RowAnalytics
transforming the delivery of health

Based on our estimate, medical error is the 3rd most common cause of death in the US

Cancer 585k
Heart disease 611k
Motor vehicles 34k
Suicide 41k
Motor vehicles 34k
Suiide 41k
COPD 149k
Suicide 41k
Motor vehicles 34k
Cancer 585k
Heart disease 611k
Medical error 251k

All causes 2,597k

© 2016 BMJ Publishing group Ltd.
Data source: http://www.cdc.gov/nchs/data/nvss/nvss64/nvss64_02.pdf

Analysis

Medical error—the third leading cause of death in the US

Medical error is not included on death certificates or in rankings of cause of death. Martin Makary and Michael Daniel assess its contribution to mortality and call for better reporting

Martin A Makary professor, Michael Daniel research fellow
Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

The annual list of the most common causes of death in the United States, compiled by the Centers for Disease Control and Prevention (CDC), informs public awareness and national research priorities each year. The list is created using death certificates filled out by physicians, funeral directors, medical examiners, and coroners. However, a major limitation of the death certificate is that it relies on assigning an International Classification of Disease (ICD) code to the cause of death. As a result, causes of death not associated with an ICD code, such as human and system factors, are not captured. The science of safety has matured to describe how communication breakdowns, diagnostic errors, poor judgment, and inadequate skill can directly result in patient harm and death. We analyzed the scientific literature on medical error to identify its contribution to US deaths in relation to causes listed by the CDC.

Death from medical care itself

Medical error has been defined as an unintended act (either of omission or commission) or one that does not achieve its intended outcome, the failure of a planned action to be completed as intended (an error of execution), the use of a wrong plan to achieve an aim (an error of planning), or a deviation from the process of care that may or may not cause harm to the patient. Patient harm from medical error can occur at the individual or system level. The taxonomy of errors is expanding to better categorize preventable factors and events. We focus on preventable lethal events to highlight the scale of potential for improvement.

The role of error can be complex. While many errors are non-consequential, an error can cascade the life of someone with a long life expectancy or accelerate an imminent death. The case in the box shows how error can contribute to death. Moving...
The Personalization of Medicine
What can Precision Medicine Deliver?

Patients with long-term conditions on 3+ drugs cost 5x more and have poorer outcomes.

Precision medicine can deliver:

- **48%** fewer ER visits
- **40%** fewer readmissions
- **75%** better patient reported outcomes

£225B+ /yr

compliance related savings to US healthcare

---

1. CMS.gov (2014)
### Healthcare Industry Megatrends

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcomes</strong></td>
<td>• Payment on results – more accurate diagnosis &amp; evidence of cost-effective patient benefit</td>
</tr>
<tr>
<td><strong>Precision Medicine</strong></td>
<td>• Giving the right treatments to the right patient at the right time, every time</td>
</tr>
<tr>
<td><strong>Patient Engagement</strong></td>
<td>• Helping patients &amp; carers be better informed and actively change behavior to manage their health</td>
</tr>
</tbody>
</table>
Whole Genome Analysis

Landscape of somatic mutations in 560 breast cancer whole-genome sequences

Serena Nik-Zainal1,2, Helen Davies1, Johan Staaf2, Mariana Ramakrishna1, Dominique Gliederl1, Xueqin Zou1, Imgo Martincocera1, Ludmil B. Alexandrov3,4,5,6, Sancha Martin1, David C. Wedge7, Peter Van Loon4,5, Young Suk Hsu1, Marcel Smul1, Arije B. Brinkmann1, Sandro Morgenella1, Miriam R. Auer1,10,11, Ole Christian Lingjærde11,12, Anita Lange10,11, Markus Ringнер1, Sung Min Ahn1, Sandrine Boyault1, Jane E. Brock1, Annelies Broeks8, Adam Butler1, Christine Desmedt1, Luc Dirix9, Serge Drouin9, Agnieszka Fatima1, John A. Focke1, Morditz Gerstung1, Gerrit K. J. Hooger1, Se Jin Jiang9, David R. Jones1, Kyung-Yong Kim1, Tari A. King1, Savitri Krishnamurthy14, Hee Jin Lee1, Jeong-Yeon Lee1, Vilong Li9, Stuart McLaren1, Andrew Menzies1, Ville Mustonen1, Sarah O’Meara1, Iris Paaporter9, Xavier Pivot7, Colin A. Purdie1, Keiran Rainie1, Kamna Ramakrishnan1, F. German Rodriguez-Gonzalez1, Gilles Romieu1, Analita M. Stewerts1, Peter T. Simpson7, Rebecca Shephard1, Lucy Stubbings7, Olafur A. Stefansson1, Ion Teague1, Stefania Tomasini1, Isabelle Trestle1, Gert C. Van den Eynden18,14, Peter Vermeulen1,14, Anne Vincent-Salomon1, Lucy Yates1, Carlos Caldas9, Laura van’t Veer7, Andrew Tutt13,56, Stijn Knappskog15,20, Renita Kiat Tee Tan15,42, Jos Jonkers6, Ake Borg1, Naoto T. Ueno1, Christos Sotiropoulou1, Alain Vial1,44, P. Andrew Fatmi1,45, Peter J. Campbell1, Paul N. Span1, Steven Van Loo16, Sanni R. Lakhana15,42, Jorunne E. Ejlifson1, Alastair M. Thompson16,48, Ewan Birney1, Hendrik G. Stukenberg1, Marc J van de Vijver17, John W. M. Marini1, Anne-Lise Berresen-Dale10,11, Andrea L. Richardson11,9, Gu Kong25, Gilles Thomas46 & Michael R. Stratton1

