



האוניברסיטה
העברית
בירושלים
THE HEBREW
UNIVERSITY
OF JERUSALEM

Detecting Cancer by Blood Samples

Yuval Cantor¹, Josh Moss^{1,2}, Tommy Kaplan¹

¹The Rachel and Selim Benin School of Computer Science and Engineering, Hebrew University, Jerusalem, Israel

²Institute for Medical Research Israel-Canada, The Hebrew University-Hadassah Medical School, Jerusalem, Israel

1

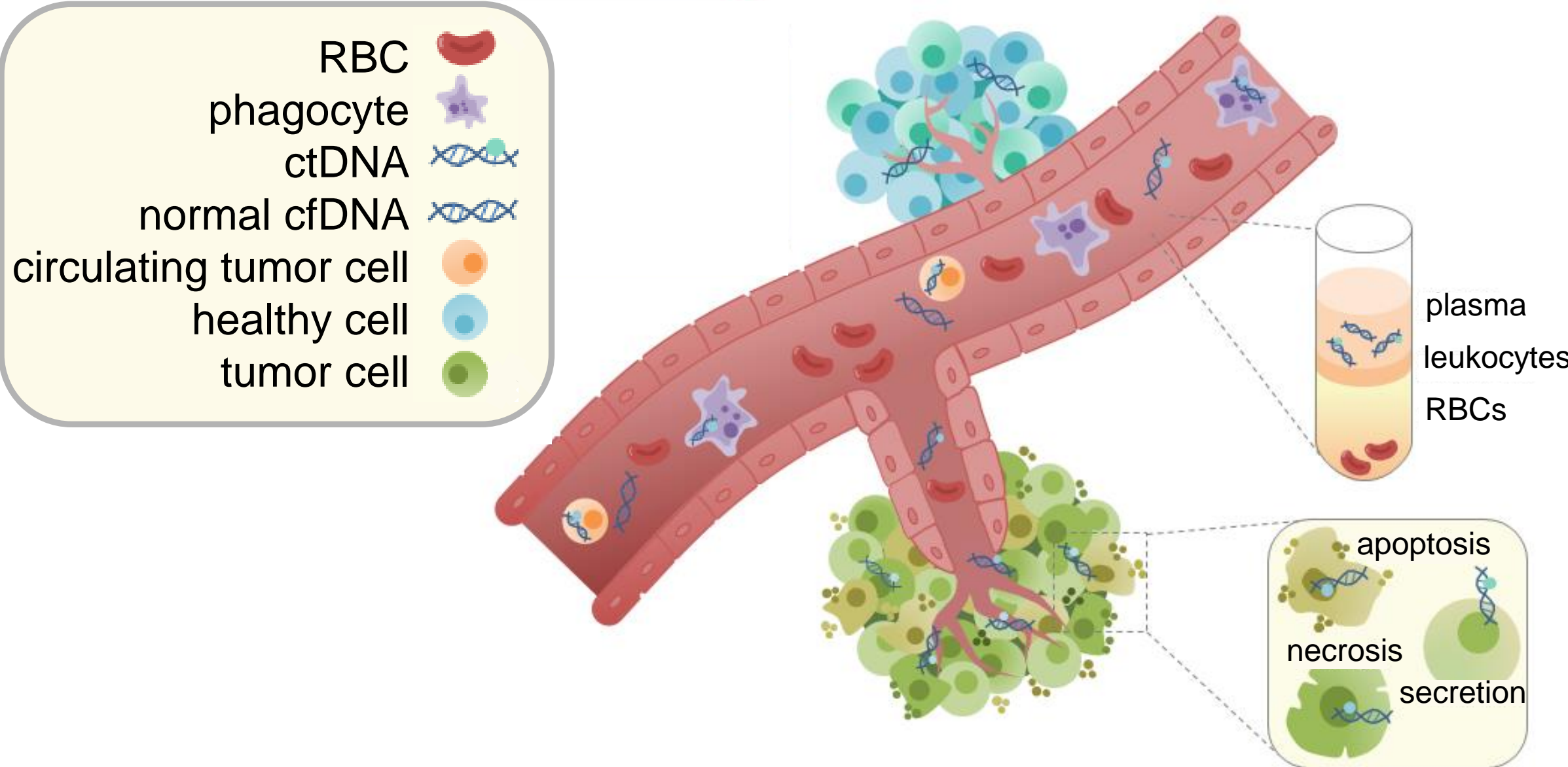
Introduction

Newly developed **blood tests** can target circulating fragments of DNA. Here, we propose applying this technology to identify **cancer** from peripheral blood.

2

Circulating Cell-Free DNA

The **human blood** is mainly composed of plasma, thrombocytes, red and white blood cells. It also consists low levels of **circulating cell-free DNA (cfDNA)**. These are short (~200bp) DNA segments, derived from apoptotic or necrotic cells, in multiple tissues.

