ORIGINAL PAPER

Gene expression signatures of angiocidin and darapladib treatment connect to therapy options in cervical cancer

Martin Koch · Michael Wiese

Abstract

Purpose   To assign functional properties to gene expression profiles of cervical cancer stages and identify clinically relevant biomarker genes.

Experimental design   Microarray samples of 24 normal and 102 cervical cancer biopsies from four publicly available studies were pooled and evaluated. High-quality microarrays were normalized using the CONOR package from the Bioconductor project. Gene expression profiling was performed using variance-component analysis for accessing most reliable probes, which were subsequently processed by gene set enrichment analysis.

Statement of translational relevance: At present, there are no generally accepted biomarkers for cervical cancer available. Here, we utilize an already approved approach for biomarker discovery on a large-scale pooled microarray dataset. Eleven highly reliable probes resulting from this analysis feature cervical cancer biomarker properties. Moreover, overlap between our gene signatures and those from studies available in the literature affirm the utility of this approach. However, the most striking findings are that the gene expression signatures were associated with signatures observed for patients treated using angiocidin and darapladib. Previous studies have shown that angiocidin induces differentiation of monocytes to macrophages and induce anti-tumoral action, whereas darapladib exerts anti-inflammatory action. Therefore, we suggest these two agents as potential candidates for novel treatment options in cervical cancer patients.

M. Koch · M. Wiese (✉)
Pharmaceutical Institute, University Bonn, An der Immenburg 4,
53121 Bonn, Germany
e-mail: mwiese@uni-bonn.de

Results   Of 22.277 probes that were subject to variance-component analysis, eleven probes had low heterogeneity, that is, a W/T ratio between 0.18 and 0.38. Seven of these probes are induced in all cervical cancer stages: they are GINS1, PAK2, DTL, AURKA, PRKDC, NEK2 and CEP55. The other four probes are induced in normal cervix: P11, EMP1, UPK1A and HSPC159. We performed GSEA of 9.873 probes exhibiting less variability, that is, having a W/T ratio of <0.75. Repeatedly, significant gene expression signatures were found that are related to treatment using angiocidin and darapladib. Additionally, expression signatures from immunological disease signatures were found, for example graft versus host disease and acute kidney rejection. Another finding comprises a gene expression signature in stage IB2 that refers to MT1-MMP-dependent migration and invasion. This gene signature is accompanied by gene expression signatures which refer to ECM receptor-mediated interactions.

Conclusion   Analysis of cervical cancer patient gene expression data reveals a novel perspective on HPV-mediated transcription processes. This novel point of view contains a better understanding and even might provide improvements to cancer therapy.

Introduction

Cervical cancer is the third most common cause of gynecological cancer death among women in the USA (Jemal et al. 2011). Worldwide, the disease accounts for the second most gynecological cancer death cases (Lowy and Schiller 2006). Cervical cancer is almost exclusively

(99 %) caused by human papilloma virus (HPV) infections. The interval between HPV infection and malignant progression takes at least 10 years (Lowy and Schiller 2006); hence, cervical cancer frequency peaks at the age of 40. However, this opens a rather long period for detection and medical intervention. At present, there is no generally accepted oncogenic marker available (Bachtiary et al. 2006). Although many microarray studies have analyzed cervical cancer expression profiles, there is little overlap in the resulting gene expression signatures. This marginal concordance may result from the diverseness of the study designs or relate to the various microarray platforms and protocols used. In addition, intratumor heterogeneity presents a different source of variation and is seldom considered. Bachtiary et al. developed a remedial measure to this situation, which is based on variance-component analysis of the genetic properties of replicate cancer biopsies (Bachtiary et al. 2006). In this preliminary study, Bachtiary et al. modeled the genetic variability of cervical cancer tumors and estimated the reliability of potential oncogenic markers based on sample size. As the authors pointed out, their results have to be corroborated by a larger study. However, an aspect, which was not taken into account, is the variability in each separate tumor stage according to the Federation International of Gynecology and Obstetrics (FIGO). Investigation of the different tumor stages would allow an even more sensitive level of prognostic capability in a potentially allocated set of oncogenic markers. Therefore, the ascertainment of expression signatures of each FIGO stage might even increase treatment opportunities since each stage may also require a different therapeutic approach.

Here, we describe a large-scale variance-component analysis of pooled microarray raw data from four publicly available cervical cancer studies. Inspired by the cutting-edge study of Bachtiary et al., we believe that intratumor heterogeneity as well as heterogeneity across tumor stages has influence on the reliability of oncogenic markers. Therefore, we wished to develop statistically as well as clinically relevant gene expression profiles of cervical cancer. In addition to variance-component analysis, we used gene set enrichment analysis (GSEA) to process the resulting profiles (Subramanian et al. 2005). From our results presented here, we derive eleven highly reliable oncogenic markers for distinction between normal and cancerous cervix tissue. Gene expression profiles of low intratumor heterogeneity were processed by GSEA and lead to expression signatures related to treatment using angiocidin and darapladib. Recently, these two agents showed anti-tumoral and anti-inflammatory properties; hence, we suggest considering angiocidin and darapladib as novel candidates for treatment of early stages cervical cancer.

