Heritability and association with distinct genetic loci of erythropoietin levels in the general population

Erythropoietin (Epo)-regulated red blood cell (RBC) homeostasis is crucial for oxygen delivery in vertebrates, and Epo over- or under-production causes erythrocytosis or anemia, respectively. Plasma Epo levels are usually very low and characterized by circadian fluctuations and strong induction upon acute exposure to low oxygen conditions such as inspiratory hypoxia or blood loss. Considering the high prevalence of various forms of anemia worldwide as well as the fact that the majority of erythrocytosis cases are of unknown origin, surprisingly little is known about the genetic determinants of circulating Epo levels. While a number of genome-wide association studies (GWAS) could link RBC traits such as RBC size, hemoglobin (Hb) content and hematocrit levels to single-nucleotide polymorphisms (SNP), only few studies analyzed circulating Epo levels. Some specific studies focused on aging populations with gradually decreasing renal function,[1] Epo levels after kidney transplantation,[2] or the (lack of) sex-specific differences in chronically anemic patients.[3] However, none of these initial analyses have been performed in a population-based cohort which would have the advantage of a high external validity of the findings. A first GWAS with circulating Epo levels was performed in a large (6,984 participants) non-family-based Dutch cohort.[4] Epo values were obtained from 6,777 participants and 2,691 were used for GWAS which identified a locus comprising the *HBS1L* and *MYB* genes as most likely targets, but no replication cohort was available for validation by an independent study.

We analyzed RBC traits in the Swiss Kidney Project on Genes in Hypertension (SKIPOGH) cohort, including 1,109 adult participants from the general population with well characterized physiological parameters.[5] Participants' characteristics by sex, including RBC traits, are shown in the *Online Supplementary Table S1*. An important aspect of the SKIPOGH cohort is its family-based design (the average family size in the SKIPOGH cohort is four) which allows for the analysis of heritability, i.e., the genetic component of a given phenotypic trait. As shown in Table 1, a high heritability of RBC traits was observed in the SKIPOGH cohort. Adding a sibling component of variance had little impact on the heritability estimates, suggesting the absence of significant dominance variance and shared environmental components across siblings.

These initial results motivated us to analyze Epo blood plasma levels which, according to our knowledge, had not been studied in such a family-based cohort before. Epo was determined in heparin-treated plasma of a total of 1,066 (96%) participants. For 1,020 samples both duplicate measurements (replicate correlation, r=0.998) and phenotypic data were obtained. Following the elimination of 14 extreme outliers, the remaining 1,006 (91%) unadjusted Epo values (Figure 1A) corresponded to the reported reference values of 2.8-17.9 IU/L (based on 2,506 samples).[4] Over a normal Hb range of 135-175 g/L and 120-155 g/L for men and women, respectively, the range of Epo was 4-24 IU/L without any significant difference between sex (*Online Supplementary Table S1*). A quadratic fit with Hb levels was found (Figure 1B), which may be explained by anemia-responsive Epo at low Hb levels and hormone-responsive Hb at high Epo levels. As shown in Table 2, Epo levels were significantly heritable, without significant sibling or marital components of variance, suggesting the absence of significant dominance variance and shared environmental variance.

We next perfomed a GWAS of $2.5x10^6$ genotyped SNP and an additional $4.0x10^6$ of imputed SNP with the mean of duplicate Epo measurements of 872 (79%) individuals, corrected for age, sex, center and familiarity. Figure 1C shows the level of significance of the association between each of the $6.5x10^6$ markers and the normalized Epo levels. No signal reached $P<5x10^{-8}$, the commonly used level of genome-wide significance (*Online Supplementary Table S2*). However, a few SNP fell into the suggestive significance zone ($P<10^{-5}$) and the top hit, lying on chromosome 15, reached $P=1.05x10^{-7}$ at rs413451. As shown in Figure 1D, the SNP identified on chromosome 15 are located within a linkage disequilibrium (LD) block comprising the last exons of mitogen-activated protein kinase kinase 5 (*MAP2K5*), the whole SKI family transcriptional corepressor 1 (*SKOR1*) gene and the 5' upstream and promoter regions of protein inhibitor of activated STAT1 (*PIAS1*). *PIAS1* is a SUMO E3 ligase affecting STAT1 and NFκB pathways. The PIAS1 locus has previously been associated with body mass index (BMI) and related phenotypes (weight, waist circumference, obesity, predicted visceral adipose tissue), as well as smoking-related phenotypes (initiation age, smoking status) and age at menarche.[6-8] According to the MR-Base PHEWAS database in the UK Biobank cohort several RBC-related phenotypes were also associated with the *MAP2K5-SKOR1-PIAS1* locus, further suggesting that it could be directly associated with Epo levels.

