Multi-Omics Analysis for Understanding the Molecular Basis of Lung Adenocarcinoma

[1]Sunil Cherukuri., [1]Srikanthi Ramachandrula, [1]Akanksha Mishra, [1]Durairaj Renu, [1]Ashwin Kalbor, [1]Shanmukh Katragadda, [1]Kamal Baid [1]Arjun Duggal, Shyam Kalakoti, [1]Rohit Gupta, [1]Aishwara Narayanan, [1]Anupama Ranjan Bhat, [1]Dipa R, [2]Danielle Fletcher, [1]Pramila Tata and [2]Vanessa Lordi

[1]Strand Life Sciences, Bangalore, India
[2]Agilent Technologies, Santa Clara, USA

Introduction

Large scale data is being generated by consortia like the TCGA and the 1000 Genomes Project in an effort to understand the molecular basis of diseases. Population based studies are key to understanding the effect of genomic variations on phenotype and are essential to pave the way for personalized medicine. Integration of basic research into medical practice becomes more effective when information from different technology platforms are incorporated into our understanding of disease process.

The current report is an attempt to analyse complex multi-omics lung adenocarcinoma data using intuitive methods coupled with powerful visualization. The motivation of this was to demonstrate that dynamic viewing of multi-omics data is key to better understand biological patterns and to arrive at testable hypothesis.

Materials and Methods

Lung Adenocarcinoma datasets were accessed from the ICGC Data Portal. All the analysis was performed using GeneSpring 13.1® (www.genespring.com) and Strand NGS (www.strand-ngs.com)

Results and Discussion

We initially examined the extensive metadata supplied by TCGA [1]. The metadata encompassed mutations, focal aberrations and copy number aberrations for the key driver genes, i.e. the oncogenes and tumour suppressors. Using the metadata framework in GeneSpring, we examined each of these classes of aberrations.

Focal aberrations: Focal copy number aberrations are known to contain driver genes [2]. They are also thought to favour tumour development and progression evolutionarily [3]. The observation that ERBB2 has a focal amplification in “oncogene-positive” subset of patients [1] was also seen by us. In addition, we noticed that some genes like the EGFR and KRAS oncogenes preferentially undergo focal amplifications, while tumour suppressors like TP53 and RB1 undergo deletion (Fig.1). However some tumour suppressors like STK11 and KEAP1, oncogenes like ERBB2 and HRAS and genes involved in splicing U2AF1 have both focal amplifications and deletions.

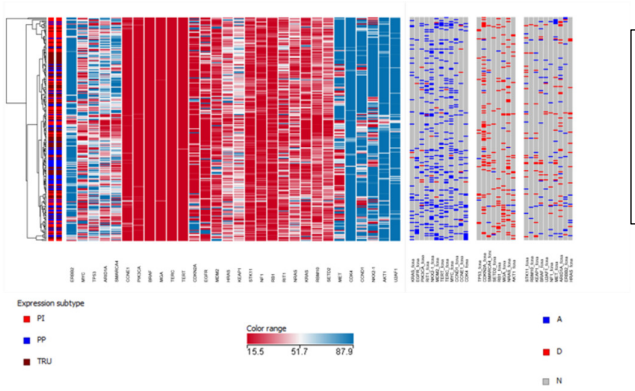

Fig1: Focal amplifications in key driver genes. While most oncogenes have focal amplifications (first panel after the cluster) and tumour suppressors have deletions (middle panel), some key drivers have both (extreme right panel). Each row corresponds to a patient sample.

Mutations: Aligning the known mutations as metadata, identified several interesting patterns. As reported in [1], mutations in p53 are diverse and span several locations, while mutations in KRAS are confined to only a few codons. This was seen very clearly in Fig.2. In addition we observe that

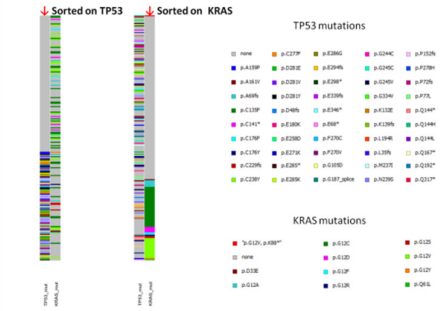

Fig.2: Patient wise alignment of mutations in TP53 and KRAS. Each line represents a patient with a mutation, while the grey represents patients with no mutations. Left panel is sorted on TP53 mutations, while the right panel is sorted on KRAS mutations. About 2/3rd of KRAS mutations are in patients who do not have TP53 mutations.

