Portrayal of the human blood transcriptome of 3,388 adults and its relation to ageing and health

Schmidt Maria
Universitat Leipzig
Henry Loeffler-Wirth
Universitat Leipzig
Lydia Hopp
Universitat Leipzig
Arsen Arakelyan
National Academy of Sciences of the Republic of Armenia https://orcid.org/0000-0002-6851-1056
Holger Kirsten
Universitat Leipzig
Knut Krohn
Universitat Leipzig https://orcid.org/0000-0001-8110-1273
Ralph Burkhardt
Universitat Leipzig
Kerstin Wirkner
Universitat Leipzig
Joachim Thiery
Universitat Leipzig
Christoph Engel
Universitat Leipzig https://orcid.org/0000-0002-7247-282X
Markus Loeffler
Universitat Leipzig
Hans Binder (binder@izbi.uni-leipzig.de)
Universitat Leipzig https://orcid.org/0000-0002-2242-4678

Research article

Keywords: gene expression analysis, self-organizing maps machine learning, transcriptome types, blood cells, population study

Posted Date: December 20th, 2019

DOI: https://doi.org/10.21203/rs.2.19387/v1

Abstract

Background: The blood transcriptome is expected to provide a detailed picture of organism's physiological state with potential impact for applications in medical diagnostics and molecular and epidemiological research. We here present the first systematic analysis of blood specimen of 3,388 adult individuals collected in the Leipzig Research Center for Civilization Diseases, together with phenotype characteristics such as disease history, medication status, lifestyle factors and body mass index (BMI).

Methods: Multidimensional Self Organizing Maps-portrayal was applied to study transcriptional states on a population-wide scale. The method permits a detailed description and visualization of the molecular heterogeneity of transcriptomes and of their association with different phenotypic features.

Results: The diversity of transcriptomes is described by about one dozen modules of co-expressed genes of different functional context. We identify two major blood transcriptome types where type 1 accumulates more men, elderly and overweight people and it upregulates genes associating with inflammation and increased heme metabolism, while type 2 accumulates women, younger and normal weight participants and it associates with activated immune response, transcriptional, ribosomal, mitochondrial and telomere-maintenance cell-functions. We find a striking overlap of signatures shared by multiple diseases, ageing and obesity driven by an underlying common pattern, which reflects the increase in inflammatory processes.

Conclusions: Our portrayal framework provides a holistic view on the diversity of the human blood transcriptome. It provides a tool for comparative analyses of transcriptional signatures and of associated phenotypes in population studies and medical applications.

1. Background

Blood is the pipeline of human organism's physiology. The accessibility and minimal invasiveness during sampling made it the most feasible source in scientific research and clinical diagnostics as they could replace more invasive and risky tests [1]. Because of utility and simplicity, blood transcriptome investigations on genome-wide scales have gained in popularity over the past years. They were applied in medical context for characterizing diseases such as ischemic stroke [2], Alzheimer's disease [3], epilepsy [4], sepsis [5-8]; in pharmacogenomics [9] and marker search [10]; and also in epidemiological investigations on ageing [11], obesity status [12, 13], lifestyle factors such as smoking and alcohol consumption [14], special nutrition [15] and in immune system characterisation [16] (see [17] and references cited therein for a broad literature survey).

Although relatively simple to extract, blood is not a simple tissue. It accounts for about 8% of body weight and composed of acellular fluid (plasma) and a mixture of multiple cell types at different stages of their life cycles, lineage and function. On transcript level, blood is a complex mixture where changes in transcript abundance can be attributed to either alterations of transcriptional regulation in the different cell types or to relative changes in composition of cell populations. Computational cell type

deconvolution techniques such as CIBERSORT [18] (see [19] for an overview) attempt to deduce cellular composition from transcriptome data using statistical methodologies and cell-specific RNA-signatures, which are usually obtained independently in calibration experiments and are assumed to remain invariant. This assumption is imprecise because blood cells can change their state, e.g. via transformation from naïve to specialized cell types involving biological processes such as chemotaxis, surface signalling and cell-cell adhesion, all governed by expression changes of associated genes [16]. Therefore, cell subset classification is to a greater or lesser extent artificial, reflecting our current ability to distinguish cells based on specific sets of available markers [11]. Since blood flows throughout the whole body it implements functions by recirculating between central and peripheral lymphoid organs as well as to and from different tissues, sites of inflammation and/or injury. The cells specifically recapitulate the influence of genetic, epigenetic, cellular and environmental factors, which can vary between individuals and between their particular constitutions over time and age. In conclusion, the blood transcriptome is expected to provide a detailed picture of organism's physiological state with individual resolution. Most of the studies mentioned above focused on only one or a few diseases and/or environmental and lifestyle conditions and didn't answered the question whether the identified gene expression signatures are condition-specific or more widely applicable. Moreover, factors contributing to transcriptome variability among nominally healthy individuals are relatively unexplored so far.

We here present the first systematic analysis of the transcriptomes obtained from whole peripheral blood specimen of more than 3,000 adult individuals collected in LIFE (-adult), the Leipzig Research Center for Civilization Diseases, together with a large collection of phenotype characteristics such as disease history, medication status, lifestyle factors and body mass index (BMI) [9](Table S 1). Our analysis aims (i), at characterizing the inter-individual variability of the blood transcriptomes in terms of a suited classification scheme; (ii), at describing the diversity of transcriptome states using a collection of modules of co-expressed genes and at characterizing their biological functions; and (iii) at associating the transcriptome landscapes with a series of phenotype data collected in a participant-matched fashion. Overall, these issues are expected to provide a holistic view on essential properties of the blood transcriptome, a methodical framework for its analysis and its possible impact for future applications.

Previously, we have developed an omics 'portrayal' methodology based on self-organizing maps (SOM) machine learning [20, 21]. It has been applied to a series of data types and diseases [22-26], among them a study about footprints of pneumonia in the blood transcriptome [8]. SOM-portrayal takes into account the multidimensional nature of gene regulation and pursues a modular view on co-expression, reduces dimensionality and supports visual perception by delivering 'personalized', case-specific transcriptome portraits. They enable a straightforward and intuitive interpretation and mutual comparison of whole transcriptome landscapes between cases and classes. By applying this method to the blood transcriptomes of thousands of individuals we aim at demonstrating that multidimensional SOM-portrayal permits a detailed description and visualization of the molecular heterogeneity of transcriptional states and of their association with different phenotypes with potential impact for applications in medical diagnostics and molecular and epidemiological research. **Figure 1** provides a schematic overview of the design of this investigation.

Figure 1

2. Methods

2.1. LIFE-adult study and phenotype characteristics

The LIFE (-adult) research project conducted one of the largest population studies in Germany focusing on extensive phenotyping of urban individuals from Leipzig city [27]. It included more than 10,000 participants in order to discover the interplay between molecular, environmental and lifestyle factors and their impact for the health status of the population. The study was approved by the ethics board of the Medical Faculty of the University of Leipzig. In this publication we analyzed transcriptomic data of whole peripheral blood (WPB) samples, which were obtained from 3,388 adult participants of the LIFE-adult study. They roughly divide equally into women and men covering an age range between about 20 and 80 years with a strong bias towards elderly persons. The LIFE-adult study overall collected a broad survey of lifestyle and health items (see [27] for details). We made use of selected lifestyle characteristics of the participants such as smoking behaviour and alcohol consumption, medication according to ATCs (Anatomical Therapeutic Chemicals) indexing and disease history of the participants collected via questionnaires, blood count data from clinical laboratory including selected serum markers and body mass index (BMI) (Table S 1).

2.2. Blood transcriptome sampling, microarray measurements and data preprocessing

We made use of pre-processed gene expression data extracted from WPB samples of individuals as provided by the LIFE data base. Participant's recruitment, blood collection, storage and mRNA preparation, microarray measurements and primary data pre-processing was realized by different groups of the LIFE center [27]. WPB was collected in Tempus Blood RNA Tubes (ThermoFisher, Waltham, MA, USA) and stored at -80 °C until further processing. RNA was isolated and then hybridized to Illumina HT-12 v4 Expression BeadChips (Illumina, San Diego, CA, USA) and measured on an Illumina HiScan device. Raw probe level data were extracted using Illumina GenomeStudio and then further pre-processed including outlier and missing value removal, log-transformation, quantile normalization and centralization of the expression value of each gene using an in-house pipeline (Supplementary File 1: supplementary methods). The final transcriptome data consists of more than 48,000 probe IDs including the expression values of 19,049 genes for each of the individuals.

2.3. Self-organizing maps (SOM) transcriptome portrayal

Pre-processed expression values were analyzed using the oposSOM pipeline, available as R-package "oposSOM" [28]. It uses SOM neuronal network machine learning to translate the high-dimensional expression data of N= 19,049 gene transcripts into K=10,000 metagene expression data per individual [20, 29]. Each metagene represents a 'micro'-cluster of co-expressed genes showing mutually similar expression profiles across the samples. Metagenes were arranged in a 100x100 two-dimensional grid coordinate system and colored according to their expression level for each sample thus providing a 'personalized' image of the blood transcriptome of each individual studied (Figure S 1A). Mean portraits of transcriptome classes (see below) were calculated by averaging metagene expression values over all portraits of the respective group. Default color scale (red to blue for maximum to minimum expression, respectively) of the portraits uses log-expression values of the metagenes [20]. Diversity of sample portraits was visualized using a graph representation called 'correlation network' as implemented in 'oposSOM' [34].

2.4. 'Spot' clustering of co-expressed genes and stratification of samples

Metagenes of similar profiles cluster together forming 'spot-like' red and blue areas of over- and under-expression in the portraits due to the self-organizing properties of the SOM. Each of the spots represents a cluster of mutually correlated genes (Figure S 1B). The spots were detected using distance-metrics clustering, which determines each spot by a 'halo' of surrounding metagenes of local maximum distances to their neighbours [30]. The spot expression patterns obtained represents a characteristic fingerprint of each particular sample. The overall collection of spot-modules detected are major nodes of the co-expression network derived from the sample series (see the spot correlation and implication networks in Figure S 1B and C, respectively). Based on their transcriptome portraits we stratified the samples into appropriate groups. In a first step the portraits were divided into 33 so-called combinatorial pattern-types (cPATs), each defined by a certain unique combination of over-expressed spots [22] (Figure S 2A). Using the cPATs we estimated the tentative number of groups (Figure S 2A - C) and used them subsequently in a K-means clustering run, which stratifies the portraits into three major transcriptome types and nine subtypes (STs, Figure S 2A - C). The transcriptome strata were further characterized by detailed statistics about spot appearance (Figure S 2C - D).

2.5. Function mining

We applied gene set analysis to the lists of genes located in each of the spot modules to discover their functional context using right-tailed Fisher's exact test [36,37]. In addition, the gene set enrichment z-score (GSZ) was used to evaluate the impact of the gene sets in the different transcriptomic strata [32,38]. The GSZ-metrics considers the mean expression of the gene set normalized by its variance, i.e. it provides high values for homogeneous gene sets reflecting activation of biological functions with high relevance for the respective transcriptional states. Gene set maps complement this analysis by visualizing the

position of the gene of a set within the SOM grid. According to their degree of accumulation in or near the spots, one can deduce their potential functional context [20].

2.6. Phenotype portrayal

Phenotype information of the participants comprises their blood cell and marker counts, BMI and information about their lifestyle (smoking and alcohol consumption), medication and disease history (Table S 1). The enrichment of categorical phenotypic characteristics in each of the transcriptomic classes (types and subtypes) was estimated using one-tailed Fisher's exact test and visualized as enrichment heatmaps. Phenotype-to-metagene correlation maps were generated by correlating each of the phenotype parameter-profiles over all participants with each of the metagene expression profiles. The matrix of correlation coefficients obtained was then visualized in the SOM-grid as 'phenotype' portraits using a red-to-blue (maximum-to-minimum correlation) color-code. The metagene of maximum correlation coefficient was marked in the SOM-grid of a phenotype overview map. Expression of each of the spots was fitted using multiple regression with the phenotype values of the participants of each of the categories as variables. Standardized regression coefficients were then visualized as heatmaps.

