Evaluation of the Tobacco Heating System 2.2. Part 9: Application of systems pharmacology to identify exposure response markers in peripheral blood of smokers switching to THS2.2.

Florian Martin, Marja Talikka, Nikolai V. Ivanov, Christelle Haziza, Julia Hoeng, and Manuel C. Peitsch

PMI R&D, Philip Morris Products S.A., Neuchatel Switzerland (part of Philip Morris International group of companies)

Running title: candidate MRTP exposure response monitored with whole blood gene expression signature

TABLE OF CONTENTS

Abstract..... 2

Abbreviations..... 3

Introduction..... 4

Materials and Methods..... 7

Results..... 10

Discussion..... 12

Conclusion..... 14

Acknowledgements..... 14

Conflict of interest statement..... 14

References..... 15

Figures..... 18

Abstract

As part of current harm reduction strategies, candidate modified risk tobacco products (MRTP) are developed to offer adult smokers who want to continue using tobacco product an alternative to cigarettes while potentially reducing individual risk and population harm compared to smoking cigarettes. One of these candidate MRTPs is the Tobacco Heating System (THS) 2.2 which does not burn tobacco, but instead heats it, thus producing significantly reduced levels of harmful and potentially harmful constituents (HPHC) compared with combustible cigarettes (CC). A controlled, parallel group, open-label clinical study was conducted with subjects randomized to three monitored groups: (1) switching from CCs to THS2.2; (2) continuous use of non-menthol CC brand (CC arm); or (3) smoking abstinence (SA arm) for five days. Exposure response was assessed by measuring biomarkers of exposure to selected HPHCs. To complement the classical exposure response measurements, we have used the previously reported whole blood derived gene signature that can distinguish current smokers from either non-smokers or former smokers with high specificity and sensitivity. We tested the small signature consisting of only 11 genes on the blood transcriptome of subjects enrolled in the clinical study and showed a reduced exposure response in subjects that either stopped smoking or switched to a candidate MRTP, the THS2.2, compared with subjects who continued smoking their regular tobacco product.

Keywords: smoking, blood, exposure response, gene expression signature, novel tobacco heating system, MRTP

Abbreviations

BoE	Biomarker of Exposure
CC	Combustible Cigarette
COPD	Chronic Obstructive Pulmonary Disease
CS	Current Smoker
FS	Former Smoker
GEO	Gene Expression Omnibus
HPHC	Harmful and Potentially Harmful Constituents
LDA	Linear Discriminant Analysis
MCC	Matthews Correlation Coefficient
MRTP	Modified Risk Tobacco Product
NCS	Non-Current Smoker (former smoker or never smoker)
NS	Never Smoker
SA	Smoking Abstinence
THS2.2	Tobacco Heating System 2.2

Introduction

The U.S. Family Smoking Prevention and Tobacco Control Act (FSPTCA) defines a Modified Risk Tobacco Product (MRTP) as "any tobacco product that is sold or distributed for use to reduce harm or the risk of tobacco related disease associated with commercially marketed tobacco products" (Family Smoking Prevention and Tobacco Control Act, 2009). This publication is part of a series of nine publications describing the nonclinical and part of the clinical assessment of a candidate MRTP, the Tobacco Heating System (THS) 2.2 and a mentholated version thereof (THS2.2M). The series of publications provides part of the overall scientific program to assess the potential for THS2.2 to be a reduced risk product. The first publication in this series describes THS2.2 and the assessment program for MRTPs (Smith et al, submitted (this issue)). This is followed by six publications that describe the nonclinical assessment of THS2.2 and THS2.2M (Kogel et al, submitted (this issue); Oviedo et al, submitted (this issue); Schaller et al, submitted (this issue)-a; Schaller et al, submitted (this issue)-b; Sewer et al, submitted (this issue); Wong et al, submitted (this issue)). The eighth publication in the series describes a clinical study to assess whether the reduced formation of Harmful and Potentially Harmful Constituents (HPHC) by THS2.2 also leads to a reduced exposure to HPHCs when the product is used in a clinical setting (Haziza, submitted (this issue)). The present study utilizes data gathered from the reduced exposure clinical study on THS2.2 to determine whether a systems pharmacology approach can identify exposure response markers in peripheral blood of smokers switching to THS2.2.

While quitting is the best way to reduce the risk of smoking related diseases, modified risk tobacco products (MRTPs) may reduce the health risks compared with combustible cigarettes (CC). Philip Morris International and its affiliates (PMI) have developed the THS2.2 as an alternative to combustible cigarette (CC) smoking with the aim to mimic nicotine delivery and other

characteristics of CC smoking (Smith et al, submitted (this issue)). Heating of tobacco produces an aerosol that is significantly less complex and contains lower levels and fewer harmful and potentially harmful constituents (HPHC) compared with smoke resulting from the burning of tobacco (Borgerding & Klus, 2005; Forster et al, 2015). In fact, it has been reported that a short term switching to an earlier version of THS demonstrated beneficial impact on biomarkers of exposure (Sakaguchi et al, 2014), as well as peripheral markers such as the amounts of neutrophils, lymphocytes, platelets, red blood cells and hematocrit, within three days (Roethig et al, 2010).

The newer version of THS used in this series of studies (Smith et al, submitted (this issue)), the THS2.2, was tested in a randomized controlled study in confinement that was conducted in Poland. This study showed that adult smokers that switched to THS2.2 showed reduced levels of biomarkers of exposure to a certain number of HPHCs (Haziza, submitted (this issue)).

Since chronic exposure to cigarette smoke leads to a multifaceted response, no single response marker is sufficient to assess the system-wide effect of smoking cessation or switching to a candidate MRTP. We have therefore conducted this study to identify additional potential exposure response markers that can be used in clinical studies aimed at assessing the effects of switching to THS2.2.

