CD33 in Immunotherapy

Ionelia Buzatu

Cell

Genetic Inactivation of CD33 in Hematopoietic Stem Cells to Enable CAR T Cell Immunotherapy for Acute Myeloid Leukemia

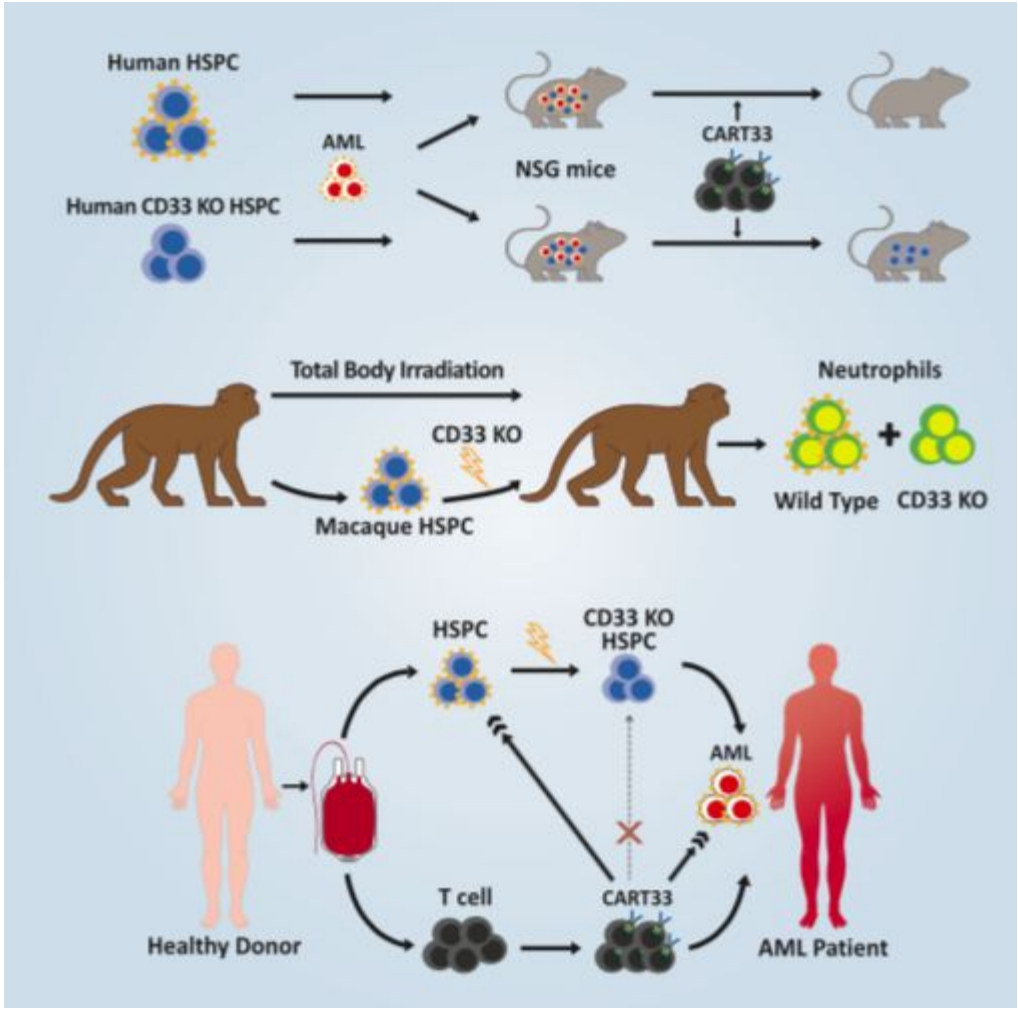
SUMMARY

The absence of cancer-restricted surface markers is a major impediment to antigen-specific immunotherapy using chimeric antigen receptor (CAR) T cells. For example, targeting the canonical myeloid marker CD33 in acute myeloid leukemia (AML) results in toxicity from destruction of normal myeloid cells. We hypothesized that a leukemia-specific antigen could be created by deleting CD33 from normal hematopoietic stem and progenitor cells (HSPCs), thereby generating a hematopoietic system resistant to CD33-targeted therapy and enabling specific targeting of AML with CAR T cells. We generated CD33-deficient human HSPCs and demonstrated normal engraftment and differentiation in immunodeficient mice. Autologous CD33 KO HSPC transplantation in rhesus macaques demonstrated long-term multilineage engraftment of gene-edited cells with normal myeloid function. CD33-deficient cells were impervious to CD33-targeting CAR T cells, allowing for efficient elimination of leukemia without myelotoxicity. These studies illuminate a novel approach to antigen-specific immunotherapy by genetically engineering the host to avoid on-target, off-tumor toxicity.

