Kim, J Mol Biol Methods 2020, 3:1
DOI: 10.37532/jmbm.2020.3(1).107

Journal of Molecular Biology and Methods

Research Article

A SCITECHNOL JOURNAL

# Post-zygotic Mosaic Mutation in Normal Tissue from Breast Cancer Patient

**Ryong Nam Kim***

*Seoul National University Bio-MAX/NBIO, Seoul, Korea*

***Corresponding author:** Ryong Nam Kim, Seoul National University Bio-MAX/NBIO, Seoul, South Korea; E-mail: ryongnamkim@gmail.com

**Received date:** April 7, 2020; **Accepted date:** April 22, 2020; **Published date:** April 30, 2020

## Abstract

Even though numerous previous investigations had shed fresh light on somatic driver mutations in cancer tissues, the mutation-driven malignant transformation mechanism from normal to cancerous tissues remains still mysterious. In this study, we performed whole exome analysis of paired normal and cancer samples from 12 breast cancer patients in order to elucidate the post-zygotic mosaic mutation that might predispose to breast carcinogenesis. We found a post-zygotic mosaic mutation PIK3CA p.F1002C with 2% variant allele fraction (VAF) in normal tissue, whose respective VAF in a matched breast cancer tissue, had increased by 20.6%. Such an expansion of the variant allele fraction in the matched cancer tissue may implicate the mosaic mutation in association with the causation underlying the breast carcinogenesis. The post-zygotic mosaic mutation is estimated to be deleterious by well-established variant annotation software programs, SIFT_pred, Polyphen2_HDIV_pred, Polyphen2_HVAR_pred, LRT_pred, MutationTaster_pred, PROVEAN_pred, fathmm. MKL_coding_pred, MetaSVM_pred, and MetaLR_pred. In addition, we discovered 61 deleterious and pathogenic mutations, including 22 stop-gain, 12 splicing site, 13 frame shift and 7 non synonymous mutations, in those patients. By performing mutational signature analysis, we identified three mutational signatures underlying breast carcinogenesis, including APOBEC cytidinedeaminase and defective DNA mismatch repair. Taken together, these results suggest that, in addition to the somatic driver mutations, post-zygotic mosaic mutation may be a critical target that is worth deserving prior attention in ascertaining the causation underlying breast carcinogenesis in the upcoming future.

**Keywords:** Post-zygotic; Mosaic mutation; Breast cancer

## Introduction

Biological mosaicism had been originally given its name from the fact that it is morphologically analogous to the intricate images created by craftsmen using small pieces of colored tiles or glass. It means an individual who has developed from a single fertilized egg and has two or more populations of cells with different genotypes [1]. The generation of genetically distinct cells from a single zygote can be caused by post zygotic de novo mutational events as the cause of mosaicism, and such mutations can result in sporadic disease [2]. De novo mutational events can also occur pre-zygotically [2]. Beyond the above-mentioned general descriptions, there are more specific types of mosaicism, which can be classified depending on which parts of the body have the variant cells and the potential for mutational transmission to offspring: germline mosaicism, somatic mosaicism and gonosomal Mosaicism [3].

Post zygotic mutations that can be classified as somatic mosaic mutations arise at the first division of the zygote or later in different somatic cell lineages [4]. The post zygotic mosaic mutations are possible drivers underlying diverse diseases including cancers [4]. These postzygotic mutations have the potential to lead to a broad range of cellular phenotypes and can thereby affect the relative fitness of the cells [4]. Most mosaic mutations can be expected to be neutral or decrease the fitness of the affected cells relative to wild-type cells [4]. However, some mosaic mutations will lead to a proliferative advantage and clonal expansion in the affected lineages, tissues or organs and could predispose to the development of neoplasia and/or cause dysfunctions [4]. Following the expansion, the prevalence of the clone increases, and it becomes readily detectable in analyses of bulk DNA as detectable Mosaicism [4]. By contrast, neutral mutations or mutations with mild negative effects will remain at low frequencies as cryptic Mosaicism [4].

Recent advancements in next generation and targeted deep sequencing technologies have accelerated the discoveries of postzygotic mosaic mutations in blood and normal tissues from cancer patients [5,6]. Especially, some investigators found that the postzygotic mosaic mutations are associated with the causation of carcinogenesis in patients [4].

