A CONCEPTUAL HISTORY OF THE "REGULATORY GENOME": FROM THEODOR BOVERI TO ERIC DAVIDSON

Maria I. Arnone[1], Paola Oliveri[2] and Pedro Martinez*[3,4]

[1]Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Napoli, Italy.

[2]Centre for Life's Origins and Evolution (CLOE), University College London, London, UK.

[3]Departament de Genètica, Universitat de Barcelona, Av. Diagonal 645, 08028 Barcelona, Spain.

[4]Institut Català de Recerca i Estudis Avancats (ICREA), Barcelona, Spain.

***Corresponding author:**

Pedro Martinez

Departament de Genètica

Universitat de Barcelona

Av. Diagonal 645

08028 Barcelona, Spain

email: pedro.martinez@ub.edu

ABSTRACT

The formalization of the idea of "Regulatory Genome" is a recent one. However, it stems from a long tradition in the study of how the genetic information is transferred between generations. Theodore Boveri suggested for the first time that the whole genome participates in the shaping of individuals. Through a long lineage of researchers, we have learned how this whole-genome activity is regulated, in space and time. It is, however, due to the insights and experimental approaches taken by different researchers, among them Eric Davidson and associates, that we understand the mechanistic basis of this regulation. Whole batteries of regulatory genes interact through their cis-regulatory modules, generating a precise pattern of cross-controlled gene activity (Gene Regulatory Networks). How these genes are deployed in development and evolution has become an area of vibrant research. Here we revisit the history of this intellectual endeavour, taking as key defining points along this historical trajectory the contributions of Theodor Boveri and Eric Davidson.

KEYWORDS

Embryos, Chromosomes, Genome, cis-regulation, Gene Regulatory Network

INTRODUCTION

The study of animal development from a fertilized egg to a completely differentiated embryo or larva has been, and still is, one needed for understanding the mechanisms that control life. The recent approach to the study of development using the gene regulatory network concept, which has been for the first time introduced, as such, by Eric H. Davidson in 1990 (Davidson, 1990) is rooted in a long history of "systems-level" approaches to our understanding of development. Davidson's contributions were the result of the successful integration of several traditions in biological thinking. His wide interests, ranging from the physical nature of the regulatory apparatus, to the role of cell lineage and cellular interactions in the development of animals, are key when it comes to understanding the powerful influence that he had on moulding a new approach to the study of development and evolution. How the different traditions were incorporated into Davidson's thinking about the mechanics of development is a fascinating topic. We can learn much by adopting a historical perspective on the issue that starts with the contributions of classic embryologists. An obvious example here would be that of another scientist who, like Davidson, dedicated his scientific life to decipher the role of the nucleus in development, Theodor H. Boveri.

Boveri and Davidson are therefore at the core of this review that aims at reconstructing the path that brought from the first demonstration of the so-called individuality of chromosomes, by Boveri, to the theory of gene regulatory networks of Davidson. In particular, we will analyse here, on the one side, the background and reasoning beyond the Boveri's theory of individual chromosomes, and on the other the 'gene regulation' theory of Britten and Davidson. In the course of this analysis, we will highlight many commonalities between these two scientists' approaches, from the practical fact that

they both used the sea urchin embryo as experimental tool to build up their models, to the intellectual consideration that they were both scientists thinking ahead of their times. The authors of this review were members of the Davidson laboratory during a period spanning two decades, the 90's and the early 2000's. This was a period in which the experimental paradigm underlying the study of gene regulation changed enormously. We saw the transformation of the field from the "gene by gene" study of regulatory systems to a systems-level approach involving scores of transcription factors and DNA elements linked through a vast array of regulatory interactions. The term Gene Regulatory Network, GRN, arose in developmental biology lexicon during this period to describe this systems-level view of these regulatory interactions. However, and in spite of the technical limitations at different times, the Davidson's laboratory worked within the paradigm that all genome was involved, through the complex interplay of transcription factors and regulatory sequences, in the process of development (and, indirectly, in evolution).

THE NUCLEUS AND THE GENOME: GENESIS OF BOVERI'S VIEWS

Historically, the theory that different sets of genes are expressed in different embryonic lineages is derived from studies of the role of whole genomes during development. As stated by Boveri and others in the late 19th century, all embryo nuclei are functionally equivalent. Specifically, and as proposed by researchers such as Weismann (1885), Hertwig (1885), Nageli (1884) and Strasburger (1884), a fundamental result of fertilization is the formation of a diploid genome from each parent haploid contribution. The genomic determinants of development are then located in chromosomes and are responsible for the characteristics of the developing individual. However, it was only thanks to Boveri's experiments, performed at the Zoological Station of Naples on the

fate of sea urchin dispermic eggs, that the fundamental insight of the individuality of the chromosomes was first demonstrated. The analysis of such series of experiments, often cited as "sea urchin dispermy" (reviewed by Sander, 1993) allowed what nowadays is considered "a key epistemological milestone in the history of embryology" (Pederson, 2006), a fact that has been emphasized by Davidson numerous times in his papers (Davidson, 1968; Davidson, 1976; Davidson, 1985; Davidson 2006).The relevance of the approach was highlighted early on by researchers such as Edmund B. Wilson, who called the experiment: "Boveri's crowning achievement, whether in respect to excellence of method or importance of result" (Wilson, 1918, pp. 74-75).

We cannot fully appreciate the conceptual progress achieved by Boveri through the analysis of sea urchin dispermy and other experiments without first briefly reviewing the knowledge at his time concerning the organization of chromatin and its role in embryonic development. At the turn of the twenty-century, it was known that the number of chromosomes remains constant over successive mitoses, but nobody ever considered them as different from each other in any aspect, their different sizes being hardly noted. August Weismann, for example, thought that each of the prophase chromosomes contained the entire genome, in several copies, which had accumulated during previous generations (Weismann, 1982). Boveri himself, just before starting his famous dispermy experiments thought that "each of the chromosomes brought together in the egg cell [at fertilization] contains all chromatin of the species" (Boveri, 1902, p. 43). It was thanks to the notable experiments that Hans Driesch performed in 1892 at the Naples Zoological Station that Boveri changed his view. It is worth describing here, for the sake of our brief conceptual reconstruction of the history of the regulatory genome, the reasoning that allowed Boveri to demonstrate, for the first time, the individuality of the chromosomes. In a review of 1904, he wrote: "From some years

past… certain reservations had crept up in me because of the pathological development of dispermic eggs, as demonstrated specifically for the sea urchin by Driesch (Driesch, 1892). Driesch raised a considerable number of dispermic sea urchin eggs individually and noted that they all ended up as strongly pathological blastulae (so-called stereoblastule); not a single one was capable of gastrulation. On the condition that all chromosomes are equivalent, and taking into account all [previous] observations and experiments on echinoid development, I was unable to conceive of a cause for this pathological development. So, when the discovery of Herbst (Herbst, 1900) had provided a method by which to separate sea urchin blastomeres safely and without damage, it was obvious that dispermic eggs should be used for checking on the problem of chromosomal equivalence" (Boveri, 1904, pp. 44-45). Continuing with Boveri's words: "In a dispermic egg the division of both sperm centrosomes as a rule gives rise to four [spindles] poles. The egg divides simultaneously into four cells. The question of interest to us is this: how are the chromosomes distributed among the four primary blastomeres? We shall assume for simplicity's sake a chromosome number of four in each pronucleus. Each of the twelve chromosomes will be arranged randomly between any two of the four poles. Fig. 48a represents one of the conceivable cases; Fig. 48b shows the subsequent stage after division of the egg into four cells. One can see that the chromatin content differs between the four blastomeres in number and kind; it is only in the lower left cell that all four kinds are present (Boveri, 1904, pp. 44-46). In Fig. 1 is reported a reproduction of the above mentioned Fig. 48a and b, which Boveri used to illustrate the rationale beyond his long series of experiments where he concluded that: "The four cells arising by simultaneous division of a dispermic egg are essentially equivalent in all properties of their cytoplasm, then the offspring of all four cells must be aberrant in the same way; if it results from the anomalous chromatic complement,

