Cell lineage trees: the central structure plus key dynamics of biological aging and formulating the limiting problem of comprehensive organismal rejuvenation

Attila Csordas
AgeCurve Limited
attila@agecurve.co.uk
Cambridge, UK

Abstract

The argument makes the case for cell lineage trees and cell tree dynamics to be considered as the central structure and process of understanding organismal level, multicellular biological aging. The limiting challenge of counteracting biological aging is comprehensive organismal rejuvenation. The central theoretical problem of comprehensive biological rejuvenation is to find an algorithm to restore the balance and maintain the healthy dynamics of the aging organismal cell lineage tree. The most comprehensive medical solution of biological aging needs to use individual cell lineage trees as a central tool for diagnosis and treatment.

Introduction

What makes biological aging such a hard problem to crack conceptually and experimentally is its essentially temporal aspect. While the chronological clock is ticking and time, measured in decades, is passing, biological processes change their course for the worse. There's more and more low level inflammation, there's less and less, healthy, tissue-specific stem cells, et cetera. Aging can only be made visible by either already having data or inferring back

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27821v7 | CC BY 4.0 Open Access | rec: 19 Sep 2019, publ: 19 Sep 2019

the past biological history of an organism, to predict a future trajectory starting from the assessment of the present state. Aging defies a static view and forces a dynamic understanding. But dynamics is the methodologically hard part. In the context of human aging, any convincing study takes long decades, consider the Baltimore Longitudinal Study of Aging (BLSA), going strong since 1958. On the other hand and in order to come up with effective interventions counteracting different components of biological aging, the default, accepted and approved tool would be to run very carefully planned longitudinal randomised clinical trials.

This fundamental scientific and technological problem set concerning biological aging is propagated into current medicine, where all the individual human dimensions are added to it. And yet, to paraphrase Dobzhansky, most processes and interventions in human biomedicine make sense (only) in the light of biological aging. Be it the reference value ranges of disease biomarkers, the severity of the same disease, vaccination programmes, surgery options, eligibility for particular kinds of drugs, availability of participating in clinical trials, all these depend on chronological age and age group and more or less accelerated biological aging status.

In terms of the scientific understanding of biological aging, the mainstream view, the hallmark framework, contains a lists of 9 main processes that are separate but interconnected (López-Otín et al., 2013). The hallmark framework is a top level, process-based view but essentially, it is a list, a big static table without overall dynamics. It is also not clear how comprehensive this characterisation of molecular and cellular aging is. Due to the plurality of the main processes I will sometimes use the plural version agings, instead of the singular aging, or just aging(s).

What we don't have though is a comprehensive organismal level view. We lack a model that is comprehensive, interconnected and quantitative/mathematical enough to help us design combined interventions and provide a well quantified understanding on how far those treatments are going to push healthy longevity and under what intervention windows. What vehicle, concept, method will yield this necessary dynamical understanding and central intervention enabler, is a big question.

In this perspective piece I would like to propose a concept/biological structure that can act as the central vehicle, the glue of understanding biological aging, one that can pull together most known and perhaps several unknown aspects of this very broad and complex phenomenon. Then, by formulating the limiting problem of comprehensive biological rejuvenation, I would like to

use this biological structure & accompanying dynamics to offer a perspective and instrument to work towards comprehensive organismal rejuvenation in order to counteract the functional decline associated with biological aging, at the deepest level.

The macro-argument will go as follows:

1. The cell lineage tree is the central concept/structure and the dynamics of the cell lineage tree is the central process of (understanding) organismal level, multicellular biological aging(s).
2. The limiting challenge of counteracting biological aging(s) is comprehensive organismal rejuvenation.
3. The central problem of comprehensive biological rejuvenation is to find an algorithm to restore the balance and maintain the healthy dynamics of the aging organismal cell lineage tree.
4. The most comprehensive medical solution counteracting biological aging(s) needs to use individual cell lineage trees as a central tool for developing new diagnosis and treatment options.

Cell lineage tree is the central structure and cell tree dynamics is the central process of organismal level, multicellular biological aging

In this section there are 5 subsections: First, we frame a list of requirements that makes a concept/structure/process central in terms of modelling, understanding and representing organismal aging. Second, we show how cell lineage tree and cell tree dynamics meet these requirements. Third, we raise objections against the centrality of tree dynamics in aging and show some of the limits of this centrality while addressing these objections. Fourth, we summarise briefly the history of cell lineage tree research, Fifth, we mention current projects and initiatives trying to capture aging dynamics, and their relations to cell lineage trees and tree dynamics.

1. The requirements to qualify as a central concept/process of biological aging

What do we expect from a central concept, structure, process of organismal biological aging? This question is about the potential expressiveness, explanatory, predictive and modelling power of a scientific tool. We list 10 requirements, grouped according to 6 bigger concepts, highlighted in bold. We don't aim to provide an exhaustive list, listing all necessary conditions together being sufficient to provide a definition of centrality in biological aging. Some requirements will (intentionally) overlap as well and logically and conceptually are not independent from each other, so might seem redundant. What we aim to provide is a starter list of requirements to assess centrality of a tool in aging research. The main aim is not a methodological enquiry with epistemological finesse but suggesting a new theoretical perspective or frame to scientists, technologists and doctors working with biological aging with enough practicalities to motivate increasing adoption. Hence the requirements are listed not in the order of weighted relevance but more in a didactic manner trying to ease understanding.

Biological intuition

- It should be an actual biological structure, concept, process, not some higher level abstraction, so it can serve biological intuition.

Quantitative, mathematical, statistical, easily expressible with computer science

- It should be simple.
- It should be deeply quantitative, mathematical.

Evolutionary

- It should be able to incorporate basic evolutionary concepts and processes.

Temporality, spatiotemporality

- It should be deeply temporal, dynamic, also cover development and being able to catch acceleration and deceleration of aging.
- It could be easily interpreted in spatial terms.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27821v7 | CC BY 4.0 Open Access | rec: 19 Sep 2019, publ: 19 Sep 2019

Originality, biological scalability, comprehensiveness

The requirements listed here extend on the spatiotemporality requirements. Comprehensiveness can be interpreted spatially as including the whole body, temporally as including the whole temporal trajectory for biological aging and conceptually/logically as including all the different layers embedded into each other.

- It should provide an original way to comprehensible show how aging disrupts key biological processes.
- It should be able to cover, represent, express most hallmark aging processes.
- Scalability: talking about multicellularity requires that it is close to the cell concept but can be extended into upper (tissue, organs, scale-up) and lower level (organellar, molecular, scale-down) structures.

Biomedical representative, expressive power, life historical resolution

- It should provide near exhaustive coverage of life events, corresponding to biological and medical history.

2. How cell lineage tree and cell tree dynamics meet the requirements for centrality in biological aging

Here we progress according to the order the requirements have been listed in the previous section.

Biological intuition: what are cell lineage trees?

- It should be an actual biological structure, concept, process, not some higher level abstraction, so it can serve biological intuition.

All normal multicellular organisms, generated both by sexual or asexual reproduction, start with one cell and grow into a multicellular organism by subsequent divisions, where parent cells give

rise to two offspring cells. This process with all its component cells represents a very elegant mathematical structure, a so called binary tree. Binary means 2 is the maximum number of offsprings.These particular biological, multicellular binary trees are called cell lineage trees. I will refer to these cell lineage trees sometimes just simply as cell trees as a convenient shorthand. Lineage trees can be 'invariant' (more like low variability eg. C. elegans, leech, in reality) or variant (high variability), eg. fish (Zebrafish) or mammals (mouse, humans). To restrict discussion to the intervention target the main cell tree object of our perspective is going to be the individual human cell lineage tree. Every human individual starts with the zygote as the root of the tree and transitions into all other subsequent states of human life, from early to middle to late life, capturing processes like development, aging and death. I use the neutral term 'transition' to be able to incorporate all kinds of changes in the tree trajectory, growth, shrinking, dramatic re-arrangement.

