Hardman-Smart Jon (Orcid ID: 0000-0002-1653-7908)
Torrelo Antonio (Orcid ID: 0000-0002-5940-6916)
van Geel Michel (Orcid ID: 0000-0002-0273-8020)
Parren Lizelotte (Orcid ID: 0000-0003-1854-9758)
Frank Jorge (Orcid ID: 0000-0003-1439-8577)
Zhang Xue (Orcid ID: 0000-0001-8274-7674)

Germline intergenic duplications at Xq26.1 underlie Bazex-Dupré-Christol basal cell carcinoma susceptibility syndrome

Running head: Intergenic Xq26.1 duplications underlie Bazex-Dupré-Christol syndrome

Yanshan Liu,[1*] Siddharth Banka,[2,3*] Yingzhi Huang,[1*] Jonathan Hardman-Smart,[4,5] Derek Pye,[4] Antonio Torrelo,[6] Glenda M. Beaman,[2,3] Marcelo G. Kazanietz,[7] Martin J. Baker,[7] Carlo Ferrazzano,[8] Chenfu Shi,[8] Gisela Orozco,[8] Stephen Eyre,[8] Michel van Geel,[9,10] Anette Bygum,[11,12] Judith Fischer,[13] Zosia Miedzybrodzka,[14,15] Faris Abuzahra,[16] Albert Rübben,[17] Sara Cuvertino,[2] Jamie M. Ellingford,[2,3] Miriam J. Smith,[2,3] D. Gareth Evans,[2,3] Lizelotte J.M.T Weppner-Parren,[18] Maurice A.M. van Steensel,[19,20] Iskander H. Chaudhary,[21] D. Chas Mangham,[22] John T. Lear,[4,23] Ralf Paus,[4,24,25] Jorge Frank,[26†] William G. Newman[2,3†] and Xue Zhang[1†]

[1]McKusick-Zhang Center for Genetic Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100005, China.
[2]Division of Evolution, Infection and Genomics, Faculty of Biology, Medicine and Human Sciences, University of Manchester, Manchester, M13 9PL, United Kingdom.
[3]Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Manchester, M13 9WL, United Kingdom.
[4]The Centre for Dermatology Research, University of Manchester, MAHSC, and National Institutes of Health Biomedical Research Center, Manchester, M13 9PL, United Kingdom.
[5]St John's Institute of Dermatology, Kings College London, London, WC2R 2LS, United Kingdom.
[6]Department of Dermatology, Hospital Infantil Universitario Niño Jesús, 28009 Madrid, Spain.
[7]Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States.
[8]Centre for Genetics and Genomics Versus Arthritis Division of Musculoskeletal and Dermatological Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PL, United Kingdom.
[9]Department of Dermatology, University Hospital Maastricht, 6229 Maastricht, The Netherlands.
[10]GROW School for Oncology and Developmental Biology, Maastricht University Medical Center +, 6229 Maastricht, The Netherlands.
[11]Department of Clinical Genetics, Odense University Hospital, 5230 Odense, Denmark
[12]Hospital Clinical Institute, University of Southern Denmark, 5230 Odense, Denmark.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bjd.21842

[13]Institute of Human Genetics, Medical Center, University of Freiburg, 79106 Freiburg, Germany.
[14]School of Medicine, Medical Sciences, Nutrition and Dentistry, University of Aberdeen, Aberdeen AB25 2ZD, United Kingdom.
[15]Medical Genetics Department, NHS Grampian, Foresterhill, Aberdeen, AB25 2ZD, United Kingdom.
[16]Department of Dermatology, Zaandam Medical Center, 1502 Zaandam, The Netherlands.
[17]Department of Dermatology and Allergology, University Hospital of RWTH Aachen, 52062 Aachen, Germany.
[18]Department of Dermatology, Jeroen Bosch Hospital, 5223 's-Hertogenbosch, the Netherlands.
[19]Skin Research Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore 138543, Singapore.
[20]Lee Kong Chian School of Medicine, Nanyang Technological University (NTU), Singapore 636921, Singapore.
[21]Department of Pathology, Royal Liverpool University Hospital, Liverpool, L7 8XP, UK
[22]Adult Histopathology, Laboratory Medicine, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, M13 9WL, United Kingdom.
[23]Department of Dermatology, Salford Royal NHS Foundation Trust, Manchester, M6 8AD, United Kingdom.
[24]Dr Phillip Frost Department of Dermatology & Cutaneous Surgery, University of Miami Miller School of Medicine, Miami, FL 33125, United States.
[25]Monasterium Laboratory, Nano-Bioanalytik Zentrum, D - 48149 Münster, Germany
[26]Dept. of Dermatology, Venereology and Allergology, University Medical Center Göttingen, 37075 Göttingen, Germany

*,† Equal contribution

Correspondence: Xue Zhang

Email: xuezhang@pumc.edu.cn;

Funding: XZ is supported by the National Natural Science Foundation of China (NSFC) [grant number 81788101], the CAMS Innovation Fund for Medical Sciences (CIFMS) [grant numbers 2021-I2M-1-018 and 2016-I2M-1-002] and National Key Research and Development Program of China [grant number 2016YFC0905100]. MJS, DGE, RP and WGN are supported by the Manchester NIHR

Biomedical Research Centre (IS-BRC-1215-20007). ZM is supported by NHS Grampian Biorepository (NHS Scotland). JME is funded by a postdoctoral research fellowship from the Health Education England Genomics Education Programme (HEE GEP). GO and CS are funded by Wellcome Trust award (207491/Z/17/Z for GO, 215207/Z/19/Z for CS). GO, CS and SE are supported by Versus Arthritis award (21754) and NIHR Manchester BRC. SE is supported by Medical Research Council award (MR/N00017X/1).

Conflicts of interests: None to declare.

Data availability: Datasets of this study have not been deposited in a public repository due to privacy and ethical restrictions, but are available from the corresponding authors on request.

Ethics statement: The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national), and informed consents were obtained.

What is already known about this topic?

Bazex-Dupré-Christol syndrome (BDCS) is a rare X-linked basal cell carcinoma susceptibility syndrome that linked to an 11.4 Mb interval on chromosome Xq25-q27.1.

Loss-of-function variants in *ACTRT1* and its regulatory elements were suggested to cause BDCS.

What does this study add?

BDCS is caused by small tandem noncoding intergenic duplications at chromosome Xq26.1.

The Xq26.1 BDCS duplications likely dysregulate *ARHGAP36*, the flanking centromeric gene.

ACTRT1 loss-of-function variants are unlikely to cause BDCS.

What is the translational message?

This study provides the basis for accurate genetic testing for BDCS which will aid precise diagnosis and appropriate surveillance and clinical management.

ARHGAP36 may be a novel therapeutic target for all forms of sporadic BCCs.

Abstract

Background: Bazex-Dupré-Christol syndrome (BDCS; MIM301845) is a rare X-linked dominant genodermatosis characterized by follicular atrophoderma, congenital hypotrichosis and multiple basal cell carcinomas (BCCs). Previous studies have linked BDCS to an 11.4 Mb interval on chromosome Xq25-q27.1. However, the genetic mechanism of BDCS remains an open question.

Objectives: To investigate the genetic etiology and molecular mechanisms underlying BDCS.

Methods: We ascertained multiple individuals, from eight unrelated families, affected with BDCS (F1-F8). Whole exome (F1 and F2) and genome sequencing (F3) were performed to identify putative disease-causing variants within the linkage region. Array-comparative genomic hybridization and quantitative PCR were used to explore copy number variations (CNV), followed by long-range gap-PCR and Sanger sequencing to amplify the duplication junctions and to define the head-tail junctions. Hi-C was performed on dermal fibroblasts from two affected individuals with BDCS and one control. Public datasets and tools were used to identify regulatory elements and transcription factor binding sites within the minimal duplicated region. Immunofluorescence was performed in hair follicles, BCCs and trichoepitheliomas from BDCS patients and sporadic BCCs. The *ACTRT1* variant (p.Met183Asnfs*17), previously proposed to cause BDCS, was evaluated with allele frequency calculator.

Results: In eight BDCS families, we identified overlapping 18-135kb duplications (six inherited and two *de novo*) at Xq26.1, flanked by *ARHGAP36* and *IGSF1*. Hi-C showed the duplications didn't affect the topologically associated domain (TAD), but may alter the

interactions between flanking genes and putative enhancers located in the minimal duplicated region. We detected ARHGAP36 expression near the control hair follicular stem cells compartment, and found increased ARHGAP36 levels in hair follicles in telogen, BCCs and trichoepitheliomas from patients with BDCS. ARHGAP36 was also detected in sporadic BCCs from individuals without BDCS. Our modelling showed the predicted *ACTRT1* variants maximum tolerated minor allele frequency in control populations to be orders of magnitude higher than expected for a high-penetrant ultra-rare disorder, suggesting loss-of-function of *ACTRT1* variants to be an unlikely cause for BDCS.

Conclusions: Noncoding Xq26.1 duplications cause BDCS., The BDCS duplications most likely lead to dysregulation of ARHGAP36. ARHGAP36 is a potential therapeutic target for both inherited and sporadic BCCs.

Introduction

Bazex-Dupré-Christol syndrome (BDCS, also called Bazex syndrome or follicular atrophoderma and basal cell carcinomas; MIM 301845) is a rare X-linked disorder characterized by congenital hypotrichosis, follicular atrophoderma (seen as "ice pick" marks, usually on the dorsum of hands and feet) and susceptibility to develop basal cell nevi and basal cell carcinomas (BCCs). The nevi and BCCs generally occur on sun-exposed areas, including the head, neck and face from the second decade of life [1]. Other features reported in some individuals with BDCS include persistent milia, hyperpigmented macules, hypohidrosis, and trichoepitheliomas (benign tumors arising from basal cells in the hair follicles, which rarely transform to BCCs) [2-9].

Clinically, BDCS overlaps with basal cell nevus syndrome (Gorlin syndrome; MIM 109400), which predisposes to multiple BCCs and is caused by heterozygous germline variants in *PTCH1* [10,11] or *SUFU* [12]. Both genes encode members of the hedgehog signaling pathway, and variants in them result in dysregulated overexpression of Gli transcription factors. BCC is the most common skin cancer [13], with somatic mutations in genes encoding key components of Hedgehog-Patched-Gli signalling often present in sporadic BCCs [14]. BCCs arise from hair follicle stem cells and/or other epithelial stem cells reprogrammed to a follicular differentiation [15,16]. Of note, X-linked inheritance pattern is unusual for a cancer predisposition syndrome. The 'two-hit model' for familial cancers caused by loss of heterozygosity of the remaining functional allele in an individual with a germline loss-of-function variant is unlikely to occur for an X-linked tumor suppressor gene. Hence, we hypothesized that BDCS is caused by genetic

variants resulting in altered Hedgehog-Patched signaling or -Gli activity in follicular stem cells via a mechanism that does not result in direct loss of function.

Methods

Patient ascertainment

The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national), and informed consents were obtained. We identified patients and families diagnosed with BDCS based on clinical history and examination.

Sanger sequencing

A single primer pair (ACTRT1-F: TAGGTATGATTTGCTTTCCTTGGC, ACTRT1-R: CAACCTAAAGATTCATGACATGACTC) was designed to amplify the full length of *ACTRT1*, which encompassed the single coding exon and its 5' and 3' untranslated regions (UTRs). At least one affected individual from each family underwent Sanger sequencing.

All coding exons of *ARHGAP36* (exons 2-12) were amplified and Sanger sequenced on an ABI3730xl DNA Analyser (Life Technologies, Paisley, UK).

Exome and Genome sequencing

Whole exome sequencing was performed by using SureSelect kits (Human All Exon 50Mb for Families 1 and 2 and Human All Exon v.5 for Family 3) (both Agilent, Santa Clara, CA, USA) were used for whole exome sequencing. Paired-end sequencing (~100bp) was performed on Illumina HiSeq2000 (Family 1 and 2) or HiSeq2500 platforms (Family 3). For the first two families, a minimum of 4.3 Gb of high-quality

mappable data was generated, yielding a mean depth of coverage of 40-fold and 84% of target bases sequenced at 10x coverage. A minimum of 4.5 Gb of sequence was generated, yielding a mean depth of coverage of 80-fold and 95% of target bases sequenced at 20x coverage for family 3. The sequence data were mapped to the human reference genome (hg19) using Burrows Wheeler Aligner [17]. Variant calling was performed using the GenomeAnalysisToolKit-v2.4.7. (GATK) software [18].

Genome sequence data was generated by Complete Genomics (Mountain View, California, USA) as described previously [19]. Bioinformatics (alignment to the hg19 reference genome, local de novo assembly and variant calling) was performed using version 2.5 of the Complete Genomics pipeline [20].

Array-comparative genomic hybridisation (a-CGH)

A-CGH was performed using Affymetrix SNP6.0 microarray as described previously [21].

Real-time quantitative PCR (qPCR)

To validate the genomic duplications detected by aCGH and determine the boundaries of duplications in the three families and other families, qPCR was performed as previously described [22]. For families 1, 2 and 7, the primers for qPCR were designed according to aCGH results. For the other families, primer pair XQM located in the middle of the shared duplicated region (chrX:130348186-130348315) was designed. The quantification of the target regions was normalized to an assay from chromosome 21. The relative copy number (RCN) was determined with the comparative ΔΔCT method, using DNA from a normal male as the calibrator. All assays were repeated three times. A ~2-fold RCN indicated duplication in samples from males and ~3-fold

RCN for samples from females. Primers used for each family were listed in Supp Table 4.

Long-range Gap-PCR

A series of qPCR primers was designed to walk through to refine the duplication boundaries. Once the boundaries were estimated by qPCR, the forward primer from the centromeric side of the duplication region that was nearest to the determined boundaries, as well as the reverse primer from the telomeric side were chosen to perform long-range gap-PCR in order to amplify the duplication junctions. F3, F5 and F8 used the same primers pair as F1. Sanger sequencing was conducted to define the precise breakpoints. Primers used for each family were listed in Supp Table 4.

