Intratumor heterogeneity of HMCN1 mutant alleles associated with poor prognosis in patients with breast cancer

Chie Kikutake1, Minako Yoshihara1,2, Tetsuya Sato1,2, Daisuke Saito1,2 and Mikita Suyama1,2

1Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan
2AMED-CREST, Japan Agency for Medical Research and Development, Fukuoka 812-8582, Japan

Correspondence to: Mikita Suyama, email: mikita@bioreg.kyushu-u.ac.jp

Keywords: breast cancer; variant allele frequency; lymph node metastasis; genetic variant; next-generation sequencing

Received: November 21, 2017   Accepted: August 15, 2018   Published: September 07, 2018

ABSTRACT

Human breast cancers comprise a complex and highly heterogeneous population of tumor cells. Intratumor heterogeneity is an underlying cause of resistance to effective therapies and disease recurrence. To explore prognostic factors based on intratumor heterogeneity, we analyzed genomic mutations in breast cancer patients registered in The Cancer Genome Atlas. We calculated the variant allele frequency (VAF) at each mutation site and evaluated the associations of VAFs with the prognosis of breast cancer. VAFs of HMCN1 correlated with the prognosis and lymph node status. Although the detailed function of HMCN1 remains unknown, it is located in extracellular matrix and the mutation in the gene might be associated with cancer cell invasion and metastasis. This finding suggests that HMCN1 is a potential metastatic factor and can be a candidate gene for targeted breast cancer therapy.

INTRODUCTION

Breast cancer is the most common type of cancer affecting women worldwide. In 2012, approximately 1.7 million cases of breast cancer were newly diagnosed [1]. Changes in dietary habits and a reduced birth rate can increase the risk of breast cancer. Breast cancer is a clinically heterogeneous disease for which four basic therapeutic or molecular subtypes have been classified based on the expression status of three receptors: estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 [2, 3]. Immunohistochemistry is used to classify these four tumor subtypes and ensure that effective treatment is provided to each patient.

Despite recent therapeutic advances, tumor recurrence and drug resistance remain major challenges in the field of breast cancer. These challenges are mainly attributed to intratumor heterogeneity [4], which is characterized by subclonal diversity within a tumor that originates from the accumulation of various somatic mutations during cell division and proliferation [5–7]. Intratumor heterogeneity has already been identified in several types of cancer, including breast, prostate, kidney, brain, liver, and lung cancers [8]. Drug-resistant subclones may develop via clonal evolution and reside at low frequencies within a tumor; after drug therapy, however, these subclones become the main population, leading to recurrence [9–11].

Intratumor heterogeneity can be most directly evaluated from DNA sequences using next-generation sequencing (NGS). One of the commonly used methods to analyze heterogeneity is the sequencing of samples from multiple regions of the same tumor [10, 12]. Ultra-deep sequencing can also be used to detect mutations with extremely low allele frequencies. Variant allele frequency (VAF), calculated as the proportion of reads with mutations at the variant site, is used as an index of heterogeneity [13, 14]. VAFs in a tumor can be used to determine the cellular prevalence of a mutation within a sample and estimate subpopulation frequencies and the tumor evolutionary process [15–17]. For example, deep sequencing was used to evaluate mutational processes of 21 breast cancers, leading to the finding that every tumor harbored a distinct subclonal lineage [10]. The use of NGS and analytical methods to define clonal heterogeneity has also provided insights into the genetic processes underlying breast cancer metastasis [18]. Recent studies also showed that clonal distribution based on VAF correlated with prognosis...
Additionally, heterogeneity can be evaluated using large datasets generated by The Cancer Genome Atlas (TCGA) or the International Cancer Genome Consortium [21, 22]. Despite these advances, it remains highly challenging to identify tumor genetic factors associated with tumor growth or metastasis, as tumors exhibit considerable heterogeneity.

