Bioinformatic Analysis of Frequently Reported Somatically Mutated Genes of Triple-negative Breast Cancer and Comparison with Top Mutated Genes from the BRCA-UK Project

Allan L. Hilario1,*, John Robert C. Medina2, Jose B. Nevado, Jr.1,3, and Catherine Lynn T. Silao3,4

1Department of Biochemistry and Molecular Biology, College of Medicine
2Department of Biostatistics and Epidemiology, College of Public Health
3Institute of Human Genetics; 4Department of Pediatrics, College of Medicine
University of the Philippines, Ermita, Manila 1000 Philippines

Triple-negative breast cancer (TNBC) is a breast cancer subtype with uncertain causation except for some subtypes with known genetic etiology. It is associated with heritability and is considered a genetically heterogeneous disease. TNBC is associated with a poorer prognosis, is more aggressive in clinical course, has poorer treatment response, and has poorer survival. This study was done to review the differentially somatically mutated genes in published literature according to the frequency of articles cited, GO and STRING analyses performed, and comparison done with the genetic profile of the BRCA-UK Project TNBC cohort. A literature search was done in PubMed using specified mesh words. The top 54 frequently reported genes were characterized using GO and STRING protein-protein interaction (PPI) analyses. The set of genes cited in literature was compared with the top 50 somatically mutated genes in the BRCA-UK Project cohort. The top ten most frequently reported mutated genes in the literature were TP53, PIK3CA, PTEN, PALB2, RB1, ATM, AKT1, CDH1, EGFR, and KMT2C. GO analysis showed enrichment in themes associated with cell proliferation and enhanced cellular metabolism. STRING PPI analysis further showed involvement in proto-oncogene activation, cell proliferation, cell cycle regulation, evasion of apoptosis, loss of tumor suppression, epithelial-mesenchymal transition (EMT), and metastasis. It also suggested involvements in interacting signaling pathways such as PI3K/AKT/mTOR, NOTCH, BRCA2, RAS/MAPK, and PK pathways. Interestingly, upon comparison with the BRCA-UK Project cohort, the number of dissimilar genes were more than the shared genes, which include TP53, TTN, SYNE1, and USH2A. The genetic profile of TNBC from literature provides a tumor microenvironment that reflects the aggressive biological behavior and poor clinical outcomes of TNBC. This can be used to elucidate the cancer biology of TNBC and provides biomarkers for possible prognosticators and targets for drug development and future research.

Keywords: bioinformatics, genetic mutations, review, triple-negative breast cancer

*Corresponding Author: alhilario@up.edu.ph
INTRODUCTION

TNBC is a peculiar type of breast cancer because of its lack of expression of the estrogen receptor (ER), the progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2). Its biology is enigmatic as it is highly heterogeneous genotypically and phenotypically. Like most cancers, its etiology is unknown. However, some TNBC subtypes have known genetic causation while the majority is uncertain. There are also some subtypes that are associated with mutations in BRCA1 and BRCA2 genes, leading to the assumption that 70% of them are hereditable. The biology and clinical course of TNBC are generally aggressive and poor with a very high predisposition for metastasis. TNBC represents approximately 15–20% of all breast cancer (Al-Mahmood et al. 2018). In the Philippines, TNBC has a prevalence of 8% with similar grim biological and clinical outlook, with patients typically presenting in late to metastatic stages (Plasilova et al. 2016).

TNBC like any breast cancer is diagnosed by histopathology. The determination of ER, PR, and HER2 using immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) is part of the clinical practice standard. These markers are used to classify them into ER ±, PR ±, and HER2 ± tumors. These classifications greatly influence prognostic, treatment planning, and overall management. Prognostic and treatment planning also depend on age, menopausal status, tumor size, histologic grade, pathologic staging, nodal and distant metastatic status, and Ki-67 expression levels, which is low for good prognosis and high for poor prognosis (Sporikova et al. 2018; Tyagi and Dhesy-Thind 2018).

