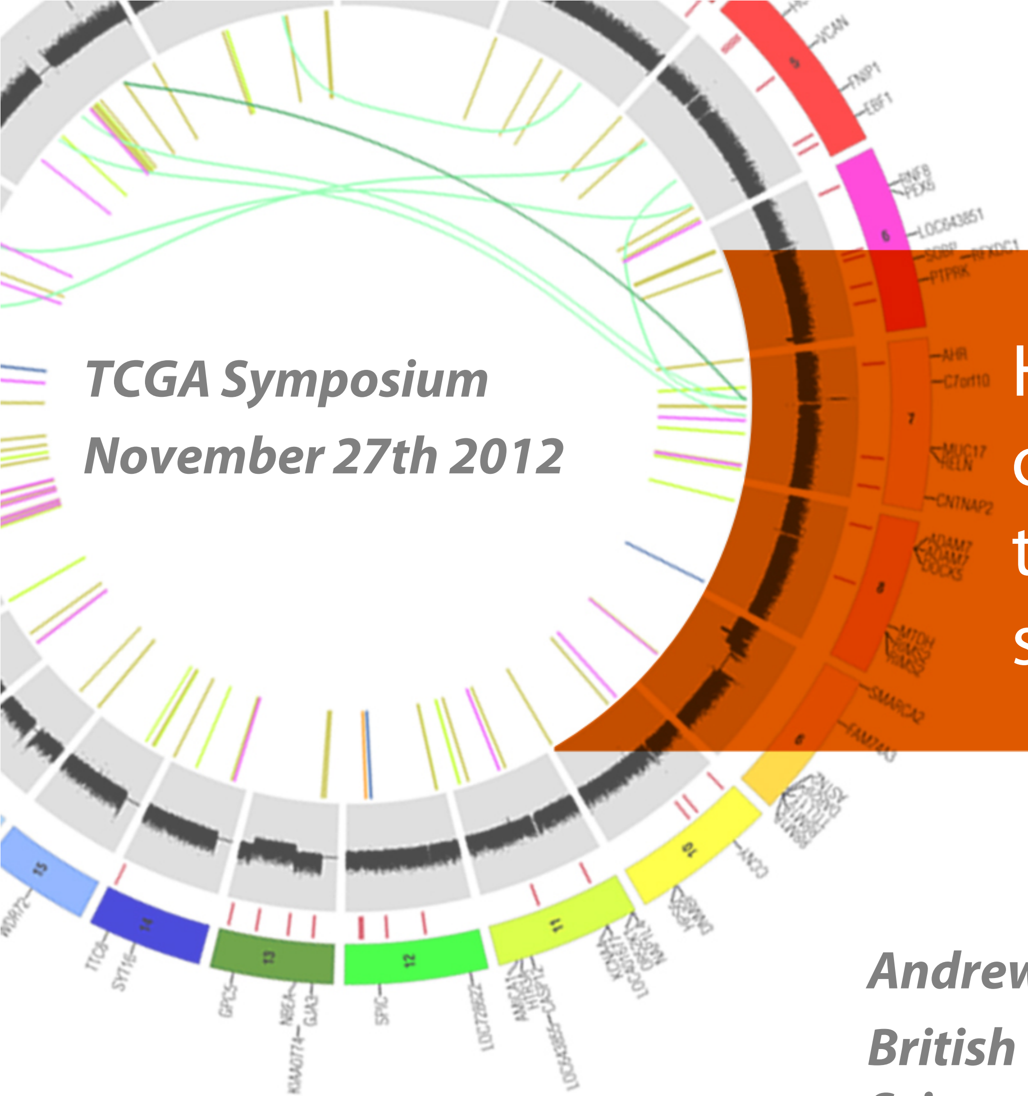


The Cancer Genome Atlas

TCGA Symposium

November 27th 2012

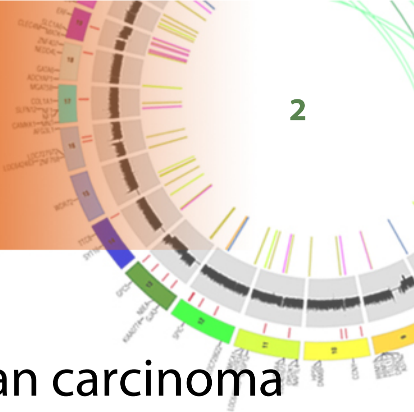
High-grade serous ovarian adenocarcinoma transcriptome sequencing

Andrew J. Mungall, Ph.D.

British Columbia Cancer Agency Genome Sciences Centre

amungall@bcgsc.ca

High-grade serous ovarian cancer



- Most deaths from advanced-stage, high-grade serous ovarian carcinoma
- Integrated genomic analyses of ovarian carcinoma (TCGA Research Network (2011) *Nature* **474**:609-615):
  - 489 tumours: mRNA, miRNA, DNA copy number and methylation.
  - 316 cases: exome T/N sequencing.
  - Simple mutational spectrum, *TP53* in 96% tumours
  - High frequency of somatic copy number aberrations

- Aim of this study
  - Transcriptome (mRNA & miRNA) sequencing: subtypes, structural variants and alternatively spliced transcripts.

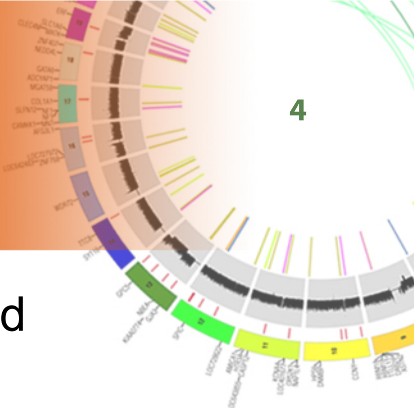
Transcriptome datasets

- 490 tumour samples (15 TSSs)
- 420 RNA-seq libraries sequenced
  - 420 submitted to CGHub and DCC
  - 300 expression datasets passed QC* submitted to DCC
- 485 miRNA-seq
  - All passed QC[#] & data submitted to CGHub and DCC
- Analyses:
  - Unsupervised NMF consensus clustering
  - miRNA anti-correlations with mRNA isoform expression
  - Trans-ABySS & UC-fusion-finder identification of gene fusions

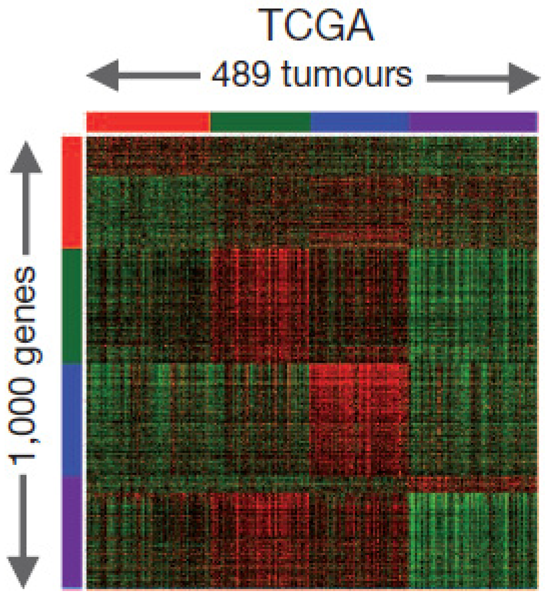
*>5Gb total; >21,000 genes; <20% rRNA; <20% mitochondrial; >0.6 5′/3′ ratio etc.

[#] >750,000M miRNA aligned reads

Sequence-based mRNA expression profiling suggests two additional groups



Microarray-based



Tumour/gene groups

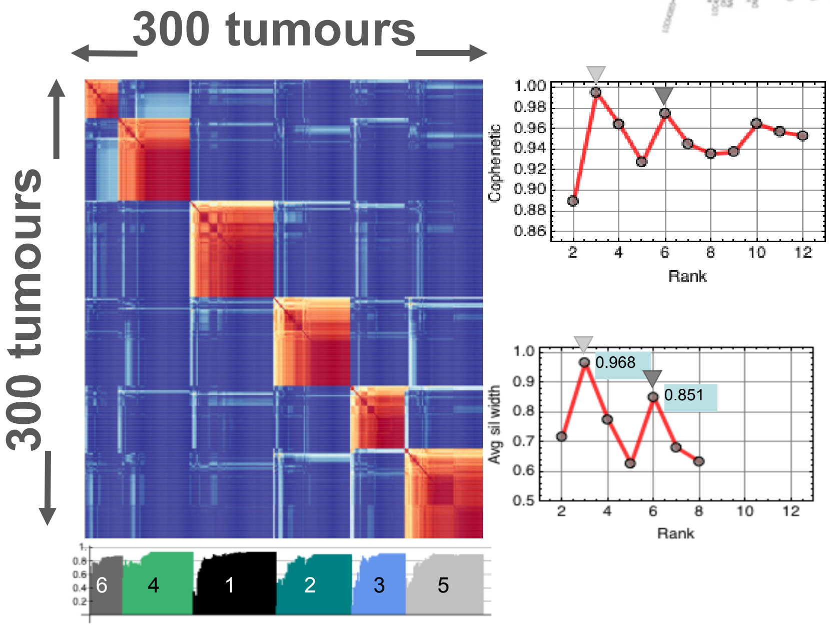
Differentiated 🟥
Immunoreactive 🟩
Mesenchymal 🟦
Proliferative 🟪

