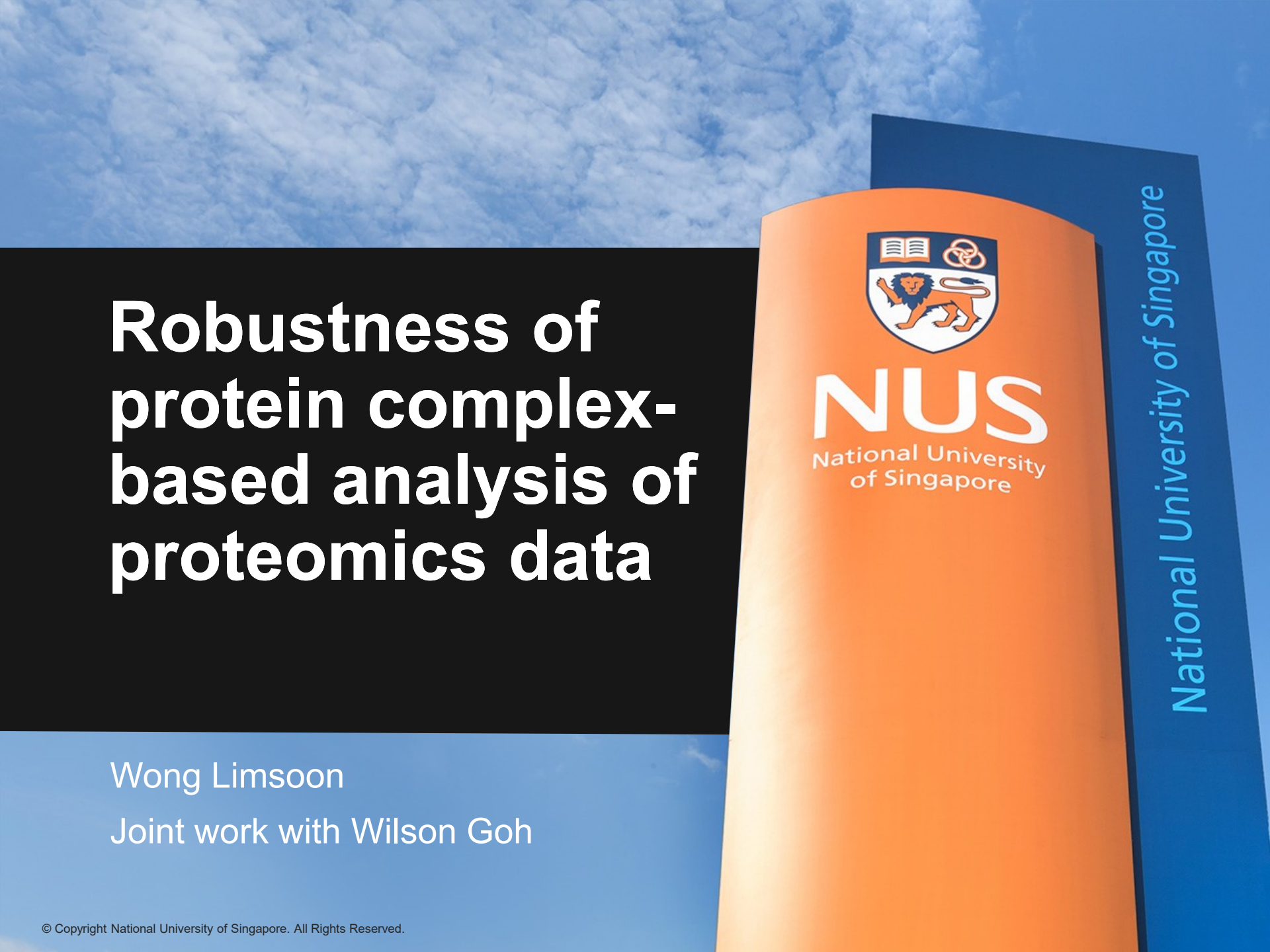


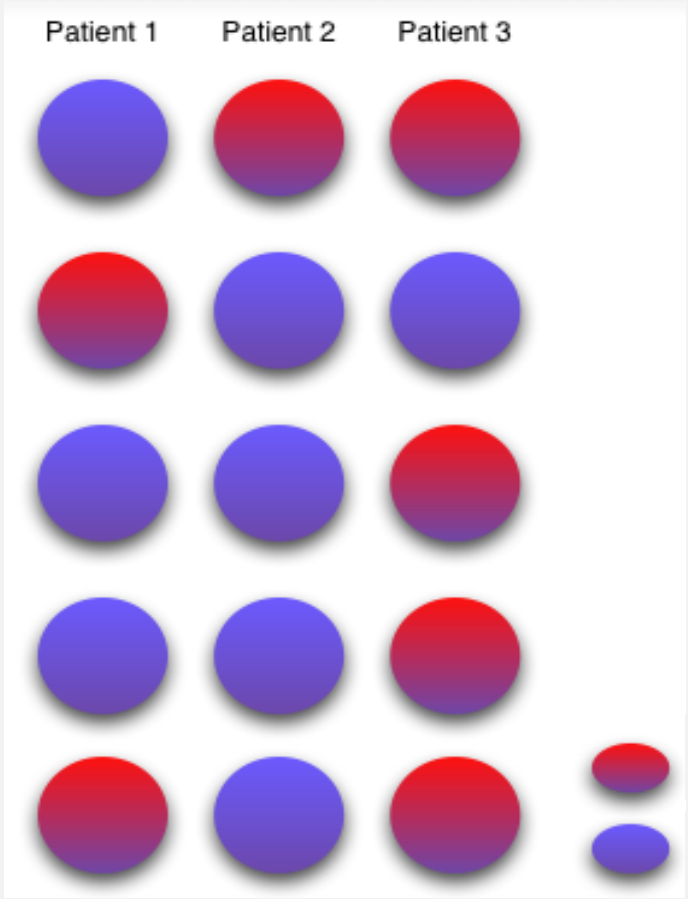
Robustness of protein complex-based analysis of proteomics data

Wong Limsoon

Joint work with Wilson Goh

Using protein complexes to enhance proteomics: Basic ideas

Intuition



Suppose the failure to form a protein complex causes a disease

If any component protein is missing, the complex can't form

Diff patients suffering from the disease can have a diff protein component missing

Construct a profile based on complexes?

… and some math

Postulate: Chance of a protein complex being present ≈ fraction of its member proteins being reported in the screen

Say a proteomics screen has 75% reliability; {A, B, C, D, E} is a complex; and screen reports A, B, C, D only

The complex has 60% (= 0.75 * 4 / 5) chance to be present

E has >60% chance to be present, as presence of complex implies presence of its constituents … improving coverage

A, B, C, and D each has 90% (= 100% * 0.6 + 75% * 0.4) chance of being present, i.e. >75% … removing noise

Reference complexes

In this talk, human complexes (of size at least 5) from CORUM are used as reference complexes

It is possible to use subnetworks generated from pathway and PPI databases. However these such subnetworks vary significantly depending on databases and generation algorithms used

So I do not consider these…

Improving consistency in proteomic profile analysis

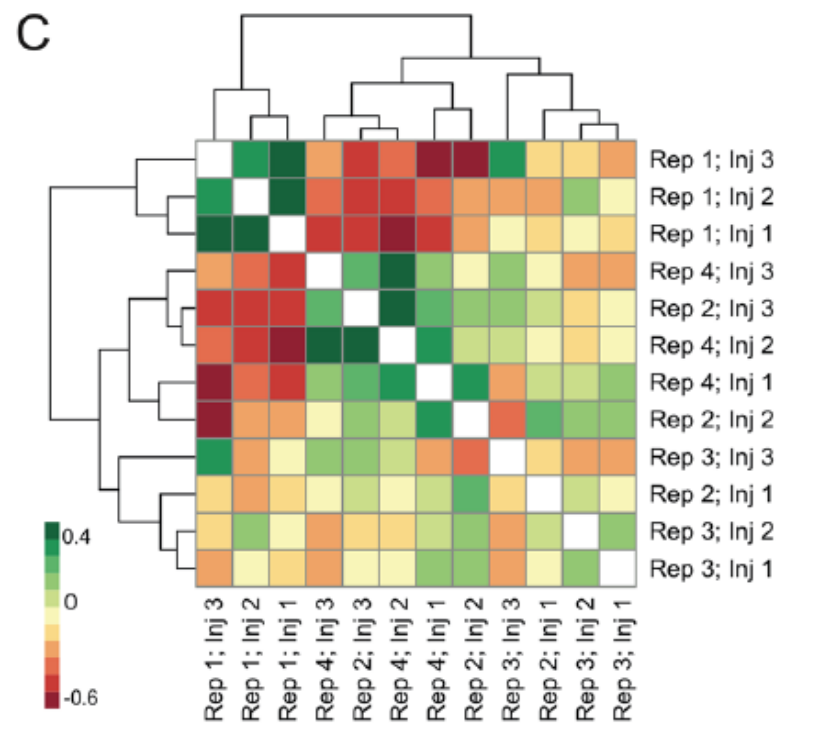
Proteomic profiles sometimes not sufficiently consistent

A human kidney tissue, digested in quadruplicates, analyzed in triplicates

Guo et al. *Nature Medicine*, 21(4):407-413, 2015

Correlation betw replicates is not good (~0.4)

Technical replicates of the same biological replicate are not tightly clustered



qPSP, a simple use of complexes

Features are complexes

Feature values are fuzzy weighted proportion of proteins in a complex

$$score(C,S_i) = \Sigma_{p \in C} fs(p,S_i) / |C|$$

Complex C is significant if $\{score(C,S_i) \mid S_i \in A\}$ is very different by t-test from $\{score(C,S_i) \mid S_i \in B\}$

