SUPPLEMENTARY INFORMATION

Integrated functional genomic analysis of microRNAs in breast cancer, prostate cancer and neuroblastoma

Nikolic I, Elsworth B, Dodson E, Wu SZ, Gould CM, Mestdagh P, Marshall GM, Horvath LG, Simpson KJ and Swarbrick A

SUPPLEMENTARY METHODS

Screen transfection conditions

Supplementary Table S1. Transfection conditions across a panel of cancer cell lines

Cell Line	Cancer type	Subtype	Media	Transfection Reagent	Volume (µl)	Cells/well
MDA-MB-231	Breast	Triple-negative	RPMI/10%FCS,HEPES, insulin	DF4	0.05	1100
MDA-MB-468	Breast	Triple-negative	RPMI/10%FCS,HEPES, insulin	DF4	0.05	1200
SK-BR-3	Breast	HER2+	RPMI/10%FCS,HEPES, insulin	DF4	0.05	1500
BT474	Breast	HER2+/ER+	RPMI/10%FCS,HEPES, insulin	DF4	0.03	1500
SHEP	Neuroblastoma		DMEM/10%FCS	DF1	0.05	800
KELLY	Neuroblastoma	MYCN amp	RPMI/10%FCS	DF1	0.03	1100
PC3	Prostate	Castrate-resistant	RPMI/10%FCS,HEPES, insulin	DF1	0.05	900
DU145	Prostate	Castrate-resistant	RPMI/10%FCS,HEPES, insulin	DF1	0.03	700
MCF10A	Normal	Epithelial	DMEM F12/5% Horse serum, EGF, cholera toxin, hydrocortisone, insulin	DF3	0.04	500
IMR-90	Normal	Fibroblast	MEM/10%FCS, Na Pyruvate, Non Essential Amino Acids	DF1	0.008	200

Database construction

The application was built using the Groovy-based web application framework Grails (http://grails.org), which runs on the Java platform. Groovy scripts were used to insert all into a PostgreSQL relational database (http://www.postgresql.org/). Additional functionality was provided using various JavaScript libraries including Biodalliance (http://www.biodalliance.org/), CanvasXpress (http://canvasxpress.org/), D3 (http://d3js.org/) and Highcharts (http://www.highcharts.com/).

Supplementary Table S2. Summary of public data implemented in the database

Data	Version	Publication	Data Location
miRBase	21	Kozomara & Griffiths-Jones, 2014 (1)	http://www.mirbase.org/ftp.shtml
Essential genes	n/a	Koh *et al*, 2011 (2)	http://dpsc.ccbr.utoronto.ca/cancer/download.html
Expression data	n/a	Klijn *et al*, 2015 (3)	http://research-pub.gene.com/KlijnEtAl2014/
TargetScan	June 2012	Lewis *et al,* 2005 (4)	http://www.targetscan.org/
MiRTarBase	Nov 2013	Hsu *et al,* 2014 (5)	http://mirtarbase.mbc.nctu.edu.tw/
StarBase	Sep 2013	Li *et al,* 2014 (6)	http://starbase.sysu.edu.cn/
MiRDB	April 2013	Wong and Wang, 2015 (7)	http://mirdb.org/miRDB/
Diana microT-CDS	Jul 2013	Paraskevopoulou *et al,* 2013 (8)	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=microT_CDS/index
Cardiac	n/a	Eulalio *et al,* 2012 (9)	http://www.nature.com/articles/nature11739
GenCode	21	Harrow *et al,* 2012 (10)	http://www.gencodegenes.org/
KEGG	Nov 2014	Kanehisa *et al,* 2014 (11)	http://www.genome.jp/kegg/

Classification of microRNA in microRNA families

MiRNA family classification used in the database was imported from the miR base (miFam.dat file), the central online repository of miRNA nomenclature, sequence data, annotation, and target prediction. The miR family classification was done using the single-linkage methods to cluster the precursor sequences based on BLAST hits, and then adjusting (merging and/or splitting) manually the clustered families by multiple sequence alignment. The aim is to group miRNA that have a common ancestor into the same family.

SUPPLEMENTARY REFERENCES

1. Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research*, **42**, D68-73.
2. Koh, J.L., Brown, K.R., Sayad, A., Kasimer, D., Ketela, T. and Moffat, J. (2012) COLT-Cancer: functional genetic screening resource for essential genes in human cancer cell lines. *Nucleic acids research*, **40**, D957-963.
3. Klijn, C., Durinck, S., Stawiski, E.W., Haverty, P.M., Jiang, Z., Liu, H., Degenhardt, J., Mayba, O., Gnad, F., Liu, J. *et al.* (2015) A comprehensive transcriptional portrait of human cancer cell lines. *Nature biotechnology*, **33**, 306-312.
4. Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15-20.
5. Hsu, S.D., Tseng, Y.T., Shrestha, S., Lin, Y.L., Khaleel, A., Chou, C.H., Chu, C.F., Huang, H.Y., Lin, C.M., Ho, S.Y. *et al.* (2014) miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic acids research*, **42**, D78-85.
6. Li, J.H., Liu, S., Zhou, H., Qu, L.H. and Yang, J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research*, **42**, D92-97.
7. Wong, N. and Wang, X. (2015) miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic acids research*, **43**, D146-152.
8. Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Vlachos, I.S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T. and Hatzigeorgiou, A.G. (2013) DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic acids research*, **41**, W169-173.
9. Eulalio, A., Mano, M., Dal Ferro, M., Zentilin, L., Sinagra, G., Zacchigna, S. and Giacca, M. (2012) Functional screening identifies miRNAs inducing cardiac regeneration. *Nature*, **492**, 376-381.
10. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S. *et al.* (2012) GENCODE: the

reference human genome annotation for The ENCODE Project. *Genome research*, **22**, 1760-1774.

11. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research*, **42**, D199-205.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure S1. Viability assay dynamic range across the panel of cancer cell lines. The technical performance of the screen was verified using robust positive and negative controls in each cell line screened, and quantified using Z' score as described in Materials and Methods. The graphs display viability values of different controls normalised to NT wells, a negative control that shows no phenotype in any of the cell lines. Negative controls also include mock, mimic nontargeting control (mimct), and inhibitor nontargeting control (inhct) while positive controls include genes that upon knock down induce variable levels of death in different cell lines (PLK1, TOX, KIF11, COPB2).

