Learning from differential expression

Japanese Toxicogenomic Project

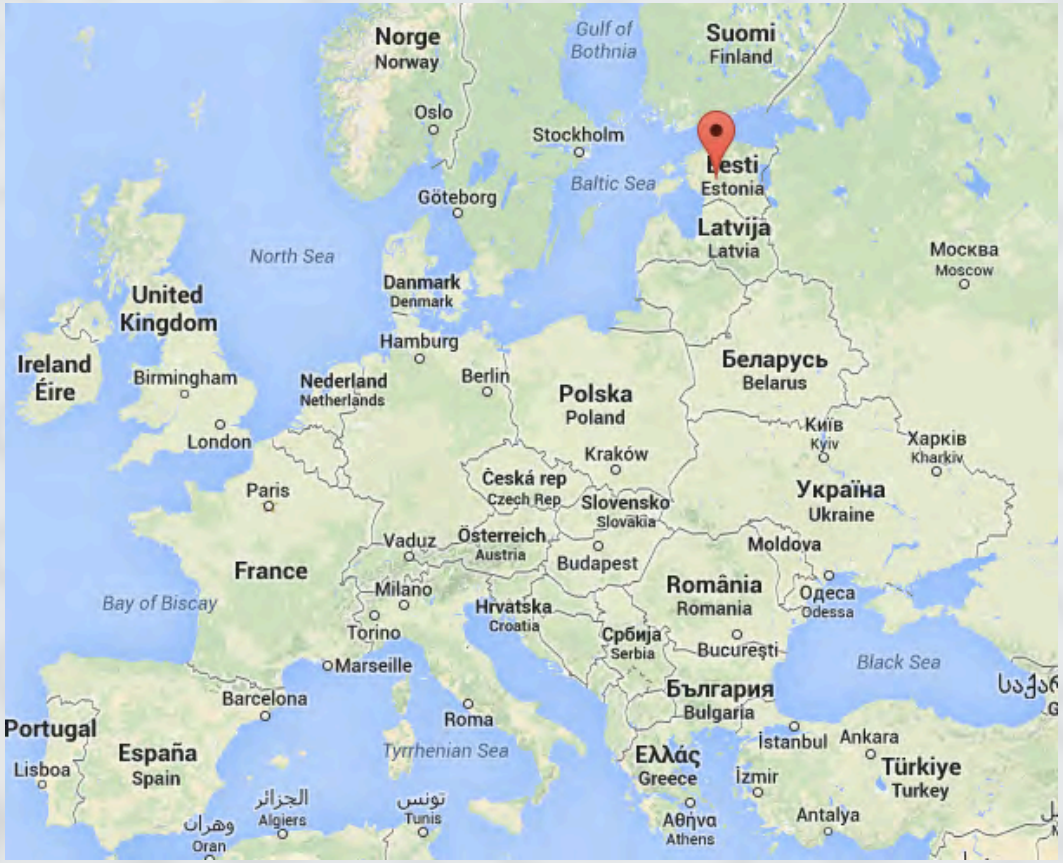
Hendrik Luuk, PhD

Dept. of Physiology
University of Tartu
Estonia

Outline

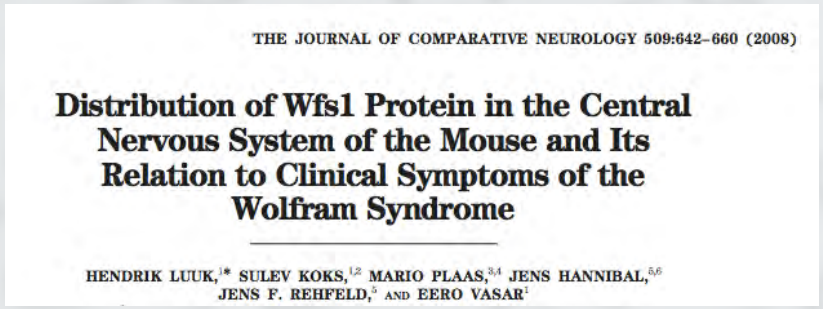
- Personal background
- Japanese Toxicogenomic Project
- Analytical objective
- Methods
- Results
- Discussion

University of Tartu



Personal background

- neurosciences PhD (University of Tartu, 2009)
- Post-doc 2009-2012 (BBH @ Copenhagen)



THE JOURNAL OF COMPARATIVE NEUROLOGY 509:642–660 (2008)

