Potential new biomarkers for squamous carcinoma of the uterine cervix

Peter A van Dam,1,2 Christian Rolfo,1,2,3 Rossana Ruiz,4 Patrick Pauwels,2,5 Christophe Van Berckelaer,2 Xuan Bich Trinh,1,2 Jose Ferri Gandia,3 Johannes P Bogers,6 Steven Van Laere6

ABSTRACT

**Aim** An in silico pathway analysis was performed in an attempt to identify new biomarkers for cervical carcinoma.

**Methods** Three publicly available Affymetrix gene expression data sets (GSE5787, GSE7803, GSE9750) were retrieved, vouching for a total 9 cervical cancer cell lines, 39 normal cervical samples, 7 CIN3 samples and 111 cervical cancer samples. An Agilent data set (GSE7410; 5 normal cervical samples, 35 samples from invasive cervical cancer) was selected as a validation set. Predication analysis of microarrays was performed in the Affymetrix sets to identify cervical cancer biomarkers. We compared the lists of differentially expressed genes between normal and CIN3 samples on the one hand (n=1923) and between CIN3 and invasive cancer samples on the other hand (n=628).

**Results** Seven probe sets were identified that were significantly overexpressed (at least 2 fold increase expression level, and false discovery rate <5%) in both CIN3 samples respective to normal samples and in cancer samples respective to CIN3 samples. From these, five probes sets could be validated in the Agilent data set (P<0.001) comparing the normal with the invasive cancer samples, corresponding to the genes DTL, HMGB3, KIF2C, NEK2 and RFC4. These genes were additionally overexpressed in cervical cancer cell lines respective to the cancer samples. The literature on these markers was reviewed

**Conclusion** Novel biomarkers in combination with primary human papilloma virus (HPV) testing may allow complete cervical screening by objective, non-morphological molecular methods, which may be particularly important in developing countries.

INTRODUCTION

Persistent infections with oncogenic human papilloma virus (HPV) play a central role in the carcinogenesis of carcinoma of the uterine cervix.[1,4] However, only a small fraction of high-risk HPV-positive women will eventually develop a clinically relevant lesion as the majority of HPV infections induce low-grade precursor lesions that are cleared spontaneously.[4] The molecular biological events involved in progression from low-grade lesions to high-grade lesions and invasive cancer are not well understood. Induction of genomic instability and global disruption of gene expression, particularly in the HPV E6 and E7 oncoproteins, has been implicated in HPV-associated carcinogenesis. Novel biomarkers that allow monitoring of these essential molecular events in histological and cytological specimens are likely to improve the detection of lesions that have a high risk of progression, in both primary screening and triage settings.[1,4]

Although polyvalent HPV vaccines seem promising to eradicate cervical cancer in the future, this disease will be with us for the next decades as these vaccines are at present not generally available in the Western and developing world, while not being entirely effective.[5] This implicates that screening will remain important as a strategy to reduce mortality caused by carcinoma of the cervix uteri. Rapid molecular methods for detecting HPV DNA have become...
commercially available over the last decade and have been introduced for HPV-based cervical cancer screening in some countries. Despite several advantages, HPV detection has a low positive predictive value for cervical cancer implicating that HPV-positive women need to be triaged with additional testing to determine optimal risk stratification and management. Disease-specific biomarkers such as p16(INK4A), HPV(E6/E7) mRNA, topoisomerase 2α, Ki67 or novel methylation assays have been evaluated in various settings but are at this point not sufficiently validated to be incorporated in HPV-based screening programmes. In the PALMS Study it was recently demonstrated that p16/Ki67 dual-stained cytology combines superior sensitivity and non-inferior specificity over cytology for detecting CIN2+ in a group of 27349 women attending routine cervical cancer screening. Reduced expression of the tumour suppressor CDKN2A leads to downstream overexpression of p16 in premalignant cervical cells. Some small panels such as methylation status of cell adhesion molecule 1 (CADM1) and T-lymphocyte maturation associated protein (MAL) showed to have promise as triage tests for HPV-positive women. However, a recent study assessing DNA methylation status of MAL, ADCYAP1, PAX1 and CADM in 205 patients with low-grade or high-grade CIN and cervical cancer demonstrated that ADCYAP1 and PAX1 had a relatively better discriminatory ability than did methylated MAL and CADM1, which illustrates that the best panel still needs to be discovered.

