




OPEN

## Eight-year longitudinal study of whole blood gene expression profiles in individuals undergoing long-term medical follow-up

Yoshio Sakai<sup>1,8</sup>, Alessandro Nasti<sup>2,8</sup>, Yumie Takeshita<sup>3</sup>, Miki Okumura<sup>3</sup>, Shinji Kitajima<sup>4</sup>, Masao Honda<sup>1</sup>, Takashi Wada<sup>4</sup>, Seiji Nakamura<sup>5</sup>, Toshinari Takamura<sup>3</sup>, Takuro Tamura<sup>6</sup>, Kenichi Matsubara<sup>7</sup> & Shuichi Kaneko<sup>1,2</sup>

Blood circulates throughout the body via the peripheral tissues, contributes to host homeostasis and maintains normal physiological functions, in addition to responding to lesions. Previously, we revealed that gene expression analysis of peripheral blood cells is a useful approach for assessing diseases such as diabetes mellitus and cancer because the altered gene expression profiles of peripheral blood cells can reflect the presence and state of diseases. However, no chronological assessment of whole gene expression profiles has been conducted. In the present study, we collected whole blood RNA from 61 individuals (average age at registration, 50 years) every 4 years for 8 years and analyzed gene expression profiles using a complementary DNA microarray to examine whether these profiles were stable or changed over time. We found that the genes with very stable expression were related mostly to immune system pathways, including antigen cell presentation and interferon-related signaling. Genes whose expression was altered over the 8-year study period were principally involved in cellular machinery pathways, including development, signal transduction, cell cycle, apoptosis, and survival. Thus, this chronological examination study showed that the gene expression profiles of whole blood can reveal unmanifested physiological changes.

### Abbreviation

cDNA Complementary DNA

Blood circulates continuously throughout the body via the peripheral circulatory system. Peripheral blood contains a wide variety of cells. In particular, leukocytes consist of phenotypically and functionally miscellaneous cells such as granulocytes, which are myeloid-lineage cells, and lymphocytes, which are associated with the lymphatic system<sup>1</sup>. These cells are indispensable for protecting the body from harmful exogenous pathogens as well as endogenous emerging diseases such as cancer; thus, they potentially respond to miscellaneous alterations of the physiological condition of the body.

Gene expression profiles reflect the specific characteristics of cells as well as the physiological condition of the host<sup>2,3</sup>. In this sense, as the blood circulates throughout the body, gene expression analysis of whole blood cells is potentially a novel tool to assess an individual's biological characteristics as well as the presence of "silent" diseases that do not result in detectable clinical signs and symptoms. Previously, we reported that comprehensive analysis of the gene expression profiles of peripheral blood cells can help us to understand a patient's condition, including immunological features in diabetes and in various cancers of the digestive system such as hepatocellular

<sup>1</sup>Department of Gastroenterology, Kanazawa University Hospital, 13-1 Takara-machi, Kanazawa, Ishikawa 920-8641, Japan. <sup>2</sup>System Biology, Graduate School of Advanced Preventive Medical Sciences, Kanazawa University, Kanazawa 920-8641, Japan. <sup>3</sup>Department of Endocrinology and Metabolism, Graduate School of Medical Sciences, Kanazawa University, Kanazawa 920-8640, Japan. <sup>4</sup>Department of Nephrology, Kanazawa University Hospital, Kanazawa 920-8641, Japan. <sup>5</sup>DNA Chip Research Inc, Tokyo 105-0022, Japan. <sup>6</sup>Research and Development Center for Precision Medicine, University of Tsukuba, Tsukuba 305-8550, Japan. <sup>7</sup>iLAC Co., Ltd, Tsukuba 305-0821, Japan. <sup>8</sup>These authors contributed equally: Yoshio Sakai and Alessandro Nasti. ✉email: yoshios@m-kanazawa.jp

Clinical parameters	Males (n = 34) and Females (n = 27); n = 61					
	2008		2012		2016	
	Average	±SD	Average	±SD	Average	±SD
Age (years)	50.6	10.0	54.6	10.0	58.6	10.1
Height (cm)	164.2	8.2	163.9	8.2	163.7	8.2
Weight (kg)	61.6	9.6	61.2	9.1	61.2	9.3
BMI (kg/m <sup>2</sup> )	22.8	2.6	22.8	2.5	22.8	2.7
Fat ratio (%)	27.0	6.0	26.7	6.1	26.9	6.1
Fat mass (Kg)	n/a	n/a	16.2	4.4	16.5	4.4
Fat free ratio (%)	n/a	n/a	44.6	7.8	45.1	8.2
Muscle ratio (%)	n/a	n/a	42.2	7.4	42.7	7.8
Bone mass (Kg)	n/a	n/a	2.4	0.4	2.4	0.4
Systolic blood pressure (mmHg)	123	16	122	14	121	13
Diastolic blood pressure (mmHg)	78	11	75	9	73	11
HbA1c (JDS) (%)	5.1	0.3	5.1	0.3	5.2	0.4
Total cholesterol (mg/dL)	203	35	199	32	201	36
TG (mg/dL)	130	113	111	71	103	64
HDL-cholesterol (mg/dL)	59	15	56	14	58	15
LDL-cholesterol (mg/dL)	118	28	115	26	117	32
AST (GOT) (IU/L)	22	8	25	14	23	8
ALT (GPT) (IU/L)	22	13	24	16	21	12
Gamma GTP (IU/L)	47	54	54	112	33	26
White blood cell count (× 10 <sup>3</sup> /ml)	5.6	1.6	5.1	1.5	5.0	1.3
Hemoglobin count (g/dL)	14.2	1.4	14.1	1.2	13.8	1.6
Hematocrit count (%)	42.6	3.4	42.6	3.1	41.7	4.0
Platelet count (× 10 <sup>4</sup> /ml)	23.5	5.1	23.2	5.6	23.7	6.1
Hypertension (n patients out of 61)	11/61	n/a	16/61	n/a	19/61	n/a
Dyslipidemia (n patients out of 61)	9/61	n/a	13/61	n/a	13/61	n/a
Diabetes (n patients out of 61)	10/61	n/a	14/61	n/a	19/61	n/a

