Artificial Intelligence Applications in Computational Biology

4D Genome, Precision Medicine, and Systems Pharmacology

Ivet Bahar

Laufer Center for Physical & Quantitative Biology
Biochemistry and Cell Biology Department, Renaissance School of Medicine, SBU

May 6, 2024– AI workshop, SBU and BNL

Two major groups of computations:

Systems Pharmacology

- Input: data on protein-small molecule interactions
- Output: drug – protein –pathway mapping, including repurposable drugs and side effects in the cell
- Provides a systems level view

Genome-Scale Modeling

- Proteome/genome-scale, or 'omics' data
- Broad range of applications, from precision medicine to disease networks
- High predictive power, but no insights into mechanisms/causality unless complemented by physics-based models

REFERENCE

Available online at www.sciencedirect.com

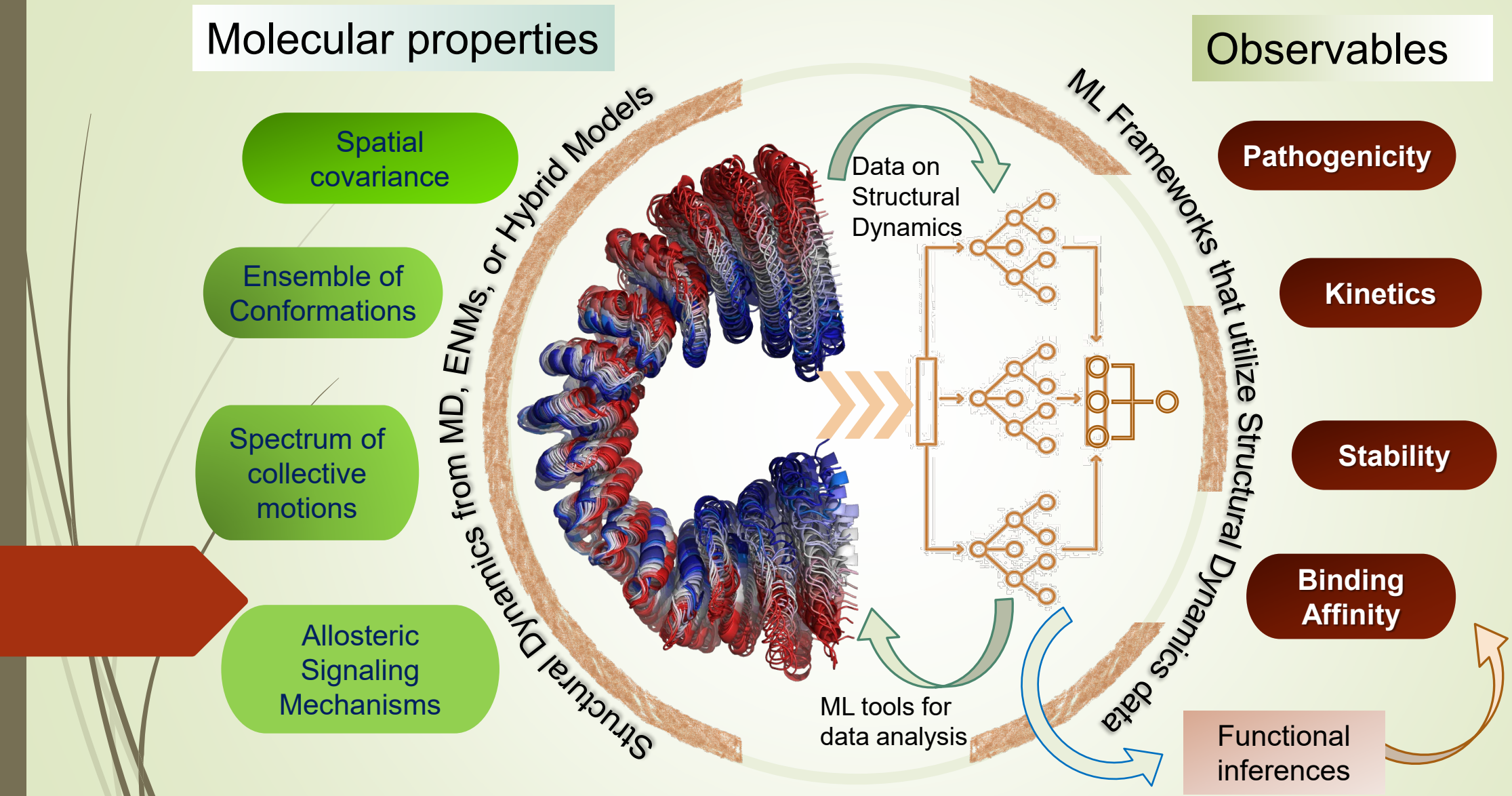
ScienceDirect

Current Opinion in Structural Biology

Mutually beneficial confluence of structure-based modeling of protein dynamics and machine learning methods

Anupam Banerjee[1], Satyaki Saha[1], Nathan C. Tvedt[1,2], Lee-Wei Yang[3,4] and Ivet Bahar[1]

Banerjee, Saha et al. (2023) *Curr Opin Struct Biol* 78:102517



Banerjee, Saha et al. (2023) *Curr Opin Struct Biol* 78:102517

Plan

Quantitative Systems Pharmacology

- Druggability simulations, pharmacophore modeling for drug discovery
- ML/AI-based assessment of alternative drugs, systems-level effects

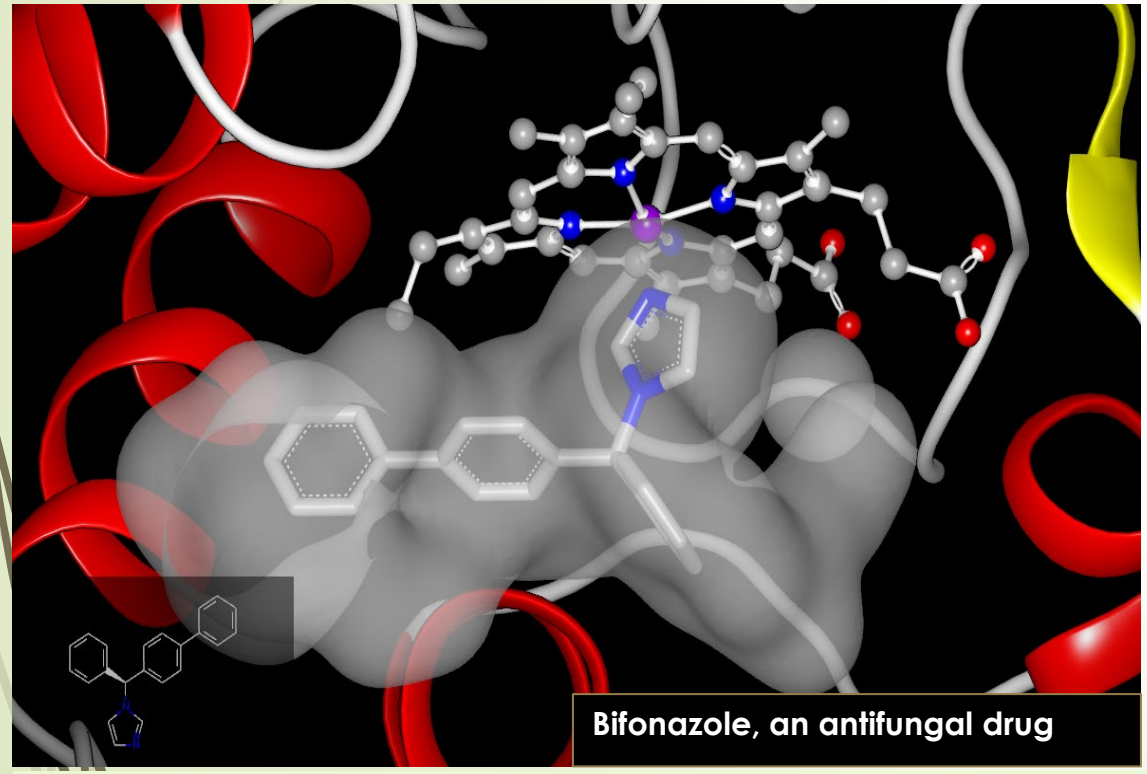
Genome/Proteome-Scale Models and Predictions

- ML-based prediction of the pathogenicity of point mutations and deletions
- 4D modeling the structure and dynamics of the genome

Quantitative Systems Pharmacology (QSP)

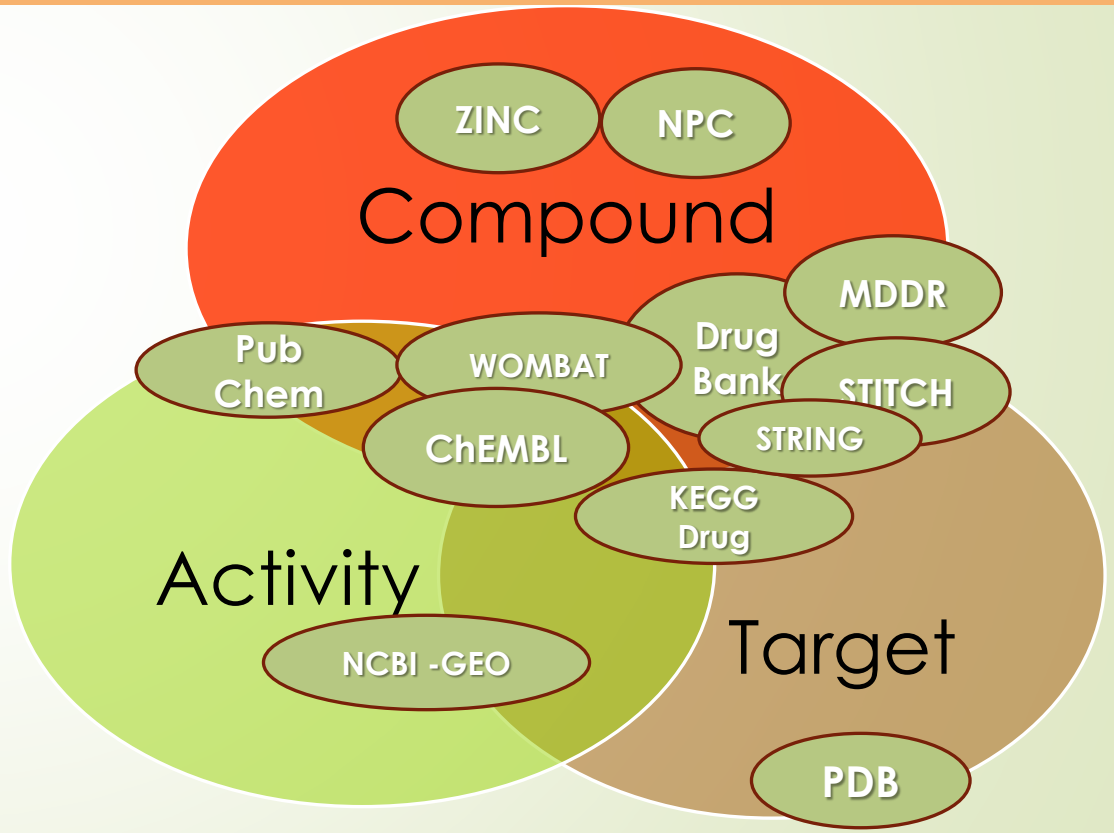
Two major groups of QSP computations:

