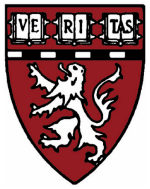
Using web-based annotation tools for bioinformatic analyses of proteomics data

Kasper Lage, PhD

Massachusetts General Hospital

Harvard Medical School

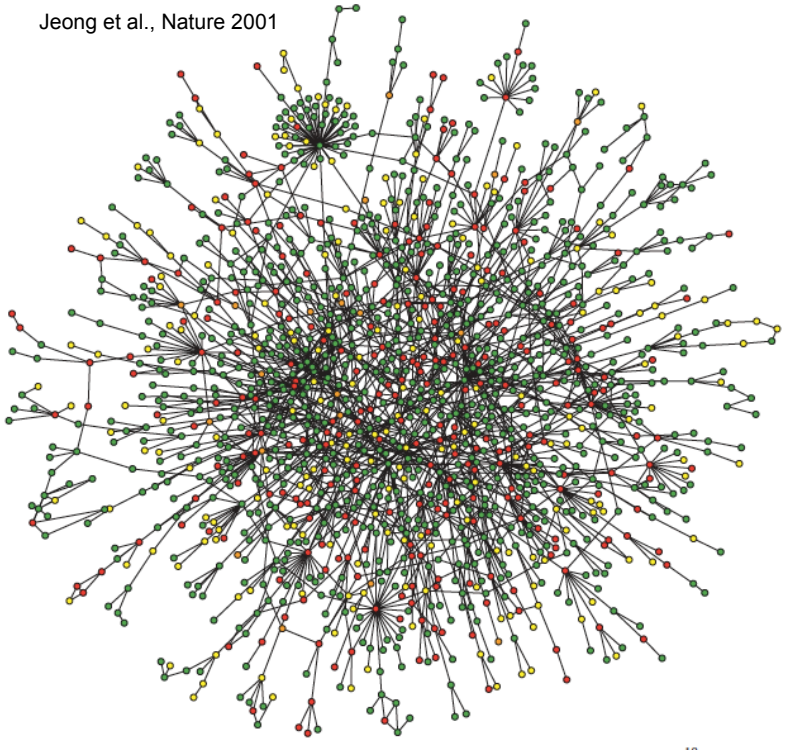
the Broad Institute



Overview of this session

1)	Biological networks
2)	Annotating genetic and proteomic data using biological networks
3)	Tissue-specific networks with disease resolution
4)	Emerging resources at the Broad Institute

What is a biological network?



1) Gene expression correlations

2) Protein-protein interactions

3) Co-mentioning in text

4) Synthetic lethality

5) TF binding

6) Pathway database mining

7) Epigenetic data

8) All of the above

Building a human protein-protein interaction network - InWeb

Email lage.kasper@mgh.harvard.edu if you want to use the data.

MINT

BIND

DIP

GRID

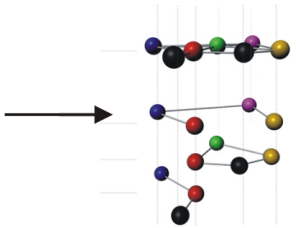
HPRD

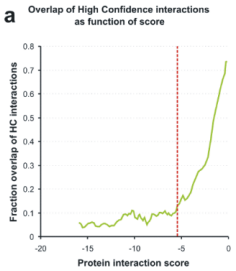
KEGG

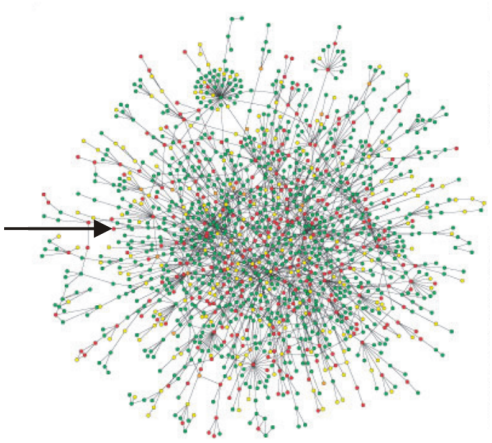
REACTOME

IntAct

Curated







Download and reformat PPI databases

Trans - organism PPI transferral

Automated scoring of all interactions

433,000 interactions between 10,300 human proteins updated every six months

Lage, Karlberg et al., Nat. Biotech., 2007

Lage, Hansen et al. PNAS, 2008

Lage et al., Mol Syst Biol, 2010

Rossin et al., PLoS Genetics, 2011

Lage et al., PNAS, 2012

Social human networks are a good model for understanding biological networks

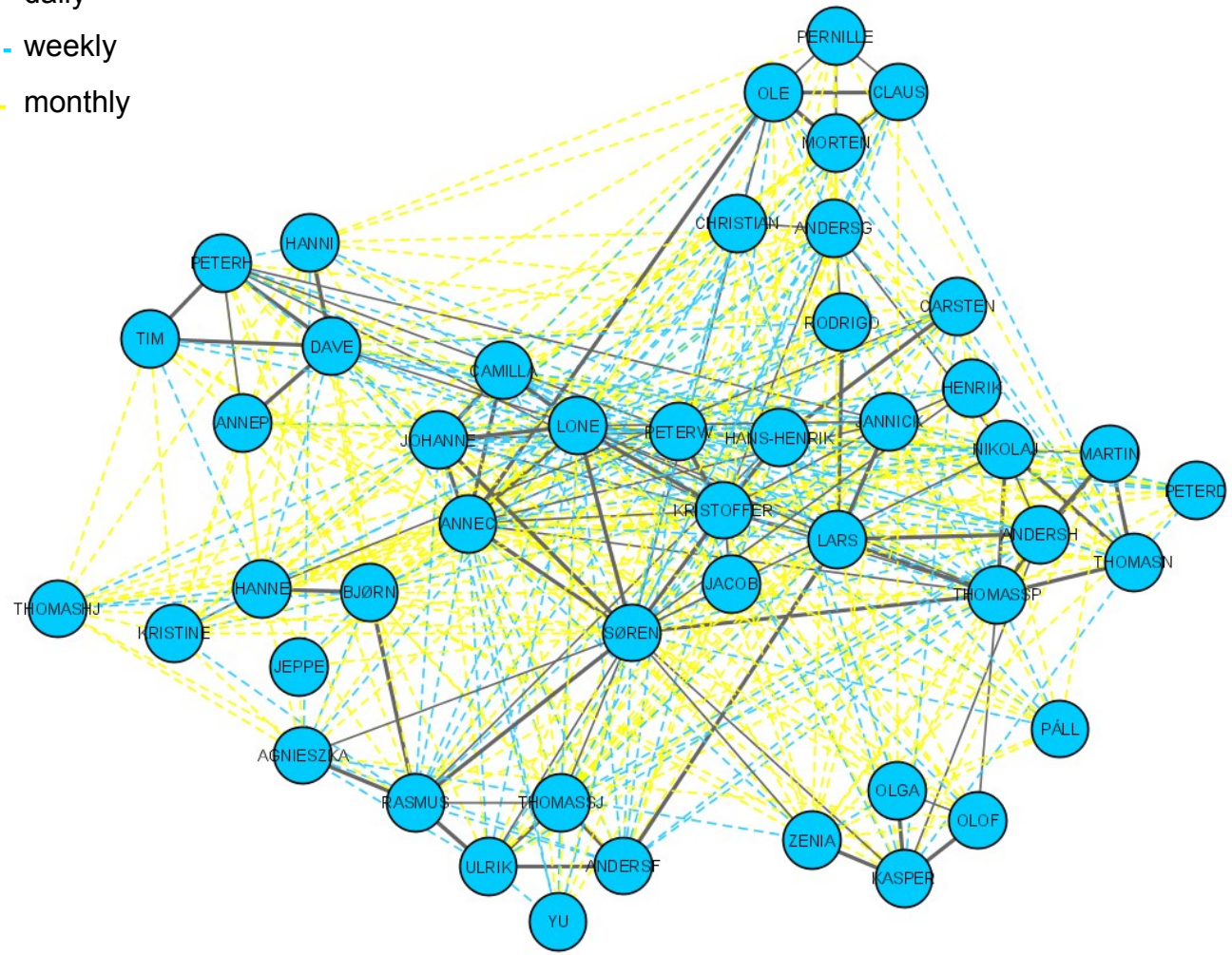
Social networks

People are represented by “nodes”, work related interactions by “edges”