Distribution of Wfs1 Protein in the Central Nervous System of the Mouse and Its Relation to Clinical Symptoms of the Wolfram Syndrome

HENDRIK LUUK,[1*] SULEV KOKS,[1,2] MARIO PLAAS,[3,4] JENS HANNIBAL,[5,6] JENS F. REHFELD,[5] AND EERO VASAR[1]

ELSEVIER

Behavioural Brain Research

Volume 198, Issue 2, 17 March 2009, Pages 334–345

Research report

Wfs1-deficient mice display impaired behavioural adaptation in stressful environment

Hendrik Luuk[1], Mario Plaas[1], Sirli Raud[1], Jürgen Innos[1], Silva Sütt, Helena Lasner, Urho Abramov, Kaido Kurrikoff, Sulev Kõks, Eero Vasar

ELSEVIER

Biochemical and Biophysical Research Communications

Volume 424, Issue 4, 10 August 2012, Pages 717–723

Circadian rhythms and food anticipatory behavior in Wfs1-deficient mice

Hendrik Luuk[a, b, c], Jan Fahrenkrug[a], Jens Hannibal[a]

PLOS ONE | Subject Areas | For Authors | About Us | Search | advanced search

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

4,004 VIEWS | 16 CITATIONS | 16 SAVES

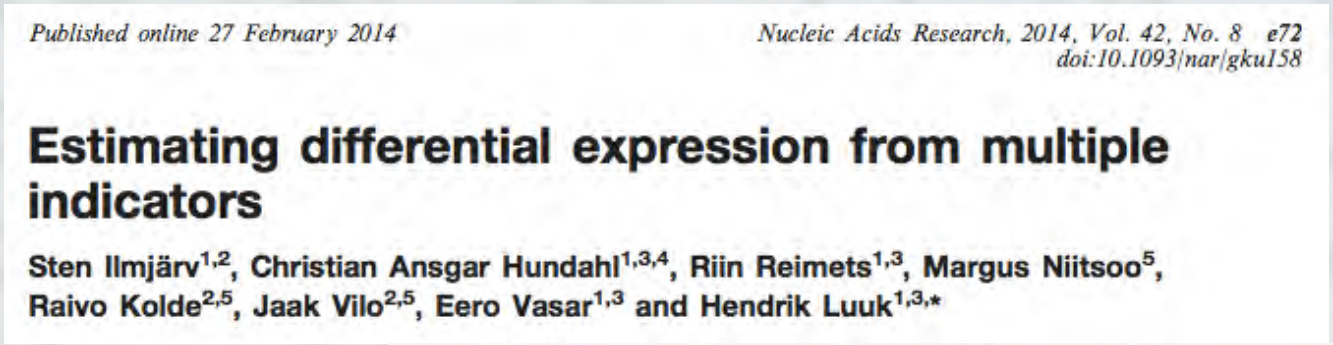
Neuroglobin-Deficiency Exacerbates *Hif1A* and c-FOS Response, but Does Not Affect Neuronal Survival during Severe Hypoxia *In Vivo*

Christian Ansgar Hundahl, Hendrik Luuk, Sten Ilmjärv, Birgitte Falktoft, Zindy Raida, Jonas Vikesaa, Lennart Friis-Hansen, Anders Hay-Schmidt