To take advantages of new technologies, such as whole genome microarrays that can profile thousands of molecular species, a sample needs to contain enough cellular material. Frequently, the primary target tissue of an exposure is not easily available for such profiling. Among the easily accessible body fluids, peripheral blood contains sufficient cellular material to allow transcriptomics profiling and can be applied to establish systemic exposure response markers. Blood-based gene signatures have previously been used to identify a disease (Poliska et al, 2011; Rotunno et al, 2011; Zander et al, 2011; Zeller & Blankenberg, 2013) or subjects exposed to

harmful environmental inputs (Bushel et al, 2007; Chauhan et al, 2014; Dressman et al, 2007; Forrest et al, 2005). We have recently identified a whole blood-based gene signature with the potential to distinguish subjects who smoked from those who stopped smoking (former smokers, FS) or never smoked (non-smokers, NS) from current smokers (CS). The signature consists of 11 genes, *LRRN3*, *SASH1*, *PALLD*, *RGL1*, *TNFRSF17*, *CDKN1C*, *IGJ*, *RRM2*, *ID3*, *SERPING1* and FUCA1, some of which have been previously described in the context of smoking. Validated on independent datasets, the signature of smoking exposure response could distinguish current smokers from non-current smokers with remarkable accuracy and the signature translated well across species (Martin et al, 2015).

In this study, the goal was to apply the previously developed signature (Martin et al, 2015) to determine whether the impact of switching to THS2.2 could already be detected in the whole blood transcriptome after 5 days, and thereby provide a sensitive and non-invasive tool for assessing the exposure response in clinical studies. We used this signature, instead of extracting the signature from the data collected during the 5 day clinical study, to test whether a transcriptomics based exposure response signature that was derived from a dataset comprising current and never smokers, and able to distinguish current smokers from long-term ex-smokers, could also serve as an indicator of cessation and switching in short-term studies. Therefore, we applied the signature to the whole blood gene expression data collected from smokers who were randomized to continue smoking CCs, to abstain from smoking, or to switch to THS2.2 for 5 days in confinement ((Haziza, submitted (this issue)). Herein, we show the results from blood transcriptomics analysis of subject samples at baseline and after 5 days of the different study arms, including smoking abstinence, continuous smoking and switching to THS2.2. Based on Linear Discriminant Analysis (LDA), we

show that the whole blood-based exposure response signature of the THS2.2 switching groups is closer to the smoking abstinence groups than to the continuous smoking groups.

Materials and Methods

The clinical study on reduced exposure.

The reduced exposure in confinement study conducted in Poland was a randomized, controlled, open-label, 3-arm parallel group, single-center study to demonstrate reductions in exposure to selected smoke constituents found in CC (Haziza, submitted (this issue)). The study included healthy smokers that switched to THS2.2, abstaining from smoking, or continued to use CC, for 5 days in confinement (Figure 1). The CC study arm involved 41 participants, the smoking abstinence arm involved 39 participants and the THS2.2 arm involved 80 participants. The study protocol was approved by the local ethics committee and adhered to the principles of the Declaration of Helsinki (World Medical Association (WMA), 2013), Good Clinical Practice guidelines (International Conference on Harmonisation, 1996) and national regulations. The study has been registered in ClinicalTrials.gov under the identifier NCT01959932 (https://clinicaltrials.gov/ct2/show/NCT01959932?term=philip+morris&rank=15).

In addition to the Informed Consent Form (ICF) for the participation in the study, subjects were provided with information, and asked for their consent to collect blood samples for bio-banking and subsequent transcriptomics profiling. The blood sampling for transcriptomics and the data related to these samples were anonymized. Anonymized data and samples were initially single or double coded where the link between the subjects' identifiers and the unique code(s) was subsequently deleted.

RNA isolation

Samples were fully randomized for the RNA extraction and Affymetrix gene expression microarray. Total RNAs (including microRNAs) were isolated using the PAXgene Blood miRNA Kit (Qiagen) according to the manufacturer's instructions. The concentration and purity of the RNA samples were determined using an UV spectrophotometer (NanoDrop ND1000; Thermo Fisher Scientific, Waltham, MA, USA) by measuring the absorbance at 230, 260, and 280 nm. The RNA integrity was further checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Only RNAs with an RNA integrity number 6< were processed for further analysis.

RNA preparation and Affymetrix hybridization

Affymetrix probe sets targeting the 3' ends of transcripts were prepared from 80 ng of RNA using the Ovation® Whole Blood Reagent and Ovation® RNA Amplification System V2 (NuGEN, San Carlos, CA, USA). The quantity of amplified cDNA was measured with a Spectramax UV reader (Molecular Devices, Santa Clara, CA, USA). The cDNA quality was determined by assessing the size of un-fragmented cDNA using a Fragment Analyzer (Ankeny, IA, USA). The size distribution of the final fragmented and biotinylated product was also monitored using an Agilent 2100 Bioanalyzer (Agilent Technologies). After labeling the cDNA, the fragments were hybridized to the GeneChip® Human Genome U133 Plus 2.0 Array (Affymetrix) according to the manufacturer's guidelines. After the investigation of artifacts on the chip scan, the data were processed through a standard quality control pipeline. Briefly, raw data files (.CEL) were read using the ReadAffy function of the *affy* package (Gautier et al, 2004) from the Bioconductor suite

of microarray analysis tools (Gentleman et al, 2004) available for the R statistical environment (R Development Core Team, 2007).

Population level analysis

For the population-level analysis (i.e., study of the average fold-changes), the data were subsequently processed using the custom Chip Description File environment HGU133Plus2_Hs_ENTREZG v16. (Dai et al, 2005) and normalized using frozen-Robust Microarray Analysis (fRMA) (McCall et al, 2010) for all arrays passing quality control checks. An overall linear model, accounting for the longitudinal design of the study, was fitted for each comparison using the time of sampling as blocking to generate raw *P*-values for each gene based on moderated t-statistics (Smyth, 2005). The Benjamini-Hochberg False Discovery Rate method was used to correct for multiple testing effects.

