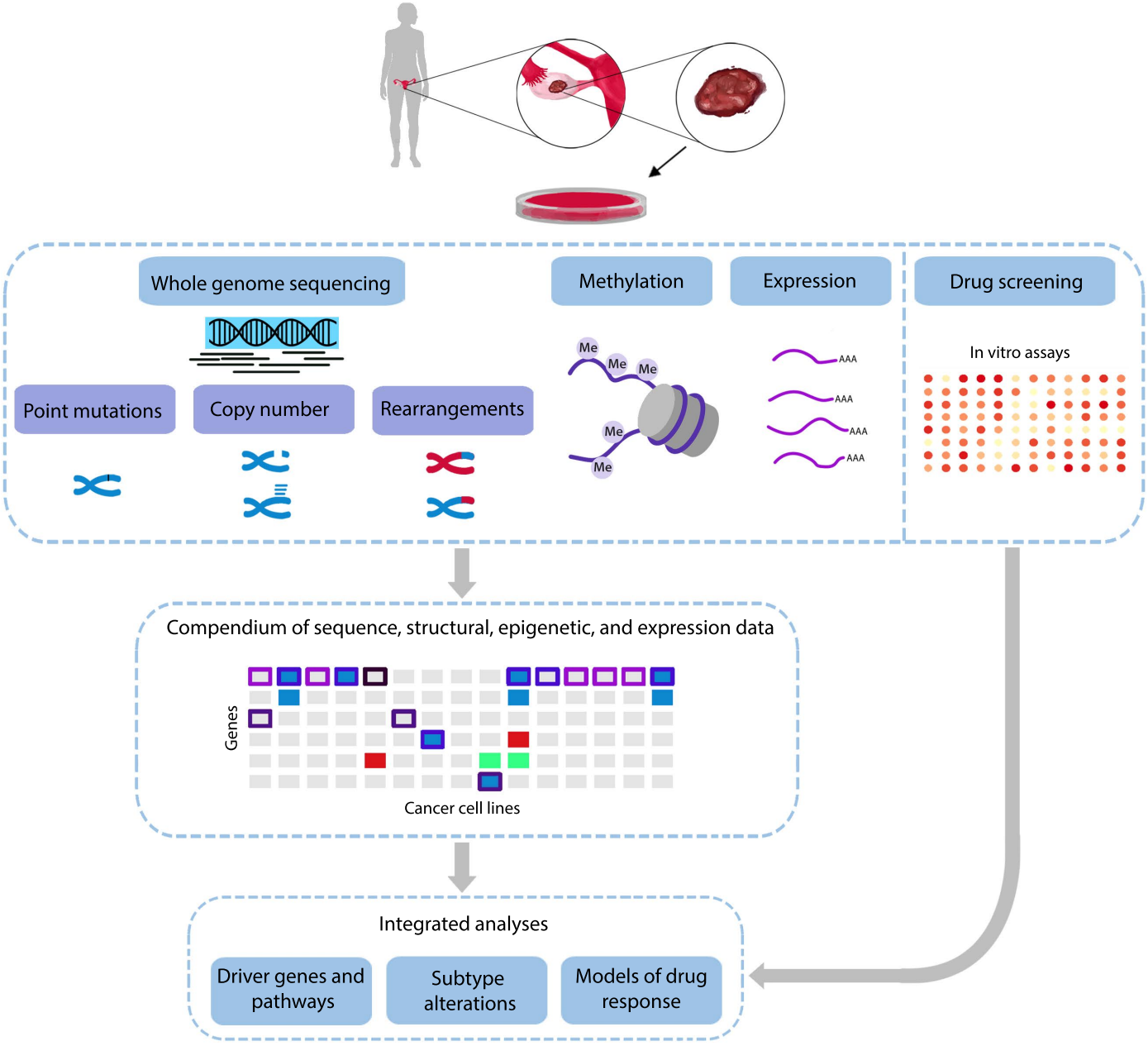
Integrated genomic, epigenomic, and expression analyses of ovarian cancer cell lines

Ovarian cancer

- Overall survival for late-stage ovarian cancer is poor with few new options for treatment
- Drug response ex-vivo can be as good or better as a biomarker than mutation
- We assembled a collection of 45 ovarian cancer cell lines of mixed subtypes (serous, clear cell, mucinous, endometrioid, mixed, and undifferentiated)
- We assessed the response of these cell lines to PARP inhibitor BMN673, PI3K inhibitor GNE-493, and MEK inhibitor MEK162
- In parallel, we characterized genomic, epigenomic, and expression changes



Papp, Hallberg et al., Cell Reports (In press)

Goals

- **Identify likely somatic structural variants**
- Integrate methylation and expression platforms
- Identify mutations associated with response to drug

Limitations

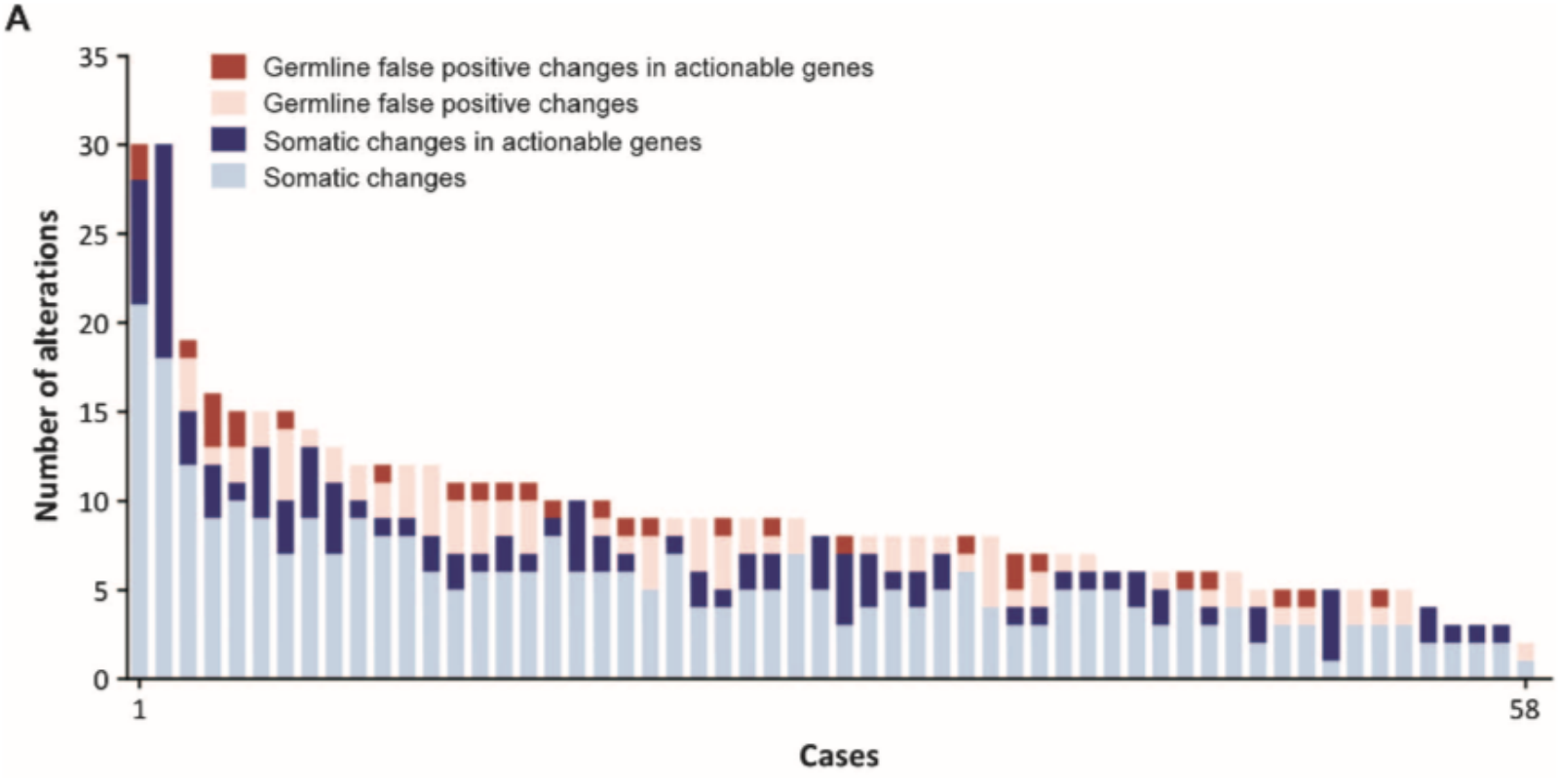
Match tumor-normal

- Identification of somatic mutations
- Identification of clinically actionable germline mutations

Tumor-only

- *Likely* somatic mutations
- *Likely* germline mutations of uncertain significance

