# Title: The blood transcriptional signature of recombinant human erythropoietin administration and implications for anti-doping strategies

**Authors:** Jérôme Durussel[1], Diresibachew W. Haile[2], Kerli Mooses[3], Evangelia Daskalaki[4], Wendy Beattie[1], Martin Mooses[3], Wondyefraw Mekonen[2], Neford Ongaro[5], Edwin Anjila[5], Rajan K. Patel[1], Neal Padmanabhan[1], Martin W. McBride[1], John D. McClure[1], and Yannis P. Pitsiladis[6*].

**Affiliations:** [1]Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom; [2]Department of Medical Physiology, Addis Ababa University, Addis Ababa, Ethiopia; [3]Faculty of Sport and Exercise Sciences, University of Tartu, Tartu, Estonia; [4]Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, United Kingdom;[5]Department of Medical Physiology, School of Medicine, College of Health Sciences, Moi University, Eldoret, Kenya; and [6]Centre for Sport and Exercise Science and Medicine (SESAME), University of Brighton, Eastbourne, United Kingdom.

***Corresponding Author:** Professor Yannis P. Pitsiladis
Professor of Sport and Exercise Science, Director of the Centre
Centre for Sport and Exercise Science and Medicine (SESAME)
Hillbrow, Denton Road
Eastbourne, BN20 7SP, United Kingdom
Email y.pitsiladis@brighton.ac.uk
Telephone +44 (0) 1273 643707 (Reception)
Fax +44 (0) 1273 643704

**Running Head:** Transcriptional signature of rHuEPO

**Key Words:** Blood doping | Erythropoiesis | mRNA

**ABSTRACT:**

**Background:** Recombinant human erythropoietin (rHuEPO) is frequently abused by athletes as a performance-enhancing drug, despite being prohibited by the World Anti-Doping Agency. Although the methods to detect blood doping, including rHuEPO injections, have improved in recent years, they remain imperfect.

**Methods:** In a proof-of-principle study, we identified, replicated and validated the whole-blood transcriptional signature of rHuEPO in endurance-trained Caucasian males at sea-level (n = 18) and Kenyan endurance runners at moderate altitude (n = 20), all of whom received rHuEPO injections for four weeks.

**Results:** Transcriptional profiling shows that hundreds of transcripts were altered by rHuEPO in both cohorts. The main regulated expression pattern, observed in all participants, was characterised by a "rebound" effect with a profound up-regulation during rHuEPO and a subsequent down-regulation up to four weeks post administration. The functions of the identified genes were mainly related to the functional and structural properties of the red blood cell. Of the genes identified to be differentially expressed during and post rHuEPO, we further confirmed a whole blood 34-transcript signature that can distinguish between samples collected pre, during and post rHuEPO administration.

**Conclusion:** By providing biomarkers that can reveal rHuEPO use, our findings represent an advance in the development of new methods for the detection of blood doping.

**INTRODUCTION**

Erythropoietin is a hormone that regulates the production of new red blood cells (erythropoiesis) (18). Administration of recombinant human erythropoietin (rHuEPO) improves sporting performance (10) and hence is frequently subject to abuse by athletes, despite being prohibited by the World Anti-Doping Agency (WADA). The direct detection of rHuEPO doping is challenging because rHuEPO is structurally very similar to endogenous erythropoietin and rapidly disappears from the circulation. The Athlete Biological Passport (ABP) was introduced as a new tool to provide indirect evidence for the use of erythropoiesis-stimulating agents such as rHuEPO by identifying abnormal intra-individual variability in selected blood parameters over time (34). While approaches to detect blood doping, including rHuEPO injections, are improving, the declarations of confessed dopers emphasize the need for further refinement of current detection methods.

The typical goal of ‘omics’-based research is the identification of new and specific molecular biomarkers for the diagnosis, monitoring and prognosis of diseases. The same approach can be used for monitoring the efficacy of drug therapy (22). Here, we applied an ‘omics’-based approach in human volunteers to generate a transcriptional signature of rHuEPO for anti-doping purposes. We first describe the specific alterations of the blood transcriptional profiles induced by rHuEPO administration in two distinct and independent groups. We then validate a whole blood 34-transcript signature using another quantitative gene technology and performed in a subset of fifty selected genes. Finally, we demonstrate the implications of our findings for anti-doping strategies. The identification of the blood transcriptional signature of rHuEPO provides the basis for the development of new and improved models to detect blood doping. Our study also provides a broad range of transcriptional biomarkers relevant for understanding the variability in responsiveness to rHuEPO therapy.

## METHODS

### Experimental design

Eighteen endurance-trained Caucasian males at sea-level (SCO; 26.0 ± 4.5 years, 74.8 ± 7.9 kg, 179.8 ± 5.4 cm) (10) and twenty east African endurance runners from the Kalenjin tribe in Kenya (KEN; 26.4 ± 4.1 years, 56.6 ± 4.7 kg, 171.8 ± 6.4) at moderate altitude (~2150 m) participated in the study. All subjects underwent a medical assessment and provided written informed consent to participate. Subjects were requested to maintain their normal training but abstain from official sporting competition for the duration of the research study (16). This study was approved by the Ethics Committees of the University of Glasgow (Scotland, UK) and Moi University (Kenya) and conformed to the Declaration of Helsinki. The subjects subcutaneously self-injected 50 IU·kg$^{-1}$ body mass of rHuEPO (NeoRecormon, Roche) every second day for four weeks (Figure 1A). Daily oral iron supplementation (~ 100 mg of elemental iron, Ferrous Sulphate Tablets, Almus) was given during the four weeks of rHuEPO administration.

### Blood sampling and RNA extraction

After ten minutes of rest in the supine position, 3 mL of whole blood was collected into Tempus Blood RNA tubes (Life Technologies) from an antecubital vein in triplicate at baseline, during rHuEPO administration and for four weeks after rHuEPO administration. The Tempus tubes were vigorously mixed immediately after collection, incubated at room temperature for three hours and stored between -20˚C and -80˚C before RNA extraction. Eight time points per subjects were selected for gene expression analysis (Figure 1A-B). RNA was extracted using the MagMAX for stabilized blood tubes RNA isolation kit (Life Technologies) or the Tempus Spin RNA isolation kit (Life Technologies). RNA yield was

determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific) and the RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent technologies) showing a quality of RNA integrity number of 8.5 ± 0.9 (95% confidence intervals (CI) 8.4 to 8.6 and ranging from 5.0 to 9.8). RNA was stored at -80°C until further analysis.

**Microarray experiment and data analysis**

Eighteen subjects in each cohort were selected for the microarray experiment. Extracted RNA was prepared using the Illumina TotalPrep RNA Amplification Kit (Life Technologies). The RNA amplification consists of synthesizing cRNA by in vitro transcription from the cDNA produced by reverse transcription from 500 ng of RNA. 750 ng of the purified labeled cRNA samples were then randomly hybridized to the Illumina HumanHT-12 v4.0 Expression BeadChip arrays, which contain more than 47,000 probes, following the manufacturers' recommended procedures (Illumina). The Bead arrays were scanned on the Illumina BeadArray Reader and raw intensity values were saved in Illumina GenomeStudio software. Following the manufacturer's instructions Illumina GenomeStudio quality control metrics were used to assess the quality of the samples; all samples passed these assessments.

For normalisation, Illumina GenomeStudio software was used to subtract background and to perform quantile normalisation. Using R software environment (http://www.r-project.org/), normalised probe level intensities were then log2 transformed after adding a small constant to avoid negative values as well as to reduce the influence of probes with low signal intensities close to the background noise. The constant's value was chosen after inspection of scatter plots of the two baseline samples from each individual. To remove the array effect a linear model was fitted to the each gene's data using array as the predictor, and the coefficients for each array were removed from each value. These normalised data were

used for all downstream analyses. Rank Products analysis (4) with a 5% false discovery rate (3) was used to identify the differentially expressed transcripts during and post rHuEPO administration compared to baseline values. An additional 1.5 fold-change threshold was applied for a more stringent analysis. The differences in genes expression between the two cohorts by time-interaction were analysed using R-package Limma (linear models for microarray) (33) with a 5% false discovery rate (3). Functional analysis of differentially expressed genes was performed using Ingenuity Pathway Analysis (Ingenuity Systems) with Fisher's exact test to calculate a *P* value determining the probability that each biological function and/or network assigned to the identified genes was due to chance alone, with a Benjamini-Hochberg correction for multiple testing applied. All microarray data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-2874.