INTRODUCTION

Recent advances in cancer immunotherapy with chimeric antigen receptor (CAR) T cells have enabled eradication of cells expressing a specific surface antigen. While this approach has been successful in targeting CD19 in B cell neoplasms (Davila et al., 2014; Kochenderfer et al., 2015; Lee et al., 2015; Maude et al., 2014; Turtle et al., 2016), it does not discriminate between normal and malignant B cells, and thus, the feasibility of this therapy rests on the tolerability of prolonged B cell aplasia. However, most malignancies do not have an expendable normal tissue counterpart, and whether the success of CAR T cells can be extrapolated beyond B cell neoplasms will depend on the ability to develop strategies to mitigate toxicity to normal cells.

The majority of acute myeloid leukemia (AML) patients relapse despite intensive therapy. AML cell surface antigens are shared with normal myeloid progenitors (Levine et al., 2015; Taussig et al., 2005); therefore, targeting AML also generates toxicity to the myeloid system (Gill et al., 2014; Leong et al., 2017; Mardiros et al., 2013; Pizzitola et al., 2014; Tashiro et al., 2017). While strategies to produce transient CAR T cells are being explored in clinical trials to avoid long-term myeloablation while targeting AML (e.g., #NCT03126864, #NCT02159495), this negates a fundamental strength of this therapy—namely, its long-term anti-tumor immune surveillance. Therefore, in the absence of a truly AML-specific antigen, novel approaches are required to definitively target AML while sparing normal hematopoiesis.


Human HSPC
Human CD33 KO HSPC
AML
NSG mice
CART33
Total Body Irradiation
CD33 KO
Macaque HSPC
Neutrophils
Wild Type
CD33 KO
HSPC
CD33 KO HSPC
AML
T cell
CART33
Healthy Donor
AML Patient

My pipeline: analisy.sh

```
#!bin/bash

#!!!reminder: nohup will auto redirect output errors to nohup.out

# store stdout and stderr
mkdir -p /home/buzatu1695848/data/redirection/

echo "Downloading fastq files"
bash downlowad_ftps.sh &> /home/buzatu1695848/data/redirection/out_downlowad_ftps.out

echo "Counting unique and total sequences"
bash count_seqFastq.sh &>  /home/buzatu1695848/data/redirection/out_count_seqFastq.out

echo "Doing the Quality Check"
bash quality_check.sh &>  /home/buzatu1695848/data/redirection/out_quality_check.out

echo "Doing the Alignment"
bash mapping.sh &> /home/buzatu1695848/data/redirection/out_mapping.out
mv /home/buzatu1695848/data/fastq/*.bam /home/buzatu1695848/data/bams

echo "Feature Counting now"
bash feature_count.sh &> /home/buzatu1695848/data/redirection/out_rowcounts.out


echo "Building of rowcounts table and DE analysis"
Rscript DESeqAnalysis.r &> /home/buzatu1695848/data/redirection/out_DE.out

echo "DONE"
```

Preparing for the fastq url retrieval: downlowad_ftps.sh

```
#!/bin/sh

cd /home/buzatu1695848/data/fastq

wget -nc  --header="Host: www.ebi.ac.uk" --header="User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14
_2) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/71.0.3578.98 Safari/537.36" --header="Accept: text/html,a
pplication/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8" --header="Accept-Language: en-U
S,en;q=0.9,it;q=0.8" --header="Referer: https://www.ebi.ac.uk/ena/data/view/PRJNA453558" --header="Cookie:
JSESSIONID=B7C097392F9B91F6F231310320E548B2; X-Mapping-ceffhojj=08190EABCB735AD60C95FE3D5A378754; X-Mapping
-ejmejipg=1388625A69E68E7C666D736DCA1CF14A" --header="Connection: keep-alive" "https://www.ebi.ac.uk/ena/da
ta/warehouse/filereport?accession=PRJNA453558&result=read_run&fields=study_accession,sample_accession,secon
dary_sample_accession,experiment_accession,run_accession,tax_id,scientific_name,instrument_model,library_la
yout,fastq_ftp,fastq_galaxy,submitted_ftp,submitted_galaxy,sra_ftp,sra_galaxy,cram_index_ftp,cram_index_gal
axy&download=txt" -O "PRJNA453558.txt" -c

######################## get the url of all samples from the ena table
awk -F ' ' '{print $13}' PRJNA453558.txt > ftp_samples.txt

for i in $(cat /home/buzatu1695848/data/ftp_samples.txt);
do
        wget -nc ${i}
done;


gunzip /home/buzatu1695848/data/fastq/*.gz
```

