Identification of Copy Number Gain and Overexpressed Genes on Chromosome Arm 20q by an Integrative Genomic Approach in Cervical Cancer: Potential Role in Progression

Luigi Scotto,[1][†] Gopeshwar Narayan,[1][†][‡] Subhadra V. Nandula,[1][†] Hugo Arias-Pulido,[2,3] Shivakumar Subramaniyam,[1] Achim Schneider,[4] Andreas M. Kaufmann,[4] Jason D. Wright,[5] Bhavana Pothuri,[5] Mahesh Mansukhani,[1] and Vundavalli V. Murty[1,6]*

[1]Department of Pathology, Columbia University Medical Center, New York, NY
[2]Department of Tumor Molecular Biology, Instituto Nacional de Cancerología, Bogotá, Colombia
[3]Division of Hematology/Oncology, The University of New Mexico Cancer Center, Albuquerque, NM
[4]Department of Gynecology, Charité Universitätsmedizin Berlin, Hindenburgdamm 30, Berlin, Germany
[5]Department of Gynecologic Oncology, Columbia University Medical Center, New York, NY
[6]Institute for Cancer Genetics, Columbia University Medical Center, New York, NY

Recurrent karyotypic abnormalities are a characteristic feature of cervical cancer (CC) cells, which may result in deregulated expression of important genes that contribute to tumor initiation and progression. To examine the role of gain of the long arm of chromosome 20 (20q), one of the common chromosomal gains in CC, we evaluated CC at various stages of progression using single nucleotide polymorphism (SNP) array, gene expression profiling, and fluorescence in situ hybridization (FISH) analyses. This analysis revealed copy number increase (CNI) of 20q in >50% of invasive CC and identified two focal amplicons at 20q11.2 and 20q13.13 in a subset of tumors. We further demonstrate that the acquisition of 20q gain occurs at an early stage in CC development and the high-grade squamous intraepithelial lesions (HSIL) that exhibit 20q CNI are associated ($P = 0.05$) with persistence or progression to invasive cancer. We identified a total of 26 overexpressed genes as consequence of 20q gain ($N = 14$), as targets of amplicon 1 ($N = 9$; two genes also commonly expressed with 20q gain) and amplicon 2 ($N = 6$; one gene also commonly expressed with 20q gain). These include a number of functionally important genes in cell cycle regulation (E2F1, TPX2, KIF3B, PIGT, and B4GALT5), nuclear function (CSEL1), viral replication (PSMA7 and LAMA5), methylation and chromatin remodeling (ASXL1, AHCY, and C20orf20), and transcription regulation (TCEA2). Our findings implicate a role for these genes in CC tumorigenesis, represent an important step toward the development of clinically significant biomarkers, and form a framework for testing as molecular therapeutic targets.   © 2008 Wiley-Liss, Inc.

INTRODUCTION

About 500,000 new cases of cervical cancer (CC) are diagnosed worldwide every year and the majority of affected women with advanced stages of cancer die (Waggoner, 2003). This failure of response to treatment of advanced CC is due to the lack of understanding of its biology at molecular level and targeted treatment regimens. Despite the documented etiologic role of HPV infection, the molecular basis of the genetic changes in progression in the multistep process of cervical tumorigenesis is poorly understood (Gius et al., 2007). CC cells exhibit highly complex karyotypic alterations (Harris et al., 2003). Molecular characterization of these complex chromosomal alterations is therefore important in understanding the genetic basis of CC,

which may ultimately facilitate in identification of critical genes in CC development.

Chromosomal gain or amplification is a common cellular mechanism of gene activation in tumori-

genesis. Specific genes amplify in distinct tumor types and their identification could be of prognostic significance as the copy number increase (CNI) of a variety of genes correlate with tumor progression and treatment resistance (Schwab, 1999). A large number of chromosomal regions are gained or amplified in CC suggesting that gene amplification and CNI of specific genes is a common genetic alteration in this tumor (Mitra et al., 1994; Narayan et al., 2003b, 2007; Rao et al., 2004). Using comparative genomic hybridization (CGH) approaches, we and others previously showed that chromosome 20 was one of the recurrently gained chromosomes and a high level CNI at 20q11.2 and 20q13.1 regions has also been reported in invasive CC (Harris et al., 2003; Narayan et al., 2003b, 2007; Rao et al., 2004; Wilting et al., 2006). These data suggest that the 20q genetic alterations contribute to the CC pathogenesis. The identification of target genes of 20q gain is likely to have a significant impact on the understanding of the pathogenesis and clinical management of CC.

To unravel the role of chromosome 20 gain in CC tumorigenesis, we characterized the chromosome 20 alterations using high throughput genomic and expression approaches.

MATERIALS AND METHODS

Tumor Specimens and Cervical Cancer Cell Lines

A total of 253 specimens were used in the present study in various investigations. These include 9 cell lines, 153 untreated primary tumors, 71 pap smears, and 20 normal cervical tissues. The cell lines (HT-3, ME-180, CaSki, MS751, C-4I, C-33A, SW756, HeLa, and SiHa) were obtained from American Type Culture Collection (ATCC, Manassas, VA) and grown in tissue culture as per the supplier's specifications. Twenty age-matched normal cervical tissues from hysterectomy specimens obtained from Columbia University Medical Center (CUMC), New York, were used as controls after enrichment for epithelial cells by microdissection. Cytologic specimens were collected using the ThinPrep Test Kit (Cytc Corporation, Marlborough, MA). After visualization of the cervical os the ectocervix was sampled with a spatula and endocervical cells obtained with a brush rotated 360°. Exfoliated cells were immediately placed in PreservCyt Solution (Cytc Corporation, Marlborough, MA) for routine processing by cytopathologist. Pap smears were collected from normal and precancerous lesions by simultaneous preparation of slides from the same spatula for both cytology

and FISH. FISH slides were immediately fixed in 3:1 methanol and acetic acid, and stored at 4°C until hybridization. A total of 71 pap smears with the diagnosis rendered by a cytopathologist as normal/squamous metaplasia/ASCUS ($N = 32$), LSIL ($N = 14$), or HSIL ($N = 25$) obtained from CUMC were used for FISH analysis. The diagnosis of all HSILs was also confirmed by a biopsy. Of the 153 primary tumors, 117 were obtained as frozen tissues and 36 specimens as formalin-fixed paraffin-embedded tissues. All primary invasive cancer specimens were obtained from patients evaluated at CUMC, Instituto Nacional de Cancerologia (Santa Fe de Bogota, Colombia) (Pulido et al., 2000), and the Department of Gynecology of Campus Benjamin Franklin, Charité-Universitätsmedizin Berlin (Germany) with appropriate informed consent and approval of protocols by institutional review boards. All primary tumors were diagnosed as squamous cell carcinoma (SCC) except five that were diagnosed as adenocarcinoma (AC). Clinical information such as age, stage and size of the tumor, follow-up data after initial diagnosis and treatment was collected from the review of institutional medical records. Tissues were frozen at −80°C immediately after resection and were embedded with tissue freeze medium (OTC) before microdissection. All primary tumor specimens were determined to contain at least 60% tumor by examining hematoxylin and eosin (H&E) stained adjacent sections. High molecular weight DNA and total RNA from tumor, normal tissues, and cell lines were isolated by standard methods. The integrity of all RNA preparations was tested by running formaldehyde gels and any samples that showed evidence of degradation were excluded from the study.

