Supplementary Information

Warren *et al*, Heterozygous Variants of CLPB are a Frequent Cause of Severe Congenital Neutropenia

Contents:

Supplemental Table 1. Characteristics - SCNIR Sequencing Cohort

Characteristic	SCN (N = 56)	Cyclic (N = 29)
Age -- years*, Median (range)	16.0 (0.1 - 39.1)	22.0 (4.8 - 77.8)
Deceased -- no (%)	5 (8.9)	1 (3.4)
Female sex-- no (%)	24 (42.9)	14 (48.3)
G-CSF treatment -- no (%)	55 (98.2)	25 (86.2)
G-CSF Dose (mcg/kg/day), median (range)	4.89 (0.34 - 99.31)	1.38 (0.05 - 8.89)
No affected family member-no (%)	42 (75)	26 (89.7)
ELANE mutation -- no (%)	11 (19.6)	4 (13.8)

*Age at last contact

**G-CSF median among patients being treated.

Dose information on G-CSF not available on 1 SCN patient

Table S2. Characteristics of Patients with Cyclic Neutropenia and *CLPB* c.1882 C>T Variant of Uncertain Significance

Sample ID#	Diagnosis	Protein p.	cDNA c.	Genomic* g.	VAF	gnomAD	Gender	Age at Diagnosis (Years)	Pre-G-CSF ANC	Median G-CSF dose (mcg/kg/day)	Bone Marrow Biopsy	Splenomegaly (Y/N)	AML/MDS (Y/N)	Infections	Neurological	Cataracts	Other	Urine 3-MGA?
SCNIR-33	Cyclic Neutropenia	R628C	1882 C>T	Chr11:72293609 G>A	0.46	0.002014	F	10.6	0.50	0.13	Normal	No	No	Sinusitis, otitis, URI, UTI	Negative	Negative	mouth ulcers, gingivitis	N/A
SCNIR-81	Cyclic Neutropenia	R628C	1882 C>T	Chr11:72293609 G>A	0.4	0.002014	F	24.8	0.40	1.06	N/A	Yes	N/A	Colitis/enteritis, esophagitis	Negative	Negative	Mouth ulcers	no unusual organic acids found
Fr-0205	Cyclic Neutropenia	R628C	1882 C>T	Chr11:72293609 G>A	0.47	0.002014	M	0.08	<1	No	Maturation arrest	Yes	No	Esophagitis	Brain imaging abnormalities (calicifications)	Yes	Steatorrhea, steatosis, learning difficulties; hyperuricemia, growth retardation, kyphoscoliosis	N/A

SCNIR: Cases identified throught he SCNIR North America registry; Fr: Cases identified through the French SCN registry
URI: upper respiratory infection; UTI: urinary tract infection
N/A: Not available

Supplementary Table 3. ACMG Evidence for Pathogenicity of ATP Binding Pocket Variants

	Strong				Moderate						Supporting						
	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	**Total Evidence**	**Classification**
T388K		X			X	X							X			1 Strong, 2 Moderate, 1 Supporting	Likely Pathogenic
N496K			X		X	X				X			X			1 Strong, 3 Moderate, 1 Supporting	Pathogenic
E557K			X		X	X				X			X			1 Strong, 3 Moderate, 1 Supporting	Pathogenic
R561G			X		X	X				X			X			1 Strong, 3 Moderate, 1 Supporting	Pathogenic
R561Q		X	X		X	X							X			2 Strong, 2 Moderate, 1 Supporting	Pathogenic
R620C			X		X	X				X			X			1 Strong, 3 Moderate, 1 Supporting	Pathogenic

PS1: Same amino acid change in known pathogenic variant
PS2: De novo with both maternity and paternity confirmed in a patient with the disease and no family history
PS3: Functional studies (in vivo or in vitro) support a damaging effect
PS4: Prevalence of the variant in affected individuals is significantly increased compared with controls
PM1: Located in a mutational hot spot and/or critical functional domain (specifically provides as an example active site of an enzyme)
PM2: Absent from controls or at extremely low frequency if recessive
PM3: For recessive disorders, proven to be in *trans* with a known pathogenic variant
PM4: In-frame deletion/insertions in a nonrepeat region or stop-loss variant causes protein length change
PM5: Novel missense change at an amino acid residue where a different missense change is pathogenic
PM6: Assume *de novo* but maternity and paternity are not confirmed
PP1: Coseggregation in multiple affected family members in a gene known to cause the disease
PP2: Missense variant in a gene with a low rate of benign missense variation
PP3: Multiple lines of *in silico* evidence predict deleterious effect
PP4: Phenotype of the patient or family history highly suggestive of monogenic etiology
PP5: Reputable source reporting pathogenic variant, but evidence is not available for independent evaluation

