Personalized Medicine and IT

Data-driven Medicine in the Age of Genomics

www.intel.com/healthcare/bigdata

Ketan Paranjape
General Manager, Life Sciences
Intel Corp.
@Portlandketan
The Central Dogma of Biology

Central Dogma of Biology:

DNA       RNA       Protein

Genomic DNA is blueprint set of instructions

Messenger RNAs (mRNAs) are the specific, short-lived, gene transcripts

Proteins perform structural and catalytic functions

transcription a.k.a. “gene expression”

Translation occurs in ribosomes: (1) mRNA attaches to ribosome, (2) polypeptides are produced, polypeptides are folded in to proteins

Typical gene

Direct protein assembly

Protein

DNA

Messenger RNA

MicroRNA gene

Binds to messenger RNA

Protein assembly is blocked

MicroRNA

The Violinist's Thumb

Health & Life Sciences at Intel
Where information and care meet
Acknowledgments

*Other names and brands may be claimed as the property of others.

It Takes a Village ...

Health & Life Sciences at Intel
where information and care meet
Big Data Analytics in Health and Life Sciences

Today: Many disparate data types, streams...

- Meds & labs
- Claims & transactions
- Personal data
- Clinical
- Genomics
- Patient experience

Future: Integrated computing and data

Leading to better decisions
- Improved patient experience
- Healthier population outcomes
- Reduced costs
We are at an Inflection Point in Healthcare - TRENDS

Healthcare costs are **RISING**
Significant % of GDP

**Global AGING**
Average Age 60+:
growing from 10% to 21% by 2050

**US Healthcare BIG DATA Value**
$300 Billion in value/year
~ 0.7% annual productivity growth
We are at an Inflection Point in Healthcare - TRENDS

Storage Growth
Total Data Healthcare Providers (PB)

- Admin
- Imaging
- EMR
- Email
- File
- Non Clin Img
- Research

Data Explosion projected to reach 35 Zetabytes by 2020, with a 44-fold increase from 2009.

Source: McKinsey Global Institute Analysis
We are at an Inflection Point in Healthcare - TRENDS

313 Exabytes
if everyone in the US
has their genes sequenced

495 Exabytes
if every cancer patient in the US has their genes sequenced every 2 weeks

Cost of sequencing is rapidly falling...
Genomics is a Big Data Problem

The Broad Institute will produce more data than Microsoft, Facebook and Amazon combined by 2015 ....

*Other names and brands may be claimed as the property of others.
Health & Life Sciences at Intel

Where information and care meet

We are barely scratching the surface ...

Life Sciences World Map

Ontario: $1.28B annual LS investment, $148B annual LS revenue

Geisinger Health: 100,000 patients

New York Genome Center: $200M investments

Broad Institute

Mt Sinai

KACST Genomics: 100,000 patients

Genomics England: 100,000 patients, £1B UK annual LS investment

Genomics France: €6B

Beijing Genomics Institute: $1.5BB line of credit, $11B Japan investments

Genomics Qatar: 400,000 patients

KACST Genomics: 100,000 patients

NY Genome Center: $200M investments

Genomics England: 100,000 patients, £1B UK annual LS investment

Genomics Qatar: 400,000 patients

Broad Institute

Wellcome Trust Sanger

Geisinger Health: 100,000 patients

New York Genome Center: $200M investments

Broad Institute

Mt Sinai
Barriers

- Clinical
  - Lab, Clinical, CLIA, Direct to Test (DTC), Clinical efficacy
- Literacy and societal challenges
  - Education, Clinical pathway/guidelines, Transparency, Trust, Accountability
- Technical
  - Data Generation, Management, Interpretation, Storage
- Ethical
  - Clinical research, Privacy, Discrimination (GINA)
- Economic and Commercial
  - Insurance, Reimbursement, Preventative Shift, Ownership Sustainability
Charite “Real-time” Cancer Analysis – Matching proper therapies to patients using in-memory techniques

- **Challenge**: Real-time analysis of cancer patients using in-memory SAP HANA Oncolyzer database running on Intel® Xeon® family infrastructure. (3.5M Data points per Patient, Up to 20 TB of data/patient)

- **Solution**: Using structured and unstructured data to collect and analyze tables used to take up to two days -- now takes seconds

- **Benefits**: Improves medical quality in disruptive way for Patient, Doctor, Hospital, Research

http://moss.ger.1th.intel.com/sites/SAP/SAP%20account%20team%20documents/Marketing/SAP%20HANA/SAP_HANA_Chatie_Case_Study_TH.pdf
**Challenge:** Experiment processing takes 7 days with current infrastructure. Delays treatment for sick patients

**Solution:** Dell Next Generation Sequencing Appliance
- Single Rack Solution; 9 Teraflops, Lustre File Storage; Intel SW tools

**Benefits:** RNA-Seq processing reduced to 4 hour
- Includes everything you need for NGS - compute, storage, software, networking, infrastructure, installation, deployment, training, service & support
High Throughput Science: Embracing **Cloud-based Analytics** for Computational Chemistry Simulation

- **Challenge:** Sustaining 50000+ compute cores for large scale simulations, for less than a week; CapEX v. OpX

- **Solution:** Novartis leveraged software from AWS partner, Cycle Computing, and MolSoft to provision a fully secured cluster of 30,000 CPUs, powered by the Intel® Xeon® processor E5 family.
  
  - Completed screening of **3.2 million compounds** in approximately **9 hrs**, compared to **4 - 14 days** on existing resources.
Regional Health Information Network
RHIN – China (Jinzhou, Pop 3M)

**Challenge:** RHIN has challenges with scalability, performance and maintenance. Data storage is expensive

**Solution:** EMR data and healthcare services running on Apache Hadoop* and Intel® Xeon® E5 servers

**Benefits:** High performance and scalability demonstrated via POC and stress testing. Significantly reduced storage cost

1/5 Reduction in Response Time; 5x Concurrent Users

*Some names and brands may be claimed as the property of others.
Training Programs

Bioinformatics, Life Sciences, Computer Sciences, Clinicians
Develop an ICT-enabled European Strategy for Personalised Medicine

2014-2020

Driving research to unleash the potential of ICT at the point-of-care

EU R&D initiatives must address:

- Interoperability of technical standards for managing and sharing sequence data in research and clinical samples;
- Development of hardware, software and workflow algorithms to accelerate cost efficient analysis of genetic abnormalities that cause cancer and other complex diseases;
- Research to ensure convergence of Big Data and Cloud Computing infrastructure to meet the requirements of High Performance Computing and data throughout the life sciences and healthcare value chains

The eHealth Action Plan 2020 should include Personalised Medicine as a priority

- Gain knowledge of the challenges and barriers (technical, organizational, legal and political) to the adoption of ICT in support of Personalised Medicine leveraged by genomic information;
- Evaluate how to change workflows and education requirements to facilitate adoption of ICT mediated personalized medicine in clinical practice;
- Expand collaboration with other regions of the world in matters of common interest, e.g. by leveraging the eHealth MoU with the United States of America;
- Study, evaluate and disseminate technology neutral risk assessment frameworks for data privacy and security, covering the entire ICT enabled Personalised Medicine delivery chain;
- Develop effective methods for enabling the use of medical information for public health and research
Genes causing it identified & disease pathways determined

Precision medicine regime

Personalized Medicine: All in a day by 2050 2020