

2014

Whole exome sequencing identifies novel genes for fetal hemoglobin response to hydroxyurea in children with sickle cell anemia.

V. A. Sheehan

J. R. Crosby

A. Sabo

N. A. Mortier

T. A. Howard

*See next page for additional authors*Follow this and additional works at: <https://academicworks.medicine.hofstra.edu/articles>Part of the [Pediatrics Commons](#)

Recommended Citation

Sheehan VA, Crosby JR, Sabo A, Mortier NA, Howard TA, Muzny DM, Dugan-Perez S, Aygun B, Nottage KA, Flanagan JM, . Whole exome sequencing identifies novel genes for fetal hemoglobin response to hydroxyurea in children with sickle cell anemia.. . 2014 Jan 01; 9(10):Article 2711 [110740 p.]. Available from: <https://academicworks.medicine.hofstra.edu/articles/2711>. Free full text article.

This Article is brought to you for free and open access by Donald and Barbara Zucker School of Medicine Academic Works. It has been accepted for inclusion in Journal Articles by an authorized administrator of Donald and Barbara Zucker School of Medicine Academic Works. For more information, please contact academicworks@hofstra.edu.

Authors

V. A. Sheehan, J. R. Crosby, A. Sabo, N. A. Mortier, T. A. Howard, D. M. Muzny, S. Dugan-Perez, Banu Aygun, K. A. Nottage, J. M. Flanagan, and +3 additional authors



Whole Exome Sequencing Identifies Novel Genes for Fetal Hemoglobin Response to Hydroxyurea in Children with Sickle Cell Anemia

Vivien A. Sheehan^{1*}, Jacy R. Crosby^{2,3}, Aniko Sabo⁴, Nicole A. Mortier⁵, Thad A. Howard⁵, Donna M. Muzny⁴, Shannon Dugan-Perez⁴, Banu Aygun⁶, Kerri A. Nottage⁷, Eric Boerwinkle^{3,4}, Richard A. Gibbs⁴, Russell E. Ware⁵, Jonathan M. Flanagan¹

1 Hematology Center, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, United States of America, **2** The University of Texas Graduate School of Biomedical Sciences at Houston, Department of Biostatistics, Bioinformatics, and Systems Biology, University of Texas, Houston, Texas, United States of America, **3** Human Genetics Center, University of Texas, Houston, Texas, United States of America, **4** Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, United States of America, **5** Division of Hematology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States of America, **6** Steven and Alexandra Cohen Children's Medical Center of New York, New Hyde Park, New York, United States of America, **7** Department of Hematology, St. Jude Children's Research Hospital, Memphis, Tennessee, United States of America

Abstract

Hydroxyurea has proven efficacy in children and adults with sickle cell anemia (SCA), but with considerable inter-individual variability in the amount of fetal hemoglobin (HbF) produced. Sibling and twin studies indicate that some of that drug response variation is heritable. To test the hypothesis that genetic modifiers influence pharmacological induction of HbF, we investigated phenotype-genotype associations using whole exome sequencing of children with SCA treated prospectively with hydroxyurea to maximum tolerated dose (MTD). We analyzed 171 unrelated patients enrolled in two prospective clinical trials, all treated with dose escalation to MTD. We examined two MTD drug response phenotypes: HbF (final %HbF minus baseline %HbF), and final %HbF. Analyzing individual genetic variants, we identified multiple low frequency and common variants associated with HbF induction by hydroxyurea. A validation cohort of 130 pediatric sickle cell patients treated to MTD with hydroxyurea was genotyped for 13 non-synonymous variants with the strongest association with HbF response to hydroxyurea in the discovery cohort. A coding variant in *Spalt-like transcription factor*, or *SALL2*, was associated with higher final HbF in this second independent replication sample and *SALL2* represents an outstanding novel candidate gene for further investigation. These findings may help focus future functional studies and provide new insights into the pharmacological HbF upregulation by hydroxyurea in patients with SCA.

Citation: Sheehan VA, Crosby JR, Sabo A, Mortier NA, Howard TA, et al. (2014) Whole Exome Sequencing Identifies Novel Genes for Fetal Hemoglobin Response to Hydroxyurea in Children with Sickle Cell Anemia. PLoS ONE 9(10): e110740. doi:10.1371/journal.pone.0110740

Editor: Wilbur Lam, Emory University/Georgia Institute of Technology, United States of America

Received: July 17, 2014; **Accepted:** September 15, 2014; **Published:** October 31, 2014

Copyright: © 2014 Sheehan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All 171 exome sequence files are available from dbGaP(php000691.v1.p1).

