Aus der Klinik für Neonatologie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Using Genome-wide Association and Exome Sequencing Studies to Improve *In Vitro* Production of Red Blood Cells and to Elucidate Human Traits and Congenital Disorders

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Felix Caspar Giani

aus Oldenburg

Datum der Promotion: 26.06.2022

Table of Contents

Thesis Summary.. 3

1. Abstract.. 3
   1.1 Abstract Deutsch.. 4
2. Introduction.. 6
   2.1 Using human genetic variation to improve cell and tissue replacement therapies........ 7
   2.2 High-throughput screening of increasing numbers of genetic associations........ 8
   2.3 Exome sequencing to characterize a severe mitochondrial disorder........ 9
3. Methods.. 10
4. Results.. 12
   4.1 Targeted application of human genetic variation can improve red blood cell production from stem cells........ 12
   4.2 Systematic functional dissection of human genetic variation affecting red blood cell traits........ 14
   4.3 Mutations in the substrate binding glycine-rich loop of the mitochondrial processing peptidase-α Protein (PMPCA) cause a severe mitochondrial disease........ 17
5. Discussion.. 18
   5.1 Utilizing human genetic variation studies to improve *in vitro* production of RBCs... 18
   5.2 MPRA to systematically screen for functional GWAS variants........ 19
   5.3 *PMPCA* mutations in a mitochondrial disorder........ 20
6. References.. 21

Eidesstattliche Versicherung.. 26

Declaration of Contribution.. 27

Publication 1.. 29

Main Article: Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells........ 30

Supplemental Information.. 37

Publication 2.. 61

Main Article: Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits........ 62

Supplemental Information.. 79

Publication 3.. 99

Main Article: Mutations in the Substrate Binding Glycine-rich Loop of the Mitochondrial Processing Peptidase-α Protein (PMPCA) Cause a Severe Mitochondrial Disease........ 100

Curriculum Vitae.. 111

Publication List.. 113

Acknowledgements.. 114

1. Abstract

Introduction

Fundamental advances in genotyping and gene sequencing technologies have led to a rapidly increasing identification of common and rare genetic variants associated with human traits and diseases. However, molecular mechanisms underlying these associations often remain elusive. In this thesis, three functional studies are summarized that were performed to follow-up on such association signals:

1) Common and rare genetic variants in the *SH2B3* gene are associated with elevated red blood cell (RBC) counts. Thus, it was investigated whether perturbation of *SH2B3* can improve the *in vitro* production of RBCs, an alternative approach to obtain RBCs for transfusion.
2) A high-throughput reporter assay of common genetic variants associated with RBC traits in a recent genome-wide association study (GWAS) was used to identify novel casual variants and to investigate their mechanisms of action.
3) Exome sequencing and follow-up studies were performed to identify rare genetic variants in a family affected by an enigmatic mitochondrial disorder.

Methods

Patient-derived cells, cDNAs, shRNAs and genome editing approaches were utilized to modulate the expression of candidate genes. Primary cells and cell lines were cultured and analyzed using *in vitro* and *in silico* approaches including flow cytometry, Western blotting and microscopic imaging. To screen common genetic variants for regulatory function a massively parallel reporter assay (MPRA) was utilized.

Results

1) Perturbation of *SH2B3* resulted in markedly increased yields of *in vitro* produced RBCs without compromising maturation or quality of the obtained RBCs.
2) MPRA identified endogenous regulatory elements and 32 MPRA functional variants of which 10-16 are casually related to the GWAS phenotype. For a subset of these variants target genes were identified and one variant was mechanistically linked back to the phenotype.
3) Exome sequencing identified *PMPCA* mutations in patients affected by a rare mitochondrial disorder which are suggested to result in impaired processing of a mitochondrial precursor protein.

Conclusion

1) Perturbation of *SH2B3* can augment the *in vitro* production of RBCs, pointing to the potential to use insight from genetic variation to improve cell and tissue replacement therapies.
2) We demonstrate that MPRAs can be an elegant tool to narrow down GWAS-nominated common variants to a reduced set of leads for further investigation.
3) Exome sequencing and follow-up studies represent a promising strategy to identify the molecular etiology of previously enigmatic genetic disorders.

1.1 Abstract Deutsch

Einleitung

Moderne DNA-Sequenzierungsmethoden haben zur Entdeckung einer stetig wachsenden Anzahl häufiger und seltener genetischer Varianten geführt, die mit bestimmten humanen Merkmalen oder Krankheiten assoziiert sind. Die zugrunde liegenden molekularen Mechanismen sind jedoch in den meisten Fällen unklar, da zu deren Verständnis weiterführende funktionelle Untersuchungen notwendig sind. In der vorliegenden Arbeit werden drei solcher *follow-up* Studien vorgestellt:

1) Varianten des humanen *SH2B3* Gens sind mit erhöhten Erythrozytenzahlen assoziiert. Daher wurde untersucht, ob eine Inaktivierung von *SH2B3* die *in vitro* Produktion von reifen roten Blutzellen (RBZs) verbessern kann, einem alternativen Verfahren zur Herstellung von Erythrozytenkonzentraten.
2) Es wurde eine Multiplex-Untersuchung häufiger genetischer Varianten durchgeführt, die in einer genomweiten Assoziationsstudie (GWAS) mit Erythrozyten-Merkmalen assoziiert sind, um kausale Varianten und zugrunde liegende Mechanismen zu identifizieren.
3) Exom-Sequenzierungen und zellbiologische *follow-up* Untersuchungen wurden durchgeführt, um die Ätiologie einer mitochondrialen Erkrankung zu ergründen.

Methodik

Von Patienten isolierte Zellen, shRNA-vermittelter *Knockdown* und *Genome Editing* wurden angewendet, um die Expression von Kandidatengenen in primären Zellen und Zelllinien zu modulieren und den resultierenden Phänotyp untersuchen zu können. Mehrere bioinformatische und zellbiologische Methoden wurden verwendet, darunter Durchflusszytometrie, Western Blots und Lichtmikroskopie. Um gleichzeitig eine Vielzahl von genetischen Varianten einer GWAS zu

untersuchen, wurde ein kürzlich entwickelter *Massively Parallel Reporter Assay* (MPRA) angewendet.

Ergebnisse

1) Ein *Knockdown* des *SH2B3* Gens führte zu einer deutlich ansteigenden Anzahl *in vitro* produzierter RBZs, ohne die Differenzierung oder Qualität der Zellen zu beeinträchtigen.
2) Mittels MPRA konnten endogen regulatorische Elemente und so 32 *MPRA Functional Variants* identifiziert werden, von diesen stehen ca. 10-16 in kausalem Zusammenhang mit dem Phänotyp. Für einen Teil dieser Varianten werden zugrundeliegende Mechanismen und Zielgene beschrieben und es wird ein funktioneller Zusammenhang einer Variante mit dem ursprünglichen GWAS-Phänotyp aufgezeigt.
3) Mittels Exom-Sequenzierung wurden Mutationen des *PMPCA* Gens entdeckt, die mit einer schweren mitochondrialen Erkrankung assoziiert sind und zur gestörten Prozessierung eines mitochondrialen Precursor-Proteins führen.

Schlussfolgerung

1) Eine genetische Inaktivierung von *SH2B3* erhöhte die Ausbeute *in vitro* produzierter RBZs. Dies zeigt das Potential auf, genetische Variationsstudien gezielt zur Verbesserung von Zell- und Gewebeersatztherapien zu nutzen.
2) Mittels MPRA können GWAS-Befunde deutlich eingegrenzt werden, was eine gezieltere funktionelle Untersuchung einzelner Varianten ermöglicht.
3) Exom-Sequenzierung und gezielte *follow-up* Untersuchungen erlauben die molekulare Ätiologie zuvor unverstandener genetischer Erkrankungen zu charakterisieren.

2. Introduction

At the turn of the millennium, the human genome project climaxed in the publication of an initial sequence of the human genome – a genome 25 times as large as any genome sequenced before and eight times as large as the sum of these genomes [1]. The paper attracted much attention from both science and the popular media [1,2]. Since its publication, numerous studies followed to better understand the vast complexity of the human genome. For example, the HapMap Consortium and the 1000 Genomes Project utilized single nucleotide polymorphism (SNP) genotyping and whole genome sequencing of (many more than 1000) individuals from different ancestry in order to assess the variation in the human genome and patterns of inheritance [3-5].

Genome-wide association studies (GWAS) used these resources to find variants that are associated with a certain human trait or disease [6]. Today, the ever-growing GWAS catalog contains more than 3800 publications and lists more than 125000 SNP-trait associations (https://www.ebi.ac.uk/gwas, March 2019). However, follow-up studies on GWAS are complex and the molecular mechanisms underlying these associations have only been identified in a few cases [7-11]. A major reason is that most GWAS variants only have a moderate effect size which is likely due to the fact that the variant otherwise might have been selected against over the course of evolution [2,12]. However, while these variants may only result in subtle differences in gene expression patterns, experimental perturbation of these genes may result in a much more drastic phenotype and drugging of underlying pathways could be highly beneficial for patients [10]. For example, the *HMGCR* locus has a variant with modest effect size, however, it encodes the target of statins. These drugs are taken by millions in order to reduce their cholesterol levels and myocardial infarction risk [2].

Advances in genetics, such as the 1000 Genomes Project, were particularly propelled by the introduction of novel sequencing approaches [2]. The human genome project still essentially utilized the technique introduced by Frederick Sanger in the late 1970s. However, in the upcoming decade modern next generation sequencing approaches became about 5000-fold faster and up to 100000-fold cheaper [1,2]. Indeed, whole genome sequencing or the targeted sequencing of the protein coding regions (exome sequencing) resulted in the increased identification of previously undiscovered mutations in Mendelian disorders. The approach is now also frequently used for diagnostic purposes in the clinical setting [13,14]. About a century after the rediscovery of Mendel's laws in the early 20th century, the Online Mendelian

Inheritance in Man (OMIM) database currently contains more than 24000 entries and lists over 5000 phenotypes with a known molecular basis (www.omim.org, March 2019).

In this thesis, three follow-up studies of genetic findings from GWAS and exome sequencing studies are summarized: First, we explored whether insight from genetic variation can be used to improve the *in vitro* production of red blood cells, an alternative approach to obtain red blood cells for transfusion [15]. Second, we performed a high-throughput screening of GWAS loci to identify novel causal variants and to investigate potential mechanisms of action. Third, we performed exome sequencing in a family affected by an enigmatic mitochondrial disorder to characterize its etiology.

2.1 Using human genetic variation to improve cell and tissue replacement therapies.

Stems cells are a potential source for cell and tissue production in regenerative medicine [16]. For instance, *in vitro* differentiated red blood cells (RBCs) derived from hematopoietic stem and progenitor cells represent an alternative to donor-derived RBCs for transfusion [15]. *In vitro* produced, mature RBCs can be generated from various stem cell sources and have already been successfully transfused to humans in clinical trials [17]. However, the yield of fully mature red cells remains low, making the process inefficient and expensive [18]. While numerous studies focus on improving RBC production by optimizing culture conditions [18,19], we decided to explore the potential of a complementary approach: the modulation of genes that regulate RBC production.

A recent GWAS has identified a common coding SNP (rs 3184504) in the *SH2B3* gene, which is thought to be a hypomorphic allele and which is associated with increased RBC counts and hemoglobin levels in humans [6]. Consistently, rare loss-of-function mutations in *SH2B3* are associated with considerably higher elevations of hemoglobin levels and erythrocytosis [20,21]. *SH2B3* encodes a negative regulator of cytokine signaling in hematopoietic cells. However, *Sh2b3* deficient mice have normal RBC counts and hemoglobin levels, suggesting cross-species differences and human-specific effects [22,23].

Major aims of this project were

1. to knockdown *SH2B3* expression in $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) derived from mobilized adult peripheral blood and cord blood, and to assess the potential increase in cell expansion,

2. to investigate the effects on differentiation upon *SH2B3* knockdown,
3. to determine the quality of the *in vitro* produced RBCs,
4. to assess to what extent these results can be reproduced in erythroid cells derived from human embryonic stem cells (hESCs).

2.2 High-throughput screening of increasing numbers of genetic associations.

In contrast to the aforementioned SNP in the *SH2B3* gene, 85% – 90% of GWAS hits tag non-coding variants [24] and current hypotheses suggest that many of these variants alter transcription factor binding and gene transcription [25,26]. Thus, identifying target genes and underlying biochemical pathways is particularly challenging. Furthermore, each GWAS hit usually represents one of hundreds of variants that are inherited collectively across a large genomic region called linkage disequilibrium (LD) block [27]. Consequently, a GWAS-nominated variant is likely just in LD with the actual causal variant, without having a direct biological effect of its own [28]. Also, keeping in mind the small effect size of most variants, moving forward from a statistical association and to establish the potential causality of a variant is complex [9]: Firstly, putative functional variants (PFVs) need to be identified, which usually requires genetic fine-mapping. Secondly, target genes need to be identified, which may be facilitated by expression quantitative trait loci (eQTL) studies to search for SNPs that affect the expression of a particular gene or by creating isogenic cellular models by genome editing. Once a target gene is found, it can be modulated in order to recreate the original phenotype and to elucidate underlying mechanisms of action.