Summarized from Richards *et al.*[14]

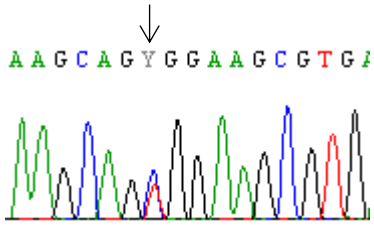
Supplementary Table 5. ACMG Evidence for Benign Classification of non-ATP Binding Pocket Variants

	Stand-alone	Strong				Supporting								
	BA1	BS1	BS2	BS3	BS4	BP1	BP2	BP3	BP4	BP5	BP6	BP7	**Total Evidence**	**Classification**
R327W			X	X									2 Strong	Benign
R603H			X	X									2 Strong	Benign

BA1: Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes, or ExAC
BS1: Allele frequency is higher than expected for the disorder
BS2: Observed in a healthy adult individual (if recessive, dominant, or X-linked) with full penetrance expected by early age
BS3: Functional studies (in vivo or in vitro) show no damaging effect
BS4: Lack of segregation in affected family members
BP2: Missense variant in a gene for which truncating variants primarily cause disease
BP3: In-frame deletions/insertions in a repetitive region that does not have a known function
BP4: Multiple *in silico* predictions suggest no impact
BP5: Variant found in a case with an alternative genetic basis for disease
BP6: Reputable source reporting benign variant, but evidence is not available for independent evaluation
BP7: Synonymous variant with *in silico* predictions to have no impact on splicing AND the nucleotide is not highly conserved

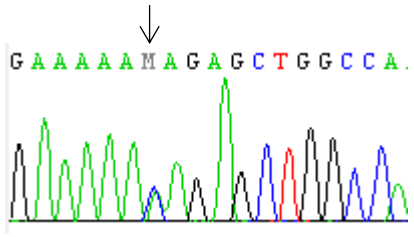
Summarized from Richards *et al.*[14]

Fr-1408



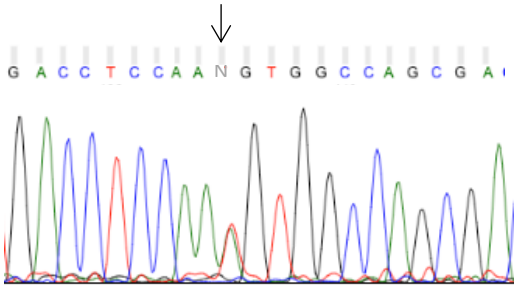
c.979C>T
p.R327W

Fr-0019



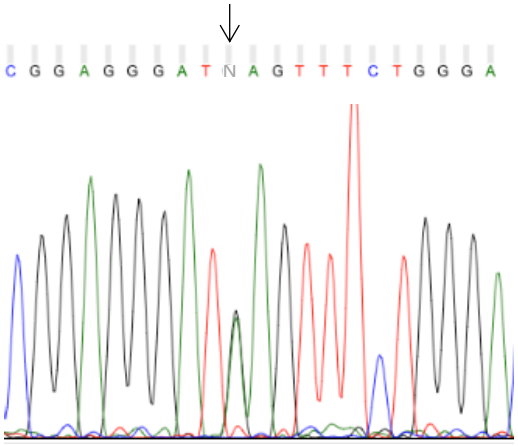
c.1163C>A
p.T388K

SCNIR-19



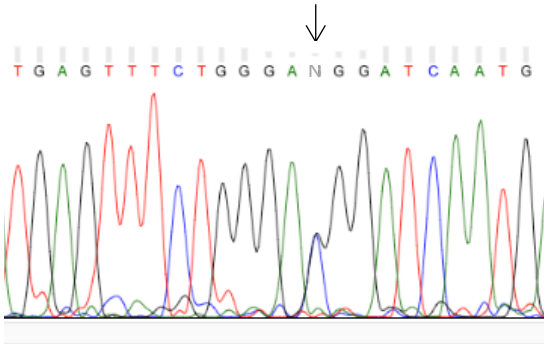
c.1488T>A
p.N496K

SCNIR-73



c.1669G>A
p.E557K

SCNIR-2

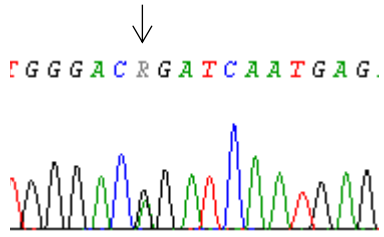


c.1681C>G
p.R561G

Supplementary Figure 1.

