AI is Not Enough

Explainable Biology for Improved Therapies
Artemis

The Hospital for Sick Children in Toronto, the Neonatal Intensive Care unit

- Monitoring premature infants' vital sign data using IBM InfoSphere Streams
- Each infant generates >1,200 points of data per second
- Nearly 90 million points of data per day
MDA: IBM WATSON & CANCER

MD Anderson Taps IBM Watson to Power "Moon Shots" Mission Aimed at Ending Cancer, Starting with Leukemia (Houston - 18 Oct 2013)

4 YEARS AND $62M LATER

Watson can't cure cancer ... or all the stuff that breaks IT projects
University spends $62m on AI trial, gets the usual trials that come with failure

How IBM Watson Overpromised and Underdelivered on AI Health Care

After its triumph on Jeopardy!, IBM’s AI seemed poised to revolutionize medicine. Doctors are still waiting

By Eliza Strickland
IBM WDD vs IID & NAViGaTOR

Literature vs Literature + Data + Predictions
1,862 proteins connected by 2,324 edges
Challenges in Data-Driven Medicine

- **Technical – ethical – legal**
- **Heterogeneity** – samples, assays, analysis
- **Biases** – data collection, confounding factors, ground truth, bad platforms
  - Unknowns – lifestyle, (co-)treatments, missing data
  - Fraud (75%), mistakes (<25%)
- **Statistical significance vs biological/clinical importance**
- **AI/ML issues** – transfer learning, drift of cohorts
  - Performance on the cohort (accuracy) vs confidence in a single prediction
- Importance of **explanation & modeling**, not just classification

AI: **Algorithm – Data – System – Use**
Data Quality

Medical-AI Systems and Applications: Future Challenges and Research Directions

Andreas Holzinger\textsuperscript{a,b,*}, Matthias Dehmer\textsuperscript{c,d}, Frank Emmert-Streib\textsuperscript{e}, Natalia Díaz-Rodríguez\textsuperscript{i}, Chris Biemann\textsuperscript{l}, Igor Jurisica\textsuperscript{f,g,h,i}

![Graph a)](image)

![Graph b)](image)
Ubiquitin ligase RNF8 suppresses Notch signaling to regulate mammary development and tumorigenesis

Li Li, ..., Anne Hakem, Razq Hakem

AI & our Goals?

> Expertise, process, standards!

<table>
<thead>
<tr>
<th>Test Dataset</th>
<th>Human Benchmark</th>
<th>Metric</th>
<th>Clinical Decision (%)</th>
<th>AI Decision (%)</th>
<th>( \Delta ) (%)</th>
<th>95% CI (%)</th>
<th>p-value</th>
<th>Comparison</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK first reader</td>
<td></td>
<td>sensitivity</td>
<td>62.69</td>
<td>65.42</td>
<td>2.70</td>
<td>(-3.0, 8.5)</td>
<td>0.0043</td>
<td>noninferiority</td>
<td>402</td>
</tr>
<tr>
<td></td>
<td></td>
<td>specificity</td>
<td>92.93</td>
<td>94.12</td>
<td>1.18</td>
<td>(0.29, 2.8)</td>
<td>0.0096</td>
<td>superiority</td>
<td>25,115</td>
</tr>
<tr>
<td>UK second reader</td>
<td></td>
<td>sensitivity</td>
<td>69.40</td>
<td>69.40</td>
<td>0.00</td>
<td>(-4.89, 4.89)</td>
<td>0.0225</td>
<td>noninferiority</td>
<td>402</td>
</tr>
<tr>
<td></td>
<td></td>
<td>specificity</td>
<td>92.97</td>
<td>92.13</td>
<td>-0.84</td>
<td>(-1.97, 0.282)</td>
<td>2e-13</td>
<td>noninferiority</td>
<td>25,113</td>
</tr>
<tr>
<td>UK Consensus</td>
<td></td>
<td>sensitivity</td>
<td>67.39</td>
<td>68.12</td>
<td>0.72</td>
<td>(-3.49, 4.94)</td>
<td>0.0039</td>
<td>noninferiority</td>
<td>414</td>
</tr>
<tr>
<td></td>
<td></td>
<td>specificity</td>
<td>96.24</td>
<td>96.24</td>
<td>-3.35</td>
<td>(-4.06, -2.63)</td>
<td>3e-6</td>
<td>noninferiority</td>
<td>25,442</td>
</tr>
<tr>
<td>USA reader</td>
<td></td>
<td>sensitivity</td>
<td>48.10</td>
<td>57.50</td>
<td>9.40</td>
<td>(4.45, 13.85)</td>
<td>0.0004</td>
<td>superiority</td>
<td>553</td>
</tr>
<tr>
<td></td>
<td></td>
<td>specificity</td>
<td>80.83</td>
<td>86.53</td>
<td>5.70</td>
<td>(2.62, 8.64)</td>
<td>0.0002</td>
<td>superiority</td>
<td>2,185</td>
</tr>
<tr>
<td>USA reader</td>
<td></td>
<td>sensitivity</td>
<td>48.10</td>
<td>56.24</td>
<td>8.14</td>
<td>(3.54, 12.5)</td>
<td>0.0006</td>
<td>superiority</td>
<td>553</td>
</tr>
<tr>
<td></td>
<td></td>
<td>specificity</td>
<td>80.83</td>
<td>84.29</td>
<td>3.47</td>
<td>(0.6, 5.98)</td>
<td>0.0212</td>
<td>superiority</td>
<td>2,185</td>
</tr>
</tbody>
</table>
Integrative Computational Biology

