Supplementary Information

Supplementary Methods

CD34+ve Separation

Fresh blood was sourced either as 50 ml peripheral blood collected from CDA-I patients using EDTA Vacuettes (Becton Dickson) or 9 ml leukocyte cones (NHS Blood & Transport) for healthy controls. Blood was diluted with PBS and overlaid onto Histopaque-1077 (Sigma) and centrifuged for 30 min at 630 rcf (no brake). Peripheral Blood Mononuclear Cells (PBMCs) were washed in PBS and MACS buffer (PBS, 2 mM EDTA, 0.5% BSA) and stained with Human CD34 Microbead kit (Miltenyi Biotec) following the manufacturer's instructions for 30 minutes (4^{o}C) before being passed successively through one or two LS Columns (Miltenyi Biotec), for patients or healthy donors respectively. For each column three MACs buffer washes were conducted before cells were eluted and counted. The CD34+ fraction was cryopreserved in freezing media consisting of 90% FBS (Gibco) and 10% DMSO.

Next Generation Capture-C

Erythroblasts differentiated using a published method [1] of were crosslinked with 2% formaldehyde (10 minutes, room temperature); quenched with cold glycine; washed in phosphate buffered saline; resuspended in cold lysis buffer (Tris 10mM, NaCl 10mM, NP40 0.2%, cOmplete protease inhibitor cocktail (Roche)) and snap frozen to -80ºC. Cells were thawed on ice, washed in ddH2O and Dounce homogenised on ice (x 40 strokes). Cells were then resuspended with 0.25% SDS and restriction enzyme buffer and incubated at 37ºC for 1h at 1400rpm on a Comfort Thermomixer (Eppendorf) followed by a further incubation of 1 h following the addition of Triton X100 (final concentration 1.67%). An overnight digestion was performed using DpnII (500U /ml (NEB) at 37ºC / 1400 rpm). The digested chromatin was ligated overnight (Fermentas HC Ligase final concentration 10U/ml) at 16ºC at 1400 rpm on the Thermomixer. The samples were then de-crosslinked overnight at 65ºC with Proteinase K (Roche) followed by a 30 min incubation at 37ºC with RNAse (Roche). Phenol/Chloroform extraction was then performed followed by an ethanol precipitation and a wash with 70% ethanol. Digestion efficiencies were assessed by gel electrophoresis (1% agarose) and RT-PCR (TaqMan), which showed digestion efficiencies in excess of 70%. DNA content of the Dpn II 3C libraries were quantitated using a Qubit fluorometer (Life Technologies). 5-10ug of each library was sheared using a Covaris S2 in ddH2O. Covaris settings used were: duty cycle 10%; intensity 5; cycles/burst 200; time, 6 cycles of 60

seconds; set mode frequency sweeping; temperature 4 to 7º C. Following shearing DNA was purified using AMPureXP beads (Agencourt) and DNA quality assessed on a Bioanalyser 2100 using a DNA High Sensitivity Chip (Agilent). DNA end repair and adapter ligation was performed using the NEB Next or NEB Ultra DNA sample preparation reagent kits, depending on the amount of DNA available, using the standard protocol. Biotinylated capture oligonucleotides were designed to the ends of the viewpoint fragments (Supp Table 3). Where possible 1-2ug of each adapter ligated library were hybridized with the biotinylated capture oligonucleotides, using the Nimblegen SeqCap reagents and an adapted protocol. The quality of the resultant captured library was assessed by Agilent TapeStation or Bioanalyser (D1000). The resulting libraries were sequenced using Illumina MiSeq (150 bp paired-end reads). Reads were mapped and analysed using CCseqBasic5 (github.com/Hughes-Genome-Group/CCseqBasic5) as previously described [2] with the following custom settings (--bowtie2 --globin 2). Briefly, CCseqBasic5 trims adaptor sequences, flashes read pairs, in silco digests fragments and uses bowtie2 to map reads before identifying capture and reporter reads (Supp Table 2).

