

## Recapitulation of erythropoiesis in congenital dyserythropoietic anaemia type I identifies defects in differentiation and nucleolar abnormalities

by Caroline Scott, Damien J. Downes, Jill M. Brown, Robert Beagrie, Aude-Anais Olijnik, Matthew Gosden, Ron Schwessinger, Christopher A. Fisher, Anna Rose, David J.P. Ferguson, Errin Johnson, Quentin A. Hill, Steven Okoli, Raffaele Renella, Kate Ryan, Marjorie Brand, Jim Hughes, Noemi B.A. Roy, Douglas R. Higgs, Christian Babbs, and Veronica J. Buckle

Haematologica 2020 [Epub ahead of print]

*Citation: Caroline Scott, Damien J. Downes, Jill M. Brown, Robert Beagrie, Aude-Anais Olijnik, Matthew Gosden, Ron Schwessinger, Christopher A. Fisher, Anna Rose, David J.P. Ferguson, Errin Johnson, Quentin A. Hill, Steven Okoli, Raffaele Renella, Kate Ryan, Marjorie Brand, Jim Hughes, Noemi B.A. Roy, Douglas R. Higgs, Christian Babbs, and Veronica J. Buckle.*

*Recapitulation of erythropoiesis in congenital dyserythropoietic anaemia type I identifies defects in differentiation and nucleolar abnormalities.*

*Haematologica. 2020; 105:xxx*

*doi:10.3324/haematol.2020.260158*

### *Publisher's Disclaimer.*

*E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval; the final version of the manuscript will then appear in print on a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.*

**Recapitulation of erythropoiesis in congenital dyserythropoietic anaemia type I (CDA-I) identifies defects in differentiation and nucleolar abnormalities**

Caroline Scott<sup>1</sup>, Damien J. Downes<sup>1</sup>, Jill M. Brown<sup>1</sup>, Robert Beagrie<sup>1</sup>, Aude-Anais Olijnik<sup>1</sup>, Matthew Gosden<sup>1</sup>, Ron Schwessinger<sup>1</sup>, Christopher A. Fisher<sup>1</sup>, Anna Rose<sup>1</sup>, David J.P Ferguson<sup>2</sup>, Errin Johnson<sup>3</sup>, Quentin A. Hill<sup>4</sup>, Steven Okoli<sup>5</sup>, Raffaele Renella<sup>6</sup>, Kate Ryan<sup>7</sup>, Marjorie Brand<sup>8</sup>, Jim Hughes<sup>1</sup>, Noemi B.A. Roy<sup>9</sup>, Douglas R. Higgs<sup>1</sup>, Christian Babbs<sup>1</sup> and Veronica J. Buckle<sup>1</sup>.

<sup>1</sup>Weatherall Institute of Molecular Medicine, Oxford University, Oxford, United Kingdom; <sup>2</sup>Ultrastructural Morphology Group, NDCLS, John Radcliffe Hospital, Oxford, United Kingdom; <sup>3</sup>Sir William Dunn School of Pathology, Oxford University, Oxford, United Kingdom; <sup>4</sup>Leeds Teaching Hospital NHS Trust, United Kingdom; <sup>5</sup>Imperial College, The Commonwealth Building, The Hammersmith Hospital, Du Cane Rd, London, United Kingdom; <sup>6</sup>Pediatric Hematology-Oncology Research Laboratory, CHUV-UNIL Lausanne Switzerland; <sup>7</sup>Department of Haematology, Manchester Royal Infirmary, Oxford Rd, Manchester, United Kingdom; <sup>8</sup>Sprott Center for Stem Cell Research, Ottawa Hospital Research Institute, Ottawa, Canada and <sup>9</sup>Department of Haematology, Oxford University Hospitals NHS Trust, Churchill Hospital, Old Rd, Headington, and NIHR Biomedical Research Centre, Oxford, United Kingdom.

Corresponding authors: Dr Caroline Scott ([caroline.scott@imm.ox.ac.uk](mailto:caroline.scott@imm.ox.ac.uk)) and Prof Veronica Buckle ([veronica.buckle@imm.ox.ac.uk](mailto:veronica.buckle@imm.ox.ac.uk)), Weatherall Institute of Molecular Medicine, Oxford University, Oxford, UK, OX3 9DS. Tel 441865222392

## **Abstract:**

The investigation of inherited disorders of erythropoiesis has elucidated many of the principles underlying the production of normal red blood cells and how this is perturbed in human disease. Congenital Dyserythropoietic Anaemia type 1 (CDA-I) is a rare form of anaemia caused by mutations in two genes of unknown function: *CDAN1* and *CDIN1* (previously called *C15orf41*), whilst in some cases, the underlying genetic abnormality is completely unknown. Consequently, the pathways affected in CDA-I remain to be discovered. To enable detailed analysis of this rare disorder we have validated a culture system which recapitulates all of the cardinal haematological features of CDA-I, including the formation of the pathognomonic 'spongy' heterochromatin seen by electron microscopy. Using a variety of cell and molecular biological approaches we discovered that erythroid cells in this condition show a delay during terminal erythroid differentiation, associated with increased proliferation and widespread changes in chromatin accessibility. We also show that the proteins encoded by *CDAN1* and *CDIN1* are enriched in nucleoli which are structurally and functionally abnormal in CDA-I. Together these findings provide important pointers to the pathways affected in CDA-I which for the first time can now be pursued in the tractable culture system utilised here.

## Introduction:

Many key discoveries in the process of erythropoiesis have come from identification and analysis of individuals with forms of inherited anaemia.<sup>1-4</sup> In some cases, because of the rarity of the disease and limited access to primary erythroid progenitors and precursors, progress can only be made by developing appropriate models of the disease or, ideally, methods that only require access to peripheral blood. Here we have utilised such an approach to study Congenital Dyserythropoietic Anaemia type 1 (CDA-I).

CDA-I is a rare autosomal recessive disease associated with ineffective erythropoiesis and macrocytic anaemia. Disease severity is commonly mild (not necessitating treatment) to moderate (requiring Interferon or occasional blood transfusion) but patients with severe disease can be transfusion-dependent from birth or even present as *hydrops fetalis*.<sup>5, 6</sup> Light microscopy reveals abnormalities in erythroblast nuclei, including binucleate cells and inter-nuclear bridging.<sup>5</sup> A diagnostic feature of CDA-I is the “Swiss-cheese” or spongy pattern of abnormal chromatin in up to 50% of erythroblasts obtained from bone marrow aspirates, visualised using transmission electron microscopy (TEM).<sup>6</sup> In ~90% of patients, bi-allelic mutations in *CDAN1* (encoding Codanin-1) or *CDIN1* (previously *C15orf41*) (encoding CDIN1) are causative,<sup>7, 8</sup> with the genetic basis of the remaining ~10% of patients yet to be determined. Both Codanin-1 and CDIN1 are widely expressed and appear to be essential to life<sup>5</sup> but their precise functions are unclear. CDIN1 comprises a helix-turn-helix binding domain and a predicted nuclease domain<sup>8</sup> whilst Codanin-1 has sequence similarity with a scaffold protein, CNOT1, involved in mRNA stability and translational control.<sup>9</sup> The two proteins form a complex where Codanin-1 is required for stability of CDIN1<sup>9-11</sup> and

both directly interact with the histone chaperone ASF1.<sup>12-14</sup> CDA-I is predominantly an erythroid-restricted disease but most of the structural and functional assessments of Codanin-1 and CDIN1 have been performed in non-erythroid cells and some characteristics described for the proteins are not recapitulated in patient-derived erythroblasts.<sup>9, 11</sup> Here we analyse the distribution and role of these proteins in erythroid cells.

Differentiation of CD34<sup>+</sup> haematopoietic stem and progenitor cells (HSPCs) from peripheral blood has been used to study normal erythropoiesis<sup>15-19</sup> and to elucidate disease mechanisms in a number of haematological disorders including Diamond Blackfan Anaemia (DBA),<sup>2, 3</sup> Hereditary Spherocytosis,<sup>20</sup> Congenital Dyserythropoietic Anaemia type II<sup>21</sup> and Myelodysplastic Syndrome (MDS).<sup>16</sup> To fully understand the defects that arise in patients who do not generate sufficient mature cells, any culture system must recapitulate terminal erythropoiesis through to enucleation and erythroblasts from controls and patients should be stage matched.<sup>22</sup> Here we use an *ex vivo* three-phase culture system<sup>15</sup> (broadly expansion, differentiation and enucleation) whereby CD34<sup>+</sup> HSPCs, obtained from peripheral blood of healthy donors and patients with TEM-confirmed CDA-I, are successfully differentiated into reticulocytes. We use immunophenotyping by FACS (fluorescence activated cell sorting), morphological assessment, single cell proteomics using CyTOF (mass cytometry time of flight) and ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) to enable direct comparisons of cell populations.

This approach successfully recapitulates the pathognomonic feature of 'spongy' heterochromatin in patient erythroblasts. We find a delay in terminal erythroid

differentiation and increased proliferation of CDA-I erythroblasts, associated with widespread changes in chromatin accessibility. We demonstrate that CDIN1 and Codanin-1 are enriched in nucleoli, which are structurally and functionally abnormal in CDA-I. These findings provide important indicators to the pathways affected in CDA-I, which can now be pursued in the tractable model of erythropoiesis utilised here.

## **Methods:**

***Patient Recruitment:*** Subjects were referred for next generation sequencing through the Oxford Molecular Diagnostic Centre. If a molecular diagnosis of CDA-I was made, patient consent was obtained for entry into this research study approved by the Wales Research Ethics Committee (REC5) (13/WA/0371), with written consent and compliance with the Declaration of Helsinki.

***Isolation and Differentiation of CD34<sup>+</sup> HSPCs:*** Peripheral blood mononuclear cells were isolated from 50 mL of EDTA anti-coagulated peripheral blood from three healthy donors and ten CDA-I patients using Histopaque. The CD34<sup>+</sup> HSPCs were extracted with the Human CD34 Microbead Kit (Miltenyi Biotec), according to the manufacturer's instructions.  $1 \times 10^5$  frozen CD34<sup>+</sup> HSPCs were recovered into Phase I media of a three-phase protocol<sup>15</sup> (see Supplemental Figure 2A and Supplemental Methods) and monitored by cytopsin (see Supplemental Methods) and FACS (Supplemental Table 1 and Supplemental Figure 2).

***Transmission Electron Microscopy (TEM):***  $5 \times 10^6$  staged intermediate erythroblasts were prepared for TEM as previously described.<sup>4, 23</sup>

***Iso-electric Focusing (IEF):***  $1 \times 10^6$  cultured erythroblasts were analysed by IEF (see Supplemental Methods).

**RT-qPCR for Globins:** RNA was extracted using a Tri-reagent protocol and RT-qPCR conducted with commercial TaqMan assays (Supplemental Table 2).

**Chromatin accessibility and NFE2 CHIP:** ATAC-seq was performed as previously described.<sup>24, 25</sup> NFE2 CHIP-seq was conducted on  $5 \times 10^6$  day 10 erythroblasts with previously described modifications<sup>26</sup> using rabbit anti-NFE2 (4  $\mu$ g, sc-22827-X; Santa Cruz discontinued). For ATAC-seq library preparation and analysis see Supplemental Methods.

**Antibody labelling, barcoding and mass cytometry for CyTOF:** Samples were prepared and analysed for CyTOF as previously described<sup>27</sup> (Supplemental Table 6). Subsampled events were concatenated for Uniform Manifold Approximation (UMAP).<sup>28</sup> See Supplemental Methods.

**Data Availability:** Sequencing data generated for this work is available on the Gene Expression Omnibus (GSE125175).

**Immunofluorescence (IF):** Cells were washed and processed essentially as described previously.<sup>29</sup> See Supplemental Methods.

**Fluorescence in situ Hybridisation (FISH):** FISH probes used were p7.1 (covering most of the rDNA array) and BAC CT476834 (demarcating perinucleolar heterochromatin) and were kindly gifted by Prof B. McStay.<sup>30</sup> Probes were labelled with Cy3-dUTP (GE Healthcare) or indirectly with digoxigenin-11-dUTP (Roche).<sup>31</sup> See Supplemental Methods.

**EU labelling and Analysis:** RNA transcripts were labelled by EU incorporation (1mM for 30 mins or 2 hrs) and detected by click chemistry with Alexa488® azide using Click-iT® RNA imaging Kit (ThermoFisher Scientific). Quantitation of EU was performed using FIJI.<sup>32</sup> All images were acquired using standardized settings and maximum-intensity projected. Mean EU intensities were quantitated using a nuclear

mask demarcated by DAPI.

## **Results:**

### **CDA-I patients**

To date, ~60 mutations have been reported in *CDAN1* and *CDIN1*,<sup>5, 7, 8, 33</sup> and six mutational hotspots have been identified in the *CDAN1* gene.<sup>10</sup> There are often differences in the severity of the disease between individuals, even for those with identical mutations.<sup>34</sup> We examined erythropoiesis in ten CDA-I patients (Supplemental Figure 1A, Supplemental Table 8). These patients (excluding those receiving regular blood transfusion or venesections) have haemoglobin (Hb) levels and mean cell volumes (MCV) within the normal range (Supplemental Figure 1B), consistent with ~30% of clinical cases<sup>6</sup> but tend to have higher mean cell haemoglobin (MCH), an increased red cell distribution width (RDW) and a reduced red cell count (RBC) compared to healthy donors (Supplemental Figure 1B). In one patient (UPID6) with CDA-I, confirmed by TEM, a potentially pathogenic homozygous variant was identified in *CDAN1* although the allele frequency for this mutation is >1% in specific populations. Data from this patient was included in the *CDAN1* mutation group.

### **Establishing a suitable model system using peripheral blood-derived CD34<sup>+</sup> HSPCs**

We initially validated a three-phase *ex vivo* culture system<sup>15</sup> for differentiation of CD34<sup>+</sup> HSPCs obtained from the peripheral blood of healthy donors (Supplemental Figures 2 and 3). In addition to the morphological assessment of cultured erythroblasts (Supplemental Figure 2B), we characterised their chromatin

landscape, globin gene expression profile and the expression of erythroid proteins and transcription factors, to comprehensively evaluate differentiation status (Supplemental Figure 4). Immunophenotyping revealed the expected gain of glycophorin A (CD235a) and transferrin receptor (CD71), which typically occurred by day 7 (Supplemental Figure 2C). As maturation progressed, cells visibly haemoglobinised by day 10 (Supplemental Figure 2D) and this coincided with increasing expression of the adult globins (Supplemental Figure 4A), with the  $\alpha$ - to  $\beta$ -globin ratios remaining around one throughout differentiation (Supplemental Figure 4B). IEF confirmed that predominantly adult globin was produced (Supplemental Figure 4C), and ATAC-seq showed open chromatin at the *HBA1/2*, and *HBB* genes and their associated locus control regions, again indicative of definitive erythropoiesis (Supplemental Figure 4D-E). During the third phase of culture, erythroblasts underwent the normal final stages of differentiation, with increased levels of Band 3 (CD233) and loss of the adhesion protein  $\alpha$ -4 integrin (CD49d) (Supplemental Figure 2C). Enucleated cells were visible on cytopins at this stage (Supplemental Figure 2B).

### **Differentiating erythroblasts from healthy donors and CDA-I patients are broadly equivalent by immunophenotyping**

We studied the differentiation of erythroblasts derived from CD34<sup>+</sup> HSPCs from CDA-I patients with a variety of mutations in *CDIN1* and *CDAN1* (Supplemental Figure 1A). Flow cytometry bulk population analysis showed that differentiation of CDA-I patient HSPCs appeared to be equivalent to that of the healthy donors with loss of CD34<sup>+</sup> and gain of erythroid markers CD71 and CD235 from day 7, and expected changes in CD36, CD49d and Band 3 occurring from day 10. (Figure 1A,

and for gating strategy Supplemental Figure 3).

We also analysed the erythroblast immunophenotype by CyTOF, a next-generation flow cytometry platform that allows functional and phenotypic characterisation of cell populations.<sup>35</sup> We examined the levels of 25 erythroid transcription factors and cell surface markers (Supplemental Table 6) at day 11 in healthy donor and *CDIN1* patient-derived erythroblasts. Visualisation by uniform manifold approximation and projection (UMAP), widely used to identify distinct cell populations in cytometry data,<sup>28</sup> revealed that erythroblasts from both groups follow a continuous trajectory during differentiation with a clear separation of cells expressing, for example, low levels of CD235 (denoted in blue) from those in the tail which have high expression of CD235 (Figure 1B). Cells from healthy donors and *CDIN1* patients appear similarly distributed along the trajectory of differentiation at day 11 for all 25 markers analysed by CyTOF (Figure 1B and Supplemental Figure 5) and no significant differences in the clustering patterns were identified by K-means clustering analysis.<sup>36</sup> Together with the FACS data (Figure 1A), this indicates that the patient and healthy donor samples cannot be distinguished by day 10/11 on the basis of immunophenotype.

