Trans-ethnic meta-analysis of fetal hemoglobin genome-wide association results identifies common variants at the *KLF1* locus

Yann Ilboudo[1]; Carlo Sidore[2]; Francesco Cucca[2]; Guillaume Lettre[1]

Department of Biochemistry and Molecular Medecine, University of Montreal, Montreal Heart Institute, Montreal, Québec (1)

Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy. (2)

Contact: yann.ilboudo@umontreal.ca

Introduction

β-hemoglobinopathies are inherited disorders resulting from a mutant adult β-globin. Sickle cell disease, for example, which stems from a point mutation leading to a Glu→Val substitution at the 6th codon of the protein, and β-thalassemia which ensues from reduced or absent synthesis of β-chains are the most common blood disorders worldwide. Seminal studies reported the benefits of high levels of fetal hemoglobin (HbF) in improving complications and thus the life outcomes of β-hemoglobinopathy patients[1].

Background

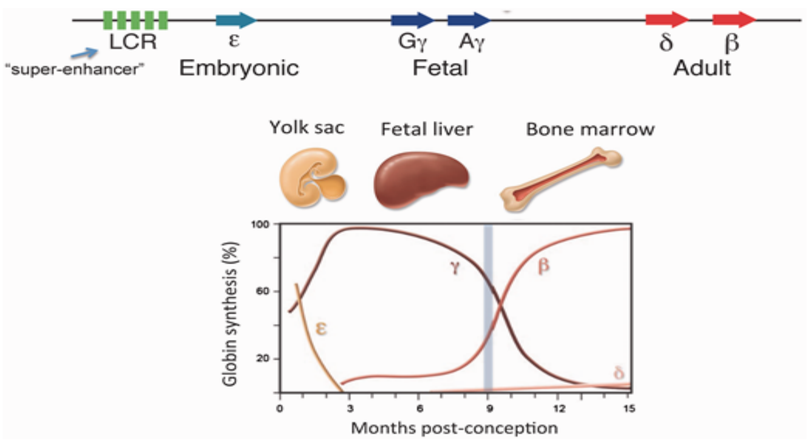

Figure 1. The spatial organization of the β-like globin genes of the human β-globin locus is illustrated above. The LCR, an erythroid super-enhancer, controls expression of the entire locus. The three sites (yolk sac, fetal liver, bone marrow) of red cell production and the relative levels of the β-like globins during pre- and post-natal development are depicted above[1].

In humans, before and after the time of birth into the adulthood three species of hemoglobin are predominant in blood (embryonic, fetal, and adult). However, certain β-globin alleles maintain high levels of fetal hemoglobin in the adult stage, these are known as hereditary persistency of fetal hemoglobin (HPFH).

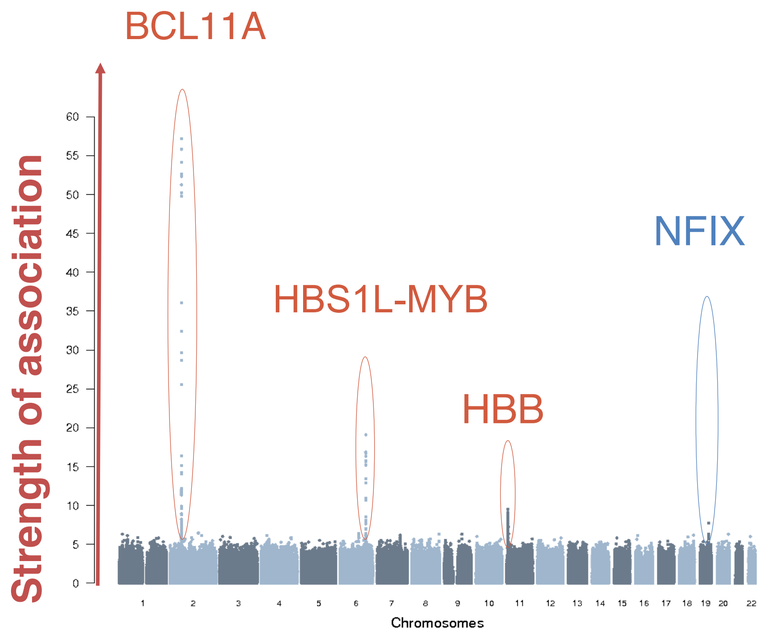

Figure 2. GWAS & Fetal Hemoglobin. Manhattan plot showing the three well known HbF loci (in orange), and the most recently discovered locus (in blue).

