Clonal haematopoiesis: a source of biological noise in cell-free DNA analyses

There is considerable interest in the development of a blood-based biomarker for early-cancer detection driven by an important clinical need; to detect cancers earlier when cure is feasible. A low false-positive rate is a vital component of any diagnostic assay being leveraged in an early-detection setting [1]. Cell-free circulating tumour DNA (ctDNA) is an analyte within blood that holds potential as a highly specific indicator of cancer. For example, Phallen et al. [2] demonstrated that the targeted error correction sequencing (TEC-seq) next-generation sequencing (NGS) assay could detect ctDNA in 62% of patients with diagnosed stage I–II breast, lung, ovarian or colorectal cancer. Using rigorous variant calling parameters, they maintained a specificity for ctDNA detection in excess of 99.9999% [2]. Other advances in this area include the CancerSEEK assay that incorporates ctDNA detection alongside protein biomarker detection [3] and the development of targeted gene panel, methylation and whole-genome assays for early-cancer detection by GRAIL [4].

One challenge to maintaining the specificity of ctDNA detection is differentiating a cancer-signal from background normal biological variation within an individual. The majority of cell-free DNA (over 80% in healthy individuals) arises from haematopoietic cells [5–7]. Normal haematopoietic cells accumulate somatic mutations during ageing which can drive clonal expansions of haematopoietic cells in the absence of dysplasia. These mutations are referred to as clonal haematopoietic mutations of indeterminate potential (CHIP) [8]. CHIP presents a biological confounding factor for early cancer detection assays predicated on characterisation of cell-free DNA as tumour DNA based on somatic variant detection [9].

Within this article Liu et al. [10] further defined the prevalence of somatic alterations present in the cell-free DNA from individuals without a diagnosis of cancer. They enriched cell-free DNA from plasma in a cohort of 259 healthy individuals using 1 of the 2 capture panels covering hotspot regions from up to 508 cancer-related genes. They leveraged an *in silico* analysis method to reduce background artefactual error in their sequencing data (errors which typically occur due to DNA damage in library preparation or incorrect base calling by the sequencing platform). To reduce these errors, Liu et al. utilised an endogenous duplex barcoding approach and achieved a background error-rate across their panel of 2×10^{7} errors per base. This error-rate is ~50-fold lower than that achieved by digital error-suppression and single strand molecular barcoding reported by Newman et al. [11] using the CAPP-Seq assay (1.5×10^{-5} errors per base). The high degree of specificity achieved with the endogenous duplex barcoding approach meant the authors could be confident regarding variant calls made using the assay. However, a disadvantage of this analytical approach is the impact requiring reads from both template DNA molecule strands to form duplex consensus read has on library complexity. For example, in this study only 6% of original DNA templates input into a library generated duplex consensus reads, limiting the sensitivity of the assay due to allele drop-out. In contrast, Phallen et al. [2] achieved a conversion efficiency of 40% using the TEC-Seq platform which utilised single-stranded exogenous barcodes. The reduction in library complexity observed with Liu et al.'s method led to a reduction in sensitivity for low-frequency variant detection, with 80% (39/49) of 0.5% frequency variants detected and only 35% (80/226) of 0.25% frequency variants detected in their validation data. This suggests that the assay would underestimate the prevalence of CHIP mutations occurring at low variant allele frequencies (<1%). This is relevant since Swanton et al. [12] demonstrated that CHIP variants can occur at variant allele frequencies <0.1%.

