Bioactivity profiles for uncharacterised compounds

Computational Applications in Secondary Metabolites Discovery (CAiSMD), March 9 (2021)

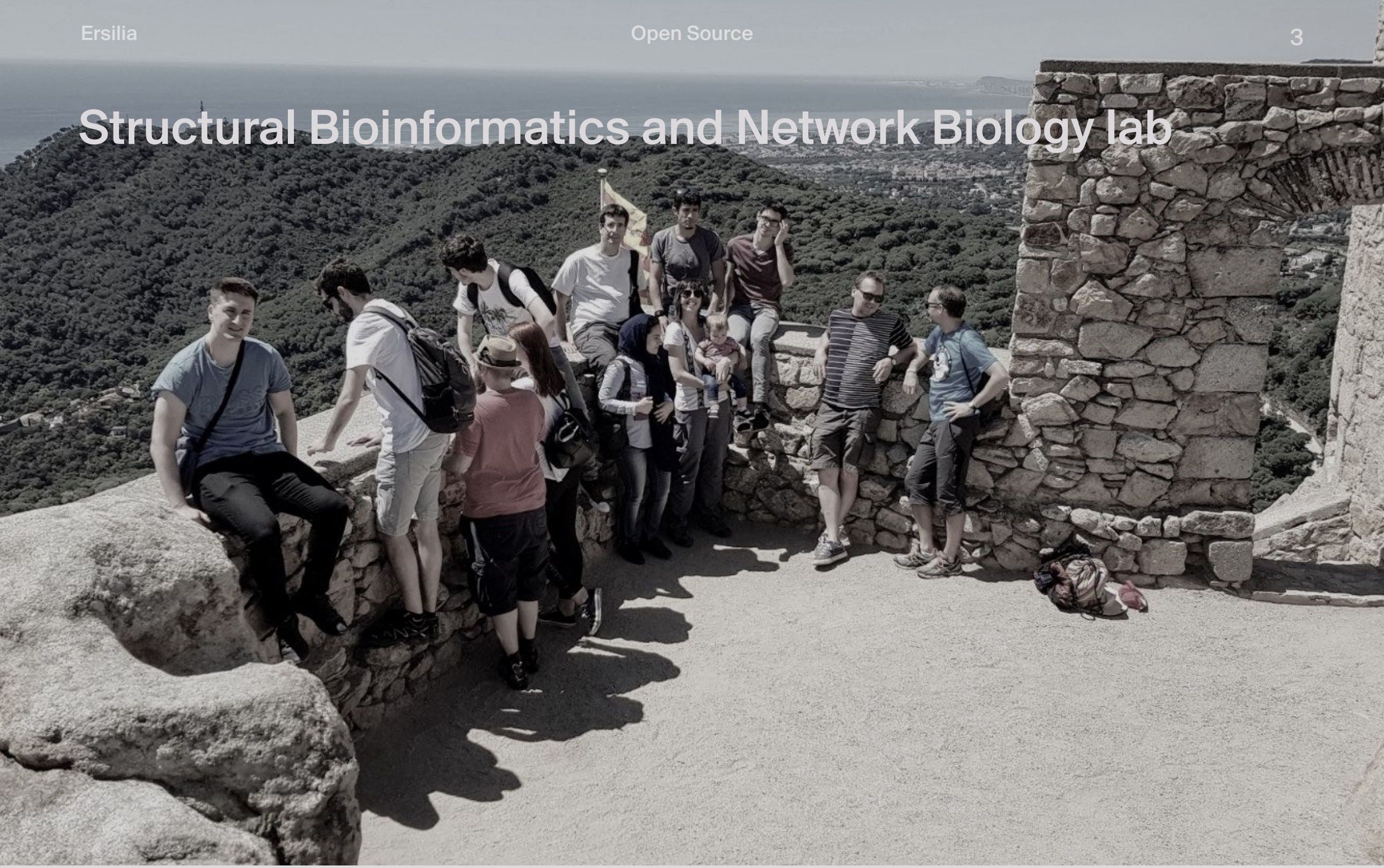
Miquel Duran-Frigola, PhD
@mduranfrigola
miquel@ersilia.io
ersilia.io

