From Integrative Informatics to Passing the Turing Test in Rare Disease Diagnosis

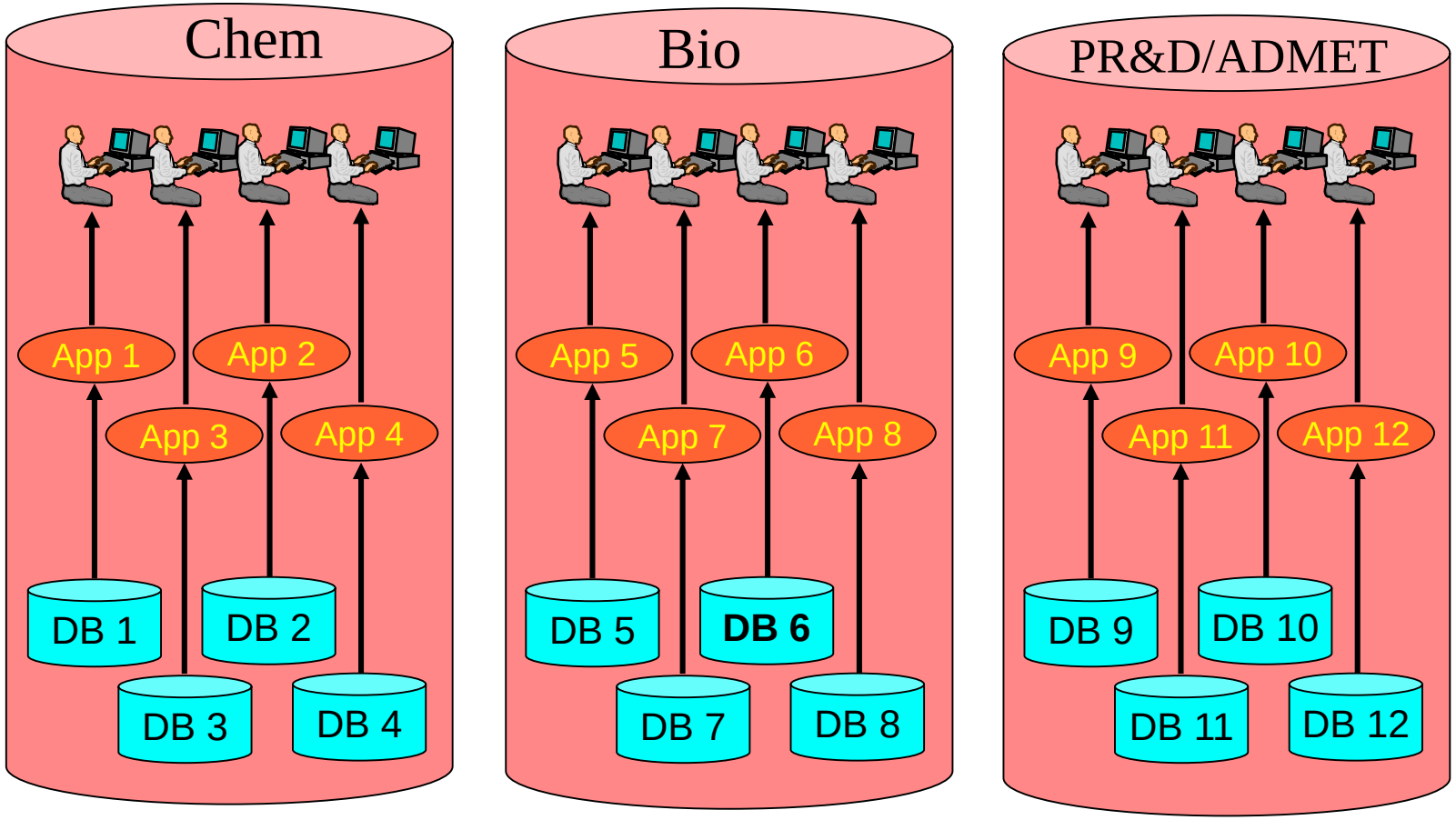
PRISME Forum

John Reynders, VP Data Sciences, Genomics, and Bioinformatics
Alexion Pharmaceuticals

