The Elizabeth H.
and James S. McDonnell III
McDonnell Genome Institute
at Washington University

@obigriffith griffithlab.org

Artificial Intelligence in Genomic Medicine

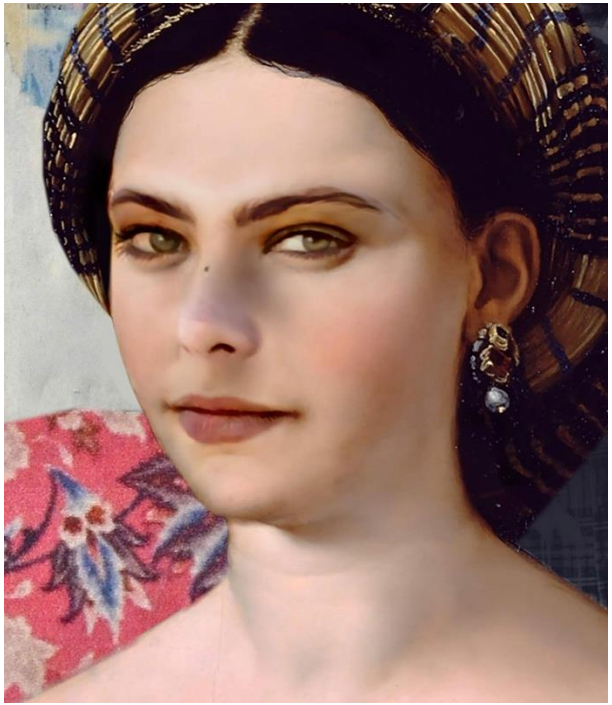
Obi L. Griffith <obigriffith@wustl.edu>

What is Artificial Intelligence?

- Artificial Intelligence (AI) is the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and other complex tasks

Examples of advanced AI appearing throughout society

- Manufacturing robots
- Automated financial investing
- Social media monitoring
- Smart assistants
- Conversational bots
  - ChatGPT



Many others…

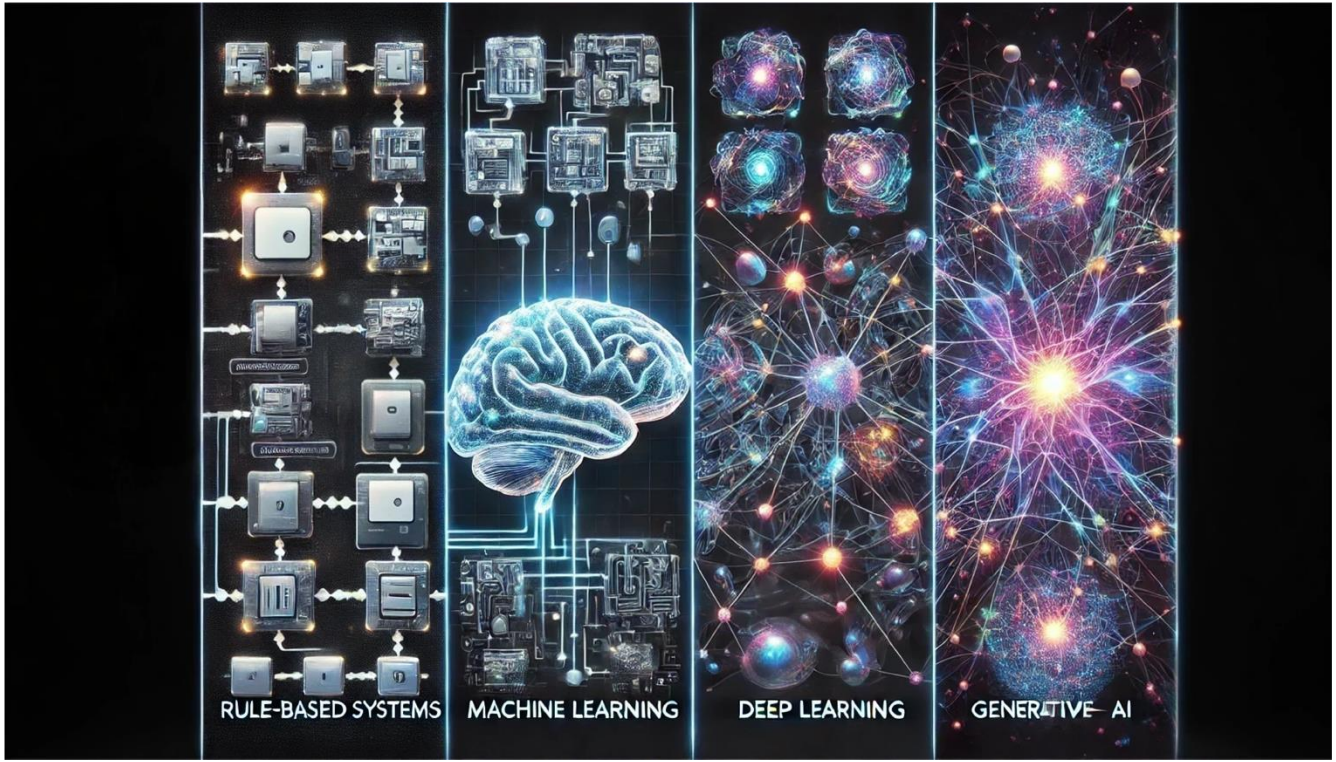
Applications in biomedical field far too many to list

- Medical image analysis
  - MRI segmentation
  - Histopathological cancer classification
  - Mammographic lesion detection
- Protein structure prediction
- Genome analysis
  - Gene expression inference
  - Enhancer prediction
  - Variant pathogenicity prediction

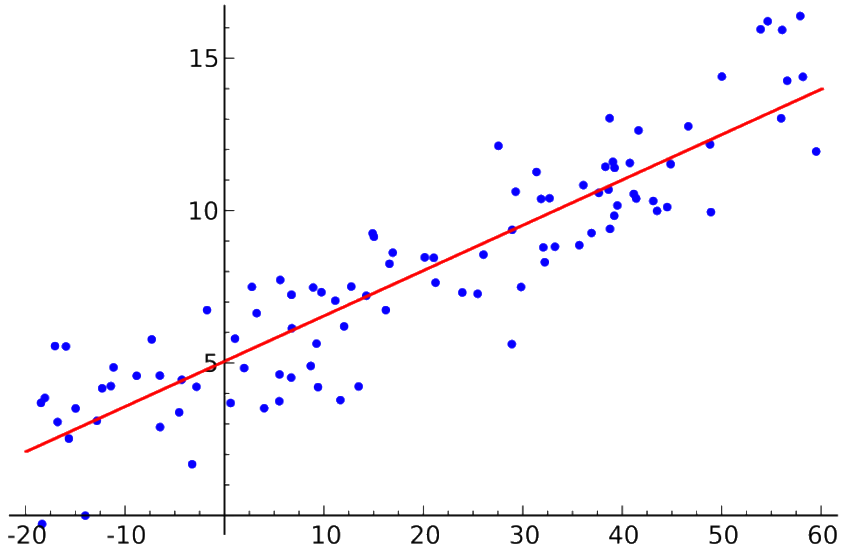
What are the key challenges of AI?

- Key Challenges of AI: knowledge representation, natural language processing (NLP), learning, perception, planning, reasoning, motion, social intelligence and general intelligence
  - Knowledge representation – catalog concepts and relationships using knowledgebases, ontologies, and structured data, for computational use
  - Natural language processing (e.g., text-mining) – allow computers to read and understand human language
  - Learning (e.g., machine learning) – develop computer algorithms that improve automatically through experience
    - Unsupervised – Find novel patterns in data without human input or labels
    - Supervised – Predict unknowns from labeled knowns (e.g., classifiers)

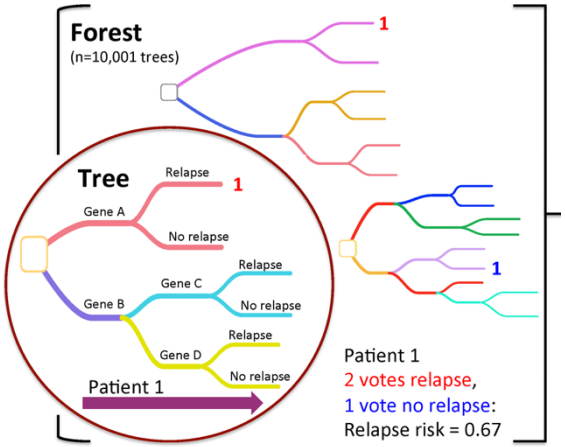
AI has gone through several eras



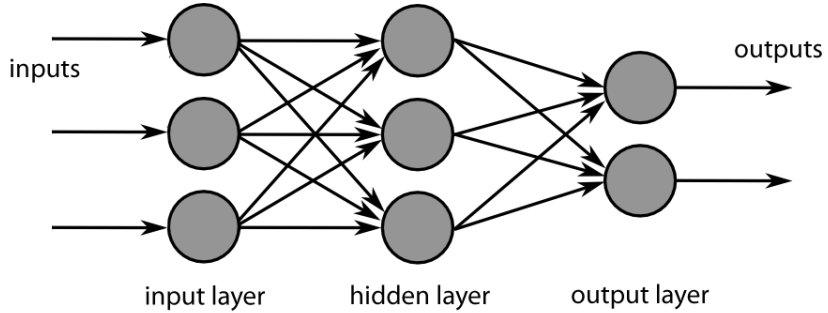
Learning -> Machine Learning -> Deep Learning



Linear regression



Decision trees, random forests, support vector machines (SVM), ...



multi-layered artificial neural networks

General considerations

- Establish training and testing datasets
  - Are there batch effects to consider?
  - Randomly split data
    - Make sure balanced for key features (sex, target class, etc)
  - Identify independent test sets
- Define target classes (e.g., Relapse vs No Relapse)
  - How is relapse defined?
  - Are the classes equal sizes? If not, how to mitigate?
- Are there features which should be filtered out?
- Are there new features we wish to design/engineer?
- Is the model overfit/overtrained?

Let's walk through an example

Griffith *et al. Genome Medicine* 2013, **5**:92
http://genomemedicine.com/content/5/10/92

Genome Medicine

RESEARCH Open Access

A robust prognostic signature for hormone-positive node-negative breast cancer

Obi L Griffith[1,2*], François Pepin[1,3], Oana M Enache[1], Laura M Heiser[1,4], Eric A Collisson[5], Paul T Spellman[1,6*] and Joe W Gray[1,4*]

https://www.biostars.org/p/85124/

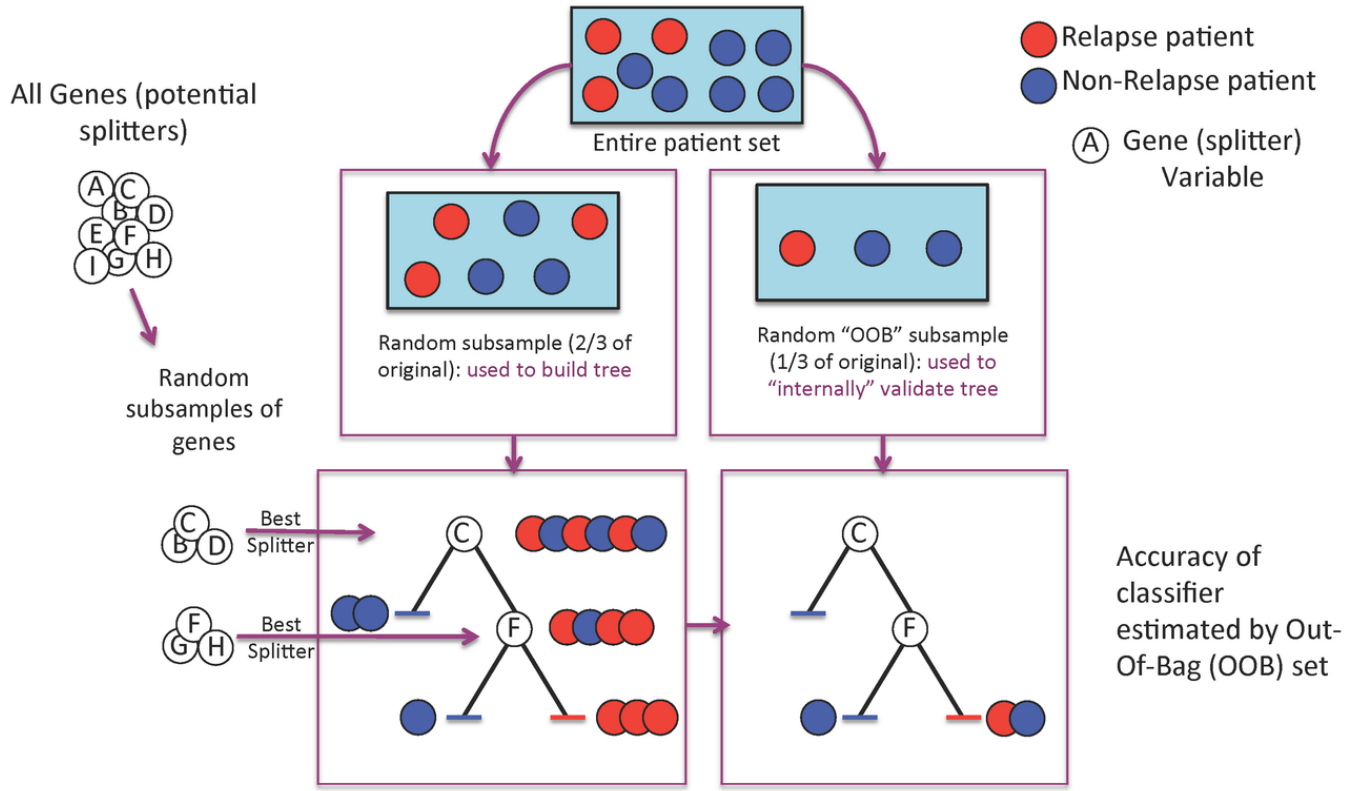
Nick Spies

Let’s walk through an example

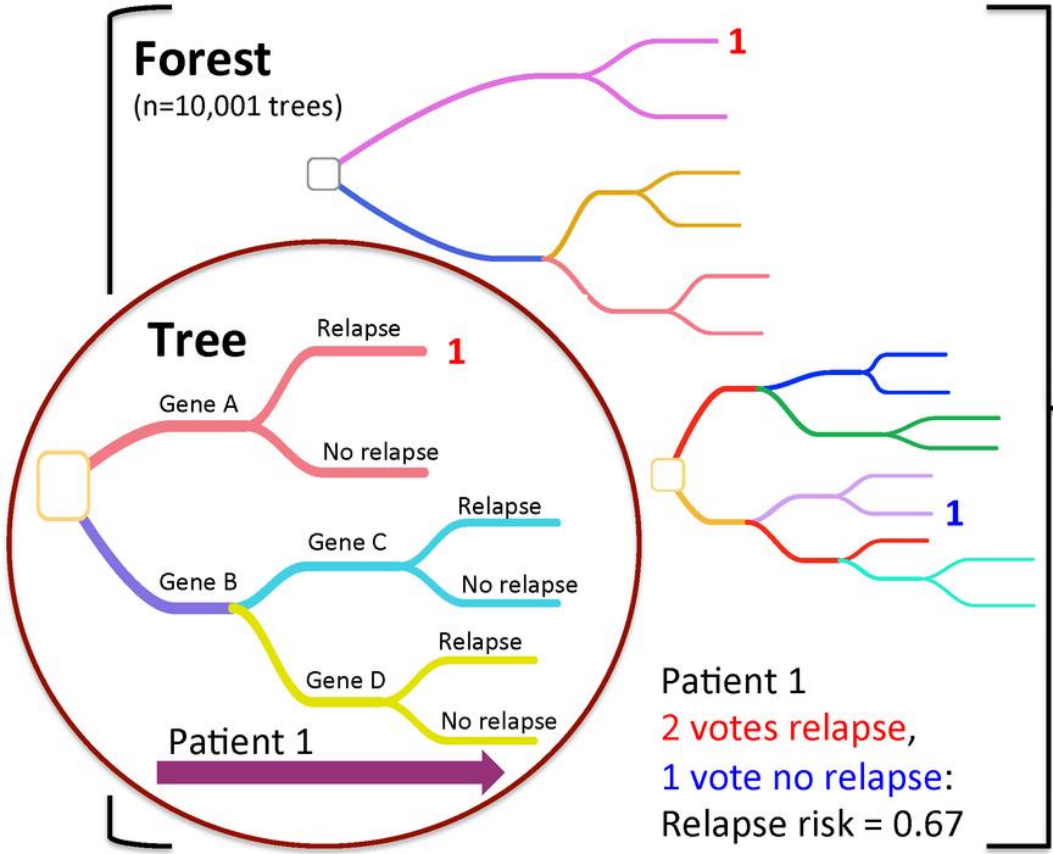
- Let’s build a classifier of relapse in ER+ breast cancer
- Suppose we have a gene expression dataset for a set of patients, with known relapse status

	Relapse	Relapse	No Relapse	No Relapse
	Patient1	Patient2	Patient3	Patient4
GeneA	4.7	5.2	1.1	0.9
GeneB	0.5	0.7	6.8	7.1
GeneC	2.3	3.1	2.4	3.2

How do you build (train) a Random Forest model?



