

Gene-specific correlation of RNA and protein levels in human cells and tissues

王凯
信息学院

MSB
2016.8.20

1 介绍

基本生物过程：基因组，基因产物，细胞表型

问题：mRNA水平和相应蛋白质水平之间的相关性

Nagaraj 2011,Vogel & Marcotte 2012,Payne 2015:
蛋白质组与转录组没有充分的相关性

Lundberg 2010,Wilhelm 2014:
人类细胞系和组织中有恒定的蛋白质-RNA比率

靶蛋白质组学方法：

已知量的稳定同位素标记的蛋白质片段的渗入；

胰蛋白酶消化和平行反应检测（PRM）分析，确定相对量肽；

从样品和标准肽之间测量的比率计算样品中的绝对蛋白质拷贝数

引入独立于组织类型的基因特异性RNA-蛋白（RTP）转化因子，可以显著增强来自RNA水平的蛋白质拷贝数的可预测性

2 结果

2.1 基因选择

2.2 组织样品使用基于组蛋白定量分析的PRM的标准化

2.3 组织和细胞系中靶基因的蛋白质拷贝数

2.4 组织和细胞系中RNA和蛋白质水平之间的直接相关性

2.5 所选基因的RNA和蛋白质水平的基因特异性相关性

2.6 基于RTP比率的蛋白质拷贝数的预测

2.1 基因的选择

表1 9个人细胞系

Cell-line	Origin
A-431	Skin
A549	Lung
HEK 293	Embryonal kidney
Hep G2	Liver
U-2 OS	Bone
HeLa	Cervix
RT4	Urinary bladder
MCF7	Pleural effusion
SH-SY5Y	Bone marrow

表2 11种人类组织样品

Tissue	ID	Gender	Age	Sections
Breast	V406	F	80	1
Colon	V420	M	84	1
Colon	V422	F	78	1
Colon	V424	F	72	1
Kidney	W801	NA	NA	3
Duodenum	V801	F	66	1
Gallbladder	V803	M	54	3
Heart	W800	NA	NA	3
Liver	V802	M	54	1
Liver	V804	F	60-65	1
Lung	V402	M	63	1
Lung	V404	F	75	1
Lung	V400	M	74	1
Prostate	V327	M	63	1
Spleen	V082	F	60	3
Tonsil	V287	M	23	1
Tonsil	V294	F	20	1
Tonsil	V301	M	20	1

靶蛋白组学分析的基因选择标准:

- (1) 细胞内或膜结合的蛋白质产物 (即非分泌蛋白)
- (2) 存在与大部分分析的组织 and 细胞中
- (3) 所分析的组织 and 细胞中具有相对高的可变性

—————> 55个基因

表3 部分基因的细胞系和组织的转录组数据 (TPM)

Gene	A431	A549	HEK	U2OS	HepG2	HeLa	RT4
AGPAT3	85.8	45.6	55.0	0.7	46.9	24.9	72.2
ALDH1A2	2.0	0.0	206.4	1.6	2.0	2.7	34.7
ANXA1	792.7	1290.0	8.8	477.3	0.9	290.7	151.5
ANXA3	337.9	68.2	3.4	52.5	16.8	110.2	34.8
ASS1	109.6	11.7	24.3	76.1	139.6	18.3	69.6
BPGM	12.4	15.5	18.5	6.3	5.3	10.3	145.5
BRD7	76.0	67.0	99.0	88.1	110.9	39.8	51.7
CANT1	103.2	128.7	54.6	109.4	36.9	44.0	70.6
CAP2	14.7	26.4	7.7	37.0	18.9	10.8	5.6
CAPG	169.3	82.3	5.3	35.8	27.5	117.0	80.6
CD81	294.0	320.6	483.5	194.2	0.5	90.2	187.0
CLPP	110.8	71.1	170.1	64.7	60.7	47.1	53.1
CRKL	39.8	25.9	60.4	83.4	32.3	18.9	30.4
DECR1	67.7	37.1	6.3	37.5	62.2	32.1	150.8
EPS8	63.5	42.0	13.4	17.9	27.1	9.6	85.2