daily

weekly

monthly



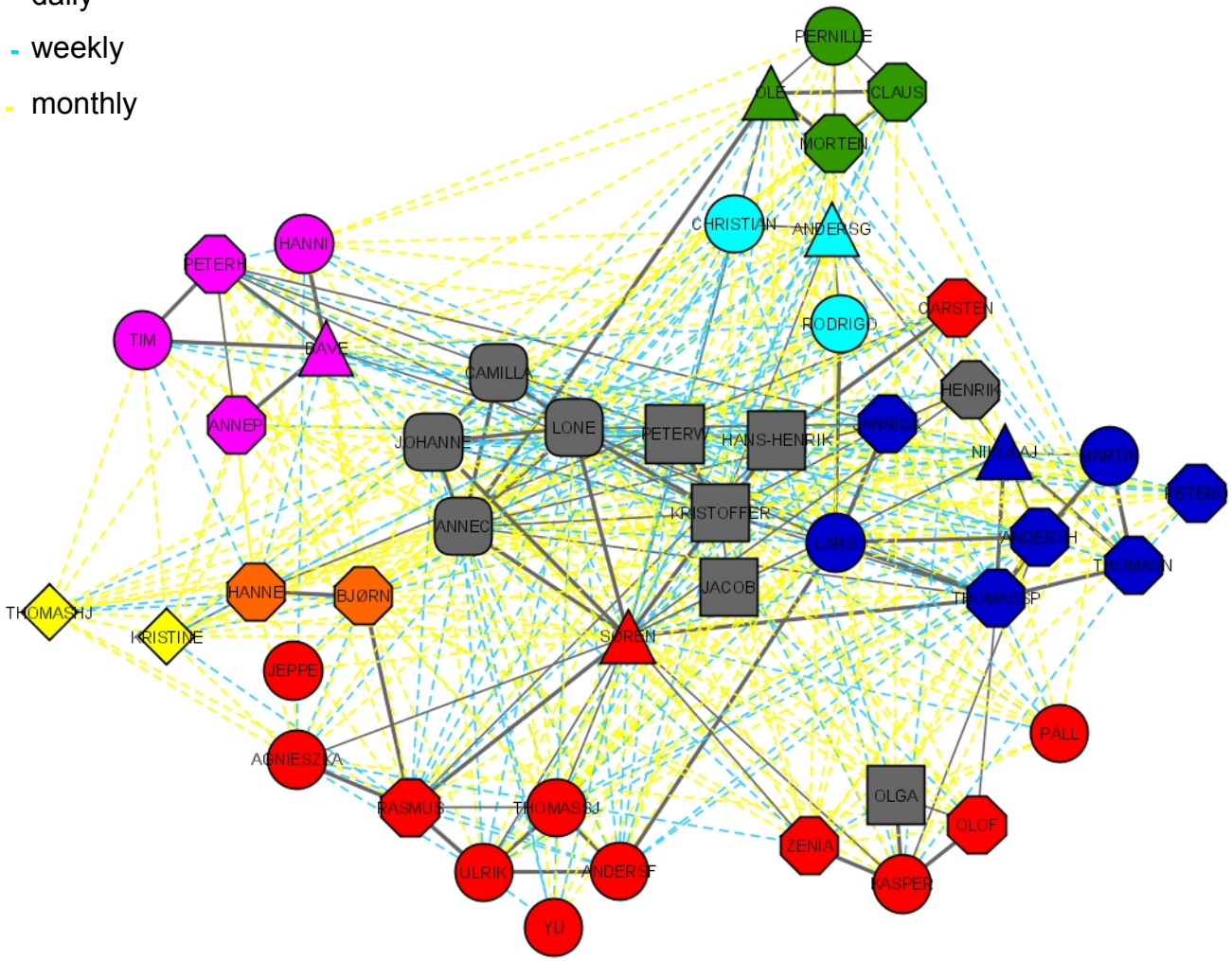
Social networks

People that work together are close to each other in the network

daily

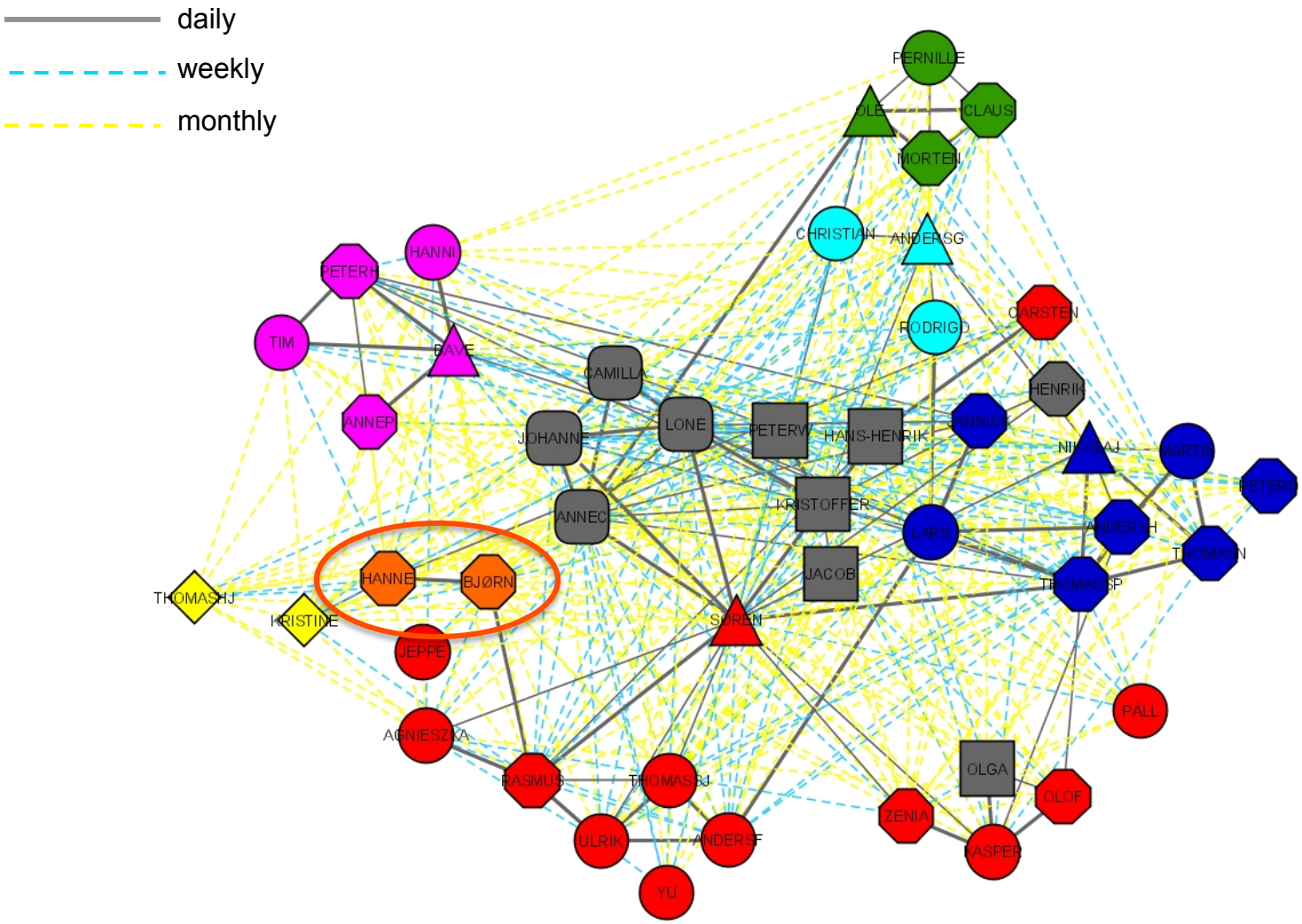
weekly

monthly



Social networks

People that work together are close to each other in the network



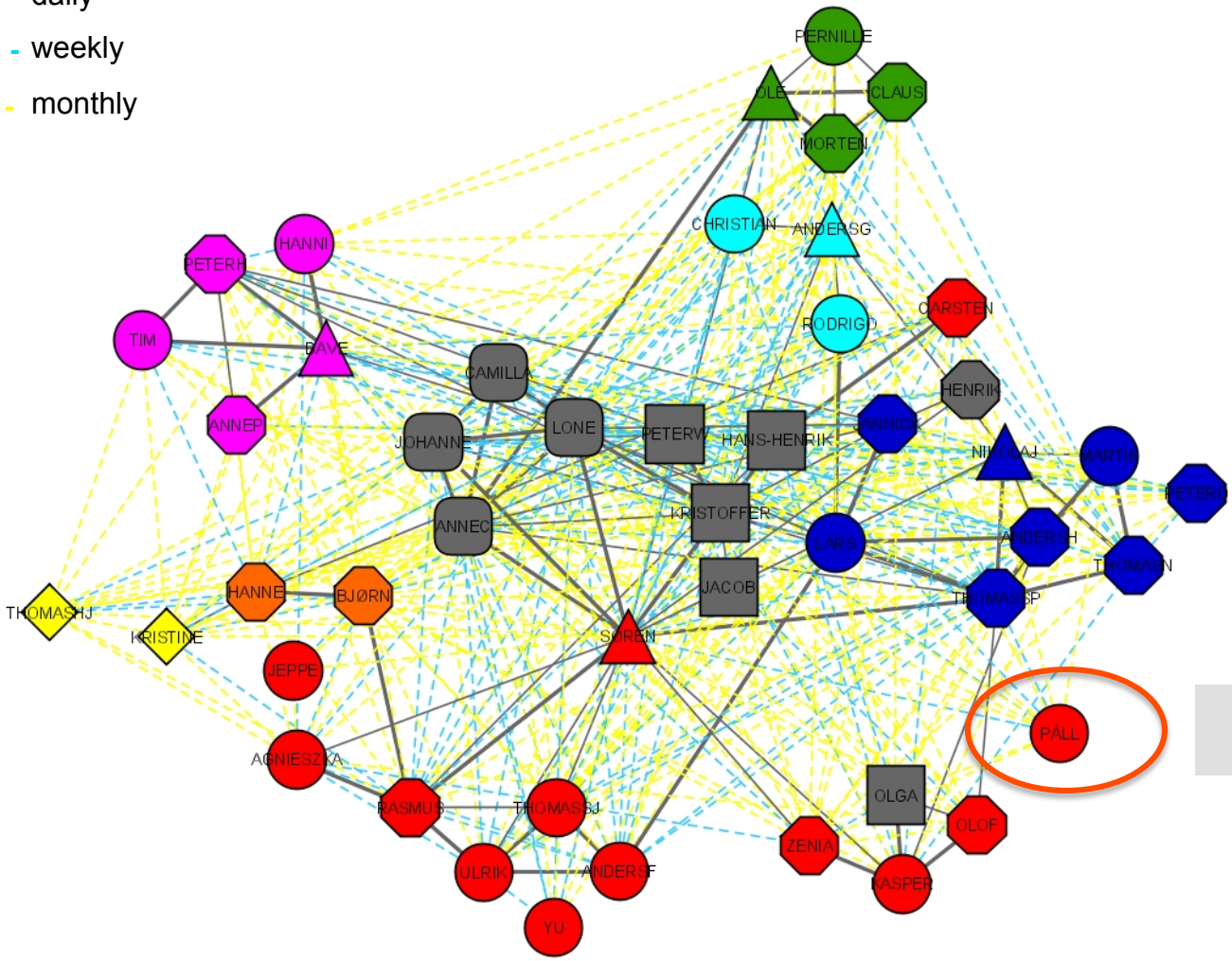
Social networks

People that work together are close to each other in the network

daily

weekly

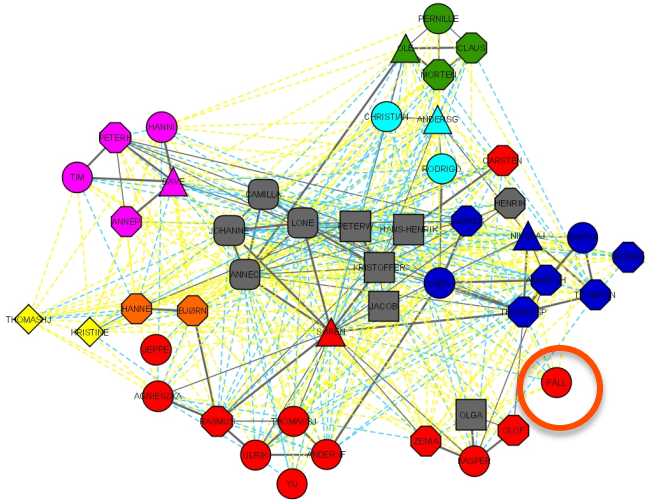
monthly



Lone wolf

Social networks

People that work together are close to each other in the network



W468–W470 *Nucleic Acids Research, 2005, Vol. 33, Web Server issue*
doi:10.1093/nar/gki463

Integrating protein annotation resources through the Distributed Annotation System

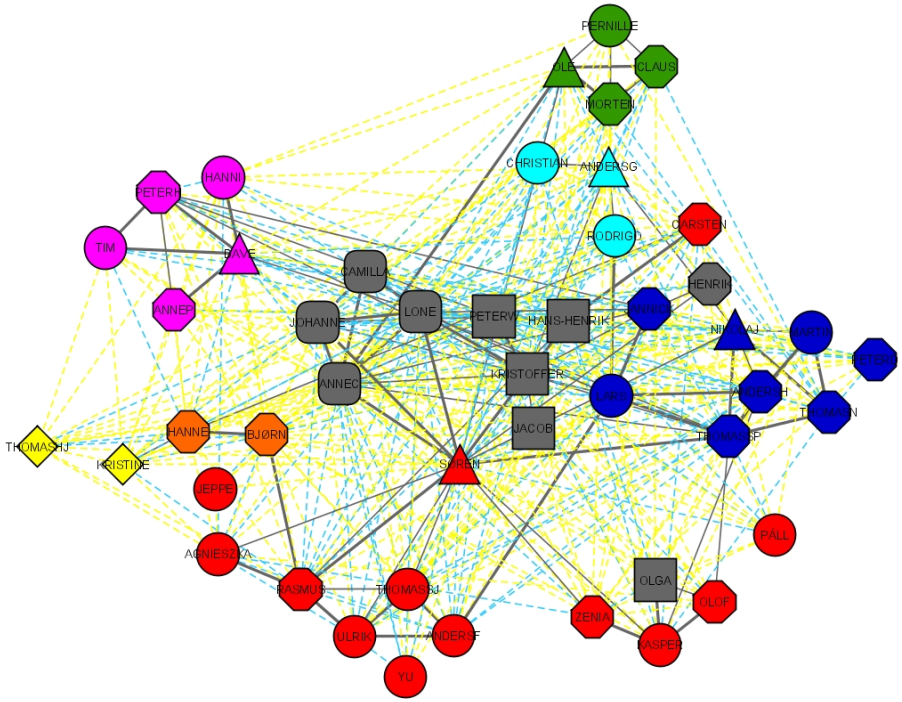
Páll Ísólfur Ólason*

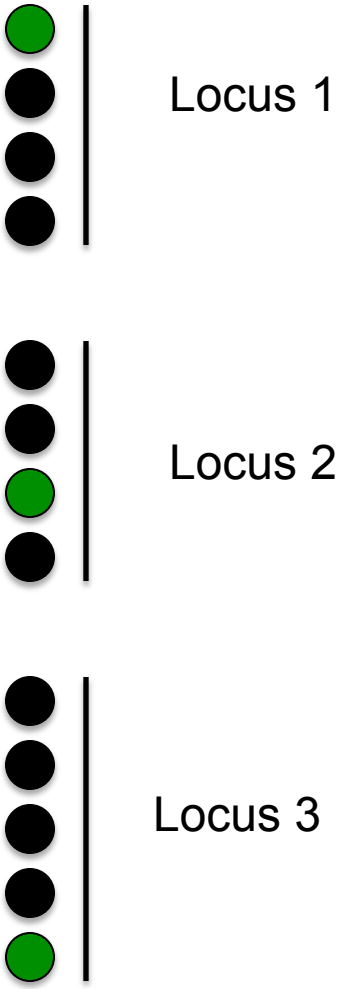
Center for Biological Sequence Analysis BioCentrum-DTU, Building 208, Technical University of Denmark, DK-2800 Lyngby, Denmark

Received February 14, 2005; Revised and Accepted April 13, 2005

Do loci connect more than a random expectation?