Modeling & Simulations

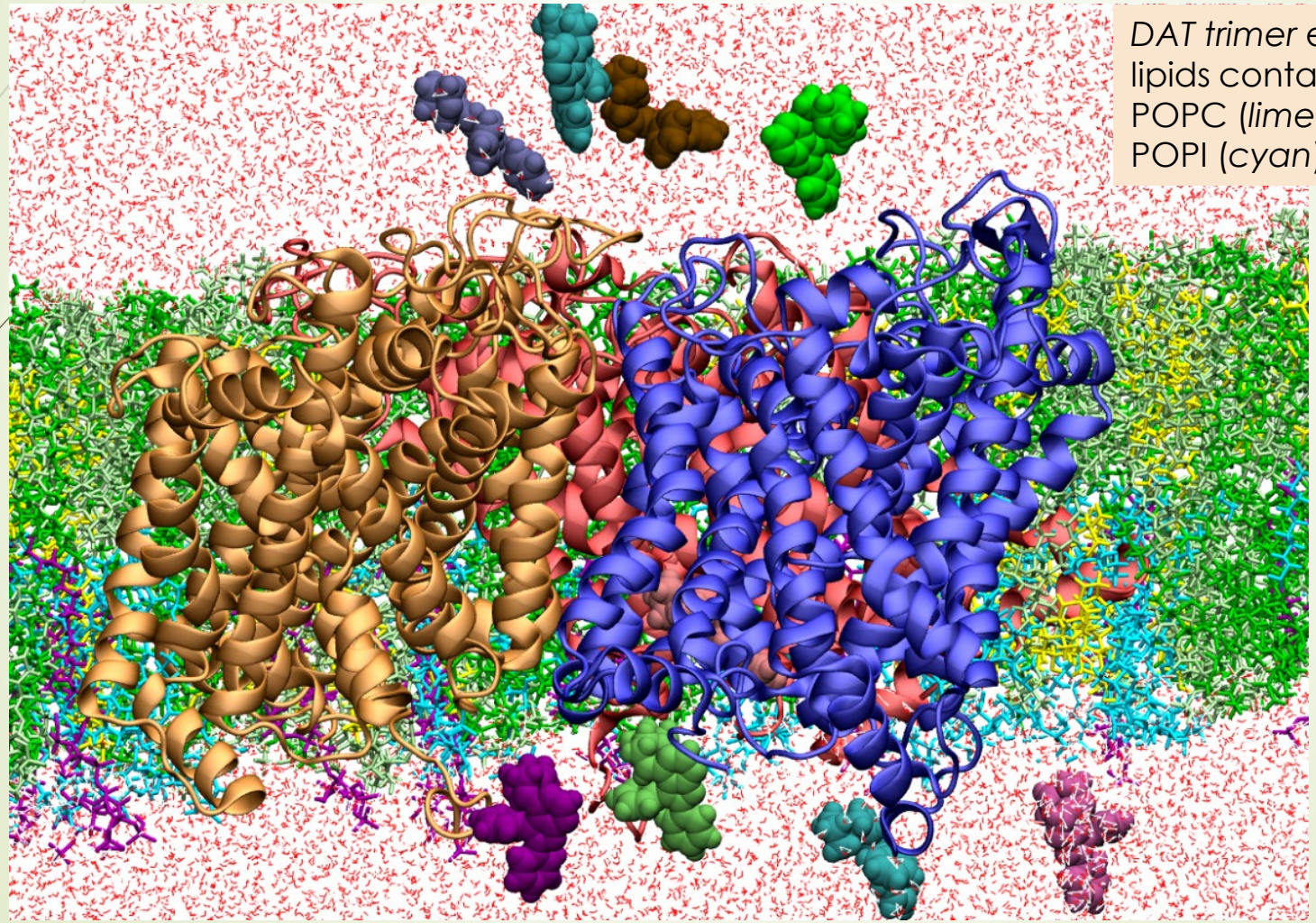


Bakan A, Kapralov AA, Bayir H, Hu F, Kagan VE, Bahar I (2015) Inhibition of Peroxidase Activity of Cytochrome c: *De Novo* Compound Discovery and Validation *Mol Pharmacol* **88**: 421-7

Data-driven Analyses



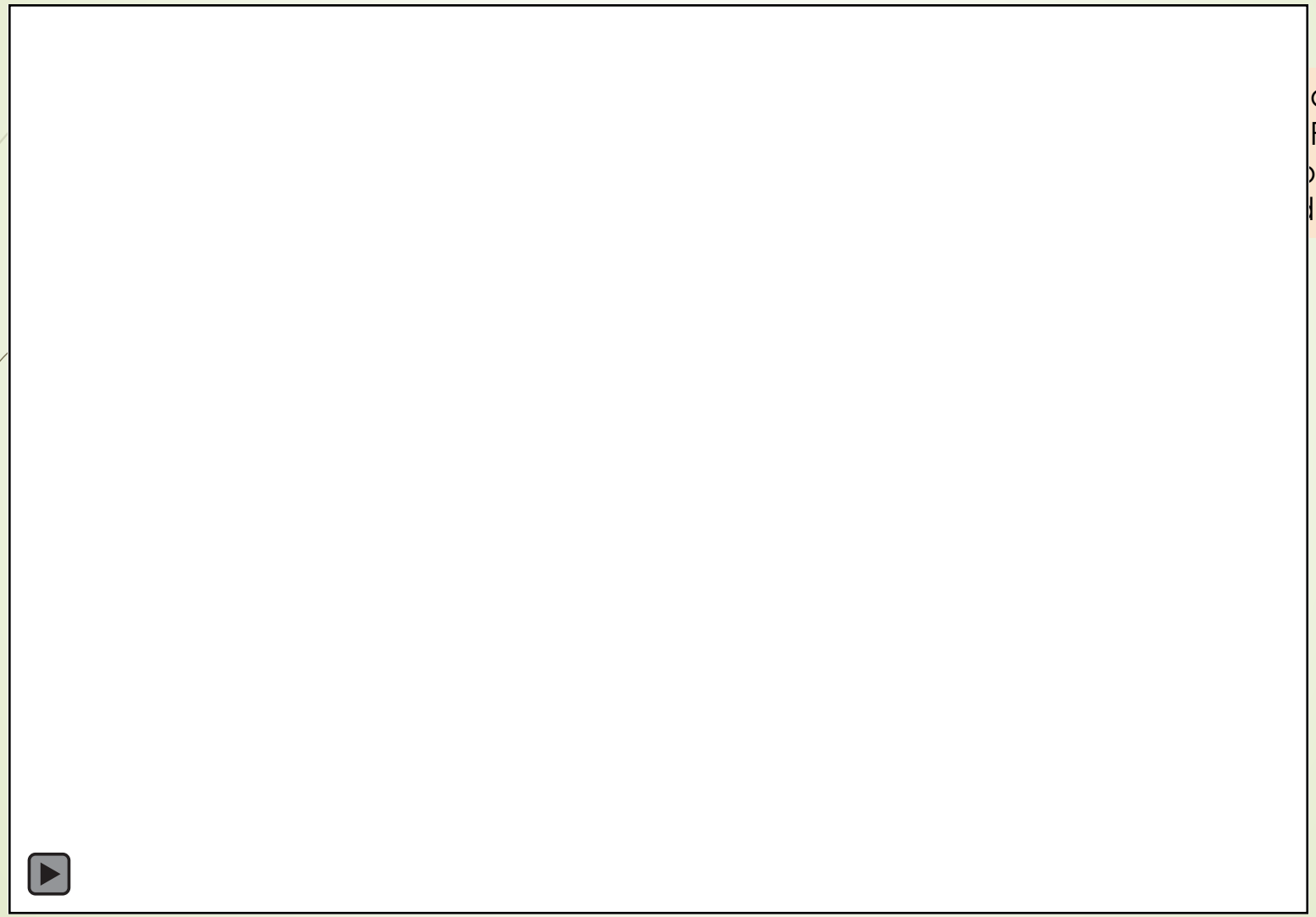
Druggability simulations



DAT trimer embedded into neuronal lipids containing POPE (*green*), POPC (*lime*), cholesterol (*yellow*), POPI (*cyan*), and PIP2 (*purple*).

Cheng MH, Ponzoni L, Sorkina T, Lee JY, Zhang S, Sorkin A, Bahar I. *Neuropharmacology* (2019)

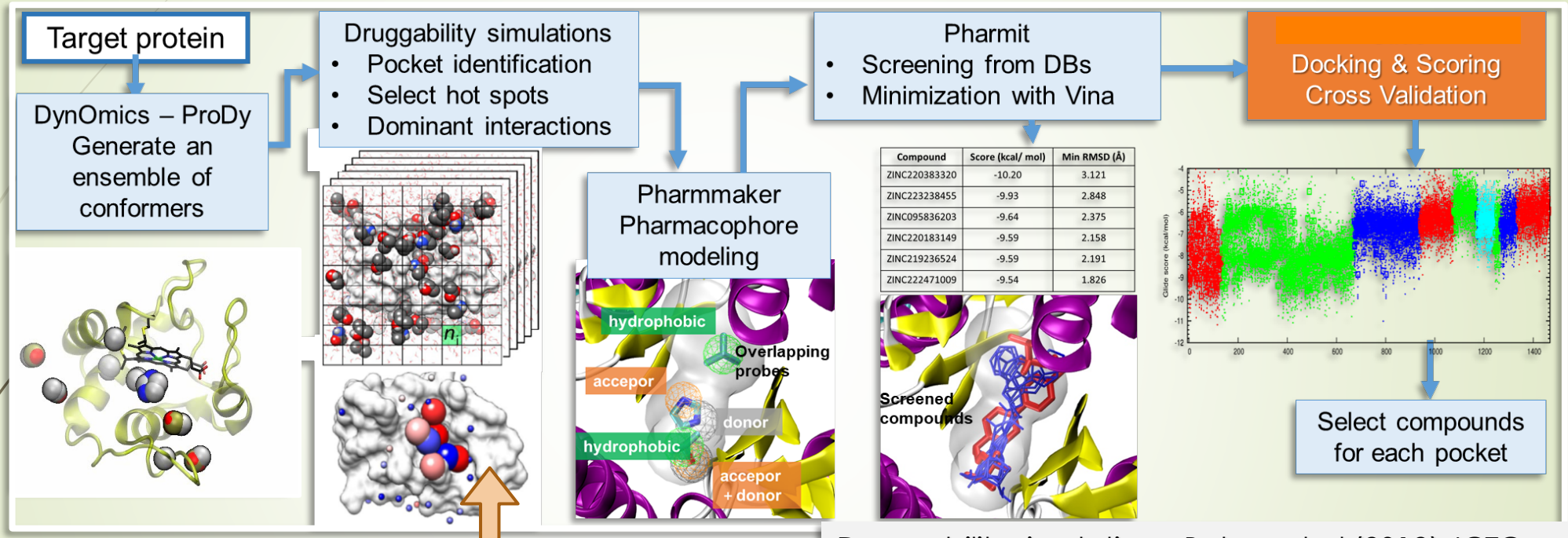
Trimerization of Dopamine Transporter Triggered by AIM-100 Binding: Molecular Mechanisms and Effect of Mutations



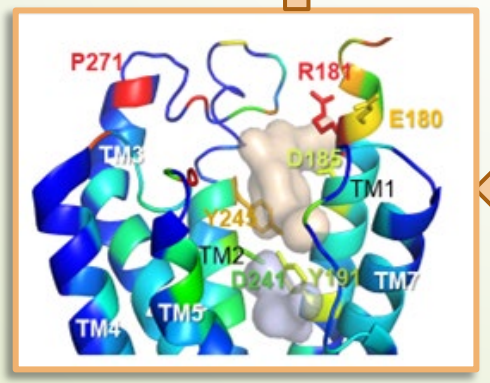
dded into neuronal
POPE (*green*),
lesterol (*yellow*),
PIP2 (*purple*).

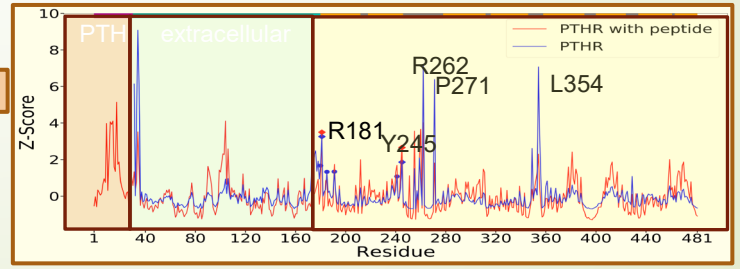
Cheng MH, Ponzoni L, Sorkina T, Lee JY, Zhang S, Sorkin A, Bahar I. *Neuropharmacology* (2019)

ProDy pipeline for identifying druggable sites and building pharmacophore models



Druggability simulations: Bakan et al (2012) *JCTC*
Pharmacophore modeling: Lee et al (2020) *Protein Sci*