However, given the exponentially increasing number of genetic associations, there is need for the development of high-throughput functional screenings [29,30]. To this end, we modified a recently developed massively parallel reporter assay (MPRA) to simultaneously screen a total of 2756 variants in high LD with 75 hits of a recent comprehensive GWAS of RBC traits [6,31]. In brief, major and minor alleles of each variant were placed into oligonucleotide constructs in linkage with unique barcodes. A plasmid library containing these constructs was then introduced into the erythro-leukemia cell line K562 and the activity of each individual allele was subsequently determined by sequencing and quantification of barcode abundance patterns.

Major aims of the project were

1. to find an appropriate cellular model for the MPRA screen,

2. to identify MPRA functional variants (MFVs) and estimate whether this set of MFVs is enriched for functional regulatory elements,
3. to investigate a subset of MFVs in greater detail by creating isogenic deletions by genome editing and to assess potential target genes,
4. to link MFVs mechanistically back to the original GWAS phenotype.

2.3 Exome sequencing to characterize a severe mitochondrial disorder.

In contrast to common genetic variation, rare genetic variants identified in Mendelian genetic disorders usually have larger effect sizes and may cause severe diseases. In Mendelian disorders, recent advances in sequencing technologies, such as exome sequencing, have allowed identification of previously unrecognized underlying mutations and to provide a more complete picture of the disease pathophysiology [13,32]. Here, a rare Mendelian disorder was investigated that affects the mitochondrion.

In the mitochondrion, a high number of precursor proteins are processed by mitochondrial processing peptidase (MPP) [33]. These precursors are encoded by nuclear DNA, get synthesized in the cytosol and are imported into the mitochondrion. MPP subsequently cleaves off amino-terminal targeting peptides (TPs) that signal the mitochondrial import. Its alpha subunit, αMPP, contains a glycine-rich loop, which is considered a key structural element that moves the substrate to the active site of the enzyme. αMPP is encoded by the *PMPCA* gene.

The starting point of this study was a large Lebanese family with two family members affected by an enigmatic disorder. These first cousins exhibited clinical findings that suggested an underlying mitochondrial genetic disorder, as outlined below. Since initial genetic tests failed to identify its molecular etiology, major aims of the project were

1. to perform exome sequencing in affected and non-affected family members to identify potential causal mutations,
2. to identify mechanisms of action of candidate mutations,
3. to assess a potential rescue in patient-derived cells by modulation of the affected genes.

3. Methods

Cell lines

K562 cells (mainly used for the MPRA screen in publication 2) were cultivated in RPMI with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S). TF-1 cells were cultured in RPMI supplemented with 10% FBS, 2 ng ml^{-1} GM-CSF and 1% P/S. 293T cells were cultured in DMEM with 10% FBS and 1% P/S. Cell culture was always performed at 37°C and 5% CO_2.

Primary erythroid cell culture

$CD34^+$ stem and progenitor cells (as used to study *in vitro* production of RBCs in publication 1) from GSCF-mobilized peripheral blood, bone marrow or cord blood were used. $CD34^+$ cells from cord blood were purified using a Microbead Kit. Cells underwent erythroid differentiation in a three-phase culture system as previously described [34], except where indicated an optimized EPO concentration of 1 IU was used. Alternatively, a 5-day expansion phase was added prior to initiating differentiation as described previously [35].

Lentiviral vector production and transduction

293T cells were transfected with the indicated constructs along with the packaging and genomic vectors VSV-G and pDelta8.9 using FuGENE® 6 reagent, according to the manufacturer's instructions. The next day, medium was changed to the appropriate medium for subsequent infection, in case of primary cell culture medium, cytokines were omitted and added later. After 24 h viral supernatant was collected and filtered at 0.45 µm. Cells were infected with lentiviral supernatant in 6-well plates in the presence of 8 µg ml^{-1} polybrene. For shRNA-mediated knockdown of *SH2B3* and *RBM38,* cells were infected on day 1 or 2 of differentiation, respectively. Cells were spun at 2000 r.p.m. for 90 min at room temperature. Media was replaced the next morning and cells were selected by adding 1 µg ml^{-1} puromycin 12 – 48 h after infection. For GATA1 and PMPCA (for rescue experiments in patient fibroblasts in publication 3) overexpression, cells were spinfected at 2500 r.p.m.. Infected K562 cells were identified by assessment of GFP expression by flow cytometry.

Generation of isogenic clonal deletions

K562 cells were transfected with Cas9 nuclease and sgRNA encoding plasmids (Table S7 in publication 2) in a ratio of 1:3 utilizing Lipofectamine® LTX Plus. Controls were transfected with Cas9 nuclease and pLKO.1-GFP in a 1:3 ratio. Puromycin-selection was performed 24 h

post-transfection for 48 h and isogenic clones were obtained by limiting dilution. PCR was performed for screening of potential clones (primer pairs flanking the guide sequence are depicted in Table S8 in publication 2), followed by Sanger sequencing of positive clones to map the deletion.

Flow cytometry

Cells were harvested, washed and stained with indicated antibodies. Propidium iodide served as a dead cell marker unless an Annexin V apoptosis assay was performed. Cells were analyzed on a BD Canto II, LSR II or LSR Fortessa. Eosin-5-maleimide (EMA) staining to assess membrane defects was performed as previously described [36]. To assess cell divisions cells were labeled using PKH26GL, cell divisions were estimated as previously described [37].

RNA extraction

RNA extraction was performed using an RNEasy® Plus Mini Kit. RNA for microarrays was extracted 48 h after infection to overexpress GATA1 in K562 cells, on day 7 of primary erythroid cell culture in *SH2B3* knockdown experiments and on day 16 of primary erythroid cell culture for RNA sequencing. For microarrays or for the assessment of target genes in selected clonal deletions of MPRA functional variants, cDNA was synthesized using the iScript™ cDNA Synthesis Kit.

May-Giemsa staining

Cells were harvested and spun on poly-L-lysine coated microscope slides and stained with May-Grünwald solution and after intermittent washing with Giemsa solution.

Western Blotting

Cells were harvested and processed as previously described [32], except that the incubation was performed at 95°C for 5 min. After SDS-PAGE using precast gels and 1% Tris-Glycine-SDS running buffer and protein transfer onto PVDF or nitrocellulose membranes, staining with the indicated primary and secondary antibodies was performed as described previously [32]. To assess SH2B3 protein levels, cells were usually harvested at day 9 of culture.

Statistical analysis

Pairwise comparisons were performed using an unpaired two-tailed Student's t-test (see also next paragraph regarding statistical analysis of the MPRA screen).

Design and analysis of an MPRA to screen GWAS variants

A total of 2756 SNPs in strong LD with 75 sentinel SNPs of a recent GWAS [6] were identified from the CEU population of the 1000 genomes project. We modified a recently developed MPRA [31]. In brief, nucleotide constructs were designed by placing major and minor alleles of each of the 2756 variants across three sliding windows into the constructs so that 1/3, 1/2 or 2/3 of the total construct length was 5’ of the variant. An oligonucleotide library containing these constructs with 14 unique barcodes each, separated by Kpnl-Xbal sites and flanked by constant primer sites, was synthesized on an Array. After PCR amplification and ligation into a plasmid backbone with a minP-luc2 insert, the oligonucleotide library was introduced into K562 or K562+GATA1 cells by nucleofection. After 48 h, total RNA was harvested, barcodes were isolated by RT-PCR, sequenced and finally counted. In order to determine the relative abundance of the constructs in the original oligonucleotide library, DNA was isolated from the library and sequenced. Consequently, activity of a tagged construct was indicated by the ratio of DNA and RNA counts. Active constructs were defined as constructs showing a significantly higher activity then the activity distribution of all constructs by a one-sided Mann-Whitney-U test. MPRA functional variants were determined by comparing the activity of constructs containing the major allele of a variant with the ones containing the minor allele by a two-sided Mann-Whitney-U test.

Additional methods

An in-depth description of all utilized methods is beyond the scope of this brief summary. A more detailed description and additional methods, primarily regarding bioinformatic approaches and experiments performed in collaborating laboratories, can be found in the attached publications and supplementary sections.

4. Results

4.1 Targeted application of human genetic variation can improve red blood cell production from stem cells.

We aimed to investigate the potential of perturbing *SH2B3* in order to improve the *in vitro* production of RBCs. Common genetic variants of *SH2B3* are associated with elevated red blood cell counts and rare loss-of-function-variants are associated with erythrocytosis [6,20,21]. To follow-up on these observations, we analyzed a group of 4678 individuals subjected to exome sequencing and identified rare putative damaging or loss-of-function variants of *SH2B3*

associated with elevated hemoglobin and hematocrit (publication 1 (P1), Figure 1A,B). In order to examine the effects of *SH2B3* suppression on RBC production, we targeted *SH2B3* with two independent shRNAs (sh82 and sh83) in adult human $CD34^+$ stem and progenitor cells that subsequently underwent erythroid differentiation to produce mature RBCs (P1, Figure 1C,D).

Although suppression of *SH2B3* resulted in essentially similar differentiation, an improved erythroid maturation was observed upon *SH2B3* knockdown, as assessed by cell surface phenotyping (earlier acquisition of CD235a; P1, Figure 1E and earlier loss of CD71; P1, Figure 1F) and cell morphology (P1, Figure 1H). Furthermore, we observed an up to 2-fold increase of enucleated RBCs ($CD235a^+$ and $Hoechst^-$; P1, Figure 1G) in *SH2B3* knockdown cultures. Additionally, microarray analysis of *SH2B3* knockdown and control cells showed essentially similar global gene expression profiles (P1, Figure 1I). However, knockdown cells displayed a gene signature reminiscent of more differentiated cells (P1, Figure 1J).

As accelerated differentiation may result in restrained proliferative capacity, we assessed cell expansion and yield of *in vitro* produced RBCs. We observed that *SH2B3* knockdown in adult HSPCs resulted in a markedly increased cell expansion (P1, Figure S1A). This resulted in an improvement of the overall yield of enucleated RBCs by a factor of 3-5 (P1, Figure 2A). This improvement could be observed regardless of the stem cell source. Indeed, *SH2B3* suppression in $CD34^+$ HSPCs from cord blood also resulted in a 2-4-fold increased yield of mature RBCs (P1, Figure 2B). Furthermore we assessed cell expansion using an alternative culture system that included a progenitor expansion step. Here, we observed an additional increase in cell expansion and yields of mature cells could be improved by up to 7-fold (P1, Figure 2C). Importantly, the observed effects were not due to suboptimal cytokine concentrations, as concentrations were utilized where cells demonstrated maximal expansion (P1, Figure S1B) [38].

We next investigated the underlying mechanisms by which knockdown of *SH2B3* enhanced cell expansion. We first assessed a potentially improved cell survival, but did not notice significant differences in the fraction of apoptotic cells between *SH2B3* knockdown cells and controls (P1, Figure S1C). Then, we assessed cell divisions upon *SH2B3* knockdown and observed 0.5-1 additional cell divisions over a 4 day period during early erythroid differentiation (P1, Figure S1D,E). SH2B3 has been described as a negative regulator of EPO and KIT signaling. Accordingly, we observed enhanced phosphorylation of STAT5 and KIT receptor and enhanced expression of early EPO responsive genes in *SH2B3* knockdown cells (P1, Figure S1F-H).

Since suppression of *SH2B3* resulted in enhanced maturation and expansion of RBCs, we investigated if *SH2B3* knockdown affected the quality of mature RBCs. Cell size and hemoglobin content upon *SH2B3* knockdown were similar to controls and within normal range. Furthermore, we observed normal hemoglobin subunits present in *SH2B3* knockdown cells and the expression of cell surface antigens was similar to controls (P1, Figure S2A,C,D). We also assessed pyruvate kinase activity in *SH2B3* knockdown and control cells, without noticing substantial differences (P1, Figure S2B). Finally, we observed similar Eosin-5-maleimide (EMA) staining, comparing *SH2B3* knockdown and control cells (P1, Figure 2D). This suggested physiologic cell membrane integrity as disorders of the RBC membrane, including hereditary spherocytosis, are clinically diagnosed by altered EMA staining.

Ultimately, we investigated the effects of ablating *SH2B3* in human embryonic stem (hES) cells. Isogenic lines with intact *SH2B3* and CRISPR Cas9-mediated deletions within *SH2B3* were differentiated *in vitro* into hematopoietic progenitor cells and finally into erythroid cells (P1, Figure 2E,S2E,F). In line with our previous observations, loss of *SH2B3* resulted in a markedly increased cell expansion by up to 3-fold (P1, Figure 2F,G,S2G), without impairing erythroid differentiation as assessed by flow cytometry (P1, Figure 2I), cell morphology (P1, Figure 2H) and globin gene expression (P1, Figure S2H).

