AstraZeneca

BIKG

Drug Discovery as a Recommendation Problem

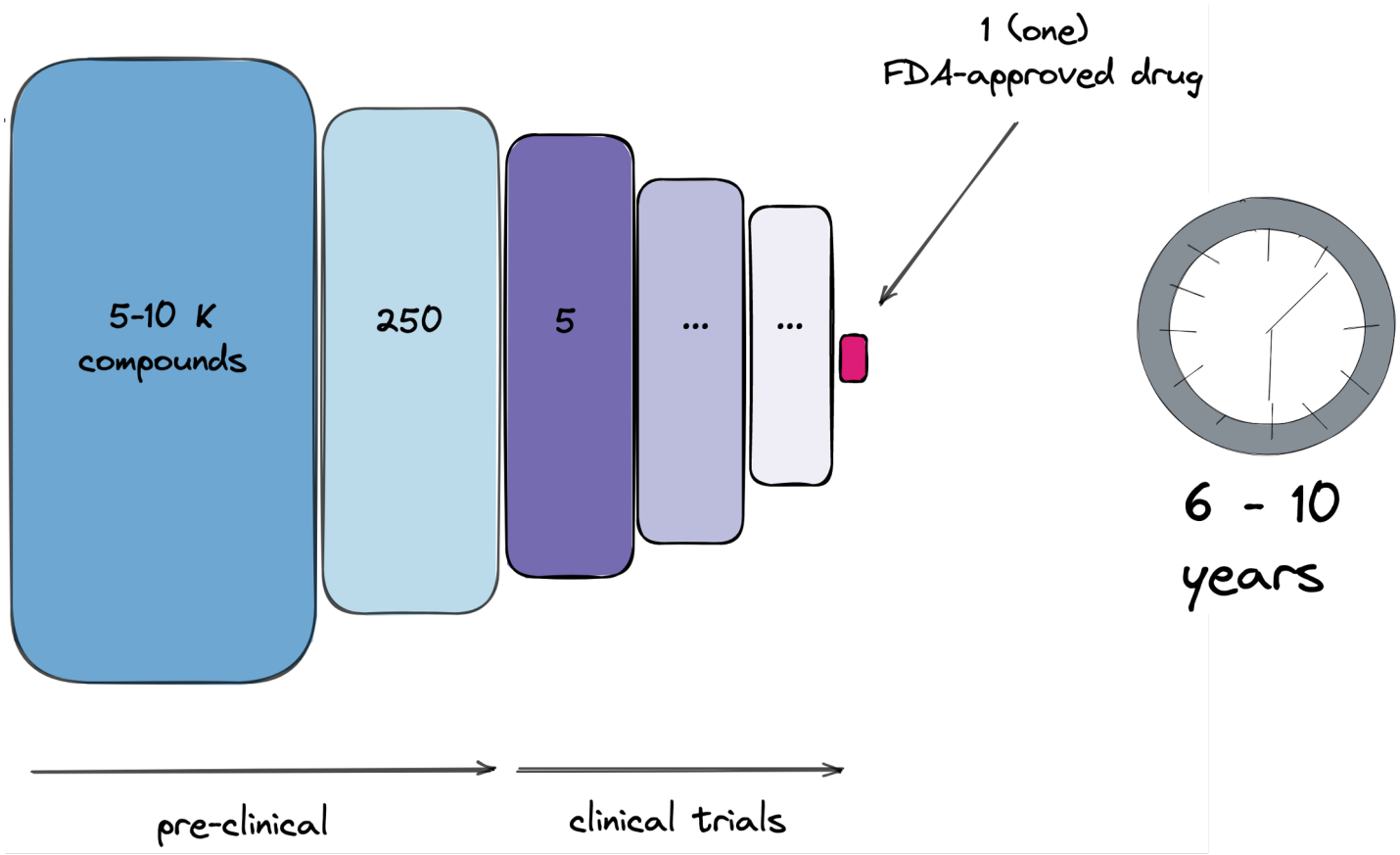
Anna Gogleva, Erik Jansson,
Greet De Baets, Eliseo Papa

27th September 2021

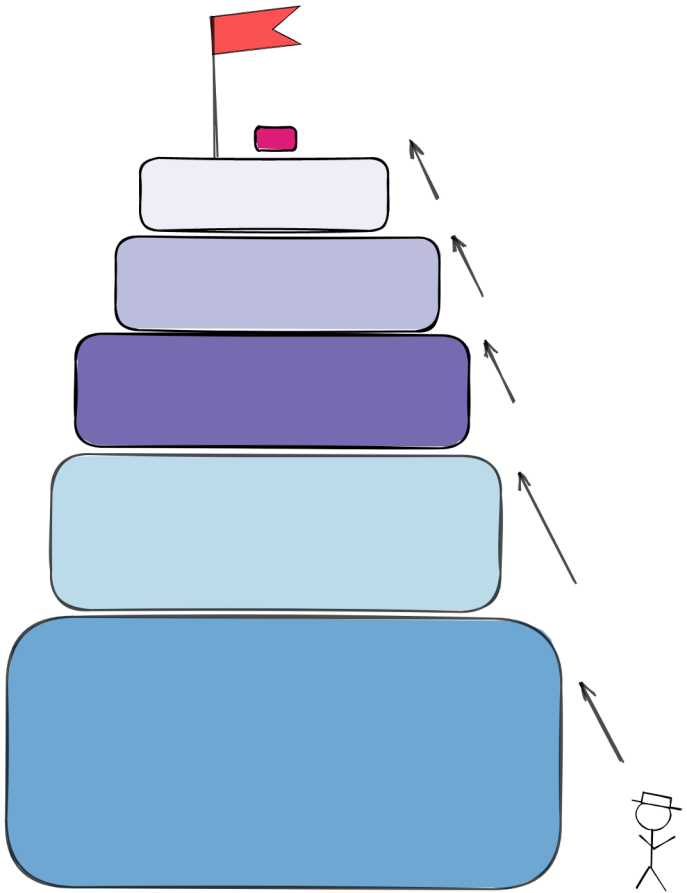
ACM RecSys'21 Amsterdam

https://astrazeneca.github.io/recsys21gogleva/

One needs to fail a lot to discover a working drug

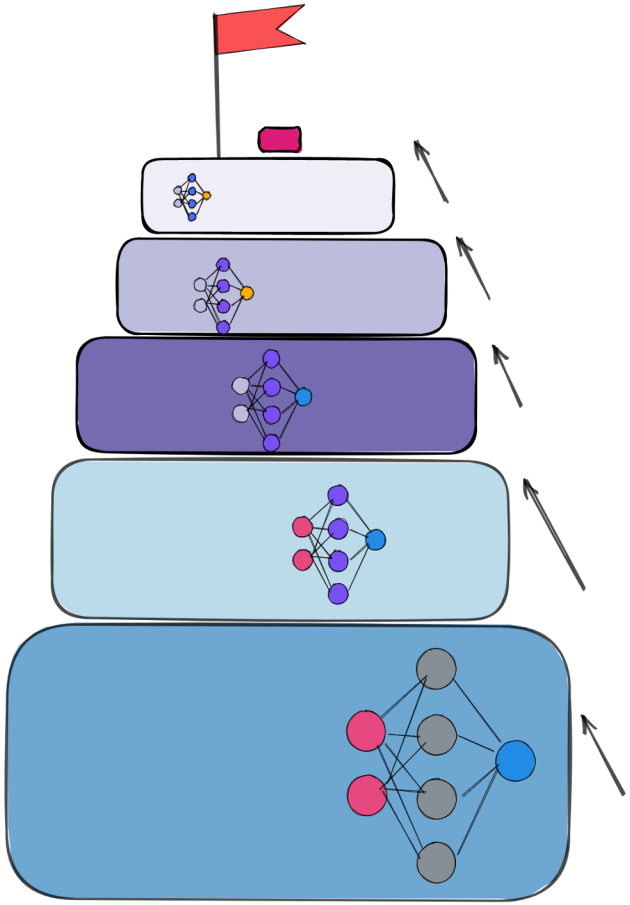


It is a tall mountain to climb

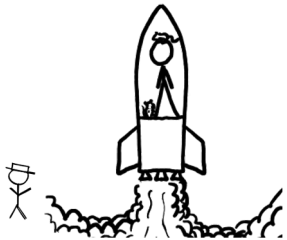


- How to develop new efficient treatments faster?
- How to make better decisions in the process?

