Research Article

The Variation of Transcriptomic Perturbations is Associated with the Development and Progression of Various Diseases

Zehua Dong,[1,2] Qiyu Yan,[1,2] and Xiaosheng Wang [1,2]

[1]*Biomedical Informatics Research Laboratory, School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing 211198, China*
[2]*Big Data Research Institute, China Pharmaceutical University, Nanjing 211198, China*

Correspondence should be addressed to Xiaosheng Wang; xiaosheng.wang@cpu.edu.cn

Background. Although transcriptomic data have been widely applied to explore various diseases, few studies have investigated the association between transcriptomic perturbations and disease development in a wide variety of diseases. *Methods*. Based on a previously developed algorithm for quantifying intratumor heterogeneity at the transcriptomic level, we defined the variation of transcriptomic perturbations (VTP) of a disease relative to the health status. Based on publicly available transcriptome datasets, we compared VTP values between the disease and health status and analyzed correlations between VTP values and disease progression or severity in various diseases, including neurological disorders, infectious diseases, cardiovascular diseases, respiratory diseases, liver diseases, kidney diseases, digestive diseases, and endocrine diseases. We also identified the genes and pathways whose expression perturbations correlated positively with VTP across diverse diseases. *Results*. VTP values were upregulated in various diseases relative to their normal controls. VTP values were significantly greater in define than in possible or probable Alzheimer's disease. VTP values were significantly larger in intensive care unit (ICU) COVID-19 patients than in non-ICU patients, and in COVID-19 patients requiring mechanical ventilatory support (MVS) than in those not requiring MVS. VTP correlated positively with viral loads in acquired immune deficiency syndrome (AIDS) patients. Moreover, the AIDS patients treated with abacavir or zidovudine had lower VTP values than those without such therapies. In pulmonary tuberculosis (TB) patients, VTP values followed the pattern: active TB > latent TB > normal controls. VTP values were greater in clinically apparent than in presymptomatic malaria. VTP correlated negatively with the cardiac index of left ventricular ejection fraction (LVEF). In chronic obstructive pulmonary disease (COPD), VTP showed a negative correlation with forced expiratory volume in the first second (FEV1). VTP values increased with H. pylori infection and were upregulated in atrophic gastritis caused by H. pylori infection. The genes and pathways whose expression perturbations correlated positively with VTP scores across diseases were mainly involved in the regulation of immune, metabolic, and cellular activities. *Conclusions*. VTP is upregulated in the disease versus health status, and its upregulation is associated with disease progression and severity in various diseases. Thus, VTP has potential clinical implications for disease diagnosis and prognosis.

1. Introduction

With the recent development of next-generation sequencing (NGS) technologies, a substantial number of multiomics data associated with various diseases have been produced, including cancer, neurological disorders, cardiovascular disease, respiratory disease, digestive system disease, metabolic disease, endocrine disease, kidney and urinary system disorders, and infectious disease. In a previous study [1], we developed an algorithm, termed DEPTH, to quantify the variation of transcriptomic perturbations (VTP) in cancer, namely intratumor heterogeneity. We found that VTP value was significantly higher in cancer than in normal controls. Moreover, VTP values increased with cancer advancement, and its

increase were associated with worse clinical outcomes in cancer patients [1]. In this study, we generalized this algorithm to a wide variety of diseases and explored the association between VTP and prognosis-associated clinical features. The disease types we analyzed included neurological disorders, infectious

2. Methods

2.1. Algorithm. The algorithm is described as follows: given a transcriptome dataset, which involves g genes and m disease samples and n normal control samples; the variation of tran-

$$
\sqrt{\frac{1}{g-1}\left(\sum_{i=1}^{g}\left(\Delta(G_i, \mathrm{DS}, \mathrm{NS}_j)^2 - \frac{1}{g}\sum_{i=1}^{g}\Delta(G_i, \mathrm{DS}, \mathrm{NS}_j)^2\right)^2\right)}
$$

$$
\cdot = \sqrt{\frac{1}{g-1}\left(\sum_{i=1}^{g}\left(\left(\mathrm{ex}(G_i, \mathrm{DS}) - \frac{1}{n}\sum_{j=1}^{n}\mathrm{ex}(G_i, \mathrm{NS}_j)\right)^2 - \frac{1}{g}\sum_{i=1}^{g}\left(\mathrm{ex}(G_i, \mathrm{DS}) - \frac{1}{n}\sum_{j=1}^{n}\mathrm{ex}(G_i, \mathrm{NS}_j)\right)^2\right)^2\right)},
$$

$$
\cdot\ \Delta(G_i, \mathrm{DS}, \mathrm{NS}_j) = \mathrm{ex}(G_i, \mathrm{DS}) - \frac{1}{n}\sum_{j=1}^{n}\mathrm{ex}(G_i, \mathrm{NS}_j), \tag{1}
$$

diseases, cardiovascular diseases, respiratory diseases, liver diseases, kidney diseases, digestive diseases, and endocrine diseases. We compared VTP values between the disease state and normal controls and analyzed correlations between VTP and disease progression or severity.

where $ex(G_i, \mathrm{DS})$ indicates the expression value of gene G_i in DS, and $\mathrm{ex}(G_i, \mathrm{NS}_j)$ indicates the expression value of G_i in the normal sample NS_j.

2.2. Datasets. We downloaded transcriptome datasets for various diseases from the NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) and analyzed these datasets with the algorithm. The datasets were associated with various types of diseases, including neurological disorders (Alzheimer's disease (AD) and schizophrenia (SCZ)), infectious diseases (COVID-19, acquired immune deficiency syndrome (AIDS), hepatitis B virus (HBV) infection, tuberculosis (TB), and malaria), cardiovascular diseases (acute myocardial infarction, dilated cardiomyopathy, idiopathic or ischemic cardiomyopathy, and heart failure), respiratory diseases (chronic obstructive pulmonary disease), liver diseases (chronic hepatitis B and liver cirrhosis), kidney diseases (nephrotic syndrome, uremia, focal segmental glomerulosclerosis and glomerular disease), digestive diseases (inflammatory bowel disease and helicobacter pylori infection), and endocrine diseases (diabetes). A description of these datasets is shown in Table 1.

2.3. Data Preprocessing. For RNA-Seq gene expression values, we normalized them by the TPM method. For microarray gene expression values, we used the normalization methods recommended by related platforms. A description of the normalization methods for the datasets analyzed was provided in Supplementary Table S1. All normalized expression values were transformed by $\log 2(x+1)$ before subsequent analyses.

scriptomic perturbations (VTP) of a disease sample DS is defined as

2.4. Statistical Analysis and Visualization. We employed the Mann–Whitney U test (one-tailed) to compare VTP values between two classes of samples, and the Kruskal-Wallis test to compare VTP values among more than two classes of samples. We utilized the Spearman method to assess the correlation between VTP values and other variables and reported the correlation coefficients (ρ) and P values. To correct for P values in multiple tests, we utilized the Benjamini and Hochberg method to calculate the false discovery rate (FDR) [2]. All statistical analyses were performed in the R programming environment (version 4.1.2). The R packages "ggplot2", "ggpubr", and "ggstatsplot" were used for data visualization.

2.5. Identifying Genes and Pathways whose Expression Perturbations Have Significant Positive Correlations with VTP across Diverse Diseases. In each dataset, we identified the genes satisfying that $|\Delta(G_i, \mathrm{DS}, \mathrm{NS}_j)|$ significantly and positively correlated with VTP values in all disease samples using a threshold of the Spearman correlation test FDR < 0.1. For each disease with n datasets analyzed, we identified the genes which satisfied the prior condition at least $n-1$ datasets. These genes were defined as the genes having significant positive correlations of expression perturbations with VTP in specific diseases. Among them, the genes identified in common in at least 5 specific diseases were defined as the genes whose expression perturbations had significant positive correlations with VTP across diseases. By inputting the genes associated with VTP across diseases into the GSEA web tool [3], we obtained the KEGG pathways [4] having

TABLE 1: Summary of the datasets analyzed.

Disease	Dataset	Platform	Sample size	
			Patients	Controls
Alzheimer's disease	GSE63063	GPL6947	145	104
		GPL10558	139	134
	GSE84422	GPL96	737	214
			328 (definite AD)	
			180 (probable AD)	
			229 (possible AD)	
		GPL97	737	214
			328 (definite AD)	
			180 (probable AD)	
			229 (possible AD)	
		GPL570	74	28
			34 (definite AD)	
			23 (probable AD)	
			17 (possible AD)	
	GSE118553	GPL10558	167	100
	GSE140831	GPL15988	599	530
	GSE158233	GPL20828	20	10
	Danio rerio dataset			
Schizophrenia	GSE38484	GPL6947	106	96
	GSE53987	GPL570	48	55
	GSE87610	GPL13667	65	72
	GSE93577	GPL13667	70	71
	GSE93987	GPL13158	102	106
Corona virus Disease 2019	GSE152075	GPL18573	430	54
	GSE157103	GPL24676	102	26
			50 (ICU)	
			42 (MVS)	
	GSE161731	GPL24676	46	19
	GSE198449	GPL24676	149	22
			85 (asymptomatic)	
			64 (symptomatic)	
Acquired immune deficiency syndrome	GSE18233	GPL6884	153	3
			16 (EC)	
	GSE87620	GPL10558	83	10
			51 (EC)	
			32 (HAART-treated)	
	GSE104640	GPL10558	188	60
Hepatitis B virus infection	GSE83148	GPL570	122	6
	GSE114783	GPL15491	3	3
	GSE121248	GPL570	70	37
Tuberculosis	GSE28623	GPL4133	71	37
			46 (TB)	
			25 (LTB)	
	GSE153340	GPL21185	18	4
	GSE152532	GPL10558	50	11
			42 (TB)	

TABLE 1: Continued.

Disease	Dataset	Platform	Sample size Patients	Controls
			8 (LTB)	
Malaria	GSE1124	GPL96	20 5 (asymptomatic) 5 (uncomplicated) 5 (severe)	5
		GPL97	20 5 (asymptomatic) 5 (uncomplicated) 5 (severe)	5
	GSE5418	GPL96	22	22
	GSE34404	GPL10558	94	61
Cardiovascular disease	GSE1145	GPL570	53	37
	GSE5406	GPL96	194	16
	GSE17800	GPL570	40	8
	GSE33463	GPL6947	49	41
	GSE48060	GPL570	31	21
	GSE62646	GPL6244	28	14
	GSE66360	GPL570	49	50
	GSE74144	GPL13497	28	8
	GSE109048	GPL17586	38	19
	GSE120895	GPL570	47	8
Respiratory disease	GSE5058	GPL570	15	24
	GSE42057	GPL570	94	42
	GSE55962	GPL13667	24	82
	GSE103174	GPL13667	37	16
	GSE112811	GPL570	20	18
	GSE151052	GPL17556	77	40
	GSE32147 Rattus norvegicus dataset	GPL6101	56	4
Liver disease	GSE14323	GPL96	58	19
	GSE77627	GPL14951	40 32 (cirrhosis) 18 (noncirrhosis)	14
	GSE135501	GPL13667	40 16 (white tongue) 24 (yellow tongue)	14
	GSE36533 Marmota monax dataset	GPL15354	11	10
Kidney disease	GSE37171	GPL570	63	20
	GSE104948	GPL22945	53	18
	GSE108113	GPL19983	269	5
	GSE133288	GPL19983	239	5
Digestive disease	GSE16879	GPL570	61	12
	GSE27411	GPL6255	12	6
Endocrine disease	GSE9006	GPL96	55	24

Table 1: Continued.