3. Results

3.1. The blood transcriptome splits into three types

SOM analysis provides one portrait for each of the 3,388 LIFE-adult participant's WPB transcriptomes (Supplementary File 2 and Figure S 1A). For an overview we performed unsupervised similarity analysis based on pairwise comparisons of these expression portraits using Pearson's correlation coefficients and their visualization as pairwise similarity heatmap (**Figure 2**A). The samples split into three major types where type 1 and type 2 show pronounced anti-correlated expression portraits while type M forms an intermediate group. The network presentation reveals that WPB transcriptomes of type 1 and type 2 split into separate clusters while type M samples overlap between them (**Figure 2**B). The functional context of activated genes was estimated using gene set analysis (**Figure 2**A, part below). Type 1 associates with functional categories related to oxygen transport, heme metabolism, neutrophil accumulation and repressed chromatin states of T-cells while type 2 relates to immune response, transcriptional activity, T-cell accumulation and active chromatin states (see below). Males and females were represented in all types. Higher percentage of men was noted in type 1 (29% versus 19% for women) and reverse relation was found for type 2 (percentage of women: 37% versus 51%; **Figure 2**C). Type 1 is higher among elderly persons compared with type 2; however, in the latter, the age-dependence is different between women and men (**Figure 2**D). The composition of types for women changes virtually monotonously with a steadily increasing percentage of type 1 in contrast to men, who show a maximum of type composition in the age range of 50 – 55 years. Note also that the age dependence of type M more resembles that of type 1 than that of type 2 which suggests functional correspondence between types M and 1 (see below). The type-

composition of men and women is virtually independent of BMI (body mass index) except for very obese persons (BMI> 35 kg/m^2) which seem to accumulate more type 1 transcriptomes (**Figure 2**D).

Taken together, we identify two major blood transcriptome types and an intermediate type partly resembling type 1. Type 1 accumulates more men, elderly participants and it upregulates genes associating with inflammation and increased heme metabolism, while type 2 accumulates women and younger participants. It associates with activated immune response and transcriptional activity. The composition of types changes in a gender- and age-specific fashion.

Figure 2

3.2. A modular map of gene activation

Clusters of genes with correlated expression profiles appear as red spot-like areas in the transcriptomic portraits, which indicates their overexpression in the respective samples (Figure S 1A). Overall we identified 13 such major overexpression spots and labelled them with capital letters A – M (**Figure 3**A, for spot lists of genes see Table S 3 and Supplementary File 3). It roughly divides into two major areas containing spots predominantly upregulated either in type 1 (and partly also type M) or type 2 samples, respectively, and a third area with mixed spot assignment as illustrated by mean portraits of the transcriptomic types (**Figure 3**B), the spot profiles (**Figure 3**C and Figure S 3) and their correlation network (**Figure 3**D). Gene maps indicate the positions of genes taken from selected functional gene sets within the SOM grid of metagenes (**Figure 3**A). For example, genes upregulated in erythrocytes and platelets accumulate in spots C and N (up in type 1), respectively, while genes associated with mitochondrial function and RNA processing are found in spot E and G. Signature genes of T cells and of ribosomal function accumulate in and near spots I and J (up in type 2). Spot H accumulates the signature of CD4 cytotoxic T lymphocytes (CTLs) including the marker genes GZMA and PRF1, which were recently found to associate with extreme longevity [31]. Genes with function in interferon (IFN) response accumulate in spot L without preferential upregulation in one of the three types.

Typically, each of the individual sample portraits shows more than one spot, which reflects the parallel activation of different transcriptional programs and/or their mutual couplings. We subsume frequently observed combinations of expressed spots as so-called combinatorial pattern types (cPATs). Overall we identified 33 cPATs, which were then used to sub-stratify each of the major transcriptomic types into three subtypes (STs, annotated by 1.1, 1.2, 1.3, M.1, M.2, M.3 and 2.1, 2.2, 2.3, respectively) differing in their mean expression portraits (**Figure 3**D) and spot expression (**Figure 3**B and Figure S 2). Part of the spot profiles show marked expression differences between the STs (e.g. spots A, B, D, F) while others change continuously (e.g. spots H- J). Most of the spots upregulate either in type 1 or 2 samples. Interestingly, spot F enriching genes encoding ribosomal subunit S26 proteins shows a specific expression patterns with strong upregulation in part of STs without preference to either type 1 or type 2. Spot co-occurrence

analysis indicates that adjacent spots are often observed together, but also spots from different areas can co-occur, especially in samples of type M, which supports their intermediate position between type 1 and type 2. Part of the STs are dominated by samples expressing only one spot while others, especially of type M, show a broader distribution owing to more heterogeneous expression patterns (Figure S 2C). The sample similarity net indicates that most samples of the different STs accumulate into well localized clouds reflecting their mutual similarity (**Figure 3**E and Figure S 2F). The ST-composition is virtually age-independent except ST 1.1, which collects an increasing percentage of men and women at an age above 65 years (Figure S 4). In summary, the diversity of transcriptional states can be described by the combinatorics of about one dozen modules of co-expressed genes of different functional context which decompose each of the transcriptional types into three subtypes.

Figure 3

3.3. Footprints of functions: cellular programs, infections, telomeres and epigenetics

Next we performed functional analysis of the transcriptome strata using gene sets taken from the functional categories ‘biological process’ [33] (**Figure 4**A), ‘hallmarks of Cancer’ offering disease characteristics in a more general context [34](Figure S 6), ‘telomere maintenance’ [35] and ‘epigenetic states’ (**Figure 4**A-E). Profiling of these signatures splits them into two major clusters either upregulated in type 1 (marked with green color in the figures) or type 2 (apricot color), respectively. Type 2 associates, for example, with activation of cell cycling, MYC-target genes, oxidative phosphorylation (oxphos) while inflammation, hypoxia, coagulation, reactive-oxygen species and pathway signaling of TNFalpha-, TGFbeta-, PI3K-Akt-MTOR-, IL6-JAK-Stat3 activate in type 1. A third cluster (blue color) accumulates signatures related to interferon (IFN) response, which eventually suggests association with viral infections. We analyzed expression signatures derived recently to differentiate between bacterial and viral infections [36-41] (**Figure 4**B and C, respectively). The former signatures associate with the ‘inflammatory’ spots A, O and, also M, which upregulate in type 1 samples. In contrast, viral signature genes accumulate strongly in the IFN-response spot L, which is found upregulated in about 10% of all samples.

We are also interested in expression profiles of genes involved in telomere length maintenance (TM) via activation of telomerase. Mean telomere length in human leukocytes is negatively correlated with lifespan and BMI [42, 43] and it associates with heart diseases, type 2 diabetes, cancer [44-46], lifestyle factors [47], diet [48] and psychological stress [49]. TM-genes are more active in type 2 transcriptomes, which suggests that they stronger counteracts telomere shortening in younger (and healthier) individuals (**Figure 4**D). TM expression associates with cell cycle activity, starvation, oxidative stress, ageing, DNA-

methylation and other functions related to spots I and J this way indicating strong mutual coupling between TM and our transcriptome types (Figure S 7).

Next we analyzed the expression sets of genes assigned to distinct chromatin states in blood cells under healthy conditions, among them T-, B- and T-regulatory-cells [77] (**Figure 4**E and Figure S 8). States involving genes with active promoter (TssA) and completed transcription (Tx) and in repressed promoter states are expected to show high and low expression levels, respectively. This relation is indeed observed in type 2 transcriptomes, however it reverses in type 1. This reversal suggests de-repression of nominally repressed states and repression of active states in type 1 transcriptomes by epigenetic chromatin re-modelling. We recently demonstrated that differentiation and adjustment of cellular programs are governed by subtle cooperation of transcription factor (TF-) networks and epigenetics, e.g., via regulation of the polycomb repressive complex 2 (PRC2) and its targets[50]. We find that signatures, related to TF-networks regulate cell function requiring relatively high expression levels of their major regulatory genes such as cell cycle, oxphos and transcription predominantly in type 2 transcriptomes (Figure S 9). On the contrary, repressive epigenetic signatures related to PRC2 function, repressive histone (H3K27me3) marks and DNA-methylation antagonistically change compared with those of the TF-networks. Interestingly, these profiles show moderate and low expression levels according to the accumulation of their signature genes in the central region of the map. In summary, type 2 transcriptomes associate with cell cycle, oxphos-metabolism, telomere maintenance and immune system activity regulated mainly via transcription factor networks, which become repressed in type 1 transcriptomes in parallel with epigenetic de-repression of inflammatory cellular programs including responses to infections.

Figure 4

3.4. Previous gene expression signatures of the blood transcriptome

Modules of co-regulated genes of a previous blood transcriptome study [53] well agree with our spot clusters and further specify functional interpretation in terms of associated blood compounds such as cytotoxic plasma-, T- and B-cells (up-regulated in type 2) and erythrocytes, platelets, neutrophils and cells of myeloid lineage (up in type 1) (**Figure 3**A and Figure S 10). Another study extracted ageing signatures of the blood transcriptome [11]. Genes of decreasing expression (‘age_dn’) accumulate near spots I and J (up in type 2) while genes of increasing expression (age_up) are found in wider areas around spots A, M and H (up in type 1) (Figure S 11). This asymmetry of the numbers of spots suggests that age_up involves a more heterogeneous collection of molecular mechanisms than age_dn (see below). Another set of signatures was obtained recently in a study of the blood transcriptomes collected from patients of sepsis framed with CAP (community acquired pneumonia) [8](**Figure 4**F). These signatures surprisingly

correspond to signatures of nominally healthy individuals, e.g. patients with less severe CAP show signatures of type 2 transcriptomes, and while more severe CAP cases show type 1 transcriptomes associating partly with activation of inflammatory and endotoxin tolerance characteristics [8].

Next, we made use of a repertoire of 382 functionally annotated expression modules extracted from a recent meta-analysis of the blood transcriptomes of 16 disease and physiological states [52](**Figure 4**G and Figure S 12). Clustering of these signatures sub-stratifies them into three of type 1-like clusters which are strongly affected by spot O (C1 in **Figure 4**G), A (C2) or C (C3), respectively. Their profiles resemble those of the different severe CAP transcriptomes and can be interpreted as inflammatory signatures which are modulated by increased and decreased erythrocyte (spot C) and thrombocyte (spot N) activation patterns, respectively. Further, the 382 modules provide a rich repertoire of functional annotations, which support interpretation of our data (see example profiles in **Figure 4**G and Supplementary File 4 for the full set of profiles). For example, age_dn modules agree with DNA-methylation signatures in the blood. Methylation of CpG's in the promoters or enhancers upon ageing obviously repress transcription of the respective downstream gene (see also Figure S 1), which is in agreement with the finding that altered methylation sites enrich in ageing genes [11]. Moreover, we find strong enrichment of 91 of these modules in at least one of our spots (Figure S 12A). Hence, the spots provide a sort of basis set of co-regulated genes, which further expands into a rich collection of functional annotations of different categories via a multitude of combinations as considered by our cPATs (see above). Correlation analysis of different previous blood signature sets [8, 11, 52, 53] and our spot profiles provides very similar patterns in support of this view on the modular structure of the blood transcriptome (Figure S 13). In summary, comparison of previous blood signatures with our data shows that our spot-modules represent a sort of minimum set describing co-expression of the blood transcriptome. It expands into a rich collection of functional annotations including molecular mechanisms, cellular programs, cell types but also lifestyle factors, diseases and ageing effects.

3.5. Blood cell signatures and seasonal effects

Gene sets implemented in blood cell deconvolution algorithms such as Cibersort [18] show the characteristic correlation patterns observed also in the other blood signatures (compare **Figure 4**H and Figure S 14). They link the expression patterns of 22 blood cell types with our spot profiles. Elevated expression (and cell fractions, Figure S 15) of monocytes, neutrophils and eosinophils is observed in type 1 transcriptomes while overall expression of T- and B-cells upregulates in type 2. Expression of M1 macrophages and dendritic cells associate with the IFN-response signature (spot L). Furthermore, signatures of monocytes, M0 and M2 macrophages are also enriched in spot L, however in combination with the inflammatory spot O. Recent studies report seasonal changes of gene expression of the blood transcriptome and of blood cell counts [54, 55]. We find a slight shift of transcriptome characteristics towards type 1 in winter compared with summer both for men and women (Figure S 16). It is characterized by increased expression levels of inflammation (spot A) and erythrocyte expression (spot C) and counts and decreased levels of thrombocyte characteristics (spot N) and reticulocyte and

eosinophil counts (Table S 5). Overall, the seasonal changes of type compositions are relatively small (less than 3% in men and 1% in women) and are not explicitly considered further.

