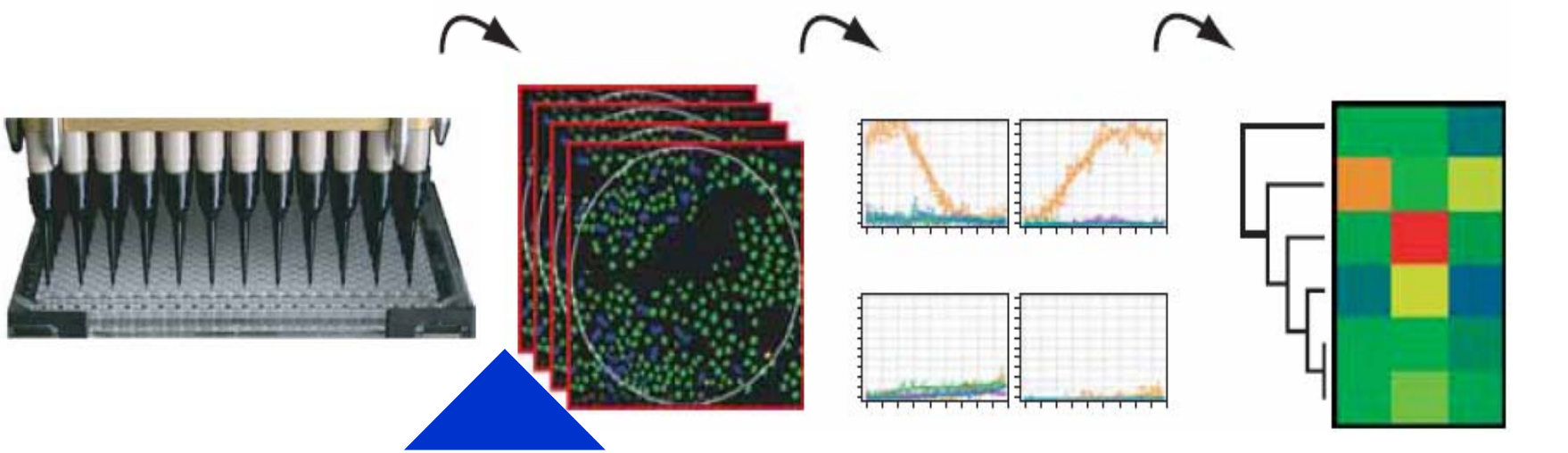
Cell Systems Biology Approach



Courtesy of BioSeek, Inc. Used with permission.

High Content Screening (HCS)

- High content screening combines automated microscopy with image analysis to capture multiple parameters of individual cells, to allow rapid, large scale and highly parallel biological research and drug discovery.



Courtesy of Jan Ellenberg. Used with permission.

(Neumann, Held et al. 2006)

Fluorescence Activated Cell Sorter (FACS) (Flow cytometry)

