Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma

Tyler Garvin

Pancreatic ductal adenocarcinoma

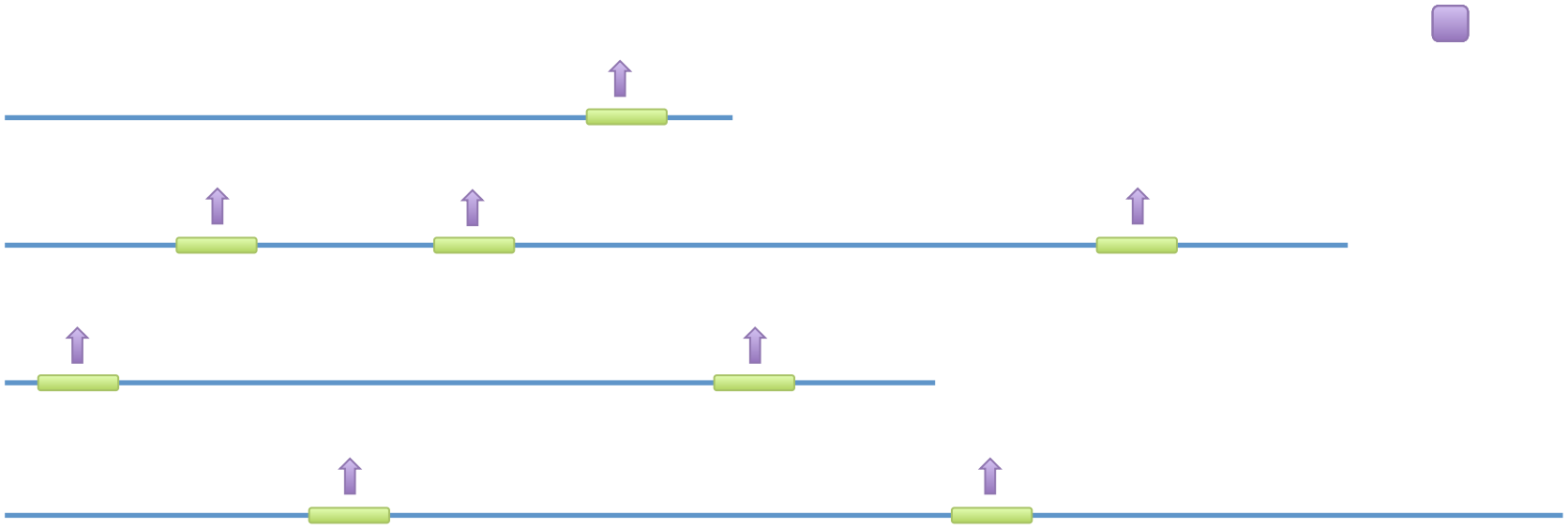
- Most common form of pancreatic cancer
- Fourth most common cause of cancer deaths worldwide
- Five year survival rate of only 6%
- Recurrent mutations in coding regions have been well established

