

RNA sequencing of whole blood defines the signature of high intensity exercise at altitude in elite speed skaters

Andrey S. Glotov^{1,2*}, Irina E. Zelenkova^{3,4}, Elena S. Vashukova², Anna R. Shuvalova², Alexandra D. Zolotareva², Dmitrii E. Polev², Yury A. Barbitoff^{2,5}, Oleg S. Glotov^{2,6}, Andrey M. Sarana^{6,7}, Sergey G. Scherbak^{6,7}, Mariya A. Rozina⁸, Victoria L. Gogotova⁸, Alexander V. Predeus^{5*}

¹Laboratory of Biobanking and genome medicine of Institute of Translational Biomedicine, Saint Petersburg State University, Russia, 199034, St. Petersburg, Universitetskaya nab., 7-9-11

²Division of Genome Medicine, D.O. Ott Research Institute of Obstetrics, Gynecology and Reproductology, Russia, 199034, St. Petersburg, Mendeleevskaya 3

³GENUD research group, FIMS Collaborating Center of Sports Medicine, Department of Physiatry and Nursing, University of Zaragoza, Zaragoza, Spain

⁴I.M. Sechenov First Moscow State Medical University (Sechenov University), Ministry of Health of Russia, Moscow, Russian Federation

⁵Bioinformatics Institute, Saint Petersburg, Russia, 194100, Kantemirovskaya st., 2A

⁶City Hospital #40, Saint Petersburg, Russia, 197706, St. Petersburg, Sestroretsk, Borisov Str., 9

⁷Department of Postgraduate Medical Education, St. Petersburg State University, Russia, 199034, St. Petersburg, Universitetskaya nab., 7-9-11

⁸Russian Skating Union, Moscow, Russia, 119991, Luzhnetskaya nab., 8, office 230

*To whom correspondence should be addressed: a.glotov@spbu.ru, predeus@bioinf.me
NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract

Objectives: Although high altitude training has been increasingly popular in endurance athletes, the molecular and cellular bases of this adaptation remain poorly understood. We aimed to define the underlying physiological changes and screen for potential biomarkers of adaptation using transcriptional profiling of whole blood. More generally, we aimed to evaluate the utility of blood RNA sequencing as a modern and sensitive method of athlete's health monitoring.

Methods: Seven elite female speed skaters were profiled before and after 1h intense exercise, on the 18th day of Live High, Train High (LHTH) training programme. Whole blood RNA sequencing (RNA-seq) with globin depletion was used to measure gene expression changes associated with high intensity exercise at high altitude. Eight public microarray datasets were used to identify genes uniquely regulated at high altitude. Gene markers derived from single cell RNA-seq data were used to evaluate the changes of individual cell types in the whole blood.

Results: Using individual cell type signatures, we were able to deconvolute the changes of finely defined cell populations from the whole blood RNA-seq. We have detected the increase in neutrophils, platelets, erythrocytes, and CD14 monocytes, and the decrease in natural killers, CD8 T cells, memory CD4 T cells, B cells, and plasmacytoid dendritic cells. The levels of naive CD4 T cells, CD16 monocytes, and myeloid dendritic cells were unchanged. Leveraging the previously published transcriptomic data allowed us to define the expression signature unique to high-altitude adaptation. Among the identified genes we highlight PHOSPHO1, which has a known role in erythropoiesis, and MARC1 with a proposed role in endogenous NO metabolism. Finally, we find that platelets and, to a lesser extent, erythrocytes are the two major cell types that uniquely respond to altitude exercise, while neutrophils represent a more generic marker of intense exercise.

Conclusions: Using publicly available data from both single-cell RNA-seq atlases and exercise-related blood profiling dramatically increases the value of whole blood RNA-seq for dynamic evaluation of physiological changes in an athlete's body. In addition to the measurement of individual gene expression changes, our approach allowed us to estimate changes of blood cell type counts from a small peripheral blood sample, without sorting or other expensive and unfeasible equipment. We also discuss a surprising parallel of hypoxia and increased thrombosis, and hypothesize about the role exercise can play in COVID-19 outcomes.

Introduction

Systemic effects of exercise have drawn substantial interest from researchers in medical, public health, and athletic fields for over a century (Gjevestad, Holven, and Ulven 2015; Nieman and Pence 2020; Nieman and Wentz 2019). In one of the earliest publications on the subject, Larrabee in 1902 described leukocytosis (a decrease in white blood cell count) in runners tested after The Boston Marathon (Larrabee 1902). With medical advances of the 20th century, it quickly became apparent that physical activity and sufficient exercise is the single most significant predictor of general population health and wellbeing (Bull et al. 2020). At the same time, biological reasons for this remained elusive, and key molecular drivers of adaptation to physical exercise are yet to be defined on the systemic level. The worldwide epidemic of obesity, and general shift towards sedentary lifestyle, especially aggravated during the current SARS-CoV-2 pandemic, make these goals ever more important (Bull et al. 2020). With an onslaught of the omics methods of the 21st century we gained access to incredibly valuable genome-wide datasets, characterizing gene expression, proteome, and metabolic changes associated with exercise. However, as it often happens with revolutionary methods, our understanding of the data still lags far behind.

Among the omics approaches used to characterize the effect of physical exertion on the human body, gene expression measurement has become the most popular. This happened for several reasons. Since first microarray experiments, and then with RNA sequencing (RNA-seq), gene expression methods have reached relative maturity, currently allowing the measurement of all expressed human genes in a particular tissue with high accuracy and reproducibility (Z. Wang, Gerstein, and Snyder 2009; C. Wang et al. 2014). They require a modest amount of biological material that can be effectively preserved without freezing, which becomes critical in many environments. Finally, both microarray and RNA-seq experiments can be done at a reasonable cost, making them accessible to many laboratories. Not all of these criteria are satisfied in case of unbiased proteomic or metabolomic profiling, both of which are still in the phase of active method development (Timp and Timp 2020; Zhang et al. 2020).

The choice of the profiled tissue is often defined by the biological questions posed. While obesity-focused studies have often used adipose tissue biopsies (Fabre et al. 2018), studies in athletes preferably profiled muscle tissue (Gustafsson et al. 1999; Hawley et al. 2018; Terry et al. 2018). Both of these approaches, however, include relatively traumatic biopsies, making them hard to perform outside of medical facilities. Peripheral blood sampling, on the other hand, has many advantages - such as well-established collection, storage, and transportation protocols, relatively minor effects on the subject's well-being, and presence of many valuable biomarkers. Blood also represents a very diverse mixture of cells, giving access to an intriguing interface between metabolic

and immune functions. Indeed, it is well documented that blood composition dynamically reacts to both physical and immunological challenges (Lewis et al. 2010). Because of these factors, gene expression of peripheral blood is used extremely widely, currently listing thousands of published studies. Depending on the study goals, researchers have profiled whole blood, white blood cells (WBCs), peripheral blood mononuclear cells (PBMCs), or sorted sub-populations of leukocytes (Donohue et al. 2019).

More specifically, the effects of short- and long-term exercise on peripheral blood gene expression have been reported in over 50 different studies, with 20+ using whole-transcriptome profiling methods reviewed in (Gjevestad, Holven, and Ulven 2015). Vast majority of the published blood whole-transcriptome datasets were generated using microarray technology. The experimental designs vary substantially; time of study changes from immediately after a bout of moderate or strenuous exercise to weeks or months of regular activity; the participants also varied in level of preparation, age, and sex. We have summarized all the relevant datasets available in the literature in Supplementary Table S1, only including the studies for which the data are openly available via databases like Gene Expression Omnibus (GEO) or ArrayExpress. It is worth noting that no RNA sequencing datasets were available at the time of our search, and no studies involved high altitude adaptation.