Supplementary Figure S2. An overview of screen results in all cell lines screened. The graphs display viability values normalised to the NT negative control for all mimics and inhibitors used in the screen as well as the positive (PLK1, TOX, KIF11, COPB2) and negative controls (mock, mimct, inhct) used for performance monitoring. The libraries' performance is consistent among different cancer cell lines—mimics induce a range of viability values while inhibitors exert less potent effects.

Supplementary Figure S3. Small RNA profiling in cancer cell lines and alternative miR inhibitor analysis. (**a**) The graphs display small RNA sequencing results in the breast cancer and neuroblastoma cell lines used in the screens. The analysis of the FastQ files was performed using Oasis 1.0 sRNA detection tool (see Materials and Methods), and raw count files of filtered miRNAs were converted to counts per million (CPM). (**b**) We applied robust z-score statistics (see Materials and Methods) to reanalyse the microRNA inhibitor data, which is able to capture smaller but statistically significant effects. The cuttoff of -2.5 corresponds to a minimum of 20% lethal activity.

Supplementary Figure S4. A snapshot of the "Target Search" page in the database listing ranked predicted microRNA targets and tools available for target enrichment analyses.

Supplementary Figure S5. Database snapshots ("KEGG" button in the "Analysis" section of the database) showing unsupervised hierarchical clustering of the disease-specific microRNA family members and the significantly enriched KEGG pathways in (**a**) breast, (**b**) prostate, and (**c**) neuroblastoma disease. Colour coding represents the

number of different target prediction algorithms within the Moonlit database predicting a particular target (mediating the enriched KEGG pathway)—red marking the genes predicted by multiple algorithms and white marking the genes with no predictions for a particular microRNA.

Supplementary Figure S6. Database snapshots ("KEGG" button in the "Analysis" section of the database) showing unsupervised hierarchical clustering of the disease-specific microRNA family members and the top 100 predicted targets commonly shared between the disease-specific microRNA family members in (**a**) breast, (**b**) prostate, and (**c**) neuroblastoma disease. Only gene targets that mapped to KEGG pathways are present in the list. Colour coding represents the number of different target prediction algorithms within the Moonlit database predicting a particular target.

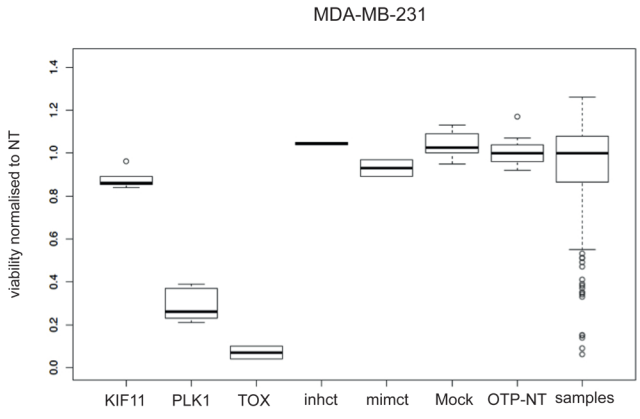
Supplementary Figure S7. MicroRNAs targeting MYCN that validated their essential function in neuroblastoma. (**a**) Unsupervised hierarchical clustering of microRNAs previously shown to be essential in MYCN-amplified neuroblastoma KELLY cell line and their predicted targets sorted according to the strength of predictions. The table was exported from the "Heatmap" tab in the Moonlit database. (**b**) Visual representation of the interconnectivity between these microRNAs and their top 25 targets. Blue circles represent individual miRs and green circles their targets. Size of the arrowheads and thickness of the connecting lines are proportional to the strength of prediction for the particular miR-target pair. The network was exported from the "Network" tab in the database.

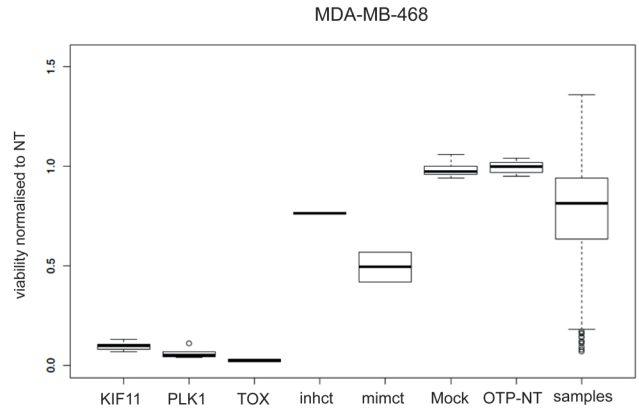
Supplementary Figure S8. Novel let-7 family targets important for neuroblastoma survival. (**a**) A snapshot of the KEGG enrichment tool page in the database. The first 100 most commonly shared targets between the MYCN-targeting microRNAs identified in our screen were investigated for the enrichment of KEGG pathways. "MicroRNAs in Cancer" and "Signalling pathways regulating pluripotency of stem cells" were the only significantly enriched pathways found. (**b**) Unsupervised hierarchical clustering of the predicted targets based on the target expression levels in a panel of analysed cancer cell lines (normalised viability cut-off <=0.3). The graph reveals a number of targets with high expression in KELLYs but lower expression in other cancer cell lines. Colour coding represents the levels of gene expression—red showing high expression level

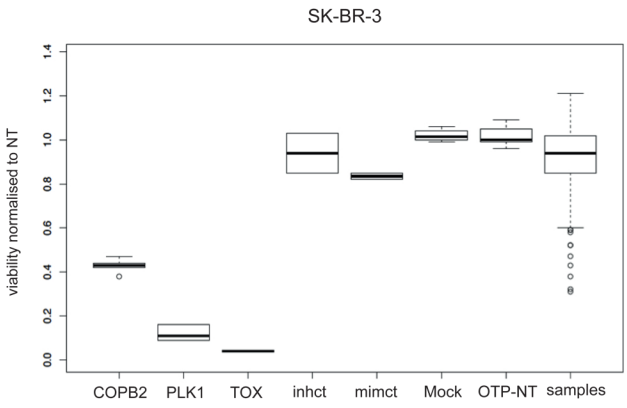
and white showing low expression level. The graph was exported from the "Expression" tab in the database.

