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

Gene Function Discovery Group

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
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.
**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
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

Genome → RNA

Protein Sequence and Structure

Biomacromolecules in cells

Sequences
- Genomes
- RNAs
- Proteins

Expression profiles
- Transcriptome
- Proteome

3D structures

Experimental data
Clinical data

Phenotypes
Significance for
- Medicine
- Biotechnology
- Agriculture/fishery
- Environment

Interpretation of link in terms of biomolecular mechanisms
… dedicated efforts can open doors to new biology

**Lipid-droplet TG**

- ATGL
- HSL
- DG + FA
- MG + FA
- MGL
- Glyc. + FA

**SET domain histone methyltransferases**

- **biochemical epigenetics**


**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**


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

  *Biology Direct 16* (2021) 4

**SUGCT/C7ORF10**

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

  *Cell Mol. Life Sci. 77* (2020) 3423
Before 2003: The transamidase complex for GPI lipid anchoring

<table>
<thead>
<tr>
<th>Protein/Component</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pig-K/gpi8</td>
<td>Cleavage of C-terminal propeptide (protease)</td>
</tr>
<tr>
<td>GPAA1/gaa1</td>
<td>?</td>
</tr>
<tr>
<td>Pig-T/gpi16</td>
<td>?</td>
</tr>
<tr>
<td>Pig-S/gpi17</td>
<td>?</td>
</tr>
<tr>
<td>Pig-U/cdc91/gab1</td>
<td>?</td>
</tr>
</tbody>
</table>
# The transamidase complex in 2021

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

<sup>1</sup>Bioessays 25 (2003) 367  
<sup>2</sup>Cell Cycle 13 (2014)1912  
<sup>3</sup>Cell 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.

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

<table>
<thead>
<tr>
<th></th>
<th>Plants</th>
<th>Microbes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*STAR owned</td>
<td>3,140 (8%)</td>
<td>105,150 (86%)</td>
</tr>
<tr>
<td>3rd party providers</td>
<td>34,002 (92%)</td>
<td>17,316 (14%)</td>
</tr>
</tbody>
</table>
A*STAR’s NOL is the world’s largest collection of microbial+fungal+plant samples and has strategic value to Singapore

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 of antifungal BII-Rafflesfungin and its biosynthetic cluster


- 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.

<table>
<thead>
<tr>
<th>Fungal strain/Cell line</th>
<th>IC50, µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candida albicans ATCC 10231</td>
<td>2.4</td>
</tr>
<tr>
<td>Candida albicans ATCC 90028</td>
<td>4.6</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae (BY14741)</td>
<td>2.7</td>
</tr>
<tr>
<td>Aspergillus fumigatus ATCC 46645</td>
<td>1.2</td>
</tr>
<tr>
<td>Aspergillus brasiliensis ATCC 16404</td>
<td>7.4</td>
</tr>
<tr>
<td>A549 lung carcinoma cells</td>
<td>16.5</td>
</tr>
<tr>
<td>HepG2 liver carcinoma cells</td>
<td>13.8</td>
</tr>
</tbody>
</table>
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