Total number of read and unique reads: count_seqFastq.sh

```
#!/bin/bash


cd /home/buzatu1695848/
echo "HELLO, doing the counting.."

## total reads and unique read for each fastq
for file in $(ls /home/buzatu1695848/data/fastq/removeDIR_after/*.fastq);
do
        echo "For " ${file##*/}
        awk '((NR-2)%4==0){read=$1;total++;count[read]++}END{for(read in count){if(!max||count[rea
d]>max) {max=count[read];maxRead=read};if(count[read]==1){unique++}};print "Total reads: "total,"\
nUniques reads: "unique}' $file
done > readsNumber.txt

echo "DONE"
```

```
For  SRR7063337.fastq
Total reads: 23968042
Uniques reads: 8949963
For  SRR7063338.fastq
Total reads: 22596988
Uniques reads: 8326445
For  SRR7063339.fastq
Total reads: 23738394
Uniques reads: 8528300
For  SRR7063340.fastq
Total reads: 22966922
Uniques reads: 8282749
For  SRR7063341.fastq
Total reads: 22056066
Uniques reads: 8202043
For  SRR7063342.fastq
Total reads: 23205463
Uniques reads: 8227435
For  SRR7063343.fastq
Total reads: 23214810
Uniques reads: 8255706
For  SRR7063344.fastq
Total reads: 23322024
Uniques reads: 8720523
For  SRR7063345.fastq
Total reads: 23016941
Uniques reads: 8826912
For  SRR7063346.fastq
Total reads: 23730733
Uniques reads: 8682535
```

Quality Check: quality_check.sh

```
#!bin/bash

mkdir -p  /home/buzatu1695848/analysis/qualityCheck

# qualtity check and report outputs

fastqc -t 10 /home/buzatu1695848/data/fastq/*.fastq -o /home/buzatu1695848/analysis/qualityCheck

rm /home/buzatu1695848/analysis/qualityCheck/*.html
unzip /home/buzatu1695848/analysis/qualityCheck/*.zip


~
~
~
```

Alignment to the HGR: mapping.sh

```
#!bin/bash


mkdir -p /home/buzatu1695848/data/bams
cd /home/buzatu1695848/data/bams

echo "Mapping..."
for file in $(ls /home/buzatu1695848/data/fastq/*.fastq)
do
        hisat2 -p 10 -x /data/public/index/hisat_hg38  -U $file | samtools view -bo ${file}.bam
done;
```

Summarizing read to genomic features: feature_count.sh

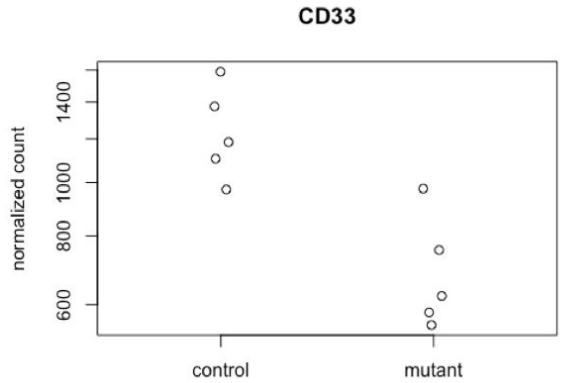
```
#!bin/bash

# GCF_000001405.38_GRCh38.p12_genomic.gff

mkdir -p /home/buzatu1695848/data/sample_counts
mkdir -p /home/buzatu1695848/data/sample_counts/summaries

echo "Counting..."
for file in $(ls /home/buzatu1695848/data/bams/*.bam)
do

        # ${file%.*}
       featureCounts -T 20 -a /data/public/genome/GCF_000001405.38_GRCh38.p12_genomic.gff -g gene  -o ${file%%.*}_counts.txt ${file}

       sed -i '1d' ${file%%.*}_counts.txt
       cut -f -1,7 ${file%%.*}_counts.txt > ${file%%.*}_rowcounts.txt
       mv  ${file%%.*}_rowcounts.txt /home/buzatu1695848/data/sample_counts
        mv  ${file%%.*}_counts.txt.summary /home/buzatu1695848/data/sample_counts/summaries
        # rm  ${file%%.*}_counts.txt
done
```