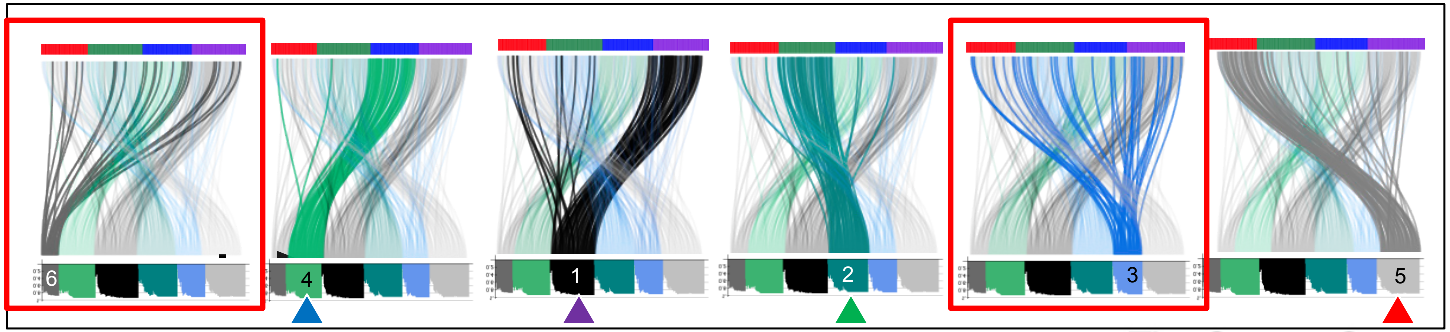
Gene expression

Low 🟩🟩🟩⬛🟥🟥🟥 High

TCGA Research Network (2011) Figure 2a

Sequence-based

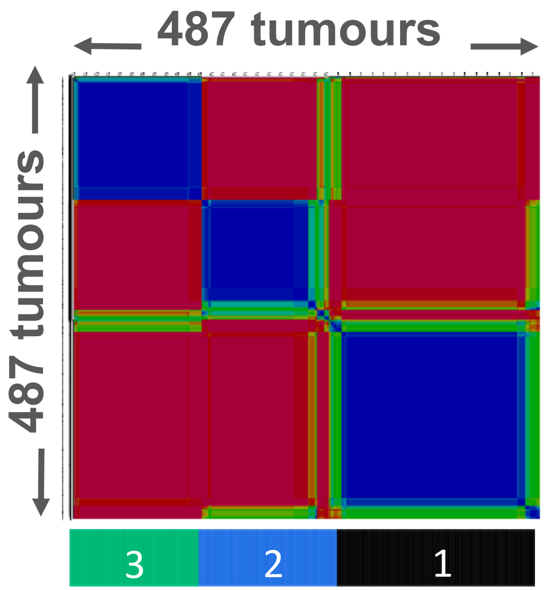




The Cancer Genome Atlas

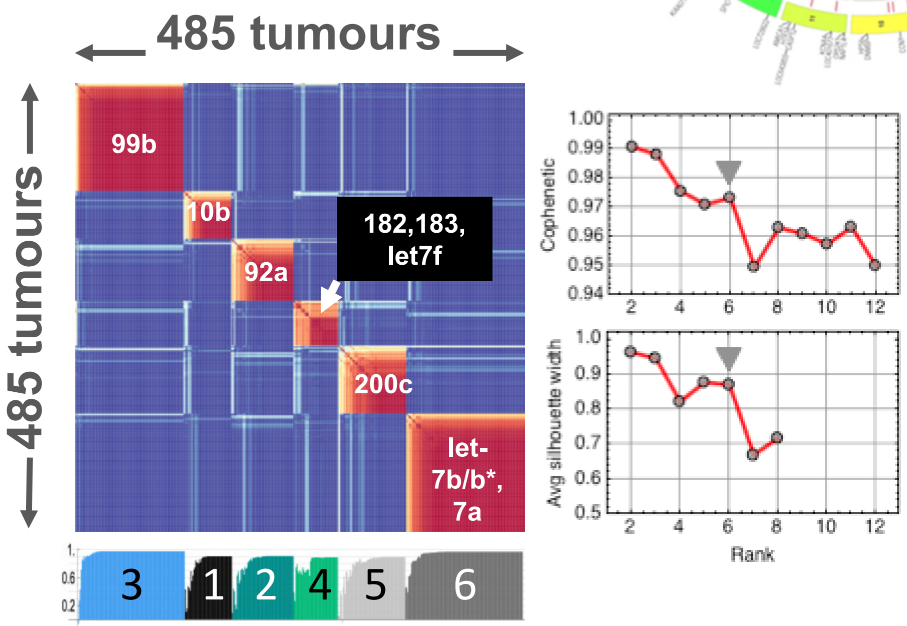
miRNA expression profiling identifies at least 3 clusters

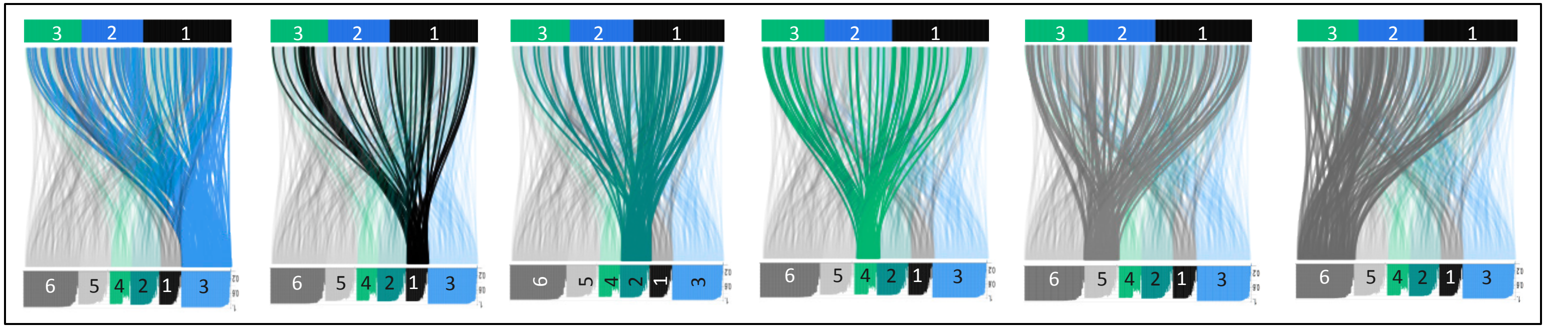
Microarray-based



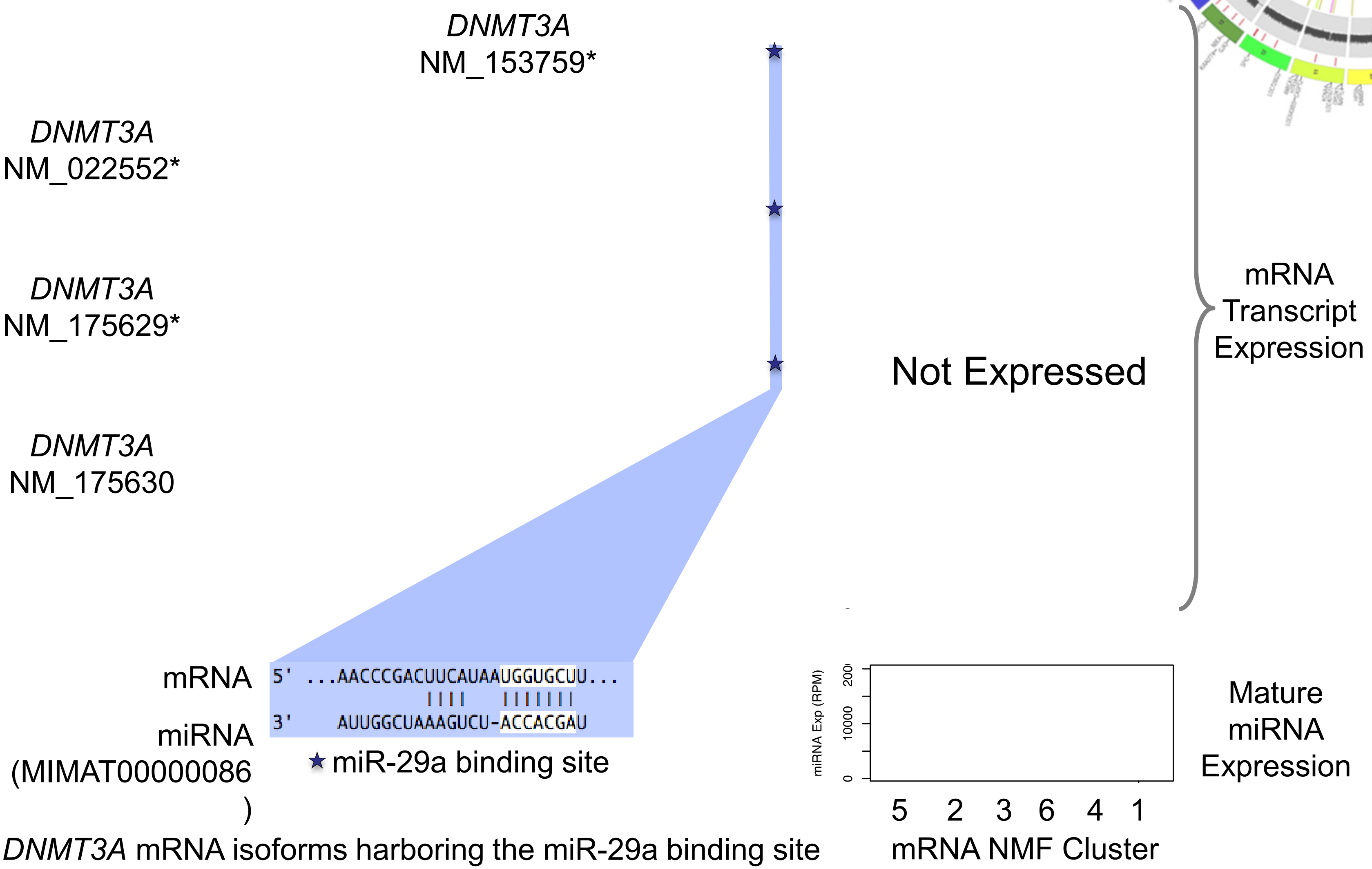
TCGA Research Network (2011) Figure S6.5

Sequence-based





Interplay between miR-29a and *DNMT3A* transcrip isoforms

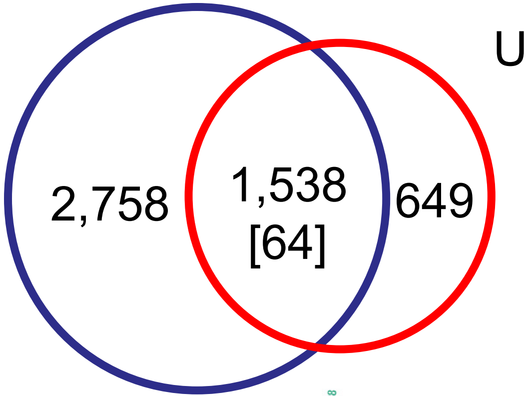


```
5' ...AACCCGACUUCAUAAUGGUGCUU...
                ||||   |||||||
3'    AUUGGCUAAAGUCU-ACCACGAU
```

Only *DNMT3A* mRNA isoforms harboring the miR-29a binding site have negatively correlated expression profiles with miR-29a.