November 21st , 2019

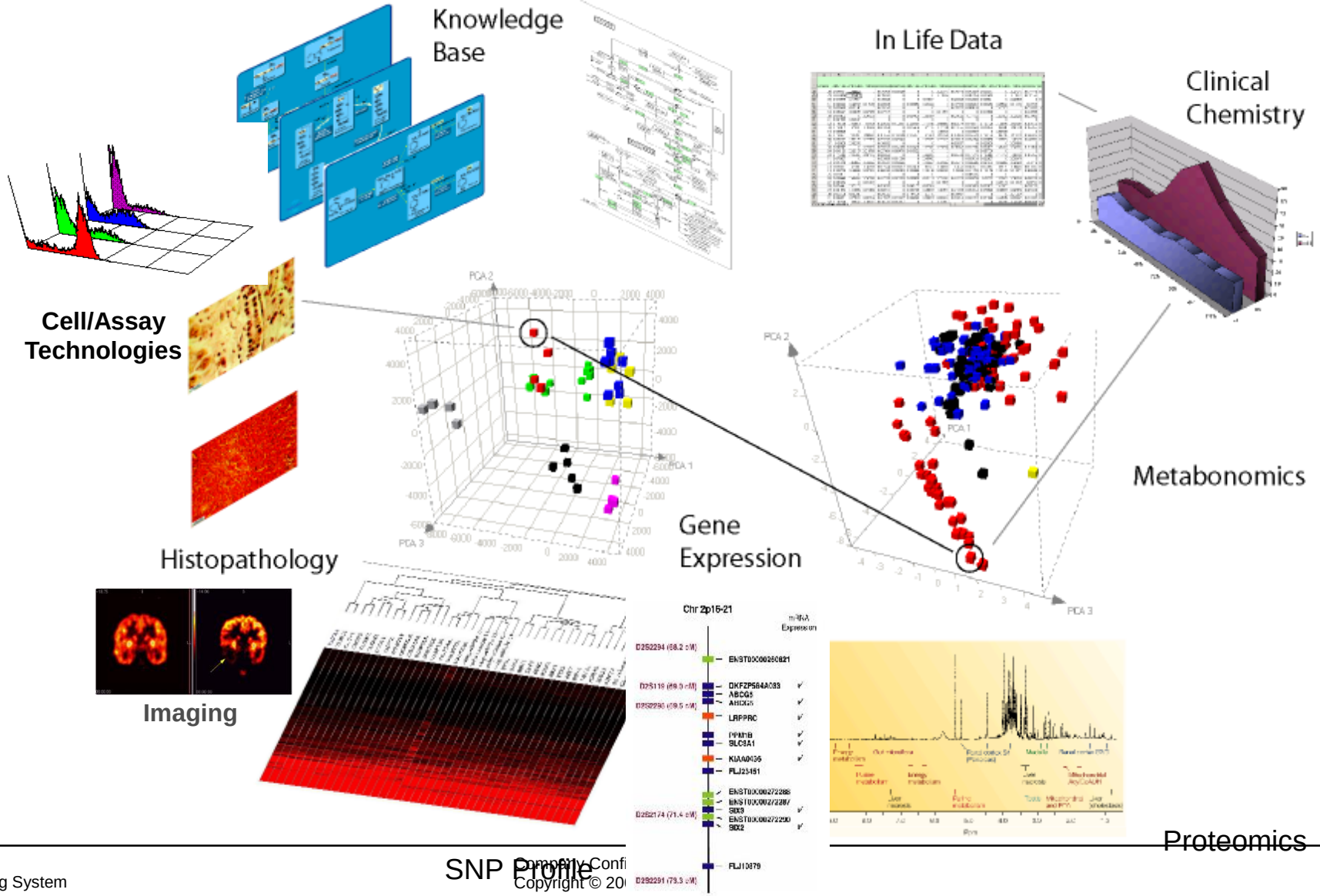
Some slides from an external presentation…

The Problem: Silos of Silos

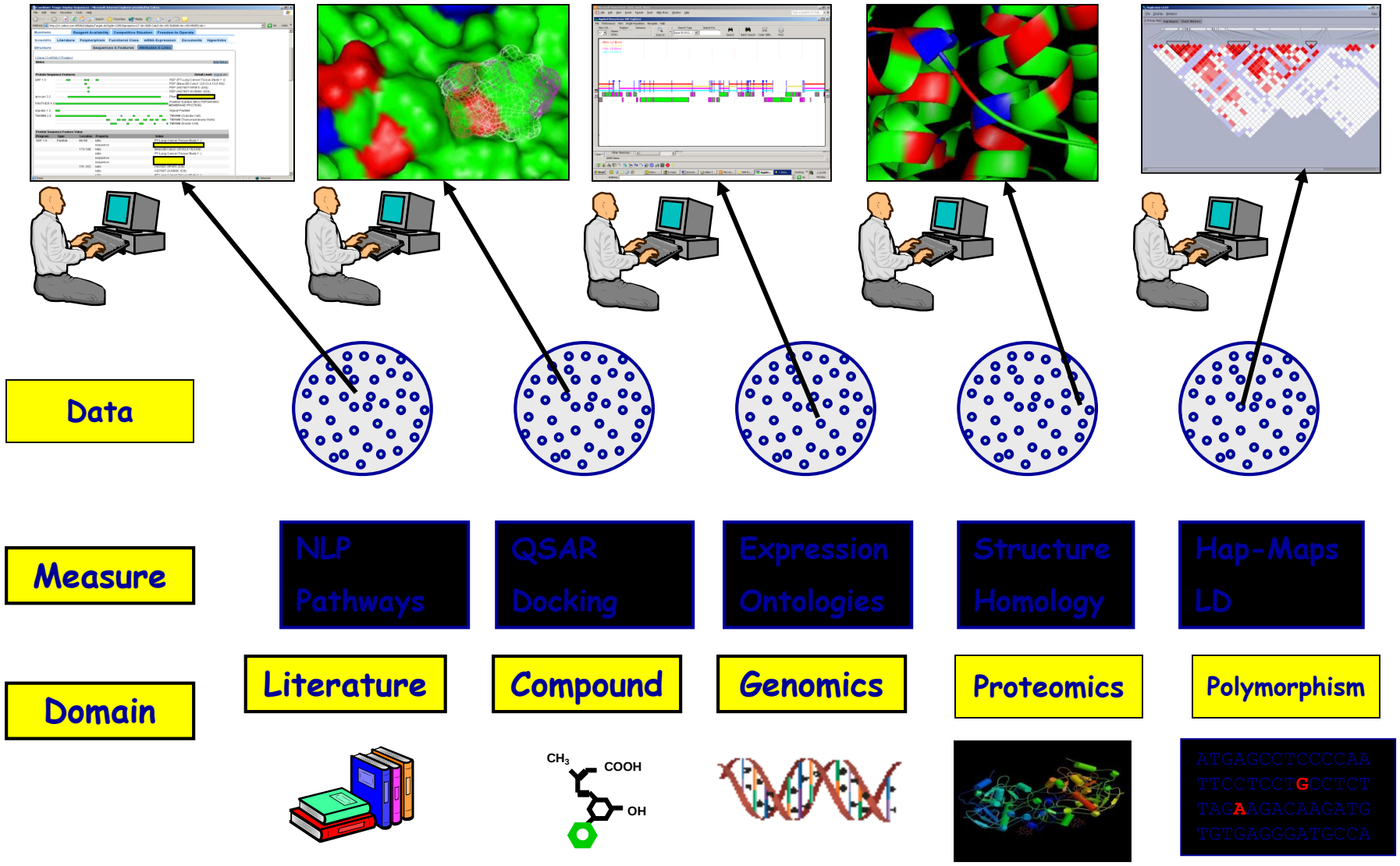


- Tools, application, and data are standalone with limited interaction
- Scientists have great difficulty finding their data and associated tools
- Asking cross-domain questions (e.g. bio+chem) very difficult
- Support becoming very impractical – hundreds of individual tools across silos