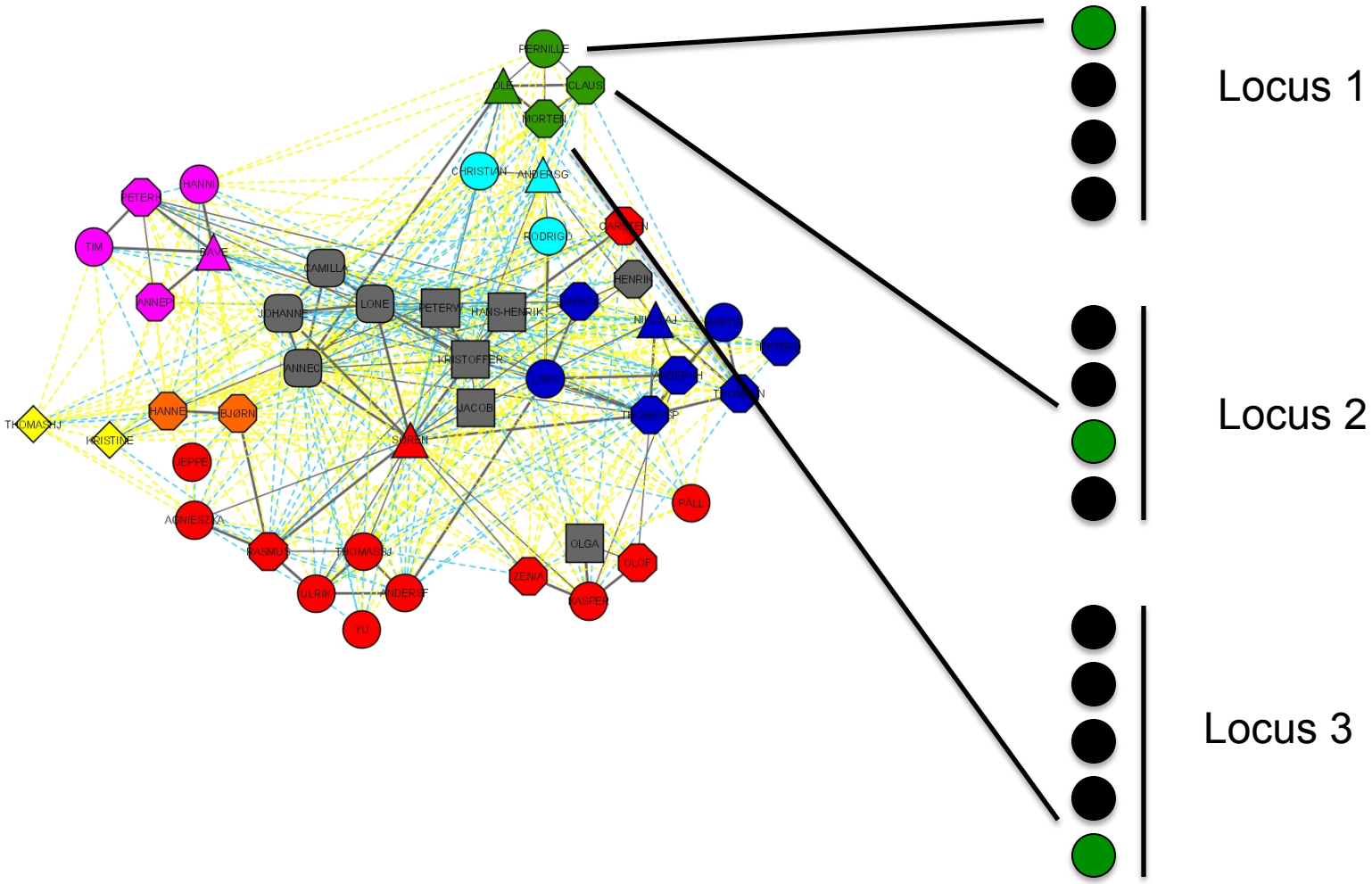
Quantify connectivity and determine the significance using permutations





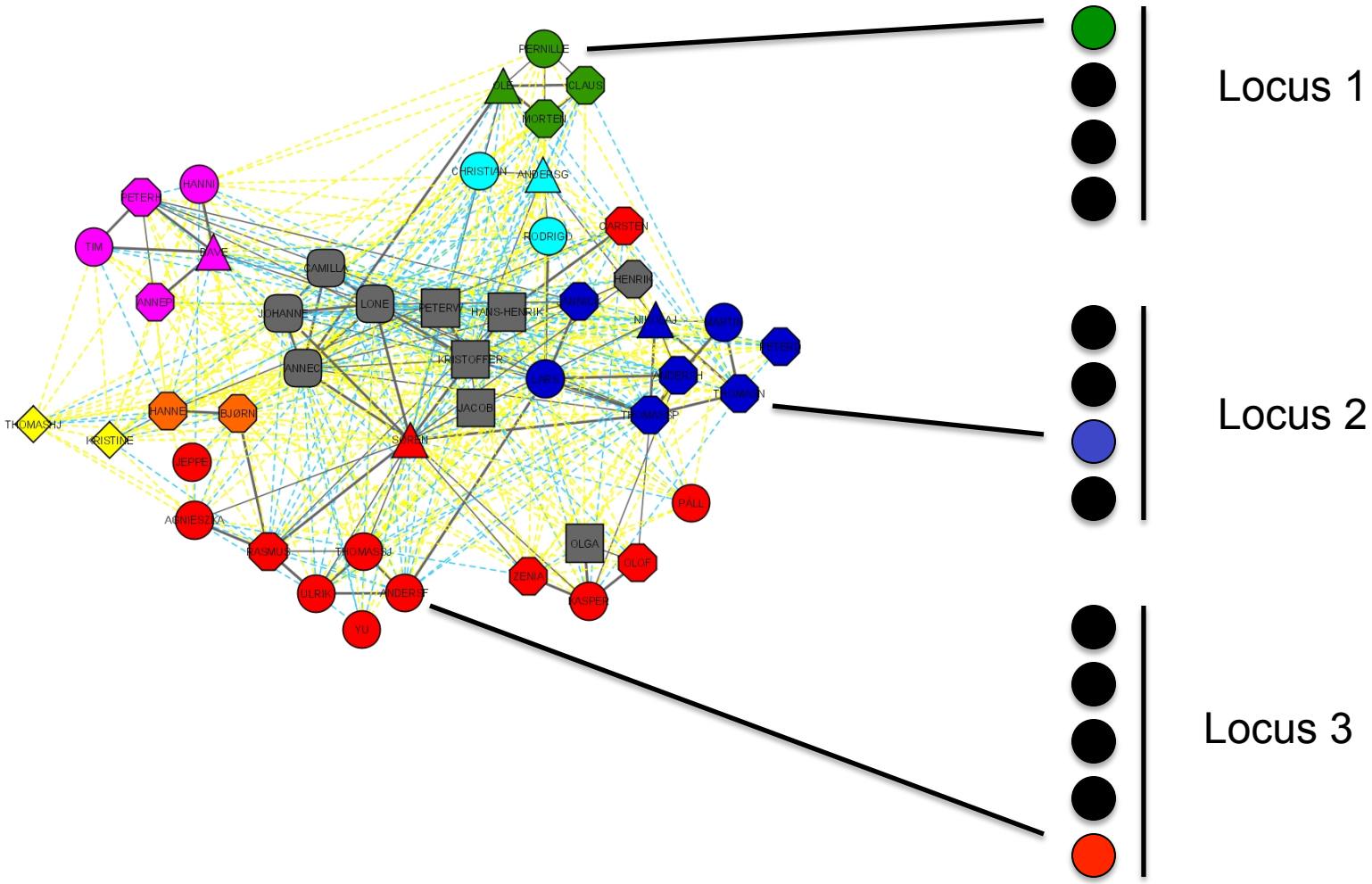
Do loci connect more than a random expectation?

Quantify connectivity and determine the significance using permutations



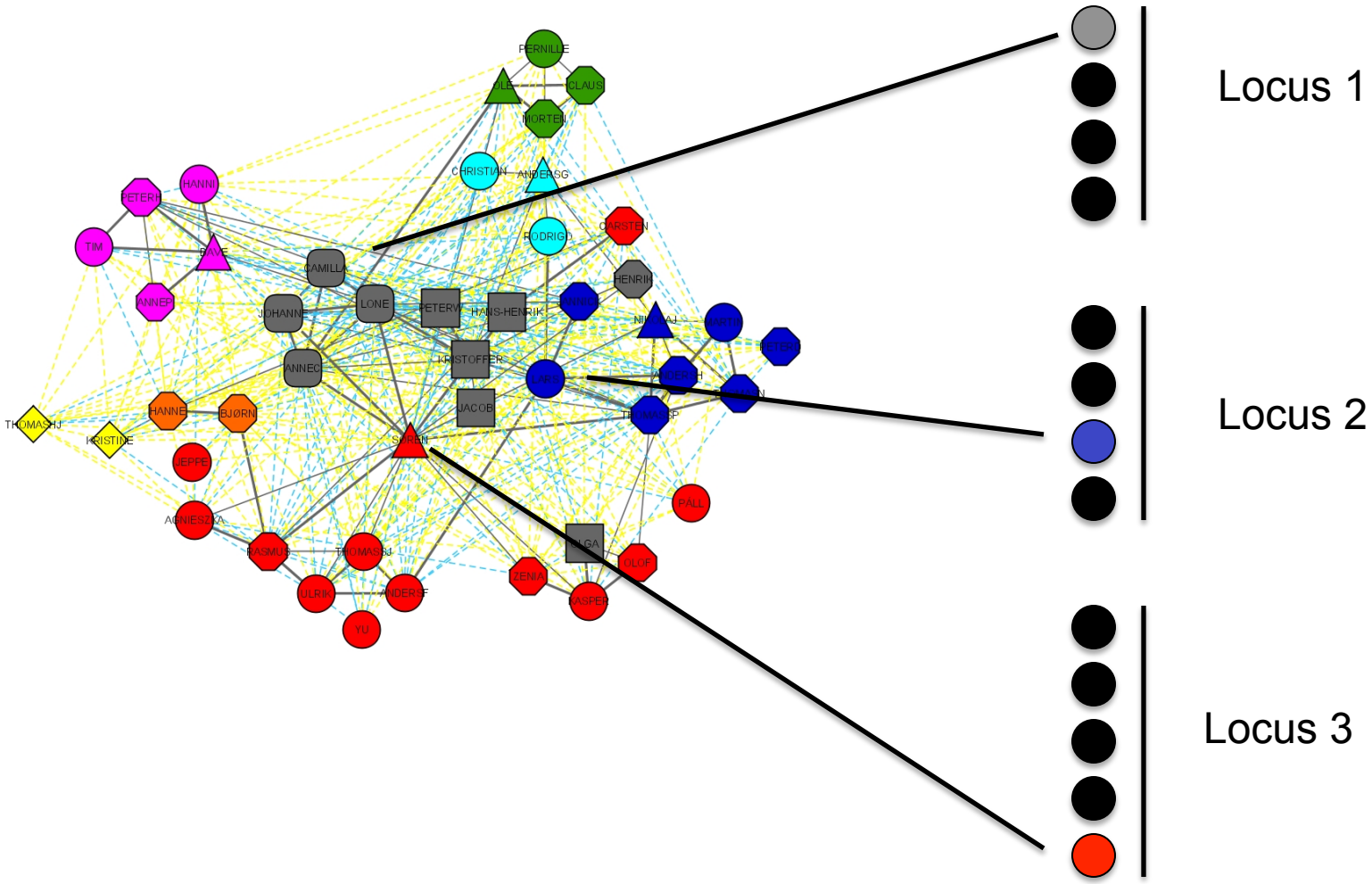
Do loci connect more than a random expectation?

