No evidence of MET and HER2 over-expression in non-small cell lung carcinoma and breast cancer, respectively, raises serious doubts on using RNA-seq profiles of tumor-educated platelets as a 'liquid biopsy' source

Sandeep Chakraborty,
R - 44/ 1, Celia Engineers, T. T. C Industrial Area, Rabale, Navi Mumbai, 400701, India.

Abstract

The prevailing excitement in the scientific and medical community about 'liquid biopsy', a minimally invasive diagnostic involving body fluid, is understandable, since it has the possibility of detecting pre-malignant and early-stage cancers, and enables the assessment of response to treatments [1]. Supplementing previous techniques that sample circulating cell-free tumor DNA, tumor cells, and microvesicles [2], a recent work has shown using RNA-seq data that tumor-educated platelets (TEP) can distinguish 228 patients with localized and metastasized tumors from 55 healthy individuals with 96% accuracy [3]. However, as demonstrated in the current work, over-expression of MET genes in non-small cell lung carcinoma (NSCLC), and HER2/ERBB2 genes in breast cancer are grossly misreported. Based on an analysis of a smaller subset of samples, it is shown that there is little, leave alone over-expression, of these genes in the samples with the specified disease. Confirmation that this is bona-fide platelet mRNA is provided by high levels of the platelet marker TMSB4X. A kmer-based method (k=32) has been used here (KEATS) to detect homologous transcripts, although the results are easily verified by a BLAST search - BLAST'ing the MET gene (Accid: NM_001127500.2) to a NSCLC sample (https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR1982781) shows almost no expression. This is in contrast to expected expression of the MET gene in another a NSCLC sample (SRR3475320) from another sample. Similar contradictions apply for HER2/ERBB2 genes with respect to breast cancer samples. This work emphasizes the neccessity of a more stringent verification framework for bioinformatic analyses, and raises serious doubts on using TEP as a possible 'liquid biopsy' candidate.

Introduction

Tumor tissue biopsy, the gold standard for cancer diagnostics, pose challenges that include access to the tumor, quantity and quality of tumoral material, lack of patient compliance, repeatability, and bias of sampling a specfic area of a single tumor [4]. This has resulted in a new medical and scientific paradigm defined by minimal invasiveness, high-efficiency, low-cost diagnostics [5], and, whenever possible, personalized treatment based on genetic and epigenetic composition [6]. The presence of fragmented DNA in the cell-free component of whole blood (cfDNA) [7], first reported in 1948 by Mandel and Metais, has been extensively researched for decades, with extremely promising results in certain niches [8]. Additionally, cfDNA derived from tumors (ctDNA) [9] have tremendous significance as a cancer diagnostic tool [10], and for monitoring responses to treatment [11]. However, detection of ctDNA, and differentiation with cfDNA, remains a challenge due the low amounts of ctDNA compared to cfDNA [1].

Recently, tumor-educated blood platelets (TEP) were proposed as an alternative source of tumor-related biological information [3, 12]. The hypothesis driving the potential diagnostic role of TEPs is based on the interaction between blood platelets and tumor cells, subsequently altering the RNA profile of platelets [13,14]. The study showed using RNA-seq data that tumor-educated platelets (TEP) can distinguish 228 patients with localized and metastasized tumors from 55 healthy individuals with 96% accuracy [3]. As validation, this study reported significant over-expression of MET genes in non-small cell lung carcinoma (NSCLC), and HER2/ERBB2 [15] genes in breast cancer, which are well-established biomarkers. Here, analysis of a subset of the samples refutes the claims of over-expression of MET and HER2/ERBB2 genes, in TEP. The rational expression profile of MET genes in another NSCLC sample (SRA:SRR3475320) from a different study, and the high counts of the platelet marker TMSB4X in the current study (also mentioned in the TEP study) further validates the discrepancy noted here. These results are easily verified by a BLAST search.

Contradictory results:

A smaller subset of lung cancer samples (DATASET:list.lung.txt,n=24) was used from the given 60 NSCLC samples in the TEP study [3]. Table 1 shows very little expression of the MET gene (Accid: NM_001127500.2) in these samples. In contrast, the RNA-seq sample from a NSCLC female patient (SRA:SRR3475320) from a different study shows reasonable expression (Table 1). A BLAST search also shows (graphically) a rational expression pattern (Fig 2). In fact, there is very little MET expression in any sample (cancer of any kind or healthy) in a random subset of 119 samples (DATASET:list.119.txt). Another verification is provided by the high level of expression of the TMSB4X gene (Fig 3), which has also been detected and mentioned in the TEP study as a platelet marker [3]. Similar contradictions apply for HER2 genes with respect to breast cancer samples (Fig 4).