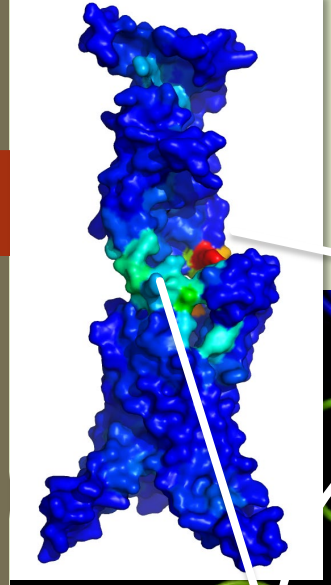


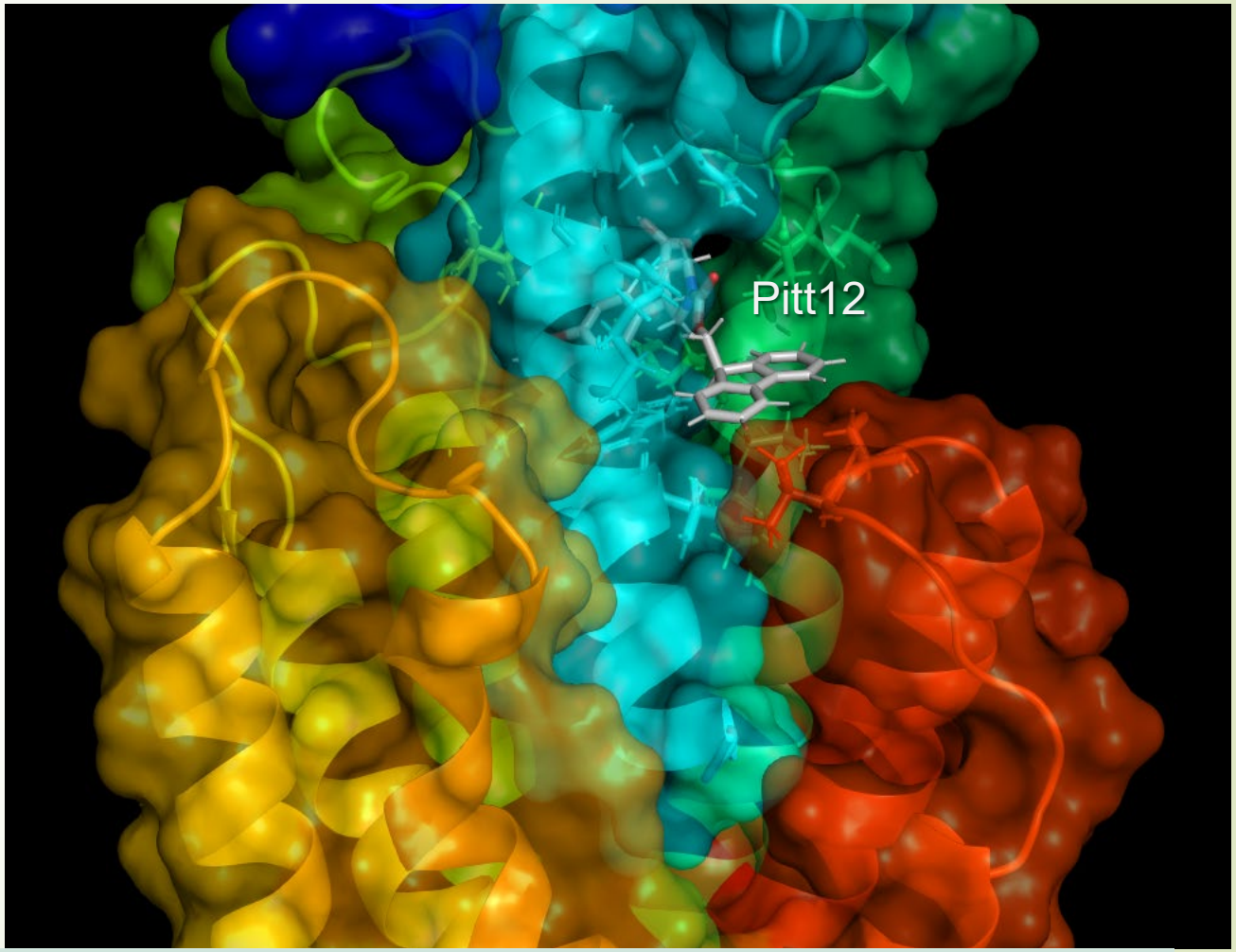


ESSA for selecting essential sites from amongst druggable sites

Kaynak et al (2020)
Comp Struct Biotech J

Discovery of a new modulator of PTHR signaling





Structure-guided identification of druggable allosteric sites in receptors (such as this Class B GPCR) can aid the design of novel therapeutic candidates for metabolic diseases,

Nat Chem Biol. 2020 October ; 16(10): 1096–1104. doi:10.1038/s41589-020-0567-0.

Allosteric interactions in the parathyroid hormone GPCR–arrestin complex formation

Lisa J. Clark[1,2,†], James Krieger[3,†], Alex D. White[1,4], Vasyl Bondarenko[5], Saifei Lei[1], Fei Fang[1], Ji Young Lee[3], Pemra Doruker[3], Thore Böttke[6], Frederic Jean-Alphonse[1], Pei Tang[1,3,5], Thomas J. Gardella[7], Kunhong Xiao[1], Ieva Sutkeviciute[1], Irene Coin[6], Ivet Bahar[3,*], Jean-Pierre Vilardaga[1,*]

JiYoung Lee, PhD

nature chemical biology

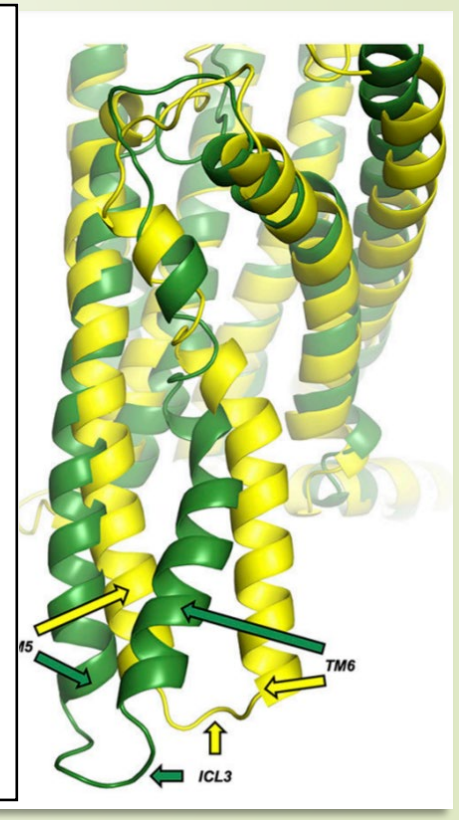
ARTICLES

https://doi.org/10.1038/s41589-021-00929-w

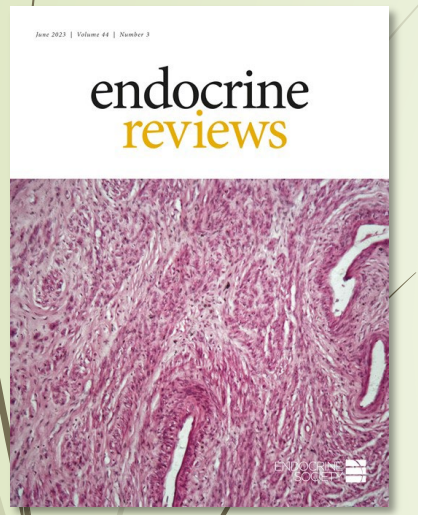
Check for updates

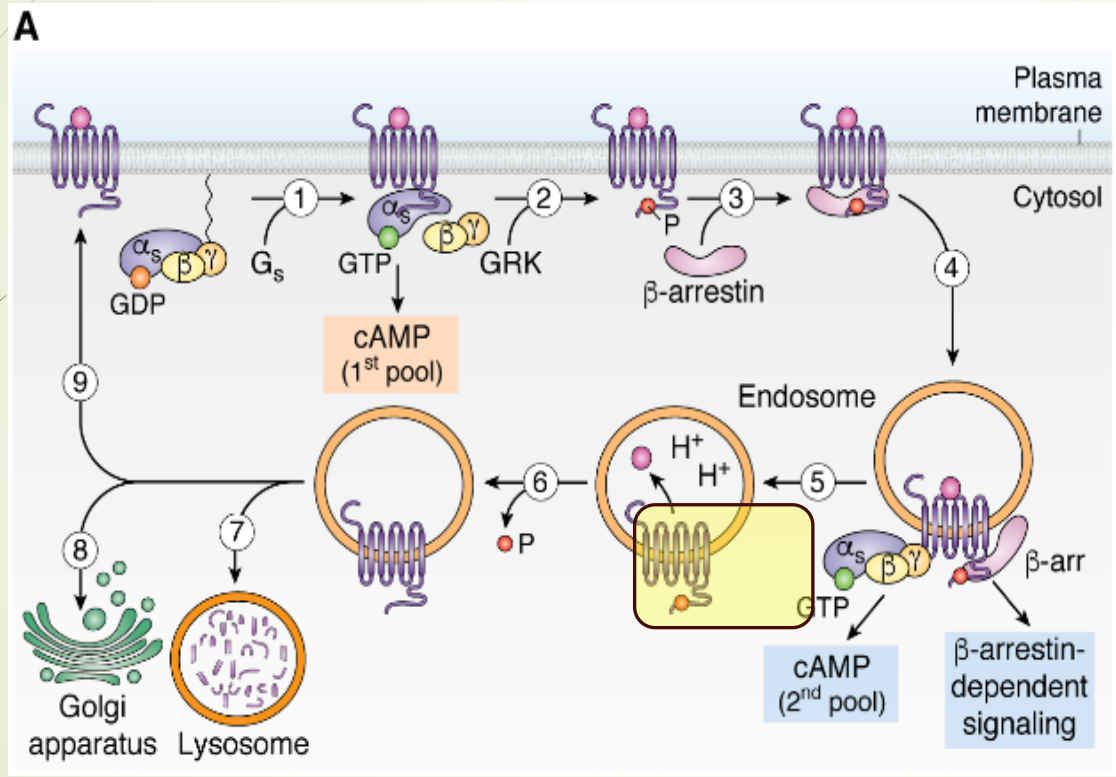
Precise druggability of the PTH type 1 receptor

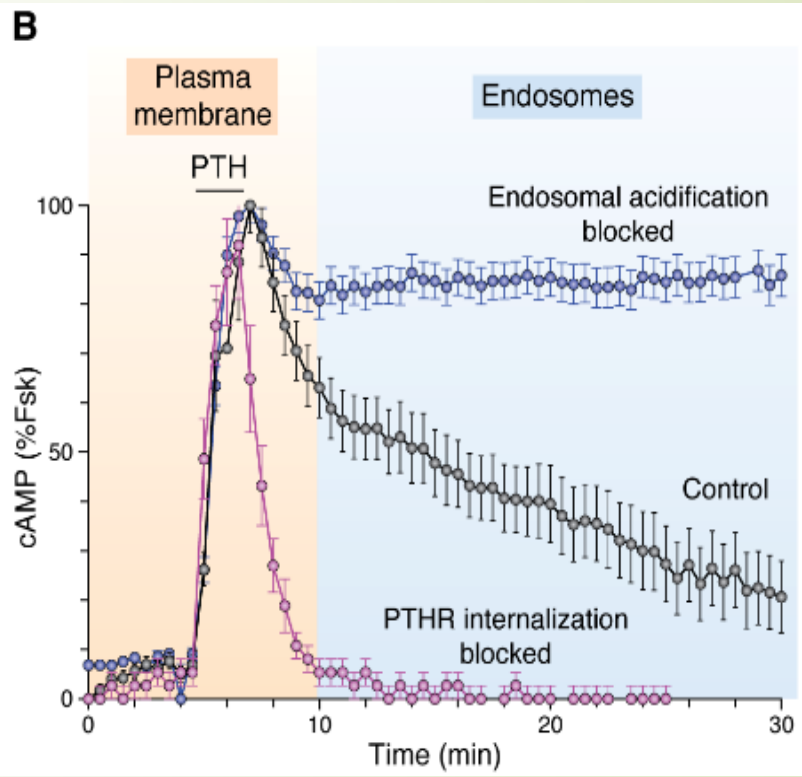
Ieva Sutkeviciute[1✉], Ji Young Lee[2,9], Alex D. White[1,9], Christian Santa Maria[3], Karina A. Peña[1], Sofya Savransky[1,4], Pemra Doruker[2], Hongchun Li[2,6], Saifei Lei[1,7], Burak Kaynak[2], Chialing Tu[3], Lisa J. Clark[1,8], Subramaniam Sanker[1], Thomas J. Gardella[5], Wenhan Chang[3], Ivet Bahar[2✉] and Jean-Pierre Vilardaga[1✉]



PTHR signaling and prolonged cAMP production



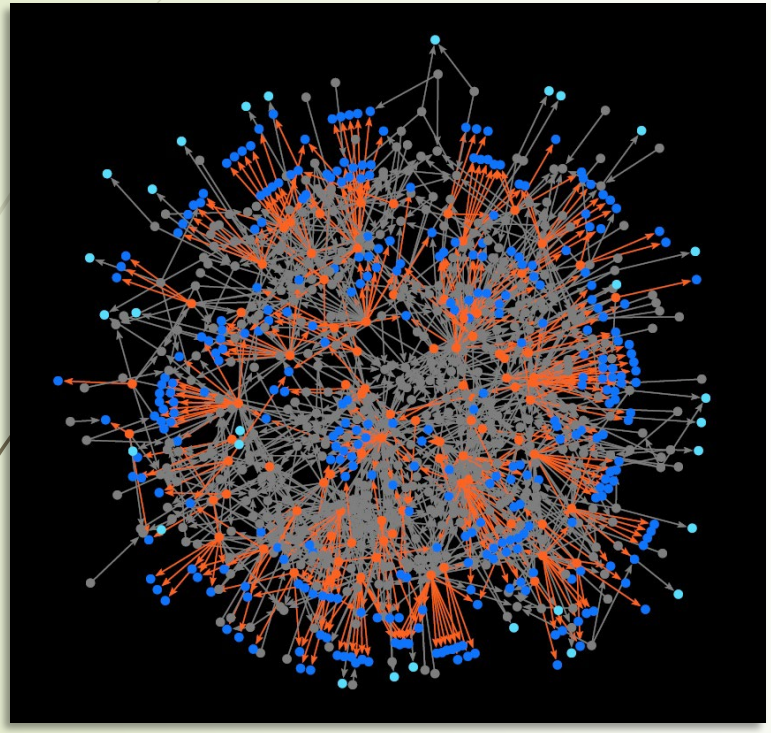


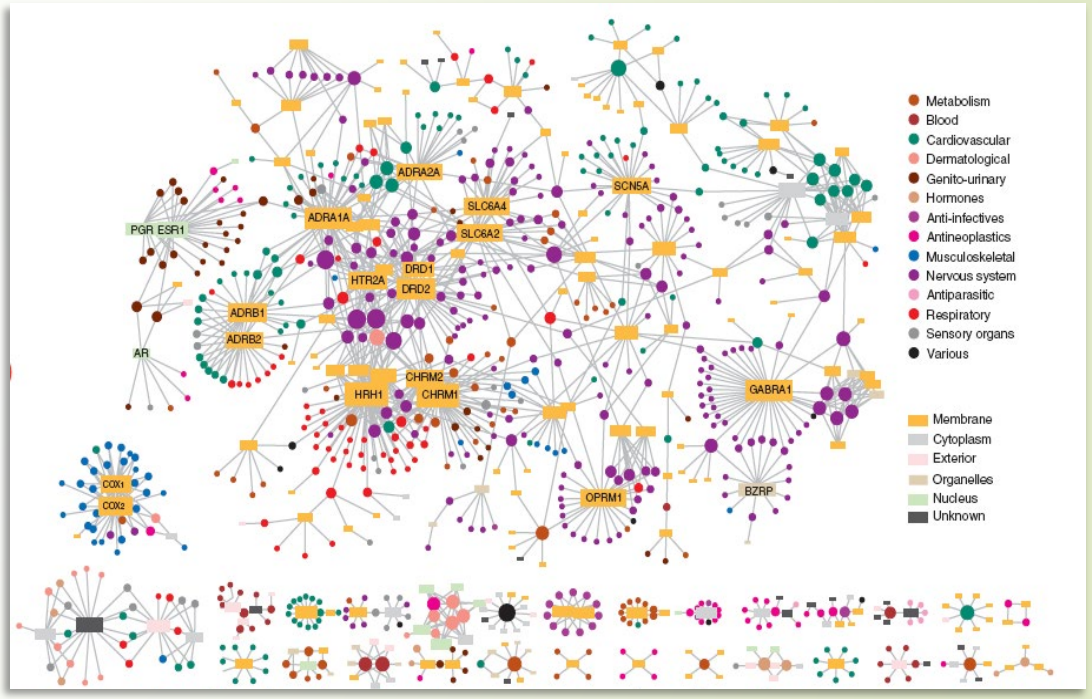


JP Vilardaga, LJ Clark, AD White, I Sutkeviciute, JY Lee, I Bahar (2023)
Endocr Rev 44, 474–491, https://doi.org/10.1210/endrev/bnac032