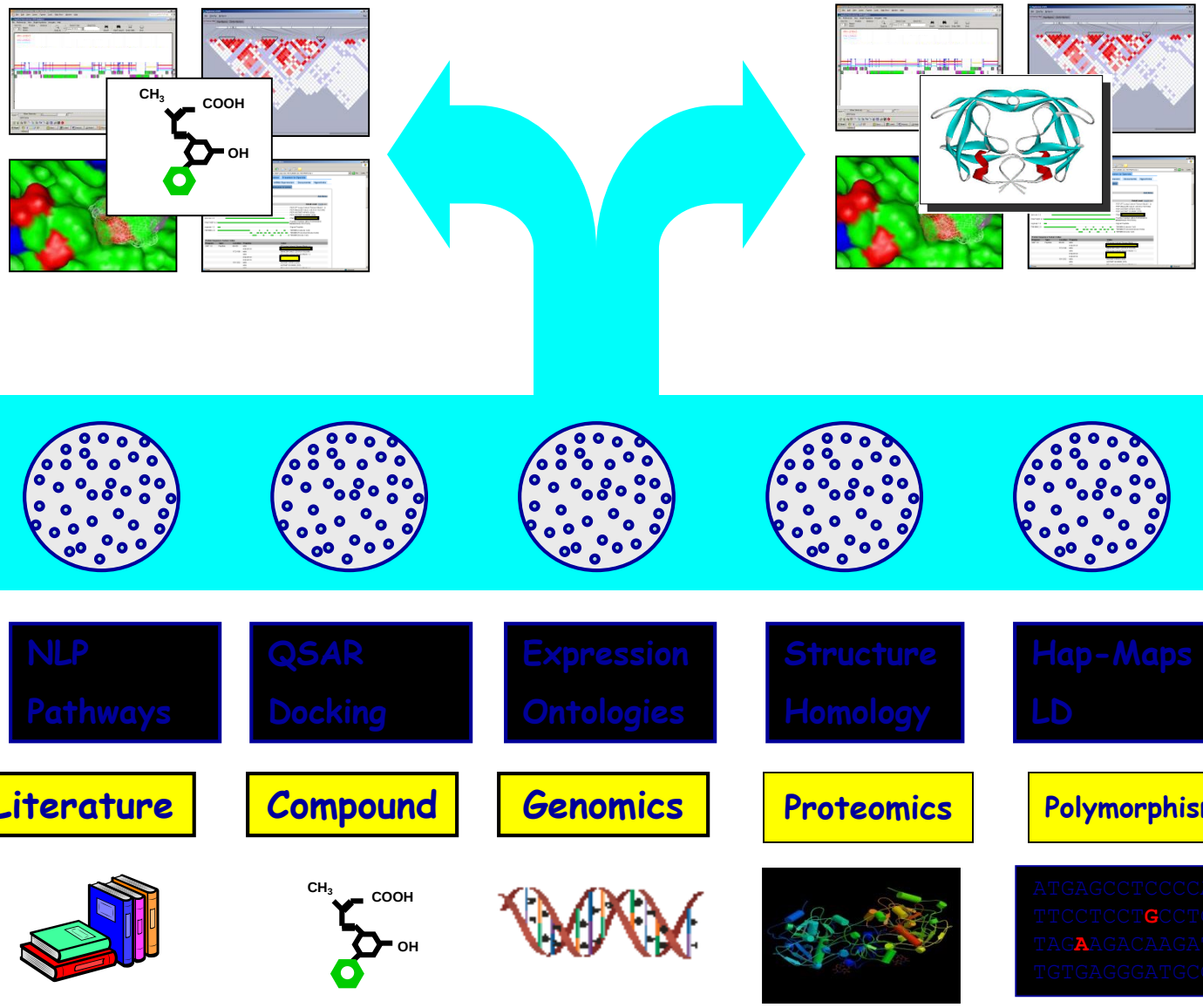
Genomics and Integrative Biology: High Content-High Resolution



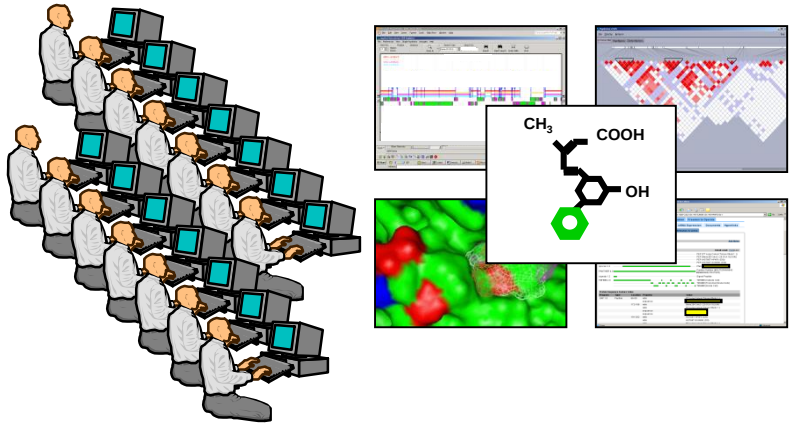
Our Collective Challenge:
We must go beyond domain-specific tools for user-specific tasks…

