Inferring NF2 Treatment Based on Similarity to Characterized Cancers

C3 Cancer Clustering Collaborative

Simas Glinskis[1], Natalia Antropova[1], Mohit Nalavadi[2], Avantika Lal[3], Joy Rimchala[4], Oren Schaedel[5]

1. University of Chicago 2. Santa Clara University 3. Stanford University

4. Intuit Inc., MIT 5. Ancestry.com

Background

Neurofibromatosis type 2 (NF2) is a genetically determined disorder that affects one in 40,000 individuals worldwide. NF2 is characterized by benign tumors in the nervous system. The condition effects region of the auditory-vestibular nerve, which carries sensory information from the ear to the brain. Even though the tumors are benign, they lead to auditory, balance, and visual problems.

Cause: mutation in NF2 gene

Current treatment: surgical removal is the most common way of treatment. Radiation might by used. Currently, extensive research is happening to develop drugs to shrink tumors or stop them from growing.

NF2 Project

The project gives a special opportunity to work with genomic data of Onno Faber, an NF2 patient.

Motivation

NF2 is rare genetic disease with limited research resources, and is consequently under-researched. Onno's dataset is unique because it is a high definition dataset for the NF2 tumor. The purpose of this exploratory analysis is to advance our understanding of the NF2 mutation and find correlates with existing cancer treatment pathways and medications. Understanding how the molecular mechanisms of NF2 relates to other well characterized cancers can help lead to future treatments of the disease.

Methods Outline

1. Filter the VCF data for patient's tumor
2. Download point mutations data for other brain cancers from The Cancer Genome Atlas (TCGA)
3. Extract mutated genes, gene functions, and mutational signatures from TCGA and NF2 datasets
4. Run clustering sequence-based and gene-based algorithms on all cancers (TCGA + NF2)
5. Identify patient cluster

Methods

1. VCF Pre-processing

We were given VCF files of tumor/normal pair from an NF2 patient and a reference sequence as a systematic control. The following information was extracted from the files, see Figure 1.

Tumor mutations were also filtered based on the variant quality score>30. We extracted the mutations present in the tumor genome but not in the blood. Further, we selected mutations with >15 reads supporting the alternate allele.

```
out = list()
for record in vcf_file:
    if qual_flag == True and record.QUAL<=30.0:
        continue
    d = {}
    d['POS'] = record.POS # the 1- based position of the variation on the sequence
    d['REF'] = record.REF # reference base at the given position of the given reference sequence
    d['ALT'] = record.ALT # list of alternative alleles at this position
    d['CHROM'] = record.CHROM # chromosome number on which variant is being called
    d['GT'] = record.genotype(cancer_ID)['GT'] # genotype
    d['AD'] = record.genotype(cancer_ID)['AD'] # allelic depths for the ref and alt alleles
    d['DP'] = record.genotype(cancer_ID)['DP'] # approximate read depth
    d['GQ'] = record.genotype(cancer_ID)['GQ'] # confidence that the assigned genotype is correct
    d['PL'] = record.genotype(cancer_ID)['PL'] # normalized, phred-scaled likelihoods for genotypes
    out.append(d)
return out
```

Figure 1. Python code selecting the specified information from VCF

2. TCGA Data Download

We downloaded annotated point mutation data for 290 glioblastomas and 286 gliomas from TCGA.

3. Mutated Genes, Gene Functions

We identified mutated genes in these samples as well as GO (biological process) annotations for the mutated genes.

4. Mutational Signatures

Cancer mutational signatures are known patterns of trinucleotide changes in the genome (e.g. a TCG trinucleotide being changed to a TTG trinucleotide). These signatures are associated with specific mutations or mutation-causing processes. We used 30 known signatures from the COSMIC database.

In each cancer sample, we identified the base before and after each point mutation and calculated the frequency of all possible trinucleotide conversions – for example, TCG -> TTG (Fig.2). We identified known mutational signatures in each tumor using the deconstructSigs R package.