2.2 组织样品使用基于组蛋白定量分析的PRM的标准化

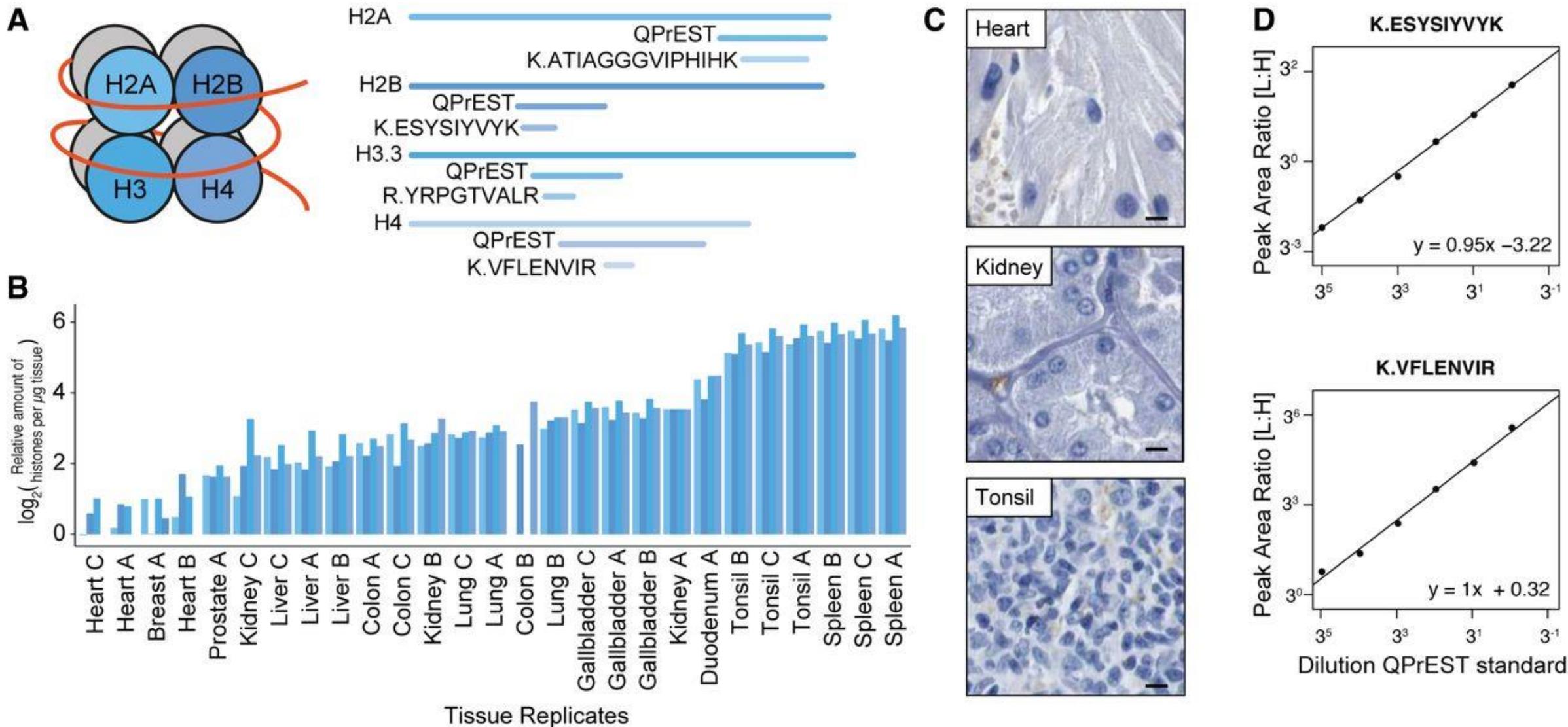


图1 使用组蛋白丰度进行标准化测定细胞计数

2.3 组织和细胞系中靶基因的蛋白质拷贝数