**Table 1.** Clinical features of the 61 individuals which were examined for expression of whole blood cells in 2008, 2012 and 2016.

carcinoma, colon cancer, and pancreas cancer<sup>4-9</sup>. These studies strongly suggest that certain changes in gene expression reflect alterations in the immune condition of local lesions. However, despite our comprehensive investigations of the gene expression profiles of whole blood cells using complementary DNA (cDNA) microarray analysis at a specific time point, it is not yet known how the gene expression profiles of whole blood cells change over an extended period of time in individuals without a serious disease such as malignancies. It is extremely important to determine whether and how the gene expression profiles of peripheral blood are affected, especially at the early stage of a serious illness. To assess the alterations in gene expression in peripheral blood cells associated with disease, we need to characterize the stable gene expression profiles of healthy individuals over an extended period of time.

Based on these backgrounds, in the present study, we collected blood samples from 61 individuals who were undergoing regular medical health check-ups or medical follow-up examinations over an 8-year period and comprehensively analyzed their gene expression profiles using a cDNA microarray. We found that 1509 genes were stably expressed over this 8-year period, and nearly all of these genes were involved in immune-related pathways. We identified 3251 genes whose expression was changed over the same period and were involved in the development, signal transduction, cell cycle, apoptosis, survival, chemotaxis, and immune response pathways. These results imply that the physiological and structural homeostatic features of the body change over time, whereas the immune system, which involves central immune regulation, antigen-presenting cells, and related immune cells, is maintained in a relatively stable state.

## Methods

**Participants.** The participants were individuals who regularly visited the Public Central Hospital of Matto Ishikawa for medical check-ups. Except for some participants who had a pre-existing clinical condition such as hypertension, dyslipidemia, or diabetes (Table 1), all of the other participants were considered to be disease-free. Every 4 years from 2008 to 2016, all participants, who were fully informed of the nature of the study and provided written informed consent to participate, were registered. The study protocol was approved by the internal review boards of the Public Central Hospital of Matto Ishikawa and Kanazawa University, and the study was conducted in accordance with the principles of the Declaration of Helsinki.

**Study design and enrollment.** The study was designed to analyze the gene expression features of whole blood of the registered individuals over an 8-year period from 2008. The number of registered participants was 61.

**Whole blood RNA isolation and gene expression analysis using cDNA microarray analysis.** Peripheral whole blood was collected periodically from the participants in 2008, 2012, and 2016, placed in PAXgene Blood RNA Tubes (PreAnalytiX GmbH, Hombrechtikon, Switzerland) to stabilize the RNA, and stored as per the manufacturer's protocol. Total RNA was extracted using a PAXgene Blood RNA Kit (PreAnalytiX), amplified, and labeled with Cy3 by using a Low Input Quick Amp Labeling Kit (Agilent Technologies, Santa Clara, CA). cRNA hybridization was performed using an Agilent SurePrint G3 Human Gene Expression 8 × 60 K v2 Microarray (Design ID:039494) and detected with a G2600D DNA Microarray Scanner (Agilent Technologies). The intensity value of each scanned feature was quantified using Agilent Feature Extraction Software (version 11.5.1.1, <https://www.agilent.com/en/genomics-software-downloads>; Agilent Technologies). To confirm the reproducibility and stability of gene expression by the DNA microarray method, the extracted blood RNA samples were analyzed twice for the transcribed mRNAs (Supplementary Fig. S1); signal intensity was adjusted using the quantile normalization method<sup>10</sup>. After excluding poorly annotated probes and low-signal probes, 10,590 probes were extracted for further statistical analysis.

**Data analysis of gene expression.** The 10,590 pre-filtered genes were used to identify differentially and non-differentially expressed genes for the years 2008, 2012, and 2016. BRB-ArrayTools v4.6.1 (<http://linus.nci.nih.gov/BRB-ArrayTools.html><sup>11</sup>) was used for Class Comparison analysis; hence, for the identification of differentially/non-differentially expressed genes and for Geneset Class Comparison analysis for the identification of the involved cell types, multiple genes were reduced to one per gene symbol by using the maximally expressed gene measured by average intensity across arrays. MetaCore software (version 21.2 build 70,500, <https://portal.genego.com/>; Clarivate Analytics, Philadelphia, PA) was used, as described previously<sup>12</sup>, for enrichment analyses of the pathway maps of the differentially and non-differentially expressed genes.

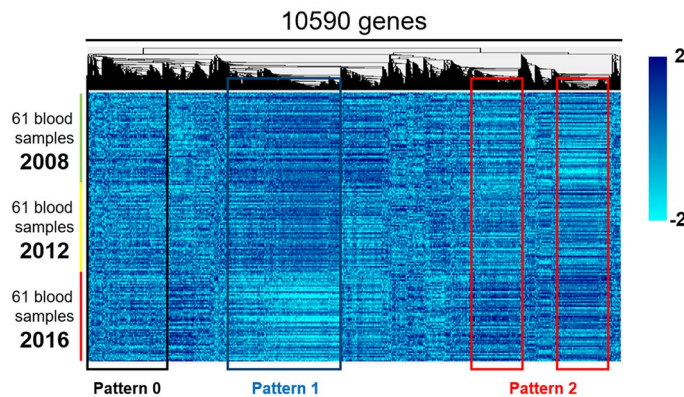
In the second analysis, we used the 10,590 pre-filtered genes to perform deconvolution analysis with the analytical tool CIBERSORTx (<https://cibersortx.stanford.edu/>)<sup>13,14</sup>, and obtained an estimation of the abundance of cell types by using the 22 pre-defined immune cell types present as a reference. Quantile normalization was applied, followed by bulk mode batch correction for the removal of technical differences between the reference signature matrix profiles and the investigated set of samples; 100 permutations were performed for significance analysis.