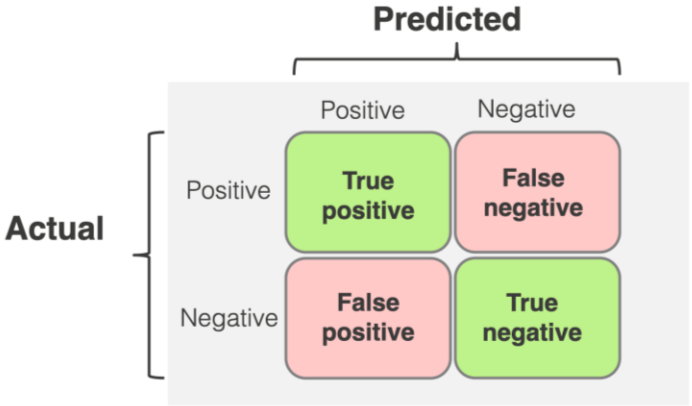
How do you predict with a Random Forest model?



- **Creates a forest of many binary decision trees**
- Each Patient traverses each tree until it reaches a terminal node
- At the terminal node each tree casts a vote (eg. “relapse”); the proportion of relapse votes from all votes is that patient’s predicted relapse risk

Class with most votes is assigned

How does our model perform? Let's look at the dreaded confusion matrix



Predicted

Actual	Relapse	NoRelapse
Relapse	49	58
NoRelapse	39	140

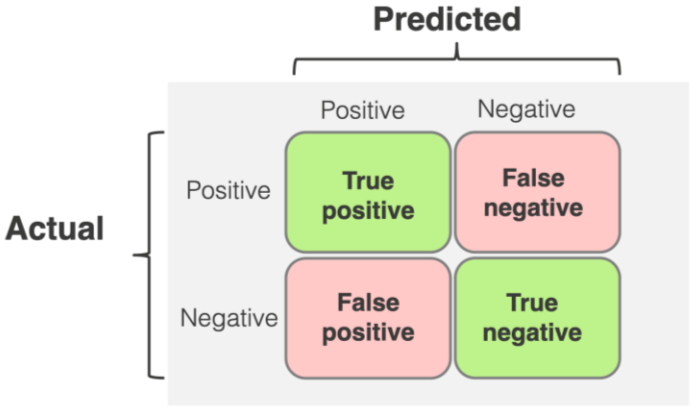
Sensitivity/Recall (true positive rate) is how well a test can identify true positives

49/(49+58) = 0.458

Specificity/Selectivity (true negative rate) is how well a test can identify true negatives

140/(39+140) = 0.782

How does our model perform? Let's look at the dreaded confusion matrix



Predicted

Actual	Relapse	NoRelapse
Relapse	49	58
NoRelapse	39	140

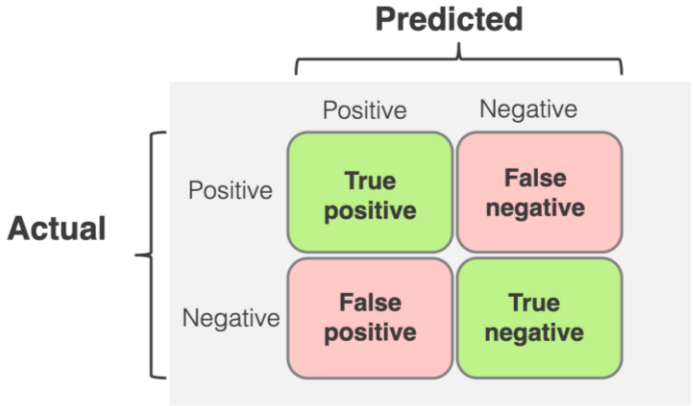
Relapse class error rate is how many true Relapses were miscalled

58/(49+58) = 0.542

No Relapse class error rate is how many true No Relapses were miscalled

39/(39+140) = 0.218

How does our model perform? Let's look at the dreaded confusion matrix

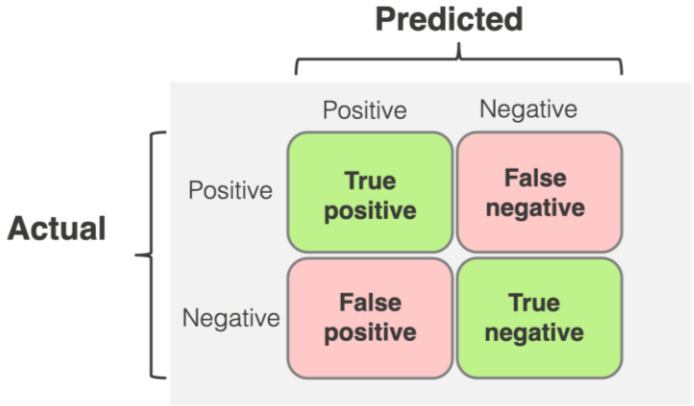


		Predicted: Relapse	Predicted: NoRelapse
Actual	Relapse	49	58
	NoRelapse	39	140

Overall accuracy (49+140)/(49+58+39+140) = 0.660

Overall error (39+58)/(49+58+39+140) = 0.340

How does our model perform? Let's look at the dreaded confusion matrix



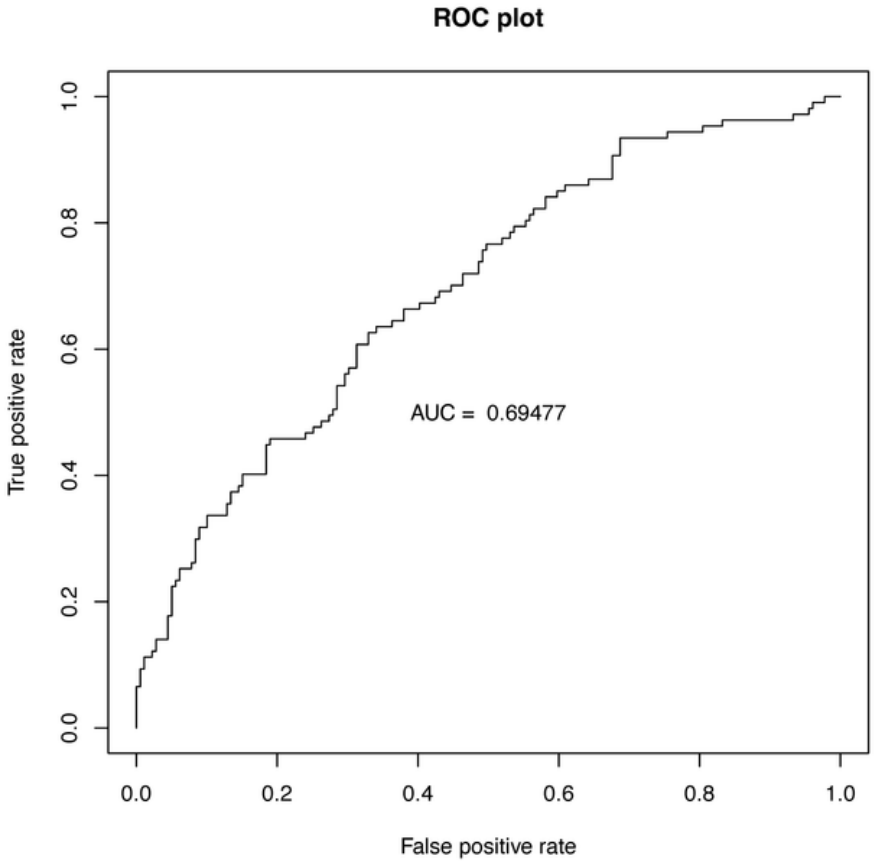
Actual \ Predicted	Relapse	NoRelapse
Relapse	49	58
NoRelapse	39	140

Precision is the share of true positive predictions in all positive predictions

$49/(49+39) = 0.557$

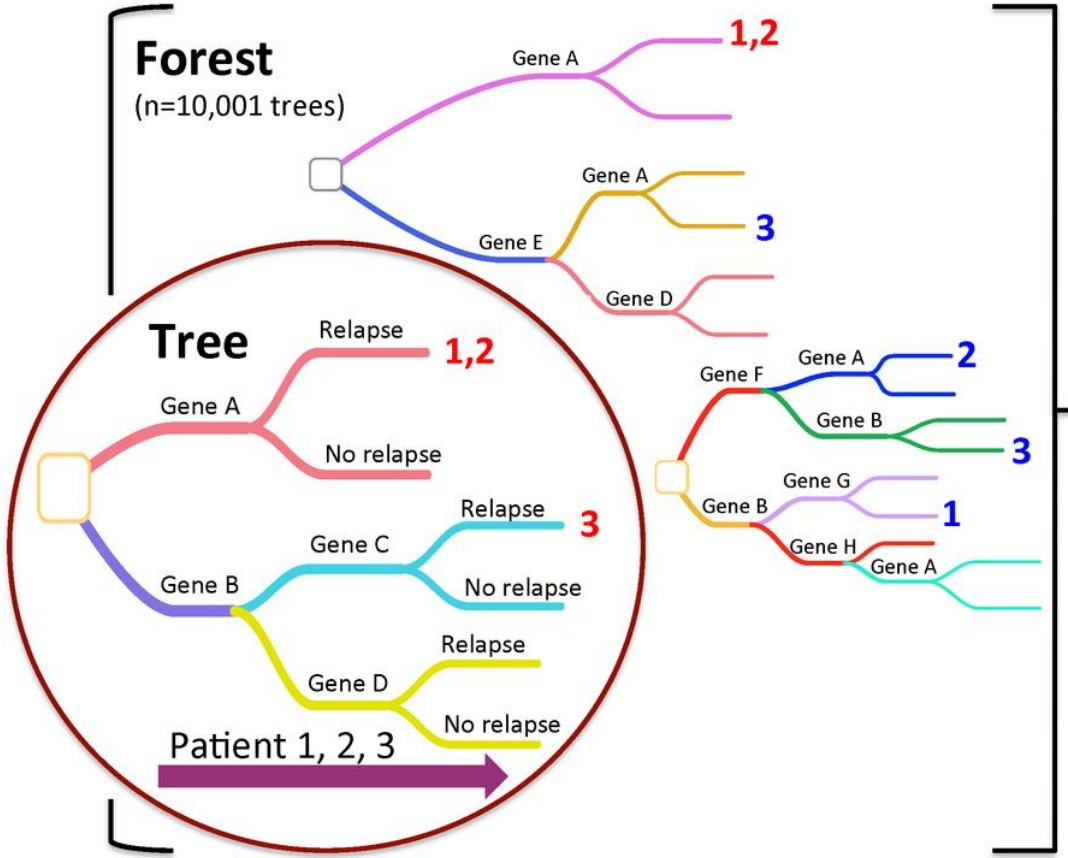
Other popular metrics: F1 score, PPV, NPV, etc

Receiver Operating Characteristic (ROC) plots TPR (Sensitivity) vs FPR for all possible model thresholds



The ROC AUC summarizes how well the model can distinguish between positive and negative classes

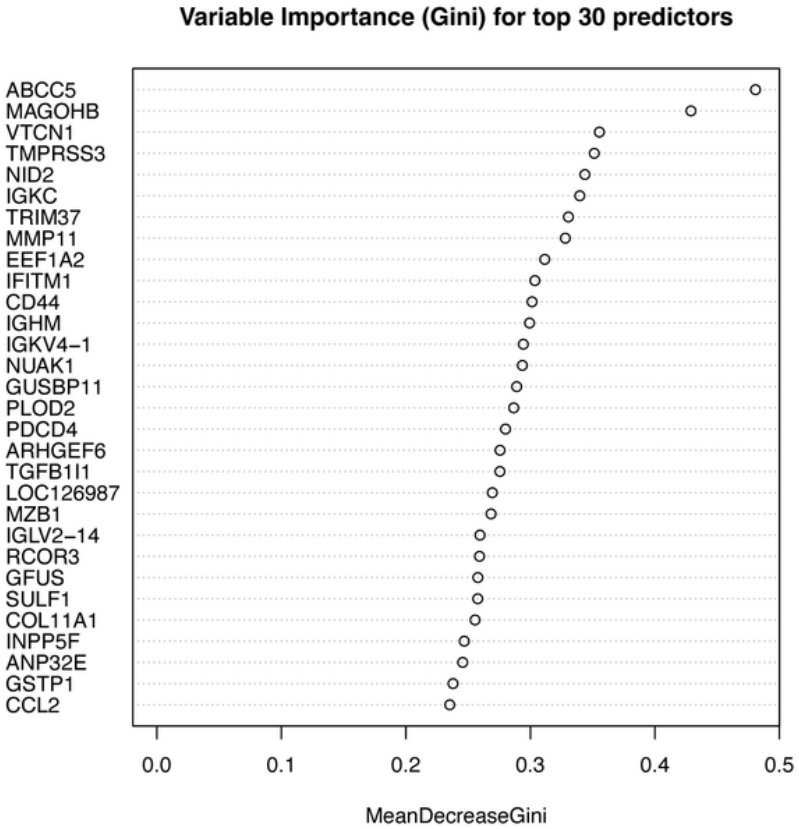
How do we know which features are important in our model



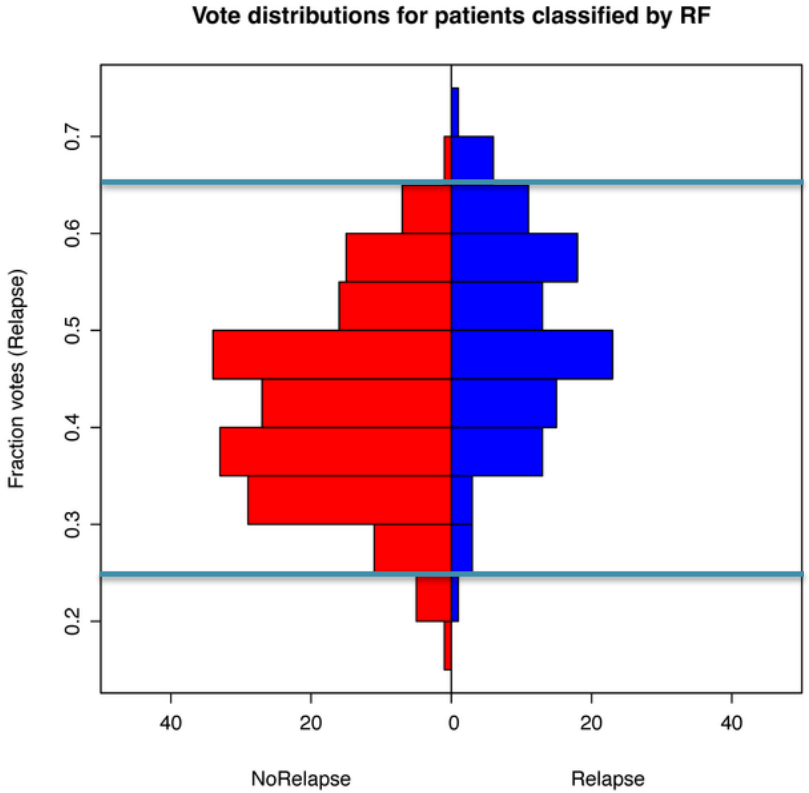
- The more often a gene is chosen as a splitter variable, the higher its "Variable Importance" – This can be used to prioritize which genes to select for an assay with limited gene measurements

Gene	Var. Imp.
Gene A	0.67
Gene B	0.20
Gene D	0.13
…	…

Variable importance could be used to prioritize genes for a targeted assay

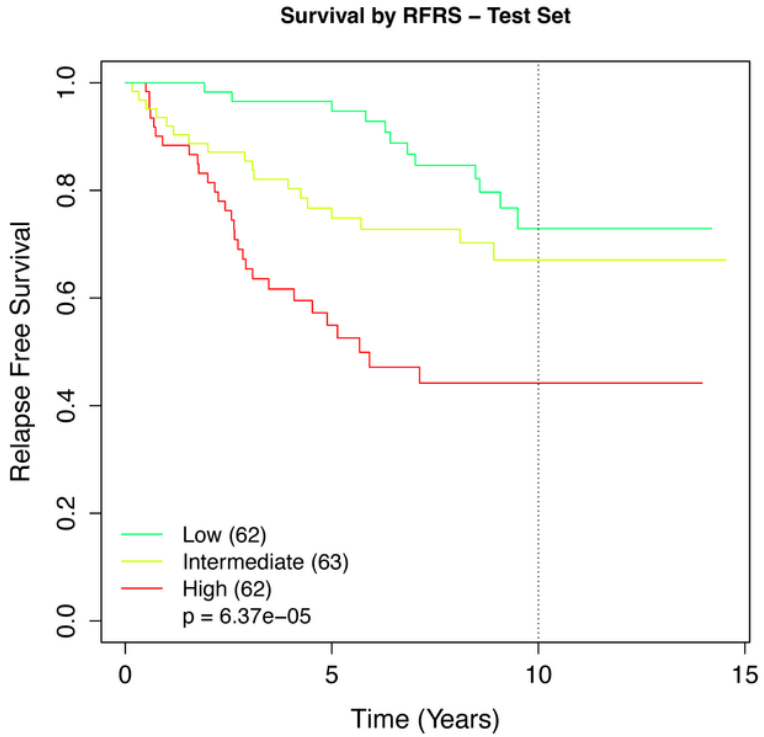


Even a model with poor overall performance may be very accurate for a subset of cases



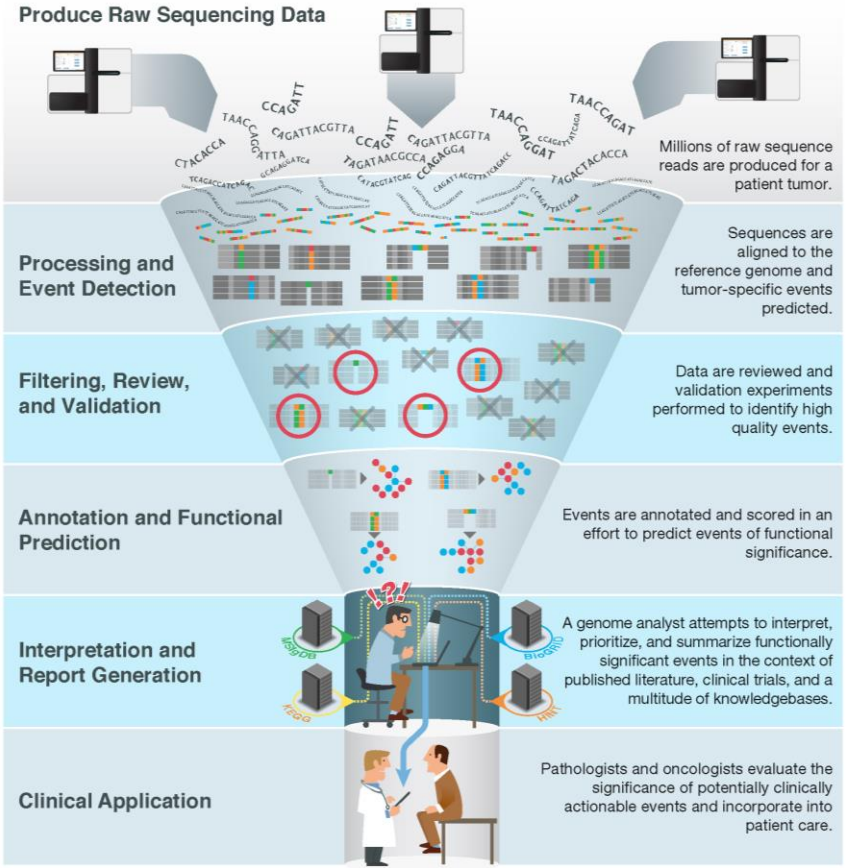
high-risk, intermediate-risk and low-risk groups can be defined

These risk groups can be related back to survival



This is how essentially how Oncotype DX, MammaPrint, and ProSigna assays were developed – used routinely in breast cancer care

How else can AI be applied to genomic medicine?