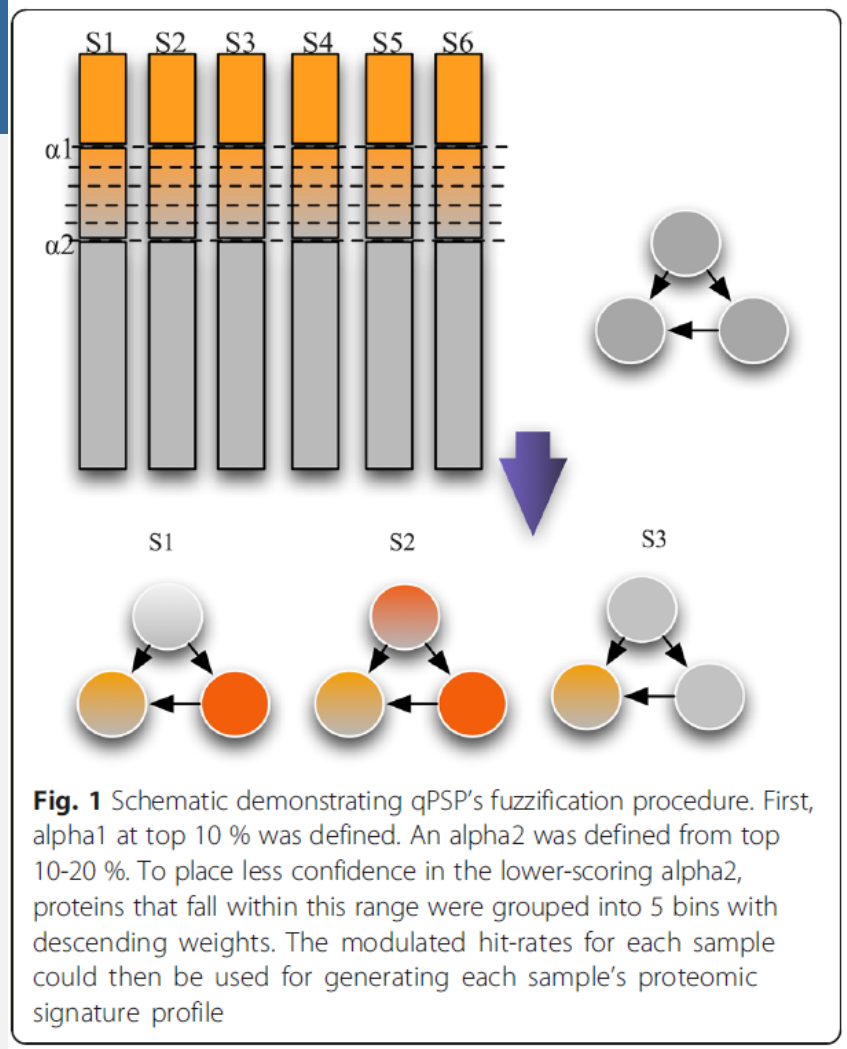


Fig. 1 Schematic demonstrating qPSP's fuzzification procedure. First, alpha1 at top 10 % was defined. An alpha2 was defined from top 10-20 %. To place less confidence in the lower-scoring alpha2, proteins that fall within this range were grouped into 5 bins with descending weights. The modulated hit-rates for each sample could then be used for generating each sample's proteomic signature profile

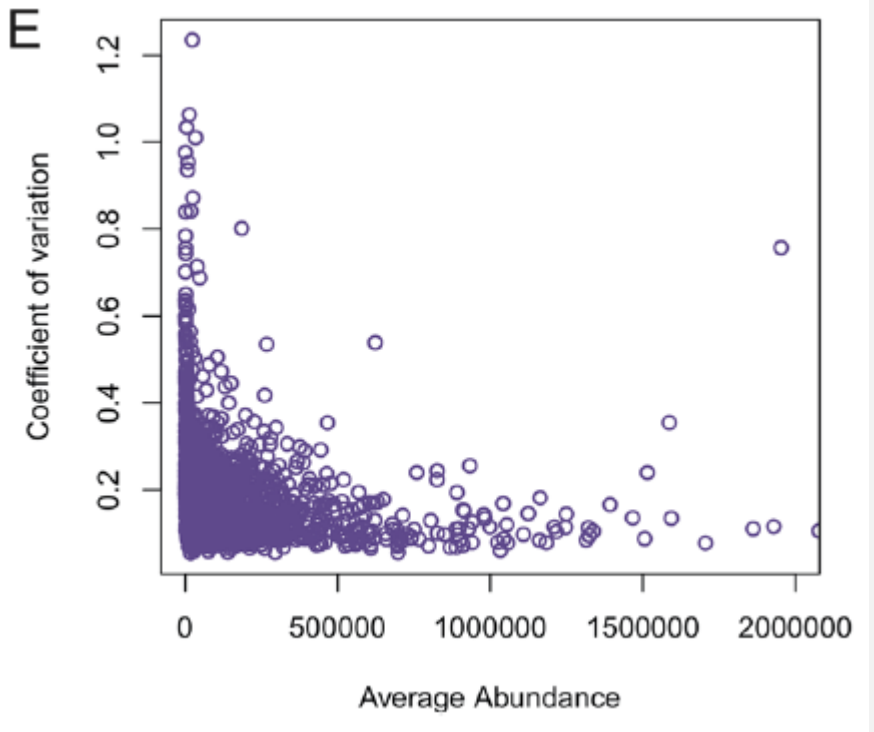
Goh et al. **Quantitative proteomics signature profiling based on network contextualization**. *Biology Direct*, 10:71. 2015

Justification for fuzzy scoring

Low-abundance proteins have very high coefficient of variation

They thus are very noisy

Fuzzy scoring mitigates this



False-positive rate analysis

12 kidney controls randomly assigned into two groups of equal size, and qPSP performed many rounds

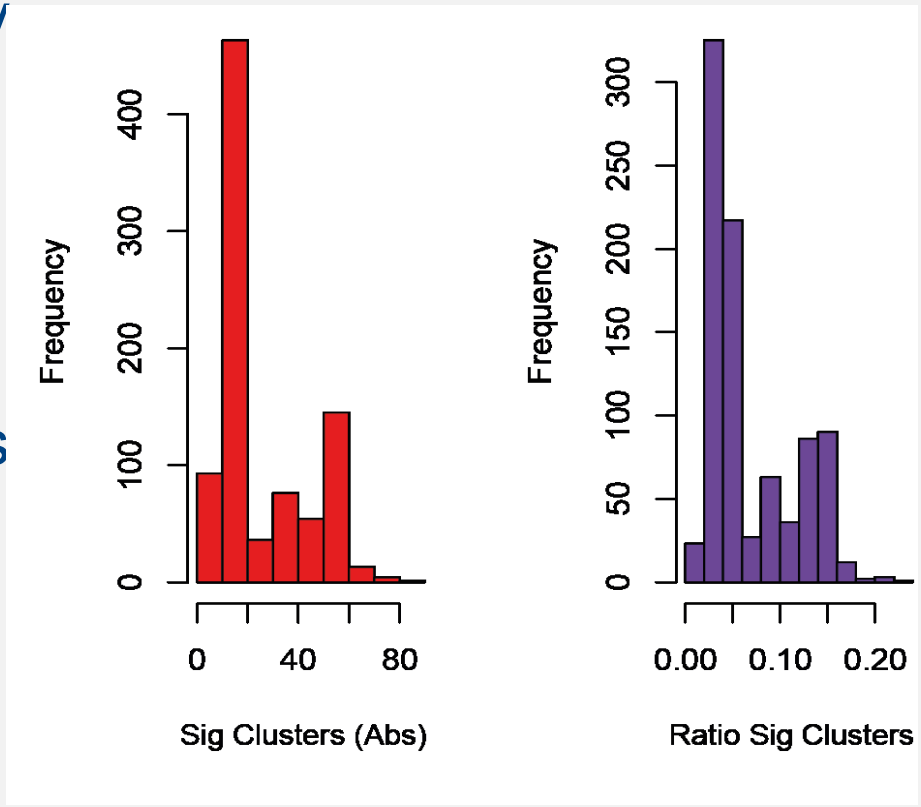
False-positive rate well within the expectation levels

Sig Clusters (Abs)

Exp: 19, Observed: 16

Sig Clusters (Ratio)

Exp: 5%, Observed: 4%



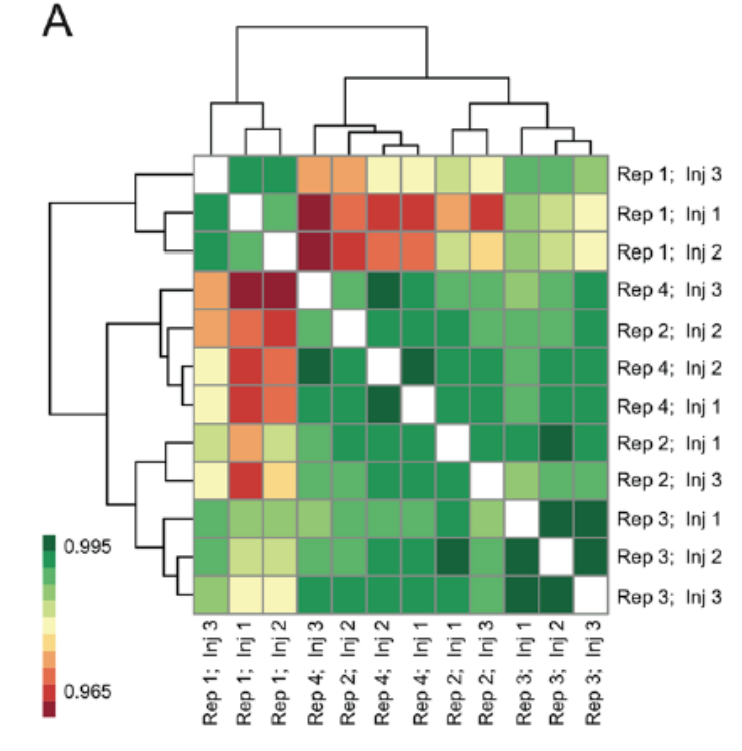
Consistency of qPSP

Clustering of benchmarking control data based on protein complexes (i.e. qPSP)

Correlation betw replicates is >0.95

Cf. 0.4 based on proteins

Also, technical replicates are better clustered



Application to renal & colorectal cancers

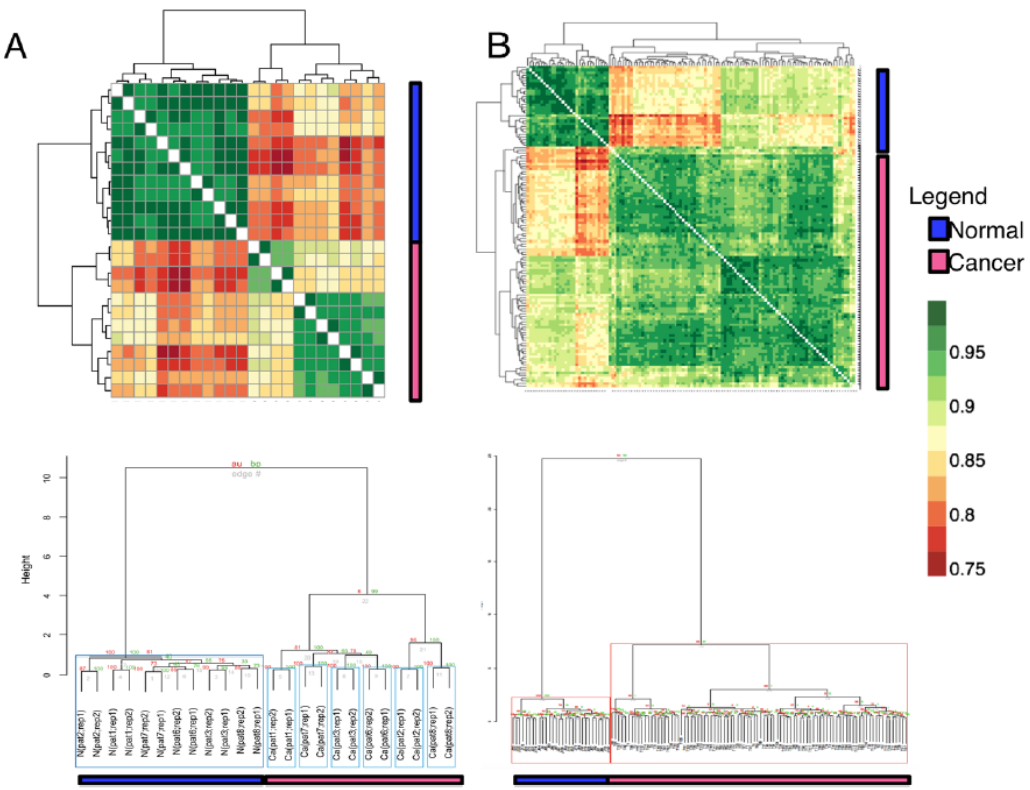


Fig. 3 qPSP strongly discriminates sample classes for renal cancer (**a**) and colorectal cancer (**b**). Clustered similarity maps at the top row showed specific and consistent segregation of non-cancer and cancer samples. The trees below the heatmaps are from bootstrap analysis (PVCLUST), which demonstrates that the discrimination between sample classes based on qPSP hit-rates is highly stable