Michael Feigin

David Tuveson

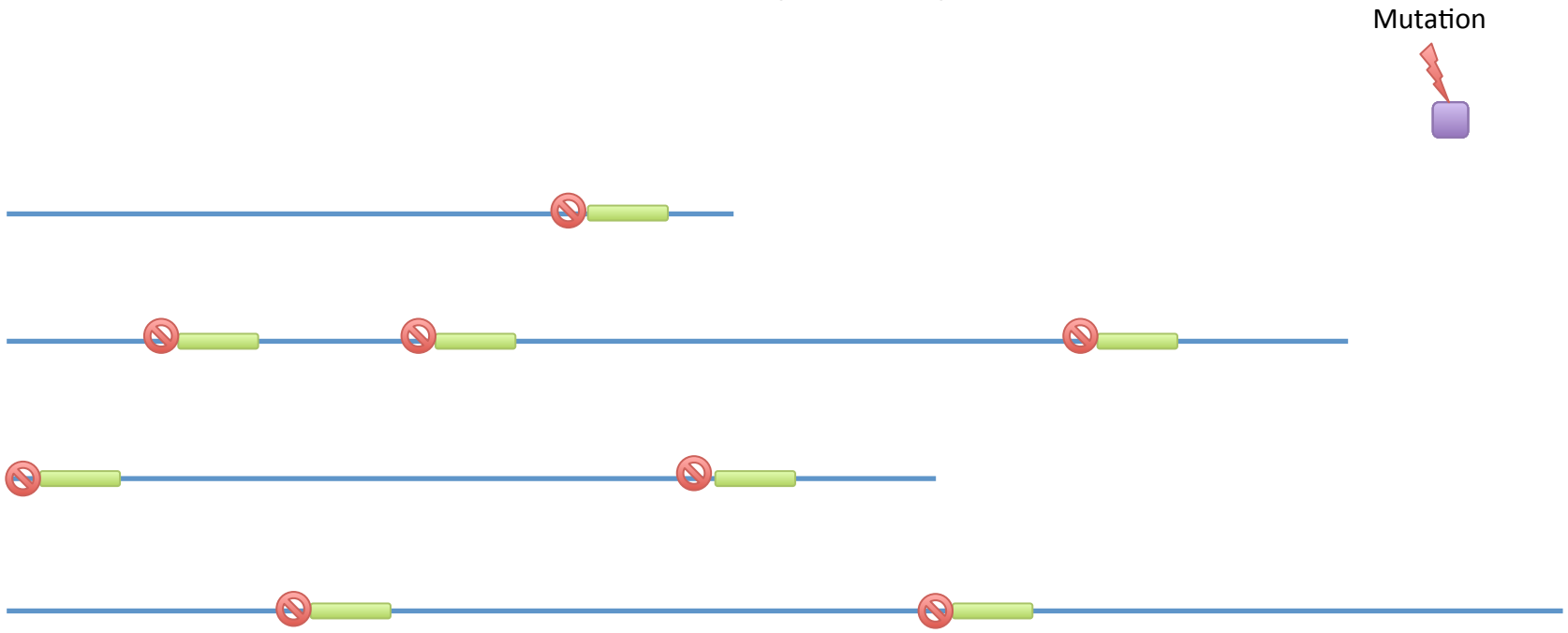
Regulatory Genome

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



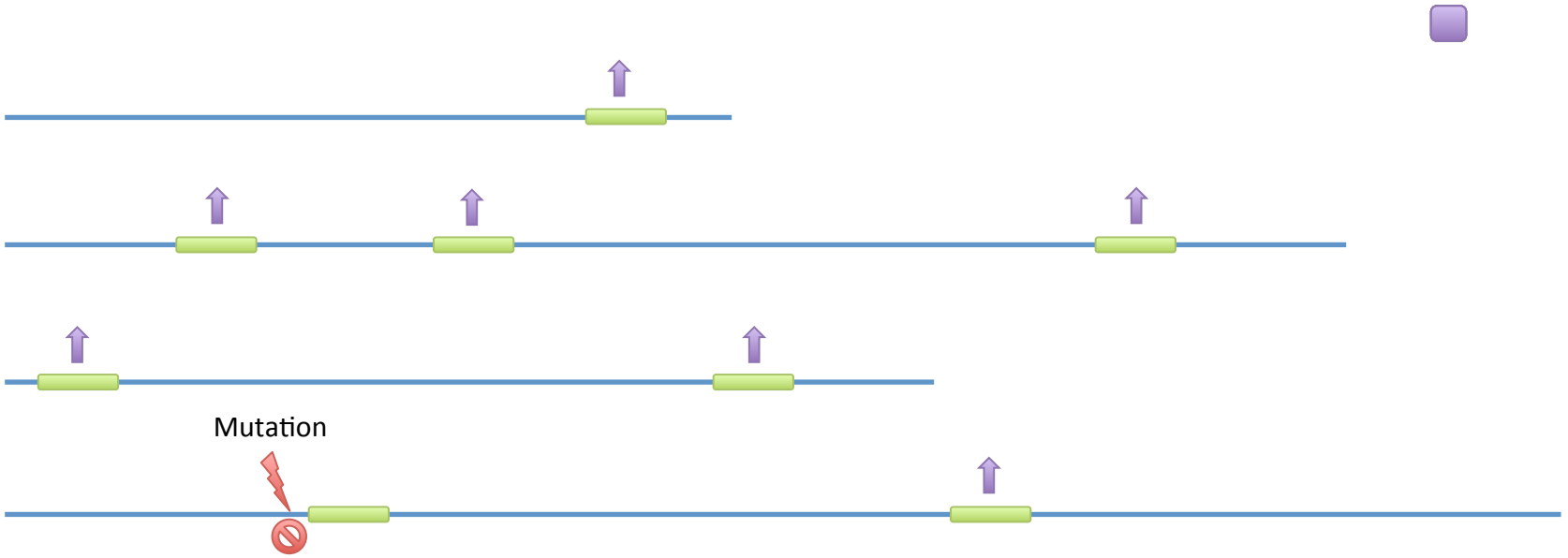
Gene-centric

- 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)

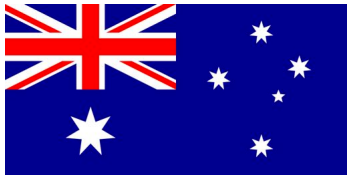
Key Terms

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.

