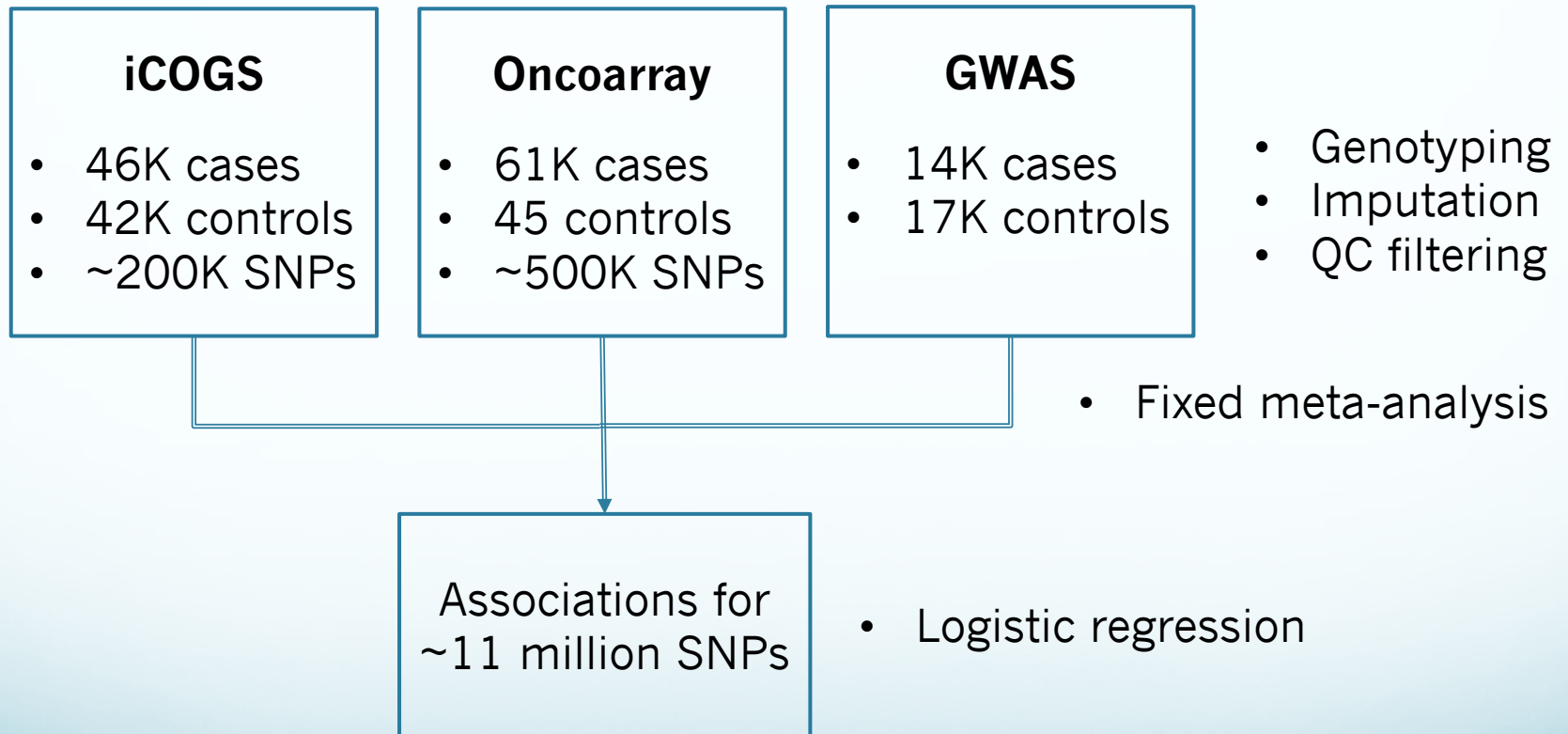


Pathway Analysis of Breast Cancer GWAS Data

By Shirley Hui

August 23, 2017

Breast Cancer GWAS Data



Pathway GSEA

- Individual genes may not reach genome-wide significance but multiple related genes in the same pathway may work together to confer disease susceptibility
- Provides more explanatory power in terms of biological pathways affected
- Complexity reduced when thousands of genes are grouped into a smaller number of pathways
- Example: GenGen by Wang et al.

