

Machine Reading for Precision Medicine

Hoifung Poon

Microsoft Research

Overview

Precision medicine

Machine reading

Self supervision

Novel neural architecture

Curation-as-a-Service (CaaS)

Medicine Today Is Imprecise

IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

1. **ABILIFY** (aripiprazole)
Schizophrenia



2. **NEXIUM** (esomeprazole)
Heartburn



3. **HUMIRA** (adalimumab)
Arthritis



4. **CRESTOR** (rosuvastatin)
High cholesterol



5. **CYMBALTA** (duloxetine)
Depression



6. **ADVAIR DISKUS** (fluticasone propionate)
Asthma



7. **ENBREL** (etanercept)
Psoriasis



8. **REMICADE** (infliximab)
Crohn's disease



9. **COPAXONE** (glatiramer acetate)
Multiple sclerosis



10. **NEULASTA** (pegfilgrastim)
Neutropenia

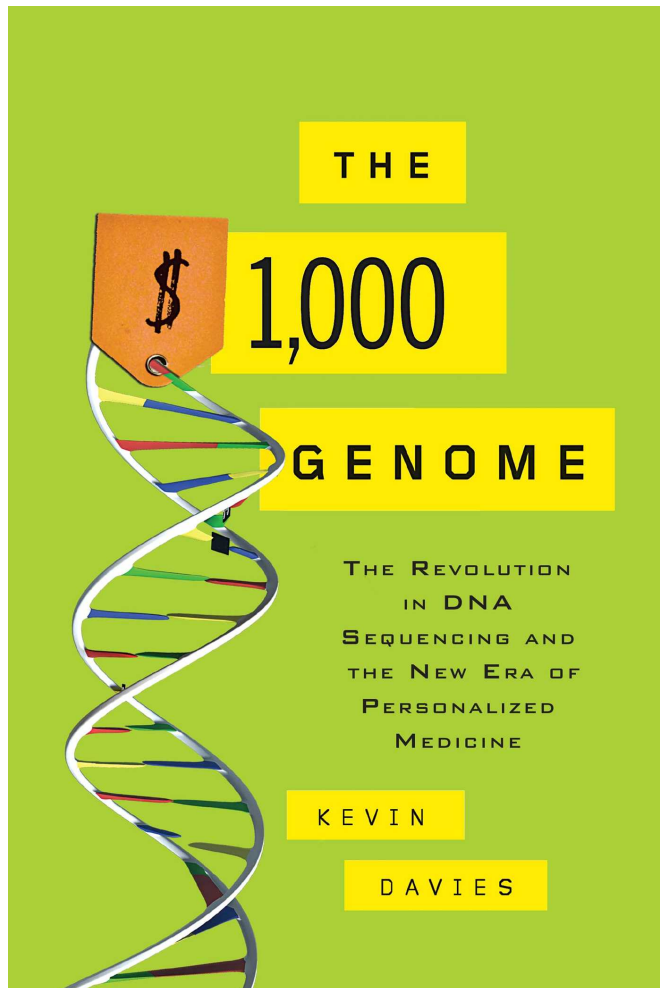


Based on published number needed to treat (NNT) figures. For a full list of references, see Supplementary Information at go.nature.com/4dr78f.

Top 20 drugs
80% non-responders

Wasted
1/3 health spending
\$1 Trillion / year

Disruption: Big Data



Accenture study: 93% of US doctors using EMRs

May 14, 2013 | IHQRE informatics, IHQRE Journal Club | EHR, EMR, Meaningful Use

2009 – 2013: 40% → 93%



Key Scenario: Molecular Tumor Board

Problem: Hard to scale

U.S. 2018: 1.7 million new cases

902 cancer hospitals

Memorial Sloan Kettering

- Sequence: Tens of thousands
- Board can review: A few hundred



$\$200 \text{ rate} \times 10 \text{ experts} \times 3 \text{ hours} \times 1.7 \text{ m} > \100 billion

OncoKB Team

OncoKB is developed and maintained by the Knowledge Systems group in the [Marie Josée and Henry R. Kravis Center for Molecular Oncology](#) at Memorial Sloan Kettering Cancer Center.

Design & Development

Debyani Chakravarty, PhD
 Jianjiong Gao, PhD
 Sarah Phillips, PhD
 Hongxin Zhang, MSc
 Ritika Kundra, MSc
 Jiaojiao Wang, MSc
 Ederlinda Paraiso, MPA
 Julia Rudolph, MPA
 David Solit, MD
 Paul Sabbatini, MD
 Nikolaus Schultz, PhD

Clinical Genomics Annotation Committee

Shrujal Baxi, MD, MPH
 Margaret Callahan, MD, PhD
 Sarat Chandarlapaty, MD, PhD
 Alexandra Charen-Snyder, MD
 Ping Chi, MD, PhD
 Daniel Danila, MD
 Mrinal Gounder, MD
 James Harding, MD
 Matthew Hellman, MD
 Alan Ho, MD, PhD
 Gopa Iyer, MD
 Yelena Janjigian, MD
 Thomas Kaley, MD
 Maeve Lowery, MD
 Antonio Omuro, MD
 Paul Paik, MD
 Michael Postow, MD
 Dana Rathkopf, MD
 Alexander Shoushtari, MD
 Neerav Shukla, MD
 Tiffany Traina, MD
 Martin Voss, MD
 Rona Yaeger, MD

Core Curators

Moriah Nissan, PhD
 Lindsay Saunders, PhD
 Tara Soumerai, MD
 Fiona Brown, PhD
 Tripti Shrestha Bhattarai, PhD
 Kinisha Gala, BSc
 Aphrothiti Hanrahan, PhD
 Anton Henssen, MD
 Phillip Jonsson, PhD
 Iñigo Landa-Lopez, PhD
 Enea Toska, PhD

Quest Diagnostics

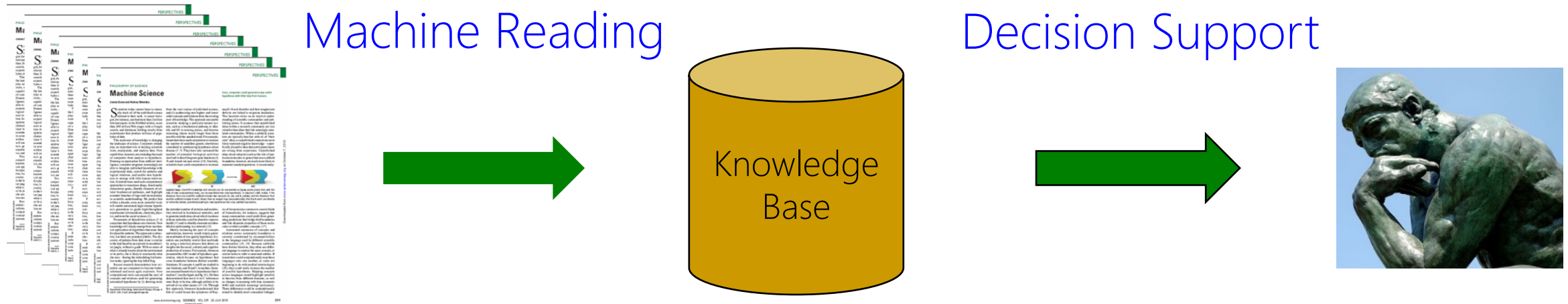
Feras M Abu Hantash, PhD
 Andrew Grupe, PhD
 Matthew Beer, BSc



Everyday
PubMed: 4000 new papers
Expert can curate <10

Information
Overload

Project Hanover



Unlock knowledge from text

Machine Reading

The deletion mutation on exon-19 of **EGFR** gene was present in 16 patients, while the **L858E** point mutation on exon-21 was noted in 10.

All patients were treated with **gefitinib** and showed a partial response.



TREAT(**Gefitinib**, **EGFR**, **L858E**)

Challenge: Variations

TP53 inhibits BCL2.

Tumor suppressor P53 down-regulates the activity of BCL-2 proteins.

BCL2 transcription is suppressed by P53 expression.

The inhibition of B-cell CLL/Lymphoma 2 expression by TP53 ...

.....

negative regulation

532 inhibited, 252 inhibition, 218 inhibit, 207 blocked, 175 inhibits, 157 decreased, 156 reduced, 112 suppressed, 108 decrease, 86 inhibitor, 81 Inhibition, 68 inhibitors, 67 abolished, 66 suppress, 65 block, 63 prevented, 48 suppression, 47 blocks, 44 inhibiting, 42 loss, 39 impaired, 38 reduction, 32 down-regulated, 29 abrogated, 27 prevents, 27 attenuated, 26 repression, 26 decreases, 26 down-regulation, 25 diminished, 25 downregulated, 25 suppresses, 22 interfere, 21 absence, 21 repress

Challenge: Ambiguity

In eubacteria and eukaryotic organelles the product of this gene, **peptide deformylase (PDF)**, removes the formyl group from the initiating methionine of nascent peptides. The discovery that a natural inhibitor of **PDF**, actinonin, acts as an antimicrobial agent in some bacteria has spurred intensive research into the design of bacterial-specific **PDF** inhibitors. In humans, **PDF** function may therefore be restricted to rapidly growing cells



Aliases for PDF Gene

Peptide Deformylase (Mitochondrial) ^{2 3 5}

Polypeptide Deformylase ⁴

EC 3.5.1.88 ⁴

PDF1A ⁴

PDF Gene (Protein Coding) ★

Peptide Deformylase (Mitochondrial)

GCID: GC16M069328 ?