In our study, we applied whole blood RNA-seq to seven elite female skaters, in order to assess the immunological and metabolic effects of high-altitude adaptation. To our knowledge, this is the first such attempt. We also put our results in the context of the previously published gene expression studies, and identified candidate genes driving the adaptation, as well as general physiological changes inferred from gene expression.

Material and methods

Subjects and study protocol

The study was conducted during the pre-season period, and each participant underwent medical evaluations including collection of medical history. Seven female elite speed skaters were enrolled and gave their informed written consent to participate in this study. All participants underwent medical examination and were deemed fit for training and competitive activity; none had a history of cardiovascular, pulmonary, or metabolic diseases. Height, weight, and several physiological parameters were recorded for each athlete. The average participant age was 25.0 ± 7.0 years. All patients were of Russian ethnicity. The study was approved by the Saint Petersburg State University Ethics Review Board for human studies (decision #40 from 07.03.2012) and was

performed in accordance to the Declaration of Helsinki.

Physiological measurements and sample collections were carried out during an altitude training camp that was carried out 1850 meters above the sea level (Font Romeu, France, 1850 meters above sea level, 21 days of stay). The study was conducted during the period after the first part of adaptation period (after 18 days from the beginning of the altitude exposure). Pre- and post-exercise data collections were carried out in the morning between 08:00 and 11:00. Key physiological and biochemical parameters are given in Supplementary Table S2.

Blood sample collection and RNA isolation

For each participant, 2.5 mL whole blood sample was collected before and after the exercise using the RNAgard Blood Tubes (Biomatrica, USA) according to the manufacturer's protocol and then stored at -20°C until further processing. Total RNA was extracted from blood using PureLink RNA Mini Kit (Thermo Fisher Scientific, Inc.) and «BioMaxi™ Precipitation Buffer» (Biomatrica, USA), according to the manufacturer's protocol. RNA concentration was measured using Quantus Fluorometer TM with QuantiFluor RNA System kit (Promega, USA). RNA quality control was performed using capillary gel electrophoresis on a QIAxcel Advanced System (Qiagen, Germany). Total RNA was depleted of globin mRNA with GLOBINclear - Human Kit (Invitrogen, USA) according to the manufacturer's protocol.

Library preparation and Illumina RNA sequencing

Fourteen samples of globin-depleted whole blood (seven skaters, before and after exercise) were sequenced using strand-specific RNA-seq with polyA selection. Libraries were prepared using TruSeq Stranded mRNA Library Prep Kit (Illumina, Inc., USA) according to the TruSeq Stranded mRNA Sample Preparation Guide # 15031047 E (Illumina, USA). Validation of the libraries was performed on the QIAxcel Advanced System (Qiagen, Germany). Library quantification was performed using Quantus Fluorometer with QuantiFluor dsDNA System kit (Promega, USA). Paired-end sequencing of the libraries was performed on HiSeq 4000 System with a reading length of 2x150 bp using HiSeq 3000/4000 PE Cluster Kit and HiSeq 3000/4000 SBS Kit (300 cycle) (Illumina, USA). The number of reads obtained per sample varied from 12.4 to 40.8 M reads, with mean of 31.3 M, and median of 32.2 M reads per sample. Raw reads and processed data were deposited in Gene Expression Omnibus database under study ID GSE164890.

Alignment and quantification

Assessment of raw read quality was done using FastQC v0.11.6. Paired-end reads were aligned using STAR v2.5.3a (Dobin et al. 2013) to the primary assembly of the human genome

(version GRCh38.p10), with GENCODE v26 annotation (Frankish et al. 2019) with pseudoautosomal (PAR) gene copies removed. STAR options “--outSAMtype BAM SortedByCoordinate --quantMode TranscriptomeSAM” were enabled, thus generating alignments to both genome and transcriptome. Overall, 94.5 - 98.6 % reads (median 96.0%) were successfully aligned, with 2.5 - 11.5% overall reads (median 2.7%) aligning to rRNA. Genome BAM files were used to generate TDF files using *igvtools* v2.3.93, and visualized using *IGV* v2.4.11 (Thorvaldsdottir, Robinson, and Mesirov 2013). Transcriptomic BAM files were used for quantification with *RSEM* v1.2.31 (Li and Dewey 2011), with “--strandedness reverse” option enabled, according to the strand-specific library preparation type, and generating expression tables in raw counts, TPM, and FPKM on a transcript and gene level. After quantification, 67.6 - 82.5% (median 81.9%) of the original reads were successfully assigned to the genes present in Gencode v26 annotation. Detailed pipeline for read quality control, alignment, and quantification is available at <https://github.com/apredeus/rnaquant>.

Differential expression and pathway enrichment analysis

Differential expression and overrepresentation pathway enrichment analysis, as well as all other bioinformatic analysis from this section on, was done in *R* v4.0.4. Per-gene expression table generated by *RSEM* was used for differential gene expression with *DESeq2* R package v1.18.1 (Love, Huber, and Anders 2014), retaining genes with $FDR < 0.1$. Variance-stabilizing *rlog* transformation from the *DESeq2* package was used to normalize the expression data for diagnostic plotting. Donor effect correction was done on *rlog*-transformed matrix using the *comBat* function from the *sva* R package (Leek et al. 2012). Pathway enrichment was done using hallmark (H) and canonical (CP) gene set collections from *MsigDB* (A. Liberzon et al. 2011; Arthur Liberzon et al. 2015) and the *clusterProfiler* package (Yu et al. 2012).

Public single cell RNA-seq dataset processing

Publicly available single cell RNA-seq datasets (GSE149938 and 10k PBMC cells from 10X Genomics) were downloaded locally and processed using the *Seurat* package (Stuart et al. 2019). Each dataset was filtered, normalized, clustered to generate coarse-grained cellular populations, and markers defining each cell type were generated.

Microarray dataset reanalysis

Eight selected microarray studies were uniformly re-processed using the *GEOquery* (Davis and Meltzer 2007) and *limma* (Ritchie et al. 2015) packages. Each dataset was visualized

(Supplementary Figures S2-S9), and a list of differentially expressed genes was generated using a pairwise *limma* linear model that included both donor and exercise.

Reproducible analysis

Detailed scripts for all performed analysis and figure generation are freely available at https://github.com/apredeus/skater_maseq. All the data files needed to reproduce every analysis done in the paper are available in the repository. Raw reads and processed data were deposited in Gene Expression Omnibus database under study ID GSE164890.

Results and discussion

Blood panel and physiological measurements

Biochemical blood parameters (see Supplementary Table S2 for the full list) and physiological measurements (fat mass and percentage, muscle mass and percentage, total fluid, phase angle, tHb-mass, total circulating blood volume (TCBV), hemoglobin, hematocrit, and percent recovery index in each microcycle) were measured throughout the adaptation period and were found to be in line with previously reported values. Due to logistic restrictions, daily physiological measurements detailing the adaptation process were not performed; instead, testing was done on several select days. At the same time, many reports detailing biochemical and physiological adaptation to high altitude have been published previously, and were not the aim of this study (Horscroft et al. 2017; Moore 2017).

Whole blood gene expression was measured in samples collected before and after a morning bout of strenuous exercise on day 18 of adaptation. The mean running time in the exercise tests was 39.0 ± 14.8 min. Heart rate was 182 ± 3 bpm at the end of the exercise. At the end of the exercise, lactate concentrations were significantly increased (3.4 ± 0.7 vs. 1.1 ± 0.2 mmol/l; t-test $p < 0.05$). Biochemical parameters immediately before and after the exercise were taken for 6 markers: total phosphate, cortisol, growth hormone, total testosterone, total T4 hormone, and CPK. Using paired Wilcoxon test we have evaluated the significance of the observed changes. We have found that cortisol and phosphate significantly decreased after the exercise, while CPK and growth hormone have increased; total T4 and testosterone remained unchanged (Supplementary Figure S2).