Ersilia

The Chemical Checker, Ersilia's backbone technology

Duran-Frigola et al. Nature Biotechnology, 2020

Structural Bioinformatics and Network Biology lab



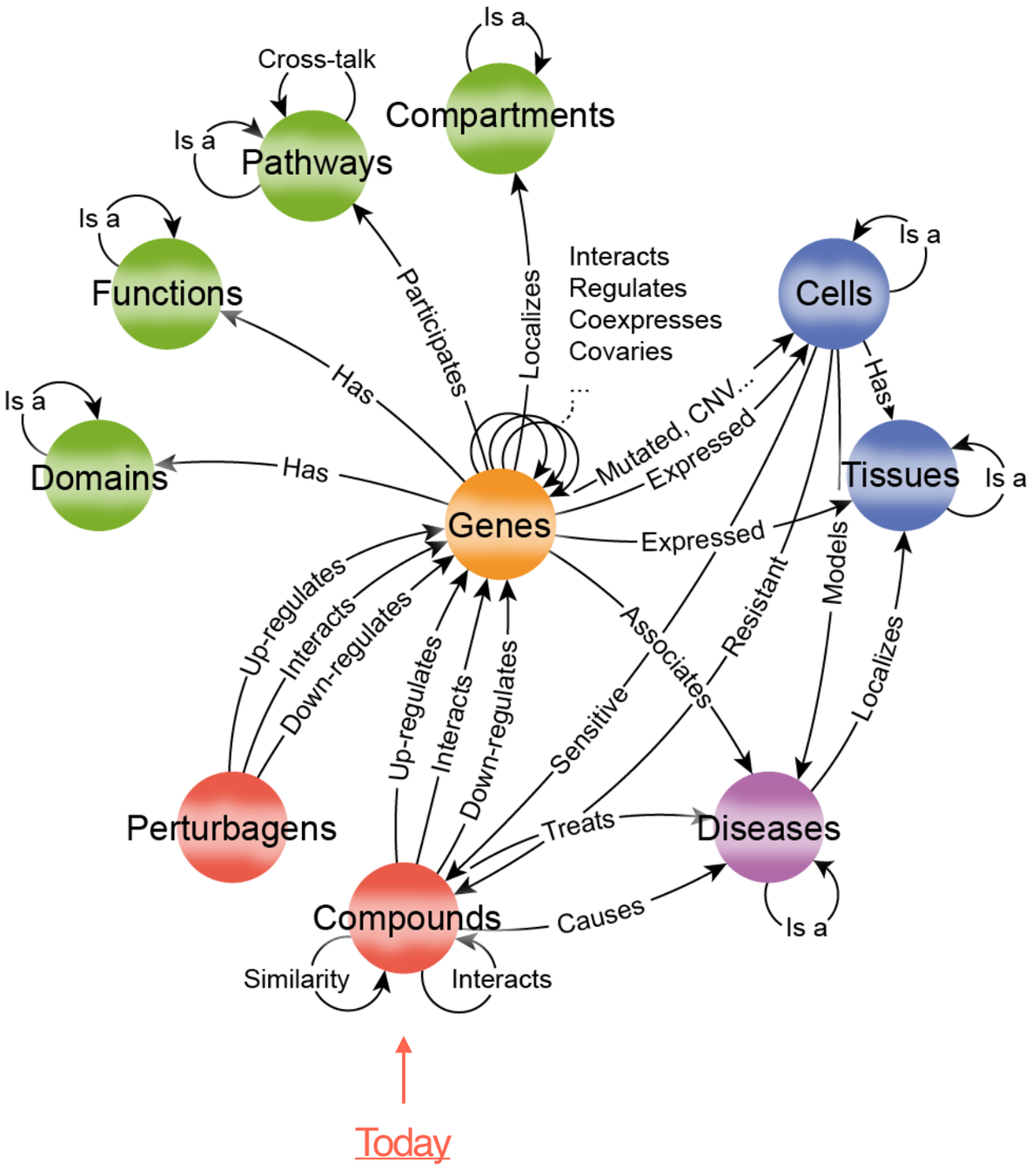
Chemical Checker

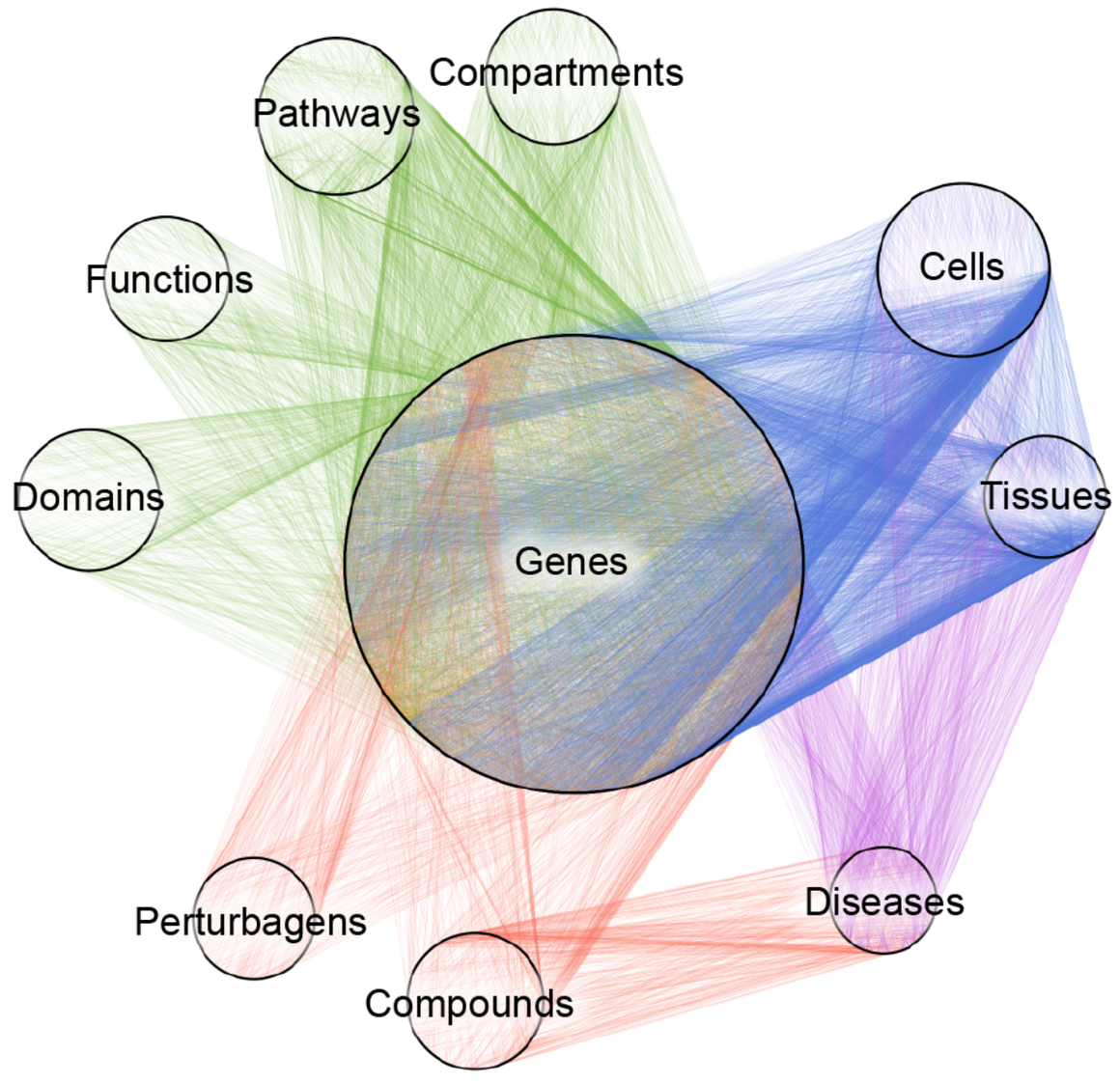
IRB BARCELONA

Structural Bioinformatics & Network Biology Dr. Patrick Aloy

— Prof. Patrick Aloy
— Dr. Eduardo Pauls
— Dr. Martino Bertoni
— Pau Badia-i-Mompel
— Oriol Guitart
— Víctor Alcalde

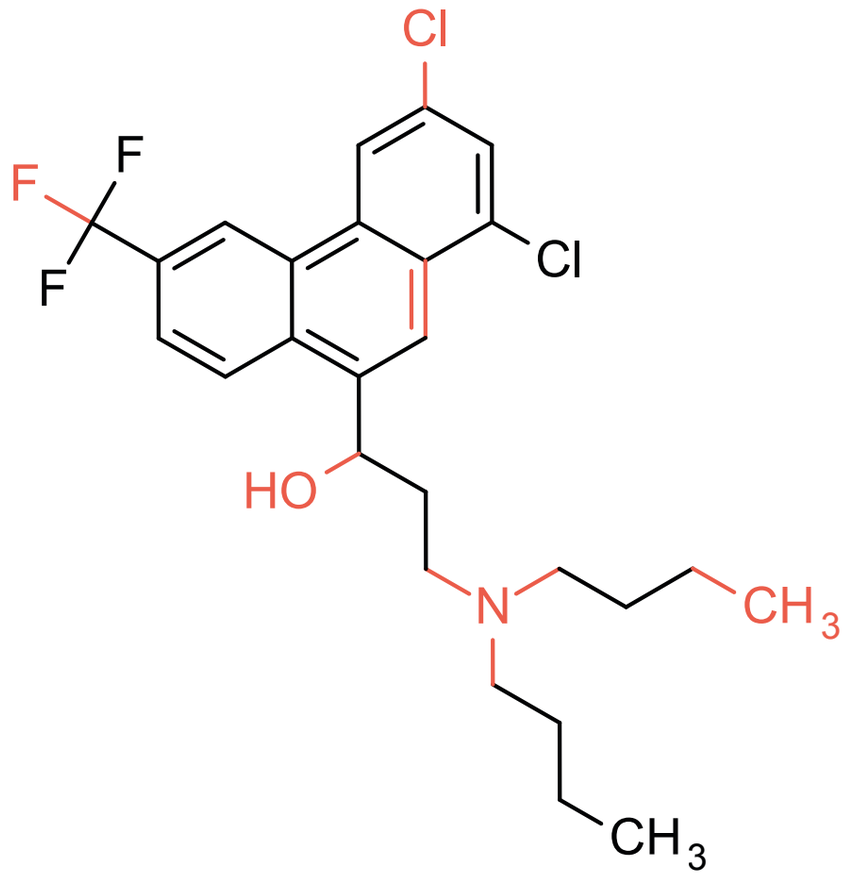
Structural Bioinformatics and Network Biology lab

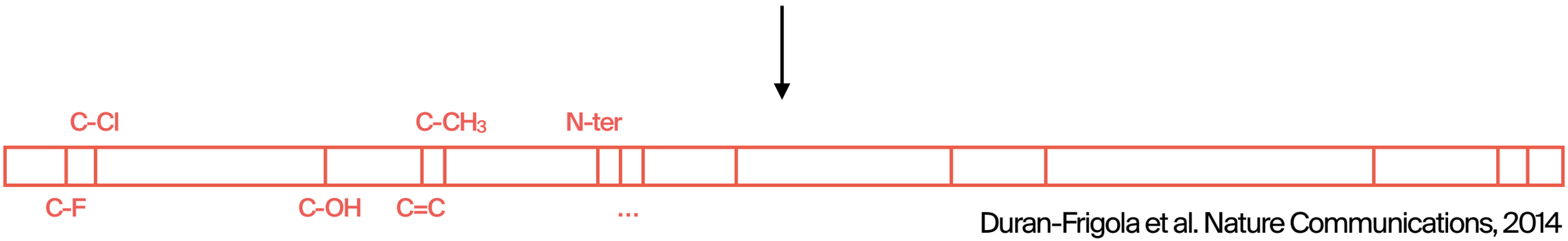




Duran-Frigola et al. WIREs Advanced Reviews, 2019
Fernandez-Torras et al, in preparation

