Impaired chromosome integrity caused by mutations in members of the BTRR complex

Dissertation
for the award of the degree
"Doctor rerum naturalium"
of the Georg-August-Universität Göttingen

within the doctoral program Molecular Medicine
of the Georg-August University School of Science (GAUSS)

submitted by
İpek Ilgın Gönenç

from Ankara, Turkey

Göttingen 2022

Thesis Committee

Prof. Dr. Bernd Wollnik

Institute of Human Genetics, University Medical Center Göttingen

Prof. Dr. Matthias Dobbelstein

Institute of Molecular Oncology, Center of Molecular Biosciences, University of Göttingen

Dr. Lukas Cyganek

Stem Cell Unit, Clinic for Cardiology and Pneumology, University Medical Center Göttingen

Members of the examination board

Prof. Dr. Bernd Wollnik (Referee)

Institute of Human Genetics, University Medical Center Göttingen

Prof. Dr. Matthias Dobbelstein (Second referee)

Institute of Molecular Oncology, Center of Molecular Biosciences, University of Göttingen

Dr. Lukas Cyganek

Stem Cell Unit, Clinic for Cardiology and Pneumology, University Medical Center Göttingen

Prof. Dr. Uwe Kornak

Institute of Human Genetics, University Medical Center Göttingen

Prof. Dr. Holger Bastians

Institute of Molecular Oncology, University Medical Center Göttingen

Prof. Dr. Wolfram-Hubertus Zimmermann

Institute of Pharmacology and Toxicology, University Medical Center Göttingen

Date of oral examination: 17 November 2022

Summary

Bloom syndrome (BS) is an autosomal recessive rare disorder clinically characterized by primary microcephaly, growth deficiency, short stature, photosensitivity, immunodeficiency, and cancer predisposition. BS is caused by biallelic, compound heterozygous, or homozygous loss-of-function (LoF) mutations in *BLM*, which encodes the BLM RecQ-like helicase. BLM has important roles during DNA replication and repair processes. Recently, additional genes associated with BS-like phenotypes were identified by whole exome sequencing strategies, namely autosomal recessively inherited LoF mutations in *TOP3A, RMI1*, and *RMI2*. *TOP3A* encodes the DNA topoisomerase III alpha, which is able to decatenate single-stranded DNA molecules, while *RMI1* and *RMI2* code for the DNA interacting proteins RecQ-mediated genome instability protein-1 and -2.

In general, BLM, TOP3A, RMI1, and RMI2 form together the BTRR multiprotein complex. The BTRR complex participates in fundamental cellular processes for DNA replication and repair. The most well-studied role of the BTRR complex is to dissolve DNA intermediates during the homologous recombination DNA repair, namely the double Holliday junctions. Failure in proper dissolution of such structures can lead to crossovers between the sister chromatids, an event called the sister chromatid exchange (SCE). Increased SCE rates are prominent cellular characteristics of BS cells. In sum, the BTRR complex is important for the protection of genome stability based on its roles in DNA replication and repair processes.

To broaden the clinical and mutational signatures of BS, I aimed to collect and characterize patients with BS-like phenotypes. For this reason, a BS flyer informing about phenotypic characteristics and mechanisms of BS, as well as the details of this project was prepared in German and Turkish languages. This flyer was printed and distributed to clinicians, human genetics institutes, and genetics centers in both countries. Prior to this doctoral thesis, eight patients with BS phenotype were diagnosed by molecular genetic approaches. During my project, I assessed the clinical and mutational findings of patients and compared the data between patients. Homozygous LoF mutations were found in *BLM* in six patients and one homozygous LoF variant was detected in *RMI1* in two patients from a consanguineous family. One of the pathogenic *BLM* variants was a novel mutation. Further, the phenotypic characteristics were milder in *RMI1*-associated

patients when compared to *BLM*-associated patients in terms of skin findings and immunodeficiency in BS.

Next, I investigated the transcriptional changes in BLM-deficient cells. Three BS patient-derived fibroblast cell lines were subjected to single-cell transcriptome sequencing (scRNAseq) using ICELL8 Single-Cell System and a high data quality was achieved. Differentially expressed genes and pathway analyses in BS samples in comparison to wild-type (WT) fibroblasts revealed highly significant terms in relation to the BS pathogenesis, such as chromosome segregation or mitosis-related terms. Many known interaction partners of BLM from the Fanconi anemia pathway were significantly upregulated. Analysis of cell cycle stages, which was possible to perform with the single-cell data, revealed that there was no difference in cell cycle stages between samples. Furthermore, several genes associated with primary microcephaly were deregulated in BS single cells, among which were *NCAPG2, NCAPH*, and *NCAPD2*. These transcripts encode members of the condensin I/II complexes, hence gene expression levels of condensin I/II complex were analyzed in detail and found highly upregulated. The overexpression of condensin I/II complex genes was a novel link in BLM deficiency. In addition, replication stress revealed a mild sensitivity of BS cells in comparison to WT according to the transcriptional changes.

I also generated isogenic BTRR complex-deficient induced pluripotent stem cell (iPSC) lines via CRISPR/Cas9 genome editing technology. Cell lines having biallelic truncating mutations in genes encoding members of the BTRR complex, namely *BLM, TOP3A*, and *RMI1*, maintained the pluripotency after genome editing experiments. The lack of full-length proteins for the corresponding gene was confirmed via Western blot for every knockout (KO) cell line. The generated cell lines were characterized in terms of cellular phenotypes of the BTRR complex deficiency. SCE rates of BLM-KO, TOP3A-KO, and RMI1-KO iPSCs, were significantly higher than the wild-type parental control. Interestingly, the SCE frequency was lower in the RMI1-KO clone than BLM- and TOP3A-KO clones, although it was still significantly higher than the wild type. Mitotic errors such as chromatin bridges and lagging chromatin were observed in the KO iPSC samples, only RMI1-KO did not show a significant difference. Ultrafine anaphase bridges (UFB) were quantified in the generated KO samples and all three BTRR complex-deficient cell lines showed significantly increased rates of UFBs, implying that the BTRR complex deficiency resulted in unresolved UFBs independent of which member of the BTRR complex was impaired. Next, transcriptional changes of BTRR complex-deficient iPSCs were determined by scRNAseq. The overall transcriptional profiles were similar in KO-iPSCs and the parental wild type, yet the TOP3A-KO clone showed a possible sensitivity to replication stress.

In summary, by the use of several different cell models for BS, the data generated in this project is of interest for further characterization of the BTRR complex. Transcriptional

changes in BS can further shed light on the pathogenesis of BS. The isogenic iPSC lines provide a source to study the roles and effects of each member of the complex while providing therapeutic screening opportunities. Future differentiation approaches will provide additional insights into disease-associated mechanisms of BS phenotypes such as microcephaly or cancer predisposition in somatic cells.

Contents

Summary i

1 Introduction 1
- 1.1 Rare diseases . . . 1
- 1.2 NGS-based gene identification strategies . . . 2
- 1.3 Single-cell transcriptome sequencing technologies . . . 3
- 1.4 Genomic instability syndromes . . . 5
- 1.5 Primary microcephaly syndromes . . . 8
- 1.6 Bloom syndrome . . . 10
  - 1.6.1 Phenotypic characteristics of BS . . . 10
  - 1.6.2 Genetic causes of BS phenotype . . . 13
- 1.7 The BTRR complex . . . 15
  - 1.7.1 BLM RecQ-like helicase . . . 16
  - 1.7.2 Topoisomerase 3A . . . 19
  - 1.7.3 RecQ-mediated genome instability proteins 1 and 2 . . . 21
- 1.8 The functional roles of the BTRR complex . . . 23
- 1.9 Cellular pathomechanisms in BS . . . 24

2 Materials and Methods 28
- 2.1 Materials . . . 28
  - 2.1.1 Antibodies . . . 28
  - 2.1.2 Cell lines . . . 29
  - 2.1.3 Reference sequences . . . 30
  - 2.1.4 Synthetic oligonucleotides . . . 30
  - 2.1.5 Solutions and buffers . . . 32
  - 2.1.6 Chemicals and expendable equipments . . . 33
  - 2.1.7 Kits . . . 34
  - 2.1.8 Non-expendable equipment . . . 35
  - 2.1.9 Software and online resources . . . 37
- 2.2 Methods . . . 39
  - 2.2.1 Cell culture . . . 39
    - 2.2.1.1 Fibroblast cell culture . . . 39
    - 2.2.1.2 iPSC culture . . . 39
      - 2.2.1.2.1 CRISPR/Cas9-based genome editing . . . 40
      - 2.2.1.2.2 Singularization of iPSCs . . . 41
  - 2.2.2 Analyses of nucleic acids . . . 42
    - 2.2.2.1 DNA extraction . . . 42
    - 2.2.2.2 Polymerase chain reaction . . . 42
    - 2.2.2.3 Agarose gel electrophoresis . . . 43
    - 2.2.2.4 Sanger Sequencing . . . 43
    - 2.2.2.5 Deep amplicon sequencing . . . 44

2.2.2.6 ArrayCGH . . . 44
2.2.3 Cytogenetic analyses . . . 45
2.2.3.1 Karyotyping . . . 45
2.2.3.2 Sister chromatid exchange assay . . . 45
2.2.4 Analyses of gene expression . . . 46
2.2.4.1 Single-cell RNA sequencing . . . 46
2.2.4.2 FACS analysis . . . 46
2.2.5 Analyses of proteins . . . 47
2.2.5.1 Protein extraction . . . 47
2.2.5.2 Western blot analysis . . . 47
2.2.5.3 Mass spectrometry . . . 48
2.2.6 Immunocytochemistry assays . . . 49
2.2.6.1 Pluripotency markers of iPSCs . . . 49
2.2.6.2 UFBs and mitotic errors . . . 49
2.2.7 Statistics . . . 50

3 Results **51**
3.1 Bloom syndrome patient cohort . . . 51
3.1.1 “Bloom syndrome and overlapping phenotypes”: The BS flyer . . . 51
3.1.2 Assessment of clinical findings of BS patients . . . 54
3.2 Characterization of the cellular phenotypes of BS patient-derived fibroblasts . . . 57
3.2.1 Confirmation of pathogenic *BLM* variants of the BS fibroblasts . . . 57
3.2.2 Defining gene expression profiles of BS cells . . . 60
3.2.2.1 Single-cell transcriptome sequencing . . . 61
3.2.2.1.1 Quality control assessments of scRNAseq data 63
3.2.2.1.2 Differentially expressed genes in BS single cells 65
3.2.2.1.3 Determination of deregulated molecular pathways in BS single cells . . . 67
3.2.2.1.4 Upregulated genes from Fanconi anemia pathway . . . 69
3.2.2.1.5 Distinct expression of genes associated with primary microcephaly . . . 71
3.2.2.1.6 Highly significant overexpression of condensin I/II complex genes in BS patient-derived single cells . . . 73
3.2.2.2 Transcriptional changes induced by replication stress in fibroblasts . . . 74
3.2.2.3 Analyses of condensin I/II complex members on a protein level . . . 75
3.2.2.4 Analysis of BS cells from a larger proteomics perspective 76
3.3 Generation and systematic characterization of the BTRR complex-deficient iPSC lines . . . 79
3.3.1 CRISPR/Cas9-based generation of the BTRR complex-deficient iPSC lines . . . 79
3.3.1.1 Design of the guideRNAs with respect to BS patient mutations . . . 80
3.3.1.2 Evaluation of genome editing efficiencies of iPSCs . . . 81

3.3.1.3 Confirmation of *BLM, TOP3A,* and *RMI1* knockouts on protein level . . . 85
3.3.1.4 Pluripotency properties of the generated knockout iPSC lines . . . 87
3.3.2 Cytogenetic and molecular cytogenetic characterizations of the BTRR complex-deficient iPSC lines . . . 89
3.3.2.1 Analyses of chromosomal integrity and CNVs in KO iPSC clones . . . 89
3.3.2.2 Establishment of increased sister chromatid exchange rates . . . 92
3.3.2.3 Aberrant chromosome segregation events in the BTRR complex deficiency . . . 95
3.3.2.4 Prevalence of the ultrafine anaphase bridges . . . 97
3.3.3 Determination of gene expression profiles of the BTRR complex-deficient iPSCs on a single-cell level . . . 98
3.3.3.1 Quality control assessments of the scRNAseq data . . . 99
3.3.3.2 Analysis of differentially expressed genes in the three KO iPSC lines . . . 102
3.3.3.3 Pathway analyses on the single-cells' transcriptome . . . 106
3.3.3.4 Determination of the effect of aphidicolin on gene expression . . . 106

4 Discussion 110
4.1 Phenotypic distinctions of BS patients . . . 111
4.1.1 Genotype-phenotype correlation in BS . . . 111
4.1.2 Genomic instability and cancer predisposition in BS patients . . . 114
4.2 Alteration of gene and protein expression signatures in BS . . . 116
4.2.1 Single-cell transcriptome alterations observed in BS patients . . . 116
4.2.2 Gene expression profiles of Fanconi anemia genes in BS . . . 118
4.2.3 Expression of genes clinically associated with microcephaly . . . 120
4.2.4 Novel genetic interaction of condensin I/II complex genes and BLM 121
4.2.5 Strategies for confirmation of the scRNAseq results . . . 123
4.3 Isogenic BTRR complex-deficient iPSCs as a cellular model for BS . . . 125
4.3.1 iPSC-based disease modeling and current BS models . . . 125
4.3.2 Genome editing by CRISPR/Cas9 on iPSCs to obtain isogenic KO cell lines . . . 127
4.3.3 Genomic instability signatures of the BTRR complex-deficient iPSCs . . . 130
4.3.4 Transcription profiles of the KO iPSC samples . . . 133
4.4 Replication stress in BS and the BTRR complex deficiency . . . 136
4.5 Future perspectives and concluding remarks . . . 139

5 Supplementary Material 142

Bibliography 149

List of Figures 164

List of Tables 166

Abbreviations **167**

Acknowledgements **171**

1 Introduction

1.1 Rare diseases

Rare diseases are defined as serious, congenital conditions that affect less than 1 in 2000 people in Europe or less than 1 in 15,000 individuals in the United States [1]. Although every single rare disease affects only a small number of individuals, rare diseases collectively affect more than 300 million individuals worldwide [2]. In the European Union, it is estimated that between 27 and 36 million people are affected by one of the more than 8000 rare diseases described, representing 6–8% of the population [3]. Rare diseases often show a high level of symptom complexity which complicates clinical diagnostics and frequently leads to a prolonged time until a definitive diagnosis is established. This often results in a severe health burden on affected patients, their families, and the healthcare system. Compared to common diseases, rare diseases do not receive a comparable level of public interest, which directly affects patient care and the development as well as the availability of therapeutic approaches. Moreover, clinical diagnosis of rare diseases is often challenging since it strongly relies on the expertise of the clinicians in dealing with these complex syndromes that often affect multiple organs.

More than 80% of rare diseases are of genetic origin [4]. Therefore, the identification of the underlying genetic cause not only provides important insights into the molecular and cellular pathomechanisms, but also helps to improve patient management in terms of existing therapeutic strategies, as well as the development of novel therapeutic approaches [5]. The possibility of therapeutic approaches is particularly important as more than half of the rare genetic diseases are present at birth or have their onset

during early childhood [5]. The diagnosis of rare diseases has become more feasible with the recent advances in the next generation sequencing (NGS) based technologies, which consequently helps to reduce the diagnostic delay and end the diagnostic odyssey of the patients while enabling better disease management and more accurate genetic counseling for additional family members.

1.2 NGS-based gene identification strategies

In the field of medical genomics, disease-causing gene identification has been greatly facilitated by advances in NGS-based technologies, which utilize high throughput parallel sequencing of patient samples in a cost and time-efficient manner. Such technologies, which are also used in diagnostics, can include NGS-based multigene panel sequencing, as well as whole exome- and whole genome sequencing (WES and WGS). Multigene panel sequencing involves sequencing of a custom-defined gene group. The target genes are defined according to known disease associations based on shared medical conditions among several disorders, e.g., primary microcephaly, hearing loss, or eye-related phenotypes. These multigene panels can be expanded over time depending on novel disease/phenotype associations. However, multigene panels can give negative results in cases of causal variants affecting unusual disease-causing genes. In this case, WES or WGS can be applied for the detection of the disease-causing variant in the progress of diagnosis. WES includes sequencing of exon-spanning regions of coding genes, which contain approximately 85% of mutations that are known as disease-causing [6]. Yet, the aggregated exonic regions correspond to approximately 1–2% of the complete genome; therefore, WES is a more cost and time-efficient NGS-based approach when compared to WGS [7].

In unsolved cases, WGS can be further applied with the aim of sequencing the complete genome to search for structural variants or distal regulatory elements which might be disease-associated [8]. Further, in the case of WES and WGS, a trio approach, that is to sequence the parents along with the index, can facilitate the detection of

disease-causing variants by revealing the state of variants as de novo or biallelic, which overall extends the diagnosis possibility. In case of variant detection, NGS-based results should be confirmed by Sanger sequencing which serves as a gold standard for verification of these variants, although this approach can be further discussed considering higher coverage possibilities. Of note, the functional analyses for novel disease-associated variants and genes are crucial to understand the molecular pathomechanisms leading to phenotypic features and the roles of respective genes on cellular processes.

NGS-based approaches still might encounter challenges in the diagnostic field in the case of disease-causing variants residing in GC-rich or repetitive regions, or structural variants that are associated with a disease phenotype [9]. This is based on the fact that most of the NGS-based diagnostic tools use short read sequencing which provides reads of 25–400 base pair length [10]. A novel technique is long read sequencing, which can produce up to 30,000 bp reads, and provides a broader perspective on the human genome. In fact, the complete human genome sequence was completed in 2022 by advances in long-read sequencing which made it possible to sequence highly repetitive regions as well as segmental duplications [11]. Although applications of long-read sequencing in the diagnostic field are yet to be established, long-read sequencing provides novel insights into human genome. Overall, NGS-based technologies are rapidly evolving and the adaptation of these techniques to clinical genetics have importance for precise diagnoses.

1.3 Single-cell transcriptome sequencing technologies

The human transcriptome is a diverse collection of RNA transcripts that can vary greatly between individuals, organs, tissues, and cells, as gene expression is unique for each cell and essential for cell fate determination and differentiation. Further heterogeneity is plausible among transcriptome profiles of cells from the same tissue or cells with the same function. It is estimated that there are approximately between 19,000 and

20,000 genes that code for proteins [11]. In addition, the transcriptome includes more genes considering the non-coding regulatory elements such as miRNAs, or lncRNAs. Telomere-to-Telomere (T2T) Consortium reported on an estimation of more than 63,000 transcribed genes from the complete human genome, i.e., T2T-CHM13 reference [11], which makes the transcriptome a complicated identity. Deciphering the human transcriptome then has a crucial meaning in disease and health.

Until recently, transcriptomics studies were generally population-based, i.e., a large number of cells were evaluated in bulk, averaging the gene expression. However, bulk RNA sequencing became a challenging approach to define the gene expression profiles of a small number of cells from a population, or for evaluating identically appearing cell groups [12]. These challenges of bulk RNAseq are today eased by the advances of single-cell technologies which greatly changed the perspective in transcriptomics in the sense of discovery of unique gene expression profiles presented by single cells. The first assay for single-cell RNA sequencing (scRNAseq) was published in 2009 [13] and since then, the applications and discoveries of scRNAseq have greatly improved. Single-cell transcriptomics eased the discovery of defining novel cell types with unique gene expression profiles. For instance, a consortium aiming to generate a reference map for the complete human body on a single-cell level was initiated in 2016, dubbed the Human Cell Atlas (HCA) [14]. This project was launched to produce an atlas including every cell type that can be found at defined positions of a human body at a specific time [15]. Further single-cell approaches are performed to compare healthy versus diseased cells in terms of gene expression to elicit pathomechanisms in disease as well as to screen for therapeutic approaches [16].

There are several methodologies developed for the experimental steps of scRNAseq, namely single-cell isolation, which is followed by reverse transcription, amplification, and sequencing. Current single-cell isolation methods include droplet-based isolation or array-based isolation of cells [17]. Droplet-based technologies make use of microfluidics which encapsulate single cells in droplets like the 10x Genomics Chromium platform, which is the most commonly used droplet-based instrument for scRNAseq. This technology enables the processing of a large number of samples in one run which can

increase to around 100,000 cells; however, empty droplets or droplets with more than one encapsulated cell should be assessed carefully [18]. Cells with irregular morphology, e.g., neurons can be also difficult to process with droplet-based technologies. The second single-cell dispensing technology includes microchambers where single cells are dispensed as an array in nanoliter volumes. ICELL8 single cell system is one of the array-based systems and uses a microchip containing 5.184 wells for single cell dispersion [19]. With this technology, the content of the microwells can be assessed by an imaging system that eliminates the empty wells or wells with more than one cell from further processing. Due to this, the array-based scRNAseq approaches have a more limited capacity for a total number of cells to be sequenced depending on the well number on a chip when compared to droplet-based techniques. Novel technologies have been developed to increase the number of single cells available for sequencing using a combination of several instruments [20]. Further, computational approaches to scRNAseq data should be adjusted depending on the single-cell platform and various strategies are developed based on RNAseq analysis steps with expanded quality control measures adjusted for single cells as well as the further normalization steps [17].

1.4 Genomic instability syndromes

Genomic instability syndromes as a subgroup of rare genetic diseases are caused by pathogenic variants in genes that are involved in various cellular functions such as DNA replication and DNA damage response (DDR) pathways. Impairment of DDR pathways can have severe consequences on cellular functions. It is estimated that on a daily basis, each human cell harbors approximately 10,000 DNA lesions [21], which need to be repaired properly to maintain genomic integrity and for proper cellular activity. If not properly repaired, DNA damage can impair different cellular processes including DNA replication, transcription, and/or chromosome segregation, which can induce genomic instability, cellular senescence, and apoptosis in affected cells. Thus, mutations in the genes encoding members of these pathways can give rise to genomic instability disor-

ders including premature aging-related syndromes and cancer predisposition disorders [22, 23].

DNA repair mechanisms include a diverse series of events orchestrated by numerous proteins. In general, there are five major pathways for DNA repair of different types of lesions; base excision repair (BER), for single nucleotide DNA damage and single-strand breaks (SSB); mismatch repair (MMR) for the nucleotide mispairings; nucleotide excision repair (NER) for the removal of pyrimidine dimers and photoproducts caused by UV light; non-homologous end joining (NHEJ) and homologous recombination (HR) for the repair of DNA double-strand breaks (DSB)s [24] (Figure 1.1).

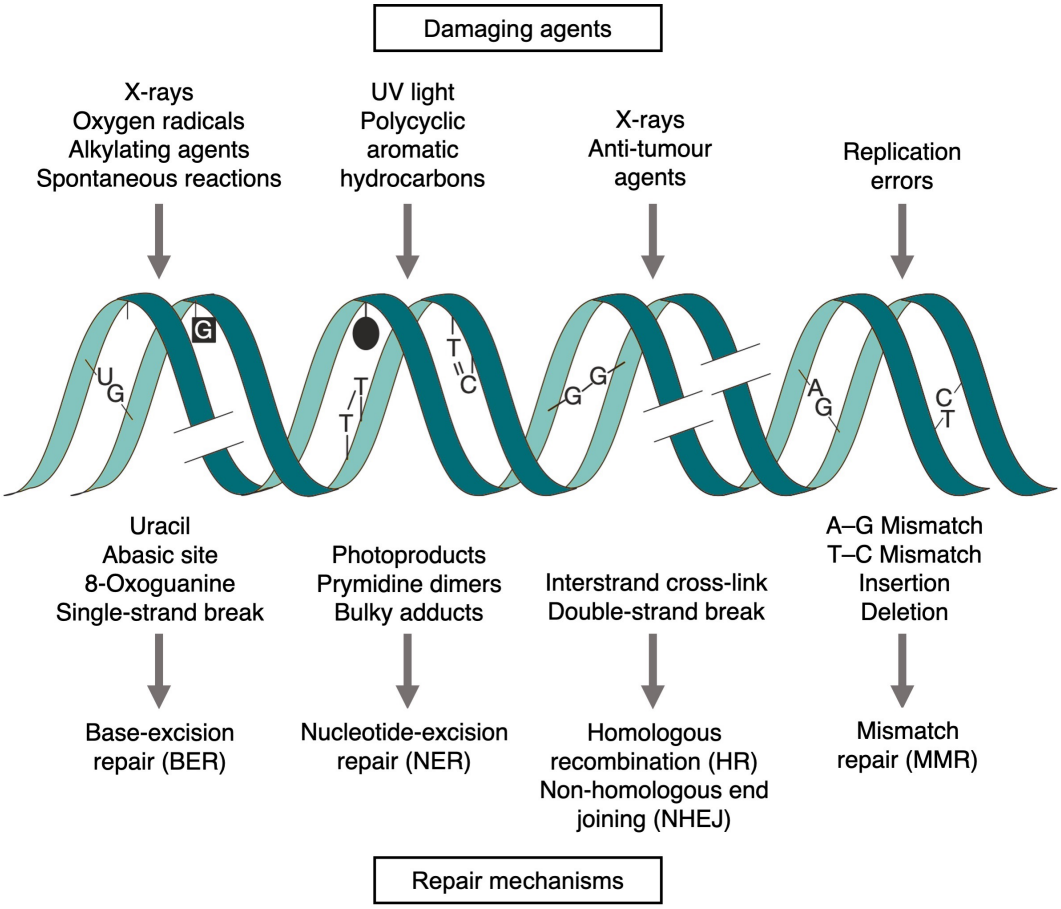


Figure 1.1. DNA damaging agents creating different types of lesions and relevant DNA repair mechanisms. Various DNA damaging agents (upper lane); consecutive DNA lesion examples (middle lane); and responsible DNA repair mechanism for the elimination of the DNA damage (bottom lane) are shown. The schematic is adapted after Hoeijmakers [25].

