

RESEARCH HIGHLIGHT

Predicting cancer drivers: are we there yet?

Vidhya G Krishnan and Pauline C Ng*

See research article: <http://genomemedicine.com/content/4/11/89>

Abstract

Genomic variants with a key role in causing cancer or affecting the response to cancer therapeutics need to be identified so that they can be targeted for therapy. The transFIC tool aims to identify somatic point mutations that drive cancer in sequencing projects. This package is available as a web service, a stand-alone program and a website. It improves the functional prediction scores generated by popular established prediction tools and will be useful to cancer researchers.

The functional impact of cancer-associated mutations

Mutations give rise to cancerous cells by affecting genes. For example, 'gain of function' mutations in oncogenes such as *EGFR* and *KRAS* promote tumor progression, and 'loss of function' mutations in the tumor-suppressor gene *TP53* promote cancer by dysregulating the cell cycle. Mutations that provide a selective growth advantage to the cancer cell are called 'driver' mutations. 'Passenger' mutations, by contrast, are present in cancer genomes but do not give such a growth advantage.

Identifying driver genes is important for clinical applications. If certain mutations are present in specific cancer-associated genes, then the cancer drugs that target these genes and their respective pathways might behave differently, thus affecting the treatment outcome. For example, the *BRAF* gene encodes a serine/threonine kinase and is known to contain activating somatic mutations in melanomas, colorectal cancer and other cancers [1]. In a metastatic colorectal cancer study, it was reported that none of the patients with *BRAF* mutations responded to treatment with the drugs panitumumab or cetuximab [2]. Thus, activating somatic mutations can affect drug sensitivities.

Identifying driver and passenger mutations

When sequencing a cancer genome, somatic single base substitutions can number in the tens of thousands [3]. Sifting through these somatic single nucleotide variants (SNVs) to pin down the few driver mutations implicated in cancer is a challenge. Most researchers concentrate on the somatic mutations that cause missense changes in gene products. This focus helps to reduce the number of mutations for further investigation.

To achieve the ultimate goal of distinguishing driver mutations from passenger mutations, one approach is to sequence many cancer samples and then identify the highly mutated genes and/or the recurrent mutations across all of the samples. The disadvantage of this approach is that many cancer samples need to be sequenced, and it is not straightforward to prioritize genes with a small number of somatic mutations. To supplement this approach, one could look at the severity of the mutations in the gene and assess whether they change the gene's function. It may be possible to detect driver genes in addition to the frequently mutated genes by using this supplementary approach. In this way, some of the genes with a smaller number of mutations would gain stronger support as cancer-causing genes as opposed to background noise [4].

In the past, researchers have used tools such as PolyPhen and SIFT to assess the effect of mutations on protein function. Although these tools are generally useful, they are not trained specifically to identify driver mutations in cancer. Recently, prediction tools that evaluate which mutations specifically drive cancer have been developed. In Table 1, we list some of these publicly available cancer-specific tools. For example, CHASM [4] ranks somatic missense SNVs according to their putative tumorigenic impact. CHASM uses a machine-learning algorithm that has been trained on approximately 50 pre-computed features to distinguish drivers from passenger mutations. CHASM uses a specific passenger mutation rate for each type of cancer. Another example is CanPredict [5], which was one of the first tools for predicting cancer-associated mutations and applies gene ontology knowledge.

*Correspondence: krishnanvg@gis.a-star.edu.sg and ngpc4@gis.a-star.edu.sg
Computational and Mathematical Biology, Genome Institute of Singapore,
60 Biopolis Street, Genome, #02-01, Singapore 138672

