Anatomical Studies of Genetic Variation Instabilities in Various Cancer

Mohammad Ghazimoradi
Tarbiat Modares University
Shirin Farivar
Shahid Beheshti University
Ehsan Zolghadr
University of Alabama, Tuscaloosa
Samaneh Montazeri
Avicenna Research Institute
Toktam Tavabe Ghavami
Tarbiat Modares University
Maryam Daryani
Tarbiat Modares University
Sadegh Babashah (babashah@modares.ac.ir)
Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, P.O. Box: 14115-154, Tehran, Iran.

Research Article

Keywords: Genome instability, Cancer, SNP

Posted Date: January 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-130530/v1

Abstract

Background: Genome instability is one of the significant markers of cancers. This features is one of the most fundamental mechanism regarding cancer cells evolution. This major mechanism has been found mostly in some of cancer types and in less extends in other types. Majority of this instability occur mostly in chromosome scale or satellites.

Results: In this regards whole exome data has been downloaded from Array express (EMBL-EBI). We investigate the amount of instability of genetic variations such as SNP, MNP and other types in various cancers. We also investigates this change in genome, chromosome and gene scale point of view in various type of cancers. Our findings might enlighten some preservative mechanism in genome scale.

Conclusion: Although genome instability on chromosomal level is well studied and has been proved, on micro scale genomic variation it might not be the case. The positive control and negative show right pattern, however on other cancers from different stages and grades the instability could not be confirm on micro scale instabilities.

Introduction

Genomic instability is one of the worst features of cancer cells (1). Although understanding of the underlying mechanism and molecular basic of this significant pathway is vitally important, a little knowledge has been obtaining in this branch of science (2). There are different types of genomic instability. Most cancers obtain chromosomal instability (CIN), refers to the high rate of alteration in chromosome structure and number in cancer versus normal cells (3). Actually it might be consider as an evolutionary approach for cancer cells to be alive and successful in the hostile environment such as our body for them (4). Abnormal chromosome structures (5) and numbers alteration associated with abnormal mitoses (6) have seen frequently. Of these chromosomal changes observed in some cells of a tumor but not in all, suggesting heterogeneity in cancer cells are the legacy of a genetically instable single cell, which acquire chromosomal abnormalities evolutionary (7, 8). It is notable that presence of CIN has also been proven in cancer cells in-vitro. Regardless of chromosomal instability which is the major form of genomic instability, other types of instability have also been reported (9, 10). One of these shifts are genomic instability, characterized by increased frequencies of base pair mutation (11). Evidence obtained by heredity cancers researches shows that loss of DNA repair genes, cell cycle checkpoints, mitosis regulators and many more pathways will increased frequencies of base pair mutation including single nucleotide polymorphism (SNP), multiple nucleotide polymorphism (MNP) and many mores (12). For example, hereditary MYH-associated polyposis, in MYH, a DNA base excision repair (BER) gene, results in increased SNPs. Microsatellite instability (MSI) characterized by change of the number of oligonucleotide repeats in tandem nucleotide repeats, repetitive motifs of 1 to 6 nucleotides (12, 13). MSI has been observed many cancers. An analysis indicates mixed result of MSI as element of prognosis in colorectal, gastric, pancreatic and esophageal cancers but a poor one in non-small cell lung cancer. SNPs are one of the evolutionary changes which could boost organism and cell to progress and cancer cells are no

exceptions. SNPs could induce naturally and spontaneously to a resultant phenomenon of DNA repair system malfunction. Regardless of causality, their position in chromosomes is determining their role. SNPs could induce in exome, introns, 3-UTR, 5-UTR and other various position in genomic content (14, 15, and 16). Generally, any change in expression or function of any transcript and gene in favor of disrupting cancerous features could be beneficial for induction and progression of cancer and malignancies. The exonal SNPs affect cancer by suppressing or overexpression of gene transcription and translation. SNPs in intron regions could affect splice variants or disrupt binding and function of long non-coding RNAs which could even indirectly upregulated other genes. SNPs in the 5'-UTR change rate of translation, although SNPs in the 3'-UTR could affect microRNA (miRNA) binding to their targets. SNPs in up or downstream of genes could boost or reduce transcription of genes affecting cis or Trans elements (17, 18).

