nature genetics

Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin

Joseph Borg[1,2,11], Petros Papadopoulos[3,11], Marianthi Georgitsi[4,11], Laura Gutiérrez[3], Godfrey Grech[1,2], Pavlos Fanis[3], Marios Phylactides[5], Annemieke J M H Verkerk[6], Peter J van der Spek[6], Christian A Scerri[1,2], Wilhelmina Cassar[1,2], Ruth Galdies[1,2], Wilfred van IJcken[7], Zeliha Özgür[7], Nynke Gillemans[3], Jun Hou[3,8], Marisa Bugeja[1,2], Frank G Grosveld[3,8,9], Marieke von Lindern[10], Alex E Felice[1,2], George P Patrinos[4] & Sjaak Philipsen[3,8]

Hereditary persistence of fetal hemoglobin (HPFH) is characterized by persistent high levels of fetal hemoglobin (HbF) in adults. Several contributory factors, both genetic and environmental, have been identified[1] but others remain elusive. HPFH was found in 10 of 27 members from a Maltese family. We used a genome-wide SNP scan followed by linkage analysis to identify a candidate region on chromosome 19p13.12–13. Sequencing revealed a nonsense mutation in the *KLF1* gene, p.K288X, which ablated the DNA-binding domain of this key erythroid transcriptional regulator[2]. Only family members with HPFH were heterozygous carriers of this mutation. Expression profiling on primary erythroid progenitors showed that KLF1 target genes were downregulated in samples from individuals with HPFH. Functional assays suggested that, in addition to its established role in regulating adult globin expression, KLF1 is a key activator of the *BCL11A* gene, which encodes a suppressor of HbF expression[3]. These observations provide a rationale for the effects of *KLF1* haploinsufficiency on HbF levels.

Hemoglobin (Hb) is composed of two α-like and two β-like globin chains, encoded by genes in the *HBA* and *HBB* clusters, respectively. Developmental regulation of globin genes results in the expression of stage-specific Hb variants (**Supplementary Fig. 1**). Persistent expression of HbF ameliorates the symptoms of β-thalassemia and sickle cell disease, and reactivation of the *HBG1* and *HBG2* genes in adults is therefore of substantial interest for the clinical management of β-type hemoglobinopathies. After birth, HbF is gradually replaced by adult hemoglobin (HbA)[4]. Residual amounts of HbF continue to be synthesized throughout life. In most adults, HbF contributes <2% to total Hb, but there is considerable variation[5]. Genetic studies have identified three loci that control HbF levels in adults: *HBB* (11p15.4)[6,7], *HBS1L-MYB* (6q23.3)[6,8,9] and *BCL11A* (2p16.1)[10,11]. Together, these loci account for <50% of the variation in HbF, indicating that additional loci are involved[5].

Genetic analysis of families in which HPFH is found is a particularly powerful approach by which to identify modifiers of HbF levels[8]. Here we describe a Maltese pedigree with HPFH. The proband (II-5; **Fig. 1a**) was referred to the clinic because of microcytosis. She presented with high HbF levels (19.5%). We recruited additional family members, and 10 of 27 tested showed HPFH (**Fig. 1a** and **Supplementary Table 1**), suggesting that inheritance of the trait was autosomal dominant. We excluded linkage to the *HBB* locus, indicating that a *trans*-acting factor was involved.

We performed a genome-wide linkage analysis on 27 family members to identify candidate loci for the HPFH modifier. We carried out whole-genome multipoint parametric linkage analysis using the Merlin program[12] with two software packages, easyLINKAGE[13] and dChip[14]. The analyses resulted in one significant linkage peak with $\log_{10}$ odds (LOD) scores of 2.7 and 4.2, respectively, on chromosome 19p13.12–13 (**Fig. 1b** and **Supplementary Fig. 2**). We performed these analyses using an autosomal dominant model, assuming a penetrance of 90% and 1% phenocopy rate. We found no evidence of significant linkage to the previously reported *trans*-acting HPFH loci at 2p16.1 (*BCL11A*)[10,11] and 6q23.3 (*HBS1L-MYB*)[6,8,9]. We further investigated these two loci by genotyping the five individual SNPs previously associated with increased HbF levels. These analyses ruled out involvement of the *HBS1L-MYB* locus and revealed that heterozygosity at SNP rs766432 in the *BCL11A* locus might have contributed to the increased HbF levels but was not the main determinant (**Supplementary Table 1**).

Individuals with HPFH had a consistent haplotype at 19p13.12–13, and the inferred haplotypes revealed that all such individuals shared one copy of an identical chromosome segment, presumably containing

[1]Laboratory of Molecular Genetics, Department of Physiology and Biochemistry, University of Malta, Msida, Malta. [2]Thalassaemia Clinic, Section of Pathology, Mater Dei Hospital, Msida, Malta. [3]Erasmus MC, Department of Cell Biology, Rotterdam, The Netherlands. [4]University of Patras, Department of Pharmacy, University Campus, Patras, Greece. [5]Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus. [6]Erasmus MC, Department of Bioinformatics, Rotterdam, The Netherlands. [7]Erasmus MC, Center for Biomics, Rotterdam, The Netherlands. [8]Netherlands Consortium for Systems Biology, Rotterdam, The Netherlands. [9]Center for Biomedical Genetics, Rotterdam, The Netherlands. [10]Erasmus MC, Department of Hematology, Rotterdam, The Netherlands. [11]These authors contributed equally to this work. Correspondence should be addressed to A.E.F. (alex.felice@um.edu.mt), G.P.P. (gpatrinos@upatras.gr) or S.P. (j.philipsen@erasmusmc.nl).