ChIP Seq

$1x10^7$ cells and nuclei were lysed for 10 minutes before chromatin was sheared by sonication in a Bioruptor (High: 10 min, 30 sec on, 30 sec off). Insoluble material was pelleted and supernatant was incubated overnight with the relevant antibody (Supp. Table 1). Antibodies were then conjugated to protein A-agarose and washed. DNA was extracted using phenol-chloroform followed by ethanol precipitation with yeast tRNA carrier. Samples were submitted to Oxford Genomics Center for indexing and sequencing (50 bp reads, Illumina Genome Analyzer II). ChIP-seq reads were mapped to the hg19 genome using NGseqBasic (V20; --blacklistFilter --noWindow --bowtie2). Sequence depth and mapped reads for each sample are provided (Supp. Table 1). For visualisation PCR-duplicate filtered bam files were converted to bigwigs with minimal smoothing using deepTools [3] (v2.2.2; bamCoverage --binSize 50 --normalizeUsingRPKM).

Western Blots of Patient Cells

All steps were performed at 4ºC. Digitonin (Sigma, D141) was added at a final concentration of 50μg/mL and cytosolic fractionation was performed for 30 mins instead of 10 mins. All lysis buffers were supplemented with 10% v:v protease inhibitor cocktail. Lysates were quantified using Qubit Protein Assay Kit (Invitrogen, Q33211). A Western-

Blot denaturing loading dye was prepared (125mM Tris-HCl (Sigma), 4% SDS (Severn-Biotech, 20-4002-10), 50% Glycerol (Sigma G5516), 0.4g Orange G (Sigma)) and mixed at a 1:10 ratio with the lysates. DL-DTT (Sigma, 43816) was added at a final concentration of 120mM. Samples were reduced and denatured by boiling for 10min at 95ºC, aliquoted in single-use samples and stored at -20ºC until use. For detection of Codanin-1 and C15orf41, nuclear lysates were used. Immediately prior SDS-PAGE, samples were boiled for 10 min at 95ºC and 12 μg of protein were loaded and run on pre-casted 4-12% NuPAGE Bis-Tris gels (Invitrogen NP0335). Samples were transferred on nitrocellulose membrane (Amersham GE Healthcare 10600008) for 1h30 at 30V. Transfer efficiency and total protein quantification were performed using REVERT Total Protein Stain (Li-Cor 926-11010) following manufacturer's instructions. Imaging in the 700nm channel (intensity=2 or 3) was performed with the Li-Cor Odyssey Classic scanner and software (Li-Cor Image Studio). A lane normalization factor was obtained by dividing the signal for each lane with the signal for the lane with the highest signal. Blocking was performed for 1h with constant agitation at room temperature with PBS-Milk 5%w:v (blocking buffer). Membrane was incubated with rabbit polyclonal antibodies directed against Codanin-1 (1:100, Bethyl Laboratories A304-951A) and C15orf41(1:50, Cusabio CSBPA897474LA01HU) in blocking buffer-0.2% Tween-20 overnight at 4ºC. Washes with PBS-0.1% Tween-20 were performed (3 x 5 mins) before incubating the membrane with near-infrared coupled secondary antibody (IRDye 800CW goat anti-rabbit IgG (H+L), 1:2000 in blocking buffer-0.2% Tween-20, Li-Cor) for 1h at room temperature in the dark. Blot was washed as above and a final wash with PBS for 10min was performed prior detection with the Li-Cor Odyssey Classic scanner and software on the 800nm channel (intensity=2 or 3). Target quantification values were normalized by dividing the target band signal by its lane normalization factor.

Immunofluorescence

Cells were washed and allowed to settle on poly-L-lysine–treated coverslips for 10 min. Cells were fixed in 4% PFA for 15 min and permeabilized in 0.2% Triton X-100 in PBS for 12 min at RT. Nonspecific sites were blocked using 10% fetal calf serum in PBS at RT for 30 min. Antibodies were prepared in blocking solution at the following concentrations: mouse anti-fibrilarin (ab-4566, AbCam Cambridge Ltd) 1:500; mouse anti–upstream binding factor (UBF) (SC13125 Santa Cruz Biotechnologies, Inc.) rabbit anti-Codanin-1 (Bethyl Laboratories A304-951A) 1:100 and rabbit anti C15orf41 (Cusabio CSBPA897474LA01HU) 1:50. Coverslips were mounted in Vectashield with DAPI

