Supplementary information

Materials and Methods

Cloning, protein expression, and purification

The DNA fragments encoding Znf4-6 (residues 731-835), Znf4-5 (residues 731-797), and Znf5-6 (residues 765-835) of B-cell lymphoma/leukemia 11A (BCL11A) were amplified via PCR from the HeLa cell cDNA library and cloned into a modified pGEX-4T1 vector (GE Healthcare), in which the thrombin protease site was substituted for a tobacco etch virus (TEV) cleavage site. All variants of Znf4-6 were generated using a MutanBEST kit (Takara) and verified via DNA sequencing.

All proteins were expressed in *Escherichia coli* BL21 (DE3) cells (Novagen). The cells were cultured at 37 °C in LeMaster and Richards minimal medium (LR medium), supplemented with 100 μM $ZnSO_4$, until the OD_{600} reached ~1.2. Next, the proteins were induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 16 °C for 24 h. The cells were harvested and lysed using a high-pressure cell disrupter in buffer A (20 mM Tris-HCl, pH 7.5, 1 M NaCl). After centrifugation, the supernatant was incubated with glutathione Sepharose (GE Healthcare) pre-equilibrated with buffer A, and the GST-tagged proteins were eluted by buffer A supplemented with 30 mM reduced glutathione. The GST tag was removed by treating the eluted protein with TEV protease in buffer B (20 mM Tris-HCl, pH 7.5, 500 mM NaCl). The cleaved protein was further purified via size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column (GE Healthcare) in buffer A. The purified proteins were dialyzed with buffer

C (20 mM Tris-HCl, pH 7.5 and 150 mM NaCl) and concentrated for subsequent experiments.

DNA synthesis and annealing

The single-stranded oligonucleotides were synthesized by General Biosystems and dissolved to 100 μM with buffer C. Next, the two complementary single-stranded oligonucleotides were mixed at equimolar amounts in a microtube, followed by incubation at 95 °C for 5 min and slowly cooling to room temperature in a water bath. The double-stranded DNA was diluted to the target concentration and used for isothermal titration calorimetry (ITC) experiments. All DNA oligonucleotides used for crystallization or ITC assays are displayed in Supplementary information, Table S1.

Crystallization, data collection, and structure determination

The Znf4-6-DNA complex was prepared by mixing the Znf4-6 protein with the 12-bp double-stranded oligo with 5′-overhang at a 1:1.2 ratio, followed by incubation at 4 °C for 6 h to form a stable protein-DNA complex. The complex was further purified via size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column in buffer C and concentrated to ~1.10 mM. The crystals were grown at 293 K via the sitting-drop vapor diffusion method by mixing 1 μL of the protein-DNA complex and 1 μL of reservoir buffer (0.2 M calcium acetate hydrate, 0.1 M sodium cacodylate trihydrate (pH 6.5), 18% w/v polyethylene glycol 8000). The crystals were cryoprotected via soaking in the mother liquor supplemented with 25% (v/v) glycerol before being flash-frozen in liquid

nitrogen.

The crystallographic data set was collected on beamline 18U1 at the Shanghai Synchrotron Radiation Facility (SSRF) at a wavelength of 0.979 Å. The diffraction data set was processed and scaled using the *HKL*-2000 program suite.[1] Experimental phasing was performed using the *AutoSol* program in the *PHENIX* suite [2, 3], and the zinc atom sites were found by the SHELX C/D program. The initial protein model was built using the Buccaneer program in the *CCP4* suite.[4] Next, the DNA model was further built using COOT,[5] and refinement was performed using the *PHENIX.Refine* program. 2Fo-Fc composite omit electron density maps were generated using PHENIX. composite_omit_map program. All structures in the figures were generated using PyMOL (DeLano Scientific LLC). The data collection and refinement statistics are listed in Supplementary information, Table S2.