Building the rowcounts table

GSM3110907 Donor A-ctrl
GSM3110908 Donor A-KO
GSM3110909 Donor B-ctrl
GSM3110910 Donor B-KO
GSM3110911 Donor C-ctrl
GSM3110912 Donor C-KO
GSM3110913 Donor D-ctrl
GSM3110914 Donor D-KO
GSM3110915 Donor E-ctrl
GSM3110916 Donor E-KO

```
> head(all)
       Geneid ctrl_A ctrl_B ctrl_C ctrl_D ctrl_E KO_A KO_B KO_C KO_D KO_E
1      DDX11L1      0      0      0      0      0    0    0    0    0    0
2       WASH7P      8     20      5     18      5    3    5    8   40    9
3    MIR6859-1      0      0      1      0      1    0    0    0    3    0
4 MIR1302-2HG      0      0      0      0      0    0    0    0    0    0
5    MIR1302-2      0      0      0      0      0    0    0    0    0    0
6      FAM138A      0      0      0      0      0    0    0    0    0    0
```

```
> filter(all, Geneid == "CD33")
  Geneid ctrl_A ctrl_B ctrl_C ctrl_D ctrl_E KO_A KO_B KO_C KO_D KO_E
1   CD33   1221   1392   1507   1120   1008  537  762  942  594  649
```

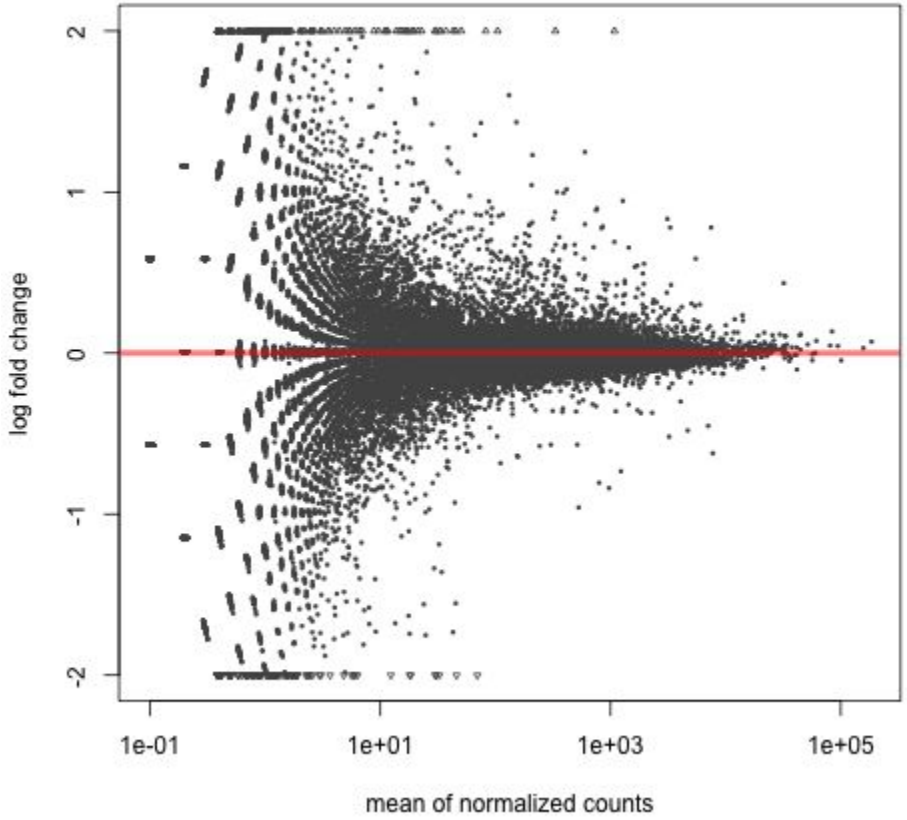


There is high similarity among the samples

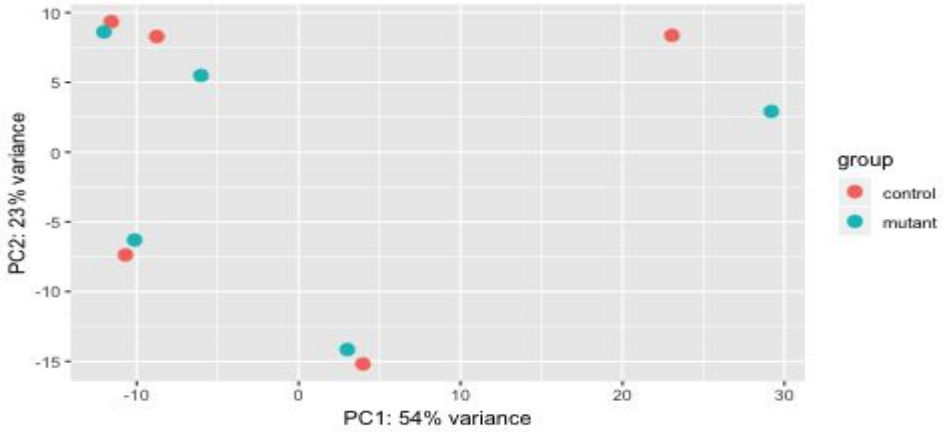
```
> m = estimateSizeFactors(model)
> sizeFactors(m)
   ctrl_A    ctrl_B    ctrl_C    ctrl_D    ctrl_E      KO_A      KO_B      KO_C
1.0317857 1.0128628 0.9486910 1.0144000 1.0384595 0.9758392 1.0114985 0.9671512
     KO_D      KO_E
1.0239434 1.0443589
```