3.6. Phenotype portrayal: Blood cell counts, lifestyle, medication and disease history

Previous blood transcriptome studies also extracted gene signatures which associate with health-related features such as BMI (body mass index) and smoking status and also with the development of different diseases such as heart failure [56], dental caries [57], schizophrenia and neoplasms [52]. We find that they predominantly upregulate in type 1 transcriptomes showing characteristics of ageing and/or inflammation (Figure S 17). The LIFE-adult study provided a series of features characterizing health and lifestyle of the participants in terms of so-called phenotypes (Table S 1). We associated them with the blood transcriptomes in a participant-matched fashion using phenotype portraits, which typically show areas of positive (colored in red) and negative (in blue) correlation between phenotype features and expression profiles in the transcriptome landscape with metagene resolution (**Figure 5**A, and for details Figure S 19 - Figure S 23). For example, phenotype associations with expression patterns of type 1 (red in the lower left part of the map) or type 2 (red in the upper right part) can be distinguished. In addition, overview maps were generated for each of the phenotype categories, which mark the metagene of maximum (and minimum) correlation for each of the phenotypes studied. Enrichment of phenotypes is evaluated in terms of the distribution of cases among the transcriptome types (**Figure 5**C, for enrichment significance evaluation using Fishers exact test see Figure S 19D - Figure S 23D).

We find that most blood count data correlate either with type 1 (e.g. erythrocytes, reticulocytes, platelets, neutrophils) or type 2 (lymphocytes) transcriptomes in agreement with the blood cell transcriptomes analyzed above. Smokers, alcohol consumers (> 30 g/day), obese and elderly people, men and participants taking different categories of medication according to the ATC (Anatomical Therapeutic Chemicals) classification and also participants with different self-reported lifetime diseases show preferences for type 1 (and partly type M) transcriptomes while younger, under- and normal-weight participants, women and non-consumers of medication associate preferentially with type 2. The degree of correlation with metagene expression is markedly higher for blood counts compared with the other phenotypes (**Figure 5**C).

Part of the blood count portraits indicate fingerprint-like correlation patterns specific for the different blood compounds (**Figure 5**A, B, Figure S 18, Figure S 19 and **Figure 4**H). The portraits of the phenotypes of the other categories partly resemble those of blood counts, this way reflecting close association between them. For example, the ‘ageing’ portrait (visualizing the correlation between age and transcriptome) can be understood as superposition of the red blood cell (RBC)- and neutrophil (NE)-phenotype portraits indicating the increased levels of RBC and NE in elderly people (see next subsection). The ‘alcohol consumption’ portrait also resembles the RBC-portrait while smoking reveals an eosinophil (EO)-like patterns. Part of the medication and disease history portraits can be interpreted similarly

reflecting, e.g., that part of medications and diseases are more prevalent in elderly people (see the mean age data of each of the phenotypes listed in Table S 1) and consequently associate with increased RBC- and NE-levels and decreased lymphocyte (LY)-counts (Figure S 22-Figure S 23).

Other phenotype portraits, e.g. those of different age ranges (see next subsection) and of different medications, cannot be simply interpreted as composite of blood count portraits. For a more detailed view we performed correlation and multiple regression analysis to estimate the particular effect of phenotypes on spot expression (part C – F in Figure S 19 - Figure S 23). We find a close relationship between high correlation coefficients and significant contributions of phenotype-coefficients (log p <-6) especially for spots located in the lower left and upper right corners of the map. These refer first of all to age, obesity, gender, RBC and white blood cell (WBC) counts and LY, medications of the groups C (cardiovascular system) and B (blood forming organs) and the previous diseases HL (hyperlipidemia), DIA (diabetes), HT (hypertension) and CAN (cancer).

In summary, phenotype portrayal visualizes fine structures of the effect of health and lifestyle factors on the blood transcriptome. They reflect alterations of blood cell composition and presumably also the specifics of the transcriptional programs activated in the different cells. The transcriptome types (and subtypes) resolve the heterogeneity of blood transcriptomes while the spot modules provide a metrics for its quantification. Overall, the phenotype portraits enable an intuitive, perception-based interpretation in terms of function and mutual associations between the different features.

Figure 5

3.7. Portrayal of ageing, obesity and of serum markers

Ageing and alterations of the BMI are accompanied by changes of the composition of transcriptome types in a gender-specific fashion (**Figure 2**D). Functional analysis shows that expression of type 1_up transcriptomes gains with age while expression of type 2_up decays on average (see the plots with age-ranked samples in Figure S 5 - Figure S 11). Plots of spot expression as a function of age and BMI reveal further details (**Figure 6**A): Spot expressions related to red blood (spot C) and platelet (spot N) characteristics increase as a function of age and BMI with differences between the mean LOESS-curves for men and women (compare the red and blue curves) in correspondence with the blood count data (Figure S 18). In turn, the expression curves of spots related to immunity (I and J) decay with age and BMI in a nearly sex-independent fashion. On the other hand, the curves show similar courses at different levels for the transcriptomic types which suggests type-independent ageing mechanisms. The ageing curves are partly non-linear where the slopes get steeper for ages above 55- 60 years (e.g. for spot A and I, indicative for inflammation and immune response, respectively) or above 65- 70 years (spot L, IFN response), which suggests altered mechanisms in elderly people above different age thresholds.

Gene maps of previous ageing signatures [11] reveal an asymmetrical distribution of ageing_up and ageing_dn genes (**Figure 6**B). The latter ones accumulate within a narrow area in and around spots I and J in the right upper corner of the map giving rise to strong correlation between signatures' expression and that of these spots. Deactivation of associated cellular functions such as immune response, telomere maintenance and/or ribosomal and mitochondrial activities with age obviously proceeds homogenously, presumably driven via mechanisms such as DNA hyper-methylation (Figure S 12). In contrast, ageing_up genes distribute much more heterogeneously between different spot-regions where each of them shows a specific profile of expression gaining with age (see curves of spots A, O, N, M, L and H in **Figure 6**A). Ageing is obviously accompanied or even driven by the activation of a multitude of inflammatory mechanisms involving different molecular and cellular components (see spot characteristics), which combine in a patient specific fashion giving rise to a relatively heterogeneous ageing_up signature.

The mean ageing portrait ('all ages' in **Figure 6**C) corresponds to the distribution of ageing_up and ageing_dn genes of the ageing signature [11] (compare the respective gene set maps with the red and blue areas in **Figure 6**B, respectively). Moreover, the ageing portrait can be roughly interpreted by the superposition of increasing RBC- and NE-like (positive correlation in red) and decaying LY-like (negative correlation in blue) contributions (compare with the cell count portraits in **Figure 6**E) in agreement with the increase/decrease of the expression of the respective landmark spots C, O and I, J, respectively. Inspection of gender- and age (decade)-stratified portraits reveals that elderly women and men (> 60 years) are similarly affected by an increase of NE- and IFN-related (found especially for subtype M.3) characteristics while the RBC-like pattern (typical for subtype 1.3) is more pronounced for mid-aged men (40 – 60 years). The mean BMI-portrait ('all BMI' in **Figure 6**D) shows characteristics of type 1 transcriptomes without the NE-like patterns and the elevated expression of spot L observed in the respective ageing portrait. Interestingly, the BMI-stratified portraits 'switch' from type 2 into type 1 for obese women and men (BMI> 30 kg/m^2), due to gained (positive) correlations between BMI and inflammatory (spot A), RBC- (spot C) and platelet (spot N) characteristics, on one hand, and decaying immune response (spots I, J) expression signatures on the other one. Interestingly, this behaviour possibly associates with the so-called obesity-paradox claiming that an intermediate BMI about 25 kg/m^2 associates with minimum health risk [58] and thus with a switch from positive to negative effect of increasing BMI on health.

For further comparison, we generated phenotype (correlation) portraits of four selected serum protein markers (**Figure 6**E). The portraits of hsCRP (human serum C-reactive protein) and of cytostatin C reflect footprints of inflammation (spot O) and IFN-response (spot L) in the blood transcriptome associated with NE-like patterns of the blood counts. The portrait of ferritin closely resembles that of RBC reflecting correspondence between the level of stored iron and erythrocyte expression (spot C). The transferrin portrait reveals a different patterns associating with diminished spots O (inflammation) and especially L (IFN-response) and enhanced spot N (thrombocytes), possibly due to the role of platelets in iron transport [59]. In summary, ageing and obesity associate with characteristic alterations of the blood transcriptome

reflecting a fine interplay between inflammatory and iron physiology as mediated by molecular (as IFN-response), cellular (e.g., WBC and RBC) and serum protein compounds.

Figure 6

4. Discussion

We ‘portrayed’ the diversity of the blood transcriptome of a cohort of more than 3,000 nominally healthy adult individuals included into the Leipzig Health Study ‘LIFE-adult’ in terms of intuitive SOM-images and classified them into three major transcriptome types. The expression patterns were decomposed into a minimum set of modules of co-regulated genes. Their functional impact was interpreted based on previous knowledge including the results of previous blood transcriptome studies. Finally, we associated the blood transcriptomes with a series of phenotype-features collected in LIFE-adult for the same participants such as age, obesity-status, blood cell count, disease history and medication by means of phenotype portraits. Overall our study provides a comprehensive characterization of the blood transcriptome taking into account the whole spectrum of transcriptional states on a population-wide scale.

4.1. SOM-portrayal reduces dimensions of the blood transcriptome

Dimension-reduction and feature extraction are important issues in high-throughput data analysis. Our machine-learning approach reduces the dimensionality of data (number of individuals x (number of genes + number of phenotypes)) into a handful transcriptome types and subtypes [60]. Their expression is governed by about one dozen expression (spot-) modules in close correspondence with a previous modularization of the blood transcriptome [53]. Moreover, data portrayal transforms high-dimensional data landscapes into easy-to-interpret images. Their visual inspection strongly supports analytic tasks on different levels of stratification ranging from individual ‘personalized’ to subtype- and type-averaged expression portraits. Our study thus provided a sort of album of transcriptomic ‘faces’ of the LIFE-participants (Supplementary File 2). Importantly, the phenotype portrayal projects low dimensional features such as age or BMI onto the high-dimensional transcriptome landscape, which generates highly granular correlation images serving as ‘fingerprint’ of the respective phenotype. Their mutual comparison helps to identify associations between them and also singular patterns.

4.2. Typing reveals parallels between a series of biological functions and health phenotypes

The tree in **Figure 7**A illustrates the similarity relations between the subtype portraits. It reveals a virtually linear arrangement of subtypes along the backbone. The portraits at the left and right margins (type 1-versus-type 2) differ mainly in antagonistic expression of genes located in opposite corners of their portraits. Our analysis thus uncovered a striking simplicity of the transcriptome. It reflects a characteristic alteration of cell components, namely a decrease in signatures of myeloid-lineage cells and increase in signatures of lymphocytes from the left to the right. This basic pattern is superimposed by transcriptional footprints of erythrocyte and thrombocytes giving rise, e.g., to gender-specific differences; of cytotoxic CTL's playing role in longevity and by patterns reflecting interferon-response, whose amplitude increases, on average, in elderly people (especially above 65 years). The transcriptional (spot-) modules diversify these basic patterns in a subtype-specific fashion (**Figure 7**C).

Transcriptome typing and modularization also paved an avenue to describe the effect of age and BMI on the blood transcriptome, and in a wider context, on human's physiology. For example, the percentage of type 1 transcriptomes in the population reflecting more inflammatory characteristics gains with age and, to a less degree, BMI in a non-linear, gender-specific fashion. In agreement with previous results [61] we find a striking overlap of signatures shared by multiple diseases, ageing and obesity driven by an underlying common pattern. It reflects the increase of inflammation and decrease of immune-responsibility (**Figure 7**B). Obesity is associated with leukocytosis and it may be regarded as a state of chronic low-grade inflammation [13, 62], which, in turn is considered a driver of many age-related disorders (inflammaging) [63].

Also mean leukocyte telomere length associates with age-related diseases [44, 45]. Telomeres are protective nucleoprotein structures that cap the ends of linear chromosomes. They shorten a bit after each cell division. In consequence, telomere length reflects cell ageing [64]. Telomere maintenance mechanisms counteract this process and thus their activation can be indicative for counteracting cell ageing [65, 66]. Expression of genes involved in telomerase-maintenance pathway [67] directly correlate with T-cell and immune response signatures suggesting that cell's ability to maintain telomeres associates with better immune responsibility and overall health constitution observed especially in younger, non-obese, non-smoking and non-alcohol consuming people. Interestingly, decay of telomere length with age [68] partly resembles the decay of the amount of type 2 transcriptomes with age and women having a higher fraction of type 2 transcriptomes with activated telomere maintenance mechanisms possess on average longer telomeres than men [68, 69].