### **Erythroblasts from CDA-I patients display delayed differentiation with increased proliferation**

Despite bulk population immunophenotyping indicating that patient and healthy donor cells are grossly stage matched in expansion and into differentiation, analysis of the proportions of morphologically-classed erythroid cells identified differences in the progression through differentiation, detectable from day 10 (Figure 2A, B). At this timepoint, significantly more erythroblasts from healthy

donors than *CDAN1* patients had reached the polychromatic stage. By day 17, there were more enucleated cells from healthy donors than *CDIN1* patients, whose erythroblasts were still delayed at the polychromatic stage, with *CDAN1* erythroblasts having progressed a little further. Therefore CDA-I patients exhibit clear delay in their erythroid differentiation. In order to firmly establish whether this represents a delay or a block, cultures would need to be continued beyond day 17 to look for persistence of precursor forms.

We also observed greater expansion in patient erythroblast numbers compared with healthy donors (Figure 2C) and this was significant even at the end of the expansion phase by day 6/7 (*CDAN1*  $P=0.0087$ , *CDIN1*  $P=0.0043$ ). The increase in viable cell counts for patient erythroblasts became especially marked in the later phase of culture, reaching significance at day 17 of  $P=0.0043$  for both patient groups.

### **Evidence of defective differentiation in patient erythroblasts, most severe for *CDIN1* mutations.**

TEM revealed that the pattern of chromatin abnormalities characteristic of CDA-I was present in patient erythroblasts by day 11 of culture. This feature was observed in all our patient samples, averaging 29% ( $\pm 7.7$  SD) of nuclei affected (Figures 3A, 3B and Supplemental Figure 7A). Furthermore, elevated expression of growth differentiation factor 15 (GDF15), a marker of ineffective erythropoiesis known to be increased in CDA-I patients,<sup>37</sup> was detected at day 10 by immunofluorescence (IF) (Supplementary Figure 7B). Therefore, *ex vivo* differentiation of CDA-I erythroblasts successfully recapitulates abnormal cellular phenotypes observed in bone marrow derived erythroblasts and these features are already apparent midway through

terminal differentiation.

We next assessed the effects of patient mutations on the enucleation stage of differentiation. Firstly, analysis of cellular morphology indicated a persistence of erythroid precursors in CDA-I cultures, particularly in those from patients with *CDIN1* mutations, together with a significant reduction in the percentage of enucleated cells (Figure 2A). Secondly, immunophenotyping of cultured erythroblasts in the enucleation phase revealed changes in Band 3 expression in CDA-I patients. In normal erythropoiesis Band 3 shows a marked increase from the pro-erythroblasts to late erythroblasts,<sup>16</sup> and while CDA-I patient erythroblasts did progressively gain Band 3, the level of protein was significantly less at day 17 than in healthy donors (Figure 3C, D). This supports other indications of a delay in the progression of differentiation in patient-derived erythroblasts. Notably the Band 3 reduction was more severe in patients with *CDIN1* mutations (n=3  $P=0.0088$ ) than in the *CDAN1* mutant cells (n=4  $P=0.0127$ ) (Figure 3D).

### **Differentiating erythroblasts from CDA-I patients display an altered regulatory landscape**

Analysis of open chromatin regions can be used to distinguish cell types and deconvolve mixed cell populations;<sup>18, 38</sup> therefore to further compare our differentiating healthy and patient cells we assayed chromatin accessibility at day 10 and day 13 using ATAC-seq.<sup>24</sup> We mapped healthy and patient cell populations against a differentiation trajectory of sorted erythropoietic cell populations and 136,698 nucleosome-depleted regions,<sup>18, 38</sup> using Principal Component Analysis (PCA) of all open chromatin regions (Supplemental Figure 6).

This mapping could not distinguish patient material from healthy donors on the differentiation trajectory.

Next we undertook a PCA analysis of our cultured erythroblast data alone (without plotting against other erythroid populations), where a distinction in the accessibility profile between healthy donor and patient samples became apparent (Figure 4A), and this was more marked by day 13. We looked at pooled day 10 and day 13 ATAC-seq data by DESeq2 analysis for differences in DNA accessibility peaks not attributable to differentiation status (Figure 4B). There were 61 peaks displaying increased accessibility in patients and 531 less accessible sites. The latter showed a marked enrichment for enhancers (strong and weak) (65%) (Figure 4C) when assessed for chromatin state annotations.<sup>26</sup> Furthermore, 40% of the 531 less accessible sites in CDA-I patients have a binding motif for the NF-E2 family of transcription factors (Figure 4D).

Consistent with motif distribution, the mean level of NF-E2 binding in ChIP-seq from healthy donor erythroblasts and K562 erythroleukemic cells was significantly higher than background at the peaks with patient-specific decreased accessibility (CDA-I down) (Supplemental Fig 8A), indicating that NF-E2 normally binds these sites. Any difference in NF-E2 binding could not be attributed to altered protein abundance as similar levels of NF-E2 were detected in healthy donors and *CDIN1* patients by CyTOF mass cytometry (Supplemental Fig 8B). Similarly, no differences were seen in abundance for either MAFG, which dimerizes with NF-E2,<sup>39</sup> or BACH1 which competitively binds with the NF-E2 motif.<sup>40, 41</sup>

Of note, the Band 3 encoding gene *Slc4a1* has erythroid-specific enhancer elements (Supplemental Figure 8C). The 5' enhancer has a binding motif for NF-E2 and is

bound by NF-E2 in ChIP analyses of three different erythroblast cultures so that reduced accessibility at this site could account for reduced levels of Band 3 observed in CDA-I patients. The decrease in accessibility at this specific site did not quite reach significance in patient erythroblasts however the ATAC was performed at day 10 and day 13 which is possibly too early to observe an effect for this gene.

### **The structure of nucleoli is disrupted in CDA-I patients**

With multiple strands of evidence for an altered pattern of differentiation in CDA-I patients, we looked for abnormal features that might be linked to the affected proteins. We have previously shown both Codanin-1 and CDIN1 endogenous proteins to be enriched in erythroblast nucleoli.<sup>10</sup> Both proteins also show nucleolar enrichment in HEK293T (human embryonic kidney cell line), G-292 (human osteocarcinoma line), mES E14 (mouse embryonic stem cells) and B16F10 (mouse melanoma cell line) (data not shown), indicating this is a common feature across a range of cell types. We therefore examined nucleolar structure in day 10/11 erythroblasts by fluorescence *in situ* hybridisation (FISH) using probes detecting the heterochromatic region adjacent to rDNA arrays (BAC-CTD), and the rDNA arrays themselves (p7.1),<sup>30</sup> where the heterochromatin normally surrounds the rDNA signal within nucleoli. In all four patients examined (UPIDs 10, 20, 22 and 25), nucleoli in a proportion of nuclei appeared more numerous, less ordered and less regular in shape, and the rDNA arrays appeared less open (Figure 5A). This was observed at a timepoint when 19-29% of nuclei in these patients exhibited abnormal chromatin distribution by TEM. Such disrupted organisation might be expected to impact on the synthesis of ribosomal RNA, which accounts for a major portion of RNA synthesis in the cell.<sup>41, 42</sup> We therefore assessed RNA synthesis in two patients by measuring

incorporation of the uridine analogue 5 ethynyl uridine (EU) into newly synthesised RNA<sup>42-44</sup> at day 10/11 of differentiation. In both cases there is a significant reduction in nuclear EU labelling (Figure 5B) despite only a percentage of nuclei apparently affected and, particularly in UPID15, there is a distinct cell population with low EU signal. We have previously shown for four patients that the two mutated proteins are not destabilized by missense and in-frame mutations and remain detectable.<sup>10</sup> Therefore, we investigated whether the normal enrichment of Codanin-1 and CDIN1 in nucleoli is also disrupted, in patients with predicted non-destabilising mutations. Using IF with day 10 erythroblasts derived from *CDAN1* patients UPID 6, 16, 20 and 22, in combination with nucleolar proteins UBF and Fibrillarin, we observed that the disrupted appearance of nucleoli recapitulated that observed by FISH (Figure 5C). Further, we were able to detect that both CDIN1 and the mutant Codanin-1 remained associated with nucleolar proteins in patient erythroblasts (Figure 5C).

### **Using the culture system to validate novel variants in CDA-I patients**

The pathogenicity of novel *CDIN1* or *CDAN1* variants identified by sequencing requires further evidence, such as chromatin abnormalities in bone marrow biopsies by TEM. TEM is not only relatively inaccessible but often requires a second bone marrow biopsy.<sup>5</sup> By contrast, peripheral blood is usually accessible and was used here in the *ex vivo* culture system to confirm the diagnosis of CDA-I in an infant with two novel mutations in *CDAN1* (UPID33) (Supplemental Figure 1).<sup>10</sup> Genetic analysis was conducted on the patient, who presented with unexplained anaemia, using the Oxford Red Cell Panel (ORCP)<sup>45</sup> (Supplemental Figure 9). Following identification of two novel variants, CD34<sup>+</sup> HSPCs from UPID33 were extracted from peripheral blood and after 11 days in *ex vivo* culture, TEM on the resulting intermediate

erythroblasts revealed 39% of erythroblasts with abnormal chromatin morphology, thus confirming diagnosis of CDA-I.

## **Discussion:**

Although understanding the cellular and molecular basis of CDA-I has the potential to elucidate new insights into the process of erythropoiesis, research is constrained by the limited studies that can be conducted using primary erythroid progenitors and precursors derived from patients with this condition. Here, using a modified *ex-vivo* culture system, we demonstrate that healthy control erythroblasts pass through the expected stages of differentiation with appropriate expression of erythroid cell surface markers<sup>16</sup> and are able to enucleate. Further, we recapitulate the cardinal haematological features of CDA-I and show by TEM that up to 40% of patient-derived erythroblasts have spongy heterochromatin, indicating that *ex vivo* culture can be used to elucidate mechanisms underlying this disease.

We staged the cultures using an array of methods including FACS and CyTOF, which rely on immunophenotyping cell populations. While such methods showed healthy and diseased erythroblasts were immunophenotypically similar during the expansion phase and then into differentiation (day 10-13), aspects of disordered erythropoiesis were already evident at this mid-differentiation timepoint. In particular at this stage we noted a delay in progression through differentiation together with an increased proliferation of CDA-I erythroid precursors, producing increased amounts of GDF15, characteristic of dyserythropoiesis,<sup>37, 46</sup>. Many of these cells exhibited the abnormal chromatin structure associated with CDA-I. These findings show that the effects of the mutant proteins start to operate early in terminal differentiation, indeed the viable cell counts would suggest that there may already be effects by the end of the expansion phase

at day 6/7.

Chromatin accessibility has become a superior approach for cell type classification, including haematopoietic lineages.<sup>38</sup> The accessibility of transcription factor motifs within chromatin changes as subsets of regulatory elements are systematically activated and repressed during commitment to different lineages.<sup>47</sup> Therefore, we used ATAC-seq as a genome-wide method to stage cell populations. When healthy donor and patient material from day 10 and day 13 cultures were aligned with a defined ATAC-seq erythroid trajectory,<sup>38</sup> both map as expected with the intermediate and later stages of erythropoiesis. A more detailed PCA revealed a distinction between healthy donors and patients, more marked by day 13, identified a reduced accessibility in patient erythroblasts at gene enhancers containing the erythroid-specific NF-E2 motif. NF-E2 (comprising NFE2-p45 and MAFG) and BACH1 (which binds the same motif) are important transcription factors for erythropoiesis and the oxidative stress pathway respectively.<sup>39-41, 48</sup> CyTOF analysis indicates that levels of these three proteins appear to be normal in patient cells at day 11. NF-E2 motifs normally increase accessibility towards late erythropoiesis,<sup>18</sup> in parallel with the level of the protein complex,<sup>49, 50</sup> and we show that in normal erythroblasts, NF-E2 does bind those sites. Together these facts suggest that reduced accessibility of this motif at enhancers could affect the later stages of erythroid differentiation. It is possible that reduced accessibility of this motif reflects a generally altered regulatory landscape due to delayed differentiation, however motifs for other erythroid-specific transcription factors such as Gata1 did not reach significance in terms of altered accessibility.

The observation of disrupted nucleoli raises interesting ideas about the roles the two

proteins may play in erythropoiesis and could explain the erythroid-specific nature of the disease. Mutations to the ribosomal proteins themselves can underlie tissue restricted disorders, including erythroid-specific disorders such as Diamond-Blackfan syndrome, Schwachmann-Diamond syndrome, Dyskeratosis Congenita and MDS.<sup>43, 51</sup> Impaired function in the nucleolus could affect the number of available ribosomes and have similar effects to these other conditions in producing anaemia. Beyond that, the nucleolus appears to have other, regulatory roles.<sup>43, 52, 53</sup> Of interest, given the importance of cell cycle described above, is the proposed role for the nucleolus in cell cycle regulation.<sup>52</sup> Another possibility is that CDIN1, with its sequence similarity to the Holliday junction resolvase family of proteins, could function in a repair pathway. The high transcription rate within nucleoli can lead to topological stress and double strand breaks<sup>53</sup> whilst partial deletion of rDNA arrays has been shown to cause disordered nucleolar structure.<sup>54</sup> Further work is required to test these possibilities.

The final stages of erythropoiesis involve nuclear condensation prior to expulsion of the pyknotic nuclei by enucleation<sup>55, 56</sup> and this process is highly organized<sup>50</sup> with chromatin condensation playing an important role.<sup>57</sup> The abnormal spongy heterochromatin observed in CDA-I could have a significant impact on the usual processes that precede enucleation, such as the selective loss of histones.<sup>55</sup> Remarkably, a substantial number of erythroblasts progress to enucleation without developing the catastrophic changes in chromatin compaction and organisation apparent in spongy nuclei. This implies that the effects of the aberrant proteins must reach a threshold within individual cells to produce the pathognomonic phenotype and could be related to the balance between euchromatin and heterochromatin under nucleolar regulation.<sup>54, 58</sup>

Two distinct types of CDA-I have been reported (CDA-Ia MIM 224120 and CDA-Ib MIM 615631)<sup>59</sup> based on the levels of Hb and MCV, with the CDA-Ib patients (caused by *CDIN1* mutations) thought to be more severely affected. In our patient cohort (excluding those regularly transfused or venesected), there is overlap between the blood indices irrespective of the mutation (Supplemental Figure 1). However, we observe a more pronounced delay in differentiation, increased proliferation and significantly reduced levels of Band 3 expression in erythroblasts cultured from *CDIN1* patients, as compared to those with mutations in *CDAN1*. This implies that there may indeed be a distinction based on patient genotype where the CDA-I phenotype is more severe when arising from *CDIN1* mutations.

In this study we provide a detailed characterisation of CDA-I erythroblasts. We recapitulate aspects of the disease pathology seen in CDA-I, including high levels of cells with spongy heterochromatin and increased GDF15 expression. We report that CDA-I patient erythroblasts have elevated levels of proliferation, together with delay in the differentiation process and reduced levels of enucleation. There are difficulties in identifying and quantifying abnormalities in this disorder since only a proportion of erythroblasts exhibit defects whilst the majority differentiate and manage to function as red cells in many patients. Further, nurturing culture conditions may diminish the abnormal phenotypes observed.<sup>21</sup> Nevertheless, ATAC-seq analysis provides clear evidence of an altered regulatory landscape during terminal differentiation. This, together with the observations of aberrant nucleolar structure and transcriptional output, gives insight into the underlying disease mechanism and highlights several new avenues for further investigation of the functional role of the two proteins in erythroid differentiation.

## **Acknowledgements:**

We thank the CDA-I patients for providing blood samples. We acknowledge the flow cytometry facility at the WIMM for providing cell analysis services and technical expertise, supported by the MRC HIU; MRC MHU (MC\_UU\_12009); NIHR Oxford BRC; Kay Kendall Leukaemia Fund (KKL1057), John Fell Fund (131/030 and 101/517), the EPA fund (CF182 and CF170) and by the WIMM Strategic Alliance awards G0902418 and MC\_UU\_12025. We also acknowledge the Electron Microscopy Facility at the Sir William Dunn School of Pathology for conducting the majority of the TEM and Raman Dhaliwal for help with sample preparation and imaging.