Patients and methods

Publicly available cervical cancer microarray studies in GEO

Microarray raw data of four publicly available cervical cancer studies were obtained from Gene Expression Omnibus [GEO, http://www.ncbi.nlm.nih.gov/geo/ (Barrett et al. 2005)]. The following studies were evaluated for high-quality arrays. Study A was performed by the Ontario Institute for Cancer Research, GSE5787 [Bachtiary et al. (2006), on Human Genome U133A Plus 2.0]. Purpose of this study was to estimate intratumor heterogeneity in CC patients. As a result, a mixed model for the reliability of biomarkers was proposed. Study B was performed by the University of Wisconsin-Madison, GSE6791 [Pyeon et al. (2007), on Human Genome U133A Plus 2.0]. Purpose of this study was a comparison of gene expression patterns of HPV$^+$ and HPV$^-$ head and neck cancers to gene expression patterns of cervical cancers. Study C was performed by the Columbia University Medical Center, GSE9750 [Scotto et al. (2008), on Human Genome U133A]. Purpose of this study was to examine the role of the gain on the long arm of chromosome 20 in early stages of CC and elucidate amplicon functions as well as potential markers. Study D was performed by the University Medical Center Groningen, GSE26511 [Nordhuis et al. (2011), on Human Genome U133A Plus 2.0]. Purpose of this study was to identify cellular tumor pathways associated with pelvic lymph node status patients with early stages of cervical cancer (Noordhuis et al. 2011).

Patients

Pooling the above enlisted four studies, we obtained in total microarray data from 102 cervical cancer biopsies and 24 microarray samples of normal cervix. All patients were between 27 and 77 years old (median 42.5). Table 1 enlists the number of patients in each assigned tumor stage according to the Federation International of Gynecology and Obstetrics (FIGO).

Data processing

An important step before the actual analysis of microarray data comprises the normalization procedure, that is,

Table 1 Enumeration of all patients and FIGO stages obtained from GEO

NC	IB	IB1	IB2	IIA	IIB	IIIB
24	12	21	15	10	25	19

to largely remove technical variation and enable comparison of the expression values. Normalization algorithms are widely implemented in the open source R-based Bioconductor project. Here, we used R version 2.12 for all statistical analysis and computing. Because we combine microarray data from four different laboratories and even different manufacturer series, an especially careful approach to normalization has to be considered. First, we performed quality control using the *Simpleaffy* package, which leads us to omit arrays having an average background intensity of more than 75 and also to the rejection of arrays in which both housekeeping probes were considered to be off ratio (>3 fold) (Wilson and Miller 2005). Additionally, an mRNA degradation check was applied. Then, all raw data were preprocessed using quantile normalization of the *RMA* package without background adjustment (Irizarry et al. 2003). Next, the GEO datasets were merged to obtain a patient data matrix using the CONOR package that provides cross-platform normalization of microarray data (Rudy and Valafar 2011).

Analysis of intrastage variance

Inspired by the study of Bachtiary et al. (2006), we wished to access the variance-components of all different stages of cervical cancer. Therefore, we applied the provided R code on the patient matrix grouped by each distinct FIGO stage of cervical cancer. Hence, in our model, W is the variance within a stage, B corresponds to the variance between stages and T is the total variance ($T = W + B$). The ratio W/T provides a measurement of the intrastage heterogeneity.

Gene set enrichment analysis (GSEA)

In 2005, Subramanian et al. introduced the gene set enrichment analysis (GSEA) (Subramanian et al. 2005). The algorithm tests whether a gene signature is significantly enriched inside of a collection of various predefined gene sets. Such predefined gene sets comprise probes from biological pathway databases like KEGG, Biocarta and Reactome as well as all layers of Gene Ontology terms and gene signatures obtained from scientific publications. The GSEA algorithm calculates an enrichment score that is accompanied by an estimation of significance which is also adjusted for multiple hypothesis testing. Here, we applied a gene set–based permutation test using 1.000 permutations and ranked all probes according to Student's t statistic. Nominal *p* value below 0.01 and FDR less than 0.05 were considered as statistically significant. Additionally, we restricted the enrichment score to values above 0.7.

Results

Quality assessment and normalization of GEO studies yields a pooled patient data matrix

The quality assessment of microarray data is as important as microarray normalization; hence, we inspected each study individually. In study A (GSE5787) from 33 arrays, only array GSM135260 had an off scale control probe, but the overall quality was sufficient. In study B (GSE6791), only 18 out of 27 possibly useful arrays were actually considered due to quality issues. Even the considered arrays had off scale control probes, however, since the average background and the present signals were in range, these 18 arrays were considered. Study C (GSE9750) had eight out of 57 arrays comprising a very low overall signal compared to the other arrays and thus were rejected from further processing (i.e. GSM246422, GSM246423, GSM246484, GSM246485, GSM246486, GSM246487, GSM246488 and GSM246489). In study D (GSE26511), all 39 arrays were considered, since these arrays were of recommendable quality on any scale. Next, all studies were subject to a second quality check regarding mRNA degradation and all arrays passed this test. In total 131, arrays were extracted that met high-quality standards. We excluded single samples of stages IA and IIIA as well as three samples from the advanced stage IVB, since these small sample sizes were not representative. This left 126 microarrays in total for further processing. Table 1 enlists all used arrays according to quantity and FIGO stage. The different arrays were normalized first within each study by *RMA* quantile normalization without background adjustment (Irizarry et al. 2003). Then, in three subsequent merging steps, using the CONOR package, all studies were bundled to yield a single pooled patient data matrix. The resulting box-plots of the normalized microarrays, and additionally, the intrasample variance are displayed in the Supplementary Information Figure S1. All gene expression values were distributed equally after quantile normalization was applied. However, the variance of the gene expression values is preserved as shown.