Population screening

Primer pair XQM, as described above, was used to determine the frequency of duplications at the disease associated locus by qPCR. The criteria defining duplication was the same as above.

Haplotype analysis

Four common single nucleotide polymorphisms (SNPs) (rs62603806, rs4240127, rs5932866, rs12559533) within the duplicated region were amplified with primer pair (SNP-F: GCACAGATGATTATGTCTGTTCC, SNP-R: CTGTCCCTACTTAGTAAATCGAG) and Sanger sequenced to generate haplotypes in the male patients of F1, F3, and F8.

Hi-C

2-5 million cells were fixed with 2% (v/v) formaldehyde for 10 minutes and then snap frozen. Hi-C libraries were generated using the Arima Hi-C kit following manufacturer's

instructions. Sequencing was performed on a NovaSeq platform with 150bp paired end reads at a target depth of 600 million reads per sample. The sequencing data was filtered and adapters were removed using fastp v0.19.4 [23]. The reads were then mapped to the GRCh38 genome with Hi-C Pro v2.11.1 [24], using default settings. Files for visualisation were created using the hicpro2juicebox.sh utility and visualised in Juicebox [25] and a modified version of coolbox [26]. Maps were normalized with the KR (balancing) algorithm. The increased interactions from the duplications disappeared after normalization which corrects for many biases including accessibility and read depth, indicating that the increase was not higher than what the increased amount of DNA would suggest.

Public datasets for GM12878, hESC and HFF were downloaded from the 4D Nucleome project website [27]. Data for the Jurkat cell line was generated in a similar way to the dermal fibroblasts.

Identification of putative enhancer and transcription factor binding sites

Putative enhancer was visualized in UCSC genome browser (http://genome-asia.ucsc.edu/) with chromHMM [28] tracks from Roadmap Epigenomics Consortium. Imputed enhancers and bivalent promoter in H1- and H9-cells were selected to visualize. The hg38 genomic co-ordinates of transcription factor binding sites for ESX1 and KLF4 from JASPAR [29] were identified.

Immunohistochemistry

Immunohistochemistry was performed on 7um frozen sections by fixation in chilled paraformaldehyde (4% v/v) for 10 minutes. After washing with phosphate buffered

saline (PBS), sections were permeabilised with PBS containing 0.1% Triton X100 for ten minutes. Primary antibodies were incubated over night at 4°C at a concentration of 1:200 (ARHGAP36, HPA002064, Sigma, Dorset, UK), 1:20 (IGSH, HPA035582, Sigma) or 1:100 (ACTRT1, HPA003119, Sigma). For dual stains cytokeratin 15 (Abcam, ab80522, Cambridge, UK), at a concentration of 1:500 was added and incubated overnight. Antibodies were detected by incubating with an Alexafluor goat anti mouse/rabbit secondary (Sigma) antibody at a concentration of 1:200 for 45 minutes. Sections were counterstained with DAPI.

For tumor tissue, paraffin sections were de-waxed using xylene and rehydrated using decreasing concentrations of ethanol (100%-50%). Antigen retrieval was performed by boiling in 10mM citrate buffer (pH 6) for 20 minutes. Following this the protocol as above was followed with P63 (Abcam-ab735) at a concentration of 1:100.

ARHGAP36 expression in sporadic BCCs

RNAseq data of 35 sporadic BCCs and paired adjacent normal tissues, 18 BCCs and 8 normal tissues, were obtained from Gene Expression Omnibus (GEO) under accession codes (GSE125285 [30] , GSE58375 [31] and GSE128786 [32]), respectively. Reads were aligned to human whole genome (hg38) with STAR (v2.7.0) [33]. Gene counts from BAM files were calculated with FeatureCounts [34]. The transcripts read per thousand bases per million mapping (TPM) were determined with edgeR (Version 3.7) [35]. Levels of *ARHGAP36* were compared with t-test for GSE125285 (paired), GSE58375 and GSE128786 (unpaired) respectively. P-values < 0.05 were considered to be significant.

Pulldown of ARHGAP36 with Rac1

The GAP domain (aa221-431) of Arhgap36 was cloned from the cDNA of Kelly cells and inserted into the FLAG-HA tag vector pCDNA3 (Addgene #52535). The Arhgap36 GAP domain was ectopically expressed in HEK293T cells via transfection using JetOPTIMUS (Polyplus #117-01). Recombinant GST tagged Rac1 with the activating Q61L mutation was then used to precipitate the Arhgap36 GAP domain as has previously been described (PMID: 16472675). Precipitated Arhgap36 GAP domain was detected via Western blotting with anti-HA antibody (Cell Signaling Technology #3724).

Allele frequency modelling

Allele frequency modelling of *ACTRT1* was performed using the allele frequency calculator from http://cardiodb.org/allelefrequencyapp [36].The maximum tolerated reference allele count (0.95 CI) for the *ACTRT1* NM_138289.3:c.547dup (p.(Met183Asnfs*17)) variant was calculated in 204,684 in alleles (https://gnomad.broadinstitute.org/variant/X-127185638-A-AT?dataset=gnomad_r2_1). The adjustable parameters were – estimated population prevalence of BDCS (1 in106) (Orpha.net); allelic heterogeneity of 33% (BDCS in 2 out of 6 families in Bal et al was attributed to this variant); and penetrance of 100% (based on description of families in the literature. Next, to account for potential inaccuracies in previous estimates, we modelled the maximum tolerated reference allele counts with various combinations of prevalence or penetrance values.

Results

BDCS is caused by small tandem intergenic duplications at chromosome Xq26.1

Previous mapping has linked BDCS to an 11.4 Mb interval on chromosome Xq25-q27.1 [37,38]. To investigate the genetic etiology of BDCS, we ascertained eight families (F1-8) with individuals affected with BDCS (Fig. 1). This cohort included five previously published families (F1 [38], F4 [39], F5 [40], F7 [3], and F8 [41]). Whole exome (F1 and F2) or genome (F3) sequencing did not identify putative disease-causing variants within the Xq26.1 locus previously linked to BDCS [38]. Array-comparative genomic hybridization in at least one affected individual from F1, F2, F3, and F7 revealed small intergenic Xq26.1 gains of varying sizes (Fig. 2A) [38]. Further, qPCR assays in affected individuals from the other four families (F4, F5, F6, and F8) were consistent with gains at this locus. qPCR analysis also confirmed that the gains segregated with BDCS in the multiplex families or were *de novo* in the two simplex cases (Fig. 2B). Long-range gap-PCR to amplify the duplication junctions followed by Sanger sequencing defined the head-tail junctions for F2, F4, F6 and F7 and confirmed these gains as tandem duplications (Fig. 2C). The duplications in F1, F3, F5, and F8 could not be differentiated further, likely due to homologous L1 elements (Fig. 2D). In the three multiplex families (F1, F3 and F8) with seemingly identically sized duplications, SNP haplotype mapping proved their likely independent origin (Fig. 2E). No gains were identified by qPCR in 215 unrelated European controls (139 females and 76 males). The largest gain was detected in F2 (135kb) and the 18kb gain in F6 defined the smallest shared overlapping region (hg38, chrX:131,207,776-131,226,336) (Fig. 2D).

No similar sized exclusively intergenic gains overlapping with the shared duplicated region defined by the eight families were detected in control populations (Fig. 3). One

entirely intergenic gain (nsv517789; hg38, chrX:131,100,726-131,206,649) of ~100kb, in an individual with no known phenotype was noted, but it did not overlap the shared duplicated region as defined by our study (Fig. 3) [42]. Another gain of >380kb (nsv528179; hg38, chrX:131,100,726-131,206,649) overlapping the shared BDCS duplicated region, in an individual with no known phenotype was noted, but it extended ~295kb beyond the most telomeric boundary of the BDCS associated duplications and encompasses *IGSF1* (Fig. 3). We also found that several other larger chromosome X duplications encompassing the region have been reported in individuals without BDCS in the DECIPHER database (Table S1, Fig. 3) [43].

BDCS duplications affect active transcription factor binding sites, but do not affect the topologically associated domain structure

None of the BDCS duplications encompass protein-coding genes (Fig. 2D, Fig. 3). To explore if these duplications altered topologically associated domains, we performed Hi-C on dermal fibroblasts from two affected individuals with BDCS (F3:I-2 and F5:II-1) and one control. This showed that a single topologically associated domain (TAD) (hg38, ~chrX:130,956,000-136,686,000) contained the BDCS duplications and the flanking genes, *ARHGAP36* and *IGSF1* (Fig.3). Examination of publicly available Hi-C and microC datasets showed this TAD to be conserved across multiple human cell lineages (Figure S1)[44,45]. No significant changes were visible in the interactions between the regions of the duplication and the two loci in the affected individuals.

Next, we interrogated the minimal duplicated region (hg38, chrX:131,207,776-131,226,336) for putative regulatory elements in human embryonic stem cell lines, from which we found potential enhancers located inside the minimal duplicated region (Fig.3). We also identified predicted motifs for the binding of transcription factors ESX1 and KLF4

(Fig.3). Of particular interest, KLF4 organizes and regulates pluripotency-associated three-dimensional enhancer networks, is expressed in hair follicle stem cells, is required for their differentiation and is expressed in BCC cells [46,47]. Together, these data suggest that the BDCS duplications are unlikely to cause its dysregulation via disruption of the TAD, but could alter enhancer-transcription factor-gene interactions, leading to dis-regulation of genes within the same TAD.

ARHGAP36 expression is dysregulated in BDCS

BDCS is considered to be a disorder of the hair follicle [5]. We, therefore, performed immunofluorescence for the two flanking genes, *ARHGAP36* and *IGSF1* in control hair follicles. *IGSF1*, the flanking telomeric gene, encodes member 1 of the immunoglobulin superfamily and loss-of-function variants in this gene cause an X-linked recessive syndrome of central hypothyroidism and testicular enlargement (MIM 300888) [48]. There is no known role of IGSF1 in Hedgehog-Patched-Gli pathway. *ARHGAP36*, the flanking centromeric gene, encodes a Smoothened (Smo)-independent positive regulator of the sonic hedgehog (shh) pathway and its expression is upregulated in medulloblastomas, a Hedgehog-Patched-Gli pathway-related tumor [49,50]. Variants in *ARHGAP36* have not been associated with any inherited disorder. Hence, the known biological role makes ARHGAP36 an excellent candidate for BDCS pathology.

In anagen, IGSF1 was present in the terminally differentiated inner root sheath (IRS), but was absent from the actively proliferating hair matrix or bulge, the main reservoir of hair follicle epithelial stem cells (Fig. S3a). In telogen, there was no evidence of IGSF1 staining (data not shown). In contrast, ARHGAP36 was present in a small number of hair follicle cells in the outer root sheath (ORS) at the level of the stem cell bulge, in both anagen (Fig. 4A) and telogen (Fig. 4C). Notably, it is at the end of telogen that the stem cells

located in the secondary hair germ and bulge regions are activated to resume hair growth [51]. There was no obvious difference regarding ARHGAP36 positive cell numbers in the ORS between hair follicles from healthy control and an individual with BDCS (F5:II-1) in anagen (Fig. 4A and 4B). In contrast, in telogen hair follicle from F5:II-1, there was a marked increase in the number of ARHGAP36 positive cells around the epithelial stem cell compartment (Fig. 4D), comparing with healthy control in telogen (Fig.4C) and the same patient in anagen (Fig. 4B). Moreover, immunofluorescence in histologically confirmed p63 positive BCC [52] from F4:III-4 showed strong staining for ARHGAP36 in a proportion of cells (Fig. 5A-B). Notably, the BCCs did not immunofluoresce for IGSF1 (Fig. S4a). We also detected striking ARHGAP36 staining of a trichoepithelioma from the same individual (Fig. 5C). Collectively, these data show dysregulation of ARHGAP36 in BDCS.

ARHGAP36 is expressed in sporadic BCCs

Next, to explore if ARHGAP36 could be relevant to sporadic BCCs, we determined the presence of ARHGAP36 in superficial (n=10), nodular (n=10), and infiltrative (n=10) sporadic BCCs. Similar to the BCCs in BDCS, ARHGAP36 was present in a small proportion of cells from all examined tumors (Fig. 5D-F), but was absent from all surrounding tissue. Next, we checked published RNAseq datasets of sporadic BCCs (GSE125285 [30], GSE58375 [31] and GSE128786 [32]), slightly higher expression of *ARHGAP36* was observed in BCCs in both public datasets (Fig.S5). However, the difference was not significant in GSE125285, which could be due to low expression level of *ARHGAP36* and small number of sample size. On the other hand, the localization of ARHGAP36 positive cells around the epithelial stem cell compartment and in

trichoepithelioma in BDCS patient, in contrast with a small proportion of positive cells in different types of sporadic BCCs, indicated the role of ARHGAP36 in the early stage of BCC formation. Together these results suggest that ARHGAP36 is expressed in sporadic BCCs.

ARHGAP36 interacts with RAC1

Individuals with BDCS show additional 'non-cancer 'phenotypes e.g. hypotrichosis, which are not seen in other disorders of the shh pathway indicating additional biological roles for ARHGAP36 in hair follicle stem cell biology. We noted that ARHGAP36 is a member of the Rho GAP family of regulatory proteins, which deactivate Rho proteins, and Rac1, one of the most important Rho GTPases [53], is essential for hair follicle stem cell function [54]. *Rac1* conditional knockdown in mice follicular stem cells results in severely defective hair development [54]. Hence, we hypothesized that aberrant or overexpression of *ARHGAP36* may inhibit activation of Rac1 and thus result in defective hair follicle phenotype present in BDCS. We, therefore, performed pulldown assays using a clone of the GAP domain of ARHGAP36 and a recombinant constitutively active RAC1 mutant. The results demonstrated binding of the two proteins and the potential effect of ARHGAP36 dysregulation impacting on Rac1 signaling (Fig.S6).