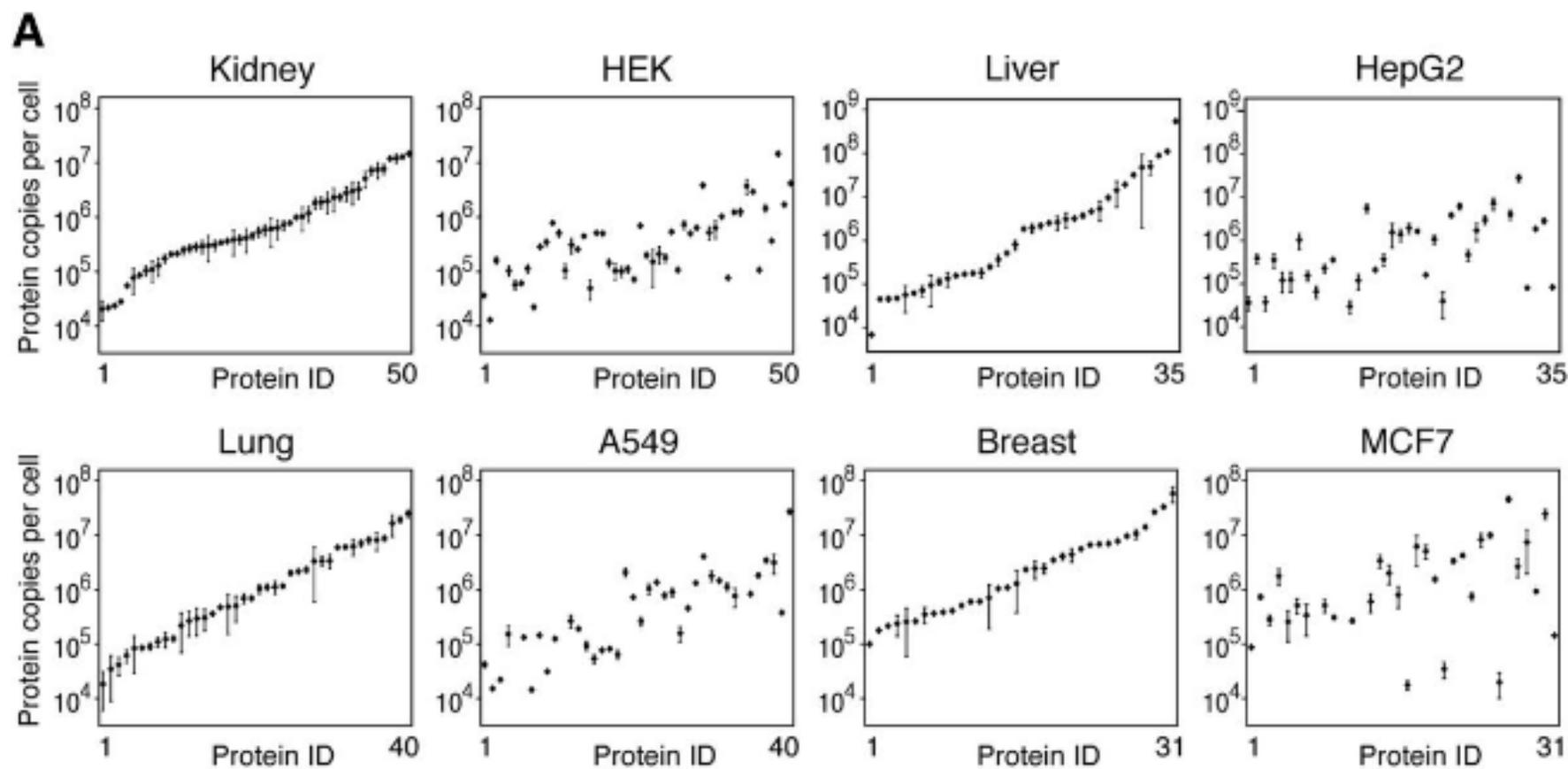


图2 4种组织和相应细胞系中蛋白质的绝对拷贝数

2.4 组织和细胞系中RNA和蛋白质水平之间的直接相关性

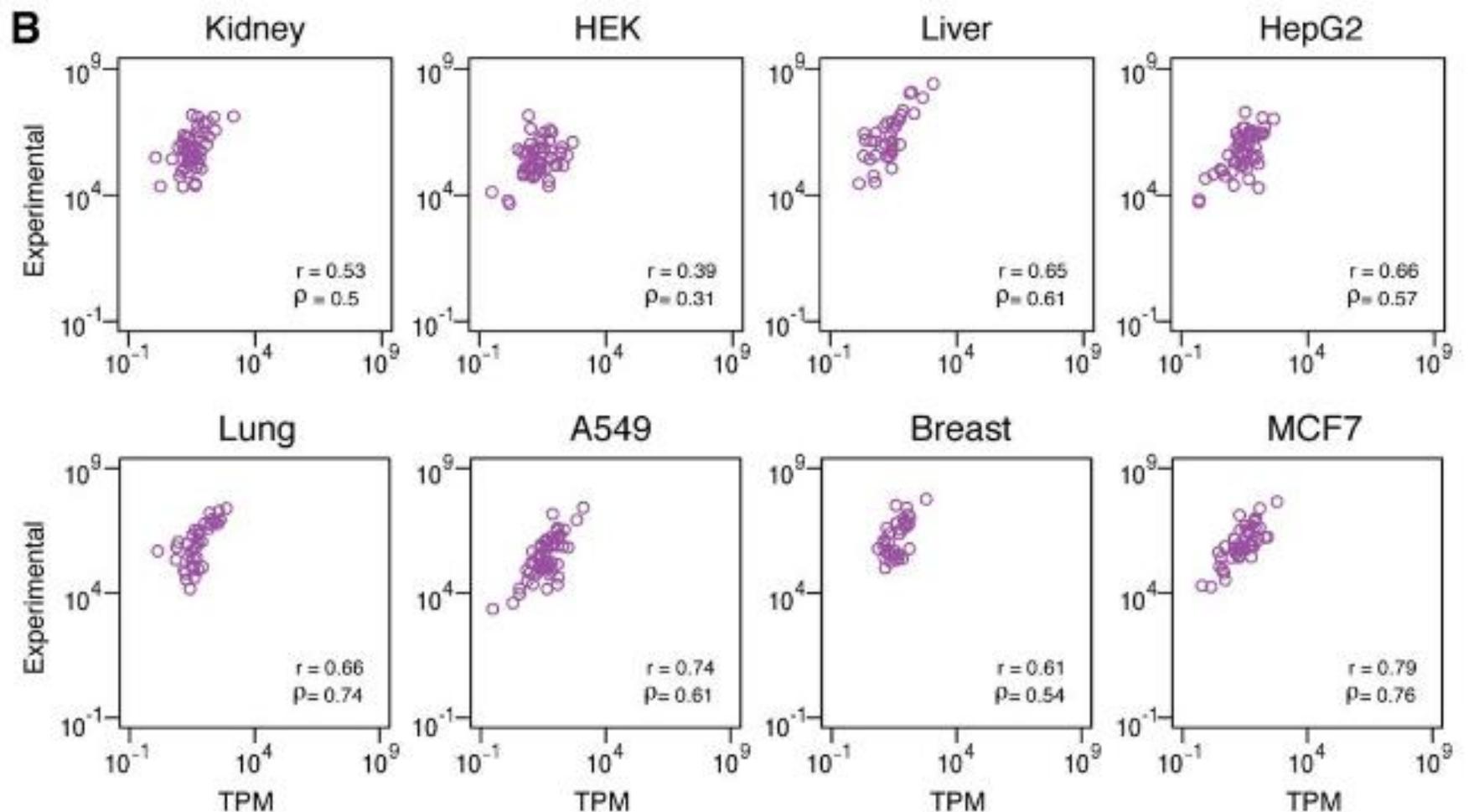