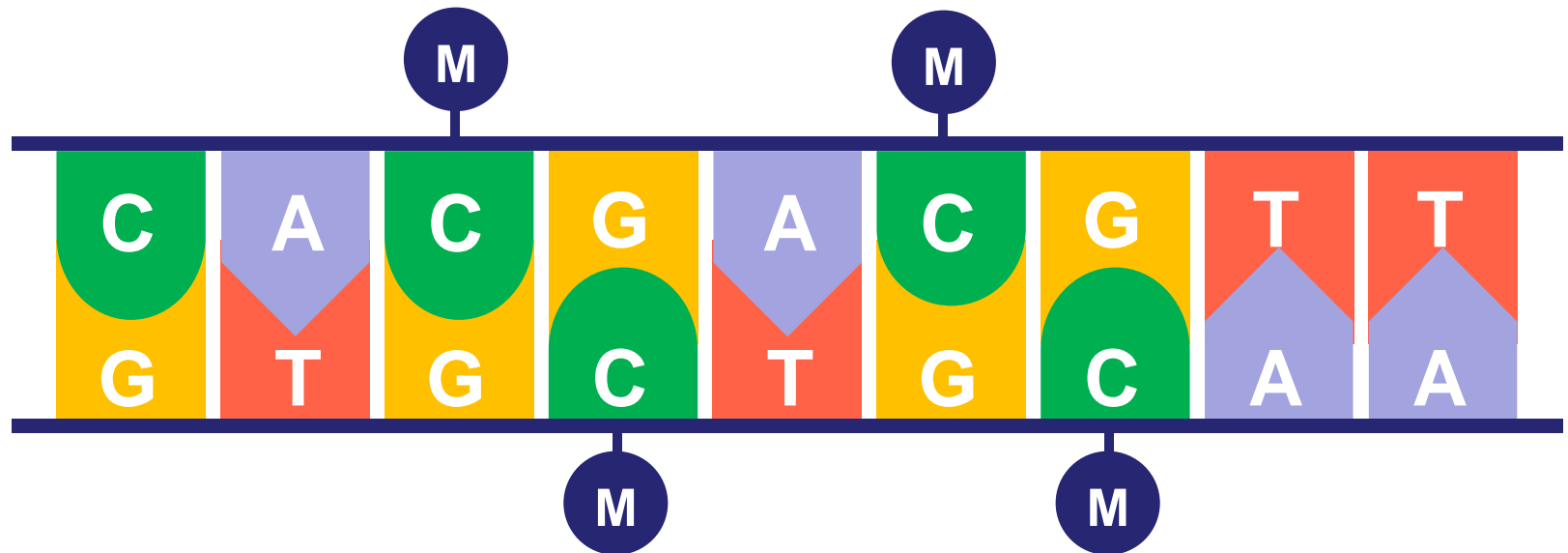


3

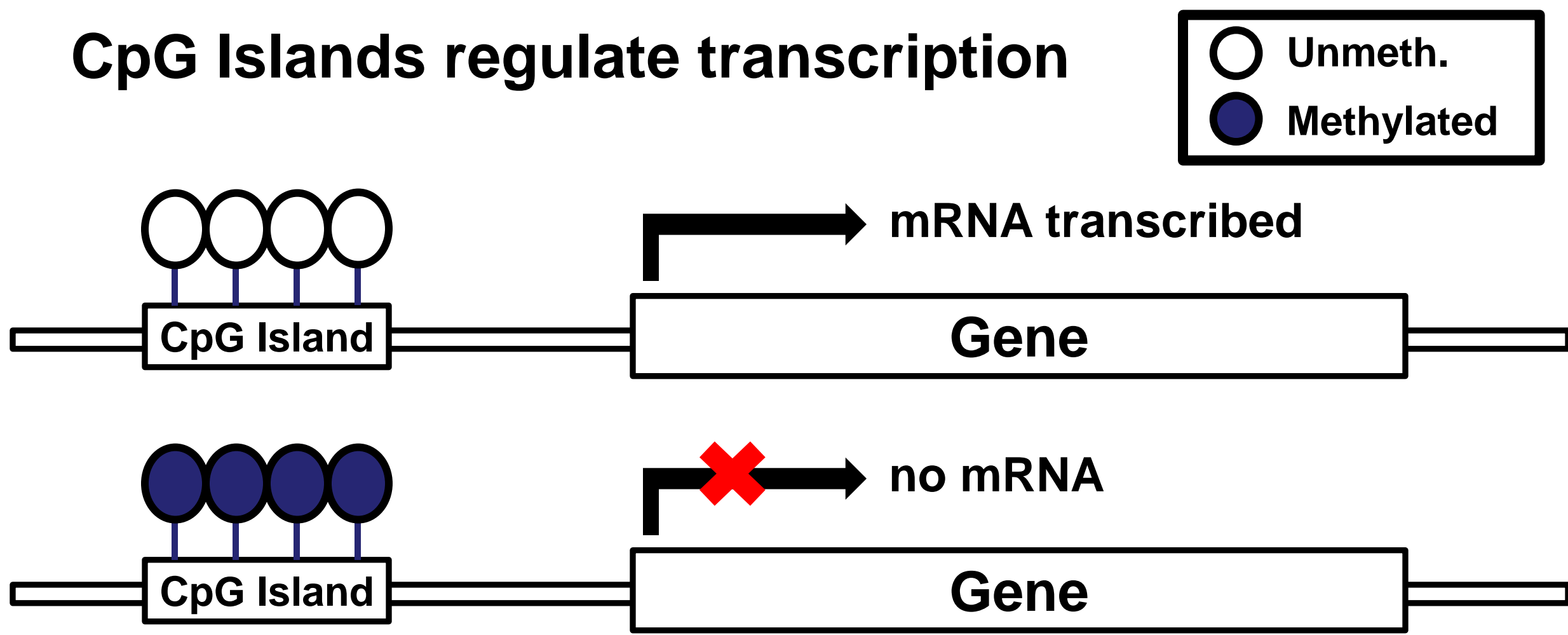
CpG Methylation Indicates cfDNA Origin

Studies have shown that **cfDNA** from different tissues differs in its **CpG methylation** patterns.

CpG Methylation: the addition of a methyl group to a **cytosine** nucleotide (when followed by guanine).



CpG Islands regulate transcription



4

Do Methylation Patterns Differ in Cancer?

Feasibility check:

- Use **Biopsies** instead of **cfDNA** (so the origin of **DNA** is known).
- Tissue-specific** markers (a step towards Pan-Cancer analysis).

Can we use **CpG methylation** patterns of DNA taken from **biopsies** to distinguish between cancer and healthy DNA?

5

TCGA Data

- Data from **TCGA** - The Cancer Genome Atlas Project. **TCGA**
- Illumina 450K methylation assay** quantifies methylation levels at specific ~450K CpGs within the human genome (out of 28M CpGs).
- 7,951** samples: **7,231 cancerous**, and **720 healthy** ones.
- First, focus on **breast cancer** (818 patients: 721 cancer, 97 healthy).

