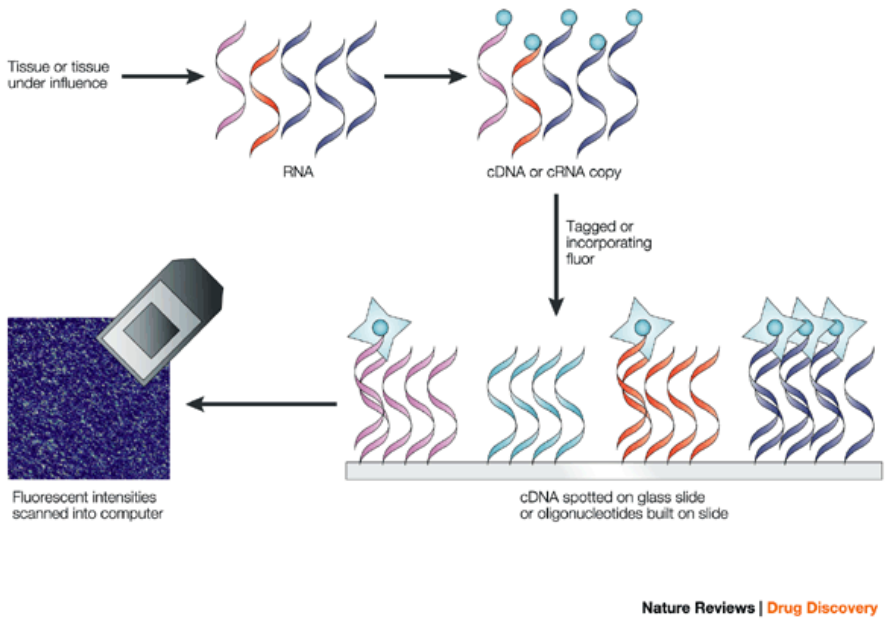
Gene expression databases

Sean Eddy, PhD

Outline

- How to measure gene expression?
  - Microarrays/RNA-seq
- What are gene expression databases?
- Which ones exist?
- How do they differ?
- How can they be used?
- What needs to be taken care of?
- What can I do with specialized databases?

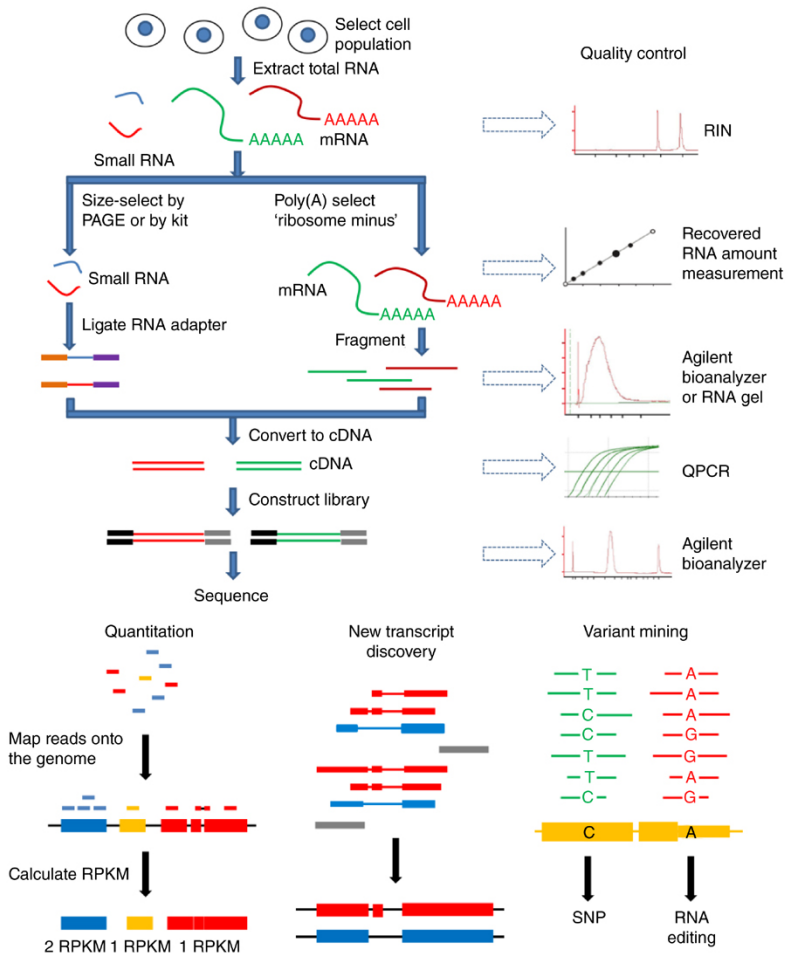
Microarrays: the beginning of high throughput gene expression



http://www.nature.com/nrd/journal/v1/n12/images/nrd961-f1.gif

- Compartmentalized chips with sequences bound to the surface
- Sample is applied to surface
- Hybridizes to complementary sequence
- Hybridizations are quantified
  - No binding, “no” signal
  - Can only find what is specifically searched for
  - Usually represented as n x m matrix

RNA-seq: the future of high throughput gene expression



Weihua Zeng & Ali Mortazavi
Nature Immunology **13**, 802–807 (2012) | doi:10.1038/ni.2407
Published online 21 August 2012

- Samples (cDNA libraries) are applied to a sequencer
- Millions of sequences are generated and mapped onto a genome
- Sequence reads are quantified
  - No sequence = no expression
  - Can find novel transcripts, splice isoforms, fusion genes, etc.
- Quality of mapping depends largely on library prep and decisions made on how RNA is initially processed.

Gene expression databases

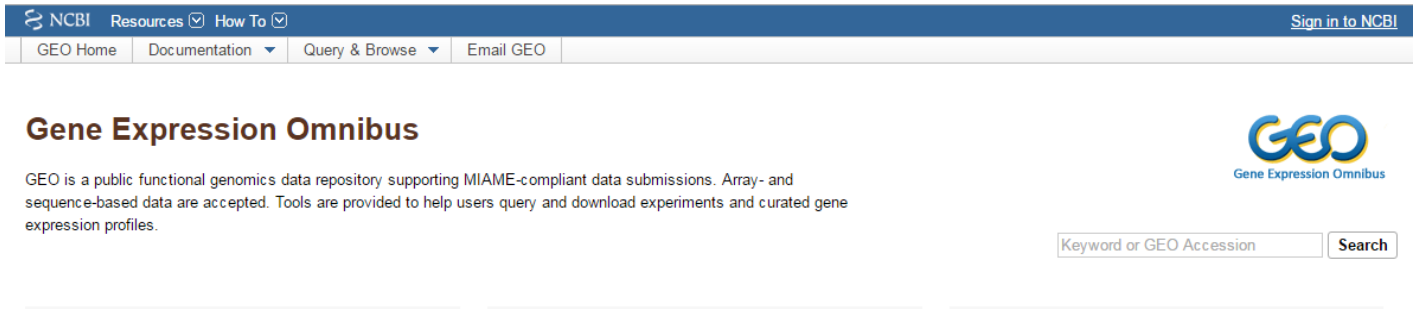
- Repositories for gene expression data
  - Mostly microarray and now RNAseq
  - Primarily for storage
  - Curated or un-curated
  - Access to data on different levels:
    - Datasets
    - Individual levels
- Integrated databases
  - Contain array data and additional data of the samples
  - Array data tends to be more annotated
  - More analytical tools
  - Smaller (more QC and curation needed)
  - Often no direct data access

Why do they exist

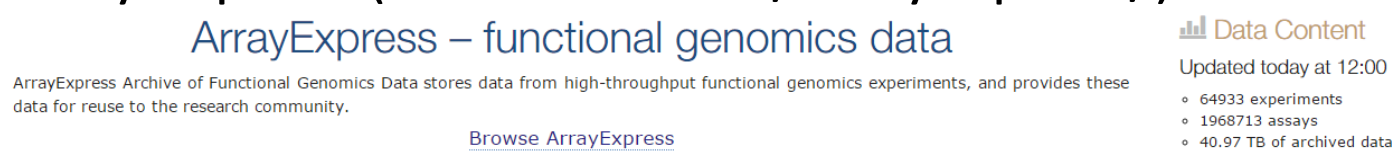
- Transparency/reproducibility of publications
  - Journals require data to be available for analysis
  - Nowadays raw data is required
  - Databases offer single resource and standardized access
- Data was generated for a specific purpose, but is not limited to that purpose
  - Can be reanalyzed in a different context
  - Can be combined with other datasets
  - Can be used as independent validation

Gene expression repository examples

- Gene expression omnibus (www.ncbi.nlm.nih/geo/)
  - 1,117,462 samples, 3848 datasets



- Array express (www.ebi.ac.uk/arrayexpress/)



- Princeton University MicroArray database (PUMAdb)
  - 40084 experiments, 6598 made public
- NCBI SRA, ENA and Princeton HTseq for NGS data

What is in a gene expression database?

- Gene expression data in different forms:
  - Resolution:
    - Gene level
    - Transcript level
    - Exon level
      - And / or raw data
  - Comprehensiveness
    - Targeted arrays
    - Whole genome arrays
  - Different platforms (microarrays, RNAseq)
- Generally only gene expression, may have limited sample information

Where does the data come from?

- Expression profiles of
  - Patients
  - Model systems
  - Cell cultures
- Data used for publication
  - Most journals now require raw data submission
  - Very coarse quality control (peer review)
  - QC depends mostly on authors
- Datasets submitted without publication
  - Little or no QC
- Most datasets are tailored towards a specific question

Example: GEO GSE32591

- Go to http://www.ncbi.nlm.nih.gov/geo/
- Enter GSE32591 into search box
- Click on "Analyze with GEO2R"
  - How would you set up the groups for analysis?
  - What do you get?
    - Does that make sense? How can results be verified?
- Go to "value distribution" tab
  - What do you see?
  - What are possible explanations?

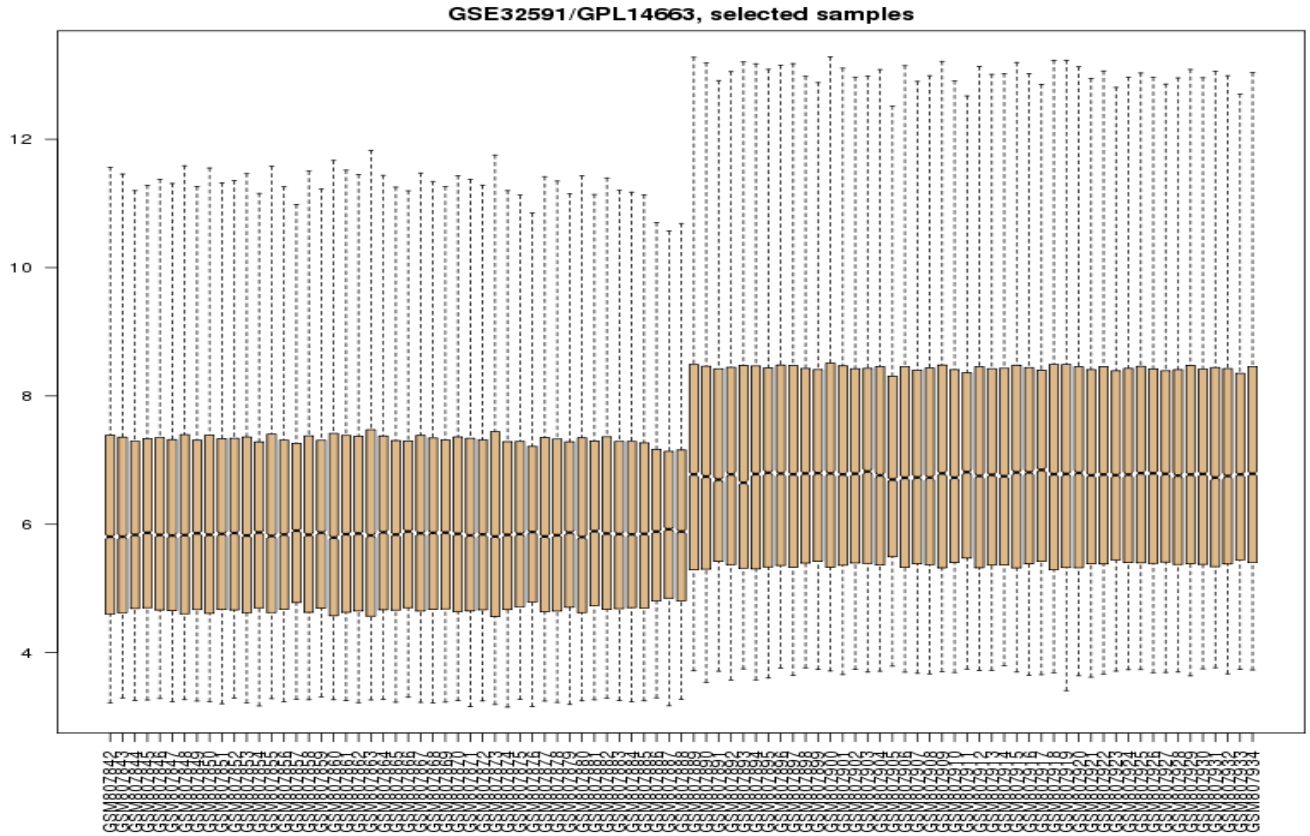
www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE32591

▸ Samples ▸ Define groups

GEO2R | Value distribution | Options | Profile graph | R script

Calculate the distribution of value data for the Samples you have selected. Distributions may be viewed graphically as a box plot or exported as a number summary table. The plot is useful for determining if value data are median-centered across Samples, and thus suitable for cross-comparison. More...

