PEP-NET

European Commission

Predictive Epigenetics: Fusing Theory and Experiment

Understanding epigenetics and our PEP-NET logo

written by Ana Fernández Palacio (ESR13), Jennifer Giannini (ESR10), Nancy Gutierrez Hernandez (ESR1), Tom Fillot (ESR9) with input from Jane Mellor and Mihaela Pruteanu

What is epigenetics?

Epigenetics has a diverse number of definitions but all of them in some way refer to heritable changes in gene expression that occur without direct changes to DNA sequence (i.e. the "*epi*" in *epi*genetics literally means "on top of genetics").

Every cell in your body has the same DNA, yet they are obviously different from one another: a neuron looks and acts in a different way than a muscle cell. Despite having the same DNA, these cells are reading and utilizing it differently, thanks to the help of epigenetics marks on the DNA. For example, adding a particular chemical group, in this case called methylation, on the DNA usually signals the cell not to read a particular gene. The addition of a different chemical group, called acetylation, to a histone (a protein around which DNA is wrapped) instead indicates that a gene should be active (Figure 1). RNA is also a fundamental component of the epigenome and may act by recruiting modifying enzymes.
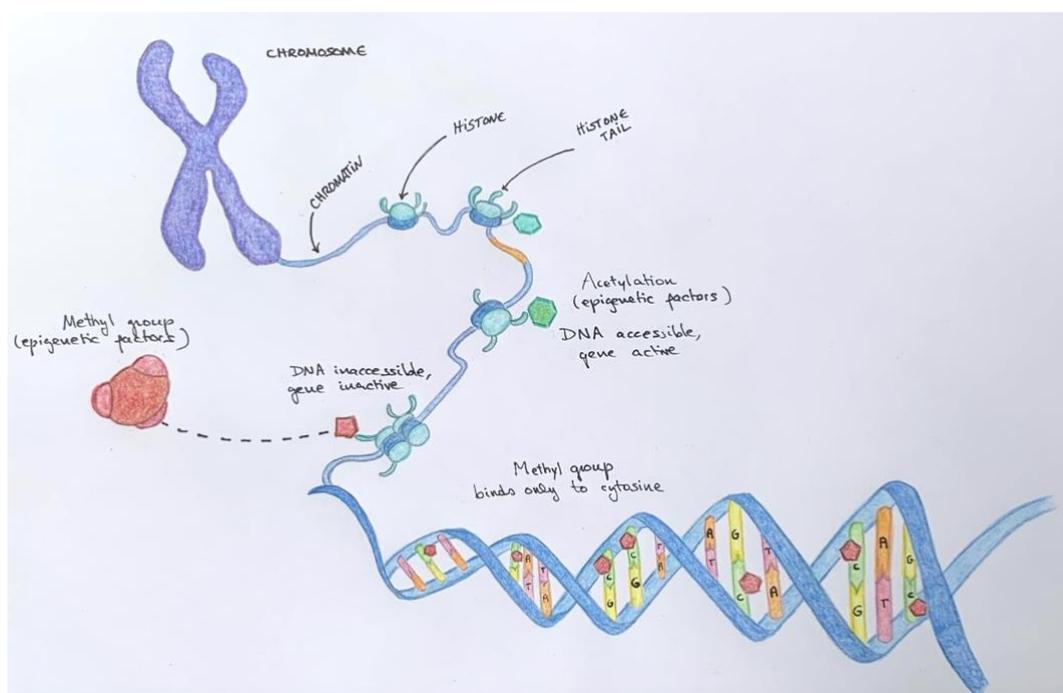

Figure 1. Epigenetic mechanisms (hand-drawn by Ana Fernández Palacio after the image from Shutterstock.com/ellepigrafica).

