Working title

Role of noncoding variants in cancer

Outline of main text:

Abstract (~100 words)

Tumor genomes contain numerous somatic sequence variants. These include single nucleotide mutations, small insertions and deletions and larger sequence rearrangements. A large majority of these variants occur in noncoding parts of the genome. Noncoding variants can effect gene expression to variable extents and may have major functional consequences causing tumor progression. Although most previous studies have focused on the identification of functional variants in protein-coding genes, many recent studies suggest that the repertoire of noncoding somatic variants contains driver events playing an important role in tumor growth. Furthermore, numerous noncoding germline variants are known to play a role in cancer susceptibility. In many instances, tumor growth relies on an intricate balance between inherited germline and acquired somatic variants. In this review, we discuss the current understanding of the role of both somatic and germline noncoding variants in cancer.

Ekta Khurana 8/6/14 4:59 PM
Deleted: genomic mutations and rearrangements

Ekta Khurana 8/6/14 5:20 PM
Deleted: there is

Ekta Khurana 8/6/14 5:22 PM
Deleted: Noncoding variants can effect expression of protein-coding genes to variable extents, and the cumulative effect of multiple variants may have major functional consequences causing tumor progression.

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 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.
2) We will discuss that noncoding mutations are much more abundant than coding by providing a sense of the scale of genomic coverage.
3) 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. Recent studies show that small changes in gene expression caused by noncoding mutations 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 in cancer can be huge. In relation to this, the current idea of binary classification of somatic mutations, i.e. drivers and passengers, may not necessarily be true. This is because the tendency of mutation to cause tumor growth may not be the extreme 100% (driver) or 0% (passenger) but rather anywhere in between. [[Mark R.: What are your thoughts about the highlighted sentences.]] We will also discuss how there is an increased interest in the cancer community to analyze noncoding mutations because many recent studies point to the role of *TERT* promoter mutations in many different cancer types.

Ekta Khurana 8/6/14 5:24 PM
Formatted: Highlight

Ekta Khurana 8/6/14 5:24 PM
Formatted: Highlight

4) 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.

Ekta Khurana 8/6/14 5:32 PM
Deleted: Although our article will focus on somatic mutations in cancer,

Ekta Khurana 8/6/14 5:32 PM
Deleted: w

Main sections

1) Noncoding annotations.

a) What are the various noncoding annotations: transcription factor binding sites, DNase I hypersensitivity sites, noncoding RNAs, etc. We will discuss that the dynamic nature of the epigenome (including various histone marks and DNA methylation) leads to differential activity of regulatory regions in different cellular states. We will also discuss large-scale efforts to annotate functional elements in the genome, such as ENCODE [1] and Roadmap Epigenomics project [2]. This section will also include a discussion of evolutionary conserved regions in noncoding genome e.g. ultraconserved elements [3] and ultrasensitive regions [4].

Ekta Khurana 8/6/14 5:38 PM
Deleted: ? For example,

Ekta Khurana 8/6/14 5:45 PM
Deleted: create a map of

b) Regulatory regions are often cell-type/tissue specific and thus sequence variants in these regions are more likely to exhibit tissue-specific effects.

c) Multiple approaches are currently used to link cis-regulatory regions to their target genes. For example: different variations of chromosome conformation capture technology [5, 6], correlation of transcription factor (TF) binding and expression across multiple cell lines [7], etc. The resulting linkages can then be studied as a comprehensive regulatory network [8].

2) Genomic sequence variants.

a) What are various types of sequence variants: single nucleotide substitutions, small insertion and deletions, and larger structural variants.

b) There are many differences in patterns of somatic variants and inherited germline variants: (i) A higher fraction of somatic variants contain large genomic rearrangements. Chromosomal aneuploidy is also often observed in cancer cells. (ii) Tumor heterogeneity makes interpretation of somatic variants more complicated. (iii) Various phenomena, such as kataegis [9] and chromothripsis [10] are characteristic only of somatic cancer variants.

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

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

a) Promoters: *TERT* promoter mutations [11-14]. UTR/promoter juxtaposed to gene, e.g. *TMPRSS2-ERG* fusion [[Mark R.: does this sound correct (does the UTR fuse or promoter) and best reference to cite??]]

b) Enhancers: Enhancer hijacking in medulloblastoma[15].

Ekta Khurana 8/7/14 5:50 PM
Formatted: Highlight

Ekta Khurana 8/7/14 5:50 PM
Formatted: Font:Not Italic, Highlight

Ekta Khurana 8/7/14 5:50 PM
Formatted: Highlight

Ekta Khurana 8/7/14 5:47 PM
Deleted: Similarly, UTR/promoter juxtaposed to gene, e.g. *TMPRSS2-ERG* [[Ask Mark R.: does this fit and best ref.??]]

Ekta Khurana 8/5/14 5:41 PM
Formatted: Font color: Auto, Highlight

c) ncRNA: expression change of ncRNA can be due to somatic variants like CNVs of ncrnas. *MALAT1* was found to be significantly mutated in bladder cancer [16]. Pseudogene deletion can effect competition for miRNA binding with the parent gene, which in turn could effect expression of the parent gene [17]. Mutations in miRNA binding sites can also effect their binding.

4) Germline inherited variants in noncoding regions that alter cancer susceptibility or patient survival.

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 will discuss the following examples and summarize them in a Table:

(i) SNPs in enhancers on chr 8q24 upstream of *MYC* are related with increased risk for multiple cancer types [18].

(ii) A SNP in *RFX6* gene intron effects *HOXB13* binding and is linked to increased prostate cancer susceptibility [19].

(iii) A SNP in miR-27a gene reduces susceptibility to gastric cancer [20].

