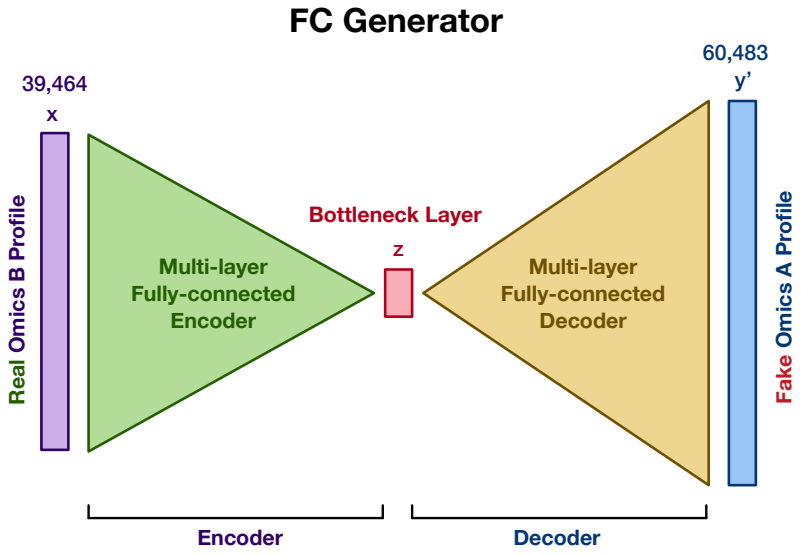
OmiTrans: generative adversarial networks based omics-to-omics translation framework

Xiaoyu Zhang, Yike Guo

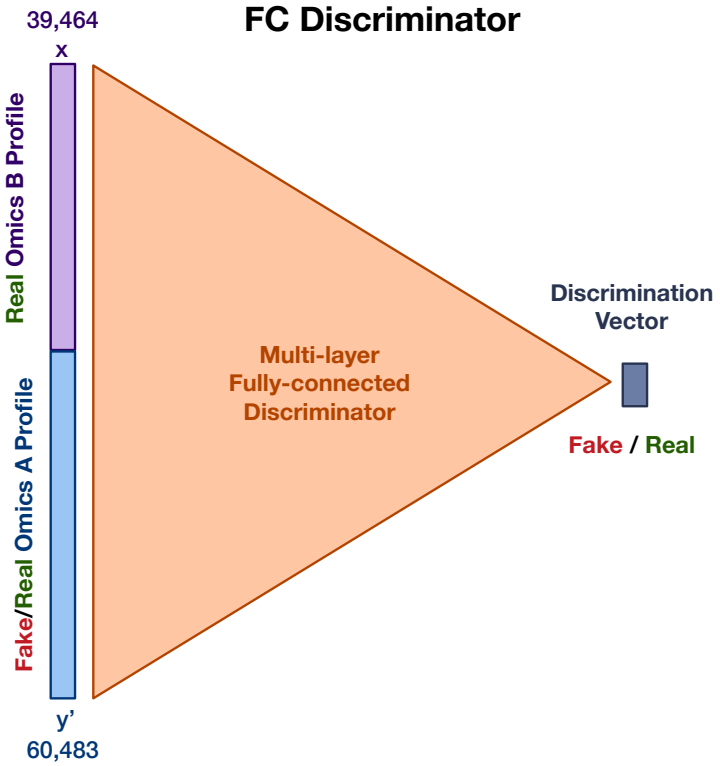
This document provides the Supplementary Figures and Tables mentioned in the manuscript.



