# *Land of hope and dreams*

## Selection of life science and translational medicine literature

***by Marco Confalonieri***

One of the most surprising results from the Human Genome Project was the relatively small number of genes in humans. The announcement of the end of the Project was done in April 2003, essentially 2 years earlier than planned probably because the human genes were less than estimated. The haploid human genome contains approximately 20,500 protein-coding genes, the same range than in mice, not so much more than the worm *C. elegans*, and significantly fewer than had been anticipated.

Protein-coding sequences account for only a very small fraction of the genome (approximately 1.5%), and the rest is associated with non-coding RNA molecules, regulatory DNA sequences, LINEs, SINEs, introns, and sequences for which as yet no function has been elucidated. Moreover, every cell in our body contains the same genetic code. But which genes are active, or "expressed", in the cell depends on its function – whether it's a pneumocyte or a neuron, for example. Which genes are expressed is controlled by tiny bits of the genome called promoters, silencer, and enhancers. Enhancer sequences are regulatory DNA sequences that, when bound by specific proteins called transcription factors, enhance the transcription of an associated gene. Regulation of transcription is the most common form of gene control, and the activity of transcription factors allows genes to be specifically regulated during development and in different types of cells. Using CAGE (Cap Analysis of Gene Expression), a recent technology established by Piero Carninci at RIKEN, it's possible to map the sets of transcripts, transcription factors, promoters and enhancers active in the majority of mammalian primary cell types.

An international consortium of researchers known as FANTOM (Figure 1 shows its history and main publications), led by the RIKEN institute in Japan, has uncovered which promoters and enhancers are used by which cells. By looking at more than 800 human tissue samples, covering nearly all cell types, they found 44,000 enhancers and 180,000 promoters that control gene expression.

A couple of papers recently published on Nature reported two atlases of genetic regulatory elements throughout the human genome. The first paper presents an atlas of transcription start sites, where RNA polymerase begins to transcribe DNA into RNA; the second paper maps active enhancers, non-promoter stretches of DNA that upregulate the transcription of certain genes.

This is the first time that a complete analysis of the enhancer RNA is performed on such a large scale. In the future, knowledge of the enhancer and promoter usage that define different cell types raises the possibility of turning one cell type into another. It could also aid in predicting whether or not a particular cancer is going to metastatize.

**1) An atlas of active enhancers across human cell types and tissues**

Andersson R, Gebhard C, Miguel-Escalada I, et al.
*Nature 2014 Mar 27;507(7493):455-61*

*Abstract*

Enhancers control the correct temporal and cell-type-specific activation of gene expression in multicellular eukaryotes. Knowing their properties, regulatory activity and targets is crucial to understand the regulation of differentiation and homeostasis. Here we use the FANTOM5 panel of samples, covering the majority of human tissues and cell types, to produce an atlas of active, *in vivo*-transcribed enhancers. We show that enhancers share properties with CpG-poor messenger RNA promoters but produce bidirectional, exosome-sensitive, relatively short unspliced RNAs, the generation of which is strongly related to enhancer activity. The atlas is used to compare regulatory programs between different cells at unprecedented depth, to identify disease-associated regulatory single nucleotide polymorphisms, and to classify cell-type-specific and ubiquitous enhancers. We further explore the utility of enhancer redundancy, which explains gene expression strength rather than expression patterns. The online FANTOM5 enhancer atlas represents a unique resource for studies on cell-type-specific enhancers and gene regulation.

**2) A promoter-level mammalian expression atlas**

The FANTOM Consortium and the RIKEN PMI and CLST (DGT)
*Nature 2014 Mar 27;507:462-470*

*Abstract*

Regulated transcription controls the diversity, developmental pathways and spatial organization of the hundreds of cell types that make up a mammal. Using single-molecule cDNA sequencing, we mapped transcrip-

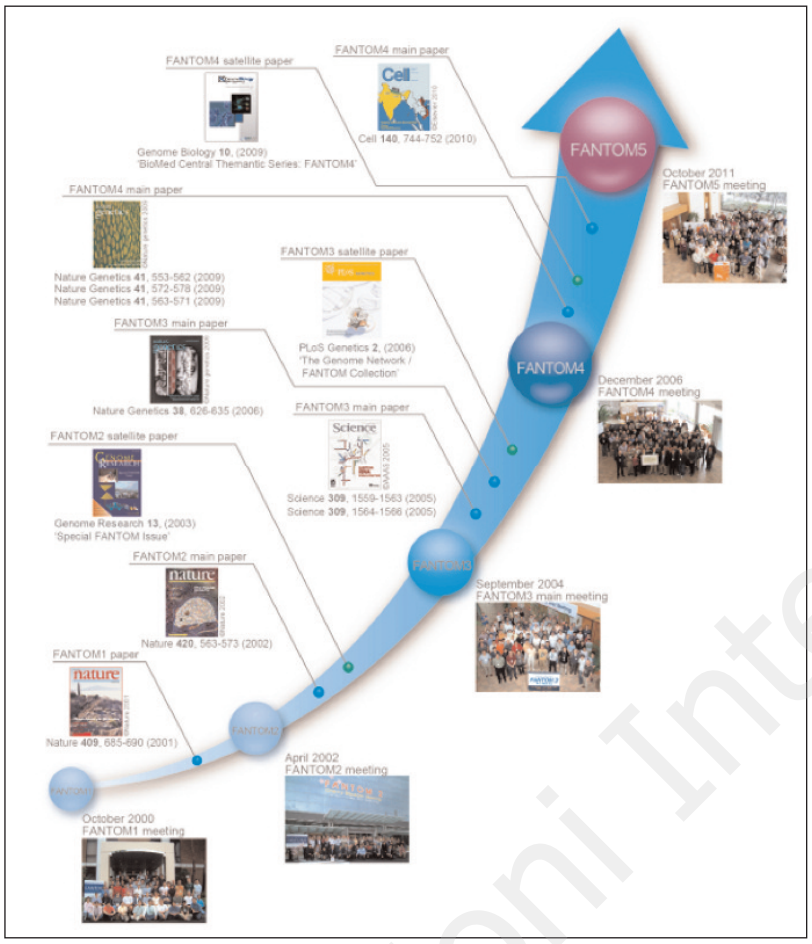


Figure 1 - FANTOM history and main publications.

tion start sites (TSSs) and their usage in human and mouse primary cells, cell lines and tissues to produce a comprehensive overview of mammalian gene expression across the human body. We find that few genes are truly 'housekeeping', whereas many mammalian promoters are composite entities composed of several closely separated TSSs, with independent cell-type-specific expression profiles. TSSs specific to different cell types evolve at different rates, whereas promoters of broadly expressed genes are the most conserved. Promoter-based expression analysis reveals key transcription factors defining cell states and links them to binding-site motifs. The functions of identified novel transcripts can be predicted by coexpression and sample ontology enrichment analyses. The functional annotation of the mammalian genome 5 (FANTOM5) project provides comprehensive expression profiles and functional annotation of mammalian cell-type-specific transcriptomes with wide applications in biomedical research.