SUPPLEMENTARY APPENDIX

# Genomic analysis of bone marrow failure and myelodysplastic syndromes reveals phenotypic and diagnostic complexity

Michael Y. Zhang,[1] Siobán B. Keel,[2] Tom Walsh,[3] Ming K. Lee,[3] Suleyman Gulsuner,[3] Amanda C. Watts,[3] Colin C. Pritchard,[4] Stephen J. Salipante,[4] Michael R. Jeng,[5] Inga Hofmann,[6] David A. Williams,[6,7] Mark D. Fleming,[8] Janis L. Abkowitz,[2] Mary-Claire King,[3] and Akiko Shimamura[1,9,10]

M.Y.Z. and S.B.K. contributed equally to this work.

[1]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA; [2]Department of Medicine, Division of Hematology, University of Washington, Seattle, WA; [3]Department of Medicine and Department of Genome Sciences, University of Washington, Seattle, WA; [4]Department of Laboratory Medicine, University of Washington, Seattle, WA; [5]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA; [6]Division of Hematology/Oncology, Boston Children's Hospital, Dana Farber Cancer Institute, and Harvard Medical School, Boston, MA; [7]Harvard Stem Cell Institute, Boston, MA; [8]Department of Pathology, Boston Children's Hospital, MA; [9]Department of Pediatric Hematology/Oncology, Seattle Children's Hospital, WA; [10]Department of Pediatrics, University of Washington, Seattle, WA, USA

 *doi:10.3324/haematol.2014.113456*
*Manuscript received on July 22, 2014. Manuscript accepted on September 15, 2014.*
*Correspondence: ashimamu@fhcrc.org*

## Supplementary Methods

**Genomics.** Libraries were prepared in 96-well format with a Bravo liquid handling robot (Agilent Technologies). One to two micrograms of genomic DNA were sheared to a peak size of 150 bp using a Covaris E series instrument. DNA was end-repaired, A-tailed, and ligated to Illumina adapters. Libraries were amplified for 5 cycles with flanking primers (PE 1.0 and PE 2.0), and 500 nanograms of purified amplified library were then hybridized to the custom biotinylated cRNA oligonucleotides (Agilent SureSelect Target Enrichment system) for 24 hours at $65^0$C. Following a series of washes with increasing stringency, the biotin cRNA-DNA hybrids were purified with streptavidin conjugated magnetic beads. Post-capture PCR was performed with 96 different 6 bp barcoded primers for 14 cycles, and the final concentration of captured library was determined with a TapeStation (Agilent Technologies).

Forty-eight barcoded libraries were pooled in equimolar amount to carry out cluster amplification. Sequencing was performed on an Illumina HiSeq 2500 with a standard paired end 2x101 bp recipe in rapid mode. Yield of Q30 base pairs ranged from 31 to 33 GB.

An average per patient of 291 ± 26 (mean ± SD) variants with $\geq$8X coverage and confirmatory reads from both minus and plus strands were detected. Variants were excluded if present in <5% of reads at the site, if the minor allele frequency was >1% on dbSNP138, the Exome Variant Server or 1000Genomes, or if present in >50% of samples; 23 ± 5 variants per sample remained.

Custom bioinformatic analysis was designed for *SBDS* given the complexities due to gene conversion from its pseudogene, *SBDSP*.[1] Reads were aligned to the *SBDS* gene in isolation, and variant allele fractions of pseudogene-derived variants were analyzed to detect the presence of gene conversion.

Consequences of splice site mutations for transcription were evaluated by subcloning and Sanger sequencing cDNA generated from patients' cells. Compound heterozygous variants were confirmed to be in *trans* by genotyping DNA from parents, or by subcloning and Sanger sequencing genomic DNA or cDNA.

