Understanding the Maintenance of Extrachromosomal DNA in Cancer

CSE280A Team 3
Jianing Wang, Hanqing Zhao

Introduction

Extrachromosomal DNA (ecDNA) is DNA that is found off the chromosomes, and it is a subtype of extrachromosomal circular DNA (eccDNA), which is known to have a circular structure. Based on recent studies of ecDNA, it is discovered to be a potential mechanism of tumor formation due to oncogene amplification or copy number alteration (CNA) in ecDNA. (Verhaak et al, 2019). Moreover, unequal segregation of ecDNA from parental to offspring tumor cells increase tumor heterogeneity, potentially providing an evolutionary advantage (Verhaak et al, 2019). However, understanding ecDNA still remains challenging in cancer research. In our project specifically, we focus on comparing mutations/CNA in cancer patients samples with(+)/without(-) ecDNA, and discover differentially mutated genes in ecDNA(+) and ecDNA(-) samples.

Methods

Data preparation

Patient sample data are downloaded from cBioPortal with TCGA (The Cancer Genome Atlas) PanCancer Atlas Studies, which is a comprehensive analysis of over 11,000 tumors from 33 of the most prevalent forms of cancer. We query over 63 tumor suppressor genes in the list we found online, selecting all mutation and copy number alterations (CNA) molecular profiles, and download the oncoprint data directly from the query results.

In the patient sample oncoprint data we have downloaded, the columns represent patient IDs and the rows represent the 63 tumor suppressor genes. There are also different tracks in the dataset, such as CNA, mutations and fusion. Under each track, we have the specific type of CNA/mutations of that track.

Another major matrix containing the TCGA patient barcode and corresponding ecDNA classification is obtained from Professor Bafna. Since the patient barcode and the patient ID in the oncoprint dataset is consistent, we can divide our patient sample into ecDNA+ and ecDNA- groups. Based on characterization of ecDNA, we have two ways to make comparison groups: One based on whether it is ‘circular’; one based on whether it has copy number alteration (CNA). So we end up having two comparison groups, one treats ecDNA+ as circular, ecDNA- as the rest that is not circular, and one treats ecDNA+ as circular, ecDNA- as no SCNA detected.

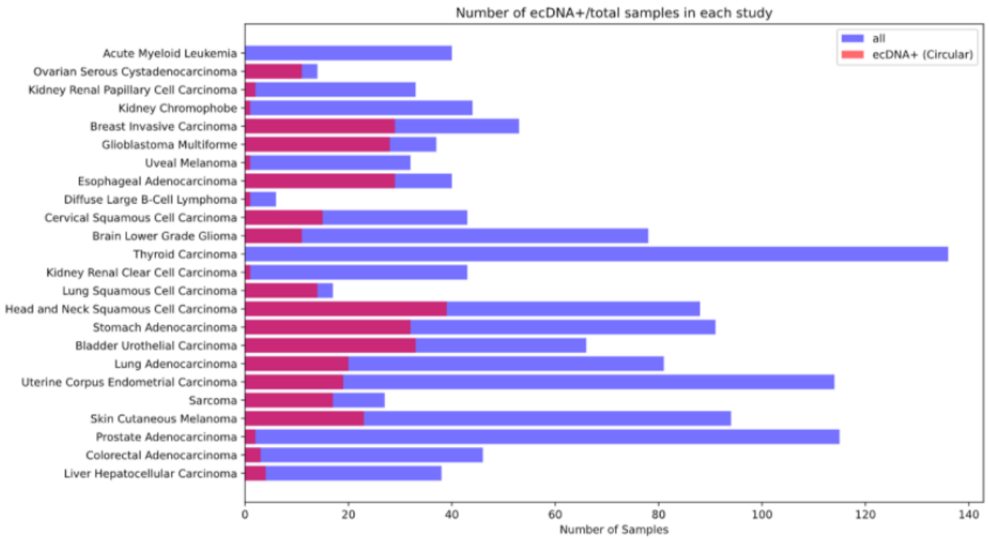

Fig. 1 Bar plot of the number of patients we used for different cancer studies in TCGA. Large difference has shown between the number of patients with ecDNA+ and the total number after filtering.

Before we process the dataset, we generate a graph showing the type of dataset in TCGA for our study after filtering with 63 tumor suppressor genes and the patients available with ecDNA classification. As we can see, some cancer studies have very few or no patients with ecDNA+, and the total number of ecDNA+ patients is not even half of the total patients. Since we combined all the studies together, we could have ignored the fact that the same gene could behave differently in different cancer types, and the uneven distribution of samples may compromise our results.