1. Biobank
   - Omics profiles across genome, proteome, metabolome can be analyzed separately or combined to find differentially expressed entities

2. HTP platforms
   - Network relationships link relevant entities within each data layer and identify better biomarkers

3. Data portals
   - Integration, Analytics
   - Layers of annotated networks; annotated with tissue, disease, network properties can further characterize potential biomarkers

4. Models
   - Discovered relationships across data layers identify combined biomarkers, drug mechanism of action and create explainable disease models

5. Precision treatments
   - Improved outcomes
   - Combined biomarkers identify clinically-relevant patient subgroups
   - Treatment tailored to patient subgroups results in improved patient outcomes
AI & Biomarkers in Precision Medicine
Positive correlation among all urine excretion rates of AngII-regulated proteins.

Urinary excretion rates of AngII signature proteins is increased in patients with IFTA (interstitial fibrosis/tubular atrophy):
- AngII, Angiotensin II
- BST1, bone marrow stromal cell antigen 1
- GLNA, glutamine synthetase
- LAMB2, laminin subunit beta-2
- LYPAL1, lysophospholipase I
- RHOB, ras homolog family member B
- TSP1, thrombospondin-1

Urine Angiotensin II Signature Proteins as Markers of Fibrosis in Kidney Transplant Recipients

- Zainaa Mohammad-Ali, MD, PhD, Tomas Boko, MD, PhD, Iyork Betru, MD, PhD, Robert Rice, MD, and Ana Kohnakova, MD, PhD, FRCP(C)
- Alexandre Tavares-Brum, BSc, Paul Yip, PhD, FCAH,(C)
- Hadije Cardinal, MD, PhD, FRCP(C)
- Marie-Josée Hébert, MD, PhD, FRCP(C)
- Yvonne Lee, MD, PhD, and Seth Joseph, MD, PhD, FRCP(C)

Quantification of angiotensin II-regulated proteins in urine of patients with polycystic and other chronic kidney diseases by selected reaction monitoring.
Translating from Kidney to Lung
Peripheral Blood RNA Sequencing Unravels a Differential Signature of Coding and Noncoding Genes by Types of Kidney Allograft Rejection

Silvia Pineda, Swastika Sun, Tara Sigdel, Mark Nguyen, Elena Crespo, Alba Tonija, Maria Menaghimi, Montse Goma, Marina Sinotes, Onno Bestard, and

Linking to Other Studies

Association between renin-angiotensin system and chronic lung allograft dysfunction
Differentially expressed microRNAs in lung adenocarcinoma invert effects of copy number aberrations of prognostic genes


Paradoxes
1,062 LUAD RNA samples
241 normal lung samples
63 LUAD aCGH samples
→ 85 paradoxical genes
TCGA LUAD: 70/85 genes validated

46/70 paradoxical genes (65.7%) are targeted by subsets of the 19/24 deregulated miRNAs ($p = 4.8E-3$)
369 co-expressed miRNA:gene pairs (27% of all miRNA:gene combinations), 362 (98.1%) of which are explanatory.
41/70 genes (58.6%) significantly associated with survival (FDR < 0.05)
Tissue-Specific Network Remodeling

Epithelial Tumour Suppressor ELF3 is a Lineage-Specific Amplified Oncogene in Lung Adenocarcinoma

Katey SS Enfield1, Erin A Marshall1, Christine Anderson1, Kevin W Ng1, Sara Rahmati2, Zhaolin Xu3, Megan Fuller1, Katy Milne4, Daniel Lu1, Rocky Shi1, David A Rowbotham1, Daiana D Becker-Santos1, Fraser Johnson1, John C English5, Calum E MacAulay1, Stephen Lam1, William W Lockwood1, Raj Chari6, Aly Karsan1, Igor Jurisica2, Wan L Lam1
Integrating Networks

miR-34a-5p:target interactions
Transcription factor:target interactions
Protein:protein interactions
multiple
Osteoarthritis and Cartilage

Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis.

