

# Integrative genomic approaches in cervical cancer: implications for molecular pathogenesis

Gopeshwar Narayan<sup>1,2</sup> & Vundavalli V Murty<sup>1</sup>

<sup>1</sup>Department of Pathology & Cell Biology, Institute for Cancer Genetics, Columbia University Medical Center, 1130 St. Nicholas Avenue, New York, NY, USA

<sup>2</sup>Banaras Hindu University, Varanasi, 221 005, India

\*Author for correspondence: Tel.: +1 212 851 4621 ■ Fax: +1 212 851 5284 ■ [vvm2@columbia.edu](mailto:vvm2@columbia.edu)

Cervical cancer (CC) as a single diagnostic entity exhibits differences in clinical behavior and poor outcomes in response to therapy in advanced tumors. Although infection of high-risk human papillomavirus is recognized as an important initiating event in cervical tumorigenesis, stratification of CC into subclasses for progression and response to treatment remains elusive. Existing knowledge of genetic, epigenetic and transcriptional alterations is inadequate in addressing the issues of diagnosis, progression and response to treatment. Recent technological advances in high-throughput genomics and the application of integrative approaches have greatly accelerated gene discovery, facilitating the identification of molecular targets. In this article, we discuss the results obtained by preliminary integrative analysis of DNA copy number increases and gene expression, utilizing the two most common copy number-gained regions of 5p and 20q in identifying gene targets in CC. These analyses provide insights into the roles of genes such as *RNASEN*, *POLS* and *SKP2* on 5p, *KIF3B*, *RALY* and *E2F1* at 20q11.2 and *CSE1L*, *ZNF313* and *B4GALT5* at 20q13.13. Future integrative applications using additional datasets, such as mutations, DNA methylation and clinical outcomes, will raise the promise of accomplishing the identification of biological pathways and molecular targets for therapies for patients with CC.

Cervical cancer (CC) is the second most common malignancy among women worldwide, accounting for an estimated 0.5 million new cases and 275,000 deaths each year [1]. The development of squamous cell carcinoma proceeds by distinct morphological changes from normal epithelium to carcinoma through low-grade squamous intraepithelial lesions (LSILs)/low-grade cervical intraepithelial neoplasia (LCIN) and high-grade SILs (HSIL)/high-grade CIN (HCIN). The biological behavior of CINs varies, where only certain high-grade lesions progress to invasive cancer when untreated [2,3]. Owing to its distinctive precancerous stages, the accessibility for early detection and variability in biologic progression, CC provides a paradigm for studying the sequential accumulation of genetic alterations in tumor progression. Although infection with high-risk human papillomavirus (HPV) types has been implicated as a major etiological factor [4], CC as a single diagnostic entity exhibits differences in clinical behavior and response to therapy where advanced tumors poorly respond to chemo/radiotherapy. The stratification of CC into subclasses for progression and response to treatment remain elusive. Existing knowledge of genetic and epigenetic alterations

remains uninformative in addressing the issues of diagnosis, tumor progression and response to treatment.

Since the first description of alterations in chromosomal content in CC, considerable efforts have been made to understand the genetic basis of its development [5,6]. These studies have largely focused on gaining information on several classes of genetic alterations, ranging from chromosomes to genes. These include alterations in chromosome copy numbers (gains, losses, amplifications and rearrangements), gene mutations, epigenetic modifications and transcriptional changes. The chromosomal translocation mechanism of deregulation, which is common in hematological and mesenchymal malignancies, is an uncommon or unknown phenomenon in CC. Despite extensive genetic and molecular studies, only a handful of leads obtained were causally implicated in CC development. A census of genomic alterations with sufficient evidence (to meet the criteria for replicate studies, more than one study had to identify the change, with losses present in >20% cases and gains in >10% cases) supporting a role in CC is presented in TABLES 1 & 2 and is discussed in the following sections.

## Keywords

- amplification ■ cervical carcinoma ■ chromosome 5p
- chromosome 20q
- chromosome alteration
- gene expression
- integrative genomics
- precancerous lesion
- single nucleotide polymorphism array

Table 1. Genomic copy number amplifications, gains and losses identified by comparative genomic hybridization and loss of heterozygosity approaches in cervical cancer.