View Export

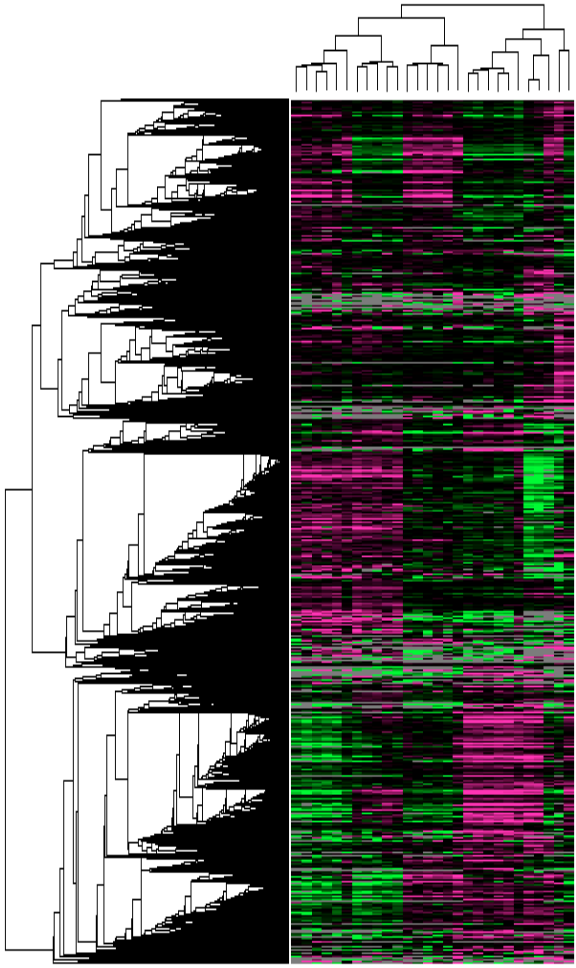


TMN:

What can be done with GEO?

What can be done with GEO?

- Programmatic access for data download
  - http://www.ncbi.nlm.nih.gov/geo/info/geo_paccess.html (GEO)
  - http://www.ebi.ac.uk/arrayexpress/help/programmatic_access.html (ArrayExpress)
- Pre-computed analyses and on the fly analyses
  - Search by gene across all GEO experiments
  - Search by experiment to retrieve cluster analysis
  - Search by gene sequence for matching expression profiles
    - Described by Barret and Edgar, Methods Mol. Biol. 2006 "Mining Microarray Data at NCBI's Gene Expression Omnibus (GEO)"
      - http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1619899/



What questions can be answered?

What questions can be answered?

- If you download: anything
  - Only limited by your knowledge, skills, resources
- Pre-computed results
  - Preselected analysis methods/ sample groups
  - Generally within one dataset
- On-the-fly analyses
  - Sets of genes that cluster in under conditions given
  - Sample properties may not be entirely transparent.

What can be answered by doing it yourself?

What can be answered by doing it yourself?

- The quality of the data
  - Is part of the data low quality?
  - Does some of the data not fit into the set (e.g. batch effect, outliers for other reasons)
  - Is it adequately processed?
- What is the relationship between expression data and non-expression variables?
  - How does my gene (of interest) associated with experimental treatments, clinical parameters?
- What are patterns across datasets?
  - Does my finding hold up across similar analyses in independent datasets?

Why do you have to do it yourself?

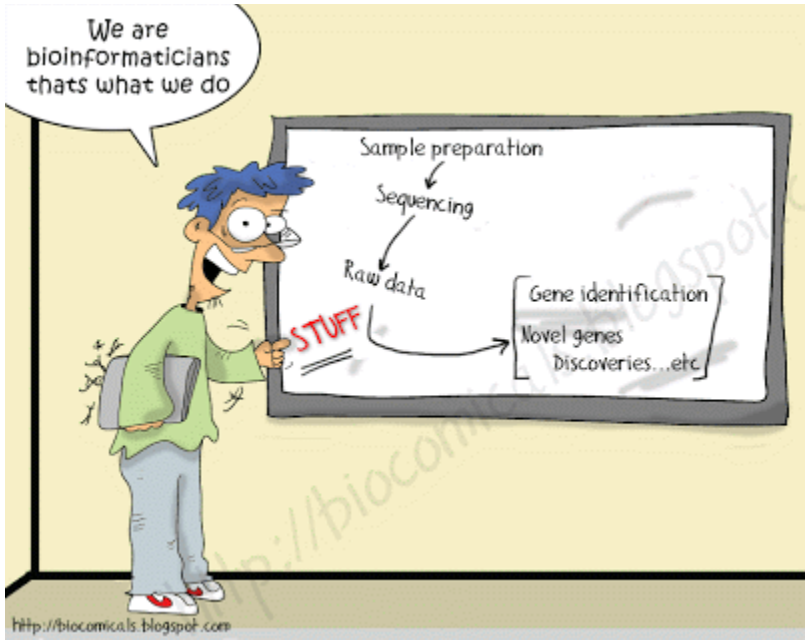
- Quality control:
  - QC parameters are often glossed over in papers and in micraorray submissions
  - For Affymetrix QC modules are available, freely available and widely accepted in the bioinformatic community
  - Other array types have distinct, but also similar properties
  - http://www.nature.com/nbt/focus/maqc/index.html
- Relations to non-expression data variables
  - Data is often not standardized within fields

Why not?

- Analysis across datasets:
  - Because:…. How?
  - Need to find a common standard for identification
  - Values need to be made comparable
    - If absolute expression values used, dynamic range can be a problem
    - Is ratios used, information about expression level lost
  - Non-expression data even worse

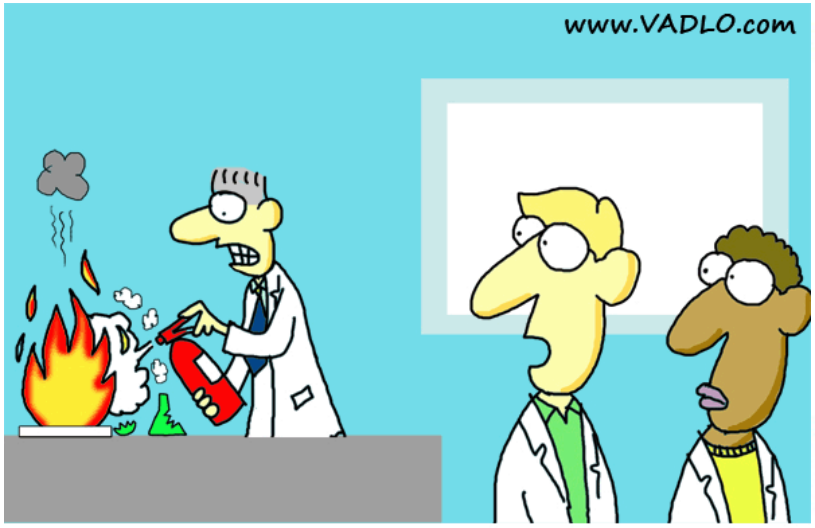
Who is the target group for doing it yourself?

- Users with experience in expression data
  - Crucial information (STUFF) is missing

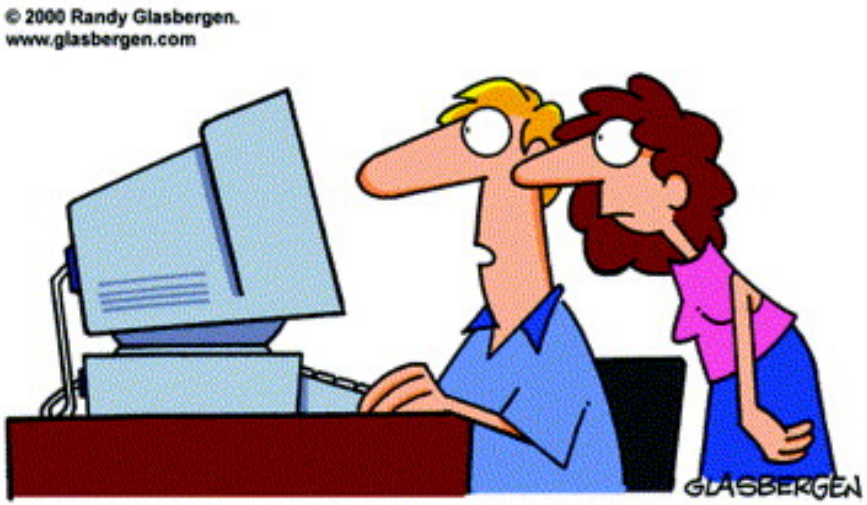


Why is this a problem?

- Excludes investigators with good hypotheses but lacking bioinformatic skills

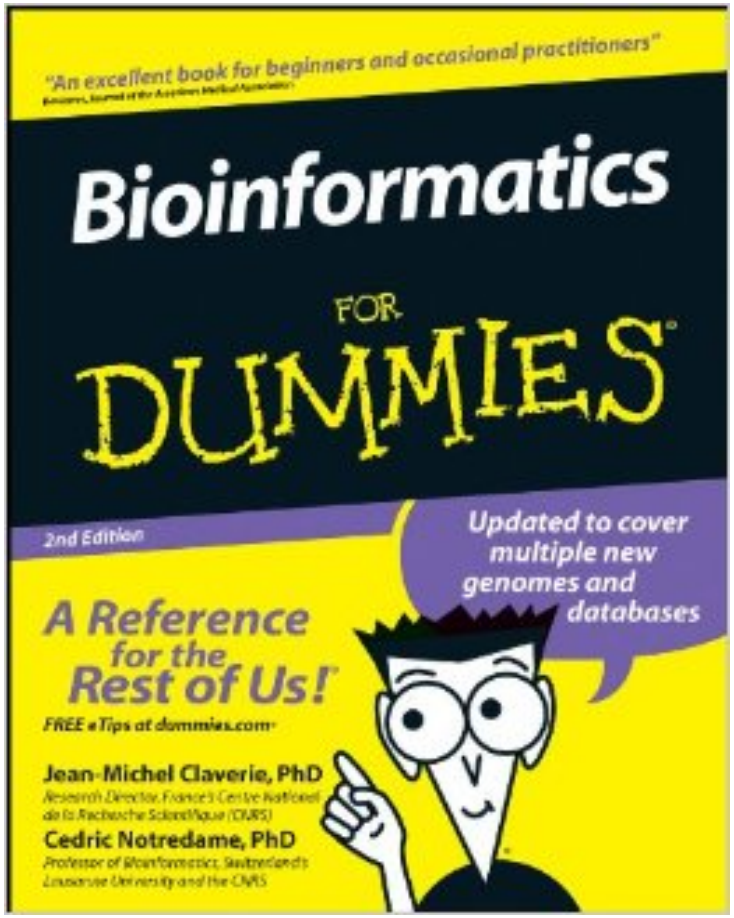


"Must be a clinical fellow."



"The computer says I need to upgrade my brain to be compatible with microarray data analysis."

How to fix that?



How to fix that?

- Specialized databases
  - Datasets are easier to find
    - Datasets relevant to specific areas are collected in one place
      - NephroSeq for renal disease
      - Oncomine for cancer
  - Datasets are standardized and expertly curated
    - Controlled vocabulary is introduced for non-expression data
    - Curation of expression possible by introducing standardized references and data transformations across datasets
      - Gene IDs/Gene Symbols as references
      - Z-transformation or median centering of log transformed expression data

Nephroseq (www.nephroseq.org)

Nephroseq

Contact

login

USER ID:

PASSWORD:

- Forgot password?
- Not a user? Register now!

Login

about

Nephroseq is supported by the Applied Systems Biology Core as part of the University of Michigan O'Brien Renal Center. The primary goal of the Applied Systems Biology Core is to provide to the renal research community a platform for integrative data mining of comprehensive renal disease gene expression data sets, in order to:

1. Define molecular characteristics/features in the circulation or kidney and associate them with known disease phenotypes so as to obtain a better understanding of the pathophysiology of a specific renal disease

2. Identify markers of disease progression and treatment response (i.e., biomarkers)

UNIVERSITY OF MICHIGAN

Welcome to Nephroseq

Developed for the renal research community, Nephroseq is a platform for integrative data mining of genotype/phenotype data, with optimized workflows that lead from search to visualization and from question to answer to next question:

- The expression of a gene is highly correlated with well-known podocyte genes. **Is the gene functionally important in glomeruli?**
- A gene is significantly differentially expressed in a subset of disease patients. **Is the gene associated with a certain phenotype, severity or sub-category of the disease?**
- A set of genes is significantly up-regulated in disease patients. **Are the disease genes inversely related to the target profile of a compound/drug?**

ABOUT NEPHROSEQ

Originally a collaborative effort, Nephroseq is now solely developed and maintained by the Applied Systems Biology Core at the University of Michigan. This resource combines a wealth of publicly available renal gene expression profiles - gathered and curated by an experienced team of data scientists, bioinformaticians, and nephrologists - with a sophisticated analysis engine and powerful web application designed for data mining and visualization of gene expression data.

Nephroseq provides researchers with a rich set of publicly available renal gene expression data, packaged with the tools and interface necessary to analyze it, all aimed at seeking answers to questions and advancing a molecular understanding of kidney disease to ultimately improve clinical outcomes.

In particular, Nephroseq provides unique access to datasets from the Personalized Molecular Nephrology Research Laboratory incorporating clinical data which is often difficult to collect from public sources.

Oncomine (www.oncomine.com www.oncomine.org)

ONCOMINE™

upgrade | contact

login

USER ID:

PASSWORD:

- Forgot password?
- Not a user? Register now!

login

news

Visit our newsroom to find out about the latest developments at Thermo Fisher Scientific.