Recently we conducted an in silico analysis looking for driver pathways on all publicly available Affymetrix data sets containing normal and pretreatment (pre)invasive cancer samples with relevant clinical information. In the present paper we addressed and validated the biomarkers which came up in this analysis and we reviewed the literature on the subject.

MATERIALS AND METHODS

Patient data sets

Patient data sets used were described previously. Briefly all publicly available Affymetrix data sets (HGU133-series: GSE5787, GSE7803, GSE9750) containing normal and pretreatment (pre)invasive cervical cancer samples with relevant clinical information were retrieved from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/gds). The only available Agilent data set (GSE741) fulfilling the same criteria was selected as a validation set. The three Affymetrix gene expression data sets vouched for a total of nine cervical cancer cell lines, 39 normal cervical samples, 7 CIN3 samples and 11 cervical cancer samples. The Agilent data set contained data of 5 normal cervical samples and 35 samples from invasive cervical cancer. No information was available on the HPV status of the samples.

Data normalisation and exploration

Data normalisation and exploration was described in detail in previous papers.

Biomarker analysis for early diagnosis

To detect potential biomarkers for early diagnosis, we performed predication analysis of microarrays comparing normal cervical samples to CIN3 samples and CIN3 samples to invasive cervical cancer samples as described previously. We also compared the lists of differentially expressed probe sets between normal versus CIN3 samples, CIN3 versus cancer samples and normal versus cancer samples (figure 1).

Immunohistochemical staining

Immunohistochemical validation for high mobility group box 3 (HMGB3) was performed on formalin-fixed paraffin-embedded sections of normal, CIN III and invasive cervical cancer samples using a polyclonal antibody against the C-terminal region of HMGB3 (Aviva Systems Biology, San Diego, USA), diluted 1/100 on a Dako autostainer.

RESULTS

Biomarker discovery for early detection

In order to identify biomarkers for potential early detection of cervical cancer, a biomarker discovery analysis of the Affymetrix microarrays was performed comparing: (A) expression profiles of the normal cervical samples versus the CIN3 samples; and (B) the normals versus the invasive cancer samples. Prediction analysis of genes comparing the top 100 most differentially expressed genes in the normal and invasive samples showed that CDKN2A, MAL, ECT2 and PPP1R3C came up as potential biomarkers (figure 2). Of these genes, CDKN2A, ECT2 and PPP1R3C could be validated in the Agilent GSE7410 data set (figure 3). Expression of CDKN2A and ECT was lower in the CIN III and invasive samples compared with the normals making these genes less suitable for early detection. Expression of PPP1R3C is higher in CIN III and invasive cancer versus the normals, but levels of expression are similar in the premalignant and malignant samples in the Affymetrix data sets, which is also inconvenient in a triage setting. Therefore, we decided to adopt a second approach which was described in detail in a previous paper. Briefly we compared the lists of differentially expressed genes between normal and CIN3 samples on the one hand (n=1923) and between CIN3 and invasive cancer samples on the other hand (n=628) and looked for genes with higher expression in invasive cancer compared with CIN III and higher expression in CIN III compared with the normals. Seven probe sets were identified that were significantly overexpressed (at least 2 fold increased expression level, and false discovery rate <5%) in both CIN3 samples respective to normal samples and in cancer samples respective to CIN3 samples. From these, six probe sets corresponded to six unique genes: Aurora kinase A (AURKA), denticleless E3 ubiquitine protein ligase homologue DTL, HMGB3, kinesin family member 2 C (KIF2C), NIMA(never in
mitosis gene a)-related kinase 2 (NEK2) and RFC4. Five probe sets could be validated in the Agilent data set (P<0.001) comparing the normal with the invasive cancer samples, corresponding to the genes DTL, HMGB3, KIF2C, NEK2 and RFC4 (figure 4). AURKA reached borderline significance (P 0.073) in a similar

Figure 1  Heat map depicting genes enriched in tumour samples compared with normal samples according to GSA analysis.