False positives with tumor-only sequencing

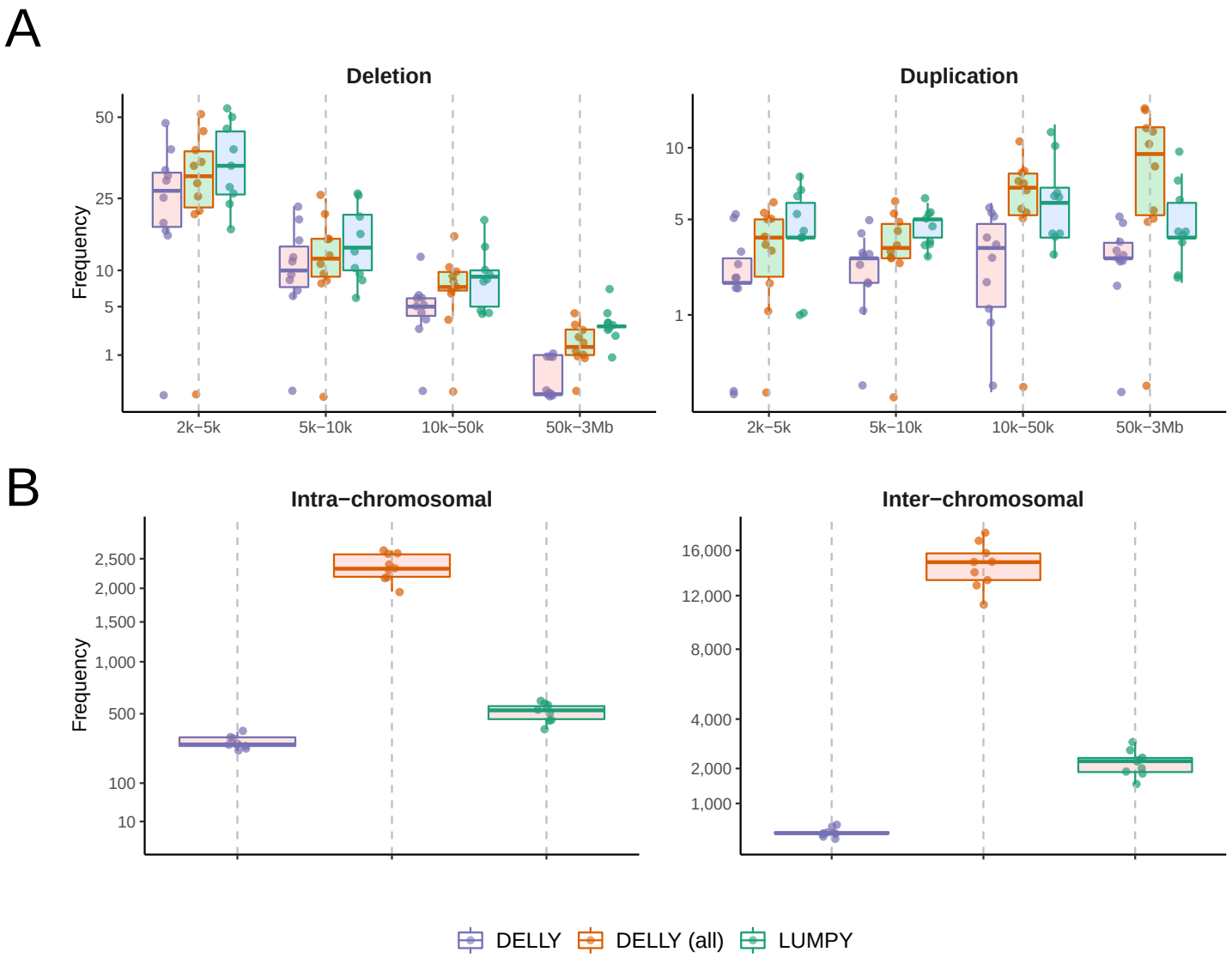


Jones et al., Sci Transl Med (2015)

Initial approach

- Prepare and sequence a collection of lymphoblastoid cell lines at the same time as the ovarian cell lines
- Use available methods to identify structural variants
- Likely somatic: mutations identified only in the tumor cell lines and in none of the lymphoblastoid cell lines
- Use LOOCV with only the lymphblastoid cell lines to estimate the number of false positive somatic mutations

Number of false positives



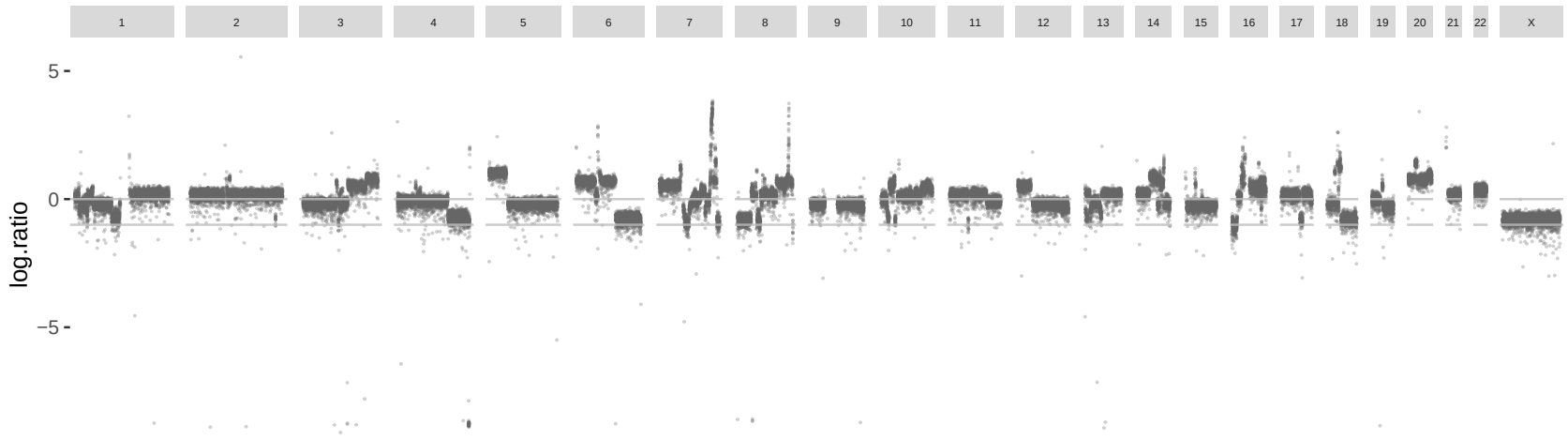
Improving specificity in tumor-only analyses

- Sequence a much larger collection of unmatched normals
- Use existing germline databases (e.g., 1000 Genomes Project, Database of Genomic Variants)
- *Improve filtering of likely spurious alignments*
- *Restrict inference to mutations more likely to be somatic* (large homozygous deletions, high copy focal amplifications, distant rearrangements)

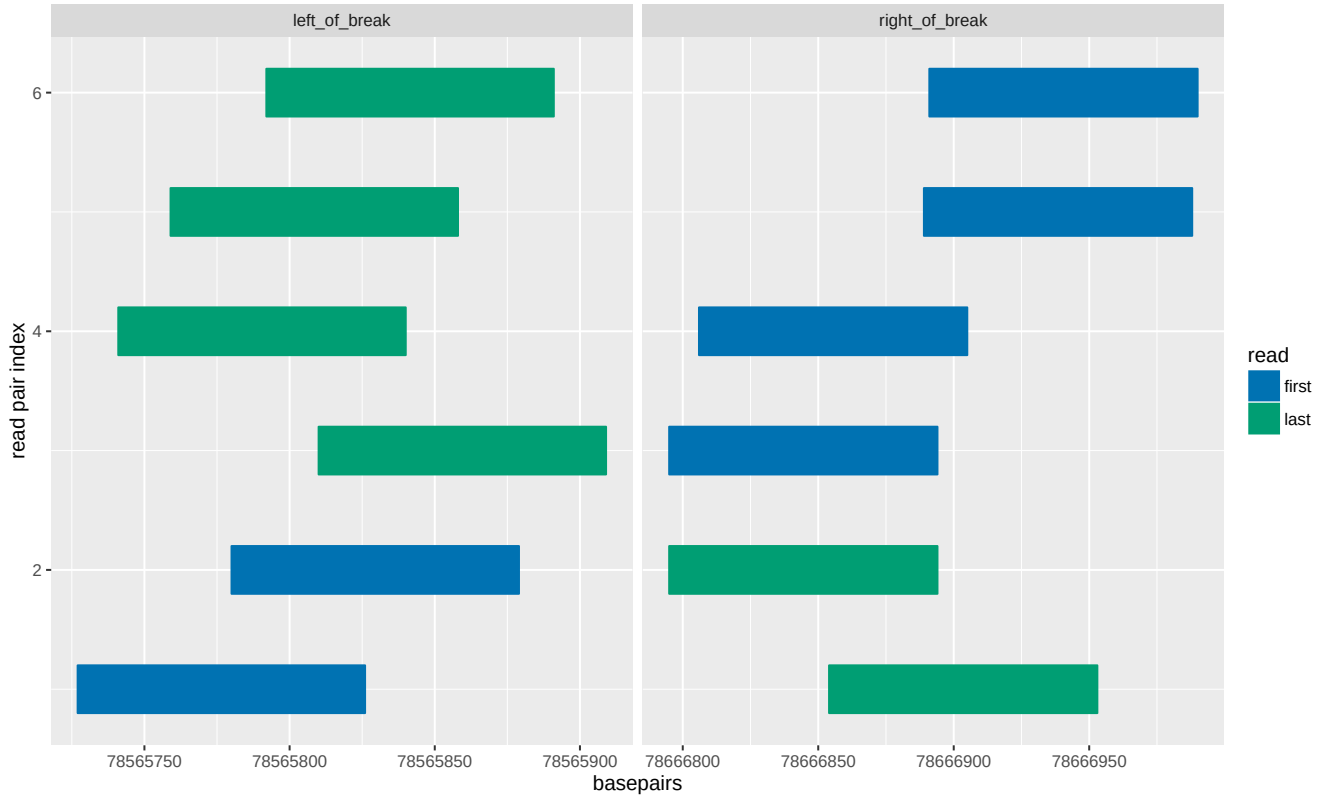
Criteria

- Amplifications/deletions must be focal: $\geq$ 2kb and $<$ 3Mb
- Amplifications: $\geq$ 3-fold modal ploidy
- Homozygous deletions
- Focal heterozygous deletions: called only if supported by both read depth and improperly paired reads
- Rearrangements: re-align all read pairs and split reads by BLAT
  - New sequence junction must be at least 10kb apart in the reference genome