Several recent studies based on TCGA data have focused on the identification of driver genes and identified pathways containing potential drug targets [23]. These studies have accelerated the development of pathway-specific inhibitory drugs. Although mutations in breast cancer driver genes such as TP53, PIK3CA, and GATA3 have been extensively investigated, somatic alterations in other genes are also believed to be associated with breast cancer [24]. To gain better insights into the extent of intratumor heterogeneity, we analyzed breast cancer genome sequencing data from TCGA. In this study, we focused on genes associated with breast cancer prognosis.

RESULTS

Identification of genes with high frequencies of mutations

We sought genes with one of four types of mutation (missense mutations, nonsense mutations, frameshift insertions, and frameshift deletions) in ≥ 50 samples derived from the 1,044 breast cancer datasets in TCGA. We identified 17 such genes (Table 1) and calculated the mean VAF for each in the sample containing mutations (Figure 1). All VAFs were adjusted for tumor purity taken from the previous study [25]. The mean VAFs of already known driver genes in breast cancer, such as TP53, PIK3CA, and CDH1 were found to be relatively high.

To examine the association of overall survival (OS) with VAFs of these 17 genes, we applied a Cox proportional hazards regression analysis with the covariates of age, tumor grade, and VAFs. In this analysis, the samples were divided into two groups using a VAF of 0.30 (i.e., 30%) as a cutoff. We used this cutoff because a previous study, which focused on samples with high tumor purity (≥ 70%), considered that a VAF of ≥ 0.25 was more likely to be clonal, whereas lower values were more likely to be subclonal [20]. Assuming that average purity is 85% (range 70–100%) then the cutoff should be 0.3 (0.25/0.85 = 0.294). We conducted this analysis without adjusting for other covariates just for a screening of genes that are possibly associated with breast cancer prognosis. We corrected P values for multiple testing using Benjamini and Hochberg false discovery rate (FDR) [26]. VAFs of HMCN1 were found to be possibly associated with breast cancer prognosis (FDR < 0.1) (Table 1), and we focused on HMCN1, for which no association with breast cancer has previously been reported.

A total of 78 somatic mutations in HMCN1 were detected in 6.1% (64/1,044) of samples (Table 2 and Figure 2). Among samples with detectable HMCN1 mutations, 9.4% (6/64) contained two distinct mutations and 3.1% (2/64) contained more than two mutations. Of the 78 HMCN1 mutations, 82.1% (64/78) were missense, whereas 10.3% (8/78) were nonsense and 7.7% (6/78) were indels. Furthermore, 69.2% (54/78) of the mutations were clustered in the Ig-like C2-type domains of HMCN1.

To evaluate the existence of any association between the mutation type and VAFs, we applied a one-

![Figure 1: Frequently mutated genes and mean variant allele frequencies (VAFs). The scatter plot depicts 17 genes harboring mutations in > 50 samples. The x-axis indicates the mean VAF, and the y-axis indicates the number of samples with mutations. The plot at right is an enlargement of the square area enclosed by dotted lines in the left plot.](image-url)
way ANOVA to the data and found that the mean VAF values did not significantly differ among the four types of mutations ($P = 0.430$) (Supplementary Figure 1). We further evaluated the associations of the four molecular breast cancer subtypes with VAFs of $HMCN1$. However, an ANOVA indicated that the mean VAF values did not significantly differ among the four subtypes ($P = 0.060$) (Supplementary Figure 2).

Expression of $HMCN1$

We compared the relative $HMCN1$ mRNA expression levels among samples with higher (VAF of $\geq 0.30$, $n = 19$) and lower VAFs (VAF of $< 0.30$, $n = 45$). As a result, we found that $HMCN1$ expression levels did not significantly differ between the two groups ($P = 0.343$) (Supplementary Figure 3A). Additionally, we compared $TP53$ and $PIK3CA$ expression levels between the VAF groups and found no significant differences in either ($P = 0.515$ and 0.300, respectively) (Supplementary Figure 3B and 3C). We also found no significant differences in the relative $HMCN1$ mRNA expression levels between $HMCN1$ mutant and wild-type samples ($P = 0.984$) (Supplementary Figure 3D).

