Supplementary manuscript of

**Multi-modal optimization to identify personalized biomarkers for detecting early warning signal of individual patients in cancer**

Jing Liang1, Zong-Wei Li1, Cai-Tong Yue1, Han Cheng2, Wei-Feng Guo1\*

1 School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China

2 School of Life Sciences, Zhengzhou University, Zhengzhou 450001, China

\* Corresponding author(s).

Email: [guowf@zzu.edu.cn](mailto:liangjing@zzu.edu.cn,)

**Part** **A:**  **The analysis of enrichment of cancer driver genes of PDNBs.**

According to the Cancer Gene Census, there are 23 ,18 genes are annotated as driver genes of BRCA and LUNG cancer tissue, respectively. We found that MMPDNB found one breast cancer driver gene in BRCA patients and six lung cancer driver genes in LUNG patients by analyzing PDNBs with the largest early warning score. By analyzing all PDNBs of each individual patients, we found that MMPDNB found two breast cancer driver genes in BRCA patients and fifteen lung cancer driver genes in LUNG patients. A key point of PDNBs with higher enrichment of cancer driver genes is that the number of genes constituting PDNB is small, which also proves the formation of PDNBs does not require a large number of genes. Although the number of cancer driver genes identified by some algorithms more than MMPDNB, the number of their PDNBs genes is much greater than that of MMPDNB.

**Table S1. Breast and lung cancer driver genes and Driver genes identified by MMPDNB**

|  |  |  |
| --- | --- | --- |
|  | **Breast cancer** | **Lung cancer** |
| **Driver genes** | |  | | --- | | AKT1\ARID1A | | ARID1B\BAP1 | | BRCA2\ CASP8 | | CDH1\CDKN1B | | CDKN2A\ CTCF | | ERBB2\ ESR1 | | FOXA1\ GATA3 | | MAP2K4\ MAP3K1 | | NCOR1\ PIK3CA | | RB1\ SALL4 | | SMARCD1\ TBX3 | | TP53 | | |  | | --- | | BRAF\ CDKN2A | | EGFR\ KDR | | KEAP1\ KRAS | | LRIG3\ MAP2K1 | | RBM10\ RET | | SMARCA4\ STK11 | | TP53\ DROSHA | | FGFR2\ NFE2L2 | | NOTCH1\ RAD21 | |
| **Driver genes contained in PDNB with the highest early warning signal score** | AKT1 | RET\ RBM10  FGFR2\ STK11  BRAF\ CDKN2A |
| **Driver genes contained in all PDNBs** | AKT1、ARID1B | SMARCA4、KDR  TP53、RET  CDKN2A、FGFR2  MAP2K1、NFE2L2  NOTCH1、RBM10  STK11、BRAF  CDKN2A、EGFR  KRAS、 |

**Part B: the analysis of drug targeted gene in multi-modal PDNB of LUSC and LUAD data.**

For LUSC patient samples, there are multi-modal PDNBs of twelve patient samples contained (41% of early patient sample with multi-modal PDNBs) drug targets for Lung Squamous Cell Carcinoma. The differential genes of multi-modal PDNBs belonging to the early stage of LUSC include 14 drug target genes, as shown in **Table S2**. Drug target genes include but not limited to PMS2, MAPK11 and RHOA. Mutations in PMS2 have been associated with hereditary nonpolyposis colorectal cancer and cancer caused by mutate in PMS2 is also distribute in the lungs. Gene MAPK11 encodes a protein that involved in the integration of biochemical signals for a wide variety of cellular processes, including cell proliferation, differentiation, transcriptional regulation, and development. Overexpression of gene RHOA is associated with tumor cell proliferation and metastasis. According to the drugs-gens networks, drug Fluorouracil acting on above three drug target genes has shown sensitivity in the clinical responsion of LUSC.