Culture, Transfection, Immunoprecipitation and Western Blots of U-2-OS Cells

Cells were grown at 37°C in a 5 % CO_2 atmosphere in Dulbecco's Modified Eagle's Medium (DMEM, Life Technologies) supplemented with 10% Fetal Bovine Serum (Life Technologies) and Penicllin/Streptomycin. U-2-OS cells expressing FLAG-HA-Codanin-1 were previously described.[4] HeLa S3 were kindly provided by P. Nakatani.

Cells were transfected with DNA using Lipofectamin2000® (Lifetechnologies) according to the manufacturer's instructions. For siRNAs, cells were transfected using Lipofectamine® RNAiMAX (Lifetechnologies) according to the manufacturer's instructions. For C15orf41, the following siRNAs were purchased from Ambion and used at 10 nM working concentration: #1: 5'- GGAACACGAGGAAACUCCAtt-3', #2: 5'- GCUGUAGAAGGGCACAUAAtt-3' and #3: 5'- UCACUGGAUUGAAAGCAAAtt-3'. The Codanin-1 targeting siRNAs were Codanin-1#1: 5'- GGAAAGAAUUCGUACCGUU-3' (sense strand) from Sigma and a Dharmacon ON-TARGET plus SMART pool (Codanin-1#2) as previously described.[4]

To harvest total protein from cells for western blots, cells were washed on ice twice in ice-cold PBS and lysed in Laemmli Sample Buffer (LSB) (50 mM Tris-HCl, 100 mM DTT, 2 % SDS, 10 % glycerol, bromphenol blue) for 5 min under rotation and collected. Benzonase (Sigma) was added to digest DNA for 1 hr at RT and the samples boiled for 10 min.

For immunoprecipitation, the total soluble proteins were extracted at 4°C with ice-cold NP40/NaCl buffer (300 mM NaCl, 0.5 % NP-40, 50 mM Tris pH 7.5, 0.2 mM EDTA, 5 % glycerol, and inhibitors (0.1 mM DTT, 5 mM sodium fluoride, 10 mM beta-glycerolphosphate, 0.1 mM PMSF, 10 µg/ml leupeptin, 10 µg/ml pepstatin, 0.2 mM sodium vanadate, trichostatin A). Cells were lysed for 20 min on ice and insoluble material removed by centrifugation. Cytosolic and nuclear extracts were obtained as previously described. [4]

FLAG-HA-tagged proteins were purified by incubation with anti-FLAG M2 magnetic beads (Sigma) for 3 hours at 4°C on a rotating wheel, and washed 8 times with low stringency washing buffer (200 mM NaCl, 20 mM Tris-HCl pH 7.8, 0.2 mM EDTA, 5 % glycerol, 0.2 % NP-40, supplemented with inhibitors). Bound proteins were eluted from the bead by boiling in LSB and analysed by western blotting.