图3 相同组织和细胞系中所有定量基因的RNA (TPM) 和蛋白质拷贝数之间的直接相关性

2.5 所选基因的RNA和蛋白质水平的基因特异性相关性

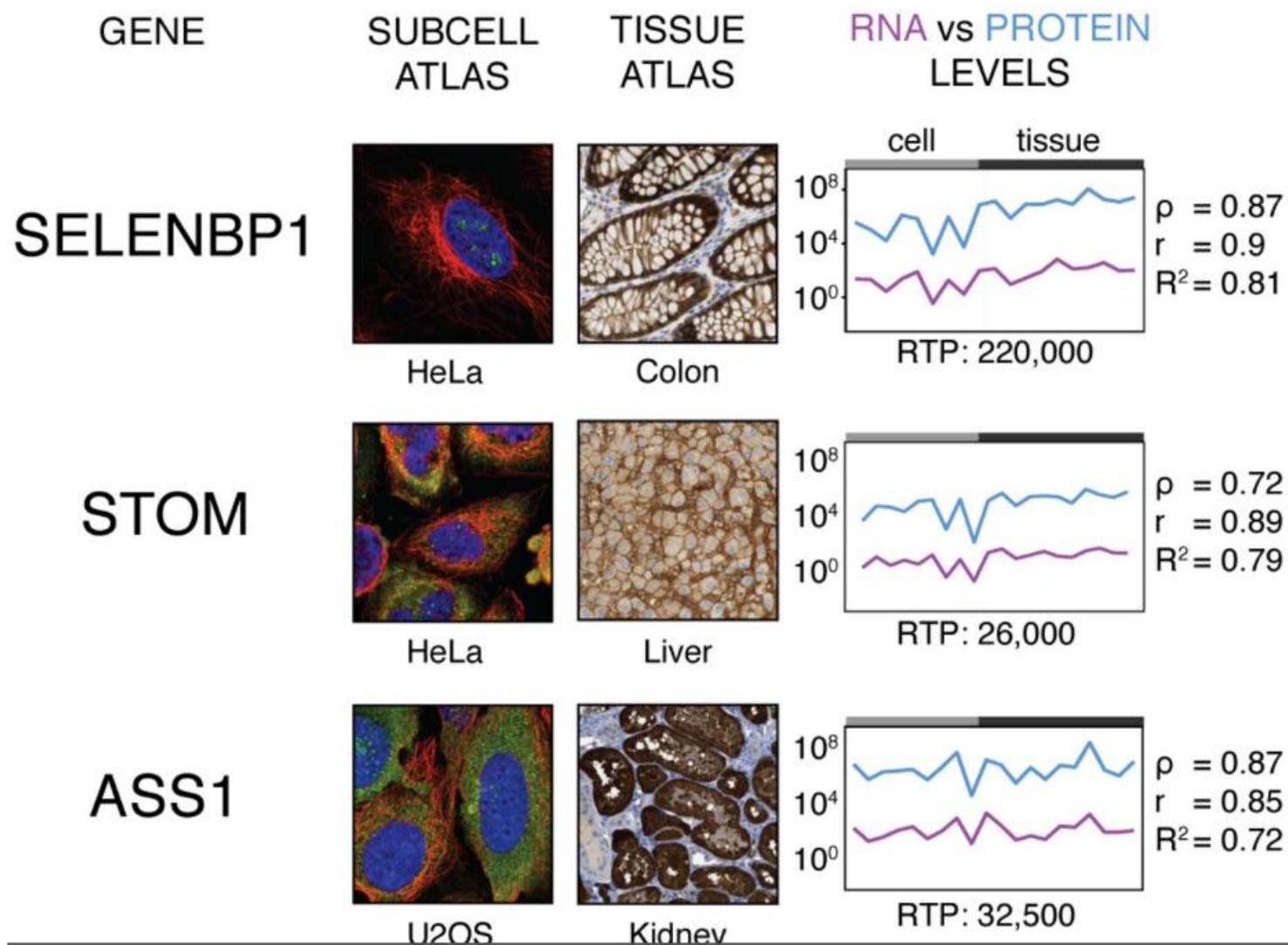


图4 三个基因的蛋白质和RNA水平