## Results

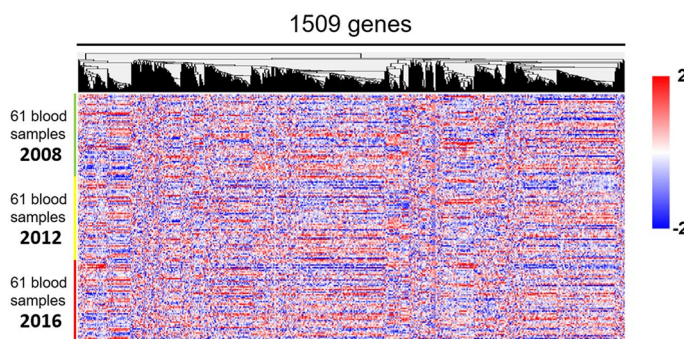
**Clinical features and overall gene expression profiles of 61 participants over an 8-year period.** The clinical features of the 61 participants are shown for 2008–2016 (Table 1); overall, there were no major differences in their clinical features after 8 years except for age. Using the cDNA microarray system, we obtained gene expression data for 2008, 2012, and 2016. The number of genes whose expression data passed the quality check was 10,590 genes; the gene expression heat-map of these genes implied a stable expression pattern (Pattern 0) over the 8-year observation period, whereas some genes demonstrated altered expression patterns. Specifically, the expression of a group of genes was downregulated in 2016 compared with 2008 and 2012 (Pattern 1), whereas the expression of a second group of genes was upregulated in 2016 compared with 2008 and 2012 (Pattern 2; Fig. 1). Among the 10,590 genes, we identified 2010, 1509, and 754 genes whose expression did not change significantly over time (F-test parametric  $p$ -value > 0.1, 0.2, and 0.5, respectively). For the following analyses, we chose the set of 1,509 genes as the most representative non-differentially expressed (stable) genes (F-test parametric  $p$ -value > 0.2; Figs. 2 and 4a). In contrast, the expression of 3,251 genes was changed over the 8-year period (F-test parametric  $p$ -value <  $1.0 \times 10^{-4}$ , false discovery rate <  $1.0 \times 10^{-3}$ ; Figs. 3a and 4a) with two major patterns of change: the first group consisted of 2,005 genes whose expression was significantly downregulated in 2016 compared with 2008 and 2012 (Pattern 1; Figs. 3b and 4a), and the second group consisted of 1093 genes whose expression was significantly upregulated in 2016 compared with 2008 and 2012 (Pattern 2; Figs. 3c and 4a); 153 genes of the 3,251 genes were not categorized (Fig. 4a) into either of these two patterns and were excluded from the following analyses. Finally, we performed Enrichment Analyses and Geneset Class Comparison of the 1509 stably expressed genes, 3,251 differentially expressed genes, 2,005 downregulated genes, and 1093 upregulated genes (Fig. 4b).

**Biological features of genes whose expression was not altered during the 8-year study period.** We next assessed the features of the 1509 stably expressed genes by analyzing which pathways they were involved in by using MetaCore software. Among the top 16 pathways (Table 2, Supplementary Table S1), all of the maps indicated immune system-related pathways. The most distinctive pathways were immune response-related, which were represented by type 1 and type 2 interferon-mediated antigen presentation and T cell regulation (Table 2, Supplementary Table S1). The other immune-related pathways were involved in miscellaneous cell roles as well as the immune cell chemotaxis system via SDF-1/CXCR4. Key genes were *CD3D*, *LCK*, and *ZAP70*, which are related to signal transduction in T cell activation<sup>15,16</sup>, and *CCL2*, which is a chemokine capable of recruiting monocytes, dendritic cells, and memory T cells<sup>17,18</sup>. These pathways suggested that the overall immune system was rather stable over an 8-year period in middle-aged people, at least in the context of conventional immune system categories.

To identify which immune-mediating cells were related to the 1,509 genes stably expressed in peripheral blood, we performed enrichment analysis by cell type (Supplementary Table S2). The identified cell types included



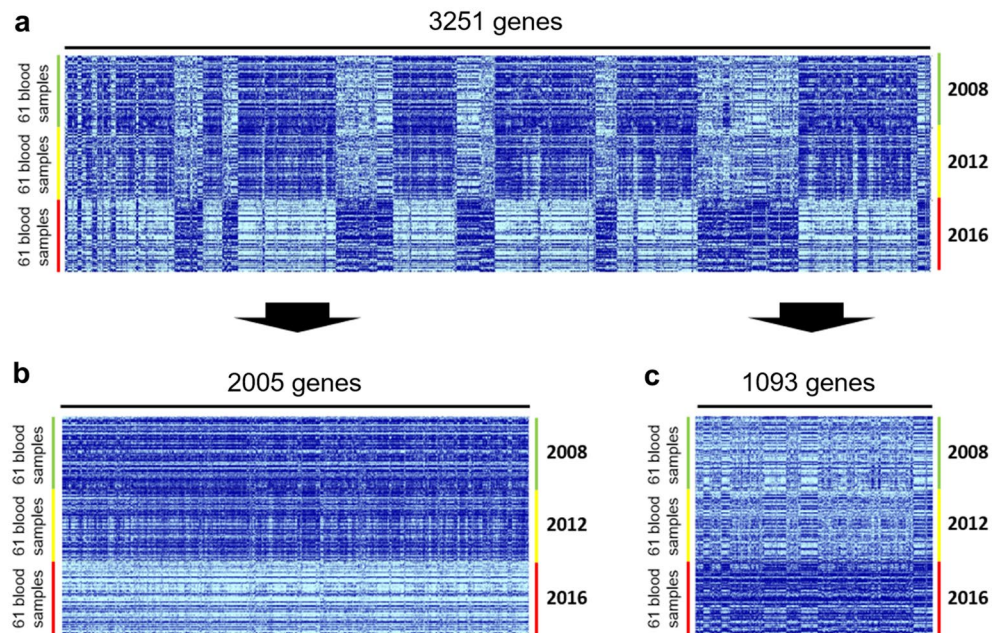
**Figure 1.** Heatmap of 10,590 genes that passed the preliminary filtering criteria, showing a distinct pattern of expression. The 10,590 genes that passed the filtration quality check suggested a pattern of stable gene expression (circumscribed by the black line; Pattern 0), whereas some genes had altered expression during the 8-year observation period. Specifically, there was a group of genes with downregulated expression in 2016 (circumscribed by the aqua blue line; Pattern 1) and a second group of genes with upregulated expression (circumscribed by the red line; Pattern 2) compared with 2008 and 2012. Blood samples from 61 individuals were used for each time point.