Microarray Analysis

The Affymetrix 250K *Nsp*I single nucleotide polymorphism (SNP) chip was used for copy number analysis as per the manufacturer's protocol. Briefly, 250 ng of genomic DNA was digested with *Nsp*I, generic linkers were added followed by PCR amplification, end-labeling, and fragmentation following standard protocols. Hybridization, washing, acquisition of raw data using GeneChip Operating Software (GCOS), and generation of .CEL files was performed by the Affymetrix Core facility at our institute. We used 79 CC cases (9 cell lines and 70 primary tumors enriched for tumor cells by microdissection) and 7 microdissected normal cervical squamous epithelial samples as controls to serve as the reference for copy number analysis. SNP data of test samples and normal cervical epi-

thelial specimens were loaded to dChip to calculate signal intensity values using the perfect match/mismatch (PM/MM) difference model followed by normalization of signals within chip and between chips using model-based expression (Li and Wong, 2001; Lin et al., 2004). DNA copy number gains were obtained as determined by dChip using analysis of signal intensity values based on the Hidden Markov Model. Arrays with >93% call rates were included in the analysis as per Affymetrix manual. Copy number data were obtained for chromosome 20 using CytoBand information files from the dChip website (http://biosun1.harvard. edu/complab/dchip/chromosome.htm#refgene). Chromosome 20 represented by a total of 5766 SNPs in the 250K NspI array with coverage of 3.6% of the genome. Both the raw copy number and $\log_2$ ratio (Signal/mean signal of normal samples at each SNP) were computed to estimate copy number changes in chromosome view. Copy numbers <1.5 were considered as deletion, 2.5–4.0 as gain, and ≥ 4.1 as amplification in the raw copy number view. All of the original data files were submitted to Gene Expression Omnibus (GEO Accession number: GSE10092).

We used Affymetrix U133A oligonucleotide microarray (Santa Clara, CA) containing 14,500 probe sets for gene expression analysis. RNA isolated from 29 CC cases (20 primary tumors enriched for tumor cells by microdissection and 9 cell lines) and 20 microdisssected normal cervical squamous epithelial cells were used for expression studies. Biotinylated cRNA preparation and hybridization of arrays was performed by the standard protocols supplied by the manufacturer. Arrays were subsequently developed and scanned to obtain quantitative gene expression levels. Expression values for the genes were determined using the Affymetrix GeneChip Operating Software (GCOS) and the Global Scaling option, which allows a number of experiments to be normalized to one target intensity to account for the differences in global chip intensity. The .CEL files obtained from the GCOS software were processed and normalized by dChip algorithm as described earlier. An average percent present call of 54% was obtained among all samples, which is expected for high quality RNA as per the manufacturer. Arrays were normalized at PM/MM probe level and a median intensity array from normal as the baseline array using invariant set normalization (Li and Wong, 2001; Lin et al., 2004). Followed by normalization, model-based expression values were calculated using PM/MM data view to fit the model for all probe sets. All original data files were deposited

to GEO (Accession number: GSE9750). To obtain a list of differentially expressed gene signatures, we compared all normal with all tumor samples using the criteria of 1.75-fold change between the group means at 90% confidence interval and a significance level of $P < 0.05$. All negative expression values for each probe set were truncated to 1 before calculating fold changes and <10% of samples with present call in each group were excluded. A total of 671 probe sets on chromosome 20 are present in U133A array representing 4.6% of the genome (3% on 20q and 1.6% on 20p). A list of differentially expressed genes identified on chromosome 20 was used in all subsequent supervised analyses using the same criteria between various groups to obtain relevant gene signatures.

Fluorescence In Situ Hybridization (FISH) and HPV Typing

FISH was performed by standard methods on frozen tissue sections fixed in 3:1 methanol: acetic acid, tissue microarrays prepared from paraffin embedded tissues, and on pap smears fixed in 3:1 methanol: acetic acid. DNA prepared from human BAC clone RP11-30F23 (20q13.1) (Open Biosystems, Huntsville, AL) was labeled by nick-translation using Spectrum Green dUTP fluorochrome (Vysis, Downers Grove, IL). Spectrum Orange-labeled chromosome 20 centromere, a Spectrum Orange-labeled D20S108 that maps to 20q12, and a Spectrum Green-labeled chromosome 11 centromere probe used as control were obtained from Vysis (Downers Grove, IL). Hybridization signals on 100–500 interphase cells on DAPI counterstained slides were scored on Nikon Eclipse epifluorescence microscope equipped with Applied Imaging CytoVision software (San Jose, CA). Scoring of FISH signals on frozen and paraffin-embedded tissue sections was restricted to tumor cells based on the identification of areas of tumor on adjacent H&E sections by the pathologist (MM). FISH signal scoring on Pap smear slides was restricted to large and atypical epithelial cells. Presence of signals suggestive of gain or amplification in at least 3% cells was considered positive and the results correlated with parallel cytomorphologic findings. Human papillomavirus types were identified as described earlier (Narayan et al., 2003a).

RESULTS

20q Gain Is a Frequent Genomic Alteration in CC

We performed Affymetrix 250K *Nsp*I SNP array analysis on a panel of 79 CC cases (70 primary tumors and 9 cell lines) to identify genome-wide
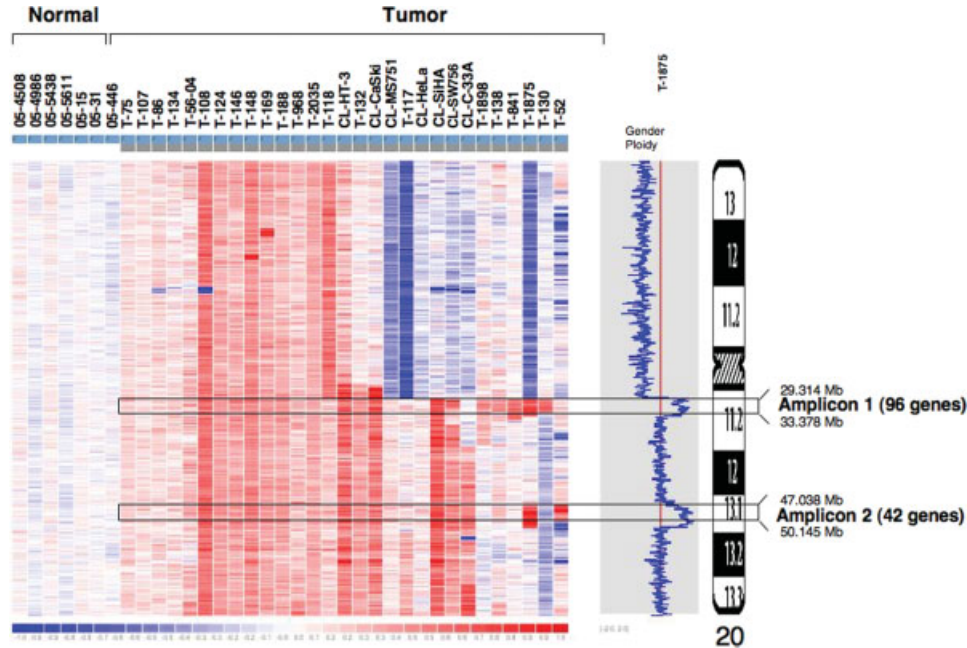
Figure 1. Identification of chromosome 20q copy number alterations and focal amplicons in invasive cervical cancer by SNP array. Patterns of copy number increase identified by 250K NspI array in log$_2$ ratio is shown from largest to smallest region on chromosome 20. Each vertical column represents a sample with genomic region representing from pter (top) to qter (bottom). Prefix "T" indicates primary tumor; "CL" indicates cell line. The blue-red scale bar (−1 to +1) at the bottom represents the copy number changes relative to mean across the samples. The intensities of blue and red indicate relative decrease and increase in copy numbers, respectively. Inferred copy number view of tumor T-1875 showing copy number changes from normal (2N) (red line) is shown on right. A G-banded ideogram of chromosome 20 is shown on extreme right. Two rectangle horizontal boxes indicate the identification of two focal amplicons. The genomic boundaries and the number of genes present in the amplicons are shown on right.

