# Unsupervised Histopathology Search Engine for Diagnosis of Primary and Metastatic Cancers (CANCERCH)

Ali Khajegili Mirabadi

## Background and rationale

Accurately determining where a cancer started (the primary site) and how it has spread (metastases) is essential for choosing the right treatment, especially for patients with rare or unusual cancers. Today, pathologists diagnose cancer by visually examining Hematoxylin & Eosin (H&E)–stained whole slide images (WSIs) and, when needed, ordering additional tests. This process can be time-consuming and challenging, and in some cases the primary site cannot be confidently identified by human review alone. Recent advances in artificial intelligence (AI) now allow us to automatically "encode" the visual patterns in WSIs, opening the door to search tools that can find similar cases across large image archives and support more objective, data-driven diagnosis.

## Objective(s)

Our objectives are to:

- Develop an unsupervised Histopathology Search Engine trained on large collections of WSIs.
- Use this system to help distinguish the origin of primary and metastatic cancers, including rare and diagnostically challenging cases.

## Methods and Results

We first train our AI models on WSIs from The Cancer Genome Atlas (TCGA), covering a broad range of tumour types. Each WSI is divided into small image patches, and deep learning is used to extract visual features that describe cell and tissue architecture. Using unsupervised learning, we organize these features into a shared "map" of histopathology, allowing the system to retrieve cases that are most similar to a query slide or region without manual labels for each cancer type. We then perform external validation on the Personalized OncoGenomics (POG) trial cohort from BC Cancer, which includes 26 different cancer types. In this setting, the model is asked to predict the primary site of the tumour using only the WSI-based features. Our preliminary results show that the method can predict the primary site directly from WSIs with **moderate accuracy**, indicating that even without additional tests, image-based search can provide useful diagnostic signals while leaving room for further improvement.

## Conclusion

An unsupervised Histopathology Search Engine trained on TCGA and externally validated on POG can turn routine WSIs into a navigable, case-based resource. By automatically finding similar cases and inferring the likely primary site from image data alone with moderate accuracy, this approach, as a proof-of-concept, provides objective support for challenging diagnoses and reveals patterns that may be difficult for human observers to systematically capture.

## Anticipated Impact

For cancer patients, this work could improve diagnostic confidence for rare and metastatic cancers by rapidly matching them to similar, well-characterized cases and suggesting candidate primary sites from WSIs alone. It may also enable more personalized treatment decisions by uncovering histological subtypes that are associated with response to specific therapies.