***ACTRT1* loss-of-function variants are unlikely to cause BDCS**

Knudson's two-hit hypothesis explains the mechanism of most inherited cancer syndromes, in which tumors occur following somatic loss of the only functional allele in an individual with a germline loss-of-function variant in a tumor suppressor gene [55]. Presence of a single X-chromosome in males and X-inactivation in females, make X-linked inherited cancer predisposition syndromes due to tumor gene suppression highly

unlikely. Previously, Bal *et al* reported loss-of-function *ACTRT1* variants in BDCS [41]. They identified a frameshift *ACTRT1* NM_138289.3: c.547dup (p.Met183Asnfs*17) variant (rs771087307), in two families with BDCS [41]. Variants in conserved non-coding elements flanking *ACTRT1* were also proposed as pathogenic. Our modelling revealed that the predicted maximum tolerated minor allele frequency (MAF) for the p.Met183Asnfs*17 variant in population control data to be ~10^4 times higher than expected and could be reconciled only if the previous prevalence and penetrance estimates were inaccurate by several orders of magnitude (Fig.S7). Furthermore, other putative loss-of-function variants in the single exon of *ACTRT1* have been reported in both males and females without BDCS (Tables S2-3). Sanger sequencing in at least one affected individual from seven out of eight families reported here (F1-F7) did not identify any rare coding variants in *ACTRT1*. As expected, the p. Met183Asnfs*17 variant was present in the affected individuals (II-1 and III-1) in F8, as reported in the original association paper [41]. Importantly, immunofluorescent staining showed that ACTRT1 was absent from both the disease-relevant regions of the hair follicles and tumors from individuals with BDCS (Fig. S3b and S4b). These combined data provide evidence that *ACTRT1* loss-of-function variants are unlikely to cause BDCS.

DISCUSSION

We show that BDCS is caused by small intergenic tandem duplications at Xq26.1 that encompass a minimum shared overlapping region of ~18Kb (hg38: chrX:131,207,776-131,226,336) (Fig. 2) and that the BDCS duplications likely cause the condition due to cell-specific and hair-follicle-cycle specific dysregulation of *ARHGAP36*. Identifying causal non-coding variants and deciphering their disease mechanism is a major

challenge in modern genomics. To the best of our knowledge, BDCS is the first example of an inherited cancer predisposition disease caused by germline CNVs that do not encompass any protein coding genes.

Our data suggest that larger Xq26.1 duplications encompassing the flanking genes in addition to the minimum common shared region may not result in BDCS. Also, smaller 'non-coding' duplications at Xq26.1 that do not overlap the shared minimum overlapping region may not cause BDCS. These observations reflect the complexity of the genetic diagnosis of BDCS and assigning pathogenicity to Xq26.1 duplications in diagnostic labs will require careful consideration of the size and location of the copy number gains at this position. Future studies will be needed to validate our findings and to refine the minimum common overlapping region for BDCS duplications.

Our Hi-C studies on fibroblasts showed that the BDCS duplications do not affect the TAD chromosomal structure. Performing Hi-C experiments on patient follicular stem cells (preferably in the appropriate stage of differentiation), rather than on dermal fibroblasts, could be more informative. However, obtaining sufficient number of stem cells from hair follicles from an individual with BDCS proved unsuccessful. The precise mechanism of how BDCS duplications result in cell-specific and hair-follicle-cycle specific dysregulation of *ARHGAP36* expression remains to be elucidated.

Franke *et al.* found duplications upstream *Sox9* led to different phenotypes. Further investigation suggested that intergenic duplication that located within the same TAD of Sox9 (Intra-TAD), which didn't disrupt the TAD structure, resulted in female to male sex reversal in humans due to increased contact of duplicated regions with the rest of TAD. In contrast, longer duplications containing the sex reversal duplications and extending to

neighboring TAD (inter-TAD) led to formation of new TAD structure (neo-TAD) that isolated from the rest of genome. The consequences of inter-TAD duplications depended on whether the flanking genes were involved [56]. We hypothesize that similar mechanisms might explain the connection between BDCS duplications and ARHGAP36 dysregulation in the present study. The existence of intra-TAD duplications increased the interactions between putative enhancer regions and *ARHGAP36*, leading to temporal and/or spatial upregulation of *ARHGAP36* (Fig.6).

Several duplications have been reported in DGV and DECIPHER that overlapped with BDCS duplications (Fig.3). We noticed two duplications, nsv517789 in DGV and 267829 in DECIPHER located within the TAD and showed no BDCS phenotype. It is noteworthy that both duplications were not validated with further experiments. We are not sure whether they are tandem duplications. Moreover, both duplications are not intergenic and encompass a flanking coding gene, *IGSF1*. The extension of duplications to *IGSF1* may disrupt structure of nested-TAD, or subTADs, which are domain-like TADs with cell-type-dynamic folding properties and boundaries, leading to distant phenotypes [57]. Further experiments are needed to elucidate the mystery underlying different range of tandem duplications within the same TAD resulted in distant phenotypes.

Interestingly, our results suggest that ARHGAP36 is also relevant to sporadic BCC pathology (Fig. 5D-F). Of note, ARHGAP36 did not co-localize with p63 (Fig. 5C-F) and the higher levels of ARHGAP36 in trichoepithelioma than in the BDCS-associated and sporadic BCCs suggests that dysregulation of ARHGAP36 may be an early step in the pathogenesis to BCC. Nevertheless, ARHGAP36 could be a biomarker and therapeutic

target for inhibition in individuals with both inherited and, the vastly more common, non-inherited forms of BCC.

The demonstration that ARHGAP36 interacts with RAC1 potentially uncovers a previously unknown role of ARHGAP36 and may explain the non-cancerous phenotypes of BDCS, namely hypotrichosis. However, further confirmatory studies will be required to prove the regulation of RAC1 by ARHGAP36 and explore the effects of this regulation. Also, it will be interesting to investigate if ectopic expression of ARHGAP36 can result in hypotrichosis.

In summary, we have shown that small intergenic non-coding tandem duplications at Xq26.1 encompassing chrX:131,207,776-131,226,336 (hg38) cause BDCS. This is the first example of an inherited cancer predisposition disease caused by germline non-coding CNVs. We propose that the duplications result in the dysregulation of *ARHGAP36* that underlies BDCS pathology. Our findings reconcile the molecular mechanism of BDCS, a tumor-predisposition syndrome, with its X-linked inheritance pattern. We also suggest that ARHGAP36 is relevant to sporadic BCCs and a potential therapeutic target.

Authors' contributions:

XZ, JF and WGN conceived the project. XZ designed and coordinated the genetics study. WGN, RP and SB designed and directed the histopathology study. YL, YH, MGK, MJB, GO, SE and JTL conducted the experiments. YL, SB, YH, JHS, DP, GMB, MGK, CF, CS, SC, JME, MJS, IHC, AT, MvG, AB, JF, ZM, FA, AR, DGE, LJMTW-P, MAMvS, DCM, JF and WGN collected clinical data. SB, YL, YH, JF, WGN and XZ wrote the manuscript.

References:

1 Bazex AD, A.; Christol, B. Genodermatose complexe de type indetermine associant une hypotrichose, un etat atrophodermique generalise et des degenerescences cutanees multiples. *Bull. Soc. Franc. Derm. Syph* 1964; **71**.

2 Plosila M, Kiistala R, Niemi KM. The Bazex syndrome: follicular atrophoderma with multiple basal cell carcinomas, hypotrichosis and hypohidrosis. *Clin Exp Dermatol* 1981; **6**: 31-41.

3 Herges A, Stieler W, Stadler R. [Bazex-Dupre-Christol syndrome. Follicular atrophoderma, multiple basal cell carcinomas and hypotrichosis]. *Hautarzt* 1993; **44**: 385-91.

4 Vabres P, de Prost Y. Bazex-Dupre-Christol syndrome: a possible diagnosis for basal cell carcinomas, coarse sparse hair, and milia. *Am J Med Genet* 1993; **45**: 786.

5 Goeteyn M, Geerts ML, Kint A *et al.* The Bazex-Dupre-Christol syndrome. *Arch Dermatol* 1994; **130**: 337-42.

6 Yung A, Newton-Bishop JA. A case of Bazex-Dupre-Christol syndrome associated with multiple genital trichoepitheliomas. *Br J Dermatol* 2005; **153**: 682-4.

7 Barcelos AC, Nico MM. Bazex-Dupre-Christol syndrome in a 1-year-old boy and his mother. *Pediatr Dermatol* 2008; **25**: 112-3.

8 Castori M, Castiglia D, Passarelli F *et al.* Bazex-Dupre-Christol syndrome: an ectodermal dysplasia with skin appendage neoplasms. *Eur J Med Genet* 2009; **52**: 250-5.

9 Kallam AR, Satyanarayana MA, Aryasomayajula S *et al.* Basal Cell Carcinoma Developing from Trichoepithelioma: Review of Three Cases. *J Clin Diagn Res* 2016; **10**: PD17-9.

10 Hahn H, Wicking C, Zaphiropoulous PG *et al.* Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell* 1996; **85**: 841-51.

11 Johnson RL, Rothman AL, Xie J *et al.* Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science* 1996; **272**: 1668-71.

12 Smith MJ, Beetz C, Williams SG *et al.* Germline mutations in SUFU cause Gorlin syndrome-associated childhood medulloblastoma and redefine the risk associated with PTCH1 mutations. *J Clin Oncol* 2014; **32**: 4155-61.

13 Peris K, Fargnoli MC, Garbe C *et al.* Diagnosis and treatment of basal cell carcinoma: European consensus-based interdisciplinary guidelines. *Eur J Cancer* 2019; **118**: 10-34.

14 Epstein EH. Basal cell carcinomas: attack of the hedgehog. *Nat Rev Cancer* 2008; **8**: 743-54.

15 Peterson SC, Eberl M, Vagnozzi AN *et al.* Basal cell carcinoma preferentially arises from stem cells within hair follicle and mechanosensory niches. *Cell Stem Cell* 2015; **16**: 400-12.

16 Youssef KK, Lapouge G, Bouvree K *et al.* Adult interfollicular tumour-initiating cells are reprogrammed into an embryonic hair follicle progenitor-like fate during basal cell carcinoma initiation. *Nat Cell Biol* 2012; **14**: 1282-94.

17 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754-60.

18 McKenna A, Hanna M, Banks E *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297-303.

19 Drmanac R, Sparks AB, Callow MJ *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010; **327**: 78-81.

20 Carnevali P, Baccash J, Halpern AL *et al.* Computational techniques for human genome resequencing using mated gapped reads. *J Comput Biol* 2012; **19**: 279-92.

21 Banka S, Blom HJ, Walter J *et al.* Identification and characterization of an inborn error of metabolism caused by dihydrofolate reductase deficiency. *Am J Hum Genet* 2011; **88**: 216-25.

22 Sun M, Ma F, Zeng X *et al.* Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. *J Med Genet* 2008; **45**: 589-95.

23 Chen S, Zhou Y, Chen Y *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018; **34**: i884-i90.

24 Servant N, Varoquaux N, Lajoie BR *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 2015; **16**: 259.

25 Durand NC, Robinson JT, Shamim MS *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* 2016; **3**: 99-101.

26 Cao M. Coolbox, https://github.com/GangCaoLab/CoolBox. In. 2020.

27 Dekker J, Belmont AS, Guttman M *et al.* The 4D nucleome project. *Nature* 2017; **549**: 219-26.

28 Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 2017; **12**: 2478-92.

29 Fornes O, Castro-Mondragon JA, Khan A *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2020; **48**: D87-D92.

30 Wan J, Dai H, Zhang X *et al.* Distinct transcriptomic landscapes of cutaneous basal cell carcinomas and squamous cell carcinomas. *Genes Dis* 2021; **8**: 181-92.

31 Atwood SX, Sarin KY, Whitson RJ *et al.* Smoothened variants explain the majority of drug resistance in basal cell carcinoma. *Cancer Cell* 2015; **27**: 342-53.

32 Sand M, Bromba A, Sand D *et al.* Dicer Sequencing, Whole Genome Methylation Profiling, mRNA and smallRNA Sequencing Analysis in Basal Cell Carcinoma. *Cell Physiol Biochem* 2019; **53**: 760-73.

33 Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* 2015; **51**: 11 4 1- 4 9.

34 Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; **30**: 923-30.

35 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; **26**: 139-40.

36 Whiffin N, Minikel E, Walsh R *et al.* Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med* 2017; **19**: 1151-8.

37 Vabres P, Lacombe D, Rabinowitz LG *et al.* The gene for Bazex-Dupre-Christol syndrome maps to chromosome Xq. *J Invest Dermatol* 1995; **105**: 87-91.

38 Parren LJ, Abuzahra F, Wagenvoort T *et al.* Linkage refinement of Bazex-Dupre-Christol syndrome to an 11.4-Mb interval on chromosome Xq25-27.1. *Br J Dermatol* 2011; **165**: 201-3.

39 Kidd A, Carson L, Gregory DW *et al.* A Scottish family with Bazex-Dupre-Christol syndrome: follicular atrophoderma, congenital hypotrichosis, and basal cell carcinoma. *J Med Genet* 1996; **33**: 493-7.

40 Torrelo A, Sprecher E, Mediero IG *et al.* What syndrome is this? Bazex-Dupre-Christol syndrome. *Pediatr Dermatol* 2006; **23**: 286-90.

41 Bal E, Park HS, Belaid-Choucair Z *et al.* Mutations in ACTRT1 and its enhancer RNA elements lead to aberrant activation of Hedgehog signaling in inherited and sporadic basal cell carcinomas. *Nat Med* 2017; **23**: 1226-33.

42 MacDonald JR, Ziman R, Yuen RK *et al.* The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* 2014; **42**: D986-92.