**Fanconi anemia pathway analysis.** Bone marrow fibroblasts were infected with a pMMP retrovirus expressing a wild-type FANCA cDNA or empty vector control. Immunoblotting for FANCD2 monoubiquitination was performed as previously described.[2]

## Supplementary Figure Legends

**Figure S1: Evolutionary and structural analysis of mutations in *GATA2*.**

(A) *GATA2* mutations identified in this study disrupt residues within an LWRR motif conserved across both zinc finger domains of the human GATA family. Missense and nonsense mutations reported in this study are indicated by red and black asterisks, respectively. Amino acid numbers correspond to the second zinc finger of GATA2. (B) Solution structure of murine GATA1 ZF1 (PDB ID: 1GNF).[3] The residue orthologous to human GATA2 Trp360 (highlighted in red with surrounding residues rendered semi-transparent) is largely buried within the GATA zinc finger.

## Table S1. Known mutations used for blinded validation of MarrowSeq

| Mutation class | Study Code | Diagnosis/Phenotype | Gene | Zygosity | Mutation[a] | Effect |
|---|---|---|---|---|---|---|
| Copy number variant | FH-142 | Thrombocytopenia | *RUNX1* | Heterozygous | Whole gene deletion | Haploinsufficiency |
| | FH-16 | Diamond-Blackfan anemia | *RPS17* | Heterozygous | Whole gene deletion | Haploinsufficiency |
| Insertion-deletion | FH-6 | Dyskeratosis congenita | *DKC1* | Heterozygous | c.1512_1514dupGAA | p.K505dup |
| | FH-11 | Diamond-Blackfan anemia | *RPL5* | Heterozygous | c.169_172delAACA | p.N57Qfs*12 |
| Single nucleotide variant – exonic | FH-28 | Diamond-Blackfan anemia | *RPS19* | Heterozygous | c.301C>T | p.R101C |
| | FH-43 | Dyskeratosis congenita (AD) | *TERT* | Heterozygous | c.2266C>T | p.R756C |
| | FH-97 | Aplastic anemia | *GATA2* | Heterozygous | c.1061C>T | p.T354M |
| | FH-130 | Dyskeratosis congenita | *TERT* | Compound heterozygous | c.1760T>C | p.I587T |
| | | | | | c.2110C>T | p.P704S |
| | FH-138 | Familial platelet disorder with predisposition to myeloid malignancy | *RUNX1* | Heterozygous | c.861C>A | p.Y287* |
| | FH-240 | Pre-B cell ALL predisposition | *PAX5* | Heterozygous | c.547G>A | p.G183S |
| | FH-158 | Shwachman-Diamond syndrome | *SBDS* | Compound heterozygous | c.183_184TA>CT | p.K62* |
| | | | | | c.258+2T>C | p.C84Yfs*4 |
| Single-nucleotide variant – intronic splice site | FH-30 | Shwachman-Diamond syndrome | *SBDS* | Homozygous | c.258+2T>C | p.C84Yfs*4 |
| | CH-130 | Shwachman-Diamond syndrome | *SBDS* | Compound heterozygous | c.258+2T>C | p.C84Yfs*4 |
| | | | | | c.653G>A | p.R218Q |
| Single nucleotide variant – promoter | FH-31 | Dyskeratosis congenita | *DKC1* | Hemizygous | c.-141C>G | Promoter mutation |

[a]RefSeq IDs: *RUNX1* NM_001754.4; *RPS17* NM_001021.3; *DKC1* NM_001363.4; *RPL5* NM_000969.3; *RPS19* NM_001022.3; *TERT* NM_198253.2; *GATA2* NM_032638.4; *PAX5* NM_016734.2; *SBDS* NM_016038.2

Asterisk denotes a protein truncation resulting from the gene mutation.