The promoter region SNP could induce down or upregulation of the transcript via alteration in binding site affinity for other proteins such as TATA box. Intronal sequences of cis-acting regulatory elements could regulate expression of genes. The 5' and 3' UTRs of mRNAs control translation. The 5'-UTR regulates translation but the 3'-UTR determines mRNA stability. Not to mention several nucleotide from each region is most important and have been neglected by many investigations. Other significant changed caused by SNPs are protein structural shifts. If a change could occur in gene coding part, a structural change could induce in protein structure. This changes could have no significant changes in protein function such as synonymous SNP, literally no effect, to termination of a protein function such as missense (17, 18, and 19). It would consist result to assume gain of function of oncogene, or loss of function of a tumors suppressor could induce cancer and malignancy which has been seen in many genes such as TP53 and RB (20).

It has been suggested all of the cancer cells have not same rate of instability with a low amount in squamosal head and neck cancer and high variation in blood related. Although the quantification and even limits of these change have been remaining to elucidate. It is notable that analysis of population of cancer cells which show heterogeneity which has been proved by single cells could be intriguing and cover many aspect of realities (21). However understanding the anatomy and underlying of cancer as single unit regardless of their micro view complexity could elucidate many question which would be beneficial for understanding cancers better.

Material And Methods

Data collection

16 samples of various cancer samples with diverging sources has been collected. Other 10 samples have been chosen from healthy donors and in some extends normal tissue adjust to cancer cells. Accession number of samples exist in Table 1. Whole exon of samples has been downloaded from ENA. And proceed for the analysis.

Table 1
Accession number of entries

1	**ERR1141789**	**10**	**ERR2528785**	**19**	**ERR3012003**
2	ERR1141790	11	ERR3426262	20	ERR1663324
3	ERR1429282	12	ERR3426261	21	ERR637414
4	ERR1629990	13	ERR3426263	22	ERR166329
5	ERR1429984	14	ERR2990062	23	ERR166323
6	ERR1629991	15	ERR2990069	24	ERR166322
7	ERR2278800	16	ERR2990152	25	ERR166319
8	ERR2528757	17	ERR3426265	26	ERR166317
9	ERR2528759	18	ERR3012007		

Whole Exon Analysis

The data have been QC by FastQC software (22) and trimmed by trimmomatic (23) if it is needed. The data have been aligned to human reference genome (hg19) by Bowtie2 (24). Variant calling has been done by Freebayes (25) and Genetic variant annotation and functional effect prediction has been done by SnpEff and SnpSift (26).

Graph Design

Graphs has been prepared by GraphPad prism 8. Heatmap has been set by heatmaper.

Statically Analysis

Statically analysis has been performing by GraphPad prism 8. T tests unpaired and two-way analysis has been performing as it fits in tests. It is notable that P values has been presented in text. For ROC curve has been performing by Wilson/Brown method. For analyzing of heatmap, Manhattan has been chosen as distance measurement method and average linkage as clustering method.

Results

To investigate an over view of genome variation in cancer 16 samples of whole exome of various cancer types has been downloaded (Table 1). The cancer group contains squamosal head and neck, hepatocellular carcinoma, acute myeloid leukemia, lymphoma, NK Malignancy, pleura lung cancer, gallbladder adenocarcinoma cancer Type. Among them head and neck squamosal cancer as more stable