Data processing

After filtering the patient samples oncoprint with ecDNA classification information, we end up having 1900 patient samples. Then, we approximately define the general standard for classifying each type of mutations and CNA:

Loss of Function (LoF)	Gain of Function (GoF)	Ambiguous
Missense Mutation (putative driver)	Amplification	splice
Truncating Mutation (putative driver)	amp_rec	splice_rec
Truncating Mutation (putative passenger)		Inframe Mutation (putative passenger)
Inframe Mutation (putative driver)		Missense Mutation (putative driver)
Deep Deletion		
homdel_rec		

Table 1: Loss of Function, Gain of Function and Ambiguous Standard for Mutations

The general standard we have is that, for putative driver mutations, we assume it has a higher potential of LoF of TSG which therefore promotes cancer development. For putative passenger mutations, we think it is of unknown significance to cancer so we treat some of them as ambiguous. Truncating and deletion are assumed to be LoF regardless of driver or passenger. Amplification is assumed to be GoF but could contain errors due to insufficient information. Finally, splice is assumed to be ambiguous after consultation with Prof. Bafna.

According to the classification of LoF and GoF in mutations & CNA, and fusion, we next generate 3 matrices with 0 and 1: Matrix with '1' if there is a LoF, '0' otherwise; Matrix 2 with '1' if there is a GoF, '0' otherwise, and Matrix 3 with '1' if there is a fusion, '0' is no fusion. Since fusion data is very sparse, we have a different way of generating this matrix which will be discussed later.

Statistical analysis

To identify genes that are preferentially mutated in the ecDNA+ (or ecDNA-) samples, it is necessary to apply statistical testing. The null hypothesis, for each gene, would be that it is not differentially mutated in ecDNA+ and ecDNA- samples. First, each row in the corresponding matrix is converted into a contingency table, which is later used for Fisher's Exact test. For fusion, we noticed that the matrix is very sparse, we combine all genes into one single row. Each non-zero entry indicates at least one of the genes that undergo some gene fusion events in the corresponding sample.

After we obtain contingency tables, we first apply two-sided Fisher's Exact Test for each gene to test if any of the genes is differentially mutated. Adjusted p-values less than 0.05 is considered as significant in this study, and we are able to classify a set of genes. Then, we perform the one-sided Fisher's Exact test, with the alternative hypothesis that the gene is preferentially mutated in ecDNA+ samples. Therefore, we can isolate genes that tend to mutate in ecDNA+ samples among all differentially mutated genes, and the rest would be the genes that are preferentially mutated in ecDNA- samples. Results will be visualized using volcano plots.

Results & Discussion

Loss-of-function mutations

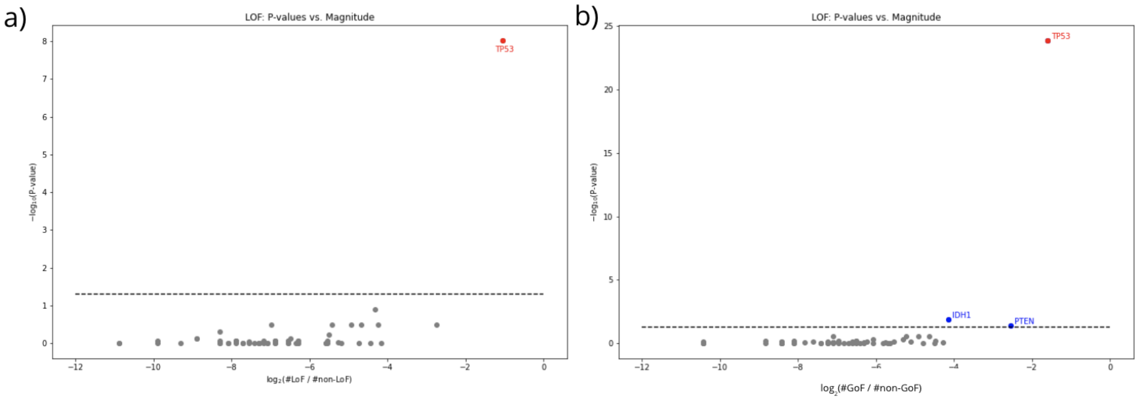

Fig. 2: Volcano plots of tumor suppressor genes considering LOF mutations/CNA. (**a**) ecDNA+ (Circular) vs. ecDNA- (others). (**b**) ecDNA+ (Circular) vs. ecDNA- (no SCNA detected). X-axis represents the magnitude (log ratio of LOF/GOF mutations). Y-axis represents the adjusted p-values from two-sided Fisher's exact test in negative log scale. Horizontal dashed line indicates p-value equal to 0.05. The red dot indicates a gene is associated with an adjusted p-value less than 0.05 from the one-sided Fisher's test, and it is marked blue otherwise.

