Genome wide copy number analyses of superficial esophageal squamous cell carcinoma with and without metastasis

Pengjiao Wang1,*, Ling Shan1,*, Liyan Xue1, Bo Zheng1, Jianming Ying1, Ning Lu1

1Department of Pathology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, P. R. China
*These authors have contributed equally to this work

Correspondence to: Ning Lu, email: nlu03@126.com
Jianming Ying, email: jmying@hotmail.com

Keywords: superficial esophageal squamous cell carcinoma, copy number alterations, metastasis, microarray

ABSTRACT

Superficial esophageal squamous cell carcinoma (ESCC) is generally considered a subtype of less invasive ESCC. Yet a subset of these superficial ESCC would have metastasis after esophagostomy or endoscopic resection and lead to poor prognosis. The objective of this study is to determine biomarkers that can identify such subset of superficial ESCC that would have metastasis after surgery using genome wide copy number alteration (CNA) analyses. The CNAs of 38 cases of superficial ESCCs originated from radical surgery, including 19 without metastasis and 19 with metastasis within 5 years’ post-surgery, were analyzed using Affymetrix OncoScan™ FFPE Assay. A 39-gene signature was identified which characterized the subset of superficial ESCC with high risk of metastasis after surgery. In addition, recurrent CNAs of superficial ESCC were also investigated in the study. Amplification of 11q13.3 (FGF4) and deletion of 9p21.3 (CDKN2A) were found to be recurrent in all 38 superficial ESCCs analyzed. Notably amplifications of 3p26.33 (SOX2OT), 8q24.21 (MYC), 14q21.1 (FOXA1) and deletion of 3p12.1 (GBE1) were only found to be recurrent in metastatic superficial ESCCs. In conclusion, using CNAs analyses, we identify a 39-gene signature which characterizes the high risk metastatic superficial ESCCs and discover several recurrent CNAs that might be the driver alterations in metastasis among superficial ESCCs.

INTRODUCTION

Esophageal squamous cell carcinoma (ESCC) is one of the most deadly tumors worldwide, with 5 year survival of only 10%. ESCC related death rates are particularly high in China, in part due to the lack of early screening tools and limited treatment options [1].

Superficial ESCC has significant better prognosis than the advanced ESCC, and can be treated effectively by endoscopic resection, including both mucosal resection and submucosal dissection [2, 3]. Endoscopic resection remains the most widely used treatment option for superficial ESCCs because it is safe and esophagus-preserving. However, 26-53% of superficial ESCCs have lymph node metastasis, and require additional esophagectomy or radiochemotherapy after endoscopic resection [4, 5]. Thus identifying biomarkers that can assess the metastatic risk in superficial ESCC is of particular importance.

Several clinicopathological features have been associated with high metastatic risk in superficial ESCC, including invasion depth, tumor size, angiolymphatic invasion and histological differentiation [4–6]. ESCCs are also characterized by CNAs [7, 8]. Among those characteristic CNAs, amplifications of 11q13.3-13.4 (CCND1), 3q26.33 (SOX2), 8q24.2 (MYC) and deletion of 3p12.1 (GBE1) have been associated with metastatic diseases [8–12]. Nevertheless, a comprehensive set of CNAs that can be used to predict metastasis risk in clinical samples are not yet available and would be important to select the appropriate treatment option for superficial ESCCs.

Genomic instabilities, in the form of chromosome instability (CIN), microsatellite instability (MIN) and point mutations, are characteristic for human cancers. CIN, including chromosome structural and number changes, is a major form of genomic instability [13]. According to their size, CNAs can be classified into focal or arm-
level CNAs. Focal CNAs are very informative and often involve important oncogenes or tumor suppressor genes [14–17]. CNAs could be detected by next-generation sequencing, comparative array genomic hybridization or single nucleotide polymorphism (SNP) microarray. The OncoScan® assay is a SNP microarray including over 220,000 SNPs across human genome with increased probe density within 891 cancer related genes and are uniquely suited to detect subtle CNAs with high sensitivity and specificity.

Here we reported the analyses of 38 superficial ESCCs originated from radical surgery with both metastatic and metastasis-free samples with Affymetrix OncoScan™ array. The CNAs landscape of superficial ESCC were determined, and the focal recurrent CNAs were compared between the metastasis and metastasis-free cases.