Gene fusions

- No tumour total RNA available for verification, so orthogonal analysis methods were used
- 1,538 calls overlap between the two methods
- 64 recurrent (≥2) gene fusions



Trans-ABySS
n=420

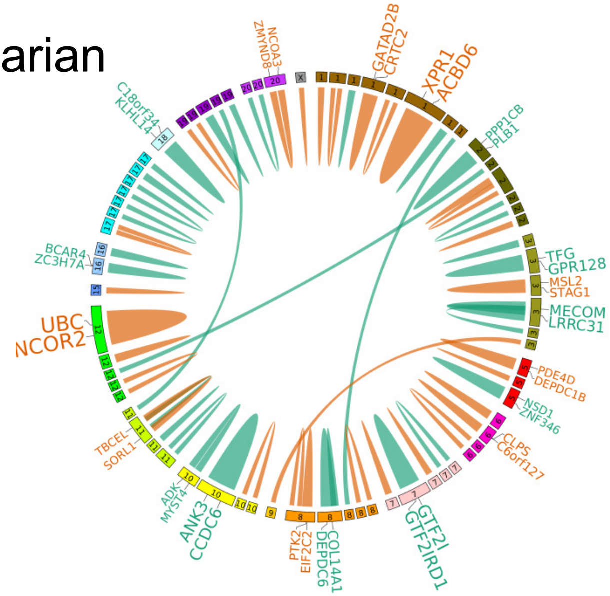
UC-fusion-finder
n=394

2,758

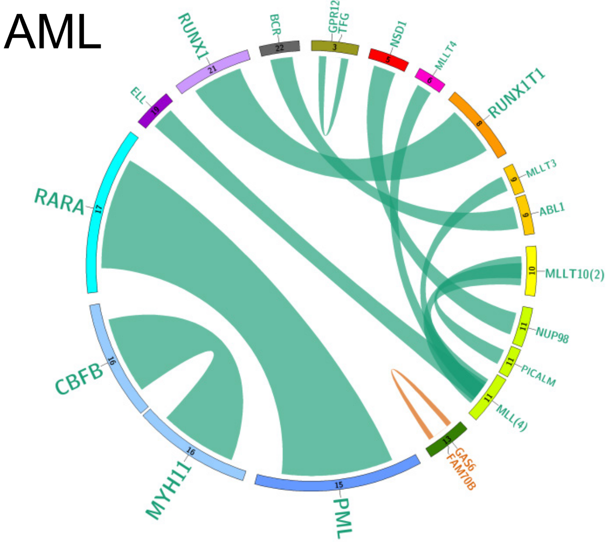
1,538
[64]

649

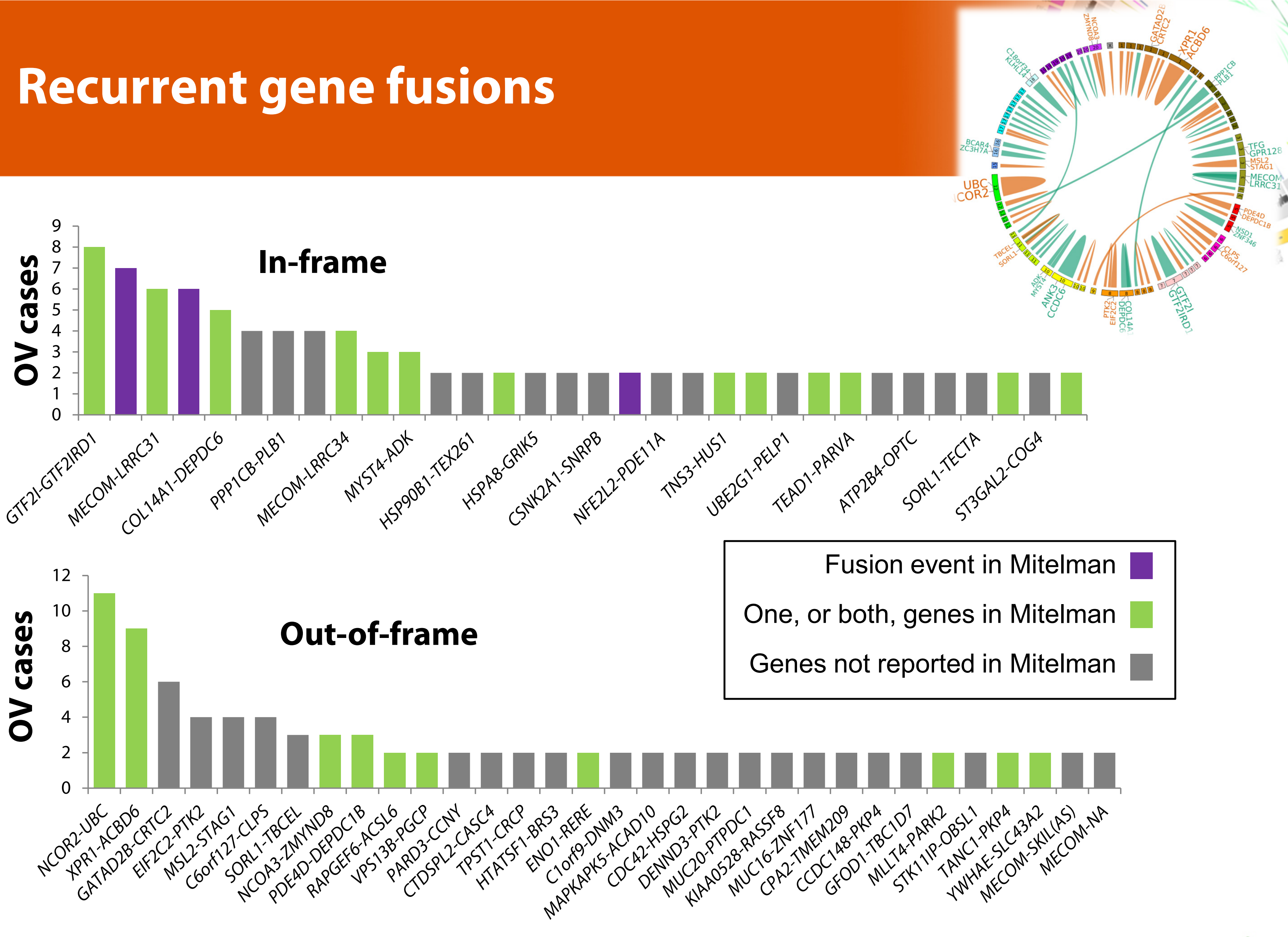
Ovarian



AML



Recurrent gene fusions



Fusion event in Mitelman 🟪

One, or both, genes in Mitelman 🟩

Genes not reported in Mitelman ◼

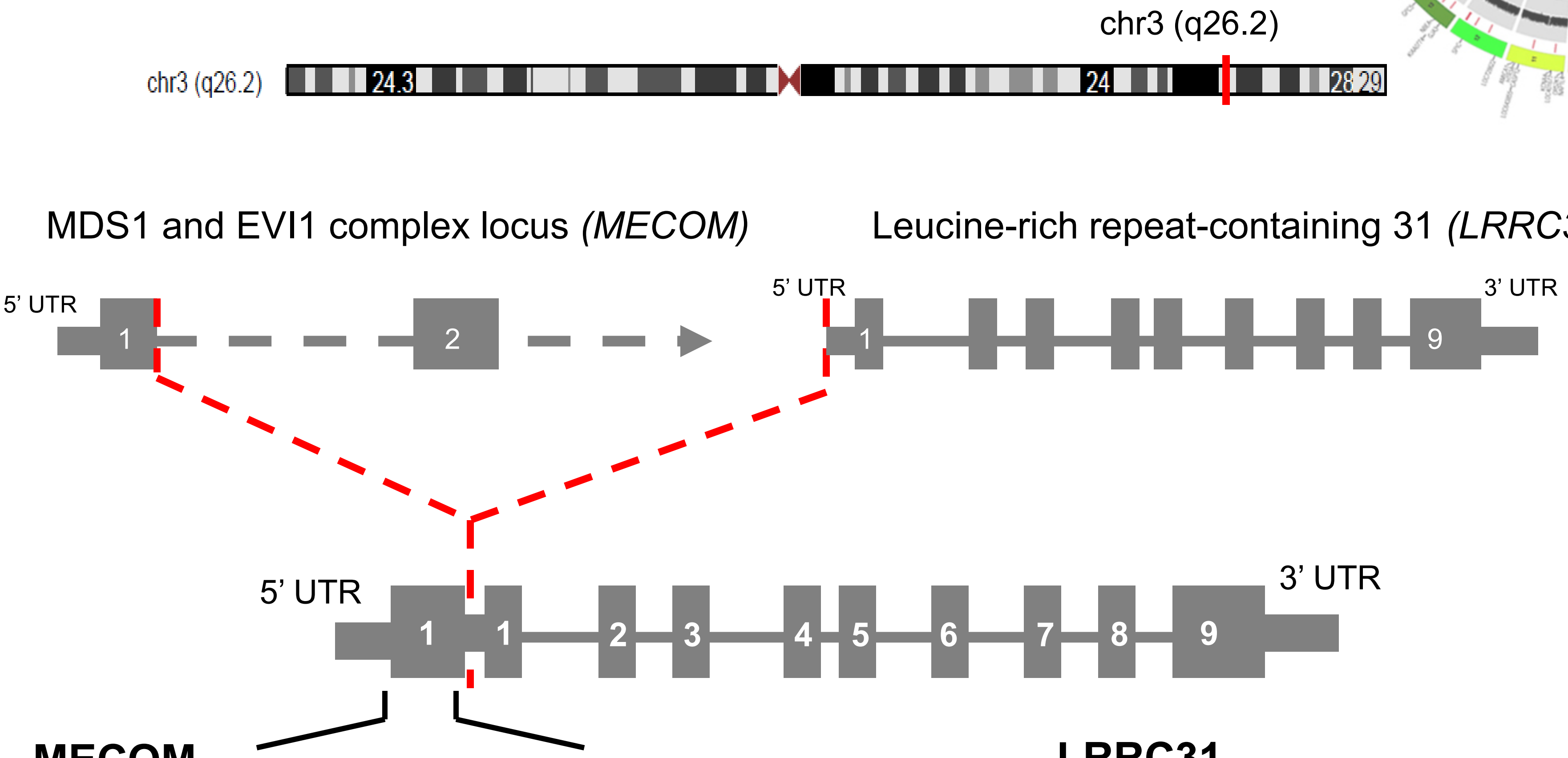
The Cancer Genome Atlas

In-frame *MECOM* fusion events

- *MDS1* and *EVI1* complex locus (*MECOM*) was focally amplified in >20% OV tumours (TCGA Res Network 2011)
- MECOM is a target of the therapeutic compounds aurintricarboxilic acid, arsenic trioxide
- We identify *MECOM* in-frame fusions with several different gene partners in at least 14 (3%) OV cases