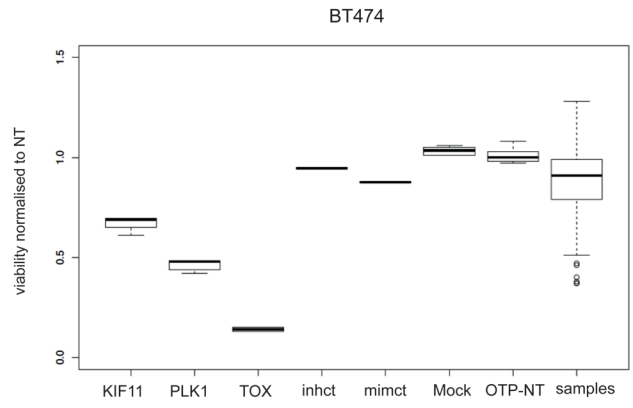
Supplementary Figure S9. Novel factors regulating neuroblastoma survival. (**a**) Unsupervised hierarchical clustering of the top most commonly shared predicted targets of the 14 miR-515 family microRNAs essential in KELLY cell line (normalised viability cut-off <=0.5) according to their expression levels in all analysed cancer cell lines. Colour coding represents the levels of gene expression—red showing high expression level and white showing low expression level. The rectangle marks the genes with higher expression in KELLY cells. The graph was exported from the "Expression" tab in the Moonlit database. (**b**) Gene expression of ISL1 in different primary cancer cohorts including multiple neuroblastoma cohorts. The graph was generated using R2 Genomic Analysis and Visulations Platform[12], and MegaSampler analysis tool. (**c**) KELLY cells were transfected with an siRNA against ISL1 or NT, and the images were taken using IncuCyte ZOOM as described in Materials and Methods.

