

Transcriptomic analysis of the organoids

Lena Morrill

March 16, 2021

Contents

1	Data	1
1.1	3' bias	1
1.2	Summary of organoids	1
2	Transcriptomics of organoids	3
2.1	Organoids <i>vs</i> normal ovarian tissue DE	3
2.2	Sensitive organoids <i>vs</i> resistant organoids DE	3
2.3	Comparison with TCGA	4
2.4	(Resistant <i>vs</i> sensitive in organoids) <i>vs</i> (complete remission or response <i>vs</i> progressive disease in TCGA)	5
3	Correlation of gene expression and copy number	5

1 Data

1.1 3' bias

Three RNASeq samples – PDO14, PDO16, and PDO18 – have a clear 3' bias. Three others – PDO4, PDO13, PDO17 – have a slight bias. Only the first three have been removed from the analysis.

1.2 Summary of organoids

PDO	ID	sWGS	RNaseq	OVO4	Cor. CN-GE	Com-ments	scDNA?	S4	WGD	CCNE1 amp	CCNE1 amp (RNA or CN)	Tan-dem dup	#seg-ments	Ploidy	Quiet	Change from CDK12 amp	CDK12 loss	Chemo-naive	
1	PDO1	151723	19844	19938	920.00	Y				✓	✓			Very low	Very high	?	×	×	
2	PDO11	54327	118976	19905	413.00	N	Very simi-lar to PDO11			✓				Very low	Med	×	×	×	
3	PDO10	119178	119127	19907	788.00	N	Very simi-lar to PDO1			✓				Med	Med	?	×	×	
4	PDO13	54276	19846	19940	297.00	Y				✓				Med	Med	?	✓	×	
5	PDO15	118947	54327	19904	333.00	N				✓				Med	Med	×	×	×	
6	PDO17	50495	50495-	19916	348.00	Y				✓				Med	Med	?	×	×	
7	PDO18	32070	32070-	19920	47.00	-				✓				Low	Med	?	×	×	
8	PDO2	23868	jB95a-	19917	75.00	Y				✓	✓			Med	Med	?	×	×	
9	PDO3	118976	119178	19906	466.00	N				✓	✓			Med	Med	×	×	×	
10	PDO9	119058	119058-	19921	466.00	Y				✓			(!)	surely also has	Med	Med	?	×	×
11	PDO14	54288	19829	19925	409.00	-				✓				High	Med	×	✓	×	
12	PDO4	151773	19848	19942	571.00	Y				✓				Very high	Med	?	×	×	
13	PDO5	119127	118947	19903	627.00	N				✓				Med	Med	?	×	×	
14	PDO6	119148	119148	19902	627.00	Y				✓				Med	Med	?	×	×	
15	PDO16	119025	19842	19936	839.00	-				✓				Low	Low	✓	✓	×	
16	PDO12	151761	19845	19939	467.00	Y				✓				Low	Low	✓	✓	×	
17	PDO7	32077	19847	19941	366.00	Y				✓				Low	Low	✓	?	×	
18	PDO8	54059	19843	19937	366.00	Y				✓				Low	Low	✓	?	×	