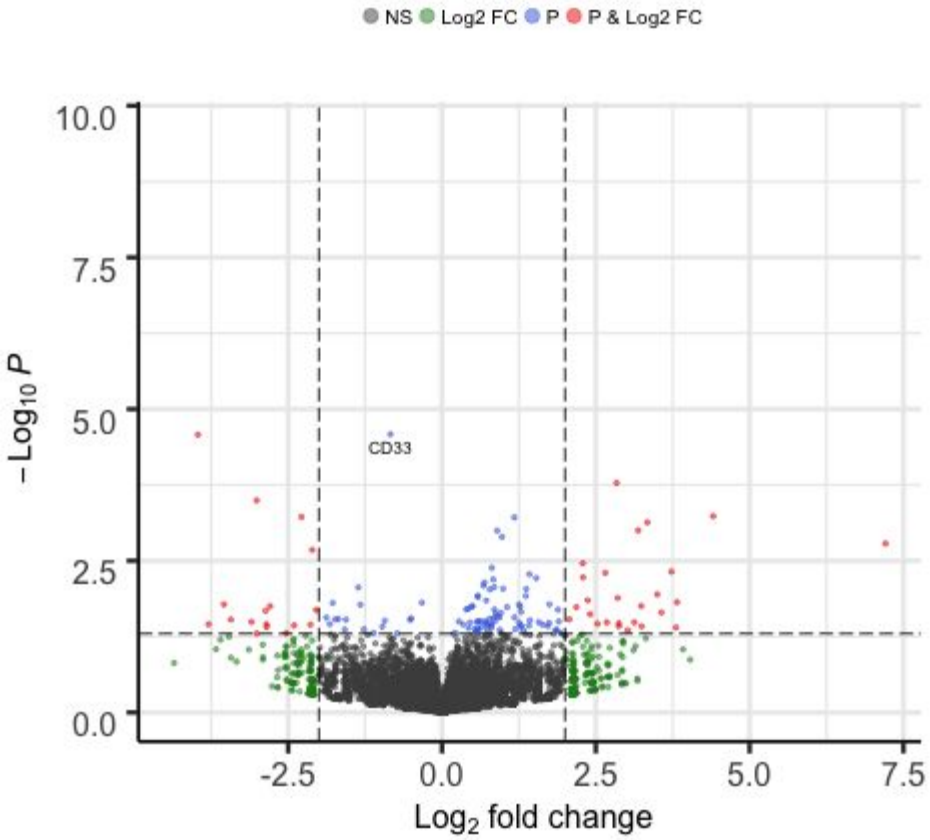
DESeq analysis

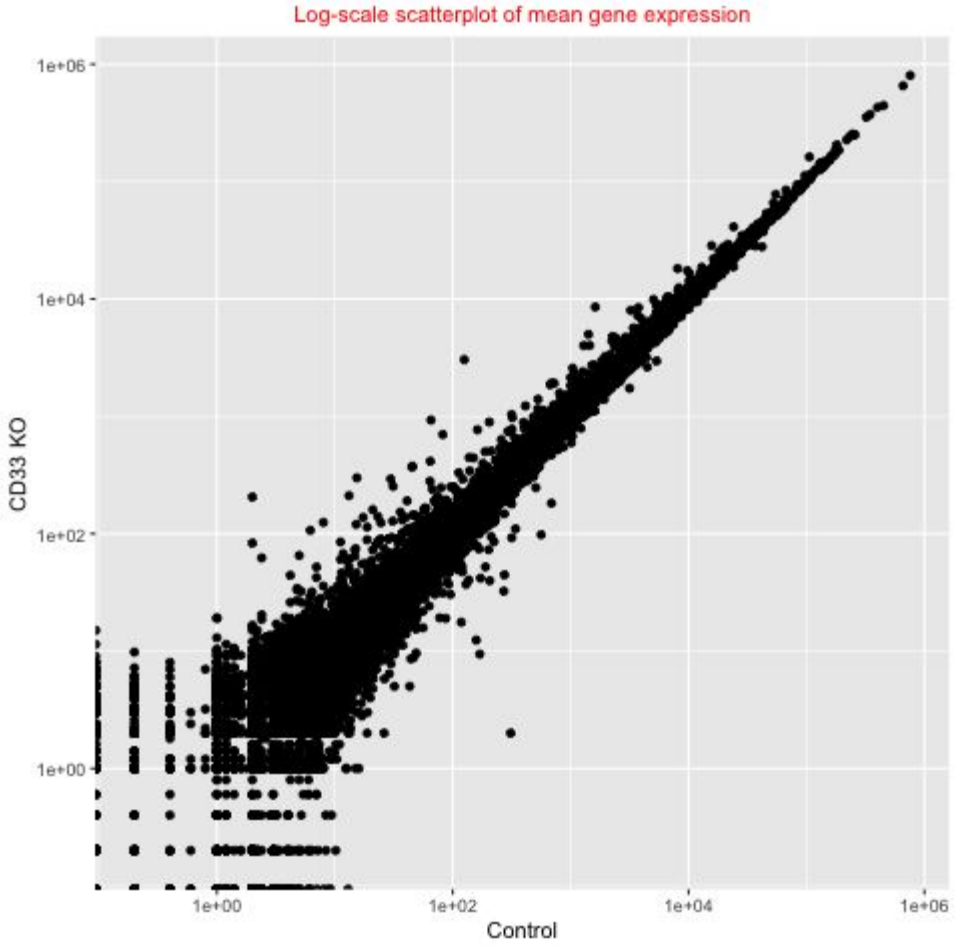
```
> table_results[1:20,]
            row   baseMean log2FoldChange     lfcSE         stat    pvalue      padj
1         DDX11L1  0.0000000             NA        NA           NA        NA        NA
2          WASH7P 11.9300825    0.210795351 0.7267117  0.290067382 0.7717647 0.9999836
3      MIR6859-1  0.4946898    0.535855828 2.6260940  0.204050513 0.8383140 0.9999836
4    MIR1302-2HG  0.0000000             NA        NA           NA        NA        NA
5      MIR1302-2  0.0000000             NA        NA           NA        NA        NA
6        FAM138A  0.0000000             NA        NA           NA        NA        NA
7          OR4F5  0.0000000             NA        NA           NA        NA        NA
8   LOC100996442  4.6769746    0.424967018 0.7766808  0.547157888 0.5842703 0.9999836
9      SEPT14P18  0.3997879   -2.123931210 2.6026067 -0.816078453 0.4144553 0.9999836
10        CICP27  0.2997882    0.583643570 2.7998824  0.208452886 0.8348754 0.9999836