1kb-bin level counts



Amplicons can be linked by read pairs



- y-axis is ordered by the start of the first read

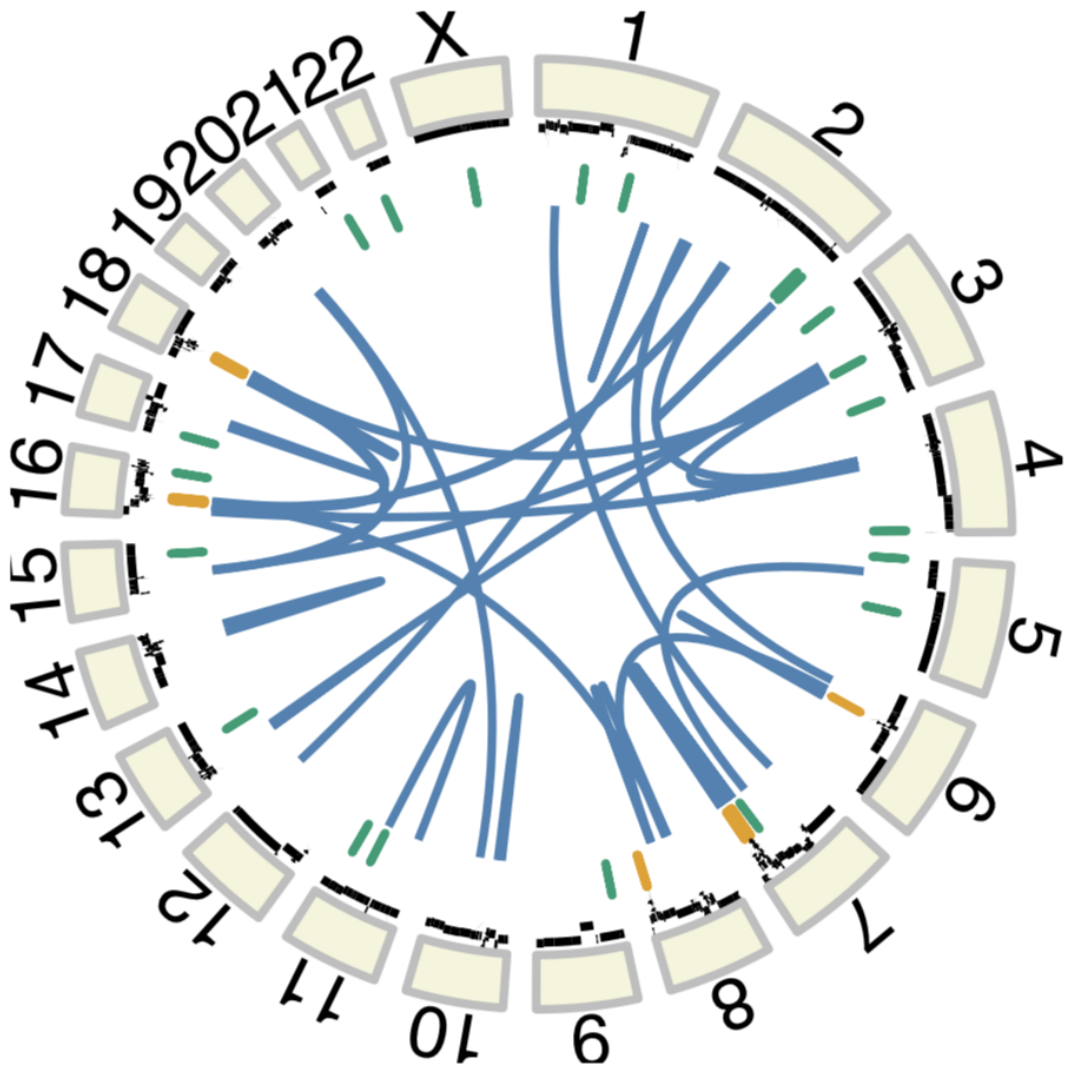
Simplifies downstream analyses

Lab ID	Cell line	Chr	Start	End	Width	Log_2 ratio	Gene	Previously implicated in cancer	Amplicon group
CGOV16T	OAW-42	chr17	73,535,001	73,768,001	233,001	1.47	MYO15B, SMIM5, SMIM6, SAP30BP, ITGB4, RECQL5, GALK1	NA	1
CGOV17T	OV-167	chr2	43,161,001	43,391,001	230,001	1.70	LOC102723854	NA	1
CGOV17T	OV-167	chr2	58,679,001	58,714,001	35,001	1.64	NA	NA	3
CGOV17T	OV-167	chr2	64,133,001	64,263,001	130,001	1.60	NA	NA	1
CGOV17T	OV-167	chr2	64,411,001	64,656,001	245,001	1.63	LOC100507006, MIR4433A, LINC00309, MIR4433B	NA	1
CGOV17T	OV-167	chr2	86,027,001	86,209,001	182,001	1.72	LOC284950, ST3GAL5-AS1, ST3GAL5	NA	4
CGOV17T	OV-167	chr2	113,774,001	113,917,001	143,001	1.53	IL36RN, IL1F10, IL1RN, IL36B	NA	5
CGOV17T	OV-167	chr2	154,543,001	154,559,001	16,001	1.73	NA	NA	6
CGOV17T	OV-167	chr11	101,039,001	101,508,001	469,001	1.70	TRPC6, MIR3920	NA	1
CGOV17T	OV-167	chr11	102,443,001	102,457,001	14,001	1.95	NA	NA	1
CGOV17T	OV-167	chr11	102,458,001	102,512,001	54,001	2.37	NA	NA	1
CGOV17T	OV-167	chr11	102,513,001	102,522,001	9,001	1.77	NA	NA	1
CGOV17T	OV-167	chr12	127,798,001	127,799,001	1,001	1.50	NA	NA	2
CGOV17T	OV-167	chr12	127,800,001	127,842,001	42,001	1.89	LOC101927616	NA	2

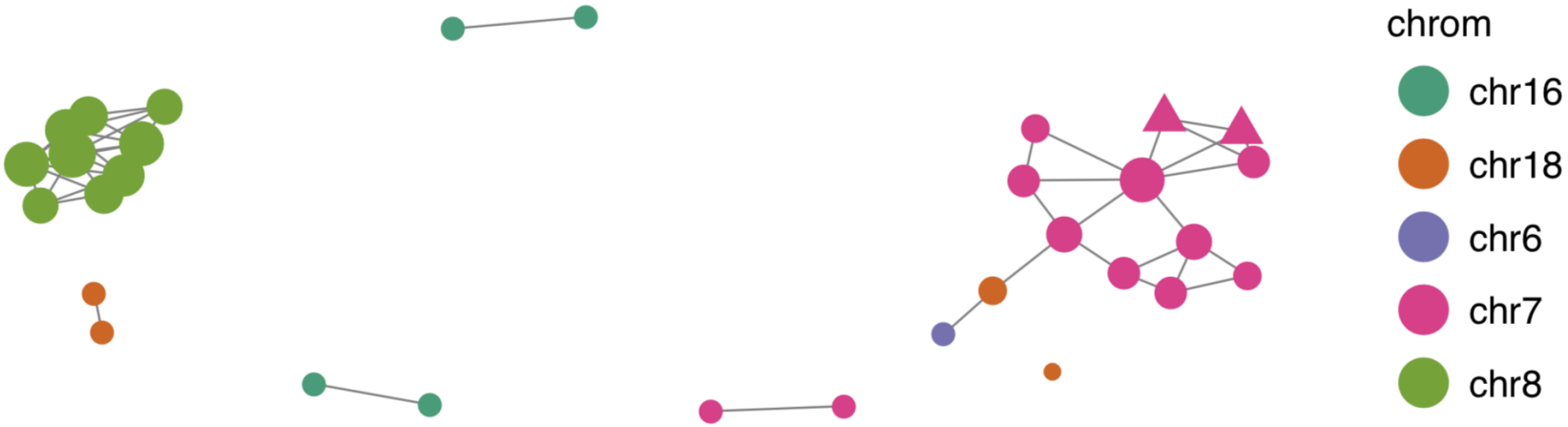
Simplifies downstream analyses

CGOV13T	Kuramochi	chr8	119,165,001	125,251,001	6,086,001	1.48	SAMD12-AS1, COLEC10, MAL2, NOV, DEPTOR, COL14A1, MTBP, LOC101927543, HAS2-AS1, ZHX2, TBC1D31, FAM83A, WDYHV1, FAM91A1, FER1L6, SAMD12, TNFRSF11B, ENPP2, TAF2, DSCC1, MRPL13, SNTB1, HAS2, LOC105375734, LINC01151, DERL1, FAM83A-AS1, MIR4663, C8orf76, ZHX1-C8orf76, ZHX1, ATAD2, FBXO32, KLHL38, ANXA13, FER1L6-AS1, FER1L6-AS2	MYC, KRAS	1
CGOV13T	Kuramochi	chr8	125,261,001	128,755,001	3,494,001	1.77	TRMT12,	MYC,	1