RESULTS

Genome wide CNA of superficial ESCCs

The whole genome CNA profiles of all 38 samples were shown in Supplementary Figure S1. The microarray data has been deposited in the Gene Expression Omnibus (GEO) with the accession number GSE78926. Weighted GII was plotted to show the genome wide CNAs. Weighted GII was 45.5% in all superficial ESCC samples, and was 49.9% and 41.1% in metastasis group and non-metastasis group, respectively. The comparison of weighted GII in metastasis and non-metastasis group was shown in Figure 1A (p = 0.07). In addition, the comparison of GII on each chromosome in non-metastasis and metastasis groups was also performed. In non-metastasis group, chromosomes 3 (65.7%), 8 (78.2%) and 20 (59.8%) have the highest GII, whereas chromosomes 4 (29.5%), 12 (28.7%) and 21(25.2%) have lowest GII. In metastasis group, chromosomes 3 (75.2%), 8 (72.1%) and 14 (60.0%) have the highest GII, whereas chromosomes 15 (35.9%), 21 (31.8%) and 22 (38.1%) have the lowest GII. Chromosome 18 has higher GII in metastasis group than in the non-metastasis group (p = 0.03) (Figure 1B).

The average number of genome segments with CNAs in all 38 samples was 218, and the average length was 13.26 Mb (minimal length was 13.14 Kb and maximal length was 242.64 Mb). In metastasis group, the average number of segments with CNAs was 217, and the average length was 13.35 Mb (minimal length was 13.16 Kb and maximal length was 239.80 Mb). In non-metastasis group, the average number of segments with CNAs was 210, and the average length was 13.05 Mb (minimal length was 13.14 Kb and maximal length was 242.64 Mb). The distribution of size of CNAs length was shown in Figure 1C and 1D.

Figure 1: The whole genome copy number alteration (CNA) profiles of all 38 samples were shown. A. Comparison of weighted genome instability index (GII) of all chromosomes between metastasis and non-metastasis groups (p = 0.07). B. Comparison of GII of Chromosome 18 between metastasis and non-metastasis groups (p = 0.03). C. Length distribution of all the CNAs in metastasis and non-metastasis groups. D. Length distribution of CNAs (< 2 Mb) in metastasis and non-metastasis groups.
Recurrent CNAs in metastasis and non-metastasis groups

We found 28 significantly recurrent focal CNAs, including 16 amplifications and 12 deletions in all 38 superficial ESCCs (Supplementary Table S1) (Figure 2). The most common CNAs were amplifications of 11q13.3 (FGF4), 3q28 (TP63), 14q21.1 (FOXA1), 8q24.21 (MYC), and deletions of 9p21.3 (CDKN2A), 22q11.23 (GSTM1), 3p13 (MITF) and 2q22.1 (LRP1B). In metastasis group, eight recurrent focal CNAs were found, including 4 amplifications and 4 deletions (Supplementary Table S2). The most common CNAs were amplifications of 11q13.3 (FGF4), 14q21.1 (FOXA1), 3q26.33 (SOX2-OT), 8q24.21 (MYC) and deletions of 22q11.23 (GSTM1), 9p21.3 (CDKN2A). In non-metastasis group, fourteen recurrent focal CNAs were found, including 12 amplifications and 2 deletions (Supplementary Table S3). The most common CNAs were amplifications of 11q13.3 (FGF4), 2q33.1 (PLCL1), 3q28 (TP63) and deletions of 22q11.23 (GSTM1) and 9p21.3 (CDKN2A). In conclusion, amplifications of 11q13.3 (FGF4), 8q24.21 (MYC) and deletions of 9p21.3 (CDKN2A) and 2q22.1 (LRP1B), which previously reported in ESCCs were also found in all 38 superficial ESCCs in our study. Furthermore, deletions of 3p13 (MITF) and 22q11.23 (GSTM1) were also found as CNAs targets of ESCC but were not previously reported. Amplification of 3q26.33 (SOX2-OT) was the most significantly recurrent CNA in metastasis group whereas amplification of 3q28 (TP63) was the most significantly recurrent CNA in non-metastasis group.