**QuantiGene Plex experiment and data analysis**

QuantiGene Plex assay (Affymetrix) was used for further confirmation of the microarray results. The QuantiGene Plex assay combines branched DNA (bDNA) signal amplification and xMAP profiling magnetic beads technologies (Luminex). The bDNA enables target RNA quantification by amplifying the reporter signal rather than target sequences and the xMAP fluorescent microspheres or "beads" are used as a support to capture specific RNA molecules (12). The MicroArray Quality Control (MAQC) project has demonstrated the good precision and accuracy of the QuantiGene Plex assay (5).

A subset of fifty identified genes was selected for confirmation using QuantiGene Plex. The selection of these fifty genes was based on the following criteria: magnitude of the changes in expression; relative expression above background level; low inter-individual responses in gene expression pattern; consistency between the Illumina probes targeting

different transcripts for the same gene; biological relevance. Eighteen and twenty subjects in the SCO and KEN cohorts were selected for the QuantiGene Plex experiment, respectively. Two different RNA inputs of 100 and 200 ng for each sample were hybridized and amplified using the QuantiGene Plex assay and following the manufacturers' recommended procedures (Affymetrix). The fluorescence signal was read on the MAGPIX (Luminex) and median fluorescence intensity (MFI) values were saved using xPonent software. Data were exported to R software environment where background subtraction, log2 transformation and normalisation to four housekeeping genes (*ACTB*, *ACTR10*, *MRFAP1* and *RAB11A*) were performed. These normalised data were used for all downstream analyses. R-package Limma (33) with a 5% false discovery rate (3) was used to confirm expression profiles.

**Blood parameters analysis**

Haematology analysis including reticulocyte percentage and haematocrit was performed from homogenized whole blood collected in EDTA tubes using fluorescence flow cytometry and hydrodynamic focusing (Sysmex XT-2000i, Sysmex). Reticulocyte percent data were exported to R software environment where log ratio compared to baseline values was calculated. The reticulocyte effect was controlled by regression modelling in the R-package Limma (33) using the regression coefficient as a scaling factor. The curve of the relationship between reticulocyte and genes which remained significantly differentially expressed after controlling for the effect of reticulocyte was fitted using polynomial local regression (loess).

**Prediction using k-nearest neighbour**

For class prediction and estimation of the discriminatory performance, we used the k-nearest neighbours method from the R-package "class", with six neighbours. The 34 genes

validated by QuantiGene Plex were used for prediction of samples collected pre, during and post rHuEPO administration as well as for prediction of samples collected at 2 days after the first rHuEPO injection. The prediction model was trained using five-fold cross-validation on thirty randomly selected subjects (training set: SCO; n = 14 and KEN; n = 16). This model was then used to predict the classification of the samples in the remaining eight subjects (test set: SCO; n = 4 and KEN; n = 4) and to determine sensitivity, specificity and 95% CI.

## RESULTS:

### rHuEPO administration profoundly alters blood transcriptional profiles

To identify changes in the blood transcriptome after rHuEPO administration, we first generated genome-wide whole-blood transcriptional profiles from the SCO cohort. The previously published phenotypic results confirmed that rHuEPO administration increased red blood cell mass and improved exercise performance (10). Transcriptional profiling shows that hundreds of transcripts were altered by rHuEPO administration (Figure 1C). Using a combination of changes in expression-level and robust statistical criteria, we identified a 41-transcript signature, distinct from baseline values, at two days after the first injection that remained altered throughout rHuEPO administration (Figure 1B-C and supplementary file Data Set S1). At the same time, a distinct 10-transcript signature was defined by differential expression after rHuEPO administration that remained differentially expressed up to four weeks post administration (Figure 1B-C and supplementary file Data Set S2).

### Similar changes in blood transcriptional profiles were observed in all individuals

In order to replicate our findings, we then applied a similar genome-wide transcriptional approach to the independent KEN cohort who underwent a similar rHuEPO administration regimen. The initial findings were replicated and only minimal significant

differences were found between the two groups. More precisely, only three transcripts (*DARC*, *JAK3* and *CCDC71*) were found to be differentially regulated between SCO and KEN.

Of the genes already differentially expressed two days after the first injection compared to baseline in the SCO and KEN cohorts, 32 transcripts were commonly regulated, while 5 transcripts were commonly regulated post administration in the two groups (Table 1, Figure 1C and supplementary files Data Sets S1-S4). The main regulated expression pattern of the identified signature was characterised by a "rebound" effect with a profound up-regulation during rHuEPO and a subsequent down-regulation up to four weeks post administration. The same transcriptional signature pattern was observed in all participants. *CA1* gene was selected to exemplify the signature pattern observed in all 32 transcripts (Figure 1D and supplementary files Data Sets S1-S4).

**Whole-blood transcriptional signature of rHuEPO reflects distinct changes in gene expression and production of new red blood cells**

The genes profoundly up-regulated during rHuEPO and subsequently down-regulated post administration were mainly related to haematological, haematopoietic and cardiovascular system development and function (Figure 2A), such as the differentiation of the red blood cells (Figure 2B) and in agreement with the observed changes in blood parameters (Figure 2C). Validation, using another quantitative gene technology and performed in a subset of fifty selected genes, confirmed the whole blood 34-transcript signature (Figure 3 and Table 2).

Although reticulocytes have shed their nucleus, they retain quantities of residual nucleic acid material (11). As such, we tested the hypothesis that the identified whole blood 34-transcript signature may reflect the relative proportion of immature red blood cells

(reticulocytes) compared to white blood cells. The rHuEPO-induced changes in expression observed in the identified genes were correlated with the changes in reticulocyte percentage (Table 2). Nevertheless, even after controlling for the effect of reticulocyte, changes in gene expression compared to baseline values still reached significance levels in several genes (Figure 4 and Table 2). In particular, four genes (*EPB42*, *SELENBP1*, *SLC4A1*, and *PITHD1*) showed that the blood transcriptional signature can be more sensitive than reticulocyte percentage in detecting changes in blood profiles induced by only one rHuEPO injection (Figure 4 and Table 2).

**Discriminatory performance and anti-doping implications**

The k-nearest neighbour prediction model was able to distinguish between samples collected pre, during and post rHuEPO administration with overall sensitivity and specificity of 79.4% (95% CI 74.8 to 83.9%) and 94.7% (95% CI 92.2 to 97.3%), respectively (Table 3). Finally, and most importantly for current anti-doping challenges, the k-nearest neighbour prediction gave a sensitivity of 58.6% (95% CI 42.7 to 74.4%) and a specificity of 97.4% (95% CI 93.8 to 100%) for specifically distinguishing samples collected at baseline and at two days after only one rHuEPO injection, which would typically remain undetectable by the current tests using standard blood parameters (1).

**DISCUSSION AND CONCLUSIONS:**

rHuEPO is frequently abused by athletes as a performance-enhancing drug, despite being prohibited by WADA. Using transcriptional profiling from whole-blood, we provide proof-of-principle for a new transcriptionally-enhanced model for the detection of rHuEPO doping. In this study, we first identified the changes in blood transcriptional profiles in perturbed erythropoiesis via rHuEPO administration in endurance-trained individuals in

Scotland. The results were then replicated in KEN and only minimal significant differences were found with SCO. Some of the only few differences observed between the two groups, such as in *DARC* and *JAK3*, may be explained by gene polymorphisms. For instance, polymorphisms in *DARC* are the basis of the Duffy blood group system which is associated with resistance to *Plasmodium vivax* malaria (44). Using another quantitative gene technology and performed in a subset of fifty selected genes, we then validated the whole blood 34-transcript signature. Finally, we tested the discriminatory performance of this blood transcriptional signature. We anticipate our results to be the starting point for a new transcriptionally-enhanced model for the detection of blood doping.