Quantify connectivity and determine the significance using permutations



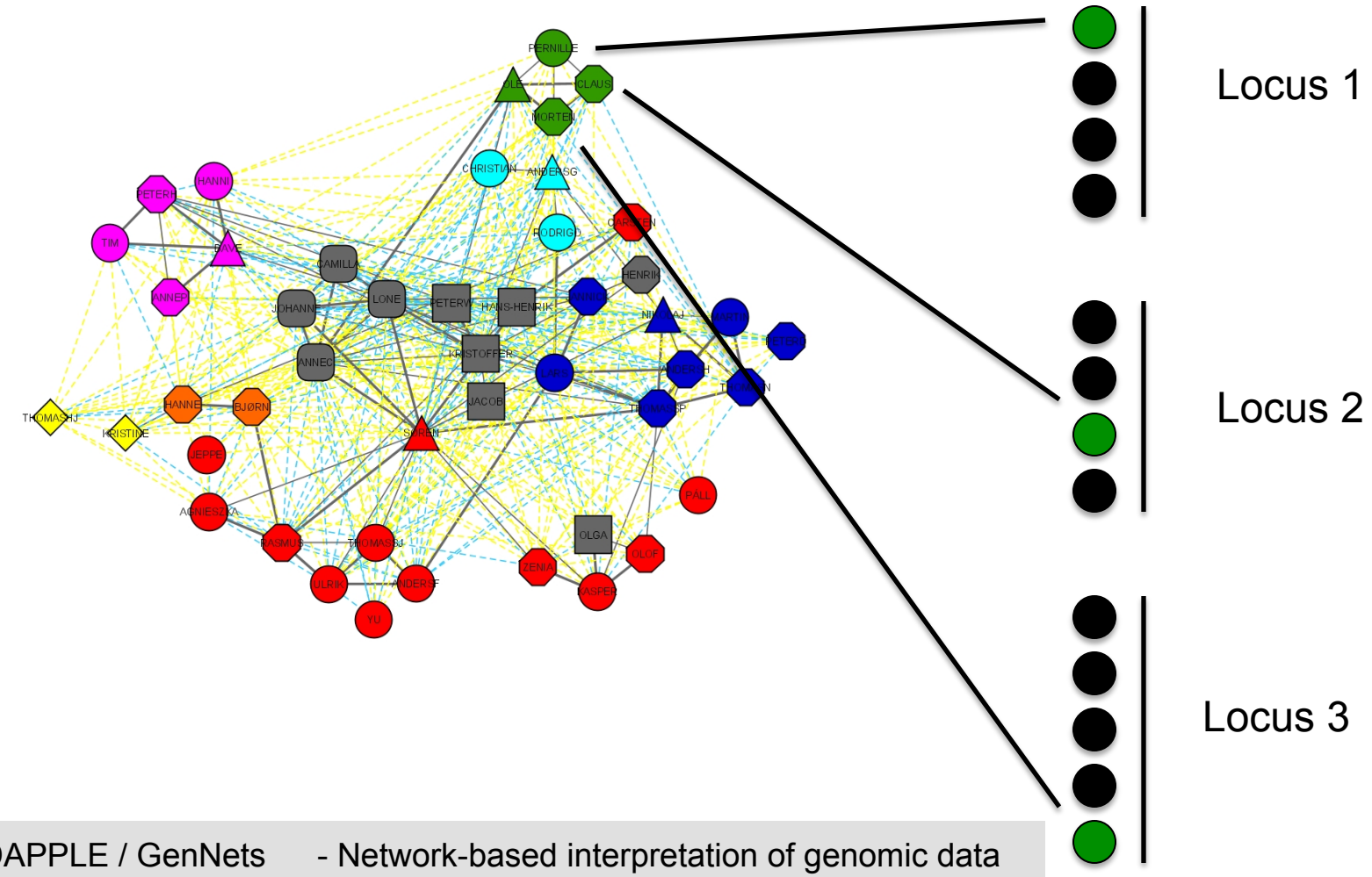
Do loci connect more than a random expectation?

Quantify connectivity and determine the significance using permutations



Do loci connect more than a random expectation?

Quantify connectivity and determine the significance using permutations



DAPPLE / GenNets - Network-based interpretation of genomic data
NetworKIN - Predicting in vivo kinase-substrate relationships

Overview of this session

1)	Biological networks
2)	Annotating genetic and proteomic data using biological networks
3)	Tissue-specific networks with disease resolution
4)	Emerging resources at the Broad Institute

An online method to make statistically tested networks from gene lists – DAPPLE

OPEN ACCESS Freely available online

PLoS GENETICS

Proteins Encoded in Genomic Regions Associated with Immune-Mediated Disease Physically Interact and Suggest Underlying Biology

Elizabeth J. Rossin[1,2,3,4,5], Kasper Lage[2,3,6,7], Soumya Raychaudhuri[1,2,8], Ramnik J. Xavier[1,2,3], Diana Tatar[6], Yair Benita[1], International Inflammatory Bowel Disease Genetics Constortium[¶], Chris Cotsapas[1,2☯], Mark J. Daly[1,2,3,4,5☯*]

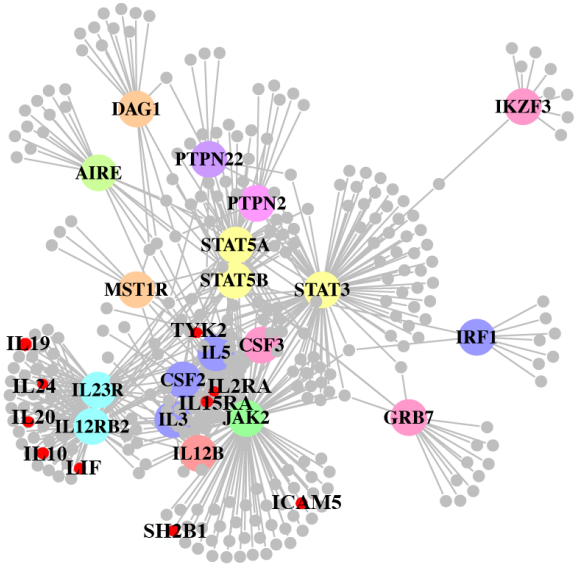
Lizzy Rossin, Daly Lab, DAPPLE

www.broadinstitute.org/mpg/dapple/

Constructing and statistical testing of protein networks

InWeb and DAPPLE analyses of GWAS data

Crohn's disease



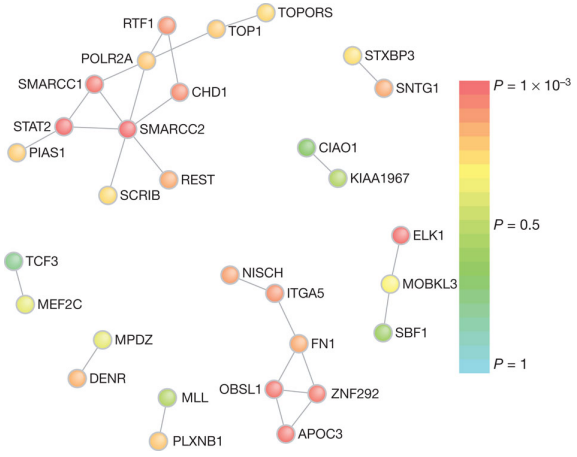
(Rossin et al., PLoS Genetics 2011)

DAPPLE: www.broadinstitute.org/mpg/dapple/

Constructing and statistical testing of protein networks

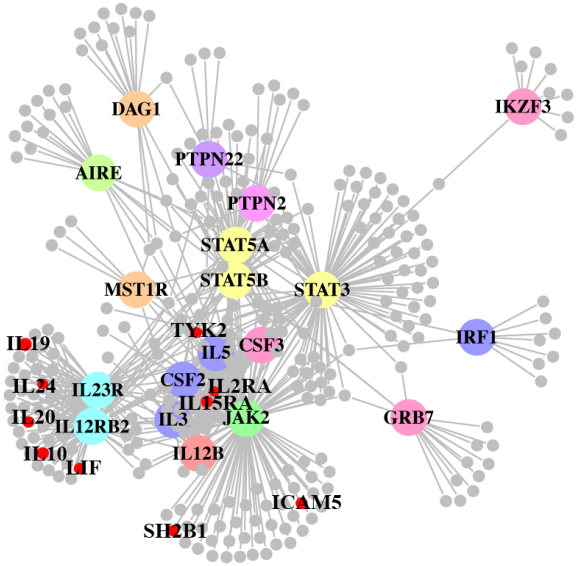
InWeb and DAPPLE analyses of GWAS data

Autism



(Neale et al., Nature 2012)
Similar findings in
(O'Roak et al., Nature 2012)

Crohn's disease

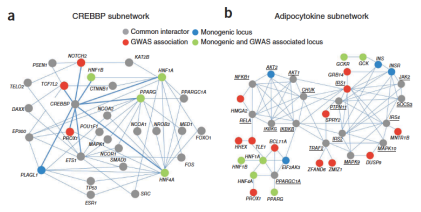


(Rossin et al., PLoS Genetics 2011)
DAPPLE: www.broadinstitute.org/mpg/dapple/

Interpreting genomes using biological networks

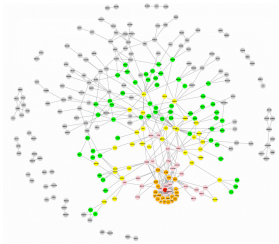
InWeb and DAPPLE analyses of exome sequencing or GWAS data

Type 2 diabetes



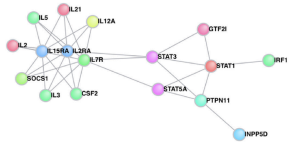
(Morris et al., Nature Genetics 2012)

Inflammatory bowel disease



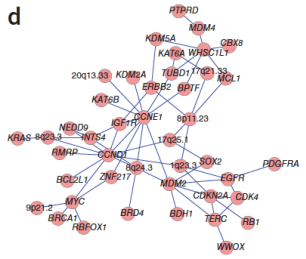
(Jostins L, Ripke S, et al., Nature 2012)

Inflammatory risk variants under recent positive selection



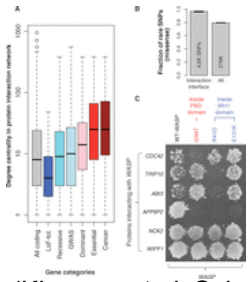
(Raj et al., AJHG 2013)

Pan cancer CNVs

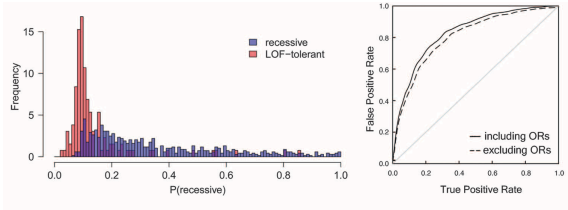


(Zack et al., Nature Genetics 2013)

Population genetic data from the 1000 Genomes Consortium

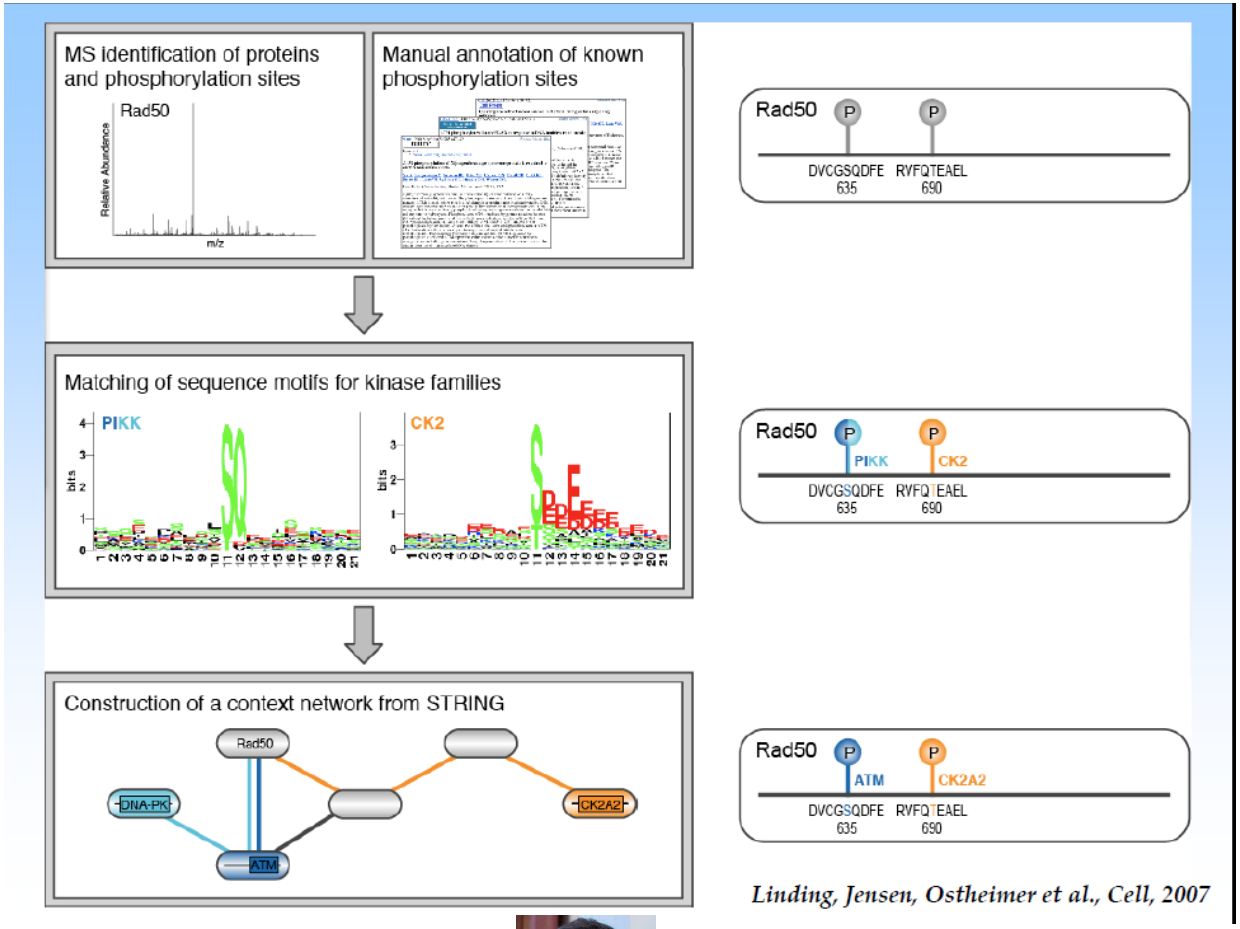


(Khurana et al.,Science 2013)
***InWeb not DAPPLE**



(MacArthur et al., Science 2012)
****Not InWeb or DAPPLE**

In Vivo Mapping of Kinases to Substrates - NetworKIN



Linding, Jensen, Ostheimer et al., Cell, 2007

Heiko Horn
hhorn@partners.org

Making networks with STRING

Info & Parameters ...

