

Repositorios públicos de datos de RNA-Seq

TCGA (*The Cancer Genome Atlas* - <https://portal.gdc.cancer.gov>): consorcio americano creado por el Instituto Nacional del Cáncer (NCI) y el Instituto Nacional de Investigación del Genoma Humano (NHGRI) que ha conseguido caracterizar molecularmente más de 20.000 cánceres primarios y compararlos con muestras sanas en 33 tipos tumorales diferentes. En su plataforma, disponible de manera pública a la comunidad investigadora, dispone de más de 2.5 petabytes de datos genómicos, epigenómicos, transcriptómicos y proteómicos. El uso de este repositorio para el estudio de expresión génica permite conocer los valores de expresión de un amplio abanico de genes en distintos tipos tumorales y compararlo en función de distintas variables clínicas recogidas o en estudios de caso-control.

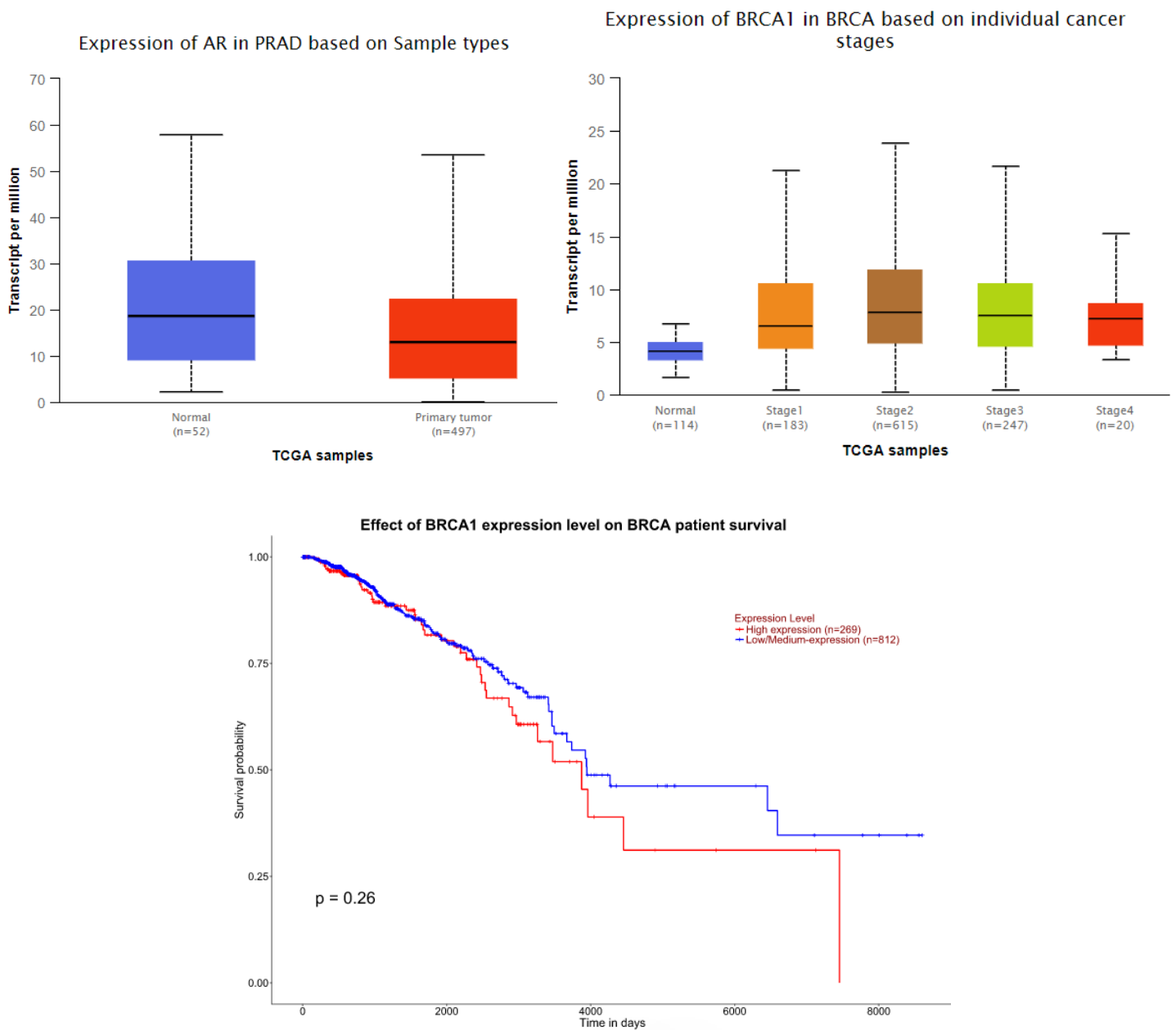
GTEX (*Genotype-Tissue Expression* - <https://gtexportal.org>): recurso público integral enfocado en el estudio de la expresión génica y su regulación de una manera específica en cada tejido. Los datos de expresión, QTLs e imágenes histológicas disponibles en su plataforma proceden de un total de 54 tejidos diferentes estudiados en muestras de casi 1000 individuos mediante técnicas como WGS, WES y RNA-Seq.