Funding: This work was supported by the National Human Genome Research Institute (U54-HG003273)(EB, RAG), National Heart, Lung and Blood Institute (U01-HL078787, R01-HL090941)(REW), Doris Duke Charitable Foundation (2010036)(REW), and the Russell and Diana Hawkins Family Foundation Discovery Fellowship (JRC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: vsheehan@bcm.edu

Introduction

Sickle cell anemia (SCA) is an inherited blood disorder, affecting 1 in 400 African Americans, causing significant morbidity and mortality. Although a monogenic disease, individuals with SCA (usually homozygous HbSS) exhibit wide variability in their laboratory and clinical phenotypes. One of the most powerful and reproducible modifiers of disease severity is an individual's endogenous level of fetal hemoglobin (HbF) [1]. If produced in sufficient amounts, HbF is able to prevent the intracellular polymerization of deoxygenated sickle hemoglobin (HbS), which is the nidus of the clinical disease process [2,3]. Pharmacologic induction of HbF is clinically beneficial, and the most widely used and safest method for increasing HbF levels in patients with SCA is treatment with hydroxyurea. Currently, it is the only FDA-

approved pharmacologic treatment for induction of HbF in adult patients with SCA, and is approved by the European Medicines Agency for both children and adults with SCA. Hydroxyurea significantly reduces pain and acute chest episodes, the need for blood transfusions and hospitalizations, and most importantly, reduces mortality [4–6]. While hydroxyurea has suspected disease modulating effects outside of HbF induction, the majority of its benefit is directly related to the amount of HbF produced in response to the drug [7,8]. There is an inverse relationship between levels of drug-induced HbF and number of pain episodes, hospitalizations, and overall mortality [9,10].

Several clinical studies have shown that individual hematological responses to hydroxyurea treatment are highly variable, with induced HbF levels ranging from 10% to greater than 30% HbF even for compliant patients on similar dosing regimens [11–14].

Previous efforts to identify predictors associated with final HbF produced in response to hydroxyurea have identified higher baseline HbF values, higher white blood cell count (WBC), and absolute reticulocyte count (ARC) as important factors [13,15,16]. However, none of these parameters can accurately predict the degree of HbF induction by hydroxyurea in an individual patient. From analysis of sibling pairs, it is known that the degree of HbF induction by hydroxyurea has a strong heritable component [17], indicating that genetic modifiers may have a large effect on drug response. Identification of specific genetic variants associated with HbF induction may elucidate reasons for this phenotypic variability and provide new insights into the drug's mechanisms of action related to HbF induction.

The aim of this study was to use a whole exome sequencing (WES) pharmacogenomics approach to identify genetic predictors of HbF response to hydroxyurea. Using two prospective pediatric cohorts with robust HbF phenotype data and standardized dose escalation regimen to MTD as a discovery cohort, we undertook a novel unbiased screen to test the entire exome for variants that are associated with hydroxyurea-induced HbF response levels (as measured by maximum %HbF at MTD [final HbF] or the change in %HbF from baseline to final [Δ HbF]). We focused on genetic variants with predicted functional effects on protein coding regions and identified several non-synonymous mutations that may influence the HbF response to hydroxyurea in children with SCA. We then validated a coding variant in *SALL2* in an unrelated, "real-world" cohort of children treated with hydroxyurea.

Results

Patient characteristics

Overall, both cohorts showed robust response to hydroxyurea with evidence of substantial individual variability in drug response (Table 1). All discovery cohort samples were genotyped for variants in *BCL11A* (rs1427407, rs4671393, rs11886868, rs7599488) and *HBSIL-MYB* (rs9399137, rs9402686); we found an association with baseline HbF for all *BCL11A* variants tested other than rs7599488. There was no significant association between the *BCL11A* variants tested and final HbF. No association with baseline, final or HbF was seen for either *HBSIL-MYB* variants tested. Linear association was performed with *BCL11A* variants as a covariate, without a significant change in associations.

At the time of hydroxyurea initiation, the average age of the 171 patients in the discovery cohort was 10.4 ± 4.5 years of age. The average age of the 130 patients in the validation cohort was 8.1 ± 4.0 years. All patients were treated under a similar dose escalation to MTD regimen according to protocol, or similar institutional guidelines [18, 19]. After a minimum of 6 months on hydroxyurea therapy, all patients reached a stable MTD (average 25.1 ± 4.5 mg/kg/day in the discovery cohort, 27.1 ± 4.3 mg/kg/day in the validation cohort) with predictable laboratory benefits (Table 1). The mean increase in HbF was $19.5 \pm 6.6\%$ in the discovery cohort and $13.9 \pm 7.0\%$ in the validation cohort, reflecting slight differences between the two groups of patients. There was evidence of consistent myelosuppression across both cohorts, however. The baseline HbF, distribution of Δ HbF at MTD and final HbF at MTD were all similar to prior reports (Figure 1, A–C). [15,20].

Whole exome sequencing

All 171 samples in the discovery cohort passed stringent WES quality control parameters with an average of 92% of the targeted

Table 1. Comparison of discovery and validation cohorts.

