# FET Consultation FET Flagships

## A 4D Human Atlas: Charting Human Development and Ageing in Health and in Disease

**Ehud Shapiro**

Department of Computer Science and Applied Math and Department of Biological Chemistry

Weizmann Institute of Science

Rehovot 76100, Israel

- *What is your background? Are you submitting this proposal as an individual, or do you represent a community or institution?*

I am submitting this proposal as an individual. An active scientist for over three decades, I always explored uncharted territories across discipline boundaries, frequently ahead of my time and often with a long-term impact that became apparent only a decade or more later. In 2010, I assembled an international team of over 150 leading scientists and submitted a proposal to the first FET Flagship call, to uncover the human cell lineage tree. See proposal website. It was ranked 8th and hence did not pass to the second round. In retrospect, the proposal was a few years ahead of its time, as its enabling technology—single-cell genomics—did not even have a name then. Now, single-cell genomics is booming and it is the true enabler to the vision presented in our original Human Cell Lineage Tree FET Flagship proposal. It is also the enabler of the vision presented in this consultation document, which expands the vision of discovering the human cell lineage tree with discovering the 3D location of the cells in the tree thus providing a 4D Human Atlas, and with a Virtual Reality component that makes the Atlas amenable to human comprehension.

### What is the challenge and the vision?

- *What is the grand S&T challenge and its underlying vision and what are the main objectives your initiative would address? Why is this a grand S&T challenge and what makes it a "game-changer"?*

The grand S&T challenge is to construct a 4D Human Atlas that charts human development and ageing in time and in space at up to cellular (and at places even molecular) resolution, starting from the fertilized egg, through embryonic development, childhood, adolescence, adulthood and ageing, in health and in various diseases. The 4D Human Atlas would integrate an ever-increasing collection of *4D maps*, each providing (a typically partial) description of the cellular dynamics of a particular individual that had particular conditions. The proposed approach to constructing this atlas is based on single-cell genomics— one of the fastest advancing branches of science today, an epitome of interdisciplinary research and the newest and most exciting frontier of biology. The key technology to enable the construction of the 4D Human Atlas is single-cell multi-omics, the integrated and simultaneous genomic, epigenomic, transcriptomic, and possibly also proteomic analysis of individual cells. Genomic analysis of individual cells

obtained from a person, by biopsies, body fluid samples and/or autopsy, can uncover somatic mutations that endow each individual cell with a unique genomics signature. Phylogenetic analysis of such cellular genomics signatures can reveal the lineage relations among the cells. Transcriptomic analysis of individual cells can reveal their types. Complementing these analyses by epigenetic analysis further refines and improves cell lineage, cell type and state analysis. By applying multi-omics to single cells and analyzing the resulting data with appropriate algorithms, a mathematical entity encoding the 4D maps that comprise the 4D Human Atlas, termed *4D-labelled cell lineage tree*, can be constructed. The leaves of the tree would represent the analysed cells, whereas the tree's internal nodes would denote their ancestral cells. Types and 4D coordinates of the leaves (sampled cells) would be determined by experimental analysis, whereas those of their ancestral cells would be inferred computationally by various techniques, including machine learning. To allow comprehension of the 4D Human Atlas, it will be transformed by special software into a visually appealing 3D Virtual Reality application. Each 4D map in the Atlas would be transformed into a corresponding 3D VR scenario that could be traversed in time and in space by scientists and the public alike. By integrating multiple 4D maps across multiple conditions of health and disease, the navigable VR application would support navigation not only in time and in space but also under multiple health and disease conditions, serving as the ultimate tool for the research and learning of human biology and medicine.