In this study, we performed whole exome analysis of matched normal and cancer samples from 12 breast cancer patients, and found causative deleterious mutations, including one postzygotic mosaic mutation, associated with breast cancer.

## Materials and Methods

### Data collection

We had obtained the raw data (fastq files) for the matched tumor and normal samples of the 12 breast cancer patients from a data repository provided by the previous investigation [7].

### Identification of somatic mutations

While the previous investigation [7] had used an old non-standard program, samtools software, for the variant calling, we have used GATK (Genome Analysis Toolkit), an advanced standard method, to overcome the liability of the approach in the previous investigation.

Raw reads in the paired-end fastq files were checked with Fast QC (v0.11.4) and low-quality reads were removed. High-quality reads in the paired-end fastq files were aligned to a human reference genome hg19 using BWA-MEM algorithm (v0.7.12) [19], and the resulting SAM files were converted to BAM files by samtools (v0.1.19) [19]. Duplicates were marked and removed from sorted BAM files by Picard (v1.115). The aligned reads were realigned and recalibrated by using Indel Realigner and Base Recalibrator commands, respectively, in GATK suite (v3.6) (Broad Institute, Cambridge, MA) [20]. Somatic single nucleotide variants were called by using MuTect2 in the GATK suite. Somatic indel variants were called by using Indelocator software (https://**www.broadinstitute.org/cancer/cga/**indelocator).

All articles published in Journal of Molecular Biology and Methods are the property of SciTechnol and is protected by copyright laws. Copyright © 2020, SciTechnol, All Rights Reserved.

*doi: 10.37532/jmbm.2020.3(1).107*

In order to identify the post-zygotic mosaic mutation, we compared the variant allele fractions of each mutation between normal and matched tumor tissues, and had chosen the candidate mutation, whose VAF in normal tissue is far less than that in the matched tumor tissue. In addition, the candidate mosaic mutation should be evaluated as deleterious by the variant annotation programs [10-15], and their allele frequency should be less than 0.01 in ExAC (Exome Aggregation Consortium) database [21]. In case of other somatic mutations, they should not be present in matched normal tissues.

Analysis of mutational signatures

In order to analyze the mutational signatures underlying the breast carcinogenesis, we had used R programming software " decompTumor2Sig " [22-24]. After analyzing the mutational signatures in those breast cancer patients, we had assessed how similar they are to the canonical COSMIC mutational signatures mentioned in the previous investigation [16].

