Marked overlap of four genetic syndromes with dyskeratosis congenita confounds clinical diagnosis.
Walne, AJ; Collopy, L; Cardoso, S; Ellison, A; Plagnol, V; Albayrak, C; Albayrak, D; Kilic, SS; Patroglu, T; Akar, H; Godfrey, K; Carter, T; Marafie, M; Vora, A; Sundin, M; Vulliamy, T; Tummala, H; Dokal, I

(c) The Authors 2016
Published version available from the Haematologica Journal website
http://www.haematologica.org

For additional information about this publication click this link.
http://qmro.qmul.ac.uk/xmlui/handle/123456789/15912

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk
Marked overlap of four genetic syndromes with dyskeratosis congenita confounds clinical diagnosis

Amanda J. Walne1*, Laura Collopy1, Shirleny Cardoso1, Alicia Ellison1, Vincent Plagnol2, Canan Albayrak3, Davut Albayrak3, Sara Sebnem Kilic4, Turkan Patiroglu5, Haluk Akar5, Keith Godfrey6, Tina Carter7, Makia Marafie8, Ajay Vora9, Mikael Sundin10,11, Thomas Vulliamy1, Hemanth Tummala1, Inderjeet Dokal1

1Centre for Genomics and Child Health, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Barts NHS Trust, London, UK
2University College London Genetics Institute, London, UK
3Department of Paediatric Haematology, Ondokuz Mayis University, Samsun, Turkey
4Department of Paediatric Immunology Uludag University, Bursa, Turkey
5Department of Paediatric Immunology Erciyes University Medical Facility, Kayseri, Turkey
6Department of Paediatric Dermatology and NIHR Southampton Biomedical Research Centre, University Hospital, Southampton and University of Southampton, Southampton, UK
7Department of Oncology and Haematology, Princess Margret Hospital, WA, Australia
8Clinical Cancer and Community Genetics, Kuwait Medical Genetics centre, Kuwait
9Department of Haematology, Sheffield Children’s NHS foundation Trust, Sheffield, UK;
10Section of Paediatric Haematology/Immunology/SCT, Astrid Lindgren Children’s Hospital, Karolinska University Hospital, Stockholm Sweden
11Division of Paediatrics, CLINTEC, Karolinska Institutet, Stockholm, Sweden.

Running head: DC or not DC? A Clinical Dilemma

*Corresponding author: Amanda Walne, Centre for Genomics and Child Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London E1 2AT, UK

Telephone +44 207 882 2458, Fax +44 207 882 2195

Email a.walne@qmul.ac.uk
Acknowledgments: The authors would like to thank all the clinicians and patients who have helped us over the years, particularly Drs Babik, Balci, Ghoulam and Leblanc. We would also like to thank the staff at Barts and The London Genome Centre for Sanger sequencing analysis. Financial support is provided by The Medical Research Council-MR/K000292/1, Children with Cancer- 2013/144 and Blood Wise-14032 (AJW, LC, SC, AE, TV, HT and ID). KMG is supported by the National Institute for Health Research through the NIHR Southampton Biomedical Research Centre.
ABSTRACT

Dyskeratosis congenita is a highly pleotropic genetic disorder. This heterogeneity can lead to difficulties in making an accurate diagnosis and delays in appropriate management. The aim of this study was to determine the underlying genetic basis in patients presenting with features of dyskeratosis congenita and who were negative for mutations in the classical dyskeratosis congenita genes. By whole exome and targeted sequencing, we identified biallelic variants in genes that are not associated with dyskeratosis congenita in 17 individuals from 12 families. Specifically these were homozygous variants in USB1 (eight families), homozygous missense variants in GRHL2 (two families) and identical compound heterozygous variants in LIG4 (two families). All patients had multiple somatic features of dyskeratosis congenita but not the characteristic short telomeres. Our case series shows that biallelic variants in USB1, LIG4 and GRHL2, the genes mutated in poikiloderma with neutropenia, LIG4/Dubowitz syndrome and the recently recognised ectodermal dysplasia/short stature syndrome, respectively, cause features that overlap with dyskeratosis congenita. Strikingly these genes also overlap in their biological function with the known dyskeratosis congenita genes that are implicated in telomere maintenance and DNA repair pathways. Collectively these observations demonstrate the marked overlap of dyskeratosis congenita with four other genetic syndromes confounding accurate diagnosis and subsequent management. This has important implications for establishing a genetic diagnosis when a new patient presents in the clinic. Patients with clinical features of dyskeratosis congenita need to have genetic analysis of USB1, LIG4 and GRHL2 in addition to the classical dyskeratosis congenita genes and telomere length measurements.
INTRODUCTION

Dyskeratosis congenita (DC) is a highly heterogeneous genetic and clinical syndrome. The classical presentation of DC is characterized by the mucocutaneous triad of abnormal skin pigmentation (hyper/hypopigmentation with atrophy and telangiectasia, termed poikiloderma), oral leukoplakia and nail dystrophy. Individuals with DC frequently develop bone marrow (BM) failure, are at a high risk of developing cancer and can develop disease features in virtually every system in the body. Pathologically DC is characterized by selective exhaustion of highly proliferative cells\(^1\) that have critically short telomeres and exhibit an abnormal DNA damage response.\(^2\)

The Dyskeratosis Congenita Registry (DCR, London) is a collection of patients that have a clinical diagnosis of DC or an overlapping phenotype as defined by Dokal et al.\(^3\) The diagnostic inclusion criteria are:

1. All three mucocutaneous features (abnormal skin pigmentation, nail dystrophy and oral leukoplakia).
2. One mucocutaneous feature plus BM failure and two other somatic features of DC.
3. Aplastic anaemia (AA), myelodysplastic syndrome (MDS) or idiopathic pulmonary fibrosis (IPF) associated with a pathogenic telomerase variant.
5. Two or more features seen in DC plus very short telomeres (<1st centile).