cancer type and blood cancer as the most instable cancer type has been choose as well to validate our analysis. Comparison among cancer and non-cancerous samples have been done. It is notable that non-cancerous samples include normal samples in addition to normal adjust cancer tissues. Moreover, our analysis has been shown that number of occurred variation in cancers are related to their type. In comparison with normal samples which their incidence is more concentrate and predictable, cancer show a high degree of divergence with highest number in blood related Cancers and lowest in cancer (Fig. 1A). After this step analysis of chromosomes of each sample seems crucial. In this essence analysis of chromosomal of cancer versus non-cancerous samples have been done. Although there was no significant result in Y chromosomes, X chromosome on the other hand shows more instability in control samples (Fig. 1B). The same situation happen with other chromosomes in these comparisons. The average of all of chromosomes of normal samples from 1 to 22, with no expectation, were more instable and significantly differ (P value < 0.0001) (Fig. 1C). To investigate the stability of all chromosomes of each sample, we draw violin plat of each sample based on their stability of their chromosomes (Fig. 1D). Once again stability of the normal group were significantly lower than cancer group (P value < 0.05). Although the two control groups of cancer such as squamosal of head and neck and blood related cancers show the most stable and instable type in all of the samples subsequently.

To understand the underlying mechanism of this variation, analysis of their types, functions and other features were important. These variations have been occurred as SNP frequently in both cancerous and non-cancerous type. It is interesting MNP occur more in cancer type and SNP slightly higher in normal cells. Other type of variation including insertion, deletion and mix changes were significantly lower that other two types. It is interesting that mix type has been occurred in cancer types more occasionally (Fig. 2A). The most interesting data of ours are related to the position of variation in genes. More than 60 percent of variation occur on exon and intron in both groups however interionic variation are significantly higher in cancer group than normal one (P value = 0.0082). In opposite changes in exonal segments are less in cancer group than normal type (P value = 0.0033) (Fig. 2B). Other group have less share in these variations except intragenic segment which cancer cell show more instability. The impact of variation has been classified to 4 groups containing the modifier, moderate, high and low. Each group represented in supplementary table1. Modifier changes affect more cancer group (P value = 0.0033) (Fig. 2C). Interestingly high changes were fewer in cancer groups (P value = 0.0086). Percentage of missense were slightly lower in cancer group versus the control group. Instead (P value = 0.0088), variation result in silencing of protein were slightly higher in cancer group (P value = 0.0411) (Fig. 2D).

To obtain deeper insights, we sort genes base on their variation index as high impact and determine their functional pathways. Genes have been categorized and classified. It has been found that more than 100 genes with high impact changes has been belonged to metabolic pathways. Further pathways such as signaling, proteoglycans and viral carcinogenesis were also significant as it has been shown (Fig. 3A). More over a list of genes whom their present was dominated in all of the cancerous samples has been prepared, and a network of them has been prepared (Table2 and Fig. 3B).

Table 2
Genes most appear in cancer group with high impact effects

CHI3L1	WARS	PTPRB	ALDH4A1	STAG2	TLR8	CHI3L2
FOLH1	LDHA	FASN	CYP11B2	XYLB	TP53	ADAM17
NMRK1	NOTUM	IDUA	PCSK9	TP73	MEN1	PTK2
CES1	ADSL	BLVRB	EGFR	PPIA	LGALS8	MAPKAPK2
SYK	UBC	NEU2	PNP	IDE	DPP4	CBS
HDAC8	GAPDH	FDFT1	EPHX2	RAB8A	OAT	FURIN
CSF1R	PKM	NAGA	PDE4B	CHIA	PPP5C	KDR
HRAS	PTGR1	CSNK1D	GRHPR	MAOB	TYMP	PFKP

Among candidates, TP53 were the key to cancer. Two significant pathways also has been link to TP53. First were metabolic pathway including FASIN and LDHA which could regulate energy pathway. Other was MEN1, a tumor suppressor (Fig. 3B and C).

In the next step, we were wondering if the sample could be classified based on their variation on their chromosomes. Furthermore, an over view of each sample based on their chromosomes also seems curial. In this essence, their heatmap has been set. It is interesting that heatmap could categorize most of the samples in their groups. In addition, most of the normal and cancerous groups show the same signature of each other (Fig. 4A). Furthermore, we perform a ROC cure analysis to see if this method could predict cancerous cells efficiently. The P-Value of Roc curve were 0.1742 with area of 0.6667 (Fig. 4B).