Liu et al. found that 60% of healthy participant cfDNA samples harboured at least one non-synonymous mutation or indel. The frequency of alterations detected increased by age supporting previous findings from Xie et al. and GRAIL [12, 13]. A total of 329 mutations across 164 samples were identified, spanning 166 genes. The most common mutations were found in genes previously associated with CHIP, particularly *DNMT3A* which was mutated in 52 independent samples. Notably only one mutation was identified in *TP53* in healthy participant cfDNA, whereas previous studies suggest that *TP53* may be more commonly mutated in CHIP [9, 13]. Possibly explanations for this discrepancy could be a limited sensitivity of Liu et al.'s assay for low-frequency *TP53* mutations or variation in age distributions or ethnicity between cohorts. These data suggest that filtering cell-free DNA analyses for alterations commonly associated with CHIP could reduce the risk of false-positives in ctDNA analyses. One hundred and twenty-five of the 329 detected mutations were indexed in the COSMIC database, yet no oncogene activating mutations were identified in this cohort. This suggests that detection of oncogene activating mutations in plasma could be specific for solid malignancies. However, Hu et al. [9] reported the detection in cell-free DNA of activating *KRAS* codon 12 mutations that localised to a peripheral blood cell population. Therefore, specificity of an oncogenic alteration detected in cell-free DNA for solid malignancies may be gene-dependent.

Liu et al. utilised non-error corrected NGS of peripheral blood cell DNA enriched with the same panels as applied to cell-free DNA for all individuals in the study to act as a comparison for cell-free DNA sequencing data. Based on these comparisons, the authors demonstrated that the variant allele frequencies of mutations in cell-free DNA and blood DNA were highly correlated ($R = 0.87$). This is consistent with observations that haematopoietic DNA makes up most of the cell-free DNA compartment [5] and reinforces the need for peripheral blood sequencing to occur to the same depth as cell-free DNA sequencing if a calling filter to remove CHIP somatic variants (a CHIP-filter) is to be applied. Expanding on this observation the authors estimated that a CHIP-filter based on the identification of a single variant read in conventional NGS data would require an original sequencing depth of 2996× to identify variants at 0.1% frequency with 95% sensitivity. This highlights a limitation of using peripheral blood exome data (typically utilised to identify and remove germline variants from tumour sequencing data) as a CHIP-filter for low-frequency cell-free DNA variant detection. In TRACERx, we orthogonally validated our patient-specific PCR-enrichment

approach with a generic error-controlled hotspot PCR-enrichment panel applied to pre-operative cell-free DNA from 28 NSCLCs [14]. We identified 13 variants not present in multi-region tumour exome data, present in cell-free DNA (variant allele frequencies ranging from 4.44% to 0.05%). These somatic variants were not detectable in germline exome data [14]. As Liu et al. highlight this does not rule out CHIP given that the raw sequencing depth achieved over these variants with the PCR-enrichment panel (65 449×) was more than that achieved with TRACERx germline exome capture (415×). Consequently, the germline exome data would have insufficient sensitivity for detection of low-frequency CHIP variants. Liu et al. also highlight that a one-read CHIP-filter based on conventional NGS of peripheral blood cell DNA would have a high false positive rate, particularly at variant allele frequencies of <0.1% and when evaluating base changes with high-background artefactual noise (e.g. G > T). The authors suggest that an optimal CHIP filter with a 95% sensitivity and specificity for CHIP variant detection of <0.1% should incorporate error-control strategies to reduce the risk of false positive CHIP variant calls. Reflecting this requirement, to maintain a specificity for ctDNA detection in excess of 99%, GRAIL sequence peripheral white blood cell DNA and cell-free DNA to the same unique coverage (60 000× original, 3000× unique) using their 507-gene panel targeted enrichment assay to remove the confounding effect of CHIP on their data [11].

In conclusion, Liu et al. provide an interesting study focussing on establishing the somatic variant profile present in cell-free DNA from healthy participants. They provide insights into strategies required to filter out CHIP associated somatic variants from cell-free DNA analyses. These strategies include variant filters based on association of a mutation with CHIP, functional annotation of a somatic variant as an oncogene activating event and deep error-controlled sequencing of peripheral blood DNA. These steps will be important to maintain the specificity of somatic variant detection as an indicator of a cancer-signal in early-detection strategies based on ctDNA.

