Reports

BioTechniques

A clinically validated human capillary blood transcriptome test for global systems biology studies

Ryan Toma[1], Nathan Duval[1], Benjamin Pelle[2], Matthew M Parks[1], Vishakh Gopu[1], Pedro J Torres[1], Francine R Camacho[1], Nan Shen[1], Subha Krishnan[1], Andrew Hatch[1], Hal Tily[1], Ally Perlina[1], Guruduth Banavar[1] & Momchilo Vuyisich*,[1]

*[1]Viome, Inc., Bellevue, WA/Los Alamos, NM/New York, NY/San Diego, CA, USA; [2]The University of Washington, Department of Genome Sciences, Seattle, CA, USA; *Author for correspondence: momo@viome.com*

BioTechniques 69: 00–00 (April 2020) 10.2144/btn-2020-0088

First draft submitted: 16 June 2020; Accepted for publication: 21 July 2020; Published online: 7 August 2020

ABSTRACT

To prevent and treat chronic diseases, including cancer, a global application of systems biology is needed. We report here a whole blood transcriptome test that needs only 50 µl of capillary (fingerprick) blood. This test is suitable for global applications because the samples are preserved at ambient temperature for up to 4 weeks and the RNA preservative inactivates all pathogens, enabling safe transportation. Both the laboratory and bioinformatic steps are automated and performed in a clinical lab, which minimizes batch effects and creates unbiased datasets. Given its clinical testing performance and accessibility to traditionally underrepresented and diverse populations, this test offers a unique ability to reveal molecular mechanisms of disease and enable longitudinal, population-scale studies.

METHOD SUMMARY

This report describes an important improvement to whole blood transcriptome analysis. While current methods are complicated and expensive, the method reported here includes at-home sample collection (from fingerprick), global shipping at ambient temperatures and pathogen inactivation at the point of collection; it uses fully automated, clinically validated and licensed laboratory and bioinformatic analyses.

KEYWORDS:

clinically validated ● population studies ● RNA-seq ● systems biology ● transcriptomics ● whole blood

Quantitative gene expression analysis provides a global snapshot of tissue function. The human transcriptome varies with tissue type, developmental stage, environmental stimuli and health/disease state [1–5]. For example, gene expression changes have been observed in venous blood of patients with ulcerative colitis and Crohn's disease [6], breast cancer [7], chronic fatigue syndrome [8] and asthma [9]. Changes in expression patterns can provide insights into the molecular mechanisms of disease onset and progression. In the era of precision medicine, methods to noninvasively assess an individual's transcriptional changes can reveal predictive markers of disease, inform the best choice of therapy and enable the development of novel therapies. For example, using systems biology approaches to interrogate the transcriptome is now a standard strategy to stratify cancer patients and select the best available therapies [10–12]. Methods to analyze the transcriptome have become a vital part of trying to better understand chronic diseases with unknown etiologies.

In the USA chronic noncommunicable diseases have become an epidemic with an ever-growing share of healthcare expenditures [13]. Many chronic diseases go through cycles of elevated disease activity (relapse) and periods of low disease activity (remission) [14–18]. The human genome has provided valuable information on disease predisposition and susceptibility, but the overall genetic contribution to many chronic diseases is weak [19]. Furthermore, many environmental or lifestyle factors increase the risk of developing a chronic disease, implicating a change in gene expression. This has been shown in large twin and epidemiological studies [20,21]. Changes in gene expression have been shown in Type I diabetes [22], Type II diabetes [23,24] and cardiovascular disease [25,26], often associated with immune regulation [27]. Unfortunately, little progress has been made in understanding the role of the transcriptome in chronic disease due to the lack of a low-cost and scalable method for which samples can be collected by anyone and anywhere. Instead, available methods rely on complex logistics of collecting peripheral (venous) blood at a clinic and transporting it frozen to a laboratory. In addition, there are very few clinically validated blood transcriptome tests available, so analyses of data from disparate methods can lead to erroneous conclusions due to biases introduced in each method.

Circulating blood is easily accessible and provides a noninvasive alternative to tissue biopsies for molecular profiling of human disease and disease risk [28]. Peripheral blood mononuclear cells have been the primary choice for transcriptomic investigations from whole

- 50 µl of capillary (fingerprick) blood
- Ambient temperature preservation
- RNA extraction and DNA removal
- PolyA selection
- Directional library preparation
- 2 × 150 Illumina sequencing
- Gene expression analysis

Figure 1. Primary components of the at-home capillary blood collection kit and outline of the sample analysis method. The lancet is used to prick the finger; the minivette is a capillary device with a piston; the sample collection tube is a microcentrifuge tube with 200 µl of RPB.

blood samples [29]. Unfortunately, methods available for scrutinizing the transcriptome from these cells are useful for research methods, but not ideal for broader applications at scale. Furthermore, methods using white blood cells fail to measure cell-free transcripts, which have been associated with certain disease conditions, particularly for cancer patients and pregnant women [30–33].

The current state of the art focuses on the whole blood transcriptome, typically requiring a venous blood draw. The availability of high-throughput sequencing technologies has enabled rapid analysis of the transcriptome from small sample volumes. Development of such methods has been used in the analysis of longevity-associated transcriptomic signatures in bats [34]. These studies often require weight-dependent blood volumes and temperature controls, making them difficult to scale. To our knowledge there is currently no blood test that is easily scalable for population-based studies of the human whole blood transcriptome. Here we describe, for the first time, a high-throughput method for interrogating the human transcriptome from a small fingerprick blood sample. The method is automated, inexpensive and clinically validated. An RNA preservation buffer mixed with the blood sample at the point of collection inactivates all pathogens (bacteria, fungi and viruses), enabling safe transportation globally at ambient temperatures for up to 28 days, while preserving RNA integrity for gene expression analysis.

Materials & methods

Ethics statement

All procedures involving human participants were performed in accordance with the ethical standards and approved by a federally accredited Institutional Review Board committee. Informed consent was obtained from all participants. A total of 508 unique participants were recruited into this study. Some participants donated blood for multiple experiments. The study consisted of a heterogeneous population of 174 males and 334 females, with a mean age of 46.1 years (females 46.0 years ± 14 SD, males 46.3 years ± 15 SD), age range of 18–93 years (females 18–93, males 19–93) (Figure 2). Health data were not collected from participants.

Sample collection

Capillary blood was collected by study participants using an at-home kit, following the included instructions (Figure 1). The kit contains a 1.5-ml screw cap tube (Sarstedt) containing 200 µl of RNA preservative buffer (RPB), a 50-µl EDTA-coated minivette (Sarstedt), a 70% isopropyl alcohol wipe (Dynarex), a gauze pad, a push-button 17-gauge disposable lancet (Acti-Lance) and an instruction sheet. Participants were instructed to fast for 8 h prior to collection and collect the samples within 1 h of waking up, without exercising or drinking (except water). The finger was wiped with a 70% isopropyl alcohol wipe and allowed to dry. Using the disposable lancet, the finger was punctured and the first drop wiped away using the gauze pad. The blood was drawn by capillary action into a 50-µl EDTA-

