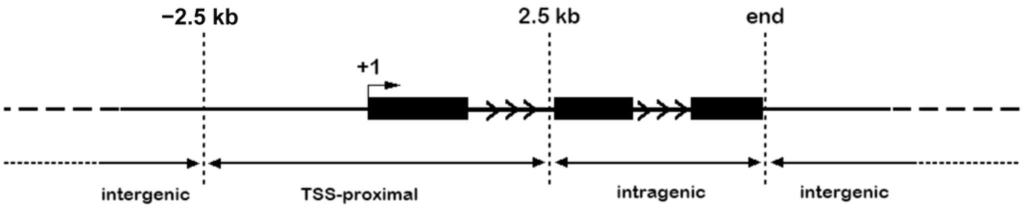
The *MYC*, *TERT*, and *ZIC1* genes Are Common Targets of Viral Integration and Transcriptional Deregulation in Avian Leukosis Virus Subgroup J-Induced Myeloid Leukosis

Yuhao Li,[1, 2] Xuemei Liu,[1, 2] Zhen Yang,[1, 2] Chenggang Xu,[1, 2] Di Liu,[1, 2] Jianru Qin,[1, 2] Manman Dai,[1, 2] Jianyong Hao,[1, 2] Min Feng,[1, 2] Xiaorong Huang,[1, 2] Liqiang Tan,[1, 2] Weisheng Cao[1, 2#], and Ming Liao[1, 2#]

National and Regional Joint Engineering Labaratory for Medicment of Zoonoses Prevention and Control,[1] and Key Laboratory of Veterinary Vaccine Innovation of the Ministry of Agriculture,[2] College of Veterinary Medicine, South China Agricultural University, Guangzhou 510642, the People's Republic of China.

Supplementary Item	**Title or Caption**
Supplementary Figure 1.	Genomic distribution of ALV-J integration sites
Supplementary Figure 2.	GSEA analyses identified gene classes that were significantly up-regulated or down-regulated in ML groups.
Supplementary Table 1.	Primer sequences.
Supplementary Table 2.	Next-generation sequencing statistics.
Supplementary Table 3.	Characterization of 241 ALV-J integration sites from ML samples.
Supplementary Table 4.	Validation of the ALV-J integration sites..
Supplementary Table 5.	GSEA over-represented molecular pathways and functions in ML groups with respect to non-tumoral samples
Supplementary Note 1.	List of genes and their gene classes that were significantly deregulated in different ML groups.



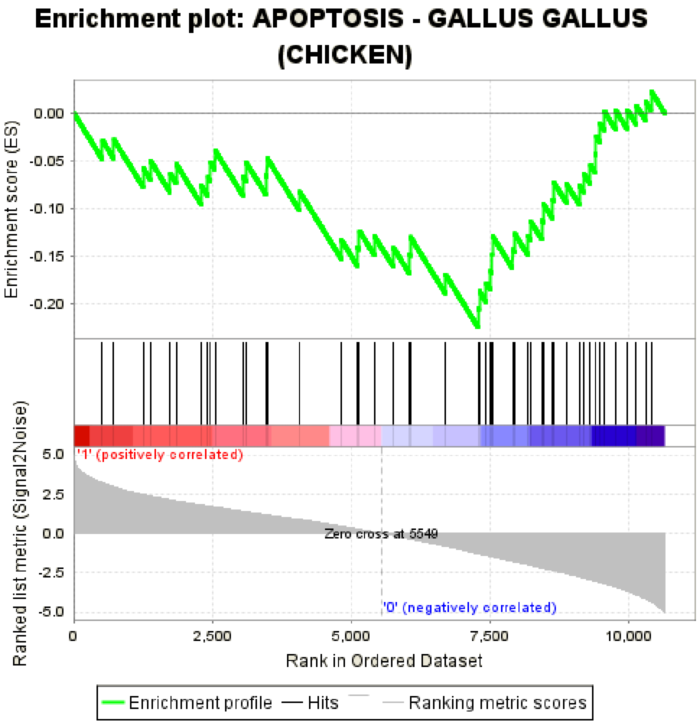
Supplementary Figure 1. Genomic distribution of ALV-J integration sites