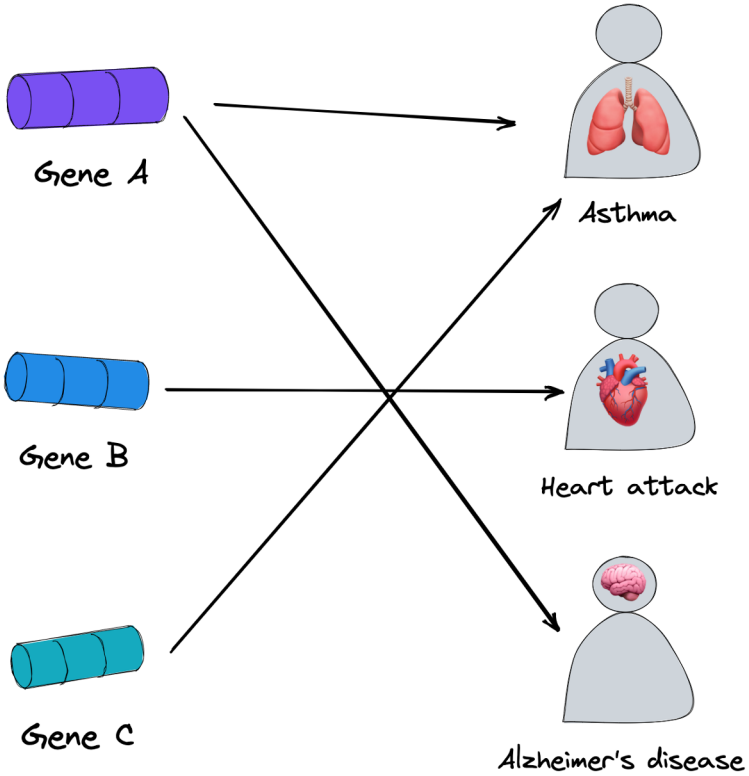
It is a tall mountain to climb



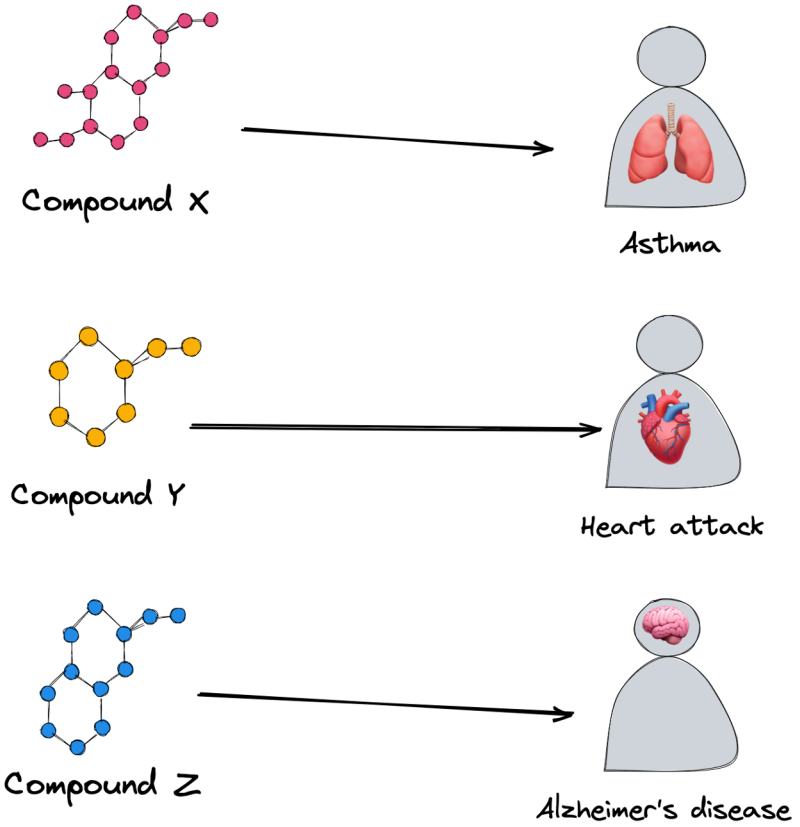
- How to develop new efficient treatments faster?
- How to make better decisions in the process?
- Recommendation systems can help in multiple places