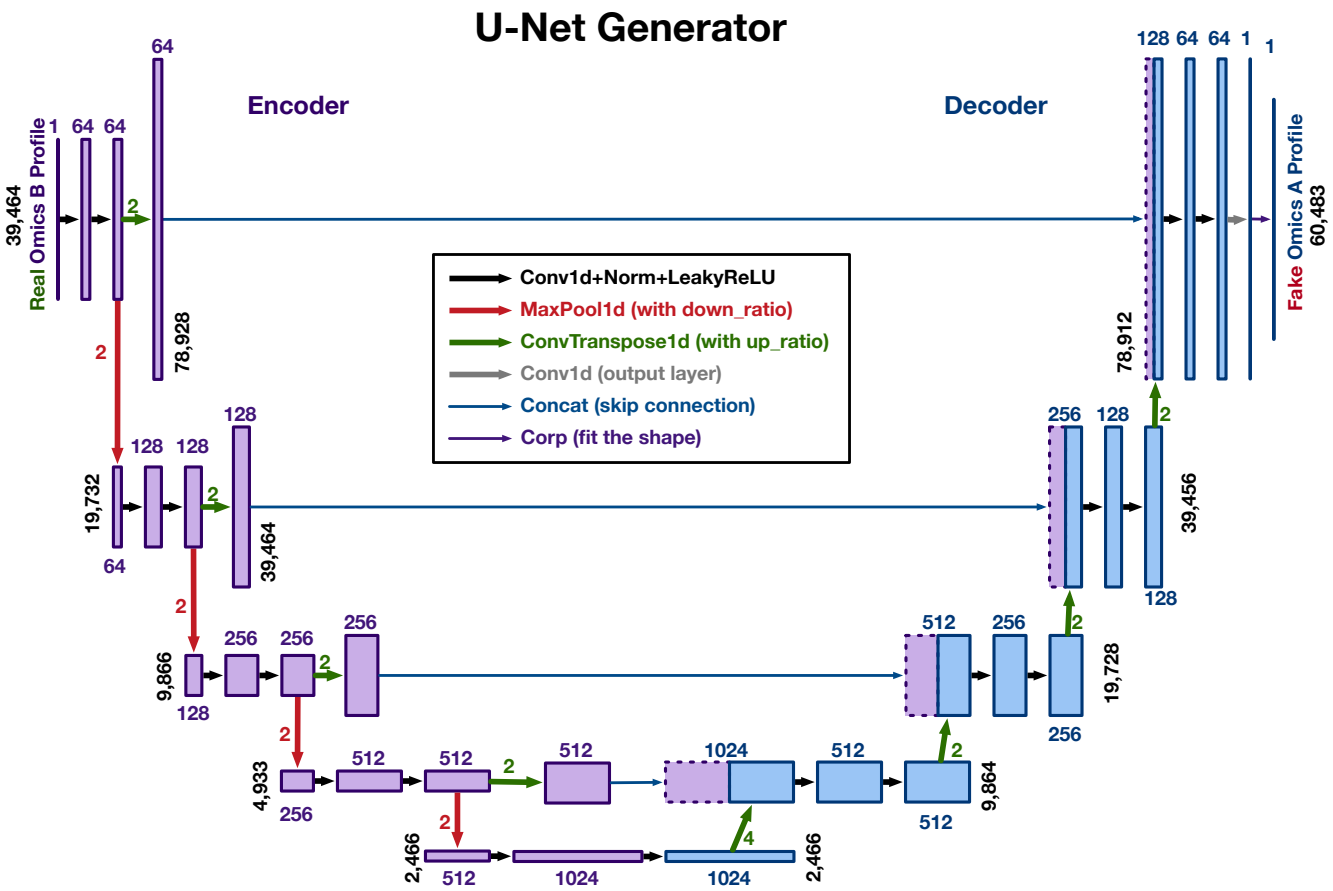
Supplementary Figure 1: The implementation of the OmiTrans generator G using a deep fully-connected (FC) encoder-decoder network. The encoder and decoder of G consist of multiple fully-connected blocks (FC + norm + dropout + activation). The dimensionality of the bottleneck layer is a hyper-parameter of the network normally set to 256. Here we use the omics translation from DNA methylation (omics B) to gene expression (omics A) as an example. The dimensionalities of the two omics types are marked in the diagram.

Supplementary Table 1: Hyper-parameters used for OmiTrans with the FC-based architecture.

