Information about genes.

Shiqiang Jin

2019-4-20

0. Heat map

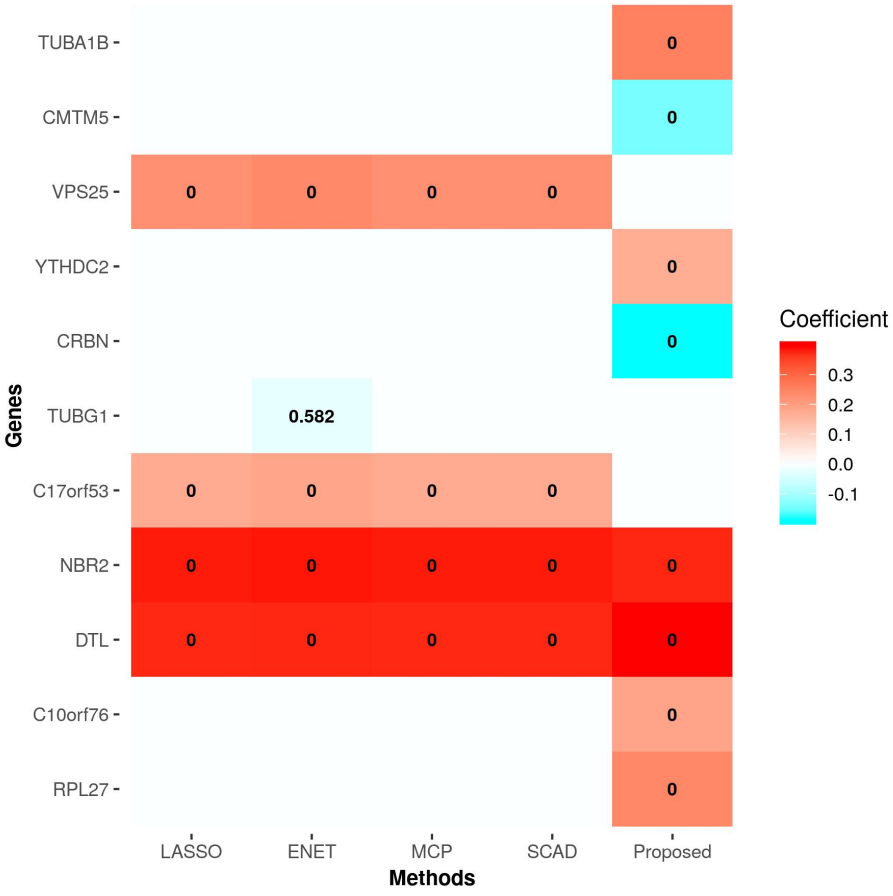

Figure 1: Heapmap

1. Genes selected by proposed method

- 1. RPL27 (Ribosomal Protein L27) in genecards.org or NCBI say:
  - Entrez Gene Summary for RPL27 Gene:

    This gene encodes a member of the L27e family of ribosomal proteins and a component of the 60S subunit. A splice site mutation in this gene has been identified in a **Diamond-Blackfan anemia (DBA)** patient. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome. [provided by RefSeq, Mar 2017]
  - GeneCards Summary for RPL27 Gene:

    RPL27 (Ribosomal Protein L27) is a Protein Coding gene. Diseases associated with RPL27 include Diamond-Blackfan Anemia 16 and Diamond-Blackfan Anemia. Among its related pathways are Viral mRNA Translation and Influenza Viral RNA Transcription and Replication. Gene Ontology (GO) annotations related to this gene include structural constituent of ribosome.
- 2. C10orf76 is descript in wiki as follows:

  C10orf76 or chromosome 10 open reading frame 76, also known as UPF0668, is a protein that in humans is encoded by the c10orf76 gene. **Its function is not currently known**, but experimental evidence has suggested that it may be involved in transcriptional regulation.
- 3. DTL in genecards.org is described as:
  - GeneCards Summary for DTL Gene:

    DTL (Denticleless E3 Ubiquitin Protein Ligase Homolog) is a Protein Coding gene. Diseases associated with DTL include **Myasthenic Syndrome, Congenital, 8**. Among its related pathways are Translesion synthesis by Y family DNA polymerases bypasses lesions on DNA template and DNA Double-Strand Break Repair. Gene Ontology (GO) annotations related to this gene include **ubiquitin-protein transferase activity**.
- 4. NBR2 in genecards.org is described as:
  - Entrez Gene Summary for NBR2 Gene:

    This gene was identified by **its close proximity on chromosome 17 to tumor suppressor gene BRCA1.** Experimental evidence indicates that the two genes share a **bi-directional**

promoter. Transcription for either gene is controlled individually by distinct transcriptional repressor factors. A short (112 amino acid) open reading frame is observed which includes a region derived from a LINE1 element. A strong Kozak signal is not observed for the putative ORF and the stop codon is more than 55 nucleotides upstream of the last splice site for the transcript, suggesting that the transcript is subject to nonsense-mediated decay. Therefore, this gene does not appear to encode a protein. Glucose starvation induces the expression of this gene and the long non-coding RNA transcribed by it functions with AMP-activated protein kinase in mediating the energy stress response. [provided by RefSeq, Aug 2016]

- GeneCards Summary for NBR2 Gene:

  NBR2 (Neighbor Of BRCA1 LncRNA 2) is an RNA Gene, and is affiliated with the non-coding RNA class. Diseases associated with NBR2 include **Adamantinoma Of Long Bones and Pancreatic Cancer 4**.