## Results

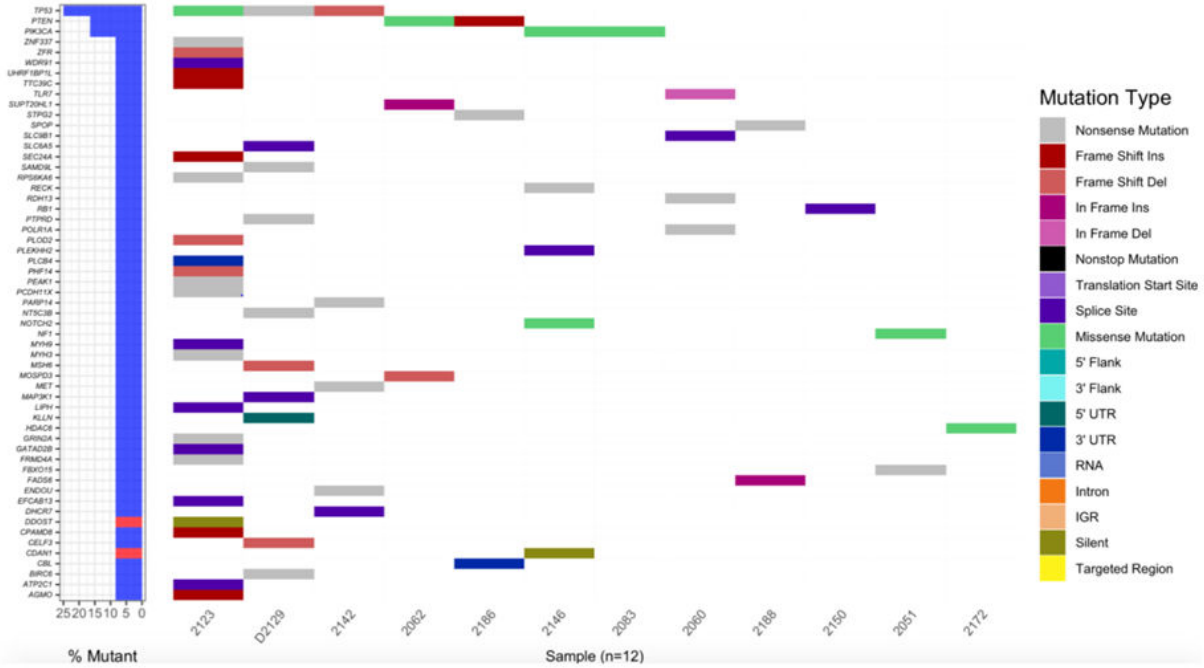


**Figure 1:** Mutational landscape showing types of deleterious mutations occurring in each of those breast cancer patients.

| Sample | Gene | Mutation Type | Change |
|---|---|---|---|
| 2051 | FBXO15 | Stop gain | p.C343X |
| 2051 | NF1 | Non synonymous SNV | p.D2674V |
| 2060 | POLR1A | Stop gain | p.E741X |
| 2060 | RDH13 | Stop gain | p.Y129X |
| 2060 | TLR7 | Non frame shift substitution | c.3147_3150C |
| 2060 | SLC9B1 | splicing | c.829+1G>T |
| 2062 | PTEN | Non synonymous SNV | p.D265Y |
| 2062 | SUPT20HL1 | Non frame shift substitution | c.1540_1540delinsGCTGCTGCTC |
| 2062 | MOSPD3 | Frame shift substitution | c.609_617C |
| 2083 | PIK3CA | Non synonymous SNV | p.E545K |
| 2123 | DDOST | synonymous SNV | p.C145C |

*doi: 10.37532/jmbm.2020.3(1).107*

| | | | |
|---|---|---|---|
| 2123 | FRMD4A | Stop gain | p.K88X |
| 2123 | PEAK1 | Stop gain | p.R1266X |
| 2123 | GRIN2A | Stop gain | p.R1206X |
| 2123 | MYH3 | Stop gain | p.Q721X |
| 2123 | ZNF337 | Stop gain | p.R523X |
| 2123 | RPS6KA6 | Stop gain | p.R462X |
| 2123 | PCDH11X | Stop gain | p.E637X |
| 2123 | TP53 | Non synonymous SNV | p.L291R |
| 2123 | TTC39C | Frame shift substitution | c.58_58delinsGA |
| 2123 | PLOD2 | Frame shift substitution | c.384_385T |
| 2123 | PHF14 | Frame shift substitution | c.1568_1570A |
| 2123 | UHRF1BP1L | Frame shift substitution | c.3216_3216delinsCA |
| 2123 | AGMO | Frame shift substitution | c.618_618delinsAT |
| 2123 | ZFR | Frame shift substitution | c.1074_1075G |
| 2123 | SEC24A | Frame shift substitution | c.2990_2990delinsGA |
| 2123 | CPAMD8 | Frame shift substitution | c.3154_3154delinsGC |
| 2123 | PLCB4 | UTR3 | c.*1011A>G |
| 2123 | GATAD2B | splicing | c.465+2T>C |
| 2123 | ATP2C1 | splicing | c.422+2T>C |
| 2123 | LIPH | splicing | c.628+2T>C |
| 2123 | WDR91 | splicing | c.1660-2A>G |
| 2123 | EFCAB13 | splicing | c.1085-2A>G |
| 2123 | MYH9 | splicing | c.3942+2T>C |
| 2142 | PARP14 | Stop gain | p.R1694X |
| 2142 | MET | Stop gain | p.S691X |
| 2142 | ENDOU | Stop gain | p.Y405X |
| 2142 | TP53 | Frame shift substitution | c.883_890A |
| 2146 | CDAN1 | synonymous SNV | p.A787A |
| 2146 | RECK | Stop gain | p.S542X |
| 2146 | NOTCH2 | Non synonymous SNV | p.A1721G |
| 2146 | PIK3CA | Non synonymous SNV | p.F1002C |
| 2146 | PLEKHH2 | splicing | c.3942-1G>A |
| 2150 | RB1 | splicing | c.2663+1G>C |
| 2172 | HDAC6 | Non synonymous SNV | p.R773H |
| 2186 | STPG2 | Stop gain | p.S211X |
| 2186 | PTEN | Frame shift substitution | c.1473_1473delinsTA |

