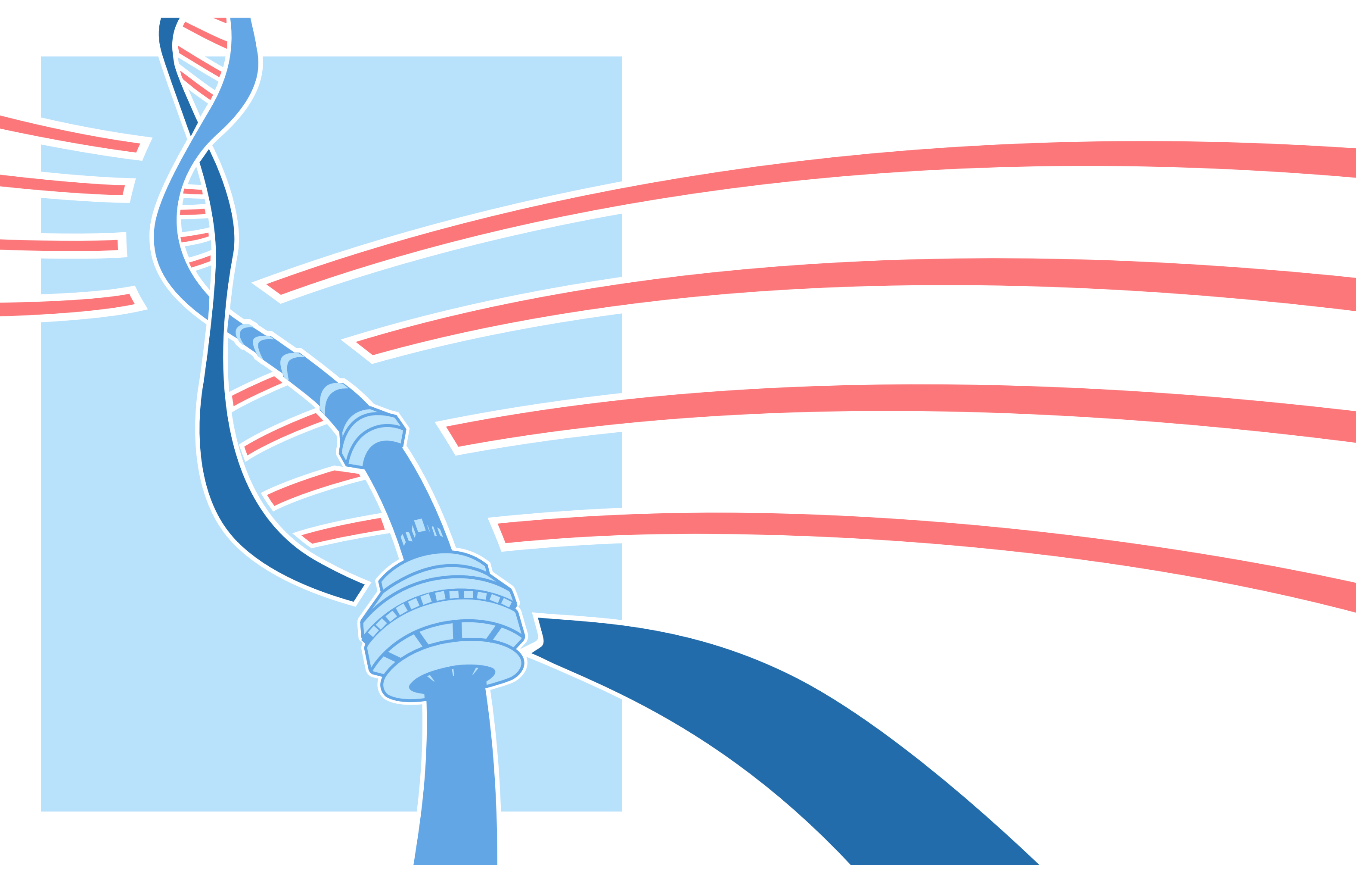


Mutational Scanning Symposium

June 13, 14, and 16, 2022
Toronto Canada

Program and Abstracts

TABLE OF CONTENTS

Sponsors..3

Agenda
- Day 1: June 13..4
- Day 2: June 14..5
- Poster Day: June 16...................................7

In Person Meeting Information...............................8

Virtual Meeting Information..................................10

Poster Day Information
- Gathertown Information..............................11
- Presenter Session and Poster Number............14

Speaker Abstracts..16

Planning Committee...27

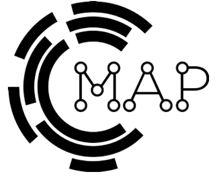
Abstracts..28

Platinum Sponsors

Gold Sponsor

Silver Sponsors

Academic Partners



Agenda

Day 1: June 13

Arrival and registration	9:00 - 9:30am (EST)
Fritz Roth, University of Toronto Welcome, Opening remarks	9:30 - 9:40am
Clare Turnbull, The Institute of Cancer Research (Keynote/Featured speaker) Conundrums and challenges of clinical variant interpretation	9:40 - 10:25am
Jacob Kitzman, University of Michigan Checking on the spellcheckers: deep mutational scans of mismatch repair factors and beyond	10:35 - 11:00am
Julia Domingo, New York Genome Center Global mapping of the energetic and allosteric landscapes of protein binding domains	11:10 - 11:35am
Lunch/Free Time	11:35 - 1:05pm
Holly Rees, Beam Therapeutics Base editing – a next-generation genome editing technology	1:05 - 1:30pm
Benedetta Bolognesi, Institute for Bioengineering of Catalunya Deep indel mutagenesis (DIM) reveals novel Aß mutations that accelerate amyloid formation and are likely pathogenic	1:40 - 2:05pm
Atlas of Variant Effects Alliance Team Presentation	2:05 - 2:30pm
Break	2:30 - 3:00pm
Alexandra Bendel, Friedrich Miescher Institute for Biomedical Research Systematic mutational scanning of the entire human leucine zipper interaction network	3:00 - 3:25pm
Timothy Yu, Fred Hutchinson Cancer Center Measuring viral escape from polyclonal antibodies	3:25 - 3:50pm
Debora Marks, Harvard Medical School Predicting the effects of genetic variation and designing biotherapeutics with neural machines	3:50 – 4:15pm

Agenda

Day 1: June 13 (continued)

Break	4:15 - 4:25pm
Lightning talks Troy McDiarmid, University of Washington Identification of enhancers to rescue haploinsufficiency in neurodevelopmental disorders via a CRISPRa-QTL approach	4:25 - 5:00pm
Sophie Moggridge, University of Washington Interrogating protein variant dysfunction using multidimensional proteomics	
Nidhi Sahni, MD Anderson Cancer Center Functional variomics: Systematic annotation of somatic mutations and gene fusions in cancer	
Moez Dawood, Baylor College of Medicine Saturation Genome Editing Reveals 10% of Missense SNVs in Functional Domains of PALB2 as Functionally Abnormal	
Alex McDonnell, University of Edinburgh Probing PAX6-DNA Interactions Using High-Throughput Yeast One-Hybrid Coupled with Deep Mutational Scanning	
Free Time	5:00 - 6:00pm
Cocktail Hour Hart House	6:00 - 7:00pm
Banquet Hart House	7:00 - 9:00pm

Day 2: June 14

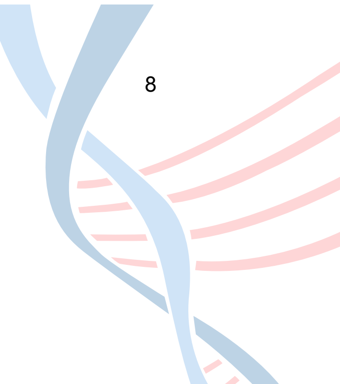
Arrivals	9:00 - 9:30am (EST)
Welcome and housekeeping	9:30 - 9:35am
Doug Fowler, University of Washington (Keynote/Featured Speaker) What do we need to scale multiplexed assays to all genes in the genome?	9:35 - 10:20am
Workshop Talks Zhenya Ivakine, Hospital for Sick Children All you need to know about saturation prime editing	10:20 – 10:45am
Kresten Lindorf-Larsen, University of Copenhagen Understanding the origins of loss of protein function using analyses of protein stability and conservation	10:45 - 11:10am

Day 2: June 14 (continued)

Jochen Weile, University of Toronto Translating multiplexed variant effect assays into evidence codes for clinical interpretation	11:10 - 11:35am
Shawn Fayer, University of Washington Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN	
Lunch (Self-organized)	11:45 - 1:30pm
Mafalda Dias - Predicting the effect of genetic variation with deep generative modeling	1:30 - 1:40pm
Lightning Talks Matt Coelho, Wellcome Sanger Institute Base editing screens map mutations affecting IFNg signalling in cancer	1:50 - 2:20pm
Willow Coyote-Maestas, University of California San Francisco DIMPLE, a method for systematic, low bias, and affordable, missense and indel scanning libraries	
Laura Haynes, University of Michgan Fingerprinting serine protease inhibitor (SERPIN) specificity	
Munmun Bhasin, Indian Institute of Science Structure and stability predictions from saturation mutagenesis	
Rick Boonen, Leiden University Medical Centre High-throughput analysis of PALB2 missense variants: linking functional impact to breast cancer risk	2:20 – 2:45pm
Break	2:45 – 3:15pm
Belinda Chang, University of Toronto Mutational landscapes and adaptive evolution of rhodopsin	3:15 – 3:40pm
Matthew Hurles, Wellcome Sanger Institute Saturation genome editing of DDX3X clarifies pathogenicity of germline and somatic variation	3:50 – 4:15pm
Closing remarks	4:15 - 4:20pm

Poster day: June 16

Opening remarks	11:00 – 11:05am (EST)
Session 1	11:05 – 11:35am
Session 2	11:35 – 12:05pm
Break	12:05 – 12:25pm
Session 3	12:25 – 12:55pm
Session 4	12:55 – 1:25pm
Closing remarks	1:25 – 1:30pm



IN PERSON MEETING INFORMATION

Symposium Location
The Symposium is located in the MaRS Centre, at 661 University Avenue (Southeast corner of University Avenue and College Street). Upon entry into the building using the doors facing University Avenue, take the escalator to the lower level to find the conference facility.

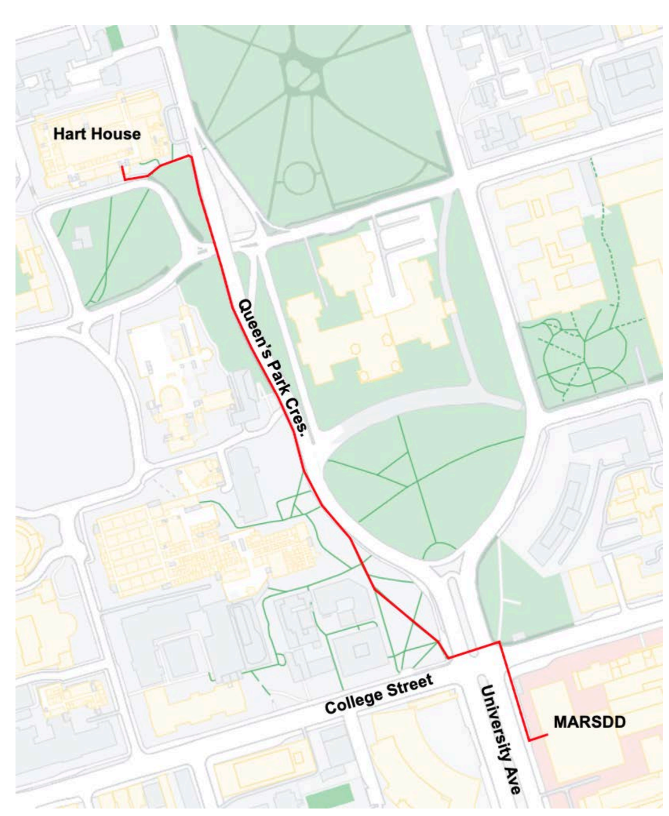
Registration will commence at 9:00am on Monday June 13th.

For lunch on both days there are few options to choose from:

* MARS Food Court located on the basement level of the building (with underground access to Toronto General Hospital Food Court.
* Mercatto Restaurant located on the ground floor of the MARSDD building.
* Food Court located in the Ontario Hydro Building on University Avenue across from MaRS Centre.
* Various restaurants on 1) Elm street, east of Bay street, 2) College Street (south side), west of University Ave, and 3) Baldwin street (see next page for map).

MaRS COVID Policy
While the wearing of masks is not mandatory in the MARSDD building or whilst in common areas (Atrium, washrooms, elevators, loading dock, mailroom, shared corridors, CIBC LIVE Lounge, etc), we request that registrants wear their masks while in the main auditorium out of respect for others with varying levels of risk.

Cocktail Hour and Banquet Location
The social hour and banquet will be held at Hart House, located on the University of Toronto campus at 7 Hart House Circle. It is a 10 minute walk from the MaRS Centre (see walking map =>).



Hart House COVID Policy
For all banquet attendees, please note that Hart House requires the wearing of masks inside the building except when eating or drinking. Masks are not required outside in the Quad of Hart House.

IN PERSON MEETING INFORMATION

MaRS Discovery District (Symposium Venue) Hart House (Banquet Venue)

University of Toronto Restaurant Coffee Shop Subway Station

[1] Chelsea Hotel [2] DoubleTree By Hilton Hotel [3] Royal Sonesta Hotel

VIRTUAL MEETING INFORMATION

Vimeo Link For Online Streaming

Please find below the vimeo link to access the live stream of the Symposium for both days. Live Streaming of the program on both June 13th and June 14th will start at 9.20am Eastern Standard Time.

https://vimeo.com/event/2189254

Technical Issues

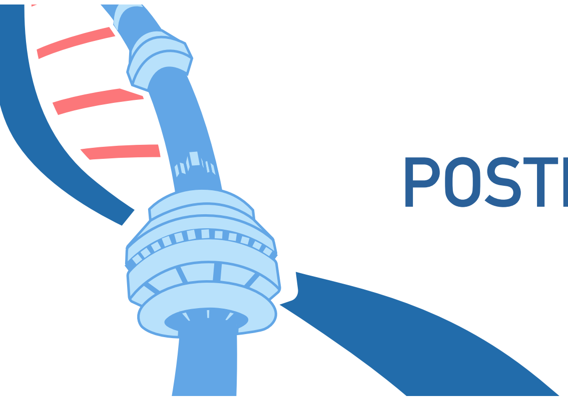
If you encounter any technical issues and need assistance during both days of the Symposium please send us an email at mss22@varianteffect.org and someone will reach out to you.

Asking questions during the symposium

We ask virtual participants to type their questions in the chat window. These will be read to the audience by the online moderators.

Interacting in Gathertown

Virtual attendees are encouraged to view the conference together and have side conversations during the entire meeting from within Gathertown. A guide to using the Gathertown space is on page 11.



POSTER DAY INFORMATION

Virtual Poster Session Details
We will be hosting a live virtual poster session on day 3 of the symposium (June 16th), and we ask you to join us during this event to connect directly with attendees via Gather Town. Instructions on how to access this virtual platform and how to present posters is below. Posters will also be available for viewing on Gathertown during the entirety of the symposium.

Live Poster Session (IMPORTANT):
When: Wednesday, June 16th
Authors at Posters: 11:00am-1:30pm EST
The posters will be available all day. However, the live poster session will be from 11:00am-1:30pm EST. During this time, we ask all presenting authors to stand at their posters during their allocated session time.

Awards:
During the live poster session, poster judges will be going around and stopping by posters for a brief amount of time. Again this year, we have three award categories (Technology & Tools, Fundamental Discovery, & Translational Application). Award winners will be announced by email on Friday June 17th.

Gathertown link: https://app.gather.town/app/tEv5md7Utw2i2bgB/DMS
Password : 22mss22

How to use our Gather Town Space:
Video tutorial*: https://www.youtube.com/watch?v=IdRi0CPohYA&feature=emb_imp_woyt
*The video tutorial has a few pro-trips at the very end that are specific to presenting a poster in Gather Town.

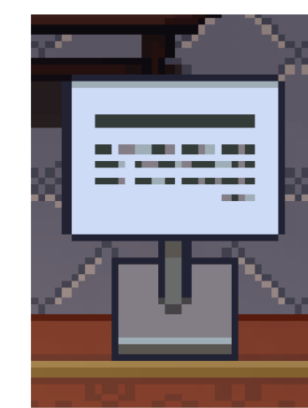
1. Setting your name and avatar
Before entering Gathertown, you will be asked to set your name and avatar. Please refer to the following link for more instructions, if needed:
https://support.gather.town/help/movement-and-basics
This link also provides information on help with camera and microphone accessibility.

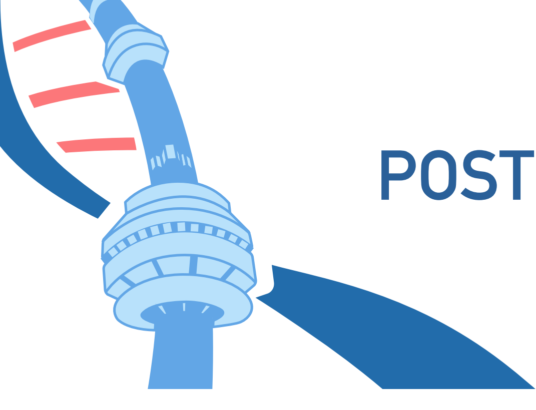
2. Interacting in the space
Upon entry into the space, your camera and microphone will be muted. You can choose to unmute and to display video by clicking on the red video and microphone icons at the bottom right of the screen, on your video display. Moving your avatar is simple. Just use the computer's arrow keys to walk around the space. When you are in close proximity to someone else, you will be able to hear their audio and their video will be displayed on the top of the screen. Likewise, your audio and video will be displayed to them (if unmuted).

3. The Poster Rooms

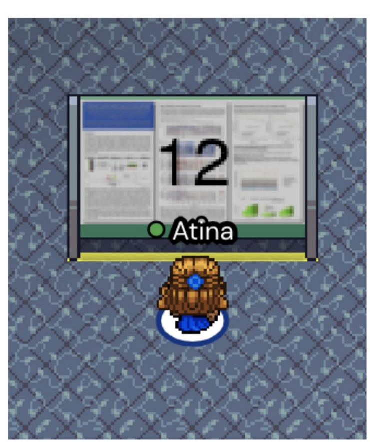
Poster list

There are 7 poster rooms, A to G. At the entryways to each poster room, you will find a podium that lists the poster titles and numbers. To view the list, walk your avatar to the podium and press x. You can scroll through the poster list then press x in the top right corner to return to the gather space.





POSTER DAY INFORMATION

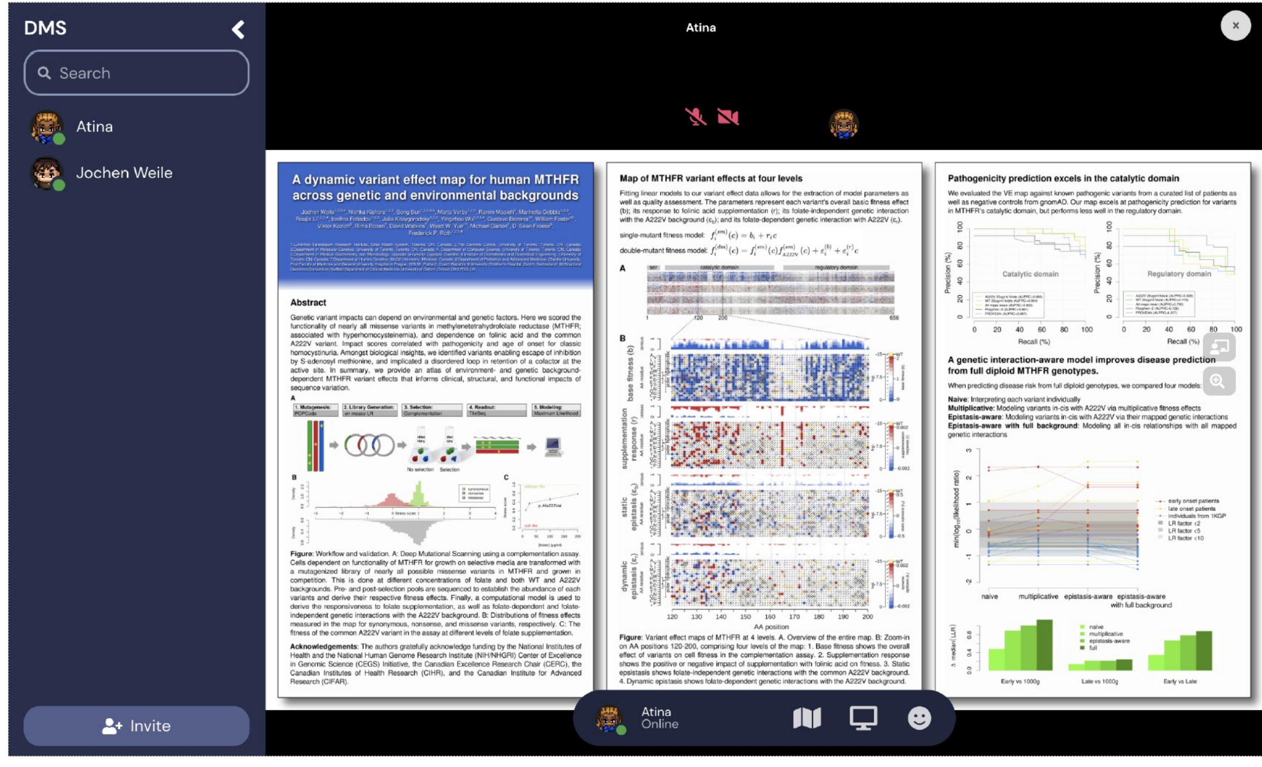


Viewing a poster

To view a poster, walk into the vicinity of the particular poster and press x to open it on your screen. Press x in the top right corner again to return to the gather space.

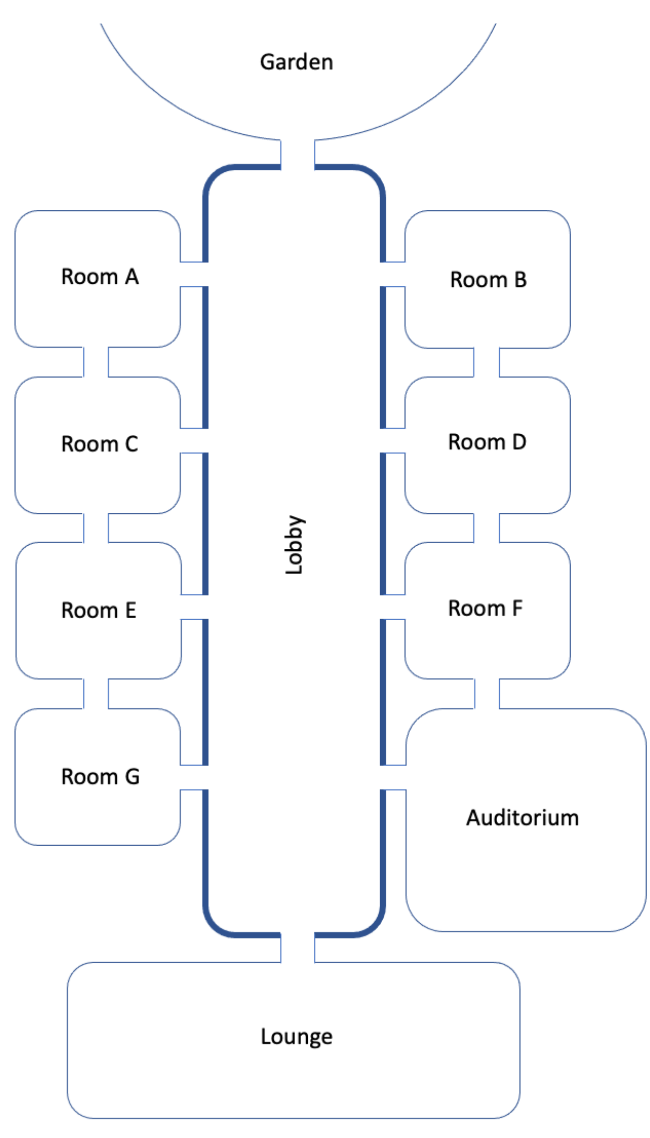
Interacting with a poster presenter

When you are in a particular poster space, you can also interact with the poster presenter if they are present in the space with you. While the poster is open, you will find a zoom icon, as well as a pointer function on the right side of the screen.



POSTER DAY INFORMATION

Map of Symposium Gathertown space



POSTER DAY INFORMATION
Presenter Room and Session

Poster Number	Session	Presenter Name	Poster Number	Session	Presenter Name
A01	1	Alexandra Bendel	C03	3	Albert Escobedo
A02	2	Munmun Bhasin	C04	4	Clayton E. Friedman
A03	3	Matthew Coelho	C05	1	Andrea Garofoli
A04	4	Moez Dawood	C06	2	Marinella Gebbia
A05	1	Julia Domingo	C07	3	Renee Geck
A06	2	Shawn Fayer	C08	4	Sebastian Gerety
A07	3	Laura Haynes	C09	1	Sarah Gersing
A08	4	Troy A. McDiarmid	C10	2	Marta Badia Graset
A09	1	Alex McDonnell	C11	3	Murat Guler
A10	2	Sophie Moggridge	C12	4	Gabe Haller
A11	3	Jochen Weile	C13	1	Shelby Hemker
A12	4	Timothy Yu	C14	2	Ryan J. Hong
A13	1	Nathan Abell	C15	3	Ian Hoskins
A14	2	Oana Nicoleta Antonescu	D01	1	Rosanna Jiang
A15	3	Anna Axakova	D02	2	David Jordan
B01	1	Dustin Baldridge	D03	3	Julian Jude
B02	2	Matthew Berg	D04	4	Nisha D. Kamath
B03	3	Christopher Bidlack	D05	1	Justin Kinney
B04	4	Maximilian Billmann	D06	2	Divya Kriti
B05	1	William Buchser	D07	3	Kevin Kuang
B06	2	Angel Fernando Cisneros Caballero	D08	4	Jessica Lacoste
B07	3	Patrick Carmody	D09	1	Jing Liu
B08	4	Michael Chambers	D10	2	Warren van Loggerenberg
B09	1	Laura Chamness	D11	3	Kaiyue Ma
B10	2	Aditya Chawla	D12	4	Stefanie Maes
B11	3	Steven Chen	D13	1	Larissa Milano
B12	4	Jishnu Das	D14	2	Molly Monge
B13	1	Natalie DeForest	D15	3	Lara Muffley
B14	2	Kristine Fredlund Degn	E01	1	Ayesha Muhammad
B15	3	Philippe C. Despres	E02	2	Vanessa Nguyen
C01	1	Bingbing Duan	E03	3	Sofie Vincents Nielsen
C02	2	Alistair Dunham	E04	4	Karen Noguera

POSTER DAY INFORMATION
Presenter Room and Session

Poster Number	Session	Presenter Name	Poster Number	Session	Presenter Name
E05	1	Victoria Offord	F08	4	Cathy Smith
E06	2	Victoria Offord	F09	1	Adrine de Souza
E07	3	Curran Oi	F10	2	Taylor Sparring
E08	4	Anthony Oppedisano	F11	3	James D. Stephenson
E09	1	Ryan Otto	F12	4	Daniel Tabet, Atina Cote
E10	2	Yeonwoo Park	F13	1	Martin Thygesen
E11	3	Jaeda Patton	F14	2	Matteo Tiberti
E12	4	Marcin Plech	F15	3	Vasileios Voutsinos
E13	1	Nicholas Popp	G01	1	Omar Wagih
E14	2	Madhuri Pulijala	G02	2	Yue Wang
E15	3	Carmen Resnick	G03	3	Andrew Waters
F01	1	Gabriel Rocklin	G04	4	Chenchun Weng
F02	2	Sarah Roelle	G05	1	Michael Xie
F03	3	Neha S. Samant	G06	2	Hongxia Xu
F04	4	Jonathan Schlebach	G07	3	Farhat Zafar
F05	1	Daniel Schmidt	G08	4	Taraneh Zarin
F06	2	James Shepherdson	G09	1	Juannan Zhou
F07	3	Nidhi Shukla			

Keynote Speaker Abstracts

Douglas M. Fowler, Ph.D. (he, him)
Associate Professor
Department of Genome Sciences
University of Washington School of Medicine

What do we need to scale multiplexed assays to all genes in the genome?

Multiplexed assays of variant effect (MAVEs) could, in principle, catalog nearly every possible variant in every gene in the human genome. However, despite recent proliferation of different types of MAVEs, many genes remain beyond reach owing to lack of an informative, generalizable assay. For example, approximately 10% of genes encode secreted proteins, most of which are not amenable to current approaches. To address this shortcoming, we developed a mammalian cell surface display method and used it to interrogate nearly all possible missense variants in the secreted clotting protein Factor IX, associated with hemophilia. A second challenge is that most MAVEs measure molecular phenotypes like ligand binding or abundance, or quantify cell growth. We developed Visual Cell Sorting to enable multiplexed assessment of variant effects on more complex cell morphological and behavioral phenotypes. We applied Visual Cell Sorting to measure the effect of thousands of variants of Lamin A, a nuclear envelope protein related to several distinct diseases, on nuclear morphology. Thus, we have expanded the range of genes to which MAVEs can be applied and the richness of phenotypes they can assess.

Clare Turnbull, Ph.D.
Professor of Translational Cancer Genetics
Institute of Cancer Research, London

Conundrums and challenges of clinical variant interpretation

Many clinical actions taken in response to detection of a genomic variant are binary, for example prophylactic mastectomy or termination of a fetus. Underlying the simple dichotomous clinical decision, there lies a variant-specific deleteriousness and a corresponding variant-specific penetrance for disease. Clinical variant classification comprises the stitching together of a range of indirect observations in individuals, populations, related species and laboratory assays in order to infer this variant-specific penetrance. Which we translate clinically into the terms ‘pathogenicity’ or ‘benignity’. I shall explore principles, pitfalls and forthcoming challenges of clinical variant interpretation, presenting exemplar analyses we have undertaken in paradigms of genetic cancer susceptibility.

Speaker Abstracts

Alexandra Bendel
Friedrich Miescher Institute for Biomedical Research

Systematic mutational scanning of the entire human leucine zipper interaction network

A fundamental question of genetic research asks how genotypes are translated into phenotypes. A major determinant at the interface between genotypes and phenotypes are protein interaction networks (PINs) that define the functional organization of a cell and thus ultimately the organismal phenotype. Mutations within the coding sequence of a gene can result in altered interaction specificities of the resulting protein and its interaction partners and can thereby cause re-wiring of PINs.
Families of conserved interaction domains can be used as models for examining specificity in protein-protein interactions (PPIs) because they combine high levels of specificity between certain members and low levels of crosstalk with the others to ensure insulation of diverse cellular processes. However, the determinants of specificity for each member's interaction profile are still unknown. Therefore, it remains challenging to predict mutation effects on PPIs de novo.
We performed a deep mutational scanning of the entire human basic leucine zipper (bZIP) domain family by measuring the effect of all possible single-amino acid substitutions in each of the 54 human bZIP domains on their ability of forming heterodimers with any of the 54 wild-type copy, homodimers with the same mutated copy, and on their protein abundance. The gathered data enables us to train deep learning models that capture the sequence-function relationships in bZIPs PPIs, in order to predict function (i.e., PPI) from sequence. This will allow us to functionally annotate germline or somatic variants reported in databases of disease-associated variants, with the aim to treat and prevent complex diseases. Moreover, it will provide us with synthetic biology tools to design novel bZIPs with potential applications in research and medicine. Finally, it will also deepen our insights into the evolution of protein interaction domains.

Munmun Bhasin
Molecular Biophysics Unit, Indian Institute of Science

Structure and stability predictions from saturation mutagenesis

Mutational scanning can be used to probe effects of large numbers of point mutations on protein function. Positions affected by mutation are primarily at either buried or at exposed residues directly involved in function, the latter are hereafter designated as active-site residues. In the absence of prior structural information, it has not been easy to distinguish between these two categories of residues. We curated and analysed a set of twelve published deep mutational scanning datasets. The analysis revealed differential patterns of mutational sensitivity and substitution preferences at buried and exposed positions. Prediction of buried-sites solely from the mutational sensitivity data was facilitated by incorporating predicted sequence-based accessibility values. The predictions were further validated by analysing the relationship between ligand binding and expression level of displayed protein using yeast surface display of a saturation mutagenesis library of the bacterial toxin CcdB. We coupled FACS and deep sequencing to reconstruct the binding and expression mean fluorescent intensity of each mutant. The reconstructed mean fluorescence intensity was utilised to distinguish between buried sites, exposed non-active-site positions, and exposed active-site positions. The method was extended to the receptor binding region of the spike protein of SARS-CoV-2, indicating its general applicability. We have also used mutational scanning to rapidly map the binding sites of the intrinsically disordered antitoxin MazE6 and delineate secondary structural features as well as residues important for binding to its partner MazF6. These studies highlight the ability of deep mutational scans to provide important structural and functional insights, in the absence of three-dimensional poor structural information; even for proteins or complexes with few homologs.

Speaker Abstracts

Benedetta Bolognesi, Ph.D.
Institute for Bioengineering of Catalunya (IBEC), Barcelona, Spain

Deep indel mutagenesis (DIM) reveals novel Aß mutations that accelerate amyloid formation and are likely pathogenic

Multiplexed assays of variant effects (MAVEs) guide clinical variant interpretation and reveal disease mechanisms. To date, MAVEs have focussed on a single mutation type - amino acid (AA) substitutions - despite the diversity of coding variants that cause disease. We have recently used Deep Indel Mutagenesis (DIM) to generate the first comprehensive atlas of diverse variant effects for a disease protein, the amyloid beta (Aß) peptide that aggregates into fibrils in Alzheimer's disease (AD) and is mutated in familial AD (fAD). We combined DIM to a selection method that reports on the rate of amyloid nucleation, i.e. the mechanism by which Aß fibrils form in the first place. The resulting comprehensive atlas identifies known fAD mutations and reveals that many variants beyond substitutions accelerate Aß aggregation and are likely to be pathogenic. Truncations, substitutions, insertions, single- and internal multi-AA deletions differ in their propensity to enhance or impair aggregation, but we identify likely pathogenic variants in all classes of mutations. Overall, mutations that increase the propensity of the peptide to aggregate into amyloid fibrils are highly enriched in the polar N-terminus of Aß, a region which remains unstructured in mature Aβ fibrils and - as a result - has been largely under-studied. This first comparative atlas highlights the importance of including diverse mutation types in MAVEs, while providing important mechanistic insights into amyloid nucleation.

Rick Boonen, Ph.D.
Leiden University Medical Center

High-throughput analysis of PALB2 missense variants: linking functional impact to breast cancer risk

Genetic testing for sequence alterations in the high-risk breast cancer susceptibility gene PALB2 frequently reveals missense variants of uncertain significance (VUS) for which the effect on protein function and consequently the associated cancer risk are unclear. We therefore developed a cDNA-based high-throughput approach in Palb2 knockout mouse embryonic stem cells that allows for the functional analysis of missense variants in human PALB2. Applying this high-throughput approach, we functionally investigated the impact of thousands of missense variants in the coiled coil (CC) and WD40 domains of PALB2 using PARP inhibitor sensitivity as a read-out for PALB2-dependent homologous recombination (HR). We identified several known, as well as new missense variants in PALB2 that impact protein function by affecting its interaction with BRCA1, its cellular localisation or its stability/expression level. Importantly, functional data analysis with available breast cancer risk data revealed that reduced HR as a result of patient-derived missense variants in PALB2 correlates with increased breast cancer risk. We conclude that our high-throughput analysis could aid in the clinical interpretation of PALB2 missense VUS, and may improve the clinical management of carriers of such variants.

Belinda Chang, Ph.D.
Professor, University of Toronto

Mutational landscapes and adaptive evolution of rhodopsin

Deep scanning mutagenesis approaches have proven to be useful for many applications in which high-throughput technologies can be applied to generate genotype to phenotype maps, but their application to evolutionary questions remains largely unexplored. Located within the photoreceptors of the vertebrate eye, rhodopsin is a prototypical G protein-coupled receptor which forms the basis for the initial step in vision, and is activated via absorption of a photon by an associated chromophore. The visual system has a surprisingly large dynamic range, and is highly specialized for the efficient sensing of light in extremely diverse light environments. How is this achieved at the molecular level? My laboratory investigates these questions via studies of natural and laboratory-created variants, combining both computational and experimental high-throughput approaches to better understand the mechanisms by which visual adaptations may occur.

Speaker Abstracts

Matt Coelho, Ph.D.
Wellcome Sanger Institute

Base editing screens map mutations affecting IFNg signalling in cancer

IFNg signalling underpins host responses to infection, inflammation and anti-tumour immunity. Mutations in the IFNg signalling pathway cause immunological disorders, haematological malignancies, and resistance to immune checkpoint blockade (ICB) in cancer, however the function of most clinically observed variants remain unknown. Here, we systematically investigate the genetic determinants of IFNg response in colorectal cancer cells using CRISPR-Cas9 screens and base editing mutagenesis. Deep mutagenesis of JAK1 with cytidine and adenine base editors, combined with pathway-wide screens, reveal loss-of-function and gain-of-function mutations with clinical precedence, including causal variants in haematological malignancies and mutations detected in patients refractory to ICB. We functionally validate variants of uncertain significance in primary tumour organoids, where engineering missense mutations in JAK1 enhanced or reduced sensitivity to autologous tumour-reactive T cells. By classifying > 300 missense variants altering IFNg pathway activity, we demonstrate the utility of base editing for mutagenesis at scale, and generate a resource to inform genetic diagnosis.

Willow Coyote-Maestas, Ph.D.
University of California San Francisco

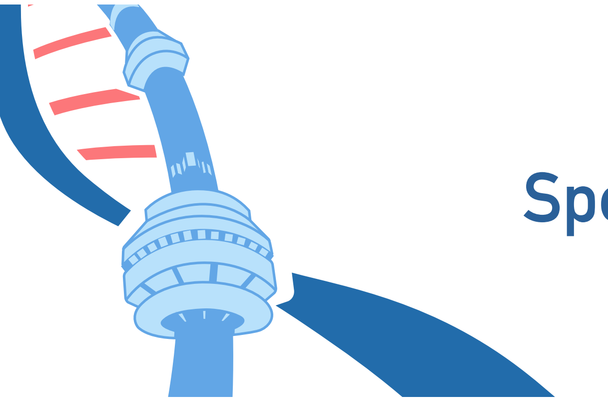
DIMPLE, a method for systematic, low bias, and affordable, missense and indel scanning libraries

Deep mutational scanning as a field has taken off resulting in numerous academic labs founded on the core technology, discoveries on fundamental protein biology, massive impacts on clinical interpretation, and companies developing new therapeutics. For DMS studies, mutational scanning library generation remains a major bottleneck due to incomplete libraries due to PCR bias, high cost of commercially available libraries, and methods that do not work on all targets. Furthermore, while indels underlie nearly 30% of disease causing mutations they are largely left out of existing libraries.
We developed a new method, Deep Indel Missense Programmed Library Engineering, to address these problems. Our libraries contain programmable indels simultaneously as all possible missense mutations within the same libraries and are based on oligo synthesis platforms that are affordable. Our libraries typically are comparable or cheaper than pcr based approaches, quicker to make, and only based on robust golden gate cloning. We anticipate that DIMPLE will be widely adopted by the field, enable new types of experiments, and further democratize mutational scanning.