Alteration	Invasive cancer	Precancer	Potential genes (region: gene)	Ref.
<b>Amplifications<sup>†</sup></b>				
Recurrent	3q21, 3q26–q29, 5p, 7q22, 8q23–q24, 8q24.3, 9p22, 9p23–24, 10q21, 11q13, 11q21, 11q22–23, 12p13, 14q12, 17q12, 17q25, 19q13.1, 20q11.2, 20q13.1	None	3q27–29: <i>RFC4</i> , <i>MUC4</i> , <i>HRASLS</i> 5p: <i>ARSB</i> , <i>RPL37</i> , <i>SKP2</i> , <i>OSMR</i> , <i>TARS</i> , <i>RAD1</i> , <i>SUB1</i> , <i>RNASEN</i> , <i>POLS</i> , <i>SDHA</i> , <i>TRIO</i> , <i>NDUFS6</i> , <i>TRIP13</i> , <i>PDCC6</i> , <i>RAI14</i> , <i>SLC12A7</i> , <i>DAP</i> , <i>MYO10</i> , <i>BASPI1</i> , <i>NNT</i> , <i>TARS</i> , <i>PAIP1</i> , <i>CCT5</i> , <i>NDUFS6</i> , <i>BRD9</i> , <i>FASTKD3</i> , <i>BXDC2</i> 8q24: <i>PTK2</i> , <i>RECQL4</i> 11q22.2: <i>MMPT1</i> , <i>MMPT3</i> 17q12: <i>ERBB2</i> 20q13.1: <i>ATP9A</i> , <i>DDX27</i> , <i>ADNP</i> , <i>B4GALT5</i> , <i>ZNF313</i> , <i>CSE1L</i> , <i>UBE2C</i> 20q11.1: <i>C20orf20</i> , <i>GSS</i> , <i>POFUT1</i> , <i>AHCY</i> , <i>TPX2</i> , <i>ASXL1</i> , <i>E2F1</i> , <i>RALLY</i> , <i>KIF3B</i> 17q21–22: <i>ABCC3</i>	Supplementary online material <sup>‡</sup>
Nonrecurrent	1q31, 2q32, 7p11, 10q24, 12q15, 16q23–q24, 17p11.2, 17q22, 18p11.2, 20p	None		
<b>Gains</b>				
Replicate studies	1q22–q23, 1q25.3–q32.1, 3q24–29, 5p12–p13, 6p, 7q11.22–q11.23, 7q31.1–q31.2, 8q24.13–q24.22, 9q33.2–q34.3, 17q25.1–q25.2, 17q21–q22, 19q13.3, 20q11.22, 20q11.21–q13.33, 20q11.21–q12, 20q13.12–q13.31, 20q12, Xp11.2–p11.3, Xq12, Xq22, Xq28	1, 3q24–29	19q13.3: <i>LIG1</i> Xp11: <i>SMC1L1</i> Xq12: <i>KIF4A</i> Xq22: <i>TMSNB</i> Xq28: <i>CSAG2</i>	Supplementary online material <sup>‡</sup>
Nonreplicate studies	1p33–35, 1p36, 1p31.3–p21.1, 1q12–31.1, 1q32.2–q44, 3q13–23, 3p26.3–p26.1, 3p14.3–p14.2, 3q21–22, 5q31–qter, 11q22, 12p13, 14q32, 15q, 16p, 16q22.1–q23.2, 19p13, 20p13–p11.21	1p36.11–p35.2, 1p31.3–p21.1, 1q25.3–q32.1, 1q32.2–q44, 3p26.3–p26.1, 3p14.3–p14.2, 3q11.2–q29, 7q31.1–q31.2, 20p13–p11.21, 20q12	1p36: <i>EPHB2</i> 1p34.3: <i>CDC48</i> 1q22–23: <i>AIM2</i> 5q31.3: <i>CENTD3</i> 14q32.3: <i>AKT1</i> 19p13.3: <i>SMARCA4</i> 3q21–23: <i>DTX3L</i> , <i>PIK3R4</i> , <i>ATP2C1</i> , <i>SLC25A36</i>	
<b>Losses</b>				
Replicate studies	2q33–q37, 3p12–23, 4p16.3–p16.1, 4, 4q, 4q28.3–q32.1, 4q13.3, 4q35.2, 4q28, 6q, 8p23.3, 8p12–21.3, 8q23.2–q23.3, 9p, 11p15.5, 11q13.3, 11q22.3–25, 13q12.11–13q14.3, 13q14.3–q21.33, 13q31.1–q31.3, 17p13.3, 18q11.2–18q23	2q35–qter, 3p12–21, 4, 4p16.3–p16.1, 4q31.21, 4q35.2	4q: <i>PCDH10</i>	Supplementary online material <sup>‡</sup>
Nonreplicate studies	1p32.3, 6p22–24, 10p, 10q11.23–q21.3, 16q24.2–q24.3, 17q25.3	11q13.3, 11q23–qter, 16q24.2–q24.3, 17p13.3, 17q25.3, 19p13.11–q12		

<sup>†</sup>Recurrent amplifications present in two or more tumors and nonrecurrent amplifications in only isolated cases.<sup>‡</sup>See [www.futuremedicine.com/doi/suppl/10.2217/FON.10.114](http://www.futuremedicine.com/doi/suppl/10.2217/FON.10.114).

Table 1. Genomic copy number amplifications, gains and losses identified by comparative genomic hybridization and loss of heterozygosity approaches in cervical cancer.

Alteration	Invasive cancer	Precancer	Potential genes (region: gene)	Ref.
<b>Losses (cont.)</b>				
Loss of heterozygosity (>20%)	1q21–42, 2q33–q37, 3p13–21.1, 3p14.1–p24, 4p, 4q34–qter, 5p15.1–15.33, 6p21.2, 6p25–21.3, 6q16.3–q27, 8p23–p21, 8q22.2, 9p21, 10p15–p14, 10q23.3, 11p15.5, 11p15.3–p15.11, 11p13–p12, 11q13, 11q23.1–q24.1, 13q12–14, 17p13–p12, 17q12–q21, 17q11.2, 18q21–q22, 19p13.3, 19q13.4, Xq21–q22	2q33–q37, 3p14.1–p24, 3p12–14, 4p, 4q, 5p15, 6p25–21.3, 6q, 10p15–p14, 11q23.3–q25, 11q22–q24, 13q12–14, 17p13, 17q12–q21	2q33–q37: CFLAR, CASP10, PPP1R7 3p13–p24: FHIT, CTNMB1, TβR-II, RASSF1A, RARβ 4p: SLIT2 6p21–p23: CD83, TAP1, NOL7, HLA class I 9p21: p16INK4a 10q23.3: PTEN 11q13–q24: MEN1, INT2, TRIM29, POUF3 13q14: RB 17p13: TP53 17q11: CCL2	Supplementary online material <sup>†</sup>

<sup>†</sup>Recurrent amplifications present in two or more tumors and nonrecurrent amplifications in only isolated cases.

<sup>‡</sup>See [www.futuremedicine.com/doi/suppl/10.2217/FON.10.114](http://www.futuremedicine.com/doi/suppl/10.2217/FON.10.114).

### Genetic & epigenetic signatures of cervical cancer

The changes listed in TABLE 1 are of several classes of genomic alterations, resulting in copy number gains and losses. Amplification ( $\geq 5$  copies) of specific regions of the genome is a common phenomenon in many human tumors targeting overexpression of dominantly acting genes [7]. To date, the extent of gene amplification, as well as the genes involved, are not completely realized in CC. However, upon review of the literature on amplifications, we identified a total of 20 recurrent and ten nonrecurrent amplifications (TABLE 1) [8,9]. A number of potential genes mapped to these amplicons have also been identified. However, so far, none of these were proven to be target genes, owing to insufficient biological evidence for contribution as cancer genes in CC.