## Table S2. Damaging mutations identified in patients with Fanconi Anemia

| Study code | Clinical complementation subtype | Gene | Mutation[a] | Effect |
|---|---|---|---|---|
| FH-9 | FA-A | *FANCA* | c.793_3C>G | p.T266Sfs*17 |
| | | | c.1741dupT | p.S581Ffs*18 |
| FH-73 | FA-A | *FANCA* | c.1360_2014dup | p.V672Gfs*31 |
| | | | c.2555C>A | p.S852* |
| FH-124 | FA-A | *FANCA* | c.1462_1535del | p.L927Afs*10 |
| | | | c.3338A>T, c.3316G>A | p.N1113I, p.E1106K |
| FH-241 | FA-A | *FANCA* | c.1535C>G | p.S512* |
| | | | c.2853-15_2856del | Splice site mutation |
| FH-42 | Non-A,C,G | *FANCD2* | c.2048T>C | p.L683P |
| | | | c.2715+1G>A | p.E906Ifs*4 |
| FH-3 | Non-A,C,G | *FANCA* | c.652T>C | p.C218R |
| | | | c.793_2014del | p.V265Lfs*8 |

[a]RefSeq IDs: *FANCA* NM_000135.2; *FANCD2* NM_033084.3
Asterisk denotes a protein truncation resulting from the gene mutation.