Further support came from grants to the Wolfson Imaging Centre Oxford (Wolfson Foundation 18272, joint MRC/BBSRC/ EPSRC MR/K015777X/1, Wellcome Trust Multi-User Equipment 104924/Z/14/Z). We would like to acknowledge Giorgio Napolitani and Michalina Mazurczyk for help in the mass cytometry facility at the WIMM, providing technical expertise and cell analysis services. The facility is supported by the MRC HIU core funded project MC\_UU\_00008 and the Oxford Single Cell Biology Consortium (OSCBC). This work was supported by the charity Congenital Anaemia Network (CAN)(UK charity no. 1176864), Blood Buddies (UK charity no. 1108692), the Medical Research Council MC\_UU\_00016/1, a Wellcome Trust Strategic Award (106130/Z/14/Z) and the National Institute for Health research (NIHR) Oxford Biomedical Research Centre Haematology Theme at Oxford University Hospitals NHS Trust and Oxford University.

Conflict of interest disclosure: The authors declare no competing financial interests.

## References

1. Migliaccio AR, Varricchio L. Concise Review: Advanced Cell Culture Models for Diamond Blackfan Anemia and Other Erythroid Disorders. *Stem Cells*. 2018;36(2):172-179.
2. Moniz H, Gastou M, Leblanc T, et al. Primary hematopoietic cells from DBA patients with mutations in RPL11 and RPS19 genes exhibit distinct erythroid phenotype in vitro. *Cell Death Dis*. 2012;3(7):e356.
3. O'Brien KA, Farrar JE, Vlachos A, et al. Molecular convergence in ex vivo models of Diamond-Blackfan anemia. *Blood*. 2017;129(23):3111-3120.
4. Renella R, Roberts NA, Brown JM, et al. Codanin-1 mutations in congenital dyserythropoietic anemia type 1 affect HP1alpha localization in erythroblasts. *Blood*. 2011;117(25):6928-6938.
5. Roy NBA, Babbs C. The pathogenesis, diagnosis and management of congenital dyserythropoietic anaemia type I. *Br J Haematol*. 2019;185(3):436-449.
6. Wickramasinghe SN. Congenital dyserythropoietic anaemias: clinical features, haematological morphology and new biochemical data. *Blood Rev*. 1998;12(3):178-200.
7. Dgany O, Avidan N, Delaunay J, et al. Congenital dyserythropoietic anemia type I is caused by mutations in codanin-1. *Am J Hum Genet*. 2002;71(6):1467-1474.
8. Babbs C, Roberts NA, Sanchez-Pulido L, et al. Homozygous mutations in a predicted endonuclease are a novel cause of congenital dyserythropoietic anemia type I. *Haematologica*. 2013;98(9):1383-1387.
9. Swickley G, Bloch Y, Malka L, et al. Characterization of the interactions between Codanin-1 and C15Orf41, two proteins implicated in congenital dyserythropoietic anemia type I disease. *BMC Mol Cell Biol*. 2020;21(1):18.
10. Olijnik AA, Roy NBA, Scott C, et al. Genetic and functional insights into CDA-I prevalence and pathogenesis. *J Med Genet*. 2020 Jun 9. [Epub ahead of print]
11. Shroff M, Knebel A, Toth R, Rouse J. A complex comprising C15ORF41 and Codanin-1- the products of two genes mutated in congenital dyserythropoietic anemia type I (CDA-I). *Biochem J*. 2020;477(10):1893-1905.
12. Noy-Lotan S, Dgany O, Lahmi R, et al. Codanin-1, the protein encoded by the gene mutated in congenital dyserythropoietic anemia type I (CDAN1), is cell cycle-regulated. *Haematologica*. 2009;94(5):629-637.
13. Ask K, Jasencakova Z, Menard P, Feng Y, Almouzni G, Groth A. Codanin-1, mutated in the anaemic disease CDAI, regulates Asf1 function in S-phase histone supply. *EMBO J*. 2012;31(8):2013-2023.
14. Ewing RM, Chu P, Elisma F, et al. Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol*. 2007;3:89.
15. Griffiths RE, Kupzig S, Cogan N, et al. Maturing reticulocytes internalize plasma membrane in glycophorin A-containing vesicles that fuse with autophagosomes before exocytosis. *Blood*. 2012;119(26):6296-6306.
16. Hu J, Liu J, Xue F, et al. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood*. 2013;121(16):3246-3253.
17. Mettananda S, Clark K, Fisher CA, Sloane-Stanley JA, Gibbons RJ, Higgs DR. Phenotypic and molecular characterization of a serum-free miniature erythroid differentiation system suitable for high-throughput screening and single-cell assays. *Exp Hematol*. 2018;60:10-20.

18. Ludwig LS, Lareau CA, Bao EL, et al. Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. *Cell Rep.* 2019;27(11):3228-3240.
19. Satchwell TJ, Hawley BR, Bell AJ, Ribeiro ML, Toye AM. The cytoskeletal binding domain of band 3 is required for multiprotein complex formation and retention during erythropoiesis. *Haematologica.* 2015;100(1):133-142.
20. Satchwell TJ, Bell AJ, Hawley BR, et al. Severe Ankyrin-R deficiency results in impaired surface retention and lysosomal degradation of RhAG in human erythroblasts. *Haematologica.* 2016;101(9):1018-1027.
21. Satchwell TJ, Pellegrin S, Bianchi P, et al. Characteristic phenotypes associated with congenital dyserythropoietic anemia (type II) manifest at different stages of erythropoiesis. *Haematologica.* 2013;98(11):1788-1796.
22. Ulirsch JC, Lareau C, Ludwig LS, Mohandas N, Nathan DG, Sankaran VG. Confounding in ex vivo models of Diamond-Blackfan anemia. *Blood.* 2017;130(9):1165-1168.
23. Riffelmacher T, Clarke A, Richter FC, et al. Autophagy-Dependent Generation of Free Fatty Acids Is Critical for Normal Neutrophil Differentiation. *Immunity.* 2017;47(3):466-480.
24. Buenrostro JD, Wu B, Litzenburger UM, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature.* 2015;523(7561):486-490.
25. 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.
26. Downes DJ, Schwessinger R, Hill SJ, et al. An integrated platform to systematically identify causal variants and genes for polygenic human traits. *bioRxiv.* 2020;813618.
27. Palii CG, Cheng Q, Gillespie MA, et al. Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. *Cell Stem Cell.* 2019;24(5):812-820.
28. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* 2018 Dec 3. [Epub ahead of print]
29. Brown JM, Leach J, Reittie JE, et al. Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol.* 2006;172(2):177-187.
30. Floutsakou I, Agrawal S, Nguyen TT, Seoighe C, Ganley AR, McStay B. The shared genomic architecture of human nucleolar organizer regions. *Genome Res.* 2013;23(12):2003-2012.
31. Brown J, Saracoglu K, Uhrig S, Speicher MR, Eils R, Kearney L. Subtelomeric chromosome rearrangements are detected using an innovative 12-color FISH assay (M-TEL). *Nat Med.* 2001;7(4):497-501.
32. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012;9(7):676-682.
33. Heimpel H, Schwarz K, Ebnother M, et al. Congenital dyserythropoietic anemia type I (CDA I): molecular genetics, clinical appearance, and prognosis based on long-term observation. *Blood.* 2006;107(1):334-340.
34. Heimpel H, Kellermann K, Neuschwander N, Hogel J, Schwarz K. The morphological diagnosis of congenital dyserythropoietic anemia: results of a quantitative analysis of peripheral blood and bone marrow cells. *Haematologica.* 2010;95(6):1034-1036.
35. Kay AW, Strauss-Albee DM, Blish CA. Application of Mass Cytometry (CyTOF) for Functional and Phenotypic Analysis of Natural Killer Cells. *Methods Mol Biol.* 2016;1441:13-26.
36. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proc Fifth Berkley Sympon Math Statist and Prob.* 1967;1:281-297.

37. Tamary H, Shalev H, Perez-Avraham G, et al. Elevated growth differentiation factor 15 expression in patients with congenital dyserythropoietic anemia type I. *Blood*. 2008;112(13):5241-5244.
38. Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 2016;48(10):1193-1203.
39. Blank V, Kim MJ, Andrews NC. Human MafG is a functional partner for p45 NF-E2 in activating globin gene expression. *Blood*. 1997;89(11):3925-3935.
40. Oyake T, Itoh K, Motohashi H, et al. Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. *Mol Cell Biol*. 1996;16(11):6083-6095.
41. Sun J, Brand M, Zenke Y, Tashiro S, Groudine M, Igarashi K. Heme regulates the dynamic exchange of Bach1 and NF-E2-related factors in the Maf transcription factor network. *Proc Natl Acad Sci U S A*. 2004;101(6):1461-1466.
42. Jao CY, Salic A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc Natl Acad Sci U S A*. 2008;105(41):15779-15784.
43. Bohnsack KE, Bohnsack MT. Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J*. 2019;38(13):e100278.
44. Warner JR. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci*. 1999;24(11):437-440.
45. Roy NB, Wilson EA, Henderson S, et al. A novel 33-Gene targeted resequencing panel provides accurate, clinical-grade diagnosis and improves patient management for rare inherited anaemias. *Br J Haematol*. 2016;175(2):318-330.
46. Tanno T, Noel P, Miller JL. Growth differentiation factor 15 in erythroid health and disease. *Curr Opin Hematol*. 2010;17(3):184-190.
47. Heuston EF, Keller CA, Lichtenberg J, et al. Establishment of regulatory elements during erythro-megakaryopoiesis identifies hematopoietic lineage-commitment points. *Epigenetics Chromatin*. 2018;11(1):22.
48. Andrews NC, Kotkow KJ, Ney PA, Erdjument-Bromage H, Tempst P, Orkin SH. The ubiquitous subunit of erythroid transcription factor NF-E2 is a small basic-leucine zipper protein related to the v-maf oncogene. *Proc Natl Acad Sci U S A*. 1993;90(24):11488-11492.
49. Gillespie MA, Palii CG, Sanchez-Taltavull D, et al. Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. *Mol Cell*. 2020;78(5):960-974.
50. Gautier EF, Ducamp S, Leduc M, et al. Comprehensive Proteomic Analysis of Human Erythropoiesis. *Cell Rep*. 2016;16(5):1470-1484.
51. Nakhoul H, Ke J, Zhou X, Liao W, Zeng SX, Lu H. Ribosomopathies: mechanisms of disease. *Clin Med Insights Blood Disord*. 2014;7:7-16.
52. Andersen JS, Lyon CE, Fox AH, et al. Directed proteomic analysis of the human nucleolus. *Curr Biol*. 2002;12(1):1-11.
53. Tsekrekou M, Stratigi K, Chatzinikolaou G. The Nucleolus: In Genome Maintenance and Repair. *Int J Mol Sci*. 2017;18(7):1411.
54. Paredes S, Maggert KA. Ribosomal DNA contributes to global chromatin regulation. *Proc Natl Acad Sci U S A*. 2009;106(42):17829-17834.
55. Zhao B, Mei Y, Schipma MJ, et al. Nuclear Condensation during Mouse Erythropoiesis Requires Caspase-3-Mediated Nuclear Opening. *Dev Cell*. 2016;36(5):498-510.
56. Baron MH, Barminko J. Chromatin Condensation and Enucleation in Red Blood Cells:

An Open Question. *Dev Cell*. 2016;36(5):481-482.

57. Ney PA. Normal and disordered reticulocyte maturation. *Curr Opin Hematol*. 2011;18(3):152-157.

58. Guetg C, Lienemann P, Sirri V, et al. The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. *EMBO J*. 2010;29(13):2135-2146.

59. Gambale A, Iolascon A, Andolfo I, Russo R. Diagnosis and management of congenital dyserythropoietic anemias. *Expert Rev Hematol*. 2016;9(3):283-296.

60. Daftari P, Gavva NR, Shen CK. Distinction between AP1 and NF-E2 factor-binding at specific chromatin regions in mammalian cells. *Oncogene*. 1999;18(39):5482-5486.

## Figure Legends

### **Figure 1. Differentiating erythroblasts from healthy donors and CDA-I patients are broadly equivalent by immunophenotyping.**

(A) Representative FACS profiles of cultured erythroblasts from a healthy donor, CDA-I patient with a mutation in *CDIN1* and CDA-I patient with a mutation in *CDAN1*. (B) UMAP plots showing CyTOF data from healthy donors (n=3) and *CDIN1*-patient derived erythroblasts (n=3) at day 11 of differentiation for the erythroid markers CD235 (glycophorin A), CD36 (Scavenger receptor), CD71 (Transferrin receptor) and transcription factor Gata1. There is no difference in the clustering patterns observed between healthy donors and *CDIN1* patients for any of the 25 markers tested (Supplementary Table 6).

### **Figure 2. Erythroblasts from CDA-I patients display delayed differentiation and increased proliferation.**

(A) Cell morphology counts from cytopins on days 10 (n=101-348), day 13 (n=92-607) and day 17 (n=63-329) of erythroblasts divided into Proerythroblast (Pro), Basophilic (Baso), Polychromatic (Poly), Orthochromatic (Ortho) and enucleated (Enuc). Data are shown as mean±SD. Statistical significance was tested using Mann-Whitney with a Benjamani-Hochberg adjustment where Q=0.05. *CDIN1* patients and *CDAN1* patients were tested separately against healthy donors for each timepoint. *CDIN1* \*P=0.0095 at day 17 for both polychromatic and enucleated erythroblasts. *CDAN1* \*P=0.0087 for day 10 basophilic erythroblasts and \*\*P=0.0043 for day 10 polychromatic erythroblasts. (B) Representative cytospin images stained with modified Wright's stain (magnification 40x) for healthy donors, *CDIN1* and *CDAN1* patients at days 10, 13 and 17, with marked examples of cell types scored for (A). (C) Proliferation of cultured erythroblasts from healthy donors (n=6), patients with mutations in *CDIN1* (n=3) and patients with mutations in *CDAN1* (n=5) showing increased proliferation in both patient cohorts. All scores are for viable cells only (see Supplemental Figure 10), normalised to 100,000 cells at day 5. Data are shown as mean±SEM. Dashed vertical lines denote the three culture phases. Statistical significance was tested using Mann-Whitney with a Benjamani-Hochberg adjustment where Q=0.05. *CDIN1* patients and *CDAN1* patients were tested separately against healthy donors for each timepoint. *CDIN1* \*\*P=0.0043 at day 7

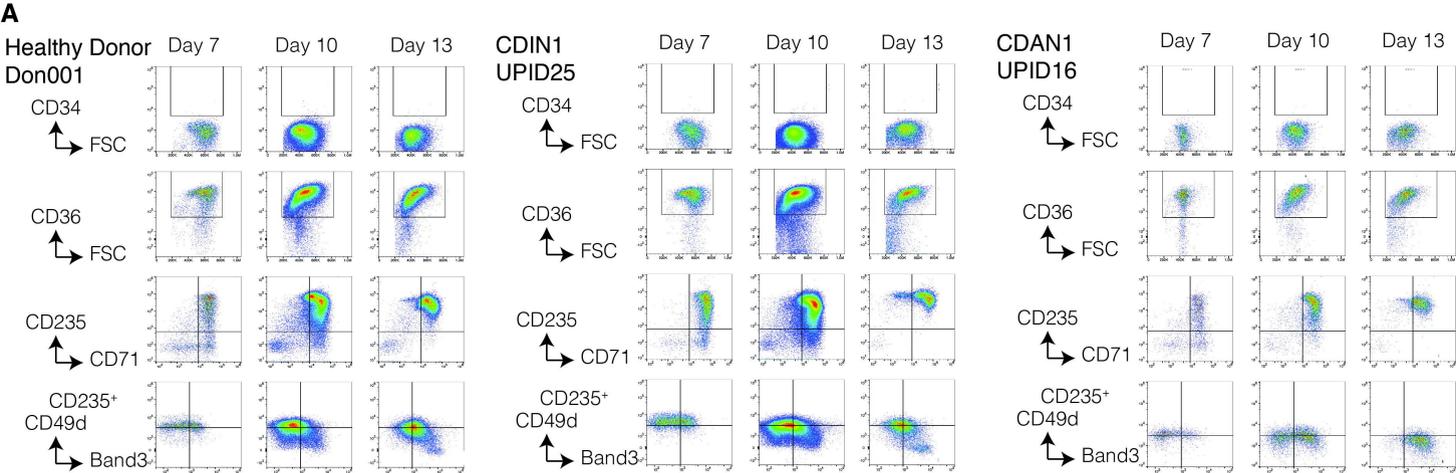
and at day 17. *CDAN1* \* $P=0.0087$  at day 6 and \*\* $P=0.0043$  at day 17. Cell counts from two patients were not scored in a comparable manner for this analysis.