Table 1. Available cancer missense mutation prediction tools

Tool ^a	User interface	URL	User input	Highlights	Output
CHASM [4]	Stand-alone software, website	http://wiki.chasmsoftware.org/index.php/Main_Page and http://www.cravat.us/	Genomic coordinates in space or Tab-delimited format. RefSeq, CCDS or Ensembl identifiers, together with the respective amino acid change	Passenger mutation rate information is available for specific types of cancer	Gene annotation CHASM score COSMIC annotation
CanPredict [5]	Website	http://research-public.gene.com/Research/genentech/canpredict	Protein sequence and a list of amino acid changes or A list of RefSeq accession identifiers with amino acid changes	Users can simultaneously analyze various combinations of mutations in a single protein sequence Batch submission is available only with protein RefSeq identifiers	Impact prediction SIFT score and alignment Pfam domain and GO analysis
transFIC [6]	Web service, stand-alone, website	http://bg.upf.edu/transfic	Genomic coordinates or Protein coordinates	Users can upload up to 300 mutations at a time and run up to 20 jobs (on the website)	Gene annotation Transformed prediction scores from SIFT, PolyPhen-2, MutationAssessor and CHASM COSMIC and/or dbSNP annotations
MutationAssessor [7]	Website, Web API	http://mutationassessor.org	Genomic coordinates or Protein coordinates	Users can analyze a list of mutations (on the website) Batch submission is available	Functional Impact score Link to three-dimensional protein structure UniProt and RefSeq identifiers Cancer Gene Census and COSMIC annotations Gene and protein domain annotations
MuSiC [8]	Stand-alone	http://gmt.genome.wustl.edu/genome-music	Mapped reads from a set of tumor and normal sample pairs in BAM format Predicted or validated SNVs and indels from the cohort in MAF format Regions of interest to users (such as exon-intron boundaries) in BED format Any available clinical information	Users can analyze whole genomes and/or exomes	Significantly mutated genes and/or pathways Annotations for known databases Links mutations to user-provided clinical information

^aThis list is not exhaustive. API, Application Programming Interface; BAM, binary SAM; BED, Browser Extensible Data; CCDS, Consensus CoDing Sequence Project; CHASM, Cancer-specific High-throughput Annotation of Somatic Mutations; COSMIC, Catalogue of Somatic Mutations in Cancer; dbSNP, Single Nucleotide Polymorphism database; GO, Gene Ontology; indel, insertion/deletion; MAF, Mutation Annotation Format; MuSiC, Mutational Significance in Cancer; PolyPhen, Polymorphism Phenotyping; RefSeq, Reference Sequence; SIFT, Sorting Intolerant From Tolerant; SNV, single nucleotide variant; transFIC, TRANSformed Functional Impact for Cancer; UniProt, Universal Protein Resource; URL, Uniform Resource Locator.

In this issue of *Genome Medicine*, Abel Gonzalez-Perez, Jordi Deu-Pons and Nuria Lopez-Bigas [6] have developed a computational method called transFIC (TRANSformed Functional Impact for Cancer) to predict somatic mutations that are putative drivers of tumorigenesis. The authors made the initial observation that cancer-associated genes are less likely to have deleterious germline variation than genes that are not involved in cancer. Based on this observation, transFIC first looks at the scores generated by a missense prediction tool such

as SIFT, PolyPhen-2, MutationAssessor [7] or CHASM. It then normalizes the initial prediction scores by taking into account a gene's tolerance to deleterious germline variation. The transformed scores are used to rank the somatic mutations that have functional effects, and mutations with higher transFIC scores are considered candidate cancer drivers. This process improves the performance of the original scores from pre-existing tools, by approximately a twofold to sevenfold increase in the Matthew's correlation coefficient, on various datasets.

In summary, the transFIC prediction tool reported by Gonzalez-Perez *et al.* has many user-friendly features to discriminate cancer driver mutations from mutations that are neutral. TransFIC could be of great use to the cancer research community because it improves the functional impact scores of four well-known tools and uses these transformed scores to prioritize mutations. This tool also has the potential to be useful in cancer resequencing projects to predict the functional impact of somatic cancer mutations.

Beyond driver mutations

The validation of driver mutations will be easier in the future because lower sequencing costs will allow the deep sequencing of tumor samples. Because driver mutations are expected to occur early in the development of cancer cells, these mutations will tend to be present at higher frequencies in a cancer sample than passenger mutations, which occur later. Deep sequencing provides better estimates of mutation frequencies compared to sequencing at medium coverage and therefore deep sequencing helps distinguish driver and passenger mutations.

