Machine learning and medicine: Where do we stand and where are we going?

Sach Mukherjee, DZNE, Bonn

Machine learning meets biostatistics

Utrecht, May 2019

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DZNE
Deutsches Zentrum für Neurodegenerative Erkrankungen in der Helmholtz-Gemeinschaft

Joint work with (among others):

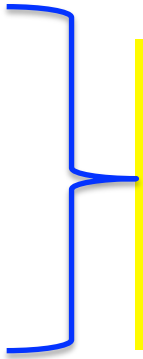
Frank Dondelinger
Steven Hill
Umberto Noe
Chris Oates
Konstantinos Perrakis
Joachim Schultze
Bernd Taschler
Stefanie Warnat-Herresthal

ML & medicine

- **Where we stand:** what's changed in recent years and some snapshots of recent work in **prediction** and **causal learning**
- **Where we're going:** how will things look in the future? Why aren't we further ahead already?

Where we stand: the high-dimensional revolution

- Last ~15 years, major developments in understanding of:
  - Regularization and high-dimensional learning
  - Highly flexible, data-adaptive models
  - Computationally feasible learning schemes
  - Hardware and associated libraries

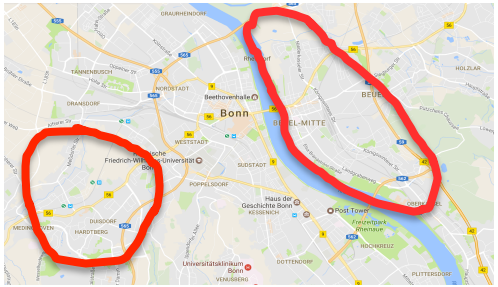


Dramatic change in how we view high-dimensional data and complex models

Where we stand: scale up of phenotyping/data acquisition

RHEINLAND STUDIE

- **Up to 30,000 people**
- **Follow-up 30 years or more**
- **Deep phenotyping**
- **Multiple modalities**



Two kinds of questions

- **Two broad kinds of questions/tasks:**

  (1) "Predictive". Can be framed in terms of minimizing some kind of expected loss, typically supervised learning set-up.

  Examples: diagnosis, prognosis, "theranostics", some pre-processing…

  (2) "Causal". Goal is to guide new interventions.

  Examples: identifying new therapies, aetiology, preventative factors …

- Depends on not only the question ("will treatment A work for patient X") but the data context
- Current/emerging biotechnological and data science tools offer promise of major changes in both areas

Prediction in medicine

Prediction in medicine

- Many medical tasks are fundamentally statistical decision problems, including:
  - Diagnosis
  - Prognosis
  - Theranostics

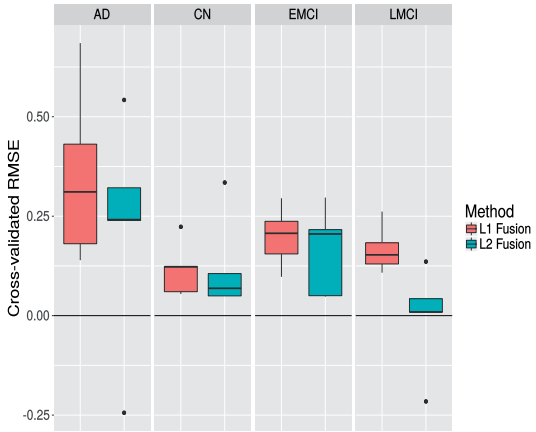
- With appropriate data, all can be viewed as supervised learning problems, with different *X*s and *Y*s

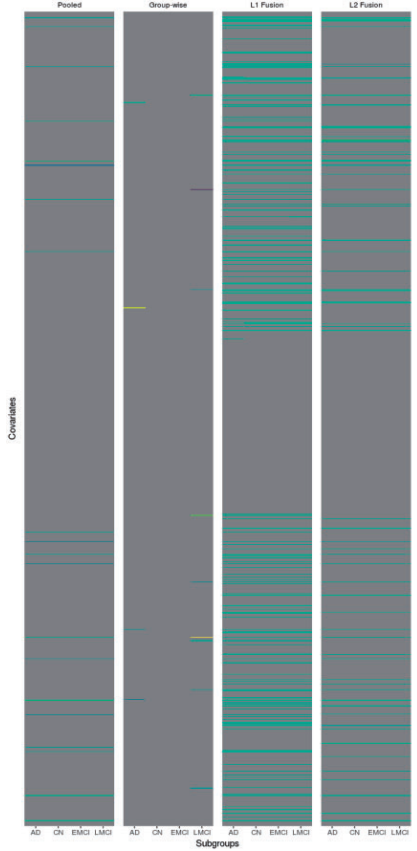
Why prediction in medicine is different

- Medical applications of supervised learning have some key features:
  - Heterogeneity (within study)
  - Batch- and population-type effects and generalizability
  - Multi-modality
  - High-dimensionality, weak first-principles information
  - Bayes' risk not known in advance – always an empirical question
  - Ethical questions