## Table S3. Clinical features of patients with idiopathic BMF/MDS

| Study code | Gender | Age (years) | Diagnosis | Family history | Congenital anomalies | Short telomeres* | Causal gene identified by MarrowSeq |
|---|---|---|---|---|---|---|---|
| **CH-103** | M | 10 | Marrow Failure, MDS | - | + | NA | *LIG4* |
| **CH-110** | M | 10 | Marrow Failure | - | - | NA | - |
| **CH-119** | M | 12 | Marrow Failure, MDS | + | - | NA | *GATA2* |
| **CH-140** | M | 10 | Marrow Failure | - | - | NA | - |
| **CH-144** | M | 1 | Marrow Failure | - | - | NA | - |
| **FH-8** | M | 7 | Marrow Failure | - | - | - | - |
| **FH-10** | M | 8 | Marrow Failure | - | + | - | - |
| **FH-13** | F | 1 | Marrow Failure, MDS | - | + | + | - |
| **FH-14** | F | 21 | Marrow Failure | + | + | + | - |
| **FH-15** | F | 9 | Marrow Failure | - | + | NA | - |
| **FH-20** | M | 1 | Marrow Failure | + | - | - | - |
| **FH-21** | F | 2 | Marrow Failure | - | + | + | - |
| **FH-22** | M | 3 | Marrow failure | - | - | NA | - |
| **FH-23** | F | 17 | Marrow Failure | - | - | NA | - |
| **FH-24** | F | 1 | Marrow Failure | + | + | NA | - |
| **FH-27** | M | 14 | Marrow Failure | + | + | - | - |
| **FH-45** | M | 2 | Marrow Failure | - | - | - | - |
| **FH-50** | M | 12 | Marrow Failure | + | - | + | - |
| **FH-53** | F | 24 | Marrow Failure | Unknown (Adopted) | - | - | - |
| **FH-54** | M | 10 | Marrow Failure | - | + | NA | - |
| **FH-55** | F | 7 | Marrow Failure | + | - | - | - |
| **FH-58** | M | 18 | Marrow Failure | + | + | - | - |
| **FH-60** | F | 1 | Marrow Failure | + | - | + | - |
| **FH-64** | M | 4 | Marrow Failure | + | + | NA | - |
| **FH-65** | M | 3 | Marrow Failure | - | - | - | - |
| **FH-66** | M | 4 | Marrow Failure | + | - | - | - |
| **FH-67** | F | 7 | MDS | + | - | - | - |
| **FH-68** | M | 9 | Marrow Failure | Unknown (adopted) | - | + | - |
| **FH-69** | F | 16 | Marrow Failure | - | - | NA | - |
| **FH-70** | M | 15 | Marrow Failure | + | + | - | *DKC1* |
| **FH-75** | F | 9 | Marrow Failure | - | - | - | - |
| **FH-78** | F | 13 | MDS, RAEB | + | - | - | - |
| **FH-79** | F | 57 | Marrow Failure | + | + | - | - |
| **FH-82** | F | 16 | Marrow Failure | - | + | - | *GATA2* |
| **FH-93** | F | 16 | Marrow Failure | - | - | + | - |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **FH-98** | M | 23 | Marrow Failure with trisomy 8 | - | - | + | - |
| **FH-105** | F | 31 | Marrow Failure | - | - | - | - |
| **FH-106** | M | 5 | Marrow Failure | + | - | NA | - |
| **FH-109** | M | 23 | Marrow Failure | - | - | NA | - |
| **FH-113** | F | 64 | Marrow Failure | + | - | NA | - |
| **FH-120** | F | 14 | Marrow Failure | + | - | - | - |
| **FH-121** | F | 17 | Marrow Failure | - | + | - | - |
| **FH-125** | M | 8 | Marrow Failure | - | + | - | - |
| **FH-126** | M | 11 | Marrow Failure | + | + | - | - |
| **FH-129** | M | 14 | Marrow Failure | - | + | + | - |
| **FH-134** | M | 4 | Marrow Failure | - | + | NA | - |
| **FH-141** | M | 10 | Marrow Failure | - | - | NA | - |
| **FH-145** | F | 1 | Marrow Failure | + | + | - | - |
| **FH-147** | M | 37 | Marrow Failure | - | - | - | - |
| **FH-149** | F | 3 | Marrow Failure | + | + | - | - |
| **FH-153** | M | 17 | Marrow Failure | + | + | NA | - |
| **FH-154** | F | 17 | Marrow Failure | - | + | - | *GATA2* |
| **FH-156** | F | 2 | MDS | - | - | NA | - |
| **FH-159** | F | 5 | Marrow Failure | + | + | - | - |
| **FH-164** | F | 11 | Marrow Failure | - | - | - | - |
| **FH-165** | F | 7 | Marrow Failure | - | - | - | - |
| **FH-166** | M | 2 | Marrow Failure | + | + | - | - |
| **FH-169** | M | 1 | Marrow Failure | - | - | - | - |
| **FH-178** | M | 12 | MDS, 5q- | Unknown (adopted) | + | - | *RUNX1* |
| **FH-179** | F | 18 | Marrow Failure | + | - | - | - |
| **FH-180** | F | 2 | Marrow Failure | - | + | NA | - |
| **FH-181** | F | 22 | Marrow Failure | + | - | NA | *GATA2* |
| **FH-182** | M | 67 | Marrow Failure | + | - | - | - |
| **FH-184** | M | 6 | Marrow Failure | + | + | NA | - |
| **FH-185** | M | 8 | Marrow Failure | + | + | - | - |
| **FH-186** | F | 6 | Marrow Failure | - | + | NA | - |
| **FH-187** | F | 20 | MDS | Unknown | - | - | - |
| **FH-189** | F | 32 | Marrow Failure | - | - | - | - |
| **FH-190** | M | 24 | Marrow Failure | + | - | + | - |
| **FH-202** | F | 12 | Marrow Failure | + | - | - | *GATA2* |
| **UW-8** | M | 18 | Marrow Failure | + | - | + | - |

F, female; M; male; MDS, myelodysplastic syndrome; NA, not available

## Table S4. Damaging mutations and variants of unknown clinical significance (VUS) at highly conserved sites in patients with previously unclassified BMF/MDS