Quantitative, multiple parameter analysis of large numbers of individual cells

Cell surface markers
Intracellular proteins
Ca++ mobilization

12 colors and 15 parameters
Sorting: 60,000 cell/second

Ref.: Jianmin Chen, MIT

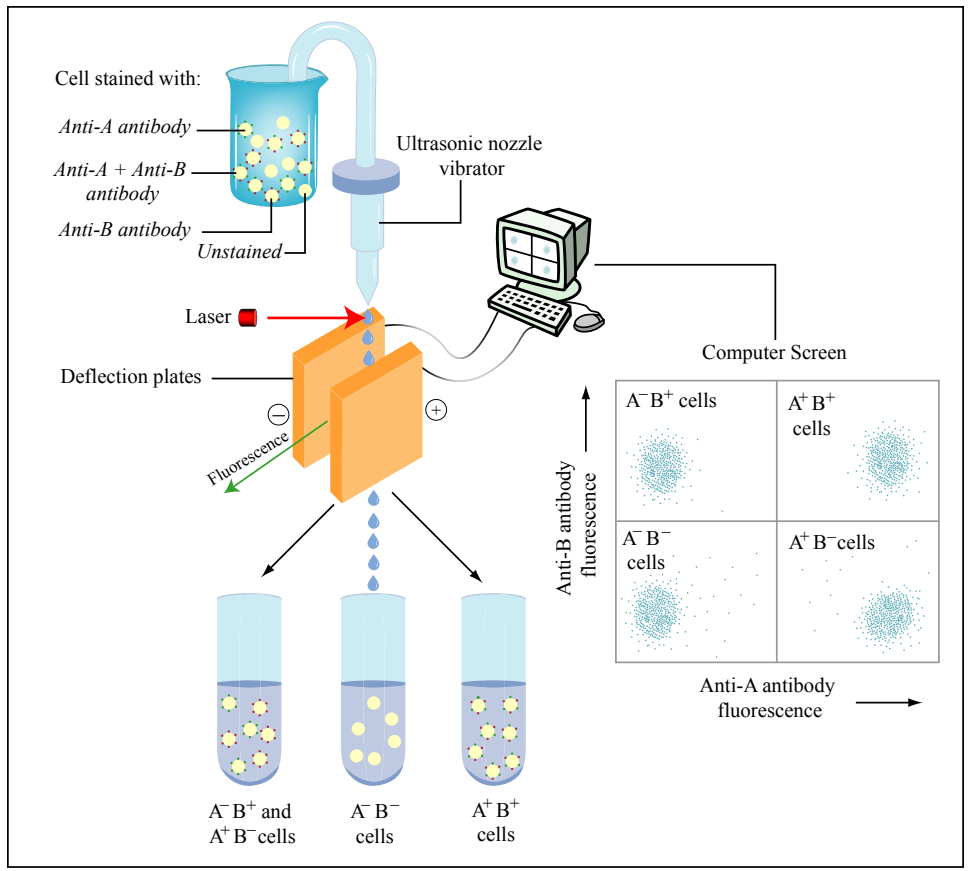


Figure by MIT OpenCourseWare

Advantages of HCS

- Combination of modern cell biology, with all its molecular tools
- High-throughput and high efficiency using automated system
- Obtain knowledge on multiple parameters in an individual experiment to give us rich information at single cell level
- Spatial and temporal information of the cells with their neighbors and environment

Applications of HCS

- Genome wide gene functional study, e.g. genome-wide RNAi approach (Pelkmans et al. 2005, Sonnichsen et al. 2005);
- Proteome sub-cellular localization (Huh et al. 2003);
- Protein-protein interaction, chemical profiling and drug screening (Perlman et al. 2004, Mitchison 2005, Abraham et al. 2004).