copy number alterations (CNA) (unpublished data). The dataset of chromosome 20 CNA from this analysis was used in the present study. Chromosome 20 CNA were found in 32 (40.5%) CC cases. Although all types of chromosome 20 changes were represented both in cell lines and primary tumors, the cell lines showed higher frequency of alterations (data not shown). Of the 32 tumors that exhibited chromosome 20 CNA, 29 (90.6%) cases showed gain and 11 (34.4%) cases showed losses. Of the latter cases that exhibited losses, 8 tumors showed concurrent CNI of 20q and loss of 20p regions. The remaining 3 cases showed only 20p deletion. Among the 29 cases that exhibited CNI, 15 showed entire chromosome 20 gain, eight had only 20q gain, whereas the remaining 6 tumors (T-1898, T-138, T-841, T-1875, T-130, and T-52) harbored regional gains on 20q (Fig. 1). Thus, these results demonstrate that one or more regions on 20q are frequent target of CNI in CC.

Identification of Focal Amplicons on 20q

To identify common minimal region(s) of CNI on chromosome 20, we examined the SNP data for smaller regions of amplification and gain. Notably, we found 9 cases with evidence of amplification

(>4-fold increase in raw copy number view), 3 on 20p, and 6 on 20q. The 20p amplicons were nonoverlapping and thus are nonrecurrent, while the 20q amplifications, mapped to 20q11.2 in 4 cases (T-1875, T-130, CL-SW756, and CL-SiHa) and 20q13.13 in 2 tumors (T-1875 and T-52), were found to be recurrent (Fig. 1). In addition, CNI (2.5–4 fold increase) of regions overlapping with 20q11.2 amplicon were also found in 3 other tumors (T-1898, T-138, and T-52). Therefore, the focal amplicon at 20q11.2 (amplicon 1) was defined based on at least 7 tumors. The 20q13.13 amplicon found in two tumors also showed focal gain of this region in an additional tumor (T-138). Thus, the identification of 20q13.13 amplicon (amplicon 2) was based on 3 tumors (Fig. 1). Therefore, we obtained evidence for the presence of two focal amplicons on 20q in CC.

FISH Validation of 20q Gain in CC

To validate the 20q CNI identified by SNP array, we performed FISH analysis using two locus specific probes (RP11-30F23 mapped to 20q13.12 and D20S108 probe mapped to 20q12) and two control probes (centromeres of chromosomes 20 and 11) on 74 invasive tumors. These include an independent panel of 36 tumors on paraffin-
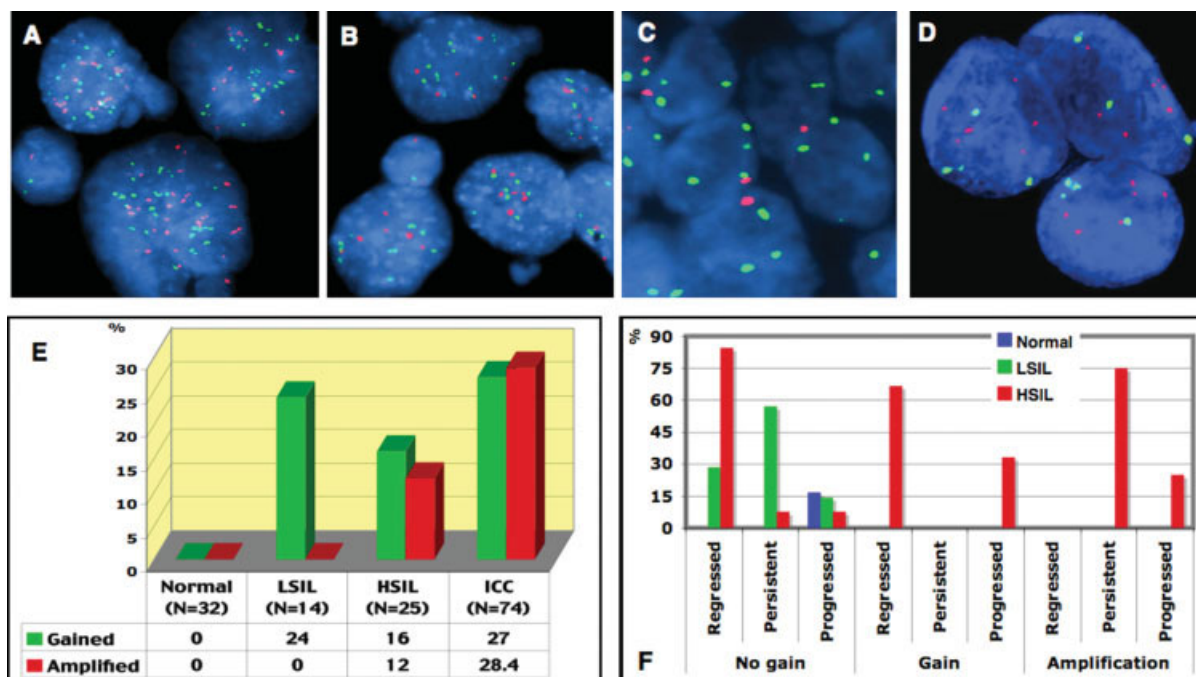
Figure 2. Role of chromosome 20 copy number gains identified by fluorescence in situ hybridization (FISH) in cervical cancer progression. A–D: FISH identification of 20q copy numbers in various stages of CC progression. A–C: invasive CC (ICC); Green signals represent BAC RP11-30F23 mapped to 20q13.1 and Spectrum Orange signals represent Chromosome 20 centromere used as control. Panel A: High-level amplification of both 20q13.1 and centromere signals on a paraffin section of an ICC. Panel B: Relative increase of copy numbers of 20q13.1 locus compared to Centromere 20 on Pap smear from an ICC. Panel C: Three to 4 copies of 20q13.1 locus and one copy of chromosome 20 centromere on a frozen section of an ICC. Panel D: FISH on Pap smear of HSIL showing 5–7 copies of D20S108 mapped to 20q12 (spectrum orange) and 3 copies of Centromere 11 (spectrum green) used as control. Panel E: Frequency of 20q gain and amplification in various stages of CC progression. Panel F: Role of 20q copy number alterations in CC progression. LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; ICC, invasive cervical cancer.