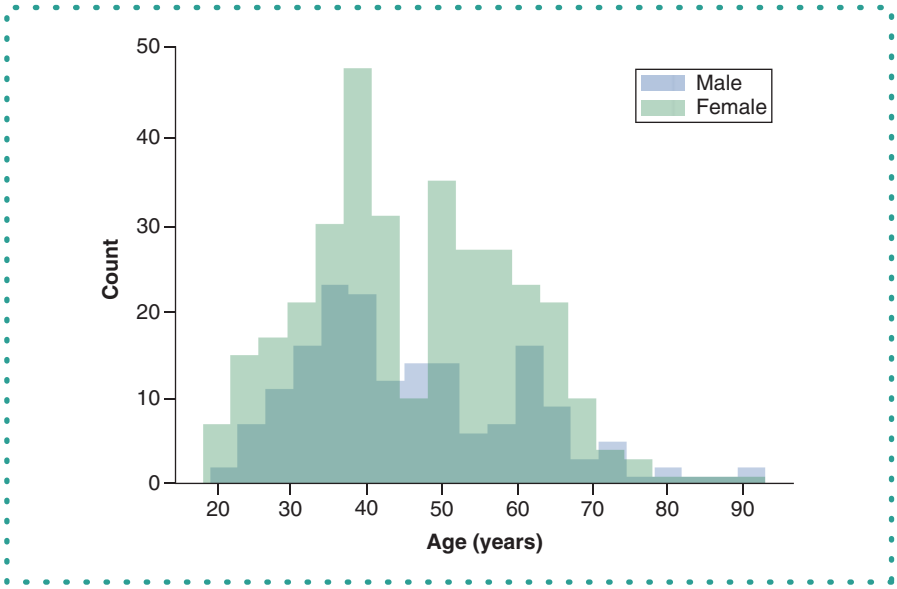

Figure 2. Age range and distribution of participants by gender.

coated minivette and dispensed into the sample collection tube that contained 200 µl of RPB. The tube was vigorously shaken for 15 s to thoroughly mix the RPB with the blood. The puncture was wiped clean and a band-aid was applied.

Venous blood was collected by a phlebotomist. Participants were instructed to fast for 8 h prior to collection and to collect the samples within 1 h of waking up, without exercising or drinking (except water). The puncture site was first cleaned with a 70% isopropyl alcohol wipe (Dynarex) and allowed to dry. Blood was then collected into 4-ml EDTA vacutainers (BD Bioscience). The blood was promptly transferred from the vacutainers to another tube and mixed with RPB at a 1:4 ratio.

Samples used to compare gene expression in venous and capillary blood were collected using the methods described above. For six participants, venous and capillary blood samples were collected within minutes of each other.

Longitudinal samples were collected weekly from ten participants over a 5-week period using the capillary blood collection method described above. Four technical replicates were collected and analyzed at each time point.

Shipping, storage and precision validation samples were collected from three participants following the capillary blood collection method described above. To validate the storage stability, samples were stored for 7, 14 and 28 days at room temperature or -80°C and compared with the technical replicates analyzed on the day of collection (day 0). A subset of replicate samples was shipped twice during storage times.

RNA extraction, library preparation & sequencing

RNA was extracted using silica beads and a series of washes, followed by elution in molecular biology-grade water. DNA was degraded using RNase-free DNase. Polyadenylated messenger RNAs (mRNAs) were isolated from the total RNA using an oligo dT-based magnetic mRNA isolation kit. mRNAs were converted to directional sequencing libraries with unique dual-barcoded adapters and ultrapure reagents. Libraries were pooled and quality controlled with dsDNA Qubit (Thermo Fisher Scientific), QuantStudio 3 qPCR (Thermo Fisher) with KAPA Library Quantification Kit (Roche) and Fragment Analyzer (Advanced Analytical). Library pools were sequenced on Illumina NextSeq or NovaSeq instruments using 300-cycle kits. Transcripts from an average of nearly 14,000 genes are detected in each sample with each sample receiving $3.42 \times 10^6 \pm 1.69 \times 10^6$ reads with a mean percentage of bases with quality score above Q30 of 87.73 ± 2.02 and a mean base quality score of 36.66 ± 0.44.

Bioinformatics

A human transcriptome catalog was created by augmenting the transcripts derived from the Ensembl release 99 [35] annotations for GRCh38 [36] with selected transcripts from RefSeq. Raw reads were mapped to this transcript catalog and quantified with *Salmon* [37]. *Salmon* was executed with parameters: 'salmon quant –libType ISF –seqBias –gcBias –discardOrphansQuasi –validateMappings'. Briefly, *Salmon* maps read against the provided transcriptome and computes per-transcript expression levels using a statistical model that accounts for mapping ambiguity, sequence-specific biases, GC-content biases and positional biases [37]. Transcript expression levels were aggregated by gene to obtain gene expression levels. In a sample, only genes with an expected read count of at least 1 were considered as expressed. Transcripts per million for genes were calculated using the expected read counts and effective read lengths as generated by the statistical model in *Salmon*.

Data analysis

Statistical parameters, including transformations and significance, are reported in the figures and figure legends. To compare pairs of samples, we report Jaccard similarity (which ignores expression and considers overlap in genes detected), Spearman correlations (which are invariant to absolute expression levels of the genes and only consider the similarity of ranked expression) and Hellinger distance (an appropriate distance measure for compositional data). Statistical analysis was performed in Python. Most variable genes present in all samples were obtained from a total of 508 unique individuals' samples by satisfying the following criteria: gene transcripts with 100% prevalence and coefficient of variability (CV) greater than 1 across all samples.

Results & discussion

Correlation of gene expression levels in capillary & venous blood

Venous blood is the current gold-standard sample for whole blood transcriptome analysis. We compared the gene expression levels obtained from the venous and capillary blood of six study participants. The correlation between gene expression in capillary and venous blood from these participants showed excellent concordance, with Spearman correlation coefficient above 0.94 for all donors (Figure 3A). Principal component analysis was performed to visualize the relatedness between the sample types (venous and capillary) as well as between the participants' samples. When assessing the first and second principal component, there is only weak grouping by either sample type or participant (Figure 3B). However, the first and third principal components reveal clustering of participant samples, though not of sample types (Figure 3C).

Method precision

For clinical studies in a large population and over time that measure gene expression changes as a function of health and disease states, method precision is a critical parameter. To determine the precision of the method, we measured the percent CV for gene expression in four technical replicates from three participants (Figure 4). This was plotted as a function of expression level (mean counts per million, CPM); as expected, the higher the expression level, the lower the variance. The distribution of transcript relative abundance is shown from 10 CPM to greater than 10^5 CPM. A small cluster of transcripts with expression levels greater than 10^5 CPM are highly abundant globins. A CV below 25% for the majority of transcripts shows low variance for this method.

Sample stability

One concern when performing transcriptomic analysis of blood samples is the facile degradation of RNA, due to self-cleavage and the action of ribonucleases [33]. To preserve RNA, our method uses an RPB with multiple functions. The RPB's detergents immediately dissolve lipid bilayers and inactivate all pathogens, its denaturants halt all enzymatic activity by disrupting the tertiary structures of proteins, and special buffering agents prevent RNA self-cleavage by keeping the 2′ oxygen protonated, thus preserving RNA integrity during sample shipping and storage. To validate the efficacy of RPB to preserve the blood transcriptome at ambient temperatures, blood samples were collected from three participants and stored at ambient temperature (22°C) or -80°C for 0, 7, 14 and 28 days, with and without shipping. Gene expression levels of four technical replicates per condition were analyzed. The correlation coefficient between gene expression levels from blood samples in all conditions was very high, with Spearman correlation coefficients above 0.81 for all conditions tested [38]. Most importantly, the correlation coefficient between 28 days at room temperature and samples analyzed at time = 0 is above 0.80, showing ability of the RPB to adequately preserve RNA. Among all comparisons, the lowest Spearman correlation for donor 1 was 0.849 (Figure 5), the lowest Spearman correlation for donor 2 was 0.865 (Supplementary Figure 1) and the lowest Spearman correlation for donor 3 was 0.811 (Supplementary Figure 1). This shows that our method can adequately preserve RNA for up to 28 days at ambient temperature. Furthermore, the slight variations in correlation coefficients are due to minor sample degradation observed after 28 days; this is random and not specific to any transcripts or pathways.