A similar situation exists for copy number changes resulting in gains ( $>2.5$  copies) or losses ( $<1.5$  copies) of specific genomic regions (TABLE 1). For simple gains of copy numbers, a large number ( $n = 21$ ) of recurrently gained/over-represented regions were shown by replicate studies. Of these, chromosome 1 and 3q24–q29 regions were also shown to occur early in precancerous lesions [10]. Gains of several other chromosomal regions have been observed in single-reported studies in invasive CC, as well as in precancerous lesions. Most of these are very large regions of genomic gains and were not further refined by analysis in additional tumors. In the absence of further studies, these gains (e.g., gains of 3q and 5p spanning larger genomic regions or even the entire chromosomal arm) remain uninformative (TABLE 1) [10–12].

A large body of data on genetic deletions has been derived either by loss of heterozygosity or comparative genomic hybridization/array-based studies. A large number of deleted regions meeting the criteria that we applied ( $>20\%$  cases with deletion and/or confirmed by more than one study) have also been reported (TABLE 1). Some of these deletions were also shown to arise at early precancerous stages. For example, 2q35–q37 regions of deletions spanning a 9.5-Mb physical distance showed deletions in high- and low-grade CINs [8]. These findings of complex patterns on genetic losses suggest loss of function of one or more proliferation-regulating genes in each of these regions and their involvement in malignant progression of cervical epithelium. Although the expression profiling allowed the identification of a subset of candidate genes, with loss accompanied

Table 2. Mutational and epigenetic promoter hypermethylation alteration of genes in cervical cancer.

Alteration	Invasive cancer	Precancerous lesions	Ref.
<b>Mutation<sup>†</sup></b>			
Replicate studies	<i>TP53</i> <sup>‡</sup> , <i>CDKN2A</i> , <i>PTEN</i>	None	Supplementary online material <sup>§</sup>
Nonreplicate studies	<i>HRAS</i> , <i>HLA B-15</i> , <i>HLA A-24</i> , <i>KRAS</i> , <i>BRAF</i> , <i>CD83</i> , <i>RNASEL</i> , <i>PIK3CA</i> , <i>LKB1</i>	None	Supplementary online material <sup>§</sup>
<b>Promoter hypermethylation</b>			
Replicate studies	<i>APC</i> , <i>CADM1</i> , <i>CALCA</i> , <i>CCNA1</i> , <i>CDKN2A</i> , <i>CDH1</i> , <i>CDH13</i> , <i>COX2</i> , <i>DAPK</i> , <i>FHIT</i> , <i>GSTP1</i> , <i>HIC1</i> , <i>MGMT</i> , <i>PCDH10</i> , <i>RARB</i> , <i>RASSF1A</i> , <i>THBS1</i> , <i>TIMP3</i>	<i>CDKN2A</i> , <i>CCNA1</i> , <i>DAPK</i> , <i>MGMT</i>	Supplementary online material <sup>§</sup>
Nonreplicate studies	<i>AHRR</i> , <i>BLU</i> , <i>C13ORF18</i> , <i>TFPI2</i> , <i>C1ORF166</i> , <i>NPTX1</i> , <i>CAGE</i> , <i>CDKN2B</i> , <i>CHFR</i> , <i>ESR1</i> , <i>TP73</i> , <i>ESR1</i> , <i>FANCF</i> , <i>HLTF1</i> , <i>IGSF4</i> , <i>MAL</i> , <i>MYOD1</i> , <i>PGR</i> , <i>POU2F3</i> , <i>PTEN</i> , <i>SOCS1</i> , <i>hTERT</i> , <i>HSPA2</i> , <i>MLH1</i> , <i>TWIST</i> , <i>SLIT1</i> , <i>SLIT2</i> , <i>SLIT3</i> , <i>ROBO1</i> , <i>ROBO3</i> , <i>SFRP1</i> , <i>SFRP2</i> , <i>SFRP4</i> , <i>SFRP5</i> , <i>SOCS2</i> , <i>SOX1</i> , <i>PAX1</i> , <i>LMX1A</i> , <i>NKX6-1</i> , <i>WT1</i> , <i>ONECUT1</i>	<i>AHRR</i> , <i>BLU</i> , <i>C13ORF18</i> , <i>CADM1</i> , <i>CCNA1</i> , <i>CDH1</i> , <i>CDH13</i> , <i>FHIT</i> , <i>GSTP1</i> , <i>LMX1A</i> , <i>MAL</i> , <i>hMLH1</i> , <i>NKX6-1</i> , <i>ONECUT1</i> , <i>PAX1</i> , <i>PCDH10</i> , <i>RARB</i> , <i>RASSF1A</i> , <i>ROBO1</i> , <i>ROBO3</i> , <i>SLIT1</i> , <i>SLIT2</i> , <i>SLIT3</i> , <i>SOCS2</i> , <i>SOX1</i> , <i>TIMP3</i> , <i>WT1</i>	
<sup>†</sup> One or more cases considered positive.			
<sup>‡</sup> Commonly seen in human papilloma virus-negative tumors.			
<sup>§</sup> See <a href="http://www.futuremedicine.com/doi/suppl/10.2217/FON.10.114">www.futuremedicine.com/doi/suppl/10.2217/FON.10.114</a> .			

by decreased expression, no known tumor-suppressor genes with evidence of mutations in the second allele have been implicated thus far in CC, with the exception of rare instances for mutations in *TP53* at 17p13, *CDKN2A* at 9p21 and *PTEN* at 10q23.3 (TABLES 1 & 2).

As stated previously, mutations in tumor-suppressor genes are infrequently reported in CC and, in most instances, the mutations have not been confirmed by independent studies (TABLE 2). In the absence of mutations, tumor-suppressor genes may be inactivated in the recurrently deleted chromosomal regions by alternative mechanisms such as epigenetic modifications. One of the most well-established epigenetic changes is the promoter DNA hypermethylation-mediated gene silencing. A large number of genes (e.g., *CDH1*, *DAPK*, *HIC1* and *PCDH10*) exhibiting promoter hypermethylation and associated downregulated expression of the gene have been reported in CC by multiple studies (TABLE 2) [13–16]. In addition, several other genes also shown to be methylated by a single study, where a subset of these cannot be completely excluded as targets in the deleted regions (TABLES 1 & 2).