Disease	Dataset	Platform	Sample size	
			Patients	Controls
			T1D (43)	
			T2D (12)	
			55	
		GPL97	T1D (43)	24
			T2D (12)	
	GSE19420	GPL570	30	12
			57	
	GSE35725	GPL570	46 (recent onset)	44
			11 (longstanding)	

Note: EC: Elite controllers. HAART: Highly active antiretroviral therapy. LTB: Latently tuberculosis. T1D: Type 1 diabetes. T2D: Type 2 diabetes.

significant positive correlations of their expression perturbations with VTP across diseases using a threshold of FDR<0.1 .

3. Results

3.1. Neurological Disorder. AD is a progressive neurodegenerative disease [5]. In four transcriptome datasets for AD (GSE63063 [6], GSE118553 [7], GSE140831, and GSE84422 [8]), the VTP values were significantly larger in AD patients than in normal controls ($P < 0.001$) (Figure 1(a)). In GSE84422, VTP values were significantly larger in define than in possible or probable AD ($P = 0.02$) (Figure 1(a)). In addition, we analyzed correlations between VTP and several measures of the degree of AD progression in GSE84422, including clinical dementia rating, Braak neurofibrillary tangle score, average neuritic plaque density, sum of consortium to establish a registry for Alzheimer's diseases (CERAD) rating scores in multiple brain regions, and sum of neurofibrillary tangles density in multiple brain regions. Notably, VTP displayed significant positive correlations with these measures ($P < 0.01$) (Figure 1(a)). Mutations in PSEN2 may result in early-onset AD. In GSE158233 [9], Barthelson et al. generated transcriptomes of two-types of PSEN2-mutated (psen2$^{T141\_L142delinsMISLISV}$ and psen2^{N140fs}) lines of zebrafish brains and transcriptomes of their wild type siblings. We observed that VTP values were remarkably greater in PSEN2-mutated zebrafish brains than in their wild type controls ($P < 0.03$) (Figure 1(a)).

Schizophrenia (SCZ) is a severe psychotic disorder characterized by relapsing incidences of psychosis [10]. In four transcriptome datasets (GSE38484 [11], GSE87610 [12], GSE93577 [13], and GSE93987 [14]) generated from SCZ patients and normal controls, VTP values were consistently greater in SCZ patients than in normal controls ($P < 0.02$) (Figure 1(b)).

Taken together, these results indicate that VTP is augmented in certain neurological disorders (such as AD and SCZ) and grows with disease progression.

3.2. Infectious Disease. COVID-19 is a highly contagious disease caused by SARS-CoV-2 infection and is currently widespread around the globe. This disease has caused more than 552 million cases and 6.3 million deaths as of July 1, 2022 [15]. In four transcriptome datasets (GSE152075 [16], GSE157103 [17], GSE161731 [18], and GSE198449 [19]) for COVID-19 patients, VTP values were significantly greater in COVID-19 patients than in normal controls ($P < 0.01$) (Figure 2(a)). Notably, in GSE157103, VTP values were significantly greater in intensive care unit (ICU) COVID-19 patients than in non-ICU patients ($P < 0.001$) (Figure 2(a)). Moreover, COVID-19 patients requiring mechanical ventilatory support (MVS) had greater VTP values than those not requiring MVS ($P < 0.001$) (Figure 2(a)). In addition, VTP displayed a significant positive correlation with the scores of the sequential organ failure assessment (SOFA) ($P = 0.006$; $\rho = 0.36$) (Figure 2(a)), which indicates the severity of ICU patients.

Acquired immune deficiency syndrome (AIDS) is a chronic condition resulting from infection with the human immunodeficiency virus (HIV) [20]. In three transcriptome datasets (GSE18233 [21], GSE87620 [22], and GSE104640 [23]) for AIDS patients, VTP values were significantly upregulated in AIDS patients versus normal controls ($P < 0.001$) (Figure 2(b)). In GSE18233, VTP correlated positively with viral loads ($P = 0.002$; $\rho = 0.28$) (Figure 2(b)). In GSE87620, the AIDS patients with highly active antiretroviral therapy had greater VTP values than elite controllers, who were the AIDS patients with undetectable levels of HIV replication not receiving antiretroviral therapy ($P = 0.01$) (Figure 2(b)). In addition, in GSE62117 [24], the AIDS patients treated with abacavir or zidovudine had lower VTP values than those without such therapies ($P < 0.05$) (Figure 2(b)).

Hepatitis B virus (HBV) infection is a major etiologic factor for hepatocellular carcinoma [25]. In three transcriptome datasets (GSE83148 [26], GSE114783 [27], and GSE121248 [28]), VTP values were greater in HBV-infected patients than in normal controls ($P < 0.01$) (Figure 2(c)).

Pulmonary tuberculosis (TB) is an infectious disease caused by Mycobacterium tuberculosis attacking lungs [29]. In three transcriptome datasets (GSE28623 [30], GSE153340 [31], and GSE152532 [32]), TB patients showed
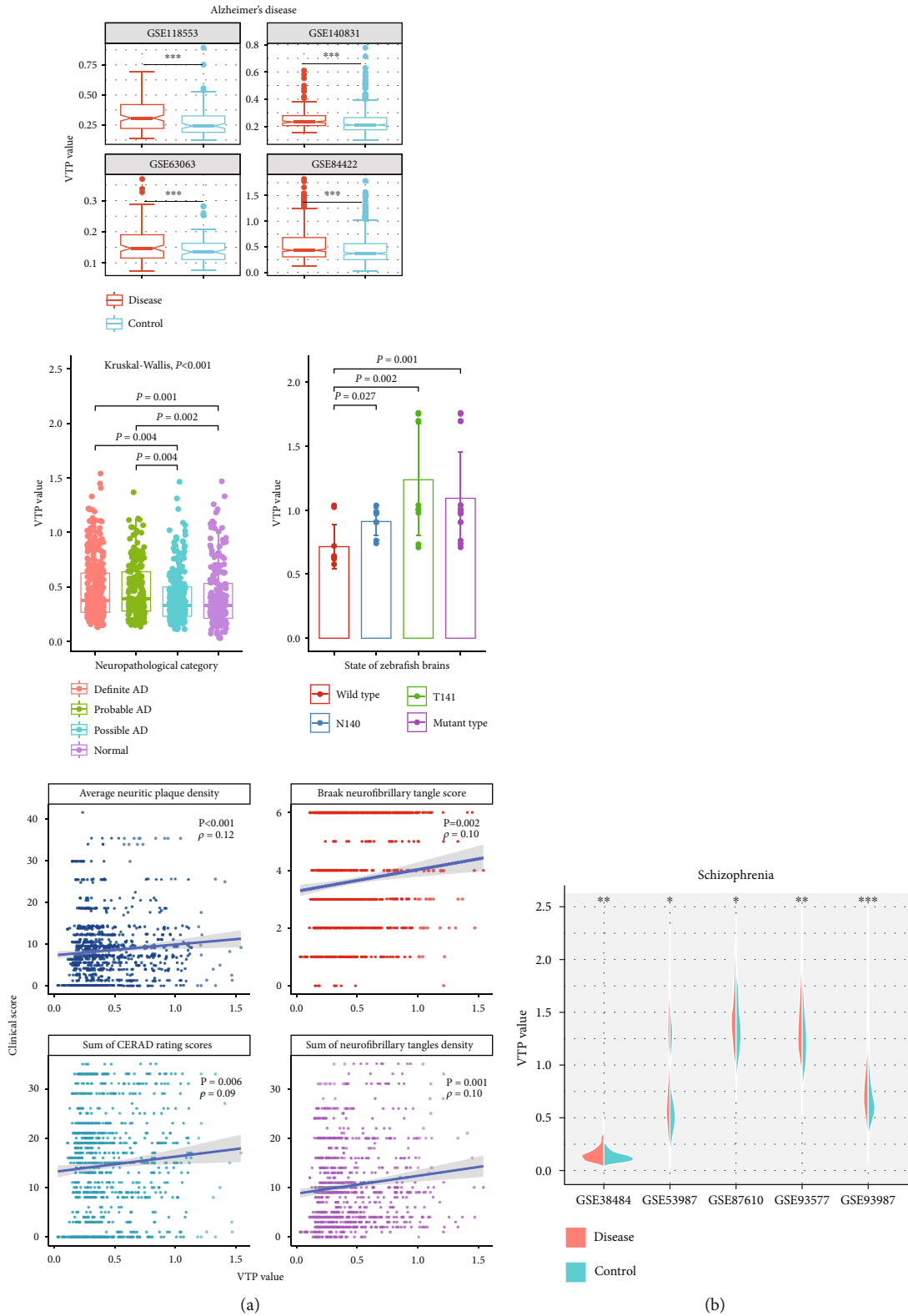
FIGURE 1: Associations between the VTP measure and disease development and progression in neurological disorder. (a) VTP values are significantly greater in AD patients than in normal controls, larger in define than in possible or probable AD, and increase with AD progression. The measures of Braak neurofibrillary tangle score, average neuritic plaque density, sum of CERAD rating scores in multiple brain regions, and sum of neurofibrillary tangles density in multiple brain regions represent the degree of AD progression. VTP values are remarkably greater in PSEN2-mutated zebrafish brains than in their wild type controls. (b) VTP values are significantly greater in SCZ patients than in normal controls. AD: Alzheimer's disease. N140: psen2^{N140fs}. T141: psen2$^{T141\_L142delinsMISLISV}$. SCZ: Schizophrenia. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. They also apply to the following figures.
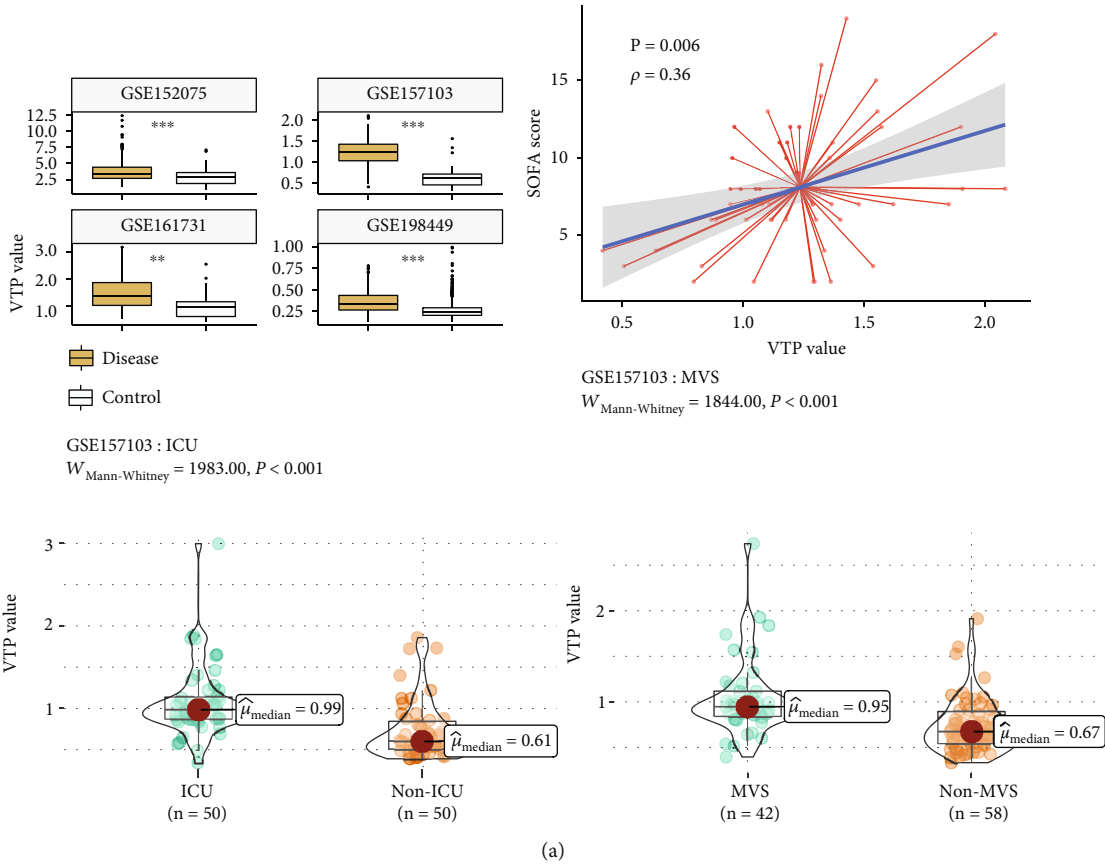
Figure 2: Continued.