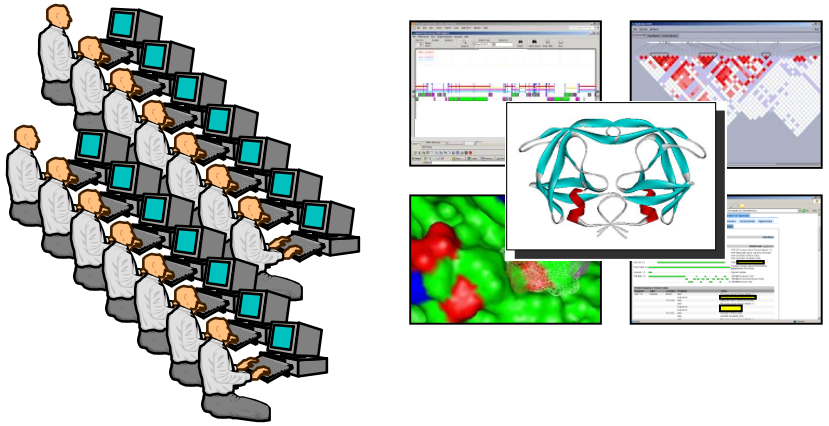


….to establish beachheads which aggregate data and tools into compound-centric and target-centric views which in turn…



…enable a broader community of scientists to improve their productivity through integrated data & tool access while..





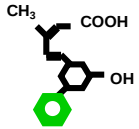
Integration

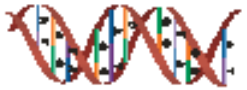
Measure

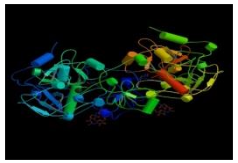
NLP Pathways	QSAR Docking	Expression Ontologies	Structure Homology	Hap-Maps LD

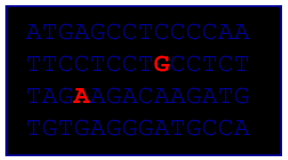
Domain

Literature	Compound	Genomics	Proteomics	Polymorphism

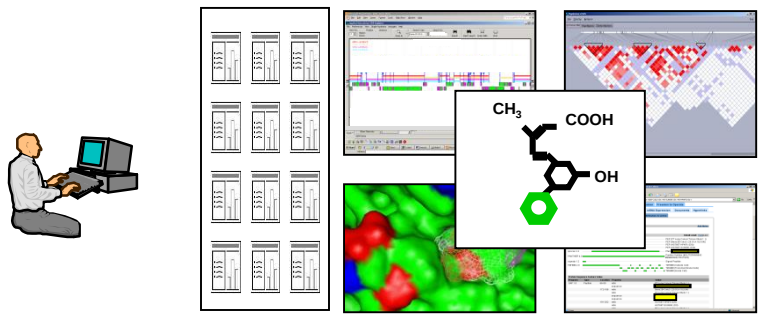


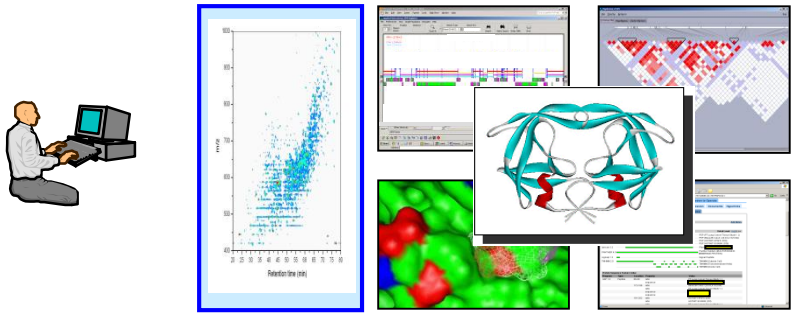


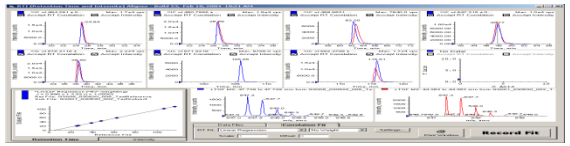


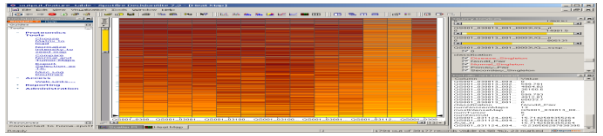


…providing the informatics and computational science tool development community a clear insertion path for new capability.









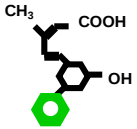
Integration

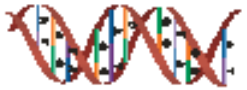
Measure

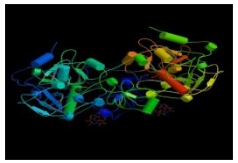
NLP Pathways	QSAR Docking	Expression Ontologies	Structure Homology	Hap-Maps LD

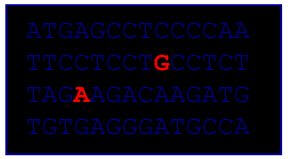
Domain

Literature	Compound	Genomics	Proteomics	Polymorphism









Do these challenges sound familiar?

So, when was this presentation?

Discovery Informatics: Biomarkers and Chemogenomics

Bio-IT World Conference + Expo
May 17th, 2005 – Boston, MA

John Reynders
Information Officer - LRL Discovery and Development Informatics

Lilly
Answers That Matter.