Individual sample prediction modeling

To achieve robustness in the prediction model, independent gene expression datasets from blood were leveraged according to previously published methodology (Martin et al, 2015). Two transcriptomics datasets obtained from whole blood samples (GSE15289) and peripheral blood mononuclear cells (PBMCs; GSE42057) were obtained from the National Center for Biotechnology Information Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/gds/?term=GEO) and processed as described above. The NOWAC study (GSE15289) (Dumeaux et al, 2010) dataset was composed of whole blood samples from post-menopausal women aged between 48 and 63 years, and included 211 NS and 74 CS. The Bahr et al. (GSE42057) (Bahr et al, 2013) dataset included PBMC samples collected from 42 control subjects and 94 subjects with chronic obstructive disease (COPD) of varying severity. All

subjects were non-Hispanic Caucasians, and either CS or FS. These subjects were used to identify genes that exhibited large changes in the average expression between samples from CS and NS or FS in each dataset. The average gene expression changes between CS and FS were used to guide signature extraction from the BLD-SMK-01 dataset (Martin et al, 2015) as follows:

L_1 and L_2 were the set of the M (M=1,000 in our study) highest fold-change genes from the two independent datasets (GSE15289 and GSE42057). The L_1 and L_2 lists were then used for *a priori* filtering of the training dataset as follows:

1. For each N from 1 to M, the performance of a linear discriminant analysis (LDA) (McLachlan, 2004) model was evaluated using 5-fold cross-validation (100 times) on $L_1[1:N] \cap L_2[1:N]$, by computing the Matthews Correlation Coefficient (MCC), thus leading to MCC(N). MCC combined all true/false positive and negative rates, thus providing a single valuation metric.
2. N was selected for which the MCC(N) was maximum: $N_{max}=\text{argmax}_N(MCC(N))$.
3. The core gene list for the signature was defined by $L_1[1:N_{max}] \cap L_2[1:N_{max}]$.
4. The model was built by computing an LDA model on the core gene list.

The establishment, validation, and translation of the exposure signature has been leveraged in many datasets and is described in detail in (Martin et al., 2015).

Results

Gene expression changes in the blood of subjects of the study arms

After checking the quality of the CEL files, there were 33, 35, and 66 profiles that remained in the CC, SA and THS2.2 arms, respectively. As described in part 8 of this series (Haziza et al., submitted (this issue)), no marked differences were observed in the age, body mass index,

distributions of genders, or daily cigarette consumption. Therefore, none of these covariates were accounted for in the gene-level models. The longitudinal effects between day 1 and day 6 in each study arm are shown as volcano plots in Figure 2. The overall transcriptomic response in the blood of subjects that switched to THS2.2 for five days included several downregulated and upregulated genes. The response was not as strong as observed for the SA arm. Differentially expressed genes were not detected in the whole blood of subjects that continued using their standard tobacco product.

Performance of the smoking exposure response signature in the cessation and THS2.2 switching arms

The approach to establish and test the performance of the smoking exposure response signature in blood samples is illustrated in Figure 3. Briefly, the genes in the signature and the associated model were established using two public datasets (GSE15289, GSE42057) and a training set as described previously (Martin et al, 2015). The identified signature was then verified in the former smokers group of the training dataset, as well as on never smokers, former smokers, asymptomatic current smokers and smokers with COPD of independent study (Martin et al, 2015). To assess the performance of the smoking exposure response signature, an LDA model was applied and the scores of the gene signature for the different arms of the study are displayed in Figure 4. There was no significant difference between day 5 of the CC arm and day 0 (baseline) of all three CC, SA and THS2.2 arms, whereas a significant decrease in the LDA score was computed for both SA and THS2.2 arms at day 5, compared with their respective baselines. The relative decrease for such a longitudinal study was computed for each subject in the study based on the signature scores.

Predicted relative change in the LDA scores derived from the blood transcriptomics at the end of the study are shown in Figure 5. The observed decrease in the score between day 0 and day 5 was

normalized to the observed 10th to 90th quantile range of the signature score in training set used for signature establishment (Martin et al, 2015). Therefore, a relative change of -100% would be achieved when a smoker's score is at the 90th quantile of the training set smoker group at day 0 and ends at the 10%-quantile of the former (>1 year) smoker group at day 5. Within this conservative range, the observed average decreases were 18% and 9% for the SA and THS2.2 arms respectively.

Discussion

To assess the reduced impact associated with candidate MRTPs compared with CC, there is a need to establish sensitive markers that detect changes in non-invasive samples derived from human subjects switching to an alternative product. Standard measures to assess the impact are the biomarkers of exposure, effective dose and biological effect (Prasad et al, 2015; Shepperd et al, 2015). However, a more global picture can be achieved by using high-throughput measurements, such as the whole blood gene expression profiling.

In this study, a previously established blood-based exposure response signature was utilized to assess changes in exposure response upon smoking abstinence and switching to THS2.2 for a short period of time. Remarkably, in our study, the changes in signature scores were evident after only 5 days of abstinence and switching to THS2.2, which demonstrated that we could provide a very sensitive readout for switching effects. In addition, the blood from the subjects who remained on cigarettes during the study, showed no statistically significant gene expression changes, confirming that the observed effects were indeed due to smoking abstinence and switching to THS2.2. Upon quantifying the relative decrease of the scores, we observed approximately a 20% decrease in the smoking abstinence arm and approximately a 10% decrease in the THS2.2 arm after 5 days. Overall, these results further support the reduced exposure

associated with switching to THS2.2 compared with continuous smoking of the CC product as described in part 8 of this series Haziza, submitted (this issue)). This is in line with the observation that the biomarkers of exposures (except those associated with nicotine exposure) were significantly reduced in the THS2.2 arm compared with the CC arm, and approached the levels observed in the SA arm (Haziza, submitted (this issue)). Importantly, the decrease in LDA score observed for the THS2.2 arm reflected the strong reduction of exposure to HPHCs in THS2.2 and not reduced consumption, because as discussed by Haziza and coworkers (Haziza, submitted (this issue)), the daily product use was not reduced in the THS2.2 arm compared with the CC arm. In fact, a slight increase in consumption was observed for the THS2.2 arm.

As discussed in (Martin et al, 2015), this eleven gene signature allows the use of qRT-PCR. While Affymetrix gene expression profiling lead to a robust signature, it may not be the method of choice for using the signature in practical applications. Therefore the assay could potentially be developed into a kit with savings in cost and time.