Further improving consistency in proteomic profile analysis

PFSNet, considering class factors

Wt of gene i in +ve class

$$\beta_1^*(g_i) = \sum_{p_j \in D} \frac{fs(e_{g_i,p_j})}{|D|}$$

$$\beta_2^*(g_i) = \sum_{p_j \in \neg D} \frac{fs(e_{g_i,p_j})}{|\neg D|}$$

-ve class wt

Score of subnet S in patient k w/ +ve class

$$Score_1^{p_k}(S) = \sum_{g_i \in S} fs(e_{g_i,p_k}) * \beta_1^*(g_i)$$

$$Score_2^{p_k}(S) = \sum_{g_i \in S} fs(e_{g_i,p_k}) * \beta_2^*(g_i)$$

For irrelevant protein complex S, the two scores above for patient p_k should be roughly equal, regardless of class

$\therefore$ Do a paired t-test on $\Delta(p_k, S) = score_1^{pk}(S) - score_2^{pk}(S)$

to decide whether S is relevant, get null distribution by permuting class labels

Goh & Wong. **Advancing clinical proteomics via analysis based on biological complexes: A tale of five paradigms**. *Journal of Proteome Research*, 15(9):3167--3179, July 2016.

ESSNet, for small datasets & low-abundance proteins

Let g_i be a protein in a given protein complex

Let p_j be a patient

Let q_k be a normal

Let $\Delta_{i,j,k} = Expr(g_i,p_j) – Expr(g_i,q_k)$

Test whether $\Delta_{i,j,k}$ is a distribution with mean 0

Goh & Wong. **Advancing clinical proteomics via analysis based on biological complexes: A tale of five paradigms**. *Journal of Proteome Research*, 15(9):3167--3179, 2016

Null hypothesis is “Complex C is irrelevant to the diff betw patients and normals, and the proteins in C behave similarly in patients and normals”

No restriction to most abundant proteins

Potential to reliably detect low-abundance but differential proteins

Methods to compare with

Network-based methods

Hypergeometric enrichment (HE)

Direct group analysis (DG), similar to GSEA

qPSP

PFSNET

ESSNET

Standard t-test on individual proteins (SP)

Simulated data

Simulated datasets from Langley and Mayr

Both D1.2 and D2.2 have 100 small-sized simulated datasets, each with 20% significant features

Effect sizes of these differential features are sampled from one out of five possibilities (20%, 50%, 80%, 100% and 200%), increased in one class and not in the other

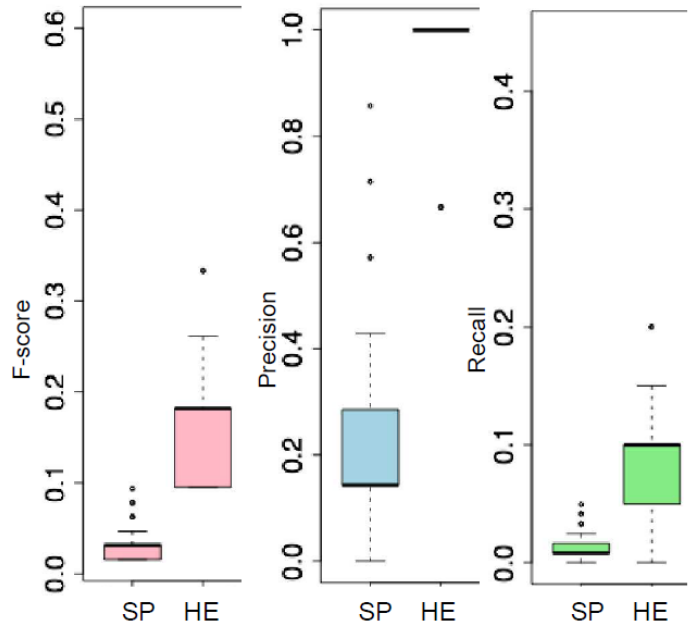
Significant artificial complexes are constructed with various level of purity (i.e. proportion of significant proteins in the complex)

Equal # of non-significant complexes are constructed too

Langley & Mayr, *J. Proteomics*, 129:83-92, 2015

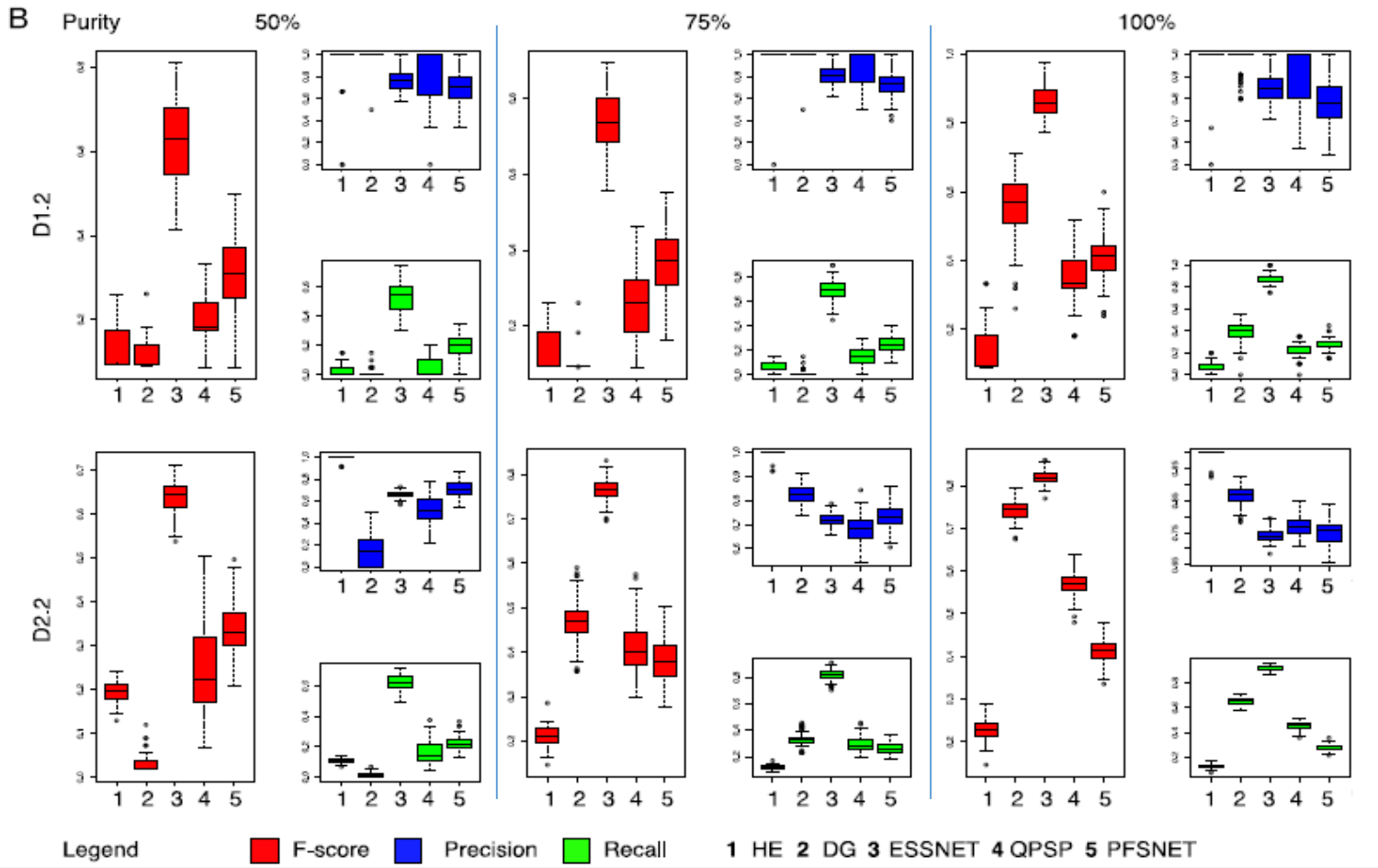
SP shows rather poor performance on simulated data

Can network-based methods do better?



Supplementary Figure 1 Single protein (SP) precision-recall performance on D1.2. The f-score (pink), precision (blue) and recall (green) shows that SP performs abysmally on simulated data. HE is shown next to SP as a reference.

qPSP / PFSNET / ESSNET perform well on simulated data



Renal cancer control data (RCC)

12 runs originating from a human kidney tissue digested in quadruplicates and analyzed in triplicates

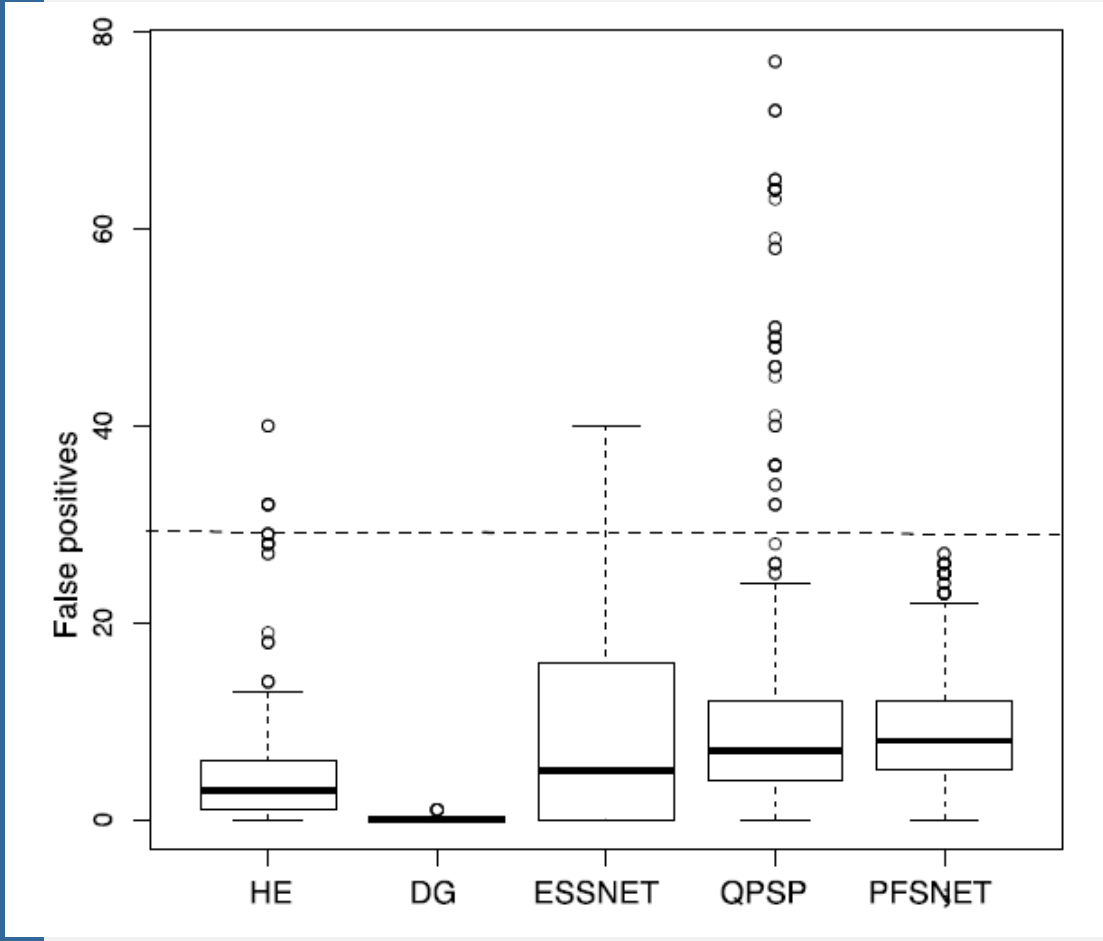
Excellent for evaluating false-positive rates of feature-selection methods

