MULTI-OMICS INTEGRATION OF TCGA-BRCA BY MOFA

MODA TEAM

[COMPANY NAME] [Company address]

Introduction

The Cancer Genome Atlas (TCGA) Breast Invasive Carcinoma (BRCA) project is a comprehensive collection of genomic, transcriptomic, and clinical data from breast cancer patients. This data provides a valuable resource for researchers interested in understanding the molecular mechanisms underlying breast cancer and developing novel therapeutic strategies. Multi-omic integration is a powerful approach for gaining a more complete picture of the complex biological processes that drive cancer development and progression. In this study, data integrated from TCGA-BRCA project across different omic to identify key molecular drivers and biomarkers of breast cancer to help develop more effective personalized treatments.

The PAM50 classification was used for TCGA breast cancer patients. PAM50 is a 50-gene signature that classifies breast cancer into five intrinsic molecular subtypes: Luminal A, Luminal B, human epidermal growth factor receptor 2 (HER2)-enriched, Basal-like and Normal-like. Each of the five molecular subtypes varies by their biological properties and prognoses.

Methodology:

This project involved the analysis of breast cancer data derived from the TCGA-BRCA project using R. Multi-omics data (RNAseq STAR counts data, microRNA data proteomics and clinical data) of 495 primary solid tumor samples was downloaded and processed using TCGABiolinks R package. The RNAseq data were divided into two matrices, one for the long noncoding RNAs and one for the mRNA. We started with features quality control, normalization, and scaling. MOFA2 (Multi-Omics Factor Analysis) R package was used for multi-omics data integration, which uses an unsupervised approach to discover the patterns that drive variation across

different molecular layers. Features included in MOFA analysis were 2281 for mRNA, 1572 for miRNA, 378 for Proteome and lncRNA for 2228.

Following MOFA, gene set enrichment analysis was performed for mRNA and Proteome using REACTOME and C5(GO: BP) from MSIGDB. On the other hand, overrepresentation analysis (ORA) was conducted on the miRNA and lncRNA. First, differential expression analysis was performed using DESeq2 Package, and the samples were grouped based on their position in factor one of MOFA, positive and negative groups. Then, ORA was performed using TAM-2 tools for miRNA and PANTHER web tool for lncRNA. The ORA analysis used REACTOME, GO: BP, and GO: MF.

Findings

The best MOFA factor was factor (1), which has the most variance contribution from each omic, followed by factor (3). There is no correlation detected between the generated factors. Factor (1) clearly differentiated between breast cancer subtypes. The basal was positive to factor one, while the luminal was negative.

The main findings are separated into two main sections as follow:

a. Features positive to factor one

The resulting enriched pathways from the four omics were mainly related to four main terms:

1. Immune modulation.

mRNA

REACTOME_INTERFERON_GAMMA_SIGNALING
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM
REACTOME_ANTIGEN_PROCESSING_CROSS_PRESENTATION

REACTOME_NEUTROPHIL_DEGRANULATION
REACTOME_SARS_COV_2_HOST_INTERACTIONS
GOBP_CYTOKINE_PRODUCTION
GOBP_CELLULAR_RESPONSE_TO_TYPE_II_INTERFERON
GOBP_HUMORAL_IMMUNE_RESPONSE
GOBP_INFLAMMATORY_RESPONSE
GOBP_POSITIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS
GOBP_TUMOR_NECROSIS_FACTOR_SUPERFAMILY_CYTOKINE_PRODUCTIO
N
GOBP_POSITIVE_REGULATION_OF_CYTOKINE_PRODUCTION

Proteome
GOBP_RESPONSE_TO_CYTOKINE
GOBP_PHAGOCYTOSIS
GOBP_LYMPHOCYTE_APOPTOTIC_PROCESS
GOBP_B_CELL_ACTIVATION
GOBP_CYTOKINE_MEDIATED_SIGNALING_PATHWAY
GOBP_T_CELL_PROLIFERATION
GOBP_NEGATIVE_REGULATION_OF_PROTEIN_SERINE_THREONINE_KINASE
_ACTIVITY
GOBP_REGULATION_OF_CELL_CYCLE_G2_M_PHASE_TRANSITION
GOBP_POSITIVE_REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_SIGNAL
ING

miRNA
Immune Response
Inflammation
Immune System(Xiao's Cell2010)
T-Cell Activation
Neutrophil Differentiation

