

Gene Function Discovery Group

Finding new gene functions and facilitating applications in medicine, biodiversity, natural products, ...



Birgit & Frank Eisenhaber

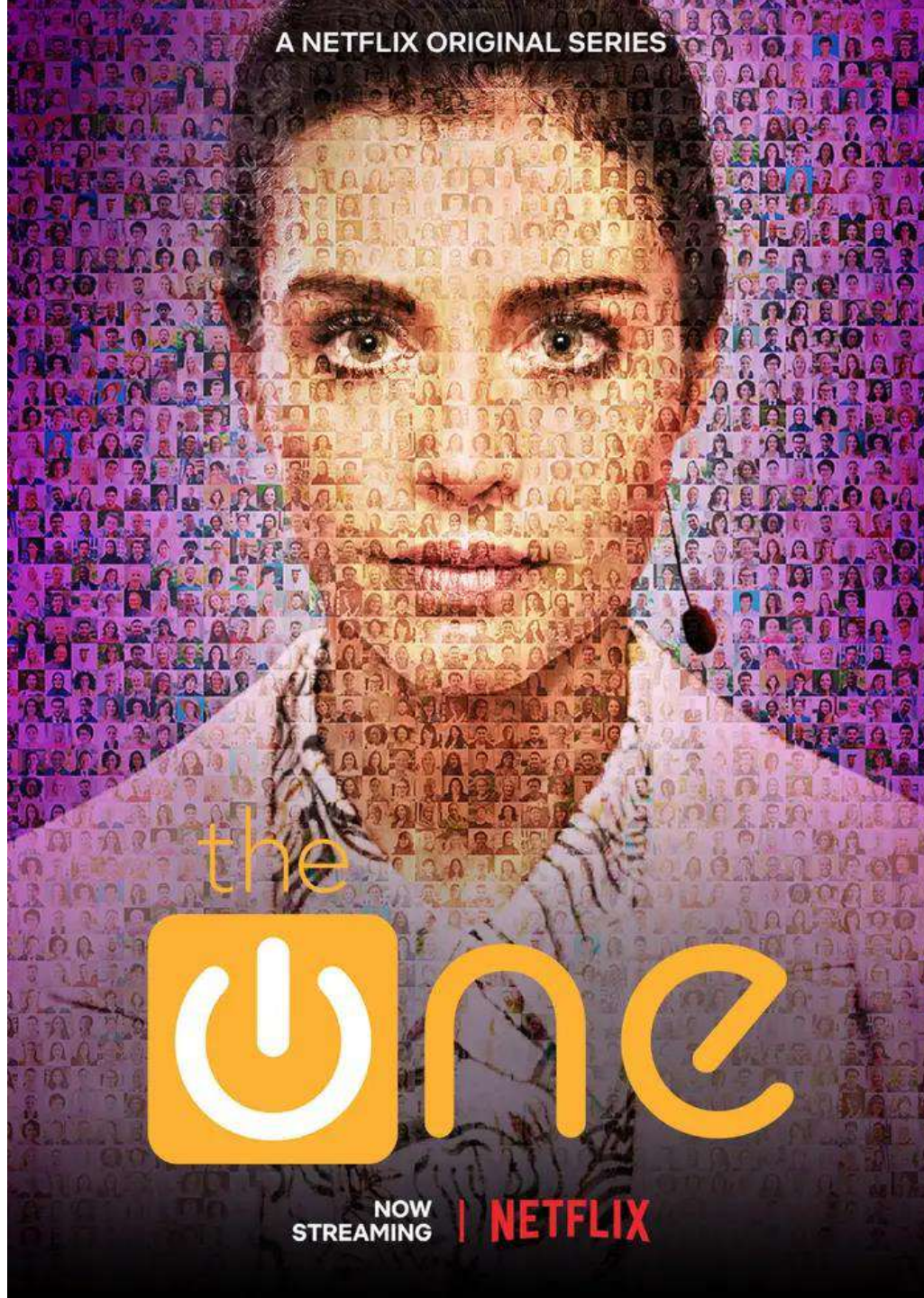
www.bii.a-star.edu.sg

birgite@bii.a-star.edu.sg

franke@bii.a-star.edu.sg

Singapore, 21st April 2021

A NETFLIX ORIGINAL SERIES



The One

Netflix 2021

1st bioinformatician in main-stream film

Main cast: **bioinformatician** and CEO Rebecca Webb

Finding the partnership match by human genome analysis

- Similarity of pheromone profile
- Matches of mutations
- **Download of millions of human genomes (1petaByte) with laptop on harddisk over home WIFI**

Reality:

- Human genome is far from functionally understood
- Partnership is favoured by non-matches of mutations and by distant immunological profiles

Most of the human genome is non-understood ...

- Human protein-coding genes (~1.5% of the genome)
 - ~4000 are not mentioned in a single article
 - ~7000 have very incomplete function characterization (~0.5% of life science literature)
 - ... but 95% of literature is about <5000 elite genes
- Only ~2500 ncRNAs are mentioned in the literature
 - 119 elite ncRNAs (~4% of all ncRNAs) are covered by 76% of the relevant literature
 - ~2200 ncRNAs (83% of all ncRNAs) attracted 5% of articles

Darkness in the Human Gene and Protein Function Space: Widely Modest or Absent Illumination by the Life Science Literature and the Trend for Fewer Protein Function Discoveries Since 2000.