To identify genes for which the expression levels were associated with the $HMCN1$ VAF, we analyzed mRNA expression levels of all annotated genes. Among the annotated genes, only two significantly exhibited

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sample count</th>
<th>Mean value of VAF</th>
<th>Hazard Ratio (95% CI)</th>
<th>$P$-value</th>
<th>FDR $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PIK3CA$</td>
<td>304</td>
<td>0.471</td>
<td>1.78 (0.729–4.348)</td>
<td>0.206</td>
<td>0.696</td>
</tr>
<tr>
<td>$TP53$</td>
<td>293</td>
<td>0.626</td>
<td>1.276 (0.455–3.581)</td>
<td>0.643</td>
<td>0.994</td>
</tr>
<tr>
<td>$TTN$</td>
<td>193</td>
<td>0.311</td>
<td>1.85 (0.843–4.06)</td>
<td>0.125</td>
<td>0.531</td>
</tr>
<tr>
<td>$MUC16$</td>
<td>113</td>
<td>0.305</td>
<td>1.768 (0.619–5.048)</td>
<td>0.287</td>
<td>0.696</td>
</tr>
<tr>
<td>$CDH1$</td>
<td>104</td>
<td>0.468</td>
<td>1.019 (0.263–3.951)</td>
<td>0.979</td>
<td>0.999</td>
</tr>
<tr>
<td>$GATA3$</td>
<td>102</td>
<td>0.398</td>
<td>0.859 (0.245–3.01)</td>
<td>0.813</td>
<td>0.999</td>
</tr>
<tr>
<td>$KMT2C$</td>
<td>86</td>
<td>0.370</td>
<td>1.327 (0.478–3.687)</td>
<td>0.587</td>
<td>0.994</td>
</tr>
<tr>
<td>$MAP3K1$</td>
<td>77</td>
<td>0.436</td>
<td>0.114 (0.013–0.985)</td>
<td>0.048</td>
<td>0.287</td>
</tr>
<tr>
<td>$HMCN1$</td>
<td>64</td>
<td>0.251</td>
<td>11.441 (2.065–63.406)</td>
<td>0.005</td>
<td>0.090*</td>
</tr>
<tr>
<td>$USH2A$</td>
<td>63</td>
<td>0.283</td>
<td>1.185 (0.245–5.74)</td>
<td>0.833</td>
<td>0.999</td>
</tr>
<tr>
<td>$RYR2$</td>
<td>62</td>
<td>0.291</td>
<td>0.059 (0.003–1.008)</td>
<td>0.051</td>
<td>0.287</td>
</tr>
<tr>
<td>$SYNE1$</td>
<td>56</td>
<td>0.272</td>
<td>1.635 (0.11–24.181)</td>
<td>0.721</td>
<td>0.999</td>
</tr>
<tr>
<td>$FLG$</td>
<td>53</td>
<td>0.264</td>
<td>0.342 (0.04–2.923)</td>
<td>0.327</td>
<td>0.696</td>
</tr>
<tr>
<td>$SPTA1$</td>
<td>52</td>
<td>0.264</td>
<td>1.965 (0.531–7.274)</td>
<td>0.312</td>
<td>0.696</td>
</tr>
<tr>
<td>$DMD$</td>
<td>51</td>
<td>0.261</td>
<td>1.87 (0.396–8.831)</td>
<td>0.429</td>
<td>0.811</td>
</tr>
<tr>
<td>$NEB$</td>
<td>50</td>
<td>0.278</td>
<td>1.124 (0.135–9.342)</td>
<td>0.914</td>
<td>0.999</td>
</tr>
<tr>
<td>$ZFHX4$</td>
<td>50</td>
<td>0.264</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1: Frequently mutated genes and mean variant allele frequencies and hazard ratios

Abbreviations: 95% CI, 95% confidence interval; VAF, variant allele frequency; FDR, false discovery rate.