Based on these criteria the phenotype of DC is highly variable. Furthermore, not all physical malformations are present at the point of diagnosis, with many additional features developing with age or, in the case of severe disease, death may occur before more typical clinical presentations become apparent.

To date pathogenic mutations have been identified in 11 genes that cause DC according to the criteria above: dyskeratosis congenita 1, dyskerin (DKC1), telomerase RNA component (TERC), telomerase reverse transcriptase (TERT), NOP10 ribonucleoprotein (NOP10), NHP2 ribonucleoprotein (NHP2), TERF1 (TRF1)-interacting nuclear factor 2 (TINF2), WD repeat containing, antisense to TP53 (WRAP53, also known as TCAB1), CST telomere maintenance complex component 1 (CTC1), regulator of telomere elongation helicase 1 (RTEL1), adrenocortical dysplasia homolog (mouse) (ACD) and poly(A)-specific ribonuclease (PARN). DKC1, TERC, TERT, NOP10 and NHP2 all encode components of the telomerase complex which is involved in telomere elongation. TINF2 and ACD encode components of the shelterin complex which is involved in telomere protection. CTC1
encodes a member of the CST complex which facilitates recruitment and docking of telomerase on to the telomere. \textit{WRAP53} is involved in telomerase trafficking and \textit{RTEL1} is involved in telomere replication, and DNA replication and repair.\textsuperscript{3-7} The final gene, \textit{PARN}, is not exclusively involved in telomere biology but is mutated in a small number of cases of DC, and is an exoribonuclease involved in the control of mRNA stability and the maturation of \textit{TERC} snoRNA.\textsuperscript{8-10} As the majority of DC genes are involved in telomere biology and DC patients usually have short telomeres, this has led to DC being classed as a telomeropathy.\textsuperscript{11} However these mutations do not explain all the cases that are present in the DCR (Data from the DCR, London, April 2016).

The advent of next generation sequencing, particularly whole exome sequencing (WES), has become an invaluable tool for determining the causal variant(s) underlying a specific disease. By sequencing all the coding regions and comparing variants in many individuals with an overlapping phenotype it is possible to elucidate potential disease genes in a way that would not have been possible five years ago.\textsuperscript{12} This can give the clinical diagnosis a genetic basis. The main problem in obtaining a definitive clinical diagnosis is the degree of phenotypic overlap that exists between many different diseases. The diagnosis, in many cases, depends on the interpretation of the initial clinician and this is often dependent on their speciality. Although subsequent opinions may be sought, the preliminary diagnosis tends to remain with the patient until proven otherwise. An accurate genetic diagnosis can aid clinical diagnosis and subsequent patient management. This has the potential to suggest alternative treatment avenues that may not have been considered and also aid the selection of healthy sibling donors for hematopoietic stem cell transplantation, should this become necessary.

Here we report 12 families with an initial diagnosis of DC who were found to have biallelic variants in one of three genes linked to a disease distinct from DC. The genes identified were \textit{USB1} (U6 snRNA biogenesis 1) associated with poikiloderma with neutropenia (PN), \textit{GRHL2} (grainyhead like transcription factor 2) which is associated with ectodermal dysplasia/short stature syndrome (ECTDS) and \textit{LIG4} (Ligase IV, DNA, ATP-dependent) which is predominantly associated with LIG4 syndrome and occasionally Dubowitz syndrome. The key presenting features of PN are poikiloderma with non-cycling neutropenia, recurrent infections, short stature and nail abnormalities\textsuperscript{13}. The typical clinical features associated with ECTDS are short stature, nail dystrophy, abnormal oral pigmentation, and keratoderma and hyperkeratosis of the hands and feet\textsuperscript{14}. LIG4 syndrome is characterized by immune deficiency and developmental and growth delay\textsuperscript{15}. Patients can also display unusual facial features, microcephaly, pancytopenia and various skin abnormalities. Dubowitz syndrome is characterised by growth failure/short stature, characteristic facial features, microcephaly, mild mental retardation and eczema\textsuperscript{16}. All of these disorders have a high degree of phenotypic overlap with DC, thus leading to difficulties in making an accurate clinical diagnosis. The identification of biallelic \textit{USB1}, \textit{GRHL2} and
LIG4 mutations in 12 different families, initially diagnosed to have DC, therefore demonstrate the marked phenotypic overlap of ‘classical DC’ with four other genetic syndromes.