Mutations in genes of the BER pathway are associated with different cancer-predisposing

and immunological disorders [26]. For instance, biallelic pathogenic variants in *MUTYH* have been identified in patients with autosomal recessive familial adenomatous polyposis type 2 (FAP2; MIM 608456). This gene encodes a DNA glycosylase that particularly plays a crucial role in the repair of oxidative DNA damage, and loss of MUTYH function leads to impaired repair of 8-oxoguanine, which is one of the most frequent base lesions generated by reactive oxygen species [27]. MMR pathway, on the other hand, is responsible for the removal of the mismatched base pairs resulting in a 100–1000 fold increase in DNA replication fidelity [28]. Heterozygous mutations in the genes responsible for encoding the members of MMR pathway such as *MLH1, MSH2, MSH6*, and *PMS2* cause a dominant cancer predisposition syndrome called Lynch syndrome (MIM 120435) which is characterized by an increased risk of gastrointestinal, uterine, and ovarian tumors [29].

Another major DNA repair pathway in mammalian cells is the nucleotide excision repair (NER) pathway, which is involved in the repair of bulky, helix-distorting DNA lesions [30]. These adducts are mainly induced by endogenous effects and substances, for instance, UV light or chemical components. In humans, several hereditary disorders, like xeroderma pigmentosum (XP) and Cockayne syndrome, are caused by pathogenic variants in genes of the NER pathway, e.g., *ERCC4* and *ERCC6*, respectively [31]. These rare autosomal recessive disorders share similar phenotypic characteristics such as skin photosensitivity and early onset skin cancer, multi-system premature aging, and microcephaly with moderate to severe impairment of neurological development [32].

DSBs belong to the most deleterious type of DNA damage. For their repair, two major repair pathways are available in higher eukaryotes: NHEJ and HR pathways. As NHEJ is independent of the presence of a homologous DNA template, it is available throughout the cell cycle making it the predominant pathway for the repair of DNA DSBs. Mutations in different components of the NHEJ pathway, such as *LIG4* and *XRCC4*, are associated with several human disorders [33, 34]. These disorders are mainly characterized by mild to severe immunodeficiency which is based on the important role of NHEJ-dependent cellular processes during B- and T-cell development [35]. Additionally, pathogenic variants in *PRKDC* and genes encoding DNA ligases give rise to se-

vere cases of immunodeficiency [36]. Microcephaly, hematological anomalies, growth retardation, and radio-sensitivity are additional clinical characteristics of patients with pathogenic variants in NHEJ-associated genes. In contrast to NHEJ, the repair of DNA DSBs by HR is dependent on the presence of a homologous template, therefore it is considered to be more accurate than repair by NHEJ and in theory, it is error-free. So far, mutations in various members of the HR pathway have been identified including rare diseases such as Bloom syndrome (BS), Werner syndrome, Ataxia Telangiectasia, and Fanconi anemia (FA). These disorders all share common clinical hallmarks such as growth deficiency, a pronounced cancer predisposition, and immunodeficiency; and are caused by bi-allelic pathogenic variants in major components of the HR DNA repair pathway.

1.5 Primary microcephaly syndromes

Microcephaly is a clinical indication of impaired brain development which is a process of rapid and specific changes of gene expression in the progenitor cells. Microcephaly is defined as a reduced occipitofrontal head circumference (OFC) of more than two standard deviations (SD) at birth, or more than three SD after six months below the mean adjusted for age, gender, and ethnicity. Primary (congenital) microcephaly is observed at birth, while secondary microcephaly is acquired postnatally [37]. Furthermore, primary microcephaly can occur as an isolated phenotypic form (MCPH), or in a syndromic form, which manifests in combination with additional clinical findings. Autosomal recessive primary microcephaly, i.e, MCPH, is a rare disease and several genes have been associated with MCPH by the genetic analyses of primary microcephaly patients [38]. Functional studies allowed to further elucidate impaired cellular mechanisms and effects leading to microcephaly. The first discovered gene associated with MCPH was the *MCPH1* gene, short for microcephalin, and was shown to have roles in DNA repair and chromosome condensation during neurogenesis [39–41]. A second gene associated with spindle pole positioning in which homozygous or compound heterozygous mutations lead to microcephaly is *WDR62*. Interestingly, the expression of this gene

is restrained to neural progenitor cells which can be an explanation of the pathomechanism of microcephaly [42, 43]. The most commonly mutated primary microcephaly gene is *ASPM* (*MCPH5*), which encodes a gene product with roles in in mitotic spindle regulation and mitosis, and compound heterozygous or homozygous LoF mutations in this gene constitute approximately 50% of MCPH cases [44, 45].

Moreover, syndromic forms of microcephaly are often combined with short stature and can include monogenic disorders such as Seckel syndrome (MIM 210600, MIM 616777), microcephalic osteodysplastic primordial short stature (MOPD) type I and II (MIM 210710 and MIM 210720, respectively), LIG4 syndrome (MIM 606593, MIM 616541), and Meier-Gorlin syndrome (MIM 224690) [46]. Further, primary microcephaly combined with intellectual disability and/or accelerated aging phenotypes can be phenotypic findings of disorders caused by impairment of proteins important for DNA replication and repair-related processes. Genomic instability is a characteristic in disorders such as Cockayne syndrome (MIM 133540), Fanconi anemia (MIM 227650), or Bloom syndrome (MIM 210900), where primary microcephaly is a prominent clinical finding.

In general, both forms of microcephaly are associated with mutations in genes encoding proteins acting on essential cellular processes for cell division and DNA management. These processes can be centriole biogenesis, microtubule dynamics, spindle pole positioning, cell cycle regulation, as well as DNA replication and repair [47]. Hence, impaired proliferation due to alterations in cell division during brain development can lead to microcephaly, defined as a reduction in brain size up to 1/3 of its volume [37]. One possible explanation for this is based on the imbalance of the symmetrical and asymmetrical divisions of the neural progenitor cells during the development of the brain [43]. The neurons of the cerebral cortex are derived from progenitor cells, which initially divide symmetrically into two identical daughter cells leading to an increase of the progenitor cell pool at the ventricular zone [48]. However, lack of proper positioning of the centrioles can lead to premature asymmetric division of the progenitor cells resulting in early production of neurons and a reduction in the size of the progenitor cell pool, leading to microcephaly [43]. The exact molecular mechanisms of this switch from symmetrical division to neuron differentiation remain generally unknown. Overall,

reduced brain size in primary microcephaly patients can result from impaired neurogenesis, which can be due to defects in mitotic processes, cytokinesis, DNA replication, and repair processes, leading to genomic instability [38].

1.6 Bloom syndrome

Bloom syndrome (BS, MIM 210900) was first described by David Bloom in 1954. He reported a child with telangiectatic erythema and short stature, and compared the phenotype to two previously presented cases suggesting that all three children were affected by the same, so far uncharacterized disorder [49]. In the following years, James German clinically characterized additional patients and expanded the phenotypic spectrum of this entity. Additionally, he established the first BS patient registry containing the clinical data of initially twenty-seven patients with similar phenotypes [50]. In the following years, German and colleagues continued to characterize the phenotype of BS patients and defined the cellular characteristics associated with BS. They could show that cytogenetic impairments, mainly aberrant chromosomes, were a key feature in BS patient-derived cells, which was used for a long time as a cellular hallmark to confirm the clinical diagnosis of BS [51]. Thanks to first advances in molecular genetic technologies, the underlying genetic cause of BS was identified and designated as *BLM* gene [52]. The *BLM* gene product is a helicase responsible for genome maintenance [53]. To date, less than 300 patients with BS have been clinically and genetically characterized through the Bloom Syndrome Registry (BSR) and a few more reported patients from around the world [54, 55]. The prevalence of BS is most prominent in the Ashkenazi Jewish population accounting for one-third of BS patients and a carrier frequency of 1% within this population [56].

1.6.1 Phenotypic characteristics of BS

The initial clinical hallmarks of BS, which led to the characterization of BS as a distinct entity, were a unique combination of short stature and sunlight-sensitive telangiectatic

erythema which was observed mainly on the face of affected individuals [49]. Currently, clinical characteristics of BS include primary microcephaly, photosensitivity associated with skin lesions, moderate to severe immunodeficiency, and a predisposition to cancer, such as hematological malignancies, colorectal cancer, or breast cancer [55, 57]. In BS patients, OFC is 3 to 8 SD below the age-, sex-, and ethnically matched means. Interestingly, BS patients do not show signs of intellectual disability or developmental delay [58]. Low OFC values are accompanied by facial dysmorphisms including a narrow face, prominent ears, and nose which overall create a facial phenotype in BS [59]. Although some phenotypic characteristics of BS, such as microcephaly and growth retardation, are shared features among the family of inherited genome instability disorders, specific clinical characteristics, like dermatological and immunological anomalies in many cases ensure the definitive diagnosis of classical BS in affected individuals. The clinical features of BS will be discussed in the following and given as a summary in Table 1.1.

Table 1.1. Clinical chracteristics of BS (MIM 210900).

Growth	Skin	Immunology	Neoplasia
Growth failure Microcephaly	Photosensitivity Hypertrichosis Café-au-lait spots	Immunoglobulin deficiency Low IgA, IgG, IgM Recurrent infections Chronic lung disease	Leukemia Lymphoma Adenocarcinoma Squamous cell carcinoma

The most prominent characteristics of BS, i.e., primary microcephaly and growth retardation, are already present at birth and can, in some cases, even be detected prenatally by week 30–32 of the pregnancy. BS patients have a reduction in birth height and weight of more than 2 standard deviations (SD) below the mean value calculated accounting for ethnicity and gender [60]. Affected individuals show growth retardation throughout life with mean average height of 148.5 cm for males and 141.5 cm for females, and a mean body weight of 41.3 kg in males and 36.6 kg in females [60]. Growth hormone production as well as secretion are normal in BS patients and there have been different trials investigating growth hormone treatment in children with BS. Several patients who

underwent growth hormone therapy developed cancer, especially leukemia and lymphoma [57]. Currently, it is unclear whether the development of cancer is caused by the administration of growth hormones or if it is a result of the overall increased risk of BS patients to develop cancer. Still, BS is currently considered a contraindication for growth hormone administration [61–63].

Dermatological features of BS particularly include malar rash, redness on the face and especially on cheeks, skin sensitivity to sunlight exposure, and café-au-lait spots in various parts of the body [64]. Most skin lesions are observed in childhood and can deteriorate over time by continuous sunlight exposure. Thus, skin protection by covering and usage of sunscreens are strongly advised in BS [62]. Café-au-lait spots, which are hypo-/hyperpigmented macules, are observed over the whole body and occur soon after birth. Additionally, hypertrichosis has been described in some patients with BS [65]. Dermatological characteristics can vary considerably among BS patients, which might complicate and delay the clinical diagnosis of BS [66].

Immunodeficiency is considered a clinical key feature of BS. Most BS patients show moderate to severe impairment of the immune system by means of a low amount of at least one class of immunoglobulins [57, 67]. Most commonly, serum levels of IgM, IgA, and IgG are severely decreased in BS patients [57]. This results in recurrent, in some cases life-threatening infections, especially of the upper respiratory tract. Pneumonia and otitis media infections were seen frequently in BS patients, but also an increased susceptibility to primary ear infection has been observed [62].

Cancer predisposition is one of the most pronounced lethal hallmarks of BS. It is the main reason of the severely reduced lifespan of BS patients, which lays in the third decade of life [54]. In the BSR which is currently the most extensively studied patient cohort available, cancer and malignancy are estimated to occur in approximately 50% of BS patients. Overall, it is estimated that the incidence to develop any type of malignancy by the age of 40 is over 80% in BS patients [68]. The spectrum of malignancies includes hematological malignancies, such as non-Hodgkin-lymphoma, acute lymphatic lymphoma (ALL), and acute myeloid lymphoma (AML), as well as solid tu-

mor. No specific type of malignancy is particularly prominent in BS, albeit rare types of tumors such as Wilms tumor were detected with a higher frequency in BS children [57]. Molecular diagnosis in BS is highly important for cancer predisposition phenotype considering surveillance, management, and possible therapy options. Thus, frequent cancer screens are advised in BS patient management, and medical treatment of BS patients as well needs to be adjusted and modified with regard to the increased cancer risk [62].

1.6.2 Genetic causes of BS phenotype

In 1994, James German and colleagues mapped the chromosomal region associated with BS to chromosome 15. By analyzing different consanguineous families, they were able to show that the underlying gene and mutation leading to BS is located on 15q26.1 [69]. In the following years, the *BLM* gene within this 250 kb region was identified and it was shown that pathogenic variants in this gene were responsible for BS [70]. Currently, there are approximately 60 different homozygous and compound heterozygous pathogenic variants in *BLM* that have been identified in BS patients. A great majority of these mutations were described by the BSR which included genetic analyses of 164 BS patients (Figure 1.2). Most of these variants are protein truncation variants, either by directly leading to the generation of a premature stop codon or by inducing it through affecting the splicing. Also, pathogenetic missense variants in the *BLM* gene have been identified (Figure 1.2). These missense variants are located in functional domains of BLM and affect highly conserved amino acids, thereby, most likely, also leading to loss of BLM function [54]. For instance, BLM has four cysteine domains which serve as ion binding domains for proper ATPase functioning [71]. Missense mutations affecting these domains were identified in BS patients such as the p.Cys1055Ser mutation which is one of the most frequent missense mutations identified in *BLM* [54]. The *BLM* gene encodes one of the five human RecQ-like helicases that is a 3´-5´ DNA helicase which unwinds DNA and is especially important during DNA replication and DNA repair processes [53].

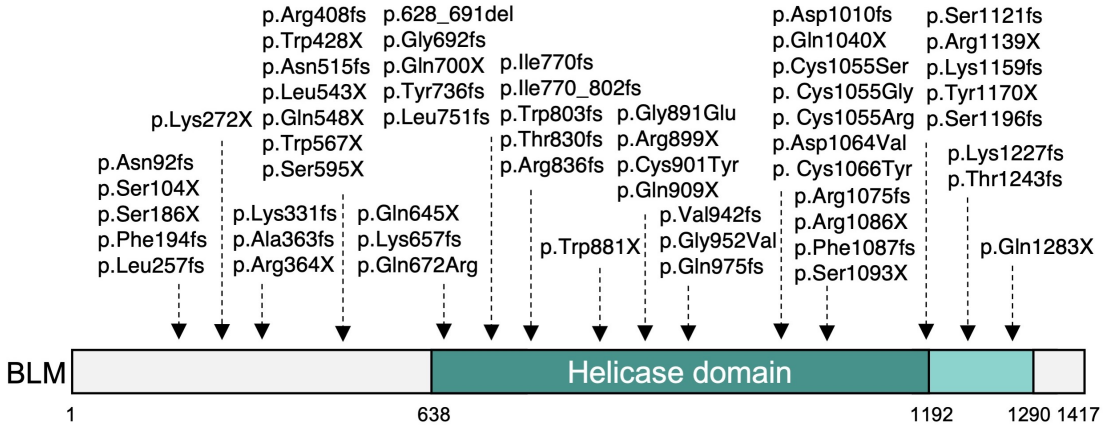


Figure 1.2. Mutation spectrum of *BLM* reported by the Bloom Syndrome Registry with respect to schematic representation of BLM helicase. Frameshifting, nonsense, and missense mutations were detected in *BLM* and reported by German *et al.* [54] in the Bloom Syndrome Registry (BSR). Conserved domains of BLM are shown with colors: helicase domain, dark green; RNase D-like C-terminal (HRDC) domain, light green. Domains are adapted from Bythell-Douglas & Deans [72].

For a long time, BS was considered a congenital disease solely caused by pathogenic variants in *BLM*. Recent advances in NGS techniques led to the identification of novel disease-causing genes in patients with BS phenotype. Approximately 20 years after the identification of *BLM* as the first causative gene for BS, pathogenic variants in a second, novel gene were associated with a BS-like phenotype. Hudson et al. identified a large homozygous deletion of 80 kb on chromosome 16 spanning the complete *RMI2* gene in two siblings [73]. Both individuals presented with a BS-like phenotype including microcephaly, growth retardation, and café-au-lait spots, although the authors reported an overall milder phenotype in terms of skin photosensitivity and immunodeficiency [73].

Two years later, an additional study consecutively reported on biallelic LoF mutations in the *TOP3A* gene in individuals presenting with a BS-like phenotype [74]. Patients were mostly diagnosed in early childhood and showed primary microcephaly and growth retardation with similar SD values as observed in BS patients. Dilated cardiomyopathy is not a canonical BS feature, yet it was present in some of the patients harboring biallelic *TOP3A* mutations. Immunodeficiency was not reported as a specific phenotype but recurrent infections were remarked in some *TOP3A*-associated patients. Furthermore, the authors presented two additional patients exhibiting a mild, BS-like phenotype both

carrying a homozygous LoF mutation in the *RMI1* gene. This variant was detected by WES in two individuals with primary microcephaly and growth retardation on a consanguineous family. The variant was predicted to introduce a premature stop codon (c.1255_1259delAAGAA;p.Lys419Leufs*5) in *RMI1* transcript matching to anterior site of a conserved domain of the protein; therefore it was reported as pathogenic [74].

On a molecular level, it was already known that the proteins encoded by these four genes, i.e., *BLM, TOP3A, RMI1,* and *RMI2* molecularly interact with each other to form a functional complex, the BTRR complex [75–77]. *TOP3A-, RMI1-*, and *RMI2*-associated patients presented common clinical phenotypic features of BS such as primary microcephaly and growth retardation. Besides the common phenotypes observed in patients, the functional connection between these genes is also reflected on the cellular level. Increased sister chromatid exchange (SCE) rates were confirmed on the *TOP3A* and *RMI2*-mutated patient-derived cells as a molecular confirmation of BS on a cellular level. Further mitotic anomalies such as chromatin bridges and micronuclei events were also detected to be higher in patient cells [73, 74]. Characterization of cellular phenotypes of *RMI1*-associated patients was not reported although depletion of RMI1 in various cell types resulted in similar cellular features as seen for the other members of the BTRR complex.

1.7 The BTRR complex

BLM forms, together with TOP3A, RMI1, and RMI2, the BTRR complex. The BTRR complex is important for the maintenance of genomic stability through its roles in dissolving DNA intermediates that occur during fundamental cellular processes such as DNA replication and DNA repair. In the following sections, each member of the BTRR complex and its function is explained in detail as well as the roles of the BTRR complex within the context of DNA replication and repair processes.

1.7.1 BLM RecQ-like helicase

Helicases are highly conserved ubiquitous enzymes that are able to unwind DNA, RNA and DNA/RNA duplexes [78]. DNA-dependent helicases break the hydrogen bonds by using the energy dispensed from the hydrolysis of mainly ATP [79]. A subgroup of DNA-helicases are the RecQ-related family DNA helicases that are named based on their similarity to the RecQ enzyme of *Escherichia coli*. The RecQ helicases are conserved proteins composed of a helicase domain and an RNase D-like C-terminal (HRDC) domain [80]. The helicase core contains the ATPase domain and the zinc-binding subdomain which are both important for the enzyme activity [81]. The ATPase domain includes residues that bind and hydrolyze ATP which is a crucial step required for the enzymatic activity of the helicase [82]. For a long time, the function of the conserved HRDC domain was unknown until recent studies in 2020 which revealed that the HRDC domain is a domain for the reinforcement of DNA binding [83, 84].

In humans, five different RecQ-like family DNA helicases are known: RECQL1, WRN (previously named RECQL2), BLM (formerly also named RECQL3), RECQL4, and RECQL5. Loss-of-function (LoF) mutations in *WRN, BLM*, and *RECQL4* have been associated with autosomal recessive disorders such as Werner syndrome, Bloom syndrome, and a spectrum of disorders including Rothmund–Thomson, RAPADILINO, and Baller–Gerold syndrome, respectively. The common phenotypic characteristics of disorders caused by RecQ helicase deficiency include short stature, photosensitivity, and immunodeficiency, although each disorder presents its own differential clinical characteristics. Recently, a homozygous mutation in *RECQL1* has been identified in patients with short stature and sunlight sensitivity [85], whereas mutations in *RECQL5* have not been described yet. An overall genomic instability profile that is observed in disorders linked to RecQ deficiency provides evidence for the importance of helicases in cellular processes of DNA replication and repair mechanisms [79].

The BLM helicase is a ubiquitously expressed, 1417 amino acid protein containing the conserved domains present in all RecQ helicases, i.e., a helicase core and an HRDC domain (Figure 1.3). In addition, BLM has a nuclear localization signal at the C-terminus

responsible for translocation of BLM to the nucleus. Studies performed with Sgs1, the yeast homolog of human BLM, revealed its diverse roles in the homologous recombination pathway such as DNA end resection, branch migration, and double Holliday junction dissolution [86]. Important interaction partners of yeast Sgs1 are yeast topoisomerases II and III [87, 88]. Similarly in humans, human BLM interacts with TOP3A where both are parts of the quaternary BTRR complex; which confirms the importance of the interaction of BLM with topoisomerases from an evolution perspective [79]. One of the well-established roles of the BTRR complex is dissolving double Holliday junctions that result in the course of the repair of DNA DSBs via the HR pathway. Furthermore, BLM helps to stabilize stalled replication forks and it can promote fork regression which induces DNA repair in vitro [89]. The helicase activity of BLM is also responsible for the dissolution of the G-quadruplexes (G4 DNA) which are four-stranded structures resulting from repetitive guanine nucleotides that interact with each other through Hoogsteen bonds [90, 91]. G4 DNA is a preferred substrate for BLM helicase, and if not resolved, can cause replication fork stalling [90]. Due to its roles in DNA replication and repair processes, BLM is predominantly expressed in the S-phase and continues to persist during G2/M, while its amount declines in G1 [92]. In addition, BLM is phosphorylated during mitosis which is important for proper sister chromatid segregation although the exact function of this activity is not completely understood [93–95].

BLM can interact with further proteins of DDR pathways such as specific Fanconi anemia (FA) proteins. The FA pathway includes a series of events orchestrated by FA proteins that are important for the removal of interstrand crosslinks (ICL) which can lead to replication fork stalling [96]. FANCM is a member of the FA pathway which is co-recruited to the stalled forks with the BTRR complex by directly interacting with members of the BTRR complex [97, 98]. Upon recruitment, FANCM binds to seven more FA proteins forming the FA core complex that is activated by BLM interaction, followed by a cascade of events which eventually leads to both repair and bypass of ICLs [97]. Roles of BLM and FA proteins have been shown for DNA replication checkpoint as well. This pathway is triggered due to replication stress by activating ATM/ATR kinases leading to a cell cycle stalling to create enough time for the repair of the lesions [99]. BLM is a

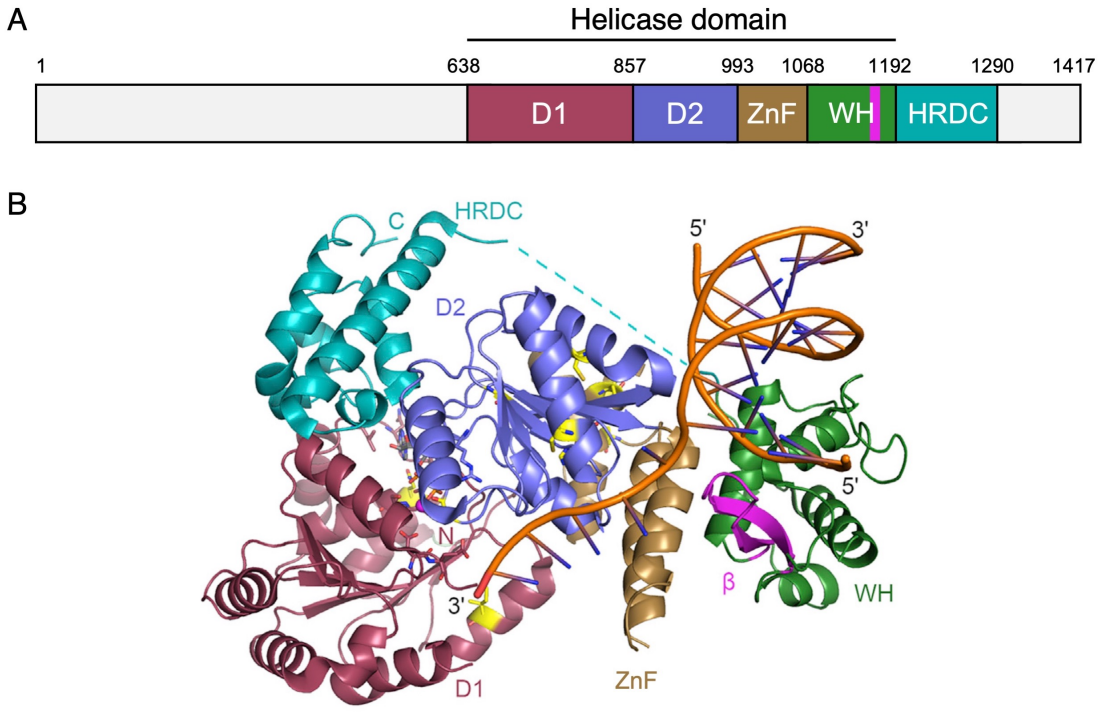


Figure 1.3. Schematic representation and crystal structure of BLM helicase. Domain colors are consistent in (A) schematic representation and in (B) crystal structure of BLM bound to DNA. DNA is shown in orange color. Abbreviations are as follows: D1 and D2, RecA-like fold domain 1 and 2; ZnF, zinc finger; WH, winged helix; HRDC, helicase and RNaseD C-terminal. Figure is adapted from Bythell-Douglas & Deans [72].

target of ATM/ATR pathways and thus is phosphorylated at the stalled replication forks, which locates the place of the stalled forks that is important for their restart [100, 101]. The stalled replication forks then can restart by lesion by-pass or fork regression and template switching, followed by HR repair of the lesion via dHJs [102]

Due to its roles in DNA replication and repair processes, BLM is predominantly expressed in the S-phase and continues to persist during G2/M, while its amount declines in G1 [92]. In addition, BLM is phosphorylated during mitosis which is important for proper sister chromatid segregation although the exact function of this activity is not completely understood [93–95].