Heterogeneity: joint learning over subtypes

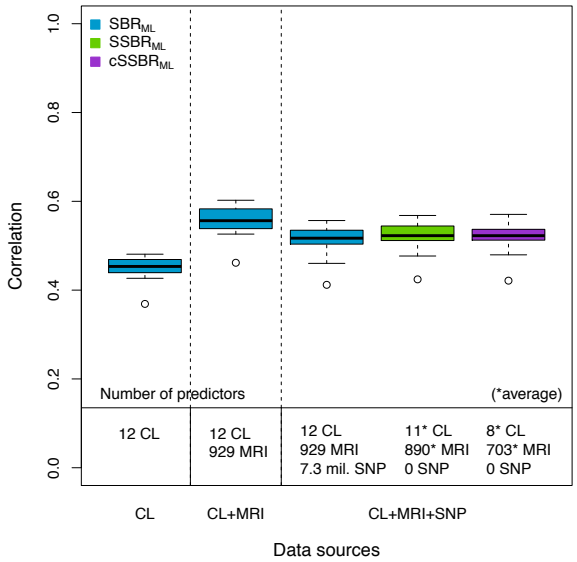
- Joint Lasso (Dondelinger & SM, *Biostatistics*, 2018), augments classical lasso penalty with between-group terms that allow for joint learning
- Can offer gains in prediction, also quite different sparsity patterns





High-dimensional, multi-modal data

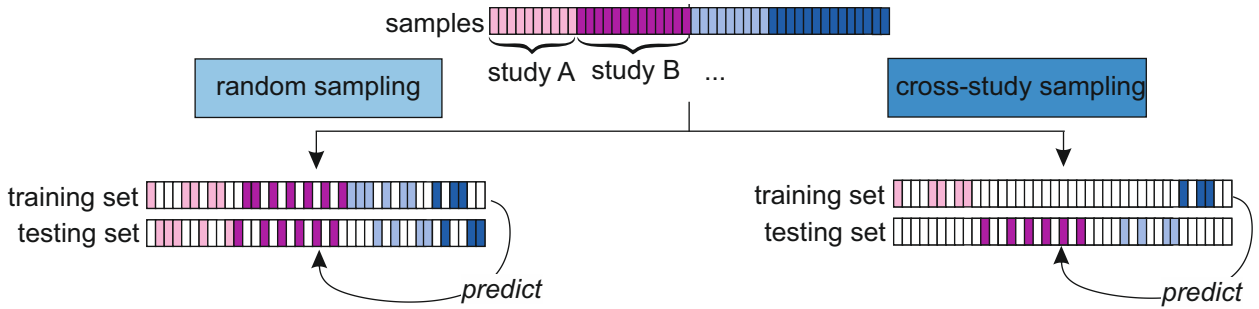
- Scalable Bayesian regression (Perrakis & SM, *JCGS*, to appear), allows for multiple modes with high total dimension
- Example from multi-modal Alzheimers' prediction



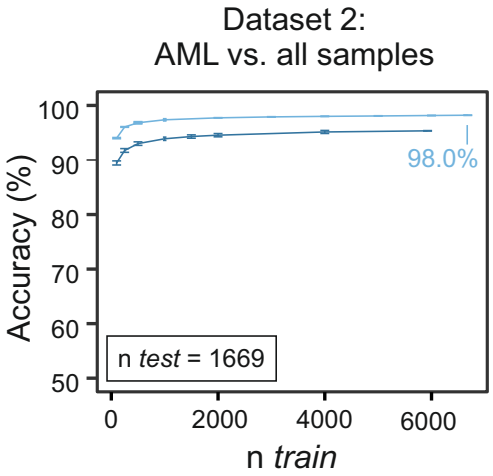
Risk estimation in practice

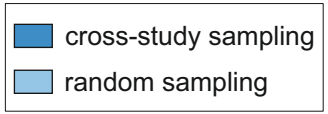
- Detailed study of one potential use-case: blood-based diagnosis of leukaemia
- Problem well known to contain gene expression signals – question is how to assess reliability/usefulness of predictors?
- Large, multi-site data, total n ~ 12000 samples, p ~ 12000 genes
- Joint with Schultze lab

