MCF10A RNA-Seq Analysis

Sebastian Kurscheid

24 November 2016

Contents

1 Introduction

1.1 Purpose of the analysis

- Determine differential expression of a set of pre-defined genes (EMT-marker genes) in human breast epithelial cell lines MCF10A/MCF10Aca1a
- Following conditions were used
  - **MCF10A WT** vs **MCF10A TGFb-treated** (induction of EMT)
  - **MCF10A WT** vs **MCF10A H2A.Z** knock-down (induction of EMT/effects of siRNAi mediated gene silencing)

Three biological replicates were used for each condition and libraries were sequenced on an Illumina Next-Seq 500 sequencer using Illumina TruSeq protocol, 76bp PE. The sequencing reads were pre-processed and aligned to the human reference genome hg19-based Ensembl Version 75 transcriptome annotation using STAR. Actual transcript quantification was performed using kallisto against and index built from Ensembl 75 (all cDNAs + ncRNAs). The sequencing data processing pipeline is implemented using snakemake https://bitbucket.org/snakemake/ and can be found at https://github.com/JCSMR-Tremethick-Lab/Breast.

Important: This is a supervised analysis, i.e. we predominantely investigated the changes in set of gene we defined prior to the experiment. The next step will be to conduct a fully unsupervised analysis and firstly identify those genes which are dysregulated, and link these back to biological pathways. This will then be a good starting point to analyse the H2A.Z ChIP-Seq results.

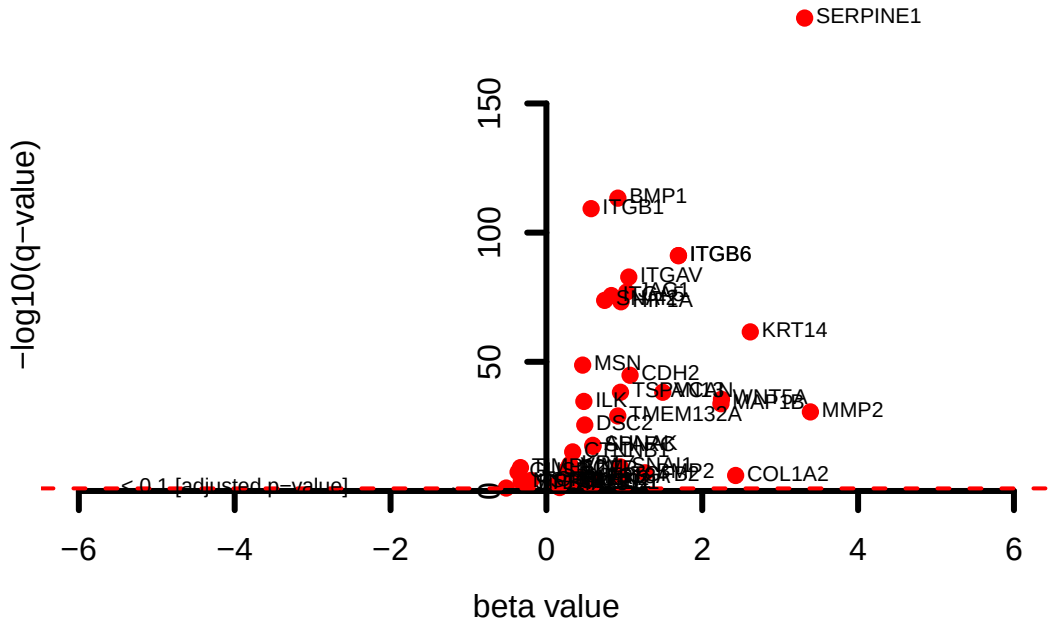
```
## [1] "MCF10A_vs_shZ"  "MCF10A_vs_TGFb"
```

2 Results

2.1 Analysis of EMT genes as defined by pPCR array

2.1.1 TGF-b treated MCF10A vs WT

Volcano plots of the two datasets, first TGFb-treated MCF10A cells vs MCF10A wt.



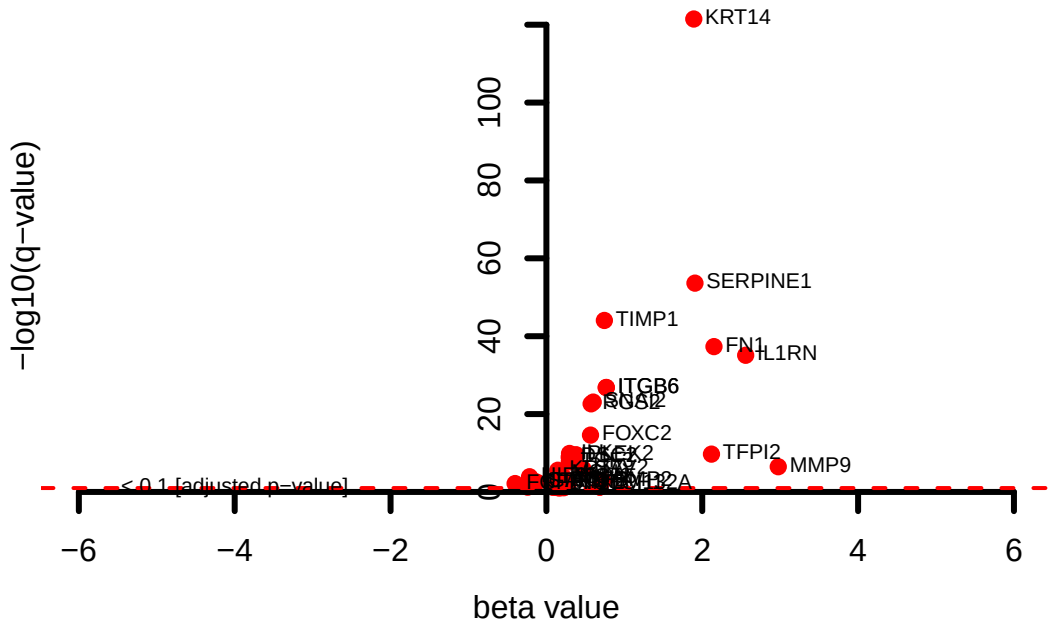
EMT genes differentially expressed in MCF10A cells upon TGFb treatment, sorted by log2fold-change (descending), and q-value (adjusted p-value).

	pval	qval	b	FC_estimated	external_gene_name
76	0.1238158	0.2129585	-3.0482981	-4.3977645	TMEFF1
77	0.1848172	0.2947829	-1.1943260	-1.7230482	OCLN
73	0.0354378	0.0738185	-0.5152484	-0.7433464	OCLN
30	0.2050421	0.3195367	-0.5001259	-0.7215292	TGFB3
65	0.0000000	0.0000001	-0.3633242	-0.5241660	GUSB
16	0.0000000	0.0000000	-0.3329524	-0.4803488	TIMP1
53	0.0000713	0.0002537	-0.3237460	-0.4670667	MST1R
54	0.0003690	0.0011838	-0.2937684	-0.4238181	STEAP1
22	NA	NA	-0.2758877	-0.3980218	GAPDH
72	0.0690325	0.1307548	-0.2750984	-0.3968831	TCF4
56	0.0002447	0.0008036	-0.2548825	-0.3677177	HPRT1
31	0.0926010	0.1672966	-0.2535487	-0.3657934	TWIST1
7	0.0000217	0.0000828	-0.1985169	-0.2863994	TCF3
23	0.2745154	0.4024652	-0.1572148	-0.2268130	PTP4A1

	pval	qval	b	FC_estimated	external_gene_name
6	0.5671969	0.6928627	-0.1550002	-0.2236181	ERBB3
71	0.5537782	0.6808669	-0.0955816	-0.1378950	MITF
64	0.7746469	0.8537572	-0.0923592	-0.1332462	ZEB2
75	0.5320018	0.6618231	-0.0922878	-0.1331431	COL5A2
46	0.3328348	0.4657680	-0.0849659	-0.1225798	NOTCH1
68	0.1007721	0.1795540	-0.0820194	-0.1183290	SMAD2
43	0.0787121	0.1460423	-0.0745919	-0.1076133	AKT1
8	0.3142945	0.4465178	-0.0625798	-0.0902836	GSK3B
69	0.9172455	0.9482366	-0.0159277	-0.0229787	FOXC2
40	0.9968169	0.9983707	0.0016867	0.0024333	IL1RN
18	0.9445484	0.9671332	0.0210207	0.0303265	TFPI2
39	0.5173825	0.6490008	0.0213116	0.0307461	RAC1
67	0.3421965	0.4761961	0.0518559	0.0748122	FKBP2
62	0.0019385	0.0055067	0.1172485	0.1691538	STAT3
17	0.1082018	0.1906169	0.1178858	0.1700732	TGFB1
42	0.4809476	0.6154037	0.1269858	0.1832018	IGFBP4
19	0.0213388	0.0472292	0.1695049	0.2445439	CAV2
50	0.2272233	0.3463181	0.2059413	0.2971105	FZD7
5	0.0000613	0.0002205	0.2254033	0.3251882	CDH1
44	0.0000000	0.0000000	0.2656055	0.3831877	EGFR
35	0.0002007	0.0006683	0.2675023	0.3859243	GNG11
63	0.0000578	0.0002084	0.2793142	0.4029652	PTK2
59	0.0000000	0.0000000	0.2832755	0.4086801	B2M
3	0.0000000	0.0000000	0.2919319	0.4211686	VIM
38	0.0000000	0.0000000	0.3018261	0.4354430	KRT7
51	0.0000051	0.0000212	0.3028277	0.4368880	F11R
14	0.4291965	0.5659161	0.3066618	0.4424195	MMP9
12	0.0000089	0.0000360	0.3226363	0.4654657	PLEK2
57	0.2594861	0.3846473	0.3300349	0.4761397	NUDT13
60	0.0000000	0.0000000	0.3375224	0.4869419	CTNNB1
74	0.0000014	0.0000061	0.4101444	0.5917133	VPS13A
29	0.0002919	0.0009494	0.4533651	0.6540675	RGS2
45	0.0000000	0.0000000	0.4652422	0.6712027	MSN
58	0.0000000	0.0000000	0.4806489	0.6934297	ILK
37	0.0000000	0.0000000	0.4923530	0.7103152	DSC2
49	0.0000000	0.0000000	0.5739156	0.8279852	ITGB1
33	0.0000000	0.0000000	0.5956615	0.8593578	AHNAK
24	0.0000000	0.0000000	0.5977801	0.8624144	SPARC
47	0.0004271	0.0013527	0.6156731	0.8882286	ZEB1
48	0.0004271	0.0013527	0.6156731	0.8882286	ZEB1
10	0.1568742	0.2581502	0.7363304	1.0623003	ESR1
2	0.0000000	0.0000000	0.7475607	1.0785021	SNAI2
52	0.0000000	0.0000000	0.8339793	1.2031778	ITGA5
1	0.0000000	0.0000000	0.9162278	1.3218373	TMEM132A
61	0.0000000	0.0000000	0.9174853	1.3236515	BMP1
11	0.0000005	0.0000022	0.9387239	1.3542923	TGFB2
32	0.0000000	0.0000000	0.9461712	1.3650365	SNAI1
21	0.0000000	0.0000000	0.9504967	1.3712769	TSPAN13
13	0.0000000	0.0000000	0.9568108	1.3803861	HIF1A
15	0.0000000	0.0000000	1.0303682	1.4865071	JAG1
41	0.0000000	0.0000000	1.0571139	1.5250930	ITGAV
66	0.0000000	0.0000000	1.0733574	1.5485274	CDH2

	pval	qval	b	FC_estimated	external_gene_name
34	0.0000000	0.0000001	1.2664850	1.8271516	BMP2
4	0.0000000	0.0000000	1.4936131	2.1548282	VCAN
26	0.0000000	0.0000000	1.6947292	2.4449774	ITGB6
27	0.0000000	0.0000000	1.6947292	2.4449774	ITGB6
36	0.0000000	0.0000000	2.2398408	3.2314071	MAP1B
25	0.0000000	0.0000000	2.2464772	3.2409815	WNT5A
55	0.0000002	0.0000010	2.4285027	3.5035888	COL1A2
70	0.0000000	0.0000000	2.6163960	3.7746616	KRT14
20	0.0000000	0.0000000	3.3140527	4.7811673	SERPINE1
9	0.0000000	0.0000000	3.3877964	4.8875571	MMP2
28	0.0000000	0.0000000	4.4350789	6.3984664	FN1