At any time point during adulthood the leaves of the cell tree form the actual human cellular components of the individual organism.The ballpark estimation is that an adult human individual body weighing 70 kg is composed of ~30 trillion human cells as the extant leaves of the cell tree (Sender et al., 2016). The extant leaves and the internal dead leaves or nodes can be labelled and also given unique identifiers.

An unlabelled cell tree encodes one default information: the topology of the tree, ie. the structure of the branching processes. If the tree is labelled this gives rise to another default information; a well-defined distance between all the different leaves.

The dynamics of tree captures all the transitions, changes, processes of the tree. Different parts of the tree are forming partial lineage trees, or subtrees. For instance, a clonally derived tumour forms a particular subtree. The unabridged, time-indexed whole-organism tree captures all the actual leaves of the tree forming the body. I will call the collection of all the actual leaves, indexed by the same time of the same organism the actual cell **foliage** as an analogy of real trees. Considered from this point-of-view, all previous layers of once extant leaves form time-indexed foliages.

I think by now it can be seen why cell lineage trees and their dynamics serve existing biological intuition well.

First, cell trees use as units the fundamental concept, process and reality of multicellular biology; the cell and its division that gives rise to more cells.

Second, they provide easy, natural analogies based on our accumulated experience on trees, that we can borrow to re-apply to animals, including ourselves. This way, thinking about cell trees also unites multicellular and even uni-cellular living things. After all, a bacterium can divide too, forming embryonic tree structures.
Third, highlighting one aspect of the previous point, trees can lend themselves to easy visual representation, just like tree dynamics can be captured as a sequence of subsequent topologies, or a movie made out of those sequences.
But in order to make thinking about cell lineage trees easy and to apply them in a biologically meaningful manner, we need to abstract away from some of their properties in order to use only properties, that matter for our investigation. In brief, we need to use mathematics.

Quantitative, mathematical, statistical, easily expressible with computer science

- It should be **simple**.

Binary trees are one of the default data structures in computer science, taught at Compute Science 101 classes right after lists and arrays, as traversable structures. Implementing different tree traversals, like depth-first search is one of the first algorithms programmers, or any kind of people learning the basics of coding, encounter.

- it should be deeply **quantitative**, **mathematical** from the ground-up

Binary trees and operations on them are an organic part of graph theory within discrete mathematics. Trees are special kind of graphs, where any two nodes (vertices) are connected by only one path and does not contain any cycles. Binary trees are the simplest meaningful tree structures where a parent node can only have up to two offspring nodes. The connections between nodes can be encoded into so called incidence matrices, leading to other kind of matrix representations as well through a series of operations. There's an immense mathematical literature dealing with all the minutiae details of the macro and micro structure of trees. Trees can be embedded into one another, and subtrees, with all their nodes can be represented as just one node as part of smaller, more compact aggregate trees that can preserve the macro structure

off the original, 'uncut' tree while abstracting away from irrelevant details economically. Most actual cell lineage trees that are recovered or going to be recovered from variable lineage trees, human cell trees being one of them, are these type of compact trees.

In terms of tree dynamics, trees can be grown random, or simulated according to different rules incorporating biological knowledge. Importantly, different statistical methods can be applied to variant cell lineage trees to handle variability, we will mention results later in the review section. Although probability theory and statistics are a branch of mathematics, luckily in biology there's already a strong computational, theoretical biology tradition with a plethora of tools that can be re-applied to cell trees too, with adjustments. This source is evolutionary biology and phylogenetics.

Evolutionary

- It should be able to incorporate basic evolutionary concepts and processes.

Evolutionary biology on the macro-scale concerns with the speciation processes and events to be able to reconstruct the tree of life. A more technical term is phylogenetics, where phylogenetic investigates the evolutionary relationships and histories between species, individuals and genes (EMBL-EBI Train Online: What is phylogenetics?).

The good news is that cell lineage trees are also evolutionary products and hence sequence evolution, selection processes, mutation rates and other core evolutionary concepts can be interpreted within them. Instead of talking about the species tree of global life we can just talk about the organismal tree of individual life and death.

In phylogenetics, there are at least four difference approaches in reconstructing species lineage trees, maximum parsimony, neighbour joining, maximum likelihood and Bayesian, each with different strengths and weaknesses and different expressional power in terms of the level and depth of tree recovery. Most of these approaches have already been applied to cell lineage trees, see literature review section. Let me just highlight one Bayesian approach that has the potential to recover intra-individual, cell lineage/cell population specific mutation rates (Csordas and Bouckaert, 2016). Mutation rate is one of the most important, if not the most fundamental, evolutionary concept, as evolution is not a constant rate, steady, balanced process and the evolutionary clock can speed up or slow down. This acceleration and deceleration of

evolutionary, biological time already hints at a crucial requirement for any effective tool in understanding the pace of biological aging: the ability to handle temporal order and structure almost natively. The ability to connect and yet separate chronological and biological time from each other.

Dynamics: temporality, spatiotemporality, change, pace

- It should be deeply **temporal.**

Cell lineage tree dynamics is deeply temporal delivering the temporal order, the subsequent waves of extant cells in a multicellular organism. Edges between cells can naturally represent time or pseudotime or some other temporal construct.

- It should be **dynamic**, also cover development and being able to catch acceleration and deceleration of aging

Cell tree dynamics makes it possible theoretically to model different, extant cell populations and track their ballpark numbers and ratios to depict and detect accelerated or decelerated or typical individual aging and tissue, organ or perhaps organismal level regenerative potentials. More on this kind of scalability in the next section.

- It could be easily interpreted as **spatial** organisation, **location** awareness.

The labels or (unique) identifiers given to the cells can be easily used to indicate spatial location within the body. On a basic level, ordered trees are ones where lineal position of offspring cells are indicated, these are the so called left and right nodes of a parent node. It's quite simple to give these already spatial motivated, directional terms an actual three dimensional spatial meaning for instance when cell tree dynamics is modelling development and anterior/ posterior is encoded into left/right. Bottom line is that binary trees are already strong spatial patterns and into the structure of these trees temporality is deeply encoded yielding such an embedded, super-flexible scientific tool that can represent many more processes and layers on the organismal biological scale.

Originality, biological scalability, comprehensiveness

The requirements listed here extend on the spatiotemporality requirements. Comprehensiveness can be interpreted spatially as including the whole body, temporally as including the whole temporal trajectory of biological aging and conceptually/logically as including all the different layers embedded into each other.

- It should provide an **original** way to comprehensible show how aging disrupts key biological processes.

I think there's little doubt that once we have a growing understanding on the larger scale shape, topology and dynamics of the human cell lineage tree we can finally start to put together the comprehensive sequence of how the regenerative potential of particular cell types, tissues and organs are gradually or continuously decrease with time. One of the hallmark aging process is stem cell exhaustion and tree dynamics is the tool that can deliver us the body-wide spatiotemporal pattern of stem cell exhaustion or more generally regenerative exhaustion. Considering how much cell lineage tree research already delivered (see literature review section) it is reasonable to assume that macro-scale structural properties of the cell tree and tree dynamics informed by detailed and representative micro-scale properties will shed new and comprehensive light on the aging process. The mathematical properties are out there that can be filled up with biological meaning and developed further.

- **Scalability**: Talking about multicellularity requires that it is close to the cell concept but can be extended into upper (tissue, organs, scale-up) and lower level (organellar, cellular, scale-down).