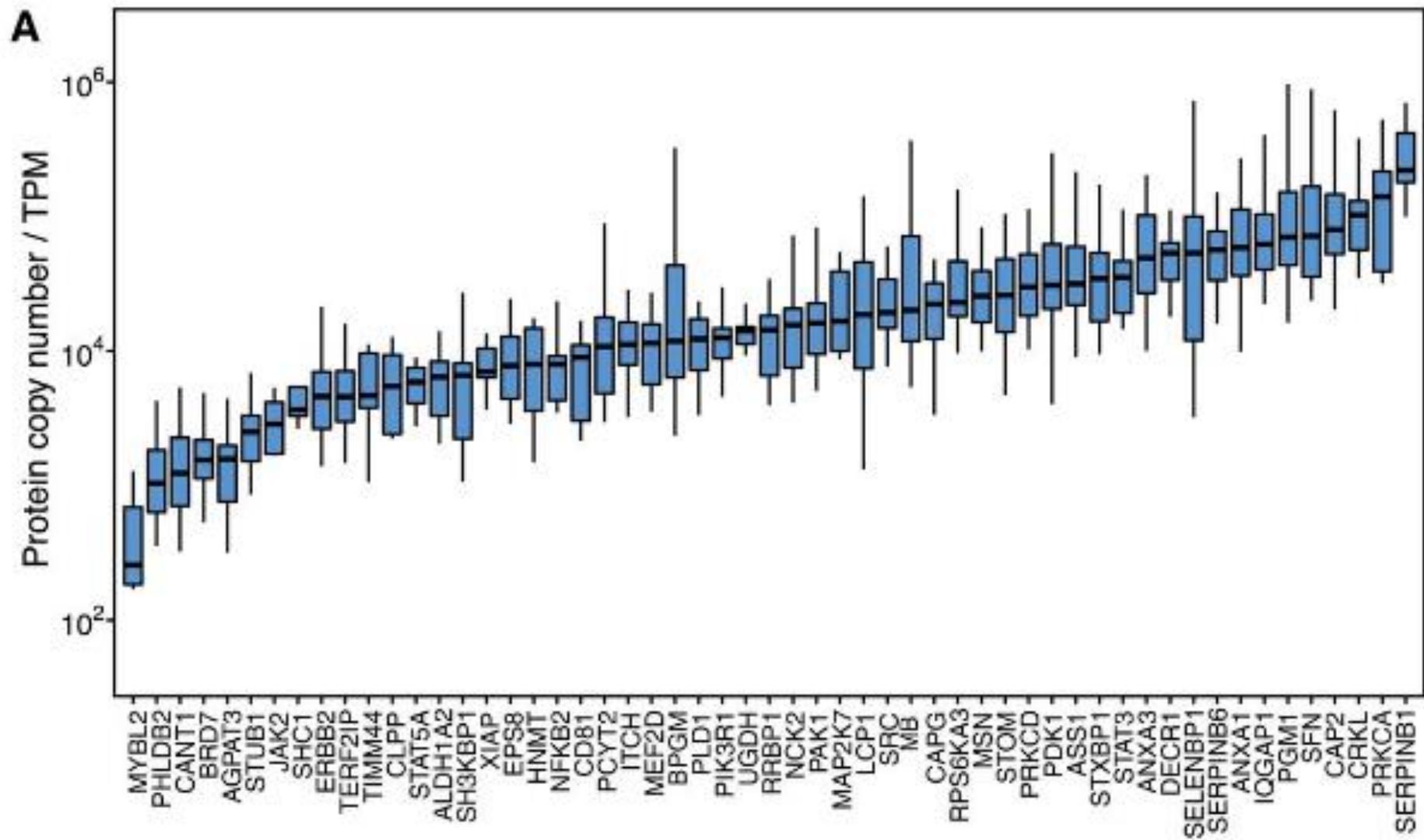


图5 55个基因的RNA和蛋白质水平的相关因子RTP

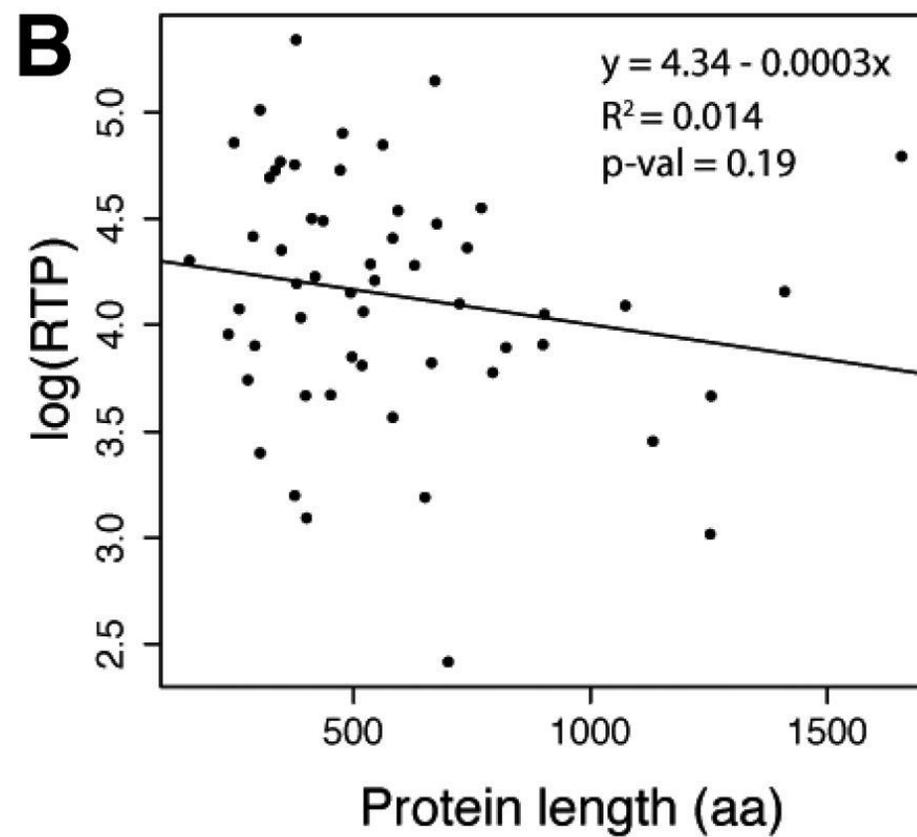
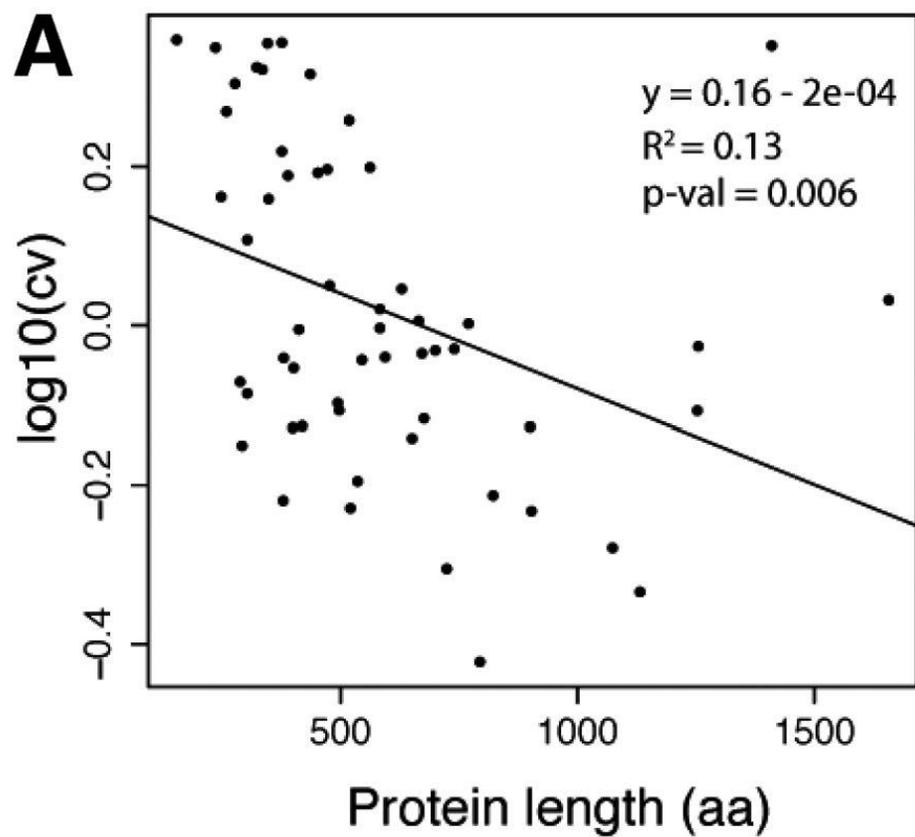