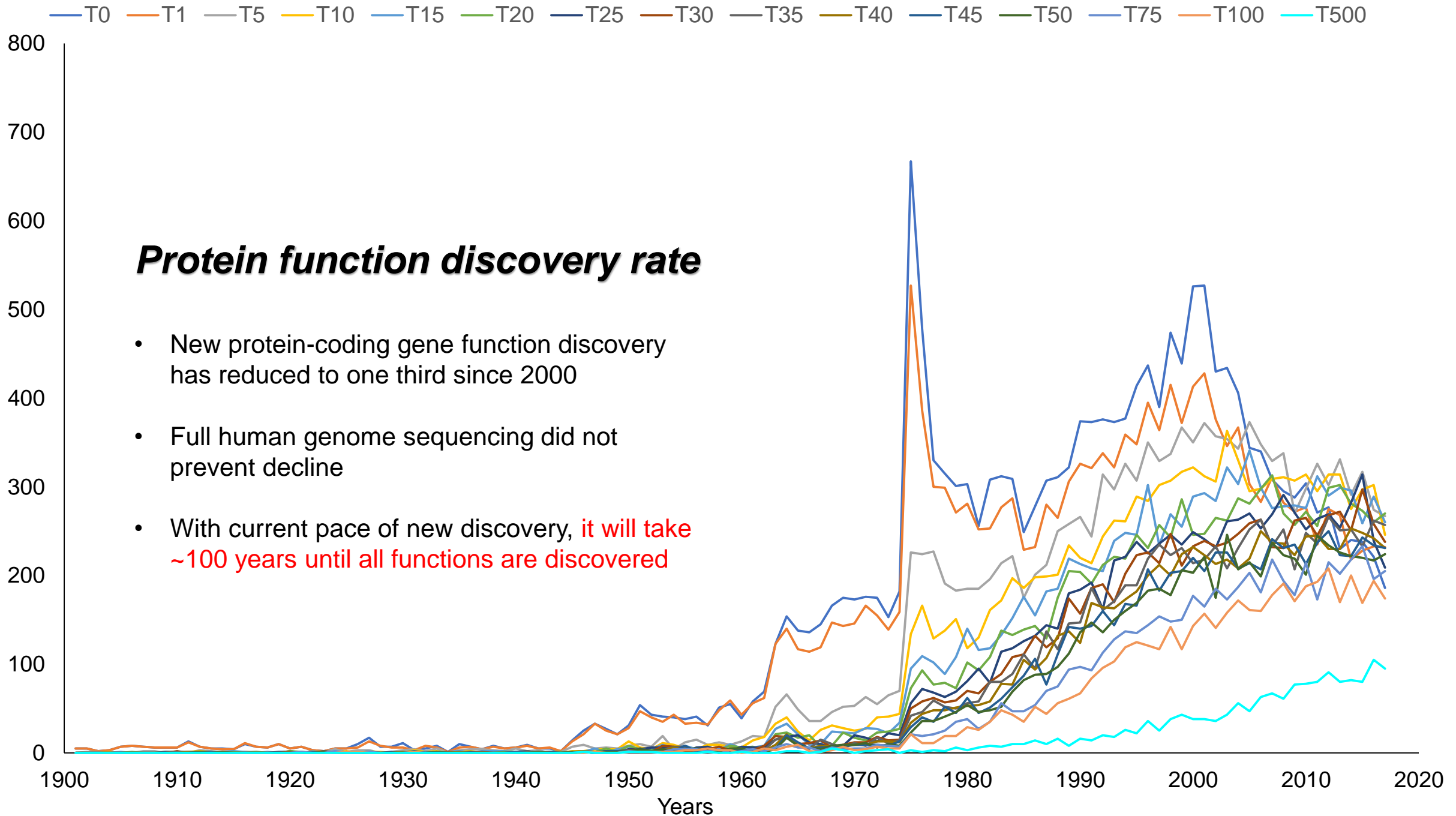
Sinha S, Eisenhaber B, Jensen LJ, Kalbuajji B, Eisenhaber F. *Proteomics*. 2018 Sep 28:e1800093. doi: 10.1002/pmic.201800093