then one should expect that the [four] cells will differ [in their fates]. The experiments yielded the latter result in a most spectacular manner" (Boveri 1904, p. 47). Boveri, who for obvious technical limitations at his time could not count any chromosome number in the blastomeres dissociated from dispermic eggs, and thus demonstrating that it was the loss of specific chromosomes what determined their aberrant fate. He was, nonetheless, able to provide an elegant series of controls and indirect evidence to support his conclusions. After pondering over the problem for several years, in 1907 Boveri eventually published his definitive paper on the matter (Boveri, 1907), which included as indirect evidence also a detailed statistical analysis of a simulation experiment based on developmental mosaicism, thus fully demonstrating what he already postulated in 1902: "What remains is that not a certain number, but a certain combination of chromosomes is required for normal development, and this cannot but mean that the individual chromosomes must possess different qualities" (Boveri, 1902, p.75).

DIFFERENTIATION AS PRODUCT OF DIFFERENTIAL GENE EXPRESSION. THE GENESIS OF DAVIDSON'S IDEAS.

It was T.H. Morgan, in 1934, who stated for the first time that differentiation could be the result of differential expression of genes in cell types. Interestingly, both Wilson (1896) and Morgan (1934) already suggested that the variation of gene expression in cell types could be ascribed to nucleo-cytoplasmic interactions. The modern form of the differential expression of genes during the specification and differentiation of lineages was built up through the 1950s, by Brachet (1949), Sonneborn (1950), Stedman and Stedman (1951), Mirsky (1951), and others. It was from Mirsky that Davidson acquired an interest in deciphering the nature of the regulatory apparatus. In fact, Davidson was

171 brought to the study of the molecular aspects of development at an early age, when, as
172 an undergraduate student, he spent time working in Heilbrunn's laboratory at the
173 University of Pennsylvania. As Ellen Rothenberg has pointed out in a recent review
174 (Rothenberg 2016), "he (Davidson) had a vast furnishing of encyclopaedic knowledge
175 of classical observational embryology from the late 1800s and 1900s"; knowledge that
176 was to grow over the years, giving him a unique (comparative) perspective on animal
177 development. Interestingly, at the time Davidson was a student, both undergraduate and
178 graduate, a model of gene regulation was being constructed, but not in animals, it was
179 (for simplicity reasons) done in bacteria. This was the time, the 50's and early 60's, of
180 the birth of Molecular Biology and the conceptualization of the "regulatory system" as
181 proposed in the Operon Model.
182
183 **BACTERIAL GENE REGULATION: THE OPERON MODEL**
184
185 The models of eukaryotic gene regulation emerged at the end of a decade of the 60's,
186 when another very important contribution to the study of gene regulation was already
187 around: the operon model, proposed by Jacob and Monod to explain the activation of
188 bacterial genes. The model, published in 1961 (Jacob and Monod, 1961), was the
189 product of the long period of innovative approaches to the study of nucleic acid
190 composition, structure and expression. It was a revolutionary period in which,
191 importantly, "a rupture in representations of life shifted from purely material and
192 energetic to the informational, resulting in a molecular vision of life supplemented by
193 an informational gaze" (Kay, 2000), where the introduction of the concept of
194 information represented biological specificity (as it is to this day by most molecular
195 biologists). The new model was based on biochemical and genetic analysis of two

196 systems: regulation of the synthesis of a bacterial enzyme, β galactosidase, on the one
197 hand; and the control of bacteriophage λ lysogeny, on the other. In the paper, the
198 authors proposed a model in which a set of structural genes was regulated in a
199 coordinated fashion by one regulator, in their case a repressor. The group of coordinated
200 genes were called an operon. The binding site for the regulator was called the operator.
201 An important aspect of the model was that repressor activity was regulated by
202 metabolites and this provided a link between gene activity and environmental cues (later
203 on, in eukaryotic models, emphasizing the role of signalling pathways would be a
204 crucial aspect for understanding cell differentiation). In the operon model, the synthesis
205 of bacterial proteins is the product of an intricate regulatory circuit. According to Yaniv
206 (2011), and in agreement with Lily Kay's observation: "such circuits resemble complex
207 control mechanisms in machines or electric circuits or even programs in computers.
208 Indeed, Jacob and Monod can be considered as promoters of the concept of cybernetics
209 in biology." (Kay, 2000).
210
211 **THE NEED OF A SUITABLE MODEL TO STUDY GENE REGULATION: SEA**
212 **URCHINS**
213
214 Davidson studied for his PhD in Alfred Mirsky's laboratory, and it was there that he
215 was introduced to molecular studies of development. His thesis explored gene activity
216 during the development of the frog *Xenopus laevis*. Importantly, in his dissertation he
217 emphasized the fact that "continuous gene action" was required for the maintenance of
218 differentiated functions in cells (Davidson, 1963; cited by Suarez-Diaz and Garcia-
219 Deister, 2015). It was in the early seventies that Davidson switched to the sea urchin
220 embryo as main model of research, an animal which, to use his own words: "…has lent

itself to the study of the role of the genome in embryonic development ever since the discovery of pronuclear fusion in these eggs by Fol (1877)". The reasons for this choice were similar to the ones of all predecessors who used this very same embryo to assess the role of the nucleus in development: the easy access to adult individuals of this animal (Boveri had many collected for him by the fishermen at the Naples Zoological Station, Davidson was initially diving to collect them in California), the abundance of its gametes, the transparency and easy manipulation of its embryos. But the reasons of the great success of the sea urchin as a model for the study of the regulatory genome certainly lays in another aspect: the easy of gene transfer by microinjection into zygotes of this animal, which, starting from the middle eighties, for the first time allowed to experimentally analyse the functioning of regulatory sequences during development (see below).

REGULATORY MODELS: THE "LONG 1970s"

In a recent, very interesting paper, by Suarez-Diaz and Garcia-Deister (2015) it is argued that the period called the "long 1970s" (1969-1983) was critical in the building of a new molecular theory of development. Eric Davidson and his collaborator Roy Britten were essential actors in this process. The period would run from the year when two fundamental models of eukaryotic gene regulation were proposed: one by Georgiev (1969) and the other by Britten and Davidson (1969, 1971); and last until the year when the "homeobox" was discovered, first in *Drosophila* (McGinnis et al. 1984; Scott and Weiner, 1984) and then in *Xenopus* (Carrasco et al, 1984). During this period, no model gained more widespread acceptance than the Georgiev and Britten-Davidson models. They both relied essentially on the new data on genome complexity derived from studies of hybridization techniques. Davidson's longstanding collaborator, Roy Britten,

had introduced those techniques at that time working in the Carnegie Institution (Department of Terrestrial Magnetism). Since 1964, he had used the technique of DNA hybridization kinetics to prove that the genome was composed of different fractions of repetitive and unique sequences. In both models, the repetitive fraction of the DNA was to play a key regulatory role.

