# Immortalization of different breast epithelial cell types results in distinct mitochondrial mutagenesis.

**Sujin Kwon[1], Susan S. Kim[3,4], Howard E. Nebeck[1,4], Eun Hyun Ahn[1,2,*]**

[1]Department of Pathology, University of Washington, Seattle, WA 98195, USA; kwonsu17@uw.edu (S.J.K); henebeck@uw.edu (H.E.N); ahneun@uw.edu (E.H.A)

[2] Institute of Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA 98109, USA; ahneun@uw.edu (E.H.A)

[3] Department of Biochemistry, University of Washington, Seattle, WA 98195, USA; ssk773@uw.edu (S.S.K)

[4] Equal contribution

* Correspondence to: ahneun@uw.edu

Received: date; Accepted: date; Published: date

**Abstract:** Different phenotypes of normal cells might influence genetic profiles, epigenetic profiles, and tumorigenicities of their transformed derivatives. In this study, we investigated whether the whole mitochondrial genome of immortalized cells can be attributed to different phenotypes (stem vs non-stem) of their normal epithelial cell originators. To accurately determine mutations, we employed Duplex Sequencing, which exhibits the lowest error rates among currently available DNA sequencing methods. Our results indicate that the vast majority of observed mutations of the whole mitochondrial DNA occur at low-frequency (rare mutations). The most prevalent rare mutation types are C→T/G→A and A→G/T→C transitions. Frequencies and spectra of homoplasmic point mutations are virtually identical between stem cell-derived immortalized (SV1) cells and non-stem cell-derived immortalized (SV22) cells, verifying that both cell types were derived from the same woman. However, frequencies of rare point mutations are significantly lower in SV1 cells (5.79x10-5) than in SV22 cells (1.16x10-4). Additionally, the predicted pathogenicity for rare mutations in the mitochondrial tRNA genes is significantly lower (by 2.5-fold) in SV1 cells than in SV22 cells. Our findings suggest that the immortalization of normal cells with stem cell features leads to decreased mitochondrial mutagenesis, particularly in noncoding RNA regions. The mutation spectra and mutations specific to stem cell-derived immortalized cells (vs non-stem cell derived) have implications in characterizing heterogeneity of tumors and understanding the role of mitochondrial mutations in immortalization and transformation of human cells.

**Keywords:** Mitochondrial DNA, rare mutation, stem cells, breast cancer, tRNA, Duplex Sequencing, next generation sequencing

## 1. Introduction

Evidence exists that distinct phenotypes of normal cell [1-5] or precancerous cell [6] originators for tumor derivatives lead to tumor heterogeneities. For example, Ince et al (2007) reported that breast tumorigenic cells transformed from two different normal epithelial cell types (BPECs; HMECs) exhibited marked differences in histopathology, tumorigenicity, and metastatic abilities. Transformed BPECs caused lung metastases, whereas transformed HMECs were nonmetastatic [4]. Transformed BPECs were up to 104-fold more tumorigenic than transformed HMECs. However, no study has investigated subclonal mutations for transformed derivatives of different normal cell types.

Here, to study consequences of normal cell origins on genetic changes in their immortalized derivatives, we examined two different immortalized (pre-neoplastic) breast epithelial cell types that

were derived from normal human breast epithelial cells (HBECs) with different phenotypes (Types I and II). Both Type I and Type II normal HBECs were isolated from breast tissue of the same woman; however, they exhibited different phenotypes. Type I HBECs display stem cell characteristics and have been characterized by: the expression of a stem cell marker octamer-binding transcription factor 4 (Oct-4) [7] and estrogen receptor (ER)-$\alpha$ [8], and luminal epithelial markers [9,10]; a deficiency in gap-junction associated intercellular communication (GJIC) [9,11]; the ability to display anchorage-independent growth [11]; the ability to differentiate into Type II (normal differentiated) HBECs [9,11]; reduced expression of maspin [12]; and the ability to form budding/ductal organoids on Matrigel in conjunction with Type II HBECs [11]. In contrast, Type II HBECs exhibit opposite phenotypes in the described features above (i.e., do not express stem cell markers; do not express ER-$\alpha$; express basal epithelial markers; express GJIC proteins; higher expression of maspin).

Both Type I (stem cell features) and Type II (without stem cell features) HBECs were transformed with Simian virus 40 large T-antigen (SV40-T) into immortal/non-tumorigenic (pre-neoplastic) cells, M13SV1 and M13SV22, respectively. SV40-T is widely used to immortalize and transform mammalian cells [13,14]. Type II HBECs are less susceptible to SV40-T transformation than are Type I HBECs, and rarely become immortal following transfection with SV40-T [8,9,11,15]. This suggests that Type I HBECs and Type I HBEC-derived immortalized cells appear to be the major target cells for breast carcinogenesis and transformation.

In the present study, we examined whether the whole mitochondrial genome of these different immortalized human breast epithelial cell types is influenced by phenotypes of their originator normal HBECs (Type I and Type II). To accurately investigate low-frequency (heteroplasmic) rare and subclonal mutations as well as high-frequency (homoplasmic) mutations, we applied Duplex Sequencing (DS), which detects mutations with unprecedented accuracy [16-19]. Unlike conventional sequencing technologies that sequence only a single strand of DNA, DS sequences both strands of DNA and scores mutations only if they are present in both strands of the same DNA molecule as complementary substitutions. This significantly lowers background error rates (eg., Next generation sequencing error rates $10^{-2}$ to $10^{-3}$; DS error rates < $5 \times 10^{-8}$) [16,17,19-21].

## 2. Results

In this study, non-stem cell-derived immortalized human breast epithelial cells (HBECs) will be referred to as SV22 and stem cell-derived immortalized HBECs will be referred to as SV1 cells. Both SV22 and SV1 cells were cultured under the same conditions, DNA was extracted, and DNA libraries were prepared for Duplex Sequencing (DS) as we have described previously [22]. The average number of nucleotides at each genome position (depth) was calculated as the total number of duplex consensus sequence (DCS) nucleotides sequenced divided by the mtDNA size of 16569 bases. The DCS depths were 1060 (all Tables and Figures except Fig S4) for SV22 cells and 1666 (all Figures and Tables) and 2738 (Fig S4) for SV1 cells. These are estimated to be equivalent to single-strand tag-based sequencing depths of approximately 5000 to 6200 and 7900 to 9800, respectively [19].