$^a$Asterisk indicates FDR < 0.1.

$^b$ZFHX4 VAF could not be analyzed because sample size is small.

Figure 2: A schematic of the domains of human $HMCN1$ (hemicentin-1). The types and positions of 78 somatic mutations are indicated above the diagram.
different expression in terms of the HMCN1 VAF. A high CA9 and CASP14 expression level (P = 0.043 and 0.024, respectively) and low MTRNR2L1 and TCN1 expression level (P = 0.024 and 0.043, respectively) were found to be significantly associated with a higher VAF (Figure 3). CA9 encodes carbonic anhydrase IX, an endogenous marker of hypoxic cells in breast cancers. CASP14 encodes caspase14, which is one of the apoptosis-related cysteine peptidase. MTRNR2L1 encodes human MT-RNR2-like 1, for which detailed functions remain unknown and TCN1 encodes a member of the vitamin B12-binding protein family, named “transcobalamin 1”.

**Relationship between intratumor heterogeneity and HMCN1 VAFs**

To investigate intratumor heterogeneity in individuals with HMCN1 mutations, we measured the

<table>
<thead>
<tr>
<th>HMCN1 Domains</th>
<th>Missense mutation</th>
<th>Nonsense mutation</th>
<th>Deletion</th>
<th>Insertion</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>VWFA domain</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Ig-like C2-type domains</td>
<td>45</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>54</td>
</tr>
<tr>
<td>TSP type-1 domains</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Nidogen G2 beta-barrel domain</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>EGF-like domains</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>64</strong></td>
<td><strong>8</strong></td>
<td><strong>3</strong></td>
<td><strong>3</strong></td>
<td><strong>78</strong></td>
</tr>
</tbody>
</table>

Abbreviations: EGF, epidermal growth factor; Ig, immunoglobulin; TSP, thrombospondins; VWFA, von Willebrand factor type A.

**Figure 3: CA9 and MTRNR2L1 mRNA expression according to HMCN1 variant allele frequencies (VAFs).** Samples were divided into two groups using a VAF cutoff of 0.30 (< 0.30, n = 45 and ≥ 0.30, n = 19). The asterisk indicates statistical significance.
number of subclones across each sample. We performed this analysis using the SciClone [16] with VAFs from somatic SNVs and copy number estimates. We compared between the higher and the lower HMCN1 VAF groups in terms of the number of subclones by Fisher’s exact test. The distributions of the number of subclones were not significantly different between the two groups (\(P = 0.347\)) (Supplementary Figure 4A).

There is another index for intratumor heterogeneity, mutant-allele tumor heterogeneity (MATH), which can be calculated from VAF distribution in a sample [27]. Previous studies have shown that higher MATH score was correlated with poor prognosis in head and neck squamous cell carcinoma and colon cancer [27, 28]. We compared between the higher and the lower HMCN1 VAF groups in terms of MATH scores by Wilcoxon signed rank test. In the 19 samples with higher HMCN1 VAF, the mean MATH was 35.014 (SD = 11.300). Meanwhile, in the 45 samples with lower HMCN1 VAF, the mean MATH was 33.919 (SD = 10.557). No significant differences between the two groups were detected (\(P = 0.771\)) (Supplementary Figure 4B). These findings indicate that the prevalence of the mutations in HMCN1 might not be involved in the status of intratumor heterogeneity.

**Associations of HMCN1 with common driver genes**

As TP53 and PIK3CA mutations are among the most common genetic aberrations in breast cancers [23], we compared the VAFs of these two driver genes with those of HMCN1. Among the 64 samples harboring mutations in HMCN1, 22 and 23 also harbored mutations in TP53 and PIK3CA, respectively, and five harbored mutations in both genes.