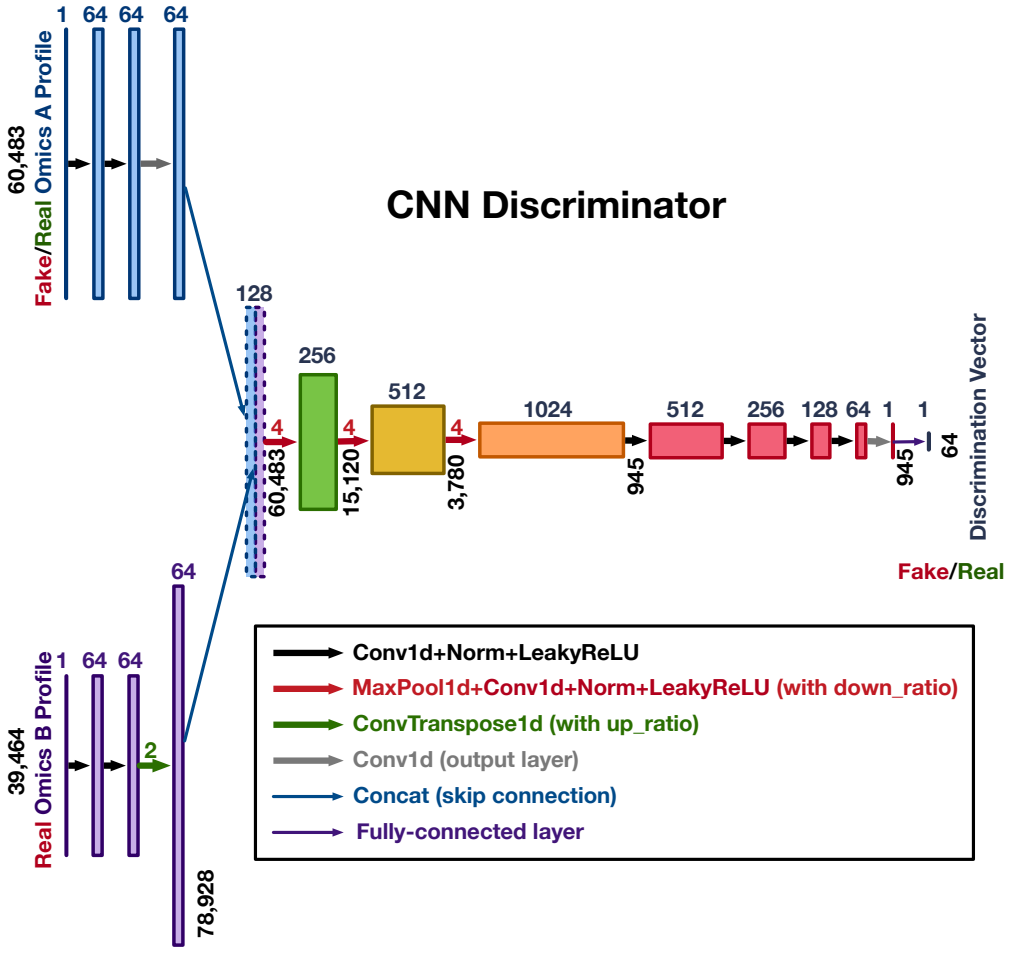
Hyper-parameter	Value
Latent dimension	256
Learning rate of the generator G	1×10^{-4}
Learning rate of the discriminator D	1×10^{-4}
Batch size	128
Epoch number - total	800
Epoch number - decay	100
Dropout rate	0.2



