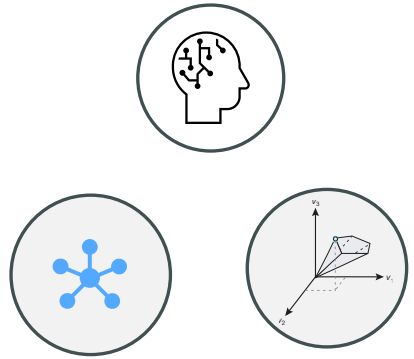
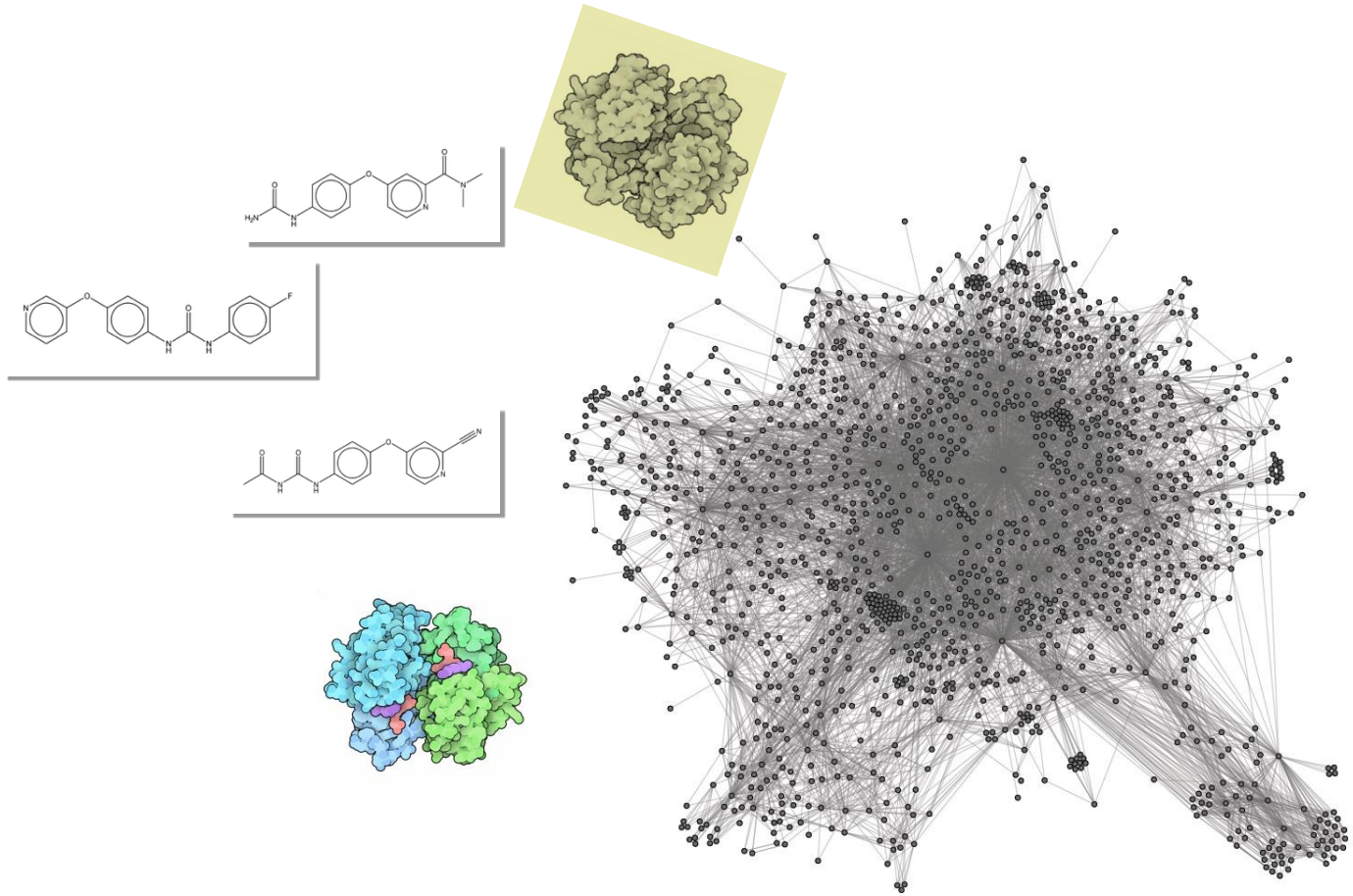
AI, NETWORKS & MODELS: A TRIFECTA RESHAPING EARLY-STAGE DRUG DISCOVERY



KARTHIK RAMAN

Department of Department of Data Science and AI
Wadhwani School of Data Science and AI
Centre for Integrative Biology & Systems medicinE (IBSE)



OVERVIEW

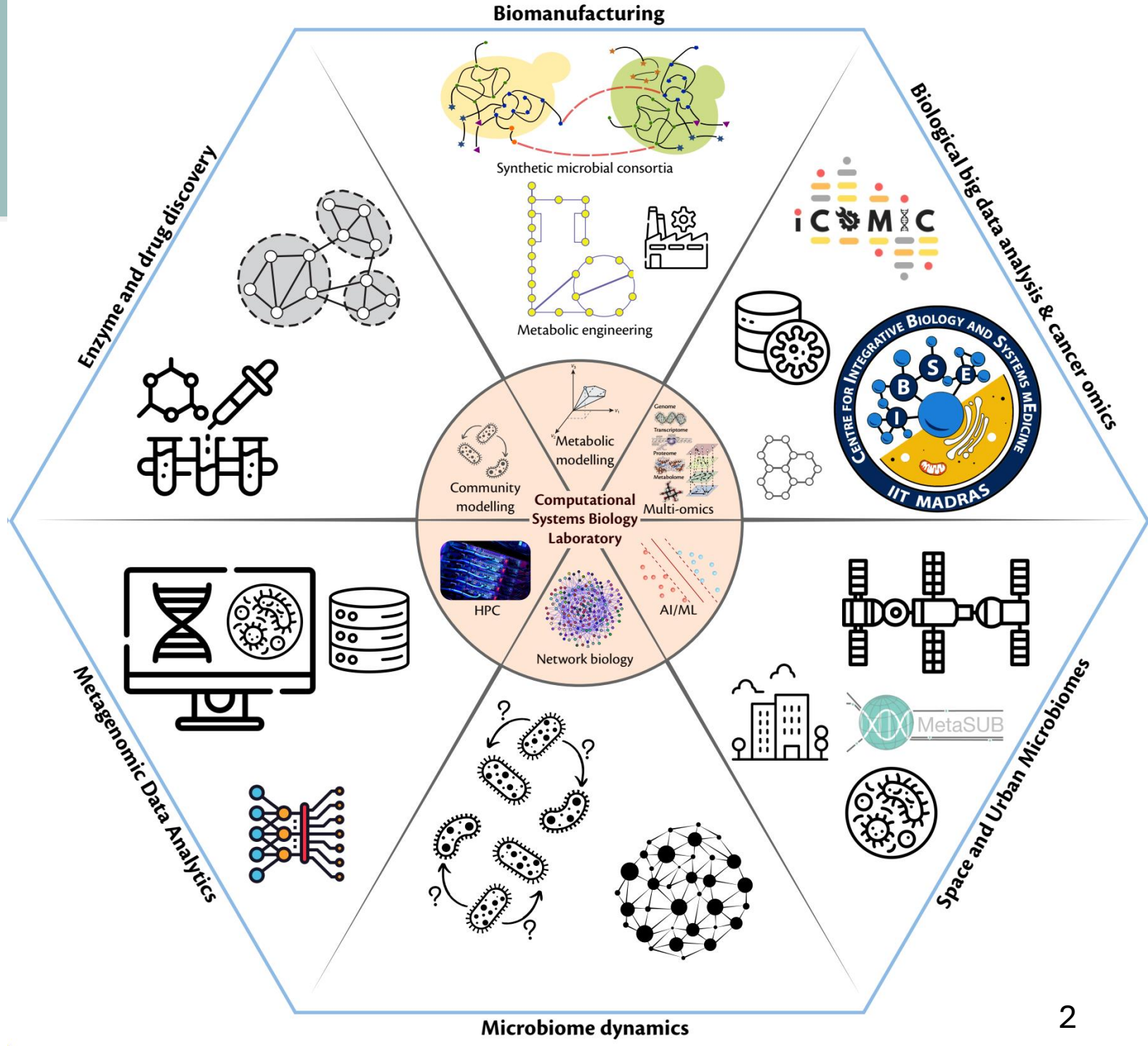
Interdisciplinary group focussing on the development of computational tools and **scalable** algorithms to

- Understand and manipulate biological networks
  - leveraging **AI, Networks & Models**
- Perform integrated analysis of multi-omic biological data

- Ultimate goal: impact in the clinic / industry / environment

We also actively drive the

- Centre for Integrative Biology and Systems mEdicine (IBSE)



KEY RESEARCH CONTRIBUTIONS

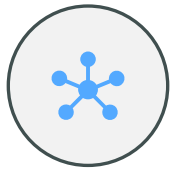
Algorithms/Theory: Networks/Constraint-based Modelling

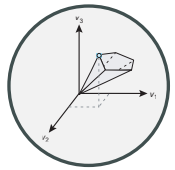
- Fast-SL: Efficient enumeration of synthetic lethals (Pratapa *et al.*, 2015)
- ReactionMiner: Predicting novel pathways (Sankar *et al.*, 2017)
- MetQuest: Enumerating all alternate pathways in metabolic networks (Ravikrishnan *et al.*, 2018)
- Metabolic Support Index (Ravikrishnan *et al.*, 2020)
- Understanding redundancy in metabolic networks (Sambamoorthy & Raman, 2018, *Bioinformatics*/ECCB)
- MinReact: a systematic approach for identifying minimal metabolic networks (Sambamoorthy & Raman, 2020)
- CAMP (Ibrahim & Raman, 2021)
- Co-FSEOF (Raajaraam & Raman, 2021)
- Control-theoretic approaches to design regulatory networks (Bhattacharya *et al.*, 2018, 2022, 2023, 2024)
- Patterning in microbial consortia (Venkatraghavan *et al.*, 2022)
- Sloppiness in biological systems (Jagadeesan *et al.*, 2023)
- Minimal Microbiome (Raghu *et al.*, 2024)
- Panera (Palanikumar *et al.*, 2024)
- Flux switching/minimal rerouting (Narasimha *et al.*, 2024)
- COSMOS (Raajaraam and Raman, 2025)