4.2 Systematic functional dissection of common genetic variation affecting red blood cell traits.

Reporter assays display cell-type-specific activity and therefore identifying an appropriate cellular model to perform an MPRA screen is crucial [11,37]. We identified three stages of glycophorin A positive human erythroid progenitor (HEP) cells, in which transcription is likely affected by GWAS variants (P2, Figure 2A,B). However, we decided not to perform our MPRA screen in primary erythroid cells, because these show heterogeneity in differentiation and cell numbers can be limiting [34]. Instead, we utilized K562 cells that have previously been successfully used in reporter assays [37] and showed gene expression profiles similar to HEPs (Figure 2E). They shared similar open chromatin features (P2, Figure 2C,D) and displayed similar occupancy by erythroid transcription factors (P2, Figure 2F). Our analyses suggested that common genetic variation associated with RBC traits likely acts in earlier- and later-stage HEPs. Thus, we performed the MPRA in both native K562 cells resembling earlier-stage and K562 cells overexpressing GATA1 (K562+GATA1) resembling later-stage HEPs, as these cells display a more differentiated erythroid gene signature (P2, Figure 2H,I).

At first, we set out to confirm that our assay (see P2, Figure 1A) can identify active regulatory elements and assessed the activity of included control elements that disrupt the binding site of the key erythroid transcription factor GATA1. Indeed, non-mutated, reference (Ref) constructs displayed an expected strong enhancer-like activity, whereas GATA1 motif disruption control mutants (Mut) showed a markedly decreased activity at a false discovery rate (FDR) of 1% (P2, Figure 3A). Next, we assessed all active constructs (ACs), which we defined as constructs showing an activity significantly greater than the activity distribution of all investigated constructs (FDR < 1%). These ACs (< 4% of all tested constructs) showed an activity distribution similar to Ref controls, while activity of inactive constructs was essentially similar to Mut controls (P2, Figure 3B). Additional findings suggested that differences in MPRA activity are likely due to differences in activity of cell-type-specific transcription factors. For instance, ACs were significantly enriched for open chromatin in erythroid cell types (P2, Figure 3E) and for chromatin occupancy of erythroid transcription factors (GATA, TAL1) and supervised learning models suggested known erythroid transcription factor-binding motifs (GATA1, TAL1, ETS/FLI1, AP-1/NFE2) as most predictive of high activity (P2, Figure 3D).

Our next goal was to identify variants displaying differential activity between the major and minor alleles and we found a total of 32 variants with allelic variation in activity, in the following termed MPRA functional variants (MFVs) (FDR < 1%). These 32 MFVs represented 31% (23/75) of the GWAS hits (P2, Figure 3G). Like ACs, MFVs were most enriched for open chromatin in HEPs and K562 cells and showed a strong enrichment for erythroid transcription factor occupancy (again including GATA1 and its co-factor TAL1) (P2, Figure 3H). Importantly, the activity of certain MFVs also showed dependency upon GATA1 protein levels when comparing the activity of each construct in K562 cells with the activity in K562+GATA1 cells (P2, Figure 3J,S3G). We applied a number of algorithms suggesting that our 32 MFVs are indeed enriched for functional regulatory variants and algorithms, such as the PICS algorithm for genetic fine mapping, and determined that our 32 MFVs are in fact enriched for PCVs. We also estimated an empirical positive predictive value (PPV) for our assay of 32% – 50% (see also supplementary methods in P2), suggesting that 10-16 MFVs represent causal trait-associated variants.

We examined a subset of these candidates in greater detail and identified several MFVs (rs3785098, rs9901219 and rs 7123861) that fall within known regulatory regions and partially overlap with transcription factor occupancy sites (P2, Figure S4A-C). We then specifically

focused on defining variants in erythroid enhancer elements bound by the key erythroid transcription factor GATA1. We identified three MFVs within such elements bound by GATA1 and multiple co-factors (rs737092, rs4490057, rs1175550) (P2, Figure 4A-C). These do not directly alter the core GATA1 binding motif, but were located in close proximity (1-3 bp) of (partially) conserved GATA1-binding motifs. In contrast, we identified rs1546723 in an erythroid enhancer that disrupts the TAL1 E-box of a GATA1/TAL1-binding motif (P2, Figure 4D).

In order to identify regulated target gene(s), we used CRISPR/Cas9 genome editing to create isogenic deletions across three MFVs (rs737092, rs1175550 and rs1546723) and assessed the expression of genes within approximately 1 megabase of distance. Consistent with previous GWAS follow-up studies, we observed that only one or two genes per MFV were fundamentally affected, most compellingly *RBM38* by rs737092, *SWIM1* by rs1175550 and *CD164* by rs1546723 (P2, Figure 5A-C).

Due to the overlap between these MFVs and a GATA1 complex, and since many MFVs displayed an increase in activity upon GATA1 overexpression, we suspected altered binding of erythroid transcription factors as an underlying mechanism of action (P2, Figure 4E-H). Indeed, FAIRE-seq and ChIP-seq data in erythroblasts suggested altered binding of TAL1 and LCB1 for rs1546723, that disrupted the TAL1 E-box (P2, Figure 6A) and altered GATA1 and TAL1 binding for rs4490057, which affected an extended GATA1 motif (P2, Figure 6B). However, the majority of PCVs do not alter any known transcription factor binding sites. Still, variants like rs1175550 and rs737092 may alter transcription factor binding, as further analyses suggested that these variants resulted in marked changes of DNA-shape adjacent to GATA1/TAL1 binding motifs (P2, Figure 6D-F).

Finally, we functionally studied the *RBM38* gene that was most substantially affected by the deletion of rs737092 in order to investigate if we can link this gene back to the associated trait. Upon knockdown of *RBM38*, we observed a delayed differentiation of primary erythroid cells that supposedly results in altered RBC size and counts, as observed in the GWAS (P2, Figure 7A,B). Additional experiments suggested that RBM38 plays a role in the regulation of the alternative splicing program in human erythropoiesis, as *RBM38* knockdown cells showed differentially spliced exons that are required for the synthesis of membrane skeletal proteins

critical for the stability of RBCs (P2, Figure 7C). Thus, we hypothesize that the phenotype observed in GWAS is due to subtle differences in RBC membrane integrity.

4.3 Mutations in the substrate binding glycine-rich loop of the mitochondrial processing peptidase-α protein (PMPCA) cause a severe mitochondrial disease.

We investigated a consanguineous Lebanese family with two members affected by a rare genetic disorder (see P3, Figure 1A for the pedigree). A 6 year old female initially presented with developmental delay and failure to thrive and later showed generalized hypotonia and muscle weakness, lactic acidemia, elevated transaminases and progressive brain atrophy (P3, Figure 1B-G). She was blind by the age of 6 due to severe optic atrophy. Her first cousin developed similar symptoms, but was more severely affected. Muscle biopsy findings and electron microscopy analysis showed abnormal mitochondria morphology (P3, Figure 1H-N), suggesting an underlying mitochondrial disorder. Initial tests included enzyme testing, mitochondrial genome sequence analysis and comprehensive sequencing analysis of mitochondrial nuclear genes, but these remained unrevealing.

In order to identify the molecular etiology of the disorder, we performed exome sequencing on the affected female proband and both parents (coverage depicted in P3, Table S1). We identified two variants in the *PMPCA* gene in the proband with appropriate parental segregation. Sanger sequencing confirmed the female proband to be compound heterozygous for two *PMPCA* missense mutations (Chr 9: 139313082; NM_015160, c.1066G>A; p.G356S and Chr 9: 139313299; NM_01560, c.1129G>A; p.A377T) (P3, Figure 2A, Table S1). Both mutations were also identified in her affected cousin, while the investigated unaffected family members were heterozygous for a single *PMPCA* variant or wildtype.

The affected amino acid residues are well conserved across evolution (P3, Figure 2B) and the mutations were not present on any public database, including the 1000 Genomes catalog. *In silico* variant prediction models suggested that the *PMPCA* variants are likely deleterious and localization of the affected residues suggested that they may impair the αMPP glycine-rich catalytic loop. Both residues were either located within or in very close proximity (20 bp downstream) to this catalytic loop and likely impair its dynamic movements and thus catalytic activity (P3, Figure 2C-E).

In order to characterize the functional effects of the mutations, immunofluorescence microscopy of proband fibroblasts and two age-matched controls was performed. Proband cells showed a reduction in PMPCA staining and showed abnormally enlarged, swollen mitochondria compared to the controls (P3, Figure 3A-F). We next assessed potentially affected processing of mitochondrial precursor proteins due to the observed *PMPCA* mutations. Frataxin is extensively processed by MPP and altered interactions of Frataxin with MPP have been described in the context of human diseases [39]. Indeed, Western blot analysis showed a reduction of mature (processed) Frataxin and an increase of immature (unprocessed) Frataxin levels in primary patient fibroblasts compared to controls (P3, Figure 3G). We performed rescue experiments in order to investigate whether the abnormal Frataxin processing observed is due to a loss-of-function of PMPCA. Indeed, overexpression of wildtype PMPCA in patient cells resulted in a robust increase of mature Frataxin and a decrease of immature Frataxin levels (P3, Figure 3H).

5. Discussion

5.1 Utilizing human genetic variation studies to improve *in vitro* production of RBCs.

We used insight from human genetic variation studies to improve the *in vitro* production of RBCs from stem cells by modulation of SH2B3 protein levels. Our approach increased the overall yield of *in vitro* produced RBCs by up to 7-fold without impairing erythroid differentiation and the quality of the mature cells. A great number of current efforts to augment this process focus on optimizing cell culture conditions [18,19], however, optimizing cytokine concentrations could be limited due to potential saturation effects. Our study therefore highlights a complementary approach, as additional intrinsic genetic modifications of erythroid cells can provide a further improvement of RBC production.

Modulation of *SH2B3* enhanced cell expansion regardless of the stem cells source that included $CD34^+$ stem and progenitor cells derived from adult peripheral blood and cord blood, as well as hES cells. Importantly, hES cells are capable of self-renewal, which could could be extremely valuable, because repeated perturbation of *SH2B3* with shRNAs will likely not be feasible in a clinical setting.

In vitro production of RBCs on an economically feasible scale will require substantial optimization of current culturing techniques [18]. Consequently, a limitation to keep in mind is the fact that today we are not able to demonstrate a benefit of *SH2B3* perturbation in such further

refined culture systems. However, comparing both investigated culture conditions, we noticed a slightly improved effect of *SH2B3* knockdown utilizing the system that included an expansion phase. Thus, effect sizes might be even higher in these more optimized culture systems.

Generally, genetic perturbations have the risk of mutagenesis and perturbation of *SH2B3* may either directly or by insertional mutagenesis increase the risk of leukemogenesis. However, as mature RBCs enucleate and thus lose their genetic material [15], these risks are minimal.

There are recurrent shortages of donor-derived transfusion products, especially for rare blood groups [18]. *In vitro* produced RBCs therefore hold the potential to meet these challenges and to improve the blood supply. Furthermore, *in vitro* production of engineered RBCs may allow for the design of unique RBCs for specific patient needs. For instance, it has been proposed to produce patient-derived RBCs after correcting the point mutation in sickle cell anemia by genome editing [40]. Such elegant approaches could help to overcome the fundamental issue of allo-antibody production against donor-derived antigens in chronically transfused patients [41]. Manipulation of SH2B3 holds the promise to increase yields of such *in vitro* produced RBCs by up to 7-fold. We envision future genetic studies on RBCs traits to reveal additional potential target genes and thus to provide additional means to manage the blood supply.

5.2 MPRA to systematically screen for functional GWAS variants.

We employed a high-throughput MPRA to perform a simultaneous functional screening of 2756 variants in strong LD with 75 sentinel variants associated with RBC traits identified in a recent GWAS [6]. Our assay was capable of identifying elements with endogenous erythroid regulatory activity and we conservatively identified 32 MFVs across 23 sentinel SNPs. We predict that between 10 and 16 of these MFVs are causally related to the original GWAS phenotype. We also suggest mechanisms of action for a subset of MFVs, identify three high-confidence target genes and ultimately link one variant back to the original phenotype.

In line with previous work, a key finding of our study is that multiple functional GWAS variants appear to alter the activity of the master erythroid transcription factor GATA1 and its co-factors [11,26,30,42]. Most of these variants did not disrupt, but were in close proximity of GATA1 binding sites. GATA1 displays a key transcription factor orchestrating signaling and differentiation of human erythroid cells and mutations or rare variants that impact GATA1, as observed in Mendelian disorders, result in a severe phenotype as erythropoiesis is markedly

perturbed [32,43]. We therefore hypothesize that common variants only result in subtle changes of master transcription factor binding, consistent with their small effect sizes that would be expected to have negligible selecting effects during the course of evolution.

