Supplementary: Direction pathway analysis of large-scale proteomics data reveals novel features of the insulin action pathway

Pengyi Yang[1,2,3*], Ellis Patrick[2*], Shi-Xiong Tan[3,4]
Daniel J. Fazakerley[3], James Burchfield[3], Christopher Gribben[3]
Matthew J. Prior[3], David E. James[3], Yee Hwa Yang[2†]

[1]Systems Biology Group, Biostatistics Branch
National Institute of Environmental Health Sciences
National Institute of Health, Research Triangle Park, NC 27709, USA

[2]School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia

[3]Diabetes and Obesity Program, Garvan Institute of Medical Research, NSW 2006, Australia

[4]Metabolism in Human Disease Unit, Institute of Molecular and Cellular Biology
A*Star, 61 Biopolis Drive, Proteos 138673, Singapore

1 *p*-value combination

There are many methodologies for combining information across studies or within pathways and the key discriminating differences between many of these methods are their assumed alternative hypotheses (Tseng *et al.*, 2012). Let τ_j represent test statistics for $j = 1, 2, ..., n$, where n are the number of tests. Assume the null hypothesis that none of the features measured by these test statistics have changed. Li and Tseng (2011) propose two broad classes of alternative hypotheses H_A and H_B. The first class of alternative hypothesis, H_A, is used to detect a series of tests in which *all* the test statistics show change. The corresponding test can be expressed as

$$\begin{cases} H_0: & \tau_j = 0 \ \forall \ j = 1, 2, ..., n. \\ H_A: & \tau_j > 0 \ \forall \ j = 1, 2, ..., n. \end{cases} \tag{1}$$

The second class of alternative hypothesis, H_B, is used to detect a series of tests in which *any* of the test statistics show change. The corresponding test can be expressed as

$$\begin{cases} H_0: & \tau_j = 0 \ \forall \ j = 1, 2, ..., n. \\ H_B: & \tau_j > 0 \text{ for at least one } j \text{ in } 1, 2, ..., n. \end{cases} \tag{2}$$

*These authors contributed equally.
†To whom correspondence should be addressed.

Here we consider four methods of p-value combination that could be used to perform the previously described tests:
Fisher: Fisher's method is defined as $\mathrm{P}\left(\chi^2_{2n} > -2\sum_{j=1}^{n}\log(p_j)\right)$ (Fisher, 1932).
Stouffer: Stouffer's method is defined as $\Phi\left(\frac{\sum_{j=1}^{n}\Phi^{-1}(p_j)}{n}\right)$ (Stouffer *et al.*, 1949).
maxP: The maximum of the p_j for $j = 1, 2, ..., n$ (Wilkinson, 1951).
OSP: A one-sided version of Pearson's method $\mathrm{P}\left(\chi^2_{2n} < -2\sum_{j=1}^{n}\log(1-p_j)\right)$ (Pearson, 1934).

The combined p-values for Fisher's and Stouffer's methods converge to zero if any one of the p_j also converges to zero making them appropriate for testing H_0 against the alternative H_B. For maxP or OSP to converge to zero all p_j must converge to zero, thus making them appropriate for testing H_0 against the alternative H_A. For our direction pathway analysis we would like to identify pathways that have had any of their proteins changed in all perturbations in the direction of interest. When put in the context of the two classes of alternative hypotheses, this would then require the tandem use of combination methods that favourable to testing H_0 against H_A when combining across experimental perturbations and H_0 against H_B when combining within a pathway, respectively.

The properties of the four methods are further illustrated in Figure 1 from a two dimensional perspective. While the arbitrary cut-offs of maxP and OSP are quite different their overall topologies are quite similar. Fisher is also seen to be quite sensitive to any change. If one of the z-scores is larger than approximately 2.4, then regardless of the sign of the other z-score, the combined p-value will be less than 0.05.

1.1 Simulation

To illustrate and distinguish the performance of the four p-value combination methods in relationship to the two classes of alternative hypotheses (H_A and H_B), we perform a simulation study to assess how each method combines information from three test statistics. The distributions of the three test statistics were chosen to describe situations of no change, mild change and strong change in none, some or all of the statistics. In our simulation study, we simulate directly from the standard normal distribution, this is equivalent to simulating test statistics (potentially from a two-sample t-test) and transforming them into z-scores. In practice, these test statistics have most likely come from multiple two-sample t-tests but could be other statistics such as regression coefficients.