Systems Biology for Drug Discovery

- Target ID and validation
- Side effects
- Molecular mechanisms for toxicity

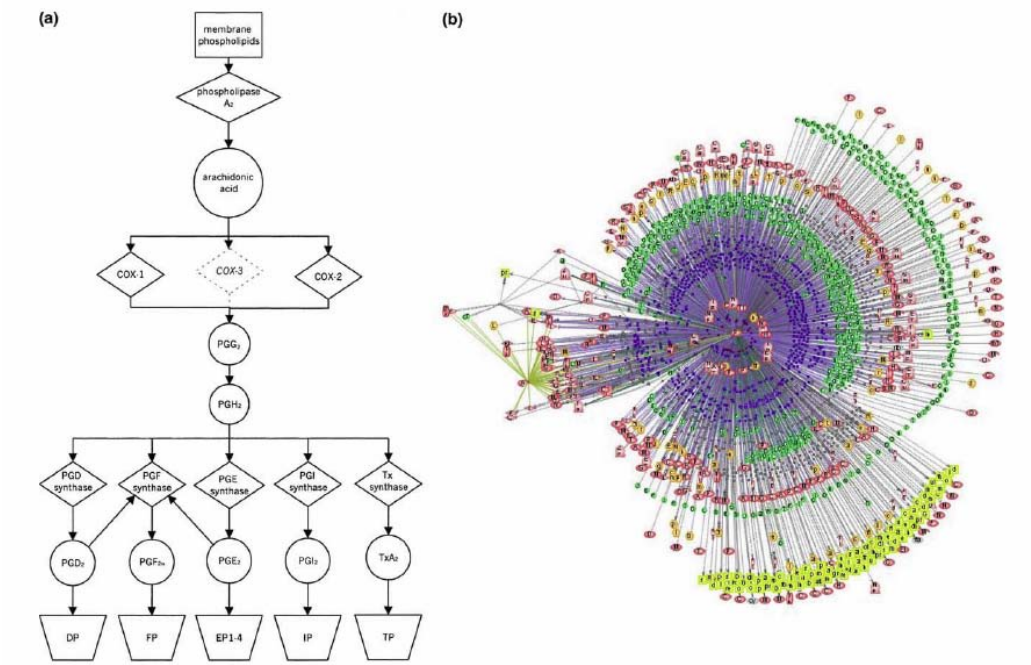


Fig. 1. Pathways involving Cox-2. Biochemical 'Pathway' of Cox-2 [25] (a) in the formation of prostanoids and the term connectivity (b) of Cox-2 as represented by the pathway-mining tool PathwayAssist. Red symbols denote proteins, yellow physiological processes and green represent small molecules.

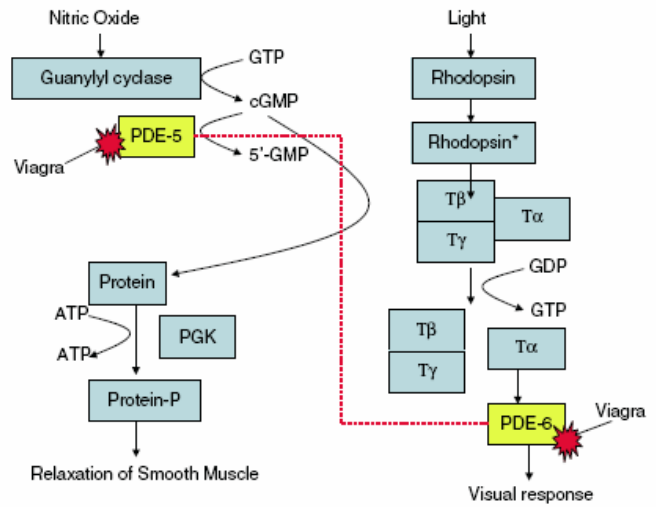
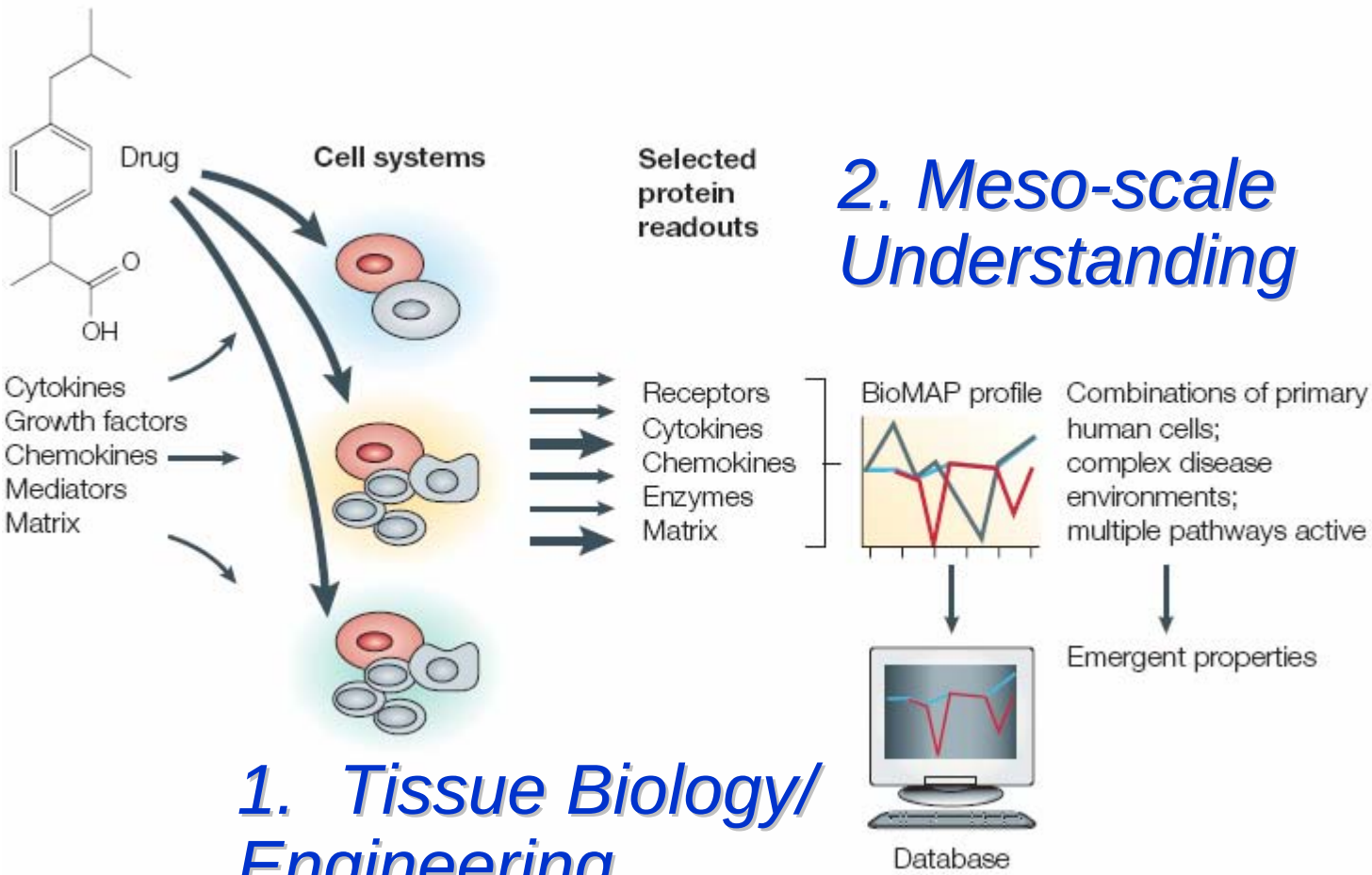