It is important to note that these proposals came in a period (1970s) in which these repetitive sequences were considered by most biologists to be "junk DNA" (Orgel and Crick, 1980). The papers were originally published in the Yearly Report of the Carnegie Institution, and became a seminal series of papers in the foundation of the field of regulatory biology. The models of Georgiev and of Britten and Davidson were developed in different contexts, though both took on board all three recognized the intrinsic complexities of eukaryotic genomes. While Georgiev focused mostly on understanding the differences in the regulation of bacterial and eukaryotic genes, Britten and Davidson worked on animal development as the subject to explain. Georgiev, as pointed out by Suarez-Diaz and Garcia-Deister, "sought to explain the newly established facts of eukaryotic genome structure, but relied on the operon hypothesis", with the assumption that "the operon in eukaryotic cells was based on 'non-informative' … and 'informative' … regions". Interestingly for Georgiev, a long DNA-like RNA was transcribed in the nuclei, including both informative and non-informative sequences. Later the "non-informative" sequences were removed and the "informative" ones transferred to the cytoplasm. A key role in the regulation of gene expression was ascribed to repressors: a marked difference from what was proposed by Britten and Davidson (who relied on activators).

The use of hybridization kinetics was extremely important in the approaches that Britten and Davidson followed in order to understand the development of tissues and the

evolution of body plans; all ultimately products of the differential expression of genes, in time and space. This experimental approach brought the two scientists together, in an effort to shed some light on the complexity of gene expression, which was what Alfred Mirsky (Davison's PhD supervisor) was trying to decipher. The collaboration started while, as mentioned before, Roy Britten was working at the Carnegie Institution and Davidson as a faculty member at The Rockefeller University. Later on they would move to California (to Caltech) where they continued to work together for the rest of their lives. Both authors emphasized the relevance of quantitative understanding and the need for causal explanations of development; in particular, explanations that provide logical links to and from the genomic regulatory code (what Davidson called "rooted" explanations). The models of development would have to be based on a systems-level approach to genome function, instead of relying on a specific set of regulators or regulatory pathways (Rothenberg, 2016). Their final goal was to propose models with predictive value, since conceptual predictability was considered a "gold standard". Britten and Davidson's model was developed over two different papers: one in 1969, dealing with cell differentiation and its genomic control; the other in 1971, in which the ideas of 1969 were extended to encompass evolutionary processes (in both cases with an explicit account of supporting data).

EUKARYOTIC VERSUS PROKARYOTIC GENE REGULATION

Now we will move back to the end of the 1960s and the early 1970s, when Britten and Davidson's model was published (in 1969). One of the most stricking aspects of their original 1969 model (see next section for the details) is the "ignorance" of the authors with respect to the contributions made by bacterial geneticists. In the case of the French school, as stated by Francois Jacob (mentioned in Morange, 2017), the Britten and

Davidson model was too speculative and removed from experimental evidence. Ellen Rothenberg (and others) have noted recently, in the first editions of Davidson's classic "Gene Activity in Early Development", the author barely mentions the contributions of Jacob and Monod to the understanding of gene regulation (their papers are not even listed in the bibliography!). The explanation for this was the notion that regulation was through different mechanisms in prokaryotes and eukaryotes: in eukaryotes, it depended very closely on the presence of repetitive sequences in the genome, which was not observed in the microbial world. This idea that the regulation of the two genomic systems was essentially different, was also very much entrenched in the first generation of microbial geneticists (Monod, Brenner or Ephrussi), with the probable exception of Crick (Suarez-Diaz and Garcia-Deister, 2015). In fact a keen observer of the development of the field of molecular biology, Conrad Hal Waddington, published a text in 1969 titled "Gene regulation in higher cells". In it, he suggested that molecular biologists trained in microbiology (referring to Jacob and Monod) did not understand the importance of differentiation or, for that matter, the role of gene regulation as "motor of organismic evolution" (Morange, 2017).

In this context of mutual ignorance, it is still surprising to reread Monod's observation that "what is true for *E. coli* is forcefully also true for elephants", which is at most a very rough approximation to reality.

THE BRITTEN AND DAVIDSON MODEL, 1969

Roy Britten and Eric Davidson published their highly influential model of genomic regulation on 25th July 1969, in the journal *Science*. Under the title "Gene Regulation for Higher Cells: A Theory", the authors produced the first model for the regulation of metazoan development that was fully rooted in the genomic sequence. The paper starts

with a clear concept: “Cell differentiation is based almost certainly on the regulation of gene activity, so that for each state of differentiation a certain set of genes is active in transcription and other genes are inactive”. This view of differentiation as a result of differential gene activity is what was inherited from researchers such as Mirsky (in the 1950s) and others; though Morgan had already considered the idea in 1934 (Morgan, 1934). The paper purports, specifically, to explain facts concerning eukaryotic gene regulation, and the authors stress this point by indicating that “this genome differs strikingly from the bacterial genome due to the presence of large fractions of repetitive nucleotide sequences”, which are transcribed in cell-specific patterns. The model is, from the outset, intended for further experimental testing (and so it has remained, in Davidson’s laboratory, for nearly 50 years). The basic components of the regulatory system (see Fig. 2 for a diagram) are defined as: **producer** genes (akin to Jakob and Monod’s “structural genes”), **receptor** genes (similar to what we would today call the cis-regulatory apparatus), **activator** RNAs (the actual regulators of gene expression; similar to our transcription factors; though Britten and Davidson assume that regulatory transactions are mediated mostly by RNA molecules). It is an important point to stress that, though it has been neglected in many modern papers, Britten and Davidson state clearly that “the role proposed for activator RNAs could well be carried out by protein molecules coded by those RNAs, without changing the formal structure of the model”. These activator RNAs are the product of the so-called integrator genes. The integrator genes are activated through the action of some initiating event, cascading into the activation of sets of producer genes. Regulatory genes, in their model, also meditate signalling events, and this is achieved through the so-called **sensor** genes.

In this model, the authors introduce the rather new concept of “gene batteries” (roughly: a set of producer genes that is turned on when a particular sensor gene activates its

downstream integrator genes). This was a concept that, in a different context, had already been suggested by Morgan in 1934.

It is a model that incorporates some insights that will be crucial later on in the analysis of many developmental systems. The model has a hierarchical nature, where development is carried forward by successive sensory–producer gene links. Gene regulation is mediated by sequence-specific binding within the nuclei and also by the activation of otherwise repressed sites, rather than by repression of active ones (stressing, once more, the differences with bacterial regulation). Moreover, Britten and Davidson emphasise that there is no need in their model for functionally correlated genes being physically linked in the genome.