*doi: 10.37532/jmbm.2020.3(1).107*

| | | | |
|---|---|---|---|
| 2186 | CBL | UTR3 | c.*6362G>A |
| 2188 | SPOP | Stop gain | p.W36X |
| 2188 | FADS6 | Non frame shift substitution | c.45_45delinsTACGGAGCCCATGGAACCG |
| D2129 | TP53 | Stop gain | p.R303X |
| D2129 | BIRC6 | Stop gain | p.R870X |
| D2129 | SAMD9L | Stop gain | p.E394X |
| D2129 | PTPRD | Stop gain | p.E776X |
| D2129 | NT5C3B | Stop gain | p.R55X |
| D2129 | CELF3 | Frame shift substitution | c.313_314A |
| D2129 | MSH6 | Frame shift substitution | c.723_724A |
| D2129 | KLLN | UTR5 | c.-774G>A |
| D2129 | MAP3K1 | splicing | c.835-2A>G |
| D2129 | SLC6A5 | splicing | c.811+1G>T |

**Table 1:** Deleterious and pathogenic mutations discovered in those breast cancer patients.

Whole exome analysis of matched normal and cancer samples from 12 breast cancer patients. We reanalyzed whole exome next-generation sequencing raw FastQ dataset of matched normal and cancer samples from 12 breast cancer patients in the public data repository from the previously published report [5-7]. Among the 12 breast cancer patients, we found 60 deleterious and pathogenic mutations, including 21 stop-gain, 7 frame shift deletion, 6 frame shift insertion, 7 non synonymous, 3 non-frame shift in del, 2 synonymous, 11 splicing, and 3 untranslated region (3'UTR and 5'UTR) mutations (Figure 1 and Table 1). Especially, carcinogenesis-associated genes, *TP53, PTEN, RB1, NF1, MSH6* and *PIK3CA*, harbored deleterious mutations in those patients. In addition, TP53 p.R303X and PIK3CA p.E545K in two breast cancer patients, respectively, had been validated as pathogenic mutations in previous clinical investigations[8,9].

Next, we performed analysis of the postzygotic mosaic mutation present in normal tissue from breast cancer patient. In this analysis, we had focused on the post zygoticmosaic mutation, whose variant allele fraction (VAF) in the breast cancer tissue had increased compared with that in the matched normal tissue. We found the postzygotic mosaic mutation, PIK3CA p.F1002C, in the matched normal and cancer samples from one breast cancer patient (Figure 2). The postzygotic mosaic mutation PIK3CA p.F1002C (VAFs in normal and cancer tissues: 2.13% and 20.6%, respectively) resides in the PIK3CA protein's functional PI3_PI4 kinase domain. This mosaic mutation caused the change of the residue F to residue C at the 1002th amino acid site that had been evolutionarily conserved in Human, Rhesus, Mouse, Dog, Elephant, Opossum, Chicken, X. tropicalis, Zebrafish and so on (as annotated in UCSC Genome Browser). In addition, PIK3CA p.F1002C is evaluated as deleterious mutation by mutation annotation softwares, SIFT_pred, Polyphen2_HDIV_pred, Polyphen2_HVAR_pred, LRT_pred, MutationTaster_pred, PROVEAN_pred, fathmm.MKL_coding_pred, MetaSVM_pred, and MetaLR_pred [10-15]. In a uterine corpus endometrial carcinoma patient in TCGA (the cancer genome atlas) data, PIK3CA p.F1002L had been discovered as a deleterious mutation at the same amino acid position, suggesting that the amino acid change at the position in the PIK3CA protein sequence might predispose to the carcinogenesis.