11     LOC729737 13.9416828   -0.069849130 0.6598776 -0.105851644 0.9157001 0.9999836
12  LOC102725121  0.6743526   -0.410735268 2.5404899 -0.161675615 0.8715613 0.9999836
13  LOC102723897 28.8389335   -0.285978385 0.6044615 -0.473112616 0.6361328 0.9999836
14     MIR6859-2  0.0000000             NA        NA           NA        NA        NA
15  LOC107985721  0.0000000             NA        NA           NA        NA        NA
16     RPL23AP21  0.7961962   -0.718127195 1.5522753 -0.462628759 0.6436305 0.9999836
17  LOC112268260  0.1950265   -1.147573576 3.1607470 -0.363070368 0.7165523 0.9999836
18        OR4F29  0.0000000             NA        NA           NA        NA        NA
19         CICP7  0.0000000             NA        NA           NA        NA        NA
20  LOC100132287  0.1957826    0.006568044 3.1607470  0.002078004 0.9983420 0.9999836
>
```


log fold change
2
1
0
-1
-2
1e-01
1e+01
1e+03
1e+05
mean of normalized counts

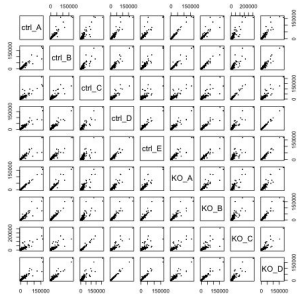
PCA, global variation seem consistent overall in all samples