Big

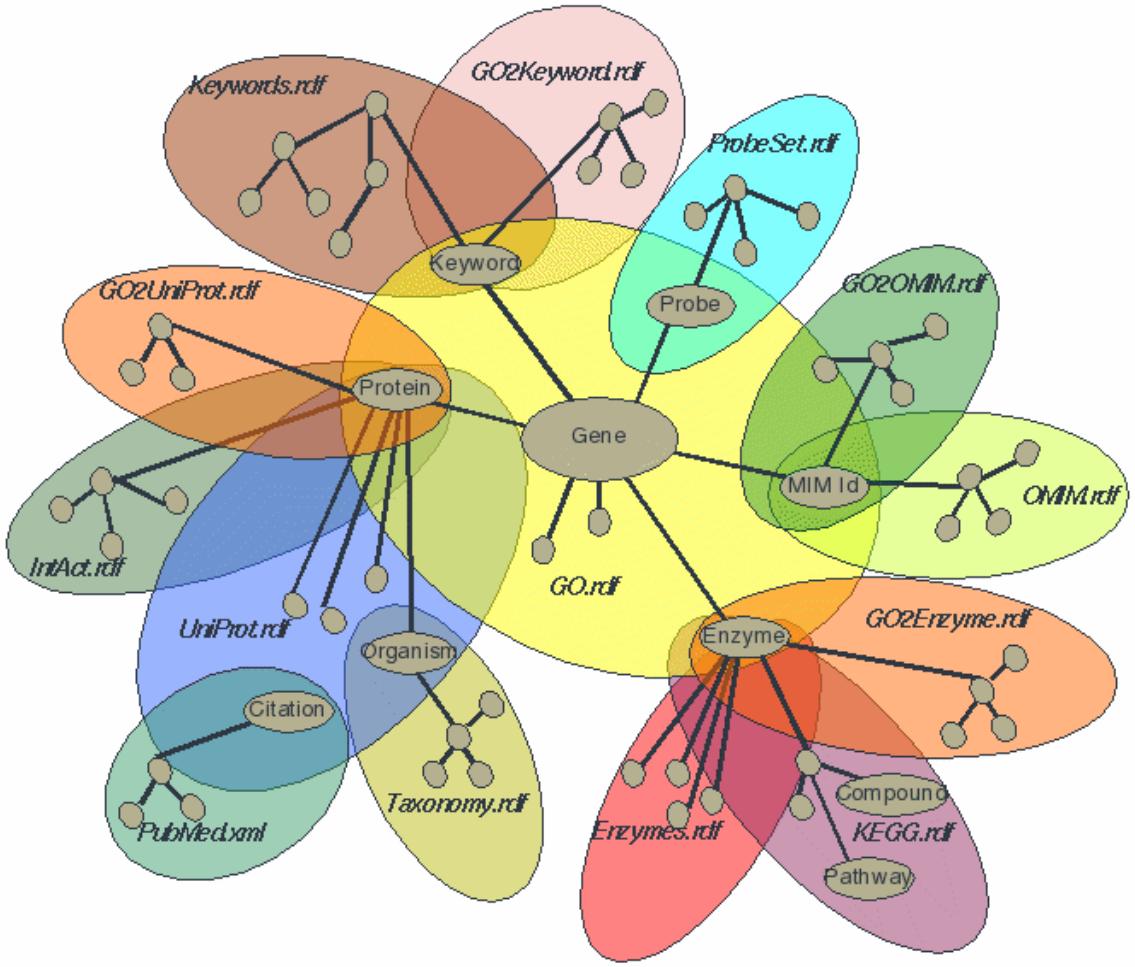
Integrative Data Sciences – Challenge 1

Big is not the issue…

"Big"

- Capture, process, filter, and manage the global and growing avalanche of internal and external scientific and clinical data

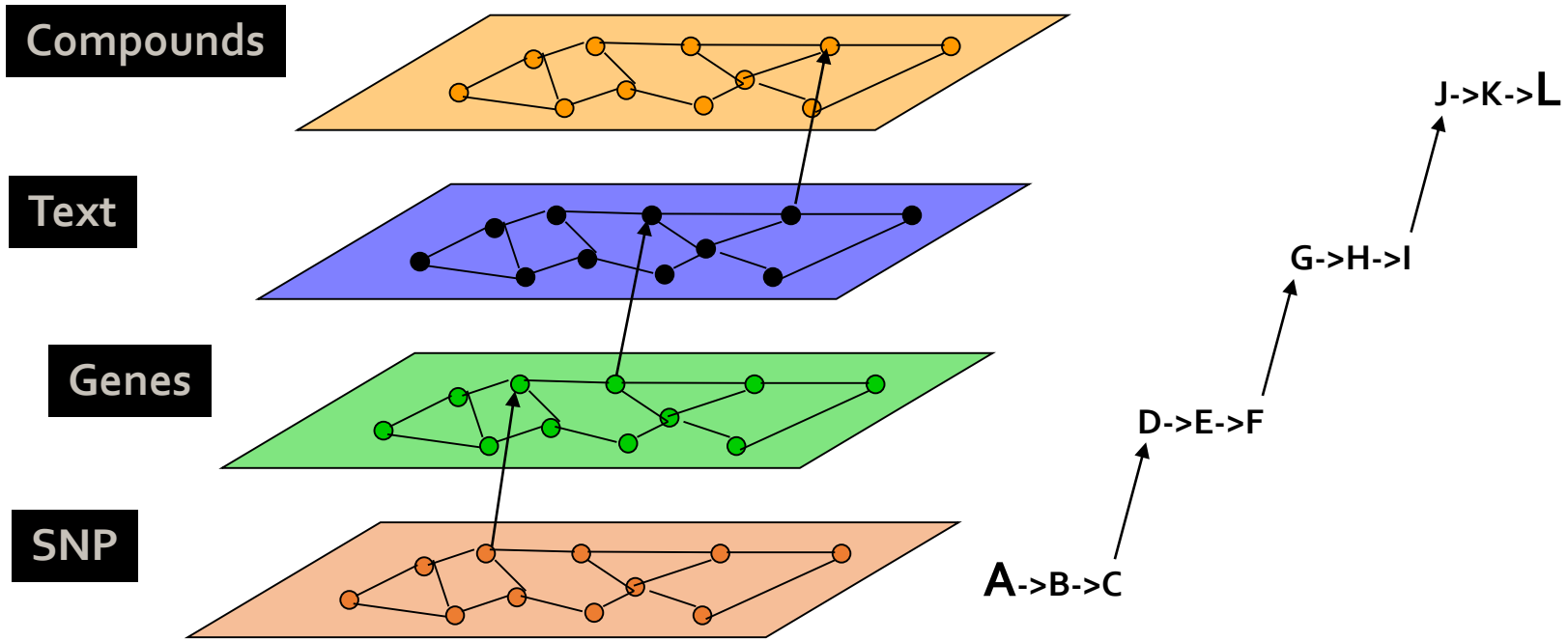
Finding Relationships Across Arbitrary Data through Ontologies