The mainstay of treatment of TNBC is surgery where the role of the breast-conserving procedure is limited. Surgery is followed by chemotherapy. Radiotherapy is done after chemotherapy due to its radiation sensitizing property. Hormonal therapy such as tamoxifen and aromatase inhibitors are ineffective due to the negative expression of ER. HER2-targeted therapy like trastuzumab is also not advised since these are HER2 negative tumors. Its triple-negative phenotype limits the treatment options to surgery, chemotherapy, and radiotherapy (Kurubanjerdjit 2020).

There are several histopathologic types under the WHO classification that have a triple-negative phenotype. The most common type is the invasive ductal carcinoma of no special type, which comprises about 75% of TNBC. It has a very aggressive behavior and poor clinical outlook. Those with similar to intermediate prognosis belong to medullary carcinoma, metaplastic carcinoma, pleomorphic lobular carcinoma, and apocrine carcinoma; and those with an almost benign behavior and good clinical outcomes are the rare types of TNBC such as adenoid cystic carcinoma, secretory carcinoma, and acinic cell carcinoma (Aktepe et al. 2016; Pareja et al. 2016).

In the past three decades, our clinical understanding of TNBC remains the same. The diagnosis, treatment, and overall prognostication continue to employ the same clinical parameters and are based on few biological markers such hormonal receptors, HER2, BRCA1/2, and Ki-67 despite several attempts to subtype TNBC molecularly. There are two major molecular subtyping based on data collected from various datasets such as The Cancer Genome Atlas and Molecular Taxonomy of Breast Cancer International Consortium, and other clinical samples such as those reported by Lehmann et al. (2011) and Burstein et al. (2015). However, the use of molecular typing has not reached wide clinical application nor contributed significantly to changing its clinical outcomes. These have just led to the targeted therapy based on molecular profiles of the subtypes, which also have not yet reached accepted evidence-based clinical use. Some are still undergoing clinical trials (da Silva et al. 2020).

Presently, the biology and clinical behavior of TNBC remain perplexing for the scientific medical community. This study was done to identify the most frequently reported mutated genes in TNBC in the literature according to the number of citing articles to form a gene set, in contrast to those mined from data repositories, and to perform bioinformatic analyses of this gene set using GO (Gene Ontology) and STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) for its PPI analysis of TNBC in the hope to provide a new light in our understanding of its biology. Findings were then compared with the most prevalent somatically mutated genes found in the Breast Triple-Negative/Lobular/Other Cancer-UK (BRCA-UK) Project uploaded in the International Cancer Genome Consortium (ICGC) to provide a clinical cohort as a comparison to our gene set generated through a systematic review approach.

MATERIALS AND METHODS

The study was cleared by the University of the Philippines Manila Research Ethics Board and was registered in the Research Grants and Administration Office (RGAO) of the National Institutes of Health – University of the Philippines Manila (NIH-UPM) with RGAO registration certificate number 2019-1131. This study was part of the TNBC Philippine Research Project.

Gene Set Generation

A literature search was done at PubMed using the query ((triple-negative breast cancer) OR (triple-negative breast carcinoma)) AND ((mutation) OR
RESULTS

Gene Set Generated

After the initial literature search, there were 1,304 articles that came out of the search. After reviewing the titles, 442 articles were excluded and 862 remained. After reviewing the abstracts, 524 articles were excluded. The remaining 338 articles were manually screened and 103 articles were further removed. At the end of the screening, 231 articles were eligible for review to identify the frequently somatically mutated genes to form the gene set for this study (Figure 1).

The 54 most frequently reported genes with somatic mutations in TNBC by citing articles are shown in Figure 2. The top 10 frequently mutated genes were TP53, PIK3CA, PTEN, PALB2, RB1, ATM, AKT1, CDH1, EGFR, and KMT2C (Table 1). Only four genes were shared by both gene sets and they were TP53, TTN, SYNE1, and USH2A (Figure 3).

GO. GO analysis of the main domain on biological process derived from the top frequently mutated genes were associated with the following GO terms: 1) positive regulation of biological process (GO:0048518); 2) cellular macromolecule metabolic process (GO:00444260); 3) regulation of response to a stimulus (GO:0048583); 4) developmental process (GO:0032502); and 5) cellular component organization or biogenesis (GO:0016043) (Figure 4).

