Mining Noncoding Mutually Exclusive Mutation Sets in Cancer

Student(s)

Tuvan Gezer
Faruk Balcı
Erkin Alaçamlı
Banu Çetinkaya

Faculty Member(s)

Öznur Taştan

Sabancı Üniversitesi

PURE
PROGRAM FOR UNDERGRADUATE RESEARCH

INTRODUCTION

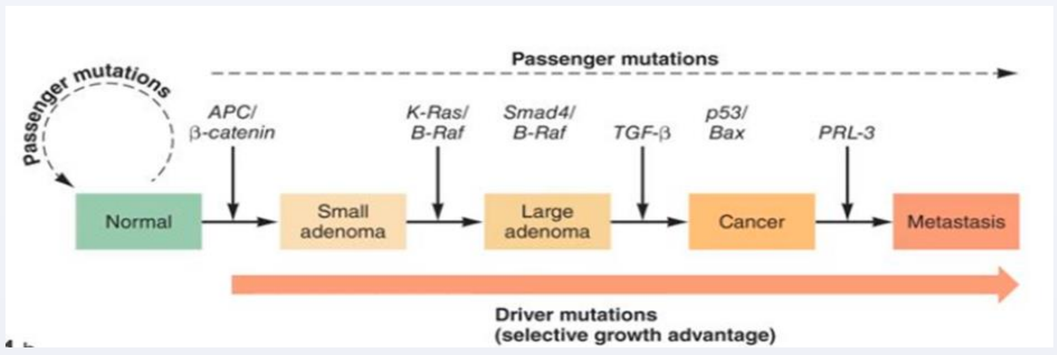

Figure 1: Figure demonstrates the driver and passengers mutations [1].

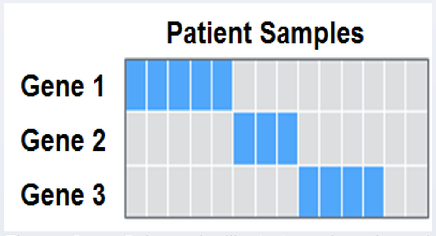

Figure 2: Schematic illustrating the idea of mutually exclusive mutation sets. Here these three genes form a MEGS.

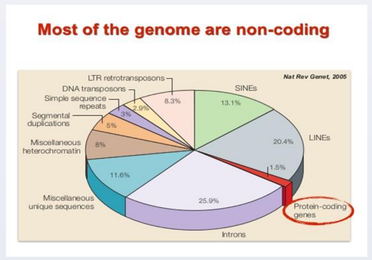

Figure 3: A large portion of the genome is non-coding. Figure is from [2].

Cancer genomes harbor many genomic alterations. With the advent of the next-generation sequencing technologies enable cataloging mutations in cancer patients at the whole genome scale. Some of these mutations act are considered as "driver" mutations as they mutations confer a fitness advantage to the tumor cell) while others are "passengers" (Figure 1). Since individual tumors exhibit a high level of diversity with different combinations of mutations, patterns that frequently observed in cancer patients may help to understand the functional relations of genes in cancer.

In this point, mutually exclusive sets are of interest. Mutually exclusive mutations are those in cancer genomes that are not found together in the same patient. Fig 2 illustrates this idea, if a gene 1 is mutated in a patient, the same patient does not harbor mutations in gene 2 or gene 3. Gene 1,2 and 3 form a mutually exclusive gene sets (MEGS). Algorithms to detect MEGS can be determined with algorithms such as MEMNAR[3]. Of these algorithms, MEMNAR uses negative association rule sets to find MEGS. These algorithms are designed to work on coding genes. However, most of the genome is non-coding (Figure 3) and most somatic variance in cancer occurs in non-coding region.

OBJECTIVES

- Applying MEMNAR to find MEGS in coding and noncoding genes.
- Investigating the clinical and functional relevance of the identified MEGS using complementary datasets.
  - Is there a relationship between different patient groups that exhibit harbour different mutations in a MEGS? That is do these mutations point to different patient subgroups.
    - Is there a relationship between age, survival date, gender and patient groups they have the same MEGS?
    - Do the mutated genes expression distributions differ from each other for different patient groups?

METHOD

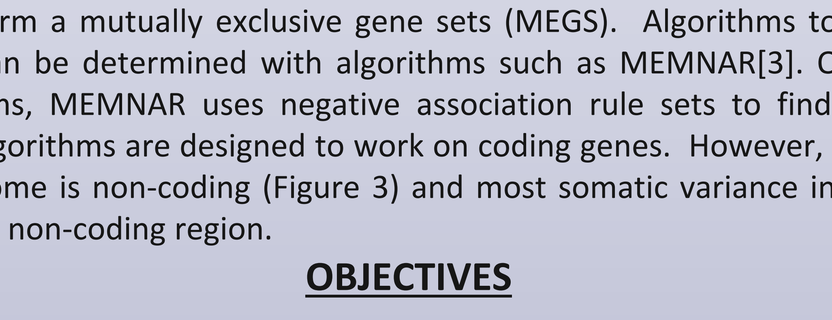

Figure 4 : Donor distributions.