Conclusion:

This raises serious doubts on using TEP as a possible 'liquid biopsy' candidate. Essentially, it refutes the hypothesis that platelets carry enough RNA-seq from tumors to make it viable as a diagnostic method. A review found it 'surprising' that although 'the tumor type was the predominant factor for the actual platelet conditioning, tumor metastasis did not significantly impact on them when compared to samples from patients without metastasis' [14]. The excitement surrounding the fact that '2016 marked the first approval of a liquid biopsy test in oncology to assist in patient selection for treatment' [16] should be tempered, and a cautious approach adopted [17, 18] with reports of 'broken promises' [19].

Finally, the current study highlights the neccessity of a more stringent verification framework for bioinformatic analyses in large scale studies [20, 21]. The black-box surrounding bioinformatic analyses should be made more transparent.

Materials and methods

A kmer-based version (KEATS [22]) of YeATS [20,23–26] was used to obtain gene counts from transcripts in the RNA-seq data. A BLAST search suffices to demonstrate the absence of MET genes in the lung cancer samples.

Competing interests

No competing interests were disclosed.

References

1. Diaz LA, Bardelli A (2014) Liquid biopsies: genotyping circulating tumor dna. Journal of Clinical Oncology 32: 579–586.
2. Quandt D, Zucht HD, Amann A, Wulf-Goldenberg A, Borrebaeck C, et al. (2017) Implementing liquid biopsies into clinical decision making for cancer immunotherapy. Oncotarget .
3. Best MG, Sol N, Kooi I, Tannous J, Westerman BA, et al. (2015) Rna-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. Cancer cell 28: 666–676.
4. Vendrell JA, Mau-Them FT, Béganton B, Godreuil S, Coopman P, et al. (2017) Circulating cell free tumor dna detection as a routine tool forlung cancer patient management. International Journal of Molecular Sciences 18: 264.
5. Han X, Wang J, Sun Y (2017) Circulating tumor dna as biomarkers for cancer detection. Genomics, proteomics & bioinformatics .
6. Sorber L, Zwaenepoel K, Deschoolmeester V, Van Schil P, Van Meerbeeck J, et al. (2016) Circulating cell-free nucleic acids and platelets as a liquid biopsy in the provision of personalized therapy for lung cancer patients. Lung Cancer .
7. Jiang P, Lo YD (2016) The long and short of circulating cell-free dna and the ins and outs of molecular diagnostics. Trends in Genetics 32: 360–371.
8. Lo YD, Corbetta N, Chamberlain PF, Rai V, Sargent IL, et al. (1997) Presence of fetal dna in maternal plasma and serum. The Lancet 350: 485–487.
9. Chen XQ, Stroun M, Magnenat JL, Nicod LP, Kurt AM, et al. (1996) Microsatellite alterations in plasma dna of small cell lung cancer patients. Nature medicine 2: 1033–1035.
10. Yi X, Ma J, Guan Y, Chen R, Yang L, et al. (2017) The feasibility of using mutation detection in ctdna to assess tumor dynamics. International Journal of Cancer 140: 2642–2647.
11. Imamura F, Uchida J, Kukita Y, Kumagai T, Nishino K, et al. (2016) Monitoring of treatment responses and clonal evolution of tumor cells by circulating tumor dna of heterogeneous mutant egfr genes in lung cancer. Lung Cancer 94: 68–73.
12. Nilsson RJA, Balaj L, Hulleman E, Van Rijn S, Pegtel DM, et al. (2011) Blood platelets contain tumor-derived rna biomarkers. Blood 118: 3680–3683.
13. Bardelli A, Pantel K (2017) Liquid biopsies, what we do not know (yet). Cancer cell 31: 172–179.

14. Feller SM, Lewitzky M (2016) Hunting for the ultimate liquid cancer biopsy-let the tep dance begin. Cell Communication and Signaling 14: 24.

15. Foulkes WD, Stefansson IM, Chappuis PO, Bégin LR, Goffin JR, et al. (2003) Germline brca1 mutations and a basal epithelial phenotype in breast cancer. Journal of the National Cancer Institute 95: 1482–1485.

16. Blumenthal GM, Pazdur R (2017) Approvals in 2016: the march of the checkpoint inhibitors. Nature Reviews Clinical Oncology 14: 131–132.