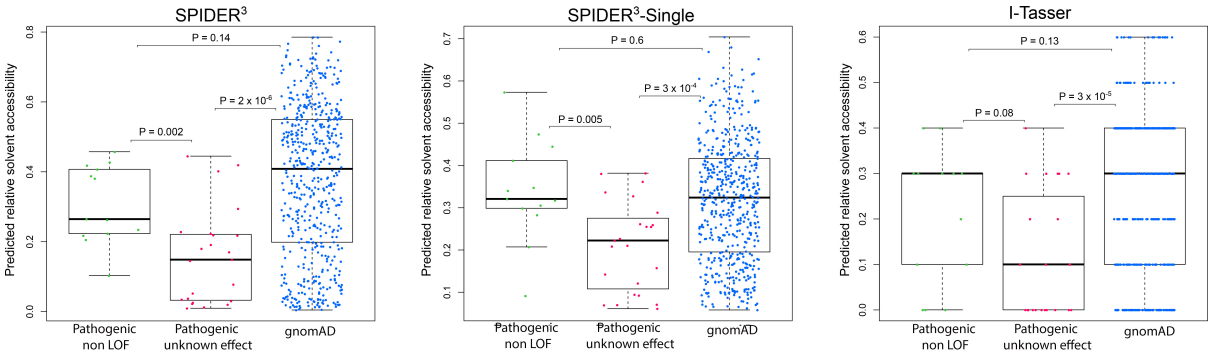
Gel Filtration

Cytosolic and nuclear extracts were applied onto a Superdex 200 10/30 GL column (Amersham) and eluted (20 mM HEPES pH 7.8, 5 mM potassium acetate, 300 mM KCl, 0.5 mM DTT, 5 % glycerol, inhibitors) at a flow of 0.3 ml/min in 400 µl fractions. 400ul

fractions were collected and the proteins precipitated with Trichloroacetic acid (TCA) and Deoxycholate (DOC).