Fig. 2. Example of how drug-specificity problems might be predicted using pathway information. The figure shows simplified parts of the smooth-muscle relaxation pathway (left) and light-sensing pathway in the eye (right). The molecular target for Viagra is PDE-5, a phosphodiesterase in smooth muscle, which is homologous to PDE-6 in the eye (indicated by a broken red line), and to which Viagra (Sidenafil) also binds, leading to a well-documented side effect of blue vision in patients (e.g. [27]).

Cell Systems Biology Approach



Courtesy of BioSeek, Inc. Used with permission.

Gaps in Bio-Medical Data

- Medical Data: physiologic and pathologic phenotypes
- Biological Data: molecular parts and interaction maps
- Lack meso-scale functional modules or steps to construct Operations Manuals at cell/tissue/systems levels; and
- Repairs Manuals

The CAR analogy

Images removed due to copyright restrictions. Two diagrams of model car assembly,
from DuraTrax "Delphi" Assembly and Operation Manual, http://manuals.hobbico.com/dtx/dtxc0012-manual.pdf.

The airplane analogy

- Understand how an airplane work:
- Observe how an airplane flying phenotype:
- Take the plane apart into individual parts
- Record the interaction maps
- What is the missing information? Why it is difficult to construct an airplane operations manual from the description of the parts and their interaction maps?
- Meso-scale studies: functional modules or steps that are small enough to be readily described by parts; and yet large enough to link with other modules to understand how the whole plane works: engine, steering, fuel, navigation systems etc.

Functional Modules vs Molecular Networks

Diagram of myosin molecule structure removed due to copyright restrictions.

Photomicrograph images of cells removed due to copyright restrictions.
See Nikon Small World (http://nikonsmallworld.com)