We have defined mtDNA mutations (variants) found in SV22 and SV1 cells as homoplasmic (90-100%), subclonal (0-10%) and rare (0-1%) based on the mutation occurrence (%) at each genome position. Maternally inherited mitochondrial mutations arising during early embryonic development are more likely to be homoplasmic [23,24]. In the present study, we focused on rare and subclonal variants, as they presumably represent *de novo* somatic variants. Rare or subclonal mutations are not accurately determined by conventional sequencing methods due to their high background error frequencies ($10^{-2}$ to $10^{-3}$) [20,21,25]. These rare and subclonal mutations, however, are accurately detectable by Duplex Sequencing [16-19].

### *2.1. Both SV22 and SV1 cells exhibit identical homoplasmic mutations, verifying that both cell types were derived from the same individual.*

Thirty-five identical homoplasmic unique mutations were found between the two cell types (Fig S1). Frequencies, types (%), and context fractions (%) of homoplasmic mutations were almost identical (Fig S1) in both cell types. T>C/A>G and C>T/G>A transitions are the only mutation types observed with T>C/A>G being more dominant than C>T/G>A (Fig S1). As homoplasmic mitochondrial mutations are more likely to be maternally inherited mutations or variants arising during early embryonic development, our finding of identical homoplasmic mutations between the two cell types verify that they were derived from the same woman.

*2.2. SV1 cells show significantly lower frequencies of rare mutations and subclonal mutations than do SV22 cells.*

We determined frequencies of rare and subclonal mutations in both cell types by Duplex Sequencing. The overall frequencies of both rare (Fig 1A) and subclonal (Fig S2A) mutations are significantly lower in SV1 cells (by 2-fold) than in SV22 cells. In addition, we determined frequencies of each point mutation type, of insertions, and of deletions. C>T/G>A and T>C/A>G transitions are the most dominant types for both cell types (Figs 1B-D, Figs S2B-D). Frequencies of each type of rare and subclonal mutation are also significantly lower in SV1 cells than in SV22 cells (Figs 1B, S2B).

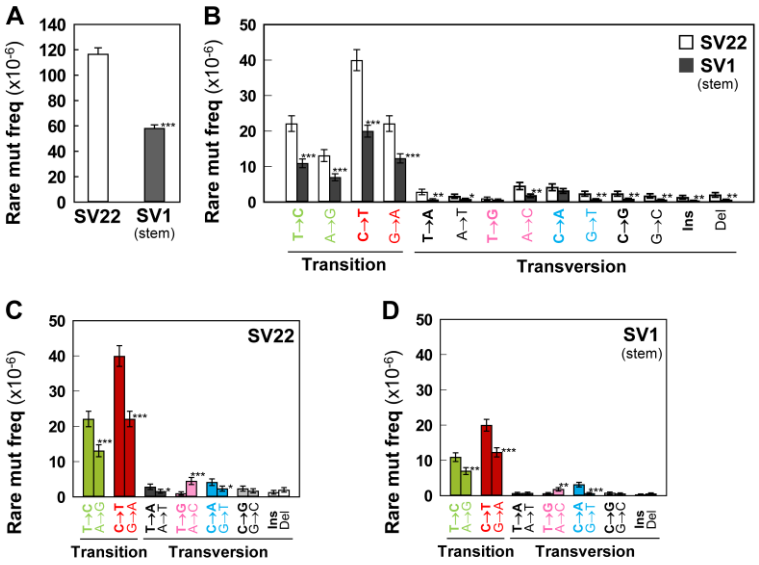


**Figure 1.** Frequencies of rare mutations in the whole mtDNA. Overall rare mutation frequency (**A**) and frequencies of rare mutation types (**B-D**) for SV22 (immortalized non-stem) and SV1 (immortalized stem) cells were determined using Duplex Sequencing. Error bars represent the Wilson Score 95% confidence intervals. Significant differences in rare mutation frequencies between two groups are indicated (* $p < 0.05$, ** $p < 5 \times 10\text{-}4$, and *** $p < 5 \times 10\text{-}10$) by the 2-sample test for equality of proportions with continuity correction.

*2.3. The strand bias of rare and subclonal mutations are observed in both cell types.*

The two strands of mtDNA are designated as heavy (H) and light (L) strands [26]. Our DS data are referenced to the L-strand. In the L-strand of both SV22 and SV1 cells, C>T are significantly more prevalent than G>A; T>C are significantly more prevalent than A>G; A>C are significantly more prevalent than T>G; C>A are significantly more prevalent than G>T (Figs 1C-D, S2C-D). These significant asymmetric distributions within the complementary mutation type reflect a strand orientation bias in both cell types.

*2.4. C>T/G>A transitions are the most prevalent mutation types followed by T>C/A>G in both cell types.*

The fraction (%) of rare and subclonal mutation types were calculated (Figs 2A, S3A). In both SV22 (non-stem) and SV1 (stem) cells, the most prevalent rare and subclonal mutation types are C>T/G>A and T>C/A>G (Figs 2A, S3A). The percentages of C>T/G>A and T>C/A>G rare mutations are similar between both cell types. In contrast, the fractions of the four rare mutation types in SV22 and SV1 cells are different by about 1.5-fold with higher fractions C>G/G>C, T>A/A>T, and T>G/A>C mutation types in SV22 cells and higher fractions of C>A/G>T mutation types in SV1 cells (Fig 2A).

To investigate influences of neighboring bases on types of rare and subclonal mutations, the bases immediately 5′ and 3′ to each mutated base were identified (i.e. the mutation occurs at the second position of each such trinucleotide). This allows classification of observed substitutions into 96 categories (4 bases X 6 substitutions x 4 bases). Numbers and fractions (%) of these mutation trinucleotides in each of the categories compose the "mutation context spectra" (MCS) of the cells. For both rare and subclonal mutations, T>G transversions in context GTA were significantly lower (p=0.02) in SV1 than in SV22 cells (Figs 2B-C, S3B-C).

