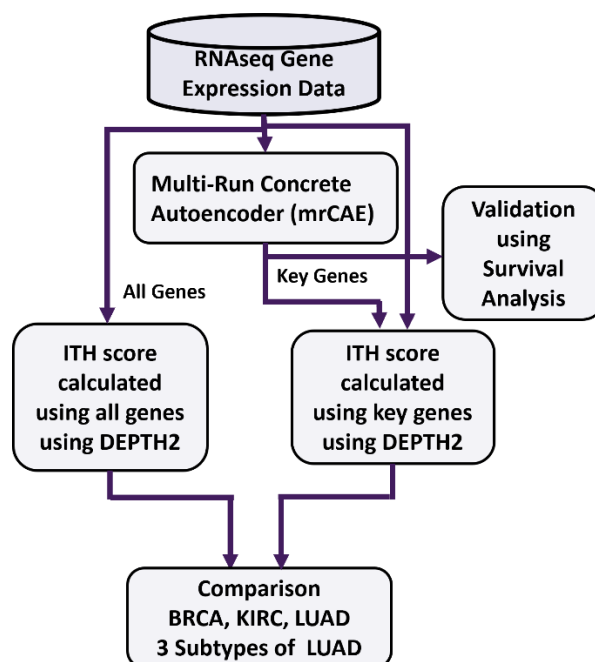
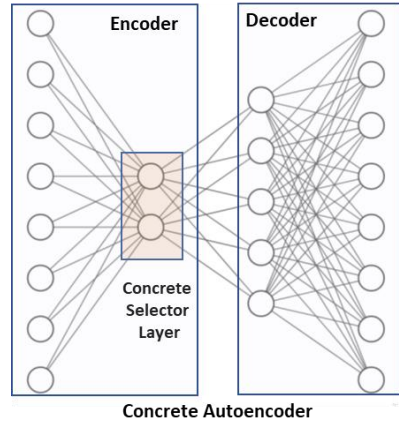


**Supplementary Materials of**

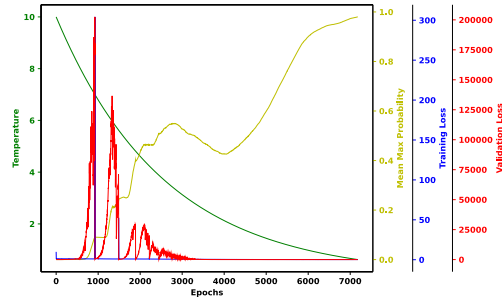
**Quantifying Intratumor Heterogeneity by Key Genes  
Selected Using Concrete Autoencoder**



**Supplementary Fig. 1.** The overall framework of this study. RNAseq gene expression data of 20K genes was used in input and multi-run concrete autoencoder was used to select a subset of key genes from 20K genes. The DEPTH2 was used to calculate ITH scores using all genes and key genes. Then ITH levels calculated from these two approaches were compared for three cancer cohorts (BRCA, KIRC, LUAD) and three subtypes of LUAD. Also, the selected key genes were used in survival analysis to identify prognostically significant genes.



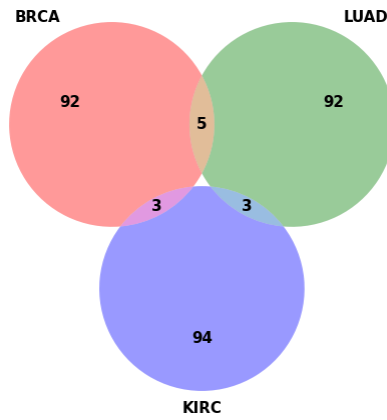
**Supplementary Fig. 2.** The general architecture of CAE. The encoder part contains the input layer and the concrete selector layer. The decoder in this figure is a 2-layer neural network, with the final layer having the same number of nodes as the input layer.



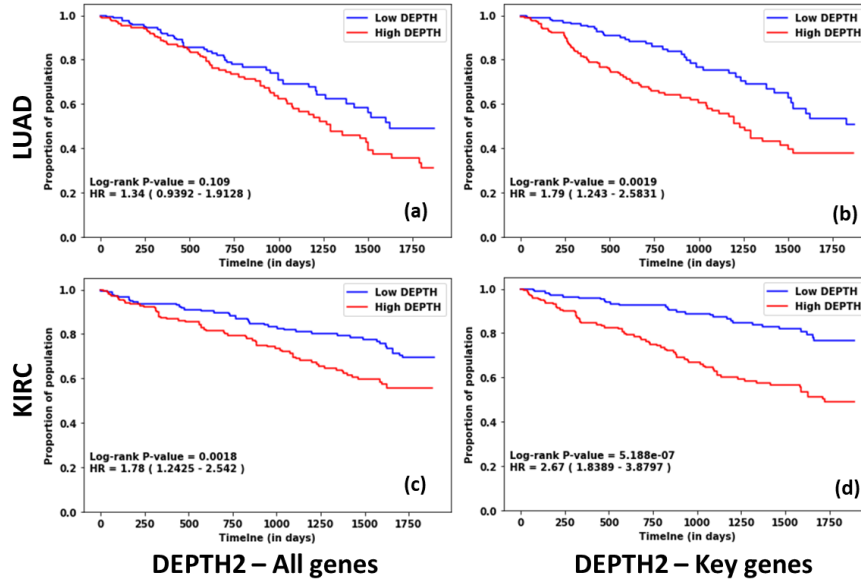
**Supplementary Fig. 3.** The characteristics curve of CAE. Reconstruction errors with training data (blue) and validation data (red) were plotted. The green and yellow curve represent the values of Temperature (T) and mean-max probability, respectively.

