2024 ESMO TAT

Targeted Anticancer Therapies

AI IN EARLY PHASE CLINICAL TRIALS: NEW ANALYTICAL FRONTIERS

André Freitas

National Biomarker Centre, CRUK-MI

Department of Computer Science

University of Manchester

IDIAP Research Institute

ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

DECLARATION OF INTERESTS

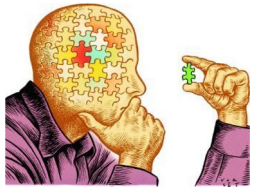
Andre Freitas

No interests to declare

Content of this presentation is copyright and responsibility of the author. Permission is required for re-use.

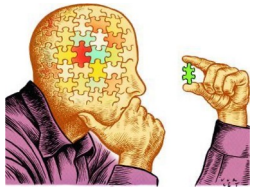
Prototypical analytical workflow

Hypotheses
Questions
New context



Prototypical analytical workflow

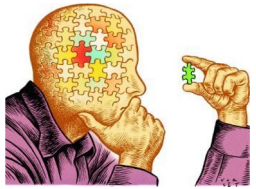
Hypotheses
Questions
New context



Select
relevant
background
knowledge

Prototypical analytical workflow

Hypotheses
Questions
New context

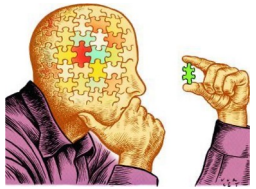


Select relevant background knowledge

Translate to an intervention

Prototypical analytical workflow

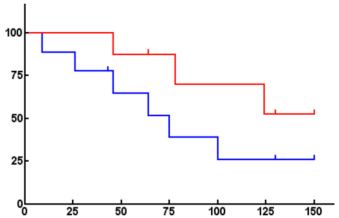
Hypotheses
Questions
New context



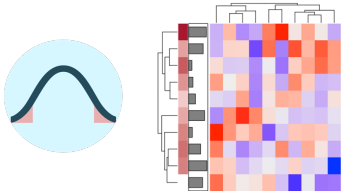
Select relevant background knowledge

Translate to an intervention

New Data
(phenomenal level)

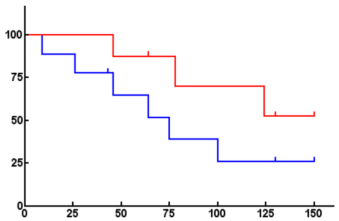


Prototypical analytical workflow



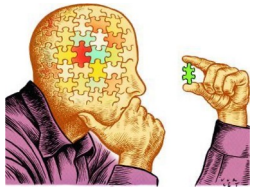
Infer
over data

New Data
(phenomenal level)



Translate to an intervention

Hypotheses
Questions
New context



Select
relevant
background
knowledge

PubMed®

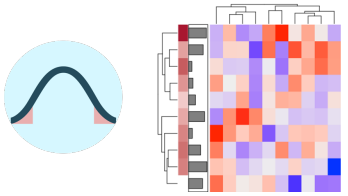
ESMO

CIViC
Clinical Interpretation of Variants in Cancer

COSMIC
Catalogue Of Somatic Mutations In Cancer

KEGG
Kyoto Encyclopedia of Genes and Genomes

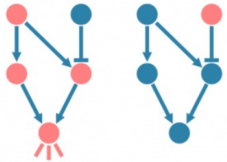
Prototypical analytical workflow



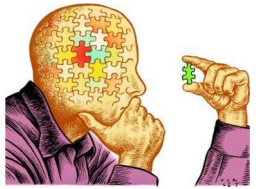
Explain & **Extend**

results

existing model



Hypotheses
Questions
New context



Select
relevant
background
knowledge

PubMed®

ESMO

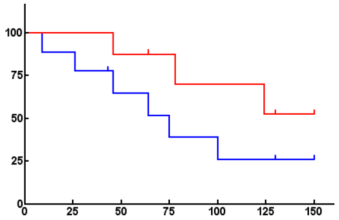
CIViC
Clinical Interpretation of Variants in Cancer

COSMIC
Catalogue Of Somatic Mutations In Cancer

KEGG
Kyoto Encyclopedia of Genes and Genomes

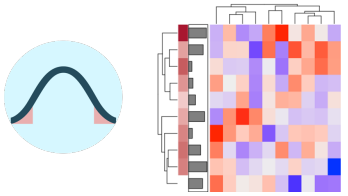
Translate to an intervention

Infer
over data



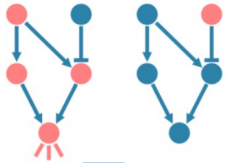
New Data
(phenomenal level)

Prototypical analytical workflow



Explain & **Extend**

results existing model



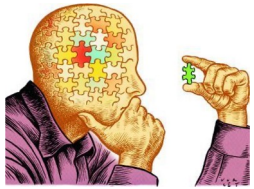
Corroborate

with previous evidence

PubMed®

ESMO

Hypotheses
Questions
New context



Select
relevant
background
knowledge

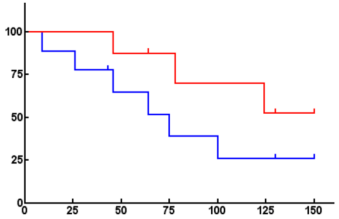
CIViC
Clinical Interpretation of Variants in Cancer

COSMIC
Catalogue Of Somatic Mutations In Cancer

Infer
over data

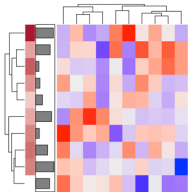
Translate to an intervention

KEGG
Kyoto Encyclopedia of Genes and Genomes



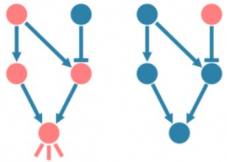
New Data
(phenomenal level)

Prototypical analytical workflow



Explain & **Extend**

results existing model



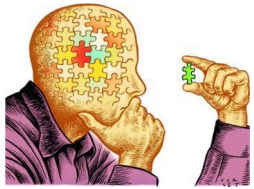
Corroborate

with previous evidence

PubMed®

ESMO

Hypotheses
Questions
New context



Select
relevant
background
knowledge

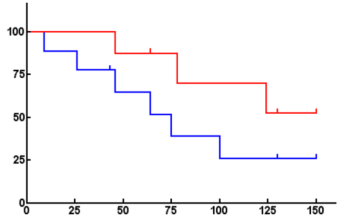
CIViC
Clinical Interpretation of Variants in Cancer

COSMIC
Catalogue Of Somatic Mutations In Cancer

Infer
over data

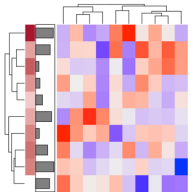
Translate to an intervention

KEGG
Kyoto Encyclopedia of Genes and Genomes



New Data
(phenomenal level)