Pathway GSEA Pipeline (1)

- Reduce number of SNPs
 - ~11M SNPs -> 1.25M SNPs based on association statistic $p < 5e-02$
- Map SNPs to Genes
 - INQUIST: Identify potential functional impact of SNPs on regulatory or coding features using publically available data (J. Beesley et al.)

Pathway GSEA Pipeline (2)

- Pathway Enrichment Score (PES):
 - Determine whether or not a gene set is overrepresented at the top or bottom of a ranked list of genes
- PES Significance:
 - Shuffle case/control status 1000 times
 - Compute chi sq statistic to test allelic association of each SNP to disease status
 - Re-compute scores to build PES null distribution and compute nominal p-values
- Normalized Enrichment Score (NES):
 - Account for differences in gene set size
 - $NES = PES / \text{mean}(\text{all permuted PES})$
- False Discovery Rate (FDR):
 - Account for multiple hypothesis testing
 - $FDR = (\% \text{ of all permuted } NES \geq NES^*) / (\% \text{ of all observed } NES \geq NES^*)$

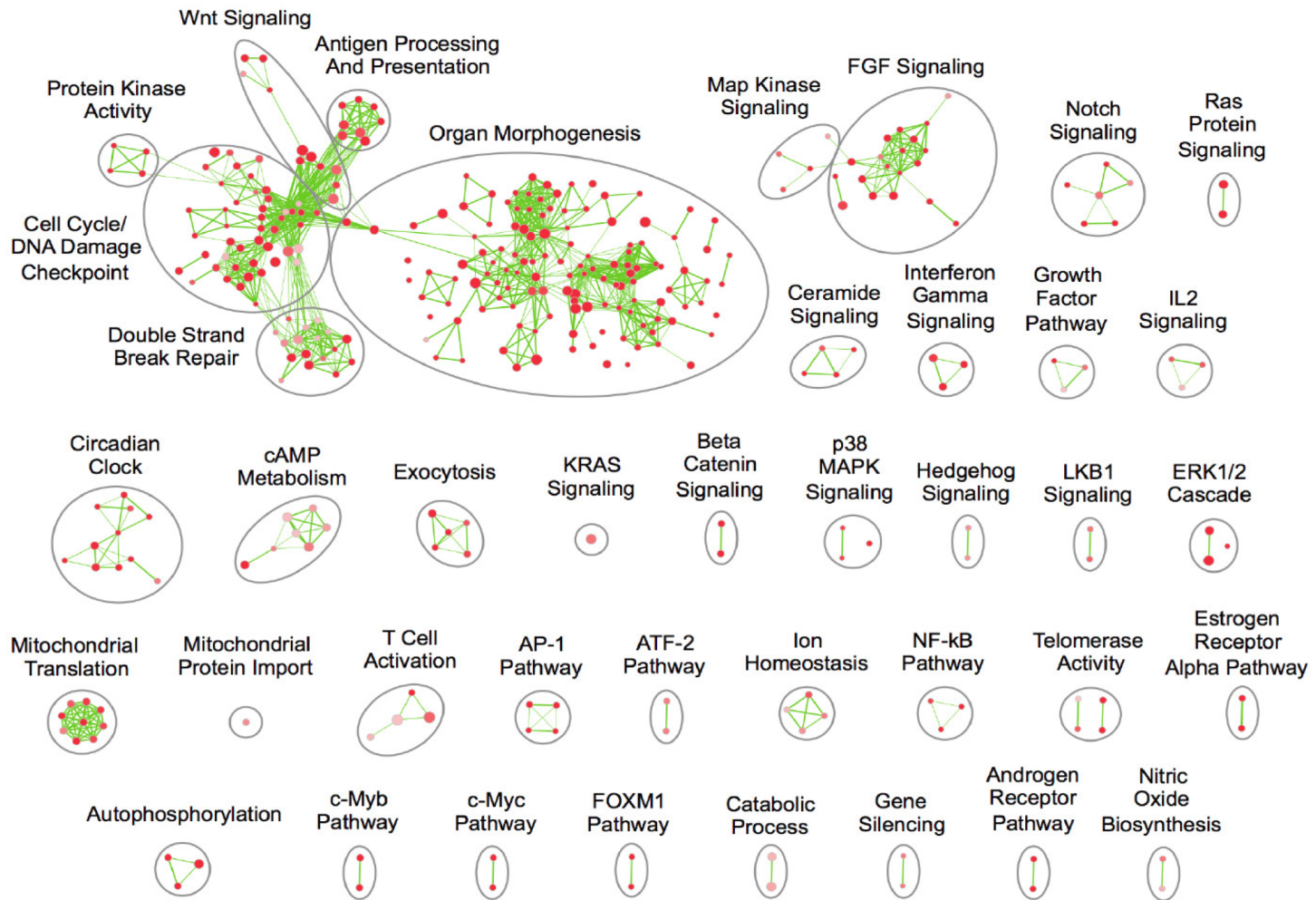
Pathway GSEA Pipeline (3)