International Cancer Genome Consortium



PACA-AU

PACA-CA

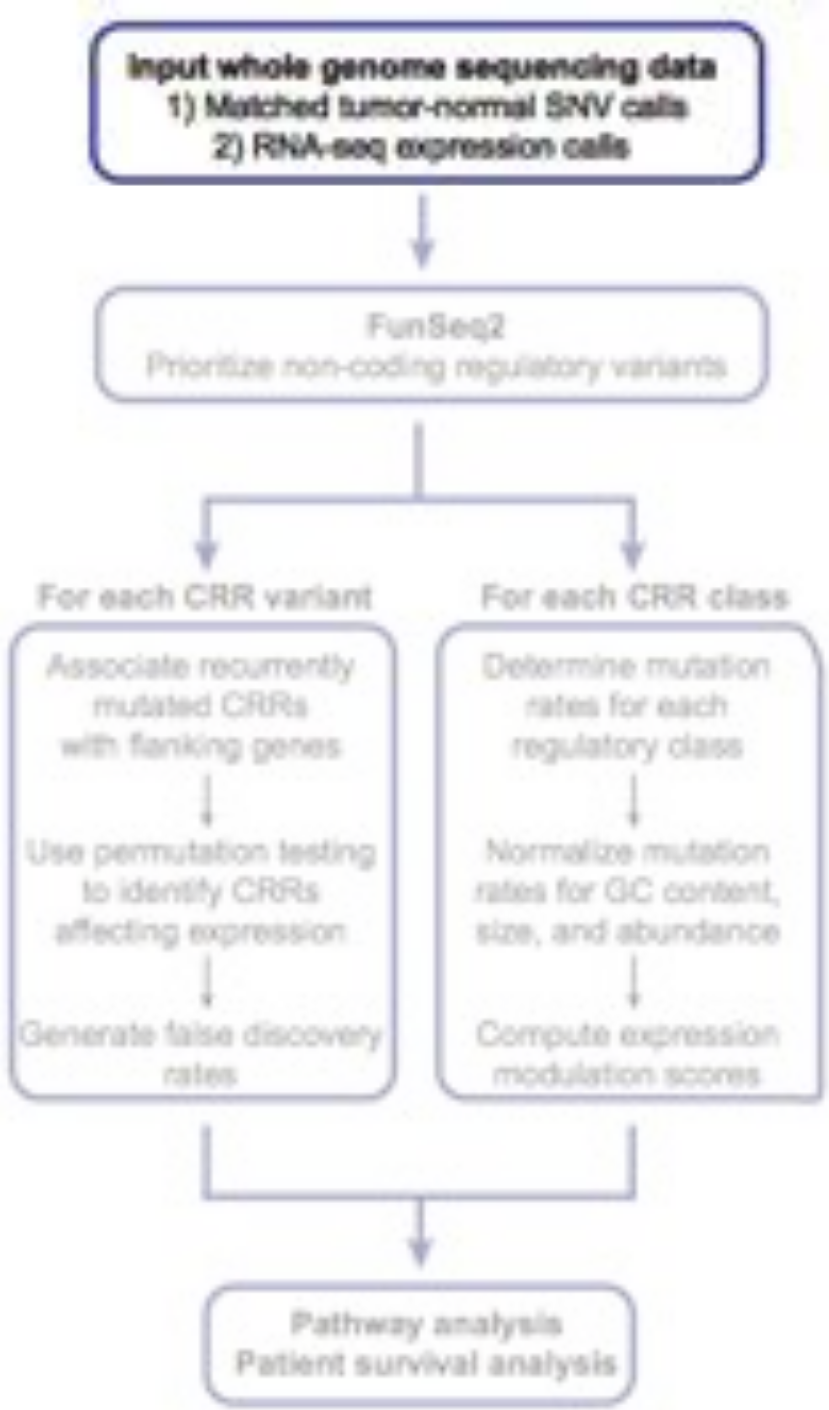
308 patients with WGS and clinical data

- Simple somatic mutations (SSMs)
- Matched tumor-normal pairs

96 patients with expression data

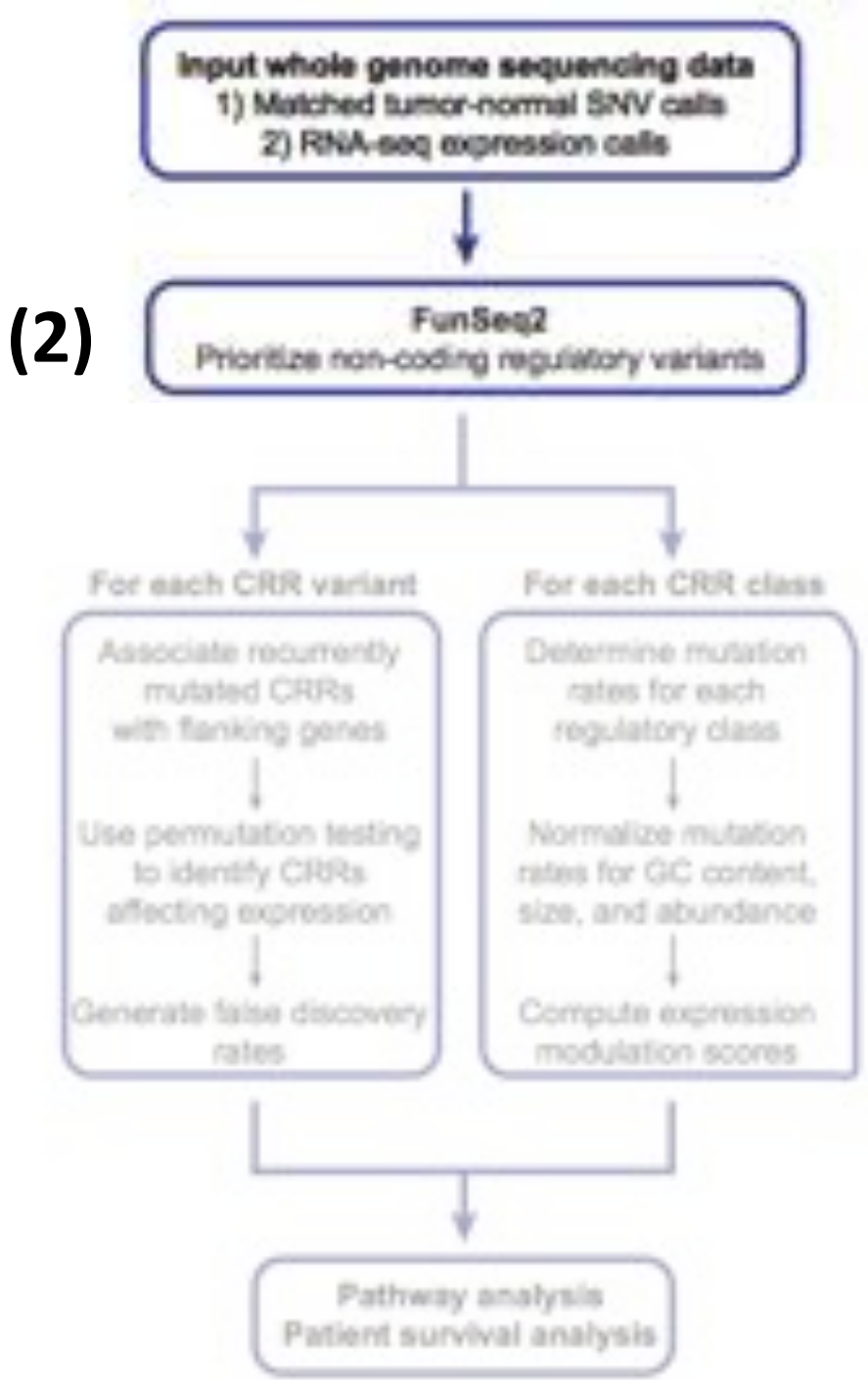
GECCO

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?