Published: December 02, 2011 • DOI: 10.1371/journal.pone.0028160

Current focus

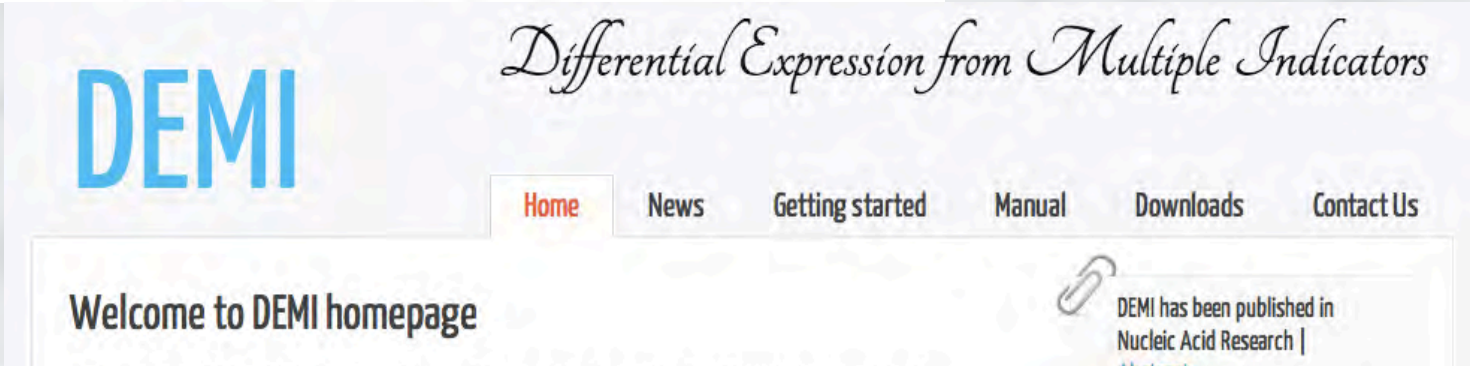
- Molecular mechanisms of therapeutic hypothermia
- Using omics-data to gain mechanistic insights



Published online 27 February 2014 *Nucleic Acids Research, 2014, Vol. 42, No. 8 e72*
doi:10.1093/nar/gku158

Estimating differential expression from multiple indicators

Sten Ilmjärv[1,2], Christian Ansgar Hundahl[1,3,4], Riin Reimets[1,3], Margus Niitsoo[5], Raivo Kolde[2,5], Jaak Vilo[2,5], Eero Vasar[1,3] and Hendrik Luuk[1,3,*]



Japanese Toxicogenomic Project

	Human (primary hepatocytes)	**Rat** (primary hepatocytes)	**Rat** (in vivo, single administration)	**Rat** (in vivo, repeated administration)
Drugs	119	131	131	131
Time points	2	3	4	4
Doses	2	3	3	3
Replicates	2	2	3	3

Japanese Toxicogenomic Project

DILI concern	No. of compounds
high	41
medium	51
nontoxic	8
NA	31
Total	131

Analytical objective

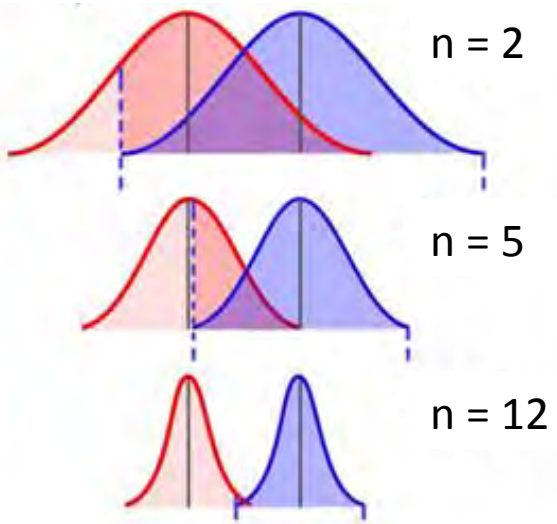
- predict DILI potential from in vitro data
- use differential expression (DE) significance values as features
- suggest optimal treatment designs