Moez Dawood
Baylor College of Medicine

Saturation Genome Editing Reveals 10% of Missense SNVs in Functional Domains of PALB2 as Functionally Abnormal

Partner and Localizer of BRCA2 (PALB2) functions in homology-directed repair of double stranded DNA breaks, promotes nuclear localization and stability of BRCA2, and enables cell cycle checkpoint functions. Pathogenic variants in PALB2 have been implicated in breast, ovarian, and pancreatic cancer as well as Fanconi anemia. In 2021, PALB2 was elevated to the ACMG's 73 medically-actionable genes in which secondary findings should be reported to patient providers. However, with 99% of >1875 PALB2 missense single nucleotide variants (SNVs) classified as Variants of Uncertain Significance (VUS) in ClinVar, the majority of variants in PALB2 still remain uninterpretable and thus untenable for improving patient care. We set out to perform Saturation Genome Editing (SGE) to determine the variant effect of all possible 10,683 SNVs in PALB2. Thus far, we have ascertained variant effects for 2600+ SNVs focusing on the N and C terminus functional domains required for binding RAD51, BRCA1, and BRCA2. Thus far, 12% of missense SNVs result in decreased PALB2 function that compromises cellular viability. These SGE variant effects are 100% consistent with published orthogonal functional assay data as well as SNVs from ClinVar with a known clinical significance of pathogenic or benign.



Speaker Abstracts

Julia Domingo, Ph.D.
New York Genome Center

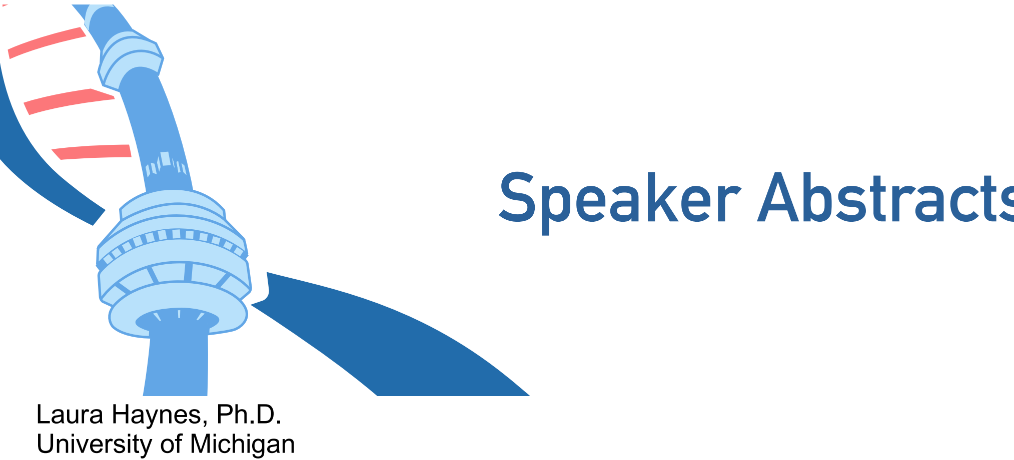
Global mapping of the energetic and allosteric landscapes of protein binding domains

Allosteric communication between distant sites in proteins is central to biological regulation but still poorly characterised, limiting understanding, engineering and drug development. An important reason for this is the lack of methods to comprehensively quantify allostery in diverse proteins. Here we address this shortcoming and present a method that uses deep mutational scanning to globally map allostery. The approach uses an efficient experimental design to infer en masse the causal biophysical effects of mutations by quantifying multiple molecular phenotypes—here binding and protein abundance—in multiple genetic backgrounds and fitting thermodynamic models using neural networks. We apply the approach to two of the most common human protein interaction domains, an SH3 domain and a PDZ domain, to produce comprehensive atlases of allosteric communication. Allosteric mutations are abundant with a large mutational target space of network-altering 'edgetic' variants. Mutations are more likely to be allosteric closer to binding interfaces, at Glycines and in specific residues connecting to an opposite surface in the PDZ domain. This general approach of quantifying mutational effects for multiple molecular phenotypes and in multiple genetic backgrounds should allow the energetic and allosteric landscapes of many proteins to be rapidly and comprehensively mapped.

Shawn Fayer
University of Washington

Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN

Clinical interpretation of missense variants is challenging because the majority identified by genetic testing are rare and their functional effects are unknown. Consequently, most variants are of uncertain significance and cannot be used for clinical diagnosis or management. Although not much can be done to ameliorate variant rarity, multiplexed assays of variant effect (MAVEs), where thousands of single-nucleotide variant effects are simultaneously measured experimentally, provide functional evidence that can help resolve variants of unknown significance (VUSs). However, a rigorous assessment of the clinical value of multiplexed functional data for variant interpretation is lacking. Thus, we systematically combined previously published BRCA1, TP53, and PTEN multiplexed functional data with phenotype and family history data for 324 VUSs identified by a single diagnostic testing laboratory. We curated 49,281 variant functional scores from MAVEs for these three genes and integrated four different TP53 multiplexed functional datasets into a single functional prediction for each variant by using machine learning. We then determined the strength of evidence provided by each multiplexed functional dataset and reevaluated 324 VUSs. Multiplexed functional data were effective in driving variant reclassification when combined with clinical data, eliminating 49% of VUSs for BRCA1, 69% for TP53, and 15% for PTEN. Thus, multiplexed functional data, which are being generated for numerous genes, are poised to have a major impact on clinical variant interpretation.



Speaker Abstracts

Laura Haynes, Ph.D.
University of Michigan

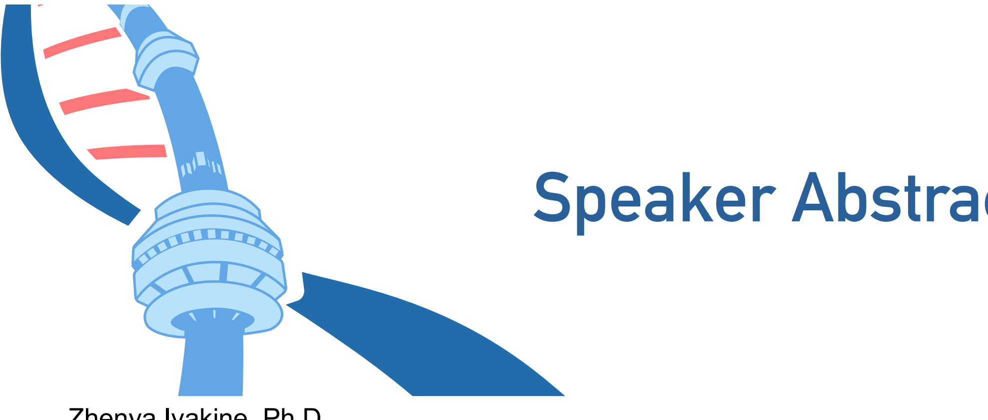
Fingerprinting serine protease inhibitor (SERPIN) specificity

Serine protease inhibitors (SERPIN) are a protein superfamily whose members are found across all kingdoms of life and are encoded by over 30 genes in humans that regulate diverse processes, including blood coagulation/fibrinolysis, the immune response, inflammation, and extracellular matrix remodeling, by specifically and irreversibly inhibiting serine proteases. In their active form, SERPINs exist in a metastable state in which a reactive center loop (RCL) extends from the globular core of the protein which serves as a bait for the target protease. When the protease attempts to cleave the peptide bond between the P1 and P1' residues of the RCL, the SERPIN and protease become covalently linked when, prior to hydrolysis of the acyl intermediate, the RCL inserts into the SERPIN's central beta-sheet resulting in a pole-to-pole translocation and structural rearrangement of the protease. This process renders both SERPIN and protease no longer functional. SERPIN specificity for its target protease has long been assumed to be driven primarily by the amino acid sequence of the RCL. However, in the present study we demonstrate that determinants of protease specificity are dispersed throughout the SERPIN's sequential and structural space. Plasminogen activator inhibitor-1 (PAI-1) is a 379 amino acid prototypic member of the SERPIN family whose canonical function is to inhibit tissue-type and urokinase-like plasminogen activators (tPA and uPA, respectively). A phage displayed PAI-1 mutational library was generated using error prone PCR that consists of $6.6x10^6$ independent clones with an average of 3-4 amino acid mutations per clone and containing 6,388 of the 7,201 possible missense mutations. The library was subsequently panned for inhibitory activity against its canonical enzyme target uPA, the procoagulant proteases thrombin and factor XIIa, as well as TMPRSS2, which primes influenza and SARS-CoV viruses for cellular entry and has been reported to be inhibited in vivo by PAI-1. More than 400 million PAI-1 DNA sequences from the input and selected libraries were determined by next generation DNA sequencing. By comparing the amino acid substitutions present in the input library to those in the selected libraries, unique patterns of enrichment and depletion of specific substitutions across PAI-1's sequence space were observed—underscoring that PAI-1 (and SERPIN) specificity is driven by exosite interactions and structural features distal from the RCL. Differences in the PAI-1 mutational landscape for each of the proteases also roughly correspond with the evolutionary diversity of the serine proteases studied, suggesting the coevolution of SERPINs with serine proteases. Finally, the results of this study form the foundation for engineering PAI-1 variants with specificity toward other non-canonical proteases.

Matthew Hurles, DMedSci FRS
Head of Human Genetics, Wellcome Sanger Institute

Saturation genome editing of DDX3X clarifies pathogenicity of germline and somatic variation

Loss-of-function of DDX3X is a leading cause of neurodevelopmental disorders (NDD) in females. DDX3X is also a somatically mutated cancer driver gene proposed to have tumour promoting and suppressing effects. We performed saturation genome editing of DDX3X, testing in vitro the functional impact of 12,776 nucleotide variants. We identified 3,432 functionally abnormal variants, in three distinct classes. We trained a machine learning classifier to identify functionally abnormal variants of NDD-relevance. This classifier has at least 97% sensitivity and 99% specificity to detect variants pathogenic for NDD, substantially out-performing in silico predictors, and resolving up to 93% of variants of uncertain significance. Moreover, functionally-abnormal variants could account for almost all of the excess nonsynonymous DDX3X somatic mutations seen in DDX3X-driven cancers. Systematic maps of variant effects generated in experimentally tractable cell types have the potential to transform clinical interpretation of both germline and somatic disease-associated variation.



Speaker Abstracts

Zhenya Ivakine, Ph.D.
Hospital for Sick Children Research Institute

All you need to know about Saturation Prime Editing

The advent of new genetic technologies and high-throughput methods has opened the doors to an era of non-hypothesis-driven biomedical research. We are now able to ask questions on a scale not previously imagined, with one such question being - what is the functional impact of every possible single base change in a gene of interest? Existing methods to answer this question include overexpression studies, as well as base editing screens and HDR-based saturation gene editing. We recently demonstrated a novel platform for this work, termed saturation prime editing (SPE), in which we harnessed (1) locus haploidization, (2) CRISPR prime editing, and (3) a FACS-based functional assay to elucidate the functionality of hundreds of single nucleotide variants in the NPC1 gene (Erwood et al., 2022). This talk will explore each of these three core aspects of our SPE platform in detail, as well as alternative functional assays and other possible applications of SPE. Applications of saturation gene editing using platforms like SPE have the potential to not only understand effects of genetic variants at scale, but also to make significant contributions to the development precision-based treatment strategies.

Jacob Kitzman, Ph.D.
Assistant Professor, University of Michigan

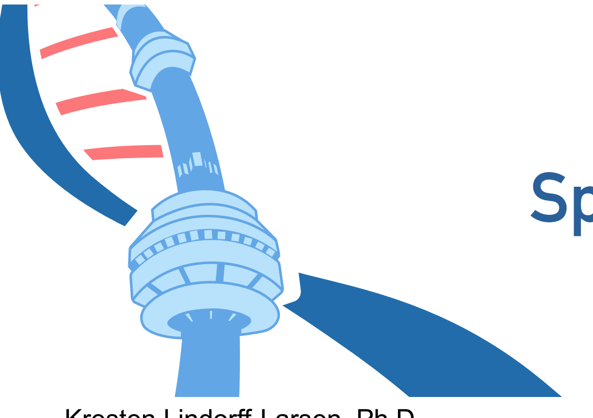
Checking on the spellcheckers: deep mutational scans of mismatch repair factors and beyond

It remains a daunting challenge to distinguish benign rare variants from those which damage gene function and contribute to disease risk, even for genes subjected to decades of biochemical and genetic study. I will share our ongoing efforts to comprehensively measure variant functional effects in several such long-studied factors involved in DNA mismatch repair (MMR). Inherited loss-of-function mutations in this key genome integrity pathway causes Lynch Syndrome, an early-onset colorectal and gynecological cancer predisposition affecting ~1:300 individuals worldwide. We have applied deep mutational scanning to generate functional effect maps for >98% of all possible missense variants in the two most-frequently mutated genes in LS, MSH2 and MLH1. I will describe the validation and active use of these maps in clinical variant interpretation pipelines. Going beyond single variant interpretation, we have extended this screen to systematically interrogate pairs and higher-order combinations of mutations, identifying a large population of variants which epistatically interact to sensitize or, conversely, buffer MMR factors against functional perturbation. Finally, I will present a complementary approach using massively parallel minigene assays to systematically identify variants which disrupt proper mRNA splicing, which we are applying to MMR genes and beyond. These approaches provide a scalable source of functional evidence to assist clinical variant interpretation, and starting point for future mechanistic studies.

Christian R Landry, Ph.D.
Professor, Universite Laval

Navigating protein fitness landscapes in multiple dimensions

Proteins could have as many fitness landscapes as they have functions. The relationship between mutations and the performance of any two functions could be positively or negatively correlated, making evolution on one landscape contingent on evolution on the other. Understanding the evolution of proteins therefore requires that we understand how mutations affect the multiple dimensions of the landscapes of protein functions. Here, I will share our recent work on the deep mutational scanning of two drug resistance enzymes in which we examine more than one dimension at a time. In one project, we examine the trade-off between resistance to an antifungal drug and activity on the canonical substrate of the drug target. We show that the trade-off between the two phenotypes is very steep and is shaped by the difference in dose response functions of drug resistance and nutrient use. The two fitness functions are nearly perfectly negatively correlated, which means no single mutation can lead to resistance while also maintaining growth in conditions that require this enzyme. In the second project, we examine how expression level affects the fitness landscape of an antibiotic resistance protein. By measuring the effect of all single mutations at optimal and several suboptimal expression levels, we show that optimal expression can mask the effects of many deleterious mutations on enzyme function. Surprisingly, some beneficial mutations also appear to be masked at optimal expression level. Protein expression level is therefore an important factor shaping the fitness landscape of proteins. These two studies illuminate how complex protein evolution is in heterogeneous growth and cellular environments.



Speaker Abstracts

Kresten Lindorff-Larsen, Ph.D.
Professor, University of Copenhagen

Understanding the origins of loss of protein function using analyses of protein stability and conservation

We have used computational and experimental approaches to determine the consequences of missense variants in proteins, with the aim of using such models both for diagnosing genetic diseases, and for providing mechanistic insight into disease. In particular, we have focused on the effects of individual amino acid changes on protein folding and stability, linking biophysical calculations with protein degradation and abundance in cells. By examining a range of proteins and diseases we have found that loss of stability and proteasomal degradation is a common origin for loss of protein function, and that predictions of changes in thermodynamic protein stability are useful to assess the pathogenicity of genetic variation. I will discuss these ideas using recent examples from our laboratories.
At the same time, our work has also revealed areas where our understanding and ability to predict the effect of amino acid changes is still imperfect. I will discuss how we are using sequence analyses and assays of variant effects experiments to understand the origins of loss of function, thus paving the way for more accurate biophysical models and machine learning methods for use in personalized medicine.

Debora Marks, Ph.D.
Associate Professor of Systems Biology
Harvard Medical School

Predicting the effects of genetic variation and designing biotherapeutics with neural machines

There's now an amazing opportunity to accelerate discovery across important 21st century challenges by using computation tightly coupled to biological experiments and clinical medicine. I will describe some recent approaches from my lab for these challenges where we have developed new machine learning methods that can exploit the enormous natural sequence diversity and our ability to synthesize DNA at scale. To demonstrate the power of these new approaches I will present recent work predicting the effects of human genetic variation on disease, anticipation of viral escape from the host immune system for vaccine design, protein design for enzyme optimization and smart antibody libraries

Troy McDiarmid, Ph.D.
University of Washington School of Medicine

Identification of enhancers to rescue haploinsufficiency in neurodevelopmental disorders via a CRISPRa-QTL approach

Neurodevelopmental disorders encompass a broad spectrum of conditions including Autism Spectrum Disorder and Intellectual Disability. As a result of rapid advances in clinical sequencing technology, thousands of variants in hundreds of genes have recently been definitively associated with an increased risk for developing neurodevelopmental disorders. For the majority of cases where a genetic etiology is known, the pathogenic mechanism is thought to be haploinsufficiency, having only one functional gene copy. Recent studies using cis regulation therapy (CRT), whereby the existing functional copy is upregulated using CRISPR activation (CRISPRa) targeting its regulatory elements, have shown to be a viable treatment option. However, to date only a handful of genes and regulatory elements have been tested due to limitations in throughput and lack of a detailed understanding of the regulatory circuitry controlling neurodevelopmental disorder risk genes. Here, we develop CRISPRaQTL screening, a high-throughput framework that combines CRISPRa perturbations with single-cell RNA sequencing to identify regulatory elements capable of potently and specifically upregulating target genes. We present results for a 494 guide, 40,000 cell experiment targeting known enhancers and putative promoters of neurodevelopmental disorder risk genes in K562 cells. CRISPRaQTL is capable of capturing target gene upregulation from single cell transcriptomes when targeting both promoters and enhancers. Current experiments are aimed at moving this CRISPRaQTL framework into iPSC-derived human neurons and scaling screening efforts to a comprehensive list of all putative regulatory elements controlling a list of all high-confidence neurodevelopmental disorder risk genes. This work establishes CRISPRaQTL as a scalable single cell framework to screen promoters and enhancers, identifies multiple functional guides targeting neurodevelopmental disorder risk gene promoters for use in CRT, and contributes to our understanding of the regulatory circuitry of neurodevelopment.

Speaker Abstracts

Alex McDonnell, University of Edinburgh

Probing PAX6-DNA Interactions Using High-Throughput Yeast One-Hybrid Coupled with Deep Mutational Scanning

PAX6 is a highly conserved transcription factor essential for the correct development of the central nervous system, the pancreas, and the eye. Heterozygous deletions, nonsense and frameshift mutations are generally well characterised as causing aniridia, while most missense variants produce a broad range of other discrete ocular pathologies. The interplay between PAX6 and its DNA targets is complicated by multiple functionally distinct subdomains, co-factors, high binding site promiscuity, and divergent spectra of disease phenotypes. Knowledge of the contributions made to binding by each residue and the impact of missense variants is key for understanding the role of PAX6 mutations in disease. Current methods of exploring this interaction grammar have been limited by relatively low-throughput techniques that are resource intensive and can only feasibly be performed on a handful of residues.

Here I used a combination of yeast one-hybrid and deep mutational scanning in competitive growth assays to probe the functional consequences of almost all the possible single amino acid variants in the paired domain of PAX6. Initial findings show clustered regions of the paired domain that are sensitive to mutation, that correlate broadly with secondary structural elements. Additionally, residues that contribute to DNA binding are overwhelmingly detrimental, and generally intolerant to mutation irrespective of the type of substitution. Finally, most clinically identified pathogenic variants appear to disrupt binding, with the exception of certain variants at the termini of the paired domain, and variants located in the linker region. The latter region also demonstrates a consistent observation of improved binding at specific residues across multiple substitutions. It is hoped that this deep mutational scan of PAX6 will aid in the modelling of existing and novel variants, and in the development of in-silico methods for pathogenicity prediction.

Sophie Moggridge, University of Washington

Interrogating protein variant dysfunction using multidimensional proteomics

Protein variants caused by single amino acid substitutions contribute to a large portion of human diseases. The molecular mechanisms of these diseases are complex because substitutions can have varied effects on protein properties such as stability, activity, conformations, localizations, and interactions. High-throughput mutagenesis methods like deep mutational scanning1,2 have accelerated variant phenotyping. However, these methods do not measure protein directly and thus are limited in the protein properties they can measure. Conversely, classical biochemical assays allow direct measurement of protein properties but require protein purification and are low throughput. To bridge this gap, we developed mass spectrometry (MS) methods that enable high-throughput analysis of many protein variants simultaneously. Biochemical assays that assess protein properties such as solubility, stability, post translational modifications, binding etc. are MS-compatible and will facilitate multidimensional analysis of mutational libraries. Implementation of these assays will greatly enhance our understanding of how single amino acid mutations affect function, and will contribute to better classifying, understanding, and treating disease.
Here I present the application of proteomic methods to determine molecular consequences of disease causing variants of human phosphoglucomutase-1 (PGM1). PGM1 was selected because it is a cytosolic, monomeric, and highly soluble protein for which the molecular effects of many variants are known. We focused on 13 missense variants known to alter a fundamental biochemical properties of PGM1, and coupled this library with three biochemical assays informative of protein function with a mass spectrometry-based readout. Solubility, thermal stability, and small molecule binding assays were performed on the PGM1 library that was recombinantly expressed in E. coli. Mass spectrometric analysis enabled the detection of all variants from the same lysate sample. The selection assays were benchmarked against a previous data set of low-throughput readouts for individual variants. The MS-based solubility assay was more quantitative than the previous method and achieved improved resolution for classification and ranking of protein variant solubilities. With the thermal stability assay, we were able to determine melting temperatures for six variants (N38Y, Q41R, G330R, E377R, E388K, L516P) which could not previously be measured using thermal denaturation circular dichroism. Use of MS-based orthogonal biochemical assays facilitates mechanistic interpretation of variant effects such as with D62H, which has wild-type-like thermal stability but is poorly soluble. Lastly, to evaluate PGM1 variants for small molecule binding, we performed a thermal stability assay in the presence of glucose 1-phosphate.
Our results successfully recapitulate published data from low-throughput assays. We have demonstrated the utility of MS-based biochemical assays that measure fundamental protein properties for the purpose of mutational scanning. The modular biochemical selection assays will contribute multi-dimensional mechanistic understanding to both novel and pre-existing mutational libraries.

References: (1): Matreyek KA, et al. Nat Genet. 2018. (2): Findlay, Gregory M., et al. Nature 2018."

Speaker Abstracts

Holly Rees, Ph.D.
Beam Therapeutics

Base editing – a next-generation genome editing technology

DNA base editors are a next-generation genome editing tool, enabling precise, targeted DNA base changes. The development, characterization and application of base editors will be discussed.

Nidhi Sahni, Ph.D.
Associate Professor, MD Anderson Cancer Center

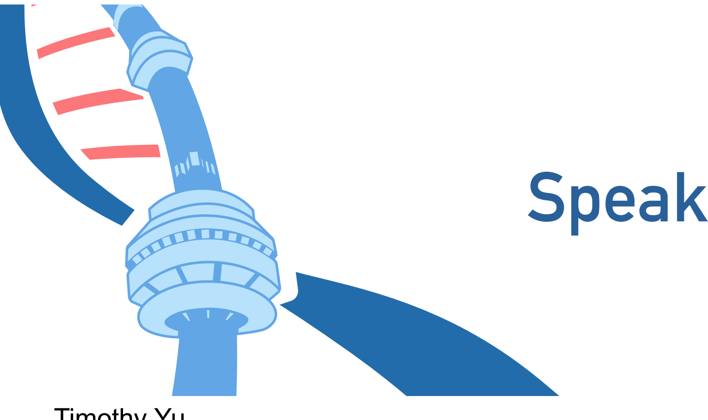
Functional variomics: Systematic annotation of somatic mutations and gene fusions in cancer

Proteins interact with other macromolecules in complex cellular networks for signal transduction and biological function. Our previous work in Mendelian disorders found a widespread phenomenon that disease-associated alleles often perturb distinct protein activities rather than grossly affecting folding and stability. In the context of cancer, the functional impact of the vast majority of somatic mutations remains unknown, representing a critical knowledge gap for implementing precision oncology. Here, we present the development of a high-throughput functional variomics platform consisting of efficient mutant generation, sensitive cell viability and drug response assays, and functional proteomic profiling of signaling effects for select aberrations. We apply the platform to annotate thousands of genomic aberrations, including point mutations, indels, and gene fusions, potentially doubling the number of driver mutations characterized in clinically actionable genes. Further, the platform is sufficiently sensitive to identify weak drivers. Our data are accessible through a user-friendly, public data portal. Our study will facilitate biomarker discovery, prediction algorithm improvement, and drug development.

Jochen Weile, Ph.D.
University of Toronto

Translating multiplexed variant effect assays into evidence codes for clinical interpretation

Variant effect (VE) maps produced via multiplexed functional assays promise to stem the ever growing tide of variants of uncertain significance (VUS). These maps excel at quantifying the impact of variants on the molecular function of a protein. However, the relationship between VE map scores and health outcomes is generally not linear. A relatively small decrease in protein function may already be sufficient to cause disease. For genes where pathogenic alleles typically act via a dominant negative mechanism such as Calmodulin, variants with a modest functional impact may be pathogenic while those that entirely eliminate protein function may be benign. Assays may have regional differences in sensitivity and precision, such as those previously observed between the catalytic and regulatory domains of MTHFR.
Clinical variant interpreters already use complex guidelines to determine evidence weights for functional data. Despite this complexity, the same weight is typically applied to all variants found damaging in a given assay and are based on simple thresholds. We propose a more data-driven and nuanced Bayesian approach that calculates the appropriate evidence weight for each variant within a VE map. In a two-step approach, we first find functions translating VE map score into a log-likelihood ratio (LLR) of pathogenicity. In the second step, LLRs can be expressed in terms of categorical evidence codes compatible with established variant interpretation guidelines. Separate LLR transformation functions can be determined for different regions of a map. We demonstrate the utility of our translation scheme for existing VE maps and evaluate reliability using leave-one-out cross-validation.



Speaker Abstracts

Timothy Yu
Fred Hutchinson Cancer Center

Measuring viral escape from polyclonal antibodies

Viral evolution is driven by mutations that enable viruses to escape antibody recognition. However, multiple mutations are often required to fully escape antibodies in human sera, which are polyclonal and can target several distinct epitopes. While deep mutational scanning has been used to measure the antigenic effects of all single mutations to viral proteins, our ability to predict antibody escape of multiply mutated viral variants remains remarkably poor. This is due to our limited understanding of the epistasis that transpires from antibody-epitope interactions, which cannot be empirically explored by single-mutant deep mutational scans. Here we address this shortcoming by harnessing recent advances in deep mutational scanning to measure the effects of combinations of mutations to influenza hemagglutinin on escape from antibodies in human sera. We then fit a biophysical model that captures the shape of epistasis to this data and can predict the escape potential of unseen, multiply mutated viral variants. We find that this approach can deconvolve an individual's unique polyclonal antibody response to reveal the locations of epitopes that are recognized by neutralizing antibodies and infer the specific mutations that impede this recognition.

Organizational Committee

Thank you to our Organizational Committee

David Adams
Anna Axakova
Atina Cote
Doug Fowler
Natasha Khalfan
Nola Klemfuss
Kevin Kuang
Uneku Haruna
Alex Hopkins
Zhenya Ivakine
Claire Major
Debora Marks
Lara Muffley
Dean Owen
Fritz Roth
Dayag Sheyhkharimli
Adrine de Souza
Lea Starita
Daniel Tabet
Olivia Taverniti
Jochen Weile
Jenny Xiao
Daniel Zimmerman

Thank you to Yvonne Shi for symposium artwork

Harnessing the power of DMS for drug development and design

Presenter: Nathan Abell (Octant, Inc)

Nathan Abell (1), Conor Howard (1), Erin Thompson (1)

1. Octant, Inc, Emeryville, CA, USA

Deep Mutational Scanning (DMS), methods using DNA synthesis, genome editing, and sequencing to create and measure all possible variants of a single protein, constitute a significant advance in our ability to probe functional protein-coding variation. These methods hold enormous potential for a variety of applications, including interpreting the pathogenicity of human variants, understanding the relationship between protein function and structure, and predicting the molecular basis of drug interactions. As a multiplexed platform company, we leverage DMS as a tool for improving drug discovery and design. This requires methods that are scalable, interrogate disease-relevant functions beyond protein expression and stability, and are sensitive enough to detect subtle effects on protein activity.

To more fully harness the potential of DMS, we have developed novel experimental and analytical techniques that allow us to not only directly assess human protein function as a consequence of every possible amino acid substitution in a single dish, but also to probe higher-level processes such as cell signaling and the effects of drug treatment. Our transcriptional reporters are engineered to assay complex and disease-relevant signaling cascades to obtain single-residue contributions to signaling biases, such as those of the G-protein-coupled receptors. In addition to technical improvements involving transcriptional amplification and barcode redundancy to improve power, we have also extended DMS to more complex experimental designs such as assessing small molecule chaperone rescue. All of these experimental advances have leaned heavily on a close, reciprocal relationship with statistical improvements necessary to address several current challenges in DMS analysis, such as variable barcode representation across samples, scaling, normalization, and detecting subtle activity changes. To this end, we developed a mixed-effects negative binomial regression model that does not require each barcode to be measured in every sample. We hierarchically structure the model fixed effects to estimate mutation consequences within each condition, in addition to comparisons across conditions, using a single regression. This framework maximizes the number of usable measurements from a DMS experiment, and thus results in more power to detect activity changes.

Collectively, these improvements have allowed us to use DMS to interrogate complex variant effects in the context of signaling bias and rescue by a molecular chaperone. We are also beginning to unravel protein structure-activity relationships (SAR) by mimicking coding variation with small molecule interactions, which has the potential to substantially improve drug design. This, combined with the fact that all possible human variants are inherently included in every experiment we execute, will facilitate novel approaches to drug development and testing. With Octant's multiplex technologies, we are building a high-throughput framework to harness the vast potential of DMS and finally realize the dream of personalized medicine.

Keywords: Deep mutational scanning, GPCR, structure-activity relationships

Interaction between phospho-regulated SLiMs and the ATG8 family: a case study on LC3B and optineurin

Presenter: Oana Antonescu (Danish Cancer Society Research Center)

Oana Antonescu (1), Valentina Sora (1,2), Matteo Lambrughi (1), Mukesh Kumar (1), Emiliano Maiani (1), Elena Papaleo (1,2)

(1) Cancer Structural Biology, Danish Cancer Society Research Center, Strandboulevarden 49, 2100, Copenhagen, Denmark

(2) Cancer Systems Biology, Section of Bioinformatics, Health and Technology Department, Technical University of Denmark, Lyngby, Denmark

The interaction between intrinsically disordered proteins and their binding partners is mediated by sequence motifs called Short Linear Motifs (SLiMs). Key features of SLiMs include their degeneration, which can be further amplified by their fine-tuned regulation by post-translational modifications (PTMs) such as phosphorylation. Non-canonical SLiMs, which bind with low affinity to their partners in their unmodified variants, can become stronger binders upon PTMs. To shed light on the details of this regulation and the related mechanisms on a class of SLiMs crucial in autophagy, we integrated biomolecular simulations with free energy calculations and high-throughput mutational scans in silico. We used the interaction between optineurin and LC3B as a test case and we achieved new insights into the conformational changes promoted by phosphorylation in optineurin. The computational predictions are being validated by high-density peptide arrays and immunoprecipitation studies of phosphorylated and mutated optineurin. The framework provided here can be applied more broadly to characterize or predict other LC3 interactors that are modulated by phosphorylation or other phospho-modulated SLiMs. Our research is funded by The Danish Council for Independent Research (DNRF125), Carlsberg Distinguished Fellowship (CF18-0314), and

NovoNordisk Founds in Bioscience and Basic Biomedicine (0065262). The calculations described in this poster were performed using DeiC (Computerome2 and Abacus) and PRACE-DECI 15th and 16th (on Archer and Archer2) HPC resources.

Keywords: SLiM, LIR, phosphorylation, IDP, PTMs

High-throughput analysis of PALB2 missense variants: linking functional impact to breast cancer risk

Presenter: Haico van Attikum (Leiden University Medical Center)

Rick A.C.M. Boonen (1), Sabine C. Knaup (1), Roberta Menafra (1), Dina Ruano (2), Noel F. de Miranda (2), Susan L. Kloet (1) and Haico van Attikum (1)

(1) Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, Leiden, 2333 ZC, The Netherlands
(2) Department of Pathology, Leiden University Medical Center, Einthovenweg 20, Leiden, 2333 ZC, The Netherlands

Genetic testing for sequence alterations in the high-risk breast cancer susceptibility gene PALB2 frequently reveals missense variants of uncertain significance (VUS) for which the effect on protein function and consequently the associated cancer risk are unclear. We therefore developed a cDNA-based high-throughput approach in Palb2 knockout mouse embryonic stem cells that allows for the functional analysis of missense variants in human PALB2. Applying this high-throughput approach, we functionally investigated the impact of thousands of missense variants in the coiled coil (CC) and WD40 domains of PALB2 using PARP inhibitor sensitivity as a read-out for PALB2-dependent homologous recombination (HR). We identified several known, as well as new missense variants in PALB2 that impact protein function by affecting its interaction with BRCA1, its cellular localisation or its stability/expression level. Importantly, functional data analysis with available breast cancer risk data revealed that reduced HR as a result of patient-derived missense variants in PALB2 correlates with increased breast cancer risk. We conclude that our high-throughput analysis could aid in the clinical interpretation of PALB2 missense VUS, and may improve the clinical management of carriers of such variants.

Keywords: PALB2, missense variants, high-throughput functional analysis, breast cancer risk

Towards a proactive missense variant atlas for the Autoimmune Regulator

Presenter: Anna Axakova (University of Toronto)

Anna Axakova (1,2,3), Amund H. Berger (4), Warren van Loggerenberg (1,2,3), Nishka Kishore (1,2,3), Atina Cote (1,2,3), Daniel Tabet (1,2,3), Aditya Chawla (1,2,3), Jochen Weile (1,2,3,5), Roujia Li (1,2,3,5), Marinella Gebbia (1,2,3), Eirik Bratland (4), Stefan Johansson (4), Frederick P. Roth (1,2,3,5).

1. The Donnelly Centre, University of Toronto, Toronto, ON, Canada
2. Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada 3. Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, ON, Canada
3. Department of Clinical Science, University of Bergen, Norway
4. Department of Computer Science, University of Toronto, Toronto, ON, Canada

Pathogenic variants of the Autoimmune Regulator (AIRE) gene cause the primary immunodeficiency disease Autoimmune Polyendocrine Syndrome Type 1 (APS-1; also called Autoimmune Polyendocrinopathy with Candidiasis and Ectodermal Dystrophy), which manifests as symptoms including chronic mucocutaneous candidiasis, hypoparathyroidism, and Addison's disease among other autoimmune symptoms. AIRE prevents autoimmunity by increasing the expression of tissue-specific genes in the niche of developing T cells, allowing for the elimination of emerging self-reactive T-cells before they can mature. Sequencing of AIRE is an increasingly common tool to diagnose APS-1, and earlier diagnosis can benefit patients. However, two thirds of clinical variants reported thus far are missense, and more than half of the missense variants found are "of unknown significance", limiting the diagnostic value of sequencing. Testing the function of variants in cultured cells can provide strong evidence for more informative variant classification, but these tests are carried out reactively, often months or years after a new variant is first observed in a patient. A proactive solution to this problem is to efficiently assess the function of all possible missense variants to generate a sequence-function map. Such maps can provide immediate and accurate evidence to diagnose disease, even for never-before-seen variants. We describe a validated scalable assay for testing AIRE missense variants in human cells, and progress towards a sequence-function map to enable more accurate clinical interpretation of AIRE missense variants. A validated sequence-function map of AIRE promises to provide proactive evidence for rapid and definitive genetic diagnosis of APS-1, and ultimately more positive patient outcomes.

Keywords: autoimmune disease, immunology, deep mutational scanning, missense variants, AIRE

Mutational Scanning to Determine Pathogenicity of Variants of Uncertain Significance in GLI2

Presenter: Dustin Baldridge (Washington University in St. Louis)

Dustin Baldridge (1), James Shepherdson (1), Barak Cohen (1)

(1) Washington University in St. Louis

Sequencing of patients with suspected monogenic disorders is a powerful approach for identifying molecular diagnoses, although the expected diagnostic yield continues to be around 30-40%. One major limitation is the abundance of Variants of Uncertain Significance (VUS) due to insufficient evidence for determining if variants are pathogenic or benign. High-throughput functional assessment of variant effects offers a scalable solution to this critical problem in genomic medicine via mutational scanning. This approach involves the simultaneous assessment of thousands of variants in a protein of interest, using a reliable, carefully calibrated, cell-based assay. To demonstrate the feasibility of establishing mutational scans for transcription factors, we have established an assay for the gene, GLI2, part of the sonic hedgehog signaling pathway. Pathogenic variants in GLI2 cause Culler-Jones syndrome, involving endocrine and skeletal patterning abnormalities. Using cells engineered with a GFP reporter and transduced with lentivirus expressing GLI2 variants, we conducted a SortSeq experiment, coupling fluorescence-activated cell sorting (FACS) to sequencing, to demonstrate the effect of these variants. Our initial results show perfect discrimination of known pathogenic and benign variants in GLI2. We use this system to assay all 170 missense variants in GLI2 that have been submitted to ClinVar, including about 100 VUS. The end result is a “look-up table” that gives clinicians and diagnostic laboratories high confidence functional evidence for reclassification of coding variants in our transcription factor of interest. We anticipate expanding this approach to include other genes in the sonic hedgehog pathway, suggesting a pathway-based approach for scaling mutational scanning to many genes in the genome.

Keywords: GLI2, sonic hedgehog

Complete energetic coupling maps of protein domains revealed by massively parallel double mutant cycles

Presenter: Toni Beltran (Centre for Genomic Regulation (CRG), Barcelona)

Toni Beltran (1) and Ben Lehner (1,2,3)

1 Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain 2 Universitat Pompeu Fabra (UPF), Barcelona, Spain 3 Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Protein folding is determined by a fine balance of thermodynamic forces that result in a relatively low net stability of the folded state. The stability of a protein fold depends on energetic interactions between residues in the folded state (“thermodynamic couplings”) resulting in a lower thermodynamic free energy than the constellation of other possible conformations.

Energetic couplings are experimentally measured as the change in free energy of a protein upon a double mutation minus the sum of the changes in free energy of the two mutations individually. This approach, introduced by Alan Fersht and co-workers in 1984, is known as a “double mutant cycle” (DMC). DMC studies, however, are typically focused on a small number of residues and a small number of mutants usually biased towards alanine, questioning the generalizability of their findings. Couplings have also been inferred from homologous protein sequences from extant species (“evolutionary couplings”), however, these are the result of selection on multiple biophysical traits, are averaged across many different sequence contexts and are incomplete. Thermodynamic couplings manifest in genetics as genetic interactions (epistasis) but the nonlinear relationship between folding/fitness and free energy changes makes quantifying energetic couplings from genetic interactions non-trivial.

Here, we combine the double mutant cycle approach with high throughput sequencing and an in vivo protein folding assay to generate the first complete energetic coupling maps for the folding of any model protein domains.

