Reference: Williams K, Christensen J, Pedersen M, Johansen JV, Cloos PAC, Rappsilber J et al. Tet1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 2011, 473(7347):343-348.

Summary of the article: Methylation of cytosines by DNA methyl transferases is an important factor contributing to the downregulation of gene expression. This epigenetic modification mostly occurs in CpG islands, expect in the ones of promotors. Tet1 katalyzes the hydoroxylation of methylated cytosines, which results in demethylation. Previous studies show that methylated cytosines frequently occur in embryonic stem cells and that this modification is a crucial factor in gene regulation.

In this paper, mouse embryonic stem cells were used in order to unravel the function of Tet1 since these cells exhibit a high Tet1 expression level. By using ChIP-seq, they found out that Tet1 preferentially binds to transcription start sites and gene bodies of CpG-rich promotors. Microarray was performed on both normal and Tet1-knockdown embryonic stem cells and showed that Tet1 represses transcription. Next, they investigated the molecular pathway of this repression by using ChIP-seq and found out that Tet1 recrutes SIN3A, thereby inhibiting transcription.

ChIP-seq analysis (GSE24841)

Location of the original ipynb: /mnt/storage/r0368030/jupyternotebooks

Files: /mnt/storage/r0368030/groepswerk

In [13]:

```
# Before starting, we created a working directory for our group work.
mkdir ~/data/groepswerk
cd ~/data/groepswerk
pwd
```

```
/home/luna.kuleuven.be/r0368030/data/groepswerk
```

In [14]:

```
# First, the SRA files were downloaded from the GEO database (GSE24841). We only downloaded
fastq-dump --split-files SRR070939
fastq-dump --split-files SRR070938
fastq-dump --split-files SRR070937
fastq-dump --split-files SRR070936
fastq-dump --split-files SRR070935
fastq-dump --split-files SRR070934
fastq-dump --split-files SRR070933
fastq-dump --split-files SRR070932
fastq-dump --split-files SRR070931
fastq-dump --split-files SRR070930
fastq-dump --split-files SRR070928
fastq-dump --split-files SRR070927
```

```
Read 16486324 spots for SRR070941
Written 16486324 spots for SRR070941
Read 21688279 spots for SRR070940
Written 21688279 spots for SRR070940
~  SRR070940_1.fastq  SRR070941_1.fastq
```

In [15]:

```
# Next, the files were renamed.
# They were renamed in order to avoid confusion and to give them appropriate names instead
mv SRR070939_1.fastq mC_Dnmt.fastq
mv SRR070938_1.fastq mC_ShTet1.fastq
mv SRR070937_1.fastq mC_shScr.fastq
mv SRR070936_1.fastq mC.fastq
mv SRR070935_1.fastq hmC_Dnmt.fastq
mv SRR070934_1.fastq hmC_shTet1.fastq
mv SRR070933_1.fastq hmC_shScr.fastq
mv SRR070932_1.fastq hmC.fastq
mv SRR070931_1.fastq IgG_Chip.fastq
mv SRR070930_1.fastq SIN3A_Scruz.fastq
mv SRR070928_1.fastq Tet1-N-shTet1.fastq
mv SRR070927_1.fastq Tet1-N-shScr.fastq
```

```
~  IgG_Dnmt.fastq  IgG.fastq
```

In [16]:

```
mkdir ~/data/ncbi
vdb-config -s /resository/user/main/public/root=~/data/ncbi
```

In [1]:

```
cd ~/data/groepswerk
pwd
```

```
/home/luna.kuleuven.be/r0368030/data/groepswerk
```

TET1 binds CpG transcription start sites

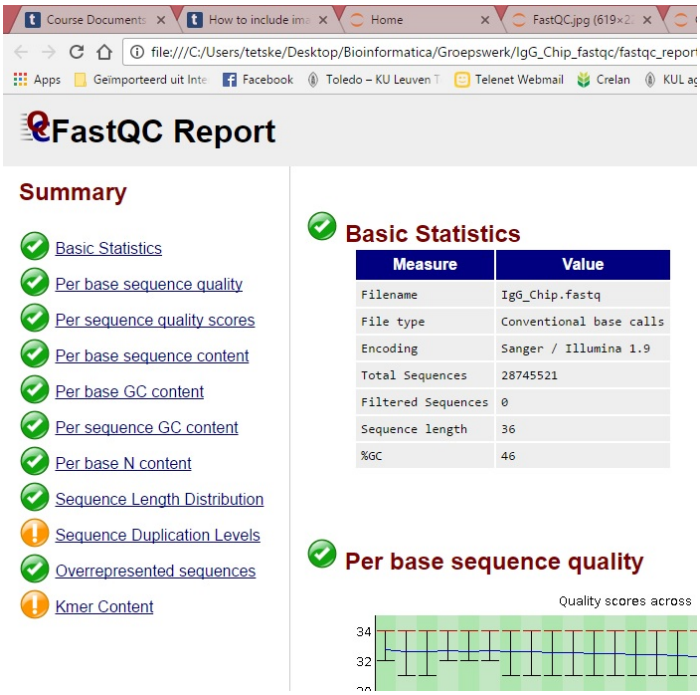
The first hypothesis of this article was to investigate which are the target genes of Tet1. Two different antibodies were used to make Tet1-knockdown cells: one antibody targeting the C-terminus and the other antibody targeting the N-terminus of the protein. We chose to only work with the ChIP-seq data of the N-terminus antibody. The analysis of the ChIP-seq data of the C-terminus antibody is similar to the analysis of data of the N-terminus antibody, and is therefore not showed in this group work.