GIFs: 44 ?



Challenge: Document-Level N-ary Relation

*“We next expressed **ALK** F1174L, **ALK** F1174L/L1198P, **ALK** F1174L/**G1123S**, and **ALK** F1174L/**G1123D** in the original SH-SY5Y cell line.”*

(... 15 sentences and 2 figures ...)

*“The 2 mutations that were only found in the neuroblastoma resistance screen (**G1123S/D**) are located in the glycine-rich loop, which is known to be crucial for ATP and ligand binding and are the first mutations described that induce resistance to TAE684, but not to **PF02341066**.”*

Challenge: Annotation Bottleneck

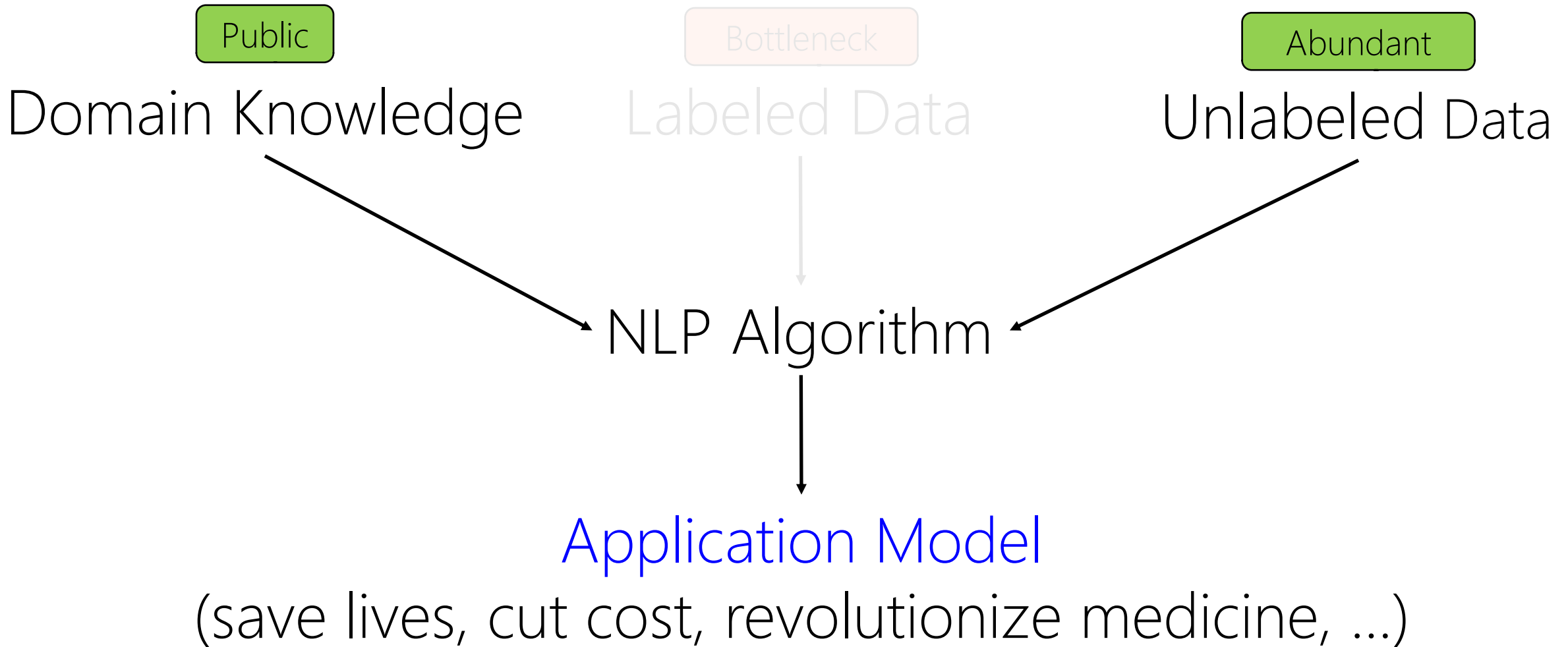
Deep learning requires many labeled examples

Hire experts to label: Not scalable

Crowdsource: Lack domain expertise

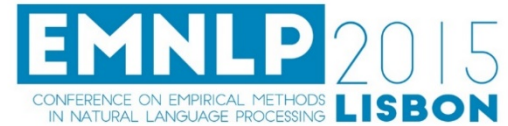
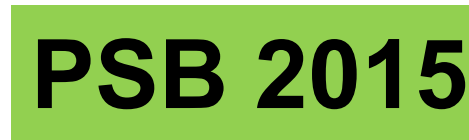
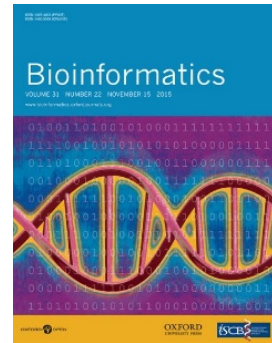
Self-Supervised Machine Reading

Self-Supervised Learning



Hanover: Self-Supervised Deep Learning

Gene Network



Molecular Tumor Board



TACL 2017

EMNLP
2018

NAACL-HLT 2019

Open Science



Example: Distant Supervision

NCI Pathway KB

Regulation	Theme	Cause
Positive	A2M	FOXO1
Positive	ABCB1	TP53
Negative	BCL2	TP53
...

Example: Distant Supervision

NCI Pathway KB

Regulation	Theme	Cause
Positive	A2M	FOXO1
Positive	ABCB1	TP53
Negative	BCL2	TP53
...

Example: Distant Supervision

NCI Pathway KB

Regulation	Theme	Cause
Positive	A2M	FOXO1
Positive	ABCB1	TP53
Negative	BCL2	TP53
...

TP53 inhibits BCL2.

Tumor suppressor P53 down-regulates the activity of BCL-2 proteins.

BCL2 transcription is suppressed by P53 expression.

The inhibition of B-cell CLL/Lymphoma 2 expression by TP53 ...

.....

Example: Distant Supervision

NCI Pathway KB

Regulation	Theme	Cause
Positive	A2M	FOXO1
Positive	ABCB1	TP53
Negative	BCL2	TP53
...

TP53 inhibits BCL2.

Tumor suppressor P53 down-regulates the activity of BCL-2 proteins.

BCL2 transcription is suppressed by P53 expression.

The inhibition of B-cell CLL/Lymphoma 2 expression by TP53 ...

.....

Example: Distant Supervision

NCI Pathway KB

Regulation	Theme	Cause
Positive	A2M	FOXO1
Positive	ABCB1	TP53
Negative	BCL2	TP53
...

TP53 inhibits BCL2.

Tumor suppressor P53 down-regulates the activity of BCL-2 proteins.

BCL2 transcription is suppressed by P53 expression.

The inhibition of B-cell CLL/Lymphoma 2 expression by TP53 ...

.....

Example: Joint Inference

In eubacteria and eukaryotic organelles the product of this gene, peptide deformylase (PDF), removes the formyl group from the initiating methionine of nascent peptides. The discovery that a natural inhibitor of PDF, actinonin, acts as an antimicrobial agent in some bacteria has spurred intensive research into the design of bacterial-specific PDF inhibitors. In humans, PDF function may therefore be restricted to rapidly growing cells.

Coreferent

Example: Joint Inference

In eubacteria and eukaryotic organelles the product of this gene, **peptide deformylase (PDF)**, removes the formyl group from the initiating methionine of nascent peptides. The discovery that a natural inhibitor of PDF, actinonin, acts as an antimicrobial agent in some bacteria has spurred intensive research into the design of bacterial-specific PDF inhibitors. In humans, PDF function may therefore be restricted to rapidly growing cells.

Apposition

Example: Joint Inference

In eubacteria and eukaryotic organelles the product of this gene, **peptide deformylase** (PDF), removes the formyl group from the initiating methionine of nascent peptides. The discovery that a natural inhibitor of PDF, actinonin, acts as an antimicrobial agent in some bacteria has spurred intensive research into the design of bacterial-specific PDF inhibitors. In humans, PDF function may therefore be restricted to rapidly growing cells.



Aliases for PDF Gene

Peptide Deformylase (Mitochondrial) ^{2 3 5}

Polypeptide Deformylase ⁴

EC 3.5.1.88 ⁴

PDF1A ⁴

PDF Gene (Protein Coding) ★

Peptide Deformylase (Mitochondrial)

GCID: GC16M069328 ?

GIFs: 44 ?



Example: Joint Inference

In eubacteria and eukaryotic organelles the product of this gene, **peptide deformylase (PDF)**, removes the formyl group from the initiating methionine of nascent peptides. The discovery that a natural inhibitor of **PDF**, actinonin, acts as an antimicrobial agent in some bacteria has spurred intensive research into the design of bacterial-specific **PDF** inhibitors. In humans, **PDF** function may therefore be restricted to rapidly growing cells



Aliases for PDF Gene

Peptide Deformylase (Mitochondrial) ^{2 3 5}

Polypeptide Deformylase ⁴

EC 3.5.1.88 ⁴

PDF1A ⁴

PDF Gene (Protein Coding) ★

Peptide Deformylase (Mitochondrial)

GCID: GC16M069328 ?

GIFs: 44 ?



