Molecular Genetic Testing and the Emergence of Computational Pathology

Mark S. Boguski, M.D., Ph.D., F.C.A.P.
Center for Biomedical Informatics, Harvard Medical School
Department of Pathology, BIDMC
May 3, 2012
Conflict of Interest Disclosure

medpage TODAY

genome HEALTH SOLUTIONS
<table>
<thead>
<tr>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is genomics?</strong></td>
</tr>
<tr>
<td>A branch of biotechnology that determines the biological functions and medical significance of all of the approximately 22,000 pairs of genes that we inherited from our parents.</td>
</tr>
<tr>
<td><strong>What is a human genome?</strong></td>
</tr>
<tr>
<td>• A “whole genome” consists of ~3 gigabytes (3 billion “base pairs”) of DNA distributed unequally among 46 chromosomes (diploid genome).</td>
</tr>
<tr>
<td>• Approximately <strong>98%</strong> of this DNA is “intergenic” (literally “between genes”), does not encode proteins and is of unknown medical relevance.</td>
</tr>
<tr>
<td><strong>What is a human exome?</strong></td>
</tr>
<tr>
<td>• Exome refers to the 2% subset of the whole genome that encompasses our ~22,000 pairs of genes.</td>
</tr>
<tr>
<td>• Because each gene, on average, is composed of 8 protein-encoding segments (“exons”), an exome corresponds to 8 x 22,000 ~ 176,000 segments of DNA.</td>
</tr>
<tr>
<td><strong>What is a human transcriptome (a.k.a. gene expression profile)?</strong></td>
</tr>
<tr>
<td>• Transcriptome refers collectively to all of the “expressed” RNA “transcripts” of genes based on the “central dogma” of molecular biology, i.e. DNA -&gt; RNA -&gt; protein.</td>
</tr>
<tr>
<td>• A transcriptome reflects what a cell is doing at a particular point in time (molecular phenotype) as opposed to what it is capable of doing (genotype).</td>
</tr>
</tbody>
</table>
Two Basic Types of Clinical Genomics

**Presymptomatic (prospective)**
Genotyping for disease risk assessment

- Newborn screening
- Carrier testing
- GWAS

**Post-diagnostic**
Genotyping for targeted therapy, Treatment optimization

- Tumor “subtyping”
Workforces for Personalized Genomic Medicine

Current Practitioners

Medical Geneticists

Genetic Counselors

Pathologists

Thousands
National Agenda

The Call to Action June 2010

Banbury Summit on Genome-Era Pathology October 13-15, 2010

The Future of Pathology in Personalized Medicine Boston, May 23-24, 2011

Location: Harvard Medical School, Countway Library of Medicine, 10 Shattuck St., Boston, MA 02115 Directions & Maps

Participants: Summit Participants
Training Residents in Genomics (TRIG) Working Group
Case History

• 78 y/o male, no prior H/O Ca, presented with throat discomfort
  – Biopsy revealed papillary adenocarcinoma
• Laser resection and lymph node dissection, 3/21 nodes positive
  – 60 Gy adjuvant radiation therapy administered
• 4 months later, PET-CT revealed numerous small bilateral pulmonary mets
  – No Tx protocol (rare tumor); immunohistochemistry indicated 2+ EGFR
• Erlotinib (Tarceva®) started
  – Lack of response and tumor progression
• Diagnostic Whole Genome and Transcriptome Analyses
Analyses of Tumor DNA for Targeted Therapy

• Genome sequencing and analysis:
  – Comparing tumor genome sequence to peripheral blood lymphocyte (normal, somatic, germline) genome
  – Mutation detection
• Transcriptome analysis:
  – Digital gene expression profiling of tumor
• Search of DrugBank.ca
  – Relate mutations and gene expression data to drugs with known therapeutic targets and mechanisms of action
Analyses of Tumor DNA for Targeted Therapy

• Genome sequencing and analysis:
  – Comparing tumor genome sequence to peripheral blood lymphocyte (normal, somatic) genome
  – Mutation detection

• Transcriptome analysis:
  – Digital gene expression profiling of tumor

• Search of DrugBank.ca
  – Relate mutations and gene expression data to drugs with known therapeutic targets and mechanisms of action

“Whole Genome Analysis” requires a variable data package of genome and/or transcriptome depending upon clinical indication and diagnostic goals
Pathology Report

Patient was switched to Sunitinib (Sutent®)
Clinical Course

In cancer, whole genome analysis will be done **not once**, but **multiple times** during the course of the disease for **tumor subtyping**, **monitoring response to therapy** and **diagnosing the reasons for recurrences or therapeutic failures**.
Conclusion:
“...whole genome characterization will become a routine part of cancer pathology.”
A baffling illness
Desperate for clues to a 4-year-old's gut-destroying disease, doctors wonder whether a pioneering DNA technique could help.