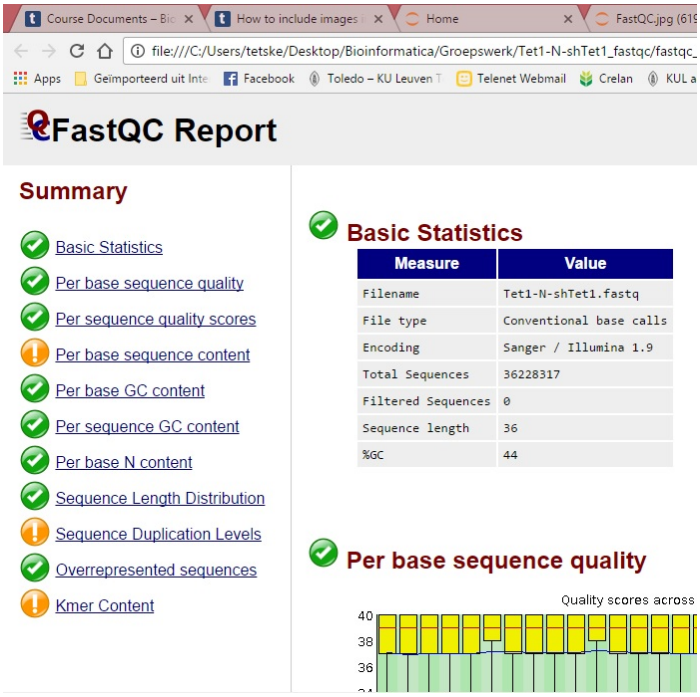
In [2]:

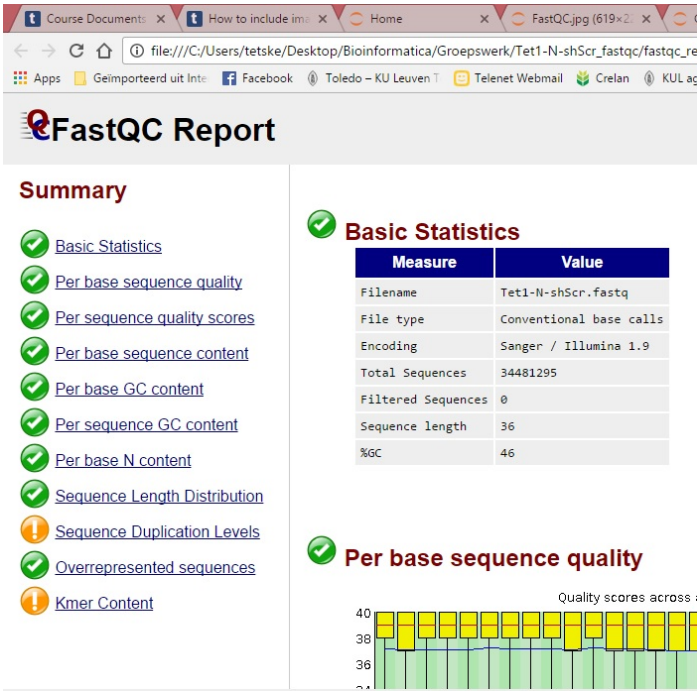
```
# Quality control of the Tet1-knockout and the normal embryonic stem cells.
# Quality control is performed on the data of Tet1, Tet1-knockdown and IgG. IgG was used as
fastqc Tet1-N-shTet1.fastq -o /mnt/storage/r0368030/groepswerk
fastqc Tet1-N-shScr.fastq -o /mnt/storage/r0368030/groepswerk
fastqc IgG_Chip.fastq -o /mnt/storage/r0368030/groepswerk
```

```
Started analysis of Tet1-N-shTet1.fastq
Approx 5% complete for Tet1-N-shTet1.fastq
Approx 10% complete for Tet1-N-shTet1.fastq
Approx 15% complete for Tet1-N-shTet1.fastq
Approx 20% complete for Tet1-N-shTet1.fastq
Approx 25% complete for Tet1-N-shTet1.fastq
Approx 30% complete for Tet1-N-shTet1.fastq
Approx 35% complete for Tet1-N-shTet1.fastq
Approx 40% complete for Tet1-N-shTet1.fastq
Approx 45% complete for Tet1-N-shTet1.fastq
Approx 50% complete for Tet1-N-shTet1.fastq
Approx 55% complete for Tet1-N-shTet1.fastq
Approx 60% complete for Tet1-N-shTet1.fastq
Approx 65% complete for Tet1-N-shTet1.fastq
Approx 70% complete for Tet1-N-shTet1.fastq
Approx 75% complete for Tet1-N-shTet1.fastq
Approx 80% complete for Tet1-N-shTet1.fastq
Approx 85% complete for Tet1-N-shTet1.fastq
Approx 90% complete for Tet1-N-shTet1.fastq
Approx 95% complete for Tet1-N-shTet1.fastq
Approx 100% complete for Tet1-N-shTet1.fastq
Analysis complete for Tet1-N-shTet1.fastq
Started analysis of Tet1-N-shScr.fastq
Approx 5% complete for Tet1-N-shScr.fastq
Approx 10% complete for Tet1-N-shScr.fastq
Approx 15% complete for Tet1-N-shScr.fastq
Approx 20% complete for Tet1-N-shScr.fastq
Approx 25% complete for Tet1-N-shScr.fastq
Approx 30% complete for Tet1-N-shScr.fastq
Approx 35% complete for Tet1-N-shScr.fastq
Approx 40% complete for Tet1-N-shScr.fastq
Approx 45% complete for Tet1-N-shScr.fastq
Approx 50% complete for Tet1-N-shScr.fastq
Approx 55% complete for Tet1-N-shScr.fastq
Approx 60% complete for Tet1-N-shScr.fastq
Approx 65% complete for Tet1-N-shScr.fastq
Approx 70% complete for Tet1-N-shScr.fastq
Approx 75% complete for Tet1-N-shScr.fastq
Approx 80% complete for Tet1-N-shScr.fastq
Approx 85% complete for Tet1-N-shScr.fastq
Approx 90% complete for Tet1-N-shScr.fastq
Approx 95% complete for Tet1-N-shScr.fastq
Approx 100% complete for Tet1-N-shScr.fastq
Analysis complete for Tet1-N-shScr.fastq
Started analysis of IgG_Chip.fastq
Approx 5% complete for IgG_Chip.fastq
Approx 10% complete for IgG_Chip.fastq
Approx 15% complete for IgG_Chip.fastq
Approx 20% complete for IgG_Chip.fastq
Approx 25% complete for IgG_Chip.fastq
Approx 30% complete for IgG_Chip.fastq
Approx 35% complete for IgG_Chip.fastq
Approx 40% complete for IgG_Chip.fastq
Approx 45% complete for IgG_Chip.fastq
Approx 50% complete for IgG_Chip.fastq
Approx 55% complete for IgG_Chip.fastq
Approx 60% complete for IgG_Chip.fastq
Approx 65% complete for IgG_Chip.fastq
Approx 70% complete for IgG_Chip.fastq
Approx 75% complete for IgG_Chip.fastq
```

```
Approx 80% complete for IgG_Chip.fastq
Approx 85% complete for IgG_Chip.fastq
Approx 90% complete for IgG_Chip.fastq
Approx 95% complete for IgG_Chip.fastq
Approx 100% complete for IgG_Chip.fastq
Analysis complete for IgG_Chip.fastq
```