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) assays were performed on a Microcal PEAQ-ITC instrument (Malvern) at 25 °C. The titration protocol consisted of a single initial injection of 1 μL of protein, followed by 19 injections of 2 μL of protein (concentration: 200-400 μM), into the sample cell containing double-stranded DNA (concentration: 15-25 μM). Thermodynamic data were analyzed with a single-site binding model using MicroCal PEAQ-ITC Analysis Software provided by the manufacturer and summarized in Supplementary information, Table S3.

References:

1 Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997; **276**:307-326.

2 Adams PD, Afonine PV, Bunkoczi G *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:213-221.

3 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr* 2007; **40**:658-674.

4 Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 2006; **62**:1002-1011.

5 Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:486-501.

Supplementary information, Figure S1

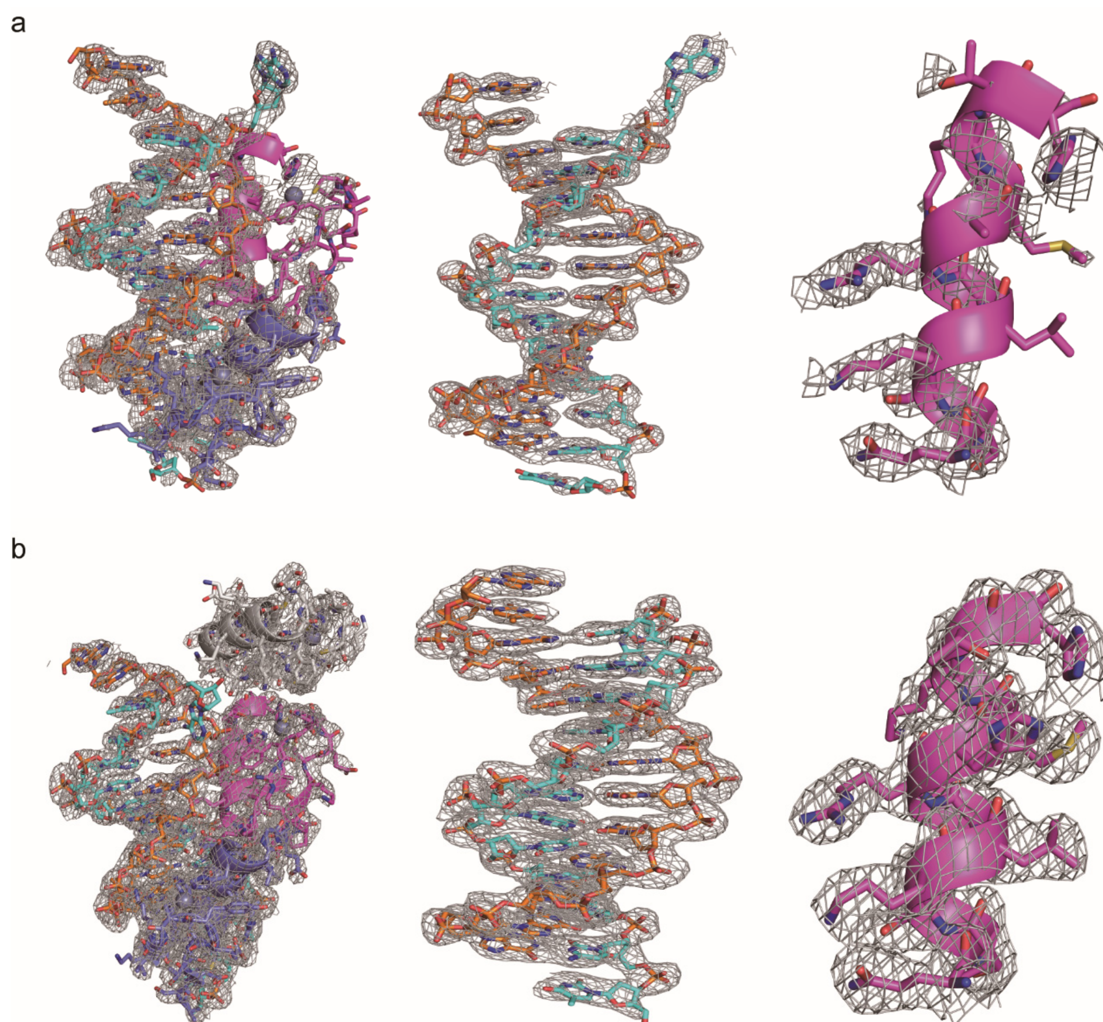

Fig. S1 2Fo-Fc composite omit electron density maps for the overall structure of both Chain A-DNA and Chain B-DNA complex, the corresponding DNA duplex, and the segments of α-helix of Znf5 of both Chain A and Chain B. **a** 2Fo-Fc composite omit electron density maps (contoured at 1σ level) of Chain B-DNA complex, the corresponding DNA duplex, and the segment