Quantitative Systems Pharmacology

Promiscuity of drugs and pleiotropy of proteins

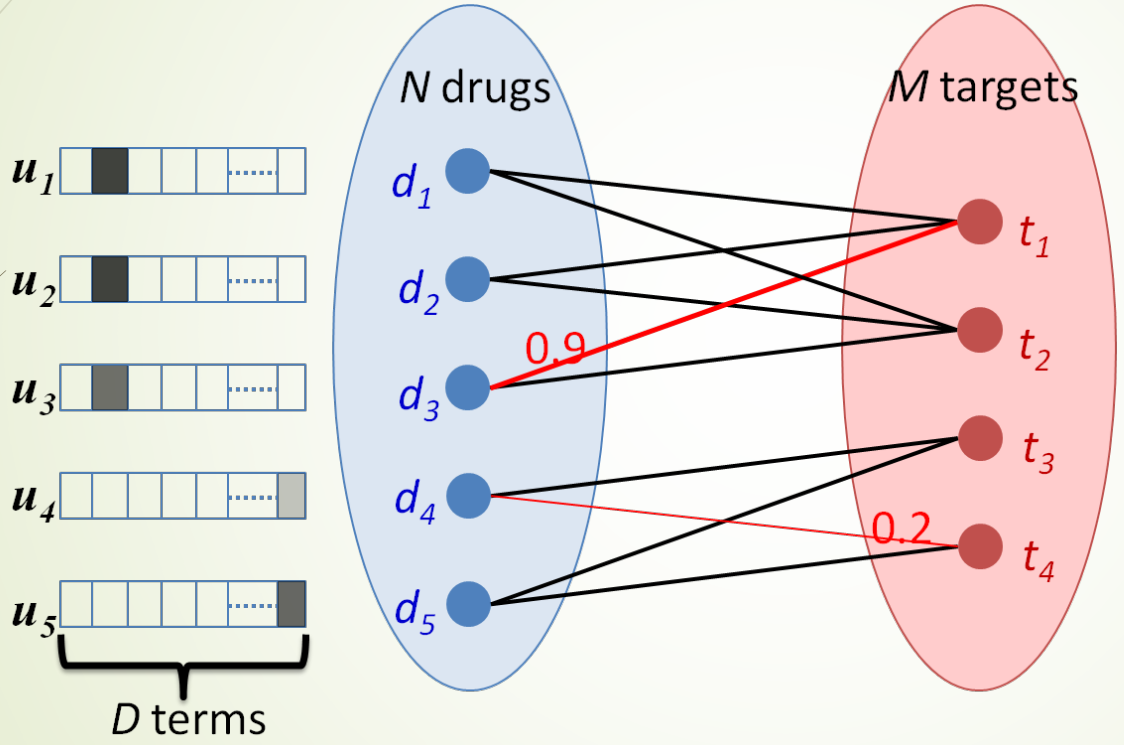




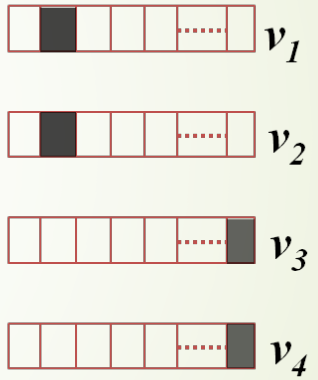
Andrew L Hopkins **Network pharmacology: the next paradigm in drug discovery** *Nature Chemical Biology* **4**, 682–690, 2008

Yıldırım, Goh, Cusick, Barabási & Vidal **Drug—target network** *Nature Biotechnology* **25**, 1119–1126, 2007)

Drug Repurposing
Probabilistic Matrix Factorization



Latent vectors characteristic of each drug (u_i) or protein (v_i), deduced from PMF analyses of the complete dataset of drug-target interactions



Cobanoglu et al. (2013) Predicting drug-target interactions using probabilistic matrix factorization *J Chem Inf Model* **53**: 3399-409

Source: **DrugBank**
1,576 targets, 1,509 drugs, 5,630 interactions

QUARTATAWEB

13,723 drugs, 5,235 targets

DRUGBANK
Open Data Drug & Drug Target Database

Bioinformatics. 2020 Jun 1;36(12):3935-3937. doi: 10.1093/bioinformatics/btaa210.

QuartataWeb: Integrated Chemical–Protein–Pathway Mapping for Polypharmacology and Chemogenomics

Hongchun Li [1 2], Fen Pei [1 3], D Lansing Taylor [1 3], Ivet Bahar [1 3]

FULL TEXT LINKS
OXFORD ACADEMIC
PMC FREE Full text

ACTIONS

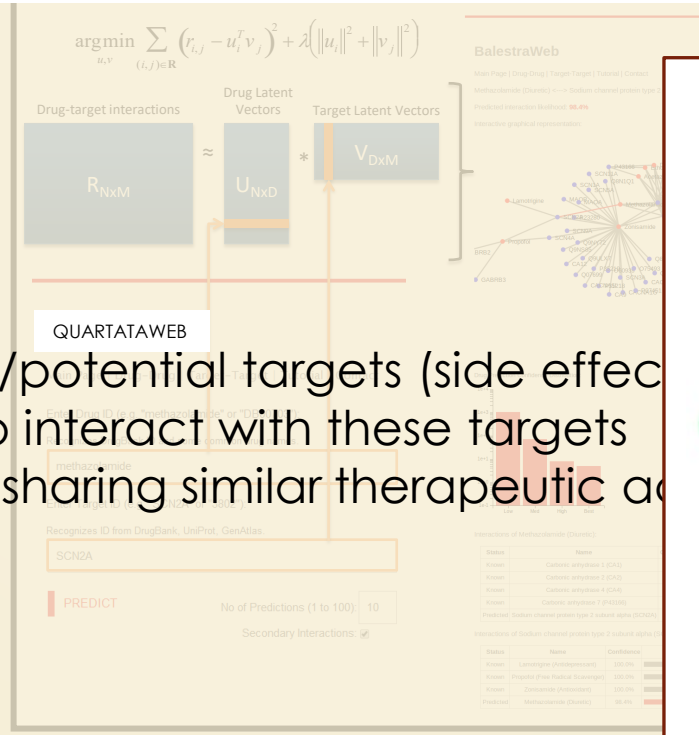
Input:

Set of query drugs

Output:

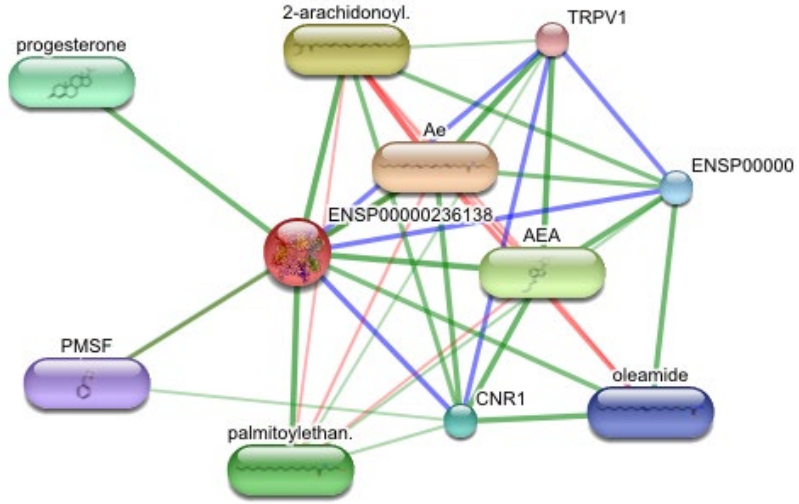
Target(s) + predicted/potential targets (side effec
Other drugs known to interact with these targets
Repurposable drugs sharing similar therapeutic a

Enriched pathways
Mechanism of action



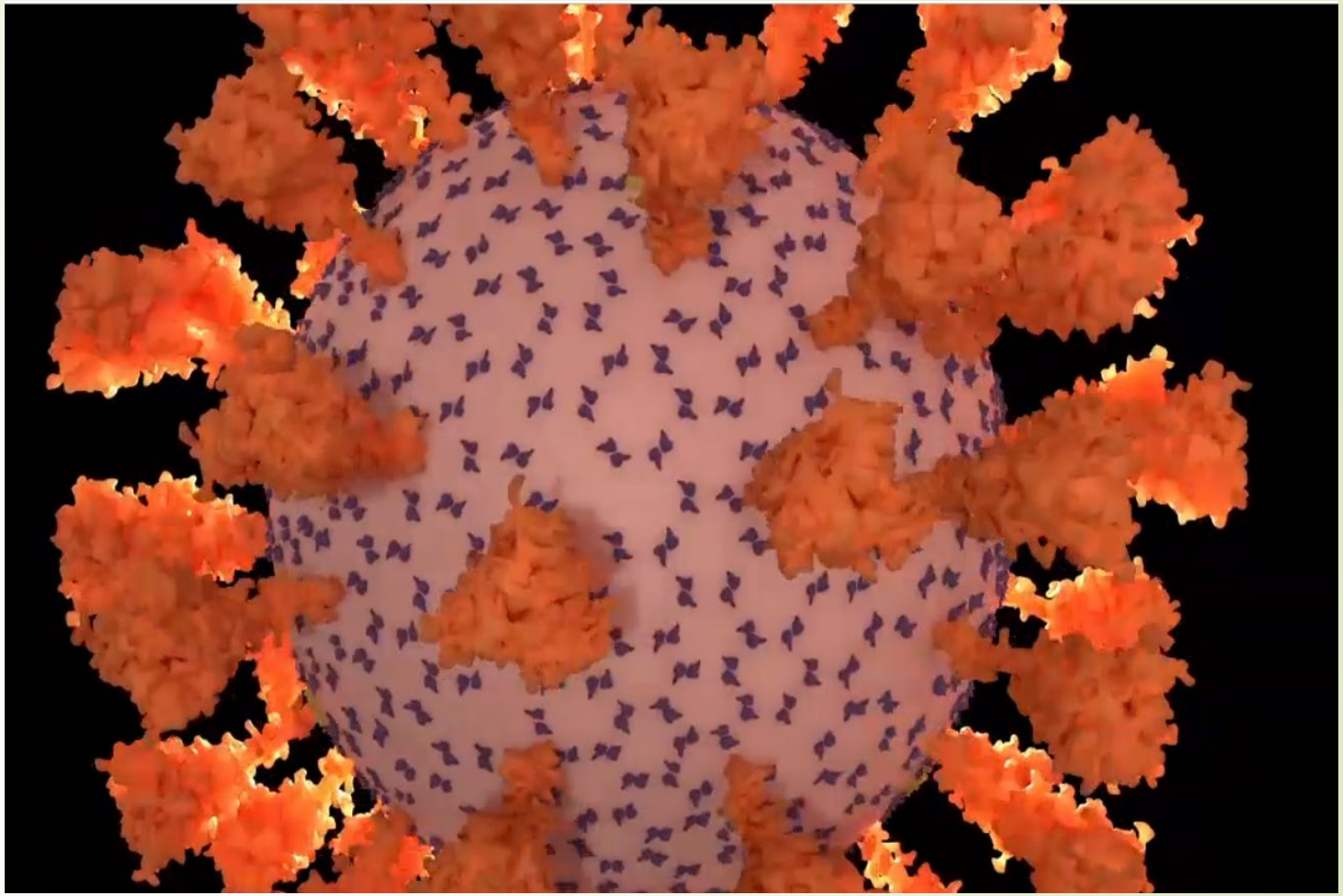
390,000 chemicals, 3.6 million proteins

STITCH 3.1



Li et al (2020) *Bioinformatics* **36**:3935-37.