Network Display - Nodes are either colored (if they are directly linked to the input - as in the table) or white (nodes of a higher iteration/depth). Edges, i.e. predicted functional links, consist of up to eight lines: one color for each type of evidence. Hover or click to reveal more information about the node/edge.

Active Prediction Methods:

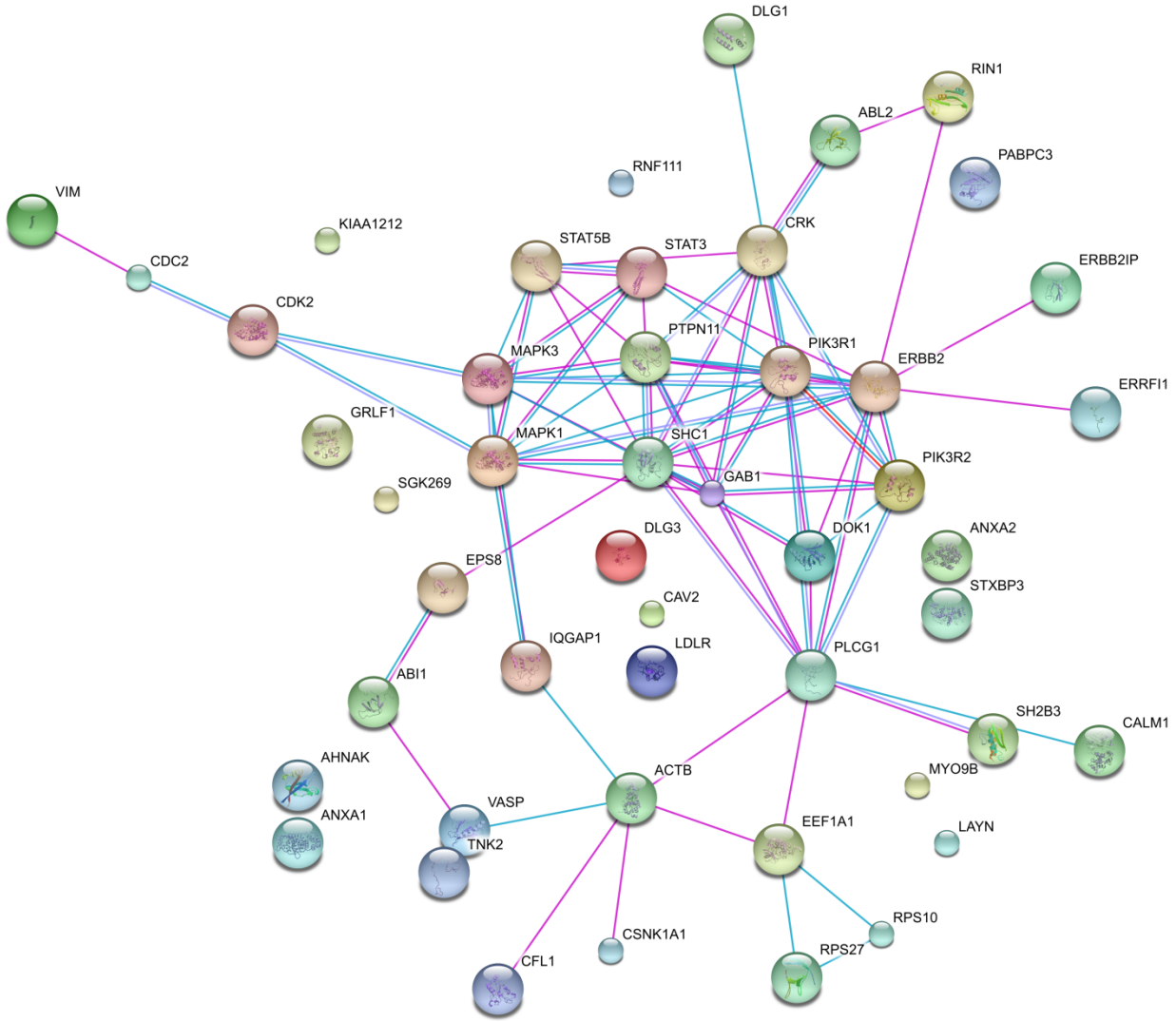
☐ Neighborhood ☑ Gene Fusion ☐ Co-occurrence
☐ Co-expression ☑ Experiments ☑ Databases ☐ Textmining

required confidence (score): medium confidence (0.400)
or custom value:

interactors shown: custom limit
custom limit: 0

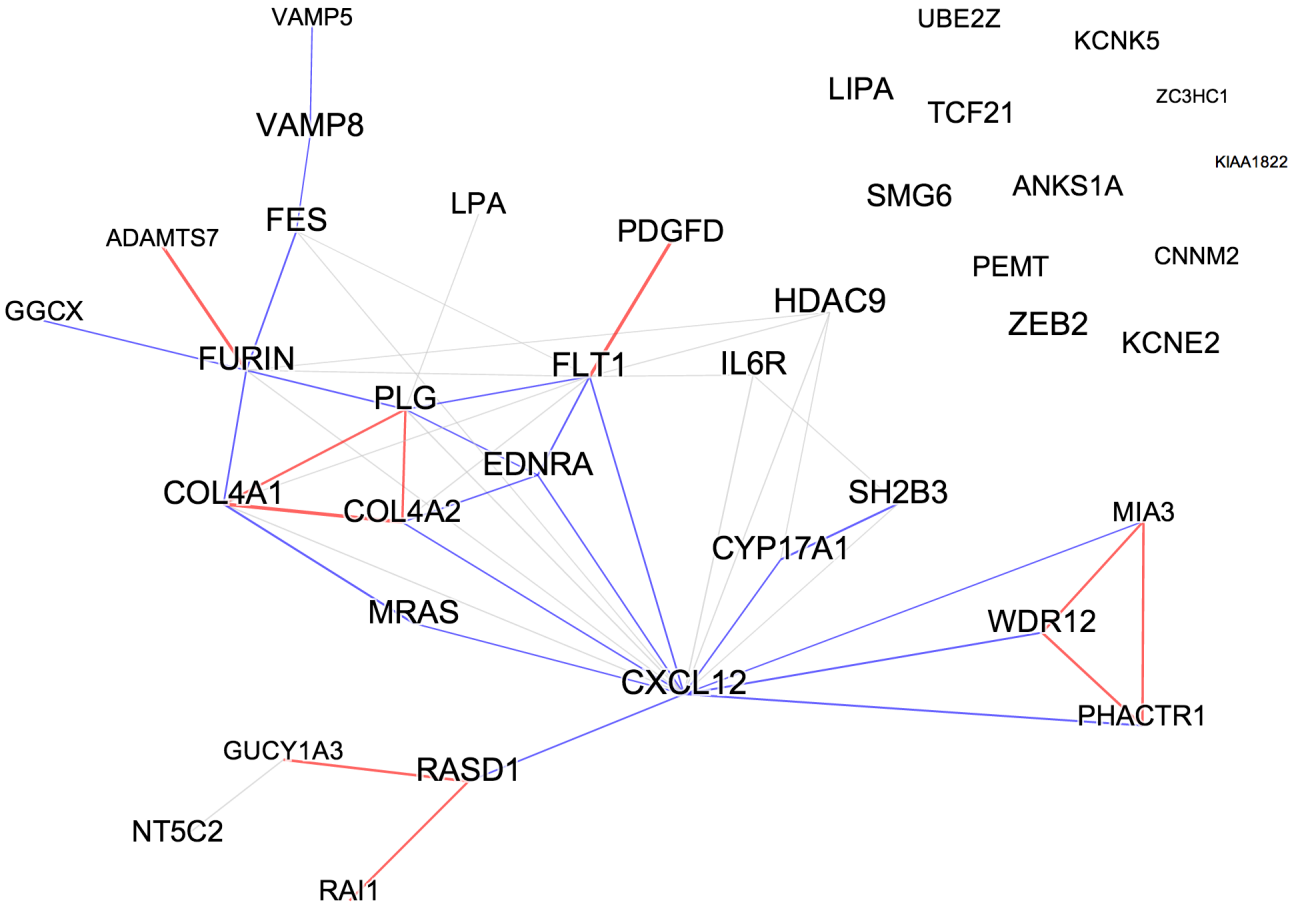
additional (white) nodes: 0

Update Parameters



A cautionary tale – text mining networks in coronary artery disease

STRING circularity and false positive associations



Visualization by Johnathan Mercer through GeNets
Collaboration with Ron Do, Sekar Kathiresan and colleagues

Overview of this session

1)	Biological networks
2)	Annotating genetic and proteomic data using biological networks
3)	Tissue-specific networks with disease resolution
4)	Emerging resources at the Broad Institute

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

Proteomic Analysis of Lysine Acetylation Sites in Rat Tissues Reveals Organ Specificity and Subcellular Patterns

Alicia Lundby,[1,2,3] Kasper Lage,[1,3,4,5] Brian T. Weinert,[1] Dorte B. Bekker-Jensen,[1] Anna Secher,[1,6] Tine Skovgaard,[1] Christian D. Kelstrup,[1] Anatoliy Dmytriyev,[1] Chunaram Choudhary,[1] Carsten Lundby,[7] and Jesper V. Olsen[1,*]

[1]Novo Nordisk Foundation Center for Protein Research, Department for Proteomics, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3b, 2200 Copenhagen N, Denmark