2 Transcriptomics of organoids

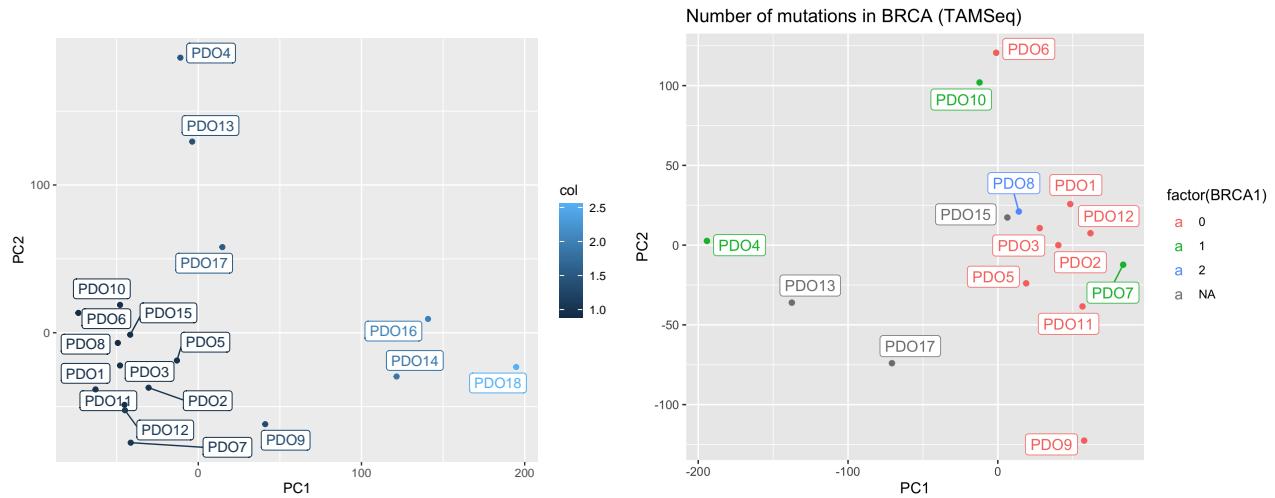


Figure 1: PCA including all samples (left) and only the ones without 3' bias (right; colour coded by number of mutations in BRCA.)