Example: Data Programming

Case Histogical type **EGFR** mutation status Brinkman index Treatment agent Response IL-8 (ng/mL) IL-10 (ng/mL) RANTES (pg/ml) 1 Ad Ex21 L858R 900 **gefitinib** PR 5.79 1.67 2 2 Ad Ex21 L858R 0 erlotinib PR 3.33 94.2 1.41 3 Ad Ex19 del 0 erlotinib PR 4.09 126 2.13 4 Ad Ex21 L858R 0 gefitinib PR 3.92 18.1 1.85 5 Ad Ex21 L858R + Ex19 del 370 erlotinib PR 21.9 1.23 2.11 6 Ad Ex19 del 0 erlotinib PR 7.05 1.37 2.56 7 Ad Ex19 del 0 gefitinib PR 4.82 1.59 2.02 8 Ad Ex21 L858R 1640 gefitinib PR 54.9 2.7 2.38 9 Ad negative 400 erlotinib SD 34.9 3.39 3.66 10 Ad negative 0 erlotinib SD 27.1 4.94 2.25 11 Ad unknown 1800 erlotinib SD NE NE NE 12 Sq Ex19 del 3840 gefitinib SD 12.6 1.1 1.98 13 Ad Ex19 del 30 erlotinib SD 5.27 0.69 2.66 14 Ad Ex19 del 0 erlotinib SD 10.5 0.85 2.5 15 Ad unknown 300 erlotinib SD 7.69 2.61 1.7 16 La negative 2080 erlotinib SD NE NE NE 17 Sq Ex21 L858R 750 erlotinib SD 11.6 56.3 3.06 18 Ad unknown 600 erlotinib SD 13.7 1.1 2.82 19 Ad Ex21 L858R 0 erlotinib SD 20.6 11.6 0.783 20 Ad Ex19 del 0 erlotinib SD 14.8 2.07 2.02 21 Ad Ex21 L858R + Ex19 del 0 gefitinib SD 3.04 1.3 1.76 22 Ad Ex19 del 200 gefitinib SD 10.9 1.59 1.91 23 Sq unknown 1560 erlotinib PD NE NE NE 24 Ad negative 1080 erlotinib PD 25.2 8.15 1.57 25 Ad unknown 0 erlotinib PD NE NE NE 26 Ad Ex18 **G719A** 2000 erlotinib PD 11.6 4.48 3.12 27 Ad unknown 1410 erlotinib PD NE NE NE 28 Ad Ex19 del 1000 gefitinib PD 9.44 2.25 2.89 29 Ad negative 1100 erlotinib PD 56.3 5.63 2.16 30 Ad negative 2000 erlotinib PD 15.6 0.85 0.292 31 Ad negative 825 erlotinib PD 8.27 1.59 2.13 32 Ad Ex21 L858R 0 erlotinib PD 5.34 7.05 1.7 33 Ad negative 1680 erlotinib PD 57.7 1.63 2.06

Too many numbers → Unlikely to describe a relation

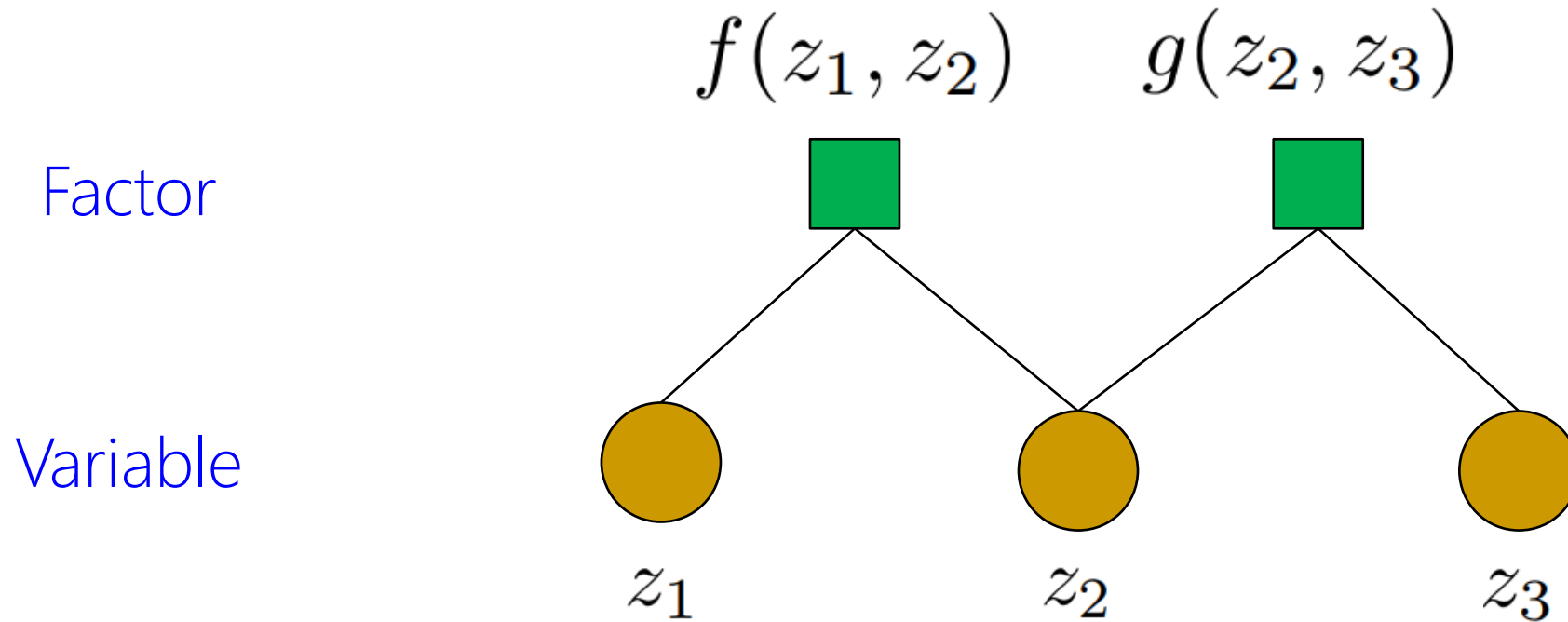
Probabilistic Logic

Distant Supervision $f_{KB}(X_i, Y_i) = \mathbb{I}[\text{In-KB}(X_i, r) \wedge Y_i = r]$

Data Programming $f_L(X_i, Y_i) = \mathbb{I}[L(X_i) = Y_i]$

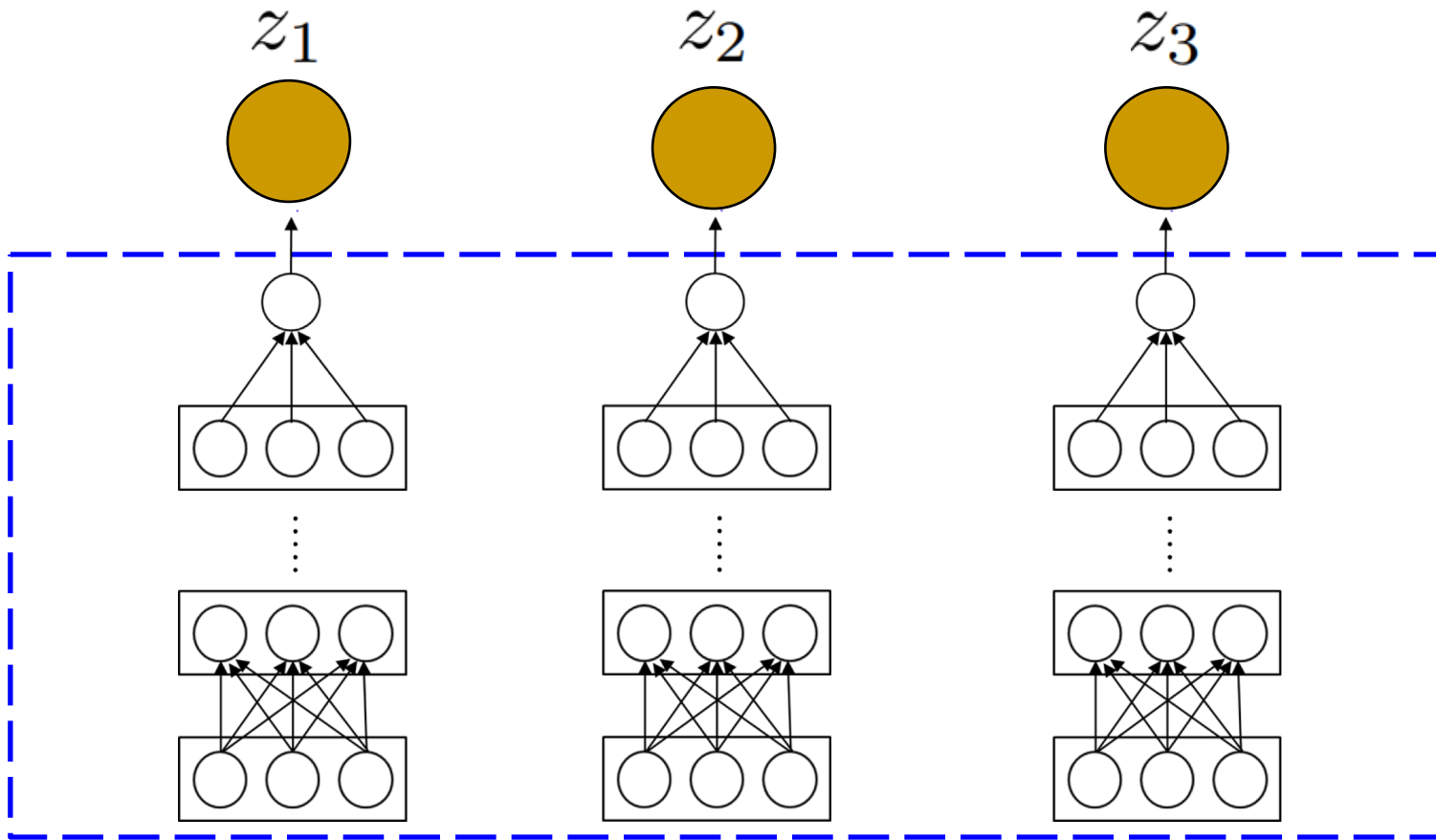
Joint Inference $f_{\text{Joint}}(X_i, Y_i, X_j, Y_j) = \mathbb{I}[\text{Coref}(X_i, X_j) \wedge Y_i = Y_j]$