Variance-component analysis of all available cervical cancer stages yields low heterogeneous probes

The patient data matrix contains 22.277 probes for each of the 126 samples that are representing one normal cervix and six histological conditions of cervical cancer according to FIGO. For accessing probes, which are less heterogeneously expressed, we performed variance-component analysis. Figure 1 displays the resulting distribution of the W/T ratio for probes from the investigated stages. The shape in red depicts the density when seven randomly
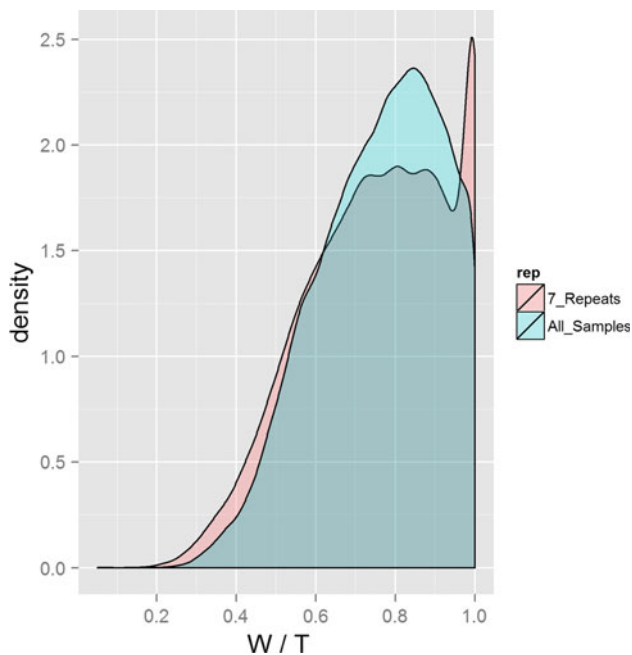
Fig. 1 The plot depicts the density against the ratio between intra stage heterogeneity (W) and total heterogeneity within stages (T). The *red shape* denotes seven replicate arrays and the shape in *blue* accesses all available replicate arrays

chosen replicates are analyzed; the shape in blue shows the distribution when all samples are incorporated into variance-component analysis. The variance-component analysis of seven randomly chosen replicates leads to a steep increase of probes which have a W/T ratio of one. However, a low ratio denotes probes that are less variable inside a specific stage. Clearly, using all available samples is of benefit and leads to a distribution that peaks at a W/T ratio of about 0.8. Next, we define a cutoff for accessing only the most reliable probes that are exhibiting less variability inside a stage. Considering probes, which have a W/T ratio of 0.75, leaves 9.873 probes for further processing.

Clustering highlights low variable probes that are constitutively expressed in all stages

Cluster analysis of 9.873 probes resulting from variance-component analysis yielded eleven probes, which are induced or repressed in cervical cancer, having a W/T ratio between 0.18 and 0.38. Seven of these probes are induced in all cervical cancer stages, these probes are GINS1, PAK2, DTL, AURKA, PRKDC, NEK2 and CEP55. The other probes are repressed in cervical cancer, namely P11, EMP1, UPK1A and HSPC159. Table 2 lists the above-stated probes, and Fig. 2 displays a heat map of these potential biomarkers, depicting induced probes in red and repressed probes in blue.

Gene set enrichment analysis of the variance-component derived expression set

Probes of 9.783 exhibit a W/T ratio of less than 0.75; these probes were analyzed using GSEA. Initially, we analyzed the normal phenotype versus all samples from stage I. Table S1 in the Supplementary Information enlists all enriched gene sets of this preliminary comparison. The majority of the enriched processes in cervical cancer are related to DNA maintenance, and there are also signatures of proliferation and metastasis. The gene signature termed "Cervical Cancer Proliferation Cluster" (CCPC) enlists 105 enriched genes which have a p value and FDR below $1e^{-4}$. This signature was found by Rosty et al. (2005); however, the authors considered genes exhibiting the highest variance; thus, in the original signature, there are 163 genes. Here, we found only genes exhibiting less heterogeneity. The gene set termed "Bidus, Metastasis Up" refers to the work of Bidus et al. (2006), the authors wished to predict lymph node metastasis in endometrioid endometrial cancer using gene expression signatures. This signature is significantly enriched in our patient matrix as

