New Computational Pipeline to Analyze Reverse Phase Protein Array Data from Adult Rats Exposed to Cigarette Smoke

Y. Xiang[1], S. Quadt-Humme[2], S. Ansari[2], A. Hengstermann[2], J. Hoeng[1], W. Schlage[2]

[1]Philip Morris International R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
[2]Philip Morris International R&D, Philip Morris Research Laboratories GmbH, Cologne, Germany

Introduction

Recent progress in functional proteomics, particularly in protein arrays such as reverse phase protein array (RPA), allows the measurement of changes in protein contents and modifications in a sensitive, high-throughput and customizable manner while requiring only small amounts of protein samples. Often, the RPA data is not normally distributed, which makes the usual linear regression model, ANOVA, and t-test unsuitable for RPA data analysis. Currently, there is no bioinformatics method specifically designed for RPA analysis. To identify proteins and pathways which play a key role in response to cigarette smoke exposure, a novel computational pipeline for RPA data analysis was developed. In this study, adult rats (n=5/group) were exposed for 28 days to conditioned fresh air (6 h/day) or to cigarette mainstream smoke (3R4F) (6 h/day at 8, 15, or 23 µg nicotine/l), and lungs and respiratory nasal epithelium samples were analyzed by RPA. Significance Analysis of Microarrays (SAM) was found to be suitable for identification of sets of differentially expressed proteins in smoke treated versus untreated animals. The empirical null distribution was generated by 150 random permutations, whereby delta was set to 0.01. To the best of our knowledge, this is the first time that SAM has been used for RPA data analysis. RPA experiment workflow and metadata capture were achieved using the caArray data management system and MAGE-TAB data exchange format with RPA-specific extensions, which enable support for commercial RPA technology and RPA-specific MAGE-TAB files.

Keywords

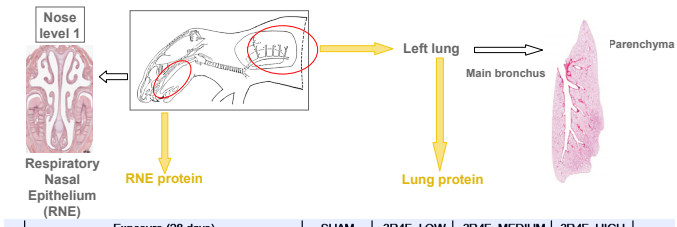
RPA: Reverse phase protein array. It is used to profile dozens or hundreds of samples (research or clinical) for the presence of antigens. Cell lysates, tissue sections, even material from laser capture microdissection, or serum samples are arrayed. Visualization can be performed with a detection antibody linked to a fluorophore or color detection reagent (1).

SAM: Significance Analysis of Microarrays (2). SAM is a technique originally developed for identifying genes that are differentially expressed in a significant manner in micro-array experiments. SAM is designed to handle data with both Gaussian and non-Gaussian distribution and it can handle small sample size problems.

ZeptoMARK Reverse Array: commercialized reverse phase protein arrays of Zeptosens (3).

SD rat: Sprague Dawley rats.

Materials & Methods



Exposure (28 days)	SHAM	3R4F_LOW	3R4F_MEDIUM	3R4F_HIGH
number of SD-rats	5	5	5	5
target nicotine concentration (µg/l)	0	8	15	23

Fig 1. Description of the overall experiment. Samples of the respiratory tract (respiratory nasal epithelium [RNE] and lung tissue) were taken at the defined sites that were also investigated in histopathology. In this study, adult SD rats (n=5/group) were exposed for 28 days to conditioned fresh air (6 h/day) or to cigarette mainstream smoke (3R4F) (6 h/day at 8, 15, or 23 µg nicotine/l).

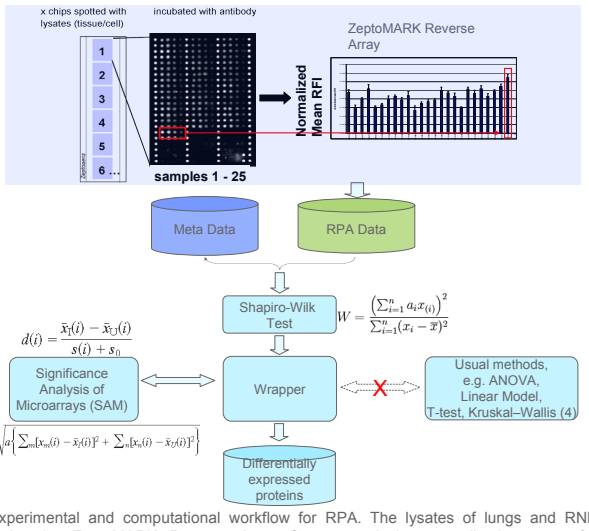
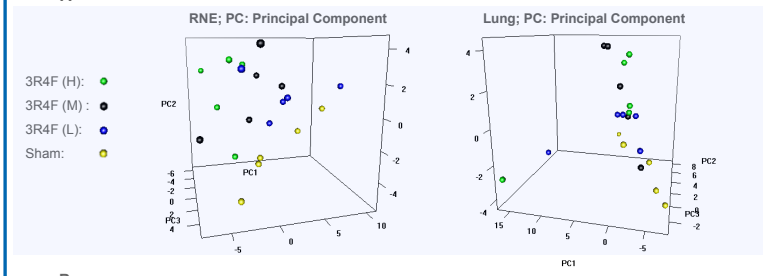


Fig 2. Experimental and computational workflow for RPA. The lysates of lungs and RNE were analyzed by the ZeptoMARK Reverse Array System and the normalized mean referenced fluorescence Intensities (RFI), which denoted protein abundances, were generated. Totally 46 and 47 antibodies were used to study the protein levels (non-phosphorylated and phosphorylated) in RNE and lung, respectively. Normality of RPA data was tested by using Shapiro-Wilk test. For Shapiro-Wilk test, $X(i)$ was the i-th observation of abundance of any specific protein in any 3R4F concentration. SAM was used for identification of differentially expressed proteins. For SAM, $\bar{x}_l(i)$ and $\bar{x}_U(i)$ were mean of abundance of protein i in group l and U respectively; $s(i)$ was the standard error of protein i; and s_0 was the adjusted factor for standard error.