Probabilistic Logic



Probability $p(z_1, z_2, z_3) \propto \exp(w_f \cdot f(z_1, z_2) + w_g \cdot g(z_2, z_3))$

Deep Learning



Variational EM

Marginal $\sim p(z_1, z_2, z_3)$



Probabilistic Labels

Knowledge

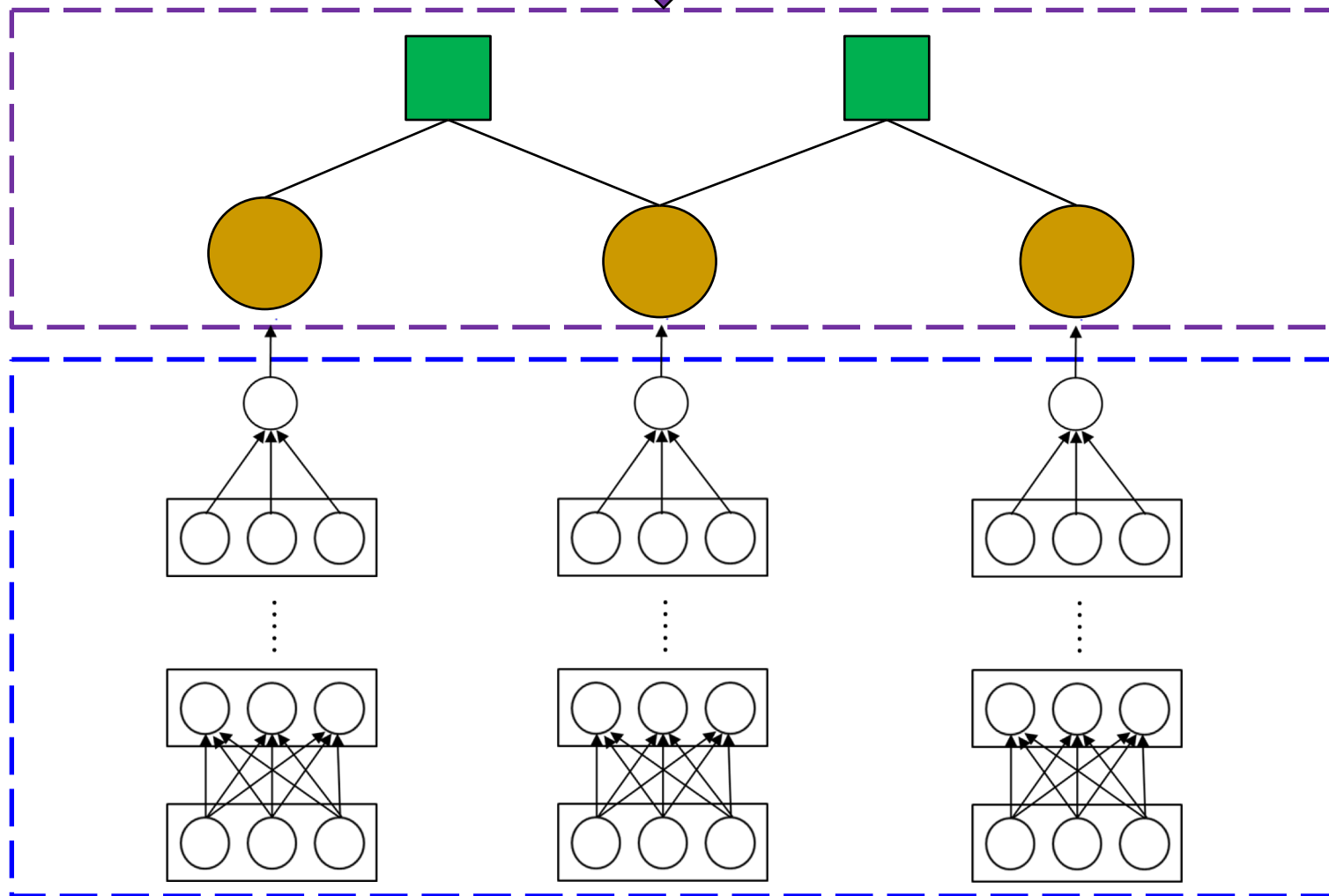
Deep Probabilistic Logic

Virtual Evidence

Latent Variable

