Integrated management and analysis of colorectal cancer research data sets

18 April 2013

Guillermo Lopez-Campos
Health and Biomedical Informatics Research (HBIR) Unit
Melbourne Medical School
Faculty of Medicine, Dentistry & Health Sciences
• Introduction
• Integrating phenotypic data
• Analysis of integrated clinical and molecular data
  – mRNA expression analysis
  – microRNA expression analysis
• Conclusion
Colorectal cancer is one of the most common tumors in western countries.

There are different classifications based on the pathological characteristics such as the invasion of intestine wall, number of affected nodes or metastasis.

Molecular data are increasing their importance for the understanding and clinical management of cancer data and gene expression is a key element in cancer.

- New high-throughput techniques are generating overwhelming amounts of data.

Data integration and analysis represent key challenges to understand the relationships between genotype and phenotype.
Integrating phenotypic data

• An overwhelming amount of data from heterogeneous data sources (>100 patients)...
  – Clinical data
    • Extracted from patient records (manually)
    • Tumour bank...
  – Molecular data:
    • Different levels: Gene, MicroRNA, Messenger RNA, Protein...
  – Different technologies
    • Microarrays, Tissue microarrays, Genotyping...
  – Different approaches
    • Systems biology and pathway analysis
  – Other Public data sources
    • From databases (GEO, ArrayExpress)...
• The goal of the study is to integrate and analyze these data to generate useful clinical information.
Integrating phenotypic Data

• Development of a “in-house” system for clinical annotation of samples used in the project.
  – Used different clinical standards
  – LIMS (Laboratory Information Management System) capabilities
  – Integrates different kinds of molecular data
    • Genotype for certain relevant genes
    • Chromosomal instability
    • Microarray data (including sample processing associated data i.e. labelling processes) in a MIAME (microarray standard) compliant format.
• mRNA microarray analysis
  – Integrated molecular data from different sources (gene expression microarrays, PCR, tissue microarrays)
  – Microarrays were analysed using unsupervised clustering techniques
  – Groups identified were validated using external data gathered from public repositories
    • Need to integrate data from different microarray platforms – Different annotation, different systematic biases
  – Analysis of relevant pathways
Analysis of integrated data - mRNA

- mRNA microarray analysis
  - High volumes of data generated in these analyses (Hi-res images from tissue microarrays) combined with the microarray data
    - These analyses included the measurement and integration of data at the protein level
    - Validated some of the results found with the microarrays
  - Integration with clinical phenotypic data
    - Different groups with different prognosis based on the molecular characteristics of the samples
Analysis of integrated data - microRNA

- Microarray analysis of microRNAs
  - MicroRNAs regulate gene expression
    - Complex interactions microRNA-mRNA
  - Integrative approach to combine previous data
    - Analysis of microRNA expression data on the same samples analysed for mRNA
    - Combination of databases and algorithms to identify microRNA and putative targets
  - Assessment and development of methods for microRNA data normalization
• Integrative analyses of different clinical and “-omics” data types allowed for new method to classify colorectal cancer with different molecular and clinical behavior based on their gene expression signatures
Conclusion

• Comprehensive analysis of colorectal cancer is using high-throughput molecular techniques

• New techniques and high-throughput methods are creating unprecedented amounts of data that need to be managed, integrated and transformed into knowledge

• Integration of clinical and molecular information is key for the success of these approaches
  – Leading to a better understanding of the mechanisms underlying diseases and might be use for prognosis and therapy
Aknowledgements

• “San Carlos” Clinical Hospital – Molecular Oncology Laboratory
  – Beatriz Perez-Villamil
  – Alejandro Romera-Lopez
  – Eduardo Diaz-Rubio

• The University of Melbourne - HBIR
  – Fernando Martin-Sanchez
Thank you for your attention!