Working title
Role of noncoding variants in cancer

Authors
Ekta Khurana, Yao Fu, Andrea Sboner (??), Mark Rubin, Mark Gerstein

Outline of main text:

Abstract (~100 words) ***[Still working]***

Tumor genomes contain numerous somatic genomic mutations and rearrangements. A large majority of these variants occur in noncoding parts of the genome. With the decreasing cost of whole-genome sequencing, it is now possible to identify noncoding variants across many cancer genomes. It is known that only a few genomic variants have a functional impact and drive tumor growth and development. Most previous studies have focused on the identification of functional mutations in protein-coding genes……

Introduction

In this part we will discuss that this topic is very timely because whole-genome sequencing of tumors is possible now and there is huge interest in knowing what the variants mean. We will discuss the following points:

1) Noncoding mutations are much more abundant than coding; we will provide a sense of the scale of genomic coverage.

2) We will discuss that many germline mutations in promoters and enhancers are known to be causal for inherited diseases and we are just beginning to explore the role of noncoding somatic mutations in oncogenesis. We will also discuss how there is an increased interest in the cancer community because many recent studies point to the role of *TERT* promoter mutations in many different cancer types.

3) We will discuss how recent studies show that small change in gene expression can have large phenotypic impact (e.g. a SNP in enhancer causing 20% change in *KITLG* expression is responsible for blond hair color). So cumulative effect of small changes in expression due to noncoding mutations can be huge. In relation to this, the current idea of driver mutations may not necessarily be true, since mutations may aid tumor growth by varying extent.

4) Although our article will focus on somatic mutations in cancer, we will also discuss germline variants that have been associated with increased cancer susceptibility, specially the cases where there is an intricate relationship between germline polymorphisms and somatic variants.

5) Noncoding regions play varied roles in cancer – e.g. besides sequence variants in these regions, epigenetic changes and expression changes of noncoding RNAs can also drive cancer. We will make the reader aware that our article focuses only on sequence variants in noncoding parts. In relation to this, we will also briefly discuss other causes of cancer, e.g. viral infections etc.

Main sections

1) Noncoding annotations and genomic sequence variants.

a) What are noncoding annotations (e.g. from ENCODE and Epigenomics Roadmap) and evolutionary conserved regions in noncoding genome

b) Regulatory regions are often cell-type/tissue specific

c) What are somatic variants and how their patterns are different from germline variants. For example, a larger fraction of somatic variants contain large genomic rearrangements and tumor heterogeneity makes interpretation of somatic variants more complicated.

2) Known cases of somatic variants playing a role in tumor development and growth.

a) A discussion of how mutations could effect gene expression, e.g. point mutations in transcription factor binding motifs and miRNAs, small insertions or deletions and larger structural variants.

b) Mutations in pseudogenes could effect competition for miRNA binding with the parent gene [[PTEN ref]], which in turn could effect expression of the parent gene.

c) Splice site disruption caused by intronic variants

This discussion can be combined with examples listed below which we will also display in a Table.

a) Promoters: *TERT* promoter mutations [[ref.]]

b) Enhancers: Enhancer hijacking in medulloblastoma (Northcott et al, Nature, 2014). Similarly, promoter juxtaposed to gene, e.g. TMPRSS2-ERG [[Mark R.]]

c) ncRNA: expression change of ncRNA can be due to somatic variants like CNVs of ncrnas [[ref.]]

Also, mutations may occur in miRNA binding sites [[ref.]]

d) Mutations in introns causing splice site disruption: Splice site mutation in the noncoding region of BRCA2 -- implications for familial breast cancer (Bakker et al., Human Mutation, 2014)

3) Germline inherited variants in noncoding regions that alter cancer susceptibility

a) There is an enrichment of GWAS variants, including those associated with cancer susceptibility, in the noncoding genome; as we sequence more populations we will identify variants that are common in those populations and related to cancer susceptibility. We discuss the following examples and summarize them in a Table:

(i) Effect of noncoding mutations and interplay between germline and somatic variants can be complex: For example, a common SNP in *TERT* promoter modifies the effects of somatic TERT promoter mutations in bladder cancer on patient survival (Rachakonda et al, PNAS, 2013).

(ii) A SNP in *RFX6* gene intron effects *HOXB13* binding and is linked to increased prostate cancer susceptibility. We will discuss the relevance of two hit hypothesis (where one allele is disabled by a germline variant and the other by somatic variant) for noncoding regions.

