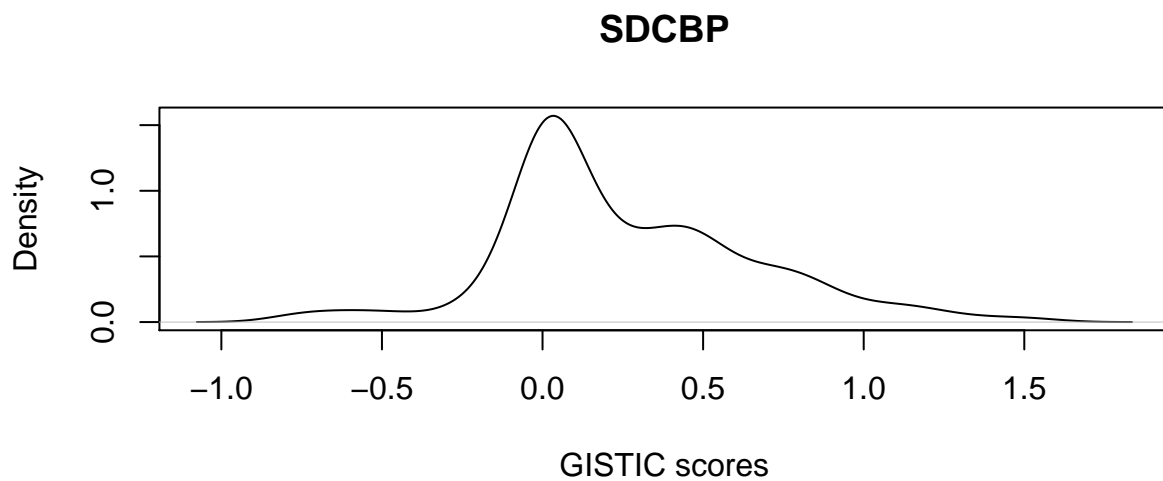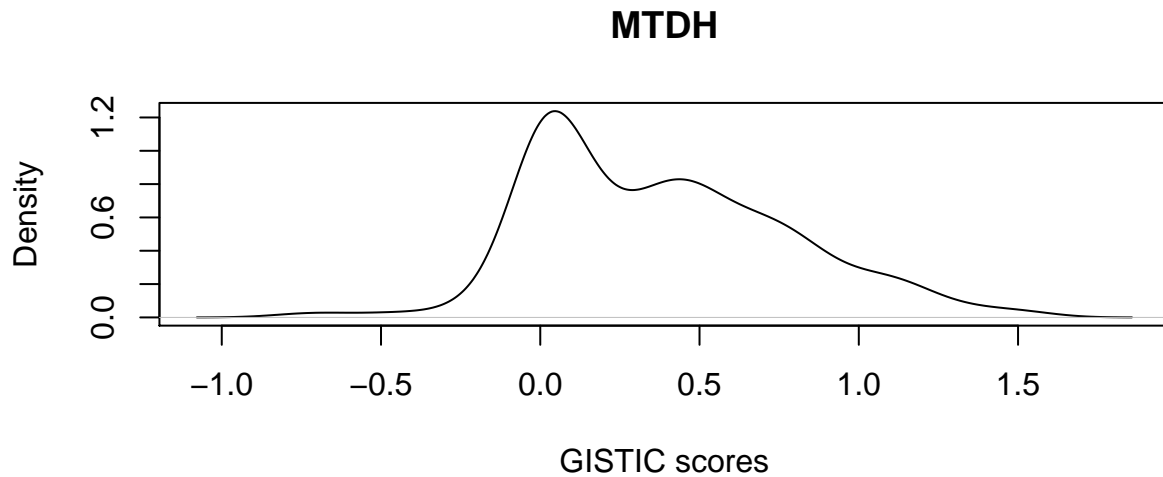# CNV analysis

Mikhail Dozmorov

2019-01-04

```r
# Cancer type
cancer = "LIHC"
# Gene(s) of interest
selected_genes = c("MTDH", "SDCBP")  # Can be multiple
```

Density plots of GISTIC scores (discretized copy number variation to values -2, -1, 0, 1, 2), see FAQ. The majority of the samples should be normal (high peak around zero).