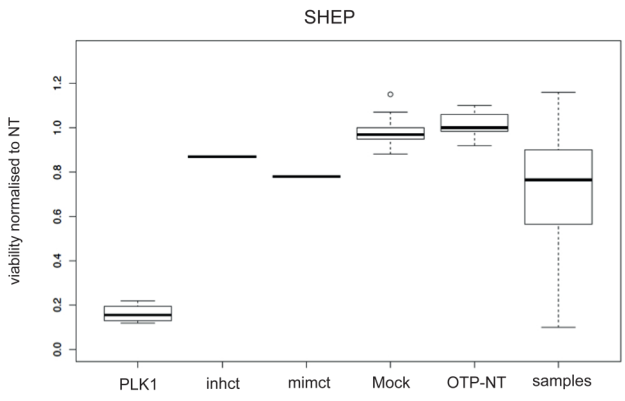
Nikolic et al. Supplementary Figure S1

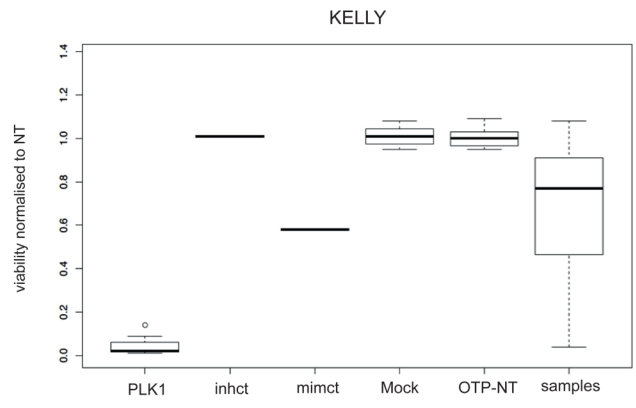


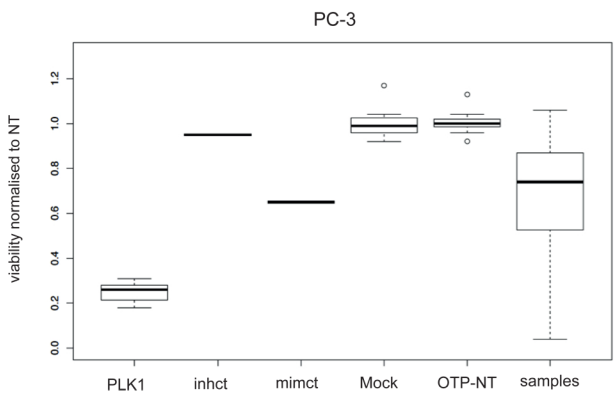


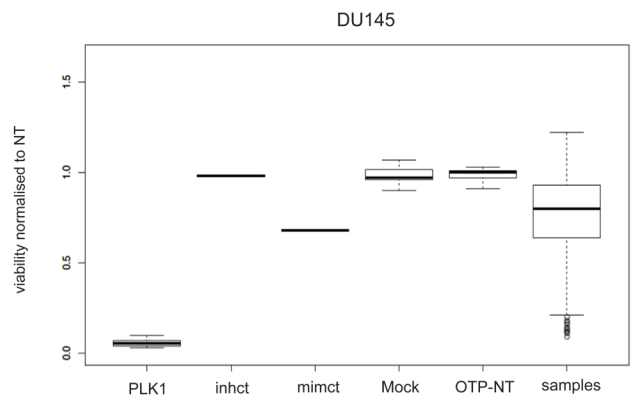




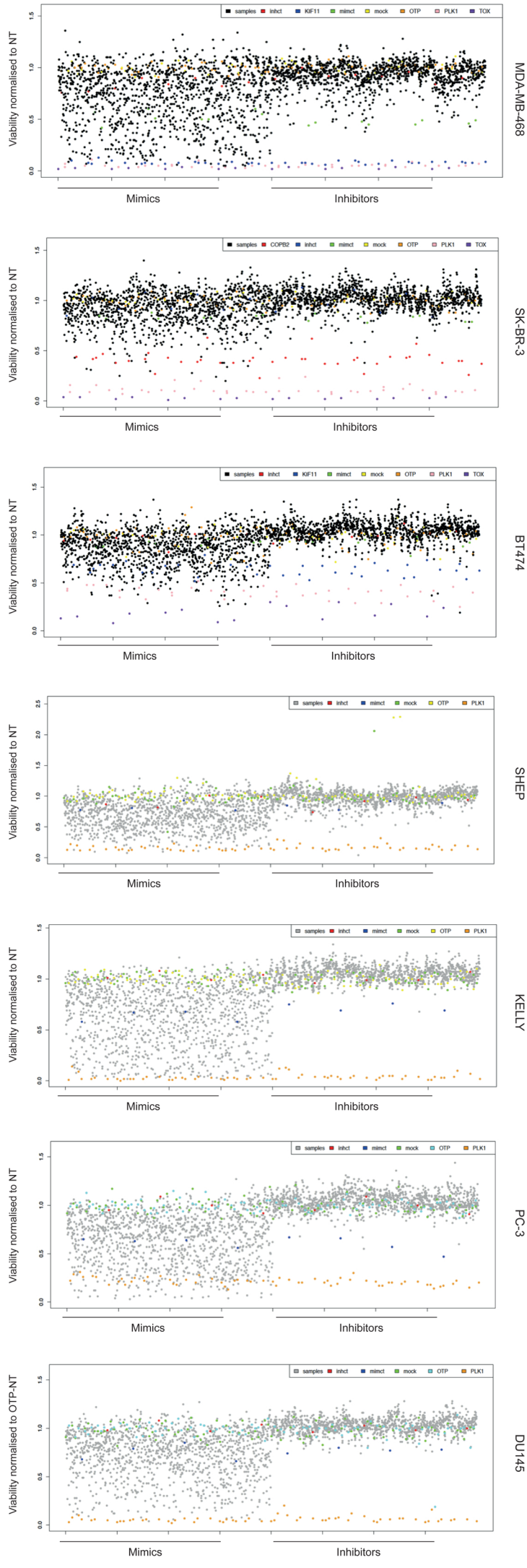






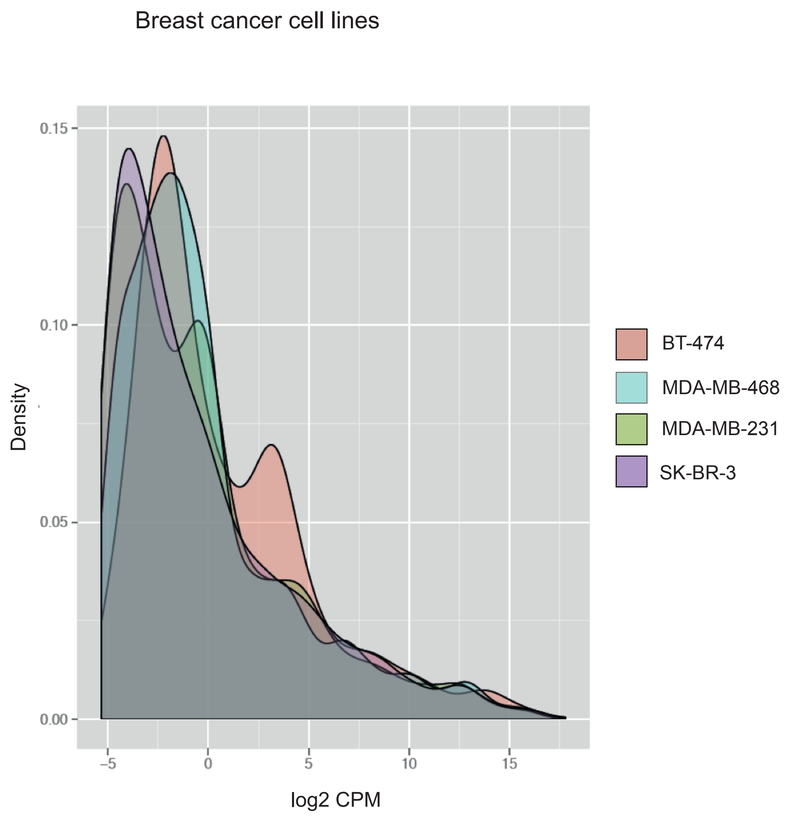


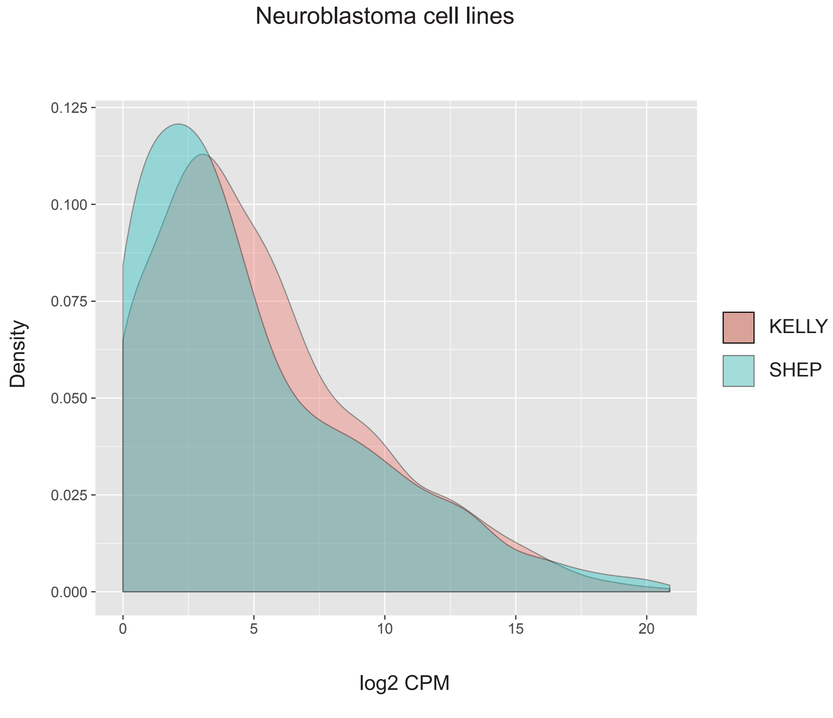
Nikolic et al. Supplementary Figure S2



Nikolic et al. Supplementary Figure S3

a





b

z-score cutoff <= 2.5 (corresponding to at least 20% killing)

Cell line	Inhibitors
MDA-MB-231	1
MDA-MB-468	64
SK-BR-3	26
BT474	29
SHEP	54
KELLY	5
PC3	10
DU145	24

Target Search results

Search Summary:

Search Results: **14 out of 14** miRNAs have target predictions | **5,613** unique targets were retrieved from the database in 6.811 seconds | **2,073** targets were filtered out.