43 Firth HV, Richards SM, Bevan AP *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 2009; **84**: 524-33.

44 Rao SS, Huntley MH, Durand NC *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014; **159**: 1665-80.

45 Krietenstein N, Abraham S, Venev SV *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell* 2020; **78**: 554-65 e7.

46 Xu M, Horrell J, Snitow M *et al.* WNT10A mutation causes ectodermal dysplasia by impairing progenitor cell proliferation and KLF4-mediated differentiation. *Nat Commun* 2017; **8**: 15397.

47 Colmont CS, Benketah A, Reed SH *et al.* CD200-expressing human basal cell carcinoma cells initiate tumor growth. *Proc Natl Acad Sci U S A* 2013; **110**: 1434-9.

48 Sun Y, Bak B, Schoenmakers N *et al.* Loss-of-function mutations in IGSF1 cause an X-linked syndrome of central hypothyroidism and testicular enlargement. *Nat Genet* 2012; **44**: 1375-81.

49 Rack PG, Ni J, Payumo AY *et al.* Arhgap36-dependent activation of Gli transcription factors. *Proc Natl Acad Sci U S A* 2014; **111**: 11061-6.

50 Beckmann PJ, Larson JD, Larsson AT *et al.* Sleeping Beauty Insertional Mutagenesis Reveals Important Genetic Drivers of Central Nervous System Embryonal Tumors. *Cancer Res* 2019; **79**: 905-17.

51 Schneider MR, Schmidt-Ullrich R, Paus R. The hair follicle as a dynamic miniorgan. *Curr Biol* 2009; **19**: R132-42.

52 Reis-Filho JS, Torio B, Albergaria A *et al.* p63 expression in normal skin and usual cutaneous carcinomas. *J Cutan Pathol* 2002; **29**: 517-23.

53 Reijnders MRF, Ansor NM, Kousi M *et al.* RAC1 Missense Mutations in Developmental Disorders with Diverse Phenotypes. *Am J Hum Genet* 2017; **101**: 466-77.

54 Castilho RM, Squarize CH, Patel V *et al.* Requirement of Rac1 distinguishes follicular from interfollicular epithelial stem cells. *Oncogene* 2007; **26**: 5078-85.

55 Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 1971; **68**: 820-3.

56 Franke M, Ibrahim DM, Andrey G *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 2016; **538**: 265-9.

57 Beagan JA, Phillips-Cremins JE. On the existence and functionality of topologically associating domains. *Nat Genet* 2020; **52**: 8-16.

Figure Legends

Fig. 1. Individuals with BDCS and pedigrees of families. Pedigrees of all families included in this study are shown. Standard symbols have been used to denote sex and affected status. * indicates individuals from whom DNA samples were available for analysis. At least one affected individual from each family underwent ACTRT1 Sanger sequencing including its full coding region and the 5' and 3' untranslated regions. (B-F) Clinical features of BDCS are shown. The photographs show a 15 years old male (F5-II:1) with hypotrichosis and wiry curly hair (B), sparse eyebrows and persistent milia (C) and follicular atrophoderma on dorsum of hand seen as "ice pick" marks (D); a 7y old female (F1-VI:2) with less severe hypotrichosis than the male, but with prominent persistent milia and a BCC in the left malar region (E); a lateral view of face in the same person (F).

Fig. 2. Small intergenic duplications of chromosome Xq26.1 in individuals with BDCS. (A) Array comparative genomic hybridisation in four individuals with BDCS, demonstrating intergenic copy number gains at the Xq26.1. The top four panels show the corresponding copy number of F1:IV-3, F2:II-2, F3:III-3 and F7:II-1 between chrX: 131,000,000-131,350,000 (hg38), respectively. The bottom panel shows the flanking RefSeq protein coding genes. (B) qPCR of an amplicon at the duplicated locus demonstrating normal copy number (white bar) in the unaffected mothers of individuals F5:II-1 and F6:II-1 consistent with *de novo* origin of the duplications in these affected (black bar) individuals. In other families the presence of the duplications segregated with the phenotype. *indicates different primer pair was used in F7 compared with other families. (C) The specific breakpoints for the duplications defined in affected individuals from families F2, F4, F6 and F7. (D) A cartoon of the duplications at Xq26.1-q26.2 from

the eight families with BDCS, with individual F6:II-1 defining the boundaries of the critical interval. (E) Long-range gap PCR in families' members of F1, F3, F5 and F8 had seemingly identically sized duplications (top). Haplotype analysis with four common single nucleotide polymorphisms (SNPs) (rs62603806, rs4240127, rs5932866, rs12559533) within the duplicated region demonstrates independent origins of F1, F3 and F8 (down).

Fig. 3. BDCS duplications do not affect the topologically associated domain structure (TADs) but could possibly alter enhancer-transcription factor-gene interactions. The first two panels showed Knight-Ruiz Matrix Balancing algorithm (KR) normalized Hi-C maps generated from dermal fibroblasts from two affected individuals with BDCS from families F3 (I-2) and F5 (II-1) and one control. The two maps from affected individuals were merged into one. The maps show that this region contains one large topologically associated domain (TAD) containing *ARHGAP36*, *IGSF1* and *OR13H1* that is not altered in patients. Red lines under each heatmap represented TADs calculated from Hi-C results. The next two panels indicated the relative location of BDCS duplications to TADs and refence genes inside each TAD. Followed two panels showed duplications near or overlapped with minimal duplicated region in DGV and DECIPHER, respectively. Minimal duplicated region were enlarged to visualize the positions of imputed enhancer (yellow) and bivalent promoter (purple) in H1- and H9-cells from chromHMM tracks in UCSC genome browser. The lowest panel showed predicted transcriptional factor binding sites of KLF4 and ESX1 (orange) inside the minimal duplicated region. The position of minimal duplicated region in genome is marked in light blue. Coordinates in hg38.

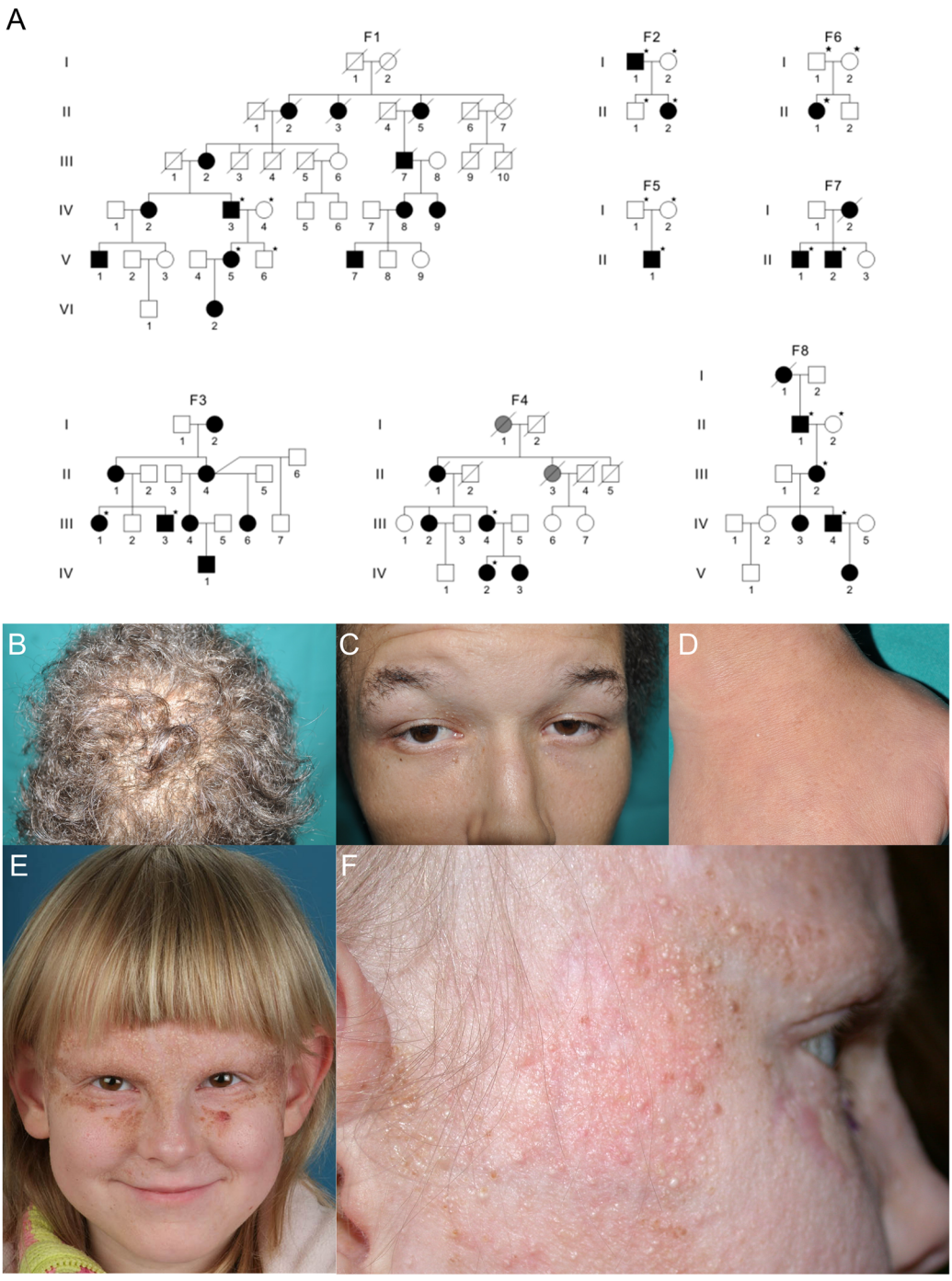
Fig. 4. Immunofluorescence of hair follicles in anagen and telogen from normal healthy skin and an individual (F5:II-1) with BDCS for ARHGAP36. A small number of cells in the stem cell bulge region stain positively (green) for ARHGAP36 in the hair follicle from normal healthy skin (A) and the skin from F5:II-1 (B). (C) A normal hair follicle in telogen stained for ARHGAP36 (green) demonstrating a small number of positively stained cells in the outer root sheath adjacent to K15 positive bulge stem cells (red). (D) A hair follicle in telogen from F5:II-1 showing an increased number of positively stained cells. Red stain for keratin 15. (DP = dermal papilla; ORS = outer root sheath; HS = hair shaft). Scale bar 50μm.

Fig. 5. Immunofluorescence of ARHGAP36 in BDCS patient and sporadic BCCs showed ARHGAP36 could be relevant to BDCS and sporadic BCC.

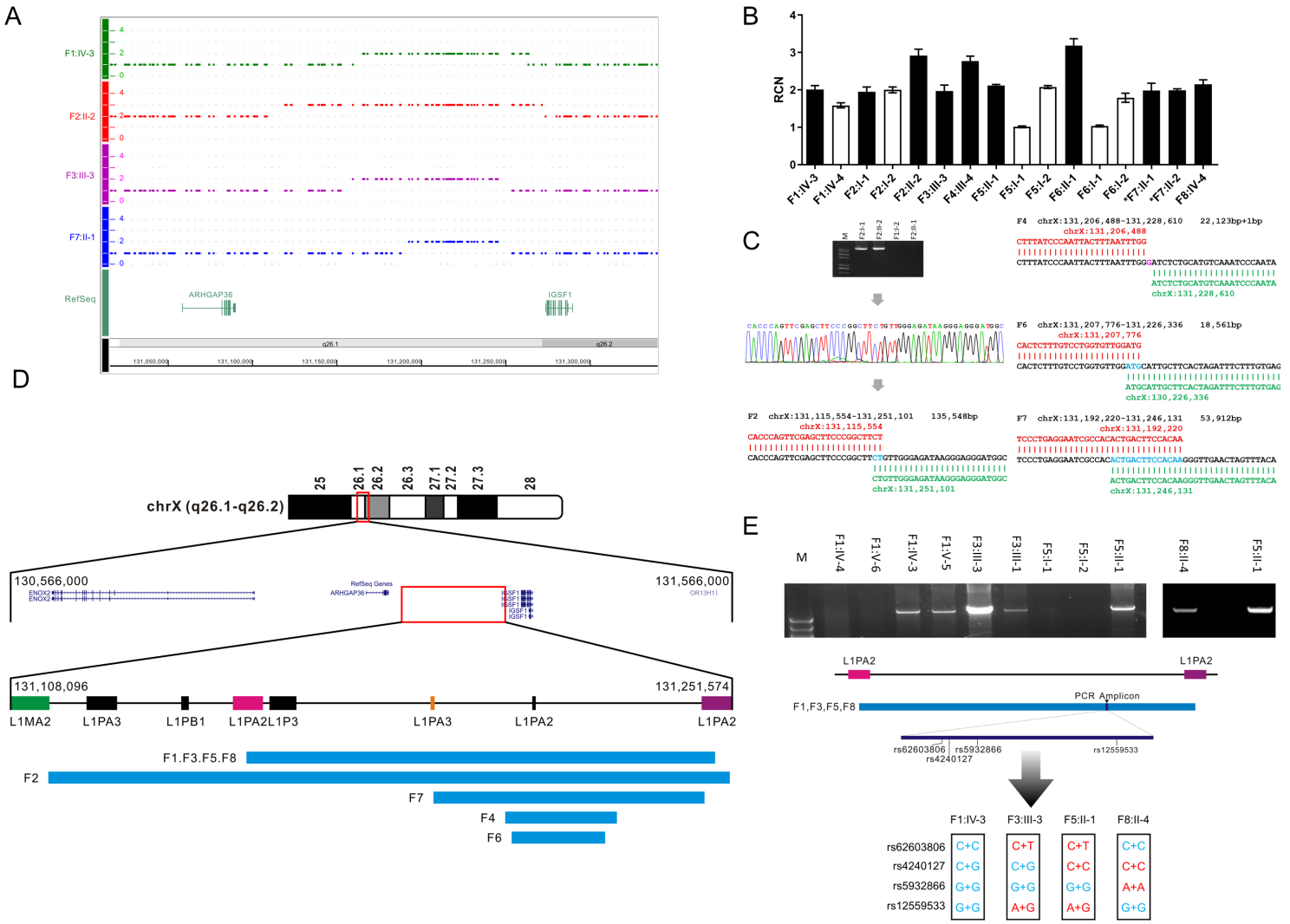
(A-C) Pathological study of BCC from an individual with BDCS (F4:III-4). (A) Haematoxylin and eosin staining of typical BCC and the adjacent normal skin tissue in F4:III-4. (B) Immunofluorescence demonstrating staining for ARHGAP36 (green) in the tumor but not in the surrounding tissue in F4:III-4. (C) Strong staining of a trichoepithelioma for ARHGAP36 and P63 (pink) in the same individual. (D-E) Pathological study of BCC from sporadic BCCs from individuals without BDCS. Different types of BCCs, including superficial (D), nodular (E) and infiltrative (F) BCCs were stained for ARHGAP36 (green) and P63 (pink). Scale bar 50μm.