**Figure 2.** Identification of 1,509 non-differentially expressed genes between the years 2008, 2012, and 2016. Among the 10,590 pre-filtered genes, the expression of 1,509 genes did not change over time and they were defined as the most representative stably expressed genes (Pattern 0; F-test parametric  $p$ -value  $> 0.2$ ).

NK cells, dendritic cells, monocytes, neutrophils, CD8+ T cells, and regulatory T cells, which are compatible with the results of the pathway map analysis described above.

**Biological features of genes with altered expression in peripheral blood cells.** We also examined the biological characteristics of genes whose expression was changed over the 8-year study period. Biological pathway map analysis of all genes whose expression was altered indicated that they had roles in the development, oxidative stress, cell cycle, signal transduction, apoptosis and survival, chemotaxis, and immune response pathways (Table 2, Supplementary Table S3), suggesting that cellular homeostatic features were affected. Pathway analysis of the 2005 downregulated genes (Pattern 1) showed that they were related to the signal transduction, apoptosis and survival, cell cycle, proteolysis, development, DNA damage, and transport pathways (Table 2, Supplementary Table S4). Key genes were *AKT2*, which is a putative oncogene<sup>19</sup>, and *CREB1*, which is involved in cell proliferation<sup>20</sup>. Pathway analysis of the 1093 upregulated genes (Pattern 2) indicated that they were related to the cytoskeleton remodeling, development, blood coagulation, signal transduction, neurophysiological process, apoptosis and survival, and immune response pathways (Table 2, Supplementary Table S5). Key genes included *PIK3CB* and *NFKB1*, which are related to cell growth and apoptosis<sup>21,22</sup>. Furthermore, the cell types involved with the downregulated genes were suggested to be B cells and T helper cells (Supplementary Table S6), whereas the upregulated genes were related to platelets and CD34+ megakaryocyte progenitors (Supplementary Table S7).

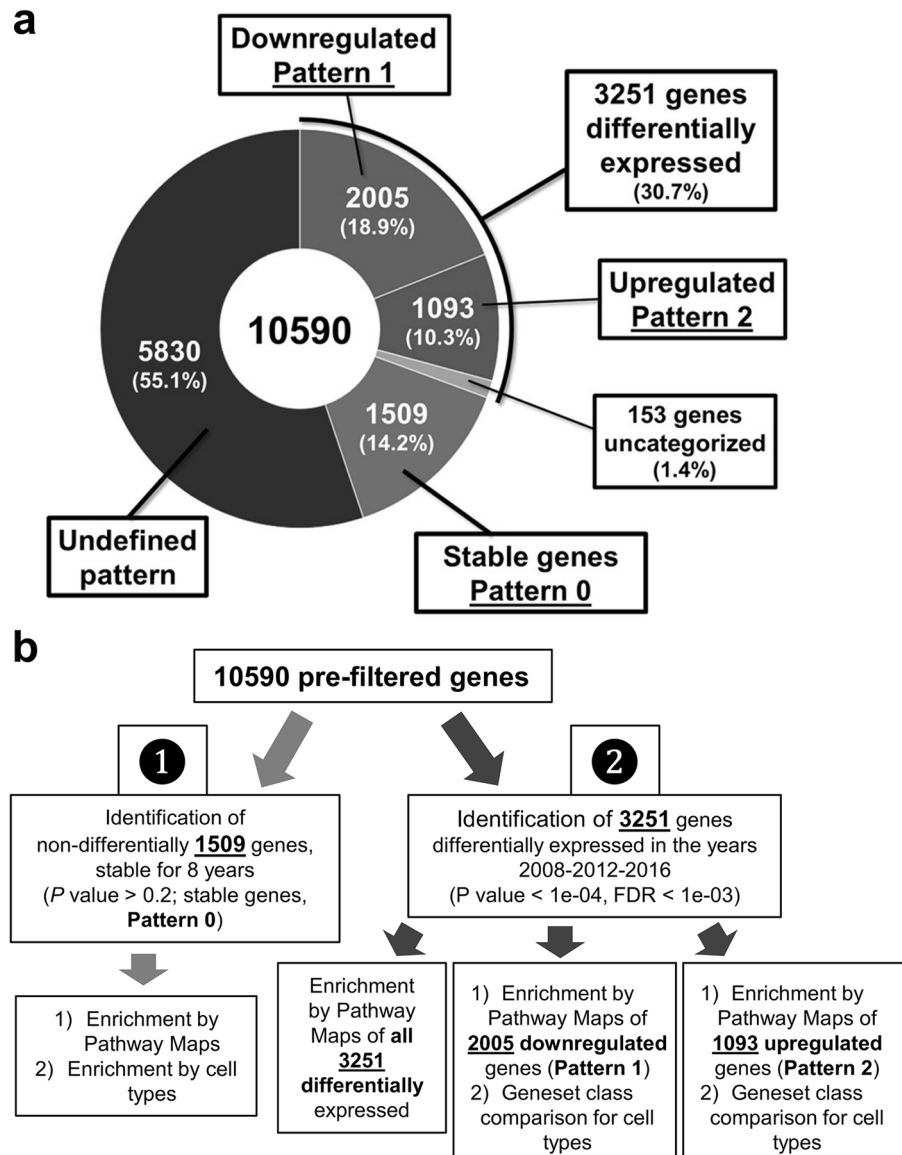


**Figure 3.** Identification of 3251 differentially expressed genes, followed by stratification of the genes into downregulated (Pattern 1) or upregulated (Pattern 2) genes in the years 2008, 2012, and 2016. (a) Heatmap of the 3251 genes with altered expression over the 8-year monitoring period (F-test parametric  $p$ -value  $< 1.0 \times 10^{-4}$ , false discovery rate  $< 1.0 \times 10^{-3}$ ). (b) Heatmap of the 2005 genes that were downregulated over time; the expression of each gene in 2016 was lower than its expression in 2008 and 2012 (Pattern 1). (c) Heatmap of the 1093 genes that were upregulated over time; the expression of each gene in 2016 was higher than its expression in 2008 and 2012 (Pattern 2).