Figure 2  (A) Biomarker discovery analysis showing prediction analysis of Affymetrix microarrays comparing the top 100 differentially expressed genes. (B) Retained genes were compared in the normal, CIN III, invasive cancer samples and the cell lines.
analysis. There was no difference in the expression of these genes in samples of patients with lymph node metastases or without lymph node metastases in the Agilent samples. The above genes were additionally overexpressed in cervical cancer cell lines respective to the cancer samples, suggesting they are cancer cell intrinsic and thus can be considered as potential biomarkers for cervical cancer tailored to early diagnosis. Genes that showed significantly lower expression in (pre)cancer samples compared with normal samples were not retained as biomarkers, as the absence of a marker for (pre)malignancy may be caused by technical errors and therefore is more difficult to use in daily practice than the presence of the marker.

Validation of the data
Data validation was discussed in detail in a previous paper. In addition we performed immunochemical staining for HMGB3 in normal cervix, CIN III and invasive carcinoma and could show absent staining in normal
cervix, absent to weak staining in CIN III and clear strong nuclear staining in invasive carcinomas (figure 5).

**DISCUSSION**

In the present in silico study we could identify six biomarkers which have the potential to be used for early detection of patients at risk to develop progressive cervical lesions: AURKA, DTL, HMGB3, KIF2C, NEK2 and RFC4. All these genes play pivotal roles in the control of proliferation and differentiation. Using a different methodology Koch and Wiese performed a similar microarray analysis of 24 normal and 102 cervical cancer biopsies from four pooled publicly available studies. They found seven probes which are induced in all cervical cancer stages and hereby confirmed the relevance of AURKA, DTL and NEK2 in addition to GINS1, PAK2, PRKDC and CEP 55.

AURKA is a member of the evolutionary conserved Aurora serine/threonine kinase family which is important to maintain genomic stability. Overexpression of AURKA promotes cell proliferation through G1/S cell cycle transition and to antiapoptosis, and can cause polyploidy and chromosomal instability, xenograft tumour growth and chemoresistance. Simultaneous inhibition of AURKA and AURKB results in a dramatic decrease in spindle microtubule stability. The human AURKA gene maps to 20q13, a region frequently amplified in breast cancers and is also overexpressed in several tumours. Although it has been suggested that AURKA expression may predict the outcome of patients with cervical cancer, its precise function and molecular mechanism in cervical cancer pathogenesis remains unclear. Using immunohistochemical staining in 180 cervical cancer tissues, Sun et al showed that AURKA is overexpressed in cervical cancer and that expression is significantly correlated with tumour size (P=0.023), lymphovascular space involvement (P<0.001) and deep invasion (P=0.014). Twu et al performed immunohistochemical staining of AURKA and AURKB in 20 samples of normal cervix, 35 CIN III samples and 95 invasive cervical carcinoma samples (76 squamous and 19 adenocarcinomas) and could show that expression of these genes is significantly increased in invasive carcinoma and CIN III. Overexpression of AURKA was higher in squamous carcinoma compared with adenocarcinoma (50% vs 21%, P=0.023). There was correlation between AURKA and AURKB expression and survival. A screen of the human kinome has identified AURKA as being synthetically lethal on the background of HPV infection.

DTL is an early checkpoint regulating gene interacting with p21. It is also known as CDT2 (CDC10-dependent transcript 2), DCAF2, L2DTL or RAMP. Checkpoint genes maintain genomic stability by arresting cells after DNA damage. Many of these genes also control cell cycle events in unperturbed cells. DTL/CDT2 is required for normal cell cycle control, primarily to prevent replication. DTL promotes genomic stability as an essential component of the CUL4-DDB1 complex that controls CDT1 levels. It has been shown that changes in the expression of TP53, which play a major role in the pathogenesis of cervical cancer, affects its downstream miRNAs and their most important gene targets MEIS1, AGTR1, DTL, TYMS and BAK1 in head and neck squamous cell carcinoma. The DTL gene was a found to be of functional relevance in the tumorigenesis of hepatocellular, gastric, colon and breast carcinoma, and rhabdomyosarcoma, and may be of prognostic importance. According to the Human Protein Atlas project (www.proteinatlas.org) low DTL expression was found in cervical cancer but was also seen in normal cervical tissue.