The html files of the quality control showed a good quality of each parameter, except for the Kmer content and the sequence duplication levels. So we concluded that the quality of the fastq-files was good enough to be used in our further analysis.

Next, the Tet1, Tet1-knockdown and the IgG data were aligned with the mouse genome. MM9 was used in the article, therefore we used this version of the mouse genome in this group work. In order to do that, the correct mm9.bt2 had to be downloaded from the bowtie2 website.

Tet1-knockdown:

In [5]:

```
# Alignment of the Tet1-knockdown data with the MM9 mouse genome using the bowtie tool.
bowtie2-align -x /mnt/storage/r0368030/groepswerk/mm9/mm9 Tet1-N-shTet1.fastq -S Tet1-N-ShT
```

```
36228317 reads; of these:
  36228317 (100.00%) were unpaired; of these:
    1661760 (4.59%) aligned 0 times
    23684391 (65.38%) aligned exactly 1 time
    10882166 (30.04%) aligned >1 times
95.41% overall alignment rate
```

In [6]:

```
# The samfile has to be converted into a bamfile, using samtools.
# This is because a samfile is too big, and convertion into a smaller bamfile is feasible t
samtools view -S -b Tet1-N-ShTet1.sam > Tet1-N-ShTet1.bam
```

In [7]:

```
# The bamfile has to be sorted.
samtools sort -O bam -o Tet1-N-ShTet1.sorted.bam Tet1-N-ShTet1.bam
```

```
[bam_sort_core] merging from 10 files...
```

In [8]:

```
# The sorted bamfile is converted into a indexed baifile.
samtools index Tet1-N-ShTet1.sorted.bam
ls
```

```
~                   IgG_Dnmt.fastq   SIN3A_Abcam.fastq
chromFa.tar.gz      IgG.fastq        SIN3A_Scruz.fastq
genome.chrom.sizes  mC_Dnmt.fastq    Tet1-N-shScr.fastq
hmC_Dnmt.fastq      mC.fastq         Tet1-N-ShTet1.bam
hmC.fastq           mC_shScr.fastq   Tet1-N-shTet1.fastq
hmC_shScr.fastq     mC_ShTet1.fastq  Tet1-N-ShTet1.sam
hmC_shTet1.fastq    mm9              Tet1-N-ShTet1.sorted.bam
IgG_Chip.fastq      mm9.fa           Tet1-N-ShTet1.sorted.bam.bai
```

Tet1:

In [2]:

```
# Alignment of the Tet1 data with the MM9 mouse genome using the bowtie tool.
bowtie2-align -x /mnt/storage/r0368030/groepswerk/mm9/mm9 Tet1-N-shScr.fastq -S Tet1-N-shSc
```

```
34481295 reads; of these:
  34481295 (100.00%) were unpaired; of these:
    1420383 (4.12%) aligned 0 times
    24604724 (71.36%) aligned exactly 1 time
    8456188 (24.52%) aligned >1 times
95.88% overall alignment rate
```

In [3]:

```
# The samfile has to be converted into a bamfile, using samtools.
# This is because a samfile is too big, and convertion into a smaller bamfile is feasible t
samtools view -S -b Tet1-N-shScr.sam > Tet1-N-shScr.bam
```

In [4]:

```
# The bamfile has to be sorted.
samtools sort -O bam -o Tet1-N-shScr.sorted.bam Tet1-N-shScr.bam
```

```
[bam_sort_core] merging from 9 files...
```

In [5]:

```
# The sorted bamfile is converted into a indexed baifile.
samtools index Tet1-N-shScr.sorted.bam
```

IgG:

In [6]:

```
# Alignment of the IgG data with the MM9 mouse genome using the bowtie tool.
bowtie2-align -x /mnt/storage/r0368030/groepswerk/mm9/mm9 IgG_Chip.fastq -S IgG_Chip.sam
```

```
28745521 reads; of these:
  28745521 (100.00%) were unpaired; of these:
    1006683 (3.50%) aligned 0 times
    19115953 (66.50%) aligned exactly 1 time
    8622885 (30.00%) aligned >1 times
96.50% overall alignment rate
```

In [7]:

```
# The samfile has to be converted into a bamfile, using samtools.
# This is because a samfile is too big, and convertion into a smaller bamfile is feasible t
samtools view -S -b IgG_Chip.sam > IgG_Chip.bam
```

In [8]:

```
# The bamfile has to be sorted.
samtools sort -O bam -o IgG_Chip.sorted.bam IgG_Chip.bam
```

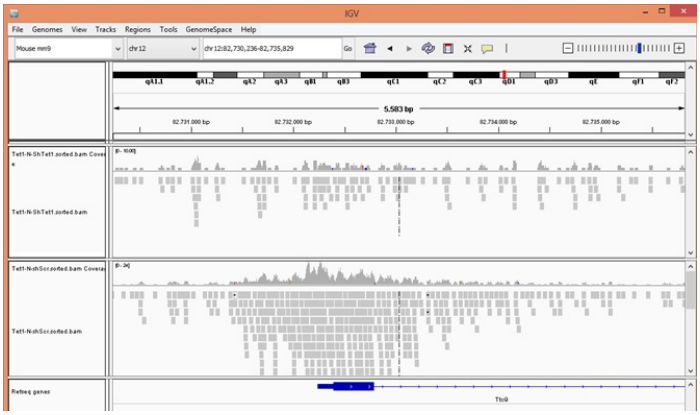
```
[bam_sort_core] merging from 8 files...
```