**Stability and variability of whole blood cell types over time.** Next, we examined the abundance of specific cell types in blood and their variability over time by deconvolution analysis (CIBERSORTx). Out of the 22 immune cell types investigated, we found 14 cell types whose frequency was unaltered (Fig. 5a) and 8 cell types whose frequency changed over time (Fig. 5b). The unaltered 14 cell types were as follows: NK resting cells, monocytes, neutrophils, CD8+ T cells, naïve CD4+ T cells, resting memory CD4+ T cells, follicular helper T cells,  $\gamma\delta$  T cells, resting dendritic cells, naïve B cells, M1 and M2 macrophages, resting mast cells, and eosinophils, which were consistent with the identified immune-mediating cells (Supplementary Table S2) related to the analysis of the 1,509 stably expressed genes. Regarding the 8 cell types that changed in frequency over time, 3 cell types increased in frequency (plasma cells, regulatory T cells, and M0 macrophages), whereas 5 cell types decreased in frequency (B memory cells, activated memory CD4+ T cells, activated NK cells, activated dendritic cells, and activated mast cells, although this last population only had a tendency to decrease in frequency over time).

## Discussion

In the present study, we observed the gene expression profiles of whole blood cells over an 8-year period in 61 individuals. The average age at registration was approximately 50 years. We observed that the expression of 1509 genes was stable in all individuals over this 8-year period. Biological pathway map analysis for these stably expressed genes indicated that they had roles in immune response-related pathways, including antigen-presenting cells and interferon signaling; consistently, the involved cells included antigen-presenting cells and lymphocyte-lineage cells. In contrast, we found 3251 genes whose expression was altered over the 8-year study period; within these genes, we observed that there were two groups that showed either downregulated expression in 2016 (Pattern 1) or upregulated expression in 2016 (Pattern 2) compared with 2008 and 2012. As a whole, the 3251 altered genes were involved in the signal transduction, apoptosis and survival, cell cycle, proteolysis, development, DNA damage, transport, cytoskeleton remodeling, blood coagulation, and neurophysiological process pathways, in addition to a minor contribution to immune response pathways, suggesting that the cellular machinery is the major altered biological process. Furthermore, by examining the abundance of specific cell types in blood and their variability over time using deconvolution analysis/digital cytometry, we confirmed that the cell types that did not change over time were related to lymphocyte-lineage cells and monocytic and granulocytic cell types.



**Figure 4.** Percentage fraction of stable and altered genes within the 10,590 genes and relative flow analysis chart. **(a)** Pie chart indicating the percentage of stably expressed genes (Pattern 0), downregulated genes (Pattern 1), and upregulated genes (Pattern 2) within all 10,590 pre-filtered genes. **(b)** Analysis flow for enrichment and Geneset Class Comparison analyses.

Cell subpopulations whose frequency increased over time were represented by plasma cells, regulatory T cells, and M0 macrophages, whereas subpopulations whose frequency decreased over time consisted of memory B cells, memory CD4+ T cells, NK cells, dendritic cells, and mast cells. These cells that changed in frequency are interesting targets for future detailed investigations.

Blood is a very useful and convenient sample for medical examination<sup>23</sup>. It consists of white blood cells, red blood cells, and platelets as cellular components; among them, white blood cells contain immune-mediating cells that respond to abnormal endogenous or exogenous pathogens or lesions<sup>24</sup>. Their reaction to lesions is

