PREMIS DE LA SECCIÓ DE CIÈNCIES BIOLÒGIQUES, IV

Animal cell type diversity, evolution and regulation

Arnau Sebé-Pedrós

Premi IEC de la Secció de Ciències Biològiques
Pius Font i Quer de Ciències de la Vida 2022

Institut d'Estudis Catalans
SECCIÓ DE CIÈNCIES BIOLÒGIQUES

Animal cell type diversity, evolution and regulation

PREMIS DE LA SECCIÓ DE CIÈNCIES BIOLÒGIQUES, IV

Animal cell type diversity, evolution and regulation

Arnau Sebé-Pedrós

Premi IEC de la Secció de Ciències Biològiques
Pius Font i Quer de Ciències de la Vida 2022

Barcelona, 2023

Institut d'Estudis Catalans
SECCIÓ DE CIÈNCIES BIOLÒGIQUES

Biblioteca de Catalunya. Dades CIP

Sebé-Pedrós, Arnau, autor
[Diversitat, evolució i regulació dels tipus de cel·lulars animals. Anglès]
Animal cell type diversity, evolution and regulation. — First edition. —
(Premis de la Secció de Ciències Biològiques ; 4)
Bibliografia. — Premi IEC de la Secció de Ciències Biològiques Pius Font i Quer de Ciències de la Vida 2022
ISBN 9788499656960
I. Institut d'Estudis Catalans. Secció de Ciències Biològiques. II. Títol
III. Col·lecció: Premis de la Secció de Ciències Biològiques. Anglès ; 4
1. Cèl·lules animals 2. Genòmica 3. Diferenciació cel·lular
575.113
576.3
576.36

© Arnau Sebé-Pedrós
© 2023, Institut d'Estudis Catalans, for this edition
Carrer del Carme, 47. 08001 Barcelona

First edition: January, 2023

Cover design: Azcunce | Ventura

Edited by Flor Edicions, SL

ISBN: 978-84-9965-696-0

DOI: 10.2436/10.1500.10.1

This work is free to use, but is subject to the terms of the Creative Commons public license. The work may be reproduced, distributed and communicated provided that its authorship and the entity that publishes it are acknowledged and that no commercial use or derivative work is made. A full copy of the terms of this license can be found at: *http://creativecommons.org/licenses/by-nc-nd/3.0/es/deed.ca.*

At the proposal of the working group formed by the Mmes. and the Messrs. Teresa Cabré i Castellví (president of the Institut d'Estudis Catalans), Àngel Messeguer i Peypoch (secretary-general of the Institut d'Estudis Catalans) and Ramon Bartrons i Bach, Anicet Ramon Blanch i Gisbert, Jordi Camí i Morell, Jordi Casanova i Roca and Cèlia Marrasé Peña (members of the Secció de Ciències Biològiques), at its session held on 23 March 2022, the Institut d'Estudis Catalans resolved to award the Premi IEC de la Secció de Ciències Biològiques Pius Font i Quer de Ciències de la Vida, to Arnau Sebé-Pedrós for his work entitled *Animal cell type diversity, regulation and evolution*.

In accordance with the terms and conditions of the award, the Secció de Ciències Biològiques is publishing this reviewed version of the winning research paper, entitled *Animal cell type diversity, evolution and regulation*.

Table of contents

Summary 9

1. The functional genomics revolution: from genome sequences to cell type gene regulation 10
2. Genome regulation and the origin of multicellularity and cell type differentiation 12
3. Cell types as the fundamental building blocks of animal multicellularity 16
4. Our current understanding of animal cell type diversity 18
5. From cell type transcriptional phenotypes to cell type evolution 20
6. Summary and outlook 22

References 23

Curriculum vitae of Arnau Sebé-Pedrós 36

Summary

A fundamental question in biology is how the myriad of specialized cell behaviors observed in a multicellular organism is encoded by a single genome sequence, and which gene regulatory mechanisms orchestrate the spatiotemporal deployment and maintenance of these cell type-specific programs. Decades of genetic and, more recently, genomic analyses have identified multiple layers of regulation, including specific molecular players and their interactions. This regulation ultimately results in cell-specific gene expression profiles that define coordinated cell cooperation at the tissue and whole-organism levels. However, the diversity and evolutionary dynamics of cell types remain almost completely unexplored beyond selected tissues in a few species, and so it is the gene regulatory networks that define them. Similarly, little is known about the emergence of complex genome regulatory mechanisms that support cell type-specific programs and cellular memory, including genome spatial compartmentalization and chromatin dynamics. The advent of low-input, highly-multiplexed epigenomic and transcriptomic profiling methods, even at single-cell resolution, and the broad applicability of these techniques to diverse systems, bypasses the need to obtain large amounts of biological material by culturing or dissecting particular tissues or cell types. This paves the way for the comparative multi-level analysis of cell differentiation in species and ontogenetic stages that span an unprecedented phylogenetic breadth and represent diverse levels of biological complexity: ranging from unicellular temporal differentiation dynamics and simple multicellular behaviors (as in some protistan eukaryotes), through loosely integrated and limitedly diversified ensembles of cell types (as in early-branching animals), to organisms with elaborate tissue and body plan organization (as in bilaterian animals). The comparative approach, through the

prism of evolutionary theory and supported by a robust phylogenetic framework, holds the promise of offering far-reaching insights into the fundamental principles that govern cell biological systems and the associated molecular mechanisms of genome regulation.

1. The functional genomics revolution: from genome sequences to cell type gene regulation

The first sequenced eukaryotic genome was that of the yeast Saccharomyces cerevisiae in 1996 (Goffeau et al., 1996), soon followed by the initial genome assemblies of the two major animal model species: the nematode *Caenorhabidits elegans* in 1998 (Consortium, 1998) and the arthropod *Drosophila melanogaster* in 2000 (Adams et al., 2000). The human draft genome arrived only a year later, in 2001 (Lander et al., 2001; Venter et al., 2001). These inaugurated two decades of progressive taxonomic expansion in genome sequencing (Dunn and Ryan, 2015), albeit with strong taxonomic biases (Del Campo et al., 2014). A good example of these efforts was the sequencing, by the Joint Genome Institute (JGI), of the first representatives of early-branching animal phyla, namely the sponge *Amphimedon queenslandica* (Srivastava et al., 2010), the cnidarian *Nematostella vectensis* (Putnam et al., 2007), and the placozoan *Trichoplax adhaerens* (Srivastava et al., 2008). This bonanza of genome sequences enabled comparative genomics to flourish, providing insights into microevolutionary (genetic variation, genome-wide association studies) and macroevolutionary processes (gene family evolution, origin of de novo genes, horizontal gene transfer events, gene synteny, ancient chromosomal linkage groups, and much more). Not only that, but genome sequencing completely transformed our understanding of animal phylogenetic relationships. This was thanks to phylogenomic analyses that emerged hand-in-hand with the development of novel phylogenetic reconstruction algorithms and elaborate substitution models (Kapli et al., 2020; Lartillot et al., 2007; Schrempf et al., 2019).

Genome sequencing across eukaryotes has recently accelerated with the launch of large-scale biodiversity genome sequencing initiatives (Blaxter et al., 2022; Lawniczak et al., 2022; Lewin et al., 2022), like the Earth Biogenome Project and the Darwin Tree of Life initiative (which aims to sequence all eukaryotic species in the United Kingdom). As part of this global effort, and with similar goals and standards, the Catalan Biogenome project has also recently started the sequencing of eukaryotic species of the Catalan territories (https://www.biogenoma.cat/). From these initiatives not only a quantitative and phylogenetic expansion of available genomes is expected, but the quality of these new generation genomes is far superior, for example, to the JGI genomes of 10 years ago. Indeed,

the advent of long-read sequencing methods (e.g. Oxford Nanopore) now allows us to accurately resolve highly-repetitive regions and, together with chromosomal contact mapping methods (see below), enable chromosome-scale assemblies, even telomere-to-telomere assemblies (Manuel et al., 2020; Nurk et al., 2022). Additional functional genomics methods like full-length cDNA sequencing and 5' RNA sequencing methods have allowed significant progress in what remains, arguably, the biggest challenge in genome sequencing: the prediction of gene models (transcriptional start and end sites, untranslated regions, specific coding sequencing, alternative isoforms, etc.).

