Decision letter

...

Author response

...

Cell line contamination is a long-standing and persistent problem in academic, clinical and commercial research. The available solutions for cell line authentication are clearly not sufficient, since contamination is still a major economic burden on society despite having been recognized nearly 50 years ago. Issues with cell line contamination and authentication are very well described by Almeida et al 2016 PLOS Biology. We would like to explain a couple of points raised:

1. Contamination and rapid identification: cell line authentication benefits from being able to identify sample contamination in the earliest stage possible. The current US national guidelines by the ASN0002 state that a cell line is pure if the STRs have a match of 80% to one of the entries in the database. Only when the STR match < 56% the cell line is considered contaminated. Our method is better and more applicable than existing methods in two crucial aspects:
   1. Cell line authentication is not a static one-off test: Transfer of a couple of cells from one culture to the other by human error happens. Only the cells that rapidly proliferate compared to the original cell line will jeopardize the research done over time. Therefore the *periodic testing* of all cell lines in the lab is crucial, since only periodic testing will reveal that a cell line is contaminated. The *ATCC takes two weeks* to return test results, which means that by the time a researcher gets his/her results back, the contaminant might have overgrown the culture. Our MinION sketching method provides a rapid, and local solution for periodic cell line testing to allow researchers to by fully aware of the state of their cell line(s) at the time of an experiment. Even with a contamination of 10% our method rarely provides a robust match, indicative of contaminants in the cell line. Plus, we can robustly pick up a contamination of 20%, which is an important two-fold improvement over the ASN0002 standard that only rejects a pure cell line when < 56% matches.
   2. Sample mix-up: swapping cell lines does happen in laboratories from time to time and this would be a 100% contamination event. The MinION sketch would provide a perfect, instant sanity check prior to starting an experiment for any researcher.
2. Contamination analysis using SNPs versus STRs: cancer cell lines are typically very unstable and proliferate rapidly. STRs are more likely to incur mutations than SNPs and therefore, even a pure cancer cell line typically will result in a non-homogenous mix of repeat lengths. Hence, the ASN0002 guidelines place a cut off at an 80% match for pure cancer cell lines. Our SNP-based MinION sketch therefore enables a higher accuracy for the detection of contaminations in cancer cell lines.
3. The reference database availability: the cancer cell line encyclopedia generated by the Broad Institute provides around a 1000 reference files for the general public to use. We have converted this database to the correct format for our pipeline and provide this with our manuscript, so that our method can easily be adopted by members of the research community. This is only the start, and the database will be updated as the number of cell lines available will increase. When our method is used for forensic purposes ethical issues need to be considered – however, the appropriate (government) bodies that regulate forensic research and investigations should be able to formulate regulations that allow for the use of our method alongside relevant reference databases (see below) in forensics.
4. Cost: The costs of sequencing are continuing to plummet and this trend will most probably extend into the near future. When samples are multiplexed using 12 barcodes the costs of our MinION sketching method is currently already comparable to, and even lower than, the available cell line authentication method via the ATCC. The difference between our MinION sketching method versus the ATCC method is that researchers do not have to wait for two weeks to obtain the test results.
5. DNA concentration: for cell line authentication we do not have to worry as much about the availability of the input DNA as the forensic field has to. Even without pushing the boundaries of the possible, we were able to identify an individual using 50ng DNA. Our method can be used in addition with a PCR amplification step of informative SNPs, which could be an expansion of the method in the situation of an extremely low input DNA sample. We do concede that this would add a time consuming step in the protocol, however.
6. Expansion to other fields: Our method might very well be suitable for other fields, such as mice cell line authentication, and for testing the purity of *C. elegans*, Drosophila, plant and even yeast stocks. This paper can be the first step to make purity testing more convenient for researchers from disciplines across the life sciences – this is the reason why we think this paper would be perfect for eLife with its broad readership.

...

Why is a PCR-free approach an advantage? Our reasoning is the following:

* First, having to perform PCR adds yet another step in the protocol that takes time.
* Second, the current protocols mostly use human-specific PCR primers and therefore contaminants from other species cannot be detected [Alston-Roberts, Nature reviews 2010].
* Third, cancer cell lines are genetically unstable, and LOH or microsatellite instability result in a reduction in matching precision. Whether one chooses to amplify the DNA with a whole genome amplification approach or to use no PCR at all, either approach can be part of a robust re-identification method since there is no need to rely on a limited number of specific bi-allelic sites. The reliance on a small set of markers makes a reidentification method sensitive to allelic dropouts.

...

Second, for the forensics field:

* Multiple studies have shown that the statistical power of 48-52 SNPs is comparable to the current 10-15 STRs used (for example: Sanchez et al., 2010 and Yu et al., 2015. Moreover, Lin et al., 2004 showed in a theoretical framework that ~80 random SNPs are sufficient to discriminate individuals from each other. Our study shows that we indeed re-identify DNA using only 60-300 SNP markers. The reason for our range is A) we randomly sample from the genome and encounter variants that are more informative then others due to their allele frequency. This is immediately the advantage, as we don’t rely on specific sites in the genome, and we therefore do not suffer from allelic dropouts. B) We have to work with a relative high error-rate. Even so, we show in our manuscript that we robustly identify individuals without exceptions.
* We acknowledge some major hurdles still need to be taken before this method can be implemented in the field. However, doesn’t the reviewer agree there is space to improve our forensics methodologies in speed and efficiency? This manuscript suggests a solid way to do it, using cutting edge technology, and proposes yet another step forward to make this happen.

...

Furthermore, our work can certainly be reproduced:

* The genomes of YE001, JP001 and the HapMap samples (NA12890) are available online (https://dna.land/consent). To increase the database size, OpenSNP.org reference files can be downloaded. The DNA samples of NA12890 can be purchased – and DNA from YE001, SZ001 and JP001 can be requested, if desired, upon reasonable request.
* For cell line authentication we provide the CCLE database (1,099 reference files) in the correct format, ready to use. Our method can be tested, and our data can be reproduced using that database and DNA from any cell line in the CCLE database.