The results are presented in Fig. 2. Under both settings for ecDNA- samples, TP53 is identified to preferentially mutate in ecDNA+ samples ($p < 0.05$). TP53 is known to regulate cell division, and circular ecDNA plays a key role in the development and progression of cancer (Zeng, 2020). Therefore, the tendency to find loss-of-function mutations of TP53 in ecDNA+ samples is plausible. On the other hand, only in the circular versus no SCNA case, we classify IDH1 and PTEN as genes that are preferentially mutated in ecDNA- samples. From literature, IDH1^{R132H} mutation is a key factor in glioma development (Núñez *et al.*, 2019), and so is the loss of PTEN (Philip *et al.*, 2018). Cooperation of mutations in these two genes may indicate that there are some relationships between them, while detailed connection between these two genes and ecDNA is unknown.

Gain-of-function mutations

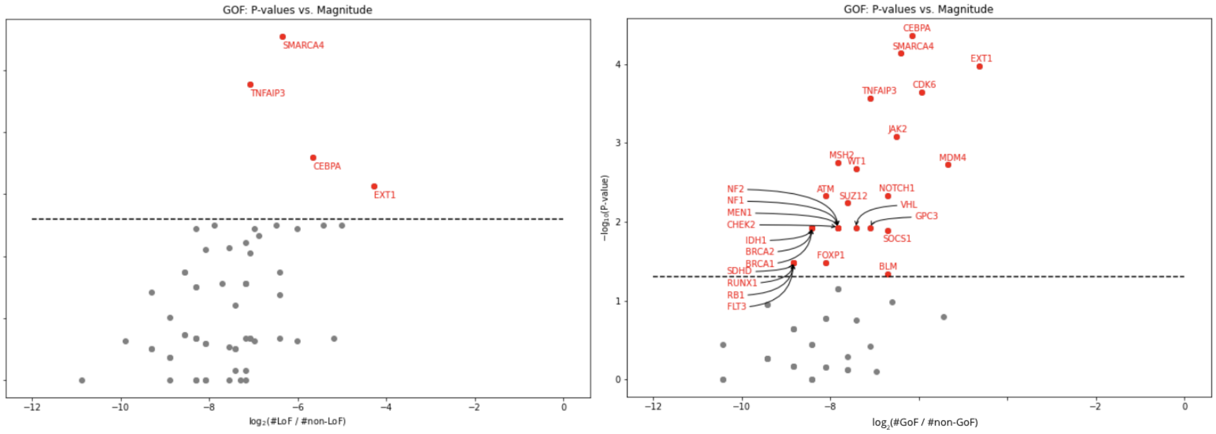

Fig. 3: Volcano plots of tumor suppressor genes considering GOF mutations/CNA. (**a**) ecDNA+ (Circular) vs. ecDNA- (others). (**b**) ecDNA+ (Circular) vs. ecDNA- (no SCNA detected). X-axis represents the magnitude (log ratio of LOF/GOF mutations). Y-axis represents the adjusted p-values from two-sided Fisher's exact test in the negative log scale. The horizontal dashed line indicates p-value equal to 0.05. The red dot indicates a gene is associated with an adjusted p-value less than 0.05 from the one-sided Fisher's test, and it is marked blue otherwise.

If we focus on gain-of-function mutations/CNA, we discover significantly more genes compared to the loss-of-function case (Fig. 2). The main reason for this is that in this study, we solely consider any form of amplification as gain-of-function mutations/CNA. However, among all significant genes we find, we fail to distinguish genes that are indeed on ecDNA, which will be definitely amplified due to the formation of ecDNA. Both settings for ecDNA- samples identify SMARCA4, TNFAIP3, EXT1 and CEBPA as genes that are preferentially mutated in ecDNA+ samples. From literature, we find that amplifications of these four genes are present in cases with breast invasive ductal carcinoma. Furthermore, SMARCA4 and CEBPA are actually located on the same chromosome. Thus, it is reasonable to infer that there is some association between these four genes, which may explain why they are identified together in this study. Nevertheless, the final list of genes are still subject to errors, since we do not take into consideration the possibility that missense mutations can also be gain-of-function mutations. Hence, further validity check on the list of differentially mutated is required before making any affirmative conclusions.