After distinguishing these two types of mutations, it is crucial to pinpoint the key cancer-causing genes and pathways. Software packages such as MuSiC [8] and MutSig [9] aid in this step by prioritizing genes that are significantly mutated. These packages identify frequently mutated genes, pathways and gene families across a group of patients for various cancer types, and they also highlight clinically relevant mutations. This entire process could allow better treatment. For example, the My Cancer Genome website [10] captures cancer variation and the reported drug responses for various cancers. This information can then make doctors aware of the outcome of a patient's drug response based on the patient's genotype. Discovery and distribution of this knowledge will lead to improved personalized cancer treatment.

Abbreviations

API, Application Programming Interface; BAM, binary SAM; BED, Browser Extensible Data; CCDS, Consensus CoDing Sequence Project; CHASM, Cancer-specific High-throughput Annotation of Somatic Mutations; COSMIC, Catalogue of Somatic Mutations in Cancer; dbSNP, Single Nucleotide Polymorphism database; GO, Gene Ontology; indel, insertion/deletion; MAF, Mutation Annotation Format; MuSiC, Mutational Significance in Cancer; MutSig, Mutation Significance; PolyPhen, Polymorphism Phenotyping; RefSeq, Reference Sequence; SIFT, Sorting Intolerant From Tolerant; SNV, single

nucleotide variant; transFIC, TRANSformed Functional Impact for Cancer; UniProt, Universal Protein Resource; URL, Uniform Resource Locator.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank Dr Francesca Menghi and Dr Joyce Suling Lin for proof-reading. We apologize if we have neglected to cite cancer tools or references because of space limitations.

Published: 26 November 2012

References

1. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, *et al.*: Mutations of the *BRAF* gene in human cancer. *Nature* 2002, **417**:949-954.
2. Di Nicolantonio F, Martini M, Molinari F, Sartore-Bianchi A, Arena S, Saletti P, De Dosso S, Mazzucchelli L, Frattini M, Siena S, Bardelli A: **Wild-type *BRAF* is required for response to panitumumab or cetuximab in metastatic colorectal cancer.** *J Clin Oncol* 2008, **26**:5705-5712.
3. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordóñez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, Beare D, Butler A, Carter RJ, Chen L, Cox AJ, Edkins S, Kokko-Gonzales PI, Gormley NA, Grocock RJ, Haudenschild CD, Hims MM, James T, Jia M, Kingsbury Z, Leroy C, Marshall J, Menzies A, *et al.*: **A comprehensive catalogue of somatic mutations from a human cancer genome.** *Nature* 2010, **463**:191-196.
4. Carter H, Chen S, Isik L, Tyekuceva S, Velculescu VE, Kinzler KW, Vogelstein B, Karchin R: **Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations.** *Cancer Res* 2009, **69**:6660-6667.
5. Kaminker JS, Zhang Y, Watanabe C, Zhang Z: **CanPredict: a computational tool for predicting cancer-associated missense mutations.** *Nucleic Acids Res* 2007, **35**:W595-W598.
6. Gonzalez-Perez A, Deu-Pons J, Lopez-Bigas N: **Improving the prediction of the functional impact of cancer mutations by baseline tolerance transformation.** *Genome Med* 2012, **4**:89
7. Reva B, Antipin Y, Sander C: **Predicting the functional impact of protein mutations: application to cancer genomics.** *Nucleic Acids Res* 2011, **39**:e118.
8. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, Koboldt DC, Mooney TB, Callaway MB, Dooling D, Mardis ER, Wilson RK, Ding L: **MuSiC: identifying mutational significance in cancer genomes.** *Genome Res* 2012, **22**:1589-1598.
9. **Cancer Genome Analysis Tool**
[<https://confluence.broadinstitute.org/display/CGATools/MutSig>]
10. **Personalized Cancer Medicine Resource**
[<http://www.mycancergenome.org/>]

doi:10.1186/gm389

Cite this article as: Krishnan VG, Ng PC: Predicting cancer drivers: are we there yet? *Genome Medicine* 2012, **4**:88.