Received 2 April; accepted 23 June; published online 1 August 2010; doi:10.1038/ng.630

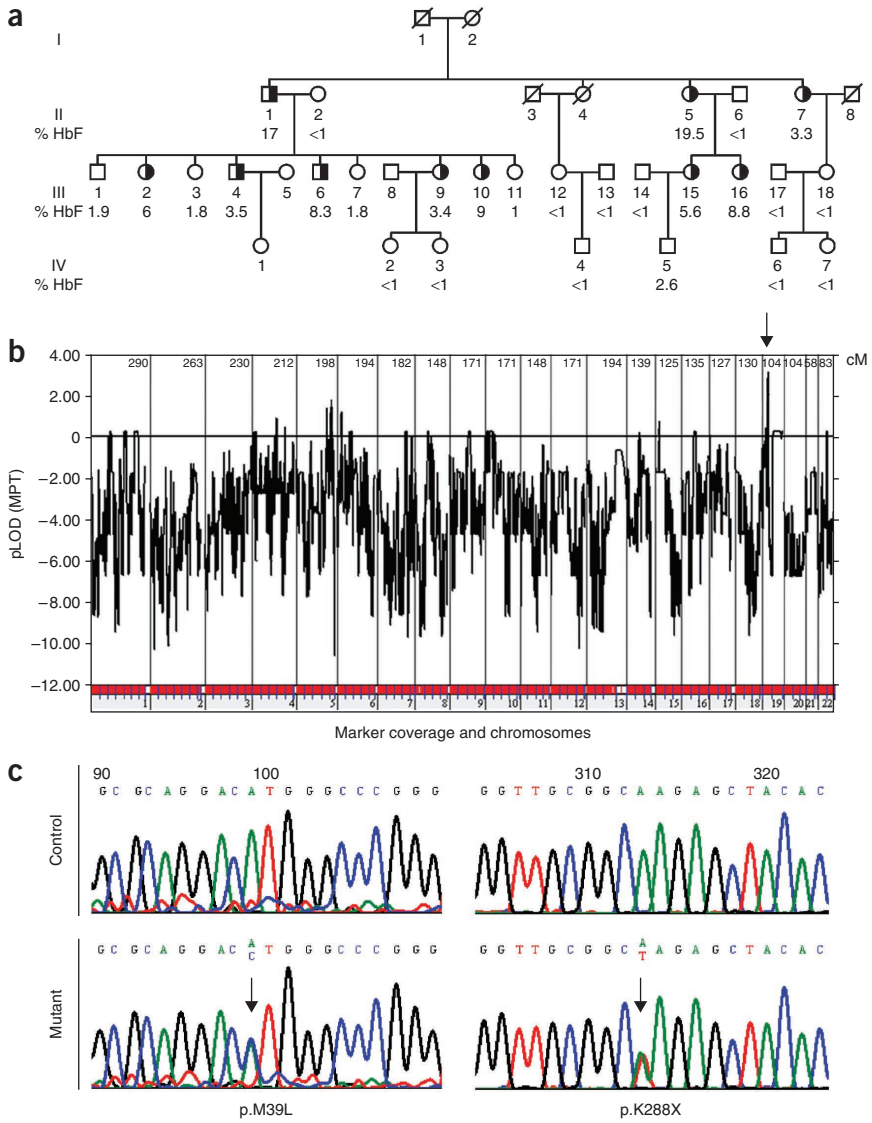

Figure 1 Chromosome 19 locus linked to HPFH in a Maltese family. (**a**) The Maltese HPFH pedigree. HbF levels are indicated as percentage of total Hb (%HbF). Individuals with HPFH are shown as half-filled symbols. (**b**) LOD scores derived from genome-wide linkage analysis. The putative HPFH locus on chromosome 19 is indicated by an arrow. pLOD, parametric LOD score; MPT, multipoint test; cM, centiMorgan. (**c**) Sequence analysis of *KLF1*. Individuals with HPFH were heterozygous for two mutations (arrows; **Supplementary Table 1**). The predicted effects of the mutations on KLF1 are shown below.

the putative HPFH locus (**Supplementary Fig. 2**). Recombination events delineating the linkage region are indicated with arrows. The distal boundary is determined by a recombination event in individuals IV-3 and IV-5 (**Supplementary Fig. 2**, white arrow). The proximal boundary is determined by individuals III-12, III-18, IV-6 and IV-7 (**Supplementary Fig. 2**, black arrow). These results narrowed the region down to a 663-kb interval between rs7247513 and rs12462609.

The *KLF1* gene, encoding a key erythroid transcriptional regulator[2], is found in this area. Mutations in *KLF1* have been reported as the molecular basis of the rare blood group In(Lu) phenotype[15] but have not been connected with HPFH. DNA sequencing revealed two linked mutations in *KLF1* that were found exclusively in all individuals with HPFH (**Fig. 1c**). The first mutation, p.M39L, is probably a neutral substitution as mouse Klf1 contains a leucine at this position[16]. The second mutation, p.K288X, is predicted to ablate the complete zinc finger domain and therefore abrogate DNA binding of the mutant protein[17]. The *KLF1* p.K288X variant was absent from a random sample drawn from the general Maltese population ($n = 400$).