2.1.2 shZ-knockdown in MCF10A vs WT



EMT genes differentially expressed in MCF10A cells upon H2AZ shRNA-mediated knock-down, sorted by log2fold-change (descending), and q-value (adjusted p-value).

	pval	qval	b	FC_estimated	external_gene_name
78	0.4549444	0.7150529	-1.1532030	-1.6637202	TMEFF1
43	0.0007403	0.0070781	-0.4018393	-0.5797316	FGFBP1
9	0.5291817	0.7689691	-0.3556788	-0.5131360	WNT11

	pval	qval	b	FC_estimated	external_gene_name
57	0.0106615	0.0589290	-0.2402392	-0.3465919	STEAP1
14	0.0000069	0.0001257	-0.2163125	-0.3120730	HIF1A
24	0.0371184	0.1489500	-0.2010768	-0.2900925	PTP4A1
33	0.1250804	0.3396231	-0.1885657	-0.2720428	TWIST1
74	0.1552760	0.3890450	-0.1734231	-0.2501967	TCF4
45	0.2008255	0.4551248	-0.1693809	-0.2443650	IGFBP4
64	0.0001818	0.0021792	-0.1479066	-0.2133841	STAT3
23	0.0002954	0.0032575	-0.1211768	-0.1748212	GAPDH
68	0.3641231	0.6373131	-0.1159052	-0.1672158	CDH2
25	0.2216042	0.4833294	-0.1001984	-0.1445557	SPARC
34	0.6788398	0.8550303	-0.0873783	-0.1260602	SNAI1
22	0.5205648	0.7638914	-0.0474862	-0.0685081	TSPAN13
77	0.7229327	0.8783603	-0.0417048	-0.0601673	COL5A2
46	0.6120269	0.8196048	-0.0241148	-0.0347903	AKT1
6	0.8851199	0.9559011	-0.0218238	-0.0314851	ERBB3
27	0.9431846	0.9787586	-0.0167579	-0.0241766	WNT5A
54	0.8568314	0.9460161	-0.0111842	-0.0161354	F11R
58	0.8414267	0.9385741	-0.0102252	-0.0147519	HPRT1
70	0.8743510	0.9522457	0.0072985	0.0105296	SMAD2
59	0.9261572	0.9717162	0.0238928	0.0344700	NUDT13
67	0.6247003	0.8257680	0.0267522	0.0385952	GUSB
37	0.6910939	0.8618255	0.0267754	0.0386287	GNG11
69	0.4816359	0.7360906	0.0485155	0.0699931	FKBP2
8	0.4201780	0.6877067	0.0510901	0.0737075	GSK3B
62	0.0642153	0.2186612	0.0628363	0.0906536	CTNNB1
52	0.1380117	0.3611313	0.0755248	0.1089593	ITGB1
3	0.0121443	0.0651116	0.0836809	0.1207259	VIM
61	0.0555951	0.1978339	0.0936810	0.1351531	B2M
41	0.0229153	0.1050470	0.1066081	0.1538030	RAC1
16	0.0131342	0.0690993	0.1255551	0.1811378	JAG1
4	0.0270051	0.1182259	0.1459597	0.2105753	VCAN
40	0.0000001	0.0000027	0.1463857	0.2111899	KRT7
5	0.0247985	0.1114055	0.1579039	0.2278072	CDH1
47	0.0001261	0.0016025	0.1623629	0.2342402	EGFR
56	0.0213672	0.0996454	0.1657639	0.2391467	MST1R
48	0.0000008	0.0000180	0.1728249	0.2493337	MSN
75	0.2726765	0.5436150	0.1792951	0.2586681	OCLN
65	0.0005299	0.0053777	0.1829166	0.2638929	PTK2
63	0.0000096	0.0001693	0.2015841	0.2908243	BMP1
73	0.2469829	0.5130110	0.2158123	0.3113514	MITF
18	0.0019794	0.0157311	0.2167953	0.3127696	TGFB1
76	0.0166553	0.0832576	0.2214058	0.3194211	VPS13A
55	0.0000108	0.0001871	0.2226421	0.3212047	ITGA5
53	0.0632819	0.2167049	0.2444590	0.3526798	FZD7
79	0.7343917	0.8856792	0.2446826	0.3530024	OCLN
38	0.3963145	0.6678061	0.2449096	0.3533298	MAP1B
44	0.0085655	0.0502244	0.2450770	0.3535714	ITGAV
49	0.0011146	0.0098606	0.2464870	0.3556055	NOTCH1
10	0.4023950	0.6728736	0.2477572	0.3574381	MMP2
35	0.0056497	0.0365302	0.2726528	0.3933548	AHNAK
1	0.0009462	0.0086456	0.2776250	0.4005282	TMEM132A
39	0.0000000	0.0000000	0.2928653	0.4225153	DSC2

	pval	qval	b	FC_estimated	external_gene_name
7	0.0000000	0.0000000	0.2938385	0.4239193	TCF3
32	0.2456404	0.5114273	0.2961117	0.4271989	TGFB3
60	0.0000000	0.0000000	0.2968793	0.4283063	ILK
13	0.0000000	0.0000000	0.3800914	0.5483560	PLEK2
20	0.0000000	0.0000011	0.4687444	0.6762552	CAV2
11	0.4241191	0.6912283	0.4826147	0.6962659	ESR1
66	0.0328612	0.1357220	0.5197709	0.7498709	ZEB2
71	0.0000000	0.0000000	0.5658477	0.8163456	FOXC2
31	0.0000000	0.0000000	0.5739157	0.8279853	RGS2
2	0.0000000	0.0000000	0.6006086	0.8664950	SNAI2
12	0.0462497	0.1737187	0.6351243	0.9162907	TGFB2
50	0.0096693	0.0548116	0.6854635	0.9889148	ZEB1
51	0.0096693	0.0548116	0.6854635	0.9889148	ZEB1
36	0.0001178	0.0015199	0.7243944	1.0450802	BMP2
17	0.0000000	0.0000000	0.7437961	1.0730710	TIMP1
28	0.0000000	0.0000000	0.7681306	1.1081782	ITGB6
29	0.0000000	0.0000000	0.7681306	1.1081782	ITGB6
26	0.1408316	0.3660486	0.9618429	1.3876460	PDGFRB
72	0.0000000	0.0000000	1.8915071	2.7288679	KRT14
21	0.0000000	0.0000000	1.9063515	2.7502839	SERPINE1
19	0.0000000	0.0000000	2.1181542	3.0558505	TFPI2
30	0.0000000	0.0000000	2.1503737	3.1023335	FN1
42	0.0000000	0.0000000	2.5573168	3.6894282	IL1RN
15	0.0000000	0.0000003	2.9755682	4.2928375	MMP9

2.2 Global analysis of differentially expressed genes

For the purpose of improved readibilit, I restrict the number of genes to the top 50 up and top 50 down-regulated genes.

2.2.1 TGF-b treated vs WT

2.2.1.1 Top 50 up-regulated genes

	pval	qval	b	estimated_FC	external_gene_name	description
12	0	0	4.435079	6.398466	FN1	fibronectin 1 [Source:HGNC Symbol;Acc:3778]
44	0	0	4.247780	6.128251	TGM2	transglutaminase 2 [Source:HGNC Symbol;Acc:1177
30	0	0	2.840006	4.097263	TAGLN	transgelin [Source:HGNC Symbol;Acc:11553]
1	0	0	2.790837	4.026327	IL32	interleukin 32 [Source:HGNC Symbol;Acc:16830]
24	0	0	2.430058	3.505833	SLC46A3	solute carrier family 46, member 3 [Source:HGNC Sy
9	0	0	2.387148	3.443927	MYL9	myosin, light chain 9, regulatory [Source:HGNC Sym
4	0	0	2.328380	3.359142	LAMA3	laminin, alpha 3 [Source:HGNC Symbol;Acc:6483]
8	0	0	2.287099	3.299587	CCDC80	coiled-coil domain containing 80 [Source:HGNC Sym
33	0	0	2.258921	3.258935	PDPN	podoplanin [Source:HGNC Symbol;Acc:29602]
32	0	0	2.215356	3.196083	KCNJ15	potassium inwardly-rectifying channel, subfamily J,
16	0	0	2.154067	3.107661	TGFBI	transforming growth factor, beta-induced, 68kDa [So
40	0	0	2.148461	3.099575	AHNAK2	AHNAK nucleoprotein 2 [Source:HGNC Symbol;Acc
15	0	0	2.137705	3.084056	GPAM	glycerol-3-phosphate acyltransferase, mitochondrial [
43	0	0	2.080737	3.001868	SELM	Selenoprotein M [Source:UniProtKB/Swiss-Prot;Acc
36	0	0	2.034350	2.934947	TUBA1A	tubulin, alpha 1a [Source:HGNC Symbol;Acc:20766]