Randomly split the 12 runs into two groups

Report of any significant features between the groups must be false positives

Guo et al. *Nature Medicine*, 21(4):407-413, 2015

All methods control false positives well



Dash line corresponds to expected # of false positives at alpha 0.05 (~30 complexes)

Renal cancer data (RC)

12 samples are run twice so we have technical replicates over 6 normal and 6 cancer tissues

Good for testing reproducibility of feature-selection methods

A good method should report similar feature sets between replicates

Can also test feature-selection stability

Apply feature-selection method on subsamples and see whether the same features get selected

Guo et al. *Nature Medicine*, 21(4):407-413, 2015

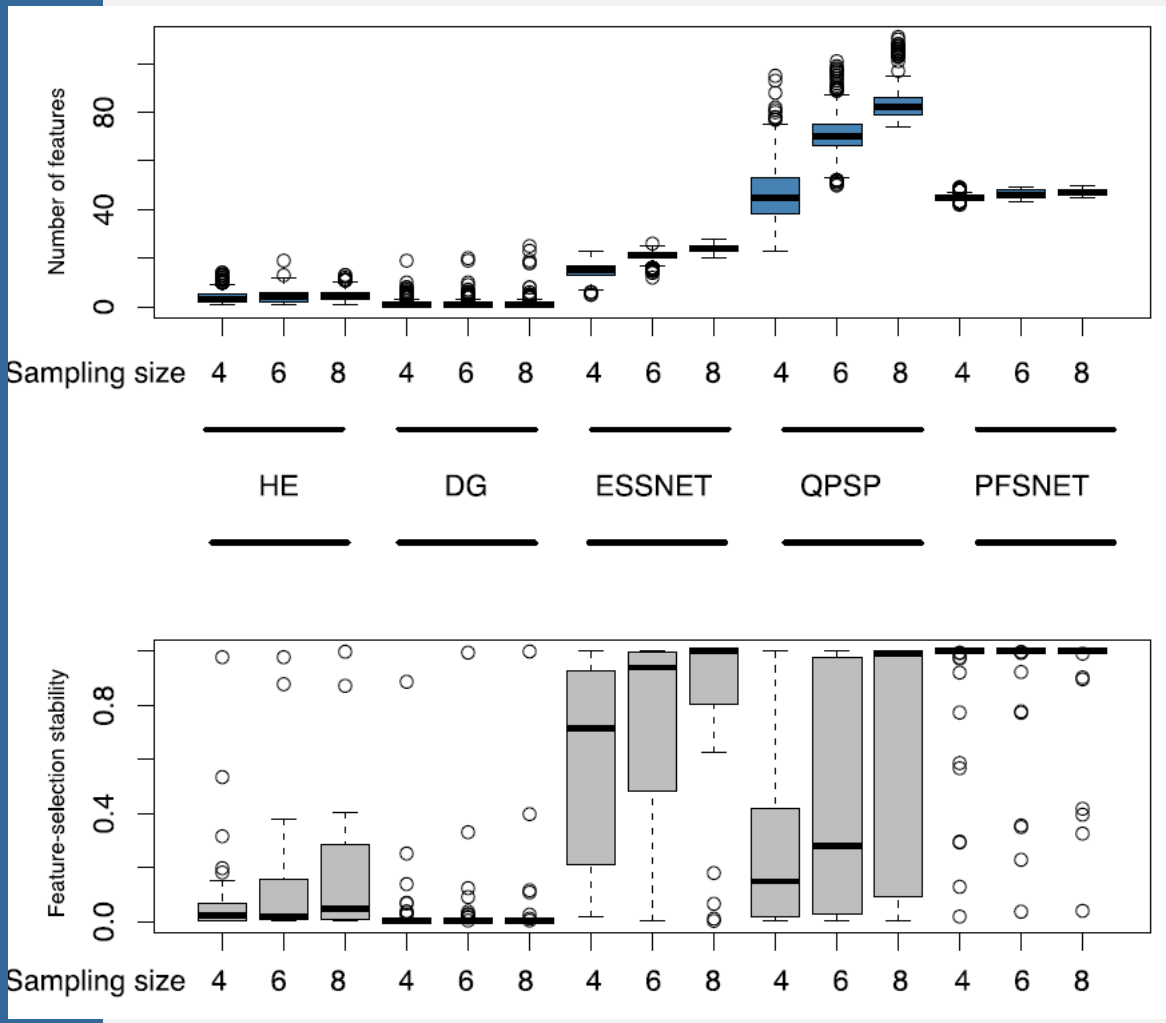
ESSNET & PFSNET show excellent reproducibility

Number of terms	HE	DG	ESSNET	QPSP	PFSNET
Replicate 1	4	1	35	86	45
Replicate 2	6	2	29	75	46
Overlaps	0.25	0.5	0.83	0.66	0.94

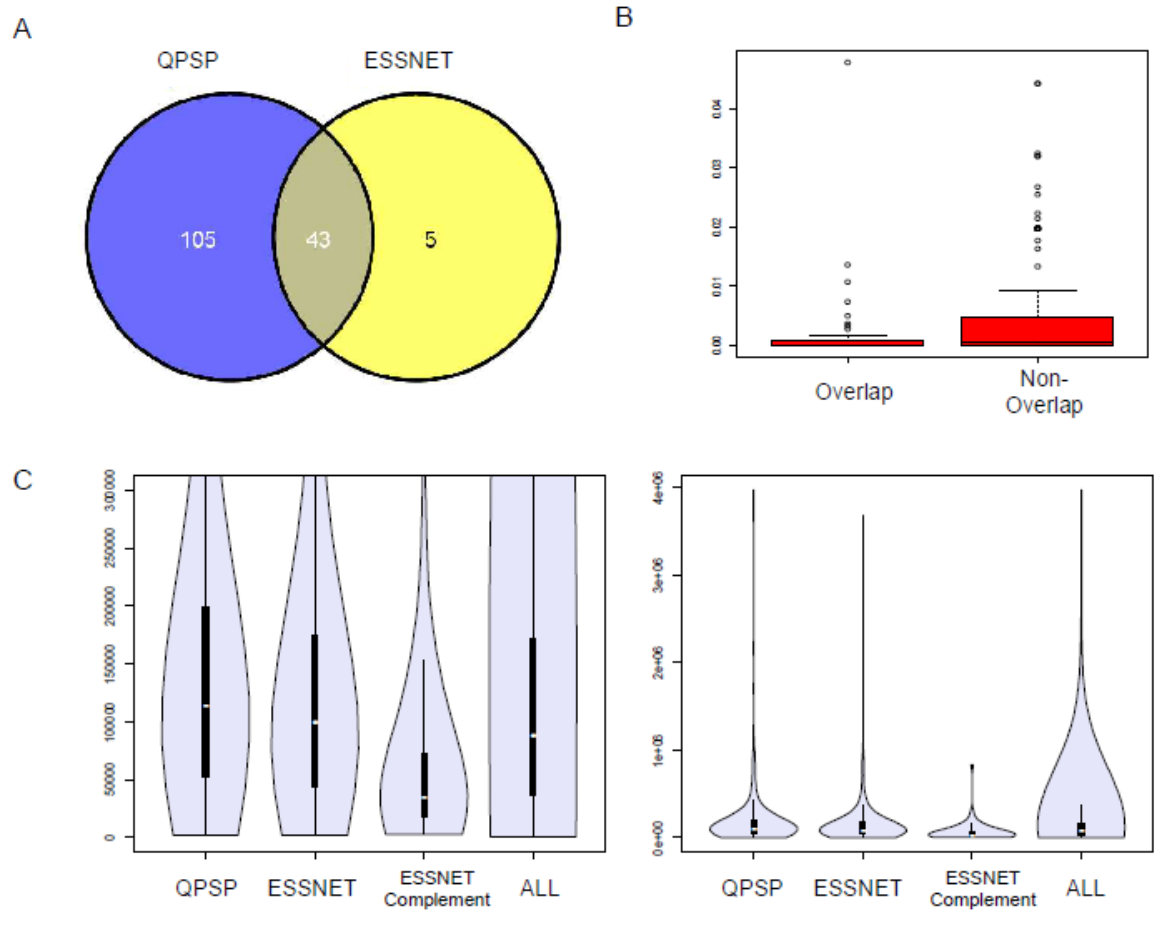
HE	DG	ESSNET	QPSP	PFSNET	
1	0.5	0.71	0.86	0.71	HE
	1	1	1	1	DG
		1	0.93	0.98	ESSNET
			1	0.90	QPSP
				1	PFSNET

This table is computed on by applying the methods on the full RC dataset

ESSNET & PFSNET show excellent stability



ESSNET can assay low-abundance complexes that qPSP cannot

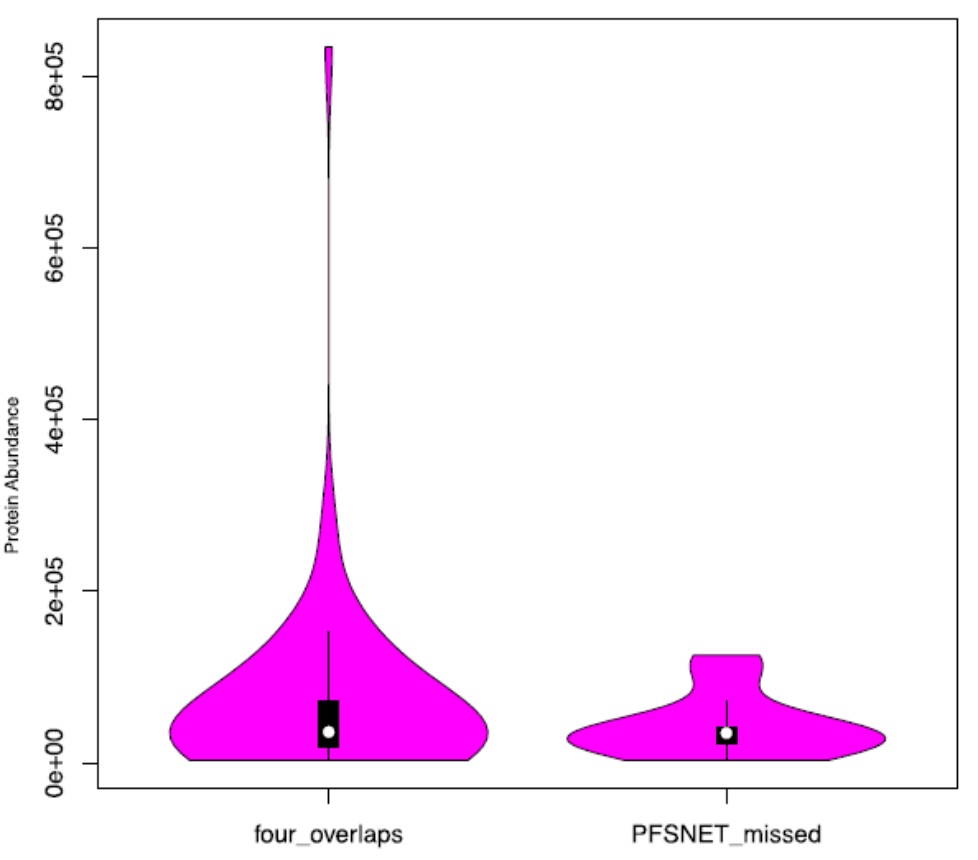


A: QPSP-ESSNET significant-complex overlaps

B: P-value distribution for overlapping and non-overlapping QPSP complexes.

