AKR1C1 as a Biomarker for Differentiating the Biological Effects of Combustible from Non-Combustible Tobacco Products

Quynh T. Tran[1*], Sangsoon Woo[2], Hong Gao[3], David Henderson[2], Wolfgang Zacharias[3], Gang Liu[1], GL Prasad[1]

[1]RAI Services Company, Winston Salem, NC 27101, USA [2]Statistical Genetics, Axio Research LLC, Seattle, WA 98121, USA [3]Department of Medicine, James Graham Brown Cancer Center, University of Louisville School of Medicine, KY 40202, USA

Abstract

Smoking has been established as a major risk factor for developing oral squamous cell carcinoma (OSCC), but less attention has been paid to the effects of smokeless tobacco products. Our objective is to identify potential biomarkers to distinguish the biological effects of combustible tobacco products from those of non-combustible ones using oral cell lines. Normal human gingival epithelial cells (HGEC), non-metastatic (101A) and metastatic (101B) OSCC cell lines were exposed to different tobacco product preparations (TPPs) including cigarette smoke total particulate matter (TPM), whole-smoke conditioned media (WS-CM), smokeless tobacco extract in complete artificial saliva (STE), or nicotine (NIC) alone. We performed microarray-based gene expression profiling and found 3456 probe sets from 101A, 1432 probe sets from 101B, and 2717 probe sets from HGEC to be differentially expressed. Gene Set Enrichment Analysis (GSEA) revealed xenobiotic metabolism and steroid biosynthesis were the top two pathways that were upregulated by combustible but not by non-combustible TPPs. Notably, aldo-keto reductase genes, AKR1C1 and AKR1C2, were the core genes in the top enriched pathways and were statistically upregulated more than eight-fold by combustible TPPs. Quantitative real time polymerase chain reaction (qRT-PCR) results statistically support AKR1C1 as a potential biomarker for differentiating the biological effects of combustible from non-combustible tobacco products.

Materials and Methods

Table 1. Treatment conditions for microarray gene expression profiling samples. Target cells were exposed for 24 h to different tobacco product preparations (TPPs) or solvent only as controls. For total particulate matter (TPM) and whole-smoke conditioned media (WS-CM), effective concentration for 30% of maximal effect (EC-30) doses were applied; for smokeless tobacco extract in complete artificial saliva (STE/CAS) , the dose with the same amount of nicotine (NIC) as that in TPM at EC-30 was applied. Also, treatment with a high dose of NIC was done. For each condition, the dilution factors and NIC equivalents delivered are listed.

	101A	101B	HGEC
TPM (EC-30) vs. **DMSO**	80 µg/mL 9.6 µg/mL NIC	150 µg/mL 18 µg/mL NIC	20 µg/mL 2.4 µg/mL NIC
Low-STE (containing same amount of NIC as those in EC-30 of TPM) vs. **CAS**	148-fold dilution 0.07% (w/v) 9.6 µg/mL NIC	79-fold dilution 0.13% (w/v) 18 µg/mL NIC	592-fold dilution 0.02% (w/v) 2.4 µg/mL NIC
WS-CM (EC-30) vs. **Media**	14-fold dilution 1.43% (v/v) 1.5 µg/mL NIC	12.5-fold dilution 1.60% (v/v) 1.7 µg/mL NIC	6-fold dilution 3.33% (v/v) 9.3 µg/mL NIC
High-STE vs. **CAS**	3-fold dilution 0.33% (w/v) 474 µg/mL NIC	3-fold dilution 0.33% (w/v) 474 µg/mL NIC	3-fold dilution 0.33% (w/v) 474 µg/mL NIC
low-NIC vs. **DMSO**	14 µg/mL NIC	N.D. *	14 µg/mL NIC
high-NIC vs. **DMSO**	474 µg/mL NIC	N.D. *	474 µg/mL NIC

Gene Expression Analysis

- To identify differentially expressed probe sets in any TPPs, a linear regression with empirical Bayes method was applied using the limma package in R [1].
- The Benjamini & Yekutieli (BY) false discovery rate (FDR) procedure was performed to correct for multiple hypothesis testing [2]. An adjusted *p*-value < 0.01 was considered to be statistically significant when applicable.
- A two tailed Student's *t*-test was used to determine the significance of the treatment effects within probe set: a value of $p < 0.05$ was considered to be statistically significant when applicable.
- To assess the association of a collection of gene sets with sensitivity to each of the treatment, we ran Gene Set Enrichment Analysis (GSEA) tool (v 2.2.2) at the default setting with "gene_set" permutation and 149 as the seed for permutation [3].