To identify differentially expressed genes, we isolated RNA from erythroid progenitors (HEPs) cultured from peripheral blood[18] from four family members with HPFH and four without, and performed genome-wide expression analysis. By comparing the results to the reported gene expression profiles of mouse *Klf1* null erythroid progenitors[19], we identified a set of common differentially regulated genes (**Supplementary Table 2**). Cluster analysis with this set of genes separated the samples from individuals with HPFH from samples from those without (**Fig. 2a**), consistent with the notion that KLF1 activity is compromised in the family members with HPFH. Deregulation of these KLF1 target genes could explain the mild hypochromic microcytic indices shown by the individuals with HPFH (**Supplementary**

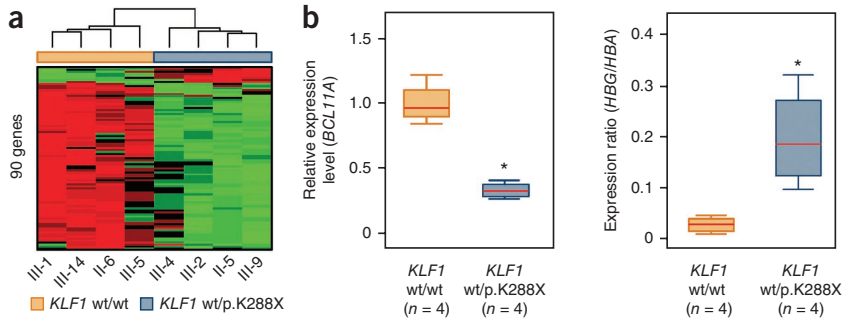

Figure 2 KLF1 target genes are downregulated in KLF1 p.K288X heterozygous HEPs. (**a**) RNA isolated from HEPs derived from healthy family members (wt/wt) and those with HPFH (wt/*KLF1* p.K288X) was used for genome-wide expression analysis. Deregulated genes common between wt/wt and wt/*KLF1* p.K288X and mouse wt/wt versus *Klf1* null mutant erythroid progenitors[19] (**Supplementary Table 2**) were used for cluster analysis. (**b**) Validation of key target genes by qPCR. Expression levels of *BCL11A* were normalized using *GAPDH* as a reference. Expression of *HBG1*/*HBG2* (*HBG*) was calculated as ratio to *HBA1*/*HBA2* (*HBA*) expression. Medians are indicated by red lines in the box plots. $^*P = 0.0209$. Error bars, s.d.

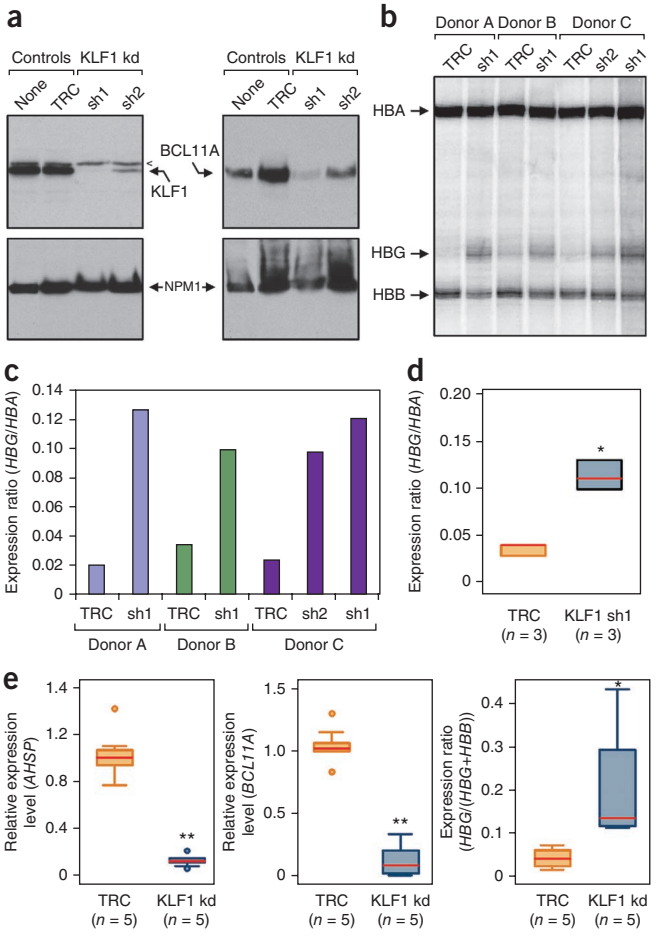

Figure 3 Increased *HBG1*/*HBG2* expression after knockdown of KLF1 in normal HEPs. (**a**) HEPs derived from normal donors were transduced with shRNA-expressing lentiviruses. Cells were harvested 5 d after transduction and nuclear extracts prepared. Top panels: KLF1 protein expression assessed by protein blot analysis. Level of BCL11A protein was decreased upon KLF1 knockdown (kd), bottom panels. NPM1 served as a loading control. None, mock transduction; TRC, control nonspecific shRNA; sh1 and sh2, two independent shRNAs targeting KLF1. A non-specific band is indicated by an arrowhead. (**b**) RNA was isolated from HEPs 5 d after transduction with the indicated lentiviruses and used in quantitative S1 nuclease protection assays to measure globin expression. Arrows indicate protected fragments diagnostic for *HBA1*/*HBA2* (*HBA*), *HBG1*/*HBG2* (*HBG*) and *HBB* mRNAs. (**c**) Quantification of data shown in **b** by phosphorimager analysis. (**d**) Box plots of *HBG*/*HBA* ratios after sh1-mediated KLF1 knockdown in HEPs derived from three independent donors. Medians are indicated by red lines. $*P = 0.0463$. (**e**) Box plots of qPCR analysis of *AHSP*, *BCL11A* and *HBG* expression after sh1/sh2-mediated KLF1 knockdown. *AHSP* is a known KLF1 target gene[28] and serves as a positive control. Expression levels of *AHSP* and *BCL11A* were normalized using *GAPDH* as a reference. Expression levels of *HBG* were calculated as ratio to total β-like globin expression (*HBG*+*HBB*) expression. Medians are indicated by red lines. Circles: points outside the range of the error bars. $*P = 0.020$; $**P < 0.003$. Error bars, s.d.