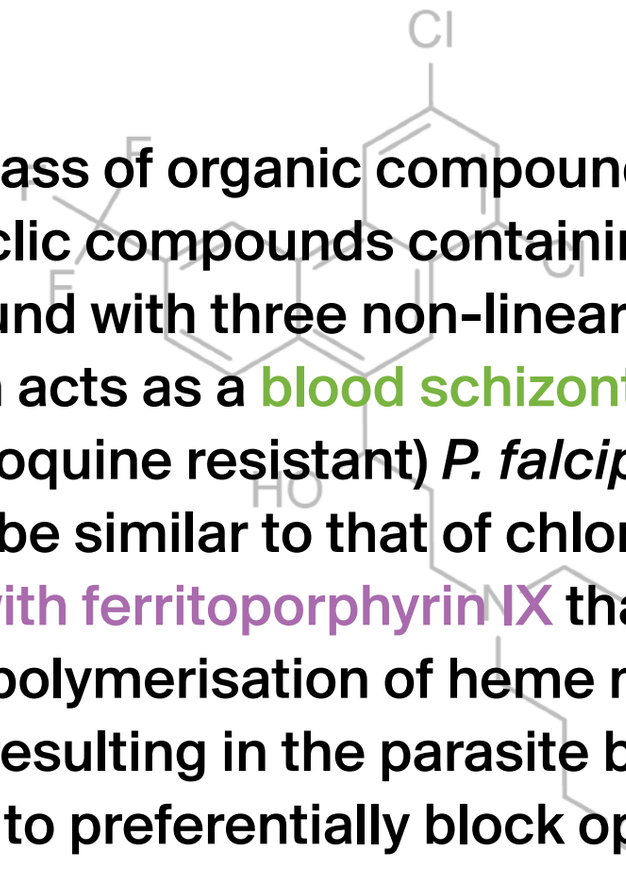
Classical chemical fingerprints



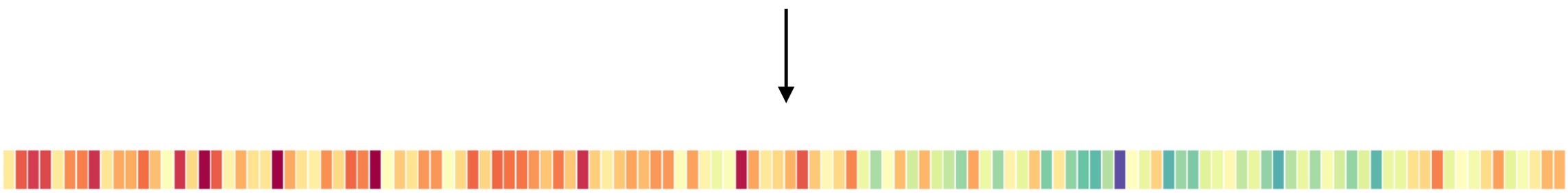


Duran-Frigola et al. Nature Communications, 2014

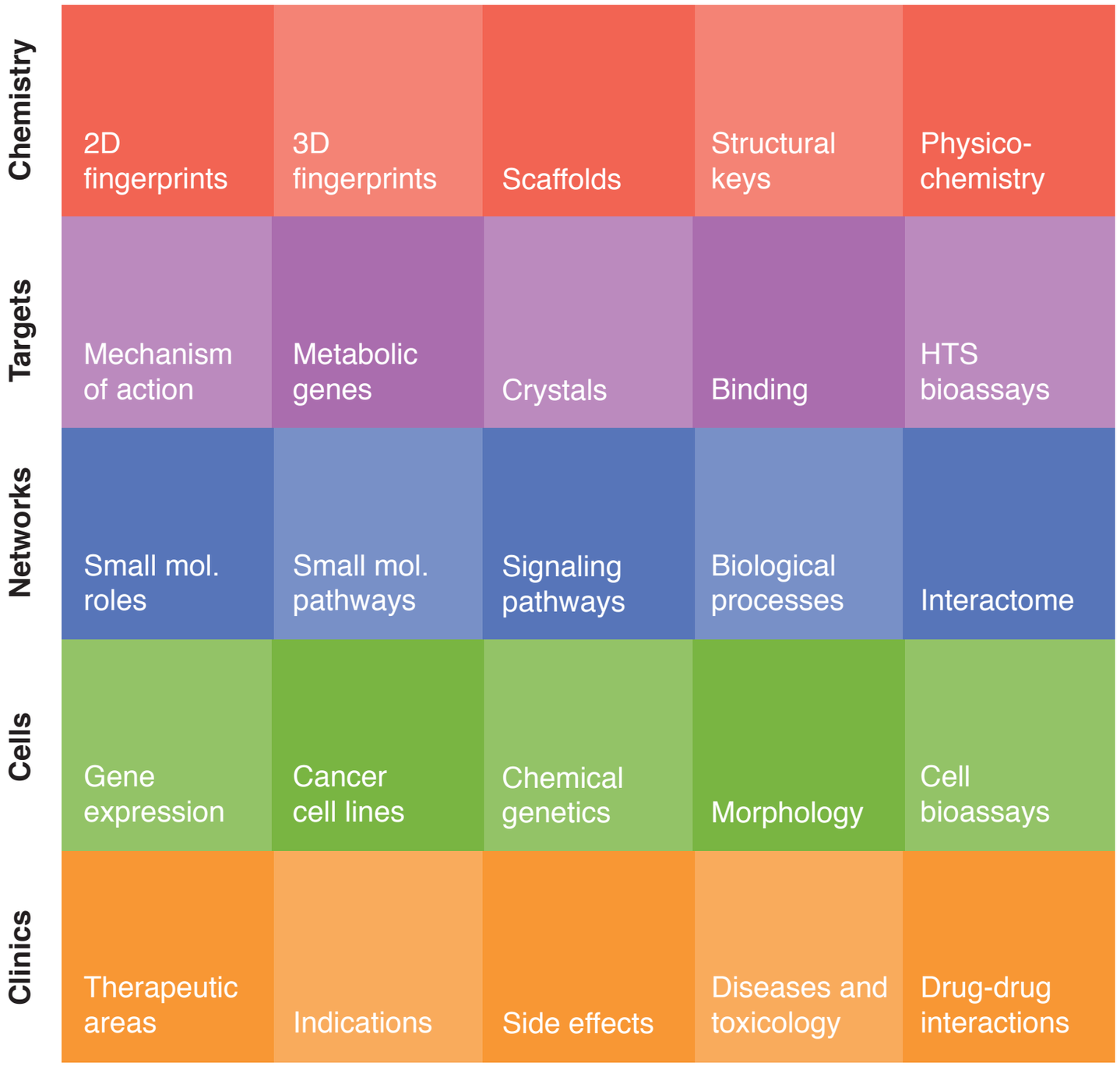
Chemical Checker signature



Halofantrine belongs to the class of organic compounds known as phenanthrenes and derivatives. These are polycyclic compounds containing a phenanthrene moiety, which is a tricyclic aromatic compound with three non-linearly fused benzene. Halofantrine is a synthetic antimalarial which acts as a blood schizonticide. It is effective against multi drug resistant (including mefloquine resistant) *P. falciparum* malaria. The mechanism of action of Halofantrine may be similar to that of chloroquine, quinine, and mefloquine; by forming toxic complexes with ferritoporphyrin IX that damage the membrane of the parasite. It appears to inhibit polymerisation of heme molecules (by the parasite enzyme 'heme polymerase'), resulting in the parasite being poisoned by its own waste. Halofantrine has been shown to preferentially block open and inactivated HERG channels leading to some degree of cardiotoxicity. Side effects include coughing noisy, rattling, troubled breathing, loss of appetite, aches and pain in joints, indigestion, and skin itching or rash, *et cetera, et cetera.*