During mitosis, BLM and FA protein interaction occurs frequently at chromatin bridges which can eventually lead to micronuclei and chromosome segregation defects [103,

104]. For instance, under-replicated DNA, e.g., from the common fragile sites, can cause DNA bridges known as the ultrafine bridges (UFB) [105, 106]. These UFBs are associated with FANCD2 foci and BLM helicase during anaphase. Currently, the exact molecular link between BLM and the FA proteins for the management of the chromatin bridges is not fully elucidated.

The link between immune deficiency and DNA repair mechanisms is defined through the affected recombination process during the production of the antigen receptors [107]. Although previous studies showed that BLM neither act on the V(D)J recombination nor is essential for somatic hypermutation of B cell receptors [108, 109], the roles of BLM in BER and alternative end-joining during the lymphocyte development have been identified [110, 111]. In addition, ablation of the murine Blm in B-lineage cells severely compromised the B cell development, maintenance, and function [109]. Similarly, proliferation and development of mice T-cell lineages were affected by Blm deficiency [112]. To date, the exact molecular pathomechanism of immunodeficiency in BS is yet to be elucidated.

1.7.2 Topoisomerase 3A

DNA topoisomerases are enzymes that control the topology of DNA by converting relaxed and supercoiled forms of DNA into each other. All living cells express topoisomerases to manage the supercoiling that arises during processes such as transcription, replication, and repair. The requirement for topological changes in DNA arose as early as the DNA became circular, e.g., in bacteria; or as it reached a certain size, like in eukaryotic cells [113]. The torsional forces are regulated by topoisomerases by either transporting DNA through a break on one strand (type I topoisomerase) or by passing larger regions through a double-stranded break (type II topoisomerases) [114]. Further classification of the topoisomerases depends on both amino-acid sequence and reaction mechanism [115]. Human DNA topoisomerase III alpha (TOP3A) belongs to the type IA topoisomerases that are a group of topoisomerases conserved from bacteria

to higher eukaryotes [116]. This type of topoisomerases nicks a single strand of DNA, transport the other DNA strand, and seal the nick before releasing the DNA.

Human TOP3A is a 1001 amino acid protein containing shared conserved domains, such as the topoisomerase-primase subdomain (TOPRIM) and the catalytic TOP IA domain [117] (Figure 1.4). Furthermore, TOP3A includes a conserved zinc-binding domain at its C-terminal end which is important for the DNA binding activity [72]. TOP3A forms part of the nuclear BTRR complex, in which it is important for relieving the torsional stress induced by the unwinding of DNA helix by BLM. The decatenation activity of TOP3A is stimulated by RMI1 [118, 119]. The role of TOP3A is crucial within the BTRR complex for processes like the dissolution of the HR intermediates at double Holliday junctions. In line with the cellular function of TOP3A, patient cells carrying biallelic LoF variants in *TOP3A* contained higher levels of unresolved DNA intermediates leading to mitosis defects such as chromatin bridges [74]. Of note, knockout of the mouse homolog Top3a severely impaired the viability in very early stage of embryonic development implying an essential role of Top3a during early development [120].

In humans, TOP3A is found in both the nucleus and the mitochondria, and the mitochondrial translocation is submitted via a mitochondrial targeting signal at the N-terminus of the protein [121]. The mitochondrial isoform of TOP3A is important especially during mitochondrial DNA (mtDNA) replication and for mtDNA maintenance due to its decatenation capacity. Impairment of TOP3A activity leads to errors in the segregation of mtDNA, and biallelic variants in *TOP3A* were recently identified in a patient with adult-onset mitochondrial disease [122]. A recent study also reported on two siblings with mitochondrial dysfunction and a BS-like phenotype who were compound heterozygous for LoF *TOP3A* mutations [123]. Previously, Martin et al. also reported on dilated cardiomyopathy and mitochondrial DNA depletion in muscle tissue in a subgroup of TOP3A patients which might be explained by impairment of mitochondrial function in these individuals [74]. Taken together, TOP3A has important roles in the maintenance of genomic as well as mitochondrial DNA material in diverse species.

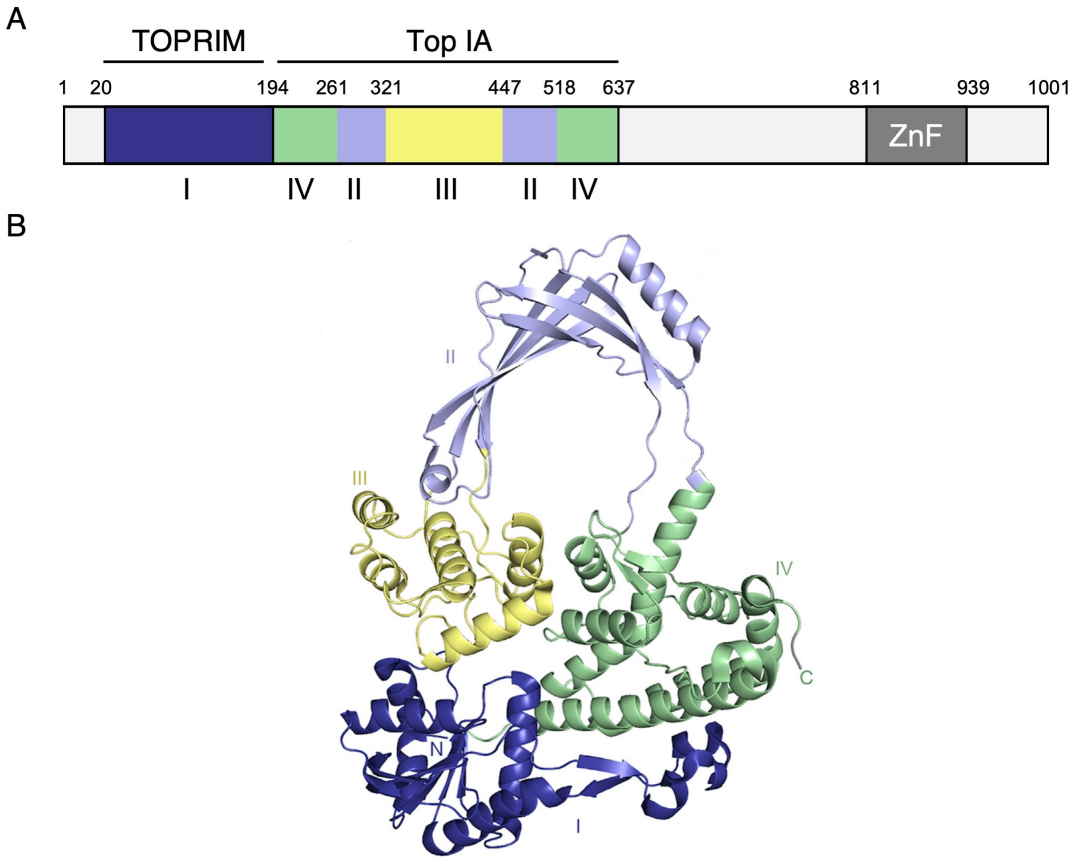


Figure 1.4. Schematic representation and crystal structure of TOP3A. Domain colors are consistent in (A) schematic representation and (B) crystal structure of TOP3A. Abbreviations are as follows: TOPRIM, topoisomerase-primase; Top IA, topoisomerase type IA central domain; and ZnF, zinc finger domain. Figure is adapted from Martin et al. [74] and from Bythell-Douglas & Deans [72].

1.7.3 RecQ-mediated genome instability proteins 1 and 2

The other two members of the BTRR complex are RecQ mediated genome instability protein 1 (RMI1) and RecQ mediated genome instability protein 2 (RMI2). Both proteins are fundamental for the maintenance of genomic stability through their roles and interaction in the BTRR complex [124, 125]. RMI1 is a 625 amino acid protein that interacts with DNA and consists of a three-helix bundle (3HB) domain and two oligonucleotide binding (OB1 and -2) domains [126] (Figure 1.5). The OB domains are conserved among species such as *Caenorhabditis elegans* and *Arabidopsis thaliana,* while the yeast homolog textitRmi1 only includes the OB1 domain [127]. RMI2 is a smaller protein of 149 amino acids that is found in higher eukaryotes and contains only

a single OB domain (Figure 1.5). In contrast to RMI1, protein orthologs of RMI2 are only present in higher eukaryotes and missing, e.g., in *C. elegans* and yeast [124, 125]. Of note, OB domains of both RMI1 and RMI2 show sequence similarities and DNA binding properties both between them and the OB domains of the RPA protein [125, 128].

RMI1 and RMI2 interact with each other and the DNA through their OB domains [126]. RMI1 interacts further with BLM and TOP3A via its OB1 domain [118, 129]. Hence, RMI1 is the member which holds the BTRR complex together by direct binding of the BTRR complex members. Furthermore, RMI1 stimulates the decatenation activity of TOP3A and can interact with FANCM in the course of ICL removals [130, 131]. During mitosis, RMI1 is required for phosphorylation of BLM which results in BLM deactivation during G1 [77, 95]. Depletion of RMI1 in human cells through shRNA resulted in reduced proliferation rates in HeLa cells [77], while complete knockout of *Rmi1* in mice resulted in embryonic lethality [132, 133]. In humans, loss of *RMI2* led to increased rates of chromatin bridges and micronuclei as observed in patient cells [73]. No mouse study on *Rmi2* knockout has been published yet. Overall, both RMI1 and RMI2 are indispensable members of a fully functioning BTRR complex [126, 129].

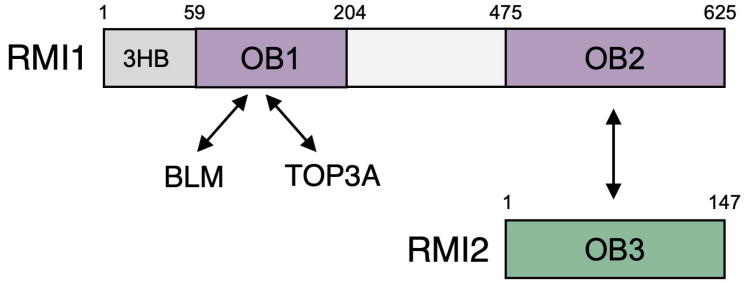


Figure 1.5. Schematic representations of RMI1 and RMI2 proteins and conserved domains. Interactions of domains with BLM, TOP3A or between RMI1 and RMI2 are shown with arrows. Abbreviations are as follows: 3HB, 3 helix bundle and OB, oligonucleotide binding domain. Figure was adapted after Hoadley *et al.* [129].

1.8 The functional roles of the BTRR complex

Several roles of the BTRR complex during the HR DNA repair process as well as DNA replication processes have been discovered. For the HR repair, a homologous DNA strand is used as a template sequence to repair DSBs on the affected allele. After localizing the sequence region surrounding the DSB at the homologous allele, the broken strand which is associated with RAD51 invades the template strand, and new DNA is synthesized according to the homologous template strand [134]. The resulting DNA complex including the template strand and the newly synthesized DNA form together with the invading strand a so-called double Holliday junction (dHJ), which needs to be solved. The entanglements can be solved by endonucleases such as GEN1 and SLX1/SLX4, but this results in crossover products between the two DNA strands in a process called resolution [135]. Alternatively, the BTRR complex can dissolve these dHJs without the formation of a potential crossover process [136]. During this process called the dHJ dissolution, the BLM helicase breaks the hydrogen bonds formed between the template and the newly synthesized DNA while moving the branches of the dHJ together [137] (Figure 1.6). Finally, TOP3A unhooks the hemicatenated DNA structure, hence leading to the separation of the two DNA strands. Currently, the BTRR complex is the only known protein complex that is capable of dissolving the dHJs into non-crossover products and this process is one of the well-documented functions of the BTRR complex [72].

The BTRR complex takes on further roles at the stalled replication forks. During DNA synthesis, replication forks may stall or collapse due to the blockages in the DNA such as breaks or DNA adducts. For instance, ICLs, which are covalently bound bases, cause replication fork stalling and can be repaired by a specific pathway, i.e., the FA pathway. The BTRR complex recruits FANCM to the stalled replication forks at upstream of the FA pathway and the FA pathway launches for the repair of the ICLs [97]. The FA pathway induces a series of events to remove the ICLs by DSB formation which is followed by HR DSB repair in a process called replication-coupled ICL repair during S-phase [138]. The resulting DSBs are repaired by HR DNA repair at the downstream

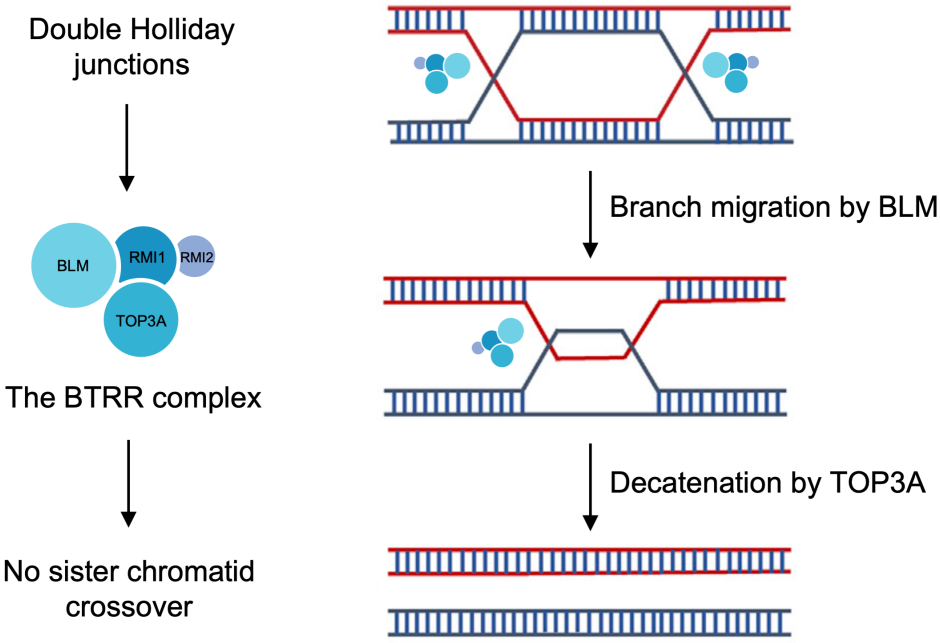


Figure 1.6. Schematic representation of double Holliday junction (dHJ) dissolution by the BTRR complex. Post-synapsis stage is shown at the right side. The branches of the dHJs are moved towards each other by the helicase activity of BLM followed by the unhooking of the hemicatenated DNA structure by TOP3A activity. The cartoon is adapted after Martin *et al.* [74].

of FA pathway by the BTRR complex. Therefore, BLM and the BTRR complex favor the restart and management of the stalled replication forks in cooperation with the FA pathway [139, 140].

Overall, the BTRR complex is important during DNA replication and DNA repair, which makes it one of the guardians of genome integrity [86]. Its interaction with various proteins from numerous DNA-related pathways reveals its importance for the cell although the complete functional definition and interaction partners of the complex remain as outstanding questions.

1.9 Cellular pathomechanisms in BS

The importance of the BTRR complex is further stressed by the effect that disruptions of the BTRR complex function have on organisms. In humans, biallelic LoF mutations in

all members of the BTRR complex are the cause of BS. On a cellular level, BS patient-derived cells show a pronounced genomic instability signature caused by unresolved DNA recombination intermediates, impaired chromosome segregations, and mitotic errors.

The dissolution of the dHJs is one of the well-documented roles of the BTRR complex. When the BTRR complex function is impaired, dHJs are resolved through alternative pathways using nucleases which can lead to crossover events between the homologous alleles. In line with this, one of the most prominent cellular phenotypes of BS cells is a significant increase in SCE rates due to increased somatic intergenic recombination events [52, 141]. The high rates of SCE were one of the first defined cytogenetic characteristics of BS [142]. The SCE assay, which is based on the visualization and quantification of the exchanges between the sister chromatids, was used as a cytogenetic diagnosis tool for BS, especially when the molecular genetic tools were limited. With the recent advances in single-cell sequencing techniques, the SCE events were located to regions rich in G4 structures that are specific targets for the BLM helicase [143]. Interestingly, high rates of SCEs were not observed in FA patients implicating the specificity of the SCEs in the BS cells although the two disorders share overlapping phenotypic features [141].

Unresolved recombination intermediates resulting from the impairment of the BTRR complex can also lead to aberrations in mitotic processes. These DNA intermediates might persist into anaphase forming bridge-like structures between the sister chromatids. The so-called anaphase bridges can be further classified based on their structure which affects the capability of DAPI staining. This depends on, e.g., the existence of histone proteins in these bridges and the DNA topology. First, chromatin bridges are DAPI-positive DNA bridges and accessorized by nucleosomes, while other bridges are not stained by DAPI due to their stretched nature [144]. DAPI-negative chromatin bridges define a separate class of bridges termed ultrafine anaphase bridges (UFB) by Ian Hickson and colleagues [105, 106]. UFBs can arise from common fragile sites, centromeric regions, or telomeric regions that are difficult to replicate during the S-phase [106, 144]. Helicases such as PICH (ERCC6L) and BLM translocate to UFBs during the

anaphase, although the exact mechanism of activity of the two helicases at the UFBs is not known [145]. In line with the localization of BLM at UFBs, increased rates of UFBs were observed in cells with impaired BTRR complex function [105, 144]. Further, high rates of UFBs in the cells derived from patients with *TOP3A* and *RMI2* mutations imply that not only BLM, but the complete BTRR complex is important in the resolution of the UFBs [73, 74]. Anaphase bridges can eventually break and lead to lagging chromatin and micronuclei (MN) which are frequently observed in BS cells during S-phase [146, 147]. MN belong together with high SCE rates and chromatin breaks, to the cellular hallmarks initially observed in BS [141, 148].

BS cells further show elevated replication fork stalling and replication stress-induced abnormalities. Since one of the cellular roles of BLM is the management of the stalled replication forks as described in the previous sections, lack of BLM in BS cells result in replication fork aberrations such as longer replication fork elongation time, abnormal replication intermediates, and an overall increase in S-phase length [149–151]. Further, BS cells undergo apoptosis dependent on the DNA damage induced by the replication fork stalling as reported previously [152]. This finding was opposed by the observations that BS cells were escaping apoptosis probably due to alternative pathways to cope with replication fork stalling such as a recombination-dependent process, which could explain the increased SCE rates of BS cells [153].

The overall genomic instability caused by the impairment of DNA replication and repair mechanisms can explain the phenotypic characteristics observed in BS such as photosensitivity, immunodeficiency, and cancer predisposition. The slow proliferation rate of BS cells resulting from increased replication lengths could, in part, explain the growth retardation and the short stature. Further, neurogenesis and brain development are impaired by the abnormalities in DNA replication and repair processes as described in previous Section 1.5. Primary microcephaly, one of the most prominent clinical characteristics of BS, can be explained by the BTRR complex deficiency through its roles in DNA replication repair especially during neurogenesis, although the complete molecular link remains elusive. To define the mutational signatures in BS and gain further insights into the cellular roles of the BTRR complex, will help to decipher the path-

omechanisms underlying BS, and improve our understanding of fundamental cellular processes of DNA replication and DNA repair.

2 Materials and Methods

2.1 Materials

2.1.1 Antibodies

Antibodies used in Western blot experiments in this thesis are given in Table 2.1 with the used dilutions.

Table 2.1. Antibodies used in Western blot experiments.

Target protein	Company	Catalog number	Dilution
Beta actin	Cell Signalling	#4970S	1:1000
BLM	Abcam	ab2179	1:1000
BLM	Abcam	ab5446	1:2000
FANCD2	Novusbio	NB100-182SS	1:10,000
Goat-IgG	Santa Cruz	sc-2020	1:10,000
Histone H3	Abcam	ab8505	1:1000
Mouse-IgG	Santa Cruz	sc-2005	1:10,000
Rabbit-IgG	Santa Cruz	sc-2004	1:10,000
RMI1	Proteintech	14630-1-AP	1:1000
RMI2	Abcam	ab122685	1:1000
SMC2	Abcam	ab229353	1:2000
SMC4	Proteintech	24758-1-AP	1:2000
TOP3A	Proteintech	14525-1-AP	1:2000
Tubulin	Abcam	ab52866	1:5000
Vinculin	Sigma	V9264	1:10,000

Antibodies used in immunocytochemistry stainings are given in Table 2.2 with the used dilutions.

Table 2.2. Antibodies used for the immunocytochemistry assays.

Name	Company	Catalog number	Dilution
FANCD2	Novusbio	NB100-182SS	1:200
PICH	Abnova	H00054821-M01	1:150
SOX2	Miltenyi Biotec	130-104-941	1:50
OCT3/4	Miltenyi Biotec	130-120-236	1:50
NANOG	Thermo Fisher Scientific	MA1-017	1:100
TRA-1-60	Abcam	ab16288	1:200
SSEA-4	Miltenyi Biotec	130-098-371	1:30
Anti-mouse IgM Alexa Fluor® 488	Abcam	ab150121	1:200
Anti-rabbit IgG(H+L) Cy5	Thermo Fisher Scientific	A10523	2 µl/ml
Anti-mouse IgG(H+L) Alexa Fluor™ 488	Thermo Fisher Scientific	A11017	0.5 µl/ml

2.1.2 Cell lines

Fibroblasts used within this doctoral thesis are given in Table 2.3 with respective reference registries.

Table 2.3. Fibroblast samples used in this study.

Cell line name	Registry ID	Reference
Goe1304 (WT1)	GM08398	Coriell Institute for Medical Research, New Jersey, USA
Goe1305 (WT2)	GM00409	Coriell Institute for Medical Research, New Jersey, USA
Goe1306	GM02520	Coriell Institute for Medical Research, New Jersey, USA
Goe1308	GM02932	Coriell Institute for Medical Research, New Jersey, USA
Goe1309	GM02548	Coriell Institute for Medical Research, New Jersey, USA

Wild-type iPSC lines were kindly provided by Dr. Lukas Cyganek and Stem Cell Unit (SCU, University Medical Center Göttingen, UMG). Wild-type and CRISPR/Cas9 genome-edited knockout iPSC lines are given in Table 2.4. The main wild-type iPSC line used in this study was described previously [154].

Table 2.4. The iPSC lines used in this thesis.

Cell line name	Registry ID	Reference
isWT1.14	UMGi014-C clone 14	Stem Cell Unit, UMG, Göttingen
isWT7.21	UMGi020-B clone 21	Stem Cell Unit, UMG, Göttingen
isWT1-BLM-KO.1D1	UMGi014-C-19 clone 1D1	Stem Cell Unit, UMG, Göttingen
isWT1-TOP3A-KO.1E10	UMGi014-C-25 clone 1E10	Stem Cell Unit, UMG, Göttingen
isWT1-RMI1-KO.1F11	UMGi014-D-1 clone 1F11	Stem Cell Unit, UMG, Göttingen

2.1.3 Reference sequences

The reference sequences for the corresponding genes are given in Table 2.5.

Table 2.5. Reference sequences

Gene	RefSeq
BLM	NM_000057.4
TOP3A	NM_004618.5
RMI1	NM_001358291.2
RMI2	NM_152308.3

2.1.4 Synthetic oligonucleotides

Custom-designed oligonucleotide sequences were ordered from Eurofins Scientific (Luxembourg). Sequence length, annealing temperature, primer dimers, and GC contents were considered while designing the primers. Lisf of the primers used in this doctoral thesis and the sequences are given in Table 2.6.

Table 2.6. Primers sequences and their applications

Primer name	Sequence (5' - 3')	Application
BLM E3_for	CCTCCACCCTCCTCTGAAAG	Deep sequencing
BLM E3_rev	TCATCAGTACGTCCTGCCTC	

Primer name	Sequence (5' - 3')	Application
BLM E4-2_for	CACCAACTGTAAAGAAATCCC	iPSC genotyping
BLM E4-2_rev	TCATCAGTACTGCCTGCCTC	
BLM E8-1_for	CAAACCACTCCTTTAACATTCG	Fibroblast genotyping
BLM E8-1_rev	TCTTTCTAAGTCATTTATTGAAGCAG	
BLM E10_for	GCCCTGCCTGAGTTATGC	Fibroblast genotyping
BLM E10_rev	GGGTTTCTGGATGAAAGTTG	
BLM E15_for	AATGGCATTGCAAGTTATCAG	Fibroblast genotyping
BLM E15_rev	ACACATCTGTATCATTTGAGTTTTAC	
BLM Intr14-15_for1	TTCTCCACCTGCTCTCCATC	Fibroblast genotyping
BLM Intr14-15_for2	TGATACTGGGAGATGCACCTG	
BLM Intr15-16_rev	CGGATGAGACAAGAGAGGCT	
TOP3A E18_for	CTGGGTCCTCAAAGGCTCTG	Deep sequencing and
TOP3A E18_rev	CATCCTTCTGCACAGTCCGT	iPSC genotyping
RMI1 E3_for2	CTTTTCTTTGACTTGCAAAAATGG	Deep sequencing and
RMI1 E3_rev2	AGGGTGGAGAATACAAATCCATG	iPSC genotyping
RMI2 E3_for	CGCCGCCACTCAAGGTGCTG	Deep sequencing
RMI2 E3_rev	CGGGCATTACCTGGGACTAG	

GuideRNA oligonucleotides were designed using CRISPOR online tool and they were ordered from Integrated DNA Technologies (Coralville, USA). The sequences and the PAM motifs of the crRNAs targetting four genes of the BTRR complex are given in Table 2.7.

Table 2.7. guideRNA sequences for target genes of the BTRR complex

Target gene	Sequence 5' - 3' (PAM)
BLM	GATGTGATTTGCATCGATGA (TGG)
TOP3A	TGTGCTGCTCACTGTCCGTA (AGG)
RMI1	GGAATGGCTATCTGAACTGC (TGG)
RMI2	TTTGTAAAATCATGGGCTAA (CGG)
RMI1[*]	AGCGGGGACTTCTCGGTCCG (GGG)
RMI2[*]	AGGGTAGTGATGGCGGACCG (CGG)

[*]Alternative crRNA oligonucleotides.