Non-obvious Relationship Analysis

Generate/Test hypotheses across heterogeneous classes of data

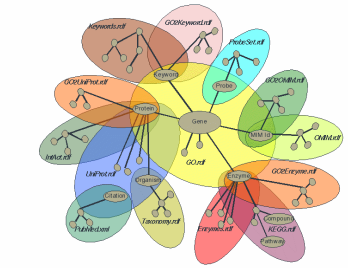
"Connect the dots" between concepts related across multiple classes of Data – but unrelated within any one class of data




Asymmetric Threat
Operational View One (OV-1) - 2013

Integrative Data Sciences – Challenge 2

Information Fusion

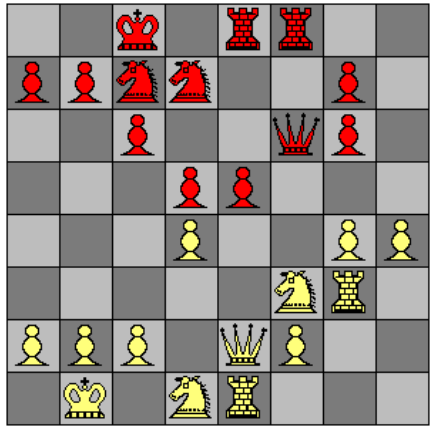


- Semantically integrate and navigate massive, heterogeneous, and distributed data sets

"Big"

- Capture, process, filter, and manage the global and growing avalanche of internal and external scientific and clinical data

Remember this Famous Match?



ALEXION

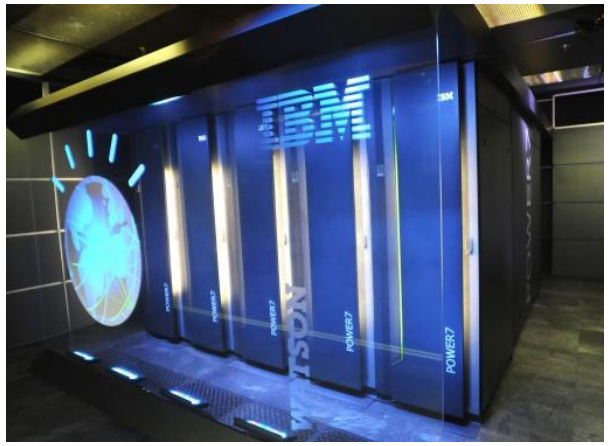
But did you know…

In 1997, Deep Blue was the 259th most powerful supercomputer capable of calculating 11.38 gigaflops
Today, 4 iphones ~ 12 gigaglops

But did you know…

In 1997, Deep Blue was the 259th most powerful supercomputer capable of calculating 11.38 gigaflops
Today, 4 iphones ~ 12 gigaglops

Any guesses on peak speed here?

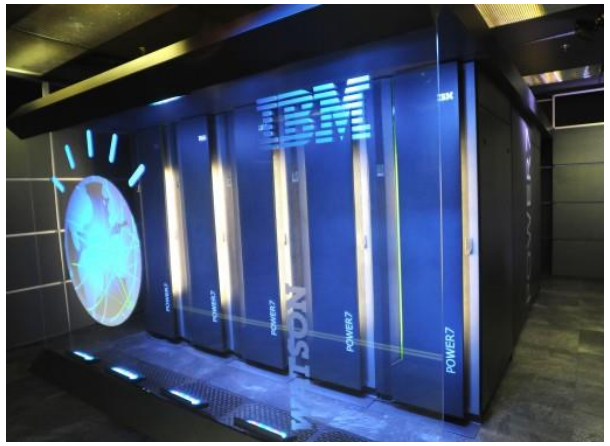


ALEXION

But did you know…

In 1997, Deep Blue was the 259th most powerful supercomputer capable of calculating 11.38 gigaflops
Today, 4 iphones ~ 12 gigaglops

Any guesses on peak speed here?

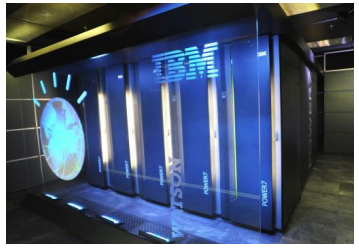


Isn't surprising how little computational power is required to match human-like reasoning in a specific domain?