We analysed whole-genome sequences of 560 breast cancers to advance understanding of the driver mutations conferring clonal advantage and the mutational processes generating somatic mutations. We found that 93 protein-coding cancer genes carried probable driver mutations. Some non-coding regions exhibited high mutation frequencies, but most have distinctive structural features probably causing elevated mutation rates and do not contain driver mutations. Mutational signature analysis was extended to genome rearrangements and revealed twelve base substitution and six rearrangement signatures. Three rearrangement signatures, characterized by tandem duplications or deletions, appear associated with defective homologous-recombination-based DNA repair: one with deficient BRCA1 function, another with deficient BRCA1 or BRCA2 function, the cause of the third is unknown. This analysis of all classes of somatic mutation across exons, introns and intergenic regions highlights the repertoire of cancer genes and mutational processes operating, and progresses towards a comprehensive account of the somatic genetic basis of breast cancer.
**Precision Medicine Use Case**

1. **SEQUENCE**
   - patient genome

2. **IDENTIFY**
   - mutations (SNPs)

3. **ANALYZE**
   - metabolic function

4. **DECIDE**
   - best drug prescription

**World's fastest and most scalable genome association studies**

**Powerful and efficient deep semantic learning & search tools**

**Most personalized clinical decision support & digital health platform**

[precision.life]
‘Traditional’ Biomarker Discovery / GWAS

• Correlate genetic markers with disease/treatment outcomes

Source: MMG 233 2014 Genetics & Genomics Wiki
The Biological Challenge

Leading Edge Perspective

An Expanded View of Complex Traits: From Polygenic to Omnigenic

Evan A. Boyle,1,* Yang I. Li,1,* and Jonathan K. Pritchard1,2,3,*

1Department of Genetics
2Department of Biology
3Howard Hughes Medical Institute
Stanford University, Stanford, CA 94305, USA
*Correspondence: eaboyle@stanford.edu (E.A.B.), yangili@stanford.edu (Y.I.L.), pritch@stanford.edu (J.K.P.)
http://dx.doi.org/10.1016/j.cell.2017.05.038

A central goal of genetics is to understand the links between genetic variation and disease. Intuitively, one might expect disease-causing variants to cluster into key pathways that drive disease etiology. But for complex traits, association signals tend to be spread across most of the genome—including near many genes without an obvious connection to disease. We propose that gene regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes and that most heritability can be explained by effects on genes outside core pathways. We refer to this hypothesis as an “omnigenic” model.
The Computational Challenge

• Current fastest supercomputer does $3 \times 10^{16}$ operations/sec
  • At $n = 6$ would take 3.1 trillion years...
  • At $n = 10$ would take $3.6 \times 10^{38}$ seconds

\[
\frac{n! \cdot 3^r}{r! \cdot (n - r)!}
\]
## Bipolar Study Findings

<table>
<thead>
<tr>
<th>Layer (# SNPs in combination)</th>
<th># Networks</th>
<th>Features</th>
<th>% Cases (cumulative)</th>
<th>% Controls (cumulative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Rare variant homozygote</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>Rare variant homozygote</td>
<td>26% (158/607)</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>Common variant homozygote/ heterozygote</td>
<td>44% (222/607)</td>
<td>0</td>
</tr>
</tbody>
</table>

The n-SNP networks are genetically distinct and have been clinically validated.
### Genomics Data Representation

**SNP Genotype**

- **SNP Genotype:**
  - 0 = homozygous ‘normal/major allele’
  - 1 = heterozygous
  - 2 = homozygous ‘variant/minor allele’

- **SNP Index 247 = rs12345678**

<table>
<thead>
<tr>
<th>SNP Genotype</th>
<th>Case #27 Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>247</td>
<td>10 20 31 40 52 61 ... n0</td>
</tr>
</tbody>
</table>

