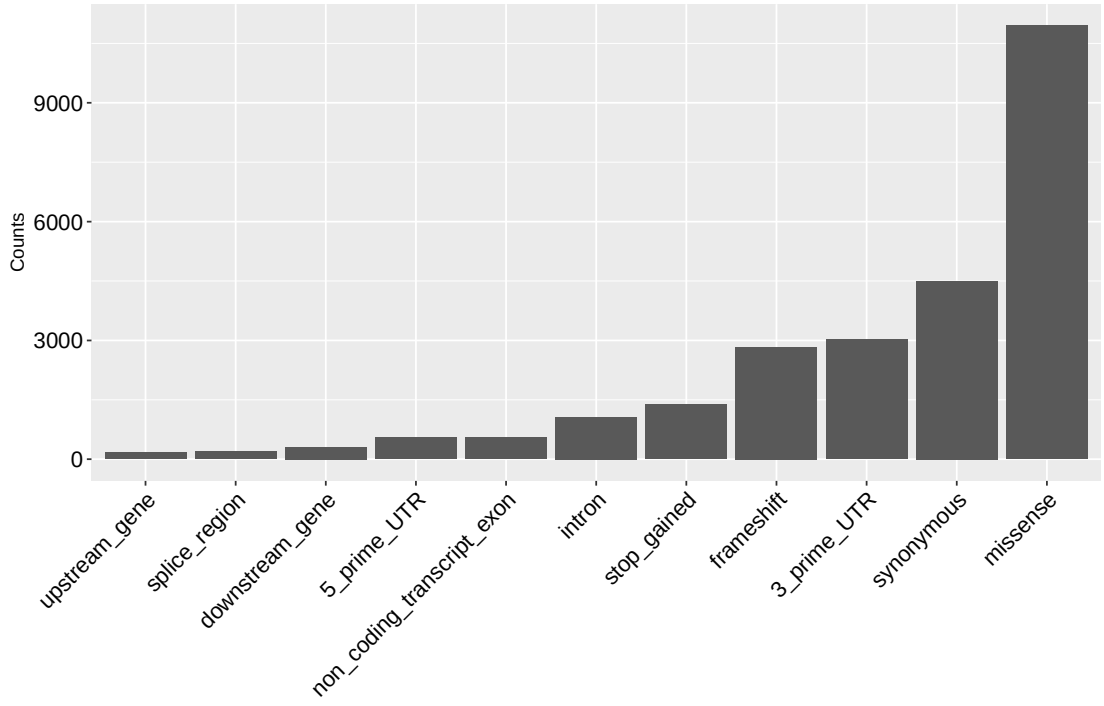
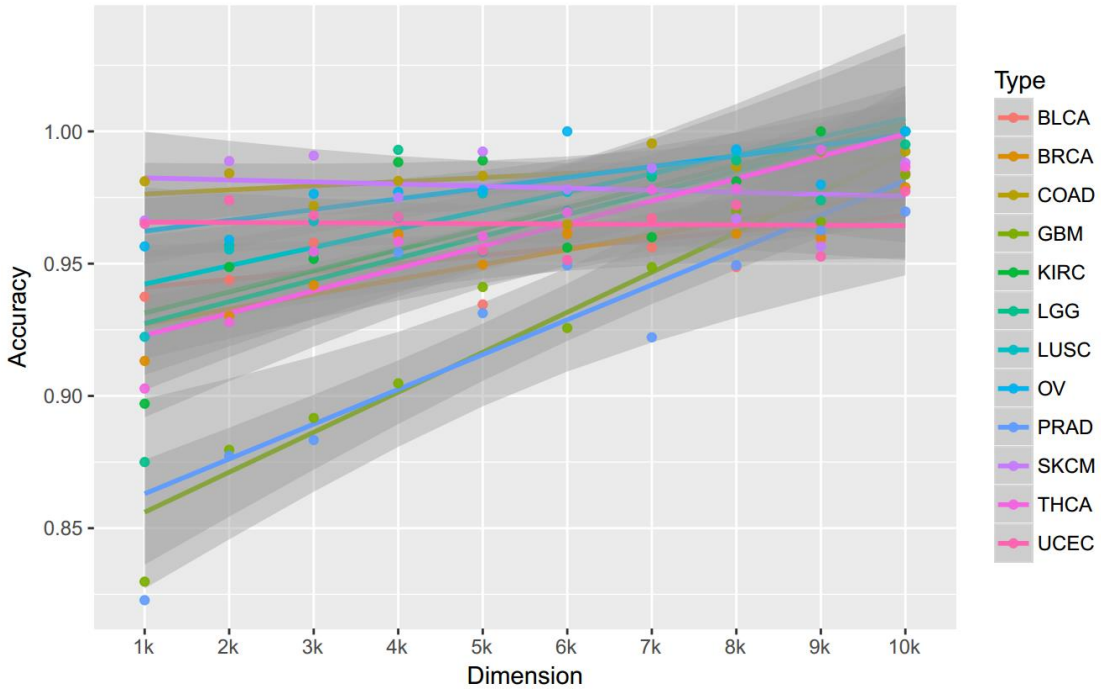
Identification of 12 cancer types through genome deep learning