Identification of repurposable drugs against SARS-CoV-2

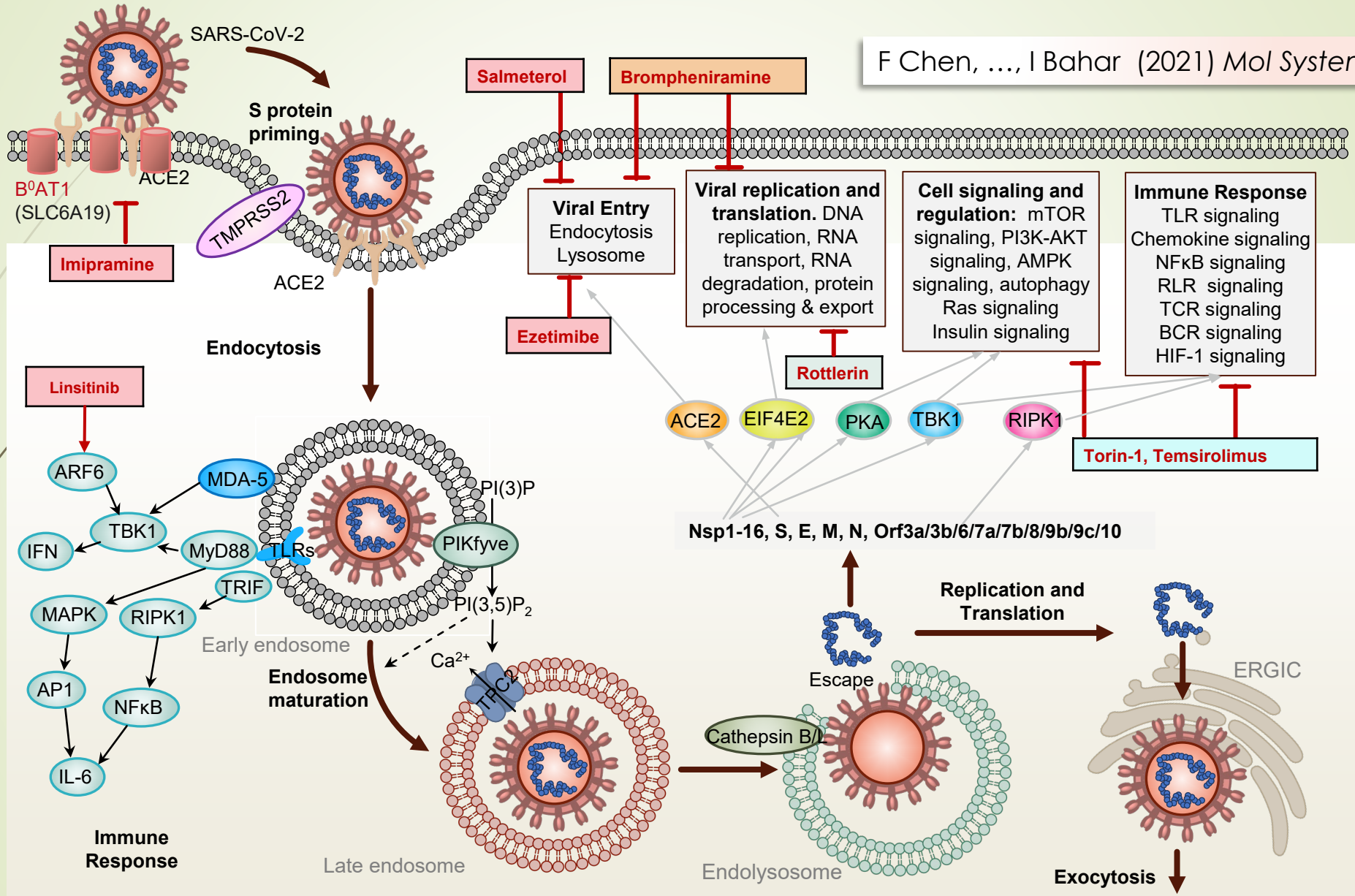


using a QSP approach

Chen F, Shi Q, Pei F, Vogt A, Porritt RA, Garcia G, Gomez AC, Cheng MH, Schurdak ME, Chan SY, Arumugaswami V, Stern A, Taylor L, Arditi M, Bahar I. A Systems-level study reveals host-targeted repurposable drugs against SARS-CoV-2 infection *Molecular Systems Biol.* 17:e10239 (2021).

Identification of drugs targeting four different modules

F Chen, …, I Bahar (2021) *Mol Systems Biology*



Genome/Proteome-Scale Analyses

Proteome-scale machine learning

Luca Ponzoni

Predicting functional impact of mutations

Luca Ponzoni & Ivet Bahar *PNAS* (2018)

Ponzoni, Penaherrera, Oltvai & Bahar *Bioinformatics* (2020)

Novel Approach

Combining physical sciences & ML

Features used for classification

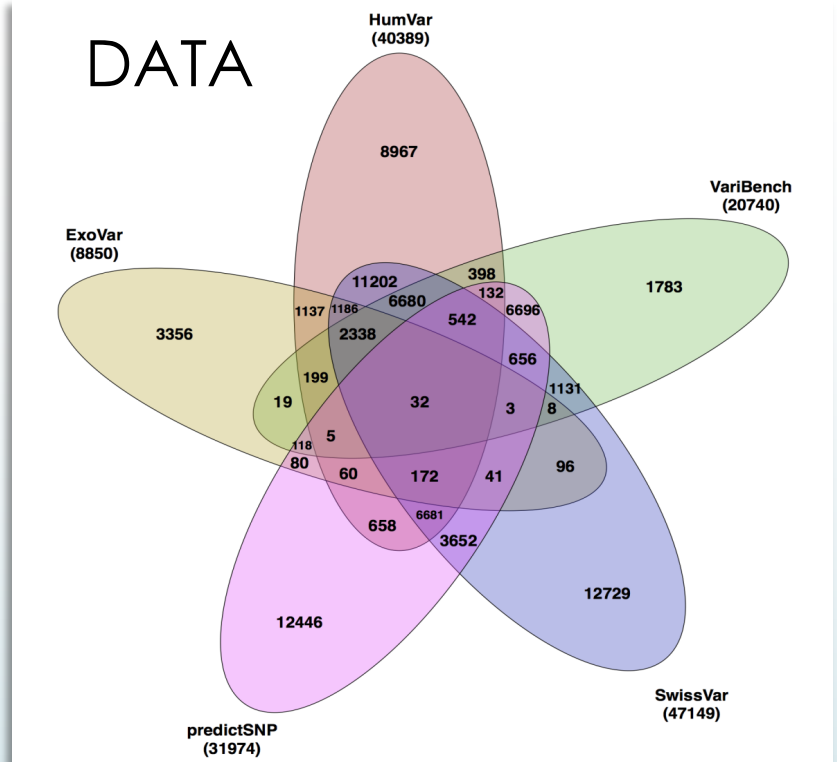
SEQuence-based features:
- conservation
- Δ conservation (wt vs mutated allele)

STRuctural feature:
- Solvent Accessible Surface Area

DYNamical features:
- GNM Mean-Square Fluctuations
- PRS analysis (allosteric effectors/sensors)
- MechStiff

Random Forest classification
- trained on 20,000 annotated human variants
- 10-fold cross-validation procedure

Aims:
1. **evaluate the accuracy attainable by combining SEQ-STR-DYN features**
2. **quantify contribution of dynamical features**



Integrated Dataset
~ 20,000 unique SAVs with known structure

Ponzoni L, Bahar I. (2018) Structural dynamics is a determinant of the functional significance of missense variants. *Proc Natl Acad Sci USA* **115**: 4164-4169

Significant improvement in predictions

Upon including DYN features

Features used for classification

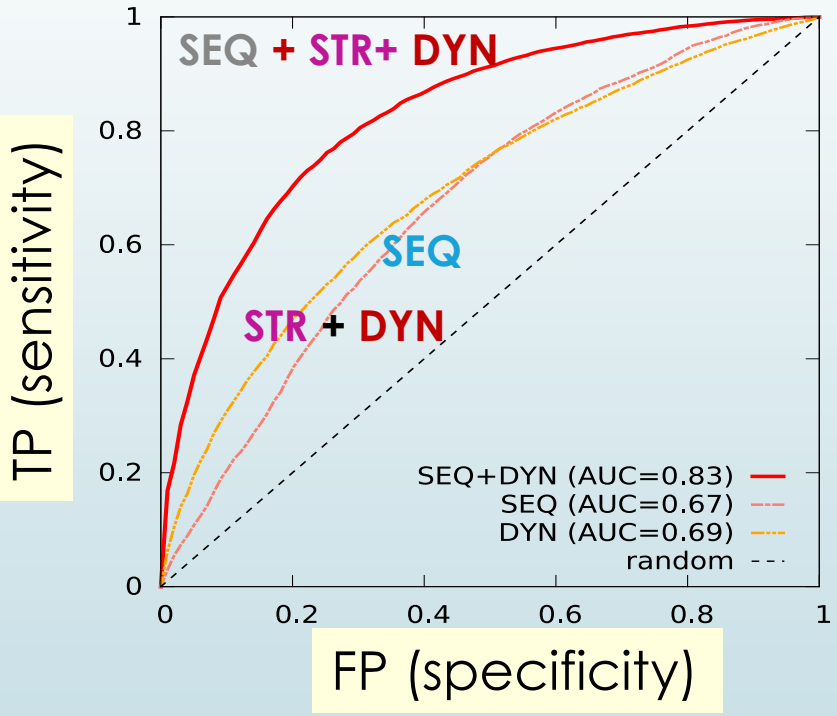
SEQuence-based features:
- conservation
- Δ conservation (wt vs mutated allele)

STRuctural feature:
- Solvent Accessible Surface Area

DYNamical features:
- GNM Mean-Square Fluctuations
- PRS analysis (allosteric effectors/sensors)
- MechStiff

Random Forest classification
- trained on 20,000 annotated human variants
- 10-fold cross-validation procedure

Aims:
1. **evaluate the accuracy attainable by combining SEQ-STR-DYN features**
2. **quantify contribution of dynamical features**



Ponzoni L, Bahar I. (2018) Structural dynamics is a determinant of the functional significance of missense variants. *Proc Natl Acad Sci USA* **115**: 4164-4169

RAPSODY

rapsody.csb.pitt.edu

RHAPSODY

Rapid High-Accuracy Prediction of SAV Outcome based on DYnamics

Home | FAQs | Download

This tool provides a prediction of pathogenicity for Single Amino acid Variants (SAVs) by employing a Random Forest classifier trained on both sequence-based and structural/dynamical features.

Option 1: Get predictions based on both sequence-based and structural/dynamical features, by uploading a:

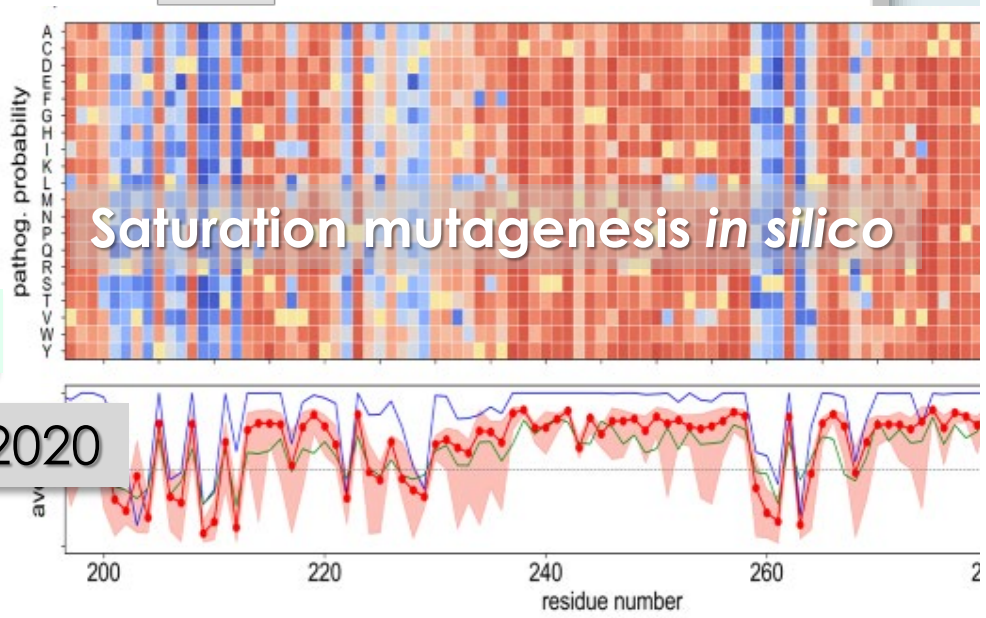
PolyPhen-2 output file (see instructions) Choose File No file chosen

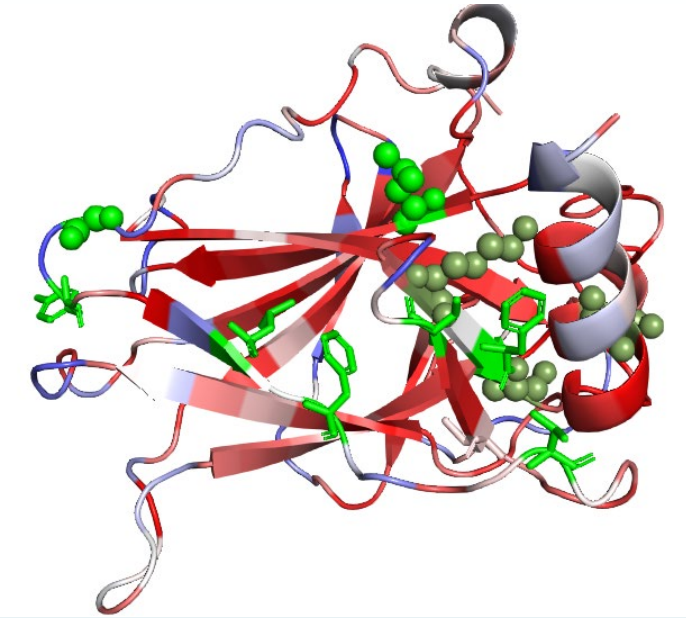
Option 2: Alternatively, you can get predictions based only on structural/dynamical features.

