Evaluation of molecular profiling platforms in clinical pharmacology

Doorn, M.B.A. van

REFERENCE LIST

1 Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; **270**:467-470
2 Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat.Genet.* 1999; **21**:33-37
3 Crow MK, Wohlgemuth J. Microarray analysis of gene expression in lupus. *Arthritis Res Ther* 2003; **5**:279-287
4 Kaab S, Barth AS, Margerie D *et al.* Global gene expression in human myocardium-oligonucleotide microarray analysis of regional diversity and transcriptional regulation in heart failure. *J Mol Med* 2004; **82**:308-316
5 van der Pouw Kraan TC, van Gaalen FA, Kasperkovitz PV *et al.* Rheumatoid arthritis is a heterogeneous disease: evidence for differences in the activation of the STAT-1 pathway between rheumatoid tissues. *Arthritis Rheum.* 2003; **48**:2132-2145
6 Waddell SJ, Stabler RA, Laing K, Kremer L, Reynolds RC, Besra GS. The use of microarray analysis to determine the gene expression profiles of Mycobacterium tuberculosis in response to anti-bacterial compounds. *Tuberculosis (Edinb.)* 2004; **84**:263-274
7 Dyrskjot L, Thykjaer T, Kruhoffer M *et al.* Identifying distinct classes of bladder carcinoma using microarrays. *Nat.Genet.* 2003; **33**:90-96
8 Neo SY, Leow CK, Vega VB *et al.* Identification of discriminators of hepatoma by gene expression profiling using a minimal dataset approach. *Hepatology* 2004; **39**:944-953
9 Bergstrom J. Percutaneous needle biopsy of skeletal muscle in physiological and clinical research. *Scand J Clin Lab Invest* 1975; **35**: 609-616
10 Luo L, Salunga RC, Guo H *et al.* Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat.Med* 1999; **5**:117-122
11 Shaw KJ, Miller N, Liu X *et al.* Comparison of the changes in global gene expression of Escherichia coli induced by four bactericidal agents. *J Mol Microbiol.Biotechnol.* 2003; **5**: 105-122
12 Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; **45**:255-268.

CHAPTER 6

Evaluation of gene expression profiles as biomarkers for the pharmacological effects of thiazolidinediones in type 2 Diabetes Mellitus patients and healthy volunteers

---

SUBMITTED

M. B. A. van Doorn[1], Q.S. Li[2], X. Liu[3], D. M. Ouwens[4], E. J. van Hoogdalem[5], A. Bittner[3], M.J.B. Kemme[1], C. R. Jones[6], J. A. Romijn[7], J. Burggraaf[1] and A. F. Cohen[1]

1 *Centre for Human Drug Research, Leiden, The Netherlands*
2 *Johnson & Johnson Pharmaceutical Research & Development, Raritan, NJ, USA*
3 *Johnson & Johnson Pharmaceutical Research & Development, La Jolla, CA, USA*
4 *Leiden University Medical Centre; Department of Molecular Cell Biology, Leiden, The Netherlands*
5 *Johnson & Johnson Pharmaceutical Research & Development, Beerse, Belgium*
6 *Biopharma – PS, Basel, Switzerland;*
7 *Leiden University Medical Centre; Department of Internal Medicine, Leiden, The Netherlands.*

Abstract

We explored the usefulness of global gene expression profiling as a method to obtain a broad array of molecular biomarkers that could be useful for the early clinical evaluation of novel thiazolidinediones in patients with type 2 Diabetes Mellitus patients (T2DM) and healthy volunteers (HVs). Additionally, we explored the baseline differences in gene expression profile between T2DM patients and HVs. We used spotted cDNA microarrays to evaluate the expression of 6268 genes in skeletal muscle, adipose tissue, and peripheral blood leukocytes (PBLs) of T2DM patients and HVs receiving rosiglitazone (RSG) 4 mg or placebo twice daily for 6 weeks. None of the genes in adipose tissue, muscle tissue or PBLs of T2DM or HVs responded to treatment with RSG *vs.* placebo. In contrast, baseline comparisons revealed several genes in muscle and adipose tissue that were significantly differentially expressed between T2DM patients and HVs. Consequently, we conclude that gene expression profiling using spotted cDNA microarray technology is currently unable to provide a broad array of molecular biomarkers for the action of thiazolidinediones that are useful for the early clinical evaluation of novel compounds. Nonetheless, the technology appears suited for disease exploration, and may yield new starting points for future hypothesis driven investigations.

Introduction

The bottleneck in delivering new or improved pharmacological treatments appears to have shifted from discovering new chemical entities i.e. 'drug discovery', to evaluating and choosing which of these warrant follow-up and valuable resource allocation, i.e. 'drug evaluation' [1]. A key element in this process is the use of suitable efficacy biomarkers in early clinical development - 'proof of concept' - studies.

For efficient early clinical drug evaluation, there is a high demand for rapid responding biomarkers that adequately, and with great specificity, indicate the presence or absence of the desired pharmacological response in relatively small groups of subjects. It has become evident that a wider array of 'mechanism-based' biomarkers (i.e. 'combinatorial biomarkers' [2]) can be used for better, 'knowledge-based' decision-making, i.e. to stop the development of nonviable drug candidates thereby freeing resources for potentially more successful candidates [3].

Newly developed (microarray-based) global gene expression analyses seek to capture the broad information about the effects of a drug in a specific molecular fingerprint. These analyses appear particularly useful for drugs, such as peroxisome proliferator activated receptor (PPAR) agonists, that are known to regulate gene expression at the cell nucleus [4] and are therefore expected to induce rapid and readily measurable changes in mRNA abundance within their target tissues.