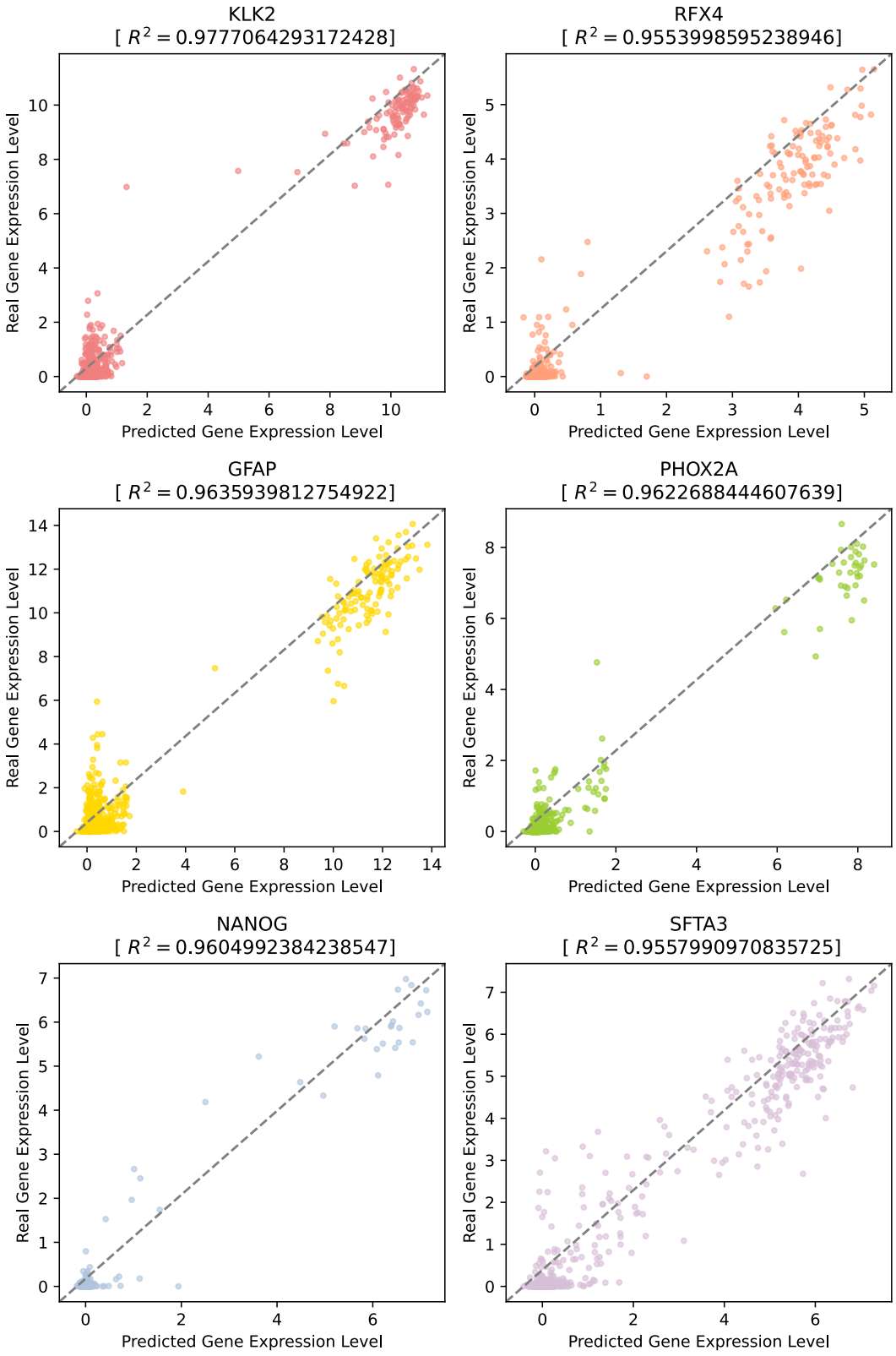
Supplementary Figure 2: The implementation of the OmiTrans discriminator D using a deep fully-connected (FC) discriminator network, which consists of multiple fully-connected blocks (FC + norm + dropout + activation). The output of the discriminator network is a discrimination vector determining whether the input omics A is fake or real. Here we use the omics translation from DNA methylation (omics B) to gene expression (omics A) as an example. The dimensionalities of the two omics types are marked in the diagram.