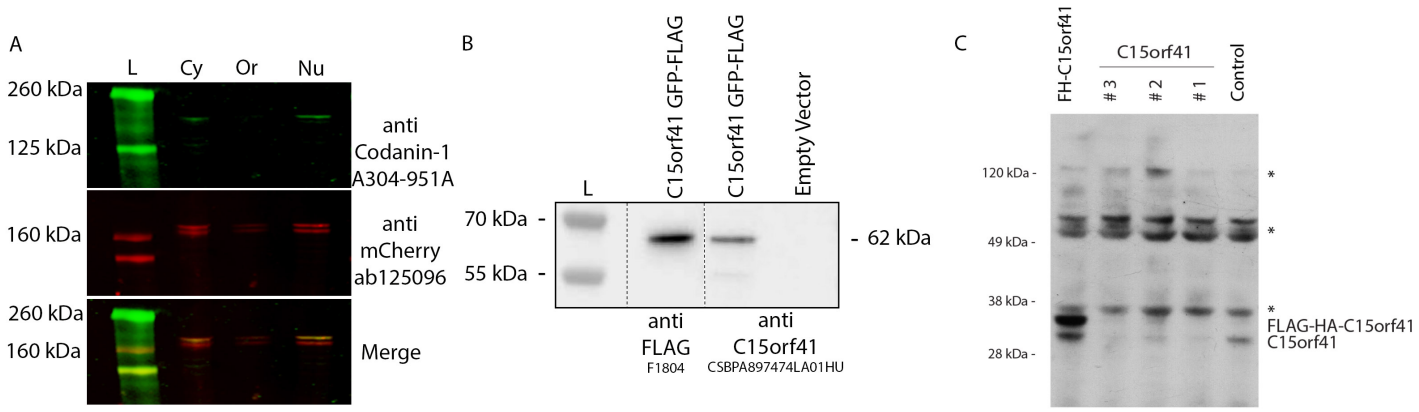
Supplementary Figures

Supplementary Figure 1



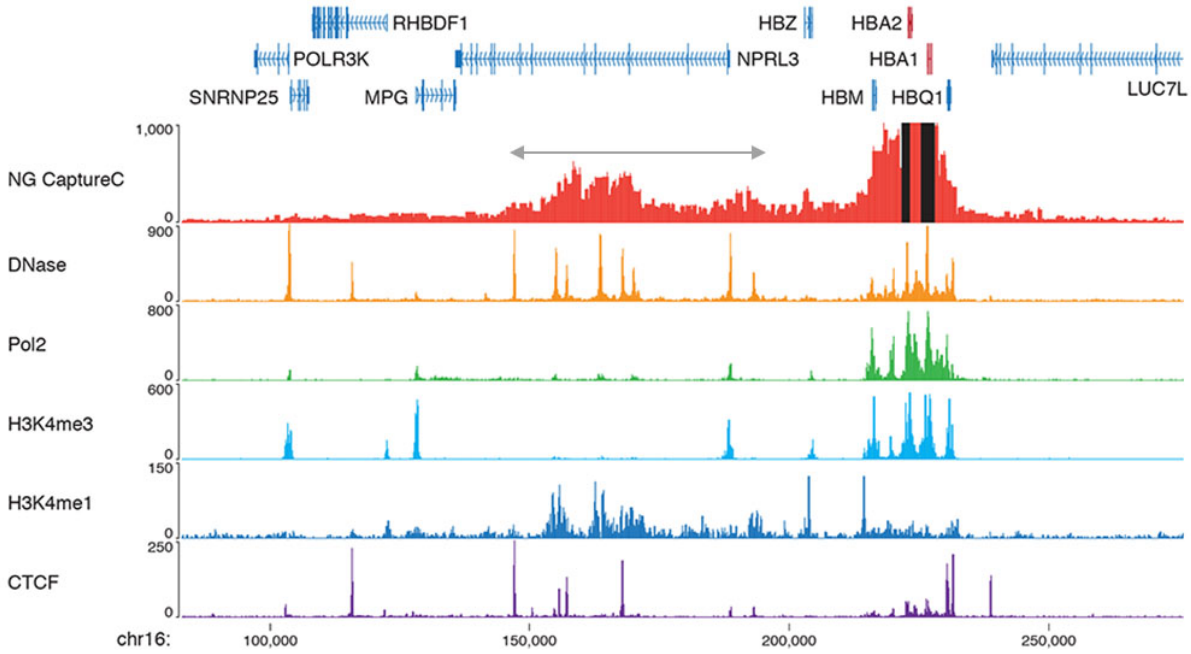
Supplementary Figure 1: Three further computational predictors (SPIDER3, SPIDER3-Single and I-Tasser) of solvent accessibility show that 'not LOF' pathogenic mutations tend to be less buried than 'mutations of unknown effect. Importantly, while most predictors of solvent accessibility utilise information from sequence alignments, SPIDER3-Single does not, demonstrating that the trends observed here are not confounded by evolutionary conservation.

Supplementary Figure 2



Supplementary Figure 2: Antibody validation. A. HEK293T cells were transfected with a plasmid encoding Codanin-1 in frame with mCherry, and harvested after 24 hours. Fractionated cell lysates were prepared, electrophoresed and blotted onto membranes. Each membrane was incubated with an anti mCherry antibody (Abcam ab125096) and an anti-Codanin-1 antibody (Bethyl Laboratories A304-951A). Membranes were imaged using the red and green channels using a Li-Cor Odyssey machine, confirming the anti-Codanin-1 antibody recognises the mCherry tagged over-expressed Codanin-1 protein. N.B. Conjugated Codanin-1 protein is larger than the native Codanin-1 protein, which is below the level of detection in HEK293T cells. B. A plasmid encoding C15orf41 in frame with green fluorescent protein (GFP) and a FLAG tag was transfected into HEK293T cells and cells were harvested after 24 hours. Cells were lysed and immunoprecipitated using anti-FLAG M2 magnetic beads (M8823). The immune-precipitated cell lysate was electrophoresed and blotted onto membranes. Each membrane was incubated with an anti-FLAG antibody (F1804) and an anti-C15orf41 antibody (CSBPA897474LA01HU) and appropriate HRP-secondary antibodies (anti-mouse P0447 and anti-goat P0048). A band of the same molecular weight can be observed in both lanes indicating the specificity of the anti-C15orf41 antibody. The tagged protein is larger (62 kDa) than the native protein owing to the presence of the GFP moiety. C. Validation of mouse monoclonal C15orf41 antibody generated in house. U-2-OS cells were transfected with the indicated siRNAs against C15orf41 or control, or a plasmid for expression of FLAG-HA-C15orf41. Total protein extracts were used in western blots. * indicates unspecific bands. L. ladder; Cy, Cytoplasmic fraction; Or, Organelle fraction; Nu, Nuclear fraction.

Supplementary Figure 3



Supplementary Figure 3: Capture-C showing chromatin interactions of the HBA1/2 promoters. The upper track shows Capture-C demonstrating chromatin interaction counts for each captured fragment (black bars). Lower tracks show normalised (RPKM) DNAseI-seq and ChIP seq for Polymerase II (Pol2), H3K4me4, H3K4me1 and CTCF. The increased region of interaction, with the enhancers is clearly visible (double headed arrow).