While we are producing genomes at an unprecedented pace and across animal life, our understanding of how these genomes "function" is still extremely limited. For example, where and how is gene regulatory information stored? What kind of regulatory circuits underlie the diversity of tissues and cell types observed in multicellular animals? In truth, we can answer these questions to a large extent for only a handful of species (again, *C.elegans*, *Drosophila*, and humans). Just a few years after the completion of these genomes, large-scale projects like ENCODE, Blueprint, and modENCODE set out to catalog the transcribed RNA species and the genome-wide distribution of regulatory sequences and chromatin modifications (e.g., histone post-translational modifications typically associated with certain transcriptional states – like activation and repression). These projects sparked some controversy, accused of being a Panglossian quest to define a function for every single base in the genome through biochemical assays, completely disregarding any evolutionary and genetic data and the possibility of non-adaptive DNA sequences (Graur et al., 2013). But beyond these considerations, these initiatives created the context for the development of an impressive toolkit of functional genomics protocols: transcription and nascent transcription (e.g. GRO-seq, CAGE-seq), accessible chromatin (e.g. DNase-seq and ATAC-seq), histone modifications and variants (ChIP-seq and variants), DNA methylation (e.g. bisulfite sequencing) and chromosomal contacts (Hi-C and variants). More recently, many of these methods went single-cell: we can now measure these features quite accurately in individual cells, and this has paved the way to understanding cellular diversity and regulation at an unprecedented resolution (as we shall see below). Most importantly for the focus of this assay: these methodologies are amenable to non-model species. So now the conditions are set for starting a comparative study of genome function. This comparative regulatory genomics should reveal the diversity, evolution and shared principles of gene regulation, cell type and tissue identity programs, chromatin mechanisms, etc. and in a way analogous to the progress in understanding genome organization and evolution, which genome sequencing has made possible.

2. Genome regulation and the origin of multicellularity and cell type differentiation

A central question in evolutionary biology is how animal multicellularity and spatial cell differentiation originated (Brunet and King, 2017; Sebé-Pedrós et al., 2017). Cell types access distinct combinations of genetic elements (genes and *cis*-regulatory sequences) and this process results in unique and stable transcriptional programs that define the cellular phenotype. Cell type programs are implemented by both transcription factors (TFs) binding to specific *cis*-elements and by epigenomic mechanisms. Together, they establish and maintain cellular identity, while often also repressing alternative cell fates. Multiple epigenomic mechanisms participate in this specific genomic interpretation and studies of vertebrates, *Drosophila* and *C.elegans* have revealed common themes in genome regulation across bilaterian animals. For example, the widespread combinatorial usage of distal enhancer elements to control gene expression, or the existence of transcriptional states defined by conserved histone marks (Bonev and Cavalli, 2016; Ernst et al., 2011; Filion et al., 2010; Ho et al., 2014; Kvon et al., 2014). Another important regulatory layer is the spatial compartmentalization of the genome, although this remains unexplored beyond a handful of species (Bonev and Cavalli, 2016; Crane et al., 2015; Rao et al., 2014; Sexton et al., 2012). In addition, comparative genomic analyses have shown that gene innovation at the origin of animals was less extensive than previously thought (Brunet and King, 2017; Richter et al., 2018; Sebé-Pedrós et al., 2017; Suga et al., 2013). This led to the hypothesis that an important animal innovation was the ability to co-regulate these genes in multiple combinations, leading to spatial cell type specialization.

In particular, three specific epigenomic mechanisms have been suggested to be key for stable cell type differentiation: (*i*) genome spatial compartmentalization (Sebé-Pedrós et al., 2017; Tanay and Cavalli, 2013), as it allows the organization of modular gene regulatory neighborhoods that are independently regulated; (*ii*) large heterochromatic domains associated with repressive histone marks (Reinberg and Vales, 2018), as they maintain parts of the genome inaccessible to the cell and stabilize/restrict cell lineage commitment; and (*iii*) the emergence of combinatorial gene regulation by distal elements (as opposed to regulation by proximal promoter elements predominant in unicellular eukaryotes (Sebé-Pedrós et al., 2016; Tunnacliffe et al., 2018)), which allows elaborate spatiotemporal patterns of gene expression (Gaiti et al., 2017; Sebé-Pedrós et al., 2017, 2018b).

(i) Chromosomal physical architecture. Bilaterian animal genomes studied to date are organized into structural chromatin territories, such as topologically associating domains (TADs) (Bonev and Cavalli, 2016; Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). These physical chromatin domains restrict the

frequency of contacts of distal regulatory elements with promoters, thus defining gene regulatory landscapes. Recent high-resolution chromatin contact maps in *Drosophila* have revealed the existence of two distinct structural elements in the genome: insulators (which define boundaries of limited cross-domain contacts) and tethering elements (which use chromatin to bring folding promoters together with other promoters or with distal enhancer elements) (Batut et al., 2022; Levo et al., 2022). The discretization of the genome into broad structural domains allows the existence of autonomous regulatory blocks, with similar prevalent chromatin features (e.g. active or repressive TADs) and within which looping regulatory interactions can occur (Dixon et al., 2016; Tanay and Cavalli, 2013). CTCF is regarded as an essential factor for TAD formation, therefore the origin of CTCF in Bilateria might be related to a bilaterian-specific structural compartmentalization of the genome. Conversely, alternative structural proteins could be acting in the formation of chromatin domain structures in non-bilaterian animals. Thus, deciphering and comparing the three-dimensional genome architectures of non-bilaterian animals will be key to resolving the evolutionary link, if any, between the origin of animal multicellularity and the emergence of specific mechanisms for genome compartmentalization and folding.

(ii) Heterochromatin. Another conserved feature across animals is the existence of transcriptional states defined by similar chromatin modifications (Ho et al., 2014; Schwaiger et al., 2014; Sebé-Pedrós et al., 2016). For example, histone marks defining transcriptionally active genes are conserved in diverse non-metazoan eukaryotes, for example H3K4me3 and H3K27ac at promoters and H3K36me3 in gene bodies. This may suggest they are ancient eukaryotic regulatory features, although we do not know whether the reading/writing/erasing mechanisms for these marks are conserved. In contrast, less is known about the definition of heterochromatic regions across animals by repressive histone marks like H3K9me2/3 (constitutive) and H3K27me3 (facultative), although these marks are also found in non-metazoan eukaryotes (Grau-Bové et al., 2022). The existence of large, regulated heterochromatic domains has been hypothesized to be important for the emergence of differentiated multicellularity (Reinberg and Vales, 2018). Heterochromatic domains maintain parts of the genome inaccessible to the cell and this would help to restrict cell lineage commitment and stabilize cell identity. But we do not know what the genome-wide distribution of different types of heterochromatin in most animal lineages is and, more importantly, which specific mechanisms (modifying enzymes, binding proteins, etc.) mediate heterochromatin function. Hence, a better understanding of heterochromatin function across animals is needed to test to what extent this represents an evolutionary regulatory novelty.

A similar repressive role is mediated by DNA methylation in vertebrates. But this modification shows a sparse phylogenetic presence/absence in metazoans

and significant differences in the genome-wide distribution (Mendoza et al., 2020). Many invertebrates present gene body methylation associated with active gene transcription, as well as methylation of silenced repetitive elements (a pattern sometimes called "mosaic" methylation). In contrast, multiple animals have completely lost DNA methylation (given the absence of DNMT1/3 enzymes encoded in their genomes) (Mendoza et al., 2020). An extreme case among metazoans is found in the vertebrate lineage, showing genome-wide high levels of methylation, with only active/accessible regulatory regions (promoters and enhancers) devoid of methylation. Interestingly, it has recently been reported that this widespread methylation pattern is also found in the demosponge *Amphimedon queenslandica*, suggesting the convergent evolution of hypermethylation in animals (Mendoza et al., 2019). In summary, while diverse chromatin modifications and their genome-wide distribution appears conserved across animals, in most cases we do not know whether the writers/erasers and the regulatory readout of these marks are conserved.

(iii) Distal enhancer regulation. Transcription factors bind to specific sequences located at gene promoters and, at least in bilaterian animals, distal enhancer elements. Enhancers are clusters of TF binding sites with specific chromatin characteristics such as depletion of nucleosomes (open chromatin sites) and particular histone marks (H3K4me1 and H3K27ac) in the flanking nucleosomes (Andersson et al., 2014; Corces et al., 2016; Heintzman et al., 2009; Rada-Iglesias et al., 2011; Thurman et al., 2012). The presence of p300, a histone acetyltransferase of holozoan origin (Sebé-Pedrós et al., 2011), has also been used to predict enhancer regions and activity (Visel et al., 2009). High-throughput approaches to identify and validate enhancer candidates and test their functions have shown that, in bilaterian animals, most enhancer elements are distal (kilobases up to megabases) to the gene promoters they regulate, and that they act through physical looping of the chromatin (Deng et al., 2012; Jin et al., 2013; Shlyueva et al., 2014). This chromatin looping is mediated by CTCF, cohesin and other structural proteins (Phillips-Cremins et al., 2013; Schmidt et al., 2010). Enhancer elements generally reside in intergenic regions and, in more compact genomes, in introns, often of genes neighboring those regulated by the enhancers. The estimated number of enhancers is in the order of thousands in animals such as *Drosophila* (Kvon et al., 2014) or humans (Andersson et al., 2014). Moreover, in *Drosophila* the vast majority of enhancers show very restricted spatial and temporal activity during development (Bonn et al., 2012; Kvon et al., 2014), emphasizing the importance of enhancer elements in orchestrating complex regulatory states. Another defining feature of *cis*-regulatory enhancer elements is their combinatorial nature and modularity: multiple binding sites occur in each enhancer (Schwarzer and Spitz, 2014) and regulatory states are generated by the

combined action of multiple enhancers on the same gene, especially in genes encoding transcription factors and other developmental regulators (Ernst et al., 2011; Levine, 2010). Overall, the combined action of both distal enhancers and, to a lesser extent, proximal promoter *cis*-regulatory elements underlie the complex spatiotemporal expression patterns observed during bilaterian development and cell type differentiation.