(iii) A SNP in miR-27a gene actually reduces susceptibility to gastric cancer (Yang et al, Oncogene, 2014)

(iv) SNPs in enhancers on chr 8q24 upstream of MYC are related with increased risk for multiple cancer types (Grisanzio & Freedman, Genes & Cancer, 2010)

(b) eQTL analysis has been used to interpret risk loci (Li et al, Cell, 2013; Xu et al, Eur J Hum Genet, 2014). This analysis also points to complex relationship of germline and somatic variants. We will also discuss why usually there is no eQTL analysis for somatic variants (since cancer is heterogeneous so these variants are rare). Cryptic effects of noncoding mutations have also been noted where germline variants exhibit allelic effects in tumor (Dermitzakis paper, Nature, 2014).

We will use the above example to discuss how the notion of driver mutations may not be binary since somatic mutations can influence cancer growth to varied extent based on the presence of other germline and somatic variants.

4) Different types of cancer

(a) Numbers of noncoding vs coding mutations in different types of cancer

(b) Summary of cancers where driver mutations have been identified in protein-coding genes vs those where causal mutations have not been identified and may lie in noncoding regions.

5) Cancer mutational signatures at whole genome level involving noncoding DNA

(a) A brief discussion of cancer mutational properties, such as kataegis [[ref.]], chromothripsis [[ref.]], and correlation of histone modifications with sequence variants [[ref.]] that include mutations of noncoding regions

6) Methods to identify noncoding somatic variants of functional

(a) Discussion of currently available computational methods to identify noncoding driver mutations from whole-genome sequencing data, for example, FunSeq, CADD, GWAVA [[ref.]]. We will also depict these in a Table with associated website links.

Conclusions/perspective

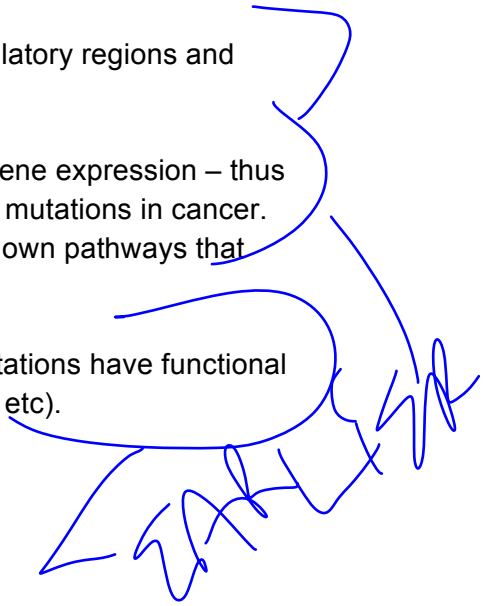
(a) Cancer arises because of accumulation of multiple mutations -- some of these drivers could be noncoding -- unknown unknowns. There is a bias in literature for driver noncoding mutations because people haven't explored these regions for cancer drivers, for example most studies have been focusing on exomes including the majority of TCGA studies.

(b) There is a debate in the community about whether we should look at noncoding/whole genomes vs exomes. Studies of somatic noncoding mutations are currently mostly for research purposes, as opposed to regular clinical use. This is primarily because current therapeutic approaches attempt to target proteins. It is possible that alternate methodologies, such as genome editing using CRISPR, may be used in future (e.g. Hu et al PNAS 2014 used Cas9 for HIV in cell lines and proposed therapeutic application). However, noncoding germline variants associated with increased cancer risk may prove to be important for preventive approaches.

(c) In relation to (b), it is very important to know the links between cis-regulatory regions and their target genes and we will discuss state-of-the-art approaches.

(d) It is important to study effects of mutations in all elements controlling gene expression – thus network approaches will be important to understand the role of noncoding mutations in cancer. We might also be able to identify new pathways or novel participants in known pathways that are important in cancer.

(e) Finally, we will discuss experimental ways to test which noncoding mutations have functional effects (e.g. CRISPR, luciferase reporter assays, high-throughput assays, etc).



Proposed display items

(1) Table of noncoding annotations
(2) Table of somatic sequence variants important in cancer
(3) Table of germline sequence variants related to altered cancer susceptibility
(4) Table of computational methods to prioritize noncoding mutations with functional effects
(5) Table of experimental techniques to validate them
(6) Schematic for role of noncoding mutations in oncogenesis

Key references

[[Will consolidate all references here]]