Noncoding gene-proximal mutational clusters in PDA

CRR	Nearest gene	Patients (%)	Gene name/protein function	shRNA
TCF12	LHX8	17 (5.52%)	LIM homeobox 8	Yes
JUND	LINC01194	16 (5.19%)	long intergenic non-protein coding RNA	NA
E2F1	BMP7	15 (4.87%)	bone morphogenetic protein 7	No
SUZ12	LHX8	15 (4.87%)	LIM homeobox 8	
WRNIP1	DUSP22	15 (4.87%)	dual specificity phosphatase 22	No
EP300	REREP3	14 (4.55%)	arginine-glutamic acid dipeptide (RE) repeats pseudogene 3	NA
SUZ12	LMX1B	14 (4.55%)	LIM homeobox txn factor	Yes (P)
SUZ12	PAX6	14 (4.55%)	paired box 6, homeodomain	Yes
TCF12	ZIC4	14 (4.55%)	zinc-finger family member 4	No
HDAC2	FANK1	14 (4.55%)	fibronectin type 3 and ankyrin repeat domains 1	No
FOXA1	REREP3	13 (4.22%)	arginine-glutamic acid dipeptide (RE) repeats pseudogene 3	NA
NFKB1, POU2F2	ST8SIA4	13 (4.22%)	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4	No
SIN3A	MIR21	13 (4.22%)	microRNA21	NA
SIN3A	VMP1	13 (4.22%)	vacuole membrane protein 1	No
SUZ12	DMRTA2	13 (4.22%)	doublesex-and Mab-3-related transcription factor A2	Yes
SUZ12	VAX2	13 (4.22%)	ventral anterior homeobox 2	Yes
SUZ12	ZIC4	13 (4.22%)	zinc-finger family member 4	
BCLAF1	DUSP22	12 (3.90%)	dual specificity phosphatase 22	
BCLAF1	MALAT1	12 (3.90%)	Metastasis Associated Lung Adenocarcinoma Transcript 1 (lncRNA)	NA
BCLAF1	VMP1	12 (3.90%)	vacuole membrane protein 1	
CDH2, JUND	ZNF595	12 (3.90%)	zinc-finger txn factor	No
CDH2, JUND	ZNF718	12 (3.90%)	zinc-finger txn factor	No
FOXA1	CDH15	12 (3.90%)	cadherin 15, type 1, M-cadherin	Yes (P)
HDAC2	CDH8	12 (3.90%)	cadherin 8, type 2	No

NCM overlap with known PDA genes

PDA gene	CRE (# patients)
KRAS	-
TP53	-
CDKN2A	-
SMAD4	-
ARID1A	-
MLL3	-
PIK3CA	-
MAP2K4	-
BRAF	-
ZIM2	JUND (6)
PEG3	TAF1 (6), FOSL2 (5)
NEB	-
FLG	-
TGFBR2	-
ATM	-
HMCN1	-
ACVR1B	-
XIRP2	-
APC	-
FBXW7	-
RB1	-
USP47	-

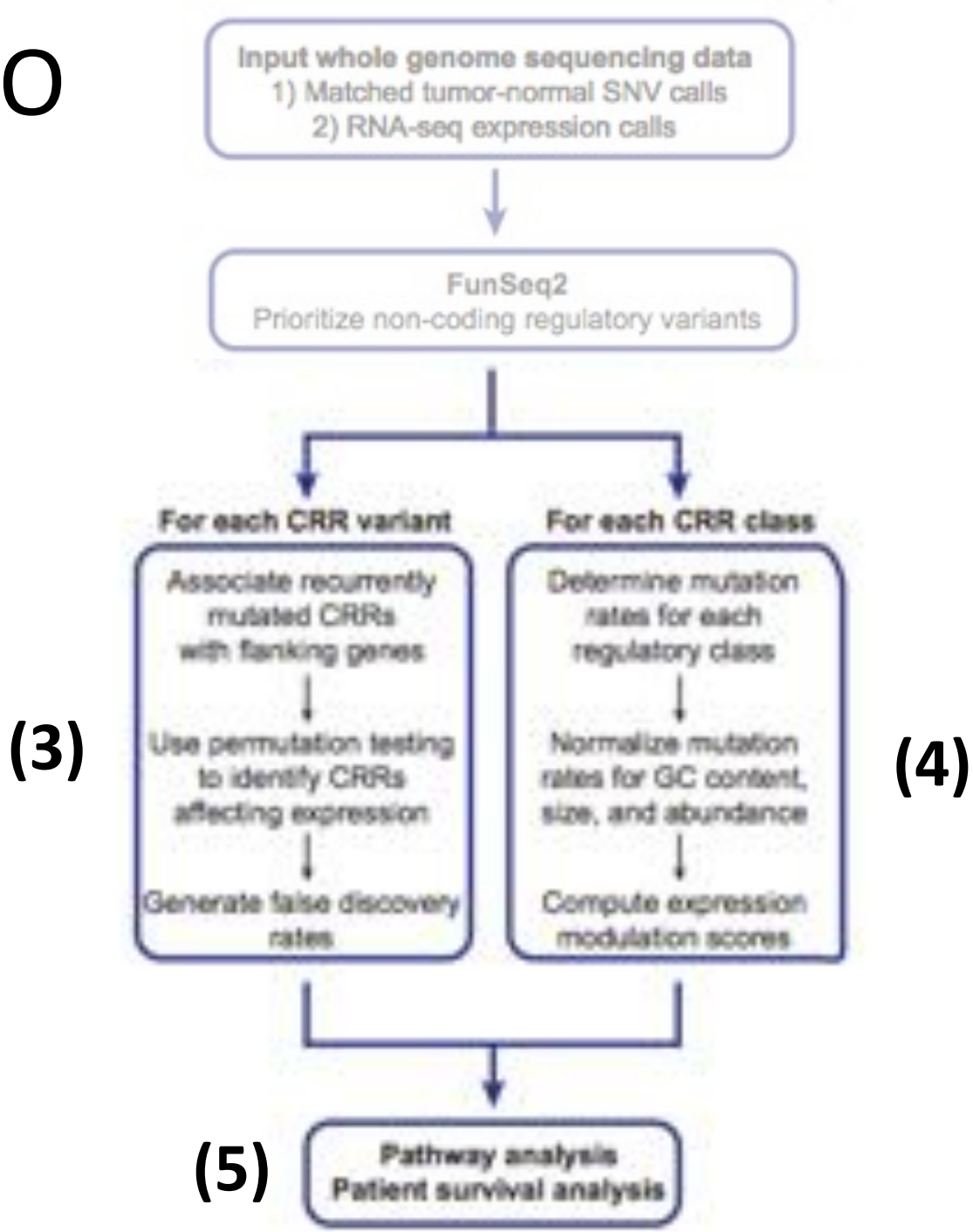
Do recurrent non-coding mutations affect known PDAC pathways?