**Figure 3. Evidence of defective differentiation in patient erythroblasts, most severe for *CDIN1* mutations.** (A) Transmission electron microscopy at day 11 of healthy and CDA-I cultured erythroblasts showing the diagnostic phenotype of abnormal chromatin in patients. Inset shows enlarged area to illustrate the pattern of euchromatin and heterochromatin and how this is disrupted in CDA-I patients. (B) Percentage of nuclei  $\pm$ SD with spongy heterochromatin at day 11, determined from large field images. Numbers of nuclei scored were 96-436 per individual. \* $P=0.0189$  with Kruskal-Wallis test. (C) FACS histograms (gated on all viable single cells in the CD235a<sup>+</sup> population (nucleated and enucleated)) of Band 3 intensity at day 17 (healthy donors  $n=6$ , *CDIN1* patients  $n=3$ , and *CDAN1* patients  $n=4$ ). (D) Median fluorescence intensity (MFI)  $\pm$ SD of Band 3-FITC at day 17 (\*\*  $P=0.0088$  and \*  $P=0.0127$  with Kruskal-Wallis test).

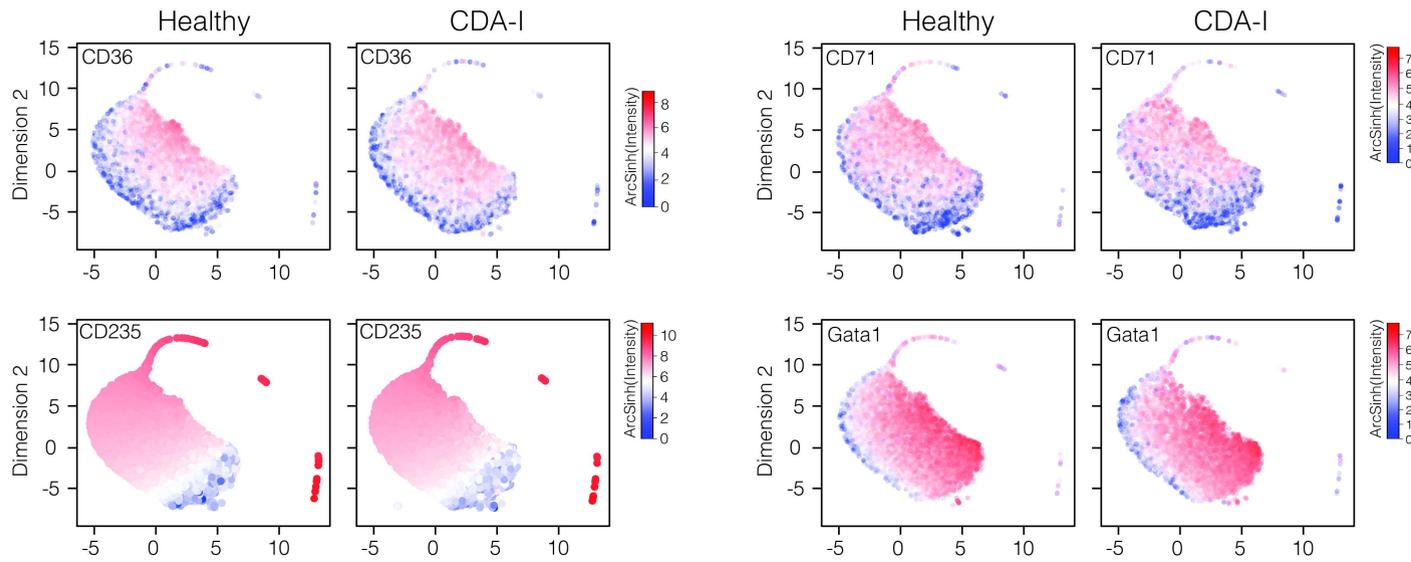
**Figure 4. ATAC-seq analysis reveals the emergence of an altered regulatory landscape in patient-derived erythroblasts.** (A) Principal component analysis (PCA) of ATAC-seq from healthy donors ( $n=6$ ) *CDIN1* patients ( $n=3$ ) and *CDAN1* patients ( $n=4$ ). The distribution of cells along PC1 follows differentiation stage and PC2 distinguishes patients and healthy donors. (B) MA plot for DESeq2 comparison of ATAC-seq from healthy donors and CDA-I patients at Day 10 and 13 of *ex vivo* differentiation with significantly different peaks ( $q<0.01$ ) highlighted as either more accessible in patients (red - up) or less accessible in patients (blue - down). (C) Comparison of chromatin state annotations<sup>26</sup> for differentially accessible peaks shows enrichment for enhancers in less accessible peaks. Strong and weak refers to the level of H3K27ac signal. (D) MEME motif discovery identified a motif matching that of NF-E2<sup>48, 60</sup> as being significantly enriched (E-value  $< 10^{-50}$ ) in ATAC-seq peaks that were less accessible in patients.

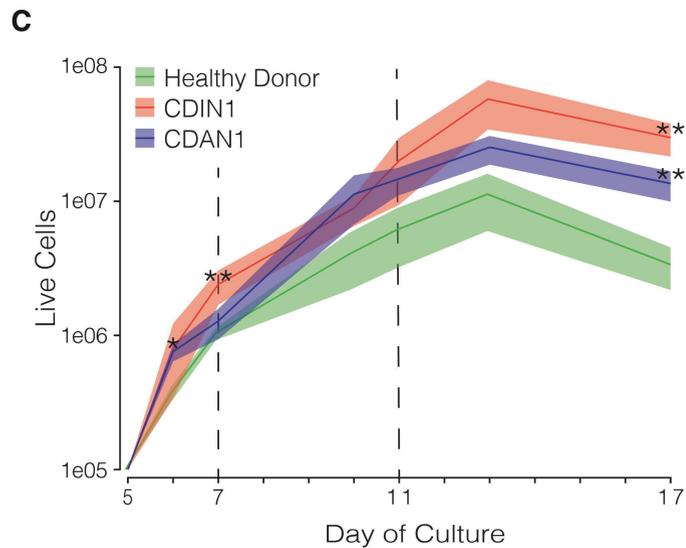
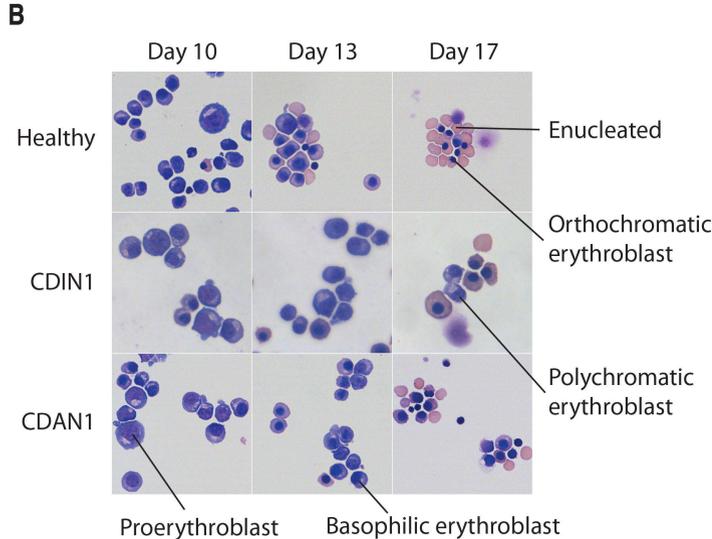
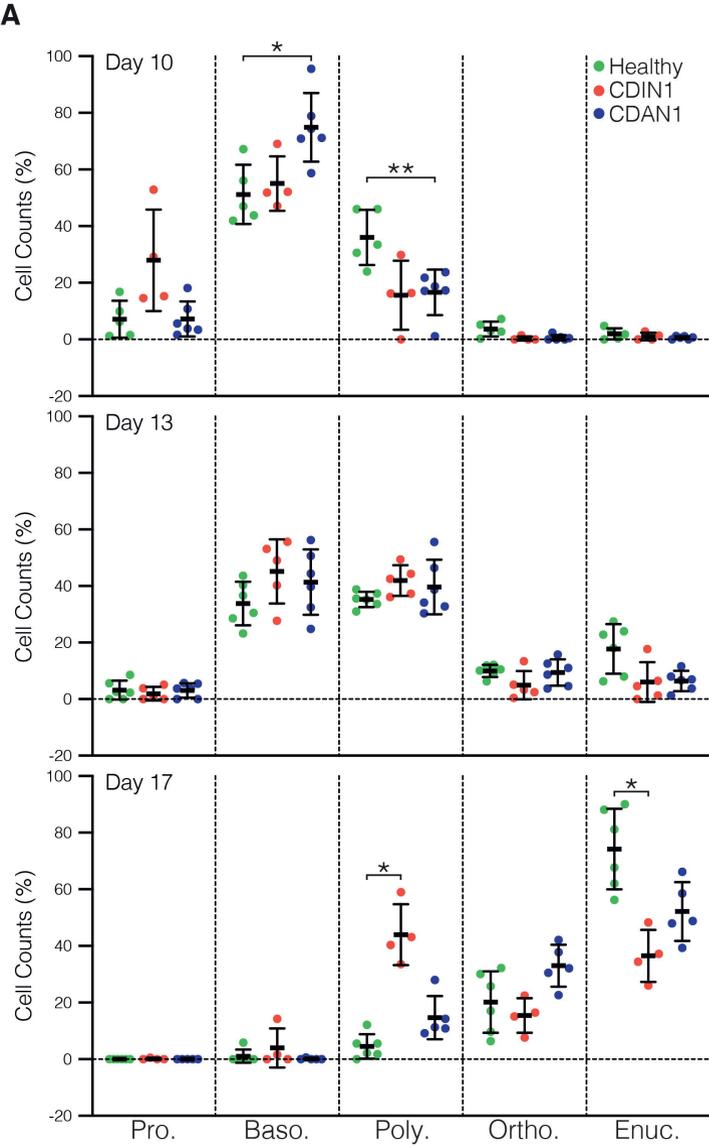
**Figure 5. Nucleolar structure is disrupted in CDA-I patients.** (A) Nucleoli detected with probe BAC-CTD (green) surrounding rDNA probe p7.1 (red) in healthy nuclei. In example patients UPID20 (*CDAN1* mutation) and UPID25 (*CDIN1* mutation) this order is disrupted. (B) Nucleolar output, judged by EU labelling, is

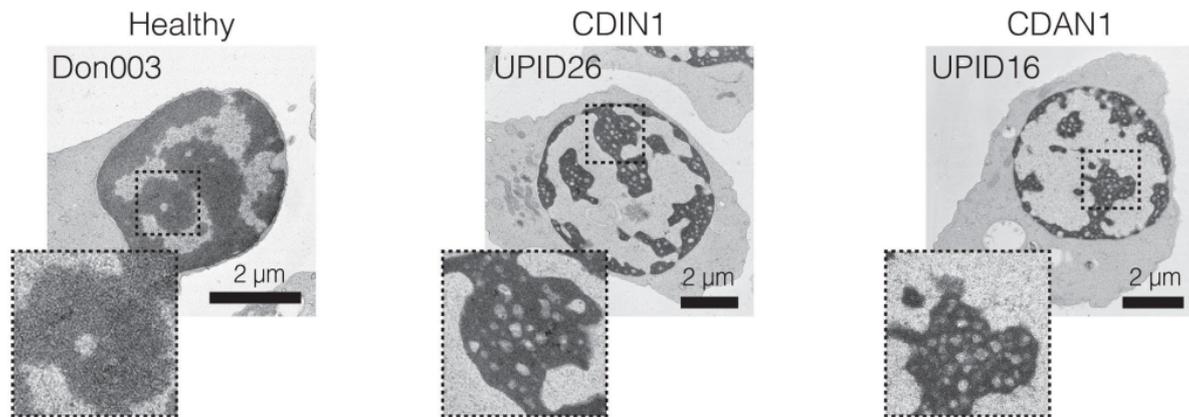
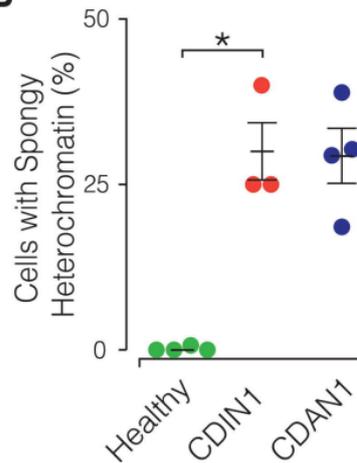
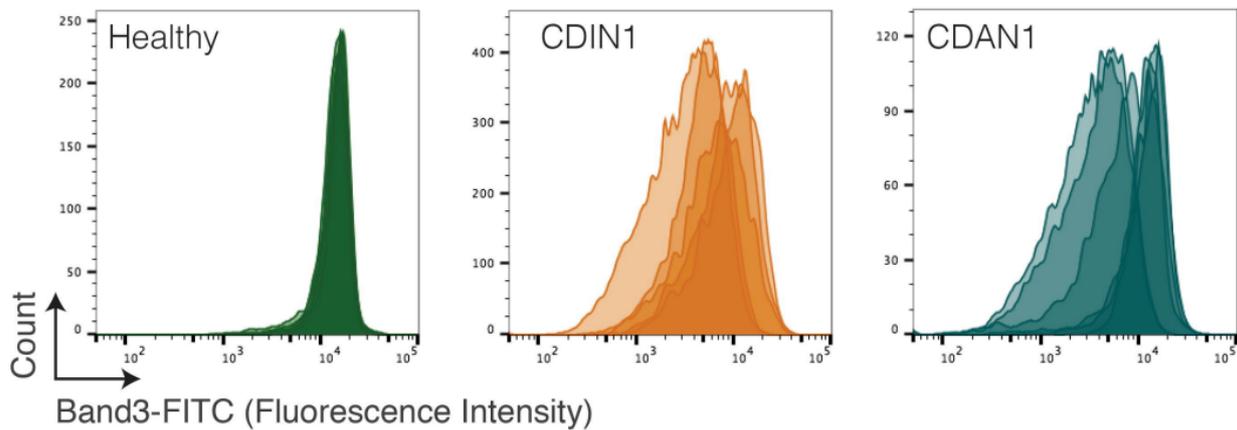
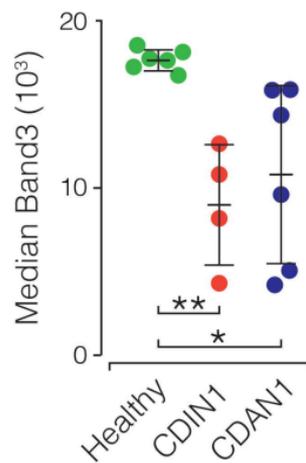
significantly reduced in two CDA-I patients with CDAN1 mutations, each compared with a healthy donor in the same experiment. (C) Despite the disrupted nucleolar structure, CDIN1 and Codanin-1 proteins (green) continue to associate with nucleolar proteins UBF and Fibrillarin (red) respectively, in two example CDA-I patients with CDAN1 mutations. All analyses are on day 10/11 cultured erythroblasts, using DAPI counterstain (blue).