17. Diamandis EP (2016) A word of caution on new and revolutionary diagnostic tests. Cancer cell 29: 141–142.

18. Best MG, Sol N, Tannous BA, Wesseling P, Wurdinger T (2016) Re: a word of caution on new and revolutionary diagnostic tests. Cancer cell 29: 143.

19. Shee K, Chamberlin M, Varn F, Bean J, Marotti J, et al. (2017). Abstract p6-07-03: Broken promise of liquid biopsy: Plasma dna does not accurately reflect tumor dna in metastatic breast cancer.

20. Chakraborty S, Martínez-García PJ, Dandekar AM (2016) Yeatsam analysis of the walnut and chickpea transcriptome reveals key genes undetected by current annotation tools. F1000Research 5.

21. Chakraborty S (2016) Rna-seq assembler artifacts can bias expression counts and differential expression analysis - case study on the chickpea transcriptome emphasizes importance of freely accessible data for reproducibility [version 2; referees: 2 not approved]. F1000Research 5.

22. Chakraborty S (2017) Cataloguing over-expressed genes in epstein barr virus immortalized lymphoblastoid cell lines through consensus analysis of pacbio transcriptomes corroborates hypomethylation of chromosome 1. bioRxiv : 125823.

23. Chakraborty S, Britton M, Wegrzyn J, Butterfield T, Martinez-Garcia PJ, et al. (2015). YeATS-a tool suite for analyzing RNA-seq derived transcriptome identifies a highly transcribed putative extensin in heartwood/sapwood transition zone in black walnut.

24. Martínez-García PJ, Crepeau MW, Puiu D, Gonzalez-Ibeas D, Whalen J, et al. (2016) The walnut (juglans regia) genome sequence reveals diversity in genes coding for the biosynthesis of nonstructural polyphenols. The Plant Journal .

25. Chakraborty S, Britton M, Martínez-García P, Dandekar AM (2016) Deep RNA-seq profile reveals biodiversity, plant–microbe interactions and a large family of NBS-LRR resistance genes in walnut (*juglans regia*) tissues. AMB Express 6: 1.

26. Chakraborty S (2017) Mcf-7 breast cancer cell line pacbio generated transcriptome has˜ 300 novel transcribed regions, un-annotated in both refseq and gencode, and absent in the liver, heart and brain transcriptomes. bioRxiv : 100974.

Table 1: **Count of MET reads identified using KEATS:** The TEP study had 60 samples - a subset (n=24) was used here. This was compared to the RNA-seq obtained from a NSCLC tissue (SRA:SRR3475320) in a different study (Bioproject:PRJNA320473).

	SRA id	KEATS counts
TEP study [3]	SRR1982781	0
	SRR2096502	1
	SRR2096503	2
	SRR1982787	2
	SRR1982760	2
	SRR1982782	4
	SRR2096517	5
	SRR1982756	5
	SRR1982761	5
	SRR1982792	5
	SRR1982793	6
	SRR2096516	8
	SRR1982795	9
	SRR1982770	10
	SRR1982759	10
	SRR1982772	12
	SRR1982762	13
	SRR1982771	13
	SRR1982780	15
	SRR1982765	16
	SRR1982791	24
	SRR1982777	42
	SRR2096501	45
	SRR1982790	46
Lung tissue	SRR3475320	4972