Discussion

Genome instability is one of the most devastating features of cancers. It has appeared that many cancer stem cells or in other word progenitors of malignancy have risen and evolve from this features.Although Genome instability is well known generally based on chromosomes instability and microsatellite rearrangements, little known regarding variations in the micro environment of the genome such as SNP, MNP and many more. Here we selected several cancer types and compare their genomic variation with each other and normal genomes. Our data suggest that there are no meaningful relationships between two groups. Although different cancer show Insurgent among them self and normal group an interesting consistency. This data show that micro variation in the genome of cancers generally is not in accordance with or MSI which is intriguing. It could be beneficial to group cancer sample in stages or grades however most of samples have lesser incidence than mean of control group which might disprove this hypothesis. We believed this data represented a population view of cancer sample. This occurrence might be due to cancer type which the molecular pathway in cancer cells trying to suppress them even in the middle of cancer. This data also could be seen in chromosomes study of cancers versus normal samples which

frankly is clearer. Other significant findings would be reducing of SNP and increase of MNP in cancer group.The increase intronal change and diminishing exonal change in cancer group also is one of the signs of suppression in cancer cells. The same situation goes for high impact and low impact changes. It is interesting that stages and grade of cancer have not been effective in our analysis which need future investigation. Pathway analysis of genes with most number of changes in order of high impact changes show meaningful changes in metabolism pathways. They might be outstanding target due to their high number genes and low value of each of them with eventually could changes many. In our study once more the impact of TP53 has been seen. Two pathways also could have most influence including MEN1 and metabolic pathway. In the last cancer group could be clustered by their chromosomal changes in right order.

Declarations

Ethics approval and consent to participate

The data has been downloaded and used from array express online database under data access policy and release policy of Array express and EMBL-EBI. All of the used data were public and could be used by others under array express and EMBL-EBI policy.

Consent for publication

Not-applicable.

Availability of data and materials

All of the data has been downloaded from Array express (EMBL-EBI) with mentioned accession number in table 1. The protocols and programs used for analyzing of the data has been provided in material and methods section.

Competing interests

none of the authors declare any financial or non-financial conflict of interest.

Funding

None.

Authors' contributions

Mohammad H Ghazimoradi conceive the idea and perform the analysis. Shirin Farivar and maryam Daryani Wrote the article and perform statistical analysis. Ehsan Zolghadr and Toktam Sadat Tavabe Ghavami edit article and perform graph design. In addition they validated the data. Samaneh montazeri collected the data. Sadegh Babashah mange the project and supervise it.

Acknowledgment

We humbly thank Dr Mohammad Mersad Ghorbani for his valuable guidance and lessons.

Data availability

All data has been downloaded from Array express (EMBL-EBI) and avaible with accession number mentioned in table 1. Generated or analyzed data during this study are included in this published article and its supplementary table files 1.

References

1. Aguilera, A., & Gómez-González, B. (2008). Genome instability: a mechanistic view of its causes and consequences. *Nature Reviews Genetics*, *9*(3), 204–217.
2. Aguilera, A., & García-Muse, T. (2013). Causes of genome instability. *Annual review of genetics*, *47*, 1–32.
3. Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability—an evolving hallmark of cancer. *Nature reviews Molecular cell biology*, *11*(3), 220–228.
4. Merlo, L. M., Pepper, J. W., Reid, B. J., & Maley, C. C. (2006). Cancer as an evolutionary and ecological process. *Nature reviews cancer*, *6*(12), 924–935.
5. Rajagopalan, H., & Lengauer, C. (2004). Aneuploidy and cancer. *Nature*, *432*(7015), 338–341.
6. Duensing, A., & Duensing, S. (2010). Centrosomes, polyploidy and cancer. In *Polyploidization and Cancer* (pp. 93–103). Springer, New York, NY.
7. Shackleton, M., Quintana, E., Fearon, E. R., & Morrison, S. J. (2009). Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell*, *138*(5), 822–829.
8. Burrell, R. A., McGranahan, N., Bartek, J., & Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, *501*(7467), 338–345.
9. Moon, S. H., Kim, J. S., Park, S. J., Lim, J. J., Lee, H. J., Lee, S. M., & Chung, H. M. (2011). Effect of chromosome instability on the maintenance and differentiation of human embryonic stem cells in vitro and in vivo. *Stem cell research*, *6*(1), 50–59.
10. Rogers, S., McCloy, R. A., Parker, B. L., Gallego-Ortega, D., Law, A. M., Chin, V. T., … Deng, N. (2018). MASTL overexpression promotes chromosome instability and metastasis in breast cancer. *Oncogene*, *37*(33), 4518–4533.
11. Bielas, J. H., Loeb, K. R., Rubin, B. P., True, L. D., & Loeb, L. A. (2006). Human cancers express a mutator phenotype. *Proceedings of the National Academy of Sciences*, *103*(48), 18238–18242.
12. Thompson, S. L., Bakhoum, S. F., & Compton, D. A. (2010). Mechanisms of chromosomal instability. *Current biology*, *20*(6), R285-R295.
13. Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N., & Szallasi, Z. (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nature genetics*, *38*(9), 1043–1048.
14. Thompson, S. L., Bakhoum, S. F., & Compton, D. A. (2010). Mechanisms of chromosomal instability. *Current biology*, *20*(6), R285-R295.