2.1 Organoids *vs* normal ovarian tissue DE

Done by Stephane

2.2 Sensitive organoids *vs* resistant organoids DE

Done by Stephane

2.3 Comparison with TCGA

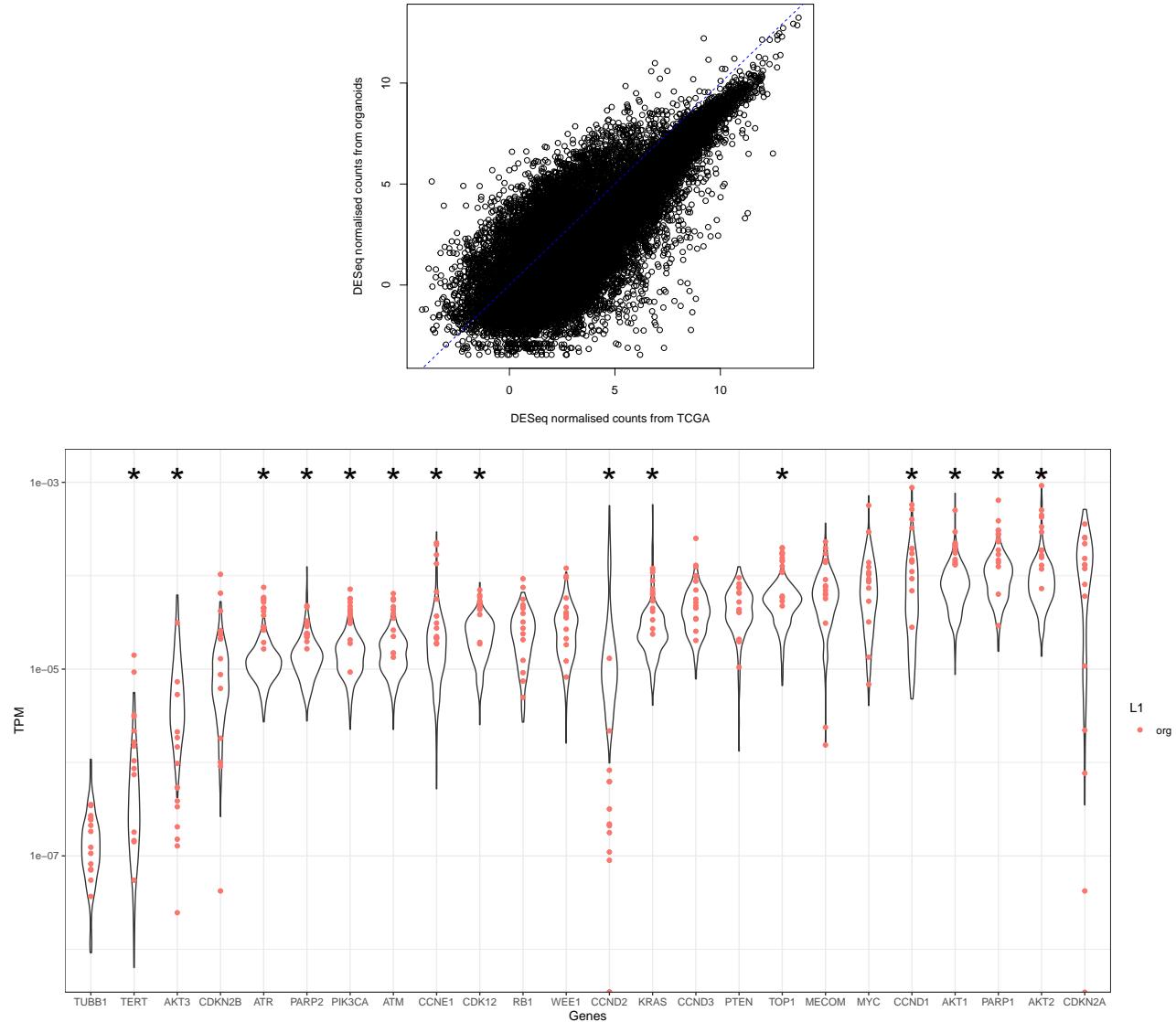


Figure 2: TPM for genes of interest in organoids (points) and in TCGA (violin plots). Genes which have a different distribution according to a Kolmogorov-Smirnov test are marked with an asterisk.

2.4 (Resistant *vs* sensitive in organoids) *vs* (complete remission or response *vs* progressive disease in TCGA)

Extremely bad correlation for both log2 fold changes and p values. As we don't expect to find a perfect match anyway (as (1) progressive disease would work as a surrogate for resistant, but there are many more factors and (2) it's an organoid comparison *vs* a primary tissue comparison) we have left this on standby.

3 Correlation of gene expression and copy number

There are two values for gene expression that we can work with:

- (Computed by Lena) the weighted average copy number value per gene
- (Computed by Stephane) the copy number value is assigned to each gene which is the closest to a copy number area of high variability (variability computed with the variance across samples)

In the genes of interest (MYC, etc.) the two are almost always the same (meaning that the relevant copy number segment is on the gene). The conclusions from Stephane's analysis is that there were some correlations in certain genes, but that it was mostly driven by only a couple of organoids (!). There were some extreme copy number values that were excluded as outliers – they all belong to the samples 118976org or 119058org, that come from the same patient.

There is no obvious correlation between raw CN values and counts (TPM or DESeq2), either pooling all organoids or doing an organoid-specific scatterplot (See Figures [scatterplot per organoid DESeq2] and [scatterplot per organoid TPM]).

This prompted us to scale the CN and GE values (scaled between organoids to a mean of zero and a standard deviation of 1). This gives a much better correlation, especially when segments of copy number of two are excluded from the analysis (see 3). However, it seems as though this trend only applies to some organoids. Organoids PDO11, PDO10, PDO15, PDO3, PDO5 show no trend between GE and CN. These organoids don't have anything special when it comes to exposures, and they are not the ones with relatively high 3' bias.

A similar approach is to compare the ranks of the organoids in CN and GE (see Figure 4).

A couple of important considerations: gene expression values are correlated (only slightly, but this can become important in whole genome analyses), and copy number regions can be extremely large, making the link between copy number and any gene very tenuous.

If we ask the question of whether genes which are in highly amplified regions have a higher gene expression than the same gene in normal fallopian tissue (average over five fallopian samples of healthy patients) the answer is overwhelmingly *not* (102174 sample/genes combinations for which it's not the case, 64621 for which it is, when the copy number is higher than 2), with most genes having actually lower expression values than normal tissue. This is confounded by compositionality/ assuming that most genes remain the same. This includes most genes of interest.