Upward scalability of the cell tree concept and tree dynamics to include coverage and representation of tissues and organs seems pretty straightforward. The mathematical representation of the tree with cells being its fundamental units, the nodes, lend itself easily to upward aggregation, mentioned already above in terms of compressing the tree to have cell

groups or subtrees as new units. Or if lineages forming tissues and organs turn out to be to diverse and dispersed around the tree then tree representation can be transformed into a more complicated graph representation as long as it can serve both large scale computation and biological intuition.

Downward scalability is a more challenging problem but it is definitely not an impenetrable and intractable one. As long as single cell measurements can provide different organellar, molecular phenotypes these can be used as labels on the extant cells and variability, patterns and association of phenotypes can be computed.

In terms of biomolecules, high throughput single cell DNA and RNA sequencing measurements are the main vehicles today of delivering cell trees and cellular trajectories and expression patterns. Technically speaking mass spectrometry based single cell resolution proteomics can also deliver cell tree topologies and dynamics by sequencing individual proteins expressed in single cells and tracking their variations, although as far as I know this has not happened yet (Budnik et al.,2018). This emerging, proof-of-concept technique has been already used for assessing macrophage heterogeneity (Specht et al., 2019). The same potential might apply to protein modifications delivered via a modified version of single cell mass spectrometry proteomics. Single-cell metabolomics, the quantitative measurement of hundreds of metabolites(amino acids, sugars, lipids) using mass spectrometry has taken off in the last couple of years (Duncan et al.,2019) but it's an open question whether this type of data can be used for cell tree inference.

In terms of organelles, let's just discuss mitochondria as they are special due to their DNA content and their huge and varied roles played in the aging process. Somatic mutations in mitochondrial DNA detected within single cells via single-cell genomics (single-cell RNA or assay for transposase accessible chromatin (ATAC) sequencing) can be used for lineage tracing in human cells (Ludwig et al. 2019). Due to their particular DNA content and the nature of mitochondrial biogenesis and decay, fission and fusion events, mitochondria form trees as well. A more distant opportunity might be advanced by single mitochondrion DNA sequencing (Morris et al. 2017) where intracellular mitochondrial trees can be overlaid upon cell lineage trees. Coalescent theory might help here provide a population level account on mitochondria embedded into cell trees.

To make my argument a bit stronger: I cannot think of any structure and dynamics other than cell tree and tree dynamics that could serve as the central layer and vehicle of overlaying all other phenotypic, genotypic information on top.

- It should be able to cover, represent, express most **hallmark aging processes.**

The already mentioned hallmark framework is perhaps the most accepted, comprehensive view on different but interconnected molecular and cellular aging processes. Below we will ask, in some cases speculatively, how cell trees are or can be used to model these processes, and what single cell measurements are in existence to potentially provide data for lineage tree reconstruction and tree dynamics simulations. There's significant overlapping with the scalability section above, but the reason to spell out some opportunities in the context of the hallmark framework is to motivate our centrality argument further.

Genomic instability: Single-cell DNA and RNA sequencing made it possible to track the accumulation of somatic mutations during the lifetime of an organism. This emerging sub-discipline is explicitly called somatic evolution and it already delivered the somewhat surprising result of showing how somatic mutant clones, harbouring cancer-associated mutations accumulate in the human esophagus with age, using only a handful of individuals (Martincorena et al.,2018). Directly utilising this kind of somatic mutation information is instrumental for cell lineage trees.

Telomere attrition: telomere shortening mechanistically tied to the chromosomal end replication problem during cell divisions, manifesting with advancing age is another genomic alteration of biological aging. As such using this information is quite natural for lineage tracing and cell trees. Single cell DNA sequencing measures the telomeres too, although to understand and track the complexity of replicative senescence, the protein components (telomerases) need be measured too.

Epigenetic alteration: Epigenetics has a narrower and a broader interpretation. The narrower includes age-associated DNA methylation changes and chromatin remodelling, the broader can include histone protein modification patterns. Single-cell epigenomics has already been used to track and assemble lineage histories of chronic lymphocytic leukaemia (Gaiti et al., 2019).

Single cell resolution histone modifications, methylation, acetylation, can potentially be tracked with an extension of the above mentioned single cell mass spectrometry proteomics approaches and be used for cell tree inference.

Loss of proteostasis: The deterioration of protein quality control systems is an integrative aging hallmark, and includes so distant mechanisms as the ubiquitin-proteasome system and lysosomal pathways, involved in chaperone mediated autophagy, micro- and macro-authophagy. Mass spectrometry based single cell proteomics can deliver some relevant information related to the protein components of these systems, for instance proteasome complexes and chaperones, but it's an open question how this information can be used for lineage tree reconstruction. No combined, cellular resolution proteostasis measurement is known and since progressive weakening of protein quality control systems lead to extracellular aggregates to contributing to significant age-associated pathology, integrating loss of proteostasis with cell tree dynamics seems like a big scientific and technological challenge. Please see more on the objection related to extracellular matrix coverage later.

Deregulated nutrient sensing: this is one of the most actively researched hallmark process due to the intervention options and dramatic effects seen in model organisms, but in order to connect it to lineage tree information, single cell metabolite measurements, mentioned above, must be applied in an experimentally and methodologically creative way.

Mitochondrial dysfunction: Already mentioned above, in the downward, organellar scalability paragraph, how single cell mitochondrial measurements are already in use to deliver lineage histories in chronic myeloid leukemia, colon cancer and hematopoiesis.

Cellular senescence: Senescent cells are the ones that became dysfunctional, stopped proliferating yet do not clear up with apoptosis and show age-associated increase. Bulk sequencing studies already started to quantify the transcriptional heterogeneity of senescent cells (Hernandez-Segura et al., 2017). Single-cell DNA/RNA sequencing approaches track those cells too, but as far as I know they have not been particularly used yet to track the lineal positions and patterns of senescent cells in cell trees. One particular representational problem is that since senescent cells go into mitotic arrest they don't contribute anymore to further tree dynamics, not being parent cells of subtrees. But exactly this property might lead to uncover important changes in cell tree topology with time, brought upon by accumulating senescent cells, and with an added spatial layer help advice potential therapies.

Stem cell exhaustion: This is a prime domain for cell lineage trees, where the most can be learned from single cell sampling and single cell molecular profiling in terms of the declining regenerative potential of a tissue during aging. For example, well known age-related decline in neurogenesis in middle-aged mice (12-14 months) was clonally analysed in vivo to potentiate the contribution of stem cell division asymmetry and quiescence to the process (Bast et al., 2018). Stem cell quiescence and differentiated cellular senescence might present themselves as same extant leaves based on single-cell genomics data alone, so additional single cell molecular profiling might be needed to dissect the different cell populations going into mitotic arrest.

Altered intercellular communication: this hallmark process refers to diverse phenomena, but accumulating, under the threshold immune system misbehaviour features as perhaps the most relevant mechanism here, various aspects of it called immunosenescence and inflammaging. A murine, transcriptome based, aging cell atlas effort additionally computed tissue-specific aging trajectories and increase immune gene expression in old tissues pointed toward the increased infiltration of different immune cells (Kimmel et al.,2019). Spatiotemporal dynamics of immune cell infiltration coupled with immunological cell lineage tree patterns might emerge as one of the most relevant, ubiquitous contributors of the aging process.

Biomedical representative, expressive power, life history resolution

- It should provide near exhaustive coverage of life events, corresponding to biological and medical history.

