Check for updates

SOFTWARE TOOL ARTICLE

REVISED InferAMP, a python web app for copy number inference from discrete gene-level amplification signals noted in clinical tumor profiling reports [version 3; peer review: 2 approved]

Previously titled: InferCNV, a python web app for copy number inference from discrete gene-level amplification signals noted in clinical tumor profiling reports

Paraic A. Kenny [1,2]

[1]Kabara Cancer Research Institute, Gundersen Medical Foundation, La Crosse, WI, 54601, USA
[2]Department of Medicine, University of Wisconsin-Madison, Madison, WI, 53705, USA

v3 **First published:** 06 Jun 2019, **8**:807 (https://doi.org/10.12688/f1000research.19541.1)
Second version: 25 Jun 2019, **8**:807 (https://doi.org/10.12688/f1000research.19541.2)
Latest published: 26 Sep 2019, **8**:807 (https://doi.org/10.12688/f1000research.19541.3)

Abstract
As somatic next-generation sequencing gene panel analysis in advanced cancer patients is becoming more routine, oncologists are frequently presented with reports containing lists of genes with increased copy number. Distinguishing which of these amplified genes, if any, might be driving tumor growth and might thus be worth considering targeting can be challenging. One particular issue is the frequent absence of genomic contextual information in clinical reports, making it very challenging to determine which reported genes might be co-amplified and how large any such amplicons might be. We describe a straightforward Python web app, InferAMP, into which healthcare professionals may enter lists of amplified genes from clinical reports. The tool reports (1) the likely size of amplified genomic regions, (2) which reported genes are co-amplified and (3) which other cancer-relevant genes that were not evaluated in the assay may also be co-amplified in the specimen. The tool is accessible for web queries at http://inferamp.org.

Keywords
cancer, genetic testing, copy number variation, gene amplification, oncology, targeted therapy

python This article is included in the Python collection.

Open Peer Review

Reviewer Status ✔ ✔

	Invited Reviewers	
	1	**2**
REVISED **version 3** published 26 Sep 2019	✔ report	✔ report
REVISED **version 2** published 25 Jun 2019	? report	? report
version 1 published 06 Jun 2019		

1 **Andrew C. Nelson**, University of Minnesota, Minneapolis, USA

2 **Oscar Krijgsman**, Netherlands Cancer Institute, Amsterdam, The Netherlands

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Paraic A. Kenny (pakenny@gundersenhealth.org)

Author roles: Kenny PA: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project was supported by the Gundersen Medical Foundation. P.K. holds the Dr. Jon & Betty Kabara Endowed Chair in Precision Oncology.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Kenny PA. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Kenny PA. **InferAMP, a python web app for copy number inference from discrete gene-level amplification signals noted in clinical tumor profiling reports [version 3; peer review: 2 approved]** F1000Research 2019, **8**:807 (https://doi.org/10.12688/f1000research.19541.3)

First published: 06 Jun 2019, **8**:807 (https://doi.org/10.12688/f1000research.19541.1)

REVISED **Amendments from Version 2**

We have improved the web tool based on reviewer recommendations. Key updates include (1) ability to input genomic results from seven additional commercial cancer genomic assay reports, (2) ability to enter custom gene lists from unsupported assays, (2) ability to select "verbose" output to provide additional data for expert users.

Any further responses from the reviewers can be found at the end of the article

Introduction

Focal somatic gene copy number changes are a widespread event in tumor evolution[1]. Although these regions of amplification may be large, encompassing many hundreds of genes, typically only one or a small number of genes within the amplified regions are involved in driving tumor growth. Identification of the key driver genes within recurrent amplicons has led to the approval of some therapies that have changed clinical practice (e.g. anti-ERBB2 agents[2]); however, targeting other amplified genes such as FGFR family members[3,4], *EGFR*[5] or *KIT*[6] has frequently proved disappointing. Nevertheless, even some of the more negative trials include occasional strong responses, indicating that sub-populations of patients with amplification of these oncogenes may experience clinical benefit if they can be identified.

With the goal of individualizing treatment for cancer patients, next-generation sequencing from tumor specimens is becoming widely adopted[7]. In addition to somatic point mutations, several of these assays report copy number changes in assayed genes. Reports for physicians typically present a list of amplified genes without providing a genomic context, leaving physicians and molecular tumor boards to hypothesize which of the listed genes might be driver genes suitable for therapeutic targeting. Given the poor response rates that have often been observed in clinical studies with amplified genes (compared to targeting genes activated by point mutation or fusion), physicians are often appropriately cautious about deciding whether a reported amplified gene may be actionable. Thus, many patients are spared receiving ineffective therapies, but a subgroup of patients who may experience clinical benefit do not get that opportunity.