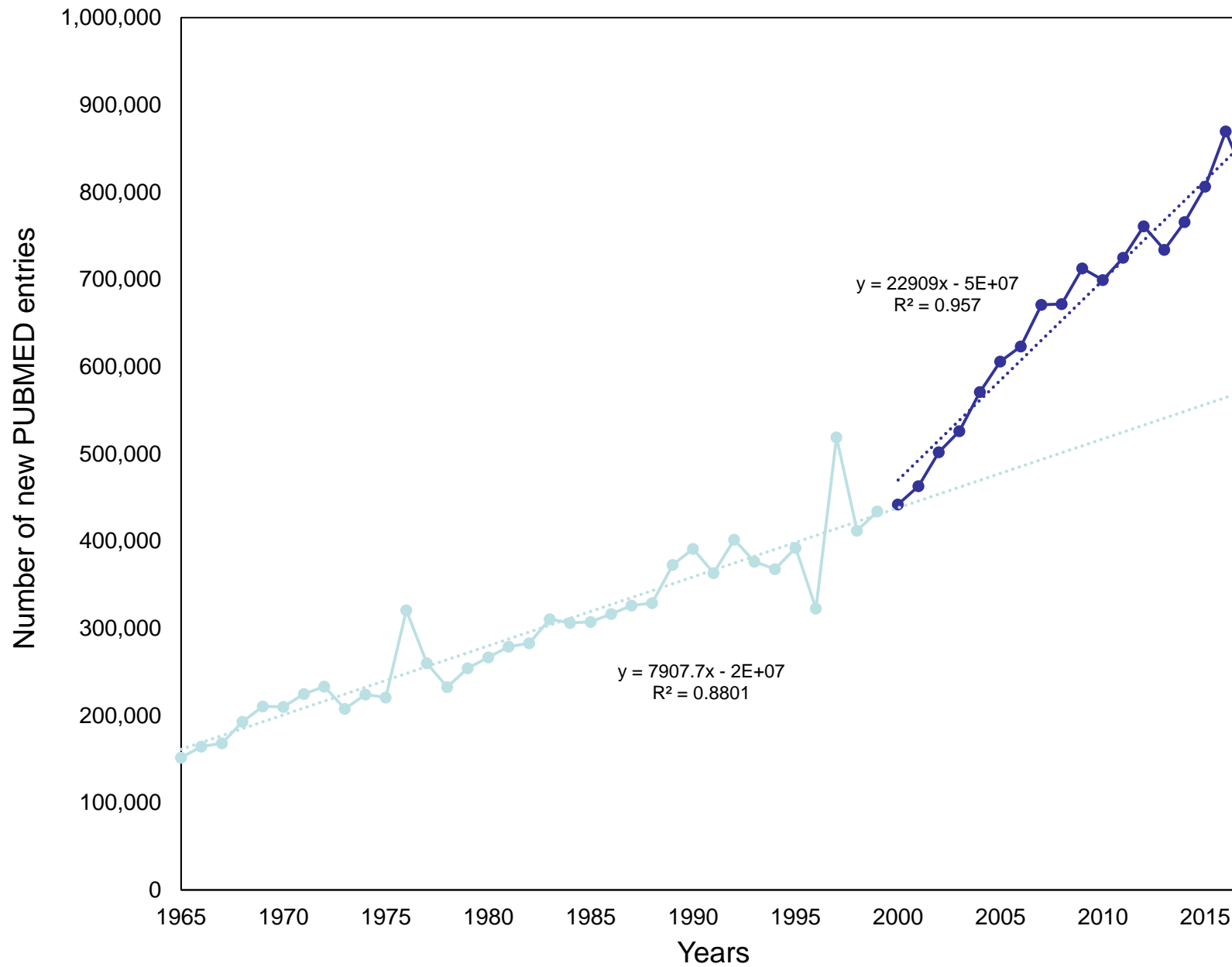


Protein function discovery rate

- New protein-coding gene function discovery has reduced to one third since 2000
- Full human genome sequencing did not prevent decline
- With current pace of new discovery, **it will take ~100 years until all functions are discovered**