In each simulation we generated $1,000,000$ observations of the three test statistics from the multivariate normal distribution $\mathbf{X}_i \sim \mathrm{N}(\mu_i, I)$ where $\mu_1 = (0, 0, 0)$. We call this simulation I and it represents the initial test results obtained from 3 different sets of experiments. With all 3 components of μ_1 equal zero, this represents the situation where test statistics were generated from comparisons with no change. We have also examine different values of μ in different simulations.

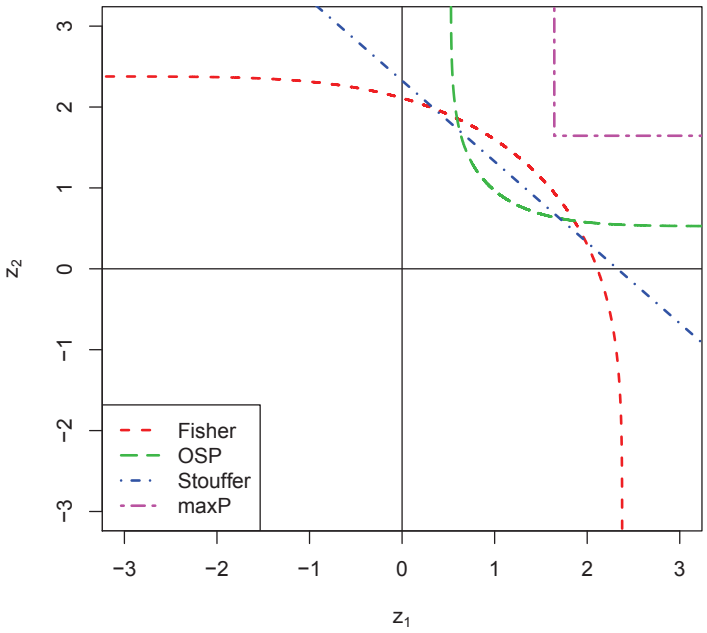

Figure 1: ***p*-value cut-offs for various combination methods** – A plot illustrating a p-value cut-off of 0.05 for various p-value combination methods in a two dimensional setting. The p-value cut-off is plotted in the negative z-score space so that a small p-value corresponds to a large positive z-score. The combination methods under consideration are Fisher (red), Stouffer (blue), maxP (pink) and OSP (green).

- $\mu_2 = (2, 0, 0)$ in simulation II ;
- $\mu_3 = (2, 2, 0)$ in simulation III;
- $\mu_4 = (4, 0, 0)$ in simulation IV and
- $\mu_5 = (2, 2, 2)$ in simulation V.

Notice, that simulation V represents the situation where all three test statistics were generated from comparisons with change. Simulations II, III and IV represent situations where at least one of the three statistics were generated from comparisons with change.

We then applied each p-value combination method to each observation in each simulation. An observation was called significant if its overall significance (combined p-value) was less than an arbitrary cut-off of 0.05. The percentage of called significance from all observations was used to characterise each method's power in performing hypothesis testing under the two classes of alternative hypotheses H_A and H_B.

In general a method with a low percentage of significance in simulation I (as this simulation is consistent with the null hypothesis of no change) and high percentage in simulations II, III, IV and V (as these are simulations representing some change) would be a good method for testing H_0 against H_B; that an observation is changed in *any* tests. While a method with a low percentage of significance in simulations I, II, III and IV (as these are simulations where not all statistics have changed) but high percentage in simulation V (a simulation where all the statistics have changed) would be more suitable for testing H_0 against H_A; that an observation is changed in *all* tests.

1.2 Simulation results

Results from the simulation study can be seen in Table 1. Focusing on the results from simulation I, Fisher, Stouffer and OSP all call five percent of the observations significant. As this simulation represents the situation where test statistics were generated from comparisons with no change, all the observations called significant are false positives. As an arbitrary cut-off of 0.05 was used, it is comforting to see that the false positive rates of Fisher, Stouffer and OSP are consistent with this. Furthermore, OSP is similar with Fisher and Stouffer in simulation V but calls much less significance in simulation II, III and IV compared to the other two methods. Together, these results demonstrate that OSP is the most suitable method at testing H_0 against H_A; that an observation is changed in *all* tests.

When considering the class of alternative hypothesis H_B, Stouffer is most powerful in detecting observations that have changed in all three tests (simulation V). It also has higher power in detecting changes from simulations II, III, and IV when compared to OSP. In comparison, Fisher has the highest percentage in simulations II, III and IV. These results suggest that both Stouffer and Fisher are most suitable for testing H_0 against H_B however while Fisher appears to be more sensitive to any changes, having the highest percentage in simulations II, III and IV, Stouffer is relatively more conservative.