Knowing that the model has implications for the evolution of animals, they stress, at the end of the paper, what is a clear mechanistic implication: "at higher grades of organization, evolution might indeed be considered principally in terms of changes in the regulatory systems" (with the clear involvement of natural selection in the changing of gene regulatory systems over time). This leads us to the second paper: an extension of the model with implications for understanding organismal evolution.

BRITTEN AND DAVIDSON'S MODEL AND THE EVOLUTION OF ORGANISMS

Davidson's interest in biological diversity (which would foster a very good grasp of the fossil record) was cemented over the years, thanks to his running the Marine Biological Laboratory "Embryology" course. This was (and still is) a forum where students and researchers explore the development of a wide range of organisms. The environment provided by the Course and their many attendants cemented in Davidson an interest in

371 evolutionary problems, a problematic that he considered intimately linked to that of the
372 regulatory control of development. In this context, an understanding of development,
373 according to Davidson, could only result from a thorough exploration of many animal
374 groups. Following in the tradition of classical embryology, more interested in specific
375 problems than model organisms, Davidson took an interest in many developmental
376 systems. In fact, he clearly states this in the Introduction to the Third edition of his
377 "Gene Activity in Early Development", where he emphasizes that "each [animal]
378 system has its strong points and its weak points as an experimental object, and it is
379 impossible to obtain anything but a partial view of the processes of oogenesis and early
380 development through the lens that any one system provides. The approach taken in this
381 review is thus comparative…." (Davidson, 1986, pp. 3). For Britten and Davidson
382 Evolution and Development were two sides of the same problematics, and were certain
383 that understanding both could only come through the study of gene regulatory processes
384 in different organisms.

The construction of a model of regulation that would address the problematics of evolutionary change was presented in 1971. As happened with the 1969 paper, that of 1971 starts with a clear position: “[we] have constructed a model for gene regulation. … Here we consider some of the implications for the processes of evolution … The purpose, as in our previous papers, … is to construct a conceptual scheme which can be tested. ... In this paper we follow the view that major events in evolution require significant changes in patterns of gene expression” (Britten and Davidson, 1971). This is a view that most people working on evolutionary mechanisms would nowadays adhere to. Britten and Davidson re-examine the structure of the gene networks, as published in 1969, to emphasize that these networks are susceptible of change, through a process that would generate new regulatory interactions. How these new interactions are generated is a problem that is treated extensively. The authors assume that these new regulatory configurations are the product of rearrangements of the genome. Since the authors work under the hypothesis that genetic function is related to the arrangement of sequences in the genome (remember that repetitive sequences are at the core of regulatory functions), it is a logical deduction that “alterations in the organization of the genome by rearrangement would be expected to have profound effects”. These rearrangements should, over evolutionary time, have the effect of constructing new “regulatory networks” (they explicitly use this terms here), with new structures being the product of changing regulatory relationships.

The paper finishes with a prediction with a typically "von Baerian" flavour, and this is that “as development progresses from stage to stage, progressively less ancient and phylogenetically more restricted genomic regulatory patterns would come into play”.

This is truly a surprisingly modern model, given the limited experimental data available at the time. It was a model, again, fully rooted in the genomic sequence, in which the whole genome participates in the control of development and evolution. Moreover, a model in which regulatory information is exclusively (hard wired) in the genomic sequences and thus available for experimental testing.

What was certainly lacking at the time was a proper understanding of how this regulatory information was encoded in the genome. There was no possibility of testing the model without a deeper knowledge of the structure and activity of the regulatory apparatus. And that was the challenge that defined Davidson's research programme for the following 50 years. Nowadays, and looking back to the initial interests of Britten and Davidson, it is surprising how the formulation of their idea of Gene Regulatory system has been successfully used to understand evolutionary processes (Hinman et al. 2003 or Dylus et al, 2016).

THE STRUCTURE OF THE GENE REGULATORY APPARATUS

The 1970s and 1980s saw a technological revolution in molecular biology, with the introduction of methods to clone and sequence DNA. This provided a unique possibility for analysis of how genes function. Later, mostly at the beginning of the 1990s, the introduction of methods to study DNA–protein interactions would provide the set of tools that would be used in Davidson's laboratory (and others) to thoroughly test the 1969 model. However, the first 20 years of enquiry were focused on individual genes ("The characteristic genomic dimension of the era's analytical methods was a few thousand base-pairs" (Galas and McCormack, 2003)). The new endeavours of Davidson's laboratory would rely on the use of a favourite system, the sea urchin *Strongylocentrotus purpuratus*, readily available in great numbers along the coast of

California. Using newly cloned genes, in the 1980s, Davidson's laboratory began systematically to analyse how cell type-specific patterns appeared. At this point, as Rothenberg has clearly stated "the transition went from a global view (of gene activity) to the selection of specific genes that could illustrate some general principles of gene regulation" (Rothenberg, 2016). This approach resulted in the first analysis and identification of regulatory sequences, the actin genes CyIIIa (Calzone et al. 1988) and CyIIa (Arnone et al. 1998), the skeletogenic gene SM50 (Makabe et al. 1995) and the most thoroughly analysed gene to date (from the regulatory point of view), the endodermal structural gene ENDO16 (Yuh et al. 1994). These represented the first comprehensive analysis of gene regulation, information that would be critical in the later development of the field of "gene regulatory networks". The period saw the painstaking dissection of regulatory segments, massive gel shift assays (for DNA–protein interactions) and the isolation of large volumes of stage-specific protein extracts (a collective endeavour of the whole laboratory) for DNA binding assays and transcription factor protein isolation. This was a period in which these main experimental approaches involved a combination of biochemistry purification techniques (taking advantage of the fact that millions or billions of synchronized embryos could be obtained in the laboratory during the breeding season; ie: Coffman et al. 1992 or Zeller et al. 1995) and in vivo cis-regulatory analysis done using transgenesis with reporter genes (either CAT or GFP; ie: Zeller et al. 1992). These techniques made the use of sea urchin embryos to dissect the regulatory regions of the genome unique. It became, for a long while, *the* system to thoroughly analyse the regulatory apparatus of any one gene. Complementary technologies such as the construction of genomic and cDNA libraries or microinjection into embryos were crucial to push forward this field of enquiry. The experimental analysis of these

regulatory regions provided some key insights: in particular, the modularity of their construction, with different modules used to regulate specific and discrete domains of the whole expression pattern, in space and time. As Rothenberg points: “the highly detailed picture that emerged was a dramatic demonstration that cis-regulatory systems could act as tiny computers”.

THE DEVELOPMENT OF THE SEA URCHIN EMBRYO. SPECIFICATION OF TERRITORIES.

It is important to point out here that the effort to characterize the regulatory elements controlling the cell-type (or territory) expression of some genes could not have led to a conceptual change the ways we understand specification and differentiation in embryos if the laboratory had not made a complementary effort to dissect, in detail, the development of the sea urchin embryo. Studies of cell lineage, cell transplantation and the role of signalling in the specification of the different territories were keys in the eventual interpretation of how the genome controls development. They also suggested, later on, the idea that gene regulatory networks were deployed in a hierarchical fashion. These important experiments were performed, and the resulting insights gained, during the 1990s were later reviewed in a highly cited paper: Arnone and Davidson 1997.