Number of protein coding genes



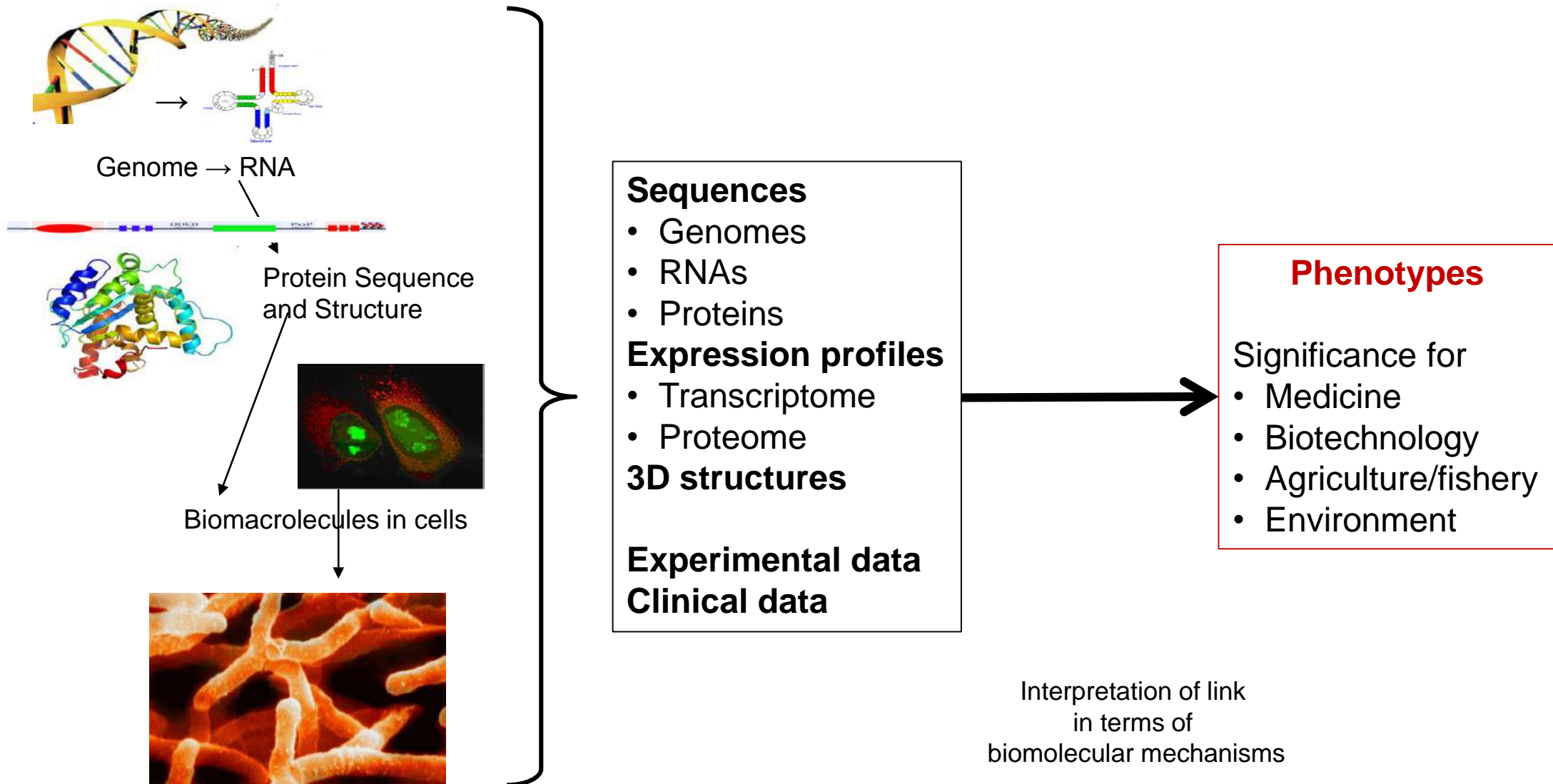


1970: 1.1 million entries
 2000: 10.7 million entries
 2017: 24.3 million entries

1965 – 2000
 Growth of annual growth
 by ~8,000 entries per year

2000 – 2017
 Growth of annual growth
 by ~23,000 entries per
 year

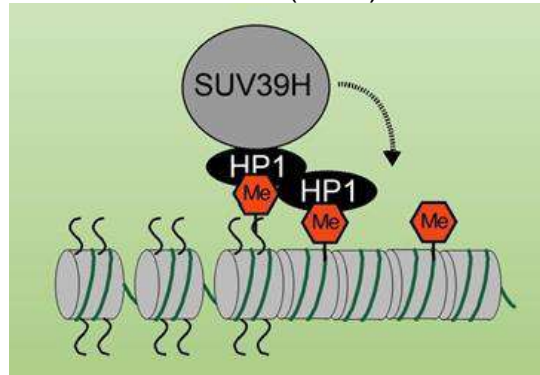
Gene Function Discovery Group



... dedicated efforts can open doors to new biology

SET domain histone methyltransferases biochemical epigenetics

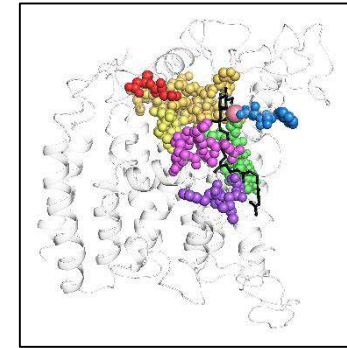
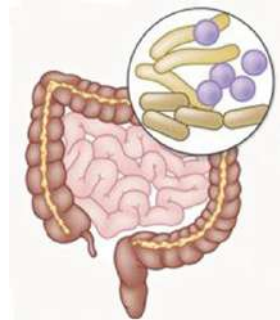
Nature 406 (2000) 593



SUGCT/C7ORF10

Glutaryl-coA synthetase for gut microbiome metabolites prevents liver/kidney inflammation metabolic syndrome/obesity