Variant calling

Manual review

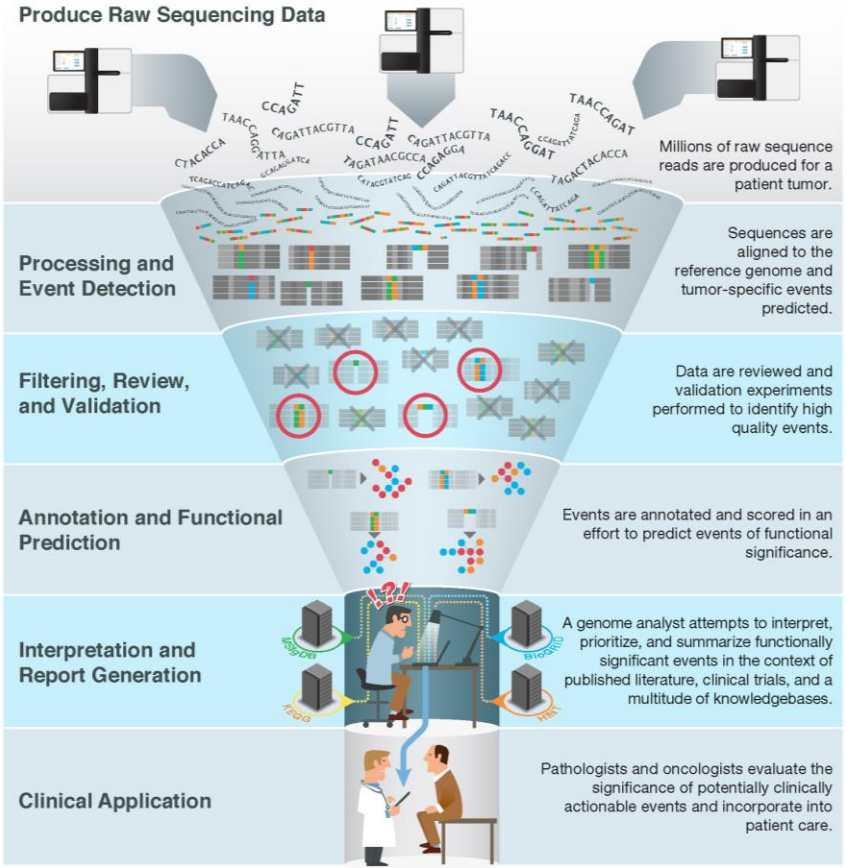
Functional prediction

Clinical Relevance

Decision-making

Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.

How else can AI be applied to genomic medicine?



Variant calling

Manual review

Functional prediction

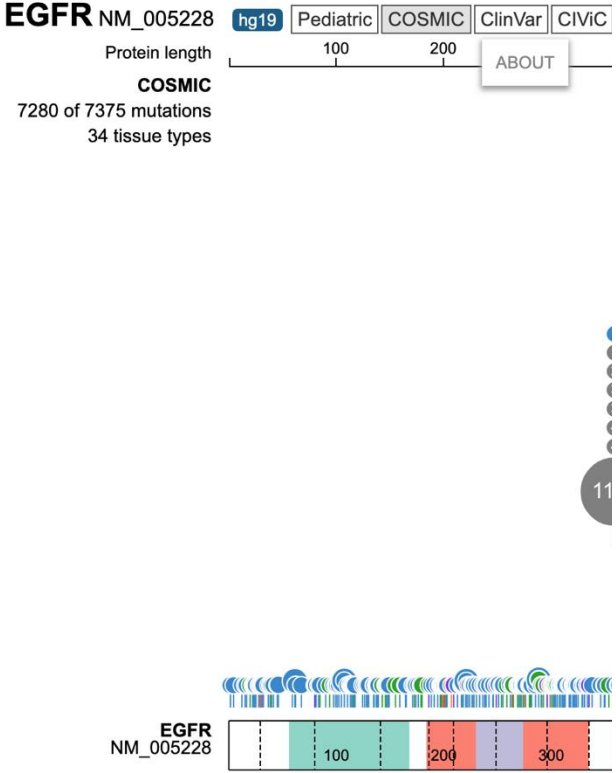
Clinical Relevance

Decision-making

Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.

I) Manual review – how do we know a cancer variant is real?

Variant calling from raw sequence data is a critical step in most cancer genomics research and clinical laboratory genomics workflows



Genomic Alterations

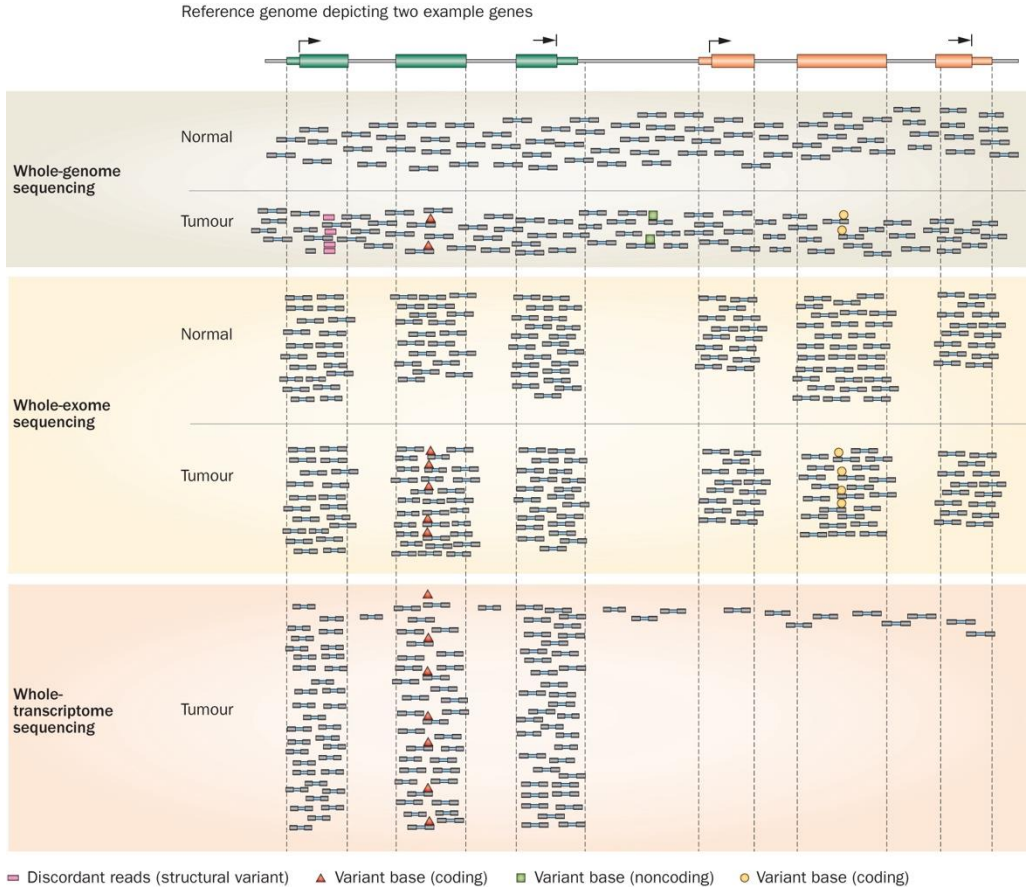
Gene Alteration	Interpretation
PIK3CA H1047R	Mutations in PIK3CA have been reported in 26% to 33% of breast cancer cases (COSMIC, Jun 2012 and Kalinsky et al., 2009; 19671852). Activating mutations in PIK3CA, such as the one seen here, may predict sensitivity to inhibitors of PI3 kinase or its downstream signaling pathway (the PI3K/Akt/mTOR pathway) (Huang et al., 2007; 18079394). The mTOR inhibitors temsirolimus and everolimus have been tested in several clinical trials in breast cancer, and have been approved by the FDA for use in other tumor types. Inhibitors of PI3K and Akt are currently in clinical trials in breast cancer, alone or in combination with other therapies. PIK3CA mutations may play a role in resistance to hormonal therapy in ER+ breast cancers (Miller et al., 2011; 22114931). Activating mutations in PIK3CA may also confer resistance to anti-Her2 therapies (Chakrabarty et al., 2010; 20581867, Kataoka et al., 2010; 19633047, Wang et al., 2011; 21676217); combined inhibition of Her2 and the PI3K pathway may be required in tumors with ERBB2 amplification and PIK3CA mutation, though this remains an area of active investigation.
CCND1 amplification	CCND1 amplification has been reported in approximately 10-15% of invasive breast cancers, more frequently in BRCA-negative cancers (Elsheikh et al., 2008; 17653856, Bane et al., 2011; 21327470). There are no approved therapies that directly target the protein product of CCND1 (Cyclin D1); however, CCND1 amplification may predict sensitivity to inhibitors of Cdk4 and Cdk6, which are currently under investigation in clinical trials. Overexpression of Cyclin D1 has also been associated with resistance to endocrine therapy in breast cancer (reviewed in Lange et al., 2011; 21613412; Musgrove and Sutherland, 2009; 19701242, Butt et al., 2005; 16113099).
CDH1 E167*	CDH1 mutations are present in approximately 17% of breast cancers, and more often in luminal type cancers (COSMIC, Jun 2012, Hollestelle et al., 2010; 19593635). Loss of the E-cadherin protein, which is encoded by the CDH1 gene, has been associated with poor prognosis in triple negative breast cancer (Kashiwagi et al., 2010; 20551954, Tang et al., 2011; 21519872). Presently, there are no targeted therapies to address loss of CDH1/E-cadherin.

Whole genome, exome and transcriptome sequencing allows us to detect and confirm many different types of ‘omic events

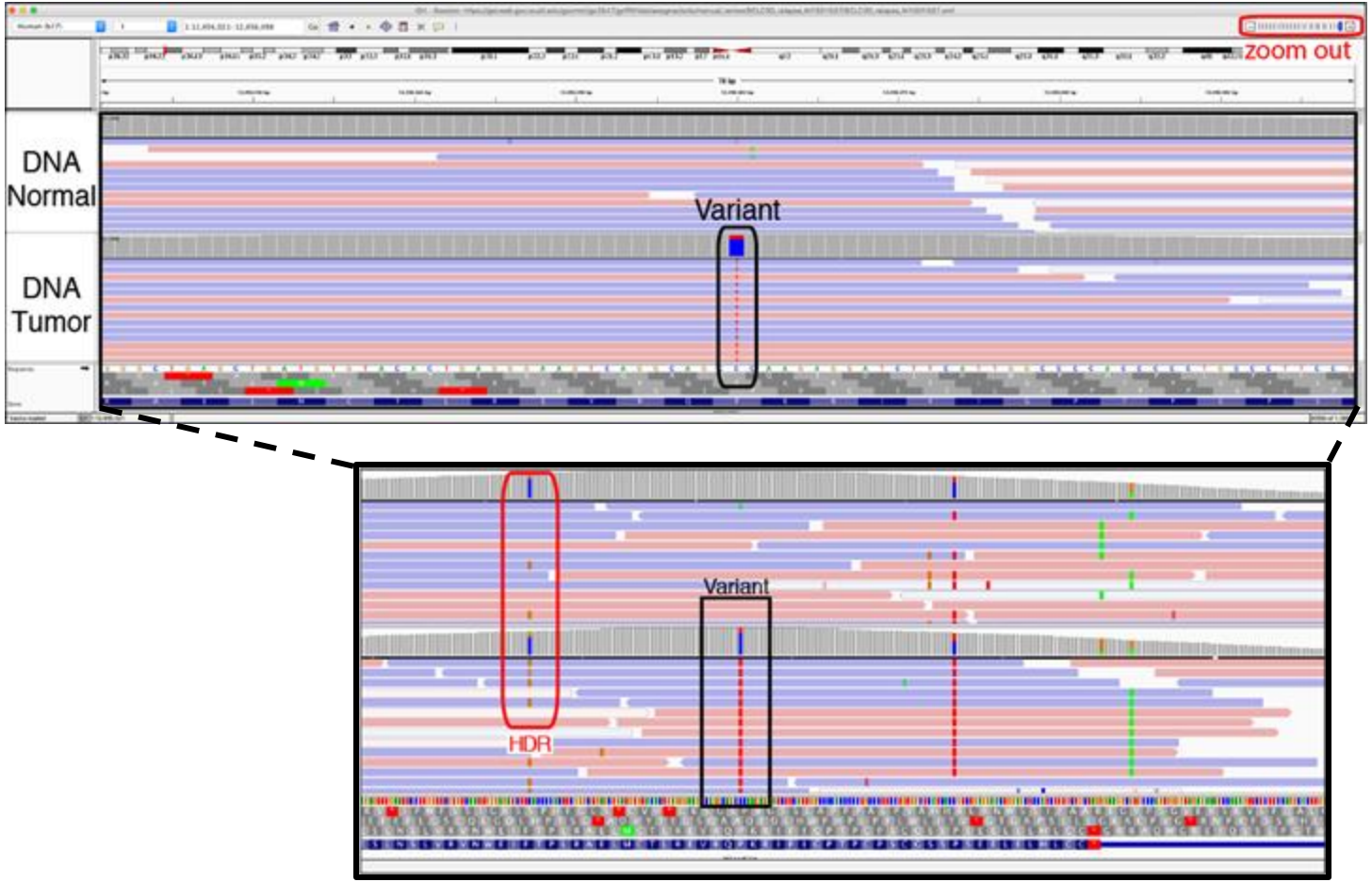
‘WGS’

‘Exome’

‘RNA-seq’



Manual review involves human inspection of sequence alignments to identify artifacts missed by automated variant callers



Literature is full of references to manual review and ad hoc filtering

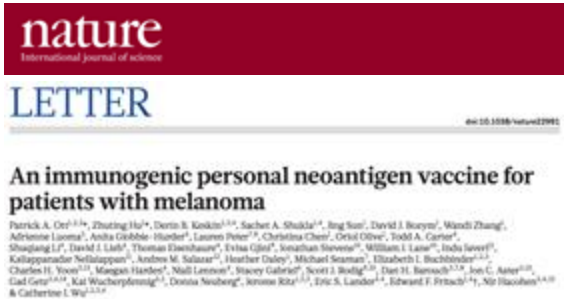
“Mutations...were called with MuTect and filtered with oxidation and panel of normal samples filters to remove artefacts...”


nature
International journal of science

ARTICLE

Recurrent and functional regulatory mutations in breast cancer

“All indels were manually reviewed in Integrative Genomics Viewer.”


nature
International journal of science

LETTER

An immunogenic personal neoantigen vaccine for patients with melanoma

- Rarely are the methods/criteria for manual review described
- Definitely not standardized

“Coverage of at least 50 reads in both tumor and normal samples, >20% of reads supporting the variant in tumor samples, and <5% of reads supporting the variant in normal samples.”


nature genetics

RNF43 is frequently mutated in colorectal and endometrial cancers

Manual review is a largely undocumented source of error, suffers from inter-reviewer variability, and requires significant resources

- We completed a study of independent/blinded manual review of a set of variants.
- High rates of inter-reviewer variability
  - Only ”fair agreement” – Kappa = 0.37
  - 77.3% good (3/3 reviewers) or acceptable (2/3 reviewers) agreement on variant calls
- Manual review is labor-intensive, time-consuming
  - ~70-100 variants per hour

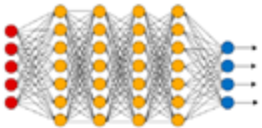
Can manual review be automated with AI?

