Presenting data: how to convey information most effectively
Centre of Research Excellence in Patient Safety
20 Feb 2015

Biomedical Informatics: helping visualization from molecules to population
Dr. Guillermo Lopez-Campos
Health and Biomedical Informatics Centre (HABIC)

Outline

• Current context - Data explosion and complexity
• Biomedical Informatics
• Visualization from theory to Implemented solutions:
  – Data, Information, Knowledge and Wisdom visualization
  – How data are currently presented
    • Genomics and other “omics”
    • Integrating different information sources
• Precision medicine and big data challenges for visualization
Current context

• Background
  - Health care, biomedical research and public health are information intensive activities:
    - Medical images and clinical records
    - DNA sequencing, molecular data
    - Literature and public databases
    - Clinical trials, biobanks, GWAS,...
  - New data types (extremely complex and heterogeneous) are being generated at an unprecedented pace

Human Genome Project
  - 10 years
  - $3 Billion

• Current MPS
  - $1000 - $10,000
  - 2 days – 12 weeks
• PubMed
  – 5000 biomedical research articles are published daily
  – Over 22 million articles
  – 6.8 million daily searches
• Sequence Databases
  – DNA...
Current context

• Main issues
  - How can we efficiently collect, store, search, integrate, analyse and visualise all of this information?
  - How can we facilitate the translation of research findings into clinical solutions?
  - Which new information processing methods will be needed to respond to the emerging research approaches?
  - The tools you use are the result of our research

Biomedical Informatics

• Informatics is the science of information
• Information is data plus meaning
• Biomedical informatics is the science of information as applied to or studied in the context of biomedicine.
• Informaticians study information (data + meaning, in contrast to focusing exclusively on data).
• Thus, practitioners must understand the context or domain (biomedicine).

Biomedical informatics (BMI) is the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, motivated by efforts to improve human health.
Different visualization techniques and methods that can be applied

A PERIODIC TABLE OF VISUALIZATION METHODS

From a theoretical perspective, biomedicine is interested in visualising Data, Information, Knowledge and Wisdom (DIKW).

DIKW some (arguable) definitions

- Data: symbols or facts
- Information: contextualised or processed data to be useful
- Knowledge: application of data and information
- Wisdom: implies understanding and using the knowledge adding value to it
In biomedical research visualization and representation of all these elements is required and needed becoming more complex as we move upper in the pyramid.

- Example: How can we convey uncertainty?
  - Relatively easy with data
  - More difficult when we move into the realm of information and knowledge.
• An example of how to visualize those elements in a biomedical domain context
  – Gene expression
    • data - Can be represented using heatmaps
    • Information – an example would be the trees generated during a clustering process where we link the genes to the conditions and we have processed the (jointly used with the heatmap
    • Knowledge – mapping the results from the experiments in pathways, networks or ontologies
    • Wisdom – Application of the knowledge / visualize it in workflows used in clinical decision support tools

• An example of how to visualize those elements in a biomedical domain context
  – Gene expression
    • data - Can be represented using box plots
• Information – Heatmaps and trees resulting from clustering analyses

“Host adaptive immunity deficiency in severe pandemic influenza”
• Other important knowledge representation and visualization tools in biomedicine are ontologies.

Anatomical Space
- Organ Cavity Subdivision
- Serous Sac Cavity Subdivision

Organ Cavity
- Serous Sac Cavity

Serous Sac
- Organ Component
- Organ Cavity

Organ Part
- Tissue
- Pleural Sac
- Pleural Cavity
- Parietal Pleura
- Visceral Pleura
- Mediastinal Pleura
- Mesothelium of Pleura
- Interlobar recess

Gene Ontology
- cellular_component
  - pigment metabolic process during pigmentation
  - pigment metabolic process during developmental pigmentation
- biological_process
  - pigmentation
  - regulation of pigmentation during development
  - negative regulation of pigmentation during development
  - negative regulation of cuticle pigmentation
  - negative regulation of eye pigmentation
- molecular_function
  - regulation of biological process
  - positive regulation of biological process
  - eye pigment precursor transport
  - positive regulation of cuticle pigmentation
  - positive regulation of eye pigmentation
• Wisdom visualization
  – It remains still a challenge to capture it and visualise it
  – How could we visualise aspects such as “How do I drive?”
• Genomes contain multiple layers of information
  – How to provide on a linear representation all that information?
    • Sequence, variants, functional annotation (genes, regulatory elements, splicing variants), proteins, similarity with other model organisms…
  – Genome browsers provide different “tracks” each of them providing one level of Information