Although the evolutionary dynamics of enhancers have been extensively studied in some bilaterians (Villar et al., 2015), the existence of such regulatory elements in other metazoan or premetazoan lineages has remained a mystery. An indirect hint to the possible existence of distal regulation across all metazoans is the deep evolutionary conservation of microsyntenic blocks across Metazoa (Irimia et al., 2012, 2013). These blocks comprise a gene (usually a developmental gene such as one encoding a TF or signaling protein) that is linked to another functionally unrelated neighboring bystander gene. This linkage is probably due to the presence of regulatory elements in the bystander gene. Interestingly, these blocks are not present in the unicellular relatives of animals (Irimia et al., 2012), suggesting that distal regulation evolved at the root of Metazoa. The first direct experimental evidence for the evolutionary conservation of the epigenetic regulatory landscape beyond Bilateria came from the landmark study of the model cnidarian *Nematostella vectensis* (Schwaiger et al., 2014). Approximately 6,000 enhancers were predicted in *Nematostella*, showing similar chromatin signatures (H3K4me1, K3K27ac and presence of the histone acetyltransferase p300) to those of bilaterian enhancers. Confirming these predictions, 12 of these *Nematostella* enhancer elements showed activity in *in vivo* reporter assays. Moreover, *Nematostella* enhancers were found to be particularly enriched close to TF genes, suggesting the existence of complex TF combinatorial regulatory networks. What remains unknown, however, is whether these *Nematostella* enhancers work through chromatin looping (through tethering elements) or non-looping proximity mechanisms. The latter is suggested by the absence of CTCF, a key protein for chromatin looping, in *Nematostella* and other non-bilaterian animals (Heger et al., 2012). Nevertheless, a non-looping proximity mechanism is inconsistent with the lack of dependence of enhancer activity on enhancer orientation or relative position to the promoter in *Nematostella* reporter assays (Schwaiger et al., 2014). Moreover, the more ancient cohesin complex (present in all animals and most eukaryotes) seems to be key to enhancer looping (Phillips-Cremins et al., 2013; Schmidt et al., 2010), and it has recently been shown that several different structural proteins, but notably not CTCF, are associated with enhancer–promoter chromatin loops in *Drosophila* (Cubeñas-Potts et al., 2016; Eagen et al., 2017). Therefore, it is possible that enhancer–promoter looping in *Nematostella* occurs, even in the absence of CTCF,

with looping mediated by cohesin and/or other structural proteins instead. More recently, enhancer elements have been reported in even earlier branching lineages such as sponges and ctenophores (Gaiti et al., 2017; Sebé-Pedrós et al., 2018a). In contrast, no distal regulatory elements are found in close extant unicellular relatives of animals (Sebé-Pedrós et al., 2016), suggesting that distal regulation was an important feature associated with the emergence of animal multicellularity. Distal enhancers would not only allow us to encode additional regulatory information, but to use it in a modular manner by employing different combinations of these elements (it remains to be proven whether this is the case in some of these invertebrate animal lineages), resulting in more complex patterns of gene expression. It is worth noting that among non-metazoan eukaryotes, distal regulatory elements have been also identified in diverse plant species (Lu et al., 2019; Ricci et al., 2019). This suggests the convergent origin of enhancers in this lineage and reinforces the association between distal regulation and differentiated multicellularity.

3. Cell types as the fundamental building blocks of animal multicellularity

Specialized cells represent the fundamental level of organization in multicellular organisms (Arendt, 2008). The morphological and molecular regularities observed in cells inspired an analogy to the diversity of organisms and their hierarchical arrangement into different taxa. This analogy suggested cellular taxonomy could be developed following similar principles to those underlying Linnaean classification: discreteness, reproducibility and hierarchy. Indeed, we can characterize a cell type as a discrete entity, with unique morphological and functional properties. And we can require a cell type to be reproducible, that is, to emerge stably across generations through embryonic development. Nevertheless, the hierarchical nature of cell types and the discrete nature of their classification remain more elusive: clearly ontogeny and cell lineages within organisms are major cell type organizing forces, but these are remodeled continuously by the plasticity and pleiotropy of gene regulatory programs across tissues. Since cell types are natural building blocks bridging molecular (gene level) and organismal (phenotypic) evolution, it is of great interest to facilitate their study as evolutionary units (Arendt et al., 2016). To this end, molecular profiling tools, particularly single-cell transcriptomics, hold the promise of bringing cell type molecular phenotyping and classification to non-model species, building systematic atlases of cells in different animal lineages. Single-cell atlases will not only advance our understanding of the molecular and cellular biology of understudied animal groups, but are the necessary first step towards a comparative biology of cell type

programs. Only through these cell type comparisons can we eventually understand and reconstruct cell type evolutionary dynamics.

Cell type classification schemes vary in their granularity and in the degree of phylogenetic and anatomical generalization. Indeed, classifications may encompass only particular organs/species or represent phylogenetically and anatomically (even organism-level) wider frameworks (Willmer, 1970). Most proposed cell type classifications are hierarchical and, with few exceptions (Xia and Yanai, 2019), use concepts and jargon borrowed from taxonomy (*clades*, *lineages*, *trees*, etc.). However, they do not explicitly consider or attempt to convey evolutionary relationships between cell types (Schwartz et al., 2020). From a historical perspective, the first efforts in cell type classification were based on the morphology, spatial tissue arrangement, and histological staining properties of cells. Using this information, multiple attempts were made in the pre-genomics era to develop global cell classification schemes (Willmer, 1970), to systematically characterize cell types in specific taxa (Baguñá and Romero, 1981; Bode et al., 1973; Simpson, 1984), and to use cell type number as a proxy for organismal complexity (Valentine et al., 1994). These classification frameworks were restricted in resolution and could not incorporate functional or ontogenetic considerations that are not readily represented morphologically. The advent of molecular profiling tools extended the ability to characterize, identify and classify cell types. Common strategies include the detection of specific proteins using antibody-based immunostaining (surface markers are still widely used for the molecular phenotyping of hematopoietic and immune cells (Novershtern et al., 2011)) or specific transcripts profiled using RNA *in situ* hybridization (more rarely also by qPCR analysis (Hirano et al., 2013)). While immunostaining is strongly constrained by the limited availability of antibodies, *in situ* hybridization with custom synthesized probes has enabled the rapid extension of molecular fingerprinting to a wide diversity of organisms, becoming a cornerstone of modern Evo-Devo studies (Ogino et al., 2011; Steinmetz et al., 2012; Tessmar-Raible et al., 2007). A major limitation of these expression profiling tools is the need to define a priori the set of gene markers to be studied and the limited scalability to dozens of markers. Molecular fingerprinting strategies have been very effective when employed in the comparative study of embryogenesis and tissue/organ-level anatomical structures (Martín-Durán et al., 2018; Sacerdot et al., 2018). Still, mapping the diversity of cell types across species in a truly systematic fashion has so far not been feasible.

A natural extension of candidate marker gene profiling is the analysis of genome-wide cell type gene expression using bulk transcriptomics. Pioneering studies in cell type bulk transcriptomics have provided systematic cell type classification schemes (Breschi et al., 2020), enabled the explicit analysis of cell type transcriptome hierarchical structure (Liang et al., 2015), and enabled the first

attempts at building phylogenies of closely related cell types based on their bulk gene expression profiles (Kin et al., 2015). However, these enrichment strategies are necessarily limited to cell lines (Breschi et al., 2020; Brown et al., 2014; Cherbas et al., 2011; Forrest et al., 2014) – virtually non-existent in the vast majority of organisms – or homogeneous cell populations isolated manually or by FACS-sorting (Alié et al., 2015; Novershtern et al., 2011; Sogabe et al., 2019), requiring dedicated methods with no guarantee of purity. High-throughput single-cell RNA sequencing methods (scRNA-seq) overcome many of these limitations, ultimately facilitating the unbiased sampling and molecular characterization of thousands of single cells, and setting the stage for *in silico* reconstruction of cell type repertoires in species that have so far been difficult to study.