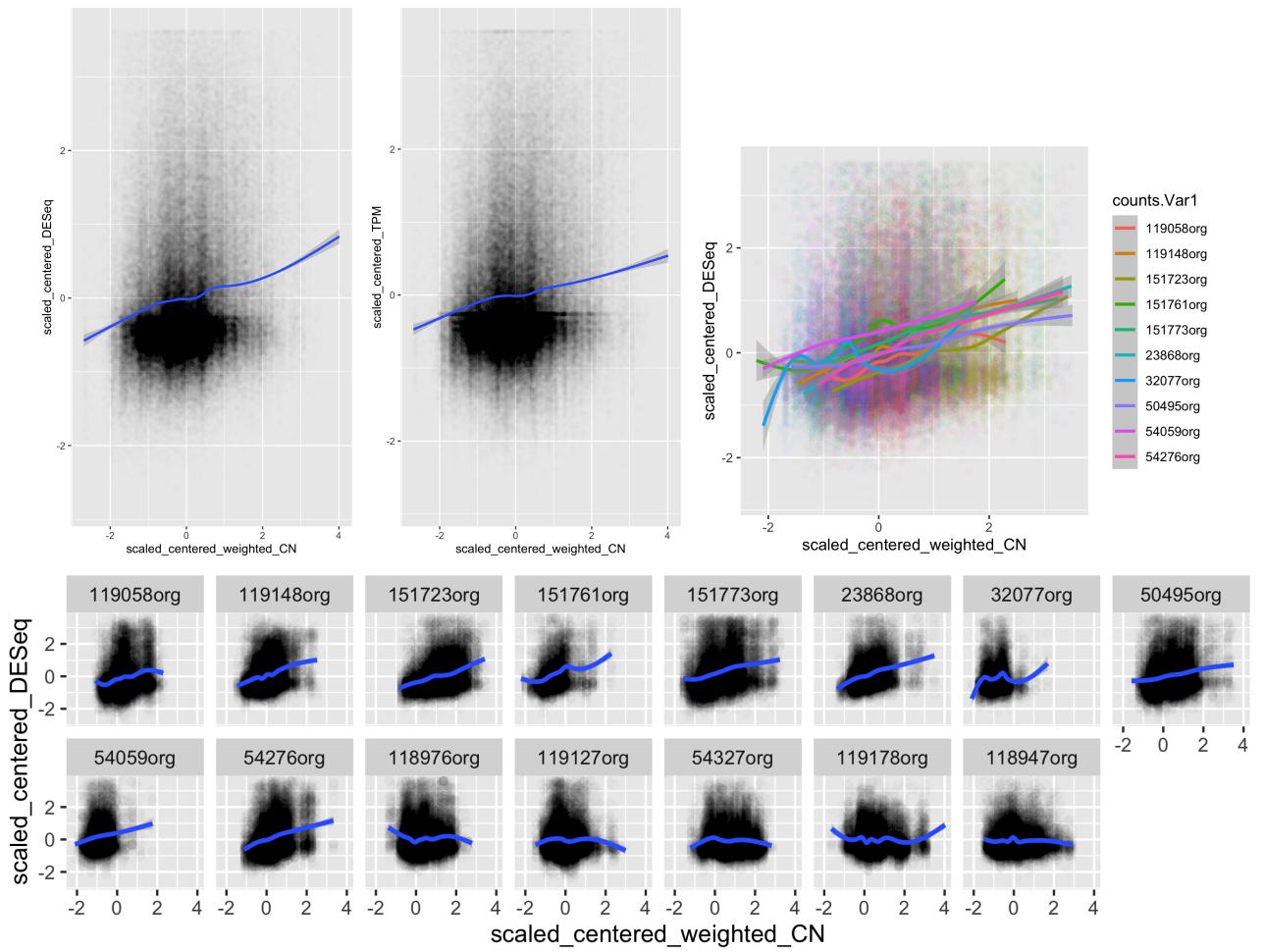


Figure 3: Scaled and centered CN and GE (DESeq2 and TPM respectively). Below: split per organoid. The same results are found when normal segments are kept.

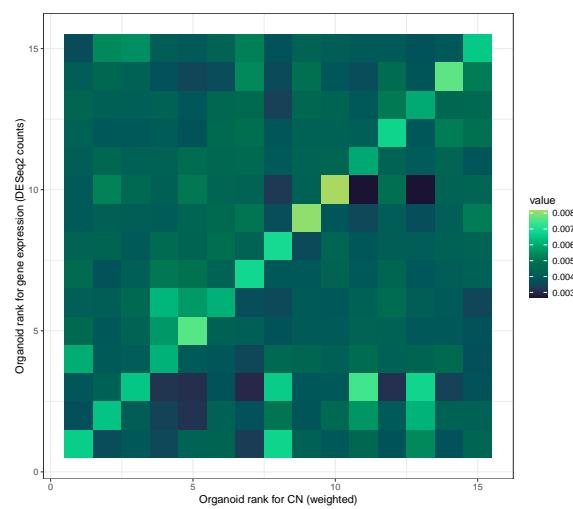


Figure 4: Relative abundances of agreement of ranks between GE and CN, for all genes. A brighter diagonal indicates a very good agreement in the rank of GE and rank of CN across organoids.