**Biological functions of the identified signature**

The whole-blood transcriptional signature of rHuEPO administration showed biologically coherent changes in gene expression. In particular, the functional and structural components of the red blood cells were dominant in the 34-transcript signature; for example, the haem synthesis (*ALAS2*, *FECH* and *SNCA*) (28, 30), the transport of oxygen and carbon dioxide (*HBD*, *BPGM*, *CA1* and *SLC4A1*) (13, 14, 24), the organisation and connectivity of the red blood cell membrane (*EPB42*, *TMOD1*, *GYPE* and *RBM38*) (15, 17, 29, 36), the control of reactive oxygen species which can potentially induce cell apoptosis (*BCL2L1* and *SELENBP1*) (26, 32), the regulation of the cell cycle and the differentiation of blood cells during erythropoiesis (*FBXO7*) (7), the purine metabolism (*GMPR* and *GUK1*), as well as genes of unknown function in the red blood cell but which were previously reported to be relevant to the red blood cell biology (*TRIM58*, *CSDA*, *DCAF12*, *FAM46C*, *STRADB, ADIPOR1*, *PITHD1* and *UBXN6*) (6, 37, 40, 43). These findings support the view that the gene set identified here is enriched for genes involved in the biological processes related to the red blood cell, particularly during late-stage erythropoiesis.

From a clinical perspective, rHuEPO and other erythropoiesis stimulating agents are widely used in the treatment of anaemia. The introduction of rHuEPO was a paradigm-shift in the treatment of patients with anaemia due to, for example, chronic kidney disease as it dramatically reduced the use of blood transfusions and their associated morbidity (20). However, there is a need to better understand the variability as well as to predict and improve responsiveness to rHuEPO therapy (20). As such, our data provides a series of candidate biomarkers relevant to investigating the variability in response to rHuEPO treatment.

**Detection of early changes in erythropoiesis**

In the bone marrow, erythropoietin binds to the erythropoietin receptor and activates the Jak2-STAT5 signalling cascade which then stimulates the survival, proliferation and differentiation of red blood cell progenitors (18) without affecting the white blood cells production (19). Following the initial development and maturation in the bone marrow, the reticulocytes are then released in the blood stream before developing into mature red blood cells. Because reticulocytes still retain quantities of residual nucleic acid material (11), the identified whole blood 34-transcript signature may reflect the relative proportion of immature red blood cells (reticulocytes) compared to white blood cells. The correlation between the rHuEPO-induced changes in gene expression with the changes in reticulocyte percentage is in agreement with previously reported correlation between red cell distribution width, which often reflects a higher percentage of reticulocytes, and a cluster of 57 genes that include several genes also identified in the present study, such as *EPB42*, *SLC4A1*, *ALAS2*, *TMOD*, *CSDA*, *GMPR* and *BPGM* (41). Nonetheless, changes in gene expression still reached significance levels in several genes even after controlling for the effect of reticulocyte.

The observation that the blood transcriptional signature, in particular four genes (*EPB42*, *SELENBP1*, *SLC4A1*, and *PITHD1*), can be more sensitive than reticulocyte

percentage in detecting early changes in erythropoiesis in whole blood is particularly relevant for the challenge that faces current anti-doping methods to detect small doses of rHuEPO. Indeed, in order to minimise the risk of being caught via the ABP, it is well recognised that athletes are now using so-called "microdoses" of rHuEPO which allegedly range from 10 to a maximum of 40 $IU \cdot kg^{-1}$ body mass. These small or "microdoses" of rHuEPO aim to increase red blood cell mass while avoiding large fluctuation in the ABP blood markers and/or "normalising" these markers after blood manipulations such as autologous blood transfusion, as well as minimising the detection window for conventional direct methods (23). However, the ABP did not detect a single case of rHuEPO when a microdosing strategy was adopted (1). The k-nearest neighbour prediction results for specifically distinguishing samples collected at baseline and at two days after only one rHuEPO injection are promising results in order to detect a small dose of rHuEPO. The k-nearest neighbour classifier is a robust method that has been shown to be very effective in gene expression classification (9). However, it is essentially a naïve machine learning algorithm that does not take into account when the sample was taken, for example. There thus remains scope for the discriminatory performance to be even further improved by using more sophisticated methods, such as the ABP adaptive Bayesian model (35), once normal physiological molecular reference profiles are generated.

**Study limitations**

The fact that the present study was not blinded and did not include a control group may limit the interpretations of the direct applications of the findings. However, participants were able to act as their own controls to an extent and any placebo effects are unlikely to have yielded the extent of differential expression and the characteristic profiles observed in differentially regulated genes. In addition, based on the recent admissions by confessed dopers, the rHuEPO dose regimen used in this study (*i.e.* 50 $IU \cdot kg^{-1}$ body mass every two

days for four weeks) does not closely reproduce the strategy of "microdoses" rHuEPO doping typically currently used by athletes (23). However, this study was not designed to exactly mimic drug taking behaviour but rather it was designed as a proof-of-principle of the transcriptomic approach to detection of rHuEPO. Further, as a consequence of the promising results discussed here, we have already designed a randomised, double-blind, placebo-controlled cross-over study involving microdosing (25).

**Implications for anti-doping strategies and conclusions**

Earlier studies, limited by small number of subjects or animals and using the serial analysis of gene expression (SAGE) method, already reported some modifications of blood gene expression after rHuEPO administration (2, 38, 39). Here we provide the first complete description of the human blood transcriptional signature of altered erythropoiesis in response to rHuEPO administration. The k-nearest neighbour approach can distinguish between samples collected pre, during and post rHuEPO administration. Moreover, a transcriptionally-enhanced model can potentially reduce the logistical challenges and cost incurred in sample collection and analysis. Current anti-doping blood samples have important logistic and cost implications since blood samples must be shipped refrigerated (4ºC) and analysed within 48 h of collection (42), whereas the blood samples, RNA stabilised in Tempus tubes (Life Technologies) of a transcriptionally-enhanced model are easy to handle and stable for 5-7 days at room temperature and for years when kept frozen prior to analysis (8, 21, 27, 31). However, further validations using rHuEPO doping regimen more similar to what is currently thought to be used, with smaller doses and more sophisticated ABP-like modelling methods as well as the generation of normal physiological reference ranges, are required before the blood transcriptional signature-based approach delivers on its promise to add a new dimension to the current detection methods used by WADA.  These promising results might

also mean that other banned substances could be detectable with transcriptomic approaches. However, their effectiveness may vary according to the substance considered and the tissue type available.

In summary, our results provide evidence to support a transcriptionally enhanced approach for the detection of blood doping. Our study also provides new insights into a broad range of transcriptional biomarkers relevant for understanding the biological mechanisms involved in erythropoiesis.

**Authorship Contributions**

Y.P.P. designed the study with input from J.D.; J.D. and E.D. recruited, sampled and collected data about subjects in SCO with input from R.K.P., N.P. and Y.P.P.; D.W.H. recruited, sampled and collected data about subjects in KEN with input from M.M., W.M., N.O., E.A., and Y.P.P; J.D. processed whole blood for microarray experiments; J.D. and W.B. processed microarray experiments supervised by M.W.Mc.; J.D. and J.D.Mc. performed microarray data analysis with input from M.W.Mc.; J.D. and M.W.Mc performed Ingenuity Pathway Analysis; J.D. processed QuantiGene Plex experiments with help from D.W.H. and W.B.; J.D., K.M. and J.D.Mc. performed QuantiGene Plex data analysis; K.M. and J.D.Mc. performed class-prediction analysis with input from J.D.; J.D. wrote the manuscript; All authors discussed the results and approved the manuscript.

**Acknowledgements**

The authors would like to thank the volunteers for their participation and cooperation; T. Chatterji, M. Anderson and other students for their help with data collection in SCO; J. McCulloch for his assistance in setting up the initial RNA extraction methods. A. Keenan for her help coordinating the shipment of equipment and consumables to Kenya. M.M. was

supported by national scholarship program Kristjan Jaak, which is funded and managed by Archimedes Foundation in collaboration with the Ministry of Education and Research. Both K.M. and M.M. were supported by European Social Fund's Doctoral Studies and Internationalisation Programme DoRa, which is carried out by Foundation Archimedes. The research was funded by grant from the World Anti-doping Agency (08C19YP).

**Disclosures:**

The authors declare no conflicts of interests.

**Data Supplements:**

The following supplementary data are available with the online version of this paper.

- **Data Set S1**: Transcripts lists related to the Venn diagrams in Figure 1C (top left panel).
- **Data Set S2**: Transcripts lists related to the Venn diagrams in Figure 1C (bottom left panel).
- **Data Set S3**: Transcripts lists related to the Venn diagrams in Figure 1C (top right panel).
- **Data Set S4**: Transcripts lists related to the Venn diagrams in Figure 1C (bottom right panel).

**REFERENCES:**

1. **Ashenden M, Gough CE, Garnham A, Gore CJ, and Sharpe K**. Current markers of the Athlete Blood Passport do not flag microdose EPO doping. *Eur J Appl Physiol* 111: 2307-2314, 2011.