For all 1,156 pairs of residues in two domains - the all-beta Pin1 WW domain and the all-alpha villin headpiece 35 (VHP35) domain - we have quantified all 19x19 double mutant cycles in 4 genetic backgrounds, resulting in over 1,600,000 DMCs. Quantifying the mutational effects and interactions in different genetic backgrounds allows us to infer the underlying causal biophysical effects of mutations and energetic couplings (dddGs).

We will present the construction of these maps and the insights they provide into the energetics of proteins and their evolution.

Keywords: protein folding, double mutant cycles, energetic couplings, epistasis

Systematic mutational scanning of the entire human leucine zipper interaction network

Presenter: Alexandra Bendel (Friedrich Miescher Institute for Biomedical Research)

Alexandra Bendel(1), Dominique Klein(1), Marcella Franco(2), Kenji Shimada(1), Guillaume Diss(1)

1) Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
2) Boston Children's Hospital, Boston, US

A fundamental question of genetic research asks how genotypes are translated into phenotypes. A major determinant at the interface between genotypes and phenotypes are protein interaction networks (PINs) that define the functional organization of a cell and thus ultimately the organismal phenotype. Mutations within the coding sequence of a gene can result in altered interaction specificities of the resulting protein and its interaction partners and can thereby cause re-wiring of PINs. Families of conserved interaction domains can be used as models for examining specificity in protein-protein interactions (PPIs) because they combine high levels of specificity between certain members and low levels of crosstalk with the others to ensure insulation of diverse cellular processes. However, the determinants of specificity for each member's interaction profile are still unknown. Therefore, it remains challenging to predict mutation effects on PPIs de novo. We performed a deep mutational scanning of the entire human basic leucine zipper (bZIP) domain family by measuring the effect of all possible single-amino acid substitutions in each of the 54 human bZIP domains on their ability of forming heterodimers with any of the 54 wild-type copy, homodimers with the same mutated copy, and on their protein abundance. The gathered data enables us to train deep learning models that capture the sequence-function relationships in bZIPs PPIs, in order to predict function (i.e., PPI) from sequence. This will allow us to functionally annotate germline or somatic variants reported in databases of disease-associated variants, with the aim to treat and prevent complex diseases. Moreover, it will provide us with synthetic biology tools to design novel bZIPs with potential applications in research and medicine. Finally, it will also deepen our insights into the evolution of protein interaction domains.

Keywords: Protein interaction networks, specificity, Deep Mutational Scanning

Engineering and characterizing alanine mistranslating tRNAs for proteome-wide alanine scanning

Presenter: Matthew Berg (University of Washington)

Matthew D. Berg (1), Ecaterina Cozma (2), Julie Genereaux (2), Ricard A. Rodriguez-Mias (1), Christopher J. Brandl (2) and Judit Villén (1)

(1) Department of Genome Sciences, University of Washington, Seattle, WA, USA
(2) Department of Biochemistry, The University of Western Ontario, London, ON, Canada

While DNA sequencing has identified millions of natural variants that alter protein sequence, determining the functional impact of these variants remains challenging. Traditional mutagenic approaches are not scalable for millions of variants and high-throughput approaches such as deep mutational scanning are limited to investigating one protein per experiment. Therefore, approaches are needed that measure the effects of many mutations simultaneously on a proteome-wide level. Mistranslation, the mis-incorporation of an amino acid that differs from what is specified by the genetic code, can stochastically introduce amino acid substitutions at the translation level generating statistical proteomes composed of thousands of protein variants. Mistranslation can be achieved with engineered tRNA variants that are mis-aminoacylated with an amino acid that differs from what is specified by the tRNA anticodon. Previously, our lab established Miro, a proteomic technology that combines mistranslation with biochemical assays and mass spectrometry to quantify the functional impact of thousands of mutations on general protein properties like solubility, thermal stability, ligand binding, protein-protein interactions and post-translational modifications. In this work, our goal was to engineer and characterize alanine mistranslating tRNA variants for use with Miro to perform proteome-wide alanine scanning experiments. As the main tRNA-Ala identity element for aminoacylation is a G3:U70 base pair in the acceptor stem, we hypothesized that tRNA-Ala variants could be engineered to mis-incorporate alanine at non-alanine codons. We mutagenized the tRNA-Ala anticodon to all 60 possible non-alanine anticodons, expressed each tRNA variant individually on centromeric plasmids in yeast and measured growth as a proxy for mistranslation. Previously, we have shown that decreased growth correlates with mistranslation frequency. We identified 36 anticodon variants representing 15 amino acids that slowed growth. For two of the variants, slow growth was only observed upon overexpression, suggesting some tRNA-Ala variants are mistranslating at low levels or the substitution they create is not very toxic. We then used mass spectrometry to quantify mistranslation frequency, finding good correlation with the effect on growth for most substitutions. We also identified that variants with synonymous anticodons mistranslate different subsets of codons. For example, a tRNA-Ala variant with the proline UGG anticodon predominantly mis-incorporates alanine at cognate CCA codons, whereas a variant with the AGG anticodon mis-incorporates at CCC and CCT codons. Applying thermal stability assays to mistranslated proteomes, we then assayed the effect of alanine mutations on protein stability for thousands of proteins simultaneously. These tRNA variants will be useful tools in protein biology, enabling proteome-wide alanine scanning experiments and expanding the throughput of variant annotation.

Keywords: Mistranslation, tRNA variants, statistical proteomes, thermal stability, alanine scanning

Structure and stability predictions from saturation mutagenesis

Presenter: Munmun Bhasin (Molecular Biophysics Unit, Indian Institute of Science)

Munmun Bhasin (1), Shahbaz Ahmed (1), Soumyanetra Chandra (1), Gopinath Chattopadhyay (1) and Raghavan Varadarajan (1)

(1) Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India

Mutational scanning can be used to probe effects of large numbers of point mutations on protein function. Positions affected by mutation are primarily at either buried or at exposed residues directly involved in function, the latter are hereafter designated as active-site residues. In the absence of prior structural information, it has not been easy to distinguish between these two categories of residues. We curated and analysed a set of twelve published deep mutational scanning datasets. The analysis revealed differential patterns of mutational sensitivity and substitution preferences at buried and exposed positions. Prediction of buried-sites solely from the mutational sensitivity data was facilitated by incorporating predicted sequence-based accessibility values. The predictions were further validated by analysing the relationship between ligand binding and expression level of displayed protein using yeast surface display of a saturation mutagenesis library of the bacterial toxin CcdB. We coupled FACS and deep sequencing to reconstruct the binding and expression mean fluorescent intensity of each mutant. The reconstructed mean fluorescence intensity was utilised to distinguish between buried sites, exposed non-active-site positions, and exposed active-site positions. The method was extended to the receptor binding region of the spike protein of SARS-CoV-2, indicating its general applicability. We have also used mutational scanning to rapidly map the binding sites of the intrinsically disordered antitoxin MazE6 and delineate secondary structural features as well as residues important for binding to its partner MazF6. These studies highlight the ability of deep mutational scans to provide important structural and functional insights, in the absence of three-dimensional poor structural information; even for proteins or complexes with few homologs.

Keywords: deep sequencing, saturation mutagenesis, protein function, activity, stability, phenotype

Deep Mutagenesis Scan of SERPINC1 to Inform Variants Affecting Thrombosis Risk

Presenter: Christopher Bidlack (University of Michigan)

Christopher Bidlack (1), Krista Golden (2), Mary Underwood (2), and Karl Desch (1,2)

1. University of Michigan, Cellular and Molecular Biology Graduate Program, Ann Arbor, MI
2. University of Michigan, Department of Pediatrics, Ann Arbor, MI

Deep Mutagenesis Scan of SERPINC1 to Inform Variants Affecting Thrombosis Risk

Venous thromboembolism (VTE) includes both formation of pathogenic thrombi in the deep veins (deep vein thrombosis) and embolization of thrombi to the arterial circulation of the lungs (pulmonary embolism). VTE is a complex genetic disorder with an estimated heritability of 40-60% from family, sibling, and twin studies. Our lab recently conducted an exome sequencing study in ~400 VTE cases and found an excess of rare likely loss of function variants in three anticoagulant genes, PROC, PROS1 and SERPINC1. Antithrombin (AT), encoded by the SERPINC1 gene, is a critical anticoagulant protein that inactivates thrombin and factor Xa (FXa). Homozygosity for SERPINC1 deficiency is lethal, while partial absence leads to an increased risk of VTE. Damaging SERPINC1 variants can be categorized into two major types of antithrombin deficiencies: Type I deficiency due to decreased antithrombin concentrations in the bloodstream (quantitative deficiency), or Type II deficiency due to altered antithrombin function without necessarily affecting quantity (qualitative defects). Since AT deficiency is a strong risk factor for VTE, a comprehensive functional database of SERPINC1 variants could guide the interpretation of sequencing results in VTE patients and lead to a better understanding of AT structure and function.

To perform a deep mutational scan, we will transfect the variant SERPINC1 cDNA library into cells bearing the Bxb1-based recombinase “landing pad” system described by Matreyek et al. To identify variants causing AT retention, we will tag our segmented SERPINC1 libraries with eGFP reporter genes, and subject our cells to fluorescence-activated cell sorting (FACS) followed by next gen sequencing to identify variants associated with poor AT secretion. To identify variants affecting antithrombin function we will add a transmembrane domain of glycophorin B (GYPB) instead of the c-terminal eGFP tag. This will allow us to perform antithrombin functional screens using flow cytometry. Specifically, we will treat library cells with thrombin or factor Xa, two canonical ligands of antithrombin, to determine which variants disrupt AT function. Thrombin-Antithrombin (TAT) complex ELISA has confirmed the functionality of our SERPINC1-eGFP construct, and Sanger sequencing has confirmed successful mutagenesis of eight known SERPINC1 variants. Efforts are underway to separate variants in SERPINC1 associated with poor secretion from reference and other functional variants. Our next steps also involve full library cloning and screens. Any variants found to affect AT secretion or function will be used to make a functional map to find mutational clustering patterns.

Keywords: SERPINC1, Antithrombin, Coagulation, Venous Thromboembolism, VTE

A reference map for genetic interactions in a human cell

Presenter: Maximilian Billmann (Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn)

Maximilian Billmann (1,2), Michael Costanzo (3), A H M Mahfuzur Rahman (1), Catherine Ross (3), Amy Tong (3), Katherine Chan (3), Henry Ward (4), Michael Arreger (3), Matej Usaj (3), Kevin Brown (3), Brenda Andrews (3,5), Charles Boone (3,5), Jason Moffat (3,5) and Chad L. Myers (1,4)

(1) Department of Computer Science and Engineering, University of Minnesota-Twin Cities, Minneapolis, MN, USA
(2) Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn, Bonn, Germany
(3) The Donnelly Centre, University of Toronto, Toronto, ON, Canada
(4) Graduate Program in Bioinformatics and Computational Biology, University of Minnesota-Twin Cities, Minneapolis, MN, USA
(5) Department of Molecular Genetics University of Toronto, ON, Canada

A major focus of systems biology and genomic medicine is to link genotype to phenotype, yet accurately predicting disease states from genome sequence remains a major challenge. Genetic interaction networks in model organisms, principally yeast, have revealed how combinations of genome variants can impact phenotypes, and highlighted the importance of reference genetic networks for understanding gene function. We have used lessons learned from yeast to systematically map genome-wide genetic interactions using CRISPR-Cas9 in human cells. We performed 298 genome-wide screens using HAP1 query cell lines each carrying a defined loss-of-function mutation in one of 223 unique genes covering diverse bioprocesses, along with 39 screens in wildtype (wt) HAP1 cells used as a basis for robust scoring of genetic interactions. Overall, we screened approximately 4 million unique gene pairs for interactions, representing the largest effort to date to study double mutant phenotypes in isogenic human cells. We developed a computational pipeline to identify quantitative genetic interactions (qGI) from these data. We identified several unexpected statistical artifacts in loss-of-function screens including frequent interactions caused by variation between wt HAP1 screens and potential clonal effects of HAP1 cells harboring a loss-of-function mutation. We describe statistical elements of the qGI scoring pipeline designed to normalize these effects and insights we gained about interpreting phenotypes from CRISPR-Cas9 screens in the process. In total, we mapped approximately 47k negative (1.2% density) and 42k positive interactions (1.1% density) among the screened gene pairs. We also describe what we have learned about the topology of negative and positive genetic interactions in human cells, the power of genetic interaction profiles to define gene function across the genome, and their connections to other types of functional relationships, many of which are conserved from yeast to human cells. Finally, we observed that genetic interaction density correlates with disease associated genetic variants, and we believe our map can guide genetic background selection for variant assay development. In summary, we performed a large number of genome-wide CRISPR-Cas9 screens in specific genetic backgrounds and developed a computational pipeline to generate a first genome-wide reference genetic interaction network in human cells.

Keywords:

Deep indel mutagenesis (DIM) reveals novel Aß mutations that accelerate amyloid formation and are likely pathogenic

Presenter: Benedetta Bolognesi (Institute for Bioengineering of Catalunya)

Mireia Seuma (1), Ben Lehner (2,3,4), Benedetta Bolognesi (1)

1= Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028, Barcelona Spain 2= Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Doctor Aiguader 88, 08003, Barcelona, Spain 3= Universitat Pompeu Fabra (UPF), Barcelona, Spain. 4= ICREA, Pg. Lluís Companys 23, Barcelona 08010, Spain

Multiplexed assays of variant effects (MAVEs) guide clinical variant interpretation and reveal disease mechanisms. To date, MAVEs have focussed on a single mutation type - amino acid (AA) substitutions - despite the diversity of coding variants that cause disease. We have recently used Deep Indel Mutagenesis (DIM) to generate the first comprehensive atlas of diverse variant effects for a disease protein, the amyloid beta (Aß) peptide that aggregates into fibrils in Alzheimer's disease (AD) and is mutated in familial AD (fAD). We combined DIM to a selection method that reports on the rate of amyloid nucleation, i.e. the mechanism by which Aß fibrils form in the first place. The resulting comprehensive atlas identifies known fAD mutations and reveals that many variants beyond substitutions accelerate Aß aggregation and are likely to be pathogenic. Truncations, substitutions, insertions, single- and internal multi-AA deletions differ in their propensity to enhance or impair aggregation, but we identify likely pathogenic variants in all classes of mutations. Overall, mutations that increase the propensity of the peptide to aggregate into amyloid fibrils are highly enriched in the polar N-terminus of Aß, a region which remains unstructured in mature Aβ fibrils and - as a result - has been largely under-studied. This first comparative atlas highlights the importance of including diverse mutation types in MAVEs, while providing important mechanistic insights into amyloid nucleation.

Keywords: Amyloid, Alzheimer's, Insertion, Deletion

Pooled Imaged-based Phenotyping with iPSC-Derived Neurons Uncovers Subtle Phenotypes of Variants from Neurodegenerative Diseases

Presenter: William Buchser (Washington University in St Louis)

Jason Waligorski, Colin Kremitzki, Marianna Vakaki, Vinay Chandrasekaran, Graham Bachman, Purva Patel, Kylan Kelley, Samah Nour, Bay Johnson, Lina Ali, William Buchser

Washington University School of Medicine Department of Genetics and McDonnell Genome Institute FIVE@MGI

Neurodegenerative diseases are still a common form of dysfunction in people of all ages, but no treatment options are available. Diagnosis and eventual treatment hinges on understanding the genetic determinants of these conditions. These neurological diseases have several hallmarks, including defects in metabolism, aggregation/autophagy, RNA splicing, and trafficking. All cells have mechanisms to deal with these stressors, but with neurons' extended length, complex morphology, and inability to be reborn from stem cells makes them susceptible to these low-level long-term stressors.
To fully evaluate the 'functional health' of neurons, neuronal morphology and intracellular activity need to be considered. Most degenerative diseases don't include neuronal cell death (till the latest stages) and since neurons don't divide, standard CRISPR screens aren't useful. Similarly, reporter-based systems suffer from only having access to a single variable and not the neuron's overall health and function. When considering the options for how to perturb neurons in the setting of a large library of perturbagens (variants), there are exogenous and endogenous approaches. Exogenous addition of genes, while sometimes more practical, is less biologically relevant than endogenous alteration, a technical key to replicating patient-derived variants and their impact. Here, we used a human iPSC-derived motor neurons with endogenous mutations to examine metabolic and degenerative disease variants in a pooled manner using Raft-Seq. This pooled image-based screening system excels at studying detailed morphology of individual neurons, which are evenly separated by the specialized plate where they adhere and extend axons. Using Raft-Seq, we can physically capture thousands of individual motor neurons and process them through a single-cell NGS pipeline to make a direct connection between phenotype and genotype. It is even possible to capture the neurons alive for further direct study. Our results show that the system can answer a range of disease-relevant neuronal morphological questions. First, we used peripheral neuropathy CMT2a mutants of the mitochondrial regulator MFN2 to find mitochondrial morphology, localization, and axonal integrity defects. We use several other mutants associated with development, dementia, and ALS to show effects on nuclear stress and neurite outgrowth. These assays are all performed in the setting of a single-cell high-content pooled-screening system. The Raft-Seq system enables mutational scanning and patient-variant screening on the most relevant cell types—mature iPSC derived neurons—and allows for the study of highly physiological phenotypes in those cells.

Keywords: MPIA, DMS, RaftSeq, Motor Neurons, Patient Variants, Stem Cell, Human Neurons, Single Cell Sequencing

Expression level shapes a mutational fitness landscape

Presenter: Angel Fernando Cisneros Caballero (Université Laval)

Angel F. Cisneros (*,1), Isabelle Gagnon-Arsenault (*,1), Alexandre Dubé (1), Philippe Després (1), Pradum Kumar (1), Christian R. Landry (1)

Université Laval

Mutation is one of the driving forces of evolution. For protein-coding genes, mutations can affect the sequence of the encoded protein and modify its function. These changes can be deleterious or beneficial, ultimately contributing to how organisms adapt to the environment and to which genotypes become more abundant. As a result, much work has focused on studying mutational fitness landscapes, that is, the distribution of effects mutations in the protein sequence have on fitness. However, fitness landscapes have been widely recognized to change depending on the environment, which shows the need for studies testing mutations in the protein sequence and other factors at the same time. Here, we use deep mutational scanning (DMS) to survey the fitness landscape of a small tetrameric enzyme. We used the plasmid-encoded DfrB1 dihydrofolate reductase to systematically test the effect of all possible single mutations on the protein under different expression levels. As expected, we find that many mutations, often found in positions that show little variation in naturally occurring sequences, are deleterious regardless of the expression level. However, many other mutations that are deleterious at low expression levels become less deleterious when expression is increased. Interestingly, these mutants tend to cluster at specific positions within the protein sequence. We are currently doing computational analyses to try to identify which properties influence the propensity of a site's fitness landscape to change with expression level. Overall, our results suggest that the expression level of a protein is a major determinant of its fitness landscape. As a result, changes in expression level could play an important role in crossing fitness valleys by making certain protein regions less sensitive to mutations.

Keywords: Expression level, protein complexes, machine learning, deep mutational scanning

Variant assay by saturation mutagenesis and scRNA-seq

Presenter: Hongxia Xu/Dan Cao (Illumina, Inc.)

Hongxia Xu(1), *Dan Cao(1)*, Ken Jean-Baptiste(1), *Francois Aguet(1)*, Victor Quijano(1), Sandy Wong(1), Mohan Sun(1), Carmen Xiao(1), Kyle Farh(1)#

(1) Illumina Artificial Intelligence Laboratory, Illumina Inc.; San Diego, California, USA. *These authors contributed equally. # Correspondence should be addressed to kfarh@illumina.com

Understanding the function of variants of uncertain significance (VUS) has the potential to improve diagnosis for rare and common diseases. Instead of developing custom assays for each gene, single-cell RNA sequencing has the potential to be a generic assay readout for phenotyping variants (Ursu et al., 2022). Here we apply the scRNA-seq approach to characterize all possible 3546 single-nucleotide variants in the TP53 gene using a randomly barcoded lentiviral library. We demonstrate successful expression of all barcoded variants in cell culture by lentiviral transduction, and linked each cell to the expressed variant by integrating amplicon sequencing data with 10X data. We used machine learning to classify variants based on their scRNA-seq expression signature, and show that our results correlate with previously published bulk assay results (Giacomelli et al., 2018).

Keywords: saturation mutagenesis, variants of uncertain significance randomly barcoded lentiviral library, scRNA-seq

Coordination of -1 Programmed Ribosomal Frameshifting by RNA and Nascent Chain Features Revealed by Deep Mutational Scanning

Presenter: Patrick Carmody (Indiana University)

Patrick J. Carmody,1 Matthew H. Zimmer,2 Charles P. Kuntz,1 Haley R. Harrington,1 Kate E. Duckworth,1 Wesley D. Penn,1 Suchetana Mukhopadhyay,3 Thomas F. Miller III,2 and Jonathan P. Schlebach1

1) Department of Chemistry, Indiana University. 2) Division of Chemistry and Chemical Engineering, California Institute of Technology. 3) Department of Biology, Indiana University.

-1 Programmed ribosomal frameshifting (-1PRF) is a translational recoding mechanism that enables the synthesis of multiple polypeptides from a single transcript. In the alphavirus structural polyprotein, -1PRF is coordinated by a “slippery” sequence in the transcript, an RNA stem-loop, and a conformational transition in the nascent polypeptide chain. To characterize each of these effectors, we measured the effects of 4,530 mutations on -1PRF by deep mutational scanning. While most mutations within the slip-site and stem-loop reduce the efficiency of -1PRF, mutagenic effects upstream of the slip-site are far more variable. Coarse grained and atomistic molecular dynamics simulations of polyprotein biogenesis suggest many of these mutations alter pulling forces on the nascent chain by perturbing its interactions with the ribosome, the translocon, and the lipid bilayer. Finally, we provide evidence suggesting the coupling between cotranslational folding and -1PRF depends on the translation kinetics upstream of the slip-site. These findings provide unprecedented insights into how -1PRF is coordinated by the interplay between structural elements within the transcript and nascent polypeptide chain.

Keywords: Programmed Ribosomal Frameshifting, Membrane Protein Folding, Virology, Folding on the Ribosome

Saturation genome editing of CTCF to resolve functional variation associated with neurodevelopmental disorder

Presenter: Silvia Casadei (Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA)

Silvia Casadei (1,2), Moez Dawood (3,4,5,2), Nahum Smith (1,2), Ian Hill (1,2), Lea Starita (1,2)

(1) Department of Genome Sciences, University of Washington, Seattle, WA
(2) Brotman Baty Institute for Precision Medicine, Seattle, WA
(3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX
(4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX
(5) Medical Scientist Training Program, Baylor College of Medicine, Houston, TX

CTCF, or CCCTC-binding factor, encodes a transcriptional regulator protein with 11 highly conserved zinc finger domains. CTCF plays fundamental roles in transcriptional regulation and chromatin architecture maintenance. CTCF belongs to a growing number of genes involved in the regulation of chromatin structure for which genetic defects have been associated with human malignancies and developmental abnormalities.

De novo gene-disruptive mutations of CTCF are associated with intellectual disability and it is hypothesized that CTCF deficiency affects expression of enhancer-regulated genes (Gregor et al., 2013). Individuals with CTCF-dependent

neurodevelopmental disorder manifest different degrees of developmental delay or intellectual disability ranging from learning difficulties to severe cognitive impairment usually presenting in childhood. CTCF is also a tumor suppressor gene somatically mutated in endometrial cancer, leukemia and Wilms tumor with mutations resulting in haploinsufficiency through nonsense-mediated decay of mutant transcripts or loss-of-function missense mutations.

The mutational spectrum of CTCF in neurodevelopmental disorder encompasses large deletions, gene disruptive SNVs or indels and missense variants, with gene disruptive events localized mostly but not only in exons 3 and 4 and missense variants affecting exons 4 thru 10 encoding the 11 zinc finger domains. ClinVar currently reports 178 overall variants of CTCF of different molecular types and clinical consequences, of which 68 interpreted as pathogenic or likely pathogenic (P/LP), 47 as uncertain significance or conflicting interpretations and 63 as benign or likely benign (B/LB).

CTCF is essential for survival of the near haploid cell line HAP1. We aim to apply saturation genome editing (SGE) to measure the consequences of all possible SNVs in CTCF coding sequence on cell survival. We will use current literature and clinical databases to calibrate observed functional data on established interpretation of pathogenicity. Given the large proportion of unresolved variation within CTCF, we believe that SGE will provide important answers to the functional interpretation of clinically relevant variants as well as reveal functional insights into this important regulator of chromatin architecture.

Keywords: saturation genome editing, variant effect, CTCF, neurodevelopment disorders, chromatin architecture

Exploring the evolutionary arms race between human PKR and vaccinia K3L

Presenter: Michael Chambers (NIH-NHGRI)

Michael Chambers (1,2), Thomas Dever (3), Meru Sadhu (1)

1. Systems Biology & Genome Engineering Section, National Human Genome Research Institute, NIH
2. Department of Microbiology & Immunology, Georgetown University
3. Protein Biosynthesis Section, National Institute of Child Health and Human Development , NIH

The interface between interacting host and viral proteins can be a battleground in which genetic variants are naturally pursued. One such scenario is found between the mammalian innate immunity protein PKR (protein kinase R) and its poxvirus antagonist K3. Exploring the impact of missense variants in both PKR and K3 will highlight residues of evolutionary constraint and opportunity while also elucidating the mechanism by which human PKR is able to subvert a rapidly evolving antagonist. We reason that paired human PKR and vaccinia K3L variants can be characterized using a combinatorial high-throughput cloning approach and a yeast growth assay. In this assay, PKR activity suppresses yeast growth, which is restored if K3 successfully inhibits PKR. By tracking barcodes from sample timepoints in the assay we will be able to quantify and characterize the impact of each PKR and K3L variant combination, highlighting points of evolutionary constraint an opportunity for PKR and K3L. This strategy would allow us to scan a vast combinatorial space in a single experiment, providing details of the evolutionary fitness landscape of PKR and K3L as well as the ability of each protein to adapt to the other.

Keywords:

Impact of Topological Energetics on Pairwise Epistasis in Integral Membrane Proteins

Presenter: Laura Chamness (Indiana University)

Laura M. Chamness (1), Charles P. Kuntz (1), Andrew G. McKee (1), Wesley D. Penn (1), Christopher M. Hemmerich (2), Douglas B. Rusch (2), and Jonathan P. Schlebach (1)

(1) Department of Chemistry, Indiana University, Bloomington IN
(2) Center for Genomics and Bioinformatics, Indiana University, Bloomington IN

Many eukaryotic membrane proteins are prone to misfolding, which compromises their function at the plasma membrane. This is particularly true for mammalian gonadotropin-releasing hormone receptors (GnRHRs), which are G-protein coupled receptors involved in reproductive steroidogenesis. Evolutionary modifications to GnRHR have contributed to improper membrane integration of the transmembrane domains (TMDs). Mechanistic considerations suggest that these energetics could generate unique epistatic patterns, where certain mutations modify the effects of others. The role of membrane protein folding energetics and topogenesis in epistasis has yet to be explored. To investigate the types of epistatic interactions that arise from cotranslational folding energetics, we have utilized deep mutational scanning (DMS) to determine how the proteostatic effects of hundreds of GnRHR mutations are modified by secondary mutations that selectively perturb either cotranslational folding or post-translational folding. Our ongoing investigations aim to identify epistatic patterns that may result from topological energetics and evaluate their impact on membrane protein fitness.

Keywords: protein folding, epistasis, topogenesis, G-protein coupled receptors, integral membrane proteins

Deep Mutational Scanning of SERPINA1 Gene For Proactive Diagnosis of Alpha-1 Antitrypsin Deficiency

Presenter: Aditya Chawla (University of Toronto)

Aditya Chawla (1,2,3), Warren von Loggerenberg(1,2,3), Daniel Tabet(1,2,3), Marinella Gebbia(1,2), Anna Axakova(1,2,3), Atina Cote(1,2,3), Frederick P. Roth (1,2,3)

1 The Donnelly Centre, University of Toronto, Toronto, ON, Canada; 2 Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada;
3 Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, ON, Canada;

The protein alpha-1 antitrypsin, encoded by SERPINA1 and secreted by the liver, blocks the pro-inflammatory action of neutrophil elastase in the lungs. The rare disorder Alpha-1 Antitrypsin Deficiency (AATD) is characterized by lung inflammation, with some aggregating missense variants causing liver damage. AATD is often misdiagnosed, necessitating accurate clinical interpretation of SERPINA1 variants. Unfortunately, 45% of clinically-interpreted SERPINA1 variants have conflicting or uncertain interpretations. Because cell-based assays can provide strong evidence for clinical variant interpretation, we aim to test the impact of all possible alpha-1 antitrypsin missense variants, thus making functional evidence immediately available even for novel variants. A scalable assay, in which GFP is fused to the C-terminus of alpha-1 antitrypsin, exhibits lower GFP intensity for known loss-of-function variants and increased GFP intensity for known aggregation variants. We constructed a mutagenized library covering 80% of all possible missense changes. Then, we integrated this library into a stable cell line to integrate one variant per cell. Finally, we performed a large-scale assay through en masse transfection and fluorescence sorting to link genotype to phenotype in a high-throughput manner. The result of this study is a variant effect map covering ~70% of missense changes and their associated fitness effects. This map will assist in the diagnosis of VUS for this AATD and offer valuable information on the structural features of the AAT protein.

Keywords:

Identifying the mechanisms that underlie the genotype to phenotype link for a G protein-coupled receptor

Presenter: Steven Chen (University of Toronto)

Steven Chen (1), Jing Liu (1), Belinda S.W. Chang (1,2,3)

Department of Cell & Systems Biology (1), Department of Ecology & Evolutionary Biology (2), Centre for the Analysis of Genome Evolution & Function (3)

Mutations that perturb G protein-coupled receptors (GPCRs) are major disease drivers because organisms rely on GPCRs to sense and communicate with their environments. However, the mechanisms that link patient genotypes to phenotype for GPCR mutations remain poorly understood because signal detection and response mediated by GPCRs is known to be mediated by complex networks of molecular interactions. To determine how mutations affect these interactions, we engineered a signal transduction system using a reporter in budding yeast that allows us to measure the activation of human rhodopsin, a prototypical GPCR found in the retina of the vertebrate eye. We tested our system against patient-derived rhodopsin variants using flow cytometry and demonstrated that our measurements closely matched published light activation measurements made using other approaches. To investigate whether rhodopsin localization could explain these functional phenotypes, we fused each of the patient-derived rhodopsin variants to a red fluorescent protein (RFP), which is spectrally distant from the green fluorescent protein (GFP) used in our fluorescent reporter. We expressed these fusion proteins in our engineered yeast to measure rhodopsin activation and expression and to track its localization in the same system. To visualize rhodopsin expression and localization in three dimensions (3D), we expressed the patient-derived rhodopsin variants in our system and used confocal microscopy to perform a quantitative analysis of the reconstructed 3D structures. This revealed that variants that are known to misfold have higher cell surface expression heterogeneity while those that are not localize to the plasma membrane in a uniform distribution throughout the cell surface. To investigate whether these phenotypes are general features of rhodopsin variants, we synthesized a variant library that encodes for single amino acid substitutions at each site in a contiguous region in rhodopsin. This region is known to be a key component of receptor structure, and to undergo pronounced conformational shifts during receptor activation. It is one of the major hotspots where pathogenic variants are known to cluster. We expressed this library in our engineered yeast system and used flow cytometry to measure rhodopsin activation and per cell protein abundance to reveal that most variants are (i) highly expressed and active, (ii) highly expressed and inactive, or (iii) weakly expressed and inactive. Very few variants were detected to be weakly expressed and active which suggests that protein abundance is still important for receptor activation for many variants. Because our approach is amenable to high-throughput studies of rhodopsin function, our upcoming work will combine fluorescence-activated cell sorting and deep sequencing to determine the mechanisms that underlying the genotype to phenotype link for each of the variants in our library.

Keywords: Pathogenic variants, Transmembrane proteins, Signaling, Yeast genetics,

Base editing screens map mutations affecting IFNg signalling in cancer

Presenter: Matt Coelho (Wellcome Sanger Institute)

Matthew A. Coelho1,8, Emre Karakoc1,8, Shriram Bhosle1, Emanuel Gonçalves1,7, Thomas Burgold2, Sarah Cooper3,8, Chiara M. Cattaneo5,8, Vivien Veninga5,8, Sarah Consonni1,8, Cansu Dinçer1, Sara F. Vieira1,8, Freddy Gibson1, Syd Barthorpe1, Claire Hardy3, Joel Rein4, Mark Thomas4, Emile Voest5,6,8, Andrew Bassett3,8 & Mathew J. Garnett1,8,*

1 Translational Cancer Genomics, 2 Gene Editing and Cellular Research and Development, 3 Cancer, Ageing and Somatic Mutation, 4 Cellular Operations and Stem Cell Informatics, Wellcome Sanger Institute, Hinxton, UK 5 Netherlands Cancer Institute, Amsterdam, Netherlands 6 Oncode Institute, Utrecht, Netherlands 7 Current address: Instituto Superior Técnico, Universidade de Lisboa, 1049-001, and, INESC-ID, 1000-029, Portugal 8 Open Targets, Cambridge, UK * Corresponding author: mathew.garnett@sanger.ac.uk

IFNg signalling underpins host responses to infection, inflammation and anti-tumour immunity. Mutations in the IFNg signalling pathway cause immunological disorders, haematological malignancies, and resistance to immune checkpoint blockade (ICB) in cancer, however the function of most clinically observed variants remain unknown. Here, we systematically investigate the genetic determinants of IFNg response in colorectal cancer cells using CRISPR-Cas9 screens and base editing mutagenesis. Deep mutagenesis of JAK1 with cytidine and adenine base editors, combined with pathway-wide screens, reveal loss-of-function and gain-of-function mutations with clinical precedence, including causal variants in haematological malignancies and mutations detected in patients refractory to ICB. We functionally validate variants of uncertain significance in primary tumour organoids, where engineering missense mutations in JAK1 enhanced or reduced sensitivity to autologous tumour-reactive T cells. By classifying > 300 missense variants altering IFNg pathway activity, we demonstrate the utility of base editing for mutagenesis at scale, and generate a resource to inform genetic diagnosis.

Keywords: Cancer, Immunotherapy, Drug resistance, Base editing, CRISPR, Functional genomics, Cancer genetics

Multiplexed assays of low-density lipoprotein receptor (LDLR) variant effects

Presenter: Atina Coté (University of Toronto)

Atina G. Coté(1,2), Daniel Tabet(1,2,3), Ashyad Rayhan(1,2), Jochen Weile(1,2), Roujia Li(1,2) Iosifina Fotiadou(2,3), Jack Castelli(1,2,3), Nishka Kishore(1,2), Song Sun(1,2,3), Jennifer Knapp(1,2), Olivia Taverniti(1,2,3), Anna Axakova(1,2,3), Daniel Zimmerman(1,2,3), Kenneth Matreyek(4), Douglas M. Fowler(5), Calum MacRae(6,7), Robert A. Hegele(8), Frederick P. Roth(1,2,3,9)

1 The Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada 2 Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, ON M5G 1X5, Canada 3 Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada 4 Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH, USA 5 Department of Genome Sciences, University of Washington, Seattle, WA, USA 6 Brigham and Women's Hospital, Boston, MA, USA 7 Harvard Medical School, Boston, MA, USA 8 Department of Medicine, Robarts Research Institute, Western University, London, ON N6A 5B7, Canada 9 Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada

Familial hypercholesterolemia (FH), characterized by elevated plasma low-density lipoprotein (LDL) cholesterol, is generally caused by sequence variants in the LDL receptor gene LDLR, encoding the receptor primarily responsible for cellular LDL-cholesterol uptake. Almost 60% of clinically-interpreted missense variants of LDLR are currently either classified as variants of uncertain significance (VUS) or with conflicting interpretations. To address this issue and to discriminate between neutral and damaging LDLR variants, we developed multiplexed human-cell based assays of LDLR variant effects (MAVEs). Specifically, we used HeLa cells in which the endogenous LDLR gene is disrupted by CRISPR. Using LDL labeled with a pH-sensitive fluorescent dye (LDL-pHrodo), flow cytometry identifies deficient LDL-pHrodo uptake that is rescued by expressing a wild-type ORF genomically-integrated at an engineered 'landing pad'. Pathogenic (i.e. FH-causing) LDLR variants rescue the uptake defect less effectively than wild-type variants. Using fluorescence activated cell sorting (FACS), we identified the relative frequencies of variants within each pHrodo fluorescence intensity bin using both tiling-based and barcode sequencing approaches. Each assay yields functional scores that correlate well with pathogenicity. To provide sub-functional classification of variant effects, we developed a second MAVE in which LDLR variant-expressing cells are sorted using immunofluorescence to measure the cell surface abundance of LDLR. We provide evidence that the resulting large-scale maps of LDLR missense variant effects correlate with variant pathogenicity, and anticipate that they will be useful in assisting in early diagnosis and improved outcomes for FH patients.

Keywords:

DIMPLE, a method for systematic, low bias, and affordable, missense and indel scanning libraries

Presenter: Willow Coyote-Maestas (UCSF)

Christian Macdonald (1), David Nedrud, Patrick Rockefeller Grimes, James Fraser (1)(3), Willow Coyote-Maestas (1)(3)

(1) Department of Bioengineering and Therapeutic Sciences, UCSF, (2) Bio-techne, (3) Quantitative Biosciences Institute, UCSF

Deep mutational scanning as a field has taken off resulting in numerous academic labs founded on the core technology, discoveries on fundamental protein biology, massive impacts on clinical interpretation, and companies developing new therapeutics. For DMS studies, mutational scanning library generation remains a major bottleneck due to incomplete libraries due to PCR bias, high cost of commercially available libraries, and methods that do not work on all targets. Furthermore, while indels underlie nearly 30% of disease causing mutations they are largely left out of existing libraries.

We developed a new method, Deep Indel Missense Programmed Library Engineering, to address these problems. Our libraries contain programmable indels simultaneously as all possible missense mutations within the same libraries and are based on oligo synthesis platforms that are affordable. Our libraries typically are comparable or cheaper than pcr based approaches, quicker to make, and only based on robust golden gate cloning. We anticipate that DIMPLE will be widely adopted by the field, enable new types of experiments, and further democratize mutational scanning.

Note: This work is currently unpublished and so we are quite excited to get it in front of a large audience, present and share the method with the broader DMS field.