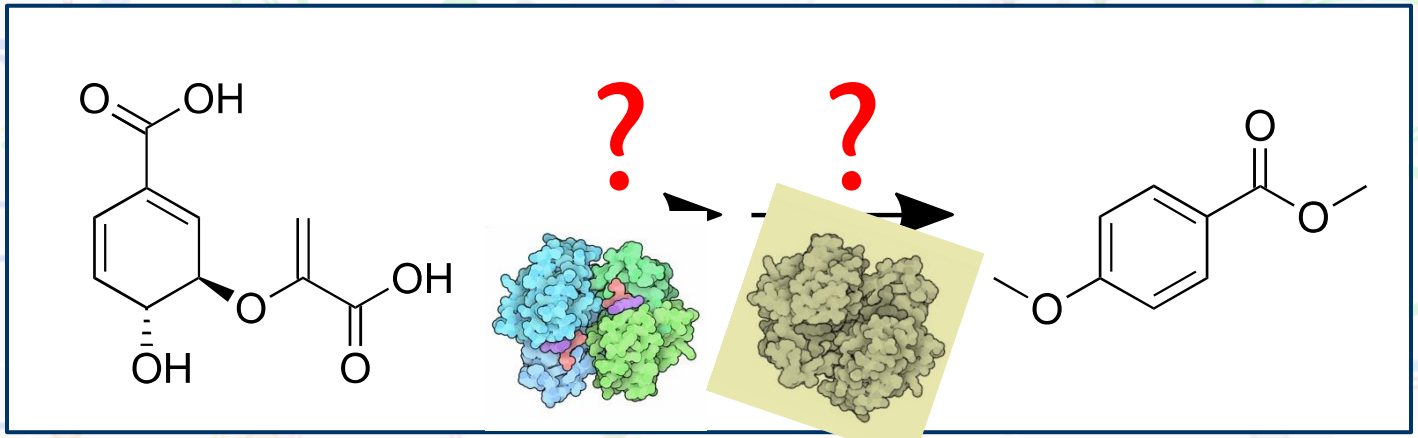
Networks / AI

- Disease Module Identification in Heterogeneous Biological Networks (Tripathi *et al.*, 2019; Choobdar *et al.*, 2019)
- Network-based features to predict essential genes across diverse organisms (Azhagesan *et al.*, 2019)
  - NetGenes (Senthamizhan, Ravindran & Raman, 2021)

HOW DO WE LEVERAGE AI, NETWORKS & MODELS TO RESHAPE EARLY-STAGE DRUG DISCOVERY?





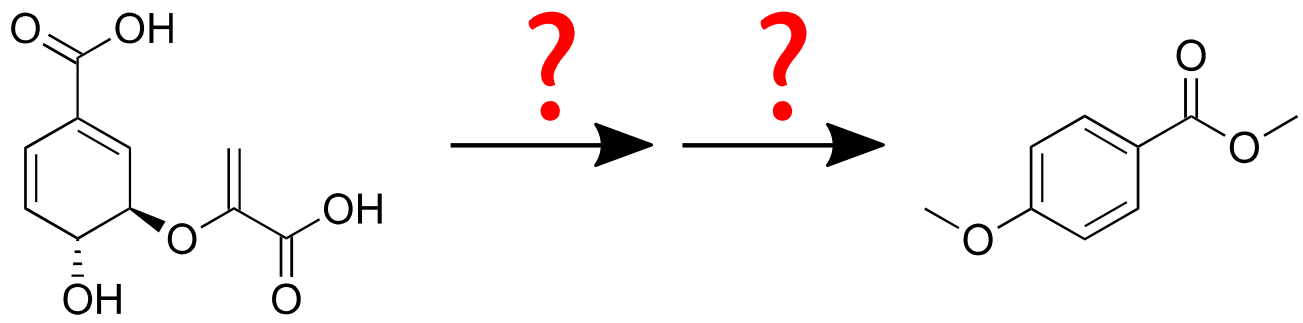


RETROSYNTHESIS
PREDICTING NOVEL METABOLIC PATHWAYS THROUGH SUBGRAPH MINING

Sankar A, Ranu S and Raman K (2017) *Bioinformatics* **33**:3955-63

/RamanLab/ReactionMiner

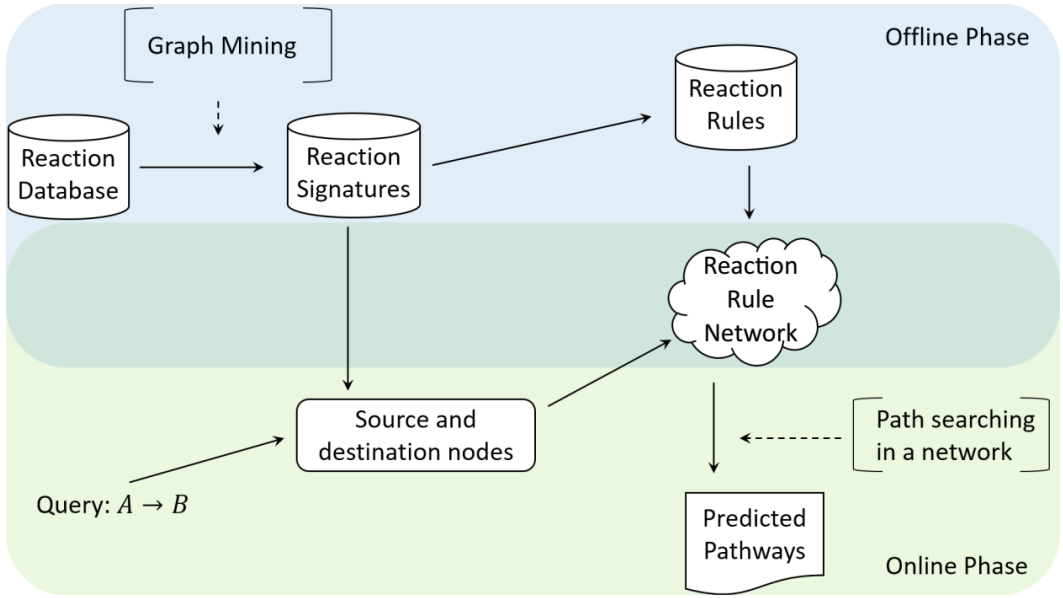
REACTION MINER: PREDICTING NOVEL PATHWAYS



- 10,000+ reactions in KEGG/BioCyc databases – represent the varied repertoire of biochemical conversions that known enzymes can catalyse
- Importing these non-native reactions into organisms has been a key driver of metabolic engineering – enabling the **retrosynthesis** of novel metabolites

CAN WE BUILD A
FULLY AUTOMATED ENGINE
TO PREDICT PATHWAYS
BETWEEN TWO MOLECULES,
RELYING ONLY ON
STRUCTURAL INFORMATION?

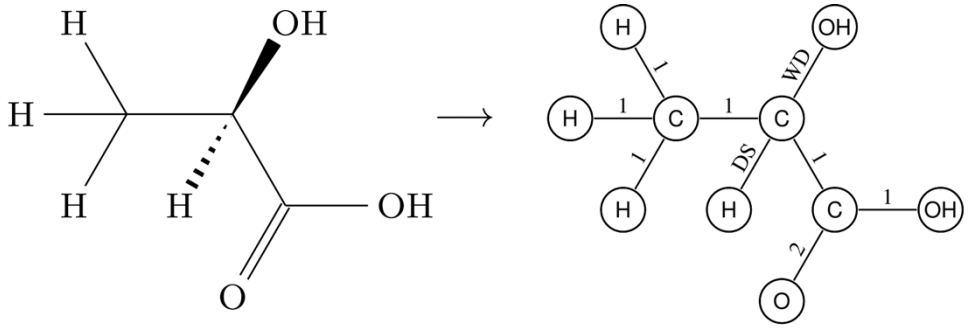
REACTION MINER: APPROACH



Sankar A, Ranu S and Raman K (2017)
Bioinformatics **33**:3955-63

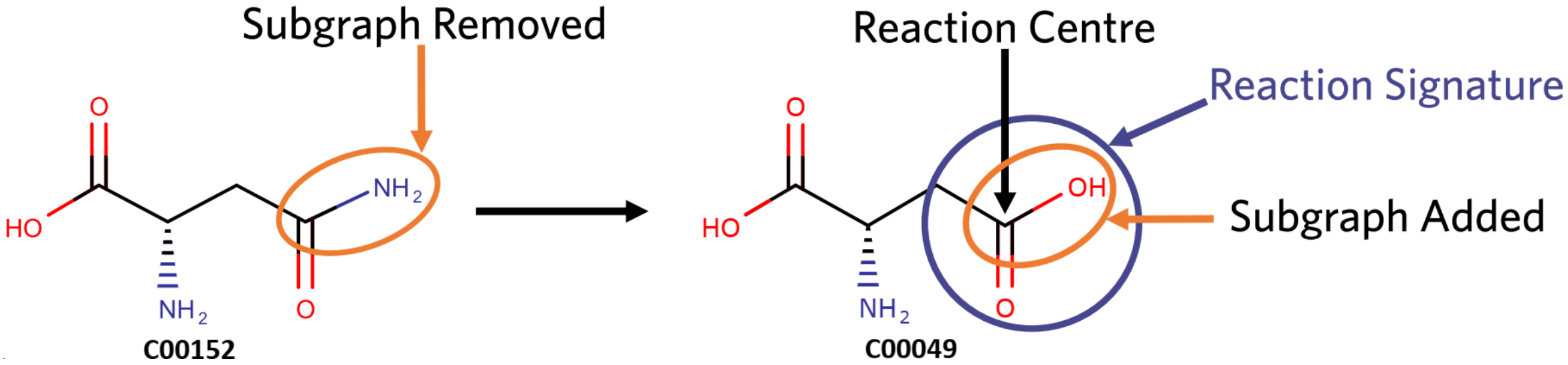
/RamanLab/ReactionMiner

- Every molecule is represented as a graph!