Based on the aforementioned discussion, the shortness of causal genetic mutations in CC, although its genomes exhibit complex chromosomal alterations is due to insufficient evidence offered by these studies to identify a cancer gene. Several supportive studies are usually required to establish the role of a specific gene, including finer physical and transcriptional mapping of the altered regions, examination of

epigenetic mechanisms, functional analysis of target genes, correlations with clinical outcome and efficacy of drugs targeted against specific genes. For most genetic alterations reported in CC, such studies are lacking and often not feasible for obtaining these data. Added to this complexity, molecular heterogeneity within CC and biological effects of multiple genes in each of the affected genomic regions constitute major obstacles in understanding its pathogenesis.

A number of studies have attempted using single high-throughput approaches to unravel the genetic, epigenetic and transcriptional alterations in CC [17–21]. As noted previously, these studies have revealed specific genomic, expression and epigenetic alterations. However, the failure to identify target genes in CC is largely owing to a lack of understanding of the relationship between how copy number and epigenetic markers influence transcription. How each change influence the other (e.g., amplification or copy number increases on gene overexpression or methylation markers and deletions on downregulated gene expression) remains largely unknown in CC. Integrative genomic analyses involving simultaneous assessment of DNA copy numbers, gene expression, mutations and methylation markers, such as cytosine methylation and histone tail modifications, have been demonstrated as potential approaches in identifying the candidate genes [22,23].

Cervical cancer genomes typically harbor multiple chromosome aberrations and epigenetic modifications, resulting in deregulated transcriptomes. These changes might play

roles in driving malignant transformation. Understanding the relationship between multi-dimensional levels of genomic modifications might expand our knowledge of the molecular basis of CC. As stated previously, CC genomes are characterized by a number of recurrent genomic copy number losses and gains. Of those exhibiting gains, chromosomal regions 3q, 5p, 20q and 1q were the most common targets [9,10]. In this article, we focus the analysis on two of these genomic regions in CC, 5p and 20q, and the outcome of integrative genomic analysis using copy number increases and gene expression is presented.

### Genomic & transcriptional analysis of chromosomes 5p & 20q

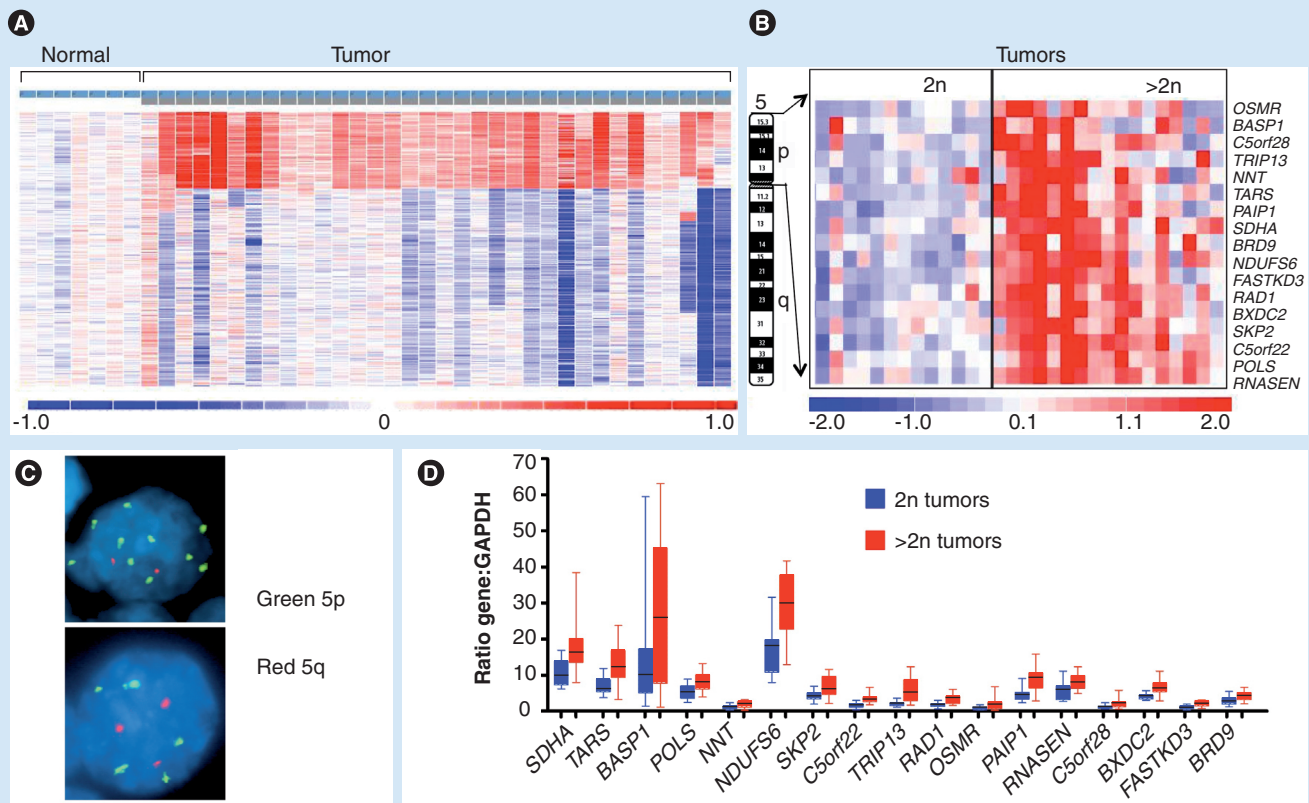
Genomic copy number alterations (CNAs) were identified utilizing Affymetrix (CA, USA) 250K *NspI* single nucleotide polymorphism (SNP) array platform and analyzed using the dChip analytical algorithms in 79 untreated primary CC samples [12,24]. Although this analysis uncovered a multitude of both known and unknown, as well as frequent and rare, altered CNAs, the 5p and 20q regions were the most significant recurrent focal copy number gains. Of these, the 5p CNA gains were found in 43% of tumors, and the 20q CNA gains or amplifications were found in 37% of tumors. The abundance of transcripts of protein-coding genes was measured by the Affymetrix U133A platform and analyzed using the dChip software algorithms among 42 CC cases [12,24]. Expression arrays were normalized using a median-intensity array from normal, as a baseline array using an invariant set normalization, as described previously [12,24]. Briefly, a list of differentially expressed genes with a twofold change was identified, with group means at 90% CIs. A list of overexpressed genes mapped to chromosomes 5 and 20 were identified further and used in subsequent supervised analyses using defined criteria to further obtain overexpressed gene signatures of specific chromosomal regions. The resulting gene expression datasets were utilized to correlate with 5p and 20q gains in order to identify expression patterns that were associated with CNAs.