The availability of a 4D Human Atlas of development and ageing in health and disease will revolutionize biology and medicine: Many fundamental open problems in human biology and medicine related to development, ageing and disease can be answered directly by observing 4D maps constructed from relevant samples. Topics include: What is the origin and dynamics of metastatic cancer? What is the cellular dynamics of ageing? How do stem cell niches in the brain develop and how are they maintained? What are the origins of immune-related diseases, such as multiple sclerosis and Alzheimer's? Are key tissues and cell types, such as beta cells, heart muscle cells and oocytes, renewed during adulthood? The 4D Human Atlas will also offer a powerful educational and research tool at all levels, from elementary school through university, medical school, medical practice and biological and medical research. Moreover, it is set to provide an essential component of personalized medicine by enabling the creation of personalized 4D maps from blood samples and/or biopsies, enabling their analysis and comparison to known maps of health and disease.

The very nature of the field of single-cell genomics in general and of single-cell multi-omics protocols in particular, and hence of the entire proposed project, is highly interdisciplinary. The dedicated assays typically involve advanced molecular biology and biochemistry, unconventional lab automation, microfluidics and sound fluidics; and cutting-edge sequencing technologies are integrated with sophisticated mathematics, algorithms and software to analyse and interpret the data. Consequently, the FET Flagship consortium will comprise experts in a wide range of disciplines, including the emerging field of single-cell genomics, computer science, human biology and medicine and leading biotechnology companies.

Here is a small sampling of the most recent achievements in the cutting-edge field of single cell genomics that are paving the way to the proposed FET Flagship project: first reconstruction of a cell lineage from whole-genome sequencing (Behjati et al., Nature 2014); first integration of RNA- and DNA-seq data of the same single cell (Dey et al., Nat Biotech 2015); enabling parallel sequencing of the mRNA and DNA of a single cell (Macaulay et al., Nat Methods 2015); first molecular classification of adult brain cell types (Zeisel et al., Science 2015); allowing parallel sequencing of the RNA & methyl DNA of a single cell (Angermueller et al., Nat Methods 2016); first demonstration of scRNA-seq of the entire ecosystem of solid human tumours (Tirosh et al., Science 2016).

FET Flagship offers the ideal platform for the 4D Human Atlas project. The visionary idea of constructing the 4D Human Atlas calls for the assemblage of a large-scale, multidisciplinary research team from a range of disciplines and backgrounds, including both academia and industry. The end result will be a powerful technology for widespread medical and scientific usage. It will position Europe as the primary focal point not only of the next wave of transformational scientific and medical technologies by establishing a strong scientific community in the field, but also go beyond to make that community the harbinger of single cell multi-omics integration technologies, as well as potentially spawning new industries based on the technologies created. It will also benefit Europe's society directly, and the world in general, by offering powerful medical diagnostic capabilities.

- *What are the main technologies, including digital technologies, which your initiative will advance?*

The initiative will advance and integrate several technologies. The core technology needed to realize this vision is single-cell multi-omics, which will probe cells at the molecular level, simultaneously inquiring the genome, epigenome, transcriptome and proteome of an individual cell. It needs to be coupled with multiple approaches to retain the 3D coordinates of the analysed cells and to visualize tissue and organ development, both prospectively and retrospectively. The amounts of data generated by high-throughput single-cell genomics make computing a critical technology. It would require incredible processing, storage and networking resources. Computational methods will need to be advanced across the board, including improved methods for the analysis single-cell multi-omics data and for describing, simulating and inferring cellular dynamics and for integrating single-cell multi-omics and 4D information.

## Why is it good for Europe?

- *Is your initiative relevant for the European industry and what is its innovation potential that would benefit Europe's economy and/or society?*

Europe is a hotbed of single-cell genomics, with pioneers and world leaders at the Sanger, Karolinska, Hubrecht and Weizmann Institutes, among others. Europe is also a leader in computational biology, high-performance computing and visualization. Bringing these technologies together to address key questions of biology and medicine could enable Europe to reap the benefits of its leadership in single-cell genomics and other fields of science and technology. While Europe's society will benefit from improved (personalized) medical diagnostic tools, its medical and scientific industries will be revitalized by new opportunities stemming from the ground-breaking technologies developed as part of the Flagship effort.