	pval	qval	b	estimated_FC	external_gene_name	description
10	0	0	1.967176	2.838035	IL1A	interleukin 1, alpha [Source:HGNC Symbol;Acc:5991
5	0	0	1.965088	2.835023	LAMC2	laminin, gamma 2 [Source:HGNC Symbol;Acc:6493]
38	0	0	1.896069	2.735450	ZBED2	zinc finger, BED-type containing 2 [Source:HGNC Sy
25	0	0	1.850017	2.669011	TPM1	tropomyosin 1 (alpha) [Source:HGNC Symbol;Acc:12
42	0	0	1.744143	2.516266	LAMB3	laminin, beta 3 [Source:HGNC Symbol;Acc:6490]
11	0	0	1.721214	2.483187	TP53I3	tumor protein p53 inducible protein 3 [Source:HGNC
28	0	0	1.700463	2.453250	TUFT1	tuftelin 1 [Source:HGNC Symbol;Acc:12422]
17	0	0	1.672445	2.412828	PMEPA1	prostate transmembrane protein, androgen induced 1
41	0	0	1.651589	2.382740	COL4A1	collagen, type IV, alpha 1 [Source:HGNC Symbol;Ac
50	0	0	1.547736	2.232911	CD24	CD24 molecule [Source:HGNC Symbol;Acc:1645]
29	0	0	1.547143	2.232055	YIPF5	Yip1 domain family, member 5 [Source:HGNC Symb
3	0	0	1.485213	2.142710	ROS1	c-ros oncogene 1 , receptor tyrosine kinase [Source:H
46	0	0	1.434930	2.070167	IER3	immediate early response 3 [Source:HGNC Symbol;A
47	0	0	1.434930	2.070167	IER3	immediate early response 3 [Source:HGNC Symbol;A
48	0	0	1.434930	2.070167	IER3	immediate early response 3 [Source:HGNC Symbol;A
49	0	0	1.434930	2.070167	IER3	immediate early response 3 [Source:HGNC Symbol;A
19	0	0	1.433151	2.067599	PALLD	palladin, cytoskeletal associated protein [Source:HGN
26	0	0	1.409544	2.033542	CYR61	cysteine-rich, angiogenic inducer, 61 [Source:HGNC S
21	0	0	1.405782	2.028115	COL4A2	collagen, type IV, alpha 2 [Source:HGNC Symbol;Ac
6	0	0	1.340888	1.934492	COL17A1	collagen, type XVII, alpha 1 [Source:HGNC Symbol;
31	0	0	1.333034	1.923162	PLOD2	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 [S
18	0	0	1.320467	1.905031	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1) [So
2	0	0	1.301809	1.878113	TIMP2	TIMP metallopeptidase inhibitor 2 [Source:HGNC S
37	0	0	1.300575	1.876333	ARF4	ADP-ribosylation factor 4 [Source:HGNC Symbol;Ac
22	0	0	1.252856	1.807489	THBS1	thrombospondin 1 [Source:HGNC Symbol;Acc:11785]
34	0	0	1.251961	1.806198	IGFBP7	insulin-like growth factor binding protein 7 [Source:H
45	0	0	1.248999	1.801924	KRT6A	keratin 6A [Source:HGNC Symbol;Acc:6443]
23	0	0	1.243594	1.794126	AMIGO2	adhesion molecule with Ig-like domain 2 [Source:HGI
14	0	0	1.173912	1.693597	F3	coagulation factor III (thromboplastin, tissue factor)
20	0	0	1.165072	1.680843	ITGB4	integrin, beta 4 [Source:HGNC Symbol;Acc:6158]
35	0	0	1.138537	1.642561	CTSB	cathepsin B [Source:HGNC Symbol;Acc:2527]
13	0	0	1.105810	1.595347	IVNS1ABP	influenza virus NS1A binding protein [Source:HGNC
27	0	0	1.036681	1.495614	PTPRF	protein tyrosine phosphatase, receptor type, F [Sourc
39	0	0	1.034722	1.492788	GAS6	growth arrest-specific 6 [Source:HGNC Symbol;Acc:4
7	0	0	0.882277	1.272857	CD59	CD59 molecule, complement regulatory protein [Sou

2.2.1.2 Top 50 down-regulated genes

	pval	qval	b	estimated_FC	external_gene_name	description
38	0	0	-2.0138652	-2.9053933	DDIT4	DNA-damage-inducible transcript 4 [Source:HGNC
31	0	0	-1.7540653	-2.5305814	PTGES	prostaglandin E synthase [Source:HGNC Symbol;A
10	0	0	-1.6298269	-2.3513432	EPCAM	epithelial cell adhesion molecule [Source:HGNC Sy
18	0	0	-1.6129140	-2.3269430	NUP210	nucleoporin 210kDa [Source:HGNC Symbol;Acc:30
39	0	0	-1.5010512	-2.1655592	FAM84B	family with sequence similarity 84, member B [Sou
40	0	0	-1.1322711	-1.6335220	SOX7	SRY (sex determining region Y)-box 7 [Source:HGI
47	0	0	-1.1216302	-1.6181703	SAPCD2	suppressor APC domain containing 2 [Source:HGN
48	0	0	-1.1216302	-1.6181703	SAPCD2	suppressor APC domain containing 2 [Source:HGN
30	0	0	-1.0772265	-1.5541093	MAL2	mal, T-cell differentiation protein 2 (gene/pseudoge
33	0	0	-1.0604286	-1.5298750	HK2	hexokinase 2 [Source:HGNC Symbol;Acc:4923]

	pval	qval	b	estimated_FC	external_gene_name	description
19	0	0	-1.0311142	-1.4875833	SLC43A3	solute carrier family 43, member 3 [Source:HGNC
58	0	0	-0.9972174	-1.4386806	DANCR	differentiation antagonizing non-protein coding RN
17	0	0	-0.9927910	-1.4322946	TRAP1	TNF receptor-associated protein 1 [Source:HGNC
29	0	0	-0.9887698	-1.4264933	PM20D2	peptidase M20 domain containing 2 [Source:HGNC
56	0	0	-0.9379174	-1.3531287	S100A14	S100 calcium binding protein A14 [Source:HGNC S
20	0	0	-0.8851409	-1.2769884	PSAT1	phosphoserine aminotransferase 1 [Source:HGNC S
11	0	0	-0.8750729	-1.2624634	AVPI1	arginine vasopressin-induced 1 [Source:HGNC Sym
3	0	0	-0.8529226	-1.2305073	TTLL12	tubulin tyrosine ligase-like family, member 12 [Sou
45	0	0	-0.8337233	-1.2028085	RCC1	regulator of chromosome condensation 1 [Source:H
37	0	0	-0.8327113	-1.2013484	HMGB2	high mobility group box 2 [Source:HGNC Symbol;
4	0	0	-0.7990502	-1.1527858	C1QBP	complement component 1, q subcomponent bindin
28	0	0	-0.7894008	-1.1388647	LYAR	Ly1 antibody reactive [Source:HGNC Symbol;Acc:
35	0	0	-0.7665610	-1.1059137	IFI16	interferon, gamma-inducible protein 16 [Source:HG
12	0	0	-0.7575778	-1.0929537	GTF3A	general transcription factor IIIA [Source:HGNC Sy
16	0	0	-0.7435653	-1.0727380	PRMT1	protein arginine methyltransferase 1 [Source:HGNC
50	0	0	-0.7364109	-1.0624163	PTMA	prothymosin, alpha [Source:HGNC Symbol;Acc:962
51	0	0	-0.7364109	-1.0624163	PTMA	prothymosin, alpha [Source:HGNC Symbol;Acc:962
52	0	0	-0.7364109	-1.0624163	PTMA	prothymosin, alpha [Source:HGNC Symbol;Acc:962
53	0	0	-0.7364109	-1.0624163	PTMA	prothymosin, alpha [Source:HGNC Symbol;Acc:962
54	0	0	-0.7364109	-1.0624163	PTMA	prothymosin, alpha [Source:HGNC Symbol;Acc:962
55	0	0	-0.7364109	-1.0624163	PTMA	prothymosin, alpha [Source:HGNC Symbol;Acc:962
36	0	0	-0.7261130	-1.0475596	TKT	transketolase [Source:HGNC Symbol;Acc:11834]
69	0	0	-0.7261130	-1.0475596	TKT	transketolase [Source:HGNC Symbol;Acc:11834]
27	0	0	-0.7201035	-1.0388897	HSPD1	heat shock 60kDa protein 1 (chaperonin) [Source:H
7	0	0	-0.6906473	-0.9963934	HSPE1	heat shock 10kDa protein 1 [Source:HGNC Symbol
26	0	0	-0.6846262	-0.9877068	SLC20A1	solute carrier family 20 (phosphate transporter), m
22	0	0	-0.6745350	-0.9731482	ANP32B	acidic (leucine-rich) nuclear phosphoprotein 32 fam
34	0	0	-0.6678214	-0.9634627	LY6E	lymphocyte antigen 6 complex, locus E [Source:HG
32	0	0	-0.6514588	-0.9398563	SSRP1	structure specific recognition protein 1 [Source:HG
57	0	0	-0.6238098	-0.8999672	PHB2	prohibitin 2 [Source:HGNC Symbol;Acc:30306]
68	0	0	-0.6238098	-0.8999672	PHB2	prohibitin 2 [Source:HGNC Symbol;Acc:30306]
23	0	0	-0.5823714	-0.8401843	HNRNPD	heterogeneous nuclear ribonucleoprotein D (AU-ric
25	0	0	-0.5740855	-0.8282303	HDGF	hepatoma-derived growth factor [Source:HGNC Sy
8	0	0	-0.5615609	-0.8101612	SET	SET nuclear oncogene [Source:HGNC Symbol;Acc:
9	0	0	-0.5615609	-0.8101612	SET	SET nuclear oncogene [Source:HGNC Symbol;Acc:
59	0	0	-0.5502668	-0.7938672	GAS5	growth arrest-specific 5 (non-protein coding) [Sour
60	0	0	-0.5502668	-0.7938672	GAS5	growth arrest-specific 5 (non-protein coding) [Sour
61	0	0	-0.5502668	-0.7938672	GAS5	growth arrest-specific 5 (non-protein coding) [Sour
62	0	0	-0.5502668	-0.7938672	GAS5	growth arrest-specific 5 (non-protein coding) [Sour
63	0	0	-0.5502668	-0.7938672	GAS5	growth arrest-specific 5 (non-protein coding) [Sour
64	0	0	-0.5502668	-0.7938672	GAS5	growth arrest-specific 5 (non-protein coding) [Sour
65	0	0	-0.5502668	-0.7938672	GAS5	growth arrest-specific 5 (non-protein coding) [Sour
66	0	0	-0.5502668	-0.7938672	GAS5	growth arrest-specific 5 (non-protein coding) [Sour
67	0	0	-0.5502668	-0.7938672	GAS5	growth arrest-specific 5 (non-protein coding) [Sour
46	0	0	-0.5433086	-0.7838286	NPM1	nucleophosmin (nucleolar phosphoprotein B23, nun
2	0	0	-0.5400433	-0.7791178	TCOF1	Treacher Collins-Franceschetti syndrome 1 [Source:
41	0	0	-0.5335299	-0.7697209	FAM211A-AS1	FAM211A antisense RNA 1 [Source:HGNC Symbo
42	0	0	-0.5335299	-0.7697209	FAM211A-AS1	FAM211A antisense RNA 1 [Source:HGNC Symbo
43	0	0	-0.5335299	-0.7697209	FAM211A-AS1	FAM211A antisense RNA 1 [Source:HGNC Symbo
44	0	0	-0.5335299	-0.7697209	FAM211A-AS1	FAM211A antisense RNA 1 [Source:HGNC Symbo
6	0	0	-0.5296341	-0.7641005	NCL	nucleolin [Source:HGNC Symbol;Acc:7667]
15	0	0	-0.5291897	-0.7634593	HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1 [So

	pval	qval	b	estimated_FC	external_gene_name	description
1	0	0	-0.5046584	-0.7280681	SLC25A5	solute carrier family 25 (mitochondrial carrier; ade
24	0	0	-0.5013588	-0.7233079	SERBP1	SERPINE1 mRNA binding protein 1 [Source:HGN
5	0	0	-0.4732020	-0.6826861	LDHB	lactate dehydrogenase B [Source:HGNC Symbol;Ac
49	0	0	-0.4730208	-0.6824247	RPS23	ribosomal protein S23 [Source:HGNC Symbol;Acc:
21	0	0	-0.4620879	-0.6666519	HNRNPA1	heterogeneous nuclear ribonucleoprotein A1 [Sourc
13	0	0	-0.4538888	-0.6548231	RPL5	ribosomal protein L5 [Source:HGNC Symbol;Acc:1
14	0	0	-0.4538888	-0.6548231	RPL5	ribosomal protein L5 [Source:HGNC Symbol;Acc:1