EC: HIV elite controllers
HAART: Patients treated with highly active antiretroviral therapy

AZT : Patients treated with zidovudine
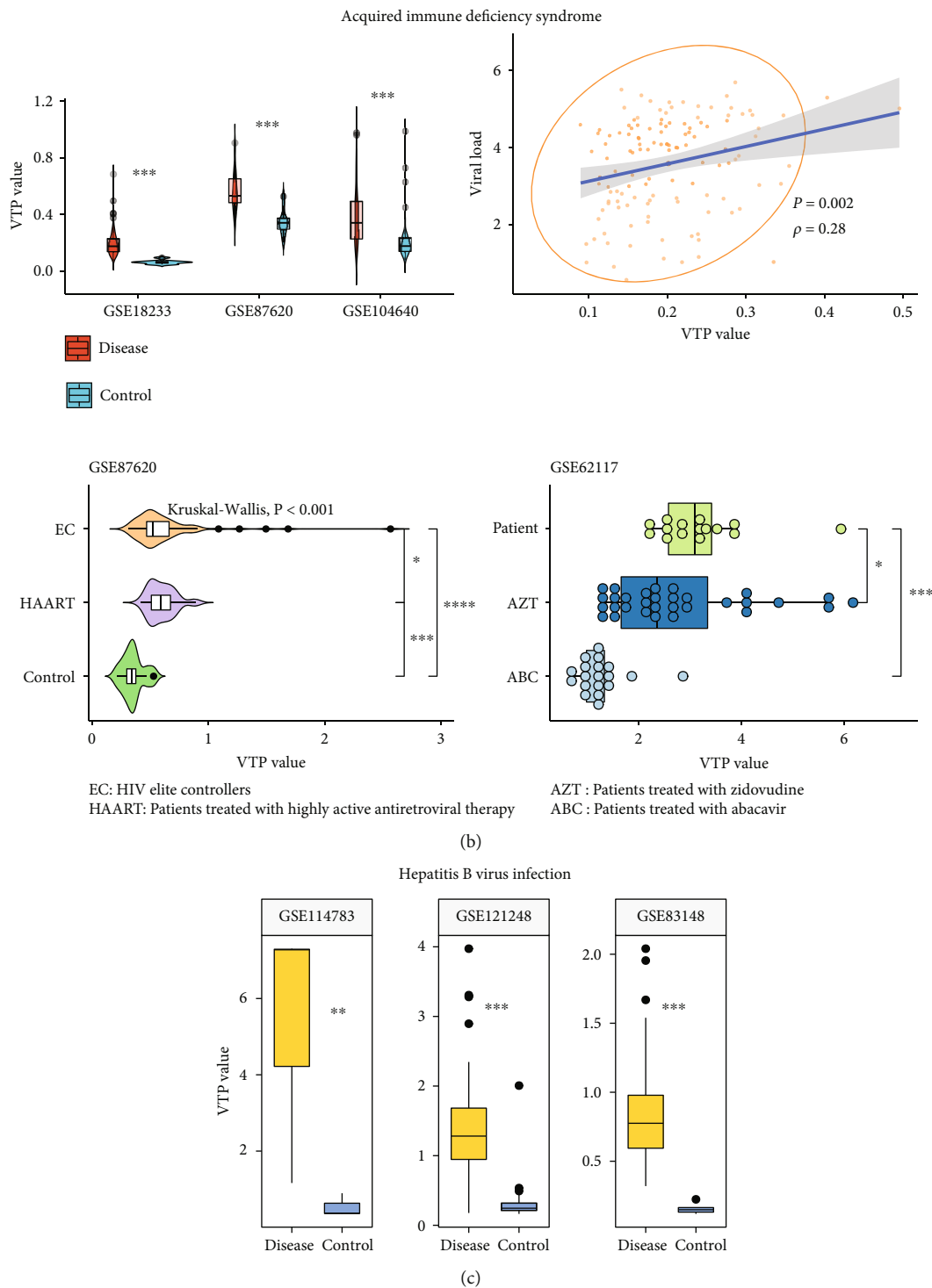ABC : Patients treated with abacavir
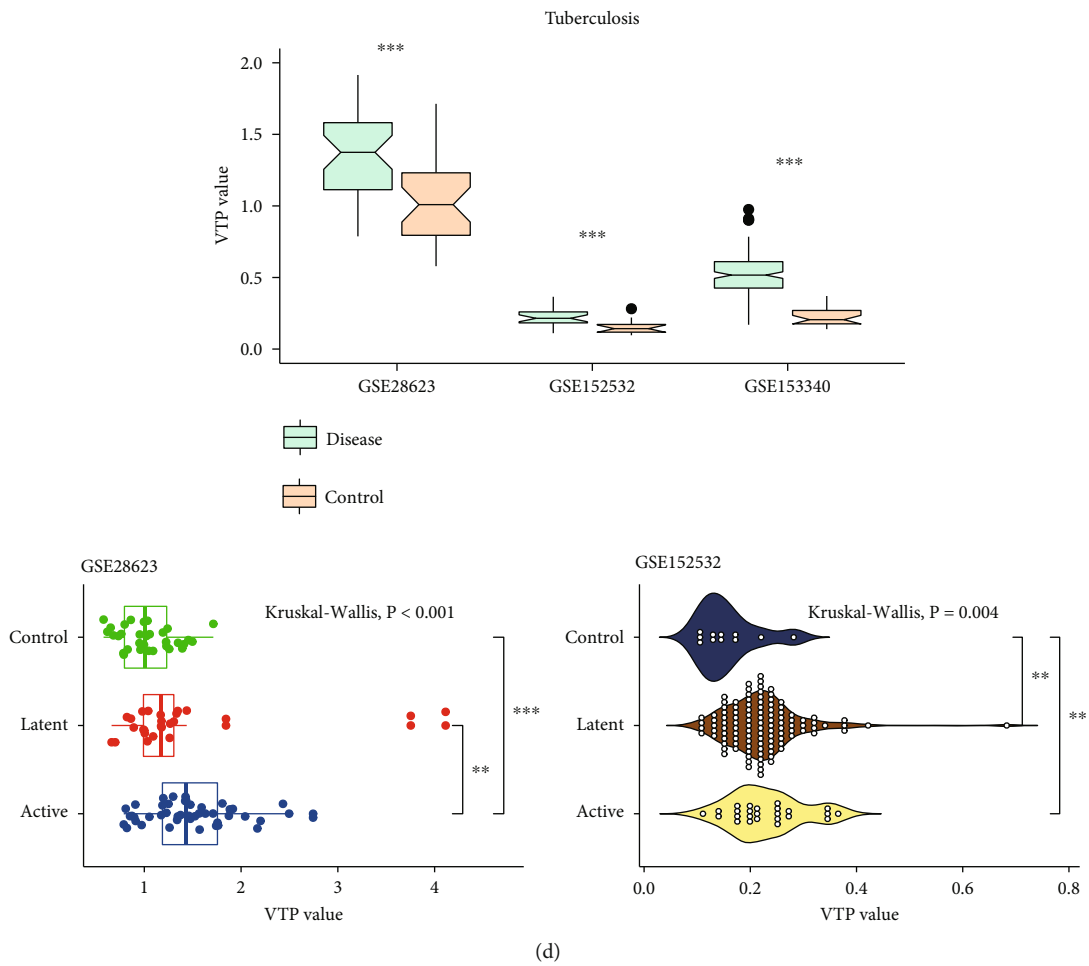
(b)

(c)