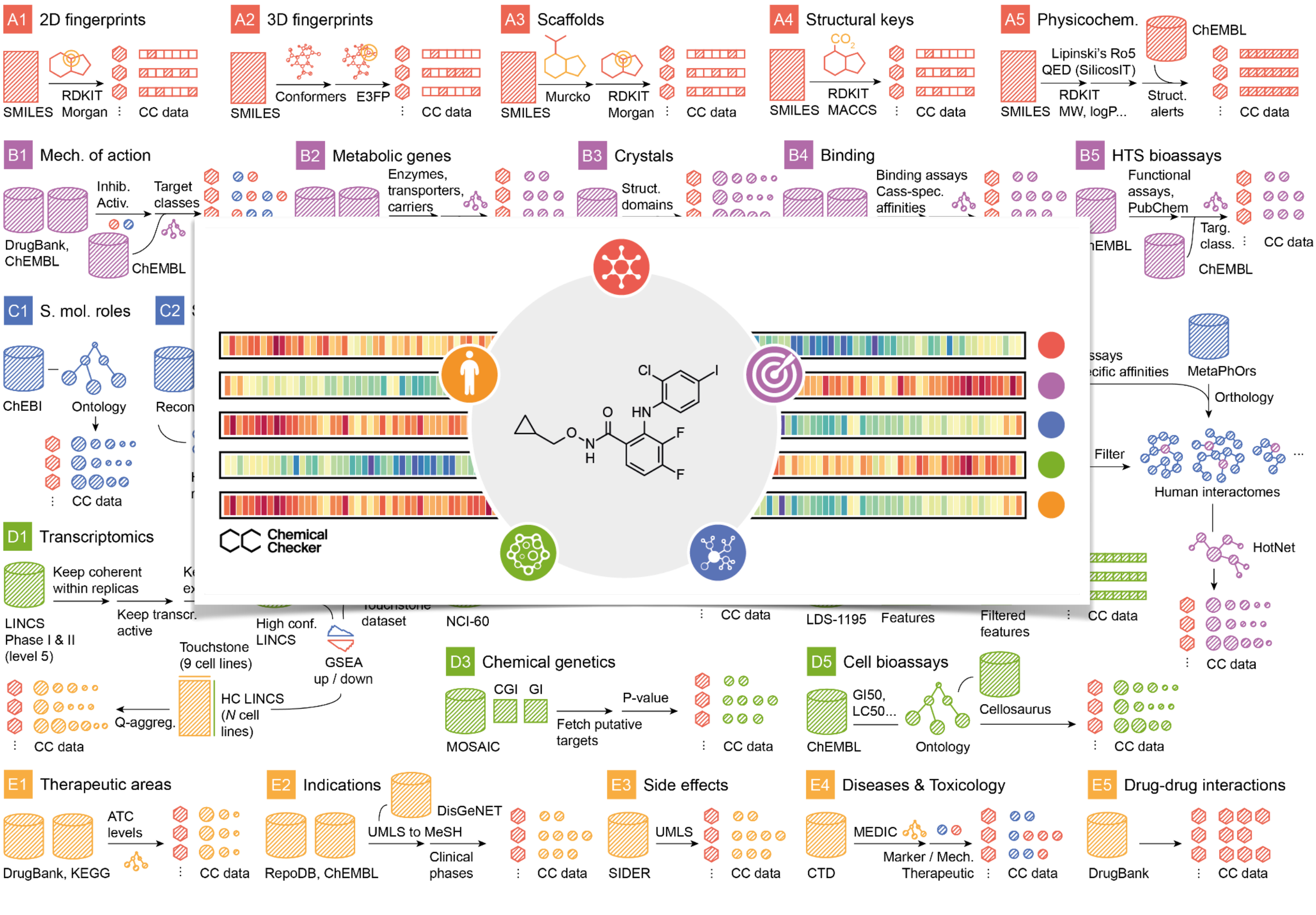
The bioactive chemical space, organised



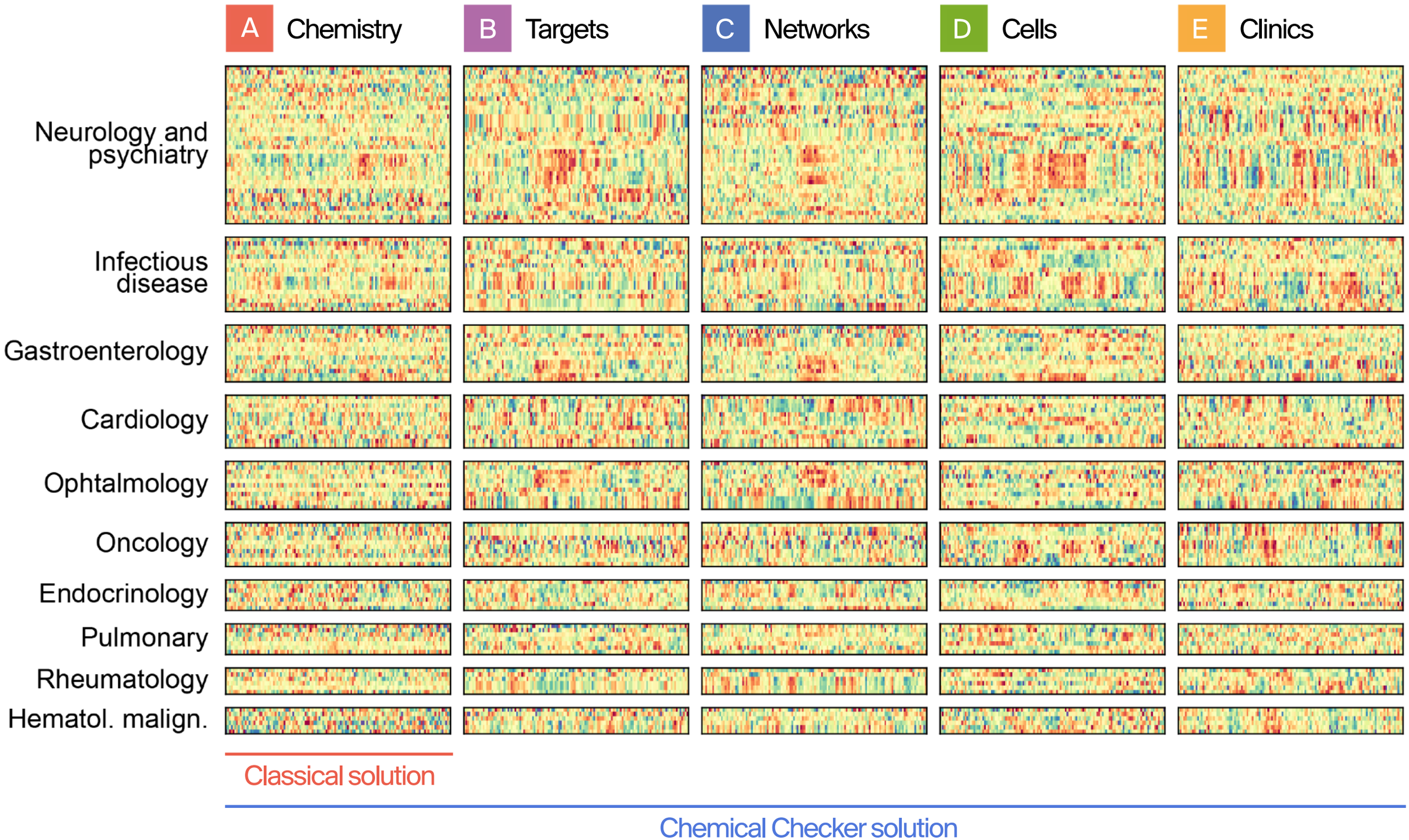
- 800k bioactive molecules
- 25 data types, from chemistry to the clinics
- The major small molecule databases are integrated
- Inference by deep learning
- chemicalchecker.org
- bioactivitysignatures.org

Chemical Checker

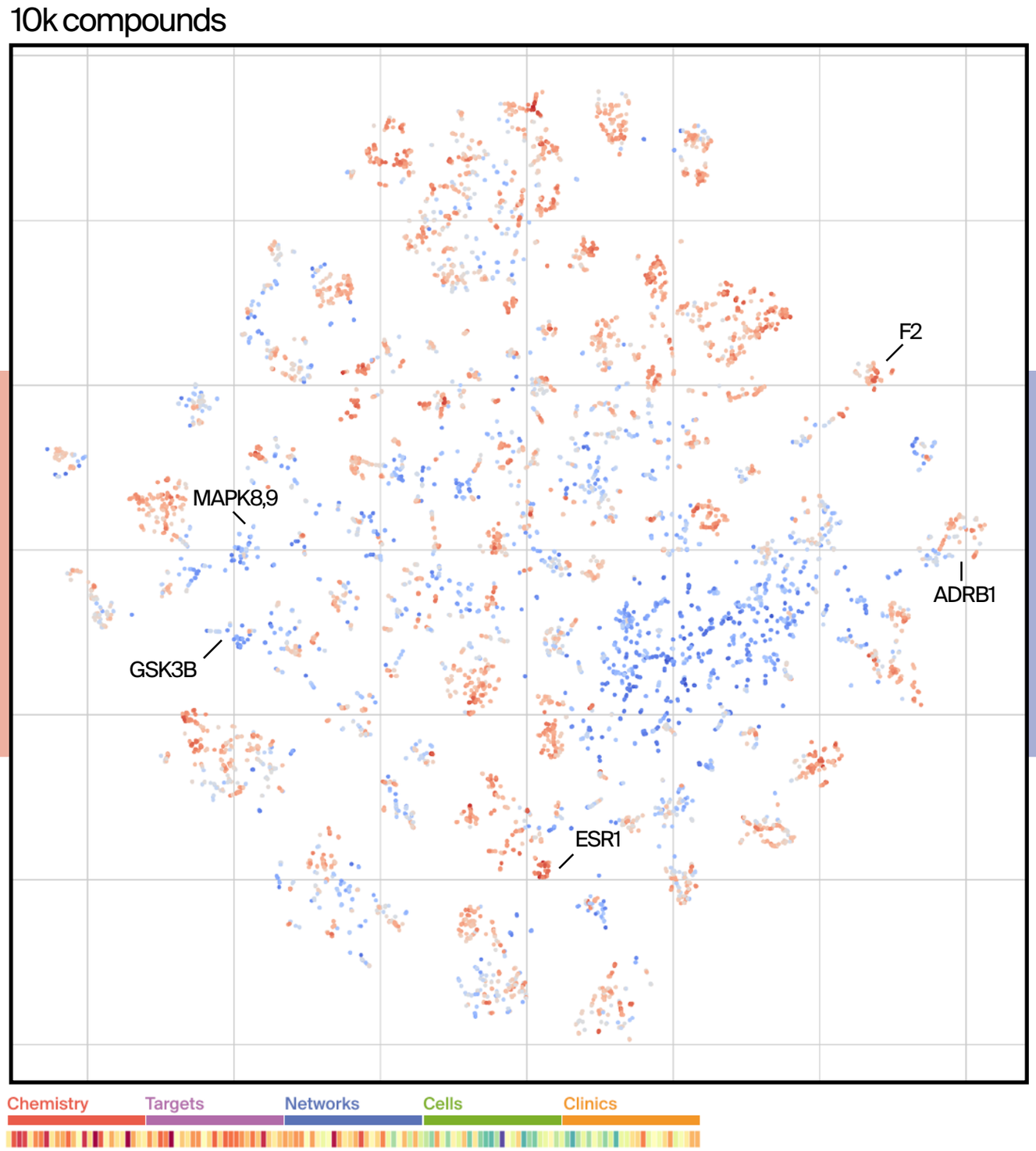
Duran-Frigola et al. Nat Biotech, 2020
Bertoni*, Duran-Frigola* et al. BioRXiV, 2020
🏆 CTD2-pancancer DREAM challenge, 2020