Gene fusion

a)

	ecDNA**+**	ecDNA**-**
Fusion	20	57
non-Fusion	315	1508

b)

Fisher's	two-sided	one-sided
p-value	0.0652	**0.0398**

c)

	ecDNA**+**	ecDNA**-**
Fusion	20	31
non-Fusion	315	1010

d)

Fisher's	two-sided	one-sided
p-value	**0.0186**	**0.0118**

Table 2: Contingency tables and p-values from Fisher's Exact test. (a-b) ecDNA+ (Circular) vs ecDNA (others). (c-d) ecDNA+ (Circular) vs ecDNA- (no SCNA detected).

Gene fusions are hybrid genes formed when 2 independent genes become juxtaposed. The effect of it can be truncating a TSG and therefore induce a LOF (Latysheva et. al, 2016), or GOF due to either creation of chimeric proteins with new or altered activities (Padmavathi, et.al). Therefore, we separate the analysis of gene fusion from the loss-of-function mutations and gain-of-function mutations. The contingency tables and the p-values from Fisher's exact tests are shown in Table 2. Significant p-values are boldfaced. In the second setting, we obtain smaller p-values compared to the first setting, while both yield p-values that support the hypothesis that some genes in the list of tumor suppressor genes tend to have gene fusion in ecDNA+ samples. Nonetheless, the frequency of gene fusion is very low in our dataset, it is hard for us to determine which gene, or which subset of genes are actually involved in gene fusion preferentially in ecDNA+ or ecDNA- samples.

Appendix

Code is available at https://github.com/insanebruce/ecDNA.

References

Verhaak, Roel G W et al. "Extrachromosomal oncogene amplification in tumour pathogenesis and evolution." Nature reviews. Cancer vol. 19,5 (2019): 283-288. doi:10.1038/s41568-019-0128-6

Yan, Y., Guo, G., Huang, J. et al. Current understanding of extrachromosomal circular DNA in cancer pathogenesis and therapeutic resistance. J Hematol Oncol 13, 124 (2020). https://doi.org/10.1186/s13045-020-00960-9

Latysheva, Natasha S, and M Madan Babu. "Discovering and understanding oncogenic gene fusions through data intensive computational approaches." Nucleic acids research vol. 44,10 (2016): 4487-503. doi:10.1093/nar/gkw282

The AACR Project GENIE Consortium. AACR Project GENIE: powering precision medicine through an international consortium. Cancer Discovery. 2017;7(8):818-831.

Padmavathi, Ganesan, et al. "Basic Concepts of Fusion Genes and Their Classification." Fusion Genes and Cancer, 2017, pp. 17–58., doi:10.1142/9789813200944_0002.

cBioPortal: The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E. Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J. Byrne, Michael L. Heuer, Erik Larsson, Yevgeniy Antipin, Boris Reva, Arthur P. Goldberg, Chris Sander and Nikolaus Schultz Cancer Discov May 1 2012 (2) (5) 401-404; DOI: 10.1158/2159-8290.CD-12-0095

Tumor suppressor gene list:
https://cancerres.aacrjournals.org/content/canres/suppl/2012/01/23/0008-5472.CAN-11-2266.DC1/T3_74K.pdf

Zeng, X., Wan, M. & Wu, J. ecDNA within tumors: a new mechanism that drives tumor heterogeneity and drug resistance. Sig Transduct Target Ther 5, 277 (2020). https://doi.org/10.1038/s41392-020-00403-4

Kim, H., Nguyen, NP., Turner, K. et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. Nat Genet 52, 891–897 (2020). https://doi.org/10.1038/s41588-020-0678-2

Núñez, Felipe J., et al. "IDH1-R132H Acts as a Tumor Suppressor in Glioma via Epigenetic up-Regulation of the Dna Damage Response." Science Translational Medicine, vol. 11, no. 479, 2019, doi:10.1126/scitranslmed.aaq1427

Philip, Beatrice, et al. "Mutant IDH1 Promotes Glioma Formation In Vivo." Cell Reports, vol. 23, no. 5, 2018, pp. 1553–1564., doi:10.1016/j.celrep.2018.03.133.