It is important to emphasize here that for Davidson, gene regulation and embryogenesis were two interlaced aspects of the same problem, how the regulatory apparatus were deployed in cell lineages to generate differential gene expression, in space and time. In a seminal paper published in 1991 (Davidson, 1991), Davidson explores the different types of embryonic development present in metazoan animals and classifies them in three different groups (Types 1,2 and 3). The idea permeating this paper is that animal

embryogenesis (and its variability) can be explained only as the result of the deployment of different regulatory strategies. Davidson promotes the idea that body part formation, in general, is regulated at a genomic level. Then these body parts, starting from the early progenitor fields, are specified through a series of finer subdivisions of the field 'into appropriately positioned regulatory state domains that generate the subparts and ultimately the cell types of the body part' (Peter and Davidson, 2015). In this context cell-lineages and GRNs are intimately linked, where cell lineages are seen as progressively moving through regulatory states configured by the GRNs.

It would not be fair to paint a picture in which only sea urchins were providing, at the time, insights into the organization and function of the regulatory apparatus. Davidson's ideas were very much the result of interaction with a series of brilliant scientists working on similar topics. Perhaps the most influential lessons were provided by the elegant work of Michael Levine with the *Drosophila* embryo (see below).

GENOMIC ANALYSIS OF GENE REGULATORY FUNCTIONS

The shift from the study of single gene regulatory functions to a large-scale, genomic approach took place during the first decade of the 21st century. The impulse for this came, obviously, from the sequencing of the *Strongylocentrotus purpuratus* genome. In this last section we will summarize, briefly, the key events in the story.

The systematic analysis of cis-regulatory sequences that the laboratory of Eric Davidson carried out during the 90's, plus those in other systems (such as those unveiling the regulation of segmental patterns in *Drosophila*) lead to the realization that transcription factors were connected to each other through complex networks. Underlying this realization is the now obvious fact that transcription factors are encoded by genes, and thus are targets of other TF regulators. In fact, and as clearly synthesized by Peter

(2016): “the main argument for constructing a model for gene regulatory networks was that if the spatial expression of differentiation genes is encoded in their cis-regulatory sequences, and read by specific transcription factors, then the same mechanism had to be responsible for the expression of these transcription factors in the right domain of the embryo at the right time in development”.

A nice insight derived from the analysis is of the first cis-regulatory domains, whether in sea urchins or *Drosophila melanogaster* (i. e. Small et al, 1991), was the fact that the regulatory apparatus for those developmental genes was organized in a modular fashion, with modules executing specific regulatory functions (spatial restriction, temporal activation or amplitude control; among others). These modules were populated of transcription factors acting in concert to specify some of the characteristic details of the expression pattern. Needless to say, the parallel work of Michael Levine was very important for Eric Davidson. They were at the time close intellectual partners and their respective research programs very much influenced by their regular interaction. The group of Eric Davidson published the very first comprehensive gene regulatory analysis in 2002 (Davidson et al, 2002). It dealt with the control of the specification of the endomesoderm layer, at the vegetal pole of the embryo, and during the firsts hours of development (before gastrulation). The model represented a paradigm shift (*sensu* Kuhn) in the field of gene regulation. It changed the way development was seen, a product of the “unfolding” of genomic instructions in which the expression of genes, in space and time, were controlled “only” through the cross-interaction of transcription factors and cis-regulatory sequences. The formalization of this “unfolding” was best represented by a Network of Regulatory Genes (GRN; see an example in Fig. 3).

THE CONSTRUCTION OF A MATURE MODEL OF GRNs.

The sequencing of the sea urchin (*S. purpuratus*) genome was of key importance for the development of a mature theory of gene regulation. It allowed transforming the field study from a "gene by gene" analysis to a comprehensive, systems-level, understanding of genomic regulation. Relevant to this transformation was the close relationship established in Caltech's Division of Biology between Eric Davidson and Leroy Hood. The latter was, from early on, a proponent of using "systems-level" approaches to understand/decode life (Ideker et al, 2001).

During the period starting in the year 2000, and previous to the sequencing of the sea urchin genome, an enormous technical effort was put into identifying extensively downstream genes of transcriptional regulators. The approached took advantage of the availability of different concurrent technologies: the amplification of cDNAs from small tissues and cDNA libraries arrayed in nylon filters (Rast et al, 2000; Rast et al., 2002; Ransick et al., 2002). This procedure was improved to the level where differentially expressed genes could be detected at levels of less than 5 molecules per embryo cell.

This herculean effort allowed, for the first time, to have access to batteries of regulated cells, a stepping-stone in the road towards understanding cross-regulation and GRNs. Needless to say, those were experiments done before any genomic sequence was known. Other factors contributing to this "paradigm shift" were the incorporation at the time of tools such as large-scale transgenesis, including BAC-derived systems, massive sequencing and, very importantly, the development of computational tools.

The progressive additions of all new generated expression data resulting from direct cis-regulatory (binding) experiments plus the analysis of gene perturbation allowed a more refined modelling of the endomesoderm (and other) networks. In parallel, more refined (Boolean) models were developed that would incorporate the cis-regulatory information

559 generated though the experiments. Eventually, and almost as a gran finale for the work
560 developed by the Davidson's group, in 2012, a Boolean model of the endomesoderm
561 network was produced that "contained a nearly complete set of instructions for
562 developmental gene expression for the first 30 h of sea urchin embryogenesis" (Peter et
563 al., 2012). Importantly, the theoretical model has a heuristic value, since it allows to
564 computationally predict the effect of specific perturbations to the system. The
565 sufficiency of the network for "explaining" development was vindicated.
566
567 **CONCLUSIONS**
568 The historical reconstruction presented in this review shows many commonalities
569 between Boveri's and Davidson's approaches to studying development: similar
570 pragmatic, experimentally based, approaches focusing on mechanisms. However, both
571 of them were guided by theoretical considerations (an *a priori* abstract modelling of the
572 process to be explained). Boveri and Davidson considered the whole genome as
573 fundamental driver of developmental processes. In this context, they attributed to the
574 whole genome, and not to any specific part of it, the responsibility of driving
575 development in a particular (forward) direction. In fact, both emphasized the importance
576 of the nucleus as the physical space where regulatory information is stored.
577 Boveri and Davidson were aware of the need of using quantitative methodologies to
578 understand development. Both had a good grasp of mathematics and were willing to use
579 it for the understanding of developmental processes. While qualitative understanding
580 was assumed as a prerequisite, a thorough, quantitative approach to the description of
581 processes was (at the end) necessary. The insistence in understanding the quantitative
582 aspects of any developmental process were at the core of their conceptions of Biology, a
583 view that those working with Davidson were fully aware.

584 The sea urchin embryo was the reference in their theories, but both Boveri and
585 Davidson applied their ideas far beyond: they knew that every system was informative
586 in many ways, and all were necessary to tackle specific problems of development and
587 (in the case of Davidson) evolution.
588 All in all, here we have shown the similar views that these two scientists had on the
589 genetic basis controlling developmental processes. In the intellectual lineage that linked
590 the work of both scientists, the conceptual seeds planted by Boveri ultimately flowered
591 and bore fruit in the theoretical work of Davidson.
592
593 **ACKNOWLEDGMENTS**
594 This paper is dedicated to our late mentor, Eric H. Davidson. He was a key influence in
595 our scientific careers, moulding the way we approach biological problems, with a
596 combination of curiosity for animal embryos and a keen appreciation of all quantitative
597 aspects of development's regulation. We should be missing al his insights.
598 It is also important to dedicate this to the group of people, friends and colleagues that
599 shared the laboratory with us over these years. They were instrumental in testing the
600 hypothesis that Britten and Davidson had postulated originally in 1969 and 1971. They
601 were all great scientists better friends.
602 All translations of Boveri's article, with additions set in square brackets, are from Klaus
603 Sander's essay (1993).
604 We would like to thank the two anonymous referees for their many insightful
605 comments.

