Understanding Genotype-Phenotype relations in Cancer via Network Approaches

Teresa Przytycka
NIH / NLM / NCBI
Genotypes | Phenotypes

Journal “Wisla” (1902) Picture from a local fare in Lublin, Poland
Molecular Phenotypes

Genotypes

EGFR

PTEN
Key challenges in cancer genotype-phenotype analysis

- **Complexity:** Multiple driver mutations are typically required for cancer progression
- **Heterogeneity:** Phenotypically similar cancer cases might be caused by different sets of driver mutations
  - Driver mutations /alterations – mutations contributing to cancer progression
  - Passenger mutations – neutral mutations accumulating during cancer progression
- Some driver mutations are rare
- Epistasis – masking of the effect of one mutation by another mutation
- Cancer evolution
Network/Systems biology view

Motivation:

- Molecules function in the context of interaction networks
- Effects of genetic alteration propagate through the interaction network affecting downstream genes
- Different driver mutations often dys-regulate common pathways
Utilizing Networks for Understanding Genotype-Phenotype effects

1. Dys-regulated Networks
2. Network based signal propagation
3. Patient-similarity Networks
1. Dys-regulated Networks

2. Network based signal propagation

3. Patient-similarity Networks

Journal "Wisła" (1902) Picture from a local fair in Lublin, Poland
Set cover approach as a method to find drivers/markers in heterogeneous data

Goal: Given a set of dysregulated genes and disease cases, find a representative set of dysregulated genes

Kim et al. PolS CB 2011/RECOMB 2010
Module Cover Approach

**Optimization problem:**
Find smallest cost set of modules so that each disease case is covered at least \( k \) times

**Cost is a function of:**
- distance in the network of genes in same module
- A similarity measure (application dependent)
- number of modules (parameterized penalty)

Kim et al. PSB 2013
Module Cover: Glioblastoma Data

Signature modules from GBM Dataset (REMBRANDT)

Kim et al. PSB 2013
Different patients groups have different signature modules
Finding networks dysregulated in multiple cancer types a Pan-Cancer analysis

- Genetic, epigenetic, expression, mRNA level,…
- ~3 thousands of cancer patients
- 12 cancer types

Questions: Are there dysregulated pathways common to many cancer types
Adding mutual exclusivity information

Thomas et al 2007

Possible explanations for ME

- any of the two drivers alone gives sufficient growth advantage
- negative genetic interactions between drivers
In tissue specific cancers mutually exclusive pairs often act in the same pathway

Example from Vandin et al. (lung adenocarcinoma data)

Thomas et al 2007
Ciriello, et al., 2012;
Vandin, et al., 2012;
Leiserson, et al., 2013
Mutual Exclusivity and PanCancer TCGA

Can Mutual Exclusivity principle help identifying common pathways dysregulated across cancer types?
Mutual Exclusivity and PanCancer TCGA

Can Mutual Exclusivity principle help identifying common pathways dysregulated across cancer types?

• Between tissues exclusivity

BETWEEN_ME
Can Mutual Exclusivity principle help identifying common pathways dysregulated across cancer types?

- Between tissues exclusivity BETWEEN_ME
- Across tissues exclusivity ACROSS_ME
Adding ME to the Module Cover

**Approach**

**Optimization problem:**
Find **smallest cost** set of modules so that each disease case is covered at least $k$ times

**Cost** is a function of:
- distance in the network of genes in same module
- Functional relation (application dependent) ; here ME
- number of modules (parameterized penalty)

Kim et al. PSB 2013
PI3K signaling (37.0%)

EGFR pathway (27.9%)

DNA damage response (11.1%)

Calcium channel (22.0%)
Calcium channel (22.0%)

Notch pathway (16.2%)

Splicing (8.1%)

MAPK signaling (11.3%)
Mutual Exclusivity Hubs

The graph illustrates the relationship between the degree and the number of nodes. The x-axis represents the degree, while the y-axis represents the number of nodes. The data points are scattered across the graph, indicating the distribution of nodes at different degrees. The labels TP53, KRAS, PIK3CA, and CTNNB1 are indicated, suggesting specific nodes of interest.
• Module cover is a universal technique to identify dysregulated pathways in a setting of a heterogeneous and complex diseases

• Can be used with various definitions of covering relation and gene connectivity relation
Genotypes

1. Dys-regulated Networks

Phenotypes

2. Network based signal propagation

3. Patient-similarity Networks

Journal "Wisla" (1902) Picture from a local fare in Lublin, Poland
1. Dys-regulated Networks

2. Network-based signal propagation

EGFR  PTEN

3. Patient-similarity Networks

Journal "Wisła" (1902) Picture from a local fair in Lublin, Poland
Information flow from genotypic changes to expression changes

Copy number aberrations or/and mutations

Gene expression

Kim et al. PolS CB 2011/RECOMB 2010
Explaining expression changes in the signature genes

1 2 .....N

Cancer Cases
CNV data

? 

1 2 .....N

Cancer Cases
Gene expression data
Selecting “signature” genes

Find smallest set of genes so that each case is “covered” (=over/under expressed)” at least specified number of times

Kim et al. PolS CB 2011/RECOMB 2010
eQTL analysis links expression variability to genotypic variability

Tu et al. Bioinformatics 2006
Suthram et al. MSB 2008
Kim et al. PoI S CB 2011/RECOMB 2010
Uncovering pathways of information flow between CNV and target gene

Tu et al. Bioinformatics 2006
Suthram et al. MSB 2008
Kim et al. PolS CB 2011/RECOMB 2010
Adding resistances differentiate likelihoods of the edges.

