

# Report 8: Contact and expression

*Cyril Matthey-Doret*

*9 novembre 2016*

## **Introduction:**

Here, I investigate the relationship between gene expression and DNA-DNA contact across 4 different cell lines (GM12878, K562, HUVEC and NHEK). This analysis is performed on elincRNAs, nelincRNAs, epcRNAs and nepcRNAs. Contact per gene was calculated in 2 ways: Gene versus chromosome computes all interactions between the gene and the chromosome, while gene versus TAD only takes interactions between the gene and the TAD it belongs to.

NOTE 1: At the moment I only have expression data for GM12878, but I will add the other cell lines later.

NOTE 2: Contact matrices for chromosome 9 for all cell lines have been normalised using SQRTVC instead of KR because the algorithm did not converge for chromosome 9 in K562 and vector was full of NAs.

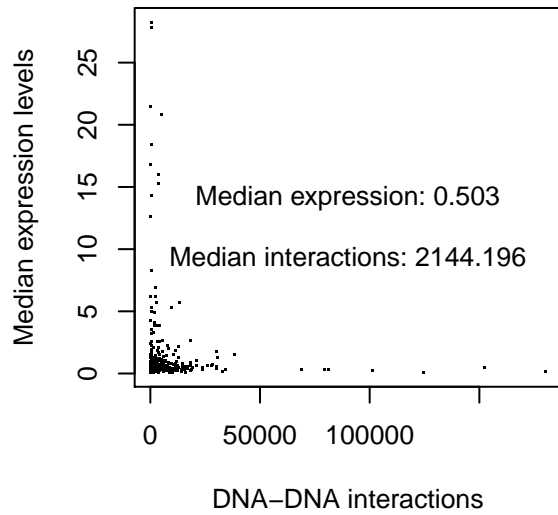
## **Results:**

### **Overview of data:**

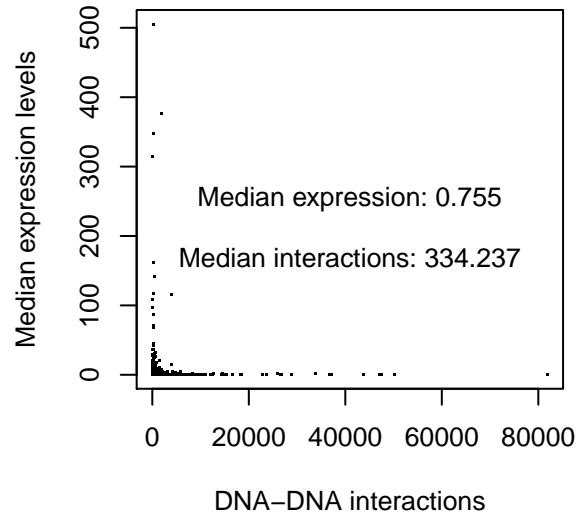
A quick overview of the data shows that enhancer-associated genes have a lower expression and higher amounts of DNA contacts on average. Patterns in the data would be easier to detect after log transforming both variables.

Gene versus chromosome:

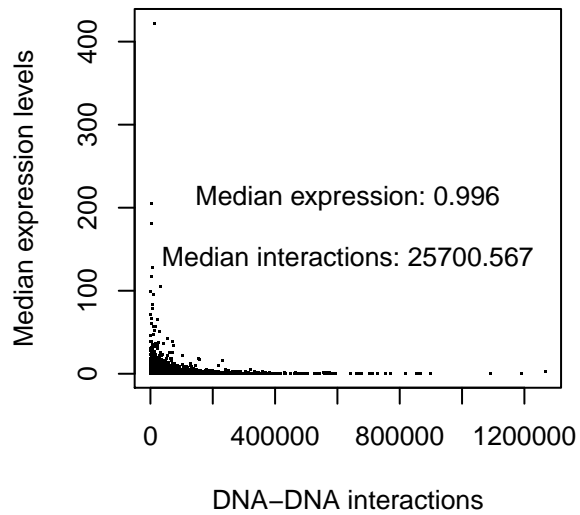
**GM12878: elincRNA**



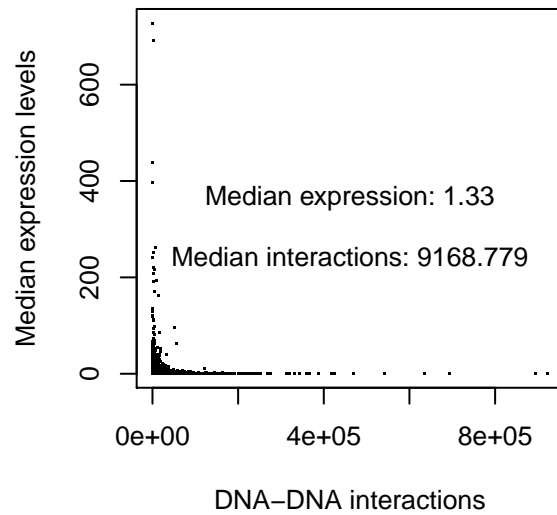
**GM12878: nelincRNA**



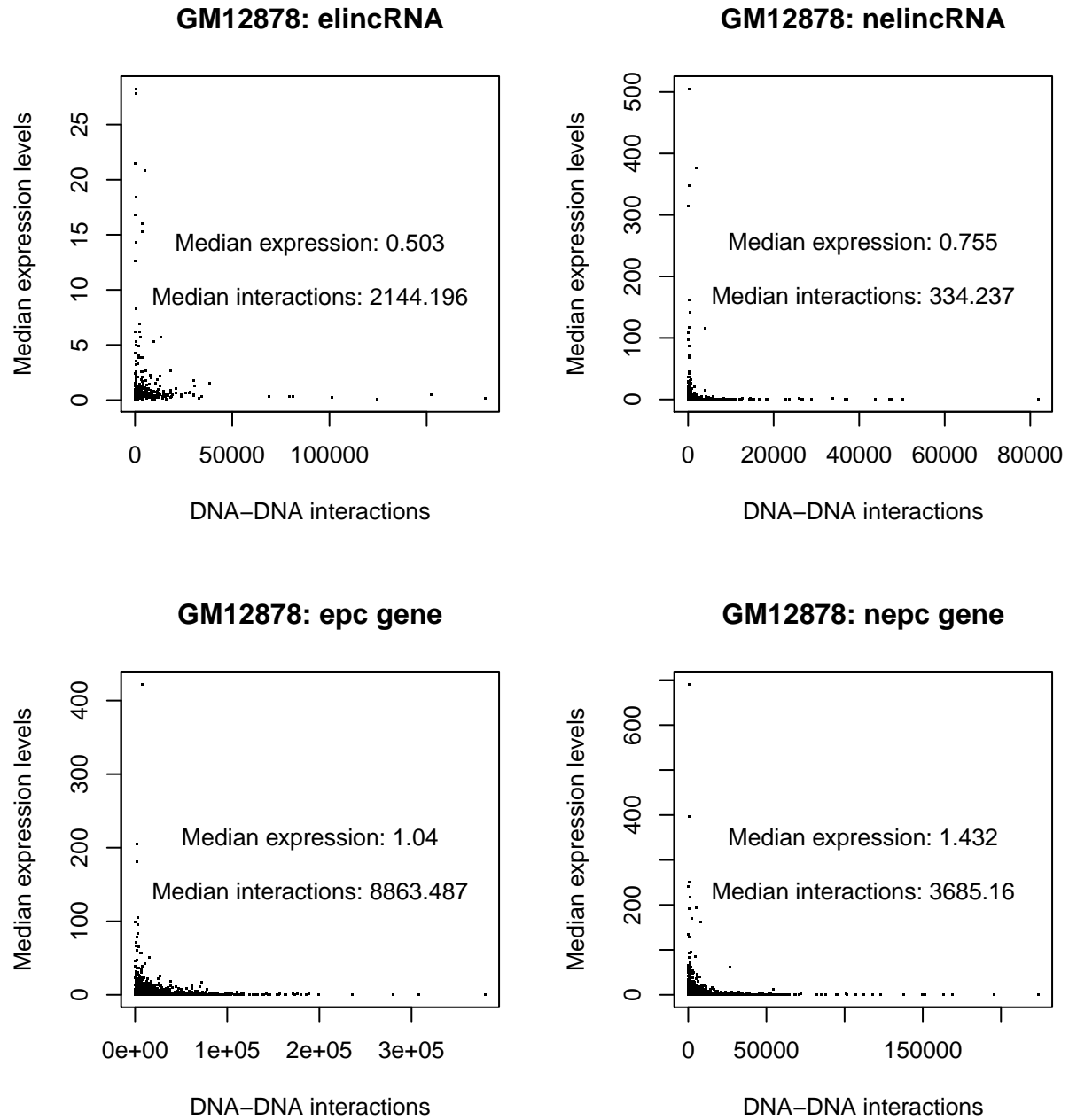
**GM12878: epc gene**



**GM12878: nepc gene**



## Gene versus TAD:

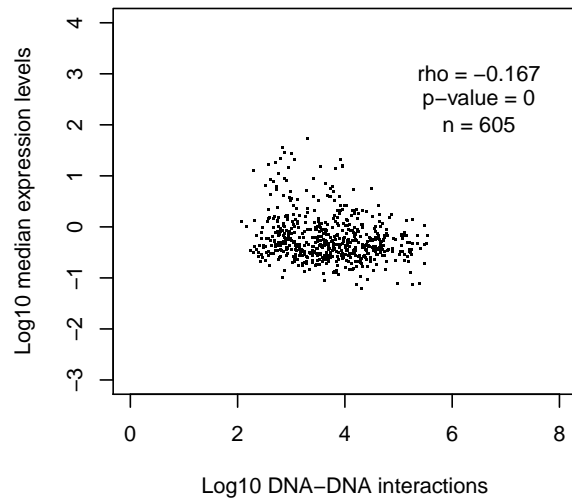


## Correlations:

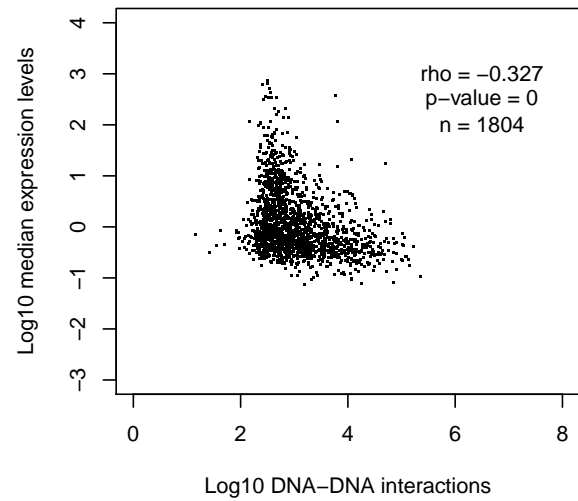
Spearman correlation was used to detect non-linear relationship between DNA contact and expression. The data was log-transformed for visualization.

Gene versus chromosome:

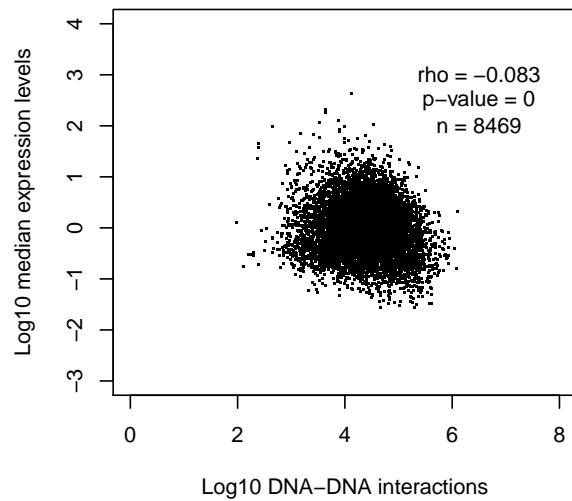
**GM12878: elincRNA**



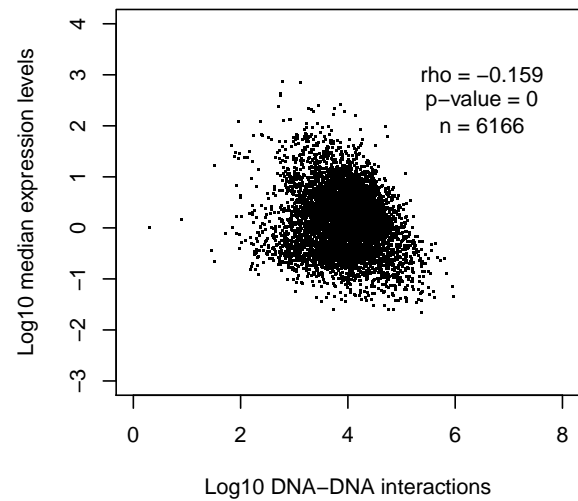
**GM12878: nelincRNA**



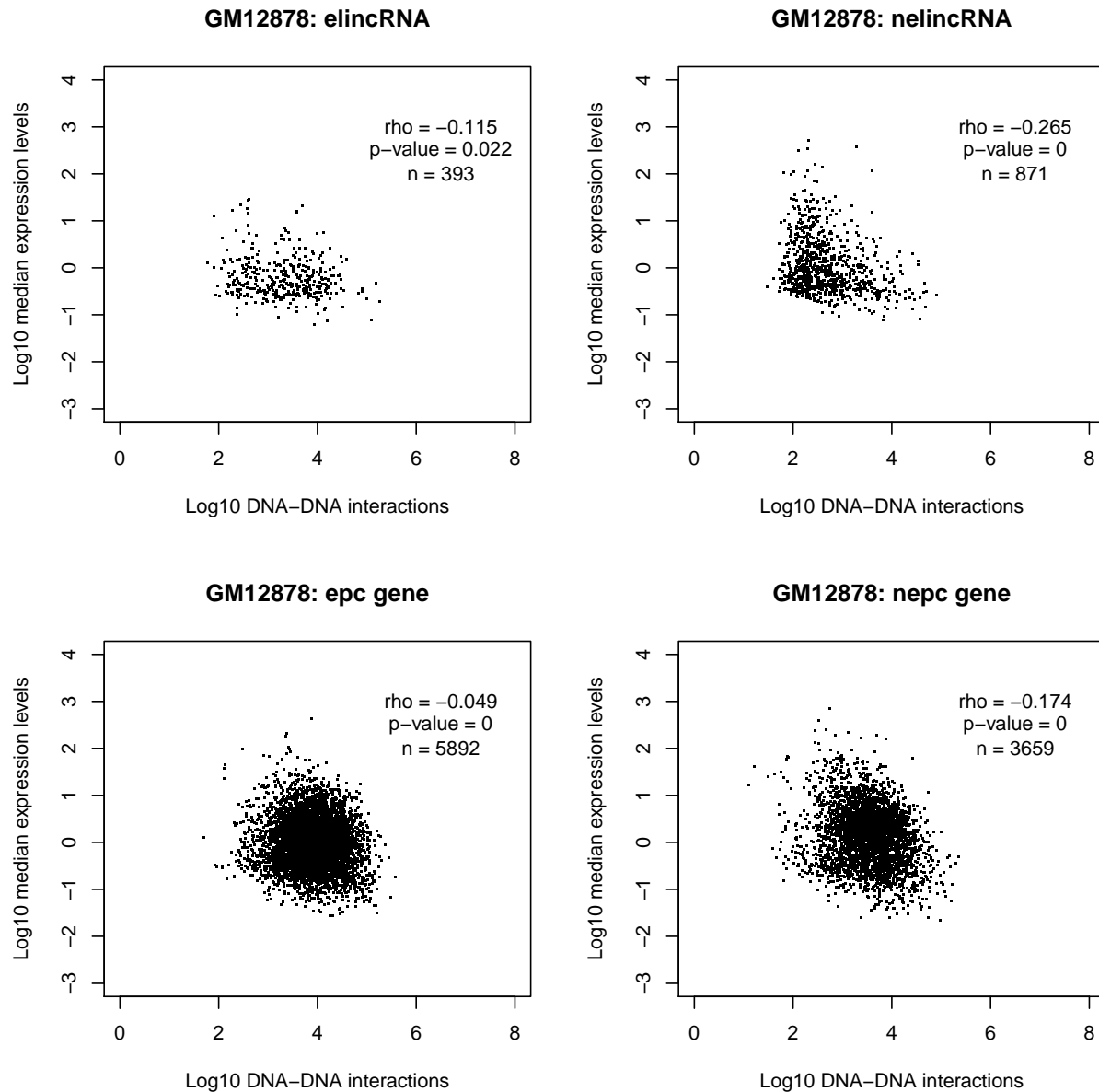
**GM12878: epc gene**



**GM12878: nepc gene**



## Gene versus TAD:



## Conclusions:

In general, there is a weak, non-linear relationship between expression and DNA-DNA contact. Genes with a lot of contact have very small expression levels, while those with little contact can have both low and high expression levels.

The correlation is weaker in elincRNAs than in nelincRNAs. This may simply be caused by the overall lower expression of enhancer-bound genes. This low expression may be due to bidirectionally transcribed RNAs that are lowly expressed. It would thus be interesting to differentiate between unidirectional and bidirectional transcription.

The relationship observed between contact and expression can in fact not really be called a correlation, but there is still an interesting difference between the patterns observed in lincRNAs and protein-coding genes. Note the graphs has been rescaled, it appears that elinc RNAs are a subset of nelincRNAs and that these 2 groups do not present significant differences.