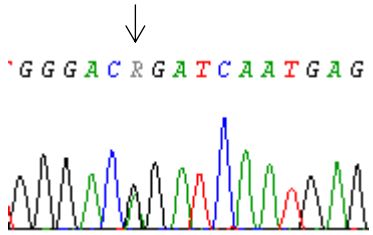
Sanger sequencing trace files of individuals identified through the SCNIR and French cohorts. Heterozygous *CLPB* variants indicated with an arrow.

Fr-0038



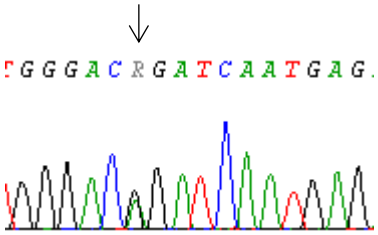
c.1682G>A
p.R561Q

Fr-1502



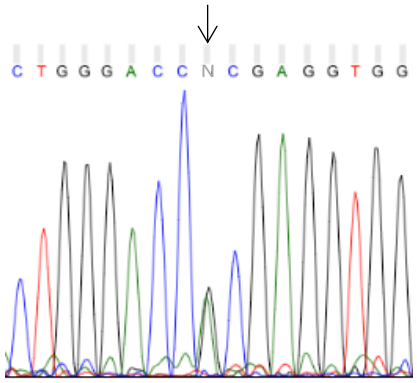
c.1682G>A
p.R561Q

Fr-0108



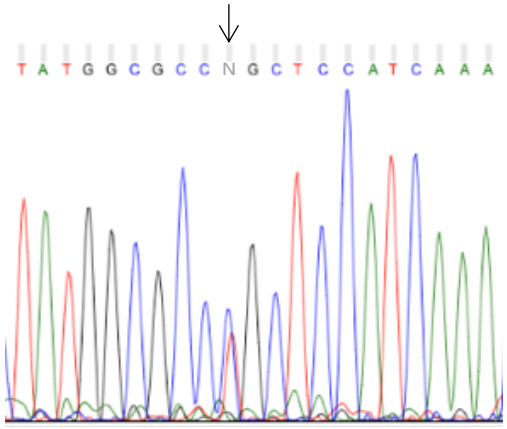