Table 1: Results for five simulations in evaluating the performance of the four p-value combination methods in relationship to the two classes of alternative hypotheses H_A and H_B. The percentage of combined p-values less than 0.05 over 1,000,000 simulations (rounded to two decimal places) are reported.

Simulation	I	II	III	IV	V
methods	$\mu_1 = (0, 0, 0)$	$\mu_2 = (2, 0, 0)$	$\mu_3 = (2, 2, 0)$	$\mu_4 = (4, 0, 0)$	$\mu_5 = (2, 2, 2)$
Fisher	0.05	0.43	0.80	0.95	0.95
Stouffer	0.05	0.31	0.75	0.75	0.97
maxP	0.00	0.00	0.02	0.00	0.26
OSP	0.05	0.17	0.46	0.20	0.93

In our application, we select Stouffer to combine protein statistics within a pathway so that any pathway results are less likely to be driven by a single protein. These simulations demonstrate the importance of having a clearly defined alternative hypothesis in mind when analysing data.

2 Extended experimental procedures

3T3-L1 fibroblasts obtained from the Howard Green Laboratory (Boston, MA) were cultured in Dulbecco's modified Eagle's medium (DMEM). Fibroblasts were differentiated to adipocytes after confluent, followed by 3 days of post differentiation in growth media plus insulin (0.35 μM). Adipocytes were used between 10 to 20 passages of post-differentiation.

2.1 Proteomic profiling

In the two sets of mass spectrometry experiments, lysates of the cells were mixed in a 1:1:1 ratio, respectively, for each experiment. Plasma membranes from the first set were purified from cell lysates using the cationic silica isolation method described (Chaney and Jacobson, 1983). Purified PM fraction were resolved by SDS-PAGE, extracted, and digested by trypin. After tryptic digestion, peptides were separated on a Dionex Ultimate 300 LC system, and analysed by a LTQ-FT Ultra mass spectrometer. For the second set of proteomic profiling, purified PM fraction were tryptic digested and the digested peptides were subjected to strong anion exchange fractionation (SAX) on SAX-Stagetips generated in-house and analysed on an Orbitrap Velos mass spectrometer (Thermo Fisher Scientific).

Collected spectra data were preprocessed using MaxQuant software version 1.2.0.18 with the mouse IPI database v3.85. False discovery rate (FDR) at both peptide and protein group level were controlled at 1%. Only proteins that were observed within all conditions were analysed. Each sample was normalised such that the median protein

expression of that sample was zero. Two sample t-tests were performed for each protein in each condition with the exception of the MK samples. As there was no replication in the MK experiment, z-scores were calculated by assuming the variance of all the proteins were equal and was equal to the average variance of the other experiment conditions.

2.2 Immunoblotting

Polyclonal rabbit antibodies raised against pThr308 Akt and pSer246 PRAS40 were purchased from Cell Signaling Technologies (Beverly, MA). Polyclonal rabbit antibody raised against 14-3-3 was purchased from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA). Syntaxin 6 antibody were a kind gift from Robert Piper (University of Iowa, Iowa). Rabbit polyclonal antibodies against Thr642 AS160 were as previously described (Tellam *et al.*, 1997). Horseradish peroxidase-conjugated secondary antibodies were from Amersham Biosciences (Buckinghamshire, UK) and IR dye 700 or 800 conjugated secondary antibodies were from Rockland Immunochemicals (Gilbertsville, PA). Dulbecco's modified Eagle's medium (DMEM) and F-12 medium were from Invitrogen. Fetal calf serum was from Trace Scientific (Melbourne, Australia. Bovine serum albumin (BSA) was from Bovogen (Essendon, Australia). Bicinchoninic acid reagent and SuperSignal West Pico chemiluminescent substrate were from Pierce (Rockford, IL). Protease inhibitor mixture tablets were from Roche Applied Science (Indianapolis, IN). The Akt inhibitor, MK-2206, was generously provided by Professor Dario Alessi (University of Dundee, Dundee, UK). Other materials were obtained from Sigma chemical Co (St Louis, MO).

Cells were washed twice with ice-cold PBS and solubilized in 2% SDS in PBS containing phosphatase inhibitors (1 mM sodium pyrophosphate, 2 mM sodium vanadate, 10 mM sodium fluoride) and complete protease inhibitor mixture. Insoluble material was removed by centrifugation at 18,000g for 10 min. Protein concentration was measured using the bicinchoninic acid method. Proteins were separated by SDS-PAGE for immunoblot analysis. After transferring proteins to polyvinylidene difluoride membranes, membranes were incubated in blocking buffer containing 5% skim milk in Tris-buffered saline and immunoblotted with the relevant antibodies overnight at 4 °C in blocking buffer containing 5% BSA, 0.1% Tween in Tris-buffered saline. After incubation, membranes were washed and incubated with horseradish peroxidase-labeled secondary antibodies and then detected by SuperSignal West Pico chemiluminescent substrate. Quantification of protein levels was performed using ImageJ software (Abràmoff *et al.*, 2004).