Differential gene expression analysis

After read alignment and quantification (see Methods), we have performed the initial evaluation of the dataset. Using 18,000 most expressed genes and principal components analysis

(PCA) plot (Figure 1A), we have assessed the difference between samples collected before and after exercise. A clear donor effect was visible from the plot, with samples belonging to the same donor being closer to each other than to samples of the same group. Thus, we have applied linear donor correction using the *comBat* function of the *sva* R package. The PCA plot after the correction (Figure 1B) has shown much clearer separation of groups. This is common in studies of human blood in particular, because blood composition varies notably between individuals. From this analysis, we conclude that it is beneficial to include donor covariate in all subsequent analyses.

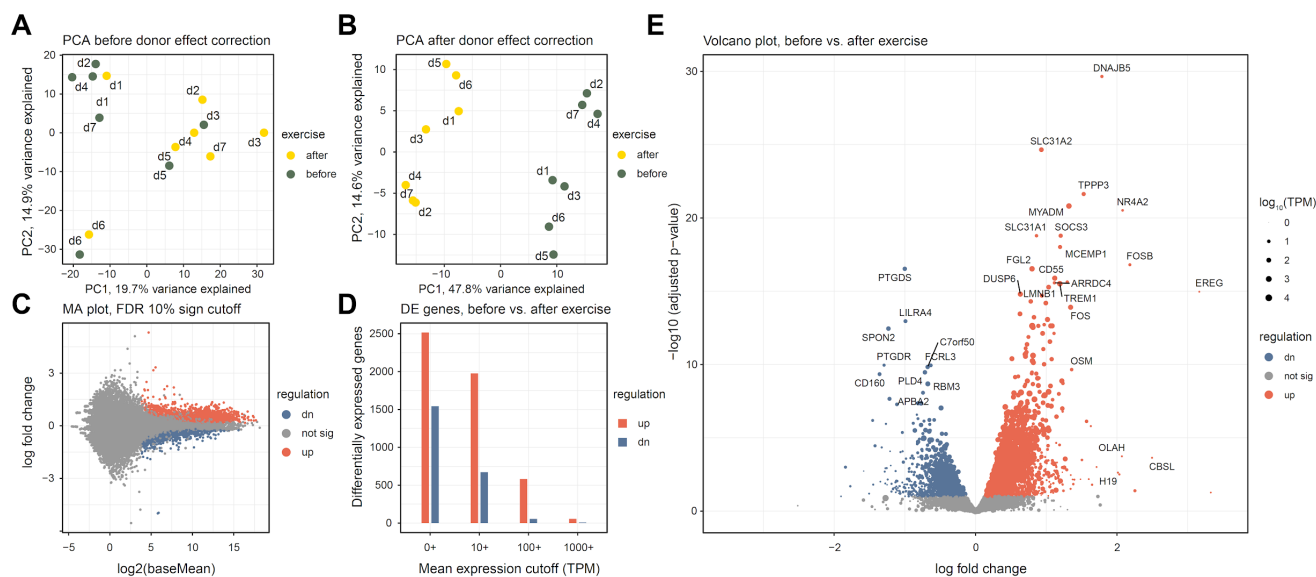


Figure 1. Overall assessment and differential expression analysis of whole blood RNA-seq from 7 altitude-adapted female skaters before and after exercise. Differential expression analysis was done using DESeq2 and “donor + exercise” design. Differentially expressed genes were reported at 10% FDR. “Up-regulated” indicates genes which expression increased after the exercise. (A, B) PCA plot of the 14 studied samples, before and after donor-effect correction using *comBat*. Top 18,000 genes were used. Read counts were normalized using the *rlog* function from the *DESeq2* package. (C) Log ratio - mean expression (MA) plot, with marked differentially expressed genes. (D) Number of differentially expressed genes depending on mean expression cutoff; TPM, transcripts per million. (E) Volcano plot of differentially expressed genes. Point size is scaled proportionally to mean gene expression.

Differential expression analysis has uncovered substantial changes in gene expression, with 2,516 genes up- and 1,542 down-regulated (see Supplementary Table S3 for a full list of genes). The difference between the up- and down-regulated gene numbers becomes more pronounced when looking at highly expressed genes; for example, when only genes with mean TPM of 100 or more were considered, 582 up-regulated and 55 down-regulated genes remained (Figure 1D). For most differentially expressed genes, expression change magnitude was modest (Figure 1C, 1E): only 104 up-regulated and 58 down-regulated genes changed their expression more than twofold.

Pathway and functional category analysis

After initial assessment of differentially expressed genes, we aimed to dissect the functional and molecular pathways regulated by the exercise. To this end, we used the molecular signature database (MsigDB) pathway collection of annotated pathways relevant to human biology, immunology, metabolism, and disease (A. Liberzon et al. 2011; Arthur Liberzon et al. 2015). We used overrepresentation analysis with hallmark (H) and canonical pathways (CP) gene set collections to define the major biological categories of interest. Figures 2A-C show top 10 representative pathways ranked by significance. Overall, inflammatory and immune pathways dominated the observed changes, with TNF α /Nf-K β , complement, interferon, IL6/JAK/STAT3 and other pathways showing strong up-regulation. On a cellular level, neutrophil degranulation and platelet activation account for a substantial fraction of up-regulated genes. Finally, hypoxia-related genes, and angiogenesis via VEGFA/EGFR2 are also strongly up-regulated in response to exercise.

Among the down-regulated genes, only one hallmark gene set (HALLMARK MYC TARGETS V2) was determined to be significant. MYC is a well-known blood oncogene that is particularly important in lymphomas (Cai et al. 2015), and has a crucial influence on cell survival and proliferation. MYC gene itself was also significantly down-regulated after the exercise (Supplementary Table S3). Together with the up-regulation of apoptosis pathway (Figure 2A), we can hypothesize that a fraction of blood cells undergoes apoptosis in response to vigorous exercise, which has been described before (Mooren et al. 2002). The majority of other pathways enriched among the down-regulated genes were related to ribosomal proteins and other components of transcriptional and translational machinery (Figure 2C). An interesting standout is the DNA repair pathway, which also appears to be down-regulated alongside its most famous member, TP53.

Since most of the observed gene expression changes were modest, gene set enrichment analysis (GSEA) may have offered additional insights, potentially highlighting metabolic processes obscured by more pronounced immune gene changes (Subramanian et al. 2005). To our surprise, however, GSEA analysis results were in exceptional agreement with the simple overrepresentation analysis (Figures 2D-F). Inclusion of broader reference gene sets, such as C2 or C7, also contributed little new biological information that was not discovered using H or CP pathways (Supplementary Table S4).