c.1682G>A
p.R561Q

SCNIR-45



c.1808G>A
p.R603H

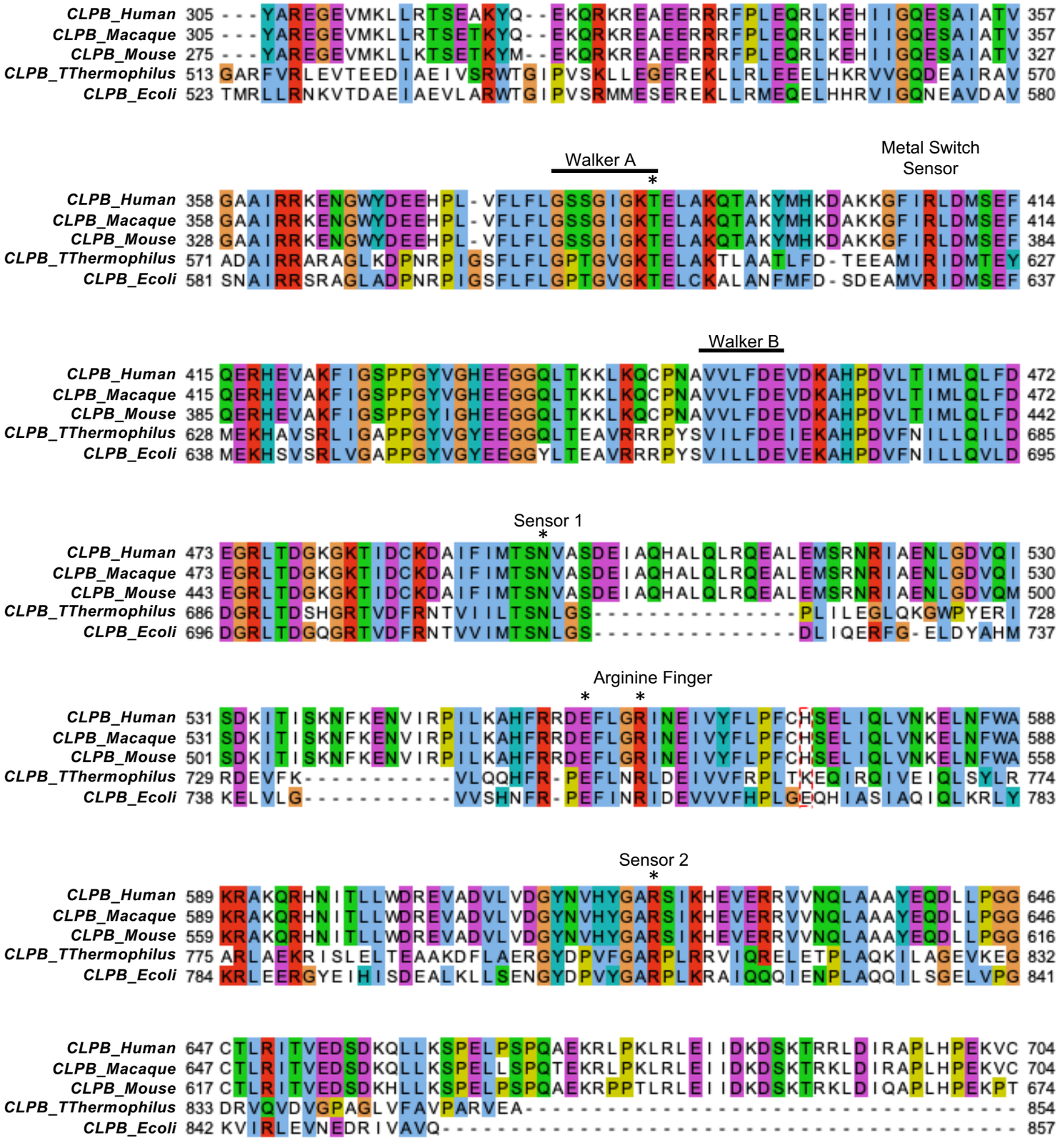
SCNIR-29



c.1858C>T
p.R620C

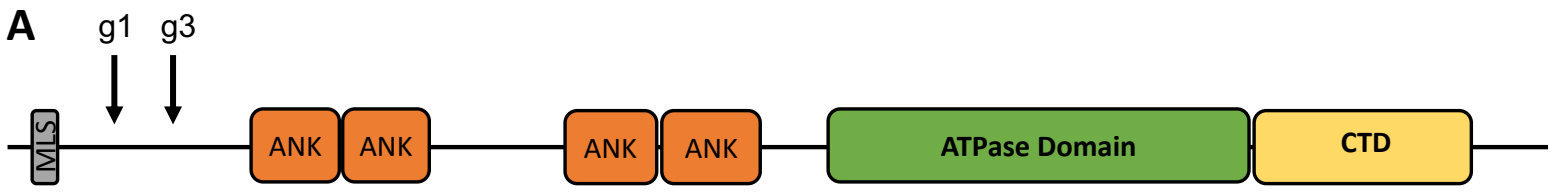
Supplementary Figure 1 (continued).

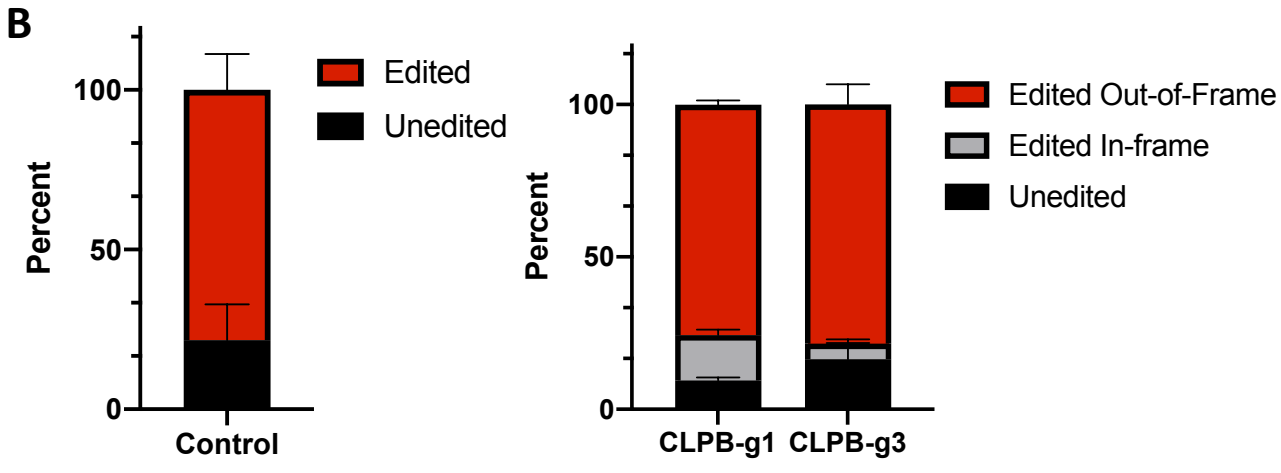
Sanger sequencing trace files of individuals identified through the SCNIR and French cohorts. Heterozygous *CLPB* variants indicated with an arrow.