Here we provide an easy-to-use web tool for analyzing clinical genomics reports of amplified genes. It determines (1) the likely size of amplified genomic regions, (2) which reported genes are co-amplified and (3) which other cancer-relevant genes that were not evaluated in the assay may also be co-amplified in the specimen.

The primary goals are to allow healthcare professionals to determine whether the amplification region surrounding a particular oncogene is relatively small and lacking in other likely candidate cancer drivers (which may indicate increased likelihood that the analyzed gene is a driver) from larger amplicons with additional candidate driver genes (which would suggest a reduced probability that the reported gene is a driver). The approach was developed and most extensively tested to analyze the widely used Foundation One test (329 genes) provided by Foundation Medicine. The web tool also allows selection of assays from several other test vendors, as well as pasting of a custom gene list from any currently unsupported assays.

Methods

Implementation

InferAMP is written in Python 2.7 with Flask and implemented as a web service running on the Google App Engine (http://inferamp.org). Additional supplied requirements are (1) the coordinates of genes in the human genome, 'coordinates.txt' (hg38, UCSC genome browser), (2) The gene list from the assay of interest, 'foundationone.txt' and (3) a file listing genes recurrently altered in cancer from COSMIC[8] (retrieved 9/18/2019), 'cosmic.txt'. This list includes all 723 Tier 1 and 2 genes in the COSMIC gene census.

An html page with a single query window allows the user to enter a comma-delimited list of genes reported as being amplified. The entry is passed to the script and parsed into individual gene names. An error check is performed to confirm that all entered gene names correspond to gene names in the genome used. The entered genes are considered to be amplified, while the other genes in the assay are considered to be not amplified.

A simplified schema of how the algorithm works is presented in Figure 1, which depicts a chromosomal region containing 30 genes. Seven cancer-relevant genes are present, five of which are evaluated by the genomic assay (Figure 1A). In this test example, three genes were reported to be amplified (Figure 1B). Running the algorithm identifies these amplified genes (8, 13, 20) as well as the nearby assayed genes that are not reported amplified (4, 27). The algorithm considers all genes located between genes reported as amplified to also be amplified (Figure 1C, red shaded region). Because not every gene is assayed, precisely delineating the boundaries of an amplicon is not possible. To address this, the algorithm determines the nearest non-amplified gene at each end of the amplicon and infers that the genes located up to, but not including that gene may be possibly amplified (Figure 1D). The script then returns an html report page listing the entered genes, the amplicons into which they fall (in many cases, several discrete genes will be consolidated into a single amplicon), and also the other cancer-relevant genes within these regions that may be co-amplified with the reported genes. All genes reported include hyperlinks to that gene's page on COSMIC. A checkbox on the form allows selection of "verbose" output which additionally includes (1) the complete list of assayed genes, (2) the identity of all genes in the inferred amplicons and (3) the identity of the assayed genes reported as non-amplified which are used to infer amplicon boundaries.

Operation

InferAMP is accessed via a web browser and has been tested on commonly used browsers such as Chrome, Firefox, Internet Explorer and Microsoft Edge.

Use cases

Three use cases taken from genomic reports of patients at our clinic are presented:

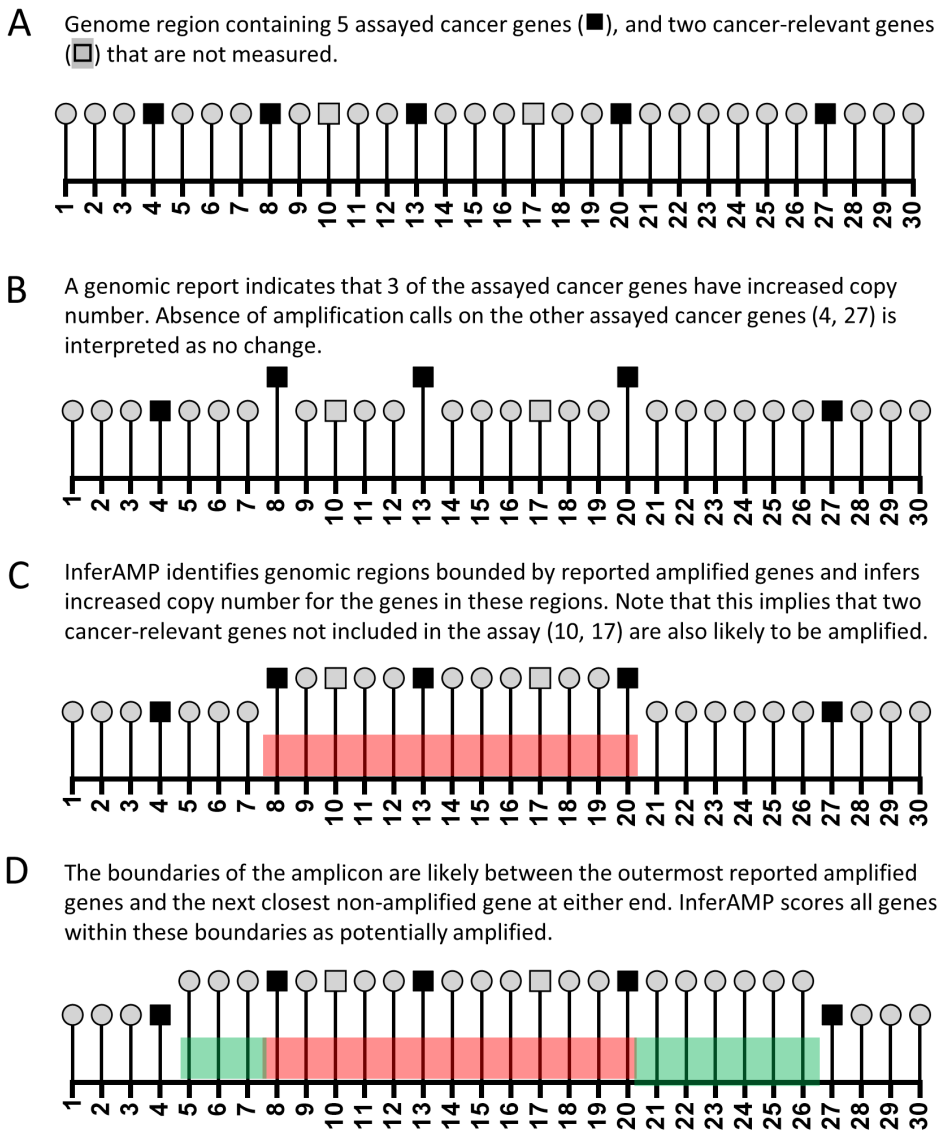

Figure 1. Schematic representation of amplicon boundary inference approach. (**A**). Schematic diagram of a model genomic region with 30 numbered genes, which include a total of 7 cancer-relevant genes. (**B**) Input scenario for algorithm: a clinical genomics report noting amplification of three genes in this region. (**C**) Copy number inference for genes in regions bounded by reported amplified genes. (**D**) Copy number inference for genes in the regions between the outermost reported amplified genes and the nearest reported non-amplified gene.

Use case 1

A case of esophageal adenocarcinoma with eight reported amplified genes (Table 1), which were resolved by InferAMP to four amplicons. The co-amplification of *FGF3*, *FGF4* and *FGF10* with *CCND1* (which is likely the driver gene in this amplicon[9]) might indicate that consideration of FGFR inhibitors may not be helpful if these FGF genes are simply co-amplified passenger genes.

Use case 2

A case of soft tissue sarcoma with five reported amplified genes (Table 2) which were resolved into three amplicons. In the absence of genomic context information, both *PDGFRA* and *KIT* might be considered as potentially druggable targets. The demonstration that these are likely co-amplified in a relatively small amplicon might provide further support to this hypothesis. Clinically, both targets are inhibited by imatinib, making joint targeting with a single agent feasible in this case.

Use case 3

The third example is a breast cancer case from our clinic with three reported amplified genes (Table 3). The report highlights one region on chromosome 5, and two regions on chromosome 7. The latter predicted amplicons share a nearby boundary at 7q22.3 suggesting the possibility that there is a regional amplification on 7q encompassing both sets of genes. In this case, *MET*

Table 1. Use case – Esophageal adenocarcinoma with reported amplification of *CCND1*, *MAP2K1*, *RICTOR*, *FGF10*, *FGF19*, *FGF3*, *FGF4* and *MCL1*.

Gene reported amplified (chromosomal location)	Number of potentially co-amplified genes (chromosomal region	Genes annotated as recurrently altered in cancer by COSMIC
MCL1 (Chr1q21.2)	8 (Chr1p12–1q23.1)	*BCL9*, *PDE4DIP*, *ARNT*, *MLLT11*, *TPM3*, *MUC1*, *LMNA*, PRCC
RICTOR (Chr5p13.1), *FGF10* (Chr5p12)	75 (Chr5p13.2–5q11.2)	*LIFR*, *IL6ST*
CCND1 (Chr11q13.3), *FGF19* (Chr11q13.3), *FGF4* (Chr11q13.3), *FGF3* (Chr11q13.3)	214 (Chr11q13.1–11q13.5)	*CCND1*, *NUMA1*
MAP2K1 (Chr15q22.31)	224 (Chr15q15.1–15q22.31)	*B2M*, *USP8*, *MYO5A*, *C15ORF65*, *TCF12*, *MAP2K1*

Table 2. Use case – Soft tissue sarcoma with reported amplification of *KIT*, *PDGFRA*, *MDM2*, *RICTOR* and *FGF10*.