9

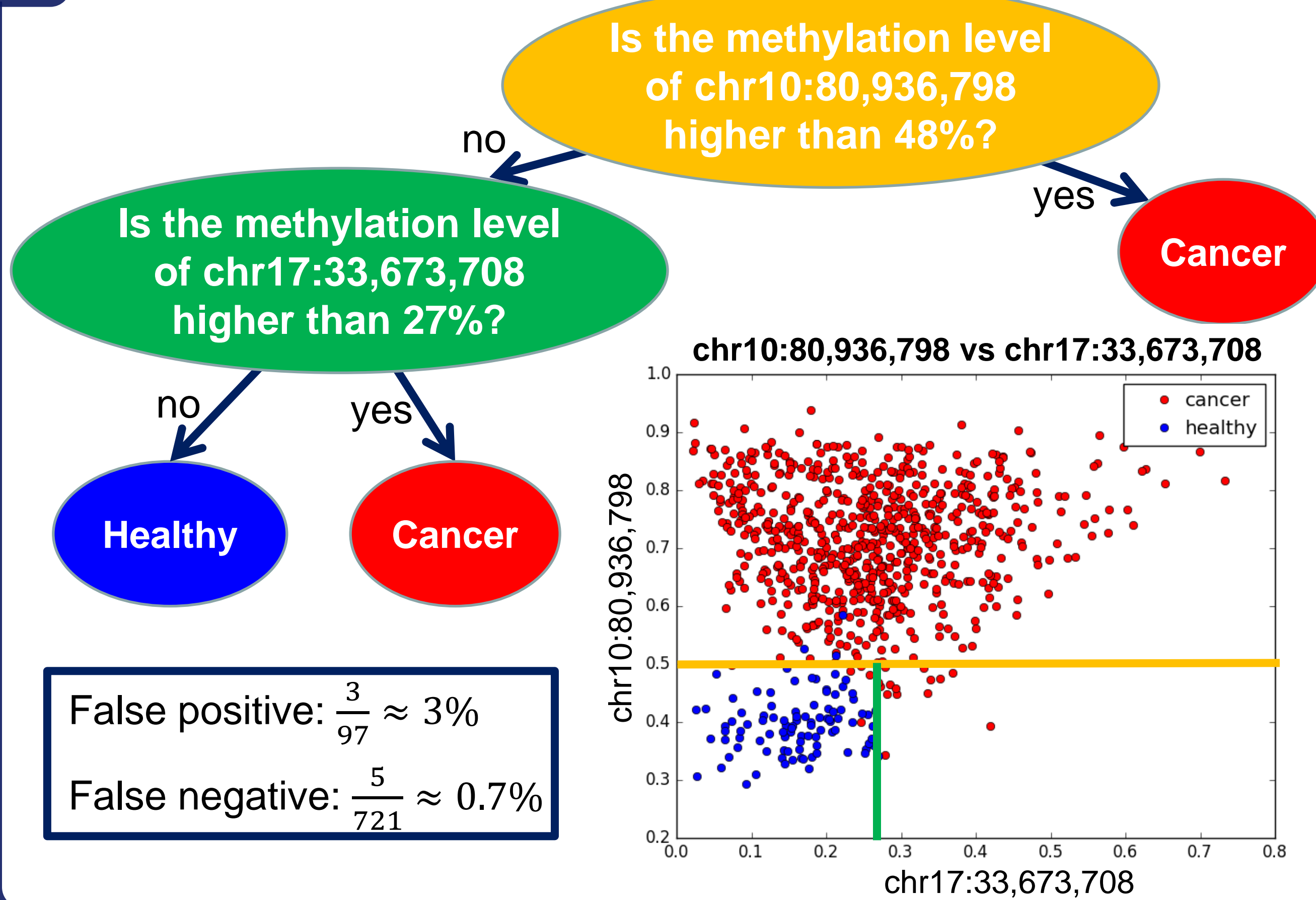
Summary and Future Plans

Given **CpG methylation patterns from biopsies**, it is possible to distinguish between healthy and cancer tissues, using only as few as 15 **CpGs** per tissue.