A1 2D fingerprints
SMILES
RDKIT Morgan
CC data
A2 3D fingerprints
SMILES
Conformers
E3FP
CC data
A3 Scaffolds
SMILES
Murcko
RDKIT Morgan
CC data
A4 Structural keys
SMILES
RDKIT MACCS
CC data
A5 Physicochem.
SMILES
Lipinski's Ro5
QED (SilicosIT)
RDKIT MW, logP...
ChEMBL
Struct. alerts
CC data
B1 Mech. of action
DrugBank, ChEMBL
Inhib. Activ.
Target classes
ChEMBL
B2 Metabolic genes
Enzymes, transporters, carriers
B3 Crystals
Struct. domains
B4 Binding
Binding assays Cass-spec. affinities
B5 HTS bioassays
Functional assays, PubChem
ChEMBL
Targ. class.
CC data
C1 S. mol. roles
ChEBI
Ontology
CC data
C2
Recon
Chemical Checker
MetaPhOrs
Orthology
Filter
Human interactomes
HotNet
CC data
D1 Transcriptomics
LINCS Phase I & II (level 5)
Keep coherent within replicas
Keep transcr. active
Touchstone (9 cell lines)
High conf. LINCS
GSEA up / down
HC LINCS (N cell lines)
Q-aggreg.
CC data
NCI-60
CC data
LDS-1195
Features
Filtered features
CC data
D3 Chemical genetics
MOSAIC
CGI
GI
P-value
Fetch putative targets
CC data
D5 Cell bioassays
ChEMBL
GI50, LC50...
Ontology
Cellosaurus
CC data
E1 Therapeutic areas
DrugBank, KEGG
ATC levels
CC data
E2 Indications
RepoDB, ChEMBL
UMLS to MeSH
DisGeNET
Clinical phases
CC data
E3 Side effects
SIDER
UMLS
CC data
E4 Diseases & Toxicology
CTD
MEDIC
Marker / Mech.
Therapeutic
CC data
E5 Drug-drug interactions
DrugBank
CC data