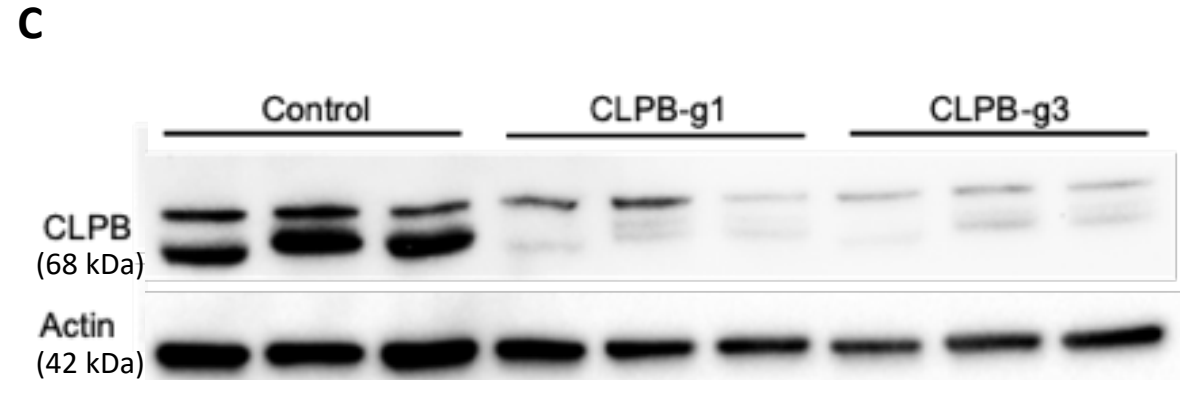


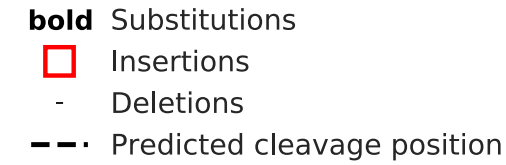
Supplementary Figure 2.

Amino acid sequence alignment generated using ClustalW and visualized in JalView. Indicated are highly conserved features as annotated and functionally validated from studies of *T. thermophilus.* Walker A and Walker B contain elements that form the ATP binding pocket[1]; the metal switch sensor proposed to control binding of the Mg^{2+} ion[2]. Sensor 1 and Sensor 2, Arginine Finger key residues coordinator ATP and facilitating hydrolysis[3]. Asterisks (*) indicate heterozygous *CLPB* mutations identified in our cohort.