Recommendation problems in drug discovery

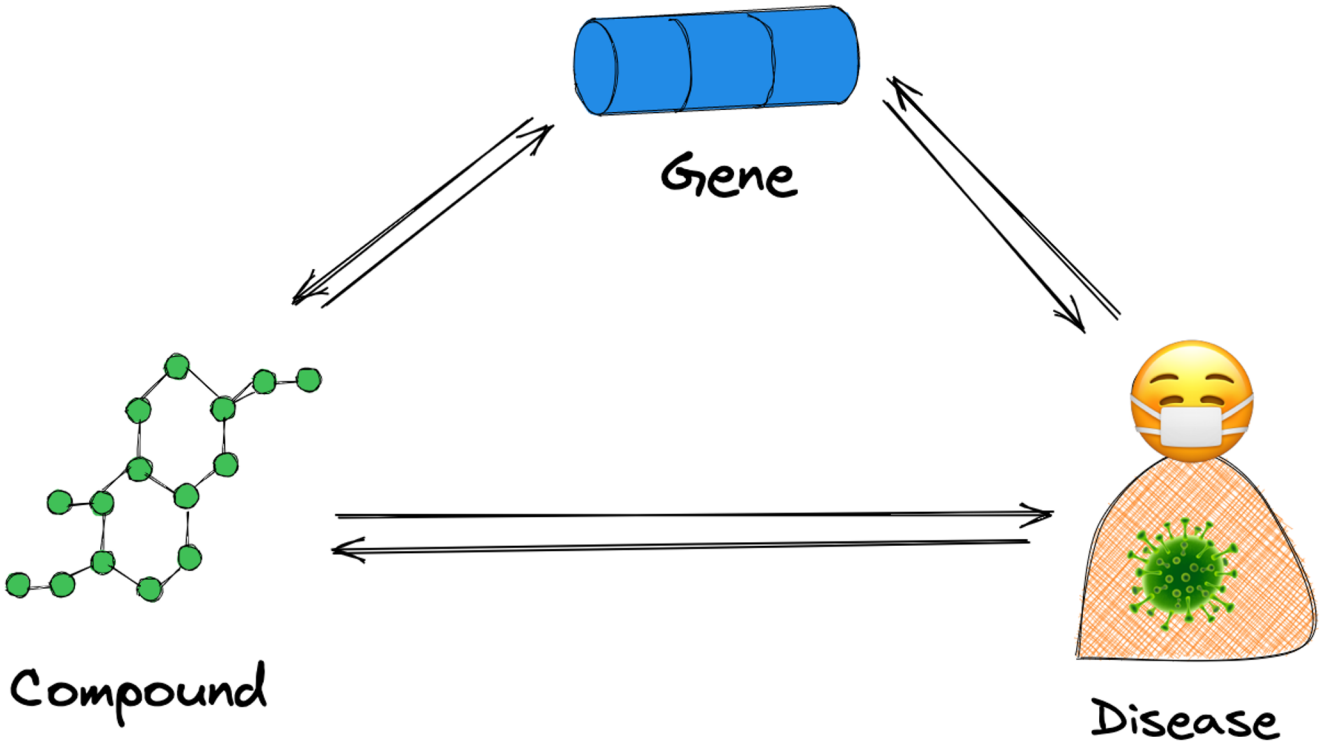


find a gene causing a disease



match a drug with a disease

Drugs, genes, diseases

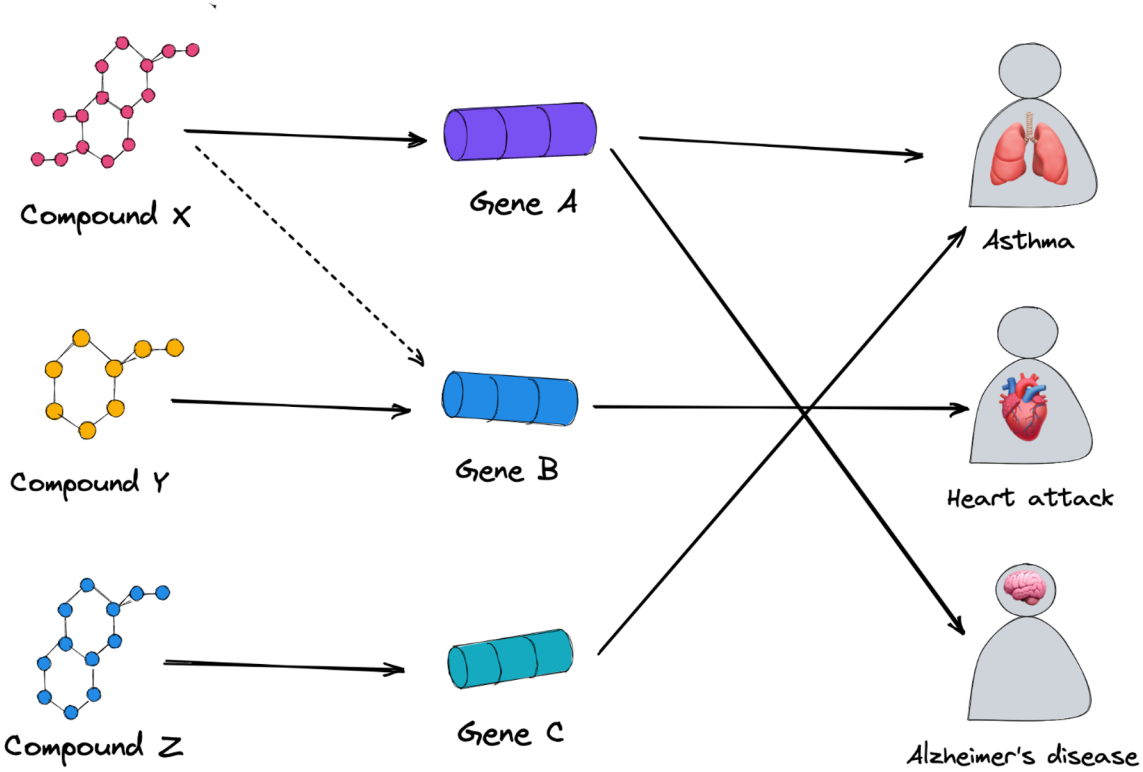


It gets complex very fast

Millions of compounds
Billions possible theoretically

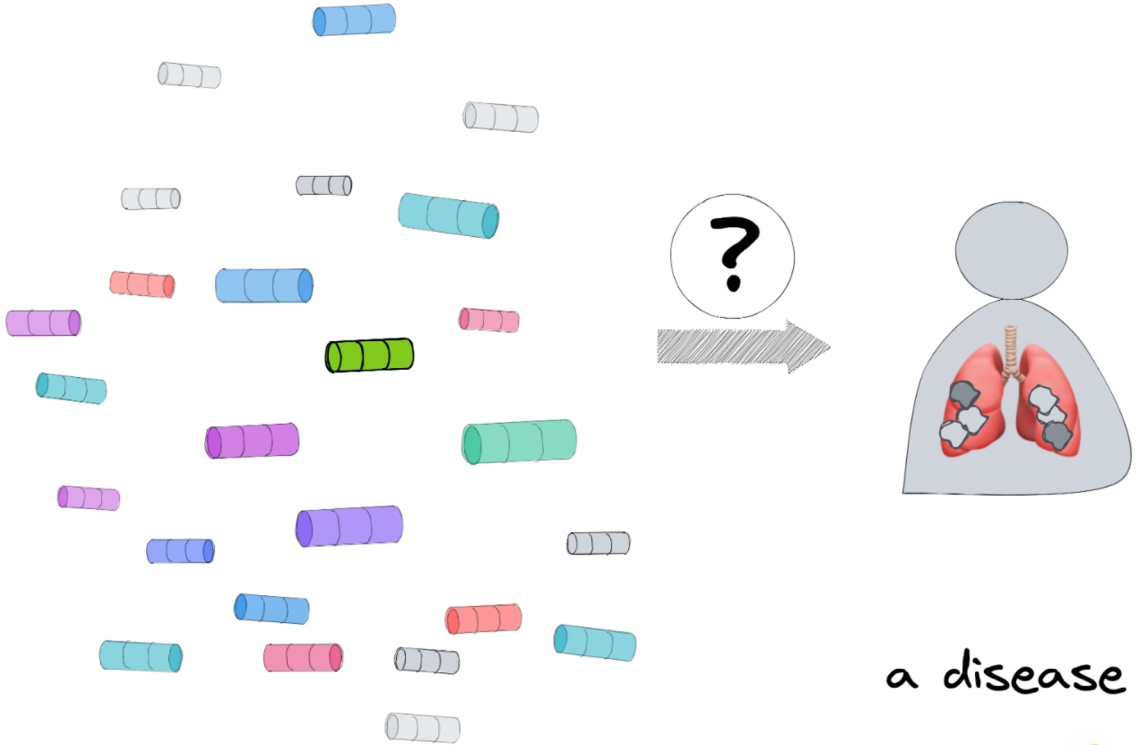
25-30 K genes,
80 K functional elements

~10 K diseases



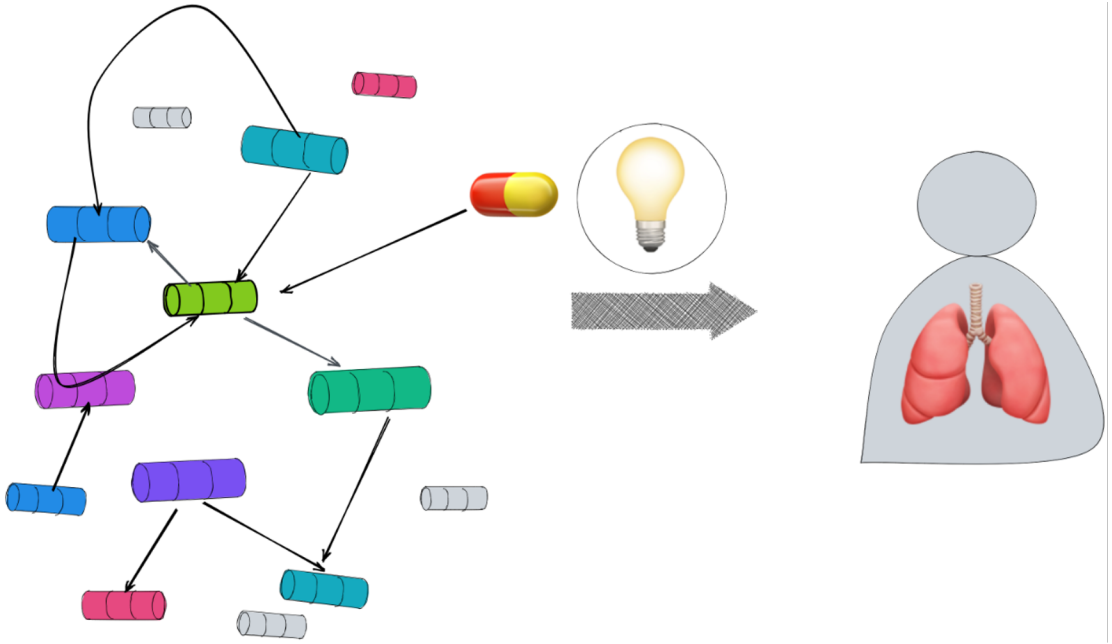
It is rarely just a single gene

- 25-30K human genes
- everything interacts with everything,
  each gene is a suspect

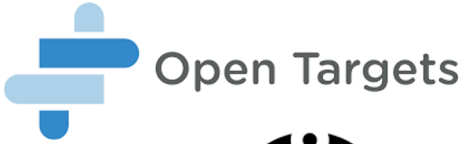


Find a molecular network behind a disease

1. disease ~ a molecular process gone awry
2. find the key molecular process
3. re-route it safely

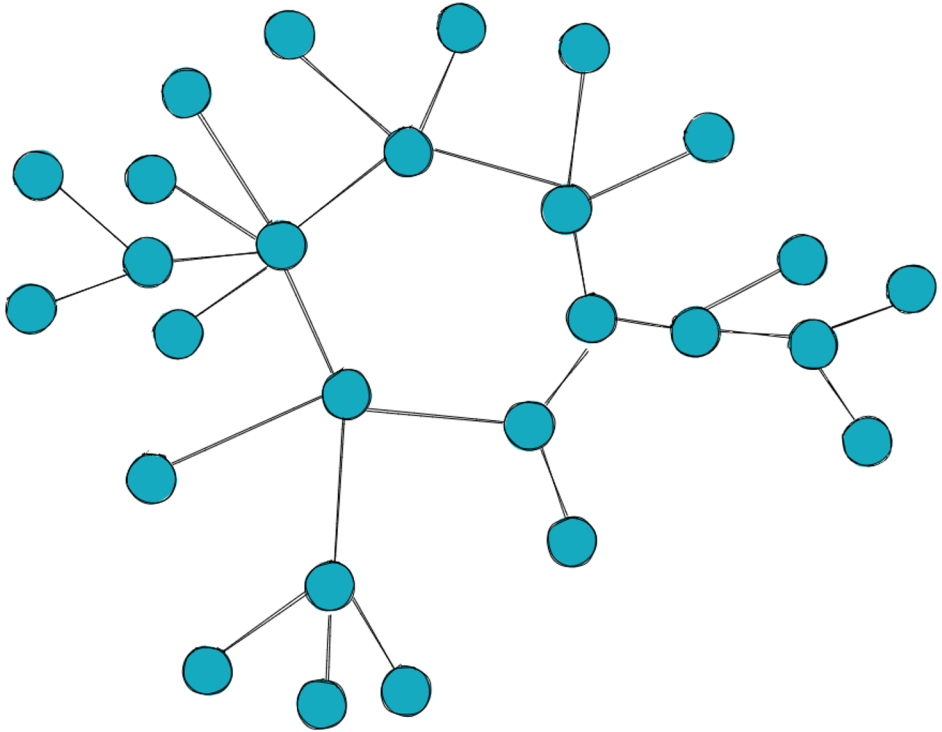


Biomedical knowledge is spread across multiple resources



Graph makes things simpler

BIKG

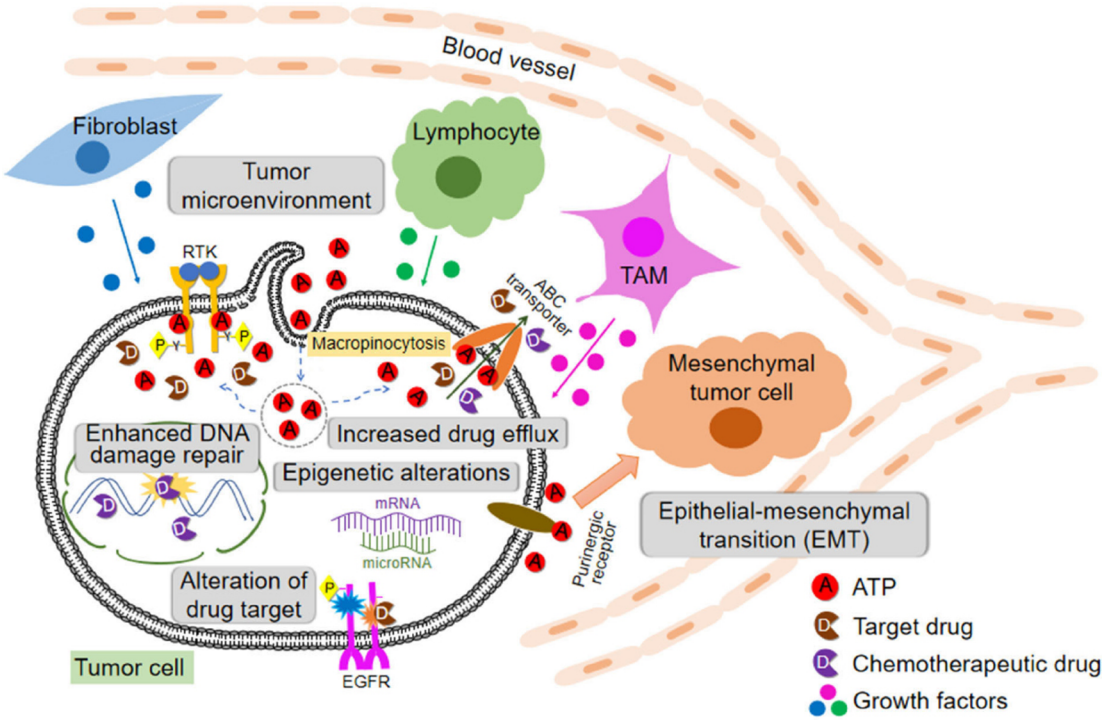


- Biomedical information often comes in forms of **networks** and **hierarchies**
- Graph is a convenient way to organise it
- BIKG (our internal knowledge graph): **60+** data sources including - omics and data extracted from the literature
- **11 M nodes, 1 B edges**
- Use graph as a source of context and features for recommenders