Keywords:

An integrative approach to uncover interaction-specific molecular phenotypes of coding and non-coding variants on underlying protein and regulatory networks

Presenter: Jishnu Das (University of Pittsburgh)

Trirupa Chakraborty1, *Swapnil Keshari1*, Jingyu Fan1, *Daniel McGrail2*, Song Yi3, Harinder Singh1^, Nidhi Sahni4^, Jishnu Das1^ [*=co-first ^=co-senior]

1 Center for Systems Immunology, Departments of Immunology and Computational & Systems Biology, University of Pittsburgh, Pittsburgh, PA 2 Department of Systems Biology, Lerner Research Institute, Cleveland Clinic, Cleveland OH 3 Departments of Oncology and Biomedical Engineering, the University of Texas at Austin, Austin, TX 4 Departments of Epigenetics & Molecular Carcinogenesis and Bioinformatics & Computational Biology, The University of Texas MD Anderson Cancer Center, Smithville, TX

Characterizing the functional impact of coding and non-coding variants in the context of underlying biological networks helps generate specific hypotheses regarding mechanisms of action of these variants. To this end, we first established an integrative experimental-computational pipeline to characterize the impact of coding variants on the underling protein interactome network. We found that Mendelian mutations typically do not disrupt protein stability, but rather result in interaction-specific perturbations. Further, the likelihood of disrupting interactions is primarily determined by the structural loci of these mutations, with interactions at the interface most likely to disrupt corresponding interactions. Variants at the same interface are more likely to be associated with the same disorder, and those at different interfaces are more likely to be associated with different disorders. We also characterized the impact of non-disease-associated population variants (ExAC SNPs) on corresponding interactions, and found that allele frequency is directly associated with the disruption rate. Even with common variants, ~10% disrupt corresponding interactions. Our framework allows us to hone in on interaction-specific molecular phenotypes that are tied to organismal phenotypes. For variants not directly associated with organismal phenotypes, we are also able to identify epistatic relationships that modulate disease risk. As part of the IGVF consortium, we are now expanding the same conceptual framework to characterize the impact of non-coding variants on dynamic gene regulatory networks coupled to a cis-regulome in primary human B cells. Our preliminary analyses suggest that GWAS-associated non-coding variants have extensive effects on open chromatin and other regulatory regions defined in the GM12878 B cell line. Many of these effects are likely tied to interaction-specific perturbations in the underlying gene regulatory network. Further and more detailed characterization of the impact of these variants as well as eQTLs on regulatory networks in primary human B cells will likely unveil actionable hypotheses regarding a wide range of autoimmune disorders, including RA, SLE, MS and IBD.

Keywords: IGVF, networks

Saturation Genome Editing Reveals 10% of Missense SNVs in Functional Domains of PALB2 as Functionally Abnormal

Presenter: Moez Dawood (Baylor College of Medicine)

Moez Dawood (1,2,3,4), Ian Hill (4,5), Aanchal Mehrotra (4,5), Silvia Casadei (4,5), Nahum Smith (4,5), Kara de Leon (4,5), Riza Daza (4,5), Riddhiman Garge (4,5), Abbye McEwen (4,5), Shawn Fayer (4,5), Lara Muffley (4,5), Jennifer Posey (2), James Lupski (1,2), Richard Gibbs (1,2), Jay Shendure (4,5,6), Douglas Fowler (4,5), Lea Starita (4,5)

(1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX
(2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX
(3) Medical Scientist Training Program, Baylor College of Medicine, Houston, TX
(4) Brotman Baty Institute for Precision Medicine, Seattle, WA
(5) Department of Genome Sciences, University of Washington, Seattle, WA
(6) Howard Hughes Medical Institute, Seattle, WA

Partner and Localizer of BRCA2 (PALB2) functions in homology-directed repair of double stranded DNA breaks, promotes nuclear localization and stability of BRCA2, and enables cell cycle checkpoint functions. Pathogenic variants in PALB2 have been implicated in breast, ovarian, and pancreatic cancer as well as Fanconi anemia. In 2021, PALB2 was elevated to the ACMG's 73 medically-actionable genes in which secondary findings should be reported to patient providers. However, with 99% of >1875 PALB2 missense single nucleotide variants (SNVs) classified as Variants of Uncertain Significance (VUS) in ClinVar, the majority of variants in PALB2 still remain uninterpretable and thus untenable for improving patient care. We set out to perform Saturation Genome Editing (SGE) to determine the variant effect of all possible 10,683 SNVs in PALB2. Thus far, we have ascertained variant effects for 2600+ SNVs focusing on the N and C terminus functional domains required for binding RAD51, BRCA1, and BRCA2. Thus far, 12% of missense SNVs result in decreased PALB2 function that compromises cellular viability. These SGE variant effects are 100% consistent with published orthogonal functional assay data as well as SNVs from ClinVar with a known clinical significance of pathogenic or benign.

Keywords: Saturation Genome Editing, SGE, Missense, Single Nucleotide Variants, SNVs, PALB2, BRCA2, BRCA1, VUS, Variants of Uncertain Significance, cancer, Fanconi anemia

Human gain-of-function variants in HNF1A confer protection from diabetes but independently increase hepatic secretion of multiple cardiovascular disease risk factors

Presenter: Natalie DeForest (UC San Diego)

Natalie DeForest (1,2), Babu Kavitha (3), Siqi Hu (1), Roi Isaac (1), Minxian Wang (4), Xiaomi Du (1), Camila De Arruda Saldanha (1), Jenny Gylys (1), Ruben Abagyan (5), Laeya Najmi (4,6), Viswanathan Mohan (7), Alnylam Human Genetics, AMP-T2D Consortium, Jason Flannick (4,8), Gina M. Peloso (9), Philip Gordts (1), Sven Heinz (1), Aimee M. Deaton (10), Amit V. Khera (4,11), Jerrold Olefsky (1), Venkatesan Radha (3), Amit R. Majithia (1)

1 Division of Endocrinology, Department of Medicine, University of California San Diego, La Jolla, CA, USA 2 Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA, USA 3 Department of Molecular Genetics, Madras Diabetes Research Foundation, ICMR Centre for Advanced Research on Diabetes, Affiliated to University of Madras, Chennai, India 4 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA 5 Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA 6 Center for Diabetes Research, Department of Clinical Science, University of Bergen, 5020 Bergen, Norway 7 Department of Diabetology, Dr. Mohan's Diabetes Specialties Centre, IDF Centre of Education, Chennai, India & Madras Diabetes Research Foundation, ICMR Centre for Advanced Research on Diabetes, Affiliated to University of Madras, Chennai, India 8 Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA 9 Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA 10 Alnylam Pharmaceuticals, Cambridge, MA, USA 11 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

Loss-of-function mutations in Hepatocyte Nuclear Factor 1A (HNF1A) are known to cause rare forms of diabetes and alter hepatic physiology through unclear mechanisms. In the general population, 1:100 individuals carry a rare protein-coding variant in HNF1A, most of unknown functional consequence. To characterize the full allelic series, we performed deep mutational scanning of 11,970 protein-coding HNF1A variants in human hepatocytes and clinical correlation with 553,246 exome-sequenced individuals. Surprisingly, we found ~1:5 rare protein-coding HNF1A variants found in the general population cause molecular gain-of-function (GOF), increasing the transcriptional activity of HNF1A by up to 50%. GOF in HNF1A conferred protection from type 2 diabetes (T2D) (OR=0.60, p=8.4 x 10-7), but not against coronary artery disease. Independently of T2D, increased hepatic expression of HNF1A promoted a pro-inflammatory and pro-atherogenic serum profile mediated in part by enhanced transcription of risk genes including PCSK9. In summary, ~1:300 individuals carry a GOF variant in HNF1A that protects carriers from diabetes but enhances hepatic secretion of metabolic disease risk factors.

Keywords: HNF1A, PCSK9, deep mutational scan, saturation mutagenesis, UK Biobank, gain-of-function, diabetes, inflammation, atherosclerosis, coronary artery disease

Examination of local and long-range effects as a result of cancer mutations in p53.

Presenter: Kristine Fredlund Degn (Technical University of Denmark)

Kristine Degn (1), Ludovica Beltrame (2), Freja Dahl Hede (1), Valentina Sora (1,2), Vincenzo Nicolaci (2), Marina Vabistevis (2), Kjeld Schmiegelow (3), Karin Wadt (4), Matteo Tiberti (2), Matteo Lambrughi (2), Elena Papaleo (1,2)*

(1) Cancer Systems Biology, Section for Bioinformatics, Department of Health and Technology, Technical University of Denmark, 2800, Lyngby, Denmark
(2) Cancer Structural Biology, Danish Cancer Society Research Center, 2100, Copenhagen, Denmark
(3) Department of Pediatrics and Adolescent Medicine, University Hospital Rigshospitalet, Copenhagen, Denmark
(4) Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Denmark

The complex context-dependent functional variability of tumor protein 53 (p53) is observable in the structural conformations of the protein. These conformational changes are mediated by the protein's binding partners and mutations of the protein itself, as it is likely to be found in tumor tissue. Some mutations impact protein stability, while others influence the protein functionality both in a local and long-range capacity. The importance of these alterations has previously been illustrated by our group when we described residues 207-213. Residues 207-213 are part of the DNA binding domain, which can recruit interactors, modulate DNA binding from long-range alteration, and finally modulate local interactions by post-translational modifications. Here, we aim to characterize >1000 somatic and germline mutations throughout the DNA binding domain with structure-based approaches to assess their impact on stability, local and allosteric effects on the residues in the 207-213 region. Initially, we conducted a high-throughput calculation of free energy with various energy functions, aiming to identify and maintain mutations with minor impact on the structural stability of the DNA binding domain according to the predictions. The mutations were analyzed in the context of their local effects: impact on DNA binding, dimerization, and intramolecular contacts with the 207-213 interaction interface, as well as their long-range effects. The long-range modulation of the 207-213 region was estimated using a coarse-grain model for allostery and molecular dynamics trajectories, subsequently analyzed by applying network theory. These predictions were further validated using enhanced sampling of the molecular simulation, allowing a higher coverage of the conformational space, all done to verify the effect and understand the specific mechanism of action. This study provides a thorough classification of cancer-related p53 mutations concerning structural stability, DNA-binding, and dimerization interface of the p53 cytosolic functions. Further, this study provides an in-depth mechanistic characterization of 15 cancer variants in the context of their interface and their recruitment effects. The workflow presented in this paper could be applied more broadly to study the effects of other p53 mutations' impact on different interfaces of protein-protein interaction or to study other specific regions of interest.

This work is part of Interregional Childhood Oncology Precision Medicine Exploration (iCOPE), a cross-Oresund collaboration between University Hospital Copenhagen, Rigshospitalet, Lund University, Region Skåne and Technical University Denmark (DTU), supported by the European Regional Development Fund. The result of this research has been achieved using the DECI-PRACE 15th and 16th HPC Grants on Archer and access to the Danish HPC Infrastructure Computerome2. The work is also supported by Danmarks Grundforskningsfond (DNR125).

Keywords: Allosteric Interactions, Therapeutic Target, Molecular Dynamics, Protein Stability, Protein Structure Network, Binding Affinity

Asymmetrical dose-responses shape the evolutionary trade-off between antifungal resistance and nutrient use

Presenter: Philippe C Després (Université Laval)

Philippe C Després (1,2,3,4), Angel F Cisneros (1,2,3,4), Emilie MM Alexander (1,2,3,4), Ria Sonigara (1,2,4,5), Cynthia Gagné-Thivierge (1,2,3,4,5), Alexandre K Dubé (1,2,3,4,5), Christian R Landry (1,2,3,4,5)

1. Département de Biochimie, de Microbiologie et de Bio-informatique, Faculté des Sciences et de Génie, Université Laval
2. Institut de Biologie Intégrative et des Systèmes, Université Laval
3. PROTEO, Le regroupement québécois de recherche sur la fonction, l'ingénierie et les applications des protéines, Université Laval
4. Centre de Recherche sur les Données Massives, Université Laval
5. Département de Biologie, Faculté des Sciences et de Génie, Université Laval

Antimicrobial resistance is an emerging threat for public health. The success of resistance mutations depends on the trade-off between the benefits and costs they incur. This trade-off is largely unknown and uncharacterized for antifungals. Here, we systematically catalog the effect of all amino acid substitutions in the yeast cytosine deaminase FCY1, the target of the antifungal 5-FC. We identify over 900 missense mutations granting resistance to 5-FC, a large fraction of which appear to act through destabilisation of the protein. The relationship between 5-FC resistance and growth sustained by cytosine deamination is characterized by a sharp trade-off, such that small gains in resistance universally lead to large losses in canonical enzyme function. We show that this steep relationship can be explained by differences in the dose-response function of 5-FC and cytosine. Finally, we generate mutants of the C. neoformans ortholog of FCY1 and show that their phenotypes are highly correlated with their S. cerevisiae equivalent. Our results provide a powerful resource and platform for interpreting drug target variants in fungal pathogens as well as unprecedented insights into resistance-function trade-offs.

Keywords: Drug resistance, Antifungal, Yeast, CRISPR-Cas9

Deep mutational scanning of entire human protein interaction domain families

Presenter: Guillaume Diss (Friedrich Miescher Institute for Biomedical Science)

Alexandra Bendel (1), Kenji Shimada (1), Sebastian Jaaks-Kraatz (1), Ben Lehner (2,3,4) and Guillaume Diss (1)

(1) Friedrich Miescher Institue for Biomedical Science (FMI), Basel, Switzerland. (2) Centre for Genomic Regulation (CRG), Barcelona, Spain. (3) Universitat Pompeu Fabra (UPF), Barcelona, Spain. (4) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

Protein interaction networks represent the backbone of the cell's functional organization. Their architecture relies on the specificity of these interactions. Mutations that alter specificity may rewire the network and the way it coordinates biological processes, and lead to diseases or functional innovations. Understanding how the sequence of a protein determines the specificity of its interaction profile thus represents an important step towards a model of the genotype-phenotype map. To address the sequence-function relationship in protein-protein interactions, we conducted a deep mutational scanning of the entire human basic leucine zipper (bZIP) interaction network. This interaction domain family, composed of 54 members in humans, has served as a model system for the study of protein interaction specificity. While the general rules of bZIP specificity are relatively well known, it remains impossible to predict interaction partners accurately. Using our established deepPCA system, we measured the effect of all single amino acid substitutions in all 54 members, on the interaction with all 54 wild-type members. This dataset of ~2M quantitative protein-protein interaction measurements will enable training a deep neural network that captures the sequence-function relationship in order to quantitatively predict interaction between bZIPs from sequence alone. I will present the dataset and the deep learning model, as well as preliminary data on two other families, basic helix-loop-helix and SH3 domains, for which we plan to collect 6 and 20M data points, respectively. We foresee that applying this strategy on a variety of domain families of different architecture will enable the development of general models that can predict the affinity of an interaction from the sequence of the two partners.

Keywords: protein-protein interaction; deep learning

Global mapping of the energetic and allosteric landscapes of protein binding domains

Presenter: Júlia Domingo (New York Genome Center (NYGC))

Andre J. Faure (1)*, Júlia Domingo (1,2)*, Jörn M. Schmiedel (1)*, Cristina Hidalgo-Carcedo (1), Guillaume Diss (1,3), Ben Lehner (1,4,5)

(1) Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Doctor Aiguader 88, 08003 Barcelona, Spain
(2) New York Genome Center (NYGC), New York, NY, US
(3) Friedrich Miescher Institute for Biomedical Research (FMI), Basel, Switzerland
(4) Universitat Pompeu Fabra (UPF), Barcelona, Spain
(5) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
• equal contribution

Allosteric communication between distant sites in proteins is central to biological regulation but still poorly characterised, limiting understanding, engineering and drug development. An important reason for this is the lack of methods to comprehensively quantify allostery in diverse proteins. Here we address this shortcoming and present a method that uses deep mutational scanning to globally map allostery. The approach uses an efficient experimental design to infer en masse the causal biophysical effects of mutations by quantifying multiple molecular phenotypes—here binding and protein abundance—in multiple genetic backgrounds and fitting thermodynamic models using neural networks. We apply the approach to two of the most common human protein interaction domains, an SH3 domain and a PDZ domain, to produce comprehensive atlases of allosteric communication. Allosteric mutations are abundant with a large mutational target space of network-altering 'edgetic' variants. Mutations are more likely to be allosteric closer to binding interfaces, at Glycines and in specific residues connecting to an opposite surface in the PDZ domain. This general approach of quantifying mutational effects for multiple molecular phenotypes and in multiple genetic backgrounds should allow the energetic and allosteric landscapes of many proteins to be rapidly and comprehensively mapped.

Keywords: Deep Mutational Scanning, allostery, binding affinity, stability, thermodynamics, neural networks

Functional dissection of the RNA polymerase trigger loop by deep mutational scanning

Presenter: Bingbing Duan (University of Pittsburgh)

Bingbing Duan (1), Chenxi Qiu (2), Sing-Hoi Sze (3), Craig Kaplan (1)

(1). Department of Biological Sciences, University of Pittsburgh; (2).Department of Genetics, Harvard Medical School ; (3)Department of Computer Science and Engineering, Texas A&M University

We have employed a deep mutational scanning “structural genetics” approach to understand RNA polymerase function. We exploit phenotypic profiling to dissect relationships among residues and how epistasis affects evolution within eukaryotic RNA polymerases. Eukaryotic multisubunit RNA Polymerases (Pol) I, II, III are structurally conserved though they have evolved to have their own regulation and produce different classes of transcripts. At the heart of these RNA polymerases is an ultra-conserved active site domain, the trigger loop (TL), that participates in transcription by switching between an open, catalytic-disfavoring state and a closed, catalytic-favoring state. Mutations in the Pol II TL residues affect substrate selection, catalysis and translocation, suggesting the TL residues are critical to every aspect of transcription elongation. Additionally, the TL’s function is shaped by interactions with adjacent protein domains. Our lab has observed complex functional residue-residue interactions in Pol II TL, indicating transcription activity is facilitated by a functional network within the TL and between the TL and adjacent domains. Furthermore, previous studies have found that identical mutations in a residue conserved between the Pol I and Pol II TLs yielded different biochemical phenotypes, implying even functions of conserved residues may be shaped by individually evolved enzymatic contexts (epistasis). In order to understand TL residue interactions that shape polymerase mechanisms, we have functionally dissected Pol II TL mutants and intra-molecular genetic interactions in a high throughput genetic phenotyping system. Through analysis of over 14000 alleles, representing single mutants, a subset of double mutants, and evolutionarily observed TL haplotypes, we have identified residue interactions within the TL and between TL and adjacent domains. Moreover, we are investigating functional residue coupling across evolution and identifying where epistasis within Pol II constraints TL residue identity and function. We are also extending this system to Pol I and III to comprehensively compare and contrast residue requirements across polymerase evolution. Our studies provide a powerful comparative system to understand the plasticity of RNA polymerase mechanisms.

Keywords: RNA Polymerase, trigger loop, deep mutational scanning, epistasis

Exploring amino acid functions in a deep mutational landscape

Presenter: Alistair Dunham (Wellcome Sanger Institute)

Alistair Dunham (1), Pedro Beltrao (2)

(1) Wellcome Sanger Institute, (2) IMSB ETH Zurich

Amino acids fulfil a diverse range of roles in proteins, each utilising its chemical properties in different ways in different contexts to create required functions. The repertoire of amino acid functions and the frequency at which they occur in proteins remains understudied. Measuring large numbers of mutational consequences, which can elucidate the role an amino acid plays, was prohibitively time-consuming until recent developments in deep mutational scanning. In this study, we gathered data from 28 deep mutational scanning studies, covering 6,291 positions in 30 proteins, and used the consequences of mutation at each position to define a mutational landscape. We demonstrated rich relationships between this landscape and biophysical or evolutionary properties. Finally, we identified 100 functional amino acid subtypes with a data-driven clustering analysis and studied their features, including their frequencies and chemical properties such as tolerating polarity, hydrophobicity or being intolerant of charge or specific amino acids. The mutational landscape and amino acid subtypes provide a foundational catalogue of amino acid functional diversity, which will be refined as the number of studied protein positions increases. The dataset also provides a good platform to study the relationship between substitutions in different context, helping to identify which measurements provide the most information about positional properties.

Keywords:

Massive combinatorial mutagenesis of protein cores and interactions for accurate sequence-to-activity prediction

Presenter: Albert Escobedo (Centre for Genomic Regulation (CRG))

Albert Escobedo (1), Ben Lehner (1,2,3)

1 - Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain. 2 - Universitat Pompeu Fabra (UPF), Barcelona, Spain. 3 - ICREA, Barcelona, Spain.

Accurate sequence-to-activity prediction would revolutionize protein engineering, drug development and clinical genetics. Two of the fundamental properties of most proteins are their stability and their binding affinity for other molecules. Towards the goal of accurate sequence-to-stability and sequence-to-affinity prediction, we are performing combinatorial mutagenesis to quantify the effects of individual mutations and their pairwise and higher order combinations on the stability and protein-binding affinity of model protein domains. This allows us to both ask questions about the genetic architecture of protein stability and affinity and to test the predictive performance of various computational methods for sequence-to-activity prediction, including biophysical models incorporating pairwise and higher order energy terms inferred from the data. To better understand and predict protein stability, we have mutagenised the cores of small globular domains quantifying the effects on stability of 80,000 different combinations of 5 different hydrophobic residues in 7 sites or 160,000 combinations of all 20 amino acids in 4 sites in a cell-based selection assay (Faure et al., Nature, 2022). Fitting thermodynamic models to this data allows us to infer the energetic effects of mutations and their pairwise and higher order energetic coupling terms and to evaluate different methods for predicting the effects of combinatorial sequence changes on protein stability. With respect to protein-binding affinity, we are combinatorially mutagenising leucine zippers as simple model protein interactions (Diss & Lehner, eLife, 2018). Quantifying the effects of very large numbers of mutation combinations allows us to build energy models that accurately predict affinity from sequence. We will present these approaches and the insights they provide into genetic architecture, protein biophysics and the grand challenge of sequence-to-activity prediction for complex changes in genotype.

Keywords: Protein, stability, binding, thermodynamics, combinatorial mutagenesis, sequence-to-activity prediction

MoCHI: a package to infer mechanistic models, free energies, energetic couplings and global epistasis from deep mutational scanning data

Presenter: Andre Faure (Centre for Genomic Regulation (CRG))

Andre J. Faure (1), Ben Lehner (1,2,3)

1. Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Doctor Aiguader 88, 08003 Barcelona, Spain
2. Universitat Pompeu Fabra (UPF), Barcelona, Spain
3. Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

The massively parallel nature of deep mutational scanning (DMS) allows the quantification of phenotypic effects of thousands of perturbations in a single experiment. We have shown previously that the underlying biophysical effects of mutations can be efficiently inferred by [1] quantifying multiple molecular phenotypes in multiple genetic backgrounds and then [2] fitting thermodynamic models to this data using neural networks. This general approach, referred to as ‘multidimensional mutagenesis’, allows the energetic landscapes of many proteins to be rapidly and comprehensively mapped. However, there are currently no software tools available that permit non-specialist researchers to easily model this type of multi-phenotype DMS data in a flexible manner.

To address this need, we have developed MoCHI, a software tool that allows the parameterisation of arbitrarily complex models using DMS data. MoCHI simplifies the task of building custom models from measurements of mutant effects on any number of phenotypes. It also allows the inference of free energy changes, as well as pairwise and higher-order interaction terms (energetic couplings) for specified biophysical models. Furthermore, when a suitable user-specified mechanistic model is not available, global nonlinearities (epistasis) can be estimated directly from the data.

Finally, MoCHI builds upon and leverages theory on ensemble (or background-averaged) epistasis to learn sparse predictive models that are more informative of the true genetic architecture of the underlying biological system. MoCHI is freely available and implemented as an easy-to-use python package relying on the PyTorch machine learning framework.

Keywords: neural networks, thermodynamic models, epistasis

Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN

Presenter: Shawn Fayer (University of Washington)

Shawn Fayer (1), Carrie Horton (7), Jennifer N. Dines (1), Alan F. Rubin (5, 6), Marcy E. Richardson (7), Kelly McGoldrick (7), Felicia Hernandez (7), Tina Pesaran (7), Rachid Karam (7), Brian H.Shirts (3, 4), Douglas M. Fowler (1, 2, 3), Lea M. Starita (1, 3)

1 Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA 2 Department of Bioengineering, University of Washington, Seattle, WA 98195, USA 3 Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA 4 Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA 98195, USA 5 Bioinformatics

Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3052, Australia 6 Department of Medical Biology, University of Melbourne, Melbourne, VIC 3010, Australia 7 Ambry Genetics, Aliso Viejo, CA 92656, USA

Clinical interpretation of missense variants is challenging because the majority identified by genetic testing are rare and their functional effects are unknown. Consequently, most variants are of uncertain significance and cannot be used for clinical diagnosis or management. Although not much can be done to ameliorate variant rarity, multiplexed assays of variant effect (MAVEs), where thousands of single-nucleotide variant effects are simultaneously measured experimentally, provide functional evidence that can help resolve variants of unknown significance (VUSs). However, a rigorous assessment of the clinical value of multiplexed functional data for variant interpretation is lacking. Thus, we systematically combined previously published BRCA1, TP53, and PTEN multiplexed functional data with phenotype and family history data for 324 VUSs identified by a single diagnostic testing laboratory. We curated 49,281 variant functional scores from MAVEs for these three genes and integrated four different TP53 multiplexed functional datasets into a single functional prediction for each variant by using machine learning. We then determined the strength of evidence provided by each multiplexed functional dataset and reevaluated 324 VUSs. Multiplexed functional data were effective in driving variant reclassification when combined with clinical data, eliminating 49% of VUSs for BRCA1, 69% for TP53, and 15% for PTEN. Thus, multiplexed functional data, which are being generated for numerous genes, are poised to have a major impact on clinical variant interpretation.

Keywords: Variant interpretation, variant of uncertain significance, clinical genetics

Multiplexed Functional Assay of MYH7 Variants Associated with Hypertrophic Cardiomyopathy in Gene-Edited Human Cardiomyocytes

Presenter: Clayton Friedman (University of Washington)

Clayton E. Friedman (1,2,3), Shawn Fayer (4), Sriram Pendyala (4), Wei-Ming Chien (1,2,3,5), Linda Tran (1,2,3), Sy-Ling Chao (1,2,3), Elaheh Karbassi (1,2,6), Charles E. Murry (1,2,3,6,7), Lea M. Starita (4,8), Douglas M. Fowler (4,7,8), Kai-Chun Yang (1,2,3,5)

1 Institute for Stem Cell and Regenerative Medicine, University of Washington, School of Medicine, Seattle, WA 98109, USA. 2 Center for Cardiovascular Biology, University of Washington, Seattle, WA 98109, USA. 3 Department of Medicine/Cardiology, University of Washington, Seattle, WA 98109, USA. 4 Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. 5 Cardiology/Hospital Specialty Medicine, VA Puget Sound HCS, Seattle, WA 98108, USA. 6 Department of Pathology, University of Washington, Seattle, WA 98109, USA. 7 Department of Bioengineering, University of Washington, Seattle, WA 98195, USA. 8 Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA.

Heterozygous missense mutations in myosin heavy chain 7 (MYH7) are a leading cause of inherited hypertrophic cardiomyopathy (HCM), accounting for ~40% of all cases. Genetic testing of HCM patients has identified greater than 1,731 MYH7 missense variants, however, ~84% are classified as variants of unknown significance (VUS) due to insufficient functional/clinical data required to interpret variant effect. In particular, while the MYH7 N-terminal S2 domain is sensitive to missense mutations, the leading in silico variant effect prediction algorithm fails to accurately predict known pathogenic variants in this domain, indicating the need for further functional analysis using in vitro disease models. Here we used CRISPR/Cas9 gene-editing to simultaneously generate heterozygous missense variants for five amino acid positions in the S2 domain within the endogenous MYH7 locus in human induced pluripotent stem cells (hiPSC). We generated a library of greater than 112 heterozygous MYH7 missense variants in hiPSCs, encoding for 78/98 amino acids (~80% library representation). Using an established, small-molecule cardiac-directed differentiation protocol, we differentiated this hiPSC variant library to cardiomyocytes and scored variant effect by measuring variant abundance across a cytotoxicity time course assay using next generation sequencing. Preliminary results indicate this multiplexed functional assay correctly called 75% of known pathogenic variants (ClinVar, 4) as functionally abnormal and 100% of synonymous variants (4) as functionally normal, indicating this assay is calibrated to accurately determine variant effect for additional MYH7 variants associated with HCM. Together, this study leverages advanced gene-editing, enabling end-to-end bulk generation and functional annotation of MYH7 variants in hiPSC-derived cardiomyocytes at 10-100X the scale of current approaches. The pipeline established here creates the opportunity for exon-scale mutational scanning to study variant effect in disease-relevant hiPSC-derived lineages to enable VUS interpretation at massive scale.

Keywords: DMS, hiPSC, Cardiomyocyte, MYH7, Disease modeling, Hypertrophic cardiomyopathy

REsolution: a portrait of the genetic variant landscape in human solute carrier transporters

Presenter: Andrea Garofoli (CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences)

Andrea Garofoli (1), Evandro Ferrada (1), Barbara Steurer (1), Tanja Osthushenrich (2), Aidan MacNamara (2), Anders Malarstig (3), Tabea Wiedmer (1), Giulio Superti-Furga (1)

(1) CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, (2) Bayer AG, (3) Pfizer Ltd.

INTRO: With over 400 members, Solute Carriers (SLCs) represent the largest superfamily among the transporter proteins in the human genome. They are the main determinants of the movement of a broad range of substrates both across the cell membrane and in its organelles. Substrates include elements necessary to sustain the cellular metabolic homeostasis, like amino acids, nucleosides, metals or ions, making it obvious how biological dysfunctions can stem from the presence of genetic alterations in SLC genes. While the correlation between the polymorphisms in SLCs and health disorders is recognized, a comprehensive assessment of the translational impact of the genetic landscape of SLCs is not yet available, showing a significantly idle potential for the development of novel precision medicine therapies. As a joint effort between public and private research partners, the REsolution consortium, funded by the Innovative Health Initiative and European Federation of Pharmaceutical Industries and Associations, was formed with the goal to fill the knowledge gap of the role of SLC variants in the rise of human diseases.

METHODS: As a first step, REsolution aims to retrieve existing SLC specific information. Publicly accessible databases, which collect medical, experimental and statistical evidence between the presence of specific variants and diseases, are programmatically mined. The data is manually annotated and weighted to serve as an essential stepping stone for the project, for driving downstream analyses and selecting few SLCs with therapeutic target potential. In the next step, this subset of SLCs is experimentally investigated based on two complementary approaches. The first one focuses on the assessment of localization and transport function of a selected set of genetic variants. Said variants include both mutations with known pathogenic effects and others with predicted, but not-yet validated phenotypic consequences. The second approach relies on Deep Mutational Scanning (DMS), where every amino acid in the protein’s chain is sequentially mutated into each of the other 19 residues and the effect on the SLC’s function is assessed and quantified. In the last step of the project, computational approaches will aid in the medical interpretation of SLC variants by integrating structural and functional annotations to the experimental readouts, providing an additional point of view towards their biological implication. Moreover, due to specific features associated with transmembrane transporters, all the data collected and generated in the previous phases will be exploited to engineer by machine learning a Variant Effect Predictor (VEP) specifically designed to infer the pathogenic impact of SLC mutations. The optimization of the prediction power caused by the fine-tuned features and model development will not only help to better understand the role of the formerly studied variants, but can act as an important base for the identification of additional potential drug targets in future studies.

CONCLUSION: REsolution was established to be a turning point for any researcher interested in the genetic landscape of SLCs. Data, tools and assays generated throughout the project will be released to the research community in the shape of a reasoned compendium and its associated, publicly accessible knowledgebase.

Keywords: SLC, genetic variants, disease association, DMS, VEP

A missense variant effect map for the CHEK2 tumour suppressor

Presenter: Marinella Gebbia (University of Toronto)

Marinella Gebbia (1,2,3), Rosanna Jiang (1,2,3), Maria Nguyen (1,2,3), Nishka Kishore (1,2,3), Michelle Gavac (1,2,3), Jochen Weile (1,2,3), Roujia Li (1,2,3), Song Sun (1,2,3), Jennifer Dines (4), Rick A Boonen (5), Haico van Attikum (5), Doug Fowler (4,6), Frederick P Roth (1,2,3)

1The Donnelly Centre, University of Toronto, Toronto, ON, Canada; 2Department of Molecular Genetics and Computer Science, University of Toronto, Toronto, ON, Canada; 3Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, ON, Canada; 4Department of Genome Sciences, University of Washington, Seattle, USA; 5Leiden University Medical Center, Leiden, The Netherlands; 6Department of Bioengineering, University of Washington, Seattle, USA

CHEK2 encodes a serine/threonine protein kinase (CHK2) which is important for pausing the cell cycle, initiating DNA repair and induction of apoptosis after DNA damage. CHK2 is also involved in proper mitotic spindle assembly and maintenance of chromosomal stability by phosphorylation of the tumor suppressor BRCA1. Consistent with its role as a ‘checkpoint gene’, both germline and somatic variants in CHEK2 have been linked to multiple cancer types. Genetic diagnosis of CHEK2-dependent cancer is made challenging by the fact that 95% of CHEK2 missense variants are classified as variants of uncertain significance. We therefore implemented a scalable assay of CHEK2 variants based on ability to complement loss of the S. cerevisiae RAD53 gene in the presence of methyl methanesulfonate DNA damage. We thereby assessed the majority of all possible amino acid substitutions, yielding a missense variant effect map for CHEK2. This map showed significant correlation small-scale assays and clinical pathogenicity, potentially enabling more informative clinical variant interpretation of CHEK2 missense variation.

Keywords: CHEK2, tumour suppressor, missense variants, variant effect map

Improving G6PD variant classification through multiplexed functional assessment

Presenter: Renee Geck (University of Washington)

Renee C. Geck (1), Maya N. Walker (1), and Maitreya J. Dunham (1)

(1) Department of Genome Sciences, University of Washington

Glucose-6-phosphate dehydrogenase (G6PD) deficiency affects over 500 million people. G6PD is important in red blood cells since it is the sole source of NADPH needed for detoxification of reactive oxygen species. Individuals with G6PD deficiency have variants with decreased activity, which can lead to hemolysis after infections or exposure to oxidants, including antibiotics, antimalarials, and treatments for tumor lysis syndrome. However, diagnosis of G6PD deficiency by activity-based assays give false negatives during hemolytic crises, and since G6PD is on the X chromosome, most heterozygous females have normal activity but must be identified to determine fetal risk and avoid triggers during pregnancy. In order to improve G6PD variant interpretation, we sought to collect all reports of G6PD variant function and genotype-phenotype associations and to measure function of additional variants by conducting a deep mutational scan (DMS).

We identified 1,123 G6PD variants in population and clinical databases, only 167 of which had functional interpretations in ClinVar. By applying ACMG guidelines to clinical and functional information found through literature review, we classified an additional 148 variants.

We are expanding these variant interpretations with additional functional data by conducting a DMS of G6PD in S. cerevisiae. We established a system to measure G6PD function by ability to rescue the growth of yeast lacking the G6PD homolog Zwf1 in oxidative stress. Direct competition of variants of known function in batch culture and in turbidostats showed that strains expressing high-activity variants out-competed strains with low-activity variants.

We obtained a site-saturated mutagenesis library of 10,784 G6PD variants from Twist Biosciences, added barcodes, and used the library to transform zwf1 S. cerevisiae. We grew the yeast library under oxidative stress in turbidostats for 72h and deeply sequenced the barcodes of samples taken at different timepoints. Our initial analyses confirm that at later timepoints some barcodes had altered frequency, indicative of selection. We are currently using PacBio long-read sequencing to associate barcodes with variants. By combining our clinical and yeast data, we will benchmark our functional assay with variants of known clinical effect, compute evidence for levels of support, and submit evidence to ClinVar, enabling interpretation of more genetic variants of G6PD with greater confidence.

We are also introducing SNVs of common background haplotypes onto our variant library and conducting scans on multiple genetic backgrounds. G6PD variants have historically been studied on a Caucasian background even though that does not represent the populations most affected by G6PD deficiency. We measured the function of select multiple missense variants in our yeast system and observed that the function is not always predicted from additive effects of the two mutations. Given this genetic complexity, our DMS on multiple backgrounds addresses an unmet need to systematically investigate the effects of common variants on rare G6PD variant activity. Altogether, our study will enable clinical interpretation of a larger number of G6PD alleles, and ensure that these interpretations are applicable to the affected populations.

Keywords: Saccharomyces cerevisiae, G6PD, deep mutational scan, turbidostat

Scaling up Saturation Genome Editing to Maximise Clinical Impact

Presenter: Sebastian Gerety (Wellcome Sanger Institute)

The Wellcome Sanger Institute Multiplex Assay of Variant Effects Team

Wellcome Sanger Institute, Wellcome Genome Campus, CB10 1SA, UK

There is an urgent clinical need to improve our ability to predict the pathogenicity of genetic variants in order to make genetic diagnoses. Predictive testing and some targeted therapies cannot be offered where variants of uncertain significance (VUS) are identified and the number of VUS is growing rapidly. Conventional strategies to resolve VUS rely on the accumulation of clinical data and variant effect prediction algorithms. Clinical data accumulate too slowly to be useful to families with rare neurodevelopmental and cancer-predisposition disorders. Variant effect prediction algorithms are confounded by circularity and error propagation and are often discordant. New approaches are urgently needed. Multiplexed assays of variant effect (MAVEs) have proven utility in clinical variant interpretation in cancer-predisposition syndromes (Fayer et al. 2021). We have successfully applied saturation genome editing (SGE, Findlay et al., 2018) to the neurodevelopmental disorder gene DDX3X. These data have over 90% sensitivity and specificity and we estimate that their incorporation in clinical interpretation could reduce the number of DDX3X-related neurodevelopmental disorder VUS by up to 90%. Similarly, we have performed SGE on the tumour-suppressor gene BAP1, demonstrating a clear separation of function between known pathogenic and benign variants associated with cancer germline predisposition.

As part of the AVE Alliance, The Wellcome Sanger Institute Multiplex Assay of Variant Effects Team aims to establish a high throughput pipeline to apply SGE to dozens of genes per year. To achieve this scale, and beyond, we must improve the efficiency and reduce the cost of MAVE techniques. Towards that aim, we have undertaken a pilot study applying SGE to 20

loci across 10 disease-associated genes. We will present results from the study, and our progress to date in refining SGE for large scale implementation.

Keywords: MAVE SGE HAP1 disease genetics

Multiplexed assessment of human glucokinase variant activity

Presenter: Sarah Gersing (University of Copenhagen)

Sarah Gersing (1), Matteo Cagiada (1), Marinella Gebbia (2,3,4), Anette P. Gjesing (8), Gireesh Seesankar (2,3,4), Amelie Stein (1), Anna L. Gloyn (6,7), Torben Hansen (8), Frederick P. Roth (2,3,4,5), Kresten Lindorff-Larsen (1), Rasmus Hartmann-Petersen (1)

1: The Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark. 2: Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada 3: Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada 4: Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON M5G 1X5, Canada 5: Department of Computer Science, University of Toronto, Toronto, ON M5T 3A1, Canada 6: Division of Endocrinology, Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA 7: Stanford Diabetes Research Center, Stanford University, Stanford, California, USA 8: Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Glucokinase (GCK) functions as the body's glucose sensor by catalyzing the rate-limiting step in insulin secretion, phosphorylation of glucose to form glucose-6-phosphate. Emphasizing the importance of GCK in glucose homeostasis, missense variants that decrease or increase GCK activity are associated with three diseases: hyperinsulinemic hypoglycemia (HH), permanent neonatal diabetes mellitus (PNDM) and GCK-maturity-onset diabetes of the young (GCK-MODY). GCK-MODY is a mild type of monogenic diabetes that usually can remain untreated, yet people with GCK-MODY frequently receive unnecessary treatment due to misdiagnosis. Consequently, diagnosing GCK-MODY can benefit economy and life quality, but is restrained by variant interpretation. To aid diagnosis, we used functional complementation in yeast to create a variant effect map of human GCK activity, comprising 97% of the possible missense and nonsense variants. Variant activity scores reflected in vitro catalytic efficiency and correlated with fasting plasma glucose levels in humans. To evaluate our map more broadly, we used evolutionary conservation analysis, which was unable to capture hyperactive variants, but supported neutral and hypoactive variants. Hypoactive variants clustered at buried residues, the active site and a surface-exposed patch involved in GCK conformational changes. In contrast, hyperactive variants concentrated at an allosteric activator site. We further explored these variants by analyzing the conformational free energies of GCK's inactive and active conformations, and found that 467 hyperactive variants destabilize the inactive state relative to the active state. Through destabilization, these variants shift the conformational equilibrium towards the active state, providing a potential mechanism for their increased activity. In conclusion, we comprehensively assess GCK variant activity to aid in GCK-MODY diagnosis, and further characterize the allosteric activator site, which may aid in the refinement of GCK-targeted drugs.