of α-helix of Znf5 of Chain B, which are shown on left, middle, and right, respectively. **b** 2Fo-Fc composite omit electron density maps (contoured at 1σ level) of Chain A-DNA complex, the corresponding DNA duplex, and the segment of α-helix of Znf5 of Chain A, which are shown on left, middle, and right, respectively. Color codes of Znf4, Znf5 and Znf6 as well as the top strand and bottom strands are defined as described in **Fig. 1b.** The residues of both Chain A and Chain B are shown in cartoon and stick representation, and the corresponding DNA duplex are shown in stick representation. The water molecules in pdb file are not shown and zinc ions are shown as brown spheres.

Supplementary information, Figure S2

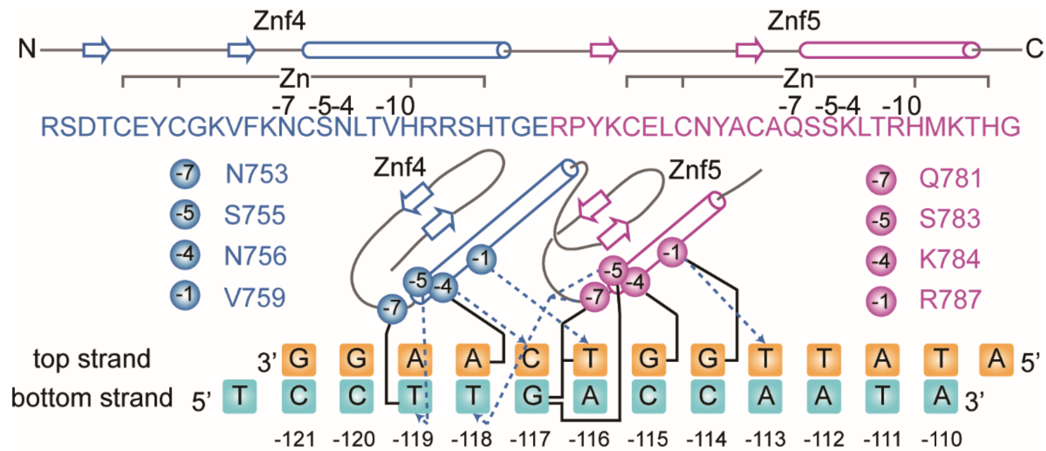

Fig. S2 Summary of the interactions between BCL11A and the γ-globin −115 HPFH region sequence. The amino acid sequence of Znf4-5 of BCL11A is shown with single-letter representation together with the secondary structure (top panel). Color codes of Znf4 and Znf5 as well as the top strand and bottom strands are defined as described in **Fig. 1b**. The residues at positions −1, −4, −5, and −7 of Znf4 and Znf5 are highlighted. The interaction details between Znf4-5 of BCL11A and double-stranded DNA are shown. The hydrogen-bonding interactions are shown as solid lines in black, and the van der Waals contacts are shown as dashed lines in blue.

