Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma

Tyler Garvin
Pancreatic ductal adenocarcinoma

- Most common form of pancreatic cancer
- Fourth most common cause of cancer deaths worldwide
- Five year survival rate of only 6%
- Recurrent mutations in coding regions have been well established
Regulatory Genome

- Signaling molecule (kinase/phosphatase)
- Transcription factor (repressor/activator)
- Chromatin remodeler
Gene-centric

• Mutations in a:
  – Signaling molecule (kinase/phosphatase)
  – Transcription factor (repressor/activator)
  – Chromatin remodeler (PRC2)
Non-coding mutations

• Mutations OUTSIDE of exons:
  – Promoter regions
  – Enhancers/Insulators (TFBS, DHS)
  – Introns
Cis-regulatory regions

- ENCODE --- Provides transcription factor binding site (TFBS) peaks for 121 "transcription factors"

- Not all of these proteins are actually transcription factors
- DNA binding proteins
- Subunits of a DNA binding protein complex (SUZ12 ~ PRC2)
Key Terms

1) Transcription factors
regulatory proteins (RPs)

2) Transcription factor binding sites
cis-regulatory regions (CRRs)

3) cis-regulatory class is all CRRs that belong to any given RP.
308 patients with WGS and clinical data
  – Simple somatic mutations (SSMs)
  – Matched tumor-normal pairs

96 patients with expression data
GECCO
Gene
Enrichment
Computational
Clustering
Operation
GECCO

(2)

1) Common variants
2) Overlap CRRs
3) Recurrence
Are there CRRs with recurrent non-coding mutations in PDAC?

<table>
<thead>
<tr>
<th>CRR</th>
<th>Nearest gene</th>
<th>Patients (%)</th>
<th>Gene name/protein function</th>
<th>shRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCF12</td>
<td>LHX8</td>
<td>17 (5.52%)</td>
<td>LIM homeobox 8</td>
<td>Yes</td>
</tr>
<tr>
<td>JUND</td>
<td>LINC01194</td>
<td>16 (5.19%)</td>
<td>long intergenic non-protein coding RNA</td>
<td>NA</td>
</tr>
<tr>
<td>E2F1</td>
<td>BMP7</td>
<td>15 (4.87%)</td>
<td>bone morphogenetic protein 7</td>
<td>No</td>
</tr>
<tr>
<td>SUZ12</td>
<td>LHX8</td>
<td>15 (4.87%)</td>
<td>LIM homeobox 8</td>
<td>No</td>
</tr>
<tr>
<td>WRNIP1</td>
<td>DUSP22</td>
<td>15 (4.87%)</td>
<td>dual specificity phosphatase 22</td>
<td>NA</td>
</tr>
<tr>
<td>EP300</td>
<td>REREP3</td>
<td>14 (4.55%)</td>
<td>arginine-glutamic acid dipeptide (RE) repeats pseudogene 3</td>
<td>No</td>
</tr>
<tr>
<td>SUZ12</td>
<td>LMX1B</td>
<td>14 (4.55%)</td>
<td>LIM homeobox ten factor</td>
<td>Yes</td>
</tr>
<tr>
<td>SUZ12</td>
<td>PAX6</td>
<td>14 (4.55%)</td>
<td>paired box 6, homeodomain</td>
<td>No</td>
</tr>
<tr>
<td>TCF12</td>
<td>ZIC4</td>
<td>14 (4.55%)</td>
<td>zinc-finger family member 4</td>
<td>Yes</td>
</tr>
<tr>
<td>HDAC2</td>
<td>FANK1</td>
<td>14 (4.55%)</td>
<td>fibronectin type 3 and ankyrin repeat domains 1</td>
<td>No</td>
</tr>
<tr>
<td>FOXA1</td>
<td>REREP3</td>
<td>13 (4.22%)</td>
<td>arginine-glutamic acid dipeptide (RE) repeats pseudogene 3</td>
<td>NA</td>
</tr>
<tr>
<td>NFKB1, POU2F2</td>
<td>ST8SIA4</td>
<td>13 (4.22%)</td>
<td>ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4</td>
<td>NA</td>
</tr>
<tr>
<td>SIN3A</td>
<td>MIR21</td>
<td>13 (4.22%)</td>
<td>vacuole membrane protein 1</td>
<td>No</td>
</tr>
<tr>
<td>SIN3A</td>
<td>VMP1</td>
<td>13 (4.22%)</td>
<td>vacuole membrane protein 1</td>
<td>Yes</td>
</tr>
<tr>
<td>SUZ12</td>
<td>DMRTA2</td>
<td>13 (4.22%)</td>
<td>doublesex-and Mab-3-related transcription factor A2</td>
<td>Yes</td>
</tr>
<tr>
<td>SUZ12</td>
<td>VAX2</td>
<td>13 (4.22%)</td>
<td>ventral anterior homeobox 2</td>
<td>Yes</td>
</tr>
<tr>
<td>SUZ12</td>
<td>ZIC4</td>
<td>13 (4.22%)</td>
<td>zinc-finger family member 4</td>
<td>Yes</td>
</tr>
<tr>
<td>BCLAF1</td>
<td>DUSP22</td>
<td>12 (3.90%)</td>
<td>dual specificity phosphatase 22</td>
<td>NA</td>
</tr>
<tr>
<td>BCLAF1</td>
<td>MALAT1</td>
<td>12 (3.90%)</td>
<td>Metastasis Associated Lung Adenocarcinoma Transcript 1 (lncRNA)</td>
<td>No</td>
</tr>
<tr>
<td>BCLAF1</td>
<td>VMP1</td>
<td>12 (3.90%)</td>
<td>vacuole membrane protein 1</td>
<td>No</td>
</tr>
<tr>
<td>CDH2, JUND</td>
<td>ZNF595</td>
<td>12 (3.90%)</td>
<td>zinc-finger b3 factor</td>
<td>No</td>
</tr>
<tr>
<td>CDH2, JUND</td>
<td>ZNF718</td>
<td>12 (3.90%)</td>
<td>zinc-finger b3 factor</td>
<td>No</td>
</tr>
<tr>
<td>FOXA1</td>
<td>CDH115</td>
<td>12 (3.90%)</td>
<td>cadherin 15, type 1, M-cadherin</td>
<td>Yes</td>
</tr>
<tr>
<td>HDAC2</td>
<td>CDH8</td>
<td>12 (3.90%)</td>
<td>cadherin b8, type 2</td>
<td>No</td>
</tr>
<tr>
<td>PDA gene</td>
<td>CRE (# patients)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRAS</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP53</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDKN2A</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMAD4</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARID1A</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLL3</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIK3CA</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP2K4</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAF</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZIM2</td>
<td>JUND (6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEG3</td>
<td>TAF1 (6), FOSL2 (5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEB</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLG</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGFBR2</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMCN1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACVR1B</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XIRP2</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APC</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBXW7</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RB1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP47</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Do recurrent non-coding mutations affect known PDAC pathways?