At present, there are no clearly defined biomarkers that are mechanism specific for *in vivo* activation of PPARγ in T2DM patients, and that are also useful in healthy volunteers (HVs) [5]. In patients with T2DM, parameters of glucoregulation are useful biomarkers, but none of these are specific to activation of PPARγ, usually take longer (6 weeks) to respond, and do not capture the pleiotropic effects that are believed to be associated with this class of drugs [6;7]. As microarray technology has shown to be an effective method to screen new compounds for *in vitro* drug effects in cell lines [8;9] it is anticipated that the technology may also have the capacity to rapidly indicate therapeutic efficacy based on tissue gene expression profiles in small groups of subjects, and thus has potential utility in early clinical development 'proof of concept' studies.

The outstanding question is whether any one or more of the necessary methods, including tissue sampling, RNA extraction and data analysis methods, yield sufficiently robust data to generate reliable results in the setting of a clinical pharmacology trial. Therefore, in the present study we sought to address the following questions:

1 Can we find statistically robust changes in gene expression in adipose tissue and skeletal muscle samples of T2DM patients after 2 and 6 weeks of *in vivo* treatment with the PPARγ agonist rosiglitazone *vs.* placebo that could serve as 'treatment fingerprint'?

2 If found, can we identify a similar PPARγ 'treatment fingerprint' in HVs who can be more easily recruited and studied than patients?

3 Can the anticipated changes in the PPARγ target tissues also be observed in peripheral blood leukocytes (PBLs) isolated from whole blood samples?

4 Can we identify potentially disease related changes ('disease fingerprint') by comparing the baseline tissue and PBLs gene expression profiles of T2DM and HVs?

Subjects and Methods

This was a randomized, double blind, placebo-controlled, multiple oral dose study. Eight male and eight female T2DM patients uncontrolled by

diet alone, aged between 40 and 75 years, with a body mass index (BMI) >25 kg/m², increased fasting plasma glucose (FPG) concentrations >7.0 mm, and C-peptide > 0.17 nmol/l were included in the study. Patients were excluded if they had a significant medical history or current symptoms of clinically relevant conditions, or had used any NSAID, thiazolidinedione or insulin preparation within two weeks of the expected study start date.

In addition, eight male and eight female HVs (as determined by medical history, physical examination and routine laboratory tests), aged between 18 and 45 years, were included.

The subjects were studied in an 8-week period consisting of 6 Visits (figure 1, p. 212). Within one week after medical screening, all subjects started with a two-week single blind placebo run-in period in which they were 'washed out' from their regular antidiabetic co-medication. At the end of the run-in period (Visit 2; baseline) they were randomly assigned to a six-week treatment with 4 mg RSG twice daily or matched placebo. Subsequently, all subjects visited the study centre after 2, 4 and 6 weeks of treatment.

On each Visit, an intravenous cannula was inserted in a forearm vein while the subject was in a supine position. Blood samples for plasma efficacy parameters (including fasting plasma concentrations of glucose, insulin, C-peptide and free fatty acids), and PBLs transcriptomic analyses were collected on all Visits after approximately 30 min of supine rest as previously described [10]. Two pre-dose blood samples were collected for PBLs gene expression analysis; one sample was collected at the beginning of the run-in period (Visit 1) and one at the end of the run-in period (Visit 2). In addition, on Visit 2, two blood samples were collected 3.5 and 10 hrs post-dose to assess potentially immediate gene expression responses in PBLs, and for RSG pharmacokinetic assessments. An oral glucose tolerance test (OGTT) was performed at baseline and at the end of the active treatment period.

Skeletal muscle and abdominal subcutaneous adipose tissue samples were collected as previously described [11]. After sampling, the tissue samples were immediately snap-frozen in liquid nitrogen and stored at -80 °C until further analysis. PBLs were isolated from whole blood samples and RNA was directly extracted and stored at -80 °C until further analysis.

The protocol for this study was approved by the Medical Ethics Committee of the Leiden University Medical Centre. All subjects gave their written informed consent.

Analytical methods for plasma markers

Concentrations of typical (biochemical) efficacy parameters were determined as previously described [10].

RNA extraction and handling

RNA was extracted from the frozen tissue samples as previously described [11]. Purity and quantity of the RNA was determined by spectrophotometric analysis at 260 and 280 nm. Integrity of the RNA was verified by agarose gel electrophoresis.

Microarrays

Human cDNAs were spotted onto Corning GAPS slides using an Amersham Biosciences Generation III spotter. Each clone was spotted in duplicate on the array, with the exception of control clones that were spotted 4 to 12 times. The custom-made microarray contained 8263 replicated spots, representing 6268 unique human genes/ESTs. The Cy3-labeled cDNA probe preparation, hybridization and subsequent washes of the arrays were performed as previously described [11]. All arrays were scanned in an Agilent scanner model G2505B (Agilent Technologies, Inc., Wilmington, DE). We performed quantification using Imagene software (BioDiscovery, El Segundo, CA).

Data normalization

VARIANCE-STABILIZING TRANSFORMATION (VST)

We hybridized each RNA sample to two microarray slides run in parallel, yielding quadruplicate data points for each sample. For analysis, the intensities of the quadruplicate data set in each sample were scaled to the 75th percentile (the 75th percentile value of each data set was set to 100), followed by log-linear hybrid transformation, a VST described by Rocke and Durbin [12]. The average of the quadruplicate transformed intensities of each spot was designated as the transformed expression intensity in the RNA sample.

NORMALIZATION AMONG SAMPLES WITHIN EACH TISSUE

A sorted nonlinear smooth spline normalization procedure [13] was applied on all samples from the same tissue. Concordance correlation coefficient was calculated as previously described [14], among all samples from the same tissue.