2. **Bailly-Chouriberry L, Noguier F, Manchon L, Piquemal D, Garcia P, Popot MA, and Bonnaire Y**. Blood cells RNA biomarkers as a first long-term detection strategy for EPO abuse in horseracing. *Drug Test Anal* 2: 339-345, 2010.

3. **Benjamini Y, and Hochberg Y**. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57: 289-300, 1995.

4. **Breitling R, Armengaud P, Amtmann A, and Herzyk P**. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* 573: 83-92, 2004.

5. **Canales RD, Luo Y, Willey JC, Austermiller B, Barbacioru CC, Boysen C, Hunkapiller K, Jensen RV, Knight CR, Lee KY, Ma Y, Maqsodi B, Papallo A, Peters EH, Poulter K, Ruppel PL, Samaha RR, Shi L, Yang W, Zhang L, and Goodsaid FM**. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* 24: 1115-1122, 2006.

6. **Debey S, Schoenbeck U, Hellmich M, Gathof BS, Pillai R, Zander T, and Schultze JL**. Comparison of different isolation techniques prior gene expression profiling of blood derived cells: impact on physiological responses, on overall expression and the role of different cell types. *Pharmacogenomics J* 4: 193-207, 2004.

7. **Ding K, Shameer K, Jouni H, Masys DR, Jarvik GP, Kho AN, Ritchie MD, McCarty CA, Chute CG, Manolio TA, and Kullo IJ**. Genetic Loci implicated in erythroid differentiation and cell cycle regulation are associated with red blood cell traits. *Mayo Clin Proc* 87: 461-474, 2012.

8. **Duale N, Lipkin WI, Briese T, Aarem J, Ronningen KS, Aas KK, Magnus P, Harbak K, Susser E, and Brunborg G**. Long-term storage of blood RNA collected in RNA stabilizing Tempus tubes in a large biobank - evaluation of RNA quality and stability. *BMC research notes* 7: 633, 2014.

9. **Dudoit S, Fridlyand J, and Speed TP**. Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Statist Assoc* 97: 77-87, 2002.

10. **Durussel J, Daskalaki E, Anderson M, Chatterji T, Wondimu DH, Padmanabhan N, Patel RK, McClure JD, and Pitsiladis YP**. Haemoglobin mass and running time trial performance after recombinant human erythropoietin administration in trained men. *PLoS One* 8: e56151, 2013.

11. **Fan H, and Hegde PS**. The transcriptome in blood: challenges and solutions for robust expression profiling. *Curr Mol Med* 5: 3-10, 2005.

12. **Flagella M, Bui S, Zheng Z, Nguyen CT, Zhang A, Pastor L, Ma Y, Yang W, Crawford KL, McMaster GK, Witney F, and Luo Y**. A multiplex branched DNA assay for parallel quantitative gene expression profiling. *Anal Biochem* 352: 50-60, 2006.

13. **Fujita T, Suzuki K, Tada T, Yoshihara Y, Hamaoka R, Uchida K, Matuo Y, Sasaki T, Hanafusa T, and Taniguchi N**. Human erythrocyte bisphosphoglycerate mutase: inactivation by glycation in vivo and in vitro. *J Biochem* 124: 1237-1244, 1998.

14. **Geers C, and Gros G**. Carbon dioxide transport and carbonic anhydrase in blood and muscle. *Physiol Rev* 80: 681-715, 2000.

15. **Heinicke LA, Nabet B, Shen S, Jiang P, van Zalen S, Cieply B, Russell JE, Xing Y, and Carstens RP**. The RNA binding protein RBM38 (RNPC1) regulates splicing during late erythroid differentiation. *PLoS One* 8: e78031, 2013.

16. **Howman D**. Scientific research using elite athletes: WADA point of view. *J Appl Physiol* 114: 1365, 2013.

17. **Huang CH, Chen Y, and Blumenfeld OO**. A novel St(a) glycophorin produced via gene conversion of pseudoexon III from glycophorin E to glycophorin A gene. *Hum Mutat* 15: 533-540, 2000.

18. **Jelkmann W**. Regulation of erythropoietin production. *J Physiol* 589: 1251-1258, 2011.

19. **Jie KE, van der Putten K, Wesseling S, Joles JA, Bergevoet MW, Pepers-de Kort F, Doevendans PA, Yasui Y, Liu Q, Verhaar MC, Gaillard CA, and Braam B**. Short-term erythropoietin treatment does not substantially modulate monocyte transcriptomes of patients with combined heart and renal failure. *PLoS One* 7: e41339, 2012.

20. **Johnson DW, Pollock CA, and Macdougall IC**. Erythropoiesis-stimulating agent hyporesponsiveness. *Nephrology (Carlton)* 12: 321-330, 2007.

21. **Matheson LA, Duong TT, Rosenberg AM, and Yeung RSM**. Assessment of sample collection and storage methods for multicenter immunologic research in children. *J Immunol Methods* 339: 82-89, 2008.

22. **Monte AA, Vasiliou V, and Heard KJ**. Omics screening for pharmaceutical efficacy and safety in clinical practice. *J Pharmacogenomics Pharmacoproteomics* S5: 2012.

23. **Morkeberg J**. Blood manipulation: current challenges from an anti-doping perspective. *Hematology Am Soc Hematol Educ Program* 2013: 627-631, 2013.

24. **Perrotta S, Borriello A, Scaloni A, De Franceschi L, Brunati AM, Turrini F, Nigro V, del Giudice EM, Nobili B, Conte ML, Rossi F, Iolascon A, Donella-Deana A, Zappia V, Poggi V, Anong W, Low P, Mohandas N, and Della Ragione F**. The N-terminal 11 amino acids of human erythrocyte band 3 are critical for aldolase binding and protein phosphorylation: implications for band 3 function. *Blood* 106: 4359-4366, 2005.

25. **Pitsiladis YP, Durussel J, and Rabin O**. An integrative 'omics' solution to the detection of recombinant human erythropoietin and blood doping. *Br J Sports Med* 48: 856-861, 2014.

26. **Prabakaran S, Wengenroth M, Lockstone HE, Lilley K, Leweke FM, and Bahn S**. 2-D DIGE analysis of liver and red blood cells provides further evidence for oxidative stress in schizophrenia. *J Proteome Res* 6: 141-149, 2007.

27. **Prezeau N, Silvy M, Gabert J, and Picard C**. Assessment of a new RNA stabilizing reagent (Tempus Blood RNA) for minimal residual disease in onco-hematology using the EAC protocol. *Leukemia Res* 30: 569-574, 2006.

28. **Sadlon TJ, Dell'Oso T, Surinya KH, and May BK**. Regulation of erythroid 5-aminolevulinate synthase expression during erythropoiesis. *Int J Biochem Cell Biol* 31: 1153-1167, 1999.

29. **Satchwell TJ, Shoemark DK, Sessions RB, and Toye AM**. Protein 4.2: a complex linker. *Blood Cells Mol Dis* 42: 201-210, 2009.

30. **Scherzer CR, Grass JA, Liao Z, Pepivani I, Zheng B, Eklund AC, Ney PA, Ng J, McGoldrick M, Mollenhauer B, Bresnick EH, and Schlossmacher MG**. GATA transcription factors directly regulate the Parkinson's disease-linked gene alpha-synuclein. *Proc Natl Acad Sci U S A* 105: 10907-10912, 2008.

31. **Shou J, Dotson C, Qian HR, Tao W, Lin C, Lawrence F, N'Cho M, Kulkarni NH, Bull CM, Gelbert LM, and Onyia JE**. Optimized blood cell profiling method for genomic biomarker discovery using high-density microarray. *Biomarkers* 10: 310-320, 2005.

32. **Silva M, Grillot D, Benito A, Richard C, Nunez G, and Fernandez-Luna JL**. Erythropoietin can promote erythroid progenitor survival by repressing apoptosis through Bcl-XL and Bcl-2. *Blood* 88: 1576-1582, 1996.

33. **Smyth GK**. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3: Article3, 2004.

34. **Sottas PE, Robinson N, and Saugy M**. The athlete's biological passport and indirect markers of blood doping. *Handb Exp Pharmacol* 305-326, 2010.

35. **Sottas PE, and Vernec A**. Current implementation and future of the Athlete Biological Passport. *Bioanalysis* 4: 1645-1652, 2012.