• Three homeobox genes implicated in PDAC
  – PAX6
  – HOXB2
  – HOXB7
GECCO

Input whole genome sequencing data
1) Matched tumor-normal SNV calls
2) RNA-seq expression calls

FunSeq2
Prioritize non-coding regulatory variants

For each CRR variant
- Associate recurrently mutated CRRs with flanking genes
- Use permutation testing to identify CRRs affecting expression
- Generate false discovery rates

For each CRR class
- Determine mutation rates for each regulatory class
- Normalize mutation rates for GC content, size, and abundance
- Compute expression modulation scores

Pathway analysis
Patient survival analysis

(3) (4) (5)
Are non-coding mutations linked to differential gene expression?

For each CRR variant

- Associate recurrently mutated CRRs with flanking genes
- Use permutation testing to identify CRRs affecting expression
- Generate false discovery rates

(3)
Are non-coding mutations linked to differential gene expression?  Yes
GECCO discovered two genes with previously unidentified clinical relevance in PDA

PTPRN2 EXPRESSION (OS)

$P = 0.0019$

Median Survival 20.9 Vs 15.0 months

n = 265

SLC12A8 EXPRESSION (DFS)

$P = 0.0490$

Median Survival 13.9 Vs 11.2 months

n = 246
Are there certain regulatory elements that are driving disease progression?

(4)

For each CRR class

- Determine mutation rates for each regulatory class
  - Normalize mutation rates for GC content, size, and abundance
  - Compute expression modulation scores
Normalized mutation frequency of CRR classes
Computing “expression modulation” scores

- Some RPs have known expression modulation
  - Strong repressors (SUZ12, CTBP2)
  - Strong activators (BDP1, BRF1)

1) Mean expression of genes flanking a CRR class ($\mu_+$)
2) Mean expression of genes NOT flanking a CRR class ($\mu_-$)
3) Ignore genes with 0 expression in > 90% of patients

$$\text{mean}(\log(\mu_+/\mu_-))$$ across all 96 patients
CRR class mutation rates sorted by activator score

![Graph showing CRR class mutation rates sorted by activator score. The x-axis represents Regulatory Protein, and the y-axis represents Mutations/kb. There are two peaks, one for Most Repressive and the other for Most Activating.]
Surprising relationships

• Mutations in the CRRs of strong *repressors* lie proximal to genes involved in known PDA pathways

• Mutations in the CRRs of strong *activators* lie proximal to genes involved in chromatin regulation.
Conclusion

• First collection of NCMs that correlate with changes of expression in PDA.

• NCMs may serve as a novel mechanism to drive key PDA tumorigenesis pathways.

• Uncover clinical outcome relationships for PTPRN2 and SLC12A8 – never implicated previously in PDA.

• There is an enrichment for NCMs in the CRRs of strong activating/repressing RPs and activator/repressor specific pathway dynamics.
Acknowledgments

Mike Schatz
Han Fang
James Gurtowski
Hayan Lee
Maria Nattestad
Srividya Ramakrishnan
Fritz Sedlazeck

Collaborators
Michael Feigin
Peter Bailey
Nicola Waddel
David Chang
Ekta Khurana
Sean Grimmond
Andrew Biankin

Committee
Dick McCombie
Dave Tuveson
Zach Lippman

John and Amy Phelan Fellowship