The non-invasive sampling of whole blood allows longitudinal monitoring of smoke exposure effects and is independent of individual variations. In addition to being very sensitive to the change in exposure status, whole blood genome profiling covers at least some of biological responses to exposure and thereby complement the biomarkers of exposure which measure HPHCs and/or their metabolites. Challenges still exist in fully establishing the blood-based signature as a reliable multi-use biomarker of exposure. Ideally, the signature should be sensitive enough to distinguish the subjects that switch in a real-world setting to multiple different products or those who continue to use both the conventional product and the candidate MRTP (Krautter & Borgerding, 2014). At this time, we have not tested the signature in such populations.

Conclusion

The whole blood-based 11-gene signature of smoking exposure response that could distinguish current smokers from non-current smokers was applied to samples derived from the reduced exposure study described in part 8 of this series Haziza, submitted (this issue). The signature performed remarkably well in predicting significant reductions in exposure response within just 5 days after subjects switched to THS2.2 or abstained from smoking. Hence, blood transcriptomics profiling serves as a complementary measure of exposure response, and unlike the biomarkers of exposure, is not limited to the direct measurements of HPHCs.

Acknowledgements.

The authors acknowledge the support of Katarzyna Jarus-Dziedzic, MD, PhD, BioVirtus Clinical Research Sp. z o.o. (Kajetany, Poland) for collecting the blood samples for the study.

Conflict of interest statement

The work reported in all nine parts of this supplement involved a candidate Modified Risk Tobacco Product developed by Philip Morris International (PMI) and was solely funded by PMI. All authors are (or were) employees of PMI R&D or worked for PMI R&D under contractual agreements.

References

Bahr TM, Hughes GJ, Armstrong M, Reisdorph R, Coldren CD, Edwards MG, Schnell C, Kedl R, LaFlamme DJ, Reisdorph N (2013) Peripheral Blood Mononuclear Cell Gene Expression in Chronic Obstructive Pulmonary Disease. ***American journal of respiratory cell and molecular biology***

Borgerding M, Klus H (2005) Analysis of complex mixtures – Cigarette smoke. ***Exp Toxicol Pathol*** **57: 43-73**

Bushel P, Heinloth A, Li J, Huang L, Chou J, Boorman G, Malarkey D, Houle C, Ward S, Wilson R (2007) Blood gene expression signatures predict exposure levels. ***Proceedings of the National Academy of Sciences*** **104: 18211-18216**

Chauhan V, Howland M, Wilkins R (2014) Identification of gene-based responses in human blood cells exposed to alpha particle radiation. ***BMC medical genomics*** **7: 1**

Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. ***Nucleic acids research*** **33: e175-e175**

Dressman HK, Muramoto GG, Chao NJ, Meadows S, Marshall D, Ginsburg GS, Nevins JR, Chute JP (2007) Gene expression signatures that predict radiation exposure in mice and humans. ***PLoS Med*** **4: e106**

Dumeaux V, Olsen KS, Nuel G, Paulssen RH, Børresen-Dale A-L, Lund E (2010) Deciphering normal blood gene expression variation—The NOWAC postgenome study. ***PLoS genetics*** **6: e1000873**

Family Smoking Prevention and Tobacco Control Act, Division a—family smoking prevention and tobacco control act. Public Law No. 111-131 (2009)

Forrest MS, Lan Q, Hubbard AE, Zhang L, Vermeulen R, Zhao X, Li G, Wu Y-Y, Shen M, Yin S (2005) Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. ***Environmental health perspectives*****: 801-807**

Forster M, Liu C, Duke MG, McAdam KG, Proctor CJ (2015) An experimental method to study emissions from heated tobacco between 100-200 degrees C. ***Chemistry Central journal*** **9: 20**

Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy---analysis of Affymetrix GeneChip data at the probe level. ***Bioinformatics*** **20: 307-315**

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. ***Genome Biol*** **5: R80**

Haziza C (submitted (this issue)) Evaluation of the Tobacco Heating System 2.2. Part 8: 5-Day randomized reduced exposure clinical trial in Poland.

International Conference on Harmonisation, 1996. ICH Harmonised Tripartite Guideline for Good Clinical Practice. Available from:

http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf (Accessed on 14 August 2014).

Kogel U, Titz B, Schlage WK, Nury C, Martin F, Oviedo A, Lebrun S, Elamin A, Guedj E, Trivedi K, Ivanov NV, Vanscheeuwijck P, Peitsch MC, Hoeng J (submitted (this issue)) Evaluation of the Tobacco Heating System 2.2. Part 7: Systems toxicological assessment of a mentholated version revealed reduced cellular and molecular exposure effects compared with mentholated and non-mentholated cigarette smoke.

Krautter GR, Borgerding MF (2014) Comparison of Consumption Patterns, Biomarkers of Exposure and Subjective Effects in Cigarette Smokers Switched to Dissolvable Tobacco (Camel Orbs), Dual Use, or Tobacco Abstinence. *Nicotine & Tobacco Research*: ntu082

Martin F, Talikka M, Hoeng J, Peitsch M (2015) Identification of gene expression signature for cigarette smoke exposure response—from man to mouse. *Human & experimental toxicology* 34: 1200-1211

McCall MN, Bolstad BM, Irizarry RA (2010) Frozen robust multiarray analysis (fRMA). *Biostatistics* 11: 242-253

McLachlan G (2004) *Discriminant analysis and statistical pattern recognition*, Vol. 544: Wiley Interscience.

Oviedo A, Lebrun S, Kogel U, Ho J, Tan WT, Titz B, Leroy P, Vuillaume G, Bera M, Martin F, Rodrigo G, Ivanov NV, Hoeng J, Peitsch MC, Vanscheeuwijck P (submitted (this issue)) Evaluation of the Tobacco Heating System 2.2. Part 6: 90-day OECD 413 rat inhalation study with systems toxicology endpoints demonstrates reduced exposure effects of a mentholated version compared with mentholated and non-mentholated cigarette smoke.

