


Annotation matters: validating the discovery of cancer drivers

Marta Rodríguez-Martínez and Jesper Q. Svejstrup 

Mechanisms of Transcription Laboratory, The Francis Crick Institute, London, UK

ABSTRACT

Advanced sequencing techniques have helped unveil numerous new, potential cancer driver mutations. However, manual curation and analysis of gene and protein annotation are essential to verify such discoveries. Our recent study of *STK19* (Serine Threonine Kinase 19), a previously identified melanoma driver, is a clear example of the importance of such detailed analysis, with both *STK19* gene and protein annotations in frequently used databases having been proven incorrect.

ARTICLE HISTORY

Received 28 July 2020
Revised 31 July 2020
Accepted 3 August 2020

KEYWORDS

STK19; annotation; cancer driver; kinase

In the past few years, targeted sequencing as well as whole genome- or exome-sequencing have allowed researchers to explore the alterations in cancer genomes, with thousands of cancer genomes having been analyzed to allow the identification of new cancer driver genes and mutations. However, the methods for uncovering cancer drivers rely on the accurate annotation of the human genome, which is often imperfect and can therefore lead to misinterpretation of the results. An illustrative example of this imperfection is found in the recent comparison of the RefSeq and Gencode human gene databases, where only 27.5% of transcripts annotated in Gencode are shared by the RefSeq annotation.¹

The gene encoding what is now known as Serine Threonine Kinase 19 (*STK19*) provides a further illustrative example of such misinterpretation.² Indeed, the identified melanoma driver mutation *STK19* D89N³ turned out not to be in the coding region of the expressed form of *STK19*, which in reality lacks the first 110 amino acids shown in database annotations. The mutation is instead non-coding and positioned near the *STK19* transcription start site (TSS), but does not affect gene or protein expression either.² It thus cannot be a cancer driver.

Cutaneous melanoma is the most aggressive form of skin cancer. It typically develops on sun-exposed skin and is linked with DNA damage arising from UV-irradiation. These melanomas have one of the highest tumor mutation burdens of all malignancies, with the alterations typically comprising UV-induced C > T transitions occurring at di-pyrimidines.^{3,4} The high background tumor mutation burden makes it challenging to identify significantly mutated genes that drive this cancer. Alkallas and coworkers recently published a mutation significance study of over 1,000 melanoma exomes, which they combined with multi-omic analysis of 470 cases from the Cancer Genome Atlas.⁵ This approach allowed them to not only identify new significantly mutated genes with co-occurring loss-of-heterozygosity and loss-of-function mutations, but also to rule out a number of false positives, identified based on the proximity of their mutation to the coding area of an unexpressed

gene isoform. Not surprisingly, *STK19* was identified as one of these false positives.⁵

Moreover, recent studies have shown that binding of the transcription factor ETS (ETS1/ETS2) can reduce DNA repair efficiency, resulting in the accumulation of specific mutation signatures near these sites and thus the TSS of certain genes.^{6,7} The mutations caused by ETS binding, which almost certainly include the misannotated *STK19* D89N driver mutation,² can be induced by simple UV-irradiation of cells, and they typically have no effect on gene or protein expression,⁸ corroborating the lack of association with malignant transformation. Intriguingly, mutations in the human *TERT* (Telomerase Reverse Transcriptase) promoter have been found in benign skin nevi, arguing against a role of these well-known mutations in cancer progression, and suggesting that further analysis is required before extending the use of this mutation as a biomarker for cutaneous melanoma.⁹

Unfortunately, in the case of *STK19*, a second potential misconception was faced due to the old annotation of the protein as a serine threonine kinase.¹⁰ The previous kinase activity studies were typically done using protein pull-downs from insect cells^{10,11} and therefore the observed activity could well have been due to contamination, or to an *STK19*-interacting kinase rather than *STK19* itself. Indeed, in spite of repeated attempts, we have been unable to detect *STK19* kinase activity,² and no evidence of *STK19* kinase activity using highly purified protein has ever been reported. Equally damagingly, the *STK19* protein has none of the domain- or sequence features expected of a protein kinase. Indeed, in their response to our recent study uncovering the misconceptions about *STK19*,² the authors of the previous study that identified *STK19* as an NRAS kinase with a role in cancer progression,¹¹ now acknowledge that *STK19* is unlikely to have intrinsic kinase activity.¹²

In summary, although *STK19* represents an unusual case in which both protein function and gene expression characteristics have been misannotated (Figure 1), it represents proof in point