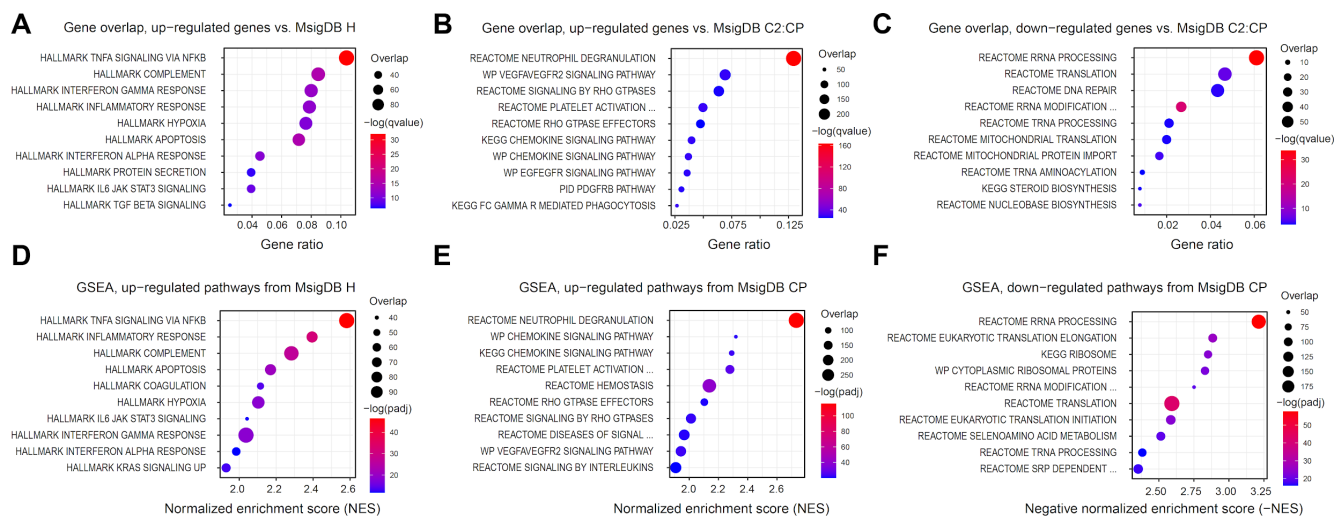


Figure 2. Gene overrepresentation and gene set enrichment (GSEA) analysis of the differentially expressed genes. Molecular signature database (MsigDB) H and CP collections were used to functionally characterize expression changes. (A-C) Top 10 significantly up- and down-regulated pathways, according to Fisher’s exact test, calculated with *clusterProfiler*. Down-regulated hallmark pathways only included one significant gene set (HALLMARK MYC TARGETS V2), and were omitted from the plot. (D-F) Top 10 significantly up- and down-regulated pathways according to GSEA, calculated with *fGSEA*. Down-regulated hallmark pathways only included one significant gene set (HALLMARK MYC TARGETS V2), and were omitted from the plot. Gene overlap indicates the number of genes in the leading edge.

Analysis of cell type composition changes based on expression signatures

While whole blood is a very rich source of metabolic and immune markers, it represents a complex tissue that is composed of numerous cell types. It is thus hard to separate effects of changes of blood cellular composition and gene expression changes within the individual cell types, both of which influence the observed differential gene expression in the bulk sample. In order to separate the two effects, we decided to leverage publicly available single cell RNA-seq datasets, generating unique expression signatures for main cell types present in whole blood. Many published datasets use human peripheral blood mononuclear cells (PBMCs). These cells, however, represent only a minor fraction of whole blood, which contains a large number of erythrocytes, platelets, and granulocytes.

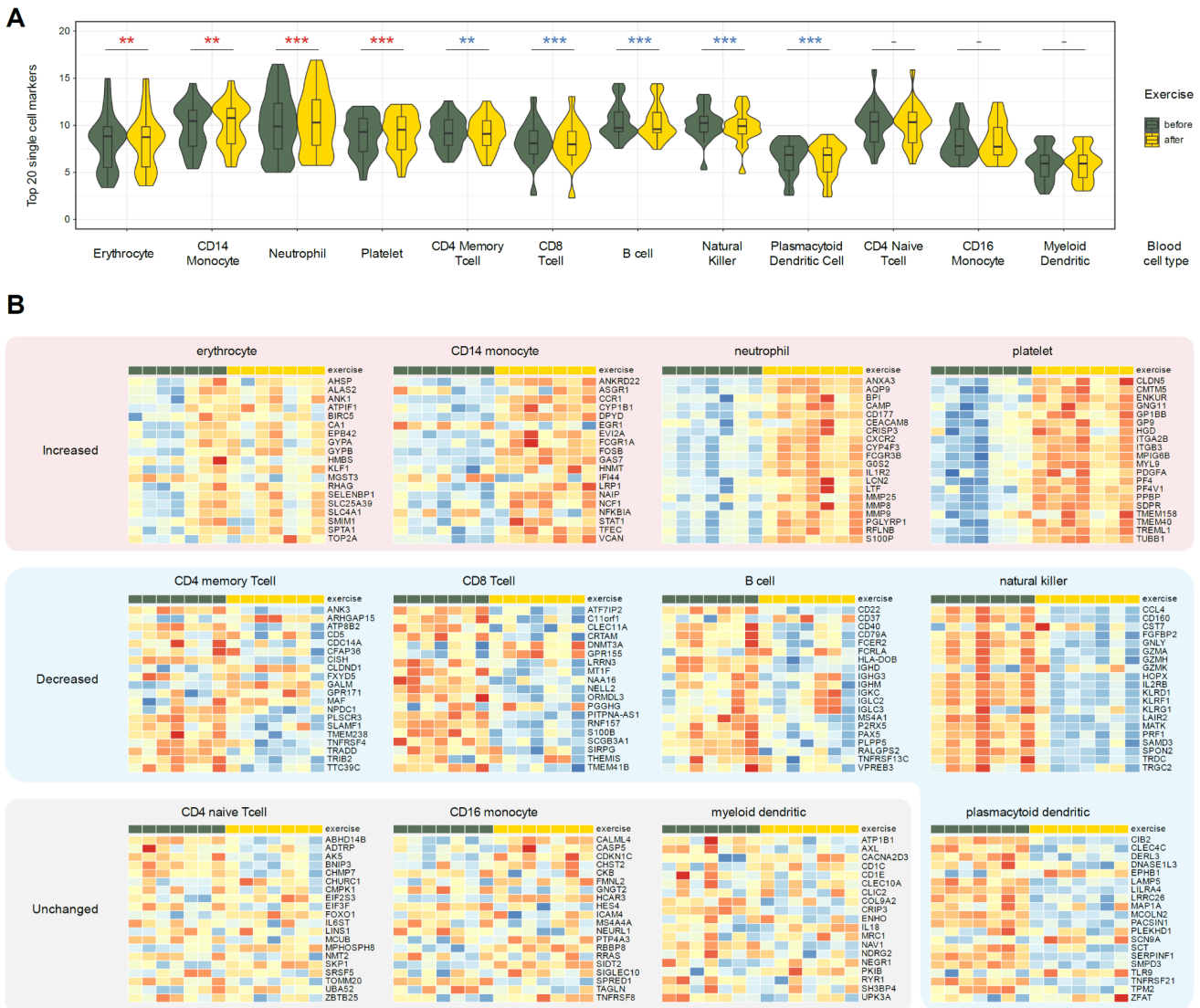


Figure 3. Using single cell markers to infer changes of the cell types from the whole blood RNA-seq data. Twenty markers specific to the 12 listed coarse-grained blood cell types were derived using two public single cell RNA-seq datasets (Methods). Donor-corrected, *rlog*-transformed expression matrix was used for all heatmap plots. (A) Observed gene expression of 20 cell type markers in all 7 profiled donors. (B) Cell types that did not display concerted change in markers: CD4⁺ naive T cells, CD16⁺ monocytes, and myeloid dendritic cells. (C) Cell types that increase after the exercise: neutrophils, CD14⁺ monocytes, platelets, and erythrocytes. (D) Cell types that decrease after the exercise: CD4⁺ memory T cells, CD8⁺ T cells, natural killer cells, B cells, and plasmacytoid dendritic cells.