Theoretically at any time point or time period of the life of an individual, multicellular organism corresponds to a particular cell tree structure, in case of time points, or temporal sequence (dynamics) of cell tree transitions from starting and ending tree structures, in case of time periods. This means that any disease that affects basic cellular structures (most, if not all disease do) will affect the shape of the cell tree. Disease onset and disease progression corresponds to possibly well-defined patterns of tree transitions. In case of cancer, which is a genuine somatic mutation driven disease delivering the temporal sequence of driver and passenger mutations in the form of an inferred cell lineage tree is a default task in the field.
As a hypothesis we can assume that each and every disease can be mapped onto a distinct cell tree pattern, as a particular tree dynamics, a sequence of transitions, forming at least an injective, one-to-one function. Although the inter-individual variation of such patterns are not known but this provides an opportunity to model/simulate chronic and acute diseases, eg. cardiovascular disease culminating in a myocardial infarction, simple or traumatic injuries, and even the process of dying due to a terminal disease ending in death. It's an open question whether we will ever develop accurate, sensitive enough and ethically unproblematic methods to trace human cellular lineages to provide us a good enough resolution to deliver these tasks. In case of a positive answer, these cell tree based encoding of medical histories can predict outcome and fine tune stratified, deeper treatment options and windows.

3. Objections against the centrality of tree dynamics in aging

Here we raise objections against the centrality of tree dynamics in aging and show some of the limits of this centrality while addressing these objections.

The first 3 objections are related to the obvious fact that the multicellular body is composed of not just cells but a complex and physiologically all too relevant extracellular space.

Extracellular Matrix

The human body (our main example of complex multicellular organisms) is way more than just the collection of self-organised cell populations. The intercellular space is filled up by different forms of extracellular matrix, ranging from well-defined molecular structures in solid tissues and organs to biological fluids. The molecular components of the ECM are secreted by different kind of cells.

The aging of the ECM is a separate research topic where convincing data accumulated for instance on collagen cross-linking due to advanced glycation end-products (AGEs). The SASP secretory phenotype of certain senescent cells is making a hostile tissue environment for instance by secreting matrix metalloproteinases in extra amounts, the main ECM enzymes digesting collagens for instance in normal tissue remodelling as well.

In the context of cell trees the obvious question is how a cell tree can account for the ECM in a quantitative manner. Theoretically it should be possible to assign through spatial coordinates and by knowing the local cell/ECM ratios a shadow spatial ECM network to the spatially interpreted cell tree, where the actual cell foliage can be mapped into corresponding ECM structures. The big question then how such a cell tree based mapping can help model the aging of the ECM. For instance, understanding extracellular aggregates through cell tree dynamics seems like a challenge. On the other hand, senescence-associated secretory **phenotype** (**SASP**) mainly manifests in the context of extracellular space by remodelling the surrounding ECM, so there's a strong connection between cellular aging phenomenon and tissue microenvironment, giving a reason to tie the two together stronger by modelling cell tree dynamics during aging.

Tissue microenvironments: stem cell niche

A more specific version of the extracellular objection focuses on stem cell niches, the tissue microenvironments that form the 'residential area' of stem and progenitor cell populations with all its local variables determining, affecting cellular behaviour. The objection is that cell tree modelling ignores these microenvironments, hence misses an important part of the regenerative story that might help us understand stem cell exhaustion and find candidates for interventions.

Part of the answer is similar to the ECM case above: if the stem cell niche spatially and anatomically is well defined like in the case of the intestinal or muscle stem cell niche then resident cellular elements can represent niches in a quantitative manner and labelling of the leaves with tissue microenvironment parameters can be incorporated into the cell tree models. In terms of modelling age associated stem cell depletions in a strong quantitative manner, cell trees are the way to go, and the niche concept seems just too abstract and not as computable. Also other systemic signals contributing to the plasticity or rigidity of stem cell pools can be taken into account easier with cell trees, than with a strictly local spatial concept.

Exosomes

Yet another version of the specific extracellular argument is the growing data on the relevance of so called microvesicles, exosomes. The objection is that stem cells exert a relevant part of their regenerative tissue potential via exosomes and microvesicles, subcellular membrane bubbles, shedded by and derived from stem cells, containing different molecular, protein, DNA factors contributing to regenerative mechanisms. Hence, cell trees ignore this mechanism. And this way they also miss out potential intervention options.

The answer is that an exosome centric view of regenerative medicine can hardly be central enough to capture the complexity of stem cell biology and it's only focusing on potential route of action, out of many. On the other hand, the cell tree concept can be extended into cellularly derived membrane vesicles, qualitatively accounting for those. In case these vesicles contain DNA, RNA (which they mostly do) then theoretically the 'lineage' histories of these vesicles can be delivered by sequencing them. This is not unlike, what I said concerning mitochondria above, the difference being that mitochondria represent intracellular networks, that can be overlaid upon cell trees, while exosomes are per definition, represent extracellular networks, derived from cells. The expressivity of the cell tree approached can be applied to capture vesicle mechanism, while it does not work the other way. Speciality of one particular process is not a sufficient reason to dismiss the arguments for the centrality of cell trees in biological aging.

Microbiome

Human bodies are made out of other living cellular things too beside all the downstream human offsprings cells of the zygote. The human microbiome is the collection of permanent guest bacteria, fungi (and viruses as non-cellular guest components) inhabiting the human host. This host and guest aggregate together is called a holobiont, and it is being increasingly acknowledged as a new unit of biological organisation (Haag, 2018), playing a role in health and disease. The estimated number of bacteria in a human 'reference person' is 38 trillion (Sender et al, 2018), a significantly larger number than the estimated 30 trillion human cells. The objection is that a human cell tree approach concerning health and aging is missing to capture the role the microbiome plays in these processes.

There are two answers to this objection. First, although it's probable that the role of the human microbiome in aging is much bigger than we currently understand, it is unlikely that future studies will fundamentally alter the central role of human cellular structures in this process, already discovered or yet to be discovered. More importantly, since bacteria undergo binary fission to carry out cell division, all the guest bacteria in the human body can be modelled with trillion different and small clonally expanded binary cell lineage trees. Fungi might undergo through open or closed mitosis. Eventually same or very similar cell tree principles are applicable, which preserves the centrality of the cell tree approach, although considers two main kinds of trees, the vast human cell tree structure (the big tree), and the vast amount smaller bacterial trees (the residents of the big tree). From the point of view of biological aging, these bacterial trees are not directly but embedded into the big tree.

Not directly age-associated, communicable diseases

The last objection states the limited medical expressibility of human cell lineage trees by highlighting that infectious diseases cannot lend themselves directly into cell tree mapping, as the main actors of the disease (bacteria, fungi, viruses) are not part of the human cell tree.

The first answer is that as far as cellular elements are concerned (bacteria, fungi), cell tree principles apply, just like mentioned in the case of the microbiome above.

The second, more important answer, is that since we can assume theoretically a unique mapping to every biologically meaningful and almost all chronologically relevant life events, all diseases, including communicable and not directly age-associated diseases, as mentioned in *Biomedical representative, expressive power, life history resolution* section above, the objection does not hold at all, as the distinctive power of cell trees covers the transitions and dynamics of the common flu for instance.

Centrality of the tree and tree dynamics from the point of view of biological aging does not mean that the approach should cover literally every phenomenon of interest, and provide a whole new level of explanation and mechanism. Yet, I tried to argue that theoretically it might be able do that nevertheless. Focusing on cell trees reinforces the centrality of cellularity in biology but all other levels can be accommodated.

4. Very brief history of cell lineage tree research

Here we summarise some landmark results, published up to 2017 on normal cell lineage trees.

The first complete and essentially invariant cell lineage tree of an animal representing 969 cells belonged to C. elegans and was delivered through watching and counting all relevant events (divisions) and cells with light microscopy (Sulston, Horvitz, 1977).