- Score and statistics computed for Oncoarray and iCOGS data sets separately
- Pathway kept for further analysis if:
 - FDR < 0.05 in either data sets, AND
 - More than one significant gene in pathway, AND
 - Involved in breast cancer as reported in five published large-scale breast cancer GWAS or literature

OR

- Contained at least 4 novel genes (i.e. genes not found among genes from mapped breast cancer themes)

Enrichment Map



Themes found in other Breast Cancer GWAS

Other BC GWAS Themes	Our Themes
Growth Related	<ul style="list-style-type: none">• Organ Morphogenesis• Growth Factor Pathways• FGF Signaling
Immune System Related	<ul style="list-style-type: none">• IL2 Signaling• T Cell Activation• Antigen Presentation and Processing
DNA Repair	<ul style="list-style-type: none">• Double Strand Break Repair
Kinase Related	<ul style="list-style-type: none">• Map Kinase Signaling• ERK1/2 Cascade
Signal Transduction	<ul style="list-style-type: none">• Notch Signaling
Compound Metabolism	<ul style="list-style-type: none">• cAMP Metabolism
Ion Transport	<ul style="list-style-type: none">• Ion Homeostasis
Circadian Clock	<ul style="list-style-type: none">• Circadian Clock

Themes found in Literature

Theme	Literature
Androgen Receptor Activity	Mishra, AK et al., Indian J Med Res (2012)
AP-1 Pathway	Shen, Q et al., Oncogene (2008)
Ceramide Signaling	Morad, EA & Cabot, MC, Nat Rev Cancer (2013)
Estrogen Receptor Alpha Pathway	Hayashi, SI et al., Endocr Relat Cancer (2003)
FOXO1 Pathway	Saba, R et al., Int J Breast Cancer (2016)
LKB1 Signaling	Rhodes, LV et al.,
Mitochondrial Translation	Sotgia, F et al., Cell Cycle (2012)
Nitric Oxide Biosynthesis	Xu, W et al., Cell Res (2002)
Telomerase Activity	Lu, L et al., Breast Cancer Res (2011)
Wnt Signaling	Howe, LR & Brown, AM, Cancer Biol Ther (2004)

Novel themes

Theme	Theme FDR	Novel Genes
Autophosphorylation	3.10E-05	FES, MAP3K11, CLK2, GRK7
Catabolic Process	0.04126	USP25, DFFA, PKP1, ZKSCAN3
Exocytosis	0.00248	SYT8, RAB3A, TC2N, FES
KRAS Signaling	0.02961	PLEK2, KCNN4, ATG10, CBX8, TSPAN1
Mitochondrial Protein Import	0.03529	HSCB, SLC25A13, SLC25A12, MTX1, TOMM70A, COQ2