embedded tissue microarrays and 38 tumors as frozen sections or pap smears (the latter include 21 tumors also studied by SNP array) (Supplementary Table 1). A total of 41 (55.4%) tumors showed evidence for increased copies of 20q (Fig. 2A–C) (Table 1). Of these, 21 tumors showed ≥ 5 signals (amplification) whereas the remaining 20 specimens showed 3–4 signals (gain). All of the tumors that exhibited 20q CNI by SNP array also showed gain by FISH (data not shown). An average of 6.5 copies (range: 3–15) of 20q was found among the 41 cases that exhibited 20q gain, whereas only 3.8 copies (range: 1–8) of the centromere 20 and 3.2 copies (range: 2–8) of the centromere 11 were present. This data, thus, suggest that the 20q CNI is independent of ploidy of the tumor. These results, therefore, validate the SNP data and establish that 20q gain is a frequent genetic alteration in CC.

Chromosome 20 CNA in Relation to Clinico-Pathologic Characteristics

Next we evaluated the association of chromosome 20 CNA with clinico-pathologic features such as histology, age, tumor stage and size, treatment outcome, and HPV type by univariate analyses (Supplementary Table 1). No significant associations could be found between chromosome 20 CNA with histological type, age, stage or size of the tumor. Although no statistically significant difference was identified between all types of chromosome 20 CNA and clinical outcome, patients who died of cancer after treatment showed an overall higher incidence of amplifications (4/35; 11.4% cases) and deletions (4/35; 11.4% cases) compared

TABLE 1. FISH Identification of Chromosome 20q Copy Number Increase Using Various Locus Specific Probes in Invasive Cervical Cancer and Precancerous Lesions

Specimen class (N)[a]	Average probe copies/cell			
	CEP 20	RP11-30F23	LSI D20S108	CEP 11
Invasive cancer (41)	3.8	5.9	6.5	3.2
HSIL (7)	3.1	5.0	4.1	2.1
LSIL (3)	3.3	4.0	3.3	2.0

HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.
[a]Only cases that showed increase in 20q signals were utilized in calculating average copies per cell.

to patients that showed complete response and remain alive (1/32; 3.1% cases each) after follow-up periods of 1-72 months. The type of HPV infection did not show significant differences when considered all types of chromosome 20 CNAs. However, a statistically significant ($P = 0.05$) correlation was observed between the HPV type and the 20q CNI. Patients negative for HPV (3 of 4; 75%), HPV 18 (7 of 11; 63.6%), and multiple HPV infection including HPV18 (3 of 4; 75%) showed a higher frequency of 20q CNI compared to patients carrying HPV 16 (8 of 35; 22.9%) and other types of HPV infection (6 of 21; 28.6%) (Supplementary Fig. 3).

Transcriptome of Chromosome 20 in CC

Because chromosome 20 is one of the commonly gained chromosome in CC, we hypothesized that increased gene dosage may induce transcriptional activation of genes relevant to cellular transformation. To understand the transcriptional consequence of 20q gain, we used gene expression profiling of chromosome 20 probe dataset from Affymetrix U133A array analysis on 20 normal (age range, 27–64 year; Mean ± SD, 46.9 ± 7.6) squamous epithelial samples (including the 7 samples used in SNP array) and 29 CC cases (20 primary tumors; age range, 28–70 year; Mean ± SD, 48.9 ± 12.3; and 9 cell lines). To identify differentially expressed gene signatures of chromosome 20 in CC, we obtained all probe sets that exhibit significant ($P < 0.05$) differences between tumors and normal using the criteria described in materials and methods. This algorithm identified 77 nonredundant probe sets with significant differences in expression levels in tumors compared to normal. These include 11 probes with decreased expression and 66 probes with increased expression (Supplementary Fig. 1). Of the 66 overexpressed probes, 52 were mapped to 20q, 13 on 20p, and one (PTMA) probe maps to chromosome 2. Because our goal was to identify overexpressed genes in relation to chromosome 20 gain, we focused only on overexpressed gene dataset in the subsequent analyses. Of the 65 overexpressed probes, 57 belonged to known genes whereas 8 probes remained as expression sequence tags of unknown genes. The majority of chromosome 20 overexpressed genes belongs to distinct functional groups such as cell cycle regulation (*PCNA*, *MYBL2*, *CDC25B*, *UBE2C*, *E2F1*, *AURKA*, *TPX2*, *SYCP2*, *KIF3B*, and *DSN10*), nuclear function (*CSTF1*, *CSE1L*, *PSMA7*, *RALY*, *TCEA2*, *SOX12*, *TCFL5*, *ASXL1*, *SNRPB*, *ARFGAP1*, *CTNNBL1*,

and *DDX27*), transferase (*TGM2*, *RPN2*, *POFUT1*, and *B4GALT5*), and endopeptidase (*WFDC2* and *MMP9*) activity (Supplementary Table 2).

Identification of Overexpressed Target Genes of 20q CNI in CC

We next asked whether the overexpression of 65 genes on chromosome 20 in CC is a generalized tumor phenomenon or a consequence of chromosome 20 CNI. To identify the relationship between chromosome 20 CNI and gene expression, we used 29 cases that were analyzed for both chromosome 20 CNA and gene expression profiles. Supervised analysis using the overexpressed gene data set between tumors with ($N = 9$) and without ($N = 17$) chromosome 20 or 20q CNI identified 14 differentially expressed genes (*SS18L10*, *TCEA2*, *LAMA5*, *STX16*, *DDX27*, *RPN2*, *PIGT*, *TCFL5*, *AHCY*, *KIAA0406*, *PSMA7*, *C20orf20*, *ASXL1*, and *CDK5RAP1*). All these genes mapped to 20q (Fig. 3). This analysis, therefore, identified the overexpressed genes target of 20q CNI in CC. To identify additional genes that may have escaped the detection by the above algorithm, we further compared the expression profiles of tumors without gain of chromosome 20 ($N = 17$) and tumors carrying only gain of 20q ($N = 6$). This algorithm identified 8 overexpressed genes on 20q. These include 6 genes (*TCEA2*, *STX16*, *DDX27*, *ASXL1*, *AHCY*, and *PSMA7*) commonly identified by both analyses and two additional genes (a hypothetical protein LOC388796 and *ADRM1*) (Supplementary Fig. 2). Thus, we have identified several upregulated genes as a consequence of 20q CNI suggesting that one or more of these genes may play a central role in CC tumorigenesis.