Rearrangements and copy number changes for one cell line

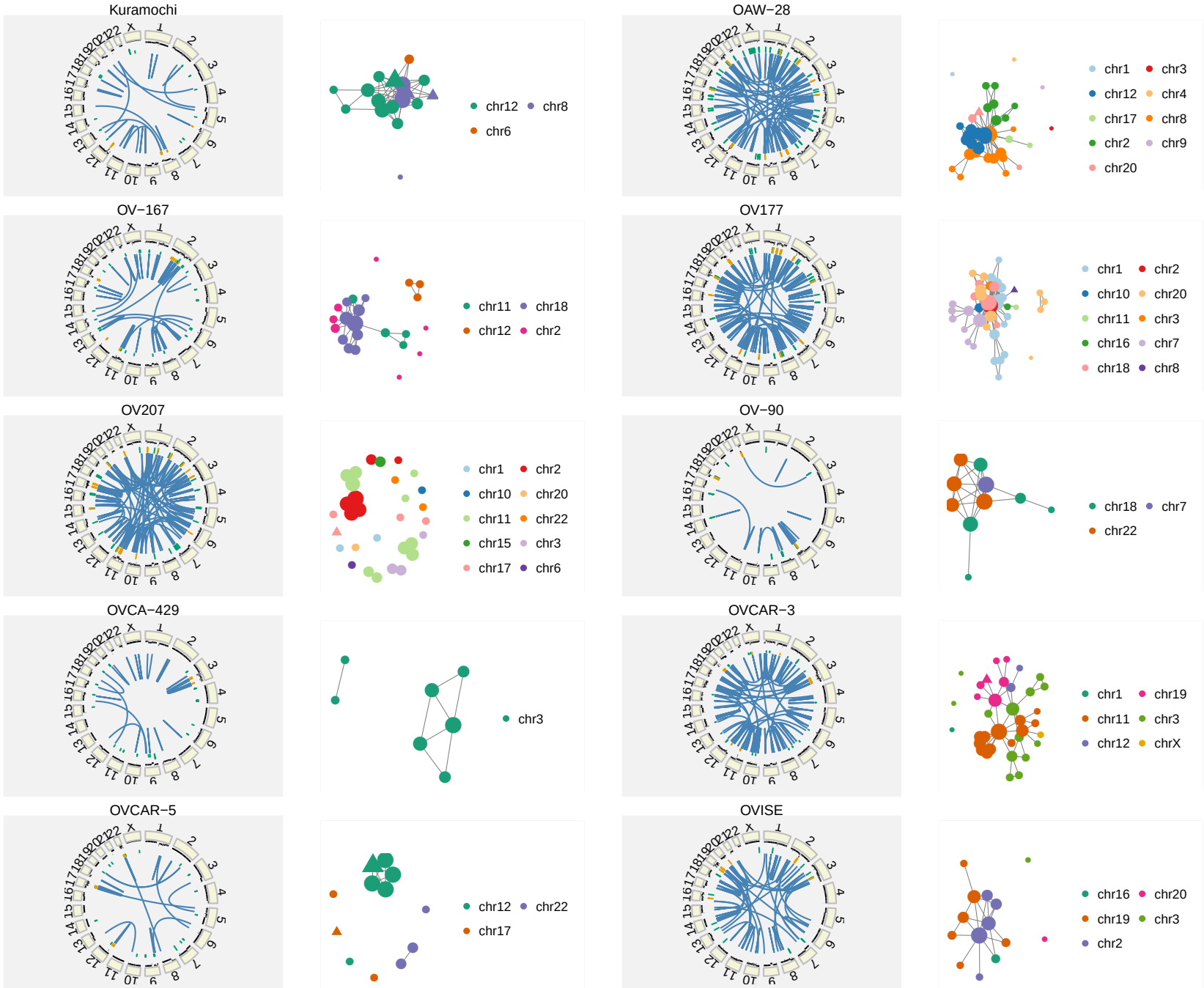


Graph structure

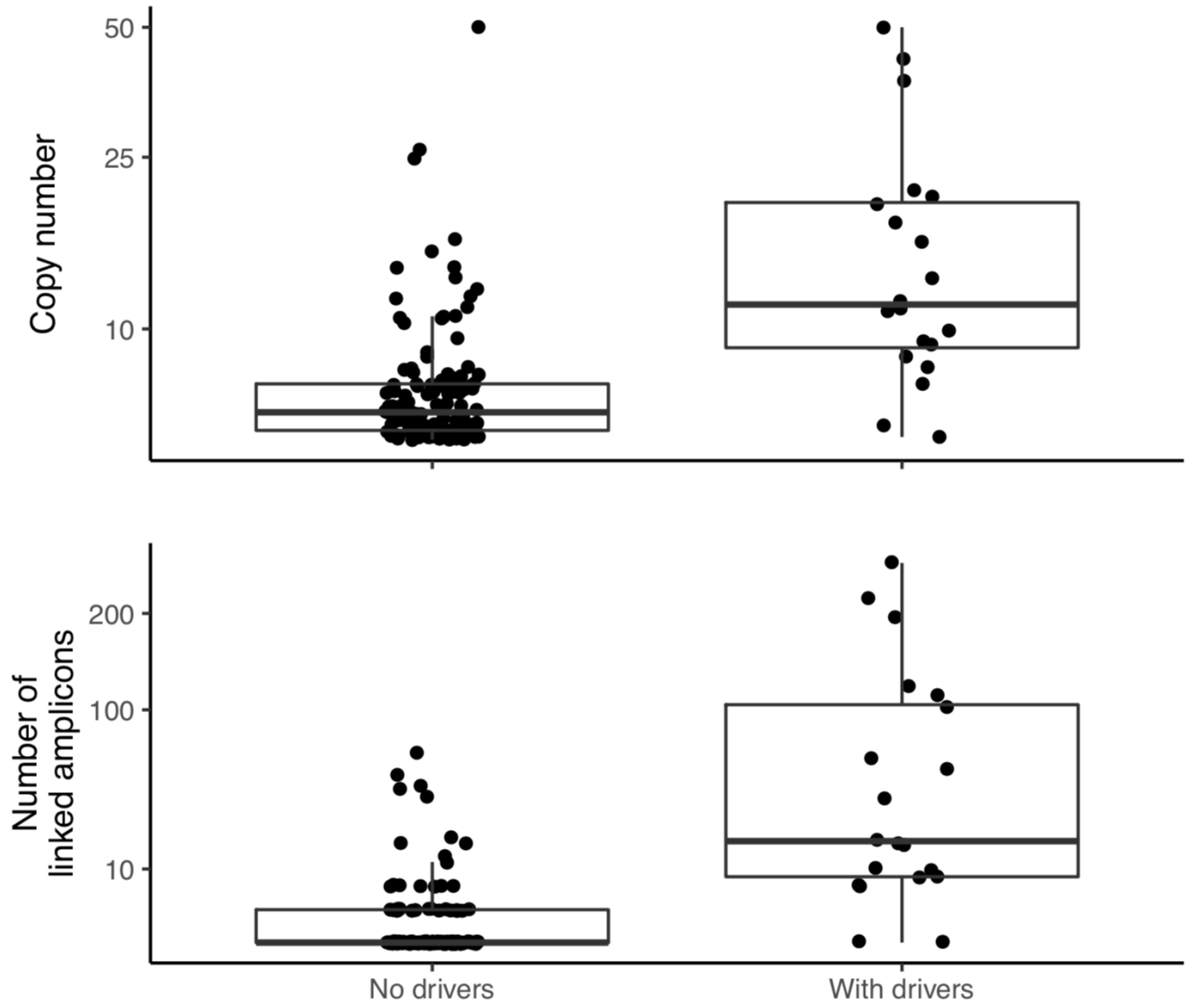


- amplicons are nodes
- an edge indicates 5 or more read pairs connecting the amplicons
- size of symbol proportional to number of connections
- triangles: amplicons containing known drivers