Longitudinal changes in the blood transcriptome

In this small study, ten participants collected capillary blood weekly for 5 weeks. On average, transcripts from 13,796 genes were detected per sample. For the genes detected, the correlation between gene expression remained high across 5 weeks for each participant. Jaccard similarity, which measures the overlap in genes detected between all samples, was stable at around 50%. This suggests that around half the genes detected in any sample may not be detected in another sample from the same donor. Our standard method includes 5×10^6 paired reads, with the expression level of many genes falling near the limit of detection. In any sample, the appearance of genes with low expression is a stochastic event; by increasing the limit of detection, we could obtain a more favorable number. The genes that are detected consistently are expressed at a relatively stable level (Figure 6).

One of the more important parameters of the blood test is its precision to distinguish very small changes among thousands of measured features. Such high precision would enable the identification of transcriptome changes, both over time and due to changes in health and disease. To assess the test precision in this context, we compared Hellinger distances among technical replicates from samples obtained on the same day from the same person (intraperson technical distances in Figure 7); biological samples from the same person collected over time (intraperson distances in Figure 7); and biological samples obtained from different people (interperson distances in Figure 7). An empirical cumulative distribution function plot of Hellinger distances shows that when taking all genes into

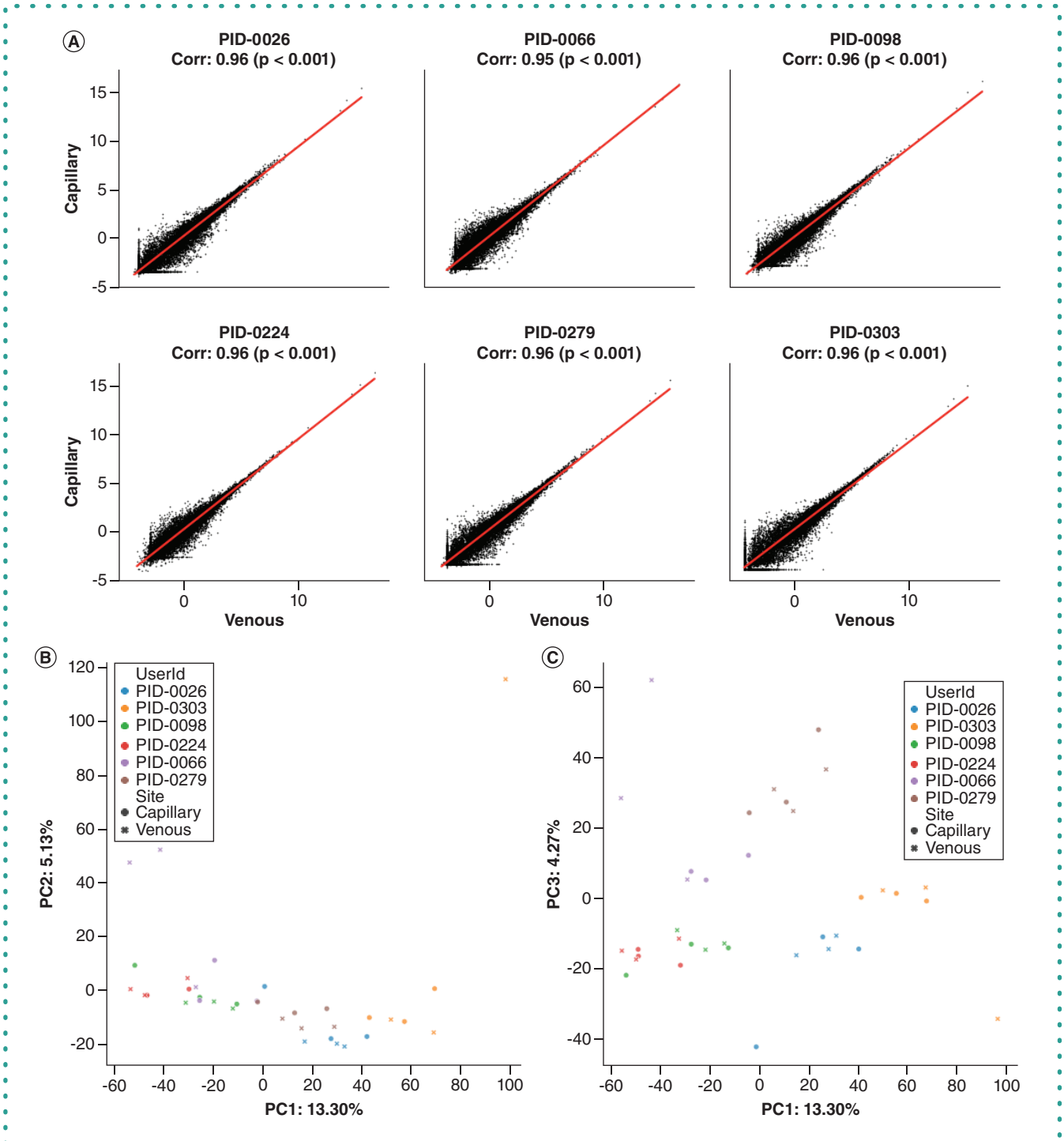

Figure 3. Correlation between capillary and venous blood sample collections. **(A)** Correlation between venous and capillary blood gene expression levels from six donors (PID-####) shows high concordance. Spearman correlation coefficients are above 0.94 for all instances with p-values <0.001. **(B)** Principal component analysis plot of the first two principal components (PC) showing now obvious grouping patterns by sample type (venous or capillary) or by participant. **(C)** Principal component analysis plot of the first and third PC showing a pattern by participant, but no discernible pattern by sample type.

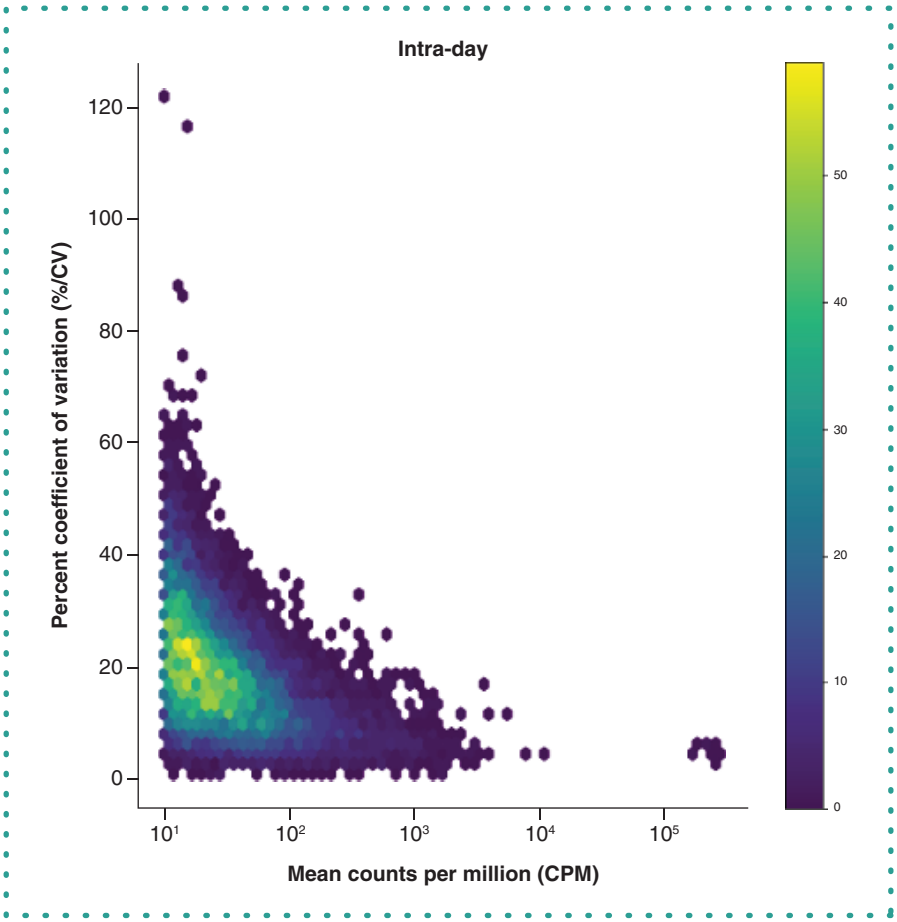

Figure 4. Method precision, shown as the percent coefficient of variation (%CV) as a function of transcript abundance (mean CPM). These data show a broad distribution of transcript abundances, from 10 CPM to $>10^5$ CPM.