Identification of Target Genes of Focal Amplicons on 20q

As stated above, we identified two focal amplicons at 20q11.2 and 20q13.13. The minimum shared region of amplicon 1 at 20q11.2 spans 4.1 Mb and harbor 96 coding genes. The amplicon 2 at 20q13.13 spans 3.1 Mb containing 42 genes (Fig. 1) (http://www.ncbi.nlm.nih.gov/mapview/maps; *Homo sapiens* Build 36.3). To identify transcriptional targets, we used all probes present on the U133A array within these two 20q amplicon intervals to compare expression levels between normal and tumor specimens that showed CNI within these regions. This analysis identified 9 overexpressed probe sets in amplicon 1 and 6 in amplicon 2 (Fig. 4). Of the 9 overexpressed transcripts in
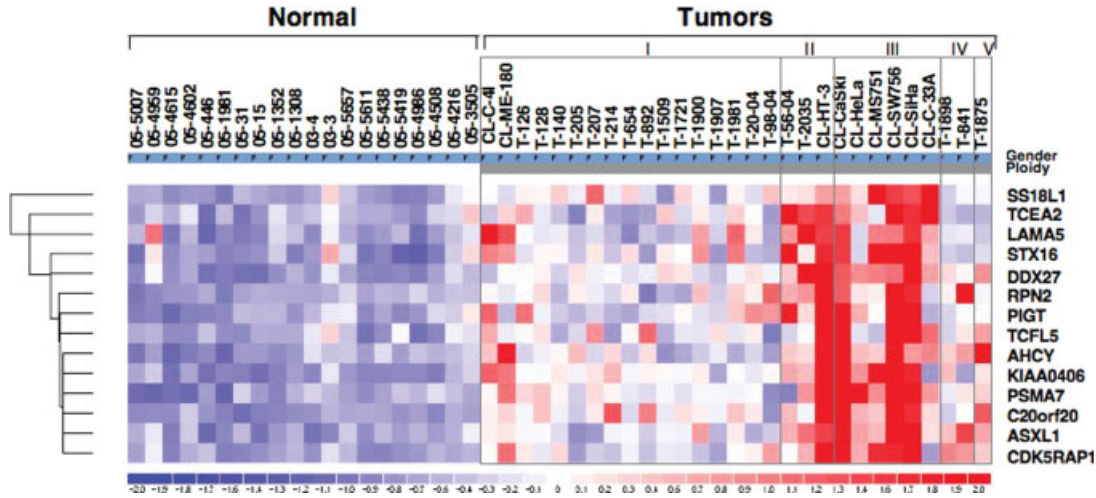
Figure 3. Supervised analysis of overexpressed genes identified as a consequence of gain of chromosome 20 or 20q in cervical cancer. Significantly differentially expressed genes were identified by filtering all of the overexpressed genes on chromosome 20 between the two tumor groups that showed gain of chromosome 20 or 20q and with out gain. In the matrix, each row represents the gene expression relative to group mean and each column represents a sample (shown on Top). T, represents primary tumor; CL, represents cell line. The dendrogram on left shows unsupervised clustering of genes differentially expressed between tumors with and without gain. The differentially expressed genes are shown on right. The scale bar (-2 to $+2$) on the bottom represents the level of expression with intensities of blue represents decrease and red for increase in expression. The groups within tumors shown at top represent no gain of chromosome 20 (I), whole chromosome 20 gain (II), only 20q gain (III), focal amplicon 1 (IV), and focal amplicon 2 (V).
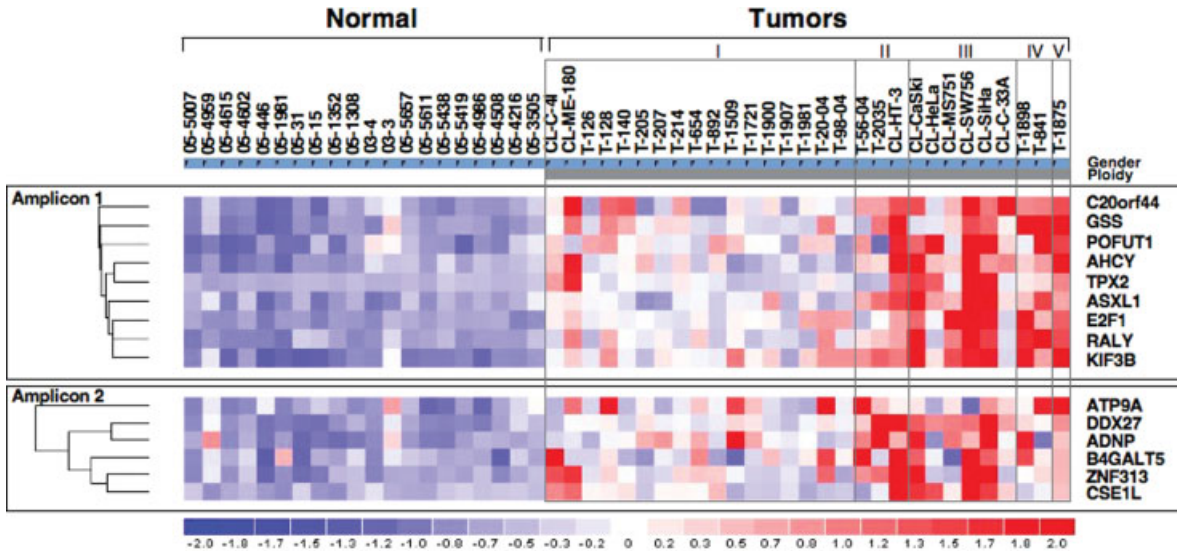

Figure 4. Supervised analysis of overexpressed genes identified as a consequence of focal amplicons on 20q in cervical cancer. Differentially expressed genes in relation to amplicons were identified by filtering all overexpressed genes on chromosome 20 between two groups; i.e., (i) using all tumors showing gain in the genomic region of Amplicon 1 and tumors without gain in the corresponding region; and (ii) using all tumors carrying gain in Amplicon 2 and tumors without gain in the same region. Top panel, Amplicon 1; Bottom panel, Amplicon 2. In the matrix, each row represents the gene expression relative to group mean and each column represents a sample (shown on Top). T, represents primary tumor; CL, represents cell line. The dendrogram on left shows clustering of genes differentially expressed between tumors with and without amplification. The differentially expressed genes are shown on right. The scale bar (-2 to $+2$) on the bottom represents the level of expression with intensities of blue represents decrease and red for increase in expression. The groups within tumors shown at top represent normal chromosome 20 (I), chromosome 20 gain (II), only 20q gain (III), focal amplicon 1 (IV), and focal amplicon 2 (V).