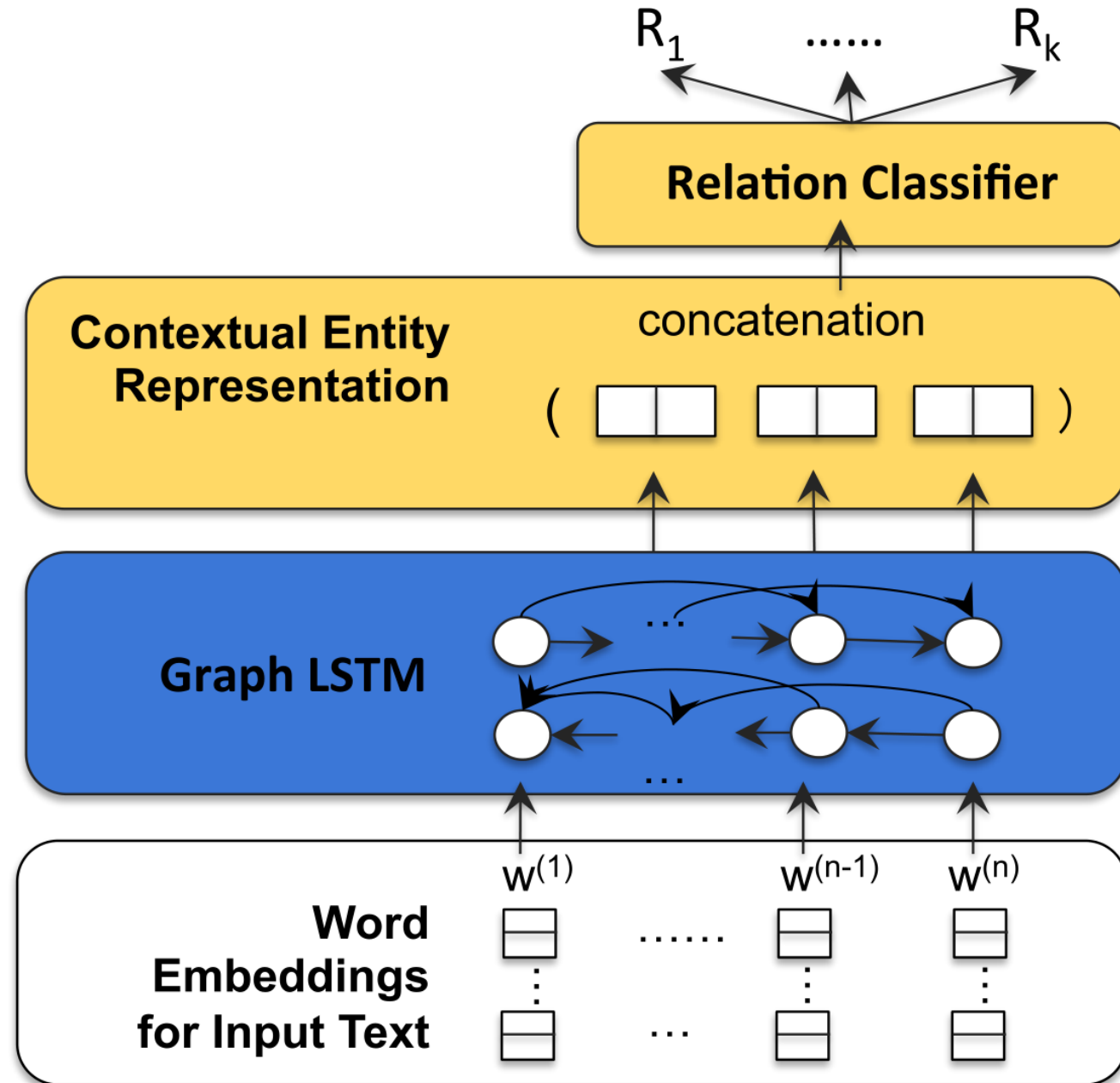
Probabilistic Logic

Deep Learning



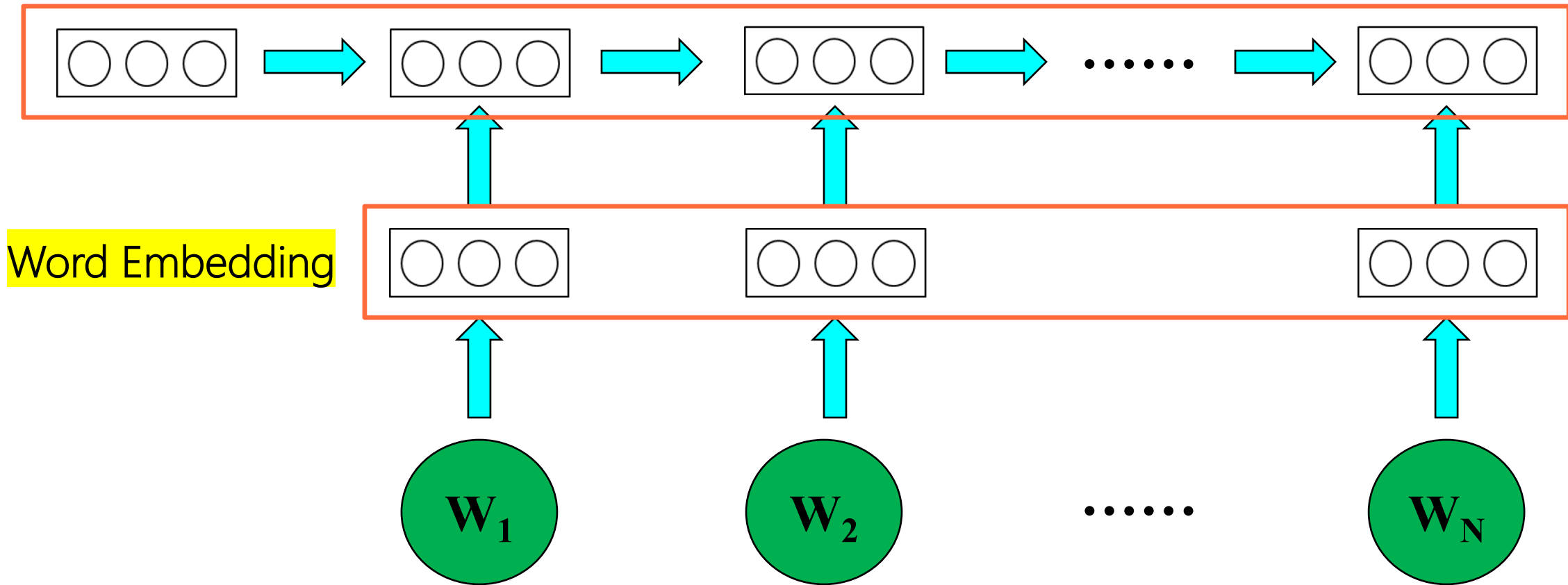
Novel Neural Architecture

Graph LSTM

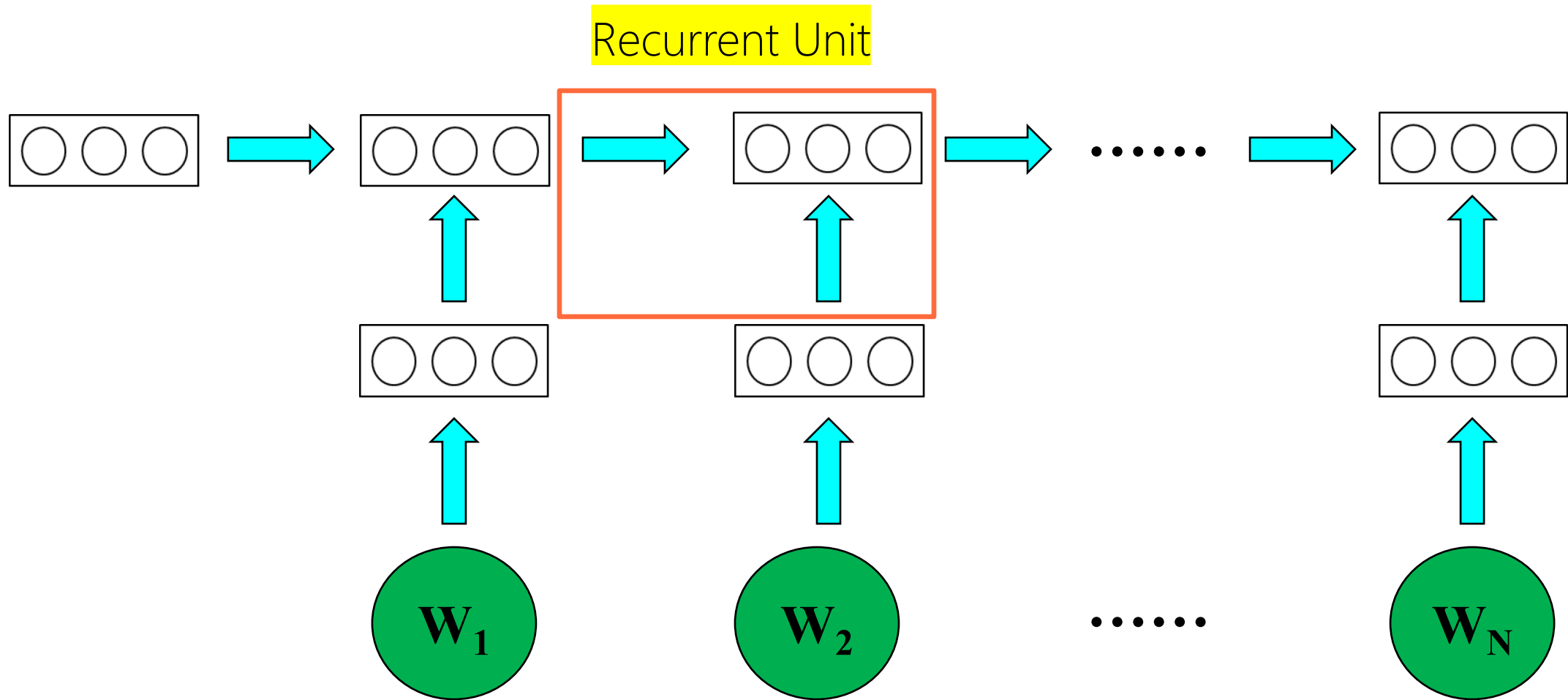


Recurrent Neural Network

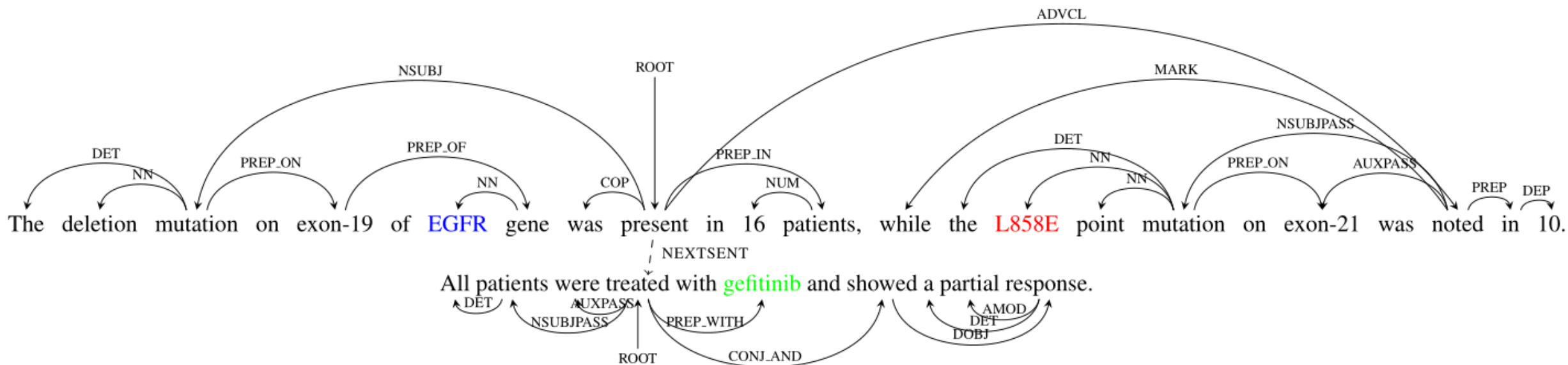
Contextual Hidden Representation



Recurrent Neural Network



Why Graph?