Results

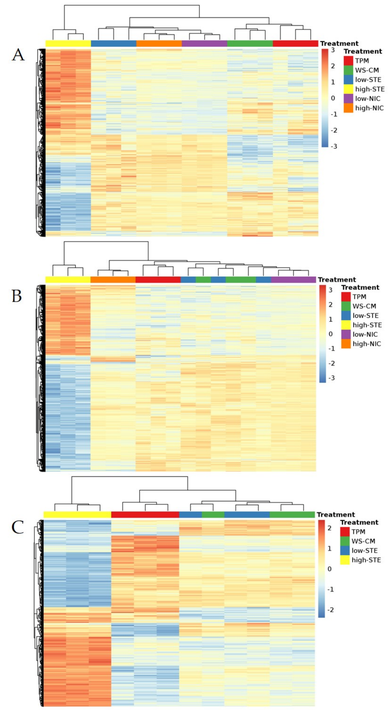

Figure 1. Heat maps of the differentially expressed probe sets in each cell line: (**A**) human gingival epithelial cell (HGEC) line; (**B**) non-metastatic 101A cell line; (**C**) metastatic 101B cell line. Hierarchical clustering was performed to cluster samples and probe sets into similar clusters. The color scale represents the log2 fold change ranging from −3 (dark blue) to 3 (dark red). Different treatments were color coded as follows: red = total particulate matter (TPM), blue = low-smokeless tobacco extract (low-STE), green = whole-smoke conditioned media (WS-CM), purple = low-nicotine (low-NIC), orange = high-nicotine (high-NIC), and yellow = high-smokeless tobacco extract (high-STE).

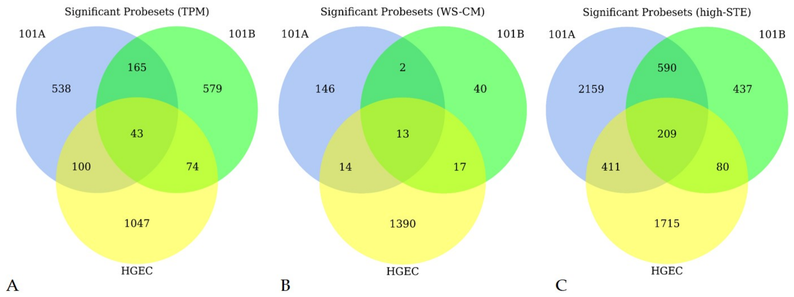

Figure 2. Venn diagrams showing the number of significant probe sets for each treatment in each cell line. (**A**) Under total particulate matter (TPM) condition; (**B**) Under whole-smoke conditioned media (WS-CM) condition; (**C**) Under high-smokeless tobacco extract (high-STE). Blue circle = 101A, green circle = 101B, and yellow circle = human gingival epithelial cell (HGEC).

Table 2. Six common genes that are robustly upregulated by total particulate (TPM) and whole-smoke conditioned media (WS-CM) in all 3 cell lines.

Affy ID	Gene Name	Log_2 FC		Adjusted *p*-Values	
		TPM	WS-CM	TPM	WS-CM
1562102_at		3.45	4.96	4.4591×10^{-7}	3.0208×10^{-15}
1555854_at	*AKR1C1, AKR1C2*	3.12	3.52	1.5291×10^{-7}	2.9126×10^{-8}
216594_x_at	*AKR1C1, AKR1C2*	2.37	3.01	1.4115×10^{-10}	1.9914×10^{-7}
204151_x_at	*AKR1C1, AKR1C2*	2.38	2.65	5.6067×10^{-12}	9.951×10^{-8}
211653_x_at	*AKR1C1, AKR1C2*	2.28	2.75	4.4159×10^{-10}	7.8189×10^{-7}
209699_x_at	*AKR1C1, AKR1C2*	2.01	2.90	1.8013×10^{-9}	3.075×10^{-7}
203665_at	*HMOX1*	3.87	4.72	1.7717×10^{-10}	8.392×10^{-12}
207528_s_at	*SLC7A11*	3.21	3.31	1.0043×10^{-9}	6.3878×10^{-10}
217678_at	*SLC7A11*	2.96	2.59	8.7581×10^{-6}	2.4987×10^{-9}
209921_at	*SLC7A11*	2.94	2.55	8.7581×10^{-6}	2.7297×10^{-9}
202831_at	*GPX2*	1.45	2.38	0.00188523	8.0032×10^{-10}
231897_at	*PTGR1*	0.80	1.73	0.00504145	5.9494×10^{-11}

Table 3. The top 3 up- and 3 down-regulated Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways from Gene Set Enrichment Analysis for each tobacco product preparation (TPP) in human gingival epithelial cells (HGEC).