Training Data

Array of 71 features obtained from paired tumor & normal BAM files

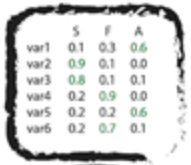
Model Development

Build random forest, deep learning, and linear reg models



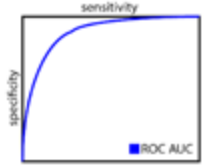
Testing Data

Array of 71 features obtained from paired tumor & normal BAM files



Predictions

Array of 3 predictions for Ambiguous, Fail, and Somatic calls.



Compare Predictions

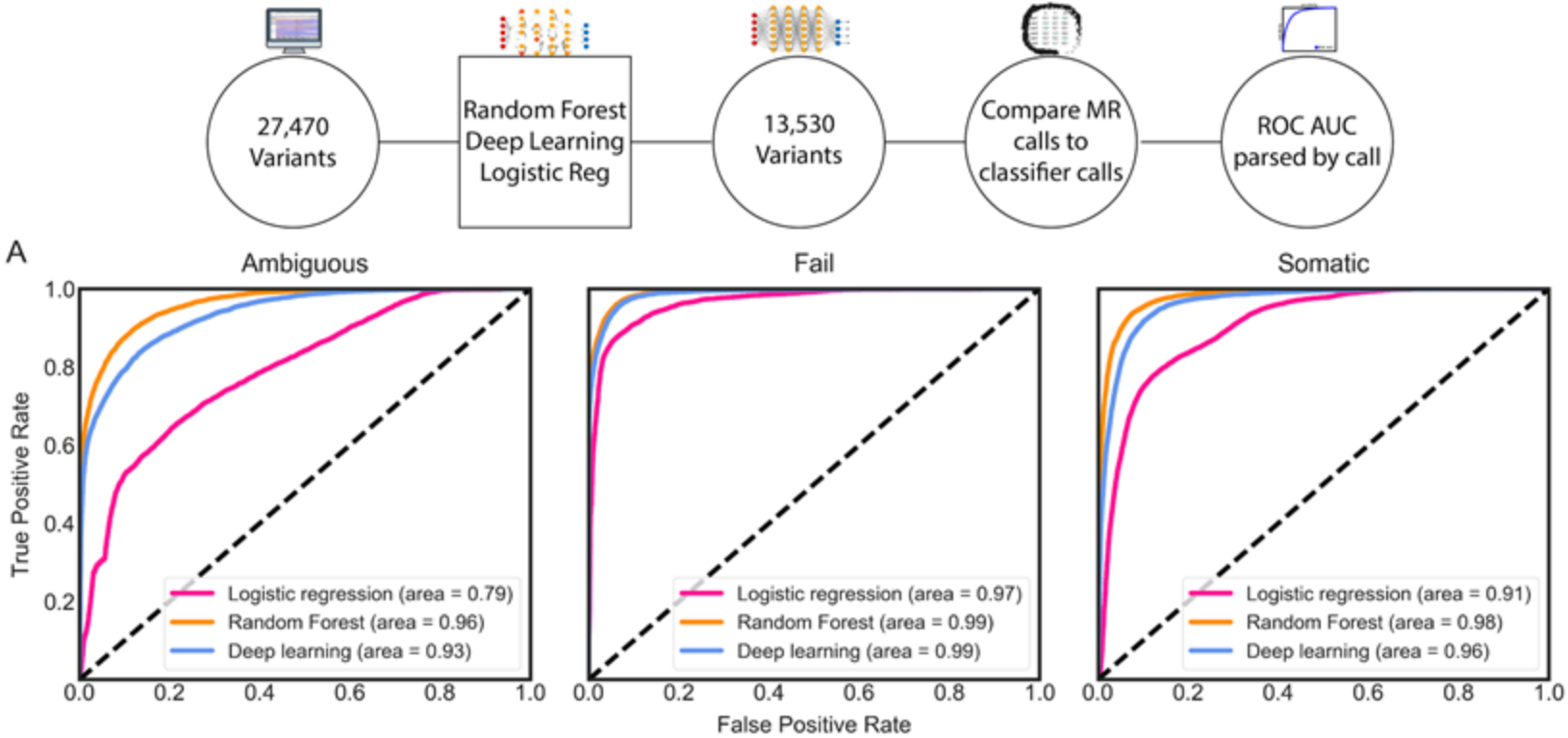
Assess sensitivity and specificity based on tuning model output

Assembled database of 40,000+ manually reviewed variants

	Training Set (n=27,470)	Testing Set (n=13,530)	Total (n=41,000)
Malignancy (410 cases)			
Leukemia (243 cases)	5,815	2,877	8,692
Lymphoma (23 cases)	1,263	628	1,891
Breast (135 cases)	8,986	4,320	13,306
Small Cell Lung (18 cases)	9,177	4,601	13,778
Glioblastoma (17 cases)	844	412	1,256
Melanoma (1 cases)	185	100	285
Colorectal (1 case)	842	419	1,261
Gastrointestinal Stromal (1 case)	70	31	101
Malignant Peripheral Nerve Sheath (1 case)	288	142	430
Sequencing Methods			
Capture Sequencing	9,479	4,755	14,234
Exome Sequencing	9,367	4,677	14,044
Whole Genome Sequencing	8,624	4,098	12,722
Variant Calls			
Somatic	12,266	6,115	18,381
Ambiguous	7,189	3,454	10,643
Fail	5,909	2,945	8,854
Germline	2,106	1,016	3,122

- Nine different tumor subtypes tested
  - 266 hematologic tumors and 144 solid tumors
- Multiple sequencing platforms tested
- Even distribution of variant calls (Somatic, Germline, Ambiguous, or Fail)
- Features used by manual reviewers extracted from alignments

Highly accurate (ROC AUC > 0.9) predictive model



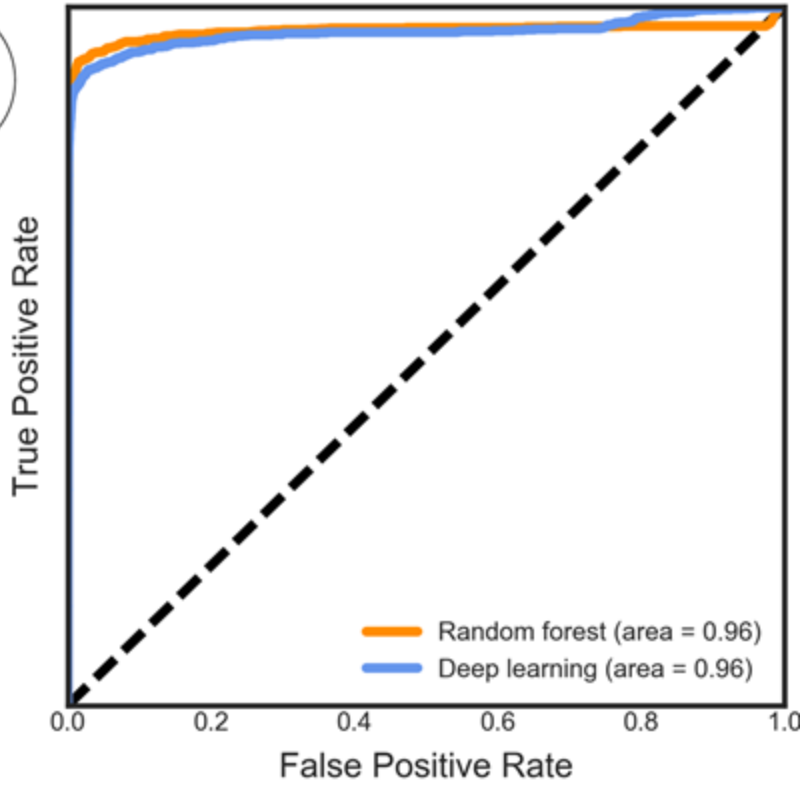
- **Random Forest and Deep Learning models out-perform Logistic Regression model**
- **Ambiguous calls are the most difficult to classify**

Excellent performance on Orthogonally validated variants

Orthogonal Validation Strategy:

- 192,241 variants were called by automated somatic variant callers.
- 1,343 true positives and 190,898 false positives were identified by 1000X sequencing
- Model predictions were compared to orthogonal sequencing results

Use of orthogonal sequencing data to classify variants performed exceptionally well

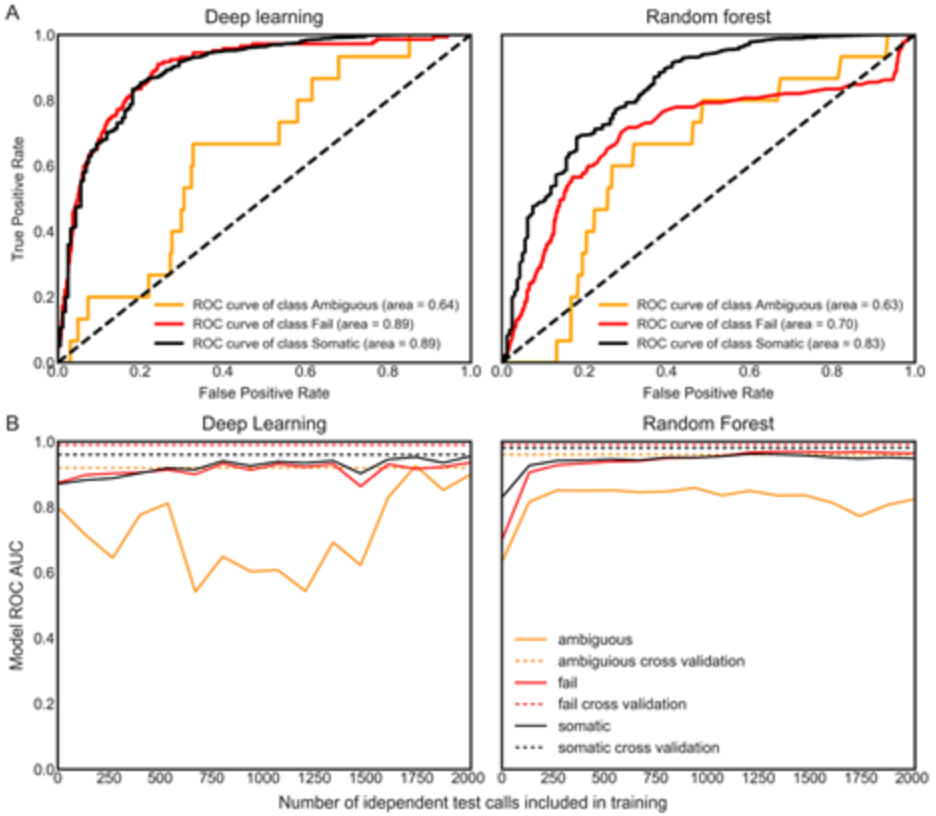


Performance is reduced on external datasets but can be recovered with a modest amount of re-training

"External Lab" Validation Strategy:

- Analyze performance on 4 SCLC cases with different manual reviewers, different sequencing platform, and different automated variant callers.
- Initial AUC showed suboptimal performance.
- Model was re-trained with 5% of manual review calls from new sequencing data.

Retraining the data with ~200 variants shows exceptional recovery of model performance.



nature genetics

TECHNICAL REPORT

https://doi.org/10.1038/s41588-018-0257-y

A deep learning approach to automate refinement of somatic variant calling from cancer sequencing data

Benjamin J. Ainscough[1,2,12], Erica K. Barnell[1,12], Peter Ronning[1], Katie M. Campbell[1], Alex H. Wagner[1], Todd A. Fehniger[2,3], Gavin P. Dunn[4], Ravindra Uppaluri[5], Ramaswamy Govindan[2,3], Thomas E. Rohan[6], Malachi Griffith[1,2,3,7], Elaine R. Mardis[8,9], S. Joshua Swamidass[10,11]* and Obi L. Griffith[1,2,3,7]*

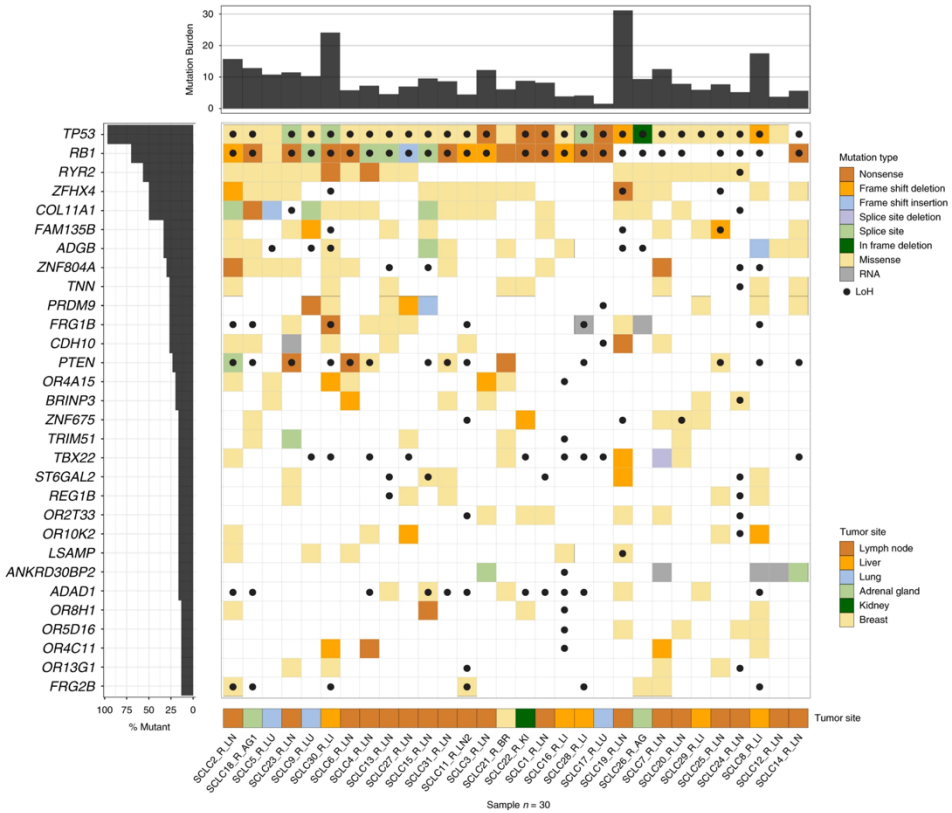
Cancer genomic analysis requires accurate identification of somatic variants in sequencing data. Manual review to refine somatic variant calls is required as a final step after automated processing. However, manual variant refinement is time-consuming, costly, poorly standardized, and non-reproducible. Here, we systematized and standardized somatic variant refinement using a machine learning approach. The final model incorporates 41,000 variants from 440 sequencing cases. This model accurately recapitulated manual refinement labels for three independent testing sets (13,579 variants) and accurately predicted somatic variants confirmed by orthogonal validation sequencing data (212,158 variants). The model improves on manual somatic refinement by reducing bias on calls otherwise subject to high inter-reviewer variability.

Josh Swamidass

Benjamin Ainscough

Erica Barnell

We have applied this model to multiple large-scale cancer genome projects, including our survey of relapsed SCLC



- High mutation burden requiring review of 100s or 1000s of variants
- Confirmed TP53 and RB1 nearly universally co-altered in SCLC
- COL11A1 next most significantly mutated, may mediate resistance to platinum chemotherapy
- Relapse samples also characterized by recurrent alterations in WNT signaling

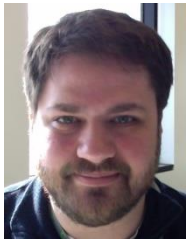
nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-06162-9 OPEN

Recurrent WNT pathway alterations are frequent in relapsed small cell lung cancer

Alex H. Wagner[1], Siddhartha Devarakonda[2,3], Zachary L. Skidmore[1], Kilannin Krysiak[1,2], Avinash Ramu[1], Lee Trani[1], Jason Kunisaki[1], Ashiq Masood[2,3,9], Saiama N. Waqar[2,3], Nicholas C. Spies[1], Daniel Morgensztern[2,3], Jason Waligorski[1], Jennifer Ponce[1], Robert S. Fulton[1], Leonard B. Maggi Jr.[2,3,4], Jason D. Weber[2,3,4], Mark A. Watson[3], Christopher J. O'Conor[5], Jon H. Ritter[5], Rachelle R. Olsen[6], Haixia Cheng[6], Anandaroop Mukhopadhyay[6], Ismail Can[6], Melissa H. Cessna[7], Trudy G. Oliver[6], Elaine R. Mardis[1,3,8,10], Richard K. Wilson[1,3,8,10], Malachi Griffith[1,2,3,8], Obi L. Griffith[1,2,3,8] & Ramaswamy Govindan[2,3]

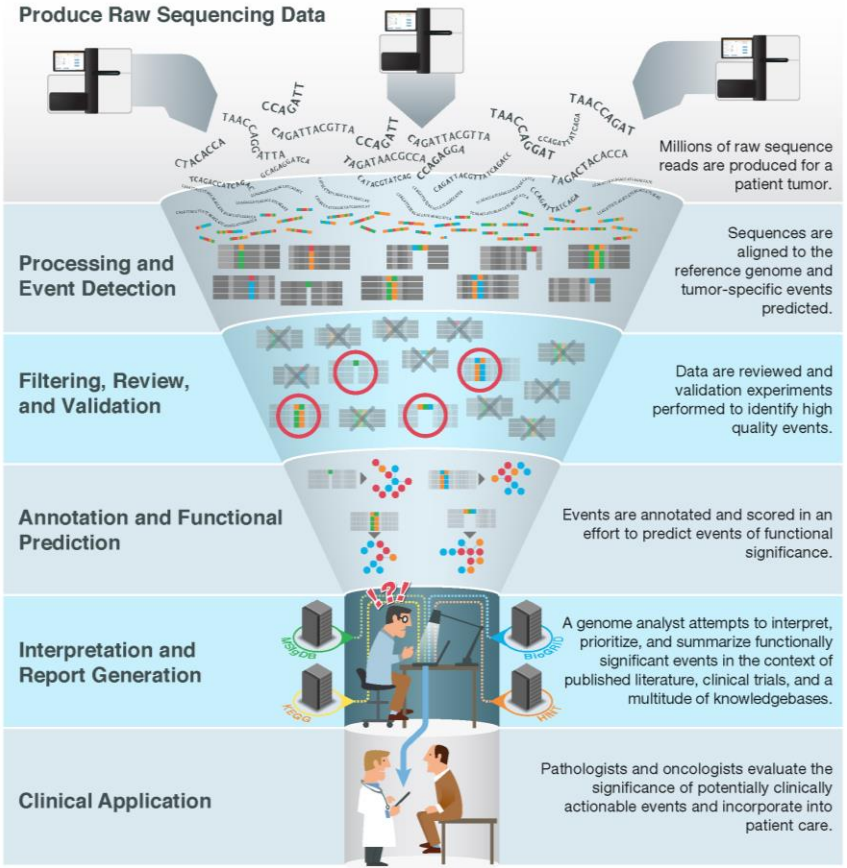


Alex Wagner

Ramaswamy Govindan

Siddhartha Devarakonda

How else can AI be applied to genomic medicine?