Overall, the development of cell type classification tools is in a way analogous to that of species phylogenetics methods: from morphological to molecular characters (nucleotide or protein sequences). Just as it is difficult to resolve species phylogenies using morphological characters alone, only with molecular data can we aim to develop phylogenetically-inclusive cell type classification schemes. But the analogy ends here: modern taxonomy is explicitly based on a species' underlying evolutionary history, very often incorporating molecular phylogenetics as a key tool for classification. Cell type taxonomies remain, to date, fundamentally Linnaean.

4. Our current understanding of animal cell type diversity

Since the first proof-of-concept scRNA-seq studies in the early 2010s (Islam et al., 2011; Jaitin et al., 2014; Tang et al., 2009), we have witnessed the rapid proliferation of scRNA-seq analyses, with an ever-growing number of cells and moving from descriptive cell type phenomenologies to perturbation assays, development and temporal differentiation dynamics, and spatial transcriptomics with single-cell resolution (Marx, 2021). Today, cataloging the full repertoire of cell type programs in human tissues and development seems within reach (Cao et al., 2020; Regev et al., 2017) and important progress has already been made in cataloguing mouse cell types (Han et al., 2018; Schaum et al., 2018). Applied to non-traditional model species, whole-organism scRNA-seq methods should pave the way for the systematic characterization and comparison of cell types across the animal and, consequently, to rapidly advance in our understanding of cell type diversity, development and evolution (Marioni and Arendt, 2017).

Given a minimally biased single-cell sampling strategy, we can employ standardized pipelines to generate gene expression profiles for thousands of cells and to group such profiles into discrete, highly-similar transcriptional cell states.

These data-driven cell groups/clusters constitute basic units that can be further developed, through biological interpretation, into cell type classification schemes. Following up on the phylogenetics analogy, scRNA-seq methodologies can have an impact on the study of cell type diversity and evolution analogous to that of whole-genome/transcriptome sequencing techniques on the resolution of the animal tree of life.

From a taxonomic perspective, whole-organism cell type atlases are currently available for seven major animal groups (in most cases represented by a single species), including a ctenophore (Sebé-Pedrós et al., 2018a), two sponges (Musser et al., 2019; Sebé-Pedrós et al., 2018a), a placozoan (Sebé-Pedrós et al., 2018a), four cnidarians (Chari et al., 2021; Hu et al., 2020; Sebé-Pedrós et al., 2018b; Siebert et al., 2019), an acoel (Duruz et al., 2021), craniates (considering mouse whole-organ single-cell transcriptomes (Cao et al., 2019; Han et al., 2018)) and platyhelminths (Fincher et al., 2018; Li et al., 2021; Plass et al., 2018). In addition, tissue-specific single-cell atlases are already available for model species, including multiple *Drosophila* (Arthorpoda) datasets (Allen et al., 2020; Croset et al., 2018; Davie et al., 2018; Hung et al., 2020; Rust et al., 2020; Slaidina et al., 2020) and *Caenorhabiditis* (Nematoda) neuronal single-cell analysis. In addition, embryonic and larval stages have been sampled in two sea urchin species (Echinodermata) (Foster et al., 2020; Massri et al., 2020), in the marine polychaete *Platynereis dumerilii* (Annelida) (Achim et al., 2018), in the tunicate *Ciona intestinalis* (Horie et al., 2018) and again both in *D. melanogaster* and *C. elegans* (Karaiskos et al., 2017; Packer et al., 2019). The most densely sampled lineage is that of vertebrates, including developmental single-cell atlases in four species (human, mouse, zebrafish and *Xenopus*) (Briggs et al., 2018; Cao et al., 2019, 2020; Farrell et al., 2018; Pijuan-Sala et al., 2019; Wagner et al., 2018) and brain single-cell data (in most cases for specific brain regions) for several mammals, reptiles and teleosts (Hodge et al., 2019; Shafer et al., 2021; Tosches et al., 2018).

It is interesting to compare the phylogenetic expansion of single-cell atlases to that of reference genomes over the past five years. In 2015, almost two decades after the publication of the first animal draft genomes, Dunn and Ryan (Dunn and Ryan, 2015) reviewed the status of genome sequencing across animal lineages. By then, 212 genomes from 14 animal phyla were available at NCBI. Today, there are 11,404 genomes from 27 animal phyla available or in progress, with many more expected in the near future in the context of large-scale biodiversity sequencing initiatives like the Darwin Tree of Life or the Vertebrate genomes project (Koepfli et al., 2015). Importantly, taxon sampling biases persist, with 86% of these genomes coming from vertebrates (6,454) and arthropods (3,383) – 95% if we include mollusks (518) and nematodes (253). In comparison, since the publication of the first high-throughput single-cell transcriptomics datasets

circa 2015 (Jaitin et al., 2014; Klein et al., 2015; Pollen et al., 2014; Zeisel et al., 2015), whole-organism single-cell atlases have been published for 12 non-model animal species. Given the fast pace in the scale and sophistication of single-cell methods in model species (and the wide availability of commercial solutions), the current taxonomic sampling of single-cell atlases across the animal tree of life seems rather modest.

5. From cell type transcriptional phenotypes to cell type evolution

Beyond cataloging cells in different organisms, the key challenge to understanding the evolution of cell types is how to compare them in order to assess whether they share a common evolutionary origin (i.e. homology) (Arendt, 2008; Arendt et al., 2016; Wagner, 2014). Complex characters like cell types or organs have multiple components, often with non-coherent evolutionary histories (Liebeskind et al., 2016; Shubin et al., 2009). This complicates evolutionary inferences, especially at large evolutionary distances (Arendt, 2008; Hejnol and Lowe, 2014; Liebeskind et al., 2016). Indeed, despite obvious phenotypic (and even molecular) similarities between cell types across animal lineages, it is still highly uncertain whether (a) these similar cell types (e.g. neurons, stem cells, muscle fibers, or epithelial cells) are homologous and were therefore present in the common ancestor of all animals or, by contrast, (b) if they arose independently in different lineages by co-opting similar effector genes (i.e. homoplasy). Ontogenetic trajectories are also often a poor indicator of evolutionary affinity at large distances as often the same cell types emerge convergently from distinct progenitors or germ layers, and vice versa (e.g. Wagner et al., 2018).

An alternative to comparing whole cell type transcriptomes is to focus on the expression of combinations of transcription factors (TFs), often called terminal selectors (Hobert, 2008), as the key regulators of cell identity programs (Arendt et al., 2016). The implicit assumption is that TFs can represent a good proxy for the gene modules employed in that particular cell type. A second underlying hypothesis when focusing on TFs is that regulatory similarities have fewer evolutionary constraints and therefore better approximate cell type homology (compared to effector gene usage). However, we still don't know enough regarding the frequency at which cell identity TFs can be replaced, especially given the intricate evolutionary history of TF gene families. Newly derived cell type atlases highlight the role of TFs from large multi-gene families (e.g. zf-C2H2, Ets or Sox TFs). Such atlases also uncover multiple expressed paralogs sharing very similar DNA-binding characteristics (Lambert et al., 2019; Nitta et al., 2015; Weirauch et al., 2014). In addition, often tens of different TFs are

expressed in a particular cell type (e.g. in sponge choanocytes), making it difficult to determine which of them are the upstream drivers of cell type identity and what the evolutionary significance is of the conservation of a few of those TFs. However, although this represents an excellent theoretical framework to try to model cell type evolution, we have very limited data to systematically implement and assess the limits of this regulatory blueprint comparison and also to understand which regulatory features (e.g.TF-gene module associations, TF binding motifs, TF expression, etc.) can be maximally informative and at which evolutionary distances.

TFs control gene expression by recognizing and binding to short sequence motifs (6–12 bps) located at *cis*-regulatory regions (promoters and enhancers) of downstream genes (Spitz and Furlong, 2012). Cell type transcriptional identity is strongly recapitulated by sequence motif enrichment (Sebé-Pedrós et al., 2018b, 2018a), representing the *cis*-regulatory embedding of the cell type program (Hobert, 2008). In addition, with a few exceptions (like zf-C2H2 TFs), the binding sequences of TFs are very often conserved across large phylogenetic distances (Lambert et al., 2019; Sebé-Pedrós et al., 2013; Weirauch et al., 2014). This opens the possibility of comparing cell types not through their gene expression profiles but instead through the set of regulatory sequences defining the cell type program. A recent study pioneered the idea of cross-species *cis*-regulatory sequence comparison (Minnoye et al., 2020). Working on melanoma cell lines in different vertebrate species, Minnoye et al. uncovered a highly-conserved *cis*-regulatory program, involving a combination of four TF binding (SOX10, TFAP2A, MITF and ETS) motifs in enhancer regions that most often showed little or no global sequence conservation. Building on this dissection of melanoma enhancer motif syntax, the authors further modeled the effect of evolutionary mutations in enhancer function, as defined by accessibility.