amplicon 1, eight belong to known genes (*GSS, POFUT1, AHCY, TPX2, ASXL1, E2F1, RALY*, and *KIF3B*) and one an open reading frame, *C20orf44*, of unknown function. Two of these genes (*AHCY* and *ASXL1*) also identified as overexpressed by the algorithm used to identify genes as a consequence of 20q gain. The overexpressed genes in amplicon 1 are functionally associated to amino acid metabolism/oxidative stress (*GSS* and *AHCY*), a notch signaling pathway (*POFUT1*), cell cycle regulation

(*TPX2*, *E2F1*, and *KIF3B*), a putative polycomb-group protein (*ASXL1*), and an RNA-binding protein (*RALY*). The overexpressed genes in the interval of amplicon 2 include nucleotide binding (*ATP9A* and *DDX27*), activity-dependent neuroprotector (*ADNP*) with a potential role in tumor proliferation, a gene encoding for UDP-Gal:beta-GlcNAc beta-1,4-galactosyltransferase (*B4GALT5*) with transferase activity, a zinc finger protein 313 (*ZNF313*), and a nuclear function protein (*CSE1L*). The *DDX27* gene identified as overexpressed within amplicon 2 was also identified as target of 20q gain. Therefore, we identified additional target-over expressed genes of relevance to various tumorigenic processes as consequence of amplifications on 20q. We have also examined the expression of these genes relative to GAPDH and found consistent overexpression with 20q gain and amplification (Supplementary Fig. 4).

Acquisition of 20q Gain Is an Early Genetic Event in CC Progression

The tumorigenic process in cervix is characterized by distinct morphological changes observed during the transition from normal epithelium to carcinoma through low-grade squamous intraepithelial lesions (LSIL) and high-grade SILs (HSIL). Currently, no biological or genetic markers are available to predict which precancerous lesions progress to invasive CC. To identify the earliest stage in CC development in which the 20q CNI occur, we used FISH assay on 71 consecutively ascertained pap smears simultaneously diagnosed by cytology as normal, squamous metaplasia or with atypical cells of undetermined significance (ASCUS) ($N = 32$), LSIL ($N = 14$), and HSIL ($N = 25$). Seven of 25 (28%) HSILs showed three or more copies of 20q (Fig. 2D). Of these, four HSILs exhibited gain while 3 showed evidence of amplification (Fig. 2E). Among the LSILs, three (21.4%) showed gain while none showed evidence of amplification (Fig. 2E) (Table 1). 20q gain was not found in any specimens diagnosed normal, squamous metaplasia or ASCUS. Thus, these data suggest that 20q gain occur as early as in LSIL stage while amplifications occur in HSIL and represent an early event in CC development.

The biological behavior of SILs varies where only a small proportion of HSILs progress to invasive cancer if left untreated and most LSILs persist or regress (Murthy et al., 1990; Schneider and Koutsky, 1992; Ostor, 1993). Morphological characterization alone does not permit the identification of HSILs at risk for progression from those that

regress or persist. Because of this, all HSILs are currently treated by surgical excision or with an ablative therapy. Identification of genetic signatures defining the subset of high-risk HSILs could alter the treatment strategies. To identify the role of 20q in progression of precancerous lesions, we obtained the follow up information (range, 1–29 months) from patients diagnosed normal ($N = 12$), LSIL ($N = 7$), and HSIL ($N = 19$). The patients with HSIL underwent standard treatment of care. Two patients initially diagnosed as normal with diploid compliment of 20q developed LSIL during 15 and 17 months follow up, respectively. No follow up data were available for the three patients diagnosed as LSIL with gain of 20q. However, one LSIL with disomy for 20q progressed to HSIL within 1 month of follow up, while two others regressed to normal. Of the 19 HSIL that had follow up information, 13 showed no gain and 6 exhibited gain or amplification of 20q. Among the 13 HSILs carrying two copies of 20q, 11 (84.6%) regressed to lower-grade or normal, while one (7.7%) persisted and one (7.7%) other progressed to invasive cancer. Of the 6 cases with increased copies of 20q, 3 patients showed gain and the 3 others showed amplification. Two of the patients with 20q gain regressed to lower grade and one (33.3%) progressed to invasive CC after 1–19 months follow up. All three (100%) patients who were diagnosed with HSIL and exhibited amplification of 20q persisted during 2–13 months follow up (Fig. 2F). Although this analysis was performed on a limited set of samples, a significant trend ($P < 0.05$) was observed among HSILs carrying 20q gain for an accompanying risk of persistence or progression (Fig. 2F). These data, thus, suggest that precancerous lesions harboring 20q CNI carry a potential risk for progression.

DISCUSSION

CC is a single diagnostic entity with differences in clinical behavior and response to therapy. Nearly one-third of patients diagnosed with invasive CC fail to respond to the current treatment protocols and die of disease. The standard histological techniques are not useful in stratifying CC into subclasses in response to treatment (Waggoner, 2003). Although several prognostic factors have been identified in CC, risk evaluation of progression and treatment response still remains elusive. A better understanding of genetic alterations in CC tumorigenesis might effectively identify therapeutic targets for successful treatment and predictive markers for progression. Toward this goal, our com-

TABLE 2. Genes Overexpressed as a Consequence of Chromosome 20q Gain and Amplification in Cervical Cancer

Gene	Description	Function	Role in cancer	Fold change[a]
E2F1	E2F transcription factor 1	Cell cycle regulation	(34)	4.7
KIF3B	Kinesin family member 3B	Chromosome segregation	(35)	3.1
TPX2	Targeting protein for Xklp2	Chromosome segregation	(36)	6.6
CSE1L	Chromosome segregation 1-like	Chromosome segregation	(37)	2.8
AHCY	S-adenosylhomocysteine hydrolase	Chromatin remodeling	(38)	3.8
ASXL1	Additional sex combs-like protein 1	Chromatin remodeling	(39)	2.8
C20orf20	Chromosome 20 open reading frame 20	Chromatin remodeling	(40)	3.9
PMSA7	Proteasome (prosome, macropain) subunit, alpha type, 7	Viral replication	(41) (42)	2.9
LAMA5	Laminin alpha 5	Extra-cellular matrix, viral replication	(16, 43)	2.8
TCEA2	Transcription elongation factor A protein 2	Transcription elongation	(44)	3.1
STX16	Syntaxin 16	Vesicular transport	–	2.6
DDX27	DEAD box protein 27	Spliceosome assembly	–	2.6
ADRM1	Adhesion regulating molecule 1	Cell Adhesion	(45)	2.6
PIGT	Phosphatidylinositol-glycan biosynthesis class T protein	Glycolipid biosynthesis	(46)	2.6
GSS	Glutathione synthase	Glutathione biosynthesis	–	2.3
POFUT1	Peptide-O-fucosyltransferase 1	Notch signaling	–	2.7
RALY	RNA-binding protein Raly	RNA splicing	–	3.2
ATP9A	ATPase class II type 9A	Ion transport	–	2.4
B4GALT5	UDP-Gal:beta-GlcNAc beta-1,4-galactosyltransferase 5	Glycosphingolipid biosynthesis	(47)	2.3
ZNF313	Zinc finger protein 313	Transcription factor	–	2.8

[a]Fold-change is calculated based on comparison between normal cervical squamous epithelium and tumors with chromosome 20 and 20q gain.

bined SNP and FISH analyses identified gain of 20q in over 50% of advanced CC confirming the previous findings that this abnormality is a common recurrent genetic alteration along with 3q and 5p gain (Heselmeyer et al., 1996; Rao et al., 2004). These data, therefore, suggest that 20q contains critical genes involved in the pathogenesis of CC.