In the 22 samples with both TP53 and HMCN1 mutations, the mean TP53 VAF was 0.697 (SD = 0.249) and the mean HMCN1 VAF was 0.288 (SD = 0.201). Meanwhile, in the 23 samples with both PIK3CA and HMCN1 mutations, the mean PIK3CA VAF was 0.442 (SD = 0.269) and the mean HMCN1 VAF was 0.230 (SD = 0.148). A paired t-test showed that VAFs of the two driver genes were significantly higher than that of HMCN1 (\(TP53; P < 0.01, \text{PIK3CA}; P < 0.01\)) (Supplementary Figure 5), indicating that mutations in HMCN1 occurred later in the tumor evolutionary process than the mutations in TP53 and PIK3CA. This finding suggests that the mutations in HMCN1 might be involved in breast cancer progression.

**HMCN1 mutations and clinical outcomes**

Next, we evaluated the association between HMCN1 mutation status and clinical variables by \(\chi^2\) test or Fisher’s exact test. We found tumor size (\(P = 0.028\)) and molecular subtype (\(P = 0.021\)) were related with the HMCN1 mutation (Table 3). To assess the relationship of the HMCN1 VAF with prognosis, the 64 samples harboring HMCN1 mutations were divided into two groups according to VAFs and subjected to an OS analysis. These groups were also compared with individuals without HMCN1 mutations (wild-type; WT). The resulting Kaplan–Meier plot shows that a higher HMCN1 VAF significantly correlated with poor prognosis (log-rank test: vs. WT; \(P = 0.022\) and vs. VAF of < 0.30; \(P = 0.015\)) (Figure 4). Concordantly, in a multivariate Cox proportional hazards regression analysis adjusted for the covariates of lymph node status, tumor grade, tumor size, and age, the VAF (\(P = 0.036\)) and lymph node status (\(P = 0.012\)) were significantly associated with poor prognosis (Table 4).

To exclude the possibility that this significant association is not attributable to bias in the impact of mutations in the two groups, we evaluated the impact of single nucleotide variants in HMCN1 on protein structure and function using PolyPhen-2 scores [29]. The Pearson correlation coefficient between VAFs and PolyPhen-2 scores of HMCN1 was -0.204 (\(P = 0.151\)), indicating no significant correlation. To analyze the relationship between prognosis and PolyPhen-2 scores, the 64 samples were divided into two groups using a PolyPhen-2 score of 0.85 as a cutoff (higher, \(n = 27\) and lower, \(n = 24\)); this score ranges from 0 to 1 and yields predictions of “probably damaging” (> 0.85), “possibly damaging” (0.85–0.15), or “benign” (< 0.15). We found that the Polyphen-2 score of nonsynonymous HMCN1 mutations did not significantly associate with breast cancer prognosis (PolyPhen-2 scores of < 0.85 vs. WT; \(P = 0.801\) and PolyPhen-2 scores of ≥ 0.85 vs WT; \(P = 0.671\)) (Supplementary Figure 6).

These results suggest that the HMCN1 VAF is an independent prognostic factor for OS, such that a higher VAF may be associated with poor survival in patients with breast cancer.