Cell Mol. Life Sci. 77 (2020) 3423

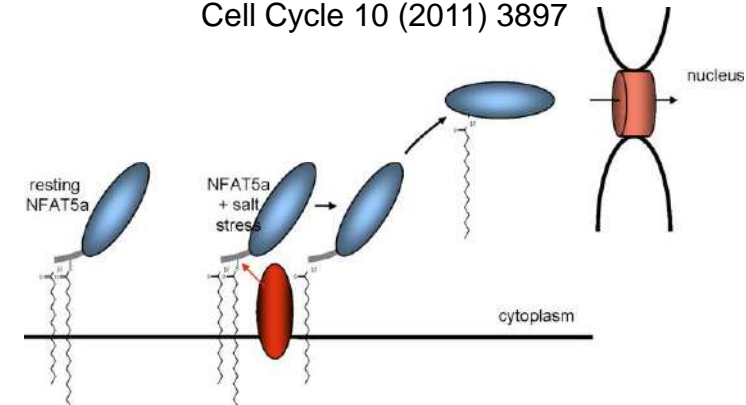


TMTCs - new O-mannosyltransferases from the GT-C/PMT clan

Biology Direct 16 (2021) 4

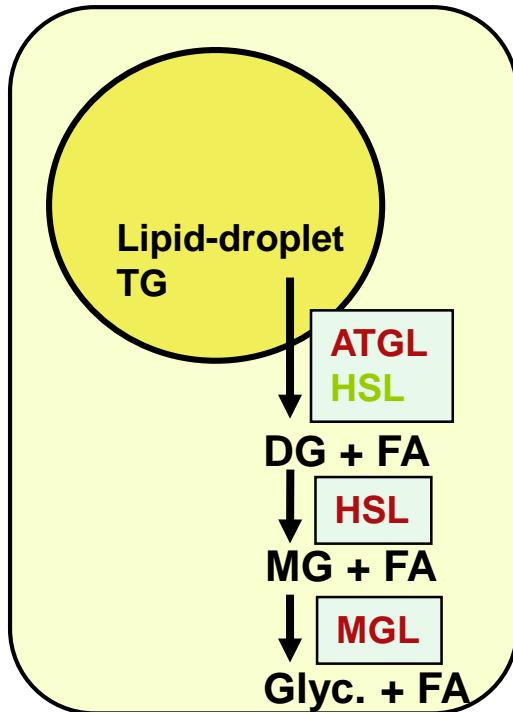
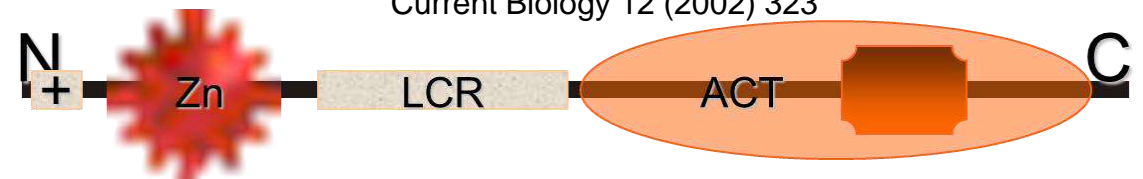
NFAT5a – osmotically sensitive transcription factor activated by LYPLA1 reversible de-palmitoylation

Cell Cycle 10 (2011) 3897



Eco1 – acetyltransferase for cohesion in cell division

Current Biology 12 (2002) 323



ATGL completed human TG catabolism pathway

Science 206 (2004) 1383

Before 2003: The transamidase complex for GPI lipid anchoring

Protein/Component	Function
Pig-K/gpi8	Cleavage of C-terminal propeptide (protease)
GPAA1/gaa1	?
Pig-T/gpi16	?
Pig-S/gpi17	?
Pig-U/cdc91/gab1	?

The transamidase complex in 2021

Protein/Component	Function
Pig-K/gpi8	Cleavage of C-terminal propeptide (protease)
GPAA1/gaa1	Peptide synthetase, attachment of the GPI-anchor via ethanolamine to C-terminal ω -site ¹
Pig-T/gpi16	Gate for regulated access to the active site of Pig-K (unusual β -propeller structure) ²
Pig-S/gpi17	?
Pig-U/cdc91/gab1	Conserved 10 TM-domain for presentation of the GPI lipid anchor (to GPAA1) ³

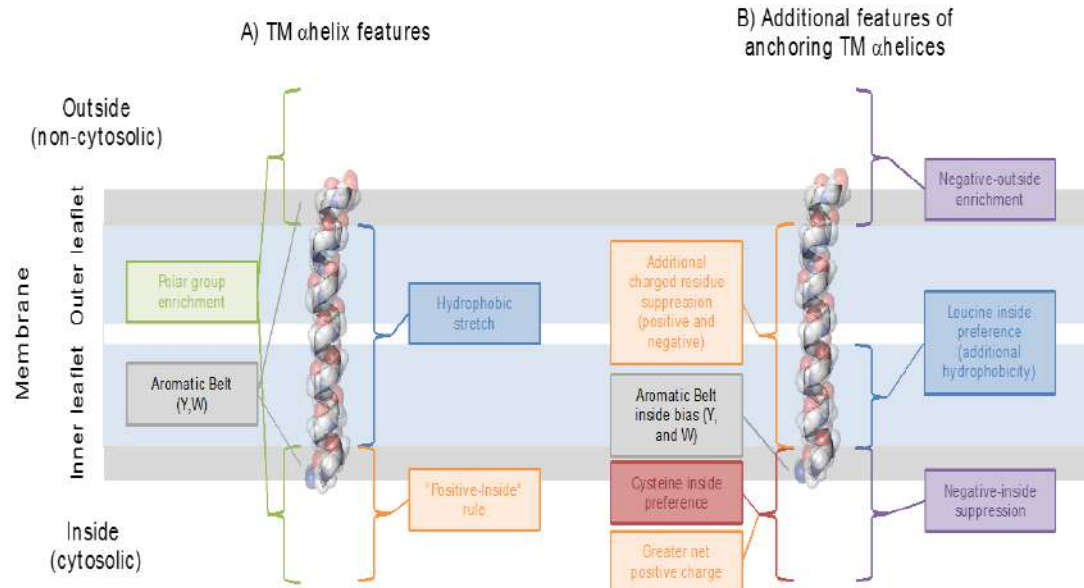
¹Bioessays 25 (2003) 367

²Cell Cycle 13 (2014)1912

³Cell Cycle 17 (2018) 874

The Negative-Outside Rule in Transmembrane Helices

James Baker (ARAP student from Manchester University 2016-2018)