About two thirds of the KRAS mutations occur in patients who do not have TP53 mutations, indicating a partial exclusivity. Lack of overlap between KRAS and EGFR mutations has been reported in lung cancers [1, 4]. TP53 and KRAS mutations have been shown to be largely mutually exclusive events in colorectal cancer [5], but our observation reveals that it is present in lung adenocarcinoma as well.

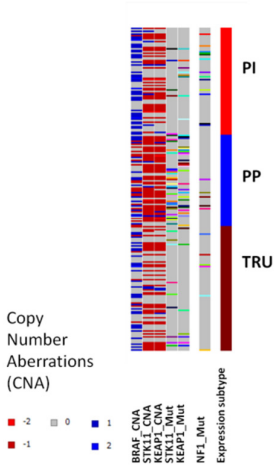

Fig.3: Patient wise alignment of copy number aberrations and mutations with the cancer subtype. Deletions in STK11 and KEAP1 are predominantly located in PP subtype. For BRAF, PP subtype contains both amplifications and deletions, while the PI and TRU contain only amplifications. Similarly, PP subtype contains the maximum mutations of STK11 and KEAP1 while PI is the subtype having NF1 mutations.

We also observed that mutations in STK11, KEAP1 and NF1 were also related to cancer subtype. STK11 and KEAP1 were more prominent in PP subtype while NF1 was more prevalent in PI subtype (Fig.3).

Copy Number aberrations: Examination of large scale genomic aberrations revealed events happening in most of the patients genomes in this cohort. As seen in Fig.4, patients who do not harbour aberrations in EGFR also do not contain deletions and amplifications in BRAF and MET. In addition patients containing amplifications in EGFR also contain amplifications in BRAF and MET. Since EGFR is on the p arm of chromosome 7 and BRAF and MET on q arm, it could indicate either 2 highly coordinated events occurring leading to the observed behaviour or result of a single event. Similar patterns were also seen for KRAS, MDM2 and CDK2 as well as for CCND1 and HRAS. STK11, KEAP1 and SMARCA4 show almost identical pattern of genomic deletions, indicating these gene deletions to be the result of a single deletion event. Chromosome 9p alteration has previously been reported [6].

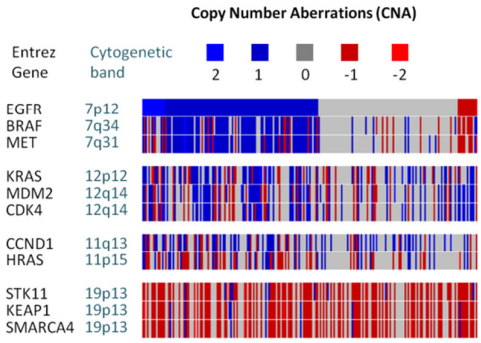

Fig.4: Similar patterns of alterations across different genes located on the same chromosome.

Effect of copy number amplifications on gene expression: Elevated levels of gene expression in cancer could be due to gene amplification, or due to its regulation by miRNAs or other transcription factors. Using expression and genomic copy number data from same patients, we performed correlation to identify genes whose over expression is possibly due to gene amplification. Thus such genes exhibit high correlation coefficient when their expression values are correlated with copy number values, as seen in Fig. 5.

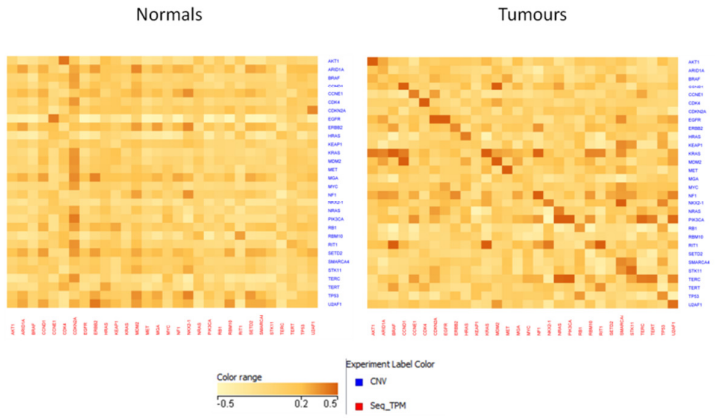

Fig.5: While normal samples do not show any correlation between their expression levels and copy number values, some of the overexpressing genes show a very high correlation with their copy number indicating that the amplification is driving the expression.