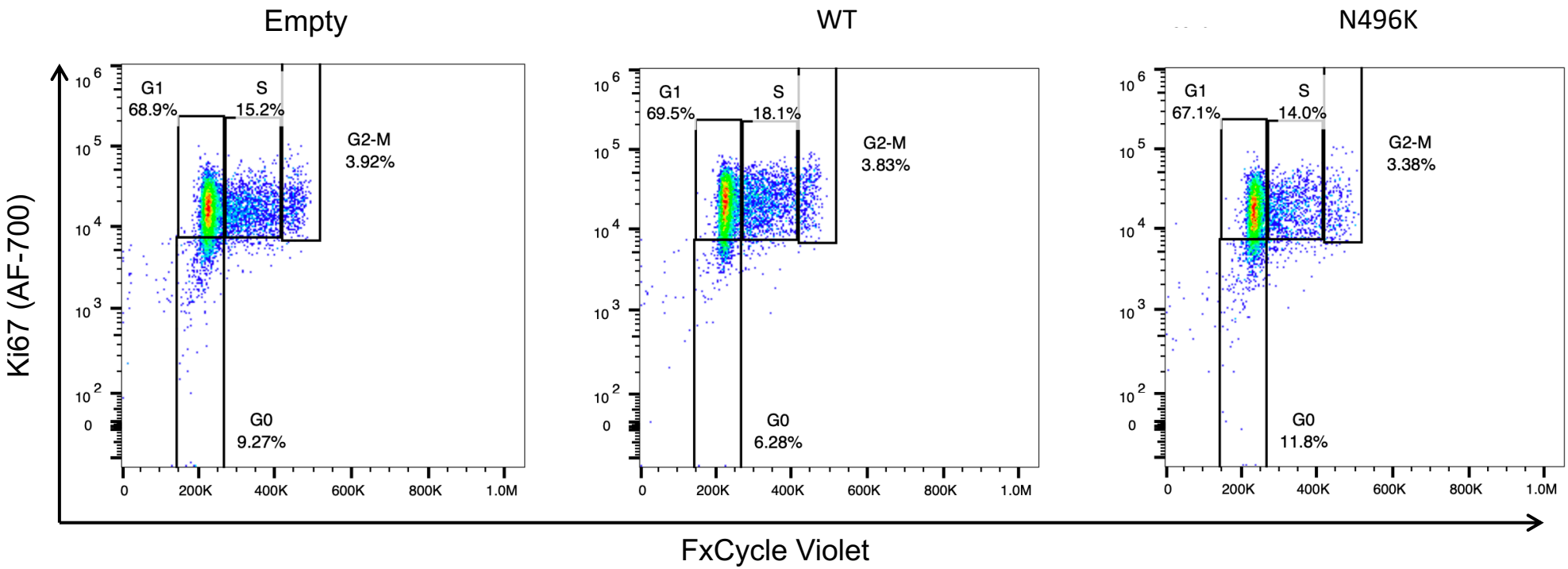




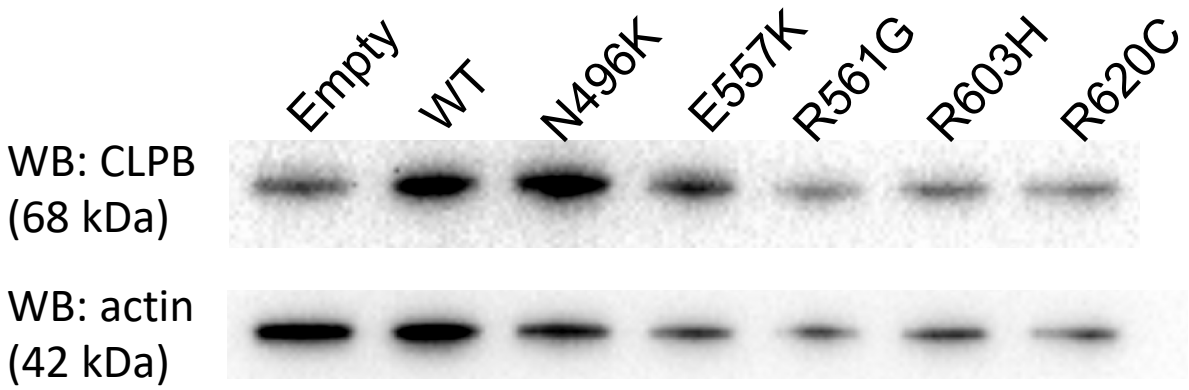




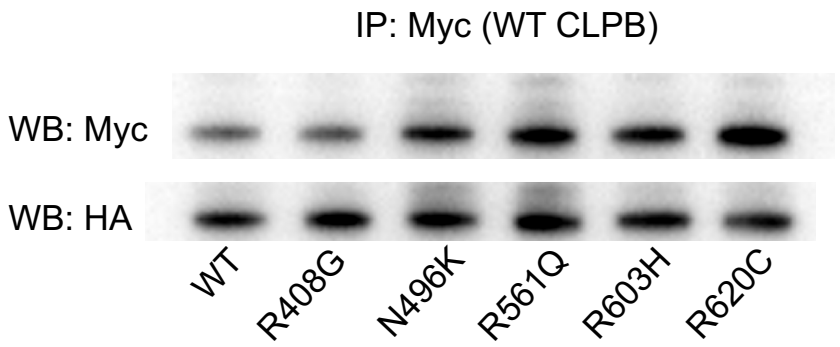
Supplemental Figure 3. Characterization of *CLPB* editing. (**A**) Schematic showing the domain architecture of CLPB and the regions targeted by small guide RNAs (sgRNA) g1 and g3, which both target an early exon. (**B-D**) Human cord blood CD34⁺ cells were nucleofected with complexes of recombinant Cas9 protein and sgRNAs targeting *CLPB* or a control guide RNA targeting an intronic region of the safe harbor locus *AAVS1*. Cells were analyzed 72 hours post nucleofection. (**B**) Percentage of edited or unedited for the control intronic *AAVS1* guide or edited out-of-frame (predicted to be deleterious), edited in-frame, or unedited for the two *CLPB* guides. (C) Western blot for CLPB from 3 independent experiments. (D) Representative sequence alignments used to quantify insertion/deletions identified in the targeted regions of *CLPB*. Data were analyzed using Crispresso2 as described[4].



Supplemental Figure 4. Cell cycle status flow cytometry. Representative flow cytometry plot showing the cell cycle markers Ki67 and FxCycle from a lentiviral experiment. CD34+ cells were transduced with CLPB lentivirus, sorted on GFP+, and cultured in neutrophil differentiation media, stained for cell surface granulocyte markers, fixed and permeabilized, followed by staining for Ki67 (Alexa-Fluor 700) and FxCycle Violet. Gating included selection for GFP+ population followed by granulocyte precursors gated as $CD14^{-}$, $CD11b^{+}$, $CD16^{-}$. Cell cycle status gates are shown above, drawn around the G0, G1, S and G2/M populations.