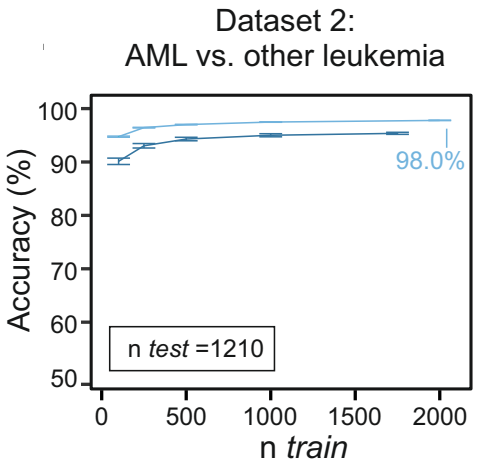
Cross-sampling to test generalization



(Warnat-Herresthal, Perrakis et al., 2019)

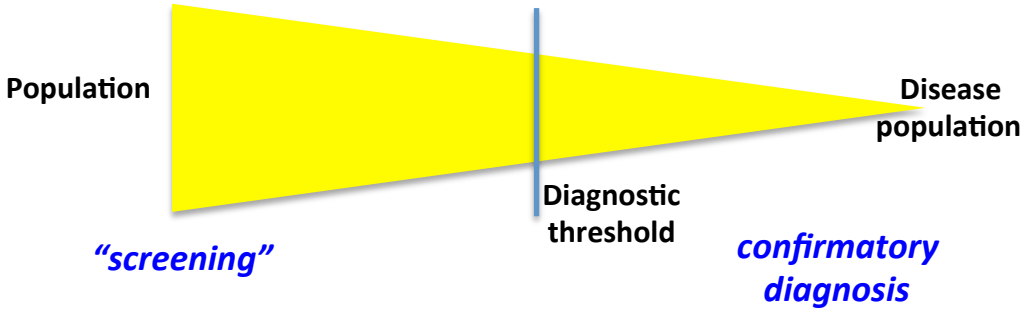




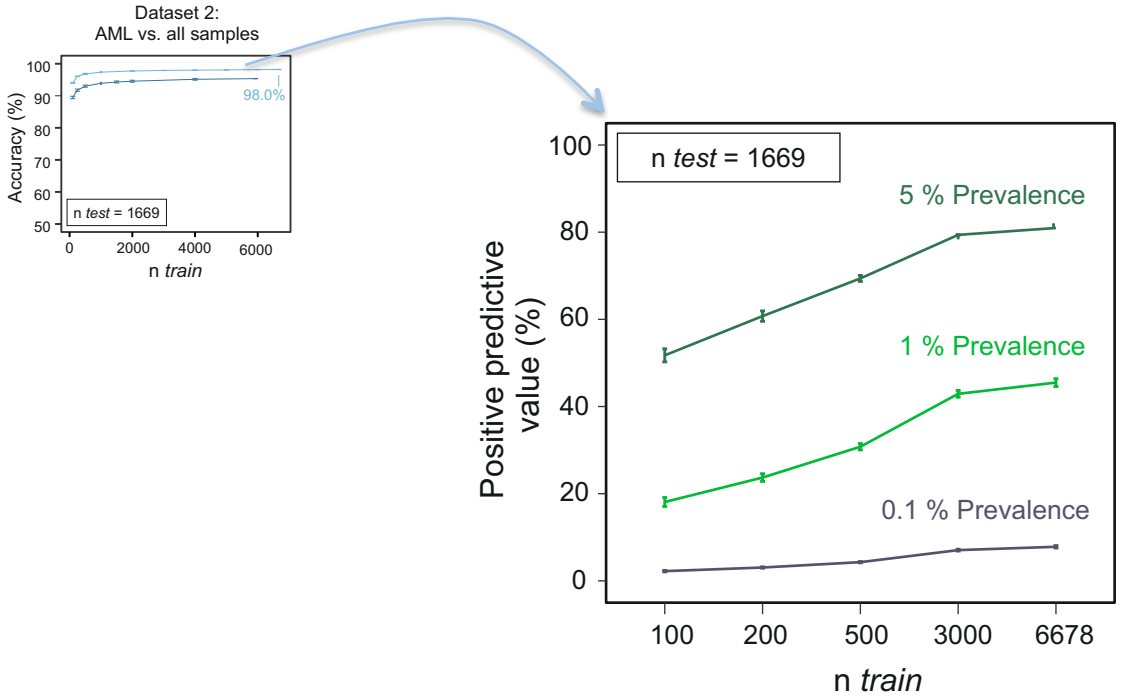


(differential diagnosis)