Did you know that you can include the fee for Oncomine Research Premium Edition in US federal grant applications? Contact us for more information.

events

Whether it's a conference, trade show, or webinar, you will find us at these events.

Oncomine™ Research Edition: 715 datasets and 86,733 samples

With Oncomine™ Research Premium Edition, you can:

Design better experiments... Answer more questions with fewer experiments, select the most promising gene or cell line, and test your hypothesis.

Gain more biological insights... Discover novel targets for therapeutic development, interrogate gene expression profiles, and identify drug and biological interactions.

Prepare to publish faster... Validate your results faster, visualize data easier and make connections to clinical significance.

The Oncomine™ Platform—from web applications to translational bioinformatics services—provides solutions for individual researchers and multinational companies, with robust, peer-reviewed analysis methods and a powerful set of analysis functions that compute gene expression signatures, clusters and gene-set modules, automatically extracting biological insights from the data. It has become an industry-standard tool cited in more than 1,100 peer-reviewed journal articles. The Oncomine Platform has been used as a foundation for ground-breaking discoveries with unique features that include:

- **Scalability** – with 700+ independent datasets
- **High quality** – with expertly curated data
- **Consistency** – with a rich, extensive and controlled ontology of terms
- **Standardized analysis** – with conventions that assure clear and consistent interpretations of results

Oncomine Research Edition remains free to the academic and nonprofit cancer research communities.

iontorrent
by Thermo Fisher Scientific

The Origin of the Oncomine Platform
The Oncomine Platform was conceived by physicians, scientists, and software engineers at the University of Michigan. It was commercialized by Arul Chinnaiyan and Dan Rhodes in February 2006 with the goal of building a version that would have a greater ability to impact drug development and clinical practice.

NephroSeq and Oncomine

- Pros:
  - Each focus on one area of interest
  - Clinical data for many individual samples available
  - Advanced analysis using integrated systems biology tools in a pre-defined automated manner
  - Meta analysis possible
  - User friendly, free accessible for academic users
  - **Hypotheses-generating**
- Cons:
  - No raw data download
  - No programmatic access
  - Only predefined analyzes

NephroSeq main Page

? HELP

https://nephroseq.org/resource/main.html

Nephroseq

Welcome, Sean Eddy. Tools ? Help Logout

search

filter

selected 26 datasets (2000 samples)

26 datasets (2000 samples)

Primary Filters
- Analysis Type
  - Coexpression Analysis (26)
  - Differential Analysis (26)
    - Demographics Analysis (23)
    - Donor Type Analysis (1)
    - Group Analysis (18)
    - Indices Analysis (3)
    - Pathology Analysis (18)
    - Risk Factor Analysis (1)
    - Tissue Type Analysis (9)
    - Treatment Analysis (1)
  - Outlier Analysis (26)
- Group
- Donor Type
- Tissue Type
- Dataset Type

Sample Filters

Dataset Filters

Concept Filters

Analysis type

Coexpression analysis
Differential analysis
Outlier analysis

Nephroseq Overview

OTHER VIEWS:

Nakagawa CKD: Gene expression data on a total of 53 CKD patients' renal biopsies (48 in discovery cohort and 5 in validation cohort) and 8 controls biopsies (5 in discovery cohort and 3 in validation cohort) were analyzed to identify genes that are differently expressed between CKD patients and controls. Published 2015/08.

Cox IgA Nephropathy: Gene expression profiling of CD14-purified monocytes from 8 IgA nephropathy samples and 9 healthy control samples was performed to identify transcripts differentially expressed in IgA nephropathy patients. Published 2015/07.

Welcome to Nephroseq!

This application is a web-based analysis engine for molecular biology researchers and clinician scientists who study renal disease and related disorders. Nephroseq gives access to renal genome-wide gene expression datasets generated by the renal research community. This tool is especially powerful because the data are already pre-analyzed and datasets include clinical data. The 3-paned user interface moves users from left to right within the application to choose data, sort and prioritize analyses, and visualize and export results.

Analyses that are available include:

- Differential Expression: Identify over- or under-expression for a particular gene
- Coexpression: View genes that are coordinately expressed with your gene of interest across a dataset
- Outlier: Identify outlier patterns where a gene is highly over-expressed in a fraction of samples
- Concept Association: Identify significant overlap between gene sets that represent underlying biology

In addition, users have the ability to upload gene lists to use as filters and export data and visualizations directly to Excel, PowerPoint and SVG.

support

ntific and password/account management
port is available from: support@nephroseq.org

ws

press releases

JANUARY 2016

Identification of urinary protein biomarker for Chronic Kidney Disease
http://www.ncbi.nlm.nih.gov/pubmed/26631632

From the June, 2015, ADA meeting: phase II clinical trial successful in Diabetic Kidney Disease
http://www.abstractsonline.com/pp8/#!/3699/presentation/12757

events

mark your calendar

11th International Podocyte Conference
APRIL 3rd - 6th
HAIFA AND JERUSALEM, ISRAEL
http://podocyte2016.org

ISN Nexus Symposium: Translational Immunology in Kidney Disease
APRIL 14th - 17th
BERLIN, GERMANY
http://www.isnnexus.org/berlin

Two Search Options

- Gene specific search:
  - Gene
- Dataset search:
  - Specific conditions/diseases

Gene Search

NPHS2: encodes podocin, a podocyte specific protein

Gene summary view

Nephroseq

Welcome, Sean Eddy. Tools Help Logout

search

NPHS2

filter

selected 11 datasets (795 samples)

Gene:

Primary Filters

- Analysis Type
  - Coexpression Analysis (11)
  - Differential Analysis (11)
    - + Demographics Analysis (10)
    - Group Analysis (6)
    - + Indices Analysis (1)
    - + Pathology Analysis (8)
    - Tissue Type Analysis (5)
    - + Treatment Analysis (1)
  - Outlier Analysis (11)
- + Group
- + Donor Type
- + Tissue Type
- + Dataset Type

Sample Filters

Dataset Filters

Concept Filters

visualize

Gene Summary

OTHER VIEWS:

GENE: NPHS2 THRESHOLD (P-VALUE): 1E-4 THRESHOLD (FOLD CHANGE): 2 THRESHOLD (GENE RANK): Top 10% DATA TYPE: All

EXPORT:

Disease Summary

Demographics

Dataset/Disease type

	Age	Body Mass Index	Race/ Ethnicity	Sex	Donor Type	Group	Indices: Activity Index Quartile	Indices: Chronicity Index Quartile	Pathology: BUN	Pathology: GFR (MDRD)	Pathology: Glomerulo-sclerosis	Pathology: Hemoglobin	Pathology: Proteinuria	Pathology: TAIF	Pathology: WHO Lupus Nephritis Class	Tissue Type	Treatment: Post-Event Treatment	Outlier
Aging																1		1
Diabetes																1		1
FSGS																		
Hypertension																		
IgAN						1												1
Lupus																		
Normal Tissue Panel																1		
Transplant																		2
Significant Unique Analyses						1										3		1 4
Total Unique Analyses	21	3	30	21	0	16	0	2	1	15	0	0	7	2	0	36	1	22

1 5 10 % 10 5 1

Cell color is determined by the best gene rank percentile for the analyses within the cell.

NOTE: An analysis may be counted in more than one cancer type.

Gene Search

Gene summary view

NEPHROMINE

Welcome, Wenjun Ju. tools help logout

search

NPHS2

filter

selected 2 datasets (365 samples)

- Gene: NPHS2
- Analysis Type: Tissue Type Analysis
- Dataset Type: Normal Tissue Panel

Primary Filters

- Analysis Type
  - Coexpression Analysis (2)
  - Differential Analysis (2)
    - Demographics Analysis (1)
    - Tissue Type Analysis (2)
  - Outlier Analysis (2)
- Group
- Tissue Type
- Dataset Type

Sample Filters

Dataset Filters

Concept Filters

visualize

Gene Summary

OTHER VIEWS:

GENE: NPHS2 THRESHOLD (P-VALUE): 0.05 THRESHOLD (FOLD CHANGE): 1.5 THRESHOLD (GENE RANK): Top 10% DATA TYPE: All

EXPORT:

Disease Summary for NPHS2

Analysis Type by Dataset Type	Demographics				Donor Type	Group	Indices		Pathology							Tissue Type	Treatment	Outlier	
	Age	Body Mass Index	Race/ Ethnicity	Sex			Activity Index Quartile	Chronicity Index Quartile	BUN	GFR (MDRD)	Glomerulo-sclerosis	Hemoglobin	Proteinuria	TAIF	WHO Lupus Nephritis Class		Post-Event Treatment		
Aging																1			1
Diabetes			2 1			3				1						1		1	
FSGS						3													
Hypertension																			
IgAN						1													1
Lupus																			
Normal Tissue Panel																22			
Transplant																			2
Significant Unique Analyses			2 1			7				1						24		1 4	
Total Unique Analyses	21	3	30	21	0	16	0	2	1	15	0					5	1	22	

22 out of 33 analyses meet your threshold for NPHS2
In 2 out of 2 datasets
Dataset Type: Normal Tissue Panel
Analysis Type: Tissue Type

1 5 10 10 5 1
%

Cell color is determined by the best gene rank percentile for the analyses within the cell.
NOTE: An analysis may be counted in more than one cancer type.

22 out of 33 analysis meet your threshold for NPHS2 in 2 out of 2 datasets

Four Basic Analysis Modes

- Differential expression
- Co-expression analysis
- Outlier analysis
  - Heterogeneity within predefined groups
- Concepts analysis
  - Gene set (Nephromine & third-party sources)

Gene Search

Differential expression (Box graph)

NEPHROMINE

Welcome, Wenjun Ju. tools help logout

search

NPHS2

filter

selected 2 datasets (365 samples)

- Gene: NPHS2
- Analysis Type: Tissue Type Analysis
- Dataset Type: Normal Tissue Panel

Primary Filters

- Analysis Type
  - Coexpression Analysis (2)
  - Differential Analysis (2)
    - Demographics Analysis (1)
    - Tissue Type Analysis (2)
  - Outlier Analysis (2)
- Group
- Tissue Type
- Dataset Type

Sample Filters

Dataset Filters

Concept Filters

datasets | concepts

ORDER BY: Under-expression: Gene Rank

ON: NPHS2

THRESHOLD BY:

P-VALUE	FOLD CHANGE	GENE RANK
0.05	1.5	Top 10%

Compare | Clear All

Lindenmeyer Normal Tissue Panel (12)

Tissue Type: Glomeruli vs. Tubulointerstitium
p = 5.68E-8 fold change = -10.623 694

Roth Normal Tissue Panel (353)

Tissue Type: Urethra vs. Kidney
p = 1.03E-4 fold change = -12.091 865

Tissue Type: Bronchus vs. Kidney
p = 1.57E-4 fold change = -11.538 881

Tissue Type: Lymph Node vs. Kidney
p = 1.84E-4 fold change = -10.851 937

Tissue Type: Papilla of the Tongue vs. Kidney
p = 3.23E-4 fold change = -8.761 954

Tissue Type: Ovary vs. Kidney
p = 2.27E-4 fold change = -9.745 964

Tissue Type: Prostate Gland vs. Kidney
p = 3.38E-4 fold change = -9.433 965

Tissue Type: Thyroid Gland vs. Kidney
p = 2.46E-4 fold change = -9.546 970

Tissue Type: Trachea vs. Kidney
p = 2.66E-4 fold change = -9.255 996

Tissue Type: Breast vs. Kidney
p = 1.70E-4 fold change = -10.594 1151

Tissue Type: Tonsil vs. Kidney
p = 2.23E-4 fold change = -9.932 1172

Tissue Type: Adrenal Cortex vs. Kidney
p = 2.87E-4 fold change = -9.031 1174

Tissue Type: Lung vs. Kidney
p = 2.05E-4 fold change = -10.044 1281

visualize

Differential Analysis

OTHER VIEWS:

GROUP BY: Tissue Type

SHOW: Only Samples in Analysis

+ PRIMARY CONCEPT: EXPORT:

NPHS2 Expression in Lindenmeyer Normal Tissue Panel

Tissue Type: Glomeruli vs. Tubulointerstitium

Lindenmeyer Normal Tissue Panel Statistics

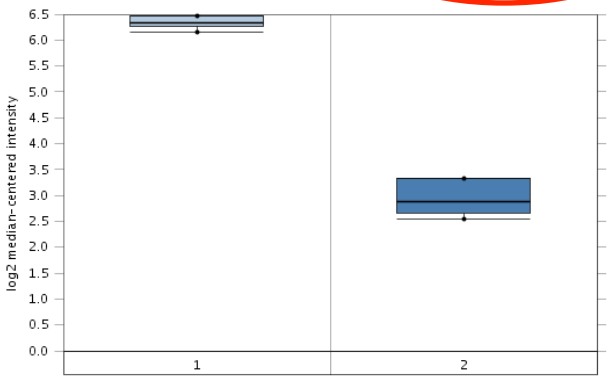
Under-expression Gene Rank: 694 (in top 6%)