Genome-wide association studies yielded the first insights into genic modulation of HbF in humans. In fact, depending on the population ethnicity, we know that taken together *BCL11A, HBSL1-MYB, HBB* account for ~50% of variability of fetal hemoglobin levels. The remaining heritability has been difficult to find, yet, a recent genome-wide scan of 6,602 individuals in the SardiNIA cohort identified *NFIX*-rs183437571 as a new locus[2].

Objective

Knowing that ~50% of heritability comes from *BCL11A, HBS1L-MYB, HBB*, are there other DNA sequences variants responsible for changing HbF levels?

Methods

Cohort Description

Cohort Name	N (by gender F/M)	Age (mean (SD))
SardiNIA	5903 (3391/2512)	43.5 (17.59)
Jamaica	89 (48/41)	NA
CSSCD(1)	353 (188/165)	14.8 (11.7)
MSH	57 (23/34)	28.5 (6.8)
GENMOD	402 (18/184)	31.2 (8.9)
Atlanta	186 (91/95)	31.4 (10.6)
Tanzania	1213 (638/575)	13.3 (7.4)
CSSCD(2)	1139 (545/594)	14.8 (12.1)

Table 1. Sample Description. The population studied here includes ~3,400 SCD patients of African ancestry and ~6000 healthy Sardinians. N: sample size; SD: standard deviation.

Methods (continued)

