Transcriptomics is one of the most important and relevant areas of bioinformatics. It allows detecting the genes that are expressed at a particular moment in time to explore the relation between genotype and phenotype. Transcriptomic analysis has been historically performed using microarrays until 2008 when high-throughput RNA sequencing (RNA-Seq) was launched on the market, replacing the old technique. However, despite the clear advantages over microarrays, it was necessary to understand factors such as the quality of the data, reproducibility and replicability of the analyses and potential biases.

The first section of the thesis covers these studies. First, an R package called NOISeq was developed and published in the public repository "Bioconductor", which includes a set of tools to better understand the quality of RNA-Seq data, minimise the impact of noise in any posterior analyses and implements two new methodologies (NOISeq and NOISeqBio) to overcome the difficulties of comparing two different groups of samples (differential expression). Second, I show our contribution to the Sequencing Quality Control (SEQC) project, a continuation of the Microarray Quality Control (MAQC) project led by the US Food and Drug Administration (FDA, United States) that aims to assess the reproducibility and replicability of any RNA-Seq analysis.

One of the most effective approaches to understand the different factors that influence the regulation of gene expression, such as the synergic effect of transcription factors, methylation events and chromatin accessibility, is the integration of transcriptomic with other omics data. To this aim, a file that contains the chromosomal position where the events take place is required. For this reason, in the second chapter, we present a new and easy to customise tool (RGmatch) to associate chromosomal positions to the exons, transcripts or genes that could regulate the events.

Another aspect of great interest is the study of non-coding genes, especially long non-coding RNAs (lncRNAs). Not long ago, these regions were thought not to play a relevant role and were only considered as transcriptional noise. However, they represent a high percentage of the human genes and it was recently shown that they actually play an important role in gene regulation. Due to these motivations, in the last chapter we focus, first, in trying to find a methodology to find out the generic functions of every lncRNA using publicly available data and, second, we develop a new tool (spongeScan) to predict the lncRNAs that could be involved in the sequestration of micro-RNAs (miRNAs) and therefore altering their regulation task.