- **Reactant-product mapping**, to identify the transformation, using subgraph edit distance

REACTION MINER: ILLUSTRATION



- **Subgraph mining**, to identify *reaction centre*, the *place of change*, and *reaction signature*, *i*.e. subgraph critical for reaction to occur
- **Reaction rules identified**, by summarizing transformations in signatures

- **Reaction Rule Network (RRN)** created
  - Nodes: reaction rules
  - Edges: connect rules that can be successively applied
- Pathway prediction and ranking—**reduces to path-finding** on the RRN! Also developed a heuristic for ranking most plausible pathways (minimal chemical

REACTION MINER: SUMMARY

- Recovers naturally occurring pathways in many cases, even in central metabolism
- Enables a realistic ranking of predicted pathways — from a large pool of possible alternate pathways
- Predicted 'correct' retrosynthetic pathways in most cases
- We use only structure information: no information on atom-atom mapping, or hand-curated rules!

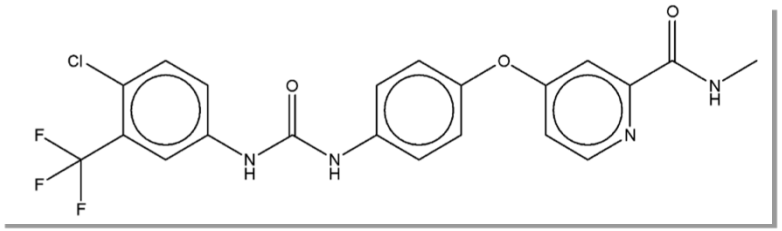
Bioinformatics, 33(24), 2017, 3955–3963
doi: 10.1093/bioinformatics/btx481
Advance Access Publication Date: 27 July 2017
Original Paper

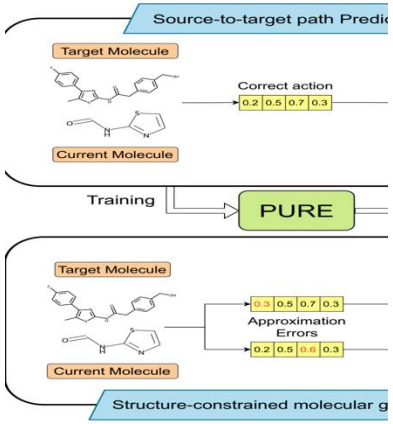
Systems biology

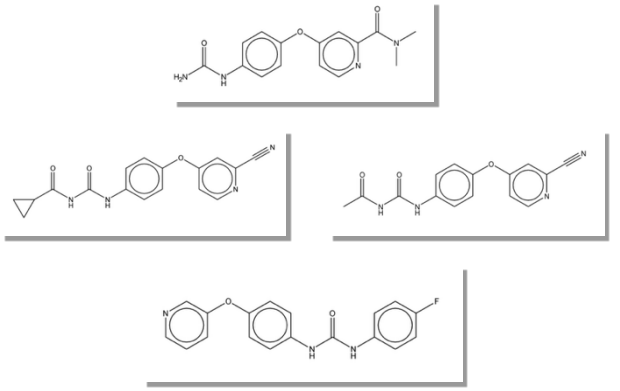
Predicting novel metabolic pathways through subgraph mining

Aravind Sankar[1,†], Sayan Ranu[1,2,*,‡] and Karthik Raman[2,3,*]

/RamanLab/ReactionMiner







PURE: POLICY-GUIDED UNBIASED REPRESENTATIONS FOR STRUCTURE-CONSTRAINED MOLECULAR GENERATION

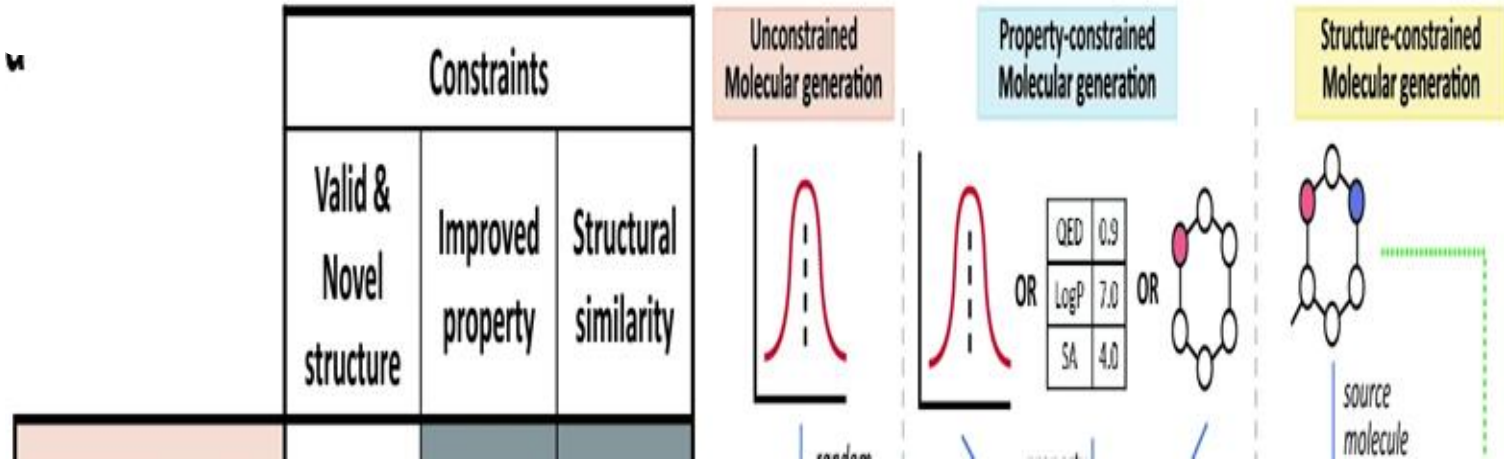
Gupta A, Barathi L, Current S, Batra R, B Ravindran, K Raman* and S Parthasarathy* (2025)

J Chemoinformatics **17**:156

/WSAI-IITM/ReactionRL

STRUCTURE-CONSTRAINED MOLECULAR GENERATION (SCMG)

- Goal-Oriented Molecular Design: SCMG aims to generate novel molecules with desired properties (e.g., drug-like, bioactive)
  - while adhering to specific structural constraints like functional groups, stereochemistry, or ring structures
- SCMG is widely used in drug discovery for generating candidate molecules that fit specific targets
  - often integrated into high-throughput screening/virtual screening workflows
- Generated molecules are evaluated using scoring functions/computational tools
  - e.g. docking simulations, synthetic feasibility models, and ADMET predictions, etc.



Choi *et al.* (2023) "COMA: …" *J Chemoinformatics* **15**:8

Challenges: Ensuring chemical validity, satisfying structural constraints, and optimizing multiple properties (e.g., drug efficacy, toxicity) ***simultaneously*** during the generation process

CURRENT CHALLENGES IN RL FOR SCMG

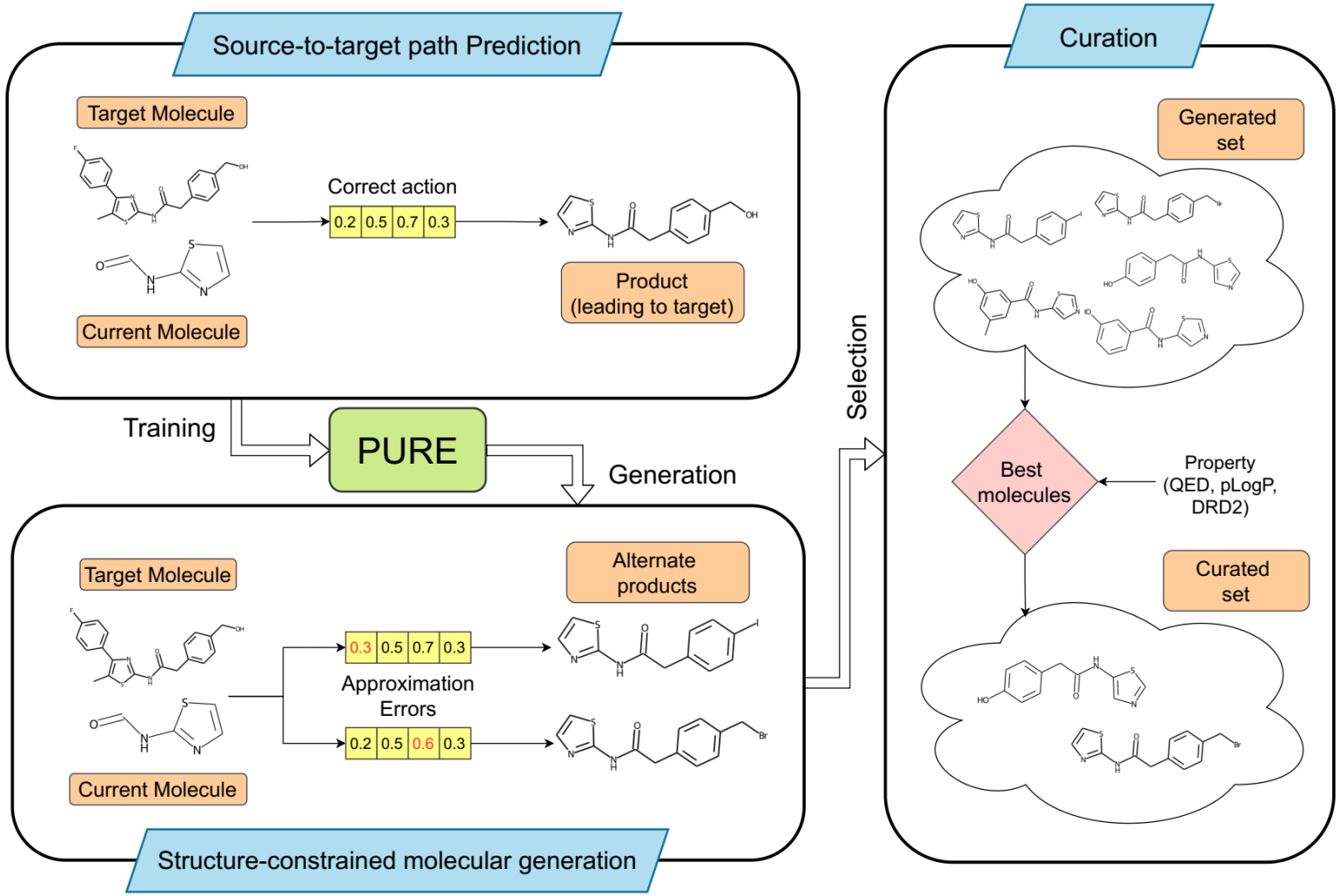
- Metric Leakage: Training exposure to evaluation metrics leads to overfitting and bias toward specific properties
- Continuous vs Discrete: Mismatch between continuous deep learning methods and discrete molecular space
- Limited Synthesizability: Generated molecules often lack practical synthesis pathways
- Sparse Data: High-quality training data for molecular properties is scarce and expensive
- Multi-objective Optimization: Difficulty balancing multiple conflicting molecular properties simultaneously

