

# Study of APOBEC3B focused breast cancer pathways and the clinical relevance

Hani Choudhary (✉ [hchoudhry@kau.edu.sa](mailto:hchoudhry@kau.edu.sa))

King Abdulaziz University

Ashwag Albukhari

King Abdulaziz University

Mohammad Mobashir

Karolinska Institute

Wesam Abdulaal

King Abdulaziz University

---

## Research Article

**Keywords:** APOBEC3B, breast cancer, expression and mutation analysis, co-expression, survival analysis, clinical relevance.

**Posted Date:** April 14th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-397072/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

APOBEC3B is considered as an enzymatic source of mutation in case of breast cancer and Human T-Cell Leukemia Virus Type 1 and Bone Leiomyosarcoma are also associated with it. The major functions controlled or affected due to APOBEC3B are gene expression, mRNA editing such as C → U conversion, and deoxycytidine deaminase activity. Here, the main goal of the study was to perform a systematic analysis of APOBEC3B associated genes and its functional impact in human breast cancer. For this purpose, the datasets have been utilized from the publicly available database such as GEO, OncoLnc, and TCGA. Based on the requirements for fetching the values, different bioinformatics approaches have been applied at different levels. Further, co-regulated genes obtained from co-expression network have been processed and the mutated genes with the pathways enrichment analysis, and the clinical relevance using survival curve analysis by using OncoLnc have been performed. In the results, we found that there are a number of critical pathways known to directly associated with breast cancer are altered because of the genes which are either overexpressed or top mutated and are associated with APOBEC3B and these pathways are cell cycle, p53 signaling, immune signaling pathways, progesterone-mediated oocyte maturation, apoptosis, critical metabolic pathways, and pathways in cancers. From the mutational and survival analysis data, we also observe that there are a number of well-known cancer associated signaling pathways (mainly cancer), immune signaling pathway, critical metabolic associated pathways, cell cycle, ubiquitin-proteasomal signaling pathways, and p53 signaling. Network-level study of pathways and their components CD40LG as the potential gene where CD40LG is directly affecting 10 pathways and most of them are the parts of immune system and known to control a number of leading human diseases including cancers.

## 1. Introduction

Genomic instability<sup>1,2</sup> is considered as the potential driver for major human diseases mainly cancer which rapidly promotes cancer progression leading to clonal selection in tumor cells<sup>3-8</sup> and further leads to drug resistance and poor clinical outcomes. Mutations are considered as among the potential genomic instability factors and play critical roles in cancer progression and clinical outcomes and drug resistance and the mutations are classified as driver and passenger mutations<sup>9-14</sup> which could be induced by multiple factors which could be internal or external factors. Furthermore, gene expression pattern also displays higher-level of aberrations and their direct functional impact<sup>15-22</sup>.

The advancement in technology coupled with appreciable progress in interdisciplinary approaches has enabled the direct characterization of significant functional mutations, genes, and pathways involved in various complex human diseases including cancer<sup>23-25</sup>. Classically, it was established that damage due to radiation and chemicals were main culprits for carcinogenic mutagenesis<sup>26-29</sup>. The enzyme family of DNA cytosine deaminases having important role in intrinsic DNA mutations and the human genomes contain information for 11 polynucleotide cytosine deaminase family enzymes that are responsible for mutations in cancer APOBEC1, activation-induced deaminase (AID), APOBEC2, APOBEC3 proteins (known

as A3A, A3B, A3C, A3D, A3F, A3G and A3H), and APOBEC4. APOBEC2 and APOBEC4 are not reported to have not mutational activity. A tissue-specific expression has been shown for APOBEC1 and AID and has implications in cancers of those tissues, hepatocytes and B cells, respectively<sup>9,30-34</sup>. These enzymes were identified independently in 2002 as DNA mutators and antiviral factors. Among all the 11 family members, APOBEC3B is considered to be responsible for mutations in breast cancer and is found over-expressed in many breast cancer cohorts<sup>31,35-38</sup>.

There are numerous causes (such as changes or mutations in DNA, gene expression pattern alterations, epigenetic changes) which may be associated with breast cancer origin<sup>39-42</sup> and the APOBEC3B mutation is one of them which is known to play pivotal role in breast cancer in the possible clinical outcomes of the patients. APOBEC3B increases the mutation load, generating clusters of closely spaced, single-strand-specific DNA substitutions with a characteristic hypermutation signature in breast cancer<sup>37</sup>. A plethora of reports is available in support of the genomic alterations by APOBEC3B in the human genome and the mutation patterns have been studied in detail<sup>37</sup>. The factors responsible for their expression regulation have also been deduced where Linda and co-workers have revealed the biochemical basis its action. It is reported that p53 and NF-κB controls the expression of APOBEC3B in an independent manner. In previous works, it has been shown that expression of these enzymes is controlled or inhibited by other genes and proteins. Analyses of *APOBEC3B* associated networks and target genes (for mutations) is a promising source of insight for discovering therapeutic targets for *APOBEC3B* mutated cancers. The associations of *APOBEC3B* hyper-expression, *APOBEC3B* interaction networks, and survival curves of patients having mutations or overexpression of these related genes have not been widely explored. A systematic analysis of *APOBEC3B* induced mutations and the related interaction networks in various types of human cancers especially in breast cancer is lacking. There have been a number of studies of genomic alterations across cancer types based on TCGA data.

However, none of them have focused on the systematic study and explored genomic aberrations of *APOBEC3B* and its related interaction networks across the breast of cancer type. Previous works have addressed the role of APOBEC3B in a discrete manner while our potential emphasis has been on the combined study of gene expression datasets and mutational datasets to evaluate the association and impact of APOBEC3B in breast cancer including the network-level understanding of these enzymes and their associated molecular components. Further, we have focused to elucidate the clinical relevance of those APOBEC3B co-expressed genes depending variation in their expression of mutation. Our second aim targets at the overall genes preferentially mutated in breast cancer and the clinical relevance. This study reveals the detailed understanding from basics of APOBEC3B and the associated genes and the inferred functions with it, mutational profiling and the associated biological functions, survival analysis for clinical relevance, and the networks of both the genes and the pathways. It leads to the conclusion that APOBEC3B is connected with a large number of genes in the human network database and these genes are present in the critical pathways for especially breast cancer. These pathways are either altered due to differential expression of the genes or mutations. Furthermore, we conclude the same from the mutated genes and the clinically significant genes.