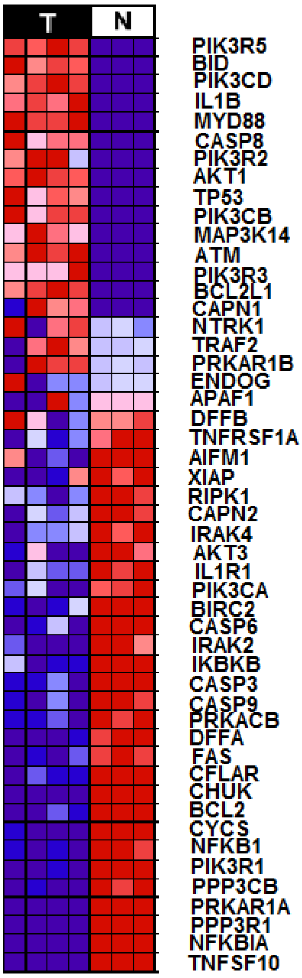
ALV-J integration sites were annotated as TSS-proximal when located at ±2.5 kb from a transcription start site (TSS) of a known gene (UCSC definition), intragenic when inside a gene at 2.5 kb from the TSS, and intergenic in any other case. Black bars represent exons of a schematic gene, arrowheads indicate the direction of transcription.

a.

Apoptosis
tumor versus normal

NES	-1.8794032
Nominal p-value	0.012244898
FDR q-value	0.03823706
FWER p-Value	0.452

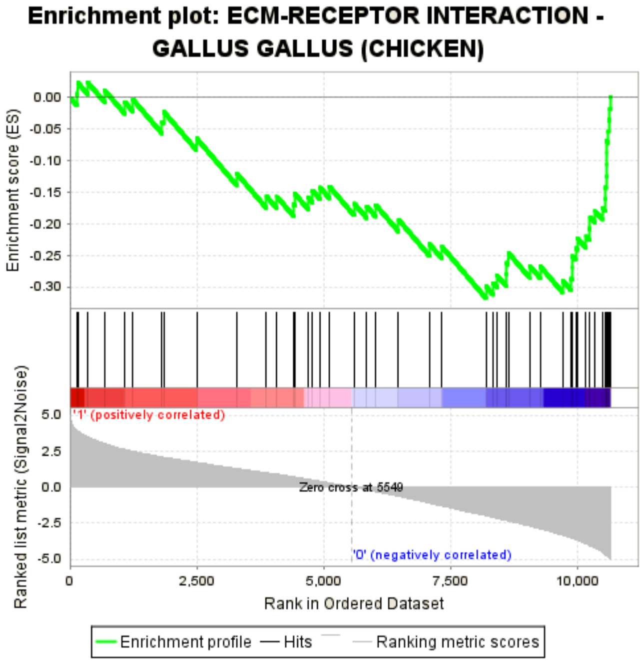


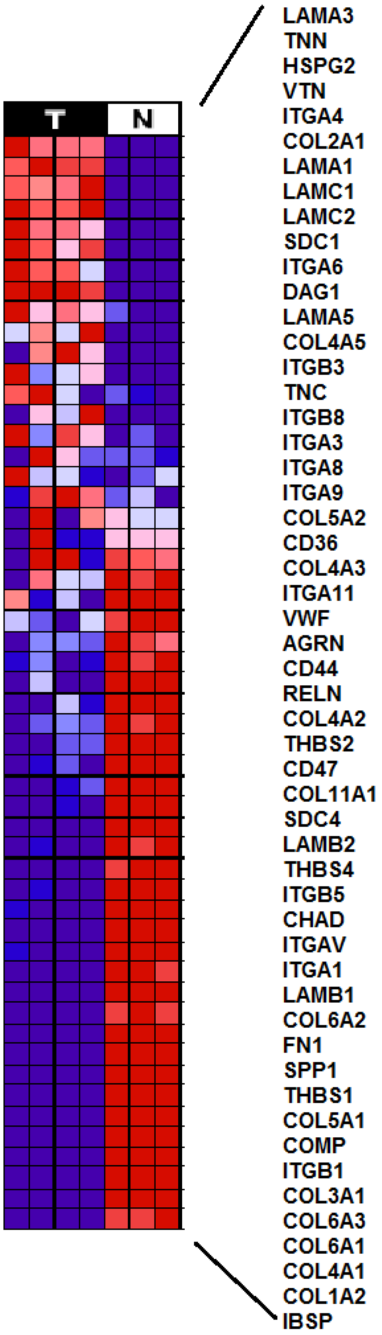


b.