Exploit rich linguistic structures

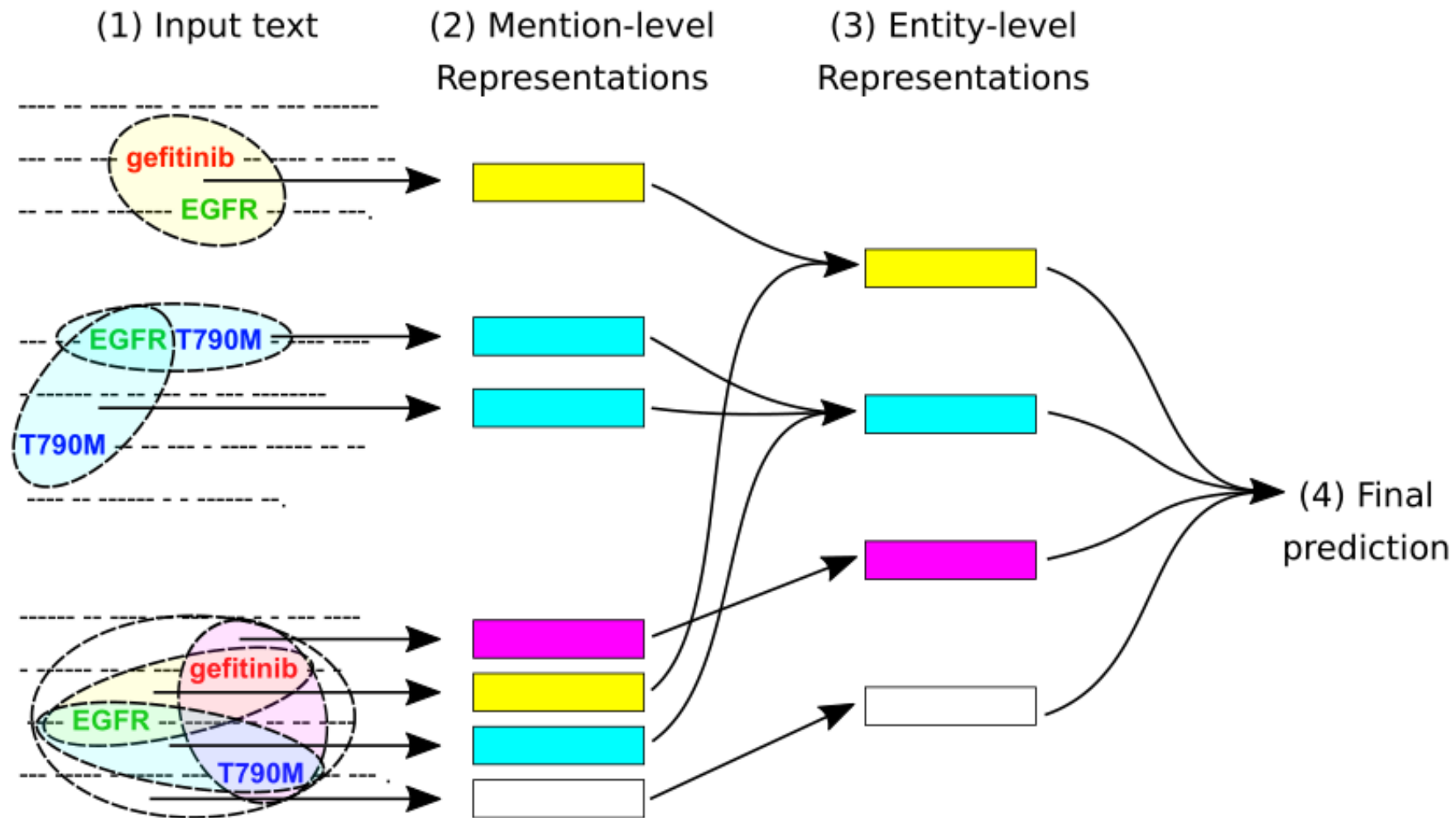
Document-Level Relation Extraction

*“We next expressed **ALK F1174L**, **ALK F1174L/L1198P**, **ALK F1174L/G1123S**, and **ALK F1174L/G1123D** in the original SH-SY5Y cell line.”*

(... 15 sentences and 2 figures ...)

*“The 2 mutations that were only found in the neuroblastoma resistance screen (**G1123S/D**) are located in the glycine-rich loop, which is known to be crucial for ATP and ligand binding and are the first mutations described that induce resistance to TAE684, but not to **PF02341066**.”*

Multiscale Representation Learning



Graph LSTM

Distant supervision: GDKD + CIVIC

Cross-sentence triples extraction yield

Machine reading: 70 X manual curation

Graph helps, esp. if syntactic parses are good

Peng et al. "Cross-Sentence N-ary Relation Extraction with Graph LSTM", *TACL-17*.

Deep Probabilistic Logic

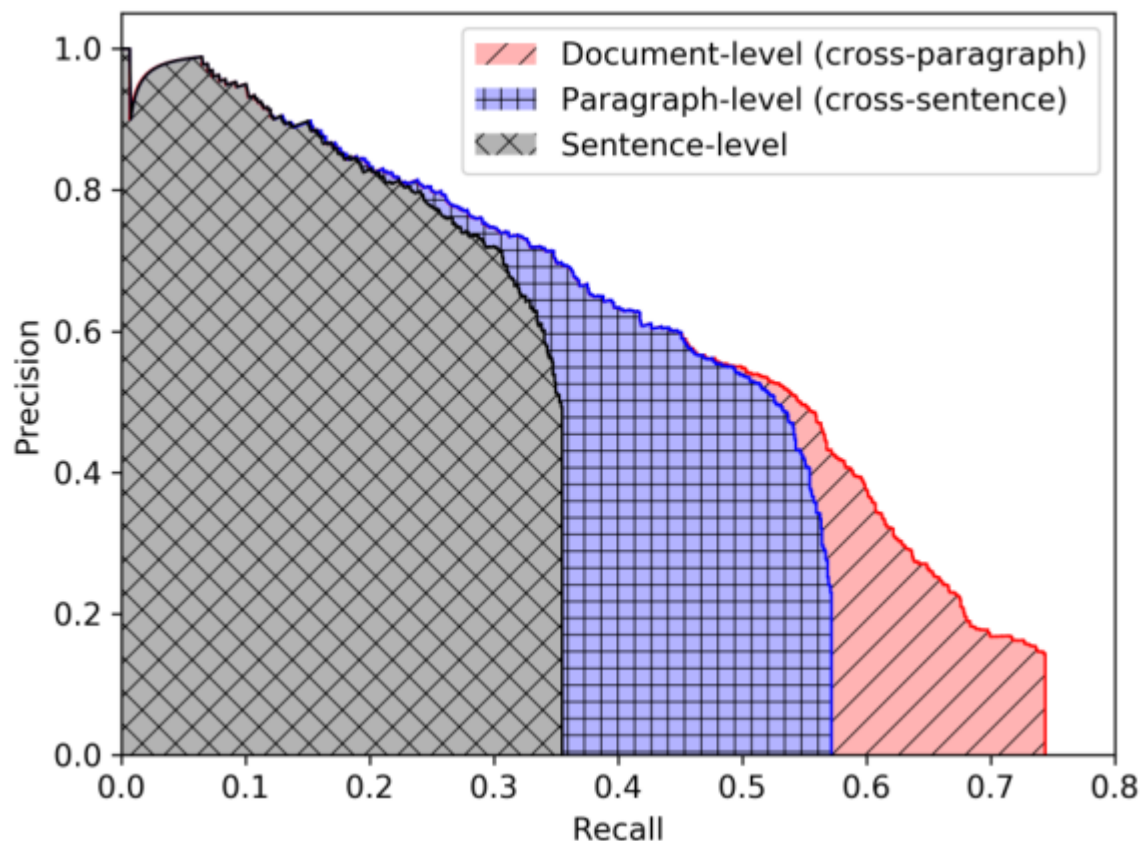
Distant Supervision + Data Programming + Joint Inference

Precision: +10 points (64% → 74%)

Recall: +25%

Wang & Poon. "Deep Probabilistic Logic: A Unifying Framework for Indirect Supervision", *EMNLP-18*.

Multiscale Representation Learning



Improve AUC by absolute 8-13 points
Compared to Wang & Poon 2018

Jia et al. "Document-Level n-ary Relation Extraction with Multiscale Representation Learning", *NAACL-19*.

Curation-as-a-Service (CaaS)

Assisted Curation

Evidence Type

- ☐ Clinical
- ☐ Preclinical Patient Derived Xenograft
- ☐ Preclinical Patient Derived Cell Culture
- ☐ Preclinical Cell Line Xenograft
- ☐ Preclinical Cell Line Culture
- ☐ Unknown

Publication Type

- ☒ Primary
- ☐ Review

Table/Text Type

- ☒ Table
- ☒ Text

Drugs

(715)

Filter

10058-f4
2-methoxyestradiol
5-fluoropyrimidine
79-6
a-1155463
a-1210477
a-395
a-485
abbv-075
abemaciclib
abiraterone
abl001
abt-263
abt-348
abt-737
ac-93253 iodide

talazoparib

Genes (13)

ATR
CTNNB1
IDH1
JAK2
KMT5B
KRAS
MAP2K1
MPL
PARP1
RAD51C
STAG2
TBCE
TP53

Variant PubMed ID Score Level of evidence

a146t 28566428 0.78

To explore whether MEKi could re-sensitize PARPi resistant cells to effects of PARPi, we developed PARPi resistant cells by culturing highly PARPi sensitive cells (UWB1.289 and A27980CP, both RAS wild type, see Fig. 2) in the continued presence of **BMN673** for 3 to 4 months, at which time drug resistant clones emerged. A2780CP PARPi resistant (A2780CP_R) and UWB1.289 PARPi resistant (UWB1.289_R) clones were highly resistant to **BMN673** and cross resistant to olaparib (Fig. 3A-B). RPPA analysis demonstrated that RAS / MAPK pathway activity (increased pMEK, pBAD, and pFOXO3a (inactive form)) was upregulated in PARPi resistant clones (Fig. 3C). Moreover, resistant clones showed lower total FOXO3a and BIM, as expected from increased RAS / MAPK pathway activity. The decreased PAR and PARP1 expression in the resistant cells could also contribute to PARPi resistance, as PARP1 expression is associated with PARPi sensitivity (22). Western blotting confirmed increased RAS / MEK pathway activity with concomitant decreases in FOXO3a and BIM in resistant cells (Fig. 3D). Overall, the signaling changes in long-term PARPi resistant cells exhibited many similarities to adaptive responses to short-term PARPi treatment (see Fig. 1). Despite increased RAS / MEK pathway activity, **KRAS** sequencing demonstrated that the resistant lines did not acquire classical activating **KRAS** mutations. However, deep NGS sequencing as well as Sanger sequencing of individual PARPi resistant clones from A2780CP_R demonstrated the presence of **KRAS** **A146T**, **KRAS** A59T and MAP2K1 A283T in 19, 11 and 6 % of cells respectively but not in A2780CP parental cells. Importantly, prolonged culture of the lines without PARPi resulted in loss of the mutant **KRAS** and MAP2K1 clones. The **KRAS** **A146T** mutant has been demonstrated to be modestly activating (30). The selection of **KRAS** mutations in a PARPi resistant line supports the concept that RAS mutations and RAS / MAPK pathway activation is a key mediator of PARP resistance. As expected by increases in RAS / MAPK activity in PARPi resistant cell lines and **KRAS** and MAPK1 mutations, A2780CP_R were markedly more sensitive and UWB1.289_R were modestly more sensitive to MEKi (Fig. 3E-F). MEKi re-sensitized both PARPi resistant clones to PARPi (Fig. 3E-F). Thus

M **Curation Worthy** Entails, but not for Curation Not Entails Clear RPi.