Reporter: 220424_at

P-value: 5.68E-8

t-Test: -26.283

Fold Change: -10.623



Legend

1. Glomeruli (6)
2. Tubulointerstitium (6)

Lindenmeyer Normal Tissue Panel

PLoS One 2010/07/12 | 12 samples | NPHS2 Information

mRNA | 12,624 measured genes | Reporter Information

Human Genome U133A Array

Reference

THE HUMAN PROTEIN ATLAS

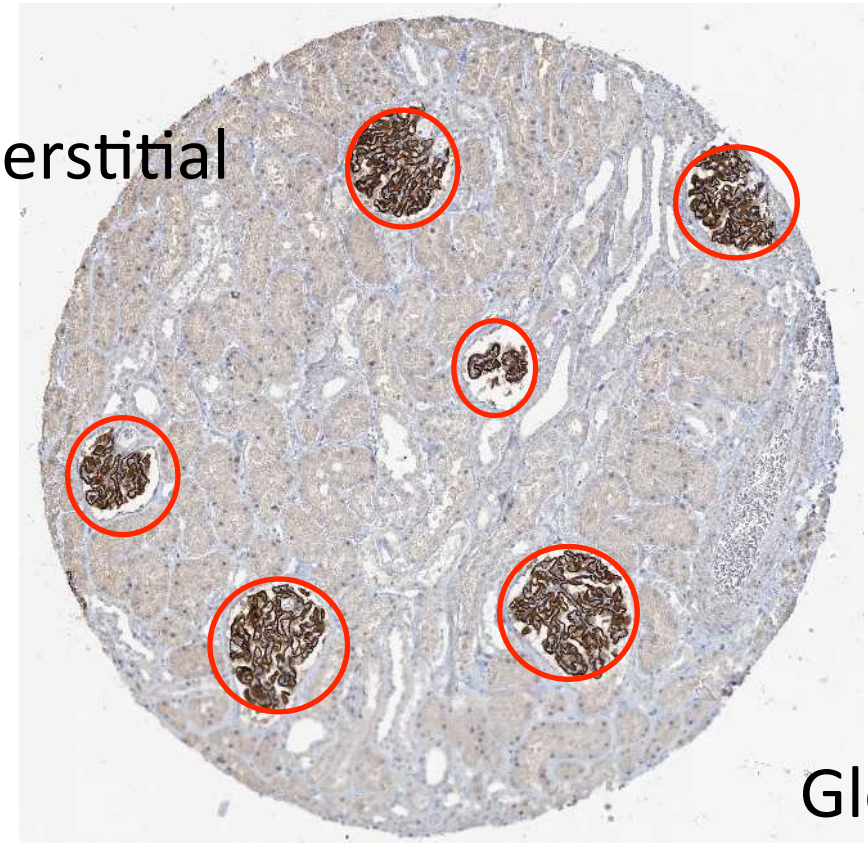
ABOUT & HELP

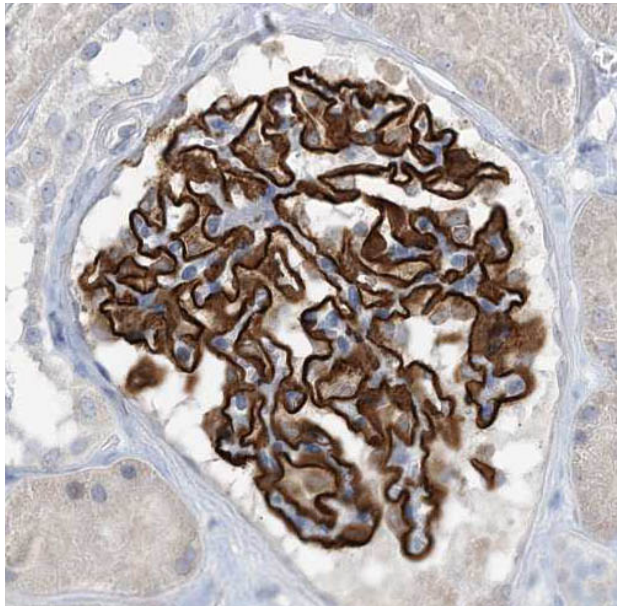
SEARCH ? »

NPHS2 | Search | Clear | Fields »

e.g. CD44, ELF3, KLK3, or use Fields to search specific fields such as protein_class:Transcription factors or chromosome:X

Tubulointerstitial





Glomeruli

Gene Search

Correlation with clinical continuous variation

Gene: VCAN
Analysis type: GFR
Dataset Type: Diabetes

search
VCAN
filter
selected 2 datasets (66 samples)
Gene: VCAN
Analysis Type: GFR (MDRD) Analysis
Dataset Type: Diabetes
Primary Filters
Analysis Type
Coexpression Analysis (2)
Differential Analysis (2)
Demographics Analysis (2)
Age Analysis (2)
Race/Ethnicity Analysis (1)
Sex Analysis (2)
Group Analysis (2)
Pathology Analysis (2)
GFR (MDRD) Analysis (2)
Proteinuria Analysis (1)
Tissue Type Analysis (1)
Outlier Analysis (2)
Group
Tissue Type
Dataset Type
Sample Filters
Dataset Filters
Dataset Name
Schmid Diabetes (1)
Woroniecka Diabetes (1)
Concept Filters

GENE RANK
Top 10%
Compare | Clear All
Woroniecka Diabetes (44)
Tubulointerstitium: GFR (MDRD) p = 7.71E-7 12
Glomeruli: GFR (MDRD) p = 0.006 447
Diabetic Nephropathy Tubulointerstitium: GFR (MDRD) p = 0.015 617
Schmid Diabetes (22)
Diabetic Nephropathy: GFR (MDRD) p = 0.069 1060

visualize
Differential Analysis
GROUP BY: GFR (MDRD) (Tubulointerstitium)
SHOW: Only Samples in Analysis

VCAN Expression in Woroniecka Diabetes
Tubulointerstitium: GFR (MDRD)

Woroniecka Diabetes Statistics
Under-expression Gene Rank: 12 (in top 1%)
Reporter: 221731_x_at
P-value: 7.71E-7
Correlation: -0.832

P=7.71E-7
Correlation: -0.832

log2 median-centered intensity
6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
1 2 3 4 5

Legend
1. < 15 ml/min/1.73m2 (3)
2. 15 - 29 ml/min/1.73m2 (4)
3. 30 - 59 ml/min/1.73m2 (7)
4. 60 - 89 ml/min/1.73m2 (5)
5. > 90 ml/min/1.73m2 (3)

Legend
1. Less Than 15 ml/min/1.73m2 (3)
2. 15 - 29 ml/min/1.73m2 (4)
3. 30 - 59 ml/min/1.73m2 (7)
4. 60 - 89 ml/min/1.73m2 (5)
5. More Than 90 ml/min/1.73m2 (3)

Woroniecka Diabetes
Diabetes 2011/09/01
mRNA
Human Genome U133A 2.0 Array
44 samples
12,603 measured genes
VCAN Information
Reporter Information

Gene Search

Outlier analysis

Outlier analysis helps to identify an expression profile where differential pattern is only seen in **a fraction of samples** of all patients within a disease type.

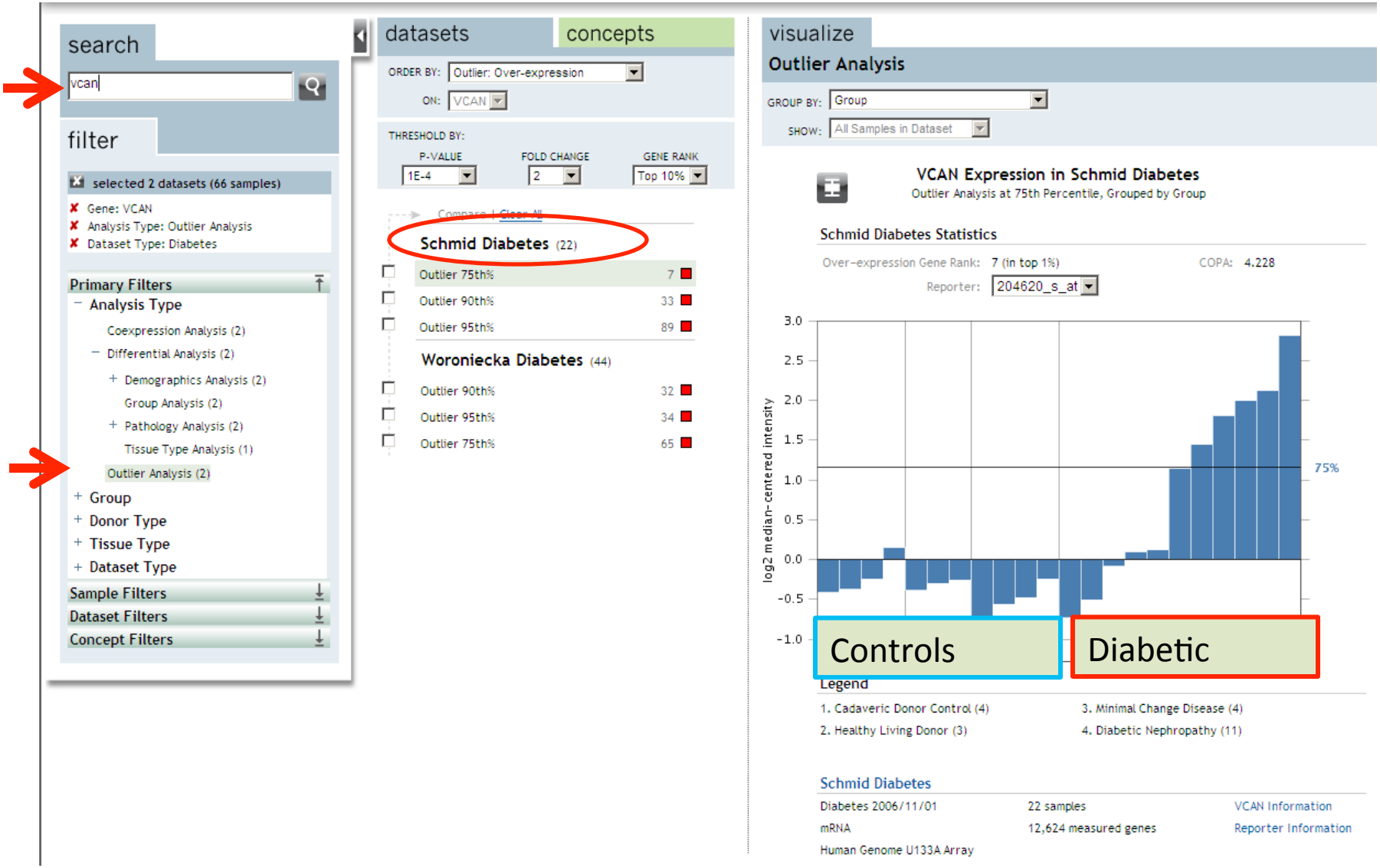
Why do we need it: 25% of patients show over-expression of a gene. This gene may not generate a significant p-value in a t-test comparing DN relative to normal kidney.

How to do it: Transform all samples within a dataset, so that genes could be ranked by their expression from high to low. The data transformation is performed at certain percentile bins (75, 90 & 95%), and a line is drawn at the percentile of that analysis to define outliers.

For example, in an outlier analysis at the 75th percentile, the system draws a line at the point at which only the top 25th percentile samples extend above it.

Gene Search

Outlier analysis



Differential expression – Dataset search

search

filter

selected 16 datasets (1121 samples)

✘ Analysis Type: Differential Analysis

Primary Filters

− Analysis Type

Coexpression Analysis (16)

− Differential Analysis (16)

+ Demographics Analysis (14)

Donor Type Analysis (1)

Group Analysis (9)

+ Indices Analysis (3)

+ Pathology Analysis (11)

Tissue Type Analysis (8)

+ Treatment Analysis (1)

Outlier Analysis (16)

+ Group

+ Donor Type

+ Tissue Type

+ Dataset Type

Sample Filters

Dataset Filters

Concept Filters

datasets | concepts

ORDER BY: Dataset Name

ON:

Compare | Clear All

Flechner Transplant (62)

- [ ] Cadaveric Donor Kidney Specimen: Acute Rejection vs. No Rejection
- [ ] Cadaveric Donor Kidney Specimen: Age
- [ ] Cadaveric Donor Kidney Specimen: GFR (MDRD)
- [ ] Cadaveric Donor Kidney Specimen: Sex
- [ ] Cadaveric Donor Peripheral Blood Lymphocyte Specimen: Acute Rejection vs. No Rejection
- [ ] Cadaveric Donor Peripheral Blood Lymphocyte Specimen: Age
- [ ] Cadaveric Donor Peripheral Blood Lymphocyte Specimen: GFR (MDRD)
- [ ] Cadaveric Donor Peripheral Blood Lymphocyte Specimen: Renal Dysfunction vs. No Rejection
- [ ] Cadaveric Donor Peripheral Blood Lymphocyte Specimen: Sex
- [ ] Cadaveric Donor Tissue Type: Kidney vs. Peripheral Blood Lymphocyte
- [ ] Kidney Specimen Donor Type: Living vs. Cadaveric
- [ ] Living Donor Kidney Specimen: Age
- [ ] Living Donor Kidney Specimen: GFR (MDRD)
- [ ] Living Donor Kidney Specimen: Renal Dysfunction vs. No Rejection
- [ ] Living Donor Kidney Specimen: Sex
- [ ] Living Donor Peripheral Blood Lymphocyte Specimen: Age
- [ ] Living Donor Peripheral Blood Lymphocyte Specimen: GFR (MDRD)
- [ ] Living Donor Peripheral Blood Lymphocyte Specimen: Renal Dysfunction vs. No Rejection
- [ ] Living Donor Peripheral Blood Lymphocyte Specimen: Sex
- [ ] Living Donor Tissue Type: Kidney vs. Peripheral

visualize

Differential Analysis

OTHER VIEWS:

GROUP BY: Group (Cadaveric Donor Kidney Specimen)

SHOW: Only Samples in Analysis

+ PRIMARY CONCEPT: EXPORT:

1 | 2 | 3 | 4 | 5 »

Over-expression

Export

Comparison of All Genes in Flechner Transplant

Over-expression in Cadaveric Donor Kidney Specimen: Acute Rejection vs. No Rejection

(log2 median-centered intensity)

Rank	P-value	Fold Change	Gene	Reporter	Gene
1	2.26E-8	1.51	NIPAL3	37850_at	NIPAL3
2	2.20E-7	1.41	STXBP5L	34130_at	STXBP5L
3	3.33E-7	1.51	KSR1	1716_at	KSR1
4	3.53E-7	1.51	SRC	1938_at	SRC
5	4.04E-7	1.68	LLGL1	804_s_at	LLGL1
6	4.45E-7	1.73	FSTL4	34518_at	FSTL4
7	6.21E-7	1.43	ARID3A	35913_at	ARID3A
8	7.71E-7	2.02	CYP1A1	1024_at	CYP1A1
9	7.93E-7	1.44	RIN1	1777_at	RIN1
10	9.63E-7	1.63	PDGFB	1573_at	PDGFB
11	1.10E-6	2.14	SKI	1918_at	SKI
12	1.28E-6	1.52	COL11A2	1027_at	COL11A2
13	1.36E-6	1.81	PTCH1	836_at	PTCH1
14	1.93E-6	1.63	ZNF646	39863_at	ZNF646
15	2.13E-6	1.55	COL2A1	598_at	COL2A1
16	2.22E-6	1.42	KISS1	1645_at	KISS1
17	2.29E-6	1.39	TBL3	41603_at	TBL3
18	2.31E-6	1.49	CYP2C18	1477_s_at	CYP2C18
19	2.44E-6	1.47	MAST1	35962_at	MAST1
20	2.52E-6	1.34	PCDHGC3	657_at	PCDHGC3
21	3.14E-6	1.40	BRF1	141_s_at	BRF1

1 | 2

Legend

1. No Rejection (5)
2. Acute Rejection (6)

Least Expressed — Most Expressed — Not measured

Note: Colors are z-score normalized to depict relative values within rows. They cannot be used to compare values between rows.