	Discovery Cohort		Validation Cohort		Discovery Cohort		Validation Cohort	
Age (years)	10.4	± 4.5	8.1	± 4.0	-	-	-	-
WBC ($\times 10^9/L$)	13.5	± 4.2	13.1	± 5.7	6.5	± 2.0	6.9	± 2.0
ANC ($\times 10^9/L$)	7.2	± 3.2	6.2	± 4.3	3.0	± 1.2	2.7	± 0.9
Hb (g/dL)	9.1	± 0.9	8.0	± 1.7	9.4	± 1.0	9.5	± 1.3
ARC ($\times 10^9/L$)	0.28	± 0.13	0.29	± 0.12	0.17	± 0.09	0.12	± 0.05
MCV (fL)	86.1	± 4.8	82.1	± 8.3	116.4	± 13.0	97.7	± 11.1
HbF (%)	8.0	± 4.9	11.9	± 5.9	27.6	± 7.3	25.8	± 8.2
Δ HbF (%)	-	-	-	-	19.5	± 6.6	13.9	± 7.0
HU dose (mg/kg/day)	-	-	-	-	25.1	± 4.5	27.1	± 4.3

WBC: white blood cell count; ANC: absolute neutrophil count; ARC: absolute reticulocyte count; MCV: mean corpuscular volume; HU: hydroxyurea. The discovery cohort was composed of 120 patients from HUSTLE and 51 from SWITCH. The validation cohort was collected from patients treated at TCCH. doi:10.1371/journal.pone.0110740.t001