15. Venkatesan, S., Natarajan, A. T., & Hande, M. P. (2015). Chromosomal instability—mechanisms and consequences. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, *793*, 176–184.
16. Kaufmann, W. K., Carson, C. C., Omolo, B., Filgo, A. J., Sambade, M. J., Simpson, D. A., … Thomas, N. E. (2014). Mechanisms of chromosomal instability in melanoma. *Environmental and molecular mutagenesis*, *55*(6), 457–471.
17. Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, *34*(3), 275–305.
18. Komar, A. A. (2007). Silent SNPs: impact on gene function and phenotype.
19. Liao, P. Y., & Lee, K. H. (2010). From SNPs to functional polymorphism: The insight into biotechnology applications. *Biochemical Engineering Journal*, *49*(2), 149–158.
20. Deng, N., Zhou, H., Fan, H., & Yuan, Y. (2017). Single nucleotide polymorphisms and cancer susceptibility. *Oncotarget*, *8*(66), 110635.
21. Pikor, L., Thu, K., Vucic, E., & Lam, W. (2013). The detection and implication of genome instability in cancer. *Cancer and Metastasis Reviews*, *32*(3–4), 341–352.
22. Andrews, S., & FastQC, A. (2015). A quality control tool for high throughput sequence data. 2010.
23. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
24. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357.
25. O. Tange. (2011). GNU Parallel - The Command-Line Power Tool.;;*Login: The USENIX Magazine*, *36*(1), 42–47.
26. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118;iso-2; iso-3. *Fly*, *6*(2), 80–92.

Tables

Supplementary Table 1 is not available with this version.