account, pairs of samples taken from an individual within a week are more similar to samples from that same individual taken in two different weeks. However, pairs of samples coming from two different participants tend to be less similar than samples taken from a single participant (Figure 7). Furthermore, the Hellinger distances show that although the samples are less similar, a substantial number of genes are stably expressed between individuals.

Most variable genes present in all samples

Using the test described, we analyzed capillary blood from a total of 508 unique individuals. Of all the genes whose transcripts were detected in 100% of the samples, those listed in Table 1 had the highest variability in expression across all samples. The ability to detect differences in expression among not only some of the seldom-detected genes, but also the most commonly expressed ones, allows for powerful and precise derivation of functional insights specific to an individual's biology.

Chronic noncommunicable diseases are the leading cause of death globally. There is strong evidence that human health and chronic diseases, including aging, are heavily influenced by nutrition and microbiome functions [39–42]. To understand the root causes of chronic diseases, we need to study the human body as an ecosystem, which requires a systems biology approach. Gene expression levels, for both the human and their microbiome, play a very important role in the onset and progression of chronic disease, and need to be studied in the context of an integrated systems biology platform. While a scalable and clinically validated gut microbiome gene expression test has been developed [43], a human blood transcriptome test with the same features and performance is also needed.

Here we describe a novel whole blood transcriptome test that is clinically validated and globally scalable. The blood sample (50 μl) can be collected anywhere by almost anyone, as it uses blood from a fingerprick. The samples can be stored and shipped without temperature controls to a reference laboratory. The method is automated, inexpensive and high throughput. The RPB inactivates pathogens; this is very important for international shipping because it prevents disease spread. The RPB preserves RNA for up to 28 days at ambient temperature (Figure 5), obviating the need for cold storage and shipping, which are expensive and not easily available.

Because the majority of published blood transcriptome studies rely on venous blood, we compared the transcriptome data obtained from venous and capillary blood and showed high gene expression level concordance between the two sample types (Figure 3). This

Table 1. Transcripts detected in all samples with the highest variation in gene expression.

Gene symbol	Ensembl ID	Z-score range	CV	Mean	Largest	Smallest	SD
ABAT	ENSG00000183044	19.5803	0.930441	10.8358	198.052	0.641576	10.8358
ATP5PF	ENSG00000154723	18.0953	0.560321	46.5863	481.141	8.79368	46.5863
BCL2	ENSG00000171791	21.1774	1.22952	15.7492	411.523	1.44519	15.7492
BRD3	ENSG00000169925	18.5145	0.893863	23.9794	402.109	5.26354	23.9794
CCR7	ENSG00000126353	17.7814	0.798884	42.9462	614.19	4.12669	42.9462
CD79B	ENSG00000007312	21.2598	1.19739	28.7157	734.581	3.58869	28.7157
COMMD6	ENSG00000188243	20.2709	0.726057	27.8127	412.507	3.16496	27.8127
COX7C	ENSG00000127184	19.6426	0.567736	128.806	1470.71	34.294	128.806
CTSH	ENSG00000103811	18.1366	0.619335	34.5955	394.646	6.04834	34.5955
CYB561A3	ENSG00000162144	21.2764	1.0695	12.3565	282.412	1.23801	12.3565
DNAJC10	ENSG00000077232	17.6809	0.724936	22.0323	284.803	2.40358	22.0323
FAU	ENSG00000149806	19.0723	0.442189	300.647	2613.72	78.1904	300.647
FCMR	ENSG00000162894	20.5402	0.853317	93.9606	1662.97	16.0941	93.9606
FCRL3	ENSG00000160856	21.7806	1.69911	19.9217	737.957	0.703213	19.9217
GGA2	ENSG00000103365	20.7858	0.97513	26.4087	537.118	1.84522	26.4087
HINT1	ENSG00000169567	18.5222	0.568632	28.1801	299.571	2.76783	28.1801
HLA-DMA	ENSG00000204257	17.5751	0.599692	38.2286	409.273	6.35635	38.2286
IRF8	ENSG00000140968	17.7098	0.667804	31.4545	374.815	2.81285	31.4545
LRMP	ENSG00000118308	17.7863	0.581489	58.342	612.327	8.92254	58.342
MARCHF1	ENSG00000145416	18.0818	0.706504	14.8847	192.4	2.24958	14.8847
NCOA3	ENSG00000124151	17.5506	0.829248	21.9864	321.37	1.38391	21.9864
POLR2I	ENSG00000105258	18.3795	0.538726	13.0067	131.221	2.43509	13.0067
PTPN2	ENSG00000175354	17.5004	0.511314	30.43	278.896	6.6027	30.43
PTRHD1	ENSG00000184924	18.7119	0.639155	14.9057	179.742	1.47349	14.9057
RABL6	ENSG00000196642	17.6621	0.812364	19.6509	284.927	2.97393	19.6509
RPL13A	ENSG00000142541	17.6341	0.452961	1220.1	10125	379.292	1220.1
RPL17	ENSG00000265681	18.5712	0.559581	200.659	2126.52	41.2519	200.659
RPL18A	ENSG00000105640	19.4296	0.580484	651.138	7511.17	167.242	651.138
RPL23A	ENSG00000198242	18.0266	0.439411	1493.11	12274.9	447.878	1493.11
RPL34	ENSG00000109475	20.0994	0.68213	407.553	5671.45	83.7351	407.553
RPL35	ENSG00000136942	19.9857	0.550939	215.598	2418.51	44.5822	215.598
RPL36	ENSG00000130255	18.2253	0.4665	131.063	1136.26	21.9482	131.063
RPS14	ENSG00000164587	18.7577	0.482776	285.318	2629.33	45.5618	285.318
RPS15	ENSG00000115268	18.2983	0.40219	544.305	4167.41	161.657	544.305
RPS19	ENSG00000105372	21.0572	0.690536	208.792	3077.14	41.1528	208.792
RPS20	ENSG00000008988	19.424	0.632074	280.895	3505.4	56.7295	280.895
RPS21	ENSG00000171858	17.8462	0.701284	56.051	705.608	4.11528	56.051
RPS23	ENSG00000186468	18.7498	0.544539	305.423	3180.79	62.4151	305.423
RPS27	ENSG00000177954	20.568	0.736969	3215.52	49447.9	707.251	3215.52
RPS27A	ENSG00000143947	18.5657	0.584048	311.26	3416.23	41.1445	311.26
RUBCNL	ENSG00000102445	21.7772	1.59506	15.5318	540.634	1.12134	15.5318
SEL1L3	ENSG00000091490	19.7646	0.864262	18.5435	317.854	1.09883	18.5435
SMDT1	ENSG00000183172	19.2955	0.811701	11.3182	178.336	1.06729	11.3182
SNRPC	ENSG00000124562	17.6885	0.552639	24.5824	244.016	3.71402	24.5824
ST6GAL1	ENSG00000073849	18.576	0.666383	36.9283	458.638	1.51209	36.9283
TLK1	ENSG00000198586	17.5502	0.631585	17.0969	190.431	0.92261	17.0969
TOMM5	ENSG00000175768	19.3365	0.650756	23.6457	300.415	2.87196	23.6457
TTN	ENSG00000155657	22.4609	3.39939	6.83862	522.35	0.200298	6.83862
UQCRQ	ENSG00000164405	18.4933	0.556532	38.5469	400.271	3.54137	38.5469
VOPP1	ENSG00000154978	19.0044	0.668412	23.6556	303.368	2.87627	23.6556
XPO7	ENSG00000130227	17.6957	0.668843	19.3292	231.02	2.24524	19.3292
ZNF266	ENSG00000174652	20.3151	1.01147	28.5009	589.015	3.37211	28.5009