Statistical Analysis

INCLUDED DATA

All subjects that dropped out before Visit 4 (three subjects: one male and two female T2DM patients) were replaced with newly recruited subjects receiving the same treatment. Data of subjects dropping out after Visit 2 (baseline) were included in the statistical analysis even when a subject was replaced. Consequently, data of subjects dropping out during the placebo run-in period (one female T2DM patient) were not used in the analysis.

This resulted in 217 blood RNA samples, 86 muscle RNA samples and 80 fat RNA samples that were analyzed by microarray. The hybridization of the blood RNA samples was carried out in five hybridization batches on chips from thirteen print lots, whereas the hybridization of the muscle RNA samples was carried out in three hybridization batches from six print lots, and the hybridization of the fat RNA samples was carried out in four hybridization batches from seven print lots.

ANALYSIS OF VARIANCE (ANOVA)

Variances in gene expression contributed by different factors were analyzed by ANOVA. For expression in skeletal muscle and adipose tissue at baseline (before treatment), ANOVA was performed on VST intensities of each clone in baseline samples using the formula, VST intensity ~ Disease status * Sex + Print lot. For differential gene expression in skeletal muscle and adipose tissue after treatment, a delta of expression in each clone between VST intensity of either post 2 week or post 6 week and that of the baseline was calculated first for each subject. Subsequently, ANOVA was performed on the delta of expression (either post 2 week or post 6 week) of each clone in available subjects using formula, Delta ~ Treatment code * Disease * Sex + Print lot.

MIXED MODEL ANALYSIS

Differentially expressed gene-lists were constructed using mixed model analysis as implemented in Partek Pro™ (Partek Inc. St. Louis, Missouri). The results obtained with Partek Pro™ were consistent with results using the SAS System (SAS Institute Inc. Cary, NC) in a small subset of genes tested. Specifically, the log-linear hybrid VST data were analyzed in a mixed model with both fixed factors, i.e. disease status, treatment regimen, time course, sex, and random factors, i.e. subject, print lot, hybridization batch. Restricted maximal likelihood (REML) method was used in variance estimation. As a first pass, a general model considering main effects of fixed factors and interaction among fixed factors were fitted for each of the 8352 clones in the microarray. Bonferroni adjusted p-values for both main effects and interactions were used to rank the genes to identify the differentially expressed genes (α=0.05).

GENE NETWORK AND FUNCTIONAL ANALYSES

Both gene network and functional analyses were performed via the use of Ingenuity Pathways Analysis (IPA; Ingenuity® Systems, www.ingenuity.com).

The Ingenuity Pathways Knowledge Base (IPKB) collects millions of gene to gene interactions, both direct and indirect, as a result of manual curation from the literature. Given a list of significantly differentially expressed genes that were selected from a data set by a particular p-value cut-off, networks consisted of 35 genes, based on the interactions collected in the IPKB, are selected to uncover the maximum number of input genes. A statistic score, equivalent to minus log P value, is given for each selected network to reflect the probability of selecting such a network by random chance. This score is dependent both on the number of input genes and on the number of genes (named as focus genes) in the selected network.

For a given list of genes, either directly selected from an entire data set, or from focus genes in a selected network, the functional analysis identifies the biological functions and/or diseases that were most significant to the selected genes. Fischer's exact test is used to calculate a p-value determining the probability that each biological function and/or disease assigned to the selected genes is due to chance alone.

Results

Baseline characteristics and effects on plasma markers

Detailed demographics, baseline characteristics of both study populations and effects on plasma markers were published previously [10]. Briefly, seven T2DM patients were withdrawn: one patient developed a clinically significant elevated triglyceride concentration (12.9 mmol/l; RSG treatment group), five patients had repetitive measurements of glucose exceeding 15 mmol/l (four patients in the placebo group and one patient in the RSG treatment group) and one patient was hospitalized (severe bronchitis) during the placebo run-in period. Three (placebo treated) T2DM patients were replaced.

Both patient and HVs groups showed good compliance and had equal exposure to RSG. We observed significant decreases in FPG, fructosamine, insulin, C-peptide, IL-6 and WBC after 6 weeks of treatment with RSG *vs.* placebo, whereas free fatty acid levels and lipid / lipoprotein parameters remained virtually unchanged [10]. There were no clinically significant changes in safety parameters.

Gene expression profiling

Thirteen blood samples and three muscle samples were identified as outliers using the methodology previously described [11] and were excluded from the statistical analysis. The total number of muscle and adipose samples with high quality microarray data is summarized in table 1. Variances in gene expression contributed by different factors were analysed by ANOVA (Methods). Most of the variance in the baseline samples of adipose and muscle tissues could be ascribed to disease status and print lot (figures 2a and 2b, p. 213). Specifically, 5 genes in skeletal muscle and 8 genes in adipose tissue were differentially expressed between the T2DM and HVs group at a Bonferroni adjusted p-value cut-off (α) of 0.05 (data not shown). In addition, ANOVA analysis of baseline samples revealed three genes that were differentially expressed between male and female in both skeletal muscle and adipose tissue (data not shown). Little variance was related to treatment code (RSG *vs.* placebo, figures 2c and 2d, p. 213), indicating the relatively weak effect of RSG treatment on gene expression levels in adipose and muscle tissue. Similar results were obtained for delta of gene expression corresponding to post 6-week treatment (data not shown).

The large variance introduced by print lot had to be corrected or taken into account in the subsequent statistical analyses. A mixed model with repetitive measurements was chosen to calibrate for print lot effect and to obtain valid conclusions in the presence of missing data points.