Gene reported amplified (chromosomal location)	Number of potentially co-amplified genes (chromosomal region)	Genes annotated as recurrently altered in cancer by COSMIC
PDGFRA (Chr4q12), *KIT* (Chr4q12)	97 (Chr4p15.31–4q12)	*SLC34A2*, *RHOH*, *PHOX2B*, *FIP1L1*, *CHIC2*, *PDGFRA*, *KIT*
RICTOR (Chr5p13.1), *FGF10* (Chr5p12)	75 (Chr5p13.2–5q11.2)	*LIFR*, *IL6ST*
MDM2 (Chr12q15)	48 (Chr12q14.1–12q15)	*LRIG3*, *WIF1*, *HMGA2*, *MDM2*

Table 3. Use Case – Triple Negative breast cancer with reported amplification of *RICTOR*, *CDK6* and *MET*.

Gene reported amplified (chromosomal location)	Number of potentially co-amplified genes (chromosomal region)	Genes annotated as recurrently altered in cancer by COSMIC
RICTOR (Chr5p13.1)	44 (Chr 5p13.2–5p12)	*LIFR*
CDK6 (Chr7q21.2)	190 (Chr 7q21.12–7q22.3)	*AKAP9*, *CDK6*, *TRRAP*, *CUX1*
MET (Chr7q31.2)	91 (Chr 7q22.3–7q32.1)	*MET*, *POT1*, *SND1*

was judged to be a possible driver amplicon, and the patient had a very strong response to a MET inhibitor[10].

Discussion

We have described a straightforward tool to provide additional genomic context to aid interpretation of amplifications in somatic cancer sequencing reports. Use of this tool may aid decision-making by healthcare professionals about therapeutic options.

The method relies on the accuracy with which test vendors report gene amplification calls. Larger gene panels will also allow better resolution of amplicon boundaries. In testing, we identified a small number of cases in which two amplicons were inferred in very close proximity (e.g. *Use case 3*), which raises the possibility that the assayed gene between the two regions is erroneously not called as amplified. In cases with two or more closely co-located amplicons, users should consider that there is a strong possibility of a regional amplification encompassing both predicted amplicons. To assist with this determination, such cases are explicitly flagged in the reports. Future assays with larger number of genes or more sensitive amplification calling algorithms will likely permit more accurate refining of the boundaries of individual amplicons.

Because the coverage across the genome is somewhat sparse, refining the amplicon boundaries is more challenging than with a more high-density approach like SNP arrays. The primary purpose is to list genes that are potentially co-amplified with a gene identified by a test vendor as possibly actionable in order to allow healthcare professionals to gain further insight into the likelihood that the listed gene is truly the driver gene in that amplicon. Accordingly, we do not distinguish in the report between genes that are likely co-amplified (red genes, Figure 1) from the boundary region genes which are possibly co-amplified (green genes, Figure 1). In any case in which a healthcare

professional might consider targeting a non-assayed gene predicted by this algorithm to be amplified (e.g. *LIFR*[11] in *Use case 2* and *Use case 3*), further clinical testing to directly confirm gene amplification would be warranted.

Although clinical NGS analysis of somatic changes in gene copy number is relatively new, comparison studies that have been performed against more established technologies such as gene-specific FISH probes and array comparative genomic hybridization have shown strong concordance[12,13]. Indeed, the ability to assess a large number of loci in parallel may be especially advantageous compared to current clinical FISH assays. Within the limits of our study (primarily focusing on the Foundation One assay and having no direct comparisons with alternative assay methodologies), we believe that our ability to easily infer contiguous amplicons involving multiple co-amplified assayed genes also supports the contention that these NGS assessments of copy number changes may be quite robust.

Data availability

All data underlying the results are available as part of the article and no additional source data are required.

Software availability

Software available at: http://inferamp.org.

Source code available from: https://github.com/paraickenny/inferAMP.

Archived source code at time of publication: http://doi.org/10.5281/zenodo.3457522[14].

License: MIT License.