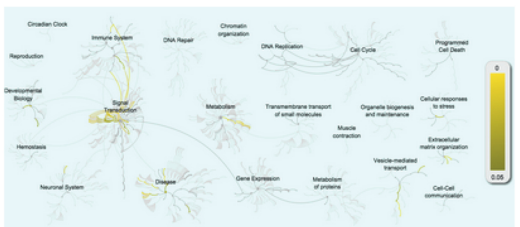
Target Counts: protein_coding - **5,405** | processed_pseudogene - **116** | antisense - **18** | lincRNA - **15** | transcribed_unprocessed_pseudogene - **13** | transcribed_processed_pseudogene - **12** | processed_transcript - **9** | miRNA - **6** | unprocessed_pseudogene - **6** | rRNA - **3** | sense_intronic - **3** | misc_RNA - **1** | snRNA - **1** | snoRNA - **1** | pseudogene - **1** | sense_overlapping - **1** | polymorphic_pseudogene - **1** | unitary_pseudogene - **1** |

Analysis:

1. Number of targets: 50

2. Analysis method

KEGG	Targets mapped to KEGG pathways
Reactome	Targets mapped to Reactome pathways
Network	Network of miRNAs and targets
Enrichr	Enrichment analysis
DAVID	Enrichment analysis
STRING	Protein-protein interactions



Target details:

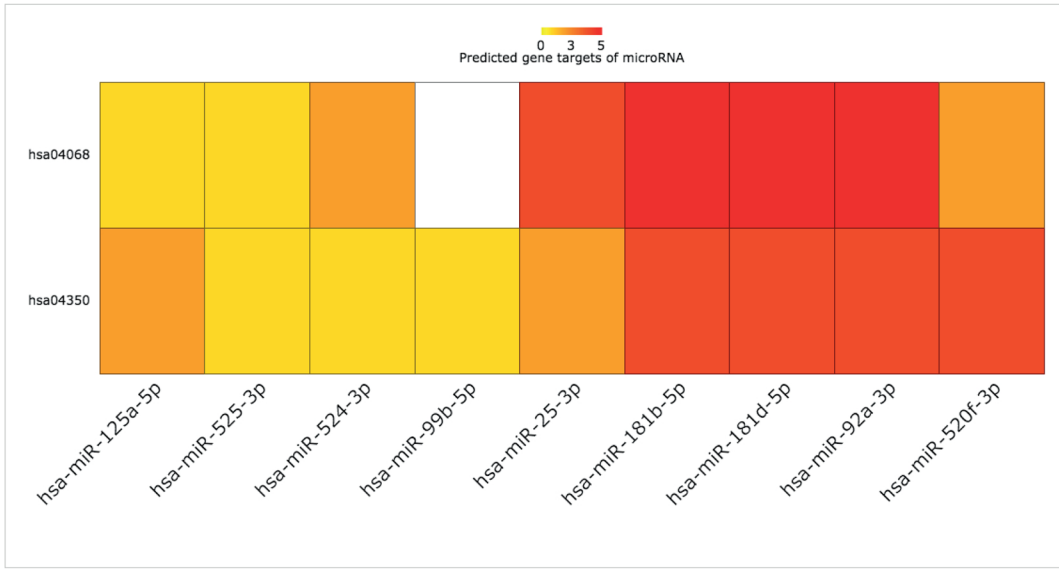
Data | Heatmap | Expression

Show 10 entries Copy | CSV | Excel | PDF | Print Filter records:

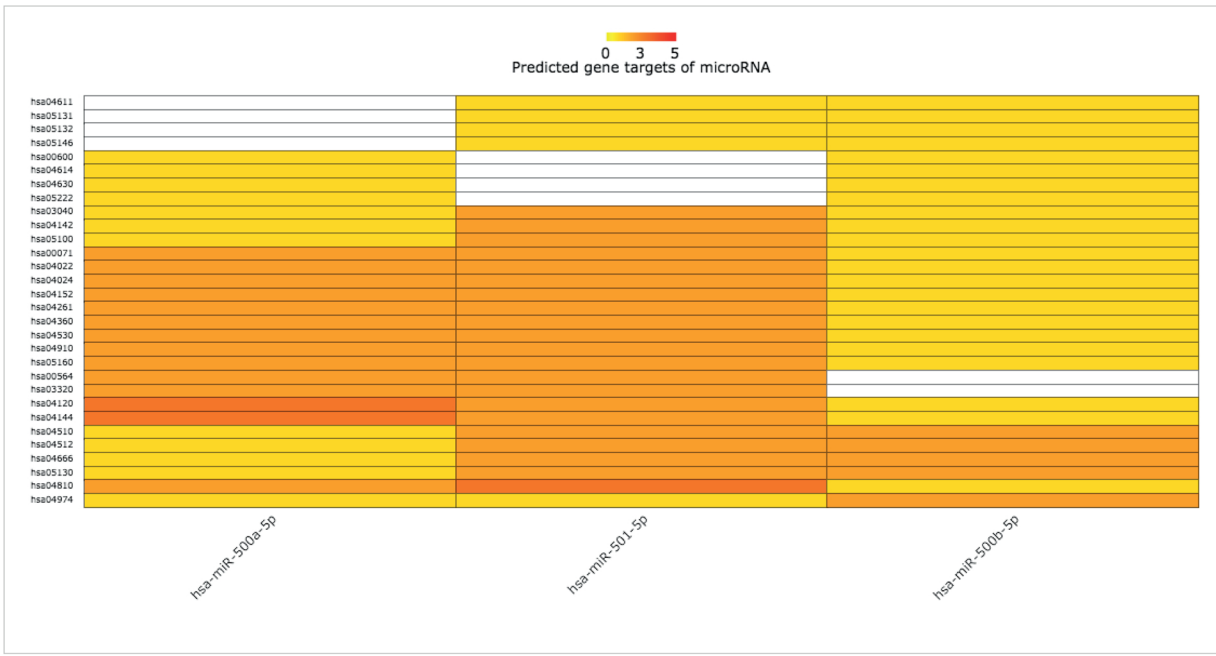
Symbol	Type	Ensembl ID	Entrez ID	Name	Location	Score	# miRNAs
FZD3	protein_coding	ENSG00000104290	7976	frizzled class receptor 3	chr8:28494205-28574268	1.30	13 / 14
GTF2I	protein_coding	ENSG00000263001	102724932	general transcription factor III	chr7:74657667-74760692	1.20	12 / 14
LIN28B	protein_coding	ENSG00000187772	389421	lin-28 homolog B (C. elegans)	chr6:104957048-105083332	1.20	12 / 14
TGFBR3	protein_coding	ENSG00000069702	7049	transforming growth factor, beta receptor III	chr1:91680343-91906335	1.20	12 / 14
ATXN1	protein_coding	ENSG00000124788	6310	ataxin 1	chr6:16299112-16761491	1.20	12 / 14
MYCN	protein_coding	ENSG00000134323	4613	v-myc avian myelocytomatosis viral oncogene neurobl ...	chr2:15940564-15947007	1.20	12 / 14
RNF44	protein_coding	ENSG00000146083	22838	ring finger protein 44	chr5:176526697-176538025	1.11	13 / 14