Thus, we have used a publicly available scRNA-seq dataset that profiled whole blood, GSE149938 (Xie et al. 2021), to define unique cell type markers for erythrocytes and neutrophils. Additionally, we have used the 10k PBMC cells dataset from 10X Genomics (Zheng et al. 2017) to define cell type markers of coarse-grained immune cell populations. Overall, we have defined markers of 12 cell types: erythrocytes, platelets, neutrophils, natural killer cells, plasmacytoid and myeloid dendritic cells, B cells, CD4⁺ memory T cells, CD4⁺ naive T cells, CD8⁺ (cytotoxic) T cells, CD14⁺ monocytes, and CD16⁺ monocytes. We have then leveraged the top 20 most discriminating markers of each cell type as a proxy allowing us to estimate the changes in particular

cell type population from bulk RNA-seq expression profile (Figure 3). Full list of all markers is available in Supplementary Table S5.

The approach was surprisingly successful, clearly identifying putative changes in individual cell type populations. Three cell types remained relatively constant: non-classical (CD16+) monocytes, naive CD4+ T cells, and myeloid dendritic cells (Figure 3B). Populations of memory CD4+ T cells, cytotoxic CD8+ T cells, natural killer (NK), B cells, and plasmacytoid dendritic cells (pDC) were decreased, albeit the latter could be defined with the least confidence due to the extreme rarity of pDCs in peripheral blood. On the other hand, populations of neutrophils, CD14+ monocytes, erythrocytes, and platelets notably increased after exercise. This agrees with neutrophil and platelet activation pathways up-regulation in our gene set analysis (Figures 2B, 2E).

Context of other exercise expression datasets

In our experiments, we have compared whole blood transcriptome profiles before and after exercise, at the expected peak of high-altitude adaptation, 18 days since the beginning of the training camp. In order to assess the influence of high-altitude adaptation, it would have been necessary to conduct similar experiments after a similar training camp at low altitude. Unfortunately, such comparison could not be performed due to logistic restrictions. In order to find the genes uniquely regulated at high altitude, and to put our data in the context of the previously published expression datasets, we have compiled a collection of relevant publicly available data (Supplementary Table S1). The following criteria had to be satisfied for inclusion: 1) whole transcriptome profiling using microarray or RNA-seq; 2) dataset is publicly available; 3) whole blood, white blood cells, or PBMCs profiled; 4) samples from the same donor before and after a period of intense exercise were available. We have selected 8 such datasets (Büttner et al. 2007; Connolly et al. 2004; Mukherjee et al. 2014; Nakamura et al. 2010; Radom-Aizik et al. 2009a; 2009b; Tonevitsky et al. 2013; Sakharov et al. 2012). The full list of considered datasets is provided in Supplementary Table S1.

We have then re-analyzed the selected datasets using the *GEOquery* and *limma* packages (see Methods). All but one of the reprocessed datasets have shown pronounced donor effects, which had to be accounted for in the fitted linear model, and significant separation of before- and after-exercise groups was observed after the donor correction (Supplementary Figures S1-S8). The outlier dataset (Tonevitsky et al. 2013) was removed from further analysis. We have then filtered the differentially expressed genes identified in at least one of the 7 datasets from our analysis, in order to identify the genes specific to the high-altitude adaptation. Taking a conservative approach, we removed all genes deemed significant in previous studies using the unadjusted p-value. This filtering

removed 65% of the differentially expressed RNA-seq genes, leaving 685 up-regulated and 795 down-regulated genes (Figure 4A).

We next tested whether the genes uniquely up- or down- regulated in our dataset have a significant enrichment of marker genes for individual blood cell types explored in the previous section. We utilized Fisher's exact test to compare the proportion of cell type markers among all DEGs and among the uniquely regulated DEGs. We discovered a dramatic enrichment of platelet marker genes in the set of altitude-specific DEGs, with up to 90% of all differentially expressed platelet markers being uniquely regulated in our dataset ($p\text{-value} = 4 * 10^{-22}$) (Figure 4B). Erythrocytes were the only other cell type that showed a significant enrichment, though to a much lower extent ($p\text{-value} = 4 * 10^{-12}$). Neutrophil marker genes, on the other hand, were not enriched among the unique DEGs despite a very strong enrichment of neutrophil markers among all DEGs (Figure 4B). These results suggest that platelets and (to a lesser extent) erythrocytes are the two major cell types that uniquely respond to altitude exercise, while neutrophils are a more generic marker of intense exercise.

All of the used published transcriptomic datasets were microarray experiments, which are known to have lower dynamic range and sensitivity than RNA-seq (C. Wang et al. 2014). Thus, to reliably define a gene signature of exercise in high-altitude adaptation and provide an additional validation of enrichment results, we have additionally prioritized differential genes by expression ($\text{TPM} \geq 10$) and magnitude of regulation (absolute $\log_2\text{FC} \geq 0.5$). This resulted in a list of 72 genes (53 up- and 19 down-regulated). We were satisfied to discover that gene expression of cell-specific markers perfectly segregated with directions of cell populations changes defined above (Figure 3). Concordantly with our earlier observations, platelet marker genes (including the canonical *PPBP* marker) were also significantly enriched among the highly expressed altitude-specific DEGs ($p\text{-value} = 2 * 10^{-4}$). In addition to platelets, the biggest cell-type specific changes were associated with increases in neutrophils, CD14+ macrophages, and decrease in natural killer cells (Figure 4C). Aside from these, 27 genes (18 up- and 9 down-regulated) were not directly attributable to any specific cell type.