Table 2 All potential biomarkers are listed above as well as their corresponding annotation

| ID	W/T	Symbol	Description	Ref seq
206102_at	0.34	GINS1	GINS complex subunit 1 (Psf1 homolog)	NM_021067
208877_at	0.35	PAK2	p21 protein (Cdc42/Rac)-activated kinase 2	NM_002577
218585_s_at	0.35	DTL	Denticleless homolog (Drosophila)	NM_016448
208079_s_at	0.37	AURKA	Aurora kinase A	NM_198437
208694_at	0.38	PRKDC	Protein kinase, DNA-activated, catalytic polypeptide	NM_006904
204641_at	0.38	NEK2	NIMA (never in mitosis gene a)-related kinase 2	NM_002497
218542_at	0.38	CEP55	Centrosomal protein 55 kDa	NM_018131
206605_at	0.18	P11	26 serine protease	NM_006025
213895_at	0.24	EMP1	Epithelial membrane protein 1	NM_001423
214624_at	0.26	UPK1A	Uroplakin 1A	NM_007000
219998_at	0.26	HSPC159	Galectin-related protein	NM_014181
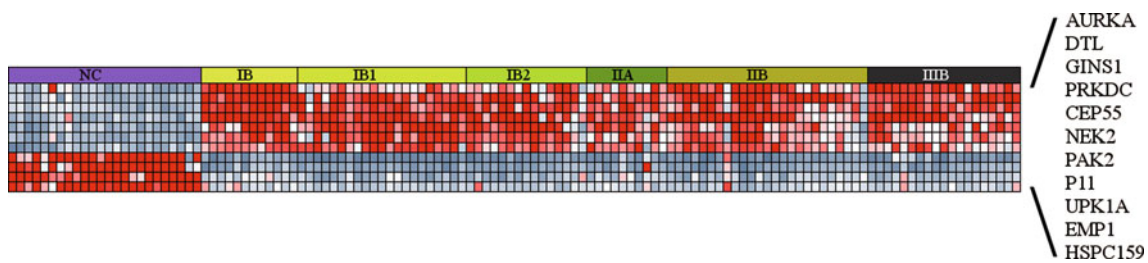
Fig. 2 Heat map of potential biomarker genes that are expressed in all FIGO stages constitutively

Table 3 All significantly enriched gene sets in stage I

IB	Gene set description	Size	ES	p value	FDR
1	RNA EXPORT FROM NUCLEUS	15	0.72	1e-4	1e-2
IB1					
1	WILENSKY RESPONSE TO DARAPLADIB	20	0.90	1e-4	1e-4
2	GAURNIER PSMD4 TARGETS	23	0.89	1e-4	1e-4
3	FLECHNER BIOPSY KIDNEY TRANSPLANT REJECTED VS OK UP	44	0.83	1e-4	1e-4
4	WIELAND UP BY HBV INFECTION	56	0.80	1e-4	1e-4
5	BROWNE INTERFERON RESPONSIVE GENES	35	0.80	1e-4	1e-4
6	REACTOME CHEMOKINE RECEPTORS BIND CHEMOKINES	23	0.78	1e-4	1e-4
7	KEGG GRAFT VERSUS HOST DISEASE	18	0.77	1e-4	1e-4
8	REACTOME TCR SIGNALING	26	0.76	1e-4	1e-4
9	CHEMOKINE ACTIVITY	16	0.76	1e-4	4e-3
10	REACTOME DOWNSTREAM TCR SIGNALING	19	0.75	1e-4	1e-4
11	G PROTEIN COUPLED RECEPTOR BINDING	17	0.74	1e-4	7e-3
12	CHEMOKINE RECEPTOR BINDING	18	0.74	1e-4	8e-3
IB2					
1	ROZANOV MMP14 TARGETS SUBSET	15	0.85	1e-4	1e-4
2	KEGG ECM RECEPTOR INTERACTION	43	0.80	1e-4	1e-4
3	REACTOME NCAM1 INTERACTIONS	28	0.78	1e-4	1e-4
4	REGULATION OF CELL MIGRATION	15	0.75	6e-3	2e-2

well, having a p value and FDR below $1e^{-4}$. These preliminary results indicate that the proposed method generates phenotype-dependent and low heterogeneous gene signatures. Next, we wished to focus on the functional properties of each stage, and therefore, we examine each single stage versus the following stage in a separate manner.

GSEA derived processes found repeatedly in stages

Tables 3 and 4 summarize the top significant-enriched gene signatures. Among, the most significant and most enriched processes are "Wilensky, Response to Darapladib" (ES 0.90) and "Gaurnier, PSMD4 Targets" (ES 0.89) in stage IB1. These signatures are repeatedly enriched in stage IIA yielding enrichment scores (ES) of 0.84 and ES 0.76. Then, the Gaurnier signature is once again enriched

in stage IIB having an ES of 0.80. We found in the literature that Wilensky et al. (2008) acquired a gene expression signature that is down-regulated after treatment with lipoprotein-associated phospholipase A_2 (Lp-PLA$_2$) inhibitor darapladib. The authors found that Lp-PLA$_2$ inhibitor reduces lesions in atherosclerosis and exerts also anti-inflammatory action by repressing a 24-gene signature that is associated with macrophage and T lymphocyte recruiting.