Comparative proteomics and mRNA profiles: As mentioned above, we observed that two thirds of KRAS mutations occur in patient samples which do not have TP53 mutations, indicating partial exclusivity. To understand the changes in expression profile due to presence and absence of KRAS and P53 mutations, the protein expression and mRNA sequencing data were clustered together. Prior to clustering, the datasets were independently scaled by auto-scaling to ensure the data from sequencing and protein array are comparable [7]. The proteins and their corresponding mRNA's were cross mapped based on common Gene Symbol annotation. The final data table included 452 columns of expression data, corresponding to 226 samples each from sequencing and protein array for the 123 gene-protein rows commonly measured between platforms. K-means clustering revealed two distinctive behaviours. The cluster 0 entities expression values increase when TP53 is mutated as compared to when KRAS is mutated. Cluster 2 on the other had has opposite behaviour with expression value in TP53 mutated samples decreasing as compared to those with KRAS mutations.

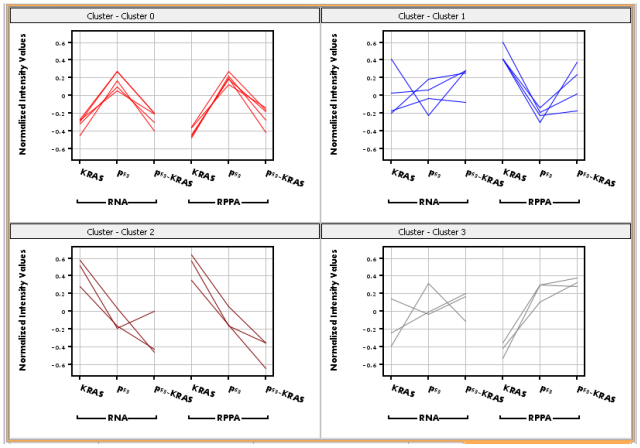

Fig.6: K-means clustering of entities (proteins and transcript expression values) as seen in patient samples with mutated KRAS, mutated TP53 and both KRAS and TP53 mutated. Cluster 0 and cluster 2 show very distinctive behaviour when both TP53 and KRAS are mutated.

More interestingly, in the small subset of patients where mutation of both KRAS and TP53 is observed, entities in cluster 0 mimic the KRAS mutation behaviour whole in cluster 2 the expression pattern in the KRAS and TP53 mutants drop down further (Fig.6).

Conclusions

We demonstrate here that use of powerful and intuitive visualizations are critical to understanding and analysing complex multi-omics data. By using this approach, we have identified new mutation patterns and relationships between the TP53 and KRAS mutations. Further we have also used the metadata framework to easily narrow down the genes where copy number changes are the result of a common event. The fact that several patients show similar aberrations also indicate that certain chromosomal events are common within a population. Investigating the TP53 and KRAS partial exclusivity further by integrating proteomic and transcriptomic data, we also identified entities which could potentially give an answer to the partial exclusive presence of the mutations.

References

[1] Comprehensive molecular profiling of lung adenocarcinoma. TCGA. Nature (2014) 511, 543–550
[2] Focal chromosomal copy number aberrations in cancer-Needles in a genome haystack. Krijgsman, O. et al. Biochim Biophys Acta. 2014 Nov; 1843(11):2698-2704

[3] Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. Davoli T et al. Cell 2013 155: 948–962

[4] EGFR and KRAS Mutations in Patients With Adenocarcinoma of the Lung. Jang, T,W. et al., Korean J Intern Med. 2009 Mar; 24(1): 48–54

[5] Mutations of APC, K-ras, and p53 Are Associated with Specific Chromosomal Aberrations in Colorectal Adenocarcinomas. Leslie A., et al. Cancer Res August 1, 2003 63; 4656

[6] Chromosomal alterations in lung adenocarcinoma from smokers and non-smokers Sanchez-Cespedes et al., Cancer Research 2001, 61, 1309 –1313.

[7] Integrating gene and protein expression data: pattern analysis and profile mining. Cox B et al. Methods 35 (2005) 303–314.