La transcriptómica es una de las áreas más importantes y destacadas en bioinformática, ya que permite ver qué genes están expresados en un momento dado para poder explorar la relación existente entre genotipo y fenotipo. El análisis transcriptómico se ha realizado históricamente mediante el uso de *microarrays* hasta que, en el año 2008, la secuenciación masiva de ARN (RNA-Seq) fue lanzada al mercado y comenzó a desplazar poco a poco su uso. Sin embargo, a pesar de las ventajas evidentes frente a los *microarrays*, resultaba necesario entender factores como la calidad de los datos, reproducibilidad y replicabilidad de los análisis así como los potenciales sesgos.

La primera parte de la tesis aborda precisamente estos estudios. En primer lugar, se desarrolla un paquete de R llamado NOISeq, publicado en el repositorio público "Bioconductor", el cual incluye un conjunto de herramientas para entender la calidad de datos de RNA-Seq, herramientas de procesado para minimizar el impacto del ruido en posteriores análisis y dos nuevas metodologías (NOISeq y NOISeqBio) para abordar la problemática de la comparación entre dos grupos (expresión diferencial). Por otro lado, presento nuestra contribución al proyecto Sequencing Quality Control (SEQC), una continuación del proyecto Microarray Quality Control (MAQC) liderado por la US Food and Drug Administration (FDA) que pretende evaluar precisamente la reproducibilidad y replicabilidad de los análisis realizados sobre datos de RNA-Seq.

Una de las estrategias más efectivas para entender los diferentes factores que influyen en la regulación de la expresión génica, como puede ser el efecto sinérgico de los factores de transcripción, eventos de metilación y accesibilidad de la cromatina, es la integración de la transcriptómica con otros datos ómicos. Para ello se necesita generar un fichero que indique las posiciones cromosómicas donde se producen estos eventos. Por este motivo, en el segundo capítulo de la tesis presentamos una nueva herramienta (RGmatch) altamente customizable que permite asociar estas posiciones cromosómicas a los posibles genes, transcritos o exones a los que podría estar regulando cada uno de estos eventos.

Otro de los aspectos de gran interés en este campo es el estudio de los genes no codificantes, especialmente los ARN largos no codificantes (lncRNAs). Hasta no hace mucho, se pensaba que estos genes no jugaban ningún papel fundamental y se consideraban como simple ruido transcripcional. Sin embargo, suponen un alto porcentaje de los genes del ser humano y se ha demostrado que juegan un papel crucial en la regulación de otros genes. Por este motivo, en el último capítulo nos centramos, en un primer lugar, en intentar obtener una metodología que permita averiguar las funciones generales de cada lncRNA haciendo uso de datos ya publicados y, en segundo lugar, generamos una nueva herramienta (spongeScan) que permite predecir qué lncRNAs podrían estar secuestrando determinados micro-RNAs (miRNAs), alterando así la regulación llevada a cabo por estos últimos.