Poliska S, Csanky E, Szanto A, Szatmari I, Mesko B, Szeles L, Dezso B, Scholtz B, Podani J, Kilty I (2011) Chronic obstructive pulmonary disease-specific gene expression signatures of alveolar

macrophages as well as peripheral blood monocytes overlap and correlate with lung function. *Respiration* 81: 499-510

Prasad G, Jones BA, Chen P, Gregg EO (2015) A cross-sectional study of biomarkers of exposure and effect in smokers and moist snuff consumers. *Clinical Chemistry and Laboratory Medicine (CCLM)*

R Development Core Team. (2007) R: A Language and Environment for Statistical Computing.

Roethig HJ, Koval T, Muhammad-Kah R, Jin Y, Mendes P, Unverdorben M (2010) Short term effects of reduced exposure to cigarette smoke on white blood cells, platelets and red blood cells in adult cigarette smokers. *Regulatory toxicology and pharmacology : RTP* 57: 333-337

Rotunno M, Hu N, Su H, Wang C, Goldstein AM, Bergen AW, Consonni D, Pesatori AC, Bertazzi PA, Wacholder S (2011) A gene expression signature from peripheral whole blood for stage I lung adenocarcinoma. *Cancer prevention research* 4: 1599-1608

Sakaguchi C, Kakehi A, Minami N, Kikuchi A, Futamura Y (2014) Exposure evaluation of adult male Japanese smokers switched to a heated cigarette in a controlled clinical setting. *Regulatory toxicology and pharmacology : RTP* 69: 338-347

Schaller J-P, Keller D, Poget L, Pratte P, Kaelin E, McHugh D, Cudazzo G, Smart D, Tricker AR, Gautier L, Yerly M, Pires RR, Bouhellec SL, Hofer I, Ghosh D, Garcia E, Vanscheeuwijck P (submitted (this issue)-a) Evaluation of the Tobacco Heating System 2.2. Part 2: Chemical composition, genotoxicity, cytotoxicity, and physical properties of the aerosol.

Schaller J-P, Pijnenburg JPM, Ajithkumar A, Tricker AR (submitted (this issue)-b) Evaluation of the Tobacco Heating System 2.2. Part 3: Influence of the tobacco blend on the formation of harmful and potentially harmful constituents in the aerosol.

Sewer A, Kogel U, Talikka M, Wong ET, Martin F, Xiang Y, Guedj E, Ivanov N, Hoeng J, Peitsch MC (submitted (this issue)) Evaluation of the Tobacco Heating System 2.2. Part 5: MicroRNA expression from a 90-day rat inhalation study indicates reduced effects of the aerosol on lung tissue compared with cigarette smoke exposure.

Shepperd CJ, Newland N, Eldridge A, Haswell L, Lowe F, Papadopoulou E, Camacho O, Proctor CJ, Graff D, Meyer I (2015) Changes in levels of biomarkers of exposure and biological effect in a controlled study of smokers switched from conventional cigarettes to reduced-toxicant-prototype cigarettes. *Regulatory Toxicology and Pharmacology* 72: 273-291

Smith M, Clark B, Luedicke F, Schaller J-P, Vanscheeuwijck P, Hoeng J, Peitsch MC (submitted (this issue)) Evaluation of the Tobacco Heating System 2.2. Part 1: Description of the System and the Scientific Assessment Program.

Smyth GK (2005) Limma: linear models for microarray data. In *Bioinformatics and computational biology solutions using R and Bioconductor*, pp 397-420. Springer

Wong ET, Kogel U, Veljkovic E, Martin F, Xiang Y, Boue S, Vuillaume G, Leroy P, Guedj E, Rodrigo G, Ivanov NV, Hoeng J, Peitsch MC, Vanscheeuwijck P (submitted (this issue)) Evaluation of the

Tobacco Heating System 2.2. Part 4: 90-day OECD 413 rat inhalation study with systems toxicology endpoints demonstrates reduced exposure effects compared with cigarette smoke.

World Medical Association (WMA) (2013) Declaration of Helsinki - Ethical principles for medical research involving human subjects.

Zander T, Hofmann A, Staratschek-Jox A, Classen S, Debey-Pascher S, Maisel D, Ansén S, Hahn M, Beyer M, Thomas RK (2011) Blood-Based Gene Expression Signatures in Non–Small Cell Lung Cancer. *Clinical Cancer Research* 17: 3360-3367

Zeller T, Blankenberg S (2013) Blood-based gene expression tests promises and limitations. *Circulation: Cardiovascular Genetics* 6: 139-140