| Variant type[a] | Study code | Sex[b] | DNA source[c] | Gene | RefSeq Transcript | Mutation | Effect | Total reads | Variant reads |
|---|---|---|---|---|---|---|---|---|---|
| Damaging | CH-103 | M | MF | *LIG4* | NM_002312.3 | c.2440C>T[d] | p.R814* | 523 | 247 |
| Damaging | CH-103 | M | MF | *LIG4* | NM_002312.3 | c.1751_1755delTAAGA[d] | p.I584Rfs*2 | 541 | 244 |
| VUS | CH-119 | M | MF | *KIT* | NM_000222.2 | c.1889A>G | p.H630R | 394 | 176 |
| VUS | CH-119 | M | MF | *SRP72* | NM_006947.3 | c.1448C>A | p.T483N | 342 | 177 |
| VUS | CH-119 | M | MF | *BCR* | NM_004327.3 | c.1889C>G | p.A630G | 491 | 222 |
| Damaging | CH-119 | M | MF | *GATA2* | NM_032638.4 | c.1078T>A | p.W360R | 317 | 153 |
| VUS | FH-68 | M | PB | *TINF2* | NM_012461.2 | c.734C>A[e] | p.S245Y | 498 | 498 |
| Damaging | FH-70 | M | MF | *DKC1* | NM_001363.4 | c.-141C>G | Promoter mutation | 92 | 92 |
| Damaging | FH-82 | F | LCL, MF | *GATA2* | NM_032638.4 | c.1084C>T | p.R362* | 354 | 182 |
| VUS | FH-105 | F | PB | *ETV6* | NM_001987.4 | c.380G>A | p.R127Q | 722 | 307 |
| VUS | FH-134 | M | LCL | *DNMT3A* | NM_022552.4 | c.347C>G | p.A116G | 299 | 129 |
| VUS | FH-147 | M | PB | *KMT2A* | NM_001197104.1 | c.11770G>T | p.D3924Y | 647 | 291 |
| VUS | FH-147 | M | PB | *U2AF1* | NM_001025203.1 | c.470A>G | p.Q157R | 607 | 261 |
| VUS | FH-147 | M | PB | *U2AF1* | NM_001025203.1 | c.101C>T | p.S34F | 648 | 283 |
| VUS | FH-149 | F | PB | *GFI1* | NM_005263.3 | c.49C>G | p.Q17E | 532 | 266 |
| VUS | FH-154 | F | BM | *ANKRD26* | NM_014915.2 | c.3899A>G | p.K1300R | 195 | 81 |
| VUS | FH-154 | F | BM | *CBL* | NM_005188.3 | c.522T>G | p.F174L | 542 | 269 |
| Damaging | FH-154 | F | BM, MF | *GATA2* | NM_032638.4 | c.1084C>T | p.R362* | 361 | 178 |
| VUS | FH-165 | F | PB | *CBL* | NM_005188.3 | c.12C>A | p.N4K | 219 | 96 |
| VUS | FH-166 | M | PB | *PRPF40B* | NM_001031698.2 | c.1649G>A | p.R550H | 407 | 189 |
| VUS | FH-169 | M | BM | *BCOR* | NM_001123385.1 | c.3308A>C | p.E1103A | 167 | 167 |
| Damaging | FH-178 | M | PB, MF | *RUNX1* | NM_001754.4 | c.567C>G | p.Y189* | 547 | 286 |
| VUS | FH-181 | F | PB | *RARA* | NM_000964.3 | c.743G>A | p.G248D | 380 | 190 |
| Damaging | FH-181 | F | PB | *GATA2* | NM_032638.4 | c.988C>T | p.R330* | 195 | 79 |
| VUS | FH-186 | M | PB | *KMT2A* | NM_001197104.1 | c.9400C>T | p.L3134F | 659 | 306 |
| Damaging | FH-202 | F | PB, MF | *GATA2* | NM_032638.4 | c.1082G>A | p.R361H | 130 | 62 |

[a]VUS, variant of unknown significance
[b]F, female; M, male;
[c]BM, bone marrow; LCL, lymphoblast cell line; MF, marrow fibroblasts; PB, peripheral blood
[d]These two mutations were experimentally determined to be in *trans*.
[e]A patient with aplastic anemia has previously been reported to harbor this variant as a heterozygote.[4] This variant is present with an allele frequency of 0.05% in 1000Genomes and 0.0326% in the Exome Variant Server. The pathogenicity of this variant as a homozygote is unknown.
Asterisk denotes a protein truncation resulting from the gene mutation.