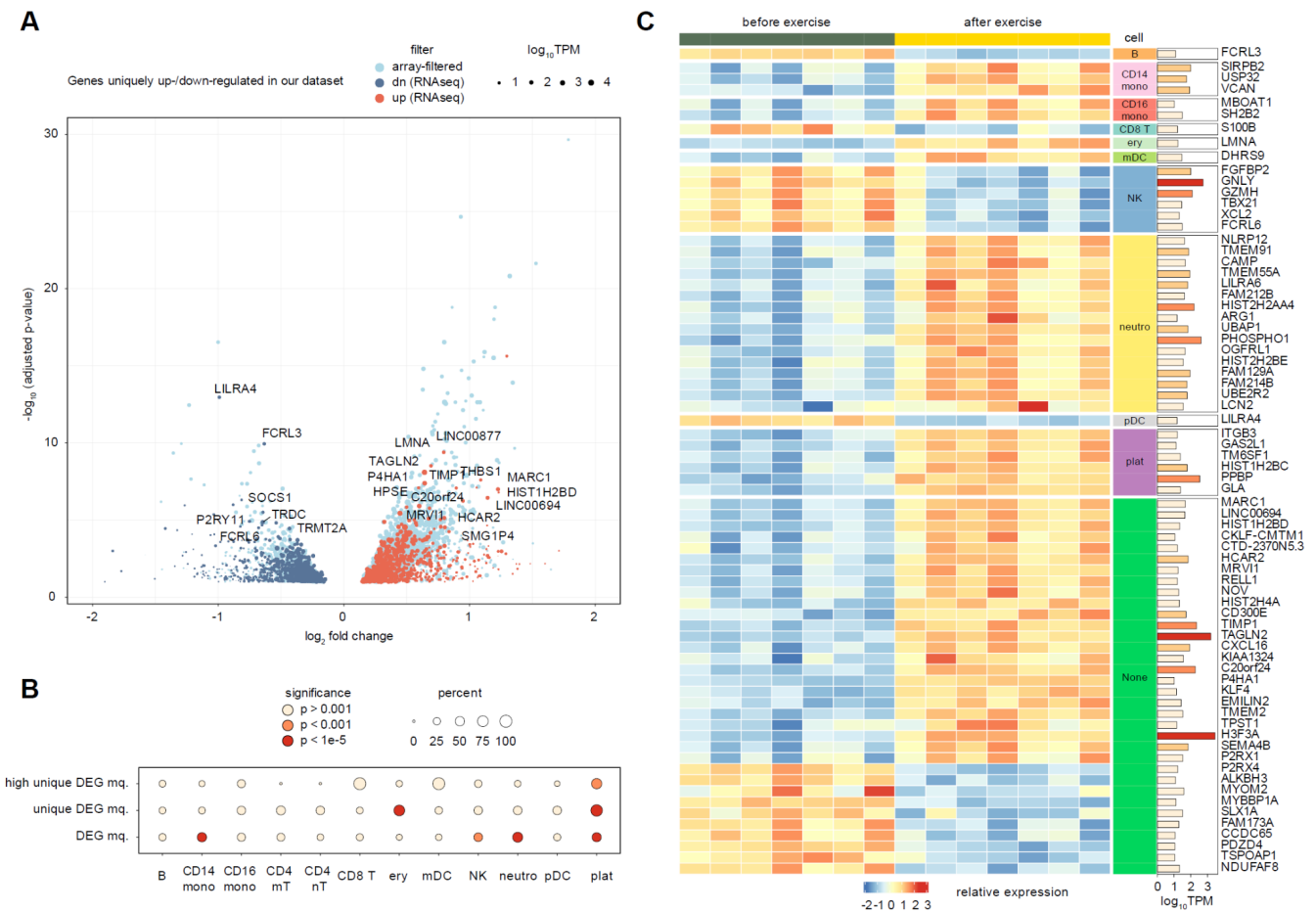


Figure 4. Analysis of genes uniquely up- or down-regulated in our dataset, as compared to 7 public microarray datasets. Full list of datasets is given in Supplementary Table S1. (A) Volcano plot of genes uniquely regulated in the whole blood of altitude-adapted skaters. Light blue points indicate differentially expressed genes previously seen in at least one other exercise dataset. Point size is scaled proportionally to mean gene expression. (B) A circle plot representing the enrichment of cell type markers among all differentially expressed genes and altitude-specific differentially expressed genes. The size of the circle is proportional to the percentage of marker genes in the target set, and the fill of the circle corresponds to the significance levels of Fisher's exact test. (C) Heatmap of highly expressed (TPM > 10) and regulated (absolute log fold change > 0.5) genes unique to our dataset. Breakdown by cell type is done based on single cell markers defined earlier. Rows are sorted by cell type and then by log fold change.

Several genes specific to high-altitude exercise (Figure 4C) stand out upon closer examination of available literature. One of the most up-regulated genes, MARC1, is a mitochondrial enzyme catalyzing the reduction of N-oxygenated molecules (Kubitza et al. 2018) that was postulated to influence the levels of endogenous nitric oxide (NO) (Kotthaus et al. 2011). Given the extremely broad and important role of NO in cardiovascular physiology in general (Loscalzo and Welch 1995), and in erythrocyte adaptation to hypoxia specifically (Zhao et al. 2018), it seems very possible that this gene plays a key role in high altitude adaptation. Human protein atlas (proteinatlas.org, (Thul and Lindskog 2018)) shows that MARC1 mRNA is enriched in sorted granulocytes and monocytes. Thus, since up-regulation of this gene was never detected in previously published exercise studies, we can hypothesize that increased MARC1 expression in these cell

types, together with general increase in granulocyte and CD14+ monocyte populations, can serve as a basis for physiological adaptation at high altitude.

Another gene of great interest is PHOSPHO1, a phosphatase which is expressed in neutrophils and eosinophils according to the Human Protein Atlas, and in neutrophils and erythrocytes according to the markers we derived from scRNAseq. Transcription of PHOSPHO1, which mediates the hydrolysis of phosphocholine to choline, was recently shown to be strongly upregulated during the terminal stages of erythropoiesis (Huang et al. 2018). Up-regulation of PHOSPHO1 caused the increased catabolism of phosphatidylcholine and phosphocholine during the terminal erythropoiesis, and its depletion caused impaired differentiation of fetal mouse and human erythroblasts. The fact that up-regulation of this gene was never detected in previous studies makes it an excellent candidate to be a key dynamic regulator of high-altitude adaptation to hypoxia.

Conclusions

Sports medicine has historically been conservative and lagged behind mainstream medicine in translation of scientific findings, often relying on “coach wisdom” or similar practices instead. This, however, has changed dramatically during the last several decades (Pujalte and Maynard 2020). There is currently a great interest in applying modern analytical techniques to athlete health surveillance, training guidance (Tanisawa et al. 2020), and even prohibited substances use monitoring (G. Wang et al. 2017). One of the debated questions is training at high elevation. There are currently numerous strategies, such as “live high, train high” (LHTH), “live high, train low” (LHTL), and many others; however, the molecular basis of this adaptation is far from understood.

In this study we have evaluated the results of whole blood RNA sequencing of elite female athletes and identified a significant number (over 4,000) genes that are up- or down-regulated as a result of vigorous exercise after high-altitude adaptation. Given the complex cellular composition of the whole blood, biological and physiological interpretation of such profound expression changes presents a formidable task. At the same time, the progress of modern gene expression profiling methods together with the growing culture of open data sharing has allowed us to make significant strides in the interpretation of the observed changes on a cellular level. The results demonstrate that the major changes associated with altitude exercise are related to innate immune response (inflammation), hypoxic stress response, and platelet activity. Using the marker genes of different blood cell types derived from public single-cell RNA sequencing data we dissected the alterations in blood cell composition, and discovered that the proportions of neutrophils, erythrocytes, CD14+ monocytes, and platelets are increased in response to exercise at altitude.

We have also leveraged a rich collection of exercise-related blood expression profiling experiments from public sources to define genes that were uniquely up- and down-regulated in our dataset. We find several notable genes that are highly expressed in blood cells and could serve as key regulatory elements responding to exercise in hypoxic conditions. These genes include PHOSPHO1, MARC1, and a number of others, including several platelet marker genes. Our analysis suggests that the majority of such platelet markers are uniquely associated with altitude exercise and are not differentially expressed in any other conditions according to published studies.

Our results provide a potential molecular link between hypoxia, platelet activity, and thrombosis. It has been long known that prolonged stay at high elevations is associated with an increased risk of thrombosis (Gupta, Zhao, and Evans 2019). Perhaps the most surprising parallel here could be related to COVID-19, which is also associated with both hypoxia and increased risk of thrombosis. It has recently been shown that physical activity influences the outcome of COVID-19 (Sallis et al. 2021). Given that thromboses are one of the dominant causes of death in COVID-19 patients, it can be hypothesized that physical activity, especially at high altitude, may serve as the pre-conditioning factor that might alleviate the relative effects of COVID-19 and prevent the negative systemic impact of platelet hyperactivity.

Acknowledgements

This research was done using equipment of Biobank of the Research Park of SPBU.

References

- Bull, Fiona C, Salih S Al-Ansari, Stuart Biddle, Katja Borodulin, Matthew P Buman, Greet Cardon, Catherine Carty, et al. 2020. "World Health Organization 2020 Guidelines on Physical Activity and Sedentary Behaviour." *British Journal of Sports Medicine* 54 (24): 1451–62. <https://doi.org/10.1136/bjsports-2020-102955>.
- Büttner, Petra, Sandy Mosig, Anja Lechtermann, Harald Funke, and Frank C. Mooren. 2007. "Exercise Affects the Gene Expression Profiles of Human White Blood Cells." *Journal of Applied Physiology* 102 (1): 26–36. <https://doi.org/10.1152/jappphysiol.00066.2006>.
- Cai, Qingqing, L. Jeffrey Medeiros, Xiaolu Xu, and Ken H. Young. 2015. "MYC -Driven Aggressive B-Cell Lymphomas: Biology, Entity, Differential Diagnosis and Clinical Management." *Oncotarget* 6 (36): 38591–616. <https://doi.org/10.18632/oncotarget.5774>.
- Connolly, Peter H., Vincent J. Caiozzo, Frank Zaldivar, Dan Nemet, Jennifer Larson, Shepin Hung, J. Denis Heck, G. Wesley Hatfield, and Dan M. Cooper. 2004. "Effects of Exercise on Gene Expression in Human Peripheral Blood Mononuclear Cells." *Journal of Applied Physiology* 97 (4): 1461–69. <https://doi.org/10.1152/jappphysiol.00316.2004>.