Methodological considerations

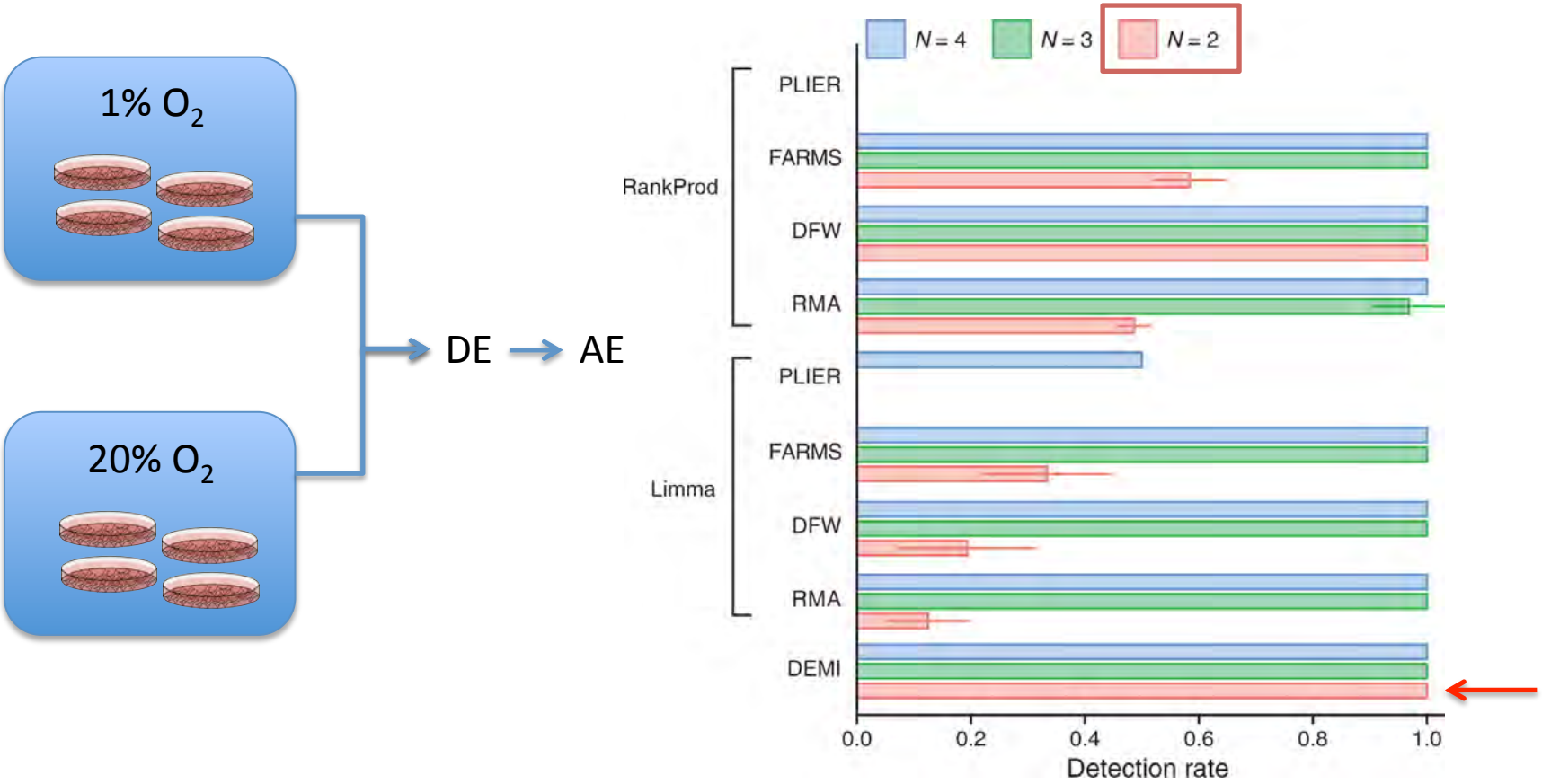
- Can we rely on DE given 2 replicates?

Distribution of individuals

Distribution of sample averages



Modified from http://www.winspc.com/datanet-enews/2013-4/index_public.php



- DE – differential expression
- AE – functional annotation enrichment

- 'cellular response to hypoxia' (GO:0071456)
- 'glycolysis' (GO:0006096)

Toy example of DEMI algorithm

- target-specific probes: the ratio of differentially expressed probes is 0.6 (9 out of 15)

	Sample A			Sample B						
	A1	A2	A3	B1	B2	B3	P_{H0}(H1: A<B)	Sig?	P_{H0}(H1: A>B)	Sig?
probe1	65.22	86.83	54.44	32.98	20.58	7.23	1	*	0.05	*
probe2	52.56	51.05	69.14	51.36	33.36	49.29	0.95	*	0.1	*
probe3	89.08	79.54	82.64	48.05	31.31	35.37	1	*	0.05	*
probe4	69.89	52.76	43.46	56.88	42.97	26.37	0.9	*	0.2	*
probe5	55.24	60.13	78.06	13.74	58.95	51.37	0.95	*	0.1	*
probe6	55.19	64.25	71.02	45.08	43.12	24.52	1	*	0.05	*
probe7	78.74	54.86	70.79	50.01	23.56	9.83	1	*	0.05	*
probe8	59.80	49.00	65.52	27.50	42.84	32.73	1	*	0.05	*
probe9	39.12	68.67	74.74	13.46	33.44	21.85	1	*	0.05	*
probe10	62.38	66.24	39.32	44.45	24.89	7.35	0.95	*	0.1	*
probe11	56.86	56.30	61.76	19.32	4.48	46.94	1	*	0.05	*
probe12	63.23	58.10	53.91	57.07	18.19	29.25	0.95	*	0.1	*
probe13	40.97	59.80	79.76	97.36	26.37	56.28	0.65	*	0.5	*
probe14	111.45	81.71	65.14	32.41	38.04	47.96	1	*	0.05	*
probe15	65.11	64.04	70.98	47.76	43.32	15.77	1	*	0.05	*
							Ratio	**0**		**0.6**