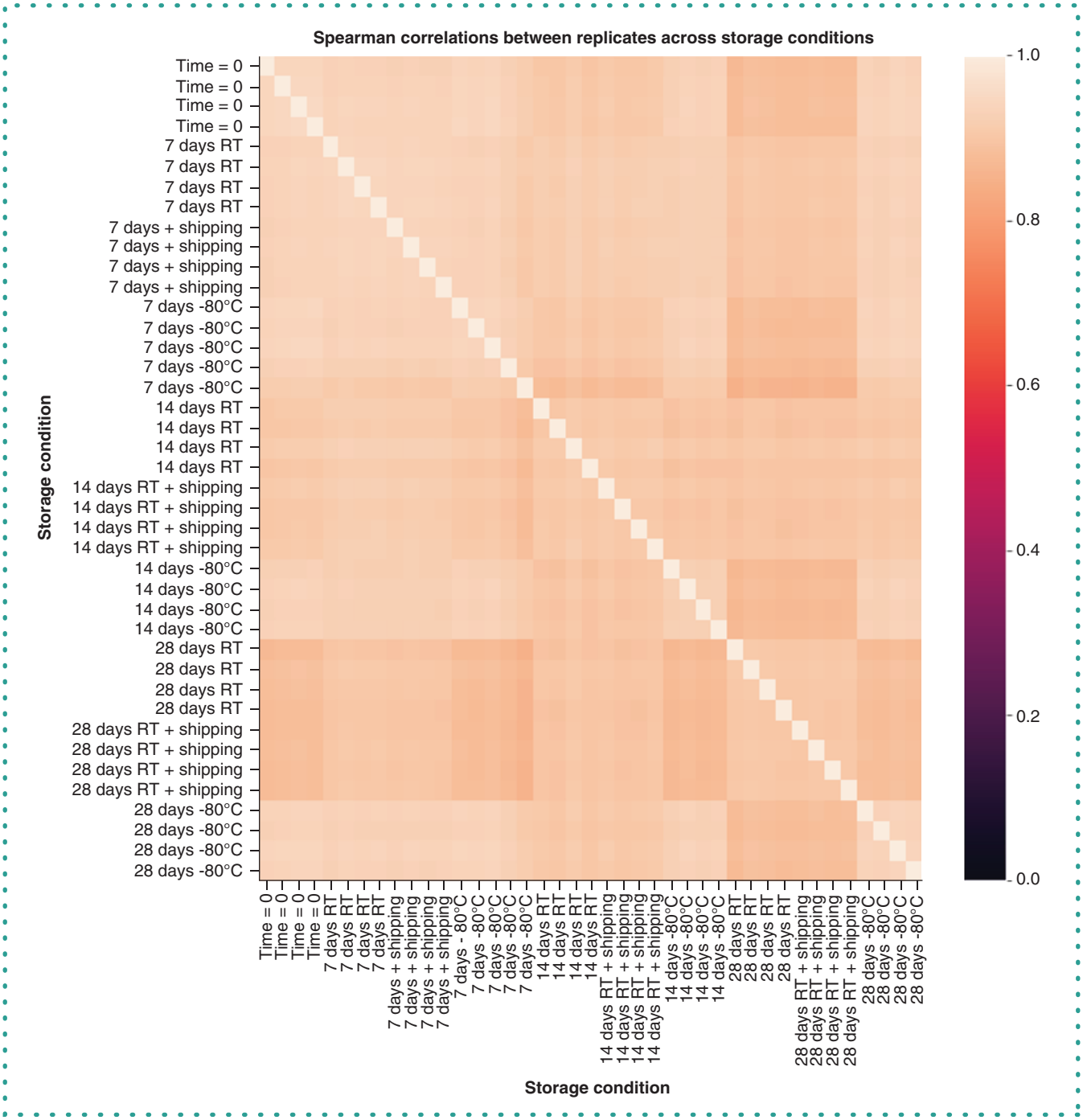

Figure 5. Heatmap showing Spearman correlation coefficients for sample stability of blood samples stored at ambient temperature and -80°C for 0, 7, 14 and 28 days, with and without shipping conditions. The Spearman correlation coefficients for all conditions in all donor samples were >0.81.

is important in case the results obtained from one sample type need to be compared with those from the other. We also show that the precision of the test is sufficient to distinguish intraperson technical replicates and longitudinal transcriptome data from other people's blood transcriptomes (Figures 6 & 7).

Gene expression patterns in individuals serve as molecular data input, which informs pathway analysis that yields numerical assessments of activity levels of specific signaling or metabolic pathway mechanisms relevant to health and disease. Taken integratively with gut microbial gene expression, complex functional profiles of cross-omic, cross-organism molecular pathway interactions offer actionable health insights, such as unique inflammatory, stress response, mitochondrial biogenesis, hormone balance, cardiometabolic, and cellular aging-related themes. For example, our results show significant variation across the dataset in two genes, *COX7C* and *UQCRQ*