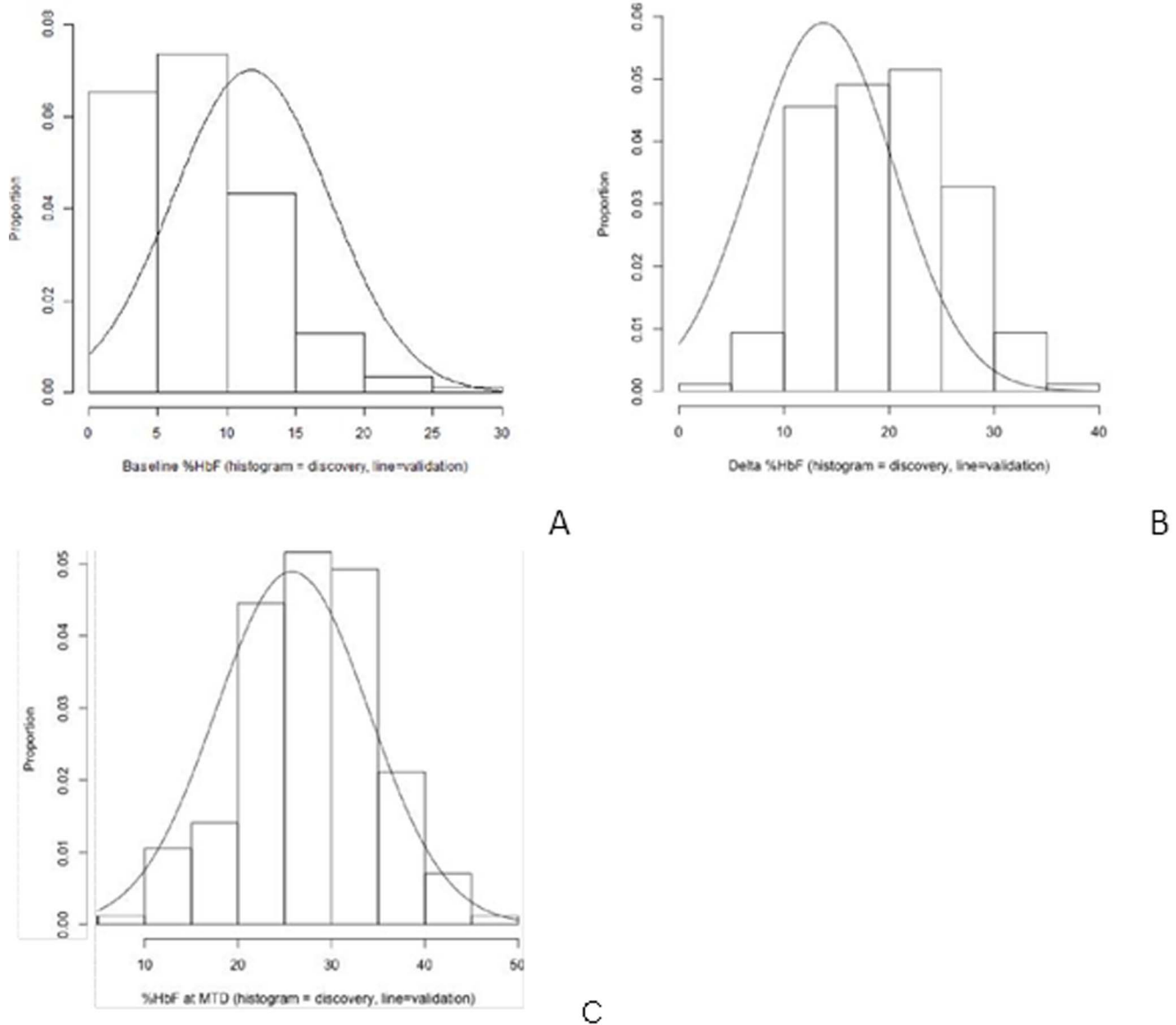


Figure 1. Comparison of discovery and validation cohorts. A, Baseline, or endogenous HbF for the discovery cohort is shown in binned histogram, and distribution of baseline HbF in validation cohort by a line plot. B, Delta HbF for the discovery cohort is shown in binned histogram, and distribution of delta HbF in validation cohort by a line plot. C, Final, or MID HbF for the discovery cohort is shown in binned histogram, and distribution of final, or MID in validation cohort by a line plot.
 doi:10.1371/journal.pone.0110740.g001

exonic regions sequenced at greater than $20\times$ coverage per individual. We identified a total of 278,639 autosomal variants, and 127,238 of these variants were non-synonymous or splice site variants expected to introduce an amino acid change in their encoded proteins (Table S1). For single variant association testing, we further filtered the non-synonymous variants ($n = 127,238$) for those with a minor allele frequency (MAF) greater than or equal to 2% ($n = 38,012$). We corrected for any population stratification using principal component analysis (PCA) performed by the EIGENSTRAT method.

As our phenotypes of interest are continuous variables, we used linear regression analysis to test the association of the 38,012 common ($\text{MAF} \geq 2\%$) non-synonymous variants using final HbF, and ΔHbF as independent, continuous variables. In addition, we attempted to find rare variants with $\text{MAF} < 2\%$ associated with

drug response by performing burden analysis with SKAT and T2 tests using the ΔHbF and final HbF phenotypes, but none of the gene level p-values were significant.

We identified 12 variants associated with ΔHbF with a p-value less than 5×10^{-4} (Table 2). In addition, we identified 13 variants associated with final HbF, also with a p-value less than 5×10^{-4} (Table 3). Although none of the p-values achieved genome wide significance level ($p < 1.3 \times 10^{-6}$), these results offered suggestive signals of potential associations. We used the existing methods of SIFT and PolyPhen2 [21] for predicting the functional impact of each non-synonymous variant to estimate which of the 25 variants had a predicted damaging or benign effect on encoded protein function (Tables 2 and 3).

From these 25 variants, we identified 13 variants with strongest association with response to hydroxyurea and predicted to

Table 2. Variants associated with Δ HbF on hydroxyurea.

Gene	SNP ID	Amino Acid Change	Protein Function Prediction	MAF (%)	Function	Beta Value	SE	P-value
<i>EML1</i>	rs141631682	Gly109Asp	Damaging	2	Microtubule assembly	10.9	2.7	7.41×10^{-5}
<i>SESN1</i>	rs2273668	Leu103Ile	Damaging	2	Peroxisome reduction	9.3	2.3	8.88×10^{-5}
<i>PAPLN</i>	rs17126352	Val416Ile	Benign	3	Metalloprotease	7.1	1.8	1.35×10^{-4}
<i>DCH52</i>	rs61746132	Pro1676Lys	Damaging	12	Calcium dependent cell adhesion	3.9	1.0	1.73×10^{-4}
<i>KRT80</i>	rs61749462	Arg364Ser	Damaging	3	Keratin 80	8.0	2.1	1.80×10^{-4}
<i>SALL2</i>	rs61743453	Pro840Arg	Damaging	4	Transcription factor	6.7	1.8	2.37×10^{-4}
<i>NOM1</i>	rs61742645	Arg779Cys	Benign	2	PP1 interacting protein	8.7	2.3	2.43×10^{-4}
<i>N4BP2L2</i>	rs35108810	Asp246Val	Damaging	15	ELA2 transcription inhibitor	-3.3	0.9	3.47×10^{-4}
<i>RNF113B</i>	rs16955011	Val92Met	Damaging	18	Ring finger protein	3.1	0.8	3.76×10^{-4}
<i>FTSJ2</i>	rs55904231	Ser11Phe	Benign	7	RNA methyltransferase	5.2	1.4	3.76×10^{-4}
<i>RHPN2</i>	rs28626308	Arg70Gln	Damaging	7	Rho GTPase binding protein	5.2	1.4	3.91×10^{-4}
<i>ADAR</i>	rs17843865	Tyr587Cys	Benign	4	Adenosine deaminase	-6.4	1.8	4.65×10^{-4}