NEK2 is a serine/threonine kinase involved in the regulation of centrosome duplication and spindle assembly during mitosis. Dysregulation of these processes causes chromosome instability and aneuploidy. There are three isoforms that result from alternate splicing of this gene, named NEK2A, NEK2B and NEK3C. NEK2A is 31% structurally identical to AURKA. Subcellular localisation analysis shows that NEK2A resides in both cell nucleus and cytoplasm. It displays a cell cycle dependent expression pattern, being low in G1, increasing through S and G2 to reach peak in late G2/M and decreasing on entry into mitosis. Overexpression of NEK2 has been reported in cervical and other cancer cell lines and several neoplastic diseases such as preinvasive and invasive breast carcinomas, lung adenocarcinomas, testicular seminomas, liver cancer, pancreatic carcinomas, prostate carcinomas, and diffuse large B cell lymphomas.
NEK2 is a bad prognostic factor in patients with breast cancer, non-small cell lung cancer, colonic carcinoma and pancreatic carcinoma. Because NEK2A has such a broad spectrum of roles in different cell processes, it is an attractive target for treatment.

HMGB3, also known as HMG4, HMG2A, is a recently discovered member of the high mobility group (HMG) superfamily of HMG proteins, and is classified with HMGB1 and HMGB2 into the HMG-box subfamily. The 80% identity between HMG-box proteins suggests similar functions at a molecular level. HMGB1 and HMGB2 can interact with DNA and subsequently bend linear DNA, thereby facilitating nucleoprotein complex formation through alteration of local chromatin architecture. HMGB3 has been reported to be overexpressed in a variety of human cancers such as gastric cancer, oesophageal squamous cell carcinoma, bladder cancer, non-small cell lung cancer and breast cancer. Overexpression of HMGB3 is correlated with aggressive behaviour and poor prognosis in almost all of these tumour types. HMGB3 has been identified in Hela cervical cancer cells. According to the Human Protein Cancer Atlas cervical cancers have variable staining for HMGB3 whereas staining is absent in normal cervix. No further data are available on the expression of HMGB3 in cervical cancer.

KIF2C encodes a kinesin-like protein that functions as a microtubule-dependent molecular motor. The encoded protein can depolymerise microtubules, thereby promoting the anaphase of mitotic chromosome segregation and may be required to coordinate the onset of sister centromere separation. KIF2C therefore plays an important role during cell proliferation and may be an essential gene in carcinogenesis. KIF2C can induce spontaneous CD4(+) T cell responses of the Th1-type which are tightly controlled by peripheral T regulatory cells, thereby enhancing CD4(+) T cell responses of the Th1-type which are necessary to suppress the development of autoimmunity. According to the Human Protein Cancer Atlas cervical cancers have variable staining for KIF2C whereas staining is absent in normal cervix.

The human replication factor C (RFC) is a multimeric protein consisting of five distinct subunits that are highly conserved through evolution. The RFC family functions as clamp loaders that load PCNA onto DNA in an ATP-dependent process during DNA synthesis. RFC is involved in DNA repair following DNA damage. The RFC4 gene that encodes for the fourth largest subunit of the RFC complex has been reported to be deregulated in diverse malignancies including prostate cancer, head and neck squamous cell carcinomas, hepatocellular carcinoma, colonic carcinomas and cervical cancer. Using cDNA array comparative genomic hybridisation Narayan et al indentified a number of over-represented and deleted genes in 29 cases of cervical cancer. This analysis exhibited frequent and robust upregulated expression for RFC4 and KIF4A among others such as EPHB2, CDC8A, MUC4, MMP1, MMP13 and AKT1 in cervical cancer compared with normal cervix. Comparing three primary isolates and two established cervical cancer cell lines to normal keratinocytes Kang et al could detect overexpression of RPA, RFC, PCNA and DNA polymerase, which seem to play a role in adeno-associated virus DNA replication. Increased expression of sine oculis homeobox homologue 1 (SIX1, a master regulator of DNA replication in cervical cancer cells) can be induced by the E7 oncoprotein of HPVs in cervical intraepithelial neoplasia and cervical cancer and can result in higher levels of expression of genes related to the initiation of DNA replication such as RFC. Recently Niu et al showed that dysregulation of CDKN2A, IL1R2 and RFC4 may contribute to cervical cancer progression and may be potential diagnostic markers. Huang et al could identify and validate a seven-gene signature (consisting of UBL3, FGF3, BM11, PDGFRA, PTPRF, NOL7 and RFC4) in, respectively, a training set (n=50) and a testing set (n=50) of invasive cervical cancer samples from 100 patients using a custom oligonucleotide microarray. Multivariate analysis showed that International Federation of Gynecology and Obstetrics (FIGO stage) and the seven-gene signature are independent prognostic factors associated with relapse-free survival of patients with cervical cancer.