Variant calling

Manual review

Functional prediction

Interpretation

Decision-making

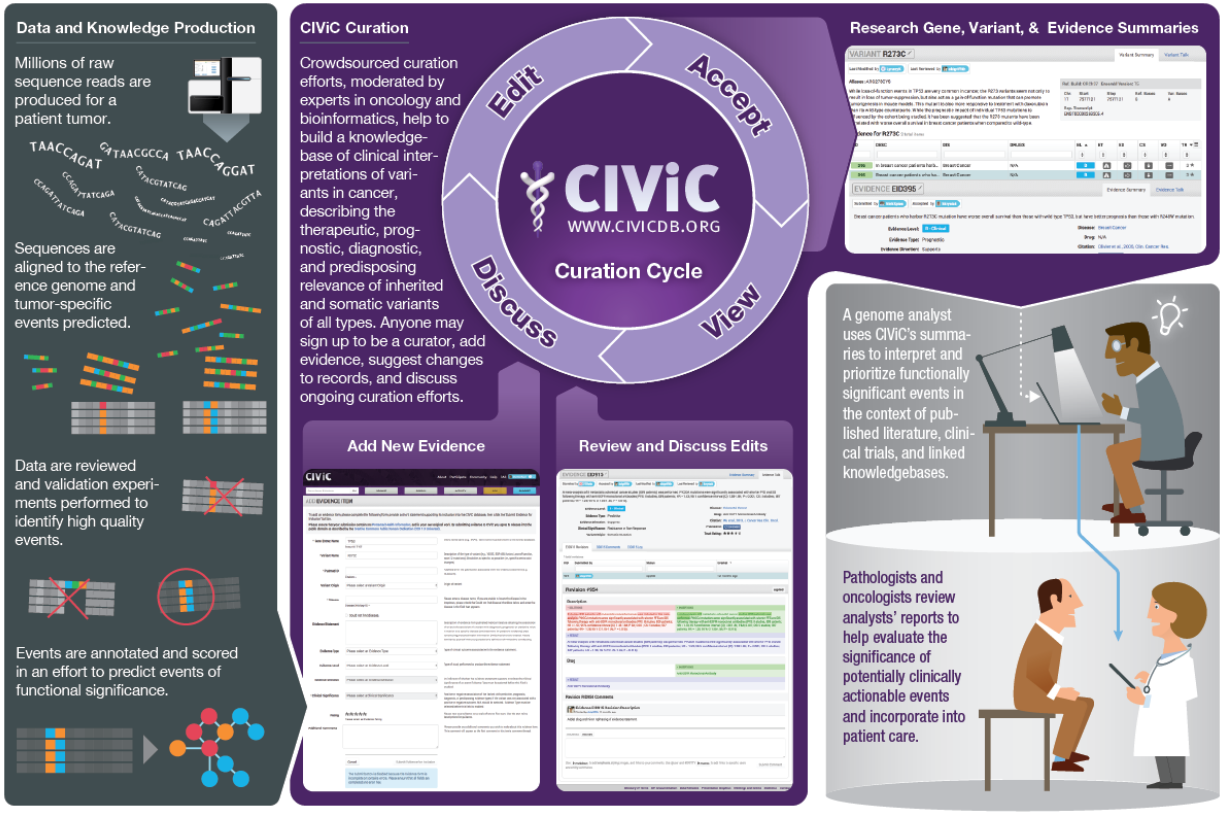
Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.

II) Interpretation – how do we know if a cancer variant is clinically significant?

What is a variant of *significance* in cancer?

- **Predictive** of therapeutic response
  - *BRAF V600E predicts sensitivity to vemurafenib*
- **Diagnostic** of tumor subtype
  - *DNAJB1-PRKACA fusion differentiates fibrolamellar hepatocellular carcinoma from conventional HCC*
- **Prognostic** of survival change
  - *TP53 mutations are associated with worse progression-free survival in lung adenocarcinoma*
- **Predisposing** for cancer development
  - *Patients with the RUNX1 Y260* mutation are associated with increased risk of developing acute myeloid leukemia*

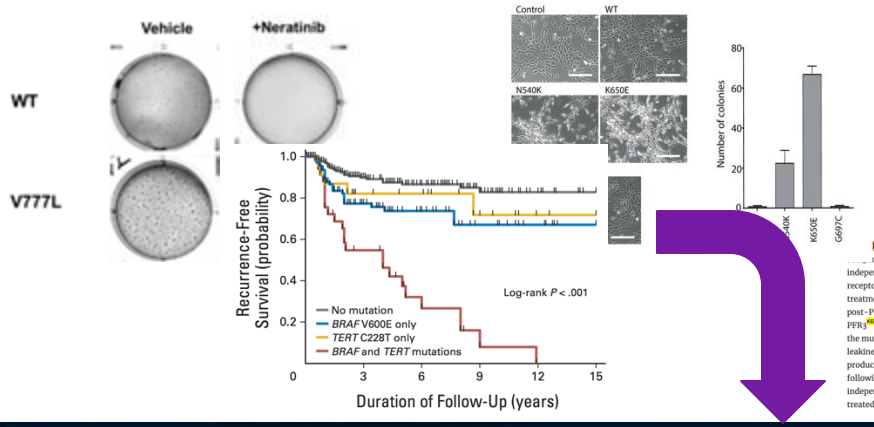
CIViC: an open knowledgebase and curation system for clinical interpretation of variants in cancer

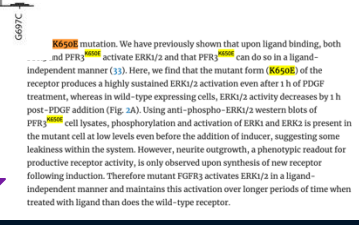


www.civicdb.org

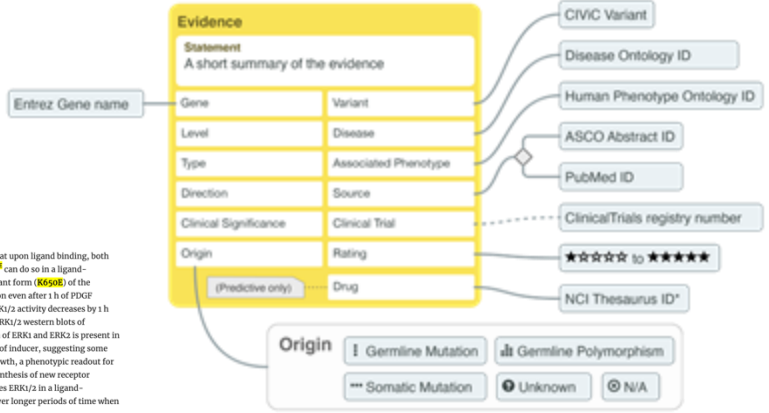
Griffith et al. Nat. Gen. 2017

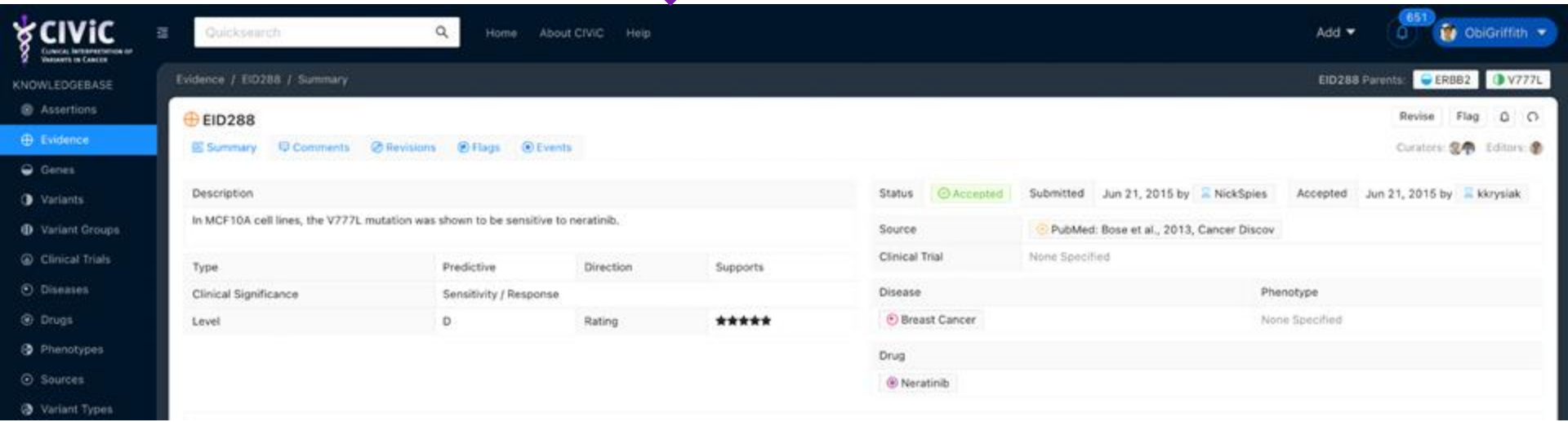
Heterogeneous evidence -> Structured data model



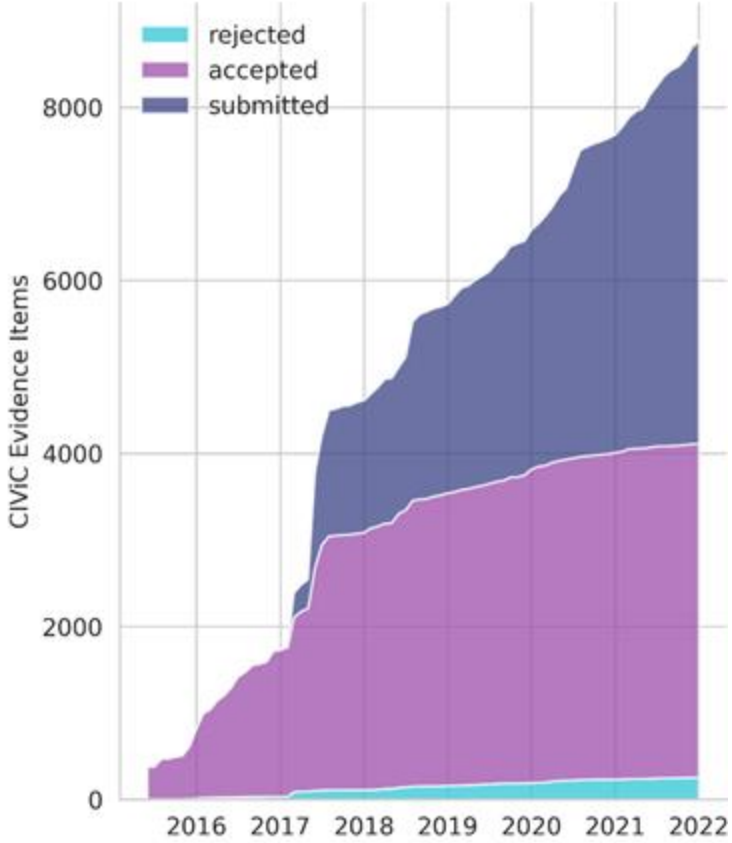


K650E mutation. We have previously shown that upon ligand binding, both ... nd PFR3 K650E activate ERK1/2 and that PFR3 K650E can do so in a ligand-independent manner (33). Here, we find that the mutant form (K650E) of the receptor produces a highly sustained ERK1/2 activation even after 1 h of PDGF treatment, whereas in wild-type expressing cells, ERK1/2 activity decreases by 1 h post-PDGF addition (Fig. 2A). Using anti-phospho-ERK1/2 western blots of PFR3 K650E cell lysates, phosphorylation and activation of ERK1 and ERK2 is present in the mutant cell at low levels even before the addition of inducer, suggesting some leakiness within the system. However, neurite outgrowth, a phenotypic readout for productive receptor activity, is only observed upon synthesis of new receptor following induction. Therefore mutant FGFR3 activates ERK1/2 in a ligand-independent manner and maintains this activation over longer periods of time when treated with ligand than does the wild-type receptor.



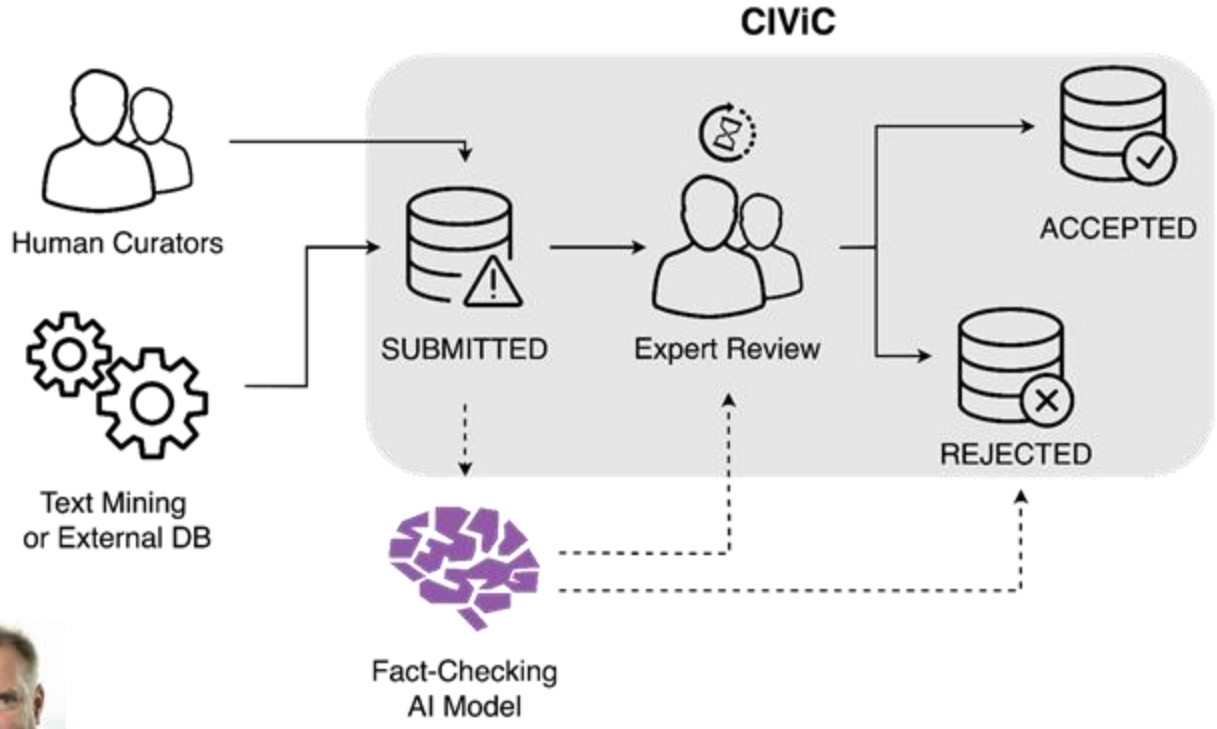


Expert moderation of submitted content is a major bottleneck



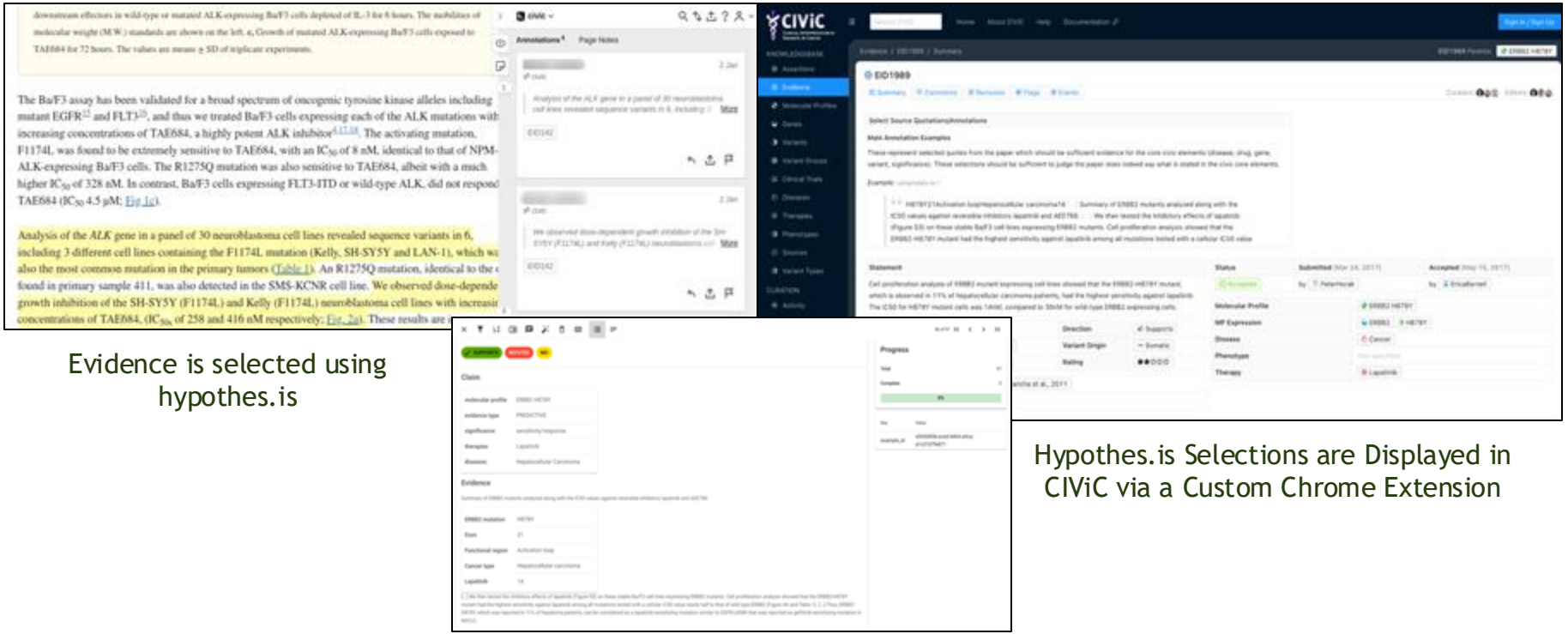
- Entries are being submitted to CIViC faster than experts can review them (right)
- Automated fact-checking could alleviate some of this burden
- Dearth of open-data scientific datasets in fact-checking and natural language interpretation