ALEXION

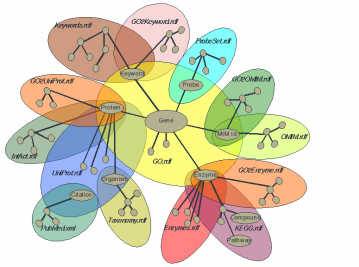
Key Themes in Artificial Intelligence in Rare Disease Diagnosis

Computational Learning



- Computational hypothesis generation, data interpretation, decision support, and acceleration of human insight to enable rare disease diagnosis

Information Fusion



- Semantically integrate and navigate complex, heterogeneous, and distributed data

“Big Data”

- Capture, process, filter, and manage a global and growing avalanche of internal and external scientific and clinical data

ALEXION

The Alexion Insight (AI) Engine

BEST PRACTICES AWARDS ★★★ Bio-IT World 2017

A Data Sciences Rare Disease Map that provides a multi-dimensional overview of the rare-disease landscape

1 Exhaustive Rare Disease Inventory

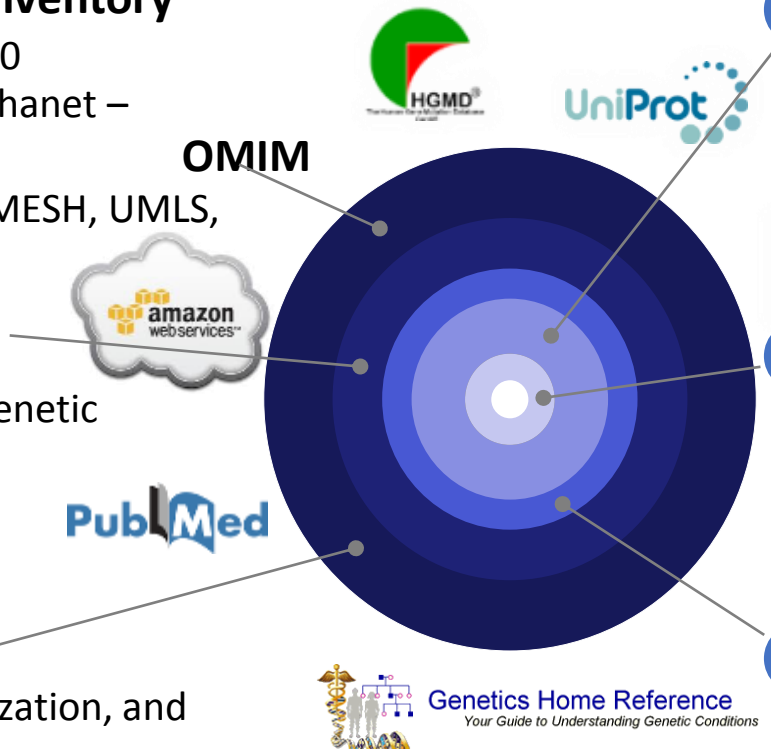
- There are approximately 9,500 rare diseases reported in Orphanet – all captured in our data graph
- Cross-referenced w/ OMIM, MESH, UMLS, MedRA, GHR, ICD-10

2 Incidence/Prevalence

- Data Mining of registry and genetic information
- Expert curation of journals
- Triage of disease rarity

3 Disease Demographics

- Ages of onset, ethnicity localization, and windows of intervention



6 Disease Biology

- Detailed mapping of genetic mutations in each target disease
- ORDO – functional disease grouping
- DisGeNet to source pathway and biomarker information on demand.

5 Intervention Opportunities

- On-Market medicines available for each disease
- Integration of ongoing trials from clinicaltrials.gov and WHO database
- Ongoing research activities linked to medicines

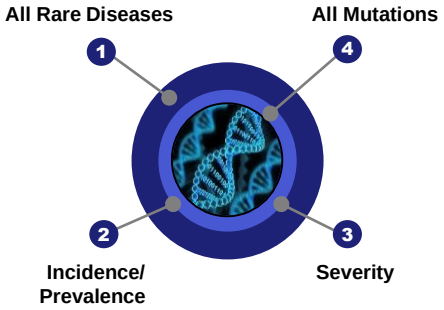
4 Phenotype & Severity

- Cross-reference every disease with HPO/SNOMED-CT terms
- Classify disease severity via tunable phenotype weighting models

AI Decision Support

Decision Support for Rare-Diseases

Leverage the Alexion Insight Engine to create a graph of every phenotype connected to every diseases



Mathematics and machine learning integrated to build a graph-based "twenty questions" framework

$$\Pr(\mathbf{Z}=d|\mathbf{X};\gamma,\alpha,\beta)\propto\Pr(\mathbf{Z}=d,\mathbf{X};\gamma,\alpha,\beta)$$

$$\Pr(\mathbf{Z},\mathbf{X},\boldsymbol{\phi},\boldsymbol{\theta};\gamma,\alpha,\beta)=$$
$$\Pr(\boldsymbol{\phi};\gamma)\Pr(\mathbf{Z}|\boldsymbol{\phi})\prod_{j=1}^{T}\prod_{k=1}^{D}\Pr(\boldsymbol{\theta}_{j,k};\alpha_{j,k},\beta_{j,k})\Pr(\mathbf{X}_j|\boldsymbol{\theta}_{j,k})$$

$$\Pr(\mathbf{Z},\mathbf{X};\gamma,\alpha,\beta)=$$
$$\int_{\boldsymbol{\phi}}\int_{\boldsymbol{\theta}}\Pr(\boldsymbol{\phi};\gamma)\Pr(\mathbf{Z}|\boldsymbol{\phi})\prod_{j=1}^{T}\prod_{k=1}^{D}\Pr(\boldsymbol{\theta}_{j,k};\alpha_{j,k},\beta_{j,k})\Pr(\mathbf{X}_j|\boldsymbol{\theta}_{j,k})d\boldsymbol{\phi}d\boldsymbol{\theta}$$

$$\sigma^2_{X_{j+1}}=\frac{\sum_{i=1}^{D}w_i*\left(Pr(X_{j+1}|Z=i)-\mu^*_{X_{j+1}}\right)^2}{\sum_{i=1}^{D}w_i}$$

$$\mu^*_{X_{j+1}}=\frac{\sum_{i=1}^{D}w_i*Pr(X_{j+1}|Z=i)}{\sum_{i=1}^{D}w_i}$$

$$w_i=\Pr(Z=i|X_j,\ldots,X_1)$$

Piloting with multiple partners to optimize Bayesian priors and correlations for rare-disease differential diagnosis

Let me tell you a story

Smart Panel

Enabling the Diagnosis of <u>any</u> patient with a rare genetic disease

I have my patient's entire exome!