2.1.5 Solutions and buffers

Cell lysis buffer		Final
TRIS/HCl (pH 7.5)	0.2 ml	0.02 M
NaCl	1.5 ml	0.15 M
EDTA	1 ml	1 mM
NP40	1 ml	1 %
SDS	0.25 ml	0.25%
dH_2O	5.8 ml	
Protease inhibitors (added just before use)	100 µl	1x

Buffer was stored at 4°C.

20x SSC		Final
NaCl	175.3 g	3.0 M
$Na_3C_6H_5O_7$ (Sodium citrate)	88.2 g	0.3 M
dH_2O	800 ml	
pH 7.0 with titration		1L

WB Transfer buffer	
5x Transfer buffer (Bio-Rad)	200 ml
100% Ethanol	200 ml
dH_2O	600 ml

Buffer was stored at 4°C.

TBST buffer	
Tween 20	1 ml
10x TBS	100 ml
dH_2O	899 ml

Fibroblast culture medium	
1x DMEM	500 ml
FBS	50 ml
P/S (10,000 U Pen/ml, 10,000 µg Strep/ml)	5 ml
Amphotericin B	1.4 ml

Fibroblast freezing medium	
FBS	9 ml
DMSO	1 ml

iPSC freezing medium	
StemMACS (or Stem Flex)	9 ml
DMSO	1 ml
10 mM TZV	20 µl

Agarose gel	
Agarose	1 g
1x TBE buffer	100 ml
GelRed	3.5 µl

2.1.6 Chemicals and expendable equipments

Chemicals and reagents used in this doctoral thesis are given in Table 2.8 .

Table 2.8. Chemicals used during the doctoral thesis

Chemical	Company
Alt-R® S.p. Cas9 Nuclease V3	Integrated DNA Technologies, Coralville, USA
Alt-R® Cas9 Electroporation Enhancer	Integrated DNA Technologies, Coralville, USA
Aphidicolin ready made solution	Thermo Fisher Scientific, Carlsbad, USA
BSA powder	Carl Roth Gmbh & Co. Kg, Karlsruhe, Germany
Colcemid	Merck Chemicals GmbH, Darmstadt,Germany
DMSO	Thermo Fisher Scientific, Carlsbad, USA
Fetal Bovine Serum	Thermo Fisher Scientific, Carlsbad, USA
GelRed nucleic acid stain	Thermo Fisher Scientific, Carlsbad, USA
GeneRuler 100 bp Ladder	Thermo Fisher Scientific, Carlsbad, USA
Giemsa stain, modified solution	Merck Chemicals GmbH, Darmstadt, Germany
Hoechst 33258	Thermo Fisher Scientific, Carlsbad, USA
Precision Plus Protein All Blue Standards	Bio-Rad Laboratories, Inc., Hercules, USA
ProLong™ Diamond Antifade Mountant	Thermo Fisher Scientific, Carlsbad, USA
Penicillin/Streptomycin	Thermo Fisher Scientific, Carlsbad, USA
Matrigel® (Growth Factor Reduced)	Corning Incorporated, Corning, USA

Chemical	Company
Milk powder, blotting grade	Carl Roth Gmbh & Co. Kg, Karlsruhe, Germany
Thiazovivin	Merck Chemicals GmbH, Darmstadt, Germany
Versene Solution (0.48 mM EDTA)	Thermo Fisher Scientific, Carlsbad, USA
VectaMount® Permanent Mounting medium	Vector Laboratories, Burlingame, USA

Expandable laboratory equipments were purchased from the following companies: Bio-Rad Laboratories Inc. (Hercules, USA), Eppendorf (Hamburg, Germany), Carl Roth Gmbh & Co. Kg (Karlsruhe, Germany), Greiner Bio-One GmbH (Frickenhausen, Germany), Lonza Group AG (Basel, Switzerland), Merck Chemicals GmbH (Darmstadt, Germany), Sarstedt AG & Co. KG (Nümbrecht, Germany), Starlab GmbH (Hamburg, Germany), Stemcell Technologies (Vancouver, Canada), Stratec Molecular GmbH (Berlin, Germany), Thermo Fisher Scientific (Carlsbad, USA).

2.1.7 Kits

Experiment kits used are given in Table 2.9 .

Table 2.9. Kits used for the corresponding applications.

Kit name	Company	Application
Amaxa P3 Primary Cell 4D-Nucleofector kit	Lonza, Basel, Switzerland	Nucleofection
BigDye™ Terminator v3.1 Cycle Sequencing kit	Thermo Fisher Scientific, Carlsbad, USA	Sanger sequencing
StemMACS™ iPS-Brew XF	Miltenyi Biotec, Bergisch Gladbach, Germany	iPSC culture medium
StemFlex™ Medium	Thermo Fisher Scientific, Carlsbad, USA	iPSC culture medium
QIAGEN DNeasy Blood & Tissue kit	QIAGEN, Hilden, Germany	DNA extraction

Kit name	Company	Application
QIAGEN Multiplex PCR Kit	QIAGEN, Hilden, Germany	PCR
MSB® Spin PCRapace	STRATEC Molecular, Birkenfeld, Germany	PCR product clean-up
Nextera XT DNA library preparation kit	Illumina, San Diego, USA	NGS library preparation
ReadyProbes Cell Viability Imaging kit	Thermo Fisher Scientific, Carlsbad, USA	ICELL8 well content assessment
Pierce BCA Protein Assay kit	Thermo Fisher Scientific, Carlsbad, USA	Protein concentration calculation
MicroBCA Protein Assay kit	Thermo Fisher Scientific, Carlsbad, USA	Protein concentration calculation
The Human Genome CGH Microarray kit 180K	Thermo Fisher Scientific, Carlsbad, USA	ArrayCGH sample preparation
WesternBright ECL HRPsubstrate	Advansta Inc., San Jose, USA	Western blot development
PCR Mycoplasma Test Kit I/C	PromoCell GmbH, Heidelberg, Germany	Mycoplasma test

2.1.8 Non-expendable equipment

The non-expandable equipment used are given in Table 2.10.

Table 2.10. Non-expandable equipments used in the doctoral thesis.

Equipment	Company
ABI PRISM® 3100 Genetic Analyzer	Advanced Biolab Service, Münich, Germany
C100 Touch Thermal Cycler	Bio-Rad Laboratories, Inc., Hercules, USA
CASY® cell counter	OLS OMNI Life Science GmbH & Co. KG, Bremen, Germany
CellenONE®	SCIENION, Berlin, Germany

Equipment	Company
Centrifuge 5418	Eppendorf, Hamburg, Germany
Centrifuge 5810R	Eppendorf, Hamburg, Germany
ChemiDoc™ Touch Imaging System	Bio-Rad Laboratories, Inc., Hercules, USA
FastGene FAS V	Nippon Genetics, Düren, Germany
Gel electrophoresis chambers	Thermo Fisher Scientific, Waltham, USA
HERA Cell 150 incubator	Thermo Fisher Scientific, Carlsbad, USA
HERA Safe safety cabinet	Thermo Fisher Scientific, Waltham, USA
Heraeus Fresco21 Centrifuge	Thermo Fisher Scientific, Waltham, USA
Heraeus Megafuge 16 Centrifuge	Thermo Fisher Scientific, Waltham, USA
ICELL8 Single-Cell System	Takara Bio, Shiga, Japan
Incucyte ®S3	Sartorius, Göttingen, Germany
Illumina HiSeq® 4000	Illumina, San Diego, USA
Lonza 4D-Nucleofector™	Lonza, Basel, Switzerland
Mitsubishi P96 printer	Mitsubishi Electric, Tokyo, Japan
Mr. Frosty™ Cryo Freezing Container	Thermo Fisher Scientific, Waltham, USA
NanoDrop One	Thermo Fisher Scientific, Waltham, USA
Primovert microscope	Carl Zeiss Microscopy GmbH, Oberkochen, Germany
Rebel	ECHO, San Diego, USA
Revolution (fluorescence microscope)	ECHO, San Diego, USA
Safe 2020 safety cabinet	Thermo Fisher Scientific, Carlsbad, USA
Stuart® Roller Mixer SRT 6D	Biocote, Staffordshire, UK
SureScan Dx G5761A Microarray Scanner	Agilent Technologies, Santa Clara, USA
ThermoMixer F1.5	Eppendorf, Hamburg, Germany
Trans-Blot Turbo Transfer System	Bio-Rad Laboratories Inc., Hercules, USA
UV irradiation system	Peqlab, VWR Life Science, Radnor, USA
Vortex-Genie 2	Scientific Industries Inc., Bohemia, USA

2.1.9 Software and online resources

Softwares used for various applications during this doctoral project are given in Table 2.11.

Table 2.11. Softwares used in the project and their application.

Software	Application
4Peaks (Nucleobytes)	Sanger sequence visualization
FinchTV Version 1.5.0	Sanger sequence visualization
GIMP Version 2.10.30	Image manipulation
Graphpad Prism 8.0 for macOS Version 8.3.0	Statistics and data graphics
IGV Version 2.8.9	Sequence analysis
Ikaros DNN (Metasystems)	Karyotype analysis
ImageJ (FIJI version 2.1.0/1.53c)	Image manipulation and processing
Inkscape 1.0	Image manipulation and production
LYX Version 2.3.6.2	Thesis text formatting
Microsoft Excel for Mac Version 16.60	Data processing
Microsoft PowerPoint for Mac Version 16.60	Image production and processing
Microsoft Word for Mac Version 16.60	Manuscript writing
R-Studio Version 21.09.1	Data analyses

Databases and web tools used within this project are given in Table 2.12.

Table 2.12. Databases and online web tools used during the thesis.

Name	URL
BioRender	https://biorender.com
CasAnalyzer	http://www.rgenome.net/cas-analyzer/#!
CRISP-ID	http://crispid.gbiomed.kuleuven.be
CRISPOR	http://crispor.tefor.net

Name	URL
Ensembl	https://www.ensembl.org/index.html
GeneCards	https://www.genecards.org
GnomAD	https://gnomad.broadinstitute.org
MutationTaster	https://www.mutationtaster.org
National Library of Medicine	https://pubmed.ncbi.nlm.nih.gov
NCBI	https://www.ncbi.nlm.nih.gov
OMIM	https://omim.org
KEGG	https://www.genome.jp/kegg/
Reactome	https://reactome.org
UCSC In-Silico PCR	https://genome.ucsc.edu/cgi-bin/hgPcr
WebGestalt	http://www.webgestalt.org
Institute of Human Genetics (UMG, Göttingen)	https://www.humangenetik-umg.de

2.2 Methods

2.2.1 Cell culture

Regular mycoplasma tests were performed using PCR Mycoplasma Test Kit I/C (PromoCell) according to manufacturer's instructions for each cell line used in this doctoral thesis.

2.2.1.1 Fibroblast cell culture

Cryopreserved cells were thawed on a water bath at 37°C as quick as possible. Outside surface of the cryovial was cleaned with 70% ethanol and the vial was taken inside of the safety cabinet. Thawed cell suspension was added to 10 ml cell culture medium on a falcon and cells were centrifuged at 1000 rpm for 5 minutes. Cells in the pellet were resuspended in 10 ml fibroblast culture media and placed in a 100 mm round dish. Next day, the plate was washed once with 10 ml Dulbecco's phosphate-buffered saline (DBPS) and the medium was changed. When fibroblast culture reached around 80% of confluency, a weekly split was performed with splitting ratio of 1:5 or 1:8. Briefly, the cells were washed once with DPBS and 2 ml 0.05% Trypsin-EDTA was added to the plate. After an incubation of 4 minutes at 37°C in the incubator, 8 ml of fresh medium was added to the plate and the cells were collected on a falcon tube. Cells were centrifuged at 1000 rpm for 5 minutes and the pellet was resuspended in 10 ml of fresh medium. Corresponding volumes for the correct splitting ratio were added to 10 ml medium on new 100mm round plates and cells were stored in the incubator of 37°C and supplemented with 5% CO_2. A medium change was performed every third day until the next split.

Cryopreservation of the cells was performed as the following. Cells were collected in a falcon tube as in the splitting procedure using trypsin. Cells were centrifuged at 1000 rpm for 5 minutes and the pellet was resuspended in 2-3 ml fibroblast freezing medium (90% FBS, 10% DMSO). Two or three cryovials were frozen from one confluent 100mm round dish and vials were placed in a Mr.Frosty box containing 100% isopropanol. The Mr.Frosty was stored in a -80°C freezer overnight and the cryovials were taken to a -150°C freezer for long term storage.

2.2.1.2 iPSC culture

Generally, iPSCs used in this thesis were cultured on 6-well plates coated with Matrigel. In brief, Matrigel was aliquoted according to the manufacturer's instructions and aliquots

were stored at -20°C. 0.25 ml Matrigel aliquots were resuspended in cold and sterile DPBS to make either 1:60 or 1:120 Matrigel dilution for plate coating. 1 ml of diluted Matrigel was added to each well and 6-well plates were either stored at 4°C for up to two weeks sealed with Parafilm or minimum of half an hour at 37°C before use.

Cryopreserved cells were thawed in a 37°C water bath as quick as possible. When there was still a small frozen fraction in the vial, the vial was cleaned with ethanol and taken into the sterile safety cabinet. Thawed cells were added to 5 ml iPSC culture medium (StemMACS) on a 15 ml falcon and centrifuged for 400 G for 7 minutes. After the centrifugation, cells in the pellet were resuspended carefully in 900 µl of medium containing 2 µM TZV and added to 2 ml medium with 2 µM TZV in one well of a 6-well plate coated with 1:60 Matrigel. Medium was changed on the next day without TZV and a daily medium change was performed until the cells reached 90% confluency.

Before the split of the cells, 6-well plates were coated with 1:120 Matrigel as described above. Medium on the cells was removed and 1 ml of Versene (0.48 M EDTA) was added to each well. The plate was incubated at room temperature for 5 minutes and Versene was aspirated carefully. Cells were washed off the well with 1000 µl micropipette in 1 ml medium containing 2 µM TZV with minimum resuspending. Cells were split to two wells containing medium with 2 µM TZV with a ratio of 1:8, 1:10, or 1:12 depending on the cell line. A medium change was performed on the next day without TZV. 2 µM TZV was used on each regular split to minimize the spontaneous differentiation of the iPSCs.

Cryopreservation of the iPSCs were performed as the following. Confluent cells on a 6-well plate were detached from the plate with Versene and resuspended in 1ml culture medium. Two cryovials were labeled with the cell line information and 500 µl of cell suspension was added to each cryovial. 500 µl of freezing medium containing 20% DMSO and 4µM TZV was added dropwise to the cells in the cryovials. The vials were stored in Mr.Frosty overnight at -80°C and transferred to -150°C for long term storage.

2.2.1.2.1 CRISPR/Cas9-based genome editing

CRISPR/Cas9-based genome editing experiments were performed in collaboration with Stem Cell Unit (UMG). Brightfield images of 70% confluent iPSC culture were taken before the transfection. Cell culture medium was changed to medium with 2 µM TZV half an hour before the start of the experiment. The crRNA oligos were designed with CRISPOR online tool for the four genes of the BTRR complex and sequences are given in Table 2.7.

CRISPR-RNA mix was prepared by mixing 5 µl crRNA and 5 µl tracrRNA in a PCR tube followed by an incubation at 95°C for 5 minutes. 20 µg of Cas9 nuclease was mixed with

6 µl of CRISPR-RNA mix from the previous step and incubated at room temperature for 20 minutes (RNP mix). Amaxa P3 Primary Cell 4D-Nucleofector kit (Lonza) was used for the transfection. In brief, nucleofector solution, nucleofector supplement and 1 µM Electroporation enhancer (IDT) were added to the RNPmix according to manufacturer's instructions and incubated at room temperature for 5 minutes. The resulting nucleofector mix was stored at 37°C while the cells were prepared. Next, cells were collected from one well in StemFlex with 2 µM TZV by the Versene detachment method. The number of cells was counted with Casy cell counter (OIS OMNI Life Science). As the negative control, 0.5 million cells were plated to one well of the 6-well plate covered with 1:60 Matrigel which was prepared earlier. For the transfection, 2 million of cells were centrifuged at 200 G for 3 minutes. The pellet was resuspended in the nucleofector mix carefully and the cells were loaded to the nucleoporation cuvette without any bubbles. The cells were nucleofected with the CA-137 program via the 4D-Nucleofector (Lonza). Next, 200 µl of StemFlex was added to the cells in the transfection cuvette with the dropper pipette. Using the same dropper, the cells were collected from the transfection cuvette and distributed dropwise to the remaining two wells of the 6-well plate. The plate containing the transfected cells and the negative control was placed in the incubator at 37°C and 5% CO_2. After two days, the medium was changed to StemFlex with 2 µM TZV. When the culture reached a 90% confluency, half of the cells were frozen with the above-described cryopreservation procedure for iPSCs and the other half was harvested for DNA extraction.

2.2.1.2.2 Singularization of iPSCs

Bulk transfected cells were thawed in StemFlex and split at least once before the singularization experiment. Conditioned medium (daily used medium from the cells) was collected 2–3 days before the singularization experiment and stored at 4°C until the day of the experiment. Two or three 96-well plates were coated with 1:60 Matrigel (60 µl/well) beforehand and kept in the fridge. The singularization medium was prepared by mixing fresh medium with the conditioned medium 1:1 and it was filtered by Steri-Flip (Merck Millipore). The singularization medium was supplemented with 1x Revita (100x stock solution) and 1x P/S (100x stock solution). 100 µl of prepared medium was distributed to each well of the 96-well plate (100 µl/well) with the multipipette and the plate was stored at 4°C until cells were ready. 70% confluent cell culture was treated with 1x Revita for 20 minutes in the incubator. The cells were washed once with DPBS and 1 ml TrypLE was added. Cells were incubated for 5 minutes at 37°C. TrypLE was aspirated and cells were collected off the well with 1 ml singularization medium. Cells were immediately placed on ice and the cell number was counted with CASY. 100,000 cells were resuspended in degassed 1 ml cold DPBS. Cells were singularized using CellenONE (ScienION) and 96-well plates containing cells were placed in the incuba-

tor. Approximately 4 days after singularization, 100 µl fresh medium containing P/S was added to the wells. Approximately one week after the singularization, complete medium was changed to fresh medium containing P/S. Ten days after singularization images of each well were taken using Incucyte and the medium was changed to StemFlex without antibiotics. Single cell colonies were passaged by Versene method to 12-well plates containing 1 ml of medium with 2 µM TZV approximately two weeks after singularization date. When cells reached 80% confluency, ⅘ of cells from each well were cryopreserved while ⅕ cells were collected for the DNA analysis. DNA was extracted using the DNeasy kit (QIAGEN). The region of interest was PCR-amplified and Sanger sequenced. Genotype analyses of the clones were done using CRISP-ID online tool.

2.2.2 Analyses of nucleic acids

2.2.2.1 DNA extraction

A fraction of cell suspension (at least 100 µl) was set apart in an eppendorf tube during cell split or cell freezing for the purpose of DNA extraction. Cell suspension was centrifuged for maximum speed for 1 minute. DNA was extracted from the pellet using the DNeasy Blood and Tissue kit (QIAGEN).

2.2.2.2 Polymerase chain reaction

PCR reactions were performed for amplifying the region of interest of the gDNA samples. Alternatively, QIAGEN Multiplex PCR kit was used according to manufacturer's instructions for the difficult-to-amplify regions. Regular PCR reactions and QIAGEN multiplex were prepared with the following chemicals at the indicated concentrations.

Generic PCR reaction

Solution	Volume
10x Taq Buffer	2.5 µl
$MgCl_2$ (25 mM)	1.5 µl
dNTPs (2.5 mM each)	2 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
Taq Polymerase (5 U/µl)	0.15 µl
DNA (50 ng/µl)	1 µl
dH_2O	15.85 µl

QIAGEN Multiplex PCR reaction

Solution	Volume
5x MM master mix	5 µl
DNA (50 ng/µl)	1 µl
dH_2O	19 µl

Touchdown-like PCR program (TOUCH58) was used for the amplification of the DNA regions of interest for 58°C annealing temperature reactions. Alternatively, TOUCH58Q program was used when the QIAGEN PCR kit was used for the difficult-to-amplify products. Detailed steps of the reactions are as the following.

TOUCH58

Temperature	Time	
94°C	3 min	
94°C	15 sec	
62°C	20 sec	x 3
72°C	30 sec	
94°C	15 sec	
60°C	20 sec	x 3
72°C	30 sec	
94°C	15 sec	
58°C	20 sec	x 32
72°C	30 sec	
72°C	10 min	
10°C	5 min	

TOUCH58Q

Temperature	Time	
94°C	15 min	
94°C	30 sec	
62°C	30 sec	x 3
72°C	1 min	
94°C	30 sec	
60°C	30 sec	x 3
72°C	1 min	
94°C	30 sec	
58°C	30 sec	x 32
72°C	1 min	
72°C	10 min	
10°C	30 min	

2.2.2.3 Agarose gel electrophoresis

PCR fragments were evaluated for correct base pair size via agarose gel electrophoresis. 0.5% TBE was used for preparing 2% agarose gel and GelRed nucleic acid gel stain was used for the visualization of the bands. Samples were mixed with 6x gel loading dye (Thermo Fischer Scientific) and gel was run at 100 V until the bands of the ladder separated enough. Gene Ruler 100 bp Plus (Thermo Fisher Scientific) was used for the control sizes. Images of the gels were taken with FastGene FAS V (Nippon Genetics) and printed with Mitsubishi P96 printer (Mitsubishi Electric).

2.2.2.4 Sanger Sequencing

PCR products of approximately 400 bp length were cleaned from primers with Exo/Sap enzymes by incubating at 37°C for 20 minutes followed by an incubation at 80°C for 15 minutes according to the manufacturer's instructions. Reactions for Sanger sequencing were prepared with BigDye™ Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific) according to manufacturer's instructions. Sanger sequencing PCR program was prepared as the following steps.

Sanger Sequencing PCR		
Temperature	Time	
96°C	30 sec	
96°C	10 sec	
55°C	5 sec	x 39
60°C	4 min	
12°C	10 min	

Final Sanger PCR products were diluted 1:1 with water to make a final volume of 20 µl. Samples were sequenced at the diagnostics department of the Institute of Human Genetics (University Medical Center Göttingen) with ABI PRISM® 3100 Genetic Analyzer (Advanced Biolab Service). The electropherograms of the sequences were visualized with 4Peaks or FinchTV softwares.

2.2.2.5 Deep amplicon sequencing

Deep amplicon sequencing was used for the determination of the CRISPR/Cas9 transfection efficiencies on the iPSC lines. Primers generating approximately 350 bp long PCR products flanking the Cas9 cut site were designed for each gene of the BTRR complex. The sequences of the primers are given in Table 2.6. After regular PCR reaction the band sizes were confirmed on agarose gel. The PCR product was cleaned by using Msb Spin PCRapase (Stratec Molecular) kit and the NGS library was generated by using SureSelect QXT (Agilent Technologies) library preparation kit. The generated library was sequenced on an Illumina NextSeq550 sequencer at the diagnostics department of the Institute of Human Genetics. The results were evaluated by the Cas-Analyzer online tool.

2.2.2.6 ArrayCGH

Genomic DNA from each KO-iPSC sample was extracted from one 6-well containing approximately 2 million cells using the DNeasy extraction kit (QIAGEN). The oligo-array CGH assay used for this doctoral thesis was a high-resolution genomic hybridization which compared the KO-iPSC sample DNA to pooled, sex-specific DNA from a control group (Human Genomic DNA Male or Female; Promega Corporation, USA). The Human Genome CGH Microarray Kit 180K (Agilent Technologies) was used for the analysis. Data were acquired using the SureScan Dx G5761A Microarray Scanner (Agilent) and extracted using Agilent Feature Extraction Software (Agilent). The data determined were analyzed using the Agilent Cytogenomics software at the diagnostics department of the Institute of Human Genetics Göttingen.

2.2.3 Cytogenetic analyses

2.2.3.1 Karyotyping

Cells were grown on a T25 flask for karyotyping. Flasks with 75% confluency were treated with 1 µM of colcemid during 3 hours. Cell culture medium containing colcemid was collected on a 15 ml falcon and cells were washed once with DPBS. The DPBS of washing was collected in the same falcon containing the medium. Cells were collected from the flask using Trypsin according to regular cell culture methods and added to the saved medium. Cells were centrifuged for 1000 rpm for 10 minutes and the pellet was resuspended in 500 µl of the supernatant via vortexing. 5 ml of warm (37°C) 0.075 M KCl solution was added dropwise by the help of glass pipette to the resuspended cells while vortexing. Cell solution was incubated at room temperature for 13 minutes, then centrifuged for 10 minutes at 1000 G. Cell pellet was resuspended in 1 ml supernatant solution while vortexing. Next, 3 ml of ice cold methanol and acetic acid (3:1 vol/vol) containing fixative solution was added dropwise to the solution while rotating the tube in hand. Another 3 ml of the fixative was added while vortexing for two times. Cells were incubated at room temperature in fixative solution for 10 minutes and centrifuged for 10 minutes at 1000 G. Pellet was resuspended in 0.5 ml fixative solution from the previous step via vortexing and 5 ml of cold fixative was added. Cell suspension was centrifuged and last step of fixation was repeated. On the last step, cell pellet was resuspended in 1 ml of fixative solution by glass pipette and dropped to cold microscope slide stored in water. Microscope slides were aged over the weekend with the help of hot plate and chromosomes were stained by GTG banding according to regular cytogenetic protocol of routine diagnostics at the Institute of Human Genetics (UMG).