Ekta Khurana 8/7/14 5:51 PM
Deleted: actually

(iv) A common SNP in *TERT* promoter modifies the effects of somatic *TERT* promoter mutations in bladder cancer on patient survival [21].

(v) Splice site mutation in the intron of *BRCA2* has implications for familial breast cancer [22].

(b) eQTL analysis has been used to interpret risk loci [23, 24]. 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 [25].

These examples illustrate how the effect of noncoding mutations and interplay between germline and somatic variants can be complex. 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. We will also use the above examples 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.

5) Different types of cancer [[More text from Mark R.]]

Ekta Khurana 8/5/14 5:41 PM
Formatted: Font color: Auto, Highlight

Ekta Khurana 8/7/14 5:51 PM
Deleted: *Ask*

Ekta Khurana 8/5/14 5:41 PM
Formatted: Font color: Auto, Highlight

Ekta Khurana 8/7/14 5:51 PM
Deleted: *for more text*

(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.

6) Computational methods to identify noncoding somatic variants with functional consequences.

(a) Discussion of currently available computational methods to identify noncoding driver mutations from whole-genome sequencing data, for example, FunSeq [4], CADD [26] and GWAVA [27]. We will also depict these in a Table with associated website links.

7) Experimental approaches to understand the functional effects of noncoding mutations. [[Mark R.: should we add something else here ?]]

Finally, we will discuss experimental ways to test which noncoding mutations have functional effects (e.g. genome editing using CRISPR, luciferase reporter assays, high-throughput assays, etc). We will also discuss the scale and cost of all the techniques and summarize them in a Table.

Conclusions/perspective

(a) Cancer arises because of accumulation of multiple mutations -- some of these drivers could be noncoding. There is a bias in literature for driver noncoding mutations because people haven't explored these regions for cancer drivers to the same extent, 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. Cas9 was used for HIV in cell lines and proposed for therapeutic application [28]). However, noncoding germline variants associated with increased cancer susceptibility should be important for risk assessment and potentially for preventive approaches.

(c) In relation to (b), it is very important to know the links between cis-regulatory regions and their target genes. Although many approaches exist (as discussed under 'Main sections'), this remains a very active and important area of research., especially the development of high-throughput choromosomal capture technologies.

(d) Even when the links between regulatory regions and target genes are known, 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.

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

Ekta Khurana 8/7/14 5:57 PM
Deleted: *[[Ask*

Ekta Khurana 8/5/14 5:41 PM
Formatted: Font color: Auto, Highlight

Ekta Khurana 8/7/14 5:56 PM
Deleted:

Ekta Khurana 8/5/14 5:41 PM
Formatted: Font color: Auto, Highlight

Ekta Khurana 8/7/14 5:52 PM
Deleted: *for more text*

Ekta Khurana 8/5/14 5:41 PM
Formatted: Font color: Auto, Highlight

(6) Schematic for role of various noncoding annotations and sequence variants in them in oncogenesis

Ekta Khurana 8/7/14 6:01 PM
Deleted: role of noncoding mutations

Key references

1. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
2. Chadwick, L.H. The NIH Roadmap Epigenomics Program data resource. *Epigenomics* **4**, 317-24 (2012).
3. Bejerano, G. et al. Ultraconserved elements in the human genome. *Science* **304**, 1321-5 (2004).
4. Khurana, E. et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* **342**, 1235587 (2013).
5. Hughes, J.R. et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**, 205-12 (2014).
6. de Laat, W. & Dekker, J. 3C-based technologies to study the shape of the genome. *Methods* **58**, 189-91 (2012).
7. Yip, K.Y. et al. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol* **13**, R48 (2012).
8. Gerstein, M.B. et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100 (2012).
9. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-93 (2012).
10. Stephens, P.J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27-40 (2011).
11. Huang, F.W. et al. Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-9 (2013).
12. Horn, S. et al. TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959-61 (2013).
13. Killela, P.J. et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc Natl Acad Sci U S A* **110**, 6021-6 (2013).
14. Heidenreich, B., Rachakonda, P.S., Hemminki, K. & Kumar, R. TERT promoter mutations in cancer development. *Curr Opin Genet Dev* **24**, 30-7 (2014).
15. Northcott, P.A. et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428-34 (2014).
16. Kandoth, C. et al. Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-9 (2013).
17. Poliseno, L. et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033-8 (2010).
18. Grisanzio, C. & Freedman, M.L. Chromosome 8q24-Associated Cancers and MYC. *Genes Cancer* **1**, 555-9 (2010).
19. Huang, Q. et al. A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nat Genet* **46**, 126-35 (2014).

20. Yang, Q. et al. Genetic variations in miR-27a gene decrease mature miR-27a level and reduce gastric cancer susceptibility. *Oncogene* **33**, 193-202 (2014).
21. Rachakonda, P.S. et al. TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. *Proc Natl Acad Sci U S A* **110**, 17426-31 (2013).
22. Bakker, J.L. et al. A Novel Splice Site Mutation in the Noncoding Region of BRCA2: Implications for Fanconi Anemia and Familial Breast Cancer Diagnostics. *Human Mutation* **35**, 442-446 (2014).
23. Li, Q. et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* **152**, 633-41 (2013).
24. Xu, X. et al. Variants at IRX4 as prostate cancer expression quantitative trait loci. *Eur J Hum Genet* **22**, 558-63 (2014).
25. Ongen, H. et al. Putative cis-regulatory drivers in colorectal cancer. *Nature* (2014).
26. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
27. Ritchie, G.R., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of noncoding sequence variants. *Nat Methods* **11**, 294-6 (2014).
28. Hu, W. et al. RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc Natl Acad Sci U S A* (2014).