Patient-specific pathway analysis using PARADIGM identifies key activities in multiple cancers

Josh Stuart, UC Santa Cruz
TCGA Symposium
National Harbor, Nov 18, 2011
Flood of Data Analysis Challenges

Structural Variation

Expression

Copy Number Alterations

DNA Methylation

Exome Sequences

$2^n$ combos
Flood of Data Analysis Challenges

Multiple, Possibly Conflicting Signals

This is What it Does to You
Analysis of disease samples like automotive repair (or detective work or other sleuthing)

Sleuths use as much knowledge as possible.
Much Cell Machinery Known: Gene circuitry now available.

Curated and/or Collected
Reactome
KEGG
Biocarta
NCI-PID
Pathway Commons
Integration key to correct interpretation of gene function

- Expression not always an indicator of activity
- Downstream effects often provide clues

Expression of 3 transcription factors:

- High expression (but inactive) → Inference: TF is OFF (expression reflects activity)
- Low expression (but active) → Inference: TF is ON (low-expression but active)
- High expression (reflects activity) → Inference: TF is ON (high expression but inactive)
Integration key to correct interpretation of gene function

- Need multiple data modalities to get it right.

BUT, targets are amplified

Expression $\rightarrow$ TF ON

Copy Number $\rightarrow$ TF OFF

Lowers our belief in active TF because explained away by cis evidence.
Probabilistic Graphical Models: A Language for Integrative Genomics

Inferring Cellular Networks Using Probabilistic Graphical Models

- Generalize HMMs, Kalman Filters, Regression, Boolean Nets, etc.
- Language of probability ties together multiple aspects of gene function & regulation
- Enable data-driven discovery of biological mechanisms
- Seminal work: J. Pearl, D. Heckerman, E. Horvitz, G. Cooper, R. Schacter, D. Koller, N. Friedman, M. Jordan, …
- Recent work: E. Segal, E Schadt, A. Hartemink, D. Pe’er, …
Integration Approach: Detailed models of gene expression and interaction
Integration Approach: Detailed models of expression and interaction

Two Parts:

1. Gene Level Model (central dogma)
2. Interaction Model (regulation)
PARDIGM Gene Model to Integrate Data

1. Central Dogma-Like Gene Model of Activity

2. Interactions that connect to specific points in gene regulation map
Integrated Pathway Analysis for Cancer

- Integrated dataset for downstream analysis
- Inferred activities reflect neighborhood of influence around a gene.
- Can boost signal for survival analysis and mutation impact
Ovarian: FOXM1 pathway altered in majority of serous ovarian tumors
FOXM1 central to cross-talk between DNA repair and cell proliferation in Ovarian Cancer

Ovarian: IPLs statify by survival time
MYC is characteristically altered in CRC

- Cohort-wide disruption of C-MYC
- Common downstream consequence of WNT and TGFB pathway alterations.
Pathway Signatures: Differential Subnetworks from a “SuperPathway”
Pathway Signatures: Differential Subnetworks from a “SuperPathway”

Pathway Activities

Pathway Activities
Pathway Signatures: Differential Subnetworks from a “SuperPathway”
Triple Negative Breast Pathway Markers Identified from 50 Cell Lines

980 pathway concepts
1048 interactions

One large highly-connected component (size and connectivity significant according to permutation analysis)

Characterized by several “hubs’

- IL23/JAK2/TYK2
- HIF1A/ARNT tetramer
- Myc/Max
- P53
- ER
- FOXA1

Higher activity in ER-
Lower activity in ER-

Sam Ng, Ted Goldstein
Master regulators predict response to drugs: PLK1 predicted as a target for basal breast

- DNA damage network is upregulated in basal breast cancers
- Basal breast cancers are sensitive to PLK inhibitors

Ng, Goldstein
Heiser et al. 2011 PNAS
• HDAC Network is down-regulated in basal breast cancer cell lines
• Basal/CL breast cancers are resistant to HDAC inhibitors

Heiser et al. 2011 PNAS

Ng, Goldstein
Predicting the Impact of Mutations On Genetic Pathways

Inference using all neighbors

Inference using downstream neighbors

Inference using upstream neighbors

PATHWAY DISCREPANCY

Sam Ng
RB1 Loss-of-Function (GBM)

Discrepancy Score
PARADIGM downstream
PARADIGM upstream
Expression
Mutation

RB1

Sam Ng

The Cancer Genome Atlas
RB1 Network (GBM)
TP53 Network

PARADIGM upstream
Expression
NFE2L2 Mutation

Sam Ng
Pathway discrepancy gives orthogonal view of the importance of mutations

- Enables probing into infrequent events
- Can detect non-coding mutation impact (pseudo FPs)
- Can detect presence of pathway compensation for those seemingly functional mutations (pseudo FPs)
- Extend beyond mutations
- Limited to genes w/ pathway representation
Correlates to mutations?