### Integrative genomic & transcriptional profiles identify target genes of 5p gains

We identified gains of the entire 5p chromosomal arm in CC and no minimal regions of amplification or gain could be delineated (FIGURE 1A). The duplications of entire chromosomal arms resulting

from isochromosome formation in human tumors are not uncommon (e.g., i[12p] in male testicular germ cell tumor [25] and i[5p][10] in several types of adenocarcinomas and squamous cell carcinomas [101,102]). Since 5p gain was one of the most commonly affected regions in CC genomes, which was validated by fluorescence *in situ* hybridization assay on a large independent cohort of tumor specimens (FIGURE 1C), we hypothesize that the increased dosage of 5p may result in deregulation of genes that may confer oncogenic properties to its host cell [9]. To identify target genes of gain, we performed supervised analyses to compare and filter the overexpressed 5p gene set between 5p gains and diploid tumors to see what extent these two platforms will facilitate the identification of target genes. This analysis (using a significance level of  $p < 0.05$  and at least twofold increased expression) identified 17 overexpressed genes associated with 5p gain (FIGURE 1B). In addition, these genes showed several-fold increased expression relative to *GAPDH* in tumors with 5p gains (>2 copies) compared with tumors showing only two copies (FIGURE 1D). Therefore, these genes represent copy number-driven target overexpressed genes, which probably provide growth advantages and/or invasion conferred by chromosome 5p gains.

This analysis identified concurrent 5p gains, with overexpression of potentially relevant genes to cellular processes associated with tumorigenesis, such as signal transduction (*OSMR*), nucleic acid binding, DNA repair, mitotic cycle (*BASP1*, *TARS*, *PAIP1*, *BRD9*, *RAD1*, *SKP2* and *POLS*), oxidative phosphorylation (*NNT*, *SDHA* and *NDUFS6*), HPV 16 E1 binding protein (*TRIP13*), ribosomal synthesis (*BXDC2*) and miRNA processing (*RNASEN*). The top overexpressed gene by this analysis was *RNASEN* (Drosha), which executes the initial step in miRNA processing by cleaving pri-miRNA to release pre-miRNA, and plays a major role in tumor progression and prognosis [26]. Muralidhar *et al.*, using a similar integrative approach, have also identified that Drosha copy number increases associated with the overexpression of this gene in CC, and Drosha overexpression was further shown to influence expression of miRNAs implicated in other cancer types [27]. Another gene of importance in cancer is *OSMR*, which has been shown to be gained and overexpressed in CC and is associated with adverse clinical outcome [28,29]. Oncostatin M is a cytokine related to the IL-6 family of cytokines, and its biological activity is mediated through the receptor complex. Upon ligand binding, oncostatin M receptor activates signaling pathways implicated in cancer, such as



**Figure 1. Integrated analyses of chromosome 5p genomic alterations in cervical cancer. (A)** Copy number alterations in  $\log_2$  ratio of chromosome 5 identified by 250K *NspI* single nucleotide polymorphism (SNP) array in normal and tumor tissue. Each vertical column represents a sample with genomic regions representing from pter (top) to qter (bottom). The blue–red scale bar (-1.0 to +1.0) at the bottom represents the copy number changes relative to the mean across the samples. **(B)** Supervised analysis of overexpressed genes identified as a consequence of gain of chromosome 5p. G-banded ideogram of chromosome 5 is shown on the left. Significantly, differentially overexpressed genes identified between tumors showed more than two copies of 5p, and tumors with two copies of 5p are shown. In the matrix, each row represents the gene expression relative to group mean and each column represents a sample. Overall, as expected, cell lines exhibited higher levels of expression differences than primary tumors (data not shown). The scale bar (-2.0 to +2.0) on the bottom represents the level of expression. **(C)** Fluorescence *in situ* hybridization identification of 5p gains in invasive cancer. Green signals represent the 5p15.2 probe and red signals represent the probe mapping to the 5q31 region used as control. **(D)** Relative expression of differentially expressed genes as a consequence of 5p gain in relation to glyceraldehyde 3-phosphate dehydrogenase in normal and tumors with and without 5p gain is shown in box plot distribution. Middle line across the box represents median value, the upper hinge represents the 75th percentile value and the lower hinge the 25th percentile. The minimum and maximum value data points are shown below and above the box, respectively. GAPDH: Glyceraldehyde-3-phosphate dehydrogenase.