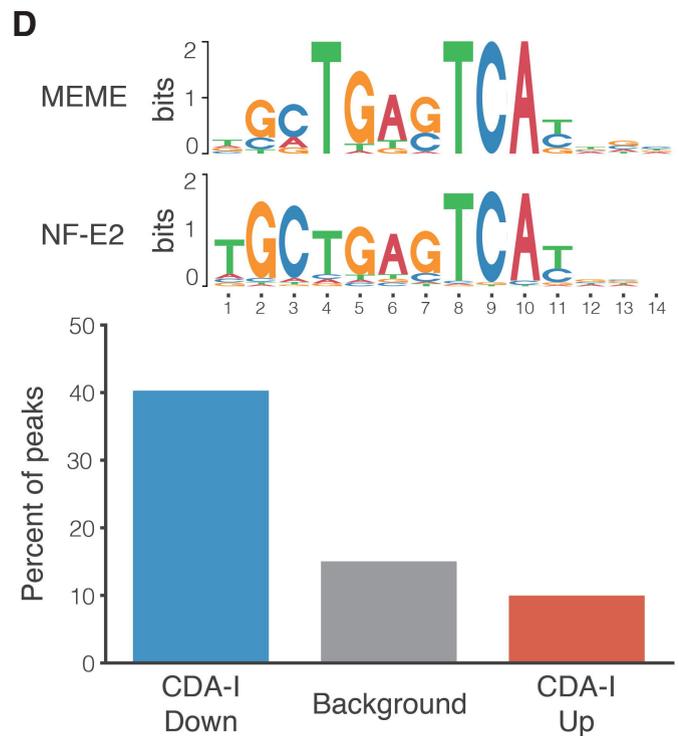
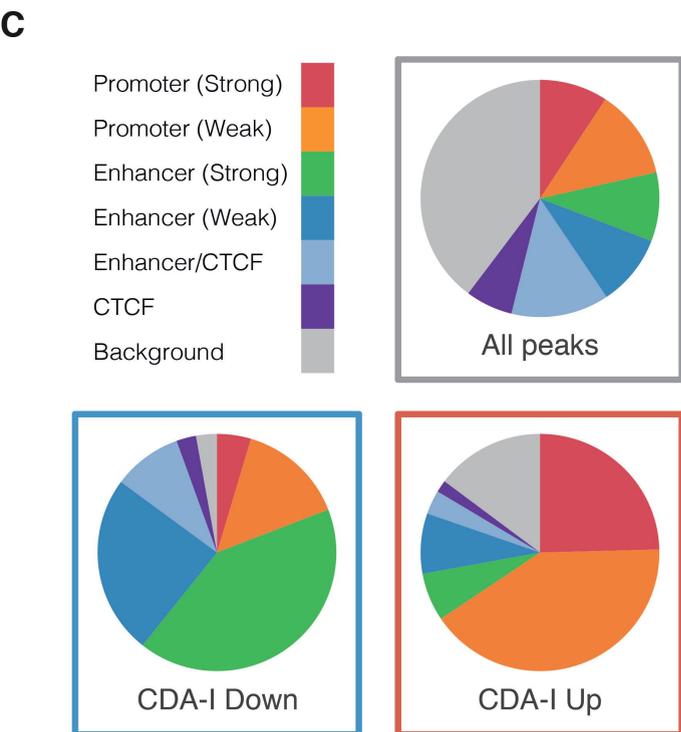
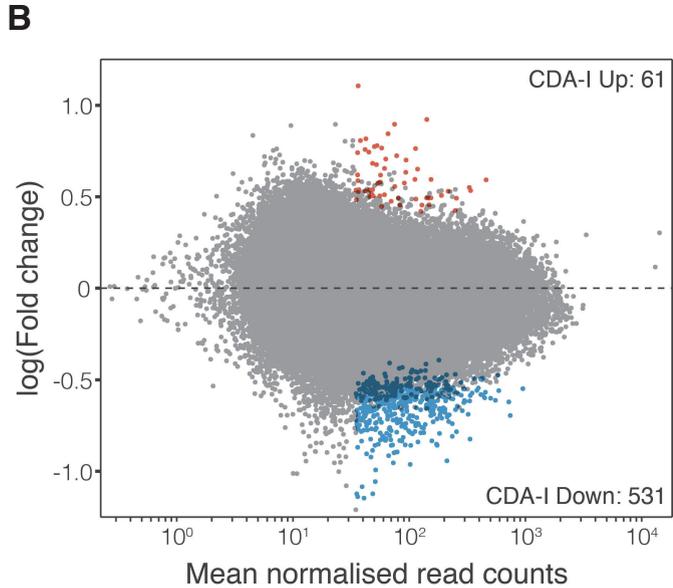
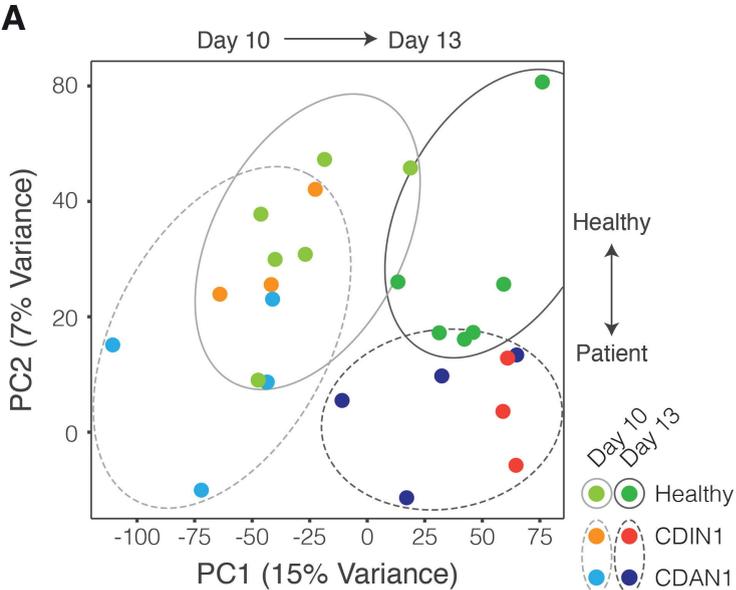


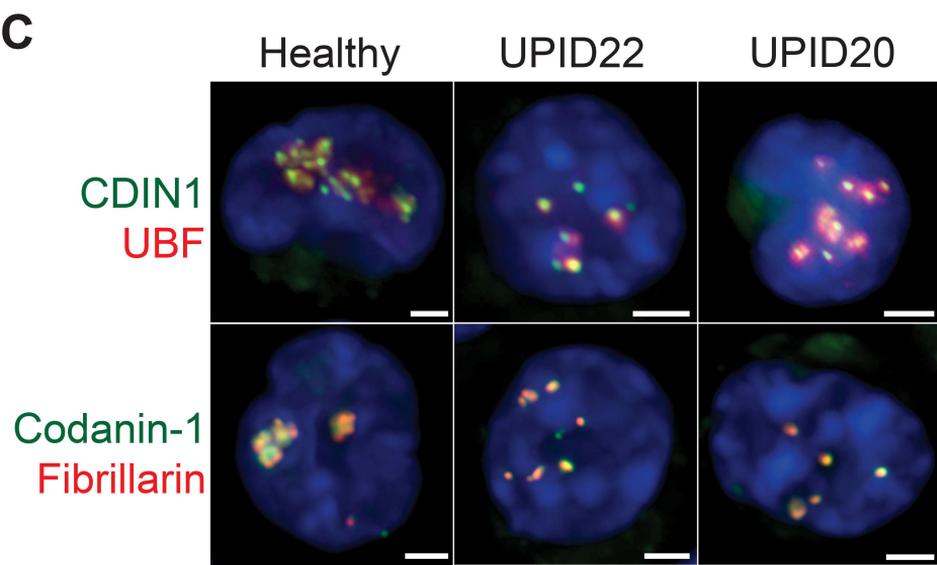
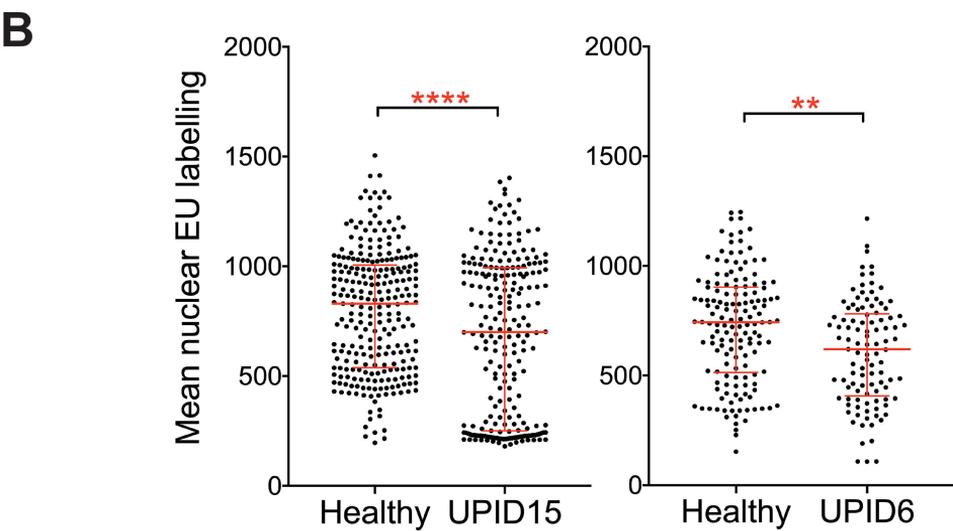
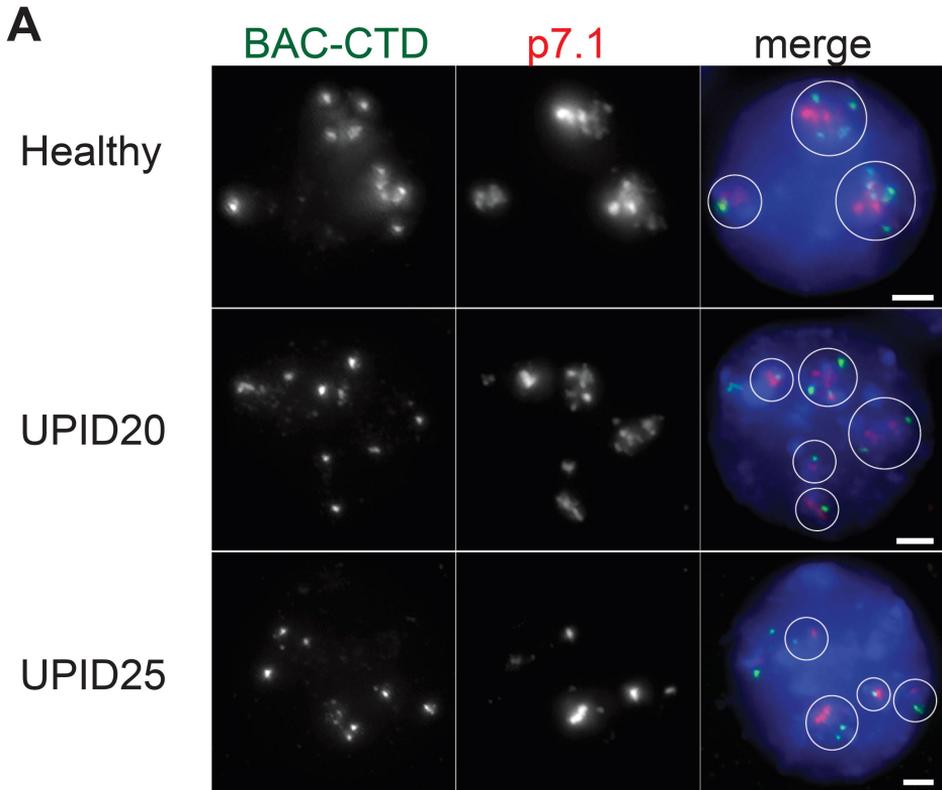
**B**





**A****B****C****D**





**Recapitulation of erythropoiesis in congenital dyserythropoietic anaemia type I (CDA-I) identifies defects in differentiation and nucleolar abnormalities**

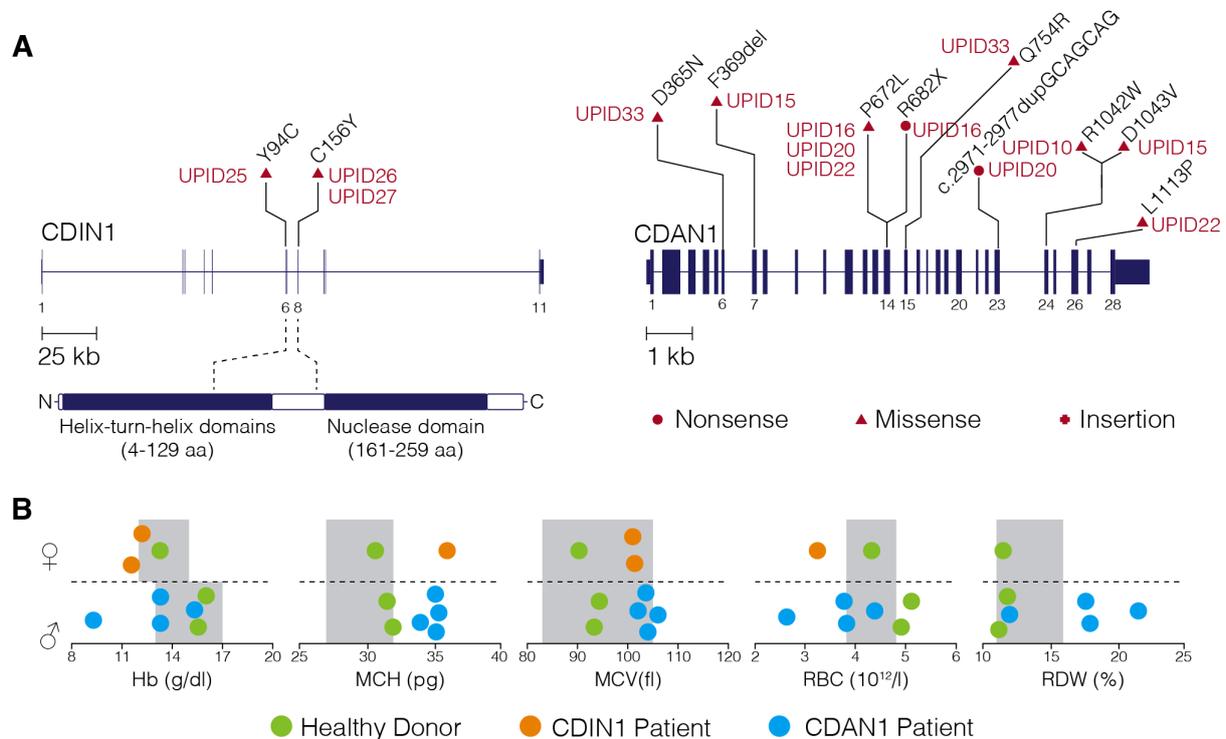
**Supplemental Data**

Caroline Scott<sup>1</sup>, Damien J. Downes<sup>1</sup>, Jill M. Brown<sup>1</sup>, Robert Beagrie<sup>1</sup>, Aude-Anais Olijnik<sup>1</sup>, Matthew Gosden<sup>1</sup>, Ron Schwessinger<sup>1</sup>, Christopher A. Fisher<sup>1</sup>, Anna Rose<sup>1</sup>, David J.P. Ferguson<sup>2</sup>, Errin Johnson<sup>3</sup>, Quentin A. Hill<sup>4</sup>, Steven Okoli<sup>5</sup>, Raffaele Renella<sup>6</sup>, Kate Ryan<sup>7</sup>, Marjorie Brand<sup>8</sup>, Jim Hughes<sup>1</sup>, Noemi B.A. Roy<sup>9</sup>, Douglas R. Higgs<sup>1</sup>, Christian Babbs<sup>1</sup> and Veronica J. Buckle<sup>1</sup>.

<sup>1</sup>Weatherall Institute of Molecular Medicine, Oxford University, Oxford, United Kingdom; <sup>2</sup>Ultrastructural Morphology Group, NDCLS, John Radcliffe Hospital, Oxford, United Kingdom; <sup>3</sup>Sir William Dunn School of Pathology, South Parks Rd, Oxford, United Kingdom; <sup>4</sup>Leeds Teaching Hospital NHS Trust, United Kingdom; <sup>5</sup>Imperial College, The Commonwealth Building, The Hammersmith Hospital, Du Cane Rd, London, United Kingdom; <sup>6</sup>Pediatric Hematology-Oncology Research Laboratory, CHUV-UNIL Lausanne Switzerland; <sup>7</sup>Department of Haematology, Manchester Royal Infirmary, Oxford Rd, Manchester, United Kingdom; <sup>8</sup>Sprott Center for Stem Cell Research, Ottawa Hospital Research Institute, Ottawa, Canada and <sup>9</sup>Department of Haematology, Oxford University Hospitals NHS Trust, Churchill Hospital, Old Rd, Headington, and NIHR Biomedical Research Centre, Oxford, United Kingdom.

