

Relatório: Contagem, expressão e diversidade

Jean Resende

2023-07-05

Introdução

Em alguma reunião apresentei dois artigos que trabalharam com dados de RNA-Seq do TCGA. Apresentei os métodos que eles utilizaram e os resultados obtidos, a fim de direcionar nossas análises seguindo nessa linha de pesquisa. Um desses artigos (Tumor-Infiltrating B- and T-Cell Repertoire in Pancreatic Cancer Associated With Host and Tumor Features) focou em um tipo de câncer e nos dois tipos de receptores (TCR e BCR). Comecei a seguir sua metodologia para ter uma direção do que fazer, como fazer e como interpretar os resultados. Além disso, estou estudando outros artigos que seguem essa linha de pesquisa para incorporar outras análises conforme o avanço nas discussões.

Na última semana foquei em como apresentar a diferença de TCR e BCR em ACC considerando o grupo esteroide (LSP=low steroid e HSP=high steroid), assim como o artigo de referência que avaliou no primeiro momento a diferença desses receptores para tumor e não tumor. Os autores trabalharam com expressão e diversidade. Para a expressão dos receptores eles definiram como $TCR/BCR = M_i/(N_i + M_i)$ em que i corresponde a cada amostra, M é o número de leituras que mapeiam para um BCR/TCR específico e N é o número de leituras que mapeiam para qualquer outra coisa no transcriptoma. Esse cálculo é usado para definir a normalização da expressão desses receptores pelo total de leituras, neste e outros trabalhos (Pineda et al., 2021, Yu et al., 2022, Selitsky et al., 2019). Para a diversidade, os autores utilizaram o cálculo da Entropia de Shannon. Também calculei a Entropia de Shannon, seguindo o algoritmo dos desenvolvedores do TRUST4 (ferramenta que estou utilizando para extrair e quantificar os receptores).

Resultados

Iniciei a avaliação da presença de TCR e BCR nos grupos LSP e HSP de ACC comparando as contagens brutas, ou seja, a quantidade de vezes que o TRUST4 encontrou uma mesma sequência de TCR/BCR (Figura 1A). Calculei a expressão de TCR e BCR dividindo o counts pelo número de leituras da amostra conforme a fórmula mencionada acima (Figura 1B). Em seguida calculei a diversidade medida pela entropia de Shannon a fim de considerar não apenas o número, mas também a distribuição de tamanho dos clones (Figura 1C).

Observei uma expressão e diversidade maior de IGH, IGK, IGL, TRA, e TRB em amostras LSP, condizendo com o heatmap que discutimos anteriormente. Para TRD e TRG as contagens eram baixas, a maioria foi zero. Essa baixa contagem de TRD e TRG também foi observado em outros trabalhos inclusive este artigo de referência.

A expressão do TCR foi mais baixa do que BCR em LSP [média (TCR) = 5,04e-07; média (BCR) = 5,35e-05; $p = 1,04e-07$], assim como em HSP [média (TCR) = 2,78e-08; média (BCR) = 1,28e-05; $p = 1,33e-09$]. Embora a contagem, expressão e diversidade tenha sido significativamente correlacionada, especialmente em HSP (Figuras 1D, E).

Vale ressaltar que a correlação em LSP foi mais alta que as correlações em HSP

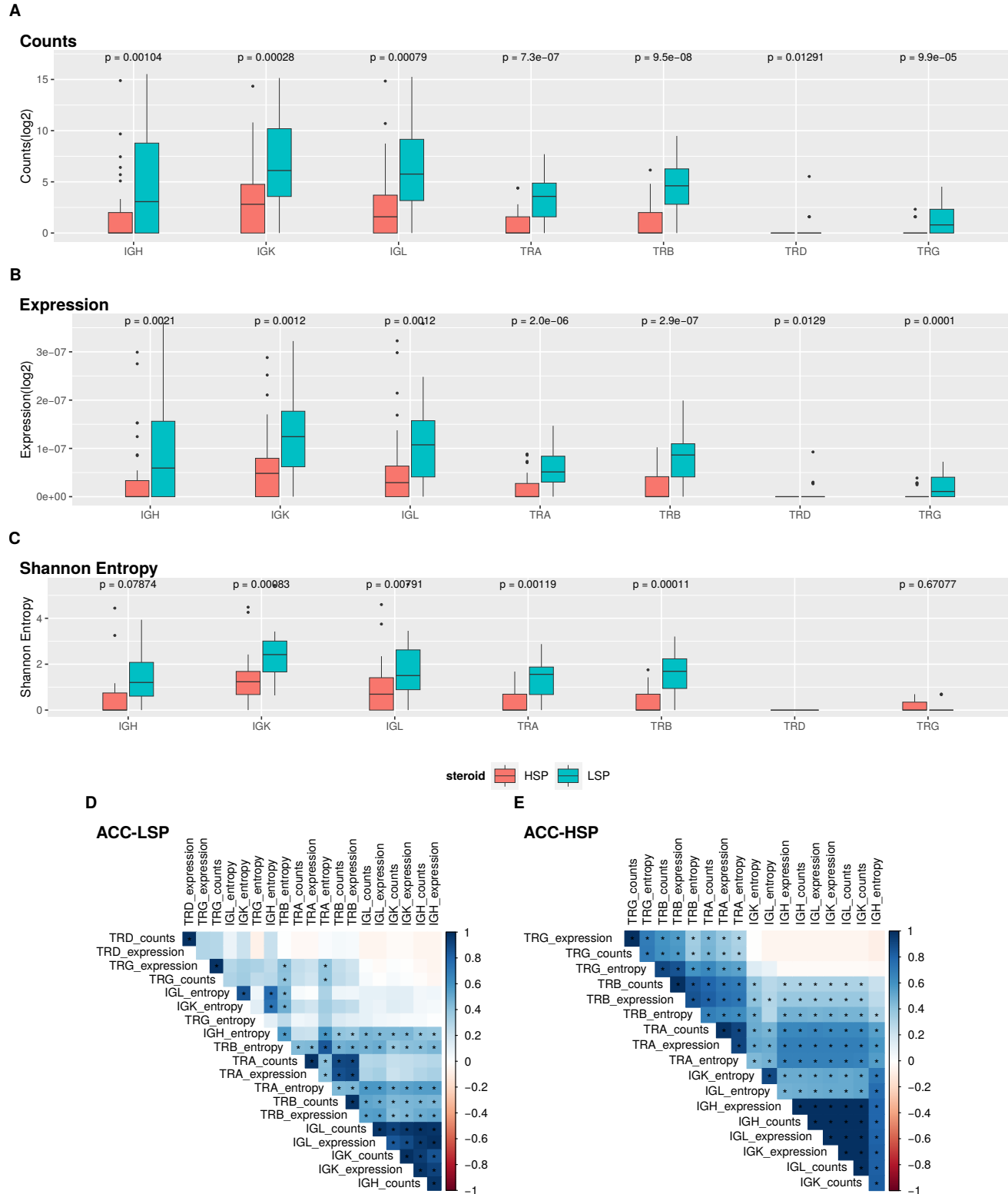


Figure 1: Análise de diversidade. Boxplots mostrando contagem de IGH, IGK, IGL TRA e TRB (A), expressão (B) e entropias de Shannon (C). Os valores de p correspondem às diferenças estatísticas do teste de Wilcoxon. Gráfico de correlação para LSP (D) e HSP (E). Os números correspondem ao coeficiente de Pearson e os asteriscos indicam as correlações com valor p abaixo de 0,05.