- Davis, S., and P. S. Meltzer. 2007. "GEOquery: A Bridge between the Gene Expression Omnibus (GEO) and BioConductor." *Bioinformatics* 23 (14): 1846–47. <https://doi.org/10.1093/bioinformatics/btm254>.
- Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Donohue, Duncan E., Aarti Gautam, Stacy-Ann Miller, Seshamalini Srinivasan, Duna Abu-Amara, Ross Campbell, Charles R. Marmar, Rasha Hammamieh, and Marti Jett. 2019. "Gene Expression Profiling of Whole Blood: A Comparative Assessment of RNA-Stabilizing Collection Methods." Edited by André Paul Gerber. *PLOS ONE* 14 (10): e0223065. <https://doi.org/10.1371/journal.pone.0223065>.
- Fabre, Odile, Lars R Ingerslev, Christian Garde, Ida Donkin, David Simar, and Romain Barrès. 2018. "Exercise Training Alters the Genomic Response to Acute Exercise in Human Adipose Tissue." *Epigenomics* 10 (8): 1033–50. <https://doi.org/10.2217/epi-2018-0039>.
- Frankish, Adam, Mark Diekhans, Anne-Maud Ferreira, Rory Johnson, Irwin Jungreis, Jane Loveland, Jonathan M Mudge, et al. 2019. "GENCODE Reference Annotation for the Human and Mouse Genomes." *Nucleic Acids Research* 47 (D1): D766–73. <https://doi.org/10.1093/nar/gky955>.
- Gjevestad, Gyrð O., Kirsten B. Holven, and Stine M. Ulven. 2015. "Effects of Exercise on Gene Expression of Inflammatory Markers in Human Peripheral Blood Cells: A Systematic Review." *Current Cardiovascular Risk Reports* 9 (7): 34. <https://doi.org/10.1007/s12170-015-0463-4>.
- Gupta, Neha, You-Yang Zhao, and Colin E. Evans. 2019. "The Stimulation of Thrombosis by Hypoxia." *Thrombosis Research* 181 (September): 77–83. <https://doi.org/10.1016/j.thromres.2019.07.013>.
- Gustafsson, Thomas, Adrian Puntchart, Lennart Kaijser, Eva Jansson, and Carl Johan Sundberg. 1999. "Exercise-Induced Expression of Angiogenesis-Related Transcription and Growth Factors in Human Skeletal Muscle." *American Journal of Physiology-Heart and Circulatory Physiology* 276 (2): H679–85. <https://doi.org/10.1152/ajpheart.1999.276.2.H679>.
- Hawley, John A., Carsten Lundby, James D. Cotter, and Louise M. Burke. 2018. "Maximizing Cellular Adaptation to Endurance Exercise in Skeletal Muscle." *Cell Metabolism* 27 (5): 962–76. <https://doi.org/10.1016/j.cmet.2018.04.014>.
- Horscroft, James A., Aleksandra O. Kotwica, Verena Laner, James A. West, Philip J. Hennis, Denny Z. H. Levett, David J. Howard, et al. 2017. "Metabolic Basis to Sherpa Altitude Adaptation." *Proceedings of the National Academy of Sciences* 114 (24): 6382–87. <https://doi.org/10.1073/pnas.1700527114>.
- Huang, Nai-Jia, Ying-Cing Lin, Chung-Yueh Lin, Novalia Pishesha, Caroline A. Lewis, Elizaveta Freinkman, Colin Farquharson, José Luis Millán, and Harvey Lodish. 2018. "Enhanced Phosphocholine Metabolism Is Essential for Terminal Erythropoiesis." *Blood* 131 (26): 2955–66. <https://doi.org/10.1182/blood-2018-03-838516>.
- Kotthaus, Jürke, Bettina Wahl, Antje Havemeyer, Joscha Kotthaus, Dennis Schade, Dieter Garbe-Schönberg, Ralf Mendel, Florian Bittner, and Bernd Clement. 2011. "Reduction of N ω -Hydroxy-L-Arginine by the Mitochondrial Amidoxime Reducing Component (MARC)." *Biochemical Journal* 433 (2): 383–91. <https://doi.org/10.1042/BJ20100960>.
- Kubitza, Christian, Florian Bittner, Carsten Ginsel, Antje Havemeyer, Bernd Clement, and Axel J. Scheidig. 2018. "Crystal Structure of Human MARC1 Reveals Its Exceptional Position among Eukaryotic Molybdenum Enzymes." *Proceedings of the*

- National Academy of Sciences* 115 (47): 11958–63.
<https://doi.org/10.1073/pnas.1808576115>.
- Larrabee, R. C. 1902. “Leucocytosis after Violent Exercise.” *The Journal of Medical Research* 7 (1): 76–82.
- Leek, Jeffrey T., W. Evan Johnson, Hilary S. Parker, Andrew E. Jaffe, and John D. Storey. 2012. “The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments.” *Bioinformatics* 28 (6): 882–83.
<https://doi.org/10.1093/bioinformatics/bts034>.
- Lewis, G. D., L. Farrell, M. J. Wood, M. Martinovic, Z. Arany, G. C. Rowe, A. Souza, et al. 2010. “Metabolic Signatures of Exercise in Human Plasma.” *Science Translational Medicine* 2 (33): 33ra37–33ra37. <https://doi.org/10.1126/scitranslmed.3001006>.
- Li, Bo, and Colin N Dewey. 2011. “RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome.” *BMC Bioinformatics* 12 (1): 323.
<https://doi.org/10.1186/1471-2105-12-323>.
- Liberzon, A., A. Subramanian, R. Pinchback, H. Thorvaldsdottir, P. Tamayo, and J. P. Mesirov. 2011. “Molecular Signatures Database (MSigDB) 3.0.” *Bioinformatics* 27 (12): 1739–40. <https://doi.org/10.1093/bioinformatics/btr260>.
- Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. 2015. “The Molecular Signatures Database Hallmark Gene Set Collection.” *Cell Systems* 1 (6): 417–25. <https://doi.org/10.1016/j.cels.2015.12.004>.
- Loscalzo, Joseph, and George Welch. 1995. “Nitric Oxide and Its Role in the Cardiovascular System.” *Progress in Cardiovascular Diseases* 38 (2): 87–104.
[https://doi.org/10.1016/S0033-0620\(05\)80001-5](https://doi.org/10.1016/S0033-0620(05)80001-5).
- Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Moore, Lorna G. 2017. “Measuring High-Altitude Adaptation.” *Journal of Applied Physiology* 123 (5): 1371–85. <https://doi.org/10.1152/jappphysiol.00321.2017>.
- Mooren, F. C., D. Blöming, A. Lechtermann, M. M. Lerch, and K. Völker. 2002. “Lymphocyte Apoptosis after Exhaustive and Moderate Exercise.” *Journal of Applied Physiology* 93 (1): 147–53. <https://doi.org/10.1152/jappphysiol.01262.2001>.
- Mukherjee, Kamalika, Brittany A. Edgett, Harrison W. Burrows, Cecilia Castro, Julian L. Griffin, Adel Giaid Schwertani, Brendon J. Gurd, and Colin D. Funk. 2014. “Whole Blood Transcriptomics and Urinary Metabolomics to Define Adaptive Biochemical Pathways of High-Intensity Exercise in 50-60 Year Old Masters Athletes.” Edited by Daniel Monleon. *PLoS ONE* 9 (3): e92031.
<https://doi.org/10.1371/journal.pone.0092031>.
- Nakamura, Seiji, Michie Kobayashi, Tomohiro Sugino, Osami Kajimoto, Ryo Matoba, and Kenichi Matsubara. 2010. “Effect of Exercise on Gene Expression Profile in Unfractionated Peripheral Blood Leukocytes.” *Biochemical and Biophysical Research Communications* 391 (1): 846–51.
<https://doi.org/10.1016/j.bbrc.2009.11.150>.
- Nieman, David C., and Brandt D. Pence. 2020. “Exercise Immunology: Future Directions.” *Journal of Sport and Health Science* 9 (5): 432–45.
<https://doi.org/10.1016/j.jshs.2019.12.003>.
- Nieman, David C., and Laurel M. Wentz. 2019. “The Compelling Link between Physical Activity and the Body’s Defense System.” *Journal of Sport and Health Science* 8 (3): 201–17. <https://doi.org/10.1016/j.jshs.2018.09.009>.
- Pujalte, George G.A., and Jennifer R. Maynard. 2020. “The Increasing Importance of Sports Science and Medicine.” *Journal of International Medical Research* 48 (1): 030006051982769. <https://doi.org/10.1177/0300060519827694>.