Fusion	Recurrence	Event type	Event breakpoints
MECOM-LRRC31	6	duplication	exon1,5’UTR (5) exon1,exon3 (1)
MECOM-LRRC34	4	duplication	exon1,exon2
MECOM-CLDN1	2	duplication	exon1,exon2
MECOM-LMAN2L	1	translocation	exon3,exon1
MECOM-SLC7A14	1	duplication	exon1,exon3

MECOM-LRRC31 recurrent, in-frame fusion



MDS1 and EVI1 complex locus *(MECOM)*

Leucine-rich repeat-containing 31 *(LRRC31)*

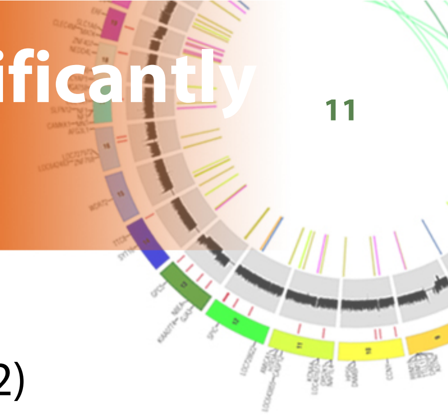
MECOM
Q03112

MRSKGRARKLAT

MAPK9, SMAD3 and
SUV39H1 interaction domain

LRRC31
Q6UY01

Known cancer-related pathways are significantly enriched with fusion genes



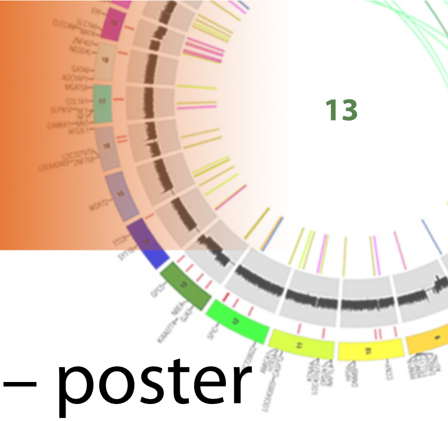
- 2,415 unique genes in 1,538 fusions
- 105 genes in COSMIC: causally implicated with cancer (p=1.8e-12)

Pathway database	Pathway	Adj. p-val (2415 genes)
KEGG; IPA	Pathways in cancer	3.0E-04; 5.1E-03
KEGG; IPA	Tight junction signaling	4.3E-02; 2.0E-02
KEGG; IPA	Cell cycle	4.4E-02; 3.3E-02
KEGG	WNT signaling	1.5E-02
KEGG	Ubiquitin-mediated proteolysis	2.6E-04
KEGG	ERBB signaling	2.4E-03
IPA	PI3K/AKT signaling	1.3E-02
IPA	TGF-b signaling	1.4E-02
IPA	Role of BRCA1 in DNA damage response	1.4E-02

Summary

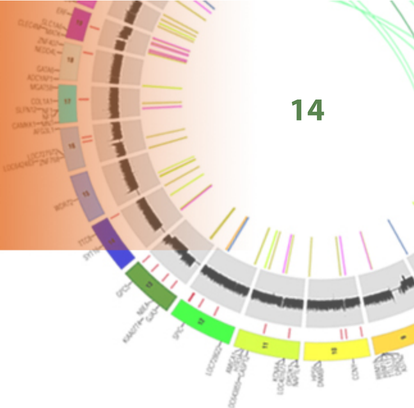
- Generated mRNA-seq and miRNA-seq for 420 and 485 of the TCGA high-grade serous ovarian adenocarcinoma cohort
- Unsupervised clustering of mRNA/miRNA expression profiles identifies additional sample groups
- An exploration of putative miRNA and mRNA interactions identifies significant expression anti-correlations including miR-29a with specific isoforms of *DNMT3A*
- In contrast to other cancers, such as AML, duplication is the primary rearrangement leading to gene fusions
- *MECOM* fusions are the most recurrent in-frame events

Future work



- Recurrent PTDs and ITDs (Lucas Swanson – poster #106 ‘Barnacle’)
- Rearrangements e.g. *MECOM*
- Differential expression and discriminatory gene analysis for unsupervised clusters and for gene rearrangements
- Further integrated analyses with our TCGA collaborators

Acknowledgements



Reanne Bowlby
Andy Chu
Hye-Jung Chun
Gordon Robertson
Karen Mungall
Readman Chiu
Kirstine Hamilton
Emilia Lim
Richard Corbett
Payal Sipahimalani
Dominik Stoll

Ka Ming Nip
Jenny Qian
Nina Thiessen
An He
Martin Krzywinski
Jacquie Schein
Richard Varhol
Angela Tam
Yongjun Zhao
Richard Moore
Inanc Birol

Poster #146

BCGSC Production Teams

Marco Marra
Steve Jones

Kevin White
Chai Bandlamudi
TCGA Research Network

BC CANCER FOUNDATION
partners in discovery

NATIONAL INSTITUTES OF HEALTH

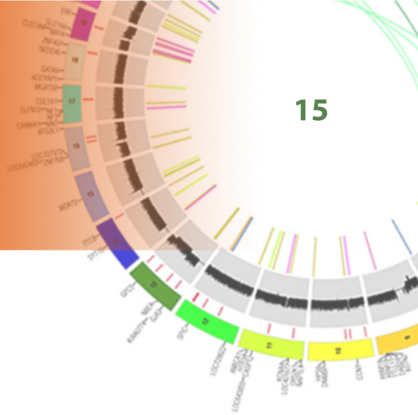
NATIONAL CANCER INSTITUTE

genome.gov
National Human Genome Research Institute
National Institutes of Health

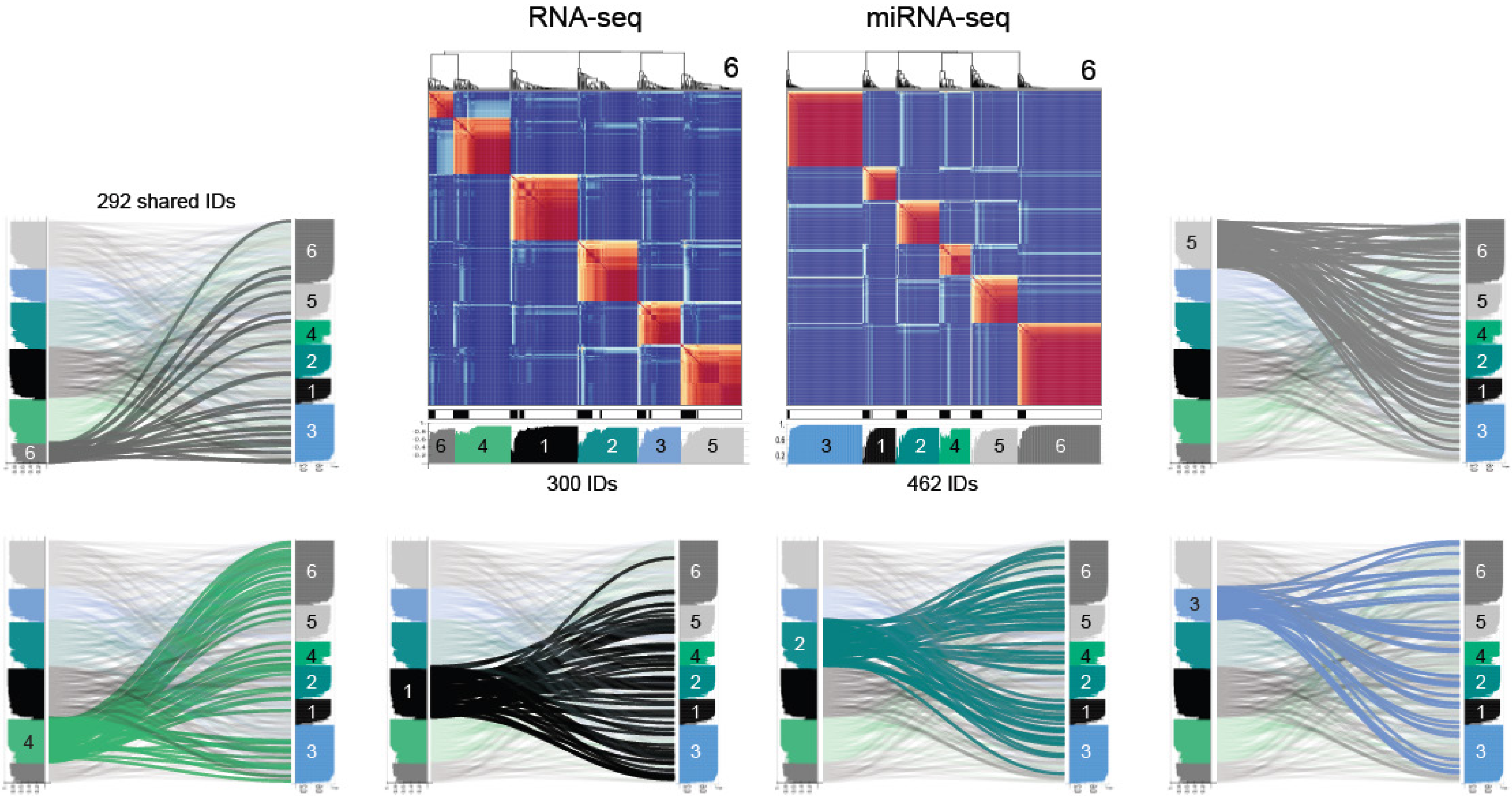
National Cancer Institute Award Number U24CA143866