2.2.3.2 Sister chromatid exchange assay

iPSCs were cultured with 10 µM BrdU during 42 hours (two replication cycles) before harvesting. On the last 3 hours, 1 µM colcemid was added to the existing medium and cell harvest was performed according to the above-described procedures for the karyotyping. After slide preparation from the same protocol of karyotyping, slides containing the metaphases were dried for two days at room temperature in dark. Next, the slides were rehydrated in DPBS for 5 seconds followed by an incubation in Hoechst 33258 dye (1 µg/ml in water) for 15 minutes in the dark. Cells were rinsed in water quickly and placed horizontally to a chamber filled with 2xSSC (pH 7) buffer which was prepared freshly. Slides were irradiated with UV-C for 0.260J/cm^2 (approximately 45 seconds). Followingly, the irradiated slides were incubated in 2xSSC buffer in a Coplin jar in a 65°C water bath for two hours. Next, the slides were removed from the water

bath and placed in another Coplin jar containing 5% Giemsa solution diluted in water for 5 minutes at room temperature. Slides were rinsed once in water and dried at room temperature before microscopy. Vectamount was used for the mounting of the slides.

2.2.4 Analyses of gene expression

2.2.4.1 Single-cell RNA sequencing

Cell harvesting for single-cell transcriptome sequencing of fibroblast samples was performed with Trypsin as in regular cell splitting. After the centrifugation, cells were resuspended (100,000 cells in 1 ml) in DPBS and brought to the NIG Core Unit (Institute of Human Genetics, UMG) on ice.

Cell harvesting of the iPSCs was performed with TrypLE. In brief, cells were washed once with DPBS and 1 ml TrypLE was added to each well. After 5 minutes of incubation at 37°C cells in TrypLE were collected and added to 5 ml medium on a falcon. Cell number was counted with a hemocytometer and cells were centrifuged at 400 G for 5 minutes.

At the NIG Core Unit cells were stained with DAPI and Propidium Iodide (ReadyProbes Cell Viability Imaging Kit, Thermo Fisher) in accordance with the manufacturer's instructions. Stainings was confirmed with fluorescent microscope and and the cell numbers were counted using CASY cell counter. Serial dilutions aiming for 1 cell/50 nl concentration were performed before single cell sorting. Single cells were sorted to a nanowell chip using ICELL8 Single-Cell System (Takara Bio). Cell lysis was done at –80°C, followed by cDNA synthesis which was done by SMART-Seq Single Cell kit (Takara Bio) in the nanowells. The library was prepared with Nextera XT DNA library preparation kit (Illumina) and sequenced on an Illumina HiSeq4000 [155].

2.2.4.2 FACS analysis

Pluripotency markers OCT3/4 and TRA-1-60 were analyzed in CRISPR/Cas9 genome edited KO-iPSC lines. 4 wells of 6-wells were harvested for each cell line for FACS. Cells were washed once with Versene and detached from the wells with Versene method. Cells were collected on a 15 ml falcon in DPBS and centrifuged at 500 G for 5 minutes. Supernantant was carefully aspirated and cell pellet was resuspended in 2 ml of 4% PFA. Cells were vortexed shortly and incubated at room temperature for 15 minutes. Next, cells were centrifuged at 500 G for 5 minutes. The PFA fixative solution was collected on a separate waste. Cell pellet was resuspended in 5 ml of DPBS and

centrifuged again as the previous steps. DPBS wash was repeated one more time. The pellet was resuspended in 1 ml of 2% BSA in DPBS and vortexed. Cell solution was strained using a 40 µm cell strainer. Cells were brought to Stem Cell Unit on ice for futher steps of the FACS experiment. Briefly cells were stained for the two pluripotency markers and DAPI and sorted at the FACS facility of UMG. The FACS data was analyzed using Flowing software.

2.2.5 Analyses of proteins

2.2.5.1 Protein extraction

One confluent well of a 6-well plate of iPSC culture was harvested for protein extraction. One or more 100 mm round dishes were harvested for protein extraction from fibroblast samples. For both of the cell samples, cells in culture were washed once with cold DPBS and were collected with a sterile cell scraper in an appropriate amount (around 80 µl) of cell lysis buffer which contained freshly added protease inhibitors (100x stock solution). Cell lysate was collected, placed in an eppendorf tube, and incubated on ice for half an hour. Benzonase (1 µl for 100 µl) was added to shear the DNA and samples were stored at -20°C overnight. Next day, samples were centrifuged at 4°C at maximum speed for 15 minutes. The supernatant containing the proteins was transferred to a clean eppendorf and protein concentration was calculated with BCA assay (Thermo Fisher Scientific) using cuvettes and the spectrophotometer function of the Nanodrop (Thermo Fisher Scientific).

2.2.5.2 Western blot analysis

Initially, 15-20 µg proteins were mixed with 4x NuPAGE® lDS Sample Buffer and 10x NuPAGE® Reducing Agent (Thermo Fisher Scientific) and incubated at 95°C for 5 minutes. Degraded protein samples were loaded to Mini-Protean TGX Stain-Free 4-20% SDS PAGE pre-prepared gels (Bio-Rad Laboratories). 5 µl of the Precision Plus Protein All Blue Standards (Bio-Rad Laboratories) was loaded as the molecular weight marker. The electrophoresis chamber was filled with 1x TGS buffer. Electrophoresis took place at 80 V for the first 15 minutes followed by 120 V until the dye front leaked to the buffer chamber. SDS-PAGE gel was removed from the chamber and transfer of the proteins to PDVF membrane was done via Trans-Blot Turbo Transfer system (Bio-Rad Laboratories) using standard molecular we program for 30 minutes. In brief, pre-cut PVDF membrane (Bio-Rad) was activated in 100% methanol for 20 seconds, rinsed in water

and equilibrated in cold transfer buffer containing 5x Transfer buffer (Bio-Rad Laboratories) and 2x 100% ethanol in water. The membrane was blocked in 5% milk in TBST for an hour. Primary antibody solutions were prepared in 2% milk in TBST. Blocked membranes were incubated with the primary antibody solution on a roller mixer at 4 °C overnight. The next day membranes were washed with 2% milk in TBST for 10 minutes three times. Secondary antibody solution was prepared in 2% milk and added to the falcons containing the membranes. The blots were incubated with secondary antibodies for one hour at room temperature on the roller mixer. Blots were washed with TBST for 10 minutes three times followed by washing with DPBS for 10 minutes two times. Membranes were developed using WesternBright ECL HRP substrate (Advansta Inc.) and visualized with ChemiDoc™ Touch Imaging System (Bio-Rad Laboratories). Quantification of the band intensities were done by using FIJI program.

2.2.5.3 Mass spectrometry

For the MS experiment, biological triplicates were prepared and approximately 1 million cells for each sample were harvested. In brief, fibroblasts were detached from 100 mm dishes with trypsin, washed with DPBS and the cell pellets were snap frozen in liquid nitrogen. Samples were transported on dry ice to TU Kaiserslautern, to the Department of Molecular Genetics to the lab of Markus Räschle where the rest of the MS experiment took place. First, pellets were thawed on ice and residues of DPBS were removed. The pellets were resuspended in appropriate amounts of lysis buffer (6 M guanidinium chloride, 10 mM TRIS(2-carboxyethyl)phosphine (TCEP), 40 mM chloracetamide, and 100 mM TRIS pH 8.5) according to the amount of cells. Samples were incubated at 98°C for 10 minutes on a shaker. Next, samples were sonicated in a water bath for 5 minutes and centrifuged at 4°C for 10 minutes at 1300 rpm. Protein amounts were calculated using Micro BCA Protein-Assay kit (Thermo Fisher Scientific) according to the manufacturer's instructions. For each sample, 25 µg of protein was diluted 1:10 with digestion buffer (10% acetonitrile, 25 mM TRIS pH 8.5). Trypsin and lysC enzymes were added with 1:50 ratio (enzyme mass/protein mass). Samples were incubated at 37°C on a heat block covered with foil for 16 hours (overnight). Next, ruptured peptides were acidified with 1% TFA for a final pH 2.0. Samples were cleaned using SDB-RPS filter which was activated with 100% acetone and equilibrated with 30% methanol, 1% TFA and 0.2% TFA. Eluted samples on MS-grade PCR tubes were vacuum dried and stored in -20°C before further continuation with the experiment. Peptides were resolubilized in 0.1% formic acid and loaded to the chromatography machine. In brief, peptides were separated by nano-high pressure liquid chromatography on an Easy nlC 1200 chromatography system. Peptides were directly sprayed into a Q Exactive HF mass spectrometer [156]. The obtained MS data were processed using MaXQuant Version and Perseus software.

2.2.6 Immunocytochemistry assays

For the immunocytochemistry experiments of iPSCs, 13 mm round coverslips of 1N thickness were washed in 100% acetone, 100% ethanol, 70% ethanol in water, each for two minutes. The coverslips were air-dried and sterilized before use. Two coverslips were placed to one well of 6-well plate and the well was coated with 1:120 Matrigel. After the Matrigel solidified, cells were split to the wells as explained above. Alternatively, cells were collected from the wells after the Versene incubation with cell scraper for larger colonies and split with 1:5 and 1:6 ratio for fixing on the second day after the split. Once colonies reached a sufficient size on the coverslips, cells were washed once with DPBS and fixed with 4% formaldehyde containing 0.1% Triton-X100 for 20 minutes. Formaldehyde solution was collected on a separate waste and the cells were washed two times with DPBS.

2.2.6.1 Pluripotency markers of iPSCs

For the pluripotency markers, cells were blocked with 1% BSA for 15 minutes after fixation. Primary antibody solutions were prepared in 1% BSA and 20 µl were dropped onto the coverslips in the wells. Cover slips in the wells were covered with Parafilm and the plates were stored at 4°C in a humidifying chamber over night. The next day, primary antibody solutions were washed three times with DPBS. Secondary antibody solutions were prepared in 1% BSA and the coverslips in the plate were incubated with secondary antibody solutions at room temperature for one hour in dark. Cover slips were washed three times with DPBS and incubated with 0.06 µM DAPI solution dissolved in water for 10 minutes in dark. Next, cover slips were washed three times with DPBS and twice with distilled water. 5 µl Prolong Diamond mounting solution was dropped onto clean microscope slides and coverslips were mounted by the help of tweezers. Coverslips were visualized with fluorescent microscope within four weeks after the mounting.

2.2.6.2 UFBs and mitotic errors

For the analysis of the mitotic errors, cells were fixed as described above. After washing PFA with DPBS, 2 ml of 0.15 M NH_4Cl was added to each well and incubated for 20 minutes at room temperature. Next, cells were washed three times with DBPS. 2 ml of 2% BSA in DPBS was added to each well for 10 minutes. The BSA solution was removed carefully with minimum movement of the plate and 2 ml 2% BSA was added again to the wells. After another 10 minutes of incubation, BSA was removed from the well. 40 µl of primary antibody solution in DPBS was prepared and 20 µl was

dropped onto each coverslip. Coverslips were covered with Parafilm and the plate was stored overnight at 4°C in a humidifying chamber. On the next day, wells containing the coverslips were washed with 2% BSA three times and incubated with secondary antibodies diluted in 2% BSA blocking solution for 45 minutes at room temperature in dark. Coverslips were washed twice with 1x DPBS. Next, coverslips were taken out of the wells with tweezers and dipped into 100% ethanol. Coverslips were placed to a clean well containing 43.75 µg/ml DAPI and 10 µg/ml RNAse solution for one minute. DAPI staining solution was removed and wells were washed once with DPBS and once with dH_2O . 0.01% NaN_3 was added to the wells. Coverslips were mounted with 5 µl Diamond mount (Thermo Fisher Scientific) into clean and dry microscopy slides. Three biological replicates were prepared for the analysis of UFBs and mitotic errors.

2.2.7 Statistics

Significance assessment of the scRNAseq data was multiple times corrected with Benjamini Hochberg correction. SCE assays, UFB counts, and chromosome segregation errors were evaluated with Microsoft Excel and GraphPad Prism 8.0 programs. For the statistics of SCE assay, unpaired nonparameteric Mann-Whitney test was used. The chromosome segregation error rates and significances were evaluated with unpaired two-tailed t-tests. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

3 Results

3.1 Bloom syndrome patient cohort

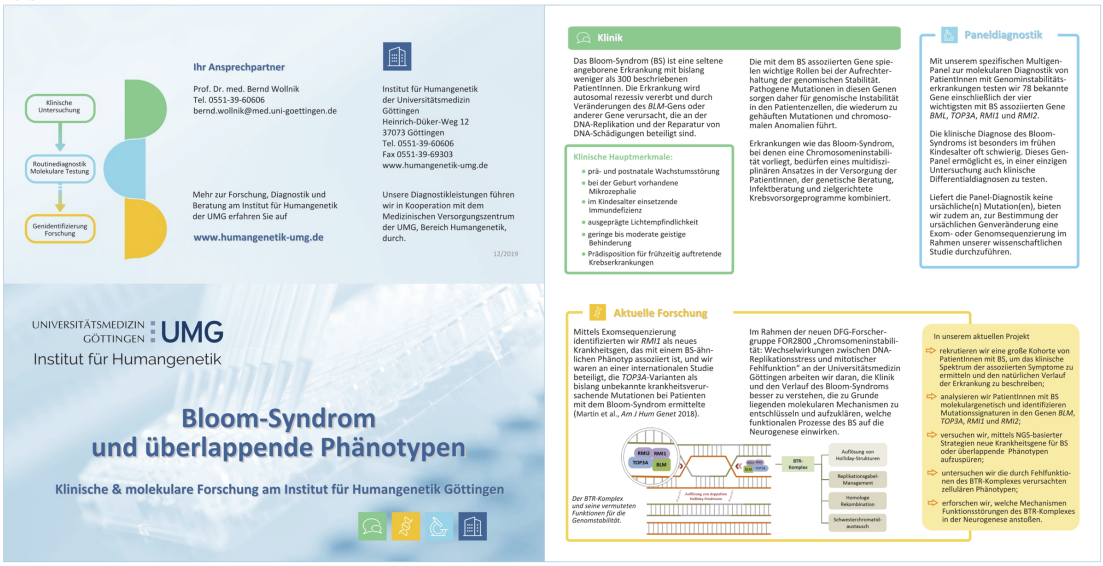
In rare diseases, systematic characterization of phenotypes and underlying genetic mutations can often be challenging due to the limited number of patients. Although the first cases of BS were reported in the 1950s, currently only around 300 patients have been reported in the literature including the most comprehensive BS patient cohort, the Bloom Syndrome Registry (BSR) with 164 reported patients [54, 157]. Autosomal recessive mutations in the *BLM* gene were initially identified in the BS patients. In the last years, the mutational signatures of BS expanded by the identification of novel pathogenic variants in genes other than BLM thanks to the advances in sequencing techniques such as WES. Homozygous or biallelic loss-of-function (LoF) mutations in *TOP3A, RMI1*, and *RMI2* were identified in patients presenting with BS-like phenotypes. In addition, it was observed that clinical findings of the BS patients varied in individuals harboring pathogenic variants not only in different genes, but also within the same gene. Hence, expanding the patient cohort of individuals presenting a BS/BS-like phenotype is especially important for the determination of clinical and mutational spectra of this rare disease.

3.1.1 "Bloom syndrome and overlapping phenotypes": The BS flyer

In the beginning of my study, I aimed to collect additional patients with a BS/BS-like phenotype to broaden the definitions of the clinical and mutational signatures of BS. For

this reason, a Bloom syndrome flyer was prepared within this project which was sent out to pediatric and medical genetics centers in Europe, especially in Germany and Turkey. This flyer contained a description of the phenotypic features of BS, the molecular pathomechanisms leading to BS, and molecular genetic diagnostic testing opportunities available at the Institute of Human Genetics of the University Medical Center Göttingen (UMG). These diagnostic tests include specific candidate gene testing, next-generation sequencing (NGS) based tests such as the microcephaly multigene panel analysis as well as WES approaches. Additionally, the flyer included information on the ongoing BS research projects at the Institute of Human Genetics. This flyer was prepared in English, German, and Turkish (Figure 3.1). Copies of the German and Turkish versions were printed and distributed to clinicians, human genetics institutes, and various genetics centers in both countries. The English version of the flyer was available for download from the homepage of the Institute of Human Genetics, UMG (https://www.humangenetik-umg.de), and was sent out via email to different international clinicians and researchers who are working on BS and/or rare genetic disorders.

A



B

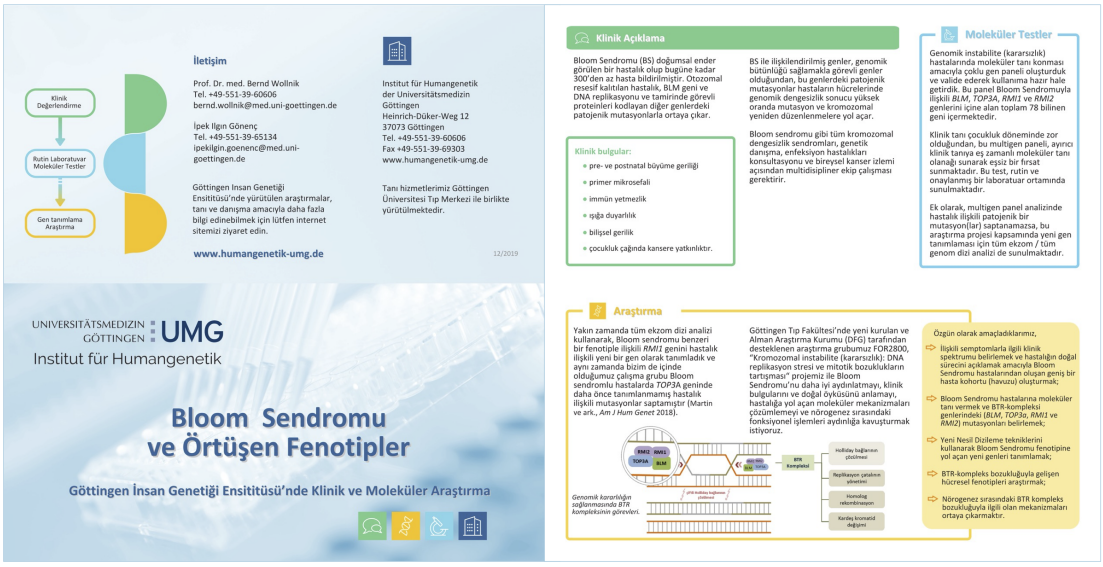


Figure 3.1. "Bloom syndrome and the overlapping phenotypes" flyer. The Bloom syndrome flyer informing about clinical characteristics and the molecular mechanisms of BS was prepared within the first year of the doctoral thesis. The project details (right side) and the contact information (left side) were also included. The flyer was printed in (A) German and in (B) Turkish languages. The printed flyers were distributed to various medical genetics centers, clinicians, and Institutes of Human Genetics in both Germany and Turkey.

3.1.2 Assessment of clinical findings of BS patients

During the last years, several individuals with BS have been diagnosed using molecular genetics approaches at the Institute of Human Genetics, UMG. In total, eight patients with BS were clinically characterized by Prof. Dr. Nursel Elcioglu at Marmara University (Istanbul, Turkey) and sent to Göttingen for genetic testing. Molecular diagnoses were performed in Göttingen using DNA extracted from blood samples of patients and their parents. Pathogenic variants in *BLM* were identified in six of these patients in homozygous state using either direct testing of the *BLM* gene (3 cases), NGS-based multigene panel analysis (1 case), or WES analysis (2 cases). Of the 3 in total identified different pathogenic variants, one was a novel variant in *BLM*. This variant, c.572_573delGA, was identified in homozygous state and predicted to induce a frameshift resulting in a premature protein truncation, p.Arg191Lysfs*4. The premature stop codon was located closer to the N terminal end of BLM helicase whose enzymatic domains are positioned afterwards. Therefore the conserved domains of BLM would lack in the case of translation of the allele containing c.572_573delGA (p.Arg191Lysfs*4) variant, thereby would most likely lead to loss of BLM function. Furthermore, a homozygous LoF mutation in *RMI1* was identified in two individuals of a consanguineous Turkish family presenting with a BS-like phenotype [74]. This variant was identified by RG Wollnik prior to my thesis.

In the course of this thesis, I analyzed and compared the clinical features of these eight patients and compared them to previously published BS patients (Table 3.1). All eight patients showed primary microcephaly with OFC varying between -3.1 SD and -7.6 SD. Intrauterine growth delay was present in 7/8 of the pregnancies in terms of mostly low birth weight (5/8), while length at birth was also affected in one patient among the patients whose data were available. Growth deficiencies continued after birth for all patients with an average height of -3.7 SD below the mean of healthy individuals of the same age, sex, and ethnicity. The average z-score for weight was -3.85 SD below the mean observed in the healthy population with a z-score between -0.9 SD and -5.5 SD within our patient cohort. Further characteristics of BS such as facial features, derma-

tological findings, and immunological properties slightly varied among the BS patients compared within this thesis. The most striking difference was observed between patients harboring homozygous *BLM* mutations and patients harboring homozygous *RMI1* mutations. Neither of the patients with *RMI1* mutations showed dermatological anomalies or immunodeficiency which are commonly observed hallmarks in BS. In general, *RMI1*-associated patients revealed a milder phenotype which was similar to the mild BS-like phenotype observed in *RMI2*-associated patients reported previously [73]. Furthermore, all *BLM*-associated patients showed immunodeficiency by low amounts of at least one class of tested immunoglobulins, i.e., IgG, IgA, IgM, and IgG1. In line with this, most of the *BLM*-associated patients were more prone to pulmonary infections, although infections were not recurrent for some of the patients (Table 3.1). Additionally, one patient was operated at the age of 5 for Wilms tumor, which was observed in BS patients at a higher frequency than in non-affected children [62]. Details of the clinical and genetic findings of all BS patients is summarized in Table 3.1, and the comparison of the clinical and genetic details of *BLM*- and *RMI1*-associated patients was part of a scientific publication which was published in the course of this thesis [158].

Table 3.1. Clinical characteristics of molecularly diagnosed BS patients. Table is adapted from Gönenc *et al.* [158].

	BLM					***RMI1***		
Patient ID	F1	F2	F3	F4	F5-1	F5-2	F6-1	F6-2
Age at exam	16y 3m	16y	3y 10m	13y 6m	8y 2m	1y 7m	7y	13y
Primary microcephaly	+	+	+	+	+	+	+	+
Intrauterine growth delay	+	+	+	+	n/a	+	+	+
Length at birth z-score	-4.3	n/a	n/a	N	N	N	n/a	n/a
Weight at birth z-score	-5	N	n/a	-3.3	-2.7	N	-1.7	-3.1
OFC at birth z-score	n/a	n/a	n/a	n/a	n/a	-3.2	n/a	n/a
Short stature	+	+	+	+	+	+	+	+
Postnatal growth delay	+	+	+	+	+	+	+	+
Height z-score	-3.9	-4.6	-4.2	-6	-1.8	-2.3	-4	-2.9
Weight z-score	-5.5	-4.7	-3.2	-4.9	-3.6	-4.2	-3.8	-0.9
OFC z-score	-6	-5.1	-4.7	-7.6	-3.1	-5.1	-5	-5
Narrow triangular face	–	+	–	+	+	–	–	–
Sunlight sensitivity	n/a	+	+	n/a	n/a	n/a	n/a	n/a
Skin lesions	–	+	+	–	+	+	–	–
Cafe-au-lait spots	–	+	+	+	+	+	–	–
Immunodeficiency	+	+	+	+	+	+	–	–
IgG	N	low	low	low	n/a	low	n/a	n/a
IgA	low	low	low	n/a	low	low	n/a	n/a
IgM	low	low	low	n/a	low	low	n/a	n/a
IgG1	n/a	N	low	n/a	n/a	N	n/a	n/a
Recurrent infections	–	+	–	–	–	+	–	–
Upper airway/ pulmonery infections	+	+	–	+	+	+	–	–

Abbreviations are as follows: y, years; m, months; n/a, not available; and N, normal.

3.2 Characterization of the cellular phenotypes of BS patient-derived fibroblasts

The BLM helicase is an important component of the cell due to its roles in the maintenance of genomic integrity. Within my doctoral thesis, three different BS patient-derived fibroblasts were analyzed to define gene and protein expression profiles, and the effects of BLM deficiency on genomic integrity.

3.2.1 Confirmation of pathogenic *BLM* variants of the BS fibroblasts

First, the pathogenic *BLM* variants in BS patient-derived fibroblasts, which were provided by the Coriell Institute (New Jersey, USA), were validated. Genomic DNA was isolated from all three different fibroblast lines and *BLM* regions/exons, in which the pathogenic variants were supposed to be located, were amplified by PCR and subjected to subsequent Sanger sequencing. For each of the BS patient-derived fibroblasts, we were able to confirm the expected biallelic homozygous or compound heterozygous pathogenic variants.

The first of the BS patient-derived fibroblasts (Goe1306) was compound heterozygous for two different truncating mutations in the *BLM* gene (Figure 3.2A). We confirmed a one base pair deletion in exon 15, c.2923delC (p.Gln975Lysfs*24), and additionally in trans, a large, 1584 base pair (bp) spanning deletion including the complete exon 15 and surrounding intronic regions of intron 14 and 15. Due to the absence of exon 15 resulting from the larger deletion on the other allele, the one base pair deletion c.2923del was observed in hemizygous state (Figure 3.2B, upper part). The second allele contained a 1584 bp long deletion spanning complete exon 15. The breakpoint of this deletion was detected by PCR and Sanger sequencing using specific forward and reverse primers within intron 14 and intron 15, respectively. On protein level, both detected variants were predicted to result inducing a frameshift and premature protein truncation, p.Gln975Lysfs*24 and p.Val942fs, respectively, most likely resulting in com-

plete loss of BLM function (Figure 3.2B). The predictions were made using in silico tools such as MutationTaster, PolyPhen, or SIFT tools.

The second BS patient-derived fibroblast (Goe1308) contained a complex indel of a 6 bp deletion and a 7 bp insertion in the exon 10, c.2207_2212delinsTAGATTC in homozygous state. This variant was predicted to induce a frameshift and also lead to a premature stop codon (p.Tyr736Leufs*5, Figure 3.2C). This specific mutation is one of the most frequent pathogenic *BLM* variants detected in patients with BS with an allele frequency of 0.0001779 in the healthy population as given by the gnomAD variant database (https://gnomad.broadinstitute.org/). It has a particular high prevalence in the Ashkenazi Jewish population where its frequency reaches 0.003772 (gnomAD). Therefore, this pathogenic *BLM* variant, which was reported by Ellis et al. and Li et al. in 1998, was originally referred to as "blmAsh" allele [159, 160]. The founder effect of this mutation was confirmed by linkage disequilibrium studies and Ellis et al. could confirm that this variant was a founder mutation within the Ashkenazi Jewish population [69].