Figures

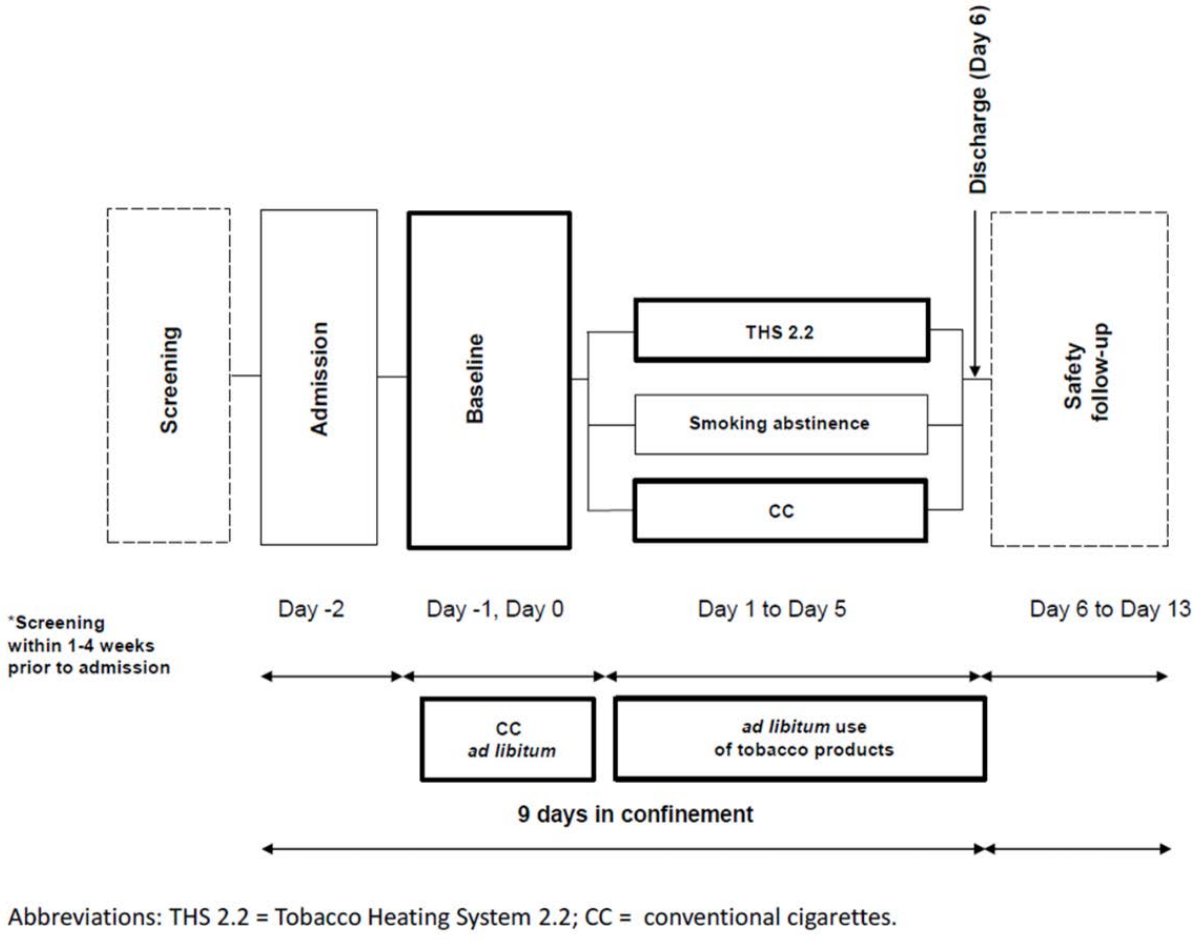


Figure 1. Study design. Open-label, 3-arm parallel group study with discharge at day 5 for each arm (CC, SA and THS).

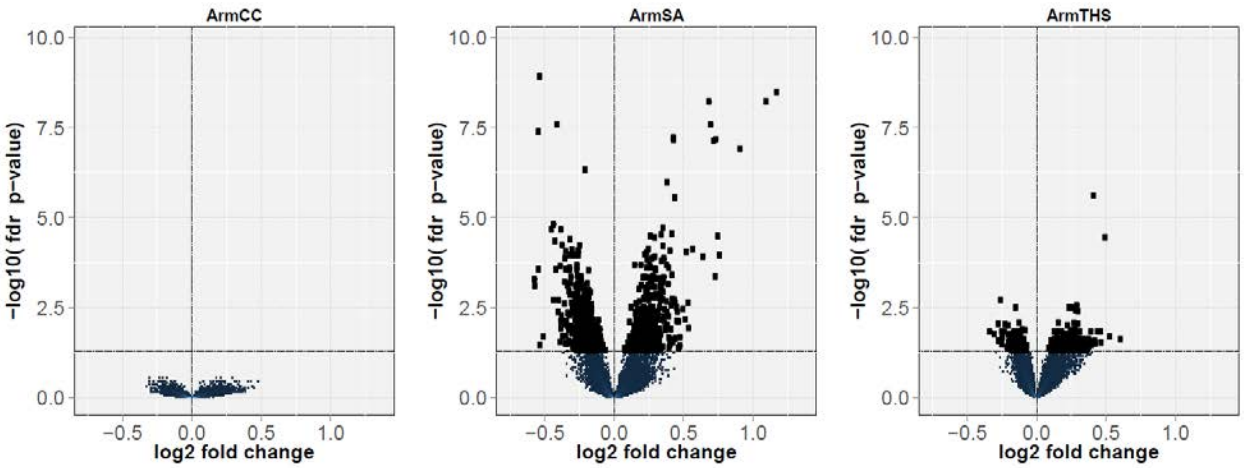
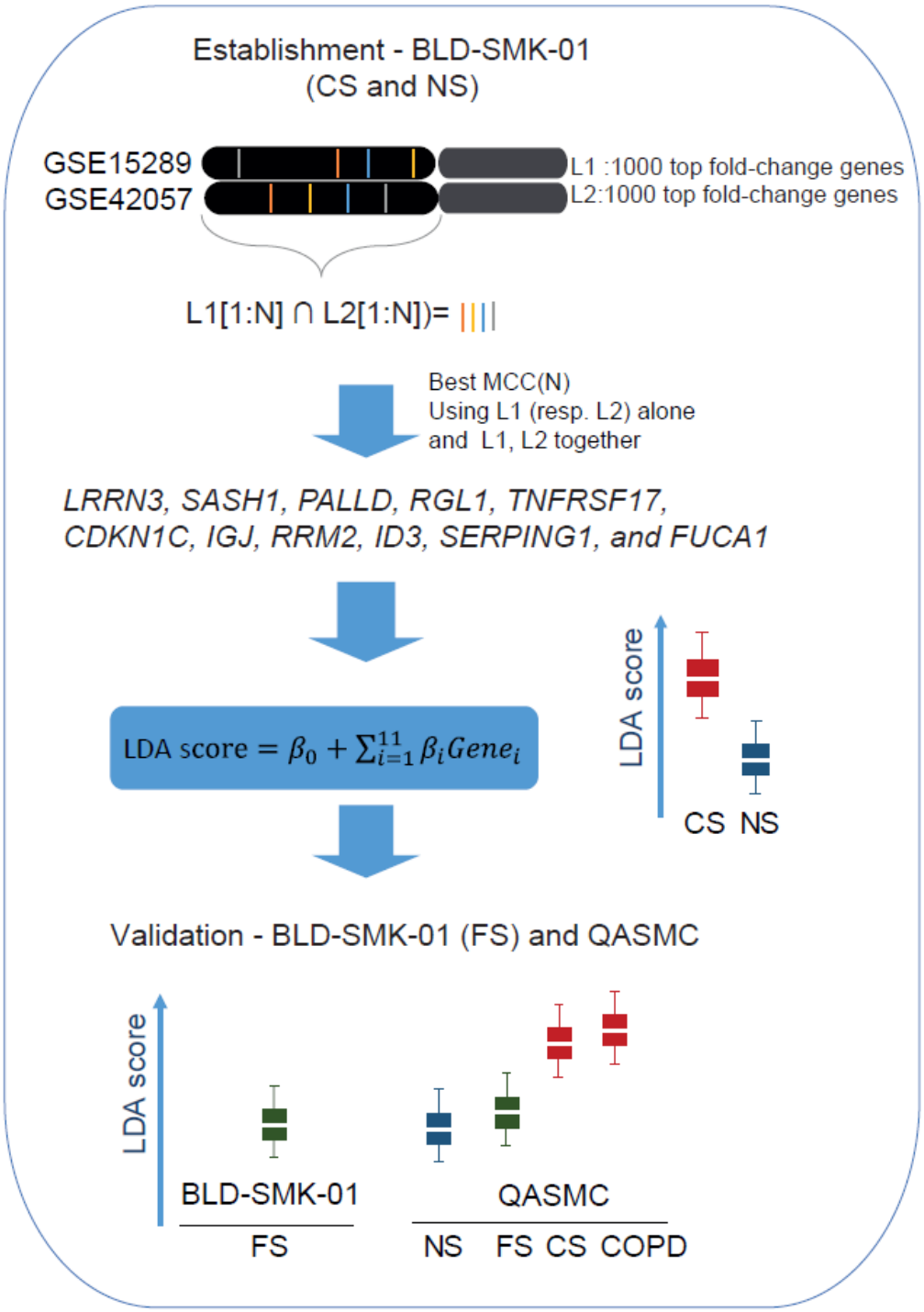