In line with the results of the initial ANOVA analysis, the mixed model analysis revealed 10 genes in skeletal muscle and 20 genes in adipose tissue, that were differentially expressed between T2DM patients and HVs at a Bonferroni adjusted p-value cut-off (α) of 0.05 (table 2). The identity of these genes is listed in table 3a and 3b. Among these genes, 2 genes were identified in skeletal muscle and 3 genes in adipose tissue, both in the ANOVA and mixed model analyses (table 3a and 3b; genes indicated with an asterisk).

The mixed model did not identify genes in any of the two tissues or PBLs that showed a significant RSG-treatment response at Bonferroni adjusted p-value cut-off of 0.05 (table 2).

TABLE 1 **Summary of tissue samples with high quality gene expression data per Visit and treatment group.** Missing data was modelled using Partek mixed model analysis. DF: Diabetic Female; HF: Healthy Female; DM: Diabetic Male; HM: Healthy Male; SK: Skeletal Muscle; AT: Adipose Tissue.

Treatment	Group	Baseline		Post 2wk		Post 6wk	
		SM	AT	SM	AT	SM	AT
Placebo	DF	4	4	6	3	1	3
Placebo	HF	3	5	3	3	2	4
Placebo	DM	4	4	5	4	1	1
Placebo	HM	3	3	4	3	4	3
Rosiglitazone	DF	4	4	3	4	3	3
Rosiglitazone	HF	4	2	2	5	3	4
Rosiglitazone	DM	4	4	4	2	3	3
Rosiglitazone	HM	3	4	3	1	4	3
	Sub-total	29	30	30	25	21	24

TABLE 2 **Significant genes in ANOVA mixed model analysis.** Number of genes in each tissue that were significant for each of the factors assessed in the mixed model ANOVA analysis after correction for multiple comparisons (Bonferroni) and the number of genes significant when using a less stringent raw p-value cut-off of p=0.001 is indicated.

# of genes	Muscle tissue		Adipose tissue		PBLs	
	Bonferroni; p=0.05	p=0.001	Bonferroni; p=0.05	p=0.001	Bonferroni p=0.05	p=0.001
Treatment	0	1	0	2	0	2
Disease	10	124	20	418	0	18
Disease*Treatment	0	4	0	0	0	9
Disease*Gender	1	65	0	22	0	7

Exploratory pathway analysis

To allow exploratory pathway analyses of differentially expressed genes using IPA, we applied a less stringent 'raw' p-value cut-off of p=0.001. At this cut-off, RSG-treatment significantly affected the expression of 1 gene in skeletal muscle tissue, 2 genes in adipose tissue, and 2 genes in PBLs (table 2). Evidently, this small number of differentially expressed genes was insufficient to allow meaningful pathway analyses.

However, when comparing baseline profiles at a raw p-value cut-off of p=0.001, a large number of differentially expressed genes between T2DM patients and HVs was identified in both adipose and skeletal muscle tissue, thus allowing subsequent gene network and canonical pathway analyses (table 2). The number of differentially expressed genes at the p=0.001 cut-off only yielded 18 differentially expressed genes in PBLs which was considered insufficient for pathway analysis (table 2).

TABLE 3A **A. Genes differentially expressed between T2DM patients and HVs in adipose tissue at baseline ('disease fingerprint'). B. Genes differentially expressed between T2DM patients and HVs in skeletal muscle at baseline ('disease fingerprint').** Identity, gene bank accession number, magnitude and direction of the genes that were differentially expressed between the two study groups (T2DM patients and HVs) in adipose tissue and skeletal muscle tissue using a mixed model analysis with (Bonferroni) correction for multiple comparisons and a p-value cut-off of 0.05. Genes with an asterisk were also selected as differentially expressed using the initial ANOVA analysis on baseline samples. *There were no genes significantly differentially expressed in PBLs.

Adipose tissue

Gene symbol	Description	Accession	Mean (N-T2DM)	Up/down in T2DM	p-value	Bonferroni corrected p-value
IDH3G*	isocitrate dehydrogenase 3 (NAD+) gamma	NM 004135.2	1.116	↑	4.99E-09	4.17E-05
AC003112	brain-specific membrane-anchored protein	3945935H1	-1.482	↓	3.24E-07	2.71E-03
KCNV1	potassium channel, subfamily V, member 1	NM 014379.2	0.375	↓	7.19E-07	6.00E-03
TRPM4	transient receptor potential cation channel, subfamily M, member 4	NM 017636.2	-1.076	↑	1.05E-06	8.77E-03
ITGA5	integrin, alpha 5 (fibronectin receptor, alpha polypeptide)	NM 002205.2	-0.660	↑	1.25E-06	1.04E-02
PLAT	plasminogen activator, tissue	NM 000930.2	-0.964	↑	1.57E-06	1.31E-02
ANK1	ankyrin 1, erythrocytic	NM 020477.1	-0.812	↑	1.92E-06	1.60E-02
SPHK1	sphingosine kinase 1	NM 021972.2	-1.094	↑	2.17E-06	1.81E-02
N.A.*	Homo sapiens cDNA: FLJ21333 fis, clone COL02535	AK 024986	0.700	↓	2.30E-06	1.92E-02
CLDN7	claudin 7	NM 001307.3	0.515	↓	2.76E-06	2.30E-02
TNR	tenascin R (restrictin, janusin)	NM 003285.1	-0.512	↑	2.82E-06	2.36E-02
TRPM4	transient receptor potential cation channel, subfamily M, member 4	NM 017636.2	-0.803	↑	3.27E-06	2.73E-02
ECHDC1	enoyl coenzyme A hydratase domain containing 1	NM 018479.2	0.796	↓	3.50E-06	2.93E-02
RAMP3	receptor (calcitonin) activity modifying protein 3	NM 005856.1	-1.080	↑	4.02E-06	3.36E-02
N.A.	N.A.	5546719H1	0.428	↓	4.41E-06	3.68E-02
ADAM15	a disintegrin and metalloproteinase domain 15 (metargidin)	NM 003815.3	-0.643	↑	4.93E-06	4.12E-02
RNF5*	ring finger protein 5	NM 006913.2	0.405	↓	5.00E-06	4.18E-02
WISP2	WNT1 inducible signaling pathway protein 2	NM 003881.2	-1.588	↑	5.22E-06	4.36E-02
N.A.	zh27d04.s1 soares_pineal_gland_N3HPG Homo sapiens cDNA clone	AA772499	0.531	↓	5.43E-06	4.53E-02
N.A.	N.A.	3406131T6	0.535	↓	5.73E-06	4.78E-02