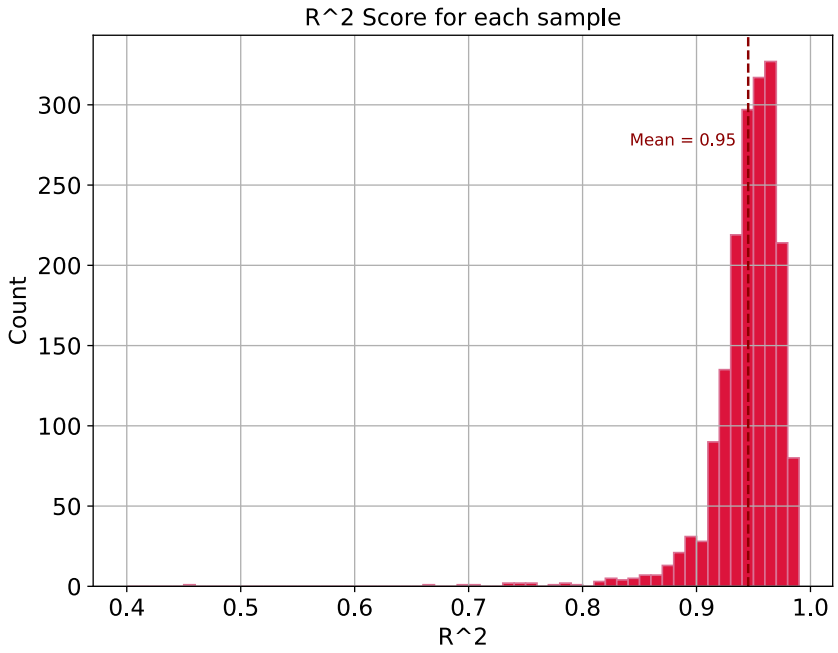
Supplementary Figure 3: The implementation of the OmiTrans generator G using a U-Net network. The channel number and feature number of each tensor in the network were marked in the diagram. Each type of operation group (e.g., conv1D + norm + LeakyReLU) was illustrated using arrows with different colours and thicknesses shown in the legend. Skip connections between each layer i and layer $n - i$ (n stands for the total number of layers in the U-Net) can also be seen in the diagram.



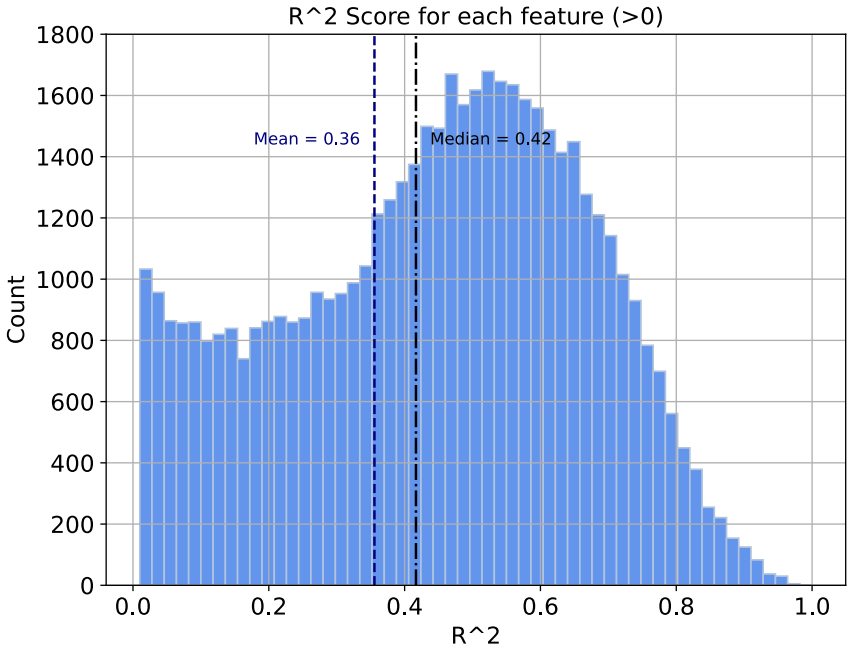
Supplementary Figure 4: The implementation of the OmiTrans discriminator D using a CNN. The channel number and feature number of each tensor in the network were marked in the diagram. Each type of operation group (e.g., conv1D + norm + LeakyReLU) is illustrated using arrows with different colours and thicknesses shown in the legend. The output of the discriminator network is a discrimination vector determining whether the input omics A is fake or real.