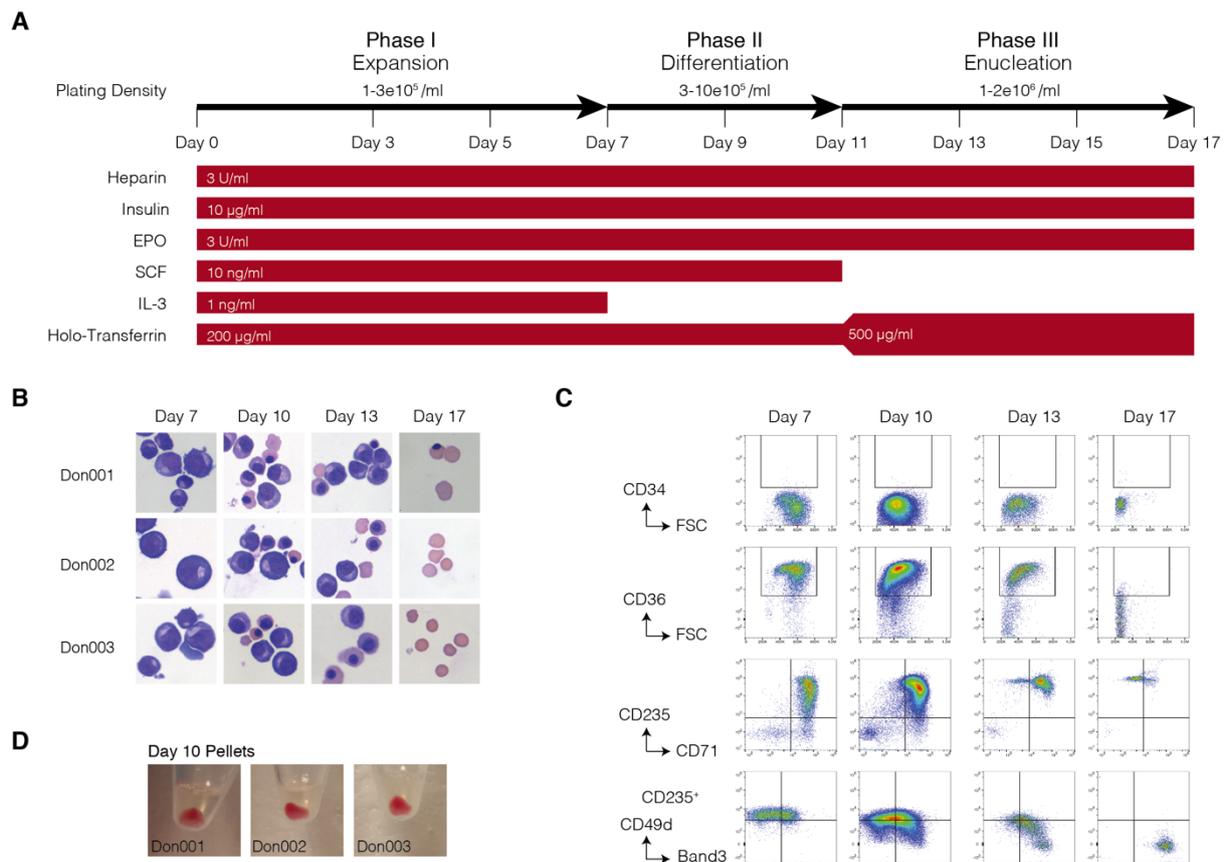
## Supplemental Figures

### Supplemental Figure 1



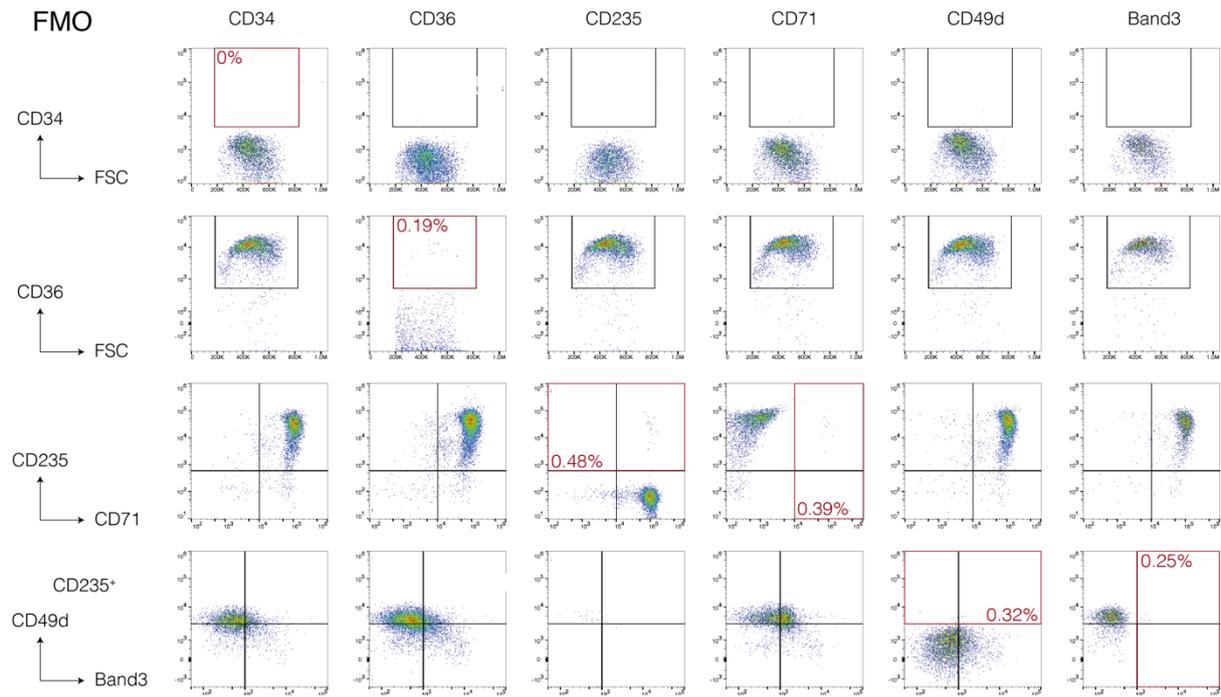
**Supplemental Figure 1. CDA-I patient mutations and clinical data.** (A) The location of pathogenic mutations associated with *CDIN1* and *CDAN1* of patients used in this study. Each patient has been given a unique personal identifier (UPID) and the type of mutation indicated. (B) Haematology of healthy donors (green dots) CDA-I patients with *CDIN1* mutations (orange dots) and with *CDAN1* mutations (blue dots) where full blood counts available. Grey boxes represent the Oxford University Hospital, NHS Foundation Trust normal ranges for adults. The gender of our cohort is also indicated. Haemoglobin (Hb), mean cell haemoglobin (MCH), mean cell volume (MCV), red blood cell count (RBC) and red cell distribution width (RDW). Data for UPID22 and UPID25 were omitted due to patients being venesected and transfused, respectively. For UPID10, 26 and 33 incomplete data were available.

## Supplemental Figure 2



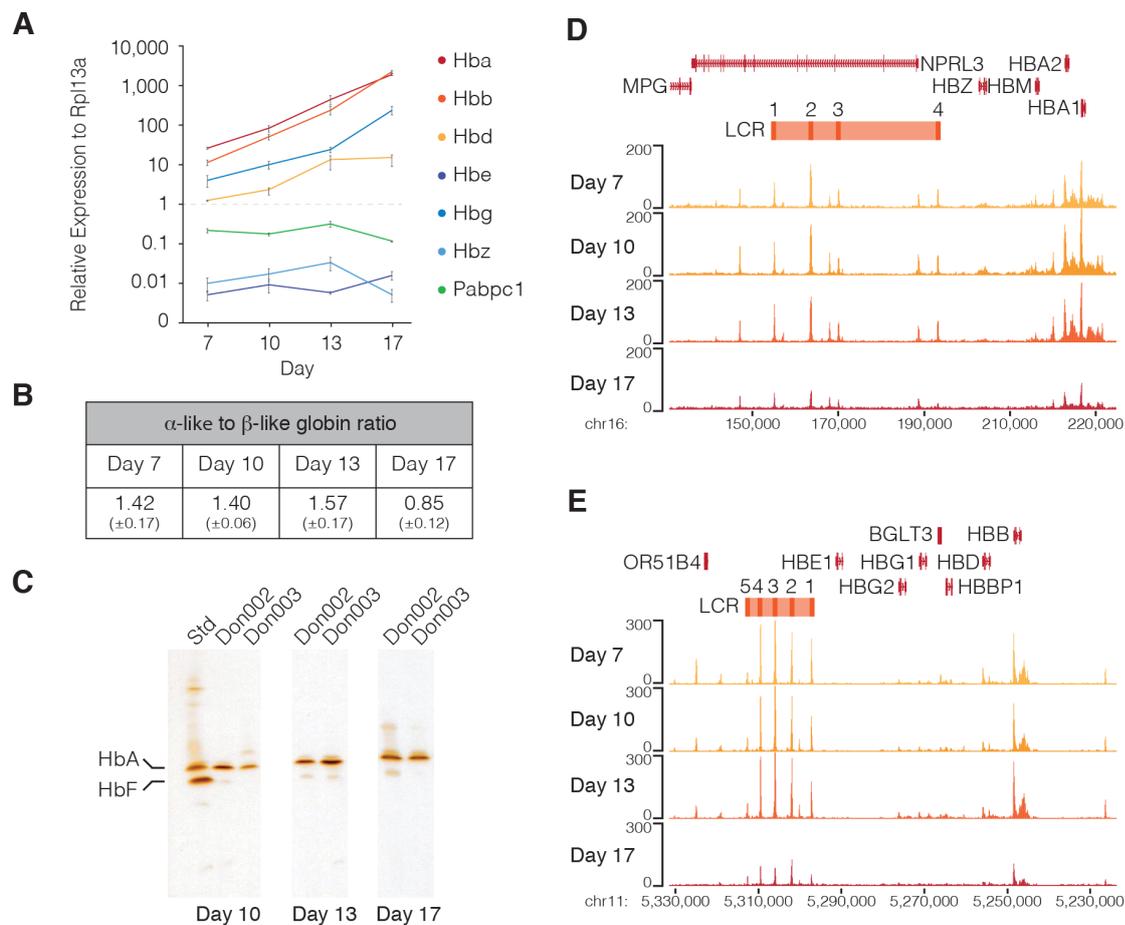
**Supplemental Figure 2. Characterisation of protocol for expansion and differentiation of CD34<sup>+</sup>HSPCs from healthy donors.** (A) Schematic of experimental approach for three-phase culture protocol. A common base media IMDM (Source BioScience UK Ltd) containing 3% (v/v) AB Serum, 10 µg mL<sup>-1</sup>, insulin, 3 U mL<sup>-1</sup> heparin (all from Sigma-Aldrich, Poole, UK), 2% (v/v) fetal bovine serum (Gibco) was supplemented as shown. (B) Representative cytopspins stained with modified Wright's stain (magnification 40x) showing cell morphology during erythroid expansion (day 7), differentiation (day 10 and 13) and enucleation (day 17). (C) Immunophenotyping of cultured erythroblasts from peripheral blood of healthy donors (n=3) using a 6-colour antibody panel at days 7, 10, 13 & 17 of differentiation. Gates were set using fluorescence minus one (FMO) controls (see Supplemental Fig 3). (D) Images of red cell pellets from day 10 cultured erythroblasts.

### Supplemental Figure 3



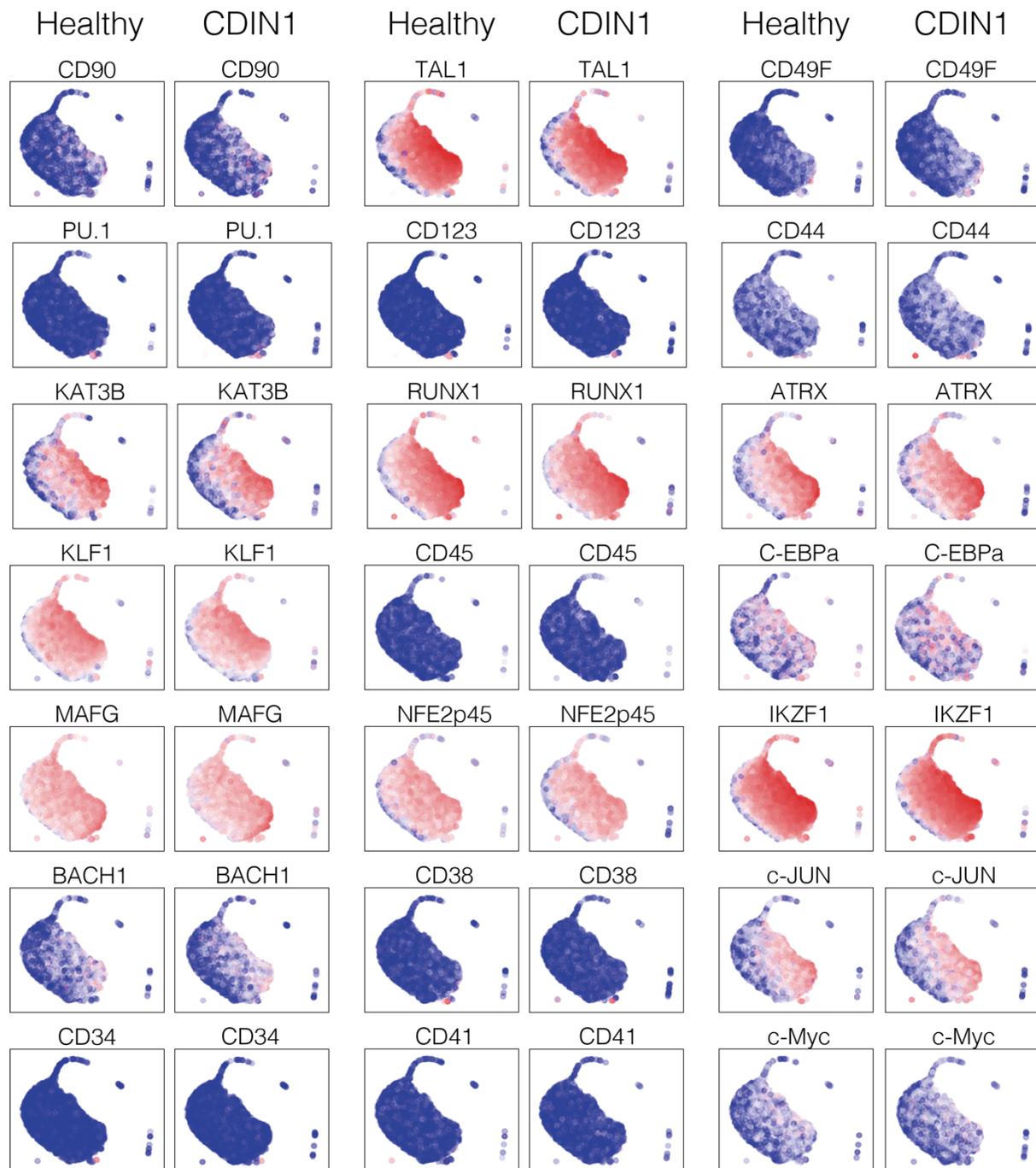
**Supplemental Figure 3. Fluorescence minus one (FMO) controls for FACS analysis.** Gating strategy for FACS analysis using FlowJo v10.4.2. Gates were set for each population on FMO. Red boxes indicated positive population for each marker and its relative percentage.

## Supplemental Figure 4



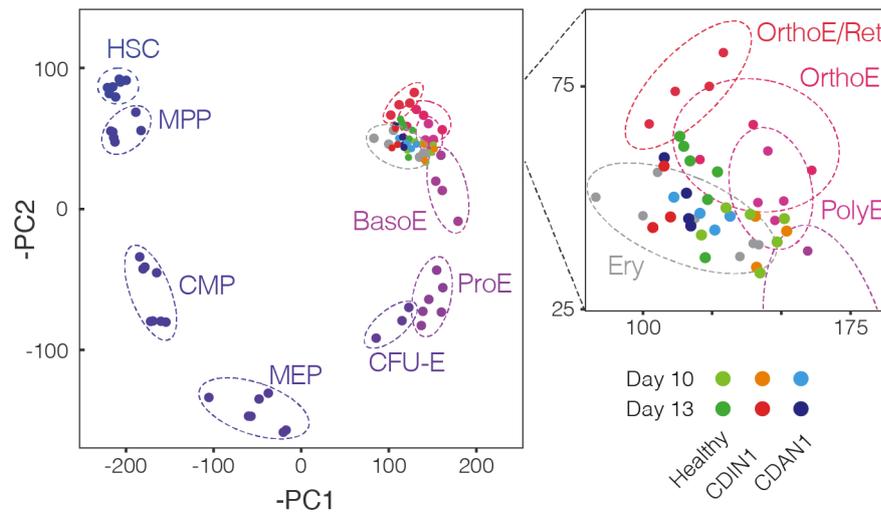
**Supplemental Figure 4. Cultured erythroblasts from healthy donors show an increase in DNA accessibility and expression of the adult globin genes during differentiation.** (A) RNA expression analysis of globins (mean  $\pm$  SEM), throughout the differentiation normalized to RPL13A. Pabpc1 was used as a housekeeping gene. (B) Ratios of the alpha-like to beta-like globins (mean  $\pm$  SEM) during erythroid differentiation. (C) IEF of erythroblasts from 2 healthy donors (Don002 and Don003) at three timepoints during *ex vivo* differentiation. HbA is adult haemoglobin and HbF is fetal haemoglobin. (D) ATAC-seq of the alpha-globin locus at four time-points throughout *ex vivo* differentiation of healthy controls (n=3, with technical replicates). (E) ATAC-seq of the beta-globin locus at four time-points throughout *ex vivo* differentiation of healthy controls (n=3, with technical replicates).