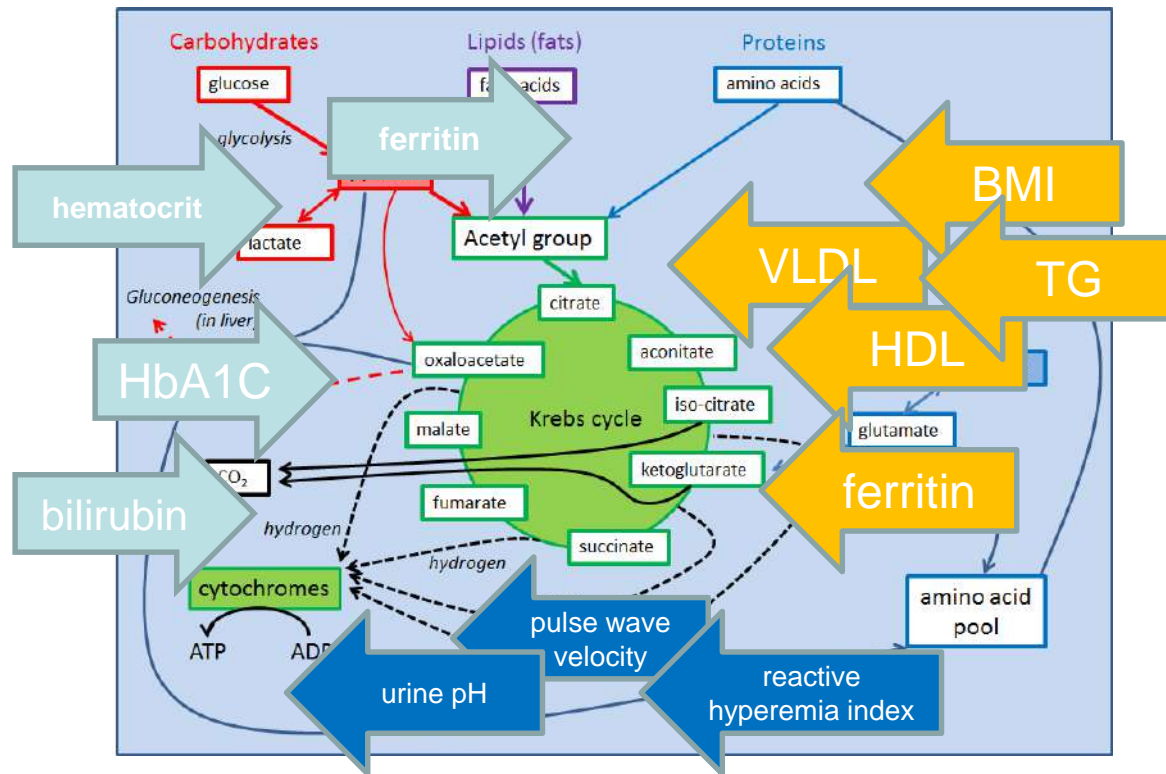
- ❑ Long-standing scientific problem of the last 25 years was finally solved, namely, the proof of the statistical bias of negative charges towards the extracellular leaflet of plasma membranes.
- ❑ This paper is a pleasure for lovers of sophisticated statistics since tools have been applied that were invented just 40 years ago

Publication: Baker, J.A., Wong, W.-C., Eisenhaber, B., Warwicker, J. and Eisenhaber, F. (2017) Charged residues next to transmembrane regions revisited: 'Positive-inside rule' is complemented by the 'negative inside depletion/outside enrichment rule'. **BMC Biol.** 2017 Aug 18;15(1):72.

Collaborations with MeshBio

<https://www.meshbio.com/>

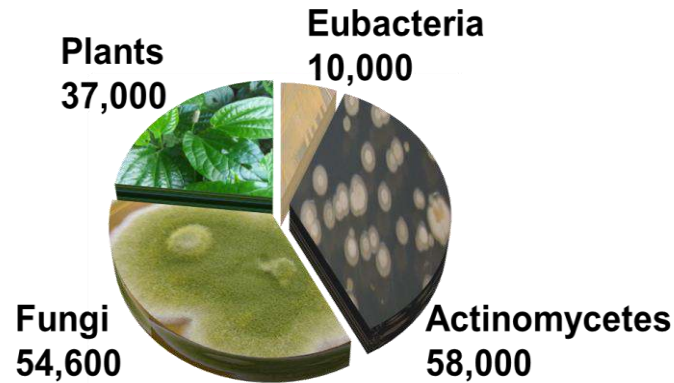
SME: Arsen Batagov, Andrew Wu
Successful recent refinancing (\$2mill.)



Innovative metabolic flux analysis (MFA) in simplified, coarse-grain gene/pathway networks

- Mapping of clinical laboratory/physical measurements into the gene network
- Computing patient-specific fluxes from her/his EHR
- In diabetes patients, vector of fluxes correlates with likelihood of complications (retinopathy/cataract)
- For a subgroup of patients, highly likely complications can be significantly predicted.