## 2. Materials And Methods

In this study, we have collected the datasets for expression and mutational profiling from a publicly available database (GEO, OncoLnc, and TCGA) for breast cancer. The list of OncoLnc for the cancer names abbreviations have been presented in Supplementary Table S1. Majority of the comparative analysis has been performed between normal and the tumor samples. The mutational datasets details are as follows: dataset 1 (from TCGA database) was breast cancer (METABRIC, Nature 2012 and Nature Communication 2016) and contains 2509 samples<sup>43,44</sup> and dataset 2 was breast cancer invasive carcinoma (TCGA, Firehose Legacy) and contains 1108 samples<sup>44</sup>. For data processing, normalization, analysis, and figure plotting, MATLAB has been used in most of this work. Similarly, MATLAB has been used for the majority of the computational analysis from normalization to statistical analysis and all kinds of data processing and to summarize our work approach, we have presented a workflow as Fig. 1a. This workflow represents the quick insight into the most of used approaches for the entire analysis which presents the details of the steps from data collection to the analysis steps while more details had been described in the next paragraphs.

After establishing the fundamental relevance of the APOBEC3B and its association with other genes and pathways, we have used KEGG pathway database and network database (FunCoup) to present a network of both the pathways and their directly associated components i.e., directly APOBEC3B-associated genes. To prepare or fetch the network(s) for the desired list of genes, the easiest way is to directly use the server or to prepare the list of genes and write the own code to fetch out the connectivity for the list of genes in the list the later was our preferred approach. Furthermore, we have accessed the TCGA datasets belonging to breast cancer where we have mapped out the co-expressed genes for APOBEC3B and presented top 25 co-expressed genes based on the correlation values and all these genes have positive correlation values (Fig. 1c) and also checked out the mutation of APOBEC3B. This have been done by searching the APOBEC3B genes in case of breast cancer in TCGA database and from the next step, the co-expression option had been selected and the threshold of +/- 0.5 (correlation value) has been applied on the correlation (spearman) values for selecting the co-expressed genes.

For the list of mutated genes obtained for large datasets of breast cancer from TCGA<sup>44</sup> and pathway enrichment analysis has been performed and the cutoff  $p - values$  were 0.05 and the steps implemented was similar to DAVID database<sup>45-47</sup> and for network drawings cytoscape<sup>48-50</sup> has been used and for enriched pathways, basic enrichment approach has been used. For survival analysis and to predict the top ranked genes in terms of survival in breast cancer OncoLnc have been used<sup>51</sup>. In survival curve analysis, we have presented the overall list of genes with significant  $p - values$ . Furthermore, for the top 100 clinically significant genes, panther database<sup>52</sup> has been used for the panther protein classification. In short we could also say that the individual files have been prepared by using the command line codes (unix) for genes and pathways list preparations based on the relevant needs for the purpose to analyze the enriched pathways or preparing the networks or for panther protein class prediction<sup>53-55</sup>.

## 3. Results

### 3.1. Understanding the role of APOBEC3B and the associated genes and pathways in breast cancer

We have designed our study to achieve the goal as mentioned in the previous section and for which the large datasets (expression and mutation) have been collected from GEO, OncoLnc, and TCGA and applied *in-silico* approach for analysis and the summarized layout have been presented in Fig. 1a. We have mapped out the cox regression analysis of APOBEC3B in different types of cancers (Table 1) and the directly APOBEC3B-associated genes and the respective pathways for all the genes by using protein-protein interaction database and finally presented a combined network of all these genes and the pathways in Fig. 1b. Furthermore, a network of co-expressed genes for APOBEC3B and the associated pathways have been presented in Fig. 1c. In Fig. 1b and 1c, we observe that p53 signaling, cell cycle, oocyte meiosis, major cancer signaling, ubiquitin-mediated proteolysis, TLR signaling, chemokine signaling, antigen processing and presentation, regulation of actin cytoskeleton, neurotrophine, MAPK, BCR signaling, a number of metabolism associated signaling, and calcium signaling pathways as the potentially affected pathways due to alteration(s) in APOBEC3B either expression or mutation or both. From the previous work, it is known that in a number of cancer normal and malignant cells, increased GSH level is associated with a proliferative response and is essential for cell cycle progression, regulation of actin cytoskeleton pathway has its role in cancer cell migration and invasion in ECM, p53 signaling pathway, MAPK kinase pathway, neurotrophine and chemokine and calcium signaling pathways which when altered in cancer cells are involved in tumor initiation, angiogenesis, progression, and metastasis. These pathways are known to be very specific and play pivotal role in cancer cell migration and proliferation and these pathways which are very important in cancer progression were observed to be strongly associated with APOBE3B.

Table 1  
Cox regression analysis for APOBEC3B.

Cox regression results for APOBEC3B						
Cancer	Cox Coefficient	P-Value	FDR Corrected	Rank	Median Expression	Mean Expression
BLCA	0.004	9.60E-01	9.82E-01	15906	362.19	603.54
BRCA	0.112	2.10E-01	5.89E-01	5758	162.02	321.98
CESC	-0.608	1.30E-05	2.55E-02	8	755.45	924.17
COAD	-0.094	3.40E-01	7.24E-01	7557	214.67	245.31
ESCA	-0.048	7.20E-01	9.84E-01	12129	239.99	314.7
GBM	0.004	9.60E-01	9.92E-01	16161	67.15	86.21
HNSC	-0.028	7.10E-01	8.90E-01	13181	467.44	691.26
KIRC	0.19	2.40E-02	6.18E-02	6421	61.4	80.76
KIRP	0.59	5.50E-04	6.78E-03	1328	50.53	73.29
LAML	-0.018	8.70E-01	9.58E-01	13732	160.13	206.39
LGG	0.216	2.00E-02	4.32E-02	7720	17.17	25.49
LIHC	0.093	2.90E-01	5.53E-01	8176	83.16	161.41
LUAD	-0.015	8.40E-01	9.23E-01	15194	135.46	255.45
LUSC	-0.117	7.80E-02	6.10E-01	2146	491.14	651.6

Further, extending our analysis we classified positive and negative co-expressing genes but threshold we have applied for correlation value was either greater than + 0.5 or less than - 0.5 and such negative and positive correlation in expression provides important information regarding the dependence of gene expression on each other.