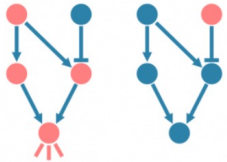
Prototypical analytical workflow



Explain & **Extend**

results

existing model



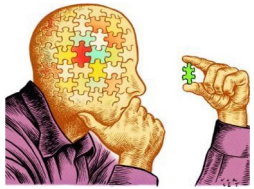
Corroborate

with previous evidence

PubMed®

ESMO

Hypotheses
Questions
New context



Evidence selection
Automating meta-analysis

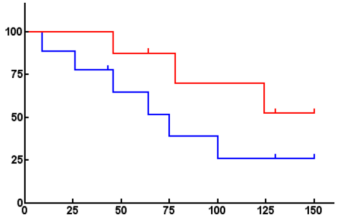
CIViC
Clinical Interpretation of Variants in Cancer

COSMIC
Catalogue Of Somatic Mutations In Cancer

Infer
over data

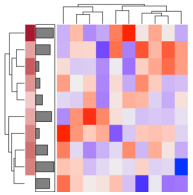
Translate to an intervention

KEGG
Kyoto Encyclopedia of Genes and Genomes



New Data
(phenomenal level)

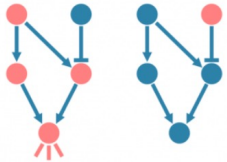
Prototypical analytical workflow



Explain & Extend

results

existing model



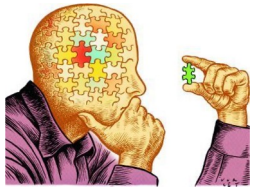
Corroborate

with previous evidence

PubMed®

ESMO

Hypotheses
Questions
New context



Evidence selection
Automating meta-analysis

CIViC
Clinical Interpretation of Variants in Cancer

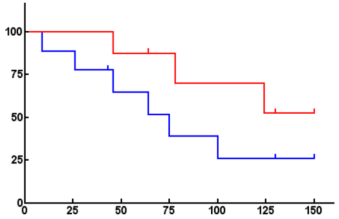
COSMIC
Catalogue Of Somatic Mutations In Cancer

Infer
over data

Automating experimental design

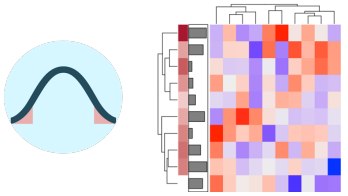
Supporting CT management

KEGG
Kyoto Encyclopedia of Genes and Genomes



New Data
(phenomenal level)

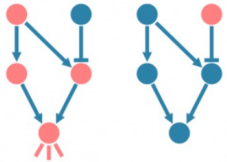
Prototypical analytical workflow



Explain & Extend

results

existing model



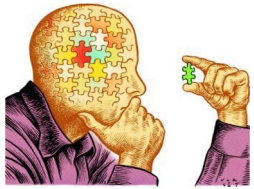
Corroborate

with previous evidence

PubMed®

ESMO

Hypotheses
Questions
New context



Evidence selection
Automating meta-analysis

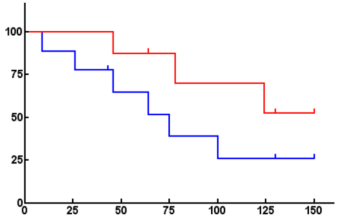
CIViC
Clinical Interpretation of Variants in Cancer

COSMIC
Catalogue Of Somatic Mutations In Cancer

Code generation
Abstraction models

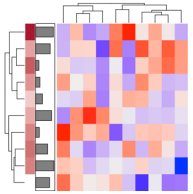
Automating experimental design

Supporting CT management

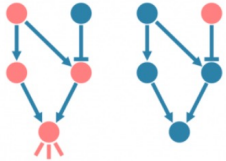


New Data
(phenomenal level)

Prototypical analytical workflow



Explanation generation
qualitative &
quantitative reasoning



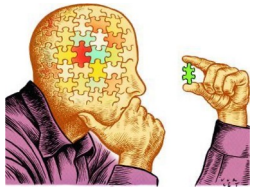
Corroborate

with previous
evidence

PubMed®

ESMO

Hypotheses
Questions
New context



CIViC
Clinical Interpretation of Variants in Cancer

Evidence selection
Automating meta-analysis

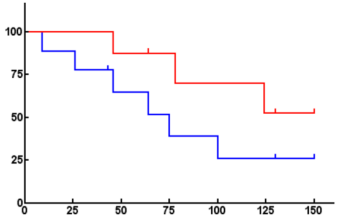
COSMIC
Catalogue Of Somatic Mutations In Cancer

Code generation
Abstraction models

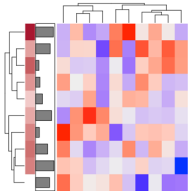
Automating
experimental design

Supporting CT
management

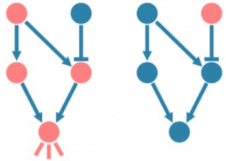
New Data
(phenomenal level)



Prototypical analytical workflow



Explanation generation
qualitative &
quantitative reasoning

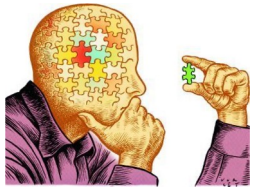


Abductive inference
Evidence selection

PubMed®

ESMO

Hypotheses
Questions
New context



Evidence selection
Automating meta-analysis

CIViC
Clinical Interpretation of Variants in Cancer

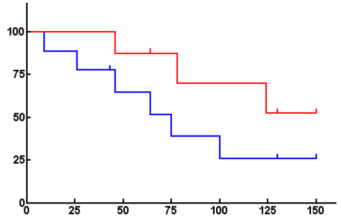
COSMIC
Catalogue Of Somatic Mutations In Cancer