Figure 2. Volcano plots for the differentially expressed genes. Volcano plots showing the estimated $\log_2$ (fold-change) against $-\log_{10}$ (adjusted *P*-value) for the longitudinal effect. *P*-values were computed based on moderated t-statistics and were adjusted by the Benjamini–Hochberg method. The horizontal line indicates the -log10 (adjusted P-value), which correspond to the false discovery rate, FDR 0.05. Left panel: Longitudinal effect (Day 5 – Day 0) for the CC arm. Middle panel: Longitudinal effect (Day 5 – Day 0) for the SA arm. Right panel: Longitudinal effect (Day 5 – Day 0) for the THS2.2 arm.



429

430 **Figure 3. Schematic overview of the signature establishment and validation.** The set of genes and the model were
431 defined using two public datasets (GSE15289, GSE42057) and the training set (BLD-SMK-01) described previously
432 (Martin et al, 2015). The obtained signature was verified in the former smokers (FS) group of the training dataset
433 (BLD-SMK-01), as well as on never smokers (NS), former smokers (FS), asymptomatic current smokers (CS) and
434 smokers with COPD (COPD) an independent study (QASMC) (Martin et al, 2015).

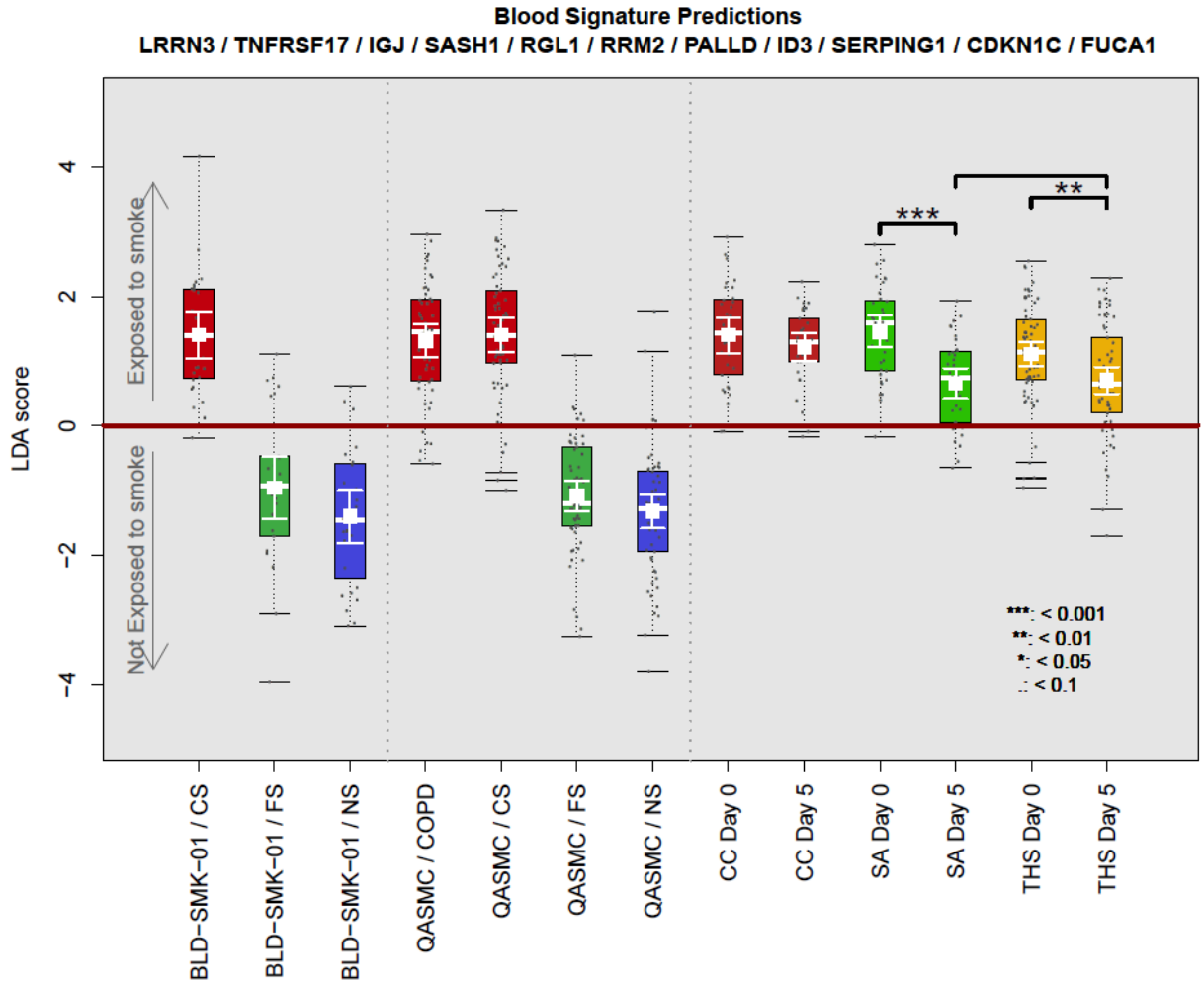
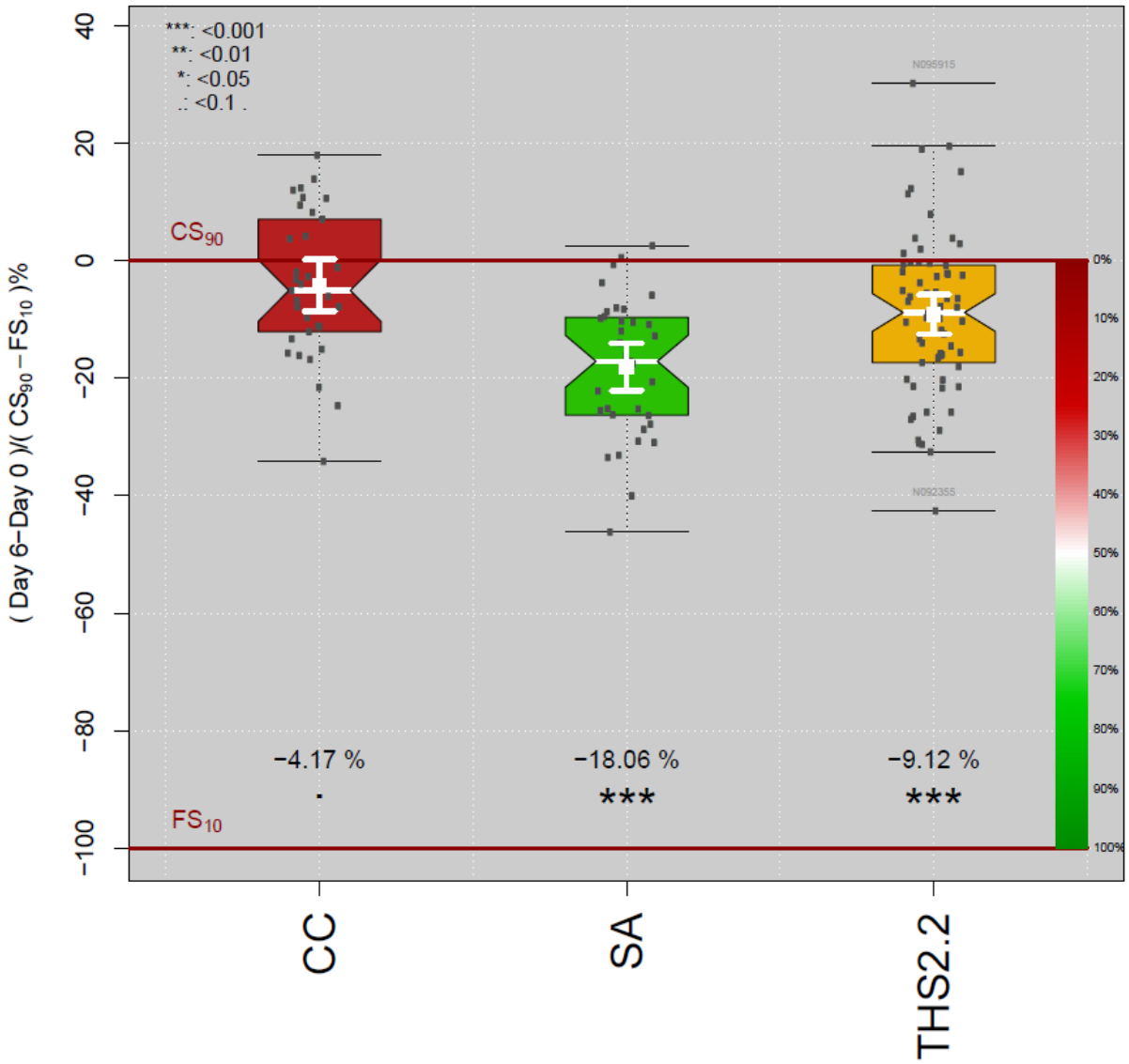


Figure 4. LDA scores for the blood-based gene signature on the different arms in the study.