ECM-recepter interaction
tumor versus normal

NES	-2.7174306
Nominal p-value	0.0
FDR q-value	5.679598E-4
FWER p-Value	0.0020

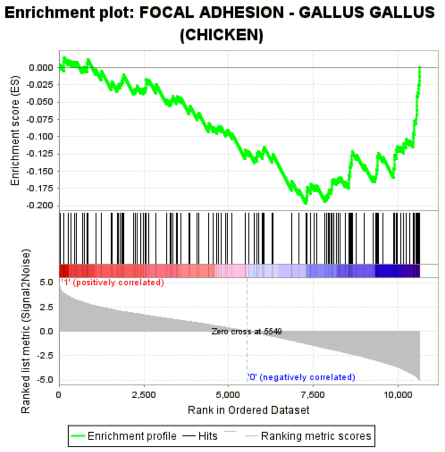


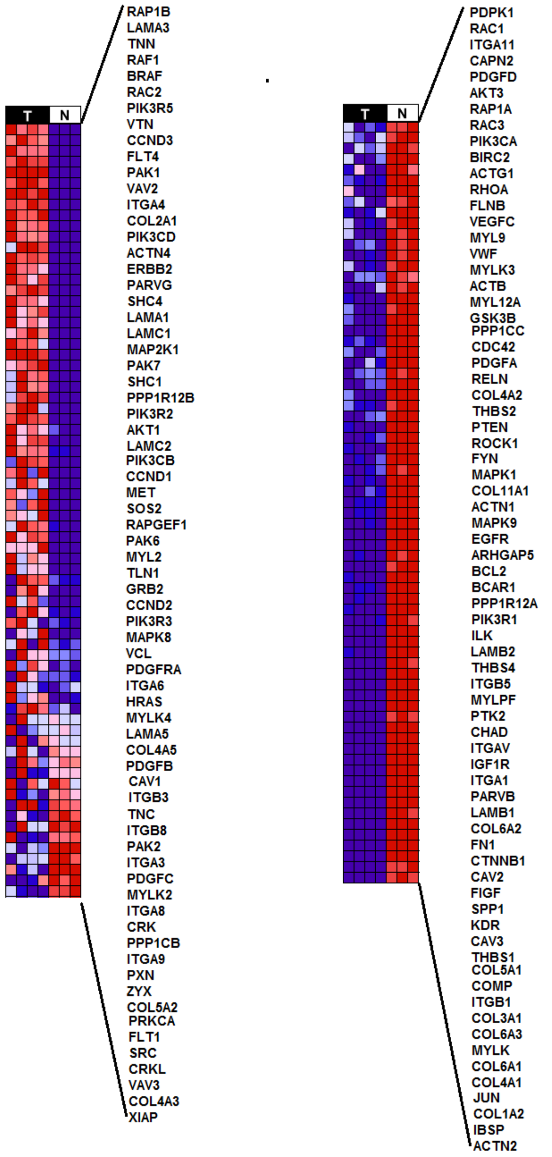


c.