Proposed Integration of Fact Checking AI into CIViC



Caralyn Reisle (Steven Jones): Poster 64

Annotation Process: Creating the Cancer-Facts Dataset

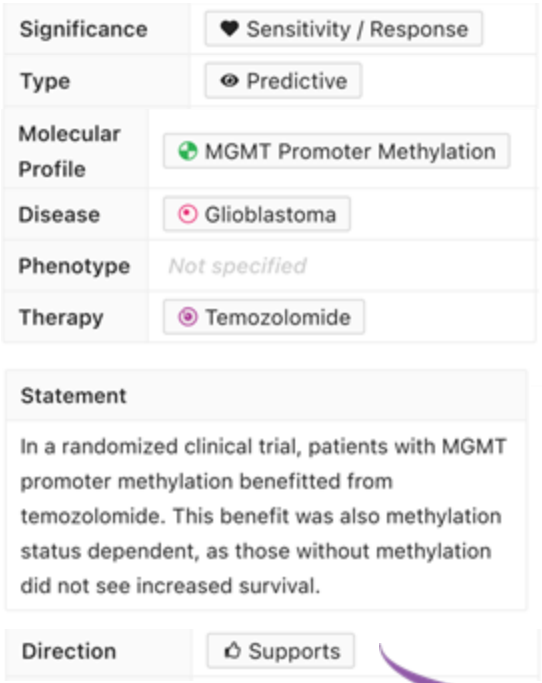


Evidence is selected using hypothes.is

Hypothes.is Selections are Displayed in CIViC via a Custom Chrome Extension

Claim/Evidence Pair Labels are Confirmed with a 2nd Round of Annotation

Evidence Item Consistency Analysis



Prompt

Given a claim and some evidence, determine the stance of the evidence toward the claim. Use the following labels: SUPPORTS, REFUTES, and NEI (not enough information). Give your reasoning.

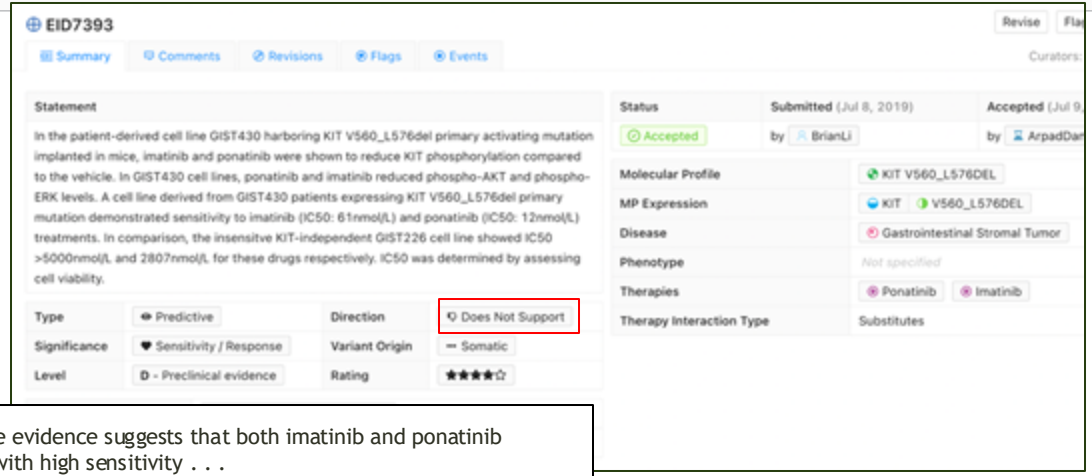
claim: molecular profile is <MP>; evidence type is <evidence type>; significance is <significance>; therapy is <therapy>; disease is <disease>

Expected Response

evidence: <statement>

label: <direction>

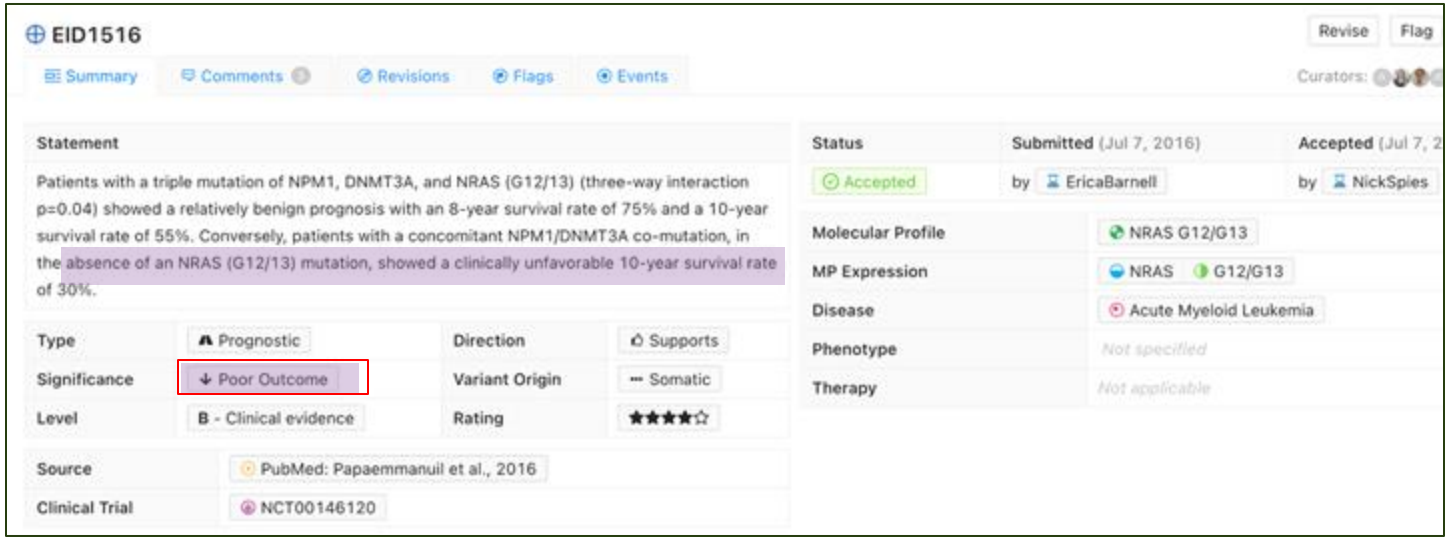
Example #1: AI fact-checking model flags errors in evidence direction field



[mosaicml__mpt-7b-chat]: SUPPORTS. The evidence suggests that both imatinib and ponatinib effectively target GIST molecular profile with high sensitivity . . .

[BioMistral__BioMistral-7B-DARE]: SUPPORTS. The cell line derived from GIST430 patients expressing KIT V560_L576del primary mutation demonstrated sensitivity to imatinib and ponatinib treatments . . .

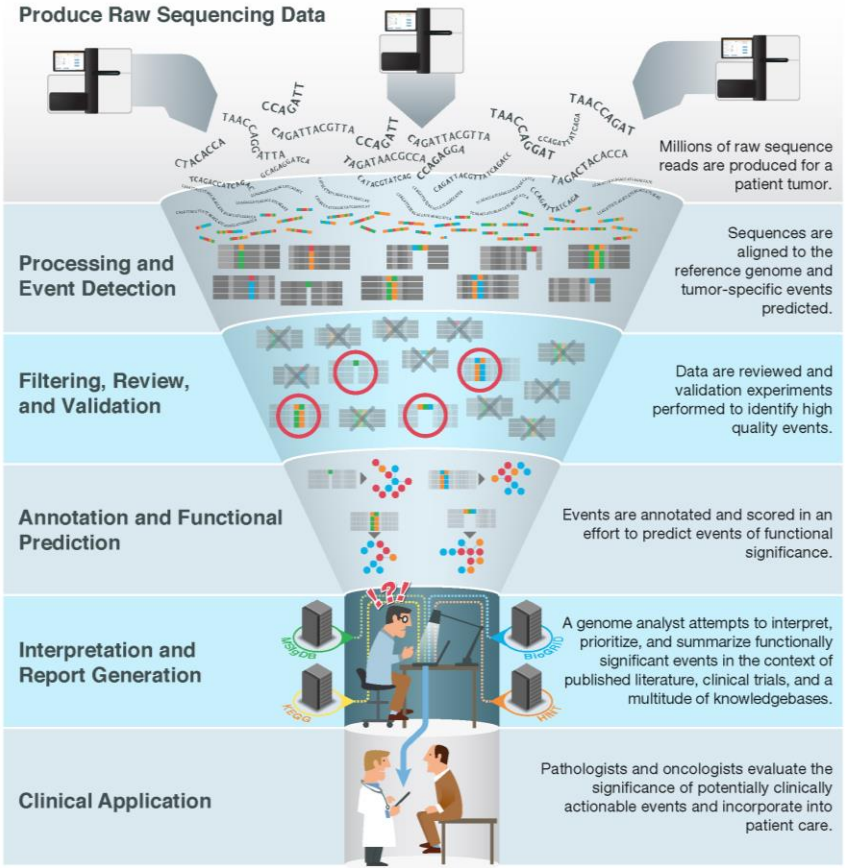
Example #2: AI fact-checking model flags errors in significance field



[BioMistral__BioMistral-7B-DARE]: All elements are present. Patients with NRAS G12/13 mutation had better outcome (8 yr OS 75%) than those without this mutation (10 yr OS 30%) . . .

How else can AI be applied to genomic medicine?

Communication
- figures
- grants
- letters
- homework



Variant calling

Manual review

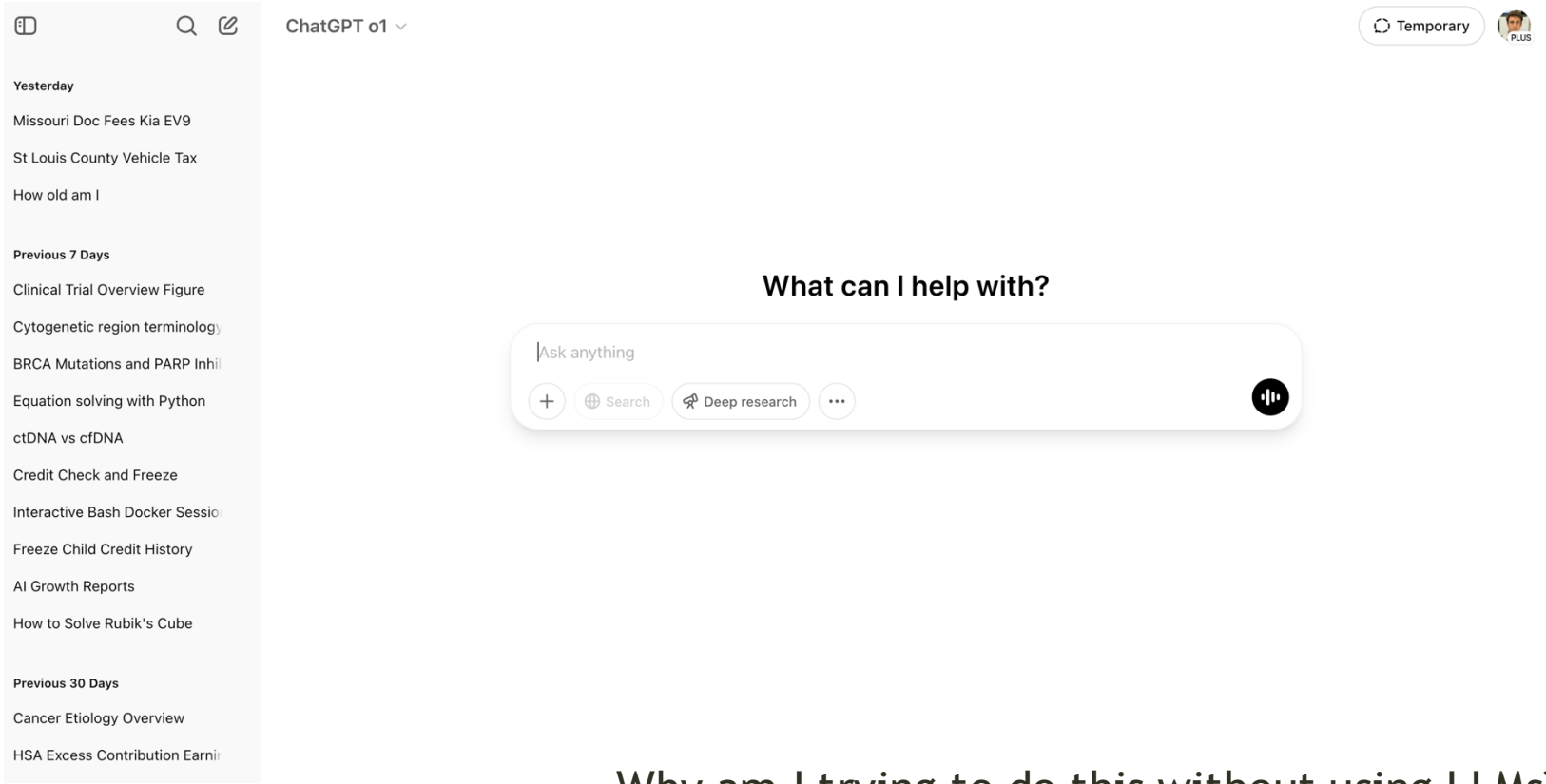
Functional prediction

Interpretation

Decision-making

Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.

What can LLMs help with?

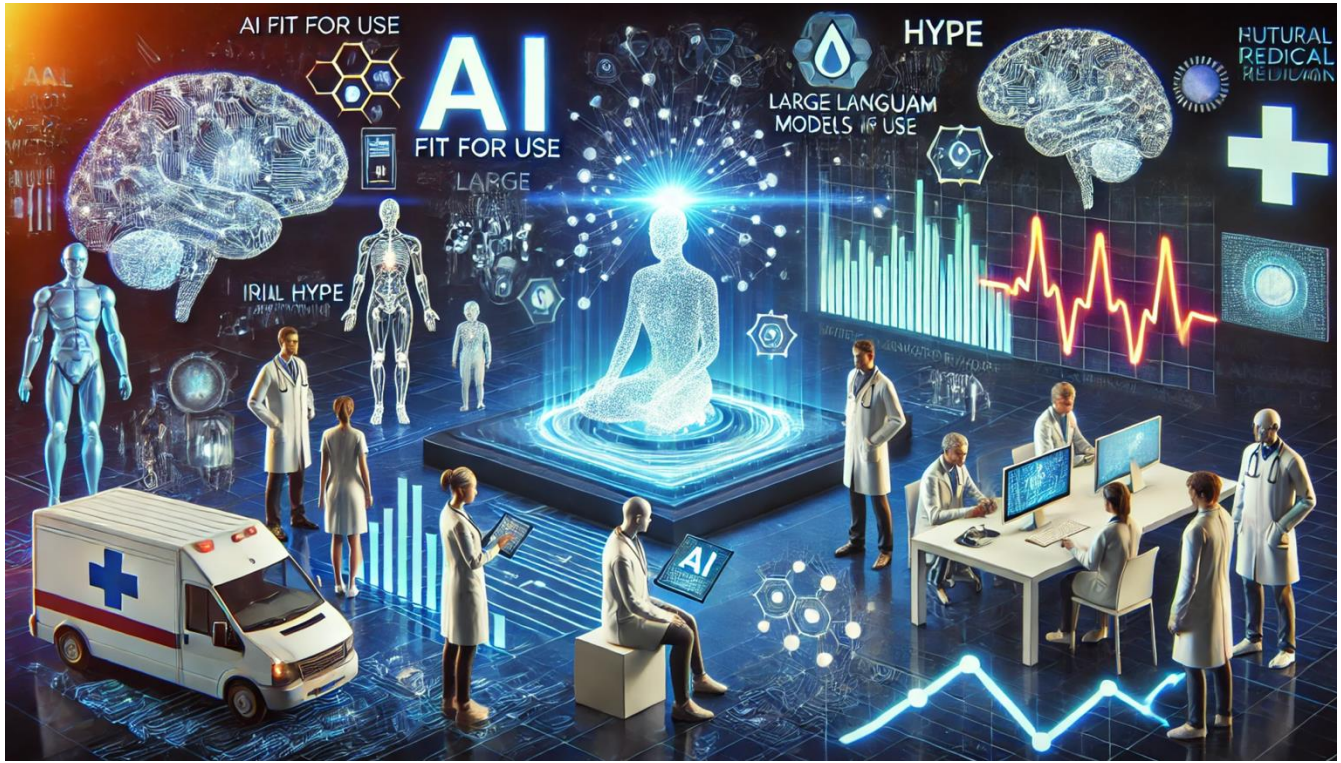


Why am I trying to do this without using LLMs?