Pathways regulated by NCMs in pancreatic ductal adenocarcinoma

Regulatory process/gene family	# genes altered	*p*-value	Representative altered genes
Regulation of transcription	135	3.9E-15	ALX4, DMRTA2, T, TWIST1, RUNX3, WWTR1
Homeobox	45	6.2E-26	LHX5, NKX2-8, HOXB4, IRX1, MSX1, VAX2
Neuron differentiation/axon guidance	53	1.1E-19	ROBO1, SLIT2, NRXN1, CTNNA2, NCAM2, BDNF
Cell adhesion	24	2.8E-4	CDH15, CDH8, CADM1, ITGB2, LAMA5, CNTN4
Wnt signaling pathway	18	4.3E-2	FZD10, FBXW11, NKD1, TCF7L1, EN2

- 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

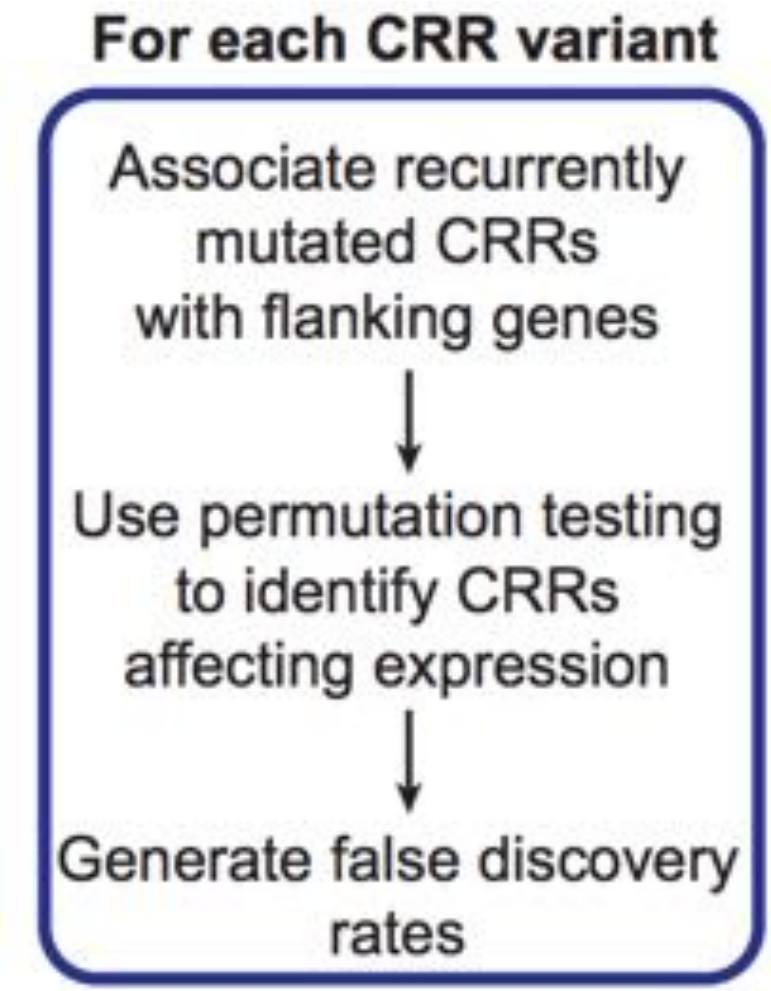
(3)

(4)

(5)

Are non-coding mutations linked to differential gene expression?

(3)