Focal adhesion
tumor versus normal

NES	-2.7084944
Nominal p-value	0.0
FDR q-value	4.5436784E-4
FWER p-Value	0.0020

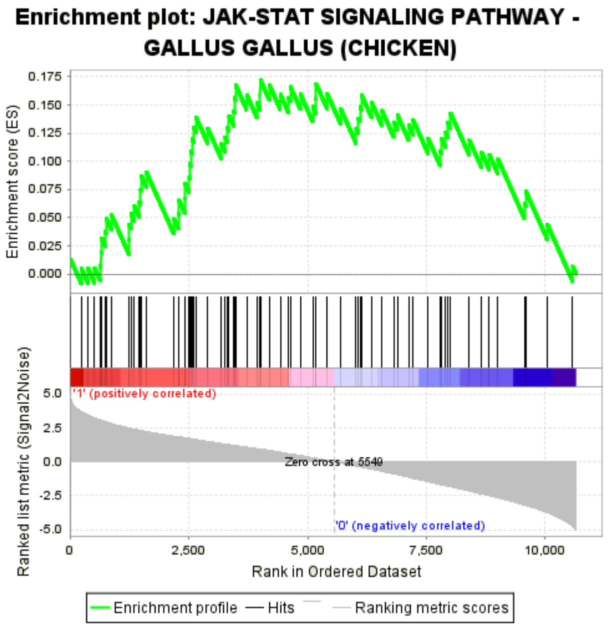


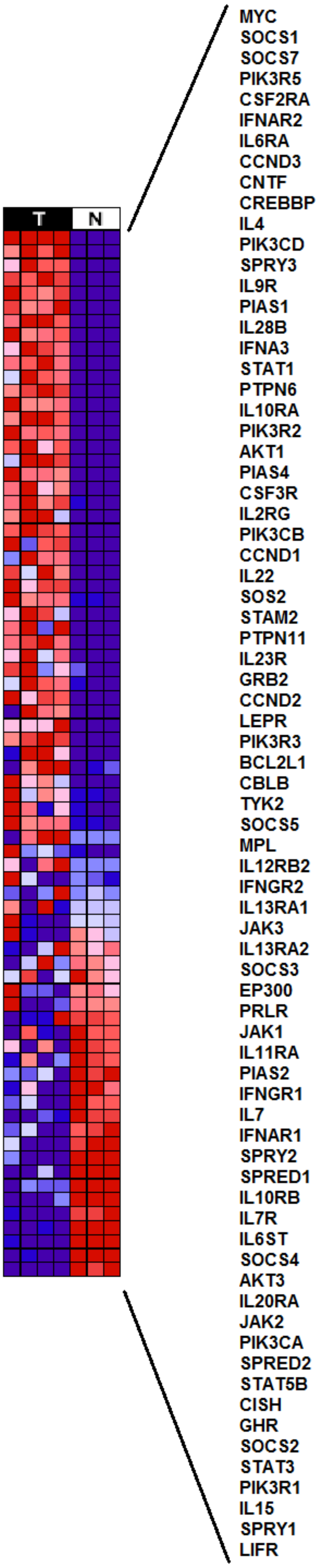


d.

JAK-STAT signaling pathway
tumor versus normal

NES	1.7702438
Nominal p-value	0.024952015
FDR q-value	0.28014386
FWER p-Value	0.663





Supplementary Figure 1. GSEA analyses identified gene classes that were significantly up-regulated or down-regulated in ML groups.

Expression profiles of ML groups were compared with respect to three pooled non-tumoral tissues from the different genetic backgrounds by gene set enrichment analysis (GSEA) (see also Supplementary Table 5). On the left, a table summarizes the GSEA statistics (NES: normalized enrichment score; FDR q-value: false discovery rate; FWER p-Value: family-wise error rate; for values <0.01, 0.0 is indicated). The enrichment plot shows the over-representation at the top/bottom of the ranked gene set. The heat map shows all significant differentially-expressed genes of the given gene set between groups (N: normal tissue; T: tumor) (from blue = low expression to red = high expression). Gene symbols are shown on the right. (a–c) Genes that were down-regulated for apoptosis, ECM-receptor interaction, and focal adhesion were found in all the ML groups. (d) Genes of the JAK-STAT signaling pathway were significantly up-regulated in ML groups.

Supplementary Tables 1-7 are provided as Excel files.

Supplementary Table 1. Primer sequences.

Sequences of primers used to perform qRT-PCR for the detection of ALV-J-induced MLs gene expression level.