Comparison of CNAs between metastasis and non-metastasis groups of 891 cancer genes

Given Oncoscan microarray has increased probe density within 891 cancer genes, these 891 cancer genes were analyzed separately. Among these 891 genes, 39 genes had significantly different CNAs between metastasis and non-metastasis groups (Mann-Whitney U-test,
Hierarchical clustering was performed on the CNA profiles of these 39 genes and two groups were obtained, including metastasis and non-metastasis groups. In general, good separated performance of metastasis and non-metastasis was obtained with a small amount of misclassification (two metastasis misclassified and five non-metastasis misclassified) (Figure 3).

The genomic distribution and copy numbers of these 39 genes was shown in Figure 4 and Supplementary Figure S1. The comparison of the 39 genes was shown in Table 1. Among the 39 genes, average copy number of FGF4 in all cases (11q13.3, Mann-Whitney U-test, \( p = 0.045 \)) were 5.14. Amplification of FGF4 happened in 68% cases in metastasis group, and 58% in non-metastasis group. FGF4 gene was located on chromosome 11q13.3, which were found to be amplified in ESCCs from previous studies. Average copy number of MRAS gene (3q22.3, Mann-Whitney U-test, \( p = 0.029 \)) in all cases was 2.61, and 68% cases had amplification of MRAS gene in metastasis and 84% cases had amplifications in non-metastasis group. Average copy number of ALK (2p21, Mann-Whitney U-test, \( p = 0.024 \)) in all cases was 2.25. Amplification of ALK gene happened in 26% of non-metastasis cases, and 63% of metastasis cases. Average copy number of CHEK1 (11q24.2, Mann-Whitney U-test, \( p = 0.017 \)) in all cases was 1.91. In non-metastasis group 3 cases (16%) had amplification of CHEK1 and 3 cases (16%) had deletion of CHEK1, and in metastasis group only one case had amplification and 10 cases (53%) had deletion of CHEK1. The most significantly enriched pathway/functional terms were “Proto-oncogene” (\( p = 7.59\times10^{-11} \), “Chromosomal rearrangement” (\( p = 8.64\times10^{-9} \), “Pathways in cancer” (\( p = 6.66\times10^{-4} \)) and “Nucleotide-binding” (\( p = 0.015 \)).

**DISCUSSION**

The main purpose of this study is to identify CNAs that could differentiate superficial ESCC patients with high risk of metastasis. Although several whole genome copy number analyses on ESCC have been reported, there has been no systematic study on superficial ESCC. In this study, we used Affymetrix OncoScan™ (SNP microarray) to analyze genome wide CNAs of 38 superficial ESCCs and compared recurrent focal CNAs level, and the CNAs of 891 cancer genes between ESCCs with and without metastasis.

The Oncoscan SNP microarray is an ideal platform to perform analyses on FFPE-derived tumor materials with increased probe density within 891 cancer genes at 50-100 kb resolution. Among the 891 cancer genes, 39 cancer genes had significantly different CNAs between the metastasis and non-metastasis cases. These genes might represent candidate biomarkers for superficial ESCC patients with high risk of metastasis after surgery. Previous studies focus on using gene expression profiles to guide prognosis prediction and diagnosis [18, 19]. Here we suggested that DNA copy number may be used for the same purpose with the advantage of being more stable [20]. FGF4, PIK3CB, MRAS, ALK, LMO2, AKT3 and CHEK1 genes were included in the 39-gene signature. Copy number of CHEK1 gene was significantly lower in metastasis than
in non-metastasis groups ($p = 0.017$). Checkpoint kinase 1 (Chk1) which encoded by CHEK1 gene trigger cell cycle arrest upon DNA damage [21] Although CHEK1 was initially thought to be a tumor suppressor gene due to its role in checkpoint activation and cell cycle arrest, several studies documented a positive correlation between Chk1 expression and tumor grade and disease recurrence, suggesting that Chk1 may promote tumor growth [22–24]. So the role of CHEK1 in ESCC should be explored further. In addition, FGF4 gene located in 11q13.3 was also recurrently amplified in superficial ESCC.

In discovery and validation experiments, there were 88.5% (23/26) cases showed the identical copy number aberrations of CCNL1 gene, and 76% (19/25) cases for PIK3CB gene. The difference of calculation of copy number alterations in these two experiments might lead to inconsistency of results.