Nikolic et al. Supplementary Figure S5

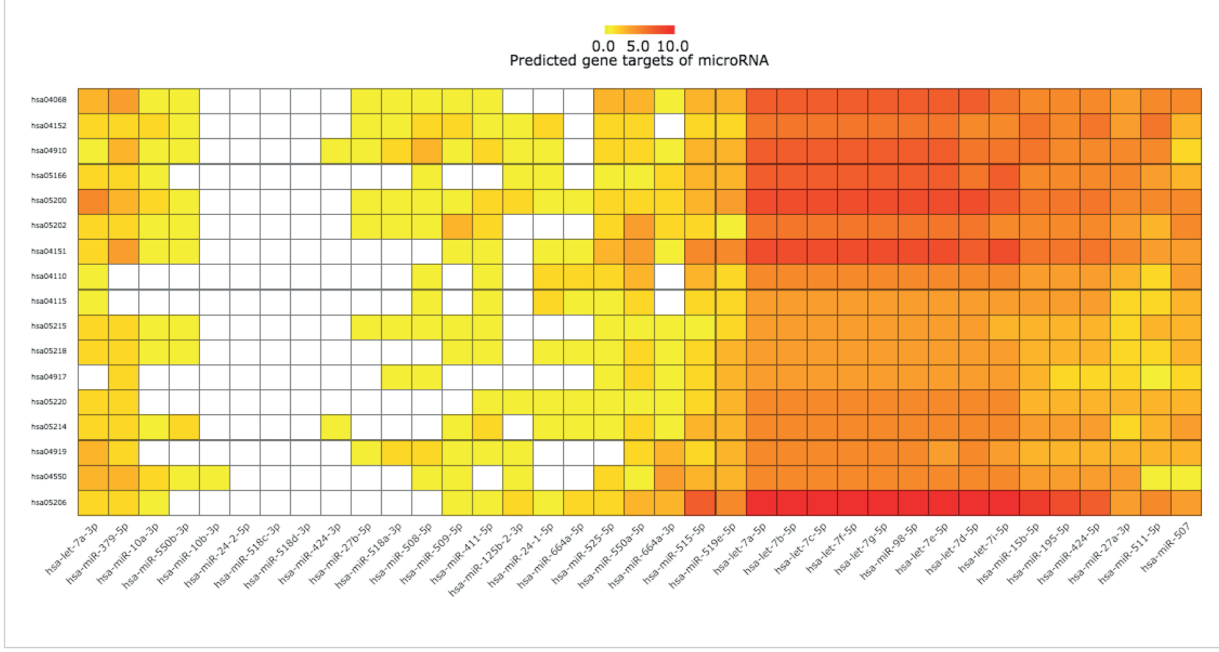
a



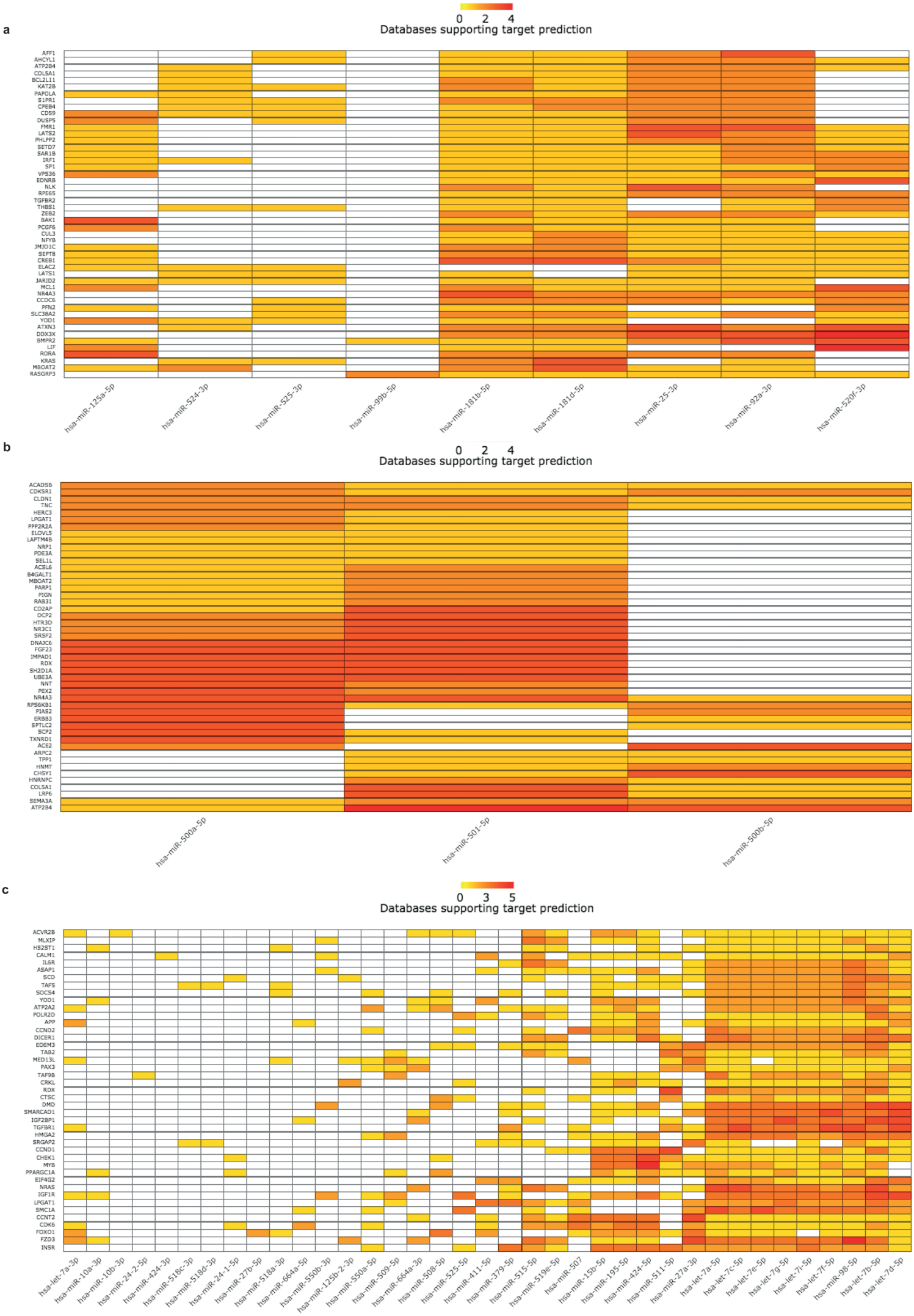
b



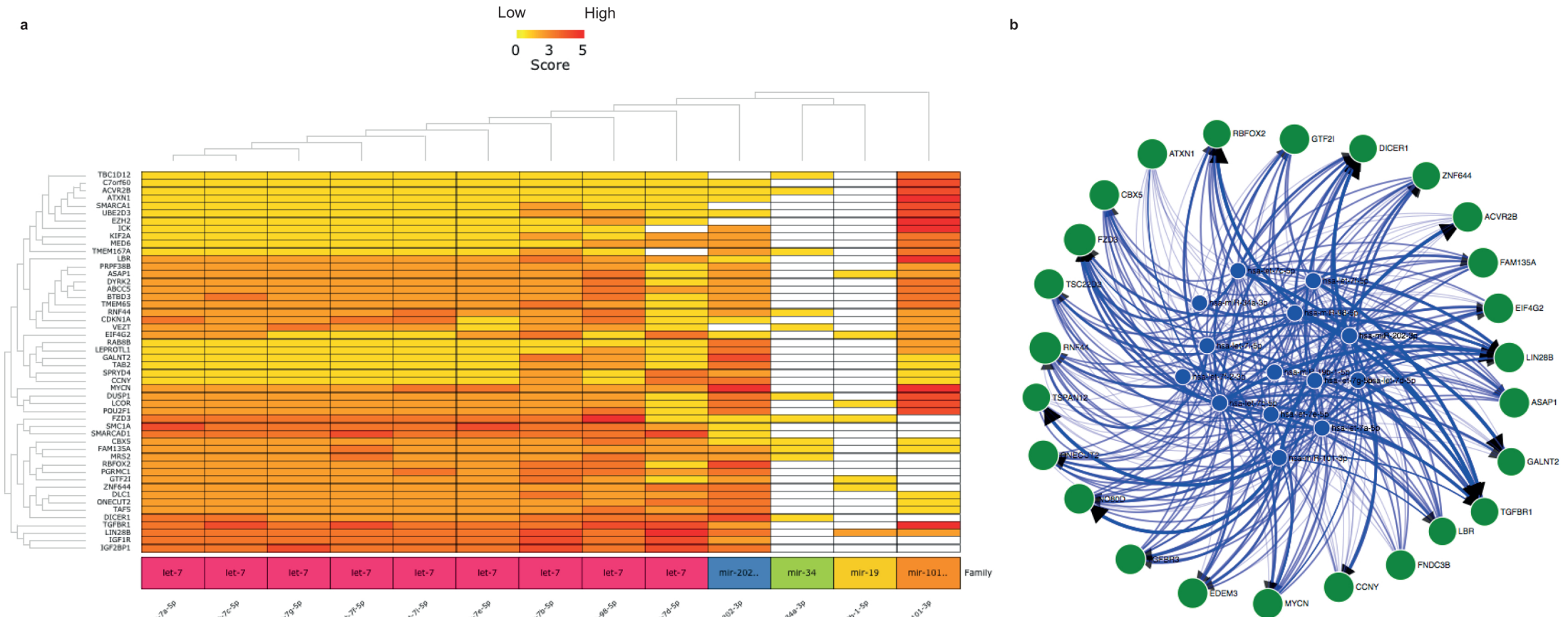
c



Nikolic et al. Supplementary Figure S6



Nikolic et al. Supplementary Figure S7



Nikolic et al. Supplementary Figure S8

a

Show 10 entries

Copy CSV Excel PDF Print

Filter records:

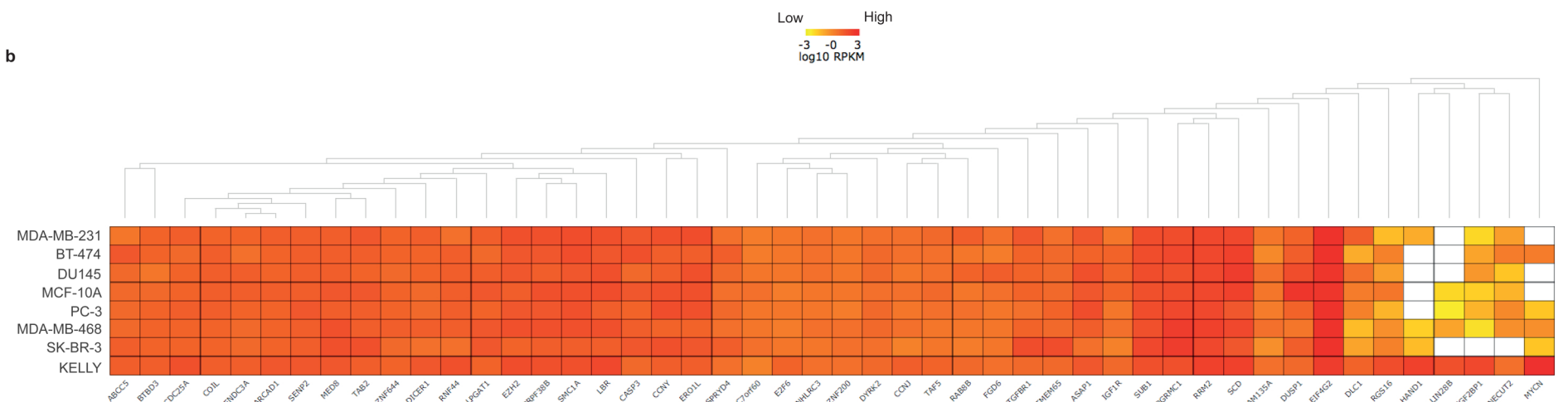
KEGG ID	Pathway name	# miRs (out of 13)	# genes matched	# genes in pathway	% genes	FET P-Value	Odds Ratio	Benjamini-Hochberg correction
hsa05206	MicroRNAs in cancer	13	8	132	6.45	2.00e-05	7.83	1.26e-03
hsa04550	Signaling pathways regulating pluripotency of stem cells	13	7	91	8.33	1.44e-05	10.04	1.81e-03
hsa05168	Herpes simplex infection	11	5	110	4.76	0.0028	5.61	0.1170
hsa05214	Glioma	11	3	49	6.52	0.0090	7.56	0.1422
hsa05220	Chronic myeloid leukemia	11	3	64	4.92	0.0186	5.69	0.1463
hsa04010	MAPK signaling pathway	12	5	173	2.98	0.0180	3.49	0.1515
hsa05215	Prostate cancer	11	3	67	4.69	0.0210	5.43	0.1554
hsa05161	Hepatitis B	11	4	99	4.21	0.0111	4.91	0.1559
hsa04726	Serotonergic synapse	12	3	61	5.17	0.0164	5.99	0.1585
hsa05166	HTLV-I infection	13	5	172	2.99	0.0176	3.51	0.1587

Showing 1 to 10 of 126 entries

First Previous 1 2 3 4 5 Next Last

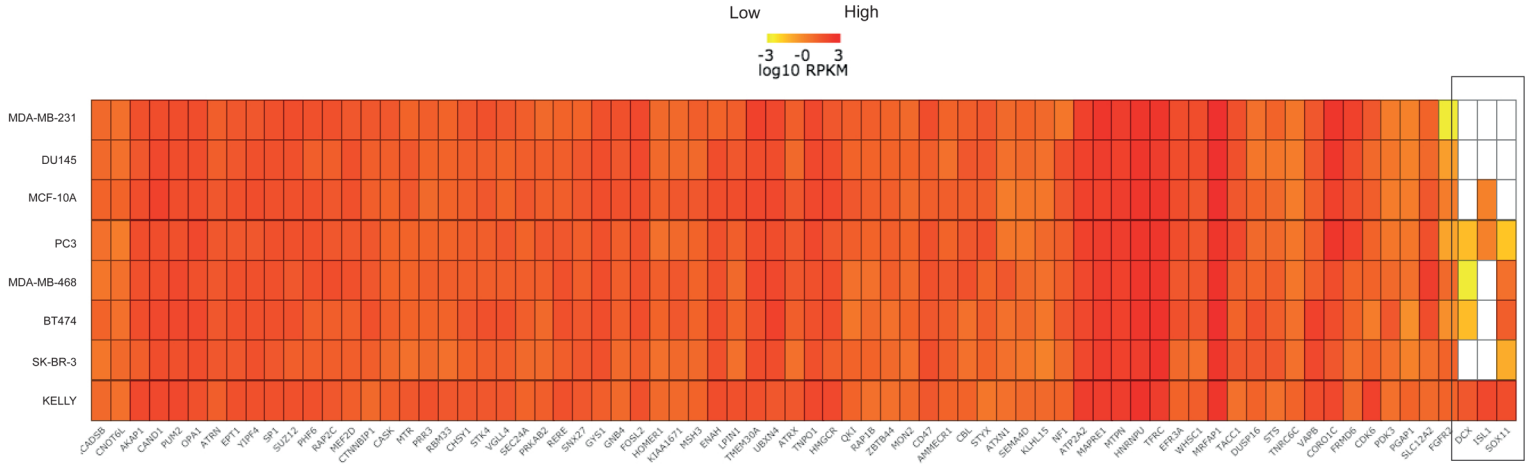
Contact: a.swarbrick@garvan.org.au

b

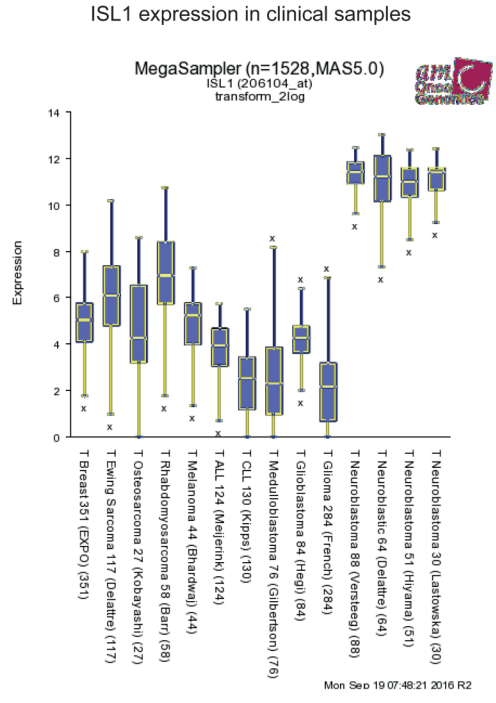


Nikolic et al. Supplementary Figure S9

a



b



c