STRING. STRING analysis of the gene set of somatically mutated genes from the literature showed genes with overlapping interactions and no speculated interaction (Figure 5). These genes – TP53, TTN, SYNE1, and USH2A – were shared genes by the two gene sets from this review and the BRCA-UK Project. Except for TP53, TTN, SYNE1, and USH2A showed no speculated interaction. There were six major clusters of interactions. These clusters of interactions involved the following: 1) TP53; 2) PTEN; 3) PIK3CA, PIK3R1, ERBB2, and NOTCH1; 4) BARD1, PALB2, ATM, and RAD50; 5) KRAS, HRAS, NRAS, EGFR, and BRAF; and 6) EGFR, CDH1, RET, and MET.

DISCUSSION

The somatically mutated genes as reported in this study provided credence to the high heterogeneity found in TNBC. This genetic complexity and heterogeneity of TNBC contribute to the intrinsic inability of established molecular subtyping to be used in clinical practice. The study reported the top ten somatically mutated genes in the literature that encompassed a wide range of

((polymorphism) OR (genetic polymorphism)) OR ((single nucleotide polymorphism) OR (single nucleotide variant) OR (single nucleotide variation))) on 27 Feb 2020 for articles published from 1999–2019 to collect extensively the reported genes associated with TNBC, which were used to generate the gene set. The inclusion criteria were as follows: 1) studies that reported genetic profile, molecular profile, and biomarkers of TNBC using techniques such as IHC, FISH, and microarray-based studies; 2) studies that are derived from human samples and/or mined from gene expression studies of human samples; and 3) review articles, case reports, and case studies with reported genetic profile, molecular profile, and biomarkers of TNBC using the above-mentioned techniques. The exclusion criteria included: 1) studies done using animal models, tumor xenografts, and cancer cell lines; 2) studies about other types of breast cancer; and 3) studies not written in English. The search was repeated on 20 Apr 2020. The most prevalent mutated genes were reported and presented according to the frequency of citing articles. Only somatic mutation, genetic alteration occurring after fertilization, was included in both data sets.

GO. The generated gene set from the top 54 somatically mutated genes in literature was subjected to GO enrichment accessible at https://geneontology.org using GO enrichment analysis by PANTHER. Significant GO terms associated with gene set were set at \( p < 10^{-7} \).

STRING. The PPIs of the gene products of the generated gene set were determined using STRING version 10.5 by STRING Consortium available at https://string-db.org. The degree of interaction between genes denoted by interconnecting lines (edges) was deduced by the weights of the lines, which are directly proportional to the degree of PPI. Significant PPI \( p \)-value and false discovery rate were set at \( p < 0.05 \).

The gene set of somatically mutated genes generated from the literature search was compared with the top 50 mutated genes of the BRCA-UK Project uploaded in the ICGC, which was accessed at https://dcc.icgc.org/projects on 21 Feb 2020 (Stratton 2019). This clinical cohort was chosen because this was not included in previous reported molecular typing in literature used in molecular subtyping and contains clinical data. Similar and dissimilar genes were presented. The data produced in this study and used to arrive at the results including the generated table and figures can be accessed freely at https://doi.org/10.5281/zenodo.3828222.
biological processes, which can affect the hallmarks of carcinogenesis of TNBC.

The gene set generated in this study was associated with protein-protein pathways associated with TNBC’s tumorigenesis, carcinogenesis, and biological characteristics. The enriched gene set also showed GO terms enrichment associated with cell proliferation and increased metabolism. Interestingly, there were more dissimilar than similar somatically mutated genes in this gene set than those found in a clinical cohort. Similar genes noted in this study—like TTN, SYNE1, and USH1A—were passenger genes in carcinogenesis except for TP53, which is a known driver gene in many cancers (Doherty et al. 2010; Li et al. 2017; Oh et al. 2020).