.

Categorisation of CDAN1 mutations:

Mutations present homozygously or in combination with a loss of function allele termed Not Loss of Function (Not LOF): P51L, F52L, G566V, E597del, N599S, P672L, L709del, R714W, R725W, T884A, R1042W, D1043V, and P1130L.

Other mutations which may or may not be loss of function alleles termed mutations of unknown effect: R335W, F360L, D365N, F369del, V373del, R397W, A412P, R623W, P671L, R688W, V690A, P694L, E698K, R725Q, G750C, Q754R, F868I, V869M, L915-922del, A944S, S1036F, L1113P, F1125S, S1227G.

Supplementary Tables

Supplementary Table 1 ChIPseq Reagents and Metadata

ChIP seq Sample	Antibody	Supplier	Sequence Length	Reads	Mapped Reads	Filtered mapped reads
CTCF	07-729	Upstate (now Millipore)	50	26875337	21730774	15055523
H3K4me1	07-436	Upstate (now Millipore)	50	15401114	7188408	6048547
H3K4me3	07-473	Upstate (now Millipore)	50	14946393	7222688	5123190
Pol2	sc-9001	Santa Cruz	50	14133732	6939310	4008517

Supplementary Table 2 Next Generation Capture C Reporter Counts

Reads			*Dpn*II containing			Mapped			Capture Contaiing			Filtered Reporter		
Total	Flashed	Unflashed	Total	Flashed	Unflashed	Total	Flashed	Unflashed	Total	Flashed	Unflashed	Total	Flashed	Unflashed
18318921	16607598	1711323	6594486	6151151	443335	7657005	6139206	1517799	3487651	3249940	237711	371426	332368	39058

Supplementary Table 3 Next Generation Capture-C Oligonucleotide Sequences

Target	Captured Fragment	Exclusion Region	Unique Interactions	Oligonucleotides 5'-3'
C15orf41	chr15:36871662-36872730	chr15:36871194-36873858	64804	GATCCTAGGTGGGCGATGCCCTGGATTCCTACCTGCCTACTGAAACTCTGGCGCGGCGCCTTTCCAAGGGCGGCCGCTACTTCCTGGTTGGTTTGCCCTTAACTGACATGGCGGCTAGTT
				GACTGTGGCTTGTACTGTTAACTGTGGAATTAAGTGACCTCAAGGGTTCTCAATCCAATCAGAGAGCAGGGTACCCAGGGATTCCACTCTTGGGTCATTGGCCATTTGAACAACCAGATC
CDAN1	chr15:43029235-43029526	chr15:43028564-43029732	67439	GATCCACCGCACGACGGCTGCGACCGACACCTCTTCTCGCAGCAGCGACTCCAAAACGGCCGCCATCCCGGTCGGGGCGCTCTGGGGCGACTGCGCAGGCGCCGGCGCGCCGCGGGCGGC
				GAGGGACCCGGCTGGCGGGGCGGGGCGGAGCCGGCTGTTGCCCAGACCCACGGAGTCAAGGCCTCGCTGAGTCCTCAACCCCGGCCGAGAGGCCGGTTGTTAGGAGGTGAAATGTGGATC
Hba-1	chr16:226254-227156	chr16:225254-228156	127940	GATCCCGCTGGAGTCGATGCGCGTCCAGCGCGTGCCAGGCCGGGGCGGGGGTGCGGGCTGACTTTCTCCCTCGCTAGGGACGCTCCGGCGCCCGAAAGGAAAGGGTGGCGCTGCGCTCCG
				AACGCCGTGGCGCACGTGGACGACATGCCCAACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCACAAGCTTCGGGTGGACCCGGTCAACTTCAAGGTGAGCGGCGGGCCGGGAGCGATC
Hba-2	chr16:222450-223352	chr16:221450-224352	127584	GATCCCGCTGGAGTCGATGCGCGTCCAGCGCGTGCCAGGCCGGGGCGGGGGTGCGGGCTGACTTTCTCCCTCGCTAGGGACGCTCCGGCGCCCGAAAGGAAAGGGTGGCGCTGCGCTCCG
				AACGCCGTGGCGCACGTGGACGACATGCCCAACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCACAAGCTTCGGGTGGACCCGGTCAACTTCAAGGTGAGCGGCGGGCCGGGAGCGATC