Variants selected were predicted damaging, with a p-value <0.001 in the discovery cohort, $n=171$, composed of patients from HUSTLE and SWITCH trials. doi:10.1371/journal.pone.0110740.t002

introduce an amino acid change that has a damaging effect on protein structure or function. We genotyped these 13 variants by TaqMan PCR in our independent validation cohort of 130 patients with SCA. We found that one of the 13 variants, located in the *SALL2* gene, maintained association with hydroxyurea response. In the discovery cohort, the P840R variant in the *SALL2* gene (rs61743453) was associated with a higher change in HbF in response to hydroxyurea ($p = 2.37 \times 10^{-4}$, beta value 6.7). In the validation cohort, this same P840R variant was associated with a higher final HbF, with a p-value of 0.05, and a beta value of 4.2. Using Fisher's combined probability test method, a meta analysis of the association of the *SALL2* variant with Δ HbF in the discovery and validation sample groups ($n=301$) leads to a combined p-value of 8.30×10^{-4} . The meta analysis of the association of *SALL2* with final HbF in the discovery and validation sample groups resulted in a combined p-value of 1.48×10^{-4} .

Discussion

Many individuals with SCA are prescribed hydroxyurea, and there is evidence that genetic modifiers affect individual response [17,22]. In order to identify novel candidate genes and variants associated with hydroxyurea response, we sequenced the exomes of 171 individuals enrolled in two prospective clinical trials and related their sequence variant data to HbF response. This discovery cohort was obtained from patients treated on protocol, with the highest level of drug compliance supervision, including monthly pill counts. Our validation cohort ($n=130$) was treated under guidelines similar to that of the discovery cohort. As expected, individual MTD was achieved at different hydroxyurea doses, within a range of 10–35 mg/kg/d, reflecting the typical range in bioavailability among patients. There was no correlation between hydroxyurea dose and HbF response ($p = 0.56$), supporting the conclusion that differences in pharmacokinetics and pharmacodynamics affect HbF levels achieved on hydroxyurea [22].

Whole exome sequencing permitted analysis of genes beyond a usual set of *a priori* biologic candidate genes for this phenotype and variants across a broad allele frequency spectrum. We identified multiple non-synonymous variants associated with Δ HbF or final HbF (Tables 2 and 3) in the discovery cohort. We then performed genotyping on a validation cohort for 13 candidate variants with the lowest p-values and were also predicted to be damaging. Of the 13 variants genotyped, the variant in *SALL2* was associated with a higher HbF in the discovery cohort, and a higher final HbF at MTD in the discovery cohort. The validated variant in *SALL2* represents a novel variant not previously implicated in α -globin expression or HbF response to hydroxyurea. Further studies of other sickle cell cohorts treated with hydroxyurea are needed to confirm this promising association.

SALL2 is a multi-zinc finger transcription factor implicated in hematopoietic cell maturation and cell cycle arrest [23]. It contains the same conserved 12 amino acid N-terminal motif as BCL11A [24,25]. This motif has been shown to be essential for binding of the nucleosome remodeling and deacetylase co-repressor complex, or NuRD, which includes the histone deacetylases HDAC1 and HDAC2. Both HDAC1 and HDAC2 have been shown to act as co-repressors of gamma globin [26]. The variant identified here (rs61743453), causes a proline to arginine change at residue 840 and is predicted to be damaging to protein function. This P840R *SALL2* variant was associated with a higher HbF response to hydroxyurea (Figure 2). Further function-

Table 3. Variants associated with final HbF on hydroxyurea.

Gene	SNP ID	Amino Acid Change	Protein Function Prediction	MAF (%)	Function	Beta Value	SE	P-value
<i>RSPH3</i>	rs61750777	Ala154Val	Damaging	6	Radial spoke protein	-6.5	1.6	6.55×10^{-5}
<i>OLR1</i>	rs11053646	Lys167Asn	Benign	2	Opioid receptor	3.4	0.9	1.22×10^{-4}
<i>SEC31B</i>	rs11819496	Arg478Thr	Damaging	2	ER transport	-12.0	3.1	1.55×10^{-4}
<i>COPE</i>	rs34510432	Arg85 His	Damaging	2	ER transport	9.9	2.6	1.88×10^{-4}
<i>RNF113B</i>	rs16955011	Val92Met	Damaging	18	Ring finger protein	3.4	0.9	1.93×10^{-4}
<i>CDHR3</i>	rs6967330	Cys529Tyr	Benign	29	Calcium dependent cell adhesion	3.1	0.8	2.58×10^{-4}
<i>ETAA1</i>	rs3770655	Pro771Ser	Benign	89	Ewing's tumor associated antigen	4.3	1.2	2.84×10^{-4}
<i>TTL10</i>	rs113596156	Arg35Gln	Benign	2	Polyglycolase	9.8	2.7	2.91×10^{-4}
<i>DOCK1</i>	rs869801	Ala1793Thr	Benign	21	Cytokinesis	-3.5	1.0	3.60×10^{-4}
<i>MYBBP1A</i>	rs899441	Lys637Glu	Benign	6	MYB associated	-5.4	1.5	4.26×10^{-4}
<i>MARCH10</i>	rs116835087	Gly587Ser	Damaging	6	Membrane-associated ring finger	-6.0	1.7	4.33×10^{-4}
<i>PKD1L1</i>	rs11972142	Thr879Ala	Benign	15	Polycystic kidney disease like	3.6	1.0	4.78×10^{-4}
<i>ADCY10</i>	rs16859886	Thr234Met	Benign	12	Adenyl cyclase	3.9	1.1	4.96×10^{-4}