BC Cancer Agency
CARE + RESEARCH
An agency of the Provincial Health Services Authority

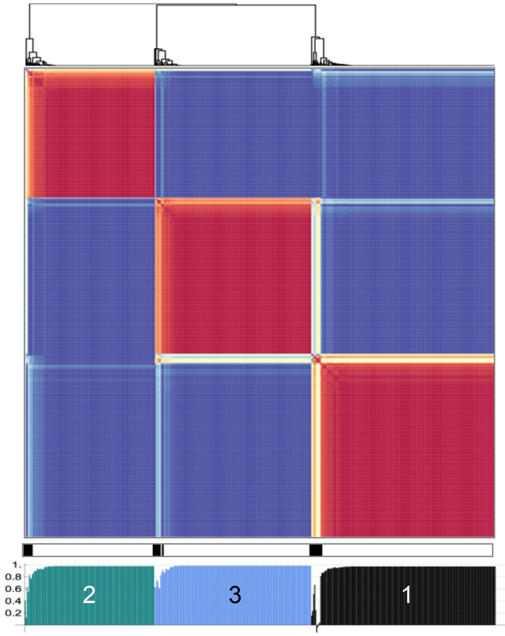
THE CANCER GENOME ATLAS



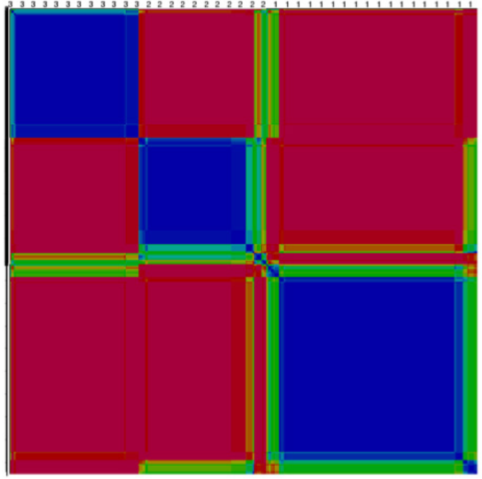
Comparison of RNA and miRNA cluster membership



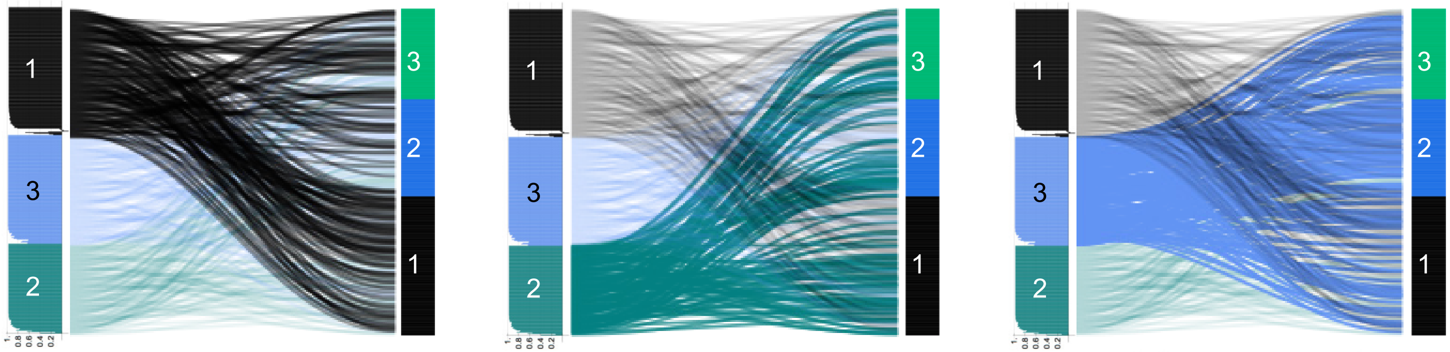
Compare miRNA-seq 3-groups to microarray 3-groups