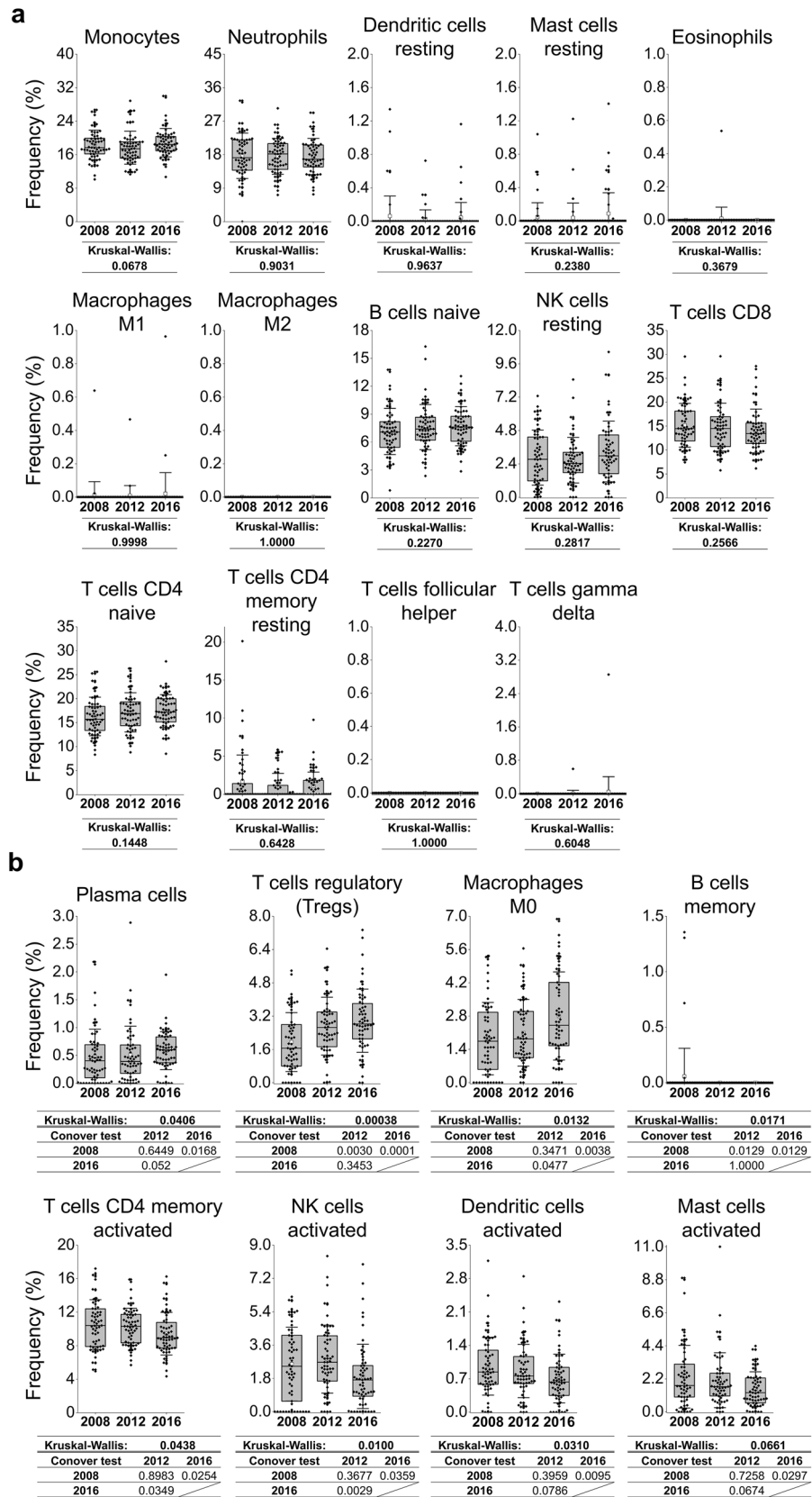
1509 non-differentially genes (P value > 0.2; stable genes; Fig. 2)	3251 genes differentially expressed (P value < 1e-04, FDR < 1e-03; Fig. 3)		
MetaCore enrichment by pathway maps of stable 1509 genes (Pattern 0; Supplementary Table S1)	MetaCore enrichment of all 3251 genes differentially expressed (Supplementary Table S3)	MetaCore enrichment by pathway maps of 2005 downregulated genes (pattern 1; Supplementary Table S4)	MetaCore enrichment by pathway maps of 1093 upregulated genes (pattern 2; Supplementary Table S5)
<p><i>Top 16 pathways related to: Immune response:</i> induction of the antigen presentation machinery by IFN-<math>\gamma</math>, IFN-<math>\alpha</math>/beta signaling via JAK/STAT, IFN-<math>\alpha</math>/beta signaling via MAPKs, antigen presentation by MHC class II, IFN-<math>\gamma</math> signaling via MAPK, inhibitory PD-1 signaling in T cells</p> <p><i>Immune related pathway (MetaCore noncategorized pathways):</i> putative role of Tregs in COPD, role of integrins in eosinophil degranulation in asthma, down-regulation of mast cell functions through ITIM-containing inhibitory receptors in asthma, SLE genetic marker-specific pathways in T cells, breakdown of CD4+ T cell peripheral tolerance in type 1 diabetes mellitus, maturation and migration of dendritic cells in skin sensitization, NETosis in SLE, neutrophil chemotaxis in asthma</p> <p><i>Chemotaxis:</i> SDF-1/CXCR4-induced chemotaxis of immune cells</p>	<p><i>Top 16 pathways related to: development:</i> thromboxane A2 signaling pathway, VEGF signaling via VEGFR2-generic cascades, role of HDAC and calcium/calmodulin-dependent kinase (CaMK) in control of skeletal myogenesis</p> <p><i>Oxidative stress:</i> ROS-induced cellular signalling</p> <p><i>Cell cycle:</i> influence of ras and rho proteins on G1/S</p> <p><i>Transition signal transduction:</i> calcium-mediated signaling, mTORC1 downstream signaling, adenosine A2B receptor signaling pathway, PKA signaling, MIF signaling pathway</p> <p><i>Immune response:</i> platelet activating factor/PTAFR pathway signaling, B cell antigen receptor (BCR) pathway</p> <p><i>Chemotaxis:</i> lysophosphatidic acid signaling via GPCRs</p> <p><i>Apoptosis and survival:</i> NGF/TrkA PI3K-mediated signaling</p>	<p><i>Top 16 pathways (DOWNREGULATED) related to:</i></p> <p><i>Signal transduction:</i> mTORC1 downstream signaling, calcium-mediated signalling</p> <p><i>Apoptosis and survival:</i> role of PKR in stress-induced apoptosis, TNFR1 signaling pathway, endoplasmic reticulum stress response pathway, NGF/TrkA PI3K-mediated signalling</p> <p><i>Chemotaxis:</i> lysophosphatidic acid signaling via GPCRs</p> <p><i>Cell cycle:</i> influence of ras and rho proteins on G1/S transition</p> <p><i>Proteolysis:</i> Putative SUMO-1 pathway</p> <p><i>Development:</i> thromboxane A2 signaling pathway, the role of GDNF ligand family/RET receptor in cell survival, growth and proliferation</p> <p><i>DNA damage:</i> ATM/ATR regulation of G2/M checkpoint: cytoplasmic signalling</p> <p><i>Transport:</i> RAN regulation pathway</p> <p><i>Immune response:</i> TLR2 and TLR4 signaling pathways</p>	<p><i>Top 16 pathways (UPREGULATED) related to:</i></p> <p><i>Immune response:</i> platelet activating factor/PTAFR pathway signalling</p> <p><i>Cytoskeleton remodeling:</i> regulation of actin cytoskeleton organization by the kinase effectors of rho GTPases</p> <p><i>Development:</i> thromboxane A2 signaling pathway, thrombospondin 1 signaling, WNT/Beta-catenin signaling in the cytoplasm</p> <p><i>Blood coagulation:</i> GPCRs in platelet aggregation</p> <p><i>Signal transduction:</i> adenosine A2B receptor signaling pathway, adenosine A3 receptor signaling pathway, MIF signaling pathway, PKA signalling</p> <p><i>Neurophysiological process:</i> constitutive and regulated NMDA receptor trafficking</p> <p><i>Apoptosis and survival:</i> Phosphorylation in TNF-<math>\alpha</math>-induced NF-<math>\kappa</math>B signalling</p> <p>Platelet activation as a result of endothelial dysfunction after stenting</p>