But where do I start looking?

And the clock is ticking....

The opportunity: Helping doctors diagnose rare genetic disorders

+

$$\sigma^2_{X_{j+1}} = \frac{\sum_{i=1}^{D} w_i * \left(Pr(X_{j+1}|Z=i) - \mu^*_{X_{j+1}}\right)^2}{\sum_{i=1}^{D} w_i}$$

$$\mu^*_{X_{j+1}} = \frac{\sum_{i=1}^{D} w_i * Pr(X_{j+1}|Z=i)}{\sum_{i=1}^{D} w_i}$$

$$w_i = \Pr(Z=i|X_j,\ldots,X_1)$$

=

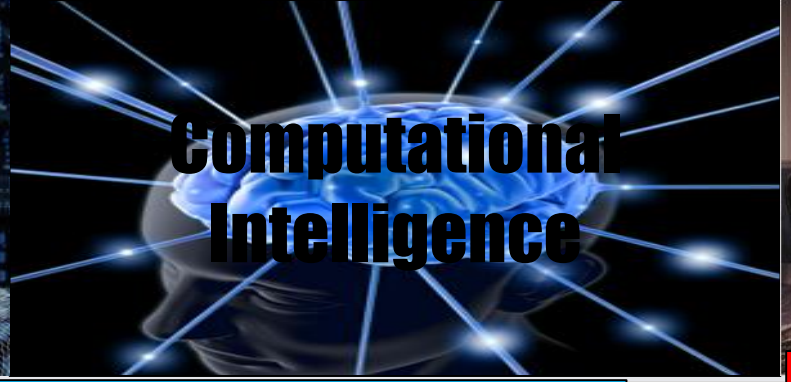
Combining our data graph with an Alexion Bioinformatics machine learning system we are researching how we can turn patient observations into a dynamic sequencing panel

The SmartPanel

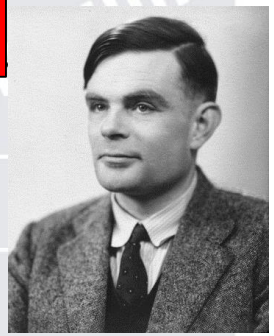
Accelerating the diagnosis of rare genetic diseases

A dynamic custom panel optimized for each patient's unique phenotype

Prioritize where first to look for a pathogenic mutation amidst three billion base pairs of a patient's DNA



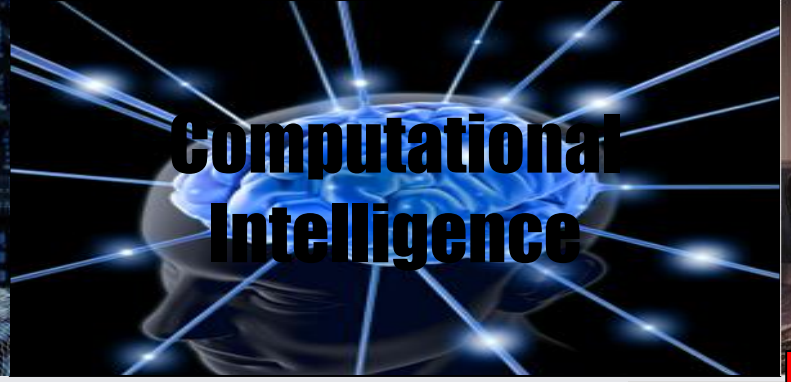
Alan Turing



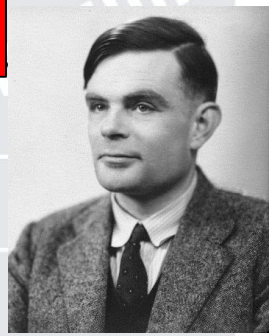
OAK RIDGE National Laboratory

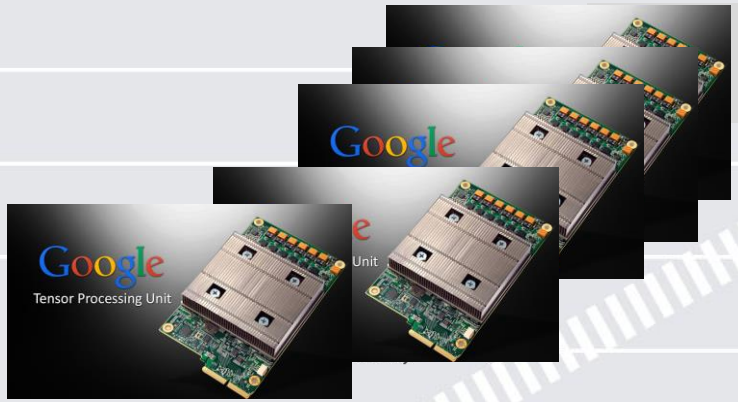
Oak Ridge Summit Supercomputer

ALEXION



Passed!





The RealySmartPanel™

Passing the Turing test in Rare Disease Diagnosis

A dynamic custom panel optimized for each patient's unique phenotype

With a Turing Equivalent method for rare-disease diagnosis against the evidence base

ALEXION


ENTER