Fig. 6. Diagram of how the duplications result in BDCS. A) shows healthy genome, where the putative enhancers regulate expression of *ARHGAP36*. B) shows duplications in BDCS patients, which occurs within the same TAD (intra-TAD) and doesn't disrupt structure of TAD. The duplications increased the interactions between

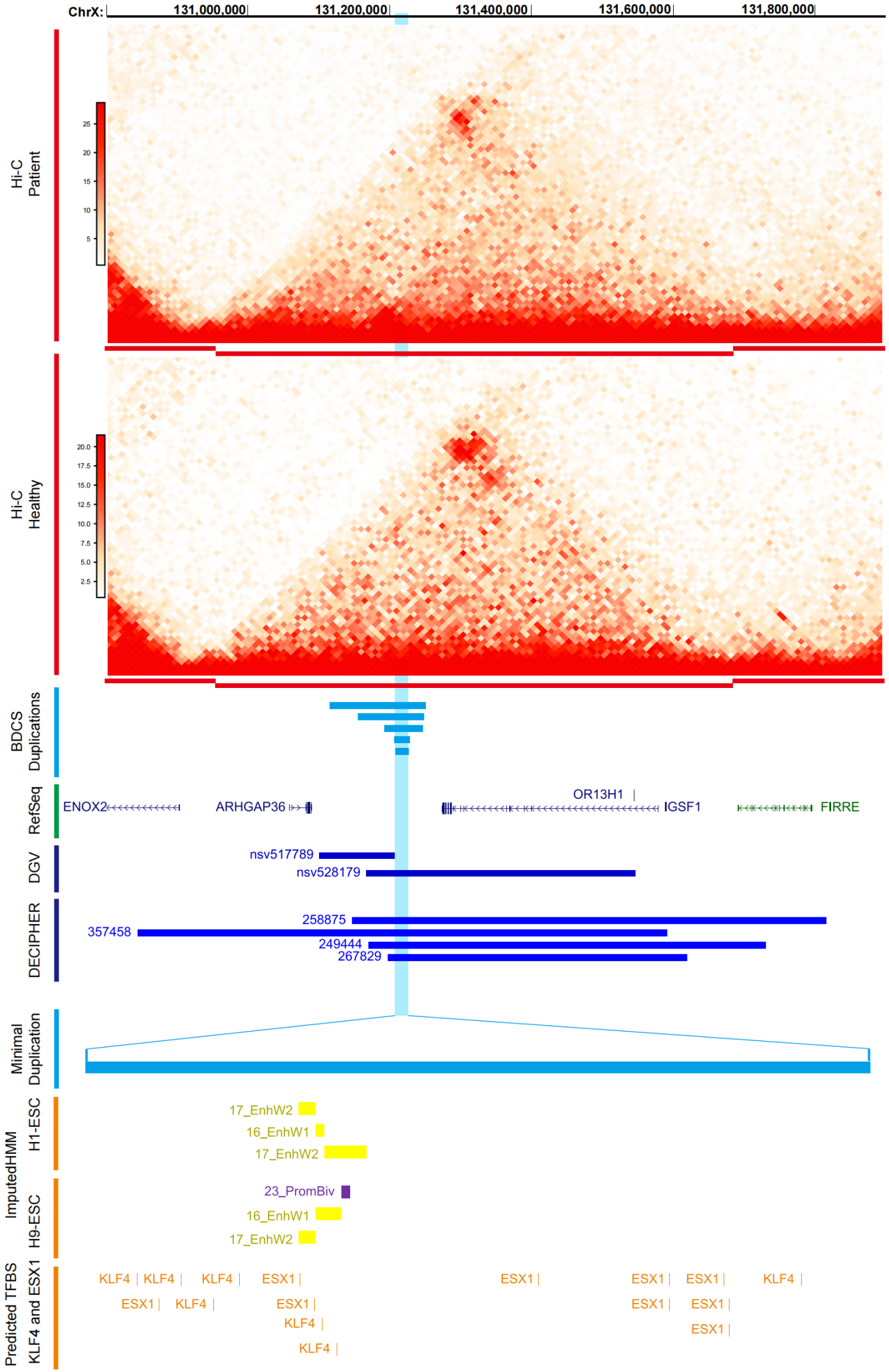
enhancers with *ARHGAP36*, leading to temporal and/or spatial disregulation of *ARHGAP36* and BDCS.



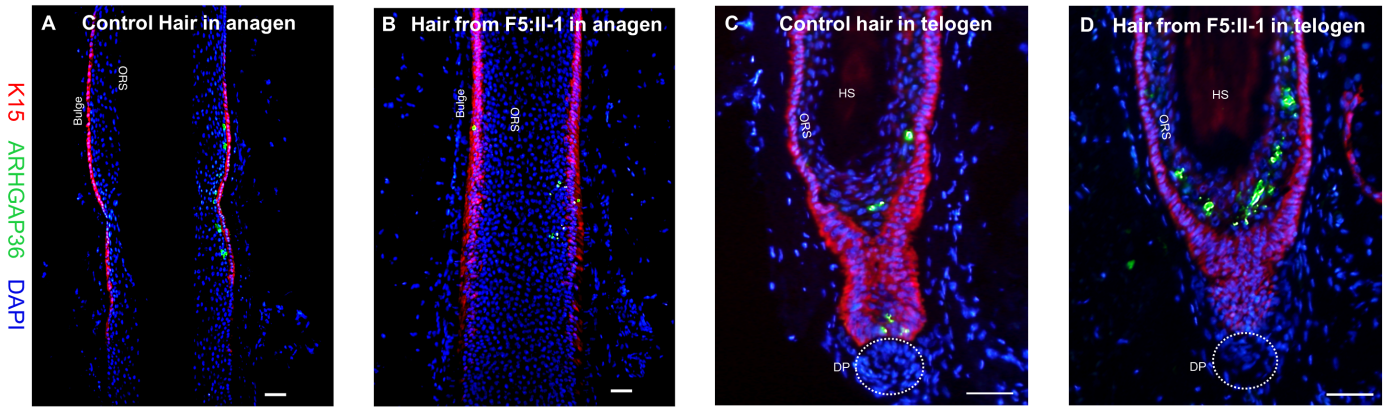
BJD_21842_figure1.tif



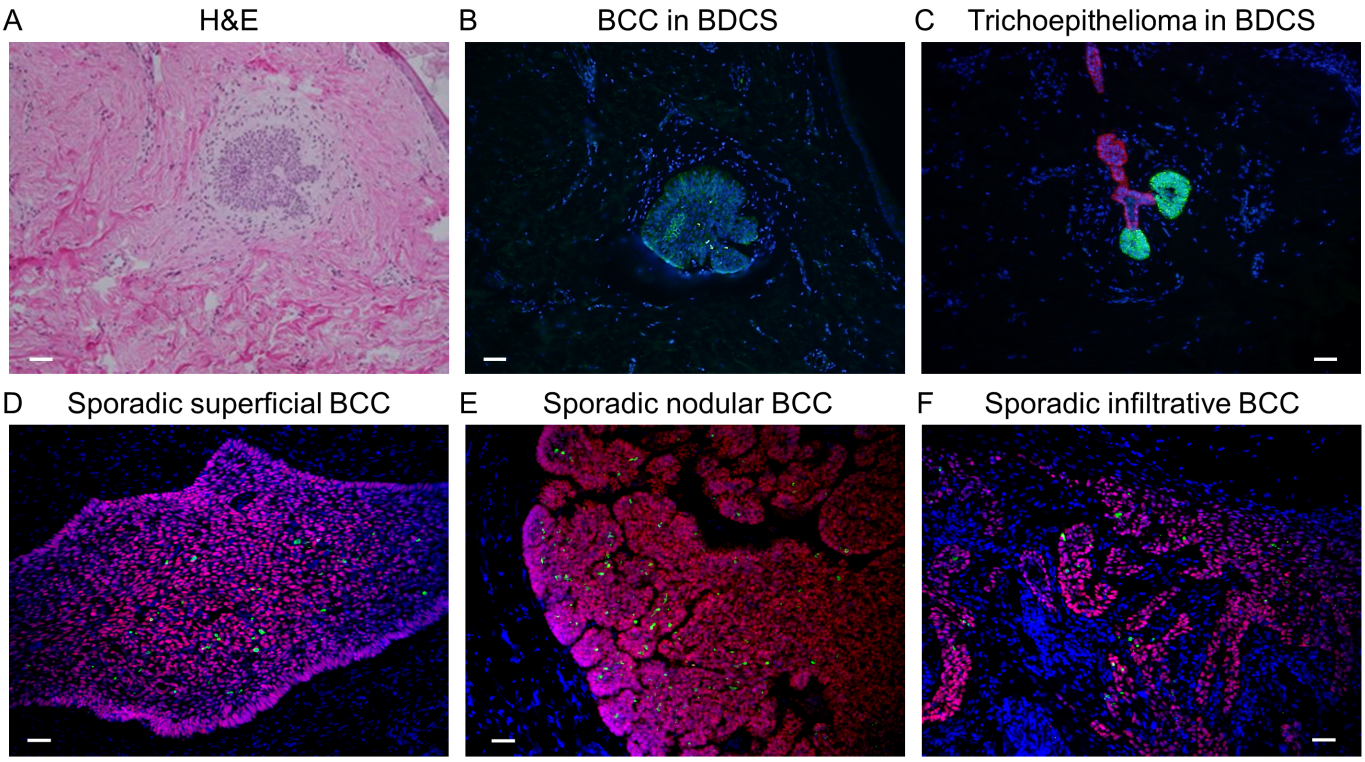
BJD_21842_Figure2.TIF



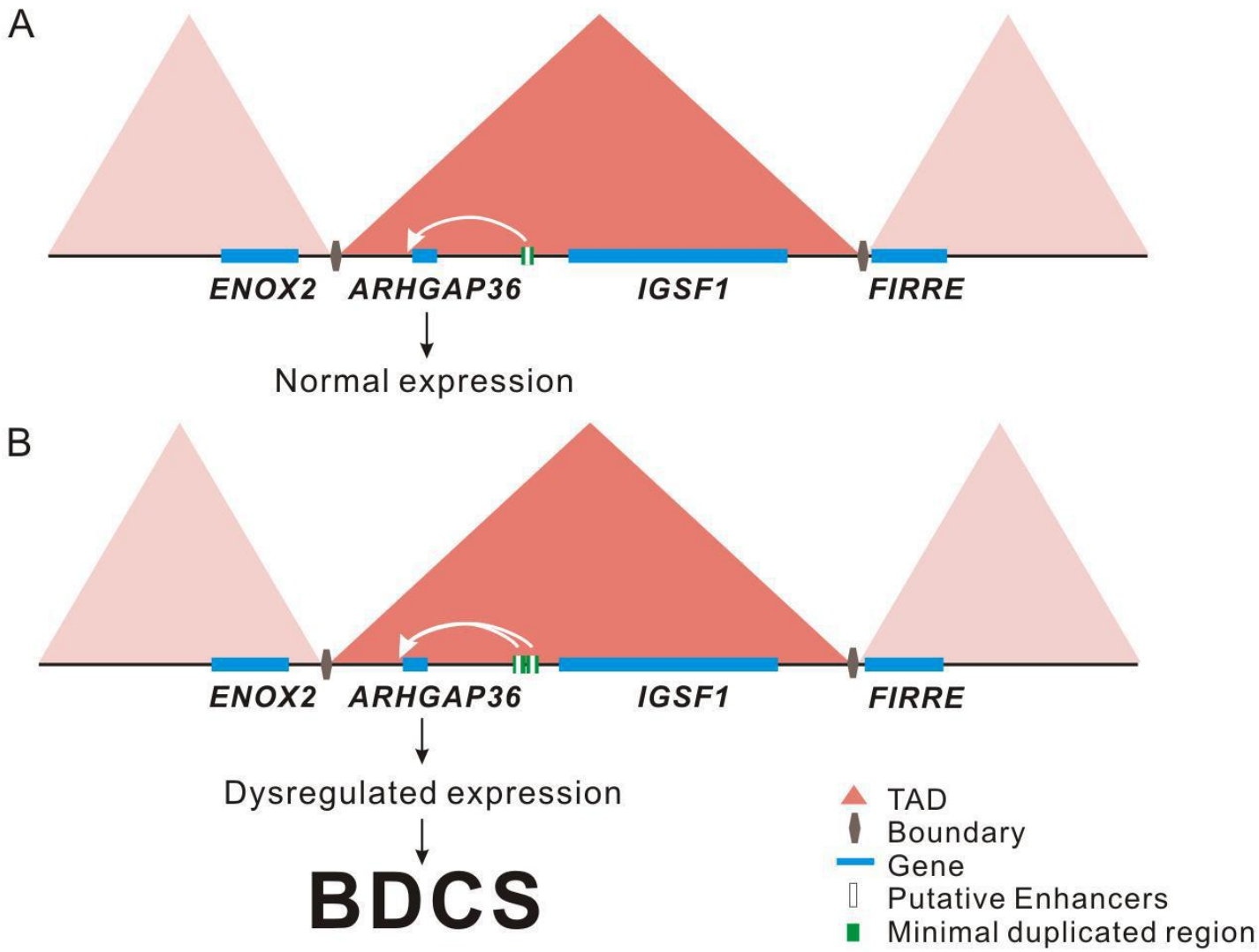
BJD_21842_figure3.tif



BJD_21842_Figure4.TIF



BJD_21842_Figure5.TIF



BJD_21842_figure6.jpg

SUPPLEMENTAL INFORMATION

Germline intergenic duplications at Xq26.1 underlie Bazex-Dupré-Christol basal cell carcinoma susceptibility syndrome

Yanshan Liu, Siddharth Banka , Yingzhi Huang, Jonathan Hardman-Smart, Derek Pye, Antonio Torrelo, Glenda M. Beaman, Marcelo G. Kazanietz, Martin J Baker, Carlo Ferrazzano, Chenfu Shi, Gisela Orozco, Stephen Eyre, Michel van Geel, Anette Bygum, Judith Fischer, Zosia Miedzybrodzka, Faris Abuzahra, Albert Rübben, Sara Cuvertino, Jamie M. Ellingford, Miriam J. Smith, D. Gareth Evans, Lizelotte J.M.T Weppner-Parren, Maurice A.M. van Steensel, Iskander H. Chaudhary, D. Chas Mangham, John T. Lear, Ralf Paus, Jorge Frank, William G. Newman, Xue Zhang.