In stage IB1 and stage IIA, we find 20 genes of the Wilensky signature up-regulated, having significant enrichment scores (p value and FDR $< 1e^{-4}$) as shown in Fig. 3. The complete Gaurnier signature consists of 59 genes, of which we found 23 (39 %) in our data that are yielding significant enrichment scores in stage IB1, IIA and IIB as shown in Fig. 4. However, here, we found a significant enrichment of a fraction of the complete signature

Table 4 All significantly enriched gene sets in stage II and III

IIA	Gene set description	Size	ES	p value	FDR
1	WILENSKY RESPONSE TO DARAPLADIB	20	0.84	1e-4	1e-4
2	BROWNE INTERFERON RESPONSIVE GENES	35	0.79	1e-4	1e-4
3	HUMORAL IMMUNE RESPONSE	15	0.77	1e-4	3e-2
4	GAURNIER PSMD4 TARGETS	23	0.76	1e-4	1e-3
5	BIOCARTA BCR PATHWAY	17	0.73	1e-4	3e-2
6	KEGG GRAFT VERSUS HOST DISEASE	18	0.72	1e-4	1e-4
7	REACTOME DOWNSTREAM TCR SIGNALING	19	0.72	4e-3	4e-2
8	REACTOME TCR SIGNALING	26	0.71	1e-4	1e-2
9	CHEMOKINE ACTIVITY	16	0.70	1e-4	1e-2
IIB					
1	GAURNIER PSMD4 TARGETS	18	0.80	1e-4	1e-4
2	METALLOENDOPEPTIDASE ACTIVITY	19	0.78	1e-4	1e-4
3	KEGG GRAFT VERSUS HOST DISEASE	23	0.71	2e-3	2e-2
4	FLECHNER BIOPSY KIDNEY TRANSPLANT REJECTED VS OK UP	44	0.70	1e-4	1e-4
IIIB					
1	SLEBOS HEAD AND NECK CANCER WITH HPV UP	30	0.78	1e-4	1e-4
2	KEGG DRUG METABOLISM CYTOCHROME P450	32	0.70	1e-4	1e-4
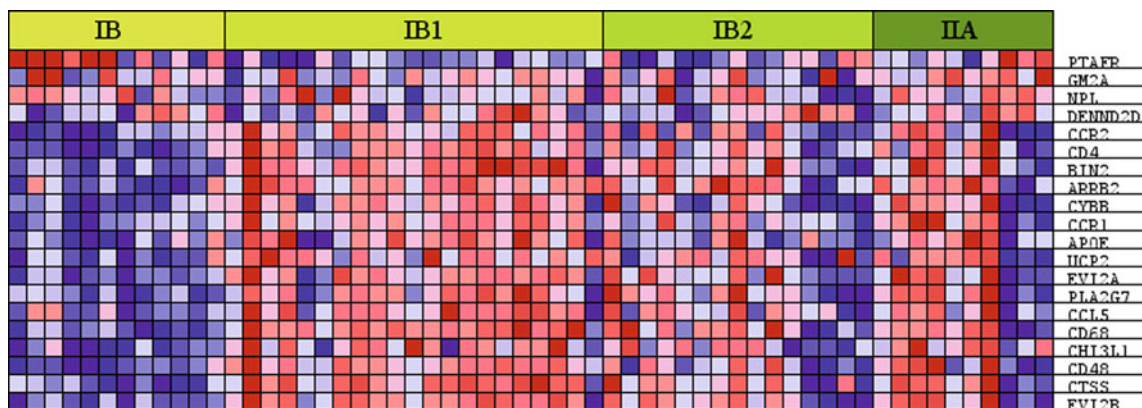
Fig. 3 Gene set enrichment analysis derived Wilensky response to darapladib signature
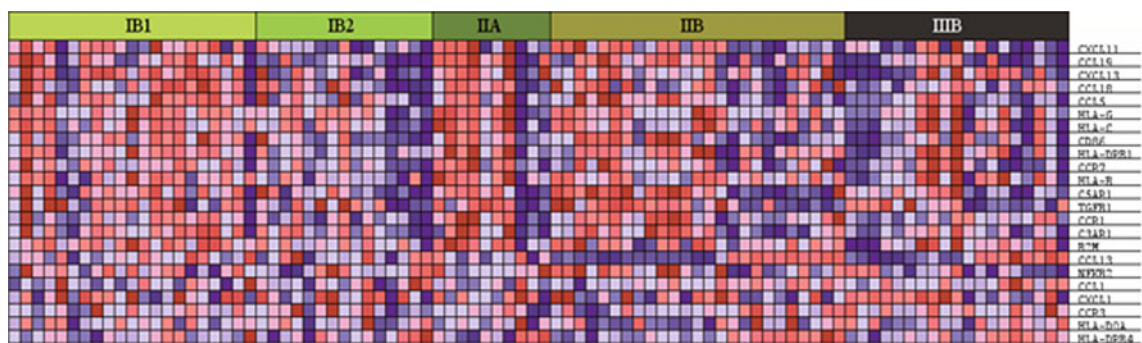
Fig. 4 Gene set enrichment analysis derived Gaurnier PSMD4 targets signature

(p value and FDR $< 1e^{-4}$). The authors showed that the activation of anti-tumoral, MMP9 secreting monocytes is induced by treatment using angiocidin, the protein product of PSMD4 (Gaurnier-Hausser et al. 2008). The Gaurnier signature is accompanied by another signature, which is called "KEGG, Graft versus Host Disease". A heat map of this gene expression signature is displayed in Figure S2 in the Supplementary Information. Graft versus host disease

(GVHD) can be characterized in three distinct phases; first, inflammatory cytokines are triggered, then interferon gamma (IFN-γ), and finally, the previously activated CTL and NK cells induce apoptosis via Fas–Fas ligand interactions. Browne et al. (2001) postulated a signature comprising 68 genes, which is specific for interferon-α responsive genes, 35 genes of this signature are enriched in our data in stage IB1 (ES 0.80) and IIA (ES 0.79). Another immunological-related gene expression signature that was first described by Flechner et al. (2004) "Biopsy Kidney Transplant Rejected VS Ok Up" was enriched in stage IB1 (ES 0.83) and stage IIB (0.70).