METHODS

Patient samples

Exome capture or targeted gene screening was performed on a series of genetically uncharacterized index cases in the Dyskeratosis Congenita Registry (held at Barts and The London). Presenting features included some/all of the classic mucocutaneous abnormalities (abnormal skin pigmentation, nail dystrophy, leukoplakia), with or without bone marrow failure. Peripheral-blood samples were obtained with written consent under the approval of our local research ethics committee (London – City and East). Genomic DNA was extracted from these peripheral blood samples for use in all the subsequent analyses (Puregene, Qiagen).

Telomere length measurement

Telomere lengths were measured using the MMqPCR method (modified from Cawthon17), on a LightCycler 480 real time thermocycler (Roche). Briefly, in each well, amplification of telomeric DNA (T) and a single copy gene (S) were quantified against standard curves obtained from dilution of a reference DNA sample. The T/S ratio, obtained in triplicate for each sample, is proportional to the telomere length. This ratio was normalized to the T/S ratio of a second reference sample that was run on every plate to give a relative T/S ratio.

Whole exome sequencing (WES)

50ng genomic DNA was subjected to library preparation and exome capture using the Nextera Rapid Capture Exome kit (Illumina). Sequencing was performed using the Illumina HiSeq 2000 system and 100bp paired end reads were generated and the data processed through the Illumina pipeline. Variants were called as described previously.8 All relevant variants were confirmed by Sanger sequencing.

Targeted gene screening

We designed an NGS assay covering the coding regions and some 5’UTR’s from 31 genes associated with genetic bone marrow failure syndromes (Supplementary Table 1). We used the Illumina TruSeq custom amplicon kit for library preparation and capture according to manufacturer’s instructions. The resultant targeted fragments were indexed by dual barcodes and then sequenced on the Illumina MiSeq platform. Read alignment was performed automatically using the online MSR: TruSeq amplicon software. Annotation of the variants was performed using ANNOVAR.
Variant filtering

Exome data from our patients were jointly called with 2500 WES internal control samples (UCL-ex consortia) with unrelated conditions to minimise artificial batch effects. Where there was a family history we assumed an autosomal recessive mode of inheritance as all parents were reportedly asymptomatic. Filtering was performed using the following criteria: All variants had to pass the Illumina filter, have a read depth of >10X, be present in the general population at a frequency of <0.0002 as reported on Exome Aggregation Consortium database, ExAC (www.exac.broadinstitute.org), 1000 genome project (www.1000genomes.org) and from the UCL-ex consortia. Variants that were predicted to be tolerated and benign by SIFT (http://sift.jcvi.org/) and Polyphen (http://genetics.bwh.harvard.edu/pph2/) were then removed from further analysis. All relevant variants were confirmed by Sanger sequencing.

RESULTS

Telomere length analysis

It has been well established that in the majority of patients with “classical” DC who have mutations in components of the telomere maintenance pathway have short telomeres, usually below the 1st centile.\textsuperscript{18, 19} In fact this measure is often used as a diagnostic screening tool to perform a differential diagnosis of DC from other bone marrow failure syndromes such as Fanconi anemia.\textsuperscript{20} Telomere lengths were therefore measured in our patients by MMqPCR and this was used as an indicator of whether we could be looking for a mutation in a known telomere biology related gene or not. As none of the patients reported in this study had short telomeres when compared with controls (Figure 1), this suggested that we were looking for mutations outside the spectrum of those usually associated with DC but they could still be associated with other bone marrow failure diseases or else have a previously unidentified association to another pathology.

Identification of causal genetic variants

Analysis of our data has identified significant biallelic variants in three genes as detailed below. Five families underwent WES (Families 4, 9-12) and the remaining underwent targeted sequencing either using a 31 bone marrow failure disease gene panel (Supplementary Table 1) or direct sequencing.

\textit{USB1} (U6 snRNA biogenesis 1)

Biallelic \textit{USB1} mutations have been previously described in patients with DC and the overlapping disease poikiloderma with neutropenia (PN).\textsuperscript{21, 22} Since our previous publication,\textsuperscript{21} we have identified homozygous \textit{USB1} variants in an additional eight families with features of DC. In seven of the eight
families, the variant was identified by targeted sequencing and the remaining family (two cases) underwent WES. The only variant that was homozygous and shared by both affected siblings was in *USB1* so this was deemed to be causal. Five out of the eight index cases had the previously documented recurring homozygous variant c.531delA, p.His179MetfsTer86 (Figure 2A) which is not described on ExAc, (accessed April 2016). The sixth index case was found to have the homozygous variant c.673C>T, p.Gln225Ter (Figure 2B), which has been reported previously as part of a compound heterozygote. The remaining two families (Family 7 and Family 8) have the same homozygous missense mutation c.623A>G, p.His208Arg (Figure 2C). This is the first report of a homozygous missense being associated with disease in *USB1* as to date all 19 disease-causing variants reported in *USB1* are predicted to be loss of function (splicing, nonsense or frame shift). Segregation analysis confirmed an autosomal recessive mode of inheritance and homozygous variants were confirmed in an additional four affected siblings within these families (Figure 2A). In silico analysis of mutations on USB1 crystal structure (PDB id: 4W7H, Figure 2D) revealed that both p.His179MetfsTer86 and p.Gln225Ter truncates several α-helices and β-sheets in USB1 (Figure 2E and 2F). Furthermore the homozygous missense change p.His208Arg identified in families 7 and 8 affects the highly conserved histidine residue H208, (Supplementary Figure 1 and Figure 2G and 2H). This change is predicted to be probably damaging by polyphen2 (http://genetics.bwh.harvard.edu/pph2/; HumVar score 1.0) and disease causing by Mutation taster (http://www.mutationtaster.org). This H208 residue is implicated in the catalytic mechanism of USB1, where it facilitates displacement of uridine nucleoside on RNA substrate to stabilise U6 small nuclear RNA, which in turn plays a critical role in RNA splicing.