**Resistance** - set to favor most likely path based on gene expression values (reversely proportional to the average correlation of the expression of the adjacent genes with expression of the target gene).
Finding subnetworks with significant current flow

Resistance - set to favor most likely path - based on gene expression values (reversely proportional to the average correlation of the expression of the adjacent genes with expression of the target gene)
Are there common functional pathways?

Cancer Cases CNV data

Cancer Cases Gene expression data

target gene/module

Kim et al. PolS CB 2011/RECOMB 2010
Gene Hubs

| MYC(110)  | E2F1(88) | E2F4(43) | CREBBP(34) | GRB2(27) | SP3(26) | ESR1(25) |
| TFAP2A(25) | NFKB1(23) | MYB(22) | JUN(22) | E2F2(22) | RELA(21) | AR(21) |
| SP1(20) | RPS27A(20) | MAPK3(19) | POU5F1(17) | HIF1A(16) | PPARA(15) | CDC42(15) |
| UBA52(13) | CDK7(13) | YBX1(13) | YWHAZ(12) | CEBPB(12) | POU2F1(12) | UBE2(11) |

Pathway Hubs

Driving Copy number aberrations

GO biological process

- cell cycle arrest
- epidermal growth factor receptor signaling pathway
- negative regulation of cell growth
- Ras protein signal transduction
- regulation of sequestering of triglyceride
- cell proliferation
- nuclear mRNA splicing, via spliceosome
- regulation of cholesterol storage
- nucleotide-excision repair
- RNA elongation from RNA polymerase II promoter
- insulin receptor signaling pathway
- transcription initiation from RNA polymerase II promoter
- N-terminal peptidyl-lysine acetylation
- phosphoinositide-mediated signaling
- positive regulation of lipid storage
- positive regulation of specific transcription from RNA polymerase II promoter
- positive regulation of epithelial cell proliferation
- base-excision repair
- negative regulation of hydrolase activity
- gland development
- positive regulation of MAP kinase activity
- regulation of nitric-oxide synthase activity
- estrogen receptor signaling pathway
- regulation of receptor biosynthetic process
- response to organic substance
- JAK-STAT cascade
- regulation of transforming growth factor-beta2 production
- G1/S transition of mitotic cell cycle
- SMAD protein nuclear translocation
1. Dys-regulated Networks
2. Network based signal propagation
3. Patient-similarity Networks

- Current flow identifies pathways that provide putative explanation of how the information flows from genetic changes to disease markers
- Other methods – shortest path
Phenotypes

1. Dys-regulated Networks

2. Network based signal propagation

3. Patient-Patient similarity Networks
Phenotype similarity network
Document similarity network

(Document are sets of words)

### Phenotypic versus explanatory features

<table>
<thead>
<tr>
<th>Phenotypic features (looks)</th>
<th>Explanatory features (words)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival time</td>
<td>– mutations, CNV, micro RNA level;</td>
</tr>
<tr>
<td>Response to drugs,……</td>
<td>– Epigenetic factors,</td>
</tr>
<tr>
<td>Gene expression profile</td>
<td>– Sex, age, environment …..</td>
</tr>
</tbody>
</table>

### Key idea

Neighbors in patient network should be described similar mixtures of topics features
Case study of GBM (Glioblastoma Multiforme)

Varhaak et al. Classification

- **Mesenchymal**
- **Classical**
- **Proneural**
- **Neural**

patient network for GMB
Based on patient’s features represent each patient as mixture of the subtypes

Subtype I
- EGFR_A 0.45
- NF1_M 0.37
- PTEN_A 0.21

Subtype II
- PDGFA_A 0.51
- IDH1_M 0.29
- M53_M 0.17

Subtype III
- mirR218_H 0.38
- ICDK2_D 0.22
- SHC1_M 0.14

Subtype IV
- CDK2B_D 0.37
- EGFR_A 0.25

Features:
- EGFR_A
- NF1_M
- CDKN2B_D

Cho et al. NAR 2013/RECOMB 2012
Visualization of subtypes distribution from a sample model
Patient-patient relationship based on 1000 models

Observation: No separate Neural group

Cho et al. NAR 2013/RECOMB 2012
And...... we have putative drivers as part of the model

- **Subtype I**
  - EGFR_A 0.45
  - NF1_M 0.37
  - PTEN_A 0.21
  - ...

- **Subtype II**
  - PDGFA_A 0.51
  - IDH1_M 0.29
  - M53_M 0.17
  - ...

- **Subtype III**

- **Subtype IV**
  - CDK2B_D 0.37
  - EGFR_A 0.25
  - ...

---

Cho et al. NAR 2013/RECOMB 2012
1. Dysregulated Networks

2. Network based signal propagation

3. Patient-patient similarity Networks

Journal "Wisła" (1902) Picture from a local fair in Lublin, Poland

Phenotypes

Genotypes
Acknowledgments

Przytycka’s group

DongYeon Cho
Phuong Dao
Jan Hoinka
YooAh Kim
Yijie Wang
Damian Wojtowicz

Support: Intramural research program NLM / NIH

My great-great-uncle (didn’t contribute to the studies)