The third BS patient-derived fibroblast sample (Goe1309) was compound heterozygous for two different truncating mutations in two separate *BLM* exons. Sanger sequencing validated a heterozygous one base pair duplication, c.1544dupA (p.Asn515Lysfs*2) located in exon 7. Additionally, I confirmed a nonsense mutation, c.2254C>T (p.Gln752*), in exon 10 of *BLM*, which was also present in the heterozygous state (Figure 3.2D). All of the mutations detected in patient fibroblasts were truncating mutations, which were predicted to result in an elimination of approximately one-third of BLM protein, most likely leading to complete loss of protein function. The missing regions at the C-terminal end of BLM include domains that are crucial for the enzyme activity of BLM such as ATPase domain and the zinc-binding domain; therefore, all variants were classified as pathogenic LoF mutations.

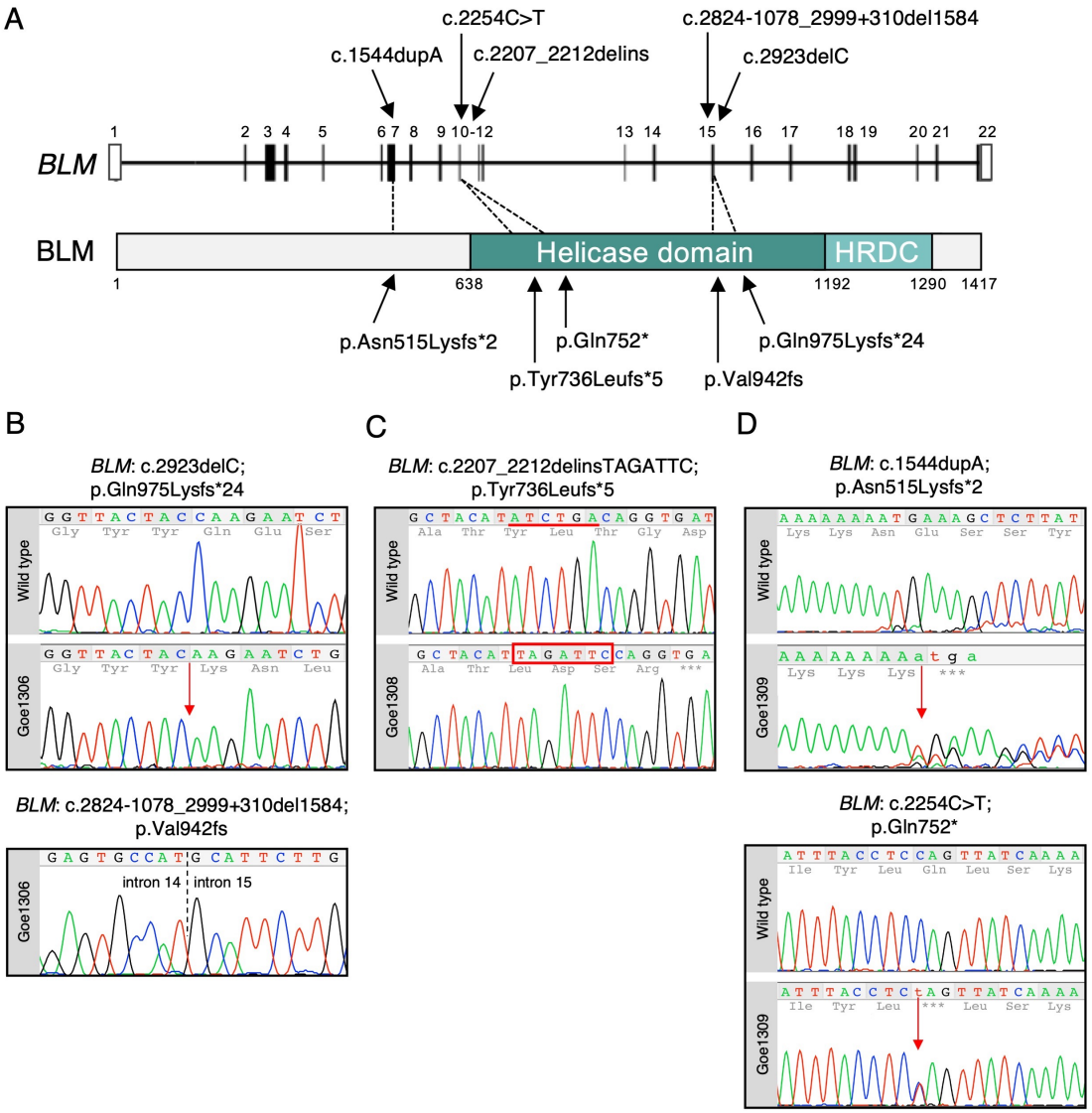


Figure 3.2. Genetic confirmation of the BS patient-derived fibroblast cell lines. (A) The locations of the *BLM* mutations of the cell lines are shown with respect to schematic representation of *BLM* gene (upper part) and BLM helicase (lower part). Exons are shown as vertical black lines, introns as horizontal black lines, and the non-coding regions of the exons as open boxes. Electropherograms of *BLM* mutations are shown: (B) Goe1306 cell line was compound heterozygous. c.2923delC mutation was observed as hemizygous due to the deletion of complete exon 15 on the other allele. The breakpoint of the 1584 bp is marked with dashed line at the bottom lane. Specific primers placed at intron 14 and intron 15 were used for explicit amplification of the allele containing the large deletion. (C) Goe1308 cell line was homozygous for the red marked mutation of the 6 bp deletion and 7 bp insertion. (D) Goe1309 cell line was compound heterozygous for the LoF mutations. Figure is adapted from Gönenc *et al.* [155].

After the confirmation of the different *BLM* mutations, the chromosome stability of the

three BS patient-derived samples were assessed. Karyotype analysis of Goe1309 revealed a normal 46,XY karyotype in 7/7 analyzed metaphases. In the second BS sample, Goe1308, a mosaicism for loss of the Y chromosome was detected. In three out of eight analyzed cells, the karyotype was 45,X. Next, the Goe1306 sample harbored chromosomal rearrangements in all five analyzed cells. The resulting karyotype in these cells was designated as 46,XX,i(9)(q10),t(9;15)(p10;q10) (Figure 3.3). In short, the q arm of chromosome 9 was duplicated and a translocation occurred between the duplicated 9q arm and the 15q arm. However, the exact positions of the duplication and translocation were not determined. The mosaic aneuploidy and chromosomal aberrations observed on BS cells could be a clonal effect due to passaging of the cells or signatures of genomic instability observed in BLM dysfunction in BS.

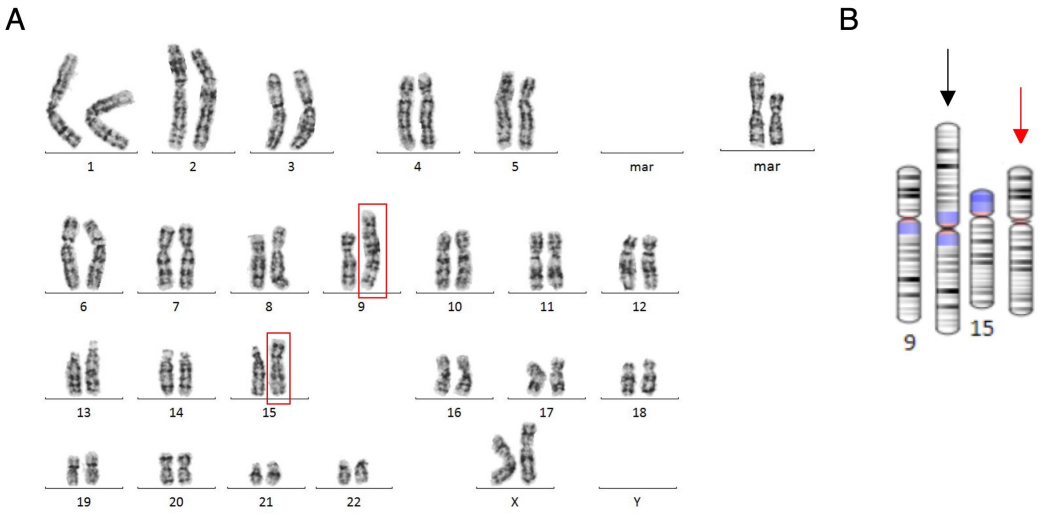


Figure 3.3. Metaphase karyotype of the BS patient-derived fibroblast sample Goe1306. (A) Representative karyotype analysis revealed a possible 46,XX,i(9)(q10),t(9;15)(p10;q10) karyotype. Derived chromosomes are marked in red and additionally shown as separate marker chromosomes. (B) Scheme of the derivative chromosomes showing duplication of 9q (black arrow) and the translocation between 9p and 15q (red arrow). In total, five metaphases were analyzed according to standard procedures established for cytogenetic diagnostics.

3.2.2 Defining gene expression profiles of BS cells

Despite the well-defined molecular functions of BLM, transcriptional profiles in BS have not been studied extensively in the literature. Interestingly, previous studies showed

that BLM specifically regulates expression of genes that contain G4 motifs in close proximity (approximately 30 bp) to the border between the first exon and intron [161]. To gain further insights into the effects of BLM deficiency on gene expression and to specifically define gene networks that are influenced by BLM, we performed single-cell transcriptome analysis in BS patient-derived fibroblasts together with aged-matched controls. Single-cell RNA sequencing allowed us to define and correct for cell cycle differences as well as to identify minor gene expression differences which would not be possible to achieve by bulk RNA sequencing. Further, transcriptional changes were validated and the effect of these changes on protein expression levels was analyzed using Western blot and proteomics analysis.

3.2.2.1 Single-cell transcriptome sequencing

To evaluate the transcriptional changes of BLM deficiency on a single-cell level, we performed NGS-based single-cell transcriptome sequencing (scRNAseq). The experimental setup for scRNAseq included three BS patient-derived fibroblast samples (Goe1306, Goe1308, and Goe1309) and two aged-matched fibroblast controls (WT1 and WT2). Additionally, two BS patient-derived samples cells (Goe1306_APH and Goe1309_APH) and one wild-type control (WT1_APH) were treated with a mild concentration of aphidicolin to induce replication stress. Aphidicolin inhibits DNA polymerase, which leads to replication stress through replication fork stalling during DNA synthesis processes. It is expected that cells harboring impaired replication fork management are more sensitive to replication stress inducers which might lead to an accumulation of aberrations in the DNA in these cells [162, 163]. Chromosome missegregation induced by replication stress was found to be linked to deregulated microtubule dynamics [164]. Cells were treated with 300 nM aphidicolin or DMSO (non-treated samples) for 24 hours, harvested, followed by single-cell sorting, library preparation, and sequencing which was carried out at the NGS Integrative Genomics (NIG) Core Unit (Institute of Human Genetics, University Medical Center Göttingen). Samples were stained with DAPI and propidium iodide (PI) prior to sorting and single-cell sorting was performed on a mi-

crochip containing 5,184 wells using the ICELL8 single-cell system (Takara Bio) [20]. All eight samples were dispensed to the same microchip and simultaneously processed in the same experimental run. The content of each well was assessed by detection of DAPI and PI signals using the image processing software from Takara Bio. Thanks to the imaging system of ICELL8 platform, only wells containing single cells were selected for further processing which was an advantageous feature when compared to the widely used scRNAseq platforms such as 10x Genomics Chromium platform. By the imaging feature of ICELL8, cells in slots showing one DAPI signal and no signal for PI were selected for library preparation and sequencing of single cells (Figure 3.4). Next, cDNA synthesis and library preparation were performed on each microwell that was selected for further processing and the generated libraries were sequenced on an Illumina HiSeq4000.

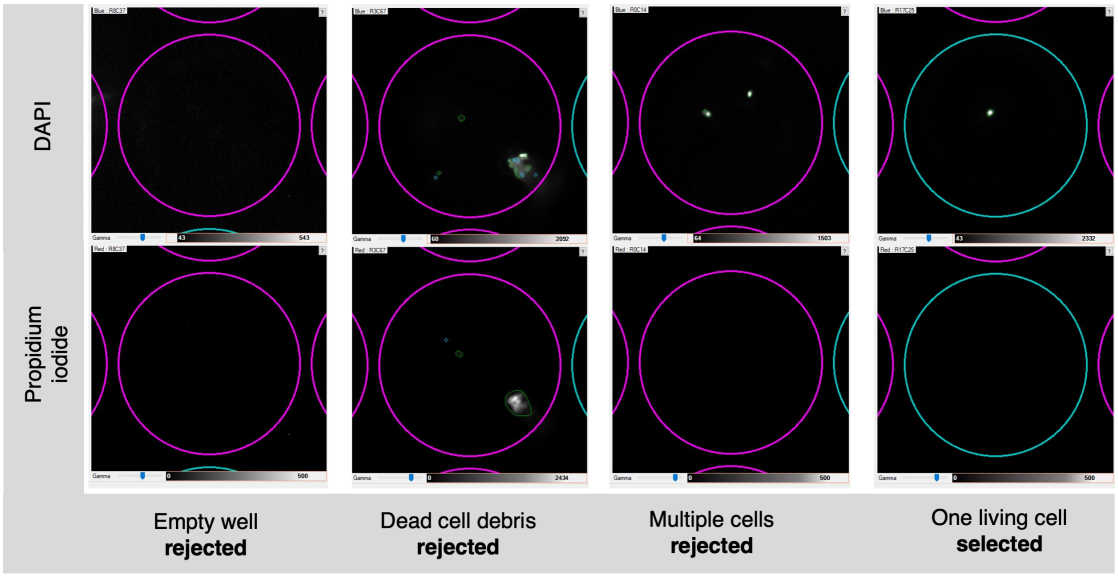


Figure 3.4. Well content assessment of the nanowell chip by fluorescent imaging. Cells were stained with DAPI (upper panel) and propidium iodide (lower panel), and dispensed to separate wells of the 5184-well containing microchip by the ICELL8 single cell system. Eight different samples were sorted to the same chip. Nanowell content was evaluated by the fluorescence signals from both dyes, DAPI and propidium iodide, respectively. The DAPI signal was used to determine the number of cells in each well, whereas the propidium iodide signal revealed the living state of each cell. Wells containing single, living cells (DAPI positive and PI negative) were selected for further processing.

3.2.2.1.1 Quality control assessments of scRNAseq data

The raw results were obtained as fastq files and processed for the quality control (QC) of the sequencing data. CogentAP (Takara Bio) was used as QC analysis pipeline, and out of 1,054 cells that were sequenced, 851 in total cells passed the quality control evaluations. Mapping of all barcoded reads to the reference genome revealed that the amount of total exonic reads was more than 70% (Figure 3.5). After the removal of the cells with less than 10,000 total reads or less than 300 genes sequenced, for all samples the mean number of reads per cell was approximately 250,000 (Figure 3.6A, upper part). Similarly, following the removal of genes with less than 100 total reads or expressed in fewer than three cells, for all samples the total gene number per cell was between 7,000 and 8,000 which was within an optimal value range (Figure 3.6A, lower part). The ribosomal RNA percentage as well as the percentage of the reads originating from mitochondrial DNA were below our quality control threshold (<10% of all uniquely, mapped reads) (Figure 3.6B). The overall quality of the data was robust and suitable for further analyses.

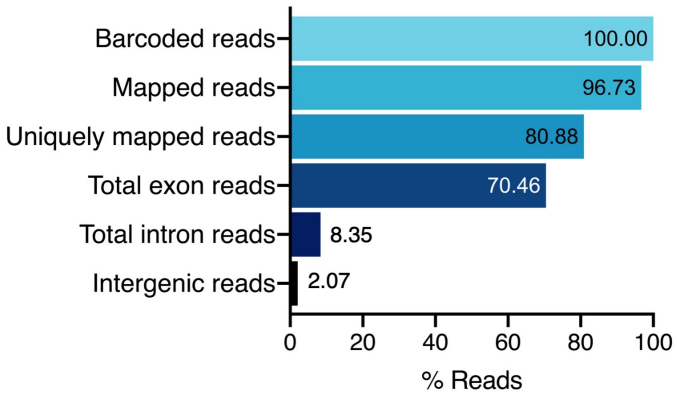


Figure 3.5. Classification of reads generated by scRNA sequencing of BS patient-derived and control fibroblasts. Evaluation of the barcoded reads revealed that after quality filtering, more than 70% of all barcoded reads mapped to exonic regions. Approximately 10% of all barcoded reads mapped to intronic or intergenic regions and therefore, were excluded.

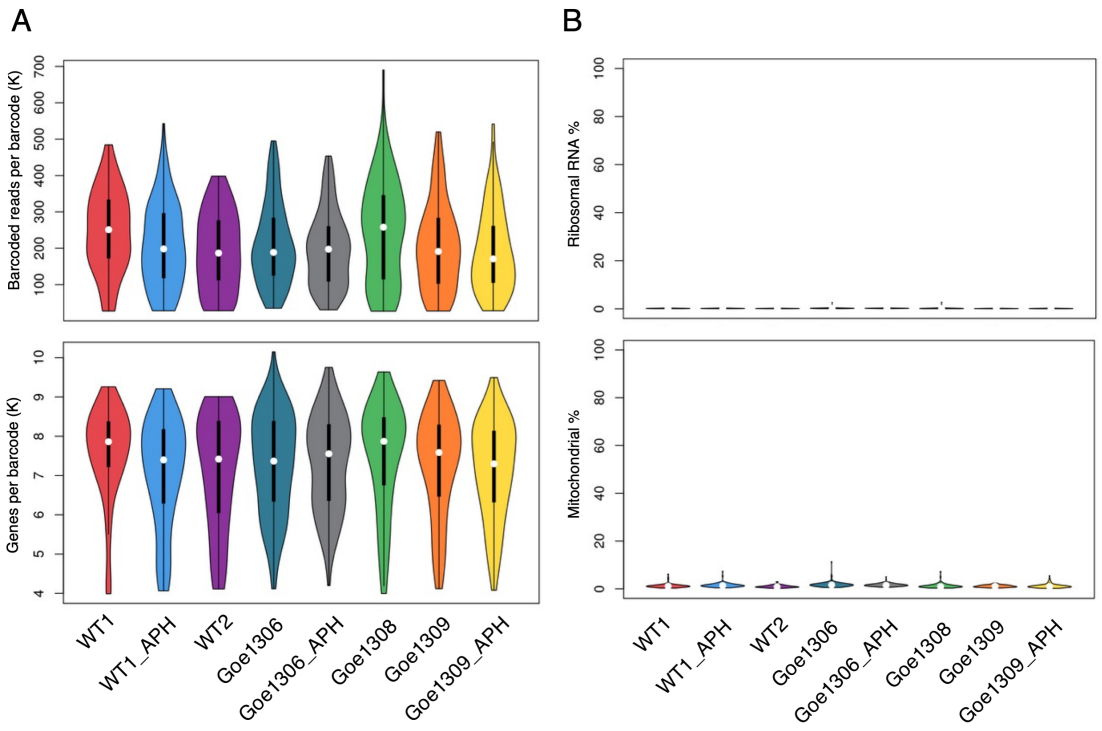


Figure 3.6. Quality control assessment of scRNAseq data. (A) After removing cells with <10,000 total reads or <300 genes, the average number of reads per barcode was around 250,000 for all eight samples (upper graph). Overall, expression of approximately 7.500 genes per cell was detected after removing genes with <100 total reads or genes expressed in fewer than three cells (lower graph). (B) Analyses of reads mapping to ribosomal RNA (upper panel) and mitochondrial RNA (lower panel) showed almost no reads indicative of ribosomal RNA, and less than 10% of reads mapping to mitochondrial RNA indicating robust scRNAseq data quality. Plots were generated using CogentAP (Cogent NGS Analysis Pipeline, Takara Bio) and adapted from Gönenc *et al.* [155].

Next, the gene expression profiles were assessed by dimension reduction analysis by UMAP [165]. The UMAP analysis showed a distinct clustering between cells originating from different samples (Figure 3.7). We also observed that cells from each sample were clustering together at close positions in the UMAP map, which indicated similar gene expression profiles within cells of the same sample. The aphidicolin-treated samples were at close positions to their non-treated sample pair suggesting only a mild effect of aphidicolin on transcription-level changes in the treated cells.

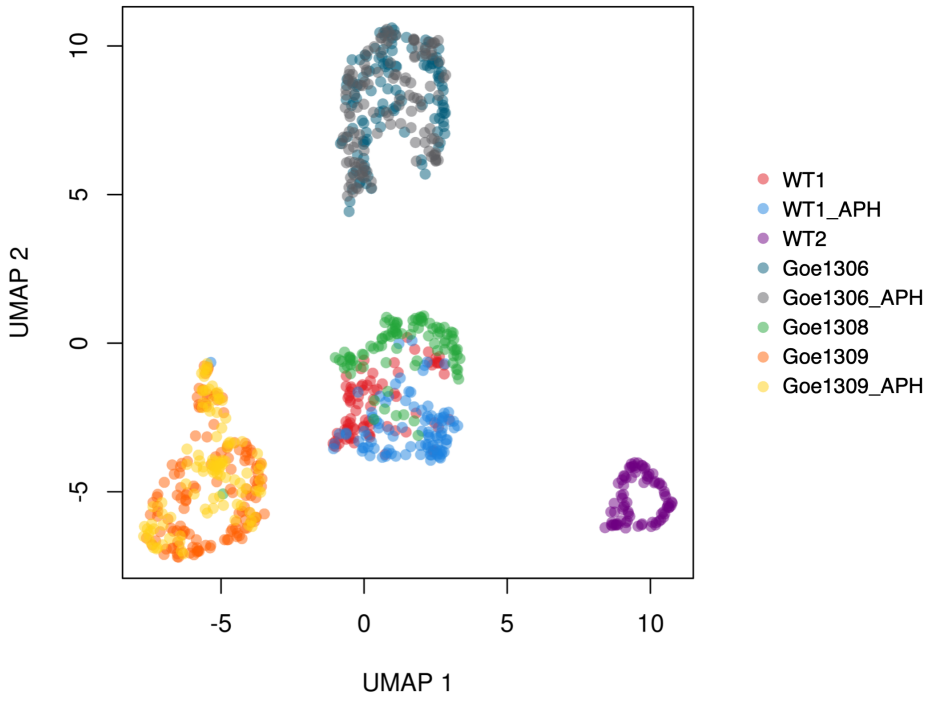


Figure 3.7. UMAP analysis showing gene expression profiles of each different fibroblast sample. Wild type, BS patient-derived (Goe1306, Goe1308, and Goe1309), aphidicolin treated wild type, and aphidicolin treated BS fibroblast cells (Goe1306_APH and Goe1309_APH) were included in the scRNAseq experiment. Cells were treated with 300nM aphidicolin for 24 hours or left untreated as control. UMAP analysis was performed using CogentAP (Cogent NGS Analysis Pipeline, Takara Bio).

3.2.2.1.2 Differentially expressed genes in BS single cells

As the quality of the scRNAseq data was robust, further transcriptional analyses were performed. First, we assessed the cell cycle stages of each samples by the Seurat vignette which is an established bionformatics pipeline used for the assessment of the cell cycle stages [166]. This approach assigns specific scores to each cell dependent on the gene expression levels of marker transcripts for cell cycle stages. According to this analysis we did not observe any cell cycle stage differences among BS and WT samples (data not shown). Next, the gene expression levels in BS single cells were tested against the WT control cells as the main part of the analyses of scRNAseq data. After multiple test corrections (Benjamini-Hochberg correction, false discovery rate [FDR] <0.05), 2,562 genes out of 30,269 mapped transcripts were evaluated as sig-

nificantly deregulated in BS patient-derived single cells (Figure 3.8). Interestingly, the BLM transcript did not show significant changes, which could be due to gene dropout events of scRNAseq [167]. Among the significantly deregulated genes, the highest fold change values did not exceed 1.5 and were mostly around 1. This observation also applied to genes that were downregulated in BS patient-derived single cells, which mostly showed fold change values around -1. The aphidicolin treatment did not result in high amounts of significantly deregulated genes, although the number of significantly deregulated genes in treated BS patient-derived single cells compared to non-treated BS patient-derived single cells was higher than the number of the significantly deregulated genes in aphidicolin treated WT cells in comparison to non-treated WT cells. This could imply that the BS patient-derived cells were more sensitive to replication stress than WT cells. The expression levels of the other members of the BTRR complex, i.e., *TOP3A, RMI1*, and *RMI2*, were not significantly altered indicating that the expression levels of members of the BTRR complex were not directly regulated and affected by the BLM helicase's presence or its abundance in cells. Detailed expression values of the members of the BTRR complex are given in Supplementary Table S1.