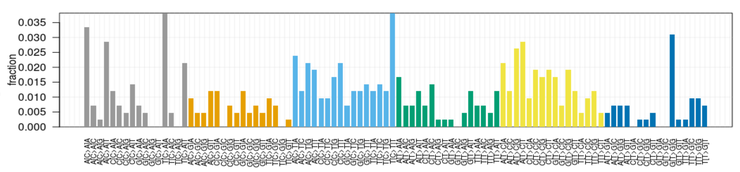

Figure 2. Frequencies of trinucleotide mutations in the NF2 tumor

5. Clustering

To insure unique and interesting clusters, it is important to maximize variance within the data. The dimensionality of the data was reduced using PCA techniques, maintaining 80 percent of the variance. Using the scaled data, and applying the KMeans algorithm, the number of clusters fit was tuned as a hyper-parameter. The number of final clusters was generated inspecting the loss profile and searching for the "elbow", where the algorithm begins to overfit.

Figure 3. Elbow in cluster loss

6. Cluster Specifics

The clustering algorithm was fit for the mutational signature and gene GO data sets respectively, and Onno's filtered genetic data was assigned to a cluster within the two data sets.

Results

- Filtering the VCF resulted in ~9000 mutations, 229 were in protein coding sequence, 43 genes.
- Onno's tumor contained mutational signatures 1,3,5,12 and 21. The most abundant signature was signature 3, which is found in another class of hereditary cancers - BRCA+ breast cancer, see Figure 3.

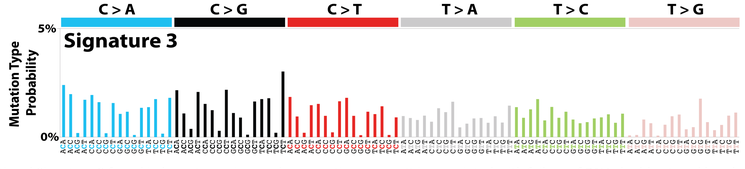

Figure 4. Mutational Signature 3 found in breast, ovarian, and pancreatic cancers. It is associated with failure of DNA double-strand break-repair by homologous recombination

- 7 signature clusters identified in the common mutational signature data: Onno fits into the least populated cluster, see Figure 5(L).
- 5 functional clusters identified in the mutated gene GO data: Onno is a strong outlier, see Figure 5(R).

Results

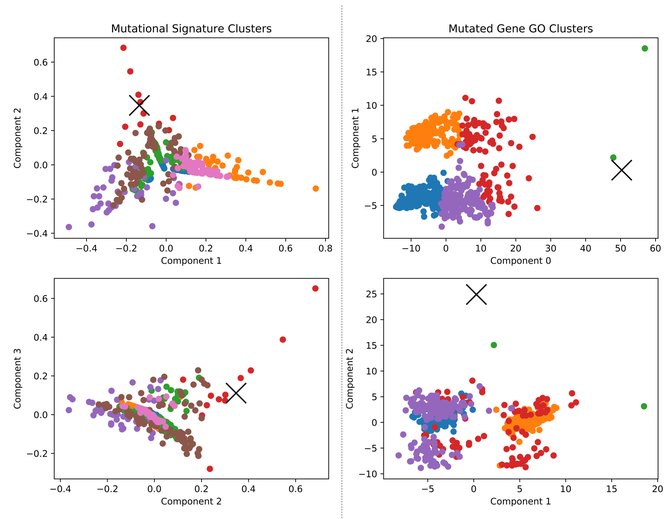

Figure 5. Clusters for common mutational signatures and mutated gene go terms respectively. Onno is represented by the X.

Conclusion

1. There are 7 mutational signature clusters of brain cancer. The NF2 patient falls into the least abundant cluster. One signature that characterizes these is signature 3, which is characteristic of defective homologous repair, especially in BRCA+ breast cancers.
2. 2. This suggests that drugs targeting homologous repair, e.g. PARP inhibitors, could be used to treat the NF2 patient.
3. The NF2 patient's tumor is functionally very different from all known brain cancers. It should be compared with other cancer types.

Future Direction

- *Broaden:* Extend the clustering analysis to cancer types beyond brain cancers
- *Deepen:* Examine similarity of Onno's to other patients assigned to the same mutational cluster
- *Explore*: Possible therapeutics. Onno's mutational signature is enriched in breast cancer mutational signature with well known therapeutics.