4.3. Parallels between health and disease and asymmetry of transcriptomic changes

Our recent study on sepsis framed by community acquired pneumonia [8], a high-grade inflammatory disease, identified three major axes of the variation of the blood transcriptome, namely the inflammatory axis (endotoxin tolerance, cytotoxic cells), a 'blood-disturbance' axis including mostly erythrocyte and thrombocyte characteristics and the IFN-response axis. These axes of variation were also found in the

blood transcriptomes of nominally heathy subjects, however with a decreased amplitude of inflammatory expression changes. Part of our modules continuously change along the subtypes (e.g. I and J: immune response; H: cytotoxic cells), others show subtype-specific activations (C, N; erythrocytes, platelets), while a third category spreads over almost all subtypes (L; IFN-response). Interestingly, also the sepsis signatures split into three type 1_up and only one type 2_up mechanism (**Figure 4**). They reflect a similar asymmetry of the heterogeneity of mechanisms as found between ageing_up and ageing_dn signatures (see above). It reflects multi-factorial activation mechanisms accompanying ageing and/or disease development (see [11] and Figure S 11). Also the phenotype portraits of ageing, obesity, disease history, medication, alcohol consumption and smoking reflect this asymmetry because the majority of them associates with type 1 showing more heterogeneous patterns than type 2 portraits.

We estimated ageing and obesity trajectories of spot expression by non-linear fits through our cross sectional data where stratification by types (and gender) reduces the variance of data points (**Figure 6**A). A recent longitudinal study revealed that individuals are more similar to their own expression profiles later in life than profiles of other individuals their own age [70]. Type-strata of the blood transcriptome thus eventually more adequately describe ageing effects. Longitudinal follow-up studies over different age ranges are required to study individual 'life-courses' of the blood transcriptome and their impact for lifetime-risk prediction.

4.4. Epigenetic background and ageing clocks

Our analysis also underlines the importance of epigenetic mechanisms, particularly of chromatin (re-) organization for changes the blood transcriptome. Using gene expression of nominally repressed and activated chromatin states as an indicator of gene activity we find a pronounced mutual switching between type 1 and type 2 transcriptomes suggesting that active states in type 2 become repressed in type 1 and that repressed states in type 2 become activated in type 1. Differences of biological mechanisms behind transcriptomic [11] and the epigenetic predictors [71] of ageing or disease progression are not completely clear [11]. We recently reported that transcriptomic and epigenetic mechanisms partly decouple in cancer development [72, 73] and cell differentiation [50]. Thus, the expression changes observed might reflect changed chromatin organization leading to altered cell function in type 1 compared with type 2 as discussed, e.g., as epigenetic mechanisms accompanying ageing [74] and inflammation [75-78] and associating with changes of DNA-methylation and histone-marks governing gene activity.

The search for reliable indicators of biological age, rather than chronological age, attracted large efforts in the last decades [79, 80]. DNA methylation-derived epigenetic clocks are currently better in estimating chronological age than transcriptomic or telomere length measures [81]. Currently it is not entirely clear how molecular clocks work, what aspect(s) of physiological or cellular aging they represent and whether age-related changes, such as telomere shortening or DNA-methylation contribute to the causes of ageing or are the results of it. There is only weak correlation between DNA-methylation and transcriptome age

predictors meaning that the transcriptomic age and the epigenetic clock describes different aspects of biological aging [11]. DNA-methylation clock was assumed to reflect the function of the epigenetic maintenance system [71]. In support of this we found that DNA-methylation maintenance methyltranferase DNMT1 is part of spot gene cluster J (Table S 3) showing decaying expression with age (age_dn signature of [11]) and correlating with DNA-methylation signatures (Figure S 12). Hypomethylation accumulates with the number of cell divisions, due to insufficient re-methylation, and also inflammatory factors that increase cell turnover led to increased methylation loss, which in turn, disrupt DNA binding patterns of transcription factors and modify their regulatory role in cell function as discussed above. Coupled transcription and DNA-methylation epidemiological studies are required to better disentangle the relation between epigenetic and transcriptomics of the blood [81].

Figure 7

5. Conclusions

We have characterized the human blood in terms of a relatively simple scheme of transcriptome types and functional gene modules, which associate with health-, lifestyle and age-related phenotypes. The large sample size of our study and wide agreement with state of the art blood-transcriptome research let us conclude that our scheme reflects basic gene expression characteristics of the human blood in a wide spectrum of conditions. It provides interpretations for underlying biological mechanisms on cellular and molecular levels. It has impact for future applications for diagnosis and prognosis via the refinement of existing and the development of novel predictors for age, lifestyle and disease outcomes. The data provides a potential basis for the establishment of reference ranges for expression values and the identification of suited transcriptome candidate markers such as the expression of single genes or of sets of them, which can be derived from our spot-modules. The individual portrayal of transcriptomes and of their associations with phenotype features in terms of easy-to interpret images offers perspectives for visual-perception based personalized diagnostics. Large scale longitudinal studies and paired transcriptome-epigenome investigations are needed to better understand lifetime courses, causal relations and mechanisms of (epi-)genomic regulation.

6. Abbreviations

A	Alimentary tract and metabolism
AP	Angina pectoris
ALC > 30	Participants consuming more than 30g alcohol per day
ALC ≤ 30	Participants consuming ess than 30g alcohol per day

ALC	Alcohol consumption
ART	Arthrosis
AST	Asthma
ATC	Anatomical Therapeutic Chemical Classification System of medication
B	Blood and blood forming organs
BA	Basophils absolute (10^9/l)
BAP	Basophils (%)
BD	Sepsis type „Blood disturbant“
BMI	Body mass index
BP	GO-term „Biological process“
C	Cardiovascular system
CAN	Cancer
CAP	Community Aquired Pneumonia
CAT	Cataract
cPAT	Combinatorial pattern type
CTL	Cytotoxic T lymphocytes
D	Dermatologics
DEP	Depression
DIA	Diabetes
DNMT1	DNA-methylation maintenance methyltranferase
EO	Eosinophils absolute (10^9/l)
EOP	Eosinophils (%)
EXSMO	Ex-smoker
G	Genitourinary system and sex hormones

GLA	Glaucoma
GO	Gene Ontology
GOU	Gout
GSZ	Gene set enrichment z-score
H	Systemic hormonal preparations, excl. sex hormones and insulins
HA	Heart attack
HCT	Hematocrit (l/l)
HEP	Hepatitis
HGB	Hemoglobin (SI units, mmol/l)
HGBK	Hemoglobin (conv. units, g/dl)
HL	Hyper-lipidaemia
HS	Sepsis type „High severety“
hsCRP	human serum C-reactive protein
HT	Hypertension
HZO	Hzoster
IFN	Interferon
J	Anti-infective for systemic use
L	Antineoplastic and immunomodulating agents
LIFE(-adult)	Leipzig Research Center for Civilization Diseases
LOESS	locally estimated scatterplot smoothing
LS	Sepsis type „Healthy & Low severety“
LY	Lymphocytes absolute (10^9/l)
LYP	Lymphocytes (%)
M	Muscular-skeletal system

MCH	Mean corpuscular hemoglobin (SI units, fmol)
MCHC	Mean corpuscular hemoglobin concentration (SI units, mmol/l)
MCHCK	Mean corpuscular hemoglobin concentration (conv. units, g/dl)
MCHK	Mean corpuscular hemoglobin (conv. units, pg)
MCV	Mean corpuscular volume (fl)
MO	Monocytes absolute (10^9/l)
MOP	Monocytes (%)
MPV	Mean platelet volume (fl)
MS	Sepsis type „Medium severety“
N	Nervous system
NE	Neutrophils absolute (10^9/l)
NEP	Neutrophils (%)
NONSMO	Non-smoker
nwt	Normal weight
ob	Obese
P	Antiparasitic products, insecticides and repellents
PLT	Platelets (10^9/l)
PRC2	Polycomb repressive complex 2
pre-ob	Pre-obese
R	Respiratory system
RBC	Erythrocytes (10^{12}/l)
RETI	Reticulocytes (/1000)
RHE	Rheuma
S	Sensory organs

SEP	Sepsis
SMO	Smoker
SMO	Smoking
SOM	Self-organizing maps
ST	Subtype
TF	Transcription factor
THY	Thyroid
TM	Telomere length maintenance
TssA	Genes with active promoter
Tx	Genes with completed transcription
uwt	Underweight
V	Various
WBC	Leucocytes (10^9/l)
WPB	Whole peripheral blood

7. Declarations

7.1. Ethics approval and consent to participate

The LIFE-adult study was approved by the ethics board of the Medical Faculty of the University of Leipzig. Written consent was obtained from all participants of LIFE-adult.

7.2. Consent to publish

Written consent of the participants to publish results of LIFE-adult was obtained.

7.3. Availability of data and materials

The data that support the findings of this study are available from the LIFE centre but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of LIFE. Secondary data are available from 'The Leipzig Health Atlas' repository (https://www.health-atlas.de/; accession number will be available after acceptance of the manuscript).

7.4. Competing interests

All authors declare that they have no competing interests.

7.5. Funding

This publication was supported by LIFE – Leipzig Research Center for Civilization Diseases, Leipzig University funded by means of the European Social Fund, the Free State of Saxony. This work was further supported by the Federal Ministry of Education and Research (BMBF) i:DSem project Leipzig Health Atlas (www.health-atlas.de, to HB, HLW, ML), the collaborative projects with Armenia PathwayMaps (WTZ ARM II-010 and 01ZX1304A, to HB and AA) and oBIG (FFE-0034, to HL-W) and the Systems Biology programme project CapSys (to HB, LH, ML).

7.6. Authors' contributions

Conceived and wrote this paper: MS, HB; Performed analysis: MS, LH, HB, HLW; downstream analysis methods development: HLW, MS, AA; preprocessing of transcriptomics data:HK; collection and curation of phenotype data: KW, CE, RB, KK; Coordinators of LIFE research centre: ML, JT. All authors read and approved the final manuscript.

7.7. Acknowledgements

Not applicable

8. References

1. Sohn E: **Diagnosis: Frontiers in blood testing**. *Nature* 2017, **549**:S16.
2. Baird AE, Soper SA, Pullagurla SR, Adamski MG: **Recent and near-future advances in nucleic acid-based diagnosis of stroke**. *Expert Review of Molecular Diagnostics* 2015, **15**(5):665-679.
3. Rembach A, Ryan TM, Roberts BR, Doecke JD, Wilson WJ, Watt AD, Barnham KJ, Masters CL: **Progress towards a consensus on biomarkers for Alzheimer's disease: a review of peripheral analytes**. *Biomarkers in Medicine* 2013, **7**(4):641-662.
4. Karsten SL, Kudo LC, Bragin AJ: **Use of peripheral blood transcriptome biomarkers for epilepsy prediction**. *Neurosci Lett* 2011, **497**(3):213-217.
5. Scicluna BP, van Vught LA, Zwinderman AH, Wiewel MA, Davenport EE, Burnham KL, Nürnberg P, Schultz MJ, Horn J, Cremer OL *et al*: **Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study**. *The Lancet Respiratory Medicine* 2017, **5**(10):816-826.
6. Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, Rautanen A, Gordon AC, Garrard C, Hill AVS *et al*: **Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study**. *The Lancet Respiratory Medicine* 2016, **4**(4):259-271.
7. Burnham KL, Davenport EE, Radhakrishnan J, Humburg P, Gordon AC, Hutton P, Svoren-Jabalera E, Garrard C, Hill AVS, Hinds CJ *et al*: **Shared and Distinct Aspects of the Sepsis Transcriptomic Response to Fecal Peritonitis and Pneumonia**. *American Journal of Respiratory and Critical Care Medicine* 2017, **196**(3):328-339.
8. Hopp L, Loeffler-Wirth H, Nersisyan L, Arakelyan A, Binder H: **Footprints of Sepsis Framed Within Community Acquired Pneumonia in the Blood Transcriptome**. *Frontiers in Immunology* 2018, **9**(1620).
9. Burczynski ME, Dorner AJ: **Transcriptional profiling of peripheral blood cells in clinical pharmacogenomic studies**. *Pharmacogenomics* 2006, **7**(2):187-202.
10. Hanash SM, Baik CS, Kallioniemi O: **Emerging molecular biomarkers—blood-based strategies to detect and monitor cancer**. *Nature Reviews Clinical Oncology* 2011, **8**(3):142-150.
11. Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN, Powell J, Reinmaa E, Sutphin GL, Zhernakova A, Schramm K *et al*: **The transcriptional landscape of age in human peripheral blood**. *Nature Communications* 2015, **6**:8570.
12. Homuth G, Wahl S, Müller C, Schurmann C, Mäder U, Blankenberg S, Carstensen M, Dörr M, Endlich K, Englbrecht 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 Medical Genomics* 2015, **8**:65.
13. Johannsen NM, Priest EL, Dixit VD, Earnest CP, Blair SN, Church TS: **Association of white blood cell subfraction concentration with fitness and fatness**. *British Journal of Sports Medicine* 2010, **44**(8):588-593.
14. Dumeaux V, Olsen KS, Nuel G, Paulssen RH, Børresen-Dale A-L, Lund E: **Deciphering Normal Blood Gene Expression Variation—The NOWAC Postgenome Study**. *PLoS Genet* 2010, **6**(3):e1000873.