A*STAR Natural Organism Library



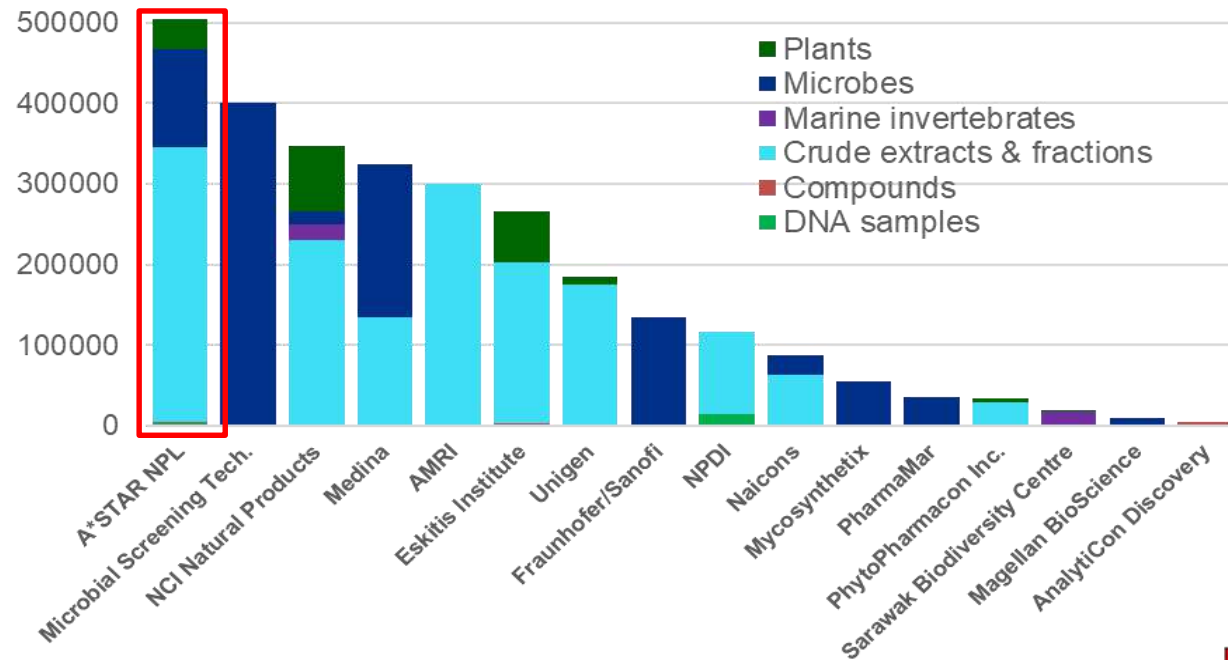
Samples from Singapore

- Actinomycetes: 11,485 strains
- Fungi: 5,846 strains
- Eubacteria: 2,185 strains
- Plants: 3,662 specimens

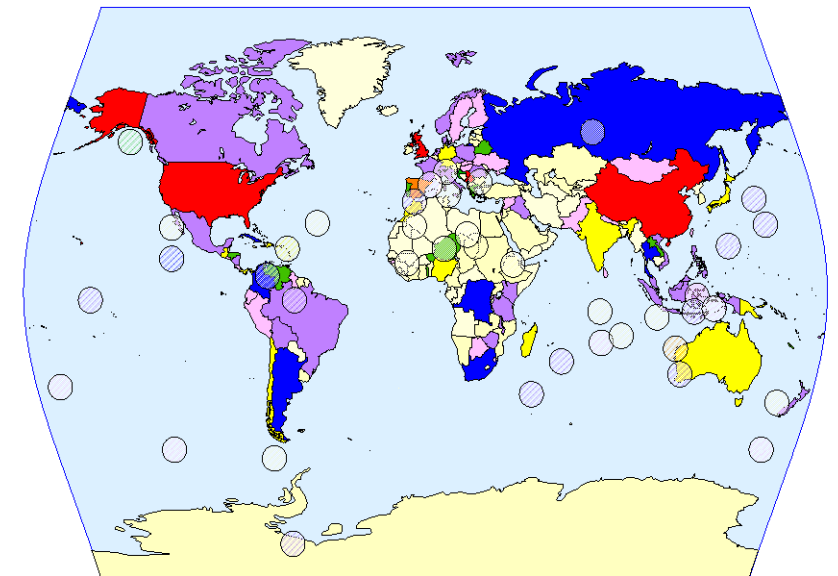
- Collected or acquired in accordance with **UN Convention on Biological Diversity**
 - ✓ IP generated from the use of supplied materials belong to A*STAR
 - ✓ Benefit sharing with supplier organisations/communities
 - ✓ 0.25-2% royalties; royalties plus milestones payment for plant samples from Kew Gardens
- **Nagoya Protocol** - international legal framework for the fair and equitable sharing of benefits arising out of commercial exploitation of biodiversity

	Plants	Microbes
A*STAR owned	3,140 (8%)	105,150 (86%)
3rd party providers	34,002 (92%)	17,316 (14%)

A*STAR's NOL is the world's largest collection of microbial+fungus+plant samples and has strategic value to Singapore



- Samples from more than 100 countries



Number of samples per country	
Red	> 10,000
Orange	5,001 – 10,000
Yellow	1,001 – 5,000
Blue	501 – 1,000
Purple	101 – 500
Green	51 – 100
Pink	10 – 50
Light Yellow	< 10

○ Location of aquatic samples

1. Based on disclosed data, absolute numbers indicated here for comparison
 Source: Ng et. al., Nat. Biotech. 36, 570-573 (2018)

Vision of NOL 2.0: *in silico* mining

The 160K Natural Organism Library, a unique resource for natural product research
Ng, Kanagasundaram, Arumugam, Hao, Eisenhaber & Eisenhaber, *Nat. Biotech.* 36, 570-573 (2018)