Yingshuai Sun[1,*], Sitao Zhu[1,*], Kailong Ma[2], Weiqing Liu[1], Yao Yue[1], Gang Hu[1], Huifang Lu[2], Wenbin Chen[2,#]

Supplementary Information



Supplementary Figure 1 | Counts of point mutation in gene component including up stream, down stream, splice, UTR, exon and intron regions. The missense mutation are twice of synonymous mutation in counts.

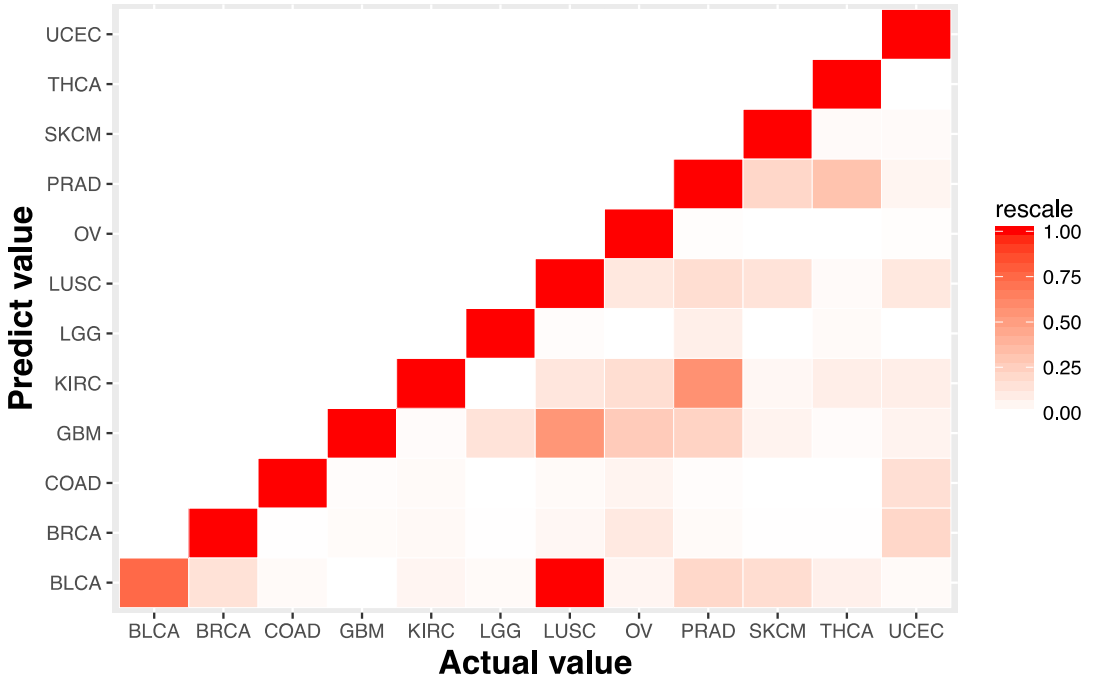


Supplementary Figure 2 | Modeling accuracy of 12 kinds of cancer in different dimensions.

The trends in modeling accuracy of 12 cancers in different dimensions. The accuracy of the model increases as the dimension increases. Due to the influence of training data selection and modeling parameters, the accuracy rate changes will produce some fluctuations, but from the whole trend, the accuracy rate increases with the increase of the dimension.