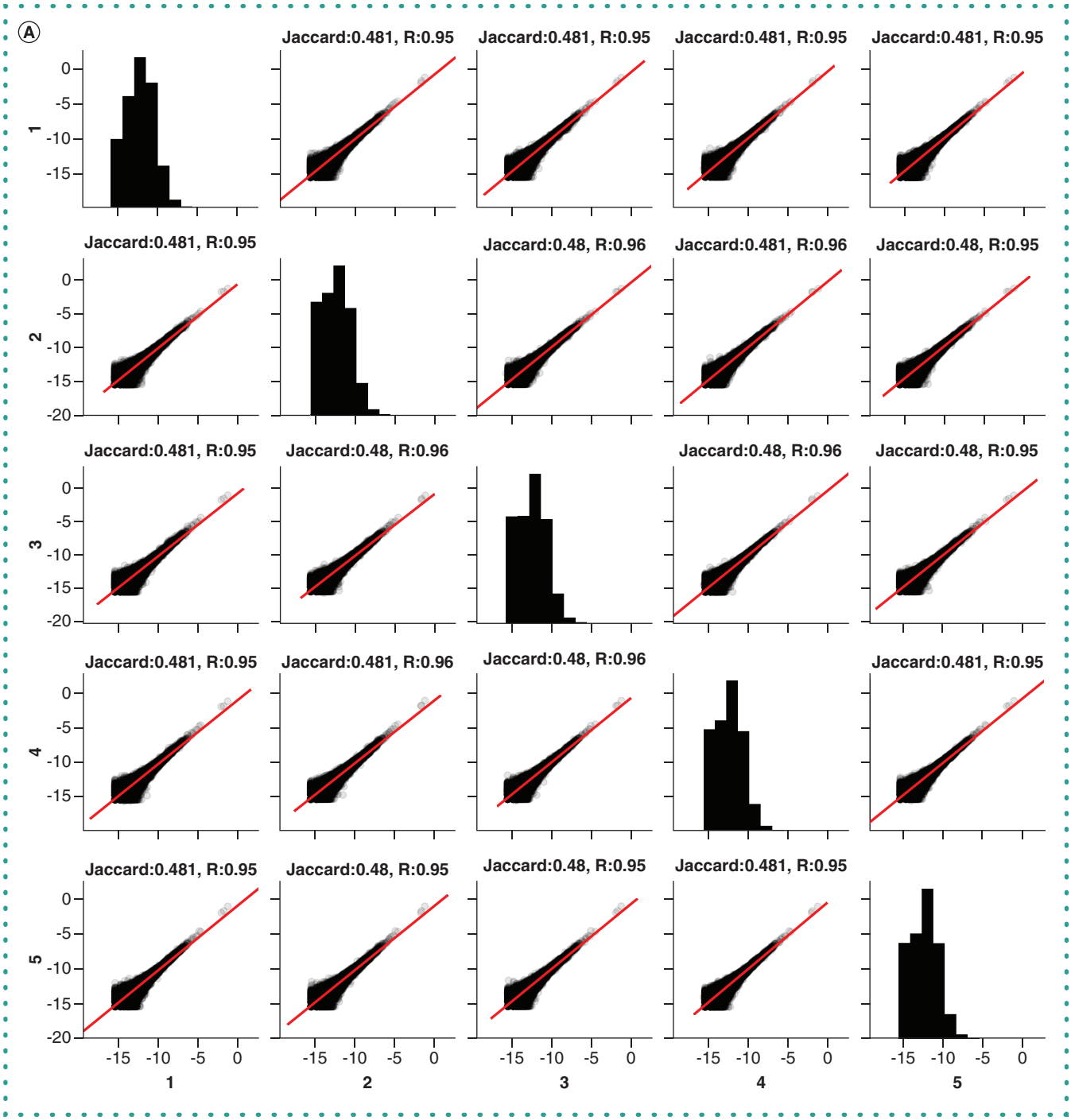

Figure 6. Longitudinal stability of whole blood transcriptome in two study participants over a 5-week period. The plots show Spearman correlation and Jaccard similarity between pairs of samples, one sample randomly selected each week over 5 weeks, for two randomly chosen participants. Correlations are calculated on log-transformed values and only for genes detected in both samples.

(SD differences 128 and 38, respectively), both of which are components of the electron transport chain of the cellular respiration process and are important for mitochondrial health. Optimal function of respiratory chain pathways is crucial for meeting the cellular demands of an energetic and healthy lifestyle. Dietary sources and supplements such as creatine, lipoic acid, CoQ10 and nicotinamide riboside help to improve energy production and mitochondrial health [44]. Foods containing resveratrol (e.g., grapes) and those containing α-lipoic acid (e.g., spinach and broccoli) also help to stimulate cellular energy production pathways through different mechanisms. Resveratrol activates mitochondrial biogenesis pathways by regulating the PGC-1α pathways, and also by regulating antioxidative enzymes, such

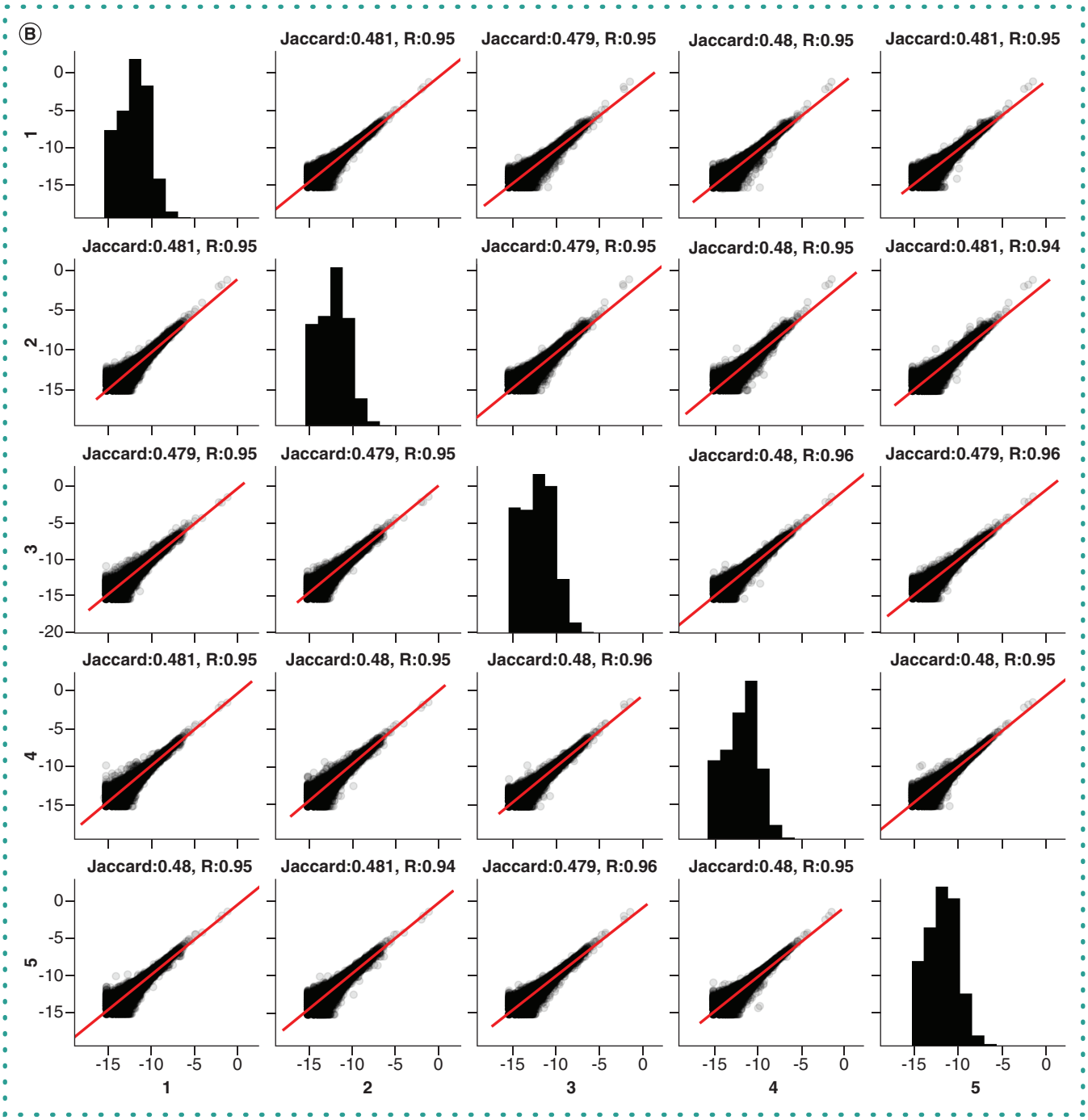

Figure 6. Longitudinal stability of whole blood transcriptome in two study participants over a 5-week period (cont.). The plots show Spearman correlation and Jaccard similarity between pairs of samples, one sample randomly selected each week over 5 weeks, for two randomly chosen participants. Correlations are calculated on log-transformed values and only for genes detected in both samples.

as NADPH oxidases, superoxide dismutases and GPx-1 [45]. On the other hand, α-lipoic acid is essential for energy production and can also act as antioxidant; it is a critical cofactor for mitochondrial energy production enzymes such as PDH, α-KGDH and BCKDC [46].