Table 1). Of note, the embryonic *Hbb-y* and *HBE1* genes were highly upregulated (**Supplementary Table 2**), whereas expression of the fetal globin repressor *BCL11A*[3] was downregulated in individuals with HPFH (**Supplementary Table 2** and **Supplementary Fig. 3**). We could not measure the expression of fetal and adult globins quantitatively on the microarrays owing to saturation effects. However, quantitative RT-PCR (qPCR) confirmed the downregulation of *BCL11A* and showed that the expression of *HBG1*/*HBG2* genes was increased in the samples from individuals with HPFH (**Fig. 2b**).

Next, we investigated the effects of KLF1 knockdown in HEPs derived from healthy donors. We obtained efficient knockdown of KLF1 with two out of five lentiviral shRNA constructs[20] tested (**Fig. 3a**). Quantitative S1 nuclease protection assays[21] showed that KLF1 knockdown led to a significant increase in *HBG1*/*HBG2* expression (**Fig. 3b**–**d**), which was confirmed by qPCR (**Fig. 3e**). In addition, we found that *BCL11A* expression was diminished after KLF1 knockdown, both at the protein (**Fig. 3a**) and at the mRNA level (**Fig. 3e**). Thus, the effects of KLF1 insufficiency on *HBG1*/*HBG2* and *BCL11A* expression in HEPs from healthy donors were similar to those observed in *KLF1* p.K288X heterozygotes, supporting the causative role of this mutation in the HPFH phenotype.

To further investigate this idea, we transduced HEPs with lentiviral vectors that expressed either the KLF1 p.K288X truncation mutant or full-length KLF1. The transgenic proteins were expressed at physiological levels in control HEPs (**Supplementary Fig. 4a**). This did not affect *HBG1*/*HBG2* expression (**Supplementary Fig. 4b**,**c**), indicating that the truncated form of KLF1 does not act as a dominant-negative factor. In HPFH HEPs, lentivirus-mediated expression of full-length KLF1 resulted in considerable downregulation of *HBG1*/*HBG2* mRNA, whereas expression of truncated KLF1 had no effect (**Fig. 4**). Levels of BCL11A protein were increased after transduction with full-length KLF1 lentivirus, whereas no such changes were observed after transduction with either GFP or truncated KLF1 lentiviral vectors (**Fig. 4a**).

The endogenous truncated KLF1 protein was not or at best barely detectable in HEPs from individuals with HPFH. This suggested that RNA transcribed from the *KLF1* p.K288X allele was subject to nonsense-mediated decay[22], further emphasizing that it was dysfunctional. Consistent with this notion, we found that *KLF1* mRNA expression was lower in HEPs from individuals with HPFH than in those from healthy donors (**Supplementary Fig. 3**).

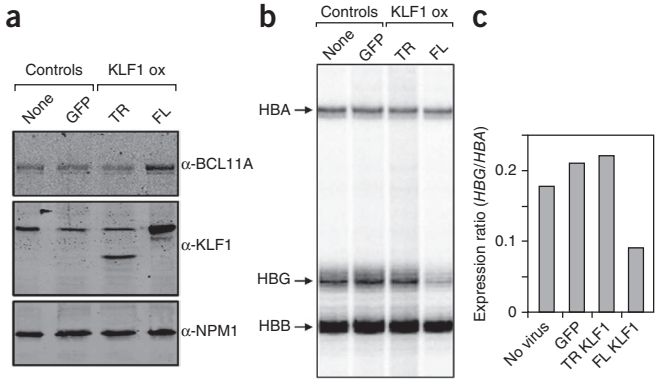

Figure 4 Expression of exogenous KLF1 in HEPs from donors with HPFH. (**a**) HEPs derived from individual II-5 were transduced with lentiviral constructs expressing GFP, KLF1 truncated at amino acid 288 (TR) or full-length KLF1 (FL). Seven days after transduction, nuclear extracts were prepared and expression of BCL11A and KLF1 was assessed by protein blot. NPM1 served as a loading control. ox, overexpression. (**b**) RNA was isolated from II-5 HEPs 7 d after transduction with the indicated lentiviruses and used for quantitative S1 nuclease protection assays to measure globin mRNA expression. Arrows indicate protected fragments diagnostic for *HBA1*/*HBA2* (*HBA*), *HBG1*/*HBG2* (*HBG*) and *HBB* mRNAs. (**c**) Quantification of data shown in **b** by phosphorimager analysis.