Toy example of DEMI algorithm

- off-target probes (the background): the ratio of differentially expressed probes is around 0.05 (expected when there is no systematic difference between A and B)

	Sample A			Sample B						
	A1	A2	A3	B1	B2	B3	P_{H0}(H1: A<B)	Sig?	P_{H0}(H1: A>B)	Sig?
probe1	66.74	53.14	54.84	55.18	42.45	47.84	0.9	*	0.2	*
probe2	35.94	59.95	46.56	51.26	63.87	65.81	0.1	*	0.95	*
probe3	26.35	56.41	54.89	57.93	56.66	34.10	0.2	*	0.9	*
probe4	87.84	54.04	44.22	50.96	38.60	70.50	0.8	*	0.35	*
probe5	55.63	52.17	56.86	70.28	34.03	35.13	0.8	*	0.35	*
probe6	75.81	51.80	66.25	45.52	63.09	39.66	0.95	*	0.1	*
probe7	34.98	51.21	50.61	40.97	33.54	63.84	0.65	*	0.5	*
probe8	62.62	53.43	59.59	34.08	45.23	37.63	1	*	0.05	*
probe9	86.47	88.12	50.31	44.60	53.09	43.84	0.95	*	0.1	*
probe10	66.57	55.94	54.54	36.99	47.68	50.44	1	*	0.05	*
probe996	46.28	31.91	36.94	44.68	47.73	62.60	0.1	*	0.95	*
probe997	63.09	75.89	45.91	64.15	42.37	62.86	0.8	*	0.35	*
probe998	27.49	50.75	38.21	35.78	13.33	50.33	0.8	*	0.35	*
probe999	61.14	66.20	34.56	50.81	12.44	50.47	0.9	*	0.2	*
probe1000	46.10	27.95	33.43	36.14	33.11	67.12	0.35	*	0.8	*
							Ratio 0.049			**0.052**

Toy example of DEMI algorithm

	H1: A < B	H1: A > B
Target (ratio)	0	0.6
Background (ratio)	0.049	0.052
P_{H0}(H1: ratio_T > ratio_B)	1	2.44E-09

- Q: What is the H_0-probability of observing a ratio of target-specific DE probes >= 0.6 when the background ratio is 0.052?
- A: 2.44E-09 (hypergeometric probability distribution)

Methods

- Arrays
  - Rat230_2
    - 157,540 probes
    - hits mapping to 14,195 rat genes
  - HG-U133_Plus_2
    - 395,023 probes
    - hits mapping to 24,809 human genes

Methods

- Performance evaluation
  - 10-times 4-fold CV
  - Matthews correlation coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

Methods

- Classifier
  - 3-layered neural network (50:10:1)

- Feature (-1..0..1): $FDR_{H1: A<B} - FDR_{H1: A>B}$

-1	expression of target X is lower in A
0	expression of target X is similar in A and B
1	expression of target X is higher in A