### 3.2. Mutational profiling and their functional impact in breast cancer

In the previous section, we have explored the fundamentally APOBEC3B associated genes and the pathways and moreover, mutational profiling of breast cancer genes have been performed and the respective enriched pathways for the two big clinical datasets from TCGA database. In this analysis, two datasets have been used and among the top-ranked mutated genes there are a large number of common genes (Fig. 2a) and as we go down the number of common genes decreases and similar to it the common enriched pathways with the respective p-values have been presented (Fig. 2b) and finally the venn diagram has been shown for both the genes and the enriched pathways (Fig. 2c). PIK3CA, TP53, MUC16, TTN, AHNK2, SYNE1, CDH1, KMT2C, GATA3, and more are among the top ranked genes and MAPK, calcium signaling, cAMP, PI3K-AKT, focal adhesion, adrenergic signaling, thyroid hormone, oxytocin, ErbB, ubiquitin, apelin, tight junction, GnRH, Ras, cGMP-PKG, cell cycle, and pluripotency of stem cells are among the commonly enriched pathways. Overall, there were 42 commonly mutated genes and 18 pathways commonly enriched, 131 genes dataset1 specific mutated genes and 41 enriched pathways while 188 mutated genes and two enriched pathways specific to dataset 2. Among the enriched pathways list for dataset 1, there are a number of pathways which directly belong to the immune system and these pathways are TCR, BCR, TLR, NK cell-mediated cytotoxicity, TNF, TGF, cytokine-cytokine receptor interaction, and leukocyte transendothelial cell migration and ubiquitin-mediated proteolysis is common to both the mutational datasets. In Fig. 2c, We have presented the detailed analysis of the clinically significant top 100 genes and the associated biological functions, where we observe that CD40LG is directly associated 10 pathways and most of them are the parts of immune system and known to control a number of leading human diseases including cancers. CD40LG (CD40 Ligand) is a Protein Coding gene. The diseases associated with CD40LG are mainly immunodeficiency with Hyper-Igm, type-1 and toxoplasmosis. It mainly acts as a ligand for integrins, mainly ITGA5:ITGB1 and ITGAV:ITGB3; both the integrins and the CD40 receptor are required for activation of CD40-CD40LG signalling which have cell-type dependent effects, such as B-cell activation, NF- $\kappa$ B signaling and anti-apoptotic signaling.

Furthermore, we have also performed analysis of APOBEC3B signature (C->T) in breast cancer and observe that this specific mutation pattern is dominantly present in missense and exon followed by upstream, synonymous, intron, 3' UTR, and 5' UTR (Supplementary data S2).

### **3.3. Functional and clinical significance of APOBEC3B and the associated genes in breast cancer**

After analyzing the APOBEC3B related genes and the pathways including the mutated genes and the altered functions because of mutation, we have analyzed the overall list of genes which are showing clinical significance in terms of overall patients survival (Fig. 3). Here, top 100 genes have been presented (Fig. 3a) followed by the respective p-values, panther protein classes for these top-ranked genes (Fig. 3b) have been shown also, and finally the associated pathways and the top-ranked genes as a network have also been shown (Fig. 3c). In terms of overall survival, MCTS1, OVOS2, MAPT-IT1, ATG4A, SLC16A2, SLC35A2, VDAC1, TBC1D24, RP11-214F16.8, and PDP1 are showing extremely high significance. From panther protein classification, majority of these top-ranked genes mainly belong to metabolite interconversion enzyme, transporter, protein modifying enzyme, defense/immunity protein, gene-specific transcriptional regulator, and membrane traffic protein classes. Most of the clinically significant genes

belong to the above mentioned protein classes where metabolite interconversion enzyme has highest number of the genes followed by transporter, and protein modifying protein classes (Fig. 3b). From the network of these top-ranked genes and the associated functions, CD40LG appears to control the 10 pathways and most of them belong to immune signaling. Among the overall functions associated with these top-ranked genes, majority of these pathways are known and well-established that they control major human diseases multiple types of cancers including breast cancer, neurodegenerative diseases, diabetes, and infection diseases (Fig. 3c). Thus leading to the conclusion that immune system and its critical components are mainly affected as a result of breast cancer. Moreover, we have also presented a supplementary data where the clinical details have been presented (Supplementary data S3).

## 4. Discussion

There are a large number previous studies where the genomic aberrations and its impact have been studied including the genomic alterations by APOBEC3B in human genome and mutation profiling<sup>35,37,56</sup>. APOBEC-induced mutations are reported in early replicating regions where chromatin is active and the single state DNA present in these replicating regions is with increased DNA fragility producing additional substrates which are of great interest for APOBEC3B activity. Alterations and distortions in early replicating and highly transcribed regions of cancer genomes arise due to chromosome breakage, such as copy number variation, chromosome rearrangements, fragility and loss of heterozygosity and similar to it, our results appear<sup>57-60</sup>. The pathway enrichment of the top mutated genes in two cohorts revealed that pathways involved in cellular signaling were the main mutational target for APOBEC3B in breast cancer patients (Fig. 3b). Enhanced DNA damage and repair is prone to increased changes of mutation by APOBEC3B. Interestingly, APOBEC-induced mutations are abundantly found in cancer genomes but no regulatory sequences are reported yet. It is reported that total of < 6% of APOBEC-induced kataegic events in the vicinity of transcriptional start sites compared to 82% for AID-induced events and while APOBECs favor early replicating regions of the genome, which in B-cells, are devoid of AID-induced DSBs<sup>61-63</sup>. Few factors affecting the expression of APOBEC3B directly or in-directly has also been screened previously. It is reported that p53 and NF- $\kappa$ B control the expression of APOBEC3B in an independent manner<sup>36,37,64</sup>. So far the studies have been mainly on the role of APOBEC3B in as an isolated study and keeping all this here, we have potentially emphasized on the integrated study of APOBEC3B associated genes, functions, expression datasets, and the clinical relevance to evaluate the association and impact of APOBEC3B in breast cancer. Here, the effort has been made to explore in depth the APOBEC3B and the associated genes followed by their functions by applying a number of approaches as an integrative study by including the network-level understanding<sup>20,65-68</sup> of the relevant genes and their associated functions. As we already know that APOBEC3B is a gene editing enzyme with cytidine deaminase activity and high expression of its mRNA in breast tumors have been shown to be associated with progressive cases and poor prognosis. In this study, we aimed to examine the relationship between the expression of APOBEC3B and the effect of neoadjuvant chemotherapy (NAC) using pretreatment biopsy tissue, and examined whether the expression of APOBEC3B influenced chemotherapy efficacy. So here, in terms of correlation of APOBEC3B expression with the clinicopathophysiological features, it has been shown APOBEC3B as a