Flechner Transplant

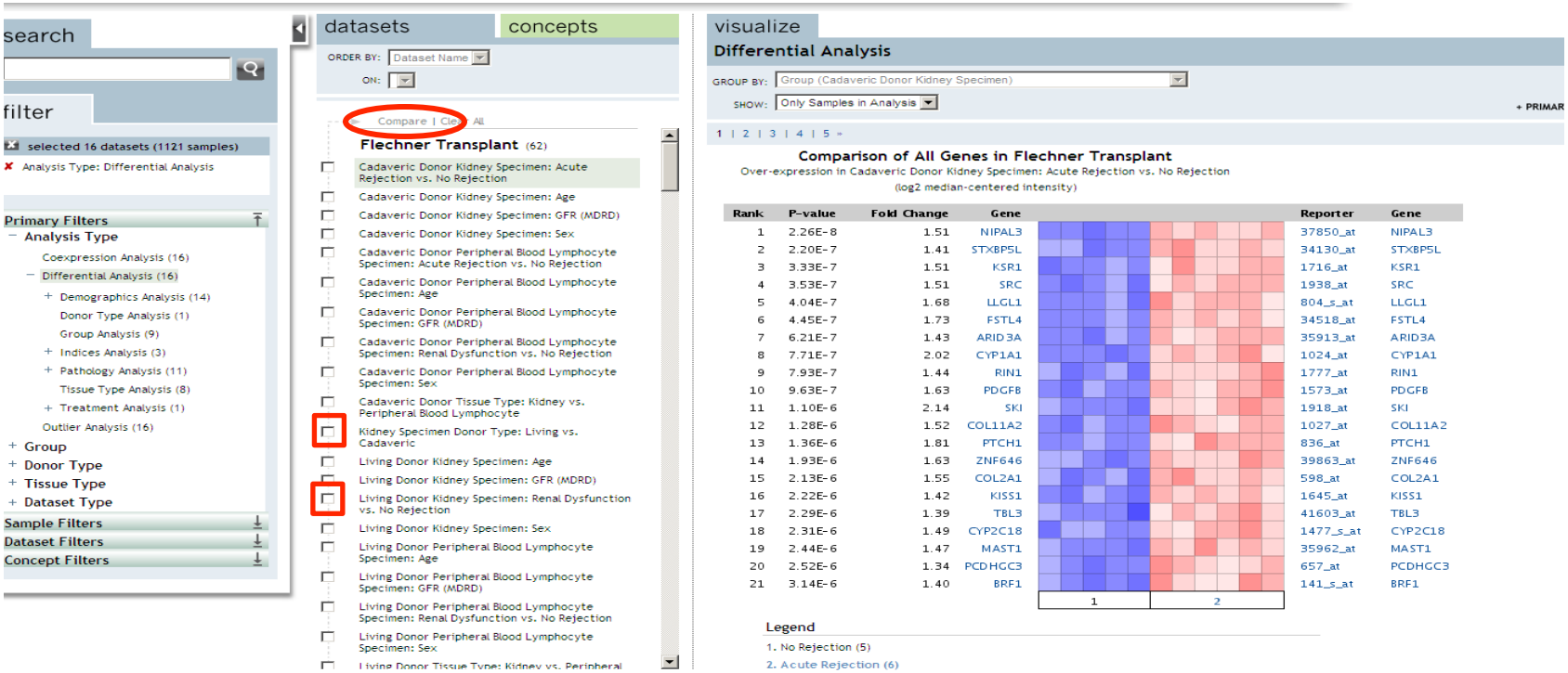
Am J Transplant 2004/09/01 — 62 samples

mRNA — 8,603 measured genes

Human Genome U95A-Av2 Array

Differential expression – dataset search – compare analysis

- Compare different analyzes
- Data is standardized on upload (centered to 0 and standardized by variance)
- all features are mapped to common identifier (EntrezGeneID)



Meta analysis

- Find out which genes are significantly more expressed in glomeruli compared to tubulointerstitium
- Can you verify that with another dataset?
- Or with more than one other dataset?
- Does it matter if the datasets are different?
- Can you imagine a use of this functionality for an exclusive filter (NOT)

Example

Firefox | heise online - IT-News, Nachrichten ... | Millionenfacher Datenklau: Provider s... | Re: Ich versteh's nicht - erklärt's mir b... | Home - SRA - NCBI | Princeton University High Throughpu... | Nephromine Main

www.nephromine.org/resource/main.html#ac%3A1N7515%2C1N9853%2C1N5129%3Bdso%3AdatasetName%3Bec%3A[1N2]%3Bepv%3A1N1.1N3%2C1N41%3Bet%3Anone%3Bpg%3A1%3Bpvf%3A1N800044980%2C1N8000914329

shortReadArchive

NEPHROMINE

Welcome, Felix Eichinger. tools help logout

search

filter

selected 4 datasets (835 samples)

- ✘ Dataset Type: Normal Tissue Panel
- ✘ Dataset Type: Podocyte

Primary Filters

- − Analysis Type
  - Coexpression Analysis (4)
  - − Differential Analysis (4)
    - + Demographics Analysis (1)
    - Group Analysis (1)
    - Tissue Type Analysis (3)
  - Outlier Analysis (4)
- + Group
- + Tissue Type
- − Dataset Type
  - Aging (2)
  - Diabetes (2)
  - Diabetes Mouse (1)
  - FSGS (1)
  - Hypertension (1)
  - IgAN (1)
  - Lupus (2)
  - Lupus Mouse (1)
  - Normal Tissue Panel (3)
  - Podocyte (1)
  - Transplant (5)

Sample Filters

Dataset Filters

Concept Filters

datasets | concepts

ORDER BY: Dataset Name

ON:

Compare | Clear All

Higgins Normal Tissue Panel (34)

- [x] Tissue Type: Glomeruli vs. All Others
- [ ] Tissue Type: Papillary Tips vs. All Others
- [ ] Tissue Type: Renal Cortex vs. All Others
- [ ] Tissue Type: Renal Medulla vs. All Others
- [ ] Tissue Type: Renal Pelvis vs. All Others
- [ ] Outlier 5th%
- [ ] Outlier 10th%
- [ ] Outlier 25th%
- [ ] Outlier 75th%
- [ ] Outlier 90th%
- [ ] Outlier 95th%

Ju Podocyte (436)

- [ ] Glomeruli: Arterial Hypertension vs. Healthy Living Donor
- [ ] Glomeruli: Diabetic Nephropathy vs. Healthy Living Donor
- [ ] Glomeruli: Focal Segmental Glomerulosclerosis vs. Healthy Living Donor
- [ ] Glomeruli: IgA Nephropathy vs. Healthy Living Donor
- [ ] Glomeruli: Lupus Nephritis vs. Healthy Living Donor
- [ ] Glomeruli: Membranous Glomerulonephritis vs. Healthy Living Donor
- [ ] Glomeruli: Minimal Change Disease vs. Healthy Living Donor

visualize

Analysis Comparison

OTHER VIEWS:

GROUP BY:

SHOW:

EXPORT:

1 | 2 | 3 | 4 | 5 »

Over-expression

Comparison of All Genes Across 3 Analyses

Over-expression

Median Rank	p-Value	Gene	1	2	3
108.5	3.13E-8	KCNH3			
124.0	1.89E-6	GPR17			
133.0	6.86E-30	MEG3			
177.0	1.22E-8	CDH19			
245.0	1.18E-22	GNAL			
245.0	2.76E-8	NRXN1			
276.5	6.55E-10	TCERG1L			
294.0	1.74E-7	CD99L2			
295.0	4.21E-8	RRBP1			
295.0	1.12E-4	FKBP15			
298.0	1.14E-4	EZR			
298.5	1.20E-7	SPOCK3			
311.0	4.73E-8	QKI			
328.0	1.65E-4	P2RX7			
329.0	4.11E-20	STXBP1			
342.5	1.31E-8	UBASH3B			
361.0	2.57E-4	HTR2A			
366.0	2.93E-4	GLE1			
370.0	7.78E-8	NEB			
372.0	3.17E-4	SEZ6L2			

RRBP1 Rank: 295
p-Value: 4.21E-8

Legend

1. Tissue Type: Glomeruli vs. All Others
   Higgins Normal Tissue Panel, Mol Biol Cell, 2004
2. Tissue Type: Glomeruli vs. Tubulointerstitium
   Lindenmeyer Normal Tissue Panel, PLoS One, 2010
3. Tissue Type: Brain vs. Kidney
   Roth Normal Tissue Panel, Neurogenetics, 2006

1 5 10 25 % 25 10 5 1 — Not measured

The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis.

Concepts Analysis

Concepts are sets of genes representing some aspect of biology.

Concepts are derived from both **Nephromine gene expression signatures** as well as **third-party sources** such as Gene Ontology, KEGG Pathways, Human Protein Reference Database, etc.

User can upload a self-defined custom concept (a set of genes) to Nephromine to explore it's association with Nephromine and third-party concepts.

Concepts Analysis

Upload Custom Concept
Manage My Concepts
Change password

Nephromine Overview

NEW IN NEPHROMINE :: MARCH 2012

OTHER VIEWS:

Nephromine: Upload My Concept - Windows Internet Explorer

http://www.nephromine.org/resource/ui/tool/concept.html?ACTION=SHOW_UPLOAD_CONCEPT

Upload My Concept

Podo-50-symbol

Download list from C-tools to the desktop, then upload

Concept Name:

Gene Set (Text File):

Category: HUGO Gene Symbol

Null Set(s):

- Affymetrix Human Genome HT U133 Plus 2.0 PM Array
- Agilent-016436 Human miRNA Microarray 1.0 (CBI v1)
- Agilent-019118 Human miRNA Microarray 2.0 G4470B (CBI v1)
- All Entrez Gene IDs
- CodeLink Human Whole Genome Bioarray

Description (Optional):

(Up to 500 characters)