Figure 1E shows the significance of the association for the SNP present at the previously associated locus

Table 1. Heritability estimates of red blood cell indices in the SKIPOGH cohort.

	Model 1			Model 2		
	h^2 ± SEM	λ	*P*	h^2 ± SEM	λ	*P*
Hemoglobin	0.40 ± 0.05	0.52	$<1.0x10^{-7}$	0.37 ± 0.06	0.52	$<1.0x10^{-7}$
Hematocrit	0.37 ± 0.06	0.57	0.001	0.33 ± 0.07	0.57	$<1.0x10^{-7}$
RBC count	0.50 ± 0.05	0.67	$<1.0x10^{-6}$	0.47 ± 0.06	0.66	$<1.0x10^{-7}$
MCV	0.68 ± 0.04	0.59	$<1.0x10^{-7}$	0.68 ± 0.05	0.59	$<1.0x10^{-7}$
MCH	0.63 ± 0.05	0.52	$<1.0x10^{-5}$	0.61 ± 0.06	0.51	$<1.0x10^{-5}$
MCHC	0.60 ± 0.06	0.94	$<1.0x10^{-5}$	0.46 ± 0.06	0.93	$<1.0x10^{-7}$
RDW	0.38 ± 0.07	0.20	<0.001	0.36 ± 0.08	0.20	<0.001

Models are adjusted for age, sex and center. Model 1, no sibship component of variance; model 2, including a sibship component of variance (which captures dominance genetic variance and shared environmental components between siblings). λ, power transformation: (λ = 0) and (λ = 1) correspond to log and no transformations, respectively. SKIPOGH: Swiss Kidney Project on Genes in Hypertension; RBC: red blood cell; MCV: mean corpuscular volume; MCH: mean cell hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW: red blood cell distribution width; h2: heritability; SEM: standard error of the mean.