**Correlations with potential prognostic factors**

We next evaluated the association of the HMCN1 VAF with individual clinical characteristics (lymph node status, tumor grade, tumor size, and age) in the 64 tumor samples, which were divided into three groups by lymph node status: N0, N1, and N2–N3. Samples were also divided into three groups by tumor grade: grades 1, 2, and 3–4. Regarding lymph node status, tumor grade, and tumor size, we examined whether a higher VAF was associated with significantly higher stages of clinical features using the Cochran–Armitage trend test. We found a significant association of a higher VAF with a much higher lymph node status (\(P = 0.029\)) (Figure 5A). By contrast, the tumor grade (\(P = 0.151\)) and tumor size (\(P = 0.283\)) were not significantly associated with the HMCN1 VAF (Figure 5B and 5C). The mean ages of patients (\(n = 64\)) in the higher and lower VAF groups were 58.05 (SD = 17.95) years and 61.41 (SD = 11.85) years, respectively. A t-test revealed no significant difference in the mean age between the groups (\(P = 0.461\)) (Figure 5D).
We also used other 15 types of cancer dataset from TCGA and examined the association between HMCN1 VAFs and OS. In the 15 types of cancer, only cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) samples showed significantly poorer OS in the samples with higher HMCN1 VAFs (n = 7, VAF ≥ 0.30) than those with lower VAFs (n = 15, VAF < 0.30) (log-rank test: P = 0.048) (Supplementary Figure 7). Although, we applied a Cox proportional hazards regression analysis with the covariates of age, tumor grade, and HMCN1 VAFs, the VAFs were not associated with poor prognosis (HR = 5.436, 95% CI: 0.543–54.432, P = 0.150).

**DISCUSSION**

Breast cancers are known to exhibit a large degree of genetic heterogeneity. In this study, we analyzed intratumor heterogeneity using VAFs calculated from a set of breast cancer cases registered in TCGA. Through a VAF-based analysis of mutations, we showed that VAFs of HMCN1 was possibly associated with breast cancer prognosis. Although the detailed function of HMCN1 in humans remains unknown, mutations in HMCN1 might be associated with cancer cell invasion and metastasis. Using the data of the DRIVE dataset, the CIMBA dataset, and the Foundation One dataset from breast cancer patients,
we could not validate our result because the HMCN1 VAF and survival time information could not be obtained. In cervical cancer from TCGA, however, survival time between the two groups of HMCN1 VAF values were also significantly different. Cervical cancer, like breast cancer, is known to associated with hormone estrogen. Therefore, this result may support the prognostic impact of HMCN1 on breast cancer. It will be possible to further evaluate the validity of our results by accumulating more cohorts.

HMCN1 encodes a large extracellular protein belonging to the immunoglobulin superfamily and comprises several distinct domains, including the von Willebrand factor and Ig-like C2-type domains [30]. HMCN1 mutations are believed to correlate with age-related macular degeneration [31]. According to another recent report, HMCN1 acts as a suppressor of gallbladder cancer metastasis [32] and is commonly mutated in certain samples of head and neck squamous cell carcinoma [33]. There are at least three possible functional implications of the mutations in HMCN1 in breast cancer metastasis. First, HMCN1 is also known as FBLN6 (fibulin 6) and one of the extra cellular matrix (ECM) proteins [34, 35].

Table 4: Multivariate Cox proportional hazards regression analysis of overall survival according to the clinical characteristics of 64 breast cancer patients

<table>
<thead>
<tr>
<th>Variables</th>
<th>Hazard Ratio (95% CI)</th>
<th>P-value *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymph node status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive vs Negative</td>
<td>97.931 (2.709–3539.805)</td>
<td>0.012</td>
</tr>
<tr>
<td>Tumor grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3–4 vs 1–2</td>
<td>9.468 (0.042–2147.487)</td>
<td>0.417</td>
</tr>
<tr>
<td>Tumor size, cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2–5 vs ≤ 2</td>
<td>1.281 (0.159–10.302)</td>
<td>0.816</td>
</tr>
<tr>
<td>&gt; 5 vs ≤ 2</td>
<td>29.032 (0.284–2966.692)</td>
<td>0.154</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 50 vs &lt; 50</td>
<td>0.114 (0.011–1.169)</td>
<td>0.068</td>
</tr>
<tr>
<td>VAF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 0.30 vs &lt; 0.30</td>
<td>17.950 (1.216–264.976)</td>
<td>0.036</td>
</tr>
</tbody>
</table>

Abbreviations: 95% CI, 95% confidence interval; VAF, variant allele frequency.

*Astarisk indicates statistical significance.