References

1. Zack TI, Schumacher SE, Carter SL, *et al.*: **Pan-cancer patterns of somatic copy number alteration.** *Nat Genet.* 2013; **45**(10): 1134–40.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Parakh S, Gan HK, Parslow AC, *et al.*: **Evolution of anti-HER2 therapies for cancer treatment.** *Cancer Treat Rev.* 2017; **59**: 1–21.
PubMed Abstract | Publisher Full Text
3. Lim SH, Sun JM, Choi YL, *et al.*: **Efficacy and safety of dovitinib in pretreated patients with advanced squamous non-small cell lung cancer with FGFR1 amplification: A single-arm, phase 2 study.** *Cancer.* 2016; **122**(19): 3024–31.
PubMed Abstract | Publisher Full Text
4. Van Cutsem E, Bang YJ, Mansoor W, *et al.*: **A randomized, open-label study of the efficacy and safety of AZD4547 monotherapy versus paclitaxel for the treatment of advanced gastric adenocarcinoma with FGFR2 polysomy or gene amplification.** *Ann Oncol.* 2017; **28**(6): 1316–24.
PubMed Abstract | Publisher Full Text
5. Sepúlveda-Sánchez JM, Vaz MA, Balañá C, *et al.*: **Phase II trial of dacomitinib, a pan-human *EGFR* tyrosine kinase inhibitor, in recurrent glioblastoma patients with EGFR amplification.** *Neuro Oncol.* 2017; **19**(11): 1522–31.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Hodi FS, Corless CL, Giobbie-Hurder A, *et al.*: **Imatinib for melanomas harboring mutationally activated or amplified *KIT* arising on mucosal, acral, and chronically sun-damaged skin.** *J Clin Oncol.* 2013; **31**(26): 3182–90.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Tan O, Shrestha R, Cunich M, *et al.*: **Application of next-generation sequencing to improve cancer management: A review of the clinical effectiveness and cost-effectiveness.** *Clin Genet.* 2018; **93**(3): 533–44.
PubMed Abstract | Publisher Full Text
8. Tate JG, Bamford S, Jubb HC, *et al.*: **COSMIC: the Catalogue Of Somatic Mutations In Cancer.** *Nucleic Acids Res.* 2019; **47**(D1): D941–D7.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Qie S, Diehl JA: **Cyclin D1, cancer progression, and opportunities in cancer treatment.** *J Mol Med (Berl).* 2016; **94**(12): 1313–26.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Parsons BM, Meier DR, Gurda GT, *et al.*: **Exceptional Response to Crizotinib in an MET-Amplified Triple-Negative Breast Tumor.** *JCO Precis Oncol.* 2017; **1**: 1–6.
Publisher Full Text
11. Hall BR, Cannon A, Thompson C, *et al.*: **Utilizing cell line-derived organoids to evaluate the efficacy of a novel LIFR-inhibitor, EC359 in targeting pancreatic tumor stroma.** *Genes Cancer.* 2019; **10**(1–2): 1–10.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Seed G, Yuan W, Mateo J, *et al.*: **Gene Copy Number Estimation from Targeted Next-Generation Sequencing of Prostate Cancer Biopsies: Analytic Validation and Clinical Qualification.** *Clin Cancer Res.* 2017; **23**(20): 6070–7.
PubMed Abstract | Publisher Full Text
13. Su D, Zhang D, Chen K, *et al.*: **High performance of targeted next generation sequencing on variance detection in clinical tumor specimens in comparison with current conventional methods.** *J Exp Clin Cancer Res.* 2017; **36**(1): 121.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Kenny P: **paraickenny/inferAMP: inferAMP command line and web tool versions (Version v1.0.3).** *Zenodo.* 2019.
http://www.doi.org/10.5281/zenodo.3457522

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 3

Reviewer Report 03 October 2019

https://doi.org/10.5256/f1000research.22814.r54388

© 2019 Nelson A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Andrew C. Nelson**
Department of Laboratory Medicine and Pathology, School of Medicine, University of Minnesota, Minneapolis, MN, USA

I appreciate the additional improvements that Dr. Kenny has made to the InferAMP tool and the edits made to clarify and strengthen the text.

This is a very useful tool for the molecular genetic pathology field and I fully approve of the manuscript.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular genetic pathology, anatomic pathology, breast cancer, gynecologic cancer, colorectal cancer, tumor microenvironment, cancer biology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 30 September 2019

https://doi.org/10.5256/f1000research.22814.r54389

© 2019 Krijgsman O. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Oscar Krijgsman**
Division of Molecular Oncology and Immunology, Netherlands Cancer Institute, Amsterdam, The Netherlands

I have no further comments after reading the revised version from the author.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, DNA copynumber profiling

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Reviewer Report 15 August 2019

https://doi.org/10.5256/f1000research.21653.r51604

© 2019 Krijgsman O. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Oscar Krijgsman
Division of Molecular Oncology and Immunology, Netherlands Cancer Institute, Amsterdam, The Netherlands

The author describes a very straight-forward tool to infer possibly amplified genomic regions based on the amplification status of single genes from a targeted panel. InferAMP reports the genes that are likely co-amplified and in addition infers the size and the number of potentially co-amplified genes that were not assessed in the assay and outputs the COSMIC genes found in the genomic region.
This tool is developed for oncologists to better understand the reported amplified genes in genomic context. The developed web-based tool is therefore important and will be the preferred way of using this tool for oncologists. It is also good to see the full code and data tables used are available on Github.