**Table 2.** Summary table of MetaCore enrichment analyses for differentially and non-differentially (stable) genes.

associated with alterations in gene expression<sup>25–30</sup>. Because humans have a biological system to maintain physiologic homeostasis, gene expression in whole blood cells should be generally steady in the absence of disease. However, as humans age, there are substantial changes to the biological function of the body<sup>31</sup>. In this sense, to monitor the features of stable health, it is important to characterize the total stable and altered gene expression profiles of peripheral blood cells over a substantial period of time. Genes with stable expression in peripheral blood have the potential to be used for monitoring; genes whose expression is altered should be studied further to determine whether the involved biological processes are related to changes in the participants' fundamental health condition over the 8-year study period.

The current study provided precious information suggesting important biological features; the stable expression of genes in whole peripheral blood cells was related mostly to the central immune regulation system, including antigen-presenting cells and interferon signaling. As we mentioned above, the average age of the participants when the study was initiated was 50 years and the observation period was 8 years. To date, we have reported that gene expression analysis of whole blood cells is beneficial for understanding diseases such as cancer and metabolic syndrome, including diabetes mellitus<sup>4,7,12,32</sup>. From 50 to 60 years of age, various common diseases may develop due to increasing age<sup>33</sup>; thus, laboratory medicine using gene expression analysis of whole blood cells might have the potential to be a novel avant-garde tool for understanding and obtaining an early diagnosis of diseases. The high number of 3251 genes whose expression was altered over the 8-year study period indicated the general constituents of tissues or cells. The importance of these biological features should be investigated further in terms of intra-individual health conditions. It follows that future analyses should be focused on a comparison of groups stratified by specific diseases or age. Taking the present study results and our previous studies of changes in gene expression in cancers of the digestive system, the chronological examination of the gene expression profiles of whole blood has the potential to be a novel laboratory-based clinical approach.

In conclusion, we reported the long-term gene expression profiles of whole blood cells in 61 individuals who underwent regular medical health check-ups or outpatient clinic visits. The current findings regarding the existence of stably expressed genes in individuals over time might be useful for detecting unmanifested diseases, specifically, by observing the altered gene expression of these typically stable genes.





◀ **Figure 5.** Deconvolution analysis for the estimation of cell types in blood and variability in cell frequency over time. We used 10,590 genes for deconvolution analysis/digital cytometry with the analytical tool CIBERSORTx. Cell frequencies were obtained by using 22 reference immune cell types. Out of the 22 immune cell types investigated, we obtained (a) 14 cell types whose frequency remained unaltered and (b) 8 cell types whose frequency changed over time; the summation of the frequencies of the 22 immune cells corresponds to the total frequency of 100%. (a, b)  $n=61$  for the years 2008, 2012, and 2016. Grey boxes indicate the interquartile range (25–75% percentile range), the median is indicated by a horizontal line within the box, and the mean is indicated by the small white square; whiskers represent  $\pm 1$  SD from the mean. The Kruskal–Wallis test and post hoc Conover test were performed as statistical analysis and are shown under each relative graph.