- 5. CRBN in genecards.org is described as:
  - Entrez Gene Summary for CRBN Gene:

    This gene encodes a protein related to the Lon protease protein family. In rodents and other mammals this gene product is found in the cytoplasm localized with a calcium channel membrane protein, and is thought to play a role in **brain development**. **Mutations** in this gene are associated with **autosomal recessive nonsyndromic cognitive disability**. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Mar 2010]

  - GeneCards Summary for CRBN Gene:

    CRBN (Cereblon) is a Protein Coding gene. Diseases associated with CRBN include **Mental Retardation, Autosomal Recessive 2 and Autosomal Recessive Non-Syndromic Intellectual Disability**. Gene Ontology (GO) annotations related to this gene include **ATP-dependent peptidase activity**.

- 6. YTHDC2 in genecards.org is described as:
  - Entrez Gene Summary for YTHDC2 Gene:

    This gene encodes a member of the DEAH (Asp-Glu-Ala-His) subfamily of proteins, part of the DEAD (Asp-Glu-Ala-Asp) box family of RNA helicases. The encoded protein binds to N6-methyladenosine, a common modified RNA nucleotide that

is enriched in the stop codons and 3' UTRs of eukaryotic messenger RNAs. Binding of proteins to this modified nucleotide may regulate mRNA translation and stability. This gene may be associated with susceptibility to **pancreatic cancer** in human patients, and knockdown of this gene resulted in reduced proliferation in a human **liver cancer** cell line. [provided by RefSeq, Sep 2016]

- GeneCards Summary for YTHDC2 Gene:

  YTHDC2 (YTH Domain Containing 2) is a Protein Coding gene. Gene Ontology (GO) annotations related to this gene include **nucleic acid binding and helicase activity**.

- 7. CMTM5 in genecards.org is described as:
  - Entrez Gene Summary for CMTM5 Gene:

    This gene encodes a member of the chemokine-like factor superfamily. This family of genes encodes multi-pass membrane proteins that are similar to both the chemokine and the transmembrane 4 superfamilies of signaling molecules. The encoded protein may **exhibit tumor suppressor activity**. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jan 2014]

  - GeneCards Summary for CMTM5 Gene:

    CMTM5 (CKLF Like MARVEL Transmembrane Domain Containing 5) is a Protein Coding gene. Gene Ontology (GO) annotations related to this gene include **cytokine activity**.

- 8. TUBA1B in genecards.org is described as:
  - GeneCards Summary for TUBA1B Gene:

    TUBA1B (Tubulin Alpha 1b) is a Protein Coding gene. Diseases associated with TUBA1B include **Kenny-Caffey Syndrome** and **Hypoparathyroidism-Retardation-Dysmorphism Syndrome**. Among its related pathways are Cytoskeleton remodeling Neurofilaments and Signaling events mediated by PRL. Gene Ontology (GO) annotations related to this gene include GTP binding and structural molecule activity. An important paralog of this gene is TUBA1A.

2. Genes selected by LASSO, MCP and SCAD:

- 1. DTL
- 2. NBR2

- 3. C17orf53 in genecards.org is described as:
  - GeneCards Summary for C17orf53 Gene:

    C17orf53 (Chromosome 17 Open Reading Frame 53) is a Protein Coding gene.

  - In wiki:

    The protein product of C17orf53 also displays a decent amount of glycation throughout the protein product. This may indicate that this protein is the start to a pathway that leads to advanced glycation end products, which have been found to be implicated in **many chronic diseases such as cardiovascular problems.** This aligns with previous research indicating that there is high expression of this protein in the heart.
- 4. VPS25 in genecards.org is described as:
  - Entrez Gene Summary for VPS25 Gene:

    This gene encodes a protein that is a subunit of the endosomal sorting complex required for transport II (ESCRT-II). This protein complex functions in sorting of ubiquitinated membrane proteins during endocytosis. A pseudogene of this gene is present on chromosome 1. [provided by RefSeq, Jul 2013]

  - GeneCards Summary for VPS25 Gene:

    VPS25 (Vacuolar Protein Sorting 25 Homolog) is a Protein Coding gene. Among its related pathways are Delta508-CFTR traffic / ER-to-Golgi in CF and Budding and maturation of HIV virion. Gene Ontology (GO) annotations related to this gene include protein homodimerization activity and protein N-terminus binding.

3. Genes selected by ENET:

- 1. DTL
- 2. NBR2
- 3. C17orf53
- 4. TUBG1
  - Entrez Gene Summary for TUBG1 Gene:

    This gene encodes a member of the tubulin superfamily. The encoded protein localizes to the centrosome where it binds to microtubules as part of a complex referred to as the gamma-tubulin ring complex. The protein mediates microtubule nucleation and

is required for microtubule formation and progression of the cell cycle. A pseudogene of this gene is found on chromosome 7. [provided by RefSeq, Jan 2009]

- GeneCards Summary for TUBG1 Gene:

  TUBG1 (Tubulin Gamma 1) is a Protein Coding gene. Diseases associated with TUBG1 include Cortical Dysplasia, Complex, With Other Brain Malformations 4 and Band Heterotopia. Among its related pathways are Regulation of PLK1 Activity at G2/M Transition and Cytoskeleton remodeling Neurofilaments. Gene Ontology (GO) annotations related to this gene include GTP binding and structural constituent of cytoskeleton.

- 5. VPS25

4. Simple comments by Shiqiang:

We can see that the 8 genes selected by our proposed methods related to a specific disease (except C10orf76) according to genecards.org. However, genes selected by LASSO, MCP, SCAD, ENET are not mentioned their mutation or themselves are related to some diseases (except DTL and NBR2 shared with proposed method).