Several previous studies have identified recurrent amplification and gain of 20q in many types of human cancers (Guan et al., 1996; Tanner et al., 1996; Hodgson et al., 2003; Hurst et al., 2004; Midorikawa et al., 2006; Koynova et al., 2007), including CC (Wilting et al., 2006; Kloth et al., 2007). Gain of 20q11.2-13.1 has also been associated with acquisition of drug resistance to tamoxifen in a human breast cancer cell line (Achuthan et al., 2001) and amplification of 20q11.2-12 is observed in human male germ cell tumors that are resistant to cisplatin therapy (Rao et al., 1998). In addition, in vitro models of HPV16 E6- and/or E7-immortalized human epithelial cells have shown to exhibit genomic instability and clonal chromosome abnormalities. Significantly, number of studies showed that the E7 transformed epithelial cells exhibit amplification of 20q. These data suggest that 20q amplification and possibly overexpression of specific genes contributes to HPV 16-E7-mediated immortalization and in overcoming cellular senescence of epithelial cells (Reznikoff et al., 1994;

Savelieva et al., 1997; Cuthill et al., 1999). Our results, demonstrating the occurrence of chromosome 20 gains early in LSIL and the concurrent overexpression of genes critical to cellular transformation, corroborate these studies. We, therefore, suggest that 20q gain represents a significant early genetic event in HPV-associated cellular transformation. Thus the evidences from both in vitro and in vivo studies implicate for the presence of one or more putative transformation genes on 20q in CC.

The genes we found upregulated in this study as a consequence of 20q gain or amplification are known to play specific roles in tumorigenic processes. The *E2F1*, *KIF3B*, *TPX2*, and *CSE1L* genes play pivotal roles in the cell cycle regulation and chromosome segregation (Table 2). Another class of genes that are upregulated as a target of 20q gain such as *AHCY*, *ASXL1*, and *C20orf20* play roles in mediating methylation and chromatin remodeling. The proteins encoded by *PMSA7* and *LAMA5* genes play a potential role in viral life cycle and replication are also overexpressed as targets of 20q gain. Specifically, laminin alpha 5 (*LAMA5*) has been shown to function as a transient receptor for HPV by binding virions and transferring them to adjacent cells by laminin 5 secreting keratinocytes, thus implicating a role for *LAMA5* in HPV viral infection and replication (Culp et al., 2006). The present study also identified over expression of a

number of other genes (*TCEA1*, *STX16*, *DDX27*, *ARDM1*, *PIGT*, *GSS*, *POFUT1*, *RALY*, *ATP9A*, *B4GALT5*, and *ZNF313*) whose function in cancer development is not well understood (Table 2).

The "high risk" HPV types (HPV 16 and HPV 18) encode transforming genes E6 and E7 that form complexes with p53 and pRB, respectively, resulting in suppression of their gene products and deregulation of the host cell cycle (Munger and Howley, 2002). Although both E6 and E7 oncoproteins of HPV16 and 18 are known to cause chromosomal instability, the specificity of each HPV type in causing targeted chromosome aberrations is unknown (Duensing and Munger, 2004). However, a number of genes have been shown to be differentially express between HPV 16 and HPV 18 infected tumors (Vazquez-Ortiz et al., 2007). To examine the relationship between HPV type and chromosome 20 copy number gains, we compared the type of HPV infection with different types of chromosome 20 abnormalities in CC and found an inverse correlation of 20q gain with HPV16. This limited data imply a relationship may exist between the presence of 20q gain and the type of HPV infection in human CC transformation, which needs to be elucidated by further studies.

A significant finding in the present study is the identification of 20q genomic CNI and the associated deregulated gene expression suggesting a functional role for this chromosomal region in the development and progression of CC. A second observation is that of the identification of two focal amplicons at 20q11.2 and 20q13.2. The HPV 16 E7 immortalized clones from human urothelial epithelial cells that show 20q13.2 amplification exhibit growth advantage over 20q gained clones (Savelieva et al., 1997) and similarly the breast cancer cells with 20q13.2 amplification exhibits high proliferative index and poor prognosis (Tanner et al., 1995). The close recapitulation of genetic alterations caused by HPV16 E7 immortalized cells and the clinical specimens from CC patients provide strong evidence that 20q gain plays a role in cervical carcinogenesis. Furthermore, the identification of this change in cervical intraepithelial lesions provide new insights into the role of 20q in the progression of CC and the diagnostic utility in identifying precancerous lesions at high-risk for progression to invasive cancer.

In conclusion, the 20q gain as a recurrent genetic change and the overexpressed genes identified here will form the basis for functional testing to develop molecular target based therapies for CC. Identification of chromosome 20 CNI in precancerous lesions may prove to be a potential prognostic molecular marker in distinguishing high-risk precancerous lesions to progress to invasive cancer.

REFERENCES

Achuthan R, Bell SM, Roberts P, Leek JP, Horgan K, Markham AF, MacLennan KA, Speirs V. 2001. Genetic events during the transformation of a tamoxifen-sensitive human breast cancer cell line into a drug-resistant clone. Cancer Genet Cytogenet 130:166–172.

Culp TD, Budgeon LR, Marinkovich MP, Meneguzzi G, Christensen ND. 2006. Keratinocyte-secreted laminin 5 can function as a transient receptor for human papillomaviruses by binding virions and transferring them to adjacent cells. J Virol 80:8940–8950.

Cuthill S, Agarwal P, Sarkar S, Savelieva E, Reznikoff CA. 1999. Dominant genetic alterations in immortalization: Role for 20q gain. Genes Chromosomes Cancer 26:304–311.

Duensing S, Munger K. 2004. Mechanisms of genomic instability in human cancer: Insights from studies with human papillomavirus oncoproteins. Int J Cancer 109:157–162.

Gius D, Funk MC, Chuang EY, Feng S, Huettner PC, Nguyen L, Bradbury CM, Mishra M, Gao S, Buttin BM, Cohn DE, Powell MA, Horowitz NS, Whitcomb BP, Rader JS. 2007. Profiling microdissected epithelium and stroma to model genomic signatures for cervical carcinogenesis accommodating for covariates. Cancer Res 67:7113–7123.

Guan XY, Xu J, Anzick SL, Zhang H, Trent JM, Meltzer PS. 1996. Hybrid selection of transcribed sequences from microdissected DNA: Isolation of genes within amplified region at 20q11-q13.2 in breast cancer. Cancer Res 56:3446–3450.

Harris CP, Lu XY, Narayan G, Singh B, Murty VV, Rao PH. 2003. Comprehensive molecular cytogenetic characterization of cervical cancer cell lines. Genes Chromosomes Cancer 36:233–241.