In [9]:

```
# The sorted bamfile is converted into a indexed baifile.
samtools index IgG_Chip.sorted.bam
```

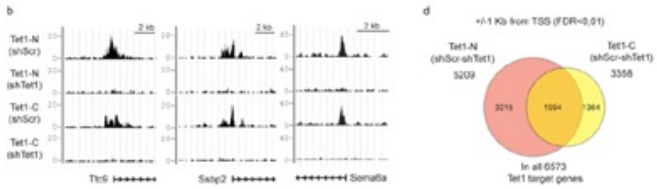
All alignments show a high percentage, which means that the coverage is good. The sorted bamfiles were used in IGV (see printscreens below).

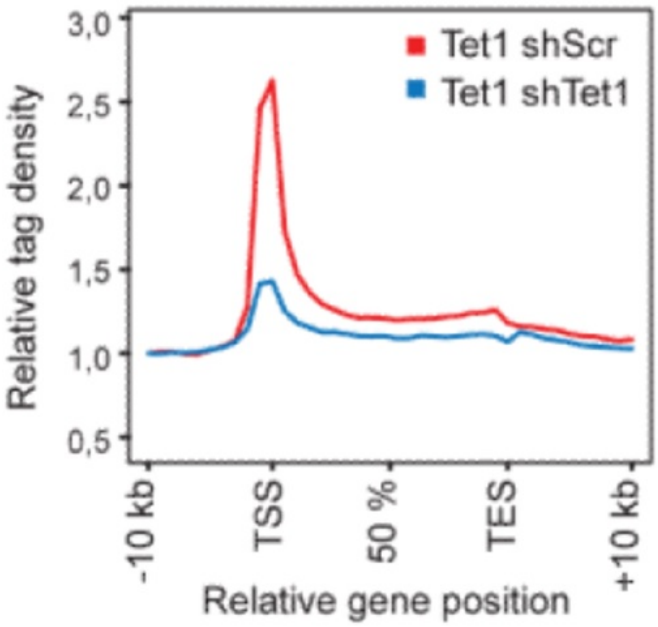
The upper track is the Tet1 knockdown and the lower track is the Tet1 wild type. We looked at the three genes that were analysed in the paper: Ttc9, Ssbp2 and Sema6. All these genes showed higher peaks in the Tet1 wildtype cells than in the Tet1 knockdown cells, as expected (and as shown in the paper). IGV also showed that Tet1 mostly binds to the promotors (transcription start sites, TSS) of these genes.











The peaks in Tet1 and Tet1-knockdown data were quantified (i.e. peakcalling) using the MACS2 tool. We used the sorted IgG_ChIP bamfile as a control/input. The command was performed over the whole mouse genome with FDR <0.01 (as done in the article).