novel predictive factor for pathological complete response to neoadjuvant chemotherapy in breast cancer<sup>69-71</sup>. Moreover, there are a number of previous studies where the functional assays of APOBEC3 have been performed<sup>72-74</sup>. Here, they have performed analysis to characterize the functional diversification of APOBEC gene family members associated with breast cancer mutagenesis for estrogen receptor (ER) status and found that both APOBEC3B and APOBEC3C mRNA levels among the APOBEC family were significantly higher in estrogen receptor negative (ER-) subtype compared with estrogen receptor positive (ER+) subtype<sup>69,71-73</sup>. Here, the APOBEC3B family members show mRNA levels extremely low or no obvious expression differences between ER + and ER - breast cancers. The expression levels of APOBEC family genes in 55 breast cancer tissues with unknown ER status were similar with those samples of ER + subtypes which suggests that many of these 55 patients may belong to ER + subtype leading to the conclusion that ER status dependent expression patterns of APOBEC3B and APOBEC3C genes. In terms of DNA methylation, highly expressed genes appear to have lower DNA methylation levels in their proximal transcription start site (TSS) regions<sup>72</sup>.

This study is an effort to elucidate the clinical relevance of APOBEC3B and breast cancer associated genes to study a relation between variation in their expression and followed by the mutational profiling and clinical relevance of breast cancer associated genes. As presented in the results section, the APOBEC3B associated genes and pathways, breast cancer mutational profiles and the enriched pathways, and the survival analysis of the breast cancer genes. This study leads to the conclusion that APOBEC3B may impact broadly in case of breast cancer. The genes and the enriched pathways are reported to play critical role in breast cancer<sup>75-83</sup>. The genes in co-relation with APOBEC3B, breast cancer mutated genes, and clinically relevant genes belong to the pathways which are potentially known to be altered in cancer especially breast cancer. Pathways responsible for the cancer cell progression and migration were observed closely associated with APOBEC3B inferring that APOBEC3B has an important contribution in the spread of breast cancer directly or indirectly. The co-relation trends generated from expression datasets suggest that some genes might be suppressed by APOBEC3B. The co-expressing analysis showed spatial patterns of gene expression indicating that their respective DNA portion is transcriptionally active. Transcription occurs on unfolded DNA which gives APOBEC3B an opportunity to target these exposed portion to induce mutations. Here, we have tried to establish a connection between the co-expression of genes with APOEBC3B and the instances of mutations occurring during this event<sup>33,35,61,62,84,85</sup>.

It is also that the study mainly focused on computational analysis by using very simplified approach to present the APOBEC3B associated genes and the pathways, mutational profiling and the altered pathways in result of mutation, survival analysis for the clinical samples, and finally, network-level understanding of the critical genes and potential pathways. This study not reveals the above mentioned goals but also gives quick insight into easy understanding of the potential genes and the critical pathways which may help in clinical and therapeutic purpose. In terms of future perspective and the application, mathematical modeling and simulations of the pathway-network model could be further

explored by using the so far prediction from this study<sup>13,86-91</sup> and integrated with experimental data also<sup>92</sup>.

## 5. Conclusions

Based on this study, we conclude that the APOBEC3B associated genes and the pathways are very specific and play pivotal role in cancer cell migration and proliferation and these pathways which are very important in cancer progression were observed to be strongly associated with APOBEC3B. PIK3CA, TP53, MUC16, TTN, AHNAK2, SYNE1, CDH1, KMT2C, GATA3, and more are among the top ranked genes and MAPK, calcium signaling, cAMP, PI3K-AKT, focal adhesion, adrenergic signaling, thyroid hormone, oxytocin, ErbB, ubiquitin, apelin, tight junction, GnRH, Ras, cGMP-PKG, cell cycle, and pluripotency of stem cells are among the commonly enriched pathways. From survival perspective, MCTS1, OVOS2, MAPT-IT1, ATG4A, SLC16A2, SLC35A2, VDAC1, TBC1D24, RP11-214F16.8, and PDP1 are showing extremely high significance and the panther protein classification, majority of these top-ranked genes mainly belong to metabolite interconversion enzyme, transporter, protein modifying enzyme, defense/immunity protein, gene-specific transcriptional regulator, and membrane traffic protein classes. Network of the top-ranked overall survival analysis based genes and the associated functions, CD40LG appears to control the 10 pathways and most of them belong to immune signaling. Among the overall functions associated with these top-ranked genes, majority of these pathways are known and well-established that they control major human diseases multiple types of cancers including breast cancer, neurodegenerative diseases, diabetes, and infection diseases.