## Data Integration Portal

- Lung, ovarian, prostate, sarcoma, head & neck, liver, OA, RA, NDD, fibrosis, kidney

<table>
<thead>
<tr>
<th>Tumor Type</th>
<th>#Datasets</th>
<th>#Samples</th>
<th>#UP genes Top 1%</th>
<th>#DOWN genes Top 1%</th>
<th>#Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA</td>
<td>87</td>
<td>11,609</td>
<td>130</td>
<td>183</td>
<td>13,193</td>
</tr>
<tr>
<td>Lung CA</td>
<td>30</td>
<td>6,270</td>
<td>51</td>
<td>625</td>
<td>16,658</td>
</tr>
<tr>
<td>Ovarian CA</td>
<td>87</td>
<td>839</td>
<td>123</td>
<td>115</td>
<td>9,942</td>
</tr>
<tr>
<td>Prostate CA</td>
<td>125</td>
<td>850</td>
<td>109</td>
<td>167</td>
<td>12,472</td>
</tr>
<tr>
<td>HN CA</td>
<td>291</td>
<td>1,437</td>
<td>90</td>
<td>78</td>
<td>9,110</td>
</tr>
<tr>
<td>Sarcoma</td>
<td>224</td>
<td>6,032</td>
<td>201</td>
<td>126</td>
<td>14,637</td>
</tr>
<tr>
<td>Liver CA</td>
<td>85</td>
<td>3,899</td>
<td>51</td>
<td>340</td>
<td>12,355</td>
</tr>
</tbody>
</table>

### Results: 11

<table>
<thead>
<tr>
<th>Adjacent Chemo</th>
<th>No</th>
<th>Unknown</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>min</td>
<td>max</td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td>ADC</td>
<td>LCUC</td>
<td>SGCCE</td>
</tr>
<tr>
<td>Recurrence</td>
<td>No</td>
<td>Unknown</td>
<td>Yes</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>Smoking History</td>
<td>CURRENTLY SMOKING</td>
<td>NEVER SMOKED</td>
<td>SMOKED IN THE PAST</td>
</tr>
<tr>
<td>Stage</td>
<td>i</td>
<td>ii</td>
<td></td>
</tr>
<tr>
<td>Time to Event</td>
<td>min</td>
<td>max</td>
<td></td>
</tr>
</tbody>
</table>
OsteoDIP: A web-based gene and non-coding RNA expression database


Deregulated in at least one study:

- 8,905 genes
- 402 lncRNAs
- 56 microRNAs
- 58 circRNAs
<table>
<thead>
<tr>
<th>Gene/Mir</th>
<th>Uniprot</th>
<th>Total Up</th>
<th>Total Dow</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXCL1</td>
<td>P09341</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>DNASE1L3</td>
<td>Q13609</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>IL6</td>
<td>P05231</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>MMP9</td>
<td>P14780</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>ACAN</td>
<td>P16112</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>ACKR3</td>
<td>P25106</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ACP2</td>
<td>P11117</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ACP5</td>
<td>P13686</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ACSL4</td>
<td>O60488</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ACTG2</td>
<td>P63267</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>HTRA1</td>
<td>Q92743</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>TNFAIP6</td>
<td>P98066</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>CCND1</td>
<td>P24385</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>COL1A2</td>
<td>P08123</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>FN1</td>
<td>P02751</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>RCAN1</td>
<td>P53805</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>TGFB1</td>
<td>Q15582</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>VCAN</td>
<td>P13611</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>ADAM12</td>
<td>O43184</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>BASP1</td>
<td>P80723</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>CD55</td>
<td>P08174</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>CILP</td>
<td>O75339</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>CLEC3B</td>
<td>P05452</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>COL1A1</td>
<td>P02452</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>COL2A1</td>
<td>P02458</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>COL5A1</td>
<td>P20908</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>CRLF1</td>
<td>Q75462</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>ITGB1</td>
<td>P05556</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>MARCKS</td>
<td>P29966</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>MMP1</td>
<td>P03956</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>MMP2</td>
<td>P08253</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>MMP9</td>
<td>P14780</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>MSX2</td>
<td>P35548</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>PSAT1</td>
<td>Q9Y617</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>PTN</td>
<td>P21246</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>TF</td>
<td>P02787</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>THBS2</td>
<td>P35442</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>THBS3</td>
<td>P49746</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>TMSB4X</td>
<td>P62328</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
Dexamethasome (hormone)
Progesterone (hormone)
Estradiol (hormone)
Trichlostatin A (HDAC inhibitor)
JQ1 compound (BET inhibitor)
AI in Drug Discovery Pipeline