For LUAD patient samples, there are multi-modal PDNBs of sixteen patient samples (70% of early patient sample with multi-modal PDNB) contained drug targets for Lung Adenocarcinoma. The differential genes of multi-modal PDNBs belonging to the early stage of LUAD include 18 drug target genes, as shown in **Table S3**. Drug target genes include but not limited to FGF5, FGF4 and MAX. Gene FGF5 is identified as oncogene. Proteins encoded by FGF5 are involved in biological processes, including cell growth and tumor growth. Drugs Dasatinib and PD-0325901 can show sensitivity in the clinical responsion of LUAD by acting on FGF5. The function of proteins encoded by FGF4 is similar to FGF5. But, gene FGF4 and FGF3, another oncogenic growth factor, are located closely on chromosome 11. Co-amplification of both genes was found in various kinds of human tumors. Drugs Sorafenib and PD-0325901 can show sensitivity in the clinical responsion of LUAD by acting on FGF4. Mutations of gene MAX have been reported to be associated with hereditary pheochromocytoma. Proteins encoded by MAX are able to form homodimers and heterodimers with other family members, which include Mad, Mxi1 and Myc. Myc is an oncoprotein implicated in cell proliferation, differentiation and apoptosis. Drug Gefitinib and Thapsigargin acting on target genes MAX has shown sensitivity in the clinical responsion of LUAD.

**Table S2.** Drug targeted genes and effective drugs obtained by MMPDNB in the early stages of LUSC cancer

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cancer | Cancer Stage | Sample with MMDNB | Differential Genes in MMDNB | Drug Target Gene | Drug |
| LUSC | Stage ia | 1 | UPF2\SMG1 | SMG1 | PIK-93 |
| 12 | MMP1\CMA1 | CMA1 | Fluorouracil |
| 21 | GRHL3\CUL5 | CUL5 | Fluorouracil |
| Stage ib | 19 | RGS6\GTPBP4\NF2\STMN2 | NF2 | Afatinib |
| 22 | PMS2\MLH3 | PMS2 | Fluorouracil |
| 40 | SIRT1\MEF2D | SIRT1 | CX-5461 |
| 41 | CDC42\IQGAP3 | IQGAP3 | Dasatinib |
| 47 | CTPS2\COPS2\COPS3\WNK4\SLC12A3\GMPS\- | WNK4 | Fluorouracil |
| Stage iia | 6 | MAPK11\DUSP6 | MAPK11\DUSP6 | Fluorouracil |
| 29 | ABCG2\HIF1A\NOD1\NOD2 | ABCG2 | Fluorouracil |
| 33 | CHN1\RHOA | RHOA | Fluorouracil |
| 37 | HLA-DOB\HLA-DQA2\PLXNB3\MICAL1\- | PLXNB3\MICAL1 | Dasatinib、 Crizotinib |

**Table S3.** Drug targets and effective drugs obtained by MMPDNB in the early stages of LUAD cancer

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cancer | Cancer Stage | Sample with MMDNB | Differential Genes in MMDNB | Drug Target Gene | Drug |
| LUAD | Stage ia | 50 | PRPS1L1\PRPSAP2 | PRPS1L1 | Tipifarnib、Dasatinib、BLEOMYCIN |
| 68 | FGF18\FGF5 | FGF5 | Mitomycin C、Dasatinib、  PD-0325901 |
| 76 | CYP4A11\SIRPG\PPARA\CD47 | CYP4A11 | Bortezomib |
| 77 | NEDD4L\SCN5A | NEDD4L | GSK-650394 |
| 88 | NUDT21\CAPRIN1 | CAPRIN1 | Bicalutamide |
| 99 | PLRG1\CDC40\HSPA6\HSPA4 | HSPA6 | Bortezomib |
| 101 | TH\PSMA8 | PSMA8 | Bortezomib |
| 105 | SOX2\FGF4 | FGF4 | Sorafenib、  PD-0325901 |
| Stage ib | 55 | P4HB\MTTP | P4HB | Thapsigargin、Dacomitinib |
| 56 | ARNTL2\SIM1 | ARNTL2 | Dasatinib |
| 66 | F2RL1\ST14 | F2RL1 | Bortezomib |
| 80 | SNAI1\MAX | SNAI1\MAX | Gefitinib、  Thapsigargin |
| 90 | PRPH\LMNB1\RPL22\RPL11 | RPL11 | Neratinib |
| 91 | RPL4\RPL37 | RPL37 | Thapsigargin |
| 97 | VPS13B\EXOC2 | VPS13B | SB-216763 |
| 100 | KCNN4\KCNN2\GRWD1\RFWD2 | KCNN4\GRWD1 | CX-5461\Thapsigargin |

**Part** **C: Prognostic prediction of drug targeted gene of three type cancer data.**

For LUSC patients, none of drug targeted genes in multi-modal PDNBs cannot serve as survival risk markers. However, we found that there are two drug targeted genes in 18 drug targeted genes of LUAD which can actually divide all patients into discriminative high-risk and low-risk groups (**Fig. S1**).



**Figure S1. The survival analysis by drug targeted genes on TCGA LUAD data.**