Statistical Analysis & Annotation

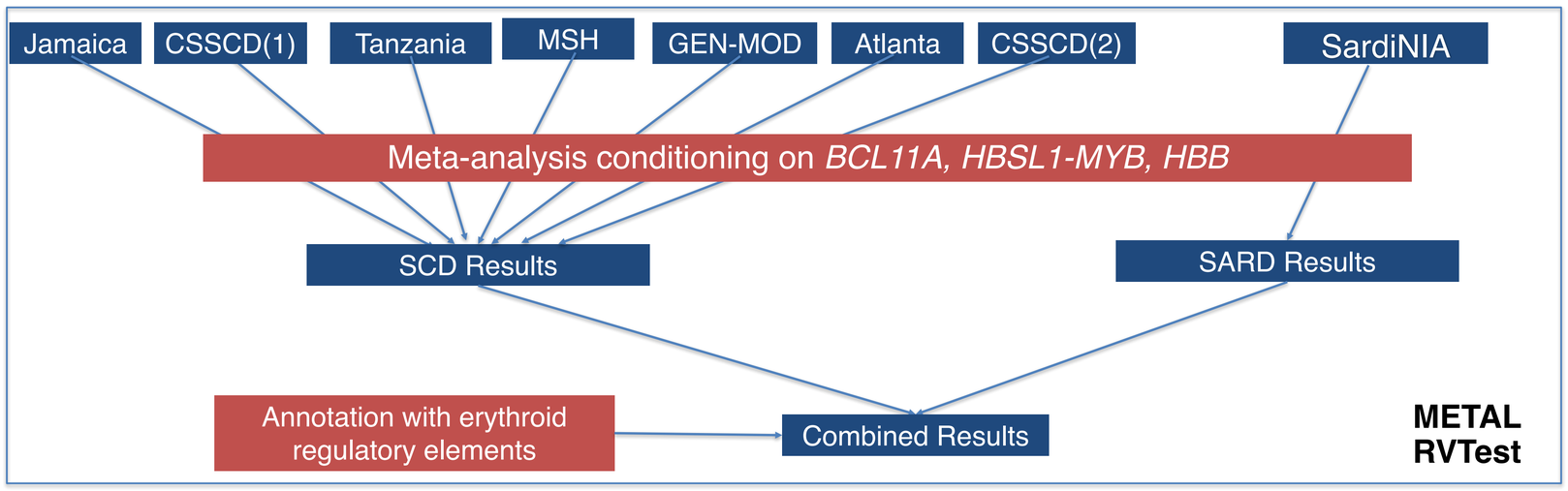

Figure 3. Conditional Meta-Analysis. The approach we took consisted in removing the GWAS signal from the three main HbF loci association for each cohort by conditional analysis, and then to combine all the summary statistics results. Both RVTest[3] and METAL[4] were used to perform the association tests.