MTDH

SDCBP

Heatmap of GISTIC scores across samples (X axis) in the corresponding gene(s). Blue/red insicate dele-
tion/amplification, respectively.
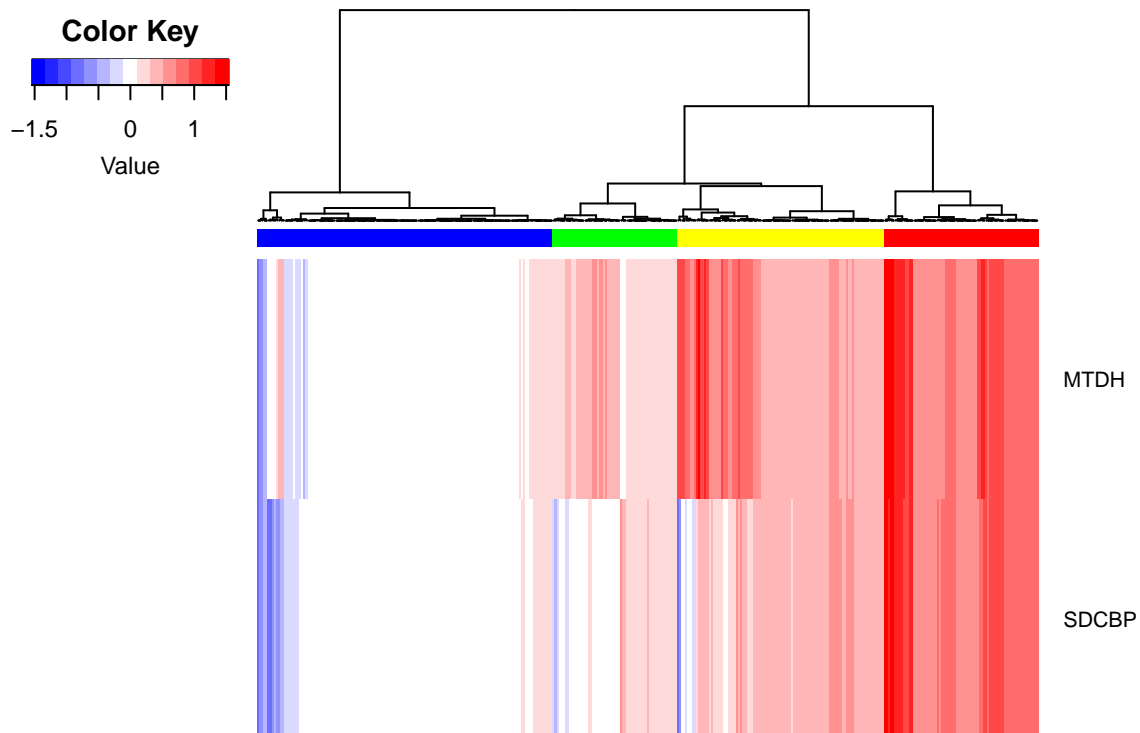
!!! Critical. Define which clusters to compare.

```
# Red cluster vs. other ind_up <- h_clust_color == 'red' ind_lo <- !(ind_up)
# Red cluster vs. blue ind_up <- h_clust_color == 'red' ind_lo <-
# h_clust_color == 'blue' Red and Yellow clusters vs. blue
ind_up <- names(h_clust_color)[h_clust_color == "red" | h_clust_color == "yellow"]
ind_lo <- names(h_clust_color)[h_clust_color == "blue"]
```