Figure 4: Kaplan–Meier analysis of overall survival according to HMCN1 variant allele frequencies (VAFs). Samples were divided into three groups using a VAF cutoff of 0.30 (< 0.30, red, n = 45 and ≥ 0.30, blue, n = 19) or WT (black, n = 968). The log-rank test was used to evaluate the statistical significance of the difference between the two survival curves (VAF of ≥ 0.30 vs. VAF of < 0.30, VAF of < 0.30 vs WT and VAF of ≥ 0.30 vs WT).

WT (n = 968) vs WT; Log-rank P = 0.233

VAF < 0.3 (n = 45) vs WT; Log-rank P = 0.022*

VAF ≥ 0.3 (n = 19) vs VAF < 0.3; Log-rank P = 0.015*

www.oncotarget.com
fibulins are shown to be involved in basement membrane and formation of stable cell-to-cell interactions, leading to organization and stabilization to ECM structure [36]. When HMCN1 does not function properly in cancer cell, sufficient cell adhesion might be inhibited and as a result of promoting cancer invasion due to instability of HMCN1 caused by the variants in the gene. For example, previous study reported that epigenetically silenced fibulin 5 promotes invasion and metastasis in lung cancer [37]. Second, HMCN1, which contains estrogen receptor binding site, seems to be associated with postpartum depression symptoms [38], and one of its functions is suggested to be cell adhesion [39]. Therefore, HMCN1 mutations may be associated to cancer proliferation and metastasis because of the disruption of these functions. Finally, other studies have shown that HMCN1 might interact with DDX1 [40], a DEAD box protein with RNA helicase activity [41, 42]. Notably, the expression of DDX1 was reported to decrease under hypoxic conditions [43], and intratumor hypoxia is associated with cancer metastasis and, consequently, patient mortality [44, 45]. Indeed, an earlier report found that DDX1 correlated with ovarian tumor metastasis and progression [46]. In the current study, we found that CA9, the expression of which is also associated with tumor hypoxia [43, 47], was expressed at significantly higher levels in patients with higher HMCN1 VAFs than in those with lower VAFs. Therefore, the HMCN1 VAF may indicate the metastatic potential of a breast cancer.

Although metastasis is the main cause of death among breast cancer patients, factors involved in metastasis remain poorly characterized. It is more difficult to identify genetic factors associated with metastasis, a complex process, than to identify driver genes [48]. Differences in genetic heterogeneity between metastatic and primary tumors may affect treatment efficacy and thus represent one of the biggest obstacles toward cure for breast cancer. Using VAFs to explicitly address heterogeneity, we successfully identified HMCN1 as a possible metastatic factor. Although further experimental validation is needed to determine the involvement of HMCN1 in metastasis, this approach could be used to screen genes that have not previously been investigated.

In this study, we focused on nonsynonymous mutations and indels. However, intratumor heterogeneity may also be caused by copy number variants or mutations.
in noncoding regions, such as those in cis-regulatory elements and splice sites [49–52]. Moreover, epigenetic alterations can also promote cancer progression [53]. As genome-wide epigenetic datasets from normal cells grow rapidly [54], epigenome data analyses of cancer cells will allow evaluations of the impacts of epigenetic factors on intratumor heterogeneity.

In conclusion, to our knowledge, this is the first study to identify HMCN1 as a potential metastatic factor in breast cancer using a comparative analysis of genomic and transcriptomic data registered in TCGA. In addition to the standard classification of breast tumors based on the four molecular types, the use of VAFs, which reflect tumor evolution, might provide further genetic profile information that can be used to characterize tumor samples. Our approach allows us to identify new diagnostic markers or candidate genes for targeted therapy and is therefore expected to facilitate precision medicine.