15. Burton KJ, Pimentel G, Zangger N, Vionnet N, Drai J, McTernan PG, Pralong FP, Delorenzi M, Vergères G: **Modulation of the peripheral blood transcriptome by the ingestion of probiotic yoghurt and acidified milk in healthy, young men**. *PLOS ONE* 2018, **13**(2):e0192947.
16. Chaussabel D, Pascual V, Banchereau J: **Assessing the human immune system through blood transcriptomics**. *BMC Biology* 2010, **8**(1):84.
17. Chaussabel D: **Assessment of immune status using blood transcriptomics and potential implications for global health**. *Seminars in Immunology* 2015, **27**(1):58-66.
18. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles**. *Nature Methods* 2015, **12**:453.
19. Finotello F, Rieder D, Hackl H, Trajanoski Z: **Next-generation computational tools for interrogating cancer immunity**. *Nature Reviews Genetics* 2019.
20. Wirth H, Löffler M, von Bergen M, Binder H: **Expression cartography of human tissues using self organizing maps**. *BMC Bioinformatics* 2011, **12**(1):306.
21. Binder H, Wirth H: **Analysis of large-scale OMIC data using Self Organizing Maps**. In: *Encyclopedia of Information Science and Technology Third Edition.* Edited by Khosrow-Pour M: IGI global; 2014: 1642-1654.
22. Loeffler-Wirth H, Kreuz M, Hopp L, Arakelyan A, Haake A, Cogliatti SB, Feller AC, Hansmann M-L, Lenze D, Möller P *et al*: **A modular transcriptome map of mature B cell lymphomas**. *Genome Medicine* 2019, **11**(1):27.
23. Nikoghosyan M, Hakobyan S, Hovhannisyan A, Loeffler-Wirth H, Binder H, Arakelyan A: **Population Levels Assessment of the Distribution of Disease-Associated Variants With Emphasis on Armenians – A Machine Learning Approach**. *Frontiers in Genetics* 2019, **10**(394).
24. Bilz NC, Willscher E, Binder H, Böhnke J, Stanifer ML, Hübner D, Boulant S, Liebert UG, Claus C: **Teratogenic Rubella Virus Alters the Endodermal Differentiation Capacity of Human Induced Pluripotent Stem Cells**. *Cells* 2019, **8**(8):870.
25. Kunz M, Löffler-Wirth H, Dannemann M, Willscher E, Doose G, Kelso J, Kottek T, Nickel B, Hopp L, Landsberg J *et al*: **RNA-seq analysis identifies different transcriptomic types and developmental trajectories of primary melanomas**. *Oncogene* 2018, **37**(47):6136-6151.
26. Hopp L, Willscher E, Wirth-Loeffler H, Binder H: **Function Shapes Content: DNA-Methylation Marker Genes and their Impact for Molecular Mechanisms of Glioma**. *Journal of Cancer Research Updates* 2015, **4**(4):127-148.
27. Loeffler M, Engel C, Ahnert P, Alfermann D, Arelin K, Baber R, Beutner F, Binder H, Brähler E, Burkhardt R *et al*: **The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany**. *BMC Public Health* 2015, **15**:691.
28. Löffler-Wirth H, Kalcher M, Binder H: **oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on bioconductor**. *Bioinformatics* 2015, **31**(19):3225-3227.
29. Wirth H, von Bergen M, Binder H: **Mining SOM expression portraits: Feature selection and integrating concepts of molecular function** *BioData Mining* 2012, **5:18**.

30. Vesanto J: **SOM-based data visualization methods**. *Intelligent Data Analysis* 1999, **3**(2):111-126.
31. Hashimoto K, Kouno T, Ikawa T, Hayatsu N, Miyajima Y, Yabukami H, Terooatea T, Sasaki T, Suzuki T, Valentine M *et al*: **Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians**. *Proc Natl Acad Sci USA* 2019, **116**(48):24242-24251.
32. Hopp L, Wirth H, Fasold M, Binder H: **Portraying the expression landscapes of cancer subtypes: A glioblastoma multiforme and prostate cancer case study**. *Systems Biomedicine* 2013, **1**(2):99-121.
33. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *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* 2005, **102**(43):15545-15550.
34. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov Jill P, Tamayo P: **The Molecular Signatures Database Hallmark Gene Set Collection**. *Cell Systems* 2015, **1**(6):417-425.
35. Barthel FP, Wei W, Tang M, Martinez-Ledesma E, Hu X, Amin SB, Akdemir KC, Seth S, Song X, Wang Q *et al*: **Systematic analysis of telomere length and somatic alterations in 31 cancer types**. *Nat Genetics* 2017, **49**:349.
36. Coates PJ, Rundle JK, Lorimore SA, Wright EG: **Indirect Macrophage Responses to Ionizing Radiation: Implications for Genotype-Dependent Bystander Signaling**. *Cancer Research* 2008, **68**(2):450-456.
37. Foster SL, Hargreaves DC, Medzhitov R: **Gene-specific control of inflammation by TLR-induced chromatin modifications**. *Nature* 2007, **447**:972.
38. Néemeth ZH, Leibovich SJ, Deitch EA, Vizi ES, Szabó C, Haskó G: **cDNA Microarray Analysis Reveals a Nuclear Factor-κB-Independent Regulation of Macrophage Function by Adenosine**. *Journal of Pharmacology and Experimental Therapeutics* 2003, **306**(3):1042-1049.
39. Pena OM, Hancock DG, Lyle NH, Linder A, Russell JA, Xia J, Fjell CD, Boyd JH, Hancock REW: **An Endotoxin Tolerance Signature Predicts Sepsis and Organ Dysfunction at Initial Clinical Presentation**. *EBioMedicine* 2014, **1**(1):64-71.
40. Sweeney TE, Wong HR, Khatri P: **Robust classification of bacterial and viral infections via integrated host gene expression diagnostics**. *Science Translational Medicine* 2016, **8**(346):346ra391-346ra391.
41. Andres-Terre M, McGuire Helen M, Pouliot Y, Bongen E, Sweeney Timothy E, Tato Cristina M, Khatri P: **Integrated, Multi-cohort Analysis Identifies Conserved Transcriptional Signatures across Multiple Respiratory Viruses**. *Immunity* 2015, **43**(6):1199-1211.
42. Gielen M, Hageman GJ, Antoniou EE, Nordfjall K, Mangino M, Balasubramanyam M, de Meyer T, Hendricks AE, Giltay EJ, Hunt SC *et al*: **Body mass index is negatively associated with telomere length: a collaborative cross-sectional meta-analysis of 87 observational studies**. *The American Journal of Clinical Nutrition* 2018, **108**(3):453-475.
43. Rode L, Nordestgaard BG, Bojesen SE: **Peripheral Blood Leukocyte Telomere Length and Mortality Among 64 637 Individuals From the General Population**. *JNCI: Journal of the National Cancer Institute* 2015, **107**(6).

44. Haycock PC, Heydon EE, Kaptoge S, Butterworth AS, Thompson A, Willeit P: **Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis**. *BMJ : British Medical Journal* 2014, **349**:g4227.
45. Oeseburg H, de Boer RA, van Gilst WH, van der Harst P: **Telomere biology in healthy aging and disease**. *Pflugers Arch* 2010, **459**(2):259-268.
46. Polonis K, Sompalli S, Becari C, Xie J, Covassin N, Schulte JP, Druliner RB, Johnson AR, Narkiewicz K, Boardman AL *et al*: **Telomere Length and Risk of Major Adverse Cardiac Events and Cancer in Obstructive Sleep Apnea Patients**. *Cells* 2019, **8**(5).
47. Townsend MK, Aschard H, De Vivo I, Michels KB, Kraft P: **Genomics, Telomere Length, Epigenetics, and Metabolomics in the Nurses' Health Studies**. *Am J Public Health* 2016, **106**(9):1663-1668.
48. Leung CW, Fung TT, McEvoy CT, Lin J, Epel ES: **Diet Quality Indices and Leukocyte Telomere Length Among Healthy US Adults: Data From the National Health and Nutrition Examination Survey, 1999–2002**. *American Journal of Epidemiology* 2018, **187**(10):2192-2201.
49. Epel ES, Prather AA: **Stress, Telomeres, and Psychopathology: Toward a Deeper Understanding of a Triad of Early Aging**. *Annual Review of Clinical Psychology* 2018, **14**(1):371-397.
50. Thalheim T, Hopp L, Binder H, Aust G, Galle J: **On the Cooperation between Epigenetics and Transcription Factor Networks in the Specification of Tissue Stem Cells**. *Epigenomes* 2018, **2**(4):20.
51. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J *et al*: **Integrative analysis of 111 reference human epigenomes**. *Nature* 2015, **518**(7539):317-330.
52. Altman MC, Rinchai D, Baldwin N, Whalen E, Garand M, Kabeer BA, Toufiq M, Presnell S, Chiche L, Jourde-Chiche N *et al*: **A Novel Repertoire of Blood Transcriptome Modules Based on Co-expression Patterns Across Sixteen Disease and Physiological States**. *bioRxiv* 2019:525709.
53. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, Stichweh D, Blankenship D, Li L, Munagala I *et al*: **A Modular Analysis Framework for Blood Genomics Studies: Application to Systemic Lupus Erythematosus**. *Immunity* 2008, **29**(1):150-164.
54. De Jong S, Neeleman M, Luykx JJ, ten Berg MJ, Strengman E, Den Breeijen HH, Stijvers LC, Buizer-Voskamp JE, Bakker SC, Kahn RS *et al*: **Seasonal changes in gene expression represent cell-type composition in whole blood**. *Human molecular genetics* 2014, **23**(10):2721-2728.
55. Goldinger A, Shakhbazov K, Henders AK, McRae AF, Montgomery GW, Powell JE: **Seasonal Effects on Gene Expression**. *PLOS ONE* 2015, **10**(5):e0126995.
56. Tan F-L, Moravec CS, Li J, Apperson-Hansen C, McCarthy PM, Young JB, Bond M: **The gene expression fingerprint of human heart failure**. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 2002, **99**(17):11387-11392.
57. McLachlan JL, Smith AJ, Bujalska IJ, Cooper PR: **Gene expression profiling of pulpal tissue reveals the molecular complexity of dental caries**. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 2005, **1741**(3):271-281.

58. Wild SH, Byrne CD: **Body mass index and mortality: understanding the patterns and paradoxes**. *BMJ* 2016, **353**.

59. Brieland JK, Vissers MCM, Phan SH, Fantone JC: **Human platelets mediate iron release from transferrin by adenine nucleotide-dependent and -independent mechanisms**. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1989, **978**(2):191-196.

60. Binder H, Hopp L, Lembcke K, Wirth H: **Personalized Disease Phenotypes from Massive OMICs Data**. In: *Big Data Analytics in Bioinformatics and Healthcare*. Edited by Baoying W, Ruowang L, William P. Hershey, PA, USA: IGI Global; 2015: 359-378.

61. Wang L, Oh WK, Zhu J: **Disease-specific classification using deconvoluted whole blood gene expression**. *Scientific Reports* 2016, **6**:32976.

62. Herishanu Y, Rogowski O, Polliack A, Marilus R: **Leukocytosis in obese individuals: possible link in patients with unexplained persistent neutrophilia**. *European Journal of Haematology* 2006, **76**(6):516-520.

63. Wu X, Hakimi M, Wortmann M, Zhang J, Böckler D, Dihlmann S: **Gene expression of inflammasome components in peripheral blood mononuclear cells (PBMC) of vascular patients increases with age**. *Immun Ageing* 2015, **12**:15-15.