2. Cell cycle and Mitotic division.

miRNA
REACTOME_CELL_CYCLE_MITOTIC

Proteomic
REACTOME_CELL_CYCLE_MITOTIC
REACTOME_CELL_CYCLE
REACTOME_CELL_CYCLE_CHECKPOINTS
REACTOME_MITOTIC_G1_PHASE_AND_G1_S_TRANSITION
GOBP_MITOTIC_CELL_CYCLE_PROCESS
GOBP_MITOTIC_CELL_CYCLE_CHECKPOINT_SIGNALING

miRNA
Cell Proliferation
Cell Cycle
Regulation of Akt Pathway

3. Cell migration

mRNA
REACTOME_COLLAGEN_DEGRADATION
REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX
GOBP_REGULATION_OF_SUBSTRATE_ADHESION_DEPENDENT_CELL_SPREADING
GOBP_POSITIVE_REGULATION_OF_MONONUCLEAR_CELL_MIGRATION
GOBP_COLLAGEN_METABOLIC_PROCESS
GOBP_CELL_ADHESION
GOBP_EXTRACELLULAR_MATRIX_DISASSEMBLY

Proteome
GOBP_REGULATION_OF_CELL_CELL_ADHESION
GOBP_CELL_CELL_ADHESION
miRNA
cilium-dependent cell motility

4. Response to microbial infection

mRNA
REACTOME_INFLUENZA_INFECTION
REACTOME_SARS_COV_2_MODULATES_HOST_TRANSLATION_MACHINERY
REACTOME_SARS_COV_1_MODULATES_HOST_TRANSLATION_MACHINERY
REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS
GOBP_ANTIMICROBIAL_HUMORAL_RESPONSE
GOBP_DEFENSE_RESPONSE_TO_OTHER_ORGANISM

GOBP_ANTIMICROBIAL_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_ANTIMICROBIAL_PEPTIDE
GOBP_DEFENSE_RESPONSE_TO_GRAM_POSITIVE_BACTERIUM

Proteome
GOBP_RESPONSE_TO_BACTERIUM
GOBP_RESPONSE_TO_MOLECULE_OF_BACTERIAL_ORIGIN

miRNA
Response to Hypoxia

b. Features Negative to factor one

The resulted enriched pathways from the four omics were mainly related to three main terms:

1. **Lipid metabolism**

mRNA

REACTOME_SPHINGOLIPID_METABOLISM
REACTOME_METABOLISM_OF_LIPIDS
GOBP_MEMBRANE_LIPID_BIOSYNTHETIC_PROCESS
GOBP_LIPID_METABOLIC_PROCESS
GOBP_LIPID_OXIDATION
GOBP_STEROID_METABOLIC_PROCESS
GOBP_LIPID_BIOSYNTHETIC_PROCESS
GOBP_LIPID_MODIFICATION
GOBP_MEMBRANE_LIPID_METABOLIC_PROCESS
GOBP_STEROID_BIOSYNTHETIC_PROCESS
GOBP_SPHINGOLIPID_METABOLIC_PROCESS
GOBP_REGULATION_OF_LIPID_METABOLIC_PROCESS
GOBP_FATTY_ACID_METABOLIC_PROCESS
GOBP_CELLULAR_LIPID_METABOLIC_PROCESS

Proteome

GOBP_LIPID_METABOLIC_PROCESS
GOBP_INOSITOL_LIPID_MEDIATED_SIGNALING

miRNA

Adipocyte Differentiation
Lipid Metabolism

2. **Hormone-related pathways**

mRNA

GOBP_MAMMARY_GLAND_EPITHELIUM_DEVELOPMENT
GOBP_GLANDULAR_EPITHELIAL_CELL_DIFFERENTIATION
GOBP_HORMONE_TRANSPORT

Proteome

REACTOME_SIGNALING_BY_INSULIN_RECEPTOR

REACTOME_INSULIN_RECEPTOR_SIGNALLING_CASCADE
GOBP_INSULIN_LIKE_GROWTH_FACTOR_RECEPTOR_SIGNALING_PATHWAY
GOBP_MAMMARY_GLAND_DEVELOPMENT
GOBP_MAMMARY_GLAND_EPITHELIUM_DEVELOPMENT
GOBP_RESPONSE_TO_STEROID_HORMONE
GOBP_CELLULAR_RESPONSE_TO_INSULIN_STIMULUS

miRNA

Hormone-mediated Signaling Pathway
mammary gland epithelial cell proliferation (GO:0033598)
response to interferon-alpha (GO:0035455)

lncRNA

Insulin processing (R-HSA-264876)
mammary gland epithelial cell proliferation (GO:0033598)
hormone metabolic process (GO:0042445)

3. **Pathways related to immunity and migration.**

mRNA

Cell Proliferation
T-Cell Differentiation
Extracellular Matrix Remodeling
Inflammation

miRNA

Cell Cycle
Cell Differentiation
T-Cell Differentiation
Extracellular Matrix Remodeling
Inflammation

LncRNA

Degradation of the extracellular matrix (R-HSA-1474228)
negative regulation of viral entry into host cell (GO:0046597)