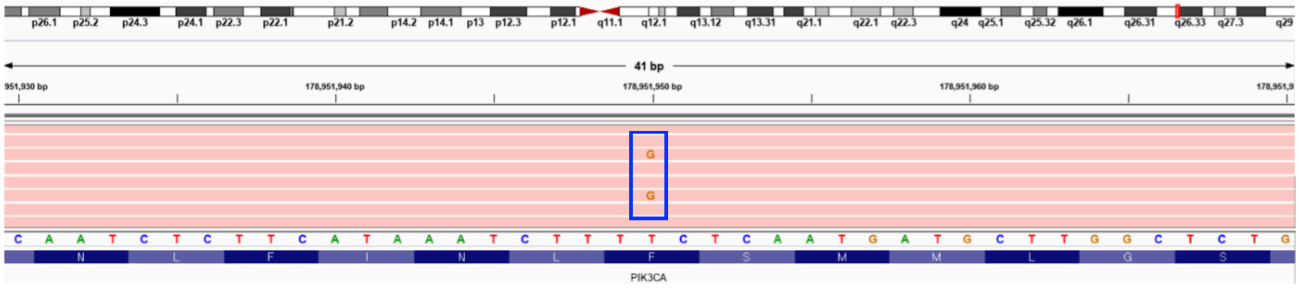


**Figure 2:** The post-zygotic mosaic mutation occurring in normal tissue from breast cancer patient. The picture is snapshot from visualized plots of the BAM files using IGV.

*doi: 10.37532/jmbm.2020.3(1).107*

## Mutational signatures

Among those 12 breast cancer patients, we discovered three major mutational signatures (signatures 1, 2 and 3), which match the known COSMIC_2, COSMIC_5 and COSMIC_26 mutational signatures, respectively, in the previous investigation (Figure 3 A and B)[16]. The COSMIC_2 mutational signature had been known to be accidently caused by APOBEC cytidinedeaminase that canonically converts cytosine to uracil during RNA editing and the restriction of retrovirus or retrotransposon, and this mutational signature is functionally associated with development of diverse cancer types including breast cancer [16,17]. Until now, little is known about the cause underlying the COSMIC_5 mutational signature. In case of the COSMIC_26 mutational signature, this mutational pattern is due to defective DNA mismatch repair [18], which is the causation underlying breast carcinogenesis [16].

As shown in Figure 4, the breast cancer patients 2015, 2060, 2062, 2083, 2123, 2142, 2146, 2150, 2172, 2183, 2186, 2188, and D2129 had signature 2, signature 1, signature 2, signature 1, signature3, signature 2, signature 2, signature 1, signature 2, signature 1, signature 2, signature 1, and signature 2 as their dominant mutational signature patterns, respectively. This result suggests that, although the three different mutational signatures occur in each of those patients in different proportions, a major signature might vary for each patient.