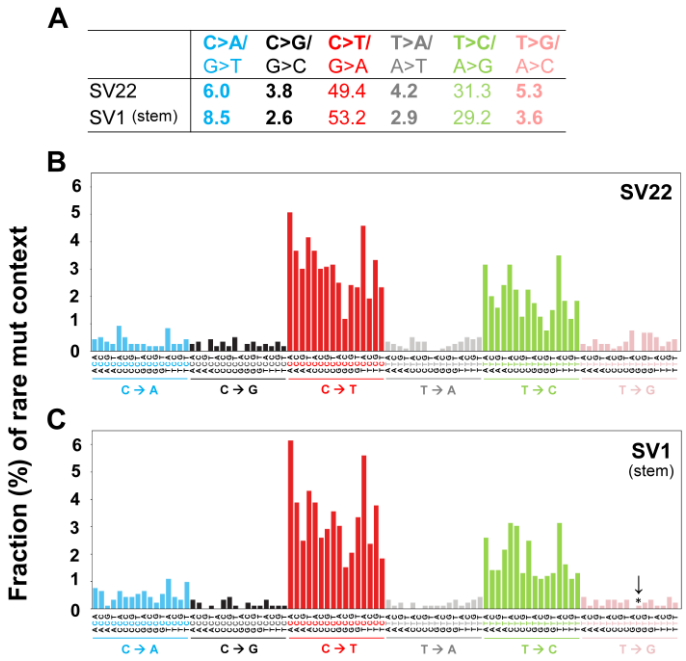


**Figure 2.** Types and sequence context spectra of rare unique mutations in the whole mtDNA. Fractions (%) of rare mutation types (A) and fractions (%) of rare mutation context spectra (**B,C**) for SV22 (immortalized non-stem) and SV1 (immortalized stem) cells were determined using Duplex Sequencing. Trinucleotide contexts (**B,C**) are the mutated base surrounded by all possibilities for its immediate 5′ and 3′ bases. To keep the graph concise, these point mutation trinucleotides are complemented as necessary to always depict the reference base as the pyrimidine of its pair. The fraction (%) of each specific trinucleotide out of all 96 possible trinucleotide contexts depicts the contribution of each genome sequence context to each point mutation type. Significant differences in fractions (%) of mutation context types between the two groups are indicated (* $p < 0.05$) by the 2-sample test for equality of proportions with continuity correction).

*2.5. The decreased mitochondrial mutagenesis of SV1 cells occurs mainly in the noncoding RNA region (rRNA and tRNA) of the mitochondrial genome*

The human mitochondrial genome consists of three regions: protein coding (13 genes), non-coding (nc) RNA (2 rRNA and 22 tRNA genes), and control/overlapped regions. We examined which of the three regions is more vulnerable to mitochondrial mutagenesis (Fig 3A, Table S1) in SV22 and SV1 cells.

Of the three categorized regions, the ncRNA regions exhibit the most significant differences in the number of rare mutations between SV22 and SV1 cells. The number of rare unique mutations in

the ncRNA regions is significantly lower in SV1 cells than in SV22 cells (Fig 3A, Table S1; p = 2.6x10-15). However, the sizes of the three regions are different, contributing to a bias in the number of observed unique mutations. In particular, the number of bases of the protein coding regions (11341) is greater than that of the ncRNA (4020) and the control/overlapped regions (approximately 1772) [27]. Thus, to compare mutation prevalence per the same genome size between protein-coding regions and ncRNA regions, we estimated the number of rare unique mutations per 1000 bases and found an even more significantly higher number of rare unique mutations in ncRNA regions than in the protein coding regions. This indicates that the ncRNA regions are more susceptible to mitochondrial mutagenesis than the protein-coding regions. Within the ncRNA region itself, there is a significantly lower number of rare unique mutations in SV1 cells than in SV22 cells (Fig 3B, Table S2; p = 1x10-4). Genome position and clonality (%) for each of the rare mutations observed in SV22 and SV1 cells are presented in Figure 4A,B. The average and median clonalities (%) of rare mutations are still about 1.7-fold significantly lower in SV1 cells (0.17% ± 0.006) (Fig 4D) than in SV22 cells (0.29% ± 0.007) (Fig 4C), even after excluding variants that were mutated only once (singlet mutations).

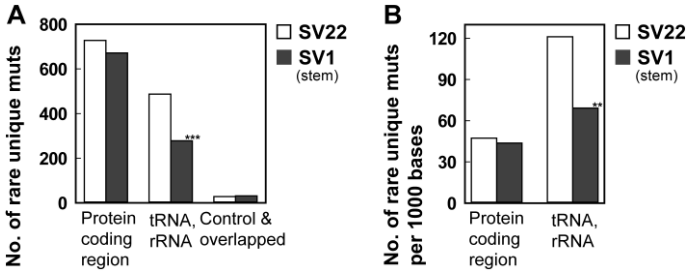


**Figure 3.** Number of rare unique mutations considered (individually and per 1000 bases) by genome positional category in the whole mtDNA. Number of rare unique mutations by positional category (**A**) and number of rare unique mutations per 1000 bases by positional category (**B**) were determined using ANNOVAR for SV22 (immortalized non-stem) and SV1 (immortalized stem) cells. Significant differences in numbers of rare mutations between the two groups are indicated (* $p < 0.05$, ** $p < 5 \times 10\text{-}4$, and *** $p < 5 \times 10\text{-}10$) by the 2-sample test for equality of proportions with continuity correction.