The real work on normal and variant cell lineage trees has been started in 2005-06 by two groups, one in Weizmann Institute of Science at Rehovot, in Israel and another in University of Washington, Seattle, in the USA. First, theoretical results have been delivered to show that based on known microsatellite mutation rates cell lineage trees can be confidently reconstructed from sampling leave cells in trees with mitotic depth of less than 40 divisions. The results have been used in human cell lines to reconstruct cell trees based on microstaellite instability using neighbour-joining algorithm (Frumkin et al, 2005). Second, polyguanine repeat DNA sequences have been used in cultured mouse NIH 3T3 cells to reconstruct cell trees with the help of both Bayesian and neighbour-joining algorithms (Salipante, Horwitz, 2006). The method was explicitly called phylogenetic fate mapping based on methods used, borrowed from phylogenetics. The breakthrough in vivo result has relied on bulk sequencing 25 organoid cell

lines of clonal, endodermal origin extracted from two old mice using 35 somatic base substitutions to reconstruct the cell lineage tree structures with maximum parsimony method (Behjati et al, 2014).

Meanwhile, a separate branch of research focused on dissecting intra-tumour heterogeneity, using bulk sequencing methods and delivering mutational signatures. Most of these tumour phylogenies are not strict cell lineage trees though (Alves et al., 2017).

5. Current cellular trajectory projects and initiatives trying to capture aging dynamics

Here we consider recent papers published in 2018-19 and a project in progress that might turn out to be relevant to understanding aging dynamics at a single cell resolution. There are different sources of potentially relevant research, from domain specific fields like cancer somatic evolution and developmental biology, to different methodological developments within computational biology and international projects like the Human Cell Atlas. High-throughput, largely single cell resolution biology is undergoing huge progress these days.

Somatic evolution started in tumour biology and intra-tumour heterogeneity eventually evolved into delivering proper cell lineage trees (Gaiti et al.,2019) and the results related to cancer-associated mutational accumulation with normal human esophagus with age (Martincorena et al.,2018).

Methodologically speaking, one crucial development is the so called diffusion pseudotime approach, abbreviated as DPT (Haghverdi et al.,2016). Single-cell profiling methods destroy the cells they analyse, hence the main challenge is to reliably estimate the temporal dynamics and ordering of cellular trajectories. Since hundred thousands of single cells are profiled in the latest experiments, robustness and scalability are added requirements for these methods. Diffusion pseudotime is a random-walk based Euclidean distance calculation in the ‘diffusion map space’ that delivers pseudotemporal ordering and identifies branching points by calculating expression similarity with a weighted nearest-neighbour graph approach.

One line of relevant research is developmental biology reimagined as single cell transcriptomics trajectories. A version of DPT, called the URD approach, used single-cell RNA-

seq (scRNA-seq) of ~40 000 cells during zebrafish embryogenesis to deliver developmental trajectories dubbed as developmental specification trees (Farrell et al 2018). Note that these snapshot DPT studies did not provide quantitative cell population estimates and hence cell flux along transcriptional trajectories can be confounded by cell population changes due to proliferation or death rates. A new mathematical and computational method, called pseudodynamics, finally made the first steps in the direction of dissecting transcriptional developmental trajectories from cell population dynamics and confidently establish these two components from scRNA-seq time series data (Fischer et al., 2019).

In terms of specific aging related studies, let's just mention two fresh ones. Single cell RNA-seq has been used on 50,000+ cells across threes tissues (kidney, lung, spleen) in four young and three old mice to track cell identity and tissue-specific aging trajectories (Kimmel et al.,2019), mentioned before already. To dissect the influence of cell identity from tissue specificity in aging trajectory, an embedded space was created by non-negative matrix factorisation to compare different trajectories by cosine similarity. The magnitude (acceleration and deceleration) of aging was assessed with optimal transport analysis. Another study used again murine scRNA-seq from 17 organs and plasma proteomics at 10 different ages to show consistent transcriptional shifts during aging (Schaum et al., 2019). From the point-of-view of comprehensive understanding of asynchronous aging approaches this kind of data is a good step in the right direction.

One important methodological, statistical development is the introduction of so called lineage variability maps capturing second-order variation in variable cell lineage trees (Hicks et al.,2019). Lineage variability maps make it possible to handle variability of individual linage maps, where lineage maps are no other than labelled lineage trees with phenotypes of the measured leave cells recorded. This technique will likely find its way into deciphering tree dynamics patterns relevant in late life.

Lastly, let's mention the Human Cell Atlas project, an international, big, Human Genome Project style consortium aiming to create comprehensive reference maps of all types of human cells to understand health and disease at a cellular resolution (humancellatlas.org). The name of the project, Atlas, primarily hints at a system of spatial coordinates aiming for a cellular level spatial understanding of histological and anatomical information. The term 'lineage tree' is mentioned only two times and the term 'aging' is not mentioned even once in the body of the

Human Cell Atlas' white paper (Regev et al.,2017). Nevertheless, in the paper, as a thought experiment, an imaginary Ultimate Cell Atlas turns out to be an all encompassing entity including temporality finally, highlighted at the third place in priorities in the following form as capturing 'every cell at every moment of a person's lifetime (by adding another dimension for time relating the cells by a lineage)' (Regev et al.,2017). Although temporality is not naturally a top priority of the Human Cell Atlas at this current stage, it seems probable that the methods and questions asked will develop into capturing dynamics of development and diseases. My personal estimation is that it might turn out to be the case that one of the main outcomes of the HCA in retrospect will be related to the understanding of biological aging and providing ample amount of raw data to assemble the first near complete human cell tree. This way it can serve as a fuel to a next big international scientific project, one focusing specifically on aging and healthy longevity. Indeed, Ehud Shapiro already envisioned a Human Cell Lineage Tree project back in 2010 (Shapiro, 2010) and some elements of this plan found their way into the Human Cell Atlas of which Shapiro is a consortium member. Let me credit here Shapiro with another important quote from a review paper (Shapiro et al.,2013).: 'Central unresolved problems in human biology and medicine are in fact questions about the human cell lineage tree: its structure, dynamics and variability during development, growth, renewal, ageing and disease.' As far as I know the idea of the central role of cell lineage trees and tree dynamics in aging has not been worked out in a detailed manner besides little hints like this. The current perspective piece can be considered as a way to motivate and unfold some of the avenues that can be taken to expose this idea and research program seriously and give proper focus to it.

The limiting challenge of counteracting biological aging(s) is comprehensive organismal rejuvenation.

Going back to our main argument, let's ask several, related scientific/technological questions to motivate the next point. What is the ultimate challenge of aging related interventions? What is the ultimate medical and technological target when it comes about keeping aging related functional decline at bay? How far can we go?

Is it the delay of biological aging in all levels and all compartments? Or is it systemic rejuvenation? Or to rephrase the question: where does maximum delay of aging ends and where does rejuvenation begin, where is the grey zone, what is the difference?

Answering these questions is currently not possible in a conceptually reliable, quantitatively predictive and experimentally productive way in the lack of system wise, comprehensive aging biomarkers, systemic computational models and longitudinal studies. But, by thinking on them we can zero in on the challenge.

The systemic breakdown of homeostasis and disrupting of the key balances that biological aging involves strongly suggests that the ultimate challenge lies in being able to comprehensively act upon all the key processes affected by it. The question is: what is needed to periodically (or near-continuously) synchronise diverse body aging rates across different tissues and organs in order to reach - again - prime biological state and maintain this state as long as possible? The problem is both finding the overlapping time windows to act and the spatial embeddedness of the structures to rejuvenate. How can a comprehensive aging regimen be sufficient for a sustained systemic one?

This is the first phrasing of the ultimate problem of organismal rejuvenation.

Why is this such a hard problem to crack?

First of all, please note that ultimate is suggestive here of an upper limit problem. But I also use the term 'limiting' challenge instead. I chose this usage of 'limiting' in order to highlight the relevance of both lower and upper limit problems of organismal rejuvenation. Limiting problems at this point seem to be linked with the complexity of aging, highlighted also by the plural aging(s) concept to indicate the diversity of the separate processes involved in this phenomenon, yet currently inextricably and densely connected, that we still recognise as one, from the point of view of the eventual outcome.