GSEA derived processes found uniquely in stages

The export of RNA from the nucleus is the only process, which is within significant cutoff values, that is, FDR and ES in stage IB. However, GSEA found also many less enriched processes in this stage; among them was a gene signature (ES 0.56) that was first described by Pyeon et al., the authors of study B (GSE6791). In stage IB1, we found a gene signature of 107 genes which is induced during HBV viral clearance that was introduced by Wieland et al. (2004). In our data, we found 56 of those genes induced significantly (p value and FDR $< 1e^{-4}$). The investigation of stage IB2 yielded unique gene signatures that were not found in other stages. We find a gene signature that was introduced by Rozanov et al. (2008) "MMP14 Targets Subset". The authors found that this gene expression signature refers to MT1-MMP-dependent migration and invasion. Additionally, we find also processes which point to "ECM Receptor Interaction" from KEGG and "NCAM1 Interactions" from Reactome (p value and FDR $< 1e^{-4}$) as well as "Regulation of Cell Migration" from the Gene Ontology (p value $6e^{-3}$ and FDR $< 2e^{-2}$). In stage IIA, we find uniquely the enrichment of immunologic gene expression signatures, that is, "Humoral Immune Response" and T cell-mediated processes like TCR signaling from the Reactome and also "Downstream TCR Signaling". Despite the processes already mentioned in stage IIB, there is one additional gene signature significantly enriched there, namely the GO term "Metalloendopeptidase activity" (ES 0.78). In stage IIIB there are two processes that meet the cutoff values. The first signature was introduced by Slebos et al. (2006), "Head and Neck Cancer in HPV Up", and refers to 89 genes which are up-regulated in HPV positive head and neck cancers. The second significantly enriched gene signature is presented by "Drug Metabolism Cytochrome P450" from the KEGG pathway database (ES 0.70).

The gene set enrichment analysis of variance component derived genes, yields a characterization of transcriptional programs in each cervical cancer stage. While the approach produces overlap with gene signatures from different studies, it only highlights robust signatures, omitting high variable and therefore actually unreliable gene expression values.

Discussion

We present a functional investigation of gene expression profiles of cervical cancer stages and identification of potential clinically relevant biomarker genes. For this purpose, gene expression profiles of FIGO staged cohorts were recruited from four publicly available microarray studies. After quality checks and normalization, we obtained gene expression profiles of 126 patients for further processing by variance-component analysis and GSEA. This approach avoids the analysis of highly variable genes, which tend to result in speculative expression signatures. Instead, a conservative approach explores reliable biomarkers genes, which are constitutively expressed. Using variance-component analysis, we found eleven genes of potential biomarker use for cervical cancer. Seven probes are constitutively up-regulated in all cervical cancer stages. Among these genes are four kinases, namely PAK2, NEK2, AURKA and PRKDC. Furthermore, we found the induction of the ligase DTL and a gene which encodes the centrosomal protein 55 (CEP55). Additionally, we found up-regulated gene expression of GINS1 which is essential for DNA replication in yeast and xenopus eggs (Ueno et al. 2005). In summary, these genes are related to cell cycle and DNA-dependent processes. On the other hand, there are four genes, which are frequently down-regulated in cervical cancer patients. Among these genes, there is the already proposed tumor marker P11 (Laneve et al. 2008) and the tumor suppressor EMP1 (Zhang et al. 2011). Another down-regulated gene is UPK1A which encodes a cell surface protein and HSPC159 which encodes a galectin-related protein. In conclusion, the utility of the proposed biomarkers must be evaluated by future studies, however, since these genes are derived from actual patient data and are of low variability, they present excellent candidates.

Gene set enrichment analysis (GSEA) allows for connecting already published gene expression signatures to expression signatures of distant phenotypes or biopsies. Interestingly, treatment-related gene expression signatures may be revealed across otherwise unrelated indications, thus leading to the discovery of novel therapeutical options for already established agents. GSEA of the variance-component processed patient matrix comprised 9.873 probes, which had a W/T ratio of less than 0.75. It is striking that the results generated by our analysis did overlap with three cervical cancer studies from the recent

literature (Rosty et al. 2005; Pyeon et al. 2007; Slebos et al. 2006). Additionally, there is also overlap in the resulting gene expression signatures with two other virus-related gene signatures from HBV and HCMV studies (Browne et al. 2001; Wieland et al. 2004). These results acknowledge our conservative approach, since, as seen in the literature, other approaches tend to produce only marginal overlap in the resulting gene expression signatures (Wolfer and Ramaswamy 2011).