A tiny chunk of the database



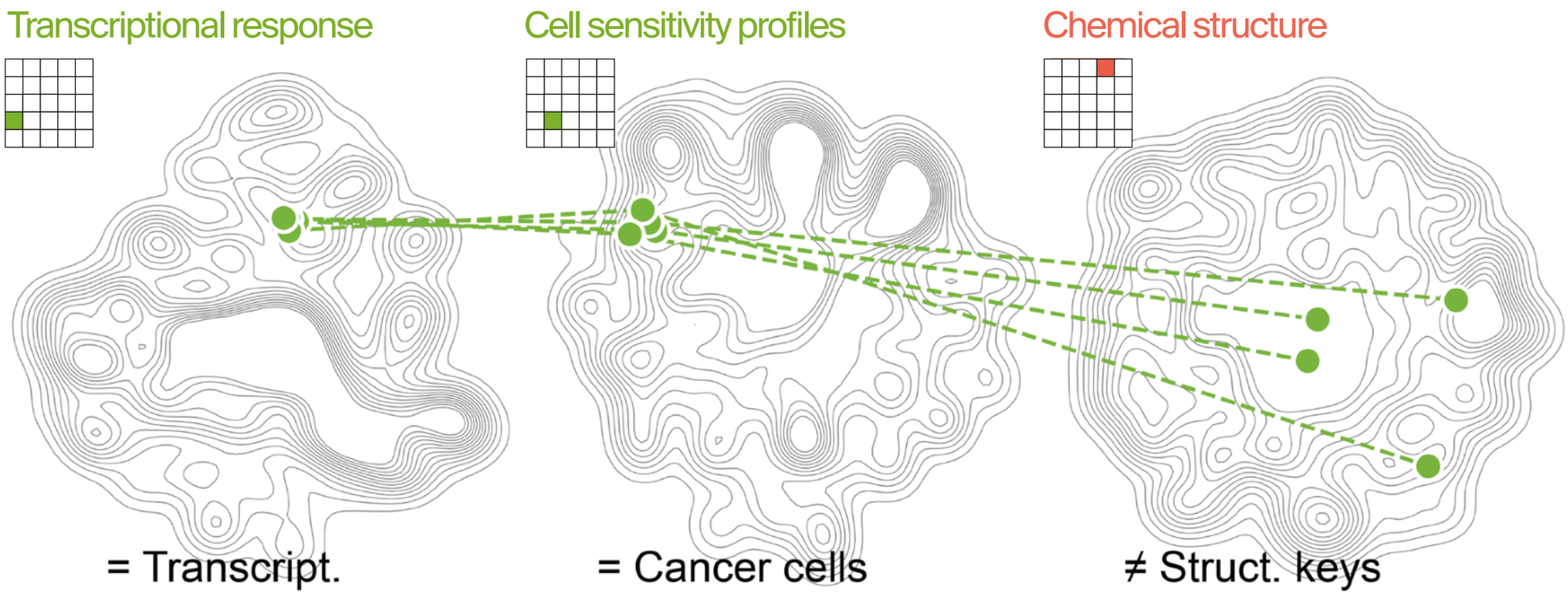
Beyond chemical similarity

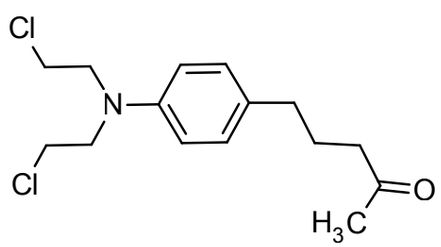


Chemically similar

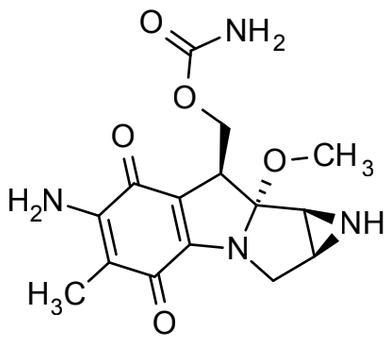
Chemically diverse

Complex similarity searches

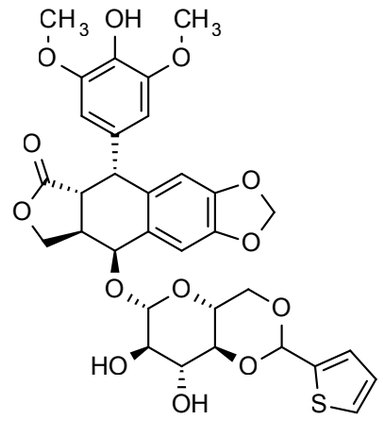




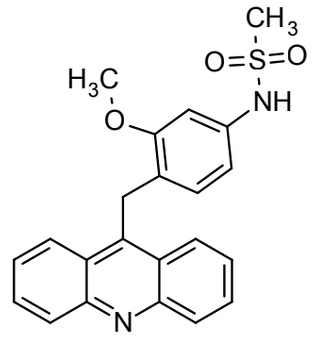
Chlorambucil



Mitomycin

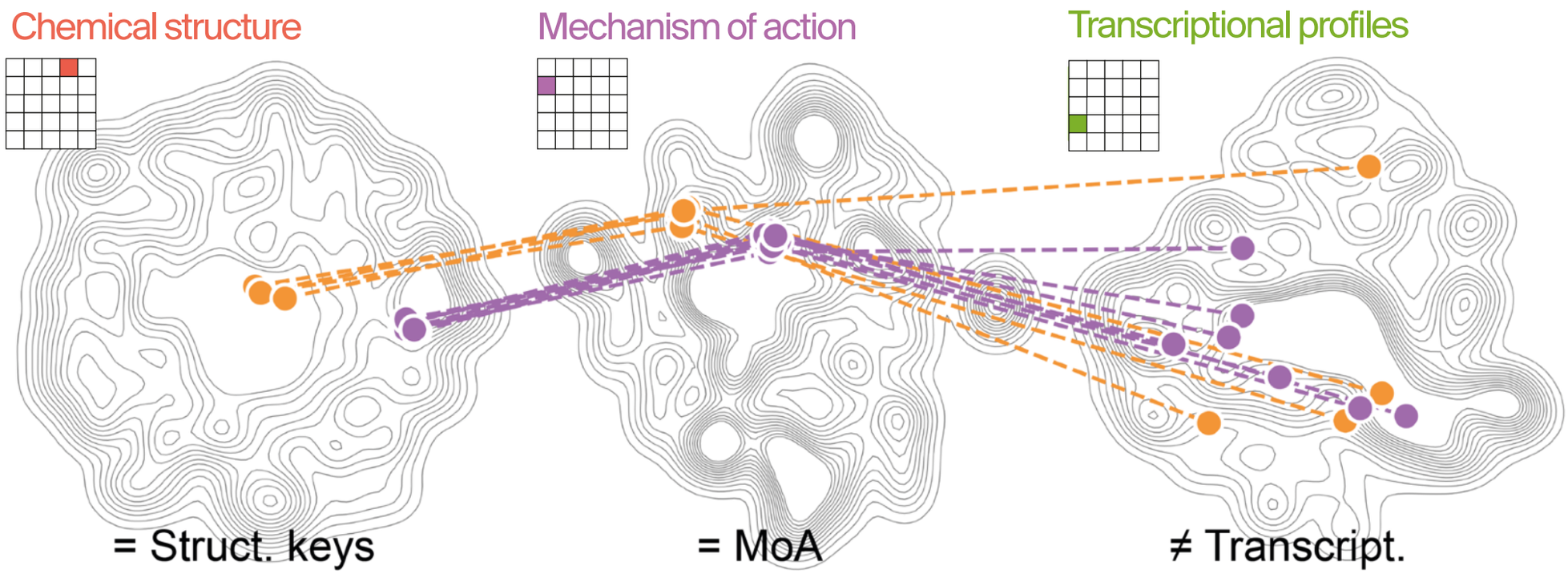


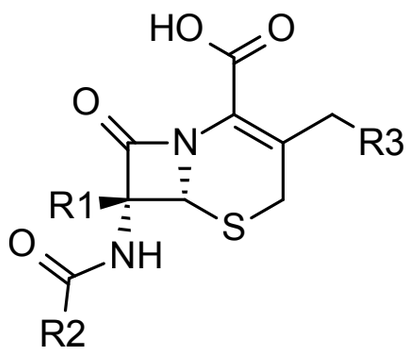
Teniposide

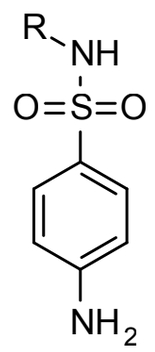


New molecule

Complex similarity searches

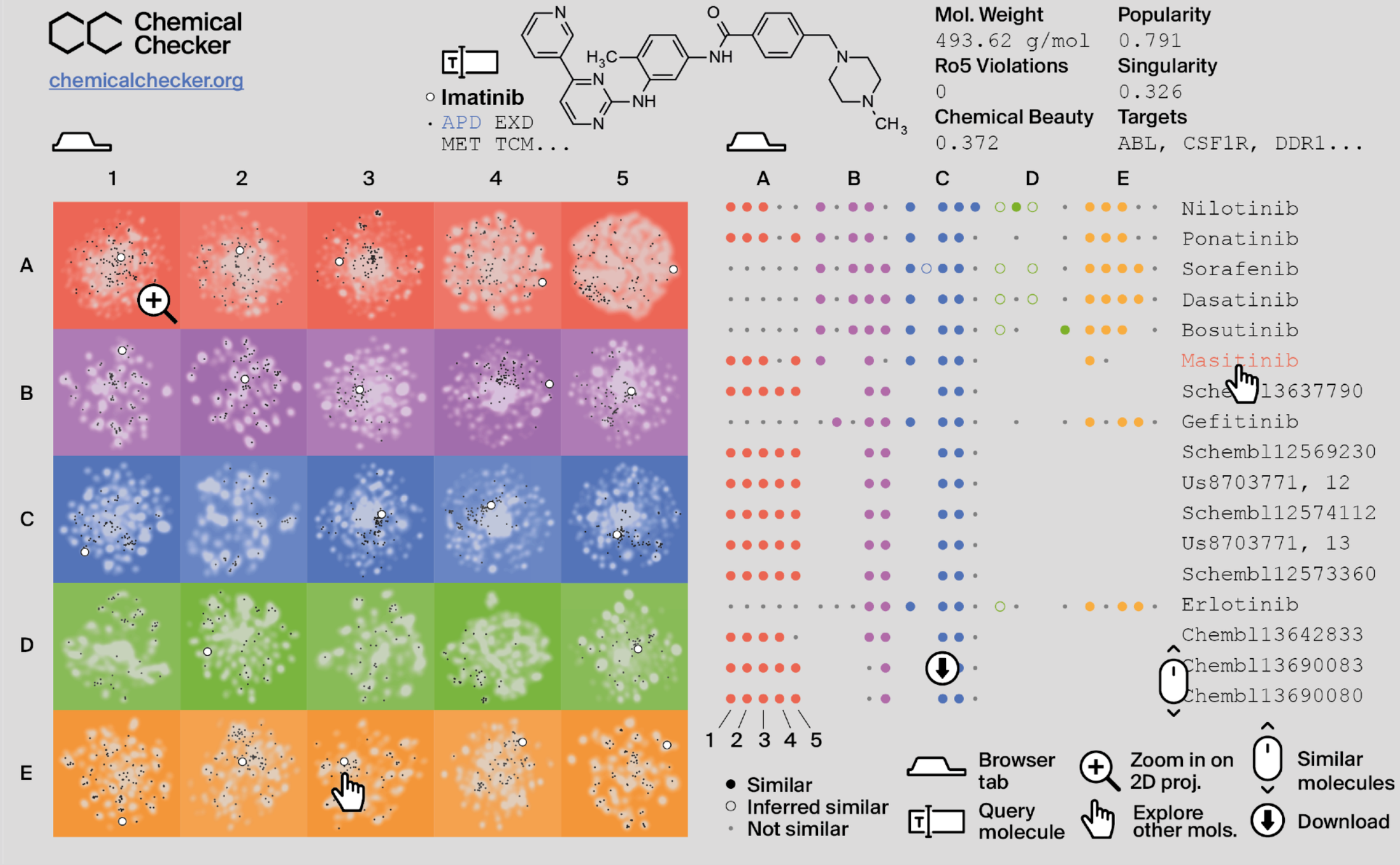






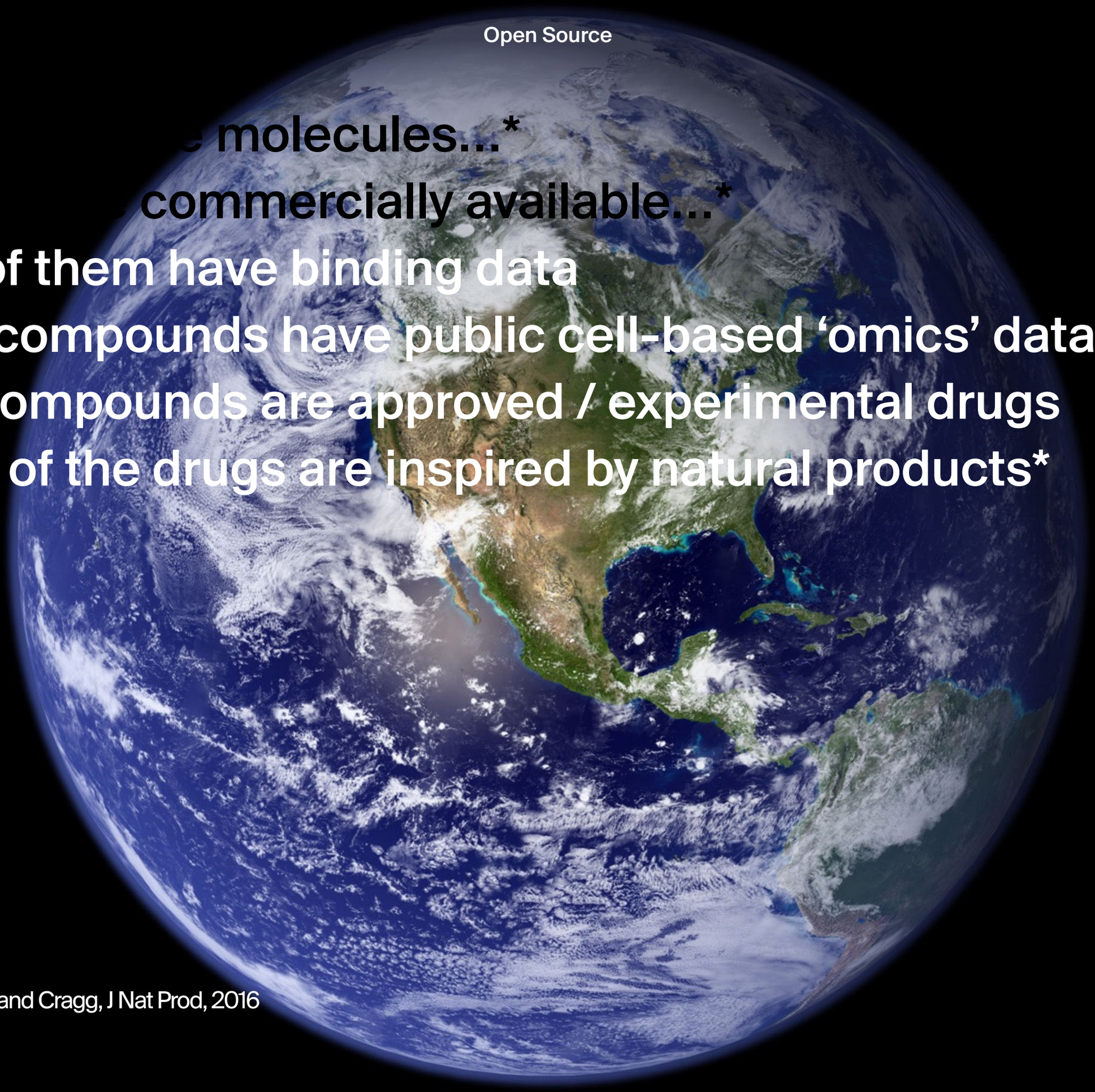
Antibiotics

A web-based interface



10^{60} possible molecules...*
200M are commercially available...*

Bertoni*, Duran-Frigola* et al. BioRXiV, 2020