Supplementary Table 2. Next-generation sequencing statistics

Data represent averages of duplicate runs. Reads matches to the ALV-J target were determined by paired-end alignment of all raw sequence data to the ALV-J genome using the Burrows-Wheeler Aligner and Short Oligonucleotide Analysis software package. ‘Reads mapped to viral genome’ was calculated by comparing the sequenced frequency of ALV-J sequences compared to the total reads. ‘Reads mapped to chicken genome’ represents the average number of sequenced reads that were fully matched to the chicken genome.

Supplementary Table 3. Characterization of 241 ALV-J integration sites from ML samples

Each row represents an independent integration. (a) ALV-J-induced chicken tumor identifier. (b) Sequence identifier from Illumina HiSeq2000 pyrosequencing. (c) Matching identity of the raw sequence reads and the UCSC galGAL4 genomic sequences. (d) Chromosome to which the integration maps. (e) Position of the integration site on the chromosome in nucleotides. (f) Orientation of ALV-J with respect to the direction of transcription of the targeted gene, indicated as forward or reverse. (g) Ref Seq gene identifier of the gene closest to ALV-J integration. (h) Ref Seq gene symbol of the gene closest to ALV-J integration. (i) Distribution of ALV-J integration sites in the genome of ML samples. Integrations were annotated for each patient as ‘TSS-proximal’ when they occurred within a distance of ±2.5 kb from the TSS of at least one gene, as ‘intragenic’ when inserted into at least one gene at a distance of 2.5 kb from the TSS, and as ‘intergenic’ in all other cases (see Figure 3a). Integration sites are sorted for Chromosome and Integration Locus.

Supplementary Table 4. Validation result for the ALV-J integration sites.

Twenty-three integration sites at the six affected genes for PCR analysis. Verification of ALV-J integrations with respect to the sequence of the targeted gene identifier from Illumina HiSeq2000 pyrosequencing, indicated as yes or no.

Supplementary Table 5. GSEA over-represented molecular pathways and functions in ML groups with respect to non-tumoral samples.

Over-represented gene-sets of Cellular Process by GSEA using setting: Real. Dataset: expression levels of ML groups were compared with respect to the expression of pooled non-tumoral samples; Trend: up-regulation (Up) or down-regulation (Down) of genes belonging to a given pathway is indicated; Gene Set: Description of significantly over-represented gene sets belonging to specific Canonical Pathway; Size: size of the given Canonical Pathways gene set. ES: Enrichment Score; NES:

Normalized Enrichment Score; NOM p-val: Nominal *P*-value; FDR q-val: False Discovery Rate; FWER p-val: Family Wise Error Rate corrected *P*-value.

Supplementary Note 1. List of genes belonging to gene classes significantly deregulated in different ML groups.

We used GSEA analyses to identify gene classes that were significantly up-regulated or down-regulated in ML groups. In **Figure 5b** we showed the heat map representation of the over-expressed genes of Oxidative Phosphorylation in MLs. The list of genes showed in the heat-map, from top to bottom is:

PROBE	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
ATP6V0E2	948	2.721667	-0.07928894	No
COX7A2	1600	2.239329	-0.13043955	No
TCIRG1	1850	2.06686	-0.14350395	No
ATP6V0A4	2579	1.671428	-0.20194969	No
NDUFA10	2827	1.52627	-0.2148246	No
COX8A	2927	1.478789	-0.21367773	No
ND6	3149	1.356025	-0.22408935	No
ATP6V1G3	3548	1.155021	-0.2512703	No
NDUFA9	3881	0.992072	-0.27219826	No
COX15	3884	0.991223	-0.26186144	No
ND4L	4509	0.655861	-0.310454	No
ATP8	4654	0.56324	-0.31357053	No
ND1	4672	0.553792	-0.3046548	No
ND5	4944	0.397635	-0.31980354	No
ATP6V0D1	5153	0.268774	-0.32898352	No
COX1	5300	0.17753	-0.33228952	No
ND2	5365	0.132001	-0.32782668	No
COX2	5376	0.125711	-0.31824777	No
ATP6	5493	0.043668	-0.31871152	No
ND3	5553	-0.00408	-0.31377497	No
NDUFS5	5744	-0.11414	-0.3212496	No
ATP6V0A1	5767	-0.1311	-0.3128076	No
COX3	5865	-0.20134	-0.31147125	No
PPA1	5915	-0.23982	-0.30558726	No
CYTB	5938	-0.25337	-0.29714528	No
ATP6V1H	6020	-0.31378	-0.29429305	No
LHPP	6242	-0.46544	-0.30470467	No
ND4	6290	-0.50044	-0.29863122	No
NDUFV3	6441	-0.6074	-0.3023162	No
NDUFB4	6486	-0.64103	-0.29595852	No
COX11	6503	-0.65563	-0.28694806	No
PPA2	6695	-0.80102	-0.29451743	No
COX10	7127	-1.13599	-0.32482484	No