图6 RTP变异系数与RTP比率与蛋白质长度关系图

2.6 基于RTP比率的蛋白质拷贝数的预测

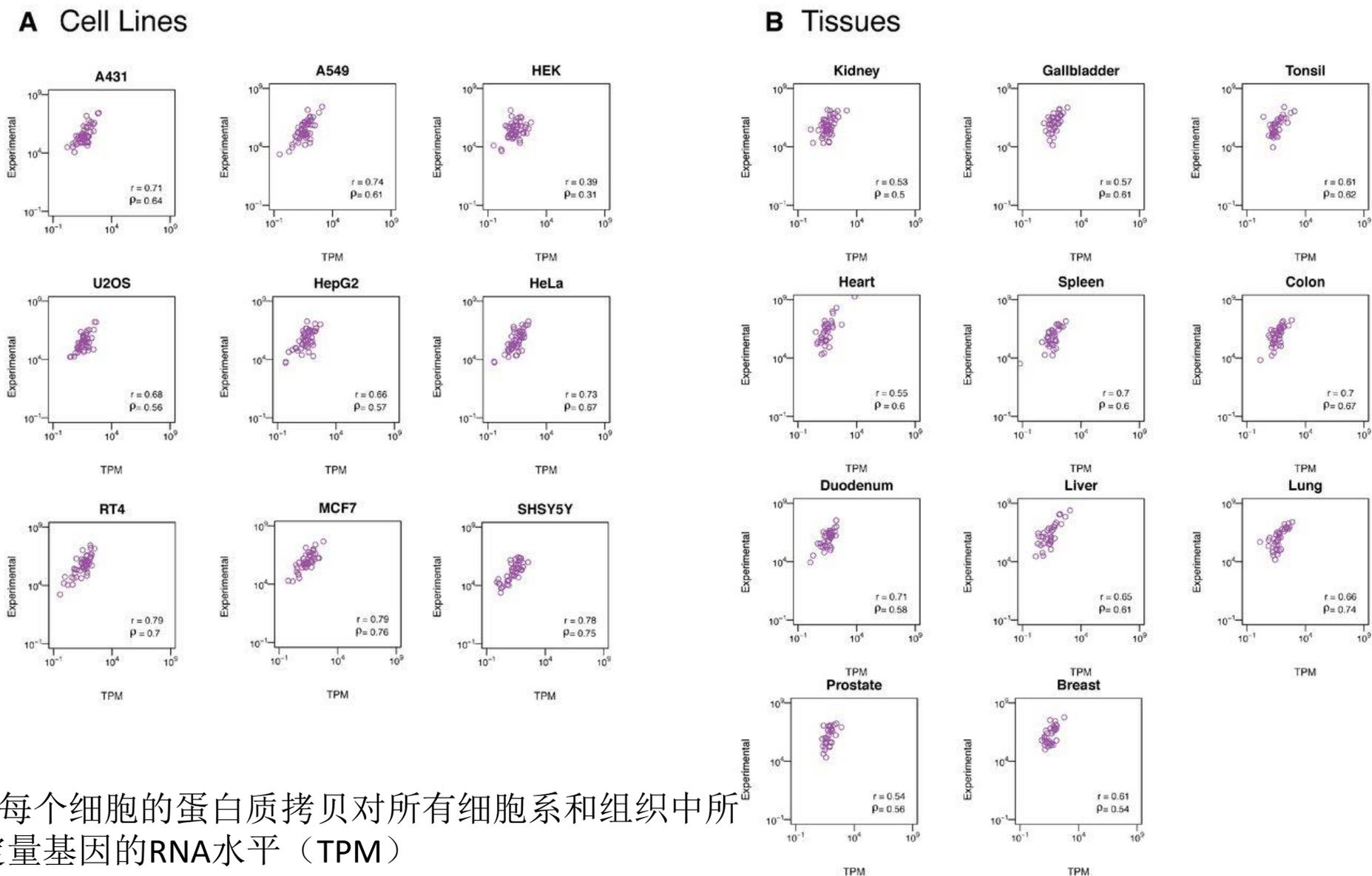
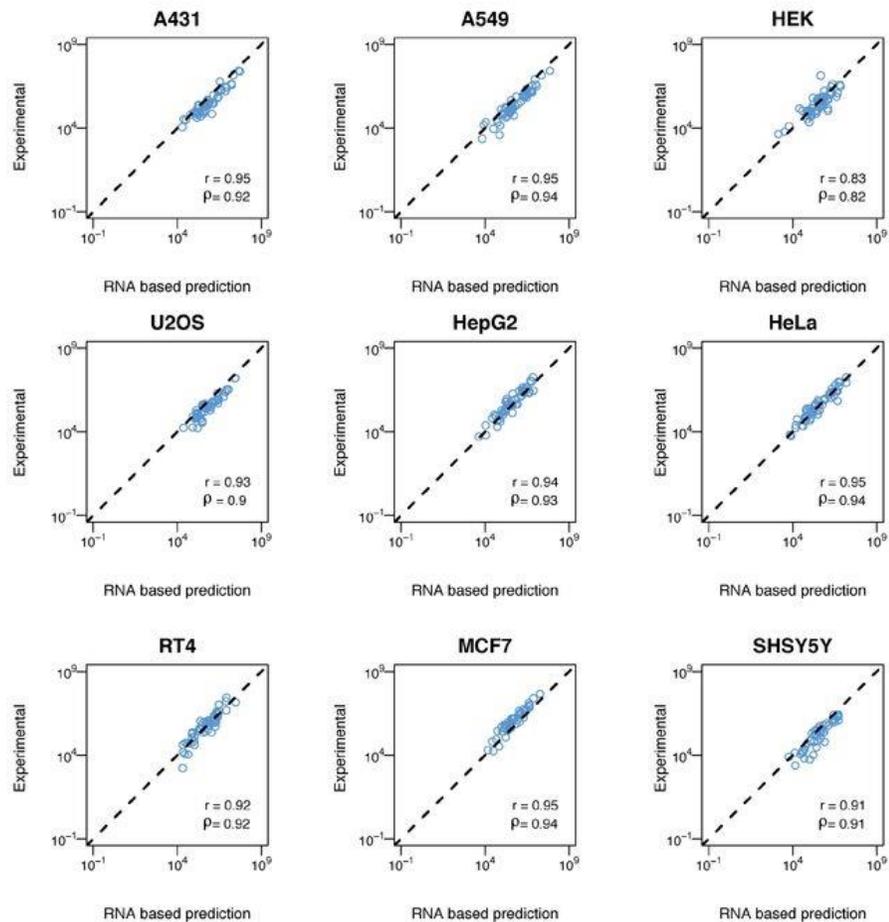


