

FET Flagship Proposal

Uncovering the Human cell lineage tree in health and disease

Ehud Shapiro
Weizmann Institute of Science

Contact: ehud.shapiro@weizmann.ac.il

Uncovering the Human cell lineage tree in health and disease

Ambition

The Human cell lineage tree captures the history a person's cells since conception; it has 100 trillion leaves and branches ($\approx 100,000$ bigger than the Human genome); it is unknown; we should know it

Impact

Complete knowledge of the Human cell lineage tree would answer many pressing open questions in biology and medicine

Integration

Achieving this goal would require an information-centric approach, developing and integrating methods from biology, biochemistry, robotics, microfluidics, computer science and statistics

Plausibility

A proof of concept method and system exists, preliminary scientific results are startling

Support

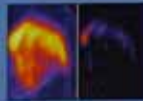
An international effort on the scale of the Human Genome Project is required. Significant advances towards the goal, with major scientific, medical and technological windfall, can be achieved in 10 years with a budget of €1Bn

The only known cell lineage tree

- A cell lineage tree captures the history of an organism's body cells since conception.
- Science knows the cell lineage tree of only one organism, the 1 millimeter, 1,000-cells worm *C. elegans*



July 15, 2008
Volume 68
Number 14
Pages 5501-6030



Tumor
Microstructure-Defined
Temporal Diffusion
Spectroscopy
Page 5941



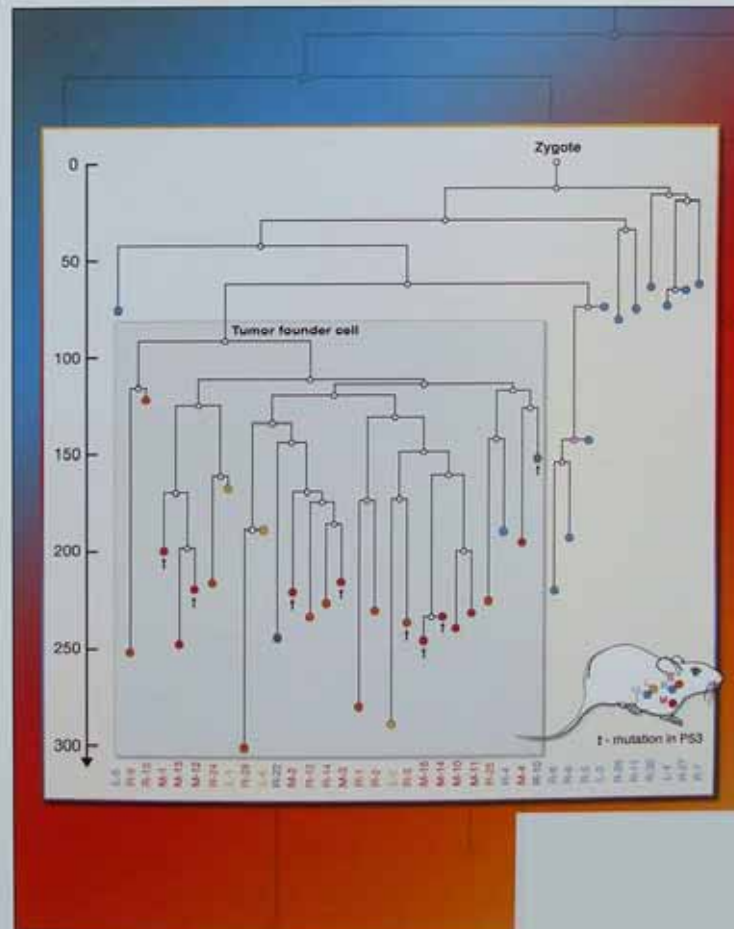
Merlin Is a Potent
Inhibitor of
Glioma Growth
Page 5733