GEO (*Gene Expression Omnibus* - <https://www.ncbi.nlm.nih.gov/geo>): repositorio público internacional que cuenta con datos resultantes de estudios de microarray, de NGS (Next Generation Sequencing) y otras formas de datos de genómica funcional de alto rendimiento. Además, dispone de herramientas para facilitar a los usuarios la consulta y descarga de experimentos; así como, el análisis de perfiles de expresión génica de interés.

Plataforma	Tipos de datos disponibles	Tipo de muestras	Magnitud
TCGA	Genómica, transcriptómica, epigenómica y proteómica	Muestras de tejido sano y de distintos tipos tumorales	> 10.000 muestras
GTEX	Variantes genómicas, transcriptómica	Muestras de distintos tipos de tejido humano sano	≈ 1.000 individuos (54 tejidos diferentes)
GEO	Microarray, RNA-Seq, Variantes Genómicas, Arrays de proteínas		

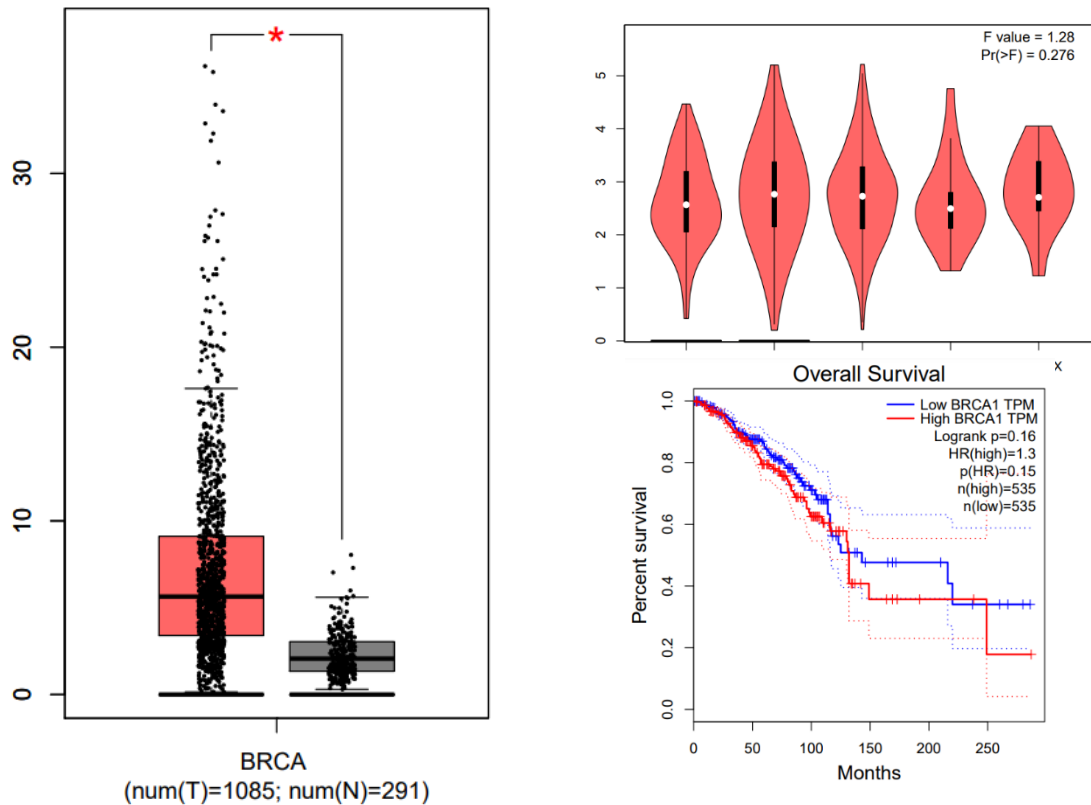
Herramientas de análisis para datos públicos de RNA-Seq:

UALCAN (*University of Alabama at Birmingham CANcer data analysis portal* - <https://ualcan.path.uab.edu/>): se trata de un recurso web completo y fácil de usar que te permite analizar datos ómicos sobre el cáncer. Así, accede a datos disponibles en TCGA, MET500, COTAC, CBTTTC o GTEx, entre otros, y devuelve gráficos de alta calidad usando JavaScript y CSS. Esto permite a los usuarios conocer el perfil de expresión e información de supervivencia del paciente para genes que codifican proteínas, miRNA o lincRNA, pudiendo identificar biomarcadores o realizar una validación in silico de posibles genes de interés.

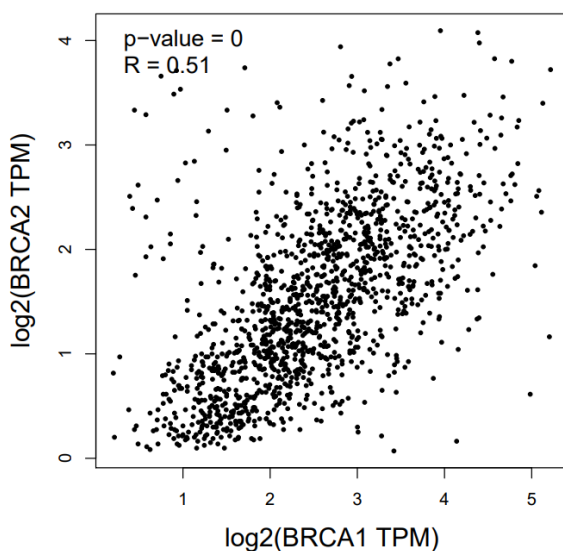


Ejemplos de gráficas obtenidas mediante la plataforma UALCAN referente a la expresión del gen BRCA1 en carcinoma de mama invasivo (BRCA), empleando los datos recogidos

GEPIA (*Gene Expression Profiling Interactive Analysis* - <http://gepia.cancer-pku.cn/index.html>): servidor web interactivo que accede a los datos disponibles en TCGA y GTEx y permite el análisis de datos de expresión por RNA-Seq procedente de 9.736 muestras tumorales y 8.587 muestras de tejido sano. Asimismo, cuenta con funciones personalizables como el análisis de expresión diferencial tumor/sano, perfiles en función del tipo tumoral o del estadio de la enfermedad, análisis de supervivencia o análisis de correlación génica.