Several limitations of our assay should be considered. Firstly, we estimate that our assay has a PPV of 32% – 50% and a sensitivity of 14% – 22%, meaning that at best only half of the variants identified will be causal, while a great number of casual variants will be missed by our approach. Secondly, our assay will likely miss variants that only have a minimal independent effect, but may act jointly with other variants. Thirdly, an MPRA screen, as performed here, can only be employed if the primary cell type is known, in which common genetic variants associated with a certain trait are likely to act. Concluding, MPRAs can only narrow down GWAS-nominated variants to a reduced set of leads that subsequently require further investigation.

Anemia is a global health burden and often caused by defective erythropoiesis [44]. An improved understanding of how human genetic variation acts during this process can therefore allow for the discovery of novel therapeutic options [10]. We envision our approach to be applied for additional GWAS follow-up studies in order to accelerate our understanding of human physiology and pathophysiology.

5.3 *PMPCA* mutations in a mitochondrial disorder.

By combining exome sequencing with *in silico* and *in vitro* follow-up studies, we describe causal *PMPCA* mutations in a mitochondrial disorder that is associated with altered processing of Frataxin. Mutations of the Frataxin encoding gene result in Friedreich's ataxia and have previously been shown to alter the interaction of Frataxin with MPP [45-46]. Here, we suggest *PMPCA* mutations affecting the alpha subunit of MPP to result in impaired catalytic activity of the enzyme and thus in impaired processing of Frataxin. Concordantly, we demonstrate that ectopic expression of wildtype PMPCA rescues the abnormal Frataxin processing that we observed in patient cells.

PMPCA mutations have previously been described to cause abnormal mitochondrial protein processing in patients with cerebellar ataxia [47]. We observe a more severe phenotype that may be due to the mutations that perturb the glycine-rich catalytic loop. This loop is well conserved and its depletion has been shown to result in decreased enzyme activity [33,48]. During substrate translocation, the catalytic loop undergoes considerable conformational changes [33]. This

fundamental flexibility of the loop may be impaired by the identified mutations due to replacement of residues with bigger size and enhanced polarity.

The impaired processing of Frataxin and the reduced levels of its mature form may contribute to the phenotype we observed in patients with *PMPCA* mutations. We have investigated additional mitochondrial proteins, such as PINK1, that is potentially affected by *PMPCA* mutations [49] and members of the electron transport chain family. However, we did not observe abnormal processing in proband cells. Still, more than 1000 proteins are known to undergo mitochondrial TP processing, which justifies future investigation.

Previous work has shown that exome sequencing provides an elegant tool to identify thus far unrecognized mutations in Mendelian disorders. For instance, mutations in the erythroid transcription factor GATA1 have recently been identified in patients affected by Diamond Blackfan Anemia (DBA) [13]. The finding has caused some controversy, as the DBA phenotype has been previously linked to ribosome mutations, exclusively. However, follow-up studies suggested these ribosomal mutations to result in impaired expression of the full-length version of GATA1 and thus, provided a more complete picture of the pathophysiology in DBA [32]. Here, we provide an additional example of how exome sequencing allows for the identification of previously unknown rare mutations in Mendelian disorders, and elucidation of novel therapeutic means, such as the identification of candidate genes for gene therapy.

6. References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide

M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921.

2. Lander ES. Initial impact of the sequencing of the human genome. Nature. 2011 Feb 10;470(7333):187-97.
3. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature. 2010 Oct 28;467(7319):1061-73.
4. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, De Pristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56-65.
5. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. Nature. 2010 Sep 2;467(7311):52-8.
6. van der Harst P, Zhang W, Mateo Leach I, Rendon A, Verweij N, Sehmi J, Paul DS, Elling U, Allayee H, Li X, Radhakrishnan A, Tan ST, Voss K, Weichenberger CX, Albers CA, Al-Hussani A, Asselbergs FW, Ciullo M, Danjou F, Dina C, Esko T, Evans DM, Franke L, Gögele M, Hartiala J, Hersch M, Holm H, Hottenga JJ, Kanoni S, Kleber ME, Lagou V, Langenberg C, Lopez LM, Lyytikäinen LP, Melander O, Murgia F, Nolte IM, O'Reilly PF, Padmanabhan S, Parsa A, Pirastu N, Porcu E, Portas L, Prokopenko I, Ried JS, Shin SY, Tang CS, Teumer A, Traglia M, Ulivi S, Westra HJ, Yang J, Zhao JH, Anni F, Abdellaoui A, Attwood A, Balkau B, Bandinelli S, Bastardot F, Benyamin B, Boehm BO, Cookson WO, Das D, de Bakker PI, de Boer RA, de Geus EJ, de Moor MH, Dimitriou M, Domingues FS, Döring A, Engström G, Eyjolfsson GI, Ferrucci L, Fischer K, Galanello R, Garner SF, Genser B, Gibson QD, Girotto G, Gudbjartsson DF, Harris SE, Hartikainen AL, Hastie CE, Hedblad B, Illig T, Jolley J, Kähönen M, Kema IP, Kemp JP, Liang L, Lloyd-Jones H, Loos RJ, Meacham S, Medland SE, Meisinger C, Memari Y, Mihailov E, Miller K, Moffatt MF, Nauck M, Novatchkova M, Nutile T, Olafsson I, Onundarson PT, Parracciani D, Penninx BW, Perseu L, Piga A, Pistis G, Pouta A, Puc U, Raitakari O, Ring SM, Robino A, Ruggiero D, Ruokonen A, Saint-Pierre A, Sala C, Salumets A, Sambrook J, Schepers H, Schmidt CO, Silljé HH, Sladek R, Smit JH, Starr JM, Stephens J, Sulem P, Tanaka T, Thorsteinsdottir U, Tragante V, van Gilst WH, van Pelt LJ, van Veldhuisen DJ, Völker U, Whitfield JB, Willemsen G, Winkelmann BR, Wirnsberger G, Algra A, Cucca F, d'Adamo AP, Danesh J, Deary IJ, Dominiczak AF, Elliott P, Fortina P, Froguel P, Gasparini P, Greinacher A, Hazen SL, Jarvelin MR, Khaw KT, Lehtimäki T, Maerz W, Martin NG, Metspalu A, Mitchell BD, Montgomery GW, Moore C, Navis G, Pirastu M, Pramstaller PP, Ramirez-Solis R, Schadt E, Scott J, Shuldiner AR, Smith GD, Smith JG, Snieder H, Sorice R, Spector TD, Stefansson K, Stumvoll M, Tang WH, Toniolo D, Tönjes A, Visscher PM, Vollenweider P, Wareham NJ, Wolffenbuttel BH, Boomsma DI, Beckmann JS, Dedoussis GV, Deloukas P, Ferreira MA, Sanna S, Uda M, Hicks AA, Penninger JM, Gieger C, Kooner JS, Ouwehand WH, Soranzo N, Chambers JC. Seventy-five genetic loci influencing the human red blood cell. Nature. 2012 Dec 20;492(7429):369-75.
7. Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L, Sabo PJ, Vierstra J, Voit RA, Yuan GC, Porteus MH, Stamatoyannopoulos JA, Lettre G, Orkin SH. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science. 2013 Oct 11;342(6155):253-7.

8. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puviindran V, Abdennur NA, Liu J, Svensson PA, Hsu YH, Drucker DJ, Mellgren G, Hui CC, Hauner H, Kellis M. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med. 2015 Sep 3;373(10):895-907.
9. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet. 2013 Nov 7;93(5):779-97.
10. Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science. 2008 Dec 19;322(5909):1839-42.
11. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Koteliansky V, Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S, Rader DJ. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010 Aug 5;466(7307):714-9.
12. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747-53.
13. Sankaran VG, Ghazvinian R, Do R, Thiru P, Vergilio JA, Beggs AH, Sieff CA, Orkin SH, Nathan DG, Lander ES, Gazda HT. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. J Clin Invest. 2012 Jul;122(7):2439-43. doi: 10.1172/JCI63597.
14. Lewis R. Exome sequencing comes to the clinic. JAMA. 2015 Apr 7;313(13):1301-3.
15. Migliaccio AR, Whitsett C, Papayannopoulou T, Sadelain M. The potential of stem cells as an in vitro source of red blood cells for transfusion. Cell Stem Cell. 2012 Feb 3;10(2):115-9.
16. Fox IJ, Daley GQ, Goldman SA, Huard J, Kamp TJ, Trucco M. Stem cell therapy. Use of differentiated pluripotent stem cells as replacement therapy for treating disease. Science. 2014 Aug 22;345(6199):1247391.
17. Giarratana MC, Rouard H, Dumont A, Kiger L, Safeukui I, Le Pennec PY, François S, Trugnan G, Peyrard T, Marie T, Jolly S, Hebert N, Mazurier C, Mario N, Harmand L, Lapillonne H, Devaux JY, Douay L. Proof of principle for transfusion of in vitro-generated red blood cells. Blood. 2011 Nov 10;118(19):507-9.
18. Rousseau GF, Giarratana MC, Douay L. Large-scale production of red blood cells from stem cells: what are the technical challenges ahead? Biotechnol J. 2014 Jan;9(1):28-38.
19. Huang X, Shah S, Wang J, Ye Z, Dowey SN, Tsang KM, Mendelsohn LG2, Kato GJ2, Kickler TS3, Cheng L. Extensive ex vivo expansion of functional human erythroid precursors established from umbilical cord blood cells by defined factors. Mol Ther. 2014 Feb;22(2):451-463.
20. Lasho TL, Pardanani A, Tefferi A. LNK mutations in JAK2 mutation-negative erythrocytosis. N Engl J Med. 2010 Sep 16;363(12):1189-90.
21. Spolverini A, Pieri L, Guglielmelli P, Pancrazzi A, Fanelli T, Paoli C, Bosi A, Nichele I, Ruggeri M, Vannucchi AM. Infrequent occurrence of mutations in the PH domain of LNK in patients with JAK2 mutation-negative 'idiopathic' erythrocytosis. Haematologica. 2013 Sep;98(9):e101-2.
22. Bersenev A, Wu C, Balcerek J, Tong W. Lnk controls mouse hematopoietic stem cell self-renewal and quiescence through direct interactions with JAK2. J Clin Invest. 2008 Aug;118(8):2832-44.
23. Velazquez L, Cheng AM, Fleming HE, Furlonger C, Vesely S, Bernstein A, Paige CJ, Pawson T. Cytokine signaling and hematopoietic homeostasis are disrupted in Lnk-deficient mice. J Exp Med. 2002 Jun 17;195(12):1599-611.
24. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, Xu H, Zang C, Ripke S, Bulik-Sullivan B, Stahl E; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium, Kähler AK, Hultman CM, Purcell SM, McCarroll SA, Daly M, Pasaniuc B, Sullivan PF, Neale BM, Wray NR, Raychaudhuri S, Price AL; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am J Hum Genet. 2014 Nov 6;95(5):535-52.
25. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal

MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317-30.

26. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome Res. 2012 Sep;22(9):1748-59.

27. Raychaudhuri S. Mapping rare and common causal alleles for complex human diseases. Cell. 2011 Sep 30;147(1):57-69.

28. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Hum Mol Genet. 2015 Oct 15;24(R1):R111-9.

29. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, Lee H, Oskolkov N, Fadista J, Ehlers K, Wahl S, Hoffmann C, Qian K, Rönn T, Riess H, Müller-Nurasyid M, Bretschneider N, Schroeder T, Skurk T, Horsthemke B; DIAGRAM+Consortium, Spieler D, Klingenspor M, Seifert M, Kern MJ, Mejhert N, Dahlman I, Hansson O, Hauck SM, Blüher M, Arner P, Groop L, Illig T, Suhre K, Hsu YH, Mellgren G, Hauner H, Laumen H. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell. 2014 Jan 16;156(1-2):343-58.

30. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJ, Shishkin AA, Hatan M, Carrasco-Alfonso MJ, Mayer D, Luckey CJ, Patsopoulos NA, De Jager PL, Kuchroo VK, Epstein CB, Daly MJ, Hafler DA, Bernstein BE. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015 Feb 19;518(7539):337-43.

31. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG Jr, Kinney JB, Kellis M, Lander ES, Mikkelsen TS. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol. 2012 Feb 26;30(3):27-7.

32. Ludwig LS, Gazda HT, Eng JC, Eichhorn SW, Thiru P, Ghazvinian R, George TI, Gotlib JR, Beggs AH, Sieff CA, Lodish HF, Lander ES, Sankaran VG. Altered translation of GATA1 in Diamond-Blackfan anemia. Nat Med. 2014 Jul;20(7):748-53.

33. Kučera T, Otyepka M, Matušková A, Samad A, Kutejová E, Janata J. A computational study of the glycine-rich loop of mitochondrial processing peptidase. PLoS One. 2013 Sep 13;8(9):e74518.

34. Hu J, Liu J, Xue F, Halverson G, Reid M, Guo A, Chen L, Raza A, Galili N, Jaffray J, Lane J, Chasis JA, Taylor N, Mohandas N, An X. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. Blood. 2013 Apr 18;121(16):3246-53.