FIGURE 2: Continued.
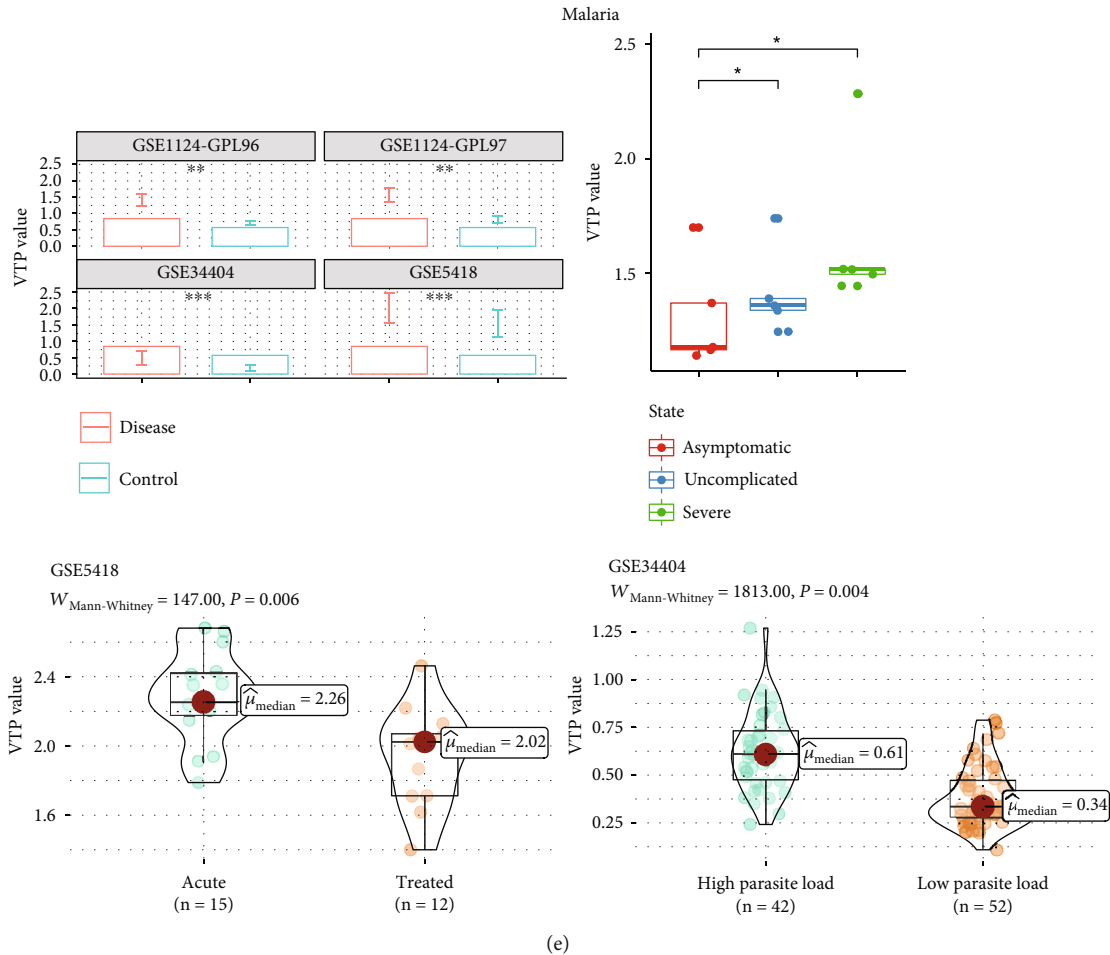
(d)

Figure 2: Continued.

FIGURE 2: Associations between the VTP measure and disease development and progression in infectious disease. (a) VTP values are significantly greater in COVID-19 patients than in normal controls and increase with disease severity. (b) VTP values are significantly greater in AIDS patients than in normal controls, greater in AIDS patients without treatment than in AIDS patients with treatment, and increase with disease severity. (c). VTP values are significantly greater in HBV-infected patients than in normal controls. (d) VTP values are significantly greater in TB patients than in normal controls and increase with disease progression. (e) VTP values are significantly greater in malaria patients than in normal controls and increase with disease severity. ICU: intensive care unit. MVS: mechanical ventilatory support. SOFA: sequential organ failure assessment. AIDS: acquired immune deficiency syndrome. HBV: hepatitis B virus. TB: tuberculosis.

greater VTP values than normal controls ($P < 0.001$) (Figure 2(d)). Moreover, in GSE152532 and GSE28623 [30], VTP values likely followed the pattern: active TB > latent TB > normal controls (Figure 2(d)).

Malaria is a serious disease caused by a parasite and is a major cause of death globally [33]. In three transcriptome datasets (GSE1124 [34], GSE5418 [35], and GSE34404 [36]), malaria patients had greater VTP values than normal controls ($P < 0.01$) (Figure 2(e)). Moreover, in GSE34404, the high parasitemia group had significantly larger VTP values than the low parasitemia group of malaria patients ($P = 0.004$) (Figure 2(e)). In GSE1124, VTP values likely followed the pattern: malaria associated with severe anemia > uncomplicated malaria > asymptomatic infection (Figure 2(e)). In addition, in another transcriptome dataset (GSE5418), VTP values were greater in clinically apparent than in presymptomatic malaria ($P = 0.006$) (Figure 2(e)).

Collectively, these results support that VTP is upregulated in infectious diseases and increases with disease severity.

3.3. Cardiovascular Disease.
Heart disease is the leading cause of death worldwide [37]. In numerous transcriptome datasets of heart disease, such as GSE1145, GSE5406 [38], GSE17800 [39], GSE48060 [40], GSE66360 [41], GSE109048 [42], and GSE120895 [43], VTP values were significantly greater in patients than in normal controls ($P < 0.05$) (Figure 3(a)). In GSE17800, VTP had a significant negative correlation with the cardiac index of left ventricular ejection fraction (LVEF) ($P = 0.035$; $\rho = -0.33$) (Figure 3(b)). In GSE62646 [44], the VTP values calculated based on gene expression patterns in leukocytes from acute myocardial infarction patients followed the pattern: the 1st day of myocardial infarction > after 4 − 6 days > after 6

(a)

GSE17800 (Dilated cardiomyopathy)

$\log_e (S) = 9.56$, $P = 0.035$, $\hat{\rho}_{Spearman} = -0.33$, $CI_{95\%}$ [$-0.59, -0.02$], $n_{pairs} = 40$

(b)

GSE62646 (Acute myocardial infarction(MI))
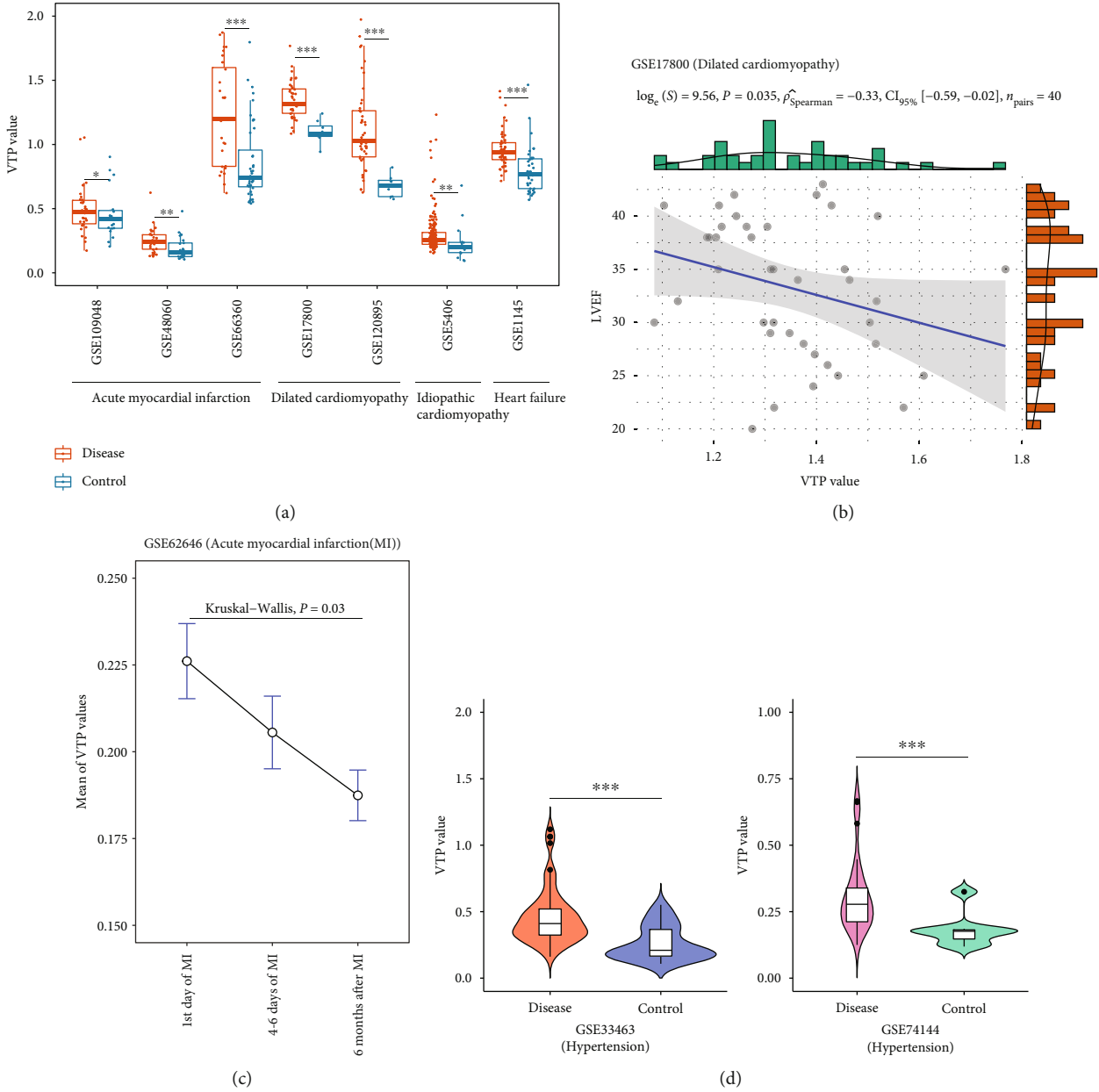
Kruskal–Wallis, $P = 0.03$

(c)

(d)

FIGURE 3: Associations between the VTP measure and disease development and progression in cardiovascular disease. (a) VTP values are significantly greater in heart disease patients than in normal controls and increase with disease severity. (b) VTP values correlate negatively with the cardiac index of LVEF. (c) VTP values decrease with the remission of acute myocardial infarction. (d) VTP values are significantly greater in hypertension patients than in normal controls. LVEF: left ventricular ejection fraction.

months (Figure 3(c)). In addition, in two transcriptome datasets (GSE33463 [45] and GSE74144) for hypertension, VTP values were significantly larger in patients than in normal controls ($P < 0.001$) (Figure 3(d)). Altogether, these results suggest that VTP is upregulated in cardiovascular diseases and decreases with disease remission.