InferAMP is a very easy tool to use and functions as advertised. The associated manuscript is well readable and explains the functionality satisfactory. However, after reading the manuscript and testing InferAMP I have a few questions that need to be answered.

Points to be addressed:

1. The web-based tool is now only suitable for FoundationOne assays. Although this is very useful for many oncologists not all institutes use this assay. I am aware that the the command-line version of the tool has the possibility to provide different genes set and therefore works with additional assays. However, a command-line version will not be suitable for oncologists. Additional functionality of the web-based would greatly improve the usability of InferAMP.

2. The output of InferAMP includes the genes that are possibly co-amplified and the genes in the amplicon that are mentioned in the COSMIC database. In addition, the results mention the potentially co-amplified genes (for example 214 genes when running CCND1, FGF3, FGF4 and FGF19, Use case 1). A list of these genes is currently not available. It would be useful to output these lists of genes in addition to the COSMIC as not all targetable genes will necessarily be in the COSMIC list.

3. Currently the COSMIC list is used to provide an additional rationale to prioritize genes and identify genes suitable for targeting. The rationale for using this list is not provided in the text. A little more background and explanation for choosing this list would be helpful. In addition, which COSMIC list is used exactly? COSMIC cancer census list[1] or all genes mentioned in COSMIC?

Points for consideration but not necessary for this manuscript:

1. Furthermore, I was wondering whether germline CNVs could affect the results of targeted panels, especially since patient-matched blood reference samples are not often used. If so, would it not be nice to add CNV data from for example the Database of Genomic Variants (http://dgv.tcag.ca)? This could identify amplifications that are not somatic but merely 'normal' differences between individuals (CNVs).

References

1. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA: The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers.*Nat Rev Cancer*. **18** (11): 696-705 PubMed Abstract | Publisher Full Text

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, DNA copynumber profiling

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Sep 2019

Paraic Kenny, University of Wisconsin-Madison, Madison, USA

I greatly appreciate the thoughtful and supportive feedback and suggestions. These have been very helpful in revising the manuscript.

Major points:
1. We have added gene lists corresponding to seven additional cancer genomic reports (Foundation CDx, Trusight, Tempus, Caris etc.). We have also added a text box into which users can paste custom lists of genes for other unsupported assays.
2. We have added a new checkbox to the input page allowing users to request "verbose" output. If this is selected, all genes in the inferred amplicon are listed (i.e. not just the COSMIC genes).
3. We used the full COSMIC cancer gene census list (723 Tier 1 and 2 genes). The goal is to indicate to users where other potentially cancer-relevant genes are likely co-amplified with the reported genes so that users may consider whether the reported gene is truly the cancer driver in that region.

Minor Points:
1. Germline CNVs are an interesting question, but beyond the scope of the current project. This is a problem best left to the developers/interpreters/reporters of the various genetic assays. Our goal has been to provide an easy to use tool with which to assess genes reported as amplified after having passed some QC by commercial testers or clinical labs.

Competing Interests: No competing interests were disclosed.

Reviewer Report 11 July 2019

https://doi.org/10.5256/f1000research.21653.r50604

© 2019 Nelson A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Andrew C. Nelson
Department of Laboratory Medicine and Pathology, School of Medicine, University of Minnesota, Minneapolis, MN, USA

In this manuscript, Dr. Kenny describes a useful and straightforward web-based informatics tool that enables genetics professionals to rapidly and potentially more accurately infer genomic relationships of copy number alterations (CNA) disparately reported as individual genes in clinical genomics reports. This will aid both clinicians in working/teaching conferences (such as molecular tumor boards) and translational researchers reviewing archived clinical data.

The rationale for the tool and the overview of the method is clear. The web interface is easy to use and provides a clear result report; code is available at Github.

I have several points for improvement, clarification, and commentary on the manuscript and the software tool:

Major Considerations

1. I recommend that the assayed genes without amplification which are used by the algorithm as boundary genes be explicitly reported in the interface. In Fig 1D, these are represented as genes 4

and 27; as a minor point, I believe the legend for Fig 1D needs to be updated to be congruent with the figure. In the current interface report, the region of potential contiguous amplifications are only reported as cytobands. Specifically reporting the identity of these boundary genes would help genetics professionals to more precisely quality control the interpretation of both the original genomics report and the interface output.