**Supplementary Table 1.** The hyperparameters and its range of values for concrete autoencoder. The selected values in some cases are marked as bold.

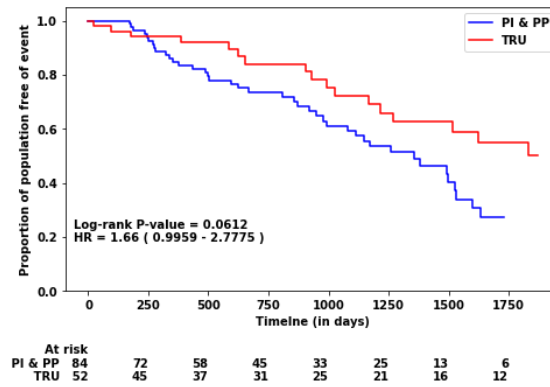
Item	Value
Activation Function	LeakyReLU
Dropout Rate	10%
Temperature	10 to 0.1
Decoder Architecture	[(200, 200), ( <b>300, 300</b> ), (400, 400)]
Learning Rate	[0.001, <b>0.002</b> , 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01]
Epochs Selected	3500 (BRCA), 9000 (KIRC), 8000 (LUAD)
Mean-max probability Threshold	0.98
Average # of epochs to reach mean-max probability $\geq 0.98$	3066 (BRCA), 8191 (KIRC), 7250 (LUAD)



**Supplementary Fig. 4.** The characteristics curve of CAE. Reconstruction errors with training data (blue) and validation data (red) were plotted. The green and yellow curve represent the values of Temperature (T) and mean-max probability, respectively



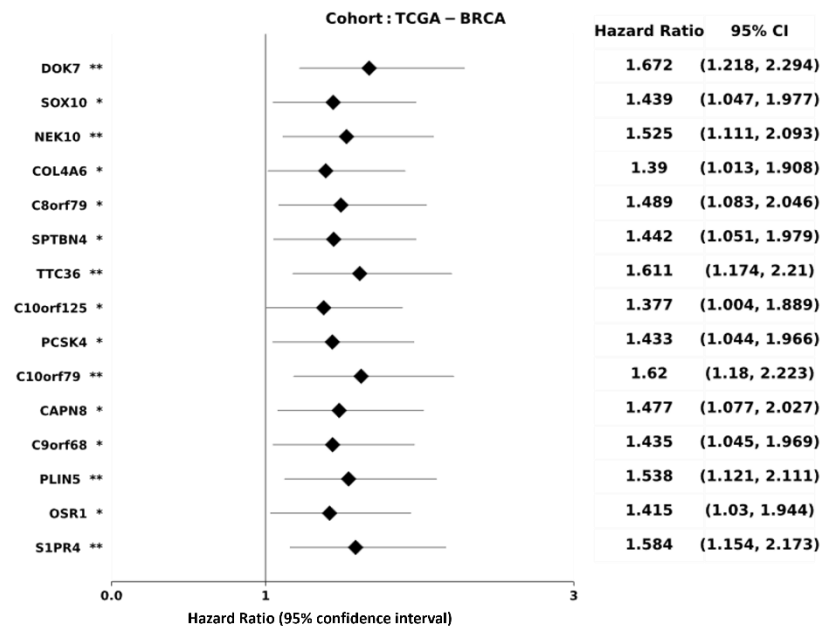
**Supplementary Fig. 5.** Prognostic analysis using DEPTH2 score. Kaplan-Meier curves for LUAD (a-b) and KIRC (c-d) are shown. Left column: KM plots using DEPTH scores derived from all genes. Right column: KM plots using DEPTH scores derived from key genes.



**Supplementary Fig. 6.** Kaplan-Meier Curve comparing the survival of TRU subtype vs PI and PP subtypes combined. It is evident from survival analysis that TRU subtype is prognostically favorable and has a higher chance of survival than PI and PP subtypes combined.

**Supplementary Table 2.** Correlation of DEPTH score distribution calculated using all genes and key genes for three molecular subtypes of LUAD and their p-values. The Pearson and Spearman correlation coefficients and the corresponding p-values are listed.

Correlation	PP (all genes) vs PP (key genes)		TRU (all genes) vs TRU (key genes)		PI (all genes) vs PI (key genes)	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Pearson	0.6031	0.0001	0.7756	3.57E-12	0.7946	7.32E-13
Spearman	0.6091	0.0001	0.7466	5.94E-11	0.8078	7.32E-13



**Supplementary Fig. 7.** Forest plot of key genes for BRCA. Summary of survival analyses of key genes that are prognostically significant. It shows the Hazard Ratio of the value (diamond shape) and its 95% confidence interval (straight line crossing the diamond). The logrank P-value is represented by the number of stars alongside the gene (\* -  $P \leq 0.05$ , \*\* -  $P \leq 0.01$ , \*\*\* -  $P \leq 0.001$ , \*\*\*\* -  $P \leq 0.0001$ ).