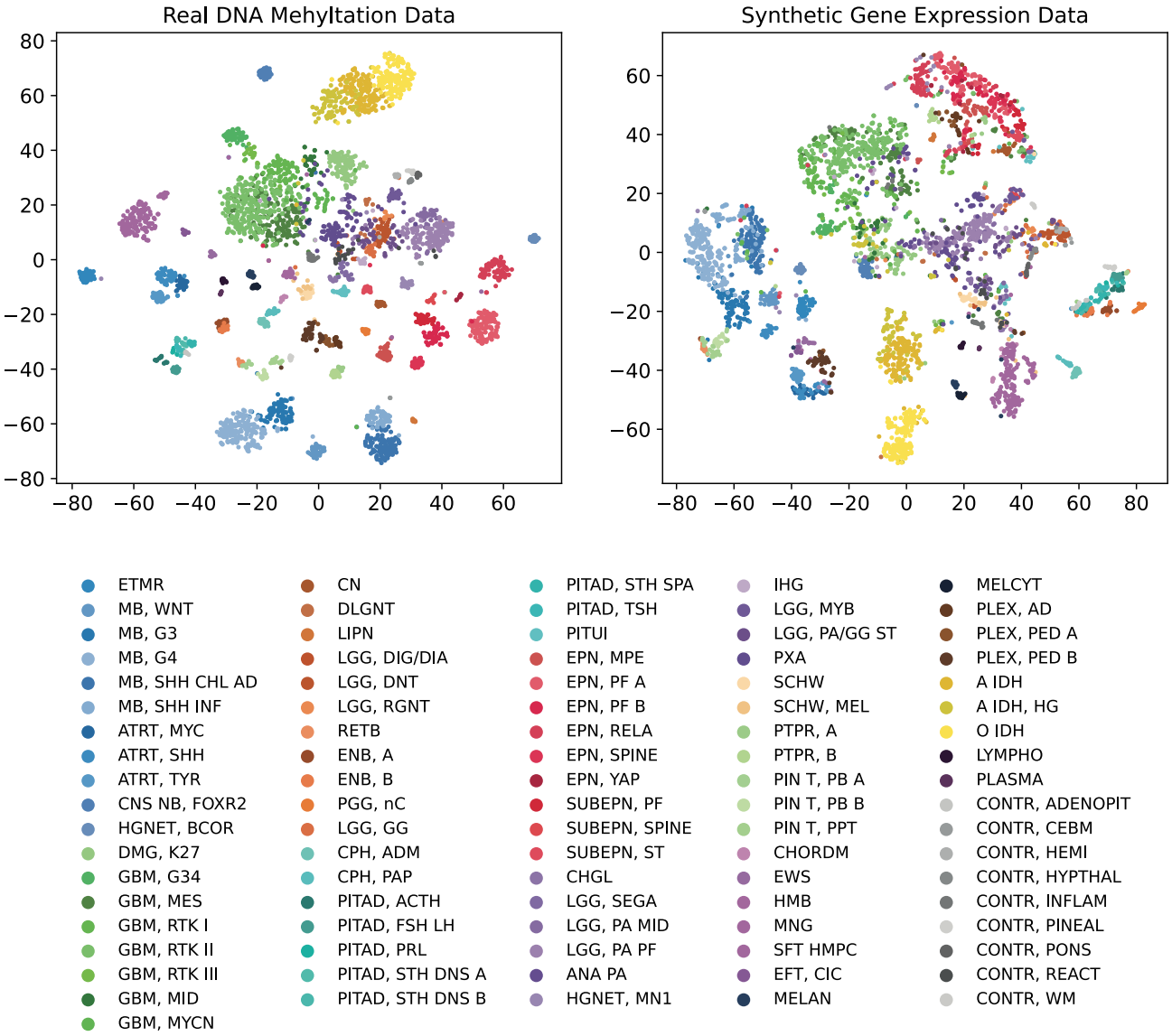
Supplementary Figure 5: Example genes with high reconstruction performance using FC-based OmiTrans. The y-axis represents the real gene expression levels, and the x-axis indicates the synthetic gene expression levels. The gene name and the corresponding R^2 value can be seen in the title of each subgraph.



Supplementary Figure 6: Histogram of the samplewise R^2 scores (R^2_s).



Supplementary Figure 7: Histogram of the featurewise R^2 scores (R^2_f) that larger than zero.



Supplementary Figure 8: The scatter graphs of the original DNA methylation data of the BTM dataset, the fake gene expression data of BTM synthesised by a OmiTrans generator trained on the GDC training set. Samples with different brain tumour types or normal control regions were marked with different colours, brain tumour types and normal control regions belonging to the same upper-level class were marked in similar colours, as shown in the legend. Full name of each brain tumour type and normal control region is listed in Supplementary Table 3.

Supplementary Table 2: Tumour type information of the multi-omics datasets from Genomic Data Commons (GDC). Sample numbers of each tumour type and the normal control for RNA-Seq gene expression profiling (A) and DNA methylation profiling (B) are shown in corresponding columns.