Our findings raise the possibility that darapladib may be a novel adjuvant treatment option for consideration in cervical cancer therapy, since we found repeatedly gene expression signatures which point to darapladib (Wilensky et al. 2008). Despite these results present significant findings, they are, however, limited to the investigation of the transcription status of cervical cancer patients: whether these findings are of clinical impact has to be tested in an independent clinical cohort. The lipoprotein-associated phospholipase A_2 (Lp-PLA$_2$) inhibitor darapladib represses 24 genes of which 20 are induced in our data. Wilensky et al. (2008) showed that darapladib reduces lesions in atherosclerosis and exerts also anti-inflammatory action. Incident HPV infections either manifest as cervical intraepithelial neoplasia (CIN) lesions or become undetectable within 36 months (Insinga et al. 2005). Therefore, it might be tempting to test whether the formation of lesions in cervical cancer can be reduced by darapladib and if this leads to clearing the HPV infection.

Another potential treatment option in therapy of cervical cancer might be presented by the novel angiogenic inhibitor angiocidin. Gaurnier et al. (2008) found that angiocidin, the protein product of PSMD4, induces in total 59 genes in THP-1 cells, which subsequently transform the cells into anti-tumoral, MMP9 secreting monocytes. In the current study, we found 23 of these genes significantly induced in stage IB1, IIA and IIB. A repeated activation of these genes might be in response to virus life cycle signals, activating the immunological defense mechanism for viral clearance. However, this immunologic defense seems to be activated only partially, since not all 59 genes are induced. As a matter of speculation, one might hypothesize that this is due to the influence of virus proteins, as seen in the literature, the virus protein E6 is able to alter host physiology to great extent, that is, by degrading p53 (Rosty et al. 2005). However, the Gaurnier signature is always accompanied by gene expression signatures related to dramatic immunological processes, that is, "Graft Versus Host Disease" and "Biopsy Kidney Transplant Rejected Versus OK Up". These immunological processes may provide a more concrete basis for considering a malfunction in monocyte activation. However, since our investigation was limited to the transcriptome, there is need for further evaluating if these effects are bound to alteration of immunologic

markers or whether there is an actual change in the number and diversity of the accumulating immune cells.

Apart from the repeating gene expression patterns in all other cervical cancer stages, we find that stage IB2 exhibits a rather unique expression profile. The most significantly enriched signature is presented by a gene set that was introduced first by Rozanov et al. (2008), that is, the "MMP14 Targets Subset". The authors found that expression of MT1-MMP is highly associated with migration, invasion and metastasis in general. Gene sets that are pointing to locomotion were found to be significantly enriched in stage IB2, for example regulation of cell migration from the Gene Ontology and NCAM1 interactions from the Reactome. Evidence for a transcriptional activation of an invasive process was found in the transcription of gene signatures related to ECM receptor interaction from KEGG, since Rozanov et al. proposed that MT1-MMP activation also induces activation of ECM maintenance. Tumor cells balance out metallopeptidase-digested cellular matrix by the activation of ECM maintenance (Rozanov et al. 2008). This is evidence for a transcriptional program, which transposes a tumor of stage IB2 into stage II, where it will invade beyond the uterus.

The analysis of cervical cancer patient gene expression data presents a novel perspective on HPV-mediated transcription processes. There is a growing magnitude of published gene expression signatures and biological pathways, which in future could provide a better basis for understanding of how HPV evades the immune response permanently. This knowledge may have implications for other tumorigenic viruses also. Thus, established gene expression signatures may harbor the next major treatment opportunity in cancer therapy.

References

Bachtiary B, Boutros PC, Pintilie M, Shi W, Bastianutto C, Li J-H, Schwock J, Zhang W, Penn LZ, Jurisica I, Fyles A, Liu F–F (2006) Gene expression profiling in cervical cancer: an exploration of intratumor heterogeneity. Clin Cancer Res 12:5632–5640. doi:10.1158/1078-0432.CCR-06-0357

Barrett T, Suzek TO, Troup DB et al (2005) NCBI GEO: mining millions of expression profiles–database and tools. Nucleic Acids Res 33:D562–D566. doi:10.1093/nar/gki022

Bidus MA, Risinger JI, Chandramouli GV, Dainty LA, Litzi TJ, Berchuck A, Barrett JC, Maxwell GL (2006) Prediction of lymph node metastasis in patients with endometrioid endometrial cancer using expression microarray. Clin Cancer Res 12(1):83–88. doi:10.1158/1078-0432.CCR-05-0835

Browne EP, Wing B, Coleman D, Shenk T (2001) Altered cellular mRNA levels in human cytomegalovirus-infected fibroblasts: viral block to the accumulation of antiviral mRNAs. J Virol 75(24):12319–12330. doi:10.1128/JVI.75.24.12319-12330.2001

Flechner SM, Kurian SM, Head SR, Sharp SM, Whisenant TC, Zhang J, Chismar JD, Horvath S, Mondala T, Gilmartin T, Cook DJ, Kay SA, Walker JR, Salomon DR (2004) Kidney transplant rejection and tissue injury by gene profiling of biopsies and peripheral blood lymphocytes. Am J Transplant 4:1475–1489. doi:10.1111/j.1600-6143.2004.00526.x

Gaurnier-Hausser A, Rothman VL, Dimitrov S, Tuszynski GP (2008) The novel angiogenic inhibitor, angiocidin, induces differentiation of monocytes to macrophages. Cancer Res 68(14):5905–5914. doi:10.1158/0008-5472.CAN-07-6179