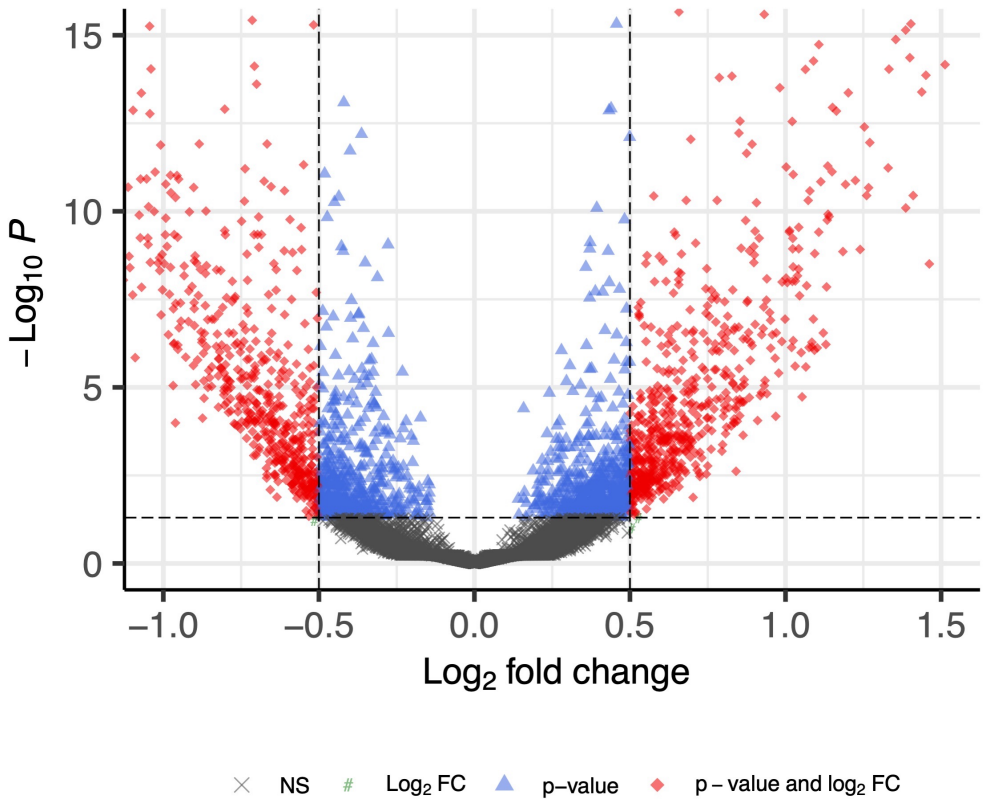


Figure 3.8. Differentially expressed genes in BS single cells. Volcano plot shows the DEGs in BS cells when compared to WT single cells. Plot was generated using using EnhancedVolcano package from Bioconductor for R-programming language and adapted from Gönenc *et al.* [155].

3.2.2.1.3 Determination of deregulated molecular pathways in BS single cells

Next, we took a detailed look at significantly deregulated genes in BS patient-derived single cells in comparison to wild type cells, and we performed pathway analyses by over-representation analysis (ORA). The first pathway analysis was done using gene ontology (GO) terms for biological processes via WebGestalt [168]. The pathways were listed according to the false discovery rate (FDR) value which was set to 0.05 (Figure 3.9A). Among the top most significant denominations of GO terms, we detected terms which were strongly related to the known BS pathogenesis. 'Mitotic sister chromatid segregation' was the most significant term which was followed by 'mitosis' and further cell cycle-related terminologies. This was in line with the cellular effects observed in BLM-deficient cells. The analysis of deregulated pathways using the KEGG database

revealed a similar result. The highly significant pathways, which were associated with BS and BLM deficiency, were pathways such as 'DNA replication', 'p53 signaling pathway', and the 'Fanconi anemia (FA)' pathway (Figure 3.9B). Taken together, pathway analysis of the differentially expressed genes in BS patient-derived single cells presented a detailed view to the molecular aberrations in BS cells and reflected the cellular phenotype associated with BLM deficiency.

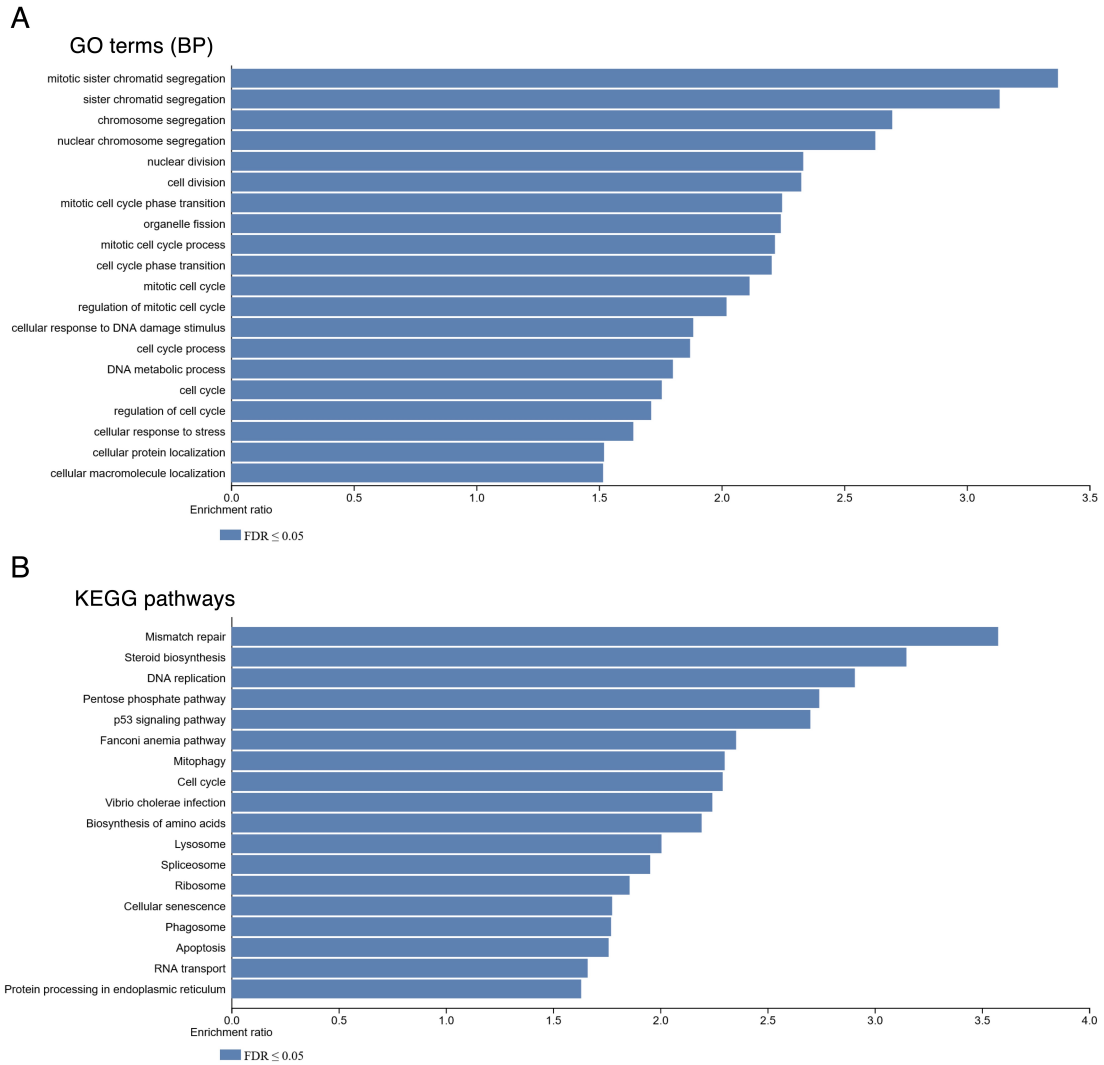


Figure 3.9. Pathway analyses of differentially expressed genes in BS single cells compared to wild-type single cells. (A) Over-representation analysis (ORA) of the significantly deregulated genes in BS single cells using Gene Ontology (GO) enrichment analysis revealed deregulated pathways associated with BS and BLM deficiency in the top 20 most significant denominations. (B) ORA using the KEGG database pathway showed significant deregulation of the pathways related to molecular processes associated with BLM function and BS pathogenesis. Multiple test corrections were performed using the Benjamini-Hochberg procedure (false discovery rate ≤ 0.05) and plots were generated using WebGestalt.

3.2.2.1.4 Upregulated genes from Fanconi anemia pathway

Biallelic pathogenic variants in numerous genes of the Fanconi anemia (FA) pathway are associated to the FA disorder, clinically characterized by microcephaly, short stature, immunodeficiency, and prominent cancer predisposition [169]. Importantly, BS

and FA have several overlapping phenotypic hallmarks as such, due to common molecular pathways affected in both disorders. BLM plays an important role during replication fork management which is mediated by its interactions with proteins of the FA pathway. Therefore, we analyzed gene expression values of known BLM interaction partners specifically within the FA pathway. Interestingly, we found that most of the genes belonging to the FA pathway were significantly upregulated in the BS patient-derived single cells (Figure 3.10). The most significantly overexpressed gene was *RPA3* followed by *FANCD2* and its interaction partner *FANCI*. The interactions of BLM and RMI1 with the single-stranded DNA binding protein RPA through the BTRR complex have been well documented [131, 170]. FANCD2 is especially important for BLM regulation as they interact during replication fork management [140]. Furthermore, FANCD2 foci are observed at the origins of the ultrafine anaphase bridges [106] that occur in BS cells at a higher frequency [74]. The values of differentially expressed genes of the FA pathway in BS patient-derived cells are shown in Table S1. Overall, upregulation of several genes of the FA pathway in BS cells provided detailed insights to BLM dysfunction through the known interaction partners of the BLM helicase.

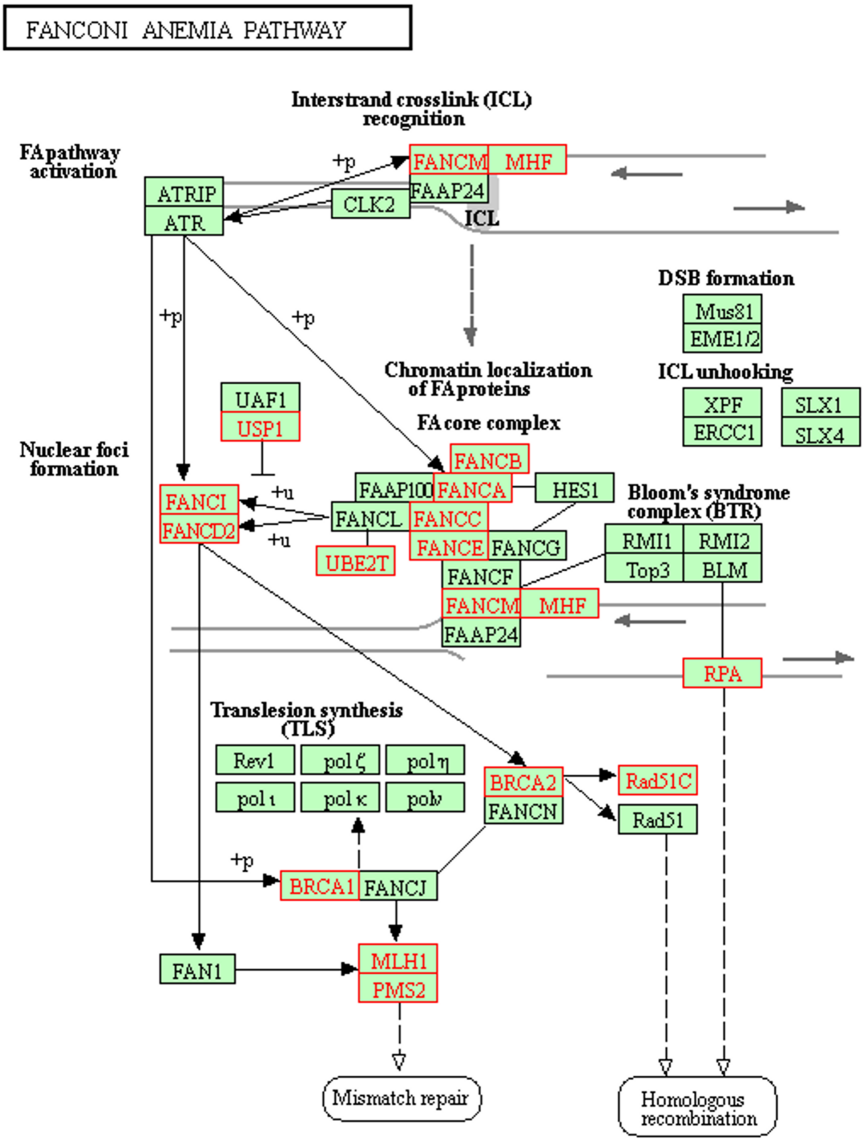


Figure 3.10. KEGG-pathway analysis of the differentially expressed genes of Fanconi anemia (FA) pathway. Several significantly deregulated genes in BS single-cell RNAseq analysis are associated with the FA pathway (indicated in red). Genes of the FA pathway that were not significantly deregulated are marked in green. The FA pathway is rendered from the KEGG database and visualized using WebGestalt. +p, phosphorylation; +u, ubiquitination.

3.2.2.1.5 Distinct expression of genes associated with primary microcephaly

Primary microcephaly is one of the most prominent clinical characteristics of BS. Therefore, we investigated the expression levels of the genes in which pathogenic variants have been previously associated with autosomal recessive forms of primary micro-

cephaly [38]. Indeed, we detected 12 genes significantly deregulated among the 27 microcephaly-associated genes (Figure 3.11). Well-described genes such as *CDK5RAP2, CENPE, SASS6*, and *ASPM* were among these deregulated genes. Detailed values for all analyzed genes are given in Supplementary Table S1. Interestingly, most of the significantly deregulated genes of this group were overexpressed in BS patient-derived single cells. Of note, the detected significant genes from the group had roles in mitotic processes or DNA replication pathways, which overall could provide a link to how BLM deficiency causes microcephaly in BS.

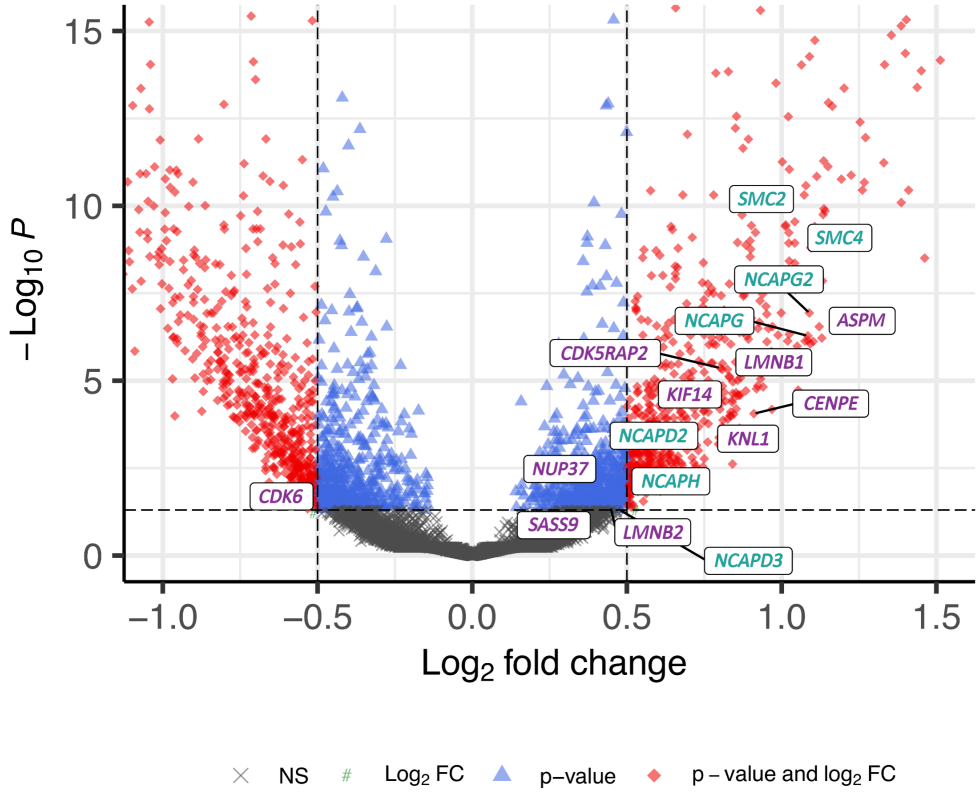


Figure 3.11. Volcano plot analysis of gene expression in BS single cells visualizing genes of the condensin I/II complexes and microcephaly-associated genes. Genes encoding members of the condensin I/II complexes are shown in turquoise, genes clinically associated with microcephaly are shown in purple. Volcano plot was generated using the EnhancedVolcano package from Bioconductor for R-programming language and adapted from Gönenc *et al.* [155].

3.2.2.1.6 Highly significant overexpression of condensin I/II complex genes in BS patient-derived single cells

Chromosome condensation is an important process throughout the cell cycle and it is regulated mainly by condensin I and II complexes. Condensin I/II complexes are composed of two SMC components (SMC2 and SMC4) and additional non-SMC proteins, e.g., NCAPG, NCAPG2, NCAPH, NCAPH2, NCAPD2, and NCAPD3, which interact with each other forming the two complexes. Both complexes are able to bind DNA and have differential contributions to the establishment and maintenance of mitotic chromosome condensation [171]. Failure in appropriate chromosome condensation results in bulky genetic material, which can lead to genomic instability like in the form of chromatin bridges [172].

Interestingly, autosomal recessive mutations in the members of the condensin I and II complexes such as *NCAPH*, *NCAPD2*, *NCAPD3*, or *NCAPG2* have been associated with primary microcephaly [173, 174]. Of note, phenotypes caused by mutations in *SMC2* and *SMC4* are not yet described. The overexpression of these three genes was already observed in our data when we analyzed the microcephaly-related gene expression. Next, we evaluated the expression levels of all condensin I/II complex members and we observed overexpression of all condensin I/II complex members except for one. Alterations in gene expression levels of each member of the condensin I/II complex member and their significance values are summarized in Table 3.2. The fold changes of the members were around 1 which indicates a doubling in gene expression in BS patient-derived single cells compared to WT cells. Especially, the *SMC2* and *SMC4* genes were among the most highly upregulated genes of all condensin I/II complex members (Figure 3.12), which is particularly interesting as, so far, no direct link between BLM and condensin I/II complex expression has been described.

Table 3.2. Differential expressions of condensin I/II complex genes in BS single cells in comparison to WT cells.

Symbol	LogFC	AveExpr	Adj.P.Val	Description
SMC2	1.0425	5.9421	$2.85x10^{-10}$	Structural maintenance of chromosomes 2
SMC4	1.0425	6.4643	$1.16x10^{-9}$	Structural maintenance of chromosomes 4
NCAPG2	1.0894	5.6399	$1.15x10^{-7}$	Non-SMC condensin II complex subunit G2
NCAPG	1.0847	5.6721	$5.23x10^{-7}$	Non-SMC condensin I complex subunit G
NCAPD2	0.7609	5.6985	$5.69x10^{-4}$	Non-SMC condensin I complex subunit D2
NCAPH	0,7499	4,9869	0,00170	Non-SMC condensin I complex subunit H
NCAPD3	0,4757	4,4978	0,04889	Non-SMC condensin II complex subunit D3
NCAPH2	0.0207	4.9657	0.92426	Non-SMC condensin II complex subunit H2

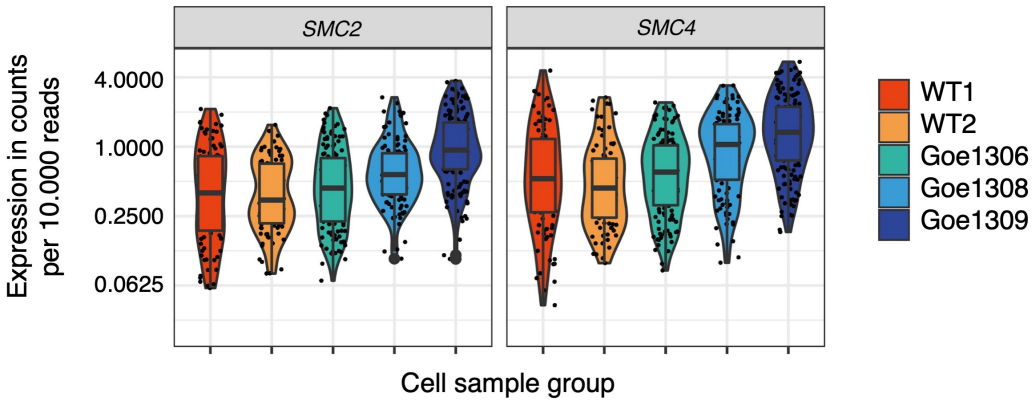


Figure 3.12. Expression levels of SMC2 and SMC4 in control and BS patient-derived single fibroblasts. Gene-specific expression levels of wild-type (WT1 and WT2) and BS patient-derived (Goe1306, Goe1308 and Go1309) samples are shown. The expression levels of both, SMC2 and SMC4 were overall higher in BS patient-derived fibroblasts compared to wild-type cells. Violin plots were generated using the R-package ggplot. Figure is adapted from Gönenc *et al.* [155].

3.2.2.2 Transcriptional changes induced by replication stress in fibroblasts

Two BS and one WT fibroblast samples were treated with aphidicolin to induce replication stress. Aphidicolin induces replication stress by inhibiting DNA polymerase which leads to replication fork stalling and further arrest in G1/S phase. The concentration of aphidicolin was 300 nM applied for 24 hours which was a mild treatment to induce

replication stress while allowing the cells to continue in cell cycle without the G1/S arrest. We compared the aphidicolin-treated samples to their non-treated sample pair and performed the same DEG and pathway analyses as described above. The aphidicolin treatment resulted in several significantly DEGs, therefore we performed ORA using gene ontology terms as well. The total number of significantly deregulated GO pathways in aphidicolin-treated BS fibroblasts was higher (152) than DEGs in the aphidicolin treated WT cells (123) when compared to their non-treated sample pair. This could imply that the BS patient-derived cells were more sensitive to replication stress than the WT cells. Of note, we did not continue with the analyses of comparing aphidicolin treated BS samples to WT samples since the effect of aphidicolin was masked by the highly significant transcription differences between BS and the WT cells. Further, the pathway analyses revealed chromatin and cell division related terms in treated BS cells but these terms were not significant in replication-stress induced WT cells. These significant GO terms were ‘mitotic nuclear division’, ‘mitotic sister chromatid segregation’, ‘sister chromatid segregation’, ‘nuclear division’, and ‘nuclear chromosome segregation’, which also contained several enriched condensin genes. This could further point to a link between BLM deficiency and condensin genes’ expression through replication stress.

3.2.2.3 Analyses of condensin I/II complex members on a protein level

To prove an involvement of condensin I/II complex members in BS, we further analyzed the protein expression levels of condensin I/II complexes in BLM-deficient cells by Western blot analysis. BS patient-derived cells and control fibroblasts, which were used in the scRNAseq experiment were harvested, analyzed and subjected to Western blot analysis using antibodies against SMC2 and SMC4. For both proteins, we detected higher protein expression in at least 2 out of 3 BS patient-derived fibroblast samples (Figure 3.13, upper part). In a series of Western blot experiments using biological replicates of these samples, we were able to verify this tendency of higher protein levels of SMC2 and SMC4 (Figure 3.13, lower part). For non-SMC proteins of the condensin I/II

complexes such as NCAPG2 and NCAPH2, protein detection in WT samples as well as BS patient-derived fibroblasts failed possibly due to low protein amounts in cell lysates and/or low antibody specificity for these proteins.

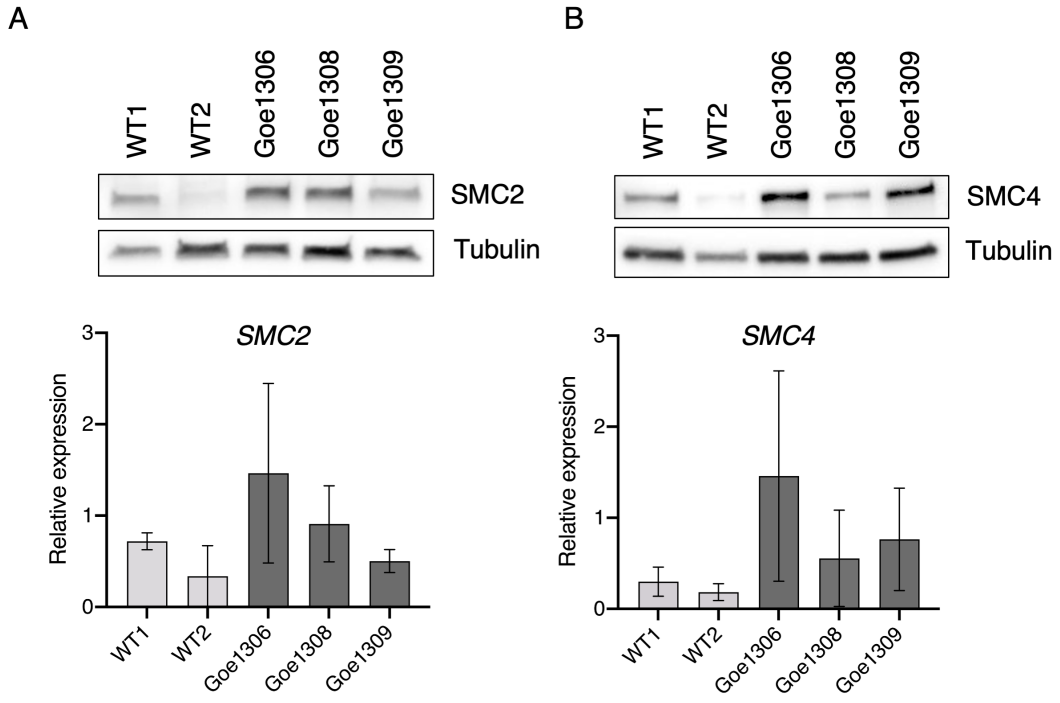


Figure 3.13. Analyses of SMC2 and SMC4 protein expression levels. Representative Western blots (upper part) and their quantification (lower part) revealed higher protein expression levels of SMC2 (A) and SMC4 (B) in the BS patient-derived fibroblasts (Goe1306, Goe1308, and Goe1309) compared to wild-type (WT1 and WT2) cells. Expression levels were normalized using alpha-tubulin as loading control. n=3; mean ± SD.