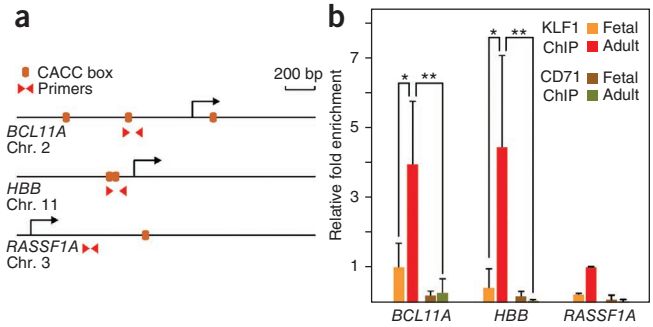

Figure 5 KLF1 binds to the promoter of the *BCL11A* gene *in vivo*. (**a**) Schematic drawings of the promoter areas of the *BCL11A*, *HBB* and *RASSF1A* genes. Positions of potential KLF1 binding sites (CACC boxes) and PCR primers used are shown. Arrows indicate transcription start sites. (**b**) ChIP analysis of KLF1 binding to the *BCL11A* promoter in human fetal liver cells and adult HEPs. The *HBB* promoter served as a positive control[26]. *RASSF1A* was used as a negative control, and the unrelated CD71 antibody served as a control for the specificity of the KLF1 antibody. $*P < 0.05$; $**P < 0.01$. Error bars, s.d.

KLF1 preferentially activates the *HBB* gene at the expense of *HBG1/HBG2* gene expression by interacting directly with regulatory elements in the *HBB* promoter[23–25]. The molecular analysis of the Maltese HPFH-affected family is consistent with this function of KLF1. In addition, our results also suggest a new potential mechanism by which KLF1 might tip the balance from *HBG1/HBG2* to *HBB* expression: through activation of the gene encoding the *HBG1/HBG2* repressor BCL11A[3]. The promoter region of the *BCL11A* gene contains several putative KLF1 binding sites (CACC boxes; **Fig. 5a**). We performed chromatin immunoprecipitation (ChIP) assays to investigate whether KLF1 was bound to the *BCL11A* promoter *in vivo*. We used human fetal liver erythroid progenitors, which express high levels of *HBG1/HBG2*, and HEPs from adult peripheral blood in which the *HBG1/HBG2* genes are suppressed. In adult HEPs, we observed strong binding of KLF1 to the *BCL11A* promoter (**Fig. 5b**). This was similar to the binding of KLF1 to the *HBB* promoter, which served as a positive control[26]. Neither promoter seemed to be bound by KLF1 in fetal liver-derived erythroid progenitors. ChIP reactions with the unrelated CD71 antibody were negative in all cases. We conclude that in adult HEPs KLF1 is bound to the *BCL11A* promoter *in vivo*.

Diminished KLF1 activity, mediated either through mutation of one *KLF1* allele (as occurs in the Maltese individuals with HPFH) or experimentally through shRNA-mediated knockdown in HEPs from normal donors, results in decreased *BCL11A* expression. Conversely, BCL11A levels were increased upon restoration of KLF1 activity in HEPs from Maltese family members with HPFH. This identifies KLF1 as a dual regulator of fetal-to-adult globin switching in humans (**Supplementary Fig. 5**). First, it acts on the *HBB* locus as a preferential activator of the *HBB* gene[27]. Second, it activates expression of *BCL11A*, which in turn represses the *HBG1/HBG2* genes. This dual activity ensures that, in most adults, HbF levels are <2% of total Hb.

In conclusion, we have identified haploinsufficiency for KLF1 as a cause of HPFH. We suggest that attenuation of KLF1 activity may be a fruitful approach to raise HbF levels in individuals with β-type hemoglobinopathies.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession code. The microarray expression data can be found at GEO under accession number GSE22109.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the family members for their cooperation; T. van Gent for help with statistical analysis; T. van Dijk for expertise in lentiviral technology; N. Papazian for preparation of human fetal liver samples; and M. Pizzuto for administrative support. The population samples from Malta were obtained from the Malta BioBank (EuroBioBank). Recombinant human erythropoietin (Epo) was a kind gift from Ortho-Biotech, and recombinant human Stem Cell Factor (SCF) was a kind gift from Amgen. Supported by the University of Malta and Mater Dei Hospital (A.E.F.); the Malta Government (J.B.); the European Molecular Biology Organization (J.B., P.P.); the Netherlands Scientific Organization (VENI 863.09.012 to L.G. and DN 82-294 and 912-07-019 to S.P.); the Netherlands Genomics Initiative, Erasmus MC (MRace; 296088 to S.P.); the Landsteiner Foundation for Blood Transfusion Research (0615 to S.P.); the Centre for Biomedical Genetics (F.G.G.); the European Commission FP6 EuTRACC consortium (037445 to F.G.G.); the US National Institutes of Health (NIH) (R01-HL073455 to F.G.G.); the Research Promotion Foundation of Cyprus (ΠΔΕ046_02 to G.P.P.); and the European Commission FP7 GEN2PHEN (200754 to G.P.P.).