Survival analysis

Survival comparison between the selected clusters. Cluster 1 corresponds to samples defined by "ind_up", cluster 2 - to "ind_dn".
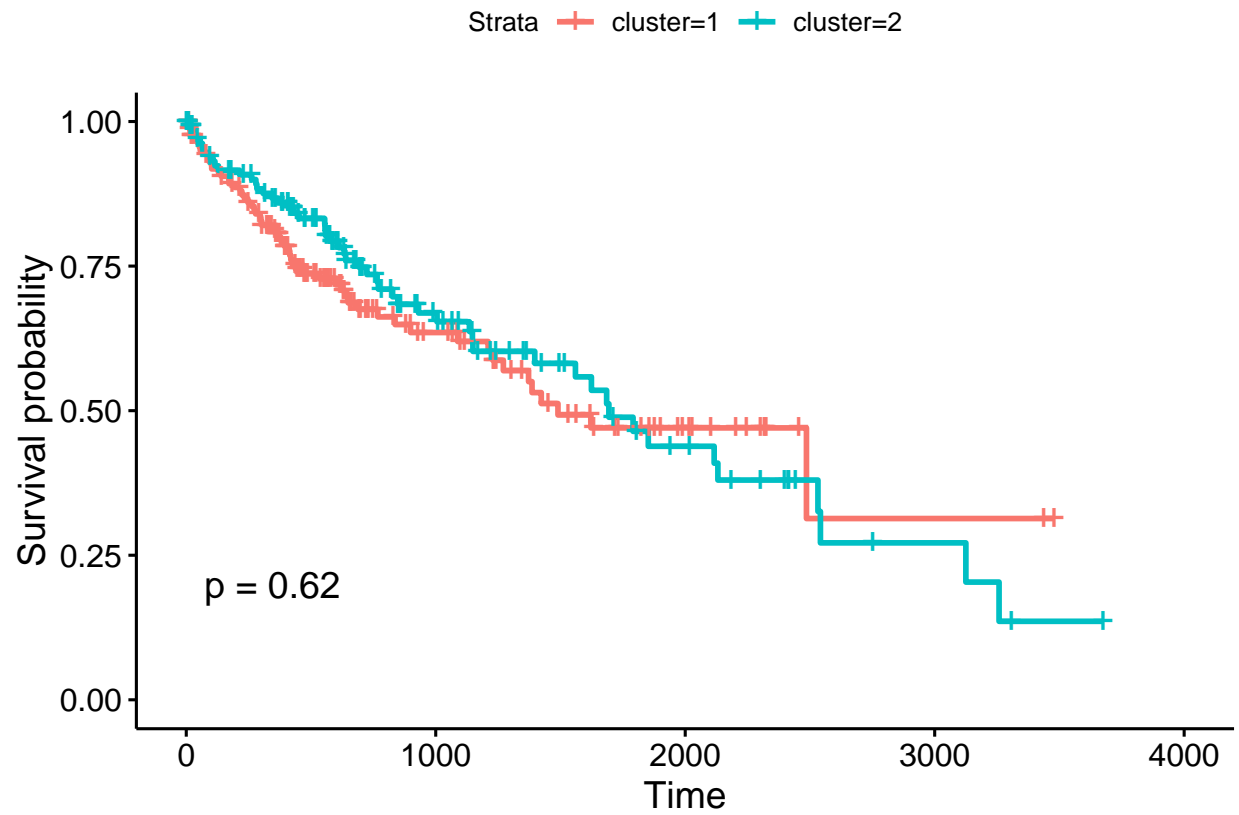
Differential expression analysis

We split LIHC cohort into 167 x 139 groups manually separated previously. We have a total of 7019 differentially expressed genes at FDR corrected p-value 0.05. 2492 are upregulated, 4527 are downregulated.