Epigenetics was coined by the embryologist and theoretical biologist, Conrad Waddington. Waddington wnated to understand how different parts of the embryo establish their identity, but by far what he is most remembered for today is "the epigenetic landscape" (Figure 2).
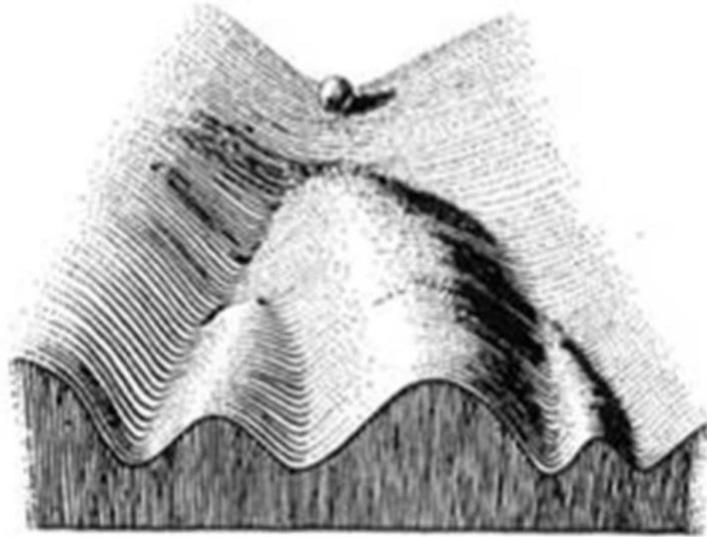

Figure 2. Waddington's epigenetic landscape (from Bhattacharya, S. et al. BMC Syst Biol **5,** 85 (2011), distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0)

First presented in his book called "The Strategy of the Genes", the epigenetic landscape is a concept that brings together all of the essential aspects of developmental biology in a single visual metaphor. In this drawing, the cell is represented by a ball rolling down a slope.

Now, let's put this in a biological context: if you imagine a zygote, the cell that originates after a sperm fertilizes an egg, this unique cell will give rise to a complete organism with different cell types: neurons, muscle cells, immune cells, etc. all with distinct functions. In the epigenetic landscape, the zygote would be the ball that is at the top of the hill. As it rolls down and continues to divide, its daughter cells commit to specific cell types until the end of development. Thus, the whole organism arose from individual cells taking different paths down the hill, which end up with a specific identity that will be kept for the rest of their life.

The **logo** of our network (Figure 3) features Waddington's epigenetic landscape, which will be explained further by the end of the article.

With this drawing, Waddington was the first to emphasize that a more general, complete point of view was necessary to understand how humans develop from a single cell. He focuses on the landscape itself, and how it is created by both the genome and the surrounding environment.

Why should you care about epigenetics?

If we search on the internet for the term "epigenetics", we can find harmless (yet misleading) articles like epigenetic pancakes. However, this also represents a window of opportunity for dishonest sellers offering miraculous products or recipes with the promise of gaining the upper hand on the battle against various health conditions or spreading the false idea that we are in control of the fate of our health by changing our diet and exercise habits alone. On the other hand, scientific research on epigenetics has made many interesting discoveries, some of

which are examples in nature where the influence of an organism's environment has a direct impact on its embryonic development.

Bees
When it comes to bees, we can recognize different castes in their social structure: queen, workers, and drone bees. They all play different roles in the colony and have distinct physical characteristics: compared to its worker counterparts, the queen bee is bigger, lives longer and it is fertile. The amazing thing is that all of them have the same DNA sequence, so what makes them different? Their epigenome. A few years ago, the Maleszka lab described the differences in DNA methylation in the brain of queen and worker bees. During larval development, larvae that eat "royal jelly" for a longer period develop into queen bees. In 2011, Astrid Dpannhoff and collaborators investigated the composition of royal jelly and described a component with the function of inhibiting the activity of a histone deacetylase. Histone deacetylases are a category of enzymes with the function of erasers, which modify the epigenome. But hold on, there is no scientific evidence on the effects of royal jelly in humans!