Supplementary information, Figure S3

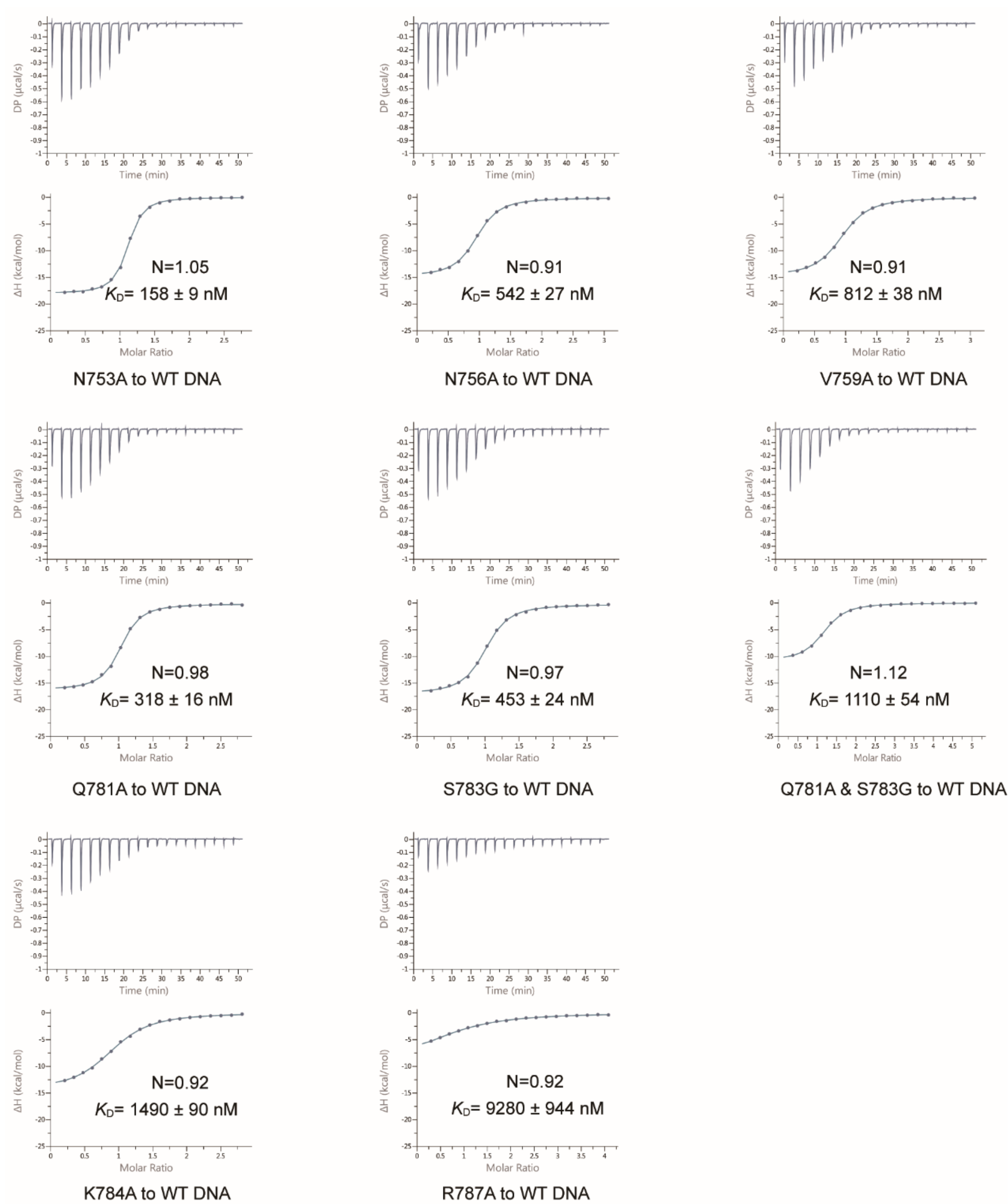

Fig. S3 The fitted curves of ITC experiments between mutant BCL11A and the wild-type γ-globin −115 HPFH region sequence are shown.