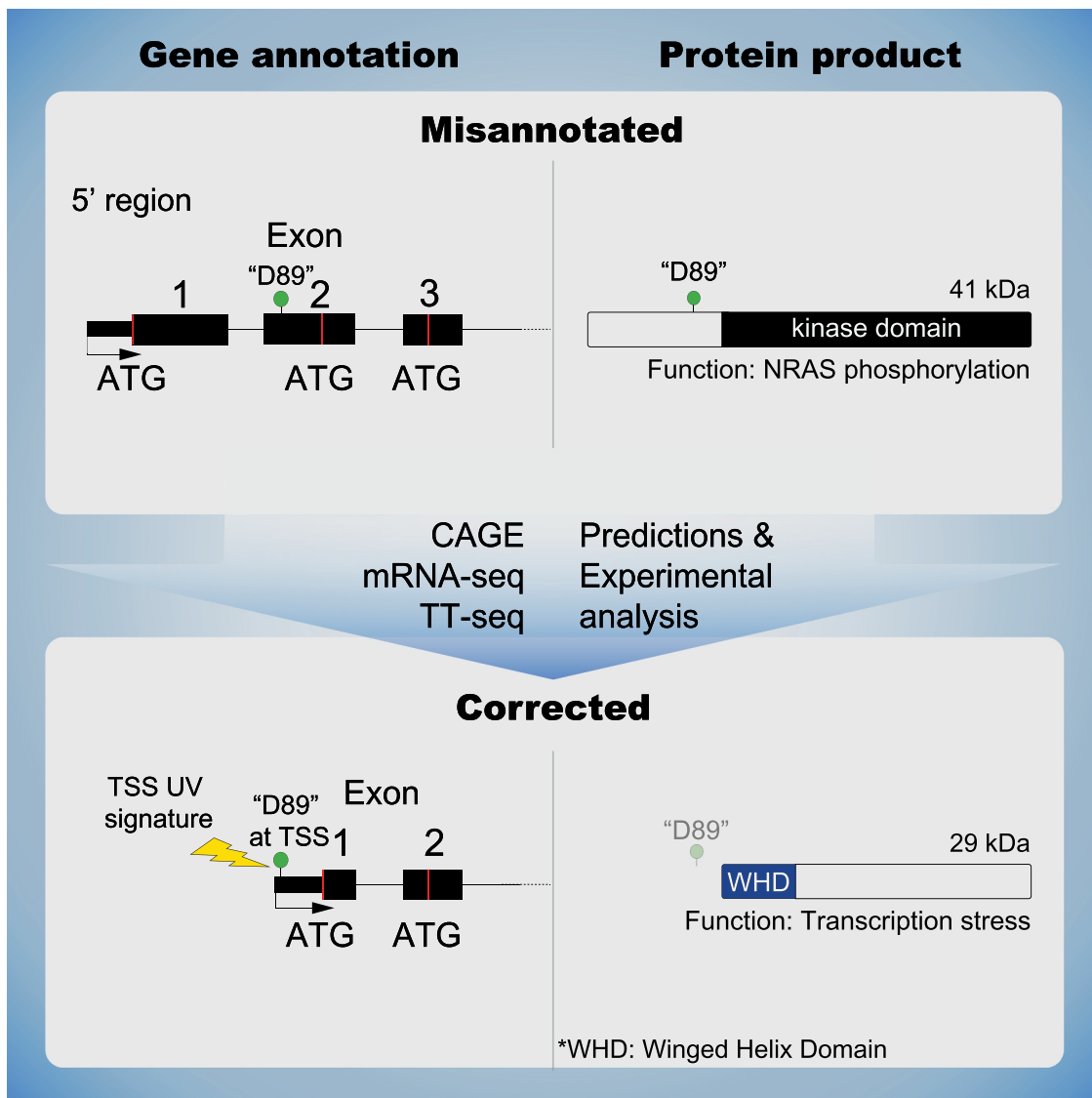


Figure 1. Schematic representation of the current (misannotated) and corrected STK19 gene and protein annotations. ATGs are indicated by red lines and the position of the (non-existent) D89 by a green pin. For gene annotation (left panel), only the 5' region of the gene is shown, as this is where the annotation correction takes place. In the right panel, depicting protein annotation, the respective sizes of the encoded protein, as well as the proposed protein function, is indicated next to the diagram. The central arrow indicates the methods and tools used in² to correct the current gene and protein annotation.

that such mistakes do occur, and that careful manual analysis of basic gene and protein characteristics should be carried out once potential cancer drivers have been identified by sophisticated algorithms. In the case of STK19, basic misconceptions unfortunately led the field to believe that STK19 represents a great opportunity to develop targeted therapies for aggressive melanoma, with two separate studies reporting progress in producing small molecule inhibitors.^{11,13} Our analysis thus also underscores the importance of carefully analyzing newly discovered cancer drivers before resources are dedicated to therapy development.