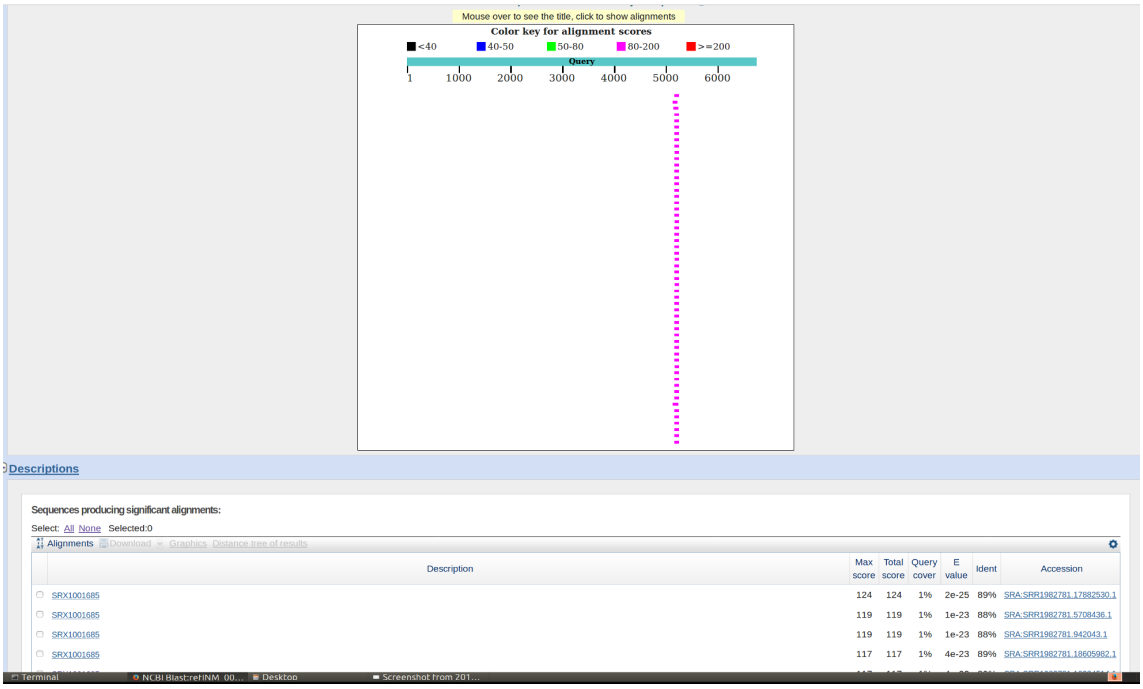

Figure 1: **Example lung cancer sample from the TEP study [3] (SRA:SRR1982781) - with no MET expression (Accid: NM_001127500.2):** The best matching transcript (SRR1982781.17882530) has low % identity match, and matches to a different loci in another chromosome with 100% identity.

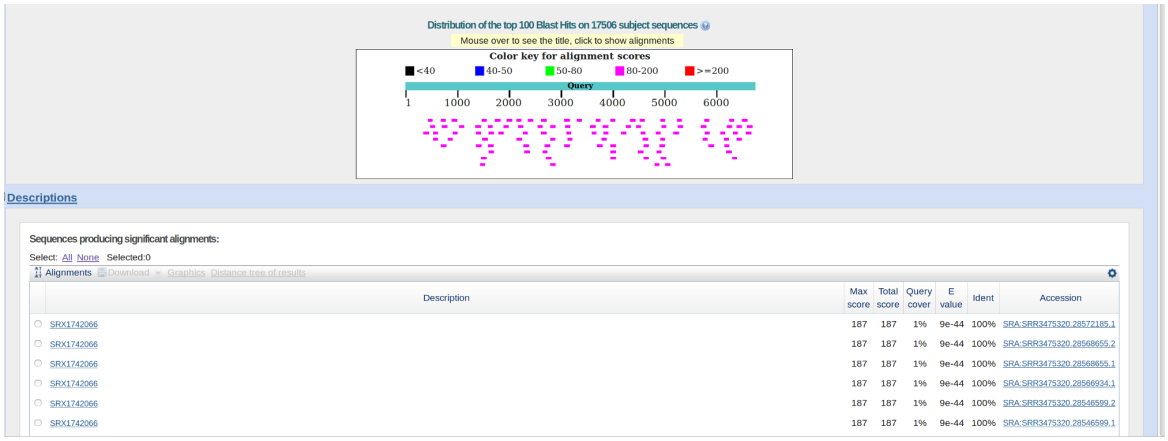

Figure 2: **BLAST results for a RNA-seq sample from a NSCLC patient (SRA:SRR3475320) from a different study shows reasonable expression**: The RNA-seq was based on tissue acquisition.