Insinga A, Monestiroli S, Ronzoni S, Gelmetti V, Marchesi F, Viale A, Altucci L, Nervi C, Minucci S, Pelicci PG (2005) Inhibitors of histone deacetylases induce tumor-selective apoptosis through activation of the death receptor pathway. Nat Med 11:71–76. doi:10.1038/nm1160

Irizarry R, Hobbs B, Collin F et al (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264. doi:10.1093/biostatistics/4.2.249

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. CA Cancer J Clinic 61 (2):69–90. doi:10.3322/caac.20107

Laneve P, Gioia U, Ragno R, Altieri F, Di Franco C, Santini T, Arceci M, Bozzoni I, Caffarelli E (2008) The tumor marker human placental protein 11 is an endoribonuclease. J Biol Chem 283(50):34712–34719. doi:10.1074/jbc.M805759200

Lowy DR, Schiller JT (2006) Prophylactic human papillomavirus vaccines. J Clin Invest 116(5):1167–1173. doi:10.1172/JCI28607

Noordhuis MG, Fehrmann RS, Wisman GB, Nijhuis ER, van Zanden JJ, Moerland PD, Ver Loren van E, Volders HH, Kok M, ten Hoor KA, Hollema H, de Vries EG, de Bock GH, van der Zee AG, Schuuring E (2011) Involvement of the TGF-beta and beta-catenin pathways in pelvic lymph node metastasis in early-stage cervical cancer. Clin Cancer Res 17(6):1317–1330. doi:10.1158/1078-0432.CCR-10-2320

Pyeon D, Ma Newton, Lambert PF, den Boon JA, Sengupta S, Marsit CJ, Woodworth CD, Connor JP, Haugen TH, Smith EM, Kelsey KT, Turek LP, Ahlquist P (2007) Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. Cancer Res 67:4605–4619. doi:10.1158/0008-5472.CAN-06-3619

Rosty C, Sheffer M, Tsafrir D, Stransky N, Tsafrir I, Peter M, de Cremoux P, de La Rochefordiere A, Salmon R, Dorval T, Thiery JP, Couturier J, Radvanyi F, Domany E, Sastre-Garau X (2005) Identification of a proliferation gene cluster associated with HPV E6/E7 expression level and viral DNA load in invasive cervical carcinoma. Oncogene 24(47):7094–7104. doi:10.1038/sj.onc.1208854

Rozanov DV, Savinov AY, Williams R, Liu K, Golubkov VS, Krajewski S, Strongin AY (2008) Molecular signature of MT1-MMP: transactivation of the downstream universal gene network in cancer. Cancer Res 68(11):4086–4096. doi:10.1158/0008-5472.CAN-07-6458

Rudy J, Valafar F (2011) Empirical comparison of cross-platform normalization methods for gene expression data. BMC Bioinform 12(1):467. doi:10.1186/1471-2105-12-467

Scotto L, Narayan G, Nandula SV, Arias-Pulido H, Subramaniyam S, Schneider A, Kaufmann AM, Wright JD, Pothuri B, Mansukhani M, Murty VV (2008) Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. Genes Chromosom Cancer 47(9):755–765. doi:10.1002/gcc.20577

Slebos RJ, Yi Y, Ely K, Carter J, Evjen A, Zhang X, Shyr Y, Murphy BM, Cmelak AJ, Burkey BB, Netterville JL, Levy S, Yarbrough WG, Chung CH (2006) Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. Clin Cancer Res 12(3 Pt 1):701–709. doi:10.1158/1078-0432.CCR-05-2017

Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545–15550. doi:10.1073/pnas.0506580102

Ueno M, Itoh M, Kong L, Sugihara K, Asano M, Takakura N (2005) PSF1 is essential for early embryogenesis in mice. Mol Cell Biol 25(23):10528–10532. doi:10.1128/MCB.25.23.10528-10532.2005

Wieland S, Thimme R, Purcell RH, Chisari FV (2004) Genomic analysis of the host response to hepatitis B virus infection. Proc Natl Acad Sci USA 101(17):6669–6674. doi:10.1073/pnas.0401771101

Wilensky RL, Shi Y, Mohler ER 3rd, Hamamdzic D, Burgert ME, Li J, Postle A, Fenning RS, Bollinger JG, Hoffman BE, Pelchovitz DJ, Yang J, Mirabile RC, Webb CL, Zhang L, Zhang P, Gelb MH, Walker MC, Zalewski A, Macphee CH (2008) Inhibition of lipoprotein-associated phospholipase A2 reduces complex coronary atherosclerotic plaque development. Nat Med 14(10):1059–1066. doi:10.1038/nm.1870

Wilson CL, Miller CJ (2005) Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. Bioinformatics 21:3683–3685. doi:10.1093/bioinformatics/bti605

Wolfer A, Ramaswamy S (2011) MYC and metastasis. Cancer Res 71:2034–2037. doi:10.1158/0008-5472.CAN-10-3776

Zhang J, Cao W, Xu Q, Chen WT (2011) The expression of EMP1 is downregulated in oral squamous cell carcinoma and possibly associated with tumour metastasis. J Clin Pathol 64(1):25–29. doi:10.1136/jcp.2010.082404