I asked DALL-E to make a figure for a review paper

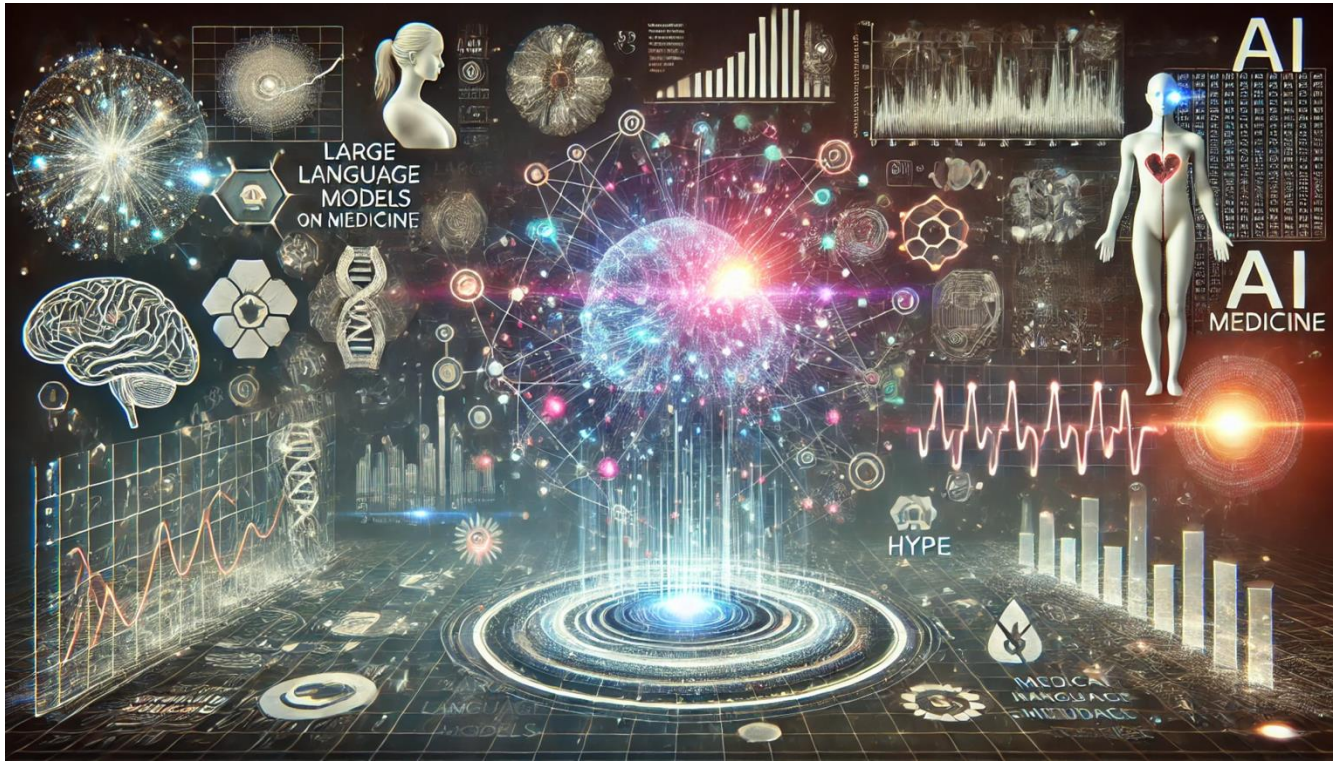
I would like to create a high-level conceptual image to summarize a paper I wrote. The paper is titled "Title: Looking Forward to AI and Medicine: Where Are We, and Where Are We Going?" The key themes of the paper are: (1) Why this moment of AI hype could be different. The emergence of LLMs; (2) The importance of Fit for use – why this should be evaluated differently; (3) The need for evaluation – how it is different in medicine (data sharing, regulation, etc); (4) What we expect to happen - After initial hype, we enter a period of diminishing returns in the short/medium term but then improvements accelerate eventually leading to transformation of many areas of medical practice.

Interesting but weird - what are these people doing?



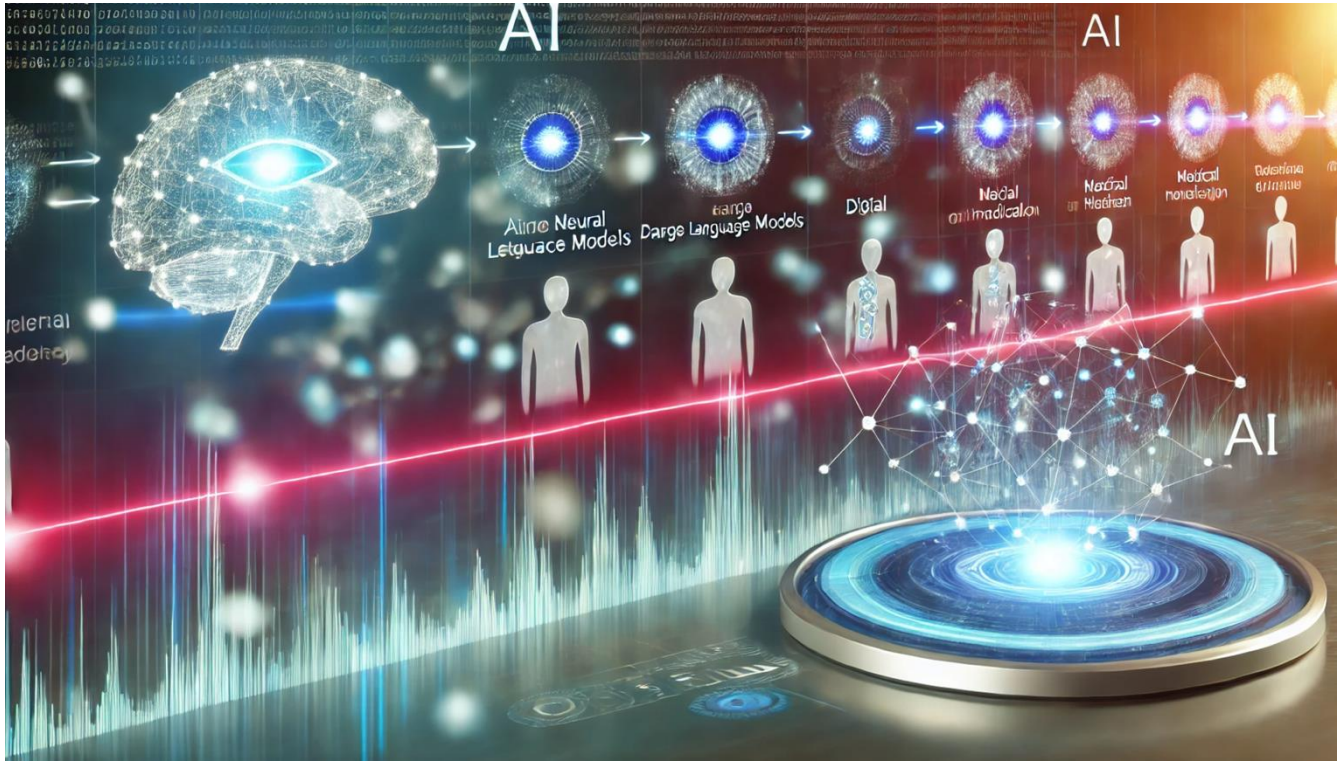
Please create new versions without human figures. Something more conceptual

Cooler but with some problem elements

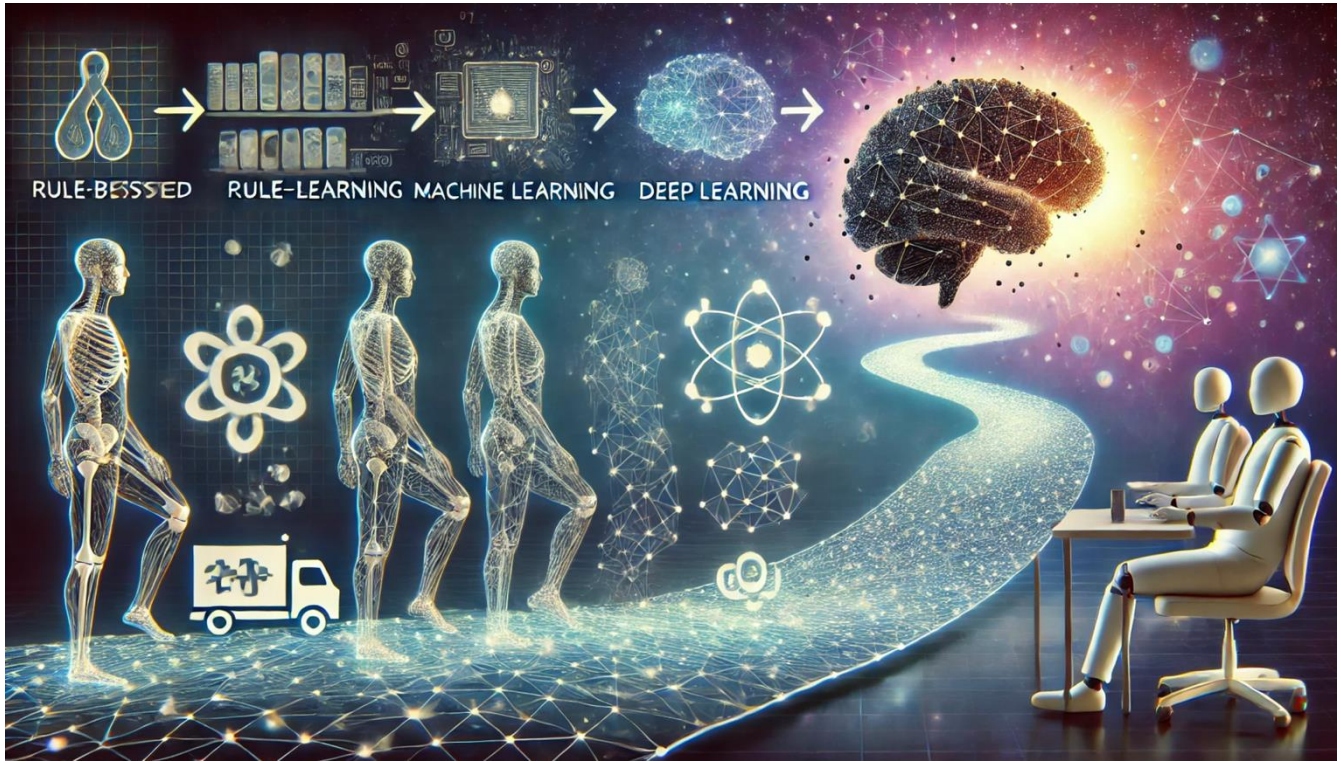


This is great. However, I don't like the red line going down the person on the right. Also, please remove the word "Hype" from the image

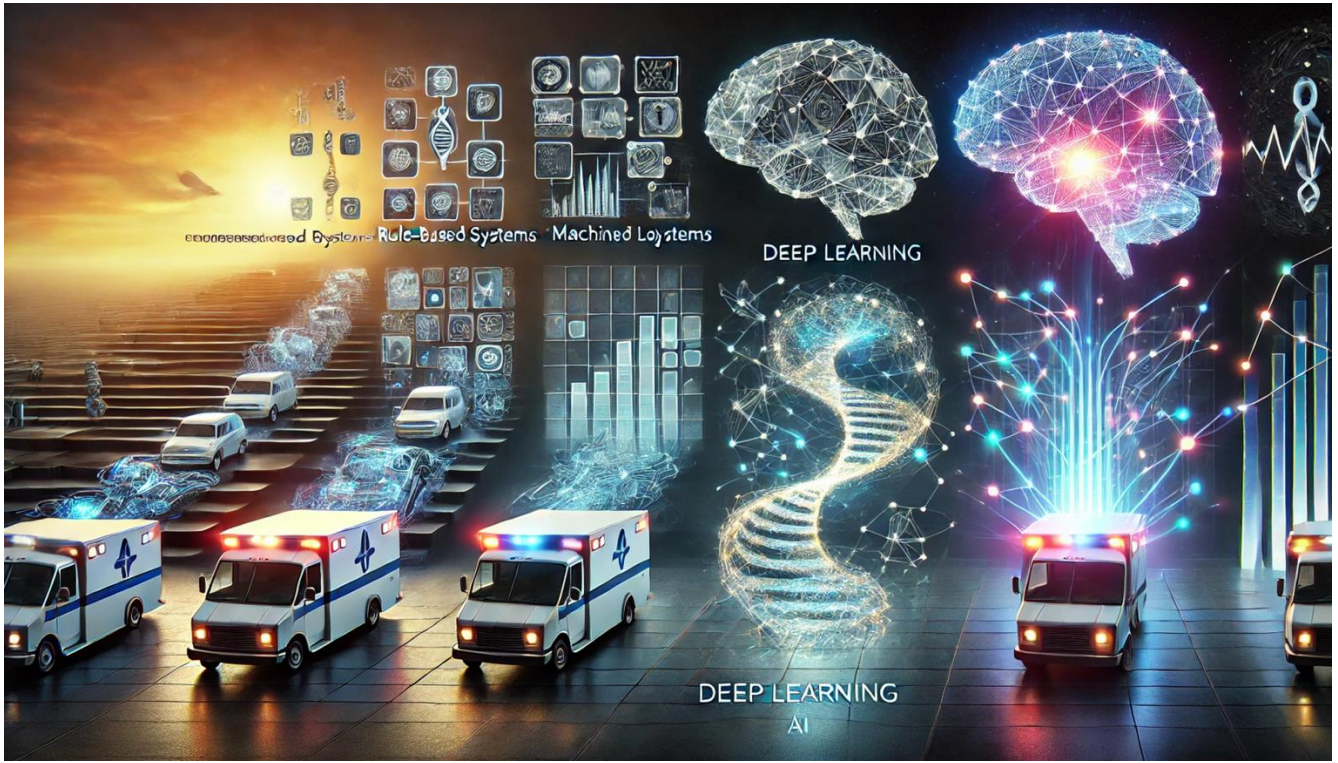
Very different given minor suggestion, but interesting new direction



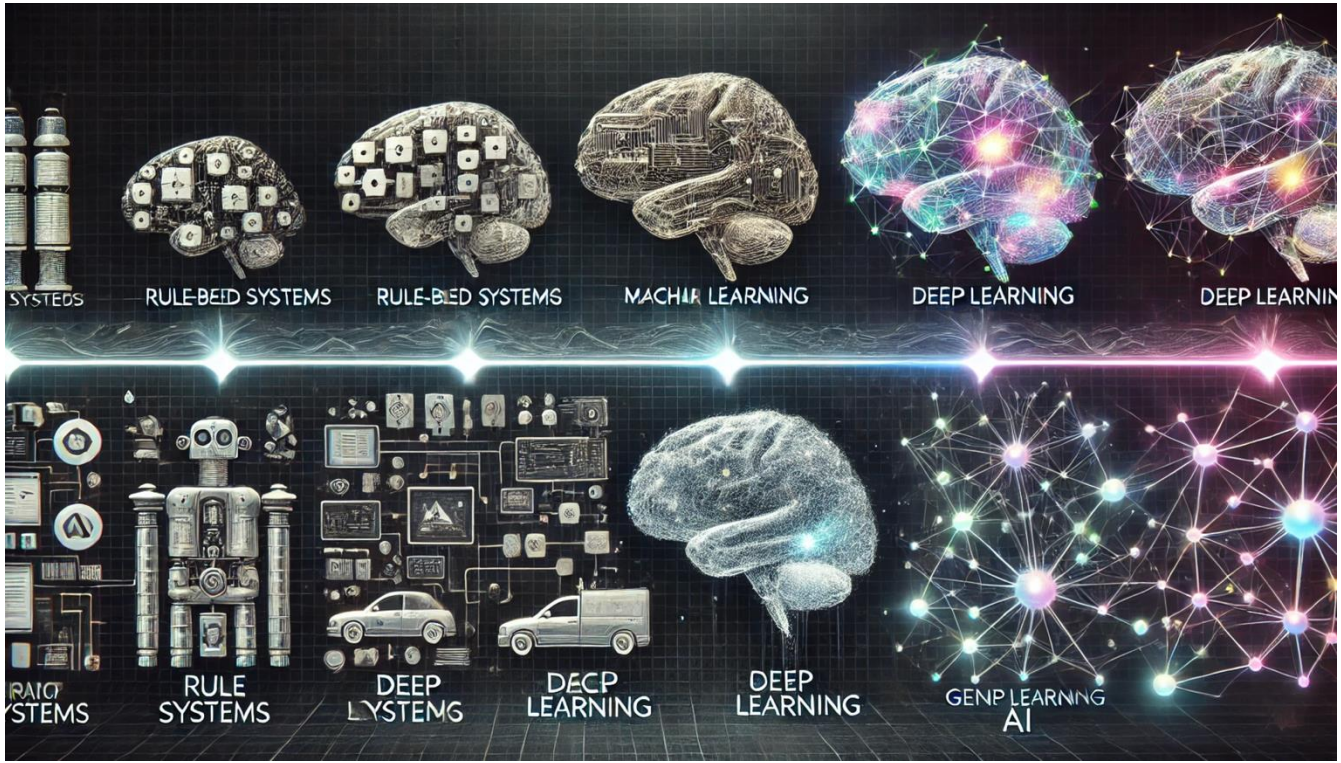
I like this last version which includes both AI concepts and a timeline. Create similar figures which graphically depict the evolution from rule-based systems -> machine learning -> deep learning -> generative AI