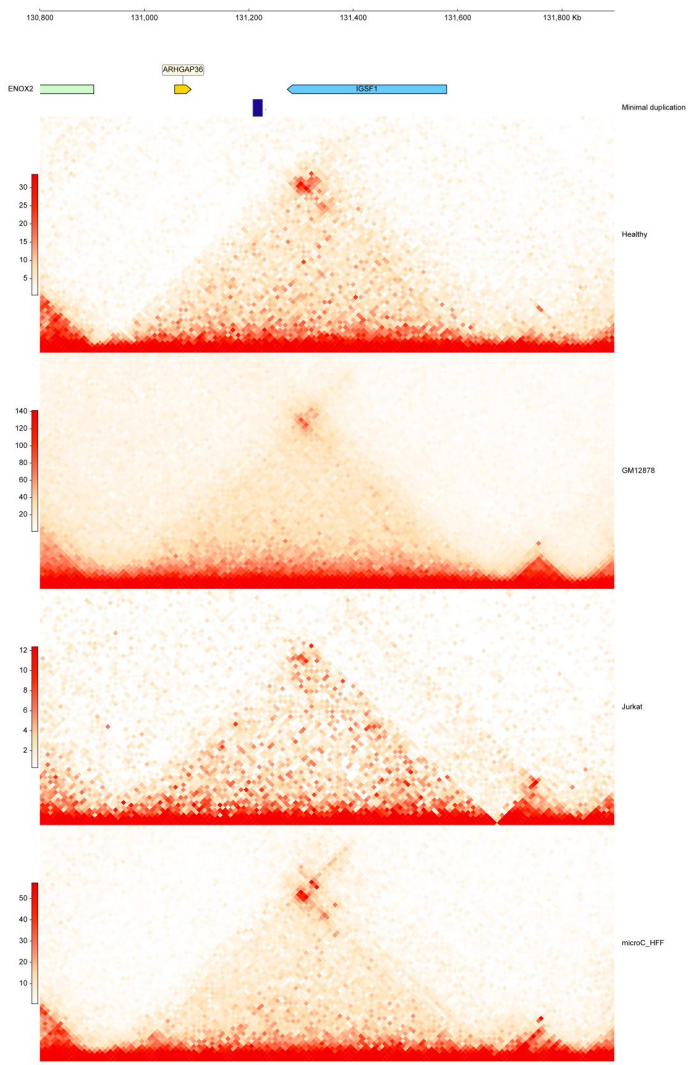


Fig. S1. The TAD is conserved across multiple cell lineages. Hi-C maps and micro-C maps obtained from publicly available databases for GM12878 (B cell lymphoblastoid), Jurkat (lymphocytes) and Human foreskin fibroblasts (HFF). These maps show that the TAD enclosing this region is conserved across multiple cell types. Coordinates in hg38.

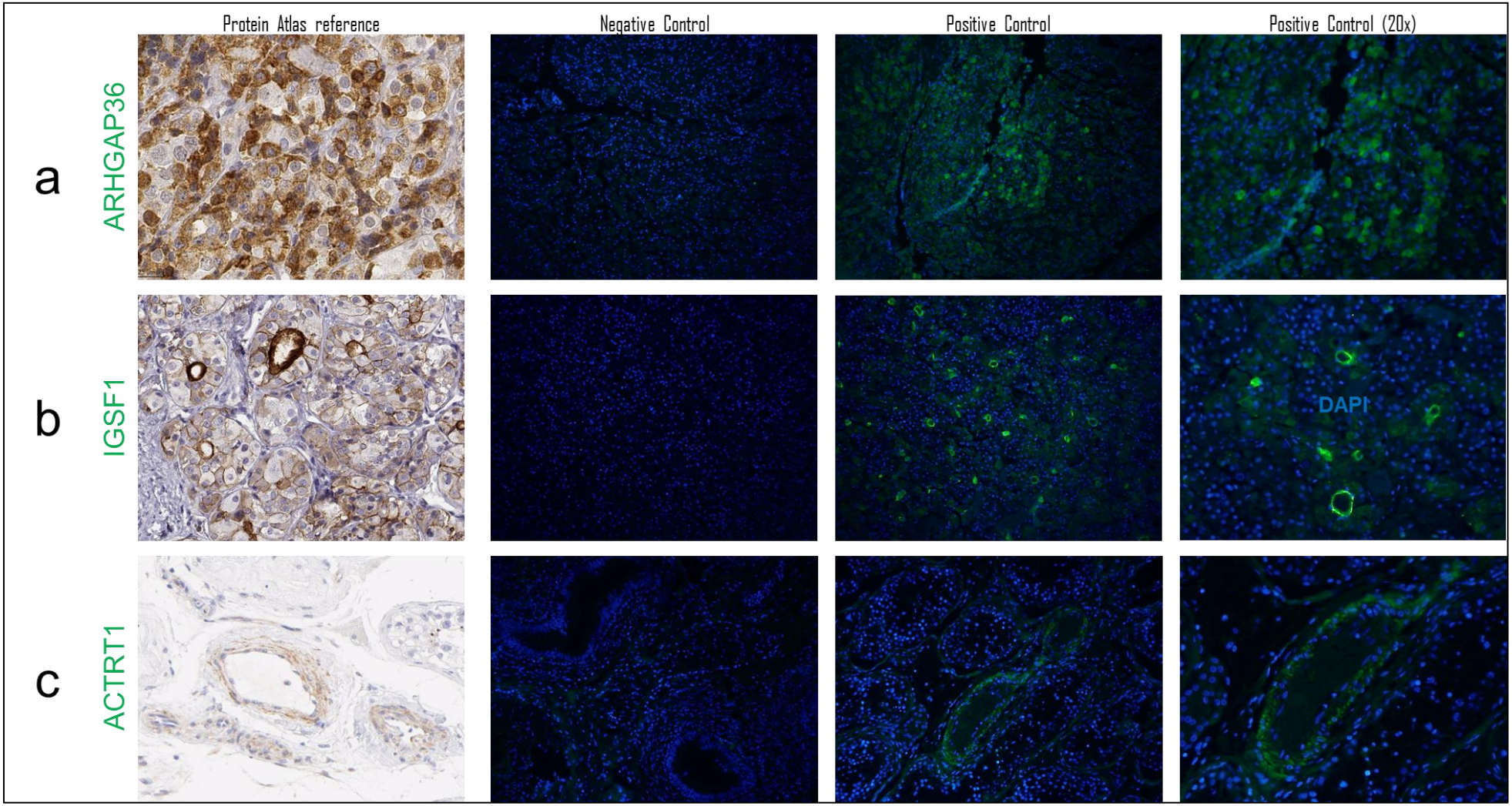


Fig.S2. Immunohistochemistry staining patterns of ARHGAP36, IGSF1 and ACTRT1. Immnuohistochemistry staining figures from the Protein Atlas Reference (https://www.proteinatlas.org) in pituitary gland for a) ARHGAP36 and b) IGSF1 and testes for c) ACTRT1 were used as reference. Brown positive staining for the respective proteins. Negative (without antibody) controls for the three proteins in the respective tissue. Immunofluorescence for the three antibodies show staining patterns comparable to the reference data supporting the specificity of the antibodies. DAPI=blue. Green stain relates to the presence of a) ARHGAP36 b) IGSF1 and c) ACTRT1.

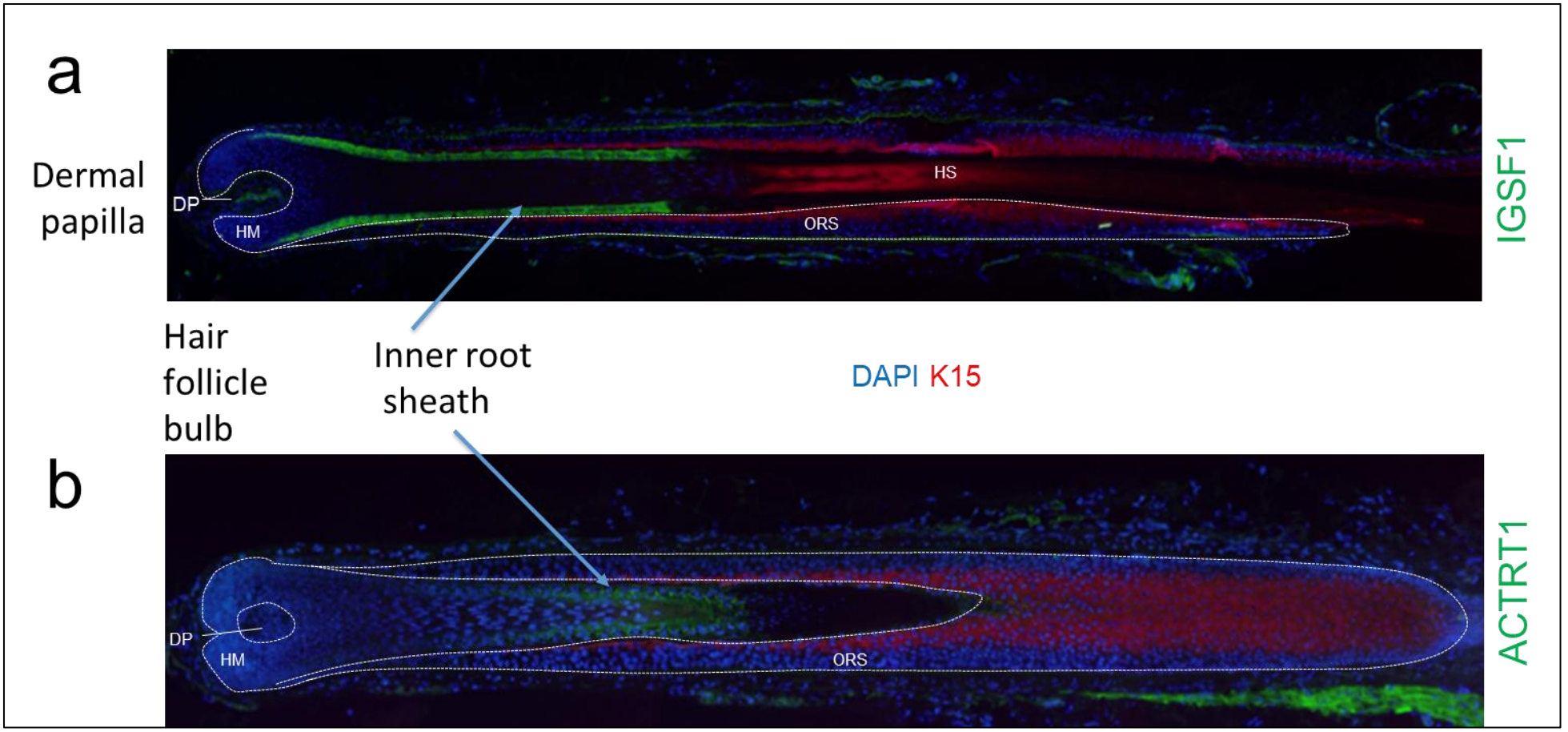


Fig. S3. Expression and localisation of IGSF1 and ACTRT1 in anagen in hair follicle from normal healthy skin. Immunofluorescence of hair follicles in anagen from normal skin for IGSF1 and ACTRT1 (green) with staining of the inner root sheath marked by arrows. Note no staining for either IGSF1 or ACTRT1 at the stem cell bulge region. DAPI =blue and Keratin K15=red. K15 is a marker of epidermal stem cells. (DP = dermal papilla; ORS = outer root sheath; HS = hair shaft)