ATP6V1B2	7308	-1.29947	-0.33135206	No
ATP5G3	7313	-1.30206	-0.32120472	No
NDUFB2	7630	-1.54408	-0.34061682	No
ATP6V0E1	7684	-1.5992	-0.33511183	No
COX7A2L	8114	-1.9129	-0.36522976	Yes
NDUFB8	8188	-1.98429	-0.3616196	Yes
ATP6V1A	8217	-2.00365	-0.35374603	Yes
COX5A	8308	-2.07536	-0.3517465	Yes
NDUFA11	8364	-2.10563	-0.34643096	Yes
ATP6V1E1	8615	-2.3288	-0.3595901	Yes
ATP6V0A2	8780	-2.45829	-0.36460146	Yes
NDUFS3	8824	-2.49051	-0.35814905	Yes
NDUFA12	8865	-2.53064	-0.3514124	Yes
COX6A1	8868	-2.53283	-0.34107557	Yes
NDUFB3	8881	-2.55185	-0.33168614	Yes
ATP6V0C	8883	-2.55293	-0.32125458	Yes
NDUFC2	8895	-2.56272	-0.3117704	Yes
COX7C	8897	-2.56628	-0.30133885	Yes
ATP5A1	8925	-2.58488	-0.29337054	Yes
ATP6V1D	8948	-2.60505	-0.28492856	Yes
ATP5B	8976	-2.62215	-0.27696028	Yes
NDUFA2	8989	-2.63318	-0.26757085	Yes
UQCRB	9060	-2.69661	-0.26367646	Yes
ATP6V1G1	9082	-2.71258	-0.25513974	Yes
UQCRC1	9116	-2.75001	-0.2477399	Yes
ATP5J	9138	-2.77487	-0.23920316	Yes
ATP6V0D2	9158	-2.80334	-0.23047695	Yes
NDUFS4	9167	-2.80883	-0.22070856	Yes
ATP5J2	9186	-2.82266	-0.2118876	Yes
NDUFB6	9251	-2.88011	-0.20742476	Yes
NDUFS8	9280	-2.90471	-0.19955121	Yes
ATP6AP1	9380	-3.00189	-0.19840434	Yes
ATP5I	9421	-3.04073	-0.19166769	Yes
COX4I1	9425	-3.0424	-0.1814256	Yes
NDUFAB1	9460	-3.08464	-0.1741205	Yes
NDUFV1	9506	-3.1366	-0.16785757	Yes
ATP5F1	9510	-3.14174	-0.15761548	Yes
ATP5C1	9539	-3.16698	-0.14974193	Yes
ATP5A1W	9619	-3.25551	-0.14670023	Yes
ATP6V1C2	9623	-3.25652	-0.13645813	Yes
NDUFA7	9698	-3.32849	-0.13294272	Yes
UQCRQ	9762	-3.3862	-0.12838513	Yes
ATP5H	9785	-3.40937	-0.11994313	Yes
UQCRH	9817	-3.45424	-0.11235382	Yes