C: Sampling abundance distribution. The left panel is a zoom-in of the right. The y-axis is the protein abundance while the four categories are the distribution of abundances of complexes found in QPSP, ESSNET, ESSNET unique (complement), and all proteins in RC.

ESSNET can assay low-abundance complexes that PFSNET cannot

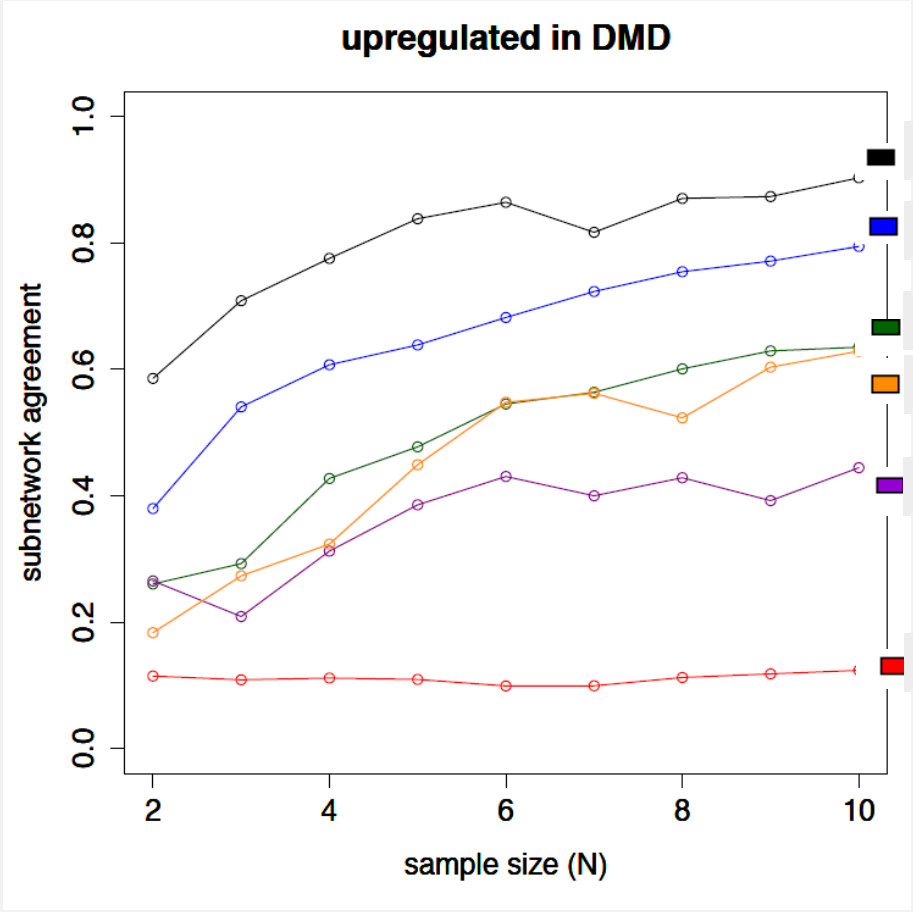


Of the 5 ESSNET-unique complexes, PFSNET can detect 4; the missed complex consists entirely of low-abundance proteins.

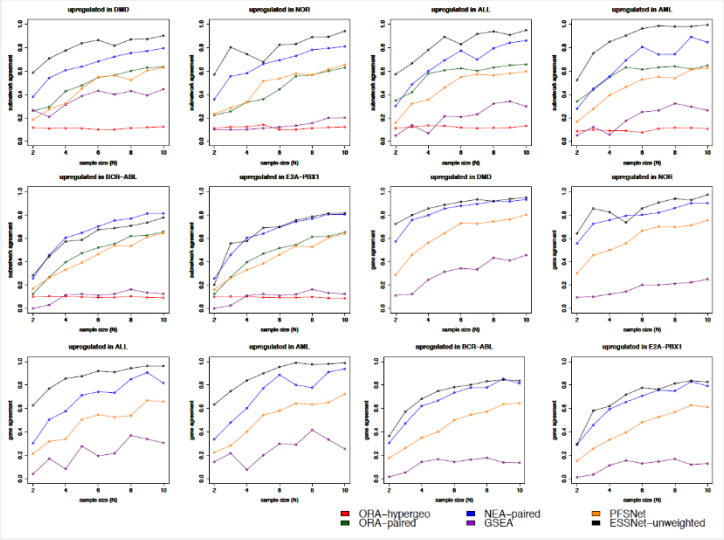
If p-value threshold is adjusted by Benjamini-Hochberg 5% FDR, PFSNET can detect only 3 of the 5 ESSNET-unique complexes while ESSNET continues to detect them all.

These methods work well on gene expression data too, using gene regulatory pathways instead of protein complexes

Reproducibility on gene expression datasets



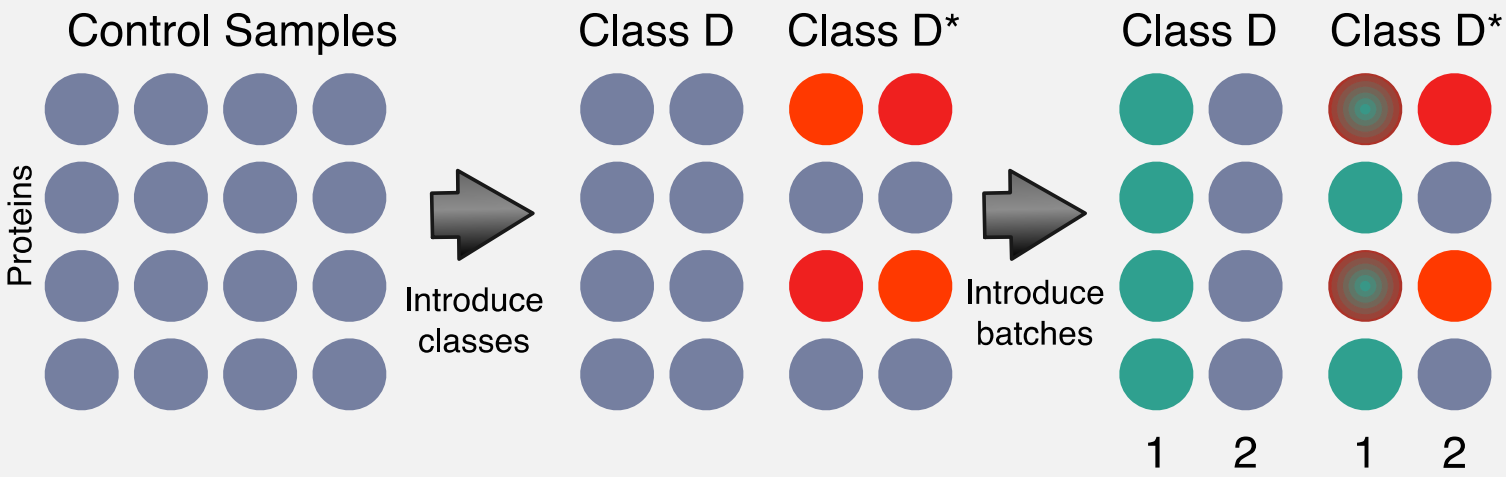
Lim et al. **A quantum leap in the reproducibility, precision, and sensitivity of gene expression profile analysis even when sample size is extremely small**. *Journal of Bioinformatics and Computational Biology*, 13(4):1550018, August 2015.



Batch effects obfuscate –omics analysis

Simulated data

Simulations



Real one-class data from a multiplex experiment (no batches); n = 8

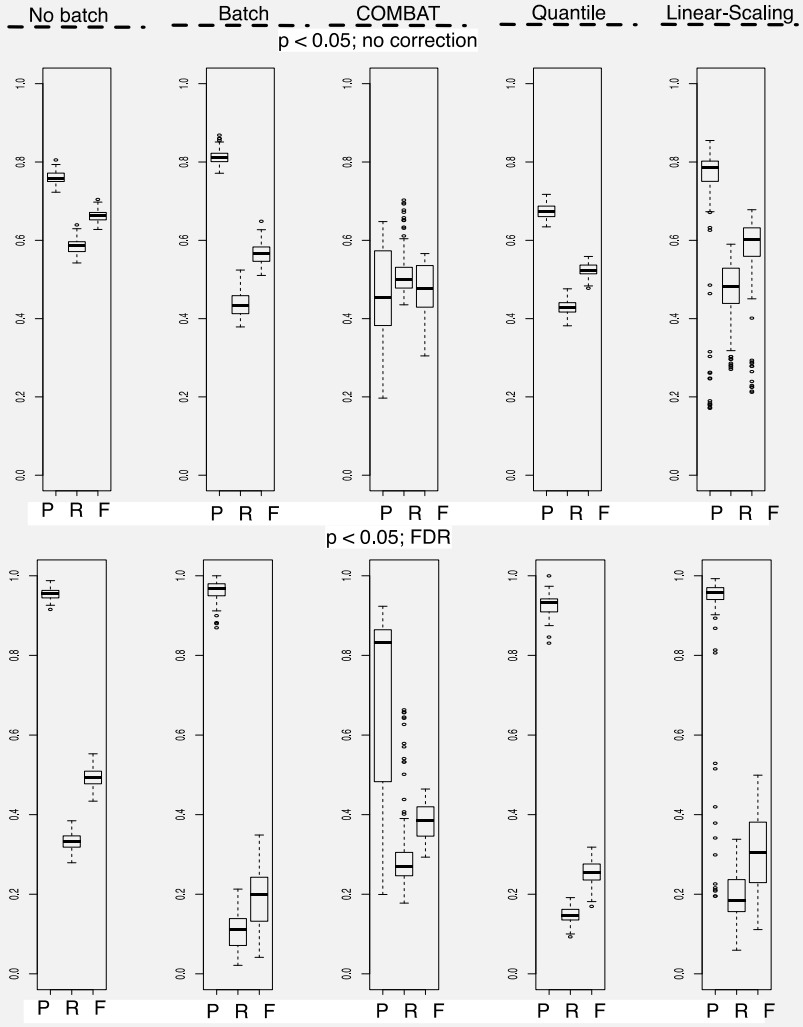
Randomly assigned into two phenotype classes D and D*, 100x