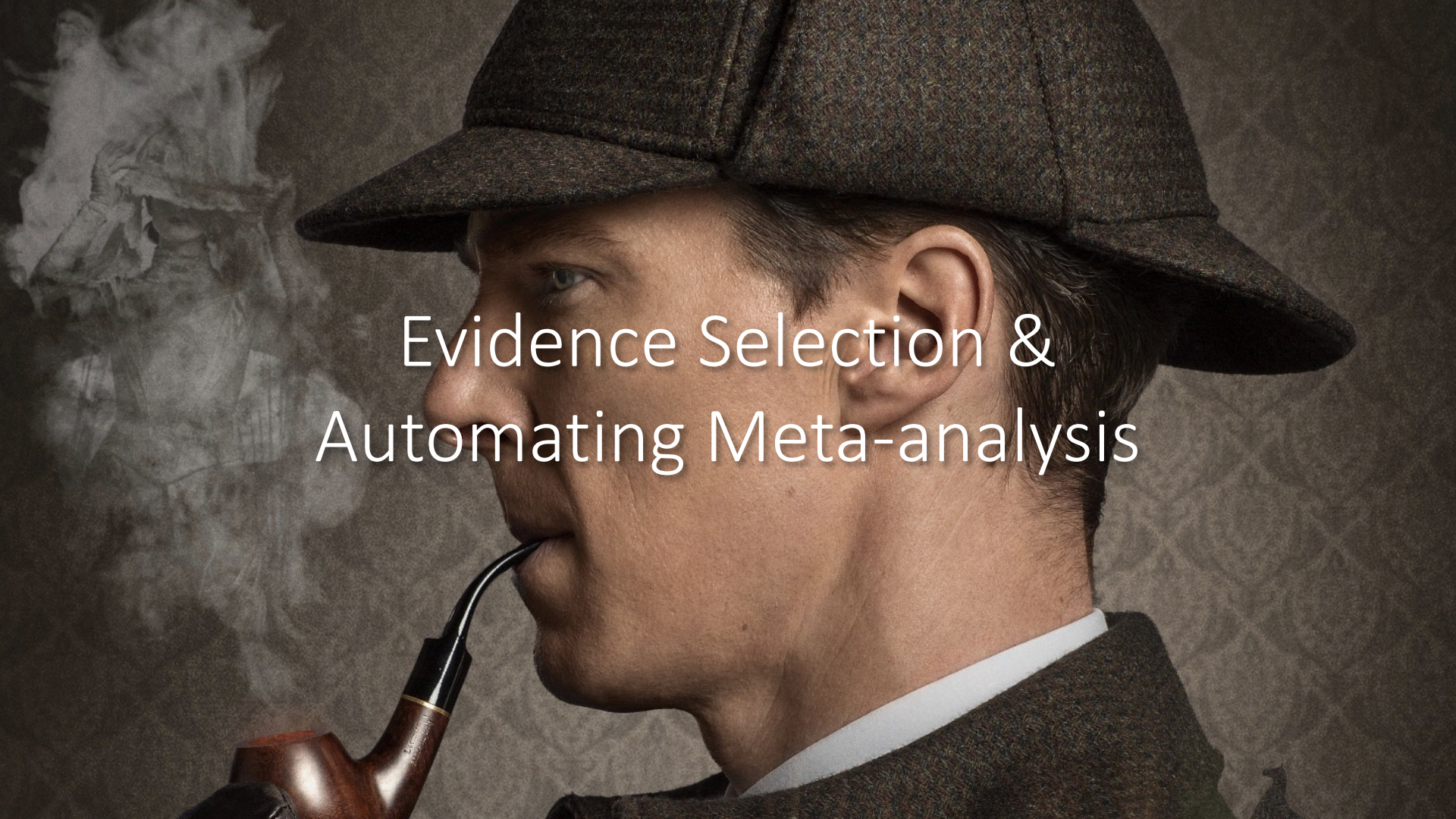
Code generation
Abstraction models

Automating
experimental design

Supporting CT
management

New Data
(phenomenal level)

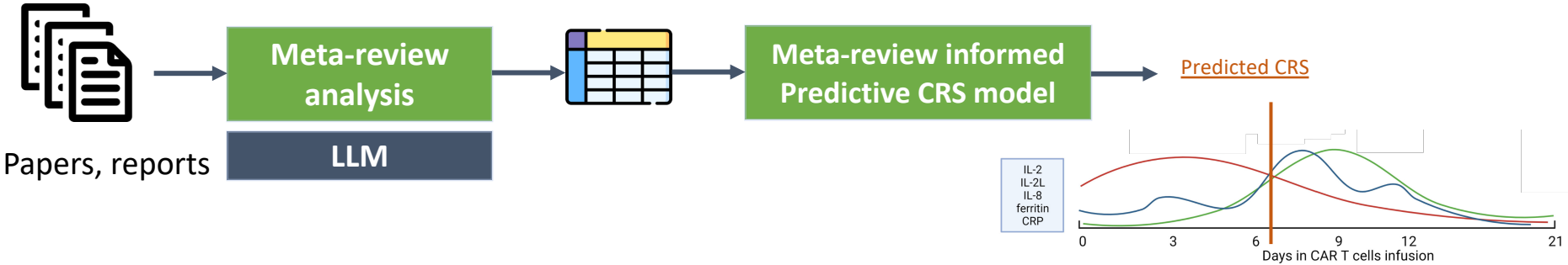




Evidence Selection & Automating Meta-analysis

Extracting evidence from the literature at scale

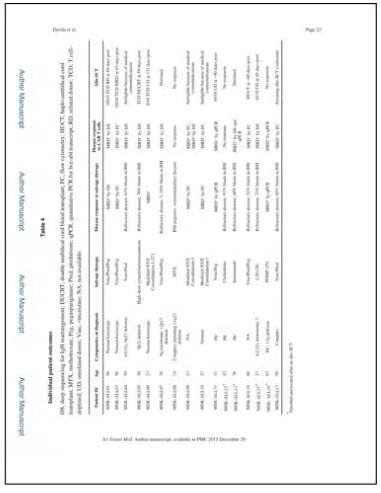
Predicting toxicity: Cytokine Release Syndrome (CRS) events for CAR-T cell therapies



	Study	IL2	IL4	IL6	IL8	IL10	IL15	IL2Rα	TNF–α	IFN–γ	GM-CSF
1	Jacobson et al. [37]	R	R	R	R	R	R	R	R	R	R
2	Hong et al. [38]	R	MV	R	MV	R	MV	MV	R	R	MV
3	Yan et al. [39]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
4	Topp et al. [40]	R	R	R	R	R	R	R	R	R	R
5	Shah et al. [41]	MV	MV	R	R	R	R	R	R	R	R
6	Liu et al. [29]	R	R	R	MV	R	MV	MV	R	R	MV
7	Sang et al. [13]	MV	MV	R	MV	MV	MV	MV	MV	R	MV
8	Yan et al. [42]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
9	Zhao et al. [43]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
10	Neelapu et al. [44]	R	MV	R	R	R	R	R	MV	R	R
11	Hay et al. [24]	MV	MV	R	R	R	R	MV	MV	R	MV
12	Turtle et al. [45]	MV	MV	R	MV	R	MV	MV	R	R	MV
13	Hu et al. [15]	MV	MV	R	MV	R	MV	MV	MV	R	MV
14	Teachey et al. [18]	R	R	R	R	R	MV	MV	R	R	R
15	Porter et al. [16]	R	MV	R	MV	MV	MV	R	MV	R	MV
16	Davila et al. [5]	MV	MV	R	MV	R	MV	MV	MV	R	R
17	Kalos et al. [46]	R	R	R	R	R	R	R	R	R	MV

Bogatu et al. (JBI, 2023)

Davila Marcoa L., et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in b cell acute lymphoblastic leukemia Sci. Transl. Med., 6 (224) (2014), 10.1126/scitranslmed.3008226 (ISSN: 1946-6234, 1946-6242)



Davila et al. Page 23

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4

Individual patient outcomes

DS, deep sequencing for IgH rearrangement; DUOXT, double umbilical cord blood transplant; FC, flow cytometry; HUCT, haplo-umbilical cord transplant; MTX, methotrexate; PNy, polymorphonuclear; PNeg, pancytopenic; qPCR, quantitative PCR for bcr-abl transcript; RD, refused donor; TCD, T cell-depleted; TKI, tyrosine kinase inhibitor; Y, yes; yrs, years; ↑, increased; ↓, decreased.