Phrasing limiting problems are not unlike finding the maxima and minima points (the so called extremes) of a mathematical function. Concerning the two kinds of limit problems, the upper limit problems of organismal rejuvenation can be thought of as finding the global maximum of the particular rejuvenation function, while lower limit problems can be viewed as finding the local minimum of the same function, at a different resolution and time period, and in a multidimensional space. Please note, that this is offered here as a restricted mathematical analogy but there's no intention to turn this into an actually applicable tool in a perspective piece.

So I don't specify how multidimensional this rejuvenation function should be or whether the y axis should be the inverse of biological age in order for the extremes of a function be more directly translatable to our problem. In what follows I discuss 5 type of problems, in the context of the hallmark framework.

(1a) A lower limit problem means the local minimal amount of interventions needed in one process to not block the effect of a rejuvenating intervention in another process. (1b) Alternatively, it can be that an intervention designed to act upon one subprocess of a hallmark process should also be enabled by another intervention acting upon another subprocess. (1a) A theoretical example of the earlier, more relevant lower limit problem, related to main hallmark processes, can be that for instance by not fixing the loss of proteostasis blocks the ongoing intervention against inflammaging and immunosenescence. Another example might be that without mitochondrial rejuvenation, fixing the skewed stem cell differentiation potential is not yielding any benefit in delaying stem cell exhaustion. (1b) An example of the local subprocess lower limit problem can be for instance that while the amount of stem cell quiescence and proliferative potential is being restored to a biologically younger status but the skewed differentiation potential of these stem cells is not fixed, for instance satellite cell numbers are returning to younger states but still tend to differentiate towards a fibrogenic lineage, instead of a myogenic one.

(1c) Making the minimal limiting problem worse is that not only separate interventions might depend on each to reach the target rejuvenation in just one process but these interventions might directly antagonise each other. Here, a perennial problem might be fixing proliferative potential of stem cells in local niches but also giving rise to more tumours. Same applies to fixing cellular senescence.

(2) In general, its unlikely that any magic bullet interventions are developed effective across several hallmark processes at the same intervention window.

(3) Another problem is that we don't know how comprehensive the list of hallmark aging processes are, it might be that some major processes are yet to be discovered. In the context of cell lineage trees it seems likely that with advanced knowledge a whole new class of problems will arise, that can not be simply fitted into the problem of stem cell exhaustion for instance, as they also explain new elements of cellular senescence, say.

(4) Related to this problem is how detailed our knowledge are on the component subprocesses within each process, and how new emerging details will modify our understanding of the process or help re-classify them. For instance, the misbehaviour of nutrient sensing pathways might be enriched with other types of, yet unknown, metabolic dysregulations with age.

Likely, new knowledge in the 2 problems related to lacking comprehensive details of the hallmark process framework will affect phrasing and solving both local minimum and global maximum problems.

(5) Finally I'd like to discuss global, upper limit problems, that was referred originally as the ultimate challenge. Here we will focus on stem cell exhaustion and review briefly some results (Schultz and Sinclair, 2016). The reason is that one global upper limiting problem is preservation and maintenance of regenerative potential across all tissues and this is the problem cell trees are most intimately seem to be related to in the first place.

The decline and function of local stem cell pools and niches around the body, most expressed in highly mitotic tissues, can be related to loss of stem cell numbers, altered proliferative potential and accompanying quiescence status, skewed differentiation potential and compromised engraftment and regeneration potential amongst others.

The most obvious upper limit problem, limiting healthy functioning of the body, is the exhaustion of the stem cell reservoirs by progressively declining numbers of stem cells. In the lack of enough normal stem cells tissue homeostasis gets severely compromised with time and might be predicative of eventual collapse. However attractive this theory, it only applies to just some adult stem cell niches, for instance the occurrence of satellite cells decreases with age (Schultz and Sinclair, 2016) and these cells are crucial for the regeneration of skeletal muscle fibres in response of injury. In several other stem cell pools the problem seems to be a compensatory increase of these cells at the price of declining differentiation and regenerative potential. For instance, hematopoietic (HSCs) and intestinal stem cell cells (ISCs) increase in number but yield skewed differentiation potential in case of HSCs, express more senescent markers, or display declining mitochondrial function in case of ISCs. Behind the frequency problem are problems with maintaining the protective quiescent states or the proper timing of the exit from dormant status to differentiate (Tümpel and Rudolph, 2019).

With the listing of these 5 problems I wanted to motivate why the limiting problem of organismal rejuvenation is comprehensive rejuvenation, both in the sense of local lower and global upper limit problems.

The central problem of comprehensive biological rejuvenation is to find an algorithm to restore the balance and maintain the healthy dynamics of the aging organismal cell lineage tree.

Now we are in a position to put together the two main premises of our macro-argument so far to formulate the central problem of the limiting challenge on counteracting biological agings(s) comprehensively.

In the first point I have argued that cell lineage trees are the central structure and cell tree dynamics is the central process of (understanding) organismal level, multicellular biological aging(s). Besides showing how cell trees meet a list of requirements to qualify for a central role in biological aging I also noted that I cannot think of any other structure and dynamics that could serve as the central layer and vehicle of overlaying all other phenotypic, genotypic information on top.

In the second point I have argued that the limiting challenge of counteracting biological aging(s) is comprehensive organismal rejuvenation. If this is the case then any kind of minimal solution to a lower limit problem will include designing interventions acting upon at least 2 processes or subprocesses. On the other hand any upper limit problem needs a robust, comprehensive, system wide engineering approach.

None of the current proposed, list based solutions deal with how to quantitatively combine all these technologies together to reach comprehensiveness. We simply just don't know, even in a theoretical sense, how far we can get on this road. No blueprint, no models, no strong quantitative handles, no simulations. But I think after following the argument so far we are in a position to delineate a central problem of comprehensive biological rejuvenation, that might prove workable. Workable as in there's lots of progress in the related fields that already shows the contours of a scientific and technological trajectory potentially leading to comprehensive, organismal, biological rejuvenation. It's like hypothetical reverse engineering: we know what the

end result should look like, and counterfactually we assume that such a product exits and from this position we are making our way back to look for the conditions to develop such a product, which in this case, is an intervention regimen to counteract biological aging(s). In short, we are looking for an algorithm that is continuously and asynchronously maintains, resets a biologically younger, already existing state comprehensively, throughout the organism, in all the relevant departments, in all the relevant time windows, acting upon all the relevant processes.

From point one and two it follows that the central problem of comprehensive biological rejuvenation is to find an algorithm to restore the balance and maintain the healthy dynamics of the aging organismal cell lineage tree.

My task here is not to formulate the general problem in a rigorous way, but to formulate it at all. Nevertheless we can substantiate the quantitative challenge a bit further in phrasing the problem in a quarter-mathematical way, with invoking some of the mathematical features of the cell lineage tree.

Let's assume one knows all the relevant properties of a dynamic, aging, mitotically active, at places and times overactive but in other positions also progressively shrinking cell lineage tree, including all the relevant properties of its organ, tissue level subtrees. The question can then be phrased the following way. What is the minimal, limiting set of replacement trees and how and when the new founder cells, giving rise to these replacement trees, should be triggered in the organism in order to form an aggregate tree, and a corresponding cumulative tree dynamics, made out of the original aging tree plus the triggered replacement trees that can exhibit the same, balanced properties as the original tree in a well-defined, previously existing healthy state or time window?