Rearrangements and connected amplicons

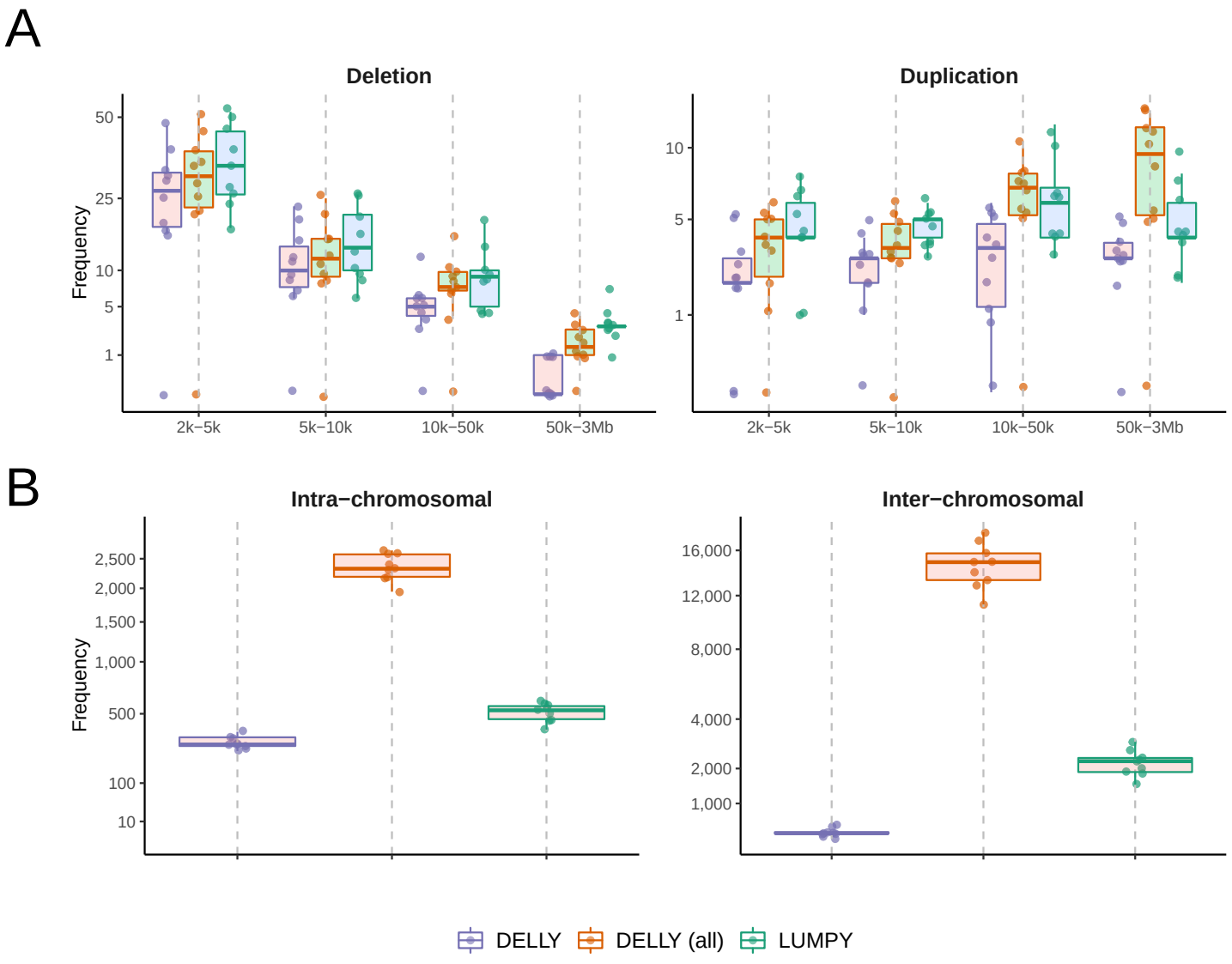


Drivers were more connected and had higher copy number

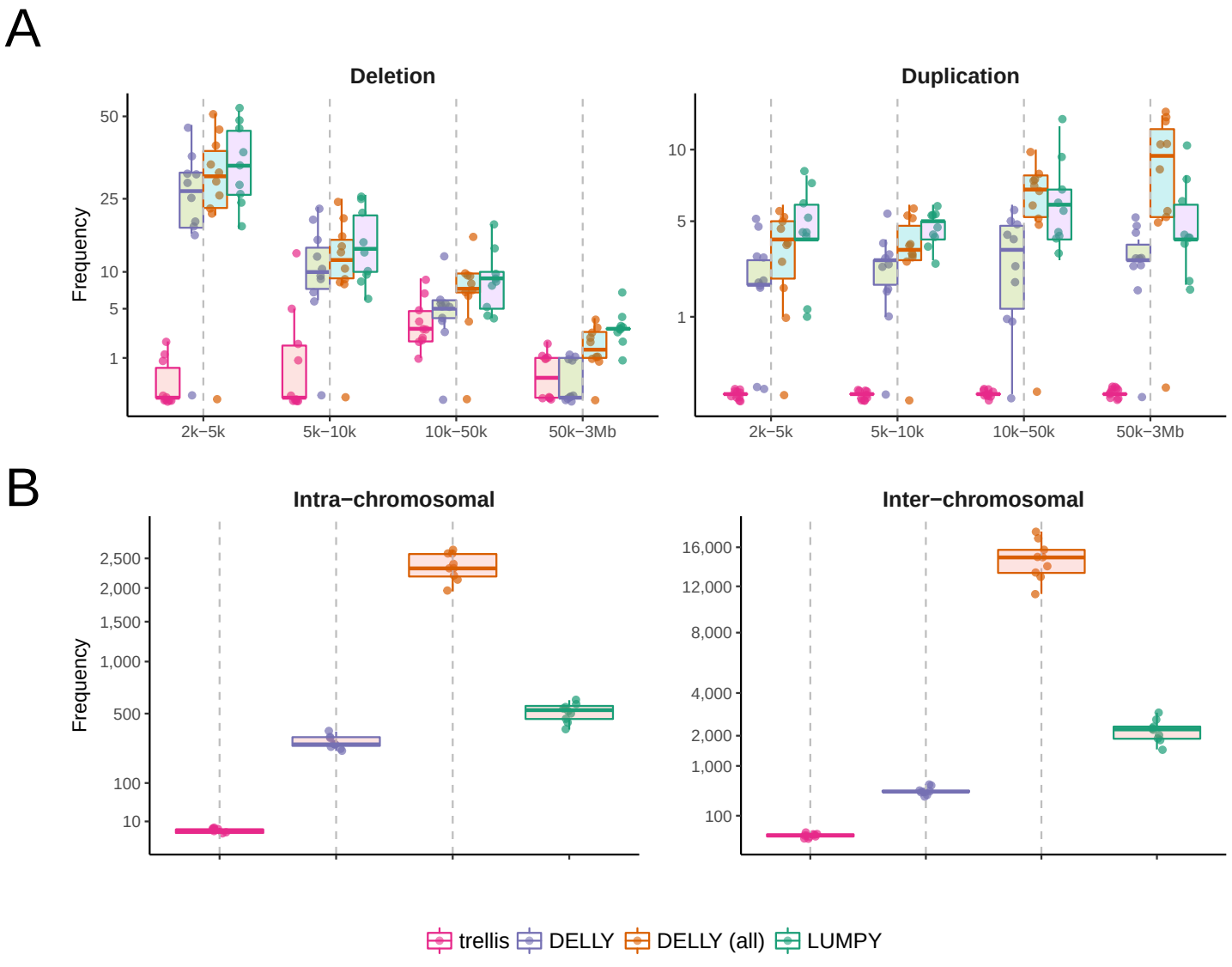


- Implication for tumor phylogenetic analyses

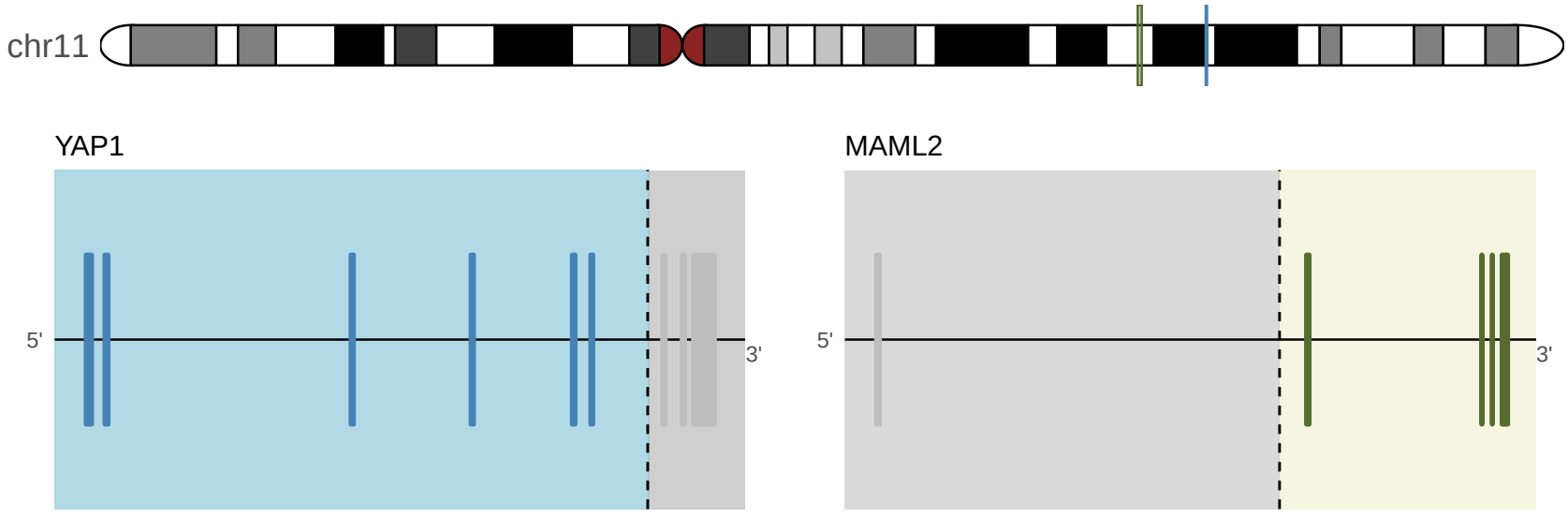
False positives



False positives



In-frame fusions

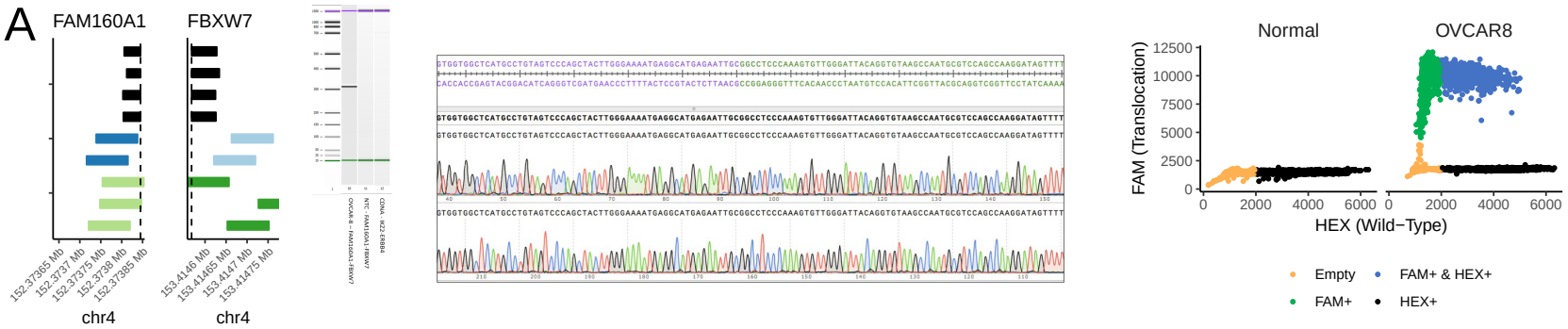


- The YAP1-MAML2 fusion has been previously identified in nasopharyngeal carcinoma
  - Same exons but different sequence junction

In-frame fusions

- Nine in-frame fusions involving clinically relevant genes were identified
- Sequence junctions were intronic
- No recurrent fusions were identified