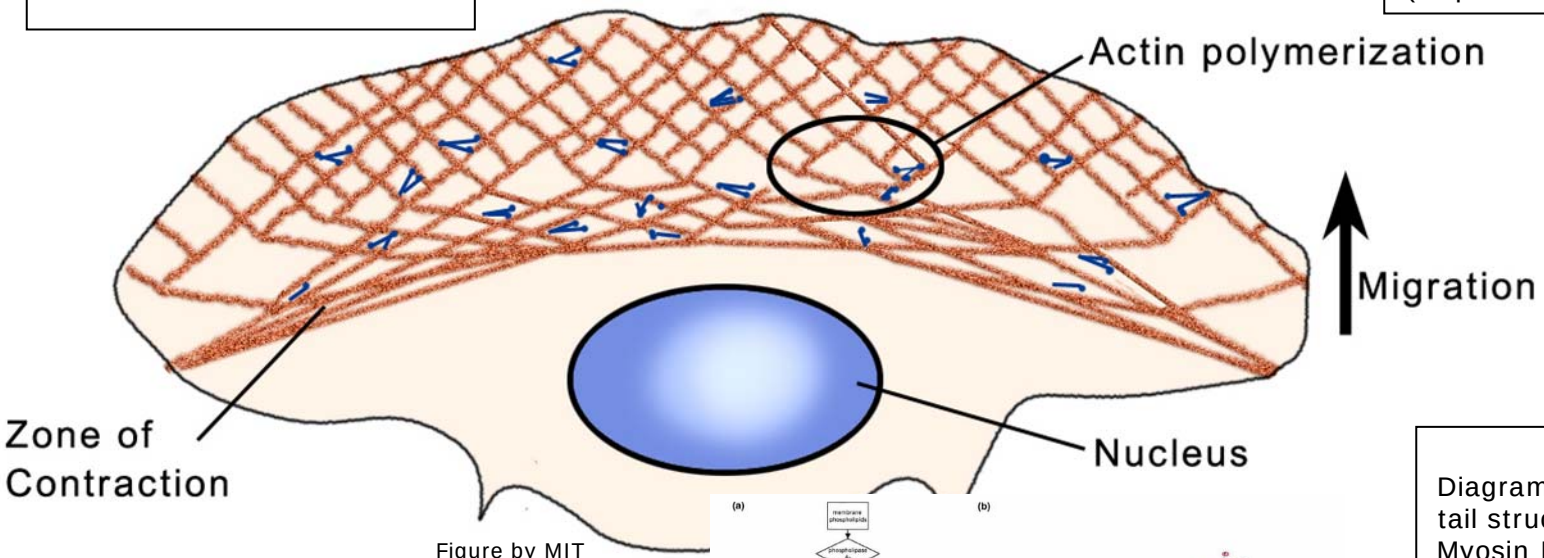
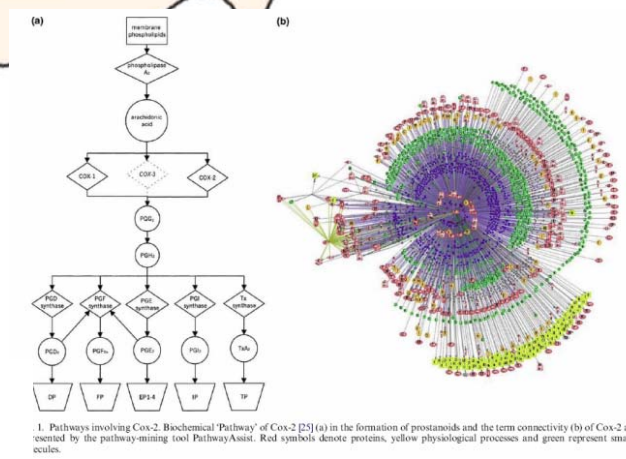


Figure by MIT OpenCourseWare.



Diagrams of head / neck / tail structure for Myosin I, Myosin II, and Myosin V. (from Alberts et al, *Molecular Biology of the Cell*) removed due to copyright restrictions.

Image courtesy of Elsevier, Inc., http://www.sciencedirect.com.
Used with permission. Source: Apic, G., et al. FEBS Letters 579 no. 8 (2005): 1872-1877.

Fill Gaps in Bio-Medical Data

- Meso-scale studies: a functional module or step of a larger process is typically defined by a group of chemical reactions or molecular networks.
- Link Molecular Networks to Physiology/Pathology
- Improve understanding of HOW things work
- Drug development
- Regenerative medicine
- Unique opportunities for computational approach to handle the complexity of constructing Operations Manual or Repairs Manual.

Things to Ponder…

Medical Practices

↑

Phenotypic Processes

↑

Functional Modules or Steps?

↑

Molecular networks

Next lecture: example in liver fibrosis

Image removed due to copyright restrictions.
Photo of liver with fibrosis.