30 JUNE 2011 | VOL 474 | NATURE | 609



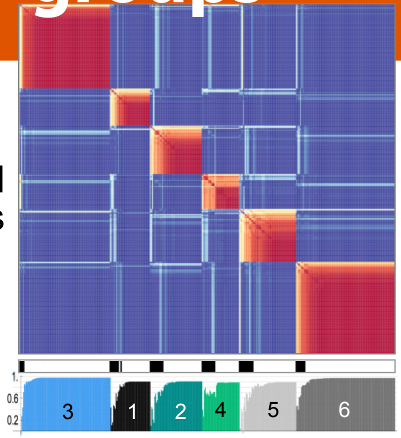
miRNA-seq 3 groups

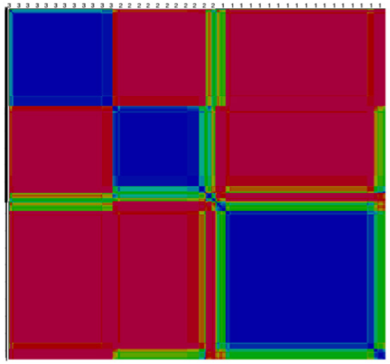


Micro-array 3 groups

Compare miRNA-seq 6-groups to microarray 3-groups

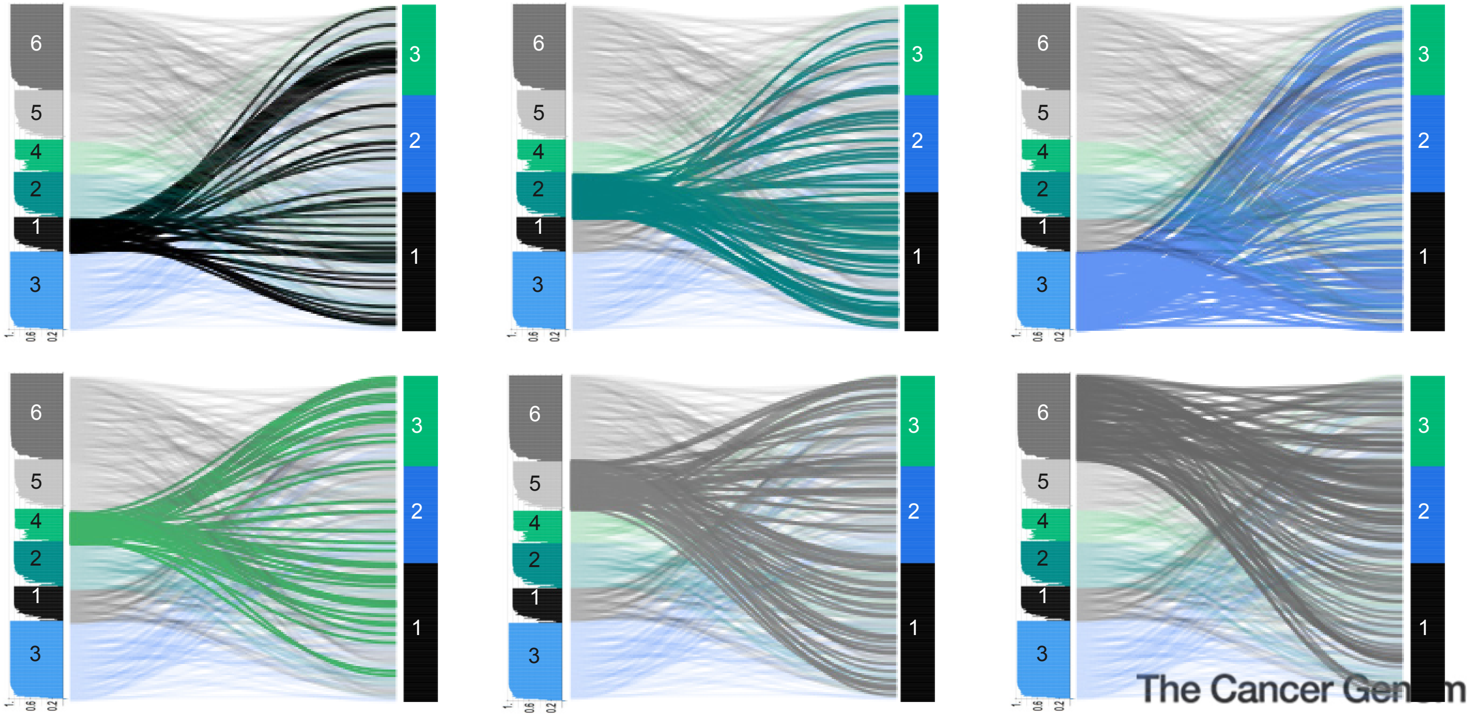
miRNA-seq
6 groups





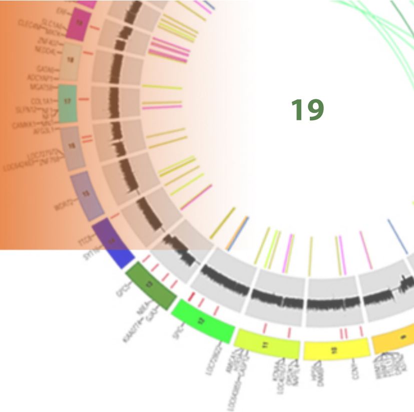
microarray
3 groups

30 JUNE 2011 | VOL 474 | NATURE | 609

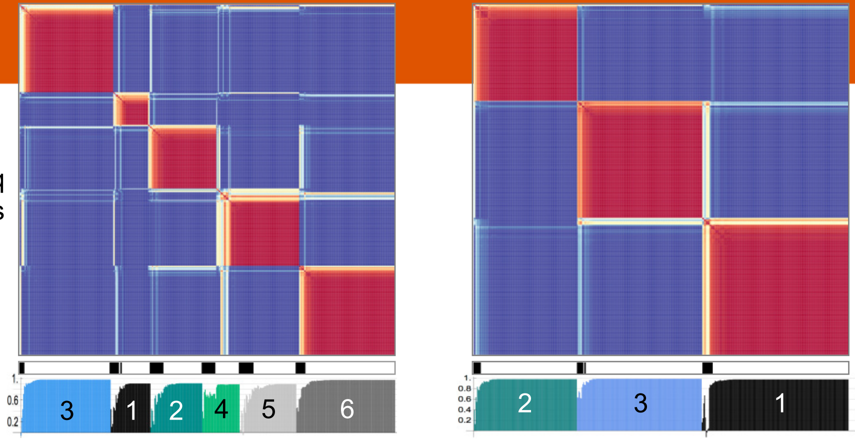


The Cancer Genome Atlas

Compare miRNA-seq 6 to 3-groups

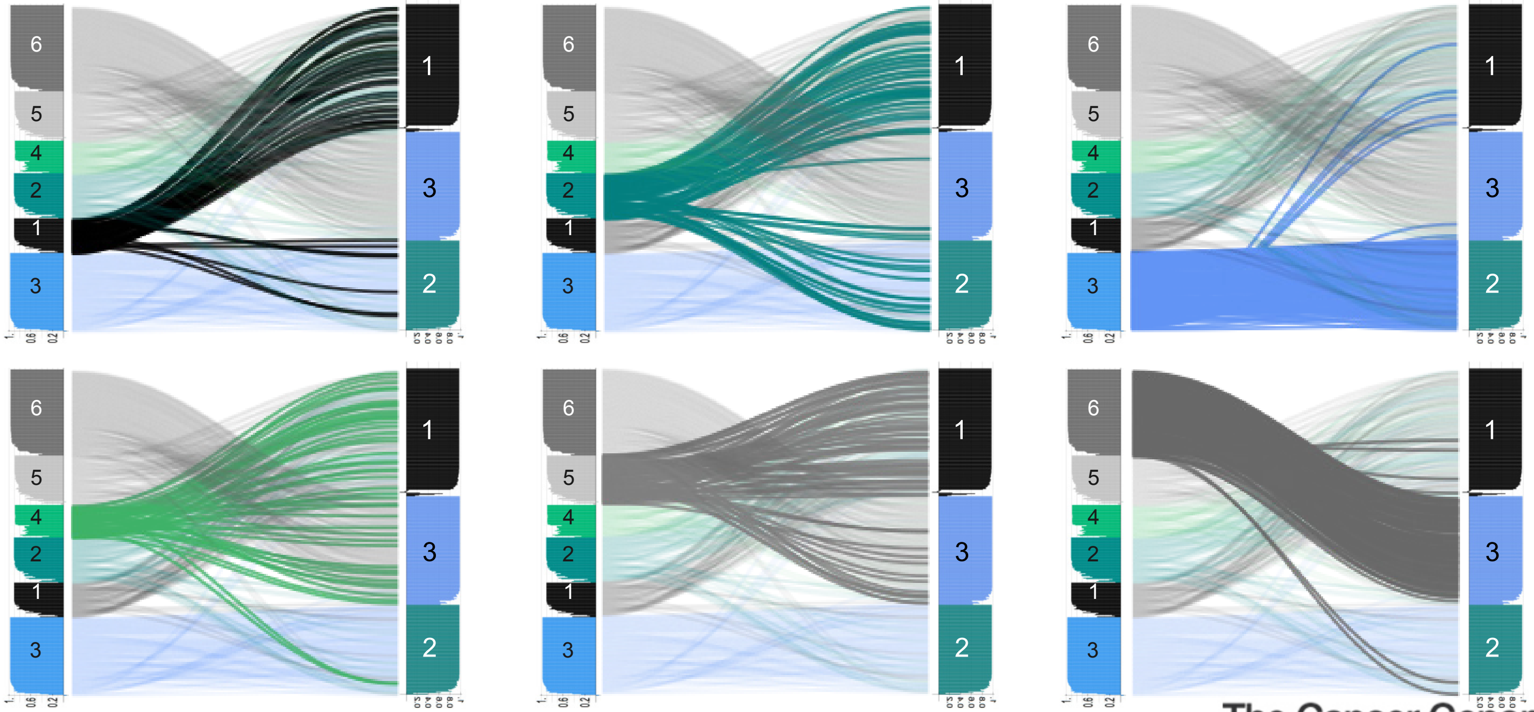


miRNA-seq
6 groups



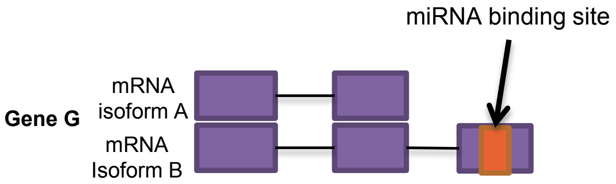
miRNA-seq
3 groups

462 samples



Putative miRNA:mRNA Isoform-specific Interactions

miRNA:Gene pairs where only mRNA transcript isoforms with miRNA binding sites have negatively correlated miRNA and mRNA expression profiles

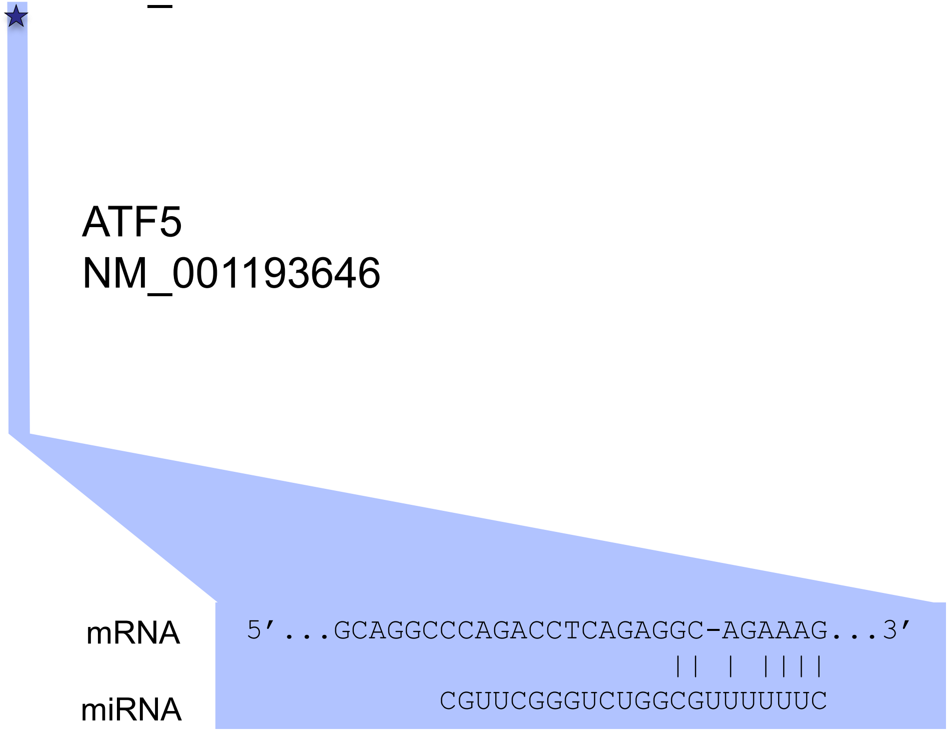


These interactions suggest an interplay between alternative isoform expression (AIE) & miRNA-mediated repression (MMR)

Top 19 miRNA:Gene Pairs That Display AIE-MMR Interplay

miRNA Name	MIMAT ID	Gene Name	With MBS (Max Q-val)	Without MBS (Max Q-val)	With MBS Mean Rho	Without MBS Mean Rho
hsa-mir-129	MIMAT0000242	ATF5	3.22E-05	0.077171081	-0.310180981	0.143693947
hsa-mir-509-3	MIMAT0004975	CHN1	1.31E-13	0.934810527	-0.452791643	0.01070408
hsa-let-7b	MIMAT0000063	EGFLAM	1.14E-05	0.999946714	-0.305326604	0.040647623
hsa-let-7i	MIMAT0000415	GABPB1	1.89E-05	0.865677253	-0.3064691	0.011480833
hsa-mir-186	MIMAT0000456	GABPB1	9.75E-06	0.705176441	-0.335491053	0.089241001
hsa-mir-877	MIMAT0004950	PPP2R2A	2.71E-05	0.000496779	-0.31261063	0.270502113
hsa-mir-509	MIMAT0002881	RCAN2	1.42E-06	0.963815897	-0.31998578	0.000418671
hsa-mir-331	MIMAT0000760	SEC14L1	1.30E-06	0.957171386	-0.33796075	0.132589096
hsa-mir-1180	MIMAT0005825	SET	7.09E-06	0.950153414	-0.301748309	0.016205006
hsa-mir-29b	MIMAT0000100	TIPRL	4.73E-06	0.486361605	-0.31272921	0.062702709
hsa-mir-9	MIMAT0000441	CHN1	9.70E-07	0.117736333	-0.348815743	0.193468941
hsa-mir-514	MIMAT0002883	CHN1	1.39E-12	0.981482288	-0.437334338	0.019167667
hsa-mir-509-3	MIMAT0004975	CHN1	1.31E-13	0.934810527	-0.452791643	0.01070408
hsa-let-7i	MIMAT0000415	GABPB1	1.89E-05	0.865677253	-0.3064691	0.011480833
hsa-mir-186	MIMAT0000456	GABPB1	9.75E-06	0.705176441	-0.335491053	0.089241001
hsa-mir-1228	MIMAT0005583	MCM7	4.74E-05	0.915797117	-0.30606887	0.014435698
hsa-mir-509	MIMAT0002881	RCAN2	1.42E-06	0.963815897	-0.31998578	0.000418671
hsa-mir-331	MIMAT0000760	SEC14L1	1.30E-06	0.957171386	-0.33796075	0.132589096
hsa-mir-29b	MIMAT0000100	TIPRL	4.73E-06	0.486361605	-0.31272921	0.062702709

Interplay Between mir-129 and ATF5 mRNA Isoforms



ATF5
NM_012068*

ATF5
NM_001193646

mRNA

miRNA
(MIMAT0000242)

```
5'...GCAGGCCCAGACCTCAGAGGC-AGAAAG...3'
                         ||  |  ||||
                 CGUUCGGGUCUGGCGUUUUUUC
```

★ mir-129 binding site

mRNA
Transcript
Expression

M)

Mature
miRNA
Expression

mRNA NMF Cluster

Only ATF5 mRNA isoforms harboring the miR-129 binding site have negatively correlated expression profiles with miR-129.