Early success story:

graph-based recommendations

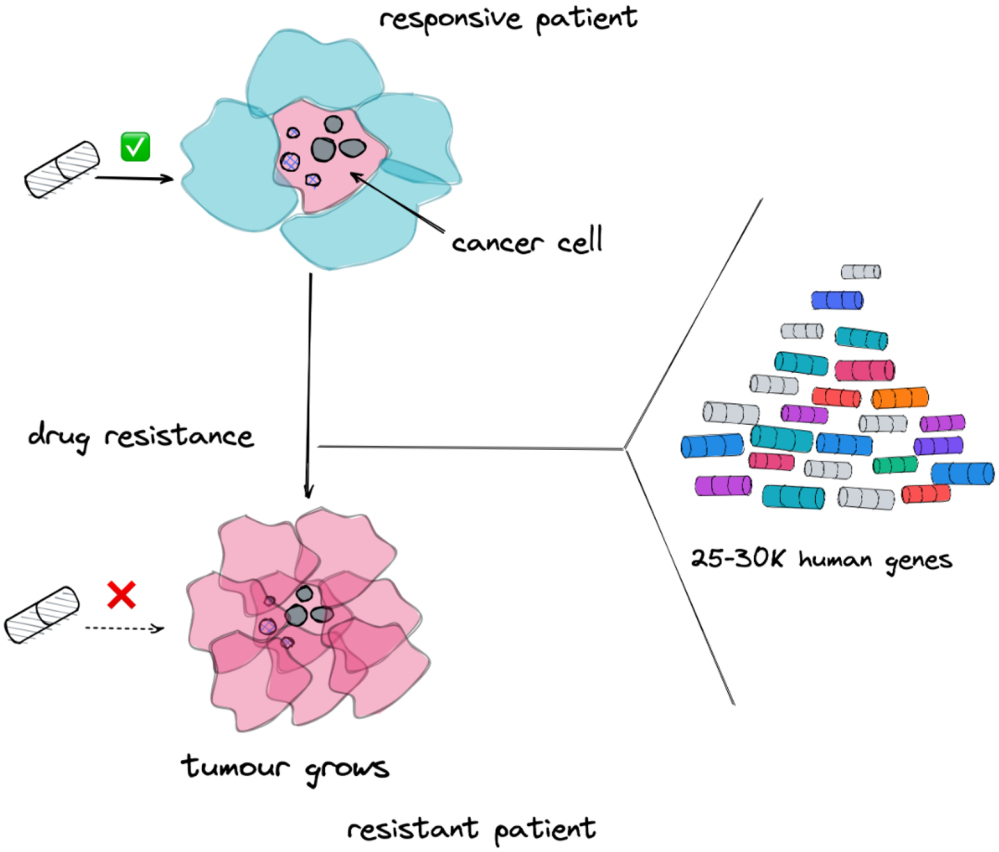
Applied recommendation problem #1: contextualize experimental data



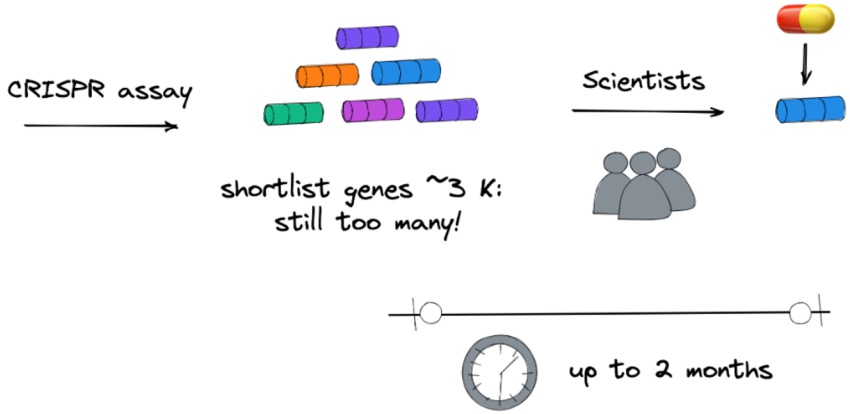
- Drug resistance in lung cancer
- Occurs in a sub-population of patients
- Resistance landscape is complex

X Wang, H Zhang, X Chen - Cancer Drug Resistance, 2019

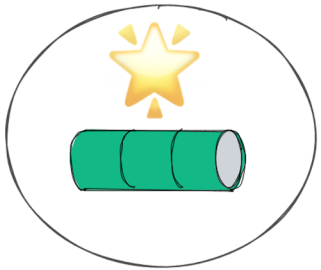
How to help scientist find key genes faster?



- key gene == target
- find a target gene ➡️ develop new treatments to overcome the resistance



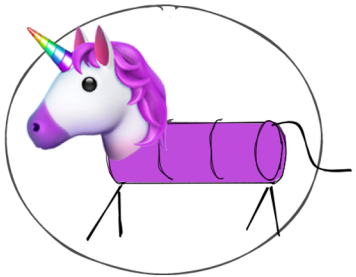
An ideal target



- [x] Expression
- [x] Pathway/complex enrichment
- [x] Effect size
- [x] Druggability
- [x] Mode of action
- [x] Translation in models
- [x] Internal assets
- [x] Bench validation
- [x] Consistency in assays
- [x] Clinical relevance
- [x] Literature support
- [x] Novelty

...