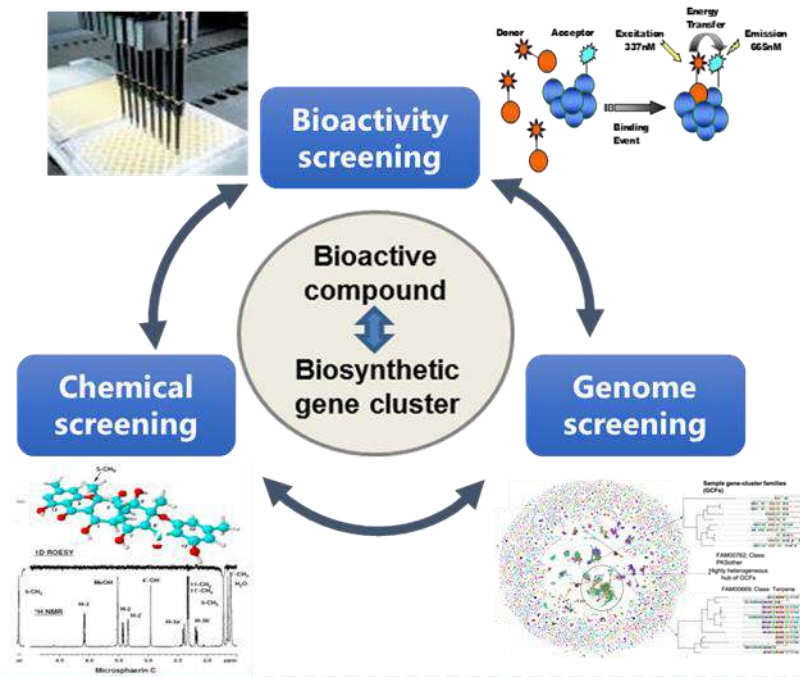
State of the art: biological high-throughput screening with subsequent analytical chemistry

- Depletion of biological material
- Tests only at very few expression conditions/concentrations
- Problem: assay development
- Very expensive with regard to time (~1 year) and lab resources

Vision:

- OMICS characterization of the NOL (genome sequencing etc.)
- Curation of iNOL
- Dedicated experimental follow-up for selected hits of the *in silico* screen
- NRF application pending (BII+GIS+SIFBI+ others)

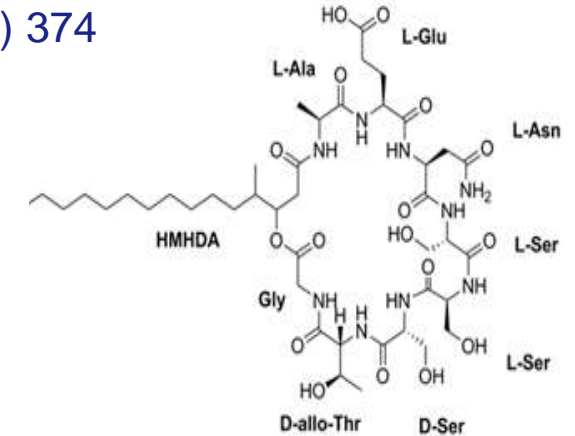
Discovery using multiple screening approaches; *in silico* screening follow by focused experimental validation



Discovery of antifungal **BII-Rafflesfungin** and its biosynthetic cluster

Ping-pong of experiment and computation: Sinha et al. BMC Genomics 20 (2019)374 20 (2019) 374

- ❑ Biosynthetic gene cluster (NRPS-t1PKS cluster 'BIIRfg') discovered in *Phoma* sp. F3723 genome.
- ❑ Antifungal activity-guided isolation yielded a new compound, BII-Rafflesfungin
- ❑ The structure of BII-Rafflesfungin was determined as cyclic lipodepsipeptide BII-Rafflesfungin [HMHDA-L-Ala-L-Glu-L-Asn-L-Ser-L-Ser-D-Ser-D-allo-Thr-Gly].
- ❑ New Stachelhaus codes for Ala, Glu, Asn, Ser, Thr, and Gly.
- ❑ Mechanism for BII-Rafflesfungin biosynthesis
 - formation of the lipid part by BIIRfg_PKS
 - followed by activation and transfer of the lipid chain by a AMP-ligase on to the first PCP domain of the BIIRfg_NRPS gene to
 - initiate the peptide synthesis. The CT domain terminates the peptide synthesis.



Fungal strain/Cell line	IC50, μ M
<i>Candida albicans</i> ATCC 10231	2.4
<i>Candida albicans</i> ATCC 90028	4.6
<i>Saccharomyces cerevisiae</i> (BY4741)	2.7
<i>Aspergillus fumigatus</i> ATCC 46645	1.2
<i>Aspergillus brasiliensis</i> ATCC 16404	7.4
A549 lung carcinoma cells	16.5
HepG2 liver carcinoma cells	13.8



Gene Function Discovery Group

Thank you!

Birgit & Frank Eisenhaber

www.bii.a-star.edu.sg

birgite@bii.a-star.edu.sg, franke@bii.a-star.edu.sg

Singapore, 21st April 2021



Group members and guests: Fernanda Maurer-Stroh, Konstancja Urbaniak, Erwin Tantoso, Wu Shuang, Swati Sinha, Chun Teck Lim, Chai Yee Wilson Kwo