REINFORCEMENT LEARNING FOR SCMG

- Goal: Generate drug-like molecules with a required scaffold and specific bioactivity
  - State: Initial molecule with a predefined scaffold (e.g., benzene ring)
  - Actions: Add a side chain, modify a bond, or introduce a heteroatom
- Reward:
  - High reward for molecules that retain the scaffold, have low toxicity, and high docking scores
  - Penalty for invalid molecules or those violating constraints
- Policy: Agent learns to prioritize actions that produce desired molecular properties
- Environment: A molecule generation and evaluation platform e.g. RDKit, integrated with scoring models

PURE: POLICY-GUIDED UNBIASED REPRESENTATIONS

- Addresses limitations in existing deep learning solutions for SCMG
- Combines self-supervised learning with a policy-based reinforcement learning approach
- Mitigates the issue of "metric-leakage" by eliminating reliance on external molecular metrics during training
- Demonstrates competitive or superior performance compared to state-of-the-art SCMG methods across multiple benchmarks



KEY INNOVATIONS IN PURE

- Policy-Guided: Uses reinforcement learning policy for intelligent molecular transformations
- Unbiased: Trained without exposure to evaluation metrics, avoiding metric-specific overfitting
- Representation Learning: Learns task-specific similarity through goal-conditioned RL framework
- Synthesisability: Uses reaction rules from USPTO-MIT dataset for synthesizable transformations
- Decoupled Design: Separates similarity learning from property optimization for flexibility

REACTION RULE MINING: USPTO-MIT DATASET

- Data Source: 84,968 reaction rules extracted from USPTO-MIT chemical reaction database
- Signature Extraction: Identify reactant and product signatures with 2-atom neighbour context
- Reaction Centres: Specific atoms where chemical transformations occur
- Validity Filter: Retain only single-centre reactions for reliability and simplicity
- Synthesizability: All transformations correspond to known synthetic chemistry

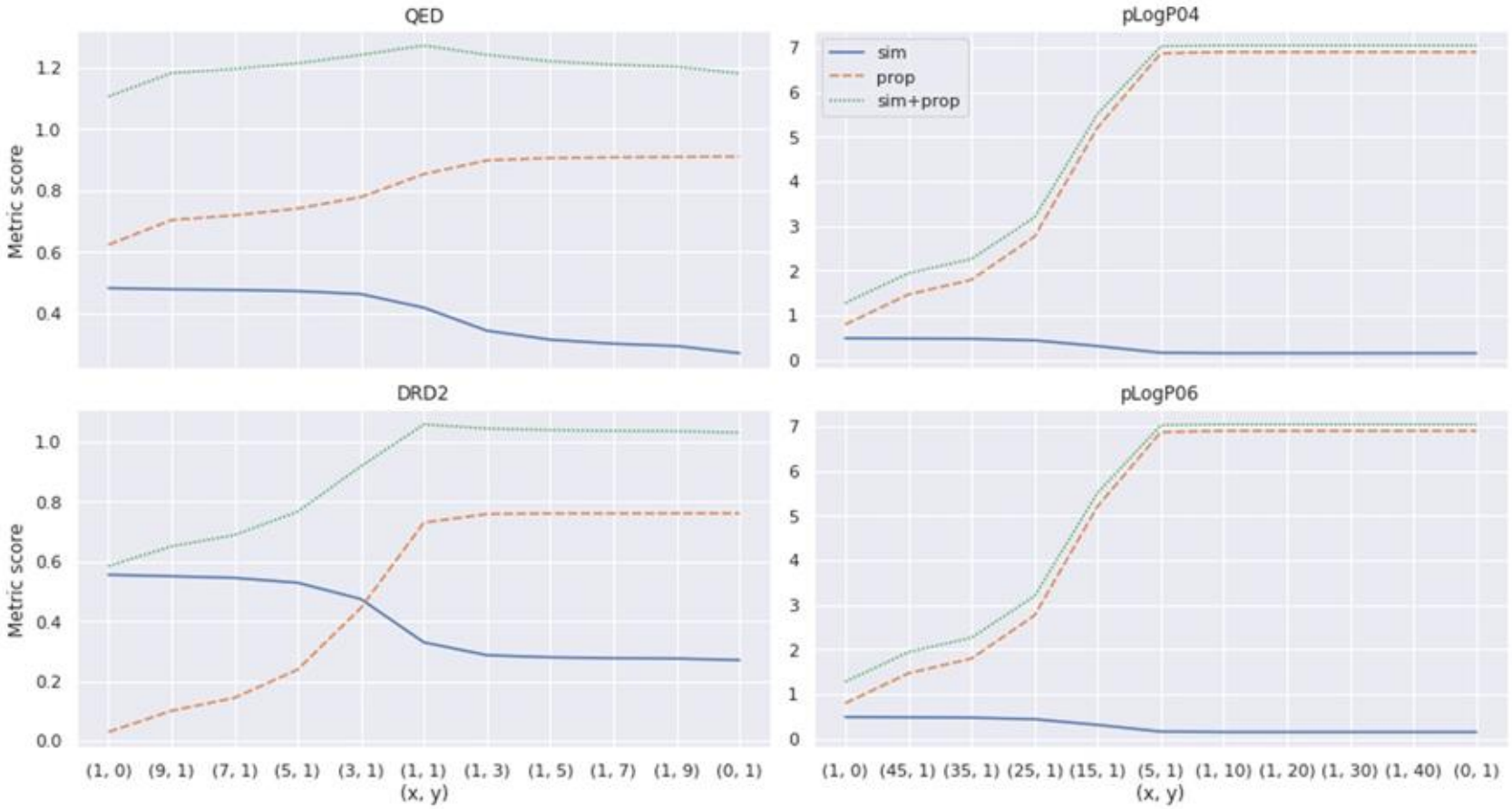
EVALUATION METRICS

- Validity: Ratio of chemically valid generated molecules following RDKit rules
- Property Score: Average target property values (QED, DRD2, pLogP) of generated molecules
- Improvement: Property score difference between generated and source molecules
- Similarity: Average Tanimoto similarity between generated and source molecules
- Novelty: Fraction of generated molecules not present in training dataset
- Diversity: Average Tanimoto dissimilarity among generated molecules

BENCHMARKING RESULTS

- PURE outperforms state-of-the-art models on multiple metrics (validity, novelty, diversity) for benchmarks like QED and DRD2, achieving higher quality molecular generation
- PURE achieves the best property scores for QED and LogP benchmarks, and excels in novelty and diversity for QED with certain coefficient combinations
- The model allows for flexible post-generation tuning by adjusting coefficients (similarity and property scores), instead of using fixed weights as in earlier methods
- PURE adapts to different tasks by decoupling coefficient selection from the generation process, reducing computational costs and improving molecule selection for specific targets
- For LogP benchmarks, PURE closely matches or surpasses the performance of other methods, achieving excellent property scores with significant improvement over baseline models

PURE METRICS



Plots show the effect of parameters x and y on the desired property metrics on the generated set of molecules for each benchmark; 'sim' represents Tanimoto similarity, 'prop' represents the desired property for the benchmark (QED, DRD2 or pLogP) and 'sim+prop' shows the linear sum of the two

CASE STUDY METHODOLOGY: SORAFENIB

WORKFLOW

From Sorafenib: 4,42,064 distinct ligands

Molecule generation → **Docking simulations** → **PLIP analysis** → **ADME profiling**