These are OK but please don't include any vehicles, include conceptual elements more like the previous two images, and make sure that the four types of AI are correctly spelled and represented in order: (1) "rule-based systems"; (2) "machine learning"; (3) "deep learning"; (4) "generative AI"

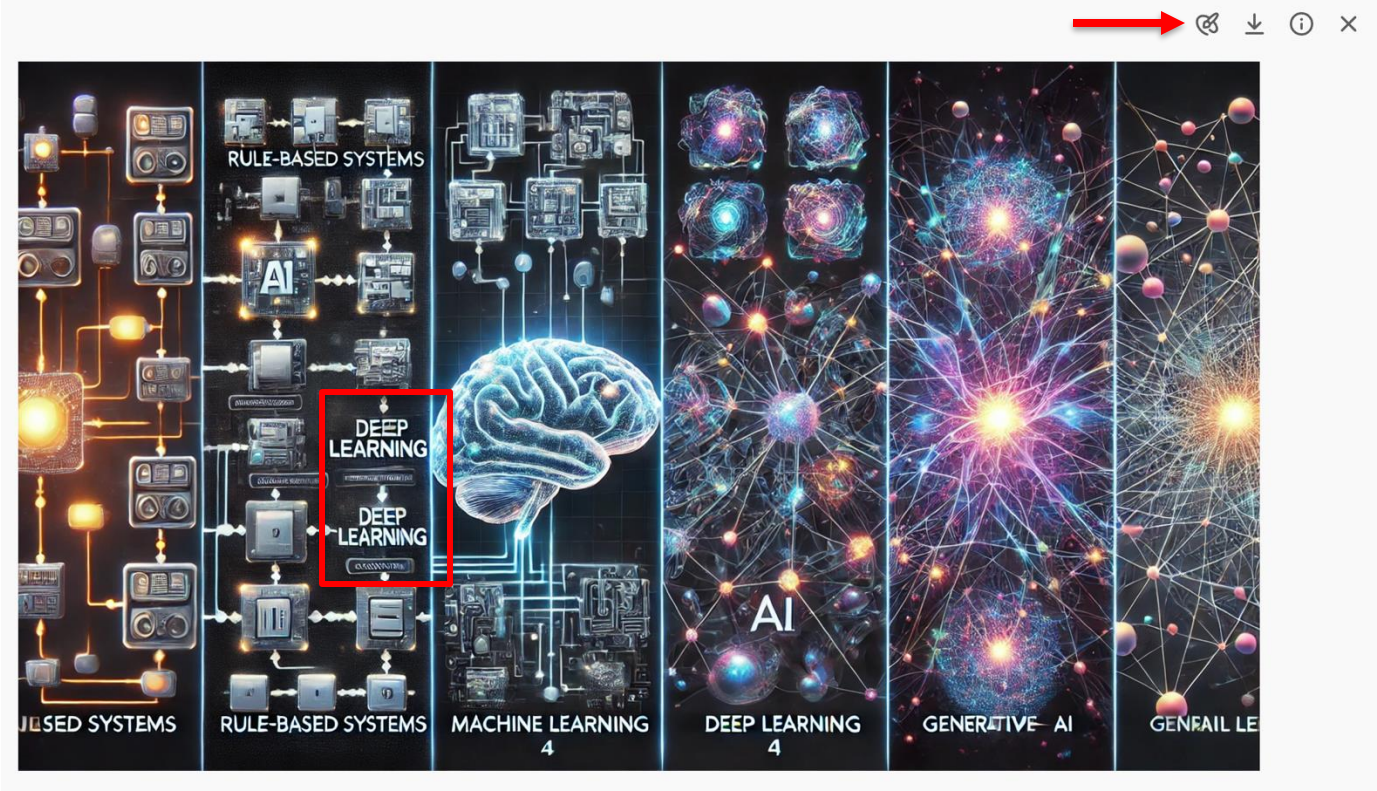


These are better but now create versions without vehicles and just limit to the 4 stages of AI



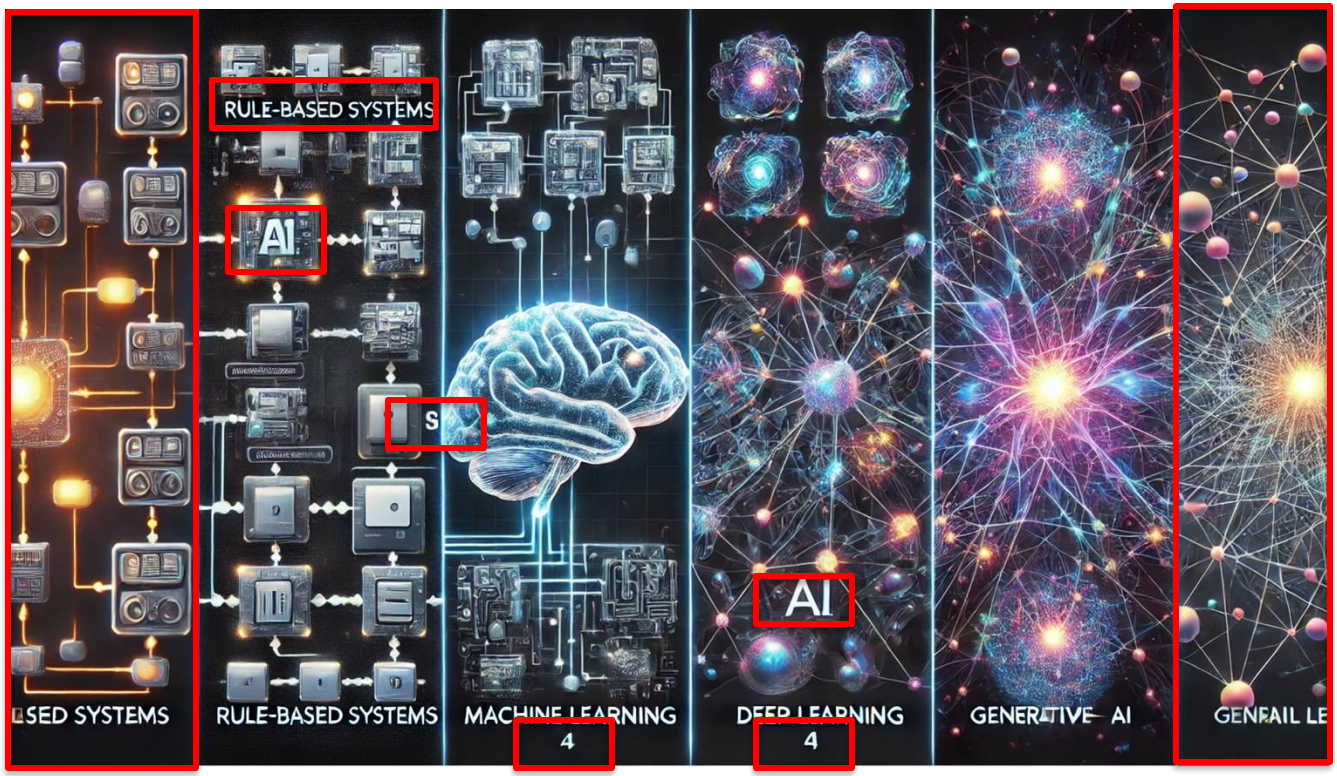
These look great but still don't have the 4 stages, correctly spelled, just appearing once each. Please just depict these four things in order: (1) "rule-based systems"; (2) "machine learning"; (3) "deep learning"; (4) "generative AI"

”Select and area to edit”



remove the words deep learning from this section and replace with graphics similar to other parts of the panel

~10 more iterations



Remove the, outer panels and highlighted letters but otherwise leave the image as is

Final Figure: The different stages of artificial intelligence

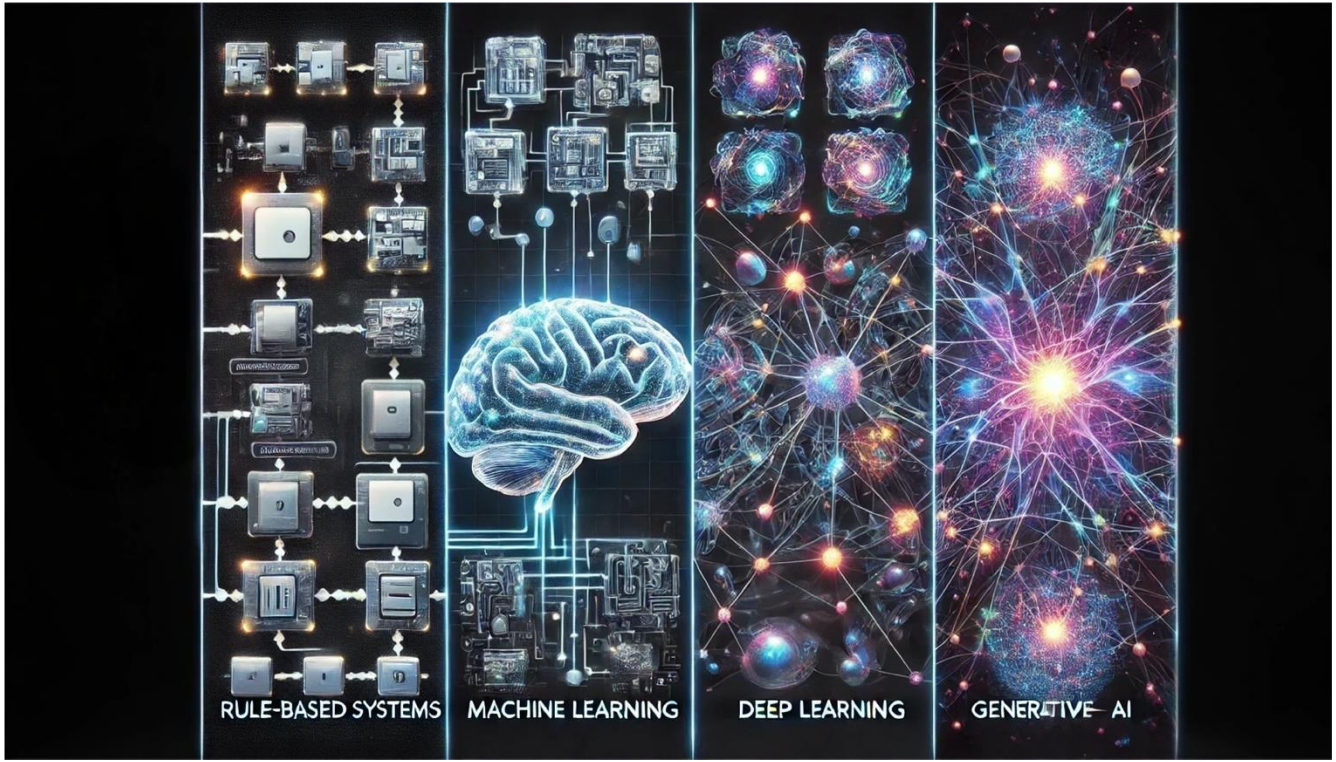


Figure 1. The different stages of artificial intelligence. This image was created (with significant prompt iteration and targeted edit requests) by DALL-E, an AI image generator.

Published in Missouri Medicine - Jan/Feb 2025

Looking Forward to AI and Medicine: Where Are We, and Where Are We Going?

by Adam Wilcox, PhD, Malachi Griffith, PhD & Obi Griffith, PhD

WashU

We speculate that the integration of AI, particularly generative AI, into medicine is expected to progress at a slower pace than in fields such as research.

Adam Wilcox, PhD, FACMI, (pictured), is Director, Center for Applied Health Informatics, Professor of Medicine, Division of General Medicine & Geriatrics, Washington University School of Medicine, St. Louis, Missouri. Malachi Griffith, PhD, and Obi Griffith, PhD, are both Associate Professors of Medicine (Oncology) and Genetics and Assistant Director of the McDonnell Genome Institute at Washington University, St. Louis, Missouri.

Abstract

Artificial intelligence (AI) has emerged as a significant area of interest in medicine, with the potential to influence various aspects of health care. However, the real benefits of applications of AI to medicine often become obscured by the considerable attention, and hype, around its capabilities. To better understand AI's role in medicine, it is important to contextualize its development and review the stages of AI evolution that have contributed to its present state.

Introduction

My (AW) first exposure to AI was taking a course in the topic as part of my graduate studies in a new field called "Medical Informatics." Most of the students in the class were training as computer scientists, and during that semester were particularly interested in a chess match between then-world chess champion Garry Kasparov and IBM's Deep Blue supercomputer. Though Kasparov won that match, the next year Deep Blue would be the victor, marking the first time a computer program defeated a world champion in a tournament-style match. What I didn't recognize at the time was that I wasn't witnessing the beginning of a new era of AI capability, but rather the apex achievement of a rule-based era in the field that was rapidly declining. By the time I completed research for my dissertation, both I and much of the world had moved on to a new approach to AI known as machine learning.[1] These transition points of different eras in the field are important for understanding both where it has been and the significance of where it may be going.

Artificial Intelligence History

The earliest applications of AI that demonstrated practical capabilities beyond proof-of-concepts were rule-based systems. Initially, such systems used simple rules and logic to simulate human decision-making. By the 1970s, expert systems were developed that used large sets of rules to make decisions in specialized domains. Perhaps the peak expert system was Deep Blue. With a complex but well-defined problem and sufficient computing power to search and apply rules it could outperform any human in that task. However, in general these systems were limited in their ability to handle ambiguity, learn from data, and adapt to complex problems. Their dependence on hand-crafted rules made these systems inflexible and difficult to scale.

Years later, after considerable disappointment in rule-based systems, and disinterest in AI generally, another stage of AI emerged in the 1990s with machine learning. Machine learning allowed systems to learn from data, so that

rules could be created and managed flexibly according to the data that were already collected and expected to be available in their application. To a degree, machine learning is a form of automated analysis of data, where algorithms that mimic different approaches to analyzing data are applied automatically. This made them much more flexible and scalable, though they were dependent on both the availability of training data and processing capabilities of the computers generating and applying the developed models. If sufficient data were not available to indicate patterns for learning, machine learning algorithms were limited in what could be discovered.[2]

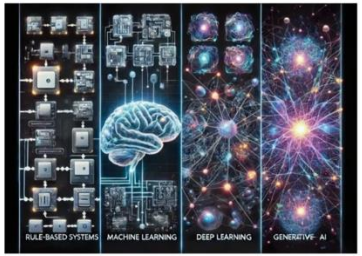


Figure 1. The different stages of artificial intelligence. This image was created (with significant prompt iteration and targeted edit requests) by DALL-E an AI image generator.

The next major breakthrough came in the 2010s with the rise of deep learning, which is a subset of machine learning.[3] This was in large part due to the rapid growth of available data that could be used for machine learning and increased processing capabilities; deep learning coincided with the era of "Big Data" and the emergence of graphics processing units (GPUs) for processing of data. Deep learning used a particular type of machine learning algorithms, artificial neural networks that were composed of multiple layers. These networks were capable of learning high-level features from raw data, which allowed greater abstraction and more powerful modeling of complex datasets.

In addition, neural network models once developed are highly efficient, because they use direct numeric calculations with the variables among the nodes in the networks. With more data and processing capabilities, deep learning allowed discovery of patterns that could not be identified during the previous machine learning era without substantial use of domain knowledge in the process.[1] Not surprisingly, the areas of greatest impact with deep learning were in the domains with the largest amounts of available data, including image and speech recognition, natural language processing, and autonomous systems like self-driving cars. These data sources were available due to growth of image capture, increase of text information ability, and improved data collection.

The current era of AI, coming to prominence in the 2020s, is defined by generative AI, which extends the capabilities of AI models beyond pattern recognition and prediction to the creation of new data (Figure 1). Generative AI models such as large language models (e.g., ChatGPT) are built on transformer architectures, which are a type of deep learning model that is more efficient and powerful than previous models.[4] This allows higher processing capability, making generative AI algorithms the most advanced that have been applied. Importantly, generative AI provides a different kind of interaction than previous modes of AI like deep learning, which were applied to specialized domains and workflows. Because generative AI creates new content rather than simply classifying content already defined, it can provide more immediate value and assist in tasks that are directly useful to a wide range of users. In fact, almost immediately after its release in late 2023, ChatGPT became the most widely and rapidly adopted technology ever.[5]

Artificial Intelligence in Medicine

The application of AI to medicine has evolved alongside the broader field of AI, though progress has often been shaped by the availability of data, regulatory constraints, and the complexity of medical tasks. Expert systems were widely developed for medical settings because of their potential to standardize complex

The role of AI in genomic medicine

- AI is likely to play an increasing role in nearly every aspect of genome-guided (and other types of) medicine
- AI has great potential to increase predictive accuracy, improve reproducibility, automate laborious tasks
- General AI is not yet attained (attainable?) – narrow AI applications will mostly augment/assist physicians and scientists in their work

Thank you!