- Obtained mutation, expression and clinical data of 1058 Glioblastoma Multiforme patients with non-coding mutations from ICGC[4] (Figure 4). There are 55.754 different mutated genes.
- We applied MEMNAR to identify MEGS. MEMNAR mines for positive association rules with high confidence and support as the first step, apply negative association rules(then prune and join rules that satisfy mutually exclusivity.
- In second step patients are grouped based on the mutation they bear. For example for Figure 2, three patients groups as P1, P2 and P3 will be formed, where P1 is patient group that have mutations in gene 1.

RESULTS

Identified Mutually Eclusive Sets

- 62 MEGSs are found after mutliple hypothesis test correction ($p < 0.05/1331$).
- The most significant MEGS is shown in Figure 5.

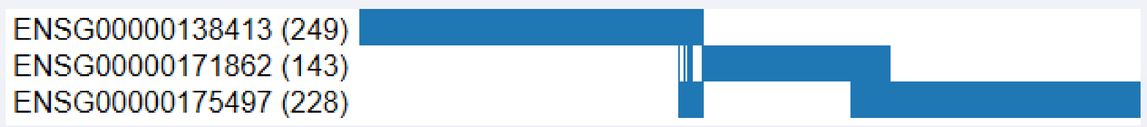

Figure 5

Expression Level Analysis

- Changing levels of gene expressions can influence cancer development. To understand if the patients in the same MEGS but harbouring different mutations are different in their expression patterns we checked if there is a differentially expressed gene among these groups of patients.
- To find differentially expressed genes we used t-test and ask the question if the two expression levels of the two patients groups differ significantly.

GENE	P-VALUE
PHYHIPL	5.305e-73
C10orf4	3.509e-70
GLUD1	1.521e-64
FCHSD2	4.767e-61
H2AFY2	9.055e-61
UBXN1	9.895e-61
FAM133A	1.547e-60
RPL3	1.589e-59
PCDH15	2.336e-58

Table 1:Differentially expressed genes

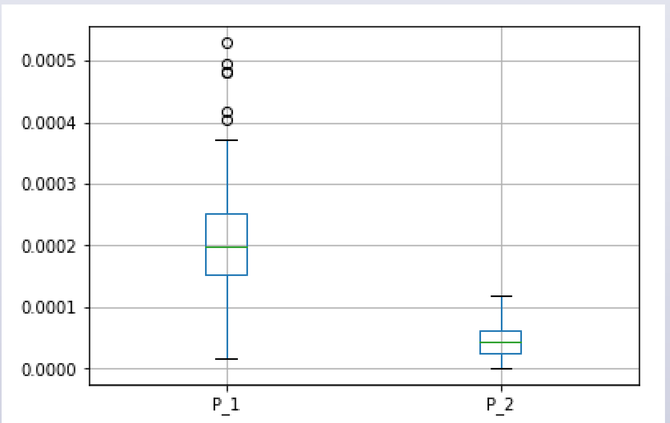

Figure 6: Expression level differences of the gene PHYHIPL on patient sets with different mutually exclusive mutations.

Clinical Data Analysis

- We set out to check if the patients groups constitute different groups based on their clinical information.
- In clinical data 445 female and 613 male patient's information is examined.
- Mutations' most occurence is found based on gender. These percentages' meanings evaluated with chi-square test.

CONCLUSION and FUTURE work

- We found significant correlations between certain exclusive mutation sets and expression levels of other genes (5970 genes with p value < 0.0005), some of the highly correlated genes are listed above.
- The work can be extended in future directions.
  - A more indepth analysis of the clinical variables will be conducted, such as the assation of the mutations with survival of the patients
  - The MEGSs will be analyzed in the light of the known functional information of the coding and nocoding genes.

REFERENCES

- Leiserson MDM, Blokh D, Sharan R, Raphael BJ (2013) Simultaneous Identification of Multiple Driver Pathways in Cancer. PLoS Comput Biol 9(5): e1003054. https://doi.org/10.1371/journal.pcbi.1003054
- Tawny N. Cuykendall, Mark A. Rubin, Ekta Khurana. Non-coding genetic variation in cancer, Current Opinion in Systems Biology, Volume 1, 2017, Pages 9-15
- Esteller, Manel.Non-coding RNAs in human disease, Nature Reviews Genetic 2011/11/18/online 12/861