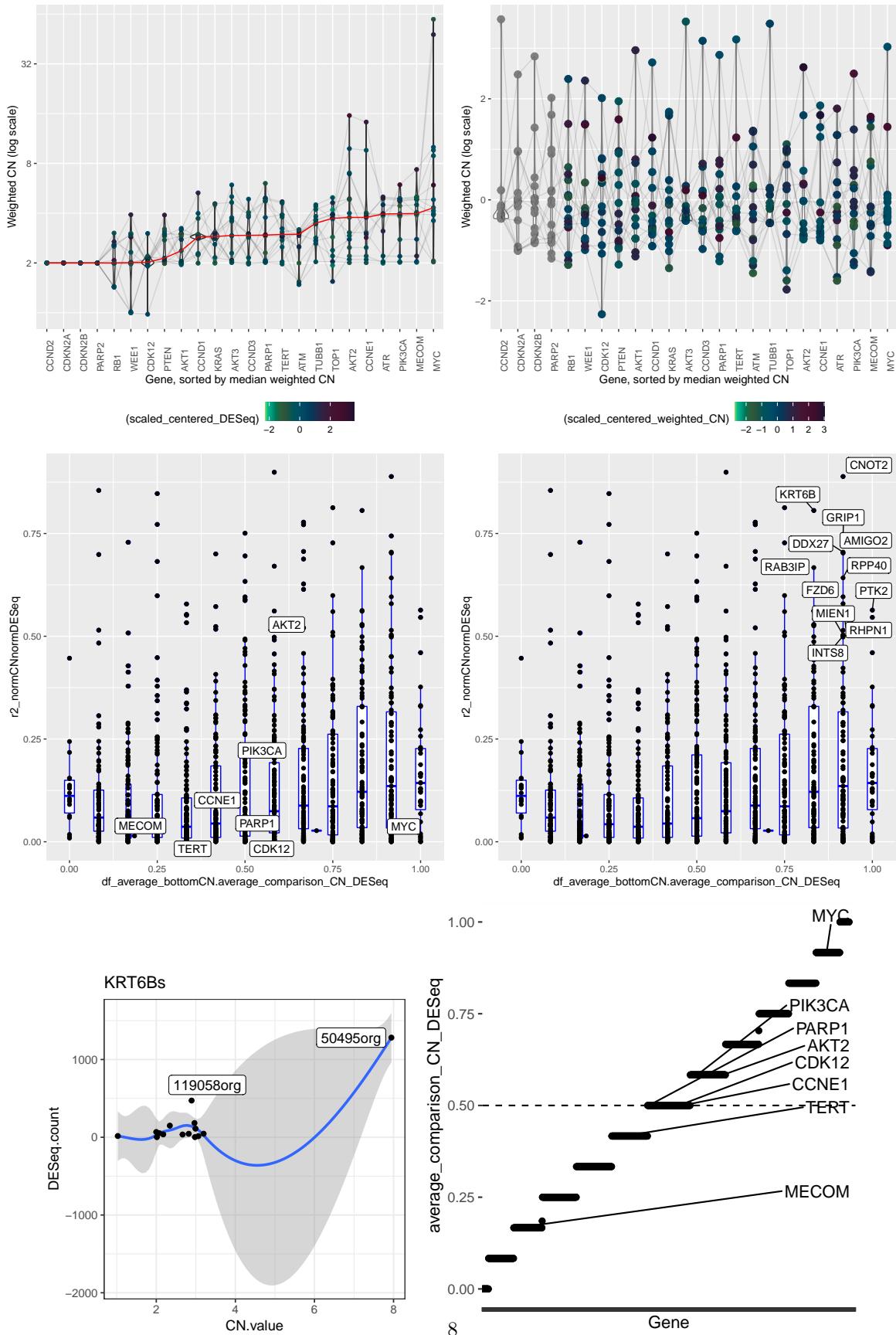


Figure 5: Fraction of samples with gene expression higher than the bottom three samples of lowest copy number. Higher values indicate a correlation between copy number and gene expression.