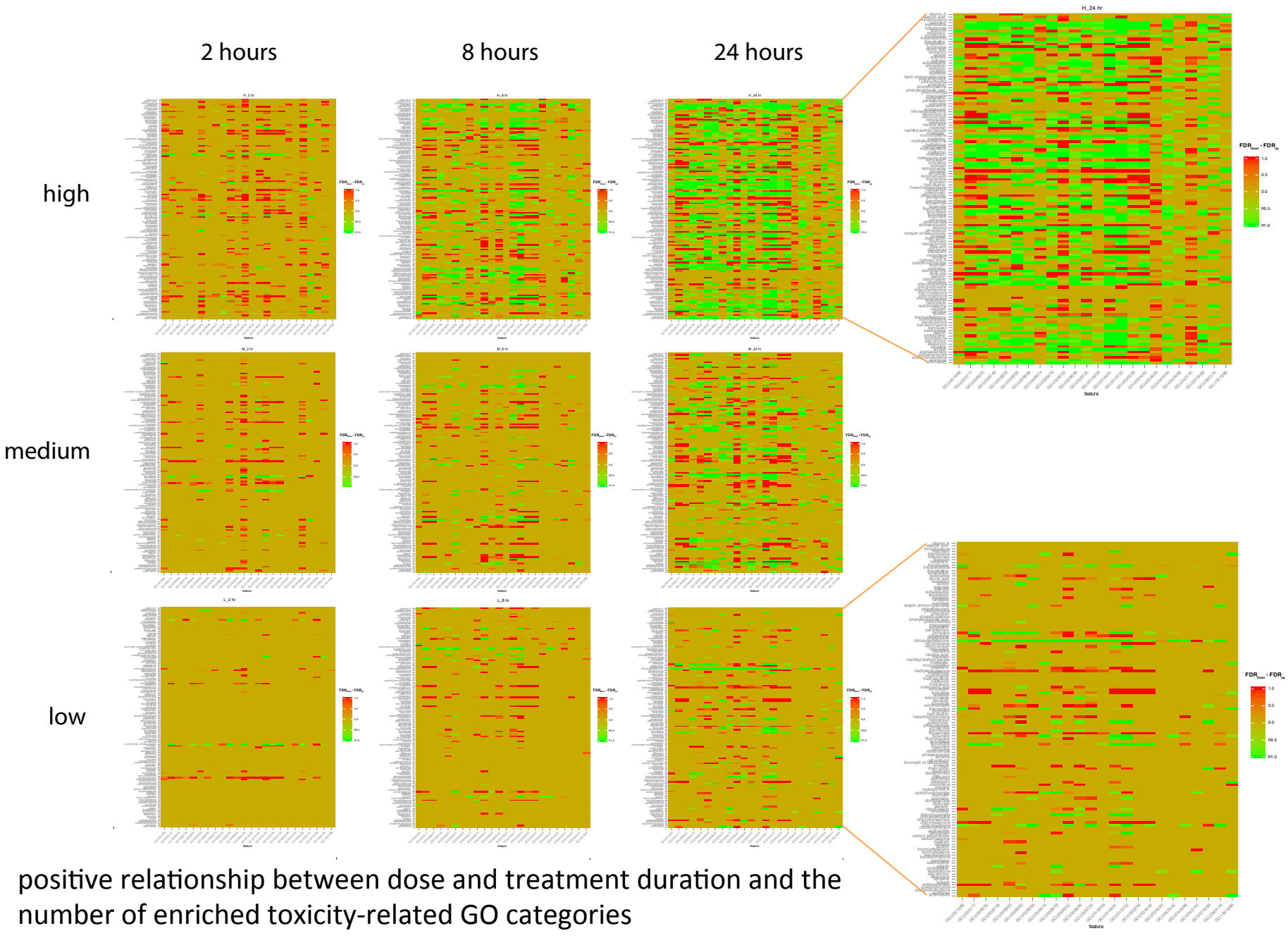
Methods

- Feature space
  - #dose * #timepoint * #gene
  - e.g. 3 * 3 * 14,195 = 127,755 features
- Filtering
  - keep 5% of features with highest variance from each treatment-reference pair
- Selection
  - recursive feature elimination
  - 50 features

Results

- Proof-of-concept: is treatment effect on differential expression more potent at higher doses and longer treatment periods?

- Dose and treatment duration correlate with the no. of enriched toxicity-related GO categories