**Supplementary Table 3.** List of key genes selected by multi-run concrete autoencoder. The genes in bold possess prognostic capability.

Cancer Type	List of Genes
BRCA	<p><i>AQP5, AQP7, ARHGAP20, BCL2A1, C10orf125, C10orf79, C12orf60, C16orf79, C3orf54, C8orf79, C9orf68, CAPN8, CARD9, CCDC160, CD80, CDC42EP5, CDKL5, CEACAM5, CECR2, CHRM1, CHST6, CIDEA, CNIH2, COL4A6, CXorf21, CYP39A1, DEPDC7, DIO1, DOK5, DOK7, EN1, FAM54A, FAM83A, FMN1, FUT3, FXYP1, GPC2, GPR1, GPR84, GPRIN2, GRP, HK3, HOTAIR, HOXC11, HRASLS2, IGDCC3, KCNJ3, KCNK15, KLHDC1, KLRG2, KREMEN2, KRT6A, LOC145837, LPPR3, LRRC26, LRRC31, MAGEA12, MS4A14, MSII, MUC16, MYT1, NEK10, NHLRC4, NXPH4, OBP2B, OSR1, PCDHGA5, PCSK4, PIF1, PLA2G4F, PLAC9, PLIN5, PRAM1, PREX2, PRKG1, PRTG, PVALB, RBM24, ROR1, RSPO4, RUNDC2A, RUNDC3A, S100A7, S100A8, SIPR4, SOX10, SOX11, SOX8, SPTBN4, SSPO, TCAP, TLR6, TLR8, TMPRSS6, TREM1, TRH, TTC36, USHBP1, ZBP1, ZNF695</i></p>
KIRC	<p><i>ABCC6P1, ADAMTS18, ANO2, AQP10, ARHGEF4, ASPDH, ASXL3, ATRNL1, B3GALT5, BEST4, BTBD11, C10orf93, C16orf74, C1QL4, C1orf116, C3orf36, C3orf55, C5orf46, C8orf22, CCDC160, CDH3, CDKL5, CHGB, CLIC3, CPXM1, CXCL5, CYP17A1, CYP3A7, CYorf15A, DAO, DGKI, EEFA12, EPN3, EPS8L3, FABP1, FAM153B, FAM153C, FAM160A1, FAM162B, FCGR1C, FGA, FOLR3, GCSH, GK3P, GPR84, GRHL2, GRRP1, HSH2D, IGF2BP3, IGLON5, IL1RL1, IL6, IQSEC3, IRX6, ITGAD, KIF14, KISS1R, KLHL32, LAX1, LOC100128842, LOC100131434, LOC221442, LOC388387, MAGED4B, MEI1, MIAT, MPZL3, MYH7B, NAT8B, NCF1B, NHSL2, NKAIN4, NRTN, PPP1R14D, PRPH, PRR15L, PRR7, RASL11B, REL, RGS7BP, RHBDL1, SEPT3, SGK196, SIT1, SLC34A1, SLC6A20, SPEF1, SPINK13, SST, STK33, SYCE1L, TBX6, TEX15, TM6SF2, TMEM72, TTTY15, TUBA3D, VSX1, ZNF354C, ZNF366</i></p>
LUAD	<p><i>ABCA6, ABCA8, ABCC2, ABO, ADAMTS8, ANXA8L2, APCDD1L, ASCL1, BCL2L14, C17orf55, C19orf59, C1orf182, C6orf176, CARNIS1, CATSPER1, CCDC150, CD1A, CD80, CDH15, CDH2, CENPI, CHAC2, CNGA3, CSMD2, CXCL6, CYP2D7P1, DLEC1, DLX3, DPCR1, EME2, FAM54A, FBN2, FCGR1C, FGB, FGFBP1, GABRP, GDF10, GGTLC1, GJB3, GKN2, GPC2, GPR87, HAL, HNF4A, HOXB8, HOXB9, HPGDS, HS3ST3A1, INHA, IP6K3, ITGA8, KIF1A, KRT23, LOC100131434, LOC100272228, LOC284578, LOC388588, LOC440356, LOC440905, LRRC36, MAGEA3, MAGEA6, MESPI, MESP2, MSII, MT1M, MYH7B, MYO7B, NCF1C, NEURL, NPW, NUP210L, NWD1, OIP5, ORM2, PITX2, PLA2G10, PPAPDC1A, PPAL4C, PRAME, PYHINI, S100A3, SCN7A, SHISA3, SLAMF1, SMC1B, SNTN, SPC25, SPIB, SSPO, STAC, TCN1, TDRD10, THEMIS, TNFSF15, TUBB2B, VILI, WBSR26, WIF1, ZYG11A</i></p>