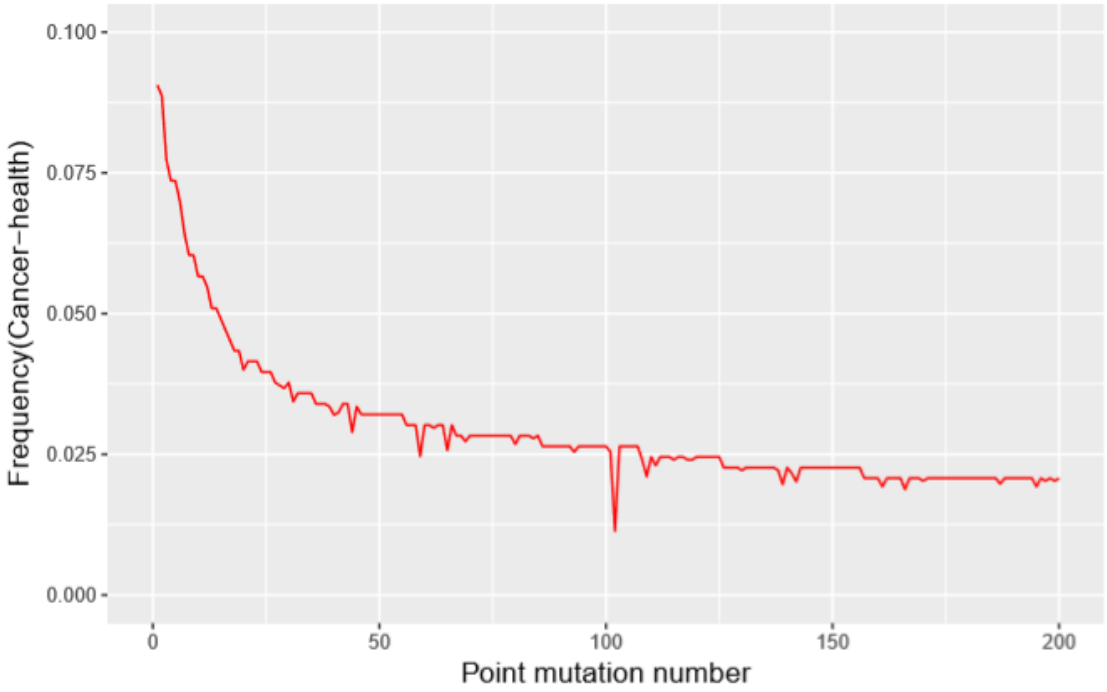
Supplementary Table 1 | The performance of cancer stage dataset for GDL model evaluation.

Cancer	Stage	Cancer	Health	Accuracy(%)	Sensitivity(%)	Specificity(%)
BLCA	I	2	51	93.23	50.00	98.04
	II	18	51	98.55	100.00	98.04
	III	26	51	98.70	100.00	98.04
	IV	30	51	98.43	98.68	98.04
BRCA	I	26	115	97.87	92.31	98.26
	II	135	115	98.40	98.52	98.26
	III	49	115	98.17	97.96	98.26
	IV	2	115	98.29	100.00	98.26
COAD	I	18	167	98.92	100.00	99.40
	II	42	167	99.52	100.00	99.40
	III	25	167	99.48	100.00	99.40
	IV	10	167	98.87	100.00	99.40
KIRC	I	53	28	100.00	100.00	100.00
	II	6	28	100.00	100.00	100.00
	III	18	28	100.00	100.00	100.00
	IV	6	28	100.00	100.00	100.00
LUSC	I	54	33	100.00	100.00	100.00
	II	39	33	100.00	100.00	100.00
	III	15	33	100.00	100.00	100.00
	IV	0	33	100.00	--	100.00
SKCM	I	10	93	99.03	100.00	98.92
	II	19	93	99.11	100.00	98.92
	III	32	93	98.40	96.88	98.92
	IV	3	93	98.96	100.00	98.92
THCA	I	71	51	97.54	97.18	98.04
	II	5	51	98.21	100.00	98.04
	III	22	51	98.63	100.00	98.04
	IV	0	51	98.04	--	98.04