Supplementary information, Figure S4

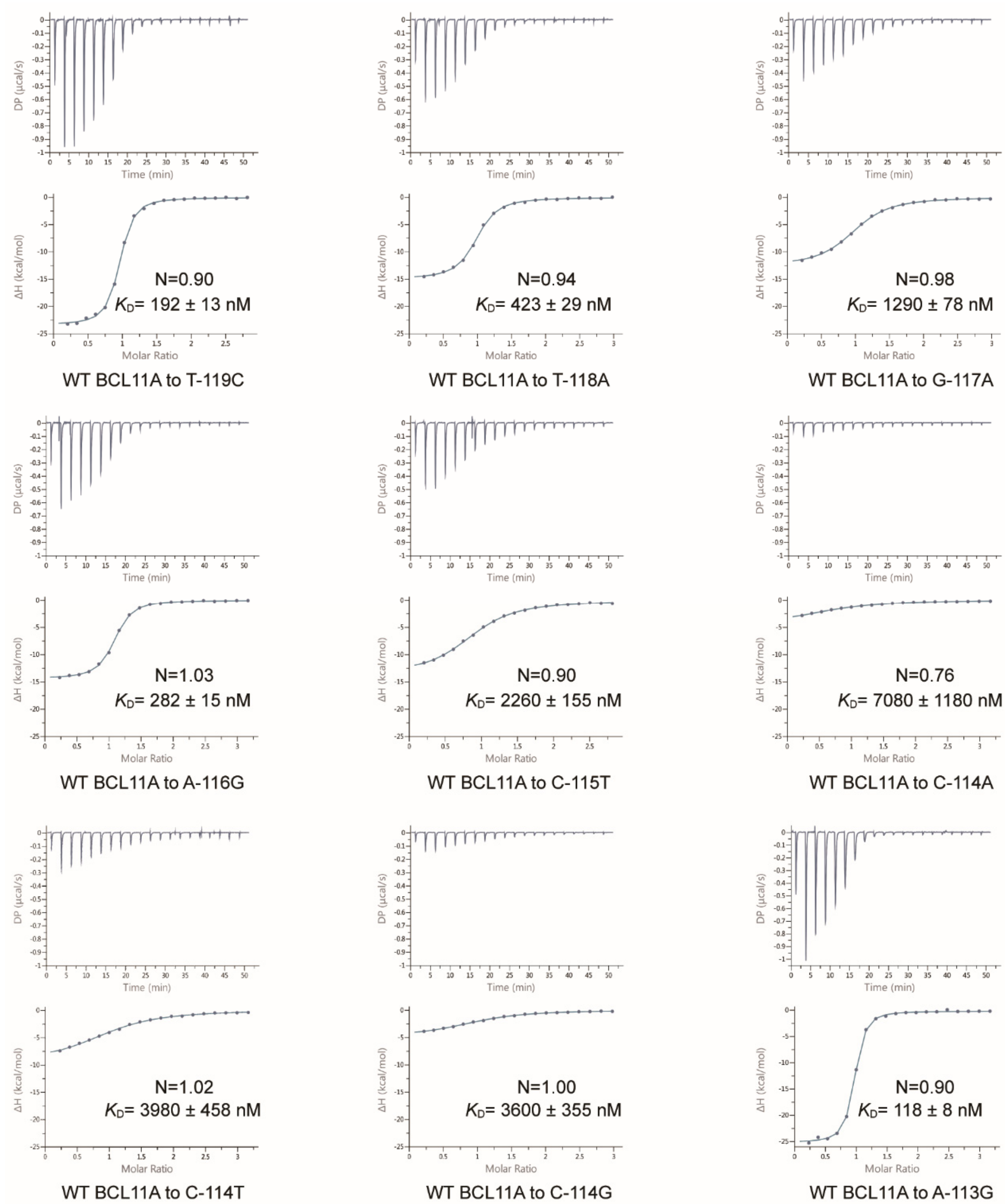

Fig. S4. The fitted curves of ITC experiments between wild-type BCL11A and the mutant γ-globin −115 HPFH region sequence are shown.