The different elements that constitute a cell type gene expression program (TFs, effector genes, TF binding sites, regulatory connections, etc.) do not necessarily have congruent evolutionary histories (Shubin et al., 2009; Tschopp and Tabin, 2017), just as gene trees are not always in agreement with species trees (as a result of horizontal gene transfers, incomplete lineage sorting, etc.). By combining sequence motif analysis with high-resolution chromatin accessibility data (Vierstra et al., 2020), we should be able to systematically reconstruct cell type gene regulatory networks in non-model species. Disentangling and comparing regulatory programs in multiple closely related species will enable the development of quantitative models of cell type evolution, including evolutionary rates of distinct regulatory characters: TF usage/replacement, sequence motifs, regulatory interactions, gene module composition (Feregrino and Tschopp, 2021), and more. These models should constitute the basis of future cell type phylogenetics

and will help address important questions in cell type evolution: are these evolutionary rates universally conserved (Carvunis et al., 2015) or are there particularly "fast-evolving" cell type programs? How robust are cell type genetic networks and which components are particularly evolvable? Finally, identifying slow-evolving regulatory characters (e.g. sequence motifs) could help formulate better cell type homology hypotheses.

6. Summary and outlook

Whole-organism single-cell transcriptomics holds the promise of developing comprehensive catalogs of cell types in phylogenetically diverse systems. Reference cell type molecular atlases will crucially advance our understanding of the biology and evolutionary history of unsampled animal groups (Dunn et al., 2015). This is similar to how the sequencing and annotation of genomes of unsampled animal lineages uncovers novel biology and enables phylogenomics and comparative genomics studies. The most immediate challenge will be to develop methodological standards to build cell type maps in the most unbiased and consistent manner. Only with dense and technically-compatible phylogenetic sampling will we be able to start a systematic comparative study of cell type programs. Based on this sampling, cell type comparative biology will enable the development of cell type phylogenetic models and can increase understanding of the genetic changes associated with cellular novelty. Overall, this can offer transformative insights linking classic models of (genomic) molecular evolution with an intermediate molecular phenotype: cell types and their associated gene regulatory networks.

Beyond this, the major limiting factor for this comparative study of cell programs is the lack of regulatory genome data on which to model these cell identity programs in the vast majority of species. This limitation can be overcome thanks to the advent of additional single-cell and functional genomics tools. While single-cell transcriptomics allows the molecular characterization of cells, by combining this phenomenology of cell types with chromatin profiling and *in vitro* TF binding motif analysis (Jolma et al., 2013; Weirauch et al., 2014), we can go one step further and characterize the *cis*-regulatory landscape (the set of accessed *cis*-regulatory sites) and genetic interactions underlying these cell types (Stegle et al., 2015; Tanay and Regev, 2017). Importantly, these methodologies are particularly suited to studying non-standard model species, as they bypass the need to obtain large amounts of biological material and require very limited experimental handling. These unique characteristics pave the way for the comparative multi-level analysis of cell type regulation in metazoan species that span an unprecedented phylogenetic breadth.

References

Achim, K., Eling, N., Vergara, H.M., Bertucci, P.Y., Musser, J., Vopalensky, P., Brunet, T., Collier, P., Benes, V., Marioni, J.C., et al. 2018. "Whole-Body Single-Cell Sequencing Reveals Transcriptional Domains in the Annelid Larval Body." *Molecular Biology and Evolution* 35: 1047–1062.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. "The Genome Sequence of *Drosophila* melanogaster." *Science* 287: 2185–2195.

Alié, A., Hayashi, T., Sugimura, I., Manuel, M., Sugano, W., Mano, A., Satoh, N., Agata, K., and Funayama, N. 2015. "The ancestral gene repertoire of animal stem cells." *Proceedings of the National Academy of Sciences* 112: E7093–E7100.

Allen, A.M., Neville, M.C., Birtles, S., Croset, V., Treiber, C.D., Waddell, S., and Goodwin, S.F. 2020. "A single-cell transcriptomic atlas of the adult *Drosophila* ventral nerve cord." *Elife* 9: 1–32.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. 2014. "An atlas of active enhancers across human cell types and tissues." *Nature* 507: 455–461.

Arendt, D. 2008. "The evolution of cell types in animals: emerging principles from molecular studies." *Nature Reviews Genetics* 9: 868–882.

Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D., et al. 2016. "The origin and evolution of cell types." *Nature Reviews Genetics* 17: 744–757.

Baguñá, J., and Romero, R. 1981. "Quantitative analysis of cell types during growth, degrowth and regeneration in the planarians Dugesia mediterranea and Dugesia tigrina." *Hydrobiologia* 84: 181–194.

Batut, P.J., Bing, X.Y., Sisco, Z., Raimundo, J., Levo, M., and Levine, M.S. 2022. "Genome organization controls transcriptional dynamics during development." *Science* 375: 566–570.

Blaxter, M., Archibald, J.M., Childers, A.K., Coddington, J.A., Crandall, K.A., Di Palma, F., Durbin, R., Edwards, S. V., Graves, J.A.M., Hackett, K.J., et al. 2022. "Why sequence all eukaryotes?" *Proceedings of the National Academy of Sciences U. S. A.* 119.

Bode, H., Berking, S., David, C.N., Gierer, A., Schaller, H., and Trenkner, E. 1973. "Quantitative analysis of cell types during growth and morphogenesis in Hydra." *Wilhelm Roux. Archiv für Entwicklungsmechanik der Organismen* 171: 269–285.

Bonev, B., and Cavalli, G. 2016. "Organization and function of the 3D genome." *Nature Reviews Genetics* 17: 661–678.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Pérez-González, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. 2012. "Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development." *Nature Genetics* 44: 148–156.

Breschi, A., Muñoz-Aguirre, M., Wucher, V., Davis, C.A., Garrido-Martín, D., Djebali, S., Gillis, J., Pervouchine, D.D., Vlasova, A., Dobin, A., et al. 2020. "A limited set of transcriptional programs define major cell types." *Genome Research* 30: 1047–1059.

Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., and Klein, A.M. 2018. "The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution." *Science* 5780: eaar5780.

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., et al. 2014. "Diversity and dynamics of the *Drosophila* transcriptome." *Nature* 512: 1–7.

Brunet, T., and King, N. 2017. "The Origin of Animal Multicellularity and Cell Differentiation." *Developmental Cell* 43: 124–140.

Campo, J. del, Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., and Ruiz-Trillo, I. 2014. "The others: our biased perspective of eukaryotic genomes." *Trends in Ecology & Evolution* 29: 252–259.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. 2019. "The single-cell transcriptional landscape of mammalian organogenesis." *Nature* 566: 496–502.

Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R., Zhang, F., et al. 2020. "A human cell atlas of fetal gene expression." *Science* 370: eaba7721.

Carvunis, A.-R., Wang, T., Skola, D., Yu, A., Chen, J., Kreisberg, J.F., and Ideker, T. 2015. "Evidence for a common evolutionary rate in metazoan transcriptional networks." *Elife* 4: 1–22.

Chari, T., Weissbourd, B., Gehring, J., Ferraioli, A., Leclère, L., Gao, F., Chevalier, S., Copley, R.R., Houliston, E., Anderson, D.J., et al. 2021. "Whole Animal Multiplexed Single-Cell RNA-Seq Reveals Plasticity of Clytia Medusa Cell Types." *BioRxiv* 14–18.

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B.D., Carlson, J.W., Landolin, J.M., Kapranov, P., Dumais, J., et al. 2011. "The transcriptional diversity of 25 *Drosophila* cell lines." *Genome Research* 21: 301–314.

Consortium, *C. elegans* S. 1998. "Genome sequence of the nematode *C. elegans*: A platform for investigating biology." *Science* 282: 2012–2018.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. 2016. "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution." *Nature Genetics* 48: 1193–1203.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. 2015. "Condensin-driven remodelling of X chromosome topology during dosage compensation." *Nature* 523: 240–244.

Croset, V., Treiber, C.D., and Waddell, S. 2018. "Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics." *Elife* 7: 1–31.

Cubeñas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P., and Corces, V.G. 2016. "Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture." *Nucleic Acids Research* 45: 39–53.

Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, Ł., Aibar, S., Makhzami, S., Christiaens, V., Bravo González-Blas, C., et al. 2018. "A Single-Cell Transcriptome Atlas of the Aging *Drosophila* Brain." *Cell* 174: 982-998.e20.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G. A. 2012. "Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor." *Cell* 149: 1233–1244.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. 2012. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." *Nature* 485: 376–380.

Dixon, J.R., Gorkin, D.U., and Ren, B. 2016. "Chromatin Domains: The Unit of Chromosome Organization." *Molecular Cell* 62: 668–680.

Dunn, C.W., and Ryan, J.F. 2015. "The evolution of animal genomes." *Current Opinion in Genetics & Development* 35: 25–32.

Dunn, C.W., Leys, S.P., and Haddock, S.H.D. 2015. "The hidden biology of sponges and ctenophores." *Trends in Ecology and Evolution* 30: 282–291.

Duruz, J., Kaltenrieder, C., Ladurner, P., Bruggmann, R., Martínez, P., and Sprecher, S.G. 2021. "Acoel Single-Cell Transcriptomics: Cell Type Analysis of a Deep Branching Bilaterian." *Molecular Biology and Evolution* 38: 1888–1904.

Eagen, K.P., Lieberman Aiden, E., and Kornberg, R.D. 2017. "Polycomb-Mediated Chromatin Loops Revealed by a Sub-Kilobase Resolution Chromatin Interaction Map." *BioRxiv*.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. 2011. "Mapping and analysis of chromatin state dynamics in nine human cell types." *Nature* 473: 43–49.

Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., and Schier, A.F. 2018. "Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis." *Science* 3131: eaar3131.

Feregrino, C., and Tschopp, P. 2021. "Assessing evolutionary and developmental transcriptome dynamics in homologous cell types." *Developmental Dynamics* dvdy.384.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., Castro, I.J. de, Kerkhoven, R.M., Bussemaker, H.J., et al. 2010. "Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells." *Cell* 143: 212–224.

Fincher, C.T., Wurtzel, O., De Hoog, T., Kravarik, K.M., and Reddien, P.W. 2018. "Cell type transcriptome atlas for the planarian Schmidtea mediterranea." *Science* 360: eaaq1736.

Forrest, A.R.R., Kawaji, H., Rehli, M., Kenneth Baillie, J., De Hoon, M. J.L., Haberle, V., Lassmann, T., Kulakovskiy, I. V., Lizio, M., Itoh, M., et al. 2014. "A promoter-level mammalian expression atlas." *Nature* 507: 462–470.

Foster, S., Oulhen, N., and Wessel, G. 2020. "A single cell RNA sequencing resource for early sea urchin development." *Development* 147: dev191528.

Gaiti, F., Jindrich, K., Fernández-Valverde, S.L., Roper, K.E., Degnan, B.M., and Tanurdžić, M. 2017. "Landscape of histone modifications in a sponge reveals the origin of animal *cis*-regulatory complexity." *Elife* 6: 1–33.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. 1996. "Life with 6000 Genes." *Science* 274: 546–567.

Grau-Bové, X., Navarrete, C., Chiva, C., Pribasnig, T., Antó, M., Torruella, G., Galindo, L.J., Lang, B.F., Moreira, D., López-Garcia, P., et al. 2022. "A phylogenetic and proteomic reconstruction of eukaryotic chromatin evolution." *Natural Ecology & Evolution* 6: 1007–1023.

Graur, D., Zheng, Y., Price, N., Azevedo, R.B.R., Zufall, R.A., and Elhaik, E. 2013. "On the Immortality of Television Sets: "Function" in the Human Genome According to the Evolution-Free Gospel of ENCODE." *Genome Biology & Evolution* 5: 578–590.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. 2018. "Mapping the Mouse Cell Atlas by Microwell-Seq." *Cell* 172: 1091-1107.e17.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. 2012. "The chromatin insulator CTCF and the emergence of metazoan diversity." *Proceedings of the National Academy of Sciences* 109: 17507–17512.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. 2009. "Histone modifications at human enhancers reflect global cell-type-specific gene expression." *Nature* 459: 108–112.

Hejnol, A., and Lowe, C.J. 2014. "Animal evolution: Stiff or squishy notochord origins?" *Current Biology* 24: R1131–R1133.

Hirano, M., Guo, P., McCurley, N., Schorpp, M., Das, S., Boehm, T., and Cooper, M.D. 2013. "Evolutionary implications of a third lymphocyte lineage in lampreys." *Nature* 501: 435–438.

Ho, J.W.K., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M.Y., Appert, A., et al. 2014. "Comparative analysis of metazoan chromatin organization." *Nature* 512: 449–452.

Hobert, O. 2008. "Regulatory logic of neuronal diversity: terminal selector genes and selector motifs." *Proceedings of the National Academy of Sciences U. S. A.* 105: 20067–20071.

Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. 2019. "Conserved cell types with divergent features in human versus mouse cortex." *Nature* 573: 61–68.

Horie, R., Hazbun, A., Chen, K., Cao, C., Levine, M., and Horie, T. 2018. "Shared evolutionary origin of vertebrate neural crest and cranial placodes." *Nature* 560: 228–232.

Hu, M., Zheng, X., Fan, C.-M., and Zheng, Y. 2020. "Lineage dynamics of the endosymbiotic cell type in the soft coral Xenia." *Nature* 582: 534–538.

Hung, R.-J., Hu, Y., Kirchner, R., Liu, Y., Xu, C., Comjean, A., Tattikota, S.G., Li, F., Song, W., Ho Sui, S., et al. 2020. "A cell atlas of the adult *Drosophila* midgut." *Proceedings of the National Academy of Sciences* 117: 1514–1523.

Irimia, M., Tena, J.J., Alexis, M.S., Fernández-Miñán, A., Maeso, I., Bogdanovic, O., Calle-Mustienes, E. de la, Roy, S.W., Gómez-Skarmeta, J.L., Fraser, H.B., et al. 2012. "Extensive conservation of ancient microsynteny across metazoans due to *cis*-regulatory constraints." *Genome Research* 22: 2356–2367.

Irimia, M., Maeso, I., Roy, S.W., and Fraser, H.B. 2013. "Ancient *cis*-regulatory constraints and the evolution of genome architecture." *Trends in Genetics* 29: 521–528.

Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lonnerberg, P., and Linnarsson, S. 2011. "Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq." *Genome Research* 21: 1160–1167.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., et al. 2014. "Massively

parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types." *Science* 343: 776–779.

Jin, F., Li, Y., Dixon, J.R.J., Selvaraj, S., Ye, Z., Lee, A.Y.A., Yen, C.-A., Schmitt, A.D., Espinoza, C. A, and Ren, B. 2013. "A high-resolution map of the three-dimensional chromatin interactome in human cells." *Nature* 503: 290–294.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. 2013. "DNA-binding specificities of human transcription factors." *Cell* 152: 327–339.

Kapli, P., Yang, Z., and Telford, M.J. 2020. "Phylogenetic tree building in the genomic age." *Nature Reviews Genetics* 21: 428–444.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. 2017. "The *Drosophila* Embryo at Single Cell Transcriptome Resolution." *Science* 3235: 117382.

Kin, K., Nnamani, M.C., Lynch, V.J., Michaelides, E., and Wagner, G.P. 2015. "Cell-type Phylogenetics and the Origin of Endometrial Stromal Cells." *Cell Reports* 10: 1398–1409.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. 2015. "Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells." *Cell* 161: 1187–1201.

Koepfli, K.-P., Paten, B., and O'Brien, S.J. 2015. "The Genome 10K Project: A Way Forward." *Annual Review of Animal Biosciences* 3: 57–111.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. 2014. "Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*." *Nature* 512: 91–95.

Lambert, S.A., Yang, A.W.H., Sasse, A., Cowley, G., Albu, M., Caddick, M.X., Morris, Q.D., Weirauch, M.T., and Hughes, T.R. 2019. "Similarity regression predicts evolution of transcription factor sequence specificity." *Nature Genetics* 51: 981–989.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. "Initial sequencing and analysis of the human genome." *Nature* 409: 860–921.

Lartillot, N., Brinkmann, H., and Philippe, H. 2007. "Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model." *BMC Evolutionary Biology* 7: S4.

Lawniczak, M.K.N., Durbin, R., Flicek, P., Lindblad-Toh, K., Wei, X., Archibald, J.M., Baker, W.J., Belov, K., Blaxter, M.L., Marquès-Bonet, T., et al. 2022. "Standards recommendations for the Earth BioGenome Project." *Proceedings of the National Academy of Sciences* 119: e2115639118.

LEVINE, M. 2010. "Transcriptional enhancers in animal development and evolution." *Current Biology* 20: R754-63.