36. **Ursitti JA, and Fowler VM**. Immunolocalization of tropomodulin, tropomyosin and actin in spread human erythrocyte skeletons. *J Cell Sci* 107 ( Pt 6): 1633-1639, 1994.

37. **van der Harst P, Zhang W, Mateo Leach I, Rendon A, Verweij N, Sehmi J, Paul DS, Elling U, Allayee H, Li X, Radhakrishnan A, Tan ST, Voss K, Weichenberger CX,**

**Albers CA, Al-Hussani A, Asselbergs FW, Ciullo M, Danjou F, Dina C, Esko T, Evans DM, Franke L, Gogele M, Hartiala J, Hersch M, Holm H, Hottenga JJ, Kanoni S, Kleber ME, Lagou V, Langenberg C, Lopez LM, Lyytikainen LP, Melander O, Murgia F, Nolte IM, O'Reilly PF, Padmanabhan S, Parsa A, Pirastu N, Porcu E, Portas L, Prokopenko I, Ried JS, Shin SY, Tang CS, Teumer A, Traglia M, Ulivi S, Westra HJ, Yang J, Zhao JH, Anni F, Abdellaoui A, Attwood A, Balkau B, Bandinelli S, Bastardot F, Benyamin B, Boehm BO, Cookson WO, Das D, de Bakker PI, de Boer RA, de Geus EJ, de Moor MH, Dimitriou M, Domingues FS, Doring A, Engstrom G, Eyjolfsson GI, Ferrucci L, Fischer K, Galanello R, Garner SF, Genser B, Gibson QD, Girotto G, Gudbjartsson DF, Harris SE, Hartikainen AL, Hastie CE, Hedblad B, Illig T, Jolley J, Kahonen M, Kema IP, Kemp JP, Liang L, Lloyd-Jones H, Loos RJ, Meacham S, Medland SE, Meisinger C, Memari Y, Mihailov E, Miller K, Moffatt MF, Nauck M, Novatchkova M, Nutile T, Olafsson I, Onundarson PT, Parracciani D, Penninx BW, Perseu L, Piga A, Pistis G, Pouta A, Puc U, Raitakari O, Ring SM, Robino A, Ruggiero D, Ruokonen A, Saint-Pierre A, Sala C, Salumets A, Sambrook J, Schepers H, Schmidt CO, Sillje HH, Sladek R, Smit JH, Starr JM, Stephens J, Sulem P, Tanaka T, Thorsteinsdottir U, Tragante V, van Gilst WH, van Pelt LJ, van Veldhuisen DJ, Volker U, Whitfield JB, Willemsen G, Winkelmann BR, Wirnsberger G, Algra A, Cucca F, d'Adamo AP, Danesh J, Deary IJ, Dominiczak AF, Elliott P, Fortina P, Froguel P, Gasparini P, Greinacher A, Hazen SL, Jarvelin MR, Khaw KT, Lehtimaki T, Maerz W, Martin NG, Metspalu A, Mitchell BD, Montgomery GW, Moore C, Navis G, Pirastu M, Pramstaller PP, Ramirez-Solis R, Schadt E, Scott J, Shuldiner AR, Smith GD, Smith JG, Snieder H, Sorice R, Spector TD, Stefansson K, Stumvoll M, Tang WH, Toniolo D, Tonjes A, Visscher PM, Vollenweider P, Wareham NJ, Wolffenbuttel BH, Boomsma DI, Beckmann JS, Dedoussis GV, Deloukas P, Ferreira MA, Sanna S, Uda M, Hicks AA,**

**Penninger JM, Gieger C, Kooner JS, Ouwehand WH, Soranzo N, and Chambers JC**. Seventy-five genetic loci influencing the human red blood cell. *Nature* 492: 369-375, 2012.

38. **Varlet-Marie E, Audran M, Ashenden M, Sicart MT, and Piquemal D**. Modification of gene expression: help to detect doping with erythropoiesis-stimulating agents. *Am J Hematol* 84: 755-759, 2009.

39. **Varlet-Marie E, Audran M, Lejeune M, Bonafoux B, Sicart MT, Marti J, Piquemal D, and Commes T**. Analysis of human reticulocyte genes reveals altered erythropoiesis: potential use to detect recombinant human erythropoietin doping. *Haematologica* 89: 991-997, 2004.

40. **von Lohneysen K, Scott TM, Soldau K, Xu X, and Friedman JS**. Assessment of the red cell proteome of young patients with unexplained hemolytic anemia by two-dimensional differential in-gel electrophoresis (DIGE). *PLoS One* 7: e34237, 2012.

41. **Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, and Brown PO**. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci U S A* 100: 1896-1901, 2003.

42. **World Anti-Doping Agency**. Athlete Biological Passport: Operating Guidelines & Compilation of Required Elements https://wada-main-prod.s3.amazonaws.com/resources/files/wada_abp_operating_guidelines_2014_v5.0_ en.pdf. [October 16, 2014].

43. **Zhang X, Zhang W, Ma SF, Miasniakova G, Sergueeva A, Ammosova T, Xu M, Nekhai S, Nourai M, Wade MS, Prchal JT, Garcia JG, Machado RF, and Gordeuk VR**. Iron deficiency modifies gene expression variation induced by augmented hypoxia sensing. *Blood Cells Mol Dis* 52: 35-45, 2014.

44. **Zimmerman PA, Ferreira MU, Howes RE, and Mercereau-Puijalon O**. Red blood cell polymorphism and susceptibility to Plasmodium vivax. *Adv Parasitol* 81: 27-76, 2013.

**Figure Captions:**

**Figure 1. Identification of a distinct whole-blood transcriptional signature of rHuEPO.** (A) Experimental design and time points selected for transcriptomic analysis. (B) Time points, comparisons and colour codes. Labels for comparisons: A=EPO3 (2 days into EPO dosing) vs. Base1 (first baseline measurement), B=EPO4 (2 weeks into EPO) vs. Base1, C=EPO5 (4 weeks into EPO) vs. Base1, D=Post6 (1 week post EPO) vs. Base1, E=Post7 (2 weeks post EPO) vs. Base1, F=Post8 (4 weeks post EPO) vs. Base1. (C) Venn diagrams of transcripts up- (↑) and down-regulated (↓) during (top panels) and after (bottom panels) rHuEPO administration compared to baseline in SCO (n = 18; left panels) and KEN (n = 18; right panels). See supplementary files Data Sets S1-S4 for the lists of transcripts. (D) Individual changes in *CA1* gene, selected as a typical example. Changes are reported as log ratios compared to the average baseline values.

**Figure 2. Haematological related functions are dominant in the rHuEPO transcriptional signature.** (A) Top three regulated Physiological System Development identified by Ingenuity Pathway Analysis of the transcripts altered by rHuEPO across the time points. Labels for comparisons: A=EPO3 (2 days into EPO dosing) vs. Base1 (first baseline measurement), B=EPO4 (2 weeks into EPO) vs. Base1, C=EPO5 (4 weeks into EPO) vs. Base1, D=Post6 (1 week post EPO) vs. Base1, E=Post7 (2 weeks post EPO) vs. Base1, F=Post8 (4 weeks post EPO) vs. Base1 (Figure 1B). (B) Differentiation of the red blood cells altered by rHuEPO (two time points are presented). (C) Changes in reticulocyte and

Jérôme Durussel 7/1/2016 21:09
**Deleted:**
Jérôme Durussel 7/1/2016 21:09
**Deleted:** see
Jérôme Durussel 7/1/2016 21:09
**Deleted:** for the time points comparisons

haematocrit in SCO (n = 18; circle) and KEN (n = 20; square) across the time points. The median values of each time point are represented by the coloured lines for SCO (right) and KEN (left).

**Figure 3.** Heat map of changes in expression in the fifty genes further validated using QuantiGene. Changes are reported as the median log ratio compared to baseline values (n = 18 and 20 in SCO and KEN, respectively).

**Figure 4.** Relationship between changes in reticulocyte and genes that remained significantly differentially expressed after controlling for the effect of reticulocyte.