AUTHOR CONTRIBUTIONS

F.G.G., A.E.F., G.P.P. and S.P. designed experiments. J.B., P.P., M.G., L.G., G.G., P.F., M.P., C.A.S., W.C., R.G., Z.Ö., N.G. and M.v.L. performed experiments. J.B., P.P., M.G., L.G. and G.G. analyzed results. P.J.v.d.S., F.G.G., A.E.F., G.P.P. and S.P. supervised data analysis. P.J.v.d.S., W.v.IJ. and M.B. provided expertise, analysis tools and infrastructure. A.J.M.H.V., J.H. and M.B. analyzed data. J.B., P.P., M.G., F.G.G., M.v.L., A.E.F., G.P.P. and S.P. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Serjeant, G. Geographic heterogeneity of sickle cell disease. in *Disorders of Hemoglobin* (eds. Steinberg, M.H., Forget, B.G., Higgs, D.R. & Nagel, R.L.) 895–906 (Cambridge Univ. Press, 2001).
2. Bieker, J.J. Probing the onset and regulation of erythroid cell-specific gene expression. *Mt. Sinai J. Med.* **72**, 333–338 (2005).
3. Sankaran, V.G. *et al.* Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322**, 1839–1842 (2008).
4. Stamatoyannopoulos, G. & Grosveld, F. Hemoglobin switching. in *The Molecular Basis of Blood Diseases* (eds. Stamatoyannopoulos, G., Majerus, P.W., Perlmutter, R.M. & Varmus, H.) 135–182 (WB Saunders Company, Philadelphia, 2001).
5. Thein, S.L., Menzel, S., Lathrop, M. & Garner, C. Control of fetal hemoglobin: new insights emerging from genomics and clinical implications. *Hum. Mol. Genet.* **18**, R216–R223 (2009).
6. Craig, J.E. *et al.* Dissecting the loci controlling fetal haemoglobin production on chromosomes 11p and 6q by the regressive approach. *Nat. Genet.* **12**, 58–64 (1996).
7. Gilman, J.G. & Huisman, T.H. DNA sequence variation associated with elevated fetal G gamma globin production. *Blood* **66**, 783–787 (1985).
8. Close, J. *et al.* Genome annotation of a 1.5 Mb region of human chromosome 6q23 encompassing a quantitative trait locus for fetal hemoglobin expression in adults. *BMC Genomics* **5**, 33 (2004).
9. Garner, C. *et al.* Haplotype mapping of a major quantitative-trait locus for fetal hemoglobin production on chromosome 6q23. *Am. J. Hum. Genet.* **62**, 1468–1474 (1998).
10. Lettre, G. *et al.* DNA polymorphisms at the *BCL11A, HBS1L-MYB*, and β-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc. Natl. Acad. Sci. USA* **105**, 11869–11874 (2008).
11. Menzel, S. *et al.* A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat. Genet.* **39**, 1197–1199 (2007).
12. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
13. Hoffmann, K. & Lindner, T.H. easyLINKAGE-Plus—automated linkage analyses using large-scale SNP data. *Bioinformatics* **21**, 3565–3567 (2005).
14. Leykin, I. *et al.* Comparative linkage analysis and visualization of high-density oligonucleotide SNP array data. *BMC Genet.* **6**, 7 (2005).
15. Singleton, B.K., Burton, N.M., Green, C., Brady, R.L. & Anstee, D.J. Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. *Blood* **112**, 2081–2088 (2008).

16. Miller, I.J. & Bieker, J.J. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol. Cell. Biol.* **13**, 2776–2786 (1993).
17. Feng, W.C., Southwood, C.M. & Bieker, J.J. Analyses of β-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J. Biol. Chem.* **269**, 1493–1500 (1994).
18. Leberbauer, C. *et al.* Different steroids co-regulate long-term expansion versus terminal differentiation in primary human erythroid progenitors. *Blood* **105**, 85–94 (2005).
19. Pilon, A.M. *et al.* Failure of terminal erythroid differentiation in EKLF-deficient mice is associated with cell cycle perturbation and reduced expression of E2F2. *Mol. Cell. Biol.* **28**, 7394–7401 (2008).
20. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
21. Antoniou, M. Induction of erythroid-specific expression in murine erythroleukemia (MEL) cell lines. *Methods Mol. Biol.* **7**, 421–434 (1991).
22. Mühlemann, O., Eberle, A.B., Stalder, L. & Zamudio Orozco, R. Recognition and elimination of nonsense mRNA. *Biochim. Biophys. Acta* **1779**, 538–549 (2008).
23. Donze, D., Townes, T.M. & Bieker, J.J. Role of erythroid Kruppel-like factor in human γ- to β-globin gene switching. *J. Biol. Chem.* **270**, 1955–1959 (1995).
24. Drissen, R. *et al.* The active spatial organization of the β-globin locus requires the transcription factor EKLF. *Genes Dev.* **18**, 2485–2490 (2004).
25. Zhou, D., Pawlik, K.M., Ren, J., Sun, C.W. & Townes, T.M. Differential binding of erythroid Krupple-like factor to embryonic/fetal globin gene promoters during development. *J. Biol. Chem.* **281**, 16052–16057 (2006).
26. Bottardi, S., Ross, J., Pierre-Charles, N., Blank, V. & Milot, E. Lineage-specific activators affect beta-globin locus chromatin in multipotent hematopoietic progenitors. *EMBO J.* **25**, 3586–3595 (2006).
27. Wijgerde, M. *et al.* The role of EKLF in human β-globin gene competition. *Genes Dev.* **10**, 2894–2902 (1996).
28. Drissen, R. *et al.* The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability. *Mol. Cell. Biol.* **25**, 5205–5214 (2005).