GC001 FOXA1 transcription factor network
GC002 Validated targets of C-MYC transcriptional repression
GC003 Validated targets of C-MYC transcriptional repression

GC004
GC005
GC006 Chemokine receptors bind chemokines
GC007 HIF-2-alpha transcription factor network
GC008 LKB1 signaling events

GC009
GC010 P2Y receptors
GC011 Olfactory Signaling Pathway
GC012 Ion transport by P-type ATPases

GC013 Circadian Clock
What about when we don’t have pathway information for a gene?

Clinical information on samples

Pathway Inferred Levels
Mutation Association to Pathways

- What pathway activities is a mutation’s presence associated?
- Can we classify mutations based on these associations?
Mutation Association to Pathways

- What pathway activities is a mutation’s presence associated?
- Can we classify mutations based on these associations?

(Note: CRC figure below; soon for BRCA)
Mutation Association to Pathways

- What pathway activities is a mutation’s presence associated?
- Can we classify mutations based on these associations?

PARADIGM Signatures

Mutations

Usage Hints
Click to select node
- use arrow keys to navigate tree

GerChemokine receptors bind chemokines; G alpha
GerHIF-2-alpha transcription factor network; HIF
GerReceptor-ligand binding initiates the second
GerCircadian Clock; Glucuronidation
GerCyclin D associated events in G1
GerValidated targets of C-MYC transcriptional
GerC-MYC transcription factor network; FOXM1 to
GerVARIOUS_PATHWAYS
GerVARIOUS_PATHWAYS
GerVARIOUS_PATHWAYS
GerVARIOUS_PATHWAYS
GerErbB1 downstream signaling; ERK/MAPK targets
GerVARIOUS_PATHWAYS
GerVARIOUS_PATHWAYS
GerVARIOUS_PATHWAYS
GerFOXA1 transcription factor network; Notch-ma
GerVARIOUS_PATHWAYS
GerVARIOUS_PATHWAYS
GerVARIOUS_PATHWAYS
GerRegulation of gene expression in early pancy
GerIL4-mediated signaling events; Direct p53 eff
GerDirect p53 effectors; Validated transcription
GerFOXA2 and FOXA3 transcription factor network
GerTransport of organic anions
GerLRKB1 signaling events; Axonal growth inhibit
Mutation Association to Pathways

- What pathway activities is a mutation’s presence associated?
- Can we classify mutations based on these associations?

(Note: CRC figure below; soon for BRCA)
Mutation Association to Pathways

- What pathway activities is a mutation’s presence associated?
- Can we classify mutations based on these associations?

PARADIGM Signatures

Mutations

PIK3CA, RTK pathway, KRAS

(Note: CRC figure below; soon for BRCA)
Mutation Association to Pathways

• What pathway activities is a mutation’s presence associated?
• Can we classify mutations based on these associations?

(Evidence for AHNAK2 acting PI3KCA-like?)
Pan-Cancer: Pathway signatures will connect molecular subtypes across tissues

- Projection of CRC modulated pathways onto GBM and OVCA
Global Pan-Cancer Map

1382 tumor samples:
- 377 OV
- 69 KIRC
- 251 GBM
- 339 BRCA
- 117 LUSC
- 21 LUAD
- 67 READ
- 141 COAD

unpublished
Is there a basal disease? – BRCA vs OVCA

- TCGA ovarian more like basal than luminal breast
Summary

• Model information flow to accurately model gene activity using multi-modal data.
• Focus first on known biology
  • Now going after novel biology (new genes and interactions)
• Patient stratification into pathway-based subtypes
• Sub-networks are predictive markers and can be used to simulate scenarios (like drug inhibition)
• Even rare mutations can be assessed for biological significance.
• Enables multi- and pan-cancer analyses
Connecting the dots: A drug for “rare toe carcinoma” (RTCA)

- TCGA cataloging many signatures of tumors: mutation spectrum, altered genes, and pathway activities
  - E.g. patient presents w/ RTCA and has HER2 amplification
- Subtypes, and ultimately single samples can be connected by these signatures
  - RTCA signature checks out w/ PAM50
- We should also engage signatures from external datasets to inform TCGA data (e.g. Connectivity Map)
  - Signature matches lapatinib sensitivity signature
- Provide a basis to bootstrap clinical findings
  - Prescribe lapatinib to RTCA patient
Shout out to the Broad Team

- PARADIGM now included in Firehose
  - Public now can access CPU-intensive results

- Special THANKS to Daniel DeCara.
Acknowledgments

UCSC Cancer Genomics
- Kyle Ellrott
- Brian Craft
- Chris Wilks
- Chris Szeto
- Amie Radenbaugh
- Mia Grifford
- Sofie Salama
- Steve Benz
- Tracy Ballinger

UCSC Genome Browser Staff
- Mark Diekins
- Melissa Cline
- Jorge Garcia
- Erich Weiler

Buck Institute for Aging
- Christina Yau
- Sean Mooney
- Janita Thusberg

Collaborators
- Laura Esserman, UCSF
- Joe Gray, LBL
- Laura Heiser, LBL
- Eric Collisson, UCSF

Funding Agencies
- NCI/NIH
- SU2C
- NHGRI
- AACR
- UCSF Comprehensive Cancer Center
- QB3

Broad Institute
- Gaddy Getz
- Mike Noble
- Daniel DeCaro