Keywords:

Deep mutagenesis to understand yeast prion nucleation

Presenter: Marta Badia Graset (Institute for Bioengineering of Catalonia)

Marta Badia Graset (1), Ben Lehner (2, 3, 4) and Benedetta Bolognesi (1)

(1) Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Barcelona, Spain.
(2) Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain.
(3) Universitat Pompeu Fabra (UPF), Barcelona, Spain.
(4) ICREA, Pg. Lluís Companys, Barcelona, Spain.

Prions are misfolded proteins capable to promote conformational changes of other protein isoforms. When prion proteins switch from a soluble (non-prion) state to a misfolded (prion) state, they can bind to each other forming small nuclei that can rapidly incorporate other monomers and form amyloid-like aggregates. During the last decades, yeast prions have been intensively studied, and they serve as a good model to understand amyloid formation and aggregation. In yeast, prion switching and maintenance are driven by prion domains, which are mostly glutamine/asparagine rich and can form amyloids in vitro.

One of the most historically famous yeast prions is Sup35, which is an essential translation termination factor. The first 123 residues of Sup35 correspond to its prion domain, which is sufficient for prion formation and maintenance, but the mechanisms by which this protein starts misfolding, nucleating, and forming amyloid aggregates still need to be elucidated.

Our lab exploits an assay that is able to capture Sup35 nucleation in yeast cells. By using deep mutagenesis, we built a library of the QN-rich region of the Sup35 prion domain, and we assessed the impact of hundreds of single mutants on Sup35 nucleation.

The mutational landscape of Sup35 nucleation reveals that the majority of single mutations in the QN-rich region of Sup35 reduce the nucleation of new aggregates. The impact of mutations on nucleation is more evident in the central core of the library (from residues 17 to 25) suggesting that this region builds the inner core of the self-templating conformation of Sup35. In contrast, outside this core, we found mutations that increase nucleation. For example, an increase of aromaticity in specific positions enhances the nucleation of Sup35.

This comprehensive dataset highlights how, thanks to deep mutagenesis, we can uncover unknown features that will help us to understand the mechanism of nucleation of QN-rich prions.

Keywords: prion, amyloid, nucleation, amyloid formation

In Silico Analysis of Synonymous and Non-Synonymous SNPs of the Human Cereblon (CRBN) Gene: An Important Gene in Immunomodulatory Drugs (IMiDs)-Based Treatment of Multiple Myeloma Patients

Presenter: Murat Güler (Genomic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany)

Murat Güler(1), Angelica Macauda(1), Federico Canzian(1)

(1)Genomic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

Background: Immunomodulatory drugs that target cereblon, a component of the E3 protein ligase complex encoded by the CRBN gene, are used to treat multiple myeloma (MM) and myelodysplastic syndrome. Because the CRBN gene is so important in IMiD metabolism, it is of interest to assess the effects of single nucleotide polymorphisms (SNPs) on cereblon, which can potentially affect IMiD treatment response in MM patients Methods: The CRBN SNPs were collected from dbSNP with filtering validation status by ALFA option. The nucleotide sequence of CRBN and amino acid sequence were downloaded from NCBI and UniProt respectively. The structure of the CRBN-pomalidomide complex was downloaded from RCSB PDB and the CRBN structure was obtained from AlphaFold. Genetic variants were analyzed with different tools for variant prioritization. The missense variants were scored by 12 different tools (PANTHER, SIFT, PolyPhen, CADD, BayesDel, ClinPred, LRT, M-CAP, MetaLR, MetaSVM, REVEL, fathmm-XF) and PTM position. The potential structural effects of amino acid changes were predicted by PremPLI (protein- IMiD complex) and INPS-MD (only protein) tools by prediction of protein stability free energy change (ΔΔG). The noncoding and non-exonic variants were scored by 10 tools (CADD, Deepsea, eigen_PC, fathmm, funseq2, fitcons_Fitness, gerp_Base.RS, gwava, phast, ReMM) and CpG region, TFBS and miRNA binding site. After calculating scores by different tools, we assigned to each SNP deleterious-benign status (0 and 1) for each prediction, then we calculated a Zscore and cumulative distribution function of normal distribution of Z-scores (one-tailed, 95% CI). Results: dbSNP has 12611 CRBN SNPs validated by ALFA: 267 deletions, 1394 deletion-insertions, 78 insertions, and 10872 single nucleotide variants. After removing duplicates, we found 314 non-synonymous and 8602 synonymous and non-coding SNPs. The experimentally detected PTM sites are extremely conserved, as only eight of them have SNPs, all of them being rare in all populations (minor allele frequency between 0-0.00002639). The 27 of 314 non-synonymous SNPs passed nominal p-value (<0.05), but after Bonferroni correction (0.05/314) none of them were significant. However, PremPLI and INPS-MD tools-based prediction of protein stability free energy change (ΔΔG) values were significantly different between these 27 SNP and the others. The INPS-MD predicted ΔΔG mean in significant SNPs is -1.048kcal/mol (min -3.659-max 0.645), while in the non-significant group the mean was -0.376kcal/mol (min -3.361-max 1.617), t-Test p=1.61x10-6. The PremPLI-predicted ΔΔG mean in the top 27 SNPs was 0.415kcal/mol (min 0.01-max 0.76), while in nonsignificant group the mean was 0.309kcal/mol (min -0.93 - max 0.9), t-Test p=0.041. The 107 of 8602 synonymous and other non-coding SNPs passed p-value threshold after multiple test correction (0.05/8602). The gnomAD global allele frequency of these SNPs is ranging from 1.16x10-4 to 8.34x10-6. Summary/Conclusion: As a result, the PTM sites of cereblon are clearly functionally significant. Therefore, malfunctions in proteins involved in the post-translational modification of cereblon should have an impact on its activity. Furthermore, protein stability ΔΔG predicts also deleteriousness of SNPs. Even though significant SNPs are very rare, they should be tested to assess their functional effect in MM cells.

Keywords: Imids, Multiple myeloma, Pharmacogenetics, SNP

Massively parallel functional testing of SGCB missense variants causing Limb-Girdle Muscular Dystrophy

Presenter: Gabe Haller (Washington University)

Chengcheng Li 1, Jackson Wilborn 2, Sara Pitmman 1, Jil Daw 1, Conrad Weihl, 1, Gabe Haller 2,1,3

1. Department of Neurology, Washington University School of Medicine, Saint Louis, MO USA
2. Department of Neurosurgery, Washington University School of Medicine, Saint Louis, MO USA

3. Department of Genetics, Washington University School of Medicine, Saint Louis, MO USA

Mutations in α-, β-, γ-, and δ-sarcoglycan cause sarcoglyanopathies, a subset of limb-girdle muscular dystrophy (LGMD) with devastating effects for patients including muscle wasting, progressive weakness, degeneration of skeletal muscle and often premature death. Accurately diagnosing patients with LGMD before symptom onset or early in the course of the disease has the potential to enable the use of preventative gene therapy or other therapeutics and in the majority of cases can only be done in presymptomatic cases through genetic testing. We focused on SGCB because it is the most frequently mutated sarcoglycan gene in LGMD. We employed deep mutational scanning to measure the effects of all possible missense variants of the SGCB protein simultaneously. We expressed libraries of variants in cultured human cells via lentivirus and measured their effects using a fluorescence-based assay that measures the capacity of β-sarcoglycan to localize to the plasma membrane. We have analyzed the functional effect of >98% of the 6340 possible amino acid changes (20,288 mutant DNA sequences) and have found the distribution of scores to be bimodally distributed and almost perfectly concordant with established assessments of pathogenicity reported in ClinVar. Using our results, >30% of VUS in ClinVar would be reclassified as 'Pathogenic' and the remainder 'Benign'. Further, we show that variants that fail to properly traffic also hinder the ability of other sarcoglycans to traffic to the membrane. We predict that these results will be useful for clinical interpretation of SGCB variants, improve LGMD diagnosis and enable wider use of potentially life-saving gene therapy.

Keywords: SGCB, Limb-Girdle Muscular Dystrophy, LGMD, DMS, Sarcoglycanopathy,

Fingerprinting serine protease inhibitor (SERPIN) specificity

Presenter: Laura Haynes (Life Sciences Institute, University of Michigan)

Laura M Haynes (1), Matthew L Holding (1), David Siemieniak (1,2), and David Ginsburg (1,2)

(1) Life Sciences Institute, University of Michigan
(2) Howard Hughes Medical Institute

Serine protease inhibitors (SERPIN) are a protein superfamily whose members are found across all kingdoms of life and are encoded by over 30 genes in humans that regulate diverse processes, including blood coagulation/fibrinolysis, the immune response, inflammation, and extracellular matrix remodeling, by specifically and irreversibly inhibiting serine proteases. In their active form, SERPINs exist in a metastable state in which a reactive center loop (RCL) extends from the globular core of the protein which serves as a bait for the target protease. When the protease attempts to cleave the peptide bond between the P1 and P1' residues of the RCL, the SERPIN and protease become covalently linked when, prior to hydrolysis of the acyl intermediate, the RCL inserts into the SERPIN's central -sheet resulting in a pole-to-pole translocation and structural rearrangement of the protease. This process renders both SERPIN and protease no longer functional. SERPIN specificity for its target protease has long been assumed to be driven primarily by the amino acid sequence of the RCL. However, in the present study we demonstrate that determinants of protease specificity are dispersed throughout the SERPIN's sequential and structural space. Plasminogen activator inhibitor-1 (PAI-1) is a 379 amino acid prototypic member of the SERPIN family whose canonical function is to inhibit tissue-type and urokinase-like plasminogen activators (tPA and uPA, respectively). A phage displayed PAI-1 mutational library was generated using error prone PCR that consists of 6.6x106 independent clones with an average of 3-4 amino acid mutations per clone and containing 6,388 of the 7,201 possible missense mutations. The library was subsequently panned for inhibitory activity against its canonical enzyme target uPA, the procoagulant proteases thrombin and factor XIIa, as well as TMPRSS2, which primes influenza and SARS-CoV viruses for cellular entry and has been reported to be inhibited in vivo by PAI-1. More than 400 million PAI-1 DNA sequences from the input and selected libraries were determined by next generation DNA sequencing. By comparing the amino acid substitutions present in the input library to those in the selected libraries, unique patterns of enrichment and depletion of specific substitutions across PAI-1's sequence space were observed—underscoring that PAI-1 (and SERPIN) specificity is driven by exosite interactions and structural features distal from the RCL. Differences in the PAI-1 mutational landscape for each of the proteases also roughly correspond with the evolutionary diversity of the serine proteases studied, suggesting the coevolution of SERPINs with serine proteases. Finally, the results of this study form the foundation for engineering PAI-1 variants with specificity toward other non-canonical proteases.

Keywords: SERPIN, protease, evolution, coagulation

Development of a multiplexed assay of variant effects for DNA repair gene MUTYH

Presenter: Shelby Hemker (University of Michigan)

Shelby L Hemker (1), Bala Bharathi Burugula (1), Grace E Clark (1), Sajini Jayakody (1), Jacob O Kitzman (1,2)

1) Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA
2) Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, USA

Disruption of MUTYH causes MYH-associated polyposis, a recessive condition for which up to 1:50 individuals are carriers, depending upon ancestry group. Nearly 800 missense mutations in MUTYH have been reported in ClinVar, 95% of which are classified as variants of uncertain significance (VUS). To resolve these reported VUS, and as-yet unseen patient variants in MUTYH, we are developing a multiplexed assay of variant effects (MAVE) to predict pathogenicity of all MUTYH missense variants. The normal function of MUTYH is to repair A:8oxoG lesions, a product of oxidative damage. We have modified an existing fluorescent reporter cell-based assay for MUTYH variant function, in which proper A:8oxoG to C:G repair initiated by MUTYH yields fluorescent protein expression. To perform this assay, we generated a CRISPR-induced MUTYH knockout cell line and then introduced different MUTYH variants: a wildtype (WT) copy, a truncating nonsense variant (Q363X), and two pathogenic missense variants - one previously demonstrated to have partial loss of function (G368D) and one with nearly total loss of function (Y151C). The Q363X variant exhibited practically no repair of the 8oxoG lesion, while the Y151C and G368D variants were able to repair the lesion at approximately 5% and 30% of WT levels, respectively. Thus, this assay has the dynamic range to distinguish between various levels of functional impairment, and is amenable to high-throughput screens such as deep mutational scans. Using this reporter as a functional read-out, we will complement MUTYH knockout cells with libraries comprising every MUTYH missense mutation (n=9,899). The pathogenicity predictions of these assays will be 1) compared to clinically identified variants from ClinVar, 2) refined based upon evolutionary conservation, and 3) used to identify structural mechanisms underlying hotspots of functional constraint by mapping data to 3D structure.

Keywords: DNA repair, MAVE, FACS

Multi-Functional Assessment of ACE2 Variants Reveal Insights Into COVID-19 Disease Pathophysiology

Presenter: Ryan J. Hong (University of British Columbia)

Warren M. Meyers (1, 2, 3), Ryan J. Hong (1, 2, 3), Wun Chey Sin (1,2), Kurt Haas (1, 2, 3)

1 Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, BC, Canada

2 Department of Cellular and Physiological Sciences, University of British Columbia, Vancouver, BC, Canada

3 School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada

Confounding features of COVID-19 is the diversity in disease severity – ranging from asymptomatic to life-threatening complications, and why distinct ethnic or regional groups exhibit increased infection rates and poorer outcomes. One likely source of population-based susceptibility is genetic variation of the human receptor for the SARS-CoV-2 virus, angiotensin-converting enzyme 2 (ACE2). ACE2 is an enzyme involved in the renin-aldosterone-angiotensin system in both the body and brain, and single nucleotide missense variants of ACE2 have been linked to the comorbidities associated with increased susceptibility and worse outcomes of COVID-19, including obesity, diabetes, and cardiovascular diseases. To investigate the molecular mechanisms underlying a potential role for ACE2 variants in COVID-19, we have developed a multi-assay platform to assess impact of ACE2 variants identified in different ethnic populations or associated with disease on protein stability, surface delivery, catalytic reactions associated with COVID-19 comorbidities, and binding to the SARS-CoV-2 S1 receptor binding domain (RBD). Further, given the rapid emergence of natural mutations of the SARS-CoV-2 virus underlying COVID-19, we will also test how variants of the RBD alters binding to ACE2 and ACE2 functions. Results may explain population differences in disease distribution.

Keywords: COVID-19, SARS-CoV-2, ACE2, functional genomics, therapeutics

satmut_utils: a simulation and variant calling package for multiplexed assays of variant effect

Presenter: Ian Hoskins (The University of Texas at Austin)

Ian Hoskins (1), Frederick Roth (2, 3, 4), Can Cenik (1)

1. Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712, USA
2. Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Ontario, Canada
3. The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada
4. Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, Ontario, Canada

The impact of thousands of individual genetic variants on molecular phenotypes for disease-relevant genes remains unknown. Multiplexed assays for variant effect (MAVEs) are a highly scalable approach to annotate the relevant variants. However, current software methods for MAVEs lack standardized annotation, can require cumbersome configuration, and do not easily scale to large target regions.

Here, we present satmut_utils as a flexible solution for 1) simulation of saturation mutagenesis data; and 2) unparalleled quantification of variants across four orders of magnitude from multiplexed assay data. satmut_utils offers several

improvements including support for multiple experimental strategies, unique molecular identifier-based consensus deduplication, and machine learning-based error correction. We developed a rigorous simulation workflow to validate the performance of satmut_utils and carried out the first benchmarking of existing software for variant calling. Finally, we used satmut_utils to determine the mRNA abundance of thousands of coding variants in cystathionine beta-synthase by two library preparation methods. We identified an association between variants near chemical cofactor binding sites and decreased mRNA abundance. We also found a correlation between codon optimality and the magnitude of variant effects, emphasizing the potential of single-nucleotide variants to alter mRNA abundance.

In summary, satmut_utils enables high-performance analysis of saturation mutagenesis data, achieves unprecedented specificity through novel error correction approaches, and reveals the capability of single-codon variants to alter mRNA abundance in native coding sequences.

Keywords: MAVE, DMS, mutagenesis, variant calling, SNP, CBS, codon optimality

All you need to know about Saturation Prime Editing

Presenter: Evgueni Ivakine (The Hospital for Sick Children)

Evgueni Ivakine (1,2)

1 Program in Genetics and Genome Biology, The Hospital for Sick Children; 2 Department of Physiology, University of Toronto

The advent of new genetic technologies and high-throughput methods has opened the doors to an era of non-hypothesis-driven biomedical research. We are now able to ask questions on a scale not previously imagined, with one such question being - what is the functional impact of every possible single base change in a gene of interest? Existing methods to answer this question include overexpression studies, as well as base editing screens and HDR-based saturation gene editing. We recently demonstrated a novel platform for this work, termed saturation prime editing (SPE), in which we harnessed (1) locus haploidization, (2) CRISPR prime editing, and (3) a FACS-based functional assay to elucidate the functionality of hundreds of single nucleotide variants in the NPC1 gene (Erwood et al., 2022). This talk will explore each of these three core aspects of our SPE platform in detail, as well as alternative functional assays and other possible applications of SPE. Applications of saturation gene editing using platforms like SPE have the potential to not only understand effects of genetic variants at scale, but also to make significant contributions to the development precision-based treatment strategies.

Keywords: saturation gene editing, prime editing,

Systematic mapping of missense variant function in HMG-CoA Reductase

Presenter: Rosanna Jiang (University of Toronto)

Rosanna J Jiang(1,2), Song Sun(1,2), Marinella Gebbia(1,2), Nishka Kishore(1,2),Yingzhou Wu(1,2), Roujia Li(1,2), Jochen Weile(1,2), Atina G Cote(1,2), Albi Celaj(1,2), Marta Verby(1,2), Frederick P Roth(1,2)

(1)Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, Ontario, Canada; (2)Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Ontario, Canada.

HMG-CoA reductase (HMGCR) is an enzyme that catalyzes the rate-limiting reaction of cholesterol biosynthesis. High levels of low-density lipoprotein cholesterol (LDL-C) increase the risk of coronary heart disease, myocardial infarction and stroke. LDL-C can be lowered via statin drugs that inhibit HMGCR activity. Substantial inter-individual variation in response to statin therapy has been reported, but the genetic component of this variation is incompletely understood. The yeast Saccharomyces cerevisiae has been shown to be a reliable model for studying the impact of human missense variants on metabolic enzymes. Here we describe progress towards en masse yeast-based assays to test the function of all possible missense variants in HMGCR, with the goal of understanding structure-sequence-function relationships.

Keywords: HMGCR, LDL-C

Exploration of the sequence space of an SH3 domain from Saccharomyces cerevisiae

Presenter: David Jordan (Université Laval)

David Jordan (1,2,3,4), Alexandre K. Dubé (1,2,3,4,5), Christian Landry (1,2,3,4,5)

1)Département de biochimie, microbiologie et bioinformatique, Université Laval, Québec, Canada 2)Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada 3) Quebec Network for Research on Protein Function,

Engineering, and Applications (PROTEO), Université Laval, Québec, Canada 4) Université Laval Big Data Research Center (BDRC_UL), Québec, Canada 5) Département de biologie, Université Laval, Québec, Canada

In order to detect external stimuli and respond appropriately, cells possess a complex and robust signaling network. One feature of these networks are interactions between organized protein domains and short stretches of amino acids in intrinsically disordered regions, known as motifs, present on other proteins. An example of this are SH3 domains which bind motifs of the form PXXP, where P represents a proline and X represents any amino acid. SH3 mediated interactions are conserved across the eukaryotic tree of life. SH3 domains are usually composed of around 60 amino acids, and have a conserved structure.

In the model organism Saccharomyces cerevisiae there are 28 different SH3 domains. The SH3 domain of the membrane protein Sho1 plays a role in the response to osmotic stress by binding the motif on the scaffold protein Pbs2, which starts a phosphorylation cascade which allows the cell to respond to osmotic stress.

Previous work has focused on the Pbs2 motif, but little attention has been given to the SH3 domain of Sho1. To better understand the structure and the determinants of binding, one can explore the sequence space surrounding the wild-type sequence of the SH3 domain. This should clarify the role different amino acids play in both binding and structure.

First, mutant versions of Sho1 were constructed where the SH3 domain was replaced by the 27 other SH3 domains present in S. cerevisiae. Then, the interaction between these SH3-swapped mutants and an array of cytoplasmic proteins including Pbs2 was measured. This was done using a protein-fragment complementation assay (PCA), a technique which measures the strength of the interaction between two given proteins. Two proteins of interest are tagged with different fragments of the enzyme dihydrofolate reductase (DHFR). When the proteins of interest interact, the DHFR is reconstituted and functional, allowing for growth in presence of the selective agent methotrexate. Cell growth in methotrexate can therefore be tied to interaction strength. This PCA showed that no other SH3 domains were able to interact with Pbs2, indicating that the SH3 domain of Sho1 has unique characteristics which allow binding.

Next, a deep mutational scanning (DMS) was undertaken on the SH3 domain. The effects of all the mutations in the DMS library were measured by PCA on the binding of Sho1 with Pbs2 and with a control protein, Ybt1. By measuring the interaction of Sho1 with both Pbs2 and Ybt1, the effect of mutations on binding and on structure can be determined. Interactions with Pbs2 measure the affinity of the SH3-motif pair specifically. Interactions with Ybt1 serve as a control and a proxy for structure and stability, as Ybt1 interacts with Sho1 at a locus other than the SH3 domain. An SH3 mutation which negatively affects the interaction with Ybt1 affects either the structure or the stability of the entire protein.

Our assay identified positions which have both positive and negative effects on binding with Pbs2 when mutated. It also identified positions which have a negative effect on binding with Ybt1 when mutated, suggesting that these residues are important for Sho1 structure and stability. A negative correlation was observed between tolerance of mutations and the level of conservation of a position. Follow-up investigations will confirm the effect of selected mutations, and measure the effect of these mutations on growth in the presence and absence of osmotic stress.

Keywords: Deep Mutational Scanning, Protein-protein interactions

Comprehensive Mutational Characterization of Calcium Binding by the STIM1 EF-hand

Presenter: Nisha D. Kamath (Case Western Reserve University School of Medicine)

Nisha D. Kamath (1), Kenneth A. Matreyek (1)

(1) Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH 44118

Intracellular calcium signaling regulates a wide range of biological processes, including gene transcription, muscle contraction, and cell death. Amplification of intracellular calcium signaling occurs through store operated calcium entry (SOCE), where reduced calcium in the ER and subsequent mobilization of ER transmembrane protein “sensor” STIM1 activates the plasma transmembrane channel ORAI1, resulting in massive influx of calcium from outside the cell. Accordingly, coding mutations in STIM1 and ORAI1 can dysregulate SOCE, with loss-of-function variants causing channelopathies and immunodeficiency, whereas gain-of-function variants are associated with tubular aggregate myopathy and Stormorken Syndrome. A functionally important domain of STIM1 is the ~12 amino acid canonical EF-hand, responsible for sensing the depletion of ER calcium to trigger SOCE. This motif is fundamental to calcium signaling and is found in over 66 calcium-binding protein subfamilies across the genome. Although the essential EF-hand residues that coordinate calcium ions are known, it is unknown how other residues in the domain influence calcium binding. To answer this, we will conduct a deep mutational scan of the STIM1 EF-hand to learn the importance of each residue of the STIM1 EF-hand domain on calcium binding. We will express a pooled site-saturation mutagenesis library of 600 unique missense variants representing every possible amino acid substitution of the 30 amino acid STIM1 EF-hand domain in engineered HEK 293T cells. We are developing a high-throughput assay based on calcium-induced apoptosis. We hypothesize that variants with normal calcium binding will survive, whereas those with reduced calcium binding will trigger uncontrolled calcium flux and cause cell death. We will use high-throughput sequencing to simultaneously count each variant within the library before and after selection and give each variant a calcium binding score. This will create a

comprehensive sequence-function map detailing the effect of each residue of the EF-hand on calcium binding. These scores will be reflective of long-term calcium dynamics associated with calcium-induced apoptosis, while individual SOCE events occur at vastly shorter timescales. We will thus compare our scores with perturbations to SOCE observed in a panel of cells encoding phenotypically diverse STIM1 variants, by using fluorescent GCaMP calcium biosensors to measure intracellular calcium concentrations in real time following stimulation. The developed apoptosis assay will be further applied to study the EF-hand domains of other calcium-binding proteins such as STIM2 and calmodulin to understand the “rules” by which calcium binding occurs in the EF-hand. We will harness machine learning algorithms such as random forests and apply them to this dataset to predict the effects of single mutations in the EF-hand on calcium binding. Our study will elucidate the fundamental sequence constraints of EF-hand motifs, while revealing how perturbations to calcium dynamics conferred by SOCE at the cellular level correlate with disease manifestations in people.

Keywords: STIM1, EF-hand, calcium, calcium-binding protein

Learning genotype-phenotype maps from multiplex assays of variant effect

Presenter: Justin Kinney (Cold Spring Harbor Laboratory)

Ammar Tareen (1), Mahdi Kooshkbaghi (1), Anna Posfai (1), William T. Ireland (2,3), David M. McCandlish (1), Justin B. Kinney (1,*)

(1) Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724
(2) Department of Physics, California Institute of Technology, Pasadena, CA, 91125
(3) Present address: Department of Applied Physics, Harvard University, Cambridge, MA, 02134 (*) To whom correspondence should be addressed: jkinney@cshl.edu

Multiplex assays of variant effect (MAVEs) are a family of methods that includes deep mutational scanning experiments on proteins and massively parallel reporter assays on gene regulatory sequences. Despite their increasing popularity, a general strategy for inferring quantitative models of genotype-phenotype maps from MAVE data is lacking. Here we introduce a broadly applicable information-theoretic framework for learning genotype-phenotype maps—including biophysically interpretable models—from MAVE datasets. This approach has been implemented in an easy-to-use, neural-network-based Python package called MAVE-NN. We demonstrate MAVE-NN in multiple biological contexts, and highlight the ability of our approach to deconvolve mutational effects from otherwise confounding experimental nonlinearities and noise.

Keywords:

Checking on the spellcheckers: deep mutational scans of mismatch repair factors and beyond

Presenter: Jacob Kitzman (University of Michigan)

Bala Burugula(1), Shelby Hemker(1), Anthony Scott(1,2), Xiaoyan Jia (1), Sajini Jayakody (1), Grace Clark (1), Jacob Kitzman(1,3)*

1. Department of Human Genetics, University of Michigan Medical School
2. Department of Internal Medicine, University of Michigan Medical School
3. Department of Computational Medicine and Bioinformatics, University of Michigan Medical School

It remains a daunting challenge to distinguish benign rare variants from those which damage gene function and contribute to disease risk, even for genes subjected to decades of biochemical and genetic study. I will share our ongoing efforts to comprehensively measure variant functional effects in several such long-studied factors involved in DNA mismatch repair (MMR). Inherited loss-of-function mutations in this key genome integrity pathway causes Lynch Syndrome, an early-onset colorectal and gynecological cancer predisposition affecting ~1:300 individuals worldwide. We have applied deep mutational scanning to generate functional effect maps for >98% of all possible missense variants in the two most-frequently mutated genes in LS, MSH2 and MLH1. I will describe the validation and active use of these maps in clinical variant interpretation pipelines. Going beyond single variant interpretation, we have extended this screen to systematically interrogate pairs and higher-order combinations of mutations, identifying a large population of variants which epistatically interact to sensitize or, conversely, buffer MMR factors against functional perturbation. Finally, I will present a complementary approach using massively parallel minigene assays to systematically identify variants which disrupt proper mRNA splicing, which we are applying to MMR genes and beyond. These approaches provide a scalable source of functional evidence to assist clinical variant interpretation, and starting point for future mechanistic studies.

Keywords: Lynch Syndrome, deep mutation scanning, missense, splicing assay

A systematic strategy to link human genetic variations to drug response

Presenter: Divya Kriti (University of British Columbia, Vancouver, Canada)

Divya Kriti (1), Corey Nislow (1,2,3)

1. Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, Canada
2. Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, Canada
3. Genome Science and Technology, University of British Columbia, Vancouver, Canada

Introduction: Every individual reacts differently to drugs, and the associated effect varies based on their genetic makeup. To understand this modulation, we need to accurately assess an individual's underlying genetic variation while prioritizing the mutations that affect the phenotype. This research study will develop and deploy a DMS-derived approach called "Variant Technology" or V-Tech to build a comprehensive map of all clinically actionable human variants in crucial drug targets.

Objectives: The overall objective is to combine the advancements in next-generation sequencing with chemical biology, to get an improved understanding of genomics and how it drives an individual's health and response to drugs. The specific aims include (i) identifying genetic mutations responsible for variations in drug efficacy, (ii) functional impact assessment of identified variants, (iii) developing a predictive model of the effect of variants on drug response, (iv) evaluating the impact of new variants in patients.

Methods: V-Tech works by screening genetically engineered humanized yeast cells in a parallel functional assay that quantifies drug resistance and sensitivity, and later bioinformatically assessing how variants in human drug targets modulate response to drug exposure. This systematic strategy involves (i) characterization of drug-target pairs and data collection on existing variants, (ii) performing massively parallel screens of variants for each drug-target pair, and (iii) data analysis to correlate all actionable variants with drug response.

Results: A well-conserved drug target Dihydrofolate reductase (DHFR) for Methotrexate was interrogated using V-Tech. For this, a site-saturated variant library of DHFR was designed with each of the 187 positions in the ORF substituted with all possible amino acids. The library consists of more than 3K variants featuring non-synonymous and yeast-specific codons. After being cloned into appropriate yeast deletion strains, libraries will be exposed to varying concentrations of Methotrexate. Samples will be collected over time for deep sequencing and our in-house variant analysis pipeline will then decode the associations between variants and their responses to the drug. All the bioinformatics analysis results coupled with published clinical variant data will help develop a predictive model.

Conclusion: Currently, there are profound gaps in knowledge about factors that influence the effectiveness of drugs. V-Tech can effectively bridge this gap by quantitating all possible variants in chosen drug-target pairs, by drug response and establishing their significance, which can further help translate variants of unknown significance into actionable variants. With the adoption of this technology by the medical community, an era will emerge where the right treatment is given to the right patient at the right dosage while allowing for reduced therapeutic costs and health care sustainability.

Keywords: Personalized Medicine, V-Tech, DMS, Pharmacogenomics, Drug response, Clinically actionable variants, Humanized yeast

Assessing computational variant effect predictors via a prospective human cohort

Presenter: Kevin Kuang (University of Toronto)

Da Kuang(1,2), Roujia Li(1,2), Yingzhou Wu(1,2), Jochen Weile(1,2), Robert A Hegele(3), Frederick P Roth(1,2)

1 Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, Canada 2 Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Canada 3 Departments of Medicine and Biochemistry, Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, Canada

Computational predictors can help interpret the likelihood of pathogenicity of human genetic variants, especially for the majority of variants where no experimental data are available. However, the lack of high-quality, independent test sets has limited our confidence in identifying the best-performing predictors. To address this issue, we evaluated missense variant effect predictors using rare variants and traits from a prospective cohort that had not previously been used for predictor training. We considered 139 gene-trait combinations for which a systematic study based on ~200K UK Biobank participants had previously reported a rare-variant burden association. Using an evaluation set of 35,525 rare missense variants and the relevant associated traits, we assessed the correlation of participant traits with scores derived from 20 computational variant effect predictors. We found that two predictors—VARITY and REVEL—outperformed all others according to multiple performance measures. Applying the two best-performing predictors, we performed a new burden test using rare variants in the ~200K UK Biobank participants and identified 18 gene-trait combinations that had not been found in previous burden testing efforts. Of these 18, 16 were supported by literature evidence. We expect that this overall approach and specific findings will help guide the selection of variant effect

predictors, for both research and clinical purposes, while providing an independent benchmarking strategy that can be applied to additional cohorts and predictors.

Keywords:

Systematic phenotyping of rare pathogenic disease variants

Presenter: Jessica Lacoste (University of Toronto)

Jessica Lacoste1,2, Marzieh Haghighi3, Syed Haider1, Zhen-Yuan Lin4, Shantanu Singh3, Anne-Claude Gingras3,4, Anne Carpenter3, Mikko Taipale1,2

1 Donnelly Centre for Cellular and Biomolecular Research, Toronto, Canada; 2Department of Molecular Genetics, University of Toronto, Canada; 3Broad Institute of Harvard and MIT, Cambridge, MA, USA; 4Lunenfeld-Tanenbaum Research Institute, Toronto, Canada

There are over 7,000 known rare diseases and over 80% have a simple Mendelian pattern of inheritance. Despite our knowledge of the genetic mutations that cause rare diseases, little is known of the cellular consequences of these genetic mutations. Functional studies on rare disease have fallen behind gene discovery since advances in genome sequencing have far outpaced the throughput in functional follow-up studies. To bridge the gap between genetic and functional information, I systematically phenotyped a human mutation open reading frame (hmORF) collection using immunofluorescence and high-content automated confocal microscopy and discovered 258 mislocalized variants, indicating that mislocalization plays an important role in Mendelian disease pathology. 54 unique mislocalization patterns were observed. The most common type of mislocalization was from the late secretory pathway (plasma membrane, Golgi apparatus, or vesicles) to the ER (61%). It is possible that many of the mislocalized secretory pathway variants fail to pass quality control in the ER and are retained and sorted for degradation. To support this rationale, I performed BioID on 16 ER-retained variants and discovered that all 16 variants associated more with ER quality control factors than their wild-type counterparts. To elucidate the degradation pathways of the ER-retained variants, I used immunofluorescence to measure protein abundance of each ER-retained variant after inhibition of the proteasome, autophagy, and VCP. I discovered that many ER-retained variants are degraded by the proteasome (31%) or by both the proteasome and autophagy (17%). Interestingly, half of the ER-retained variants were unaffected by inhibition of proteasome or autophagy. In this study, I show that protein mislocalization and protein degradation are key molecular mechanisms in human disease. The data collected from this study, combined with previously generated data on protein-protein interactions, protein-DNA interactions and protein stability, will generate a unique, publicly available database for molecular phenotypes in Mendelian disease. Importantly, this resource will provide a platform for future efforts in drug discovery, and more specifically, for correcting the localization pattern of ER-retained variants with pharmacological chaperones. The database for molecular phenotypes in Mendelian disease will be a transformative and invaluable resource for the scientific and medical community.

Keywords:

Navigating protein fitness landscapes in multiple dimensions

Presenter: Christian Landry (Université Laval)

Christian R Landry

Université Laval

Proteins could have as many fitness landscapes as they have functions. The relationship between mutations and the performance of any two functions could be positively or negatively correlated, making evolution on one landscape contingent on evolution on the other. Understanding the evolution of proteins therefore requires that we understand how mutations affect the multiple dimensions of the landscapes of protein functions. Here, I will share our recent work on the deep mutational scanning of two drug resistance enzymes in which we examine more than one dimension at a time. In one project, we examine the trade-off between resistance to an antifungal drug and activity on the canonical substrate of the drug target. We show that the trade-off between the two phenotypes is very steep and is shaped by the difference in dose response functions of drug resistance and nutrient use. The two fitness functions are nearly perfectly negatively correlated, which means no single mutation can lead to resistance while also maintaining growth in conditions that require this enzyme. In the second project, we examine how expression level affects the fitness landscape of an antibiotic resistance protein. By measuring the effect of all single mutations at optimal and several suboptimal expression levels, we show that optimal expression can mask the effects of many deleterious mutations on enzyme function. Surprisingly, some beneficial mutations also appear to be masked at optimal expression level. Protein expression level is therefore an important factor shaping the fitness landscape of proteins. These two studies illuminate how complex protein evolution is in heterogeneous growth and cellular environments.

Keywords: Drug resistance; Evolution; Trade-off; Enzymes;

Understanding the origins of loss of protein function using analyses of protein stability and conservation

Presenter: Kresten Lindorff-Larsen (University of Copenhagen)

Kresten Lindorff-Larsen

Linderstrøm-Lang Centre for Protein Science, University of Copenhagen, Denmark

We have used computational and experimental approaches to determine the consequences of missense variants in proteins, with the aim of using such models both for diagnosing genetic diseases, and for providing mechanistic insight into disease. In particular, we have focused on the effects of individual amino acid changes on protein folding and stability, linking biophysical calculations with protein degradation and abundance in cells. By examining a range of proteins and diseases we have found that loss of stability and proteasomal degradation is a common origin for loss of protein function, and that predictions of changes in thermodynamic protein stability are useful to assess the pathogenicity of genetic variation. I will discuss these ideas using recent examples from our laboratories. At the same time, our work has also revealed areas where our understanding and ability to predict the effect of amino acid changes is still imperfect. I will discuss how we are using sequence analyses and assays of variant effects experiments to understand the origins of loss of function, thus paving the way for more accurate biophysical models and machine learning methods for use in personalized medicine.

Keywords:

Investigating evolutionary pathways using combinatorial libraries

Presenter: Jing Liu (University of Toronto)

Jing Liu (1), Steven K. Chen (1), Belinda S.W. Chang (1,2,3)

Department of Cell & Systems Biology (1), Department of Ecology & Evolutionary Biology (2), Centre for the Analysis of Genome Evolution & Function (3)

To what extent neutral mutations affect protein evolution remains controversial and is partially attributed to the lack of experimental tools available to evolutionary investigators. Indeed, advances have been made in developing statistical evolutionary models to infer ancestral sequences using extant sequences. Testing of these inferred ancestral pathways has provided much of the existing data that continues to fuel the debates concerning evolutionary trajectories. However, in almost all cases, these studies of ancestral reconstruction have focused on a very limited number of reconstructed sequences, often restricted to key nodes along a phylogenetic tree. To investigate pathways of molecular evolution more comprehensively requires testing many more sequences to identify mutations that are deleterious, neutral, or confer functional gains.

Here, we investigate four sites in rhodopsin, a sensory protein important for organismal physiology and behavior. It has been proposed that these four sites co-evolved in a functional trade-off between rhodopsin thermal and active state stability. Howe ver, to date only a few pathways have been investigated.