Validation of fusions



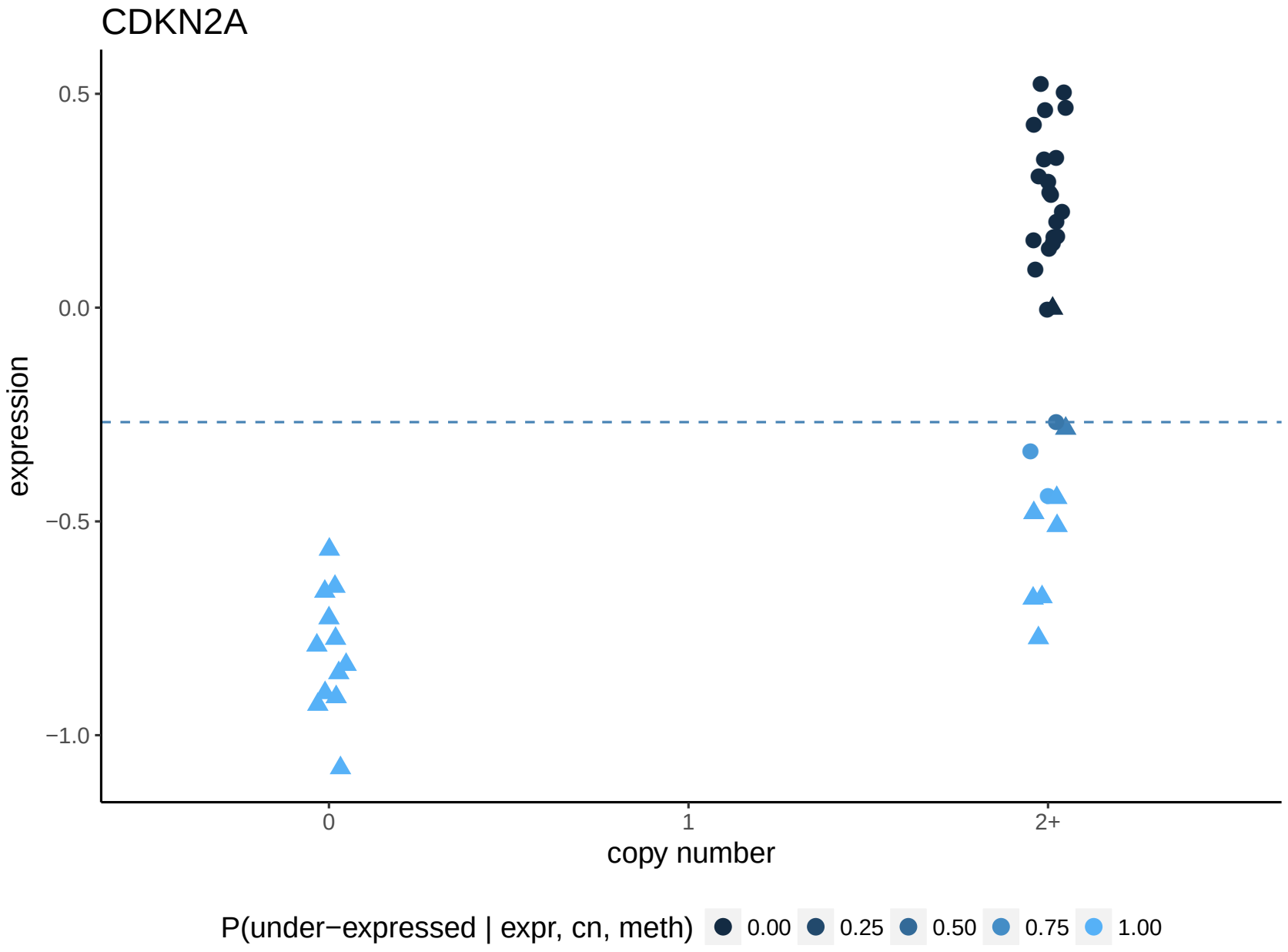
Goals

- Identify likely somatic structural variants
- **Integrate methylation and expression platforms**
- Identify mutations associated with response to drug

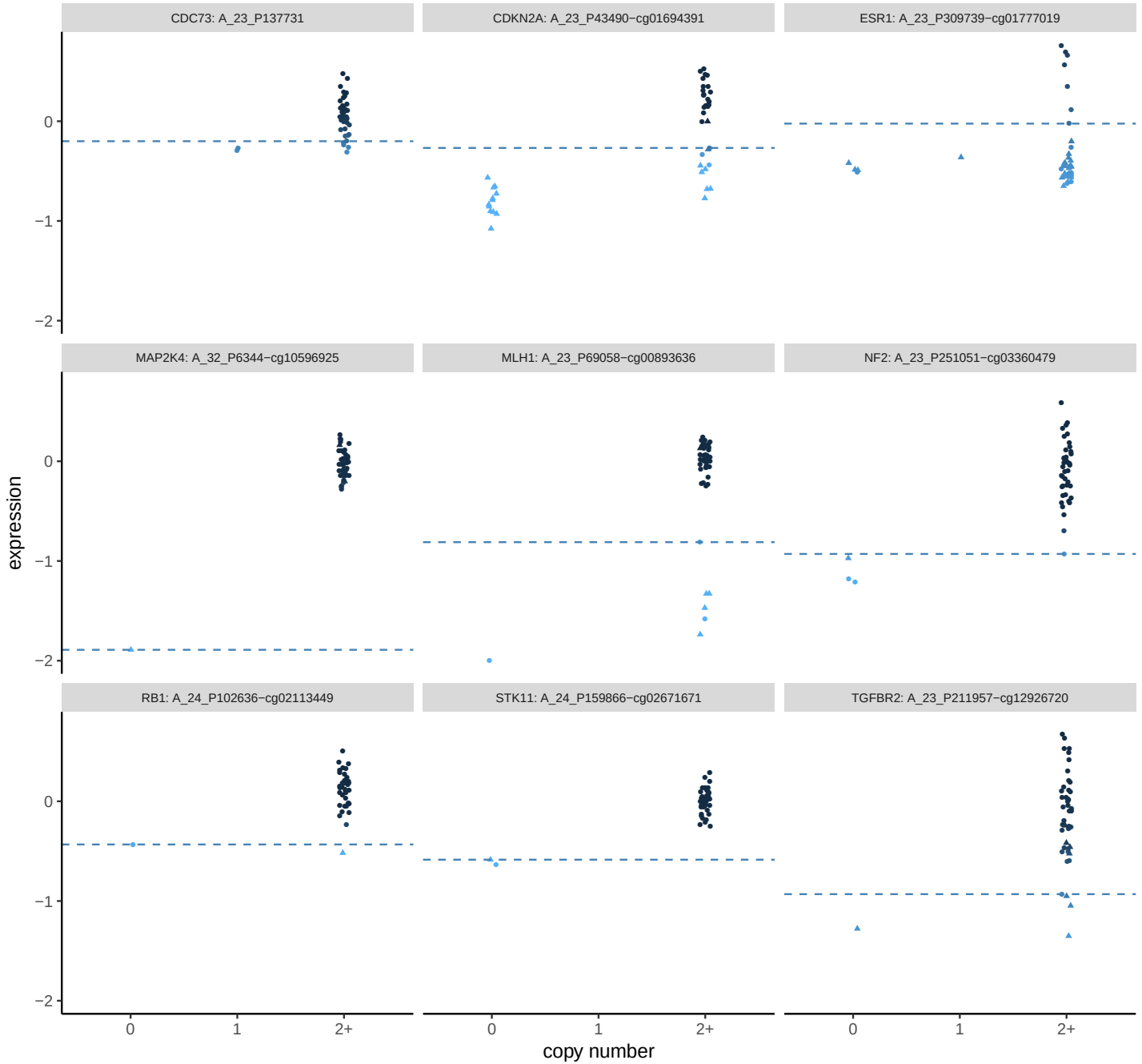
Integration anchored by DNA analyses

- Processed Infinium MethylationEPIC arrays with `minfi`
- Focused on genes with recurrent deletions or amplifications to anchor interpretation of other platforms
- Are there additional ovarian cancer cell lines that we can identify as under- or over-expressed?

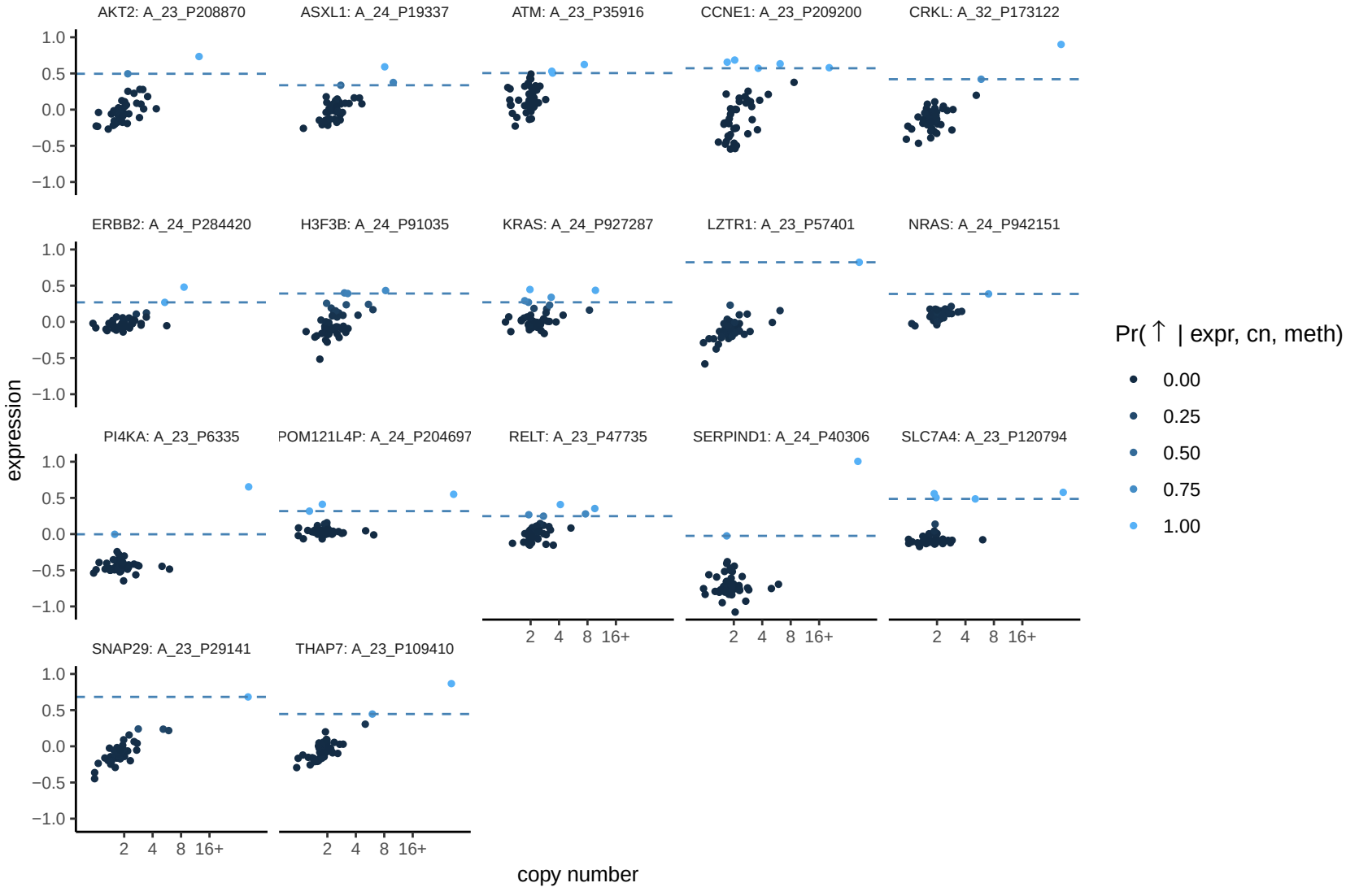
Integration of methylation & expression platforms