## Declarations

**Author Contributions:** Conceptualization, H.C.; methodology, H.C., A.A., M.M., and W.H.A.; software, H.C.; validation, H.C., A.A. and W.H.A.; formal analysis, H.C.; investigation, H.C.; resources, H.C.; data curation, H.C., A.A., M.M., and W.H.A.; writing—original draft preparation, H.C., A.A., M.M, and W.H.A.; writing—review and editing, H.C.; visualization, H.C., A.A., and W.H.A.; supervision, H.C.; project administration, H.C.; funding acquisition, H.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU), Jeddah, grant number G: 1437-130-655” and “The APC was funded by G: 1437-130-655.

**Acknowledgments:** We are grateful to the Department of Biochemistry, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia and Cancer Metabolism and Epigenetic Unit, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia for providing us all the facilities to carry out the entire work.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. BECKMAN, R. & LOEB, L. Genetic instability in cancer: Theory and experiment. *Seminars in Cancer Biology*. **15**, 423–435 (2005).
2. Joenje, H. & Patel, K. J. The emerging genetic and molecular basis of Fanconi anaemia. *Nature Reviews Genetics*. **2**, 446–457 (2001).
3. Swanton, C. Cancer evolution: the final frontier of precision medicine? *Annals of Oncology*. **25**, 549–551 (2014).
4. Michor, F., Iwasa, Y. & Nowak, M. A. Dynamics of cancer progression. *Nature Reviews Cancer*. **4**, 197–205 (2004).
5. Paguirigan, A. L. *et al.* Single-cell genotyping demonstrates complex clonal diversity in acute myeloid leukemia. *Science Translational Medicine* **7**, (2015).
6. Yizhak, K. *et al.* RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science*. **364**, eaaw0726 (2019).
7. Cooper, C. S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nature Genetics*. **1–9**, <https://doi.org/10.1038/ng.3221> (2015).
8. Wheeler, H. E., Maitland, M. L., Dolan, M. E., Cox, N. J. & Ratain, M. J. Cancer pharmacogenomics: strategies and challenges. *Nature Reviews Genetics*. **14**, 23–34 (2012).
9. Pon, J. R. & Marra, M. A. Driver and Passenger Mutations in Cancer. *Annu. Rev. Pathol. Mech. Dis.* **10**, 25–50 (2015).
10. Vogelstein, B. *et al.* Cancer genome landscapes. *Science*. **339**, 1546–1558 (2013).
11. Parsons, D. W. *et al.* An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science*. **321**, 1807–1812 (2008).
12. Klijn, C. *et al.* Analysis of Tumor Heterogeneity and Cancer Gene Networks Using Deep Sequencing of MMTV-Induced Mouse Mammary Tumors. *PLoS ONE*. **8**, e62113 (2013).
13. Huang, S., Ernberg, I. & Kauffman, S. Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in Cell & Developmental Biology*. **20**, 869–876 (2009).
14. Wang, E. Cancer Letters. *Cancer Lett.* **340**, 261–269 (2013).
15. Stein, S. *et al.* Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nature Medicine*. **16**, 198–204 (2010).
16. Gazy, I. & Kupiec, M. Genomic instability and repair mediated by common repeated sequences. *Proceedings of the National Academy of Sciences* **110**, 19664–19665(2013).
17. Habermann, J. K. *et al.* The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. *Int. J. Cancer*. **124**, 1552–1564 (2009).

18. Grubor, V. *et al.* Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). *Blood*. **113**, 1294–1303 (2009).
19. Restifo, N. P., Smyth, M. J. & Snyder, A. Acquired resistance to immunotherapy and future challenges. *Nature Reviews Cancer*. **16**, 121–126 (2016).
20. Zhao, B., Pritchard, J. R., Lauffenburger, D. A. & Hemann, M. T. Addressing Genetic Tumor Heterogeneity through Computationally Predictive Combination Therapy. *Cancer Discov*. **4**, 166–174 (2014).
21. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *British Journal of Cancer*. **108**, 479–485 (2013).
22. Alexander, S. & Friedl, P. Cancer invasion and resistance: interconnected processes of disease progression and therapy failure. *Trends in Molecular Medicine*. **18**, 13–26 (2012).
23. Farazi, T. A. *et al.* Bioinformatic analysis of barcoded cDNA libraries for small RNA profiling by next-generation sequencing. *Methods*. **58**, 171–187 (2012).
24. Rizzo, J. M. & Buck, M. J. Key Principles and Clinical Applications of ‘Next-Generation’ DNA Sequencing. *Cancer Prevention Research*. **5**, 887–900 (2012).
25. Robasky, K., Lewis, N. E. & Church, G. M. The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews Genetics*. **15**, 56–62 (2013).
26. Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews Cancer*. **9**, 153–166 (2009).
27. Heenan, E. J. & Sancar, A. DNA Damage: Repair. (John Wiley & Sons, Inc 2007).  
doi:10.1002/9780470048672.webc131
28. Raffoul, J. J., Heydari, A. R. & Hillman, G. G. DNA Repair and Cancer Therapy: Targeting APE1/Ref-1 Using Dietary Agents. *Journal of Oncology*. **2012**, 1–11 (2012).
29. Polo, S. E. & Jackson, S. P. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes Dev*. **25**, 409–433 (2011).
30. Forment, J. V. & O’Connor, M. J. *Accepted Manuscript. #startpage#*.  
<https://doi.org/10.1016/j.pharmthera.2018.03.005> (2018).
31. Harris, R. S. Cancer mutation signatures, DNA damage mechanisms, and potential clinical implications. **5**, 1–1 (2013).
32. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*. **173**, 291–3046 (2018).
33. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nature Genetics*. **47**, 1402–1407 (2015).
34. Watson, I. R., Takahashi, K., Futreal, P. A. & Chin, L. Emerging patterns of somatic mutations in cancer. *Nature Reviews Genetics*. **14**, 703–718 (2013).

35. Ferreira, M. A. *et al.* Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. *Nature Communications*. **1–18**, <https://doi.org/10.1038/s41467-018-08053-5> (2019).
36. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nature Reviews Genetics*. **15**, 585–598 (2014).
37. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature*. **500**, 415–421 (2013).
38. Burrell, R. A. & Swanton, C. ScienceDirectThe evolution of the unstable cancer genome. *Current Opinion in Genetics & Development*. **24**, 61–67 (2014).
39. WARBURG, O. Origin of Cancer Cells. *Science*. **123**, 309–314 (1956).
40. Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics*. **24**, 227–235 (2000).
41. Kittler, R. *et al.* A Comprehensive Nuclear Receptor Network for Breast Cancer Cells. *CellReports*. **3**, 538–551 (2013).
42. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nature Genetics*. **33**, 49–54 (2002).
43. Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nature Communications*. **7**, 1–15 (2016).
44. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell*. **173**, 400–41611 (2018).
45. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. **37**, 1–13 (2009).
46. Li, Y., Vongsangnak, W., Chen, L. & Shen, B. Integrative analysis reveals disease-associated genes and biomarkers for prostate cancer progression. *BMC Medical Genomics*. **7**, S3 (2014).
47. Chen, Y. *et al.* Identifying potential cancer driver genes by genomic data integration. *Sci. Rep.* **3**, (2013).
48. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. **13**, 2498–2504 (2003).
49. Newman, R. H. *et al.* Construction of human activity-based phosphorylation networks. *Mol. Syst. Biol*. **9**, 1–12 (2013).
50. Kamburov, A., Stelzl, U., Lehrach, H. & Herwig, R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res*. **41**, D793–800 (2013).
51. Anaya, J. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *PeerJ Computer Science*. **2**, e67 (2016).
52. Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res*. **44**, D336–42 (2016).
53. Kumar, P. P. *et al.* In-silico study reveals immunological signaling pathways, their genes, and potential herbal drug targets in ovarian cancer. *Informatics in Medicine Unlocked* **100422** (2020).

54. Warsi, M. K., Kamal, M. A., Baeshen, M. N., Izhari, M. A. & Mobashir, A. F. A. M. Comparative Study of Gene Expression Profiling Unravels Functions associated with Pathogenesis of Dengue Infection. *Current Pharmaceutical Design* **26 IS**, 1–8
55. Kamal, M. A. *et al.* Gene expression profiling and clinical relevance unravel the role hypoxia and immune signaling genes and pathways in breast cancer: Role of hypoxia and immune signaling genes in breast cancer. *jimsa* **1**(2020).
56. Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **1–22**(2017). doi:10.1038/nature20788
57. Biesecker, L. G. & Spinner, N. B. A genomic view of mosaicism and human disease. *Nature Reviews Genetics*. **14**, 307–320 (2013).
58. Aguilera, A. & García-Muse, T. Causes of Genome Instability. *Annu. Rev. Genet.* **47**, 1–32 (2013).
59. Speleman, F. *et al.* Copy number alterations and copy number variation in cancer: close encounters of the bad kind. *Cytogenet Genome Res.* **123**, 176–182 (2009).
60. Gaillard, H., García-Muse, T. & Aguilera, A. Replication stress and cancer. *Nature Reviews Cancer*. **15**, 276–289 (2015).
61. Roberts, S. A. & Gordenin, D. A. Hypermutation in human cancer genomes: footprints and mechanisms. *Nature Reviews Cancer*. **14**, 786–800 (2014).
62. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nature Reviews Genetics*. **15**, 585–598 (2014).
63. Nik-Zainal, S. *et al.* Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell*. **149**, 979–993 (2012).
64. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering Signatures of Mutational Processes Operative in Human Cancer. *CellReports*. **3**, 246–259 (2013).
65. Agell, L. *et al.* A 12-Gene Expression Signature Is Associated with Aggressive Histological in Prostate Cancer. *The American Journal of Pathology*. **181**, 1585–1594 (2012).
66. *A generic methodological framework for studying single cell motility in high-throughput time-lapse screening data.* (The Febs ... doi:10.1093/bioinformatics/btv225/-/DC1
67. Stelzl, U. *et al.* A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell*. **122**, 957–968 (2005).
68. Stoevesandt, O. *et al.* A Network Analysis of Changes in Molecular Interactions in Cellular Signaling. *Molecular & Cellular Proteomics*. **6**, 503–513 (2006).
69. Fujiki, Y. *et al.* APOBEC3B gene expression as a novel predictive factor for pathological complete response to neoadjuvant chemotherapy in breast cancer. *Oncotarget*. **9**, 30513–30526 (2018).
70. Mao, Y. *et al.* APOBEC3B expression and its prognostic potential in breast cancer. *Oncol Lett*. <https://doi.org/10.3892/ol.2020.11433> (2020).
71. Kim, Y. S. *et al.* Clinical implications of APOBEC3A and 3B expression in patients with breast cancer. *PLoS ONE*. **15**, e0230261 (2020).

72. Zhang, Y., Delahanty, R., Guo, X., Zheng, W. & Long, J. Integrative genomic analysis reveals functional diversification of APOBEC gene family in breast cancer. *Hum. Genomics*. **1–12**, <https://doi.org/10.1186/s40246-015-0056-9> (2015).
73. Matsumoto, T. *et al.* protein kinase A inhibits tumormutator APOBEC3B through phosphorylation. *Sci. Rep.* **1–12** <https://doi.org/10.1038/s41598-019-44407-9> (2019).
74. Yamazaki, H. *et al.* APOBEC3B reporter myeloma cell lines identify DNA damage response pathways leading to APOBEC3B expression. *PLoS ONE*. **15**, e0223463 (2020).
75. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell*. **173**, 321–33710 (2018).
76. Guttridge, K. H. F. D. G. D., Glass, D. J. & Guttridge, D. C. Cancer Cachexia: Mediators, Signaling, and Metabolic Pathways. *Cell Metabol.* **16**, 153–166 (2012).
77. Brechbiel, J., Miller-Moslin, K. & Adjei, A. A. Crosstalk Between Hedgehog and Other Signaling Pathways as a Basis for Combination Therapies in Cancer. *Cancer Treat. Rev.* **40**, 750–759 (2014).
78. Eroles, P., Bosch, A., Pérez-Fidalgo, J. A. & Lluch, A. Cancer Treatment Reviews. *Cancer Treat. Rev.* **38**, 698–707 (2012).
79. da Silva, H. B. *et al.* Dissecting Major Signaling Pathways throughout the Development of Prostate Cancer. *Prostate Cancer*. **2013**, 1–23 (2013).
80. Oren, M. & Aylon, Y. The Hippo Signaling Pathway and Cancer. (Springer Science & Business Media 2013).
81. Altomare, D. A. & Testa, J. R. Perturbations of the AKT signaling pathway in human cancer. *Oncogene*. **24**, 7455–7464 (2005).
82. Wilson, F. H. A Cluster of Metabolic Defects Caused by Mutation in a Mitochondrial tRNA. *Science*. **306**, 1190–1194 (2004).
83. Bansal, M. *et al.* a community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol.* **1–12**, <https://doi.org/10.1038/nbt.3052> (2014).
84. Nakamura, H. *et al.* Genomic spectra of biliary tract cancer. *Nature Genetics*. **1–11**, <https://doi.org/10.1038/ng.3375> (2015).
85. Real, F. X., Boutros, P. C. & Malats, N. Next-generation Sequencing of Urologic Cancers: Next Is Now. *Eur. Urol.* **66**, 4–7 (2014).
86. Mobashir, M., Schraven, B. & Beyer, T. Simulated evolution of signal transduction networks. *PLoS ONE*. **7**, e50905 (2012).
87. Mobashir, M., Madhusudhan, T., Isermann, B., Beyer, T. & Schraven, B. Negative Interactions and Feedback Regulations Are Required for Transient Cellular Response. *Sci. Rep.* **4**,
88. nez, F. A. J., nez, R. J. & Cobaleda, C. Seminars in Cancer Biology. *Seminars in Cancer Biology*. **20**, 98–106 (2010).
89. Ferrell, J. E. Jr. The Biochemical Basis of an All-or-None Cell Fate Switch in *Xenopus* Oocytes. *Science*. **280**, 895–898 (1998).

90. Chatterjee, A., Kaznessis, Y. N. & Hu, W. S. Tweaking biological switches through a better understanding of bistability behavior. *Current Opinion in Biotechnology*. **19**, 475–481 (2008).
91. Iwamoto, K., Hamada, H., Eguchi, Y. & Okamoto, M. Mathematical modeling of cell cycle regulation in response to DNA damage: Exploring mechanisms of cell-fate determination. *Biosystems*. **103**, 384–391 (2011).
92. Mustafa, S. & Mobashir, M. LC-MS and docking profiling reveals potential difference between the pure and crude fucoidan metabolites. *International Journal of Biological Macromolecules*. **143**, 11–29 (2020).

## Table 1

Table 1: Cox regression analysis for APOBEC3B.

### Cox regression results for APOBEC3B

Cancer	Cox Coefficient	P-Value	FDR Corrected	Rank	Median Expression	Mean Expression
BLCA	0.004	9.60E-01	9.82E-01	15906	362.19	603.54
BRCA	0.112	2.10E-01	5.89E-01	5758	162.02	321.98
CESC	-0.608	1.30E-05	2.55E-02	8	755.45	924.17
COAD	-0.094	3.40E-01	7.24E-01	7557	214.67	245.31
ESCA	-0.048	7.20E-01	9.84E-01	12129	239.99	314.7
GBM	0.004	9.60E-01	9.92E-01	16161	67.15	86.21
HNSC	-0.028	7.10E-01	8.90E-01	13181	467.44	691.26
KIRC	0.19	2.40E-02	6.18E-02	6421	61.4	80.76
KIRP	0.59	5.50E-04	6.78E-03	1328	50.53	73.29
LAML	-0.018	8.70E-01	9.58E-01	13732	160.13	206.39
LGG	0.216	2.00E-02	4.32E-02	7720	17.17	25.49
LIHC	0.093	2.90E-01	5.53E-01	8176	83.16	161.41
LUAD	-0.015	8.40E-01	9.23E-01	15194	135.46	255.45
LUSC	-0.117	7.80E-02	6.10E-01	2146	491.14	651.6

## Figures



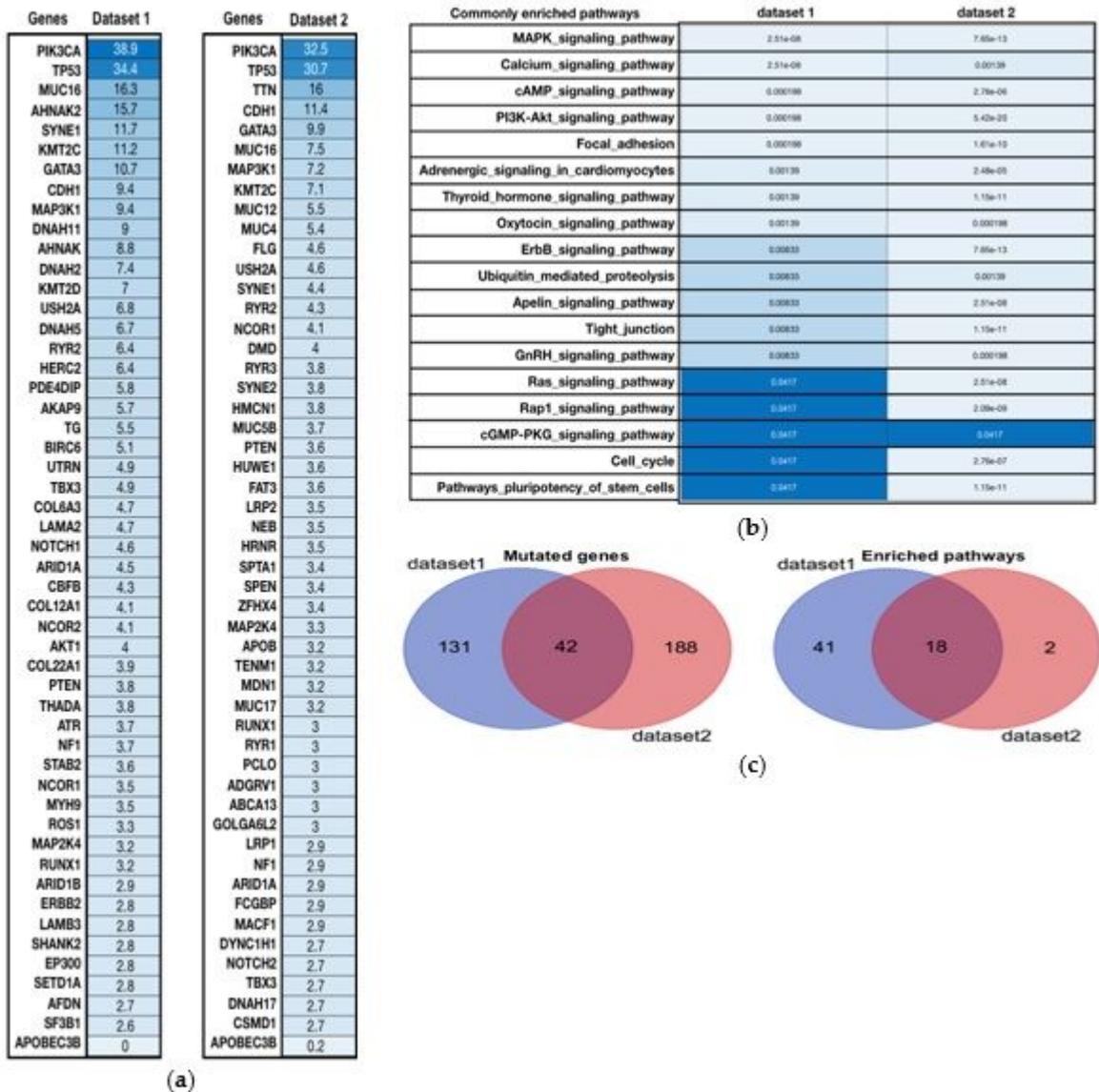


Figure 2

Breast cancer mutation and its functional impact. (a) Top ranked mutated genes in the clinical breast cancer samples (TCGA database), (b) commonly enriched pathways, and (c) venn diagram to present the commonly top ranked mutated genes and the enriched pathways among the TCGA database.

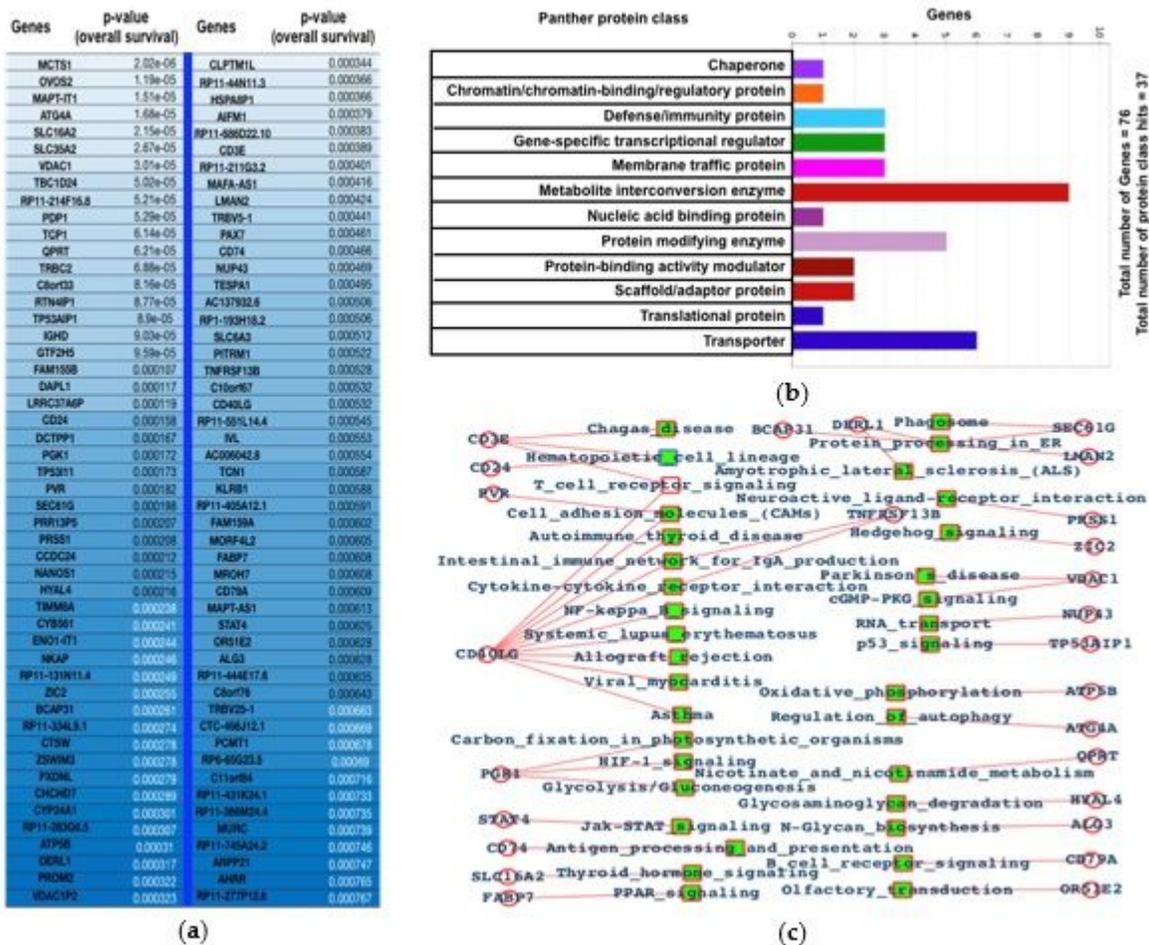


Figure 3

Clinically significant genes and their functional classification. (a) Top 100 genes which are highly significant in terms of the overall survival as presented here the respective p-values. (b) Panther classification of the genes. (c) Pathway–gene association network of the top 100 clinically significant genes.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryDataS1.xlsx](#)
- [SupplementaryDataS2.pdf](#)
- [SupplementaryDataS3.pdf](#)
- [SupplementaryTable1.docx](#)