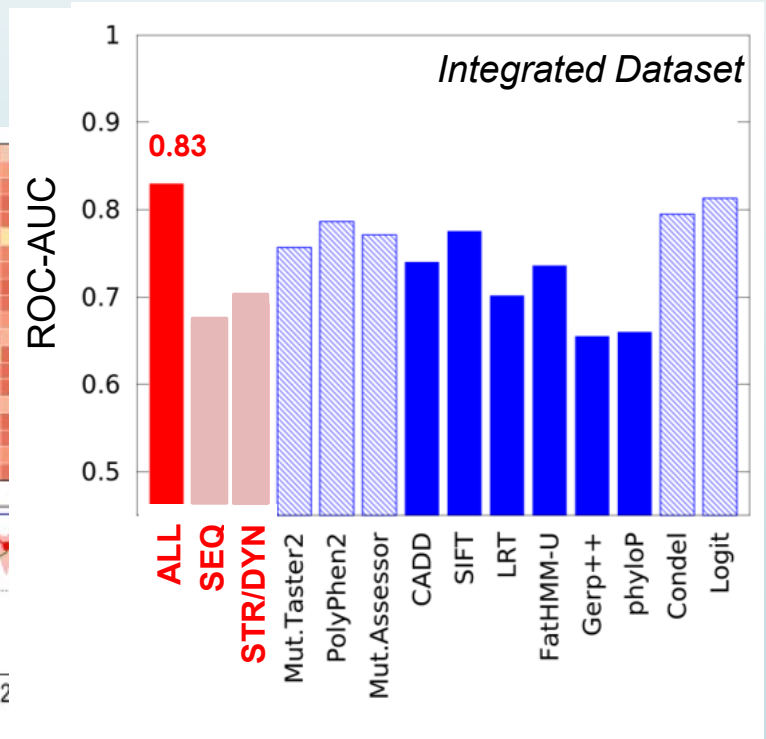
2.1: single query (e.g.: P17516 135 G E)

2.2: batch query Choose File No file chosen

email (optional):





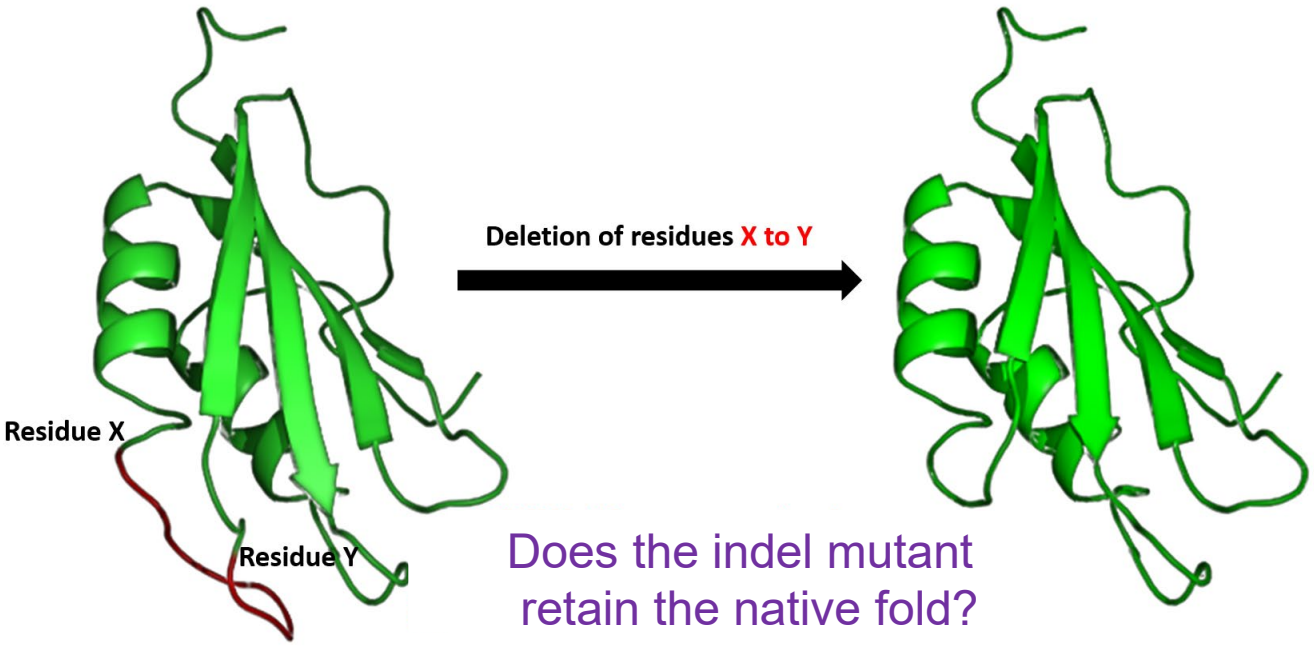


Ponzoni & Bahar, PNAS 2018

Ponzoni et al. *Bioinformatics* 2020

Effect of Amino Acid Deletions (AA-Del) on the Fold

Can the protein retain its original fold despite a deletion?



Does the indel mutant retain the native fold?

Challenges:

1. No existing dataset listing effect of stability on foldability
2. No computational framework to predict foldability

Anupam Banerjee, PhD

Banerjee & Bahar (2023) Structural Dynamics Predominantly Determine the Adaptability of Proteins to Amino Acid Deletions. *IJMS* **24**, 8450

Positive-Unlabeled Learning-based Random Forest Classifier

Database

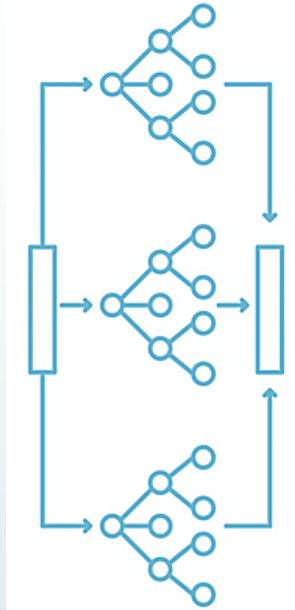
87 positive + 4,350 unlabeled mAA-dels in loops/coils

66 positive + 3,299 unlabeled mAA-dels in other regions

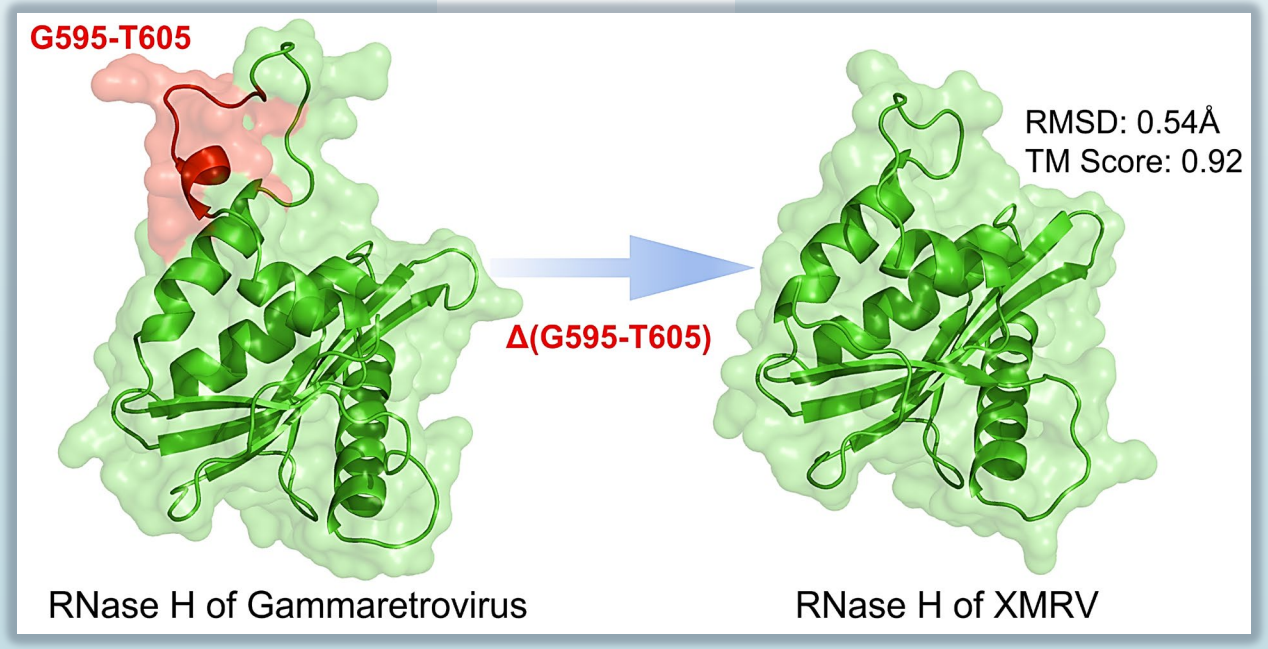
Up to m = 23 deleted amino acids

- Signaling effectiveness
- Sensitivity
- Mechanical Stiffness
- MSF in Global Modes
- MSF in Local Modes
- Essentiality (ESSA score)
- **PROFOUND features** *(41 features)*

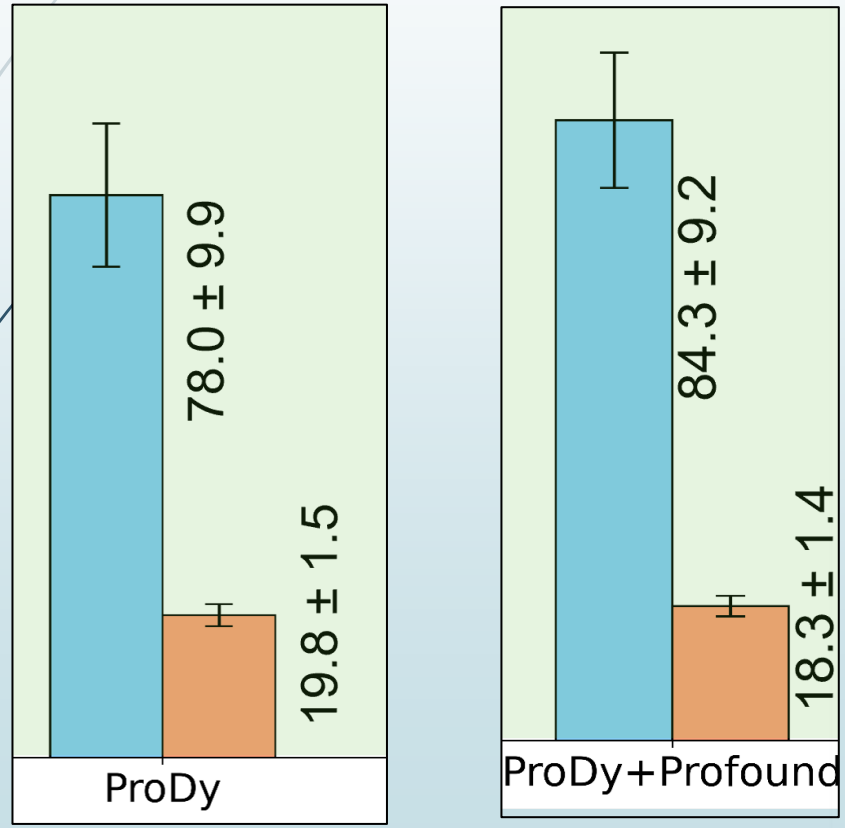
Encoding mAA-del



Fold stability - Classification



Recall rate of 84.3% is obtained for detecting TPs. Dynamics-only features yield 78.0%

