There are over 200 distinct cell types in the human body and virtually all our cells share the same genome. One of the biggest questions in modern biology is how this astounding complexity of a multicellular organism is established during development and subsequently maintained in the adult life.

The identity of each cell is primarily defined by which genes are active in it. This results in a production of a defined set of proteins which in turn can fulfill a specified group of tasks such as, for example, the immune response in T or B cells, or the transduction of electrochemical signal in neurons. Failure to express the right set of genes can have catastrophic consequences for individual cells, tissues and the whole organism. For instance, lack of expression of a single gene can perturb the formation of immune cells and thereby compromise defense against pathogens; the formation of palate can be disturbed by a diminished expression of a key morphogen. Likewise, a profoundly disturbed pattern of genome activity can result in cancer. Understanding how gene expression is regulated is thus key to unveil the molecular bases of cell identity.

Concert action of a dedicated group of DNA regulatory elements (DREs) determines the level of gene activity. DREs include: promoters – DNA sequences that flank the start of the gene and that allow its expression; enhancers – distal DREs which activate promoters over vast genomic distances; silencers – DREs that dampen the activity of promoters; and insulators – DNA sequences that orchestrate the regulatory transactions between promoters and enhancers.

According to the prevailing models, functional interactions between DREs rely on a formation of physical contacts between them. Yet, DREs are often separated by long stretches of intervening DNA. Thus, genomes must adopt a three-dimensional structure that allows the efficient and specific formation of regulatory interactions. The development of high throughput sequencing based technologies to probe for chromatin structure, including Hi-C, opened news avenues in genomics and molecular biology and allowed, for the first time, to obtain snapshots of how chromatin is organized in the nucleus. High resolution Hi-C revealed partitioning of the mammalian genomes into domains of preferential contacts termed topologically associating domains (TADs). TAD boundaries act as strong insulators that block interactions between DREs thus favoring intra domain contacts. Current models posit that TAD structures underlie the precision in gene activation. Genetic deletion of TAD boundaries can lead to pathological conditions and a miss-regulation of gene expression. Remarkably, TAD boundaries often interact with each other forming structures reminiscent of loops.

CTCF, a unique multi-task DNA binding protein, underlies the formation of TAD boundaries and loops. Removal of CTCF protein leads to a disruption of both TADs and loops and as a consequence results in the loss of insulation. How CTCF functions to fulfil these tasks is an open question of a fundamental importance for molecular and developmental biology.

The genomic coordinates of TAD boundaries and loop anchors are overall preserved during development. However, we and others have recently found that chromatin structure undergoes a global consolidation during the earliest stages of development. The establishment of a mature chromatin topology coincides with the irreversible loss of cell plasticity whereby the embryonic stem cells commit towards one of the major three germ layers of the developing organism. Chromatin consolidation features: strengthening of TAD boundaries an in the number of chromatin loops. These discoveries suggest a general modification of the activity of CTCF dependent regulatory elements in development. However, the molecular mechanisms underlying these phenomena are currently unknown. Likewise, it remains unclear how the developmental enhancement of loops and TAD boundaries translates to the to the real-time dynamics of chromatin interactions in the developing cells.

Our research project is directed towards addressing these two questions. We will identify the factors that regulate CTCF binding and functions during the process of embryonic stem cell differentiation. We will identify CTCF partner proteins in the developing embryonic stem cells. We will take advantage of CRISPR-Cas9 genome editing technology to perturb the partners of CTCF. Next, using state-of-the-art high throughput sequencing technologies allowing to map protein binding in the genome and to infer the three-dimensional chromatin structure, we will determine the molecular bases of the implication of the most relevant CTCF partners in its functions. In a parallel research axis, we will use innovative microscopy techniques to measure the spatiotemporal dynamics of chromatin contacts in embryonic stem cells and their differentiated progeny. Our results will lead to a new understanding of how chromatin structure and gene expression is regulated during development.