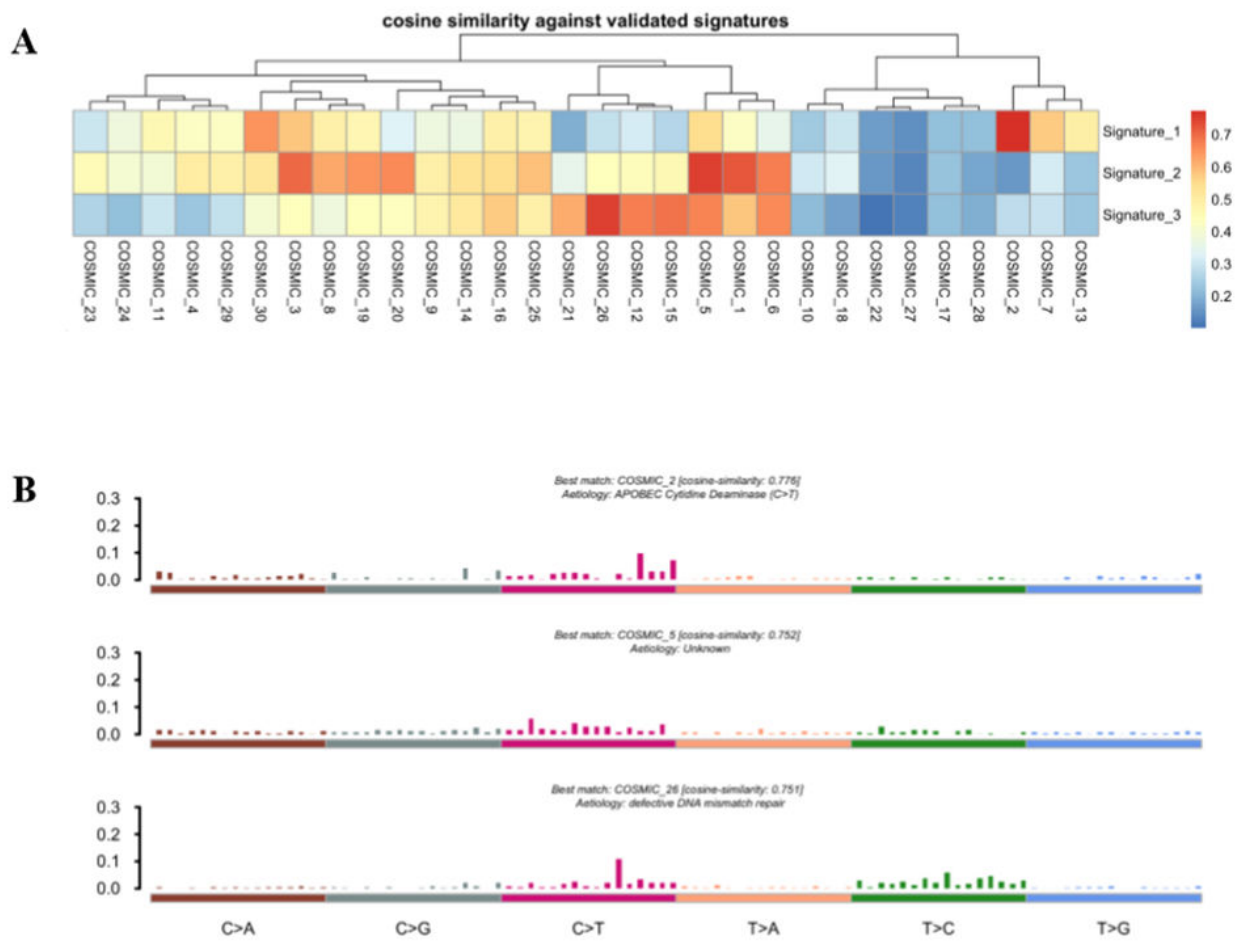


**Figure 3:** Mutational signatures in those breast cancer patients. A: Similarities of the three mutational signatures to the previously validated signatures. For example, the densest red color for each signature means that a given signature is most similar to the previously validated COSMIC signature corresponding to the densest red square. B: The known etiologies of the best matched COSMIC signatures, and the relative proportions of nucleotide substitutions with possible cases of adjacent nucleotides.

*doi: 10.37532/jmbm.2020.3(1).107*

Relative Proportion of Three Mutational Signatures in 12 Breast Cancer Patients

| | 2051 | 2060 | 2062 | 2083 | 2123 | 2142 | 2146 | 2150 | 2172 | 2183 | 2186 | 2188 | D2129 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Signature_1 | 0.14303990612736 | 0.97568190131493 | 0.118003863455893 | 0.794354650233176 | 0.0145821042551442 | 0.00159409959099449 | 0.00048969082163184 7 | 0.549135603108231 | 0.302678270656699 | 0.742046156809836 | 0.0759996136076657 | 0.546159353177794 | 0.383608482506016 |
| Signature_2 | 0.583398469286934 | 0.0243180977958 24 | 0.683384419138356 | 0.0588526695941633 | 0.00049586930551 2025 | 0.979089583897491 | 0.632230854587819 | 0.323272006454373 | 0.549702926163458 | 0.203191731728281 | 0.589195998974678 | 0.299585615385074 | 0.542405866186707 |
| Signature_3 | 0.273561624585706 | 8.89246143432963E-10 | 0.198611717405751 | 0.14679268017266 | 0.98492202643934 4 | 0.0193163165115147 | 0.367279454590549 | 0.127592390437395 | 0.147618803179844 | 0.0547621114618827 | 0.334804387417656 | 0.154255031437132 | 0.0739856513072764 |