STAT, PI3/AKT and angiogenic factor VEGF, and mediates inhibition of tumor growth [30]. The other genes identified in this analysis as a consequence of 5p gains possess functions related to nucleic acid binding, DNA repair and mitotic cell cycle (*BASP1*, *TARS*, *PAIP1*, *BRD9*, *RAD1*, *SKP2* and *POLS*) and nuclear genes (*NNT*, *SDHA* and *NDUFS6*) encoding mitochondrial proteins that play a role in oxidative phosphorylation. Downen *et al.* showed that upregulation of *SKP2* gene transcription relates to 5p gains in CC cell lines [31]. *SKP2* is an F-box family protein that plays a critical role in G1/S cell cycle progression and degrades CDKN1B (p27kip). However, a similar integrated gene dosage and expression

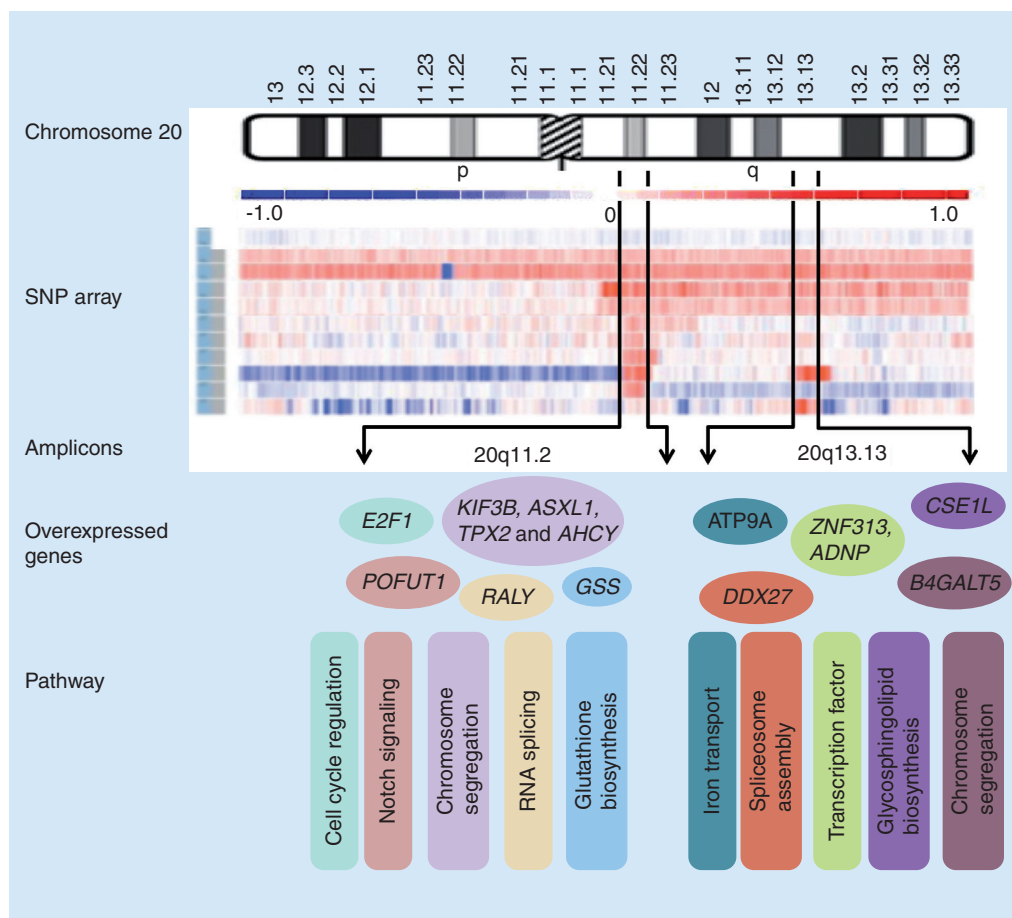
analysis by Lando and coworkers found a different set of target genes on 5p than we identified in our study [32]. Although the exact role of these overexpressed genes on 5p in CC remains unknown, their identification provides a basis for dissecting the signaling cascades involving their role individually or synergistically as oncogenes in regulating the transformation in CC.

#### Identification of target genes of 20q amplicons by integrative genomic analysis

Chromosome 20q has been reported to be one of the most common targets of NCAs in invasive CC [9,33] and arise at CIN stage 2/3 of tumor

development [24,34,35]. Chromosome 20q gains were also shown to be associated with HPV E-7-mediated immortalization of human epithelial cells [36]. These data suggest that 20q amplification is an early change in CC development, and the concurrent overexpression of specific gene(s) on this genomic region might be critical to transformation. Our copy number SNP array analysis of chromosome 20 identified two recurrent and nonoverlapping focal amplicons on 20q at 20q11.2 and 20q13.13 (FIGURE 2) [24]. The minimum shared region of amplicons at 20q11.2 spans a 4.1-Mb genomic region, and the amplicons at 20q13.13 span a 3.1-Mb physical distance (FIGURE 2). Since chromosome 20q is one of the commonly gained regions in CC genomes, we hypothesize that the amplicons located within 20q may induce transcriptional

activation of specific genes relevant to cellular transformation. Integrative genomic CNAs and expression data analysis identified eight overexpressed genes in amplicon 20q11.2 and six in amplicon 20q13.13 (FIGURE 2). The eight overexpressed genes in amplicon 20q11.2 are *GSS*, *POFUT1*, *AHCY*, *TPX2*, *ASXL1*, *E2F1*, *RALY* and *KIF3B*. These genes are functionally associated with 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 20q13.13 amplicons include nucleotide binding (*ATP9A* and *DDX27*), activity-dependent neuroprotector (*ADNP*) with a potential role in tumor proliferation,



**Figure 2. Chromosome 20q amplification target genes and their pathways identified by integrated analysis of copy number and gene expression.** A G-banded ideogram of chromosome 20 is shown at the top. Representative cases of SNP array copy numbers by 250K *Nspl* array in  $\log_2$  ratio are shown. Each horizontal column in the SNP array represents a sample with genomic region, representing pter (top) to qter (bottom). The blue–red scale bar (-1.0 to +1.0) at the top represents the copy number changes relative to mean across the samples. Downward arrows show two focal amplicons on 20q, identified by SNP array. Significantly overexpressed genes identified and the pathways the genes connected as a consequence of 20q amplifications are shown in the figure. SNP: Single nucleotide polymorphism.

a gene encoding for UDP-Gal:β-GlcNAc β-1,4-galactosyltransferase (*B4GALT5*) with transferase activity, a zinc finger protein 313 (*ZNF313*) and a nuclear function protein (*CSE1L*). The genes that we found to be upregulated as a consequence of chromosome 20q amplifications are known to play specific roles in tumorigenic processes. For example, *E2F1*, *KIF3B*, *TPX2* and *CSE1L* genes play pivotal roles in cell cycle regulation and chromosome segregation (FIGURE 2). Therefore, the genes identified by this approach provide a basis for testing their significance in relation to HPV infection, a functional role in tumor initiation and progression of CC. Recently, Lando and coworkers, using integrative analysis of gene dosage and expression, also found three of the genes (*POFUT1*, *KIF3B* and *AHCY*) that we identified to be overexpressed as targets of 20q gain [32]. However, Wilting and coworkers, in a similar approach utilizing a smaller sample size and whole-genome analysis using differential gene locus mapping and array comparative genomic hybridization expression integration tool, did not identify any of the genes we identified in our study [37]. Therefore, these studies highlight the importance of the application of appropriate algorithms of integrative genomic approaches to identify gene targets that are biologically relevant to cervical carcinogenesis.