---

**States Layer 3**

**A**

- [3, 6, 8, 9, 14, 56]

**B**

- [2, 7, 9, 10, 23]

**C**

- [1, 4, 56, 99, 113]

---

**(SNP Genotypes)**

**(Case Indices)**

*where nCases => minCases (e.g. 5 above) and nControls <= maxControls (e.g. 0)*
Synomics Example – Breast Cancer Study

14,777 People with BRCA1/2 mutations
200K SNPs per person

• All participants have BRCA 1 and/or BRCA 2 mutations
  • 3,850 affected by breast cancer (cases)
  • 10,927 non-affected (controls)

• Seeking combinations of multiple SNPs associated with:
  • disease risk
  • disease protective effect
  • therapy response
# Synomics – Transforming Genomic Medicine

14,777 People with BRCA1/2 mutations

200K SNPs per person

## Current GWAS (1,000 node supercomputer) vs synomics (on single IBM Minsky with 4x Nvidia P100)

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Current GWAS</th>
<th>synomics (on single IBM Minsky with 4x Nvidia P100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 SNPs</td>
<td>$10^{11}$</td>
<td>6-8 months</td>
</tr>
<tr>
<td>6 SNPs</td>
<td>$10^{32}$</td>
<td>-</td>
</tr>
<tr>
<td>17 SNPs</td>
<td>$10^{84}$</td>
<td>-</td>
</tr>
</tbody>
</table>
Synomics - Breast Cancer (BRCA1)
# BRCA2 affected / non-affected

<table>
<thead>
<tr>
<th>False Discovery Rate (%)*</th>
<th>SNP Genotypes</th>
<th>Cases</th>
<th>Penetrance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2,113</td>
<td>799</td>
<td>50.7</td>
</tr>
<tr>
<td>10</td>
<td>1,320</td>
<td>627</td>
<td>39.8</td>
</tr>
<tr>
<td>5</td>
<td>868</td>
<td>513</td>
<td>32.6</td>
</tr>
<tr>
<td>1</td>
<td>142</td>
<td>221</td>
<td>14.0</td>
</tr>
</tbody>
</table>

* Using Benjamini-Hochberg correction for multiple testing
Synomics Example – Breast Cancer Study

Key Findings:

Found co-occurring sub-clusters of 3, 4, 5, or 6 SNP variants
  • co-occurring in later layers of analysis (8 SNPs+)
  • SNPs associated with same pathways show disease functional units
  • opportunity to identify combinatorial therapies

Detected 17 SNP networks in up to 103 cases and 0 controls
Very high (>25%) penetrance for good clinical relevance

Identified disease protective & disease risk associated factors
  • BRCA1/2 status may suggest risk, but other variants in combination confer an overall greater protective effect

Currently analysing phenotypic and clinical features
Genomics Data Representation

Merged Networks (based on sharing at least x% SNP genotypes)

higher nCases with lower nSGs and high densities is better, i.e. a small number of highly interconnected SNP genotypes
Real World Personalisation Challenges

PERSONALISED MEDICINE 101

- Clinical status
- Phenotype
- Co-morbidities
- Co-prescriptions
- Lifestyle & environment
- (Clonal) heterogeneity
- Polygenic disease aetiology

More accurate assessment of diagnosis and response to treatment. Molecular profiling is used to determine the appropriate therapy.

Adapted from: sarahcannon.com
• 15,000 MND patients / 7,500 controls
  • 40% whole genomes sequenced
  • 2 petabytes

• Multi-factor late-onset disease
  • Only 5-10% genetically determined heritability
  • 6 independent factors required to trigger disease
  • Imaging, epigenetics, lifestyle, diet, environment, clinical history, co-morbidities
Biological Interpretation
Biological Annotation

• **Query:** rs3734805 rs9383935 rs9383589 c6_pos151989450 rs4648881 rs9383936 CD14+CD16- monocyte CD8+/ab T fetal thymus naive B cell
  - Full context of all cell types in which epigenetic activation occurs
  - Literature search (keywords) gave no relevant results (Google/PubMed)
  - Deep semantic search identified 36 relevant papers including:
    “In vivo modulation of the distribution of thymocyte subsets: effects of estrogen...”