The diagnostic threshold



- ML methods have low marginal cost → opens up possibility of moving the diagnostic threshold, i.e. invoking predictor earlier
- Invoking predictor earlier → larger population "at risk", lower prevalence, implications for positive predictive value (PPV)…



- Small gains can mean large differences in PPV, hence depending on application, may need very good predictors

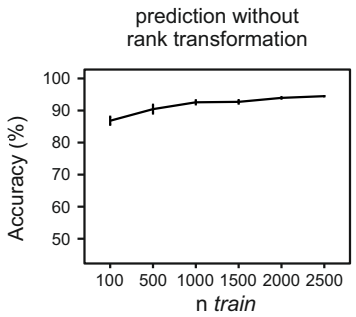
Technological change

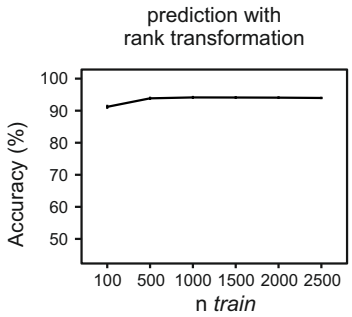
- Data acquisition does not stay fixed over time → biotechnologies change
- Can “old” results still be used?

- Train entirely on one technology/generation
- Test on another, disjoint with respect to study/technology/ samples/normalization
- Covers first and second gen microarrays and RNA-seq

Setting ①

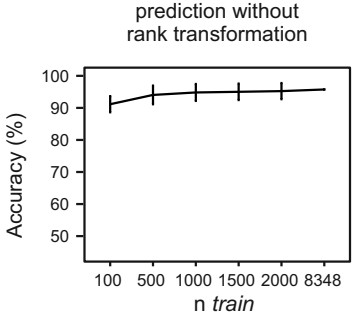
Gen1 → Gen2

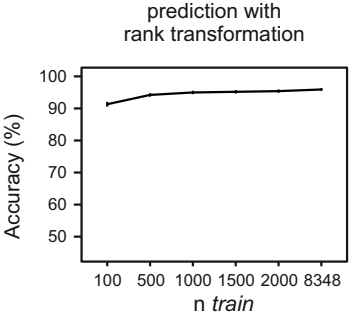




Setting ②

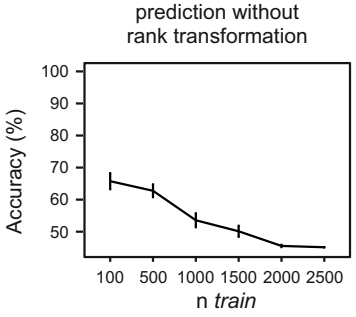
Gen2 → Gen3

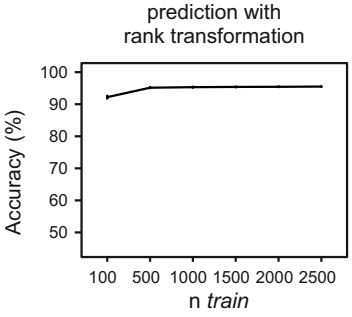




Setting ③

Gen1 → Gen3



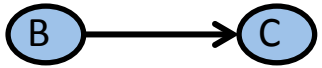


Towards scalable causal learning

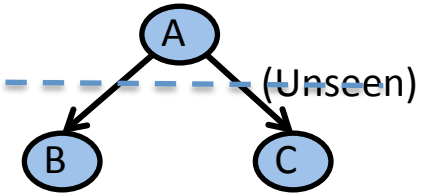
Causal models in biomedicine

- Scientists (rightly) point out that there is life beyond prediction
- Can we make this statement precise?
- Yes: point is that some biomedical questions are *causal or mechanistic*, cannot be directly addressed by multivariate modelling or prediction
- Causal ideas are needed to scale up molecular study of disease processes

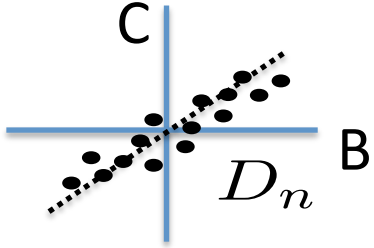
But what is so different about causality?