Top 50 the most differentially expressed genes are shown
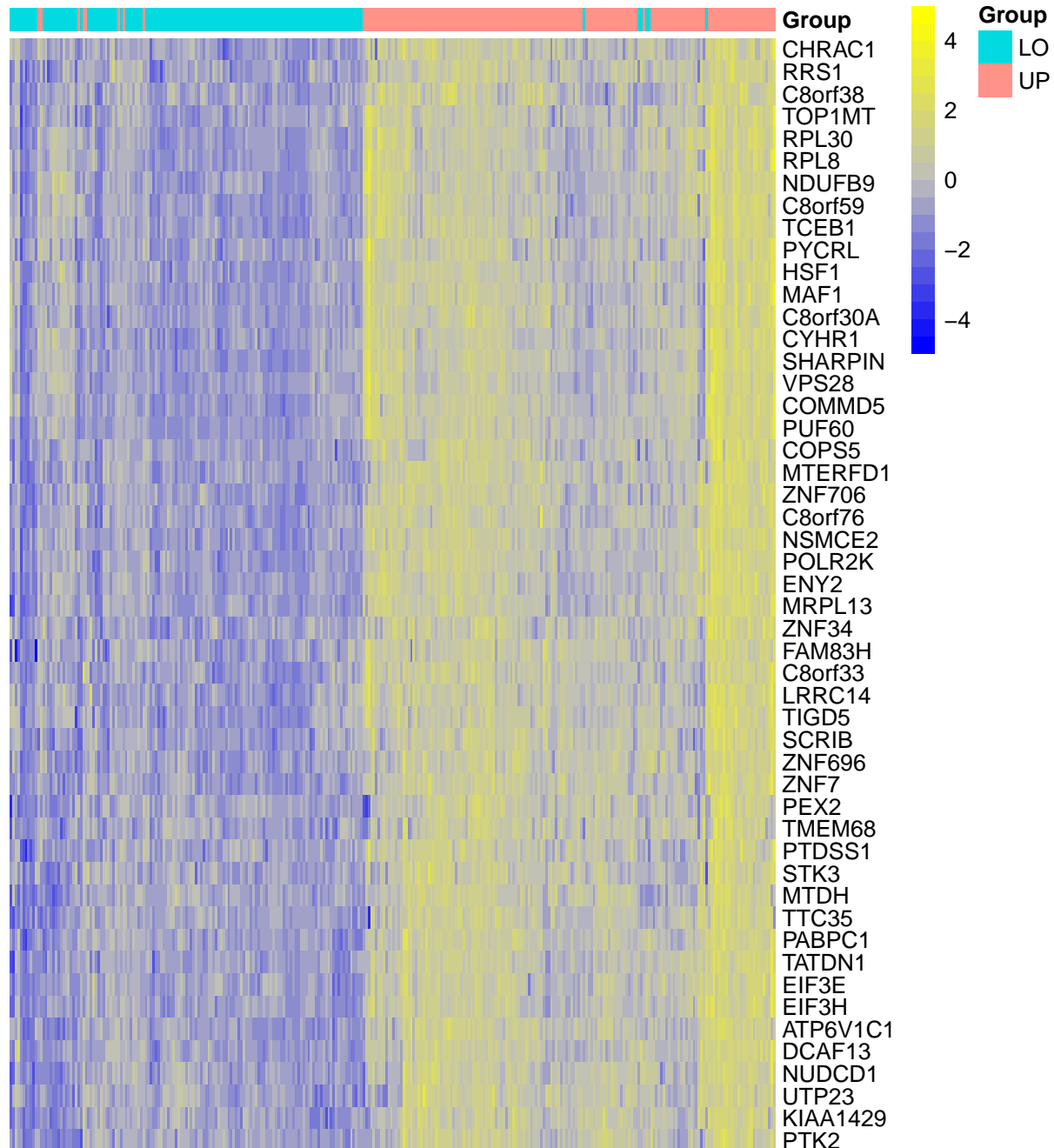
Results are stored in the Excel file results/LIHC_MTDH-SDCBP_DEGs_0.01.xlsx

- Legend for gene lists: "Gene" - gene annotations; "logFC" - log fold change; "AveExpr" - average expression, log2; "t" - t-statistics; "P.Val"/"adj.P.Val" - non-/FDR-adjusted p-value, "B" - another statistics.