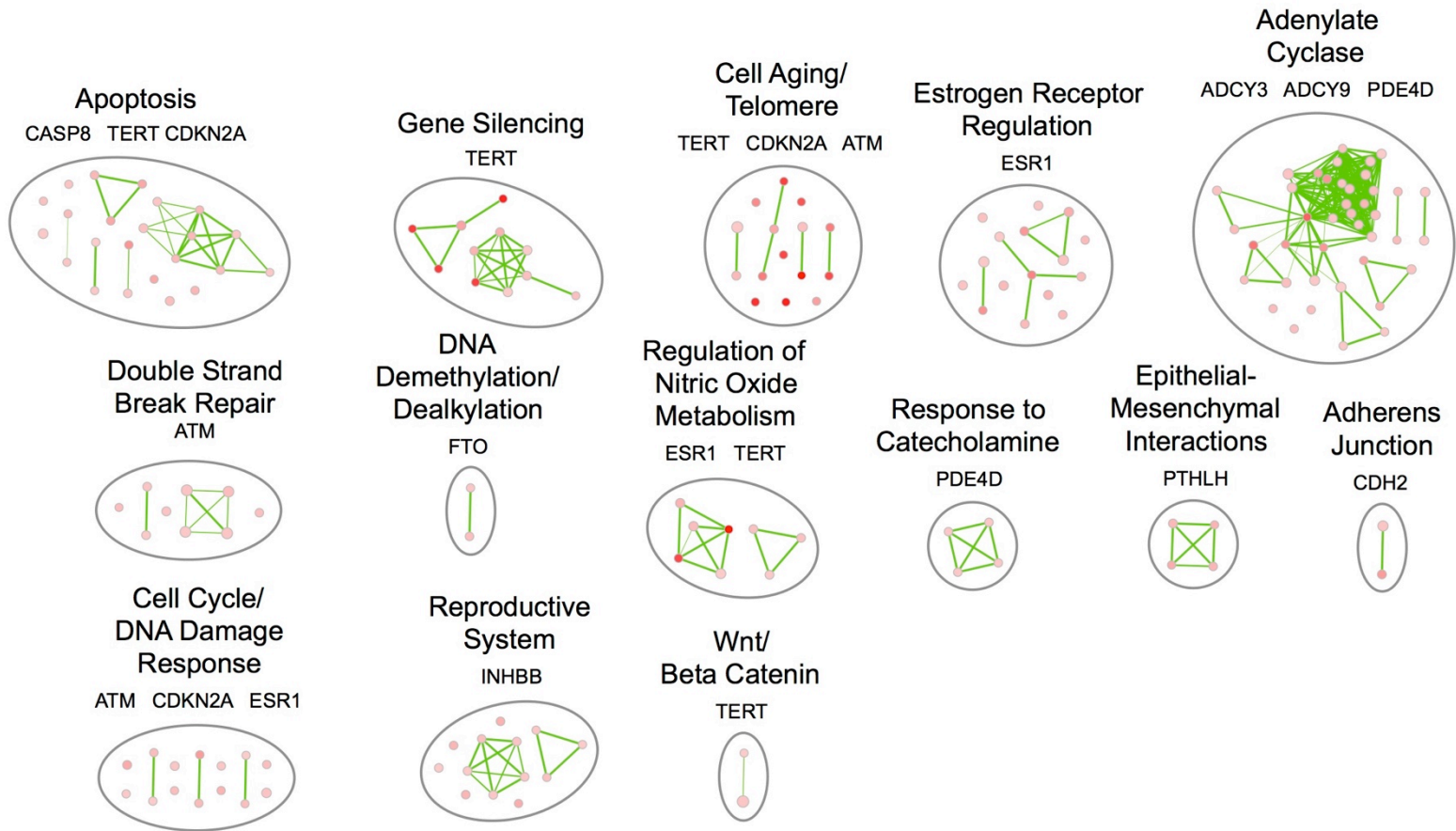
ER- BRCA1 GWAS

- Oncoarray
 - 15,566 BRCA1 mutation carriers, 7784 with breast cancer, European
- iCOGS
 - 3342 BRCA1 mutation carriers, 1630 with breast cancer, European
- ~470K SNPs genotyped
- Total ~17 million SNP associations

Pathway GSEA Pipeline

- SNPs to Genes
 - SNPs within 500kb away from a gene were mapped to the closest gene
 - Gene significance assigned as most significant SNP among all SNPs mapped
- Pathway GSEA
 - Compute PES using all associations
 - Define PES threshold $PES \geq 0.41$ (TPR=0.2, FDR=0.14)
- Pathways confirmed to be relevant to breast cancer based on literature search and other GWAS results

ER- BRCA1 Enrichment Map



Acknowledgements

Asha Rostamianfar

Gary Bader

Authors: Jonathan Beesley, Joe Dennis, Doug Easton,
Lesley McGruffog, Kyriaki Michailidou, Roger Milne

Genome Canada/CIHR Genomics and Personalized
Health