64. Mather KA, Jorm AF, Parslow RA, Christensen H: **Is Telomere Length a Biomarker of Aging? A Review**. *The Journals of Gerontology: Series A* 2010, **66A**(2):202-213.

65. Shawi M, Autexier C: **Telomerase, senescence and ageing**. *Mechanisms of Ageing and Development* 2008, **129**(1):3-10.

66. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, Hottenga JJ, Fischer K, Esko T, Surakka I *et al*: **Identification of seven loci affecting mean telomere length and their association with disease**. *Nat Genetics* 2013, **45**:422.

67. Nersisyan L, Hopp L, Loeffler-Wirth H, Galle J, Loeffler M, Arakelyan A, Binder H: **Telomere length maintenance and its transcriptional regulation in Lynch syndrome and sporadic colorectal carcinoma**. *Frontiers in Oncology* 2019, **in press**.

68. Lapham K, Kvale MN, Lin J, Connell S, Croen LA, Dispensa BP, Fang L, Hesselson S, Hoffmann TJ, Iribarren C *et al*: **Automated Assay of Telomere Length Measurement and Informatics for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort**. *Genetics* 2015, **200**(4):1061-1072.

69. Gardner M, Bann D, Wiley L, Cooper R, Hardy R, Nitsch D, Martin-Ruiz C, Shiels P, Sayer AA, Barbieri M *et al*: **Gender and telomere length: Systematic review and meta-analysis**. *Experimental Gerontology* 2014, **51**:15-27.

70. Balliu B, Durrant M, Goede Od, Abell N, Li X, Liu B, Gloudemans MJ, Cook NL, Smith KS, Knowles DA *et al*: **Genetic regulation of gene expression and splicing during a 10-year period of human aging**. *Genome Biol* 2019, **20**(1):230.

71. Horvath S: **DNA methylation age of human tissues and cell types**. *Genome Biol* 2013, **14**(10):R115-R115.

72. Binder H, Willscher E, Loeffler-Wirth H, Hopp L, Jones DTW, Pfister SM, Kreuz M, Gramatzki D, Fortenbacher E, Hentschel B *et al*: **DNA methylation, transcriptome and genetic copy number signatures of diffuse cerebral WHO grade II/III gliomas resolve cancer heterogeneity and development**. *Acta Neuropathologica Communications* 2019, **7**(1):59.
73. Hopp L, Löffler-Wirth H, Galle J, Binder H: **Combined SOM-portrayal of gene expression and DNA methylation landscapes disentangles modes of epigenetic regulation in glioblastoma**. *Epigenomics* 2018, **10**(6):745-764.
74. Ciccarone F, Tagliatesta S, Caiafa P, Zampieri M: **DNA methylation dynamics in aging: how far are we from understanding the mechanisms?** *Mechanisms of Ageing and Development* 2018, **174**:3-17.
75. Ray D, Yung R: **Immune senescence, epigenetics and autoimmunity**. *Clin Immunol* 2018, **196**:59-63.
76. Lorente-Sorolla C, Garcia-Gomez A, Català-Moll F, Toledano V, Ciudad L, Avendaño-Ortiz J, Maroun-Eid C, Martín-Quirós A, Martínez-Gallo M, Ruiz-Sanmartín A *et al*: **Inflammatory cytokines and organ dysfunction associate with the aberrant DNA methylome of monocytes in sepsis**. *Genome Medicine* 2019, **11**(1):66.
77. Daniel S, Nylander V, Ingerslev LR, Zhong L, Fabre O, Clifford B, Johnston K, Cohn RJ, Barres R, Simar D: **T cell epigenetic remodeling and accelerated epigenetic aging are linked to long-term immune alterations in childhood cancer survivors**. *Clinical Epigenetics* 2018, **10**(1):138.
78. Busslinger M, Tarakhovsky A: **Epigenetic Control of Immunity**. *Cold Spring Harbor Perspectives in Biology* 2014, **6**(6).
79. Jylhävä J, Pedersen NL, Hägg S: **Biological Age Predictors**. *EBioMedicine* 2017, **21**:29-36.
80. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD: **DNA Methylation Clocks in Aging: Categories, Causes, and Consequences**. *Molecular Cell* 2018, **71**(6):882-895.
81. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, Christensen BC, Gladyshev VN, Heijmans BT, Horvath S *et al*: **DNA methylation aging clocks: challenges and recommendations**. *Genome Biol* 2019, **20**(1):249.

Figures

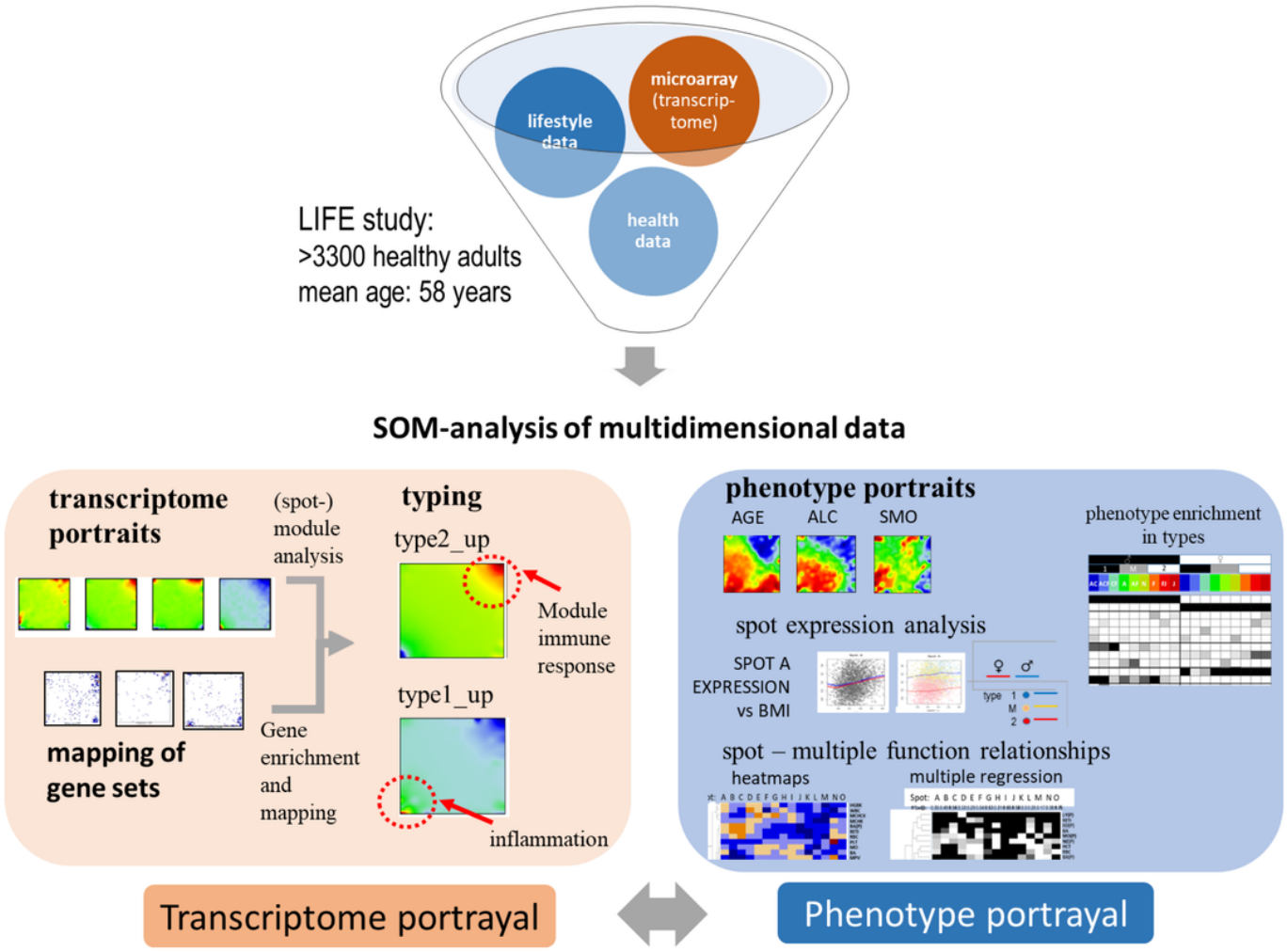


Figure 1

Schematic overview about portrayal approach applied in this study: Subject-matched transcriptome and ‘phenotype’ data of the LIFE-adult study were analyzed using SOM machine learning, which provides ‘personalized’ data portraits. Transcriptome portraits were used for classification, gene-module extraction and functional interpretation. Phenotype features such as blood cell counts or BMI provide phenotype (versus-transcriptome correlation) portraits. Further spot-module expression profiles were used for detailed pattern analyses. Phenotypic features analyzed in this study were listed in Table S 1.

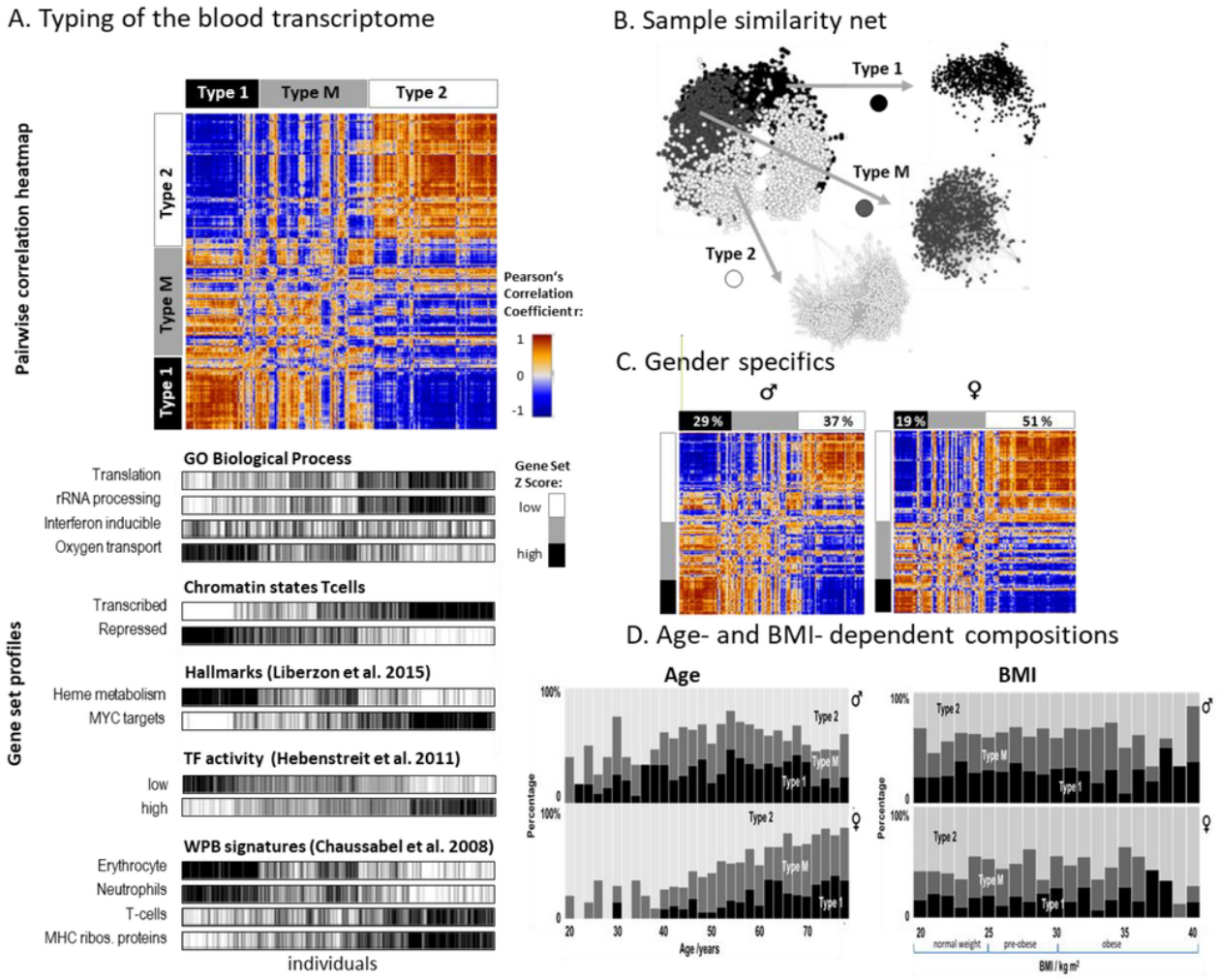


Figure 2

Stratification of the blood transcriptome into three types: (A) The pairwise correlation heatmap divides the samples into three transcriptome types according to the mutual correlations between their SOM expression portraits. The part below associates the samples of all three types to selected functional categories by means of gene set overexpression. The samples are sorted in the clusters according to different combinatorial pattern types (cPATs) which gives rise to the striped patterns in the map (see below). (B) Network presentation of the similarity relations illustrates that the samples of type 1 and type 2 form almost separate data clouds while type M overlaps with both. (C) Stratification of the samples into women and men shows essentially similar correlation heatmaps and thus similar expression patterns. Women however are more frequently found in type 2. (D) The composition of types changes with age in a gender specific way but virtually not with BMI (body mass index). The percentage of Type 2 women permanently decreases with age while the relative amount of type 2 men older than 55 years again increases.