2.2.2 H2AZ knock-down vs WT

2.2.2.1 Top 50 up-regulated genes

	pval	qval	b	estimated_FC	external_gene_name	description
35	0	0	5.1024858	7.3613310	SPRR1A	small proline-rich protein 1A [Source:HGNC Symbo
30	0	0	5.0921496	7.3464190	SPRR3	small proline-rich protein 3 [Source:HGNC Symbol;
50	0	0	4.9738268	7.1757153	RPPH1	ribonuclease P RNA component H1 [Source:HGNC
37	0	0	4.8961673	7.0636763	KRT4	keratin 4 [Source:HGNC Symbol;Acc:6441]
36	0	0	4.7521420	6.8558917	CRCT1	cysteine-rich C-terminal 1 [Source:HGNC Symbol;A
24	0	0	4.6174433	6.6615625	S100A7	S100 calcium binding protein A7 [Source:HGNC Sy
34	0	0	3.6712498	5.2964939	SPRR1B	small proline-rich protein 1B [Source:HGNC Symbo
15	0	0	3.2820785	4.7350384	PI3	peptidase inhibitor 3, skin-derived [Source:HGNC S
46	0	0	3.1484276	4.5422209	SBSN	suprabasin [Source:HGNC Symbol;Acc:24950]
22	0	0	2.5573168	3.6894282	IL1RN	interleukin 1 receptor antagonist [Source:HGNC Sy
14	0	0	2.1503737	3.1023335	FN1	fibronectin 1 [Source:HGNC Symbol;Acc:3778]
43	0	0	2.0825334	3.0044606	KRT16	keratin 16 [Source:HGNC Symbol;Acc:6423]
10	0	0	1.9063515	2.7502839	SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasmino
44	0	0	1.8915071	2.7288679	KRT14	keratin 14 [Source:HGNC Symbol;Acc:6416]
19	0	0	1.7182342	2.4788880	KRT17	keratin 17 [Source:HGNC Symbol;Acc:6427]
3	0	0	1.4177393	2.0453655	IL32	interleukin 32 [Source:HGNC Symbol;Acc:16830]
31	0	0	1.3952169	2.0128725	CLDN1	claudin 1 [Source:HGNC Symbol;Acc:2032]
8	0	0	1.1952577	1.7243924	MYL9	myosin, light chain 9, regulatory [Source:HGNC Sy
16	0	0	1.1536047	1.6642998	PMEPA1	prostate transmembrane protein, androgen induced
38	0	0	1.1248123	1.6227612	SPHK1	sphingosine kinase 1 [Source:HGNC Symbol;Acc:11
27	0	0	1.1191836	1.6146406	TAGLN	transgelin [Source:HGNC Symbol;Acc:11553]
51	0	0	1.0991677	1.5857638	CD24	CD24 molecule [Source:HGNC Symbol;Acc:1645]
23	0	0	1.0124296	1.4606271	SLC46A3	solute carrier family 46, member 3 [Source:HGNC S
17	0	0	0.9226449	1.3310952	SDCBP2	syndecan binding protein (syntenin) 2 [Source:HGN
26	0	0	0.9074709	1.3092037	COL8A1	collagen, type VIII, alpha 1 [Source:HGNC Symbol
4	0	0	0.8497352	1.2259088	TYMP	thymidine phosphorylase [Source:HGNC Symbol;Ac
28	0	0	0.8394319	1.2110443	GPR153	G protein-coupled receptor 153 [Source:HGNC Sym
6	0	0	0.8224498	1.1865443	FSTL3	follistatin-like 3 (secreted glycoprotein) [Source:HG
12	0	0	0.7681306	1.1081782	ITGB6	integrin, beta 6 [Source:HGNC Symbol;Acc:6161]
13	0	0	0.7681306	1.1081782	ITGB6	integrin, beta 6 [Source:HGNC Symbol;Acc:6161]
40	0	0	0.7611651	1.0981292	JUN	jun proto-oncogene [Source:HGNC Symbol;Acc:620
41	0	0	0.7501589	1.0822506	MYADM	myeloid-associated differentiation marker [Source:H
2	0	0	0.7477386	1.0787588	TNFRSF12A	tumor necrosis factor receptor superfamily, member
25	0	0	0.7473602	1.0782129	RHOB	ras homolog family member B [Source:HGNC Symb
9	0	0	0.7437961	1.0730710	TIMP1	TIMP metallopeptidase inhibitor 1 [Source:HGNC
48	0	0	0.7204303	1.0393612	FADS3	fatty acid desaturase 3 [Source:HGNC Symbol;Acc:
11	0	0	0.7135840	1.0294841	CYTH1	cytohesin 1 [Source:HGNC Symbol;Acc:9501]
49	0	0	0.7126299	1.0281076	LINC00707	long intergenic non-protein coding RNA 707 [Source

	pval	qval	b	estimated_FC	external_gene_name	description
21	0	0	0.6686057	0.9645941	EPHB2	EPH receptor B2 [Source:HGNC Symbol;Acc:3393]
32	0	0	0.6090896	0.8787305	DUSP7	dual specificity phosphatase 7 [Source:HGNC Symb
1	0	0	0.5776948	0.8334374	SLC7A2	solute carrier family 7 (cationic amino acid transpo
33	0	0	0.5530118	0.7978273	CTSB	cathepsin B [Source:HGNC Symbol;Acc:2527]
42	0	0	0.5276258	0.7612031	GAS6	growth arrest-specific 6 [Source:HGNC Symbol;Acc
45	0	0	0.5215442	0.7524293	GJB3	gap junction protein, beta 3, 31kDa [Source:HGNC
18	0	0	0.5182116	0.7476212	TUBA4A	tubulin, alpha 4a [Source:HGNC Symbol;Acc:12407
29	0	0	0.4981172	0.7186313	PEA15	phosphoprotein enriched in astrocytes 15 [Source:H
39	0	0	0.4367009	0.6300262	PTRF	polymerase I and transcript release factor [Source:H
5	0	0	0.4249671	0.6130980	LIMA1	LIM domain and actin binding 1 [Source:HGNC Sy
20	0	0	0.4038286	0.5826015	TNS4	tensin 4 [Source:HGNC Symbol;Acc:24352]
47	0	0	0.4005200	0.5778282	SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), me
7	0	0	0.3938584	0.5682176	CD59	CD59 molecule, complement regulatory protein [Sou

2.2.2.2 Top 50 down-regulated genes

	pval	qval	b	estimated_FC	external_gene_name	description
21	0	0	-2.0496383	-2.9570031	ADM2	adrenomedullin 2 [Source:HGNC Symbol;Acc:28898
23	0	0	-1.9846578	-2.8632559	CHAC1	ChaC, cation transport regulator homolog 1 (E. co
4	0	0	-1.5703041	-2.2654700	ASNS	asparagine synthetase (glutamine-hydrolyzing) [Sou
39	0	0	-1.3233719	-1.9092221	SEMA6B	sema domain, transmembrane domain (TM), and c
8	0	0	-1.2254337	-1.7679271	PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondr
49	0	0	-1.1881104	-1.7140809	SLC6A9	solute carrier family 6 (neurotransmitter transporte
6	0	0	-1.1713273	-1.6898682	PHGDH	phosphoglycerate dehydrogenase [Source:HGNC Sy
28	0	0	-1.0523542	-1.5182261	PSAT1	phosphoserine aminotransferase 1 [Source:HGNC S
10	0	0	-0.9368049	-1.3515238	TRIB3	tribbles pseudokinase 3 [Source:HGNC Symbol;Acc
35	0	0	-0.9200383	-1.3273346	H2AFZ	H2A histone family, member Z [Source:HGNC Sym
32	0	0	-0.9041619	-1.3044299	SLC7A11	solute carrier family 7 (anionic amino acid transpo
18	0	0	-0.9036152	-1.3036412	STC2	stanniocalcin 2 [Source:HGNC Symbol;Acc:11374]
41	0	0	-0.8422814	-1.2151551	DDIT4	DNA-damage-inducible transcript 4 [Source:HGNC
36	0	0	-0.7931437	-1.1442645	GPT2	glutamic pyruvate transaminase (alanine aminotra
34	0	0	-0.7837435	-1.1307028	TSC22D3	TSC22 domain family, member 3 [Source:HGNC S
29	0	0	-0.7149083	-1.0313947	ALDH1L2	aldehyde dehydrogenase 1 family, member L2 [Sou
31	0	0	-0.7043755	-1.0161991	PSPH	phosphoserine phosphatase [Source:HGNC Symbol
7	0	0	-0.7027026	-1.0137856	SCD	stearoyl-CoA desaturase (delta-9-desaturase) [Sour
33	0	0	-0.6723065	-0.9699333	ACSL1	acyl-CoA synthetase long-chain family member 1 [
25	0	0	-0.6473998	-0.9340004	SESN2	sestrin 2 [Source:HGNC Symbol;Acc:20746]
42	0	0	-0.6149965	-0.8872524	PXK	PX domain containing serine/threonine kinase [Sou
17	0	0	-0.5941170	-0.8571296	CARS	cysteinyl-tRNA synthetase [Source:HGNC Symbol;
45	0	0	-0.5670660	-0.8181033	NIPSNAP1	nipsnap homolog 1 (C. elegans) [Source:HGNC Sy
24	0	0	-0.5556528	-0.8016376	ARHGEF16	Rho guanine nucleotide exchange factor (GEF) 16
3	0	0	-0.5553555	-0.8012086	MTHFD2	methylenetetrahydrofolate dehydrogenase (NADP+
30	0	0	-0.5485623	-0.7914081	WARS	tryptophanyl-tRNA synthetase [Source:HGNC Sym
16	0	0	-0.5452225	-0.7865898	GARS	glycyl-tRNA synthetase [Source:HGNC Symbol;Ac
50	0	0	-0.5232495	-0.7548894	APOL6	apolipoprotein L, 6 [Source:HGNC Symbol;Acc:148
37	0	0	-0.5152614	-0.7433650	MARS	methionyl-tRNA synthetase [Source:HGNC Symbo
38	0	0	-0.5152614	-0.7433650	MARS	methionyl-tRNA synthetase [Source:HGNC Symbo
15	0	0	-0.5111301	-0.7374049	DNASE2	deoxyribonuclease II, lysosomal [Source:HGNC Sy
14	0	0	-0.4970217	-0.7170508	WDR83OS	WD repeat domain 83 opposite strand [Source:HG

	pval	qval	b	estimated_FC	external_gene_name	description
5	0	0	-0.4950638	-0.7142261	AARS	alanyl-tRNA synthetase [Source:HGNC Symbol;Ac
51	0	0	-0.4882800	-0.7044391	PEG10	paternally expressed 10 [Source:HGNC Symbol;Acc
46	0	0	-0.4852357	-0.7000471	EIF4EBP1	eukaryotic translation initiation factor 4E binding
40	0	0	-0.4671883	-0.6740103	SLC3A2	solute carrier family 3 (amino acid transporter hea
9	0	0	-0.4589282	-0.6620934	CDC25B	cell division cycle 25B [Source:HGNC Symbol;Acc:
27	0	0	-0.4488746	-0.6475892	YARS	tyrosyl-tRNA synthetase [Source:HGNC Symbol;A
43	0	0	-0.4221496	-0.6090331	CKS1B	CDC28 protein kinase regulatory subunit 1B [Sour
26	0	0	-0.4184890	-0.6037520	NUP210	nucleoporin 210kDa [Source:HGNC Symbol;Acc:30
19	0	0	-0.4033343	-0.5818884	TFCP2L1	transcription factor CP2-like 1 [Source:HGNC Sym
22	0	0	-0.4011661	-0.5787603	ATF4	activating transcription factor 4 [Source:HGNC Sy
12	0	0	-0.3805752	-0.5490539	SFRP1	secreted frizzled-related protein 1 [Source:HGNC S
2	0	0	-0.3749004	-0.5408669	PSMA4	proteasome (prosome, macropain) subunit, alpha t
11	0	0	-0.3731684	-0.5383683	SLC7A5	solute carrier family 7 (amino acid transporter ligh
47	0	0	-0.3527527	-0.5089146	IARS	isoleucyl-tRNA synthetase [Source:HGNC Symbol;
48	0	0	-0.3475641	-0.5014289	LONP1	lon peptidase 1, mitochondrial [Source:HGNC Sym
44	0	0	-0.3443693	-0.4968199	SHMT2	serine hydroxymethyltransferase 2 (mitochondrial)
1	0	0	-0.3209688	-0.4630601	SARS	seryl-tRNA synthetase [Source:HGNC Symbol;Acc
13	0	0	-0.2975210	-0.4292321	SLC1A5	solute carrier family 1 (neutral amino acid transpo
20	0	0	-0.2877494	-0.4151346	EIF2S2	eukaryotic translation initiation factor 2, subunit 2