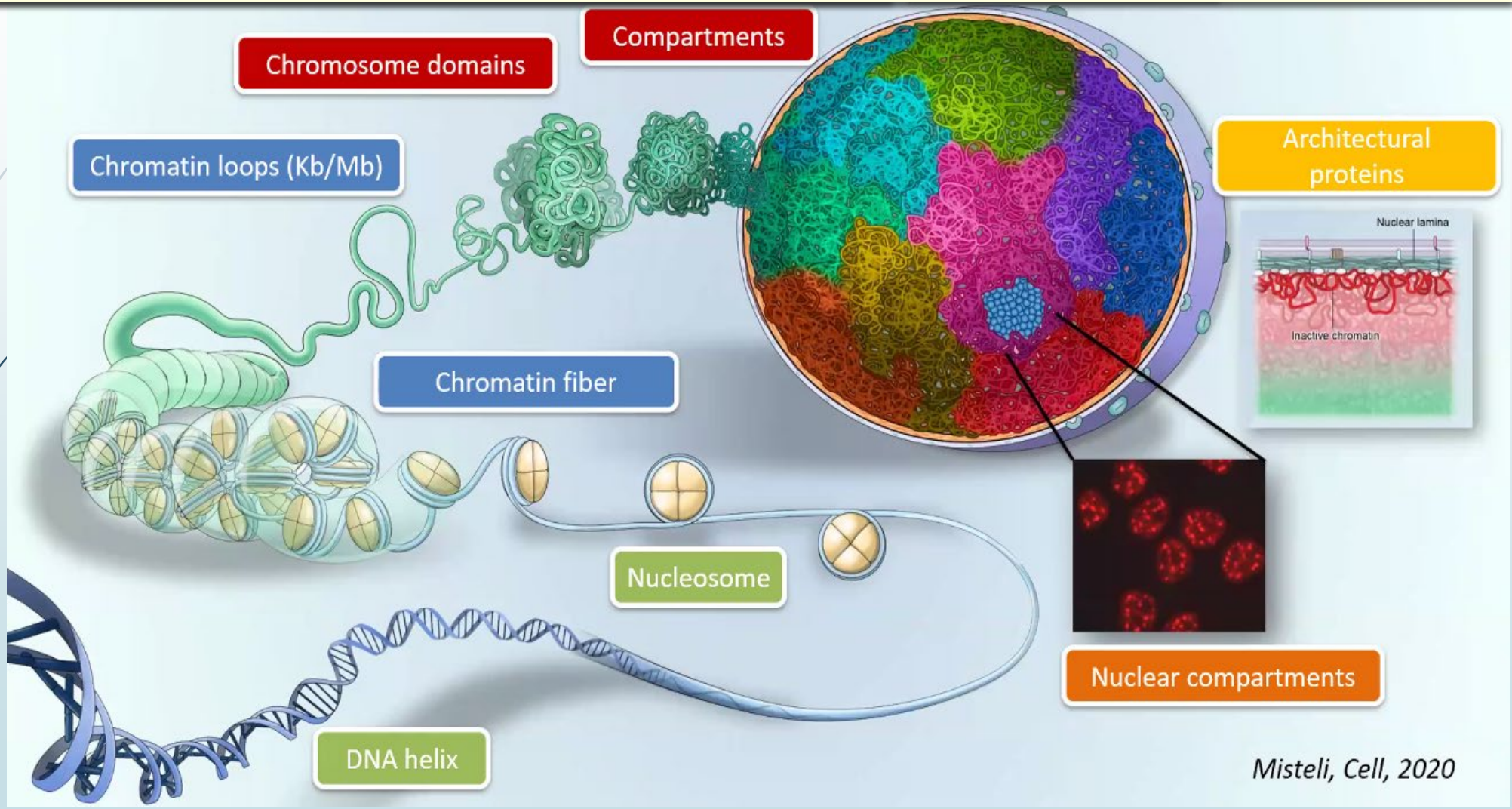


Intrinsic Dynamics is a major determinant of the fold adaptability to a deletion (of 2-23 residues).

Banerjee & Bahar (2023) *IJMS* **24**, 8450

27

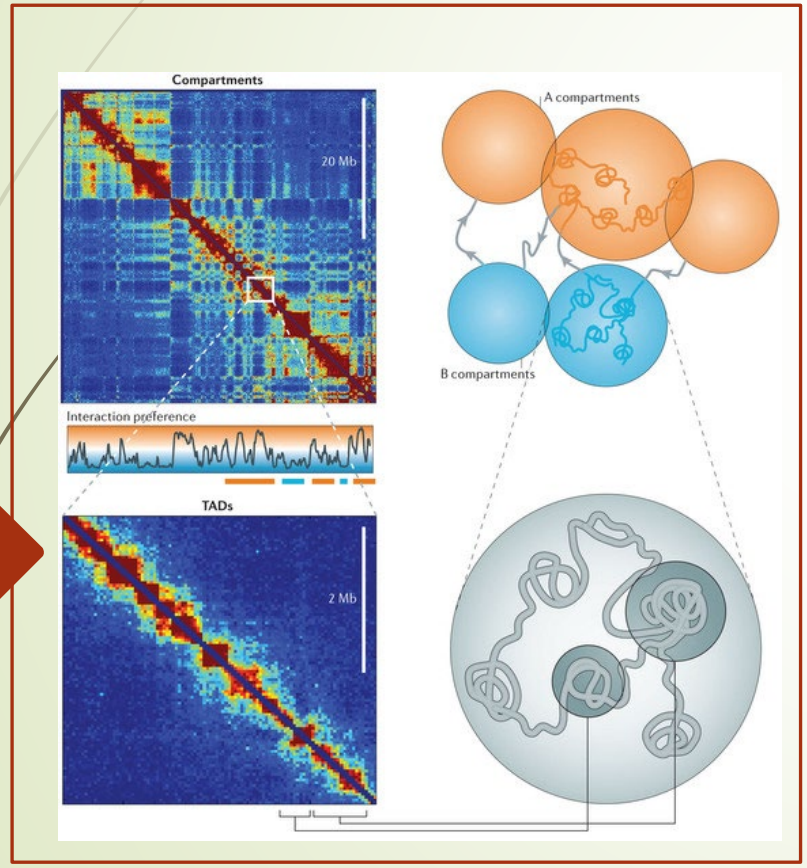
Or relate interloci topology to gene expression patterns?



Tom Misteli, *Cell* 183,28-45, 2020

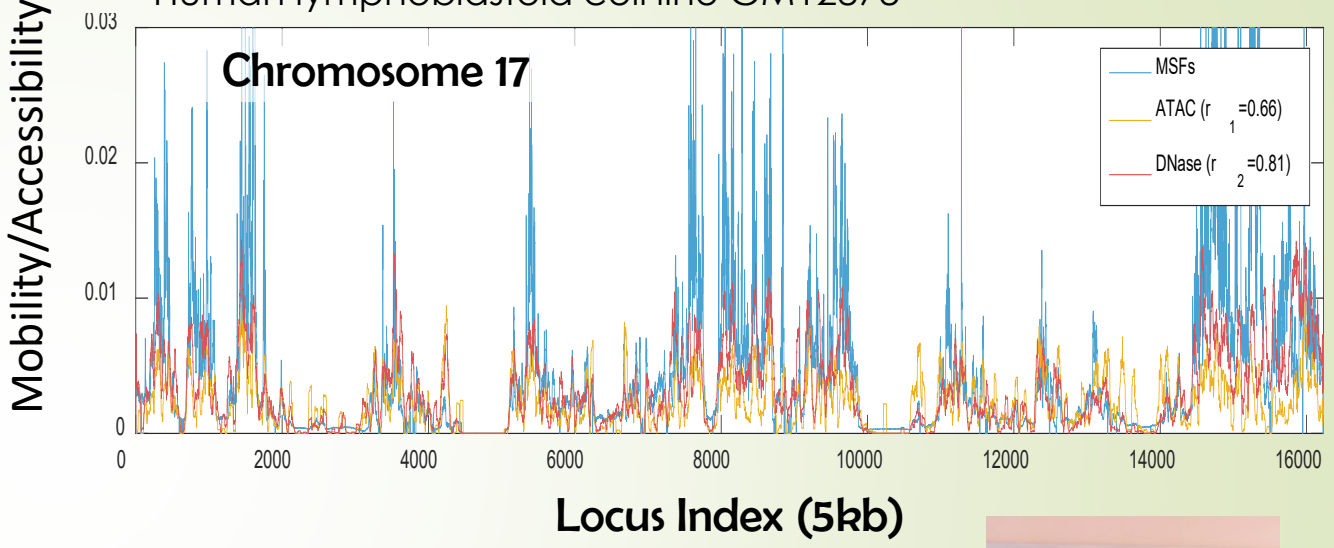
Chromatin accessibility and 3D spatial fluctuations