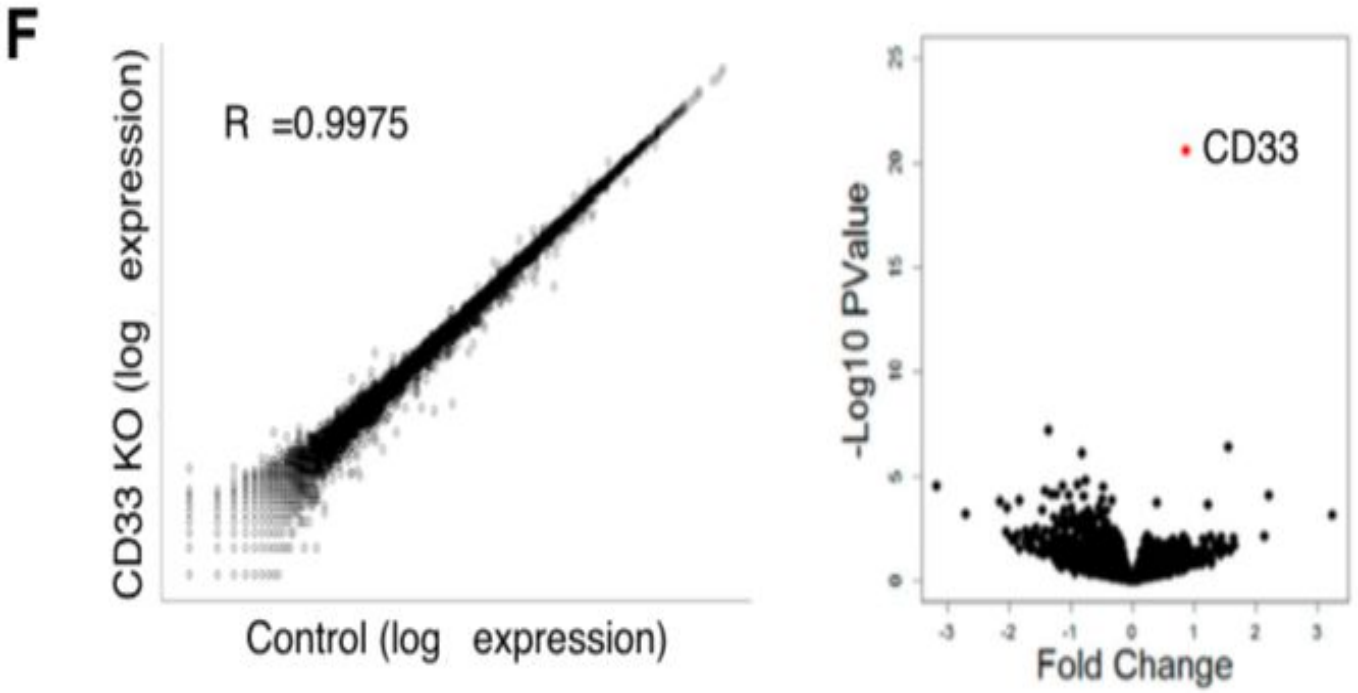

NS
Log2 FC
P
P & Log2 FC
10.0
7.5
5.0
2.5
0.0
−Log10 P
CD33
-2.5
0.0
2.5
5.0
7.5
Log2 fold change




```
> meancounts[1:15,]
         Geneid controlmean treatedmean
1        DDX11L1          0.0          0.0
2         WASH7P         41.6         57.8
3      MIR6859-1          2.0          3.0
4    MIR1302-2HG          0.0          0.0
5      MIR1302-2          0.0          0.0
6        FAM138A          0.0          0.0
7           OR4F5          0.0          0.0
8   LOC100996442         20.0         19.0
9      SEPT14P18          4.0          0.0
10        CICP27          1.0          2.0
11     LOC729737         70.2         61.6
12 LOC102725121          4.0          1.4
13 LOC102723897        127.2        100.0
14     MIR6859-2          0.0          0.0
15 LOC107985721          0.0          0.0
```

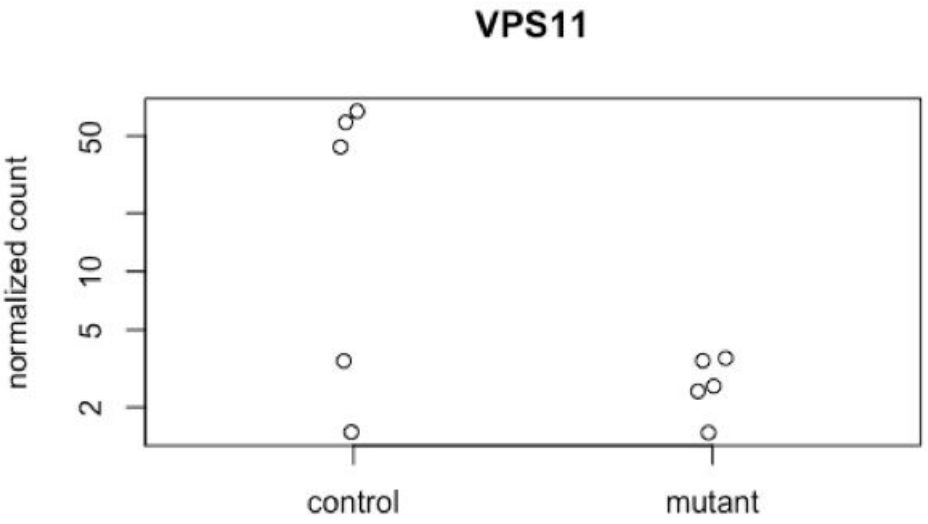


Plots from the RNA-seq analysis in the paper...