2.2.3 Plots of genes used in qPCR validation of experiment

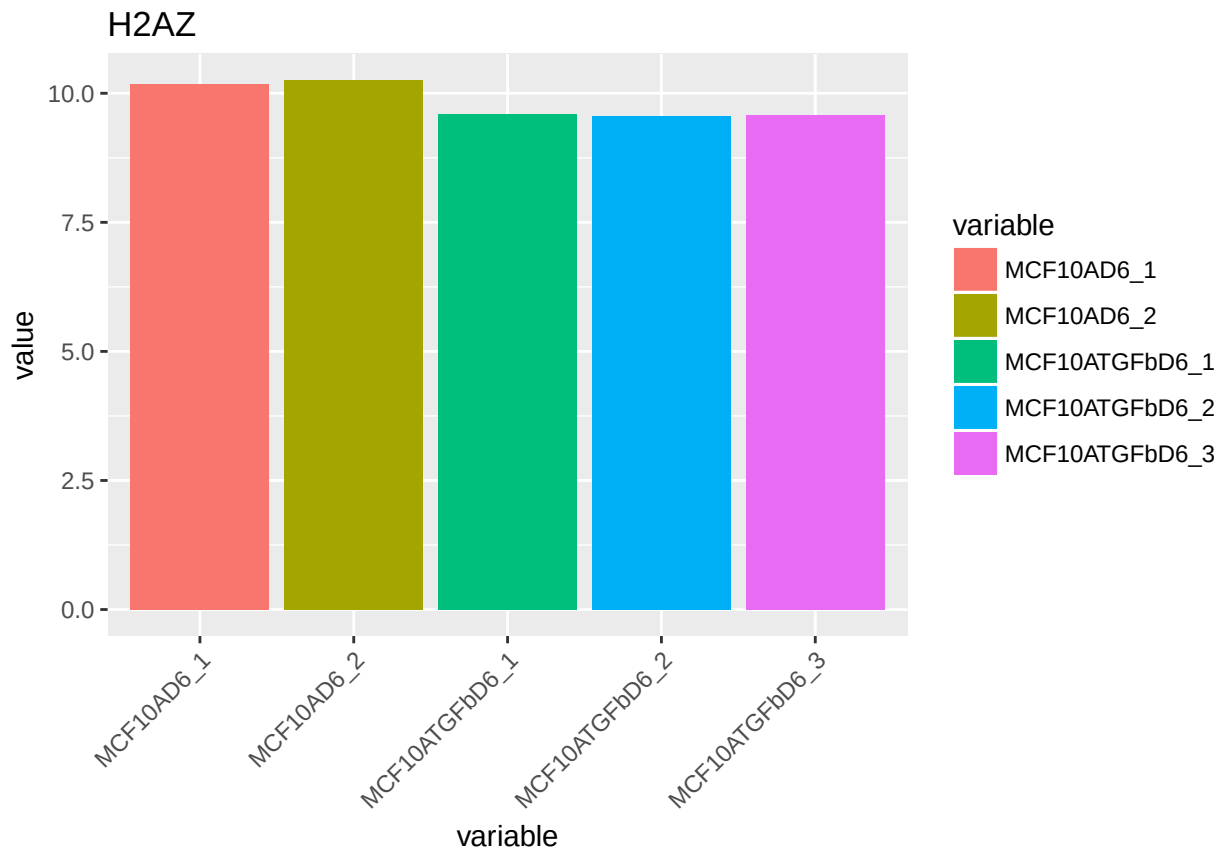
2.2.3.1 TGFb-treated MCF10A vs WT

```
# list of IDs
geneList <- c(H2AZ = "ENSG00000164032",
              ECadherin = "ENSG00000039068",
              NCadherin = "ENSG00000170558",
              EpCAM = "ENSG00000119888",
              Fibronectin = "ENSG00000115414",
              TGFb1 = "ENSG00000105329")

geneExp <- as.data.frame(resultsCompressed[["MCF10A_vs_TGFb"]]$kallisto_table_genes)
samples <- grep("MCF", colnames(geneExp))
w1 <- geneExp$ensembl_gene_id %in% geneList
geneExp <- geneExp[w1, c(1, samples)]
gdata <- melt(geneExp)
```

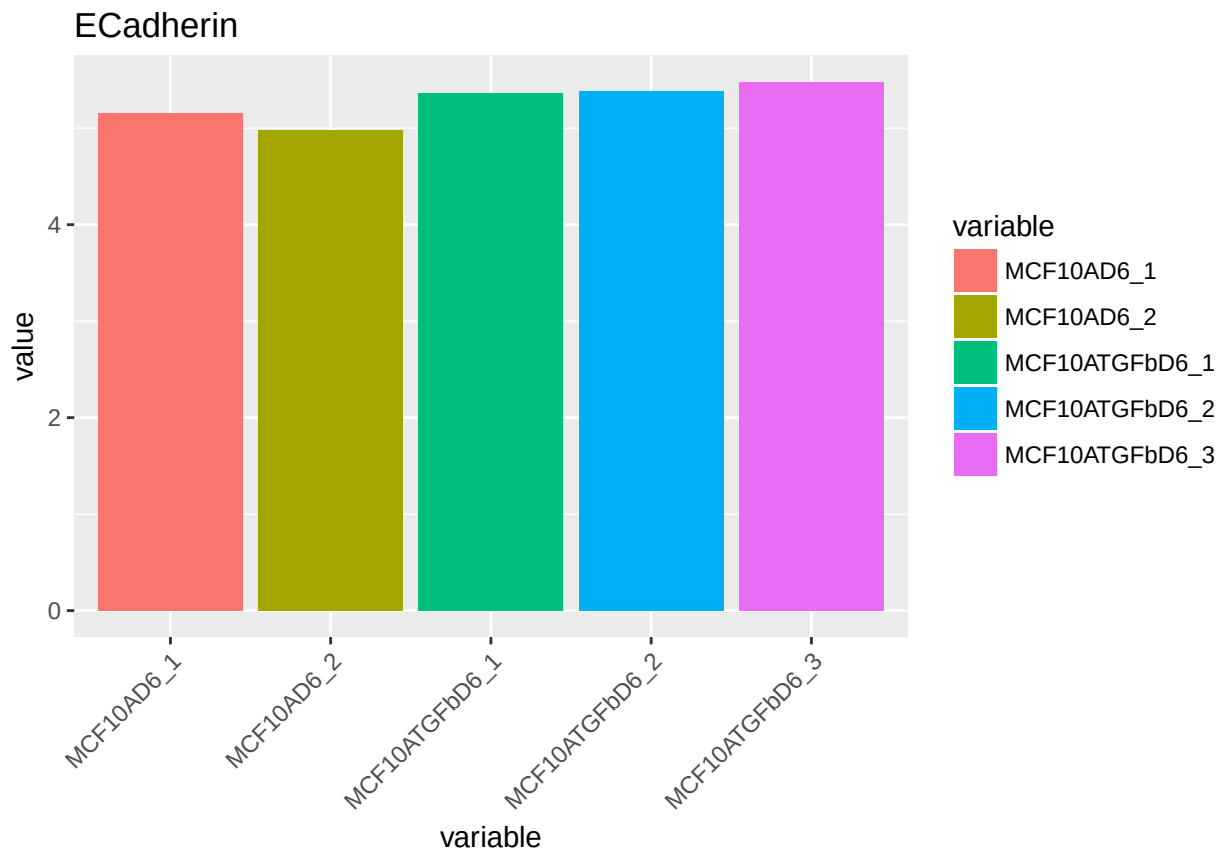
```
## Using ensembl_gene_id as id variables
```

```
gdata$value <- log2(gdata$value)

sapply(names(geneList), function(x){
  s <- geneExp$ensembl_gene_id == geneList[x]
  dat <- geneExp[s, ]
  dat <- melt(dat)
  dat$value <- log2(dat$value + 1)
  p1 <- ggplot(dat, aes(x = variable, y = value, fill = variable)) + geom_bar(stat="identity") + ggtitl
  plot(p1)
})
```

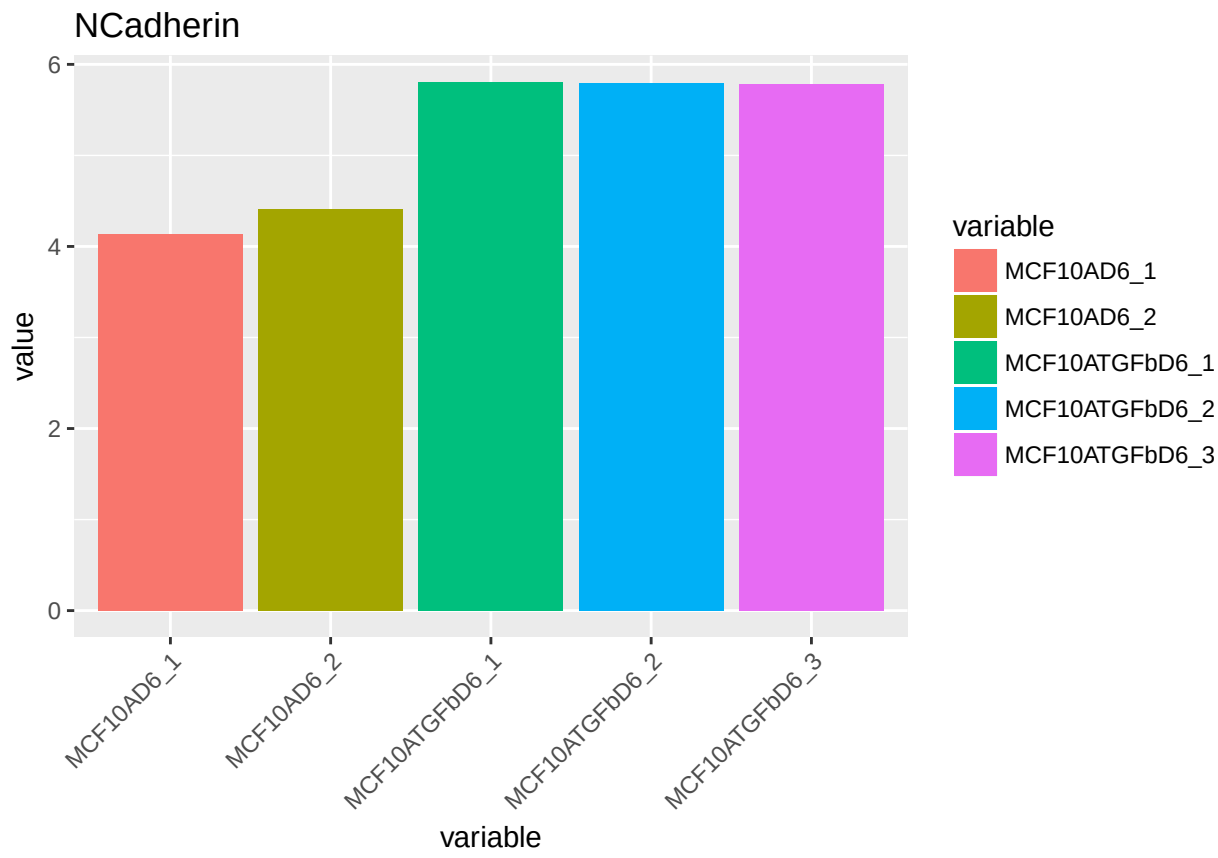
```
## Using ensembl_gene_id as id variables
## Using ensembl_gene_id as id variables
```



