

BUILDING SEX-SPECIFIC SYNTHETIC LETHALITY NETWORK IN CANCER - TEMP NAME

ALEXANDER TURCO

July 26, 2023

¹ Department of, University, , ON, Canada

Contents

| | |
|---|----------|
| Abstract | 3 |
| Introduction/Background Information | 4 |
| What are Synthetic Lethal Interactions? | 4 |
| Synthetic Lethal Interactions are Harnessed for Precision Oncology | 4 |
| Building Pan-Cancer Synthetic Lethality Networks | 4 |
| Human Sex Differences add An Additional Layer of Complexity | 4 |
| Building Pan-Cancer Synthetic Lethality Networks in a Sex Specific Manner | 4 |
| Materials and Methods | 5 |
| TCGA Data | 5 |
| Pre-filtering and Normalization of Raw RNA-seq Count Data | 6 |
| Data Quality Assessment (PCA, NPMANOVA) | 6 |
| Differential Gene Expression Analysis with DESeq2 | 6 |
| Results | 6 |
| Discussion | 6 |
| References | 7 |

Abstract

Introduction/Background Information

What are Synthetic Lethal Interactions?

Synthetic Lethal Interactions are Harnessed for Precision Oncology

Building Pan-Cancer Synthetic Lethality Networks

Human Sex Differences add An Additional Layer of Complexity

Building Pan-Cancer Synthetic Lethality Networks in a Sex Specific Manner

Materials and Methods

TCGA Data

RNA sequencing (RNA-seq) data was obtained from The Cancer Genome Atlas (TCGA). Raw STAR (Spliced Transcripts Alignment to a Reference) aligned counts for tumor tissue and healthy tissue samples were collected. The Cancer Genome Atlas contains genomic information which spans 33 cancer types. For the purpose of this study, only 12 of the 33 cancer types were considered. We first filtered out sex-biased cancers which include breast invasive carcinoma (BRCA), cervical cell carcinoma (CESC), ovarian serous cystadenocarcinoma (OV), prostate adenocarcinoma (PRAD), testicular germ cell tumors (TGCT), uterine corpus endometrial carcinoma (UCEC), and uterine carcinosarcoma (UCS). The reason for this was due to the fact that we are already aware of sex biases in these cancer types. Next, we filtered out blood cancers as well as cancers which lacked normal tissue gene expression samples. This included adrenocortical carcinoma (ACC), lymphoid neoplasm diffuse large b-cell lymphoma (DLBC), glioblastoma multiforme (GBM), acute myeloid leukemia (LAML), and brain lower grade glioma (LGG). The reason for this was due to the fact that we did not have adequate control samples to compare to tumor samples. Finally, we filtered out cancers that did not have any matching pairs of samples (NT and TP from same individual), as well as cancers that had less than 10 matched sample pairs across both males and females. This included mesothelioma (MESO), skin cutaneous melanoma (SKCM), thymoma (THYM), uveal melanoma (UVM), cholangiocarcinoma (CHOL), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), rectum adenocarcinoma (READ), and sarcoma (SARC). We selected matched tumor-normal sample pairs to help control for genetic background and other individual-specific factors that could influence gene expression. Once cancer types were selected, two pan-cancer raw count gene expression matrices were created, one for males and one for females.

Table 1: List of 12 TCGA Cancer Types With Number of Matched Tumor-Normal Samples in Males and Females.

| TCGA information | | |
|------------------|------------------------|----------------------|
| Cancer Type | Matched Female Samples | Matched Male Samples |
| BLCA | 9 | 10 |
| COAD | 21 | 20 |
| ESCA | 5 | 8 |
| HNSC | 14 | 29 |
| KICH | 12 | 13 |
| KIRC | 20 | 52 |
| KIRP | 10 | 22 |
| LIHC | 22 | 28 |
| LUAD | 34 | 24 |
| LUSC | 14 | 37 |
| STAD | 10 | 23 |
| THCA | 42 | 17 |
| TOTAL | 213 | 283 |

Pre-filtering and Normalization of Raw RNA-seq Count Data

Raw count gene expression matrices for males and females were pre-filtered to remove genes unlikely to exhibit differential expression. For each matrix, we calculated the 90th quantile of overall gene expression as a threshold. For each gene in the matrix, we checked to see whether its expression was greater than the threshold in at least 1 sample. We removed genes where no samples showed an expression value greater than the quantile threshold.

The pre-filtered matrices were then normalized using the DESeq2 package in R (Love et al. [2014](#)). RNA-seq data must be normalized in order to account for factors that prevent the direct comparison of expression measures. The DESeq2 R package employs a median of ratios normalization method to account for the inherent biases associated with RNA-seq data.

Data Quality Assessment (PCA, NPMANOVA)

Differential Gene Expression Analysis with DESeq2

Results

Discussion

References

- Cheng K, Nair N U, Lee J S, and Ruppin E (2021). Synthetic lethality across normal tissues is strongly associated with cancer risk, onset, and tumor suppressor specificity. *Science advances* 7(1), eabc2100.
- Lee J S, Nair N U, Dinstag G, Chapman L, Chung Y, Wang K, Sinha S, Cha H, Kim D, Schperberg A V, et al. (2021). Synthetic lethality-mediated precision oncology via the tumor transcriptome. *Cell* 184(9), 2487–2502.
- Love M I, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15(12), 1–21.
- Shen J P and Ideker T (2018). Synthetic lethal networks for precision oncology: promises and pitfalls. *Journal of molecular biology* 430(18), 2900–2912.
- Shohat S and Shifman S (2022). Gene essentiality in cancer cell lines is modified by the sex chromosomes. *Genome Research* 32(11-12), 1993–2002.