35. Sankaran VG, Menne TF, Šćepanović D, Vergilio JA, Ji P, Kim J, Thiru P, Orkin SH, Lander ES, Lodish HF. MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. Proc Natl Acad Sci U S A. 2011 Jan 25;108(4):1519-24.

36. King MJ, Behrens J, Rogers C, Flynn C, Greenwood D, Chambers K. Rapid flow cytometric test for the diagnosis of membrane cytoskeleton-associated haemolytic anaemia. Br J Haematol. 2000 Dec;111(3):924-33.

37. Sankaran VG, Ludwig LS, Sicinska E, Xu J, Bauer DE, Eng JC, Patterson HC, Metcalf RA, Natkunam Y, Orkin SH, Sicinski P, Lander ES, Lodish HF. Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. Genes Dev. 2012 Sep 15;26(18):2075-87.

38. Abkowitz JL, Sabo KM, Nakamoto B, Blau CA, Martin FH, Zsebo KM, Papayannopoulou T. Diamond-blackfan anemia: in vitro response of erythroid progenitors to the ligand for c-kit. Blood. 1991 Nov 1;78(9):2198-202

39. Koutnikova H, Campuzano V, Koenig M. Maturation of wild-type and mutated frataxin by the mitochondrial processing peptidase. Hum Mol Genet. 1998 Sep;7(9):1485-9..

40. Huang X, Wang Y, Yan W, Smith C, Ye Z, Wang J, Gao Y, Mendelsohn L, Cheng L. Production of Gene-Corrected Adult Beta Globin Protein in Human Erythrocytes Differentiated from Patient iPSCs After Genome Editing of the Sickle Point Mutation. Stem Cells. 2015 May;33(5):1470-9.

41. Vichinsky EP, Earles A, Johnson RA, Hoag MS, Williams A, Lubin B. Alloimmunization in sickle cell anemia and transfusion of racially unmatched blood. N Engl J Med. 1990 Jun 7;322(23):1617-21.

42. Soccio RE, Chen ER, Rajapurkar SR, Safabakhsh P, Marinis JM, Dispirito JR, Emmett MJ, Briggs ER, Fang B, Everett LJ, Lim HW, Won KJ, Steger DJ, Wu Y, Civelek M, Voight BF, Lazar MA. Genetic

Variation Determines PPARγ Function and Anti-diabetic Drug Response In Vivo. Cell. 2015 Jul 2;162(1):33-44.
43. Deindl P, Klar M, Drews D, Cremer M, Gammella E, Gassmann M, Dame C. Mice over-expressing human erythropoietin indicate that erythropoietin enhances expression of its receptor via up-regulated Gata1 and Tal1. Haematologica. 2014 Oct; 99(10):e205–e207.
44. Sankaran VG, Weiss MJ. Anemia: progress in molecular mechanisms and therapies. Nat Med. 2015 Mar;21(3):221-30.
45. Palau F. Friedreich's ataxia and frataxin: molecular genetics, evolution and pathogenesis. Int J Mol Med. 2001 Jun;7(6):581-9.
46. Cavadini P, Adamec J, Taroni F, Gakh O, Isaya G. Two-step processing of human frataxin by mitochondrial processing peptidase. Precursor and intermediate forms are cleaved at different rates. J Biol Chem. 2000 Dec 29;275(52):41469-75.
47. Jobling RK, Assoum M, Gakh O, Blaser S, Raiman JA, Mignot C, Roze E, Dürr A, Brice A, Lévy N, Prasad C, Paton T, Paterson AD, Roslin NM, Marshall CR, Desvignes JP, Roëckel-Trevisiol N, Scherer SW, Rouleau GA, Mégarbané A, Isaya G, Delague V, Yoon G. PMPCA mutations cause abnormal mitochondrial protein processing in patients with non-progressive cerebellar ataxia. Brain. 2015 Jun;138(Pt 6):1505-17.
48. Nagao Y, Kitada S, Kojima K, Toh H, Kuhara S, Ogishima T, Ito A. Glycine-rich region of mitochondrial processing peptidase alpha-subunit is essential for binding and cleavage of the precursor proteins. J Biol Chem. 2000 Nov 3;275(44):34552-6.
49. Greene AW, Grenier K, Aguileta MA, Muise S, Farazifard R, Haque ME, McBride HM, Park DS, Fon EA. Mitochondrial processing peptidase regulates PINK1 processing, import and Parkin recruitment. EMBO Rep. 2012 Apr 13(4):378-85.

Eidesstattliche Versicherung

„Ich, Felix Giani, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Using Genome-wide Association and Exome Sequencing Studies to Improve *In Vitro* Production of Red Blood Cells and to Elucidate Human Traits and Congenital Disorders“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)“ des ICMJE -*www.**icmje**.org*) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) entsprechen den URM (s.o) und werden von mir verantwortet.

Meine Anteile an den ausgewählten Publikationen entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Betreuer/in, angegeben sind. Sämtliche Publikationen, die aus dieser Dissertation hervorgegangen sind und bei denen ich Autor bin, entsprechen den URM (s.o) und werden von mir verantwortet.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum Unterschrift

Declaration of Contribution to the attached Publications / Anteilserklärung an den erfolgten Publikationen

Felix Caspar Giani contributed as follows to the publications mentioned below / hatte folgenden Anteil an den folgenden Publikationen.

Publication 1:

Giani FC*, Fiorini C*, Wakabayashi A*, Ludwig LS, Salem RM, Jobaliya CD, Regan SN, Ulirsch JC, Liang G, Steinberg-Shemer O, Guo MH, Esko T, Tong W, Brugnara C, Hirschhorn JN, Weiss MJ, Zon LI, Chou ST, French DL, Musunuru K, Sankaran VG.
*these authors contributed equally
Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells.
Cell Stem Cell, 2016, Jan 7;18(1):73-8.

Contribution / Beitrag im Einzelnen

Felix Giani conceptualized the study with Kiran Musunuru and Vijay Sankaran. He established various protocols for the Sankaran laboratory, including a cell culture system for the production of muture enucleaded RBCs with Aoi Wakabayashi. He established the Eosin-5-maleimide binding-test (Fig 2 D), PKH-Labeling with input from Leif Ludwig (Fig S1 D-E) and magnetic isolation of CD34+ cells from cord blood. He designed, performed and analyzed a majority of the experiments utilizing CD34+ stem and progenitor cells, including retroviral infections for *SH2B3* knockdown, flow cytometry assessment of differentiation and enucleation (Fig 1 E-G), cytospin preparation (Fig 1 H), Western blotting to assess SH2B3 and cell counts to assess cell expansion (Fig 2 A-B, Fig S1 A). He purified RNA and aided in the analysis of gene expression upon *SH2B3* knockdown. He in part analyzed data from human pluripotent stem cell work. He wrote the manuscript with Claudia Fiorini, Aoi Wakabayashi, Mitchell Weiss, Leonard Zon, Kiran Musunuru and Vijay Sankaran with input from all authors.

Publication 2:

Ulirsch JC, Nandakumar SK, Wang L, **Giani FC**, Zhang X, Rogov P, Melnikov A, McDonel P, Do R, Mikkelsen TS, Sankaran VG.
Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits.

Cell, 2016 Jun 2;165(6):1530-45.

Contribution / Beitrag im Einzelnen

Felix Giani designed and performed experiments for large scale production of GATA1+ K562 and appropriate controls to resemble early and later stage human erythroid progenitor cells (as for instance in Fig 3 B, Fig 4 E-H). He coordinated these experimental procedures with Li Wang to organize seamless transition to subsequent work at the collaborating laboratory at Broad institute and reviewed the manuscript.

Publication 3:

Joshi M, Anselm I, Shi J, Bale TA, Towne M, Schmitz-Abe K, Crowley L, **Giani FC**, Kazerounian S, Markianos K, Lidov HG, Folkerth R, Sankaran VG, Agrawal PB.
Mutations in the substrate binding glycine-rich loop of the mitochondrial processing peptidase-α protein (PMPCA) cause a severe mitochondrial disease.
Cold Spring Harb Mol Case Stud, 2016 May;2(3):a000786.

Contribution / Beitrag im Einzelnen

Felix Giani designed and performed experiments for lentiviral production for ectopic PMPCA expression and designed rescue experiments with Mugdha Joshi (Fig 3 H). He provided expertise, provided reagents and reviewed the manuscript.

____

Unterschrift, Datum und Stempel des betreuenden Hochschullehrers der Charité - Universitätsmedizin Berlin

____

Datum und Unterschrift des betreuenden Hochschullehrers der Harvard Medical School

____

Unterschrift des Doktoranden

Publication 1:

The original article and supplementary online material are included on the following pages.

Giani FC*, Fiorini C*, Wakabayashi A*, Ludwig LS, Salem RM, Jobaliya CD, Regan SN, Ulirsch JC, Liang G, Steinberg-Shemer O, Guo MH, Esko T, Tong W, Brugnara C, Hirschhorn JN, Weiss MJ, Zon LI, Chou ST, French DL, Musunuru K, Sankaran VG. *co first-author

Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells.

Cell Stem Cell. 2016 Jan 7;18(1):73-8. https://doi.org/10.1016/j.stem.2015.09.015.

Publication 2:

The original article and supplementary online material are included on the following pages.

Ulirsch JC, Nandakumar SK, Wang L, **Giani FC**, Zhang X, Rogov P, Melnikov A, McDonel P, Do R, Mikkelsen TS, Sankaran VG.
Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits.
Cell. 2016 Jun 2;165(6):1530-45. https://doi.org/10.1016/j.cell.2016.04.048.

Publication 3:

The original article is included on the following pages.

Joshi M, Anselm I, Shi J, Bale TA, Towne M, Schmitz-Abe K, Crowley L, **Giani FC**, Kazerounian S, Markianos K, Lidov HG, Folkerth R, Sankaran VG, Agrawal PB.
Mutations in the substrate binding glycine-rich loop of the mitochondrial processing peptidase-α protein (PMPCA) cause a severe mitochondrial disease.
Cold Spring Harb Mol Case Stud. 2016 May;2(3):a000786. https://doi.org/10.1101/mcs.a000786.

COLD SPRING HARBOR
Molecular Case Studies

RESEARCH ARTICLE

Mutations in the substrate binding glycine-rich loop of the mitochondrial processing peptidase-α protein (PMPCA) cause a severe mitochondrial disease

Mugdha Joshi,[1,2,3,10] Irina Anselm,[4,10] Jiahai Shi,[5,6] Tejus A. Bale,[7] Meghan Towne,[1,3] Klaus Schmitz-Abe,[1] Laura Crowley,[1,3] Felix C. Giani,[8,9] Shideh Kazerounian,[1] Kyriacos Markianos,[1] Hart G. Lidov,[7] Rebecca Folkerth,[7] Vijay G. Sankaran,[8,9] and Pankaj B. Agrawal[1,2,3]

[1]Division of Genetics and Genomics, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA; [2]Division of Newborn Medicine, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA; [3]Gene Discovery Core, Manton Center for Orphan Disease Research, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA; [4]Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA; [5]Whitehead Institute for Biomedical Research, MIT, Cambridge, Massachusetts 02142, USA; [6]Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China; [7]Department of Pathology, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA; [8]Division of Hematology/Oncology, Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, Massachusetts 02115, USA; [9]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

Corresponding author: pagrawal@enders.tch.harvard.edu

© 2016 Joshi et al. This article is distributed under the terms of the Creative Commons Attribution-NonCommercial License, which permits reuse and redistribution, except for commercial purposes, provided that the original author and source are credited.

Ontology terms: diffuse cerebellar atrophy; generalized hypotonia due to defect at the neuromuscular junction; hydrocephalus; mitochondrial encephalopathy

Published by Cold Spring Harbor Laboratory Press

doi: 10.1101/mcs.a000786

Abstract We describe a large Lebanese family with two affected members, a young female proband and her male cousin, who had multisystem involvement including profound global developmental delay, severe hypotonia and weakness, respiratory insufficiency, blindness, and lactic acidemia—findings consistent with an underlying mitochondrial disorder. Whole-exome sequencing was performed on DNA from the proband and both parents. The proband and her cousin carried compound heterozygous mutations in the *PMPCA* gene that encodes for α-mitochondrial processing peptidase (α-MPP), a protein likely involved in the processing of mitochondrial proteins. The variants were located close to and postulated to affect the substrate binding glycine-rich loop of the α-MPP protein. Functional assays including immunofluorescence and western blot analysis on patient's fibroblasts revealed that these variants reduced α-MPP levels and impaired frataxin production and processing. We further determined that those defects could be rescued through the expression of exogenous wild-type *PMPCA* cDNA. Our findings link defective α-MPP protein to a severe mitochondrial disease.