Table 1 summarises the clinical presentation of these families and the similarity between the clinical presentation of DC and PN is further highlighted in Figure 3. The degree of overlap between the reticular pigmentation observed on the legs (Figure 3A-C), nail dystrophy of the toes (Figure 3D-F) is marked. Similar pigmentation is seen on the trunk (Figure 3G and 3H) and nail dystrophy and blistering is shown on the hands (Figure 3I). In addition to highlighting the similarity between the cutaneous presentation of DC and PN, panels B, C, E, F and H, I also emphasise the difference in presentation that can be caused by mutations within the same gene. The blood results detailed in Table 1 show that although there is neutropenia in all the patients, the overall blood pathology is more global rather than being restricted to just one lineage. This investigation has expanded the repertoire of *USB1* variants identified to date, by adding a recurrent novel homozygous missense variant.

*GRHL2* (grainyhead like transcription factor 2)

The index case from both families 9 and 10 underwent WES, and as both families were consanguineous, we assumed an autosomal recessive mode of inheritance. Using the filtering
approach described, it was noted that both families had homozygous non synonymous variants in the gene \textit{GRHL2} c.1445T>A p.Ile482Lys in Family 9 and c.1213C>A p.Pro405Thr in Family 10 (other homozygous calls that fulfilled the filtering criteria are detailed in Supplementary Table 2). Segregation analysis confirmed that both variants were inherited in an autosomal recessive manner (Figure 4A). Both of these single base substitutions affect highly conserved nucleotides (Supplementary Figure 2). As neither variant is reported on ExAc (accessed April 2016) and both variants are predicted by polyphen2 to be probably damaging (HumVar scores p.P405T - 0.995 and p.I482K - 0.917) and disease causing by Mutation Taster, \textit{GRHL2} is the disease causing gene in these two families. \textit{GRHL2} is a member of a highly conserved family of transcription factors that are essential for epithelial development. At the molecular level, the GRHL transcription factors regulate the expression of proteins involved in cell proliferation, differentiation, adhesion and polarity. These factors adopt a DNA-binding immunoglobulin fold homologous to the DNA-binding domain of key tumour suppressor p53.\textsuperscript{26} Recently Petrof and colleagues described homozygous mutations in \textit{GRHL2} in two families with ectodermal dysplasia/short stature syndrome (ECTDS)\textsuperscript{14}. One of these was p.Ile482Lys, which is the same variant as identified in Family 9. All the variants described by us and Petrof et al., affect the DNA binding domain of the protein (Figure 4A). Phenotypically there are many shared features between our two families (Families 9 and 10) and the ECTDS families reported in the literature (Table 2). Another recent study\textsuperscript{27} examined the role of GRHL2 in the developing kidney and clarified the involvement of GRHL2 in a network that controls the development of lumen expansion and barrier formation in renal epithelia, which may explain the previously reported renal agenesis seen in Family 10.\textsuperscript{25} This suggests that homozygous variants in the gene \textit{GRHL2}, previously linked to ECTDS, are disease-causing in these two families.

\textit{LIG4} (Ligase IV, DNA, ATP-dependent) Analysis of exome data from the index cases of Families 11 and 12 failed to reveal any homozygous variants that passed the filtering strategy. However, a biallelic analysis showed that they both shared the same compound heterozygous variants in \textit{LIG4} namely c.2440C>T, p.Arg814Ter and c.1270_1274 delAAAAG, p.Lys424ArgfsTer20 (Figure 4B, other biallelic calls are listed in Supplementary Table 2). Although both variants observed in Families 11 and 12 are reported on ExAc at very low frequency (0.000082 and 0.00014, respectively) in the heterozygous state, the likelihood of them occurring together (as is the case here) by chance is $2.4 \times 10^{-8}$. \textit{LIG4} is involved in DNA non-homologous end joining and V(D)J recombination.\textsuperscript{28} Biallelic mutations in \textit{LIG4} are rare with only 28 cases being reported to date in the literature,\textsuperscript{29} and are usually associated with LIG4 syndrome and more recently in Dubowitz syndrome (DS).\textsuperscript{16} Both LIG4 syndrome and DS belong to a group of disorders that are associated with impaired DNA damage response mechanisms.\textsuperscript{28} LIG4 syndrome is a hereditary disorder associated with impaired DNA double-strand break repair mechanisms. It is
characterized by growth restriction, developmental delay, microcephaly, facial dysmorphism, pancytopenia, variable immune deficiency and an increased predisposition to leukemia. DS is a rare multiple congenital syndrome characterized by cognitive delay, growth failure, microcephaly, distinctive facial dysmorphism, immune defects, pancytopenia, hematological malignancy and neuroblastoma. The degree of clinical overlap between LIG4 syndrome, DS and our two patients is marked (Table 2). Furthermore the fact that both of the variants in LIG4 described here have been previously documented to be pathogenic and have been shown to occur in the same combination suggests that these are the causal genetic defects in these two unrelated individuals.