To exhaustively investigate alternative pathways, we constructed a rhodopsin gene variant library where these four sites were combinatorially mutated using a NNK mutagenesis scheme. Since the number of variant sequences is enormous (1.05 x 106 sequences), we used a method from statistical theory to approximate that, assuming random sampling, sampling 1.52 x 107 s equences is required to achieve full coverage. We then conducted an in silico simulation that not only confirmed the statistical approximation but also revealed that sampling requirements increase exponentially as coverage approaches 100%. This infor mation will be invaluable for overcoming technical bottlenecks in deep scanning mutagenesis studies involving combinatorial l ibraries. Future directions include functional characterization of our gene variant library by deep scanning mutagenesis using an engineered yeast system developed to measure rhodopsin activation.

Keywords: Protein evolution, yeast genetics, in silico simulations

Systematic proactive variant effect maps for CPOX

Presenter: Warren van Loggerenberg (University of Toronto)

Warren van Loggerenberg(1,2,3,4), Aditya Chawla(1,2,3), Jochen Weile(1,2,3,4), Marinella Gebbia(1,2,3), Song Sun(1,2,3,4,*) and Frederick P. Roth(1,2,3,4,5)

1Donnelly Centre, University of Toronto, Toronto, Ontario, Canada; 2 Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada; 4Department of Computer Science, University of Toronto, Toronto, Ontario, Canada; 5Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, Massachusetts, USA; *Present address: Sanofi, Inc., Toronto, ON, Canada.

Genome sequencing is a tool for diagnosing hereditary coporphyria (HCP), which can be caused by sequence variation in coproporphyrinogen oxidase (CPOX). However, sequence-based diagnosis is severely limited by the difficulty of distinguishing tolerated from damaging variants. Although more prevalent variants can be statistically associated with disease, most of the missense variants in CPOX (52.6% of those in ClinVar) have been interpreted as “variants of uncertain significance” (VUS). Here, we validate a yeast-based functional complementation assay for CPOX missense variation. Using saturation codon-mutagenesis, en masse growth selection, and sequencing, we applied the assay to all possible amino acid substitutions in CPOX, thus generating a comprehensive ‘variant effect map’. High-quality measurements of functional impact were obtained for 84% of all possible amino-acid substitutions, and for >90% of substitutions that are accessible by a single-nucleotide change. Importantly, we demonstrated the potential clinical utility of the CPOX variant effect map, showing that it can outperform all assessed computational methods in reliably identifying pathogenic missense variants. We show that yeast-based functional complementation assays for heavy metal tolerance can identify variants of CPOX associated with an atypical metal-induced porphyrinogenic response in humans. Thus, we are proceeding to generate an atlas of the impacts of CPOX variants across environmental contexts to enable more rapid and sensitive genetic diagnosis of HCP, support therapeutic intervention for an even larger number of individuals, and perhaps allow a gateway for ‘genome-informed career counselling’.

Keywords:

Using high-thoughput assays to improve the interpretation of rare variants in the context of the dystroglycanopathies

Presenter: Kaiyue Ma (Yale University)

Kaiyue Ma(1), Shushu Huang(1), Nicole Lake(1) , Keryn Woodman(1), Angela Lek(1), Monkol Lek(1)

Yale University

Dystroglycanopathies are caused by defective glycosylation of Alpha-dystroglycan (Alpha-DG). Pathogenic variants in related enzymes are typically ultra-rare with many unique to affected families. This results in challenges in interpreting pathogenicity, where typically over 50% of variants are variants of uncertain significance (VUS). We aim at developing a high-throughput functional assay to characterize all possible variants of these enzymes with FKRP as our initial focus. We developed a platform cell line by knocking out FKRP and overexpressing Alpha-DG in the HAP1 cell line. Employing the IIH6C4 antibody that detects glycosylated Alpha-DG, we developed a flow-cytometry based assay to characterize the functional impact on Alpha-DG glycosylation. We extended this assay to be high throughput by building a lentivirus pool carrying all possible FKRP single nucleotide variants (SNVs) to transduce our platform cell line followed by sorting cells into high and low Alpha-DG glycosylation groups, which were then examined by Next Generation Sequencing (NGS). The highlight of our saturation mutagenesis strategy is that We developed a 7-step unbiased and cost-effective cloning work flow, which introduces the variants in pooled PCR primers and employs two-way extensions to generate the full-length gene. In our evaluation, no transition/transversion bias or variant location bias was observed. This workflow uses only common reagents and equipment and can be easily applied to genes of all sizes. We developed a computational pipeline to determine the function score of each variant based on its enrichments in high and low groups. We characterized 99% of all possible FKRP SNVs and our result showed the expected trend for nonsense variants to be the most damaging. Among the missense variants, known pathogenic and benign variants clustered closer to nonsense variants and synonymous variants, respectively. Thus, we anticipate our function scores can be used as another line of evidence to interpret VUS. In summary, here we report a high-throughput functional assay whose experimental workflow and analytical pipeline can be adapted and applied to all the Alpha-DG glycosylation enzymes. Future work will involve validation by orthogonal functional assays independent of the IIH6C4 antibody.

Keywords: saturation mutagenesis; high throughput; functional assay; FKRP; dystroglycanopathies

Development of deep mutational scanning reporter cell lines with a wild-type to mutant switch and flexible reporter read-outs.

Presenter: Stefanie Maes (VIB-UGent Center for Medical Biotechnology)

Stefanie Maes (1), Nick Deploey (1), Karolien De Bosscher (1), Frank Peelman (1)

(1) VIB-UGent Center for Medical Biotechnology

We used two medically important intracellular drug targets, namely the glucocorticoid receptor (GR) and p53, as relevant protein paradigms to develop deep mutational scanning reporter cell lines.

We have developed HEK293 cell lines for deep mutational scanning of the GR protein that use a FACS-based read-out. In this way, we determine the effects of GR mutations on both gene activation and gene repression. Development of cell lines with the proper reporter gene response requires stable introduction of one copy of the wild-type GR protein, coupled to the fluorescent protein tdTomato via a self-cleaving peptide. The reporter gene is introduced via the PiggyBac transposon system. After selecting a clonal cell line with the correct response, a GR mutant library can be introduced using the Bxb1 integrase. In

each cell, one GR mutant displaces the wild-type GR and tdTomato from their promoter, resulting in the loss of tdTomato fluorescence and in a switch in expression from wild-type GR to a mutant version. Random integration of additional GR mutants is monitored via a CD20 surface expression marker. The system nicely replicates the gene activation profile of well-characterized GR mutants.

Similarly, we have developed HEK293 cell lines for deep mutational scanning using the MAPPIT mammalian two-hybrid protein-protein interaction assay. As an example, we determine the effects of mutations in a p53 MAPPIT bait on its interactions with MDM2 and MDM4 MAPPIT prey proteins using a STAT3-induced reporter gene. The MAPPIT system properly detects mutations that affect the p53-MDM2 and p53-MDM4 interactions. Development of the cell lines with a proper STAT3 reporter response for deep mutational scanning requires expression of the wild-type p53 bait protein. A clonal cell line with the proper STAT3 reporter response is developed in a similar fashion as described above and not only contains the wild-type p53 bait protein coupled to tdTomato, but also one of the prey proteins. A wild-type to mutant switch is again accomplished using Bxb1 integrase.

In our FACS-based reporter assays, we use the alpha subunit of the IL-5 receptor without its intracellular tail as the reporter. In addition, cells that contain multiple mutants by the means of random integration can be detected through a CD20 surface marker. While detection of CD20 and IL-5 receptor alpha surface expression via FACS requires the use of fluorescently labeled antibodies, this offers flexibility. It facilitates the use of more labels in a single FACS experiment and allows to easily adapt the fluorescent labels to match a specific cell sorter. Moreover, it allows the enrichment or depletion of cells via MACS if required.

We conclude that the wild-type to mutant switch in our deep mutational scanning system is useful when expression of the wild-type protein of interest is required for the development of proper reporter cell lines. In addition, our implementation of surface expression markers in the system offers several advantages.

Keywords: FACS, PiggyBac, Bxb1, HEK293, protein-protein interaction, mammalian two-hybrid, MAPPIT, p53, glucocorticoid receptor

Identification of enhancers to rescue haploinsufficiency in neurodevelopmental disorders via a CRISPRa-QTL approach

Presenter: Troy A. McDiarmid (Department of Genome Sciences, University of Washington School of Medicine, 15th Avenue NE, Seattle, WA, USA)

Troy A. McDiarmid (1,8), Florence M. Chardon (1,8), Nick Page (3,4,5,8), Beth Martin (1), Silvia Domcke (1), Samuel G. Regalado (1), Jean-Benoît Lalanne (1), Diego Calderon (1), Lea M. Starita (1,6), Stephan J. Sanders (3), Nadav Ahituv (4,5), and Jay Shendure (1,2,6,7)

1. Department of Genome Sciences, University of Washington School of Medicine, 15th Avenue NE, Seattle, WA, USA
2. Howard Hughes Medical Institute, Seattle, WA, USA
3. Department of Psychiatry, Kavli Institute for Fundamental Neuroscience, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA
4. Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA
5. Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA
6. Brotman Baty Institute for Precision Medicine, NE Pacific Street, Seattle, WA, USA
7. Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA
8. These authors contributed equally to this work

Neurodevelopmental disorders encompass a broad spectrum of conditions including Autism Spectrum Disorder and Intellectual Disability. As a result of rapid advances in clinical sequencing technology, thousands of variants in hundreds of genes have recently been definitively associated with an increased risk for developing neurodevelopmental disorders. For the majority of cases where a genetic etiology is known, the pathogenic mechanism is thought to be haploinsufficiency, having only one functional gene copy. Recent studies using cis regulation therapy (CRT), whereby the existing functional copy is upregulated using CRISPR activation (CRISPRa) targeting its regulatory elements, have shown to be a viable treatment option. However, to date only a handful of genes and regulatory elements have been tested due to limitations in throughput and lack of a detailed understanding of the regulatory circuitry controlling neurodevelopmental disorder risk genes. Here, we develop CRISPRaQTL screening, a high-throughput framework that combines CRISPRa perturbations with single-cell RNA sequencing to identify regulatory elements capable of potently and specifically upregulating target genes. We present results for a 494 guide, 40,000 cell experiment targeting known enhancers and putative promoters of neurodevelopmental disorder risk genes in K562 cells. CRISPRaQTL is capable of capturing target gene upregulation from single cell transcriptomes when targeting both promoters and enhancers. Current experiments are aimed at moving this CRISPRaQTL framework into iPSC-derived human neurons and scaling screening efforts to a comprehensive list of all putative regulatory elements controlling a list of all high-confidence neurodevelopmental disorder risk genes. This work establishes CRISPRaQTL as a scalable single cell framework to screen

promoters and enhancers, identifies multiple functional guides targeting neurodevelopmental disorder risk gene promoters for use in CRT, and contributes to our understanding of the regulatory circuitry of neurodevelopment.

Keywords: Multiplex single-cell CRISPR screens, Neurodevelopmental disorders, cis-regulatory element screening, enhancers, cis-regulatory therapy, crisprQTL

Probing PAX6-DNA Interactions Using High-Throughput Yeast One-Hybrid Coupled with Deep Mutational Scanning

Presenter: Alex McDonnell (University of Edinburgh)

Alex McDonnell, Grzegorz Kudla, David FitzPatrick, Joe Marsh

University of Edinburgh

PAX6 is a highly conserved transcription factor essential for the correct development of the central nervous system, the p ancreas, and the eye. Heterozygous deletions, nonsense and frameshift mutations are generally well characterised as causin g aniridia, while most missense variants produce a broad range of other discrete ocular pathologies. The interplay between P AX6 and its DNA targets is complicated by multiple functionally distinct subdomains, co-factors, high binding site promiscuity, and divergent spectra of disease phenotypes. Knowledge of the contributions made to binding by each residue and the impac t of missense variants is key for understanding the role of PAX6 mutations in disease. Current methods of exploring this intera ction grammar have been limited by relatively low-throughput techniques that are resource intensive and can only feasibly be performed on a handful of residues.

Here I used a combination of yeast one-hybrid and deep mutational scanning in competitive growth assays to probe the fu nctional consequences of almost all the possible single amino acid variants in the paired domain of PAX6. Initial findings sho w clustered regions of the paired domain that are sensitive to mutation, that correlate broadly with secondary structural eleme nts. Additionally, residues that contribute to DNA binding are overwhelmingly detrimental, and generally intolerant to mutation irrespective of the type of substitution. Finally, most clinically identified pathogenic variants appear to disrupt binding, with the exception of certain variants at the termini of the paired domain, and variants located in the linker region. The latter region als o demonstrates a consistent observation of improved binding at specific residues across multiple substitutions. It is hoped tha t this deep mutational scan of PAX6 will aid in the modelling of existing and novel variants, and in the development of in-silico methods for pathogenicity prediction.

Keywords: PAX6, Deep Mutational Scanning, DMS, Multiplexed Assays of Variant Effect, MAVE, MAVEs, Variant, Edinburgh

Evaluation of an online novel course ‘Female Genital Mutilation (FGM): Health, Law, and Socio-Cultural Sensitivity’.

Presenter: Zainita Meherally (University of Aberdeen)

Zainita Meherally (1), Heather Morgan(1,2), Victoria Kinkaid (1,2), Grace Manji (1), Sulemanas Saibu(1)

1- university of Aberdeen 2- Futurelearn

Abstract Background An online course: ‘Female Genital Mutilation (FGM): Health, Law, and Socio-Cultural Sensitivity’ was developed and delivered by multidisciplinary medics/researchers/teachers from the University of Aberdeen via the FutureLearn platform. It was designed for health and social care professionals, educators, police, and anyone with an interest in women’s health. This study aimed to understand different aspects of FGM covered in the course and evaluate the delivery of and responses to the course. Methodology The course used videos, articles, audio, and practical activities. The course was run for the first time this year, between 08/Feb (Zero Tolerance for FGM Day) and 08/Mar (International Women’s Day). A team of MSc Global Health students evaluated the course from participant perspectives shared in the comments section, activities, and online surveys. Results Posts in the comments section indicate a gap in knowledge and eagerness to learn. Participant numbers rose from 100+ at the beginning (week-1) to 400+ by the end of the course (week-4). 88 participants engaged in discussion boards. However, the participation in the comments section dropped from 793 comments (week-1) to 209 comments (week-4) during the course. 199 participants completed tests and 71% answered correctly. Engagement with the course was high. The team also received a request from the UK Association of Professional Piercers to reflect more on consensual and legal piercing of women’s genitalia, classified as type 4 FGM and will update content with their support. Conclusion The course coordinators actively seek feedback to make content current, relevant, and applicable in different settings. The evaluation feedback will improve the next run in 2023.

Keywords: female genital mutilation, Futurelearn

Systematic assessment of allostery in a model GPCR

Presenter: Taylor Mighell (Centre for Genomic Regulation)

Taylor L. Mighell (1), Ben Lehner (1, 2, 3)

(1) Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain
(2) Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain
(3) Universitat Pompeu Fabra (UPF), Barcelona, Spain

G-protein coupled receptors (GPCRs) are biologically and medically relevant as they bind thousands of endogenous and synthetic ligands (including a third of FDA approved drugs), and transmit information allosterically across the cell membrane. Allostery is a phenomenon whereby perturbation at one site of a molecule causes some distal effect. GPCRs activated by "balanced" agonists engage two canonical signaling pathways through G-proteins and β-arrestins, while "biased" agonists preferentially signal through one pathway. Because of their biological and medical relevance, and the central role of allostery in their function, GPCRs are an attractive model for answering fundamental questions. However, they are recalcitrant to traditional biochemical approaches as they are large, transmembrane, and dynamic. Decades of work in structural biology have identified some active or inactive GPCR structures, but these snapshots are ill-equipped to identify allosteric mechanisms in such dynamic proteins. As an alternative approach, we are developing deep mutational scanning assays to ascertain high resolution functional data for all single and a targeted set of double missense mutations in beta-2 adrenergic receptor (b2AR), a model class A GPCR with established pharmacological tools, including balanced and biased agonists. First, we established an assay to measure the surface abundance of b2AR with an N-terminal (i.e. extracellular) HA-tag followed by fluorescent antibody staining and sort-seq. Measuring surface abundance enables disambiguation of effects on folding/trafficking from signaling. Application of this assay to a pilot library of ~4,000 single and double mutants showed high reproducibility (average replicate Pearson r=0.88) and complete separation between early-truncating and wild-type synonyms. Next, we adapted a previously developed assay to measure ligand-dependent signaling, which relies on 3 genetically encoded components: 1. A fusion of b2AR and the Gal4 transcription factor, linked by a protease cleavage site. 2. A signal transducer (e.g. β-arrestin) fused to a protease. 3. Gal4-UAS response element upstream of a transcriptional barcode. Agonist binding activates the receptor, recruiting the signal transducer and bringing protease into proximity with the cleavage site resulting in liberation of Gal4, which moves to the nucleus and drives transcription of the reporter barcode. RNA-seq is then used to quantitate barcode abundance which reports on signaling activity. An ideal platform would enable readout through both canonical signaling pathways. Toward this end, we validated the use of both β-arrestin and Nb80, a G-protein-mimic nanobody that specifically binds to the active form of b2AR. Since we measure receptor abundance and signaling through both canonical pathways, these data will highlight all residues important for allosteric transmission as well as those with differential contribution. Repeating the experiment with balanced and biased agonists or in the presence of allosteric modulators will add information that could uncover mechanisms of biased agonism and inform future efforts to design drugs with improved signaling characteristics. Double mutants will reveal genetic interactions between residues, which could also help define allosteric networks. The framework developed here should be applicable to many GPCRs and could represent a general paradigm for understanding allosteric proteins.

Keywords: GPCR, allostery, epistasis

Functional analysis of homologous recombination genes

Presenter: Larissa Milano (Université Laval - CHU de Quebec)

Larissa Milano*, Gemma Montalban*, Charles Joly-Beauparlant, Amélie Rodrigue, Martine Dumont, Yan Coulombe, Penny Soucy, Arnaud Droit, Jacques Simard(=), and Jean-Yves Masson(=)

CHU de Québec – Université Laval, Québec City, QC, Canada

- These authors contributed equally (=) co-corresponding authors

Homologous recombination (HR) is an essential pathway to accurately repair DNA double-strand breaks (DSBs). HR deficiency, forces cells to rely on mutagenic DSB repair pathways for survival, leading to genomic instability and cancer. BRCA1 and BRCA2 are the most altered HR genes in hereditary breast and ovarian cancer, while PALB2 and RAD51C are emerging cancer predisposition genes. Inherited mutations in PALB2 are associated with a predisposition for ovarian, breast and pancreatic cancers. In particular, it has been reported that mutations in PALB2 increase risk of breast cancer up to 8-9 fold. PALB2 was identified as a BRCA2 interacting protein, essential for BRCA2 anchorage to nuclear structures and for its function in DSB repair. We have shown that RAD51C is at the heart of two RAD51 paralogs complexes, RAD51C-XRCC3 and RAD51B-RAD51C-RAD51D-XRCC2. Inherited variants which result in the truncation of the RAD51C and RAD51D proteins, but not variants in the other paralogs, have been linked to a predisposition to ovarian cancer. Our laboratory focuses on both the regulation and activities of PALB2 and the RAD51 paralogs during the DNA damage response and the effect of cancer-causing mutations. Primarily, we work on deciphering the functions of PALB2 and the RAD51 paralogs in HR. Due to the challenges of predicting the functional consequences of HR variants and their different biological effects, we have developed imaging and

biochemical tools for the functional assessment of single variants in HR genes. We are currently implementing deep mutational scans for regions of PALB2/RAD51 paralogs enriched in predicted deleterious missense mutations. Altogether, these approaches will improve functional assessment of missense variants in HR genes and facilitate clinical and therapeutic management of carriers.

Keywords: PALB2, RAD51C, Breast and ovarian cancer, Homologous recombination

Interrogating protein variant dysfunction using multidimensional proteomics

Presenter: Sophie Moggridge (University of Washington, Seattle, WA, USA)

Sophie Moggridge (1), Ricard Rodriguez-Mias (1), Kyle Hess (1), Judit Villén (1)

1. Center for the Multiplexed Assessment of Phenotype, Department of Genome Sciences, University of Washington, Seattle, USA

Protein variants caused by single amino acid substitutions contribute to a large portion of human diseases. The molecular mechanisms of these diseases are complex because substitutions can have varied effects on protein properties such as stability, activity, conformations, localizations, and interactions. High-throughput mutagenesis methods like deep mutational scanning1,2 have accelerated variant phenotyping. However, these methods do not measure protein directly and thus are limited in the protein properties they can measure. Conversely, classical biochemical assays allow direct measurement of protein properties but require protein purification and are low throughput. To bridge this gap, we developed mass spectrometry (MS) methods that enable high-throughput analysis of many protein variants simultaneously. Biochemical assays that assess protein properties such as solubility, stability, post translational modifications, binding etc. are MS-compatible and will facilitate multidimensional analysis of mutational libraries. Implementation of these assays will greatly enhance our understanding of how single amino acid mutations affect function, and will contribute to better classifying, understanding, and treating disease. Here I present the application of proteomic methods to determine molecular consequences of disease causing variants of human phosphoglucomutase-1 (PGM1). PGM1 was selected because it is a cytosolic, monomeric, and highly soluble protein for which the molecular effects of many variants are known. We focused on 13 missense variants known to alter a fundamental biochemical properties of PGM1, and coupled this library with three biochemical assays informative of protein function with a mass spectrometry-based readout. Solubility, thermal stability, and small molecule binding assays were performed on the PGM1 library that was recombinantly expressed in E. coli. Mass spectrometric analysis enabled the detection of all variants from the same lysate sample. The selection assays were benchmarked against a previous data set of low-throughput readouts for individual variants. The MS-based solubility assay was more quantitative than the previous method and achieved improved resolution for classification and ranking of protein variant solubilities. With the thermal stability assay, we were able to determine melting temperatures for six variants (N38Y, Q41R, G330R, E377R, E388K, L516P) which could not previously be measured using thermal denaturation circular dichroism. Use of MS-based orthogonal biochemical assays facilitates mechanistic interpretation of variant effects such as with D62H, which has wild-type-like thermal stability but is poorly soluble. Lastly, to evaluate PGM1 variants for small molecule binding, we performed a thermal stability assay in the presence of glucose 1-phosphate. Our results successfully recapitulate published data from low-throughput assays. We have demonstrated the utility of MS-based biochemical assays that measure fundamental protein properties for the purpose of mutational scanning. The modular biochemical selection assays will contribute multi-dimensional mechanistic understanding to both novel and pre-existing mutational libraries.

References: (1): Matreyek KA, et al. Nat Genet. 2018. (2): Findlay, Gregory M., et al. Nature 2018.

Keywords: proteomics, stability, binding, solubility, PGM1

How to use the principles of deep mutation scanning without doing mutational scanning

Presenter: Molly Monge (National Human Genome Research Institute, National Institutes of Health)

Molly Monge (1), Apoorva Ravi Shankar (1), Meru Sadhu (1)

Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

Improving the reproducibility and statistical power of molecular biology experiments is of key concern. We are developing a high-throughput framework that will enhance reproducibility in molecular biology experiments by using hundreds to thousands of experimental replicates while minimizing confounding factors. We will use this framework to study the nucleotide excision repair (NER) factor-I complex, which repairs bulky DNA lesions. In yeast, the NER factor-I complex is required for the survival of cells exposed to genotoxic drugs such as methyl methanesulfonate (MMS) and is encoded by the genes RAD1, RAD10, and RAD14. Homologs of these genes also exist in humans, named ERCC4, ERCC1, and XPA, respectively. While the human

genes have been individually shown to function in yeast, the degree to which different combinations of yeast and human proteins can perform nucleotide excision repair is unknown. Using thousands of replicates, we will precisely quantify how interchangeable these proteins are. In our approach, we generate thousands of distinct, DNA-barcoded plasmids, each carrying one of 27 possible combinations of yeast and human NER genes. We will precisely determine the functionality of the 27 NER complex combinations by measuring the abundances of their many representative barcodes before and after MMS exposure. We see enormous potential for pooled barcoding to improve the reproducibility of everyday experiments.

Keywords:

Atlas of Variant Effects (AVE) Alliance: A year in review

Presenter: Lara Muffley (Department of Genome Sciences, University of Washington, Seattle, WA USA)

AVE Alliance Members, AVE Executive Committee, Lara Muffley (1)

(1) Genome Sciences, University of Washington, Seattle, WA USA

The Atlas of Variant Effects is an international collaborative effort with a mission to propel systematic measurements of variant impact on functional elements of human and pathogen genomes, towards diagnosing and treating human disease and understanding genes, gene products and their regulation.

The Alliance (which was formed in 2020) serves as a key forum to catalyse the assembly of a nucleotide-resolution atlas (collection of comprehensive variant effect maps) by expanding the community of scientists who are developing and applying experimental and analytical technologies to generate variant effect maps. Here we outline our near and long term goals, current initiatives and summarize progress to date.

Keywords: alliance, multiplexed assays of variant effect, variants of uncertain significance, functional genomics, atlas, consortium

Deep mutational scan of a cardiac potassium channel gene, KCNE1

Presenter: Ayesha Muhammad (Vanderbilt University Medical Center)

Ayesha Muhammad (1,2), Tao Yang (3), Marcia A. Blair (3), Lynn D. Hall (3), Kenneth A. Matreyek (4), Douglas M. Fowler (5), Dan M. Roden (1,3,6,7), Andrew M. Glazer (1,3)

1. Vanderbilt Genetics Institute, Vanderbilt University Medical Center 2. Medical Scientist Training Program, Vanderbilt University 3. Department of Medicine, Vanderbilt University Medical Center 4. Department of Pathology, Case Western Reserve University School of Medicine 5. Department of Genome Sciences, University of Washington 6. Department of Biomedical Informatics, Vanderbilt University Medical Center 7. Department of Pharmacology, Vanderbilt University Medical Center

Introduction: KCNE1 encodes a 129-amino acid transmembrane subunit that interacts with KCNQ1 to conduct IKs, a potassium current important for cardiac repolarization. Loss-of-function KCNE1 variants can cause long QT syndrome and lead to fatal arrhythmias. Objective: We conducted two different massively parallel assays on a comprehensive KCNE1 library to measure cell-surface abundance and K+ influx-dependent cell death. Methods: To assess surface expression, an extracellular HA-tag was cloned in KCNE1. We used degenerate primers in inverse PCRs to create and barcode a library of KCNE1 variants. For the surface abundance assay, conducted in triplicate, the library was expressed in a single-variant-per-cell format using a Bxb1 landing pad in HEK293 cells, which also stably expressed wild-type KCNQ1. Cells were stained with a fluorescent anti-HA antibody and FACS-sorted into 4 equal groups of increasing surface abundance. A weighted average of variant frequency across the four groups, measured by high-throughput sequencing of barcodes, was calculated. Scores were linearly transformed such that the median of synonymous variants was 1 and of early nonsense variants was 0. To assess K+ influx, we used landing-pad cells that stably expressed a gain-of-function KCNQ1 variant, S140G. Unlike wild-type IKs channels, which are closed at the resting membrane potential of HEK293 cells, KCNQ1 S140G biases IKs channels into the open state. Thus, functional IKs channels allow constant cellular K+ influx that decreases cell fitness. S140G cells expressing the library were grown for two weeks, and high-throughput sequencing performed at day 10. We quantified cell survival using a log2-transformed ratio of variant frequency in cells to that in the original library, linearly transformed as above. Results: The library represented 88.1% of 2,451 possible missense KCNE1 variants. Surface abundance scores, obtained for 1,886 missense variants, were normally distributed (0.93±0.37, mean±SD). Patterns of decreased surface abundance were observed as follows: in the extracellular domain, most variants at positions 35-38; in the transmembrane domain, most polar or charged variants at positions 47-59; and in the cytoplasmic domain, most aromatic or larger aliphatic variants at positions 79-114. Nonsense variants before residue 57 (early nonsense) reduced surface expression, but those after residue 57 (late nonsense) had WT-like scores, indicating that KCNE1's latter half is dispensable for surface trafficking. We also identified variants with a gain-of-function phenotype. Most variants at positions 61-65 had increased trafficking. Channel function for 4 novel and 4

previously-studied variants, measured by manual patch clamp, correlated highly with trafficking scores (ρ=0.71, p=0.08). For the K+ influx assay, late nonsense variant scores (0.30±0.44) were more similar to early nonsense variant scores (0.09±0.22) than to synonymous variant scores (0.99±0.27), indicating that the late nonsense variants likely disrupt KCNQ1/KCNE1 interaction to alter IKs channel function. Missense variants showed a bimodal distribution (median=0.75, SD=0.57). Conclusion: We conducted a deep mutational scan to interrogate ~2,000 variants in a clinically important ion channel subunit, KCNE1. To validate these results, we will conduct more replicates of the S140G assay. Our data provide structural insights and functional information to facilitate clinical interpretation of KCNE1 variants.

Keywords: Arrhythmia, Ion Channel, Cell Surface Abundance, Fitness assay

Molecular determinants of Hsp90 recognition of Src kinase revealed by deep mutational scanning

Presenter: Vanessa Nguyen (University of Washington)

Vanessa Nguyen (1, 2), Ethan Ahler (1), Douglas Fowler (1)

1: Department of Genome Sciences, University of Washington, 2: Department of Bioengineering, University of Washington

Hsp90 is a molecular chaperone involved in the conformational and regulatory control of numerous signaling proteins, and more than 60% of human kinases. However, the mechanistic and biochemical details of how Hsp90 recognizes its diverse array of client proteins remains poorly understood. Here, we utilize deep mutational scanning of the Src catalytic domain in the presence of the Hsp90 inhibitor radicicol to determine the molecular determinants of Hsp90 client recognition. Our data showed a striking correlation between Src kinase activity and client strength. Hydrophobic variants in the αF pocket and β1-β2 strands, regulatory regions that autoinhibit Src activity, led to hyperactivation of Src and a strong functional dependence on Hsp90. Interestingly, variants at the other regulatory regions, the SH2/CD and SH3/CD interfaces, were not nearly as activating nor dependent on Hsp90. Variants with similar levels of hyperactivation had divergent client dependencies that corresponded to the opening of the global kinase conformation through mutations in Src regulatory regions. Other factors previously suggested to drive Hsp90 recognition like surface hydrophobicity and stability of the kinase had little correlation with client status. Thus, we show that Hsp90 recognition is a function of both kinase hyperactivity and the global conformation of the kinase. Our approach, which was able to identify biochemical trends in hotspots for Hsp90 recognition and the subsequent functional outcomes, could be broadly applied to study other aspects of protein quality control in the cell.

Keywords: protein chaperones, protein folding, kinase, Hsp90

Investigating drug resistance mechanisms using multiplexed assays and structural stability calculations

Presenter: Sofie Vincents Nielsen (University of Copenhagen)

Sofie V. Nielsen (1), Henrike Zschach(1), Aleksandra Panfilova (1), Matteo Cagiada(2) and Amelie Stei(1)

1: Section for computational and RNA biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark 2: Linderstrøm-Lang Centre for Protein Science , Department of Biology, University of Copenhagen, Copenhagen, Denmark.

Understanding drug resistance has numerous implications for human health. Comprehensive, systematic profiling of drug resistance can significantly improve our ability to understand, and with time, predict potential resistance trajectories. Here, we present a multiplexed assay of variants effect (MAVE) experiment of human dihydrofolate reductase (DHFR) in which a library of DHFR variants are scored for their ability to sustain growth in the presence of the competitive inhibitor methotrexate (MTX). Cells expressing non-functional DHFR variants are depleted from the culture while WT-like variants remain, and possibly, MTX-resistant variants will increase in the population. Interestingly, we observe a substantial number of variants that appear beneficial compared to WT DHFR. Known resistant variants at positions L23 and F32 are identified as resistant in our screen, and two variants commonly found in the population, N30K and M140L, are, as expected, WT-like in the MAVE. When comparing our MAVE data to the computationally predicted loss of stability or evolutionary sequence energy, we surprisingly find that MTX resistant variants are predicted to be highly unfavored by nature and some appear to gain their drug resistance by compromising their stability.

Keywords: Protein-Drug interaction, drug-resistance, structural biophysics, dihydrofolate reductase

Impact of the Endoplasmic Reticulum Membrane Protein Complex (EMC) on the Mutational Tolerance of Rhodopsin

Presenter: Karen Noguera (Indiana University)

Karen Noguera (1), Andrew G. McKee (1), Charles P. Kuntz (1), Wesley D. Penn (1), Jonathan P. Schlebach (1)

(1) Indiana University

The proteostasis network features a spectrum of partially redundant chaperones and insertases that assist nascent membrane proteins (MPs) as they establish their orientation within the membrane (topology) and form their native structure. However, it remains unclear how preferential chaperone interactions are specified by the molecular features of MP substrates and how mutations can remodel these interactions. Recent investigations of the endoplasmic reticulum membrane protein complex (EMC) suggest it facilitates the cotranslational membrane insertion of the first transmembrane domain (TMD) of certain GPCRs and that it can function as a holdase for misassembled TMDs in polytopic MPs. To explore the nature of these interactions, we have employed deep mutational scanning to compare the plasma membrane expression of ~1,200 rhodopsin variants in HEK293T cells and in ΔEMC6 HEK293T cells. Consistent with emerging perspectives, trends within TMD1 suggest EMC dependence is sensitive to changes in solvation energetics. However, we also find that EMC impacts the expression of numerous TMDs 2 & 7 variants. A statistical analysis shows the depletion of EMC enhances the expression of temperature- and retinal-sensitive variants that disrupt a native helical kink in TMD2 or perturb the topological energetics of TMD7. Together, these findings reveal how EMC engages this substrate and how this relates to its impact on mutational tolerance.

Keywords:

pyCROQUET: Crispr Read to Oligo QUantification and Evaluation Tool

Presenter: Victoria Offord (Wellcome Sanger Institute)

Keiran Raine (1), Emre Karakoc (1), Victoria Harle (1), Adam Butler (1), David J. Adams (1), Victoria Offord (1)

Wellcome Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK

Well established tools exist for the quantification of single guides in CRISPR screens. However, these software packages are poorly adapted to computing the abundance of long, highly similar reads such as those generated by saturation genome editing (SGE). Furthermore, no tool exists to determine and assess the abundance of guides in complex combinatorial CRISPR screens.

We present pyCROQUET, a python Crispr Read to Oligo QUantification and Evaluation Tool, which may be used to quantify both single guide and multiplex CRISPR-based screens and SGE experiments. pyCROQUET is comprised of three context-specific modes: ‘single-guide’, ‘dual-guide’ and ‘long-read’ all of which produce counts per guide, quality control statistics and read-to-guide alignments.

Whilst pyCROQUET ‘single-guide’ mode is similar in functionality to existing tools, the ‘dual-guide’ mode fulfils the more pressing need for a tool to quantify combinatorial or multiplexed CRISPR screens. Furthermore, the long-read mode is tailored towards the quantification of SGE screens and, in addition to the per-guide counts, statistics and read-to-guide alignments generated in other modes, produces unique read counts that are utilised by SGE analysis pipelines, such as QUANTS (https://github.com/cancerit/QUANTS).

pyCROQUET is available as a Python 3 module and Docker container, fulfilling the need for a portable application for the robust quantification of CRISPR-based functional genomics experiments.

pyCROQUET is available under the GNU General Public License as: ● Python3 module (https://pypi.org/project/pycroquet/) ● Repository on GitHub (https://github.com/cancerit/pycroquet/) ● Docker on quay.io (https://quay.io/repository/wtsicgp/pycroquet)

Keywords: CRISPR software SGE quantification

QUANTS: an open source, portable pipeline for the quantification of saturation genome editing experiments

Presenter: Victoria Offord (Wellcome Sanger Institute)

Victoria Offord (1), Hong Kee Tan (1), Andrew J. Waters (1), Mark Thomas (1), Matthew E. Hurles (1), David J. Adams (1)

Wellcome Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK

Saturation genome editing (SGE) is a powerful CRISPR/Cas9-based method used to systematically assess the functional consequences of genetic variants at scale. Whilst tools exist for the analysis of SGE screens, these packages serve specific functions such as adapter trimming or the merging of reads. With the evolution of the technology, SGE screens are increasing in scale and there is now a necessity for end-to-end tools which enable robust screen analysis.

QUANTS (https://github.com/cancerit/QUANTS) is an open source pipeline which incorporates well-established tools into a single, accessible workflow. QUANTS is portable and infrastructure independent, deploying its software components through

publicly accessible Singularity and Docker containers. These containers are embedded in reusable modules within the Nextflow nf-core framework. This enables QUANTS to be scalable and flexible, with the potential to add new tools as the field evolves.

QUANTS facilitates the analysis of both paired and single-end high throughput sequencing data, accepting FASTQ and CRAM file formats. Common analysis steps from previously published screens have been integrated as reusable modules. The majority encompass the pre-processing of sequencing data such as: read trimming (cutadapt), read merging (SeqPrep or FLASH2), read transformation and filtering by read length (SeqKit). Quality control metrics and plots for each of these stages is produced using FastQC and compiled into a single, user-friendly MultiQC report. Furthermore, QUANTS incorporates pyCROQUET (https://github.com/cancerit/pycroquet), a novel tool for the quantification of unique read sequences and oligo libraries.

We present QUANTS, a portable Nextflow pipeline which fulfils the current need for the robust analysis and quantification of SGE screens. QUANTS is available under the GNU General Public License as a repository on GitHub (https://github.com/cancerit/QUANTS).

Keywords: QUANTS SGE pipeline Nextflow quantification

Deep mutational scanning in whole organisms

Presenter: Curran Oi (University of Washington)

Curran Oi (1), Elizabeth Morton (1), Philip Abitua (1), Douglas Fowler (1,2), Celeste Berg (1), Christine Queitsch (1), Stanley Fields (1,3)

(1) Department of Genome Sciences, University of Washington, (2) Department of Bioengineering, University of Washington, (3) Department of Medicine, University of Washington

Abstract: Although a single mutation in a gene can lead to cancer, many other mutations in that gene are benign. How do we determine which mutations are pathogenic and which are benign? Only ~2% of the approximately 5 million protein-coding variants in the human genome have clinical annotations, half of which are “variants of uncertain significance.” High throughput functional screening of variants is thus needed for the millions of variants that have not been annotated. Deep mutational scanning has proven useful in classifying variants of BRCA1, PTEN, and other genes through screens carried out in cell culture. While this approach has the benefit of scale, it does not enable mutations to be studied in a developmental or tissue-specific context. However, human diseases, such as cancer, are tissue-specific, making it desirable to study mutations in the context of whole organisms. We are seeking to advance deep mutational scanning technology by establishing it in whole organisms. We use Caenorhabditis elegans, Drosophila melanogaster, and Ciona robusta as model organisms for this work. Our objective is to establish the technique robustly in at least one of these three organisms. Using a combination of genetic tools, we will introduce a library of variants of a test gene into each of the three organisms. In C. elegans, we will inject a purified processive base editing protein into the worms; in D. melanogaster, we will inject a pool of plasmids representing a variant library into embryos; and in C. robusta, we will electroporate a library of plasmids expressing variants into embryos. The resulting variant organisms will be collected and the barcodes they contain sequenced to determine the efficiency of library integration. After this step, we plan deep mutational scanning of three disease-relevant genes (let-60 in C. elegans, myc in D. melanogaster, and ß-catenin in C. robusta) and benchmark the results against databases of annotations for clinically-characterized gene variants. This work should develop tools for performing deep mutational scanning of developmental and disease-related genes in whole organisms.