## Supplemental Figure 5



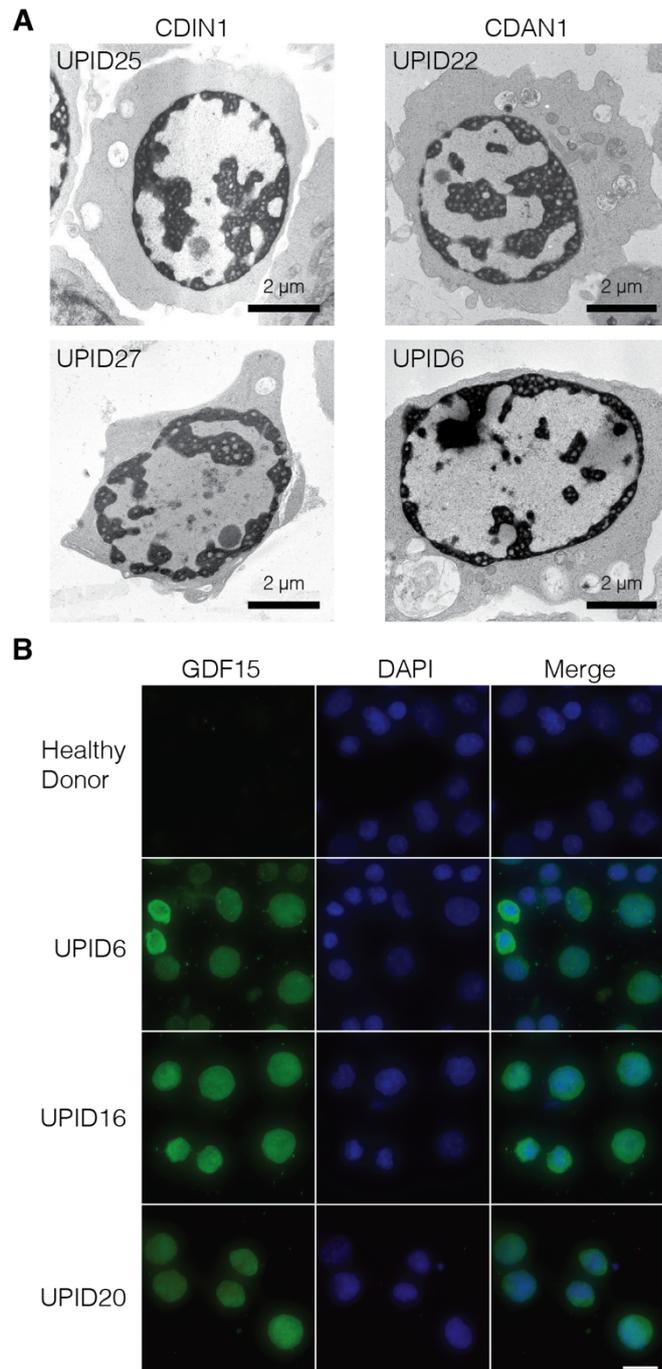
**Supplemental Figure 5. UMAP plots of CyTOF data from day 11 cultured erythroblasts.** UMAP plots showing the panel of CyTOF markers (Supplemental Table 6) analysed from healthy donors (n=3) and *CDIN1*-patient derived erythroblasts (n=3) at day 11 of differentiation. Plots are graded by colour from minimum (blue) to maximum (red) signal intensity. UMAP plots for CD235, CD71, CD36 and GATA1 shown in Figure 1B.

## Supplemental Figure 6



**Supplemental Figure 6. DNA accessibility for CDA-I patient and healthy donor cultured erythroblasts mapped against a trajectory for normal erythropoiesis.** PCA comparison of ATAC-seq for *ex vivo* differentiated erythroblasts from CDA-I patients (*CDIN1* patients n=3 and *CDAN1* patients n=4) and healthy donors (n=6) using a trajectory of immunophenotyped sorted cell-types shows a clear overlap between groups. PCA was performed using 136698 peaks from haematopoietic stem cells (HSC), multi-potent progenitors (MPP), common myeloid progenitors (CMP), myeloid-erythroid progenitors (MEP) and bulk erythrocytes (Ery) from bone marrow or peripheral blood<sup>1</sup> and erythroid colony forming units (CFU-E), pro-erythroblasts (ProE), basophilic erythroblasts (BasoE), polychromatic erythroblasts (PolyE), orthochromatic erythroblasts (OrthoE) and reticulocytes (Ret) from *ex vivo* culture.<sup>2</sup> Healthy and CDA-I patient counts for PC1 and PC2 were then calculated and mapped relevant to the sorted populations.

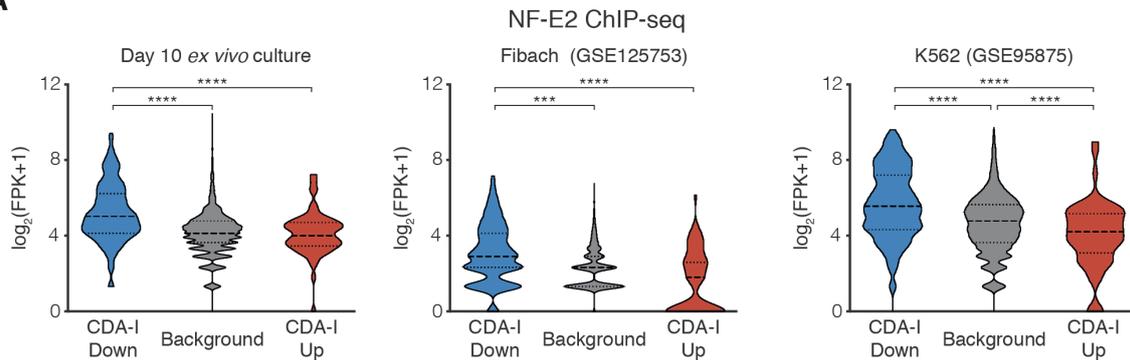
## Supplemental Figure 7



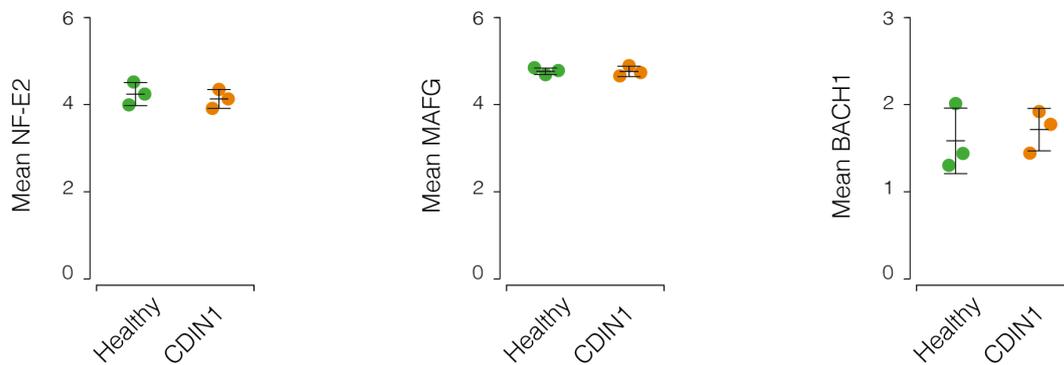
**Supplemental Figure 7. *Ex vivo* cultured patient erythroblasts show the diagnostic features of CDA-I.** (A) Representative examples of electron micrographs of abnormal nuclei seen in patients with mutations in *CDIN1* (n=2) and *CDAN1* (n=2). (B) Immunofluorescence of day 10 cultured erythroblasts from a healthy donor and *CDAN1* patients (UPID6, 16 and 20). GDF15 is detected with Alexa488 and DAPI was used as a nuclear counter stain. The merged images are shown in the right-hand panel. Bar is 8 $\mu$ m.

## Supplemental Fig 8

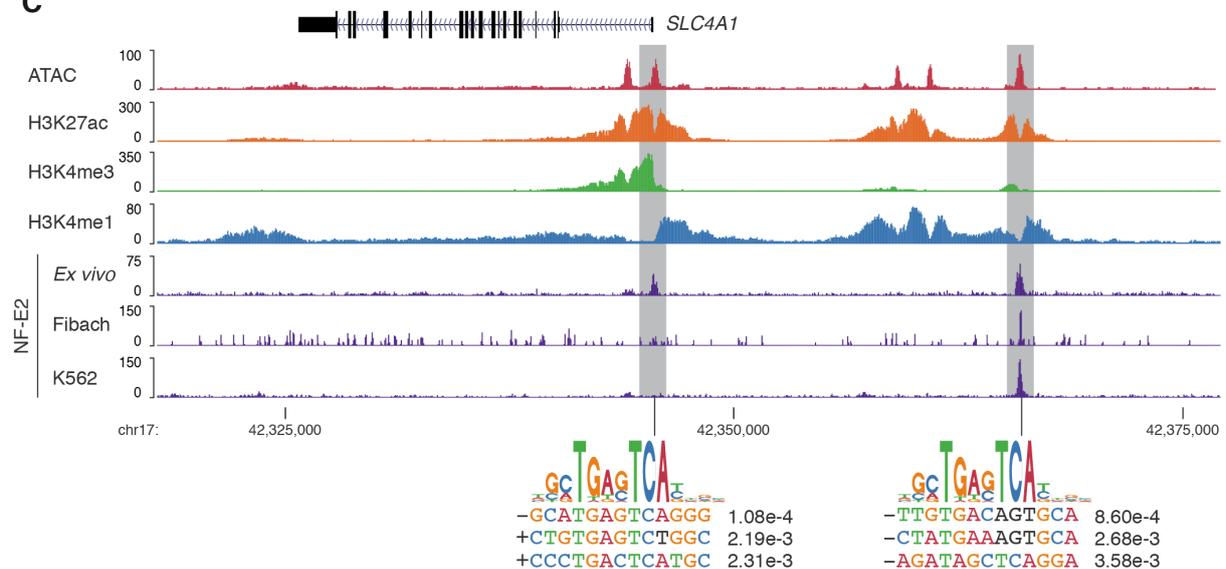
**A**



**B**



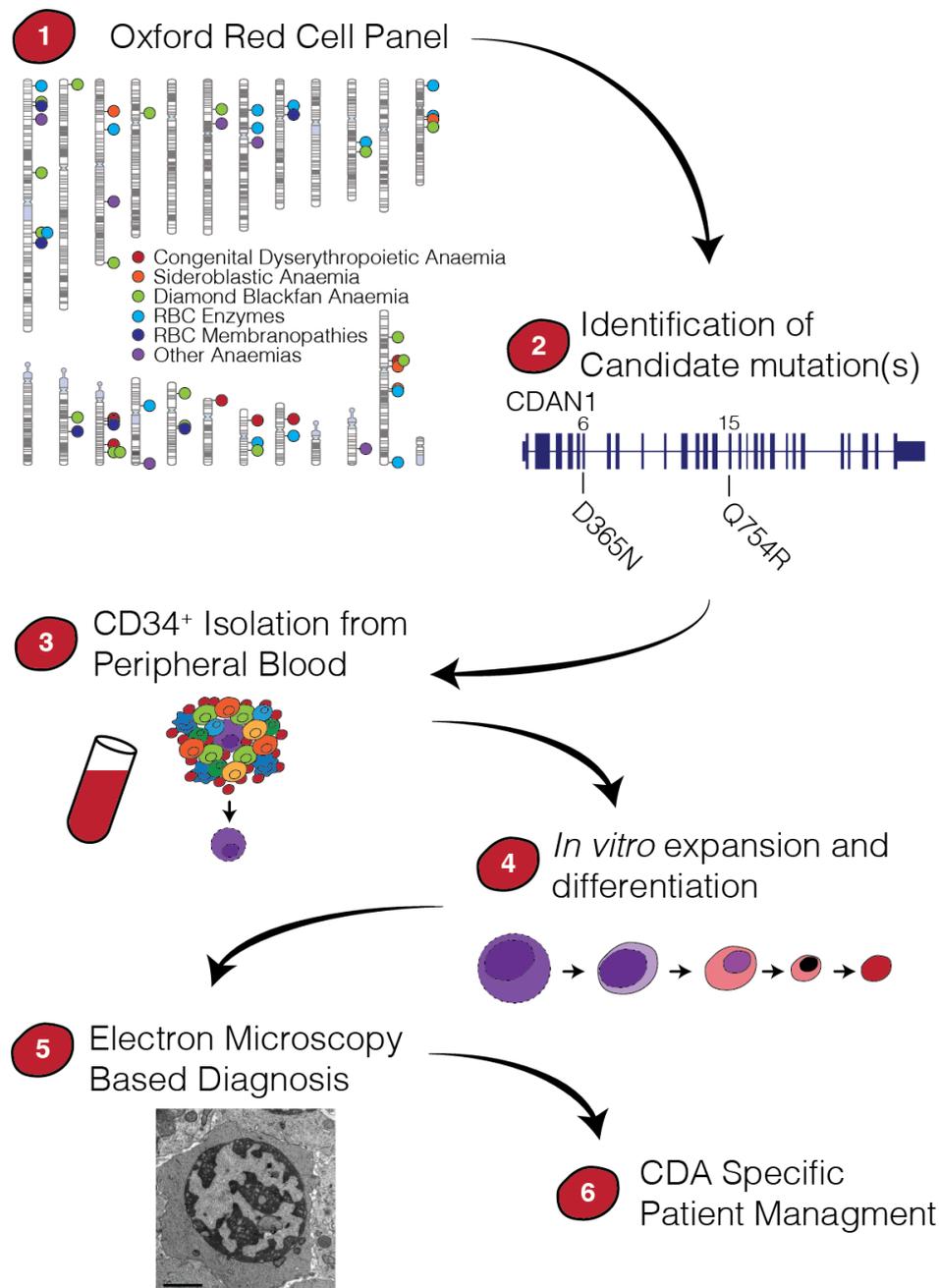
**C**



**Supplemental Fig 8. NF-E2 regulation of differentially accessible peaks and Band 3.** A) Counts of NF-E2 ChIP-seq reads over ATAC-seq peaks that had decreased accessibility (n=531), unchanged accessibility (n=3,742), or increased accessibility (n=61) in CDA-I patients. ChIP-seq is from day 10 *ex vivo* differentiation of CD34<sup>+</sup> HSPCs, erythroblasts generated from peripheral blood mononuclear cells using the Fibach method<sup>3</sup> (GSE125753), and erythroleukemia K562 cells (GSE95875). Adjusted p-values (\*\*\*\* p<0.0001, \*\*\* p =0.001) are from a Kruskal-

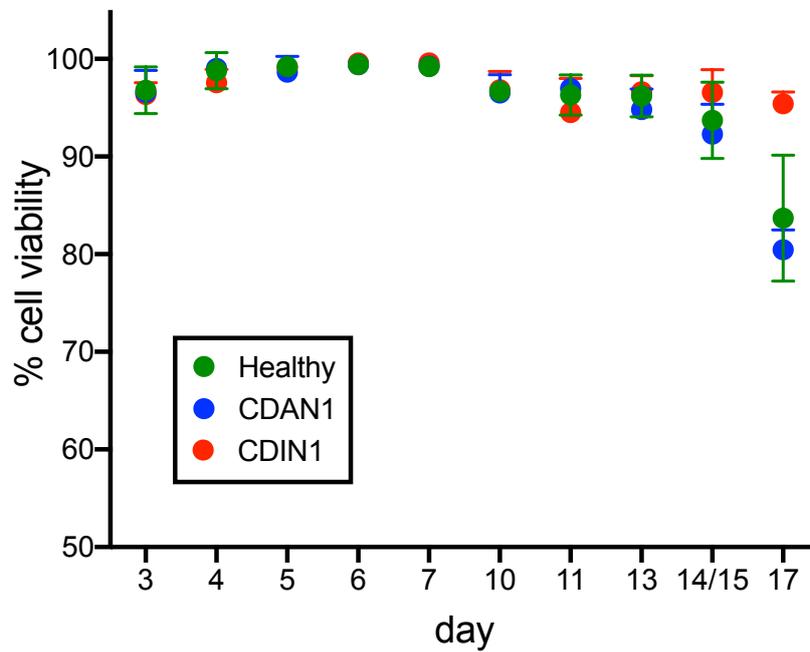
Wallis test with Dunn's multiple comparison test correction show significantly more NF-E2 at open chromatin peaks with decreased accessibility in CDA-I patients. FPK: Fragments per kilobase. B) Mean CyTOF signal for the two protein components of NF-E2 (NFE2-p45 and MAFG) and for the transcription factor BACH1 (which binds to the same motif) at day 11 of *ex vivo* differentiation of healthy donor (n=3) and CDIN1 patient (n=3) CD34<sup>+</sup> HSPCs. Bars show mean and one standard deviation. C) Chromatin landscape at the Band 3 encoding gene, *Slc4a1*, at day 10 of *ex vivo* differentiation from healthy donors, showing open chromatin (ATAC), promoter sites (H3K4me3), enhancers (H3K4me1), active transcription (H3K27ac) and NF-E2 binding. NF-E2 signal is also shown from Fibach and K562 erythroid cells. Grey bars highlight the NF-E2 bound peaks (promoter left and 5' enhancer right) with motif analysis below. P-values for FIMO comparison with NFE2\_HUMAN.H11MO.0.A are shown.