DISCUSSION

Historically the diagnosis of disease is defined by the presenting clinical features but it is becoming more apparent that this can give conflicting diagnoses as shown in Tables 1 and 2. Recent advances in whole exome and targeted sequencing have enabled a genetic diagnosis to be obtained more readily than ever before. This now enables the clinician to consider the possibility of diseases with overlapping phenotype based on a genetic result rather than solely relying on the clinical presentation. In this report we have elucidated the underlying genetic causal variant(s) in 12 families (comprising 17 affected individuals) who were given a clinical diagnosis of dyskeratosis congenita. Initially we undertook whole exome sequencing on uncharacterised patients who presented with sufficient clinical features to be diagnosed as DC. As targeted sequencing methodologies improved, we then chose to use this approach as a screening tool to identify patients who had mutations in the known DC and bone marrow failure genes (supplementary Table 1). The data reported here represent a sub-set of all the cases analysed, where we believe the disease-causing gene has been identified and where the identified gene has been previously associated with another genetic syndrome. By using a combination of these two strategies we identified disease-causing variants in the genes USB1 (eight families, 12 cases), GRHL2 (two families, three cases) and LIG4 (two families, two cases). Clinically all patients show a high degree of overlap with the phenotypic profile for DC as described by Dokal et al and the other diseases with which the disease-causing variants have been associated. This presents a problem in defining the disease: is it best to use phenotype or genotype?

Significantly short telomeres are often used as a biomarker for DC. A telomere length below the 1st centile is considered diagnostic for DC but this measurement is not routinely reported in patients with other bone marrow failure diseases. Figure 1 shows how telomeres in patients with variants in the known telomere genes (DKC1, TERC, TERT and TINF2) are significantly short when compared with a control population. Telomere lengths in all affected cases reported here were not significantly short when compared with controls regardless of the underlying genetic variant (Figure 1). This measure
should perhaps be used where a diagnosis of classical DC is suspected as a screening tool to decide the best course of action to determine a genetic diagnosis. Reduced telomere length has also been associated with an increased risk of cancer and DC is regarded as a cancer prone syndrome. Patients with *USB1* mutations tend to have an earlier presentation of myelodysplasia (MDS) than the general population as most reports describe disease in children. In this study and in our previous study, we report 6 patients with MDS and 2 with acute myeloid leukemia suggesting there is an increased cancer risk despite the patients not having short telomeres. Based on the small number of cases in our series, there does not appear to be an increased cancer risk for patients with either *GRHL2* or *LIG4* mutations. It is notable however that in the literature *LIG4* syndrome is described as a disorder with predisposition to leukemia. Although an increased cancer risk is not observed in our families this may be due to the small numbers of patients described herein.

Due to the extensive overlap in clinical features between DC and PN as described in Table 1, telomere length seems to be one way of separating these two diagnostic definitions. Given that these cases do not have short telomeres but do possess DC features (Table 1) present us with perplexing situation in accurately defining these patients. We propose it is timely to combine features of these two entities as a new syndrome, “USB1 deficiency syndrome”, characterized by bone marrow failure, abnormal skin pigmentation (poikiloderma), nail dystrophy, growth restriction, cancer predisposition and normal length telomeres.

It is also notable that the genes mutated in these overlapping syndromes have similarities with the biological functions of the known DC genes (Figure 5). *USB1* has a role in snRNA processing, as is the case for dyskerin, NHP2 and NOP10. It also functions as an exoribonuclease reminiscent to the function of PARN that has been recently found to be mutated in both DC and pulmonary fibrosis. Both *LIG4* and *RTEL1* are involved in the maintenance of genomic stability as they are both involved in pathways that repair double strand breaks. *LIG4* is an ATP-dependent DNA ligase that joins double strand breaks during non-homologous end joining and *RTEL1* (a protein that is mutated in a subset of DC) is an ATP-dependent DNA helicase, which has an important role in telomere-length regulation. Both function via interactions with another shelterin component TRF2 (Telomere repeat binding factor 2). Phenotypically many of the features seen in *LIG4*/Dubowitz syndrome are also seen in DC patients with biallelic *RTEL1* mutations (Table 2). This supports the idea that mutations in *LIG4* can give rise to phenotype that is reminiscent of DC as caused by a gene that has a known association with telomere biology. *GRHL2* affects TERT expression by acting on its promoter thereby providing an important biological link to classical DC. While these proteins do have different functions or interactions, we speculate that it is this similarity as suggested in Figure 5 that gives rise to the overlapping phenotypes. It is also notable that many of the DC genes have functions other than telomere maintenance. It is therefore possible that those overlapping shared functions may also be contributing to the similar clinical features observed here. In the future it will be interesting to
determine whether therapeutic agents that work in one disease have a beneficial effect in any of the overlapping syndromes.