Tumour Type	Abbreviation	Count A	Count B
Breast invasive carcinoma	BRCA	1,104	794
Brain lower grade glioma	LGG	529	534
Thyroid carcinoma	THCA	510	515
Head and neck squamous cell carcinoma	HNSC	502	530
Prostate adenocarcinoma	PRAD	499	503
Lung adenocarcinoma	LUAD	526	471
Skin cutaneous melanoma	SKCM	471	473
Uterine corpus endometrial carcinoma	UCEC	548	436
Bladder urothelial carcinoma	BLCA	411	416
Liver hepatocellular carcinoma	LIHC	374	380
Lung squamous cell carcinoma	LUSC	501	370
Stomach adenocarcinoma	STAD	375	395
Kidney renal clear cell carcinoma	KIRC	535	323
Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC	306	309
Colon adenocarcinoma	COAD	471	309
Kidney renal papillary cell carcinoma	KIRP	289	276
Sarcoma	SARC	263	265
Pheochromocytoma and paraganglioma	PCPG	183	184
Pancreatic adenocarcinoma	PAAD	178	185
Esophageal carcinoma	ESCA	162	186
Testicular germ cell tumours	TGCT	156	156
Thymoma	THYM	119	124
Acute myeloid leukemia	LAML	151	140
Rectum adenocarcinoma	READ	167	99
Mesothelioma	MESO	86	87
Uveal melanoma	UVM	80	80
Adrenocortical carcinoma	ACC	79	80
Kidney chromophobe	KICH	65	66
Uterine carcinosarcoma	UCS	56	57
Lymphoid neoplasm diffuse large B-cell lymphoma	DLBC	48	48
Cholangiocarcinoma	CHOL	36	36
Ovarian serous cystadenocarcinoma	OV	379	10
Glioblastoma multiforme	GBM	168	153
Normal control		741	746
Total		11,068	9,736

Supplementary Table 3: Detailed tumour type information of the GSE109381 Brain Tumour Methylation (BTM) dataset.

Methylation Class	Abbreviation	Count
Embryonal tumor with multilayered rosettes	ETMR	49
Medulloblastoma, WNT	MB, WNT	48
Medulloblastoma, subclass group 3	MB, G3	99
Medulloblastoma, subclass group 4	MB, G4	181
Medulloblastoma, subclass SHH A (children and adult)	MB, SHH CHL AD	126
Medulloblastoma, subclass SHH B (infant)	MB, SHH INF	65
Atypical teratoid/rhabdoid tumor, subclass MYC	ATRT, MYC	31
Atypical teratoid/rhabdoid tumor, subclass SHH	ATRT, SHH	51
Atypical teratoid/rhabdoid tumor, subclass TYR	ATRT, TYR	39
CNS neuroblastoma with FOXR2 activation	CNS NB, FOXR2	43
CNS high grade neuroepithelial tumor with BCOR alteration	HGNET, BCOR	26
Diffuse midline glioma H3 K27M mutant	DMG, K27	117
Glioblastoma, IDH wildtype, H3.3 G34 mutant	GBM, G34	54
Glioblastoma, IDH wildtype, subclass mesenchymal	GBM, MES	160
Glioblastoma, IDH wildtype, subclass RTK I	GBM, RTK I	108
Glioblastoma, IDH wildtype, subclass RTK II	GBM, RTK II	261
Glioblastoma, IDH wildtype, subclass RTK III	GBM, RTK III	22
Glioblastoma, IDH wildtype, subclass midline	GBM, MID	33
Glioblastoma, IDH wildtype, subclass MYCN	GBM, MYCN	33
Central neurocytoma	CN	23
Diffuse leptomeningeal glioneuronal tumor	DLGNT	12
Cerebellar liponeurocytoma	LIPN	11
Low grade glioma, desmoplastic infantile astrocytoma / ganglioglioma	LGG, DIG/DIA	8
Low grade glioma, dysembryoplastic neuroepithelial tumor	LGG, DNT	56
Low grade glioma, rosette forming glioneuronal tumor	LGG, RGNT	12
Retinoblastoma	RETB	19
Esthesioneuroblastoma, subclass A	ENB, A	24
Esthesioneuroblastoma, subclass B	ENB, B	16
Paraganglioma, spinal non-CIMP	PGG, nC	20
Low grade glioma, ganglioglioma	LGG, GG	26
Craniopharyngioma, adamantinomatous	CPH, ADM	25
Craniopharyngioma, papillary	CPH, PAP	20
Pituitary adenoma, ACTH	PITAD, ACTH	19
Pituitary adenoma, FSH/LH	PITAD, FSH LH	23
Pituitary adenoma, prolactin	PITAD, PRL	8
Pituitary adenoma, STH densely granulated, group A	PITAD, STH DNS A	9
Pituitary adenoma, STH densely granulated, group B	PITAD, STH DNS B	13
Pituitary adenoma, STH sparsely granulated	PITAD, STH SPA	17
Pituitary adenoma, TSH	PITAD, TSH	11
Pituicytoma / granular cell tumor / spindle cell oncocytoma	PITUI	30
Ependymoma, myxopapillary	EPN, MPE	45
Ependymoma, posterior fossa group A	EPN, PF A	127
Ependymoma, posterior fossa group B	EPN, PF B	59
Ependymoma, RELA fusion	EPN, RELA	90
Ependymoma, spinal	EPN, SPINE	34
Ependymoma, YAP fusion	EPN, YAP	11
Subependymoma, posterior fossa	SUBEPN, PF	40
Subependymoma, spinal	SUBEPN, SPINE	12
Subependymoma, supratentorial	SUBEPN, ST	20
Chordoid glioma of the third ventricle	CHGL	12
Low grade glioma, subependymal giant cell astrocytoma	LGG, SEGA	22