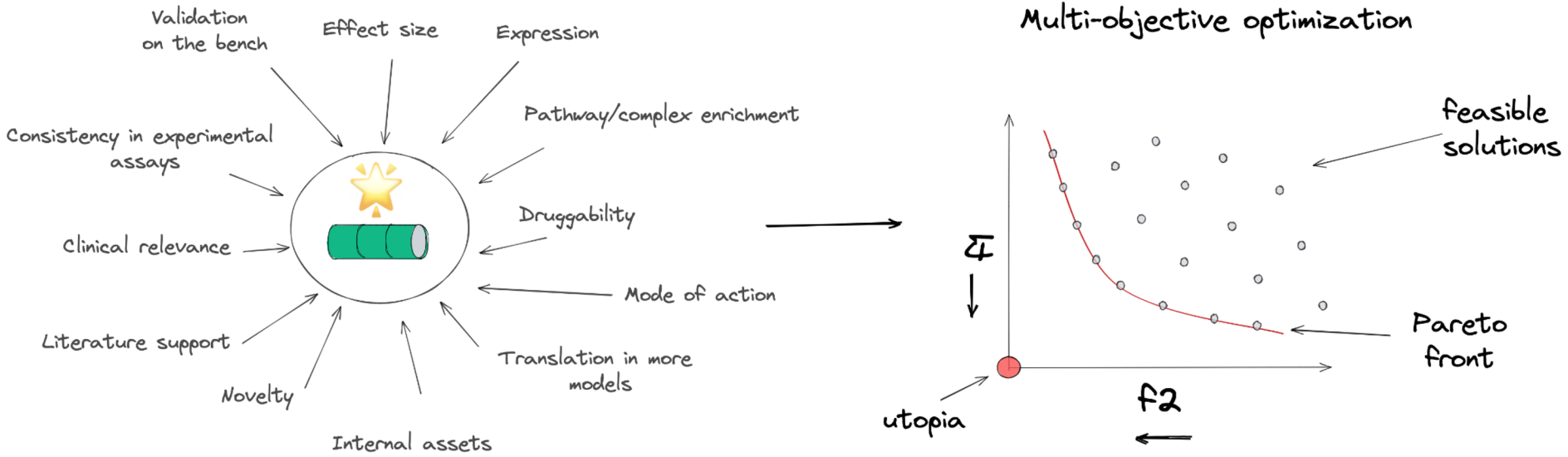
An ideal target does not exist



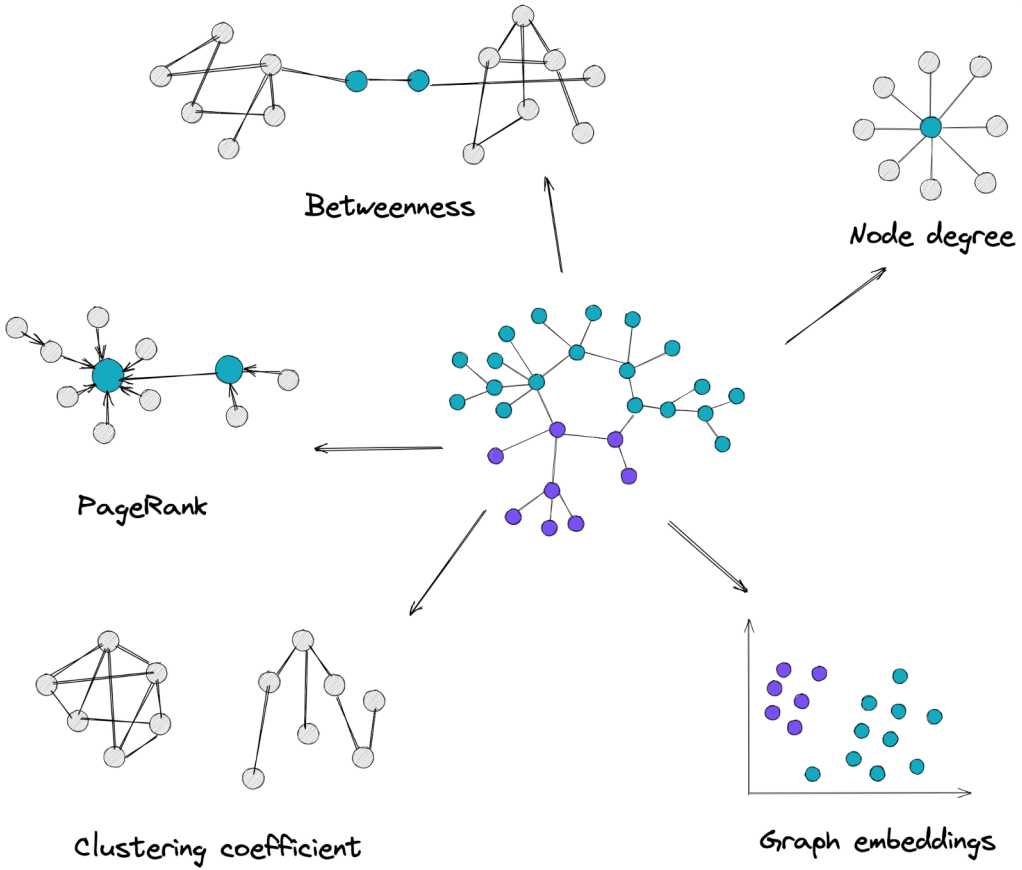
- [x] Expression
- [x] Pathway/complex enrichment
- [x] Effect size
- [x] Druggability
- [x] Mode of action
- [x] Translation in models
- [x] Internal assets
- [x] Bench validation
- [x] Consistency in assays
- [x] Clinical relevance
- [x] Literature support
- [x] Novelty

...

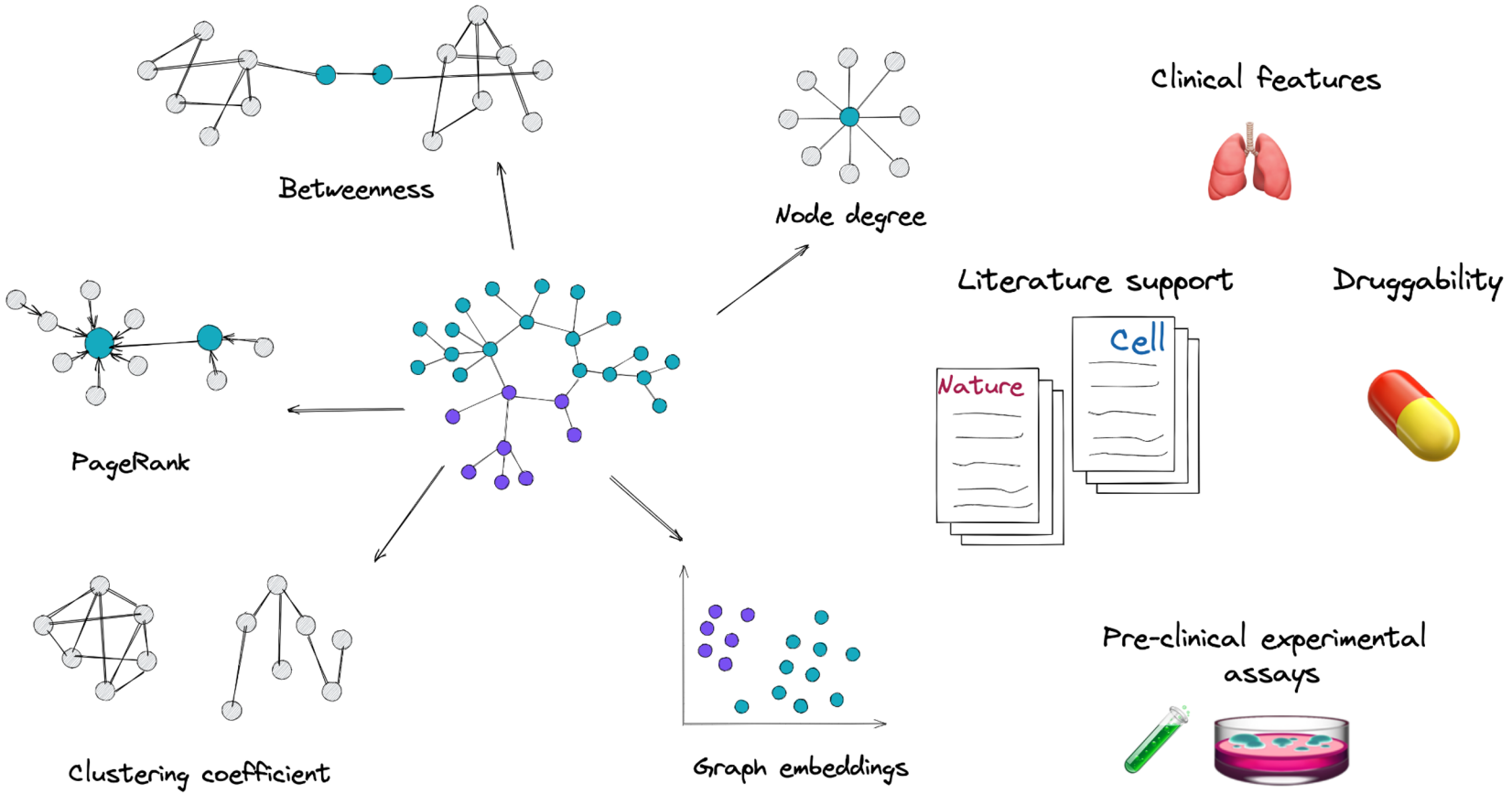
Target selection as an optimization problem