In [14]:

```
# Peakcalling Tet1 knockdown.
mkdir macs2_Tet1_poging2
macs2 callpeak -t Tet1-N-ShTet1.sorted.bam -c IgG_Chip.sorted.bam --outdir macs2_ShTet1_pog
```

```
INFO  @ Sun, 29 Jan 2017 18:45:22:
# Command line: callpeak -t Tet1-N-ShTet1.sorted.bam -c IgG_Chip.sorted.bam
 --outdir macs2_ShTet1_poging2 -n Tet1_poging2 -g mm -q 0.01
# ARGUMENTS LIST:
# name = Tet1_poging2
# format = AUTO
# ChIP-seq file = ['Tet1-N-ShTet1.sorted.bam']
# control file = ['IgG_Chip.sorted.bam']
# effective genome size = 1.87e+09
# band width = 300
# model fold = [5, 50]
# qvalue cutoff = 1.00e-02
# Larger dataset will be scaled towards smaller dataset.
# Range for calculating regional lambda is: 1000 bps and 10000 bps
# Broad region calling is off

INFO  @ Sun, 29 Jan 2017 18:45:22: #1 read tag files...
INFO  @ Sun, 29 Jan 2017 18:45:22: #1 read treatment tags...
INFO  @ Sun, 29 Jan 2017 18:45:22: Detected format is: BAM
INFO  @ Sun, 29 Jan 2017 18:45:22: * Input file is gzipped.
INFO  @ Sun, 29 Jan 2017 18:45:26:  1000000
INFO  @ Sun, 29 Jan 2017 18:45:31:  2000000
INFO  @ Sun, 29 Jan 2017 18:45:35:  3000000
INFO  @ Sun, 29 Jan 2017 18:45:39:  4000000
INFO  @ Sun, 29 Jan 2017 18:45:43:  5000000
INFO  @ Sun, 29 Jan 2017 18:45:47:  6000000
INFO  @ Sun, 29 Jan 2017 18:45:52:  7000000
INFO  @ Sun, 29 Jan 2017 18:45:56:  8000000
INFO  @ Sun, 29 Jan 2017 18:46:00:  9000000
INFO  @ Sun, 29 Jan 2017 18:46:05:  10000000
INFO  @ Sun, 29 Jan 2017 18:46:09:  11000000
INFO  @ Sun, 29 Jan 2017 18:46:13:  12000000
INFO  @ Sun, 29 Jan 2017 18:46:17:  13000000
INFO  @ Sun, 29 Jan 2017 18:46:22:  14000000
INFO  @ Sun, 29 Jan 2017 18:46:26:  15000000
INFO  @ Sun, 29 Jan 2017 18:46:30:  16000000
INFO  @ Sun, 29 Jan 2017 18:46:35:  17000000
INFO  @ Sun, 29 Jan 2017 18:46:39:  18000000
INFO  @ Sun, 29 Jan 2017 18:46:44:  19000000
INFO  @ Sun, 29 Jan 2017 18:46:48:  20000000
INFO  @ Sun, 29 Jan 2017 18:46:52:  21000000
INFO  @ Sun, 29 Jan 2017 18:46:57:  22000000
INFO  @ Sun, 29 Jan 2017 18:47:01:  23000000
INFO  @ Sun, 29 Jan 2017 18:47:05:  24000000
INFO  @ Sun, 29 Jan 2017 18:47:09:  25000000
INFO  @ Sun, 29 Jan 2017 18:47:14:  26000000
INFO  @ Sun, 29 Jan 2017 18:47:18:  27000000
INFO  @ Sun, 29 Jan 2017 18:47:22:  28000000
INFO  @ Sun, 29 Jan 2017 18:47:27:  29000000
INFO  @ Sun, 29 Jan 2017 18:47:31:  30000000
INFO  @ Sun, 29 Jan 2017 18:47:35:  31000000
INFO  @ Sun, 29 Jan 2017 18:47:40:  32000000
INFO  @ Sun, 29 Jan 2017 18:47:44:  33000000
INFO  @ Sun, 29 Jan 2017 18:47:48:  34000000
INFO  @ Sun, 29 Jan 2017 18:47:52:  35000000
INFO  @ Sun, 29 Jan 2017 18:47:54:  36000000
INFO  @ Sun, 29 Jan 2017 18:47:55: #1.2 read input tags...
INFO  @ Sun, 29 Jan 2017 18:47:55: Detected format is: BAM
INFO  @ Sun, 29 Jan 2017 18:47:55: * Input file is gzipped.
INFO  @ Sun, 29 Jan 2017 18:47:59:  1000000
```

```
INFO  @ Sun, 29 Jan 2017 18:48:03:  2000000
INFO  @ Sun, 29 Jan 2017 18:48:07:  3000000
INFO  @ Sun, 29 Jan 2017 18:48:11:  4000000
INFO  @ Sun, 29 Jan 2017 18:48:15:  5000000
INFO  @ Sun, 29 Jan 2017 18:48:19:  6000000
INFO  @ Sun, 29 Jan 2017 18:48:23:  7000000
INFO  @ Sun, 29 Jan 2017 18:48:27:  8000000
INFO  @ Sun, 29 Jan 2017 18:48:31:  9000000
INFO  @ Sun, 29 Jan 2017 18:48:36:  10000000
INFO  @ Sun, 29 Jan 2017 18:48:40:  11000000
INFO  @ Sun, 29 Jan 2017 18:48:44:  12000000
INFO  @ Sun, 29 Jan 2017 18:48:48:  13000000
INFO  @ Sun, 29 Jan 2017 18:48:52:  14000000
INFO  @ Sun, 29 Jan 2017 18:48:56:  15000000
INFO  @ Sun, 29 Jan 2017 18:49:01:  16000000
INFO  @ Sun, 29 Jan 2017 18:49:05:  17000000
INFO  @ Sun, 29 Jan 2017 18:49:09:  18000000
INFO  @ Sun, 29 Jan 2017 18:49:13:  19000000
INFO  @ Sun, 29 Jan 2017 18:49:17:  20000000
INFO  @ Sun, 29 Jan 2017 18:49:21:  21000000
INFO  @ Sun, 29 Jan 2017 18:49:25:  22000000
INFO  @ Sun, 29 Jan 2017 18:49:29:  23000000
INFO  @ Sun, 29 Jan 2017 18:49:34:  24000000
INFO  @ Sun, 29 Jan 2017 18:49:38:  25000000
INFO  @ Sun, 29 Jan 2017 18:49:42:  26000000
INFO  @ Sun, 29 Jan 2017 18:49:46:  27000000
INFO  @ Sun, 29 Jan 2017 18:49:50:  28000000
INFO  @ Sun, 29 Jan 2017 18:49:51: #1 tag size is determined as 36 bps
INFO  @ Sun, 29 Jan 2017 18:49:51: #1 tag size = 36
INFO  @ Sun, 29 Jan 2017 18:49:51: #1  total tags in treatment: 34566557
INFO  @ Sun, 29 Jan 2017 18:49:51: #1 user defined the maximum tags...
INFO  @ Sun, 29 Jan 2017 18:49:51: #1 filter out redundant tags at the same
 location and the same strand by allowing at most 1 tag(s)
INFO  @ Sun, 29 Jan 2017 18:49:52: #1  tags after filtering in treatment: 29
951934
INFO  @ Sun, 29 Jan 2017 18:49:52: #1  Redundant rate of treatment: 0.13
INFO  @ Sun, 29 Jan 2017 18:49:52: #1  total tags in control: 27738838
INFO  @ Sun, 29 Jan 2017 18:49:52: #1 user defined the maximum tags...
INFO  @ Sun, 29 Jan 2017 18:49:52: #1 filter out redundant tags at the same
 location and the same strand by allowing at most 1 tag(s)
INFO  @ Sun, 29 Jan 2017 18:49:53: #1  tags after filtering in control: 2466
5059
INFO  @ Sun, 29 Jan 2017 18:49:53: #1  Redundant rate of control: 0.11
INFO  @ Sun, 29 Jan 2017 18:49:53: #1 finished!
INFO  @ Sun, 29 Jan 2017 18:49:53: #2 Build Peak Model...
INFO  @ Sun, 29 Jan 2017 18:49:53: #2 looking for paired plus/minus strand p
eaks...
INFO  @ Sun, 29 Jan 2017 18:49:56: #2 number of paired peaks: 1459
INFO  @ Sun, 29 Jan 2017 18:49:56: start model_add_line...
INFO  @ Sun, 29 Jan 2017 18:49:56: start X-correlation...
INFO  @ Sun, 29 Jan 2017 18:49:56: end of X-cor
INFO  @ Sun, 29 Jan 2017 18:49:56: #2 finished!
INFO  @ Sun, 29 Jan 2017 18:49:56: #2 predicted fragment length is 46 bps
INFO  @ Sun, 29 Jan 2017 18:49:56: #2 alternative fragment length(s) may be
 46 bps
INFO  @ Sun, 29 Jan 2017 18:49:56: #2.2 Generate R script for model : macs2_
ShTet1_poging2/Tet1_poging2_model.r
WARNING @ Sun, 29 Jan 2017 18:49:56: #2 Since the d (46) calculated from pai
red-peaks are smaller than 2*tag length, it may be influenced by unknown seq
uencing problem!
WARNING @ Sun, 29 Jan 2017 18:49:56: #2 You may need to consider one of the
```

```
 other alternative d(s): 46

WARNING @ Sun, 29 Jan 2017 18:49:56: #2 You can restart the process with --n
omodel --extsize XXX with your choice or an arbitrary number. Nontheless, MA
CS will continute computing.
INFO  @ Sun, 29 Jan 2017 18:49:56: #3 Call peaks...
INFO  @ Sun, 29 Jan 2017 18:49:56: #3 Pre-compute pvalue-qvalue table...
INFO  @ Sun, 29 Jan 2017 18:51:58: #3 Call peaks for each chromosome...
INFO  @ Sun, 29 Jan 2017 18:53:04: #4 Write output xls file... macs2_ShTet1_
poging2/Tet1_poging2_peaks.xls
INFO  @ Sun, 29 Jan 2017 18:53:04: #4 Write peak in narrowPeak format fil
e... macs2_ShTet1_poging2/Tet1_poging2_peaks.narrowPeak
INFO  @ Sun, 29 Jan 2017 18:53:04: #4 Write summits bed file... macs2_ShTet1
_poging2/Tet1_poging2_summits.bed
INFO  @ Sun, 29 Jan 2017 18:53:05: Done!
```