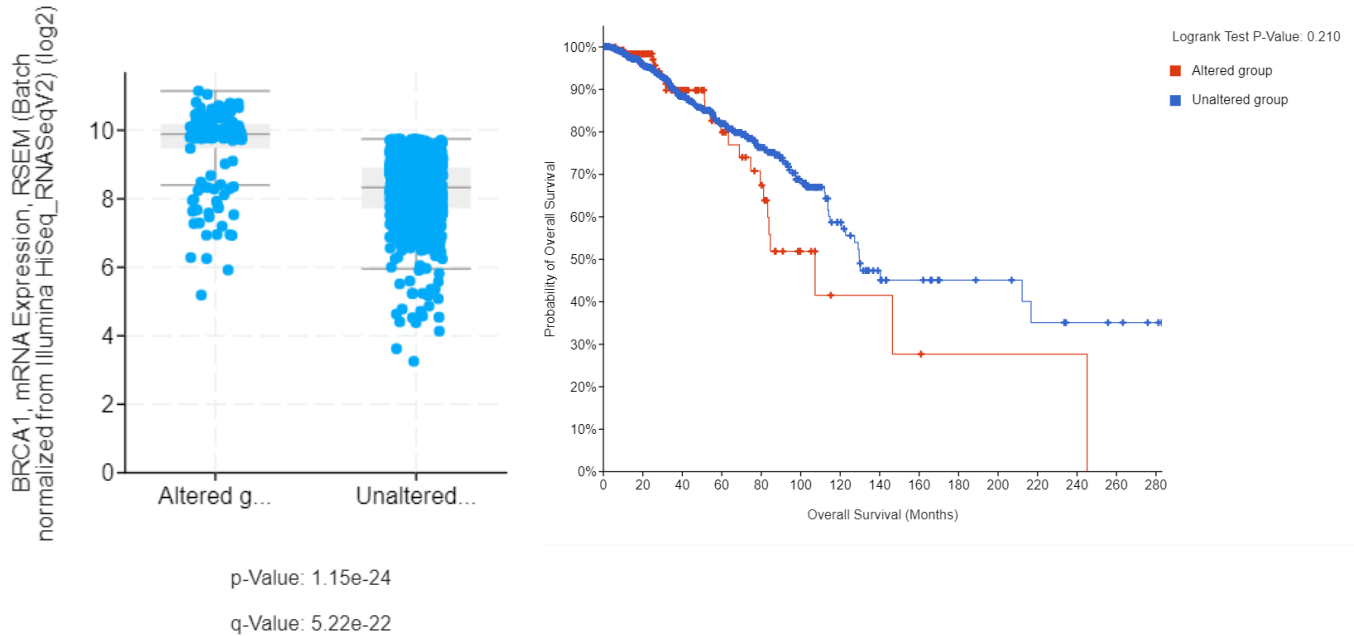


Ejemplos de gráficas obtenidas mediante la plataforma GEPIA referente a la expresión del gen BRCA1 en BRCA, empleando los datos recogidos en TCGA

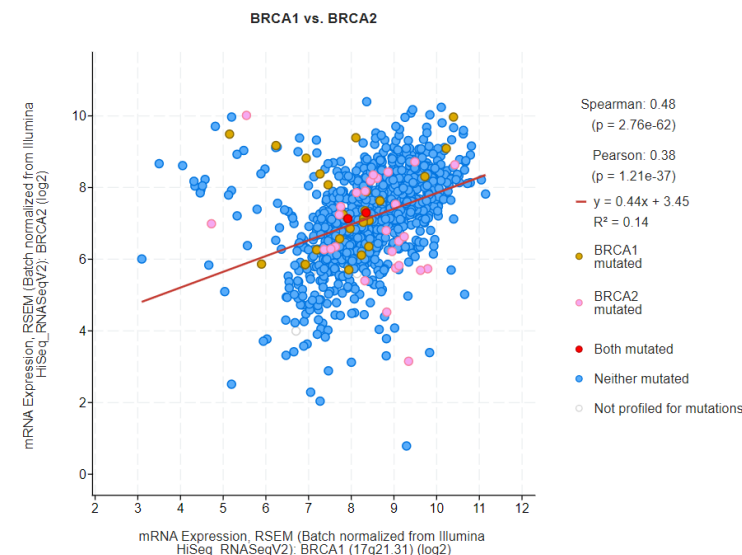


Ejemplo de gráfica obtenida mediante la plataforma GEPIA referente a la correlación en la expresión de los genes BRCA1 y BRCA2 tejido mamario sano y tumoral empleando los datos recogidos en TCGA y GTEx

cBioportal for Cancer Genomics (<https://www.cbioportal.org>): recurso web que permite la exploración interactiva de conjuntos de datos multidimensionales de la genómica del cáncer, permitiendo un acceso rápido, intuitivo y de alta calidad a los perfiles moleculares y las variables clínicas de los proyectos de genómica del cáncer a gran escala.



Ejemplos de gráficas obtenidas mediante la plataforma cBioportal referente a la expresión del gen BRCA1 en BRCA, empleando los datos recogidos en TCGA



Ejemplo de gráfica obtenida mediante la plataforma cBioportal referente a la correlación en la expresión de los genes BRCA1 y BRCA2 tejido mamario sano y tumoral empleando los datos recogidos en TCGA.