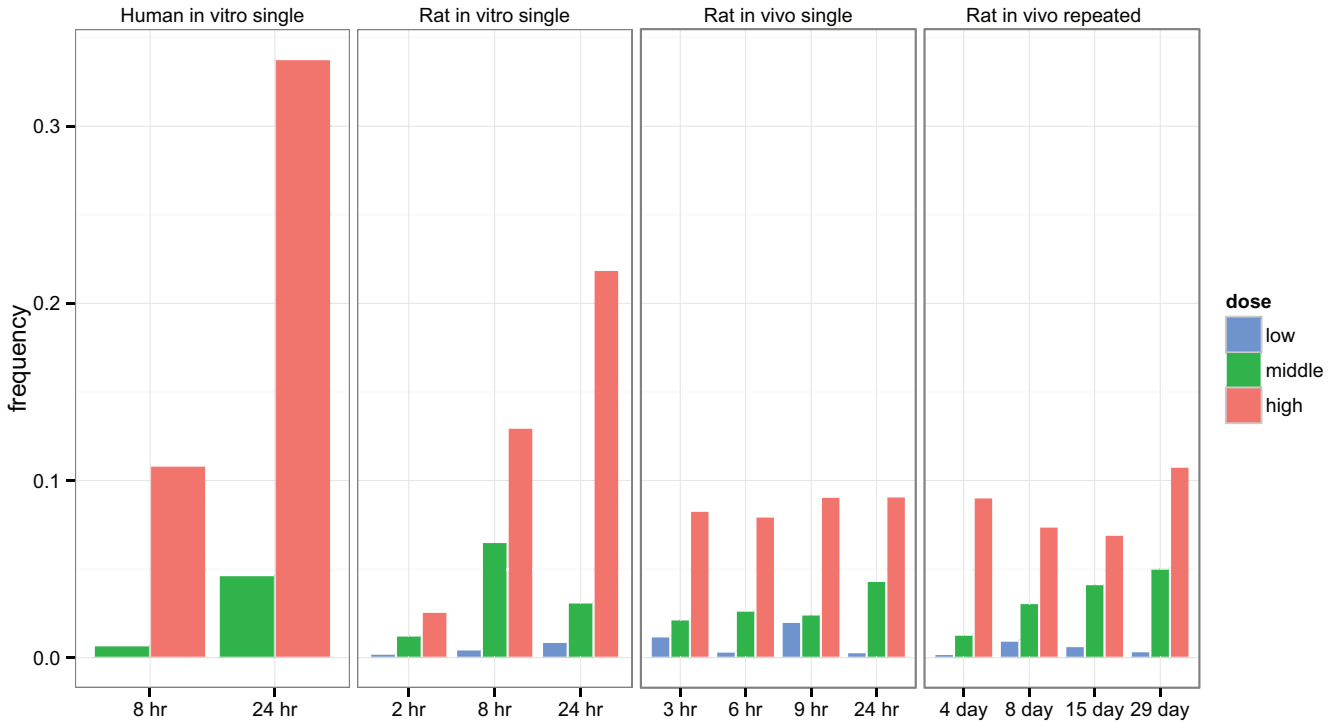
positive relationship between dose and treatment duration and the number of enriched toxicity-related GO categories

Toxicity-related categories in Gene Ontology

accession	name
GO:0001666	response to hypoxia
GO:0005777	peroxisome
GO:0006635	fatty acid beta-oxidation
GO:0006749	glutathione metabolic process
GO:0006805	xenobiotic metabolic process
GO:0006955	immune response
GO:0006956	complement activation
GO:0006974	response to DNA damage stimulus
GO:0006979	response to oxidative stress
GO:0008202	steroid metabolic process
GO:0009636	response to toxic substance
GO:0009725	response to hormone stimulus
GO:0016209	antioxidant activity
GO:0016491	oxidoreductase activity
GO:0019752	carboxylic acid metabolic process
GO:0030258	lipid modification
GO:0034440	lipid oxidation
GO:0034620	cellular response to unfolded protein
GO:0042440	pigment metabolic process
GO:0043068	positive regulation of programmed cell death
GO:0043161	proteasomal ubiquitin-dependent protein catabolic process
GO:0070469	respiratory chain
GO:0080134	regulation of response to stress
GO:1901988	negative regulation of cell cycle phase transition

based on conceptual relevance and variance of enrichment significance in JTGP data

Frequency of treatments among RFE-selected features



- The more frequent the treatment, the more informative were the features for distinguishing nontoxic compounds from compounds with high DILI potential based on RFE