Figures

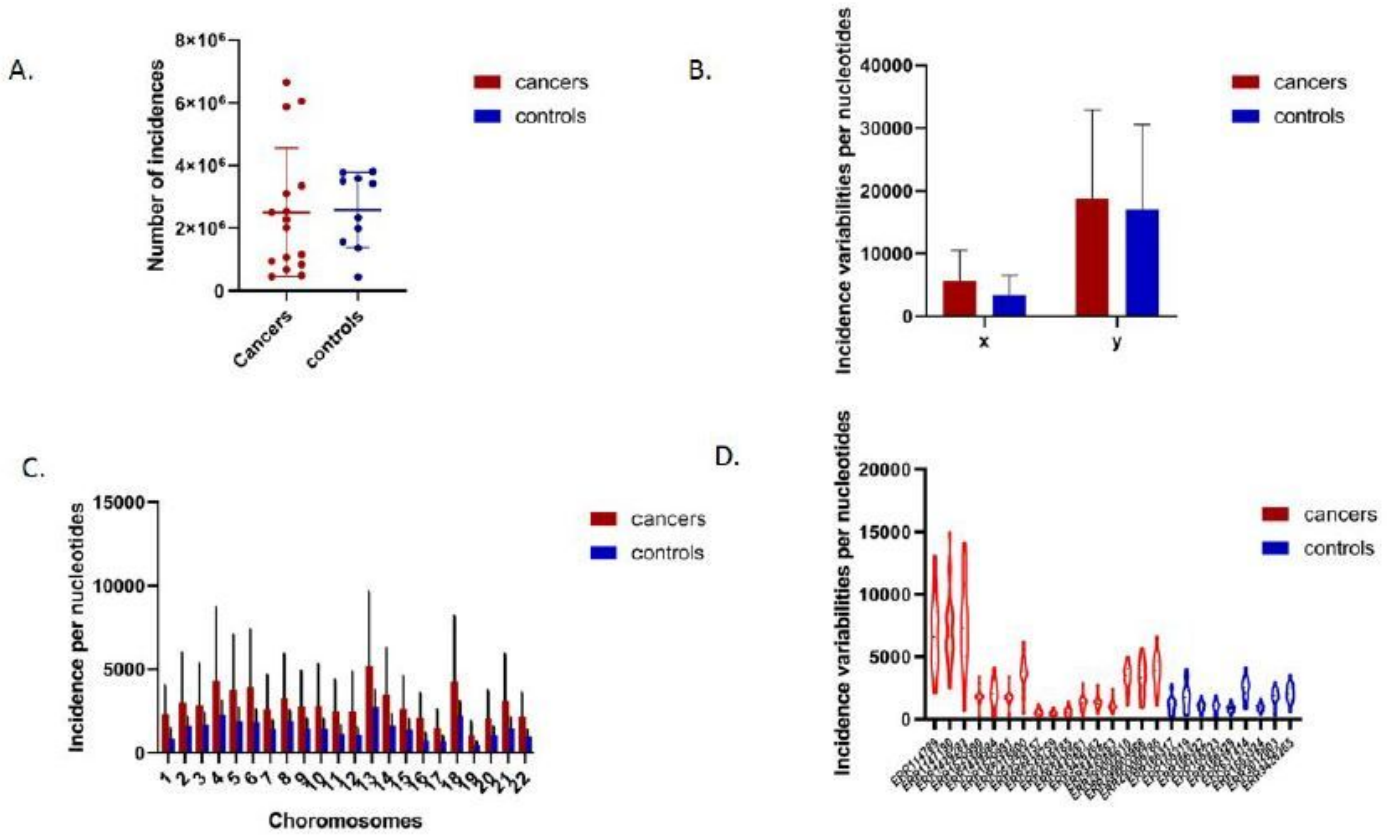


Figure 1

An overview of genomic variations. A. Number of incidents per samples in cancerous and non-cancerous groups. B. variation in sexual chromosomes. C. variation in somatic chromosomes. D. stability of each sample.