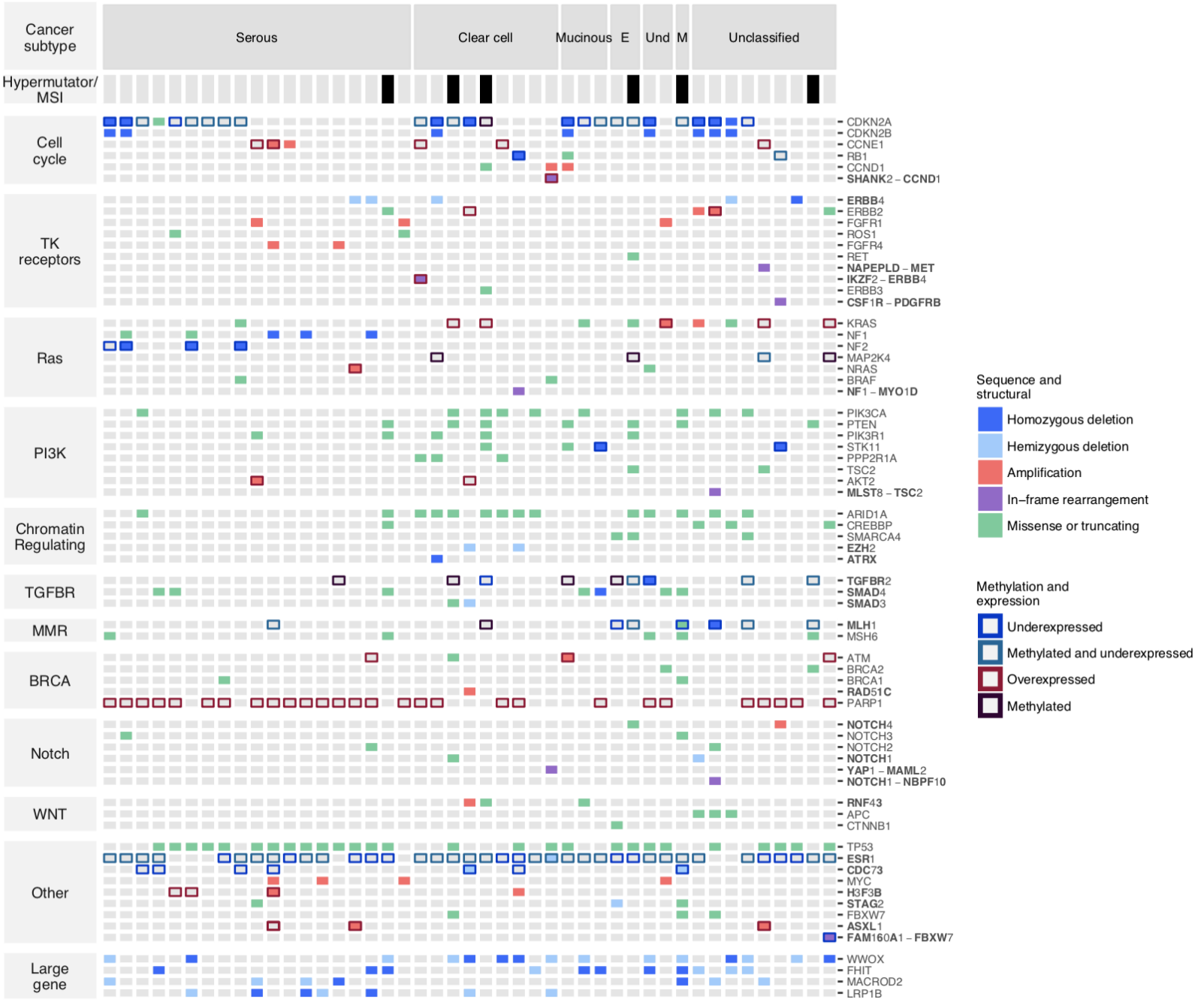


Integration of methylation & expression platforms



Expression of amplified genes




Cancer subtype
Serous
Clear cell
Mucinous
E
Und
M
Unclassified
Hypermutator/ MSI
Cell cycle
CDKN2A
CDKN2B
CCNE1
RB1
CCND1
SHANK2 – CCND1
TK receptors
ERBB4
ERBB2
FGFR1
ROS1
FGFR4
RET
NAPEPLD – MET
IKZF2 – ERBB4
ERBB3
CSF1R – PDGFRB
Ras
KRAS
NF1
NF2
MAP2K4
NRAS
BRAF
NF1 – MYO1D
PI3K
PIK3CA
PTEN
PIK3R1
STK11
PPP2R1A
TSC2
AKT2
MLST8 – TSC2
Chromatin Regulating
ARID1A
CREBBP
SMARCA4
EZH2
ATRX
TGFBR
TGFBR2
SMAD4
SMAD3
MMR
MLH1
MSH6
BRCA
ATM
BRCA2
BRCA1
RAD51C
PARP1
Notch
NOTCH4
NOTCH3
NOTCH2
NOTCH1
YAP1 – MAML2
NOTCH1 – NBPF10
WNT
RNF43
APC
CTNNB1
Other
TP53
ESR1
CDC73
MYC
H3F3B
STAG2
FBXW7
ASXL1
FAM160A1 – FBXW7
Large gene
WWOX
FHIT
MACROD2
LRP1B
Sequence and structural
Homozygous deletion
Hemizygous deletion
Amplification
In–frame rearrangement
Missense or truncating
Methylation and expression
Underexpressed
Methylated and underexpressed
Overexpressed
Methylated

Goals

- Identify likely somatic structural variants
- Integrate methylation and expression platforms
- **Identify mutations associated with response to drug**

Drugs evaluated

- PARP inhibitor BMN673: target cancers with defective DNA-damage repair
- PI3K inhibitor GNE-493
- MEK inhibitor MEK162

Feature selection for drug sensitivity analyses

- Required at least 3 or more cell lines with an alteration (point mutation, structural variant, expression, or methylation)
- Bayesian model averaging:

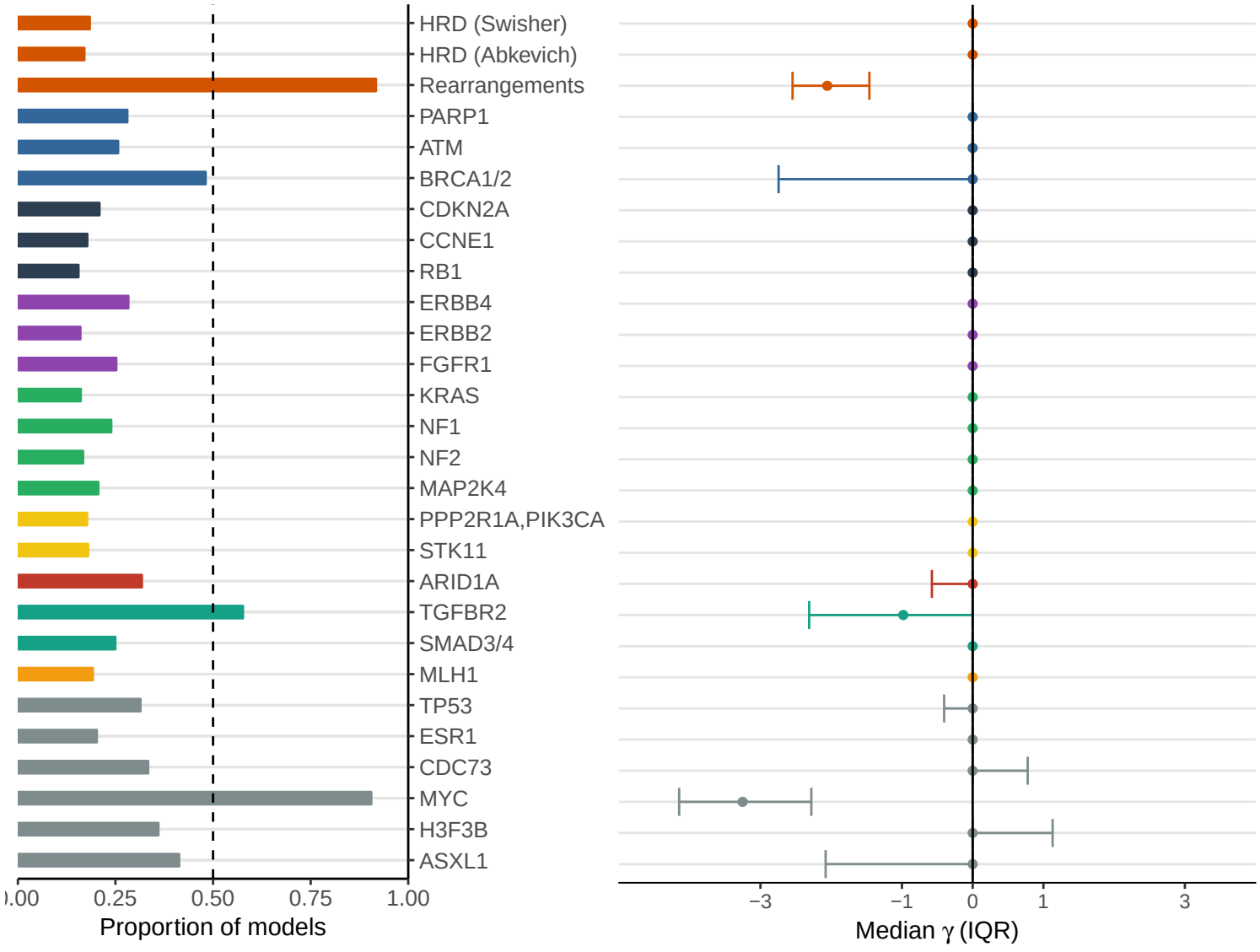
$$\log C_i = \gamma_1 x_{i,1} + \ldots + \gamma_p x_{i,p}, + \varepsilon_i, \text{ where}$$