Cell Lineage
Analysis of a
Mouse Tumor
Page 5924

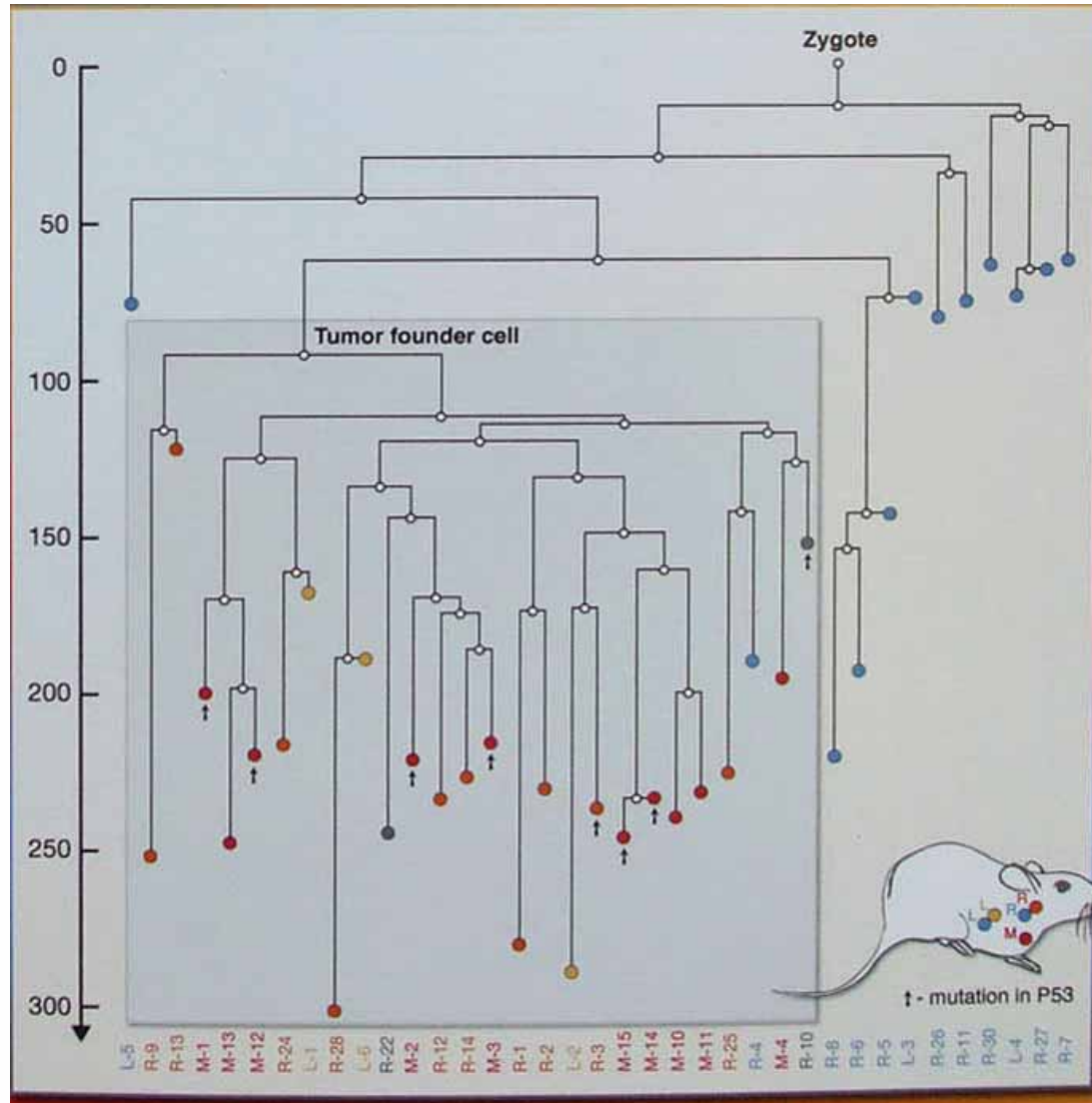


Cancer Research

Laboratory for
Biomolecular
Computers

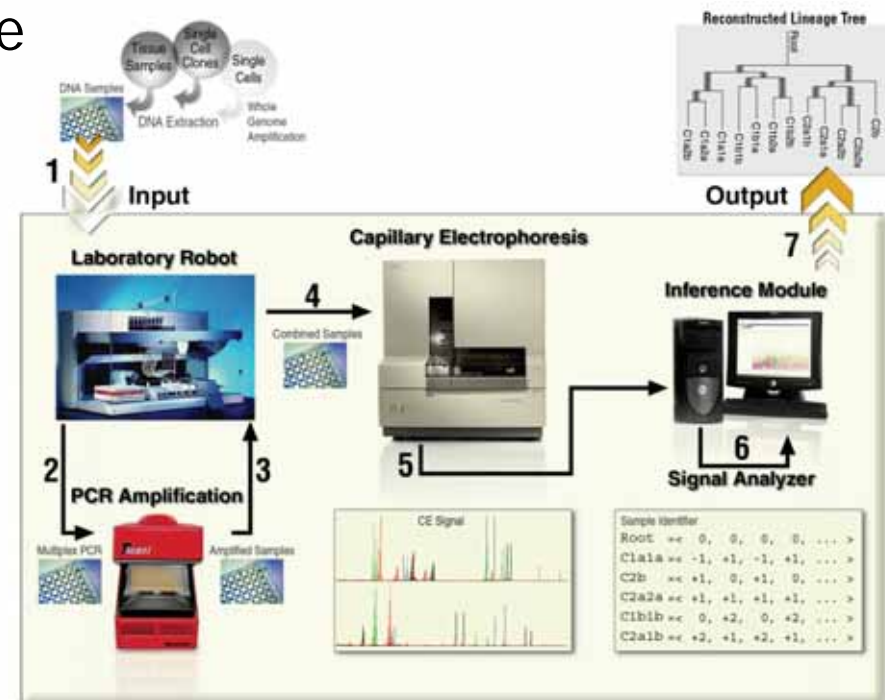


Lineage tree of cells sampled from a mouse with cancer

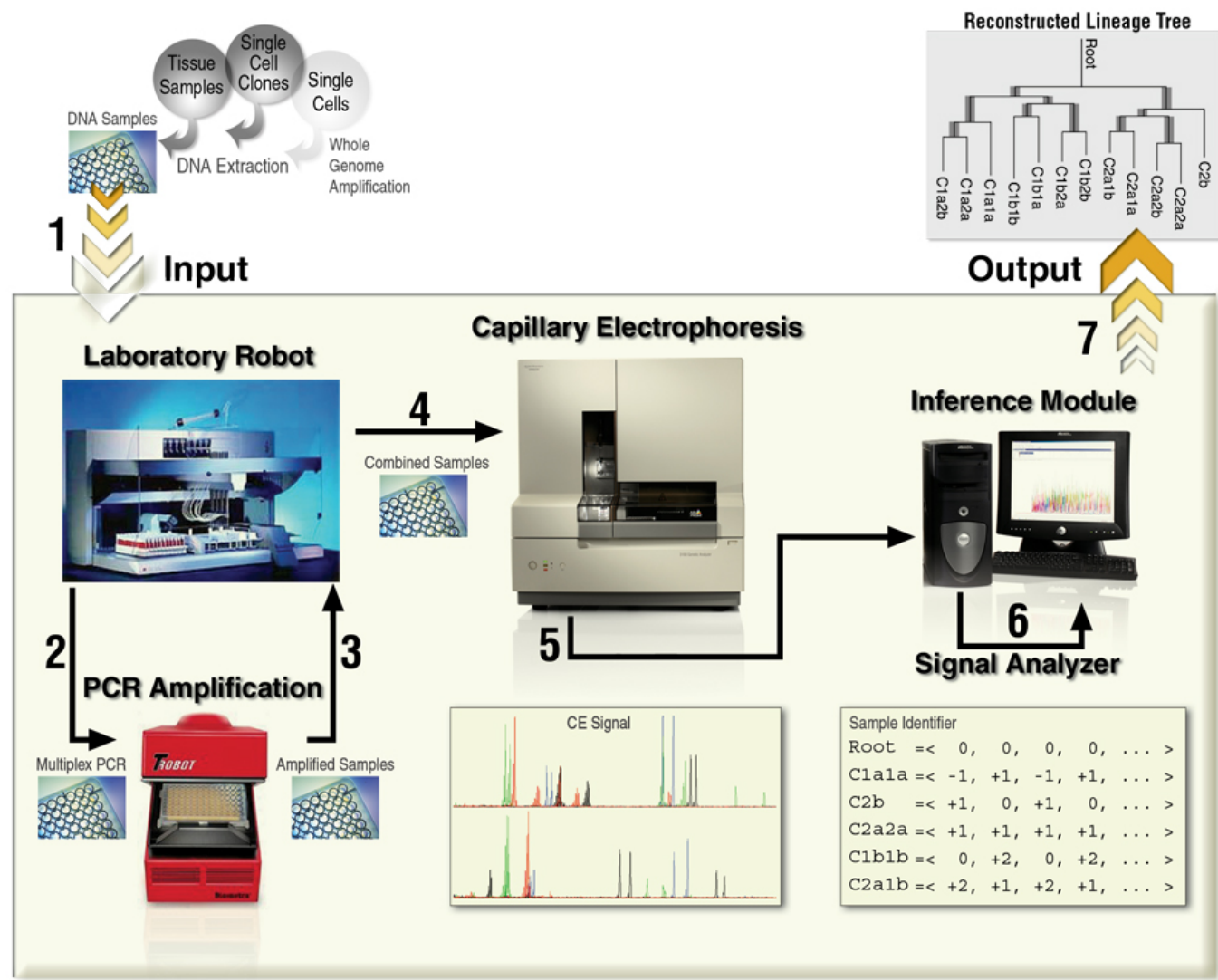


Uncovering the Human cell lineage tree in health and disease: **Concept**

- Our cell lineage tree is implicitly encoded in our cells' genomes via mutations that occur when body cells divide.
- Theoretically, it could be reconstructed by sequencing every cell in our body, at a cost of €100,000 trillion.
- Analyzing only highly-mutable fragments of the genome is sufficient for cell lineage reconstruction. This can be done at reasonable cost and at high throughput.
- Further improved by harnessing single-molecule NG sequencing technologies



Structure of a cell lineage analysis system



Uncovering the Human cell lineage tree in health and disease: **Ambition**

- The Human cell lineage tree captures the history of our body cells since conception.
- The cell lineage tree of an adult has about 100 trillion leaves and 100 trillion branches ($\approx 100,000$ bigger than the Human genome);
- The Human cell lineage tree is largely unknown.
- **Our ambition is to uncover the Human cell lineage tree in health and disease**

Uncovering the Human cell lineage tree in health and disease: **Impact**

Let's "Up the Ante"

Make money in poker:

Increase the stake, raise the bet

Prove a hard theorem in Mathematics:

To prove a hard theorem, formulate an even more general and powerful theorem and prove it. The original theorem then follows as a corollary.

The Human cell lineage tree in Biology and Medicine:

Many key problem in biological and medical research are actually specific questions about the Human cell lineage tree, in health and disease:

1. Which cells give rise to which other cells?
2. Which cells renew during adulthood?

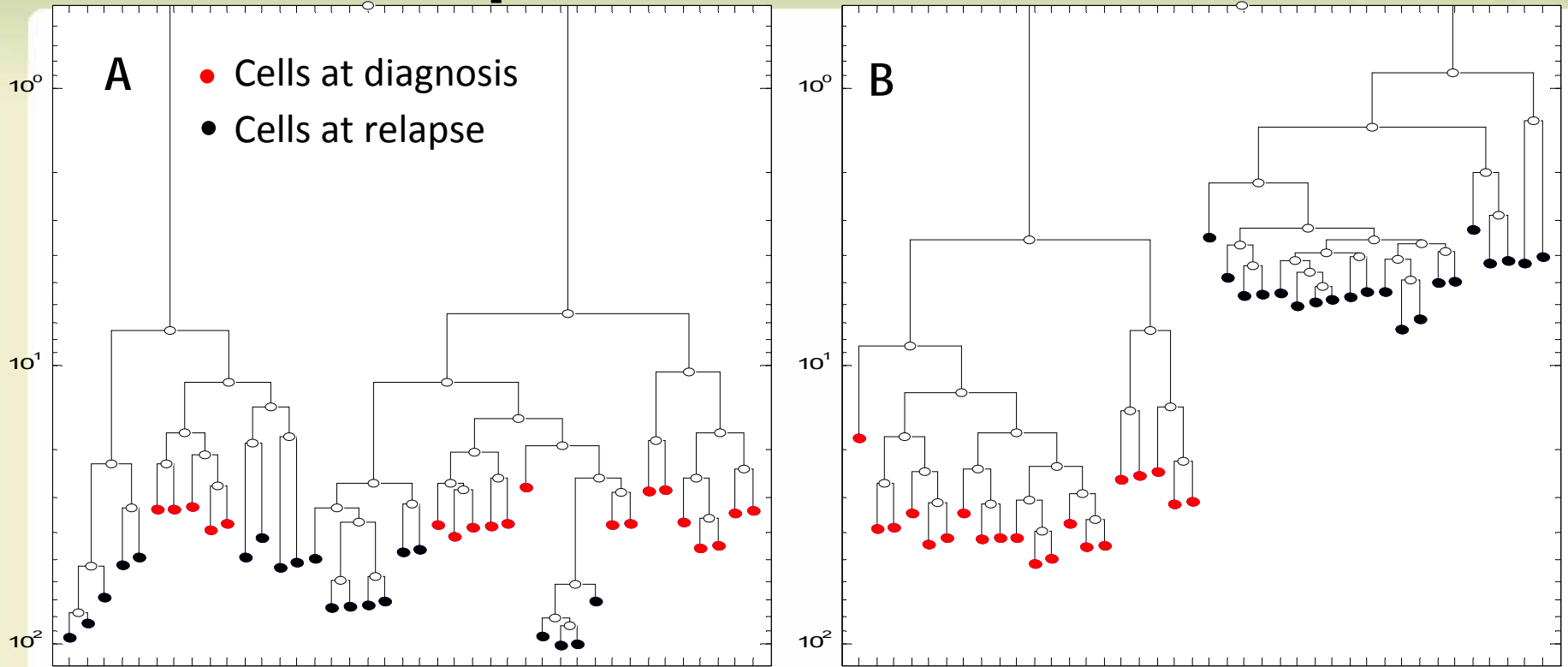
Uncovering the Human cell lineage tree in health and disease will answer all these questions

Uncovering the Human cell lineage tree in health and disease: **Impact**

Complete knowledge of the Human cell lineage tree in health and disease would answer many pressing open questions in biology and medicine:

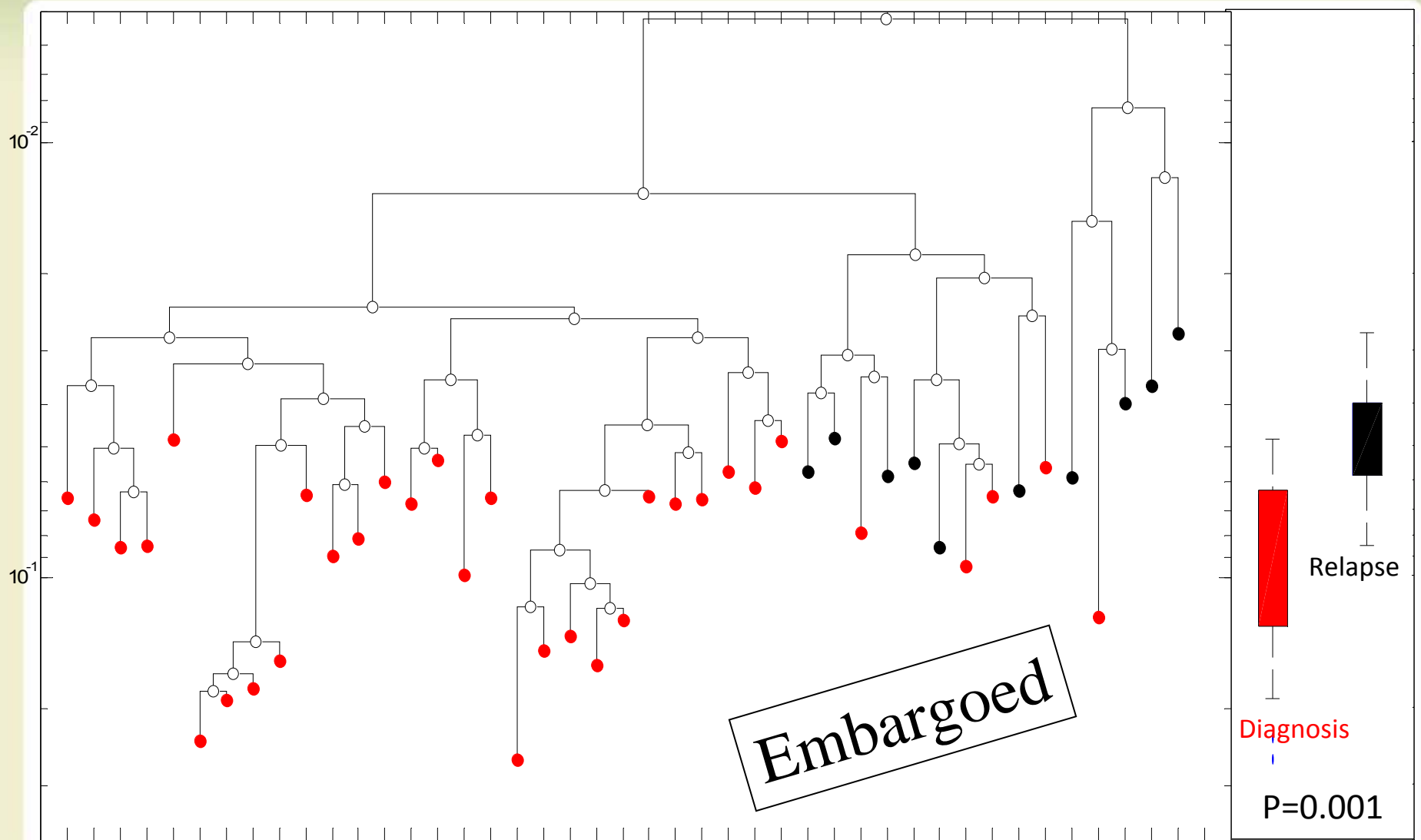
- Cancer: Which cancer cells initiate relapse after chemotherapy?
 - The answer will affect cancer therapy strategy.

Which cells initiate relapse? Two possible scenarios



- A. Ordinary cells that escape chemotherapy by accident
- B. Rarely-dividing cells unharmed by chemotherapy

Patient X at diagnosis and relapse



Uncovering the Human cell lineage tree in health and disease: **Impact**

Complete knowledge of the Human cell lineage tree in health and disease would answer many pressing open questions in biology and medicine:

- Cancer: Which cancer cells initiate relapse after chemotherapy?
 - The answer may affect cancer therapy strategy.
- Cancer: Which cancer cells can metastasize?
 - The answer may affect cancer therapy strategy.
- Diabetes: Do insulin-producing beta cells renew in healthy adults?
 - If they do, a regenerative approach to curing diabetes may be possible.
- Fertility: Do eggs renew in adult females?
- Brain: What is the normal development path of the brain?
- Brain: Which cells renew in healthy and in unhealthy adult brain?
- And many many more...

Uncovering the Human cell lineage tree in health and disease: **Integration**

Uncovering the Human cell lineage tree would require an **information-centric approach** that relies on developing and integrating methods from:

- **Biology:** Single-cell isolation, identification and characterization through multiple methods from all organs and systems in health, disease, and post-mortem.
- **Biochemistry:** Comprehensive molecular characterization of individual cells.
- **Robotics and microfluidics chips:** High-throughput genomic and other analyses of individual cells
- **Computer science:** Extremely large scale data capture, analysis and management tools, methods and algorithms, cell lineage analysis and reconstruction algorithms, interactive data analyses and presentation.
- **Statistics:** Putting its best tools to the utmost test.

Uncovering the Human cell lineage tree in health and disease: **Integration**

Structure and Budget
≈€50-100M/yr

PIs, individual labs
≈€0.2-0.5M/yr



Research

Computer Science, Math & Statistics, Biochemistry, Biology

National Cell Lineage Analysis Centers
National Labs
≈€2-5M/yr



Automation, biochemistry, microfluidics

International Data warehouse
Google/SAP/IBM/MS...
≈€10M/yr



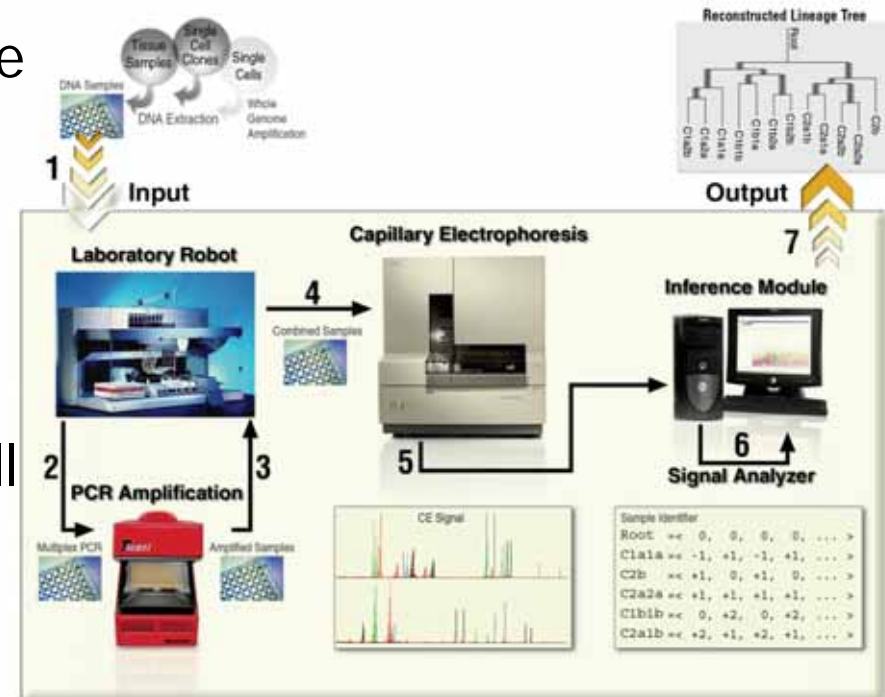
Data repository, data presentation

Bits

Uncovering the Human cell lineage tree in health and disease: **Plausibility**

A proof-of-concept method and system demonstrating plausibility exist, based on the following principles:

- Our cell lineage tree is implicitly encoded in our cells' genomes via mutations that occur when body cells divide.
- Theoretically, it could be reconstructed by sequencing every cell in our body, at a cost of €100,000 trillion.
- Analyzing only highly-mutable fragments of the genome is sufficient for cell lineage reconstruction. This can be done at reasonable cost and at high throughput.
- Preliminary scientific results obtained from small-scale cell lineage analysis are startling.



Uncovering the Human cell lineage tree in health and disease: **Support**

The proposed Human cell lineage project is similar to the Human genome project in ambition, scope, science and technology gaps at the outset, and potential impact upon successful completion.

- An international collaborative effort on the scale of the Human Genome Project is required.
- Cell lineage analysis centers to be established at each participating state.
- Collaboration within science, especially between computer science and biology, and between science and industry on high-throughput supporting technologies, is needed.

Uncovering the Human cell lineage tree in health and disease

Ambition

The Human cell lineage tree captures the history a person's cells since conception; it has 100 trillion leaves and branches ($\approx 100,000$ bigger than the Human genome); it is unknown; we should know it

Impact

Complete knowledge of the Human cell lineage tree would answer many pressing open questions in biology and medicine

Integration

Achieving this goal would require an information-centric approach, developing and integrating methods from biology, biochemistry, robotics, microfluidics, computer science and statistics

Plausibility

A proof of concept method and system exists, preliminary scientific results are startling

Support

An international effort on the scale of the Human Genome Project is required. Significant advances towards the goal, with major scientific, medical and technological windfall, can be achieved in 10 years with a budget of €1Bn