Hybrid feature set: source features from the graph

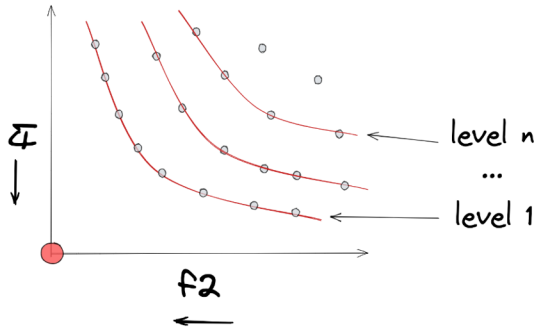


Hybrid feature set: combine with clinical features

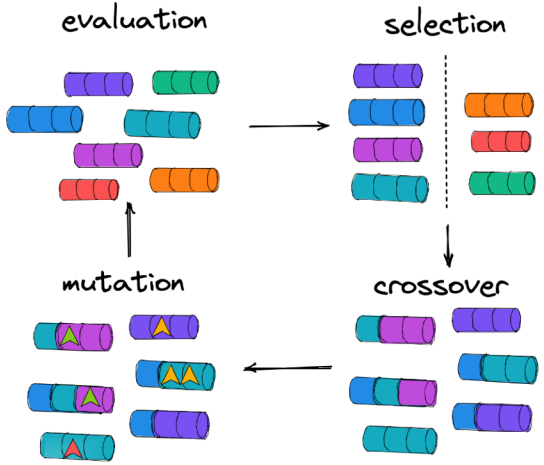


Approaches

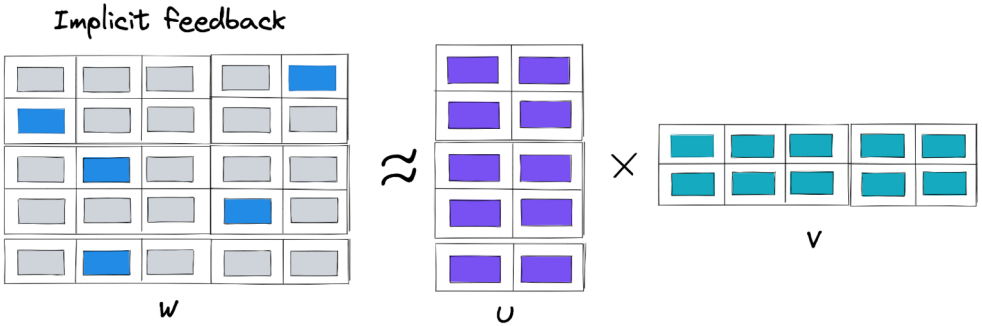
1 Compute exact Pareto front



2 Evolutionary algorithms

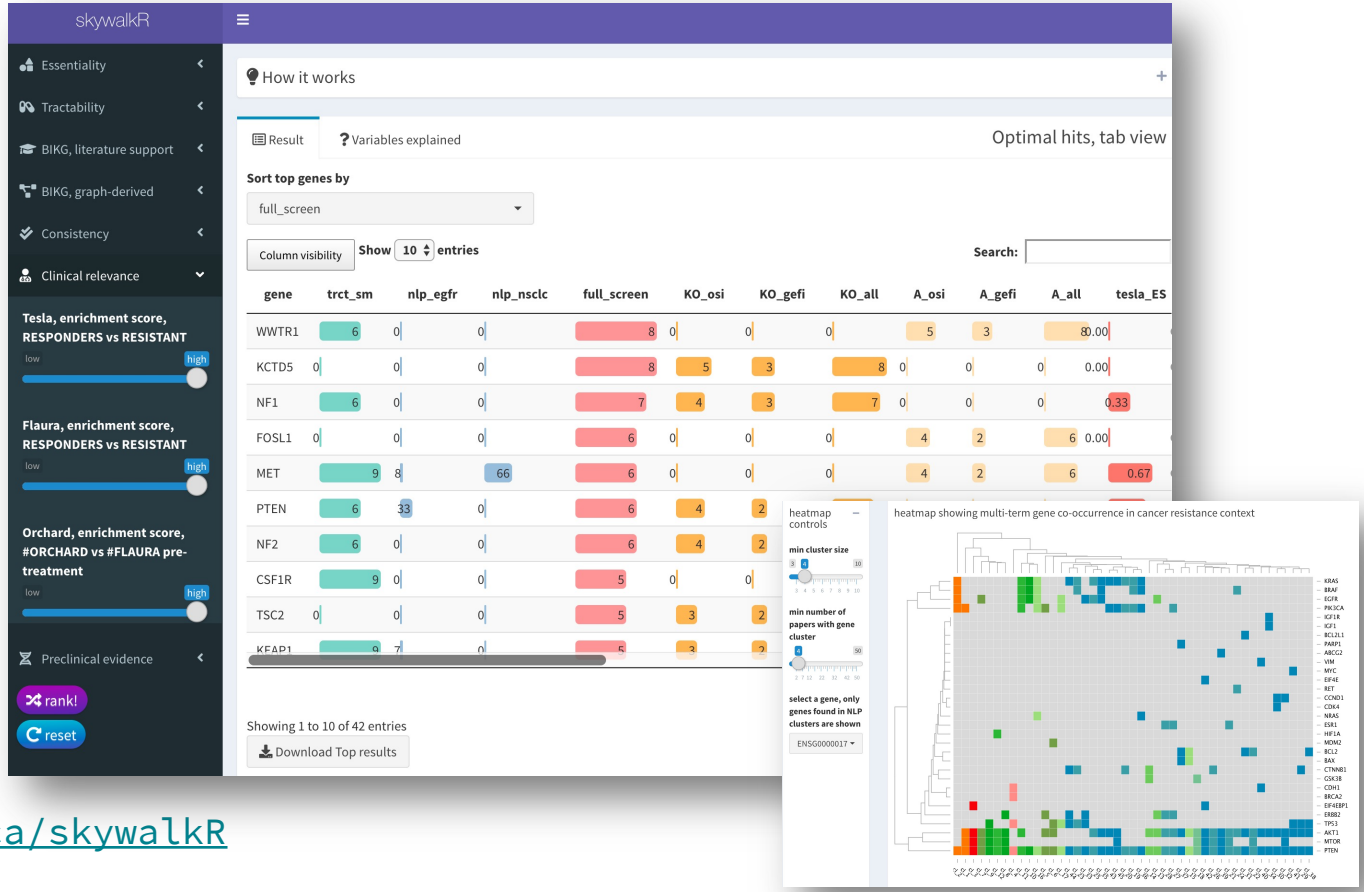


3 Matrix factorization



SkywalkR, interactive interface

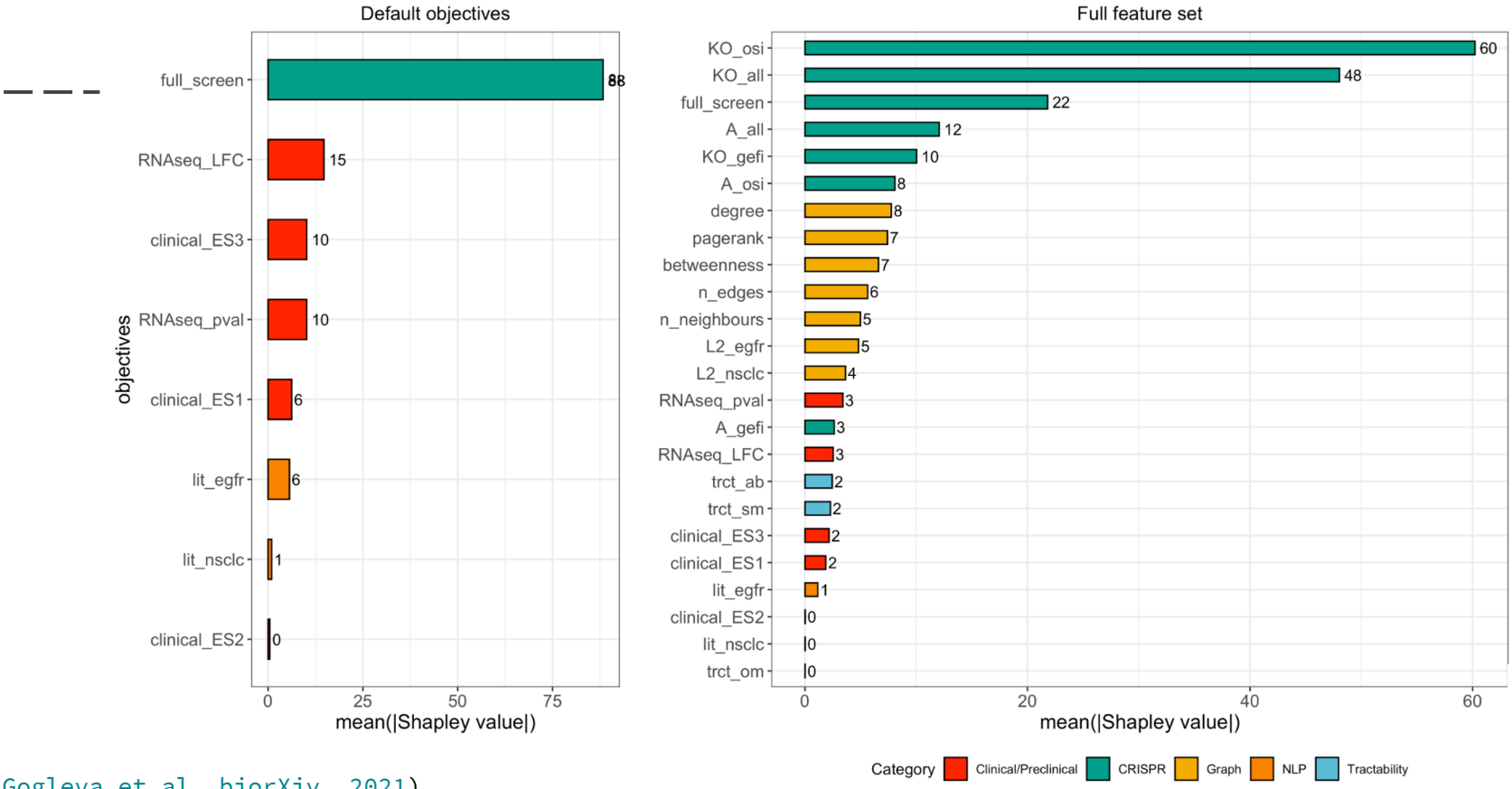
- select a subset of objectives
- set optimization directions
- explore trade-offs