**Figure 4:** Relative proportion of the three mutational signatures in each of the 12 breast cancer patients.

## Discussion

In this investigation, we discovered a post-zygotic mosaic mutation, PIK3CA p.F1002C, as well as deleterious and pathogenic mutations and mutational signatures in 12 breast cancer patients. Recent study suggested that mosaic copy number alterations existed in blood and normal tissues adjacent to tumor tissues from patients with diverse cancer types [5]. Intriguingly, the authors identified that the haplotype-phased mosaic allelic fractions of the copy number alterations are much greater in tumor tissues compared with matched blood and normal tissues in some patients. In addition, the genomic regions with the mosaic allelic copy number alterations harbored canonical cancer-associated genes, suggesting that such mosaic copy number alterations might contribute to the causation of carcinogenesis. We convincingly suggest that mosaic mutations, whose variant allele fractions in normal tissue are far less than in matched tumor tissue and whose functional impacts are deleterious, deserve prior attention in making decision for prioritizing variants and ascertaining drug targets for diagnosing and treating cancer patients in the upcoming years.

## Acknowledgement

This investigation has been supported by a research grant (NRF-2017R1D1A1B04033856 (Ryong Nam Kim)) funded by the Ministry of Education and the National Research Foundation in Korea.

## References

1. Biesecker LG, Spinner NB (2013) A genomic view of mosaicism and human disease. Nat Rev Genet 14: 307-320.
2. Veltman JA, Brunner HG (2012) De novo mutations in human genetic disease. Nat Rev Genet13: 565-575.
3. Happle R (1993) Mosaicism in human skin, Understanding the patterns and mechanisms. Arch Dermatol 129: 1460-1470.
4. Forsberg LA, Gisselsson D, Dumanski JP (2017)Mosaicism in health and disease - clones picking up speed. Nat Rev Genet 18: 128-142.
5. Jakubek YA, Chang K, Sivakumar S, Yu Y, Giordano MR, et al. (2020) Large-scale analysis of acquired chromosomal alterations in non-tumor samples from patients with cancer. Nat Biotechnol 38: 90-96.
6. Dou Y, Kwon M, Rodin RE, Cortes-Ciriano I, Doan R, et al. (2020) Accurate detection of mosaic variants in sequencing data without matched controls. Nat Biotechnol 38: 314-319.
7. Lee JH, Zhao XM, Yoon I, Lee JY, Kwon NH, et al. (2016) Integrative analysis of mutational and transcriptional profiles reveals driver mutations of metastatic breast cancers. Cell Discov 2: 16025.
8. Hao JJ, Lin DC, Dinh HQ,Mayakonda A, Jiang YY, et al. (2016) Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma. Nat Genet 48: 1500-1507.
9. Leontiadou H, Galdadas I, Athanasiou C, Cournia Z (2018) Insights into the mechanism of the PIK3CA E545K activating mutation using MD simulations. Sci Rep-Uk 8.
10. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073-1082.
11. Adzhubei IA Schmidt S, Peshkin, L,Ramensky VE, Gerasimova Aet al. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248-249.
12. Chun S, Fay JC (2009) Identification of deleterious mutations within three human genomes. Genome Research 19: 1553-1561.
13. Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nature Methods 11: 361-362.
14. Choi Y, SimsGE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. Plos One 7.
15. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E,et al. (2015)Comparison and integration of deleteriousness prediction methods for non synonymous SNVs in whole exome sequencing studies. Hum Mol Genet 24: 2125-2137.
16. Nik-ZainalS, Davies H, Staaf J, Ramakrishna M,Glodzik D, et al. (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534: 47-54.
17. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, et al. (2013) An APOBEC cytidinedeaminase mutagenesis pattern is widespread in human cancers. Nat Genet 45: 970-976.
18. Steele CD, Tarabichi M, Oukrif D, Webster AP, Ye H, et al. (2019) Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways. Cancer Cell 35: 441-456.
19. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.
20. Van der AuweraGA, Carneiro MO, Hartl C, Poplin R, Del Angel G, et al. (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. CurrProtoc Bioinformatics 43: 111011-111033.
21. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536: 285-291.
22. Kruger S, Piro RM (2019)decompTumor2Sig: identification of mutational signatures active in individual tumors. BMC Bioinformatics 20:152.

*doi: 10.37532/jmbm.2020.3(1).107*

23. Thorvaldsdottir H, Robinson JT,Mesirov JP (2013) Integrative Genomics Viewer (IGV): high- performance genomics data visualization and exploration. Brief Bioinform 14: 178-192.

24. Kim RY (2019) mutational mosaicism in breast cancer. Journal of pharmaceutical sciences & emerging drugs 7.