By: Mark Johnson and Kathleen Gallagher, photos by Gary Porter
December 19, 2010
Pathologist’s New “Microscope”
Data Production

Data Annotation

Cloud Computing Services

Data Interpretation

“Send outs”

Report generation
New Paradigm for the Laboratory Diagnosis of Cancer

Clinical Genome Analysis means “Big Data”

One patient’s genome
NGS data per patient
1 medium-sized hospital
IBM Watson memory?
Every person on earth

See also: Multiples of bits · Orders of magnitude of data

Wikipedia, 2011
Biomedical Cloud Computing With Amazon Web Services

A. Prototyping

Amazon Web Services
- 7 GB memory
- 1.7 TB disk space
- $0.68/CPU hr.

Work flow overview
- Truncated test set of NGS reads.
  - 2 files with 10,000 reads per file.
  - [3 GB ref. genome + 2.2 MB read files]
- Align reads and determine SNP calls using MAQ.
  - [5 hours]
- Final alignment output file
  - [1 MB]

Cost: $3.85

B. Developing Scalable Application

Amazon Web Services
- Cluster management software
- Development and testing.

Work flow overview
- Test set of NGS reads.
  - 32 files with 1 million reads per file.
  - [3.34 GB read files]
- Align reads and determine SNP calls using MAQ.
  - [12 hours]
- Final alignment output file
  - [1.3 GB]

Cost: $49.60

C. Scaled Application

Amazon Web Services

Work flow overview
- Whole genome set of NGS reads.
  - 606 files with ~7 million reads per file.
  - [370 GB read files]
- Align reads and determine SNP calls using MAQ.
  - [10 hours * 38 instances]
- Final alignment output file
  - [142 GB]

Cost: $320.10

Use your local computer to connect to instances using secure shell (ssh) and transfer data using secure copy (scp).

V. Fusaro, PLoS Comp Bio, 2011
From WGA to Clinical Action

- Paired-end sequencing RNA-seq
- Validation (FISH, RT-PCR, Sanger)
- Validation Gene Expression
- Treatement Plan Prepared
- Database
  - Oncogene/Tumor Suppressor Detection
    - Amplified CNV
    - Over-expressed
    - Deleterious, LOH
  - Variants Identified
  - FDA Approved On-Label
  - FDA Approved Off-Label
  - Clinical Trial Underway
- Medical Impact Report Generated

D Wall, CBMI, HMS
Pathologists and the third wave of medical genomics

1\textsuperscript{st} wave: The Human Genome will yield many new drugs
2\textsuperscript{nd} wave: Genome-wide association data will help us manage complex diseases
3\textsuperscript{rd} wave: Precision diagnostics will result in better outcomes
Delivering Personalized Genomic Medicine as the New Standard of Care

WHY
Patients deserve faster and more complete access to new medical technologies that enable personalized medicine.

- Patients are more knowledgeable and empowered than ever and demand access to information that could improve their outcomes.
- Healthcare providers and delivery systems are overwhelmed by the rapid emergence and complexity of new diagnostic modalities and the "big data" challenges they engender.
- Payers, benefits managers, and pharmaceutical R&D organizations new approaches to assess the comparative effectiveness of technologies.
- Traditional processes and core channels impede the diffusion of innovation in healthcare.

HOW
We are developing new multi-sided platforms and networks for both traditional and new stakeholder groups to:

- Accelerate the diffusion of personalized genomic medicine in cancer and other diseases.

WHAT
- Genomic Medicine Readiness™ Assessments
- Comparative effectiveness research
- Individual case management and/or referrals
- Patient/physician education, marketing and communication

Service Lines