github.com/AstraZeneca/skywalkR

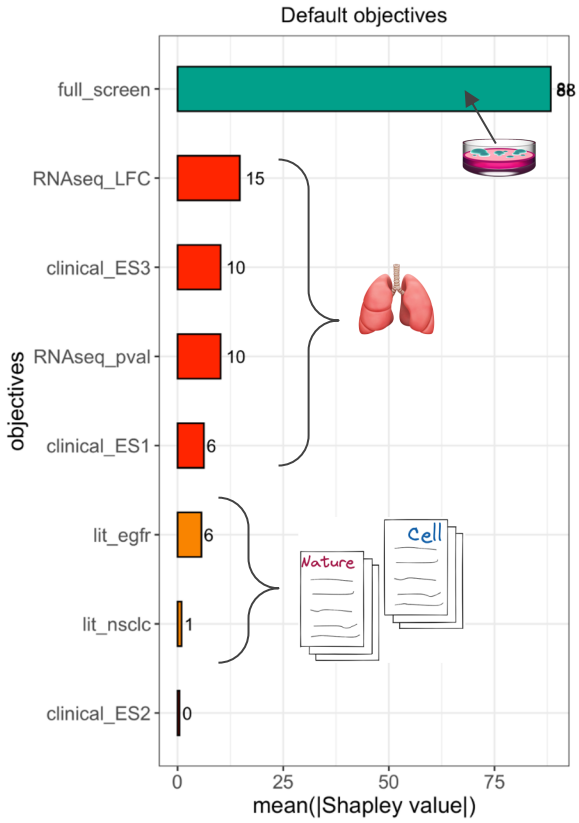
Imperfect validation

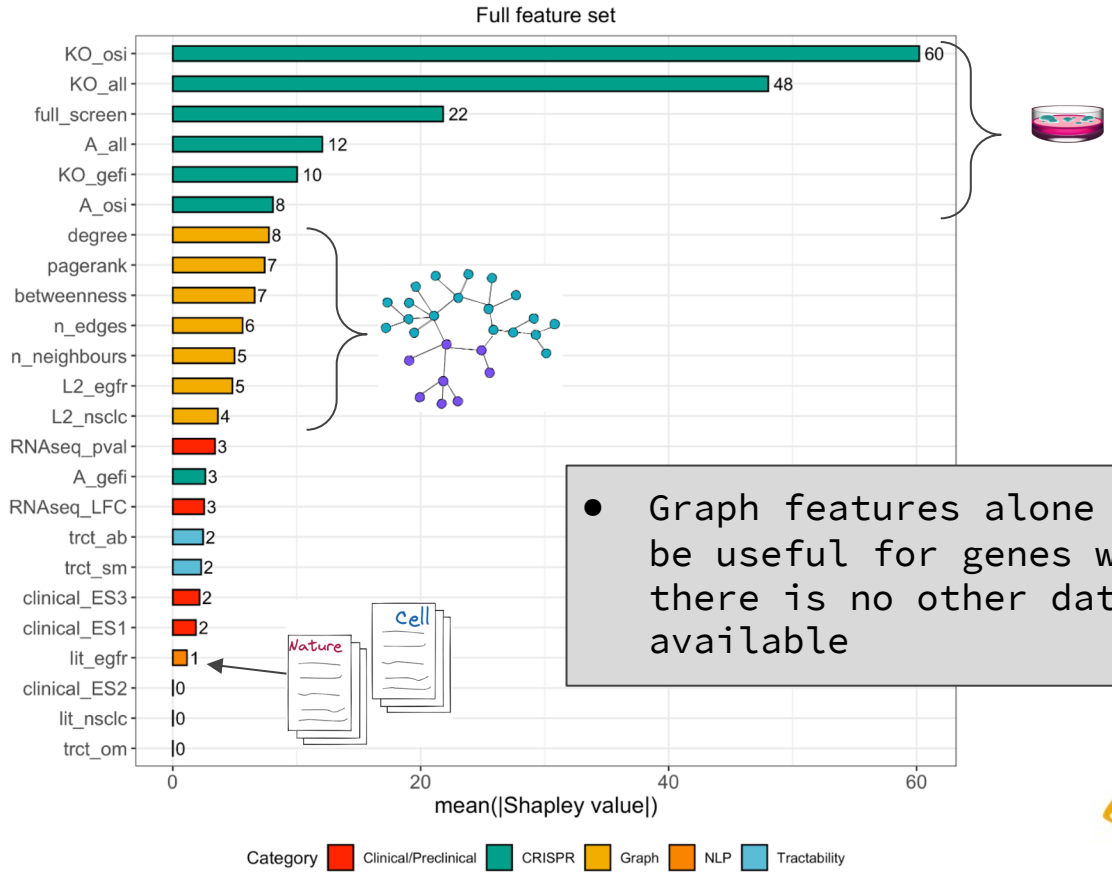
Model domain scientist as a black box classifier



(Gogleva et al, biorXiv, 2021)

Graph-derived features follow clinical in unbiased setting

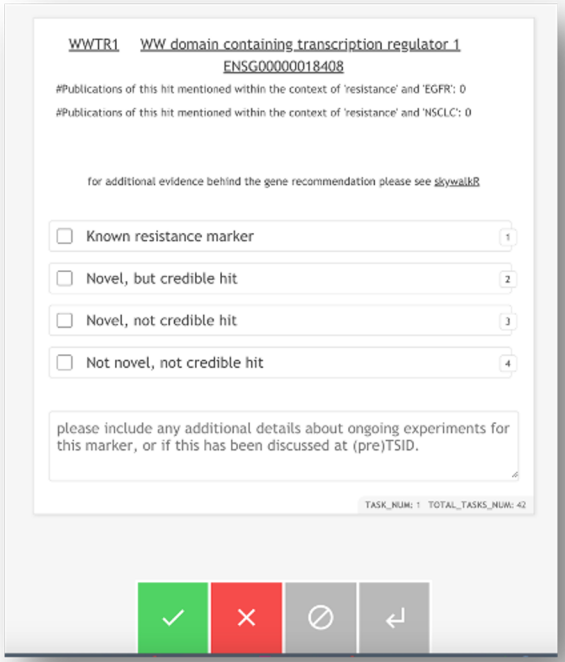




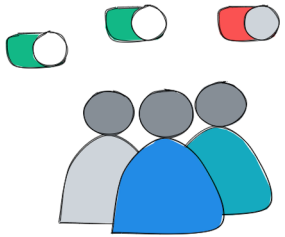
- Graph features alone can be useful for genes when there is no other data available

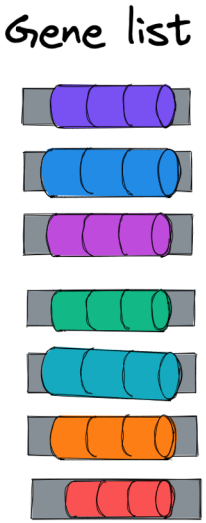
(Gogleva et al, biorXiv, 2021)

Annotation by the experts

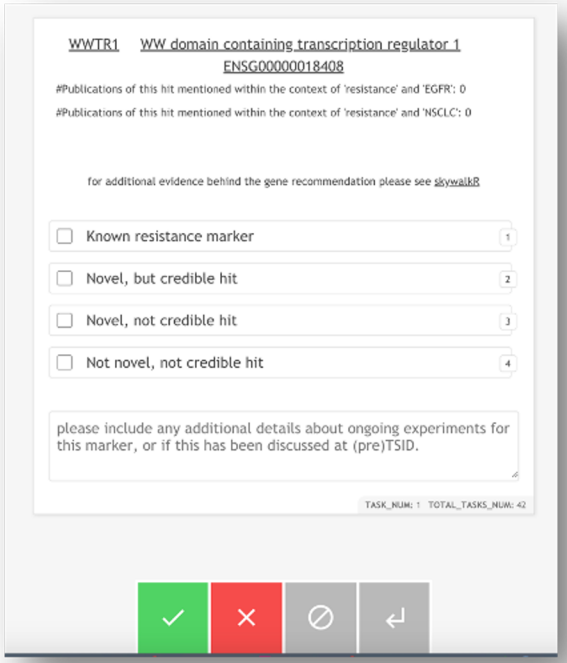


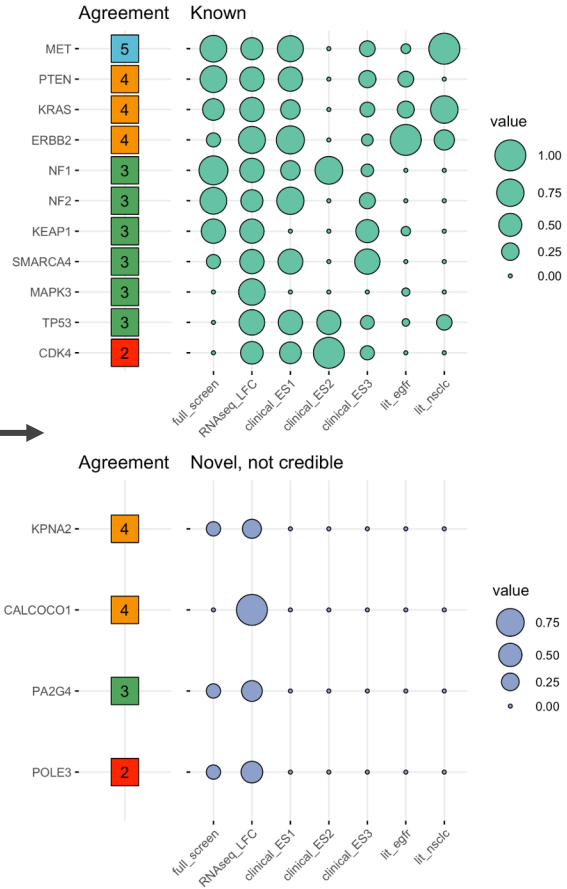
prodigy





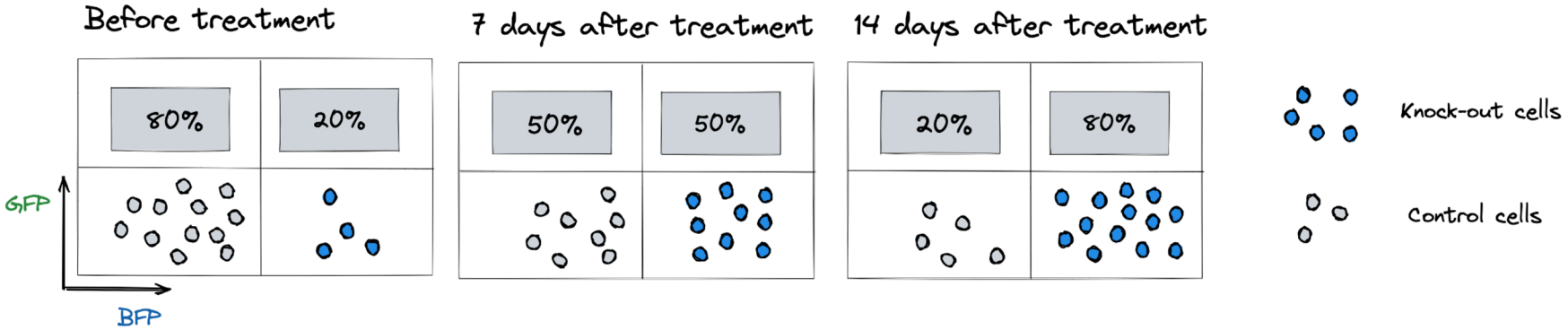
Most of recommendations are ‘novel & credible’