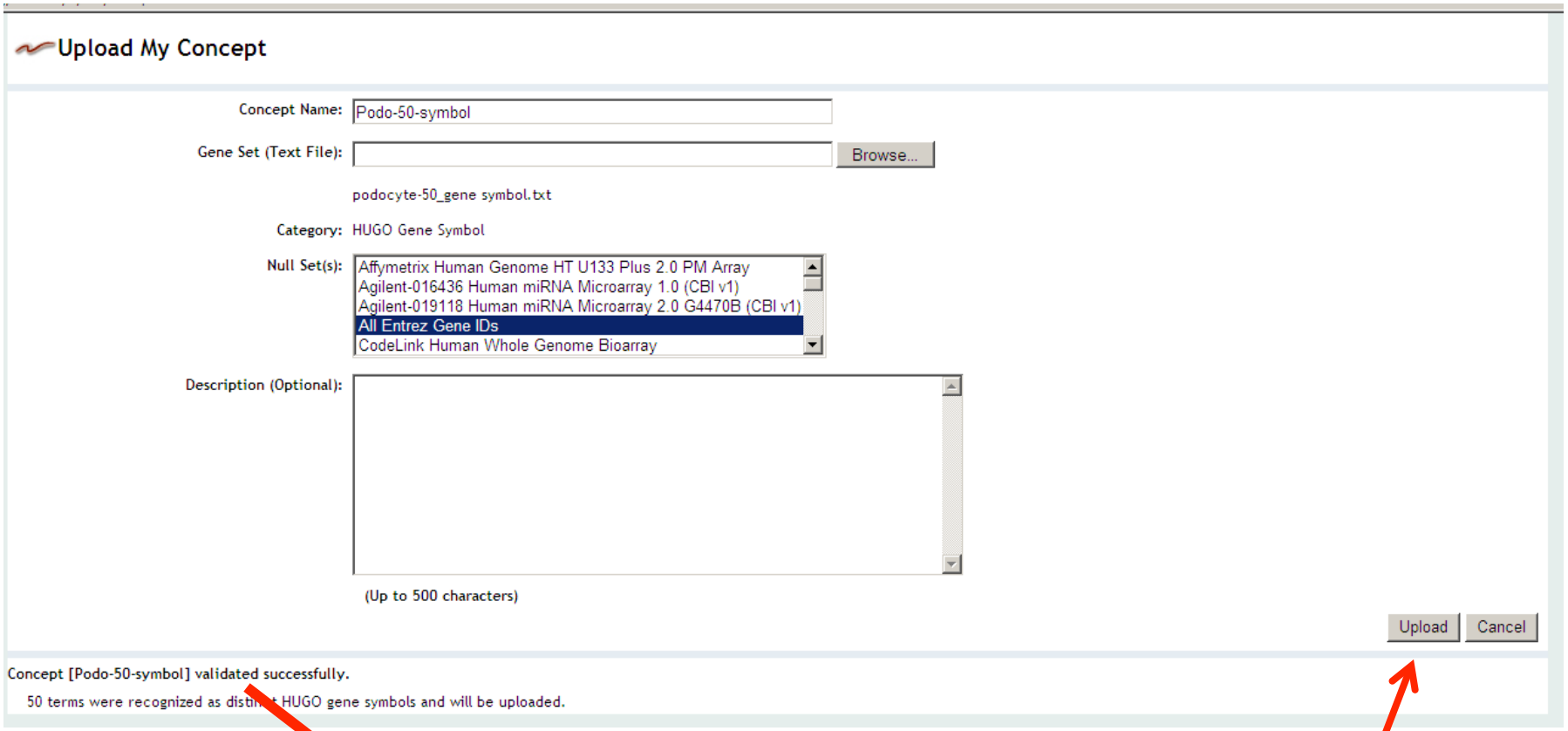
The press "validate"

Done

Internet

100%

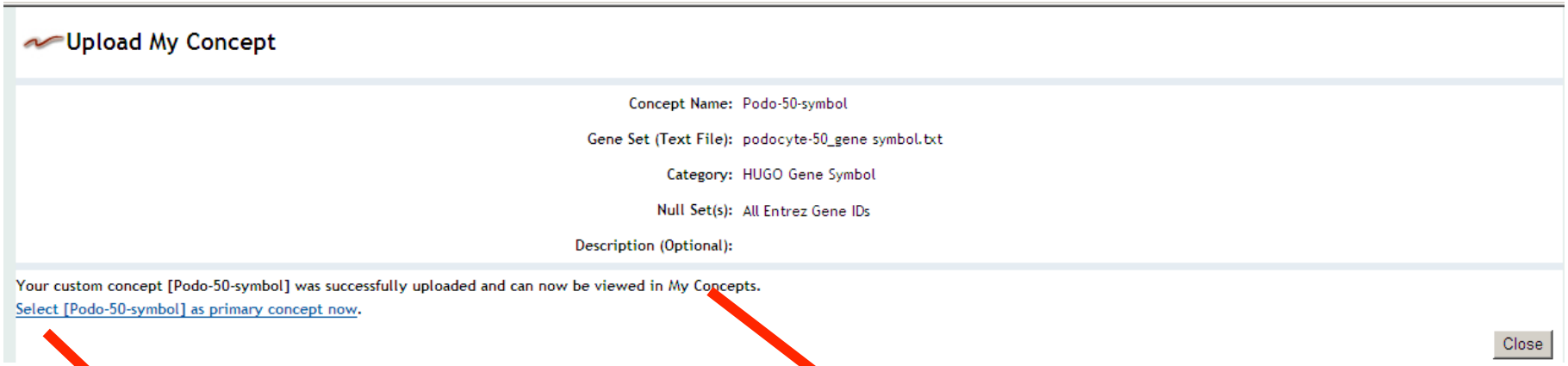
Concepts Analysis Upload



Then press “Upload”

Concept (Podo-50-symbol) validated successfully.

Concepts Analysis Upload



Concept (Podo-50-symbol) was successfully uploaded and can be viewed in My Concepts

Select (Podo-50-symbol) as primary concept now.

Concepts Summary View

Nephromine Concept Summary

4 concepts meet your threshold and are associated with the primary concept

search

filter

selected 16 datasets (1121 samples)

✘ Concept: Podo-50-symbol - My Concepts

Primary Filters

- Analysis Type
  - Coexpression Analysis (16)
  - Differential Analysis (16)
    - + Demographics Analysis (14)
    - Donor Type Analysis (1)
    - Group Analysis (9)
    - + Indices Analysis (3)
    - + Pathology Analysis (11)
    - Tissue Type Analysis (8)
    - + Treatment Analysis (1)
  - Outlier Analysis (16)
- + Group
- + Donor Type
- + Tissue Type
- + Dataset Type

Sample Filters

Dataset Filters

Concept Filters

visualize

Concept Summary

THRESHOLD (ODDS RATIO): 2.0 THRESHOLD (P-VALUE): 1E-4 DATA TYPE: All

Associated Concept Summary for "Podo-50-symbol - My Concepts"

Nephromine Concept Summary

Concept Type by Dataset Type	Demographics				Donor Type	Group	Indices		Pathology							Tissue Type		Treatment	Nephromine Clusters
	Age	Body Mass Index	Race/ Ethnicity	Sex			Activity Index Quartile	Chronicity Index Quartile	BUN	GFR (MDRD)	Glomerulo-sclerosis	Hemoglobin	Proteinuria	TAIF	WHO Lupus Nephritis Class			Post-Event Treatment	
Aging																	1		2
Diabetes						1				2							1		1
FSGS						4													1
Hypertension																			
IgAN						1											1		1
Lupus																			
Normal Tissue Panel																1	6		3
Transplant																			1
Significant Unique Concepts						6				2						1	9		9

Red: Over-expression Blue: Under-expression

Other (Non-Nephromine) Concept Summary

Biological Annotations	Pathway Concepts	Regulatory Concepts	Connectivity Map v2 Drug Signatures	Literature-defined Concepts	Mutation Concepts	My Concepts	shRNA Concepts
7	7		3	17		3	

Other (Non-Nephromine) Concept Summary

Concepts Analysis

COLL FSGS vs. Normal kidney

Nephromine Gene Expression Signatures

P=1.54E-18, q=1.15E-14, Odds=18

Top 5% Under-expressed

Hodgin FSGS

search

filter

Outlier Analysis (1)

+ Group

+ Dataset Type

Sample Filters

Dataset Filters

Concept Filters

datasets | associated concepts

Primary Concept: Podo-50-symbol - My Concepts

ORDER BY: p-Value

THRESHOLD BY:

ODDS RATIO: 2.0 P-VALUE: 1E-4

Group: Focal Segmental Glomerulosclerosis vs. Normal Kidney

Nephromine Gene Expression Signatures

p=2.21E-8 q=2.56E-5 Odds=6.1

Top 10% Under-expressed

Hodgin FSGS

Group: Focal Segmental Glomerulosclerosis vs. Minimal Change Disease and Normal Kidney

Nephromine Gene Expression Signatures

p=7.15E-5 q=0.032 Odds=3.9

Top 10% Under-expressed

Hodgin FSGS

visualize

Concept Association Results

OTHER VIEWS:

EXPORT:

Primary Concept:

Name: Podo-50-symbol

Concept Type: My Concepts

Size: 50 genes

Null List: All Entrez Gene IDs (42490)

ssociated Concept:

+ primary concept

Name: Group: Collapsing Focal Segmental Glomerulosclerosis vs. Normal Kidney - Top 5% Under-expressed (Hodgin FSGS)

Concept Type: Nephromine Gene Expression Signatures

Size: 956 genes

Null List: Affymetrix Human X3P Array (19139)

teraction:

alue: 1.54E-18 Q-value: 1.15E-14 Odds Ratio: 18.0 Size: 24 genes

TN4	actinin, alpha 4
AGRN	agrin
CD2AP	CD2-associated protein
CD80	CD80 molecule
CLIC5	chloride intracellular channel 5
DAG1	dystroglycan 1 (dystrophin-associated glycoprotein 1)
EZR	ezrin
FYN	FYN oncogene related to SRC, FGR, YES
LRRC7	leucine rich repeat containing 7
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)
MAGI2	membrane associated guanylate kinase, WW and PDZ domain containing 2
MME	membrane metallo-endopeptidase
NES	nestin
NPHS1	nephrosis 1, congenital, Finnish type (nephrin)
NPHS2	nephrosis 2, idiopathic, steroid-resistant (podocin)
PLCE1	phospholipase C, epsilon 1
PODXL	podocalyxin-like
PTPRO	protein tyrosine phosphatase, receptor type, O
SCEL	sciellin
SULF1	sulfatase 1
SYNPO	synaptopodin
TCF21	transcription factor 21
TJP1	tight junction protein 1 (zona occludens 1)
WT1	Wilms tumor 1

Concepts Analysis

datasets | associated concepts

ORDER BY: Dataset Name

ON:

Compare | Clear All

Hodgin FSGS (30)

- Group: Collapsing Focal Segmental Glomerulosclerosis vs. Focal Segmental Glomerulosclerosis
- Group: Collapsing Focal Segmental Glomerulosclerosis vs. Minimal Change Disease and Normal Kidney
- Group: Collapsing Focal Segmental Glomerulosclerosis vs. Normal Kidney
- Group: Focal Segmental Glomerulosclerosis vs. Minimal Change Disease and Normal Kidney
- Group: Focal Segmental Glomerulosclerosis vs. Normal Kidney
- Group: Minimal Change Disease vs. Normal Kidney

visualize

Differential Analysis

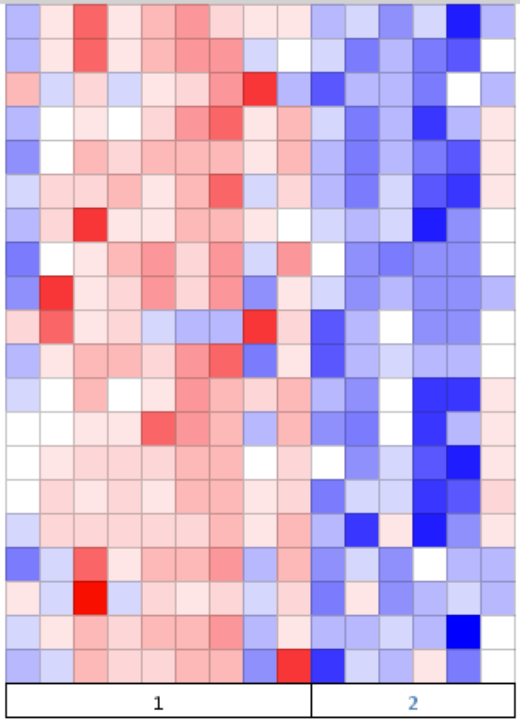
GROUP BY: Group

SHOW: Only Samples in Analysis

1 | 2 | 3 »

Comparison of Concept: "Podo-50-symbol - My Conc
Under-expression in Group: Collapsing Focal Segmental Glomerul
(log2 median-centered intensity)

Rank	P-value	Fold Change	Gene	Reporter
99	0.001	-2.02	ACTN4	g3157975_3p_at
100	0.001	-3.74	SYNPO	g6005797_3p_at
142	0.002	-1.75	MAGI2	Hs.229355.0.A1_3p_at
156	0.002	-2.07	TJP1	g4507516_3p_at
171	0.002	-2.56	PODXL	g4885556_3p_at
194	0.002	-2.63	CLIC5	g8393146_3p_at
234	0.003	-2.78	NES	g13375818_3p_at
272	0.003	-2.19	SULF1	Hs.70823.0.S3_3p_at
278	0.003	-2.18	NPHS1	207673_3p_at
359	0.004	-1.25	LRRC7	Hs2.157325.1.S1_3p_s_at
385	0.005	-2.67	TCF21	g4507394_3p_at
504	0.006	-2.46	NPHS2	g7657614_3p_at
580	0.008	-1.58	DAG1	g4758115_3p_a_at
590	0.008	-3.31	PLCE1	g7705940_3p_s_at
594	0.008	-1.57	FYN	g181171_3p_a_at
658	0.009	-1.97	WT1	g13386509_3p_a_at
696	0.009	-2.99	EZR	g340216_3p_a_at
707	0.009	-1.13	CD80	Hs2.838.3.S1_3p_at
717	0.009	-2.39	MAFB	Hs.169487.0.S1_3p_a_at
747	0.010	-2.40	CD2AP	g11321633_3p_at