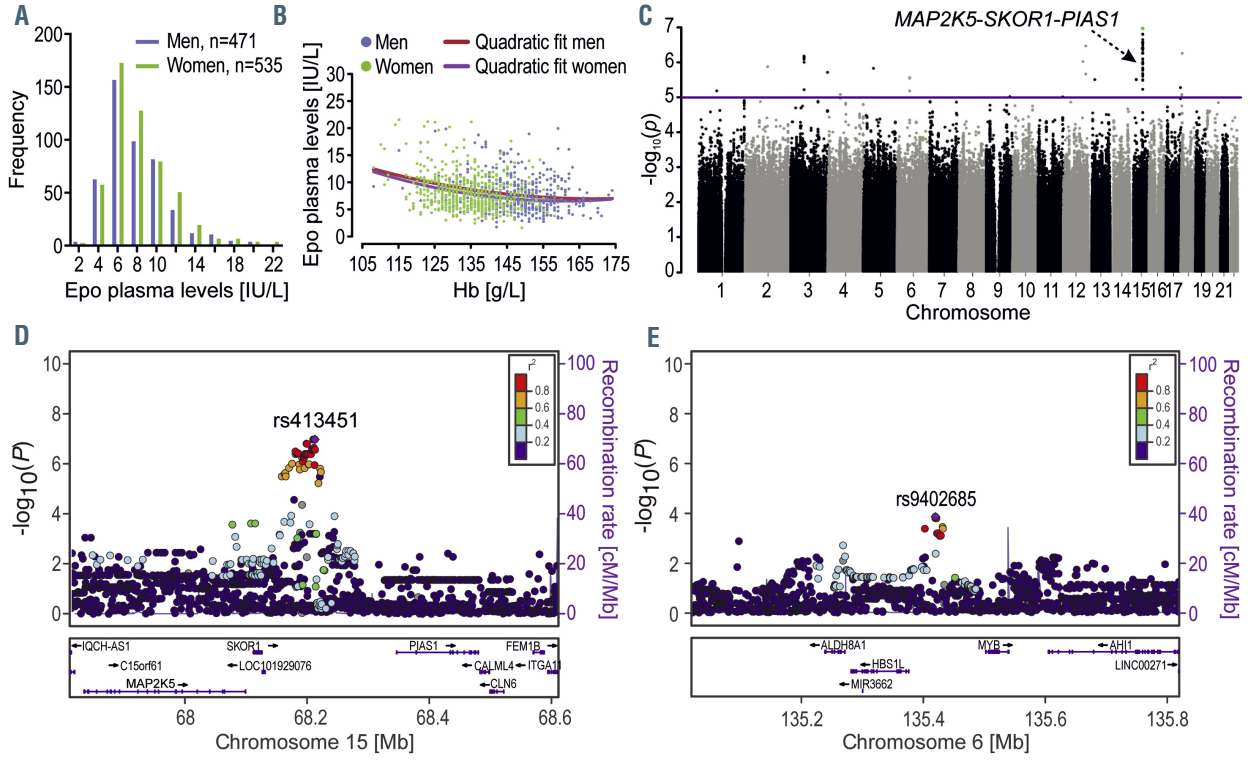

Figure 1. Genome-wide association studies of erythropoietin plasma levels in the SKIPOGH cohort. (A) Histogram distribution of the mean values of duplicate erythropoietin (Epo) measurements in the blood plasma of 1,006 participants. (B) Correlation between Epo and hemoglobin (Hb) levels in the SKIPOGH cohort. (C) Manhattan plot showing the $-\log_{10}$(*P*-value) of the association between Epo plasma levels and 6.4×10^6 single-nucleotide polymorphisms (SNP) in 872 samples following correction for sex, age and center. The blue line shows the indicative suggestive threshold of $P<10^{-5}$. Markers are ranked by chromosome and positions. The green dot on chromosome 6 shows the top SNP at the *HBS1L-MYB* locus, and the green dot on chromosome 15 shows the top SNP at the *MAP2K5-SKOR1-PIAS1* locus. (D) Region around the top SNP rs413451 in the *MAP2K5-SKOR1-PIAS1* locus. (E) Region around the top SNP rs9402685 in the *HBS1L-MYB* intergenic locus which is the reference SNP found in our study. SKIPOGH: Swiss Kidney Project on Genes in Hypertension.

Table 2. Heritability of plasma erythropoietin levels in the SKIPOGH cohort.

Variance components	Adjustment	Heritability ± SD	*P*
P* - M - S	Age, sex, center	0.40 ± 0.07	<0.001
P* - M	Age, sex, center	0.41 ± 0.07	<0.001
P* - S	Age, sex, center	0.39 ± 0.07	<0.001
P*	Age, sex, center	0.40 ± 0.06	<0.001
P* - M - S	Fully adjusted	0.51 ± 0.08	<0.001
P* - M	Fully adjusted	0.52 ± 0.07	<0.001
P* - S	Fully adjusted	0.48 ± 0.07	<0.001
P*	Fully adjusted	0.49 ± 0.06	<0.001

Shown is the heritability of ln(Epo) ± standard deviation. Narrow sense heritability was estimated from family data using the ASSOC program in the Statistical Analysis in Genetic Epidemiology software package (Case Western Reserve University). SKIPOGH: Swiss Kidney Project on Genes in Hypertension; Epo: erythropoietin; SD: standard deviation; P: polygenic; M: marital; S: sibling. Fully adjusted: age, sex, center, current smoker (yes/no), hemoglobin level, eGFR (ckd-epi formula). *Only the polygenic component of the variance was significantly different from 0 in all models.

HBS1L-MYB. Interestingly, the top common SNP of our study (rs9402685) and the associated *HBS1L-MYB* locus showed a robust association with Epo levels ($P=1.46 \times 10^{-4}$), confirming the previously published results.[4] The SNP rs1617640 of the *EPO* locus itself has been reported to be associated with low Epo serum levels in predialysis chronic kidney disease patients,[9] but neither this SNP nor any other SNP of the *EPO* locus (lowest $P=0.17$ at rs7789679) were associated with Epo levels in our study.