Expressed mutations and RNA-editing

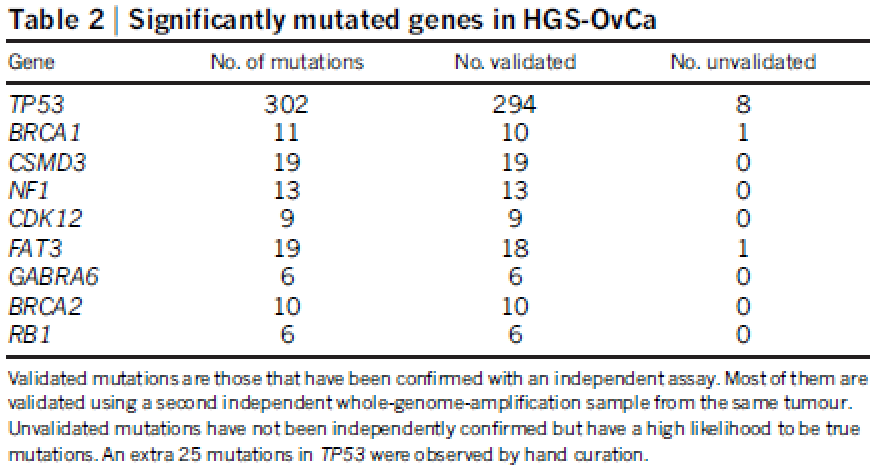


Table 2 | Significantly mutated genes in HGS-OvCa

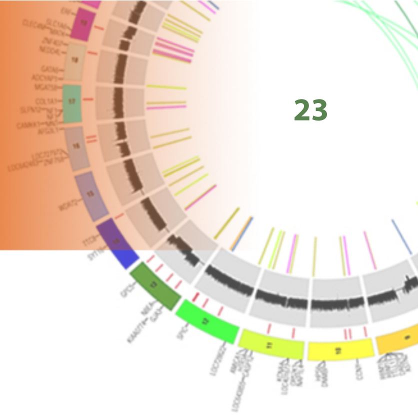
Gene	No. of mutations	No. validated	No. unvalidated
TP53	302	294	8
BRCA1	11	10	1
CSMD3	19	19	0
NF1	13	13	0
CDK12	9	9	0
FAT3	19	18	1
GABRA6	6	6	0
BRCA2	10	10	0
RB1	6	6	0

Validated mutations are those that have been confirmed with an independent assay. Most of them are validated using a second independent whole-genome-amplification sample from the same tumour. Unvalidated mutations have not been independently confirmed but have a high likelihood to be true mutations. An extra 25 mutations in *TP53* were observed by hand curation.

TCGA Research Network (2011) *Nature* **474**:609-615.

- Focus on gene mutations listed in marker paper: *TP53* (96%), *NF1*, *BRCA1* & *BRCA2*, *RB1*, *CDK12*
- Look for ITDs, PTDs and SNVs
- Can we find evidence for TP53 mutations in the 4% of patients (~16?) missing from the marker paper?

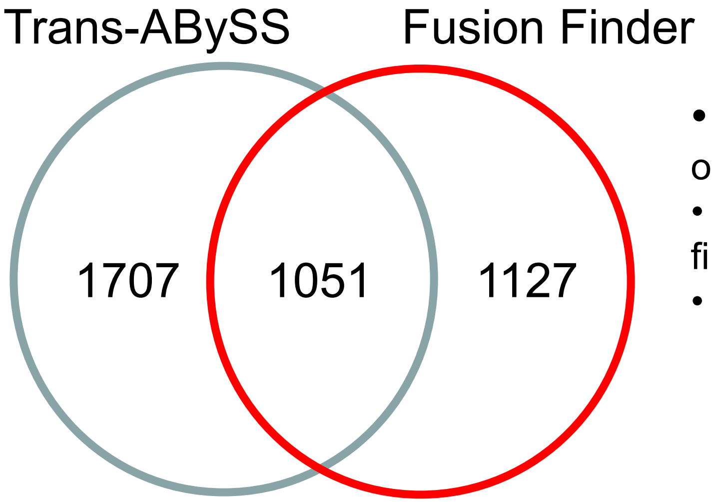
Orthogonal gene fusion detection



- 3007 events in total

Sense and anti-sense fusions reported.

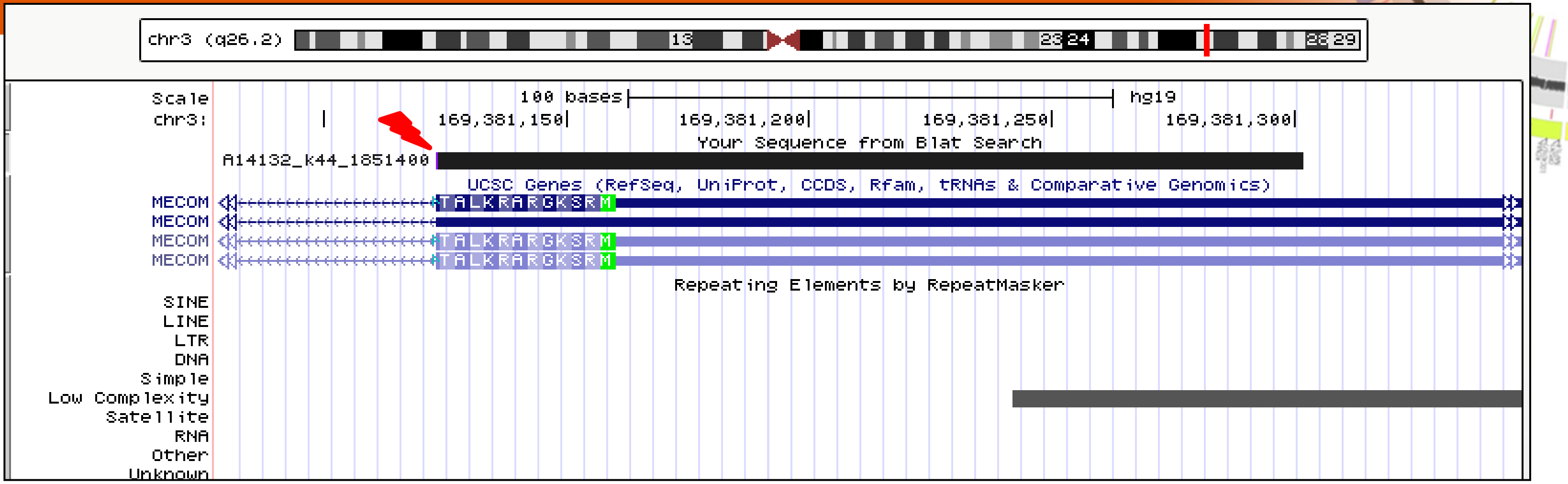
- 3007 events in total
  - Includes 2758 gene fusions
  - Includes 249 Large-scale rearrangements (one or no annotated genes)

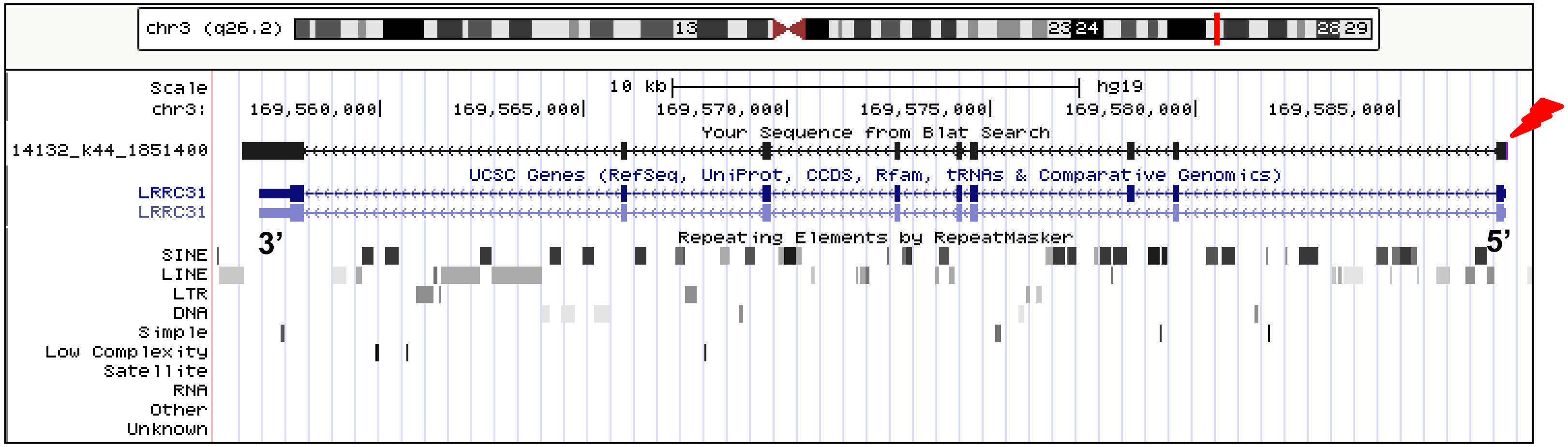


- All but 15 are in sense orientation.
- 2178 events in Chai's excel file
- Of the 1127:
  - 487 are short contigs (e.g. <150bp) or have insufficient read evidence for Trans-ABySS calls.
  - 360 have poor contig-to-genome alignments
  - 280 putative novel FF events