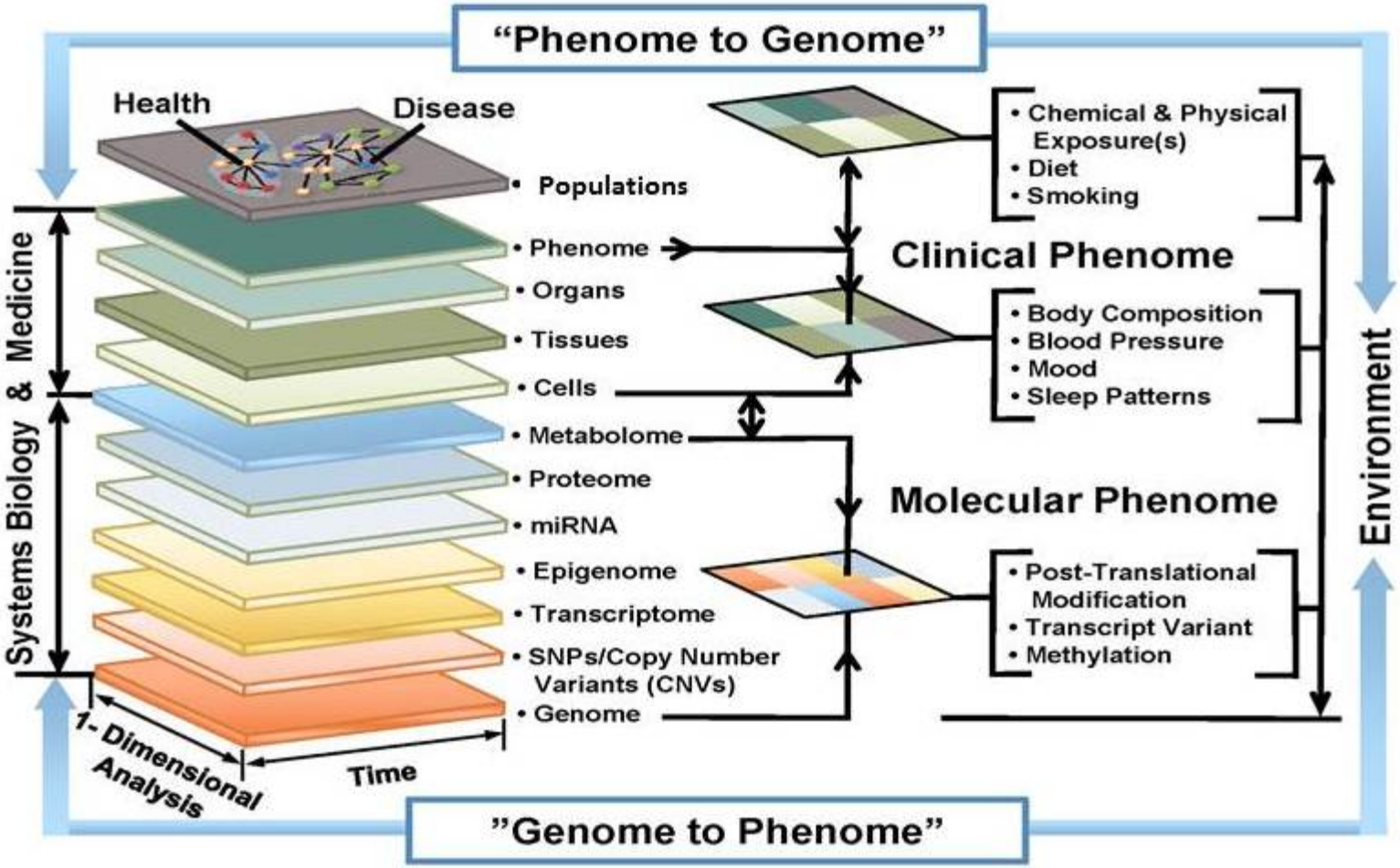
Legend

1. Normal Kidney (9)
2. Collapsing Focal Segmental Glomerulosclerosis (6)

PowerPoint
Publication-quality graphic (SVG)
Excel - Analysis Comparison
Excel - Analysis Gene List
Excel - Dataset Detail

tranSMART

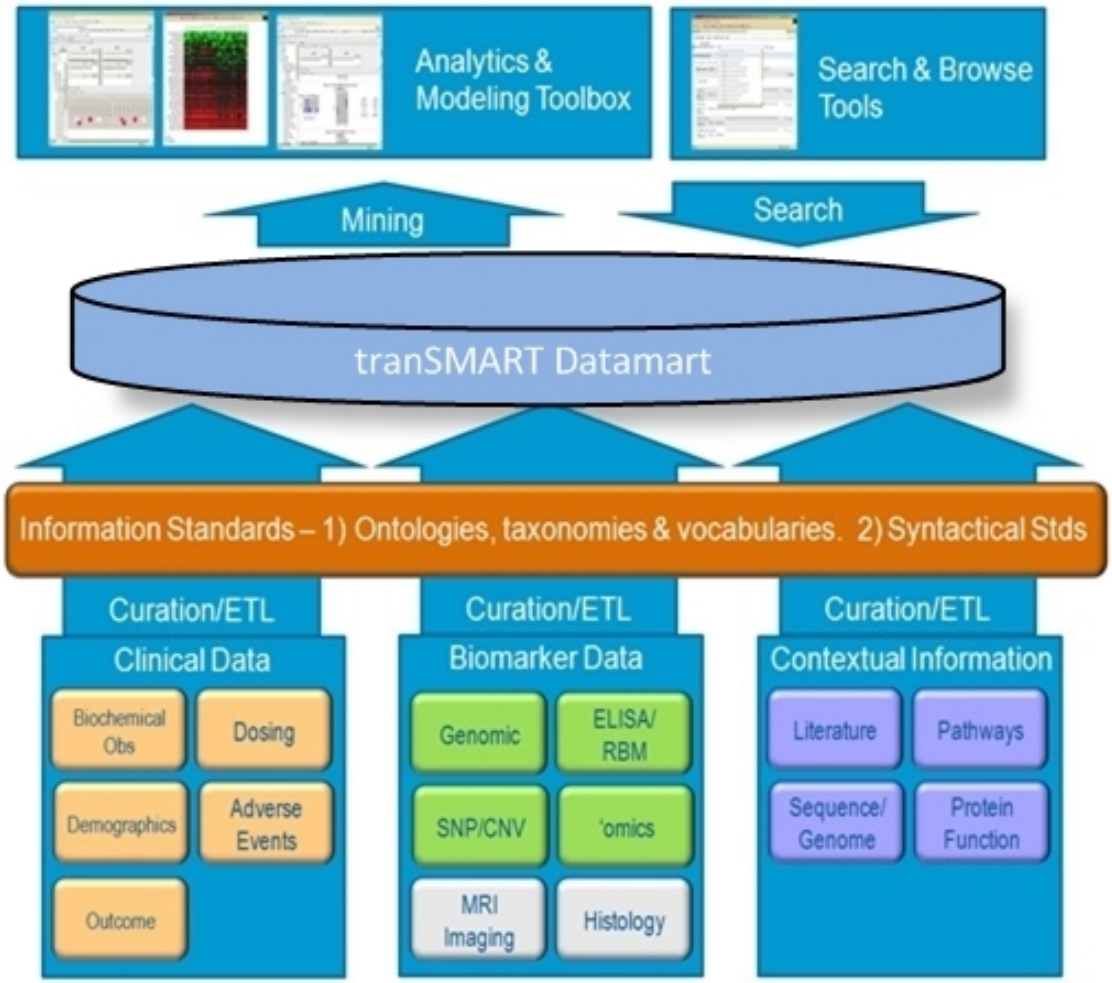
The Translational Challenge: Data Integration & Analysis



Athey and Omenn, 2009

MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN

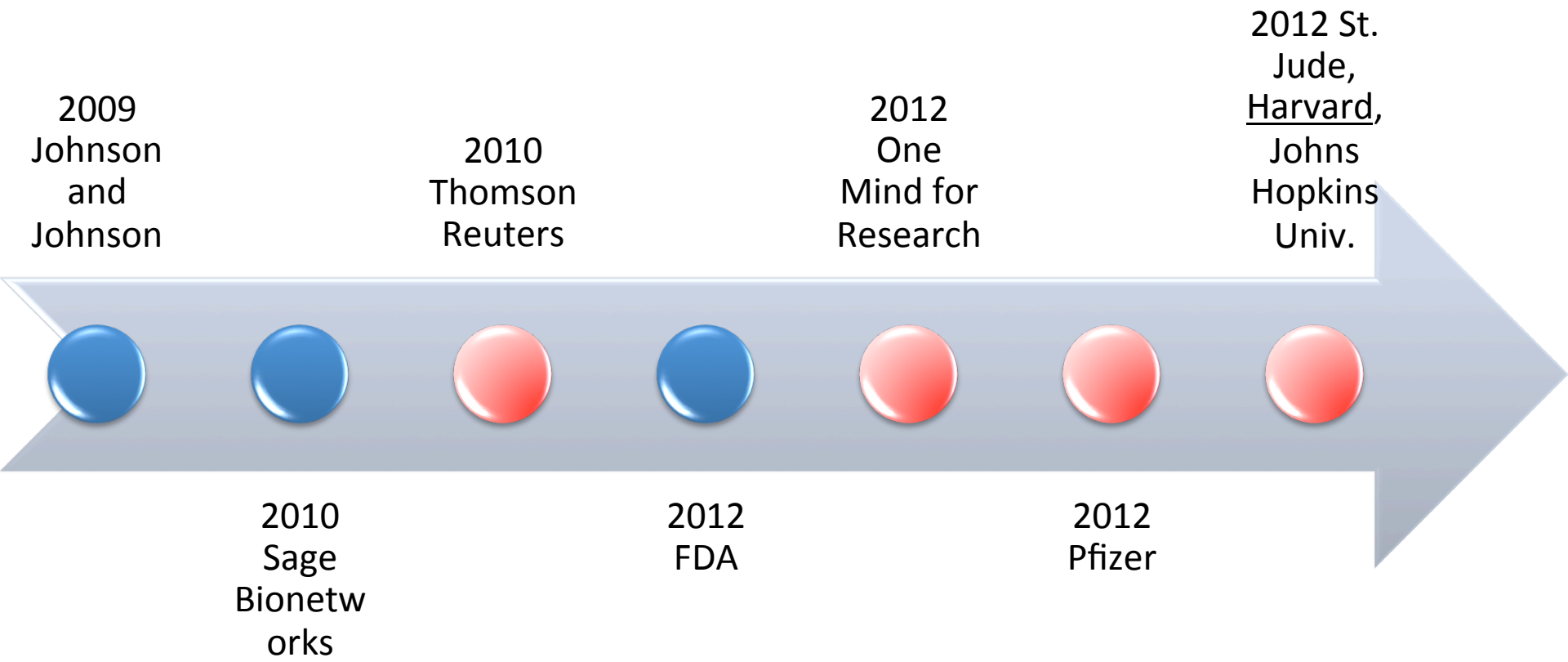
tranSMART Platform: Enabling Translational research



tranSMART –
A platform and community

- Open-source and open-data translational biomedical research community
- Biomedical Researchers, Developers, Service Providers
- Clinician Researchers

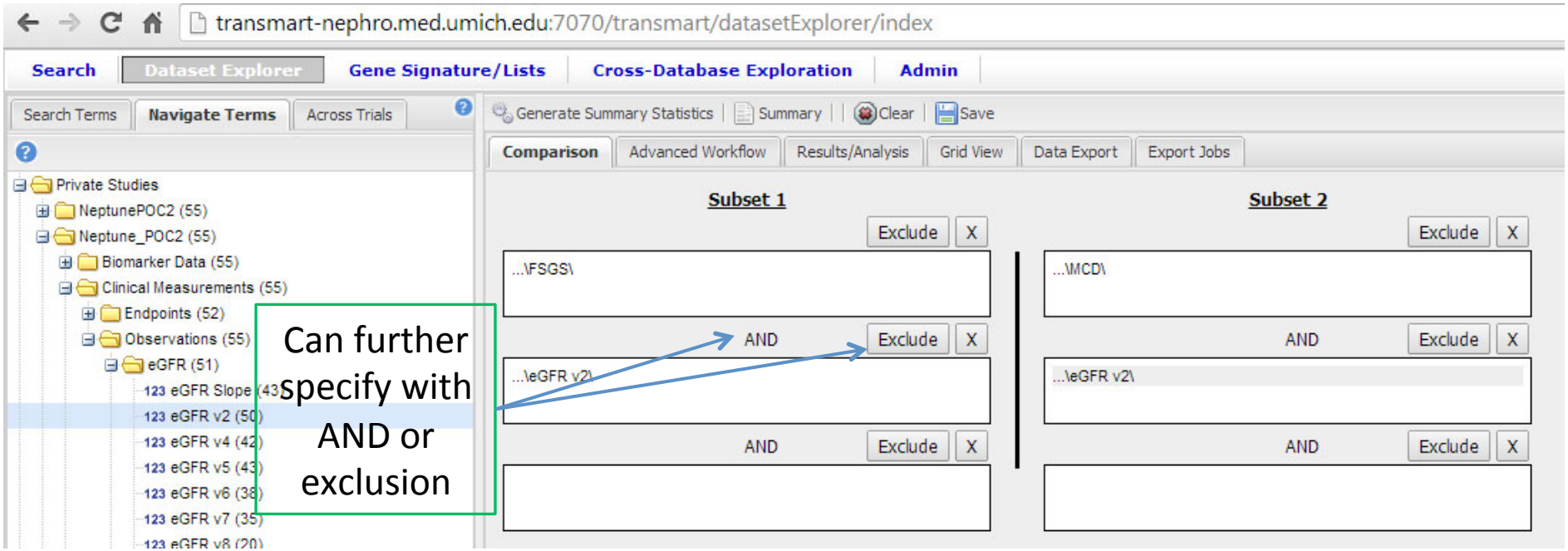
tranSMART Platform: Academics and industry



tranSMART: controlled vocabulary

- 01 Demographics (460)
  - Ethnicity (458)
  - Gender (460)
  - Pediatric or Adult (460)
  - Race (456)
  - Race includes Black? (438)
  - 123 Age (460)
  - 123 Age of Onset (years) (402)
  - 123 Kidney disease duration (months) (373)
- 02 Disease Cohort (460)
- 03 Clinical Data and Measurements (460)
  - 01 Measurements (460)
  - 02 Endpoints (460)
  - 03 Genotype (87)
  - 04 Morphology (292)
  - 05 Morphometry (76)
- 04 Medical History and Medications (460)
  - Medical History (460)
  - Medications (445)
- 05 Socio-economic Status and Healthcare (459)
  - Healthcare (456)
  - Socio-economic Status (459)
- 06 Biospecimens Collected (451)
- 07 Proteomics (310)
  - Serum (310)
  - Urine (310)
- 08 Gene Expression Data (264)