Patient ID	Age	Composition of Remission	Salvage therapies	Disease response at salvage therapy	Primary outcome (CR, CR with CRI, or NR)	OS/DFS
MSK-ALL01	20	Neuroblastomatosis	7-arm Blinatumomab	MRD+ by DS	CR with CRI	OS/DFS 8/03 at 4 days post
MSK-ALL07	56	Neuroblastomatosis	7-arm Blinatumomab	MRD+ by FC	CR with CRI	OS/DFS 9/03 at 5 days post
MSK-ALL04	64	↑LDH, Ph+ blastosis	Clofarabine	Refractory disease, 50% blastosis in BM	MRD+ by FC	Died 9/03 at 5 days post
MSK-ALL02	20	Ph+ blastosis	Ph+ ALL HUCT	MRD+	MRD+ by FC	Died 9/03 at 5 days post
MSK-ALL06	23	Neuroblastomatosis	Ph+ ALL HUCT, Dasatinib, Nelarabine, CNS radiation	MRD+	MRD+ by FC	Died 9/03 at 5 days post
MSK-ALL07	56	Ph+ blastosis, CR1?	7-arm Blinatumomab	Refractory disease, 5–10% blastosis in BM	MRD+ by FISH	Pleural
MSK-ALL08	74	↑Cytopenias, Dysplasia?	MTX	Mid-response, ↓extramedullary disease	No response	No response
MSK-ALL09	23	↑Cytopenias, Dysplasia?	Rituximab, Dasatinib, Blinatumomab	MRD+ by FC	MRD+ by FC, Refractory disease	Indefinite clearance of medical
MSK-ALL10	21	Normal	Rituximab, Dasatinib, Blinatumomab	MRD+ by FC	MRD+ by FC	Indefinite clearance of medical
MSK-ALL11	72	Ph+	7-arm Blinatumomab	MRD+ by qPCR	MRD+ by qPCR	10/13 ICA, 4/9 days post
MSK-ALL12	62	Ph+	Clofarabine	Refractory disease, 80% blastosis in BM	No response	No response
MSK-ALL13	24	Ph+	Ph+ ALL HUCT	Refractory disease, 80% blastosis in BM	MRD+ by FC and qPCR	10/13 ICA, 4/9 days post
MSK-ALL14	66	N/A	7-arm Blinatumomab	Refractory disease, 15% blastosis in BM	MRD+ by FC	10/13 ICA, 4/9 days post
MSK-ALL15[a]	27	↑CLL/Lymphocytosis, ↑ IgH/IgK ratio	Ibrutinib	Refractory disease, 15% blastosis in BM	MRD+ by DS	10/13 ICA, 4/9 days post
MSK-ALL16[a]	65	Ph+, ↑ blastosis	Ph+ ALL HUCT	MRD+ by qPCR	MRD+ by qPCR	No response
MSK-ALL17	48	↑Cytopenias	Nelarabine	Refractory disease, 50% blastosis in BM	MRD+ by FC	Alemtuzumab, 11/13 ICA, 4/9 days post

[a]Enrolled under medical rather than pediatric IRB.

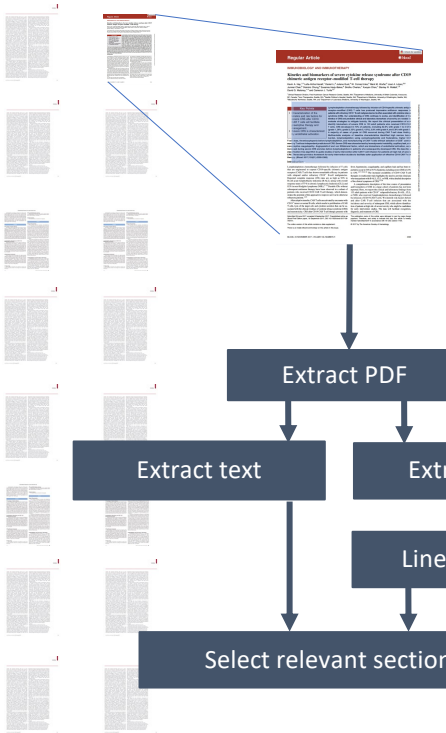
Biomark	Davila e	Hay et a	Hong et	Hu et al	Jacobso	Kalos et	Liu et al	Neelapu	Porter e	Sang et	Shah et	Teachey	Topp et	Turtle et	Yan et a	Yan et a	Zhao et
IL-2																	
IL-4																	
IL-6																	
IL-8																	
IL-10																	
IL-15																	
IL-2Rα																	
TNF-α																	
IFN-γ																	
GM-CS																	
CRP																	
ferritin																	
D-dime																	
VWF																	
Ang-2																	
MCP-1																	
granzy																	
TNFRp																	
MIP1β																	
MIP1α																	
IL-17																	
sIL6R																	

Meta-review

~ 460 papers

17 highly aligned papers

Parameter extraction



Extract PDF

Extract text

Extract tables

Linearise tables

Select relevant sections

context window

chain of prompts

Mistral 7B

LLM

Table builder

KB-query

NCI thesaurus

e.g. **TNF-α:**
'tumor necrosis factor–α',
'Tumor necrosis factor–α',
'TNF-α', 'TNFα', 'TNF-a', 'TNFa', 'TNF',
'Tumor necrosis factor alpha',
'tumor necrosis factor alpha'