There are many **CpG sites** to choose from, which gives us opportunity to consider different CpG features (e.g. proximity, presence in blood).

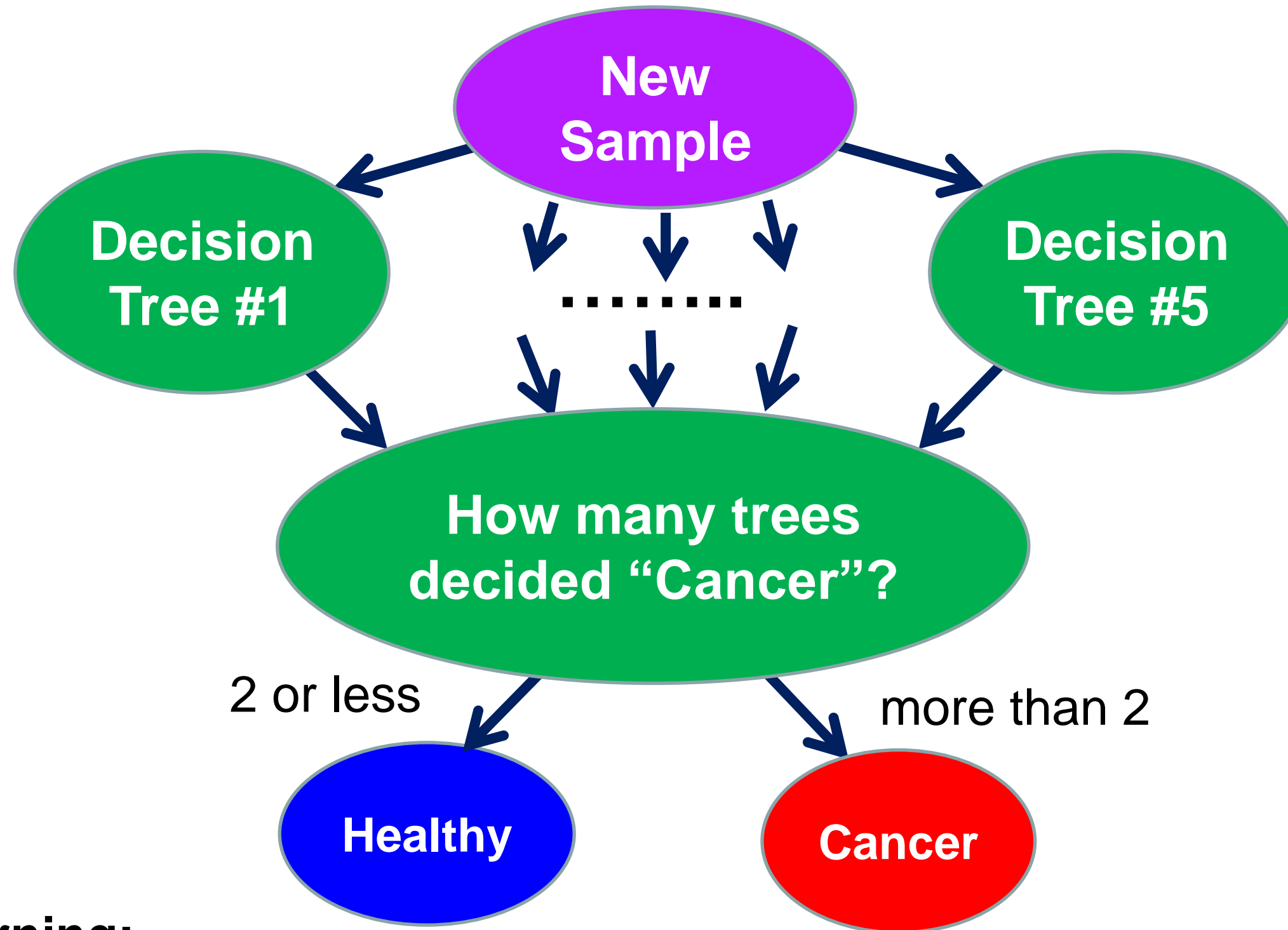
6

Decision Tree Classifiers



7

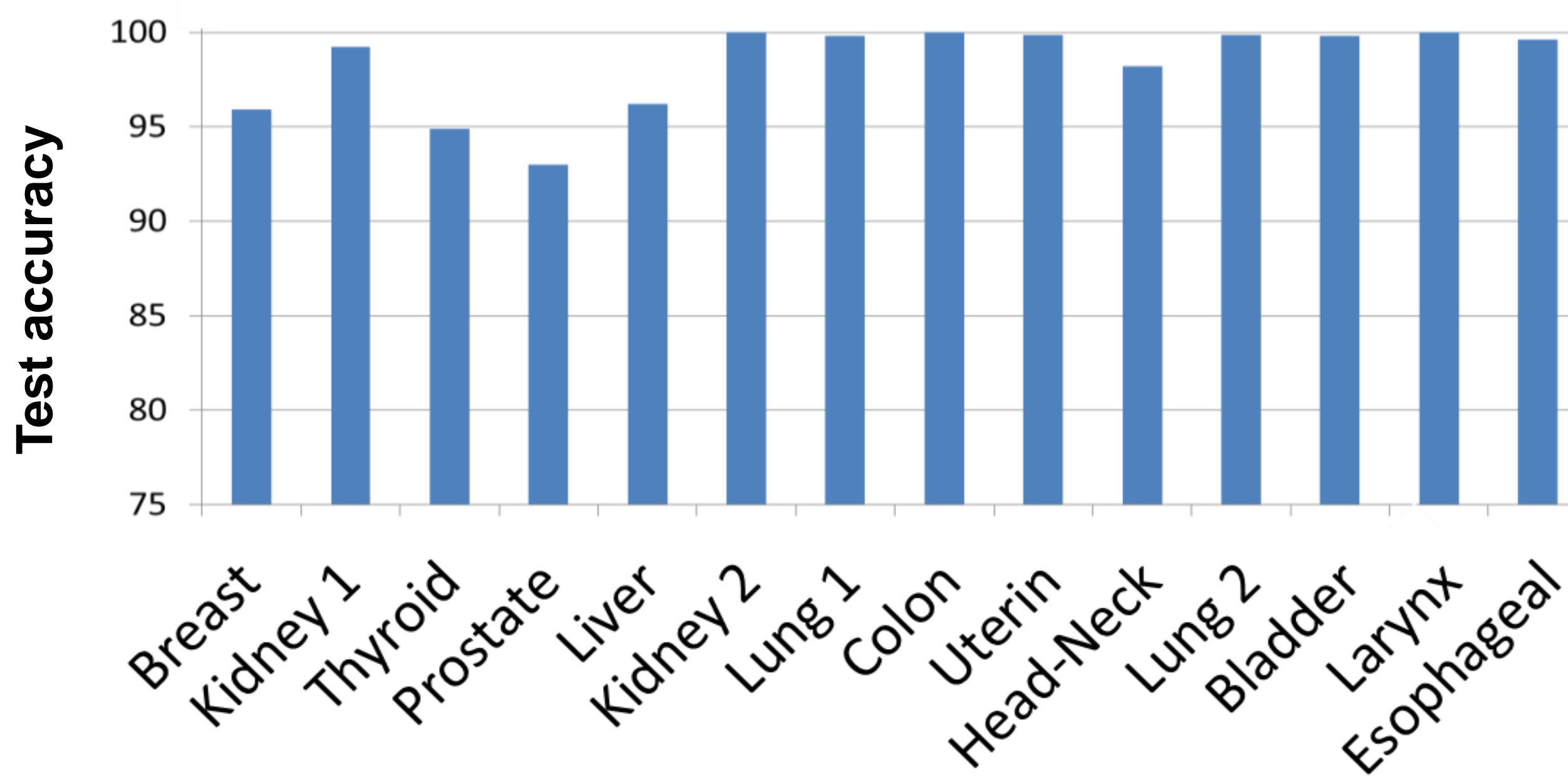
Forest Classifiers Improve Accuracy



Learning:

- Samples divided into **train** (5/6) and **test data** (1/6).
- We learned **five trees**, each trained on 4 of the 5 parts.

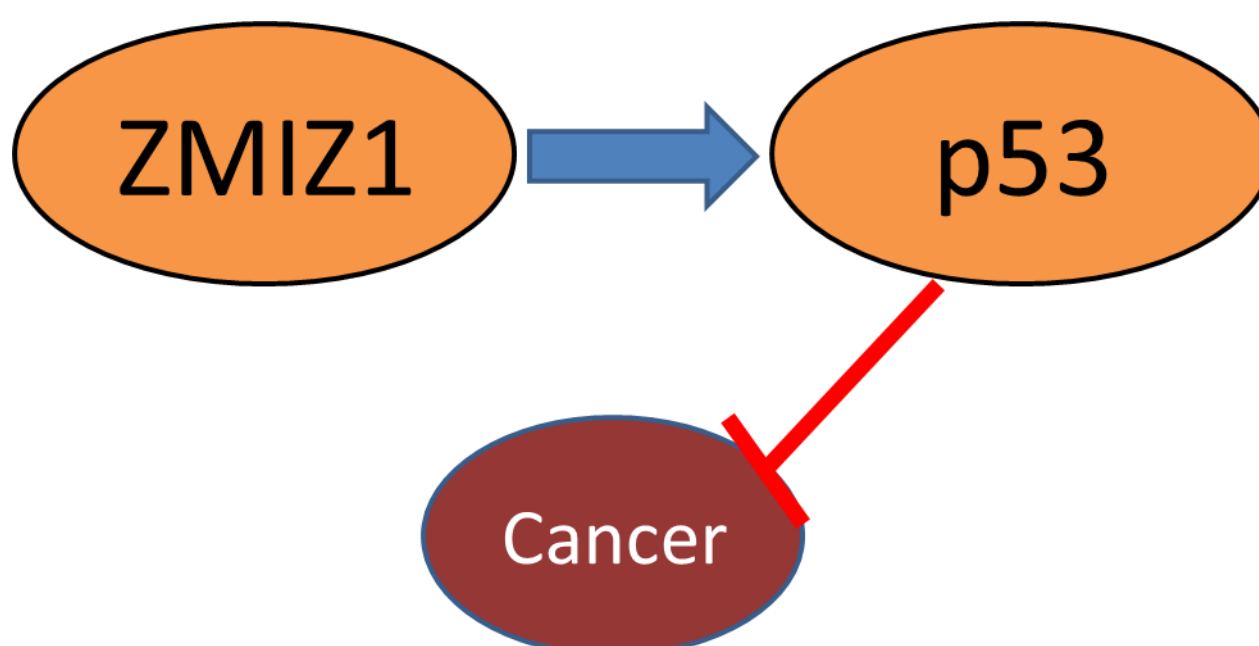
Success on Test Data - Different Tissues



8

Biology Example

CpG site at ZMIZ1's intron, a **coactivator** of tumor suppressor **p53**, over-methylated in **Breast Cancer** DNA.



However, a lot of CpG sites were found to be good classifiers, while many of them are not directly related to cancer.

Next:

- Sub-classifiers** within tissue – sub-type, cancer stage, prognosis, etc.
- Multi-classifier:** Pan-cancer classifier with tissue resolution.
- Blood:** Check accuracy with **CpG** methylation patterns found in blood as **test**. Find which sites are best for that purpose.