20% biological features are assigned as differential, and a randomly selected effect size (20%, 50%, 80%, 100% and 200%) added to D*

Half of D and D* are assigned to batch 1, and the other half assigned to batch 2. A randomly selected batch effect (20%, 50%, 80%, 100% and 200%) is added to all features in batch 1

Batch effects obfuscate

Batch-effect correction confound

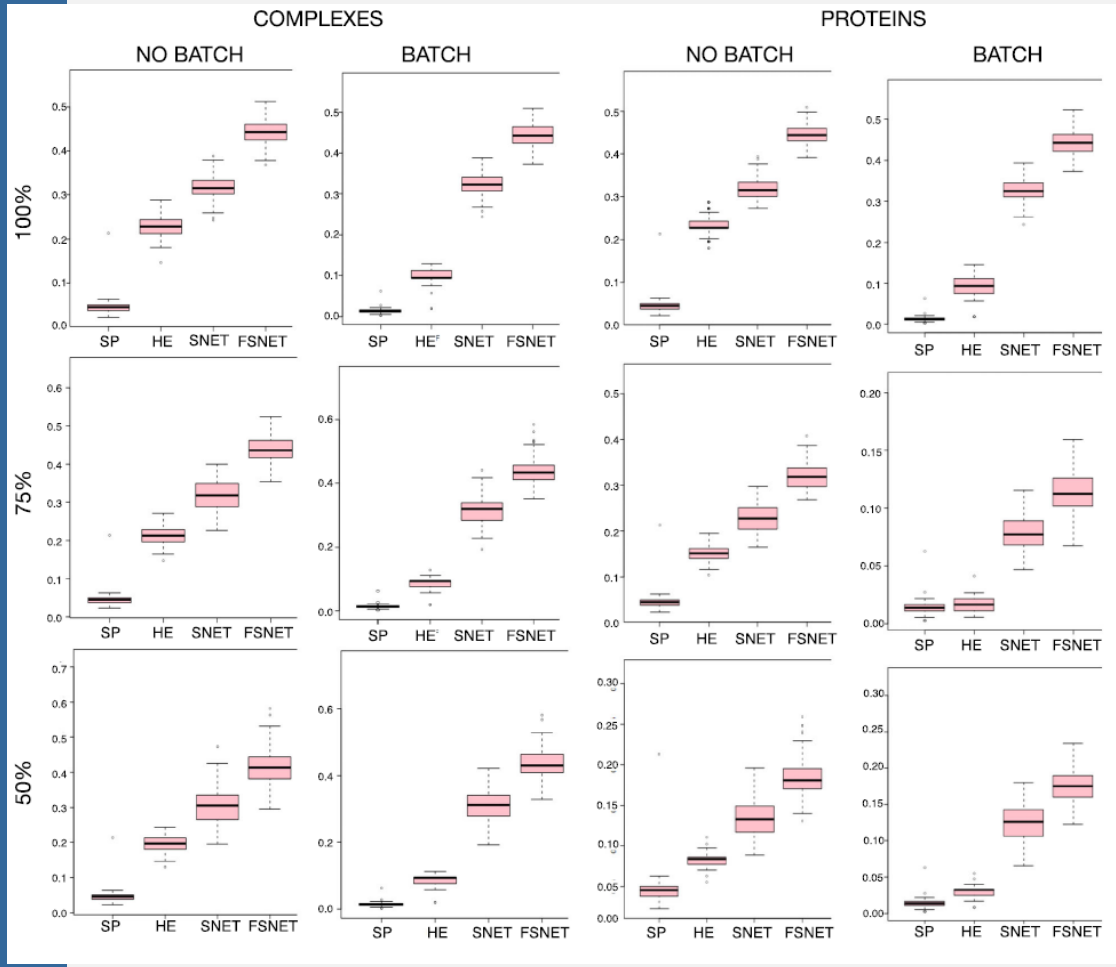


P: Precision R: Recall F: F-measure

Feature selection via t-test

Providing batch-effect resistance in proteomic profile analysis

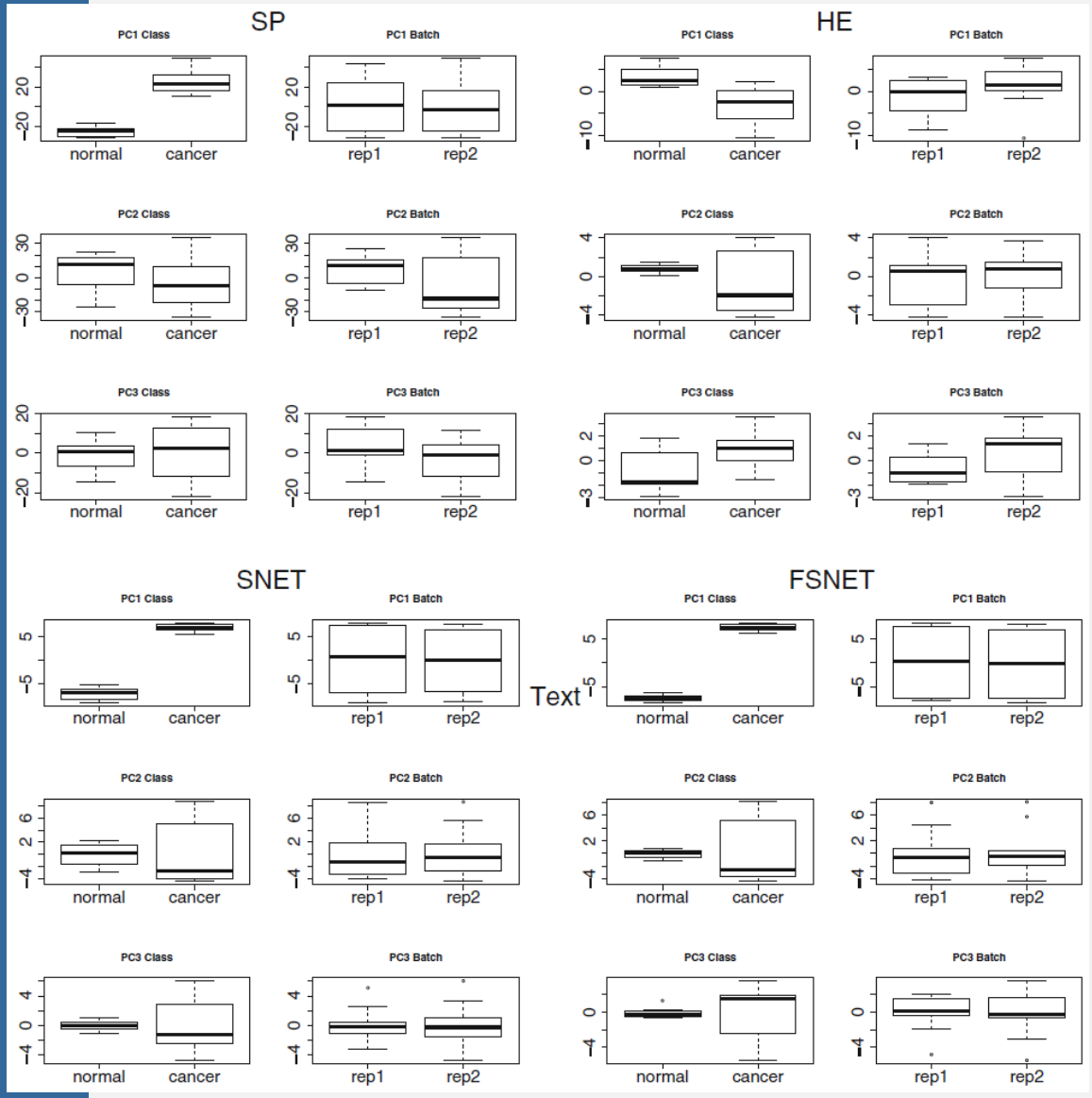
Protein complex-based feature selection is resistant to batch effects



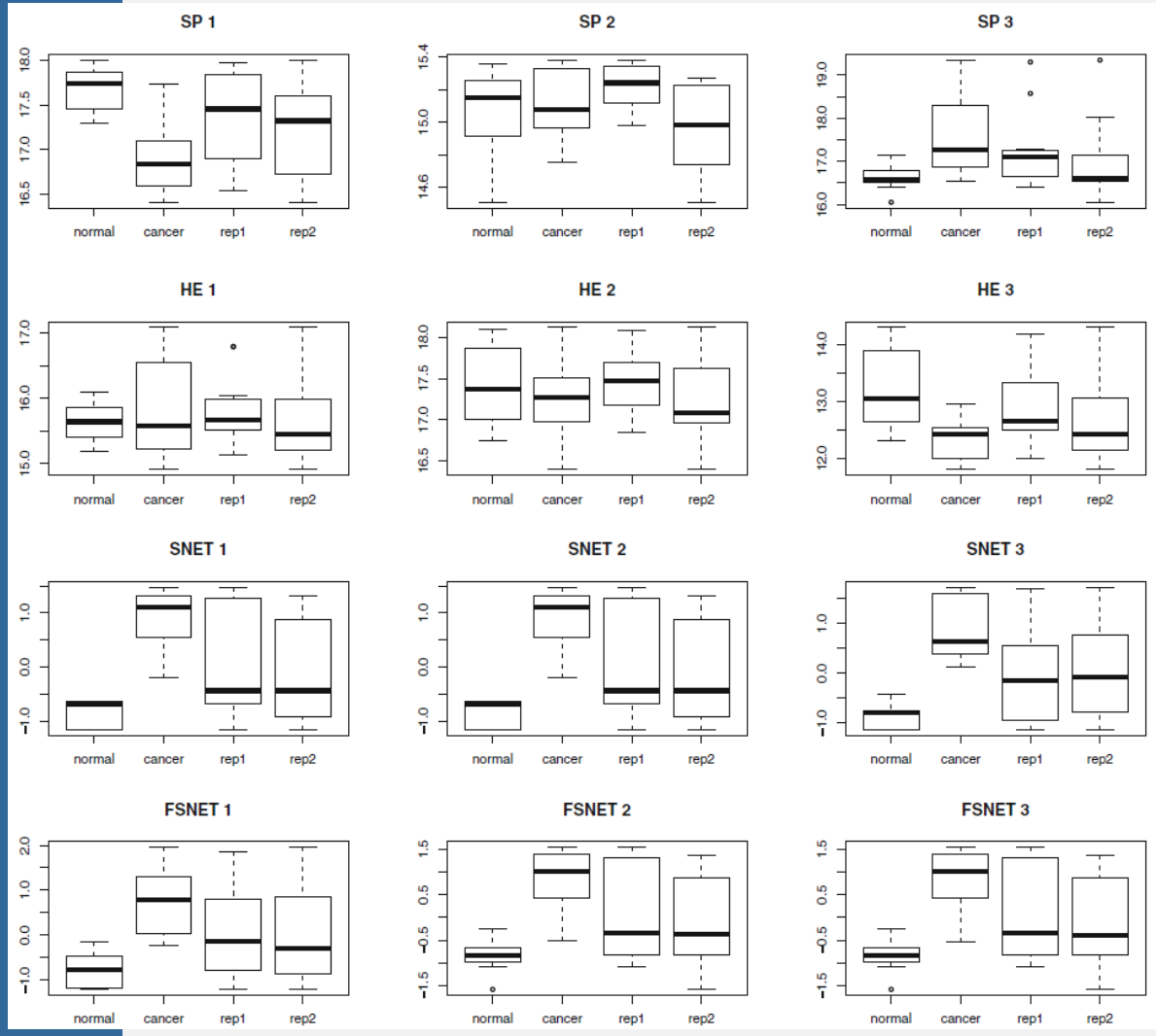
Wilson Wen Bin Goh, Limsoon Wong. **Protein complex-based analysis is resistant to the obfuscating consequences of batch effects---a case study in clinical proteomis**. *BMC Genomics*, 18(Suppl 2):142, March 2017.