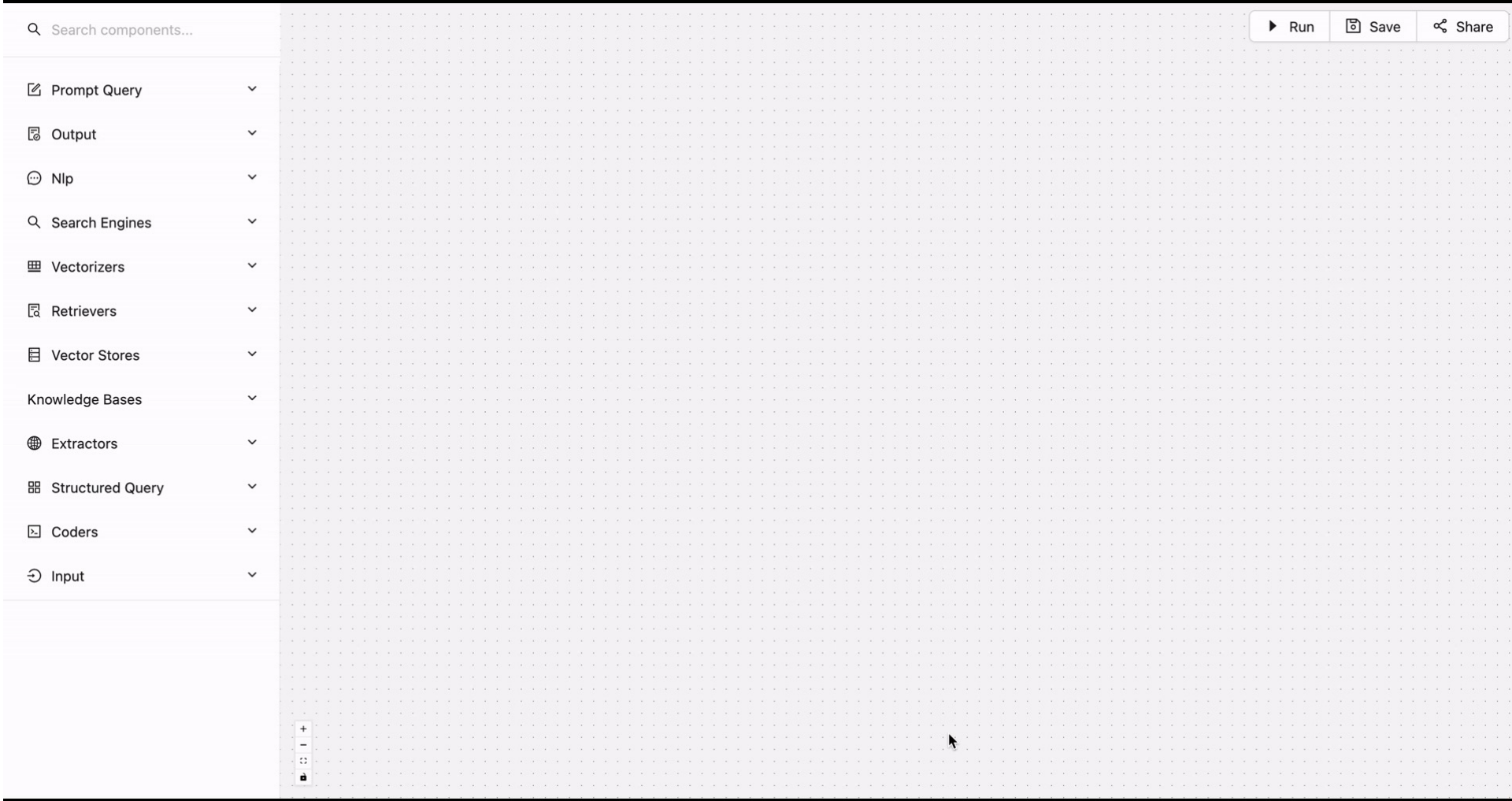
	Study	IL2	IL4	IL6	IL8	IL10	IL15	IL2Rα	TNF–α	IFN–γ	GM-CSF
1	Jacobson et al. [37]	R	R	R	R	R	R	R	R	R	R
2	Hong et al. [38]	R	MV	R	MV	R	MV	MV	R	R	MV
3	Yan et al. [39]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
4	Topp et al. [40]	R	R	R	R	R	R	R	R	R	R
5	Shah et al. [41]	MV	MV	R	R	R	R	R	R	R	R
6	Liu et al. [29]	R	R	R	MV	R	MV	MV	R	R	MV
7	Sang et al. [13]	MV	MV	R	MV	MV	MV	MV	MV	R	MV
8	Yan et al. [42]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
9	Zhao et al. [43]	MV	MV	R	MV	MV	MV	MV	MV	MV	MV
10	Neelapu et al. [44]	R	MV	R	R	R	R	R	MV	R	R
11	Hay et al. [24]	MV	MV	R	R	R	R	MV	MV	R	MV
12	Turtle et al. [45]	MV	MV	R	MV	R	MV	MV	R	R	MV
13	Hu et al. [15]	MV	MV	R	MV	R	MV	MV	MV	R	MV
14	Teachey et al. [18]	R	R	R	R	R	MV	MV	R	R	R
15	Porter et al. [16]	R	MV	R	MV	MV	MV	R	MV	R	MV
16	Davila et al. [5]	MV	MV	R	MV	R	MV	MV	MV	R	R
17	Kalos et al. [46]	R	R	R	R	R	R	R	R	R	MV

325x efficiency gain

Demo Wysocki, Wysocka, Carvalho, Bogatu, Miranda

Lunar

AI coordination infrastructure


Search components...
Prompt Query
Output
Nlp
Search Engines
Vectorizers
Retrievers
Vector Stores
Knowledge Bases
Extractors
Structured Query
Coders
Input
Run
Save
Share

Matching patients to clinical trials

64 years old woman with:

- multiple myeloma,
- s/p allogeneic transplant with recurrent disease and with systemic amyloidosis (involvement of lungs, tongue, bladder, heart),
- on hemodialysis for ESRD who represents for malaise, weakness, and generalized body aching x 2 days.
- she was admitted with hypercalcemia and treated with pamidronate 30mg, calcitonin, and dialysis.
- patient was initially treated with melphalan and prednisone, followed by VAD regimen, and autologous stem cell transplant.
- with relapse of her myeloma, she received thalidomide velcade and thalidomide, which were eventually also held due to worsening edema and kidney function.