(possibly latent) sequence of mechanistic events



***no** sequence of mechanistic events linking B and C*



Issue does not go away asymptotically, not solved by more data
Widespread in high dimensions

BIOINFORMATICS ORIGINAL PAPER

Vol. 28 no. 21 2012, pages 2804–2810
doi:10.1093/bioinformatics/bts514

Systems biology

Advance Access publication August 24, 2012

Bayesian Inference of Signaling Network Topology in a Cancer Cell Line

Steven M. Hill[1,2,3], Yiling Lu[4], Jennifer Molina[4], Laura M. Heiser[5], Paul T. Spellman[6], Terence P. Speed[7,8], Joe W. Gray[6], Gordon B. Mills[4,*] and Sach Mukherjee[1,3,*]

The Annals of Applied Statistics
2015, Vol. 9, No. 1, 507–524
DOI: 10.1214/15-AOAS806

INFERRING NETWORK STRUCTURE FROM INTERVENTIONAL TIME-COURSE EXPERIMENTS

BY SIMON E. F. SPENCER*, STEVEN M. HILL† AND SACH MUKHERJEE†,‡,1

Journal of Machine Learning Research 17 (2016) 1-23

Estimating Causal Structure Using Conditional DAG Models

Chris. J. Oates CHRISTOPHER.OATES@UTS.EDU.AU
School of Mathematical and Physical Sciences
University of Technology Sydney
NSW 2007, Australia

Jim Q. Smith J.Q.SMITH@WARWICK.AC.UK
Department of Statistics
University of Warwick
Coventry, CV4 7AL, UK

Sach Mukherjee SACH.MUKHERJEE@DZNE.DE
German Center for Neurodegenerative Diseases (DZNE)
53175 Bonn, Germany

nature methods
Techniques for life scientists and chemists

- An analysis of causal network inference
- Lattice light sheet-PAINT imaging of thick specimens
- HomoFRET based sensors for NADP+
- Membrane protein-ligand binding detected by MS
- Measuring activity in multiple brain regions

Inferring causal molecular networks: empirical assessment through a community-based effort

Steven M Hill[1,28], Laura M Heiser[2–4,28], Thomas Cokelaer[5,27], Michael Unger[6,7], Nicole K Nesser[8], Daniel E Carlin[9], Yang Zhang[10,27], Artem Sokolov[9], Evan O Paull[9], Chris K Wong[9], Kiley Graim[9], Adrian Bivol[9], Haizhou Wang[10,27], Fan Zhu[11], Bahman Afsari[12], Ludmila V Danilova[12,13], Alexander V Favorov[12–14], Wai Shing Lee[12], Dane Taylor[15,16], Chenyue W Hu[17], Byron L Long[17], David P Noren[17], Alexander J Bisberg[17], HPN-DREAM Consortium[18], Gordon B Mills[19], Joe W Gray[2–4], Michael Kellen[20], Thea Norman[20], Stephen Friend[20], Amina A Qutub[17], Elana J Fertig[12], Yuanfang Guan[11,21,22], Mingzhou Song[10], Joshua M Stuart[9], Paul T Spellman[8], Heinz Koeppl[6,7,27], Gustavo Stolovitzky[23], Julio Saez-Rodriguez[5,24] & Sach Mukherjee[1,25–27]

Cell Systems
Article

OPEN ACCESS
CellPress

BIOINFORMATICS

Vol. 30 ECCB 2014, pages i468–i474
doi:10.1093/bioinformatics/btu452

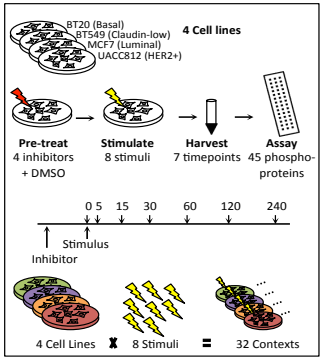
Causal network inference using biochemical kinetics

Chris J. Oates[1,*], Frank Dondelinger[2], Nora Bayani[3], James Korkola[4], Joe W. Gray[4] and Sach Mukherjee[2,5,*]