LEVO, M., RAIMUNDO, J., BING, X.Y., SISCO, Z., BATUT, P.J., RYABICHKO, S., GREGOR, T., and LEVINE, M.S. 2022. "Transcriptional coupling of distant regulatory genes in living embryos." *Nature* 605: 754–760.

LEWIN, H.A., RICHARDS, S., LIEBERMAN AIDEN, E., ALLENDE, M.L., ARCHIBALD, J.M., BÁLINT, M., BARKER, K.B., BAUMGARTNER, B., BELOV, K., BERTORELLE, G., et al. 2022. "The Earth BioGenome Project 2020: Starting the clock." *Proceedings of the National Academy of Sciences* 119: e2115635118.

LI, P., NANES SARFATI, D., XUE, Y., YU, X., TARASHANSKY, A.J., QUAKE, S.R., and WANG, B. 2021. "Single-cell analysis of Schistosoma mansoni identifies a conserved genetic program controlling germline stem cell fate." *Nature Communications* 12: 485.

LIANG, C., FORREST, A.R.R., and WAGNER, G.P. 2015. "The statistical geometry of transcriptome divergence in cell-type evolution and cancer." *Nature Communications* 6: 6066.

LIEBESKIND, B.J., HILLIS, D.M., ZAKON, H.H., and HOFMANN, H. A. 2016. "Complex Homology and the Evolution of Nervous Systems." *Trends in Ecology & Evolution* 31: 127–135.

LU, Z., MARAND, A.P., RICCI, W.A., ETHRIDGE, C.L., ZHANG, X., and SCHMITZ, R.J. 2019. "The prevalence, evolution and chromatin signatures of plant regulatory elements." *Nature Plants* 5: 1250–1259.

MARIONI, J.C., and ARENDT, D. 2017. "How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology." *Annual Review of Cell and Developmental Biology* 33: 537–553.

MARTÍN-DURÁN, J.M., PANG, K., BØRVE, A., LÊ, H.S., FURU, A., CANNON, J.T., JONDELIUS, U., and HEJNOL, A. 2018. "Convergent evolution of bilaterian nerve cords." *Nature* 553: 45.

MARX, V. 2021. "Method of the Year 2020: spatially resolved transcriptomics." *Nature Methods* 18: 1–1.

MASSRI, A.J., GREENSTREET, L., AFANASSIEV, A., ESCOBAR, A.B., WRAY, G.M., SCHIEBINGER, G., and MCCLAY, D.R. 2020. "Developmental Single-cell transcriptomics in the Lytechinus variegatus Sea Urchin Embryo." *BioRxiv* 2020.11.12.380675.

MENDOZA, A. de, LISTER, R., and BOGDANOVIC, O. 2020. "Evolution of DNA Methylome Diversity in Eukaryotes." *Journal of Molecular Biology* 432: 1687–1705.

MENDOZA, A. de, HATLEBERG, W.L., PANG, K., LEININGER, S., BOGDANOVIC, O., PFLUEGER, J., BUCKBERRY, S., TECHNAU, U., HEJNOL, A., ADAMSKA, M., et al. 2019. "Convergent evolution of a vertebrate-like methylome in a marine sponge." *Nature Ecology & Evolution* 3: 1464–1473.

Minnoye, L., Taskiran, I.I., Mauduit, D., Fazio, M., Van Aerschot, L., Hulselmans, G., Christiaens, V., Makhzami, S., Seltenhammer, M., Karras, P., et al. 2020. "Cross-species analysis of enhancer logic using deep learning." *Genome Research* 30: 1815–1834.

Musser, J.M., Schippers, K.J., Nickel, M., Mizzon, G., Kohn, A.B., Pape, C., Hammel, J.U., Wolf, F., Liang, C., Hernández-Plaza, A., et al. 2019. "Profiling cellular diversity in sponges informs animal cell type and nervous system evolution." *BioRxiv* 1–36.

Nitta, K.R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen, J., Deplancke, B., Furlong, E.E.M., et al. 2015. "Conservation of transcription factor binding specificities across 600 million years of bilateria evolution." *Elife* 4: 1–20.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. 2012. "Spatial partitioning of the regulatory landscape of the X-inactivation centre." *Nature* 485: 381–385.

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. 2011. "Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis." *Cell* 144: 296–309.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V, Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. 2022. "The complete sequence of a human genome." *Science* 376: 44–53.

Ogino, K., Tsuneki, K., and Furuya, H. 2011. "Distinction of cell types in Dicyema japonicum (phylum Dicyemida) by expression patterns of 16 genes." *Journal of Parasitology* 97: 596–601.

Packer, J.S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J., et al. 2019. "A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution." *Science* 365: eaax1971.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.T., Hookway, T. A., Guo, C., Sun, Y., et al. 2013. "Architectural protein subclasses shape 3D organization of genomes during lineage commitment." *Cell* 153: 1281–1295.

Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. 2019. "A single-cell molecular map of mouse gastrulation and early organogenesis." *Nature* 566: 490–495.

Plass, M., Solana, J., Wolf, F.A., Ayoub, S., Misios, A., Glažar, P., Obermayer, B., Theis, F.J., Kocks, C., and Rajewsky, N. 2018. "Cell type atlas

and lineage tree of a whole complex animal by single-cell transcriptomics." *Science* 1723: eaaq1723.

Pollen, A.A, Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P., et al. 2014. "Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex." *Nature Biotechnology* 32: 1053–1058.

Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V., et al. 2007. "Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization." *Science* 317: 86–94.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. 2011. "A unique chromatin signature uncovers early developmental enhancers in humans." *Nature* 470: 279–283.

Rao, S.S.P., Huntley, M.H., Durand, N.C., and Stamenova, E.K. 2014. "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping." *Cell* 159: 1–16.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. 2017. "The human cell atlas." *Elife* 6: 1–30.

Reinberg, D., and Vales, L.D. 2018. "Chromatin domains rich in inheritance." *Science* 361: 33–34.

Ricci, W.A., Lu, Z., Ji, L., Marand, A.P., Ethridge, C.L., Murphy, N.G., Noshay, J.M., Galli, M., Mejía-Guerra, M.K., Colomé-Tatché, M., et al. 2019. "Widespread long-range *cis*-regulatory elements in the maize genome." *Nature Plants* 5: 1237–1249.

Richter, D.J., Fozouni, P., Eisen, M.B., and King, N. 2018. "Gene family innovation, conservation and loss on the animal stem lineage." *Elife* 7: 1–43.

Rosa, P.M.G. de la, Thomson, M., Trivedi, U., Tracey, A., Tandonnet, S., and Blaxter, M. (2020). *A telomere to telomere assembly of* Oscheius tipulae *and the evolution of rhabditid nematode chromosomes*. BioRxiv 2020. 05.15.097428.

Rust, K., Byrnes, L.E., Yu, K.S., Park, J.S., Sneddon, J.B., Tward, A.D., and Nystul, T.G. 2020. "A single-cell atlas and lineage analysis of the adult *Drosophila* ovary." *Nature Communications* 11: 5628.

Sacerdot, C., Louis, A., Bon, C., Berthelot, C., and Roest Crollius, H. 2018. "Chromosome evolution at the origin of the ancestral vertebrate genome." *Genome Biology* 19,: 166.

Schaum, N., Karkanias, J., Neff, N.F., May, A.P., Quake, S.R., Wyss-Coray, T., Darmanis, S., Batson, J., Botvinnik, O., Chen, M.B., et al. 2018. "Sin-

gle-cell transcriptomics of 20 mouse organs creates a Tabula Muris." *Nature* 562: 367–372.

Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., and Odom, D.T. 2010. "A CTCF-independent role for cohesin in tissue-specific transcription." *Genome Research* 20: 578–588.

Schrempf, D., Lartillot, N., and Szöllősi, G. 2019. "Scalable empirical mixture models that account for across-site compositional heterogeneity." *Mol. Biol. Evol.* 37: 3616–3631.

Schwaiger, M., Schonauer, A., Rendeiro, a. F., Pribitzer, C., Schauer, A., Gilles, a. F., Schinko, J.B., Renfer, E., Fredman, D., and Technau, U. 2014. "Evolutionary conservation of the eumetazoan gene regulatory landscape." *Genome Research* 24: 639–650.

Schwartz, G.W., Zhou, Y., Petrovic, J., Fasolino, M., Xu, L., Shaffer, S.M., Pear, W.S., Vahedi, G., and Faryabi, R.B. 2020. "TooManyCells identifies and visualizes relationships of single-cell clades." *Nature Methods* 17: 405–413.

Schwarzer, W., and Spitz, F. 2014. "The architecture of gene expression: integrating dispersed *cis*-regulatory modules into coherent regulatory domains." *Current Opinion in Genetics & Development* 27: 74–82.