Keywords:

Using Mutational Scanning in MUS81 to Identify Dominant Synthetic Lethal Interactions

Presenter: Anthony Oppedisano (University of British Columbia)

Anthony Oppedisano (1)

University of British Columbia (1)

The concept of synthetic lethality (SL) holds great promise for the selective killing of cancer cells. However, after more than 20 years of searching for SL-based anti-cancer therapies, only one example has reached the clinic: PARP inhibitors. Successful PARPi’s are unique in that they cause PARP to “trap” on DNA by disrupting key mechanistic steps required to dissociate from DNA. Trapped PARP-DNA complexes are toxic and require the concerted action of repair factors to facilitate resolution. However, when the machinery required to repair DNA-protein lesions are absent, such as in certain cancer cells, the lesion is irreparable resulting in trapping-mediated SL. Trapping inhibitors have the potential to be effective even in the presence of uninhibited target, in effect phenocopying a dominant mutation. Conversely, dominant mutations could be used to find and model trapping inhibitors. It is possible that other DNA repair proteins can be trapped by small molecules resulting in SL

expanding the range of cancers that can be targeted by trapping inhibitors. Key properties of a good trapping candidate are: 1. Dynamic association with DNA, 2. Enriched activity in tumour cells and 3. Genetic interactions with cancer-associated mutations. Mus81 is a DNA repair enzyme associated with increased survival and drug resistance of tumours. As an endonuclease involved primarily in replication restart and resolving recombination intermediates, transient interactions with DNA are imperative for Mus81 function. Importantly, cells that lack MUS81 or carry catalytically inactivating mutations often require other DNA repair enzymes for viability and tolerance to replication stress, highlighting its value as a target for trapping in anti-cancer therapies. I generated libraries of mutant Saccharomyces cerevisiae MUS81 to screen for mutations that, when expressed in wild-type cells, cause dominant sensitivity to methylmethanesulfonate (MMS), a DNA alkylating agent that induces replication stress similar to the molecular environment of a tumour. Screening for mutants in a WT MUS81-buffered system allows for destabilizing and nonsense alleles to be filtered out, allowing only mutations with a strong dominant growth defect to be observed. Dominant mutations clustered at the interface of the catalytic domain. These mutations tend to be conserved between species, further suggesting they are critical for function. Additionally, expressing these dominant mutants in cells lacking endogenous MUS81 cause a greater sensitivity to DNA damage than loss of MUS81. Expressing these mutants in SGS1-deficient cells carrying endogenous MUS81 causes a strong synergistic growth defect that becomes lethal upon mild MMS treatment. Similar observations are found in other cell lines that increase substrate abundance for Mus81, such as srs2Δ, suggesting mutant-induced toxicity relies on DNA interaction. The screen for dominant mutations in MUS81 identified residues on the 3D structure of MUS81 that cluster around the catalytic interface and appear imperative for function. To determine whether these mutations enhance Mus81 retention on DNA, thereby trapping, biochemical assays will be used to assess DNA-protein stability and catalytic activity. Future results from this screen may implicate other non-catalytic domain residues that may promote DNA trapping. Using this approach, we can map critical regions on the protein that may guide the design of small molecule inhibitors for trapping.

Keywords: Mutational Scanning, DNA repair, Cancer therapeutics, DNA trapping, DNA repair enymes, Dominant Synthetic Lethality

Predicting the effect of gene expression and environmental variation on E. coli growth rate

Presenter: Ryan Otto (The University of Texas Southwestern Medical Center, Dallas, TX 75390, USA)

Ryan M. Otto (1), Agata Turska-Nowak, Philip M. Brown (1), Kimberly A. Reynolds (1, 2)

1 Lyda Hill Department of Bioinformatics, The University of Texas Southwestern Medical Center, Dallas, TX 75390, USA 2 Department of Biophysics, The University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

Mapping variation in gene expression and environment to microbial growth rate is a critical advancement in bioengineering and medicine. However, our ability to predict the relationship between expression variation and phenotypes like growth is limited by three factors. First, the relationship between gene expression and growth rate is often strongly non-linear. Second, expression-fitness relationships for a single gene can change based on genetic and environmental context. Finally, many studies of gene-gene and gene-environment interactions focus on discrete perturbations with strong effects – like total knockouts, strong overexpression, or complete nutrient depletion. By neglecting intermediate phenotypes, we miss information that is relevant to organismal variation, evolution, and human disease. To address these challenges, we developed a methodology to make gradated changes in gene expression and quantify their effects on growth rate in varied environments. We found that a mechanism-independent model trained on a sparse sampling of these data that includes only pairwise interactions can predict the combinatorial effects of expression and environmental perturbations. Specifically, we quantified the relationship between gene expression and growth rate by collecting over 25,000 expression-fitness measurements following repression of 9 single genes and 36 gene pairs. We achieved this throughput using titrated CRISPR interference. By iteratively introducing mismatches into the homology region of an sgRNA, we gradually release repression of the targeted gene. This generates a library of sgRNAs that produce unique, intermediate knockdowns that can be efficiently multiplexed for combinatorial studies. We quantified each sgRNA's effect on gene expression with qPCR and measured growth rates using an NGS-based assay. To then relate gene expression to growth rate, we extended a mathematical model originally designed to predict the effect of combinatorial drug therapies. We show that this model can be modified to capture nonlinearity between expression and growth rate as well as context-dependence due to gene-gene and gene-environment interactions. This model drastically outperforms non-epistatic models in predicting growth rate; the model's accuracy approaches the limit of experimental variability. Critically, while this model was trained on growth rate data from a combinatorial landscape of pairwise knockdowns, its parameters are robust to subsampling as little as 20% of our data. We then used this framework to investigate interactions between gene expression and environmental context within a metabolic pathway. To test our ability to predict higher-order combinations of genetic and environmental perturbations, we measured microbial growth following 110 pairwise expression perturbations in 22 distinct environments, creating a 4th order growth rate landscape. Using only pairwise gene-gene, gene-environment, and environment-environment interactions, our model successfully predicts the effects of higher-order perturbations, drastically decreasing the amount of input data required to generate such a landscape. This high-throughput, unit-agnostic framework can be applied in future work to map the fitness effects of altering gene expression across entire metabolic pathways, or even genomes, in varied environments using sparsely sampled, low-order measurements.

Keywords: Gene expression, CRISPRi, Epistasis

Simplicity of experimental protein genotype-phenotype maps

Presenter: Yeonwoo Park (University of Chicago)

Yeonwoo Park (1), Brian P. H. Metzger (2), Joseph W. Thornton (1,2,3)

(1) Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL, USA
(2) Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA
(3) Department of Human Genetics, University of Chicago, Chicago, IL, USA

Learning the mapping from genotype to phenotype is a central goal in biology. Existing studies disagree on the complexity of protein genotype-phenotype (GP) maps, some studies finding widespread, high-order epistasis but others reporting epistasis limited in extent and order. These studies examine different GP maps using different methods, often using formalisms that are inadequate for learning the global structure of a GP map. Here we developed a formalism that optimally captures the global structure of a GP map and used it to analyze 15 experimental GP maps encompassing 4 to 16 sites in an enzyme, fluorescent protein, interface, or transcription factor. We found surprising simplicity in every dataset: Modeling only the independent effects and pairwise interactions of mutations is sufficient to predict phenotype with high accuracy (r2 > 0.92 in cross-validation). The fraction of possible effects and interactions that needs to be measured to achieve 90% prediction accuracy decreases with the size of GP map, reaching as low as 1 in 10,000. Consistent with this sparsity, a sparse learning method can be used to estimate the important effects and interactions from a small sample of genotypes and predict the phenotype of unobserved genotypes. Overall, our analyses reveal the extraordinary simplicity of available protein GP maps and suggest that sparse experimental characterization and statistical learning may allow elucidating the GP map of an entire protein.

Keywords: Genotype-phenotype map, protein, epistasis, sparse learning

Characterizing the genetic determinants of specificity and evolvability in an ancestral transcription factor-DNA complex

Presenter: Jaeda Patton (University of Chicago)

Santiago Herrera Álvarez (1,4), Jaeda E. J. Patton (2,4), Joseph W. Thornton (1,3)

(1) Department of Ecology and Evolution, University of Chicago
(2) Committee on Genetics, Genomics and Systems Biology, University of Chicago
(3) Department of Human Genetics, University of Chicago
(4) Equal contribution

Deep mutational scanning enables empirical characterization of genotype-phenotype maps at large scale, providing insight into the genetic determinants of phenotype and how this shapes the mutational paths that are available to evolving molecules. Studies that analyze combinatorial libraries of mutations have so far only focused only on mutations within a single protein or DNA sequence, but most biological molecules function in complexes, and intermolecular epistasis is known to be important for interaction specificity and evolution. To address this gap, we are performing combinatorial intermolecular deep mutational scanning on ancestral steroid receptor DNA binding domains (DBDs) and their DNA response elements (REs). We have constructed libraries of DBD variants that contain all possible combinations of amino acids at four sites in the protein's DNA recognition helix (RH). Using a yeast-based sort-seq assay, we are testing these libraries for the ability to bind all 4^2=16 possible RE sequences that vary at two evolutionary variable sites. This approach will allow us to decompose RH and RE mutation effects into their main and epistatic contributions to binding affinity and to understand the genetic mechanisms by which binding specificity is encoded at this interface. Additionally, by testing RH libraries in the context of two ancestral DBD sequences that span a phylogenetic interval in which binding specificity evolved from one to another of the 16 REs, we will be able to compare the mutational accessibility of all possible derived RE binding functions before and after the historical change in function and ask to what extent mutational accessibility of novel functions in sequence space is concordant with the historical evolutionary outcome. Finally, this dataset will allow us to connect the genetic contributions of mutations in the RH and RE to their effects on the genotype-phenotype map and thereby the accessibility of different DNA specificity phenotypes during evolution of the complex.

Keywords: evolution; ancestral sequence reconstruction; epistasis; transcription factor

Identification of Pathogenic mutations and exploration of molecular mechanisms of human disease with Deep Mutational Scanning in yeast

Presenter: Marcin Plech (University of Edinburgh)

Marcin Plech, Grzegorz Kudla, Joseph Marsh

University of Edinburgh

Quick and precise diagnosis and detailed understanding of molecular mechanisms of rare genetic disorders will improve quality of life of patients and their families. Today with the advent of Deep Mutational Scanning (DMS), which combines Next Generation Sequencing technologies with high-throughput screening assays, we are overcoming many limitations that stifled the progress of genetic disease researcher. Our team focuses on investigating damaging protein variants of human genes in yeast competition experiments. Most studies to date have only aimed at identification of loss of function (LOF) mutations as this was the perceived source of genetic disorders. However, pathogenic mutations can exert a wide range of effects on proteins, including gain of completely new functions (GOF) and dominant negative effects (DN) which compromise whole protein complexes by incorporation of singular defective peptides. We have performed DMS of the human GTP cyclohydrolase (hGCH1) gene implicated in the DOPA-responsive dystonia, a neurological movement disorder. An assay based on complementation of the yeast GCH1 homologue, FOL2, by wild-type and mutant GCH1, resulted in extremely good detectability of human pathogenic mutations as our empirical measurements have outperformed every single currently available variant effect predictor (including supervised and unsupervised VEPs). Most deleterious mutations were found in substrate binding residues, ligand pockets, and in the interfaces between the two pentamers that form the native GCH1 homodecamer. These results demonstrate the functional importance of the decameric configuration of the protein. In our experiments we have utilised a set of vectors allowing variable expression levels of GCH1 variants to assess the mechanistic distinction between LOF, DN, or GOF status. Our results revealed statistical signatures of DN mutations but at this point are not conclusive and require further tests. We also ran our experiments with and without co-expression of the WT allele of hGCH1 to annotate the mode of inheritance of individual mutations (dominant or recessive). Studying the same library under different expression and stoichiometric regimes gives us the power to classify deleterious mutations more accurately. Our ultimate goal is to build inference models that combine the mode of inheritance with other factors to predict the pathogenicity of mutants and accurately assess the risk of individual missense variants.

Keywords: genetic disorder, GCH1, mode of inheritance, complex assembly, yeast competition screen, variant prioritisation

A novel method to profile missense variation at scale for extracellular proteins reveals biochemical features important for expression and secretion of Factor IX

Presenter: Nicholas Popp (University of Washington, Department of Genome Sciences)

Nicholas Popp (1,2), Rachel Powell (1), Alexis Chang (1), Kerry Lannert (3), Jill Johnsen (3,4,5), and Douglas Fowler (1,6,7)

(1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195
(2) Medical Scientist Training Program, University of Washington School of Medicine, Seattle, WA 98195
(3) Bloodworks Research Institute, Seattle, WA 98102
(4) Division of Hematology, Department of Medicine, University of Washington School of Medicine, Seattle, WA 98195
(5) Washington Center for Bleeding Disorders, Seattle, WA 98101
(6) Department of Bioengineering, University of Washington School of Medicine, Seattle, WA 98195
(7) Brotman Baty Institute for Precision Medicine, Seattle, WA 98195

Multiplexed Assays of Variant Effect (MAVEs) are a powerful tool for assessing the effects of variation in proteins to understand sequence-function relationships, inform biological structures, and infer pathogenicity in clinical settings. Currently, the genes that can be studied with MAVEs are limited to those which encode for intracellular or membrane-bound proteins, because functional selection on the protein and subsequent genomic DNA sequencing readout for identifying variants must be spatially connected within an individual cell. We hypothesized that cell-surface display, as has been used in yeast and phage, could be adapted for use in MAVEs in mammalian cells. By fusing an extracellular protein to a single pass transmembrane domain, protein variants can be displayed on the surface of mammalian cells and assayed for function, thus re-establishing the spatial connection between functional selection and genomic DNA sequencing readout necessary for MAVEs.

As proof of concept, we first assayed Factor IX, a secreted coagulation protein for which mutations lead to hemophilia B, for surface staining with anti-factor IX antibodies. We then applied this technique to a library of nearly all possible missense variants in Factor IX to profile each variant's effect on Factor IX expression and secretion. We find that 37.7% of the profiled missense variants can be classified as poorly secreted, and that our secretion assay alone explains approximately half of the variance of assayed Factor IX antigen levels in patients with hemophilia B. We further find that cysteine substitutions have strong negative consequences on Factor IX expression and secretion, highlighting the importance of disulfide bonds and oxidative conditions in the extracellular space. These findings contrast with data from cytoplasmic proteins where cysteine is generally well-tolerated, and instead, proline is the most detrimental substitution for protein abundance and stability. Lastly, we expand our cell surface display system to other antibodies and proteins, showing the generalizability of our method for assaying variation at scale for secreted proteins.

Keywords: secretion, abundance, Factor IX, coagulation, hemophilia, high-throughput functional genomics

Variations in genes associated with recurrent miscarriages in the HGDP-CEPH Dataset and euploid recurrent miscarriage samples

Presenter: Madhuri Pulijala (Monash University Malaysia)

Madhuri Pulijala (1), Silvia Buonaiuto (2), Keang Peng Song (1), Vincenza Colonna (3), Qasim Ayub (1, 4)

1 School of Science, Monash University Malaysia, 47500 Bandar Sunway, Malaysia. 2 University of Campania Luigi Vanvitelli, 81100, Caserta, Italy. 3 Institute of Genetics and Biophysics, National Research Council, Naples, 80111, Italy. 4 Monash University Malaysia Genomics Facility, 47500 Bandar Sunway, Malaysia.

Spontaneous pregnancy loss before the third trimester occurs in ~15 % of all pregnancies and it may recur in up to 4% of women. Recurrent miscarriages are usually defined as the loss of two or more pregnancies within 20 - 24 weeks of gestation. A large proportion of recurrent miscarriages that are classified as idiopathic may in reality occur due to single nucleotide variants. We identified 972 genes associated with recurrent miscarriages or stillbirths, and used the publicly available whole genome sequenced Human Genome Diversity Project Centre d'Etude du Polymorphisme Humain (HGDP-CEPH) dataset comprising of 54 worldwide populations to characterise 2,071,033 variants in these genes. Ensembl's Variant Effect Predictor was used to assess the impact of these variants. As expected for an apparently healthy adult population dataset, no high impact variants were found but 105,212 moderate impact missense mutations were identified of which 25 were annotated as probably damaging by PolyPhen and 37% as deleterious by SIFT. No homozygous deleterious or loss of function variations were observed in these genes, which is in line with our hypothesis that such mutations in these candidate genes would be incompatible with live births. Variants in these genes are being examined in 3 euploid whole genome sequenced miscarriage samples from Malaysia and these will be compared with a European dataset that comprised of 6 euploid miscarriage samples. The latter showed a mean of 6.5 homozygous deleterious mutations per sample. These results will aid in understanding the genetic architecture and mutational spectrum in recurrent miscarriages and highlight genes that are essential for early human development.

Keywords: Recurrent miscarriages, spontaneous miscarriage, HGDP-CEPH data, missense mutations

Mapping Sequence-Function Landscapes in the Dihydrofolate Reductase Family

Presenter: Carmen Resnick (University of Oregon, Knight Campus)

Samuel Hinton (1), Calin Plesa (1)

(1) Phil and Penny Knight Campus for Accelerating Scientific Impact

Dihydrofolate reductase (DHFR) is an essential enzyme in the folic acid synthesis pathway and has been the subject of intense study in the past few decades. Despite the wide diversity of homologs, research attention has primarily focused on DHFR proteins from particular organisms, as well as their mutants. In this study we explore DHFR expression through a knockout E. coli strain ER2566 ΔfolA ΔthyA. We focus on the ability of DHFR to both rescue metabolic function and tolerate treatment against the antibiotic trimethoprim, which will allow us to understand how antibiotic resistance emerges given many different evolutionarily divergent starting points. Changes in the mutational landscape of DHFR allows for varying survival rates when in the presence of antibiotic inhibitors. We carry out a broad mutational scan using a library of nearly 5,000 DHFR homologs synthesized using DropSynth gene synthesis. Variant fitness is determined in a multiplex survival assay in the knockout strain which allows conditional selection dependent on external supplementation.

We aim to collect quantitative fitness data on which mutations impact DHFR's activity, both in the presence and absence of inhibitors, in order to elucidate sequence-function relationships and understand how correlations between the fitness landscapes vary as a function of evolutionary distance between homologs. This data can be applied towards the development of narrow-spectrum and targeted antibiotics and mitigation of resistance through understanding the pathways from which antibiotic resistance arises.

Keywords: Antibiotic Resistance, E. coli, High-throughput scan, DHFR, Dihydrofolate Reductase, Multiplex Gene Synthesis, Trimethoprim

Insights from millions of folding stability measurements

Presenter: Gabriel Rocklin (Northwestern University)

Kotaro Tsuboyama (1), Gabriel Rocklin (1)

(1) Northwestern University Department of Pharmacology & Center for Synthetic Biology

Protein sequence variants often have altered folding stability, and these folding stability changes are a major cause of biological phenotypes. Improving our understanding of folding stability would provide a mechanistic basis for interpreting variant effects and accelerate the design of new protein structures. However, folding stability has historically been challenging to measure, and the limited availability of stability data has also hindered the development of predictive models. We recently developed a new high-throughput assay for folding stability based on cDNA display, protease sensitivity, and next-generation sequencing. Employing libraries of ~800,000 sequences, our method is fast, precise, accurately reproduces known stabilities, and displays a >10,000-fold (>5 kcal/mol) dynamic range. Using this approach, we measured folding stability for all single muntants (including deletions and G/A insertions) for ~300 natural protein domains and ~200 de novo designed domains 40-72aa in length. These data reveal how individual amino acids contribute to folding stability in different contexts and how evolutionary conservation diverges from the biophysical requirement for stability. Our full dataset of over 2 million stability measurements on small domains should prove valuable for training the next generation of computational models to predict folding stability.

Keywords: folding stability, protein design, biophysics

Developing high-throughput infection assays for characterizing ACE2 receptor dependencies of diverse SARS-like coronavirus receptor binding domains

Presenter: Sarah Roelle (Case Western Reserve University School of Medicine)

Sarah M. Roelle, Nidhi Shukla, Anh T. Pham, Anna M. Bruchez, Kenneth A. Matreyek

Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA

Viral spillover from animal reservoirs can trigger public health crises and cripple the world economy. Knowing which viruses a re primed for zoonotic transmission can focus surveillance efforts and mitigation strategies for future pandemics. Host factor c ompatibility, such as the successful engagement of receptor protein orthologs during viral entry, is necessary for cross-specie s transmission. The large number of sequence differences that oftentimes occur between related viruses require each viral en try protein to be tested separately. Owing to long-term evolutionary conflicts, host receptor proteins often exhibit high sequenc e variation from positive selection. Thus, we need new methods to test pairwise combinations of viral entry protein and host re ceptor orthologs to determine the matrix of virus-host compatibility that underlies potential viral cross-species transmission. To pilot our approach, we focused on the SARS-like coronaviruses due to their pandemic potential. The clade 1 sarbecovirus es, including SARS-CoV and SARS-CoV-2, enter cells via engagement of ACE2. The receptors for the distantly related clade 2 viruses sequenced in Asia, and a handful of clade 3 viruses found in Europe and Africa, remain largely uncharacterized. Usi ng a Bxb1 recombinase-based transgenic system, we stably expressed either human ACE2 or ACE2 from various Rhinoloph us horseshoe bat species in HEK 293T cells. To determine their compatibility with different viral entry proteins, we mixed thes e cells with GFP reporter pseudoviruses coated with unmodified SARS-CoV and SARS-CoV-2 spike proteins, or chimeric spi ke proteins with the receptor binding domains (RBD) of the various clades 2 and 3 sarbecoviruses were swapped in. To incre ase throughput, we created a duplex infection assay, wherein cells within the same well were either ACE2-expressing or ACE 2-null and were differentially fluorescently-tagged, so that we could use flow cytometry as a readout of infection of the two po pulations. Having the experimental and control cells together in the same well allowed for each sample to have its own intern al control. Our experimental approach revealed clear differences separating the ACE2-dependent and ACE2-independent vir uses. The clade 2 RBDs exhibited very weak infection, indicating that the clade 2 sarbecoviruses likely infect via a different re ceptor. All clade 3 viruses used at least one ACE2 ortholog we tested. Similar to SARS-CoV and SARS-CoV-2, RBDs from th e clade 3 viruses, Khosta-1 and BtKY72, infected cells of nearly all tested bat ACE2 orthologs. Interestingly, human ACE2 all owed for infection by Khosta-2 and BtKY72 RBDs, in addition to the already known clade 1 viruses. We found sarbecovirus R BDs encoding a lysine at the position homologous to SARS-CoV-2 Q493 were generally incompatible with human ACE2. Our results suggest that clade 3 sarbecoviruses from Africa and Europe use Rhinolophus ACE2 for entry, and some of their spike proteins appear especially primed to contribute to zoonosis under the right conditions.
To further improve throughput for future studies, we are developing a multiplex infection assay capable of simultaneously testi ng dozens to hundreds of orthologs and protein variants in a single infection experiment using a sequencing-based readout . Thus, our study provides a new genetic approach for characterizing receptor utilization during viral entry, which will be used to create an atlas of virus-host compatibility in the future.

Keywords:

Functional variomics: Systematic annotation of somatic mutations and gene fusions in cancer

Presenter: Nidhi Sahni (MD Anderson Cancer Center)

Daniel J. McGrail (1), Patrick Kwok-Shing Ng (1), Gordon B. Mills (2,3), S. Stephen Yi (4,5), and Nidhi Sahni (6,7,8*)

1Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA 2Department of Cell, Developmental and Cancer Biology, School of Medicine, Oregon Health & Science University, Portland, OR 97201, USA 3Precision Oncology, Knight Cancer Institute, Portland, OR 97201, USA 4Department of Oncology and

Department of Biomedical Engineering, Livestrong Cancer Institutes, Dell Medical School, The University of Texas at Austin, Austin, TX 78712, USA The University of Texas at Austin, Austin, TX 78712, USA 5Oden Institute for Computational Engineering and Sciences (ICES), and Interdisciplinary Life Sciences Graduate Programs (ILSGP), College of Natural Sciences, The University of Texas at Austin, Austin, TX 78712, USA 6Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Smithville, TX 78957, USA 7Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA 8Quantitative and Computational Biosciences Program, Baylor College of Medicine, Houston, TX 77030, USA

*Correspondence: nsahni@mdanderson.org (N.S.)

Proteins interact with other macromolecules in complex cellular networks for signal transduction and biological function. Our previous work in Mendelian disorders found a widespread phenomenon that disease-associated alleles often perturb distinct protein activities rather than grossly affecting folding and stability. In the context of cancer, the functional impact of the vast majority of somatic mutations remains unknown, representing a critical knowledge gap for implementing precision oncology. Here, we present the development of a high-throughput functional variomics platform consisting of efficient mutant generation, sensitive cell viability and drug response assays, and functional proteomic profiling of signaling effects for select aberrations. We apply the platform to annotate thousands of genomic aberrations, including point mutations, indels, and gene fusions, potentially doubling the number of driver mutations characterized in clinically actionable genes. Further, the platform is sufficiently sensitive to identify weak drivers. Our data are accessible through a user-friendly, public data portal. Our study will facilitate biomarker discovery, prediction algorithm improvement, and drug development.

Keywords: Functional variomics, somatic mutations, gene fusions, cancer, network biology, precision oncology

Comprehensive mutational analysis to investigate functional dependencies in viral proteins

Presenter: Neha S Samant (University of Massachusetts Chan Medical School)

Neha Samant(1), Gily Nachum(1), Tenzin Tsepal(1), and Daniel N.A. Bolon(1)

1 University of Massachusetts Medical School, Worcester, MA.

Viruses make up over two-thirds of all new human pathogens, out of which, HIV-1; the causative agent of Acquired Immunodeficiency Syndrome (AIDS) still remains the leading cause of death from infectious diseases, claiming about 37 million deaths since the beginning of the HIV pandemic. HIV protease (PR) plays a critical role in viral maturation by processing viral polyproteins at specific cut sites, making protease inhibitors (PIs) an important component of Highly Active Anti-Retroviral Therapy (HAART). These cut sites are highly diverse in sequence, the amino acids adjacent to the scissile bond are referred to as position(P) 4-3-2-1/1'-2'-3'-4 (scissile bond indicated by /). P4-P4' directly contact the PR active site, a few recent studies indicate that substrate positions beyond P4 and P4' can also contribute to substrate binding. PIs potently inhibit replication of the wildtype HIV-1, and the virus responds through the accumulation of mutations in both PR and some cut-sites. PR accumulates many drug resistant mutations and studies have shown that cut sites in Gag co-evolve with drug resistant mutations in PR, indicating that the cut site mutations may compensate for reduced enzyme activity of PR. While there is extensive evidence that cut-sites are critical for viral replication and subject to selection pressure by PI's, the wide variation in cut-site sequences, along with poorly understood impact of positions beyond P4-P4' has made it challenging to elucidate the features that underlie cut-site recognition and processing by PR. To further probe the role of positions distal to the scissile bond and investigate the determinants of cut site recognition, we generated all amino acid mutations at 12 positions in three selected sites, MA/CA, Nc/p1 and p1p6 (A.A region P6-P6') and analyzed the mutational effects on proteolytic cleavage by wild-type PR using a yeast display based assay. The resulting cut-site fitness landscapes reveal both general biophysical features that underlie cutting efficiency in all the three cut sites as well as features that depend on the rest of the sequence background of the cleavage site. We observe strong correlation of our yeast-display results for cut-site variants with WT-like or lower function, while the variants that cut faster than WT by yeast-display had a wide array of function when measured using a natural substrate, which may imply that the difference in cutting is due to features in the yeast-display assay that differ from natural substrates. We propose a model that, the artificial disruption of contacts in the yeast display setup provides an opportunity for amino acid changes at positions towards the end and past the active site to recoup binding energy with protease. However, in the natural MA/CA substrate, the effect of these amino acid changes has varied impacts because it disrupts the natural interactions of MA/CA with protease. This work provides a new and generalizable approach to investigating the relationship between viral proteases and their substrates. We show that long-range and short-range sequence interdependencies are prevalent for HIV-1 cut-site efficiency, but that consistent and complex biophysical properties have large impacts at sites adjacent to the scissile bond. This study provides framework to examine evolutionary constraints and biochemical mechanisms of viral proteins that can contribute to the evolution of drug resistance.

Keywords:

Pharmacological Profiling of Cystic Fibrosis Variants by Deep Mutational Scanning

Presenter: Jonathan Schlebach (Indiana University, Bloomington)

Andrew G. McKee, Francis J. Roushar, Wesley D. Penn, Charles P. Kuntz, Laura M. Chamness, and Jonathan P. Schlebach

Department of Chemistry, Indiana University, Bloomington, USA

Over 400 mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene are known to cause cystic fibrosis (CF). Though all of these mutations cause a loss of CFTR function, they vary considerably with respect to the nature of their mechanistic effects on the structure and function of the CFTR chloride channel. Moreover, these variants exhibit considerable differences in their response to current therapeutics. Here we utilize deep mutational scanning to compare the plasma membrane expression of 129 CFTR variants in HEK293T cells. We identify 82 variants, including 10 variants of unknown significance (VUS), that exhibit diminished plasma membrane expression relative to WT. These findings confirm that most CF variants enhance CFTR misfolding and degradation. Of these apparent class II variants, we identify 77 that measurably respond to VX-661 and/ or VX-445- two different FDA-approved "corrector" molecules designed to rescue the expression of misfolded CFTR variants. Nevertheless, the magnitude of their effects varies widely across the spectrum of clinical variants. Our preliminary analyses suggest the most responsive mutants are structurally coupled to the corrector binding pocket. Together, our findings identify previously uncharacterized CF variants that are potentially amenable to therapeutic intervention and provide insights into the molecular basis for the mutation-specific efficacy of CFTR modulators.

Keywords: cystic fibrosis, CFTR, corrector, proteostasis, misfolding

Multi-Phenotye Deep Mutational Scanning in a K+ Channel

Presenter: Daniel Schmidt (University of Minnesota)

Willow Coyote-Maestas (1), David Nedrud (1), Yungui He (2), Daniel Schmidt (2)

(1) Department of Biochemistry, Molecular Biology & Biophysics , University of Minnesota, Minneapolis, MN, 55455, USA
(2) Department of Genetics, Cell Biology & Development, University of Minnesota, Minneapolis, MN, 55455, USA

A longstanding goal in protein science and clinical genetics is to develop quantitative models of sequence, structure, and function relationships and delineate the mechanisms by which mutations cause disease.

Here, we use multi-phenotype Deep Mutational Scanning (DMS) in the Inward Rectifier K+ channel Kir2.1 to map determinants of folding, assembly, and trafficking, as well as regulation by allosteric ligands and ion conduction.

Our data provide high-resolution information on a cotranslationally-folded biogenic unit, trafficking and quality control signals, and segregated roles of different structural elements in fold-stability and function. We show that Kir2.1 trafficking mutants are underrepresented in variant effect databases, which has implications for clinical practice. By comparing fitness scores with expert-reviewed variant effects, we can predict the pathogenicity of 'variants of unknown significance' and disease mechanisms of know pathogenic mutations.

Our study in Kir2.1 provides a blueprint for how multi-phenotype DMS can help us understand the mechanistic basis of genetic disorders and the structure-function relationships of proteins.

Keywords: membrane protein; ion channel; folding; trafficking; pathogenicity prediction

Deep Mutational Scanning of Retinal Transcription Factors

Presenter: James Shepherdson (Washington University in St. Louis)

James Shepherdson (1, 2, 3); David Granas (1, 2), Michael White (1, 2), Barak Cohen (1, 2)

(1) Department of Genetics,
(2) Center for Genome Sciences and Systems Biology,
(3) Medical Scientist Training Program,

Washington University School of Medicine in St. Louis

The transcription factors Cone-Rod Homeobox (CRX) and Neural Retina-specific Leucine zipper (NRL) are master regulators of photoreceptor cell fate. Sequence variants in CRX and NRL can cause Retinitis Pigmentosa, Cone-Rod Dystrophy, and Leber Congenital Amaurosis, all inherited causes of vision loss and blindness. In particular, CRX is the only gene implicated in

the pathogenesis of all three of these diseases, which present with both rod- and cone-centric phenotypes of varying age of onset and severity. Several CRX and NRL variants have been reported to cause severe dominant disease, and yet these mutations are adjacent to variants which are benign or only cause mild, recessive disease. Thus, determining which mutations in CRX and NRL are pathogenic and quantifying their effect on functional activity is prerequisite to interpreting patient variation and predicting patient phenotypes. Routine clinical sequencing is on the rise, but our ability to interpret patient variants in genes like CRX and NRL has not kept pace. In ClinVar, the NIH-sponsored database of human genetic variation, CRX and NRL alone have over 100 reported VUS, with nearly 95% of all possible variants not yet observed in a patient (and thus also uncharacterized). Without a robust catalog of human genetic variation, advances in patient sequencing cannot be translated into clinical guidance or therapies for patients with uncharacterized variants. We are using Deep Mutational Scanning to measure the effects of CRX and NRL variants on protein abundance, as measured by Variant Abundance by Massively Parallel sequencing (VAMP-seq), as well as the effects of CRX and NRL variants on transcriptional activity, as measured in a fluorescent transcriptional reporter assay. Our DMS libraries are introduced into an engineered HEK 293 cell line carrying a genomic landing pad for single-copy integration and expression. Ultimately, we aim to leverage the ease of access and genetic manipulations of mouse retina to validate our cell culture findings in the native transcriptional environment of CRX. We aim for this work to establish a catalog of functional variation in CRX, and an extensible system for DMS of retinal transcription factors.

Keywords:

Quantitative characterization of ACE2 cell surface abundance and susceptibility of cells to SARS-like coronavirus entry.

Presenter: Nidhi Shukla (Case Western Reserve University)

Nidhi Shukla, Sarah M. Roelle, Anna M. Bruchez, Kenneth A. Matreyek

Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA

SARS-like coronaviruses, including SARS-CoV and SARS-CoV-2, encode spike proteins that bind human ACE2 on the cell surface to enter target cells and cause infection. The efficiency of virus entry depends on ACE2 sequence and expression level in target cells. A small fraction of humans encodes variants of ACE2, thus altering the biochemical properties at the protein interaction interface. All humans possess cells with vastly differing amounts of ACE2 on the cell surface, ranging from cell types with high expression in the gut and lungs to lower expression in the liver and pancreas. Mastering our understanding of spike-ACE2 interaction and infection requires experiments precisely perturbing both variables. Thus, we developed a synthetic cell engineering approach compatible with high throughput assays for pseudotyped virus infection capable of assessing both variables individually and in combination. We adapted an engineered HEK 293T DNA recombinase landing pad cell line capable of expressing transgenic ACE2 sequences at highly precise levels. Infection with lentiviruses pseudo-typed with the spike proteins of SARS-like coronaviruses revealed that high ACE2 abundance could mask the effects of impaired binding thereby making it challenging to know the role of affinity altering mutations during infection. Reducing ACE2 translation rates with a suboptimal Kozak sequence, hence reducing its steady-state abundance, yielded sufficient dynamic range to study ACE2 sequence alterations, as two human ACE2 variants at the binding interface exhibited reduced infection. These experiments showed that we need finer detail in defining how ACE2 expression determines infectibility. We thus created an ACE2 Kozak library consisting of ~4,096 Kozak variants. Combining fluorescence-activated cell sorting and high-throughput DNA sequencing (FACS-seq) revealed that the library spanned two orders of magnitude of ACE2 abundance. Challenging this library of cells with spike pseudo-typed lentiviruses revealed how ACE2 abundance correlated with infection rate. The library-based experiments yielded a dynamic range wider than traditional single sample infection assay, suggesting that the multiplex approach is more representative of infection dynamics in vivo.
Now that we have our data revealing the relationship between ACE2 abundance and infectivity, our next goal is to expand the comparison to physiologically relevant cells. We believe modulating protein abundance levels will be key to creating maximally informative assays for protein function. These results lay the groundwork for being able to simultaneously test the impacts of protein abundance and sequence in combination in many future projects.

Keywords: SARS-like coronaviruses, ACE2, Kozak sequence, FACS-seq

Benchmarking splice variant predictive algorithms with saturation screens.

Presenter: Cathy Smith (University of Michigan)

Cathy Smith (1), Jacob O. Kitzman (1,2)

1 Department of Computation Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA 2 Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA

Splicing is a critical step in mRNA maturation with roles in gene regulation and proteome diversification, and its disruption is implicated across a variety of human disease genes. Although a substantial fraction of variants outside of canonical sites may

disrupt splicing, their identification remains challenging due to the degeneracy and redundancy of the underlying sequence code. The over-representation of canonical splice mutations sites among reported clinical variants also makes it difficult to establish a representative set of splice-disruptive variants to evaluate computational splice effect predictors. Massively parallel splicing assays (MPSAs) avoid this ascertainment bias by measuring the impact of all possible point mutations to nominate candidate splice disruptive variants. Using data from four of these high-throughput functional screens, we compared experimentally measured splicing outcomes from BRCA1, POU1F1, RON, and FAS with predictions from seven contemporary variant prediction algorithms. We also compared assessments of 312 MLH1 colorectal cancer patient variants collected from the literature with predictions from the same set of tools. As expected, the algorithms correctly predicted >84% of canonical sites as splice disruptive, however at positions outside of splice sites less than half (45.8%) of the splice altering variants identified by MPSA were predicted as such informatically. Pangolin and SpliceAI - two deep learning tools trained on annotated splice sites - corresponded best with the measured splicing outcomes, both at intronic and exonic positions (prAUC range: .658 (HAL) to .890 (Pangolin); prAUC mean = .779). We use the results of each measured screen and MLH1 patient variants to propose a clinical threshold for each tool that best identifies splice disruptive variants. As a measure of genome-wide specificity, we then scored randomly selected genic variants with each tool, and found that at optimal cutoffs, the tools call 4.9-43.0% (mean = 21.2%) of coding and proximal intronic SNVs as pathogenic.

Keywords: Splicing, Benchmarking, Computational Prediction, Massively Parallel Splicing Assay, Variant Effect Prediction

Large-scale testing of AGXT variant effects

Presenter: Adrine de Souza (University of Toronto / Lunenfeld-Tanenbaum Research Institute)

Adrine de Souza(1,2), Nishka Kishore(1), Roujia Li(2), Warren van Loggerenberg(1), Ashyad Rayhan(2), Chandra L. Tucker(3), Marinella Gebbia(1), Atina G. Cote(1,2) Frederick P. Roth(1,2)

1 Donnelly Centre, University of Toronto, Ontario, Canada 2 Lunenfeld-Tanenbaum Research Institute, Mt. Sinai Hospital, Toronto, Ontario, Canada 3 Department of Pharmacology, University of Colorado School of Medicine, Aurora, Colorado, USA

Sequence variation in the AGXT gene can disrupt the liver enzyme alanine:glyoxylate aminotransferase (AGT). Loss of AGT function can cause accumulation of oxalate crystals in the kidneys, causing kidney stones and the rare disease Primary Hyperoxaluria Type 1 (PH1). Over 25% of clinically-observed AGXT missense variants are deemed variants of uncertain significance (VUS). To support more informative variant annotation by systematically measuring variant effects of AGXT, our group is deploying a Multiplexed Assay of Variant Effect (MAVE) to test nearly all AGXT missense variants in a cell-based assay: Saccharomyces cerevisiae lacking key genes in the glycine metabolism (SHM1, SHM2, GLY1 and AGX1) cannot grow in ethanol media without either functional AGXT or supplemented glycine. Validating the assay, we found that benign AGXT alleles robustly supported growth in ethanol media, while pathogenic variants did not. In the background of the common minor allele Pro11Leu and Ile340Met (global minor allele frequency 10%), our assay detected a functional impact for 4 out of 5 pathogenic variants and 0 out of 4 benign variants, thus achieving 100% precision with 80% recall. We present progress towards variant effect maps of both major and minor alleles of AGXT, with the goal of discovering AGT molecular structure/function relationships and expanding the number of patients with a definitive genetic diagnosis who can benefit from already-approved PH1 therapies.