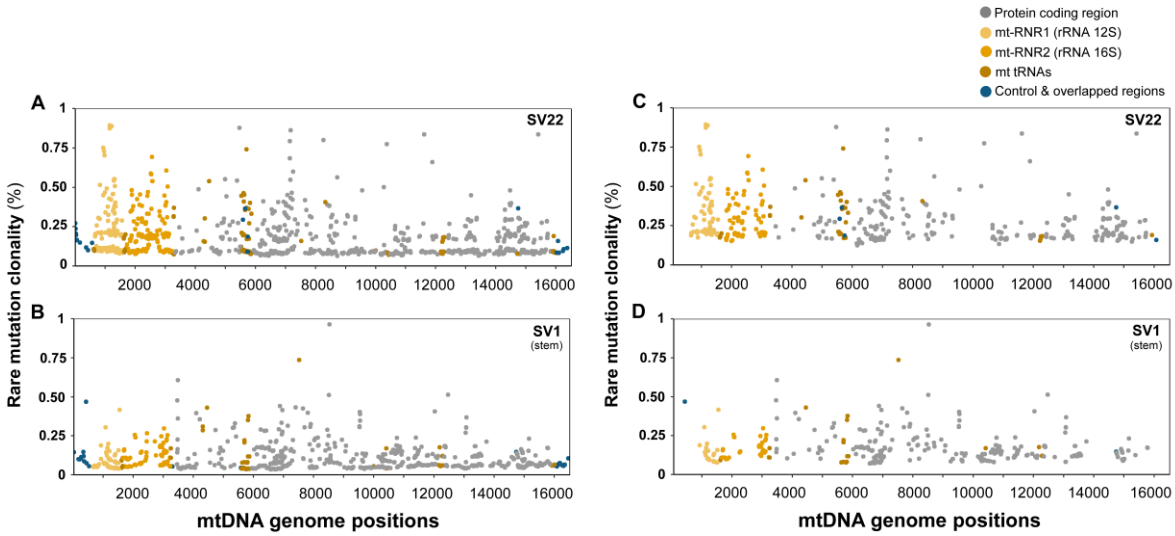


**Figure 4**. Genomic positions and clonalities (%) of rare unique mutations in the whole mtDNA. Rare mutation clonalities (%) by genomic position including singlets (**A,B**) or excluding singlets (**C,D**) were determined for SV22 (immortalized non-stem) and SV1 (immortalized stem) cells using Duplex Sequencing. Singlets are defined as variants that are mutated only once in nucleotides sequenced at a specific genome position.

*2.6. Proportions (%) of nonsynonymous mutations in mitochondrial protein coding regions and their predicted pathogenicity scores of subclonal mutations are similar between SV22 and SV1 cells*

We compared the proportion (%) of nonsynonymous mutations (causing changes in amino acids) of mitochondrial protein coding regions between SV22 and SV1 cells. Overall, no significant differences in rare or subclonal mutations were observed in the mitochondrial protein coding region between the two cell types (Table S3).

*2.7. Predicted pathogenicity scores of missense mutations of mitochondrial protein coding regions are similar between SV22 and SV1 cells.*

Predicted pathogenicity scores of the subclonal missense mutations were evaluated using MutPred [28]. MutPred provides a general *g* score for each missense mutation, where a higher *g* score indicates a higher chance that an amino acid substitution is deleterious. The average of the *g* score sums for each of the 13 mtDNA protein-coding genes was calculated for subclonal missense mutations present only in SV22 cells (4.25 ± 0.73) and for those present only in SV1 cells (4.15 ± 0.69) but not for those present in both cell types.

*2.8. Predicted pathogenicity scores of rare mutations of mitochondrial tRNA genes tend to be lower in SV1 cells than in SV22 cells.*

Variants in mt-tRNA genes are a common cause of mitochondrial disease [29]. Thus, we calculated pathogenicity values for rare and subclonal mutations in mt-tRNA genes using the mitochondrial tRNA informatics predictor (MitoTIP) program of MITOMAP [29]. MitoTIP provides a raw score, which is derived from three sub-scores: variant history and conservation score, position score, and secondary structure score. A higher score represents a higher probability of pathogenicity.

We compared the raw pathogenicity values for rare and subclonal mutations in each of 22 mt-tRNA genes found specifically in SV22 cells vs in SV1 cells. Eighty-three rare mutations were found in SV22 cells only; while 31 rare mutations were present in SV1 cells only. The exact same mutations specific to each cell type were observed in the subclonal range. The average pathogenicity score of mt-tRNA mutations tends to be lower in SV1 cells than in SV22 cells (Fig 5; $p = 0.082$). In addition, we found four known/confirmed pathogenic tRNA variants (m.5650 G>A, m.5521 G>A, m.5690 A>G, m. 1630 A>G) among the 83 variants in SV22 cells, whereas confirmed pathogenic tRNA variants were not found in SV1 cells.

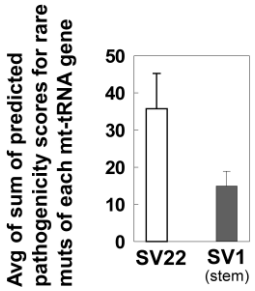


**Figure 5.** Predicted pathogenicity of rare unique mutations of mt-tRNAs. Predicted pathogenicity scores of rare variants in mt-tRNAs were obtained using MitoTIP and then were totaled for each region of tRNAs. Only the mutations present exclusively in each sample (mt-tRNA mutations present only in SV22 cells vs mt-tRNA mutations present only in SV1 cells) were included. The sums of predicted pathogenicity scores from each mt-tRNA region were averaged. Error bars represent the standard error of the mean (SEM).

*2.9. Duplex Sequencing identifies many rare or subclonal mutations that are specific to SV22 and to SV1 cells.*

To compare positions of mutations between SV22 and SV1 cells, only genome positions that had DCS read counts at least 100 or higher in both cell types were considered. Mutations occurring

at the same genome positions were scored only once and are referred to as unique mutations. We have identified rare and subclonal unique mutations that are present only in SV22 cells, only in SV1 cells, or in both cell types.

A total of 792 rare and 797 subclonal mutations were found in SV22 cells only (but not in SV1 cells); 527 rare and 528 subclonal mutations were found in SV1 cells only (but not in SV22 cells) (Table S4). 447 rare mutations were found in both SV22 and SV1 cells, of which 88 are more prevalent in SV22 than in SV1 and 9 are more prevalent in SV1 than in SV22 (prevalence meaning by at least a 3-fold difference) (Table S4). Similarly, 459 subclonal mutations were found in both SV22 and SV1 cells, of which 92 are more prevalent in SV22 than in SV1 and 11 are more prevalent in SV1 than in SV22 by at least a 3-fold difference (Table S4).

A detailed list of subclonal unique mutations, in which variant reads are at least two or greater, is presented in Table S5. Mutations were considered singlets if they have a variant read of 1, and these were excluded in Table S5. A total of 190 non-singlet subclonal mutations were observed only in SV22 cells but not in SV1 cells, whereas 122 non-singlet subclonal mutations were observed only in SV1 cells but not in SV22 cells (Table S5A, S5B). Among the identified subclonal mutations present in both SV22 and SV1 cells, 92 subclonal mutations were more highly mutated in SV22 cells than in SV1 cells by at least 3-fold (Table S5C). In contrast, only 11 subclonal mutations were more highly mutated in SV1 cells than in SV22 cells by at least 3-fold (Table S5D).

### *2.10. Independent DNA library preparation experiments of Duplex Sequencing produce reproducible results.*

Two independent DNA library preparation experiments were performed for SV1 cells. Consistent results of frequencies (Fig S4A), types (%) (Fig S4B), and sequence context spectra (%) of rare mutations (Fig S4C-D) were obtained from the two independent experiments. This validates the reproducibility of our Duplex Sequencing data.