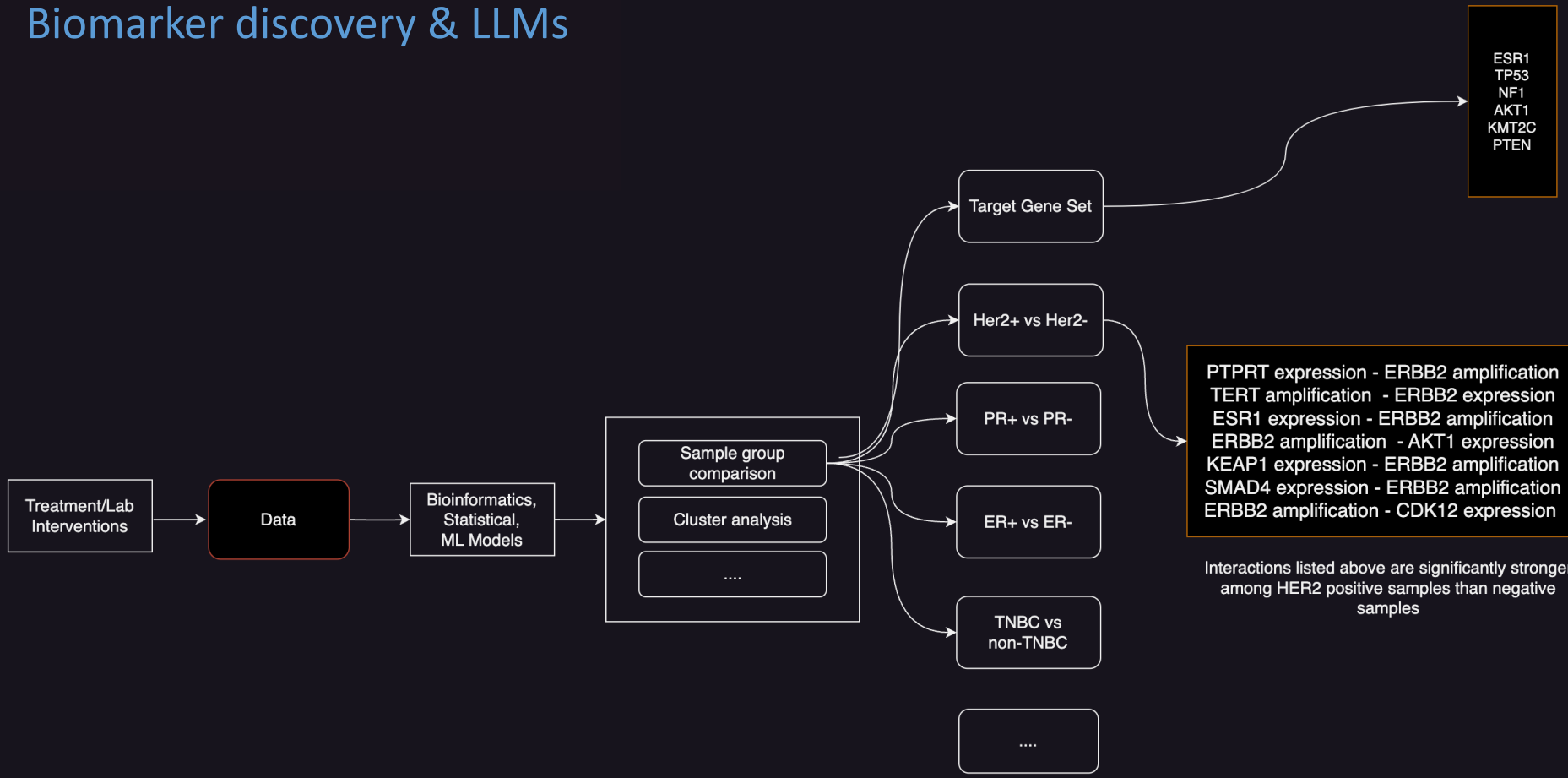
~ 375,600 CTs

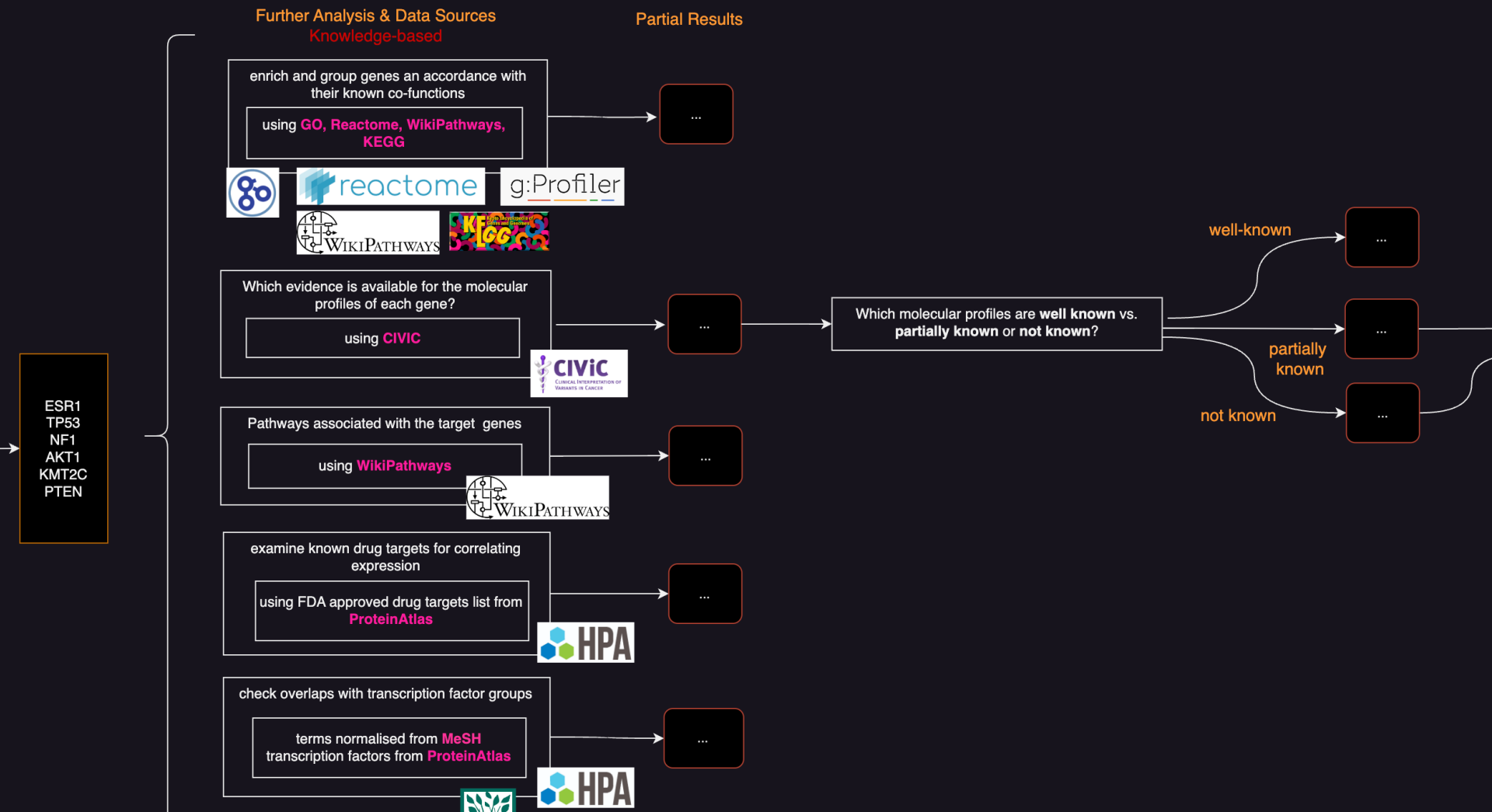
CLINICAL TRIALS **TRANSFORMATION** INITIATIVE

Demo from Bogatu, Jullien
Jullien et al. (Semeval 2022, EMNLP 2023)