Similarity to sorafenib > 0.4 - 19,101

↓

BA to ABCG2 (DeepPurpose) < 4.7 - 8,218

↓

Molecular weight < 500 Da - 1,744

Ligand preparation

↓

Grid estimation

↓

Preliminary analysis - 20 molecules

Extended to 1744 molecules

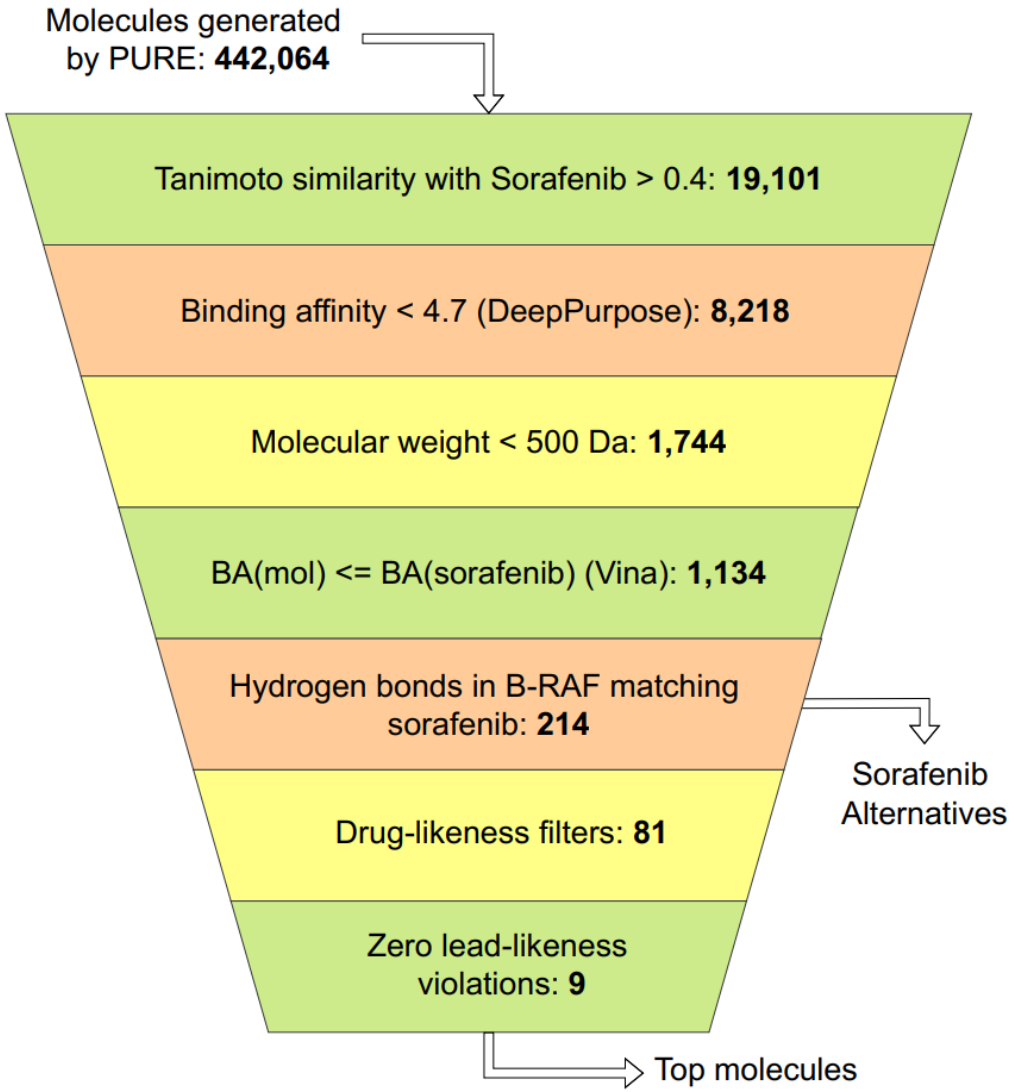
↓

Iterative simulations for significance of conformations

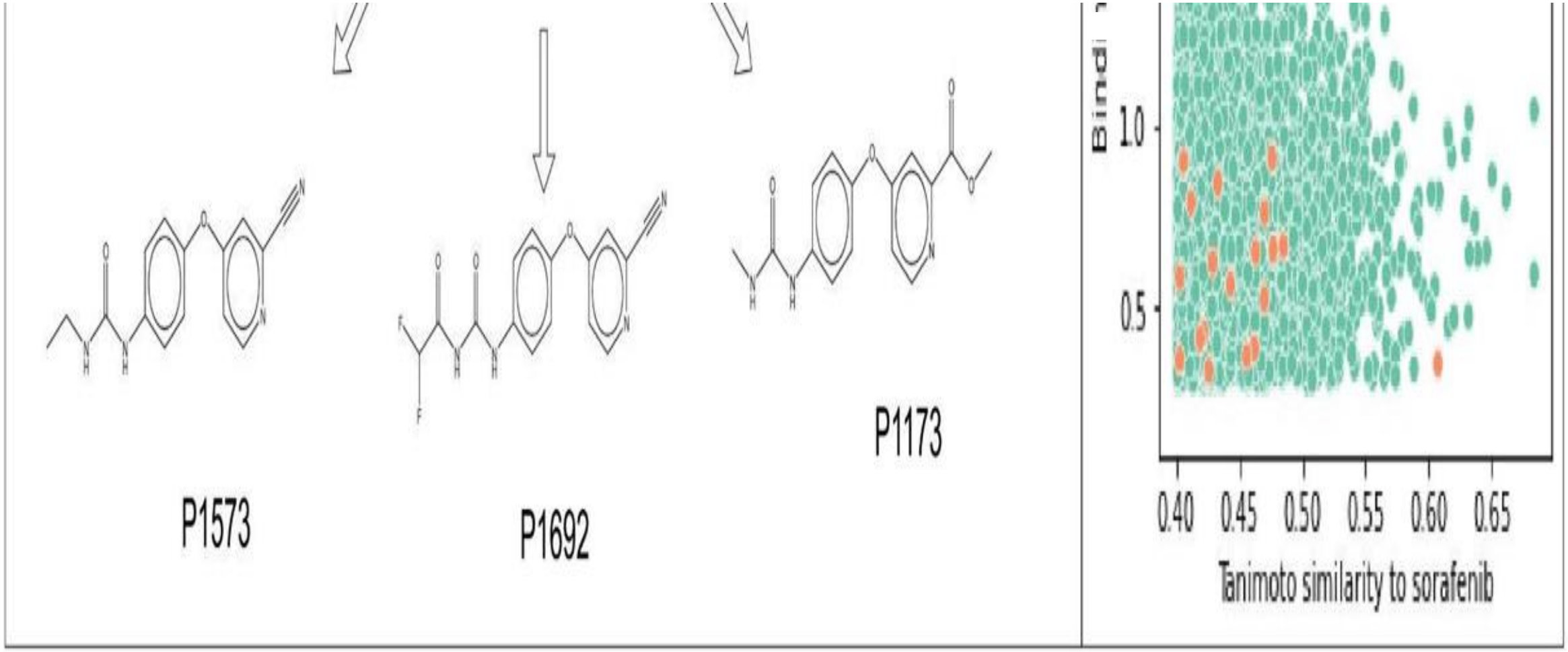
- 5 major interactions
- Higher affinity towards ABCG2 and interaction with BRAF

- 9 leadlike molecules

OVERALL WORKFLOW: SORAFENIB



LEAD LIKE CANDIDATES



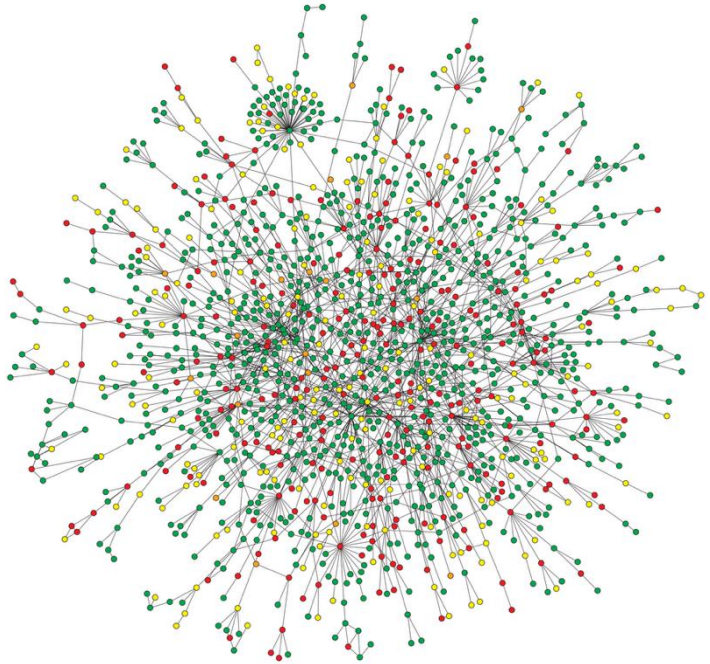
LIMITATIONS AND CHALLENGES

- Property Dependence: Relies on property coverage in generated molecular set rather than direct optimization
- Sparse Properties: Performance may suffer when target properties are extremely rare in chemical space
- Computational Cost: Requires generation of large molecular sets for adequate property coverage
- Reaction Rule Limitations: Restricted to transformations present in USPTO-MIT dataset