The findings of this study raise a potential issue as to what constitutes DC. Taking into account the clinical features, germline genetic mutation(s) and telomere length we can recognise three categories of patients: “Category 1”. Patients with sufficient clinical features (such as abnormal skin pigmentation, nail dystrophy, leukoplakia and bone marrow failure) to be labelled as DC, who have a germline mutation(s) in a telomere biology gene (such as TERT, TERC, DKC1, TINF2, RTEL1 and so forth) and who have short telomeres (described as being less than 1st centile). “Category 2”. Patients with some clinical features of DC (but not sufficient to be classified as DC) who are found to have a germline mutation(s) in one of the telomere biology genes and who have short telomeres. “Category 3”. Patients with sufficient clinical features of DC but who harbour a germline mutation(s) in a non-telomere biology gene (such as USB1, LIG4 and GRHL2) and have normal length telomeres. Category 1 patients can be considered to represent “classical or pure DC”. Category 1 and Category 2 patients can both be considered to represent disorders of telomeres – “telomeropathies”. Category 3 could be considered as “DC-like” or “DC-overlap”.

In summary, in nine out of the 12 families highlighted in this study, the biallelic variants identified in USB1, GRHL2 and LIG4 have been shown to be pathogenic in previous studies. In the remaining families we have provided strong genetic and in-silico evidence that the identified mutations are pathogenic. This study has demonstrated that the pleotropic clinical phenotype of DC markedly overlaps with the recognised disease entities LIG4 syndrome, Dubowitz syndrome and PN, as well as the recently recognised ECTDS. In doing so it has substantiated the last category as a significant disease entity. The marked overlap of features of DC with PN, ECTDS, LIG4 and Dubowitz syndromes has important implications for establishing genetic diagnosis when a new patient presents in the clinic. Specifically in patients with a clinical diagnosis of DC the genetic analysis needs to include GRHL2, LIG4 and USB1 in addition to the classical DC genes.

All work in this study was conducted with informed, written consent and approval of the local ethics committee.
References


### Table 1: Overlap between DC, PN and patients with homozygous variants in *USB1*

<table>
<thead>
<tr>
<th>Syndrome/ Family (OMIM)</th>
<th>DC (#305000, #127500, #224230)</th>
<th>PN (#604173)</th>
<th>Family 1</th>
<th>Family 2&lt;sup&gt;st&lt;/sup&gt;</th>
<th>Family 3&lt;sup&gt;rd&lt;/sup&gt;</th>
<th>Family 4</th>
<th>Family 5</th>
<th>Family 6</th>
<th>Family 7</th>
<th>Family 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consanguinity</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Country of Origin</td>
<td>Turkey</td>
<td>Turkey</td>
<td>Turkey</td>
<td>Turkey</td>
<td>Turkey</td>
<td>Australia</td>
<td>Afghanistan</td>
<td>Afghanistan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at report (years)</td>
<td>24,14</td>
<td>12,7,5</td>
<td>7</td>
<td>9, 7</td>
<td>6</td>
<td>10</td>
<td>4</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nail dystrophy</td>
<td>&gt;60% cases**</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Abnormal skin pigmentation</td>
<td>&gt;60% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>BM failure</td>
<td>&gt;60% cases</td>
<td>Yes&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Yes&lt;sup&gt;h&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Leukoplakia</td>
<td>&gt;40% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Developmental delay</td>
<td>&gt;20% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Microcephaly</td>
<td>&gt;20% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Growth restriction</td>
<td>&gt;20% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hair loss</td>
<td>&gt;20% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pulmonary disease</td>
<td>&gt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cancer</td>
<td>~10% cases</td>
<td>~10% cases</td>
<td>~10% cases</td>
<td>~10% cases</td>
<td>~10% cases</td>
<td>~10% cases</td>
<td>~10% cases</td>
<td>~10% cases</td>
<td>~10% cases</td>
<td>~10% cases</td>
</tr>
<tr>
<td>Abnormal dentition</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Gonadal abnormalities</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Deafness/ear abnormalities</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Eye abnormalities</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Abnormal facies</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Skeletal abnormalities</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Immune deficiency</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Short telomeres</td>
<td>&lt;1&lt;sup&gt;st&lt;/sup&gt; centile</td>
<td>Unknown</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
</tbody>
</table>