ONLINE METHODS

Molecular genetic analysis. The proband (II-5; **Fig. 1a**) was referred to the clinic because of microcytosis. She presented with 19.5% HbF, and therefore additional family members were approached to participate (**Supplementary Note**). Blood samples were obtained with informed consent and standard hematological indices were determined (**Supplementary Table 1**). Genomic DNA was extracted from ~1×10^6 cells from whole blood using a modified salting out procedure[29]. Control DNA samples isolated from 400 random Maltese individuals were available from the Laboratory of Molecular Genetics, Biomedical Sciences Building, University of Malta. The family members were genotyped in the *HBB*, *HBD* genes and the *HBG1*/*HBG2* gene promoters to detect point mutations and small insertions/deletions leading to β-thalassemia, δ-thalassemia or HPFH, respectively, using routine procedures[30]. Gap PCR was carried out to detect possible genomic rearrangements leading to deletional HPFH or δβ-thalassemia[31]. This excluded linkage of the HPFH phenotype to the *HBB* locus. Occurrence of common α-thalassemic mutations (SEA, 3.7 and 4.2 deletions) was also excluded. The NspI mapping 250K set (Affymetrix) was used to analyze 27 DNA samples from the HPFH family, starting with 250 ng of genomic DNA per array. Individual SNPs in the *HBS1L-MYB* (rs28384513, rs9399137, rs4895441) and *BCL11A* (rs766432, rs11886868) loci[32] were genotyped manually.

DNA linkage analysis. Multipoint parametric linkage analysis was performed using the Merlin v1.0.1 program[12] with two software packages (EasyLinkage v5.05 Beta[13] and dChip[14]) to calculate parametric LOD scores. Parametric analysis was carried out using an autosomal dominant mode of inheritance. Penetrances used for the dominant model were 0.01, 0.90 and 0.90 for the wild-type homozygote, mutant heterozygote and mutant homozygote, respectively. We assumed a disease allele prevalence frequency of 0.0001 and a phenocopy rate of 1%. A co-dominant allele frequency algorithm was used for the analysis. These analyses were carried out using the sex-averaged 500K Marshfield genetic map provided with the easyLINKAGE software package[13]. A Mendelian inheritance check was performed for all family members using the program PedCheck[33] and incompatibilities were omitted from the analysis. This increased the power and accuracy. The analysis was performed by taking HbF as a quantitative hematological value and classifying family members as 'affected' with HbF >2%, and 'nonaffected' with HbF <2%. Replicates of the linkage analysis and inferred haplotypes were constructed and visualized using dChip[14]. DNA from 400 random Maltese individuals was used to check for the presence of the *KLF1* p.K288X mutation in the population.

Cell culture. Human erythroid progenitor cells (HEPs) were cultured as described[18] in the presence of recombinant human Epo (1 U/ml, gift from Ortho-Biotech), recombinant human SCF (50 ng/ml, gift from Amgen) and dexamethasone (5×10^{-7} M; Sigma). Cells were counted with an electronic cell counter (CASY-1, Schärfe System).

Transcription profiling. A minimum of 1.5×10^6 HEPs were harvested at day 12 of culture and RNA was extracted with Trizol reagent (Sigma) and purified using the RNeasy Mini Kit (Qiagen), including an on-column DNaseI digestion, according to the manufacturer's instructions. RNA yield was determined using the 2100 Bioanalyzer (Agilent Technologies). We analyzed 8–10 μg of total RNA by microarrays using cells from day 12 of culture. The quality of the total RNA samples and the resulting cRNA was assessed on the Bioanalyzer. Fragmented biotinylated cRNA was prepared and 15 μg hybridized to HG-U133 plus 2 GeneChips according to the manufacturer's protocols (Affymetrix). The data files have been deposited in MIAME-compliant format in the NCBI GEO database (GSE22109). Single array expression analysis was performed using the Affymetrix GeneChip Operating Software (GCOS). A global scaling strategy was used to give an average target intensity of 500 for each array. Data from all eight arrays were filtered to exclude probe sets called either absent or marginal in all arrays. Control probe sets with the prefix AFFX were also removed before subsequent data analysis. Filtered data were transformed to a log2 scale and analyzed to determine differentially expressed genes. A 1.5-fold change threshold and test statistic of $P < 0.05$ were used as cutoff. A list of genes differentially expressed in mouse *Klf1* null erythroid progenitors ($P < 0.05$)[19] was downloaded from http://data.genome.duke.edu/EKLFDef.

Quantitative S1 nuclease protection assays. To measure globin mRNA levels directly, we used quantitative S1 nuclease protection assays[21]. The probe fragment for detection of *HBG1*/*HBG2* mRNAs was amplified by PCR using the primers S1-HBG-S and S1-HBG-A (**Supplementary Table 3**). Sizes of probes/protected fragments are: *HBA1*/*HBA2*: 700 nt/218 nt; *HBG1*/*HBG2*: 350 nt/165 nt; *HBB*: 525 nt/155 nt (ref. 21). Quantification was performed using a Typhoon Trio Phosphorimager (GE Healthcare) and corrected for specific activity of the probes.

qPCR analysis. Total RNA (1 μg) isolated from HEPs was converted to cDNA using SuperScript II reverse transcriptase according to the manufacturer's instructions (Invitrogen). Expression of mRNAs was analyzed by qPCR. Amplification reactions were performed with primers designed with Primer Express software v2.0 (Applied Biosystems). All amplifications used SYBR Green PCR Master Mix (Applied Biosystems). qPCR was performed with an Optical IQ Thermal Cycler (Bio-Rad Laboratories) with the following conditions: 50 °C for 2 min and 95 °C for 10 min, followed by 45 cycles of 95 °C for 15 s and 62 °C for 45 s. All reactions were performed in triplicate. Gene expression levels were calculated with the $2^{(-\Delta\Delta C(T))}$ method[34]. Target gene expression was normalized to *GAPDH* expression, unless indicated otherwise. Primers used are listed in **Supplementary Table 3**.