**Table 1. List of genes identified by the microarray experiment**

**1A List of the 32 genes (Figure 1C)**

| Gene symbol | NCBI | Illumina ID | Gene name |
|---|---|---|---|
| ALAS2 | NM_000032.1 | ILMN_1708323 | aminolevulinate, delta-, synthase 2 |
| BCL2L1 | NM_138578.1 | ILMN_1654118 | BCL2-like 1 |
| BPGM | NM_001724.3 | ILMN_2352921 | 2,3-bisphosphoglycerate mutase |
| C1ORF128 | NM_020362.2 | ILMN_1784207 | chromosome 1 open reading frame 128 |
| CA1 | NM_001738.1 | ILMN_1652431 | carbonic anhydrase I |
| E2F2 | NM_004091.2 | ILMN_1777233 | E2F transcription factor 2 |
| EPB42 | NM_000119.1 | ILMN_1814397 | erythrocyte membrane protein band 4.2 |
| FAM46C | NM_017709.2 | ILMN_1713266 | family with sequence similarity 46, member C |
| FECH | NM_000140.2 | ILMN_1774091 | ferrochelatase |
| GMPR | NM_006877.2 | ILMN_1729487 | guanosine monophosphate reductase |
| GYPB | NM_002100.3 | ILMN_1683093 | glycophorin B |
| GYPE | NM_002102.2 | ILMN_1695187 | glycophorin E |
| HBD | NM_000519.3 | ILMN_1815527 | hemoglobin, delta |
| IFI27 | NM_005532.3 | ILMN_2058782 | interferon, alpha-inducible protein 27 |
| KRT1 | NM_006121.2 | ILMN_1735712 | keratin 1 |
| LOC100131164 | XM_001721919.1 | ILMN_3285762 | PREDICTED: similar to anion exchanger |
| LOC441455 | XR_041340.1 | ILMN_3278170 | PREDICTED: misc_RNA |
| MARCH8 | NM_145021.4 | ILMN_2336335 | membrane-associated ring finger (C3HC4) 8 |
| OSBP2 | NM_030758.3 | ILMN_1781966 | oxysterol binding protein 2 |
| RAP1GAP | NM_002885.1 | ILMN_1776519 | RAP1 GTPase activating protein |
| RBM38 | NM_183425.1 | ILMN_2404049 | RNA binding motif protein 38 (transcript variant 2) |
| RBM38 | NM_017495.4 | ILMN_1704079 | RNA binding motif protein 38 (transcript variant 1) |
| SELENBP1 | NM_003944.2 | ILMN_1680652 | selenium binding protein 1 |
| SIAH2 | NM_005067.5 | ILMN_1801313 | seven in absentia homolog 2 |
| SLC4A1 | NM_000342.1 | ILMN_1772809 | solute carrier family 4, anion exchanger, member 1 |
| SLC6A10P | NM_198857.1 | ILMN_1704446 | solute carrier family 6, member 10 |
| SNCA | NM_007308.1 | ILMN_1701933 | synuclein, alpha (transcript variant NACP112) |
| SNCA | NM_000345.2 | ILMN_1766165 | synuclein, alpha (transcript variant NACP140) |
| TMOD1 | NM_003275.1 | ILMN_1736911 | tropomodulin 1 |
| WDR40A | NM_015397.1 | ILMN_1786328 | WD repeat domain 40A |
| XK | NM_021083.2 | ILMN_1759117 | X-linked Kx blood group |
| YOD1 | NM_018566.3 | ILMN_1678919 | YOD1 OTU deubiquinating enzyme 1 homolog |

**1B List of the 5 genes (Figure 1C)**

| Gene symbol | NCBI | Illumina ID | Gene name |
|---|---|---|---|
| ALAS2 | NM_001037968.1 | ILMN_2367126 | aminolevulinate, delta-, synthase 2 |
| LOC389599 | XM_372002.3 | ILMN_1714765 | PREDICTED: similar to amyotrophic lateral sclerosis 2 |
| SELENBP1 | NM_003944.2 | ILMN_1680652 | selenium binding protein 1 |
| SNCA | NM_007308.1 | ILMN_1701933 | synuclein, alpha |
| TRIM58 | NM_015431.2 | ILMN_1705458 | tripartite motif-containing 58 |

Jérôme Durussel 7/1/2016 21:05
Deleted: 2A

List of the 32 and 5 genes differentially and commonly expressed in both groups during and post rHuEPO administration identified by the microarray experiment (Figure 1C).