In [15]:

```
ls macs2_ShTet1_poging2
```

```
Tet1_poging2_model.r           Tet1_poging2_peaks.xls
Tet1_poging2_peaks.narrowPeak  Tet1_poging2_summits.bed
```

In [16]:

```
cat macs2_ShTet1_poging2/Tet1_poging2_summits.bed | wc -l
```

```
314
```

The sorted bam-file of the Tet-knockdown was used as input, the bamfile of the normal embryonic stem cell was used as control. We performed this command over the whole mouse genome with FDR<0.01 (like the article). This results in 314 peaks for the Tet1 knockdown and 37518 peaks for the Tet1 wildtype. The narrowpeak and the bed files were analysed in IGV.

In the paper, they used the MEDIPS tool to call peaks. This tool is specially designed for methylation (epigenetics) data. However, we were not able to install the MEDIPS package so we could not compare the MACS2 results with the MEDIPS results.

In [17]:

```
# Peakcalling wildtype Tet1.
mkdir macs2_Tet1_2_poging2
macs2 callpeak -t Tet1-N-shScr.sorted.bam -c IgG_Chip.sorted.bam --outdir macs2_Tet1-N-ShSc
```

In [18]:

```
cat macs2_Tet1-N-ShScr/Tet1_2_poging2_summits.bed | wc -l
```

```
37518
```

We took a look at the narrowpeak files in IGV:

The upper track is the Tet1 wildtype and the lower track is the Tet1 knockdown. The printscreens show that there are only peaks in the Tet1 wildtype cells and that these peaks are located in the TSS regions of the Ttc9, Ssbp2 and Sema6 genes.







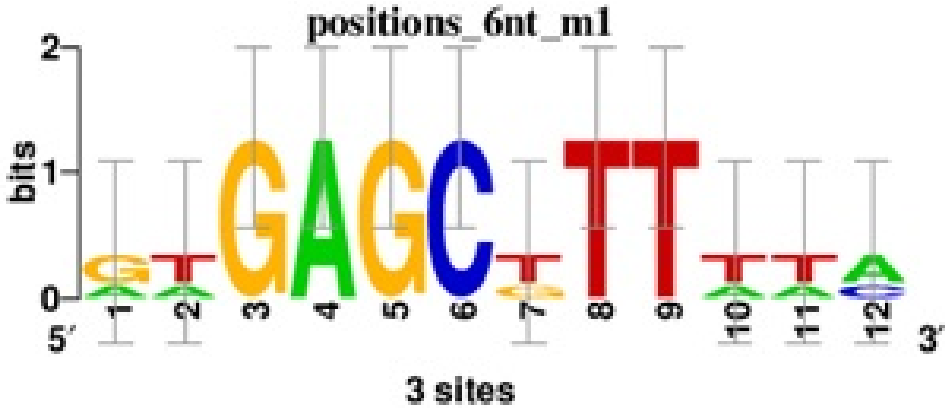
In [4]:

```
cat macs2_Tet1-N-ShTet1/Tet1_2_2_summits.bed | wc -l
```