The recurrent focal CNAs which determined by GISTIC2.0 were more likely to identify cancer-causing genes [14, 25]. Among all 38 superficial ESCC samples, recurrent focal amplifications of 11q13.3 (FGF4 and FGF19), 8q24.21 (MYC), 7q22.1 (CDK6) and deletions of 9p21.3 (CDKN2A), 3p12.3 (MITF), 2q22.1 (LRP1B), 13q14.2 (RB1) which were previously reported in all stage of ESCC [8, 11, 12, 26–30] were also found in our study. This indicated that those CNAs might be the driver events in early stage of ESCC. In addition, amplification of 11q21.1 (FOXA1), which has not been previously reported as associated with ESCCs, was found in 20 cases (52.6%) in our study. Forkhead box protein A1 (FOXA1), the representative member of the Forkhead-box (FOX) proteins subfamily, is a DNA-binding transcription factor. Amplification of FOXA1 gene has been found in lung cancer, esophageal adenocarcinoma, [31] estrogen receptor

Figure 4: The genomic distribution and copy number of 39 genes identified by Mann-Whitney U-test in 38 esophageal squamous cell carcinoma genomes. A. The red lines marked on chromosomes represented the position of the 39 genes on genome. B. The copy numbers of the part of the 39 genes in metastasis and non-metastasis groups were shown. The copy numbers of other part of the 39 genes were shown in Supplementary Figure S2.
Table 1: The significantly different copy numbers of thirty nine genes in metastasis and non-metastasis groups

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Position</th>
<th>Average CN</th>
<th>P value</th>
<th>Non-metastasis</th>
<th>Metastasis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average CN</td>
<td>Amplification</td>
<td>Deletion</td>
<td>Average CN</td>
</tr>
<tr>
<td>TPM3</td>
<td>Chr1:154117779-154165725</td>
<td>2.32</td>
<td>0.021</td>
<td>2.16</td>
<td>8 (0.42)</td>
</tr>
<tr>
<td>MUC1</td>
<td>Chr1:155148299-155172706</td>
<td>2.28</td>
<td>0.023</td>
<td>2.14</td>
<td>8 (0.42)</td>
</tr>
<tr>
<td>RAB25</td>
<td>Chr1:156020966-156050295</td>
<td>2.31</td>
<td>0.029</td>
<td>2.16</td>
<td>8 (0.42)</td>
</tr>
<tr>
<td>SDHC</td>
<td>Chr1:161274165-161344535</td>
<td>2.35</td>
<td>0.021</td>
<td>2.17</td>
<td>8 (0.42)</td>
</tr>
<tr>
<td>FCGR2B</td>
<td>Chr1:161622904-161658444</td>
<td>2.34</td>
<td>0.022</td>
<td>2.17</td>
<td>8 (0.42)</td>
</tr>
<tr>
<td>PBX1</td>
<td>Chr1:164518596-164831060</td>
<td>2.32</td>
<td>0.049</td>
<td>2.17</td>
<td>8 (0.42)</td>
</tr>
<tr>
<td>PRRX1</td>
<td>Chr1:170623312-170718541</td>
<td>2.30</td>
<td>0.023</td>
<td>2.14</td>
<td>7 (0.37)</td>
</tr>
<tr>
<td>TNN</td>
<td>Chr1:175026993-175127202</td>
<td>2.30</td>
<td>0.023</td>
<td>2.14</td>
<td>7 (0.37)</td>
</tr>
<tr>
<td>RFWD2</td>
<td>Chr1:175903966-176186370</td>
<td>2.28</td>
<td>0.025</td>
<td>2.12</td>
<td>7 (0.37)</td>
</tr>
<tr>
<td>ABL2</td>
<td>Chr1:179058461-179122224</td>
<td>2.30</td>
<td>0.023</td>
<td>2.14</td>
<td>7 (0.37)</td>
</tr>
<tr>
<td>RNASEL</td>
<td>Chr1:182532768-182568394</td>
<td>2.29</td>
<td>0.023</td>
<td>2.14</td>
<td>7 (0.37)</td>
</tr>
<tr>
<td>LAMC2</td>
<td>Chr1:183145173-183220406</td>
<td>2.29</td>
<td>0.023</td>
<td>2.14</td>
<td>7 (0.37)</td>
</tr>
<tr>
<td>TPR</td>
<td>Chr1:186270785-186354457</td>
<td>2.28</td>
<td>0.045</td>
<td>2.14</td>
<td>7 (0.37)</td>
</tr>
<tr>
<td>PTGS2</td>
<td>Chr1:186630943-186659559</td>
<td>2.30</td>
<td>0.034</td>
<td>2.14</td>
<td>7 (0.37)</td>
</tr>
<tr>
<td>TGFB2</td>
<td>Chr1:218508675-218627961</td>
<td>2.28</td>
<td>0.045</td>
<td>2.14</td>
<td>7 (0.37)</td>
</tr>
<tr>
<td>CDC42BPA</td>
<td>Chr1:227167565-227515826</td>
<td>2.30</td>
<td>0.043</td>
<td>2.16</td>
<td>6 (0.32)</td>
</tr>
<tr>
<td>OBSCN</td>
<td>Chr1:228385860-228558951</td>
<td>2.30</td>
<td>0.043</td>
<td>2.16</td>
<td>6 (0.32)</td>
</tr>
<tr>
<td>RHOU</td>
<td>Chr1:228770393-228892416</td>
<td>2.30</td>
<td>0.043</td>
<td>2.16</td>
<td>6 (0.32)</td>
</tr>
<tr>
<td>AKT3</td>
<td>Chr1:243641534-244016584</td>
<td>2.26</td>
<td>0.047</td>
<td>2.14</td>
<td>6 (0.32)</td>
</tr>
<tr>
<td>NCOA1</td>
<td>Chr2:24797345-25003570</td>
<td>2.23</td>
<td>0.036</td>
<td>2.10</td>
<td>5 (0.26)</td>
</tr>
<tr>
<td>DNMT3A</td>
<td>Chr2:25445829-25485184</td>
<td>2.23</td>
<td>0.036</td>
<td>2.10</td>
<td>5 (0.26)</td>
</tr>
</tbody>
</table>