Novel biomarkers in combination with primary HPV testing are needed to enable complete cervical screening by objective, non-morphological molecular methods, which may be particularly important in developing countries. The potential biomarkers in this paper should be further validated in cytology and histological samples of patients with normal, cervical, intraepithelial neoplasias and primary and recurrent cervical cancer. Particularly assessing the diagnostic performance of cytology and/or HPV positivity versus a quantitative polymerase chain reaction (PCR) measuring the expression of a panel of the above genes warrants furthers study. In an interesting study proving this concept Nischalke et al could demonstrate that measuring IGF2BP3, HOXB7 and NEK2 mRNA levels by PCR in addition to cytology has the potential to improve diagnostic precision to detect malignant biliary disorders from brush cytology specimens. Sensitivity and specificity were highest when the three diagnostic markers were combined with routine cytology. As the expression levels are not mutually exclusive between the normal and cancer samples, an algorithm using multiple biomarkers will have to be developed in the future to use this test for molecular screening of cervical cancer. Further studies are also needed to clarify whether some of the retained genes may also be used as therapeutic targets.
41. Nischalke HD, Schmitz V, Luda C, et al. Detection of IGF2BP3, HOXB7, and NEK2 mRNA expression in brush cytology specimens as a new diagnostic tool in patients with biliary strictures. *PLoS One* 2012;7:e42141.

42. Xia J, Franqu Machin R, Gu Z, et al. Role of NEK2A in human cancer and its therapeutic potentials. *Biomed Res Int* 2015;2015:862461.

43. Nemeth MJ, Curtis DJ, Kirby MR, et al. Hmgb3: an HMG-box family member expressed in primitive hematopoietic cells that inhibits myeloid and B-cell differentiation. *Blood* 2003;102:1298–306.

44. Nemeth MJ, Cline AP, Anderson SM, et al. Hmgb3 deficiency derepresses proliferation and differentiation of common lymphoid and myeloid progenitors. *Blood* 2005;105:627–34.

45. Nemeth MJ, Kirby MR, Bodine DM. Hmgb3 regulates the balance between hematopoietic stem cell self-renewal and differentiation. *Proc Natl Acad Sci U S A* 2006;103:13783–8.

46. Tang D, Kang R, Zeh HJ, et al. High-mobility group box 1 and cancer. *Biochim Biophys Acta* 2010;1799:131–40.

47. Goug Y, Cao Y, Song L, et al. HMGB3 characterization in gastric cancer. *Genet Mol Res* 2013;12:6032–9.

48. Elgami OA, Park JK, Gusev Y, et al. Tumor suppressor function of mir-205 in breast cancer is linked to HMGB3 regulation. *PLoS One* 2013;8:e76402.

49. Gao J, Zou Z, Gao J, et al. Increased expression of HMGB3: a novel independent prognostic marker of worse outcome in patients with esophageal squamous cell carcinoma. *Int J Clin Exp Pathol* 2015;8:345–52.

50. Li M, Cai Y, Zhao H, et al. Overexpression of HMGB3 protein promotes cell proliferation, migration and is associated with poor prognosis in urinary bladder cancer patients. *Tumour Biol* 2015;36:4785–92.

51. Song N, Liu B, Wu JL, et al. Prognostic value of HMGB3 expression in patients with non-small cell lung cancer. *Tumour Biol* 2013;34:2599–603.

52. Guggenheimer ER, Xu D, Zhang CX, et al. Photoaffinity isolation and identification of proteins in cancer cell extracts that bind to platinum-modified DNA. *Chembiochem* 2009;10:141–57.