**MATERIALS AND METHODS**

**Datasets**

A total of 1,080 RNA-seq and variant datasets from breast cancers were downloaded from TCGA (https://portal.gdc.cancer.gov/). For variant data, we used VCF files generated by comparing matched tumor–normal pairs using the Mutect2 software package. We also downloaded the associated clinical patient data. Similarly, dataset from other 15 types of cancer were obtained from TCGA: Bladder urothelial carcinoma (BLCA; n = 416), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC; n = 307), colon adenocarcinoma (COAD; n = 605), glioblastoma multiforme (GBM; n = 938), head and neck squamous cell carcinoma (HNSC; n = 512), kidney renal clear cell carcinoma (KIRC; n = 697), lower grade glioma (LGG; n = 938), liver hepatocellular carcinoma (LIHC; n = 378), lung adenocarcinoma (LUAD; n = 587), lung squamous cell carcinoma (LUSC; n = 503), ovarian serous cystadenocarcinoma (OV; n = 443), prostate adenocarcinoma (PRAD; n = 503), skin cutaneous melanoma (SKCM; n = 472), thyroid carcinoma (THCA; n = 504), and uterine corpus endometrial carcinoma (UCEC; n = 604).

**Mutation analysis**

In this study, we only considered mutations with a coverage depth of ≥20. We extracted gene mutations [i.e., nonsynonymous substitutions (missense and nonsense mutations) and indels (frameshift insertions and frameshift deletions)] observed in ≥50 samples. We used this cutoff because the lower limit of the average mutation rate for significantly mutated genes was approximately 2–4% [23]. VAFs were calculated as the proportion of variant allele reads to total reads at the mutation site. When a sample harbored multiple mutations in the same gene, the larger VAF was used as the VAF for the gene. The VAF was adjusted for tumor purity estimate. This estimate, which was derived from immunohistochemistry analysis, was downloaded from the previous study [25].

The number of subclones in a tumor cell were inferred by both VCF files and DNA copy number variation data using the R package SciClone with default settings [16].

**Statistical analysis**

Statistical analysis was conducted using the R software, version 3.3.1 (R Project for Statistical Computing, Vienna, Austria), and JMP Pro, version 13.0 (SAS Institute Inc., Cary, NC, USA). A χ² test or Fisher’s exact test (when ≥1 cells had an expected frequency of ≤5 in any clinical group) was used to evaluate the relationships between the mutation status and clinical variables. We also used the test for comparison the number of subclones. OS was estimated using the Kaplan–Meier method in the R survival package (version 2.41–3). For the multivariate analysis, adjusted hazard ratios (HRs) with 95% confidence intervals (95% CIs) were calculated using a Cox proportional hazards regression model. We used edgeR (version 3.16.5), a Bioconductor package (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html), to detect genes differentially expressed between two groups. For each gene, the R exactRankTests package (version 0.8–28) was used to evaluate the difference in both expression levels and MATH between groups of samples. For categorical data such as tumor grade, tumor size category, and lymph node status, we used a one-sided Cochran–Armitage trend test to evaluate the existence of a linear relationship in terms of VAFs. We used Welch’s t-test to compare continuous data between the two groups. An analysis of variance (ANOVA) model was used to compare the mean values of more than two groups. P-values were considered statistically significant at <0.05 (*P < 0.05, **P < 0.01).

**Author contributions**

Chie Kikutake: Conception and design of study, writing original draft, and data analysis. Minako Yoshihara: Data interpretation, review, and supervision. Tetsuya Sato: Data interpretation, review, and supervision. Daisuke Saito: Data interpretation, review, and supervision. Mikita Suyama: Data review and interpretation, writing review and editing final draft, supervision, and project administration. All authors reviewed and approved the article.

**ACKNOWLEDGMENTS**

This work was supported by the “Advanced Computational Scientific Program” of the Research Institute for Information Technology, Kyushu University. The results shown here are partly based...

CONFLICTS OF INTEREST

The authors declare they have no competing interests.

FUNDING

This study was supported by JSPS Grants-in-Aid for Scientific Research [Grant number: 17H03619 to M.S. and 17KT0128, 17K07257 to T.S.] and The Shin-Nihon of Advanced Medical Research [to T.S.].

REFERENCES