Low grade glioma, subclass hemispheric pilocytic astrocytoma and ganglioglioma	LGG, PA MID	55
Low grade glioma, subclass midline pilocytic astrocytoma	LGG, PA PF	154
Anaplastic pilocytic astrocytoma	ANA PA	46
CNS high grade neuroepithelial tumor with MN1 alteration	HGNET, MN1	26
Infantile hemispheric glioma	IHG	15
Low grade glioma, MYB/MYBL1	LGG, MYB	27
Low grade glioma, subclass posterior fossa pilocytic astrocytoma	LGG, PA/GG ST	45
(Anaplastic) pleomorphic xanthoastrocytoma	PXA	67
Schwannoma	SCHW	31
Melanotic schwannoma	SCHW, MEL	12
Papillary tumor of the pineal region group A	PTPR, A	9
Papillary tumor of the pineal region group B	PTPR, B	23
Pineoblastoma group A / intracranial retinoblastoma	PIN T, PB A	9
Pineoblastoma group B	PIN T, PB B	23
Pineal parenchymal tumor	PIN T, PPT	20
Chordoma	CHORDM	11
Ewing sarcoma	EWS	17
Hemangioblastoma	HMB	27
Meningioma	MNG	149
Solitary fibrous tumor / hemangiopericytoma	SFT HMPC	18
CNS Ewing sarcoma family tumor with CIC alteration	EFT, CIC	13
Melanoma	MELAN	18
Melanocytoma	MELCYT	19
Plexus tumor, subclass adult	PLEX, AD	23
Plexus tumor, subclass paediatric A	PLEX, PED A	16
Plexus tumor, subclass paediatric B	PLEX, PED B	49
IDH glioma, subclass astrocytoma	A IDH	172
IDH glioma, subclass high grade astrocytoma	A IDH, HG	87
IDH glioma, subclass 1p/19q codeleted oligodendroglioma	O IDH	163
Lymphoma	LYMPHO	14
Plasmacytoma	PLASMA	8
Control tissue, pituitary gland anterior lobe	CONTR, ADENOPIT	9
Control tissue, cerebellar hemisphere	CONTR, CEBM	8
Control tissue, hemispheric cortex	CONTR, HEMI	13
Control tissue, hypothalamus	CONTR, HYPTHAL	9
Control tissue, inflammatory tumor microenvironment	CONTR, INFLAM	24
Control tissue, pineal gland	CONTR, PINEAL	12
Control tissue, pons	CONTR, PONS	12
Control tissue, reactive tumor microenvironment	CONTR, REACT	23
Control tissue, white matter	CONTR, WM	9
Total		**3,905**