- Radom-Aizik, Shlomit, Frank Zaldivar, Szu-Yun Leu, and Dan M Cooper. 2009a. "Brief Bout of Exercise Alters Gene Expression in Peripheral Blood Mononuclear Cells of Early- and Late-Pubertal Males." *Pediatric Research* 65 (4): 447–52. <https://doi.org/10.1203/PDR.0b013e3181993473>.
- Radom-Aizik, Shlomit, Frank Zaldivar, Szu-Yun Leu, and Dan M. Cooper. 2009b. "A Brief Bout of Exercise Alters Gene Expression and Distinct Gene Pathways in Peripheral Blood Mononuclear Cells of Early- and Late-Pubertal Females." *Journal of Applied Physiology* 107 (1): 168–75. <https://doi.org/10.1152/japplphysiol.00121.2009>.
- Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7): e47–e47. <https://doi.org/10.1093/nar/gkv007>.
- Sakharov, Dmitry A., Diana V. Maltseva, Evgeniy A. Riabenko, Maxim U. Shkurnikov, Hinnak Northoff, Alexander G. Tonevitsky, and Anatoly I. Grigoriev. 2012. "Passing the Anaerobic Threshold Is Associated with Substantial Changes in the Gene Expression Profile in White Blood Cells." *European Journal of Applied Physiology* 112 (3): 963–72. <https://doi.org/10.1007/s00421-011-2048-3>.
- Sallis, Robert, Deborah Rohm Young, Sara Y Tartof, James F Sallis, Jeevan Sall, Qiaowu Li, Gary N Smith, and Deborah A Cohen. 2021. "Physical Inactivity Is Associated with a Higher Risk for Severe COVID-19 Outcomes: A Study in 48 440 Adult Patients." *British Journal of Sports Medicine*, April, bjsports-2021-104080. <https://doi.org/10.1136/bjsports-2021-104080>.
- Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019. "Comprehensive Integration of Single-Cell Data." *Cell* 177 (7): 1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>.
- Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, et al. 2005. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences* 102 (43): 15545–50. <https://doi.org/10.1073/pnas.0506580102>.
- Tanisawa, Kumpei, Guan Wang, Jane Seto, Ioanna Verdouka, Richard Twycross-Lewis, Antonia Karanikolou, Masashi Tanaka, et al. 2020. "Sport and Exercise Genomics: The FIMS 2019 Consensus Statement Update." *British Journal of Sports Medicine* 54 (16): 969–75. <https://doi.org/10.1136/bjsports-2019-101532>.
- Terry, Erin E, Xiping Zhang, Christy Hoffmann, Laura D Hughes, Scott A Lewis, Jiajia Li, Matthew J Wallace, et al. 2018. "Transcriptional Profiling Reveals Extraordinary Diversity among Skeletal Muscle Tissues." *ELife* 7 (May): e34613. <https://doi.org/10.7554/eLife.34613>.
- Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov. 2013. "Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration." *Briefings in Bioinformatics* 14 (2): 178–92. <https://doi.org/10.1093/bib/bbs017>.
- Thul, Peter J., and Cecilia Lindskog. 2018. "The Human Protein Atlas: A Spatial Map of the Human Proteome: The Human Protein Atlas." *Protein Science* 27 (1): 233–44. <https://doi.org/10.1002/pro.3307>.
- Timp, Winston, and Gregory Timp. 2020. "Beyond Mass Spectrometry, the next Step in Proteomics." *Science Advances* 6 (2): eaax8978. <https://doi.org/10.1126/sciadv.aax8978>.
- Tonevitsky, Alexander G, Diana V Maltseva, Asghar Abbasi, Timur R Samatov, Dmitry A Sakharov, Maxim U Shkurnikov, Alexey E Lebedev, Vladimir V Galatenko, Anatoly I Grigoriev, and Hinnak Northoff. 2013. "Dynamically Regulated MiRNA-MRNA

- Networks Revealed by Exercise.” *BMC Physiology* 13 (1): 9.
<https://doi.org/10.1186/1472-6793-13-9>.
- Wang, Charles, Binsheng Gong, Pierre R Bushel, Jean Thierry-Mieg, Danielle Thierry-Mieg, Joshua Xu, Hong Fang, et al. 2014. “The Concordance between RNA-Seq and Microarray Data Depends on Chemical Treatment and Transcript Abundance.” *Nature Biotechnology* 32 (9): 926–32. <https://doi.org/10.1038/nbt.3001>.
- Wang, Guan, Jérôme Durussel, Jonathan Shurlock, Martin Mooses, Noriyuki Fuku, Georgie Bruinvels, Charles Pedlar, et al. 2017. “Validation of Whole-Blood Transcriptome Signature during Microdose Recombinant Human Erythropoietin (RHuEpo) Administration.” *BMC Genomics* 18 (S8): 817. <https://doi.org/10.1186/s12864-017-4191-7>.
- Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. “RNA-Seq: A Revolutionary Tool for Transcriptomics.” *Nature Reviews Genetics* 10 (1): 57–63.
<https://doi.org/10.1038/nrg2484>.
- Xie, Xiaowei, Mengyao Liu, Yawen Zhang, Bingrui Wang, Caiying Zhu, Chenchen Wang, Qing Li, et al. 2021. “Single-Cell Transcriptomic Landscape of Human Blood Cells.” *National Science Review* 8 (3): nwaa180. <https://doi.org/10.1093/nsr/nwaa180>.
- Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. “ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters.” *OMICS: A Journal of Integrative Biology* 16 (5): 284–87.
<https://doi.org/10.1089/omi.2011.0118>.
- Zhang, Xi-wu, Qiu-han Li, Zuo-di Xu, and Jin-jin Dou. 2020. “Mass Spectrometry-Based Metabolomics in Health and Medical Science: A Systematic Review.” *RSC Advances* 10 (6): 3092–3104. <https://doi.org/10.1039/C9RA08985C>.
- Zhao, Yajin, Xiang Wang, Milody Noviana, and Man Hou. 2018. “Nitric Oxide in Red Blood Cell Adaptation to Hypoxia.” *Acta Biochimica et Biophysica Sinica* 50 (7): 621–34. <https://doi.org/10.1093/abbs/gmy055>.
- Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. “Massively Parallel Digital Transcriptional Profiling of Single Cells.” *Nature Communications* 8 (1): 14049.
<https://doi.org/10.1038/ncomms14049>.