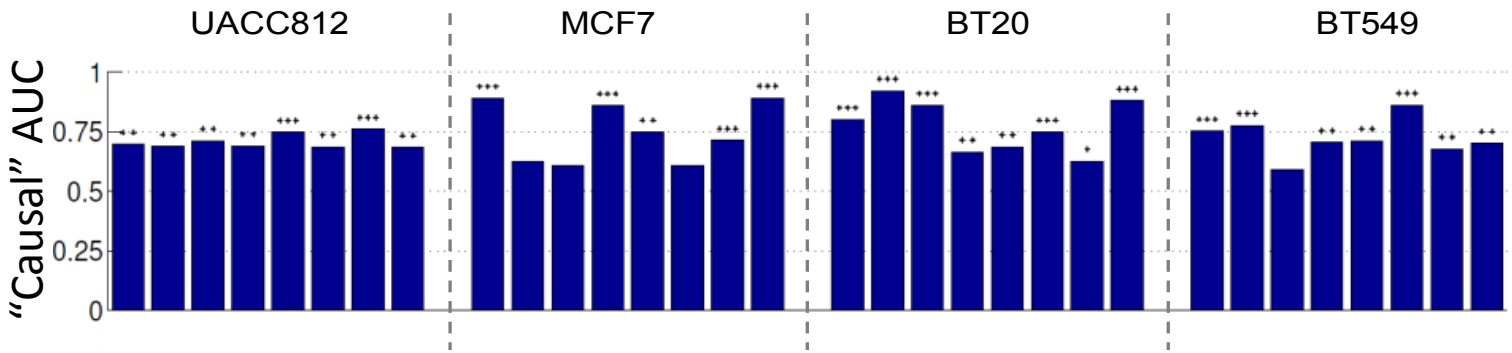
Context Specificity in Causal Signaling Networks Revealed by Phosphoprotein Profiling

Steven M. Hill,[1,12] Nicole K. Nesser,[2,12] Katie Johnson-Camacho,[2] Mara Jeffress,[3] Aimee Johnson,[4] Chris Boniface,[2] Simon E.F. Spencer,[5] Yiling Lu,[6] Laura M. Heiser,[7] Yancey Lawrence,[2,13] Nupur T. Pande,[8,9] James E. Korkola,[7] Joe W. Gray,[7,8,10] Gordon B. Mills,[6] Sach Mukherjee,[1,11,14,*] and Paul T. Spellman[2,15,*]

Testing agreement with *unseen* interventions



(Hill et al., 2016)



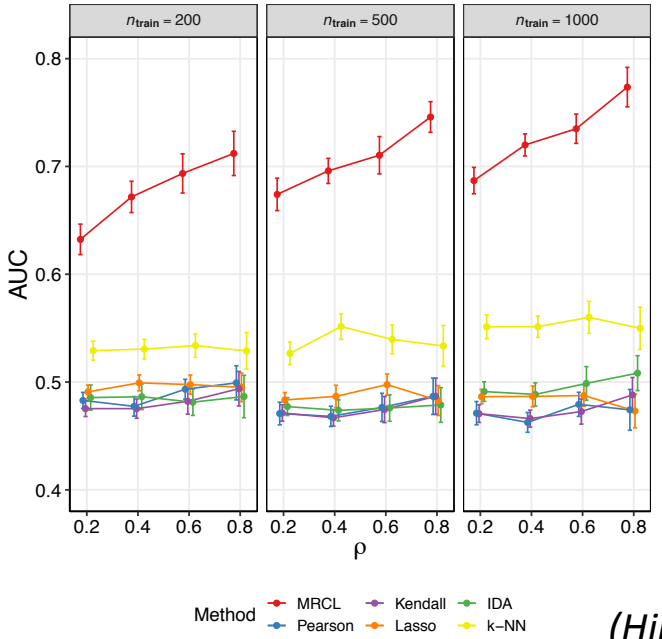
(Hill et al., 2017)

Towards truly *scalable* causal learning

- Casual learning: very hard problem, progress exciting, but existing approaches do not always scale well in terms of *p* or human overhead
- Recently pursuing new approach, based on *causal manifolds*
- Idea is to bypass graphical models whilst learning asymmetric relationships at scale

- Some examples using large scale experimental data…

Causal manifold learning



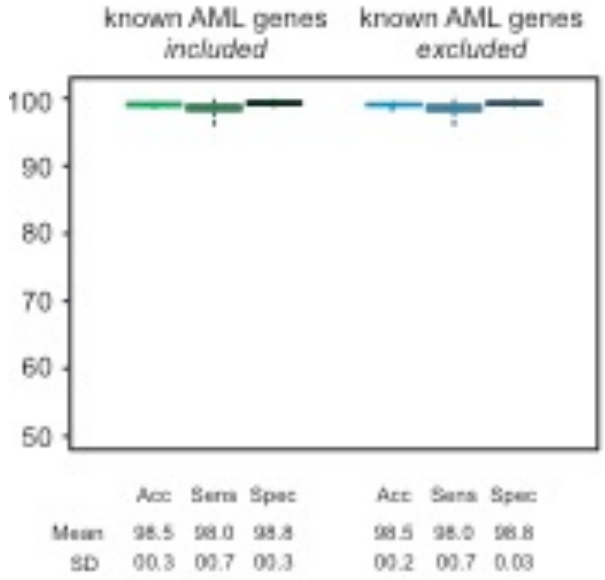
(Hill et al., arXiv:1612.05678 [stat.ML])