3.4. Respiratory Disease. Respiratory diseases are the diseases affecting the organs and tissues involved in gas exchange in air-breathing animals [46]. Some of the most common respiratory diseases include obstructive lung disease, restrictive lung disease, and respiratory tract infections. In many tran-

scriptome datasets of respiratory diseases, such as GSE112811, GSE42057 [47], GSE55962 [48], GSE103174, and GSE151052, VTP values were significantly larger in patients than in normal controls ($P < 0.05$) (Figure 4(a)). In chronic obstructive pulmonary disease (COPD), forced expiratory volume in the first second (FEV1) and ratio of FEV1 to forced vital capacity (FVC) are crucial in evaluating the severity of disease [49]. In GSE103174, which is a transcriptome dataset for COPD, VTP showed negative correlations with both FEV1 ($P = 0.018$; $\rho = -0.39$) and FEV1/FVC ($P = 0.067$; $\rho = -0.31$) (Figure 4(b)). The transcriptome dataset GSE32147 [50] is gene expression profiles in lung
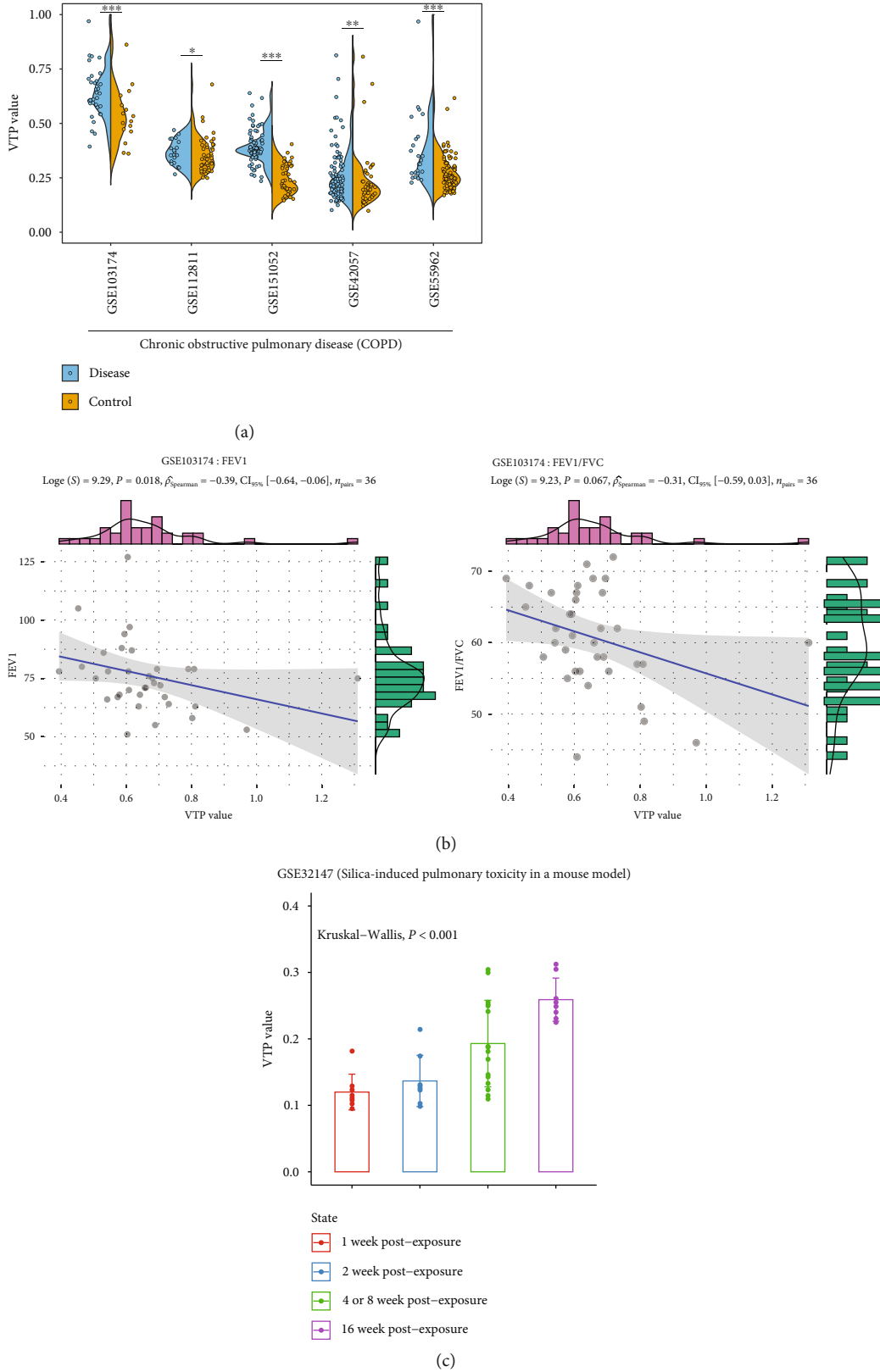
(a)

GSE103174 : FEV1
Loge (S) = 9.29, P = 0.018, $\hat{\rho}_{Spearman}$ = −0.39, $CI_{95\%}$ [−0.64, −0.06], n_{pairs} = 36

GSE103174 : FEV1/FVC
Loge (S) = 9.23, P = 0.067, $\hat{\rho}_{Spearman}$ = −0.31, $CI_{95\%}$ [−0.59, 0.03], n_{pairs} = 36

(b)

GSE32147 (Silica-induced pulmonary toxicity in a mouse model)

(c)

Figure 4: Associations between the VTP measure and disease development and progression in respiratory disease. (a) VTP values are significantly greater in respiratory disease patients than in normal controls. (b) VTP values correlate negatively with FEV1 and ratios of FEV1/FVC in COPD. (c) VTP values increase steadily with the progression of silica-induced pulmonary toxicity in a mouse model. COPD: chronic obstructive pulmonary disease. FEV1: forced expiratory volume in the first second. FVC: forced vital capacity.
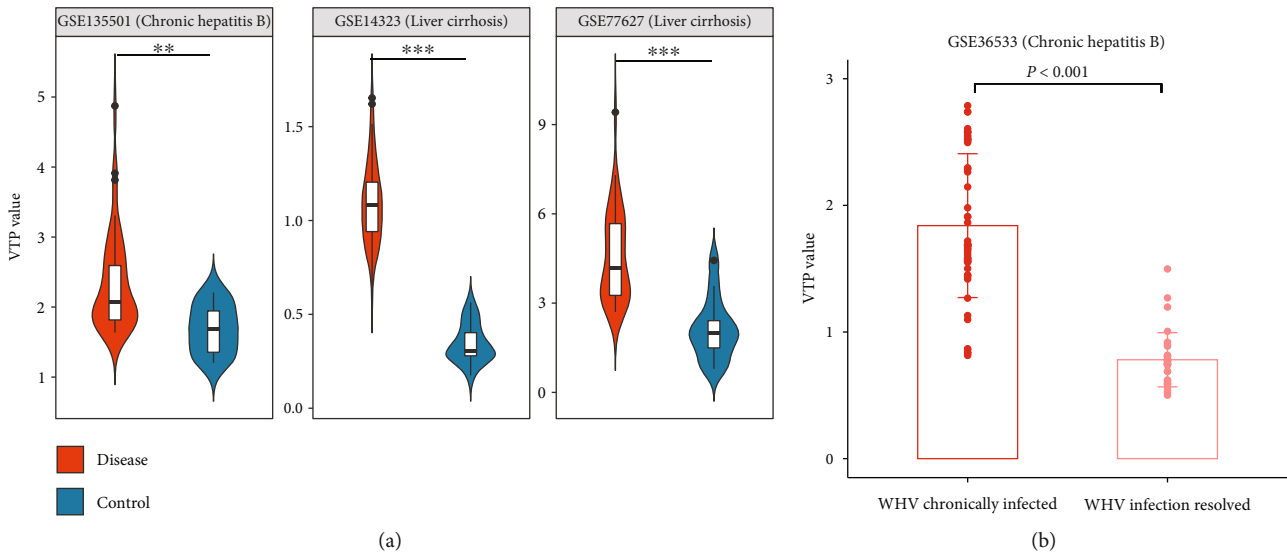
FIGURE 5: Associations between the VTP measure and disease development and progression in liver disease. (a) VTP values are significantly greater in liver disease patients than in normal controls. (b) VTP values are greater in WHV chronically infected than in infection resolved woodchuck in an animal model. WHV: woodchuck hepatitis virus.
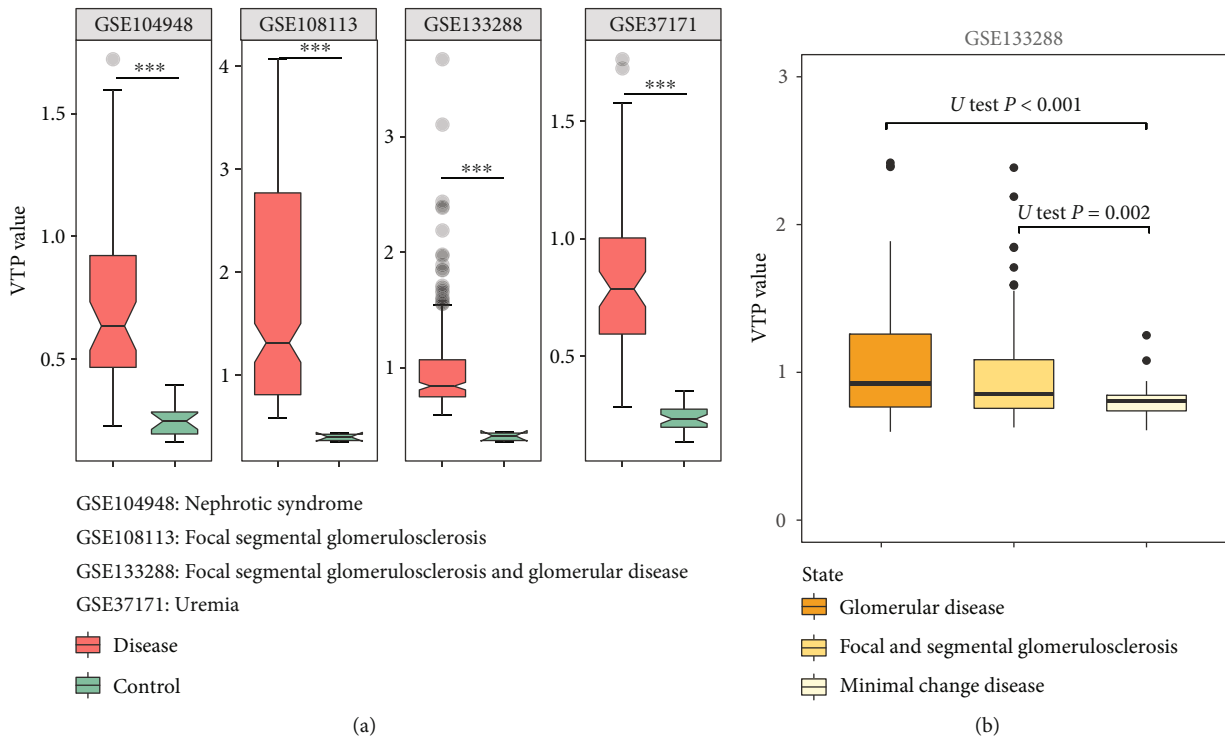