Variants selected were predicted damaging, with a p-value <0.001 in the discovery cohort, n=171, composed of patients from HUSTLE and SWITCH trials.
doi:10.1371/journal.pone.0110740.t003

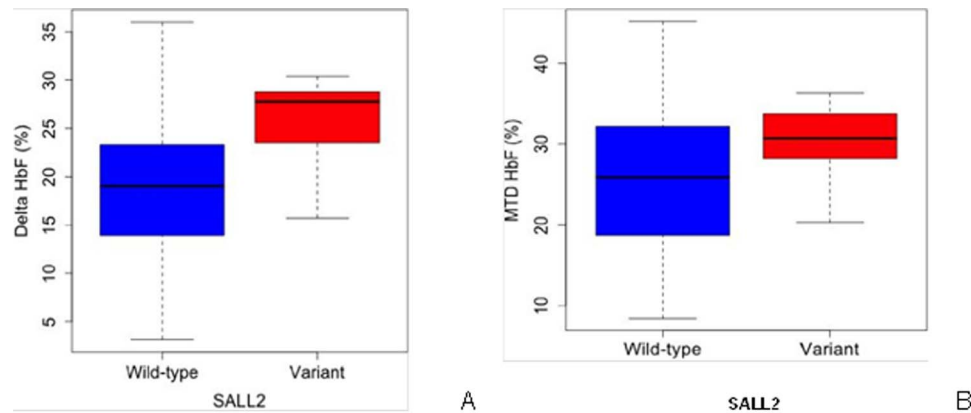


Figure 2. Effect of *SALL2* variant rs61743453 on HbF response to hydroxyurea. A, Effect of rs61743453 on delta HbF in discovery cohort. B, Effect of rs61743453 on MTD HbF in validation cohort. Variant refers to the Pro840Arg variant; no individuals were homozygous for this change. doi:10.1371/journal.pone.0110740.g002

al studies will help establish the role of *SALL2* in HbF induction in the context of hydroxyurea.

Despite the relatively small sample size, we used the best genotype-phenotype pairs available from prospectively treated pediatric patients from two clinical trials for the discovery cohort. We assembled a validation cohort from patients treated with hydroxyurea according to standard of care and expert guidelines in a pediatric hematology center with an established sickle cell program.

This study may have failed to detect loci with modest effects because of low statistical power and some associations identified in the discovery cohort may have failed validation given the small size of the validation cohort. Accordingly, all of the mutations identified in the discovery cohort may represent coding variants worth pursuing in future studies. The data presented here bodes well for the success of larger collaborative efforts aimed at identifying genetic modifiers of hydroxyurea response, and serves as a call for coordinated collaboration among pediatric sickle cell centers to increase sample size and increase the odds for novel discovery and translational potential.

Methods

Subjects

The discovery cohort was composed of 171 unrelated children with SCA; 120 were enrolled in the Hydroxyurea Study of Long-Term Effects (HUSTLE, NCT00305175); and 51 were enrolled in the NHLBI-sponsored Stroke with Transfusions Changing to Hydroxyurea (SWITCH, NCT00122980). HUSTLE was a single center trial investigating long term effects of hydroxyurea in SCA, while SWITCH was a multi-center trial investigating the use of hydroxyurea on stroke prevention. HUSTLE and SWITCH study patient samples were used with approval from the Baylor College of Medicine Internal Review Board, protocol H-29047. Patients and their families in both clinical trials provided written consent for DNA sample collection, storage and sequencing. Texas Children's Hospital Hematology Center patients and their families provided written consent to whole exome sequencing, posting of sequences to dbGAP, and data collection under BCM Internal Review Board protocol H31356. All DNA samples and data in this study were denormalized for analysis. All 171 individuals had a known baseline HbF level measured at greater than 3 years of age, were initially treated with hydroxyurea at 20 mg/kg, and then

dose-escalated to mild myelosuppression using a standardized regimen [18,19].

The validation cohort contained 130 unrelated children with SCA followed at the Texas Children's Hospital Hematology Center (TCHHC). All patients receiving hydroxyurea at TCHHC were approached for enrollment in an Internal Review Board-approved protocol for genetic analysis. They were treated with hydroxyurea using institutional guidelines rather than a specific protocol, but all were escalated to MTD following a standardized regimen as previously described [19]. All patients were treated with hydroxyurea for at least 6 months prior to the designated MTD timepoint. Total HbF levels for both discovery and validation cohorts were measured by HPLC.