Supplemental Figure 5. Expression of CLPB protein in primary HSPC. Human primary CD34+ HSPCs were lentivirally transduced with empty vector or CLPB utilizing an IRES-GFP for monitoring transduction as described in the main text. 72 hours post-induction, all cells were collected and sorted on GFP-positivity directly into fetal calf serum followed by protease inhibitor treatment, lysis, and western blot analysis as described in the supplementary methods section.



Supplemental Figure 6. WT and CLPB variant proteins associate equally well *in vitro*. To ask whether CLPB variant proteins can interact with WT CLPB (which normally assembles as a homohexamer), we co-transfected plasmids expressing Myc-tagged WT CLPB with HA-tagged WT or mutant CLPB into 293T cells. Myc- or HA-tagged CLPB plasmid contained an IRES followed by a different fluorescent protein and expression was monitored by BFP-positivity (WT Myc-tagged CLPB) and GFP-positivity (WT or mutant HA-tagged CLPB). Double positive cells were sorted into Tris-NP40 lysis buffer with protease inhibitors, followed by immunoprecipitation using anti-Myc magnetic beads. Immunoprecipitated proteins were separated by gel electrophoresis following by transfer to PVDF membrane, and blots were probed for the Myc epitope (WT CLPB) and HA epitope (WT or mutant CLPB). Blot is representative of 3 independent experiments.

Supplemental Materials and Methods

Alignments

Primary amino acid sequences for CLPB protein from human (*Homo sapiens*, NP_001245321.1), macaque (*Macaca mulatta*, NP_001248231.1), mouse (*Mus musculus*, NP_001350920.1), Thermus thermophilus (Q9RA63), and E. coli (NP_417083.1) were aligned using ClustalW and visualized in JalView.

Western Blot

Edited cells were collected 72 hours after nucleofection and placed in fetal calf serum with the protease inhibitor diisopropyl fluorophosphate, with subsequent lysis using RIPA buffer (Cell Signaling) and protease inhibitor cocktail (Thermo Fisher). Following BCA protein assay quantification (Lamba Biotech), lysates were reduced with beta-mercaptoethanol and boiled prior to electrophoretic separation using a 4-12% Bis-Tris gel (Invitrogen). Proteins were transferred onto a PVDF membrane (Bio-rad), blocked with 5% milk and blotted for CLPB (Thermo Fisher), Myc epitope tag (Cell Signaling), HA epitope tag (Abcam), or beta-actin (Sigma). Bands were visualized using host species specific secondary antibodies conjugated to HRP and detected with high-sensitivity HRP substrate (Thermo Fisher). Blots were visualized with the ChemiDox XRS System (Bio-Rad).

Cytospin

Cells from liquid neutrophil differentiation cultures were centrifuged onto glass slides, allowed to dry briefly, then stained using a haematoxylin/eosin stain kit (Vector Laboratories). Slides were imaged using a color camera and 63X objective on the Zeiss Axio Observer.D1 microscope. Images were processed using the Zeiss Zen software. Samples were counted by two observers who were blinded to the sample identity, at least 100 cells were counted per slide.

Co-Immunoprecipitation

Plasmids were constructed using the pMND lentiviral backbone as described in the main text. This plasmid uses the MND promoter to drive WT CLPB followed by a C-terminal Myc epitope tag, internal ribosomal entry site (IRES), and Blue Fluorescent Protein. Addditionally using the same backbone WT or mutant CLPB followed by a C-terminal HA epitope tag, IRES, Green Fluorescent Protein. Plasmid DNA were prepared using a Gigaprep kit (Zymo Research). 293T cells were cultured in Dulbecco's Modified Eagle Medium (Gibco). One day prior to transfection, cells were seeded at 50% confluence in a 6-well plate. The next day, plasmid DNA was mixed with linear polyethylenimine (Sigma) at a 1:3 ratio in Optimem (Gibco), incubated at room temperature for 15 minutes, and added dropwise onto cells. 72 hours later, cells were lifted using TrypLE (Gibco), washed with phosphate buffered saline, and incubated in in cold buffer for 30 minutes (50mM TrisHCl pH 7.4, 150mM NaCl, 10% glycerol, and protease inhibitor (Pierce)). Supernatant was collected following centrifugation at 12,000g for 15 minutes and protein concentration was determined using a detergent compatible Bradford assay (Pierce). Equal total protein was incubated with anti-Myc magnetic beads according to the manufacturers instructions (Pierce). Proteins were eluted using Laemmli reducing buffer and resolved using western blot as described above.