2.3 Live cell microscopy

42mm and 10mm Glass coverslips (PeCon GmbH, Erbach, Germany) were incubated at room temperature for 120 min with a 1:50 dilution of Matrigel in ice cold PBS. Coverslips were washed twice with PBS prior to use. After 7-9 days post-differentiation, adipocytes were trypsinised with 5x Trypsin/EDTA for 5-10 min at 37 °C, washed twice with PBS and resuspended in Electroporation Solution (20mM Hepes, 135mM KCl, 2mM MgCl2, 0.5% Ficol 400, 1% DMSO, 2 mM ATP and 5 mM Glutathione, pH7.6) along with 5-20μg of plasmid DNA. Cells were electroporated at 200mV for 20ms using an ECM 830 Square

Wave Electroporation System, (BTX Molecular Delivery Systems, Massachusetts, USA) and seeded onto matrigel coated coverslips. Adipocytes were maintained in DMEM supplemented with 10% FCS until required.

Coverslips were then mounted in a perfusion open/closed chamber (POC; PeCon GmbH, Erbach, Germany) containing modified KRP buffer (120 mM NaCl, 0.6 mM Na2HPO4, 0.4 mM NaH2PO4, 6 mM KCl, 1.2 mM MgSO4, 12.5 mM HEPES, 1 mM CaCl2, 10mM Glucose, 1x MEM Amino Acids Solution, 20 mM GlutaMAX, 0.2% (w/v) BSA, pH7.4) and placed in a heated stage microscope insert 'P' (PeCon GmbH; Erbach, Germany) on an Axiovert 200M (Carl Zeiss Microscopy GmbH, Germany) equipped with a large incubator XL (PeCon GmbH, Erbach, Germany) maintained at 37 ^{o}C.

Healthy, suitably transfected cells were identified by brightfield and fluorescence using an appropriate objective (Typically a Zeiss A-Plan 20x/0.45). TdTomato and pHluorin were simultaneously excited using a 488/5nm bandpass filter. Emitted fluorescence was filtered by a 500nm LP filter and then split (568nm dichroic with 525/25nm and 607/70nm bandpass filters) onto two halves of an iXon DU-888D EMCCD camera (Andor, Belfast, N. Ireland) using a custom configured optosplit II (Cairn Research, Kent, UK). In this configuration, bleed through from green:red was measured at less than 3% and as such was considered negligible. All images were acquired using μManager (Edelstein *et al.*, 2010) and analyzed using ImageJ (Abràmoff *et al.*, 2004) and Cell Profiler 2.0 (Carpenter *et al.*, 2006).

References

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image processing with imagej. *Biophotonics International*, **11**(7), 36–42.

Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J., *et al.* (2006). Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, **7**(10), R100.

Chaney, L. K. and Jacobson, B. S. (1983). Coating cells with colloidal silica for high yield isolation of plasma membrane sheets and identification of transmembrane proteins. *Journal of Biological Chemistry*, **258**(16), 10062–10072.

Edelstein, A., Amodaj, N., Hoover, K., Vale, R., and Stuurman, N. (2010). Computer control of microscopes using μmanager. *Current Protocols in Molecular Biology*, pages 14–20.

Fisher, R. (1932). Statistical methods for research workers.

Li, J. and Tseng, G. (2011). An adaptively weighted statistic for detecting differential gene expression when combining multiple transcriptomic studies. *The Annals of Applied Statistics*, **5**(2A), 994–1019.

Pearson, K. (1934). On a new method of determining "goodness of fit". *Biometrika*, **26**(4), 425–442.

Stouffer, S., Suchman, E., DeVinney, L., Star, S., and Williams Jr, R. (1949). The american soldier: adjustment during army life.(studies in social psychology in world war ii, vol. 1.).

Tellam, J. T., Macaulay, S. L., McIntosh, S., Hewish, D. R., Ward, C. W., and James, D. E. (1997). Characterization of munc-18c and syntaxin-4 in 3t3-l1 adipocytes putative role in insulin-dependent movement of glut-4. *Journal of Biological Chemistry*, **272**(10), 6179–6186.

Tseng, G. C., Ghosh, D., and Feingold, E. (2012). Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res*, **40**(9), 3785–3799.

Wilkinson, B. (1951). A statistical consideration in psychological research. *Psychol Bull*, **48**(3), 156–158.