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

Map acetylation sites across 16 tissues

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

Map acetylation sites across 16 tissues

Make tissue-specific networks

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

Map acetylation sites across 16 tissues

Make tissue-specific networks

Map OMIM diseases onto network

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

Map acetylation sites across 16 tissues

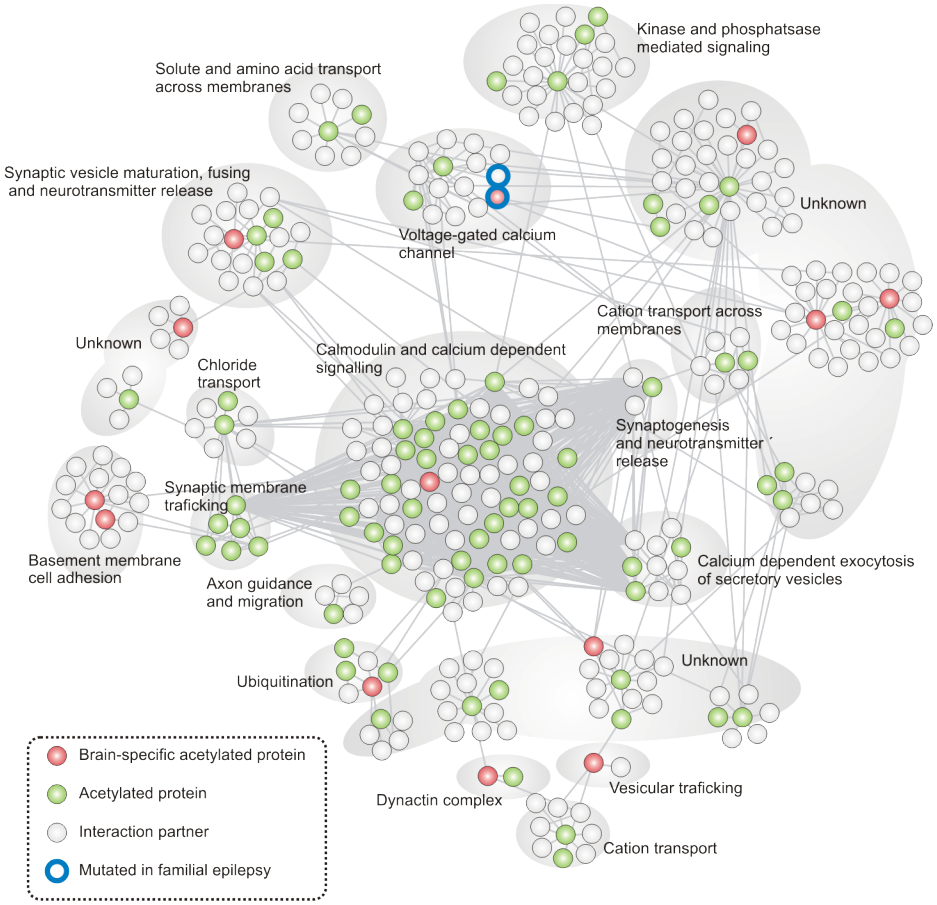
Make tissue-specific networks

Map OMIM diseases onto network

Network based on brain-specific acetylation

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications



Map acetylation sites across 16 tissues

Make tissue-specific networks

Map OMIM diseases onto network

Network based on brain-specific acetylation

Lundby et al., 2012 – Jesper Olsens Lab, CPR, University of Copenhagen

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

ARTICLE

Received 22 Nov 2011 | Accepted 25 Apr 2012 | Published 6 Jun 2012

DOI: 10.1038/ncomms1871

Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues

Alicia Lundby[1,2,*], Anna Secher[1,3,*], Kasper Lage[1,4,5,6], Nikolai B. Nordsborg[7], Anatoliy Dmytriyev[1], Carsten Lundby[8] & Jesper V. Olsen[1]

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

Map phosp. sites across 14 tissues

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

Map phosp. sites across 14 tissues

Make tissue-specific networks

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

Map phosp. sites across 14 tissues

Make tissue-specific networks

Map OMIM diseases onto network

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications

Map phosp. sites across 14 tissues

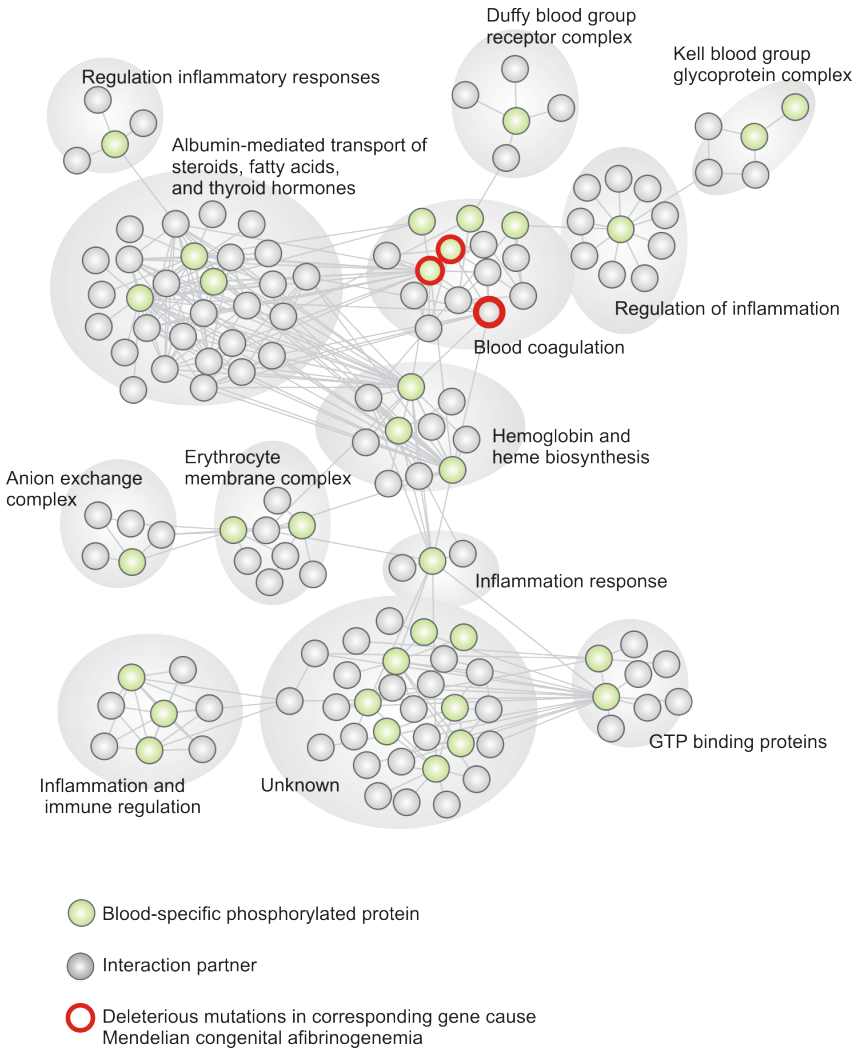
Make tissue-specific networks

Map OMIM diseases onto network

Network based on blood-specific phosphorylation

Tissue-specific networks with disease resolution

Interpreting tissue-specific posttranslational modifications



Map phosp. sites across 14 tissues

Make tissue-specific networks

Map OMIM diseases onto network

Network based on blood-specific phosphorylation

Lundby et al., 2012 – Jesper Olsens Lab, CPR, University of Copenhagen

Overview of this session

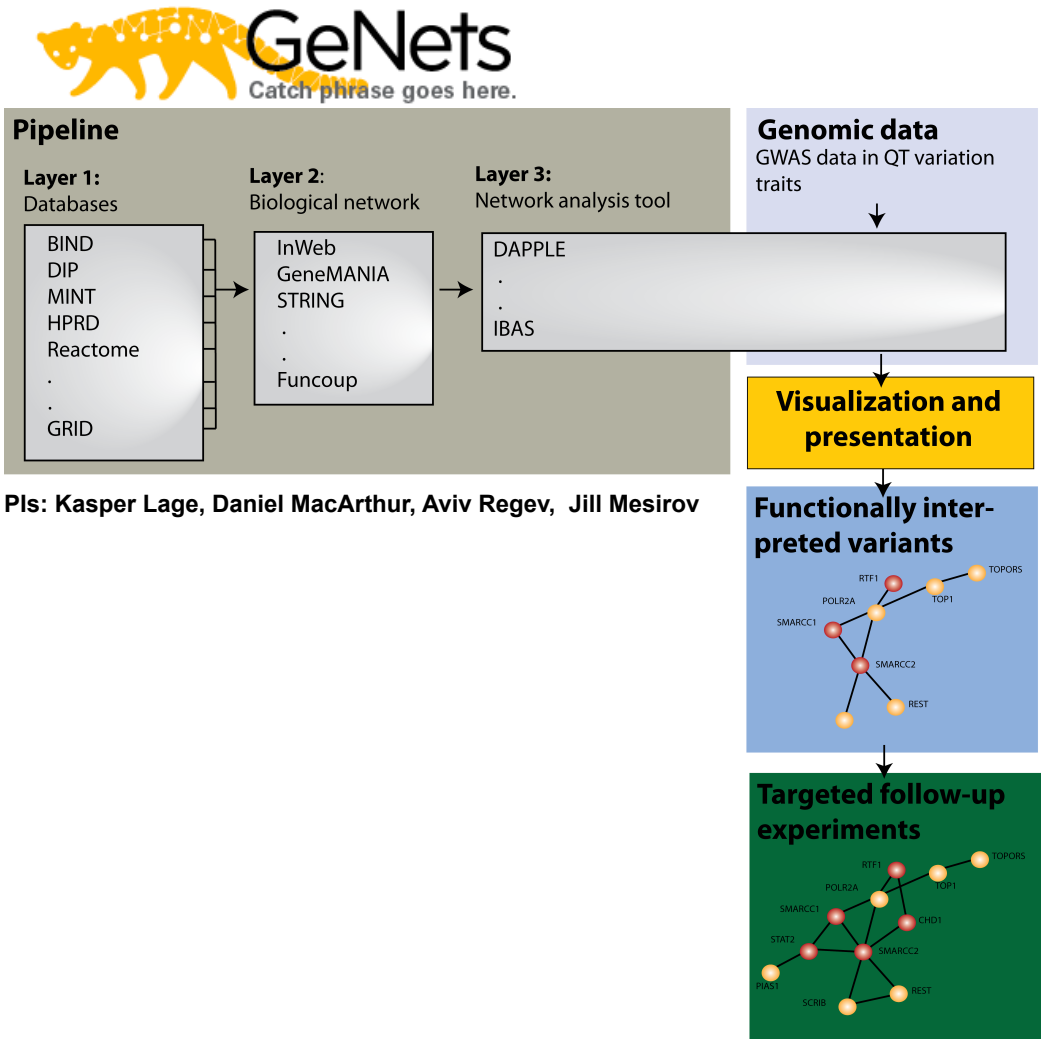
1)	Biological networks
2)	Annotating genetic and proteomic data using biological networks
3)	Tissue-specific networks with disease resolution
4)	Emerging resources at the Broad Institute

Broad Institute SPARC, next gen. sys bio platform – GeNets

iwww.broadinstitute.org/genets

Broad Institute SPARC, next gen. sys bio platform – GeNets

iwww.broadinstitute.org/genets



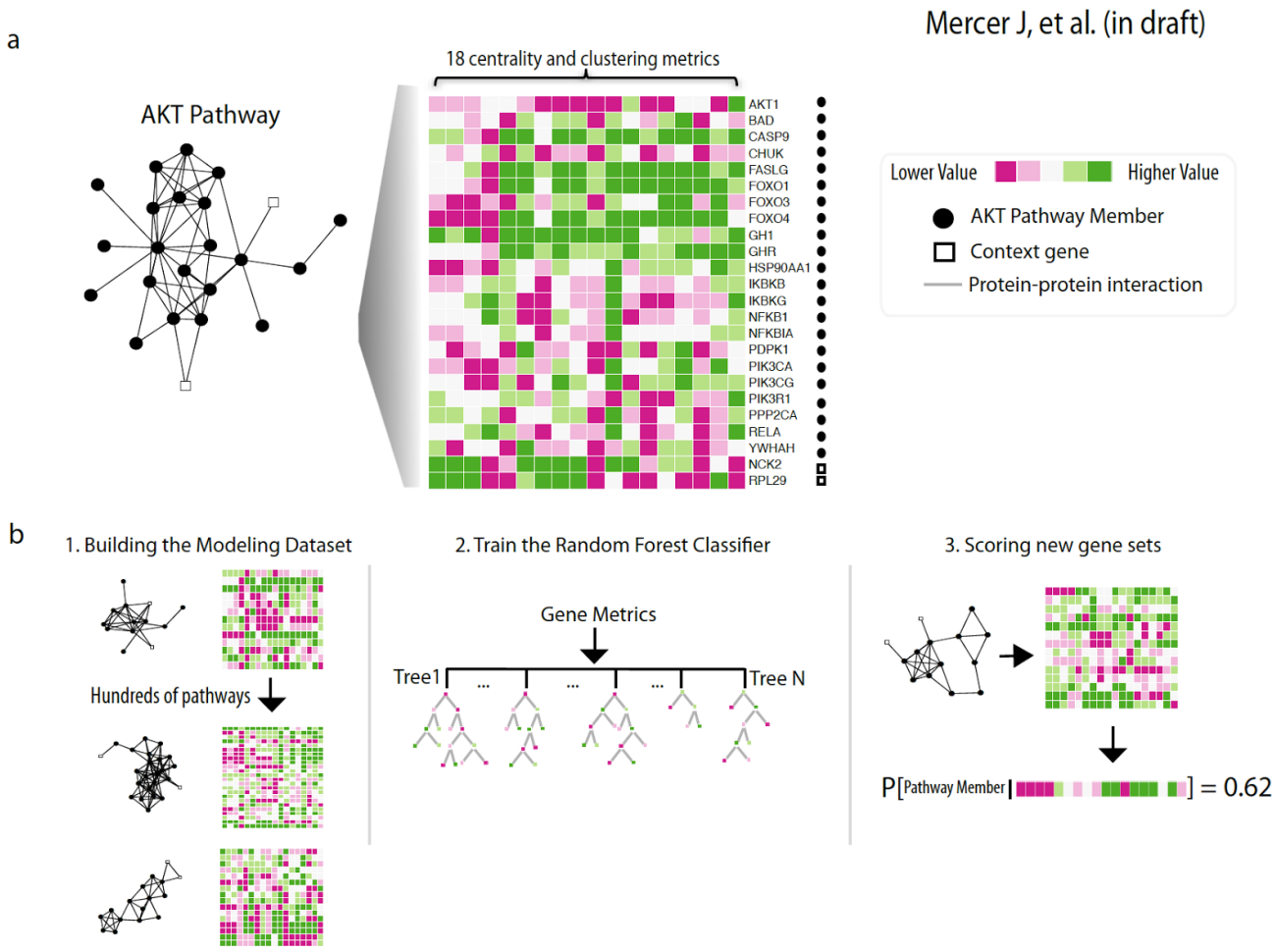
PIs: Kasper Lage, Daniel MacArthur, Aviv Regev, Jill Mesirov