- Diagnostics
- Biomarker discovery
- Drug repurposing
- Drug target discovery
- Drug development
- Testing
- Clinical trial
- Drug delivery optimization

Advanced AI tools for a specific Use Case
Advanced AI systems with multiple models
End-to-end AI Platforms

Expertise in AI

Clinical pipeline
Preclinical pipeline
Validated R&D Use cases

Target Discovery Lead Discovery Lead Optimization Pre-Clinical Phase I Phase II Phase III FDA
OpenPandemics – COVID-19

Dr. Stefano Forli
Principal Investigator
Scripps Research

- Rapid deployment
- 300 days – 100K volunteers

300,000,000 simulations
20,000 compounds screened
70 promising compounds
25 compounds in lab testing
Limitations of AI Speedup

> Grüenenthal Thalidomide (Contergan)

“As a result of the thalidomide scandal, the Health Law in West Germany was strengthened and new requirements for pharmaceutical testing were created; the Federal Ministry of Health was established in 1962.”

https://en.wikipedia.org/wiki/Grunenthal
MoA of Antipsychotic Drugs

IID NAVIGaTOR
TP53

Capturing variation impact on molecular interactions in the IMEx Consortium mutations data set

Authors: S Makhija, D D Taylor, R K Gibb, C GęcTEL-Taylor

Taxol-induced bcl-2 phosphorylation in ovarian cancer cell monolayer and spheroids.
PPIs & Mutation

TP53 mutations in high grade serous ovarian cancer and impact on clinical outcomes: a comparison of next generation sequencing and bioinformatics analyses

Towards a unified open access dataset of molecular interactions
Deep longitudinal multiomics profiling reveals two biological seasonal patterns in California

M. Reza Salehi\textsuperscript{1,5}, Ahmed A. Metwally\textsuperscript{1,5}, Wenyu Zhou\textsuperscript{1}, Sophia Miriam Schüssler-Fiorenza Rose\textsuperscript{1}, Sara Ahadi\textsuperscript{1}, Kevin Contrepois\textsuperscript{1}, Tejaswini Mishra\textsuperscript{1}, Martin Jinye Zhang\textsuperscript{2}, Łukasz Kidziński\textsuperscript{3}, Theodore J. Chu\textsuperscript{4} & Michael P. Snyder\textsuperscript{103}
BMI – Measurements … and Causality

Figure 8: Proportion of overweight or obese individuals

Source: CCHS (2017)

http://www.thetimes.co.uk/tto/health/news/article3764026.ece
Comprehensive & Systematic

An atlas of human metabolism


The proteome landscape of the kingdoms of life

Innovative Omics

An atlas of the protein-coding genes in the human, pig, and mouse brain

Precision Medicine

> Systematically profile and characterize the “X” spectrum of diseases

> Data availability, quality, annotation
   » Standard protocols for sample collection/profiling
   » Detailed clinical information
   » Curated databases with relationships

> Explainable models
   » Increase trust, improve robustness
   » Proper train-test-validate workflow
   » Identify (sub)-cohorts, biomarkers, “missing” data, treatments

... from data to models to insights & treatments
Neuro: Sandmo, Filipcik, Cente, Hanes, Poschmann, Schulze, Tator, Kalia, Tartaglia, Paus, …
Cancer: Shepherd, Sound-Tsao, Lam, Singh, Kamel-Reid, Reis, Oza, Hedley, Mes-Masson, Jurisicova, Kaur, Kislinger, Rogatto, Dick, Minden, Salmena, Mego, …
Immunology: Lapointe, Winer, Tsai, Marques, …
Transplant: Konvalinka, Keshavjee, Martinu, …
Ph. Activity: Jurisicova, Cameron, Kotsopoulos, …
Other: Stagljar, Maestro, Mills, Ricard-Blum, Hermjakob, Orchard, Tyers, Porras Millan, McCulloch, Glogauer, Rigoutsos, Holzinger, Gunning, DeTitta, Luft, Snell, …