The genetic landscape of TNBC paints more aggressive biology and a vicious clinical outcome consistent with this breast cancer. The mutations in EGFR, PTEN, PIK3CA, AKT1, RB1, ATM, and KMT2 might explain the observed high cell proliferative state and dysregulated cell signaling, which lead to uncontrolled cell proliferation in TNBC. This state of high cell proliferation exposes tumor cells to the vulnerability of the DNA to genetic alterations,
Table 1. Biological functions of 10 most frequently reported genes mutated in TNBC.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene name</th>
<th>Biological function</th>
<th># of citing articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>Tumor protein 53</td>
<td>Tumor suppression</td>
<td>98</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>Phosphatidylinositol 4,5-bisphosphate-3-kinase catalytic subunit alpha</td>
<td>Cell signaling</td>
<td>77</td>
</tr>
<tr>
<td>PTEN</td>
<td>Phosphatase and tensin homolog</td>
<td>Cell signaling and tumor suppression</td>
<td>49</td>
</tr>
<tr>
<td>PALB2</td>
<td>Partner and localizer of BRCA2</td>
<td>Tumor suppression</td>
<td>29</td>
</tr>
<tr>
<td>RB1</td>
<td>RB transcriptional RB transcriptional kinase</td>
<td>Cell proliferation</td>
<td>27</td>
</tr>
<tr>
<td>ATM</td>
<td>ATM serine/threonine kinase</td>
<td>Cell proliferation</td>
<td>24</td>
</tr>
<tr>
<td>AKT1</td>
<td>AKT serine/threonine kinase 1</td>
<td>Cell proliferation</td>
<td>22</td>
</tr>
<tr>
<td>CDH1</td>
<td>Cadherin 1</td>
<td>Maintenance of epithelial features</td>
<td>22</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor Receptor</td>
<td>Cell proliferation</td>
<td>16</td>
</tr>
<tr>
<td>KMT2C</td>
<td>Lysine methyltransferase 2D</td>
<td>Cell proliferation</td>
<td>16</td>
</tr>
</tbody>
</table>

Source: https://www.ncbi.nlm.nih.gov/genbank
Figure 3. Venn diagram of the two sets of genes: A) top 50 mutated genes from the BRCA-UK Project (blue) and B) The 54 most reported mutated genes cited from the literature (yellow). Source: http://icgc.org.

Figure 4. GO terms in GO:biological process most significantly enriched ($p < 10^{-7}$) in the ranked list of genes mutated in TNBC cited in the literature. Source: https://geneontology.org.
which can lead to significant genetic mosaicism within
the tumor itself (Davis et al. 2014). The loss of function in
the mutated genes (TP53, PTEN, and RB1) contributes to
the evasion of apoptosis and progression to metastasis (Liu et
al. 2017). Also, a mutation in the CDH1 gene promotes
EMT, which leads to invasiveness and metastasis (Li et
al. 2019). The genetic perturbations in these genes play
a significant role in the highly aggressive, proliferative,
and metastatic potential of TNBC.

Characteristically, the number one mutated gene in this
study was TP53, which is also the topmost mutated gene
in the BRCA-UK Project (Stratton 2019). Central to
the biology of cancer metastasis, the loss of function of
TP53 is considered the last hurdle for tumors to cross
the divide between benign and malignant tumors or
localized to metastatic tumors. The mutation in TP53 is
widespread in various breast cancer types and even across
different cancers (Powell et al. 2014). The role of TP53
mutation plays a significant role in TNBC metastasis.
The loss of function of this tumor suppressor gene leads
to heightened cell migration and invasion beyond the
basement membrane. This generally results in increased
metastatic potential. The mutation in TP53 also results in
the fields effect among its tumor population. This fields
effect allows the creation of more clones within the tumor
itself (Meric-Bernstam et al. 2018; Lochhead et al. 2015).

GO analysis showed enrichment of genes associated with
GO terms responsible for diverse biological processes
consistent with cell proliferation and increased cellular
metabolism. These findings are consistent with noted cell
proliferation and increased metabolism in cancer cells.
Also, cancer cells especially in large tumors may have to
adapt to a low-nutrient environment in order to survive
and allow cancer cell metabolism to sustain cell growth,
division, and proliferation (Sun et al. 2020).