In order to be general and abstract enough for further mathematisation, please note that the term 'replacement trees' is used above in a source- and process-neutral way, so it can include all, internal and external, solutions regenerative medicine can offer. New founder cells establishing replacement trees are the substitute, proxy cells that play the role in maintaining a previous, healthier biological state, and these trees can be 'introduced' by both internally triggering the already existing stem and progenitor cell populations or can be actually introduced externally via cell therapy or even as a result of building a complex tissue engineering product. Due to centrality premise number one, the replacement trees can be designed to represent and factor in the hallmark aging processes and due to limiting premise two, the central comprehensive

rejuvenation task can represent, glue together and act upon any limiting combination of these relevant processes. For instance, translating the combined problem of senescent, malfunctioning and exhausted local stem cell niches into cell tree proportions, cell type distributions and ratios seems like a feasible and attractive task.

Next, I will substantiate the above proposition by anticipating a separate biomedical field emerging and mention some examples where progress with cell tree techniques in the broader sense has already been done or where re-phrasing the biomedical problem in terms of cell trees might be a realistic application medicine can benefit from in the shorter term.

The most comprehensive medical solution of biological aging(s) needs to use individual cell lineage trees as a central tool for diagnosis and treatment

This point is a corollary of the previous, spelling out some of the details, giving more specialised content to the more general proposition above in the context of biomedicine and medical practice.

The practical conclusion is that the central design/algorithm of the ultimate, personalised, most robust and comprehensive medical solution counteracting aging and interventions maintaining healthy longevity need to use and focus on individual cell lineage trees both as diagnostics and as therapeutics.

Any kind of stem cell based comprehensive regenerative medicine aiming to act upon at least two aging related processes as a lower limit problem, or on an upper limit problem needs to rely on the cell tree lineage and trajectory approach. This new emerging field of comprehensive regenerative medicine will be in the position to advance continuous, gradual and consecutive regeneration of most tissues and organs of the human body combining cell therapy, gene editing and tissue engineering. The reason am not calling it ‘systemic regenerative medicine’ instead is that the term ‘systemic’ in this context can only refer to approaches using the systemic circulation for therapeutic delivery.

What applications of cell trees within biotechnology, biomedicine or within actual medical practice might deliver results in the short term? The trajectory of the implementation should start with applications focusing on current, clinically classified diseases.

Let's start with the oldest, best known and most successful composite stem cell therapy, bone marrow transplantation. The two main forms of transplants are autologous or allogeneic transplants, the donor being the same person as the recipient in the former case, while the donor being a different person in the latter case. Re-phrasing it with the cell tree approach, autologous transplantation preserves the original binary cell lineage tree as the transplant cells are the direct offsprings of the donor, no matter whether they might spend some part of their cellular lives outside the body, undergoing different modifications, or whether they are the results of in vitro passages. In case of allogeneic transplantation, the new founder cells and their derivative replacement trees, to use the terms from the previous section, are not part of the original cell tree but after the successful transplant a new composite tree emerges, where the new bone marrow repopulates the ecological niche occupied by the older, depleted bone marrow. These composite trees form so called forests in a mathematical sense, the co-existing collection of disjoint binary trees.

Somatic mutations accumulate in clonal hematopoiesis with age, similarly to the somatic evolution results mentioned earlier related to human esophagus. Bulk whole-exome DNA sequencing data from blood of more than 17,000 people showed this phenomenon, observed eg. in 18.4% of cases between 90-108 years of age, associated with increase in all-cause mortality and different cardiovascular adverse events (Jaiswal et al., 2014).

This acquisition of mutations in HSCs is called 'clonal hematopoiesis of indeterminate potential' (CHIP) and recent theoretical and animal studies started to model and investigate how these age-related, pre-cancerous changes can affect both the safety and efficiency of bone marrow stem cell transplants, depending whether they are present in the donor or recipient or both (Park et al., 2019). According to my current knowledge, hight throughput, single cell transcriptomics studies have not been published yet delivering cell lineage trees and trajectories for bone marrow transplants. Since clonal composition is not part of the clinical assessment prior to autologous transplantation, it's easy to see that cell tree methods can offer a way to select safer and better transplants.

Generally to move from bone marrow transplantation to other regenerative stem cell therapies for other indications than leukemias, the benefits of standardising and personalising cell therapies with cell trees seems ripe for successful application.

Concerning cell therapy, a big question is to select a genetically heterogenous enough starter cell population for grafting, that can increase success of transplant. The same applies in tissue engineering when designing three dimensional products, where scaffolds need be populated with cells in strict ratios at different stages and spatial positions. On the other hand the use of different stem cells, eg. induced pluripotent stem cells, as disease models might also have diversity requirements that can be handled with cell tree approaches.

Lastly, intra-tumor heterogeneity is already a heavily researched field and reconstruction of cancer cell lineage trees based on somatic mutations have the potential to personalise cancer treatment by exposing the temporal sequence of how driver mutations have been introduced during the clonal development of the tumour. Cancer can be re-phrased as primarily an uncontrolled cell lineage tree disease, with accumulating mutations and proliferative potential gone wild. The biology of cancer and aging already has numerous and complicated connections and through cell lineage trees and trajectories cancer and aging research can form an important methodological alliance, where cancer research signals important ways ahead for geroscience.

Future prospects

Ehud Shapiro is quite pessimistic in terms of how quickly our knowledge on cell trees in the context of aging will match up with that of genomics: ‘Obtaining knowledge of the human cell lineage tree in development, aging, and disease on par with our current knowledge of the human genome will take decades’ (Shapiro, 2019). I think it is not going to take that long if there are concerted research efforts focusing on the role cell trees can play in delivering a comprehensive understanding of biological aging. One big enabler of results might turn out to be developing new mathematical concepts that can lead to new concrete metrics on how the changes of tree topology with age encode information on actual biological aging. Although the mathematics of binary trees is particularly well established and in phylogenetic methods is well accounted for, still the domain of systemic biological aging warrants a review of how deeply

current concept can capture comprehensive and local changes in tree topology at the same time. This fundamental conceptual job is not something that can be easily replaced with just applying new machine learning methods, but these two approaches need to be developed in a parallel manner and inform each other.

Acknowledgements

I am grateful to Alexandra Bause for suggesting the stem cell niche and exosome objections.

Literature

Alves JM, Prieto T, Posada D. Multiregional Tumor Trees Are Not Phylogenies. Trends Cancer. 2017 Aug;3(8):546-550. doi: 10.1016/j.trecan.2017.06.004.

Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, Martincorena I, Petljak M, Alexandrov LB, Gundem G, Tarpey PS, Roerink S, Blokker J, Maddison M, Mudie L, Robinson B, Nik-Zainal S, Campbell P, Goldman N, van de Wetering M, Cuppen E, Clevers H, Stratton MR. Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature. 2014 Sep 18;513(7518):422-425. doi: 10.1038/nature13448.

Bast L, Calzolari F, Strasser MK, Hasenauer J, Theis FJ, Ninkovic J, Marr C. Increasing Neural Stem Cell Division Asymmetry and Quiescence Are Predicted to Contribute to the Age-Related Decline in Neurogenesis. ***Cell Rep.*** 2018 Dec 18;25(12):3231-3240.e8. doi: 10.1016/j.celrep.2018.11.088.

Budnik B, Levy E, Harmange G, Slavov N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. ***Genome Biol.*** 2018 Oct 22;19(1):161. doi: 10.1186/s13059-018-1547-5.

Csordas A, Bouckaert R: Bayesian analysis of normal mouse cell lineage trees allowing intra-individual, cell population specific mutation rates. ***biorXiv***, doi: https://doi.org/10.1101/040733.

Duncan KD, Fyrestam J, Lanekoff I. Advances in mass spectrometry based single-cell metabolomics. ***Analyst***. 2019 Jan 28;144(3):782-793. doi:10.1039/c8an01581c.

EMBL-EBI Train Online. What is phylogenetics? https://www.ebi.ac.uk/training/online/course/introduction-phylogenetics/what-phylogenetics

Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. ***Science***. 2018 Jun 1;360(6392). pii: eaar3131. doi: 10.1126/science.aar3131.