Heselmeyer K, Schrock E, du Manoir S, Blegen H, Shah K, Steinbeck R, Auer G, Ried T. 1996. Gain of chromosome 3q defines the transition from severe dysplasia to invasive carcinoma of the uterine cervix. Proc Natl Acad Sci USA 93:479–484.

Hodgson JG, Chin K, Collins C, Gray JW. 2003. Genome amplification of chromosome 20 in breast cancer. Breast Cancer Res Treat 78:337–345.

Hurst CD, Fiegler H, Carr P, Williams S, Carter NP, Knowles MA. 2004. High-resolution analysis of genomic copy number alterations in bladder cancer by microarray-based comparative genomic hybridization. Oncogene 23:2250–2263.

Kloth JN, Oosting J, van Wezel T, Szuhai K, Knijnenburg J, Gorter A, Kenter GG, Fleuren GJ, Jordanova ES. 2007. Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer. BMC Genomics 8:53.

Koynova DK, Jordanova ES, Milev AD, Dijkman R, Kirov KS, Toncheva DI, Gruis NA. 2007. Gene-specific fluorescence in-situ hybridization analysis on tissue microarray to refine the region of chromosome 20q amplification in melanoma. Melanoma Res 17:37–41.

Li C, Wong WH. 2001. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. Proc Natl Acad Sci USA 98:31–36.

Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. 2004. dChipSNP: Significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics 20:1233–1240.

Midorikawa Y, Yamamoto S, Ishikawa S, Kamimura N, Igarashi H, Sugimura H, Makuuchi M, Aburatani H. 2006. Molecular karyotyping of human hepatocellular carcinoma using single-nucleotide polymorphism arrays. Oncogene 25:5581–5590.

Mitra AB, Murty VV, Pratap M, Sodhani P, Chaganti RS. 1994. ERBB2 (HER2/neu) oncogene is frequently amplified in squamous cell carcinoma of the uterine cervix. Cancer Res 54:637–639.

Munger K, Howley PM. 2002. Human papillomavirus immortalization and transformation functions. Virus Res 89:213–228.

Murthy NS, Sehgal A, Satyanarayana L, Das DK, Singh V, Das BC, Gupta MM, Mitra AB, Luthra UK. 1990. Risk factors related to biological behaviour of precancerous lesions of the uterine cervix. Br J Cancer 61:732–736.

Narayan G, Arias-Pulido H, Koul S, Vargas H, Zhang FF, Villella J, Schneider A, Terry MB, Mansukhani M, Murty VV. 2003a. Frequent promoter methylation of CDH1, DAPK, RARB, and HIC1

genes in carcinoma of cervix uteri: Its relationship to clinical outcome. Mol Cancer 2:24.

Narayan G, Pulido HA, Koul S, Lu XY, Harris CP, Yeh YA, Vargas H, Posso H, Terry MB, Gissmann L, Schneider A, Mansukhani M, Rao PH, Murty VV. 2003b. Genetic analysis identifies putative tumor suppressor sites at 2q35-q36.1 and 2q36.3-q37.1 involved in cervical cancer progression. Oncogene 22:3489–3499.

Narayan G, Bourdon V, Chaganti S, Arias-Pulido H, Nandula SV, Rao PH, Gissmann L, Durst M, Schneider A, Pothuri B, Mansukhani M, Basso K, Chaganti RS, Murty VV. 2007. Gene dosage alterations revealed by cDNA microarray analysis in cervical cancer: Identification of candidate amplified and overexpressed genes. Genes Chromosomes Cancer 46:373–384.

Ostor AG. 1993. Natural history of cervical intraepithelial neoplasia: A critical review. Int J Gynecol Pathol 12:186–192.

Pulido HA, Fakruddin MJ, Chatterjee A, Esplin ED, Beleno N, Martinez G, Posso H, Evans GA, Murty VV. 2000. Identification of a 6-cM minimal deletion at 11q23.1-23.2 and exclusion of PPP2R1B gene as a deletion target in cervical cancer. Cancer Res 60:6677–6682.

Rao PH, Houldsworth J, Palanisamy N, Murty VV, Reuter VE, Motzer RJ, Bosl GJ, Chaganti RS. 1998. Chromosomal amplification is associated with cisplatin resistance of human male germ cell tumors. Cancer Res 58:4260–4263.

Rao PH, Arias-Pulido H, Lu XY, Harris CP, Vargas H, Zhang FF, Narayan G, Schneider A, Terry MB, Murty VV. 2004. Chromosomal amplifications, 3q gain and deletions of 2q33-q37 are the frequent genetic changes in cervical carcinoma. BMC Cancer 4:5.

Reznikoff CA, Belair C, Savelieva E, Zhai Y, Pfeifer K, Yeager T, Thompson KJ, DeVries S, Bindley C, Newton MA, Sekhon G, Waldman F. 1994. Long-term genome stability and minimal genotypic and phenotypic alterations in HPV16 E7-, but not E6-, immortalized human uroepithelial cells. Genes Dev 8:2227–2240.

Savelieva E, Belair CD, Newton MA, DeVries S, Gray JW, Waldman F, Reznikoff CA. 1997. 20q gain associates with immortalization: 20q13.2 amplification correlates with genome instability in human papillomavirus 16 E7 transformed human uroepithelial cells. Oncogene 14:551–560.

Schneider A, Koutsky LA. 1992. Natural history and epidemiological features of genital HPV infection. IARC Sci Publ (119):25–52.

Schwab M. 1999. Oncogene amplification in solid tumors. Semin Cancer Biol 9:319–325.

Tanner MM, Tirkkonen M, Kallioniemi A, Holli K, Collins C, Kowbel D, Gray JW, Kallioniemi OP, Isola J. 1995. Amplification of chromosomal region 20q13 in invasive breast cancer: Prognostic implications. Clin Cancer Res 1:1455–1461.

Tanner MM, Tirkkonen M, Kallioniemi A, Isola J, Kuukasjarvi T, Collins C, Kowbel D, Guan XY, Trent J, Gray JW, Meltzer P, Kallioniemi OP. 1996. Independent amplification and frequent coamplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. Cancer Res 56:3441–3445.

Vazquez-Ortiz G, Garcia JA, Ciudad CJ, Noe V, Penuelas S, Lopez-Romero R, Mendoza-Lorenzo P, Pina-Sanchez P, Salcedo M. 2007. Differentially expressed genes between high-risk human papillomavirus types in human cervical cancer cells. Int J Gynecol Cancer 17:484–491.

Waggoner SE. 2003. Cervical cancer. Lancet 361:2217–2225.

Wilting SM, Snijders PJ, Meijer GA, Ylstra B, van den Ijssel PR, Snijders AM, Albertson DG, Coffa J, Schouten JP, van de Wiel MA, Meijer CJ, Steenbergen RD. 2006. Increased gene copy numbers at chromosome 20q are frequent in both squamous cell carcinomas and adenocarcinomas of the cervix. J Pathol 209:220–230.