Exome sequencing

Exome sequencing for the SCNIR North America cohort was performed using 250 ng of input DNA for dual-indexed library construction with the KAPA HTP kit for Illumina (Roche) followed by hybridization-capture enrichment with the IDT xGen Exome Panel (IDT). Amplified libraries were pooled and sequenced on a NovaSeq 6000 instrument (Illumina) to obtain a target exome coverage of >100x in paired-end 150 bp reads. Sequencing was performed on 90 samples, 4 of which did not pass quality control thresholds and were therefore not analyzed; 2 samples were found to be identical, therefore only 1 was utilized for further analysis. Data were aligned to

genome build GRCh38 and joint variant analysis across all samples was performed with the Genome Analysis Toolkit using the recommended best practices approaches[5]. Annotation was performed with VEP using release 88, and an initial variant file contained all variants passing quality filters and present in the ExAC exome variant database[6] non-Finnish European (NFE) population at a frequency <1%. SIFT[7], Polyphen[8], LRT[9], Mutation Taster[10], and CADD[11] were used for functional annotations to determine predicted deleterious effects. Mean expression (in transcripts per million or TPM) was derived from data using 3 healthy donors with populations defined as previously reported[12]. The following criteria were used to identify potentially pathogenic variants: 1) variants that altered amino acid sequence, including missense, nonsense, or splice site variants; 2) missense variants predicted to be deleterious based on a CADD score ≥ 15; 3) variants with a frequency of < 0.0025 in the ExAC database[6]; and 4) variants in genes that are highly expressed in granulocyte precursors (TPM > 2). CNV analysis was performed with cnvkit[13] using default parameters to generate a reference copy number profile across all samples, followed by the 'batch' command for identification of CNVs in each sample.

Exome sequencing for the French SCN registry cohort was performed on trios (proband-parents) after excluding the genes classically involved in SCN by targeted high throughput sequencing. Library preparation, exome capture and sequencing were performed by IntegraGen SA (Evry, France). Exome sequencing was performed from 250 ng of genomic DNA using the SureSelect XT Clinical Reasearch Exome - 54 Mb (Agilent). Sequencing was performed on a HiSeq4000 (Illumina) and 75 bp paired-end reads were generated. Bioinformatics analyses were done using an in-house pipeline. Sequence reads were mapped to the human genome build (hg19/GRCh37) using bwa-mem 0.7.10. Raw data quality was assessed with FastQC v0.11.5 and mapping file quality was assessed with picard-tools 1.121. All analyzed samples exceeded the standard 90% coverage at 20X threshold. Single-nucleotide variants and short

insertions/deletions were called with GATK 3.8 and annotated with SnpEff 4.3r. Functional predictions scores from dbNSFP v3 were added: CADD, SIFT, Polyphen3, REVEL, M-CAP, dbscSNV, as well as gnomAD allele frequency and number of heterozygous/homozygous, OMIM gene inheritance, HGMD and ClinVar annotations. An in-house script allowed us to select variants of interest, sorted by encoded categories: de novo, X-linked homozygous, homozygous and compound heterozygous. Variants were filtered based on their impact on the gene (missense, nonsense, frameshift, splice site-altering variants).

Variants identified through exome sequencing were confirmed using Sanger sequencing. Briefly, freshly prepared genomic DNA was amplified with primers specific for the exon of interest and submitted to commercial sequencing platforms (Genewiz, SCNIR samples). For samples Fr-0019 and Fr-0038, parentage was confirmed using the Powerplex 16 system (Promega) according to the manufacturer's instructions.

References

1. Lee, S. *et al.* The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell* **115**, 229–240 (2003).
2. Zeymer, C., Fischer, S. & Reinstein, J. trans-Acting arginine residues in the AAA+ chaperone ClpB allosterically regulate the activity through inter- and intradomain communication. *Journal of Biological Chemistry* **289**, 32965–32976 (2014).
3. Wendler, P., Ciniawsky, S., Kock, M. & Kube, S. Structure and function of the AAA+ nucleotide binding pocket. *Biochim Biophys Acta* **1823**, 2–14 (2011).
4. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature biotechnology* **37**, 224–226 (2019).
5. Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr Protoc Bioinform* **43**, 11.10.1-11.10.33 (2013).

6. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
7. Sim, N.-L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* **40**, W452–W457 (2012).
8. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248–9 (2010).
9. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res* **19**, 1553–1561 (2009).
10. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575–576 (2010).
11. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886–D894 (2018).
12. Network, C. G. A. R. *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England Journal of Medicine* **368**, 2059–2074 (2013).
13. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *Plos Comput Biol* **12**, e1004873 (2016).
14. Richards, S., Aziz, N. Bale, S., Bick, D., Das, S., Fastier-Foster, J., Grody, W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H., on behalf of the ACMG Laboratory Quality Assurance Committee. Standards and Guidelines for the Interpretation of Sequence Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* **17**, 405-423 (2015).