• Genome Browsers
  • Systems to navigate and visualize genomes and their annotations
  • Complex views using great amounts of information
  • Integrates phenotypical (clinical) and molecular (DNA sequence and variants, gene expression, protein sequence…) information
    – Disease names can be used in queries
  • Allows customized views of different tracks

http://www.ensembl.org/
http://genome.ucsc.edu/
• Ensembl (http://ensembl.org)

http://genome.ucsc.edu/cgi-bin/hgGateway
• Other “omics” techniques
  – Gene expression – most common ways to present information already presented
  – Pathways and networks. Understanding systems by representing the relationships among the individual elements
    – Pathways (metabolic, signaling…)
    – Regulatory networks/genetic networks
    – Protein protein interactions
    – Others…
How data are presented in the molecular world

• Interaction networks - Cytoscape environment and its plug-ins - others


• Integrating multiple levels of "-omics" data and information
  – Circos, OmicCircos and others
  – Uses circles, creating different layers of circles representing different datasets (each one used for one kind of omic data)
  – Facilitates the representation of links or connections between elements
  – Creates infographics and its commonly used to present omics data
How data are presented in the molecular world

- Complex multi-omics visualizations


The integrative Personal Omics Profile (iPOP)
Follow up for 14 months
Multi-omics plus other clinical/physiological data

Presenting molecular data for clinical use

- Development and application of genomic information for clinical purposes:
  - How can we present information to clinicians
  - Concept of actionable information

Source: http://www.personalis.com/ac_e_platform/analysis.php

PartnersHealthcare. Personalized medicine.
Presenting molecular data for clinical use

• Presenting molecular information for molecular biology boards


Integrating different information sources

• It is not simply visualizing the data but also facilitating access to the available information and presenting unified views of disparate data sources
  – How much do we know about a certain molecule/gene/protein
Precision medicine and big data challenges for visualization

What is precision medicine?

• Refers to the use of molecular and other disruptive technologies to improve diagnosis identifying the causes of the diseases rather than simply using their symptoms
• It is INFORMATION driven
• New technologies implies broader variety of data sources
  – Including the exposome, patient collected data and continuous data
• Patient centric
• Multilayered approaches for new biomedical information systems
Connecting different levels of biomedical information

Multiple data sources at different levels

From atom to environment…

… with a patient/citizen centric and longitudinal approach

Role of informatics - Measuring the exposome

Environment-Wide Association Study on Type 2 Diabetes Mellitus

266 environmental Factors

Future: combined GWAS-EWAS?

(Patel et al. 2010 PloS One)
Precision medicine and big data challenges for visualization - New taxonomy of diseases

Stratification of disease – ICD 11 – US Nat Academy – **Towards Precision Medicine**  
**New taxonomy based on human molecular biology**

- Is there an intuitive way to visualize and integrate the elements of the equation: Phenome = Genome * Exposome?

- Could we visualise these interactions to close existing gaps in the use of information?

- Could we represent all the collected information at certain data point in an intuitive manner?

skin, colon, parathyroid – BRAF Mutation  
MD Anderson CC – Breast, Ovarian, Uterine, Cervical – PIK3CA Mutation trial
Proposal for a new model for biomedical data representation INDIV-3D

- Inspired in the geographical information systems (GIS) and the information commons suggested by the US Nat. Academy
- Theoretical approach:
  - It is possible to define a **three dimensional model** to represent and store biomedical data.
  - It enables the representation of virtually **any kind of biomedical data**
  - Such model provides a **longitudinal** perspective of the information levels
  - **Common coordinates for comparison** among individuals
Model characteristics

• This model provides insight about the volume of information captured for each individual during his life as well as at defined time points
• Use of ontologies will increase the connectivity among levels and provide ground for semantic reasoning
• It can store continuous data (such as those captured by new mobile technologies and sensors)
• Provides a flexible framework for adding new data types
• Data can be captured by time, information level or individual data point

Conclusions

• Biomedicine requires visualization tools for data, information knowledge and wisdom.
• There are tools and ways to visualise data, information and knowledge although wisdom visualisation remains elusive.
• “Omics” approaches generally imply integration of several layers of complexity and therefore require multi-layered visualisation approaches
• Presenting “omics” data for clinicians requires a reduction and simplification to actionable items.
• Precision medicine and new data sources represent new visualization challenges.
Thank you for your attention!

guillermo.lopez@unimelb.edu.au

Ex. Comparing patients

Populational representation

Data retrieval & comparison

Individual models