Fischer DS, Fiedler AK, Kernfeld EM, Genga RMJ, Bastidas-Ponce A, Bakhti M, Lickert H, Hasenauer J, Maehr R, Theis FJ. Inferring population dynamics from single-cell RNA-sequencing time series data. Nat Biotechnol. 2019 Apr;37(4):461-468. doi: 10.1038/s41587-019-0088-0.

Frumkin D, Wasserstrom A, Kaplan S, Feige U, Shapiro E. Genomic variability within an organism exposes its cell lineage tree. ***PLoS Comput Biol.*** 2005 Oct;1(5):e50.

Gaiti F, Chaligne R, Gu H, Brand RM, Kothen-Hill S, Schulman RC, Grigorev K, Risso D, Kim KT, Pastore A, Huang KY, Alonso A, Sheridan C, Omans ND, Biederstedt E, Clement K, Wang L, Felsenfeld JA, Bhavsar EB, Aryee MJ, Allan JN, Furman R, Gnirke A, Wu CJ, Meissner A, Landau DA. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. ***Nature***. 2019 May;569(7757):576-580. doi: 10.1038/s41586-019-1198-z.

Haag KL. Holobionts and their hologenomes: Evolution with mixed modes of inheritance. ***Genet Mol Biol.*** 2018;41(1 suppl 1):189-197. doi: 10.1590/1678-4685-GMB-2017-0070.

Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. Nat Methods. 2016 Oct;13(10):845-8. doi:10.1038/nmeth.3971.

Hernandez-Segura A, de Jong TV, Melov S, Guryev V, Campisi J, Demaria M.Unmasking Transcriptional Heterogeneity in Senescent Cells. ***Curr Biol.*** 2017 Sep 11;27(17):2652-2660.e4. doi: 10.1016/j.cub.2017.07.033.

Hicks DG, Speed TP, Yassin M, Russell SM. Maps of variability in cell lineage trees. ***PLoS Comput Biol***. 2019 Feb 12;15(2):e1006745. doi:10.1371/journal.pcbi.1006745.

Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burtt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen HA, Ladenvall C, Getz G, Correa A, Banahan BF, Gabriel S, Kathiresan S, Stringham HM, McCarthy MI, Boehnke M, Tuomilehto J, Haiman C, Groop L, Atzmon G, Wilson JG, Neuberg D, Altshuler D, Ebert BL. Age-related clonal hematopoiesis associated with adverse outcomes. ***N Engl J Med.*** 2014 Dec 25;371(26):2488-98. doi: 10.1056/NEJMoa1408617.

Kimmel JC, Penland L, Rubinstein ND, Hendrickson DG, Kelley DR, Rosenthal AZ. A murine aging cell atlas reveals cell identity and tissue-specific trajectories of aging, ***biorXiv*** doi: https://doi.org/10.1101/657726

López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. ***Cell***. 2013 Jun 6;153(6):1194-217. doi: 10.1016/j.cell.2013.05.039.

Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, Pelka K, Ge W, Oren Y, Brack A, Law T, Rodman C, Chen JH, Boland GM, Hacohen N, Rozenblatt-Rosen O, Aryee MJ, Buenrostro JD, Regev A, Sankaran VG. Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. ***Cell***. 2019 Mar 7;176(6):1325-1339.e22. doi: 10.1016/j.cell.2019.01.022.

Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, Cagan A, Murai K, Mahbubani K, Stratton MR, Fitzgerald RC, Handford PA, Campbell PJ, Saeb-Parsy K, Jones PH. Somatic mutant clones colonize the human esophagus with age. ***Science***. 2018 Nov 23;362(6417):911-917. doi: 10.1126/science.aau3879.

Morris J, Na YJ, Zhu H, Lee JH, Giang H, Ulyanova AV, Baltuch GH, Brem S, Chen HI, Kung DK, Lucas TH, O'Rourke DM, Wolf JA, Grady MS, Sul JY, Kim J, Eberwine J. Pervasive within-Mitochondrion Single-Nucleotide Variant Heteroplasmy as Revealed by Single-Mitochondrion Sequencing. ***Cell Rep***. 2017 Dec 5;21(10):2706-2713. doi:10.1016/j.celrep.2017.11.031.

Park DS, Akuffo AA, Muench DE, Grimes HL, Epling-Burnette PK, Maini PK, Anderson ARA, Bonsall MB. Clonal hematopoiesis of indeterminate potential and its impact on patient trajectories after stem cell transplantation. ***PLoS Comput Biol***. 2019 Apr 26;15(4):e1006913. doi: 10.1371/journal.pcbi.1006913.

Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, Clevers H, Deplancke B, Dunham I, Eberwine J, Eils R, Enard W, Farmer A, Fugger L, Göttgens B, Hacohen N, Haniffa M, Hemberg M, Kim S, Klenerman P, Kriegstein A, Lein E, Linnarsson S, Lundberg E, Lundeberg J, Majumder P, Marioni JC, Merad M, Mhlanga M, Nawijn M, Netea M, Nolan G, Pe'er D, Phillipakis A, Ponting CP, Quake S, Reik W, Rozenblatt-Rosen O, Sanes J, Satija R, Schumacher TN, Shalek A, Shapiro E, Sharma P, Shin JW, Stegle O, Stratton M, Stubbington MJT, Theis FJ, Uhlen M, van Oudenaarden A, Wagner A, Watt F, Weissman J, Wold B, Xavier R, Yosef N; Human Cell Atlas Meeting Participants. The Human Cell Atlas. ***Elife***. 2017 Dec 5;6. pii: e27041. doi: 10.7554/eLife.27041.

Salipante SJ, Horwitz MS. Phylogenetic fate mapping. Proc Natl Acad Sci U S A. 2006 Apr 4;103(14):5448-53.

Schaum N, Lehallier B, Hahn B, Hosseinzadeh B, Lee SE, Sit R, Lee DP, Losada PM, Zardeneta ME, Pálovics R, Fehlmann T, Webber J, McGeever A, Zhang H, Berdnik D, Tan W, Zee A, Tan M, The Tabula Muris Consortium, Pisco A, Karkanias J, Neff NF, Keller A, Darmanis S, Quake SR, Wyss-Coray T. The murine transcriptome reveals global aging nodes with organ-specific phase and amplitude ***biorXiv,*** doi: https://doi.org/10.1101/662254

Schultz MB, Sinclair DA. When stem cells grow old: phenotypes and mechanisms of stem cell aging. Development. 2016 Jan 1;143(1):3-14. doi: 10.1242/dev.130633.

Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. ***PLoS Biol***. 2016 Aug 19;14(8):e1002533. doi:10.1371/journal.pbio.1002533.

Shapiro E. Uncovering the Human Cell Lineage Tree in Health and Disease, 2010 http://www.lineage-flagship.eu/uploads/file/Draft_Pilot_Flagship_Shapiro_Rehovot_v3.pdf

Shapiro, E., Biezuner, T., Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. ***Nat. Rev. Genet.*** 14(9), 618-30 (2013).

Shapiro E. On the journey from nematode to human, scientists dive by the zebrafish cell lineage tree. ***Genome Biol***. 2018 May 29;19(1):63.

Specht H, Emmott E, Koller T, Slavov N: High-throughput single-cell proteomics quantifies the emergence of macrophage heterogeneity. ***biorXiv***, doi**:** https://doi.org/10.1101/665307

Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev Biol. 1977 Mar;56(1):110-56. PubMed PMID: 838129.

Tümpel S, Rudolph KL. Quiescence: Good and Bad of Stem Cell Aging. Trends Cell Biol. 2019 Jun 24. pii:S0962-8924(19)30081-9. doi: 10.1016/j.tcb.2019.05.002.