Results

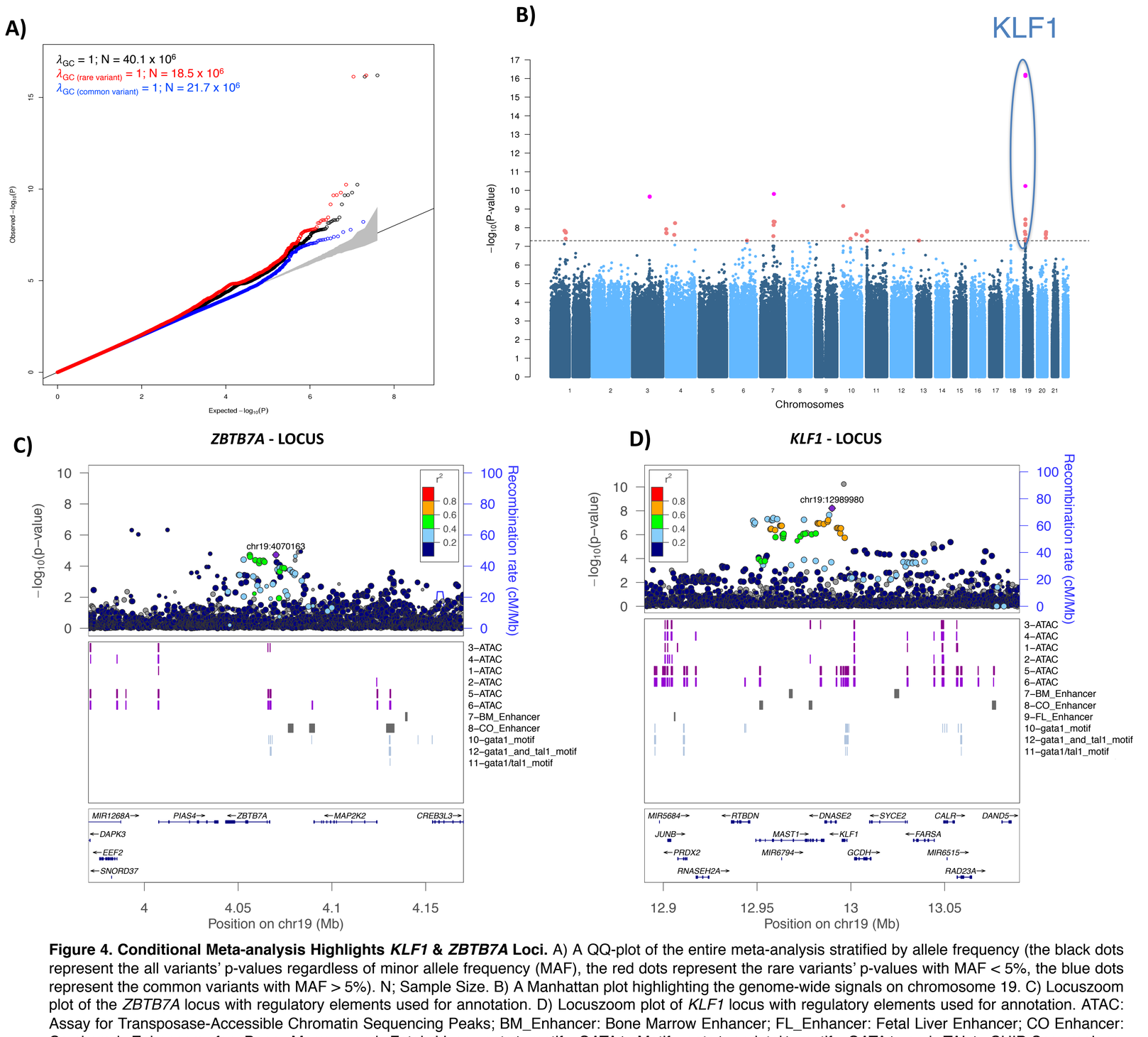

Figure 4. Conditional Meta-analysis Highlights *KLF1* & *ZBTB7A* Loci. A) A QQ-plot of the entire meta-analysis stratified by allele frequency (the black dots represent the all variants' p-values regardless of minor allele frequency (MAF), the red dots represent the rare variants' p-values with MAF < 5%, the blue dots represent the common variants with MAF > 5%). N; Sample Size. B) A Manhattan plot highlighting the genome-wide signals on chromosome 19. C) Locuszoom plot of the *ZBTB7A* locus with regulatory elements used for annotation. D) Locuszoom plot of *KLF1* locus with regulatory elements used for annotation. ATAC: Assay for Transposase-Accessible Chromatin Sequencing Peaks; BM_Enhancer: Bone Marrow Enhancer; FL_Enhancer: Fetal Liver Enhancer; CO Enhancer: Co-shared Enhancers for Bone Marrow and Fetal Liver; gata1_motif: GATA1 Motif; gata1_and_tal1_motif: GATA1 and TAL1 CHIP-Seq regions; gata1/tal1_motif: GATA1 and TAL1 co-bound regions.

Conclusions

We present the largest genetic study of HbF, where one novel genome-wide significant locus is on chromosome 19p13 (rs4804210/chr19:12989980, $P_{combined}$=6.1x10^{-9}, $P_{SardiNIA}$=1.6x10^{-6}, P_{SCD}=6.4x10^{-4}). HbF association results at this locus are independent from the previously reported HbF signal at the nearby *NFIX* gene[2]. Plus using the GTEx resource, we found that rs4804210 (and its linkage disequilibrium proxies) is an eQTL for *DNASE2* and *KLF1*. Further, this region physically interacts with the promoter of *CALR* as determined by Hi-C methodology. Although additional functional work is required, *KLF1* represents a strong candidate causal gene at this locus: (1) it encodes a key erythroid transcription factor, (2) it is mutated in patients with hereditary persistence of fetal hemoglobin, and (3) it regulates the expression of *BCL11A*. Another noteworthy locus is, on chromosome 19q13.11 (rs59394312/chr19:4070163, $P_{combined}$=1.9x10^{-5}, $P_{SardiNIA}$=7.5x10^{-3}, P_{SCD}=5.2x10^{-4}). Although, the *ZBTB7A* variant is not genome-wide significant the gene is known to increases HbF levels independently of *BCL11A*[5]. While results from the meta-analysis have yet to be replicated, they highlight the need for increase sample size for traits such as fetal hemoglobin to discover new loci.

References

1. Smith EC, Orkin SH. Hemoglobin genetics: recent contributions of GWAS and gene editing. *Hum Mol Genet.* 2016;25(R2):R99-R105.
2. Danjou F, Zoledziewska M, Sidore C, et al. Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. *Nat Genet.* 2015;47(11):1264-1271.
3. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics.* 2016;32(9):1423-1426.
4. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
5. Masuda T, Wang X, Maeda M, et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science.* 2016;351(6270):285-289