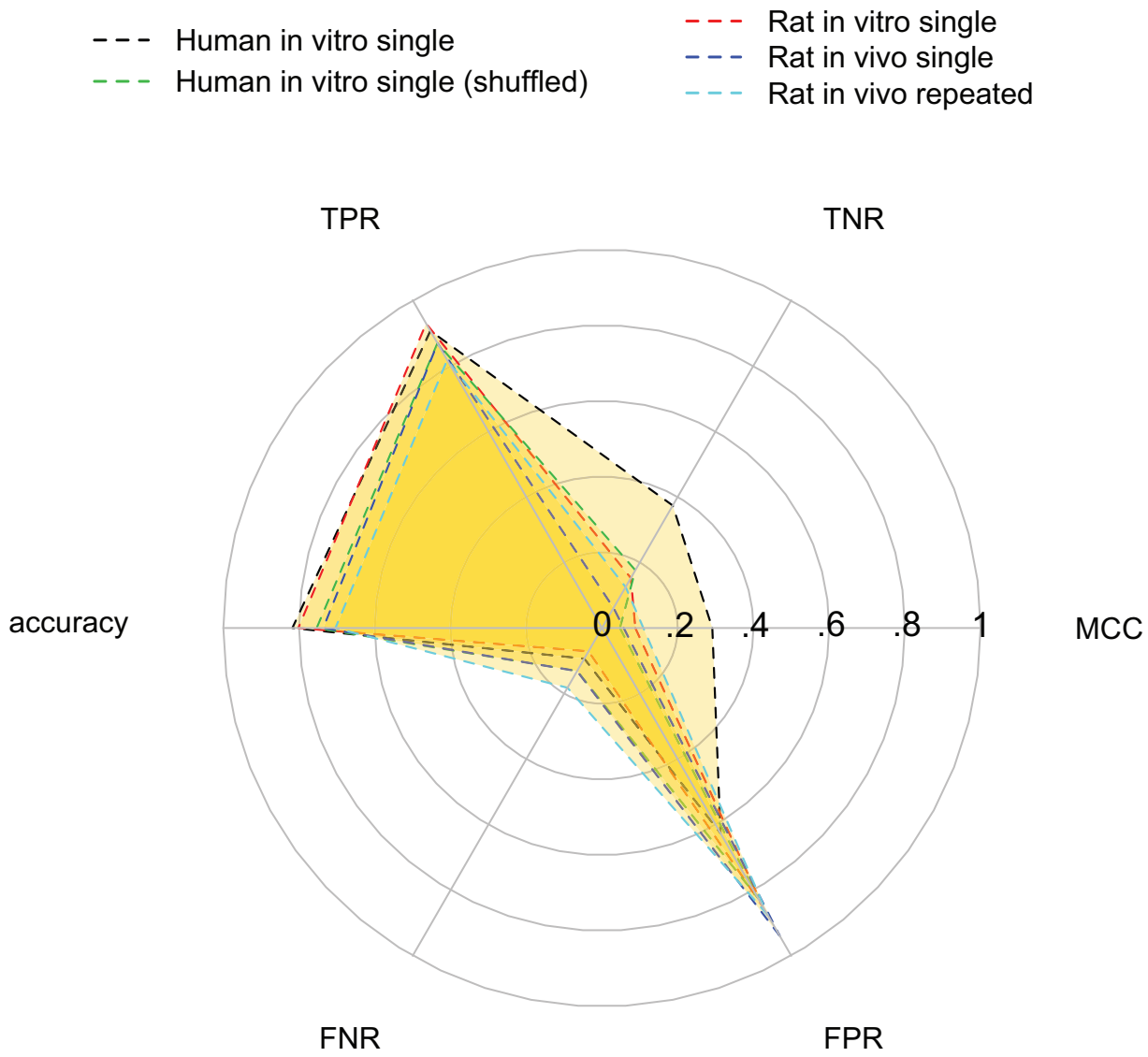
Results

- Differential expression from low dose and short treatment period is less informative than higher doses and longer periods

Results

- Classification performance

	Training		Validation	
	mean	**95% ci**	**mean**	**95% ci**
Human in vitro single	0.936	0.006	0.2914	0.0355
Human in vitro single (shuffled)	0.904	0.007	0.0474	0.0272
Rat in vitro single	0.984	0.004	0.0881	0.0292
Rat in vivo single	0.912	0.018	-0.0685	0.0172
Rat in vivo repeated	0.771	0.038	-0.0593	0.0135


Human in vitro single
Human in vitro single (shuffled)
Rat in vitro single
Rat in vivo single
Rat in vivo repeated
TPR
TNR
accuracy
0
.2
.4
.6
.8
1
MCC
FNR
FPR

Discussion

- DE appears to be informative w.r.t. DILI
- distinguishing responses to nontoxic and toxic agents non-trivial
- classification performance higher in simpler system (presumably less biological and technical var)

Discussion

- Low dose and short treatment period are less informative than higher doses and longer periods

- Human arrays yield more gene-level features – suggesting the use of human cell lines for in vitro studies

Acknowledgements

- Sten Ilmjärv - DEMI R package
- Xpressomics
- BIIT group (Jaak Vilo & co.)
- Estonian research foundation (ETF9099)