FIGURE 6: Associations between the VTP measure and disease development and progression in kidney disease. (a) VTP values are significantly greater in kidney disease patients than in normal controls. (b) VTP values correlate positively with disease severity in kidney disease.
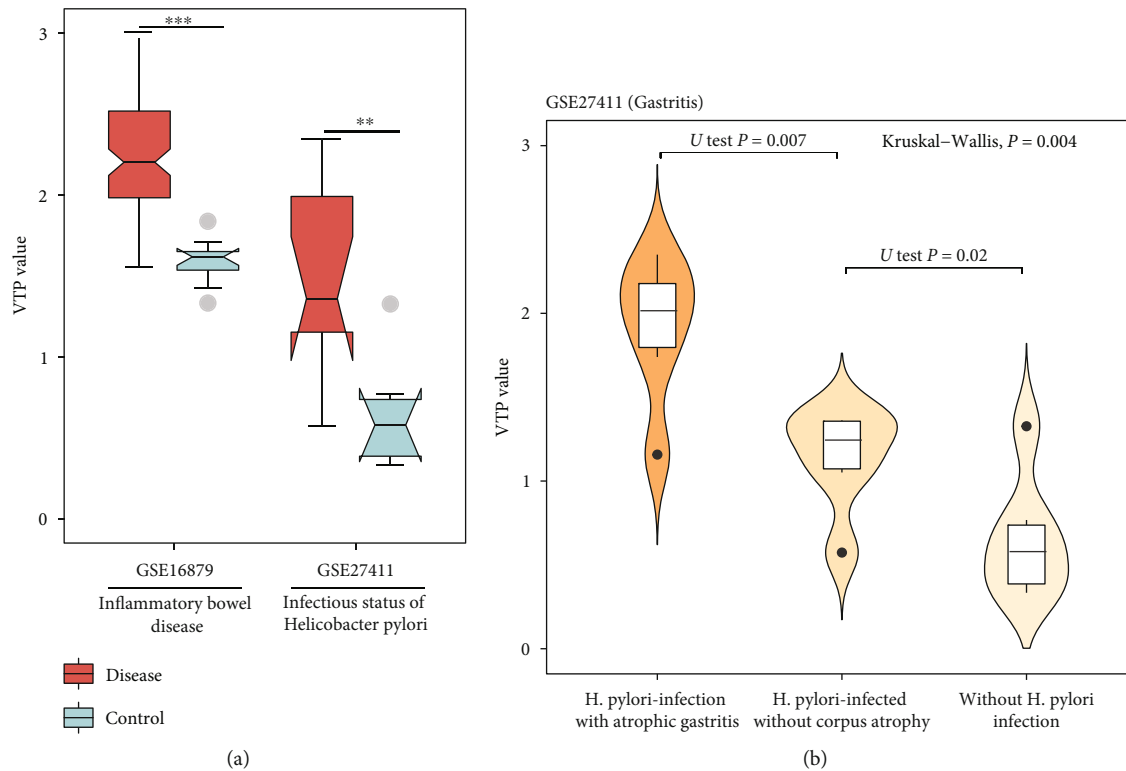
FIGURE 7: Associations between the VTP measure and disease development and progression in digestive disease. (a) VTP values are significantly greater in digestive disease patients than in normal controls. (b) VTP values correlate positively with disease severity in atrophic gastritis.

samples of rats exposed to crystalline silica. We observed that VTP values increased steadily with the progression of silica-induced pulmonary toxicity: 1 week of exposed to crystalline silica < 2 weeks < 4 or 8 weeks < 16 weeks (Figure 4(c)).

Collectively, these results support that VTP is upregulated in respiratory diseases and is negatively associated with their clinical outcomes.

3.5. Liver Disease.
In three transcriptome datasets (GSE14323 [51], GSE77627, and GSE135501) for liver diseases, VTP values were significantly larger in patients than in normal controls ($P < 0.01$) (Figure 5(a)). The transcriptome dataset GSE36533 [52] is gene expression profiles in woodchuck infected with woodchuck hepatitis virus (WHV), an animal model for studying the human HBV. Notably, VTP values are greater in WHV chronically infected than in infection resolved woodchuck ($P < 0.001$) (Figure 5(b)).

3.6. Kidney Disease.
In four transcriptome datasets (GSE37171 [53], GSE104948 [54], GSE108113 [54], and GSE133288 [55]) for kidney disease, VTP values were significantly greater in patients than in normal controls ($P < 0.001$) (Figure 6(a)). In addition, in GSE133228, VTP values were significantly larger in focal segmental glomerulosclerosis and glomerular disease than in minimal change disease ($P < 0.01$) (Figure 6(b)). It indicates that VTP values increase with disease progression in kidney disease.

3.7. Digestive Disease.
In two transcriptome datasets (GSE16879 [56] and GSE27411 [57]) for digestive disease, VTP values were significantly larger in patients than in normal controls ($P < 0.01$) (Figure 7(a)). GSE27411 is a transcriptome dataset for patients with different stages of Helicobacter pylori (H. pylori) infection. Interestingly, we found that VTP values were significantly different among different stages of H. pylori infection and followed the pattern: without current H.pylori infection < H.pylori − infected without corpus atrophy < with current or past H.pylori − infection with corpus-predominant atrophic gastritis (Figure 7(b)). These results collectively support that VTP is upregulated in digestive diseases and increases with disease severity.

3.8. Endocrine Disease.
Diabetes is a metabolic disease that causes high blood sugar to cause many chronic health problems, such as cardiovascular diseases, vision damage, and kidney disease [58]. In two transcriptome datasets (GSE9006 [59] and GSE19420 [60]) for diabetes, VTP values were significantly greater in patients than in normal controls ($P < 0.05$) (Figure 8(a)). Moreover, in the transcriptome dataset GSE35725 [61] for diabetes, VTP values were significantly greater in recent onset diabetes patients than in long-standing diabetes patients ($P < 0.001$) (Figure 8(b)).

3.9. Genes and Pathways whose Expression Perturbations Correlate Positively with VTP across Diseases.
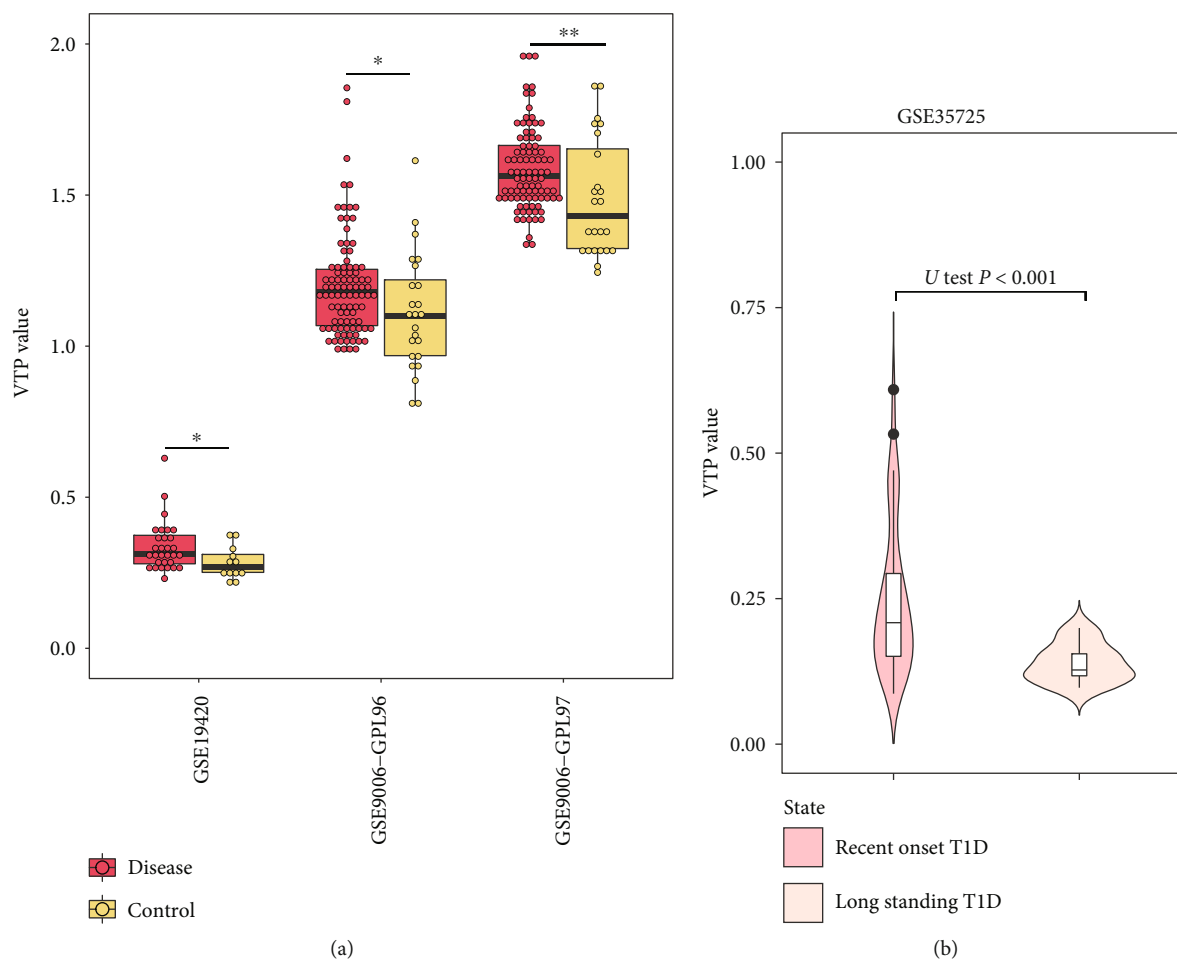We identified

FIGURE 8: Associations between the VTP measure and disease development and progression in endocrine disease. (a) VTP values are significantly greater in diabetes patients than in normal controls. (b) VTP values are significantly greater in recent onset diabetes patients than in longstanding diabetes patients. T1D: Type 1 diabetes.

369 genes whose expression perturbations showed significant positive correlations with VTP values across diseases (Supplementary Table S2). Notably, many of these genes are involved in immune regulation (such as CD2, CD247, CD300A, CD2AP, CD28, CD47, CD53, CD7, and CXCR2), cell cycle (such as CCND2, CDK4, and SKP2), and metabolism (such as LDHA, LDHB, PDHA1, GLO1, and ME2). Furthermore, we identified 58 KEGG pathways showing significant positive correlations of expression perturbations with VTP across diseases. Notably, many of these pathways are immune pathways, including natural killer cell-mediated cytotoxicity, T cell receptor signaling, B cell receptor signaling, chemokine signaling, cell adhesion molecules, Fc gamma R-mediated phagocytosis, leukocyte transendothelial migration, Fc epsilon RI signaling, hematopoietic cell lineage, Toll-like receptor signaling, Jak-STAT signaling, cytokine-cytokine receptor interaction, intestinal immune network for IgA production, and NOD-like receptor signaling (Figure 9). The 58 pathways also included many metabolism-related pathways, such as pyruvate metabolism, inositol phosphate metabolism, propanoate metabolism, cysteine and methionine

metabolism, fructose and mannose metabolism, riboflavin metabolism, β-alanine metabolism, and nicotinate and nicotinamide metabolism. Moreover, many pathways regulating cell growth and division were included in the list of the 58 pathways. Such pathways included MAPK signaling, Wnt signaling, calcium signaling, ErbB signaling, oocyte meiosis, and cell cycle. In addition, the 58 pathways also included many specific diseases-associated pathways, such as leishmania infection, AD, vibrio cholerae infection, epithelial cell signaling in Helicobacter pylori infection, amyotrophic lateral sclerosis, viral myocarditis, pathogenic Escherichia coli infection, arrhythmogenic right ventricular cardiomyopathy, pancreatic cancer, non-small-cell lung cancer, acute myeloid leukemia, colorectal cancer, glioma, and chronic myeloid leukemia.