<sup>a</sup> p.H179MfsTer<sup>86</sup> in Family 1; <sup>b</sup> p.H179MfsTer<sup>86</sup> in Family 2; <sup>c</sup> p.H179MfsTer<sup>86</sup> in Family 3; <sup>d</sup> p.H179MfsTer<sup>86</sup> in Family 4; <sup>e</sup> p.H179MfsTer<sup>86</sup> in Family 5; <sup>f</sup> p.H179MfsTer<sup>86</sup> in Family 6; <sup>g</sup> p.H179MfsTer<sup>86</sup> in Family 7; <sup>h</sup> p.H179MfsTer<sup>86</sup> in Family 8.
All clinical features reported within a family are detailed. DC-Dyskeratosis congenita; PN-poikiloderma with neutropenia; * Mutations identified in CTC1, DKC1, NHP2, NOP10, PARN, RTEL1, TERC, TERT, TINF2, TPP1 and WRAP53; ** as calculated from the index cases of the 1st 400 families included in the DCR, London; †Term used to cover short stature, intra uterine growth restriction and low birth weight; ‡Parents reported as being from the same village but no statement of consanguinity. The blood counts for the index case at presentation are as follows: a-hemoglobin 106g/l, leucocytes 2.3x10⁹/l, neutrophils 0.3x10⁹/l, platelets 246x10⁹/l and bone marrow showed decreased cellularity; b-hemoglobin 125g/l, leucocytes 3.2x10⁹/l, neutrophils 0.1x10⁹/l, platelets 311x10⁹/l and bone marrow showed decreased cellularity with approximately 10% blasts; c-hemoglobin 113g/l, leucocytes 3.2x10⁹/l, neutrophils 0.8x10⁹/l platelets 308x10⁹/l; d-hemoglobin 115g/l, leucocytes 3.4x10⁹/l, neutrophils 0.8x10⁹/l, platelets 138x10⁹/l; e-hemoglobin 121g/l, leucocytes 3.68x10⁹/l, neutrophils 0.7x10⁹/l, platelets 269x10⁹/l; f-hemoglobin 70g/l, leucocytes 2.1x10⁹/l, neutrophils 1.1x10⁹/l, platelets 5x10⁹/l and bone marrow showed increased cellularity, decreased magakaryocytes, dysmyelopoiesis with nuclear changes and abnormal granularity; g-hemoglobin 115g/l, leucocytes 4.5x10⁹/l, neutrophils 0.9x10⁹/l, platelets 204x10⁹/l; h-hemoglobin 114g/l, leucocytes 1.7x10⁹/l, neutrophils 0.8x10⁹/l, platelets 112x10⁹/l. MDS—myelodysplastic syndrome.
Table 2: Clinical overlap between the different genetic syndromes and the families with biallelic variants identified in this study

<table>
<thead>
<tr>
<th>Syndrome/Family (OMIM reference)</th>
<th>DC (#305000, #127500, #224230)</th>
<th>ECTDS (#616029)</th>
<th>Family 9</th>
<th>Family 10&lt;sup&gt;15&lt;/sup&gt;</th>
<th>LIG4 syndrome (#606593)</th>
<th>Dubowitz syndrome (#223370)</th>
<th>Family 11</th>
<th>Family 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consanguinity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country of origin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family history</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at report (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nail dystrophy</td>
<td>&gt;60% cases**</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal skin pigmentation</td>
<td>&gt;60% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental delay</td>
<td>&gt;20% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microcephaly</td>
<td>&gt;20% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth restriction†</td>
<td>&gt;20% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oesophageal stricture</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gonadal abnormalities</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ataxia</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immune deficiency</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deafness/ear abnormalities</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney abnormalities</td>
<td>&lt;10% cases</td>
<td>Recurrent infections</td>
<td>Renal agenesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal facies</td>
<td>&lt;10% cases</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short telomeres</td>
<td>&lt;1&lt;sup&gt;st&lt;/sup&gt; centile</td>
<td>Unknown</td>
<td>Normal</td>
<td>Normal</td>
<td>Unknown</td>
<td>&lt;1&lt;sup&gt;st&lt;/sup&gt; centile&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Normal</td>
<td>Normal</td>
</tr>
</tbody>
</table>

All clinical features reported within a family are detailed. ECTDS – Ectodermal dysplasia/short stature syndrome. * Mutations identified in CTC1, DKC1, NHP2, NOP10, PARN, RTE1L1, TERC, TERT, TINF2, TPP1 and WRAP53;** as calculated from the index cases of the 1<sup>st</sup> 400 families in the DCR, London; †Term used to cover short stature, intra uterine growth restriction and low birth weight. The blood counts for the index case at presentation are as follows: a - hemoglobin 141g/l, leucocytes 7.2x10<sup>9</sup>/l, platelets 307x10<sup>9</sup>/l (normal blood counts); b-hemoglobin 89g/l, leucocytes 4.9x10<sup>9</sup>/l, platelets 216x10<sup>9</sup>/l and bone marrow showed reduced cellularity with megaloblastic changes; c-hemoglobin 96g/l, neutrophils 0.5x10<sup>9</sup>/l, platelets 20x10<sup>9</sup>/l and bone marrow showed reduced cellularity and reduced megakaryocytes; d-pancytopenia with B-lymphopenia and bone marrow showed reduced cellularity. e- reported in 3 patients only
**Figure Legends**

**Figure 1**: Telomere lengths in patients with *GRHL2*, *LIG4* and *USB1* mutations are not short

Relative T/S ratios are within the normal range. Open circles – controls (n=130); green diamonds – patients with biallelic *USB1* variants (n=12), blue triangles – patients with homozygous *GRHL2* variants (n=2); Inverted red triangles – patients with compound heterozygous *LIG4* variants (n=2). For comparison short telomeres are seen in patients with mutations in the known telomere associated genes *DKC1* (n=50, open squares), *TERC* (n=66, open triangles), *TERT* (n=36, open inverted triangles) and *TINF2* (n=33, open diamonds). Black line represents the median for each data set.