## 3. Discussion

Ince et al (2007) reported that transformation of different normal human breast epithelial cell types led to distinct neoplastic phenotypes. This finding aligns with that the same oncogenes can have quite different phenotypic consequences depending on the cell origin [30,31]. Taken together, these suggest that the pre-existing differences in originator normal cell types significantly influence phenotypes of their transformed cells. Gordon et al (2014) demonstrated that immortalization of human fetal lung fibroblasts increased DNA methylation at gene promoters and caused large-scale changes in gene expression. While these studies [4,32] have examined tumorigenicity, histopathology, and metastatic behavior of human breast transformed cells or DNA methylation and gene expression of immortalized human fibroblasts, no study has examined the subclonal mutations of the whole mitochondrial genome for immortalized derivatives of phenotypically different normal cells.

In the current study, using Duplex Sequencing, we systemically examined both subclonal and clonal mutations of the whole mitochondrial genome for two types of immortalized HBECs [SV22 (non-stem) vs SV1 (stem)]. Our data indicate that both rare and subclonal mutation frequencies are significantly lower in SV1 cells than in SV22 cells (Figs 1, S2). Previously, we compared rare mutations of the whole mitochondrial genome between paired normal HBECs (non-stem vs stem cells) from three independent women and we observed that normal stem cells from two women exhibited significantly lower frequencies of rare mutations than the matching non-stem cells [22]. Our previous [22] and current findings suggest that the significantly low frequencies of rare and subclonal mutations of the whole mitochondria genome in stem cell-derived immortalized (SV1) cells can be attributed to mitochondrial mutations of their parental originator normal cell types. It supports the idea that distinct normal breast epithelial cell types lead to different mitochondrial mutagenesis in their immortalized/transformed cells.

Mechanisms for the lower frequencies of subclonal mutations in stem cell-derived immortalized cells (SV1) (Fig S2B) than in non-stem cell-derived immortalized cells (SV22) are unclear. A possible mechanism might be associated with lower levels of reactive oxygen species

(ROS). The lower ROS levels in stem cells could lead to reduced mtDNA damage and accumulation of subclonal mutations [33,34]. Stem cells might have more efficient systems of DNA repair and mitophagy for removing damaged mtDNA, therefore, lowering the number of mtDNA mutations generated [34-36]. These lower ROS levels and reduced mtDNA mutations of normal stem cells could be continually passed down to their immortalized cells.

Our results indicate that rare mutations are more prevalent in the non-coding RNA (ncRNA) regions (mt-RNR1; mt-RNR2; mt-tRNAs) than in the protein coding and control regions (Fig 3). Furthermore, the largest decrease in rare mutation burden in SV1 cells than in SV22 cells is observed in the ncRNA region (Fig 3). In ncRNA regions, the pathogenicity of mt-tRNA mutations can be predicted using the analysis tool MitoTIP (Fig 5). Our results indicate that the lower rare mutation burden of ncRNA genes in SV1 cells is accompanied by lower predicted pathogenicity of mt-tRNA mutations.

Furthermore, we demonstrate that four known pathological mt-tRNA gene mutations, m.5650 G>A (mt-tRNAAla) [37], m.5521 G>A (mt-tRNATrp) [38], m.5690 A>G (mt-tRNAAsn) [39], and m.1630 A>G (mt-tRNAVal) [40,41] were present in SV22 cells. However, no known pathogenic mutations were found in the mt-tRNA genes of SV1 cells. Pathogenicities of these four mt-tRNA mutations were confirmed by several studies [37-41]. All four confirmed pathogenic mutations are related to causing myopathy, a common mitochondrial disease effectuated by impacting protein translation or affecting tRNA stability [37-41]. Cytochrome c oxidase (COX)-deficiency was commonly reported with all four mutations [37-41]. The m.5650 G>A mutation causes a phenotype of pure myopathy [37] and disrupts interaction between mitochondrial alanyl-tRNA synthetase and the mt-tRNAAla aminoacyl acceptor stem. The m.5521 G>A mutation may impact protein translation [38]. The m.5690 A>G mutation manifests in clinical presentations of chronic progressive external ophthalmoplegia and ptosis [39]. The m.1630 A>G mutation impairs oxygen consumption, which affects the stability of mt-tRNAVal and reduces the levels of subunits of the electron transport chain [41]. Horvath et al (2009) reported that the m.1630 A>G mutation may cause mitochondrial neurogastrointestinal encephalomyopathy such as gastrointestinal dysmotility and cachexia [40].

In human breast cancer, tRNA mutations can cause significant consequences due to their important roles in protein translation [42]. Mt-tRNA point mutations are typically caused by the loss of mt-tRNA stability, which leads to defective mitochondrial translation and respiratory chain deficiency [39] through: aberrant processing of mRNA transcripts by RNases P and ZL, impaired post-transcriptional mt-tRNA modification such as specific base modifications, 3-end additions of -CCA sequence and mt-tRNA aminoacylation, as well as compromised interaction of mt-tRNA with both mtEF-Tu (mitochondrial elongation factor Tu) and the mitoribosome [43]. Our results suggest that stem cell-derived immortalized (SV1) cells might possess a more stable mitochondrial genome, exhibited by lower subclonal mutation burden and lower predicted pathogenicity than non-stem cell-derived immortalized (SV22) cells.

In summary, we examined whether different phenotypes of originator normal human breast epithelial cells (non-stem vs stem) can lead to different profiles of mitochondrial mutations for their immortalized derivatives. Our results indicate that the vast majority of mutations in the cells are rarely occurring mutations, which are not detectable by conventional DNA sequencing methods, but are accurately detectable by Duplex Sequencing. The most prevalent rare mutation types are C>T/G>A and A>G/T>C transitions. Immortalized stem cells (SV1) exhibit lower frequencies of rare mutations than immortalized non-stem cells (SV22). The reduced mitochondrial rare mutation burden of immortalized stem cells mainly occurs in ncRNA regions of the whole mtDNA, and these are accompanied by reduced predicted pathogenicity. Our findings suggest that phenotypes of originator parental normal cells significantly influence and direct mutational profiles of the whole mitochondrial genome in their immortalized derivatives. Our results have implications in investigating genetic changes of mitochondrial genomes acquired during cellular immortalization and in characterizing immortalized stem (vs non-stem) cells, which represent in vitro preneoplastic stages of breast carcinogenesis.