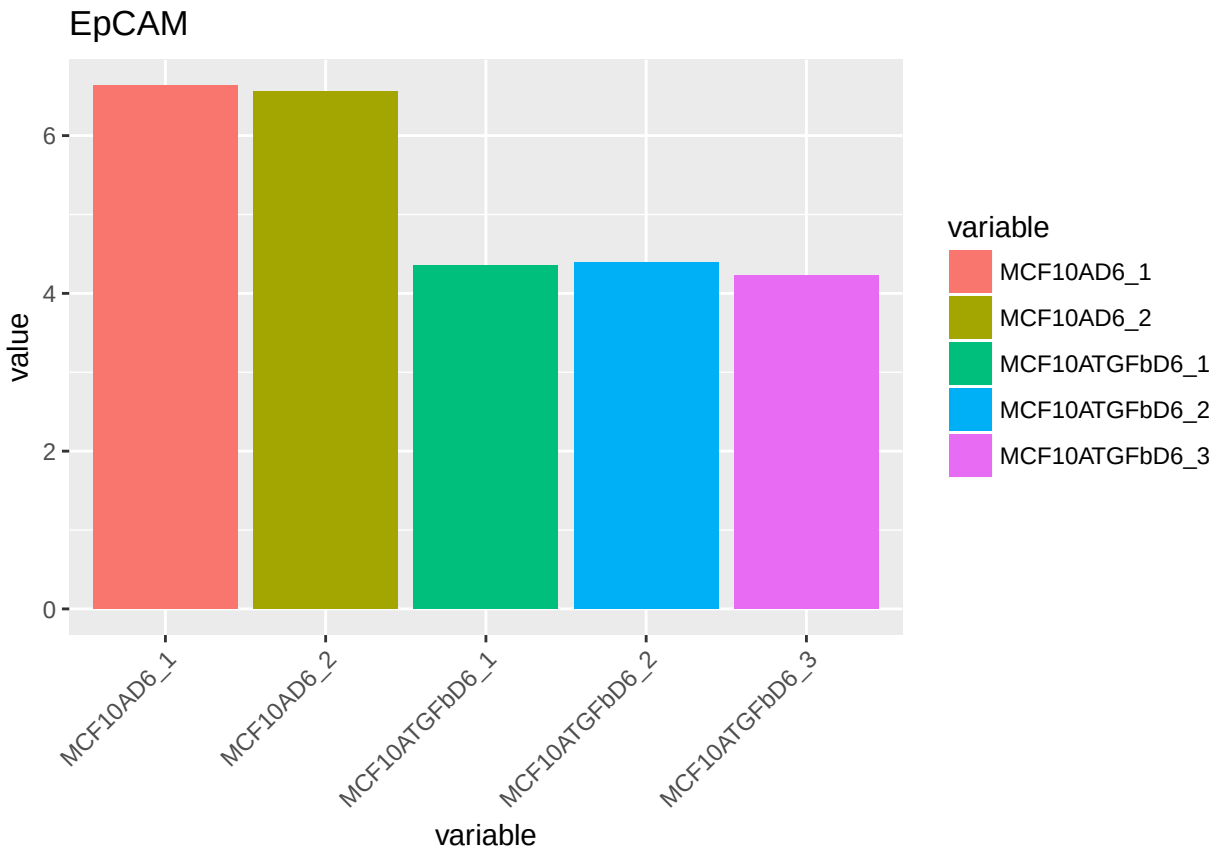

```
## Using ensembl_gene_id as id variables
```



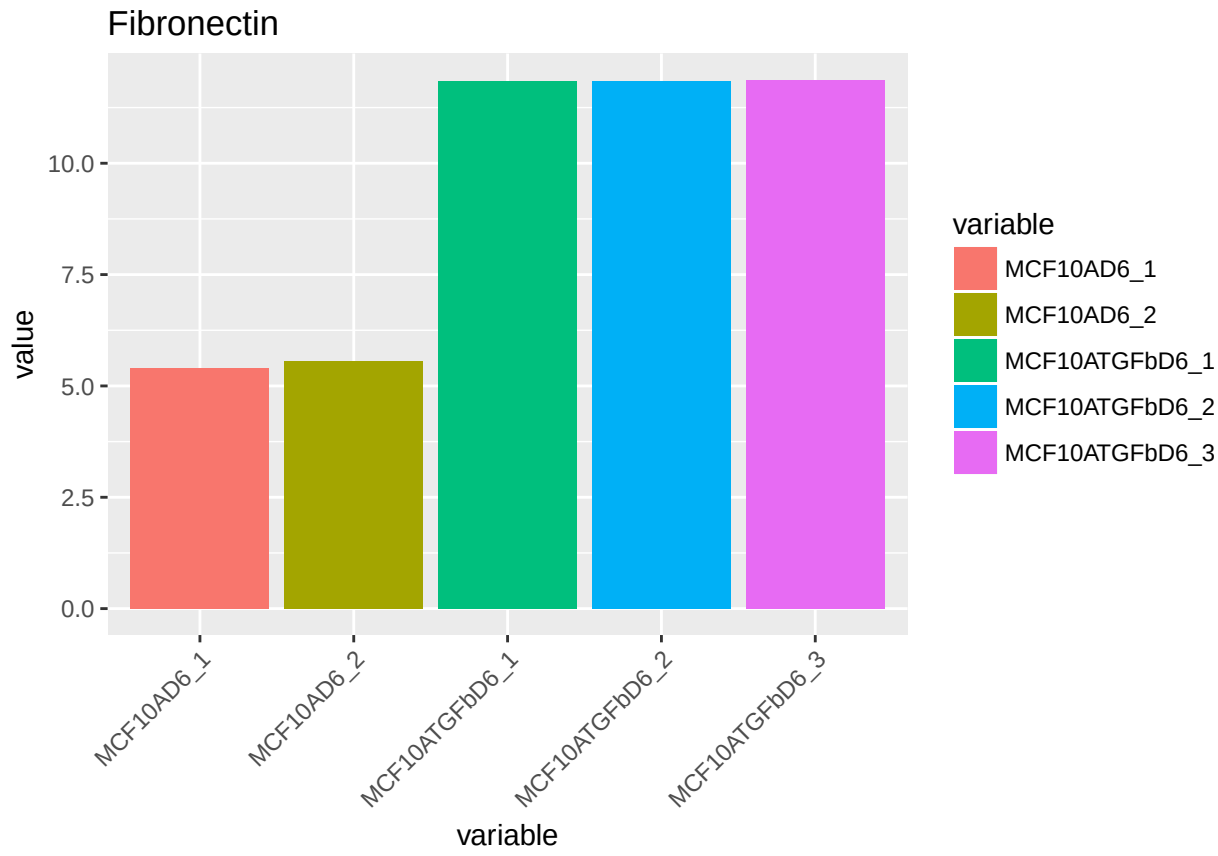

```
## Using ensembl_gene_id as id variables
```

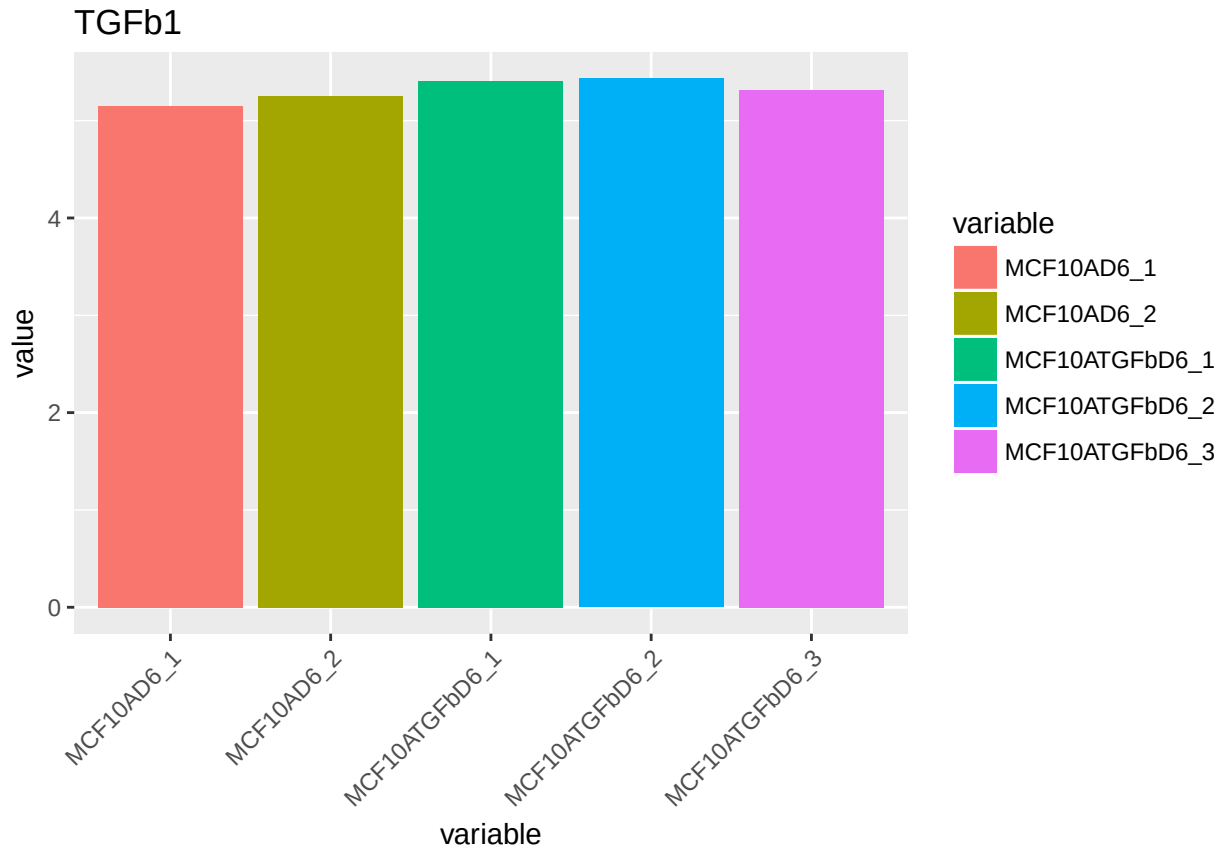



```
## Using ensembl_gene_id as id variables
```




```
## Using ensembl_gene_id as id variables
```


Fibronectin
10.0
7.5
5.0
2.5
0.0
value
MCF10AD6_1
MCF10AD6_2
MCF10ATGFbD6_1
MCF10ATGFbD6_2
MCF10ATGFbD6_3
variable
variable
MCF10AD6_1
MCF10AD6_2
MCF10ATGFbD6_1
MCF10ATGFbD6_2
MCF10ATGFbD6_3