4. Discussion

Although transcriptomic data have been widely applied to biomedical science, few studies have explored the association between transcriptomic perturbations and disease development and progression in a wide variety of diseases. For the
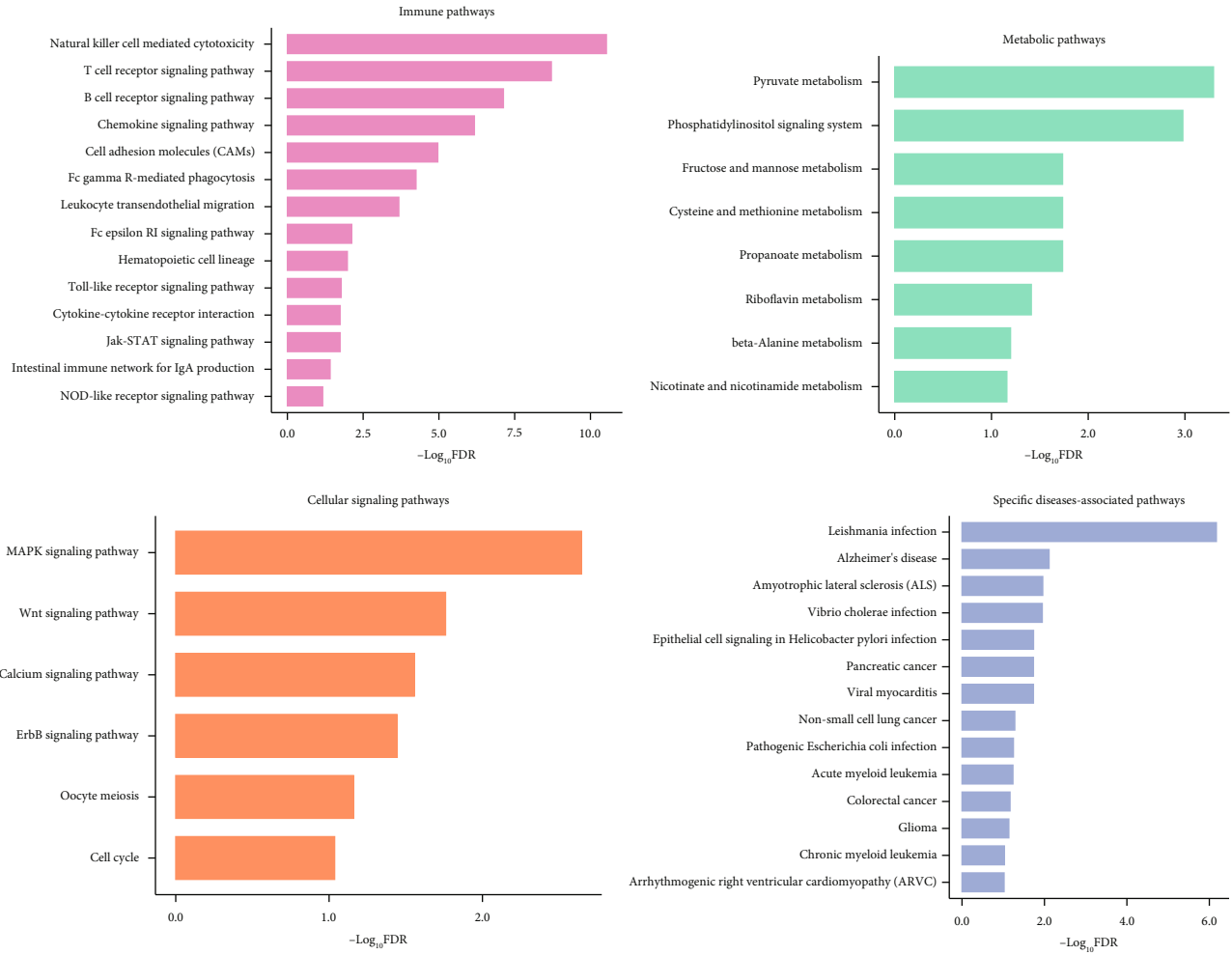
FIGURE 9: Pathways whose expression perturbations correlate positively with VTP scores across diseases.

first time, we investigated the association between the VTP and various diseases' onset and progression. Our analysis suggests that VTP values are upregulated in various diseases relative to their normal controls, and that VTP values increase with disease progression. Thus, this analysis uncovers a common characteristic of transcriptomic perturbations across various human diseases. In fact, the VTP measure reflects the asynchronous degree of transcriptomic perturbations in a disease status relative to the health status. Our results indicate that the asynchronous degree of transcriptomic perturbations is positively associated with disease progression or severity. That is, the higher asynchronous degree of transcriptomic perturbations suggests more unfavorable clinical outcomes in disease. This is consistent with the findings in cancer [1]. An intriguing question is whether the variation of perturbations in other molecules, such as genome, proteome, and metabolome, has similar associations with disease development and progression.

We identified numerous genes and pathways whose expression perturbations correlated positively with VTP scores across diseases. These genes and pathways are mainly involved in the regulation of immune, metabolic, and cellular

activities. It is justified since deregulated immune, metabolic, and cellular activities have been associated with various diseases. Our data suggest that the disordered perturbations of the molecules modulating immune, metabolic, or cellular activities are associated with the development and progression of various diseases. Interestingly, by searching for the database of publicly available GWAS summary statistics (https://www.ebi.ac.uk/gwas/), we found that many of the 369 genes, which displayed significant expression perturbations' correlations with VTP values across diseases, had genetic variants that are statistically associated with the risk of the diseases we analyzed (Supplementary Table S3). For example, there were 16 genes, including RDX, PIP4K2A, PILRA, LPXN, LILRB2, ITGAX, IQGAP2, FOXN2, CR1, CELF2, CDC42SE2, CD2AP, PDK4, PARP8, HSPA6, and BNIP3, whose genetic variants are statistically associated with the risk of AD. Six genes (TKT, TCF4, SWAP70, DDHD2, ARHGAP31, and LTB) showed significant associations of genetic variants with the risk of cardiovascular disease. Notably, FOXN2 had genetic variants statistically associated with the risk of both AD and SCZ, and NOTCH2 displayed genetic variants that are

statistically associated with the risk of both endocrine disease and kidney disease. These data support the relevance of many of these genes with the diseases.

This study has several limitations. First, although we have analyzed numerous datasets for various diseases, more datasets are needed to be analyzed to bolster the validity of this analysis. Second, the mechanism underlying the association between VTP and disease development and progression needs to be explored. Finally, the prospect of translating the present findings into clinical practice remains unclear. Nevertheless, our further study is to implement further investigations to overcome these limitations.

5. Conclusions

VTP is upregulated in the disease relative to health status, and its upregulation is associated with disease progression and severity in various diseases. The molecules whose abundance perturbations correlate positively with VTP are mainly involved in the regulation of immune, metabolic, and cellular activities. Thus, VTP has potential clinical values in disease diagnosis and prognosis.

Data Availability

All data associated with this study are available within the paper and the database of NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/).

Ethical Approval

Ethical approval and consent to participate were waived since we used only publicly available data and materials in this study.

Consent

Not applicable.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

ZD performed data analyses, visualization, and manuscript editing. QY performed data analyses, visualization, and manuscript editing. XW conceived this study, designed analysis strategies, and wrote the manuscript. All the authors read and approved the final manuscript. Zehua Dong and Qiyu Yan contributed equally to this work.

Acknowledgments

Supplementary Materials

Supplementary 1. Table S1: a description of the normalization methods for the datasets analyzed.

Supplementary 2. Table S2: 369 genes having significant positive correlations of expression perturbations with VTP values across diseases.

Supplementary 3. Table S3: the genes displayed significant expression perturbations' correlations with VTP values across diseases and had genetic variants statistically associated with a risk of the diseases.

References

[1] M. Li, Z. Zhang, L. Li, and X. Wang, "An algorithm to quantify intratumor heterogeneity based on alterations of gene expression profiles," *Communications biology*, vol. 3, no. 1, p. 505, 2020.

[2] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society B*, vol. 57, no. 1, pp. 289–300, 1995.

[3] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.

[4] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Research*, vol. 45, no. D1, pp. D353–D361, 2017.

[5] J. A. Soria Lopez, H. M. Gonzalez, and G. C. Leger, "Alzheimer's disease," *Handbook of Clinical Neurology*, vol. 167, pp. 231–255, 2019.

[6] S. Sood, I. J. Gallagher, K. Lunnon et al., "A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status," *Genome Biology*, vol. 16, no. 1, p. 185, 2015.

[7] H. Patel, A. K. Hodges, C. Curtis et al., "Transcriptomic analysis of probable asymptomatic and symptomatic alzheimer brains," *Brain, Behavior, and Immunity*, vol. 80, pp. 644–656, 2019.

[8] M. Wang, P. Roussos, A. McKenzie et al., "Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease," *Genome Medicine*, vol. 8, no. 1, p. 104, 2016.

[9] K. Barthelson, S. M. Pederson, M. Newman, H. Jiang, and M. Lardelli, "In-frame and frameshift mutations in zebrafish Presenilin 2 affect different cellular functions in young adult brains," *Journal of Alzheimer's Disease Reports*, vol. 5, no. 1, pp. 395–404, 2021.

[10] J. van de Leemput, J. L. Hess, S. J. Glatt, and M. T. Tsuang, "Genetics of schizophrenia: historical insights and prevailing evidence," *Advances in Genetics*, vol. 96, pp. 99–141, 2016.

[11] S. de Jong, M. P. M. Boks, T. F. Fuller et al., "A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes," *PLoS One*, vol. 7, no. 6, article e39498, 2012.

[12] D. Arion, Z. Huo, J. F. Enwright, J. P. Corradi, G. Tseng, and D. A. Lewis, "Transcriptome alterations in prefrontal

pyramidal cells distinguish schizophrenia from bipolar and major depressive disorders," *Biological Psychiatry*, vol. 82, no. 8, pp. 594–600, 2017.

[13] D. Datta, J. F. Enwright, D. Arion et al., "Mapping phosphodiesterase 4D (PDE4D) in macaque dorsolateral prefrontal cortex: postsynaptic compartmentalization in layer III pyramidal cell circuits," *Frontiers in Neuroanatomy*, vol. 14, article 578483, 2020.

[14] D. Arion, J. P. Corradi, S. Tang et al., "Distinctive transcriptome alterations of prefrontal pyramidal neurons in schizophrenia and schizoaffective disorder," *Molecular Psychiatry*, vol. 20, no. 11, pp. 1397–1405, 2015.

[15] *COVID-19 Dashboard*, Johns Hopkins University, 2022, https://coronavirus.jhu.edu/map.html.