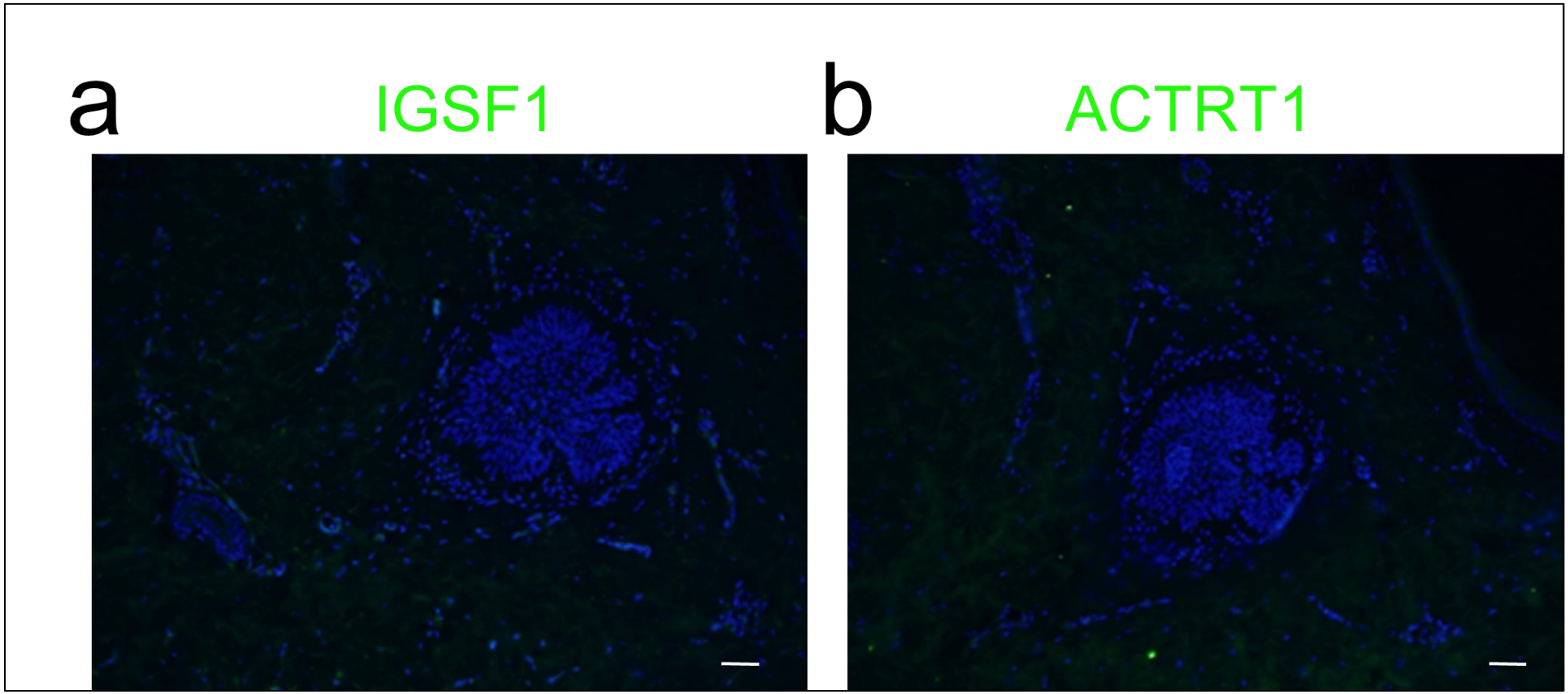


Fig. S4. Immunofluorescence of BCC tumors from an individual with BDCS (F5:II-1). Absence of immunofluorescence for (a) IGSF1 and (b) ACTRT1 respectively in a BCC.

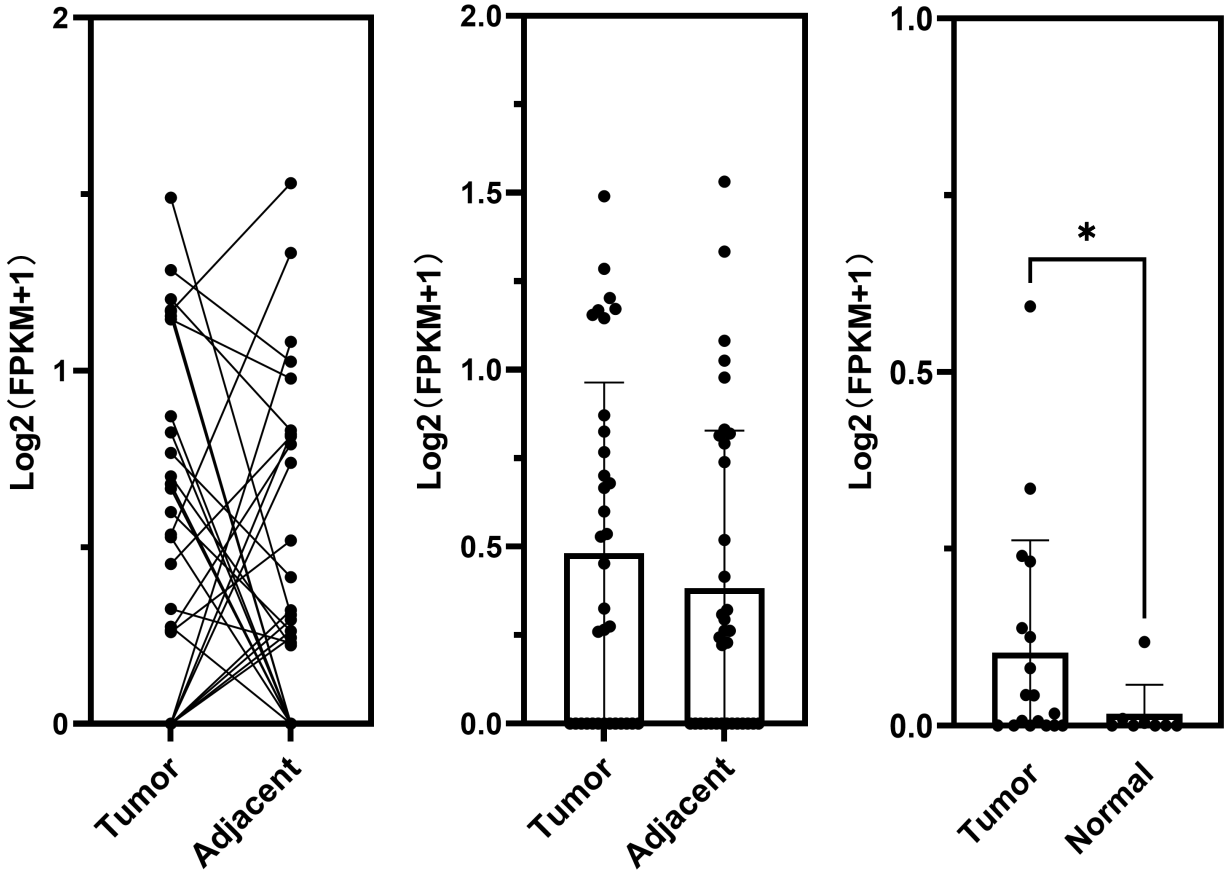


Fig. S5: Expression of *ARHGAP36* in published sporadic BCCs. Paired comparison of dataset GSE125285 showed slightly higher expression of ARGHAP36 in BCC than paired adjacent tissues (left and middle figures). Unpaired comparison between BCCs and normal tissues from GSE58375 and GSE128786 suggested higher expression in BCC tissues ($P<0.05$) (right figure). Each dot represents expression of each sample.

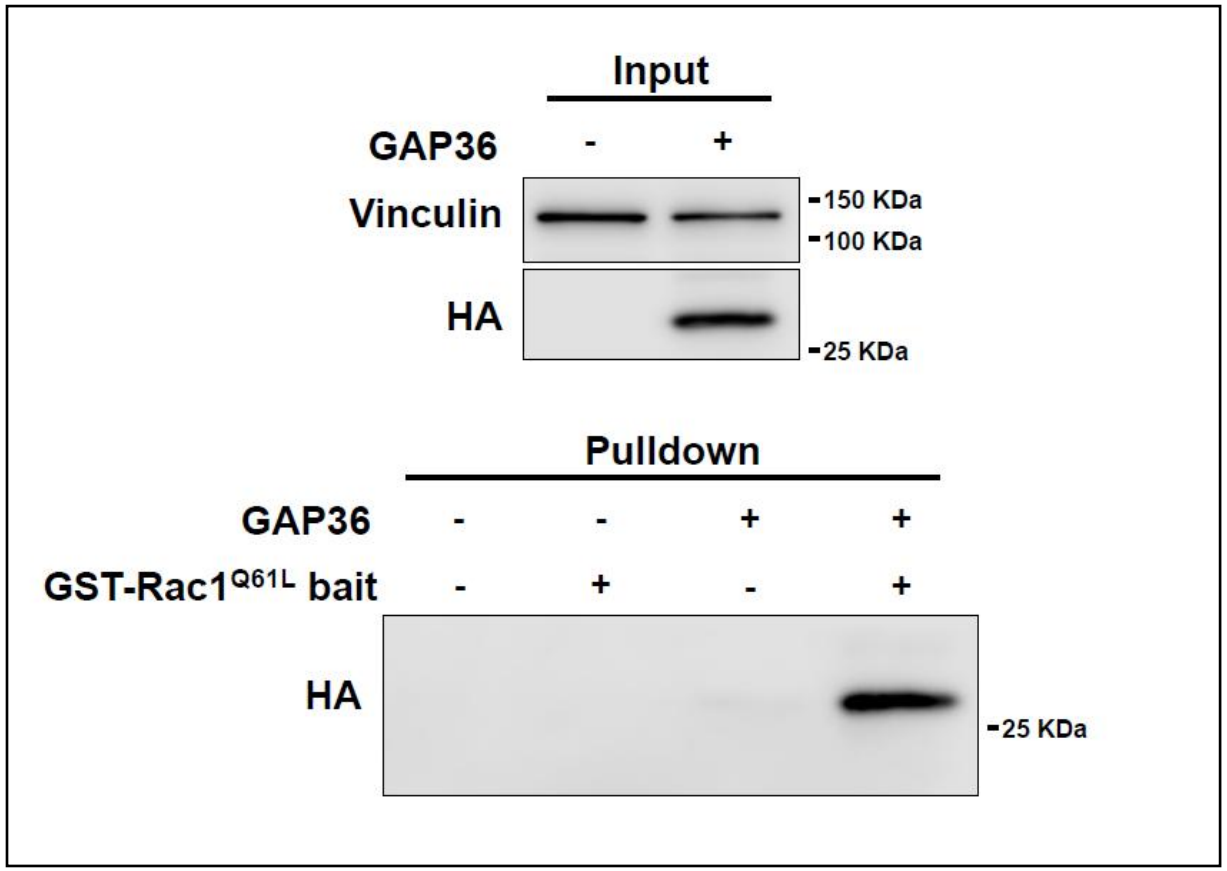


Fig. S6: ARHGAP36 interacts with RAC1.

Pulldown assays using a clone of the GAP domain of ARHGAP36 and a recombinant constitutively active RAC1 mutant. The FLAG-HA tagged GAP domain of Arhgap36 (GAP36) was ectopically expressed in HEK293T cells and precipitated using a recombinant GST tagged active Rac1Q61L mutant. The top panel shows the input of control or FLAG-HA-GAP36 expressing cells and the bottom panel shows the precipitated GAP36, detected with anti-HA.

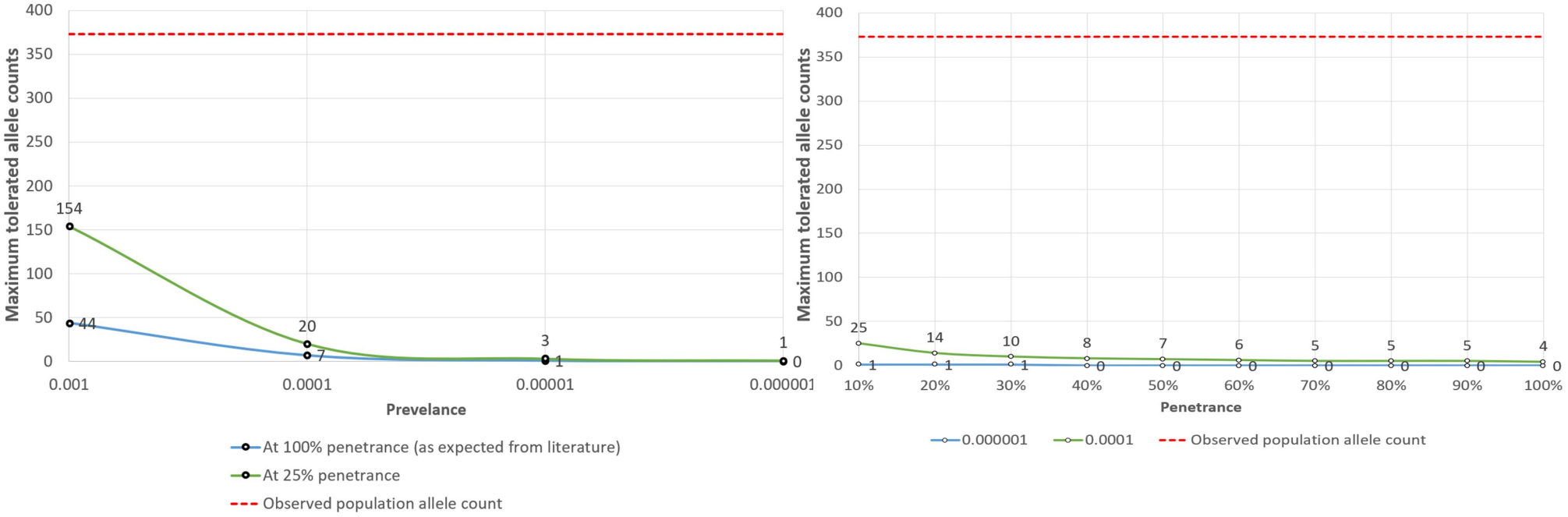


Fig. S7. *ACTRT1* loss-of-function variants are unlikely to cause BDCS. Modelling of maximum tolerated allele counts (MTAC) for the *ACTRT1* NM_138289.3:c.547dup (p.(Met183Asnfs*17)) variant is shown. The left and right panels show MTACs against different levels of possible prevalence and penetrance respectively. In both panels, the blue lines model MTACs at the prevalence (left panel) or penetrance (right panel) levels consistent with the existing literature. The green lines model MTACs at much lower hypothetical constraints. Allele count estimates are for 204,684 alleles in the reference population at 0.95 confidence interval, 33% allelic heterogeneity and genetic heterogeneity of 1. Note that the observed population allele count (=373) of the variant, shown in broken red line, is significantly higher than estimated maximum tolerated allele counts across all the modelled scenarios.

