Comparisons

Load results

```
gse5145 <- read.csv("results/GSE5145/array5145_genes.csv")
gse5145_gse <- read.csv("results/GSE5145/array5145_genesets.csv")
```

Microarray: GSE5145

```
gse145483 <- read.csv("results/GSE145483/array145483_genes.csv")
gse145483_gsa <- read.csv("results/GSE145483/array145483_genesets.csv")
```

Microarray: GSE145483

```
gse69284_2.5h <- read.csv("results/GSE69284/genes2.5h.csv")
gse69284_4h <- read.csv("results/GSE69284/genes4h.csv")
gse69284_24h <- read.csv("results/GSE69284/genes24h.csv")
gse69284_2.5h_gsa <- read.csv("results/GSE69284/gsa_1.csv")
gse69284_4h_gsa <- read.csv("results/GSE69284/gsa_2.csv")
gse69284_24h_gsa <- read.csv("results/GSE69284/gsa_3.csv")
```

RNAseq: GSE69284

```
gse189984_4 <- read.csv("results/GSE189984/genes4h.csv")
gse189984_8 <- read.csv("results/GSE189984/genes8h.csv")
gse189984_24 <- read.csv("results/GSE189984/genes24h.csv")
gse189984_48 <- read.csv("results/GSE189984/genes48h.csv")
```

RNAseq: GSE189984

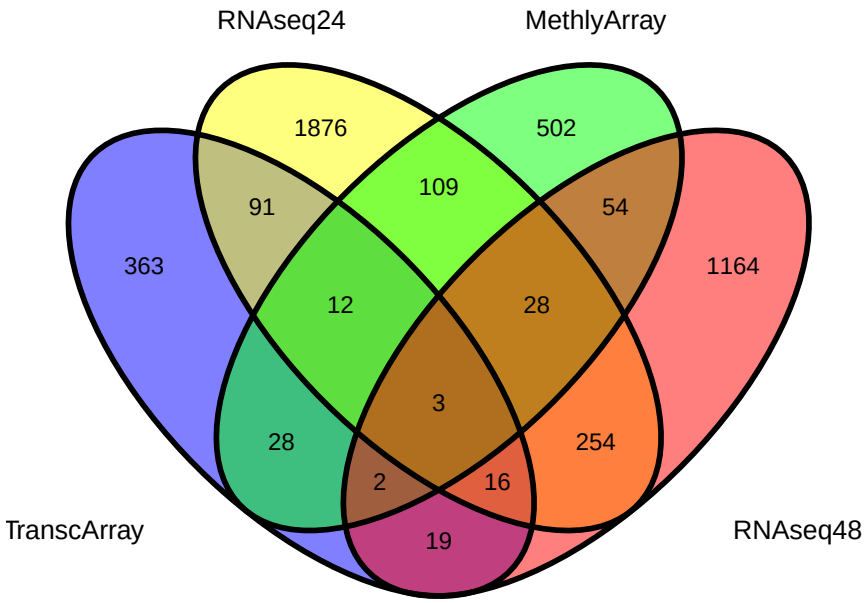
Common genes

```
library(ggvenn)

# filter na values
gse189984_48 <- gse189984_48[!is.na(gse189984_48$gene_symbol),]
gse69284_24h <- gse69284_24h[!is.na(gse69284_24h$gene_symbol),]
gse5145 <- gse5145[!is.na(gse5145$Gene.symbol),]
gse145483 <- gse145483[!is.na(gse145483$Gene),]

# List of items
x <- list(
  TranscArray = gse5145$Gene.symbol,
  RNAseq24 = gse69284_24h[gse69284_24h$FDR < 0.05,]$gene_symbol,
  MethlyArray = gse145483[gse145483$adj.P.Val < 0.05,]$Gene,
  RNAseq48 = gse189984_48[gse189984_48$FDR < 0.05,]$gene_symbol)

ggvenn(x, show_percentage = FALSE, text_size = 3, set_name_size = 3.4)
```



Venn diagram

There is a clear distinction in the number of common genes when comparing the experiments that were performed on lymphocytes (monocytes, THP-1, PBMC) against the experiment that was perfomed on smooth muscle cells (Microarray: GSE5145). Many common genes between GSE69284 (RNAseq) and GSE145483 (methylation array) can be observed, this can be explained by the fact that these experiments were both performed on monocyte-derived cell lines. This demonstrates the pleiotropic and cell type-dependent nature of Vit D-regulated gene expression.

Get the common genes:

```
Reduce(intersect, list(gse5145$Gene.symbol,
          gse69284_24h[gse69284_24h$FDR < 0.05,]$gene_symbol,
          gse145483[gse145483$adj.P.Val < 0.05,]$Gene,
          gse189984_48[gse189984_48$FDR < 0.05,]$gene_symbol))
```

All the data sets

```
## [1] "GRAMD4" "RBPJ"   "LRRC8A"
```

Only three highly significant genes that are present in all four data sets were found:

- The **GRAMD4** gene product is know as the Death-Inducing Protein (DIP) and is linked to the recruitment of E3 ligases and other factors that play an important role in (anti-tumor) immune responses and cell death.
- The **RBPJ** gene (also knows as CBF1) encodes the Recombination Signal Binding Protein For Immunoglobulin Kappa J Region, which is a downstream effector of the NOTCH signaling pathway. These highly conserved pathways have varied biological functions, including neurogenesis and embryonic

development. Dysregulated notch signaling is linked to many diseases, including leukemia and multiple sclerosis.