FSNET is PFSNET using only one-class score; SNET is FSNET with alpha1 = alpha2

Batch effects "avoid" top PCs produced from protein complexes selected by SNET & FSNET

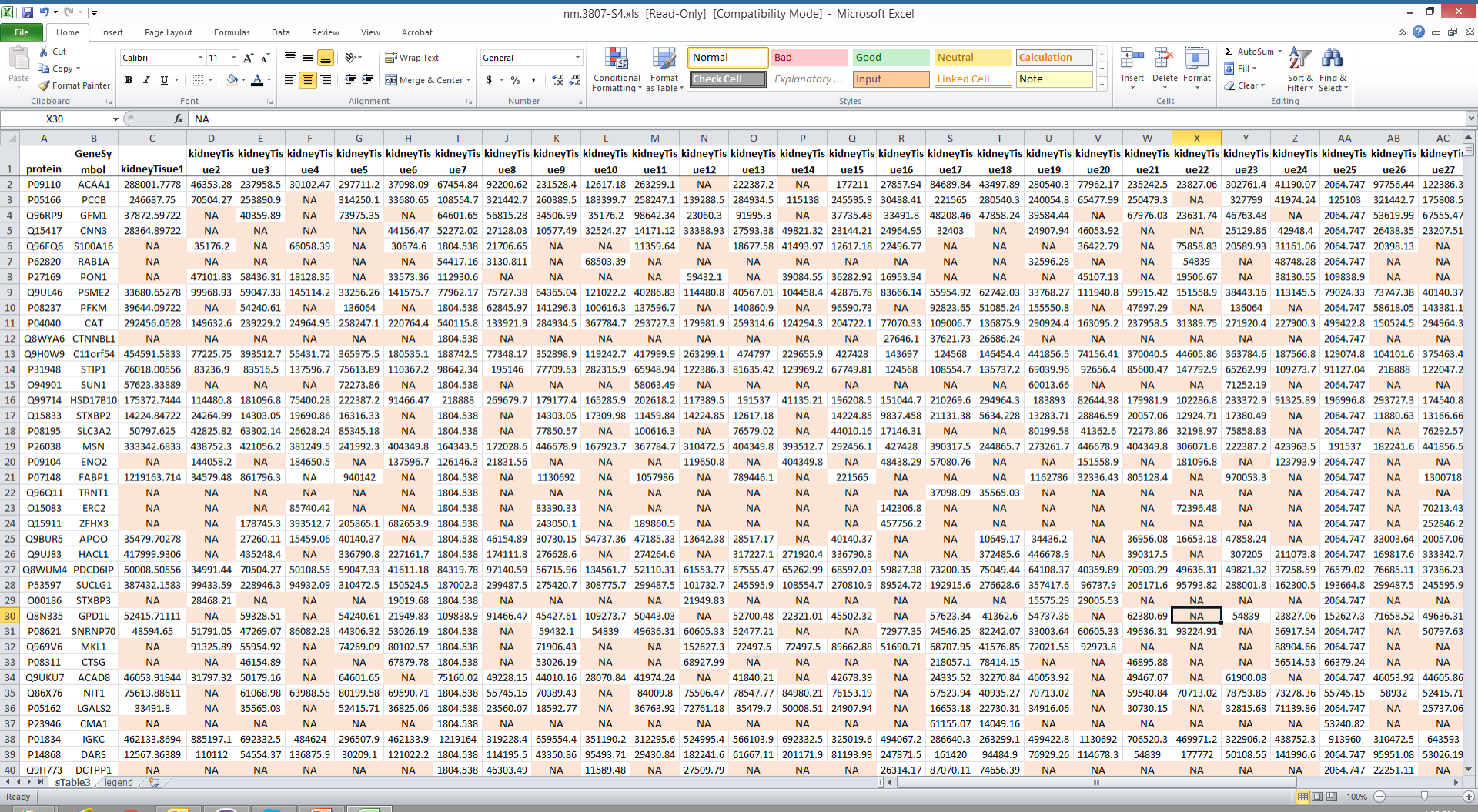


Batch effects "avoid" top protein complexes selected by SNET & FSNET



Recovering missing proteins from proteomic screen

Lots of missing values in real proteomics datasets



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	protein	GeneSymbol	kidneyTisue1	kidneyTis ue2	kidneyTis ue3	kidneyTis ue4	kidneyTis ue5	kidneyTis ue6	kidneyTis ue7	kidneyTis ue8	kidneyTis ue9	kidneyTis ue10	kidneyTis ue11	kidneyTis ue12	kidneyTis ue13	kidneyTis ue14	kidneyTis ue15	kidneyTis ue16	kidneyTis ue17	kidneyTis ue18	kidneyTis ue19	kidneyTis ue20	kidneyTis ue21	kidneyTis ue22	kidneyTis ue23	kidneyTis ue24	kidneyTis ue25	kidneyTis ue26	kidneyTis ue27
2	P09110	ACAA1	288001.7778	46353.28	237958.5	30102.47	297711.2	37098.09	67454.84	92200.62	231528.4	12617.18	263299.1	NA	222387.2	NA	177211	27857.94	84689.84	43497.89	280540.3	77962.17	235242.5	23827.06	302761.4	41190.07	2064.747	97756.44	122386.3
3	P05166	PCCB	246687.75	70504.27	253890.9	NA	314250.1	33680.65	108554.7	321442.7	260389.5	183399.7	258247.1	139288.5	284934.5	115138	245595.9	30488.41	221565	280540.3	240054.8	65477.99	250479.3	NA	327799	41974.24	125103	321442.7	175808.5
4	Q96RP9	GFM1	37872.59722	NA	40359.89	NA	73975.35	NA	64601.65	56815.28	34506.99	35176.2	98642.34	23060.3	91995.3	NA	37735.48	33491.8	48208.46	47858.24	39584.44	NA	67976.03	23631.74	46763.48	NA	2064.747	53619.99	67555.47
5	Q15417	CNN3	28364.89722	NA	NA	NA	NA	44156.47	52272.02	27128.03	10577.49	32524.27	14171.12	33388.93	27593.38	49821.32	23144.21	24964.95	32403	NA	24907.94	46053.92	NA	NA	25129.86	42948.4	2064.747	26438.35	23207.51
6	Q96FQ6	S100A16	NA	35176.2	NA	66058.39	NA	30674.6	1804.538	21706.65	NA	NA	11359.64	NA	18677.58	41493.97	12617.18	22496.77	NA	NA	NA	36422.79	NA	75858.83	20589.93	31161.06	2064.747	20398.13	NA
7	P62820	RAB1A	NA	NA	NA	NA	NA	NA	54417.16	3130.811	NA	68503.39	NA	NA	NA	NA	NA	NA	NA	NA	32596.28	NA	NA	54839	NA	48748.28	2064.747	NA	NA
8	P27169	PON1	NA	47101.83	58436.31	18128.35	NA	33573.36	112930.6	NA	NA	NA	NA	59432.1	NA	39084.55	36282.92	16953.34	NA	NA	NA	45107.13	NA	19506.67	NA	38130.55	109838.9	NA	NA
9	Q9UL46	PSME2	33680.65278	99968.93	59047.33	145114.2	33256.26	141575.7	77962.17	75727.38	64365.04	121022.2	40286.83	114480.8	40567.01	104458.4	42876.78	83666.14	55954.92	62742.03	33768.27	111940.8	59915.42	151558.9	38443.16	113145.5	79024.33	73747.38	40140.37
10	P08237	PFKM	39644.09722	NA	54240.61	NA	136064	NA	1804.538	62845.97	141296.3	100616.3	137596.7	NA	140860.9	NA	96590.73	NA	92823.65	51085.24	155550.8	NA	47697.29	NA	136064	NA	2064.747	58618.05	143381.1
11	P04040	CAT	292456.0528	149632.6	239229.2	24964.95	258247.1	220764.4	540115.8	133921.9	284934.5	367784.7	293727.3	179981.9	259314.6	124294.3	204722.1	77070.33	109006.7	136875.9	290924.4	163095.2	237958.5	31389.75	271920.4	227900.3	499422.8	150524.5	294964.3
12	Q8WYA6	CTNNBL1	NA	NA	NA	NA	NA	NA	1804.538	NA	NA	NA	NA	NA	NA	NA	NA	27646.1	37621.73	26686.24	NA	NA	NA	NA	NA	NA	2064.747	NA	NA
13	Q9H0W9	C11orf54	454591.5833	77225.75	393512.7	55431.72	365975.5	180535.1	188742.5	77348.17	352898.9	119242.7	417999.9	263299.1	474797	229655.9	427428	143697	124568	146454.4	441856.5	74156.41	370040.5	44605.86	363784.6	187566.8	129074.8	104101.6	375463.4
14	P31948	STIP1	76018.00556	83236.9	83516.5	137596.7	75613.89	110367.2	98642.34	195146	77709.53	282315.9	65948.94	122386.3	81635.42	129969.2	67749.81	124568	108554.7	135737.2	69039.96	92656.4	85600.47	147792.9	65262.99	109273.7	91127.04	218888	122047.2
15	O94901	SUN1	57623.33889	NA	NA	NA	72273.86	NA	1804.538	NA	NA	58063.49	NA	NA	NA	NA	NA	NA	NA	NA	60013.66	NA	NA	71252.19	NA	2064.747	NA	NA	
16	Q99714	HSD17B10	175372.7444	114480.8	181096.8	75400.28	222387.2	91466.47	218888	269679.7	179177.4	165285.9	202618.2	117389.5	191537	41135.21	196208.5	151044.7	210269.6	294964.3	183893	82644.38	179981.9	102286.8	233372.9	91325.89	196996.8	293727.3	174540.8
17	Q15833	STXBP2	14224.84722	24264.99	14303.05	19690.86	16316.33	NA	1804.538	NA	14303.05	17309.98	11459.84	14224.85	12617.18	NA	14224.85	9837.458	21131.38	5634.228	13283.71	28846.59	20057.06	12924.71	17380.49	NA	2064.747	11880.63	13166.66
18	P08195	SLC3A2	50797.625	42825.82	63302.14	26628.24	85345.18	NA	1804.538	NA	77850.57	NA	100616.3	NA	76579.02	NA	44010.16	17146.31	NA	NA	80199.58	41362.6	72273.86	32198.97	75858.83	NA	2064.747	NA	76292.57
19	P26038	MSN	333342.6833	438752.3	421056.2	381249.5	241992.3	404349.8	164343.5	172028.6	446678.9	167923.7	367784.7	310472.5	404349.8	393512.7	292456.1	427428	390317.5	244865.7	273261.7	446678.9	404349.8	306071.8	222387.2	423963.5	191537	182241.6	441856.5
20	P09104	ENO2	NA	144058.2	NA	184650.5	NA	137596.7	126146.3	21831.56	NA	NA	NA	119650.8	NA	404349.8	NA	48438.29	57080.76	NA	NA	151558.9	NA	181096.8	NA	123793.9	2064.747	NA	NA
21	P07148	FABP1	1219163.714	34579.48	861796.3	NA	940142	NA	1804.538	NA	1130692	NA	1057986	NA	789446.1	NA	221565	NA	NA	NA	1162786	32336.43	805128.4	NA	970053.3	NA	2064.747	NA	1300718
22	Q96Q11	TRNT1	NA	NA	NA	NA	NA	NA	1804.538	NA	NA	NA	NA	NA	NA	NA	NA	NA	37098.09	35565.03	NA	NA	NA	NA	NA	NA	2064.747	NA	NA
23	O15083	ERC2	NA	NA	NA	85740.42	NA	NA	1804.538	NA	83390.33	NA	NA	NA	NA	NA	NA	142306.8	NA	NA	NA	NA	NA	72396.48	NA	NA	2064.747	NA	70213.43
24	Q15911	ZFHX3	NA	NA	178745.3	393512.7	205865.1	682653.9	1804.538	NA	243050.1	NA	189860.5	NA	NA	NA	NA	457756.2	NA	NA	NA	NA	NA	NA	NA	NA	2064.747	NA	252846.2