SUPPLEMENTARY TABLES

Table S1. Copy number gain variants encompassing shared BDCS duplication region in DECIPHER and associated clinical phenotypes.

ID	**Location (hg38)**	**Size**	**Inheritance**	**Genotype**	**Disruption of TAD**	**Pathogenicity**	**Phenotypes**
258875	chrX:131146354-131816185	669.9Kb	Inherited from normal parent	Hemizygous	Yes	Unknown	Blepharophimosis, Cleft palate, Depressed nasal tip, Epicanthus, Intellectual disability
249444	chrX:131169899-131730235	560.4Kb	NA	Unknown	Yes	Unknown	Unknown
267829	chrX:131196585-131619679	423.1Kb	Inherited from normal parent	Hemizygous	No	Unknown	Inguinal hernia, Intellectual disability, Microcephaly, Plagiocephaly
357458	ChrX:130843799-131591104	747.31kb	NA	Heterozygous	Yes	Uncertain	Abnormality of the musculoskeletal system

Table S2. Putative loss-of-function variants of *ACTRT1* in public databases.

position	dnSNP	Major Allele	Minor Allele	Frequency#	Consequence	HGVS
ChrX:128051143	rs1569354431	T	-	0.000005 (1/183050)	Frameshift	NM_138289.4:c.1064del, NP_612146.1:p.Gln355fs
ChrX:128051093	rs867661986	G	A	NA	Stop_gained	NM_138289.4:c.1114C>T, NP_612146.1:p.Gln372Ter
ChrX:128051202	rs768201050	-	ATCA	0.000016 (3/183193)	Stop_gained	NM_138289.4:c.1002_1005dup, NP_612146.1:p.Arg336Ter
ChrX:128051222	rs766490344	T	A	0.000005 (1/183246)	Stop_gained	NM_138289.4:c.985A>T, NP_612146.1:p.Lys329Ter
ChrX:128051268	rs773801438	C	-	0.000005 (1/183215)	Frameshift	NM_138289.4:c.939del, NP_612146.1:p.Arg313fs
ChrX:128051350	rs1223259859	-	CTTCA	0.000008 (1/125568)^	Stop_gained	NM_138289.4:c.853_857dup, NP_612146.1:p.Cys286Ter
ChrX:128051483	rs747807748	TGT	T	0.000005 (1/183135)	Stop_gained	NM_138289.4:c.723_724del, NP_612146.1:p.Tyr241Ter
ChrX:128051511	rs1174530093	C	-	0.000005 (1/183221)	Frameshift	NM_138289.4:c.696del, NP_612146.1:p.Ser233fs
ChrX:128051585	rs771674435	GAGT	-	0.000119 (15/125568)^	Frameshift	NM_138289.4:c.619_622del, NP_612146.1:p.Leu207fs
ChrX:128051592	rs1397178516	G	T	0.000005 (1/183275)	Stop_gained	NM_138289.4:c.615C>A, NP_612146.1:p.Cys205Ter
ChrX:128051660	rs771087307	-	T	0.001822(373/204684)	Frameshift	NM_138289.4:c.547dup, NP_612146.1:p.Met183fs
ChrX:128051663	rs760509819	AG	-	0.000024 (3/125568)^	Frameshift	NM_138289.4:c.541_542CT, NP_612146.1:p.Cys182fs
ChrX:128051852	rs767426435	G	A	0.000011 (2/183183)	Stop_gained	NM_138289.4:c.355C>T, NP_612146.1:p.Arg119Ter
ChrX:128051861	rs776692535	T	-	0.00004 (1/183206)	Frameshift	NM_138289.4:c.346del, NP_612146.1:p.Arg116fs
ChrX:128051880	rs1569354678	G	-	0.000005 (1/183222)	Frameshift	NM_138289.4:c.327del, NP_612146.1:p.Glu110fs
ChrX:128051885	rs759455373	-	T	0.000005 (1/183243)	Frameshift	NM_138289.4:c.322dup, NP_612146.1:p.Met108fs
ChrX:128051898	rs764199744	TT	-	0.000005 (1/183229)	Frameshift	NM_138289.4:c.308_309del, NP_612146.1:p.Gln103fs
ChrX:128051904	rs1556035539	-	G	0.00108 (11/10195)*	Frameshift	NM_138289.4:c.303dup, NP_612146.1:p.Ser102fs
ChrX:128051928	rs761972772	AA	-	0.000027 (5/183180)	Stop_gained	NM_138289.4:c.278_279del, NP_612146.1:p.Leu92_Phe93insTer
ChrX:128051930	rs767686485	GA	-	0.000005 (1/183194)	Frameshift	NM_138289.4:c.274_275CT[1], NP_612146.1:p.Leu92_Phe93insTer
ChrX:128051932	rs750649097	A	-	0.000005 (1/183181)	Frameshift	NM_138289.4:c.275del, NP_612146.1:p.Leu92fs
ChrX:128051939	rs759293030	T	A	0.000005 (1/183169)	Stop_gained	NM_138289.4:c.268A>T, NP_612146.1:p.Lys90Ter
ChrX:128052163	rs754366510	T	-	0.000011 (2/176668)	Frameshift	NM_138289.4:c.44del, NP_612146.1:p.Asp15fs

Frequency data was from GnomAD_exome if not marked. *: GO-ESP, NHLBI Grand Opportunity Exome Sequencing Project (ESP). ^: TOPMED, NHLBI Trans-Omics for Precision Medicine WGS.

Table S3. Microdeletions encompassing *ACTRT1* in DECIPHER and associated clinical phenotypes.

Patient ID	Location	Genes	Size	Inheritance / Genotype	Pathogenicity	Phenotypes
407109	ChrX:128050021-128224023	*ACTRT1*	174Kb	Maternally inherited/Heterozygous	Uncertain	Not available
254266	ChrX:127824452-128235789	*ACTRT1*	411.34Kb	Unknown/Heterozygous	Unspecified	Atrial septal defect, Abnormality of the outer ear, Blepharophimosis, High palate, Thin upper lip vermilion, Wide nasal bridge, Recurrent urinary tract infections, Delayed speech and language development, Dysarthria, Intellectual disability, Short attention span, Clinodactyly of the 5th finger, Joint laxity, Micrognathia, Nasal speech
260788	ChrX:127645238-128307787	*ACTRT1*	662.55Kb	Inherited from normal patient/Hemizygous	Unspecified	Abnormality of cardiovascular system morphology
279086	ChrX:127645238-128307787	*ACTRT1*	662.55Kb	Unknown/Hemizygous	Unspecified	Proptosis, Neonatal hypotonia, Delayed gross motor development
2214	ChrX:126983897-128224455	*ACTRT1*	1.24Mb	Inherited from normal patient/Heterozygous	Unspecified	Deeply set eye, Hypertelorism, Epicanthus, Abnormality of the palmar creases, Delayed speech and language development, Clinodactyly of the 5th finger
1641	ChrX:125305481-128665958	7 genes	3.36Mb	Inherited from normal patient/Hemizygous	Unspecified	Truncal obesity, Abnormal eyelid morphology, Macrodontia, Thick lower lip vermilion, Thick upper lip vermilion, Intellectual disability, Tapered finger

Table S4. Primers used for qPCR and long-range gap-PCR in BDCS families.

Family	name	sequence	location
F1	XLB4F	CAGACACCAGCTGTGCAATGTA	chrX:131146130-131146278
	XLB4R	TGGTCCTTGGATCCCATAGC	
	F1LB2-1F	CATTGAGCCATGCACCTTGT	chrX:131151244-131151369
	F1LB2-1R	CAATTGGTGCGCAAGACACT	
	F1LB2-2F	TCCATGCCAATAGAATGCTGTT	chrX:131158606-131158720
	F1LB2-2R	AATGCTCACCCAAATCACAGAA	
	F1LB2-3F	ATGGCTCTGACAGGGAATGACT	chrX:131164971-131165081
	F1LB2-3R	GCAGGACCACAATGTTTTGGA	
	XQMF	TGCCTCCAGTAGTCCATTGACA	chrX:131214207-131214336
	XQMR	CCAAACCTCATCCAGAGTGCTT	
	RB3-1F	TCAGACATTTTCATTTGCCAGTCTT	chrX:131244139-131244253
	RB3-1R	TGACCAGCAAACAGGAAATGC	
	RB3-3F	TGGAAAAATCAAAGAGTGAAAATCCT	chrX:131252219-131252311
	RB3-3R	TGCGCTTCGTTGTTGATTACA	
F3, F5, F8	XQMF	TGCCTCCAGTAGTCCATTGACA	chrX:131214207-131214336
	XQMR	CCAAACCTCATCCAGAGTGCTT	
F2	F2LB2-1F	TCCGAGATTCTTTTGACCCACTA	chrX:131099907-131100014
	F2LB2-1R	CAGCAGAGCTGCAATATTACTGAAA	
	F2LB2-2F	TTTCTAAAGCGTTGGCACCAT	chrX:131118694-131118839
	F2LB2-2R	TCAGACCTGCCAGTACGGTTT	
	F2LB2-3F	CACTGAATTCCTTTGGTGACTTTG	chrX:131119457-131119567
	F2LB2-3R	GCAACATTGGTTTGATGAAAGAGA	
	XQMF	TGCCTCCAGTAGTCCATTGACA	chrX:131214207-131214336
	XQMR	CCAAACCTCATCCAGAGTGCTT	
	RB3-1F	TCAGACATTTTCATTTGCCAGTCTT	chrX:131244139-131244253
	RB3-1R	TGACCAGCAAACAGGAAATGC	
	RB3-3F	TGGAAAAATCAAAGAGTGAAAATCCT	chrX:131252219-131252311
	RB3-3R	TGCGCTTCGTTGTTGATTACA	
F4	F4B1-9F	CAGGCTGCTATGGTATCAAATTTG	chrX:131204570-131204668
	F4B1-9R	ATCTGAGCAGGAGAGCTTATGCA	
	F4B1-10F	ACTCCTGGGATAGGGCCAGAT	chrX:131208157-131208246
	F4B1-10R	CTTGACAGCCACTGCCATATTG	
	XQ-F1	TGGTGTTAATTGGTCGCATCTG	chrX:131213229-131213318
	XQ-R1	AATAACTCGCCCTATGCCCATCT	
	XQ-F2	TCTCACCATTAGGGATGGACTGT	chrX:131218349-131218438
	XQ-R2	CCTGCCGATCATCTCACAAA	

	XQ-F3	GTGGTGCCCAAAGACCCATA	chrX:131222340-131222429
	XQ-R3	GGTCACAGGGCAGACTGATCTC	
	F4RF1	TGGCCCCAACGTTCCTAGA	chrX:131223274-131223380
	F4RR1	TCCCTCATAGCATTGATCACAATAG	
	F4B1-12F	TTCTATGCCTGCTCCATTCC	chrX:131226175-131226271
	F4B1-12R	GGGCTAGCACCATGAGGGTAT	
	XQ-F4	TTATCACAGGTTCCAGATGGAGAA	chrX:131229882-131229972
	XQ-R4	TGATGAAAACCACCCTCCTATCTAC	
F6	F4B1-9F	CAGGCTGCTATGGTATCAAATTTG	chrX:131204570-131204668
	F4B1-9R	ATCTGAGCAGGAGAGCTTATGCA	
	F4B1-10F	ACTCCTGGGATAGGGCCAGAT	chrX:131208157-131208246
	F4B1-10R	CTTGACAGCCACTGCCATATTG	
	F4B1-12F	TTCTATGCCTGCTCCATTCC	chrX:131226175-131226271
	F4B1-12R	GGGCTAGCACCATGAGGGTAT	
	F4B1-13F	TGATGCTTCCCCCACAAAAGA	ChrX:131228910-131229002
	F4B1-13R	TCATTCTGTGATCCCCCACAT	
F7	F4B1-6F	TTGATTCCAAGTTGGTTACCCCATA	chrX:131191389-131191478
	F4B1-6R	TGGCTTTGAAGCATGGGAAA	
	F7B1-1F	AGGTATTGTCATTGGCCCACAT	chrX:131193586-131193678
	F7B1-1R	GCATGAAGTTGCCTTTAAGATCAG	
	F4B1-7F	CAGGGCCATAGGTGGATCAT	chrX:131195926-131196015
	F4B1-7R	CAAATAACATTCAGCCCAGTTTCC	
	F4B1-8F	CCCTATAGCAAGCCCTGTTCTAAA	chrX:131199022-131199113
	F4B1-8R	GGCCTCCAGATACAATCCATAATTA	
	F4B1-9F	CAGGCTGCTATGGTATCAAATTTG	chrX:131204570-131204668
	F4B1-9R	ATCTGAGCAGGAGAGCTTATGCA	
	RB3-1F	TCAGACATTTTCATTTGCCAGTCTT	chrX:131244139-131244253
	RB3-1R	TGACCAGCAAACAGGAAATGC	
	RB3-3F	TGGAAAAATCAAAGAGTGAAAATCCT	chrX:131252219-131252311
	RB3-3R	TGCGCTTCGTTGTTGATTACA	
For Long-range gap-PCR			
Family	**name**	**sequence**	
F1，F3，F5 and F8	RB3-1F	TCAGACATTTTCATTTGCCAGTCTT	
	F1LB2-2R	AATGCTCACCCAAATCACAGAA	
F2	RB3-1F	TCAGACATTTTCATTTGCCAGTCTT	
	F2LB3-4R	TGCCTGAGATAGCCATAATAGAAATC	
F4	F4B1-12F	TTCTATGCCTGCTCCATTCC	
	F4B1-10R	CTTGACAGCCACTGCCATATTG	
F6	F4B1-10R	CTTGACAGCCACTGCCATATTG	
	F4B1-12F	TTCTATGCCTGCTCCATTCC	
F7	RB3-1F	TCAGACATTTTCATTTGCCAGTCTT	

	F7B1-1R	GCATGAAGTTGCCTTTAAGATCAG	
Primers located outside the duplication region are colored in red.			