KEGG PATHWAY	TPM	WS-CM	Low-STE	High-STE	Low-NIC
	q-value (NES) *	*q*-value (NES) *	*q*-value (NES) *	*q*-value (NES) *	*q*-value (NES) *
METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450	0 (2.41)	0 (2.37)	--	0.0095 (1.90)	--
STEROID_HORMONE_BIOSYNTHESIS	0 (2.36)	0.0014 (2.05)	--	0.0050 (2.00)	--
RETINOL_METABOLISM	0.0018 (2.02)	--	--	--	--
PHENYLALANINE_METABOLISM	--	--	0.0475 (−1.73)	--	--
PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	--	0.0616 (1.78)	--	--	--
SYSTEMIC_LUPUS_ERYTHEMATOSUS	--	--	--	--	--
CELL_CYCLE	0 (−2.48)	0 (−2.30)	--	--	--
DNA_REPLICATION	0 (−2.30)	0 (−2.18)	--	--	--
OOCYTE_MEIOSIS	--	0.0007 (−2.05)	--	--	--
MISMATCH_REPAIR	0.0016 (−1.98)	--	--	--	--
GLYCINE_SERINE_AND_THREONINE_METABOLISM	--	--	0.0273 (−1.85)	--	--
OLFACTORY_TRANSDUCTION	--	--	0.0286 (−1.80)	--	--
ARACHIDONIC_ACID_METABOLISM	--	--	--	0.0050 (1.95)	--
CYTOSOLIC_DNA_SENSING_PATHWAY	--	--	--	0.0186 (−1.93)	--
PEROXISOME	--	--	--	0.0942 (−1.75)	--
RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY	--	--	--	0.0676 (−1.74)	--
FATTY_ACID_METABOLISM	--	--	--	--	--

* NES: normalized enrichment score. It reflects the degree to which a gene set is overrepresented at the top or bottom of a ranked genes list and accounts for differences in gene set size and in correlations between gene sets and the expression dataset. Positive NES indicates gene set enrichment at the top of the ranked list while a negative NES indicates gene set enrichment at the bottom. * FDR *q*-value is the estimated probability that a gene set with a given NES represents a false positive finding. The suggestive FDR cutoff of 25% was utilized.

Table 4. Microarray results and real time quantitative RT-PCR (qRT-PCR) results with Student's *t*-test *p*-value for *AKR1C1* and *AKR1C2* mRNA level in HGEC. qRT-PCR was performed using the TaqMan® Gene Expression system according to the manufacturer's instruction.

TPPs	Microarray		qRT-PCR			
	AKR1C1/AKR1C2		*AKR1C1*		*AKR1C2*	
	Fold Change	Adj. *p*-Value	Fold Change	*p*-Value	Fold Change	*p*-Value
TPM	8.69	<0.001	16.43	<0.001	1.95	0.3121
WS-CM	11.47	<0.001	33.28	<0.001	2.07	0.2719
low-STE	1.08	1	1.23	0.5689	1.61	0.8227
high-NIC	1.59	0.5413	1.61	0.2052	0.59	0.4112

Conclusions

In summary, our global analysis of gene expression profiles in TPP-treated versus untreated human oral cavity cells revealed that two members of the *AKR1* gene family are highly induced by short-term exposure to combusted tobacco product components. Induction of some of these genes was previously observed in lung, colon, or rectal tumor tissues of smokers [4,5,6]. However, their selective induction by combusted, but not by non-combusted TPPs, has not been characterized before. Our results suggest that *AKR1C1*, and possibly *AKR1C2*, may be potential biomarkers for reduced harm effects in oral cavity cells when treated with non-combustible (ST/CAS; NIC) compared to combustible (TPM; WS-CM) tobacco products.

References

1. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47.
2. Benjamini, Y.; Yekutieli, D. The Control of the False Discovery Rate in Multiple Testing under Dependency. *Ann. Stat.* **2001**, *29*, 1165–1188.
3. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; Mesirov, J.P. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550.
4. Gao, H.; Prasad, G.L.; Zacharias, W. Differential cell-specific cytotoxic responses of oral cavity cells to tobacco preparations. *Toxicol. In Vitro* **2013**, *27*, 282–291.
5. Nagaraj, N.S.; Beckers, S.; Mensah, J.K.; Waigel, S.; Vigneswaran, N.; Zacharias, W. Cigarette smoke condensate induces cytochromes P450 and aldo-keto reductases in oral cancer cells. *Toxicol. Lett.* **2006**, *165*, 182–194.
6. Khariwala, S.S.; Hatsukami, D.; Hecht, S.S. Tobacco carcinogen metabolites and DNA adducts as biomarkers in head and neck cancer: Potential screening tools and prognostic indicators. *Head Neck* **2012**, *34*, 441–447.

This work has been published in the *Genes* journal (PMID: 28467356)
***Corresponding author** (trann1@rjrt.com)

Conflict of interest: Three of the authors (Q.T.T., G.L. and G.L.P.) are full time employees of RAI Services Company. This research was funded by RAIS. None of the authors has any other conflict of interest to declare