C_i denotes the log IC_{50} for cell line i

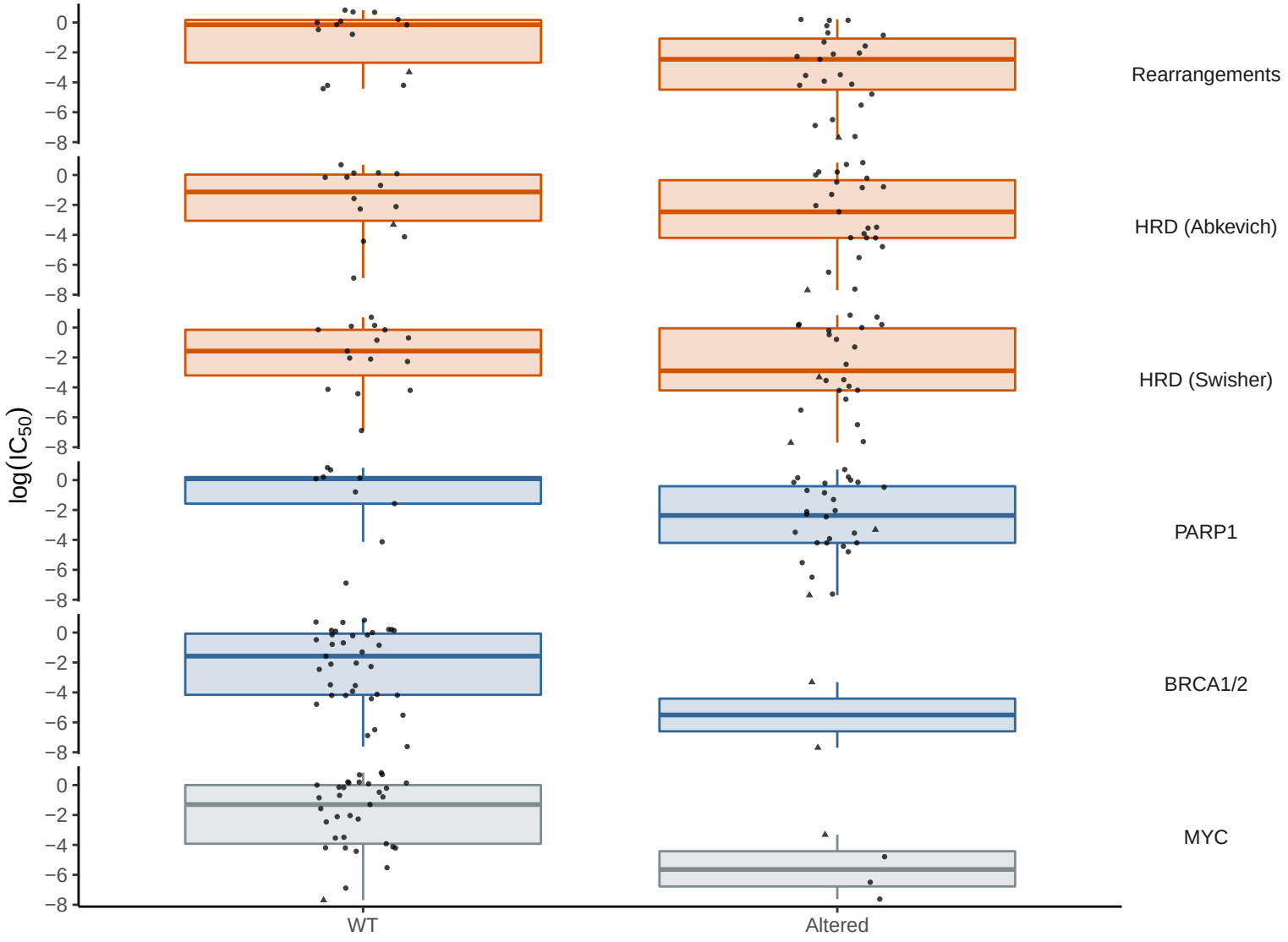
$x_{i,j}$ an alteration (binary for most variables)

- Regression coefficient for feature j is the product of a binary indicator and a real number
- Used MCMC to explore model space

Rearrangements associated with increased sensitivity to PARP inhibition



Univariate analyses



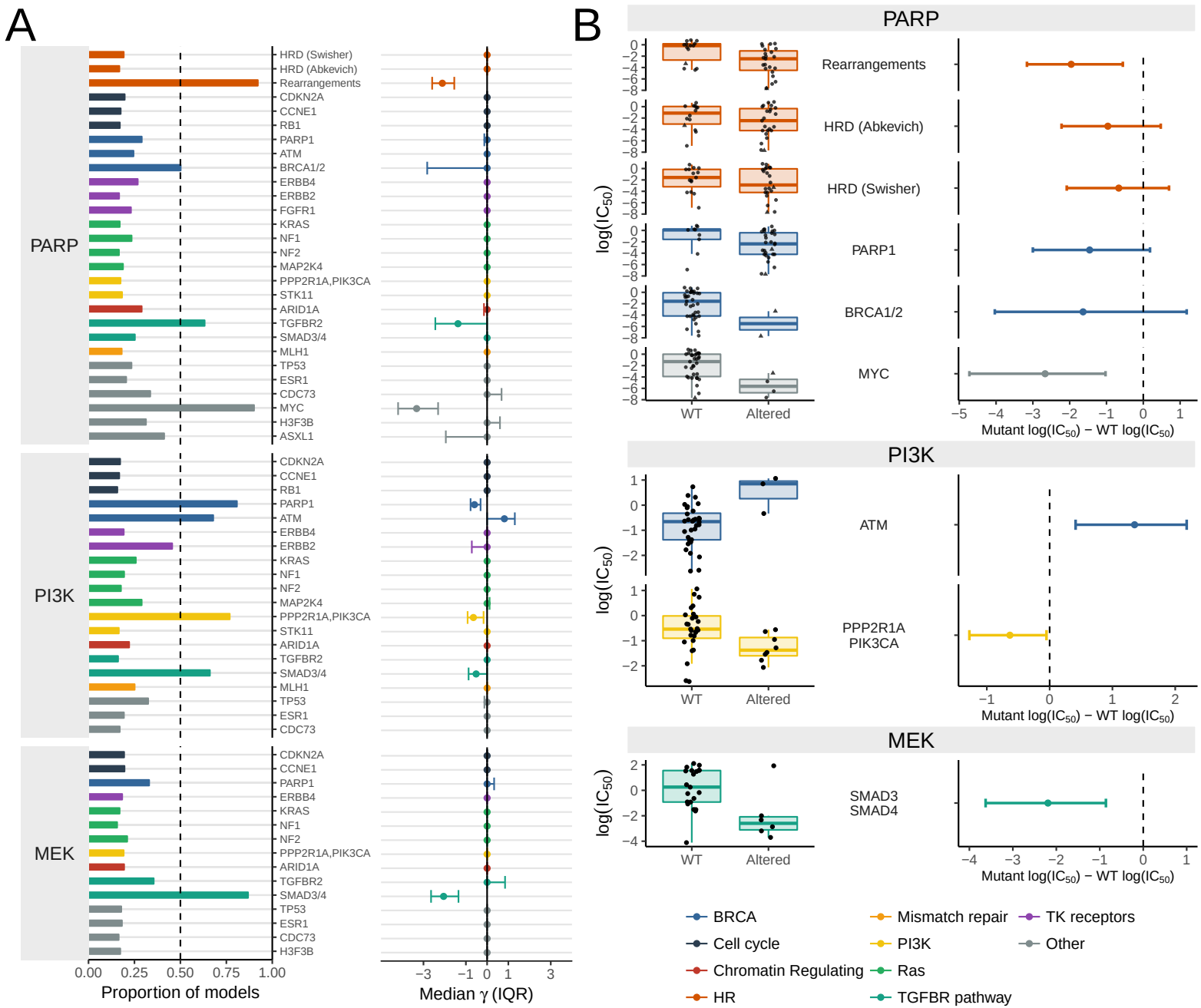
MYC amplification associated with increased sensitivity to PARP inhibition

- We found increased sensitivity of endometrioid and serous tumors to *MYC* (chr 8q)
- Amplification of *MYCN* (chr 2p) has been associated with increased sensitivity to PARP inhibitors in neuroblastoma
- *MYCN* amplified neuroblastoma cell lines were more sensitive to the PARP inhibitor BYK204165

Colicchia et al. 2017 (Oncogene)

Hallett et al., 2016 (Molecular Oncolology)

PARP, PI3K, and MEK inhibitors



Conclusions and summary

- We identified likely somatic variants in 45 widely used ovarian cell lines
- Grouped amplicons simplified downstream analysis and interpretation of the potential drivers
- Analysis of gene expression and integration of methylation platforms identified additional samples that would have been missed in a DNA-only analysis
- Structural variant analyses from whole genome sequencing may improve measures of homologous recombination defects
- *MYC* amplification may be useful for identifying patients responsive to PARP inhibitors
- R package trellis: https://github.com/cancer-genomics/trellis

Acknowledgements

Sidney Kimmel Cancer Institute, Johns Hopkins University

- **Victor Velculescu**
- **Eniko Papp**
- **Dorothy Hallberg**
- Daniel Bruhm
- Vilmos Adleff
- Michael Noe
- James White
- Doreen Palsgrove
- Stephen Baylin
- Hari Easwaran
- Ioannis Kagiampakis

Jonsson Comprehensive Cancer Center, David Geffen School of Medicine at UCLA

- **Gottfried Konecny**
- **Dennis Slamon**
- Dylan Conklin

Perelman School of Medicine, Unniversity of Pennsylvania

- Ronny Drapkin

Norris Comprehensive Cancer Center, University of Southern California

- Michael Press