C. Abbosh[1], C. Swanton[1,2] & N. J. Birkbak[2,3*]
[1]Cancer Research UK Lung Cancer Centre of Excellence London and Manchester, University College London Cancer Institute, London; [2]Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK; [3]Department of Molecular Medicine, Aarhus University, Aarhus, Denmark
(*E-mail: nbirkbak@clin.au.dk)

Acknowledgements

NJB is a fellow of the Lundbeck Foundation and acknowledges funding from Aarhus University Research Foundation (AUFF). CS is supported by The Francis Crick Institute (FC001169), MRC (FC001169), the Wellcome Trust (FC001169), Cancer Research UK (TRACERx and CRUK Cancer Immunotherapy Catalyst Network), the CRUK Lung Cancer Centre of Excellence, Stand Up 2 Cancer (SU2C), the Rosetrees and Stoneygate Trusts, NovoNordisk Foundation (16584), the NIHR BRC at University College London Hospitals, and the CRUK-UCL Centre, Experimental Cancer Medicine Centre, and in part by the Breast Cancer Research Foundation.

Funding

None declared.

Disclosure

CA and CS submitted a patent with University College London (UCL) business PLC (provisional patent number 1618485.5) based on a phylogenetic approach to analysis of circulating tumour DNA. CS has received grant/research support from Pfizer, AstraZeneca, BMS and Ventana; personal fees from Boehringer Ingelheim, Eli Lily, Servier, Novartis, Roche-Genentech, GlaxoSmithKline, Pfizer, BMS, Celgene, AstraZeneca, Illumina, Sarah Cannon Research Institute; has stock options in Achilles Therapeutics, ApoGen Biotechnologies, EPIC Bioscience, and GRAIL; and is a co-founder of Achilles Therapeutics. NJB declares no competing interests.

References

1. Aravanis AM, Lee M, Klausner RD. Next-generation sequencing of circulating tumor DNA for early cancer detection. Cell 2017; 168(4): 571–574.
2. Phallen J, Sausen M, Adleff V et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med 2017; 9(403). doi: 10.1126/scitranslmed.aan2415.
3. Cohen JD, Li L, Wang Y et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 2018; 359(6378): 926–930.
4. Klein EA, Hubbell E, Maddala T et al. Development of a comprehensive cell-free DNA (cfDNA) assay for early detection of multiple tumor types: the Circulating Cell-free Genome Atlas (CCGA) study. J Clin Oncol 2018; 36(Suppl 15): 12021–12021.
5. Moss J, Magenheim J, Neiman D et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 2018; 9(1): 5068.
6. Lui YY, Chik KW, Chiu RW et al. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. Clin Chem 2002; 48: 421–427.
7. Snyder MW, Kircher M, Hill AJ et al. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell 2016; 164(1–2): 57–68.
8. Steensma DP, Bejar R, Jaiswal S et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood 2015; 126(1): 9–16.
9. Hu Y, Ulrich BC, Supplee J et al. False-positive plasma genotyping due to clonal hematopoiesis. Clin Cancer Res 2018; 24(18): 4437–4443.
10. Liu J, Chen X, Wang J et al. Biological background of the genomic variations of cf-DNA in healthy individuals. Ann Oncol 2019; 30(3): 464–470.
11. Newman AM, Lovejoy AF, Klass DM et al. Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol 2016; 34(5): 547–555.
12. Swanton C, Venn O, Aravanis A et al. Prevalence of clonal hematopoiesis of indeterminate potential (CHIP) measured by an ultra-sensitive sequencing assay: exploratory analysis of the Circulating Cancer Genome Atlas (CCGA) study. J Clin Oncol 2018; 36(Suppl 15): 12003.
13. Xie M, Lu C, Wang J et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med 2014; 20(12): 1472–1478.
14. Abbosh C, Birkbak NJ, Wilson GA et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 2017; 545(7655): 446–451.

doi:10.1093/annonc/mdy552
Published online 15 January 2019