While the recently reported SNP rs1130864 of the C-reactive protein (*CRP*) locus did not associate with Epo levels in our study ($P=0.69$), it associated with altered Epo levels in dried neonatal blood spots.[10] In the same study, *EPO* SNP heritability was approximately zero. However, at this age blood volume and hematocrit are different from the adult stage, the liver-to-kidney switch of Epo synthesis is still ongoing, and the method for Epo determination was much less precise, altogether explaining the discrepancy.

The result obtained with our top SNP of the *HBS1L-MYB* locus (rs9402685) was meta-analyzed with results publicly available from Beverborg *et al.*[4] The combined *P*-value reached 1.78×10^{-23}, which was more significant than in any of the two individual studies (1.46×10^{-4} and 1.09×10^{-20}, respectively). The intergenic locus between the *HBS1L* (GTP-binding elongation factor) and *MYB* (myeloblastosis oncogene) genes had previously been reported to be associated with deregulated HbF in a

Chinese β-thalassemia anemia population, a favorable genetic environment for the selection of otherwise erythrocytosis-causing mutations.[11] Several erythropoietic transcription factors have been shown to be prevented from binding to the mutant locus, resulting in lowered *Myb* gene activation and increased HbF synthesis.[12] A link to increased Epo levels, suggesting secondary (Epo-dependent) rather than primary (Epo-independent) erythrocytosis, has not been made in these original reports. GWAS performed in a UK Biobank cohort and a Japanese population showed significant associations between the *HBS1L-MYB* locus and RBC-related phenotypes.[13,14] Together with our direct replication of the association with circulating Epo levels in a Swiss cohort, these results further confirm the implication of this locus in erythropoiesis, maybe both upstream as well as downstream of Epo. However, it is currently not known whether the *HBS1L-MYB* locus also contributes to the heritable genetic determinants triggering Epo levels.

A gene score and a pathway analysis, run with the PASCAL algorithm based on our GWAS results, failed to show any significant pathway after applying multiple testing corrections.

For a candidate-based approach, we selected 33 genes known to influence *EPO* gene expression. The association gene scores from PASCAL could be retrieved from 30 of these 33 genes. Bonferroni correction applied to the number of genes observed led to a significance threshold of 1.67×10^{-3}. The *OS9* gene was significantly associated with a gene score association *P*-value of 1.47×10^{-3} (*Online Supplementary Table S3*). OS-9 is known to interact with both HIF-1α and HIF prolyl-4-hydroxylases, promoting HIF-1α degradation. Interestingly, a *OS9* gene variant has previously been reported to be associated with erythrocytosis in a single patient.[15]

The top SNP of the *MAP2K5-SKOR1-PIAS1* locus (rs413451) was subjected to a phenome-wide association study (PHEWAS) using the MR-Base database of the UK Biobank cohort. SNP rs413451 was most significantly associated with BMI-related phenotypes. Interestingly, Hb concentration ($P=6.35 \times 10^{-6}$), reticulocyte count ($P=7.28 \times 10^{-6}$), hematocrit ($P=1.16 \times 10^{-4}$), reticulocyte fraction of RBC ($P=2.04 \times 10^{-4}$) and RBC count ($P= 2.23 \times 10^{-4}$) were also highly associated with rs413451. Circulatory Epo levels were not available in the UK Biobank. However, the GWAS Atlas database showed a preponderance of BMI-related phenotypes for most significant studies in the database of published GWAS.

In summary, our study revealed the heritability of circulating Epo levels, validated a previously published association with the *HBS1L-MYB* locus, and identified an association with the *MAP2K5-SKOR1-PIAS1* locus. From the list of candidate Epo-regulatory genes, *OS9* showed the highest association with circulating Epo levels. However, the two latter associations require replication, and the functional implication of all three loci in Epo regulation needs to be further investigated. Regarding the idiopathic nature of the majority of erythrocytosis cases, we suggest that especially in patients with high Epo levels, indicative of secondary erythrocytosis, these loci should be considered for further investigation.