(Continued)
(ER)-positive breast cancer [32] and anaplastic thyroid cancer [33]. And the levels of FOXA1 protein have been correlated with the prognosis of breast cancer and gastric cancer [34, 35]. Deletions of 22q11.23 (GSTT1, GSTTP1, and GSTT2) occurred in 20 superficial ESCC samples (57.9%) in our study. Glutathione S-transferase (GST) theta 1 (GSTT1) which encoded by GSTT1 gene is a member metabolizing dimeric phase II enzymes superfamily. These enzymes play a vital role in cellular defense system by catabolism of a broad range of xenobiotics and carcinogen [36, 37]. Many previous studies have found that GSTT1 gene deletion polymorphisms increases susceptibility to

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Position</th>
<th>Average CN</th>
<th>P value</th>
<th>Non-metastasis</th>
<th>Metastasis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Average CN</td>
<td>Amplification</td>
</tr>
<tr>
<td>ALK</td>
<td>Chr2:29405639-30154477</td>
<td>2.25</td>
<td>0.024</td>
<td>2.12</td>
<td>5 (0.26)</td>
</tr>
<tr>
<td>MRAS</td>
<td>Chr3:138056489-138134377</td>
<td>2.61</td>
<td>0.029</td>
<td>2.37</td>
<td>13 (0.68)</td>
</tr>
<tr>
<td>PIK3CB</td>
<td>Chr3:138361539-138436463</td>
<td>2.62</td>
<td>0.029</td>
<td>2.37</td>
<td>13 (0.68)</td>
</tr>
<tr>
<td>FOXL2</td>
<td>Chr3:138653065-138675982</td>
<td>2.62</td>
<td>0.043</td>
<td>2.39</td>
<td>13 (0.68)</td>
</tr>
<tr>
<td>GMPS</td>
<td>Chr3:155578324-155665520</td>
<td>2.74</td>
<td>0.042</td>
<td>2.51</td>
<td>15 (0.79)</td>
</tr>
<tr>
<td>CCNL1</td>
<td>Chr3:156855585-156888482</td>
<td>2.75</td>
<td>0.032</td>
<td>2.51</td>
<td>15 (0.79)</td>
</tr>
<tr>
<td>LMO2</td>
<td>Chr11:33870122-33901371</td>
<td>2.15</td>
<td>0.031</td>
<td>2.05</td>
<td>2 (0.11)</td>
</tr>
<tr>
<td>FGF4</td>
<td>Chr11:69577796-69600171</td>
<td>5.14</td>
<td>0.045</td>
<td>3.15</td>
<td>11 (0.58)</td>
</tr>
<tr>
<td>ZBTB16</td>
<td>Chr11:113920430-114131397</td>
<td>1.90</td>
<td>0.049</td>
<td>1.97</td>
<td>2 (0.11)</td>
</tr>
<tr>
<td>PAFAH1B2</td>
<td>Chr11:117004999-117051761</td>
<td>1.90</td>
<td>0.049</td>
<td>1.97</td>
<td>2 (0.11)</td>
</tr>
<tr>
<td>PCSK7</td>
<td>Chr11:117065787-11712811</td>
<td>1.90</td>
<td>0.049</td>
<td>1.97</td>
<td>2 (0.11)</td>
</tr>
<tr>
<td>H2AFX</td>
<td>Chr11:118954584-118976177</td>
<td>1.89</td>
<td>0.041</td>
<td>1.95</td>
<td>2 (0.11)</td>
</tr>
<tr>
<td>CHEK1</td>
<td>Chr11:125485030-125537042</td>
<td>1.91</td>