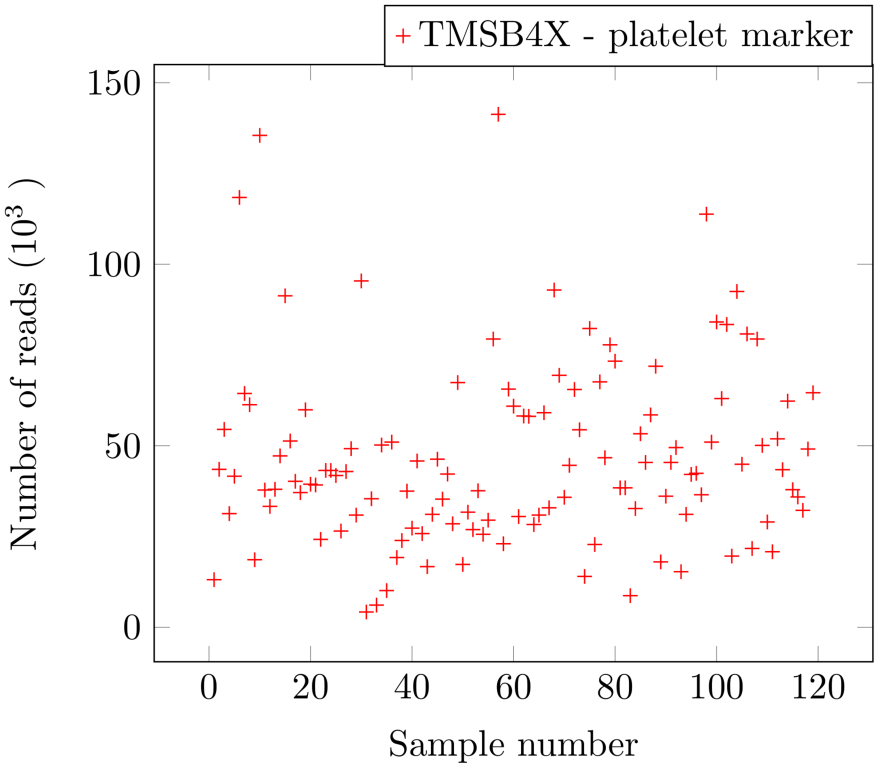

Figure 3: **Read counts of the platelet marker TMSB4X in a subset (n=119) of the samples used in the TEP study [3]:**

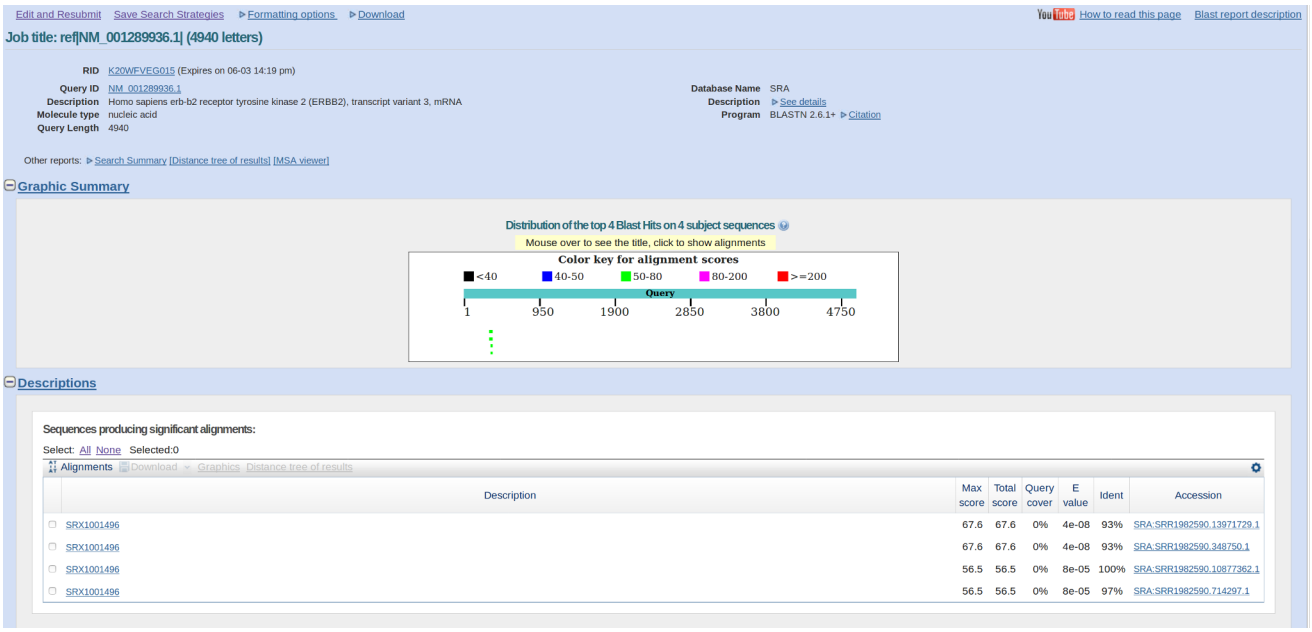

Figure 4: **No HER2/ERBB2 expression in a breast cancer sample from the TEP study:** HER2/ERBB2 sequence was obtained from Accid:NM_001289936.1.