Plants
Now let's move to plants, which, because of their environment, have to respond very drastically to temperature changes. In *Arabidopsis thaliana,* one of the most studied species of plants, flowering is controlled by a gene called FLC, whose expression is linked to temperature. In winter time, after long exposure to cold weather, the FLC gene prevents the plant from flowering. When spring comes and the temperature increases, the FCL gene is repressed which allows the plant to flower. The FCL gene is switched on and off by the interplay between proteins and RNA molecules.

(*Svenja Reeck, PhD student in our network, is studying the effect of warmth in flowering).*

Mammals - Mice and rats
A well-studied example in mammals is the agouti viable yellow (Avy) version of the gene that influences coat colour and metabolism in the mouse. Near the agouti gene in the genome, there is also a type of genetic element called retrotransposon, in this case called the "IAP". When the IAP is unmethylated, the agouti gene is abnormally expressed leading to these mice developing yellow coats, obesity and diabetes.

Not only nutrition, but also some chemical compounds included in today's diets can lead to changes in DNA methylation at several genomic locations in mice and rats, leading to fertility issues. Specifically, when Vinclozolin, an endocrine disruptor widely used in vegetable and fruit production, is administered by intravenous injection to male mice and rats, it causes problems in testicular development and sperm production and maturation. These effects are observed to be paternally inherited for up to three generations. For instance, in rats, this transgenerational effect is correlated with some slight alterations in DNA methylation levels of promoter regions in sperm (promoter = genomic sequences needed for gene expression to start).

Humans
Nutrition is one of the most modifiable factors that can have an impact on human health, especially during early developmental stages. For instance, diet can affect epigenetic mechanisms such as DNA methylation, histone modification, and RNA interference pathways.

Changes in DNA methylation levels of certain genes involved in metabolism can result in persistent metabolic and physiological changes. These changes, which are specific to the locations of particular genes, can be driven by environmental stimuli or toxic compounds. Ultimately, this can lead to different levels of susceptibility to chronic diseases later in life,

especially those related to metabolism. This concept is referred to as "developmental origins of health and disease''.

Several studies have shown a correlation between a mother's unbalanced diet during prenatal stages with abnormal fetal growth and risk of cardiovascular disease during adulthood. The epigenetic link lies in some blood biomarkers that affect the intake of nutrients which are sources of methyl-donor groups. The abundance levels of these biomarkers can be used to predict changes postnatally in the methylation patterns of particular genomic sequences in infants (in this case, called epialleles). A metastable epiallele refers to a genetic sequence (locus) that can switch from one epigenetic state to another from one generation of cells to the next, perhaps affecting long-term epigenetic regulation of that locus.

It has also been proposed that nutrition and lifestyle not only affect human fetal development but also throughout adulthood. There are several studies showing that monozygotic (identical) twins (twins that developed from the same embryo) become physically different as they age. The group of Manell Esteller measured the DNA methylation and histone acetylation levels in a group of 80 twins of differing ages and they found that young twins are epigenetically indistinguishable during the first years of life, but older twins have large differences in both of these epigenetic marks, also leading to differences in their gene expression. So it is not uncommon to find older pairs of monozygotic twins that are very different from each other, even one twin having developed a disease while the other remains healthy.

What is our approach and logo?

Now that you know a bit more about the foundation of epigenetics and have learned about some interesting examples, we'll explain a bit about our PhD program's approach to researching epigenetics:
Our paramount aim is to incorporate mathematics (and more specifically mathematical modeling) into traditional experimental epigenetics. The three main reasons we seek to include mathematics in our approach are as follows:

1. **Mathematical models provide scientists with greater predictive power:**
   - When planning and interpreting large-scale epigenetic experiments, it can be difficult to know where to start. Parsing through past literature and available datasets for a given subject, and deciding its relevance for you, can often feel like finding a needle in a haystack. To alleviate this, a well-developed mathematical model can incorporate past findings and predict what may be most interesting to investigate next. Then, as more information is acquired, this process can be repeated. In short, mathematical models help guide our experiments in an informed and meaningful way and help answer questions like:
     - "Which experiments should we do next?"
     - "Which targets are best pursued further? Which are less exciting?"