Supplementary Table 4 Allele Frequencies of *C15orf41* mutations

C15orf41 Mutation	Allele frequency
p.Gln7Ter	4.0657E-06
p.Gln13ProfsTer41	4.0615E-06
c.101+2T>C	8.2703E-06
c.102-1G>T	3.1851E-05
c.147+1G>A	4.0319E-06
p.Arg54LysfsTer20	6.3137E-06
c.212+1G>C	1.2682E-05
p.Tyr72Ter	5.0752E-06
p.Val77GlyfsTer15	5.031E-06
c.273+1G>A	5.7051E-06
c.274-2A>T	4.3754E-05
p.Gln109Ter	4.0699E-06
c.346+1G>A	8.276E-06
c.347-1G>T	1.342E-05
p.Gln130AspfsTer34	4.2583E-06
p.Ile157HisfsTer8	4.0204E-06
c.477-2A>G	4.0371E-06
p.Leu181Ter	4.0505E-06
c.545-1G>T	4.1854E-06
p.Gly190ValfsTer16	4.0905E-06
p.Cys207Ter	4.0201E-06
P20R	na
Y94C	na
C156Y	na
L178Q	na
H230P	na
Y236C	3.1853E-05

Supplementary Table 4: Loss of function alleles present in the gnomAD database (orange shading) and missense mutations (yellow shading) shown with allele frequencies. In cases where a mutation has been identified in a patient and is absent from the gnomAD database the frequency is shown as na (not applicable).

Supplementary Table 5 – Allele Frequencies of *CDAN1* mutations

CDAN1 Mutation	Allele Frequency
p.Glu11Ter	1.11318E-05
p.Ser14CysfsTer27	3.22144E-05
p.Leu57Ter	6.04259E-06
p.Glu59Ter	6.17147E-06
p.Gly120ProfsTer26	1.76442E-05
p.Val140GlnfsTer14	na
p.Pro146AlafsTer9	na
p.Pro146ArgfsTer9	na
p.Gly155GlnfsTer11	4.14814E-06
p.P185AfsTer9	na
c.570-2A>G	3.98032E-06
p.Pro235Ter	3.97681E-06
p.Arg251Ter	3.9769E-06
c.774-1G>T	4.02123E-06
p.Gln261HisfsTer45	8.02253E-06
p.Arg295SerfsTer20	3.98375E-06
p.His377GlnfsTer18	3.18512E-05
p.Arg410Ter	3.97722E-06
p.Tyr413Ter	3.97662E-06
p.Leu422GlyfsTer16	3.97627E-06
p.Arg447Ter	3.97624E-06
p.Phe459TrpfsTer28	6.3955E-05
p.Glu472GlyfsTer7	3.97621E-06
c.1534-2A>C	3.19122E-05
p.Ala519CysfsTer19	3.99336E-06
p.Ala519LeufsTer13	3.99377E-06
p.Leu524Ter	7.97315E-06
p.Ser532ValfsTer60	3.98314E-06
p.Met533TyrfsTer5	3.98206E-06
p.Leu542CysfsTer50	3.98448E-06
p.Gln546Ter	3.98346E-06
p.Leu549TyrfsTer23	3.18918E-05
p.Met550ArgfsTer42	3.18776E-05
c.1740-2dupA	1.9917E-05
p.Gln596Ter	3.97697E-06
p.Glu597AspfsTer14	7.95323E-06
p.Asn599GlufsTer11	7.95273E-06
p.Glu622Ter	3.9764E-06
p.Arg649GlyfsTer92	3.18776E-05
p.Gln660Ter	3.98064E-06
p.Arg682Ter	1.09361E-05
p.Glu711AspfsTer30	3.98333E-06
p.Leu719ProfsTer15	3.98235E-06