Ethics Statement

All patients and their families gave informed consent for genomic DNA sample collection, storage, and sequencing. The WES genomic analyses were approved by the Baylor College of Medicine Institutional Review Board. All DNA samples and data in this study were denormalized for analysis.

Whole exome sequencing

DNA concentrations were quantified using picogreen fluorescent detection method (Quant-iT, Invitrogen). For each DNA sample, the entire exome was captured using the NimbleGen VCRome 2.1 capture reagent followed by sequencing on an Illumina platform using standard chemistries. The sequencing reads were mapped to Hg19 reference genome using the BWA [27]. Sample level genome variants were identified and annotated using the Human Genome Sequencing Center's integrated Mercury pipeline which includes quality score recalibration and insertion/deletion (Indel) realignment, genome variant identification by AtlasSNP [28], and annotation using Cassandra software. A project level variant call format (VCF) was generated for all the samples, and included variants that were present in at least one sample. Variants with more than 5% missed genotyping calls were excluded from analysis. Error threshold for alignment was two base errors per read with penalties for indels (maximum of 1) are much more costly than the penalties for SNVs (maximum of 2). We allowed for read-trimming to 35 bp.

WES genotyping of variants of interest with heterozygosity scores less than 0.45 were verified by TaqMan genotyping. TaqMan genotyping assays were performed on an Applied Biosystem's StepOne instrument (AB, Foster City, CA). After 40

amplification cycles, threshold cycle values were automatically calculated, and the individual SNP genotypes were called by the StepOne v2.0 software (AB, Foster City, CA).

Statistical Analysis

Linear regression analysis was used to test the association of the filtered variants using final HbF and Δ HbF as independent, continuous variables. The Δ HbF and final HbF phenotypes both had normal distributions, indicating they were suitable for linear regression analysis (Figure 1); values were adjusted for age and gender. Principal component analysis (PCA) was performed using the EIGENSTRAT method, and applied to all models. Quality control filtering steps, including SNP missingness check, removal of sex chromosomes, monomorphic site, synonymous and intronic variant removal, excess heterozygosity filter and minor allele frequency (MAF) cut-off of 2% and the impact of these filtering steps on the total number of SNPs, are described in Table S1.

To analyze the effect of rare variants (MAF<2%) on the phenotypes Δ HbF and final HbF, we used two gene based tests, a simple burden test (T2) and SKAT [29,30]. In SKAT and T2 testing, a collection of rare variants within a single gene are tested for association with the phenotype. The T2 considers those rare variants with MAF less than 2% and assumes that the effects of all variants are in the same direction. SKAT is a kernel-based test that considers rare variants having effects in either direction [29,31]. Both tests considered only non-synonymous variants.

Validation

Genomic DNA from 130 patients from TCHHC collected as a validation cohort were genotyped by TaqMan or Sanger

sequencing for 13 SNPs with the lowest p-values that were identified as associated with HbF response to hydroxyurea, non-synonymous, and damaging. The relationship between genotype and HbF response to hydroxyurea was analyzed with a one directional t-test.

Supporting Information

Table S1 Quality control filters used in WES analysis. Sex chromosomes were removed, as gender did not impact HbF response to hydroxyurea. Sites with heterozygous to homozygous ratio >0.4 were removed. Variants with MAF < 2% were analyzed by burden testing.
(PDF)

Acknowledgments

We thank all the patients and families, as well as the clinical investigators and research staff, for their participation in the SWITCH and HUSTLE trials. We would like to thank the patients, their families, and the research staff at Texas Children's Hematology Center for their participation in FWES, which provided the validation cohort.

Author Contributions

Conceived and designed the experiments: VAS JMF EB RAG REW. Performed the experiments: TAH AS DMM SDP. Analyzed the data: JRC EB. Contributed reagents/materials/analysis tools: RAG EB. Wrote the paper: VAS JMF REW EB. Clinical trial data analysis and sample collection: BA KAN NAM REW.