TABLE 3B

Muscle tissue

Gene symbol	Description	Accession	Mean (N-T2DM)	Up/down in T2DM	p-value	Bonferroni corrected p-value
BMPR2	bone morphogenetic protein receptor, type II (serine/threonine kinase)	NM 001204.4	-0.583	↑	4.93E-07	4.12E-03
AXIN2*	axin 2 (conductin, axil)	NM 004655.2	-0.958	↑	5.00E-07	4.18E-03
TNFSF7	tumor necrosis factor (ligand) superfamily, member 7	NM 001252.2	-0.652	↑	9.31E-07	7.77E-03
LR8	LR8 protein	NM 0140202	-0.506	↑	1.04E-06	8.69E-03
COL13A1	collagen, type XIII, alpha 1	NM 005203.3	-0.532	↑	1.46E-06	1.22E-02
PLCE1	phospholipase C, epsilon 1	NM 016341.2	-0.802	↑	2.81E-06	2.35E-02
ADAM1	ADAM metallopeptidase domain 1 (fertilin alpha)	Y09232	-0.763	↑	4.06E-06	3.39E-02
ICEBERG*	ICEBERG caspase-1 inhibitor	NM 021571.2	-0.803	↑	4.75E-06	3.97E-02
ATP5G1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 1	NM 005175.2	0.479	↓	4.91E-06	4.10E-02
N.A.	N.A.	1512407H1	-0.715	↑	4.99E-06	4.17E-02

Subsequently, we varied the selection criterion (p-value cut-off for the number of genes in the input list) in order to attain the highest score for the top ranking network(s) in the adipose tissue and skeletal muscle dataset. A top score of 59, corresponding with 200 focus genes, was reached for two gene networks in the adipose tissue dataset. Similarly, a top score of 26 corresponding with 100 focus genes was reached for 1 network in the skeletal muscle dataset (table 4).

TABLE 4 **Variation of selection criterion to attain the highest score for the top ranking gene network(s) in the adipose tissue and skeletal muscle datasets.** The number of input genes derived from the baseline comparison between the T2DM patients and HVs group at baseline, number of genes from this list that is covered in Ingenuity Pathway Analysis (focus genes), top score of gene network and the number of focus genes in each gene network are given in relation to the p-value cut-off that was applied. NA: not available; IPA: Ingenuity Pathway Analysis.

Selection Criterion	Number of genes	Number of genes covered in IPA	Top score of gene network	Number of focus genes in gene network
Adipose tissue				
p < 0.001	418	273	54	35
p < 0.00059	**308**	**200**	**59**	**35**
p < 0.0001	111	67	27	17
Muscle tissue				
p < 0.01	213	104	26	18
p < 0.006	**207**	**100**	**26**	**18**
p < 0.001	124	58	24	15

The top ranking networks and indications of potentially important canonical pathways within each top network for both adipose tissue and skeletal muscle tissue are presented in figure 3a; p. 214, figure 3b; p. 215 and figure 4; p. 216, respectively. In addition, tables 5 and 6 provide an overview of the different functions covered by the genes within the two top ranking networks for each tissue.

Subsequently, a global pathway analysis using all networks constructed from the top 200 and top 100 focus genes for adipose tissue and skeletal muscle, respectively, identified several potentially disease related canonical pathways. The top 10 disease related canonical pathways for adipose tissue and skeletal muscle are presented in figures 5a and 5b; p. 217, respectively.

In adipose tissue, the functional analysis identified several genes implicated in lipid and carbohydrate metabolism as well as several

T2DM-associated genes (table 5). In addition, the 2 top rating gene networks and canonical pathways identified in the adipose tissue dataset, harbour genes (i.e. INSR, SLC2A4, PPARG and RPS6KB1) that are either components of the insulin signalling system or have been characterized as modulators of insulin action (figures 3b; p. 215 and 5a; p. 217). Moreover, the chemokine, NF-KB, and P38MAPK canonical pathways or components thereof, like the up regulated pro-inflammatory cytokine TNFα (central component of adipose tissue gene network 1; figure 3a; p. 215), have been recognized as negative regulators of molecular insulin action [15]. Importantly, PPARγ, the target for TZD class drugs which activation is associated with improvements in insulin sensitivity, was identified in gene network 2 (figure 3b; p. 215).

Although the number of focus genes that could be used to build gene networks and canonical pathways in skeletal muscle was smaller compared to adipose tissue, genes implicated in lipid metabolism (table 6), and canonical pathways involved in the negative modulation of insulin signalling were identified, including the IL-6, NF-KB, and P38MAPK-pathways (figure 5b; p. 217).