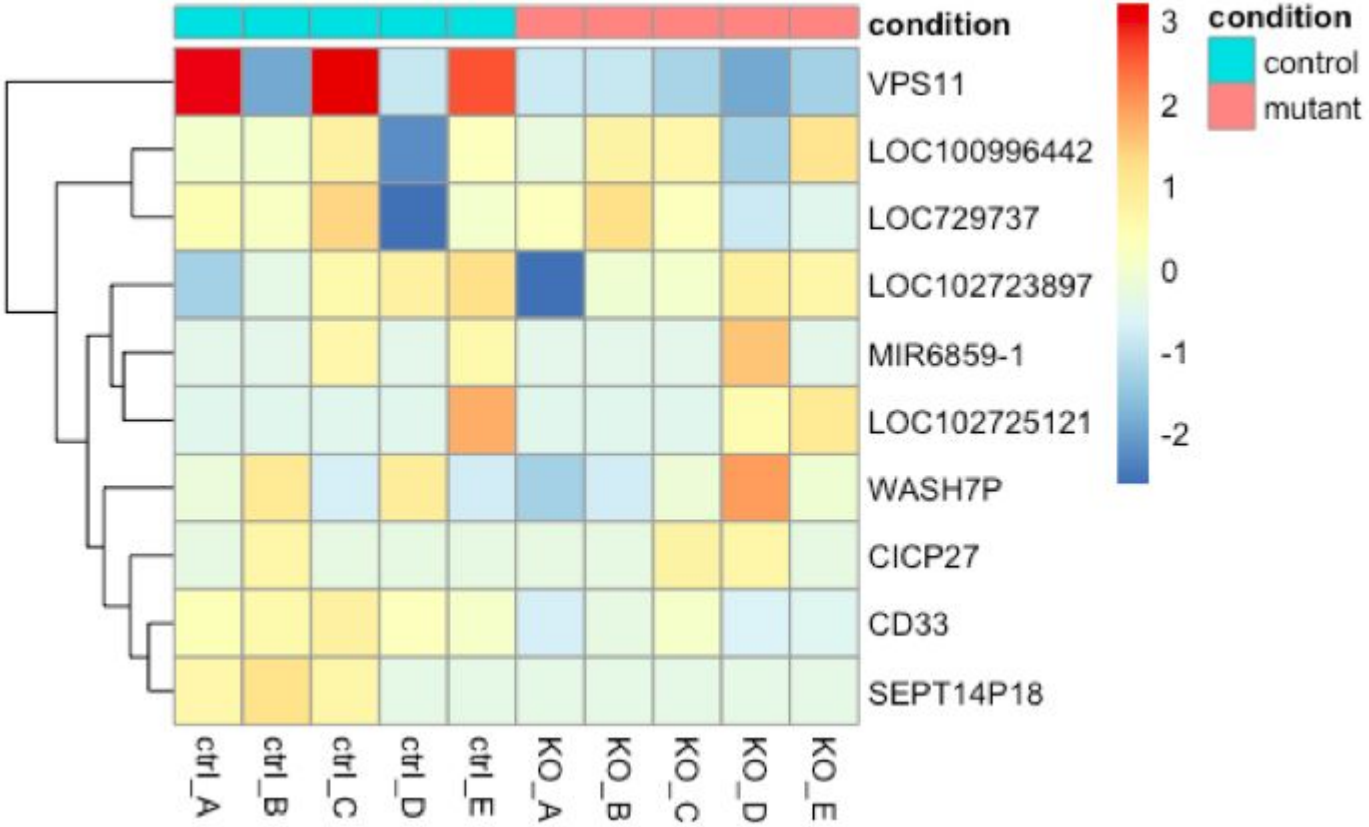


The most highly significantly expressed genes (ordered by adjusted p-value)

```
> head(result_model[order(result_model$padj), ], 5)
                row    baseMean log2FoldChange     lfcSE       stat       pvalue      padj
27094         VPS11  18.2836800     -3.9673495 0.9440287 -4.2025729 2.638982e-05 0.3791557
40680          CD33 970.0153619     -0.8384135 0.1993164 -4.2064459 2.594181e-05 0.3791557
2            WASH7P  11.9300825      0.2107954 0.7267117  0.2900674 7.717647e-01 0.9999836
3         MIR6859-1   0.4946898      0.5358558 2.6260940  0.2040505 8.383140e-01 0.9999836
8     LOC100996442   4.6769746      0.4249670 0.7766808  0.5471579 5.842703e-01 0.9999836
```



Heatmap of the 10 most differentially expressed genes



Thank you