The PPI of the top 54 differentially expressed genes in
this study showed six major clusterings. The first cluster
revolved around the TP53 gene. This gene is the most
frequently mutated gene in breast cancer in general
and TNBC as a subset with 30% mutant forms, which
could result in negative or positive outcomes in TNBC
(Shahbandi et al. 2020). It is involved in the dysregulation
of cell proliferation, evasion of apoptosis, and cellular
invasion and metastasis. Similarly, PTEN – a tumor
suppressor gene – is associated with both primary and
metastatic TNBC affecting the AKT pathway, which
results in evasion of apoptosis and cell proliferation (Shen et al. 2019).

One notable protein clustering is the affection of several signaling pathways, which are widely and significantly involved in tumorigenesis and carcinogenesis. These include the P13K/AKT/mTOR, NOTCH, BRCA2, RAS/MAPK, and PK pathways. These pathways are involved in cell proliferation and cell survival. The P13K/AKT/mTOR pathway is responsible for chemoresistance, making chemotherapy useless for some TNBC cohorts (Khan et al. 2019). NOTCH pathway is associated with cell differentiation and cell proliferation. It promotes EMT and down-stream activation of the AKT pathway. It is also considered a factor in chemoresistance for some chemotherapeutics agents (Xiao et al. 2019). The BRCA2 pathway is activated by proto-oncogene activation. The BRCA2 protein is co-stabilized by PALB2 and affects RAD50 in downward signaling. This results in cell proliferation and loss of tumor suppression (Chartron et al. 2019).

The RAS/MAPK pathway is involved in various tumors, especially solid tumors. It regulates cell cycle continuation, cell proliferation, and other signaling pathways. Its initiating activation is the interaction of the growth factor to EGFR and the subsequent activation of its intracellular signaling, causing the transcription of genes involved in the cell cycle, cell growth, and proliferation. This pathway is also activated by the RAS family of genes like KRAS, HRAS, and NRAS (Peng et al. 2019). Lastly, the PK (protein kinase) pathway – which is associated with the AKT/PKB pathway – is responsible for cell proliferation. It is activated by growth factors, RET, and MET proteins. There are significant interactions in its downstream signaling with AKT, ATM, and ERBB proteins. There are also significant interconnections with some pathways like the P13K/AKT/mTOR, RAS/MAPK, and AKT/PKB pathways (Iida et al. 2020).

This study showed that a robust genetic profile can be collected from the gene set reported in the literature, which can be subjected to bioinformatic analyses and used to elucidate the biology of TNBC and explain its observed clinical outcomes. The clinical applicability of molecular subtyping in TNBC remains highly unutilized. Future research directions should unravel not only the genetic profile but also its epigenetic and microRNA profiles. TNBC remains a highly heterogeneous disease in terms of histologic diversity and molecular complexity, which may continue to challenge our ability to provide a clinically meaningful molecular subtyping. This peculiarity of TNBC continues to elude the medical community’s efforts to significantly change the way the medical community diagnoses, treats, and prognosticates this specific disease.

CONCLUSION

TNBC is a heterogeneous breast cancer type with divergent genotypes and phenotypes and is associated with aggressive biological behavior, poor prognosis, poor response to standard treatment, and high metastatic potential and mortality. The most frequently cited mutated genes in literature are numerous and belong in wide and overarching biological processes. The top mutated genes included TP53, PIK3CA, PTEN, PALB2, RB1, ATM, AKT1, CDH1, EGFR, and KMT2C. GO and STRING analyses showed various GO terms with biological consequences and protein interactions involved in proto-oncogene activation, growth factors and other ligands for cell proliferation, cell cycle, cell survival, evasion of apoptosis, loss of tumor suppression, EMT, and metastasis. It also affected several interacting signaling pathways such as the P13K/AKT/mTOR, NOTCH, BRCA2 pathway, RAS/MAPK, and PK pathways. Comparison with the gene profile of the BRCA-UK Project cohort showed more dissimilar genes than shared genes such as TP3, TTN, SYNE1, and USH2A. The genetic profile gathered from the literature in this study provides a tumor microenvironment that supports the biology, prognosis, and clinical outcomes of TNBC and can be used for prognostication and drug discovery for targeted therapeutics.

ACKNOWLEDGMENTS

This study was partially funded by the NIH-UPM.

STATEMENT ON CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES