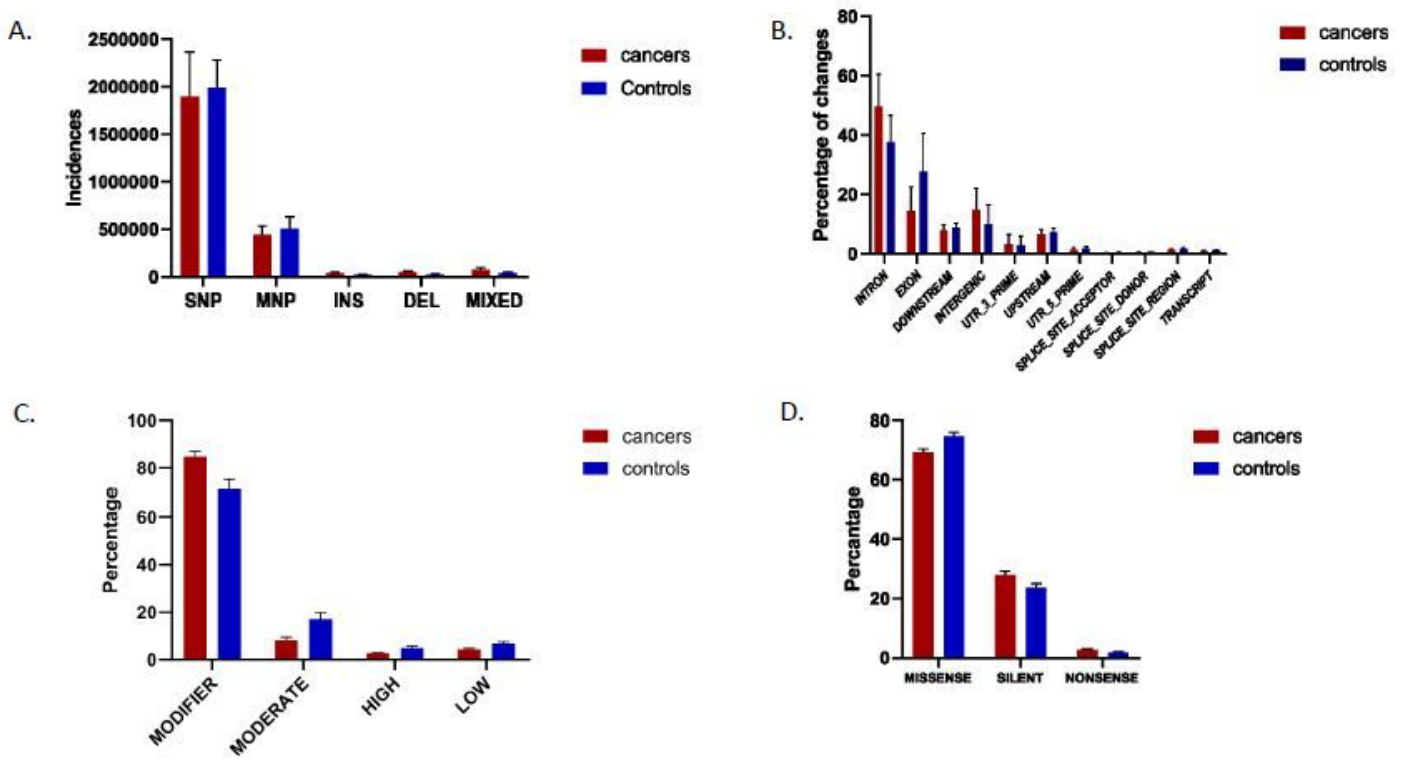
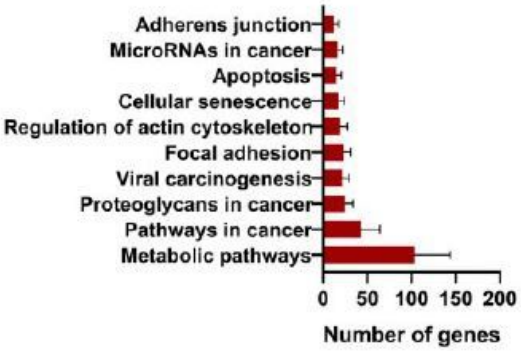


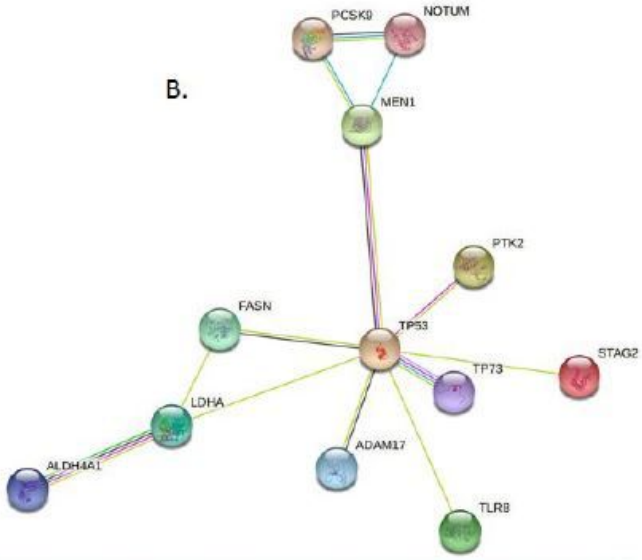
Figure 2

Classification of variations. A. type of each variations. B. Position of each variations on genes. C. Effects of each variations. D. Function classification of each variation.