1) **Only available internally at the Broad Institute for 13 months.**
2) **250 users**
3) **>3,600 Analyses (since October 2014).**
4) **Industry collaborations (Merck & Pfizer).**
5) **Best post. award (Broad & CBBO Retreat)**
6) **Patent application submitted.**
7) **Presentations across many programs and platforms.**

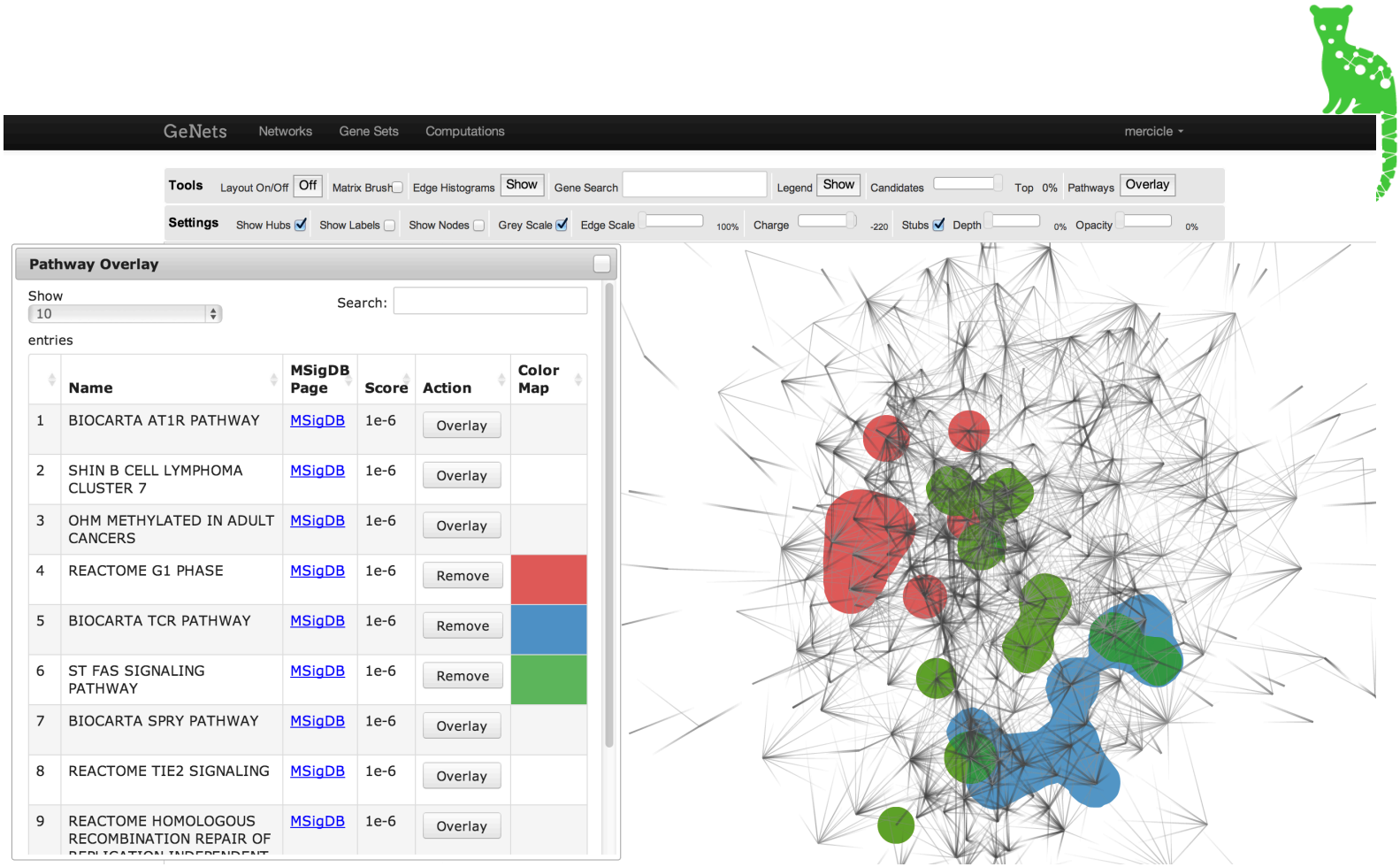
Jonathan Mercer

mercer@broadinstitute.org

Quack - A machine learning algorithm identifies nontrivial pathway signal in any functional network based on modeling thousands of known pathways

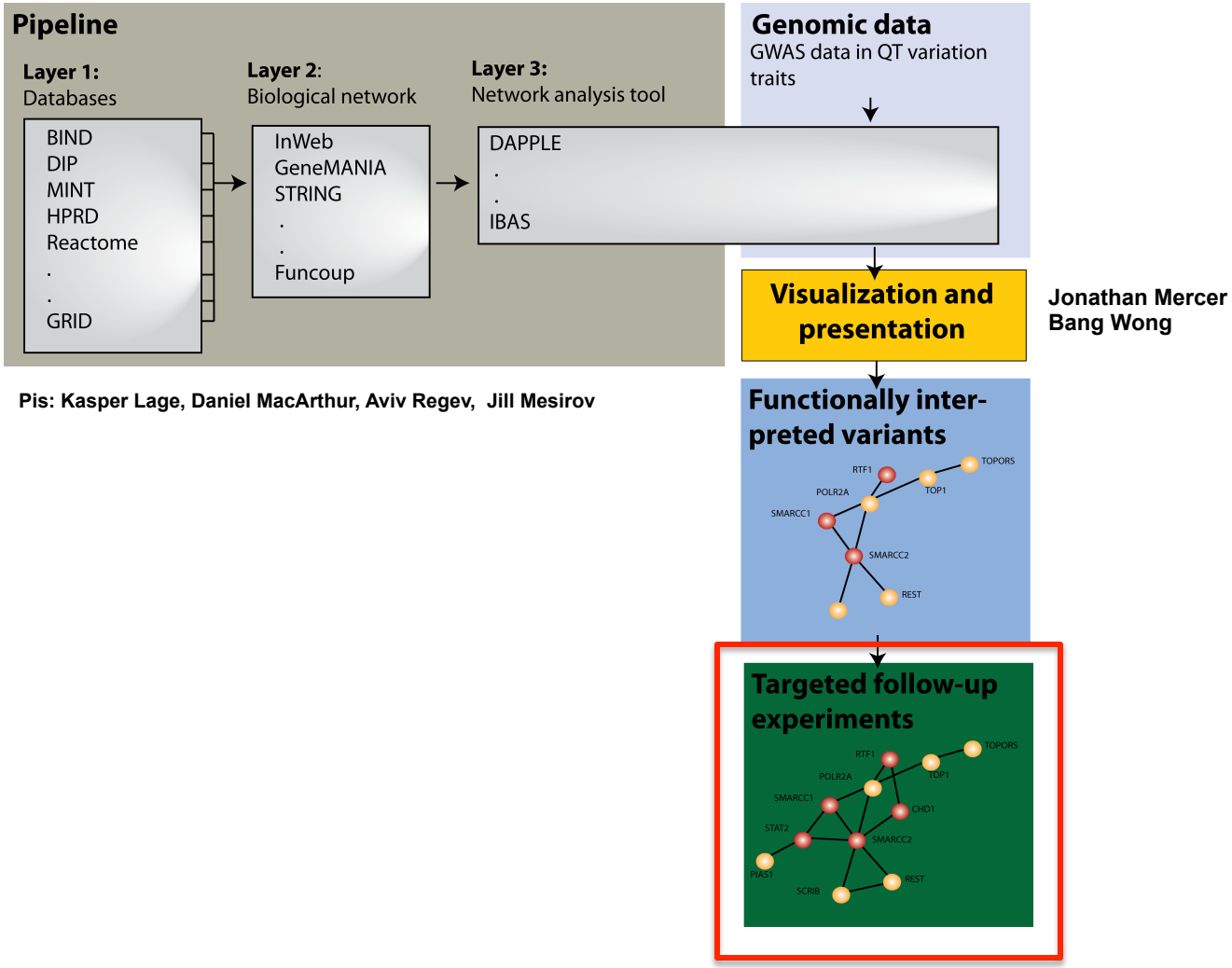


Jonathan Mercer


GeNets
Networks
Gene Sets
Computations
mercicle
Tools
Layout On/Off
Off
Matrix Brush
Edge Histograms
Show
Gene Search
Legend
Show
Candidates
Top 0%
Pathways
Overlay
Settings
Show Hubs
Show Labels
Show Nodes
Grey Scale
Edge Scale
100%
Charge
-220
Stubs
Depth
0%
Opacity
0%
Pathway Overlay
Show
10
entries
Search:
Name
MSigDB Page
Score
Action
Color Map
1 BIOCARTA AT1R PATHWAY MSigDB 1e-6 Overlay
2 SHIN B CELL LYMPHOMA CLUSTER 7 MSigDB 1e-6 Overlay
3 OHM METHYLATED IN ADULT CANCERS MSigDB 1e-6 Overlay
4 REACTOME G1 PHASE MSigDB 1e-6 Remove
5 BIOCARTA TCR PATHWAY MSigDB 1e-6 Remove
6 ST FAS SIGNALING PATHWAY MSigDB 1e-6 Remove
7 BIOCARTA SPRY PATHWAY MSigDB 1e-6 Overlay
8 REACTOME TIE2 SIGNALING MSigDB 1e-6 Overlay
9 REACTOME HOMOLOGOUS RECOMBINATION REPAIR OF MSigDB 1e-6 Overlay

John Mercer

What's next?