<td>0.017</td>
<td>2.00</td>
<td>3 (0.16)</td>
</tr>
<tr>
<td>CCNB1P1</td>
<td>Chr14:20769528-20807533</td>
<td>2.36</td>
<td>0.039</td>
<td>2.14</td>
<td>6 (0.32)</td>
</tr>
<tr>
<td>TRA@</td>
<td>Chr14:22080057-23031075</td>
<td>2.29</td>
<td>0.040</td>
<td>2.13</td>
<td>6 (0.32)</td>
</tr>
<tr>
<td>CTDP1</td>
<td>Chr18:77429800-77524510</td>
<td>1.91</td>
<td>0.041</td>
<td>1.98</td>
<td>2 (0.11)</td>
</tr>
<tr>
<td>OLG2</td>
<td>Chr21:34388215-34411503</td>
<td>1.94</td>
<td>0.036</td>
<td>2.04</td>
<td>4 (0.21)</td>
</tr>
<tr>
<td>ITSN1</td>
<td>Chr21:35004783-35220802</td>
<td>1.94</td>
<td>0.030</td>
<td>2.03</td>
<td>4 (0.21)</td>
</tr>
</tbody>
</table>
lymphoma, breast cancer, colon cancer and lung cancer [38–40]. So the results of our study suggested that the role of GSTT1 deletion in the etiology of ESCC should be explored further.

Amplifications of 3p26.33 (SOX2-OT), 8q24.21 (MYC), 14q21.1 (FOXA1) and deletion of 3p12.1 (GBE1) were found as the recurrent CNAs in metastasis group only, which indicated that those CNAs might be associated with metastasis in superficial ESCC. Deletion of 3p12.1 (GBE1) has been associated with prognosis in cervical cancer [41], but not in ESCC previously. Sex determining region Y-box 2 (SOX2), a key transcription factor involved in self-renewal and pluripotency of embryonic stem cells, plays an important role in tumor cell metastasis and apoptosis [42]. SOX2 gene embedded in the introns of SOX2 overlapping transcript (SOX2OT) gene, which encodes a long non-coding RNA (lncRNA) [43]. Amplifications of SOX2 have been associated with metastasis or poor prognosis in previous studies of ESCC [9, 44–46]. In our study, amplification of SOX2OT gene on 3p26.33 was found as recurrent CNA in metastasis group. SOX2OT gene encodes the lncRNA which have been demonstrated that involved in regulation of SOX2 expression and/or other related processes. And expression of SOX2 and SOX2OT were concordant in ESCC and breast cancer [47, 48]. The role of SOX2OT gene and expression in ESCC metastasis should be explored further.

In conclusion, our study constructed a 39-gene signature associated with metastasis from superficial ESCCs. The comparison of recurrent CNAs between superficial ESCCs with and without metastasis also revealed a subset of metastasis specific events. A larger set of independent samples are warranted to validate and refine this 39-gene signature from our study.