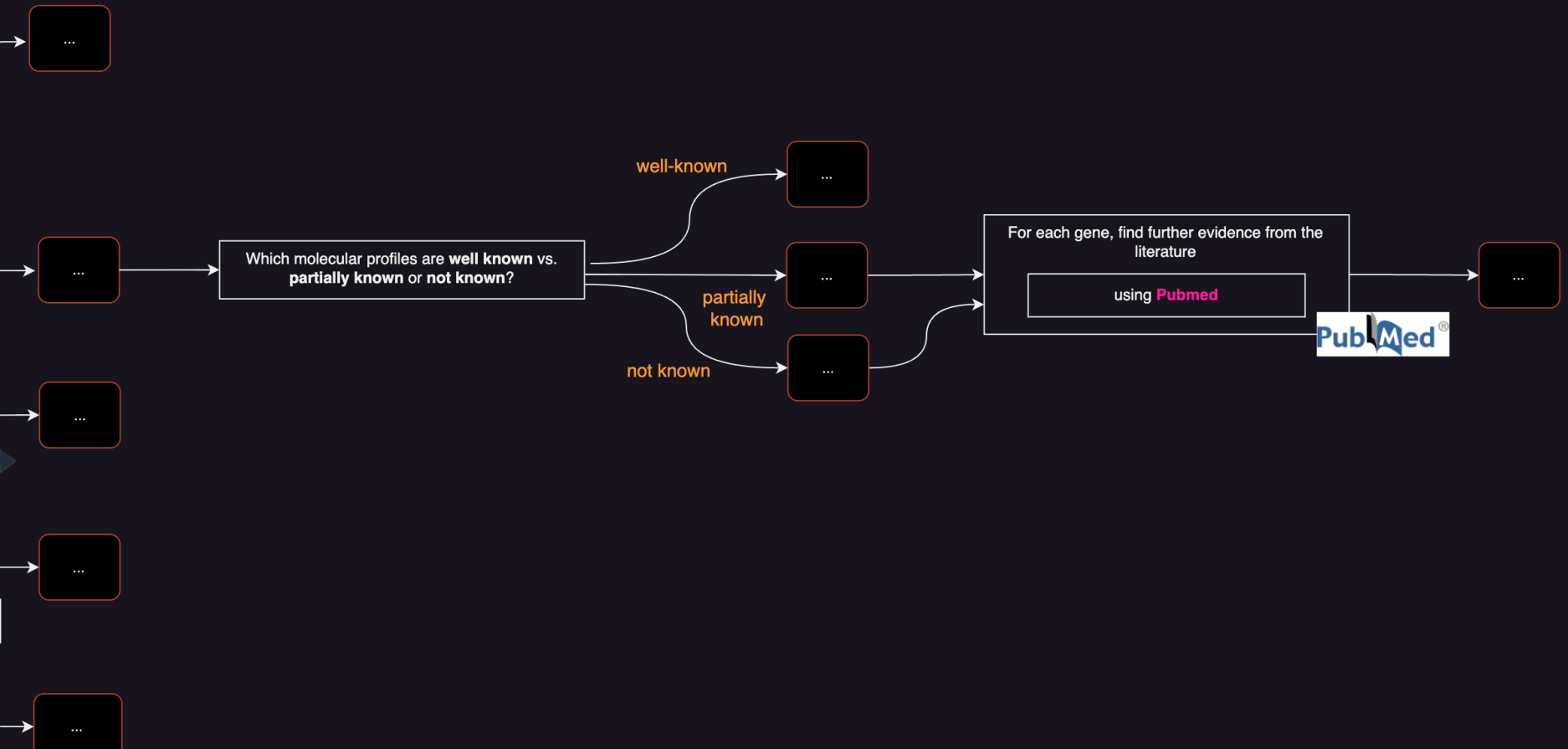
Evidence-based Scientific Reasoning

Biomarker discovery & LLMs




Further Analysis & Data Sources
Knowledge-based
Partial Results
ESR1
TP53
NF1
AKT1
KMT2C
PTEN
enrich and group genes an accordance with their known co-functions
using GO, Reactome, WikiPathways, KEGG
reactome
g:Profiler
WikiPathways
...
Which evidence is available for the molecular profiles of each gene?
using CIVIC
CIViC
...
Which molecular profiles are well known vs. partially known or not known?
well-known
partially known
not known
...
...
...
Pathways associated with the target genes
using WikiPathways
WikiPathways
...
examine known drug targets for correlating expression
using FDA approved drug targets list from ProteinAtlas
HPA
...
check overlaps with transcription factor groups
terms normalised from MeSH
transcription factors from ProteinAtlas
HPA
...

Partial Results



Further Analysis & Data Sources
Knowledge-based

Partial Results

enrich and group genes an accordance with their known co-functions

using GO, Reactome, WikiPathways, KEGG

...

Which evidence is available for the molecular profiles of each gene?

using CIVIC

Not known molecular profiles
- **Gene: MTAP:** methylthioadenosine phosphorylase
-- MolecularProfileName: MTAP Deletion; no Evidence in CivicDB

- **Gene: KMT2C:** lysine methyltransferase 2C
-- MolecularProfileName: KMT2C Loss-of-function; no Evidence in CivicDB

For each gene, find further evidence from the literature

using Pubmed

PRMT5: An Emerging Target for Pancreatic Adenocarcinoma. 7 citations
DOI: 10.3390/cancers13205136

Loss of MAT2A compromises methionine metabolism and represents a vulnerability in H3K27M mutant glioma by modulating the epigenome. 6 citations
DOI: 10.1038/s43018-022-00348-3
...

Pathways associated with the genes

using WikiPathways

...

examine known drug targets for correlating expression

using FDA approved drug targets list from ProteinAtlas

Using the Human Protein Atlas as a reference:

Eight genes overlapped between the two sets, including CHST11, STK4, CASP8, TGFBR2, CIITA, BTK, CSF1R, and CYLD. The precision of the analysis was 0.127, indicating that 12.7% of the analyzed set was also present in the reference set. The recall was 0.008, indicating that only 0.8% of the reference set was present in the analyzed set. The p-value was marginally significant, indicating that the overlap between the two sets was not entirely random.

check overlaps with transcription factor groups

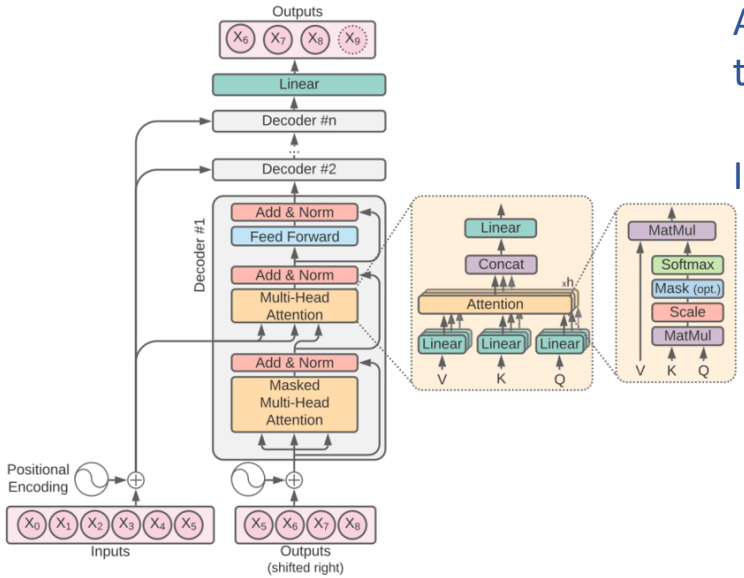
terms normalised from MeSH
transcription factors from ProteinAtlas

Using the Human Protein Atlas as a reference:

When compared the selected set of genes with the reference set of transcription factors it was found that 13 genes overlapped between the two sets, including EBF1, MAF, NFATC2, PAX5, LYL1, BCL11B, PRDM1, TCF7, IKZF1, FLI1, FOXO1, IRF4, and TFEB. The precision of the comparison was 0.206, indicating that 20.6% of the genes in the selected set were also present in the reference set. The recall was 0.009, indicating that only 0.9% of the reference set genes were also present in the selected set. The Fisher's Test resulted in a statistically significant p-value, indicating that the overlap between the two sets was not random.