Keywords: MAVE, DMS, AGXT, variant effect map, PH, Primary Hyperoxaluria, alanine:glyoxylate aminotransferase, kidney stones, functional genomics.

Deep mutational scan of the VWF C domains to define mutations associated with VWD

Presenter: Taylor Sparring (Thrombosis and Atherosclerosis Research Institute (TaARI), McMaster University)

Nicholas Popp (2), Douglas Fowler (2), Colin Kretz (1)

(1) Thrombosis and Atherosclerosis Research Institute (TaARI), McMaster University (2) University of Washington

Rationale: Von Willebrand Factor (VWF) is a multimeric glycoprotein that recruits platelets to sites of blood vessel injury. VWF is a 2300 amino acid multimeric protein that is stored in the Weibel-Palade Bodies in endothelial cells. It undergoes both constitutive and agonist induces secretion. Mutations in the VWF C domains can lead to intracellular retention and lower circulating VWF levels associated with the bleeding disorder type 1 von Willebrand disease. It’s currently not well understood why some mutations lead to decreased VWF secretion, while others do not. Objective: To identify every possible missense mutation in the VWF C domains that either increase or decrease its biosynthesis and secretion from cells. Results: Using the previously developed Bxb1 recombinase-based landing pad system, we have developed a protein secretion assay in HEK293T cells expressing VWF-eGFP. Mutations in VWF that affect biosynthesis and secretion can be identified based on cell fluorescence intensity using (FACS). Initial studies confirmed that cells expressing VWF mutants with known secretion defects have increased intracellular eGFP intensity (eGFP MFI: V1822G=157.72 5.40, C2693Y=184.06 2.10), relative to cells expressing WT VWF (eGFP MFI: WT=69.22 1.87). We used GenScript’s oligo pool synthesis platform to generate 10 845

unique oligos containing every possible missense mutation of the VWF C domains. A major limitation for deep mutational scanning of VWF is the size and complexity of the protein. We therefore used Q5 assembly to generate a C-domain mutagenesis library, which was subsequently cloned into full length VWF cDNA using appropriate restriction enzymes. These methods provide an appropriate approach for generating a deep mutational library of a very large gene of interest. HEK293T cells expressing the VWF-eGFP mutagenesis library exhibited a broad distribution of MFI compared to cells expressing WT VWF-eGFP, confirming the presence of variants that affect secretion. In preliminary studies we focused on the N-terminal region of the VWF C domains (CK domain) consisting of a total of 2169 unique oligos. Cells were sorted into four equally distributed bins of eGFP intensity and prepared for NGS to identify mutations that are enriched in cells with impaired (upper 25% eGFP) or enhanced (lower 25% eGFP) secretion. VWF mutations can lead to ER retention, causing secondary effects on cell viability. Cells expressing variants with the highest 25% eGFP intensity had upregulation of GRP78, CHOP, and ERP57 expression compared to cells expressing WT VWF. Therefore, missense mutations that cause impaired VWF basal secretion are associated with increased expression of markers in the ER stress adaptive and apoptotic UPR pathways. Conclusions: These data show that FACS is a useful method to screen for variants that impair basal VWF secretion using an unbiased deep mutational scan of the VWF C domains. These data will provide a detailed structure/function analysis of VWF in mammalian cells and identify variants associated with type 1/3 VWD.

Keywords:

The Protein Variant Effect Omnibus (ProVEO) - integrating deep mutational scanning data with protein functional annotations

Presenter: James D. Stephenson (European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), UK)

James D. Stephenson (1), Alan F. Rubin (2), Maria J. Martin (1), Douglas M. Fowler (3)

1. European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), UK, 2. Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia, 3. University of Washington, Seattle, WA, USA

Deep mutational scanning (DMS) experiments have incredible potential for the interpretation of protein function and residue interchangeability. However, their full capacity cannot be realised without the integration of other sources of knowledge about the protein and the heterogeneity of this information complicates potential integration and limits the scope for using DMS data in protein-centric data resources.

We are building the Protein Variant Effect Omnibus (ProVEO), which will provide the necessary infrastructure to map the DMS data in MaveDB to UniProt in order to integrate the rich structural and functional annotations therein. Integration of DMS data into UniProt will increase discoverability, and standardising mappings and identifiers will allow users to find relevant data both within MaveDB and via linked external resources. We are also developing new web components in UniProt to visualise DMS data in a structural context and to filter on datasets, residues or mutations to facilitate exploratory data analysis. ProVEO will be automatically updated with new datasets and all web components will be freely shared with the community, so other researchers can embed or otherwise use them in their own web-based tools. By integrating UniProt and MaveDB we aim to help a diverse set of investigators contextualize and interpret DMS data to better understand the unique role of each amino acid position and protein structure-function relationships in general.

Keywords: MaveDB, UniProt, mapping, DMS, annotations, structure, function

Multiplexed interrogation of lipoprotein lipase missense variant effects

Presenter: Daniel Tabet (University of Toronto)

Daniel Tabet (1,2,3,4), Atina G. Coté (1,2,3,4), Olivia Taverniti(1,2,3,4), Jochen Weile (1,2,3,4), Ashyad Rayhan (1,2,3,4), Roujia Li (1,2,3,4) & Frederick P. Roth (1,2,3,4,*)

(1) The Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada
(2) Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada
(3) Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada
(4) Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, ON M5G 1X5, Canada *Corresponding author

Coronary heart disease (CHD) remains a leading cause of death in North America. Due to the success of primary prevention therapeutics, CHD is an ideal candidate for genetic screening as a basis for disease risk diagnosis. Yet, much of the heritability of CHD is likely attributable to difficult-to-predict rare variation, substantially limiting the applicability of clinical genetic testing. To overcome this barrier, highly multiplexed sequence-function studies can be applied to exhaustively evaluate the effects of all coding alterations on protein function, thereby improving our capacity to determine patient disease risk from genetic sequence. We have applied this method to the gene lipoprotein lipase (LPL), which works in the homeostatic regulation of lipid

levels by hydrolyzing triglyceride-rich lipoprotein particles (i.e., chylomicrons and very low-density lipoproteins). By using a fluorescent antibody against the surface-expressed LPL protein, we can separate populations by variant impact using fluorescent activated cell sorting. We thus detect loss of protein function en masse, reliably separating pathogenic from benign variants. We expect that the resulting variant effect maps will enable a more sensitive and accurate clinical diagnosis of LPL-deficiency.

Keywords:

Precision diagnostics for GLUT1 disorders using deep mutational scanning

Presenter: Naeimeh Tayebi (Washington university in St. Louis)

Naeimeh Tayebi, Kevin McCall, Brian Leon Ricardo, Paul Hruz, Ashley Quiggle, Christina A. Gurnett

Department of Neurology, Washington University in Saint Louis School of Medicine, Saint Louis, Missouri, USA

Objective: Accurate genetic diagnosis is critical, particularly as effective therapies become available for neurological disorders. The accumulation of variants of uncertain significance (VUS) creates a growing crisis that negatively influence the ability to implement precision therapies. The goal of this research is to demonstrate the utility of a growth assay to quantify the functional impact of variants in SLC2A1, the gene responsible for GLUT1 deficiency syndrome.

Methods: 55 variants across three exons were evaluated using a growth assay in HAP1-Lig4KO cell line. We created a donor library consisting of 15 variants for exon 10, 22 variants for exon 2 and 18 variants for exon 3 in pathogenic, benign, and VUS categories. Variant libraries were introduced into the HAP1-lig4KO cell line using exon specific CRISPR/Cas9. Cell populations were harvested and sequenced at days 5 and 11.

Results: By sequencing pools of cells at days5 and 11, we could quantitative determine the impact of variants on cell viability. As proof of principle, Nonsense and known pathogenic variants in all three exons dropped out of the population, while known benign variants were not depleted. In addition, SLC2A1 missense variants in exon 10 (p.R458P) and exon 3 (p.W65R and G76S) identified in the patients with childhood onset epilepsy had intermediate functional effects in this assay.

Conclusions: Our quantitative functional data correlated with clinical phenotypes, suggesting that it can also be clinically useful for prognosis. Future work is needed to scale the assay such that every possible variant in SLC2A1 can be quantitatively determined.

Keywords:

A mutational map for protein stability in Canavan´s disease

Presenter: Martin Thygesen (University of Copenhagen)

Martin Grønbæk-Thygesen1, Vasileios Voutsinos1, Matteo Cagiada1, Kristoffer E. Johansson1, Lene Clausen1, Snehal Nariya2, Douglas M. Fowler2, Kresten Lindorff-Larsen1, Rasmus Hartmann-Petersen1

1: Department of Biology, The Linderstrøm-Lang Centre for Protein Science, University of Copenhagen, Copenhagen, Denmark. 2: Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Aspartoacylase (ASPA) is a 313-residue enzyme responsible for catalyzing the hydrolysis of N-acetyl-aspartic acid (NAA) into aspartate and acetate. The enzyme is expressed mainly in the oligodendrocytes of the brain where it plays an indispensable role in brain-metabolism. Loss-of function variants in the ASPA encoding gene leads to Canavan´s disease; an incurable and lethal neurological disease characterized by spongiform degeneration of the brain white matter. A recent study has shown that ASPA misfolding and degradation plays an important role for the disease1. To gain further information into how the disease-linked ASPA variants operate, we used Variance Abundance by Massive Parallel sequencing (VAMP-seq)2 to map the relative protein abundance for thousands of ASPA variants. The resulting mutational map comprises >98% of all possible single-site missense and nonsense APSA variants. The cellular abundance data correlate with in silico predictions of ASPA structural stability, evolutionary conservation, and separates known disease-linked variants from harmless variants. Systematic mapping of degradation signals (degrons) shows that inherent primary degrons in ASPA largely overlap with regions that are highly sensitive to mutations. The vast majority of unstable ASPA variants are degraded through the ubiquitin-proteasome system and are stabilized at lowered temperatures. In conclusion, the obtained data may provide a useful diagnostic tool, but also reveals both protein specific and general information on the specificity of the protein quality control system and the ubiquitin-proteasome pathway.

1Gersing et al. (2021) Mapping the degradation pathway of a disease-linked aspartoacylase variant. PLoS Genet. 17, e1009539.

2Matreyek et al. (2018) Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat. Genet. 50, 874-882.

Keywords: ASPA, Aspartoacylase, VAMP-seq, protein degradation, protein abundance, Canavan´s disease

MutateX, a FoldX-based pipeline for high-throughput prediction of effects of mutations on protein structural stability and binding

Presenter: Matteo Tiberti (Cancer Structural Biology, Danish Cancer Society Research Center, Strandboulevarden 49, 2100, Copenhagen, Denmark)

Matteo Tiberti (1), Thilde Terkelsen (1), Kristine Degn (2), Ludovica Beltrame (1), Tycho Canter Cremers (1), Isabelle da Piedade (1), Miriam Di Marco (1), Emiliano Maiani (1) and Elena Papaleo (1,2)

(1) Cancer Structural Biology, Danish Cancer Society Research Center, Strandboulevarden 49, 2100, Copenhagen, Denmark
(2) Cancer Systems Biology, Section of Bioinformatics, Health and Technology Department, Technical University of Denmark, Lyngby, Denmark

Large-scale cancer genomics data sets as well as collections of cancer-related mutations are becoming increasingly available. Predicting their effect on protein structural stability and binding with partners can be an important layer of evidence to predict the potential effect of cancer mutations. Such predictions can complement and synergize with tools based on other data sources. We are working towards this goal by combining a software for the retrieval and annotation of cancer mutations from different data sources with structure-based predictions, which can be performed using MutateX [1]. MutateX is a comprehensive pipeline to predict the effect of mutations on protein structural stability or binding starting from three-dimensional structures of proteins. It is designed to obtain complete (i.e. mutation of all residues to all natural amino acids) or partial mutational scans with a single command line, using the FoldX [2] method and program for the underlying calculations. It implements a complete in-silico mutagenesis protocol, including repair of the starting structure(s), mutation, and calculation of the effect on free energy of folding or binding between binding partners. MutateX also performs calculations on multiple structures from experiments or simulations, which can be useful to compensate for the sensitivity of the method with respect to using any specific single structure. As, especially in the latter case, a large number of FoldX runs becomes necessary, MutateX is able to take advantage of modern multi-core computing architectures to achieve very significant speed-ups with respect to FoldX alone. Finally, MutateX includes a suite of visualization and data manipulations tools to help make sense of the large amounts of data generated by mutational scans. MutateX is available at https://www.github.com/ELELAB/mutatex and requires the FoldX program, which can be acquired separately.

The result of this research has been achieved using the DECI-PRACE15th and 16th HPC Grants on Archer. The research has been also carried out thanks to the access to the Danish HPC Infrastructure Computerome2. The research has been supported by grants from LEO Foundation (LF17006, LF17024), Carlsberg Distinguished Fellowship (CF18-0314), Danmarks Grundforskningsfond (DNRF125), NovoNordisk Fonden in Bioscience and Basic Biomedicine (0065262) and Biotechnology-based Synthesis and Production (17OC0027588) and Hartmann Foundation (A33877). This work is part of Interregional Childhood Oncology Precision Medicine Exploration, a cross-Oresund collaboration between University Hospital Copenhagen, Rigshospitalet, Lund University, Region Skåne and Technical University Denmark, supported by the European Regional Development Fund.

[1] Matteo Tiberti, Thilde Terkelsen, Kristine Degn, Ludovica Beltrame, Tycho Canter Cremers, Isabelle da Piedade, Miriam Di Marco, Emiliano Maiani, Elena Papaleo, MutateX: an automated pipeline for in silico saturation mutagenesis of protein structures and structural ensembles, Briefings in Bioinformatics, 2022;, bbac074, https://doi.org/10.1093/bib/bbac074

[2] Javier Delgado, Leandro G Radusky, Damiano Cianferoni, Luis Serrano, FoldX 5.0: working with RNA, small molecules and a new graphical interface, Bioinformatics, Volume 35, Issue 20, 15 October 2019, Pages 4168–4169, https://doi.org/10.1093/bioinformatics/btz184

Keywords: free energy, structure, stability, binding, foldx, mutatex

Interpreting disease variants in human membrane proteins using protein sequence and structure

Presenter: Johanna K.S. Tiemann (University of Copenhagen)

Johanna K.S. Tiemann (1), Henrike Zschach (1), Kresten Lindorff-Larsen (1*), Amelie Stein (1*)

(1)Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N., Denmark

Next-generation-sequencing of human genomes reveals millions of missense variants, which can lead to loss of protein function and disease. Membrane proteins are key drivers in cell signaling and recognition and form an important group of disease-relevant proteins. Here, we collect all 10,000 variants in the human membrane proteome that are classified as benign or pathogenic, and observe an enrichment of pathogenic variants within the transmembrane region. In soluble proteins, loss of stability is often the underlying mechanism of pathogenicity. To gain mechanistic insight into membrane protein variants, we perform structure-based calculations of changes in thermodynamic stability on a subset of 15 proteins, and analyse evolutionary conservation by sequence variation. The latter helps pinpoint variants critical for function, but not stability. We find that on average, 62% of pathogenic variants lose function via loss of stability, indicating that protein stability is a driving factor also in membrane-protein-associated diseases. Our findings showcase how computational tools such as stability calculations aid in gaining mechanistic insights into variant consequences for membrane proteins.

Keywords: membrane proteins, pathogenicity, protein folding and stability, genome variability & evolution

Understanding the impact of indels on protein stability and function using deep mutational scanning

Presenter: Magdalena Topolska (Centre for Genomic Regulation, Barcelona, Spain)

Magdalena Topolska (1,2) and Ben Lehner (1,2,3)

(1) Centre for Genomic Regulation (CRG) Barcelona, Spain (2) University Pompeu Fabra (UPF), Barcelona, Spain (3) Institució Catalana de Recerca i estudis Avançats (ICREA), Barcelona, Spain

Amino acid insertions and deletions (indels) are common mutational events but there is limited information about their effects on protein stability and function. Whereas substitutions only alter the amino acid side chains of proteins, indels are ‘backbone mutations’ that modify protein length with potentially severe and difficult-to-predict effects.

To address the lack of data on the impact of indels, we are systematically quantifying their effects on the stability and protein binding of a diverse set of human protein domains.

The goal is to build a reference atlas of mutational effects for insertions and deletions. This dataset will be useful for guiding clinical variant interpretation and it should also provide general insights into indel tolerance across different folds. More importantly, it will also be a dataset of sufficient scale and diversity to allow the benchmarking and training of computational methods to predict the impact of indels, with applications in clinical genetics and protein engineering.

We will present the approach that we are using and the data from a pilot study.

Keywords:

Conundrums and challenges of clinical variant interpretation

Presenter: Clare Turnbull (Institute of Cancer Research)

Clare Turnbull

Institute of Cancer research

Many clinical actions taken in response to detection of a genomic variant are binary, for example prophylactic mastectomy or termination of a fetus. Underlying the simple dichotomous clinical decision, there lies a variant-specific deleteriousness and a corresponding variant-specific penetrance for disease. Clinical variant classification comprises the stitching together of a range of indirect observations in individuals, populations, related species and laboratory assays in order to infer this variant-specific penetrance. Which we translate clinically into the terms ‘pathogenicity’ or ‘benignity’. I shall explore principles, pitfalls and forthcoming challenges of clinical variant interpretation, presenting exemplar analyses we have undertaken in paradigms of genetic cancer susceptibility.

Keywords: clinical, variant, interpretation

A mutational atlas for Parkin

Presenter: Vasileios Voutsinos (University of Copenhagen)

Lene Clausen (1), Vasileios Voutsinos (1), Matteo Cagiada (1), Kristoffer E. Johansson (1), Martin Grønbæk-Thygesen (1), Snehal Nariya (3), Vibe H. Østergaard (2), Amelie Stein (2), Douglas M. Fowler (3), Kresten Lindorff-Larsen (1), Rasmus Hartmann-Petersen (1)

1: Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Copenhagen, Denmark.

2: Department of Biology, University of Copenhagen, Copenhagen, Denmark.

3: Departments of Genome Sciences, University of Washington, Seattle, WA, USA.

Mutations in the PRKN gene, encoding the E3 ubiquitin-protein ligase Parkin are a major cause of autosomal recessive juvenile Parkinsonism. The majority of pathogenic mutations are associated with reduced Parkin protein levels Here, we characterized abundance of 9219 out of the 9300 possible missense variants of the PRKN protein by utilizing variant abundance by massively parallel sequencing (VAMP-seq). We observe that the majority of destabilizing mutations are located within the structured domains of the protein, while the flexible linker regions are more tolerant to mutations. One exception being a region within the linker connecting the UBL domain with the RING0 domain, which is sensitive to substitutions to more hydrophobic residues, while substitutions to more hydrophilic residues result to stabilization of the protein. We observed that most low abundance variants are heat-sensitive and proteasome targets. By screening oligopeptides tiling over the entire protein, we also mapped inherent degrons in Parkin, which we found to overlap with positions that are sensitive to mutations. Finally, we demonstrate that Parkin variant abundance correlates with both the predicted structural stability and evolutionary conservation.

Keywords: VAMP-seq, Parkin

Deciphering 3’ UTR expression regulation with convolutional neural networks and endogeneous massively parallel reporter assays

Presenter: Omar Wagih (Deep Genomics)

Fraser Soares (1*), Jeff Wintersinger (1*)*, Kathleen Delfosse (1), Omar Wagih (1)

1, Deep Genomics, * Equal contribution

The 3’ untranslated regions (3’ UTRs) play a complex but critically important role in post-transcriptional regulation. These regions contain functional elements, such as binding sites for microRNAs and RNA binding proteins, that modulate expression of the products encoded by their transcripts. Mutations in these regions can disrupt transcript expression, processing, and localization, leading to disease. Consequently, understanding the role played by regulatory elements in the 3’ UTR can help with predicting variant effect for mutations in this region, while also aiding the design of therapies that modulate expression to alleviate disease.

The relationship between 3’ UTR sequence and expression regulation is complex. Functional elements in these regions interact with diverse mechanisms that often differ between cell types. Massively parallel reporter assays (MPRAs) are a powerful tool for overcoming these challenges by enabling researchers to interrogate tens of thousands of UTRs for their effects on gene expression in a single experiment. Using pooled reporter plasmids, MPRAs can demonstrate whether disease-associated genetic variants affect transcript expression, thereby providing valuable information about the functional impact of variants. Additionally, MPRAs can systematically ablate putative regulatory elements in UTRs of interest, illustrating how each portion of a UTR affects expression in a cell-type dependent manner.

Traditional MPRAs are limited to querying UTR fragments of 200 nt or less. However, because the median mature human 3’ UTR sequence is 1,000 nt in length, interactions between multiple functional elements in the UTR may be lost in these experiments. To overcome this shortcoming, we developed an improved MPRA that incorporates the full endogenous UTR for transcripts of interest, which we term the endogenous MPRA (eMPRA). Using a novel mutagenesis strategy, eMPRAs incorporate the entire UTR sequence for each gene target while incorporating mutations of interest. We demonstrate a proof-of-concept eMPRA in the low-density lipoprotein receptor (LDLR) gene, where we determined how 30,000 mutagenized versions of the full 2.5 kb 3’ UTR affected expression. Additionally, we show that data from these experiments can be combined with previously published MPRA datasets to train deep-learning models that learn to disentangle with single-nucleotide resolution the different categories of microRNA and RBP binding sites. These models can ultimately aid in the interpretation of variant effects and designing of new therapies.

Keywords: mpra,3’ utr,machine learning,deep learning,variant effect

Integration of multi-omics data with saturation mutagenesis data to assess biological impact on a systems level

Presenter: Yue Wang (University of North Carolina at Chapel Hill)

Yue Wang (1), Stephanie Ting (3), Elizabeth Brunk (1,2,3,4,5)

1. Department of Pharmacology at University of North Carolina at Chapel Hill
2. Department of Chemistry, at University of North Carolina at Chapel Hill

3. Integrated Program in Biological Genome Sciences, at University of North Carolina at Chapel Hill
4. Computational Medicine Program, at University of North Carolina at Chapel Hill
5. Lineberger Cancer Center, at University of North Carolina at Chapel Hill

Multiplexed assays for variant effects (MAVEs) enable saturation mutagenesis and functional analysis of regulatory elements or proteins of interest at nucleotide precision. In recent years, MAVEs demonstrate great clinical potential for determining the significance of variants of unknown significance (VUS). In cancer research, the majority of variants have VUS status and it remains a considerable challenge to distinguish functionally relevant somatic mutations from benign mutations. In one of the most established cancer omics databases, the Cancer Cell Line Encyclopedia (CCLE), over 1000 cancer cell lines have been mutationally profiled and the collective mutational landscape consists of over 1.05 million unique protein-coding variants. The CCLE also houses the largest repository for functional omics data (e.g. transcriptomics, proteomics, metabolomics, CRISPR loss-of-function screens, and drug response assays) in cancer cell lines. In this study, we aim to integrate the recently published MAVE data for two oncogenes, BRCA1[1] and TP53[2], with multi-omics data from the CCLE. Through integration of these complementary data types, we infer the impact of gene variation in BRCA1 and TP53 genes on downstream biochemical and regulatory networks in breast cancer cell lines. Using a semi-supervised approach, we develop inference models that predict association between the functionally-impactful variants (from MAVE assays) and activated pathways (from CCLE transcriptomics, proteomics and metabolomics data). We verify functional connections between variants and their impacted pathways through analysis of CRISPR-mediated loss of function screens[3]. Integration across MAVE and multi-omics data presents novel opportunities for utilizing saturation mutagenesis to assess biological impact on a systems level.

Reference:

[1] Giacomelli AO, Yang X, Lintner RE, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. Nat Genet. 2018;50(10):1381-1387. doi:10.1038/s41588-018-0204-y

[2] Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. Saturation editing of genomic regions by multiplex homology-directed repair. Nature. 2014;513(7516):120-123. doi:10.1038/nature13695

[3] Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol. 2013;31(3):230-232. doi:10.1038/nbt.2507

Keywords: deep-mutational scanning, TP53, BRCA1, breast cancer, variant-function relationships, CCLE

Saturation Genome Editing of Cancer Genes

Presenter: Andrew Waters (Wellcome Sanger Institute)

Andrew Waters (1), David Adams (1), Victoria Offord (1), Sofia Obolenski (1,2), Rebeca Olvera-Leon (1,3)

1-Wellcome Sanger Institute ; 2-Leiden University, 3-University of Cambridge

Pathogenic germline variants in BAP1, a deubiquitinase enzyme and tumour suppressor, are associated with uveal melanoma, mesothelioma, cutaneous melanoma and renal cell carcinoma, with the full spectrum of associated tumours yet to be determined. BAP1 is also frequently somatically mutated across a spectrum of cancer types. We have optimized the design, cell model, culture conditions and computational analysis of SGE to increase efficiency and the scale of experiments and have focused our initial studies on BAP1, in addition to other tumour suppressors such as POT1, RAD51C and the mismatch repair proteins. Whilst some BAP1 variants are known to be pathogenic, many observed variants within BAP1 are of uncertain significance or have conflicting interpretations of pathogenicity. SGE data shows a clear separation of function for known pathogenic and known benign variants, which has allowed us to attribute functional scores to all variants in the gene. We have further been able to apply these data to resolve cancer-predisposed pedigrees with profound implications for patient management.

We ultimately aim to perform SGE on all known cancer genes, in order to functionally characterize all variation in these genes at a nucleotide level and to facilitate precision cancer medicine.

Keywords: SGE, Cancer, HAP1, CRISPR-Cas9

Translating multiplexed variant effect assays into evidence codes for clinical interpretation

Presenter: Jochen Weile (University of Toronto)

Jochen Weile (1-4), Roujia Li (1-5), Rachel A Silverstein (1-4,6), Frederick P Roth (1-4)

1 The Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada 2 Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, ON M5G 1X5, Canada 3 Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada 4 Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada 5 Present address: Deep Genomics Inc., 661 University Avenue, MaRS Centre West Tower Suite 480, Toronto, ON, M5G 1M1, Canada 6 Present address: Division of Medical Sciences, Harvard Medical School, 260 Longwood Ave, Boston, MA 02115, USA

Variant effect (VE) maps produced via multiplexed functional assays promise to stem the ever growing tide of variants of uncertain significance (VUS). These maps excel at quantifying the impact of variants on the molecular function of a protein. However, the relationship between VE map scores and health outcomes is generally not linear. A relatively small decrease in protein function may already be sufficient to cause disease. For genes where pathogenic alleles typically act via a dominant negative mechanism such as Calmodulin, variants with a modest functional impact may be pathogenic while those that entirely eliminate protein function may be benign. Assays may have regional differences in sensitivity and precision, such as those previously observed between the catalytic and regulatory domains of MTHFR.

Clinical variant interpreters already use complex guidelines to determine evidence weights for functional data. Despite this complexity, the same weight is typically applied to all variants found damaging in a given assay and are based on simple thresholds. We propose a more data-driven and nuanced Bayesian approach that calculates the appropriate evidence weight for each variant within a VE map. In a two-step approach, we first find functions translating VE map score into a log-likelihood ratio (LLR) of pathogenicity. In the second step, LLRs can be expressed in terms of categorical evidence codes compatible with established variant interpretation guidelines. Separate LLR transformation functions can be determined for different regions of a map. We demonstrate the utility of our translation scheme for existing VE maps and evaluate reliability using leave-one-out cross-validation.

Keywords: MAVE, Variant interpretation, ACMG, VUS

DeepAllostery: comprehensive identification of allosteric sites in the oncogenic molecular switch RAS

Presenter: Chenchun Weng (center for genomic regulation(CRG))

Chenchun Weng (1) and Ben Lehner (1,2,3)

1 Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology Barcelona, Spain 2 University Pompeu Fabra (UPF), Barcelona, Spain 3 Institució Catalana de Recerca i estudis Avançats (ICREA), Barcelona, Spain

Proteins are usually ‘switchable’ with their enzymatic activities and molecular interactions controlled by binding partners, covalent modifications and mutations outside of their active sites. This transmission of information spatially from one site to another in a protein is termed allostery, which is central to nearly all biological regulation. In addition, many of the most effective drugs target allosteric sites to inhibit or modulate protein function.

A good example of the importance of allostery is provided by the RAS oncoproteins. Allosteric mutations switch RAS into an active state that binds diverse effector proteins to drive cancer. In addition, the first licensed RAS inhibitor is an allosteric drug that binds a cryptic pocket to lock mutant KRAS into its inactive state. However, as for nearly all proteins, there is no global map of allosteric sites in RAS proteins and allosteric regulation, in general, remains poorly understood and very difficult to predict.

To address these shortcomings, we have used mutational scanning to chart a comprehensive map of allosteric communication in KRAS. We quantied the effects of mutations on the solubility (abundance) of KRAS and its binding to six different interaction partners. Measuring how mutations interact in double mutants allows us to infer the underlying biophysical effects of mutations i.e. changes in the free energies of protein folding and binding to six different interaction partners.

The resulting energy landscapes provide important insights into binding specificity and a global map of allosteric regulation. We will present the construction of these energy landscapes and the insights they provide into protein multifunctionality and allosteric communication.

Keywords: KRAS, protein-protein interaction, allostery

Comprehensive survey of the functional impact of SNP accessible variants in PSAT1 using a high throughput yeast assay.

Presenter: Michael Xie (University of Washington)

Michael Xie (1,2), Russell Lo (1), Gareth A Cromie (1), Katherine Owens (1,3), Marty Timour (1), Julee Ashmead (1), J. Nathan Kutz (3), Richard N. McLaughlin Jr. (1), and Aimée M Dudley (1,2)

(1) Pacific Northwest Research Institute, Seattle, Washington, USA
(2) Molecular Engineering Program, University of Washington, Seattle, Washington, USA
(3) Department of Applied Mathematics, University of Washington, Seattle, Washington, USA

Loss of function mutations in genes of the serine biosynthesis pathway (PHGDH, PSAT1, and PSPH) cause a set of rare, autosomal recessive diseases known as Neu-Laxova syndrome (NLS) or serine deficiency disorders. The phenotypic spectrum of these inborn errors of metabolism includes severe neurological symptoms, failure to thrive, and lethality. However, timely detection of pathogenic variants can have an enormous, positive impact on patient health. Prenatal and postnatal supplementation with serine and glycine can ameliorate, and in some cases, completely prevent the onset of symptoms. To aid variant annotation for this rare and devastating disease, we developed a yeast “surrogate genetics” assay in which the highly conserved human protein coding sequence for the second enzyme in the pathway, phosphoserine aminotransferase (PSAT1), functionally replaces the deletion of its yeast ortholog (SER1). In this assay, yeast growth in the absence of serine provides a quantitative readout of the human enzyme’s function. Previous work from our group demonstrated that results from this assay agree well with clinical annotations and the disease literature. Here, we extend that analysis to present quantitative results for the functional impact of >80% (n=2182) of all single nucleotide polymorphism (SNP) accessible PSAT1 missense variants. Our approach leverages the availability of low-cost, large-scale gene synthesis and the development of high-throughput in vivo assays of protein function (Multiplexed Assays for Variant Effect, MAVEs) in the model organism, Saccharomyces cerevisiae. We will discuss the extent to which these results agree with a set of recently published pathogenic variants and the spectrum of Variants of Uncertain Significance (VUS) that have been detected in the human population. We will also present our results in the context of the published PSAT1 crystal structure. Taken together, our work provides an example of the ways in which large-scale functional assays in model systems can be powerfully applied to the study of rare diseases.

Keywords: Neu-Laxova syndrome, Serine Deficiency Syndrome, Rare Diseases, Inborn Errors of Metabolism, Functional Assays, Multiplexed Assays for Variant Effect

Measuring viral escape from polyclonal antibodies

Presenter: Timothy Yu (Fred Hutchinson Cancer Center)

Timothy Yu (1), Frances Welsh (1), Bernadeta Dadonaite (1), Zorian Thornton (1), William Dewitt (1), Andrea Loes (1), Frederick Matsen (1), Jesse Bloom (1)

(1) Fred Hutchinson Cancer Center

Viral evolution is driven by mutations that enable viruses to escape antibody recognition. However, multiple mutations are often required to fully escape antibodies in human sera, which are polyclonal and can target several distinct epitopes. While deep mutational scanning has been used to measure the antigenic effects of all single mutations to viral proteins, our ability to predict antibody escape of multiply-mutated viral variants remains remarkably poor. This is due to our limited understanding of the epistasis that transpires from antibody-epitope interactions, which cannot be empirically explored by single-mutant deep mutational scans. Here we address this shortcoming by extending deep mutational scanning to measure the effects of combinations of mutations to viral entry proteins on escape from antibodies in human sera. We then fit a biophysical model that captures the shape of epistasis to this data and can accurately predict antibody escape of unseen, multiply-mutated viral variants. We apply this approach to two viral proteins that are major targets of neutralizing antibodies, H3 influenza hemagglutinin and SARS-CoV-2 spike. We find that our biophysical model can deconvolve the polyclonal antibody response to reveal the locations of specific epitopes that are recognized by neutralizing antibodies and infer the mutations that hinder this recognition.

Keywords: viral escape

Humanized yeast to measure the functional impact of human genetic variation in the mevalonate kinase gene

Presenter: Farhat Zafar (Concordia University)

Farhat Zafar (1), Nishka Kishore (2,3), Frederick P. Roth (2,3), Aashiq H. Kachroo (1)

1.Centre for Applied Synthetic Biology, Department of Biology, Concordia University, Montreal, QC, CANADA. 2.The Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, ON, CANADA. 3.Lunenfeld-Tanenbaum Research Institute, Toronto, ON, CANADA.

Advances in sequencing technology have allowed the study of human genetic variation. Thousands of human genomes are currently available, identifying several thousand variants. A grand challenge is to determine what, if any, impact each missense mutation has on human health. The fact that many human genes can functionally replace their corresponding yeast equivalents provides an opportunity to use humanized yeast to model human genetic variation [1,2]. By swapping essential yeast genes

with human counterparts, we can link the fitness of the variant human protein with the fitness of the yeast cell. Therefore, mutations that have functional consequences will show phenotypic differences in the humanized yeast, such as slower growth rate. In addition, our system allows for tunable control of gene expression levels, differing from studies that look only at complementation through overexpression. We demonstrate this approach on the rare human genetic disorder known as mevalonate kinase deficiency by devising a strategy to conditionally replace yeast ScERG12 gene with the orthologous human mevalonate kinase gene, Hs-MVK. Next, we show the yeast growth as an easily measured proxy for the proper functioning of the human gene. By humanizing yeast with gene variants, followed by sequencing as a readout to inform each strain's relative growth, we can score the functional impact of human gene variation. Thus, neutral variants are easily distinguished from deleterious variants. This work, in addition to previous similar strategies [3], establishes a platform for using humanized yeast to model human genetic variation at scale. Through the application of a deep mutational scanning approach, we hope to greatly contribute to the current data on mevalonate kinase mutations. And as a result, further our understanding of mevalonate kinase deficiency. References: 1. Kachroo, A.H., et al., Science (2015). (http://dx.doi.org/10.1126/science.aaa0769) 2. Laurent, J.M., et al., (2019). bioRxiv (https://doi.org/10.1101/668335) 3. Weile, J., et al., (2017). Molecular Systems Biology (https://doi.org/10.15252/msb.20177908)

Keywords: humanization, yeast, mevalonate kinase, genetic variation, complementation assay, health, next-gen sequencing

Understanding the genetic encoding of dynamic and non-dynamic interactions in a model PDZ domain

Presenter: Taraneh Zarin (Centre for Genomic Regulation (CRG), Barcelona)

Taraneh Zarin(1,2), Ben Lehner(1,2,3,4)

(1) Center for Genomic Regulation (CRG)
(2) The Barcelona Institute of Science and Technology
(3) Universitat Pompeu Fabra (UPF)
(4) Institució Catalana de Recerca i Estudis Avançats (ICREA)

Domain-motif interactions are central to signaling and regulation in eukaryotic cells. Many such interactions take place through structured binding modes, in which the interacting macromolecules adopt a specific structure upon binding. However, a significant portion of these motifs (and domains) occur in a disordered protein context, where the macromolecule in question does not adopt a stable three-dimensional structure. Such dynamic binding modes have been observed through in vitro experiments and molecular simulations, but it is not yet clear how they are encoded and constrained in the genome. Here, we use a deep mutational scanning approach on a well-characterized PDZ domain-motif interaction that contains structured and dynamic binding modes in adjacent amino acid residues. Using combinatorially complete libraries of variants for each binding mode and double mutant cycles between the motif and domain of interest, we observe not only how each mode of binding is encoded separately, but also how they interplay with each other and the domain of interest to enable function.

Keywords: disordered, domain, peptide, binding, dynamic

Higher-order epistasis and phenotypic prediction

Presenter: Juannan Zhou (University of Florida)

Juannan Zhou(1), Mandy S. Wong (2), Wei-Chia Chen(3), Adrian R. Krainer(4), Justin B. Kinney(5), David M. McCandlish(5)

1. Department of Biology, University of Florida, Gainesville, FL, 32611
2. Skyhawk Therapeutics, Inc. 35 Gatehouse Drive Waltham, MA 02451
3. Department of Physics, National Chung Cheng University, Chiayi 62102, Taiwan (R.O.C.)
4. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724
5. Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

Contemporary high-throughput mutagenesis experiments are providing an increasingly detailed view of the complex patterns of genetic interaction that occur between multiple mutations within a single protein or regulatory element. By simultaneously measuring the effects of thousands of combinations of mutations, these experiments have revealed that the genotype-phenotype relationship typically reflects genetic interactions not only between pairs of sites, but also higher-order interactions among larger numbers of sites. However, modeling and understanding these higher-order interactions remains challenging. Here, we present a method for reconstructing sequence-to-function mappings from partially observed data that can accommodate all orders of genetic interaction. The main idea is to make predictions for unobserved genotypes that match the type and extent of epistasis found in the observed data. This information on the type and extent of epistasis can be extracted by considering how phenotypic correlations change as a function of mutational distance, which is equivalent to estimating the fraction of phenotypic variance due to each order of genetic interaction (additive, pairwise, three-way, etc.). Using these estimated variance components, we then define an empirical Bayes prior that in expectation matches the observed pattern of

epistasis, and reconstruct the genotype-phenotype mapping by conducting Gaussian process regression under this prior. To demonstrate the power of this approach, we present an application to the antibody-binding domain GB1 and also provide a detailed exploration of a dataset consisting of high-throughput measurements for the splicing efficiency of human pre-mRNA 5’ splice sites, for which we also validate our model predictions via additional low-throughput experiments.

Keywords: genotype-phenotype map, Gaussian processes, epistasis, genetic interaction, splicing