SUMMARY

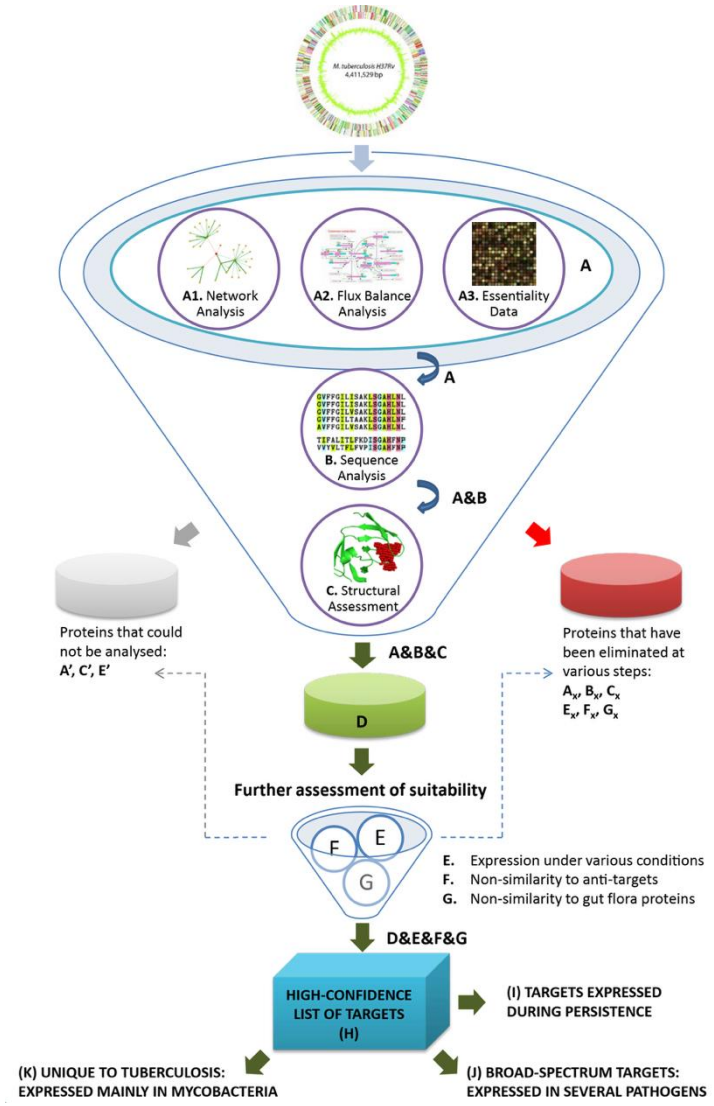
- PURE provides a new paradigm for SCMG/unbiased molecular generation
  - Surmounts many challenges of previous methods/algorithms
- Can potentially design molecules that overcome established resistance mechanisms in cancer/infectious diseases
- Systematic improvement of hit compounds through guided molecular modification, rooted in synthesisability
- Illustrated case study using sorafenib

TARGET IDENTIFICATION:
LEVERAGING NETWORKS AND MODELS TO PREDICT RELEVANT TARGETS IN PATHOGENS



Jeong H *et al.* (2001) *Nature* **411**:41–42

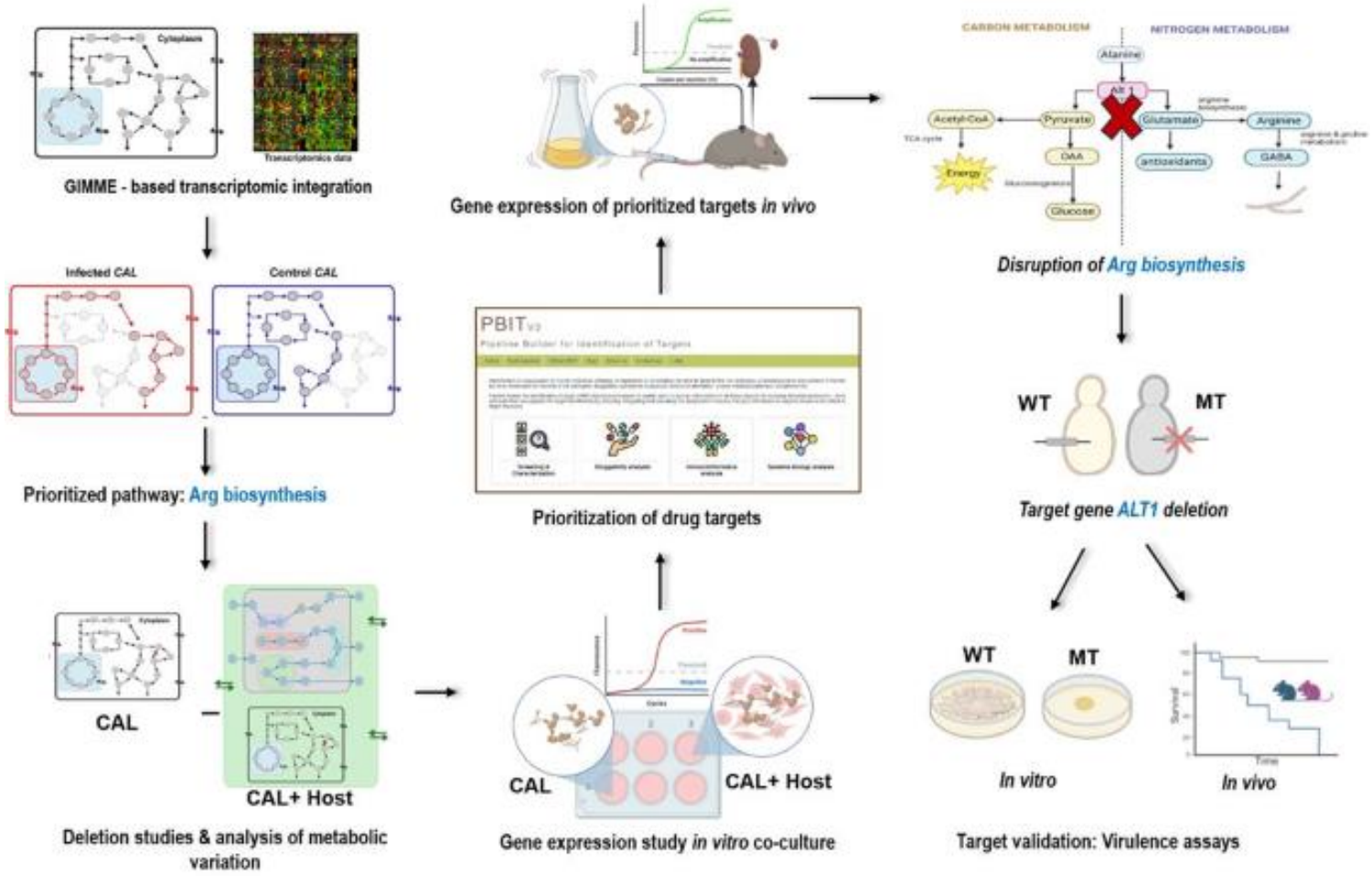
targetTB: TARGET IDENTIFICATION PIPELINE



- Focus on systems-level target identification
  - Combines network analyses
  - Metabolic modelling
- Novel structural analyses
- Generic pipeline, applicable to any organism
- Many high-confidence targets predicted for *Mycobacterium tuberculosis*
  - Includes known targets such as InhA, EmbA and FabH
  - Also passes some proposed targets: DevS, GlfT2, PanK, GlnE, Fas
- Also flags potential problems (cross-reactivity etc.) with known targets

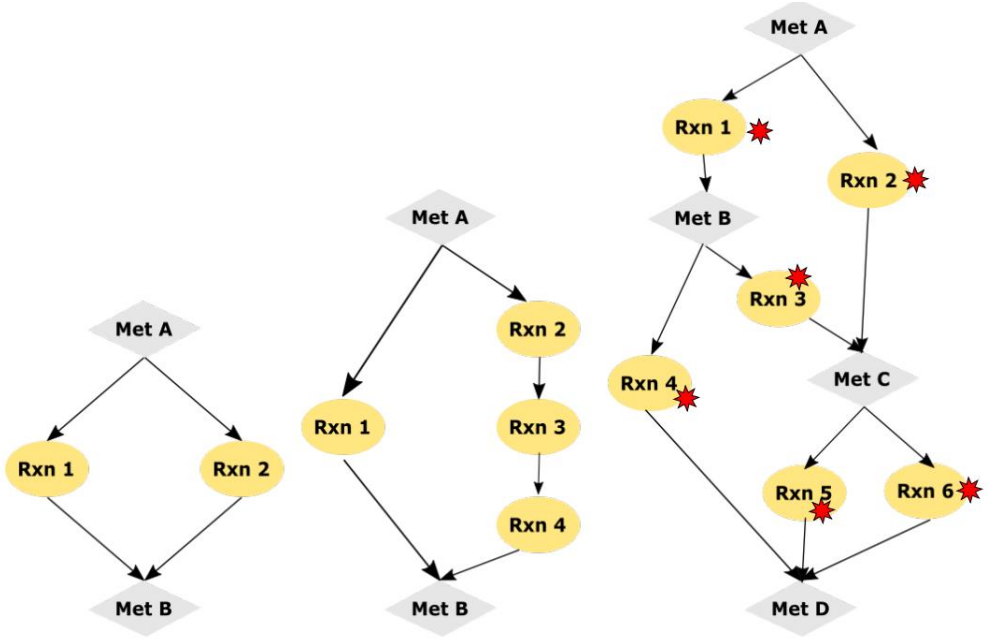
Raman *et al.* (2008) *BMC Syst Biol* **2**:109

METABOLIC VULNERABILITIES IN CANDIDA



- Global threat: Candida albicans → ~70% of Candida infections; invasive cases with drug resistance show mortality up to 63.6%!
- Systems biology approach: Built condition-specific genome-scale metabolic models and a novel host–pathogen integrated model
- **Key finding: Central role of arginine metabolism in virulence**
  - Target identified: **ALT1** (arginine biosynthesis enzyme) emerges as a critical vulnerability
  - Mechanistic insight: During host interaction, arginine synthesis upregulated via proline intermediates
- Experimental validation: ALT1 deletion → impaired virulence and reduced *in vivo* pathogenicity
- Targeting arginine metabolism is a promising antifungal strategy and demonstrates the power of integrated modelling + experiments!