Subset selection



Can further specify with AND or exclusion

Subset 1

Subset 2

Summary statistics 1

Generate Summary Statistics | Summary | Clear | Save | Export | Print

Comparison | Advanced Workflow | **Results/Analysis** | Grid View | Data Export | Export Jobs

Summary Statistics

Query Summary for Subset 1

(\\Private Studies\Private Studies\Neptune_POC2\Subjects\Medical History\Disease\dx\FSGS\ )
AND
(\\Private Studies\Private Studies\Neptune_POC2\Clinical Measurements\Observations\eGFR\eGFR v2\ )

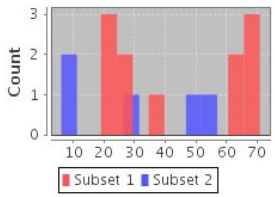
Query Summary for Subset 2

(\\Private Studies\Private Studies\Neptune_POC2\Subjects\Medical History\Disease\dx\MCD\ )
AND
(\\Private Studies\Private Studies\Neptune_POC2\Clinical Measurements\Observations\eGFR\eGFR v2\ )

Subject Totals

Subset 1	Both	Subset 2
11	0	5

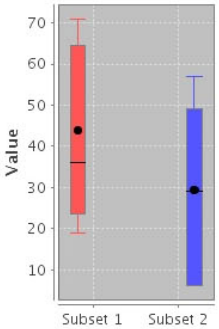
Histogram of Age



Subset 1

Mean: 43.82
Median: 36
IQR: 41
SD: 21.84
Data Points: 11

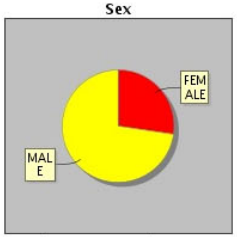
Comparison of Age



Subset 2

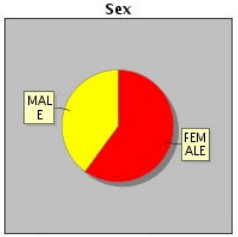
Mean: 29.4
Median: 29
IQR: 43
SD: 23.67
Data Points: 5

Sex



Category	Subset 1 (n)	Subset 1 (%n)
F	0	0%
FEMALE	3	27.3%
M	0	0%
MALE	8	72.7%
UNKNOWN	0	0%
null	0	0%
Total	**11**	**100%**

Sex



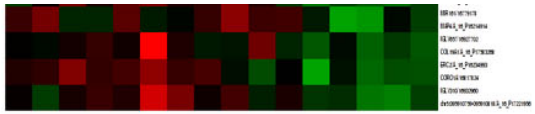
Category	Subset 2 (n)	Subset 2 (%n)
F	0	0%
FEMALE	3	60%
M	0	0%
MALE	2	40%
UNKNOWN	0	0%
null	0	0%
Total	**5**	**100%**

Race

Race

ASIAN/

Differentially expressed genes



Gene symbols

P-values

Fold change

Table of top Markers

Gene Symbol	Probe ID	Raw p-value	Bonferroni	Holm	Hochberg	SidakSD	SidakSS	BH	BY	t	t (permutation)	Raw P (permutation)	Adjusted P (permutation)	Rank	S1 Mean	S2 Mean	S1 SD	S2 SD	Fold Change
CENPE	A_16_P16795940	0.00000	0.00037	0.00037	0.00037	0.00037	0.00037	0.00018	0.00209	-5.789798	-5.789798	0.0004578755	0.2426740	1	0.38883375	-0.585704500	0.4883929	0.1823082	-0.66387358
LTF	16952883	0.00000	0.00037	0.00037	0.00037	0.00037	0.00037	0.00018	0.00209	-5.789798	-5.789798	0.0004578755	0.2426740	2	0.38883375	-0.585704500	0.4883929	0.1823082	-0.66387358
CORO1A	16817824	0.00001	0.40718	0.40716	0.40715	0.33447	0.33446	0.09503	1.00000	-4.469327	-4.469327	0.0011446886	0.8759158	3	0.63605159	-0.384358980	0.4674378	0.4016459	-1.65483734
ERC2	A_16_P16234993	0.00001	0.40718	0.40716	0.40715	0.33447	0.33446	0.09503	1.00000	-4.469327	-4.469327	0.0011446886	0.8759158	4	0.63605159	-0.384358980	0.4674378	0.4016459	-1.65483734
chr5:095910759-095910818	A_16_P17221956	0.00001	0.57016	0.57011	0.57010	0.43456	0.43454	0.09503	1.00000	-4.396770	-4.396770	0.0006868132	0.9063645	5	0.72159773	-0.020050546	0.4414995	0.2316556	-35.98893154
IGLV310	16932960	0.00001	0.57016	0.57011	0.57010	0.43456	0.43454	0.09503	1.00000	-4.396770	-4.396770	0.0006868132	0.9063645	6	0.72159773	-0.020050546	0.4414995	0.2316556	-35.98893154
MAP7	A_16_P17737416	0.00002	1.00000	1.00000	1.00000	0.68246	0.68242	0.14339	1.00000	4.242481	4.242481	0.0052655678	0.9553571	7	-0.05458949	0.881133820	0.5509338	0.3244491	-0.06195369
RN5S473	16866280	0.00002	1.00000	1.00000	1.00000	0.68246	0.68242	0.14339	1.00000	4.242481	4.242481	0.0052655678	0.9553571	8	-0.05458949	0.881133820	0.5509338	0.3244491	-0.06195369
AK092155	A_16_P17061075	0.00004	1.00000	1.00000	1.00000	0.90285	0.90282	0.23315	1.00000	4.080551	4.080551	0.0050366300	0.9809982	9	-0.06729611	0.450180580	0.3635915	0.1425495	-0.14948691

Enlarged:

Table of top Markers

Gene Symbol	Probe ID	Raw p-value	Bonferroni
CENPE	A_16_P16795940	0.00000	0.00037
LTF	16952883	0.00000	0.00037
CORO1A	16817824	0.00001	0.40718
ERC2	A_16_P16234993	0.00001	0.40718
chr5:095910759-095910818	A_16_P17221956	0.00001	0.57016

Comparisons can be saved/emailed

transmart-nephro.med.umich.edu:7070/transmart/datasetExplorer/index

Search | Dataset Explorer | Gene Signature/Lists | Cross-Database Exploration | Admin

Search Terms | Navigate Terms | Across Trials

- Private Studies
  - NeptunePOC2 (55)
    - Biomarker Data (55)
      - ST2 1 (55)
        - kidney tub (55)
    - Clinical Measurements (55)
      - Endpoints (52)
        - 123 evercr (52)
        - 123 evercrpr (52)
      - Observations (55)
        - eGFR (51)
          - 123 eGFR Slope (43)
          - 123 eGFR v2 (50)
          - 123 eGFR v4 (42)
          - 123 eGFR v5 (43)
          - 123 eGFR v6 (38)
          - 123 eGFR v7 (35)
          - 123 eGFR v8 (20)
          - 123 eGFR v9 (7)
        - Serum Creatinine (53)
          - 123 Screat v2 (53)
          - 123 Screat v4 (42)
          - 123 Screat v5 (43)

Generate Summary Statistics | Summary | Clear | Save

Comparison | Advanced Workflow | Results/Analysis | Grid View | Data Export | Export Jobs

Analysis

Analysis: Heatmap

Cohorts:
Subset 1: (\Private Studies\NeptunePOC2\Subjects\Medical History\Disease\dx\FSGS\ )
AND
(\Private Studies\NeptunePOC2\Clinical Measurements\Observations\eGFR\eGFR v2\ )
Subset 2: (\Private Studies\NeptunePOC2\Subjects\Medical History\Disease\dx\MCD\ )
AND
(\Private Studies\NeptunePOC2\Clinical Measurements\Observations\eGFR\eGFR v2\ )

Variable Selection

Heatmap Variable

Select a High Dimensional Data node from the Data Set Explorer Tree and drag it into the box.

X

...\kidney tub\

High Dimensional Data

Max rows to display : 50

Saved Comparison

ID: 1797249

Email this comparison

tranSMART – why do we care?

- Enables data exploration with low hurdles
- Integrates many different data types
- Has interfaces to real analysis tools
- Provides a consistent data set
- Can be run locally/ institutional etc
- Can possibly be "shared" across institutions
  - McMurry et al, PLOS one: *Shrine: enabling nationally scalable Multi-site disease studies*
- Go to: http://transmartfoundation.org/

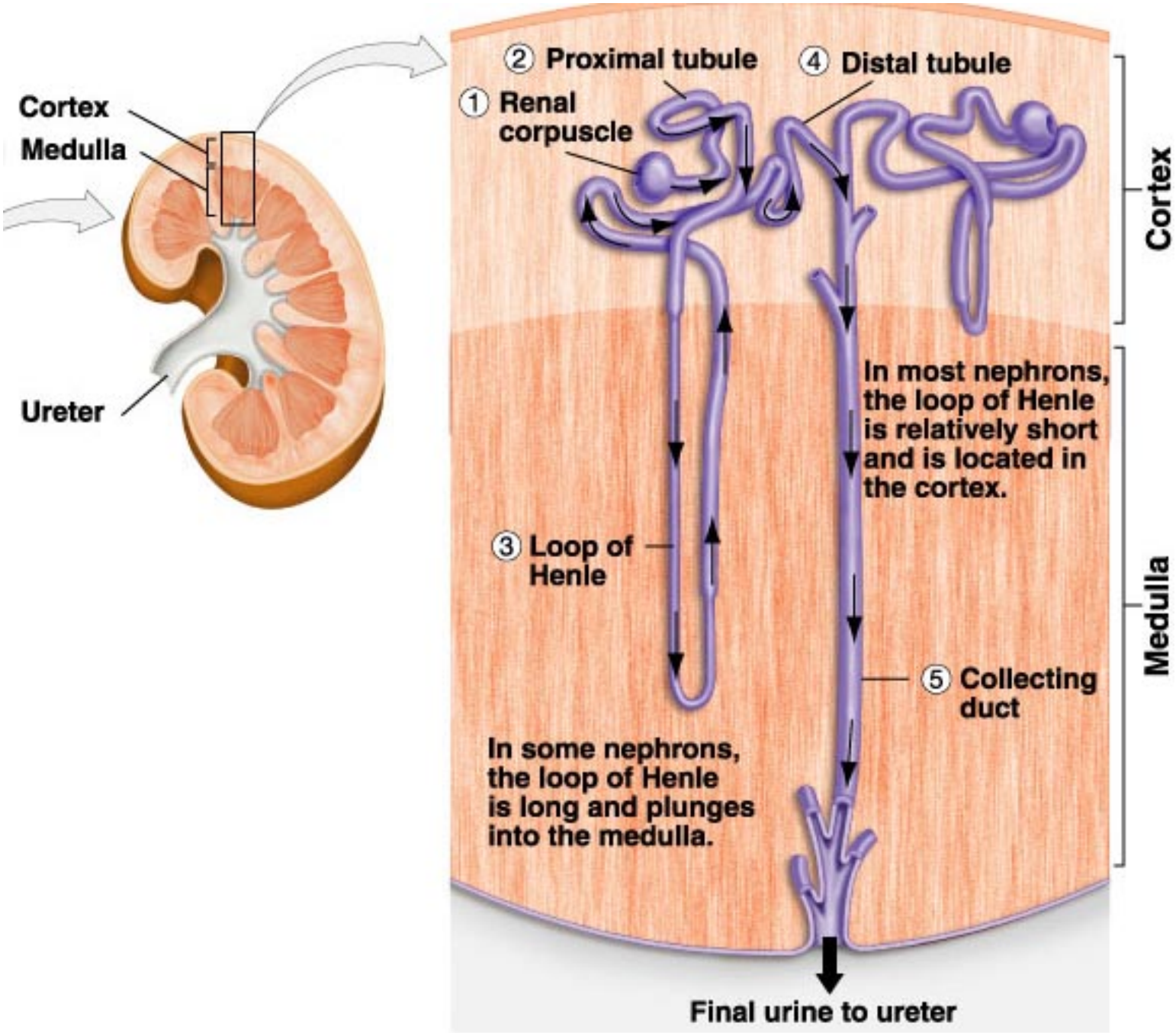
Acknowledgements

Matthias Kretzler
Felix Eichinger
Wenjun Ju
Sebastian Martini
Viji Nair
Celine Berthier
Laura Mariani
Becky Steck
Colleen Kincaid-Beal
Rachel Dull

Daniel R. Rhodes
Rodney Keteyian
Becky Steck
Colleen Kincaid-Beal
Rachel Dull

Homework for fun

- Connectivity map
  - Use Diabetes vs. control (tubulointerstitium dataset)
  - Select top 1% overexpressed as primary concept
  - Compare to significantly overlapping concepts with Connectivity map
  - Can you find potential drug candidates? Are there any drugs that work for both glom. and tub?
  - What could be optimized? How will you plan further experiments to test your hypothesis?


Cortex
Medulla
Ureter
② Proximal tubule
④ Distal tubule
① Renal corpuscle
Cortex
In most nephrons, the loop of Henle is relatively short and is located in the cortex.
③ Loop of Henle
Medulla
⑤ Collecting duct
In some nephrons, the loop of Henle is long and plunges into the medulla.
Final urine to ureter