2. **Mathematical models help scientists interpret their results more thoroughly:**
   a. **Visualization**: In highly complex biological systems where many components are simultaneously interacting, it can be difficult (even for scientists) to keep track of what's going on. A benefit of mathematical models is that they often aim to boil these complex problems down to their principal components. Then, these components are represented through variables and equations in the model. Finally, the model can be visualized through an animation. These animations are

not only beautiful, but are a tremendous help to understanding the biological systems we work with.

([Here you can see an animation](#) of stochastic simulation of RNA production and degradation in a population of cells, from Mellor lab)

b. "More bang for your buck" → **More data/meaning extracted from each experiment**) : As previously mentioned, the data we produce in our experiments can oftentimes be continually fed into a mathematical model or at least continually compared to the analogous data the model produces. In this way, modeling helps us better interpret our data: either by processing it alongside a set of mathematical constraints or by producing another dataset for us to compare our experimental data to. In fact, there are already many examples in epigenetics research where some insights and findings would have otherwise not been possible without mathematical models!

3. **New epigenetic experimental techniques require more advanced data analysis:**
   a. **Single cell technologies:** With traditional epigenetic experimental techniques, a population of cells is usually utilized. This gives scientists an idea of the epigenetic events occurring within that population as a whole. In contrast, with new and improved experimental techniques, cells from the population are isolated and analyzed on their own. This helped scientists to realize that the output of an epigenetic event can vary from cell to cell, over time, and from individual to individual. However, this creates much more data which we need help from mathematics to be able to properly analyze.

   ([Ana Fernández Palacio](#) and [Andrea Hita Ardiaca](#), PhD students in our network, are studying these aspects)

   b. **3D genome architecture:** A comparatively new subsection of epigenetics, genome architecture focuses on the regulation of our genes which occurs through interacting pieces of DNA in 3 dimensions. More specifically, DNA pieces that, although potentially located far apart in linear space (i.e. located on different chromosomes), can be close together and interact with each other often in 3D space. Thus, techniques used to study genome architecture often aim at reconstructing these 3D interactions. Understandably, this type of data is difficult to analyze and even more difficult to visualize - which is where mathematical models become vital.

   ([Alex Abraham](#), [Jennifer Giannini](#) and [Tom Fillot](#), PhD students in our network, are studying these aspects)

   In addition to academic laboratories, PEP-NET also partners with industrial laboratories who value similar research approaches. One such lab is Oxford BioDynamics (OBD), who uses a machine learning approach to predict areas of the genome which take on particular 3D architectures in disease. One of their most recent papers uses this approach to examine different severities of Covid-19 infection.
   Learn more on their website: https://www.oxfordbiodynamics.com/

Our network logo:

Our logo (Figure 3), designed by PEP-NET principal investigators Marc Rehmsmeier and Leonie Ringrose, aims at capturing exactly what is described above: a close friendship between mathematics and epigenetics.
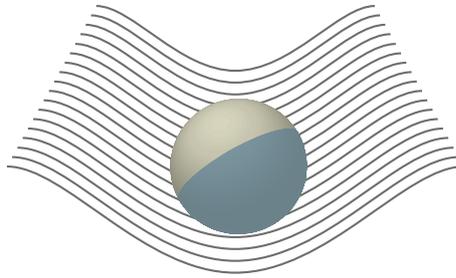