## Supplemental Figure 9



**Supplemental Figure 9. Strategy for generation of patient specific erythroblasts to validate CDA-I variants.** Mutations in *CDAN1* genes were identified from gDNA of patient peripheral blood (UPID33) using a targeted re-sequencing Oxford Red Cell Panel<sup>4</sup> and validated by demonstrating the presence of 'spongy' heterochromatin in *ex vivo* cultured erythroblasts, thus confirming the diagnosis of CDA-I.

### Supplemental Figure 10



**Supplemental Figure 10. Cell viability during *ex vivo* culture period.** Daily viability scores for *ex vivo* cultures from healthy donors (n=11) and CDA-I patients with mutations in *CDAN1* (n=6) or *CDIN1* (n=5), measured by staining with acridine orange and DAPI.

## Supplemental Tables

### Supplemental Table 1. Fluorophore conjugated antibodies used for staging the erythroid differentiation by FACS.

	Protein	Channel	Supplier	Catalogue Number
<b>CD235a</b>	Glycophorin A	PE	BD Bioscience	555570
<b>CD71</b>	Transferrin receptor	PerCP Cy5.5	Biolegend	334114
<b>CD49D</b>	$\alpha$ -Integrin	APC	BD Bioscience	561892
<b>CD34</b>	CD34	PE/Cy7	Biolegend	343616
<b>CD233</b>	Band3	FITC	IBGRL	9439FI
<b>CD36</b>	Platelet glycoprotein	APC/Cy7	Biolegend	336213
<b>Hoechst 33258</b>	Viability dye	Violet	Invitrogen	H3569

### Supplemental Table 2: TaqMan probes used for globin expression analysis.

Gene	Product Code
HBA	Hs00361191_g1
HBB	Hs00747223_g1
HBD	Hs00426283_m1
HBE	Hs00362216_m1
HBG	Hs00361131_g1
HBZ	Hs00923579_m1
RPL13a	Hs03043885_g1
PABPC1	Hs00743792_s1

**Supplemental Table 3: Next generation sequencing depth.** See file designated 'Supplemental Excel Tables'.

**Supplemental Table 4: Bed file of top 1000 ATAC peaks used for PCA plots.** See file designated 'Supplemental Excel Tables'.

**Supplemental Table 5: Bed file of ATAC-seq non-TSS nucleosome depleted regions (NDR) used for PCA plots.** See file designated 'Supplemental Excel Tables'.

### Supplemental Table 6: Pre-conjugated antibodies used for CyTOF.

Label	Antibody Target	Clone	Source	Catalog number	RRID
149Sm	CD34	581	Fluidigm	3149013B	AB_2756285
155Gd	CD36	5-271	Fluidigm	3155012B	AB_2756286
175Lu	CD71	OKT-9	Fluidigm	3175011B	AB_2756287
172Yb	CD38	HIT2	Fluidigm	3172007B	AB_2756288

143Nd	CD45RA	HI100	Fluidigm	3143006B	AB_2651156
151Eu	CD123	6H6	Fluidigm	3151001B	AB_2661794
164Dy	CD49F	G0H3	Fluidigm	3164006B	AB_2756289
161Dy	CD90	5E10	Fluidigm	3161009B	AB_2756290
153Eu	CD44	691534	Fluidigm	3153021B	AB_2756291
89Y	CD41	HIP8	Fluidigm	3089004B	AB_2756292
141Pr	CD235ab	HIR2	Fluidigm	3141001B	AB_2651154
PE	GATA1	234739	R&D Systems	C1779P	AB_2108404
156Gd*	A-PE	PE001	Fluidigm	3156005B	AB_2756294
167Er	PU1	7C6B05	Biolegend	658002	AB_2562720
160Gd	ATRX	39f	Abcam	218936	AB_2756295
176Yb	c-Myc	9E10	Fluidigm	3176012B	AB_2756296
165Ho	KLF1	1B6A3	Abcam	175372	AB_2756297
162Dy	TAL1	2TL242	Thermo	14-9101-82	AB_2572922
158Gd	RUNX1	polyclonal	Thermo	PA5-12409	AB_2184103
154Sm	NFE2p45	polyclonal	Genetex	GTX102698	AB_1950992
171Yb	BACH1	GO11-1A3	Thermo	37-0900	AB_2533297
159Tb	IKZF1	polyclonal	Thermo	PA5-23728	AB_2541228
152Sm	MAFG	polyclonal	Genetex	GTX114541	AB_10619599
173Yb	c-JUN	2HCLC	Thermo	711202	AB_2633131
166Er	KAT3B/p300	RW105	Novusbio	NB100-616	AB_10002598
145Nd	C/EBPa	polyclonal	Thermo	PA5-26487	AB_2543987

\*this antibody was used as a secondary antibody to detect the PE-labelled GATA1 antibody

**Supplemental Table 7: Antibodies used for IF.**

Antibody	Source	Dilution
goat anti-GDF15	ab-39999; Abcam	1:50
anti-goat Alexa488	A-11055; Thermo Fisher Scientific	1:300
rabbit anti-Codanin-1	Bethyl A304-951A	1:300
rabbit anti-C15orf41	Cusabio CSB-PA897474LA01HU	1:50
mouse anti-Fibrillarlin	Abcam ab4566	1:500
mouse anti-UBF	Santa Cruz F-9	1:100

donkey anti-mouse Cy3	Jackson ImmunoResearch	1:500
donkey anti-rabbit Alexa 488	ThermoFisher Scientific	1:200

**Supplemental Table 8: CDA-I patient mutations and disease severity**

<b>Patient ID</b>	<b>Gender</b>	<b>Age (yrs)</b>	<b>Gene</b>	<b>Mutation types</b>	<b>Disease severity</b>
UPID6	M	54	<i>CDAN1</i>	Unknown effect	mild
UPID10	M	46	<i>CDAN1</i>	Non LOF	moderate
UPID15	M	70	<i>CDAN1</i>	Unknown effect/Non LOF	mild
UPID16	M	50	<i>CDAN1</i>	LOF/Non LOF	mild
UPID20	M	72	<i>CDAN1</i>	LOF/Non LOF	mild
UPID22	F	30	<i>CDAN1</i>	Non LOF/Unknown effect	moderate
UPID33	M	8	<i>CDAN1</i>	Unknown effect/Unknown effect	severe
UPID25	F	46	<i>CDIN1</i>	Unknown effect	severe
UPID26	F	46	<i>CDIN1</i>	Unknown effect	mild
UPID27	F	42	<i>CDIN1</i>	Unknown effect	mild

## Supplemental Methods

*Differentiation of CD34<sup>+</sup> HSPCs:*  $1 \times 10^5$  cells were resuspended on day 0 in Phase I media (see Supplemental Figure 2A) at  $10^5$  cells  $\text{ml}^{-1}$ . Cell counts and viability was assessed throughout the differentiation using the cell count and viability assay on the Nucleocounter 3000 (Chemometec). Acridine orange was used to stain the entire population, DAPI to stain the non-viable cells and % viability calculated from these. On days 3 and 5 with additional Phase I media the cell concentration was maintained at  $2 \times 10^5$  cells  $\text{ml}^{-1}$ . On day 7, cells were counted and pelleted (400 rcf, 5 min, RT) and resuspended in Phase II media at  $2 \times 10^5$  cells  $\text{ml}^{-1}$ . Cells were counted on day 9 and diluted to  $2 \times 10^5$  cells  $\text{ml}^{-1}$  Phase II media. On day 11, cells were counted and pelleted (400 rcf, 5 min, RT) and resuspended in Phase III media at  $1 \times 10^6$  cells  $\text{ml}^{-1}$ . Cells were counted on days 13 and 15 and diluted to  $1 \times 10^6$  cells  $\text{ml}^{-1}$  in Phase III media. Live cell counts were normalized to  $1 \times 10^5$  cells  $\text{ml}^{-1}$  on day 5 to take into account different numbers of starting cells after freezing. The Mann-Whitney rank sum test with Benjamini-Hochberg multiple test correction ( $q < 0.05$ ) was used to compare cell counts of healthy and CDA-I patient samples.

*Morphological analysis using Cytospins:*  $1 \times 10^5$  cells were resuspended in 200  $\mu\text{L}$  PBS, spun (5min, 400 rpm) in a Cytospin4 (ThermoFisher), stained with modified Wright's stain and mounted in DPX (Sigma). Cytospins were imaged using an Olympus BX60 microscope with 10x and 20x objectives. Using images of the cytospins, the morphology of the cultured erythroblasts was scored with the following categories; pro-erythroblasts (Pro), basophilic erythroblasts (Baso), polychromatic erythroblasts (Poly), orthochromatic erythroblasts (Ortho) and enucleated (Enuc).

*Iso-electric Focusing:*  $1 \times 10^6$  cultured erythroblasts were lysed in Haemoglobin elution solution and 50% loaded on an iso-electric focusing (IEF) gel (RESOLVE<sup>®</sup> Haemoglobin kit, PerkinElmer, USA) then run at 1200v for 90 min at 15 °C on a water-cooled horizontal electrophoresis rig (GE Healthcare). Gels were fixed in 10% trichloroacetic acid and stained with the JB-2 system (Perkin Elmer, USA) as per manufacturer instructions.

*CyTOF:* CyTOF detects expression levels of multiple proteins in single cells by staining with antibodies conjugated to heavy metal isotopes. Levels of each isotope

are measured by mass spectrometry which eliminates the problem of spectral overlap associated with FACS and thus expands the parameters that can be measured simultaneously.

*ATAC-Seq Library preparation and analysis:* Immunoprecipitated material was indexed using NEB Next Ultra II DNA library prep kit for Illumina (New England Biolabs). ChIP-seq and ATAC-seq libraries were sequenced on the NextSeq platform (Illumina v2 chemistry) with 39-bp paired-end reads. Reads were mapped to the hg19 genome using NGseqBasic<sup>5</sup> (V20; --nextera --blacklistFilter --noWindow), which utilises Bowtie.<sup>6</sup> Sequence depth and mapped reads are provided (Supplemental Table 3). GEO repositories of sorted cell population ATAC-seq (GSE75384, GSE115684),<sup>1, 2</sup> chromatin marks (GSE125926)<sup>7</sup> and NF-E2 ChIP-seq (GSE125753, GSE95875)<sup>8</sup> (ENCODE) were analysed by the same method. For visualisation PCR-duplicate filtered replicates were merged using Samtools<sup>9</sup> (v1.3) and converted to bigwig format with minimal smoothing using deepTools<sup>10</sup> (v2.2.2; bamCoverage --binSize 10 --normalize using RPKM). ATAC-seq peaks were called from sorted hematopoietic populations (Supplemental Tables 4 and 5), healthy donor and patient samples using Macs2<sup>11</sup> (v2.0/10 callpeak -f BAMPE -g 1.87e9 -q 0.1). Peaks identified in less than three samples were discarded and remaining peaks were merged using BEDtools merge<sup>12</sup> (v2.25.0), to form a collection of peaks detected across all cell types. For differentiation trajectory plotting, principal component analysis (PCA) was performed on published datasets using scikitlearn (v0.22), then samples from *ex vivo* differentiation cultures were projected into the same space using the rotation identified from purified subpopulations. Differential accessibility was determined using DESeq2<sup>13</sup> by comparing ATAC-seq read counts in autosomal open chromatin regions from both day 10 and day 13 from two separate differentiations of three healthy donors (n=6 for each timepoint) with CDA-I patients comprised of both *CDIN1* patients (n=3 for each timepoint) and *CDAN1* patients (n=4 for each timepoint). A factorial design was used to account for changes between day 10 and day 13 as well as changes between CDA-I patients and healthy donors. Motif identification was performed using MEME, using a custom background of peaks that were accessible in *ex vivo* cultured cells but which maintained the same accessibility between CDA-I patients and healthy donors.<sup>14</sup>

*Immunofluorescence (IF):*  $1-2 \times 10^5$  cells were washed and settled on poly-L-lysine treated coverslips (5 mins). Cells were fixed (4% PFA, 15 min) and permeabilized in 0.2% Triton X100 in PBS (10 min, RT). Slides were blocked using 10% FCS in PBS (RT for 30 mins). Antibodies were prepared in blocking solution (Supplemental Table 7). Following detection, slides were washed in PBS, fixed (4% PFA) and coverslips mounted in Vectashield with  $1 \mu\text{g mL}^{-1}$  DAPI added as a nuclear counterstain.

*Fluorescence in situ hybridization (FISH):* Cells were washed in PBS, settled on clean poly-L-lysine coated slides for 5 min, then slid horizontally into a dish containing 3:1 (vol:vol) methanol : acetic acid fixative for 10 min. Following fixation slides were dried vertically and stored at  $-20 \text{ }^\circ\text{C}$  with desiccant. To perform FISH, slides were treated with  $100 \mu\text{g/ mL}$  RNase at  $37 \text{ }^\circ\text{C}$  for 1 hr, washed in 2xSSC, dehydrated through an ethanol series, denatured in 70% formamide in 2xSSC pH 7.0 at  $74 \text{ }^\circ\text{C}$  for 5 min, dehydrated through an ice-cold ethanol series and air-dried. Labelled probes (100 ng each) were denatured in hybridization buffer (Leica, KBI-FHB) at  $90 \text{ }^\circ\text{C}$  for 10 min; BACs were preannealed at  $37 \text{ }^\circ\text{C}$  for 20 min. Slides were hybridized with prepared probes at  $37 \text{ }^\circ\text{C}$  overnight. After hybridization, slides were washed as follows; 2 min in 2xSSC with 0.1% IGEPAL® CA 630 (vol/vol) to remove coverslips, 2 min 2xSSC with 0.1% IGEPAL® CA 630 (vol/vol), then 2 min exactly at  $70 \text{ }^\circ\text{C}$  in 0.4x SSC with 0.3% IGEPAL® CA 630 (vol/vol). Following 1 min in 2xSSC with 0.1% IGEPAL® CA 630 (vol/vol), slides were blocked in 3% BSA (wt/vol) in 4xSSC and digoxigenin was detected with sheep anti-digoxigenin FITC 1:50 (Roche, 11207741910) followed by rabbit anti-sheep FITC 1:100 (Vector Laboratories, FI-6000). Slides were mounted in Vectashield (Vector Laboratories) containing  $2 \mu\text{g/ mL}$  DAPI.

*Imaging equipment, settings and image restoration:* Widefield fluorescence imaging was performed at  $20 \text{ }^\circ\text{C}$  on a DeltaVision Elite system (Applied Precision) using a  $100 \times/1.40 \text{ NA}$  UPLSAPO oil immersion objective for FISH slides and a  $60 \times/1.42 \text{ NA}$  PLAPON oil immersion objective (Olympus) for immunofluorescence, a CoolSnap HQ2 CCD camera (Photometrics), DAPI (excitation 390/18; emission 435/40), FITC (excitation 475/28; emission 525/45) and TRITC (excitation 542/27; emission 593/45) filters. 12-bit image stacks were acquired with a z-step of 200 nm giving a voxel size of  $64.5 \text{ nm} \times 64.5 \text{ nm} \times 200 \text{ nm}$  ( $\times 100$  objective) or  $108.2 \text{ nm} \times 108.2 \text{ nm} \times 200 \text{ nm}$  ( $\times 60$  objective). For the immunofluorescence shown in Figure 6C, image restoration was

carried out using Huygens deconvolution Classic Maximum Likelihood Estimation (Scientific Volume Imaging B.V.).

## Supplemental References

1. Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet.* 2016;48(10):1193-1203.
2. Ludwig LS, Lareau CA, Bao EL, et al. Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. *Cell Rep.* 2019;27(11):3228-3240 e3227.
3. Fibach E, Manor D, Oppenheim A, Rachmilewitz EA. Proliferation and maturation of human erythroid progenitors in liquid culture. *Blood.* 1989;73(1):100-103.
4. Roy NB, Wilson EA, Henderson S, et al. A novel 33-Gene targeted resequencing panel provides accurate, clinical-grade diagnosis and improves patient management for rare inherited anaemias. *Br J Haematol.* 2016;175(2):318-330.
5. Telenius J, Hughes JR. NGseqBasic - a single-command UNIX tool for ATAC-seq, DNaseI-seq, Cut-and-Run, and ChIP-seq data mapping, high-resolution visualisation, and quality control. *bioRxiv.* 2018;393413.
6. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.
7. Downes DJ, Schwessinger R, Hill SJ, et al. An integrated platform to systematically identify causal variants and genes for polygenic human traits. *bioRxiv.* 2020;813618.
8. Olijnik AA, Roy NBA, Scott C, et al. Genetic and functional insights into CDA-I prevalence and pathogenesis. *J Med Genet.* 2020;in press.
9. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.
10. 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):W160-165.
11. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc.* 2012;7(9):1728-1740.
12. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841-842.
13. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
14. Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(Web Server issue):W202-208.