• Further queries identified a study where female infants with enlarged thymus treated with X-rays were observed to have higher incidence of breast cancer 36 years later
• Suggested novel disease sensitization mechanism
“Transient involution of the maternal thymus in mice is known to occur during pregnancy. Although estrogen crosses the placenta, fetal thymus gland enlarges with advancing gestational age. It is not known if fetal thymocytes are resistant to estrogen or if there are other factors that prevent estrogen from exerting an effect on the development of fetal thymocytes. Therefore we studied the effect of estrogen on isolated fetal thymic glands in vitro. All CD4 and CD8 defined T cell subsets were reduced with a disproportionate loss of CD4+ single positive (SP), CD8+ SP: CD4+CD8+ double positive (DP) cells.”
spot.my GPU enabled semantic search

* scraping & indexing infrastructure is fully scalable and distributed
spot.my automatically scales to meet demand and is fully monitored with failure alerts
spot.my GPU enabled semantic search

- Enables very fast searching of large corpora & vocabs with low RAM/CPU

STAGE 1 (cluster articles)
Identify clusters of similar stories and orphans (non-clustered)
27M papers > 5M clusters + 3M orphans
GPU with fast GRAM

STAGE 2 (cluster search)
Find closest clusters and orphans
GPU with fast GRAM

STAGE 3 (results refinement)
Full search of selected clusters to find hits
CPU & cheap RAM
other spot.my features

use keywords or whole paragraphs to search

find relevant papers even if they use different words

‘like this’ - drag & drop whole papers as queries

create subject channels and like/dislike papers to refine

iterate searches to get even better matches
Creating New Opportunities

Much deeper insight into complex diseases

- Novel (patentable) R&D / combinatorial interventions

Includes genotypic, phenotypic and clinical data

- Clinical trials design / patient stratification
- Healthcare analytics / service planning

Use of biomarker clusters in clinical decision support:

- Personalized disease risk scoring and therapy selection
- Personalized dietary and lifestyle advice
Example - Meet Albert

- High cholesterol, asthma, high blood pressure, atrial fibrillation & gout
- Simvastatin, symbicort, bisoprolol fumarate, coumadin & naproxen

- What side-effects might he expect?
- When should he call his GP?
- What’s safe/good for him to eat?
Adverse Drug Reactions

2,500 common drugs
15,000 dosage forms

500,000 drug-drug interactions
10,000 drug-disease interactions
2,000 drug-food interactions

but... these are just the first-order interactions. Our precision.diet API (built on RACE) provides fully personalized advice considering all combinatorial interactions.
Analysis of Complex Systems

- Problem space $\approx m^n$
  $m=$ no. of states $n=$ no. of dimensions

- Easy to get problem spaces of $10^{1000}$

- Searching uses short cuts
  - AI/ML, neural nets, GA

- Short cuts miss things and may still require huge CPU/RAM

- RACE Array Logic (tensor algebras) offers provably complete computation quickly using very low CPU/RAM
  - $10^{1000}$ options > 10 hits in ms
RACE Platform

1. Scalable
   very large multi-dimensional system models

2. Complete
   including all constraints in all dimensions to ensure logical consistency

3. Compact
   complete, yet compact representations of complex systems

4. Real-time
   provably complete deduction in real time even on low power devices

precision.life/race
patented (US 6,633,863 / EU 1,062,603) and patents pending
Case Study: Danish State Railways

Engineering design and verification

• Verification of railway interlocking systems (track, points, signals...)
• 12,000+ variables and a state space with $>10^{300}$ combinations
Railway Safety Problem

• **ALL** constraints (physical and logical) must be taken into account to ensure safe and economic operation

• Even small local changes, e.g. a new position of a signal or addition of new points, requires complete validation

• Manual validation of new signal interlocking systems took at least 2 man-years
RACE Solution

• Track topology and connected objects are defined from a CAD tool
• All valid states (i.e. which won’t lead to accidents) are determined via constraint resolution, giving a provably complete system state model
  • Entire Danish railway system = 26KB
• Objects functions added to optimize costs and operational efficiency
• 2 man-years validation -> 10ms on a mobile phone
Interactions KGraph/KModel

- 500,000 drug-drug interactions
- 10,000 drug-disease interactions
- 2,000 drug-food interactions

Compilation (20 secs)

RACE Engine

Interactions KModel
Interactions KModel

185,000 Food Products (ingredients, macro/micro nutrients, allergens)

HealthySwaps App

Personal Profile

Suggested alternative products that are known to be compatible with the user

Online Shopping Basket

Precision.life
Example – precision.diet

Connect to online shopping basket, use in-store or at home via barcode scanner

Identify food items that are incompatible with your prescriptions, diseases & health goals, and understand risk levels

Choose a healthier alternative from same category, all on your own phone with no sharing of your data
Better Tools for Healthcare

- Clinical and patient decision support tools
- At the point of care / in day-to-day life
- Using full power of complex, multi-trait knowledge models
- Improving patient outcomes
- Lowering the cost of care provision
WE'RE ALL GOING TO NEED BETTER JETPACKS

with apologies to TOM GAULD