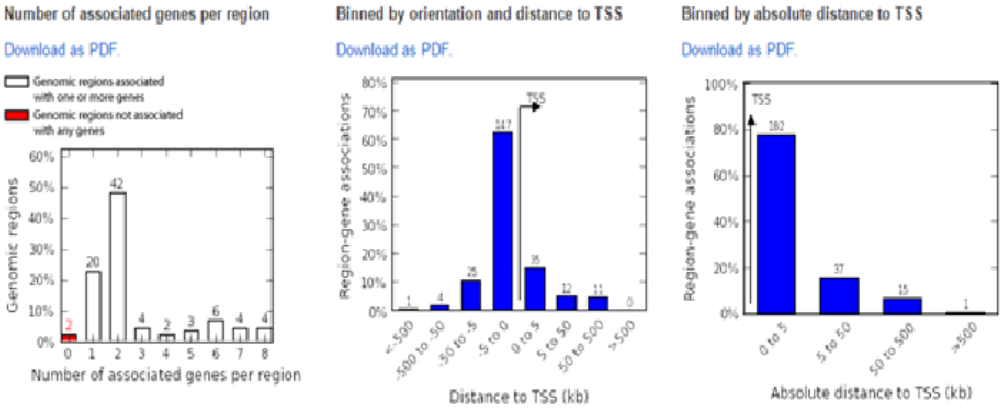
```
38532
```

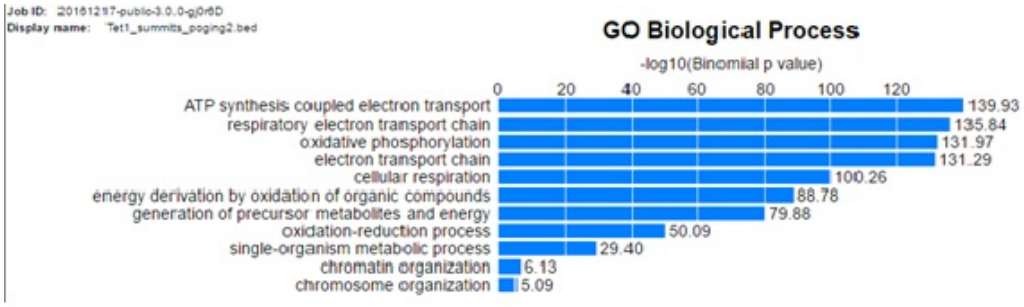
Motif analysis using RSAT.

Since we did not find the bindingmotif of Tet1 in literature, we assumed that it is not known yet. Therefore, we chose to use RSAT instead of icis-target to determine the bindingmotif of Tet1. First, the fastasequences of the peaks were requested in UCSC. These sequences were then uploaded in RSAT, which resulted in 13 motifs including the one that is depicted below.



Next, we used GREAT to link the peaks to genes. These graphs illustrate that Tet1 mostly binds to the transcription start sites of genes, as was shown in the paper. Gene Ontology (in GREAT) showed that most of the target genes of Tet1 have a function that is related to the mitochondrial energy production. In the paper, Gene Ontology showed that Ter1 target genes are involved in a variety of basic cellular processes, but no more details are given.





The second hypothesis that was tested in the article was if Tet1 associates with the SIN3A complex. The Microarray data showed that Tet1 contributes to transcriptional repression and know they wanted to know the mechanism of the repression. The purification of protein associated with double-epitope Flag-haemagglutinin-tagged Tet1 resulted in the identification of SIN3A and other components of the SIN3A co-repressor complex. It has already been shown that this repressor complex contributes to transcriptional repression by mediating histone deacetylation. In the paper, they performed ChIP-seq analysis to investigate if SIN3A and Tet1 colocalize on target genes using two different antibodies directed against SIN3A. We only analysed the ChIP-seq data from one antibody (Santa Cruz).

We first aligned the fastqfile of SIN3A-ChIP-seq data with the mm9 genome, which resulted in an alignment rate of 81.84%.

In [2]:

```
bowtie2-align -x /mnt/storage/r0368030/groepswerk/mm9/mm9 SIN3A_Scruz.fastq -S SIN3A.sam
```

```
21672699 reads; of these:
  21672699 (100.00%) were unpaired; of these:
    3934818 (18.16%) aligned 0 times
    12693630 (58.57%) aligned exactly 1 time
    5044251 (23.27%) aligned >1 times
81.84% overall alignment rate
```

In [3]:

```
samtools view -S -b SIN3A.sam > SIN3A.bam
```

In [4]:

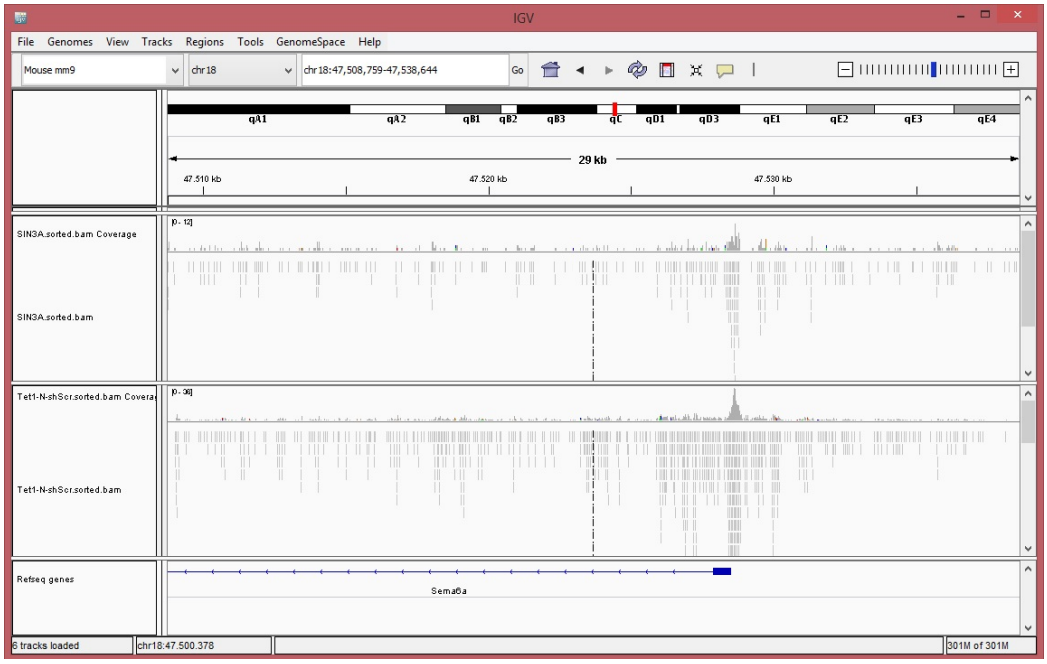
```
samtools sort -O bam -o SIN3A.sorted.bam SIN3A.bam
```