molecules...*

commercially available...*

1% of them have binding data

10k compounds have public cell-based ‘omics’ data

5k compounds are approved / experimental drugs

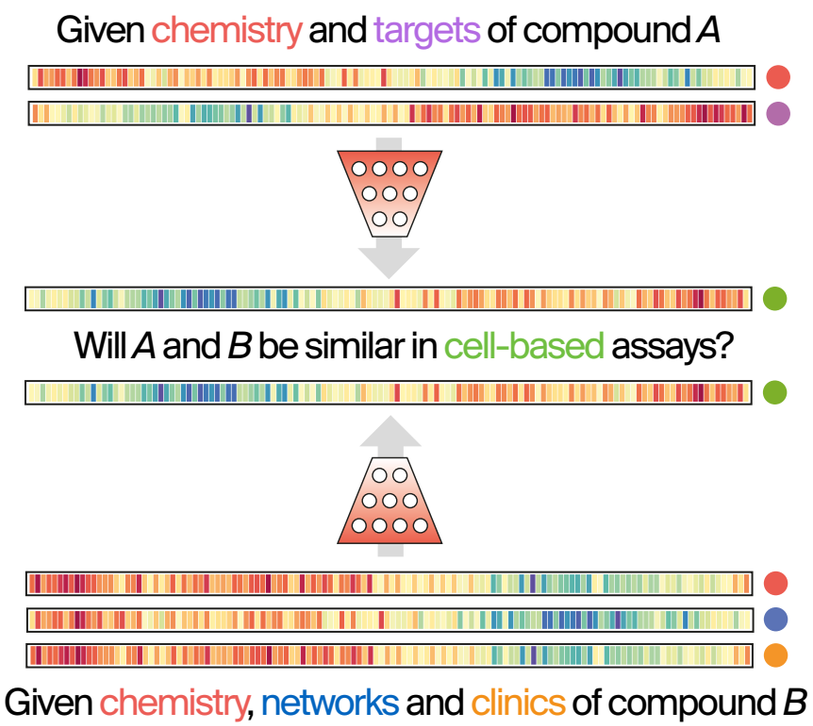
Half of the drugs are inspired by natural products*

Newman and Cragg, J Nat Prod, 2016

An honest view of the Chemical Checker

Chemistry	2D fingerprints	3D fingerprints	Scaffolds	Structural keys	Physico-chemistry
Targets	Mechanism of action	Metabolic genes	Crystals	Binding	HTS bioassays
Networks	Small mol. roles	Small mol. pathways	Signaling pathways	Biological processes	Interactome
Cells	Gene expression	Cancer cell lines	Chemical genetics	Morphology	Cell bioassays
Clinics	Therapeutic areas	Indications	Side effects	Diseases and toxicology	Drug-drug interactions