Supplementary information, Figure S5

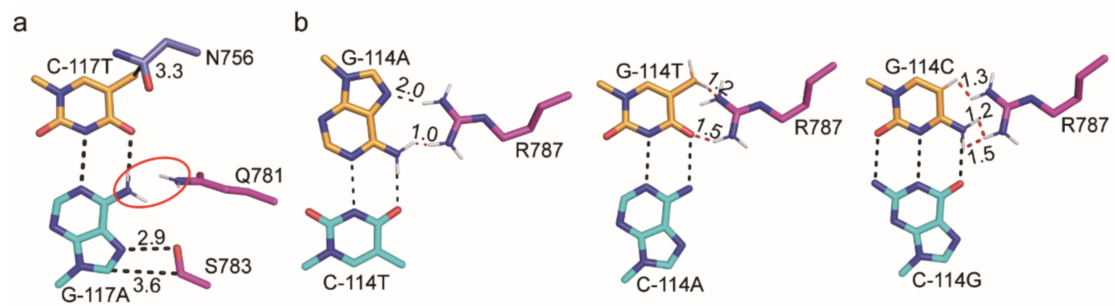

Fig. S5 a Model for the interactions of the G−117A mutant with Gln781 and Ser783. The steric clash is highlighted with red ovals. **b** Models for the interactions between G−114A/T/C mutants and Arg787. The steric clashes are marked with red dash lines. The model was generated from Chain A of the pdb file. The hydrogen atoms were added by PyMOL.

Supplementary information, Figure S6

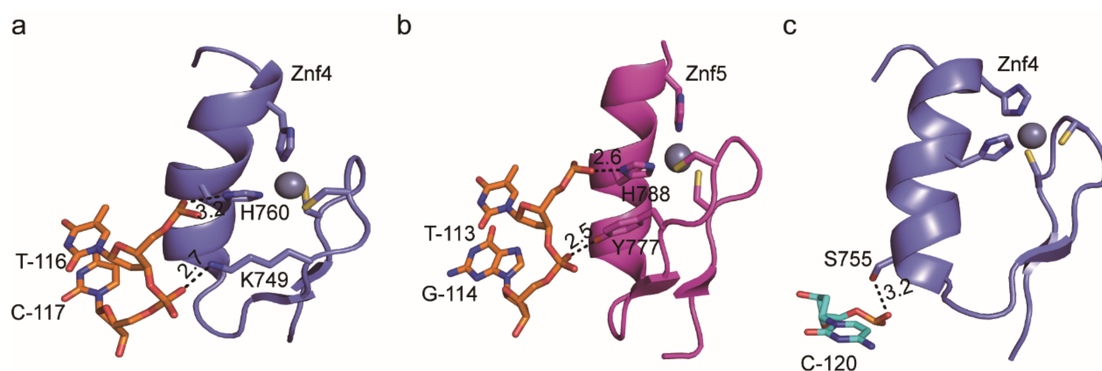

Fig. S6 a-c Summary of hydrogen-bonding contacts between Znf4-5 of BCL11A and the phosphate backbone. Color codes of Znf4 and Znf5 as well as the top and bottom strands are defined as described in **Fig. 1b**. The residues responsible for phosphate backbone interactions are presented in stick representation. Zinc ions are shown as brown spheres and coordinated via two cysteines and two histidines which are also presented in stick representation. The hydrogen-bonding interactions are depicted as black dashed lines with distances measured in Å.

Supplementary information, Table S1 Oligos used in this study

Oligonucleotides	Name	sequence(5'–3')
DNA oligos for Crystallization	TTGACCA-F(13bp)	TCCTTGACCAATA
	TTGACCA-R	ATATTGGTCAAGG
DNA oligos for Isothermal Titration Calorimetry assays	TTGACCA-F(WT)	CCTTGACCAATA
	TTGACCA-R	TATTGGTCAAGG
	CTGACCA-F(T−119C)	CCCTGACCAATA
	CTGACCA-R	TATTGGTCAGGG
	TAGACCA-F(T−118A)	CCTAGACCAATA
	TAGACCA-R	TATTGGTCTAGG
	TTAACCA-F(G−117A)	CCTTAACCAATA
	TTAACCA-R	TATTGGTTAAGG
	TTGGCCA-F(A−116G)	CCTTGGCCAATA
	TTGGCCA-R	TATTGGCCAAGG
	TTGATCA-F(C−115T)	CCTTGATCAATA
	TTGATCA-R	TATTGATCAAGG
	TTGACAA-F(C−114A)	CCTTGACAAATA
	TTGACAA-R	TATTTGTCAAGG
	TTGACGA-F(C−114G)	CCTTGACGAATA
	TTGACGA-R	TATTCGTCAAGG
	TTGACTA-F(C−114T)	CCTTGACTAATA
	TTGACTA-R	TATTAGTCAAGG
	TTGACCG-F(A−113G)	CCTTGACCGATA
	TTGACCG-R	TATCGGTCAAGG