2. The tool is currently static configured for FoundationOne (F1); it would be beneficial to configure the web interface to allow (perhaps through a drop down menu) selection of other reasonably common cancer genomics reports which offer CNA data such as: Caris, Tempus, and the Illumina products TST170 and TSO500 (which are being deployed by some academic laboratories for clinical testing). Ultimately, an option to input a text file with HGVS gene names would be beneficial.

3. I would suggest an expanded introduction or discussion about the status of clinical utility for gene amplification from comprehensive genomic profiling NGS assays. For example, only ERBB2 amplification is specifically included in the FDA premarket approval of the F1 assay (https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=p170019, accessed 07/10/2019). The most common diagnostic approach for gene amplification in clinical practice and clinical trials is FISH, frequently normalized to the centromeric copy number of the same chromosome on which the gene of interest resides (for example, see references 2-5 of the manuscript). There are, of course, significant pitfalls in the use of FISH as a longer established clinical standard, but I am not immediately aware of comprehensive accuracy assessments for somatic CNA analysis across broad numbers of genes by NGS, particularly in unpaired tests (i.e. no patient-specific normal sample for comparisons). A more in-depth review of these considerations and limitations in the introduction or discussion would equip readers to better interpret the expected output of datasets and any results generated using the tool.

Minor Comments and Commentary

1. Please clarify in the methods if the COSMIC Cancer Gene Census list in the tool has been filtered for type of alteration (i.e. SNV and small indel vs. gene rearrangement vs. CNA). Or is it the full 719 CGC list as referenced?

2. It may be valuable to consider cross-referencing the cytobands and/or cancer-associated genes within the predicted CNA region against regions/genes commonly copy number altered in pan-cancer analyses. For example: Beroukhim *et al*.[Nature 2010[1], Zack *et al*. Nature Genetics 2013[2], Ohshima et al. Scientific Reports 2017[3]. Presenting this information in the report interface in future iterations would improve the quality and utility of the software tool.

The ability to specifically annotate case results against disease-specific databases (i.e. TCGA projects) would also be valuable (see below as well).

Commentary on use cases.

The use cases provide a reasonable snapshot of how the tool functions and can be applied. Specifically, use case 1 highlights a common CNA in cancer (11q13.3) which has been described previously (Zack *et al*. Nature Genetics 2013[2]) and this specific amplicon is commonly seen in clinical genomics reports at our institution. The question of utility around FGFR inhibitors has arisen in our own molecular tumor boards based on the co-amplified FGF ligand genes, when in reality CCND1 is most likely the significant driver.

Use case 2 highlights potentially actionable genes (KIT, PDGFRA) located in the same cytoband (4q12).

However, the region of potential co-amplification crosses the Chr4 centromere (Chr4p15.31-4q12). It may be beneficial to further discuss how carefully to interpret amplicons which span both chromosomal arms. In this particular case, I infer that no reported amplified genes are present on 4p; but no boundary genes were analyzed centromeric to the PDGFRA locus. This seems worth a deeper discussion in the text.

Use case 3 highlights an important point about potential drop out of amplification calls in NGS data. Dr. Kenny proposes that the entire amplicon on 7q is amplified, which is a reasonable hypothesis. It is important to note that specific bioinformatics pipelines and wet-bench library prep methods will not have equivalent analytic sensitivity/specificity for amplification calls for every captured gene. Case 3 is also interesting because the amplicon included CDK6, which is a target for FDA approved drugs in hormone receptor positive metastatic breast cancer. It might be interesting for Dr. Kenny to further comment on whether CDK inhibitors were considered less likely based on the triple negative hormone receptor status (PMID: 30038670[4]).

I ran several of my own cases through InferAmp. Of interest, a case of high grade serous epithelial ovarian carcinoma is illustrative of the utility of the tool. This case had 8 separate gene amplifications reported by F1, including: KRAS, FGF23, FGF6, and CCND2 (all on Chr12p). The static clinical report indicated KRAS amplification was potentially relevant for MEK inhibitor therapy. Nevertheless, my molecular tumor board noted this amplification contig and noted that 12p is commonly amplified in high-grade serous carcinoma (TCGA); therefore it was unlikely to be a patient-specific driver alteration. The InferAmp quickly and accurately identified this amplicon, with the caveat that a potentially erroneous “break” was present at 12p13.1 (similar to Dr. Kenny’s use case 3).

Finally, the final sentence of the manuscript cautioning that inferred co-amplified genes should be confirmed by CLIA-validated assays cannot be emphasized enough.