图7 每个细胞的蛋白质拷贝对所有细胞系和组织中所有定量基因的RNA水平 (TPM)

A Cell Lines



B Tissues

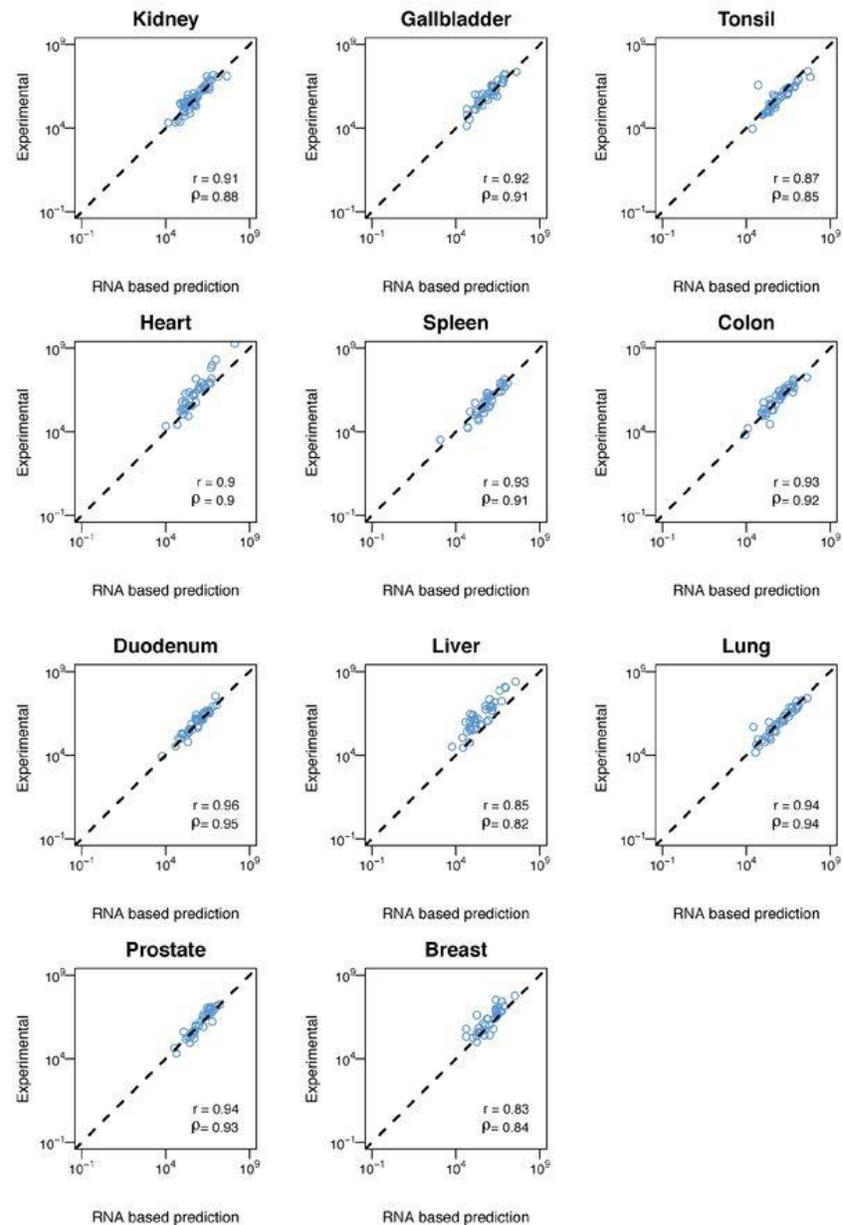


图8 基于所有细胞系和组织中所有定量基因的实验衍生的蛋白质拷贝数的TPM水平的蛋白质水平预测

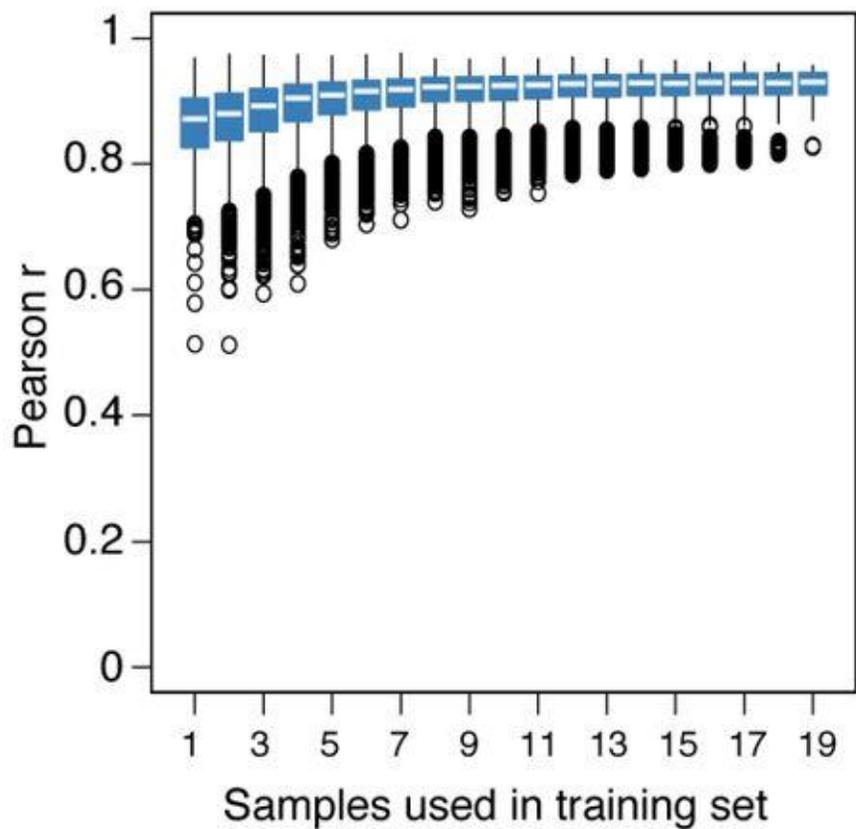
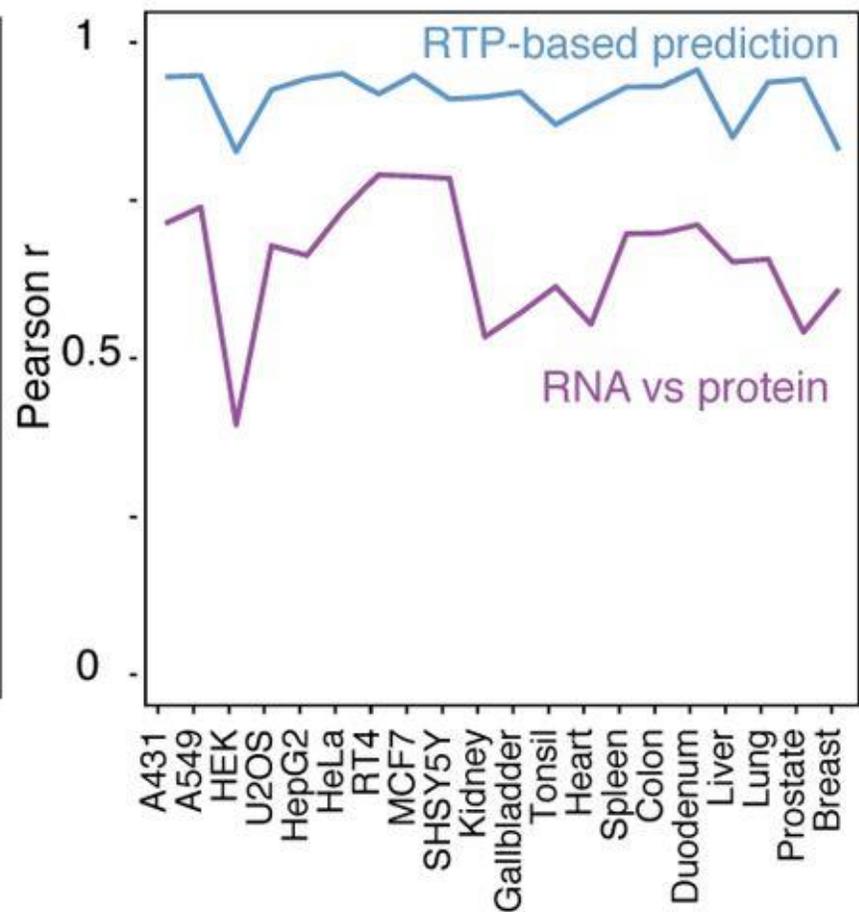
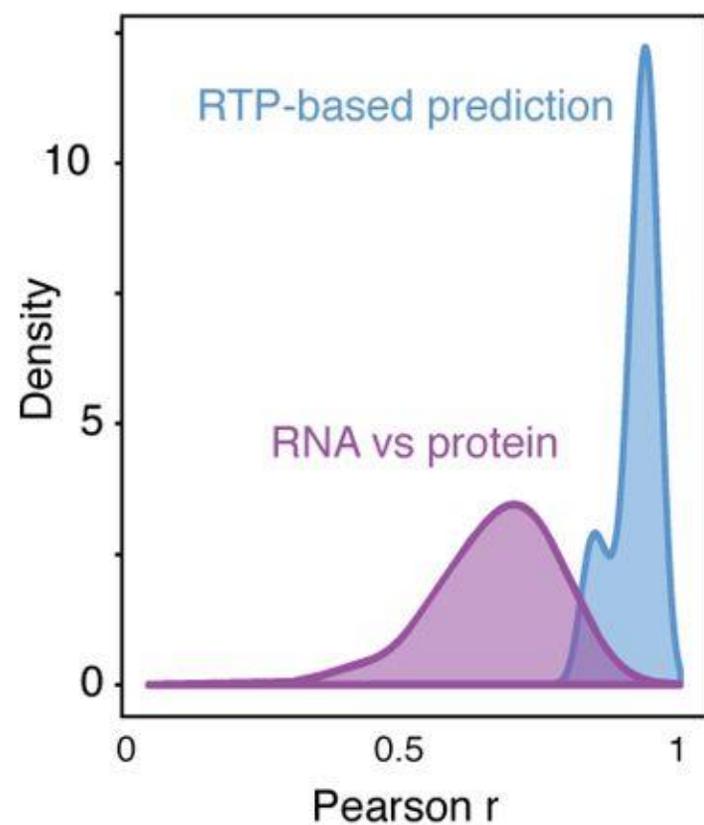
A**B****C**

图9 RNA和蛋白质水平之间的基因特异性相关性

3 讨论与总结

小结：作者在人细胞系和组织中对基因转录组预测蛋白质组进行实验和分析，最后发现基因转录组数据可以预测蛋白质拷贝数，并且引入RTP转化因子增强预测性，还暗示了基因特异性差异RTP不依赖细胞或组织。

启发：作者通过实验数据从转录组水平预测了蛋白质水平，给出了实验基础，可以通过深度学习，构建并且训练模型，从转录组水平预测蛋白质水平

改进的思考：作者分析了细胞系和组织中代表人体的一些主要器官和组织的RTP比率，需要更多实验对其他生物和细胞系组织中RTP比率

其他因素对RTP的影响，例如蛋白质修饰、非编码RNA，以及未知可能的因素，例如蛋白质长度、蛋白质翻译速率降解速率.....

作者通过组蛋白标准化来计算细胞数量，可能会出现计算误差，可以考虑用细胞计数器计数。

谢 谢！