**MATERIALS AND METHODS**

**Patients and genomic DNA extraction**

Thirty eight superficial ESCC patients, at stage T1N0M0, were collected from Cancer Hospital, Chinese Academy of Medical Sciences between 2004 and 2010. The most recent follow-up visit was dated in June 2015 and the follow was conducted via telephone interview or clinical data consultation. Among the 38 cases, 19 metastasis cases have lymph node metastasis or distant metastasis within five years from radical surgery, and the
19 metastasis-free cases have no detectably metastasis within five years after radical surgery. All the patients underwent radical resection without radiotherapy or chemotherapy prior to surgery. The clinicopathological characters of 38 patients were summarized in Table 2. The Institute Review Board of the Cancer Hospital, CAMS, agreed to waive the need for consent for this study and approved the study protocol.

Genomic DNA were extracted from formalin-fixed and paraffin-embedded (FFPE) tissues of 38 superficial ESCCs using QIAamp DNA Mini kit (Qiagen, Hilgen, Germany) according to the manufacturer’s instruction. H&E sections of all cases were reviewed to manually identify areas with minimum of 85% malignant cells for microdissection.

**Single nucleotide polymorphism (SNP) array**

Molecular Inversion Probe (MIP) based Oncoscan Array was used to detect CNAs, loss of heterozygosity, and somatic point mutations. The experiments were performed according to the user guide of Affymetrix OncoScan™ FFPE Assay Kit (Affymetrix, CA). Briefly, 75 ng FFPE DNA were hybridized to MIP probes and allowed to anneal 58°C overnight (16-18h) after denaturation at 95°C for 5 min. Then each sample was split into two tubes and gap fill reaction was performed by adding dATP (A) and dTTP (T) (A/T) in one tube and dGTP (G) and dCTP (C) (G/C) to another. After removing the un-circularized MIP probes through exonuclease treatment, the cleavage enzyme was added to linearize the gap-filled circular MIP Probes. Then the circular MIP probes were amplified by first and second round PCR. The enriched product was digested by HaeIII enzyme and the 44bp fragments were hybridized to the OncoScan™ Array for 16-18h. The hybridized arrays were washed, stained using the GeneChip® Fluidics Station 450 and scanned through GeneChip® Scanner 3000 7G (Affymetrix, CA).

**Real time PCR (qPCR)**

Real Time PCR (qPCR) was used to validate CNAs of CCNL1 and PIK3CB genes with the internal control HBB gene using SYBR-Green II fluorescence and Mx3005P System (Agilent Technologies, CA, USA) (Supplementary Table S4). The results were analysed using the MxPro QPCR software. Comparative CT method was used to calculate the copy numbers of target genes. Cases with $2^{ΔΔCT} > 1$ were considered as amplification of genes, and cases with $2^{ΔΔCT} < 1$ were considered as deletion of genes [49].

**Data and statistical analysis**

The intensity (CEL) files generated by the scanner were imported into Oncoscan Console Software (Biodiscovery, Inc., CA USA) and analyzed by the Affymetrix TuScan algorithm (a modified ASCAT algorithm) to create segmentation to differentiate between adjacent clusters of probes and determines the CNAs. Amplification was defined as copy number which was calculated using Affymetrix TuScan algorithm > 2, and copy number < 2 was considered to be deletion. In addition, the copy number information of 891 cancer genes were obtained as Gene Report text files from the Console software.

The recurrent CNAs were determined using GISTIC2.0 (Genomic Identification of Significant Targets in Cancer) with a Q-value cutoff < 0.25. And the significant recurrent focal CNAs has a 90% likelihood of containing the targeted genes.

The Mann-Whitney U-test was used to detect the significantly different CNAs of 891 cancer genes between metastasis group and non-metastasis group. And then, hierarchical clustering was performed on significantly differential CNAs of cancer genes. Pheatmap function in R pheatmap package was used to draw heat maps to visualize the clustering results.

**ACKNOWLEDGEMENTS**

All authors would like to thank Ying Yang for help in technique support, and Peng Zhang and Xuexin Yu for help in data analysis.

**CONFLICTS OF INTEREST**

Professional judgment concerning a primary interest would not be influenced by a secondary interest. There is no financial/personal interest or belief that could affect objectivity.
REFERENCES


are Coupregulated with SOX2 and OCT4 in Esophageal Squamous Cell Carcinoma. Stem Cells. 2014; 32: 126-34.