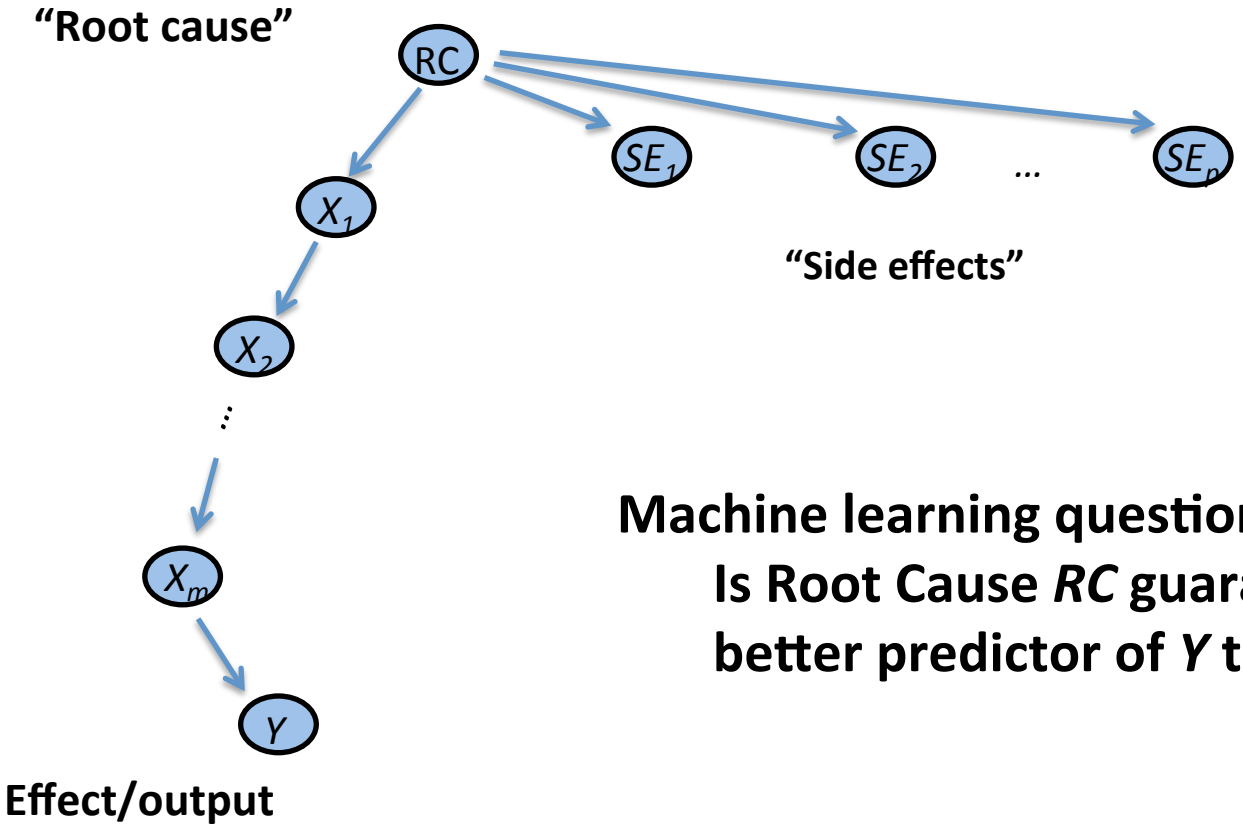
- Extensive yeast data, *p*=50, tested against experimentally-verified causal relationships
- Significantly outperforms several existing approaches
- Ongoing: scaling and testing on human-genome-wide scale problems

Causality and prediction

Causality and prediction

- Different problems, often confused in medical research
- Predictive or multivariate tools do not in general work for causal learning
- But equally mechanistic insights may not be very relevant for prediction!
- Example, go back to leukaemia data….
- Include/entirely exclude known causal drivers… → known disease drivers not needed for prediction





Machine learning question:

Is Root Cause *RC* guaranteed to be a better predictor of *Y* than e.g. side effects $\{SE_j\}$?

The prediction paradox

- In real-world systems – with measurement noise, nontrivial correlation structure etc. – *not* guaranteed that true model *class* is best predictor
- Composing mechanistic models across scales may not work, end-to-end input-output mapping may be more effective
- Real-world examples: cancer prediction, speech recognition (Jelinek: “Anytime a linguist leaves the group the recognition rate goes up”), and more

Implications:

- Be clear about nature of task!

- For prediction, more/better data and good regularization are key

ML and medicine: where are we going?