TGFb1
4
2
0
value
MCF10AD6_1
MCF10AD6_2
MCF10ATGFbD6_1
MCF10ATGFbD6_2
MCF10ATGFbD6_3
variable
variable
MCF10AD6_1
MCF10AD6_2
MCF10ATGFbD6_1
MCF10ATGFbD6_2
MCF10ATGFbD6_3

```
##        H2AZ   ECadherin NCadherin EpCAM  Fibronectin TGFb1
## data   List,1 List,1    List,1    List,1 List,1      List,1
## layout ?      ?         ?         ?      ?           ?
## plot   List,9 List,9    List,9    List,9 List,9      List,9
```

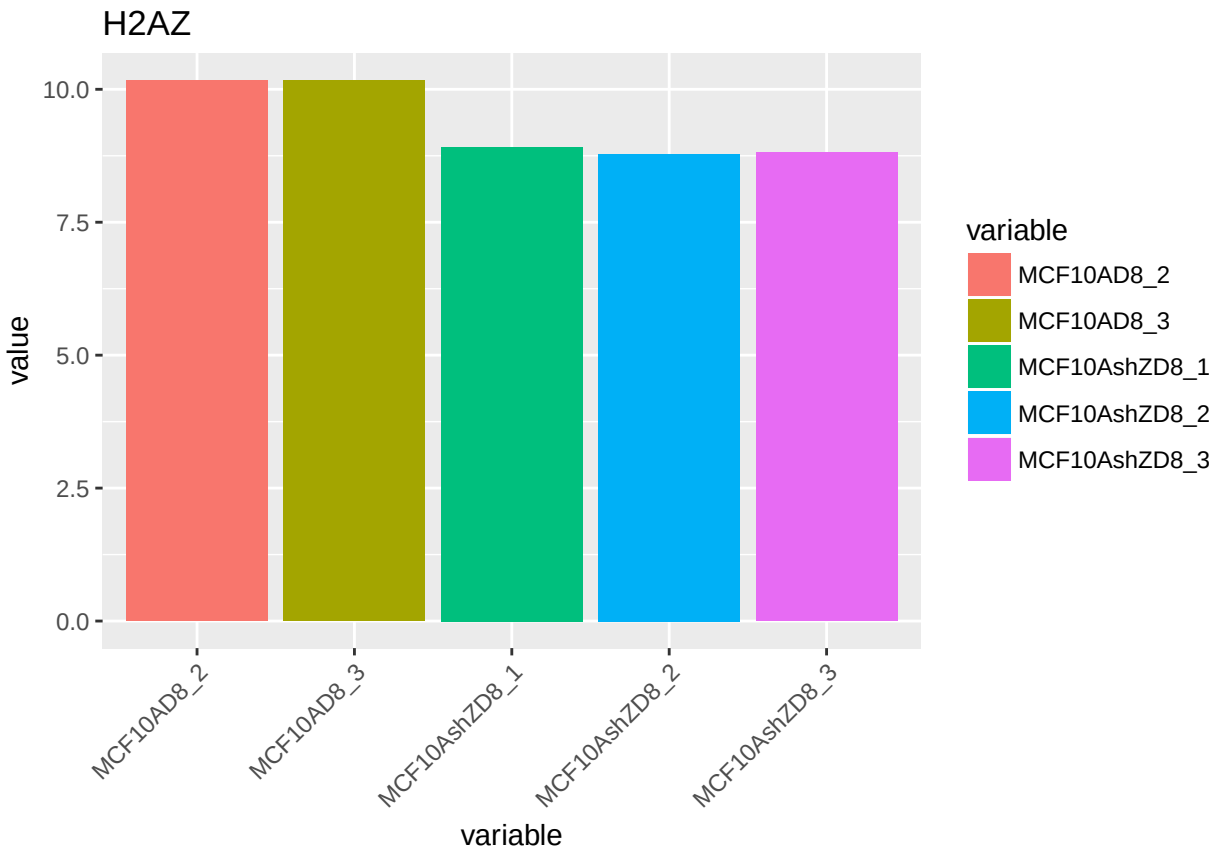
2.2.3.2 shZ-knock-down MCF10A vs WT

```
# list of IDs
geneList <- c(H2AZ = "ENSG00000164032",
              ECadherin = "ENSG00000039068",
              NCadherin = "ENSG00000170558",
              EpCAM = "ENSG00000119888",
              Fibronectin = "ENSG00000115414",
              TGFb1 = "ENSG00000105329")

geneExp <- as.data.frame(resultsCompressed[["MCF10A_vs_shZ"]]$kallisto_table_genes)
samples <- grep("MCF", colnames(geneExp))
w1 <- geneExp$ensembl_gene_id %in% geneList
geneExp <- geneExp[w1, c(1, samples)]
gdata <- melt(geneExp)
```

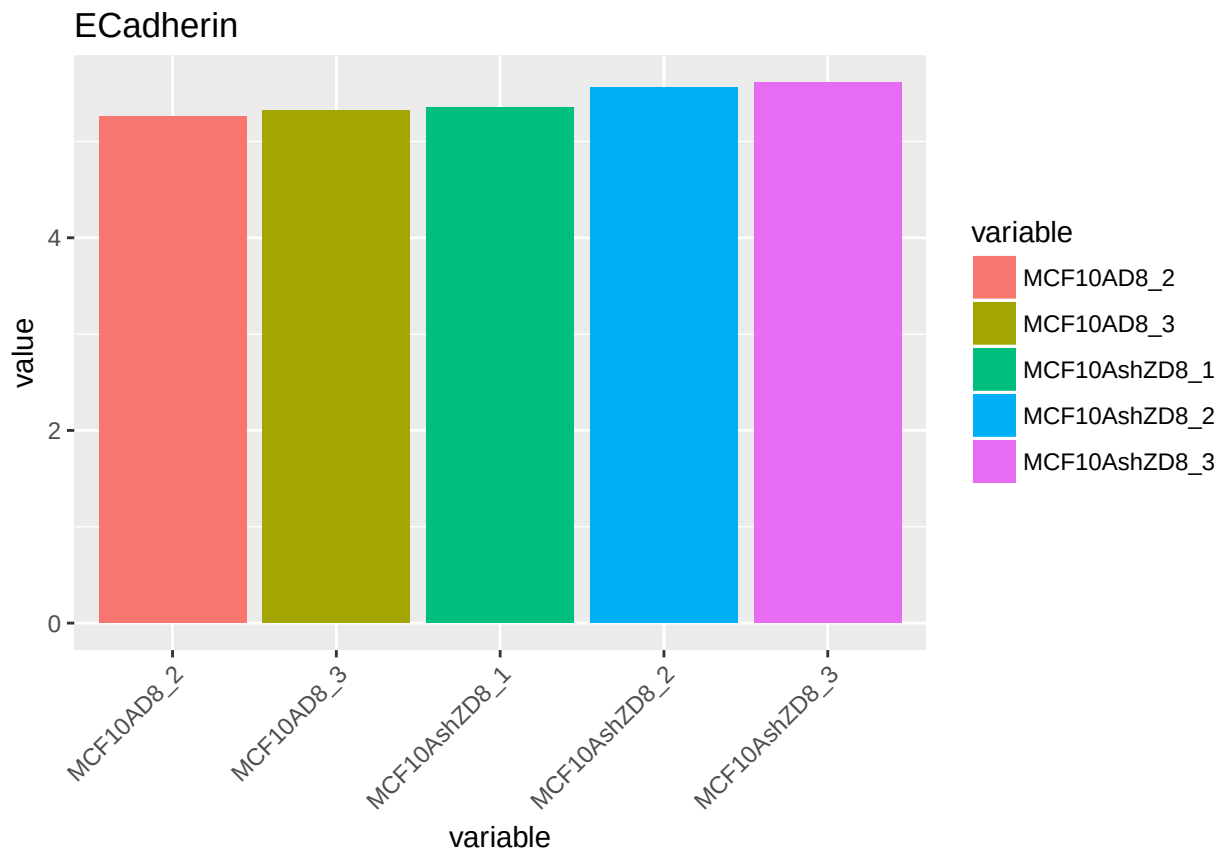
```
## Using ensembl_gene_id as id variables
```

```
gdata$value <- log2(gdata$value)

sapply(names(geneList), function(x){
  s <- geneExp$ensembl_gene_id == geneList[x]
  dat <- geneExp[s, ]
  dat <- melt(dat)
  dat$value <- log2(dat$value + 1)
  p1 <- ggplot(dat, aes(x = variable, y = value, fill = variable)) + geom_bar(stat="identity") + ggtitl
  plot(p1)
})
```

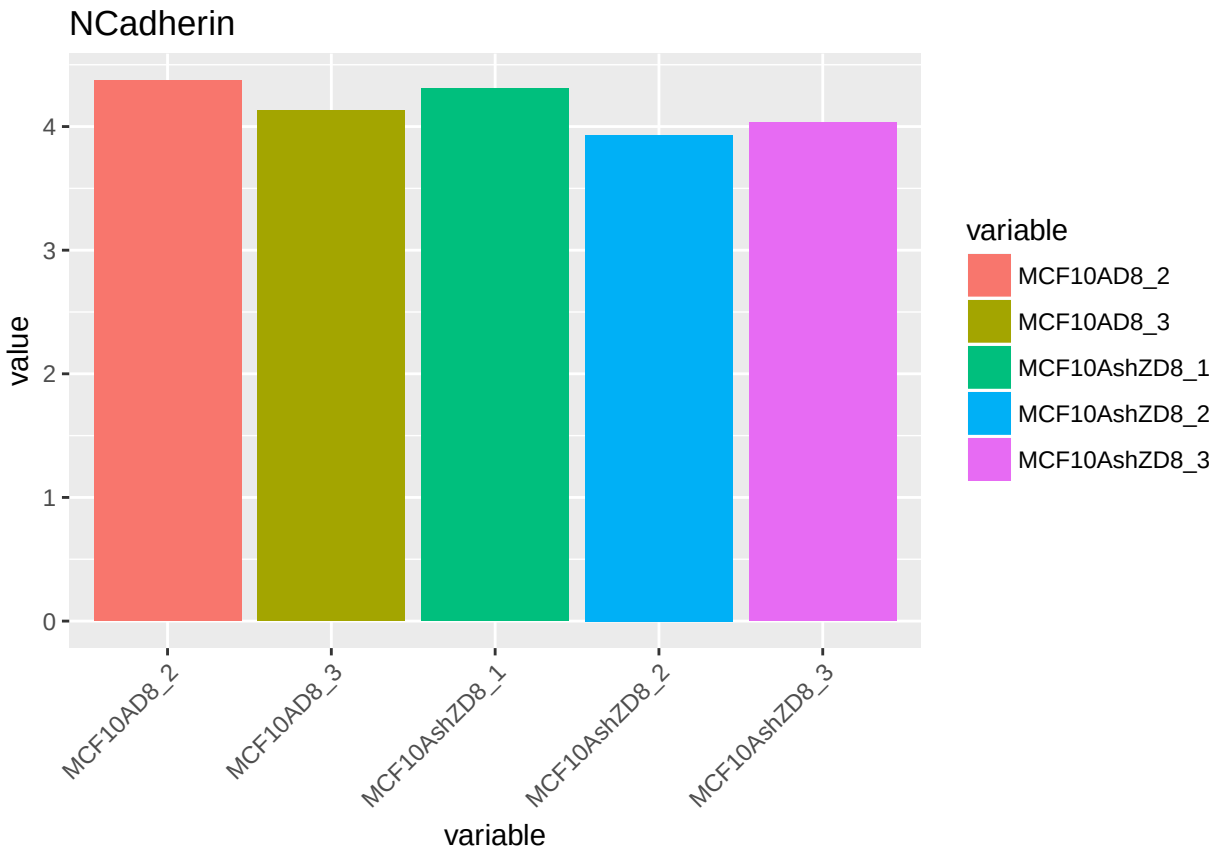
```
## Using ensembl_gene_id as id variables
## Using ensembl_gene_id as id variables
```