25	Q9BUR5	APOO	35479.70278	NA	27260.11	15459.06	40140.37	NA	1804.538	46154.89	30730.15	54737.36	47185.33	13642.38	28517.17	NA	40140.37	NA	NA	10649.17	34436.2	NA	36956.08	16653.18	47858.24	NA	2064.747	33003.64	20057.06
26	Q9UJ83	HACL1	417999.9306	NA	435248.4	NA	336790.8	227161.7	1804.538	174111.8	276628.6	NA	NA	274264.6	NA	317227.1	271920.4	336790.8	NA	NA	372485.6	446678.9	NA	390317.5	NA	307205	211073.8	2064.747	169817.6	333342.7
27	Q8WUM4	PDCD6IP	50008.50556	34991.44	70504.27	50108.55	59047.33	41611.18	84319.78	97140.59	56715.96	134561.7	52110.31	61553.77	67555.47	65262.99	68597.03	59827.38	73200.35	75049.44	64108.37	40359.89	70903.29	49636.31	49821.32	37258.59	76579.02	76685.11	37386.23
28	P53597	SUCLG1	387432.1583	99433.59	228946.3	94932.09	310472.5	150524.5	187002.3	299487.5	275420.7	308775.7	299487.5	101732.7	245595.9	108554.7	270810.9	89524.72	192915.6	276628.6	357417.6	96737.9	205171.6	95793.82	288001.8	162300.5	193664.8	299487.5	245595.9
29	O00186	STXBP3	NA	28468.21	NA	NA	NA	19019.68	1804.538	NA	NA	NA	NA	21949.83	NA	NA	NA	NA	NA	NA	15575.29	29005.53	NA	NA	NA	NA	2064.747	NA	NA
30	Q8N335	GPD1L	52415.71111	NA	59328.51	NA	54240.61	21949.83	109838.9	91466.47	45427.61	109273.7	50443.03	NA	52700.48	22321.01	45502.32	NA	57623.34	41362.6	54737.36	NA	62380.69	NA	54839	23827.06	152627.3	71658.52	49636.31
31	P08621	SNRNP70	48594.65	51791.05	47269.07	86082.28	44306.32	53026.19	1804.538	NA	59432.1	54839	49636.31	60605.33	52477.21	NA	NA	72977.35	74546.25	82242.07	33003.64	60605.33	49636.31	93224.91	NA	56917.54	2064.747	NA	50797.63
32	Q969V6	MKL1	NA	91325.89	55954.92	NA	74269.09	80102.57	1804.538	NA	71906.43	NA	NA	152627.3	72497.5	72497.5	89662.88	51690.71	68707.95	41576.85	72021.55	92973.8	NA	NA	NA	88904.66	2064.747	NA	NA
33	P08311	CTSG	NA	NA	46154.89	NA	NA	67879.78	1804.538	NA	53026.19	NA	NA	68927.99	NA	NA	NA	NA	218057.1	78414.15	NA	NA	46895.88	NA	NA	56514.53	66379.24	NA	NA
34	Q9UKU7	ACAD8	46053.91944	31797.32	50179.16	NA	64601.65	NA	75160.02	49228.15	44010.16	28070.84	41974.24	NA	41840.21	NA	42678.39	NA	24335.52	32270.84	46053.92	NA	49467.07	NA	61900.08	NA	2064.747	46053.92	44605.86
35	Q86X76	NIT1	75613.88611	NA	61068.98	63988.55	80199.58	69590.71	1804.538	55745.15	70389.43	NA	84009.8	75506.47	78547.77	84980.21	76153.19	NA	57523.94	40935.27	70713.02	NA	59540.84	70713.02	78753.85	73278.36	55745.15	58932	52415.71
36	P05162	LGALS2	33491.8	NA	35565.03	NA	52415.71	36825.06	1804.538	23560.07	18592.77	NA	36763.92	72761.18	35479.7	50008.51	24907.94	NA	16653.18	22730.31	34916.06	NA	30730.15	NA	32815.68	71139.86	2064.747	NA	25737.06
37	P23946	CMA1	NA	NA	NA	NA	NA	NA	1804.538	NA	NA	NA	NA	NA	NA	NA	NA	NA	61155.07	14049.16	NA	NA	NA	NA	NA	NA	53240.82	NA	NA
38	P01834	IGKC	462133.8694	885197.1	692332.5	484624	296507.9	462133.9	1219164	319228.4	659554.4	351190.2	312295.6	524995.4	566103.9	692332.5	325019.6	494067.2	286640.3	263299.1	499422.8	1130692	706520.3	469971.2	322906.2	438752.3	913960	310472.5	643593
39	P14868	DARS	12567.36389	110112	54554.37	136875.9	30209.1	121022.2	1804.538	114195.5	43350.86	95493.71	29430.84	182241.6	61667.11	201171.9	81193.99	247871.5	161420	94484.9	76929.26	114678.3	54839	177772	50108.55	141996.6	2064.747	95951.08	53026.19
40	Q9H773	DCTPP1	NA	NA	NA	NA	NA	NA	1804.538	46303.49	NA	11589.48	NA	27509.79	NA	NA	NA	26314.17	87070.11	74656.39	NA	NA	NA	NA	NA	NA	2064.747	22251.11	NA

Guo et al. *Nature Medicine*, *21*, 407, 2015

PROTREC

Recall this postulate:

The chance of a protein complex being present is proportional to the fraction of its protein members being correctly reported in the screen

Presence of complex implies presence of all member proteins

PROTREC: Rank predicted missing proteins by

$$\text{Prob(Protein p is present but unreported)} = \text{Max}_{\text{complex C contains p}} \text{Prob(p is present | C is present)} * \text{Prob(C is present)} + \text{Prob(p is present | C is absent)} * \text{Prob(C is absent)}$$

SWATH experiment

If there are technical replicates, they should have reported the same proteins. So we can run PROTREC on one replica, and see whether the predicted missing proteins show up in other replicas

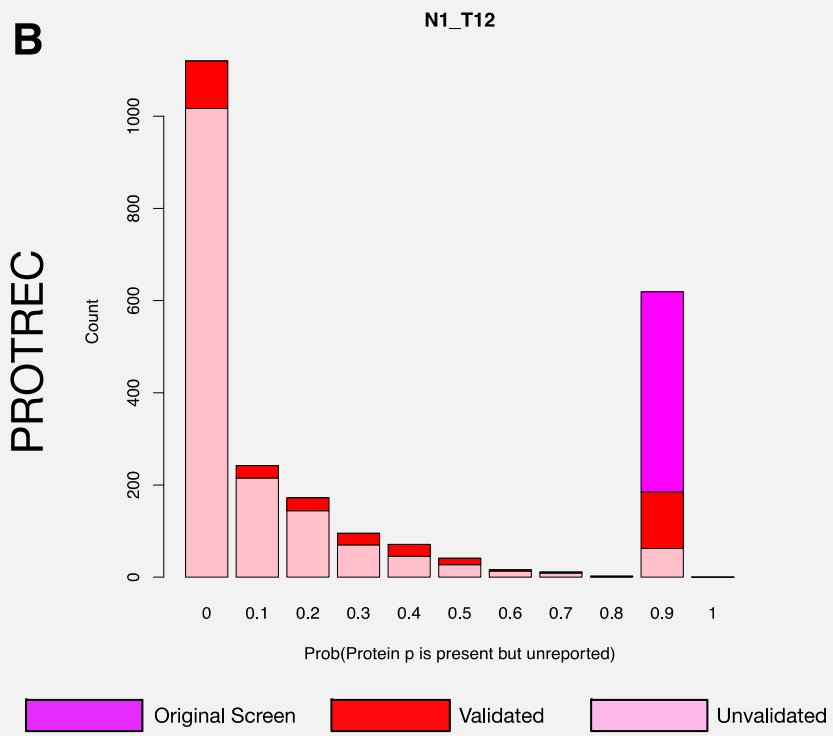
Proteomics data used: Renal cancer

Guo et al. Nature Medicine, 21(4):407-413, 2015

6 pairs of normal vs cancer ccRCC tissues

SWATH in duplicates

“Missing” proteins with high PROTREC scores are usually supported by at least one peptide in technical replicates



Closing remarks

What have we learned?

Contextualization into protein complexes can help with consistency & reproducibility issues in proteomic profile analysis

It can help with coverage issue too

It can also help with consistency & reproducibility issues in gene expression profile analysis

It can mitigate against batch effects

References

[qPSP] Goh et al. **Quantitative proteomics signature profiling based on network contextualization**. *Biology Direct*, 10:71, 2015

[PFSNET] Goh & Wong. **Evaluating feature-selection stability in next-generation proteomics**. *Journal of Bioinformatics and Computational Biology*, 14(5):1650029, 2016

[ESSNET] Goh & Wong. **Advancing clinical proteomics via analysis based on biological complexes: A tale of five paradigms**. *Journal of Proteome Research*, 15(9):3167-3179, 2016

[NETPROT] Goh & Wong. **NetProt: Complex-based feature selection**. *Journal of Proteome Research*, 16(8):3102-3112, 2017

Goh & Wong. **Protein complex-based analysis is resistant to the obfuscating consequences of batch effects---a case study in clinical proteomis**. *BMC Genomics*, 18(Suppl 2):142, 2017

[PROTREC] Goh & Wong. **Integrating networks and proteomics: Moving forward**. *Trends in Biotechnology*, 34(12):951-959, 2016