Chakraborty *et al.* (2025) *Cell Commun Signal* **23**:362



FAST-SL: RAPIDLY ENUMERATING SYNTHETIC LETHALS IN METABOLIC NETWORKS

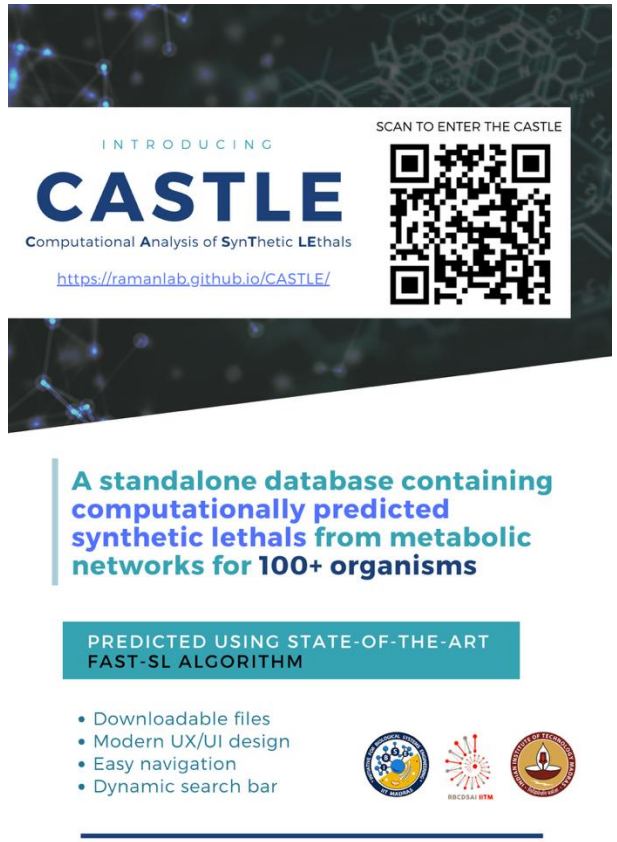
Pratapa A, Balachandran S, K Raman* (2015)

Bioinformatics **31:**3299

/RamanLab/FastSL

FAST-SL: OVERVIEW

- New perspective on a 20 year old algorithm (flux balance analysis)!
  - Very different approach from previous algorithms
- Massively reduced search space for identifying synthetic lethals
- Revealed **complex redundancies and robustness** in metabolic networks at an unprecedented scale[1]
- Adopted widely, including a ***Cell* (2019)** study[2] on *Plasmodium* liver-stage metabolism, where Fast-SL identified parasite-specific lethals



[1]**Sambamoorthy and Raman (2018)** *Bioinformatics* **34:***i981*
[2]**Stanway *et al.* (2019)** *Cell* **179:**1112

http://ramanlab.github.io/CASTLE

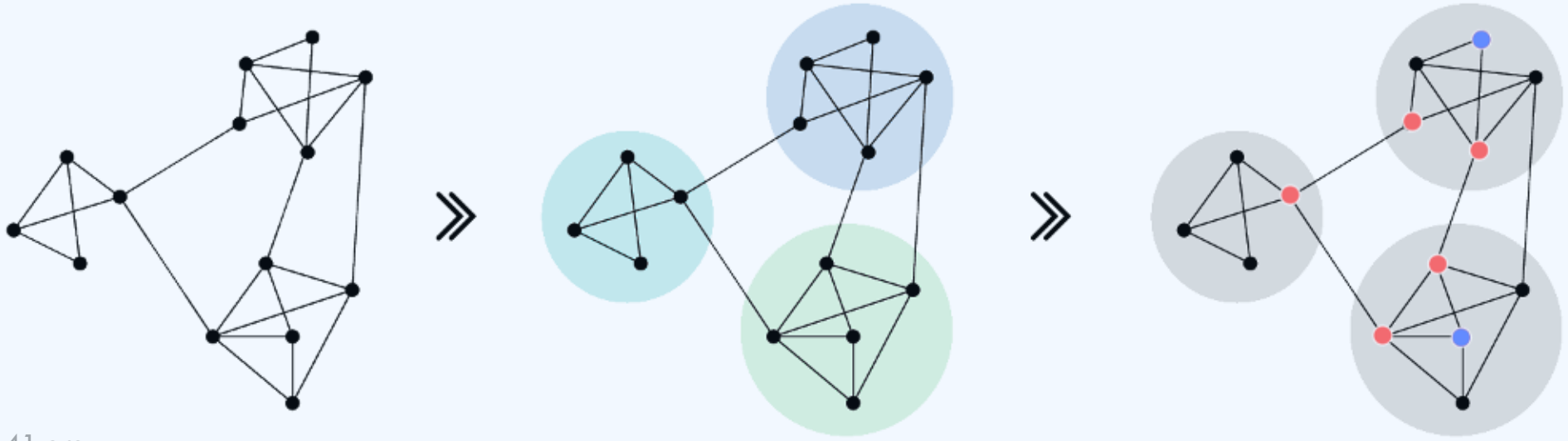


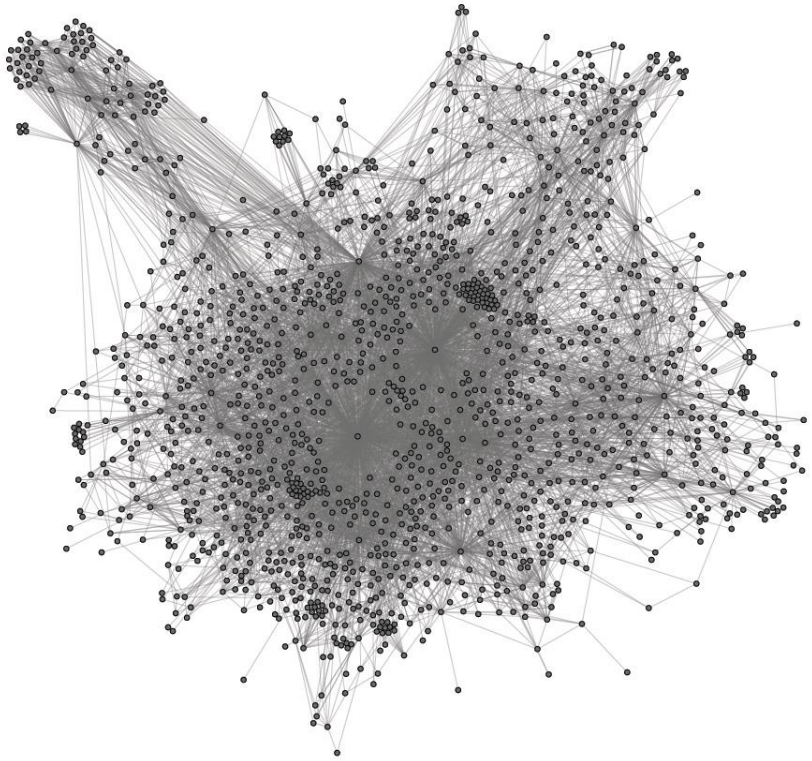
Image Courtesy: lab41.org

PREDICTING ESSENTIAL GENES IN PROTEIN-PROTEIN INTERACTION NETWORKS

K Azhagesan, B Ravindran and K Raman (2018) *PLoS ONE* **13**:e0208722

V Senthamizhan, B Ravindran and K Raman (2021) *Frontiers in Genetics* **12**:722198

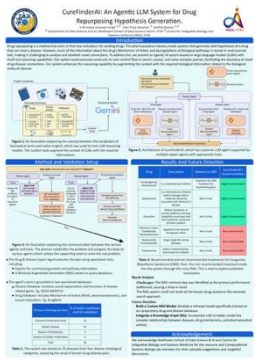
Poster!



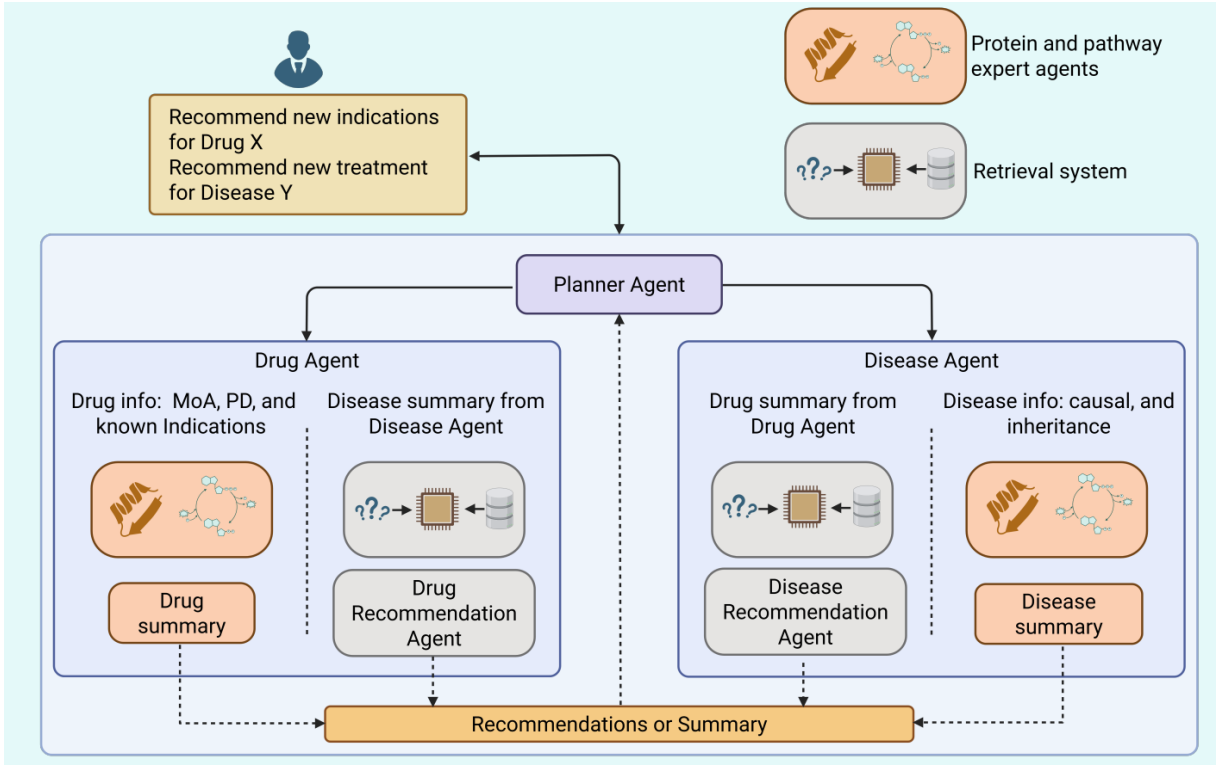
LEVERAGING SCIENTIFIC KNOWLEDGE GRAPHS FOR INDICATION EXPANSION