Sebé-Pedrós, A., Mendoza, A. de, Lang, B.F., Degnan, B.M., and Ruiz-Trillo, I. 2011. "Unexpected repertoire of metazoan transcription factors in the unicellular holozoan Capsaspora owczarzaki." *Molecular Biology and Evolution* 28: 1241–1254.

Sebé-Pedrós, A., Ariza-Cosano, A., Weirauch, M.T., Leininger, S., Yang, A., Torruella, G., Adamski, M., Adamska, M., Hughes, T.R., Gómez-Skarmeta, J.L., et al. 2013. "Early evolution of the T-box transcription factor family." *Proceedings of the National Academy of Sciences* 110: 16050–16055.

Sebé-Pedrós, A., Ballaré, C., Parra-Acero, H., Chiva, C., Tena, J.J., Sabidó, E., Gómez-Skarmeta, J.L., Di Croce, L., and Ruiz-Trillo, I. 2016. "The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity." *Cell* 165: 1224–1237.

Sebé-Pedrós, A., Degnan, B.M., and Ruiz-Trillo, I. 2017. "The origin of Metazoa: a unicellular perspective." *Nature Reviews Genetics* 18: 498–512.

Sebé-Pedrós, A., Chomsky, E., Pang, K., Lara-Astiaso, D., Gaiti, F., Mukamel, Z., Amit, I., Hejnol, A., Degnan, B.M., and Tanay, A. 2018a. "Early metazoan cell type diversity and the evolution of multicellular gene regulation." *Nature Ecology & Evolution* 2: 1176–1188.

Sebé-Pedrós, A., Saudemont, B., Chomsky, E., Plessier, F., Mailhé, M.-P., Renno, J., Loe-Mie, Y., Lifshitz, A., Mukamel, Z., Schmutz, S., et al.

2018b. "Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq." *Cell* 173: 1520-1534.e20.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. 2012. "Three-dimensional folding and functional organization principles of the *Drosophila* genome." *Cell* 148: 458–472.

Shafer, M.E.R., Sawh, A.N., and Schier, A.F. 2021. "Gene family evolution underlies cell type diversification in the hypothalamus of teleosts." *BioRxiv*.

Shlyueva, D., Stampfel, G., and Stark, A. 2014. "Transcriptional enhancers: from properties to genome-wide predictions." *Nature Reviews Genetics* 15: 272–286.

Shubin, N., Tabin, C., and Carroll, S. 2009. "Deep homology and the origins of evolutionary novelty." *Nature* 457: 818–823.

Siebert, S., Farrell, J.A., Cazet, J.F., Abeykoon, Y., Primack, A.S., Schnitzler, C.E., and Juliano, C.E. 2019. "Stem cell differentiation trajectories in Hydra resolved at single-cell resolution." *Science* 365: eaav9314.

Simpson, T.L. 1984. *The Cell Biology of Sponges* (New York, NY: Springer New York).

Slaidina, M., Banisch, T.U., Gupta, S., and Lehmann, R. 2020. "A single-cell atlas of the developing *Drosophila* ovary identifies follicle stem cell progenitors." *Genes & Development* 34: 239–249.

Sogabe, S., Hatleberg, W.L., Kocot, K.M., Say, T.E., Stoupin, D., Roper, K.E., Fernandez-Valverde, S.L., Degnan, S.M., and Degnan, B.M. 2019. "Pluripotency and the origin of animal multicellularity." *Nature* 570: 519–522.

Spitz, F., and Furlong, E.E.M. 2012. "Transcription factors: from enhancer binding to developmental control." *Nature Reviews Genetics* 13: 613–626.

Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., et al. 2008. "The Trichoplax genome and the nature of placozoans." *Nature* 454: 955–960.

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., et al. 2010. "The Amphimedon queenslandica genome and the evolution of animal complexity." *Nature* 466: 720–726.

Stegle, O., Teichmann, S.A., and Marioni, J.C. 2015. "Computational and analytical challenges in single-cell transcriptomics." *Nature Reviews Genetics* 16: 133–145.

Steinmetz, P.R.H., Kraus, J.E.M., Larroux, C., Hammel, J.U., Amon-Hassenzahl, A., Houliston, E., Wörheide, G., Nickel, M., Degnan, B.M., and Technau, U. 2012. "Independent evolution of striated muscles in cnidarians and bilaterians." *Nature* 487: 231–234.

Suga, H., Chen, Z., Mendoza, A. de, Sebé-Pedrós, A., Brown, M.W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N., et al. 2013. "The Capsaspora genome reveals a complex unicellular prehistory of animals." *Nature Communications* 4: 1–9.

Tanay, A., and Cavalli, G. 2013. "Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization." *Current Opinion in Genetics & Development* 23: 197–203.

Tanay, A., and Regev, A. 2017. "Scaling single-cell genomics from phenomenology to mechanism." *Nature* 541: 331–338.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. 2009. "mRNA-Seq whole-transcriptome analysis of a single cell." *Nature Methods* 6: 377–382.

Tessmar-Raible, K., Raible, F., Christodoulou, F., Guy, K., Rembold, M., Hausen, H., and Arendt, D. 2007. "Conserved Sensory-Neurosecretory Cell Types in Annelid and Fish Forebrain: Insights into Hypothalamus Evolution." *Cell* 129: 1389–1400.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. 2012. "The accessible chromatin landscape of the human genome." *Nature* 489: 75–82.

Tosches, M.A., Yamawaki, T.M., Naumann, R.K., Jacobi, A.A., Tushev, G., and Laurent, G. 2018. "Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles." *Science* 360: 881–888.

Tschopp, P., and Tabin, C.J. 2017. "Deep homology in the age of next-generation sequencing." *Philosophical Transactions of the Royal Society B: Biological Sciences 372*, 20150475.

Tunnacliffe, E., Corrigan, A.M., and Chubb, J.R. 2018. "Promoter-mediated diversification of transcriptional bursting dynamics following gene duplication." *Proceedings of the National Academy of Sciences*: 201800943.

Valentine, J.W., Collins, A.G., and Meyer, C.P. 1994. "Morphological complexity increase in metazoans." *Paleobiology* 20: 131–142.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. "The Sequence of the Human Genome." *Science* 291: 1304–1351.

Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., et al. 2020. "Global reference mapping of human transcription factor footprints." *Nature* 583: 729–736.

Villar, D., Berthelot, C., Flicek, P., Odom, D.T., C., Aldridge, S., Rayner, T.F., Lukk, M., and Pignatelli, M. 2015. "Enhancer Evolution across 20 Mammalian Species." *Cell* 160: 554–566.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. 2009. "ChIP-seq accurately predicts tissue-specific activity of enhancers." *Nature* 457: 854–858.

Wagner, G.P. 2014. *Homology, genes, and evolutionary innovation* (Princeton University Press).

Wagner, D.E., Weinreb, C., Collins, Z.M., Briggs, J.A., Megason, S.G., and Klein, A.M. 2018. "Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo." *Science* 4362: eaar4362.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. 2014. "Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity." *Cell* 158: 1431–1443.

Willmer, E.N. 1970. *Cytology and evolution* (New York, NY: Academic Press).

Xia, B., and Yanai, I. 2019. "A periodic table of cell types." *Development* 146: dev169854.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. 2015. "Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq." *Science* 347: 1138–1142.

Curriculum vitae of Arnau Sebé-Pedrós

Arnau Sebé-Pedrós did his PhD (2009–2014) under the supervision of Iñaki Ruiz-Trillo at the University of Barcelona, investigating the origin of animal multicellularity from a comparative and functional genomics perspective. He then spent four years (2015–2018) at the Weizmann Institute of Science (Israel), working with Amos Tanay on single-cell analysis of animal cell type diversity and genomic regulation. Since 2019 he has been a group leader at the CRG, as part of the Systems Biology programme, where his group investigates the origin and evolution of cell type programs and associated genome regulatory novelties (transcription factors, enhancer elements, chromatin architecture).

Premis de la Secció de Ciències Biològiques

Títols publicats

1 Clara Ruiz-González, *Metacomunitats microbianes: la dispersió i la connectivitat com a factors determinants de la diversitat i la funció dels microorganismes aquàtics = Microbial metacommunities: Dispersal and connectivity as key drivers of the diversity and function of aquatic microorganisms* (2020)

2 Marc Güell, *Noves funcions biològiques sintètiques i implicacions per al present i el futur de la societat* (2021)

3 Marc Güell, *New Synthetic Biological Functions and their Implications for the Present and Future of Society* (2021)

4 Arnau Sebé-Pedrós, *Diversitat, evolució i regulació dels tipus de cel·lulars animals = Animal cell type diversity, evolution and regulation* (2023)


INSTITVT D'ESTVDIS CATALANS
MCMVII