Discussion

In this study we could not identify genes in adipose tissue, skeletal muscle tissue or PBLs of T2DM and HVs that showed significant changes in expression after 6 weeks of treatment with RSG *vs.* placebo after adjustment for multiple comparisons (table 2). In contrast, we did identify several novel potentially T2DM-associated genes that were significantly differentially expressed in adipose tissue and skeletal muscle between the T2DM patients and HVs at baseline (tables 3a and 3b). In addition, subsequent exploratory analyses of these tissues corroborated previous reports with regard to the regulation of several T2DM-associated genes, functions and canonical pathways (tables 4 and 5; figures 3a, 3b; p. 215, 4; p. 216 and 5; p. 217).

Although the practical emphasis in pharmacological gene expression profiling to date has been to determine *in vitro* drug-toxicity profiles, *in vivo* drug-efficacy profiles may have utility in several important areas. One apparent application is to better and possibly faster evaluate candidate drugs in early clinical development studies with patients or HVs. In addition, studying drug effects in both T2DM patients and HVs, creates an opportunity to compare the baseline gene expression profiles and identify putative disease associated genes ('disease fingerprint')

TABLE 5 Functional analysis of top 200 focus genes within the adipose tissue data set. If the association between the input list and biological function has a p-value less than 0.05 the function and/or disease category was considered significant (Fisher's exact test).

Function	Gene symbol	p-value	Number of function genes
Lipid Metabolism			
accumulation of lipid	CCL2, DBI, IL1RN, LPL, PPARD, PPARG, TNF	4.2e-6	7
release of lipid	BTK, CCL2, CHRM4, GNA11, IL1RN, PRKCB1, TNF	1.3E-5	7
production of lipid	CCL2, F2, IL1RN, P2RX7, PLAA, RALA, TNF	1.5E-5	7
Carbohydrate Metabolism			
uptake of monosaccharide	CSF1, CTF1, IRS1, LPL, PPARG, PRKCB1, SLC2A4, TNF	6.9E-7	8
uptake of d-glucose	CSF1, CTF1, IRS1, LPL, PPARG, SLC2A4, TNF	9.1E-7	7
uptake of 2-deoxyglucose	IRS1, PPARG, PRKCb1, SLC2A4, TNF	1.6E-5	5
transport of carbohydrate	ABCC3, F2, IRS1, PDGFRB, SLC2A4, TNF	1.0E-5	6
accumulation of carbohydrate	ABCC3, CD44, CRHR1, EDNRB, GNA11	3.1E-5	5
Metabolic Disease			
diabetes associated	HLA-DRA, ICAM1, IRS1, LPL, PPARG, SLC2A4, SOD1, TNF	2.4E-6	8

TABLE 6 Functional analysis of top 100 focus genes within the skeletal muscle data set. If the association between the input list and biological function has a p-value less than 0.05 the function and/or disease category was considered significant (Fisher's exact test).

Function	Number of functional category genes	p-value functional category genes	Focus gene(s) symbol	p-value focus gene(s)
Lipid Metabolism				
release of lipid	8	3.8e-9	alox5ap, clec11a	0.016
release of eicosanoid	7	3.1E-8	ALOX5AP, CLEC11A	0.012
release of arachidonic acid	6	9.2E-8	CLEC11A	0.11
synthesis of leukotriene	3	3.7E-5	ALOX5AP	0.03
synthesis of eicosanoid	4	4.8E-5	ALOX5AP	0.10

and biological pathways. Hence, this type of study design not only has the potential to identify a broad array of mechanism-based molecular biomarkers for pharmacological effects, but it could also offer important new insights in the pathogenesis of a disease.

The main objective of the present study was to identify molecular biomarker patterns that could serve as efficacy parameters for the pharmacological effects of TZDs in human tissues and thus provide us with potentially better, more sophisticated tools for the early clinical evaluation of future candidate compounds. To this end, we conducted

a small intensive clinical study that investigated the pharmacological effects of RSG (as 'prototype' TZD) on global gene expression profiles in adipose tissue, skeletal muscle and PBLs of T2DM patients and HVs. In parallel, we investigated and demonstrated the efficacy of RSG on typical parameters of glucoregulation and inflammation as published previously [10]. Although we could demonstrate efficacy of RSG using traditional plasma biomarkers, we were unable to identify changes in tissue gene expression patterns that could serve as molecular biomarkers for the pharmacological effects of RSG *in vivo*.

We chose to study a relatively small number of subjects in a fairly short-term treatment period since our primary objective was to identify molecular biomarkers that would be useful for the early clinical evaluation of novel candidate compounds. Usually, these early clinical studies are so-called 'first in man studies', which aim to rapidly demonstrate 'proof of concept' while at the same time provide data on the pharmacokinetic profile and tolerability/safety of the new compound under investigation. Since there is usually no or little clinical experience with the compound, these studies usually enroll a small number of subjects who are studied during a limited exposure period. Consequently, if one wants to identify better, more sophisticated biomarkers to indicate clinical 'proof of concept' for the candidate compound at hand, these markers evidently need to respond rapidly in relatively small groups of subjects.

A potential interpretation issue of this study could lie within the dropout rate of subjects in the T2DM placebo group. Unfortunately, dropouts are inevitable in this type of study design, in which regular oral hypoglycaemic medication of T2DM patients was withdrawn. We chose withdraw this medication since concerns were raised with regard to the glucose lowering potential of RSG as add-on therapy, and to exclude a potential confounding effect of regular oral hypoglycaemic medication.