INTRODUCTION

The majority of mitochondrial proteins are encoded as precursor proteins from nuclear DNA, synthesized in the cytosol, and then imported into the mitochondria. These precursor

[10]These authors contributed equally to this work.

proteins contain amino-terminal extensions called targeting peptides (TPs) that function as import signals and are recognized by import receptors present on the mitochondrial membrane. Once imported into the mitochondria, the TP is cleaved off and the protein can fold and fully mature, allowing for optimal protein function. The majority of these essential cleavage reactions are catalyzed by MPP (mitochondrial processing peptidase), a matrix-localized, soluble heterodimer of α-MPP and β-MPP subunits, encoded by *PMPCA* and *PMPCB*, respectively (Kucera et al. 2013). In budding yeast, when either MPP subunit is eliminated, the cells continue to import proteins into the mitochondria, but they fail to cleave TPs, the precursor proteins accumulate, and the cells fail to grow (Geli and Glick 1990). The limited studies available on *PMPCA* and *PMPCB* have suggested that the catalytic activity of MPP is confined to the β-MPP subunit, whereas the α-MPP subunit seems to be involved in substrate recognition (Luciano et al. 1997; Gakh et al. 2002). Existing data have suggested that a flexible, glycine-rich loop within the α-MPP subunit is largely responsible for substrate affinity and MPP functionality (Taylor et al. 2001). This loop is a key structural element of MPP that moves the precursor protein substrate toward the active site through a multistep process (Kucera et al. 2013).

As appropriate cleavage by processing peptidases is necessary for protein stability, correct localization, and proper assembly of protein complexes, some have hypothesized a link between MPP dysfunction and human disease (Teixeira and Glaser 2013). One such example is the association between MPP and Friedreich's ataxia (FRDA), an autosomal-recessive neurodegenerative disease in which the mitochondrial protein frataxin is deficient (Palau 2001; Wilson 2003; Lim et al. 2007). It has been shown that frataxin is processed by MPP and mutations in frataxin alter this interaction, resulting in frataxin deficiency and impaired function (Koutnikova et al. 1998; Cavadini et al. 2000). Further evidence for this hypothesis is suggested by the finding that without proper processing by MPP, precursor proteins, generally more unstable than their processed mature form, are readily degraded (Mukhopadhyay et al. 2007). Here we describe two first cousins with a complex multisystem mitochondrial disease carrying compound heterozygous missense variants in *PMPCA* associated with a marked reduction in the levels of processed frataxin. We also demonstrated rescue in the proband's fibroblasts by overexpressing wild-type (WT) *PMPCA* using a lentiviral vector.

RESULTS

Clinical Presentation

We describe a large family with two members presenting with a severe mitochondrial disease. The pedigree is shown in Figure 1A. The proband is a 6-yr-old female of Lebanese descent who was hospitalized at the age of 6 mo with developmental delay and failure to thrive. Her neurological examination revealed bilateral ptosis, ophthamoplegia, mild generalized hypotonia, and weakness. Laboratory tests revealed lactic acidemia and elevated transaminases. A brain magnetic resonance imaging (MRI) revealed cerebellar and mild diffuse cerebral atrophy (Fig. 1B,C). Magnetic resonance spectroscopy revealed a lactate peak (data not shown). Over the subsequent years, her condition deteriorated and by 6 yr she was completely blind and had severe optic atrophy. She was minimally responsive and has been immobile because of profound generalized weakness. She has chronic hypoventilation and a tracheostomy was placed for ventilator support. She has persistent lactic acidemia, whereas her transaminases gradually normalized. MRIs over the years have shown progression of generalized brain atrophy, with diffuse parenchymal volume loss and areas of gliosis and severe ventriculomegaly involving both of the lateral ventricles as well as the third and fourth ventricles (Fig. 1D,E).

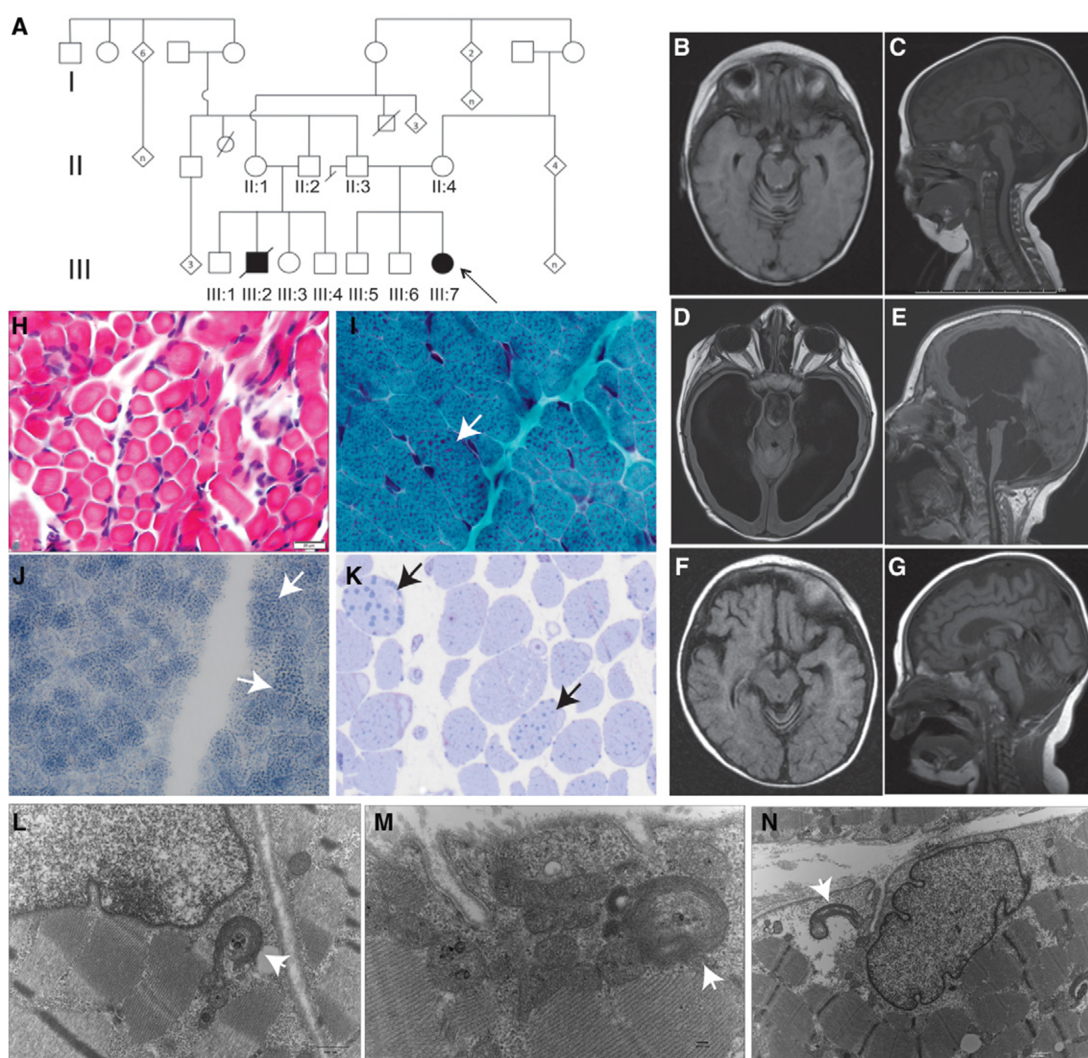

Figure 1. Clinical and histopathological findings in a family with *PMPCA* mutations. (*A*) Pedigree of the extended family carrying *PMPCA* mutations, Arrow indicates the proband. (*B*–*G*) Magnetic resonance imaging (MRI) head findings in the proband (*B*–*E*) and affected cousin (*F*,*G*). MRI at 6 mo of age revealed cerebellar atrophy with enlarged interfolial spaces (*B*,*C*), whereas at 3 yr of age there was marked cerebral and cerebellar atrophy with enlarged ventricles (*D*,*E*). The cousin's head MRI at 8 mo of age revealed cerebral and cerebellar atrophy (*F*,*G*). (*H*–*N*) Histopathological findings from the muscle biopsy in the proband (*H*), the affected cousin (*I*–*K*), and electron microscopic findings in the proband (*L*–*N*). Fiber size variation was noted on H&E (hematoxylin and eosin) staining (*H*), whereas trichrome staining revealed coarse granular staining suggesting enlarged mitochondria (marked by arrows) (*I*), and was also noted on NADH (nicotinamide adenine dinucleotide) staining (*J*). Toludine blue staining showed large prominent lipid droplets (arrows) (*K*). Electron microscopy revealed scattered structurally abnormal mitochondria with electron dense aggregates within cristae (arrows) (*L*–*N*).

Her muscle biopsy findings included mild variation in fiber size (Fig. 1H) and an unusual staining pattern with cytochrome oxidase (COX) characterized by clumping of the reaction product in both subsarcolemmal and intermyofibrillar distribution suggestive of a mitochondrial myopathy. Similar findings were present on staining by succinate dehydrogenase, and electron microscopy revealed scattered structurally abnormal mitochondria. The mitochondria were enlarged and elongated with electron dense aggregates within cristae (Fig. 1L–N).

Mitochondrial respiratory electron transport chain (ETC) enzyme analysis was performed on skin fibroblasts, and no enzymatic defect was detected in any component of the respiratory complex. Furthermore, activities of pyruvate carboxylase, phosphoenolpyruvate carboxykinase, and citrate synthase in cultured skin fibroblasts were normal. The whole mitochondrial genome sequence of the patient was negative for point mutations or deletions. Additional DNA sequencing was performed for several mitochondrial nuclear genes followed by comprehensive mitochondrial nuclear gene panel analysis of 101 genes that revealed no pathogenic mutations. Her chromosomal microarray analysis did not reveal any significant copy-number variations.

Patient 2 is the first cousin of the proband. Their fathers are brothers and their mothers are first cousins (Fig. 1A). His presentation at 6 mo of age was similar to that of the proband with developmental delay, weakness, and ataxia. He had persistent lactic academia, as well. Brain MRIs showed marked cerebellar and mild cerebral atrophy (Fig. 1F,G). He developed severe hypertrophic left ventricular cardiomyopathy and liver failure and died at 14 mo from respiratory failure in the setting of having multiple medical complications. A limited autopsy including liver and skeletal muscles was performed. Skeletal muscle findings revealed the presence of multiple fibers with subsarcolemmal clusters of abnormal mitochondria (Fig. 1I,J) and lipid droplets indicative of mitochondrial dysfunction (Fig. 1K). The reported electron microscopic (EM) abnormalities included giant mitochondria (megaconial) and mitochondria with concentric cristae. ETC analysis on the muscle showed normal activity of complexes I, II, III, and IV. Liver histopathology revealed mild hepatic fibrosis, marked cholestasis, and microvesicular steatosis. The electron microscopy revealed glycogen depletion in hepatocytes. Mitochondria appeared swollen with increased matrix and margination of cristae, with matrical inclusions of neutral fat and vacuolar structures. The ETC activity in the liver specimen was suggestive of defects in complexes II and III.

Exome Sequencing

Exome sequencing and analysis was performed on the proband and both parents as described in the Methods section and the coverage is described in Table 1. Two variants in the *PMPCA* gene were identified in the proband that segregated appropriately in the parents as described in Table 2. PMPCA was a prime candidate given that it has been ascribed a role in the mitochondria based on a large integrative genomic database of putative mitochondrial genes (Pagliarini et al. 2008). The proband and the cousin were confirmed to be compound heterozygous for the two missense *PMPCA* mutations (Chr 9: 139313082; NM_015160, c.1066G>A; p.G356S and Chr 9: 139313299; NM_01560, c.1129G>A; p. A377T) by Sanger sequencing with appropriate parental segregation of the mutant alleles. (Fig. 2A). Several unaffected siblings (III:1, III:3, III:4, III:5, and III:6) were WT or heterozygous for a single variant. The identified mutations were not present in any of the publicly available databases including the 1000 Genomes, Exome Variant Server, and Exome Aggregation Consortium (ExAC) databases. Several in silico missense variant prediction methods, including PolyPhen-2, MutationTaster, and SIFT, suggested that the missense variants we

Table 1. Whole-exome sequencing coverage

Sample	Total aligned reads	Average read coverage[a]	% Coding genes[a] with >10-fold coverage
Proband	31,863,723	41.9	84.1
Father	42,287,619	54.6	87.1
Mother	33,224,612	42.9	84.4

[a]Based upon the consensus coding DNA sequence (CCDS) database.

Table 2. *PMPCA* (NM_015160) variants

Chr:Position GRCh37 (hg19)	HGVS cDNA	HGVS protein	Predicted effect	dbSNP ID	Genotype	Parent	MAF (%) ExAC database
9:139313082	c.1066G>A	p.G356S	Missense	Absent	Heterozygous	Mother	0
9:139313299	c.1129G>A	p.A377T	Missense	Absent	Heterozygous	Father	0

HGVS, Human Genome Variation Society; dbSNP, Database for Short Genetic Variations; MAF, minor allele frequency; ExAC, Exome Aggregation Consortium.