- The **LRRC8A** gene produces the Leucine-Rich Repeat-Containing Protein 8A, which is a subunit of the cell size- and proliferation-regulating Volume-Regulated Anion Channel (VRAC).

Between monocytic data sets Because the cell type is roughly the same, more convergence in the results is to be expected.

```
common <- Reduce(intersect, list(gse145483[gse145483$adj.P.Val < 0.05,]$Gene,
                        gse69284_24h[gse69284_24h$FDR < 0.05,]$gene_symbol))
```

Perform GSA on this common set of genes:

```
library(biomaRt)
library(limma)

GRCh38 <-  useEnsembl(biomart="genes", host="https://www.ensembl.org",
                      dataset="hsapiens_gene_ensembl")

# attributes
atr <- listAttributes(GRCh38)
idmap <- getBM(attributes = c('entrezgene_id',
                              'external_gene_name'), mart = GRCh38)

# map gene symbols to entez ids
idmap <- idmap[!duplicated(idmap$external_gene_name),]
idmap <- idmap[idmap$external_gene_name %in% common,]
entrez <- idmap$entrezgene_id

res <- kegga(de=entrez,
             species.KEGG="hsa")
res <- res[order(res$N),]

res$FDR.DE <- p.adjust(res$P.DE, n=nrow(res), method="BH")
```

Analyse results:

```
head(res[order(res$P.DE),], 10)
```

```
##                                                      Pathway   N DE
## path:hsa05417                        Lipid and atherosclerosis 215  8
## path:hsa05231                    Choline metabolism in cancer  98  5
## path:hsa04064                    NF-kappa B signaling pathway 104  5
## path:hsa05169                    Epstein-Barr virus infection 202  7
## path:hsa04920                 Adipocytokine signaling pathway  69  4
## path:hsa04015                           Rap1 signaling pathway 210  7
## path:hsa04662               B cell receptor signaling pathway  82  4
## path:hsa00604 Glycosphingolipid biosynthesis - ganglio series  15  2
## path:hsa04360                                  Axon guidance 182  6
## path:hsa04728                            Dopaminergic synapse 132  5
##                       P.DE    FDR.DE
## path:hsa05417 0.001647324 0.3469820
## path:hsa05231 0.003366513 0.3469820
## path:hsa04064 0.004344331 0.3469820
## path:hsa05169 0.004781366 0.3469820
## path:hsa04920 0.005535520 0.3469820
```

```
## path:hsa04015 0.005897711 0.3469820
## path:hsa04662 0.010122481 0.4122306
## path:hsa00604 0.010177310 0.4122306
## path:hsa04360 0.011215444 0.4122306
## path:hsa04728 0.011677920 0.4122306
```

Vitamin D, its metabolites, and the Vit D receptor (VDR) play an important role as regulating factors in a wide range of highly conserved and critical pathways.The functions of these pathways inlcude:

https://pubmed.ncbi.nlm.nih.gov/25144342/
- The regulation of calcium absorption and resorption in the intestine, kidneys and bonematrix. This explains the observed link between Vit D and atherosclerosis, which is a disease that results in the infection and calcification of the artery wall. The infections present in atherosclerosis can be linked to increased immunologic responses as a result of Adipocytokine signaling, in which cytokines are secreted by adipose tissue.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3270541/
- The regulation of immune functions, specifically the innate anti-viral response and cell death. This explains the observed link between Vit D and the NF-kappa B signaling / Epstein-Barr virus infection / B cell receptor signaling. Though it should be mentioned that these findings are most likely greatly dependent on the fact that these experiments were performed on monocytic cells which have an important role in the immune response on their own. Rap1 (a small GTPase) is a general signal transduction factor that will most likely guide the monocyte in its immunogenic endavors

https://www.sciencedirect.com/science/article/pii/S0306453009002145?via%3Dihub
- Vit D and its metabolites have shown to be linked to the functioning of the central nervous system, the development of of the fetal brain, and neurogenesis. Insuficiency has also been linked to many diseases associated with the CNS, including multiple sclerosis, Alzheimer's and Parkinson's disease, seasonal affective disorder and schizophrenia.

https://pubmed.ncbi.nlm.nih.gov/33838984/
- Choline, as a precursor of key lipids (including Glycosphingolipid) has also been linked to embryonic development and neurogenesis. Cross-regulation between these pathways can be a possible explanation for the remaining findings shown in this table.

The `ggvenn` package is described in:

> Yan L (2021). *ggvenn: Draw Venn Diagram by 'ggplot2'*. R package version 0.1.9, https://CRAN.R-project.org/package=ggvenn.

The `limma` package is described in:

> Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43(7), e47.

The `biomaRt` package is described in:

> Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Steffen Durinck, Paul T. Spellman, Ewan Birney and Wolfgang Huber, *Nature Protocols* 4, 1184-1191 (2009).
>
> BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Steffen Durinck, Yves Moreau, Arek Kasprzyk, Sean Davis, Bart De Moor, Alvis Brazma and Wolfgang Huber, Bioinformatics 21, 3439-3440 (2005).