Another issue entails the use of multiple microarray print batches. These are typically needed in larger studies with multiple samples at various time points. Initial analyses indicated that a large proportion of the variance, in for example the adipose tissue data set (figure 2a; p. 213), was related to the respective print batch number. This issue is rarely encountered in smaller pre-clinical or *in vitro* studies but appears to be a major point of attention when analyzing large sample quantities on a spotted cDNA microarray platform.

To adequately deal with the large number of variables and potential interpretation issues encountered in the setting of a clinical trial, a statistical method was chosen (repeated measures mixed effect model) that could calibrate for each of these variables (including the print batch effect) and that provides an as unbiased a prediction as possible for the missing values encountered [16].

Although adjustment for multiple comparisons is obviously indicated, the conservative Bonferroni adjustment used in our primary analysis probably leads to a considerable amount of type II error and thus false negative results. For exploratory analyses, it is probably more desirable to include a number of false positives genes than to discard some truly biologically significant ones [17]. Consequently, we lowered the raw p-value to the rather non-stringent cut-off of p=0.001 in order to perform so-called 'biological filtering' (identification of global network genes or 'focus genes') of the resulting gene list using IPA. Unfortunately, also this approach failed to identify a significant number of genes that would allow the identification of gene networks or biological pathways affected by RSG treatment.

It appears that in this study design the signal-to-noise ratio is too low to adequately identify pharmacological effects on tissue gene expression levels using spotted cDNA microarray technology. This could be due to the fact that a sequence of relatively small changes in the expression of PPARγ modulated genes give rise to the previously reported more robust downstream effects as observed for several plasma markers [10]. In addition, the natural (intra- and inter-individual) variation of gene expression in time, and the modelling of missing data may be additional factors that add noise and obscure the possibly subtle biological signal. Alternatively, many of the so-called 'downstream' TZD effects may actually be controlled by posttranscriptional PPARγ-independent mechanisms, as has recently been shown for adiponectin [18;19].

The technology appears better suited for the analysis of disease related changes. Comparison of the T2DM patients and HVs at baseline identified several genes that were significantly differentially expressed in muscle and adipose tissue, even after adjustment for multiple comparisons (tables 3a, 3b and 4).

Furthermore, subsequent exploratory (IPA) analyses revealed several genes in the adipose tissue data set that have previously been identified as pivotal components of the insulin signalling system (i.e. IRS1 [insulin receptor substrate 1] and SLC2A4 [solute carrier family 2 (facilitated glucose transporter), member 4; also known as GLUT4]), or have been characterized as modulators of insulin action (i.e. RPSK6KB1 [ribosomal protein S6 kinase, 70kDa polypeptide 1], PPARG [peroxisome proliferator activated receptor gamma]) [20-24] (figure 3b; p. 215). In addition, the chemokine, NF-KB, and P38MAPK canonical pathways or components

thereof, like the pro-inflammatory cytokine TNFα, have been recognized as negative regulators of molecular insulin action [15] (figure 5; p. 217).

These confirmatory findings strengthen our belief that 'biological filtering' from a 'dirty' significant genes list is a valuable approach to identify putative T2DM-associated genes and pathways that would otherwise remain undetected using conventional statistical analyses. Unfortunately, as the biological sampling was limited in our clinical setting, we did not have sufficient RNA quantities to validate our findings using complementary approaches such as real-time RT-PCR.

Our results appear to differ from previous findings by other groups that used microarray technology, who failed to identify any differentially expressed genes between T2DM patients and controls in skeletal muscle tissue after adjustment for multiple comparisons [25;26]. This could relate to different sampling conditions, sensitivity of microarray platforms (spotted cDNA microarrays *vs.* Affymetrix oligonucleotide based microarrays), number of genes surveyed, and subsequent normalization / analysis methods, and underlines the need for proper and universal standards to allow a better comparison between studies. On the other hand, as both groups were not age-matched some of our findings may be partly attributable to differences in age. We chose to study unmatched groups since the primary objective of the study was to identify a rapid molecular treatment response in T2DM patients that would also respond in more easily studied and recruited healthy volunteers, who are usually between 18 and 30 years of age.

In conclusion, our results indicate that gene expression profiling using spotted cDNA microarrays is (in our clinical study design and wet lab design settings) not capable of capturing the broad molecular response to drug intervention with a PPARγ agonist *in vivo*. Consequently, in its current stage of development, this technology appears unable to provide more sophisticated, rapid responding molecular biomarkers for PPARγ agonists that are useful for future clinical 'proof of concept' studies with novel compounds. However, the technology will be evaluated in further clinical studies using a modified design, which may result in an improved signal to noise ratio.

Finally, we have shown that microarrays may be useful tools for disease exploration, as illustrated by the identification of several previously reported T2DM-associated genes, in parallel with the discovery of several novel putative T2DM-associated genes and pathways. These newly identified candidates may serve as new starting points for future hypothesis driven investigations, which may pave the way to a better understanding of the molecular mechanisms involved in the pathogenesis of T2DM.