### Conclusion

In this article, we have described an integrative genomic strategy utilizing information on recurrent CNAs at 5p, 20q11.2 and 20q13.13, with gene expression to identify genes relevant to genomic copy number gains and amplifications. Utilizing this approach, we demonstrated the robustness of this strategy in identifying genes and genetic pathways relevant to a specific region of genomic copy number increases in tumorigenesis. We conclusively show that a simple approach of systematic integrative genomic

analysis can lead to better molecular discoveries, which could then be used in identifying relevant therapeutic targets for CC.

### Future perspective

To gain insight into molecular-based therapeutic targets for patients with CC, it is essential to construct an integrated view of multi-dimensional genomic data from complementary technologies for CNA (gains, amplifications and deletions) with transcription profiles, mutations, miRNA, epigenetic markers and clinical end points. It would be reasonable to speculate such an approach – complemented by appropriate bioinformatics tools and functional data – is likely to advance our molecular knowledge to identify core pathways, leading to individualized molecular-based therapy. Advances in molecular knowledge will also change the detection and prediction of precancerous lesion, making the existing tests obsolete. Among the first hints at this approach is the integrative genomic analysis that we discussed in this article. A significant reduction in the incidence of CC has been already achieved by Pap smear screening. The availability of a prophylactic HPV vaccine is expected to further reduce this incidence. Almost certainly, the advances in molecular and bioinformatic technologies will further enhance our molecular understanding and continue to raise the hope for high cure rates of CC.

### Financial & competing interests disclosure

*Vundavalli V Murty has received funding from the NIH (grant number CA095647). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.*

### Executive summary

- Amplification of over 20 different chromosomal regions has been reported in cervical cancer (CC).
- CC genomes harbor multiple copy number gains and losses of specific chromosomal regions; some of these were shown to arise at early precancerous stages, suggesting a role for copy number alterations in its tumorigenesis.
- Mutational mechanisms are relatively uncommon in CC.
- Promoter hypermethylation and the associated downregulated gene expressions are frequent in CC.
- The chromosomal regions of 5p and 20q exhibit the most significant recurrent focal copy number alterations in CC, suggesting their role in tumor formation and progression.
- Integrative genomic analysis of 5p gains and 20q11.2 and 20q13.13 amplifications identified overexpressed genes as a consequence of genomic copy number increases. The genes identified are involved in cellular processes associated with specific pathways in tumorigenesis.



## Bibliography

Papers of special note have been highlighted as:

▪ of interest

▪▪ of considerable interest

1. Jemal A, Siegel R, Ward E *et al.*: Cancer statistics, 2009. *CA Cancer J. Clin.* 59(4), 225–249 (2009).
2. Murthy NS, Sehgal A, Satyanarayana L *et al.*: Risk factors related to biological behaviour of precancerous lesions of the uterine cervix. *Br. J. Cancer* 61(5), 732–736 (1990).
3. Schneider A, Koutsky LA: Natural history and epidemiological features of genital HPV infection. *IARC Sci. Publ.* 119, 25–52 (1992).
4. Dunne EF, Markowitz LE: Genital human papillomavirus infection. *Clin. Infect. Dis.* 43(5), 624–629 (2006).
5. Richart R, Ludwig AS Jr: Alterations in chromosomes and DNA content in gynecologic neoplasms. *Am. J. Obstet. Gynecol.* 104(4), 463–471 (1969).
6. Wakonig-Vaartaja R, Hughes DT: Chromosomal anomalies in dysplasia, carcinoma-*in-situ*, and carcinoma of cervix uteri. *Lancet* 2(7416), 756–759 (1965).
7. Santarius T, Shipley J, Brewer D *et al.*: A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer* 10(1), 59–64 (2010).
8. Narayan G, Arias-Pulido H, Koul S *et al.*: Genetic analysis identifies putative tumor suppressor sites at 2q35–q36.1 and 2q36.3–q37.1 involved in cervical cancer progression. *Oncogene* 22(22), 3489–3499 (2003).
  - **Demonstrates the novel observation of deletions at 2q35–p37 regions, identifying a site of tumor-suppressor gene occurrence early in cervical cancer development.**
9. Rao PH, Arias-Pulido H, Lu XY *et al.*: Chromosomal amplifications, 3q gain and deletions of 2q33–q37 are the frequent genetic changes in cervical carcinoma. *BMC Cancer* 4(1), 5 (2004).
  - **Demonstrates, for the first time, that gene amplification is a frequent phenomenon in invasive cervical cancer.**
10. Heselmeyer K, Schrock E, du Manoir S *et al.*: Gain of chromosome 3q defines the transition from severe dysplasia to invasive carcinoma of the uterine cervix. *Proc. Natl Acad. Sci. USA* 93(1), 479–484 (1996).
  - **Provides the first demonstration that gain of a specific chromosomal region occurs in cervical precancerous lesions.**
11. Heselmeyer K, Macville M, Schrock E *et al.*: Advanced-stage cervical carcinomas are defined by a recurrent pattern of chromosomal aberrations revealing high genetic instability and a consistent gain of chromosome arm 3q. *Genes Chromosomes Cancer* 19(4), 233–240 (1997).
12. Scotto L, Narayan G, Nandula SV *et al.*: Integrative genomics analysis of chromosome 5p gain in cervical cancer reveals target over-expressed genes, including Drosha. *Mol. Cancer* 7, 58 (2008).
  - **Identifies crucial gene-expression signatures as a consequence of 5p gain in cervical cancer.**
13. Dong SM, Kim HS, Rha SH, Sidransky D: Promoter hypermethylation of multiple genes in carcinoma of the uterine cervix. *Clin. Cancer Res.* 7(7), 1982–1986 (2001).
14. Narayan G, Scotto L, Neelakantan V *et al.*: Protocadherin PCDH10, involved in tumor progression, is a frequent and early target of promoter hypermethylation in cervical cancer. *Genes Chromosomes Cancer* 48(11), 983–992 (2009).
15. Narayan G, Arias-Pulido H, Koul S *et al.*: Frequent promoter methylation of *CDHI*, *DAPK*, *RARB*, and *HIC1* genes in carcinoma of cervix uteri: its relationship to clinical outcome. *Mol. Cancer* 2, 24 (2003).
16. Virmani AK, Muller C, Rathi A *et al.*: Aberrant methylation during cervical carcinogenesis. *Clin. Cancer Res.* 7(3), 584–589 (2001).
17. Hidalgo A, Baudis M, Petersen I *et al.*: Microarray comparative genomic hybridization detection of chromosomal imbalances in uterine cervix carcinoma. *BMC Cancer* 5, 77 (2005).
18. Wilting SM, Smeets SJ, Snijders PJ *et al.*: Genomic profiling identifies common HPV-associated chromosomal alterations in squamous cell carcinomas of cervix and head and neck. *BMC Med. Genomics* 2, 32 (2009).
19. Kloth JN, Oosting J, van Wezel T *et al.*: Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer. *BMC Genomics* 8, 53 (2007).
20. Lockwood WW, Coe BP, Williams AC, MacAulay C, Lam WL: Whole genome tiling path array CGH analysis of segmental copy number alterations in cervical cancer cell lines. *Int. J. Cancer* 120(2), 436–443 (2007).
21. Fitzpatrick MA, Funk MC, Gius D *et al.*: Identification of chromosomal alterations important in the development of cervical intraepithelial neoplasia and invasive carcinoma using alignment of DNA microarray data. *Gynecol. Oncol.* 103(2), 458–462 (2006).
22. Boehm JS, Zhao JJ, Yao J *et al.*: Integrative genomic approaches identify *IKBKE* as a breast cancer oncogene. *Cell* 129(6), 1065–1079 (2007).
23. The TCGA Research Network: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455(7216), 1061–1068 (2008).
24. Scotto L, Narayan G, Nandula SV *et al.*: 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 Chromosomes Cancer* 47(9), 755–765 (2008).
  - **Identifies overexpression of genes involved in specific pathways as a consequence of amplifications on chromosome 20q.**
25. Chaganti RS, Houldsworth J: Genetics and biology of adult human male germ cell tumors. *Cancer Res.* 60(6), 1475–1482 (2000).
26. Sugito N, Ishiguro H, Kuwabara Y *et al.*: *RNASEN* regulates cell proliferation and affects survival in esophageal cancer patients. *Clin. Cancer Res.* 12(24), 7322–7328 (2006).
27. Muralidhar B, Goldstein LD, Ng G *et al.*: Global microRNA profiles in cervical squamous cell carcinoma depend on Drosha expression levels. *J. Pathol.* 212(4), 368–377 (2007).
28. Narayan 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 Cancer* 46(4), 373–384 (2007).
  - **Demonstrates the utility of the whole-genome approach in identifying recurrent copy number alterations in cervical cancer.**
29. Ng G, Winder D, Muralidhar B *et al.*: Gain and overexpression of the oncostatin M receptor occur frequently in cervical squamous cell carcinoma and are associated with adverse clinical outcome. *J. Pathol.* 212(3), 325–334 (2007).
30. Liu J, Hadjokas N, Mosley B *et al.*: Oncostatin M-specific receptor expression and function in regulating cell proliferation of normal and malignant mammary epithelial cells. *Cytokine* 10(4), 295–302 (1998).
31. Downen SE, Neutze DM, Pett MR *et al.*: Amplification of chromosome 5p correlates with increased expression of *Skp2* in HPV-immortalized keratinocytes. *Oncogene* 22(16), 2531–2540 (2003).

32. Lando M, Holden M, Bergersen LC *et al.*: Gene dosage, expression, and ontology analysis identifies driver genes in the carcinogenesis and chemoradioresistance of cervical cancer. *PLoS Genet.* 5(11), E1000719 (2009).
33. Wilting SM, Snijders PJ, Meijer GA *et al.*: Increased gene copy numbers at chromosome 20q are frequent in both squamous cell carcinomas and adenocarcinomas of the cervix. *J. Pathol.* 209(2), 220–230 (2006).
34. Wilting SM, Steenbergen RD, Tijssen M *et al.*: Chromosomal signatures of a subset of high-grade premalignant cervical lesions closely resemble invasive carcinomas. *Cancer Res.* 69(2), 647–655 (2009).
35. Alazawi W, Pett M, Strauss S *et al.*: Genomic imbalances in 70 snap-frozen cervical squamous intraepithelial lesions: associations with lesion grade, state of the HPV16 E2 gene and clinical outcome. *Br. J. Cancer* 91(12), 2063–2070 (2004).
36. Klingelhutz AJ, Qian Q, Phillips SL *et al.*: Amplification of the chromosome 20q region is associated with expression of HPV-16 E7 in human airway and anogenital epithelial cells. *Virology* 340(2), 237–244 (2005).
37. Wilting SM, de Wilde J, Meijer CJ *et al.*: Integrated genomic and transcriptional profiling identifies chromosomal loci with altered gene expression in cervical cancer. *Genes Chromosomes Cancer* 47(10), 890–905 (2008).

### Websites

101. Progenetix  
[www.progenetix.com/progenetix/G\\_Cacervixca](http://www.progenetix.com/progenetix/G_Cacervixca)
102. National Cancer Institute. Recurrent chromosome aberrations in cancer  
<http://cgap.nci.nih.gov/Chromosomes/RecurrentAberrations>