Tanguy Corre,[1,2,3] Belen Ponte,[4] Edward Pivin,[1] Menno Pruijm,[5] Daniel Ackermann,[6] Georg Ehret,[7] Katharina Spanaus,[8] Murielle Bochud,[1,2] and Roland H. Wenger[2,9]

[1]Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne; [2]National Center of Competence in Research "Kidney.CH", Zurich; [3]Department of Computational Biology, University of Lausanne, Lausanne; [4]Nephrology Service, Department Medicine, Geneva University Hospital, Geneva; [5]Nephrology Service, University Hospital of Lausanne and University of Lausanne, Lausanne; [6]Department of Nephrology and Hypertension, Inselspital, Bern and University Hospital, University of Bern, Bern; [7]Cardiology, Department of Medicine, Geneva University Hospital, Geneva; [8]Institute of Clinical Chemistry, University Hospital of Zurich, Zurich and [9]Institute of Physiology, University of Zurich, Zurich, Switzerland

Correspondence:

ROLAND H. WENGER - roland.wenger@access.uzh.ch

doi:10.3324/haematol.2021.278389

Received: January 17, 2021.

Accepted: February 26, 2021.

Pre-published: April 8, 2021.

Disclosures: no conflicts of interest to disclose.

Contributions: BP, MP, DA, and GE provided materials; KS measured the samples; TC, EP and MB performed data analysis; TC, MB and RHW wrote the manuscript; MB and RHW supervised the study.

Acknowledgments: we wish to thank P. Spielmann for expert technical help.

Funding: this project was supported by the NCCR "Kidney.CH".

References

1. Ble A, Fink JC, Woodman RC, et al. Renal function, erythropoietin, and anemia of older persons: the inchianti study. Arch Intern Med. 2005;165(19):2222-2227.
2. Goch J, Birgegard G, Wikstrom B, Backman U, Wadstrom J, Danielson BG. Serum erythropoietin and erythropoiesis during six years after kidney transplantation. Nephron. 1996;74(4):687-693.
3. Jelkmann W, Wiedemann G. Lack of sex dependence of the serum level of immunoreactive erythropoietin in chronic anemia. Klin Wochenschr. 1989;67(23):1218.
4. Grote Beverborg N, Verweij N, Klip IT, et al. Erythropoietin in the general population: reference ranges and clinical, biochemical and genetic correlates. PLoS One. 2015;10(4):e0125215.
5. Moulin F, Ponte B, Pruijm M, et al. A population-based approach to assess the heritability and distribution of renal handling of electrolytes. Kidney Int. 2017;92(6):1536-1543.
6. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010;42(11):937-948.
7. Karlsson Linner R, Biroli P, Kong E, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet. 2019;51(2):245-257.
8. Perry JR, Day F, Elks CE, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature. 2014; 514(7520):92-97.
9. Yugavathy N, Huri HZ, Kun LS, et al. Clinical and genetic markers of erythropoietin deficiency anemia in chronic kidney disease (predialysis) patients. Biomark Med. 2020;14(12):1099-1108.
10. Wang Y, Nudel R, Benros ME, et al. Genome-wide association study identifies 16 genomic regions associated with circulating cytokines at birth. PLoS Genet. 2020;16(11):e1009163.
11. Stadhouders R, Aktuna S, Thongjuea S, et al. Hbs1l-myb intergenic variants modulate fetal hemoglobin via long-range myb enhancers. J Clin Invest. 2014;124(4):1699-1710.
12. Suzuki M, Yamazaki H, Mukai HY, et al. Disruption of the hbs1l-myb locus causes hereditary persistence of fetal hemoglobin in a mouse model. Mol Cell Biol. 2013;33(8):1687-1695.
13. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell. 2016;167(5):1415-1429.e1419.
14. Kanai M, Akiyama M, Takahashi A, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat Genet. 2018;50(3):390-400.
15. Camps C, Petousi N, Bento C, et al. Gene panel sequencing improves the diagnostic work-up of patients with idiopathic erythrocytosis and identifies new mutations. Haematologica. 2016; 101(11):1306-1318.