- *Are there existing international research initiatives linked to this proposal? How would this initiative position Europe with respect to other regions in the world?*

At present there are no initiatives of this scale anywhere. It would clearly position Europe as a pioneer of this vision.

## What would it take to do it?

- *What is the scale of the effort required to reach the objectives and how long will it take to do so?*

It is best to compare the proposed 4D Human Atlas initiative to the Human Genome Project. Unlike *C. elegans*, were all normal organisms have identical cell lineage trees, there is no one true human cell lineage tree and, therefore, no one true human 4D map. Much like our genomes,

the cell lineage tree of each of us is different, with our life experience shaping our growing cell lineage tree beyond our genetic differences. Hence, to uncover the 4D Human Atlas, we must discover the following regarding human cell lineage treea and their corresponding 3D maps: (i) their basic structure, dynamics and salient features shared by normal humans, (ii) variants resulting from normal genetic variance, (iii) variants associated with disease and (iv) variants associated with specific life experiences, e.g., nutrition, pregnancy, learning, exercise, and stress.

Making an analogy with the discovery of the human genome, goal (i) above is analogous to the goal of the original Human Genome Project, goal (ii) with that of the International HapMap Project, and goal (iii), restricted to cancer, with that of the International Cancer Genome Consortium. Note that while the ultimate objects sought by the various human genome projects are strings 3-billion-letters long, the 4D maps comprising the 4D Human Atlas could be up to 100-trillion-node 4D-labelled trees, with node labels encoding cell states with as much detail as practical.

Clearly, discovering the 4D Human Atlas will require the work of multiple international consortia over decades. A FET Flagship project could certainly achieve the initial vital spearheading fragment of this enormous endeavour. For example, a set of objectives that could be reached within one decade with a budget of 1Bn€ might be:

1. A 4D Atlas of Early Human Embryonic Development, say till day 30 (obtained via rapid autopsy of miscarried pregnancies).
2. A 4D Atlas of Cancer Development, for at least two types of cancer. For example, comprehensive samples of non-small cell lung cancer development in individual patients, including a resection of the primary tumour, circulating tumour cells from periodic blood samples, and metastases obtained via biopsies and rapid autopsy, could be provided by the TraceRx project.
3. A 4D Brain Atlas of Ageing, comparing between partial 4D brain maps of normal aged people and people with dementia (from brain banks).

These objectives are provided as an illustration of what might be expected from a FET Flagship project aiming towards the long term objective of a comprehensive 4D Human Atlas, and should be subject to discussion and refinement.

- *Why is Europe well positioned in terms of skills/expertise and capabilities, including industrial capabilities, to address the challenge and exploit the results? Which are the research communities to be involved?*

The key research community to lead this effort is the Single Cell Genomics community, which has tremendous European leadership. Its leaders can be found in the locations, organizing committees, and speakers of the first four such conferences: SCG2013, held at the Weizmann Institute of Science; SCG2014, held at Karolinska Institutet; SCG2015, held at Hubrecht Institute; and SCG2016, to be held at the Wellcome Trust Sanger Institute. The proposed interdisciplinary FET Flagship project will involve a broad range of research communities, including biotechnology (in particular single-cell genomics), biology and medicine (in particular, with the example objectives above, developmental biology, cancer biology and medicine, and brain research) and computer science (in particular high-performance computing, bioinformatics and machine learning).

- *Are there existing national or European research initiatives linked to this proposal? What is the added value for such an effort at the European level?*

There are no existing national or European research initiatives linked to this proposal. Pioneers and leaders of single-cell genomics have teamed up and proposed to construct a 4D Atlas of non-small cell lung cancer under the CRUK 20M£ grand challenge initiative (See CRUK website), with the challenge's winner to be announced in September 2016.

In conclusion, I think the added value to Europe to lead a cutting-edge, multi-disciplinary, international effort to construct a 4D Human Atlas that will revolutionize biology and medicine must be self-evident.