Received: 4 June 2021; Accepted: 4 August 2021

Published online: 16 August 2021

## References

- Virella, G. *Medical Immunology* 7th edn. (CRC Press, 2019).
- Park, H. J. *et al.* Transcriptomic analysis of human IL-7 receptor alpha (low) and (high) effector memory CD8(+) T cells reveals an age-associated signature linked to influenza vaccine response in older adults. *Aging Cell* **18**, e12960. <https://doi.org/10.1111/ace1.12960> (2019).
- Preininger, M. *et al.* Blood-informative transcripts define nine common axes of peripheral blood gene expression. *PLoS Genet.* **9**, e1003362. <https://doi.org/10.1371/journal.pgen.1003362> (2013).
- Honda, M. *et al.* Differential gene expression profiling in blood from patients with digestive system cancers. *Biochem. Biophys. Res. Commun.* **400**, 7–15. <https://doi.org/10.1016/j.bbrc.2010.07.123> (2010).
- Komura, T. *et al.* Inflammatory features of pancreatic cancer highlighted by monocytes/macrophages and CD4+ T cells with clinical impact. *Cancer Sci.* **106**, 672–686. <https://doi.org/10.1111/cas.12663> (2015).
- Sakai, Y. *et al.* Common transcriptional signature of tumor-infiltrating mononuclear inflammatory cells and peripheral blood mononuclear cells in hepatocellular carcinoma patients. *Cancer Res.* **68**, 10267–10279. <https://doi.org/10.1158/0008-5472.CAN-08-0911> (2008).
- Sakai, Y. *et al.* Development of novel diagnostic system for pancreatic cancer, including early stages, measuring mRNA of whole blood cells. *Cancer Sci.* **110**, 1364–1388. <https://doi.org/10.1111/cas.13971> (2019).
- Sakai, Y. *et al.* Distinct chemotherapy-associated anti-cancer immunity by myeloid cells inhibition in murine pancreatic cancer models. *Cancer Sci.* **110**, 903–912. <https://doi.org/10.1111/cas.13944> (2019).
- Takamura, T. *et al.* Gene expression profiles in peripheral blood mononuclear cells reflect the pathophysiology of type 2 diabetes. *Biochem. Biophys. Res. Commun.* **361**, 379–384. <https://doi.org/10.1016/j.bbrc.2007.07.006> (2007).
- Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193. <https://doi.org/10.1093/bioinformatics/19.2.185> (2003).
- da Huang, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucl. Acids Res.* **37**, 1–13. <https://doi.org/10.1093/nar/gkn923> (2009).
- Sakai, Y. *et al.* Association of changes in the gene expression profile of blood cells with the local tumor inflammatory response in a murine tumor model. *Biochem. Biophys. Res. Commun.* **428**, 36–43. <https://doi.org/10.1016/j.bbrc.2012.10.004> (2012).
- Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdagh, P. & De Preter, K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat. Commun.* **11**, 5650. <https://doi.org/10.1038/s41467-020-19015-1> (2020).
- Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782. <https://doi.org/10.1038/s41587-019-0114-2> (2019).
- Alarcon, B. *et al.* The CD3-gamma and CD3-delta subunits of the T cell antigen receptor can be expressed within distinct functional TCR/CD3 complexes. *EMBO J.* **10**, 903–912 (1991).
- Rudd, C. E., Trevillyan, J. M., Dasgupta, J. D., Wong, L. L. & Schlossman, S. F. The CD4 receptor is complexed in detergent lysates to a protein-tyrosine kinase (pp58) from human T lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5190–5194. <https://doi.org/10.1073/pnas.85.14.5190> (1988).
- Carr, M. W., Roth, S. J., Luther, E., Rose, S. S. & Springer, T. A. Monocyte chemoattractant protein 1 acts as a T-lymphocyte chemoattractant. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 3652–3656. <https://doi.org/10.1073/pnas.91.9.3652> (1994).
- Xu, L. L., Warren, M. K., Rose, W. L., Gong, W. & Wang, J. M. Human recombinant monocyte chemotactic protein and other C-C chemokines bind and induce directional migration of dendritic cells in vitro. *J. Leukoc. Biol.* **60**, 365–371. <https://doi.org/10.1002/jlb.60.3.365> (1996).
- Nicholson, K. M. & Anderson, N. G. The protein kinase B/Akt signalling pathway in human malignancy. *Cell Signal* **14**, 381–395. [https://doi.org/10.1016/s0898-6568\(01\)00271-6](https://doi.org/10.1016/s0898-6568(01)00271-6) (2002).
- Rao, M., Zhu, Y., Cong, X. & Li, Q. Knockdown of CREB1 inhibits tumor growth of human gastric cancer in vitro and in vivo. *Oncol. Rep.* **37**, 3361–3368. <https://doi.org/10.3892/or.2017.5636> (2017).
- Nikoletopoulou, V., Markaki, M., Palikaras, K. & Tavernarakis, N. Crosstalk between apoptosis, necrosis and autophagy. *Biochim. Biophys. Acta.* **1833**, 3448–3459. <https://doi.org/10.1016/j.bbamcr.2013.06.001> (2013).
- Zhao, Y. *et al.* PI3K positively regulates YAP and TAZ in mammary tumorigenesis through multiple signaling pathways. *Mol. Cancer Res.* **16**, 1046–1058. <https://doi.org/10.1158/1541-7786.MCR-17-0593> (2018).
- Ignatiadis, M., Sledge, G. W. & Jeffrey, S. S. Liquid biopsy enters the clinic: Implementation issues and future challenges. *Nat. Rev. Clin. Oncol.* <https://doi.org/10.1038/s41571-020-00457-x> (2021).
- Gallucci, S. & Matzinger, P. Danger signals: SOS to the immune system. *Curr. Opin. Immunol.* **13**, 114–119. [https://doi.org/10.1016/s0952-7915\(00\)00191-6](https://doi.org/10.1016/s0952-7915(00)00191-6) (2001).
- Basu, M., Wang, K., Ruppin, E. & Hannehalli, S. Predicting tissue-specific gene expression from whole blood transcriptome. *Sci. Adv.* <https://doi.org/10.1126/sciadv.abd6991> (2021).
- Crowley, E., Di Nicolantonio, F., Loupakis, F. & Bardelli, A. Liquid biopsy: Monitoring cancer-genetics in the blood. *Nat. Rev. Clin. Oncol.* **10**, 472–484. <https://doi.org/10.1038/nrclinonc.2013.110> (2013).
- Dorsey, S. G. *et al.* Whole blood transcriptomic profiles can differentiate vulnerability to chronic low back pain. *PLoS ONE* **14**, e0216539. <https://doi.org/10.1371/journal.pone.0216539> (2019).
- Laing, E. E. *et al.* Blood transcriptome based biomarkers for human circadian phase. *Elife* <https://doi.org/10.7554/eLife.20214> (2017).
- Luo, Y. *et al.* Transcriptome profiling of whole blood cells identifies PLEK2 and C1QB in human melanoma. *PLoS ONE* **6**, e20971. <https://doi.org/10.1371/journal.pone.0020971> (2011).

30. Plaza-Flórida, A. *et al.* Distinct whole-blood transcriptome profile of children with metabolic healthy overweight/obesity compared to metabolic unhealthy overweight/obesity. *Pediatr. Res.* <https://doi.org/10.1038/s41390-020-01276-7> (2020).
31. McHugh, D. & Gil, J. Senescence and aging: Causes, consequences, and therapeutic avenues. *J. Cell Biol.* **217**, 65–77. <https://doi.org/10.1083/jcb.201708092> (2018).
32. Takashima, S. *et al.* Altered gene expression in T-cell receptor signalling in peripheral blood leucocytes in acute coronary syndrome predicts secondary coronary events. *Open Heart* **3**, e000400. <https://doi.org/10.1136/openhrt-2016-000400> (2016).
33. Falandry, C., Bonnefoy, M., Freyer, G. & Gilson, E. Biology of cancer and aging: A complex association with cellular senescence. *J. Clin. Oncol.* **32**, 2604–2610. <https://doi.org/10.1200/JCO.2014.55.1432> (2014).

## Acknowledgements

We thank Dr. Aiko Ubasawa for her technical assistance. We also thank Dr. Takeshi Urabe for his great assistance to enroll participants in this study.

## Author contributions

Y.S., K.M., S.Ka.: Conception and design, data analysis and interpretation, manuscript writing and review, final approval of manuscript. A.N.: Data analysis and interpretation, manuscript writing and review. Y.T., M.O.: Conception and design, collection and assembly of data, data analysis and interpretation. S.Ki.: Data analysis and interpretation. M.H., T.W., S.N., T.Tak., T.Tam.: Conception and design, data analysis and interpretation, final approval of manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-96078-0>.

**Correspondence** and requests for materials should be addressed to Y.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021