Multi-omics data integration

The unreasonable effectiveness of transformers



Vaswani et al (NeurIPS, 2017)

Adapting **transformers for multi-omics data analysis** to improve the modelling of heterogeneous cell states/populations

Intuition: ground-up modelling of the 'cell-discourse'

Omics transformers

scGPT

scBERT

Enformer

GeneFormer

...

Brown et al (Arxiv:2005.14165, 2020)
Stienon et al (NeurIPS, 2020)

ChatGPT

Ouyang et al (NeurIPS, 2022)

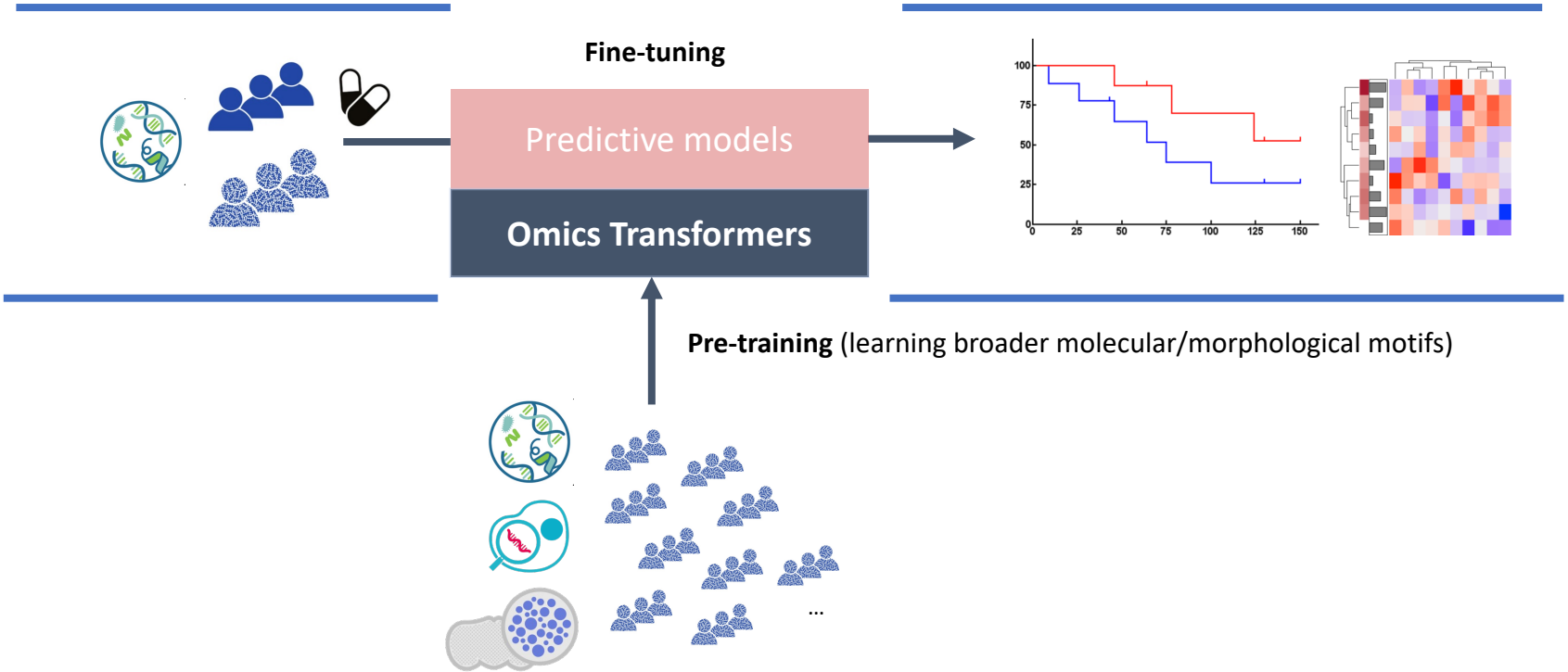
MISTRAL AI_

2017 2020 2022 2023

Figure credit : Lara-Benitez et al, 2020

Modelling complex cell-states for biomarker discovery

Modelling complex cell-states for biomarker discovery



Applications

- Biomarker discovery, improving patient stratification and tumour subtyping.
  - Improving biological understanding between responders and non-responders.
- Improving clinical trial design – integrating maximum available evidence.
- Dose optimisation.
- Improving the prediction of toxicity effects.
- Improving patient-treatment matching and patient accessibility to clinical trials.
- Improving the understanding and outcomes of clinical trials:
  - Optimising screening.
  - Reducing patient drop-out.
  - Understanding protocol deviation.
- Organising observational evidence in the clinic – structuring electronic health records.
- Major implications in terms of optimising drug discovery pipelines
  - E.g. Drug repurposing.

Concluding remarks

Take away

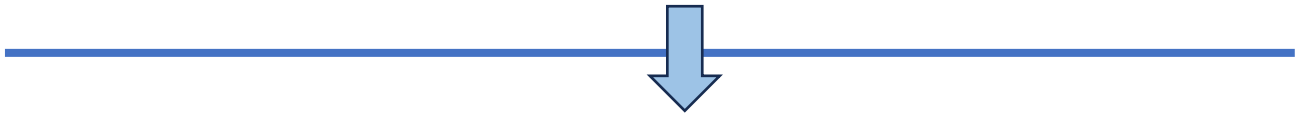
New emerging foundations for **analytical inference in oncology**

Universal framework for **integrating, organising and reasoning over heterogeneous evidence**

Significant analytical changes

Selection & integration of unstructured and structured evidence **at scale**

Better integration between **mechanistic** and **data-driven** inference



Common denominator

Addressing analytical barriers in personalised/experimental cancer medicine

More specific phenomena (personalised response, smaller cohorts)

More data per patient

($p >> n$)

Take away

Large Language Models

Are a **(monumental!)** game-changing foundation.
Transformers are an efficient substrate for modelling language, omics and reasoning.
Fluidity/lower impedance between representation modalities:

… Text, Databases, Pathways, Equations, Code …

Alone LLMs are not fit for purpose for biomedical reasoning.

Evidence-based reasoning

Biomedical AI-based reasoning **requires complex coordination pipelines**.
... and domain experts building them.
Implement once, reuse forever.
Allows using **maximum available evidence**.
Magnitude order efficiency gains.

Thank you for your attention!

Work in collaboration with:

Alex Bogatu, Caroline Dive, Danilo Carvalho, Donna Graham, Fouziah Butt, Loic Verlingue, Louise Carter, Marco Valentino, Maxime Delmas, Harriet Unsworth, Mael Julien, Magdalena Wysocka, Oskar Wysocki

Supported by:

Swiss National Science Foundation

UKRI Engineering and Physical Sciences Research Council

the ark

Innosuisse

Horizon Europe

CANCER RESEARCH UK

contact: andre.freitas@manchester.ac.uk

2024 ESMO TAT

Targeted Anticancer Therapies

andre.freitas@manchester.ac.uk

European Society for Medical Oncology (ESMO)
Via Ginevra 4, CH-6900 Lugano
T. +41 (0)91 973 19 00
esmo@esmo.org

esmo.org