53. Lu Y, Xiao X, Crystal JS, et al. Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions. *Cancer Res* 2014;74:3401–10.

54. Gnjatic S, Cao Y, Crystal JS, et al. NY-ESO-1, a novel T cell antigen recognized by T cells associated with durable tumor regressions. *Cancer Res* 2003;63:5198–206.

55. Gnjatic S, Cao Y, Reichelt U, et al. NY-ESO-1, a novel T cell antigen recognized by T cells associated with durable tumor regressions. *Cancer Res* 2003;63:5198–206.

56. Abdel-Fatah TMA, Green AR, Lementre C, et al. P4-09-11: Kinesin Family member 2C (KIF2C) is a new surrogate prognostic marker in Breast Cancer (BC). *Cancer Res* 2011;71:P4-09-11.

57. Wang J, Ma S, Ma R, et al. KIF2A silencing inhibits the proliferation and migration of breast cancer cells and correlates with unfavorable prognosis in breast cancer. *BMC Cancer* 2014;14:461.

58. Bidkhor G, Narimani Z, Hosseini Ashtiani S, et al. Reconstruction of an integrated genome-scale co-expression network reveals key modules involved in lung adenocarcinoma. *PLoS One* 2013;8:e67552.

59. Zhao F, Shu MK, Jiang L, et al. Overexpression of forkhead box protein M1 (FOXM1) in ovarian cancer correlates with poor patient survival and contributes to paclitaxel resistance. *PLoS One* 2014;9:e113478.

60. Tsunimoto T, Stillman B. Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. *Proc Natl Acad Sci U S A* 1990;87:1023–7.

61. Kim H-S, Brill SJ. Rfc4 interacts with Rpa1 and is required for both dna replication and DNA damage checkpoints in Saccharomyces cerevisiae. *Mol Cell Biol* 2001;21:3725–37.

62. Erdogan E, Klee EW, Thompson EA, et al. Meta-analysis of oncogenic protein kinase C iota signaling in lung adenocarcinoma. *Clin Cancer Res* 2009;15:1527–33.

63. Kang B, You H, Bandyopadhyay S, et al. Cervical cancer isolate PT3, super-permissive for adenov-associated virus replication, over-expresses DNA polymerase δ, PCNA, RFC and RPA. *BMC Microbiol* 2009;9:79–93.

64. Natayan G, Bourdon V, Chaganti S, et al. Gene dosage alterations revealed by cDNA microarray analysis in cervical cancer: Identification of candidate amplified and overexpressed genes. *Genes, Chromosomes and Cancer* 2007;46:373–84.

65. Slezos RJ, Yi Y, Ely K, et al. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *Clin Cancer Res* 2006;12:701–9.

66. Smith IM, Mithani SK, Liu C, et al. Novel integrative methods for gene discovery associated with head and neck squamous cell carcinoma development. *Arch Otolaryngol Head Neck Surg* 2009;135:487–95.

67. Xiang J, Fang L, Luo Y, et al. Levels of human replication factor C4, a clamp loader, correlate with tumor progression and predict the prognosis for colorectal cancer. *J Transl Med* 2014;12:320.

68. Lee C-F, Ling Z-Q, Zhao T, et al. Genomic-wide analysis of lymphatic metastasis-associated genes in human hepatocellular carcinoma. *World J Gastroenterol* 2009;15:356–65.

69. Liu D, Zhang XX, Xi BX, et al. Sine oculis homeobox homolog 1 promotes DNA replication and cell proliferation in cervical cancer. *Int J Oncol* 2014;45:1323–40.

70. Niu G, Wang D, Pei Y, et al. Systematic identification of key genes and pathways in the development of invasive cervical cancer. *Gene* 2017;618:28–41.

71. Huang L, Zheng M, Zhou QM, et al. Identification of a 7-gene signature that predicts relapse and survival for early stage patients with cervical carcinoma. *Med Oncol* 2012;29:2911–8.

72. Tilborgs H, Corthouts J, Verhoeven Y, et al. The role of nuclear factor-kappa B signaling in human cervical cancer. *Crit Rev Oncol Hematol* 2017;120:141–50.