References

- Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O, et al. (1994) Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N Engl J Med* 330: 1639–1644.
- Cheetham RC, Huehns ER, Rosemeyer MA (1979) Participation of haemoglobins A, F, A2 and C in polymerisation of haemoglobin S. *J Mol Biol* 129: 45–61.
- Powars DR, Weiss JN, Chan LS, Schroeder WA (1984) Is there a threshold level of fetal hemoglobin that ameliorates morbidity in sickle cell anemia? *Blood* 63: 921–926.
- Steinberg MH, McCarthy WF, Castro O, Ballas SK, Armstrong FD, et al. (2010) The risks and benefits of long-term use of hydroxyurea in sickle cell anemia: A 17.5 year follow-up. *Am J Hematol* 85: 403–408.
- Voskaridou E, Christoulas D, Bilalis A, Plata E, Varvagiannis K, et al. (2010) The effect of prolonged administration of hydroxyurea on morbidity and mortality in adult patients with sickle cell syndromes: results of a 17-year, single-center trial (LaSHS). *Blood* 115: 2354–2363.
- Lobo CL, Pinto JF, Nascimento EM, Moura PG, Cardoso GP, et al. (2013) The effect of hydroxycarbamide therapy on survival of children with sickle cell disease. *Br J Haematol* 161: 852–860.
- Lebensburger J, Johnson SM, Askenazi DJ, Rozario NL, Howard TH, et al. (2011) Protective role of hemoglobin and fetal hemoglobin in early kidney disease for children with sickle cell anemia. *Am J Hematol* 86: 430–432.
- Lebensburger JD, Pestina TI, Ware RE, Boyd KL, Persons DA (2010) Hydroxyurea therapy requires HbF induction for clinical benefit in a sickle cell mouse model. *Haematologica* 95: 1599–1603.
- Steinberg MH, Barton F, Castro O, Pegelow CH, Ballas SK, et al. (2003) Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: risks and benefits up to 9 years of treatment. *JAMA* 289: 1645–1651.
- Smith WR, Ballas SK, McCarthy WF, Bauserman RL, Swerdlow PS, et al. (2011) The association between hydroxyurea treatment and pain intensity, analgesic use, and utilization in ambulatory sickle cell anemia patients. *Pain Med* 12: 697–705.
- Maier-Redelsperger M, de Montalembert M, Flahault A, Neonato MG, Ducrocq R, et al. (1998) Fetal hemoglobin and F-cell responses to long-term hydroxyurea treatment in young sickle cell patients. The French Study Group on Sickle Cell Disease. *Blood* 91: 4472–4479.
- Kinney TR, Helms RW, O'Branski EE, Ohene-Frempong K, Wang W, et al. (1999) Safety of hydroxyurea in children with sickle cell anemia: results of the HUG-KIDS study, a phase I/II trial. Pediatric Hydroxyurea Group. *Blood* 94: 1550–1554.
- Zimmerman SA, Schultz WH, Davis JS, Pickens CV, Mortier NA, et al. (2004) Sustained long-term hematologic efficacy of hydroxyurea at maximum tolerated dose in children with sickle cell disease. *Blood* 103: 2039–2045.
- Ware RE, Aygun B (2009) Advances in the use of hydroxyurea. *Hematology Am Soc Hematol Educ Program*: 62–69.
- Ware RE, Eggleston B, Redding-Lallinger R, Wang WC, Smith-Whitley K, et al. (2002) Predictors of fetal hemoglobin response in children with sickle cell anemia receiving hydroxyurea therapy. *Blood* 99: 10–14.
- Green NS, Ender KL, Pashankar F, Driscoll C, Giardina PJ, et al. (2013) Candidate sequence variants and fetal hemoglobin in children with sickle cell disease treated with hydroxyurea. *PLoS One* 8: e55709.
- Steinberg MH, Voskaridou E, Kutlar A, Loukopoulos D, Koshy M, et al. (2003) Concordant fetal hemoglobin response to hydroxyurea in siblings with sickle cell disease. *Am J Hematol* 72: 121–126.
- Ware RE, Helms RW (2011) Stroke With Transfusions Changing to Hydroxyurea (SWITCH). *Blood* 119: 3925–3932.
- Ware RE (2009) How I use hydroxyurea to treat young patients with sickle cell anemia. *Blood* 115: 5300–5311.
- Steinberg MH (2001) Modulation of fetal hemoglobin in sickle cell anemia. *Hemoglobin* 25: 195–211.
- Flanagan SE, Patch AM, Ellard S (2010) Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers* 14: 533–537.
- Ware RE, Despotovic JM, Mortier NA, Flanagan JM, He J, et al. (2011) Pharmacokinetics, pharmacodynamics, and pharmacogenetics of hydroxyurea treatment for children with sickle cell anemia. *Blood* 118: 4985–4991.
- Chai L (2011) The role of HSAL (SALL) genes in proliferation and differentiation in normal hematopoiesis and leukemogenesis. *Transfusion* 51 Suppl 4: 87S–93S.
- Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, et al. (2008) Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* 322: 1839–1842.
- Lauberth SM, Rauchman M (2006) A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. *J Biol Chem* 281: 23922–23931.
- Bradner JE, Mak R, Tanguturi SK, Mazitschek R, Haggarty SJ, et al. (2010) Chemical genetic strategy identifies histone deacetylase 1 (HDAC1) and HDAC2 as therapeutic targets in sickle cell disease. *Proc Natl Acad Sci U S A* 107: 12617–12622.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

28. Challis D, Yu J, Evani US, Jackson AR, Paithankar S, et al. (2012) An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics* 13: 8.
29. Lee S, Wu MC, Lin X (2012) Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 13: 762–775.
30. Li B, Leal SM (2008) Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 83: 311–321.
31. Morgenthaler S, Thilly WG (2007) A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res* 615: 28–56.