NDUFA8	9822	-3.46131	-0.10220647	Yes
NDUFA4	9836	-3.47879	-0.0929118	Yes
SDHD	9847	-3.49228	-0.0833329	Yes
NDUFA1	9850	-3.49516	-0.07299607	Yes
NDUFB1	9869	-3.52262	-0.0641751	Yes
UQCRC2	9883	-3.54112	-0.05488043	Yes
NDUFA6	9987	-3.65515	-0.05411252	Yes
NDUFB5	9993	-3.66592	-0.04405992	Yes
SDHB	10000	-3.67091	-0.03410205	Yes
ATP5O	10074	-3.77651	-0.03049189	Yes
SDHA	10119	-3.8581	-0.02413421	Yes
COX6C	10179	-3.95012	-0.01919766	Yes
ATP5G1	10223	-4.02997	-0.01274525	Yes
NDUFA5	10263	-4.0866	-0.00591386	Yes
NDUFS6	10267	-4.09482	0.00432823	Yes
NDUFB9	10327	-4.19777	0.009264777	Yes
NDUFB10	10419	-4.36146	0.011169587	Yes
NDUFS1	10489	-4.49781	0.015158718	Yes

In **Figure 5c** we showed the heat map representation of the most up-regulated (top) and down-regulated (bottom) genes of the ERBB Signaling Pathway in ML tumors vs. normal samples. The list of the most up-regulated genes showed in the heat-map from top to bottom is:

PROBE	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
MYC	0	4.996099	0.018182	Yes
CIP1	200	3.772469	0.017581	Yes
RAF1	258	3.657594	0.030383	Yes
BRAF	355	3.461406	0.039504	Yes
ERBB3	420	3.36342	0.051645	Yes
TGFA	461	3.305974	0.066052	Yes
PIK3R5	510	3.228925	0.079703	Yes
GAB1	600	3.121755	0.089485	Yes
PAK1	821	2.865258	0.086902	Yes
PIK3CD	1243	2.480062	0.065348	Yes
ERBB2	1560	2.260876	0.053705	Yes
SHC4	1749	2.131556	0.054142	Yes
MAP2K1	1872	2.056336	0.060809	Yes
PAK7	1899	2.037929	0.076537	Yes
NRAS	2076	1.942731	0.078107	Yes
SHC1	2202	1.87839	0.084491	Yes
PIK3R2	2290	1.836239	0.094461	Yes
AKT1	2410	1.767826	0.101411	Yes
PIK3CB	2541	1.692321	0.107323	Yes
SOS2	2649	1.631977	0.115406	Yes
CAMK2D	2826	1.526355	0.116976	Yes
PAK6	2863	1.505315	0.13176	Yes
CAMK2B	3044	1.41424	0.132953	Yes
CAMK2A	3286	1.288265	0.128388	Yes
GRB2	3314	1.275653	0.144022	Yes
MAP2K4	3434	1.216933	0.150972	Yes
PIK3R3	3457	1.205967	0.167077	Yes
MAPK8	3495	1.188499	0.181767	Yes
CBLB	3712	1.07358	0.179562	Yes
HBEGF	3742	1.056185	0.195006	Yes
HRAS	4076	0.901284	0.181758	No
PAK2	4973	0.381219	0.115372	No
NRG1	5238	0.213532	0.108636	No
CRK	5770	-0.13483	0.0767	No
NRG4	5897	-0.23087	0.082989	No
EIF4EBP1	5940	-0.25432	0.097207	No
PRKCA	6045	-0.33074	0.105573	No

RPS6KB1	6183	-0.4269	0.110824	No
NCK2	6295	-0.50276	0.118529	No
SRC	6304	-0.51052	0.135956	No
CRKL	6860	-0.92467	0.101755	No
MAP2K2	7099	-1.11737	0.097473	No
KRAS	7224	-1.22523	0.103951	No
AKT3	7535	-1.46593	0.092874	No
PIK3CA	7903	-1.7552	0.076417	No
STAT5B	8005	-1.82755	0.085066	No
GSK3B	8545	-2.27235	0.052375	No
ERBB4	8643	-2.34714	0.061401	No
BTC	8667	-2.37286	0.077412	No
MAPK1	9008	-2.65165	0.063503	No
MAPK9	9361	-2.97883	0.048462	No
EGFR	9367	-2.98745	0.066172	No
PIK3R1	9558	-3.1831	0.066421	No
PTK2	9959	-3.62362	0.046849	No
JUN	10629	-4.91332	0.001888	No