This paragraph describes an observed relation between talazoparib, KRAS and a146t

Add notes

Updated by microsoft 24 days ago (Jul 25, 2019 3:48:29 PM)

a59t 28566428 0.56

To explore whether MEKi could re-sensitize PARPi resistant cells to effects of PARPi, we developed PARPi resistant cells by culturing highly PARPi sensitive cells (UWB1.289 and A27980CP, both RAS wild type, see Fig. 2) in the continued presence of **BMN673** for 3 to 4 months, at which time drug resistant clones emerged. A2780CP PARPi resistant (A2780CP_R) and UWB1.289 PARPi resistant (UWB1.289_R) clones were highly resistant to **BMN673** and cross resistant to olaparib (Fig. 3A-B). RPPA analysis demonstrated that RAS / MAPK pathway

Assisted Curation

Curation

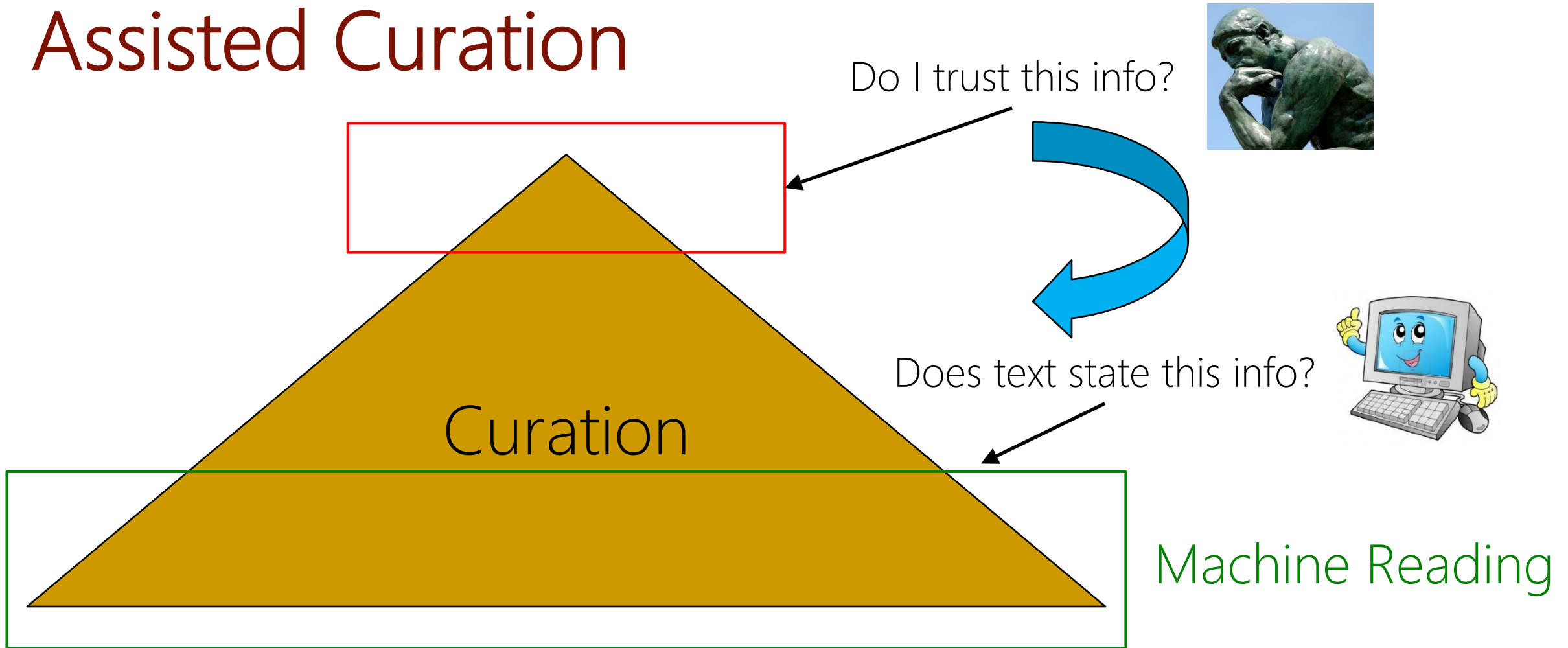
Does text state this info?



Machine Reading

Goal: Empower curators with super speed

Assisted Curation



Goal: Empower curators with super speed

Curation-as-a-Service (CaaS)

Molecular tumor board

Cancer registry

Off-label drug use

Clinical trial matching

Synthetic control

Post-market surveillance

Molecular Tumor Board



DRUG	GENE	MUTATION	RESPONSE
gefitinib	EGFR	L858R	sensitive
erlotinib	EGFR	T790M	resistant
talazoparib	KRAS	A146T	resistant

.....

Goal: Democratize cutting-edge cancer care

EMR: 60-80% In Unstructured Text



Wolters Kluwer: Health Language Blog

```
1,23224,174680,2147-12-05,,, "Discharge summary", "Report", "", "Admissi  
on Date:  [**2823-9-29**]      Discharge Date:  [**2823-10-1  
7**]
```

```
Date of Birth:  [**2768-10-11**]      Sex:  F
```

```
Service: SURGERY
```

```
Allergies:
```

```
Patient recorded as having No Known Allergies to Drugs
```

```
Attending:[**First Name3 (LF) 1**]
```

```
Chief Complaint:
```

```
headache and neck stiffness
```

```
Major Surgical or Invasive Procedure:
```

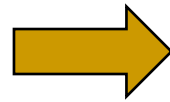
```
central line placed, arterial line placed
```

```
History of Present Illness:
```

```
54 year old female with recent diagnosis of ulcerative colitis  
on 6-mercaptopurine, prednisone 40-60 mg daily, who presents  
with a new onset of headache and neck stiffness. The patient is  
in distress, rigoring and has aphasia and only limited history  
is obtained. She reports that she was awoken 1AM the morning of  
[**2823-9-28**] with a headache which she describes as bandlike. She  
states that headaches are unusual for her. She denies photo- or  
phonophobia. She did have neck stiffness. On arrival to the ED  
at 5:33PM, she was afebrile with a temp of 96.5, however she  
later spiked with temp to 104.4 (rectal), HR 91, BP 112/54, RR  
24, O2 sat 100 %. Head CT was done and revealed attenuation  
within the subcortical white matter of the right medial frontal  
lobe. LP was performed showing opening pressure 24 cm H2O WBC of  
316, Protein 152, glucose 16. She was given Vancomycin 1 gm IV,  
Ceftriaxone 2 gm IV, Acyclovir 800 mg IV, Ambesone 183 IV,  
Ampicillin 2 gm IV q 4, Morphine 2-4 mg Q 4-6, Tylenol 1 gm ,  
Decadron 10 mg IV. The patient was evaluated by Neuro in the  
ED.
```


Real-World Evidence

1,23224,174680,2147-12-05,,, "Discharge summary", "Report", "", "Admission Summary"
 1,23224,174680,2147-12-05,,, "Discharge summary", "Report", "", "Admission Summary"
 1,23224,174680,2147-12-05,,, "Discharge summary", "Report", "", "Admission Summary"
 Date on Date: [**2823-9-29**] Discharge Date: [**2823-10-17**]
 Service: SURGERY
 Sex: F
 Service: SURGERY
 Allergies:
 Patient recorded as having No Known Allergies to Drugs
 Attending:[**First Name3 (LF) 1**]
 Chief Complaint:
 headache and neck stiffness
 Major Surgical or Invasive Procedure:
 central line placed, arterial line placed
 History of Present Illness:
 54 year old female with recent diagnosis of ulcerative colitis on 6-mercaptopurine, prednisone 40-60 mg daily, who presents with a new onset of headache and neck stiffness. The patient is in distress, rigoring and has aphasia and only limited history is obtained. She reports that she was awoken 1AM the morning of [**2823-9-28**] with a headache which she describes as bandlike. She states that headaches are unusual for her. She denies photophobia. She did have neck stiffness. On arrival to the ED at 5:33PM, she was afebrile with a temp of 96.5, however she later spiked with temp to 104.4 (rectal), HR 91, BP 112/54, RR 24, O2 sat 100 %. Head CT was done and revealed attenuation within the subcortical white matter of the right medial frontal lobe. LP was performed showing opening pressure 24 cm H2O WBC of 316, Protein 152, glucose 16. She was given Vancomycin 1 gm IV, Ceftriaxone 2 gm IV, Acyclovir 800 mg IV, Ambesone 183 IV, Ampicillin 2 gm IV q 4, Morphine 2-4 mg Q 4-6, Tylenol 1 gm, Decadron 10 mg IV. The patient was evaluated by Neuro in the ED.



What if we can unlock valuable info in EMRs?