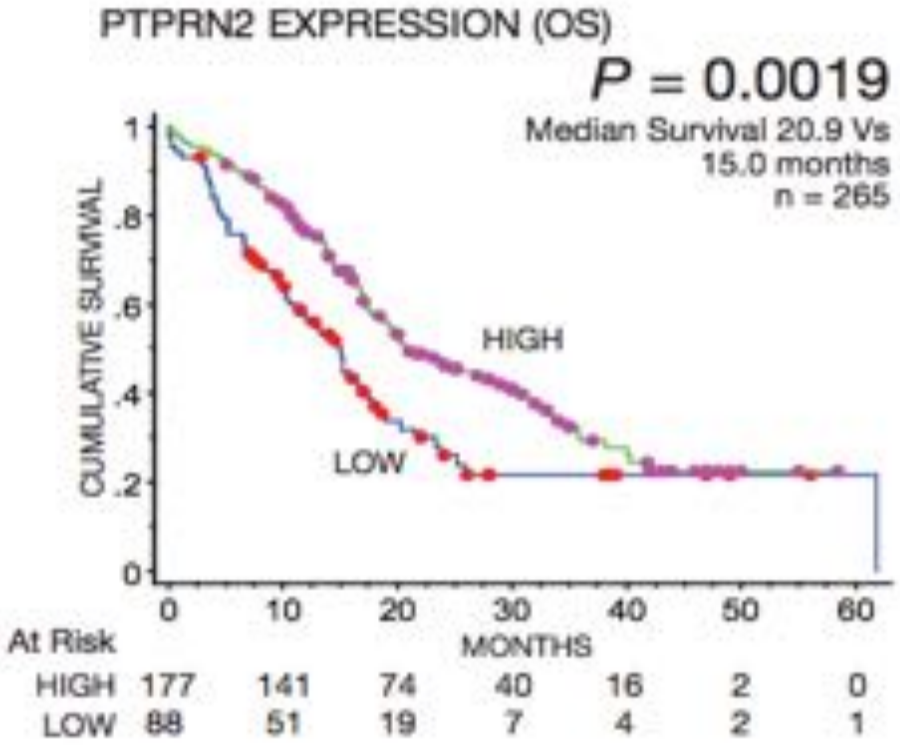


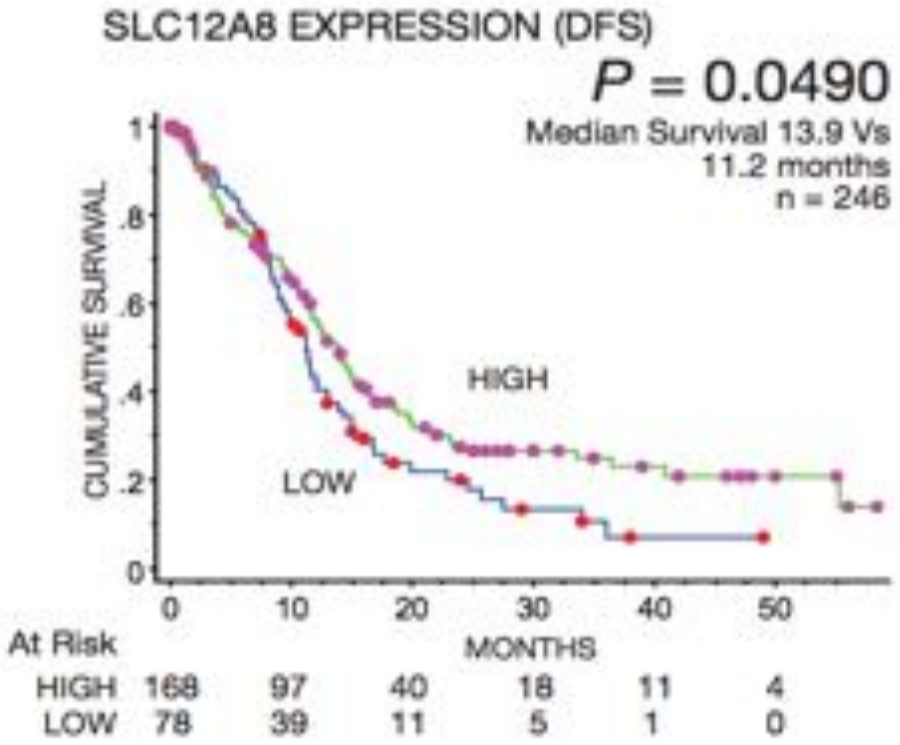
Are non-coding mutations linked to differential gene expression? Yes

Recurrent mutations correlate with gene expression changes

CRE	Gene symbol	MUT exp	WT Exp	Fold change	p-value	q-value
MAX	PTPRN2	0.82	10.92	13.38	0.00593	0.09689
FOSL2	KCNQ1	0.85	6.39	7.52	0.02456	0.18212
TAF7	SNRPN	0.46	3.4	7.39	0.00818	0.11818
NFKB1	GYPC	1.08	7.29	6.75	0.01845	0.15157
TAF1	PDPN	2.09	13.08	6.26	0.03544	0.22016
BCLAF1	PRSS12	1.07	6.46	6.06	0.01107	0.14144
MAFK	SOX5	0.29	1.63	5.60	0.02851	0.20379
POU2F2	MIR4420	8.16	40.24	4.93	0.01773	0.15157
WRNIP1	IKZF1	0.64	3.15	4.89	0.01811	0.15157
GATA3	PCLO	0.35	1.67	4.75	0.01113	0.14144
JUND	TUSC7	0.98	4.53	4.62	0.02909	0.20560
STAT1	PBX1	3.34	14.12	4.22	0.04833	0.26237
REST	MTERF4	1.46	5.78	3.96	0.02209	0.16542
GATA1	FNIP2	7.59	18.32	2.41	0.02588	0.18929
CEBPB	PNPLA8	5.69	13.62	2.39	0.01726	0.15157
EGR1	SLC12A8	4.34	7.99	1.84	0.04185	0.23823
SIN3A	FAM192A	20.31	30.48	1.5	0.01788	0.15157

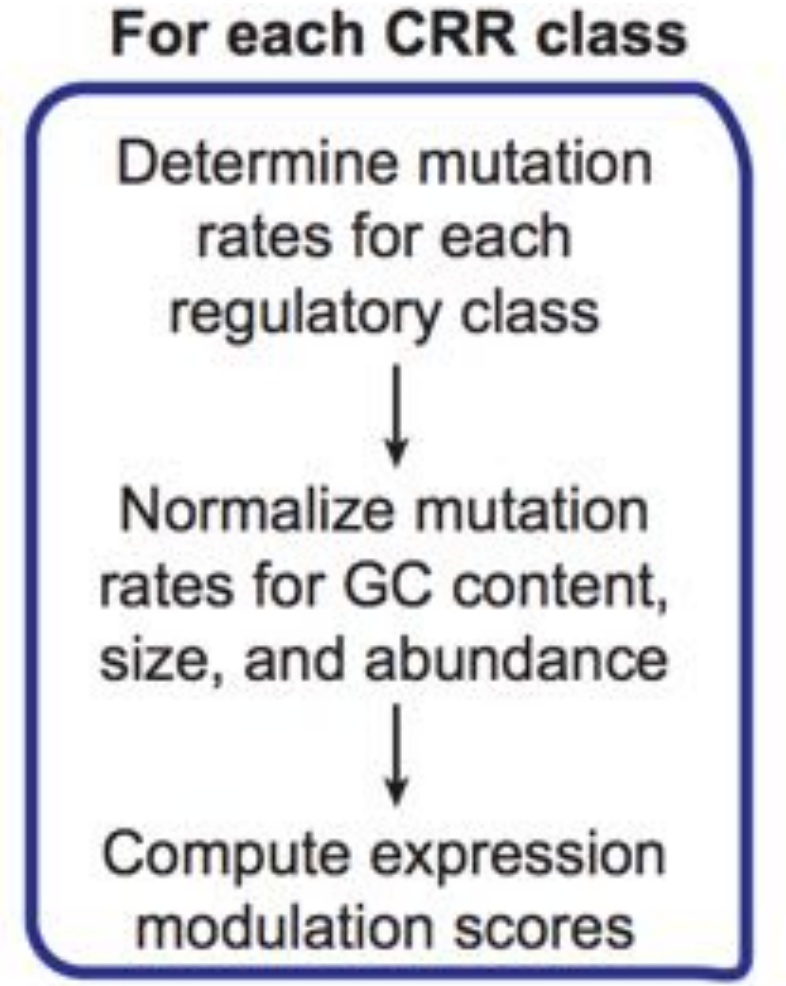
GECCO discovered two genes with previously unidentified clinical relevance in PDA