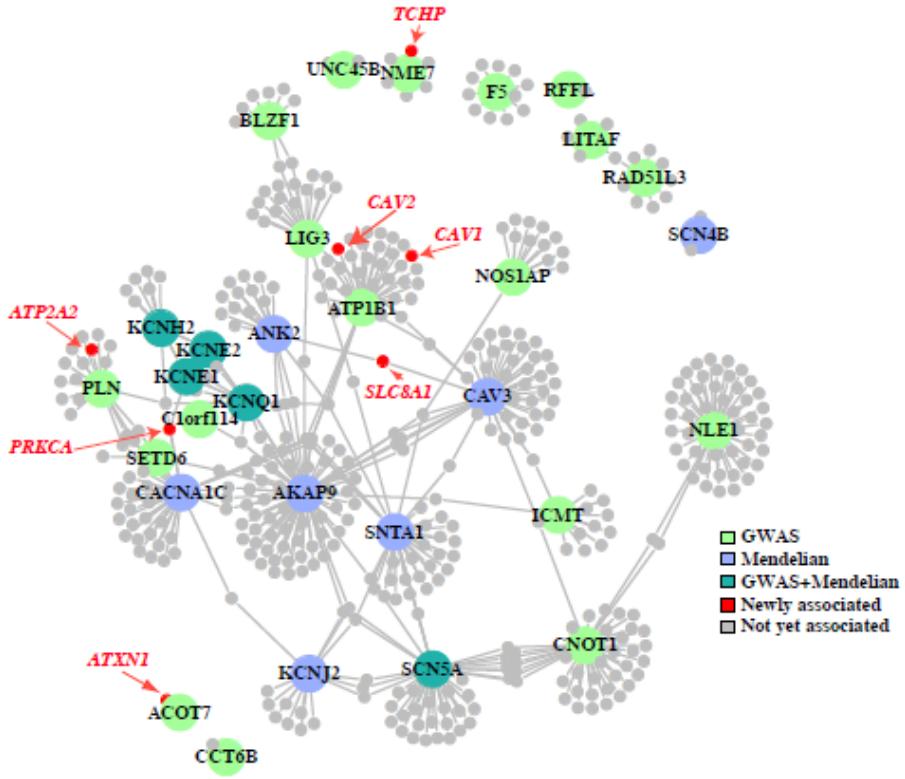
Pis: Kasper Lage, Daniel MacArthur, Aviv Regev, Jill Mesirov

Jonathan Mercer
Bang Wong

QT variation – a general approach to following up a GeNets network analysis

Validating *in silico* network models

Targeted and cost-efficient experiments motivated by network analyses



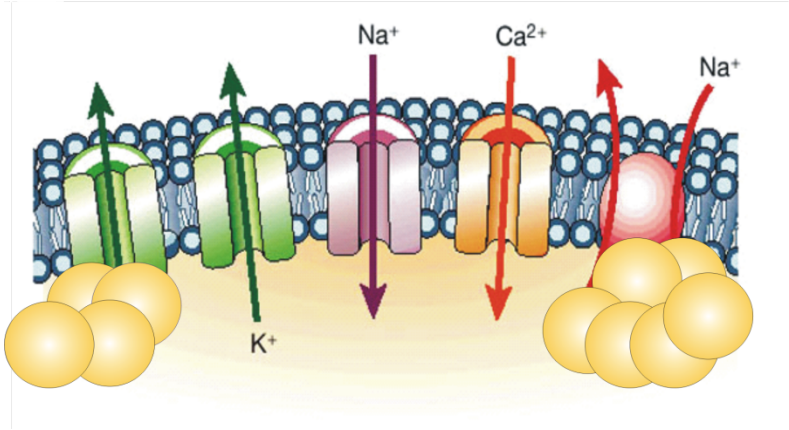
*Arking, *Pulit et al., 2014

Lundby*, Rossin* et al., 2014

Figure by Lizzy Rossin via DAPPLE

Cardiac ion channels form protein networks

Not mapped in detail in heart tissue (or any other tissue) until this work



Confirming, consolidating and expanding the QT network model

Functional, genetic and regulatory insight into heart signaling and QT variation

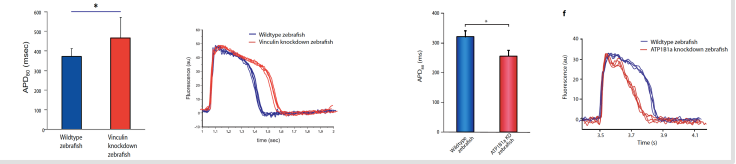
1)

2)

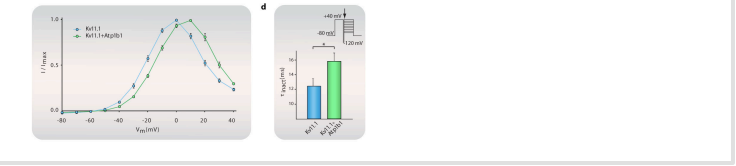
Table 1 | Genetic replication results

Locus information			Meta-analysis			Replication			Joint		
Gene	SNP	Coded allele	Effect size	s.e.	*P* value	Effect size	s.e.	*P* value	Effect size	s.e.	*P* value
Genome-wide significant loci in the joint analysis (joint $P < 5 \times 10^{-8}$)											
VCL	rs10824026	A	−0.71	0.13	5.20×10^{-8}	−0.72	0.27	4.23×10^{-3}	−0.71	0.12	1.49×10^{-9}
SRI	rs889807	T	−0.51	0.10	2.59×10^{-7}	−0.53	0.22	7.16×10^{-3}	−0.51	0.09	1.18×10^{-8}
TSPYL EIF3J/EIF3CL	rs7498491	A	−0.51	0.10	6.15×10^{-7}	−0.54	0.21	5.50×10^{-3}	−0.51	0.09	2.18×10^{-8}
Nominally significant loci in replication (replication $P < 0.05$)											
CAMK2D	rs17531033	C	0.39	0.11	3.75×10^{-4}	0.66	0.24	2.79×10^{-3}	0.44	0.10	1.11×10^{-5}
TWNK1	rs752139	T	0.46	0.10	1.31×10^{-5}	0.62	0.21	2.06×10^{-3}	0.44	0.09	1.49×10^{-6}
PKP2	rs7760812	A	−0.59	0.14	1.99×10^{-5}	−0.51	0.29	4.10×10^{-2}	−0.57	0.12	4.23×10^{-6}
CDH13	rs8043673	T	0.80	0.17	4.61×10^{-6}	0.75	0.45	4.92×10^{-2}	0.79	0.16	1.12×10^{-6}

3)



4)



1) Proteomics confirms int. partners in heart

2) Replication genot. confirms predicted SNPs

3) Knockdown confirms role in repolarization

4) Electrophysiology confirms channel reg.

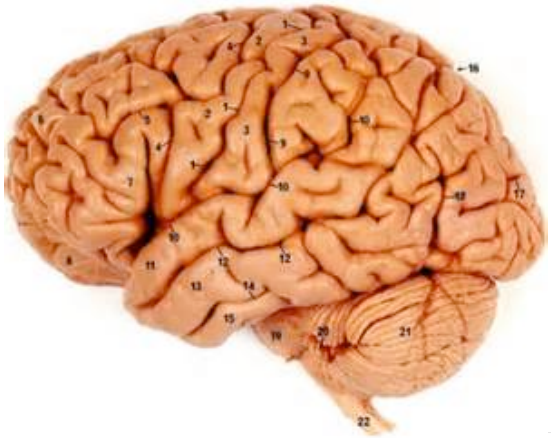
5) Exome sequencing suggests LQTS role

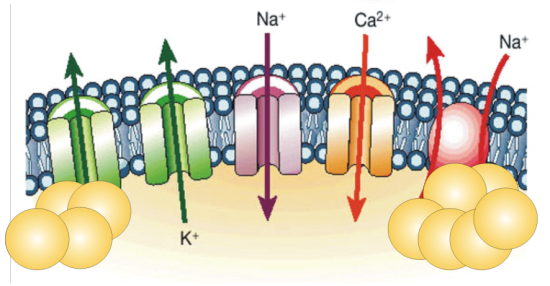
Lundby*, Rossin*, et al., 2014 Nature Methods

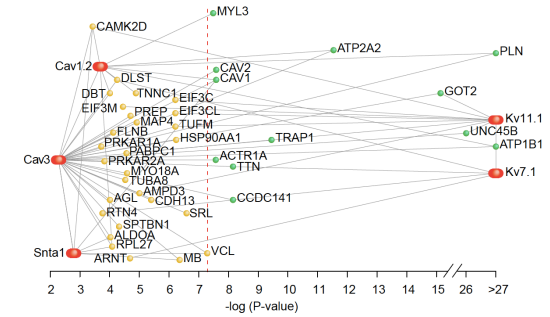
Arking*, Pulit*, et al., 2014 Nature Genetics

Repositioning this general framework for psychiatric disorders

Human brain networks perturbed by genetics in psych disorders







1) Human neurons from iPS

2) Unbiased quantitative network construction anchored in robust genetic findings

3) Rigorous follow-up experiments to validate networks emerging from the analysis

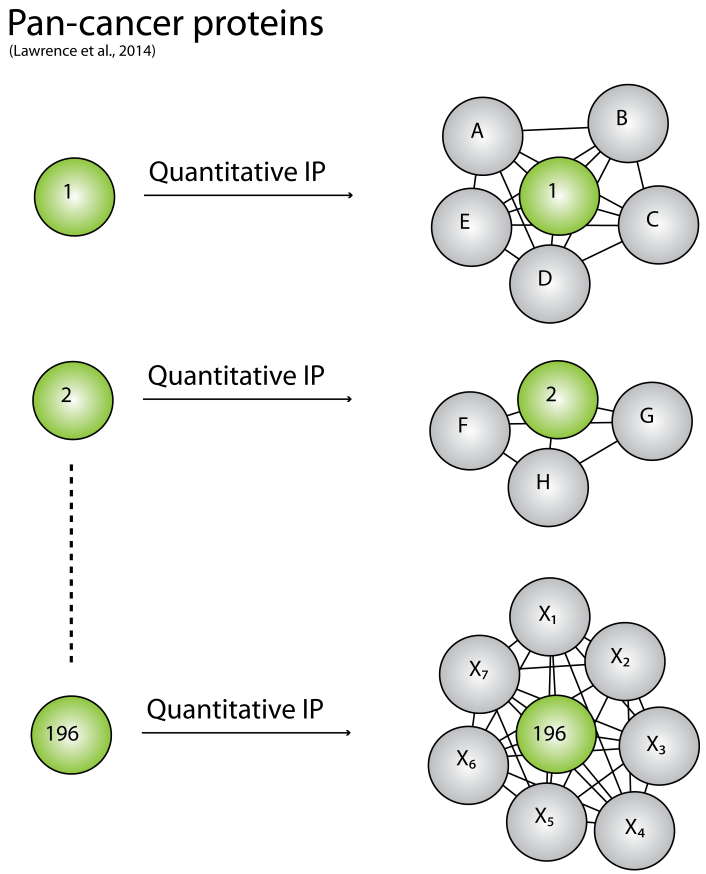
4) Therapeutic opportunities

5) Funded through a multi million grant from the Broad Institute's Stanley Center

Mark Daly, Steve Carr, Jake Jaffe, Monica Schenone, Kevin Eggan, Ben Neale, iPSYCH, Jennifer Moran, Steve Hyman

Vision: A Broad Institute Cancer Complex Compendium – CanCom^2

The first high-quality compendium of cancer-related protein networks / complexes generated through high-resolution proteomics



Steve Carr, Jake Jaffe, Monica Schenone, Ben Ebert, Matthew Meyerson, Jesse Boehm, Gad Getz, the Cancer Program

Resource List:

InWeb: A quality-controlled human protein-protein interaction network. www.cbs.dtu.dk/suppl/dgf or lage.kasper @mgh.harvard.edu

DAPPLE: A method to connect genes or proteins into statistically tested protein-protein interaction networks. www.broadinstitute.org/mpg/dapple/

DAVID: An integrated biological knowledgebase and analytical tools to extract biological meaning from gene or protein lists. http://david.abcc.ncifcrf.gov

Reactome: An open-source, open access, manually curated and peer-reviewed pathway database. http://www.reactome.org/

Ingenuity: A commercial tool for pathway analyses. http://www.ingenuity.com/

GSEA: A a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states. http://www.broadinstitute.org/gsea/

NetworKIN: A method for *predicting* in vivo kinase-substrate relationships, that augments consensus motifs with context for kinases and phosphoproteins. http://networkin.info/

STRING: STRING is a database of known and functional associations. The interactions include direct (physical) and indirect (functional) associations. http://string-db.org/

GeNets The Broad Institute Web Platform for Genome Networks. lwww.broadinstitute.org/genets

Acknowledgements

Lab @ MGH & Harvard

Heiko Horn **Jonathan Mercer**
Taibo Li **Jakob B. Jespersen**
Jessica Xin Hu Reza Kashani
Joy Paulo Bria Wallace

ATGU, MGH, Harvard, and Broad Institute

Lizzy Rossin **Mark Daly**
Chris Newton-Cheh **Paul de Bakker**
Daniel MacArthur **Ben Neale**

NNF Center for Protein Research, Uni. of CPH

Alicia Lundby
Jesper Olsen

Broad Institute and Stanley Center

Aviv Regev **Jill Mesirov**
Ed Scolnick **Jennifer Moran**
Steve Hyman **Jake Jaffe**
Bang Wong **Steve Carr**
David Altshuler **Jesse Boehm**
Todd Golub **Kevin Eggan**
Benjamin Ebert
Cmap team Achilles team

InWeb – CBS, Technical University Of Denmark

Raz Wernersson **E.Olof Karlberg**
Hans Henrik Stærfeldt **Olga Rigina**
Christopher Workman **Søren Brunak**

iPSYCH Consortium

Thomas Werge Preben Bo Mortensen
Anders Børglum Ole Mors
Merete Nordentoft

Funding

American Cancer Society
Lundbeck Foundation
MGH - Department of Surgery
Merck Research Laboratories
Broad Institute SPARC
NHLBI
NICHD
Shore Fellowship, Harvard Medical School
Jr. Faculty Dev. Award, Harvard Medical School