.UBAexactas
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

Licenciatura en Ciencias de Datos
Propuesta de tema de tesis

Uso de datos de genómica funcional y redes regulatorias de la expresión génica para analizar la heterogeneidad de células tumorales mamarias

El tratamiento el cáncer de mama triple negativo (TNBC) representa un gran desafío en oncología debido a su extensa heterogeneidad molecular, histológica y clínica. Sin embargo, los avances recientes en la secuenciación de RNA (scRNA-seq) y de mapeo de elementos regulatorios activos (scATAC-seq) en células únicas proporcionan información sin precedentes sobre la heterogeneidad celular asociada a cancer. La integración de los datos genómicos en el contexto de redes de regulación de la expression génica permite la inferencia de relaciones regulatorias, pudiendo informarnos sobre cómo la desregulación de determinados genes contribuye a la heterogeneidad y progresión de la enfermedad. Por ejemplo, nuestros resultados sugieren un rol del complejo remodelador de la cromatina BAF en las características de malignidad el TNBC, lo que lleva a preguntarnos sobre las posibles consecuencias de la presencia de heterogeneidad en los componentes de este complejo en para la biología de este tipo de tumores.

Palabras clave: `genómica, cáncer, scRNA-seq, redes regulatorias`

Conocimientos deseables

R y/o Python, inferencia estadística

¿Qué podría aprender quien realice esta tesis?

La/el estudiante se familiarizará con el manejo de datos genómicos, tanto en bulk como en células únicas, a integrarlos y a realizar extracción de conocimiento biologíco de los mismos aprovechando la existencia de numerosas bases de datos. También podrá aprender la inferencia de redes regulatorias por ingeniería reversa a partir de datos genómicos, así como su uso para analizar el funcionamiento de las células.

Dirección de la tesis

Schor, Ignacio
IFIBYBNE - DFBMC (FCEN-UBA)

Contacto: *naschor@gmail.com*

Más información en el pdf a continuación.

Uso de datos de genómica funcional y redes regulatorias de la expresión génica para analizar la heterogeneidad de células tumorales mamarias

Using functional genomics data and gene regulatory networks to assess breast cancer cell heterogenity

Introduction:

Triple-negative breast cancer (TNBC) is notorious for its significant heterogeneity, encompassing a spectrum of tumors with diverse genetic mutations, gene expression patterns, and histological subtypes. Unlike other breast cancer types with well-defined molecular drivers, TNBC lacks a singular pathway, posing challenges for targeted therapy development. This heterogeneity extends to clinical outcomes, with TNBC patients exhibiting varied prognoses influenced by factors such as age, ethnicity, and genetic predispositions like BRCA mutations. Current therapeutic strategies for TNBC primarily rely on conventional chemotherapy, which, while effective to some extent, is limited in its ability to target the specific molecular vulnerabilities of diverse TNBC subtypes.

Central to TNBC's complexity is the BAF (BRG1/BRM-associated factor) complex, a chromatin remodeling entity crucial for regulating gene expression by modifying DNA accessibility. The BAF complex exists in multiple forms with varying subunit compositions, each influencing gene activation or repression differently. Dysregulation of BAF complex subunits, characterized by frequent mutations in TNBC, disrupts normal gene expression patterns, promoting tumor growth and progression.

The emergence of single-cell RNA sequencing (scRNA-seq) has provided a powerful tool to dissect TNBC heterogeneity at a cellular resolution. By analyzing gene expression profiles of individual cells within TNBC tumors, scRNA-seq identifies distinct subpopulations characterized by unique molecular signatures. This technology not only reveals the presence of aggressive versus indolent tumor cells coexisting but we think could elucidate how variations in BAF complex activity contribute to these different phenotypes. Integration of scRNA-seq with other single-cell technologies, such as single-cell Assay for Transposase-Accessible Chromatin (scATAC-seq), offers a comprehensive view of TNBC biology by linking gene expression patterns with chromatin accessibility profiles.

During the last decade, gene co-expression matrices have been used to construct gene regulatory networks (GRN), which have uncovered key centers of gene regulation. The study of GRN in cancer context has highlighted critical master regulators and groups of genes important for malignancy, most of them not evidenced by typical pathway enrichment analysis. Advancement of scRNA-seq technology has incorporated a new viewpoint into the construction of GRN, dissecting the architecture of regulation in different cell types. Using GRN derived

from scRNA-seq data from TNBC patients could not only provide insights into tumor biology but also allows to interrogate the role and centrality of BAF complex on cancer's network.

Objectives:

- Identify the cellular populations of TNBC and their heterogeneity using patient data.
- Explore the main differences of cellular pathways and the behavior of the BAF complex in the different TNBC populations identified.
- Obtain a network of regulatory relationships between genes that accounts for the expression state of the different cellular populations.

Methods:

We will use available scRNA-seq data from tumor biopsies from TNBC patients, and perform basic quality control on it. We will apply existing tools and packages for scRNA-seq data analysis, such as the R based Seurat suite, to preprocess the data from each patient and integrate all patients and data sources into one data set. We will add external annotations and gene sets to identify cell subpopulations, specially using genes associated with relevant processes such as stemness, cell cycle control, remodeling of extracellular matrix, epithelial-mesenchymal state, etc. We will incorporate scATAC-seq data to leverage the value of our findings in subsequent objectives. We will address the BAF complex signature as a whole and dissect the expression of each of its components. We will compare the gene signature of sub-populations having different BAF signatures, by means of differential expression analysis and gene set analysis to get the cellular pathways differentially enriched on them. We will test available methods to reverse engineer GRN from scRNA-seq data (and optionally adding scATAC-seq data), such as SCENIC and PISCES. After evaluation of the GRNs, we will look for master regulators differentially activated in each cancer cell sub-population and assess the role and centrality of BAF to explain differences among sub-populations.