606 **REFERENCES**

607 Arnone MI, Davidson EH (1997). The hardwiring of development: organization and
608 function of genomic regulatory systems. Development 124: 1851-1864.

609 Boveri T (1902). Über mehrpolige Mitosen als Mittel zur Analyse des Zellkerns. Verh
610 Phys.-med Ges Würzburg [N F] 35: 67-90.

611 Boveri T (1904). Ergebnisse über die Konstitution der chromatischen Substanz des
612 Zellkerns, Gustav Fischer, Jena.

613 Boveri T (1907). Die Entwicklung dispermer Seeigel-Eier. Ein Beitrag zur
614 Befruchtungslehre und zur Theorie des Kerns (Zellenstudien VI). Gustav Fischer,
615 Jena.

616 Brachet J (1949). L' hypothèse des plasmagènes dans le développment et la
617 différenciation. Colloq. Int. CNRS 8, 145.

618 Britten RJ, Davidson EH (1969). Gene regulation for higher cells: a theory. Science.
619 165: 349-357.

620 Britten RJ, Davidson EH (1971). Repetitive and non-repetitive DNA sequences and a
621 speculation on the origins of evolutionary novelty. Q Rev Biol. 46: 111-138.

622 Carrasco AE, McGinnis, W, Gehring WJ, De Robertis EM (1984). Cloning of an *X.*
623 *laevis* gene expressed during early embryogenesis coding for a peptide region
624 homologous to Drosophila homeotic genes. Cell 37: 409–414.

625 Davidson, EH (1963). The expression of differentiated character in monolayer tissue
626 culture cells, Doctoral Dissertation. Ann Harbor: The Rockefeller Institute.

627 Davidson EH (1968). Gene Activity in Early Development. Academic Press.

628 Davidson EH (1976). Gene Activity in Early Development, 2nd edition, Academic
629 Press.

630 Davidson EH (1985). Genome function in sea-urchin embryos: fundamental insights of
631 Th. Boveri reflected in recent molecular discoveries. In: Horder TJ, Witkowski JA,
632 Wylie CC (Eds.), A History of Embryology. 8th Symposium of the British Society
633 for Developmental Biology. Cambridge Univ. Press, pp. 397–406.
634 Davidson EH (1986). Gene Activity in early development, 3rd edition, Academic Press,
635 Orlando, Florida.
636 Davidson EH (1990). How embryos work: a comparative view of diverse modes of cell
637 fate specification. Development 108: 365-389.
638 Davidson EH (2006). The Sea Urchin Genome: Where Will It Lead Us? Science 314:
639 939-940.
640 Davidson EH (2015). Genomics, "Discovery Science," Systems Biology, and Causal
641 Explanation: What Really Works?. Perspect Biol Med. 58: 165-181.
642 Davidson EH, Cameron RA, Ransick A (1998). Specification of cell fate in the sea
643 urchin embryo: summary and some proposed mechanisms. Development 125: 3269-
644 3290.
645 Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T,
646 Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY,
647 Revilla R, Rust AG, Pan Zj, Schilstra MJ, Clarke PJ, Arnone MI, Rowen L,
648 Cameron RA, McClay DR, Hood L, Bolouri H (2002). A genomic regulatory
649 network for development. Science 295, 1669-1678.
650 Driesch H (1892). Entwicklungsmechanische Studien III-V1. Z Wiss Zool 55: 1-62.
651 Dylus DV, Czarkwiani A, Stångberg J, Ortega-Martinez O, Dupont S, Oliveri P (2016).
652 Large-scale gene expression study in the ophiuroid *Amphiura filiformis* provides
653 insights into evolution of gene regulatory networks. Evodevo 7: 2.

Fol MH (1877). Sur le premier développement d'une Étoile de mer. Comptes Rendus 84, 357-360.

Galas DJ, McCormack SJ (2003). An historical perspective on genomic technologies. Curr Issues Mol Biol. 5: 123-127.

Georgiev GP (1969). On the structural organization of operon and the regulation of RNA synthesis in animal cells. J Theor Biol, 25: 473-490.

Herbst C (1900). Über das Auseinanderweichen von Furchungs- und Gewebszellen in kalkfreiem Medium. Arch Entwickelungsmech (Roux's Arch) 9: 424-436.

Hertwig O (1885). The problem of fertilization and isotropy of the egg, a theory of inheritance. (Engl. Transl.) In Voeller, BR ed (1968). "The chromosome theory of Inheritance: Classic papers in Development and Heredity". Appleton, New York.

Hinman VF, Nguyen AT, Cameron RA, Davidson EH (2003). Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. Proc Natl Acad Sci USA 100: 13356-13361.

Hood L, Rothenberg EV (2015). Developmental biologist Eric H. Davidson, 1937-2015. Proc Natl Acad Sci USA 112: 13423-13425.

Jacob F, Monod J (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3: 318-356.

Kay LE (2000). Who Wrote the Book of Life? A history of the genetic code. Stanford University Press, Stanford.

Laubichler MD, Davidson EH (2008). Boveri's long experiment: sea urchin merogones and the establishment of the role of nuclear chromosomes in development. Dev Biol 314: 1-11.

McGinnis W, Levine MS, Hafen E, Kuroiwa A, Gehring WJ (1984). A conserved sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. Nature 308: 428–433.

Mirsky AE (1951). Some chemical aspects of the cell nucleus. In "Genetics of the 20th Century" (LC Dunn, ed), p 127. Macmillan, New York.

Morange M (2017). Homage to Eric Davidson. Hist Philos Life Sci 39: 30.

Morgan TH (1934). Embryology and Genetics. Columbia Univ. Press, New York.

Nageli C (1984). Mechanisch-physiologische theorie der abstammungslehre. R Oldenburg, Munich and Leipzig.

Orgel LE Crick FH (1980). Selfish DNA: the ultimate parasite. Nature 284: 604-607.

Pederson T (2006). The sea urchin's siren. Dev Biol. 300: 9–14.

Peter IS (2016). Eric Davidson. A genomic control odyssey. Dev Biol 412: S41-4.

Peter IS, Faure E, Davidson EH (2012). Predictive computation of genomic logic processing functions in embryonic development. Proc Natl Acad Sci USA 109: 16434-16442.

Rothenberg EV (2016). Eric Davidson: Steps to a gene regulatory network for development. Dev Biol 412: S7-19.

Sander K (1993). Aneuploidy disrupts embryogenesis: Theodor Boveri's analysis of sea urchin dispermy. Roux's Arch Dev Biol 202: 247-249.

Scott MP, Weiner AJ (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of *Drosophila*. Proc Natl Acad Sci USA 81, 4115–4119.