Disclosure of potential conflict of interest

The authors declare no competing financial interests.

Funding

This work was supported by the Francis Crick Institute (FCI received its core funding from Cancer Research UK [FC001166];H2020 European Research Council [693327];Medical Research Council [FC001166]; Wellcome Trust [FC001166]).

ORCID

Jesper Q. Svejstrup  <http://orcid.org/0000-0003-4964-6147>

References

1. Pertea M, Shumate A, Pertea G, Varabyou A, Breitwieser FP, Chang YC, Madugundu AK, Pandey A, Salzberg SL. CHES: a new human gene catalog curated from thousands of large-scale

- RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* 2018;19:208. doi:10.1186/s13059-018-1590-2.
- Rodriguez-Martinez M, Boissiere T, Noe Gonzalez M, Litchfield K, Mitter R, Walker J, Kjoer S, Ismail M, Downward J, Swanton C, *et al.* Evidence that STK19 is not an NRAS-dependent melanoma driver. *Cell.* 2020;181:1395–1405 e1311. doi:10.1016/j.cell.2020.04.014.
 - Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D, Li L, Place C, *et al.* A landscape of driver mutations in melanoma. *Cell.* 2012;150:251–263. doi:10.1016/j.cell.2012.06.024.
 - Cancer Genome AN. Genomic classification of cutaneous melanoma. *Cell.* 2015;161:1681–1696. doi:10.1016/j.cell.2015.05.044.
 - Alkallas R, Lajoie M, Moldoveanu D, Vo Hoang K, Lefrançois P, Lingrand M, Ahanfeshar-Adams M, Watters K, Spatz A, Zippin JH, *et al.* Multi-omic analysis reveals significantly mutated genes and DDX3X as a sex-specific tumor suppressor in cutaneous melanoma. *Nature Cancer.* 2020;1(6):635–652. doi:10.1038/s43018-020-0077-8.
 - Perera D, Poulos RC, Shah A, Beck D, Pimanda JE, Wong JW. Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. *Nature.* 2016;532:259–263. doi:10.1038/nature17437.
 - Sabarinathan R, Mularoni L, Deu-Pons J, Gonzalez-Perez A, Lopez-Bigas N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature.* 2016;532:264–267. doi:10.1038/nature17661.
 - Fredriksson NJ, Elliott K, Filges S, Van den Eynden J, Stahlberg A, Larsson E. Recurrent promoter mutations in melanoma are defined by an extended context-specific mutational signature. *PLoS Genet.* 2017;13:e1006773. doi:10.1371/journal.pgen.1006773.
 - Colebatch AJ, Ferguson P, Newell F, Kazakoff SH, Witkowski T, Dobrovic A, Johansson PA, Saw RPM, Stretch JR, McArthur GA, *et al.* Molecular genomic profiling of melanocytic nevi. *The Journal of Investigative Dermatology.* 2019;139:1762–1768. doi:10.1016/j.jid.2018.12.033.
 - Gomez-Escobar N, Chou CF, Lin WW, Hsieh SL, Campbell RD. The G11 gene located in the major histocompatibility complex encodes a novel nuclear serine/threonine protein kinase. *J Biol Chem.* 1998;273:30954–30960. doi:10.1074/jbc.273.47.30954.
 - Yin C, Zhu B, Zhang T, Liu T, Chen S, Liu Y, Li X, Miao X, Li S, Mi X, *et al.* Pharmacological targeting of STK19 inhibits oncogenic NRAS-driven melanomagenesis. *Cell.* 2019;176:1113–1127 e1116. doi:10.1016/j.cell.2019.01.002.
 - Yin C, Zhu B, Li X, Goding CR, Cui R. A reply to “evidence that STK19 is not an NRAS-dependent melanoma driver”. *Cell.* 2020;181:1406–1409 e1402. doi:10.1016/j.cell.2020.04.029.
 - Qian L, Chen K, Wang C, Chen Z, Meng Z, Wang P. Targeting NRAS-mutant cancers with the selective STK19 kinase inhibitor chelidonine. *Clin Cancer Res.* 2020;26:3408–3419.