A.



B.



C.

Pathway	Description	Count in network	Strength	False discovery rate
hsa00250	Alanine, aspartate and glutamate metabolism	2 of 35	1.73	0.0303
hsa05230	Central carbon metabolism in cancer	2 of 65	1.46	0.0493
hsa04115	P53 signaling pathway	2 of 68	1.44	0.0493
has05202	Transcriptional misregulation in cancer	3 of 169	1.22	0.0303
hsa01100	Metabolic pathway	8 of 1250	0.78	0.0023

Figure 3

Pathway analysis of cancer grups. A. classification of varitions in cancers. B. Network of most affected genes. C. Pathway analysis of netwok.

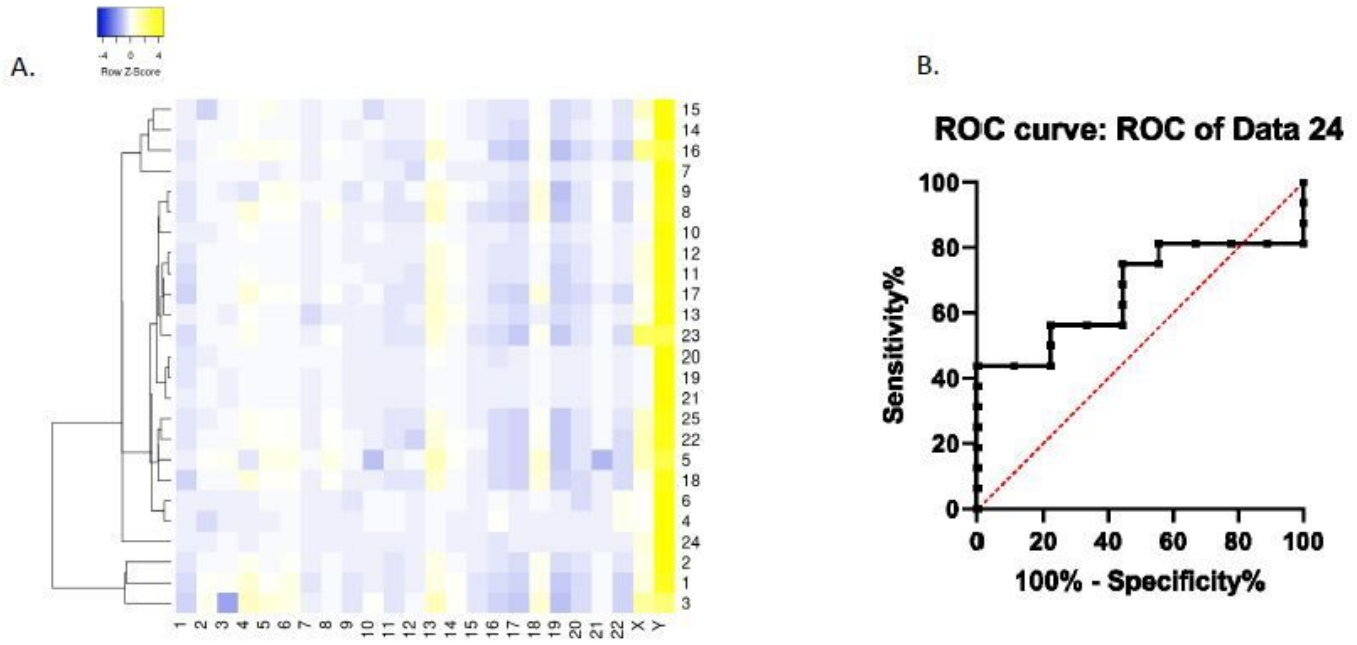


Figure 4

Clustering and Roc curve of samples. A. Heatmap of samples and their clustering (Based On number of entries). B. ROC curve chromosomes stability in cancerous and non-cancerous groups.