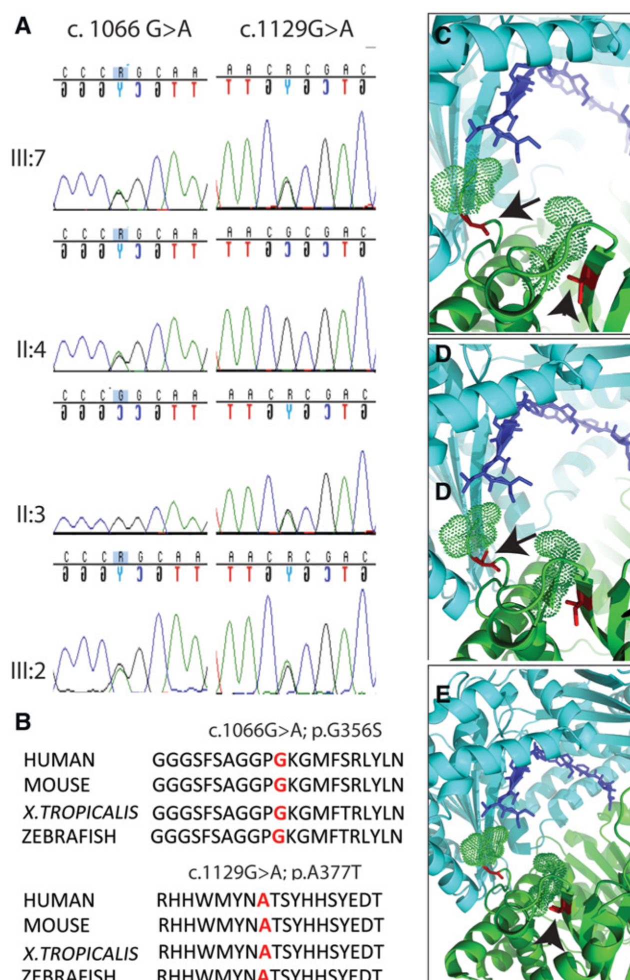

Figure 2. Molecular findings in the family with *PMPCA* mutations. (*A*) Distribution of the two heterozygous mutations c.1066G>A (p.G356S) and c. 1129G>A (p.A377T) in the proband, her affected cousin, and both their sets of parents. (*B*) Conservation of the amino acid residues across the vertebrates. (*C–E*) The crystal structure of yeast PMPCA with its substrate (PDB: 1HR8) is presented to show the location of the mutations. (*C*) WT-PMPCA is a heterodimer that consists of two subunits, α and β. The β-subunit is colored in cyan, whereas the α-subunit is colored in green. The blue stick represents a substrate. Part of the catalytic loop (residues 345–355) is shown in dots, with a few residues missing in the *middle* because of high flexibility. (*D*,*E*) The two mutation sites colored in red, depicted by arrow p.G356S (*D*) and arrowhead A377T (*E*), are located in and 20 residues downstream from the catalytic loop. The increased side chain volume caused by the mutations likely restricts the dynamic of the catalytic loop, which affects the catalytic activity of PMPCA.

identified were likely to be deleterious and the altered amino acids were extremely well conserved across evolution (Fig. 2B). The locations of the mutations suggest that both could affect the functionality of the α-MPP glycine-rich loop. The significance of these residues is further corroborated by the fact that p.A377 is conserved across vertebrates and invertebrates, whereas p.G356 is conserved across vertebrates, invertebrates, and monera.

Computer modeling revealed that the mutations were located in p.G356S or 20 residues downstream (p.A377T) of the glycine-rich catalytic loop (residues 345–358), which would be predicted to restrict the dynamic movements of this catalytic loop and thereby affect the catalytic activity or substrate binding of PMPCA (Fig. 2C–E). This loop is highly conserved among vertebrates and invertebrates, and its depletion has been shown to reduce the enzyme's affinity for substrate peptides and its activity (Nagao et al. 2000; Kucera et al. 2013). The glycine-rich loop is positioned at the opening to the space between the subunits, such that it is exposed to the external substrate and the zinc-binding pocket. The loop needs to be highly flexible as the process of substrate translocation causes it to undergo major conformational changes (Kucera et al. 2013). The intraloop p.G356S mutation may directly alter the conformation and affinity of the loop by replacing a small glycine with a bulkier and more nucleophilic serine residue. This could interfere with the flexibility of the loop, as well as with the ability of the loop to recognize and bind to substrates, change conformation following early TP cleavage in multistep processing, and maintain heterodimer stability. The downstream p.A377T mutation replaces the alanine residue with a threonine bearing an alcohol group. Moreover, the side chain of threonine points toward to the catalytic loop. The enhanced polarity and bulkier size of p. A377T mutation could additionally affect the flexibility of the catalytic loop and the stability of the heterodimer.

Functional Effects of the Mutations

To evaluate the functional effects of *PMPCA* mutations, fibroblasts from the proband and two age-matched controls were obtained. Immunofluorescence experiments were performed using anti-PMPCA and anti-mitochondrial antibodies that revealed a reduction in PMPCA staining along with several abnormally enlarged mitochondria (Fig. 3A–F).

One of the proteins extensively processed inside mitochondria is frataxin, encoded by *FXN*, and known to interact with MPP (Koutnikova et al. 1998; Branda et al. 1999; Gordon et al. 1999; Cavadini et al. 2000). Frataxin deficiency causes FRDA, an autosomal-recessive neurodegenerative disease (Palau 2001; Wilson 2003; Lim et al. 2007). An imbalance in the ratio of immature to mature frataxin and general depletion of the processed forms characterizes FRDA (Palau 2001; Wilson 2003; Lim et al. 2007). Studies have shown that *FXN* mutations alter its interaction with MPP, resulting in impaired frataxin function (Koutnikova et al. 1998). We performed experiments on cultured fibroblasts from the proband and two age-matched controls for frataxin processing and found abnormal processing in the proband. The fibroblasts from the proband contained higher levels of the immature form of frataxin (23 kDa) and lower levels of the processed or mature form (18 kDa) (Fig. 3G).

To directly show that the observed reduction in mature frataxin is due to the *PMPCA* mutations, we performed rescue experiments using lentiviral transduction of fibroblasts from Patient 1. The *PMPCA*-rescued cells from the proband revealed a reduction in the unprocessed form of frataxin (23 kDa) and an increase in the levels of the processed (mature) form (18 kDa) in comparison to the control rescue (Fig. 3H). The levels of PMPCA increased appropriately in the PMPCA-rescued cells as compared with the control (Fig. 3H). This directly demonstrates that the mutant alleles in *PMPCA* are loss of function and can be rescued by transduction of the WT gene in primary patient cells. Moreover, this also shows that the processing defect observed in the patient cells is due to loss of function of *PMPCA*.

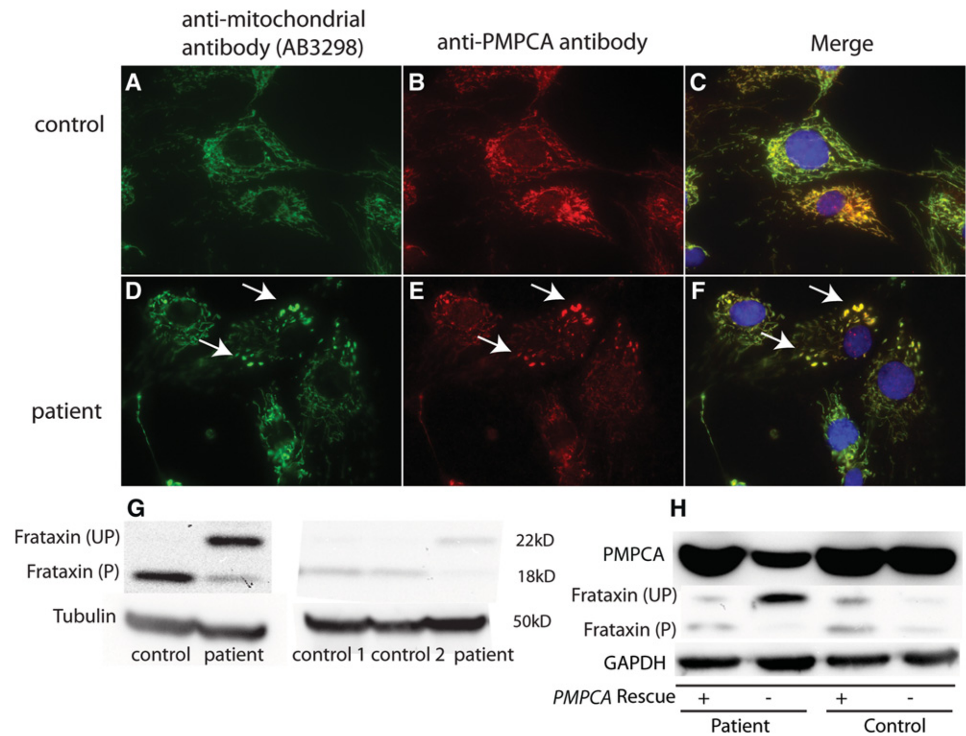

Figure 3. Effects of *PMPCA* mutations on fibroblasts cultured from the proband and rescue experiments. (*A–F*) Immunofluorescence showing swollen mitochondria (arrows) and abnormal PMPCA staining compared with control. (*G*) Western blot analysis revealed a reduction in mature frataxin (18 kDa) and an increase in unprocessed frataxin (23 kDa). (*H*) WT *PMPCA* cDNA overexpression in the proband's fibroblasts resulted in normalization of frataxin processing with high amounts of processed frataxin and reduction in unprocessed form. Empty vector rescue was used as control showing absent processed frataxin and a high amount of unprocessed frataxin. Further, PMPCA levels are much higher in the rescued cells than in the one rescued by the empty vector.

DISCUSSION

This study describes a severe mitochondrial disease due to *PMPCA* mutations in a large family, which we show is associated with altered levels of mature frataxin. Whereas a recent study described cerebellar ataxia as a presentation of *PMPCA* mutations (Jobling et al. 2015), the phenotype in our family is more severe and typical of a mitochondrial disease. The increased severity of disease in our family may be due to direct perturbation of the glycine-rich catalytic loop, which was not seen in the patients described by Jobling et al. (2015). The p.A377T variant, present in homozygous state in their patients, is downstream from the catalytic loop and may have milder effects. The proband had severe muscle weakness, extensive brain atrophy, visual impairment, and respiratory defects, and the affected cousin in addition had cardiomyopathy and liver complications and died at an early age from resultant complications. The observed phenotype may be related to impaired PMPCA function due to a reduction in its level and the resultant abnormal processing of frataxin and other mitochondrial proteins. Importantly, the observed processing defects were rescued by ectopic expression of the wild-type form of *PMPCA*. The mutations observed in both affected family members appear to affect the critical glycine-rich catalytic loop within the α-subunit of the MPP enzyme, which is responsible for moving the precursor protein substrates toward the active site on the β-subunit for processing to mature forms (Taylor et al. 2001; Kucera et al. 2013).

We have described abnormal frataxin processing in this report, but there are more than 1000 nuclear-encoded proteins that undergo mitochondrial import and amino-terminal TP processing in the mitochondria. We expect that the maturation of several additional mitochondrial proteins may be affected by an α-MPP functional deficit and therefore contribute to the severe phenotype observed in the family. We have tested several additional mitochondrial proteins for abnormal processing using the patient's fibroblasts. One of them is PINK1, potentially affected by PMPCA levels (Greene et al. 2012), but we did not find a change in PINK1 amounts in patient's fibroblasts (data not shown). Similarly, members of the electron transport chain family (COX4I1, ATP5A1, and ATP5B) were tested without any evidence of a defect in their processing (data not shown). These findings indicate that either the above proteins are processed independently of MPP or that the mutations in *PMPCA* may not affect their recognition by the MPP for processing. A search for additional mitochondrial proteins that may be affected by *PMPCA* mutations is currently underway.

In summary, by combining exome-sequencing and variant-segregation analysis across a family with a unique syndrome along with comparative genomic, structural, and functional rescue experiments, we have been able to demonstrate causality for a mitochondrial disease resulting from PMPCA disruption (MacArthur et al. 2014). Additionally, by testing a multitude of candidate mitochondrial import proteins, we will be able to establish the mechanism by which this disruption leads to specific phenotypes, as observed in these patients with a multisystemic disorder.

METHODS

Genetic Studies

The proband (Patient 1), her parents, the affected cousin (Patient 2) and his parents, and other family members were enrolled in an institutional review board (IRB)-approved study at Boston Children's Hospital (BCH) to perform whole-exome sequencing. Blood samples were collected and processed for DNA extraction by the Research Connection Biobank Core at BCH. DNA from the proband and both parents was sent for whole-exome sequencing (WES) to Axeq Technologies. Samples were prepared as an Illumina sequencing library and enriched for exonic sequences using the Illumina Exome Enrichment protocol. The captured libraries were sequenced using Illumina HiSeq 2000 Sequencers. The reads were mapped to the human genome assembly UCSC hg19 using Burrows–Wheeler Alignment (BWA version 0.5.8). Single-nucleotide polymorphisms (SNPs) and small insertions/deletions were called with SAMtools (version 0.1.7). The resulting VCF (variant call format) files were filtered to include nonsynonymous, splice site, and indel variants with an allele frequency <0.001 in the National Heart, Lung, and Blood Institute (NHLBI) Exome Variant Server (EVS) database (http://evs.gs.washington.edu/EVS/) or <0.01 in 1000 Genomes Project, phase 3 (http://www.1000genomes.org). The Variant Explorer pipeline developed by the authors (K.S.-A. and K.M.) was used for annotating all variants according to the minor-allele-frequency data described above. Polymerase chain reaction (PCR) was performed to amplify the *PMPCA* mutations using the primers PMPCA_F: CTTCATCCCCTTTGCAGTGT and PMPCA_R: TACAACGTGGCTGTCTCCG, and the amplified DNA returned was sent for Sanger sequencing. The results were analyzed using the Sequencer 5.0 software and aligned with the WT gene sequence available from NCBI Blast.