Diagnoses
Treatments
Outcomes

• • • • •

Only 3% cancer patients enroll in clinical trials

Goal: Make every patient count

Cancer Registry



Stakeholder: Providers

Required by law: report basic cancer info

Manual: slow & tedious

```
1,23224,174680,2147-12-05,,, "Discharge summary", "Report", "", "Admission Date: [**2823-9-29**] Discharge Date: [**2823-10-17**]"
Date of Birth: [**2768-10-11**] Sex: F
Service: SURGERY
Allergies:
Patient recorded as having No Known Allergies to Drugs
Attending:[**First Name3 (LF) 1**]
Chief Complaint:
headache and neck stiffness
Major Surgical or Invasive Procedure:
central line placed, arterial line placed
History of Present Illness:
54 year old female with recent diagnosis of ulcerative colitis on 6-mercaptopurine, prednisone 40-60 mg daily, who presents with a new onset of headache and neck stiffness. The patient is in distress, rigoring and has aphasia and only limited history is obtained. She reports that she was awoken 1AM the morning of [**2823-9-28**] with a headache which she describes as bandlike. She states that headaches are unusual for her. She denies photo- or phonophobia. She did have neck stiffness. On arrival to the ED at 5:33PM, she was afebrile with a temp of 96.5, however she later spiked with temp to 104.4 (rectal), HR 91, BP 112/54, RR 24, O2 sat 100 %. Head CT was done and revealed attenuation within the subcortical white matter of the right medial frontal lobe. LP was performed showing opening pressure 24 cm H2O WBC of 316, Protein 152, glucose 16. She was given Vancomycin 1 gm IV, Ceftriaxone 2 gm IV, Acyclovir 800 mg IV, Ambesone 183 IV, Ampicillin 2 gm IV q 4, Morphine 2-4 mg Q 4-6, Tylenol 1 gm, Decadron 10 mg IV. The patient was evaluated by Neuro in the ED.
```



ATTRIBUTE	VALUE
Site	Large intestine (C180)
Morphology	Glassy cell (8015)
Staging	IIA (T2 N0 M0)

Off-Label Drug Use

Stakeholder: providers, oncologists

Standard of care often fails

N-of-one trials → Find patients like X

```
1,23224,174680,2147-12-05,,, "Discharge summary", "Report", "", "Admission Date: [**2823-9-29**] Discharge Date: [**2823-10-17**]"
Date of Birth: [**2768-10-11**] Sex: F
Service: SURGERY
Allergies:
Patient recorded as having No Known Allergies to Drugs
Attending:[**First Name3 (LF) 1**]
Chief Complaint:
headache and neck stiffness
Major Surgical or Invasive Procedure:
central line placed, arterial line placed
History of Present Illness:
54 year old female with recent diagnosis of ulcerative colitis
on 6-mercaptopurine, prednisone 40-60 mg daily, who presents
with a new onset of headache and neck stiffness. The patient is
in distress, rigoring and has aphasia and only limited history
is obtained. She reports that she was awoken 1AM the morning of
[**2823-9-28**] with a headache which she describes as bandlike. She
states that headaches are unusual for her. She denies photo- or
phonophobia. She did have neck stiffness. On arrival to the ED
at 5:33PM, she was afebrile with a temp of 96.5, however she
later spiked with temp to 104.4 (rectal), HR 91, BP 112/54, RR
24, O2 sat 100 %. Head CT was done and revealed attenuation
within the subcortical white matter of the right medial frontal
lobe. LP was performed showing opening pressure 24 cm H2O WBC of
316, Protein 152, glucose 16. She was given Vancomycin 1 gm IV,
Ceftriaxone 2 gm IV, Acyclovir 800 mg IV, Ambesone 183 IV,
Ampicillin 2 gm IV q 4, Morphine 2-4 mg Q 4-6, Tylenol 1 gm,
Decadron 10 mg IV. The patient was evaluated by Neuro in the
ED.
```



Diagnoses
Treatments
Outcomes

.....

Clinical Trial Matching

Marty Tenenbaum

Late-stage melanoma (late 1990s)

Initial prognosis: 6 months

Saved by Phase III trial of Canvaxin



20% cancer trials failed due to insufficient patients

Currently, rely on manual word-of-mouth

The diagram illustrates the matching process between two structured data sets. On the left, a yellow arrow points into a box labeled "Structured Patient Info". On the right, a yellow arrow points into a box labeled "Structured Eligibility Criteria". A large green double-headed arrow connects the two boxes, with the word "Match?" written in red below it.

52

Scenario: Drug Development

Average cost of an
FDA-approved drug

Annual number of
FDA-approved drugs

Per Year

\$2.5-10 billion

×

~50

= \$125-500 billion

Phase-3 Trial

Case: New drug

Control: Standard-of-care

Thousands of patients
Cost hundreds of million
per trial

Can we get this for free?

Synthetic Control

Stakeholder: Pharma, Providers

EMR: Standard of care \Rightarrow Virtual control arm

Case study: Flatiron

- Demonstrated efficacy in pivotal study with Pfizer
- Acquired by Roche for \$2B in 2018
- Manual curation by hundreds of abstracters; 2-3 hours per patient

Can we speed up curation by 10X?

Post-Market Surveillance

Stakeholder: Pharma, Providers

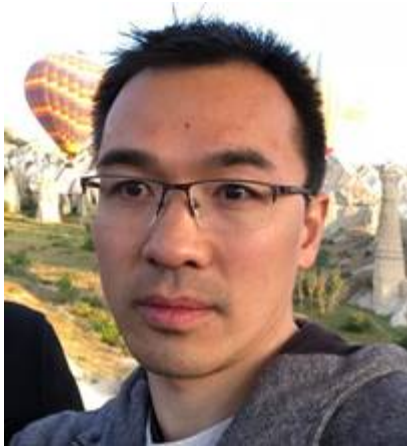
Assess drug performance in real population

FDA favors more conditional approval + PMS

Phase-5 trials: often cost more than 1-3 combined

Goal: Make drug development sustainable

Hanover Team



Cliff Wong



Tristan Naumann



Rajesh Rao

Collaborators

JAX: Susan Mockus, Sara Patterson

Fred Hutchinson: Christopher Li, Kathi Malone

Knight Cancer Institute: Brian Druker, Jeff Tyner, Steve Kurtz

U. Chicago: Andrey Rzhetsky

UCSC: Max Haeussler

MSR: Chris Quirk, Ravi Pandya, Bill Bolosky, Lucy Vanderwende, Robin Moeur, Curtis von Veh, Tony Carbary

Interns: Maxim Grechkin, Ankur Parikh, Victoria Lin, Sheng Wang, Stephen Mayhew, Daniel Fried, Violet Peng, Hai Wang, Robin Jia

Publications

EMRs

Clinical Trials

Precision Cancer Treatment

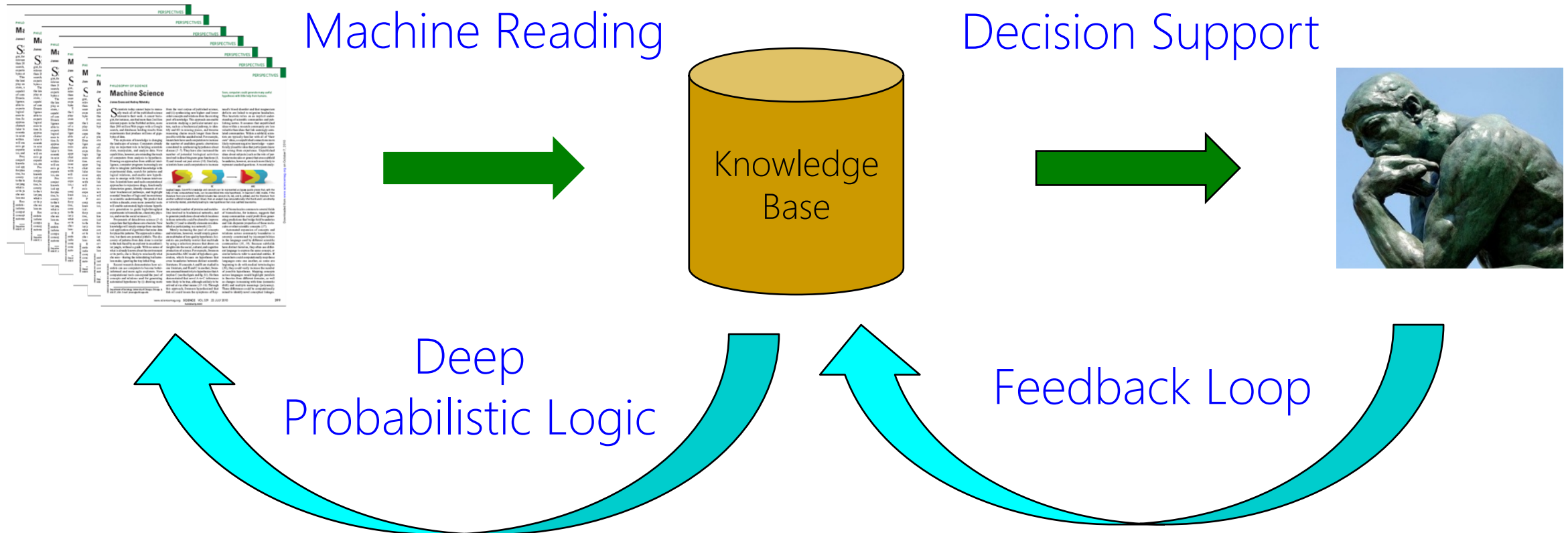
Structured Patient Info

Structured Eligibility

Molecular Tumor Board

Real-World Evidence

Clinical Trial Matching



<http://hanover.azurewebsites.net>