Where are we going?

- ML and AI methods solve decision problems using data, and ML and statistical concepts allow objective assessment of performance
- Decision problems are ubiquitous in medicine → **what would a truly ML-assisted hospital or healthcare system look like?**
  - **Data-driven decisions**, empirical assessment of *both* artificial and human intelligence based decision processes
  - **Redefine diseases**, identify subgroups, direct therapy
  - Allow **systems-level optimization**

Where are we going?

- **What would truly ML-assisted biological research look like?**
  - Near-automated data collection
  - Iterative, near-automated experimental design/active/reinforcement learning
  - Systematic, empirical link to translational goals

Claim: we are currently far away from what could be achieved even with *current* technology!

(wikipedia)

(nobelprize.org)

You can see the computer age everywhere
but in the productivity statistics (Solow, 1987)

ML/AI as “general purpose” technologies

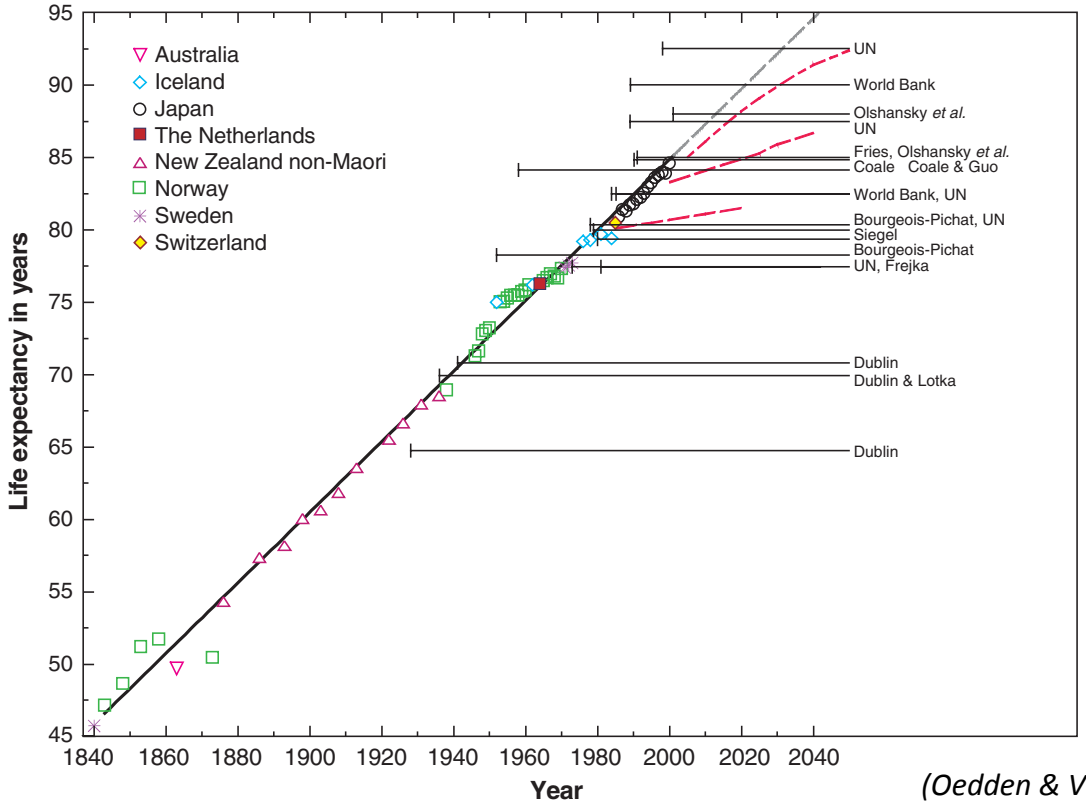
- ML and AI methods solve decision problems using data → this is extraordinarily general
- Some economists consider ML/AI as (potentially) a GPT
- Some characteristics:
  - Scalable, low marginal cost → expands scope of what’s possible
  - Potential to change entire workflows or even systems
  - BUT: seeing the gains may require many changes at once (so-called “complementary innovations”)

Why don’t we have data-driven medicine yet?

- **Why aren’t we further along the road to truly data-driven medicine?**

- Is this a specific case of the **Solow paradox?**
- Idea is that precisely because big advances require *coupled* changes, lags can be long. Borne out by economic history (see e.g. Brynjolfsson et al., 2017)

Implication: collectively need to work not only on primary innovations, but on all the things needed to take advantage of them



(Oedden & Vaupel, 2002)

Can ML/AI/genomics/phenotyping contribute to keeping this success story going?