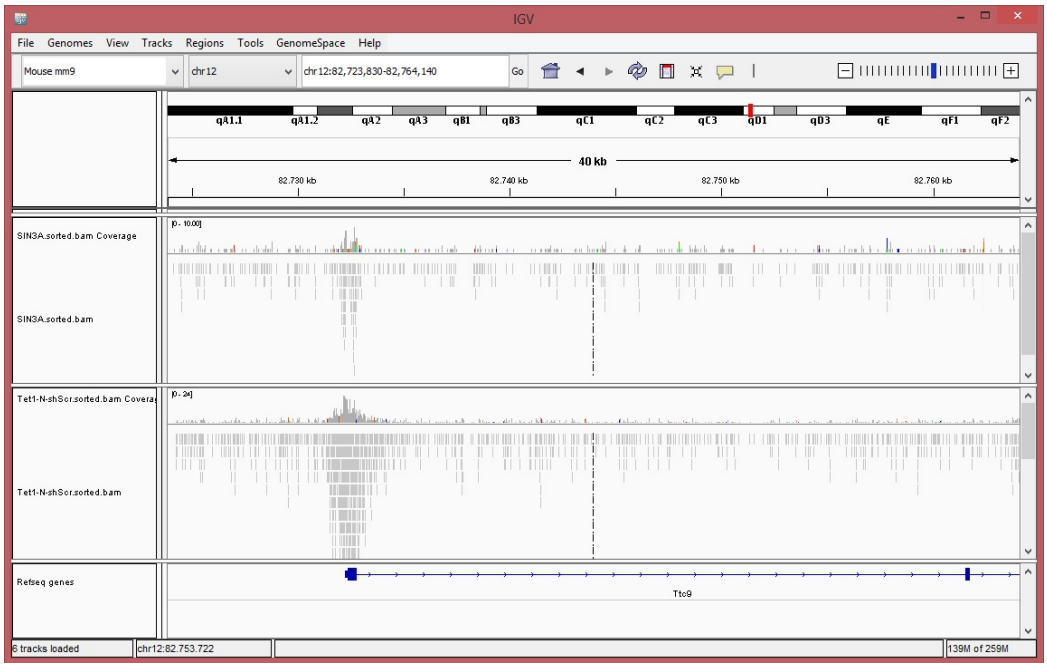
```
[bam_sort_core] merging from 6 files...
```

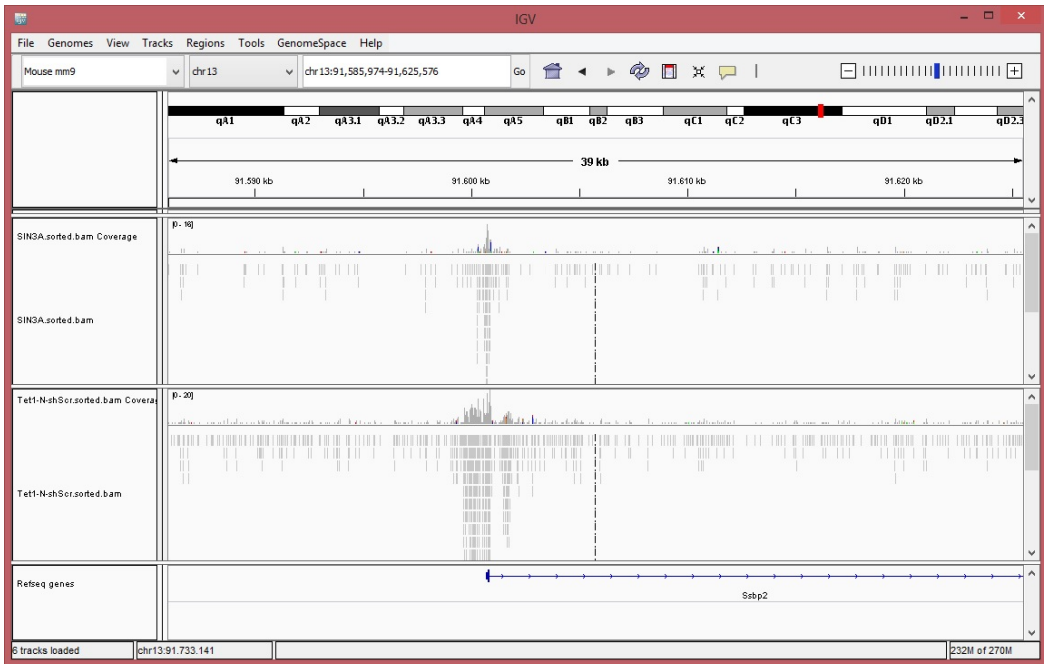
In [5]:

```
samtools index SIN3A.sorted.bam
```

Next we created bam and bai files, which were visualized in IGV together with the bam files of Tet1. Again, we looked at the same genes that are discussed in the paper and found that the peaks of SIN3A and Tet1 overlap, meaning that they have the same target genes.


IGV
File Genomes View Tracks Regions Tools GenomeSpace Help
Mouse mm9
chr18
chr18:47,508,759-47,538,644
Go
29 kb
47.510 kb
47.520 kb
47.530 kb
SIN3A.sorted.bam Coverage
SIN3A.sorted.bam
Tet1-N-shScr.sorted.bam Coverag
Tet1-N-shScr.sorted.bam
Refseq genes
Sema6a
6 tracks loaded
chr18:47,500,378
301M of 301M


IGV
File Genomes View Tracks Regions Tools GenomeSpace Help
Mouse mm9
chr12
chr12:82,723,830-82,764,140
Go
40 kb
82.730 kb
82.740 kb
82.750 kb
82.760 kb
SIN3A.sorted.bam Coverage
SIN3A.sorted.bam
Tet1-N-shScr.sorted.bam Coverag
Tet1-N-shScr.sorted.bam
Refseq genes
Ttc9
6 tracks loaded
chr12:82,753,722
139M of 259M


IGV
File Genomes View Tracks Regions Tools GenomeSpace Help
Mouse mm9
chr13
chr13:91,585,974-91,625,576
Go
39 kb
91.590 kb
91.600 kb
91.610 kb
91.620 kb
SIN3A.sorted.bam Coverage
SIN3A.sorted.bam
Tet1-N-shScr.sorted.bam Coverag
Tet1-N-shScr.sorted.bam
Refseq genes
Ssbp2
6 tracks loaded
chr13:91,733,141
232M of 270M