Small S, Kraut R, Hoey T, Warrior R, Levine M (1991). Transcriptional regulation of a pair-rule stripe in *Drosophila*. Genes Dev 5: 827-839.

Sonneborn TM (1950). The cytoplasm in heredity. Heredity 4: 11.

Stedman E, and Stedman E (1950). Cell specificity of histones. Nature 166: 780.

Strasburger E (1884). New investigations on the course of fertilization in the phanerograms as basis for a theory of heredity. (Engl. Trans.) In Voeller BR, ed. (1968) "The chromosome theory of inheritance: Classic papers in Development and Heredity." Appleton, New York.

Suárez-Díaz E, García-Deister V (2015). That 70s show: regulation, evolution and development beyond molecular genetics. Hist Philos Life Sci 36: 503-24.

Waddington CH (1969). Gene regulation in higher cells. Science, 166: 639-640.

Weismann A (1885). Die Continuität des Keimplasmas als Grundlage einer Theorie de Verersbung Fisher, Jena [The continuity of the germ plasm as the foundation of a theory of heredity. (Engl. Transl.) In "Essays upon Heredity and Kindred Biological Problems." Poulson, E.B., Schönland, S., and Shipley, A.E. eds. (1891). Vol.1, p. 162. Clarendon Press, Oxford].

Weismann A (1892). Das Keimplasma -eine Theorie der Vererbung. Gustav Fischer, Jena.

Wilson EB (1896). The cell in Development and Inheritance. Macmillan, New York.

Wilson EB (1918). Theodor Boveri. In: Roentgen CW (ed) Erinnerungen an Theodor Boveri. JBC Mohr (Paul Siebeck), Tübingen, pp 67-89.

Yaniv M (2011). The 50th anniversary of the publication of the operon theory in the Journal of Molecular Biology: past, present and future. J Mol Biol 409: 1-6.

Yuh CH, Ransick A, Martinez P, Britten RJ, Davidson EH (1994). Complexity and organization of DNA-protein interactions in the 5'-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. Mech Dev 47: 165-186.

FIGURE LEGENDS

Figure 1. Diagram by Boveri of the Simultanvier type division of a dispermic egg, illustrating the chance distribution of the three chromosome sets among the four spindle poles derived from the two spermatozoa. Adapted from Boveri (1904).

Figure 2. Diagrammatic scheme of the Britten and Davidson regulation model. Names for all gene components are those used in their original papers.

Figure 3. Diagrammatic scheme of a typical Gene Regulatory Network visualized using BioTapestry (Longabaugh et al, 2005). TF, transcription factor; TD, terminal differentiation.