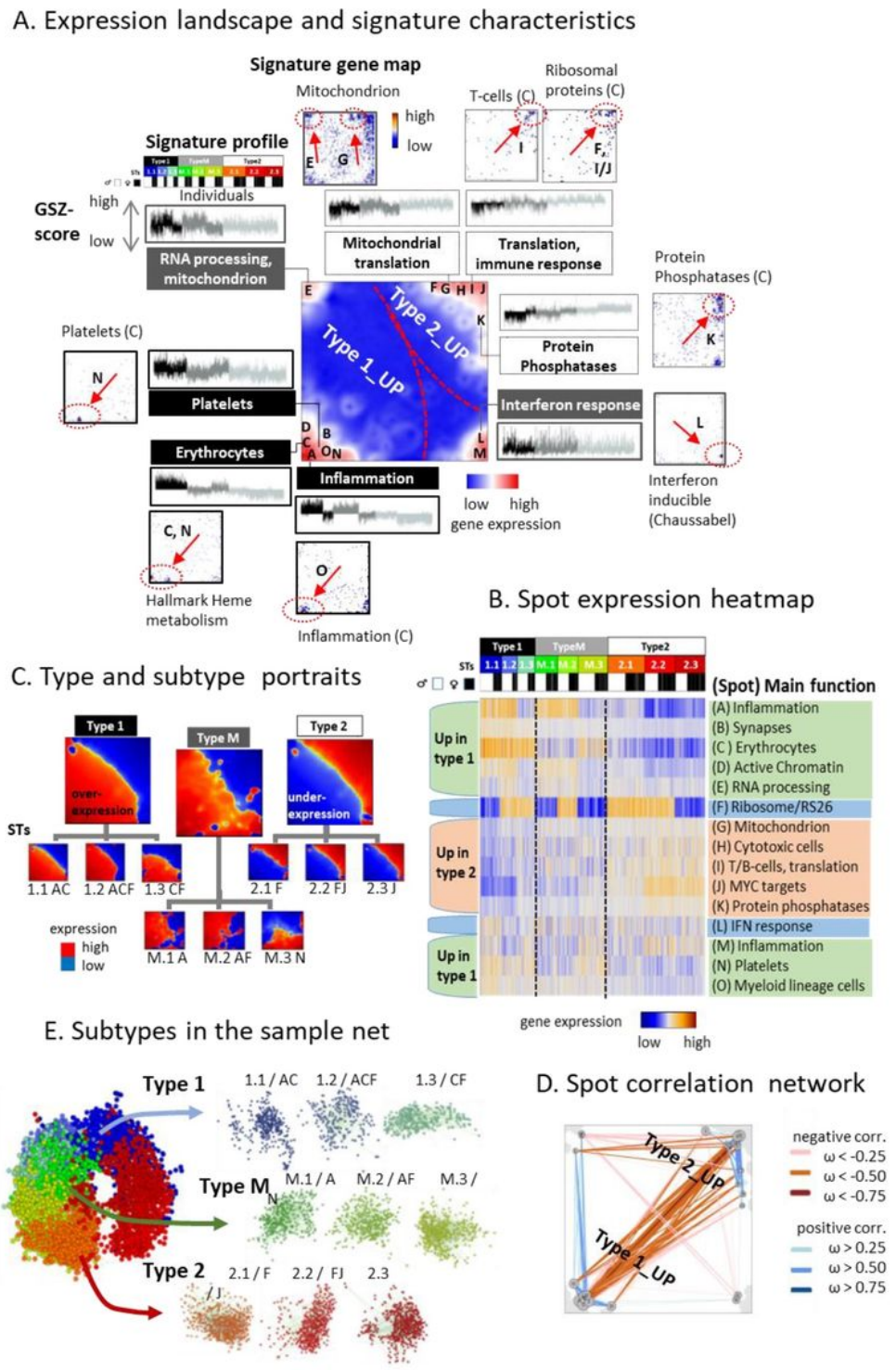


Figure 3

The landscape of the blood transcriptome: (A) The spot overexpression summary map shown in the center of part A provides an overview about spot-clusters of co-expressed genes labelled with capital letters A – M. The map divides into areas which contain spots upregulated predominantly in type 1 and type 2 transcriptomes, respectively. A selection of gene sets illustrates the functional context of the spots by showing their profiles and gene maps. These maps indicated the location of the genes of the

respective gene set by dots. Their accumulation in and near spot areas are shown by arrows and dashed ellipses (see also Table S 2). The expression profiles of the gene sets reveal their distinct up- and downregulation in the different sample types. (B) Profiles of the spots are shown as heatmap together with their major functional context (see also Table S 2). (C) The correlation map indicates positive and negative correlations between the spots by red and blue lines, respectively. Correlations were calculated using the weighted overlap measure [32]. (D) Mean portraits of types and subtypes (STs) reveal mutual up- and downregulation of genes located in different parts of the map. (E) Samples of different STs accumulate in well separated data clouds in the similarity net.

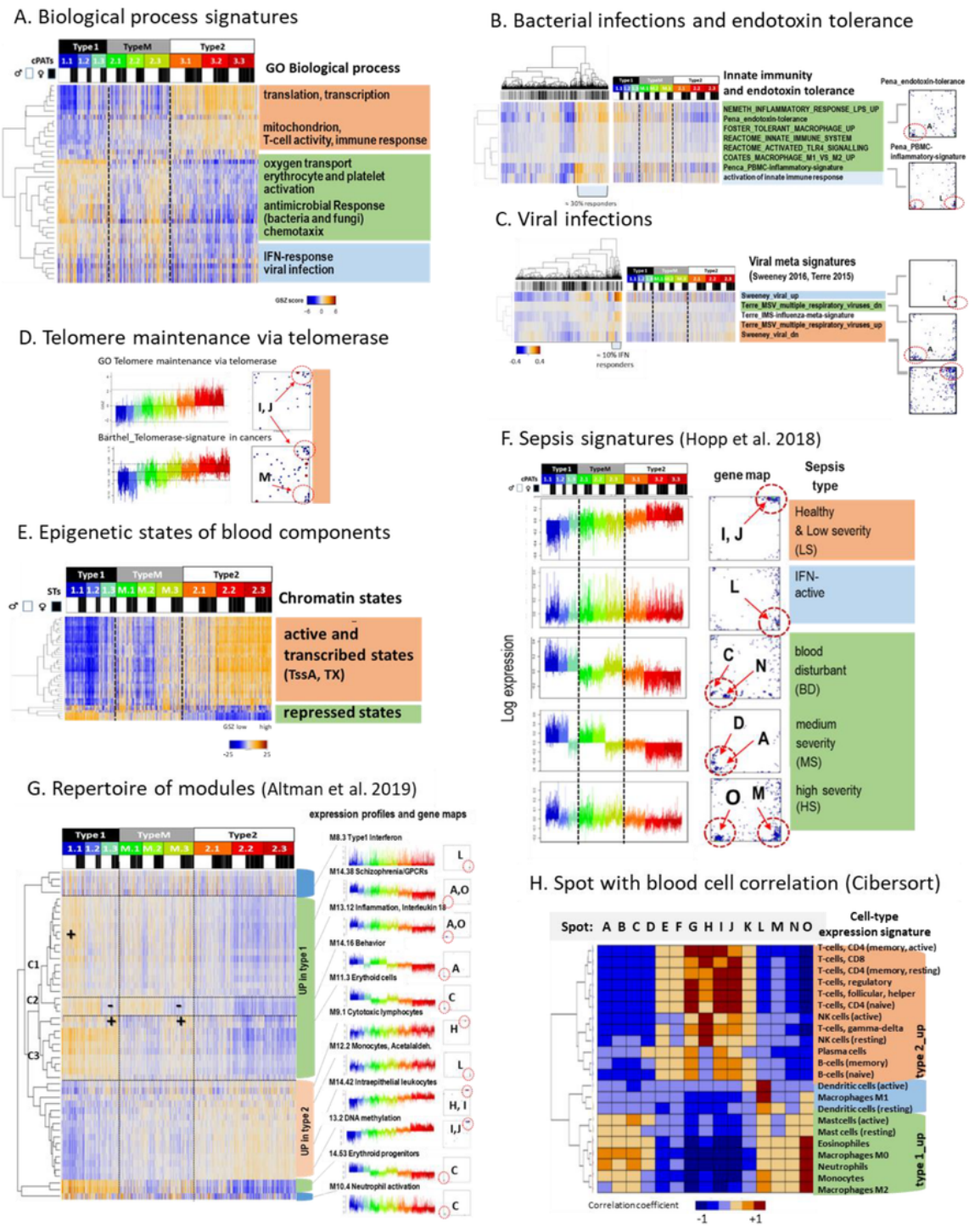
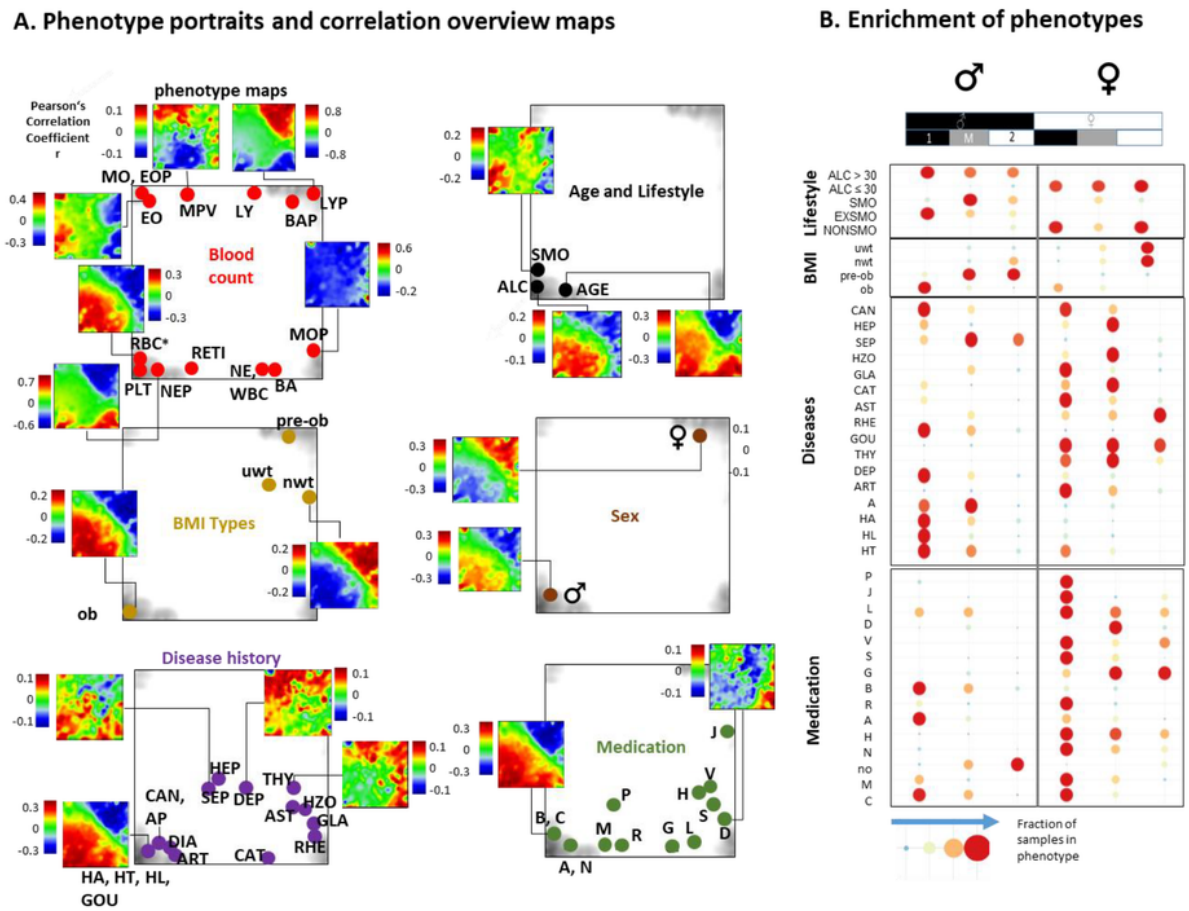