Statistical analysis. Statistical analysis of gene expression data obtained from quantitative S1 nuclease protection assays and qPCRs was performed with Mann Whitney tests using STATA data analysis and statistical software (StataCorp LP).

KLF1 expression constructs. A human *KLF1* cDNA clone (BC040000, Imagenes) was amplified by PCR with an *att*-specific set of primers (Invitrogen) to fuse the cDNA with a V5 tag at the carboxy terminus of the protein. Primers used were KLF1-F and KLF1-R1 (**Supplementary Table 3**). In parallel, part of the clone was amplified, truncating the protein at amino acid 288, with *att*-specific primers using a different reverse primer KLF1-R2. The PCR products were introduced into the lentiviral expression vector pRRLsin.sPPT.CMV.Wpre[35] modified for Gateway cloning (Invitrogen). The final clones were verified by sequencing.

Lentiviral transduction of human erythroid progenitors. Lentivirus was produced by transient transfection of 293T cells according to standard protocols[36]. Two days after transfection, the supernatant was collected, filtered and concentrated by centrifugation at 20,000 rpm for 2 h at 4 °C. HEPs cultured for 1 week were transduced in 24-well plates. We used 0.5×10^6 cells per well and sufficient amounts of virus to transduce ~80% of the cells. When appropriate, puromycin (1 μg/ml final concentration) was added to the cells after 2 d, and selection was performed for 2–3 d. At day 5–7 after transduction, cells were harvested and nuclear extracts were prepared[37]. RNA was extracted with the Trizol reagent. For knockdown experiments, clones from The RNAi Consortium (TRC[20]; Sigma) were used. The nontarget SHC002 vector was used as a control (SHC002: 5′-CAACAA GATGAAGAGCACCAA-3′). Five shRNA clones targeting *KLF1* were tested: TRCN0000016273, TRCN0000016274, TRCN0000016275, TRCN0000016276 and TRCN0000016277. Efficient knockdown of *KLF1* expression was observed with TRCN0000016276 (sh1) and TRCN0000016277 (sh2). Sequences are listed in **Supplementary Table 3**.

Protein blotting. Nuclear extracts were separated on denaturing polyacrylamide gels followed by semi-dry blotting to PVDF or nitrocellulose membranes. The membranes were probed with the following primary antibodies: BCL11A (sc-56013, Santa Cruz Biotechnology), NPM1 (ab10530, Abcam), KLF1 (ref. 26) and anti-V5-HRP (R961-25, Invitrogen). For detection, the appropriate secondary antibodies were used. The enhanced chemoluminescence kit (GE Healthcare) or the Odyssey Infrared Imaging System (Li-Cor Biosciences) was used to develop the membranes.

Chromatin immunoprecipitations. Fetal liver and adult HEPs were cultured[18] and used for ChIP reactions, which were performed as described[38] with the KLF1 antibody and a CD71 antibody (347510, BD Biosciences) as a

doi:10.1038/ng.630

negative control. qPCR was performed on the input and immunoprecipitated samples using primers for the *RASSF1A*, *HBB* and *BCL11A* genes. The relative fold enrichment was calculated as $2^{-[(CT\ x\ ChIP\ y - CT\ input\ y)-(CT\ KLF1\text{-}ChIP\ HEP\ RASSF1A - CT\ input\ HEP\ RASSF1A]}$ (where 'x' is the antibody and 'y' the sample), that is, setting the relative fold enrichment of the *RASSF1A* amplicon by the KLF1 antibody in HEPs to 1. Primers used are listed in **Supplementary Table 3**.

29. Miller, S.A., Dykes, D.D. & Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
30. Trifillis, P., Ioannou, P., Schwartz, E. & Surrey, S. Identification of four novel δ-globin gene mutations in Greek Cypriots using polymerase chain reaction and automated fluorescence-based DNA sequence analysis. *Blood* **78**, 3298–3305 (1991).
31. Craig, J.E., Barnetson, R.A., Prior, J., Raven, J.L. & Thein, S.L. Rapid detection of deletions causing δβ-thalassemia and hereditary persistence of fetal hemoglobin by enzymatic amplification. *Blood* **83**, 1673–1682 (1994).
32. Higgs, D.R. & Wood, W.G. Genetic complexity in sickle cell disease. *Proc. Natl. Acad. Sci. USA* **105**, 11595–11596 (2008).
33. O'Connell, J.R. & Weeks, D.E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63**, 259–266 (1998).
34. Schmittgen, T.D. & Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108 (2008).
35. Follenzi, A., Sabatino, G., Lombardo, A., Boccaccio, C. & Naldini, L. Efficient gene delivery and targeted expression to hepatocytes *in vivo* by improved lentiviral vectors. *Hum. Gene Ther.* **13**, 243–260 (2002).
36. Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery *in vivo*. *Nat. Biotechnol.* **15**, 871–875 (1997).
37. Andrews, N.C. & Faller, D.V. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res.* **19**, 2499 (1991).
38. Follows, G.A. *et al.* Epigenetic consequences of AML1-ETO action at the human c-FMS locus. *EMBO J.* **22**, 2798–2809 (2003).