## Table S5. Damaging carrier mutations and carrier variants of unknown clinical significance at highly conserved sites in patients with previously unclassified BMF/MDS

| Variant type[a] | Study code | Sex[b] | DNA source[c] | Gene | RefSeq Transcript | Mutation | Effect | Total reads | Variant reads |
|---|---|---|---|---|---|---|---|---|---|
| Damaging | FH-22 | M | LCL | *FANCL* | NM_018062.3 | c.1096_1099dupATTA | p.T367Nfs*13 | 574 | 241 |
| VUS | FH-50 | M | MF | *FANCL* | NM_018062.3 | c.1007_1009delTAT | p.I336_C337delinsS | 611 | 286 |
| VUS | FH-55 | F | LCL | *FANCM* | NM_020937.2 | c.4931G>A | p.R1644Q | 463 | 232 |
| VUS | FH-65 | M | MF | *FANCI* | NM_001113378.1 | c.1813C>T | p.L605F | 466 | 238 |
| Damaging | FH-66 | M | MF | *SBDS* | NM_016038.2 | c.258+2T>C | p.C84fs*3 | 742 | 304 |
| Damaging | FH-69 | F | MF | *AK2* | NM_001625.3 | c.219+5G>A | Splice site mutation | 456 | 195 |
| Damaging | FH-98 | M | LCL | *MPL* | NM_005373.2 | c.744_747dupTGGC | p.N250Wfs*13 | 382 | 151 |
| Damaging | FH-113 | F | BM | *PALB2* | NM_024675.3 | c.1163dupC | p.L389Sfs*12 | 489 | 236 |
| Damaging | FH-121 | F | BM | *CTC1* | NM_025099.5 | c.2452C>T | p.R818* | 396 | 193 |
| VUS | FH-154 | F | BM | *FANCI* | NM_001113378.1 | c.1813C>T | p.L605F | 410 | 187 |
| VUS | FH-156 | F | BM | *FANCI* | NM_001113378.1 | c.1264G>A | p.G422R | 718 | 340 |
| VUS | FH-180 | F | PB | *AK2* | NM_001625.3 | c.224G>T | p.S75I | 459 | 222 |
| VUS | FH-186 | M | PB | *AK2* | NM_001625.3 | c.49C>G | p.R17G | 122 | 63 |
| VUS | UW-8 | M | LCL | *NBN* | NM_002485.4 | c.706A>G | p.K236E | 223 | 101 |

[a]VUS, variant of unknown significance
[b]M, male; F, female
[c]BM, bone marrow; LCL, lymphoblast cell line; MF, marrow fibroblasts; PB, peripheral blood
Asterisk denotes a protein truncation resulting from the gene mutation.

## References


1. Boocock GRB, Morrison JA, Popovic M, Richards N, Ellis L, Durie PR, et al. Mutations in *SBDS* are associated with Shwachman-Diamond syndrome. Nat Genet. 2003;33(1):97–101.