Figure 4

Functional characteristics and previous signatures of the blood transcriptome: (A) Signatures of the GO-term biological process (BP) roughly group into processes upregulated in type 1 (green), type 2 (apricot color) and in samples showing high expression in all types (blue)(see also Figure S 5). (B) Gene signatures associated with bacterial [36-39] and, (C), with viral infections [40, 41] associate with distinct spots in the SOM and profiles. Clustering of samples reveals that about 30%/13% of specimen show

elevated levels of transcriptomic footprints of bacterial/viral infections. (D) Signatures of telomere length maintenance (TM) taken from GO-repository [33] and a meta-analysis of cancer induced TM [35]. Both signatures activate in type 2 (spots I, J) where the cancer-derived signature also contains genes located in spot M associated with inflammation (see also Figure S 7). (E) Expression of genes assigned to different epigenetic states of healthy blood cells (see [51]) reflects that transcriptionally active states (active promoters and transcribed genes) upregulate predominantly in type 2 while repressed states become upregulated in type 1 what suggests their reprogramming into active ones (see also Figure S 8). (F) Expression signatures extracted from the blood transcriptomes of sepsis patients framed within Community Aquired Pneumonia (CAP) [8] express clear footprints also in the LIFE-adult cohort. (G) A repertoire of functionally annotated transcriptional modules [52] reveals agreement with our typing scheme. The heatmap of the 50 most variant modules indicates concerted regulation with our types where modules overexpressed in type 1 further split into 3 clusters C1-C3 associating with specific ST expression patterns. Interestingly, modules which upregulate in type 2 show a less diverse cluster structure (see also Figure S 12). (H) Correlation between spot expression profiles and signature expression of a collection of 22 blood cell compounds taken from Cibersort [18]. They upregulate either in type 1 or type 2 blood transcriptomes or resemble the IFN-response characteristics (dendritic cells and macrophages M1). This patterns are commonly found also for collections of other signatures thus reflecting major features of the intrinsic modular structure of the blood transcriptome (see Figure S 13).



C. Glossary of phenotypes

Blood count		max r
BA(P)	Basophils absolute	0.16
	(Basophils per WBC)	(0.15)
EO(P)	Eosinophils absolute	0.36
	(Eosinophils per WBC)	(0.29)
LY(P)	Lymphocytes absolute	**0.54**
	(Lymphocytes per WBC)	**(0.77)**
MO(P)	Monocytes absolute	**0.49**
	(Monocytes per WBC)	**(0.6)**
MPV	Mean platelet volume	0.07
NE(P)	Neutrophils absolute	**0.72**
	(Neutrophils per WBC)	**(0.84)**
PLT	Platelets	0.255
RETI	Reticulocytes	**0.66**
RBC*	Red blood cell components (i.e. RBC, HGB, MCH, MCHK, MCHC, MCHCK, MCV, HCT)	
WBC	Leucocytes	0.55
RBC*		
RBC	Red blood cells	0.55
HCT	Hematocrit	0.31
HGB	Hemoglobin (conv. units)	0.34
MCH	Mean corpuscular hemoglobin (SI units)	0.21
MCHK	MCH (conv. units)	0.2
MCHC	MCH concentration (SI units)	0.21
MCHCK	MCHC (conv. units)	0.21
MCV	Mean corpuscular volume	0.13

Age and Lifestyle		max r
AGE	Age	**0.26**
AGE 30-50	Age between 30 and 50	0.16
ALC	Alcohol consumption	
ALC > 30	> 30g Alcohol per day	
ALC ≤ 30	≤ 30g Alcohol per day	0.16
EXSMO	Ex-smoker	
SMO	Smoker	
NONSMO	Non-smoker	
BMI types		
nwt	normal weight	0.21
ob	obese	0.22
pre-ob	pre-obese	0.05
uwt	underweight	0.09

Disease history		
A	Angina pectoris	0.1
ART	Arthrosis	0.08
AST	Asthma	0.05
CAN	Cancer	0.1
CAT	Cataract	0.15
DEP	Depression	0.03
DIA	Diabetes	0.18
GLA	Glaucoma	0.07
GOU	Gout	0.12
HA	Heartattack	0.1
HEP	Hepatitis	0.07
HL	Hyperlipidaema	0.17
HT	Hypertension	**0.25**
HZO	Hzoster	0.05
RHE	Rheuma	0.05
SEP	Sepsis	0.04
THY	Thyroid	0.05

Medication		max r
A	Alimentary tract and metabolism	0.16
B	Blood and blood forming organs	0.18
C	Cardiovascular system	**0.26**
D	Dermatologicals	0.05
G	Genito urinary system and sex hormones	0.06
H	Systemic hormonal preparations (exclusive sex hormones + insulin)	0.07
J	Antiinfectives for systemic use	0.04
L	Antineoplastic and immunomodulating agents	0.08
M	Musco-skeletal system	0.11
N	Nervous system	0.11
P	Antiparasitic products	0.05
R	Respiratory organs	0.08
S	Sensory organs	0.07
V	Various	0.06

Figure 5

Association of selected features (phenotypes) with the transcriptome landscape of blood: (A) Phenotype (correlation) portraits visualize the correlation between metagene expression profiles and the profiles of selected phenotypes in a red-to-blue colour scale. The correlation overview maps for each of the categories mark the metagene of maximum correlation coefficient for each of the phenotype studied (see the legend in part (C) and also Table S 1 for assignment of phenotypes). (B) Distribution of samples of

each of the phenotypes among the transcriptome types. Obese men and men consuming alcohol (>30g/day) accumulate (red circles) in type 1 transcriptomes. Also participants with different disease histories enrich in type 1 in a gender-specific fashion while medication is most prevalent in women of type 1. Type 2 and type M refer to underweight and normal weight women and to smoking men, respectively. Enrichment of blood count data is provided in Figure S 18. Correlation maps and further details are presented for each of the phenotype categories in Figure S 19 - Figure S 23.

Figure 6

Ageing and BMI characteristics of the blood transcriptome: (A) Expression of selected spots as a function of age and BMI. Separate LOESS (local weighted scatterplot smoothing) fits for women and men (red and blue curves) and for types 1, M and 2 visualize mean spot expressions as a function of age and BMI. The course is mostly non-linear and change slope at different turning points (see arrows). (B) Genes from previous ageing sets [11] spread either over a heterogeneous (age_up) or a more homogeneous (age_dn) distribution of spots (letters and dashed circles). The latter set correlates with the expression of spots I and J (correlation plots at the right). (C) The overall ageing portrait (of correlations between the ages of the participants and the transcriptome ‘metagene’ landscape) roughly can be understood as superposition of (increasing with age in red) RBC- and NE-like fingerprints and (decreasing with age in blue) LY-patterns. Stratification into decade intervals shows that NE-fingerprints apply mainly to elderly people (>60 years) while elevation of RBC is relevant mostly for participants younger than 60 years. (D) Portraits of different BMI strata reveal the continuous change from type 2 (under- and normal-weight participants) to type 1 (obese) transcriptomes. (E) Selected blood cell count and serum marker portraits reveal fingerprint-like patterns. Characteristic ‘landmark’ spots are indicated by dashed circles and the respective spot letters.

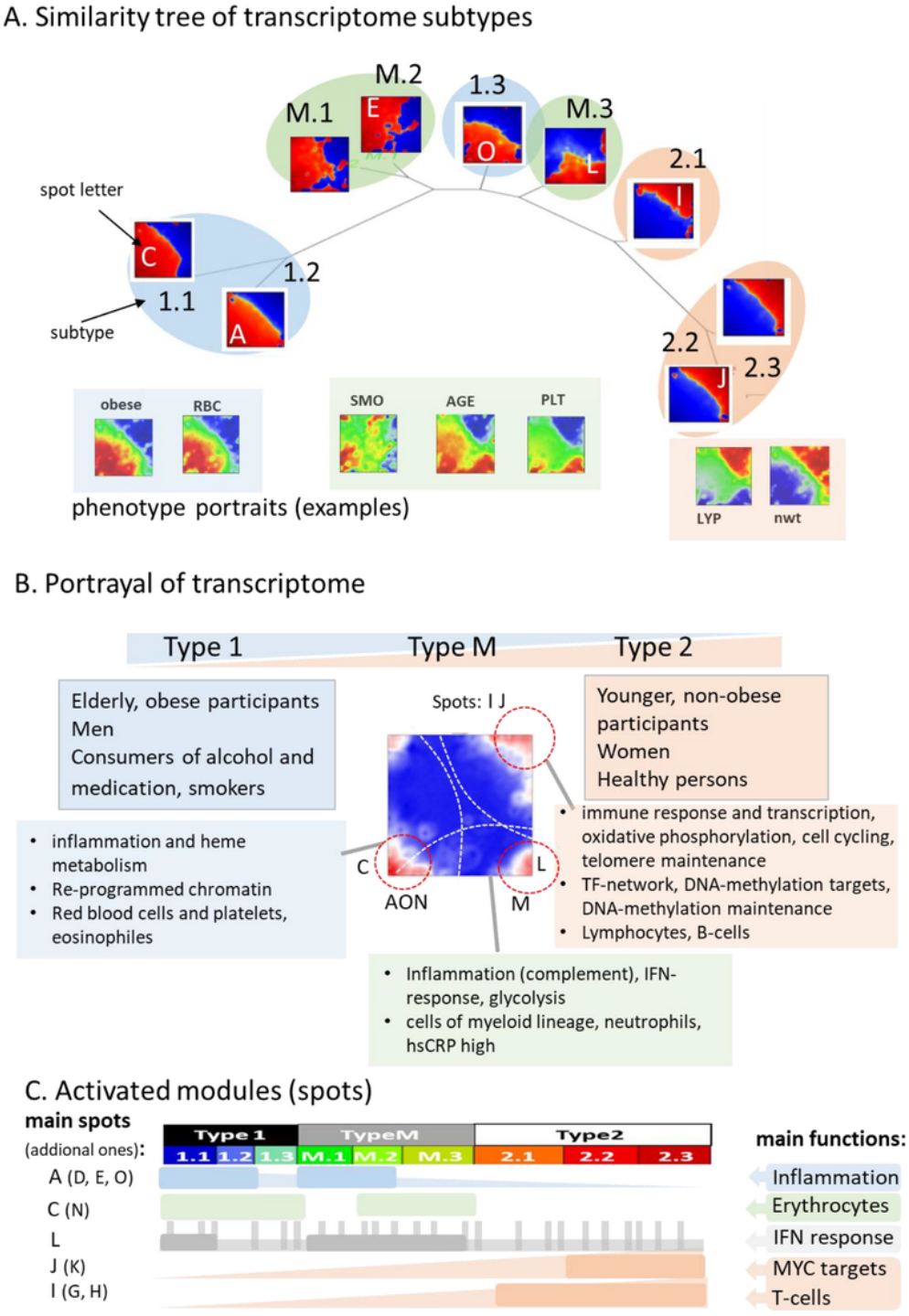


Figure 7

Portrayal of the blood transcriptome: (A) The similarity tree reflects a virtually linear arrangement of subtypes due to a continuum of transcriptional states ranging from type 1 to type 2. Selected phenotype portraits illustrate correlation of the respective features with the transcriptomes of different types. (B) Main functions and phenotypes associating with the transcriptome types. (C) The spot modules of co-expressed genes can be roughly grouped into a handful main patterns, which heterogeneously distribute

over types. Profiles of spots A and I/J change continuously among the subtypes while spot C shows specific upregulation in selected subtypes. Samples with activated spot L distribute over all subtypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryfile4Altmanbloodgenesets.pdf
- supplementaryfile3spotlists.xlsx
- Suppementaryfile1bloodtranscriptomesupplement8.pdf
- SuppFig2portraits.pdf