## 4. Materials and Methods

### *4.1. Development and culture of immortalized human breast epithelial cells.*

Immortalized cells (M13SV22, M13SV1) were derived from paired groups (non-stem vs stem) of normal human breast epithelial cells (HBECs) treated with SV40 large T-antigen (SV40-T). M13SV22 cells were derived from normal HBECs without stem cell features; M13SV1 cells were derived from normal HBECs with stem cell features. These immortalized cells were provided by Dr. Chia-Cheng Chang at Michigan State University (MSU) in East Lansing, MI, USA. A Material Transfer Agreement was approved by both MSU and University of Washington (UW). Development, characterization, and culture of normal HBEC and in vitro transformed HBECs were described previously [7,9-12,22,44-48].

### *4.2. DNA extraction, adapter synthesis, and DNA library preparation for Duplex Sequencing (DS) were carried out as we had described previously [22].*

### *4.3. Data processing.*

DS data were processed as described previously [16,17,22,49] with modifications. Previously, our in-house DS script processed and aligned two sequence read files of pair-end sequencing separately and independently before merging the two read files. For the current study, the script was modified to merge two sequence read files first then to process and align the merged file. This modification improved accuracy and efficiency of data processing. Sequence reads were aligned to the revised Cambridge Reference sequence (rCRS) reference genome (NC_012920) using BWA and the genome analysis toolkit (GATK) software as described previously [22]. BWA "mem" was used in replacement of BWA "aln". During processing, all datasets' reads were filtered using a mapping quality score of 40. Pileup-based variant calling used the default base quality score of 13. The first four bases at 5' and 3' ends of each sequence reads were clipped to eliminate potential artifactual variants commonly present at the ends of each read.

### *4.4. Data analysis for positions, frequencies, types, and context spectra of mutations and annotation of genetic variants.*

Genome positions with a DCS sequence read depth of 100 or greater were included for the data analysis. In addition, for comparing the mutations of SV22 and SV1 cells, only genome positions that had a minimum DCS read of 100 in matching positions of both samples were considered. The total number of DCS variant reads were divided by total number of DCS sequenced reads to calculate mutation frequencies for each sample. The total number of unique mutations (i.e. mutants were scored only once at each position of the genome regardless of number of variant reads observed in that position) were used for all other analyses, which includes fractions (%) of mutation types, mutation context spectra (%), and comparison of mutation positions. Genetic variant data from DS were annotated using Annotate Variation (ANNOVAR) bioinformatical software version 2017Jun01 (annovar.openbioinformatics.org) [50]. Annotations were added for each sequenced position of the whole mitochondrial genome.

### *4.5. Predicted pathogenicity.*

The MutPred program version 2.0 (http://mutpred.mutdb.org) [28] was performed to obtain pathogenicity scores of missense mutations in mt-protein coding regions as described previously [28,48]. The predicted pathogenicity scores for mt-tRNA mutations were analyzed using MitoTIP (August 2017 version) available *via* MITOMAP (www.mitomap.org). MitoTIP analysis [29] involves calculating pathogenicity scores for each possible variant through a summation of variant history, conservation score, position score, and secondary structure score. Using publicly available databases, an algorithm that estimates the importance of a position across mtDNA was generated

and used in scoring. Pathogenicity scores ranged from -5.9 to 21.8 and were assigned to each variant in mt-tRNAs by its position.

*4.6. Statistical analysis.*

Differences in mtDNA mutation frequencies and in the fraction (%) of mutation types between the two groups were analyzed by performing the prop.test for '2-sample equality of proportions with continuity correction' using an R program (version 3.4.4). To compare the MitoTIP predicted pathogenicity scores between the two groups, the Mann-Whitney U-test (Wilcoxon Rank-Sum test) was applied using Sigma Plot (version 12.0) (Systat Software, San Jose, CA). Differences between the two groups were considered significant if the p-value was less than 0.05.

**Supplementary Materials:** are available on line.

**Author Contributions:** Conceptualization, EHA; Data curation, EHA, SK, SSK, HEN, Formal analysis, SK, SSK, HEN, EHA; Funding acquisition, EHA; Investigation, EHA, SK, SSK, HEN; Methodology, EHA, HEN, SK; Project administration, EHA; Resources, EHA, SK, HEN; Software, HEN, SK; Supervision, EHA; Validation, EHA, SK, SSK, HEN; Visualization, SK, SSK, EHA, HEN; Writing - original draft, EHA; Writing - review & editing, EHA, HEN, SK, SSK.

**Funding:** The research was supported by grants from the National Institute of Environmental Health Sciences (NIEHS) P30 ES007033 sponsored-University of Washington (UW) Center for Exposures, Diseases, Genomics and Environment (EDGE) pilot grant (to EH Ahn), UW Office of Research Royalty Research Fund (to EH Ahn), and National Cancer Institute (NCI) P30 CA015704-39 Fred Hutchinson Cancer Research Center-UW Cancer Consortium Support Grant (to EH Ahn), NCI R21 CA220111 (to EH Ahn), and NCI of the National Institutes of Health (NIH) under award numbers P01 AG001751 and R33 CA181771 (to Lawrence A. Loeb). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

**Acknowledgments:** We thank Drs. Chia-Cheng Chang, Brad L. Upham, and James E. Trosko for obtaining immortalized human breast epithelial cells and Seung Hyuk (Tony) Lee for editing the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

| | |
|---|---|
| DCS | Duplex Consensus Sequence |
| DS | Duplex Sequencing |
| Mt | Mitochondrial |
| SV1 | Normal stem cell-derived immortalized human breast epithelial cells (HBECs); |
| SV22 | Normal non-stem cell-derived immortalized HBECs |