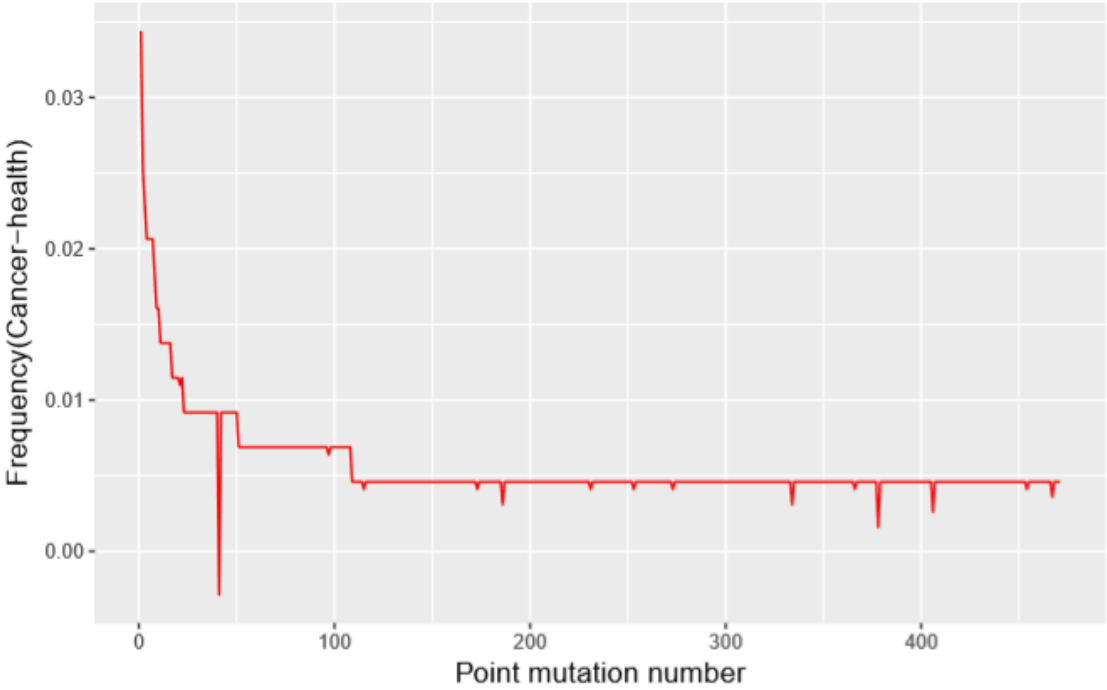


Supplementary Figure 3 | Confusion matrix of mixture model two-way judgment.

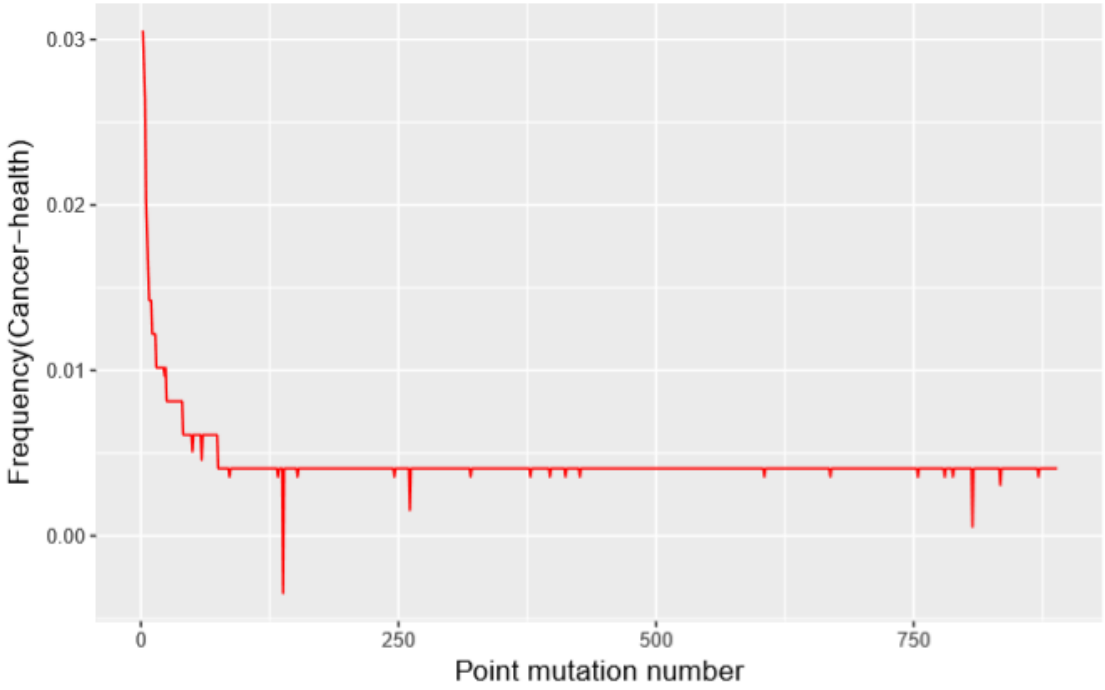
The misjudgment between LUSC and BLCA is most obvious, followed by LUSC and GBM, PRAD and KIRC.



Supplementary Figure 4 | The point mutation of uterine corpus endometrial carcinoma(UCEC). Horizontal coordinate is the specific point mutation and Vertical coordinate standards for the difference between cancer and normal.



Supplementary Figure 5 | The point mutation of ovarian carcinoma (OV). Horizontal coordinate is the specific point mutation and Vertical coordinate standards for the difference between cancer and normal.



Supplementary Figure 6 | The point mutation of lung squamous cell carcinoma (LUSC). Horizontal coordinate is the specific point mutation and Vertical coordinate standards for the difference between cancer and normal.