References

1. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, Maher E, Kaye FJ, Sasaki H, Tepper JE, Fletcher JA, Tabernero J, Baselga J, Tsao MS, Demichelis F, Rubin MA, Janne PA, Daly MJ, Nucera C, Levine RL, Ebert BL, Gabriel S, Rustgi AK, Antonescu CR, Ladanyi M, Letai A, Garraway LA, Loda M, Beer DG, True LD, Okamoto A, Pomeroy SL, Singer S, Golub TR, Lander ES, Getz G, Sellers WR, Meyerson M: The landscape of somatic copy-number alteration across human cancers.*Nature*. 2010; **463** (7283): 899-905 PubMed Abstract | Publisher Full Text

2. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhsng CZ, Wala J, Mermel CH, Sougnez C, Gabriel SB, Hernandez B, Shen H, Laird PW, Getz G, Meyerson M, Beroukhim R: Pan-cancer patterns of somatic copy number alteration.*Nat Genet*. 2013; **45** (10): 1134-40 PubMed Abstract | Publisher Full Text

3. Ohshima K, Hatakeyama K, Nagashima T, Watanabe Y, Kanto K, Doi Y, Ide T, Shimoda Y, Tanabe T, Ohnami S, Ohnami S, Serizawa M, Maruyama K, Akiyama Y, Urakami K, Kusuhara M, Mochizuki T, Yamaguchi K: Integrated analysis of gene expression and copy number identified potential cancer driver genes with amplification-dependent overexpression in 1,454 solid tumors. *Scientific Reports*. 2017; **7** (1). Publisher Full Text

4. Pernas S, Tolaney SM, Winer EP, Goel S: CDK4/6 inhibition in breast cancer: current practice and future directions.*Ther Adv Med Oncol*. 2018; **10**: 1758835918786451 PubMed Abstract | Publisher Full Text

5. Chen Y, McGee J, Chen X, Doman TN, Gong X, Zhang Y, Hamm N, Ma X, Higgs RE, Bhagwat SV,

Buchanan S, Peng SB, Staschke KA, Yadav V, Yue Y, Kouros-Mehr H: Identification of druggable cancer driver genes amplified across TCGA datasets.*PLoS One*. 2014; **9** (5): e98293 PubMed Abstract | Publisher Full Text

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular genetic pathology, anatomic pathology, breast cancer, gynecologic cancer, colorectal cancer, tumor microenvironment, cancer biology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Sep 2019

Paraic Kenny, University of Wisconsin-Madison, Madison, USA

I greatly appreciate the thoughtful and supportive feedback and suggestions. These have been very helpful in revising the manuscript.

Major:
1. We have added a new checkbox to the input page allowing users to request “verbose” output. In this case, the boundary genes are explicitly listed in the report.
2. We have added gene lists corresponding to seven additional cancer genomic reports (Foundation CDx, Trusight, Tempus, Caris etc.). We have also added a text box into which users can paste custom lists of genes for other unsupported assays. Because the accuracy of amplicon inference is related to the density/distribution of assayed genes across the genome, we added a comment on this to reports generated using panels with low gene numbers.
3. We have added a brief section to the discussion to discuss this, citing a couple of recent publications addressing the cross-comparison of NGS and FISH-based methodologies for CNA assessment.

Minor:
1. We used the 723 Tier 1 and 2 variants in the COSMIC gene census. This is now explicitly stated in the manuscript. While it may be reasonable to restrict the list to only those genes with known CNAs, we considered it best to use the broader list and trust the user to make the appropriate assessment. This avoids potentially missing situations where an oncogene that is typically activated by point mutation is activated by amplification under rare/unusual circumstances.
2. This is an excellent suggestion. We evaluated a number of CNA databases/datasets but implementing a straightforward and automated method for importing the data to annotate inferred amplicons in our reports was not feasible.

Use cases:
The point about drop-outs leading to inference of two adjacent amplicons where, in fact, there is just one true amplicon which contains an assayed gene not reported as amplified is an important one. As we now report the boundary genes (See point 1), we have implemented an additional check so that if a report contains two inferred amplicons which share a single boundary gene, this is explicitly flagged so that the user will be asked to consider the possibility that the “boundary gene” may be erroneously reported as non-amplified.

Competing Interests: No competing interests were disclosed.

Comments on this article

Version 1

Author Response 12 Jun 2019

Paraic Kenny, University of Wisconsin-Madison, Madison, USA

Soon after publishing this article, we realized that the "inferCNV" name was already in prior use for an unrelated bioinformatics tool. To avoid confusion, the tool described in this article has been renamed "InferAMP" and can be accessed at http://inferamp.org. The name of the tool will be updated in the next version of the article.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com