Shallow Whole Genome Sequencing (sWGS) prior to deep-WGS assays to estimate tumor purity

Scientific Abstract

Background and Rationale: Tumor purity (TP), defined as the percentage of cancer cells within a tissue sample, is a critical parameter for molecular assays, as it can influence the outcomes of genomic analyses commonly used to identify gene alterations and chromosomal abnormalities. Traditionally, TP has been estimated by a pathologist using hematoxylin-eosin staining histology, which is subject to interobserver variability. Genomic approaches, such as shallow whole-genome sequencing (sWGS), offer an alternative for TP estimation by leveraging tumor-specific features, including somatic copy number alterations (CNAs).

Objectives(s): This study aims to compare the accuracy of TP estimation between a sWGS-based method and histological analysis as quality control steps prior to whole-genome sequencing (WGS) in a pan-cancer cohort. Additionally, it seeks to identify tumor and sample characteristics that may influence TP estimations by both methods.

Methods and Results: We assessed a pan-cancer cohort (from Marathon of Hope Cancer Centres Network – Ontario Cancer Consortium)) in which all samples were submitted to deep-WGS (full depth of 80X) in an institutional workflow, enabling the evaluation of samples TP through histology, sWGS and deep-WGS. The samples went through a two-step TP quality control: H&E slide histology, conducted up to 2 pathologists, followed by sWGS (full depth of 0.1X) analyzed with ichorCNA (sWGS+ichorCNA) algorithm. The threshold for adequacy was TP ≥40% from histology and ≥25% from sWGS+ichorCNA. Using TP inferred by deep-WGS as the ground truth, we compared the results for TP estimation from these two methods. A total of 507 tumors were included, across 20 distinct tumor types categorized as carcinoma (85.6%), sarcoma (2.6%), and hematologic neoplasms (9.7%). The histological subtype ($p=7.51x10^{-5}$) and the preservation method of biopsy specimens (p=0.002, median TP: fresh-frozen samples=55.7%, FFPE samples=32.4%) influenced the TP outputs from sWGS+ichorCNA. TP estimation by sWGS+ichorCNA demonstrated concordance with that determined by deep-WGS (R=0.70, $p<2.2 x10^{-16}$). According to histology estimate, 78.9% of the samples had sufficient TP compared to 62.5% for sWGS+ichorCNA, demonstrating discordance between the methods. False positive samples under sWGS+ichorCNA had higher ploidy estimates than true negatives, reflecting the inability of sWGS methods to account for hyperploidy (p=0.00014). This discordance was not seen in histologic assessment (p=0.32). Further, specificity in high-ploidy cases (deep-WGS inferred ploidy ≥ 4) was lower compared to low-ploidy cases (deep-WGS inferred ploidy < 4) by 35.2%. Notably, the overall specificity of sWGS+ichorCNA was superior to pathology (sWGS+ichorCNA: 73.5% pathology: 25.9%) without compromising sensitivity (shWGS+ichorCNA: 82.1%, pathology: 81.5%)%)

Conclusion(s): In this pan-cancer analysis, sWGS+ichorCNA outperformed histology in TP estimation across multiple metrics, although notable discordance between the methods was observed. Tumor characteristics, including lineage, histological subtype, sample preservation conditions, and ploidy, significantly impacted TP estimations using CNA-based genomic assays. These factors should be carefully considered when employing shWGS+ichorCNA for TP determination.

Anticipated Impact: This study demonstrates the utility of sWGS-based estimate TP, which may improve the efficiency of sample selection in WGS workflows.

Plain language abstract

Background and rationale: When sequencing cancer tissue, it's important to know what percentage of the sample is actually cancer cells. This is called tumor purity. Traditionally, doctors look at the tissue under a microscope to estimate this before sequencing all of the genes. However, this method can vary between doctors. We looked at a newer, gene-based method using a type of DNA sequencing called shallow whole-genome sequencing (sWGS) to estimate tumor purity and compared it to the traditional method.

Objectives: This study compared the accuracy of TP estimation using sWGS with traditional microscopic evaluation in a large group of cancer samples from Marathon of Hope Cancer Centres Network – Ontario Cancer Consortium. This analysis also investigated how different factors, such as the type of cancer and how the samples were stored, might affect the TP estimates.

Methods and Results:
We examined 507 cancer samples from different types of cancer. We used three methods to estimate tumor purity: a doctor's visual assessment, sWGS, and a more detailed, expensive type of DNA sequencing (deep-WGS), which was used as the most accurate measure. We found that the sWGS method was more accurate than the doctor's visual assessment in identifying samples with low tumor purity. However, there were differences between the two methods. The type of cancer, how the sample was stored (frozen or preserved in formalin), and the number of chromosomes in the cancer cells affected the sWGS results.

Conclusion: The sWGS method is a useful tool for estimating tumor purity and may be more reliable than traditional visual assessment. However, it's important to consider factors like cancer type and sample storage as these can affect the accuracy of the sWGS results. Using sWGS can help researchers select the best samples for more detailed and expensive genetic testing, making the process more efficient.