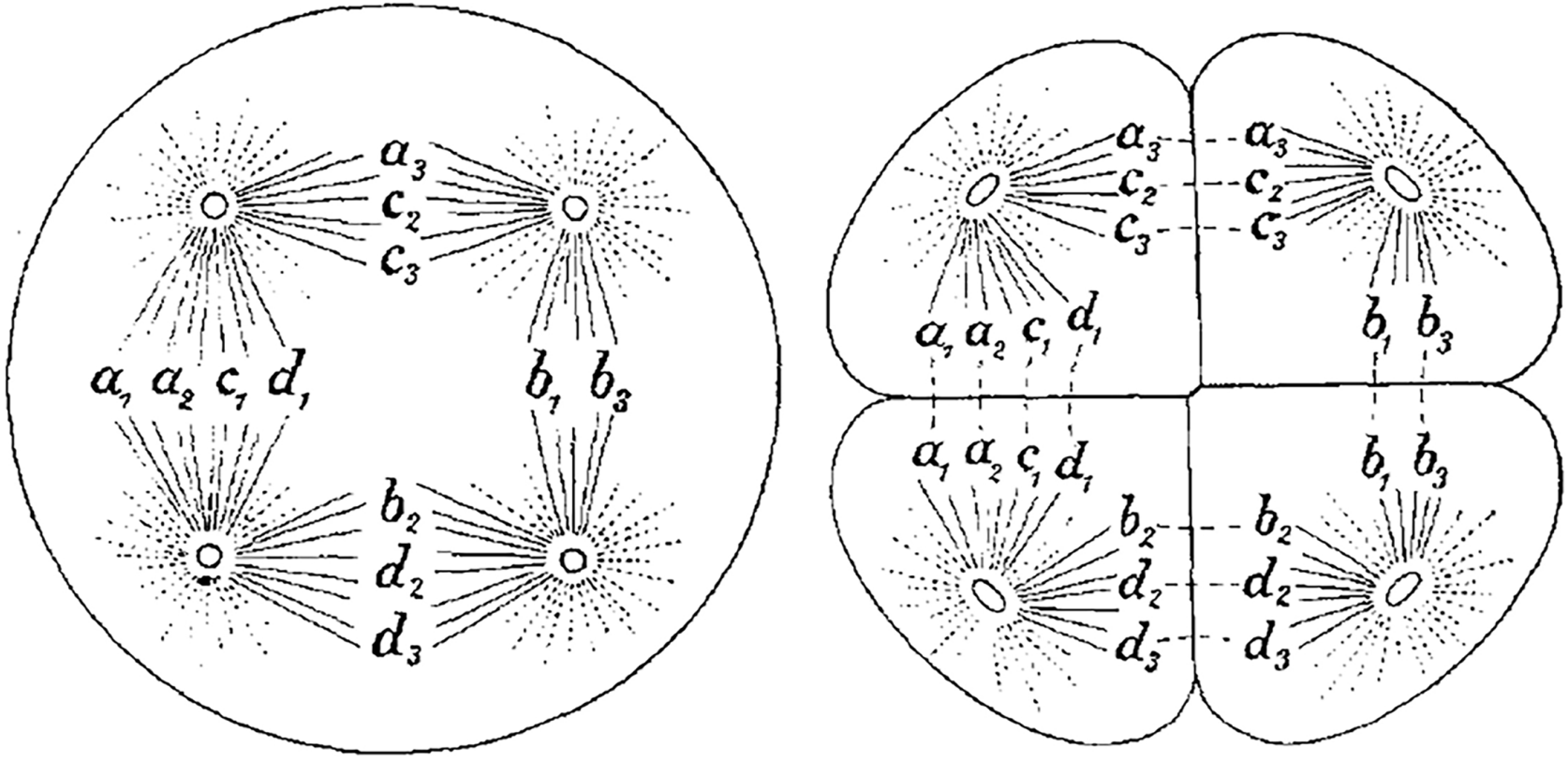
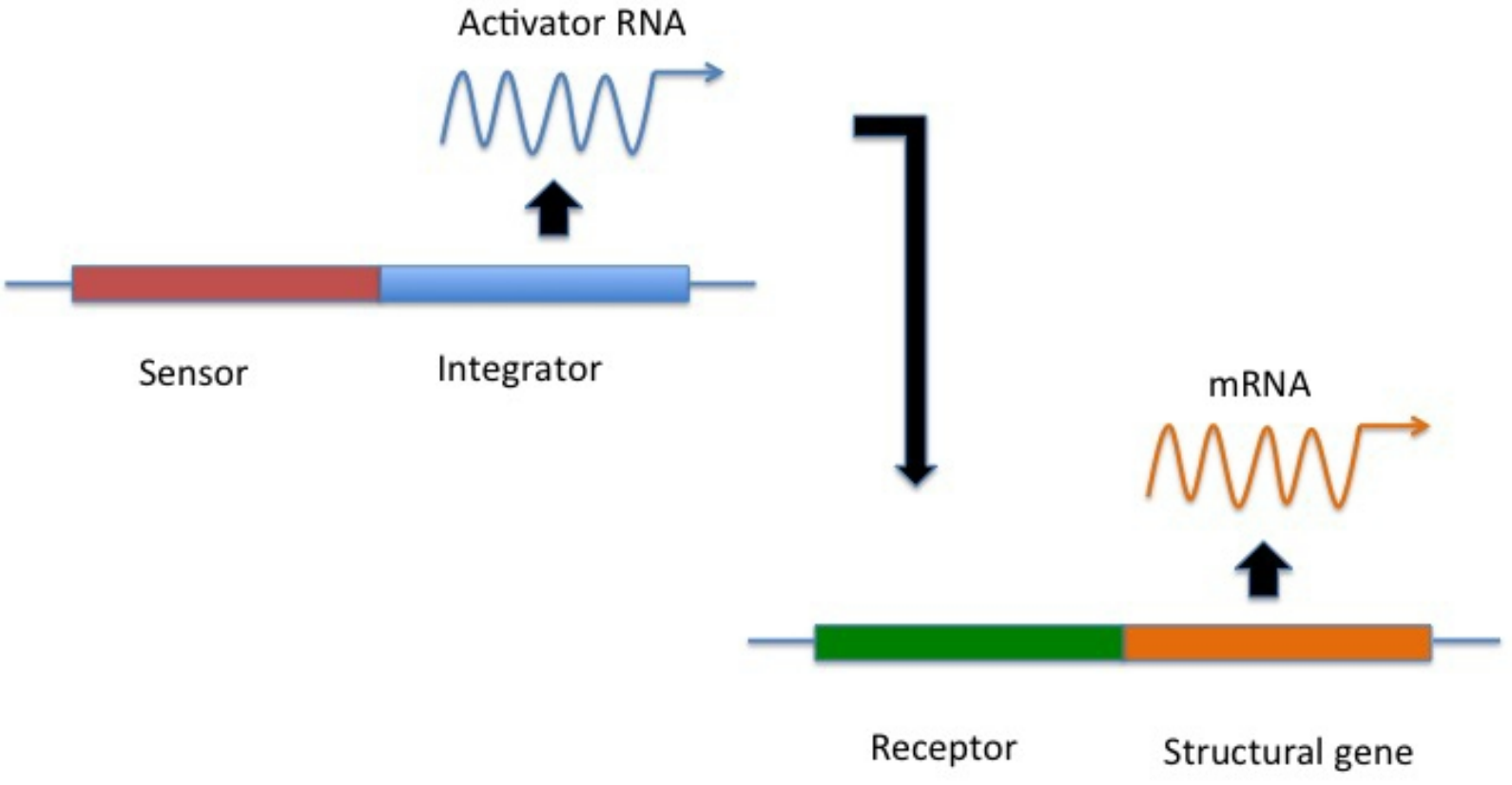
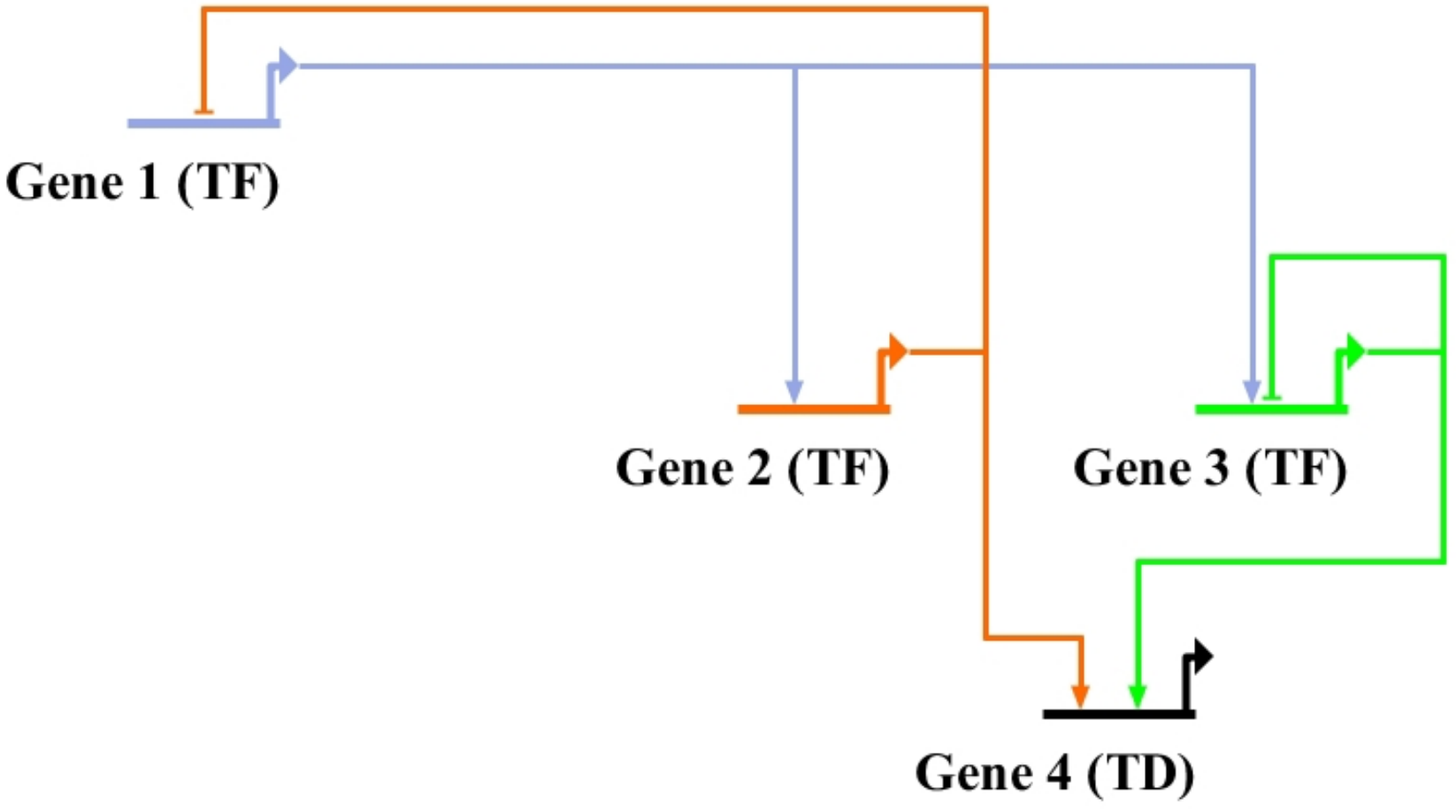


Fig. 48a und b. Schema eines Falls von Chromosomenverteilung bei der Entwicklung eines doppeltbefruchteten Eies.


Activator RNA
Sensor
Integrator
mRNA
Receptor
Structural gene


Gene 1 (TF)
Gene 2 (TF)
Gene 3 (TF)
Gene 4 (TD)