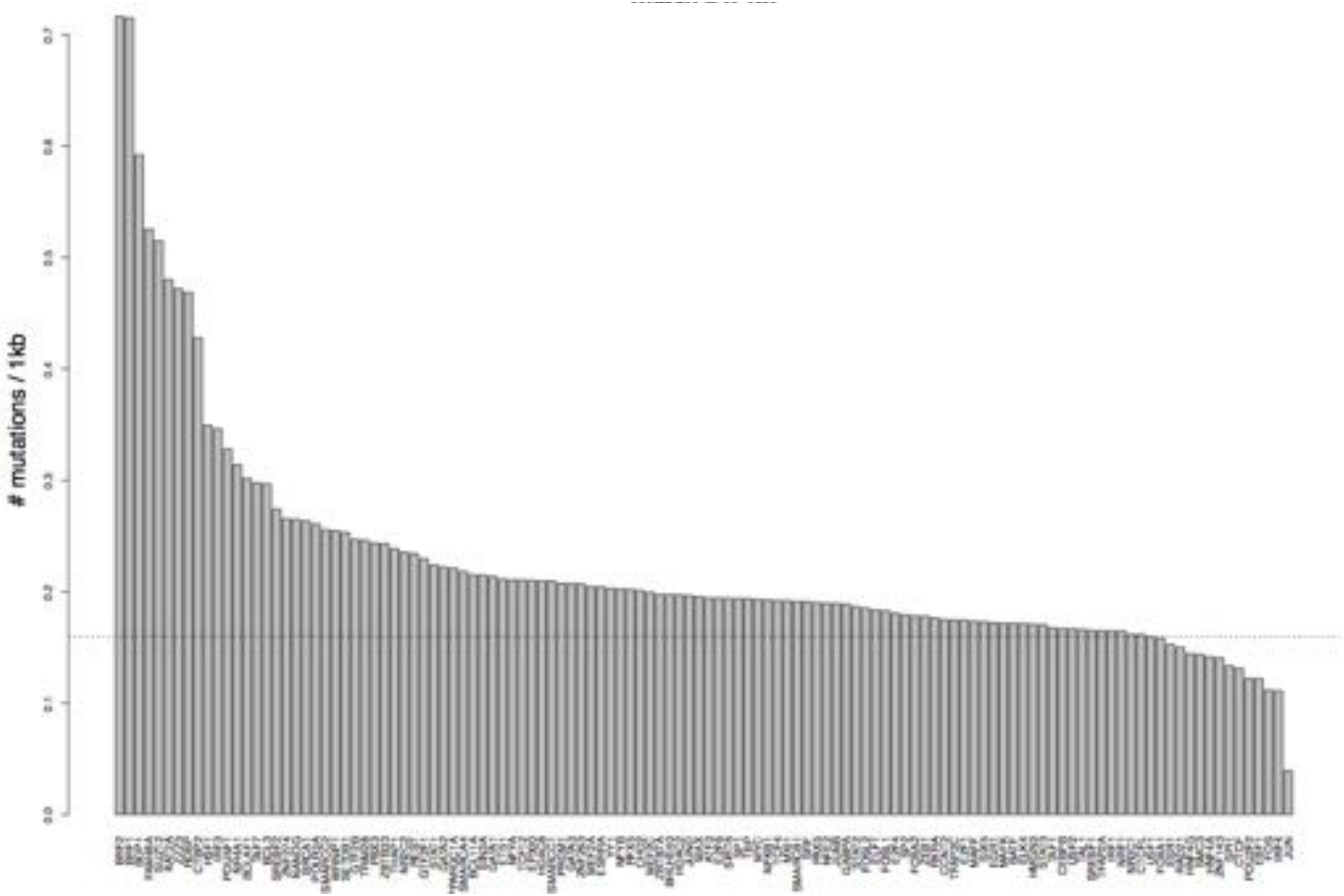


Are there certain regulatory elements that are driving disease progression?

(4)