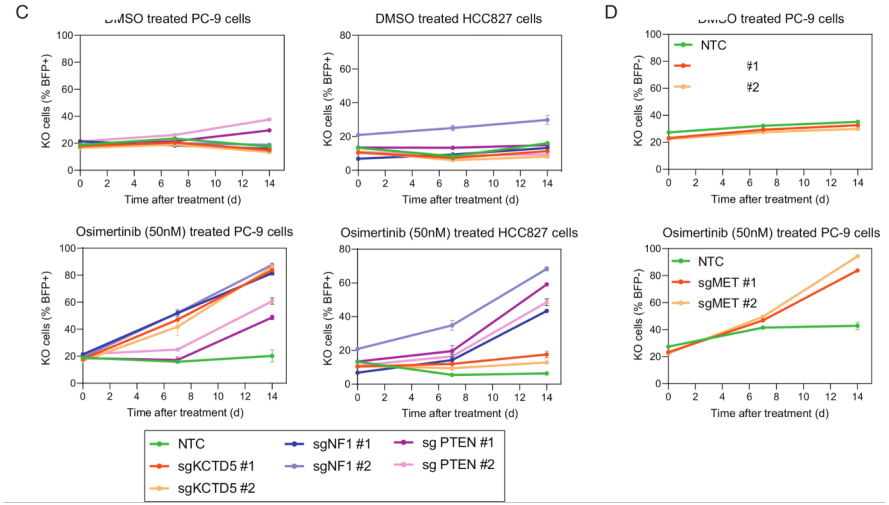


(Gogleva et al, biorXiv, 2021)

Experimental validation *in vitro*

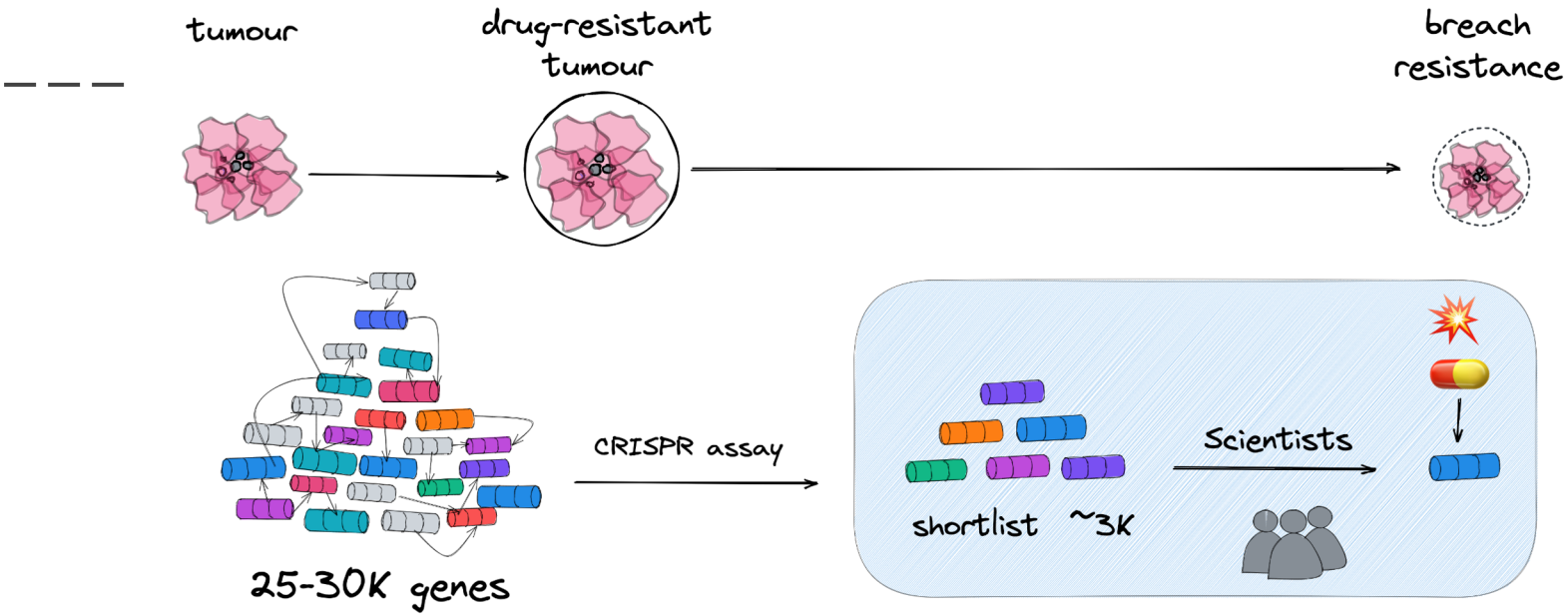


- confirmed involvement of 4 recommended genes in drug resistance
- next: test the remaining genes



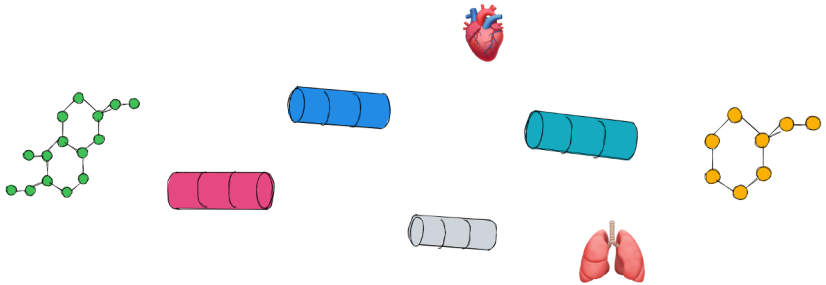
(Gogleva et al, biorXiv, 2021)

Imperfect, yet already useful recommendation system



- 🐌 -> 🏎️ re-rank lists in seconds, not months
- ⚙️ automated feature generation
- ♻️ approach can be re-used in related problems

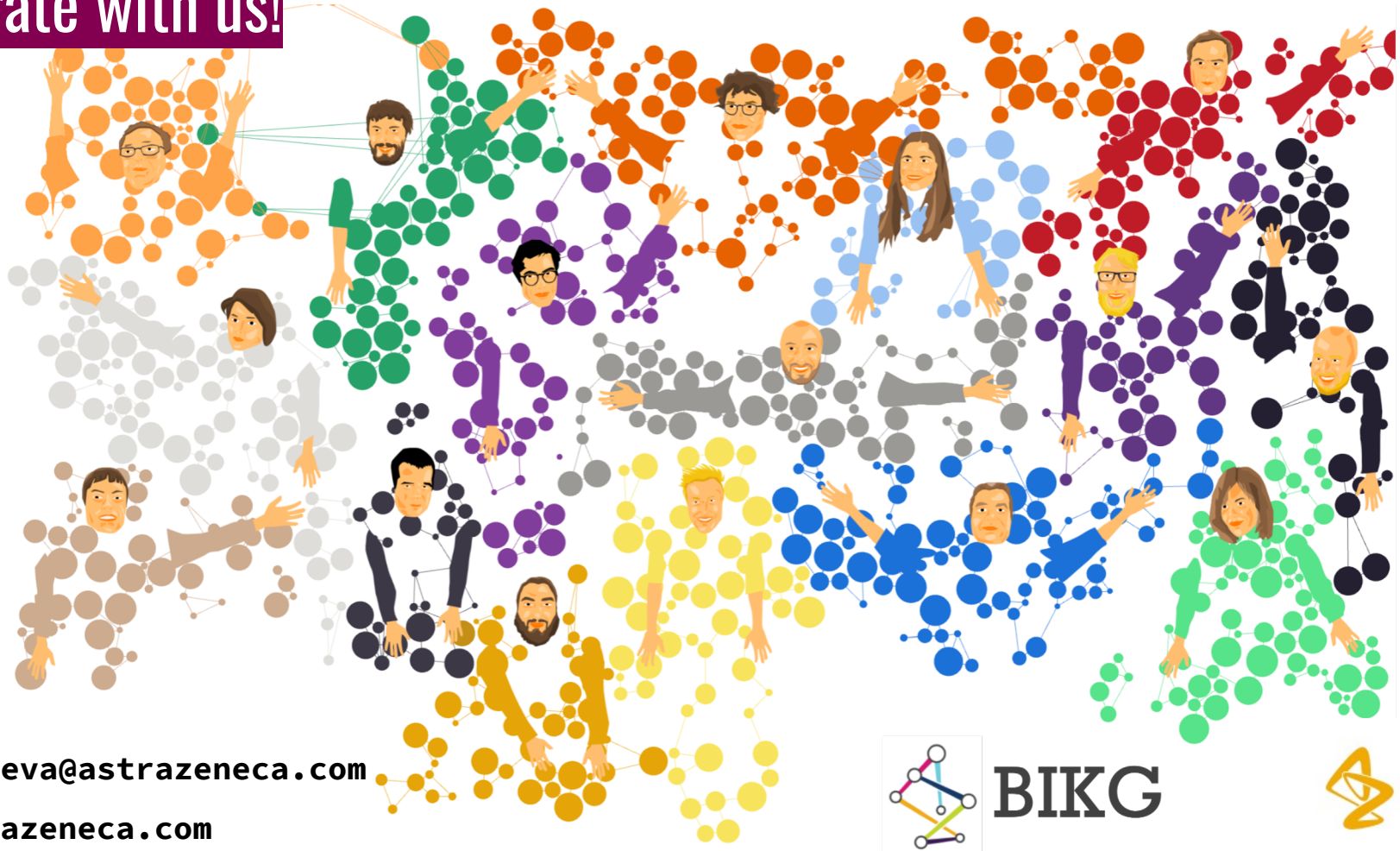
Take home message



- Drug discovery is an exciting field for recommender systems
- Relatively simple recommenders can have a lot of impact
- Need for recommenders that can operate in unsupervised or weakly supervised settings
- There are a number of challenges❇️

❇️ Read more in the extended deck:
https://astrazeneca.github.io/recsys21gogleva/

Collaborate with us!



anna.gogleva@astrazeneca.com

bikg@astrazeneca.com

BIKG