c.2174+2T>G	2.78891E-05
c.2175-2dupA	7.07019E-06
p.Lys736ArgfsTer5	3.97627E-06
p.Asn786IlefsTer44	3.98321E-06
p.Gln792Ter	3.9807E-06
c.2408-2A>G	3.98813E-06
p.Thr831AsnfsTer173	3.98406E-06
p.Gln847Ter	na
c.2541+1G>A	4.00006E-06
p.Val866GlyfsTer3	3.9763E-06
p.Val878GlnfsTer125	3.97624E-06
p.Lys879SerfsTer25	3.97624E-06
p.Glu894ValfsTer109	7.95773E-06
p.Gln898Ter	3.9795E-06
p.Gln900ArgfsTer3	3.97962E-06
c.2868+1G>A	1.42102E-05
c.2864_2868+1dupCAGCCG	3.99751E-06
c.2869-2A>G	2.78501E-05
p.Trp976Ter	3.97741E-06
p.Val993GlyfsTer13	na
p.Arg998Ter	4.29384E-06
p.Glu1003AspfsTer2	na
p.Glu1009LeufsTer24	na
p.Arg1010GlyfsTer22	4.32743E-06
p.Arg1011GlufsTer8	8.59173E-06
p.Cys1017TrpfsTer15	4.31581E-06
p.His1026ProfsTer8	9.2308E-06
p.Ile1028ArgfsTer16	4.64179E-06
p.Gly1047ArgfsTer12	8.96555E-06
p.Ser1049AspfsTer27	na
p.Ser1049PhefsTer27	2.70966E-05
p.Gln1059ProfsTer17	7.78165E-05
c.3204+1G>T	4.95756E-06
c.3204+2T>C	4.98738E-06
p.Ser1082LeufsTer4	3.9764E-06
p.Ala1086ProfsTer11	na
p.Ala1086ValfsTer106	3.97649E-06
p.Ala1101GlyfsTer90	3.98982E-06
p.Gln1109Ter	3.98038E-06
p.Arg1111Ter	1.194E-05
p.Arg1111SerfsTer77	3.97867E-06
p.Arg1111ProfsTer77	3.97899E-06
p.Glu1123ThrfsTer12	7.95855E-06
p.Glu1150Ter	3.18492E-05
p.Leu1179AlafsTer16	3.97766E-06
p.Gln1183Ter	7.95716E-06

p.Trp1184Ter	7.95906E-06
p.Glu1191Ter	1.26024E-05
p.Ala1193IlefsTer6	6.30756E-06
P51L	5.87427E-06
F52L	8.96147E-05
R335W	4.09997E-06
F360L	3.18552E-05
D365N	1.19381E-05
F369del	1.06136E-05
V373del	na
R397W	1.41E-05
A412P	na
G566V	3.98581E-06
E597del	na
N599S	7.95286E-06
R623W	1.19293E-05
P671L	na
P672L	8.49837E-05
R688W	8.12764E-06
V690A	na
P694L	8.01719E-06
E698K	na
L709del	na
R714W	7.56918E-05
R725Q	7.97E-06
G750C	na
Q754R	na
F868I	na
V869M	3.54E-05
T884A	7.96E-06
L915-922del	3.98E-06
A944S	na
S1036F	na
R1042W	9.09E-06
D1043V	1.34905E-05
R1065Q	na
L1113P	1.06E-05
F1125S	na
P1130L	1.59E-05
S1227G	na

Supplementary Table 5: Loss of function alleles present in the gnomAD database (orange shading) and missense mutations (yellow shading) shown with allele frequencies. In cases where a mutation has been identified in a patient and is absent from the gnomAD database the frequency is shown as na (not applicable).

Supplementary References

1. Davies JO, Telenius JM, McGowan SJ et al. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods* 2016;**13**(1):74-80
2. Hay D, Hughes JR, Babbs C et al. Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat Genet* 2016;**48**(8):895-903
3. Ramirez F, Ryan DP, Gruning B et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;**44**(W1):W160-5
4. Ask K, Jasencakova Z, Menard P et al. Codanin-1, mutated in the anaemic disease CDAI, regulates Asf1 function in S-phase histone supply. *EMBO J* 2012;**31**(8):2013-23