**Table 2. Subset of identified genes further analysed using QuantiGene**

| GENE.QG | NCBI Genbank: | Pattern | Corr QG vs. Illumina | Limma Sig | | | | | | Corr QG vs. Ret | Limma Ret Sig | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | A | B | C | D | E | F | | A | B | C | D | E | F |
| *EPB42* | NM_000119 | UpDown | $0.98_{(0.98 \text{ to } 0.98)}$ | $A^{-21}$ | $B^{-64}$ | $C^{-44}$ | $D^{-10}$ | $E^{-42}$ | $F^{-14}$ | $0.91_{(0.89 \text{ to } 0.93)}$ | $A^{-3}$ | $B^{-3}$ | $C^{-2}$ | $D^{-3}$ | | |
| *GMPR* | NM_006877 | UpDown | $0.98_{(0.97 \text{ to } 0.98)}$ | $A^{-15}$ | $B^{-49}$ | $C^{-33}$ | $D^{-11}$ | $E^{-53}$ | $F^{-17}$ | $0.91_{(0.88 \text{ to } 0.93)}$ | | | | $D^{-3}$ | $E^{-3}$ | |
| *SELENBP1* | NM_003944 | UpDown | $0.98_{(0.98 \text{ to } 0.99)}$ | $A^{-21}$ | $B^{-64}$ | $C^{-43}$ | $D^{-12}$ | $E^{-50}$ | $F^{-16}$ | $0.91_{(0.89 \text{ to } 0.93)}$ | $A^{-3}$ | $B^{-2}$ | | $D^{-4}$ | $E^{-2}$ | |
| *CA1* | NM_001738 | UpDown | $0.96_{(0.95 \text{ to } 0.97)}$ | $A^{-14}$ | $B^{-68}$ | $C^{-53}$ | | $E^{-46}$ | $F^{-22}$ | $0.91_{(0.88 \text{ to } 0.92)}$ | | $B^{-3}$ | $C^{-4}$ | | $E^{-2}$ | |
| *SLC4A1* | NM_000342 | UpDown | $0.97_{(0.97 \text{ to } 0.98)}$ | $A^{-19}$ | $B^{-60}$ | $C^{-41}$ | $D^{-8}$ | $E^{-59}$ | $F^{-19}$ | $0.92_{(0.90 \text{ to } 0.94)}$ | $A^{-2}$ | | | $D^{-2}$ | $E^{-3}$ | |
| *TRIM58* | NM_015431 | UpDown | $0.97_{(0.97 \text{ to } 0.98)}$ | $A^{-16}$ | $B^{-55}$ | $C^{-41}$ | $D^{-3}$ | $E^{-51}$ | $F^{-13}$ | $0.91_{(0.88 \text{ to } 0.93)}$ | | | | | $E^{-2}$ | |
| *BCL2L1* | NM_001191 | UpDown | $0.96_{(0.95 \text{ to } 0.97)}$ | $A^{-14}$ | $B^{-46}$ | $C^{-36}$ | $D^{-5}$ | $E^{-48}$ | $F^{-16}$ | $0.91_{(0.88 \text{ to } 0.93)}$ | | | | | $E^{-2}$ | |
| *CSDA* | NM_003651 | UpDown | $0.95_{(0.93 \text{ to } 0.96)}$ | $A^{-7}$ | $B^{-53}$ | $C^{-40}$ | | $E^{-55}$ | $F^{-22}$ | $0.91_{(0.89 \text{ to } 0.93)}$ | | | | | $E^{-3}$ | |
| *TMOD1* | NM_003275 | UpDown | $0.94_{(0.93 \text{ to } 0.95)}$ | $A^{-7}$ | $B^{-39}$ | $C^{-28}$ | $D^{-3}$ | $E^{-51}$ | $F^{-18}$ | $0.91_{(0.89 \text{ to } 0.93)}$ | | | | | $E^{-3}$ | |
| *BPGM* | NM_001724 | UpDown | $0.87_{(0.84 \text{ to } 0.89)}$ | $A^{-14}$ | $B^{-48}$ | $C^{-36}$ | | $E^{-48}$ | $F^{-20}$ | $0.90_{(0.87 \text{ to } 0.92)}$ | | | | | $E^{-3}$ | |
| *DCAF12* | NM_015397 | UpDown | $0.95_{(0.94 \text{ to } 0.96)}$ | $A^{-8}$ | $B^{-40}$ | $C^{-30}$ | | $E^{-55}$ | $F^{-16}$ | $0.91_{(0.89 \text{ to } 0.93)}$ | | | | | $E^{-3}$ | |
| *FAM46C* | NM_017709 | UpDown | $0.84_{(0.80 \text{ to } 0.87)}$ | $A^{-7}$ | $B^{-32}$ | $C^{-25}$ | | $E^{-51}$ | $F^{-18}$ | $0.89_{(0.86 \text{ to } 0.91)}$ | | | | | $E^{-3}$ | |
| *STRADB* | NM_018571 | UpDown | $0.96_{(0.96 \text{ to } 0.97)}$ | $A^{-5}$ | $B^{-34}$ | $C^{-26}$ | | $E^{-52}$ | $F^{-19}$ | $0.90_{(0.87 \text{ to } 0.92)}$ | | | | | $E^{-2}$ | |
| *ADIPOR1* | NM_015999 | UpDown | $0.97_{(0.96 \text{ to } 0.97)}$ | $A^{-6}$ | $B^{-39}$ | $C^{-29}$ | | $E^{-53}$ | $F^{-20}$ | $0.90_{(0.88 \text{ to } 0.92)}$ | | | | | $E^{-3}$ | |
| *HBD* | NM_000519 | UpDown | $0.93_{(0.91 \text{ to } 0.94)}$ | $A^{-5}$ | $B^{-23}$ | $C^{-18}$ | $D^{-2}$ | $E^{-54}$ | $F^{-16}$ | $0.88_{(0.85 \text{ to } 0.91)}$ | | | | | $E^{-3}$ | |
| *ALAS2* | NM_000032 | UpDown | $0.93_{(0.92 \text{ to } 0.95)}$ | $A^{-11}$ | $B^{-41}$ | $C^{-31}$ | $D^{-3}$ | $E^{-69}$ | $F^{-23}$ | $0.90_{(0.88 \text{ to } 0.92)}$ | | | | | $E^{-4}$ | |
| *PITHD1* | NM_020362 | UpDown | $0.93_{(0.92 \text{ to } 0.95)}$ | $A^{-19}$ | $B^{-63}$ | $C^{-48}$ | | $E^{-30}$ | $F^{-8}$ | $0.88_{(0.86 \text{ to } 0.91)}$ | $A^{-3}$ | $B^{-6}$ | $C^{-6}$ | | | |
| *RBM38* | NM_017495 | UpDown | $0.96_{(0.95 \text{ to } 0.97)}$ | $A^{-15}$ | $B^{-56}$ | $C^{-40}$ | | $E^{-38}$ | $F^{-10}$ | $0.92_{(0.90 \text{ to } 0.93)}$ | | $B^{-2}$ | | | | |
| *GYPE* | NM_002102 | UpDown | $0.89_{(0.87 \text{ to } 0.91)}$ | $A^{-8}$ | $B^{-39}$ | $C^{-36}$ | $D^{-2}$ | $E^{-11}$ | $F^{-10}$ | $0.85_{(0.82 \text{ to } 0.88)}$ | | $B^{-4}$ | $C^{-6}$ | $D^{-2}$ | | |
| *SNCA* | NM_000345 | UpDown | $0.92_{(0.90 \text{ to } 0.94)}$ | $A^{-7}$ | $B^{-48}$ | $C^{-40}$ | $D^{-2}$ | $E^{-49}$ | $F^{-18}$ | $0.89_{(0.86 \text{ to } 0.91)}$ | | | | $D^{-3}$ | $E^{-2}$ | |
| *YOD1* | NM_018566 | UpDown | $0.77_{(0.72 \text{ to } 0.81)}$ | $A^{-5}$ | $B^{-27}$ | $C^{-25}$ | $D^{-3}$ | $E^{-40}$ | $F^{-16}$ | $0.85_{(0.82 \text{ to } 0.88)}$ | | | | $D^{-3}$ | $E^{-2}$ | |
| *FECH* | NM_000140 | UpDown | $0.88_{(0.85 \text{ to } 0.90)}$ | $A^{-8}$ | $B^{-41}$ | $C^{-35}$ | | $E^{-37}$ | $F^{-16}$ | $0.89_{(0.87 \text{ to } 0.92)}$ | | | | | | |
| *FBXO7* | NM_001033024 | UpDown | $0.93_{(0.91 \text{ to } 0.94)}$ | $A^{-4}$ | $B^{-18}$ | $C^{-15}$ | | $E^{-38}$ | $F^{-13}$ | $0.84_{(0.80 \text{ to } 0.87)}$ | | | | | $E^{-2}$ | |
| *UBXN6* | NM_025241 | UpDown | $0.97_{(0.96 \text{ to } 0.97)}$ | $A^{-5}$ | $B^{-35}$ | $C^{-25}$ | | $E^{-53}$ | $F^{-20}$ | $0.91_{(0.88 \text{ to } 0.93)}$ | | | | | $E^{-3}$ | |
| *GUK1* | NM_000858 | UpDown | $0.97_{(0.96 \text{ to } 0.98)}$ | $A^{-4}$ | $B^{-22}$ | $C^{-15}$ | $D^{-3}$ | $E^{-55}$ | $F^{-23}$ | $0.91_{(0.89 \text{ to } 0.93)}$ | | | | | $E^{-3}$ | |
| *KRT1* | NM_006121 | UpDown | $0.77_{(0.71 \text{ to } 0.81)}$ | $A^{-3}$ | $B^{-14}$ | $C^{-10}$ | | | | $0.74_{(0.69 \text{ to } 0.79)}$ | | $B^{-4}$ | $C^{-4}$ | | | |
| *ROPN1B* | NM_001012337 | UpDown | $0.81_{(0.76 \text{ to } 0.84)}$ | $A^{-3}$ | $B^{-13}$ | $C^{-10}$ | | $E^{-6}$ | $F^{-2}$ | $0.77_{(0.72 \text{ to } 0.82)}$ | | $B^{-3}$ | $C^{-2}$ | | | |
| *SERPINA13* | NM_207378 | UpDown | $0.76_{(0.70 \text{ to } 0.80)}$ | $A^{-2}$ | $B^{-9}$ | $C^{-13}$ | | $E^{-5}$ | | $0.74_{(0.68 \text{ to } 0.79)}$ | | $B^{-2}$ | $C^{-4}$ | | | |
| *OSBP2* | NM_030758 | UpDown | $0.51_{(0.42 \text{ to } 0.59)}$ | | $B^{-9}$ | $C^{-10}$ | $D^{-3}$ | $E^{-4}$ | $F^{-11}$ | $0.52_{(0.43 \text{ to } 0.60)}$ | | $B^{-4}$ | $C^{-5}$ | | $E^{-2}$ | $F^{-4}$ |
| *HBE1* | NM_005330 | UpDown | $0.56_{(0.48 \text{ to } 0.64)}$ | | $B^{-4}$ | $C^{-4}$ | | | $F^{-4}$ | $0.54_{(0.45 \text{ to } 0.61)}$ | | $B^{-2}$ | $C^{-3}$ | | | |
| *SLC6A10P* | NM_198857 | UpDown | $0.35_{(0.24 \text{ to } 0.44)}$ | | | $C^{-2}$ | $D^{-3}$ | $E^{-7}$ | $F^{-13}$ | $0.31_{(0.20 \text{ to } 0.41)}$ | | | $C^{-3}$ | | $E^{-2}$ | $F^{-3}$ |
| *VEGFB* | NM_003377 | DownUp | $-0.26_{(-0.36 \text{ to } -0.15)}$ | | | $C^{-2}$ | $D^{-3}$ | $E^{-8}$ | $F^{-11}$ | $0.40_{(0.30 \text{ to } 0.49)}$ | | | $C^{-2}$ | | $E^{-3}$ | $F^{-3}$ |
| *TPRA1* | NM_016372 | UpDown | $0.47_{(0.37 \text{ to } 0.55)}$ | | | | | | $F^{-2}$ | $0.44_{(0.34 \text{ to } 0.53)}$ | | $B^{-2}$ | $C^{-3}$ | | | |
| *TNS1* | NM_022648 | UpDown | $0.86_{(0.83 \text{ to } 0.89)}$ | $A^{-6}$ | $B^{-7}$ | $C^{-4}$ | $D^{-2}$ | $E^{-12}$ | $F^{-3}$ | $0.78_{(0.73 \text{ to } 0.82)}$ | | | | | | |
| *SGK223* | NM_001080826 | DownUp | $0.21_{(0.09 \text{ to } 0.31)}$ | | $B^{-13}$ | $C^{-4}$ | | $E^{-5}$ | $F^{-11}$ | $0.04_{(-0.07 \text{ to } 0.16)}$ | | | | | | |
| *ACTR10* | NM_018477 | HK | $0.06_{(-0.06 \text{ to } 0.17)}$ | | $B^{-8}$ | $C^{-4}$ | | $E^{-4}$ | $F^{-2}$ | $0.29_{(0.18 \text{ to } 0.39)}$ | | | | | | |
| *MRFAP1* | NM_033296 | HK | $0.15_{(0.04 \text{ to } 0.26)}$ | | $B^{-20}$ | $C^{-16}$ | | $E^{-2}$ | $F^{-2}$ | $0.07_{(-0.05 \text{ to } 0.19)}$ | | | | | $E^{-2}$ | |
| *RAB11A* | NM_004663 | HK | $0.21_{(0.10 \text{ to } 0.32)}$ | | $B^{-12}$ | $C^{-9}$ | | $E^{-2}$ | | $0.09_{(-0.03 \text{ to } 0.20)}$ | | | | | $E^{-2}$ | |
| *ACTB* | NM_001101 | HK | $0.30_{(0.19 \text{ to } 0.41)}$ | | $B^{-16}$ | $C^{-13}$ | | $E^{-7}$ | $F^{-3}$ | $-0.38_{(-0.48 \text{ to } -0.28)}$ | | | | | | |
| *CD3D* | NM_000732 | DownUp | $0.93_{(0.91 \text{ to } 0.94)}$ | $A^{-3}$ | $B^{-51}$ | $C^{-32}$ | | $E^{-6}$ | $F^{-3}$ | $-0.63_{(-0.69 \text{ to } -0.55)}$ | | | | | $E^{-2}$ | |
| *CCR7* | NM_001838 | DownUp | $0.94_{(0.92 \text{ to } 0.95)}$ | | $B^{-48}$ | $C^{-31}$ | | $E^{-9}$ | $F^{-7}$ | $-0.60_{(-0.67 \text{ to } -0.52)}$ | | | | | | |
| *LEF1* | NM_016269 | DownUp | $0.93_{(0.91 \text{ to } 0.94)}$ | $A^{-2}$ | $B^{-46}$ | $C^{-29}$ | | $E^{-8}$ | $F^{-5}$ | $-0.62_{(-0.69 \text{ to } -0.55)}$ | | | | | $E^{-2}$ | |
| *SKAP1* | NM_003726 | DownUp | $0.87_{(0.83 \text{ to } 0.89)}$ | | $B^{-34}$ | $C^{-20}$ | | $E^{-4}$ | $F^{-3}$ | $-0.45_{(-0.54 \text{ to } -0.36)}$ | | | | | | |
| *RNF213* | NM_020914 | UpDown | $-0.47_{(-0.56 \text{ to } -0.38)}$ | $A^{-3}$ | $B^{-29}$ | $C^{-18}$ | | $E^{-3}$ | $F^{-2}$ | $-0.40_{(-0.50 \text{ to } -0.30)}$ | | | | | | |
| *PPIB* | NM_000942 | HK | $0.38_{(0.27 \text{ to } 0.47)}$ | $A^{-2}$ | $B^{-43}$ | $C^{-27}$ | | $E^{-4}$ | $F^{-2}$ | $-0.54_{(-0.62 \text{ to } -0.45)}$ | | | | | $E^{-2}$ | |
| *MIF* | NM_002415 | DownUp | $0.84_{(0.80 \text{ to } 0.87)}$ | $A^{-3}$ | $B^{-47}$ | $C^{-31}$ | | $E^{-5}$ | | $-0.60_{(-0.67 \text{ to } -0.52)}$ | | | | | $E^{-2}$ | |
| *EEF1D* | NM_032378 | DownUp | $0.66_{(0.58 \text{ to } 0.72)}$ | $A^{-2}$ | $B^{-48}$ | $C^{-33}$ | | $E^{-8}$ | $F^{-4}$ | $-0.63_{(-0.69 \text{ to } -0.55)}$ | | | | | $E^{-2}$ | |
| *LOC286444* | XR_038693 | DownUp | $0.63_{(0.55 \text{ to } 0.69)}$ | $A^{-4}$ | $B^{-55}$ | $C^{-32}$ | | $E^{-9}$ | $F^{-5}$ | $-0.65_{(-0.71 \text{ to } -0.58)}$ | | | | | $E^{-2}$ | |
| *CD247* | NM_198053 | DownUp | $0.92_{(0.90 \text{ to } 0.94)}$ | $A^{-3}$ | $B^{-42}$ | $C^{-26}$ | | $E^{-3}$ | | $-0.52_{(-0.60 \text{ to } -0.43)}$ | | | | | $E^{-3}$ | |
| *LOC100130562* | XM_001720379 | DownUp | $0.59_{(0.51 \text{ to } 0.66)}$ | $A^{-2}$ | $B^{-44}$ | $C^{-29}$ | | $E^{-6}$ | $F^{-3}$ | $-0.59_{(-0.66 \text{ to } -0.51)}$ | | | | | $E^{-2}$ | |