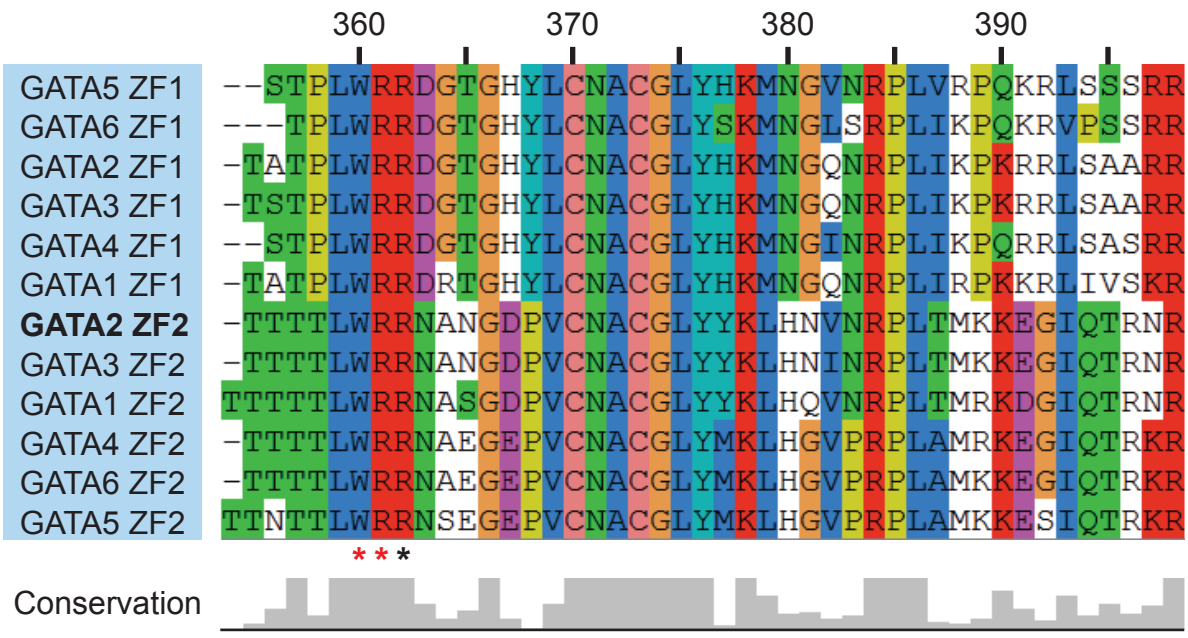
2. Shimamura A, Oca RM de, Svenson JL, Haining N, Moreau LA, Nathan DG, et al. A novel diagnostic screen for defects in the Fanconi anemia pathway. Blood. 2002;100(13):4649–54.

3. Kowalski K, Czolij R, King GF, Crossley M, Mackay JP. The solution structure of the N-terminal zinc finger of GATA-1 reveals a specific binding face for the transcriptional co-factor FOG. J Biomol NMR. 1999;13(3):249–62.

4. Walne AJ, Vulliamy T, Beswick R, Kirwan M, Dokal I. *TINF2* mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. Blood. 2008;112(9):3594–600.

# Figure S1

## A


360
370
380
390
GATA5 ZF1 --STPLWRRDGTGHYLCNACGLYHKMNGVNRPLVRPQKRLSSSRR
GATA6 ZF1 ---TPLWRRDGTGHYLCNACGLYSKMNGLSRPLIKPQKRVPSSRR
GATA2 ZF1 -TATPLWRRDGTGHYLCNACGLYHKMNGQNRPLIKPKRRLSAARR
GATA3 ZF1 -TSTPLWRRDGTGHYLCNACGLYHKMNGQNRPLIKPKRRLSAARR
GATA4 ZF1 --STPLWRRDGTGHYLCNACGLYHKMNGINRPLIKPQRRLSASRR
GATA1 ZF1 -TATPLWRRDRTGHYLCNACGLYHKMNGQNRPLIRPKKRLIVSKR
GATA2 ZF2 -TTTTLWRRNANGDPVCNACGLYYKLHNVNRPLTMKKEGIQTRNR
GATA3 ZF2 -TTTTLWRRNANGDPVCNACGLYYKLHNINRPLTMKKEGIQTRNR
GATA1 ZF2 TTTTTLWRRNASGDPVCNACGLYYKLHQVNRPLTMRKDGIQTRNR
GATA4 ZF2 -TTTTLWRRNAEGEPVCNACGLYMKLHGVPRPLAMRKEGIQTRKR
GATA6 ZF2 -TTTTLWRRNAEGEPVCNACGLYMKLHGVPRPLAMKKEGIQTRKR
GATA5 ZF2 TTNTTLWRRNSEGEPVCNACGLYMKLHGVPRPLAMKKESIQTRKR
* * *
Conservation

## B