Protein Expression Study

Western blot was performed to study the level of PMPCA protein on the fibroblasts available from the patient. Fibroblasts from two control human cell lines (ATCC-2127 and ATCC-2104)

were used as controls. Anti-PMPCA antibody (NBP1-89126, 1:100 dilution, Novus Biologicals) was used to probe the specific protein band at 58 kDa. The results of the study were visualized using enhanced chemiluminescence using the program Quantity One 4.2.1 (Bio-Rad) and were compared with appropriate age- and tissue-matched controls normalized for the loading control glyceraldehyde-3-phosphate dehydrogenase (GAPDH) or β-tubulin.

To identify the probable interactions of PMPCA with other proteins that may be processed in the mitochondria, we also performed quantitative protein analysis for CO4I1, frataxin, PINK1, and a mitochondrial antibody cocktail (ab3298, Abcam).

Immunofluorescence on Skin Fibroblasts

Cultured patient fibroblasts were transferred onto chambered slides and grown to 50%–60% confluence. The fibroblasts were fixed with 4% paraformaldehyde for 20 min and 10% Triton-X was used to permeabilize the cells before staining. The fixed slides were stored and used up to a week. Rabbit anti-PMPCA from Novus Biologicals (SAB1303187; 100 µL) made in rabbit was used as a primary antibody at a ratio of 1:100 overnight at 4°C. Anti-mitochondrial antibody (mouse) from Abcam (ab 3298) was used for costaining at 1:200 overnight at 4°C.

Abcam frataxin antibody (mouse) (17A11) (ab113691) was also used at 1:200 with Novus PMPCA for colocalization studies.

Viral Transfection of Skin Fibroblasts for Rescue

Human *PMPCA* cDNA clone (HsCD00334734) was ordered from the Plasmid ID database and the gene was cloned into HMD lentiviral transfection vector with a green fluorescent protein (GFP) tag (to assess for those cells that have successfully been lentivirally transduced) (Ludwig et al. 2014) by restriction digestion using the following primers: PMPCA_cloning_F:TAAGCAGGATCCATGGCGGCTGTGGTGCTGGCG and PMPCA_cloning_R: TGCTTACTCGAGCTACCGGAAGAGCCGTGCGCA.

0059-1 (patient) and 2127 (control) fibroblasts were each then transfected with HMD-PMPCA (rescue) viral transfection vector and an empty HMD vector (control) without the PMPCA insert as another control. Cells that were successfully transfected were separated from the other cells using fluorescence-activated cell sorting (FACS). The transfected cells (0059_HMD-PMPCA, 0059-HMD only, 2127_HMD-PMPCA and 2127_HMD only) were then allowed to grow and washed with media two to three times before splitting and kept frozen at −140°C for future use. Western blot was repeated with the transfected fibroblasts using PMPCA and frataxin antibodies to study the effects of the rescue.

ADDITIONAL INFORMATION

Ethics Statement

The Institutional Review Board of Boston Children's Hospital approved this research protocol (10-02-0053). Written informed consent was received from proband's parents, parents of the cousin, and other family members.

Data Deposition and Access

Our patient consent does not allow patient sequence data to be uploaded to a data repository. The NM_015160: c.1066G>A; p.G356S and c.1129G>A; p.A377T missense variants have been deposited to ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) under accession numbers SCV000262557 and SCV000262558 and MSeqDR (https://mseqdr.org/) under accession numbers MSCV_0000001 and MSCV_0000002.

Acknowledgments

We thank the family for their support and enrollment.

Author Contributions

M.J. drafted and revised the manuscript for content, including medical writing for content, analysis or interpretation of data, and acquisition of data. I.A. contributed to study concept and design, drafting and revising the manuscript for content, including medical writing for content, analysis, or interpretation of data. J.S. contributed to the acquisition of data, analysis and interpretation of data, and revision of the manuscript for content. T.A.B., M.T., and S.K. contributed to acquisition of data. K.S.-A. contributed to analysis or interpretation of data and statistical analysis. L.C. contributed to drafting and revising the manuscript for content. F.C.G. contributed vital reagents and acquisition of data. K.M. contributed to analysis and interpretation of data and statistical analysis. H.G.L. and R.F. contributed to analysis and interpretation of data and revised the manuscript for content. V.G.S. drafted and revised the manuscript for content and contributed vital reagents and acquired, analyzed, and interpreted data. P.B.A. contributed to the study concept and design, obtained funding, drafted and revised the manuscript for content, analysis, and interpretation of data, and supervised and coordinated study.

Competing Interest Statement

The authors have declared no competing interest.

Referees

Neil Sondheimer
Anonymous

Received November 12, 2015; accepted in revised form January 11, 2016.

Funding

This work was made possible by grants from the following institutions: P.B.A. was supported by the William F. Milton fund from Harvard Medical School, Boston, 1R01AR068429-01 from the National Institute of Arthritis and Musculoskeletal and Skeletal Diseases (NIAMS) of the National Institute of Health (NIH), and U19HD077671 from National Institute of Child Health and Human Development (NICHD)/National Human Genome Research Institute (NHGRI)/NIH. The Gene Discovery Core of The Manton Center for Orphan Disease Research, Boston Children's Hospital also supported the work. Sanger sequencing was performed by the Molecular Genetics Core Facility of the IDDRC at Boston Children's Hospital, which was supported by National Institutes of Health grant P30 HD18655.

REFERENCES

Branda SS, Cavadini P, Adamec J, Kalousek F, Taroni F, Isaya G. 1999. Yeast and human frataxin are processed to mature form in two sequential steps by the mitochondrial processing peptidase. *J Biol Chem* **274:** 22763–22769.

Cavadini P, Adamec J, Taroni F, Gakh O, Isaya G. 2000. Two-step processing of human frataxin by mitochondrial processing peptidase. Precursor and intermediate forms are cleaved at different rates. *J Biol Chem* **275:** 41469–41475.

Gakh O, Cavadini P, Isaya G. 2002. Mitochondrial processing peptidases. *Biochim Biophys Acta* **1592:** 63–77.

Geli V, Glick B. 1990. Mitochondrial protein import. *J Bioenerg Biomembr* **22:** 725–751.

Gordon DM, Shi Q, Dancis A, Pain D. 1999. Maturation of frataxin within mammalian and yeast mitochondria: one-step processing by matrix processing peptidase. *Hum Mol Genet* **8:** 2255–2262.

Greene AW, Grenier K, Aguileta MA, Muise S, Farazifard R, Haque ME, McBride HM, Park DS, Fon EA. 2012. Mitochondrial processing peptidase regulates PINK1 processing, import and Parkin recruitment. *EMBO Rep* **13:** 378–385.

Jobling RK, Assoum M, Gakh O, Blaser S, Raiman JA, Mignot C, Roze E, Durr A, Brice A, Levy N, et al. 2015. PMPCA mutations cause abnormal mitochondrial protein processing in patients with non-progressive cerebellar ataxia. *Brain* **138:** 1505–1517.

Koutnikova H, Campuzano V, Koenig M. 1998. Maturation of wild-type and mutated frataxin by the mitochondrial processing peptidase. *Hum Mol Genet* **7:** 1485–1489.

Kucera T, Otyepka M, Matuskova A, Samad A, Kutejova E, Janata J. 2013. A computational study of the glycine-rich loop of mitochondrial processing peptidase. *PLoS One* **8:** e74518.

Lim F, Palomo GM, Mauritz C, Gimenez-Cassina A, Illana B, Wandosell F, Diaz-Nido J. 2007. Functional recovery in a Friedreich's ataxia mouse model by frataxin gene transfer using an HSV-1 amplicon vector. *Mol Ther* **15:** 1072–1078.

Luciano P, Geoffroy S, Brandt A, Hernandez JF, Geli V. 1997. Functional cooperation of the mitochondrial processing peptidase subunits. *J Mol Biol* **272:** 213–225.

Ludwig LS, Gazda HT, Eng JC, Eichhorn SW, Thiru P, Ghazvinian R, George TI, Gotlib JR, Beggs AH, Sieff CA, et al. 2014. Altered translation of GATA1 in Diamond–Blackfan anemia. *Nat Med* **20:** 748–753.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, et al. 2014. Guidelines for investigating causality of sequence variants in human disease. *Nature* **508:** 469–476.

Mukhopadhyay A, Yang CS, Wei B, Weiner H. 2007. Precursor protein is readily degraded in mitochondrial matrix space if the leader is not processed by mitochondrial processing peptidase. *J Biol Chem* **282:** 37266–37275.

Nagao Y, Kitada S, Kojima K, Toh H, Kuhara S, Ogishima T, Ito A. 2000. Glycine-rich region of mitochondrial processing peptidase α-subunit is essential for binding and cleavage of the precursor proteins. *J Biol Chem* **275:** 34552–34556.

Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, Walford GA, Sugiana C, Boneh A, Chen WK, et al. 2008. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134:** 112–123.

Palau F. 2001. Friedreich's ataxia and frataxin: molecular genetics, evolution and pathogenesis (Review). *Int J Mol Med* **7:** 581–589.

Taylor AB, Smith BS, Kitada S, Kojima K, Miyaura H, Otwinowski Z, Ito A, Deisenhofer J. 2001. Crystal structures of mitochondrial processing peptidase reveal the mode for specific cleavage of import signal sequences. *Structure* **9:** 615–625.

Teixeira PF, Glaser E. 2013. Processing peptidases in mitochondria and chloroplasts. *Biochim Biophys Acta* **1833:** 360–370.

Wilson RB. 2003. Frataxin and frataxin deficiency in Friedreich's ataxia. *J Neurol Sci* **207:** 103–105.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Publication List

Giani FC, Fiorini C, Wakabayashi A, Ludwig LS, Salem RM, Jobaliya CD, Regan SN, Ulirsch JC, Liang G, Steinberg-Shemer O, Guo MH, Esko T, Tong W, Brugnara C, Hirschhorn JN, Weiss MJ, Zon LI, Chou ST, French DL, Musunuru K, Sankaran VG. Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells. **Cell Stem Cell**. 2016 Jan 7;18(1):73-8. IF 23,290

Joshi M, Anselm I, Shi J, Bale TA, Towne M, Schmitz-Abe K, Crowley L, **Giani FC**, Kazerounian S, Markianos K, Lidov HG, Folkerth R, Sankaran VG, Agrawal PB. Mutations in the substrate binding glycine-rich loop of the mitochondrial processing peptidase-α protein (PMPCA) cause a severe mitochondrial disease. **Cold Spring Harb Mol Case Stud**. 2016 May;2(3):a000786. IF N/A

Ulirsch JC, Nandakumar SK, Wang L, **Giani FC**, Zhang X, Rogov P, Melnikov A, McDonel P, Do R, Mikkelsen TS, Sankaran VG. Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. **Cell**. 2016 Jun 2;165(6):1530-45. IF 31,398

Berlin, 19.11.2019

Acknowledgements

I would like to thank Vijay Sankaran and Christof Dame for giving me the opportunity to work in an extremely fascinating research environment. I am grateful for the exquisite research training, the advice, the inspiration, the enthusiasm and the superb support inside and outside the laboratory.

I also thank my friends and labmates for support, interesting discussions, their help and the fun we had during or after work in the laboratory. I thank Anindita Basak, Christian Brandel, Jennifer Eng, Claudia Fiorini, Rajiv Khajuria, Jessica Lacy, Jacob Ulirsch and Aoi Wakabayashi. I would like to particularly emphasize my gratitude for Leif Ludwig for all the help, all scientific knowledge he shared and great support. I would like to thank Lina Sciesielski for proof reading this manuscript and giving helpful advice. I also thank Lüder Wiebusch for excellent instructions during my very first lab experiences in Berlin.

Furthermore, I thank my fellow students and close friends Frederik Fahrendorf, André Gentsch, Simon Kruse, Maximilian Nossek, Sandro Schneider and Eleni Stathopoulou.

My profound thanks go to my parents Luise and Dieter Giani for their outstanding support of my endeavors, their trust, their optimism and the unlimited moral and emotional support. I wish to also thank the other family members for their support, especially Erna Karkheck, Gerrit Meyer and Eka Meyer-Lausch.