- We observe 60 recurrent (≥2 libraries), high-confidence events when overlapping Trans-ABySS and FusionFinder results
- These include:
  - *NCOR2-UBC* (11); *XPR1-ACBD6* (9); *GTF2I-GTF2IRD1* (8); *CCDC6-ANK3* (7); *GATAD2B-CRTC2* (6); *TFG-GPR128* (6); *COL14A1-DEPDC6* (5); ***MECOM-LRRC31*** (5)

In-frame *MECOM-LRRC31* fusion found in 5/420 libraries



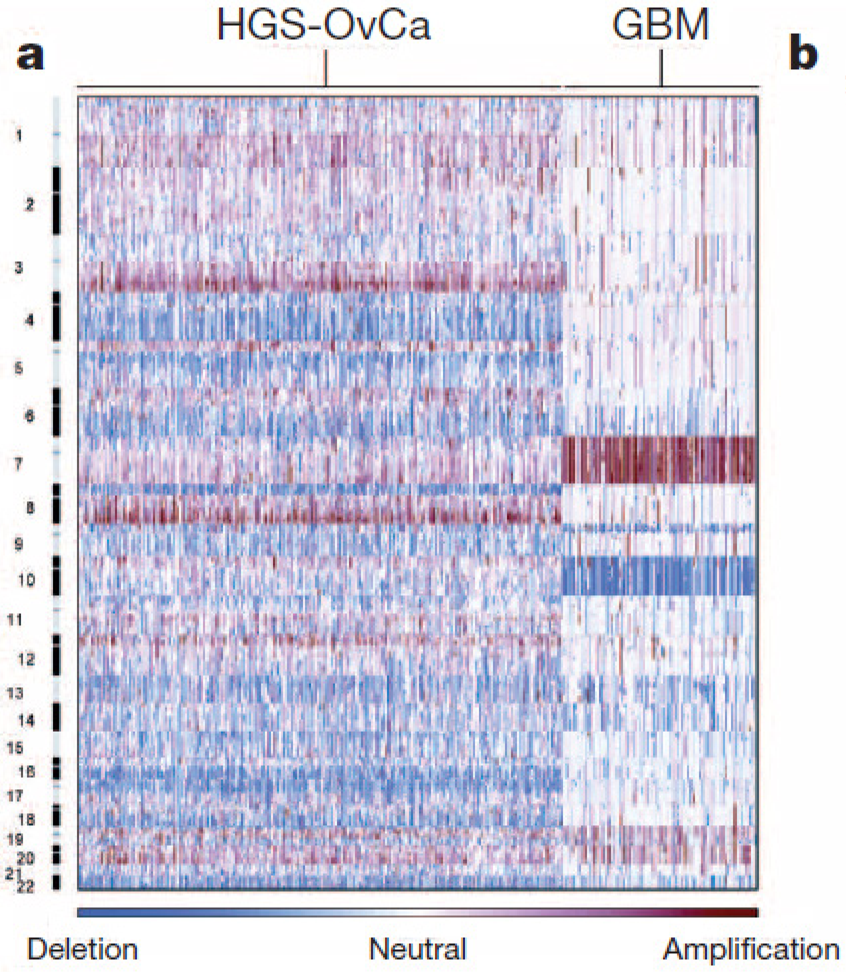


A14132_k44_1851400 180-3057bp 100.0% chr3:169556616-169587660 (-)
A14132_k44_1851400 1-179bp 100.0% chr3:169381124-169381302(-)

Summary of all MECOM events

Fusion	Recurrence	In frame	Event type	
MECOM_LRRC31	6	yes	duplication	exon1,5utr (5) (A08215,A12152,A14126,A14132,A14134) exon1,exon3 (1) A14327
MECOM_LRRC34	4	yes	duplication	exon1,exon2 (3) (A14376,A14193,A14280) exon1,exon3 (1) (A14180)
MECOM_CLDN1	2	yes	duplication	exon1,exon2 (A08245,A08257)
MECOM_LMAN2L	1	yes	translocation	exon3,exon1 (A08240)
MECOM_SLC7A14	1	yes	duplication	exon1,exon3 (A14279)
MECOM_NA	1		deletion	exon1,NA (A12038)
MECOM_NA	1		duplication	exon1,NA (A12132)
MECOM_NA	1		duplication	exon1,NA (A12095)
MECOM_SKIL (AS)	2		duplication	exon1,intron3 (A12132) intron2,exon3 (A14308)
MECOM_NA	2		inversion	exon1,NA (A08245,A08257)
SEC62_MECOM	1		inversion	exon7,intron2 (A14184)
MECOM_NA	1		translocation	exon1,NA (A12110)

“HGS-OvCa demonstrates a remarkable degree of genomic disarray”



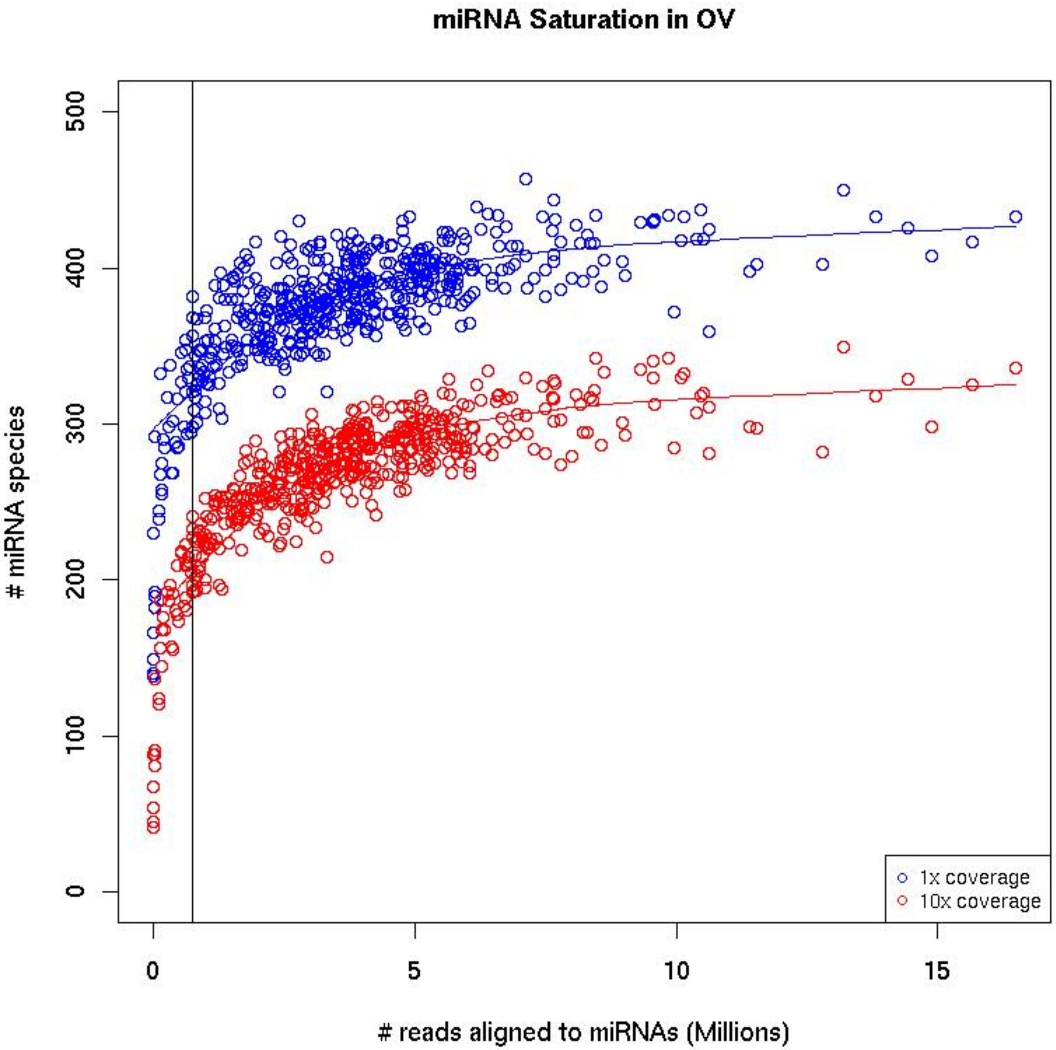
- Point 1
- Point 2

TCGA Research Network (2011) Nature 474:609-615.

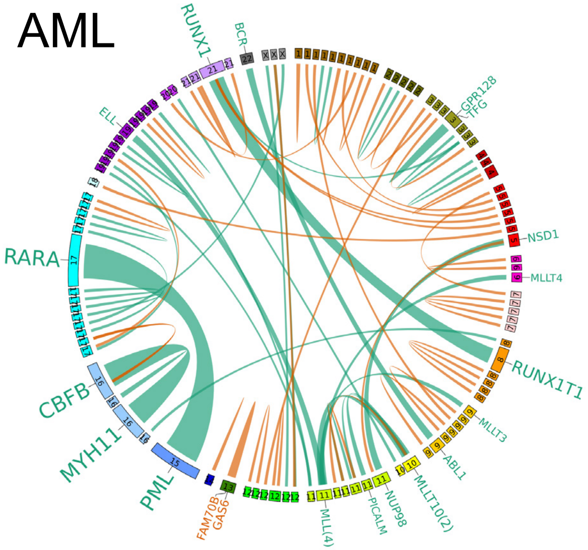
Partial and tandem duplications

- Table of most recurrent ITDs/PTDs to come from Karen
- E.g. MSLN PTD in 26/420 libraries
- ARID1A ~20/420

miRNA saturation in ovarian cancer



AML


RUNX1
BCR
GPR128
TFG
NSD1
MLLT4
RUNX1T1
MLLT3
ABL1
MLLT10(2)
NUP98
PICALM
MLL(4)
GAS6
FAM70B
PML
MYH11
CBFB
RARA
ELL