Supplementary information, Table S2 Data collection and refinement statistics

	BCL11A-γ-globin −115 HPFH region
Wavelength(Å)	0.979
Space group	$P3_221$
Cell parameters	
a, b, c (Å)	59.56, 59.56, 213.37
α, β, γ (°)	90, 90, 120
Resolution(Å)	40.00−2.50(2.54−2.50)
R_{merge} (%)	8.3(99.4)
$CC_{1/2}$	0.982(0.749)
$I/\sigma I$	23.2(2.1)
Completeness (%)	99.9(99.9)
Average redundancy	8.7(7.8)
Refinement(F>0)	
No. reflections (overall)	15839
No. reflections (test set)	744
R_{work}/R_{free}(%)	20.67/25.14
Number of atoms	
Protein	1051
DNA	1054
ZN	5
Water	69
B factors (Å^2)	
Protein	52.53
DNA	70.87
ZN	55.38
Water	39.62
r.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.557
Ramachandran plot % residues	
Favored	96.85
Allowed	2.36
Outliers	0.79

[a] Values in parentheses are for highest-resolution shell.

Supplementary information, Table S3 ITC results. K_D, dissociation constant; DP, differential power; N, binding stoichiometry; ΔH, binding enthalpy; -TΔS, binding entropy. Each K_D value is presented as fitted value ± error.

Protein	DNA	ΔH (kJ/mol)	−TΔS (kJ/mol)	N	K_D (nM)
$BCL11A^{wt}$	WT	−23.7 ± 0.143	13.9	0.93	72 ± 6
$BCL11A^{wt}$	T−119C	−23.3 ± 0.169	14.1	0.90	192 ± 13
$BCL11A^{wt}$	T−118A	−14.9 ± 0.143	6.21	0.94	423 ± 29
$BCL11A^{wt}$	G−117A	−12.6 ± 0.169	4.61	0.98	1290 ± 78
$BCL11A^{wt}$	A−116G	−14.4 ± 0.093	5.44	1.03	282 ± 15
$BCL11A^{wt}$	C−115T	−13.6 ± 0.268	5.94	0.90	2260 ± 155
$BCL11A^{wt}$	C−114A	−5.04 ± 0.520	−1.99	0.76	7080 ± 1180
$BCL11A^{wt}$	C−114T	−9.7 ± 0.416	2.33	1.02	3980 ± 458
$BCL11A^{wt}$	C−114G	−4.98 ± 0.171	−2.45	1.00	3600 ± 355
$BCL11A^{wt}$	A−113G	−25.0 ± 0.159	15.5	0.90	118 ± 8
$BCL11A^{N753A}$	WT	−18.0 ± 0.102	8.70	1.05	158 ± 9
$BCL11A^{N756A}$	WT	−14.7 ± 0.122	6.19	0.91	542 ± 27
$BCL11A^{V759A}$	WT	−14.8 ± 0.136	6.48	0.91	812 ± 38
$BCL11A^{Q781A}$	WT	−16.1 ± 0.105	7.23	0.98	318 ± 16
$BCL11A^{S783G}$	WT	−16.8 ± 0.137	8.10	0.97	453 ± 24
$BCL11A^{K784A}$	WT	−14.5 ± 0.225	6.58	0.92	1490 ± 90
$BCL11A^{R787A}$	WT	−10.0 ± 0.696	3.14	0.92	9280 ± 944
$BCL11A^{731-797}$	WT	−15.6 ± 0.077	6.58	0.90	264 ± 10
$BCL11A^{765-835}$	WT	−8.97 ± 0.585	1.96	0.87	7380 ± 761