Pioneering a Tumor-Specific Methylation Atlas (TSMA) to Identify Tissue of Origin (TOO) in Multi-Cancer Early Detection

Published 03 July 2024, BMC Journal of Translational Medicine. Click to access full article.



Minh Duy Phan, PhD.
University of Cambridge, U.K. Computer Science and Molecular Biology



Trong Hieu Nguyen, PhD.
RWTH Aachen University, Germany. Applied mathematics.

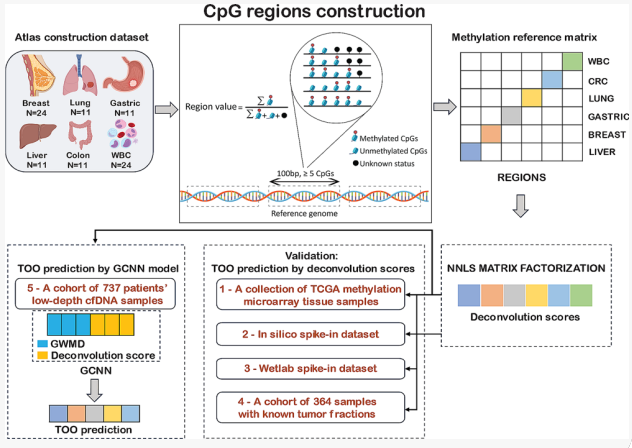
Background:
Cell free DNA (cfDNA)-based assays hold great potential in detecting early cancer signals. However, **predicting the tissue-of-origin** for multiple cancer types at an early stage remains challenging due to the low abundance of tumor-derived cfDNA and the confounding presence of various DNA components released from non-tumor sources.

Solutions:
1. DNA methylation has been proved an important epigenetic marker in cancer early detection context. Furthermore, these **DNA methylation patterns are tissue-specific** and remain stable during neoplastic transformation. An atlas of these tissue-specific methylation was constructed for five cancer types.
2. In this study, we investigated the application of **artificial intelligence and methylation atlas** to deconvolute cell types, with the goal of creating a model that can detect tumor origin in cfDNA samples for early cancer screening.

Utilizing artificial intelligence and tumor-specific methylation atlas (TSMA) to identify tissue-of-origin in ctDNA samples

ATLAS DISCOVERY

2,945 differential CpG regions between 5 tumor tissue types and WBC





TSMA TOO results

69% 1st TOO accuracy

achieved by integrating a tumor-specific methylation atlas (TSMA) with cfDNA features in a graph convolutional neural network **(an artificial intelligence technique).** The developed model combined deconvolution scores and genome-wide methylation density in a held-out validation dataset of 239 cfDNA samples.

Outcome:
Leveraging tumor-specific methylation atlas, we aim to optimize multi-cancer early detection (MCED) assays by reducing sequencing depth for ctDNA detection and enabling tumor origin prediction.

Pioneering a Tumor-Specific Methylation Atlas (TSMA) to Identify Tissue of Origin (TOO) in Multi-Cancer Early Detection

CONSTRUCTING TSMA

To construct the TSMA, the regions in which region values were significantly different among five cancer tissues and WBC were captured. Overall, a TSMA of 2,945 differential CpG regions between 5 tumor tissue types and WBC were constructed.

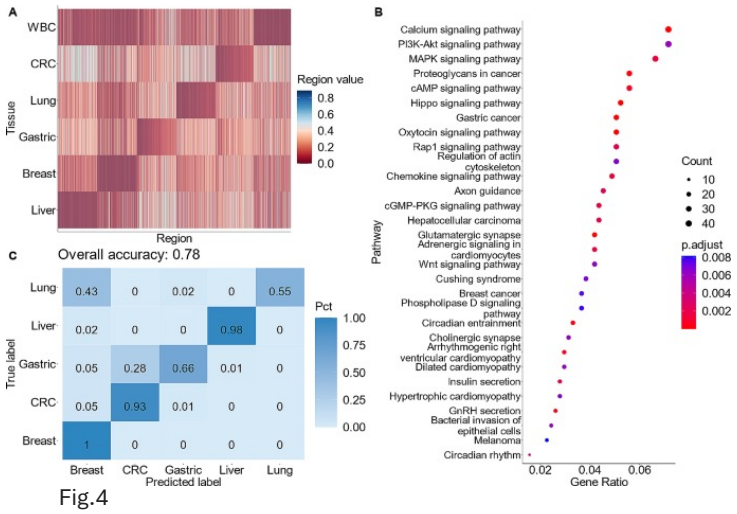


Fig.4

A) Heatmap of average region values in each cancer tissue type or WBC across 2,945 CpG regions included in the TSMA. (B) Pathway analysis reveals cancer-related pathways, which were enriched by the set of genes to which TSMA regions were mapped. (C) Prediction performance using highest deconvolution score to assign label to samples in Dataset 1 comprising of 888 colorectal samples, 1,814 lung samples, 398 gastric samples, 888 breast samples and 429 liver samples

> Specifically, we achieved accuracies of 100%, 98% and 93% for breast, liver and CRC cancer, respectively; while gastric and lung cancer exhibited lower accuracies of 66%, and 55%. These results validated our hypothesis and suggested that our TSMA has successfully captured cancer-specific signals that could be used to determine the TOO of a sample.
>
> Trong Hieu Nguyen, PhD. TMSA key researcher

Combining TSMA deconvolution scores (DS) with genome-wide methylation density (GWMD)

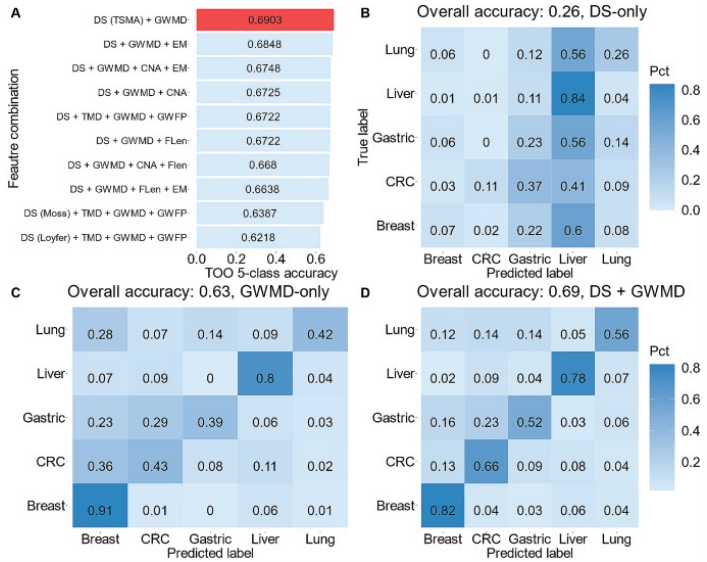


Fig.5

Here, we used the same dataset from K-DETEK assay to explore the possibility of combining deconvolution scores with other cfDNA features to enhance TOO performance. GWMD, expressed as the average methylation density across non-overlapping 1 M bins in the entire genome, when combined with deconvolution scores achieved the highest accuracy of 69% (Fig. 5A, D). This was markedly improved from deconvolution scores alone (26% accuracy, Fig. 5B), or GWMD alone (63% accuracy, Fig. 5C).

This result highlighted the integration of TSMA deconvolution scores and artificial intelligence model for TOO detection, especially when combined with GWMD.