Srikrishnan B, Yuvaram S, Aswin B, Rit Pratik Mishra, Subha Madhavan, Krishna Kilambi, Karthik Raman

(under preparation)



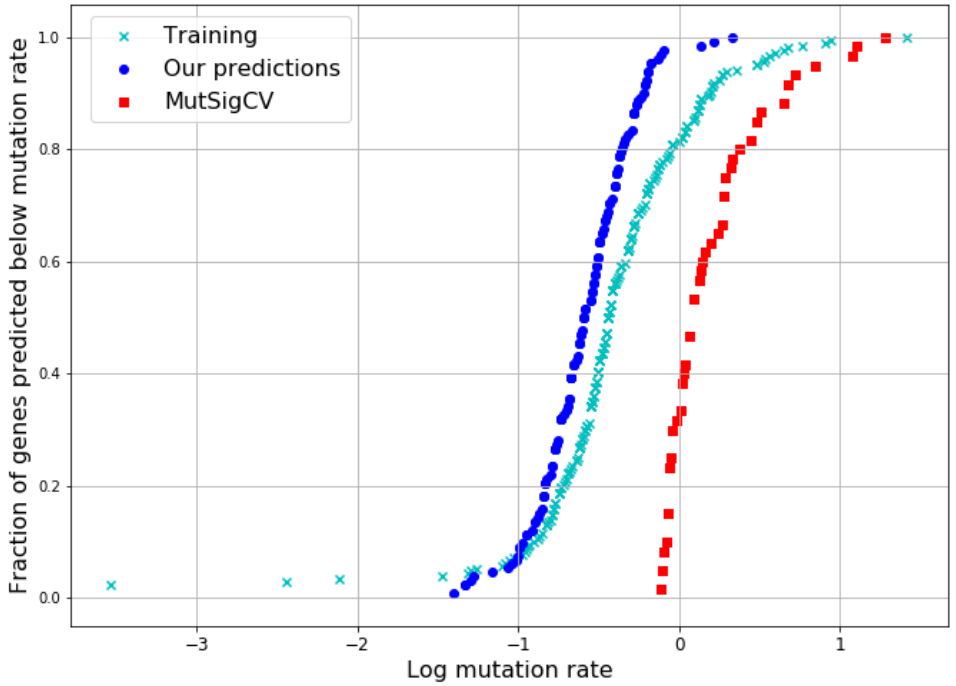
Poster!



CureFinderAI: AN AGENTIC LLM SYSTEM FOR DRUG

Yuvaram Singh, Hari Priya Narahari, Karthik Raman

(under preparation)



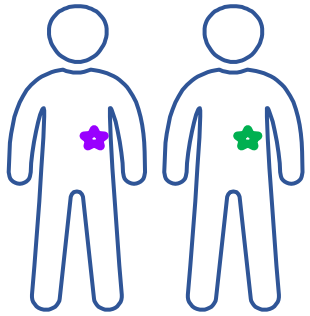
cTaG: IDENTIFY TUMOUR SUPPRESSOR GENES AND ONCOGENES USING FUNCTIONAL MUTATIONS

Sudhakar, Rengaswamy & Raman (2022)

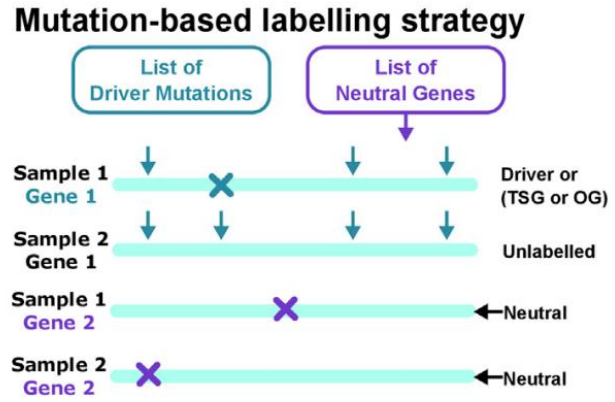
Scientific Reports **12**:5

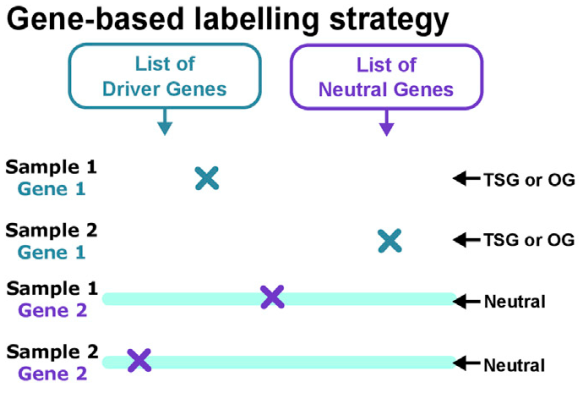
/RamanLab/cTaG

BUT TUMOURS
ARE
HETEROGENEOUS!



CAN WE IDENTIFY DRIVER GENES
IN AN INDIVIDUAL?





PIVOT: PERSONALISED IDENTIFICATION OF DRIVER OGs AND TSGs

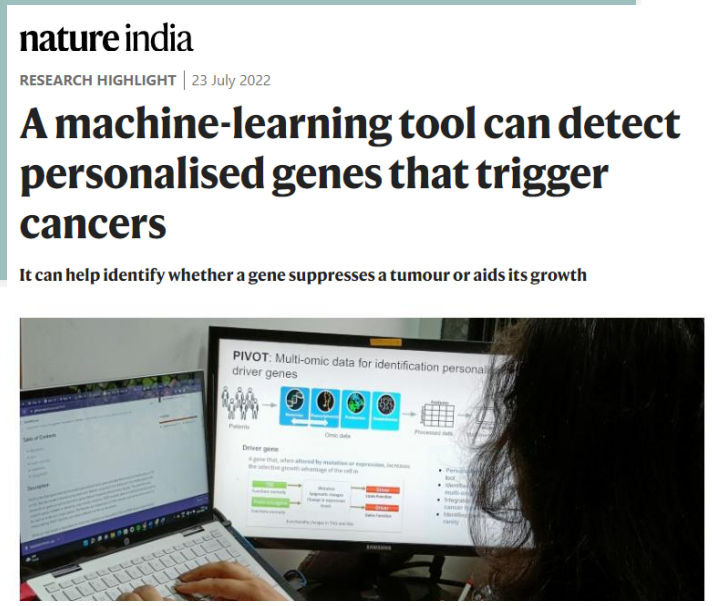
Sudhakar, Rengaswamy & Raman (2022)

Frontiers in Genetics **13**:854190

/RamanLab/PIVOT

PIVOT: SUMMARY

- PIVOT predicts driver genes in a patient using multi-omic data
  - Altered genes labelled as TSG, OG or Neutral
- The same gene can be labelled as TSG in one patient and OG in another patient of the same cancer type!
  - Such predictions are not possible with existing computational tools
- The list of rare predicted genes provides potential genes to target and study—to understand their role in progression of tumours
- First supervised ML model to labels genes as TSGs and OGs in individual patients
- Reaffirms that genes can behave differently based on the signals they receive, the environment, and other


nature india

RESEARCH HIGHLIGHT | 23 July 2022

A machine-learning tool can detect personalised genes that trigger cancers

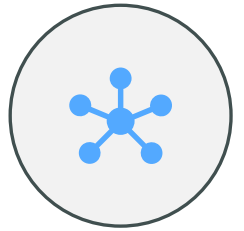
It can help identify whether a gene suppresses a tumour or aids its growth

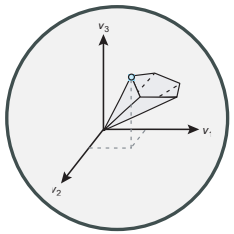
Sithara AA *et al.* (2022) *NAR Genomics & Bioinformatics* **4**:lqac053

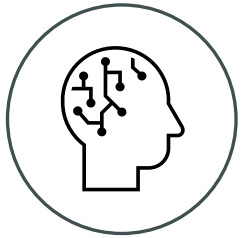
/RamanLab/iCOMIC

TAKE HOME

- **AI, Networks, and Models** are ***all*** useful in our quest to understand biology more deeply – and are reshaping drug discovery
- **AI advances: esp. RL (and going forward, GenAI)** can completely upend current discovery pipelines
- **Networks** provide a powerful paradigm to represent, analyse, and interpret complex relationships in **diverse biological systems**
- **Models can help in grounding predictions** in biological mechanisms and enabling interpretability – useful for target predictions
- Very important to understand the **domain** and get meaningful biologically relevant features for **AI/ML**[1]
  - Mutation-based features/Network-based features/Model-oriented features/…







[1]**Raman *et al.*** (2025) *Clinical and Translational Science* **13**:e70124

ACKNOWLEDGEMENTS

- Aravind Sankar
- Prof. Sayan Ranu (IITM/IITD)
- Abhor Gupta
- Barathi L
- Sean Current (OSU)
- Prof. Rohit Batra
- Prof. B Ravindran
- Prof. Srinivasan Parthasarathy (OSU)
- Kalidas Yeturu (IISc/IIT Tirupati)
- Prof. Nagasuma Chandra (IISc)
- Aditya Pratapa
- Prof. Shankar Balachandran
- Dr. Shuvecchа Chakraborty (NIRRCH)
- Indumathi P
- Yash Gune (NIRRCH)
- Prof. K V Venkatesh (IITB)
- Dr. Susan Idicula-Thomas (NIRRCH)
- Karthik Azhagesan

FUNDING

Department of Biotechnology Govt. of India

European Commission | Co-funded by the Horizon 2020 Framework Programme of the European Union

THANK YOU!