	logFC	AveExpr	t	P.Value	adj.P.Val	B
MTERFD1	1.096	8.877	18.67	2.05e-52	2.564e-48	108.3
NSMCE2	1.287	8.882	18.65	2.501e-52	2.564e-48	108.1
POLR2K	1.044	10.41	18.49	9.966e-52	6.81e-48	106.7
ZNF706	1.085	10.76	18.03	5.614e-50	2.343e-46	102.8
ATP6V1C1	1.13	10.94	18.03	5.715e-50	2.343e-46	102.8
PUF60	1.124	12.38	17.61	2.235e-48	7.636e-45	99.14
ZNF7	0.9114	8.758	17.17	1.082e-46	2.864e-43	95.32
C8orf33	1.122	10.46	17.16	1.118e-46	2.864e-43	95.28
PABPC1	1.349	14.35	16.72	5.217e-45	1.188e-41	91.5
TATDN1	1.107	8.691	16.68	7.557e-45	1.527e-41	91.13
KIAA1429	0.8398	10.39	16.67	8.193e-45	1.527e-41	91.05
C8orf76	1.057	9.282	16.64	1.046e-44	1.787e-41	90.81
COPS5	0.8384	10.72	16.37	1.197e-43	1.888e-40	88.41
TCEB1	1.092	10.07	16.25	3.392e-43	4.967e-40	87.38
SHARPIN	1.117	10.94	16.1	1.243e-42	1.699e-39	86.1

Functional enrichment analysis

Up- and downregulated genes are tested for functional enrichment jointly. . FDR cutoff of the significant enrichments - 0.05. Top 15 genes shown.

KEGG pathway enrichment analysis

Legend: "database" - source of functional annotations, "category" - name of functional annotation, "pval" - unadjusted enrichment p-value, "qval" - FDR-adjusted p-value, "genes" - comma-separated differentially expressed genes enriched in a corresponding functional category.

```
[1] "KEGG pathway run on 2000 genes without distinguishing them by directionality."
[1] "Running KEGG_2016 analysis"
```

KEGG pathway GSEA analysis

Legend: "ID", "Description" - KEGG pathway ID/description, respectively; "NES" - normalized enrichment score; "pvalue", "p.adjust" - raw and FDR-adjusted p-values, respectively; "core_enrichment" - genes enriched in the corresponding pathway.

A total of 275 KEGG pathways were detected as significantly affected at FDR 0.05. Top 15 shown.

database	category	pval	qval
KEGG_2016	Ribosome_Homo sapiens_hsa03010	4.617e-35	1.270e-32
KEGG_2016	Circadian rhythm_Homo sapiens_hsa04710	4.488e-04	4.279e-02
KEGG_2016	Pyrimidine metabolism_Homo sapiens_hsa00240	6.128e-04	4.279e-02
KEGG_2016	Purine metabolism_Homo sapiens_hsa00230	6.224e-04	4.279e-02
KEGG_2016	Ribosome biogenesis in eukaryotes_Homo sapiens_hsa03008	1.107e-03	5.071e-02
KEGG_2016	N-Glycan biosynthesis_Homo sapiens_hsa00510	8.038e-04	4.421e-02
KEGG_2016	ErbB signaling pathway_Homo sapiens_hsa04012	5.194e-03	2.040e-01
KEGG_2016	Prostate cancer_Homo sapiens_hsa05215	6.576e-03	2.136e-01
KEGG_2016	Neurotrophin signaling pathway_Homo sapiens_hsa04722	7.766e-03	2.136e-01

database	category	pval	qval
KEGG_2016	RNA degradation_Homo sapiens_hsa03018	8.595e-03	2.149e-01
KEGG_2016	p53 signaling pathway_Homo sapiens_hsa04115	7.600e-03	2.136e-01
KEGG_2016	Endometrial cancer_Homo sapiens_hsa05213	1.246e-02	2.637e-01
KEGG_2016	Epithelial cell signaling in Helicobacter pylori infection_Homo sapiens_hsa05120	1.613e-02	3.168e-01
KEGG_2016	RNA polymerase_Homo sapiens_hsa03020	1.162e-02	2.637e-01
KEGG_2016	Renal cell carcinoma_Homo sapiens_hsa05211	2.923e-02	5.038e-01

Selected pathway

Red/Green - up/downregulated genes in upper vs. lower MTDH, SDCBP expressing samples. Gray - marginal fold change, yet significant. White - gene is not differentially expressed