Figure 3. PEP-NET logo

It is also a play on Waddington's epigenetic landscape. In a way, we wish for epigenetics research to "return to its roots". Waddington himself based the epigenetic landscape image using mathematical and theoretical concepts. However, over time the field got carried away focusing mostly on experimental work, leaving the underlying mathematics largely behind. This is exemplified by comparing the number of scientific publications which include the word "epigenetics" versus those which contain "epigenetics and mathematical models" (Figure 4). As a research network, we wish to restore this gap so that the number of research publications including both "epigenetics and mathematical models" are far more numerous. The "ball" in our version of Waddington's epigenetic landscape is half grey and half blue indicating that our research is half experimental and half theoretical. Overall, we also wish to emphasize the importance in science (and beyond!) of collaboration between people from a variety of different backgrounds.
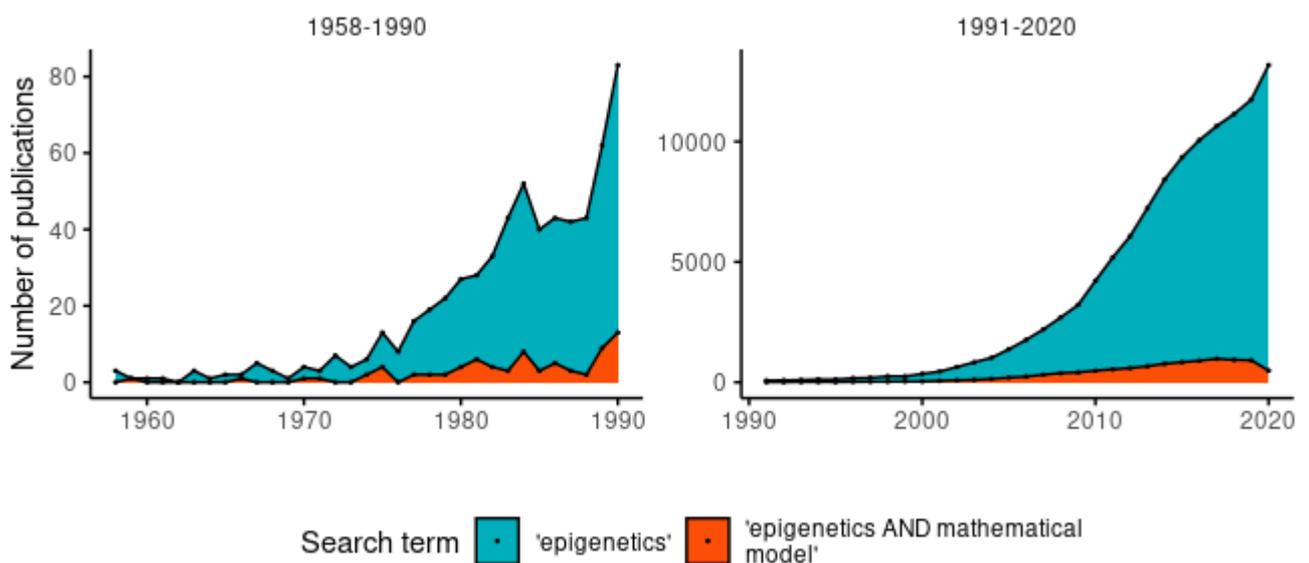

Figure 4. Number of publications per year retrieved with the Pubmed search term "epigenetics' or "epigenetics AND mathematical model" (created by Jennifer Giannini using the search terms from Ringrose, L. & Howard, M. Current Opinion in Systems Biology 2017, 3:7–14).

References:

- Bhattacharya S et al. (2011) A deterministic map of Waddington's epigenetic landscape for cell fate specification. BMC Syst Biol 27;5:85.
- Spannhoff A et al. (2011) Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. EMBO Reports. Mar 3; 12(3): 238–243.
- Feil R and Fraga M (2012) Epigenetics and the environment. Emerging patterns and implications. Nature Reviews Genetics. 13, pages 97–109.
- Lyko F et al. (2010) The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. PLOS Biology. Nov 2;8(11):e1000506.
- Fraga M et al. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. PNAS. Vol. 102, 10604–10609.
- Kempfer R and Pombo A (2020) Methods for mapping 3D chromosome architecture. Nat. Rev. Genet. 21(4):207-226.