## References

1. Böcker, W.; Moll, R.; Poremba, C.; Holland, R.; Van Diest, P.J.; Dervan, P.; Bürger, H.; Wai, D.; Ina Diallo, R.; Brandt, B.; et al. Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: a new cell biological concept. *Lab. Invest.* **2002**, *82*, 737–746.
2. Dontu, G.; Al-Hajj, M.; Abdallah, W.M.; Clarke, M.F.; Wicha, M.S. Stem cells in normal breast development and breast cancer. *Cell Prolif.* **2003**, *36 Suppl 1*, 59–72.
3. Welm, B.; Behbod, F.; Goodell, M.A.; Rosen, J.M. Isolation and characterization of functional mammary gland stem cells. *Cell Prolif.* **2003**, *36 Suppl 1*, 17–32.
4. Ince, T.A.; Richardson, A.L.; Bell, G.W.; Saitoh, M.; Godar, S.; Karnoub, A.E.; Iglehart, J.D.; Weinberg, R.A. Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell* **2007**, *12*, 160–170.
5. Keller, P.J.; Arendt, L.M.; Skibinski, A.; Logvinenko, T.; Klebba, I.; Dong, S.; Smith, A.E.; Prat, A.; Perou, C.M.; Gilmore, H.; et al. Defining the cellular precursors to human breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 2772–2777.
6. Bu, W.; Liu, Z.; Jiang, W.; Nagi, C.; Huang, S.; Edwards, D.P.; Jo, E.; Mo, Q.; Creighton, C.J.; Hilsenbeck, S.G.; et al. Mammary Precancerous Stem and Non-Stem Cells Evolve into Cancers of Distinct Subtypes. *Cancer Res.* **2019**, *79*, 61–71.

7. Tai, M.-H.; Chang, C.-C.; Kiupel, M.; Webster, J.D.; Olson, L.K.; Trosko, J.E. Oct4 expression in adult human stem cells: evidence in support of the stem cell theory of carcinogenesis. *Carcinogenesis* **2005**, *26*, 495–502.
8. Kang, K.S.; Morita, I.; Cruz, A.; Jeon, Y.J.; Trosko, J.E.; Chang, C.C. Expression of estrogen receptors in a normal human breast epithelial cell type with luminal and stem cell characteristics and its neoplastically transformed cell lines. *Carcinogenesis* **1997**, *18*, 251–257.
9. Kao, C.Y.; Nomata, K.; Oakley, C.S.; Welsch, C.W.; Chang, C.C. Two types of normal human breast epithelial cells derived from reduction mammoplasty: phenotypic characterization and response to SV40 transfection. *Carcinogenesis* **1995**, *16*, 531–538.
10. Kao, C.-Y.; Oakley, C.S.; Welsch, C.W.; Chang, C.-C. Growth requirements and neoplastic transformation of two types of normal human breast epithelial cells derived from reduction mammoplasty. *In Vitro Cellular & Developmental Biology - Animal* **1997**, *33*, 282–288.
11. Chang, C.C.; Sun, W.; Cruz, A.; Saitoh, M.; Tai, M.H.; Trosko, J.E. A human breast epithelial cell type with stem cell characteristics as target cells for carcinogenesis. *Radiat. Res.* **2001**, *155*, 201–207.
12. Ahn, E.H.; Chang, C.-C.; Schroeder, J.J. Evaluation of sphinganine and sphingosine as human breast cancer chemotherapeutic and chemopreventive agents. *Exp. Biol. Med. (Maywood)* **2006**, *231*, 1664–1672.
13. Ahuja, D.; Sáenz-Robles, M.T.; Pipas, J.M. SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene* **2005**, *24*, 7729–7745.
14. Pipas, J.M. SV40: Cell transformation and tumorigenesis. *Virology* **2009**, *384*, 294–303.
15. Sun, W.; Kang, K.S.; Morita, I.; Trosko, J.E.; Chang, C.C. High susceptibility of a human breast epithelial cell type with stem cell characteristics to telomerase activation and immortalization. *Cancer Res.* **1999**, *59*, 6118–6123.
16. Schmitt, M.W.; Kennedy, S.R.; Salk, J.J.; Fox, E.J.; Hiatt, J.B.; Loeb, L.A. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 14508–14513.
17. Kennedy, S.R.; Schmitt, M.W.; Fox, E.J.; Kohrn, B.F.; Salk, J.J.; Ahn, E.H.; Prindle, M.J.; Kuong, K.J.; Shen, J.-C.; Risques, R.-A.; et al. Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat Protoc* **2014**, *9*, 2586–2606.
18. Schmitt, M.W.; Fox, E.J.; Prindle, M.J.; Reid-Bayliss, K.S.; True, L.D.; Radich, J.P.; Loeb, L.A. Sequencing small genomic targets with high efficiency and extreme accuracy. *Nat. Methods* **2015**, *12*, 423–425.
19. Ahn, E.H.; Lee, S.H. Detection of Low-Frequency Mutations and Identification of Heat-Induced Artifactual Mutations Using Duplex Sequencing. *Int J Mol Sci* **2019**, *20*.
20. Lou, D.I.; Hussmann, J.A.; McBee, R.M.; Acevedo, A.; Andino, R.; Press, W.H.; Sawyer, S.L. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 19872–19877.
21. Fox, E.J.; Reid-Bayliss, K.S.; Emond, M.J.; Loeb, L.A. Accuracy of Next Generation Sequencing Platforms. *Next Gener Seq Appl* **2014**, *1*.
22. Ahn, E.H.; Hirohata, K.; Kohrn, B.F.; Fox, E.J.; Chang, C.-C.; Loeb, L.A. Detection of Ultra-Rare Mitochondrial Mutations in Breast Stem Cells by Duplex Sequencing. *PLoS ONE* **2015**, *10*, e0136216.
23. Taylor, R.W.; Turnbull, D.M. Mitochondrial DNA mutations in human disease. *Nat. Rev. Genet.* **2005**, *6*, 389–402.
24. Chan, D.C. Mitochondria: dynamic organelles in disease, aging, and development. *Cell* **2006**, *125*, 1241–1252.
25. Salk, J.J.; Schmitt, M.W.; Loeb, L.A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat. Rev. Genet.* **2018**, *19*, 269–285.
26. Strachan, T.; Read, A.P. *Human molecular genetics*; 2nd ed.; Wiley: New York, 1999; ISBN 978-0-471-33061-5.
27. Govatati, S.; Deenadayal, M.; Shivaji, S.; Bhanoori, M. Mitochondrial displacement loop alterations are associated with endometriosis. *Fertil. Steril.* **2013**, *99*, 1980-1986.e9.
28. Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K.A.; Lin, G.N.; Nam, H.-J.; Mort, M.; Cooper, D.N.; Sebat, J.; Iakoucheva, L.M.; et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv* **2017**.
29. Sonney, S.; Leipzig, J.; Lott, M.T.; Zhang, S.; Procaccio, V.; Wallace, D.C.; Sondheimer, N. Predicting the pathogenicity of novel variants in mitochondrial tRNA with MitoTIP. *PLoS Comput. Biol.* **2017**, *13*, e1005867.
30. Vogelstein, B.; Kinzler, K.W. Cancer genes and the pathways they control. *Nat. Med.* **2004**, *10*, 789–799.