Results

A



B

		RNE						Lung					
		3R4F (L) : sham		3R4F (M) : sham		3R4F (H) : sham		3R4F (L) : sham		3R4F (M) : sham		3R4F (H) : sham	
Protein	pathway	FC	FDR	FC	FDR	FC	FDR	FC	FDR	FC	FDR	FC	FDR
ATF3.c19	Transcription Factors	1.237	0.878	1.710	0.184	2.174	0.028	1.757	0.400	2.314	0.200	2.620	0.130
pc.Myc	MAPK Signaling, Transcription Factors	1.305	0.788	1.497	0.245	1.719	0.032	1.339	0.418	1.436	0.286	1.450	0.177
Prohibitin	Tumor Marker, Housekeeping	1.209	0.825	1.676	0.107	1.711	0.039	1.479	0.217	1.943	0.137	1.586	0.180
AMPK.alpha	Metabolism	1.423	0.845	1.301	0.341	1.565	0.041	1.264	0.562	1.241	0.551	1.386	0.249
MEK1.2	MAPK Signaling	1.252	0.888	1.687	0.282	1.932	0.051	1.733	0.202	2.090	0.201	2.006	0.152
pAkt.Ser473	Akt Signaling, Metabolism	1.115	0.888	1.307	0.565	1.996	0.062	-1.071	1.000	1.084	1.000	1.318	0.187
p38.MAPK	MAPK Signaling, Stress pathway	1.011	1.000	1.595	0.255	1.683	0.064	1.195	1.000	-1.135	1.000	1.208	1.000
pIGF.1R	Insulin Signaling, Metabolism, Receptors, Tyrosine Kinases	1.267	0.872	1.292	0.527	1.339	0.099	1.667	0.218	1.558	0.149	1.916	0.157
ikB.a	NFkB Signaling	1.043	0.853	1.471	0.314	1.453	0.111	1.326	0.565	1.231	0.561	1.527	0.290
pEGFR	MAPK Signaling, Receptors, Tyrosine Kinases	-1.143	0.864	1.233	0.537	1.371	0.126	1.016	1.000	1.120	1.000	1.113	1.000
p53	Apoptosis, Cell Cycle Control	1.468	0.886	1.907	0.522	1.309	1.000	2.823	0.161	3.639	0.142	3.361	0.127

C

Normality Test for p38.MAPK				
p38.MAPK	sham	3R4F (low)	3R4F (medium)	3R4F (high)
p value	0.693	0.049	0.692	0.938

D

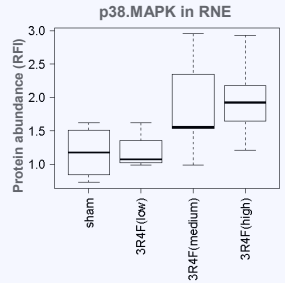


Fig 3. (A) Principal component analysis of RPA data in RNE (left panel) and lung (right panel). Overall protein abundance of rats exposed to low dose of 3R4F was the most similar to the protein abundance of rats exposed to sham for both RNE and lung. (B) Only differentially expressed proteins (cutoff of local false discovery rate (FDR) was 0.13, which was chosen according to both our previous experience and the local FDR distribution of this data set), among total 46/47 proteins, were listed in this table. (L), (M), and (H) in the header of the table denoted low, or medium, or high concentration of 3R4F respectively. A dose effect was observed: at higher concentrations of 3R4F, more proteins showed significant differences in abundance vs. sham, in both RNE and lung. Increased levels of phosphorylated protein p_c-Myc, protein MEK1.2, phosphorylated protein p_Akt-Ser473, protein p38.MAPK, and phosphorylated protein p_EGFR may be indicative of activation of MAPK and AKT pathway, which was supported by gene expression analysis. (C) Normality test for protein abundance of p38.MAPK in RNE. Protein abundance of p38.MAPK in RNE under low dose of 3R4F was not normally distributed (marked with red color). (D) Visualization of protein abundance of p38.MAPK in RNE. p38.MAPK abundance was increased in RNE of rats exposed to high 3R4F when compared to sham.

Conclusions

In conclusion, a novel computational pipeline for RPA data analysis was established. Our pipeline revealed distinct, significant changes of protein levels that were indicative of various pathways affected in respiratory nasal epithelium and lungs of SD rats exposed to high dose and medium dose cigarette smoke.

References

[1] Zhang DY, Ye F, Gao L, Liu X, Zhao X, Che Y, Wang H, Wang L, Wu J, Song D, Liu W, Xu H, Jiang B, Zhang W, Wang J, Lee P. Proteomics, pathway array and signaling network-based medicine in cancer. Cell Div. 2009 Oct 28;4:20.
[2] Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001 Apr 24;98(9):5116-21. Epub 2001 Apr 17.
[3] Zeptosens, located near Basel, Switzerland, is a division of Bayer (Schweiz) AG and part of Bayer Technology Services GmbH.
[4] Kruskal W, Wallis WA. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 1952 Dec; 47 (260):583–621.

Philip Morris International Research & Development,
Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland
T: +41 58 242 21 11, F: +41 58 242 28 11, W:
www.philipmorrisinternational.com