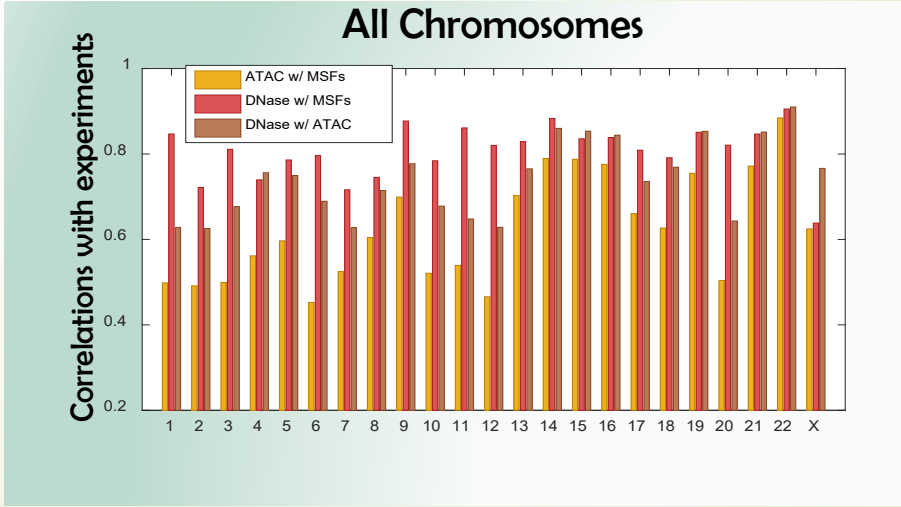
Using Hi-C data for gene loci contacts

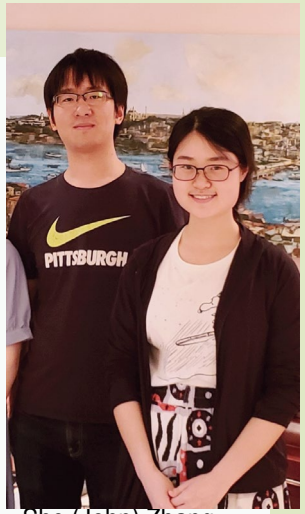


28

Human lymphoblastoid cell line GM12878





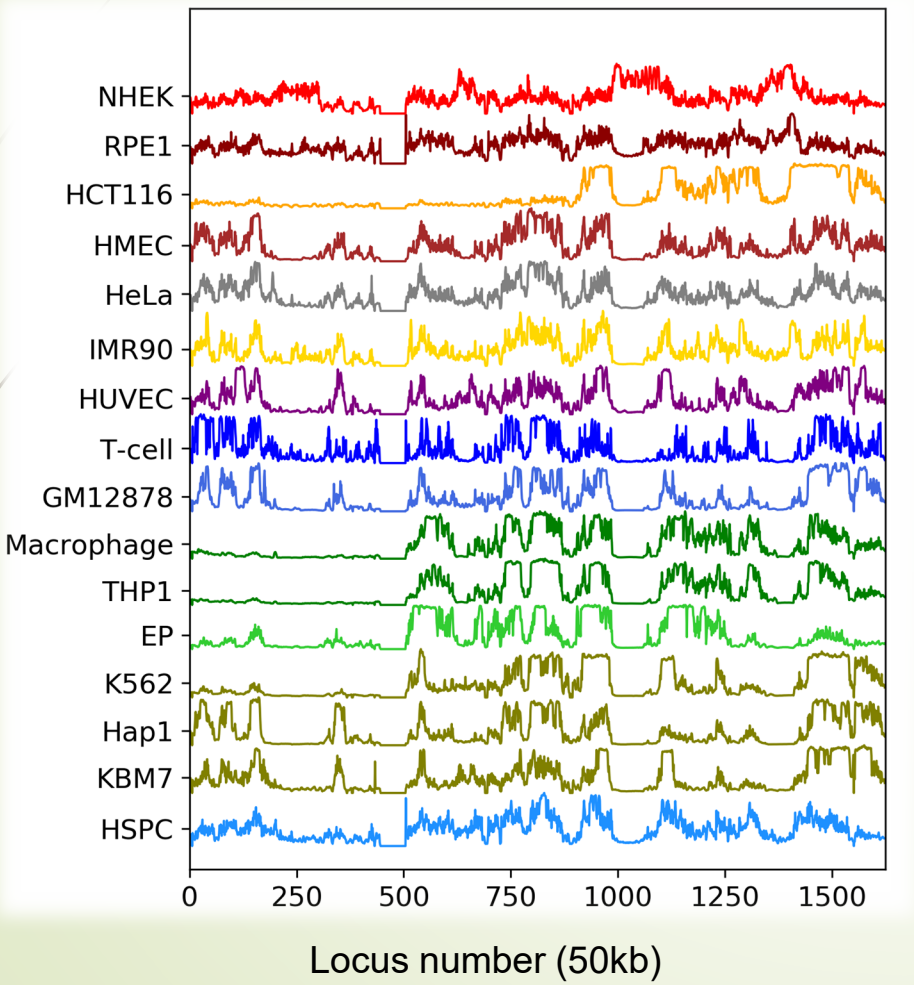


She (John) Zhang
Fangyuan (Chelsea) Chen

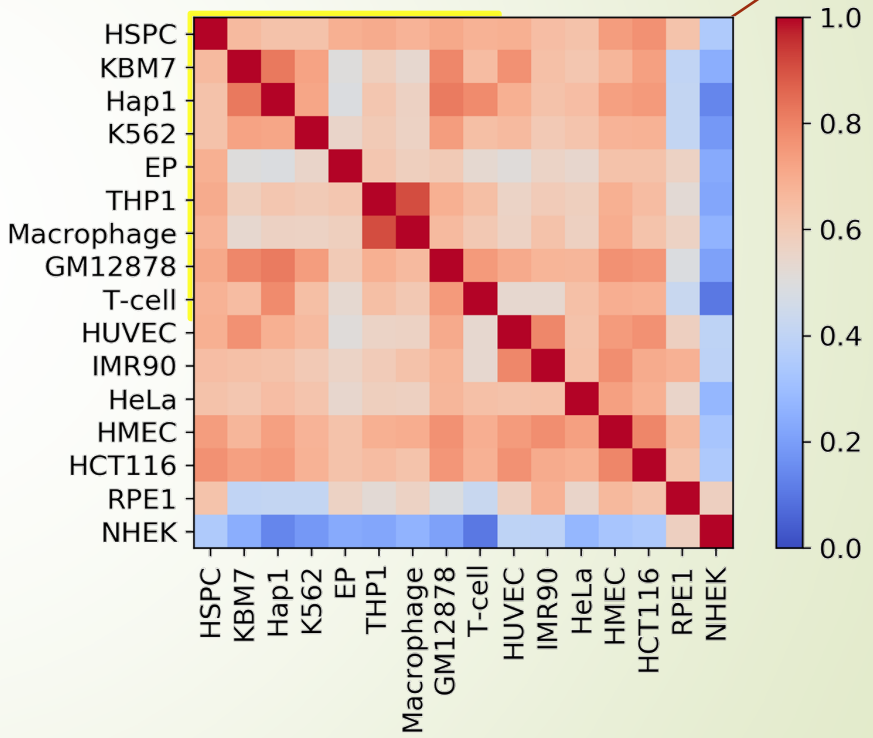
Sauerwald, Zhang, Kingsford & Bahar *Nucleic Acids Res* 2017

Analysis of chromosomal dynamics
for 16 different cell types

Chromosome 17



epidermal keratinocytes

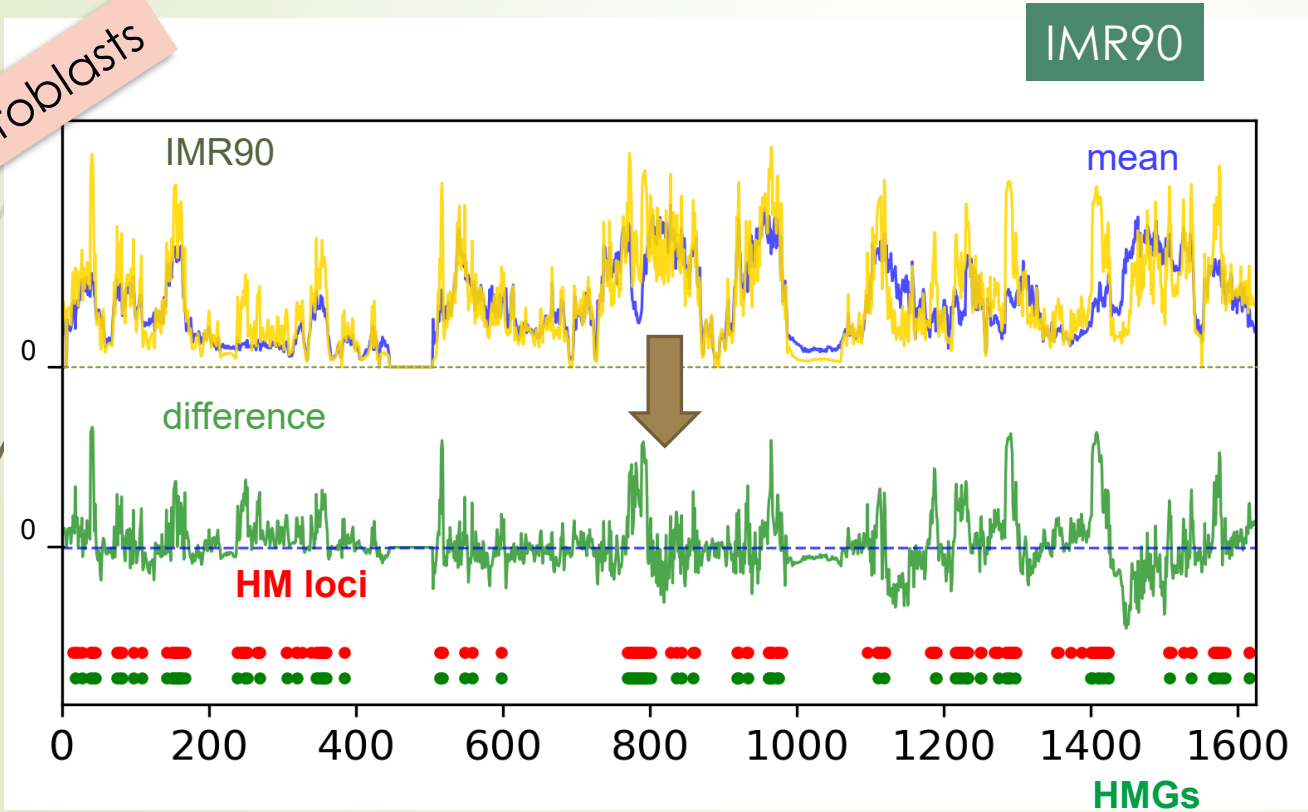


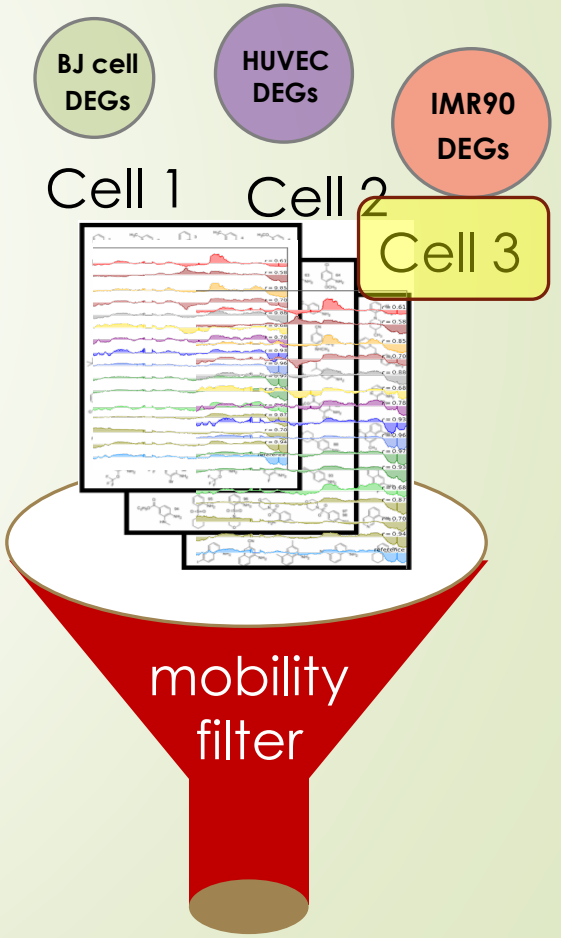
Zhang et al. (2020) *Nucleic Acids Res* 48, 1131-1145

Each cell type has a signature gene-mobility profile

Does high mobility correlate with up-regulation, or high expression? Yes!

Lung fibroblasts

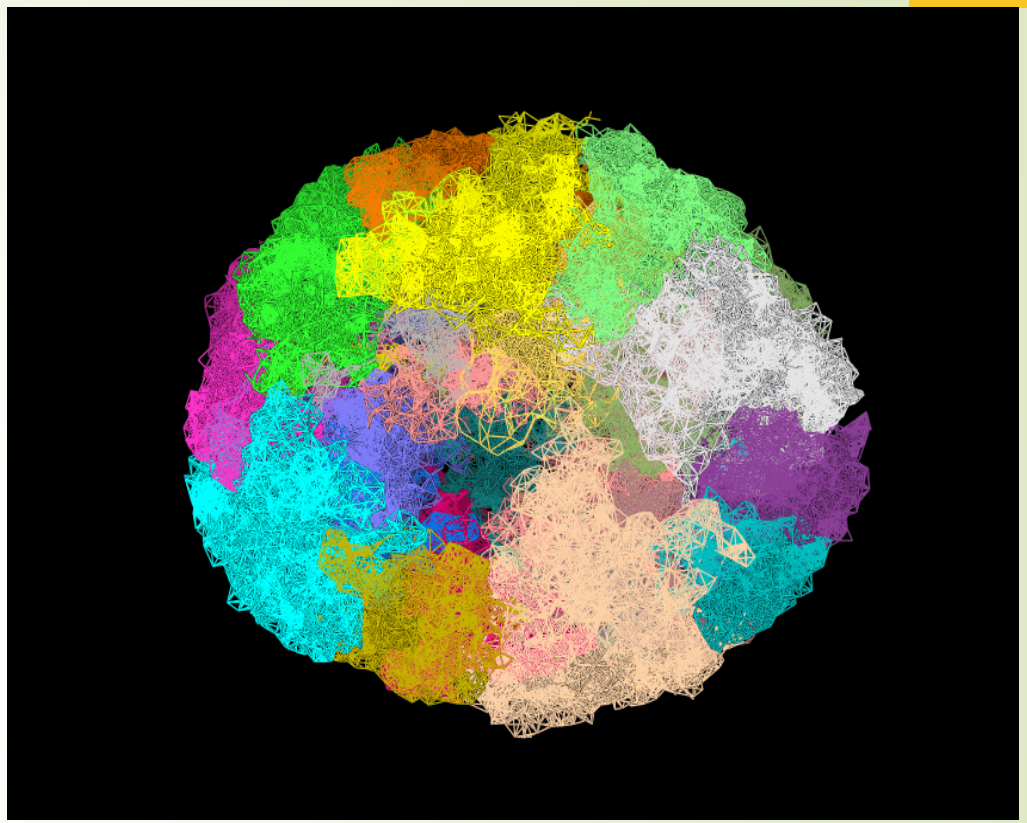




31

Concluding remarks

- First steps toward **4D Genome**
- Good agreement with DNase-seq, ATAC-seq and ChIA-PET data
- Differences in gene mobilities correlate with differences in gene expression patterns: dynamics is a determinant of cell expression signature



Softest mode of motion predicted by ANM analysis of single-cell Hi-C data (for mouse embryonic stem cells)*

*Stevens, …, Laue (2017) 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*, **544**, 59-64

REMARKS

Merger of physics-based with ML tools will help bridge microscopic & macroscopic behaviors

Mathematical accuracy (exact solution) of physics-based approaches complements the statistical accuracy of AI tools.

Banerjee et al. (2023) *Curr Opin Struct Biol* 78:102517

Two major groups of computations:

Systems Pharmacology

Genome-Scale Modeling

Both groups use

- theory and methods of computer science, machine learning, and mathematics,

AND

- models and methods of physical sciences and engineering.

Lab members

. Anupam Banerjee
. Satyaki Saha
. Carlos Ventura
. Hoang Nguyen
· Mary H Cheng
· Jiyoung Lee
·Antony Bogetti
·Xiaowei Bogetti
. Lee-Wei Yang
· Hongchun Li
· Frank H Zhang
· She Zhang
· Yan Zhang
· Luca Ponzoni
· K Mikulska-Ruminski
· James Krieger
· Murat C Cobanoglu
· Fen Pei
· Burak Kaynak
· Ahmet Bakan
· Anindita Dutta
· Lidio Meireles
· Ying Liu
· Indira Shrivastava

Collaborators

- Susan Amara (NIMH)
- Moshe Arditi (Cedars-Sinai/UCLA)
- Ali Rana & Canan Atilgan (Sabanci U)
- Tom Bartol (Salk)
- Hulya Bayir (Pitt)
- Jeff Brodsky (Pitt)
- Pemra Doruker (Bogazici/Pitt)
- James Faeder (Pitt)
- Aurelio Galli (Alabama U)
- Jenny Garcia-Olivares (NIMH)
- Joel Greenberger (UPMC)
- Ingo Greger (MRC, UK)
- Angela Gronenborn (Pitt)
- Turkan Haliloglu (Bogazici U)
- Amnon Horovitz (Weizmann)
- Valerian Kagan (Pitt)
- Carl Kingsford (Carnegie Mellon)
- David Perlmutter (Wash U)
- Gideon Schreiber (Weizmann)
- Terry Sejnowski (Salk)
- Gary Silverman (Wash U)
- Sasha Sorkin (Pitt)
- Lans Taylor (Pitt, UPDDI)
- Delany Torres-Salazar (NIMH)
- Andy VanDeMark (Pitt)
- Yong Wan (Emory)
- Sally Wenzel (Pitt, GSPH)
- David Whitcomb (UPMC)
- Eric Xing (CMU)

Thank you!

Support from NIGMS, NIDDK, NHGRI & HFSP