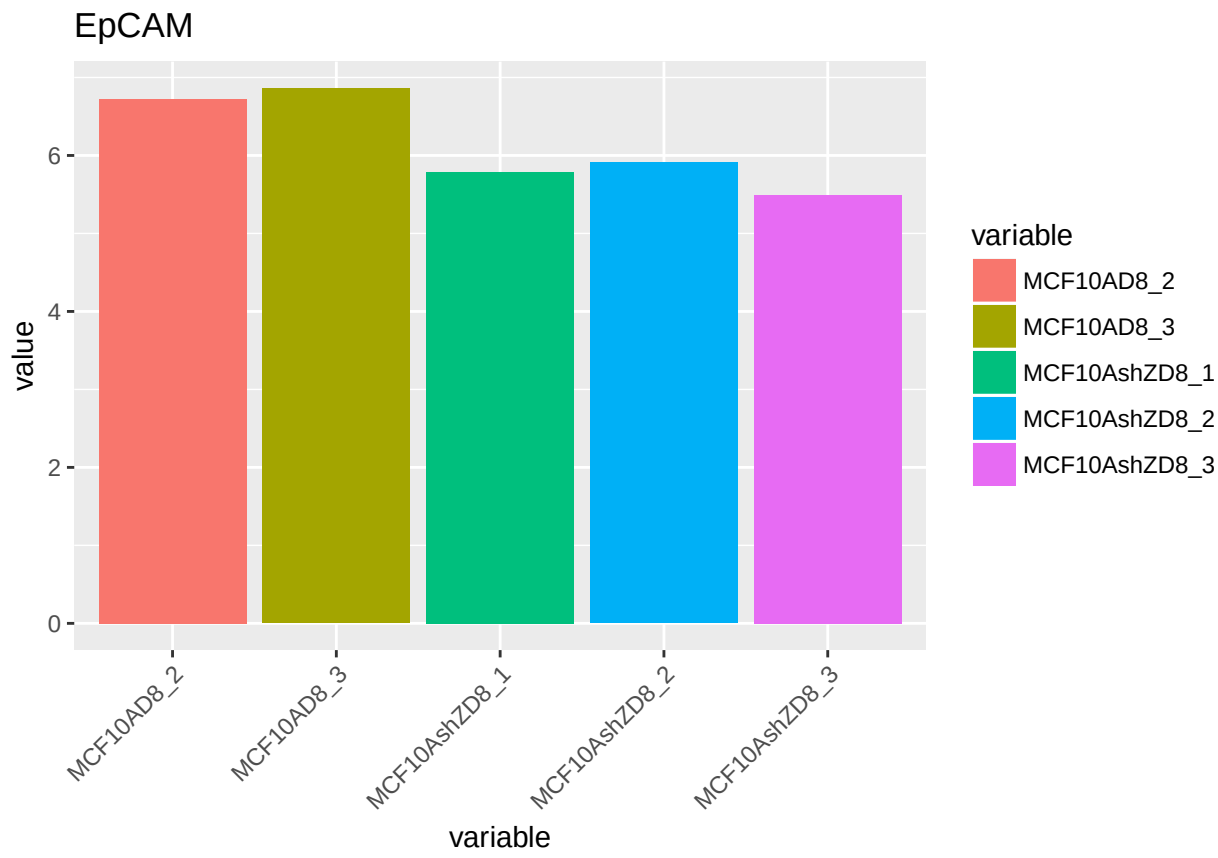

```
## Using ensembl_gene_id as id variables
```



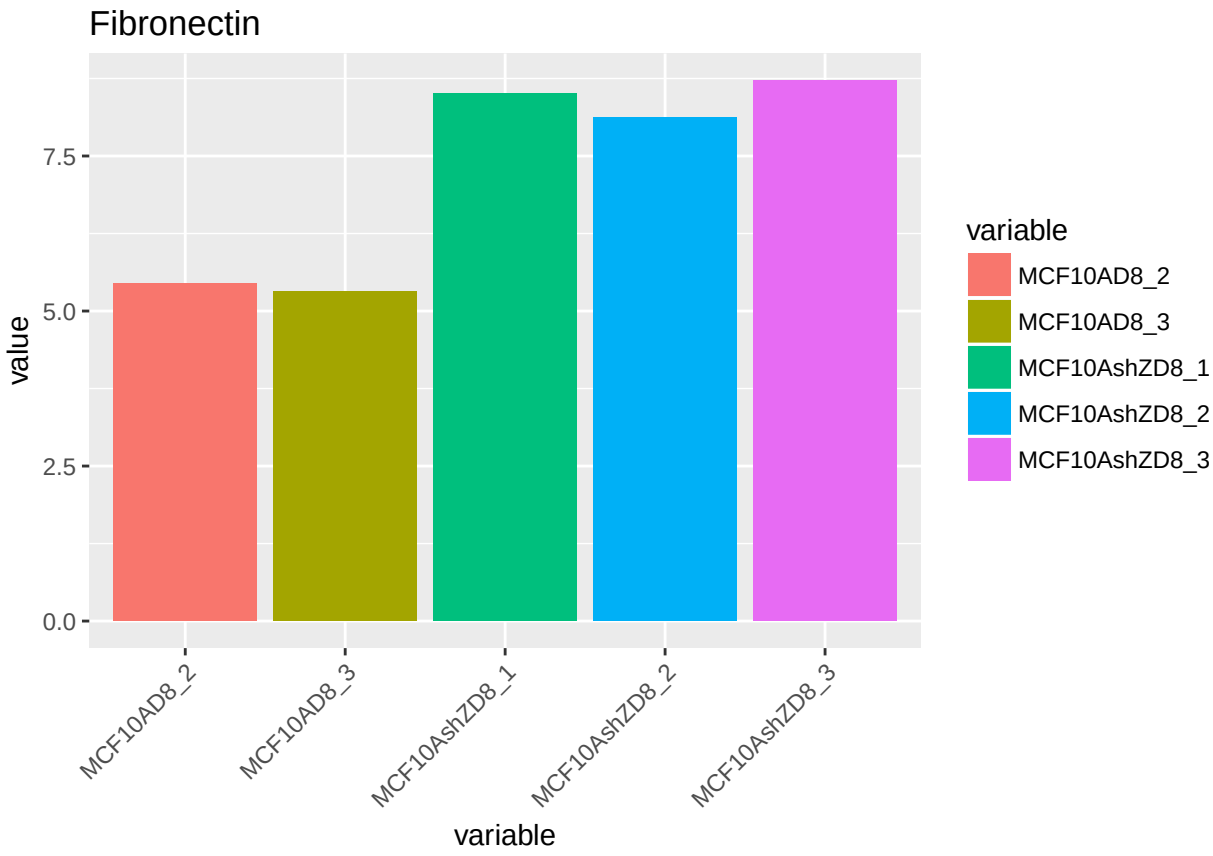

```
## Using ensembl_gene_id as id variables
```

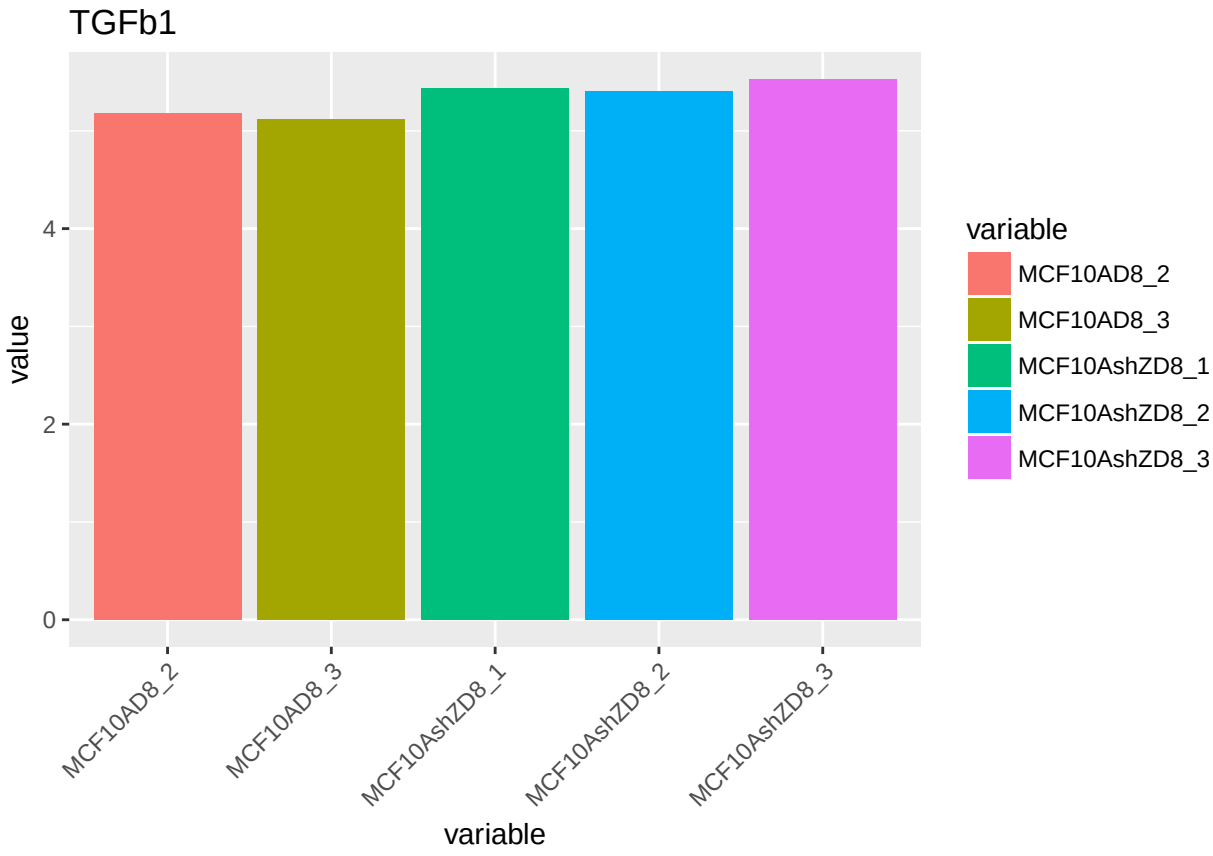



```
## Using ensembl_gene_id as id variables
```




```
## Using ensembl_gene_id as id variables
```


Fibronectin
7.5
5.0
2.5
0.0
value
MCF10AD8_2
MCF10AD8_3
MCF10AshZD8_1
MCF10AshZD8_2
MCF10AshZD8_3
variable
variable
MCF10AD8_2
MCF10AD8_3
MCF10AshZD8_1
MCF10AshZD8_2
MCF10AshZD8_3


TGFb1
4
2
0
value
MCF10AD8_2
MCF10AD8_3
MCF10AshZD8_1
MCF10AshZD8_2
MCF10AshZD8_3
variable
variable
MCF10AD8_2
MCF10AD8_3
MCF10AshZD8_1
MCF10AshZD8_2
MCF10AshZD8_3

```
##        H2AZ   ECadherin NCadherin EpCAM  Fibronectin TGFb1
## data   List,1 List,1    List,1    List,1 List,1      List,1
## layout ?      ?         ?         ?      ?           ?
## plot   List,9 List,9    List,9    List,9 List,9      List,9
```