Significant variation in gene expression was also observed for *CCR7*, *FCRL3* and *FCMR*, which are involved in the immune response. CCR7 plays a role in adaptive immunity in response to viral infection by regulating trafficking and homing of T cells and B cells to secondary lymphoid tissue during viral infection and regulation of viral pathogenesis [47,48]. Supplements (or food sources) containing quercetin (onion), Vitamin C (kiwi), β-glucan (white mushroom) and nitrate (beets) are reported to ameliorate the effect of abnormal proinflammatory responses by regulating specific proinflammatory pathways that can be measured with our technology. For example,

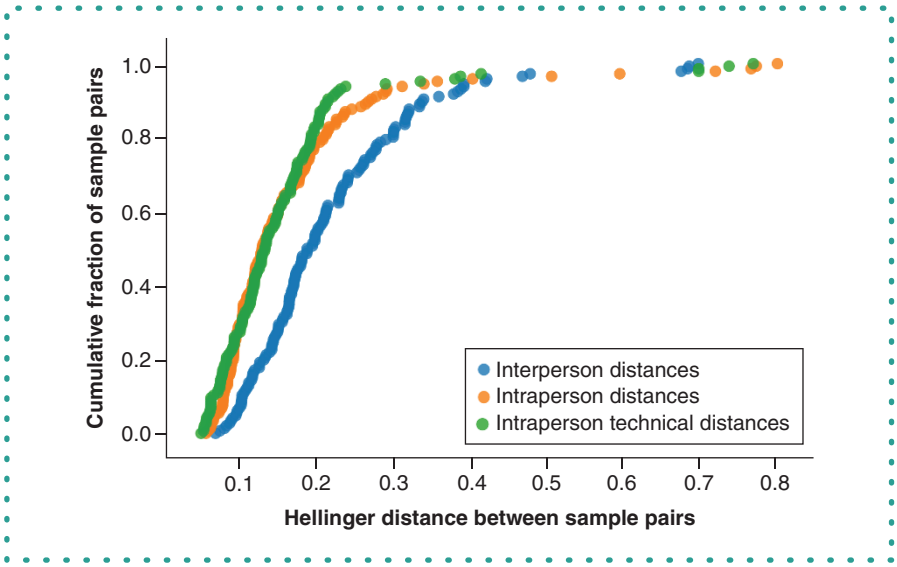

Figure 7. Empirical cumulative distribution function plots of Hellinger distance. The plots show that when taking all genes into account, pairs of samples taken from an individual within the same week (intraperson technical, green) are not obviously more similar to each other than are samples from one individual taken on two different weeks (intraperson, orange). However, pairs of samples coming from two different participants (interperson, blue) tend to be less similar.

quercetin reduces inflammation by suppressing proinflammatory prostaglandin and leukotriene biosynthesis pathways via inhibition of COX-2 and LOX-5. It also suppresses proinflammatory signaling pathways downstream of its target molecules TNF-α, IL-1α and IL-8. In addition, quercetin can also block histamine-induced inflammation [49]. Nitrate-rich food reduces inflammation by inhibiting proinflammatory signaling molecules like IL-1β, TNF-α, IL-6 and IFN-γ in various immune cells such as lymphocytes, eosinophils and monocytes [50].

The whole blood transcriptome test described here has high potential value for scientific discoveries. It overcomes the major drawbacks of the current methods – cost, visiting a clinic for a blood draw, shipping samples on dry ice and potentially spreading blood-borne pathogens – which limit their widespread use and cause exclusion of underprivileged populations. The ability of this test to be implemented anywhere (including remote rural locations) and its low cost enables it to have a widespread utility. Compared with other whole blood quantitative gene expression methods, this test displays similar precision (Figure 4), a key attribute for studying large populations across time. While this test remains expensive for very large studies, sequencing and molecular biology reagent costs continue to decrease, which will decrease the test costs even further. In addition, the Viome Research Institute offers an unlimited number of free tests to qualified applicants via its grants program, which further enables large population studies, either cross-sectional or longitudinal.

This test is integratively offered along with the previously published stool metatranscriptome test; together with upcoming saliva and vaginal metatranscriptome tests, it augments a comprehensive systems biology platform that can be applied to population-scale longitudinal studies. Large studies empowered with such data will advance the ability to identify mechanisms of chronic disease onset and progression, improve or bring new diagnostic and companion diagnostic tools, and enable precision nutrition (including foods, nutrients and probiotic supplements) and microbiome engineering in an effort to prevent and cure chronic diseases.

Future perspective

Currently, there are hundreds of chronic diseases whose combined prevalence is growing in the human population, yet whose etiologies are unknown. To understand the root causes of these complex diseases, we have to analyze the human body as a whole, including the influence of the environment, nutrition and the microbiome. This report describes an important piece of this puzzle, the whole blood transcriptome. Used in combination with the other -omics pieces, such as saliva and stool metatranscriptomics, immune repertoire, and so on, we will be able to reveal the underlying mechanisms of chronic disease. Once the root causes of chronic diseases are understood mechanistically, we will be able to prevent or cure them with a switch to a personalized diet and supplements. This will fundamentally change the way we approach their diagnosis and treatment.

Author contributions

M Vuyisich conceived and designed the method. R Toma, B Pelle, A Hatch and N Duval performed data collection. M Parks, V Gopu, P Torres, F Camacho, N Shen, S Krishnan, H Tily, A Perlina and G Banavar performed data analysis. R Toma, N Duval, M Parks, V Gopu, P

Torres, F Camacho, N Shen, S Krishnan, A Hatch, H Tily, A Perlina, G Banavar and M Vuyisich contributed to data interpretation. N Duval, M Vuyisich, R Toma, A Hatch, M Parks, H Tily, A Perlina and S Krishnan contributed to the writing of the manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Financial & competing interests disclosure

All authors are current or former employees of Viome, Inc. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