[16] N. A. P. Lieberman, V. Peddu, H. Xie et al., "In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age," *PLoS Biology*, vol. 18, no. 9, article e3000849, 2020.

[17] K. A. Overmyer, E. Shishkova, I. J. Miller et al., "Large-scale multi-omic analysis of COVID-19 severity," *Cell Systems*, vol. 12, no. 1, pp. 23–40.e7, 2021.

[18] M. T. McClain, F. J. Constantine, R. Henao et al., "Dysregulated transcriptional responses to SARS-CoV-2 in the periphery," *Nature Communications*, vol. 12, no. 1, p. 1079, 2021.

[19] A. Soares-Schanoski, N. Sauerwald, C. W. Goforth et al., "Asymptomatic SARS-CoV-2 infection is associated with higher levels of serum IL-17C, matrix metalloproteinase 10 and fibroblast growth factors than mild symptomatic COVID-19," *Frontiers in Immunology*, vol. 13, article 821730, 2022.

[20] A. C. Collier and H. H. Handsfield, "Acquired immune deficiency syndrome," *The Urologic Clinics of North America*, vol. 11, no. 1, pp. 187–197, 1984.

[21] M. Rotger, K. K. Dang, J. Fellay et al., "Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals," *PLoS Pathogens*, vol. 6, no. 2, article e1000781, 2010.

[22] F. Z. Chowdhury, Z. Ouyang, M. Buzon, B. D. Walker, M. Lichterfeld, and X. G. Yu, "Metabolic pathway activation distinguishes transcriptional signatures of CD8+ T cells from HIV-1 elite controllers," *AIDS*, vol. 32, no. 18, pp. 2669–2677, 2018.

[23] A. Quach, S. Horvath, N. Nemanim et al., "No reliable gene expression biomarkers of current or impending neurocognitive impairment in peripheral blood monocytes of persons living with HIV," *Journal of Neurovirology*, vol. 24, no. 3, pp. 350–361, 2018.

[24] M. Shahmanesh, K. Phillips, M. Boothby, and J. W. Tomlinson, "Differential adipose tissue gene expression profiles in abacavir treated patients that may contribute to the understanding of cardiovascular risk: a microarray study," *PLoS One*, vol. 10, no. 1, article e0117164, 2015.

[25] C. Shih, C. C. Yang, G. Choijilsuren, C. H. Chang, and A. T. Liou, "Hepatitis B Virus," *Trends in Microbiology*, vol. 26, no. 4, pp. 386-387, 2018.

[26] W. Zhou, Y. Ma, J. Zhang et al., "Predictive model for inflammation grades of chronic hepatitis B: large-scale analysis of clinical parameters and gene expressions," *Liver International*, vol. 37, no. 11, pp. 1632–1641, 2017.

[27] Y. Lu, Z. Fang, M. Li et al., "Dynamic edge-based biomarker non-invasively predicts hepatocellular carcinoma with hepatitis B virus infection for individual patients based on blood testing," *Journal of Molecular Cell Biology*, vol. 11, no. 8, pp. 665–677, 2019.

[28] S. M. Wang, L. L. Ooi, and K. M. Hui, "Identification and validation of a novel gene signature associated with the recurrence of human hepatocellular carcinoma," *Clinical Cancer Research*, vol. 13, no. 21, pp. 6275–6283, 2007.

[29] D. Schlossberg, "Acute tuberculosis," *Infectious Disease Clinics of North America*, vol. 24, no. 1, pp. 139–146, 2010.

[30] J. Maertzdorf, M. Ota, D. Repsilber et al., "Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis," *PLoS One*, vol. 6, no. 10, article e26938, 2011.

[31] C. Broderick, J. M. Cliff, J. S. Lee, M. Kaforou, and D. A. J. Moore, "Host transcriptional response to TB preventive therapy differentiates two sub- groups of IGRA-positive individuals," *Tuberculosis (Edinburgh, Scotland)*, vol. 127, article 102033, 2021.

[32] J. G. Burel, A. Singhania, P. Dubelko et al., "Distinct blood transcriptomic signature of treatment in latent tuberculosis infected individuals at risk of developing active disease," *Tuberculosis (Edinburgh, Scotland)*, vol. 131, article 102127, 2021.

[33] L. S. Garcia, "Malaria," *Clinics in Laboratory Medicine*, vol. 30, no. 1, pp. 93–129, 2010.

[34] A. B. W. Boldt, H. van Tong, M. P. Grobusch et al., "The blood transcriptome of childhood malaria," *eBioMedicine*, vol. 40, pp. 614–625, 2019.

[35] C. F. Ockenhouse, W. C. Hu, K. E. Kester et al., "Common and divergent immune response signaling pathways discovered in peripheral blood mononuclear cell gene expression patterns in presymptomatic and clinically apparent malaria," *Infection and Immunity*, vol. 74, no. 10, pp. 5561–5573, 2006.

[36] Y. Idaghdour, J. Quinlan, J. P. Goulet et al., "Evidence for additive and interaction effects of host genotype and infection in malaria," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 42, pp. 16786–16793, 2012.

[37] Y. Kokubo and C. Matsumoto, "Hypertension is a risk factor for several types of heart disease: review of prospective studies," *Advances in Experimental Medicine and Biology*, vol. 956, pp. 419–426, 2017.

[38] S. Hannenhalli, M. E. Putt, J. M. Gilmore et al., "Transcriptional genomics associates FOX transcription factors with human heart failure," *Circulation*, vol. 114, no. 12, pp. 1269–1276, 2006.

[39] S. Ameling, L. R. Herda, E. Hammer et al., "Myocardial gene expression profiles and cardiodepressant autoantibodies predict response of patients with dilated cardiomyopathy to immunoadsorption therapy," *European Heart Journal*, vol. 34, no. 9, pp. 666–675, 2013.

[40] R. Suresh, X. Li, A. Chiriac et al., "Transcriptome from circulating cells suggests dysregulated pathways associated with long-term recurrent events following first-time myocardial infarction," *Journal of Molecular and Cellular Cardiology*, vol. 74, pp. 13–21, 2014.

[41] E. D. Muse, E. R. Kramer, H. Wang et al., "A whole blood molecular signature for acute myocardial infarction," *Scientific Reports*, vol. 7, no. 1, p. 12268, 2017.

[42] G. Gobbi, C. Carubbi, G. M. Tagliazucchi et al., "Sighting acute myocardial infarction through platelet gene expression," *Scientific Reports*, vol. 9, no. 1, p. 19574, 2019.

[43] E. Witt, E. Hammer, M. Dörr et al., "Correlation of gene expression and clinical parameters identifies a set of genes reflecting LV systolic dysfunction and morphological alterations," *Physiological Genomics*, vol. 51, no. 8, pp. 356–367, 2019.

[44] M. Kiliszek, B. Burzynska, M. Michalak et al., "Altered gene expression pattern in peripheral blood mononuclear cells in patients with acute myocardial infarction," *PLoS One*, vol. 7, no. 11, article e50054, 2012.

[45] C. Cheadle, A. E. Berger, S. C. Mathai et al., "Erythroid-specific transcriptional changes in PBMCs from pulmonary hypertension patients," *PLoS One*, vol. 7, no. 4, article e34951, 2012.

[46] F. García-Río, B. Alcázar-Navarrete, D. Castillo-Villegas et al., "Biological biomarkers in respiratory diseases," *Archivos de Bronconeumología*, vol. 58, no. 4, pp. 323–333, 2022.

[47] T. M. Bahr, G. J. Hughes, M. Armstrong et al., "Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease," *American Journal of Respiratory Cell and Molecular Biology*, vol. 49, no. 2, pp. 316–323, 2013.

[48] R. Faner, N. Gonzalez, T. Cruz, S. G. Kalko, and A. Agustí, "Systemic inflammatory response to smoking in chronic obstructive pulmonary disease: evidence of a gender effect," *PLoS One*, vol. 9, no. 5, article e97491, 2014.

[49] R. C. Langan and A. J. Goodbred, "Office spirometry: indications and interpretation," *American Family Physician*, vol. 101, no. 6, pp. 362–368, 2020.

[50] R. Sellamuthu, C. Umbright, J. R. Roberts et al., "Molecular insights into the progression of crystalline silica-induced pulmonary toxicity in rats," *Journal of Applied Toxicology*, vol. 33, no. 4, pp. 301–312, 2013.

[51] V. R. Mas, D. G. Maluf, K. J. Archer et al., "Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma," *Molecular Medicine*, vol. 15, no. 3-4, pp. 85–94, 2009.

[52] S. P. Fletcher, D. J. Chin, Y. Ji et al., "Transcriptomic analysis of the woodchuck model of chronic hepatitis B," *Hepatology*, vol. 56, no. 3, pp. 820–830, 2012.

[53] A. Scherer, O. P. Günther, R. F. Balshaw et al., "Alteration of human blood cell transcriptome in uremia," *BMC Medical Genomics*, vol. 6, no. 1, p. 23, 2013.

[54] P. C. Grayson, S. Eddy, J. N. Taroni et al., "Metabolic pathways and immunometabolism in rare kidney diseases," *Annals of the Rheumatic Diseases*, vol. 77, no. 8, pp. 1226–1233, 2018.

[55] L. H. Mariani, S. Martini, L. Barisoni et al., "Interstitial fibrosis scored on whole-slide digital imaging of kidney biopsies is a predictor of outcome in proteinuric glomerulopathies," *Nephrology, Dialysis, Transplantation*, vol. 33, no. 2, pp. 310–318, 2018.

[56] I. Arijs, G. de Hertogh, K. Lemaire et al., "Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment," *PLoS One*, vol. 4, no. 11, article e7984, 2009.

[57] I. Nookaew, K. Thorell, K. Worah et al., "Transcriptome signatures in helicobacter pylori-infected mucosa identifies acidic mammalian chitinase loss as a corpus atrophy marker," *BMC Medical Genomics*, vol. 6, no. 1, p. 41, 2013.

[58] J. Refardt, "Diagnosis and differential diagnosis of diabetes insipidus: update," *Best Practice & Research Clinical Endocrinology & Metabolism*, vol. 34, no. 5, article 101398, 2020.

[59] E. C. Kaizer, C. L. Glaser, D. Chaussabel, J. Banchereau, V. Pascual, and P. C. White, "Gene expression in peripheral blood mononuclear cells from children with diabetes," *The Journal of Clinical Endocrinology and Metabolism*, vol. 92, no. 9, pp. 3705–3711, 2007.

[60] F. H. van Tienen, S. F. E. Praet, H. M. de Feyter et al., "Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes," *The Journal of Clinical Endocrinology and Metabolism*, vol. 97, no. 9, pp. 3261–3269, 2012.

[61] S. Gao, N. Wolanyk, Y. Chen, S. Jia, M. J. Hessner, and X. Wang, "Investigation of coordination and order in transcription regulation of innate and adaptive immunity genes in type 1 diabetes," *BMC Medical Genomics*, vol. 10, no. 1, p. 7, 2017.