The training (BLD-SMK-01) and validation (QASMC) data sets used for the signature establishment study (Martin et al. 2015) are shown on the left as the reference for typical LDA scores for CS, NC, and FS. A positive score is predictive of a CS status, and a negative score indicates a non-CS status. The mean prediction scores of each arm are marked by a white square, together with its 95%-confidence interval. The median prediction scores of each arm is indicated by horizontal segments. The key comparison t-test statistics are displayed on top of the boxplots. Abbreviations: LDA = linear discriminant analysis; CS = current smokers; NS = non-smokers; FS = former smokers; CC = conventional cigarette; SA = smoking abstinence; THS = Tobacco Heating System 2.2.



445

446 **Figure 5. Predicted relative change at the end of the study.**
447 Relative change for each subject in the study based on the signature scores. The observed change is normalized to the
448 observed 10th to 90th quantile range of the signature score in the training set described previously (Martin et al, 2015).
449 A relative decrease of 100% is therefore achieved when a smoker is at the 90%-quantile of the BLD-SMK-01 smoker
450 group at Day 0 and ends at the 10%-quantile of the BLD-SMK-01 former smoker group at day 5. A one-sample t-test
451 for the observed relative decrease was performed in each arm and the significance is displayed as: *** (p-value <0.001),
452 ** (p-value <0.01), * (p-value <0.05), . (p-value <0.1).

453

454

455