Normalized mutation frequency of CRR classes

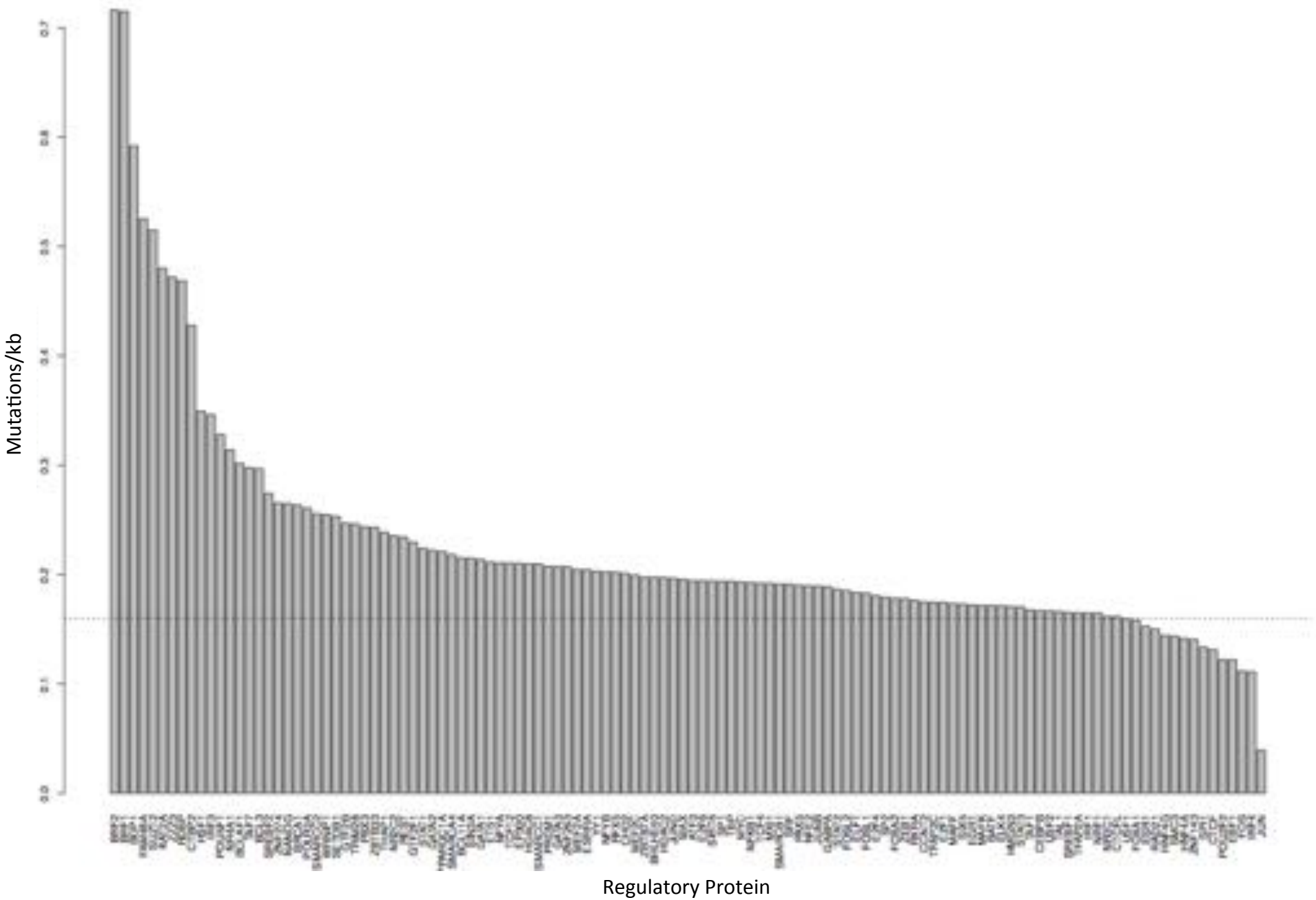


Computing "expression modulation" scores

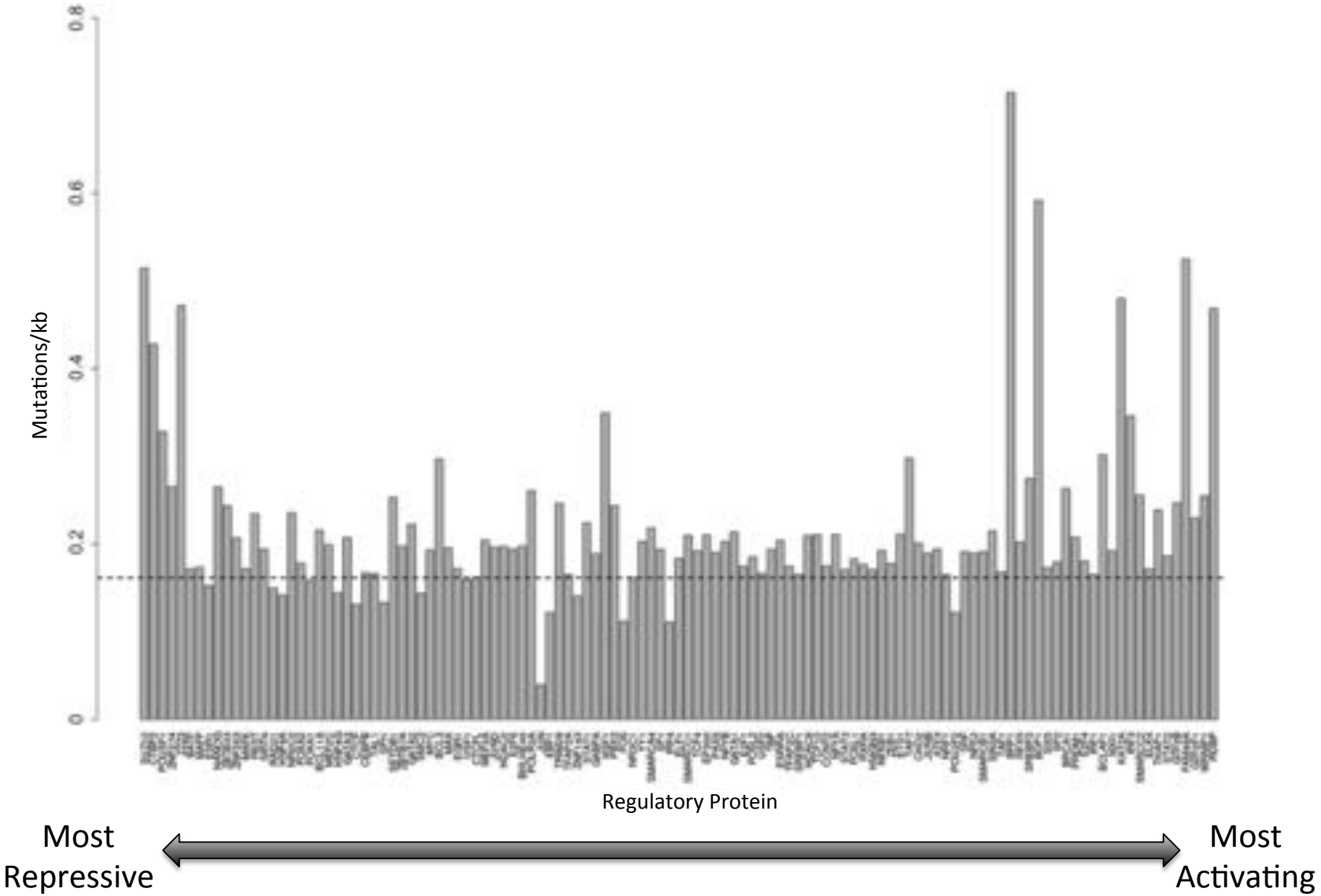
- Some RPs have known expression modulation
  - Strong repressors (SUZ12, CTBP2)
  - Strong activators (BDP1, BRF1)

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

$$\text{mean}(\log(\mu_+/\mu_-)) \text{ across all 96 patients}$$


Mutations/kb
Regulatory Protein

CRR class mutation rates sorted by activator score



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

Watson School of Biological Sciences

CSH Cold Spring Harbor Laboratory