The subset of identified genes further validated using QuantiGene (QG) Plex assay is arranged according to the heatmap (Figure 3A). The columns are: (1) Gene name; (2) Accession number; (3) Physiological identified pattern (UpDown: Up-regulated during rHuEPO and down-regulated post administration; DownUp: Down-regulated during rHuEPO and up-regulated post administration; HK: Housekeeping genes); (4 and 6) Correlation between QG and Illumina and QG and reticulocyte results, respectively ($r_{(95\%\ CI)}$); (5 and 7) Significant difference in expression ($^{\text{exponent of the FDR adjusted p-value}}$) at the time points described in Figure 1B using linear models for microarray (Limma) before and after removing the effect of reticulocyte, respectively. Labels for comparisons: A=EPO3 (2 days into EPO dosing) vs. Base1 (first baseline measurement), B=EPO4 (2 weeks into EPO) vs. Base1, C=EPO5 (4 weeks into EPO) vs. Base1, D=Post6 (1 week post EPO) vs. Base1, E=Post7 (2 weeks post EPO) vs. Base1, F=Post8 (4 weeks post EPO) vs. Base1.

**Table 3. Results of the k-nearest neighbour class prediction using the 34 transcripts for the detection of rHuEPO**

| **3A** | Baseline | rHuEPO | Post |
|---|---|---|---|
| Baseline | 23.8% | 5.4% | 10.0% |
| rHuEPO | 0.7% | 31.0% | 1.5% |
| Post | 0.7% | 1.0% | 25.8% |

Jérôme Durussel 7/1/2016 21:05
Deleted: 2

| **3B** | Base1 | Base2 | EPO3 | EPO4 | EPO5 | Post6 | Post7 | Post8 |
|---|---|---|---|---|---|---|---|---|
| Baseline | 11.9% | 11.9% | 4.7% | 0.0% | 0.7% | 4.6% | 1.5% | 3.9% |
| rHuEPO | 0.7% | 0.0% | 7.2% | 12.3% | 11.6% | 1.0% | 0.2% | 0.3% |
| Post | 0.0% | 0.7% | 0.3% | 0.3% | 7.0% | 7.0% | 10.5% | 8.4% |

Jérôme Durussel 7/1/2016 21:05
Deleted: 2

The top row represents the observed phases (3A) or time points (3B) and the first column represents the k-nearest neighbour predictions. The results highlighted with a grey background are correctly classified results in percentage. The majority of the misclassifications affecting the sensitivity of the detection of rHuEPO arose at two days after only rHuEPO injection (EPO3) and one week after ceasing rHuEPO administration (Post6), when the gene expression is crossing baseline levels due to the "rebound" effect. 34 Transcripts: *EPB42, GMPR, SELENBP1, CA1, SLC4A1, TRIM58, BCL2L1, CSDA, TMOD1, BPGM, DCAF12, FAM46C, STRADB, ADIPOR1, HBD, ALAS2, PITHD1, RBM38, GYPE, SNCA, YOD1, FECH, FBXO7, UBXN6, GUK1, CD3D, CCR7, LEF1, SKAP1, MIF, EEF1D, LOC286444, CD247, LOC100130562.*

Jérôme Durussel 7/1/2016 21:05
Deleted: 2

Jérôme Durussel 7/1/2016 21:05
Deleted: 2