**Figure 2**: Identification and segregation of disease-causing variants in *USB1*

A) The c.531delA variant is identified in five families with a DC phenotype. Representative sequencing traces are shown for the wild type (+/+), heterozygous (+/-) and homozygous (-/-) forms of the mutation.

B) Family tree and sequence traces of c.673C>T variant seen in Family 6. A representative heterozygous trace and the homozygous trace are shown.

C) Family tree and sequence traces of the novel c.623A>G homozygous variant (G/G) as seen in Family 7 and 8. A representative wild type (A/A) trace is also shown. NA – sample not available.

D) Protein crystal structure of human USB1 (PDB id: 4W7H) depicted in ribbon form indicating the both transit and terminal lobes indicated in green α helices. The position of identified residues mutated in USB1 deficiency patients is indicated as stick model in blue.

E and F) In silico modelling of truncation mutations reveals loss several α helices and β sheets in both transit and terminal lobes of USB1 secondary structure.

G and H) The invariant H-x-S motif in the active site is indicated as stick model. The missense change Histidine to Arginine at 208 position (red) introduces an extra hydrogen bond (dotted line black indicated by arrows), with Ser210 (blue), which in turn might cause aberrant oligoadenylation of RNA substrate.

**Figure 3**: Similar clinical presentation is observed between DC and PN highlighting the marked overlap between these two syndromes

Panels A, D, G are from a male patient with p.Ala353Val DKC1mutation, panels B, E and H are from a male patient with the homozygous USB1 p.His127MetfsTer86 mutation and panels C, F I are from a child with the homozygous USB1 p.His208Arg mutation . Panels A-C show similar patterns of reticular pigmentation on the legs. Panels D-F show nail dystrophy on the toes. Panels G and H show reticular pigmentation on the trunk and panel I shows pigmentation on the arm along with nail dystrophy on the figures. It is notable that the mucocutaneous features of the patient with USB1 p.His179MetfsTer86 mutation appear more like those of the patient with DKC1 p.Ala353Val mutation than the one with the USB1 p.His208Arg mutation.
Figure 4: Identification of disease-causing variants in two different genes in patients with clinical features of DC

A) Segregation and sequence traces for the homozygous *GRHL2* variants identified in Families 9 and 10. Representative heterozygous and homozygous traces are shown for both variants. Schematic representation of the protein shows the identified domains and the location of the variants. Red arrow highlights the variant identified in this study. Black arrow highlights a variant identified in previous study. Green arrow highlights the variant observed in both this and previous study. NA-sample not available.

B) Segregation and sequence traces for the compound heterozygous *LIG4* variants identified in Families 11 and 12. A representative wild type and heterozygous trace is shown for each. The nucleotides deleted are indicated by a line above the corresponding bases on the wild type (+/+) trace. Schematic representation of the protein showing the precise location of variants in different domains. Black arrows highlight variants identified in previous studies. Green arrows highlight variants observed in both this and previous studies.

Figure 5: Biological overlap of classical DC proteins (TERT, DKC1, NOP10, NHP2, PARN and RTEL1) and those (GRHL2, USB1 and LIG4) mutated in patients in this study

Proteins with similar functions are grouped together. The schematic shows that GRHL2 can be grouped broadly with the molecules which have role in telomerase expression, USB1 with small nuclear ribonucleoprotein (snRNP) processing and LIG4 with genomic stability. Previous studies have shown that biallelic mutations in *GRHL2*, *USB1* and *LIG4* are associated with ectodermal dysplasia, poikiloderma with neutropenia (or USB1 deficiency syndrome) and Lig4/Dubowitz syndrome, respectively. Mutations in DKC1, NHP2, NOP10, PARN and RTEL1 have been linked to dyskeratosis congenita.
Figure 3
Figure 4

A

GRHL2

Family 9

T/A

Heterozygous

T/A

TT

Homogenous

A/A

c.1445T>A

p.Ile482Lys

Family 10

C/A

Heterozygous

C/A

c.1213G>A

p.Pro405Thr

P405T M482K

Transactivation

Domain

DNA binding

Domain

Dimerization

Domain

Y398H

B

LIGIV

Family 11

C/T

Wildtype

C/C

T/-

Heterozygous

C/T

c.2446C>T

p.Arg814Ter

del

Wildtype

(+/+)

c.1270, 1274del(AAAA)

p.Lys424ArgfsTer20

A3V

MD248E

M248E

S205fs

Q299R

K424fs

G433R

W447C

R814X

Parn

snRNP

processing

TERC

SNRPN

NOP10

DKC1

NHP2

SNRPN

Telomerase

expression

GRHL2

TEL1

PARN

snRNP

processing

USB1

Telomerase

expression

GRHL2

TEL1

PARN

snRNP

processing

USB1

Genomic

stability

Figure 5