– Siamese neural nets 🤓



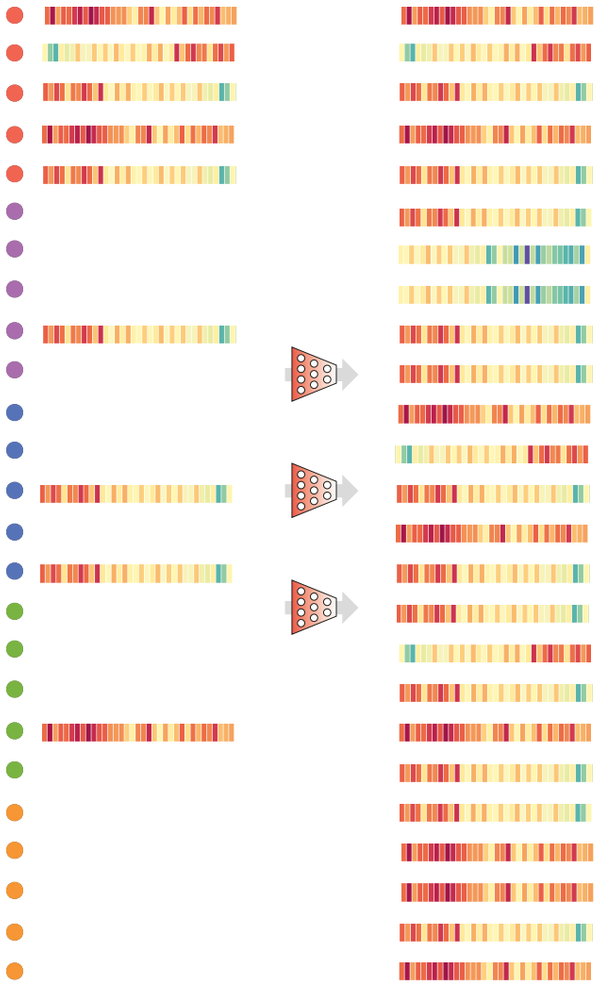
– bioactivitysignatures.org

– `pip install signaturizer`

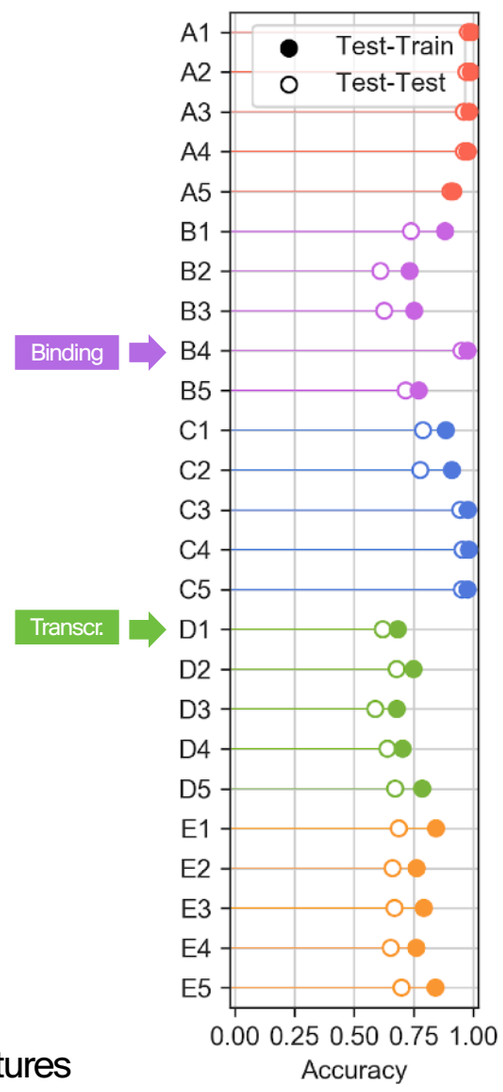
High-performance computing

Large-scale inference of bioactivity signatures

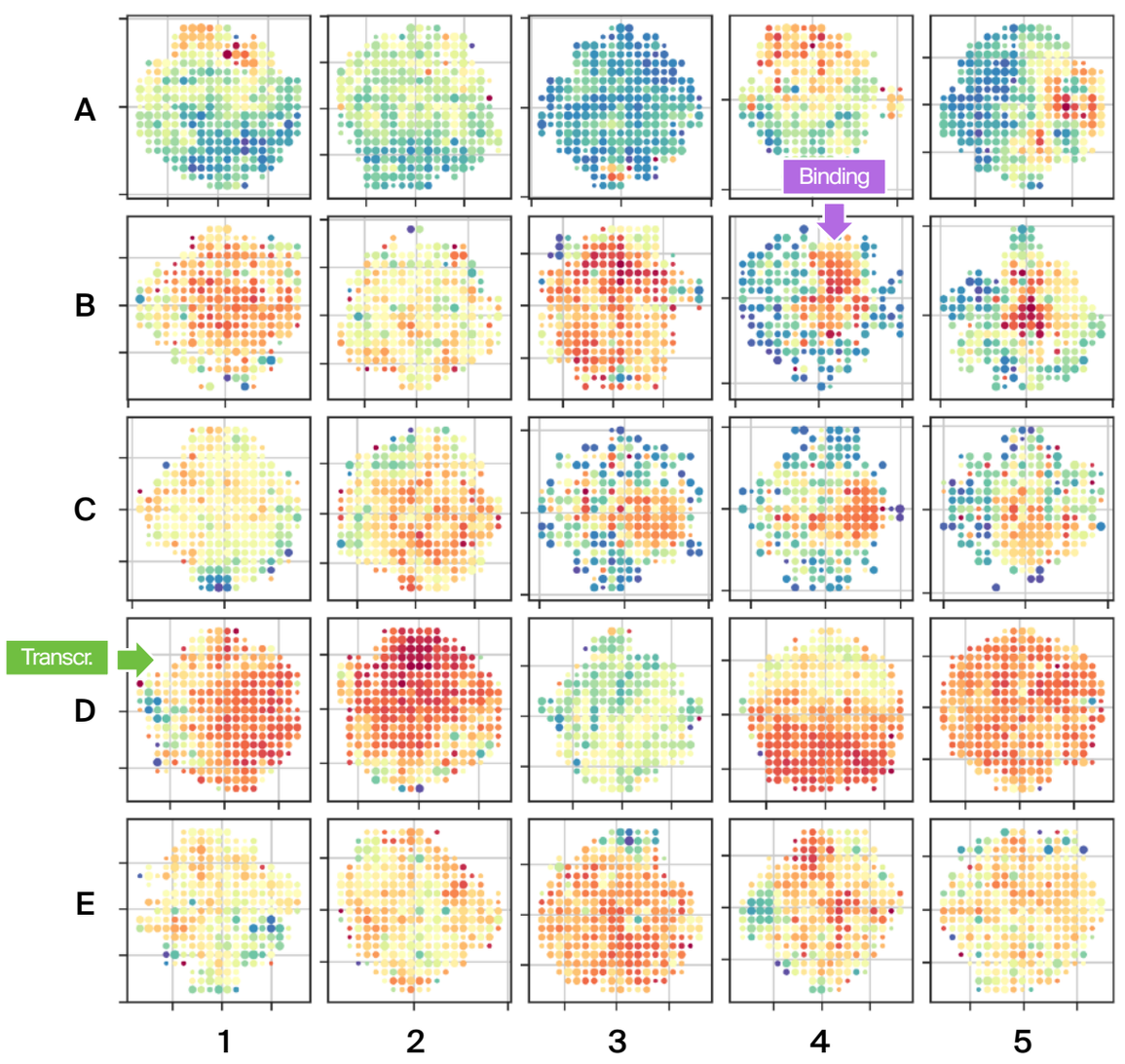
Prediction

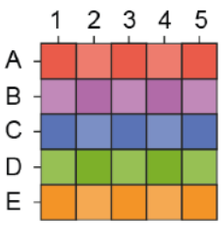


(a) Accuracy



(b) Confidence

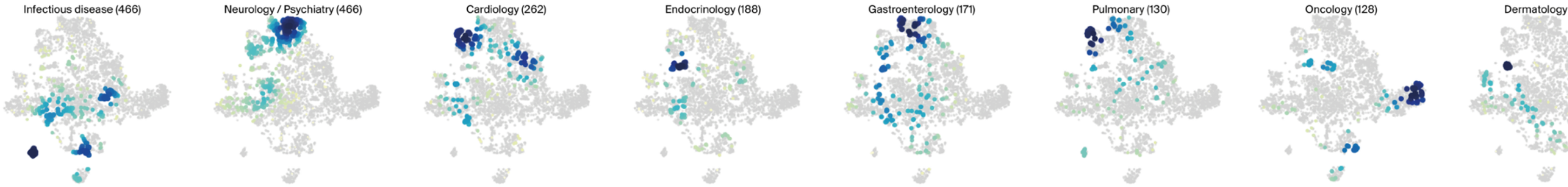




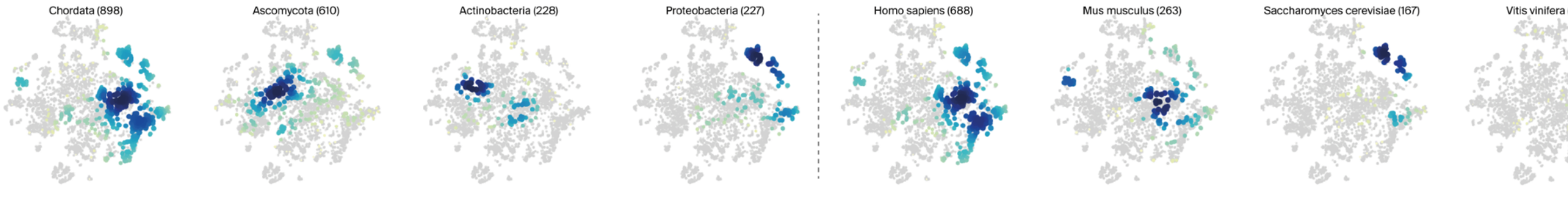
A1: 2D fingerprints
A2: 3D fingerprints
A3: Scaffolds
A4: Structural keys
A5: Physicochemistry
B1: Mechanisms of action
B2: Metabolic genes
B3: Crystals
B4: Binding
B5: HTS bioassays
C1: Small molecule roles
C2: Small molecule pathways
C3: Signaling pathways
C4: Biological proceses
C5: Interactome
D1: Transcription
D2: Cancer cell lines
D3: Chemical genetics
D4: Morphology
D5: Cell bioassays
E1: Therapeutic areas
E2: Indications
E3: Side effects
E4: Diseases & toxicology
E5: Drug-drug interactions

High
Medium
Low

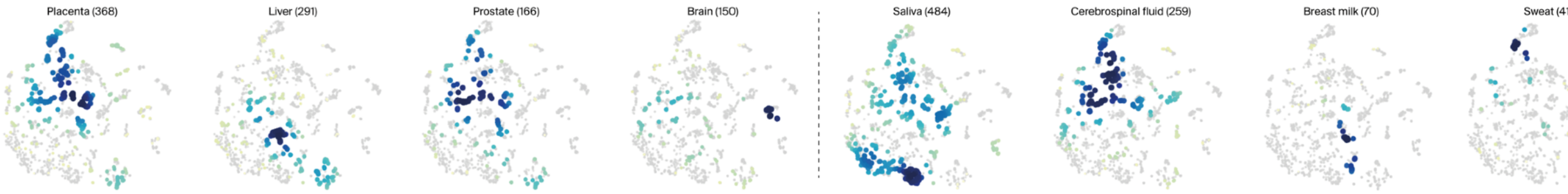
— Drug molecules



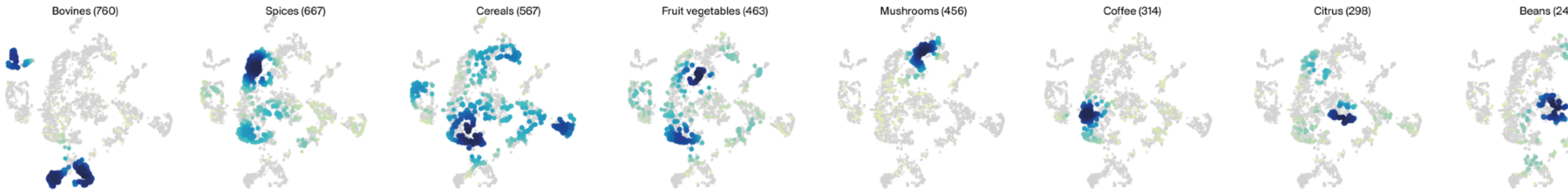
— Species metabolites



— Tissues and fluids



— Food ingredients



Take-home messages

- The CC contains compound signatures of 25 different types
- Experimental signatures are *not* always available for the compounds
- Inferred signatures are available for *every* compound (with a confidence estimate)
- CC signatures can be used to match and revert perturbational and disease signatures*
- 2D projections help visualise compound collections
- CC signatures can be used as drop-in replacements for chemical fingerprints (e.g. similarity search, SAR modelling)

* Not shown today

Thanks for your attention!

Bioactivity profiles for uncharacterised compounds

Computational Applications in Secondary Metabolites Discovery (CAiSMD), March 9 (2021)

Miquel Duran-Frigola, PhD
@mduranfrigola
miquel@ersilia.io
ersilia.io

Ersilia