REFERENCE LIST

1 Bailey WJ, Ulrich R. Molecular profiling approaches for identifying novel biomarkers. *Expert.Opin.Drug Saf* 2004; **3**:137-151

2 Koop R. Combinatorial biomarkers: From early toxicology assays to patient population profiling. *Drug Discov.Today* 2005; **10**:781-788

3 Rolan P, Atkinson AJ, Jr., Lesko LJ. Use of biomarkers from drug discovery through clinical practice: report of the Ninth European Federation of Pharmaceutical Sciences Conference on Optimizing Drug Development. *Clin Pharmacol Ther* 2003; **73**:284-291

4 Spiegelman BM. PPAR-gamma: adipogenic regulator and thiazolidinedione receptor. *Diabetes* 1998; **47**:507-514

5 Wagner JA. Early clinical development of pharmaceuticals for type 2 diabetes mellitus: from preclinical models to human investigation. *J Clin Endocrinol.Metab* 2002; **87**:5362-5366

6 Buckingham RE. Thiazolidinediones: Pleiotropic drugs with potent anti-inflammatory properties for tissue protection. *Hepatol.Res* 2005; **33**:167-170

7 Giannini S, Serio M, Galli A. Pleiotropic effects of thiazolidinediones: taking a look beyond antidiabetic activity. *J Endocrinol.Invest* 2004; **27**:982-991

8 Gunther EC, Stone DJ, Gerwien RW, Bento P, Heyes MP. Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. *Proc.Natl.Acad.Sci. U.S.A* 2003; **100**:9608-9613

9 Gunther EC, Stone DJ, Rothberg JM, Gerwien RW. A quantitative genomic expression analysis platform for multiplexed in vitro prediction of drug action. *Pharmacogenomics.J* 2005; **5**:126-134

10 van Doorn M, Kemme M, Ouwens M *et al.* Evaluation of proinflammatory cytokines and inflammation markers as biomarkers for the action of thiazolidinediones in Type 2 diabetes mellitus patients and healthy volunteers. *Br J Clin Pharmacol* 2006; **62**:391-402

11 van Doorn M, Kemme M, Ouwens M *et al.* Biopsy and RNA extraction procedures of muscle and adipose tissue for microarray gene-expression profiling.In: *Biochips as pathways to drug discovery* eds Carmen E, Hardiman G, CRC, 2006; **7**:109-122

12 Rocke DM, Durbin B. Approximate variance-stabilizing transformations for gene-expression microarray data. *Bioinformatics.* 2003; **19**:966-972

13 Shaw KJ, Miller N, Liu X *et al.* Comparison of the changes in global gene expression of Escherichia coli induced by four bactericidal agents. *J Mol Microbiol.Biotechnol.* 2003; **5**:105-122

14 Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; **45**:255-268

15 Perseghin G, Petersen K, Shulman GI. Cellular mechanism of insulin resistance: potential links with inflammation. *Int J Obes.Relat Metab Disord.* 2003; **27 Suppl 3**:S6-11

16 Little RJA, Rubin DB. *Statistical Analysis with Missing Data* J. Wiley & Sons, 1987

17 Brazeau DA. Combining genome-wide and targeted gene expression profiling in drug discovery: microarrays and real-time PCR. *Drug Discov.Today* 2004; **9**:838-845

18 Osei K, Gaillard T, Cook C, Kaplow J, Bullock M, Schuster D. Discrepancies in the regulation of plasma adiponectin and TNF-alpha levels and adipose tissue gene expression in obese African Americans with glucose intolerance: a pilot study using rosiglitazone. *Ethn.Dis.* 2005; **15**:641-648

19 Rasouli N, Yao-Borengasser A, Miles LM, Elbein SC, Kern PA. Increased plasma adiponectin in response to pioglitazone does not result from increased gene expression. *Am J Physiol Endocrinol.Metab* 2006; **290**:E42-E46

20 Carvalho E, Jansson PA, Nagaev I, Wenthzel AM, Smith U. Insulin resistance with low cellular IRS-1 expression is also associated with low GLUT4 expression and impaired insulin-stimulated glucose transport. *FASEB J* 2001; **15**:1101-1103

21 Hammarstedt A, Jansson PA, Wesslau C, Yang X, Smith U. Reduced expression of PGC-1 and insulin-signaling molecules in adipose tissue is associated with insulin resistance. *Biochem Biophys Res Commun.* 2003; **301**:578-582

22 Rondinone CM, Wang LM, Lonnroth P, Wesslau C, Pierce JH, Smith U. Insulin receptor substrate (IRS) 1 is reduced and IRS-2 is the main docking protein for phosphatidylinositol 3-kinase in adipocytes from subjects with non-insulin-dependent diabetes mellitus. *Proc.Natl.Acad.Sci.U.S.A* 1997; **94**:4171-4175

23 Um SH, Frigerio F, Watanabe M *et al.* Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity. *Nature* 2004; **431**:200-205

24 Lehrke M, Lazar MA. The many faces of PPARgamma. *Cell* 2005; **123**:993-999
25 Mootha VK, Lindgren CM, Eriksson KF *et al*. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat.Genet.* 2003; **34**:267-273
26 Patti ME, Butte AJ, Crunkhorn S *et al*. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc.Natl.Acad.Sci.U.S.A* 2003; **100**:8466-8471.

CHAPTER 7

Evaluation of global gene expression profiles as biomarkers for the pharmacological effects of fibrates in patients with type 2 Diabetes Mellitus and healthy volunteers

---

SUBMITTED

M. B. A. van Doorn[1], Q.S. Li[2], X. Liu[3], D. M. Ouwens[4], E. J. van Hoogdalem[5], A. Bittner[3], C. R. Jones[6], J. A. Romijn[7], J. Burggraaf[1] and A. F. Cohen[1]

1 *Centre for Human Drug Research, Leiden, The Netherlands*
2 *Johnson & Johnson Pharmaceutical Research & Development, Raritan, NJ, USA*
3 *Johnson & Johnson Pharmaceutical Research & Development, La Jolla, CA, USA*
4 *Leiden University Medical Centre; Department of Molecular Cell Biology, Leiden, The Netherlands*
5 *Johnson & Johnson Pharmaceutical Research & Development, Beerse, Belgium*
6 *Biopharma – PS, Basel, Switzerland;*
7 *Leiden University Medical Centre; Department of Internal Medicine, Leiden, The Netherlands.*