31. Gupta, G.P.; Massagué, J. Cancer metastasis: building a framework. *Cell* **2006**, *127*, 679–695.
32. Gordon, K.; Clouaire, T.; Bao, X.X.; Kemp, S.E.; Xenophontos, M.; de Las Heras, J.I.; Stancheva, I. Immortality, but not oncogenic transformation, of primary human cells leads to epigenetic reprogramming of DNA methylation and gene expression. *Nucleic Acids Res.* **2014**, *42*, 3529–3541.
33. Saretzki, G.; Walter, T.; Atkinson, S.; Passos, J.F.; Bareth, B.; Keith, W.N.; Stewart, R.; Hoare, S.; Stojkovic, M.; Armstrong, L.; et al. Downregulation of multiple stress defense mechanisms during differentiation of human embryonic stem cells. *Stem Cells* **2008**, *26*, 455–464.
34. Frosina, G. The bright and the dark sides of DNA repair in stem cells. *J. Biomed. Biotechnol.* **2010**, *2010*, 845396.
35. Guan, J.-L.; Simon, A.K.; Prescott, M.; Menendez, J.A.; Liu, F.; Wang, F.; Wang, C.; Wolvetang, E.; Vazquez-Martin, A.; Zhang, J. Autophagy in stem cells. *Autophagy* **2013**, *9*, 830–849.
36. Joshi, A.; Kundu, M. Mitophagy in hematopoietic stem cells: the case for exploration. *Autophagy* **2013**, *9*, 1737–1749.
37. McFarland, R.; Swalwell, H.; Blakely, E.L.; He, L.; Groen, E.J.; Turnbull, D.M.; Bushby, K.M.; Taylor, R.W. The m.5650G>A mitochondrial tRNAAla mutation is pathogenic and causes a phenotype of pure myopathy. *Neuromuscul. Disord.* **2008**, *18*, 63–67.
38. Porcelli, A.M.; Ghelli, A.; Ceccarelli, C.; Lang, M.; Cenacchi, G.; Capristo, M.; Pennisi, L.F.; Morra, I.; Ciccarelli, E.; Melcarne, A.; et al. The genetic and metabolic signature of oncocytic transformation implicates HIF1alpha destabilization. *Hum. Mol. Genet.* **2010**, *19*, 1019–1032.
39. Blakely, E.L.; Yarham, J.W.; Alston, C.L.; Craig, K.; Poulton, J.; Brierley, C.; Park, S.-M.; Dean, A.; Xuereb, J.H.; Anderson, K.N.; et al. Pathogenic mitochondrial tRNA point mutations: nine novel mutations affirm their importance as a cause of mitochondrial disease. *Hum. Mutat.* **2013**, *34*, 1260–1268.
40. Horváth, R.; Bender, A.; Abicht, A.; Holinski-Feder, E.; Czermin, B.; Trips, T.; Schneiderat, P.; Lochmüller, H.; Klopstock, T. Heteroplasmic mutation in the anticodon-stem of mitochondrial tRNA(Val) causing MNGIE-like gastrointestinal dysmotility and cachexia. *J. Neurol.* **2009**, *256*, 810–815.
41. Glatz, C.; D'Aco, K.; Smith, S.; Sondheimer, N. Mutation in the mitochondrial tRNA(Val) causes mitochondrial encephalopathy, lactic acidosis and stroke-like episodes. *Mitochondrion* **2011**, *11*, 615–619.
42. Meng, X.-L.; Meng, H.; Zhang, W.; Qin, Y.-H.; Zhao, N.-M. The role of mitochondrial tRNA variants in female breast cancer. *Mitochondrial DNA A DNA Mapp Seq Anal* **2016**, *27*, 3199–3201.
43. Levinger, L.; Oestreich, I.; Florentz, C.; Mörl, M. A pathogenesis-associated mutation in human mitochondrial tRNALeu(UUR) leads to reduced 3'-end processing and CCA addition. *J. Mol. Biol.* **2004**, *337*, 535–544.
44. Kang, K.S.; Sun, W.; Nomata, K.; Morita, I.; Cruz, A.; Liu, C.J.; Trosko, J.E.; Chang, C.C. Involvement of tyrosine phosphorylation of p185(c-erbB2/neu) in tumorigenicity induced by X-rays and the neu oncogene in human breast epithelial cells. *Mol. Carcinog.* **1998**, *21*, 225–233.
45. Park, J.-S.; Noh, D.-Y.; Kim, S.-H.; Kim, S.-H.; Kong, G.; Chang, C.-C.; Lee, Y.-S.; Trosko, J.E.; Kang, K.-S. Gene expression analysis in SV40-immortalized human breast luminal epithelial cells with stem cell characteristics using a cDNA microarray. *Int. J. Oncol.* **2004**, *24*, 1545–1558.
46. Ahn, E.H.; Chang, C.-C.; Talmage, D.A. Loss of anti-proliferative effect of all-trans retinoic acid in advanced stage of breast carcinogenesis. *Anticancer Res.* **2009**, *29*, 2899–2904.
47. Wang, K.-H.; Kao, A.-P.; Chang, C.-C.; Lee, J.-N.; Hou, M.-F.; Long, C.-Y.; Chen, H.-S.; Tsai, E.-M. Increasing CD44+/CD24(-) tumor stem cells, and upregulation of COX-2 and HDAC6, as major functions of HER2 in breast tumorigenesis. *Mol. Cancer* **2010**, *9*, 288.
48. Ahn, E.H.; Lee, S.H.; Kim, J.Y.; Chang, C.-C.; Loeb, L.A. Decreased Mitochondrial Mutagenesis during Transformation of Human Breast Stem Cells into Tumorigenic Cells. *Cancer Res.* **2016**, *76*, 4569–4578.
49. Pickrell, A.M.; Huang, C.-H.; Kennedy, S.R.; Ordureau, A.; Sideris, D.P.; Hoekstra, J.G.; Harper, J.W.; Youle, R.J. Endogenous Parkin Preserves Dopaminergic Substantia Nigral Neurons following Mitochondrial DNA Mutagenic Stress. *Neuron* **2015**, *87*, 371–381.
50. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164.