References

1. Frith MC, Pheasant M, Mattick JS. Genomics: the amazing complexity of the human transcriptome. *Eur. J. Hum. Genet.* 13(8), 894–897 (2005).
2. Lin Y, Ghazanfar S, Strbenac D *et al.* Evaluating stably expressed genes in single cells. *GigaScience* 8(9), 894–897 (2019).
3. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18(9), 1509–1517 (2008).
4. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5(7), 621–628 (2008).
5. Velculescu VE, Madden SL, Zhang L *et al.* Analysis of human transcriptomes. *Nat. Genet.* 23(4), 387–388 (1999).
6. Sipos F, Galamb O, Wichmann B *et al.* Peripheral blood based discrimination of ulcerative colitis and Crohn's disease from non-IBD colitis by genome-wide gene expression profiling. *Dis. Markers* 30(1), 1–17 (2011).
7. Suzuki E, Sugimoto M, Kawaguchi K *et al.* Gene expression profile of peripheral blood mononuclear cells may contribute to the identification and immunological classification of breast cancer patients. *Breast Cancer* 26(3), 282–289 (2018).
8. Nguyen CB, Alsøe L, Lindvall JM *et al.* Whole blood gene expression in adolescent chronic fatigue syndrome: an exploratory cross-sectional study suggesting altered B cell differentiation and survival. *J. Transl. Med.* 15(1), 102 (2017).
9. Bigler J, Boedigheimer M, Schofield JPR *et al.* A severe asthma disease signature from gene expression profiling of peripheral blood from U-BIOPRED cohorts. *Am. J. Respir. Crit. Care Med.* 195(10), 1311–1320 (2017).
10. Cabanero M, Tsao M. Circulating tumour DNA in EGFR-mutant non-small-cell lung cancer. *Curr. Oncol.* 25(0), 38–44 (2018).
11. Chauhan AK, Bhardwaj M, Chaturvedi PK. *Molecular Diagnostics in Cancer Patients*. 293–303 (2019).
12. Rosell R, Carcereny E, Gervais R *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised Phase III trial. *Lancet Oncol.* 13(3), 239–246 (2012).
13. CDC. Health and Economic Costs of Chronic Diseases (2020). https://www.cdc.gov/chronicdisease/about/costs/index.htm
14. Carlberg K, Korotkova M, Larsson L, Catrina AI, Ståhl PL, Malmström V. Exploring inflammatory signatures in arthritic joint biopsies with spatial transcriptomics. *Sci. Rep.* 9(1), 18975 (2019).
15. Haberman Y, Karns R, Dexheimer PJ *et al.* Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response. *Nat. Commun.* 10(1), 38 (2019).
16. Ostrowski J, Goryca K, Lazowska I *et al.* Common functional alterations identified in blood transcriptome of autoimmune cholestatic liver and inflammatory bowel diseases. *Sci. Rep.* 9(1), 7190 (2019).
17. Stephenson W, Donlin LT, Butler A *et al.* Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation. *Nat. Commun.* 9(1), 791 (2018).
18. Zhao C, Mao J, Ai J *et al.* Integrated lipidomics and transcriptomic analysis of peripheral blood reveals significantly enriched pathways in Type 2 diabetes mellitus. *BMC Med. Genomics* 6(Suppl. 1), S12 (2013).
19. Rappaport SM. Genetic factors are not the major causes of chronic diseases. *PLoS ONE* 11(4), e0154387 (2016).
20. Hand TW, Vujkovic-Cvijin I, Ridaura VK, Belkaid Y. Linking the microbiota, chronic disease, and the immune system. *Trends Endocrinol. Metab.* 27(12), 831–843 (2016).
21. Wilkins LJ, Monga M, Miller AW. Defining dysbiosis for a cluster of chronic diseases. *Sci. Rep.* 9(1), 12918 (2019).
22. Reynier F, Pachot A, Paye M *et al.* Specific gene expression signature associated with development of autoimmune Type-I diabetes using whole-blood microarray analysis. *Genes Immun.* 11(3), 269–278 (2010).
23. Berisha SZ, Serre D, Schauer P, Kashyap SR, Smith JD. Changes in whole blood gene expression in obese subjects with Type 2 diabetes following bariatric surgery: a pilot study. *PloS ONE* 6(3), e16729 (2011).
24. Homuth G, Wahl S, Müller C *et al.* Extensive alterations of the whole-blood transcriptome are associated with body mass index: results of an mRNA profiling study involving two large population-based cohorts. *BMC Med. Genomics* 8(1), 65 (2015).
25. Pedrotty DM, Morley MP, Cappola TP. Transcriptomic biomarkers of cardiovascular disease. *Prog. Cardiovasc. Dis.* 55(1), 64–69 (2012).
26. Wirka RC, Pjanic M, Quertermous T. Advances in transcriptomics. *Circ. Res.* 122(9), 1200–1220 (2018).
27. Chaussabel D, Pascual V, Banchereau J. Assessing the human immune system through blood transcriptomics. *BMC Biol.* 8(1), 84 (2010).
28. Liew C-C, Ma J, Tang H-C, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J. Lab. Clin. Med.* 147(3), 126–132 (2006).
29. Sen P, Kemppainen E, Orešič M. Perspectives on systems modeling of human peripheral blood mononuclear cells. *Front. Mol. Biosci.* 4, 96 (2018).
30. Ng EKO, Tsui NBY, Lau TK *et al.* mRNA of placental origin is readily detectable in maternal plasma. *Proc. Natl Acad. Sci. USA* 100(8), 4748–4753 (2003).
31. Pinzani P, Salvianti F, Pazzagli M, Orlando C. Circulating nucleic acids in cancer and pregnancy. *Methods* 50(4), 302–307 (2010).
32. Pös O, Biró O, Szemes T, Nagy B. Circulating cell-free nucleic acids: characteristics and applications. *Eur. J. Hum. Genet.* 26(7), 937–945 (2018).
33. de Souza MF, Kuasne H, Barros-Filho M de C *et al.* Circulating mRNAs and miRNAs as candidate markers for the diagnosis and prognosis of prostate cancer. *PLoS ONE* 12(9), e0184094 (2017).
34. Huang Z, Whelan CV, Foley NM *et al.* Longitudinal comparative transcriptomics reveals unique mechanisms underlying extended healthspan in bats. *Nat. Ecol. Evol.* 3(7), 1110–1120 (2019).
35. Yates AD, Achuthan P, Akanni W *et al.* Ensembl 2020. *Nucleic Acids Res.* 48(D1), D682–D688 (2019).

36. Schneider VA, Graves-Lindsay T, Howe K *et al.* Evaluation of GRCh38 and *de novo* haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* 27(5), 849–864 (2017).
37. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. *Salmon* provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14(4), 417–419 (2017).
38. Akoglu H. User's guide to correlation coefficients. *Turk. J. Emerg. Med.* 18(3), 91–93 (2018).
39. Blair SN, Horton E, Leon AS *et al.* Physical activity, nutrition, and chronic disease. *Med. Sci. Sports Exerc.* 28(3), 335–349 (1996).
40. Casamassimi A, Federico A, Rienzo M, Esposito S, Ciccodicola A. Transcriptome profiling in human diseases: new advances and perspectives. *Int. J. Mol. Sci.* 18(8), 1652 (2017).
41. Durack J, Lynch SV. The gut microbiome: relationships with disease and opportunities for therapy. *J. Exp. Med.* 216(1), 20–40 (2018).
42. Wang B, Yao M, Lv L, Ling Z, Li L. The human microbiota in health and disease. *Engineering* 3(1), 71–82 (2017).
43. Hatch A, Horne J, Toma R *et al.* A robust metatranscriptomic technology for population-scale studies of diet, gut microbiome, and human health. *Int. J. Genomics* 2019, 1–9 (2019).
44. Rak M, Bénit P, Chrétien D *et al.* Mitochondrial cytochrome C oxidase deficiency. *Clin. Sci.* 130(6), 393–407 (2016).
45. Spanier G, Xu H, Xia N *et al.* Resveratrol reduces endothelial oxidative stress by modulating the gene expression of superoxide dismutase 1 (SOD1), glutathione peroxidase 1 (GPx1) and NADPH oxidase subunit (Nox4). *J. Physiol. Pharmacol.* 60(Suppl. 4), 111–116 (2009).
46. Ong SLH, Vohra H, Zhang Y, Sutton M, Whitworth JA. The effect of alpha-lipoic acid on mitochondrial superoxide and glucocorticoid-induced hypertension. *Oxid. Med. Cell. Longev.* 2013, 1–9 (2013).
47. Yan Y, Chen R, Wang X *et al.* CCL19 and CCR7 expression, signaling pathways, and adjuvant functions in viral infection and prevention. *Front. Cell Dev. Biol.* 7, 212 (2019).
48. Kocks JR, Adler H, Danzer H *et al.* Chemokine receptor CCR7 contributes to a rapid and efficient clearance of lytic murine γ-herpes virus 68 from the lung, whereas bronchus-associated lymphoid tissue harbors virus during latency. *J. Immunol.* 182(11), 6861–6869 (2009).
49. Li Y, Yao J, Han C *et al.* Quercetin, inflammation and immunity. *Nutrients* 8(3), 167 (2016).
50. Wink DA, Hines HB, Cheng RYS *et al.* Nitric oxide and redox mechanisms in the immune response. *J. Leukoc. Biol.* 89(6), 873–891 (2011).