3.2.2.4 Analysis of BS cells from a larger proteomics perspective

In order to assess alterations in protein expression in a more general and systematic setup and to further validate the results obtain from scRNAseq data analysis, we aimed to elucidate the protein composition in BS cells using a proteomics approach. In collaboration with Markus Räschle (Technical University of Kaiserslautern), I conducted a mass spectrometry (MS) experiment using cell lysates of BS patient-derived fibroblast (Goe1306, Goe1308, and Goe1309) together with two WT controls (WT1, WT3). I performed the MS analysis during a scientific stage at the Räschle Lab in Kaiser-

slautern. PCA analysis of data showed mostly similar protein expression profiles of all triplicates belonging to the same sample with an exception for WT3 replicate 3 and BS2 replicate 1 which clustered separately than their pairs (Figure 3.14A). In total, 6,534 protein identities were detected and filtering of potential contaminants, reverse hits, and hits identified only by site resulted in 6,192 remaining protein identities. Next, the abundance of the detected proteins from the BS patient-derived sample group was compared to the WT group and the significance of the differences was assessed by multiple corrections using the Perseus (MaxQuant) pipeline [175]. In total, 148 proteins were differentially expressed in BS patient-derived samples compared to WT samples. We analyzed the gene groups, which were addressed in the scRNAseq analyses, i.e., FA pathway, microcephaly-related genes, and the condensin I/II complex members, and we were able to confirm expression of most of these proteins in the cell lysates of fibroblasts analyzed by MS. However, we were not able to finally confirm alterations also on protein level of these proteins in BS patient-derived samples when compared to WT samples after multiple test corrections (Figure 3.14B). Solely, protein expression of *CIT*, one gene of the microcephaly-related gene group, was significantly increased in BS samples in comparison to the WT samples. SMC2, SMC4, NCAPH, and NCAPD2 protein expressions were detectable in the MS data, however, the expression of these proteins was not significantly higher in BS patient-derived samples. Of note, BLM as well as the additional members of the BTRR complex (TOP3A, RMI1, and RMI2) were neither detected in BS patient-derived samples nor in WT samples. In general, protein expression changes between BS and WT samples were only very mild and not reaching significance levels for many of the detected proteins in the MS data.

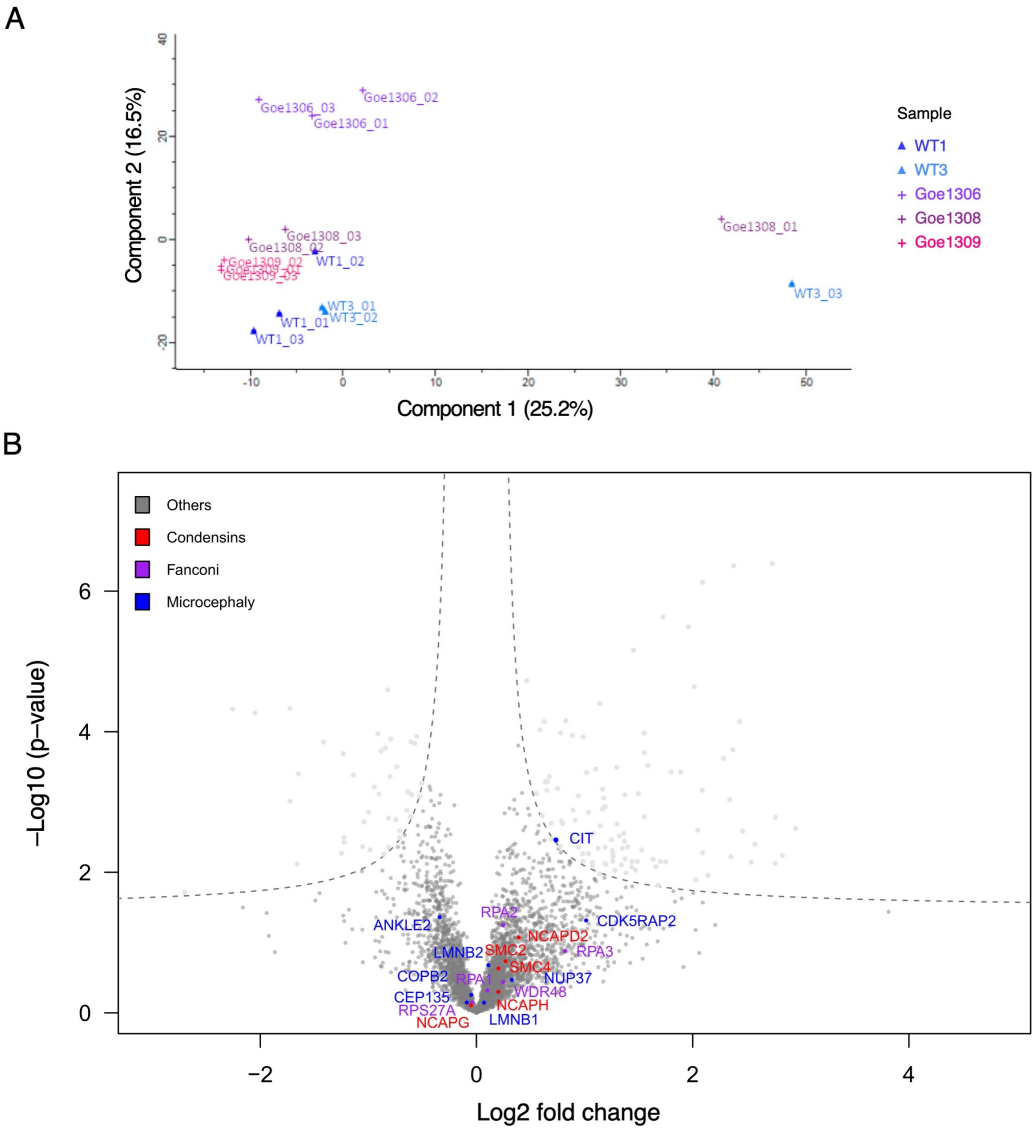


Figure 3.14. Mass spectrometry analyses of protein expression in BS patient-derived fibroblasts compared to wild-type cells. (A) Principal component analysis (PCA) score plot of mass spectrometry data obtained from three BS patient-derived fibroblast samples (Goe1306, Goe1308 and Go1309) and two wild type samples (WT1 and WT3). All samples were analyzed in triplicates, the percentage of variation, 25.2% for PC1 and 16.5% for PC2, is shown in parentheses.. (B) Volcano plot analysis of protein expression in BS patient-derived fibroblast compared to wild-type cells. Protein expression of genes of the condensin I/II complexes are labeled in red, microcephaly-associated genes in blue, and Fanconi anemia gene group in purple. Volcano plot was generated using ggplot package for R-programming language.

3.3 Generation and systematic characterization of the BTRR complex-deficient iPSC lines

To determine the functional impact of mutations in members of the BTRR complex on different cellular processes and pathways was often challenged by the restricted availability of patient material. Disease modelling through patient-derived cell systems can be also challenging due to the different genetic backgrounds of individuals. During this thesis, we used a CRISPR/Cas9-based approach to analyze the functional impact of the BTRR complex deficiency and the pathomechanisms causing BS on a cellular level. We focused on the generation of knockout cell lines for each member of the BTRR complex on an isogenic background, which provided us a solid basis for robust and reproducible experiments due to shared single genetic background of all cell lines. WT iPSCs were used and biallelic frameshifting mutations were introduced in genes encoding members of the BTRR complex.

3.3.1 CRISPR/Cas9-based generation of the BTRR complex-deficient iPSC lines

To study cellular phenotypes observed in BS cells on isogenic cellular models apart from patient-derived fibroblasts, we made use of CRISPR/Cas9 technology to introduce LoF mutations in genes coding for each member of the BTRR complex. Transfections were performed using the wild-type iPSC line isWT1.14 (from here on referred to as isWT1), which was kindly provided by the Stem Cell Unit (SCU, University Medical Center Göttingen). This WT iPSC line was generated as previously described via reprogramming of skin fibroblasts from a healthy control donor [154]. Transfections were carried out using ribonucleoprotein delivery of the crRNA and the Cas9 protein through nucleoporation. After the cells recovered from the transfection, half of the cells were cryopreserved while the other half were harvested for the assessment of the transfection efficiency (Figure 3.15). The analyses of the generated alleles were done by

NGS-based deep amplicon sequencing. According to the transfection efficiency determined by the analyses of the deep amplicon sequencing, cryopreserved cells were thawed and singularized followed by clonal analysis. A new crRNA was designed for a new transfection in case of low efficiencies, i.e., for *RMI1* and *RMI2*. Detailed results of the genome editing experiments on iPSCs are shown in the following sections of this thesis.

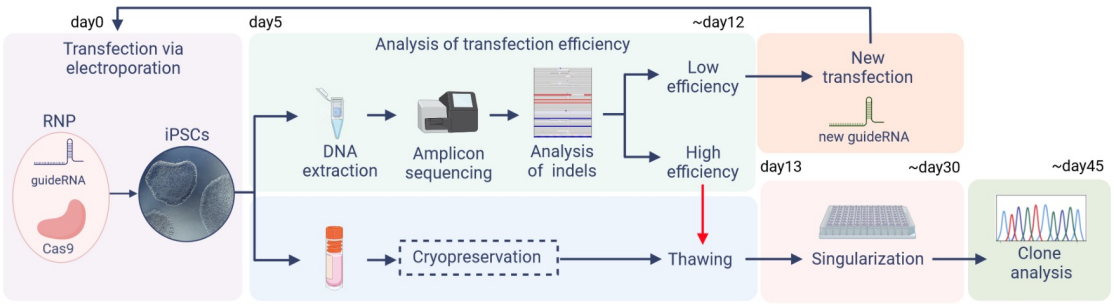


Figure 3.15. Workflow of the CRISPR/Cas9-based genome editing strategy. iPSCs were transfected with RNP containing Cas9 and guideRNAs for the corresponding target gene via nucleofection. Following transfection, half of the cells were harvested for bulk DNA extraction followed by deep amplicon sequencing. The other half of the cells were cryopreserved for later thawing depending on the indel frequencies obtained from deep amplicon sequencing. In the case of low efficiencies, new transfection strategies were planned and performed. In the case of sufficiently high efficiencies, bulk cells were thawed and singularized using the CellenONE platform at the Stem Cell Unit (UMG). Single cell clones were analyzed by Sanger sequencing when colonies reached a sufficient size. Approximation of minimum timings is given above the chart boxes. The chart was created using BioRender.

3.3.1.1 Design of the guideRNAs with respect to BS patient mutations

Mutations in all members of the BTRR complex have been identified in patients presenting with BS/BS-like phenotype. We focused on the locations of the described patient truncating mutations in the genes encoding the members of the BTRR complex and designed the guideRNAs in close proximity to those mutations for genome editing [73, 74]. The patient mutations used as reference regions for the crRNA design are given in Table 3.3. We used the CRISPOR tool to analyze possible guideRNA sequences in gDNA regions near the target regions [176]. The web tool lists possible guideRNA sequences based on various parameters such as off- and on-target scores. For all four

members of the BTRR complex, we selected guideRNA sequences with the best score at a close position to mutations identified in BS patients. For *RMI2*, a guideRNA sequence with the highest score was selected, which was located towards the end of the first coding exon of *RMI2*.

Table 3.3. BS patient mutations used as reference regions for the design of the guideRNA.

Gene	Patient mutation (homozygous)	Reference
BLM	c.572_573del;p.Arg191Lysfs*4	Gönenc et al., 2022 [158]
TOP3A	c.2428del;p.Ser810Leufs*2	Martin et al., 2018 [74]
RMI1	c.1255_1259del;p.Lys419Leufs*5	Martin et al., 2018 [74]
RMI2	Complete loss of the gene	Hudson et al., 2016 [73]

3.3.1.2 Evaluation of genome editing efficiencies of iPSCs

Using the guideRNA sequences given in Table 2.7 on page 32 and Cas9 nuclease, we aimed to introduce biallelic LoF mutations to all four members of the BTRR complex. After the first round of transfection using crRNAs for each member of the BTRR complex on wild-type iPSCs, we assessed the transfection efficiencies by performing a deep sequencing experiment with the approximately 400 bp region spanning the putative Cas9 cut site on the bulk gDNA of the transfected cells. The analysis of the generated indels was done via Cas-Analyzer [177], and calculated indel frequencies are summarized in Table 3.4 for each gene and transfection. Transfection efficiencies for *BLM* and *TOP3A* were sufficient for continuing with singularization of cells followed by single cell-derived clone analysis. Based on the deep amplicon sequencing of bulk gDNA regions, transfection efficiencies for *RMI1* and *RMI2* were calculated to be below 10%, which was not sufficient for further clone analysis. Therefore, transfections were repeated with alternative crRNA oligos (Table 2.7 see Section 2.1.4 on page 30). The efficiencies for these transfections remained very low; therefore, bulk cells from the first transfection of *RMI1* were thawed and singularized. Due to time restrictions, a new transfection for *RMI2* was not performed and this gene was left out from further

investigations. Singularization of iPSCs was carried out at the Stem Cell Unit using the CellenOne system (Scienion). The content of each well was evaluated by imaging and wells containing zero or more than one single-cell colony were excluded from further analyses. When single-cell clones reached to a sufficient size, each clone was cryopreserved and, additionally, gDNA was isolated from each clone and subjected to Sanger sequencing for the targeted regions. For *BLM*, 36 clones were analyzed (Table 3.4). 10 clones turned out to carry the wild-type sequence homozygously at the targeted region, 19 clones showed alterations on one allele, and 7 clones carried biallellic alteration at the targeted region. These 7 clones were analyzed in detail and all clones carried an in-frame deletion on at least one allele and were therefore excluded. For *TOP3A*, 98 clones were analyzed and 84 of them carried the wild-type sequence in homozygosity at the *TOP3A* target site. 11 clones were heterozygous and two clones showed compound heterozygous alterations. Similar to *BLM*, both of these clones carried at least an in-frame deletion on one *TOP3A* allele and were therefore excluded. Single-cell clone analysis for *RMI1* revealed 65 wild-type and 12 heterozygous clones for the targeted region in *RMI1*. In total, no iPSC clone harboring biallelic frameshifting mutations in *BLM, TOP3A*, or *RMI1* was obtained after the first round of transfections.

Table 3.4. Experimental flow and the transfection efficiencies determined by deep sequencing for the corresponding target genes of the BTRR complex.

Cell line type	Experiment	*BLM*	*TOP3A*	*RMI1*	*RMI2*
Wild-type	Transfection (indel %)	55	12.7	8.9	0.3
Wild-type	Transfection[a] (indel %)	-	-	5.8	7.6
Bulk transfected wild-type	Singularization (n)[b]	36	98	77	-
Heterozygous	Transfection (indel %)	92	82.7	75.2	-
Bulk transfected heterozygous	Singularization (n)	34	23	15	-
Final number of compound heterozygous clones		1	3	2	-

[a]Different guideRNA sequences were used in this transfection for the two genes.
[b]The number of the analyzed clones.

We then continued by performing a second round of transfections using heterozygous clones carrying heterozygous frameshifting variants in *BLM, TOP3A,* and *RMI1*. The same crRNA molecules were used targeting only the wild-type allele on the heterozygous clones. Therefore, clones that carry larger deletions including the PAM sites of the crRNAs were selected for each gene which facilitated crRNA binding to the wild-type allele and not the edited allele (Table 2.7). For all three genes, the transfection efficiencies were higher compared to the first round of transfections (Table 3.4). Transfected cells were singularized, cryopreserved and genome editing efficiencies were analyzed by Sanger sequencing of gDNA isolated from each clone. For *BLM*, 34 clones were analyzed in total. 11 clones showed the same *BLM* genotype as the parental iPSCs used for transfection. 23 clones showed compound heterozygous alteration at the targeted region in *BLM*. However, 22 of these clones contained in-frame deletions on the second *BLM* allele and were hence excluded. A single iPSC clone was identified carrying biallelic frameshifting variants in *BLM* (Table 3.5 and Figure 3.16A) which was selected for further experiments in the course of this thesis. For *TOP3A*, 27 single-cell clones were screened. 22 clones still contained a wild-type allele, 5 clones contained compound-heterozygous alteration at the targeted region in *TOP3A*. Of these 5 clones, three carried frameshifting variants in *TOP3A* on both alleles. The clone with the allele resulting in the earliest stop codon was selected for further studies (Table 3.5, Figure 3.16B). In the case of *RMI1*, 15 clones were analyzed. 13 clones did not show an alteration at the target region compared to the parental iPSC clone, two clones carried compound heterozygous frameshifting variants in *RMI1*, one of which was selected for further characterization (Figure 3.16C). The genotypes of the introduced alleles for each gene are summarized in Table 3.5.

In summary, we introduced truncating mutations on both alleles for three members of the BTRR complex, namely, *BLM, TOP3A,* and *RMI1*, in separate iPSC clones using a CRISPR/Cas9-based approach. All three biallelic clones harbored compound heterozygous truncating variants in the corresponding genes in close proximity to the positions of mutations identified in BS/BS-like patients (Figure 3.16). It is of interest to note that, in general, homozygous truncating variants in specific genes by CRISPR/Cas9 tech-

Table 3.5. Allele details of the iPSC clones for the corresponding genes.

Cell line name	Gene	Allele 1	Allele 2
isWT1-BLM-KO.2D7	*BLM*	wild type	c.715_716del; p.Asp239*
isWT1-BLM-KO.1D1	*BLM*	c.716del; p.Asp239Valfs*10	c.715_716del; p.Asp239*
isWT1-TOP3A-KO.1F3	*TOP3A*	wild type	c.2475_2481del; p.Lys826Alafs*54
isWT1-TOP3A-KO.1E10	*TOP3A*	c.2473dup; p.Arg825Profs*2	c.2475_2481del; p.Lys826Alafs*54
isWT1-RMI1-KO.1D9	*RMI1*	wild type	c.1228_1234del; p.Leu410Metfs*10
isWT1-RMI1-KO.1F11	*RMI1*	c.1230dup; p.Ala411Serfs*3	c.1228_1234del; p.Leu410Metfs*10

nology is feasible. The reason why the generation of knockout iPSCs for the members of the BTRR complex required a second round of transfections using heterozygous clones could be due to the sensitivity of the cells to the introduced mutations as well as transection efficiencies and strategies.

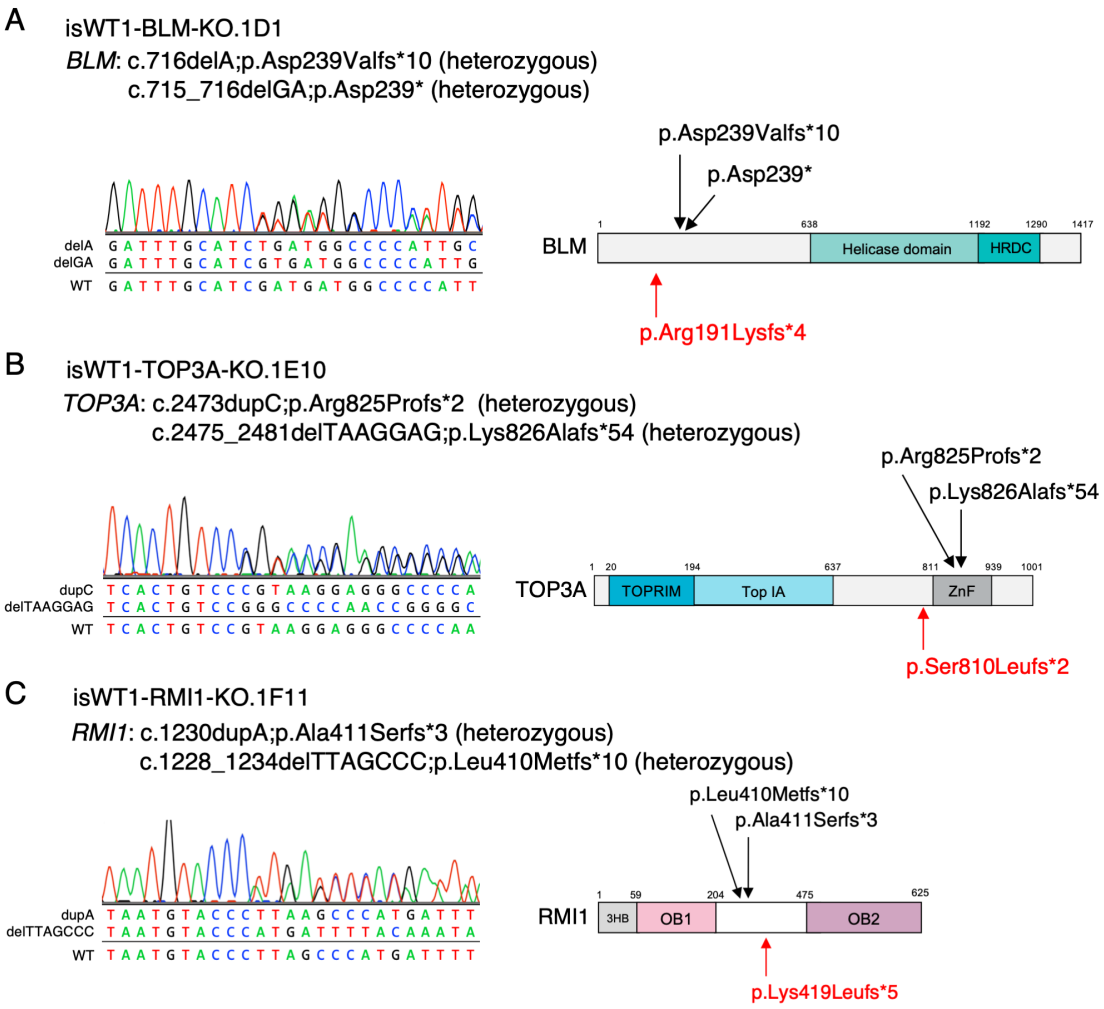


Figure 3.16. Sanger sequencing analyses of the KO iPSC clones. Sanger sequencing analyses of the CRISPR/Cas9-based KO of (A) BLM, (B) TOP3A, and (C) RMI1 in iPSCs. All three KO iPSC clones were compound heterozygous for loss-of-function mutations in the corresponding genes. All mutations were confirmed by Sanger sequencing (left panel). The positions of the truncating mutations with respect to the schematic representations of the corresponding proteins are marked in black, the positions of the corresponding mutations identified in BS and BS-like patients are marked in red (right panel).

3.3.1.3 Confirmation of *BLM, TOP3A,* and *RMI1* knockouts on protein level

To analyze the impact of the introduced frameshifting variants on a protein level, Western blot assays were performed using antibodies against each targeted member of the BTRR complex, i.e., *BLM, TOP3A,* and *RMI1*. WT, heterozygous and compound-heterozygous iPSC clones for each targeted BTRR member were analyzed and sub-

jected to Western blot analyses. In all three heterozygous clones we observed decreased protein expression of full-length BLM, TOP3A, and RMI1 compared to WT iPSCs (Figure 3.17A,B,C, middle lanes). Further, Western blot results confirmed that none of the compound heterozygous clones showed protein expression of full-length BLM, TOP3A, or RMI1 (Figure 3.17A,B,C, right lanes) as an indication of successful targeting and knockout of protein of the corresponding BTRR complex members. In the case of RMI1, a band smaller in size than the full-length RMI1 protein was detected in the compound heterozygous RMI1-KO iPSC clone. As this lower band was absent in the heterozygous clone, it is most likely that the truncated RMI1 protein was mainly expressed from the second targeted *RMI1* allele. The size difference between this band and the full-length protein was matching to the predicted size of truncated RMI1 lacking the C-terminal part due to the introduced frameshifting variant (p.Ala411Serfs*3). In line with this, we performed cDNA analysis using the RMI1-KO clone for the mutation region and observed the existence of both of the truncated mRNAs in the KO cells (data not shown). Therefore, it could be possible that the initial truncating allele that was introduced (p.Leu410Metfs*10) was not resulting in protein translation although the mRNA was stable. On the other hand, the second truncating allele (p.Ala411Serfs*3) might have been stable at both mRNA and translation level. Since the introduced *RMI1* variants were in close proximity to the *RMI1* mutation detected in BS patients, we continued analyzing this clone in the course of this thesis (Figure 3.16C).

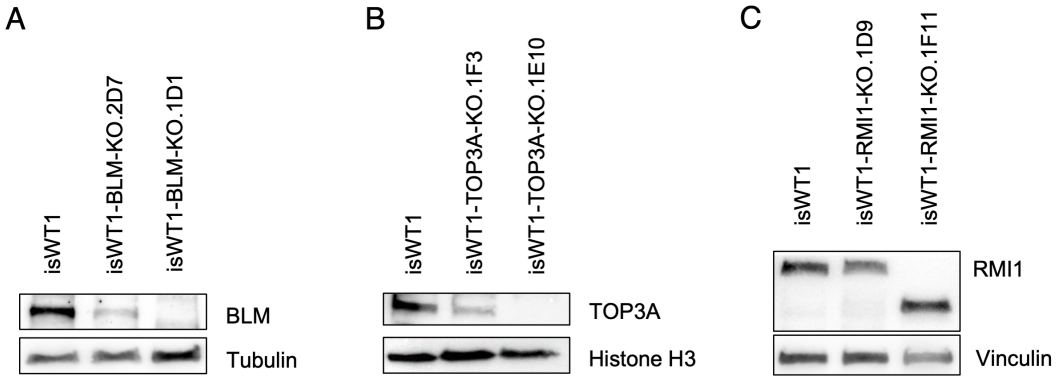


Figure 3.17. Western blot analyses of target protein expression in CRISPR/Cas9-based KO iPSC clones. Protein expression of (A) BLM, (B) TOP3A, and (C) RMI1 was analyzed using cell lysates of wild-type (left lane), heterozygous KO (middle lane), and compound heterozygous KO (right lane) iPSC clones for the corresponding genes. Alpha-Tubulin, Histone H3, and Vinculin served as loading control.

3.3.1.4 Pluripotency properties of the generated knockout iPSC lines

iPSC culture can be considered more peculiar than the commonly used cell models due to the potential of spontaneous differentiation of the cells in culture. Therefore, pluripotency has to be assessed regularly, especially after conducting experiments such as CRISPR/Cas9-based genome editing, transfections, and/or singularization. For this reason, we analyzed colony morphology and the expression of pluripotency marker genes, i.e., SOX2, OCT3/4, NANOG, SSEA4, and TRA-1-60. The expressions of the pluripotency markers were assessed by immunocytochemistry experiments. All three compound heterozygous KO clones showed a typical iPSC colony morphology (Figure 3.18A) and all five pluripotency markers were expressed according to fluorescent signals from the corresponding antibodies (Figure 3.18B). SOX2, OCT3/4, and NANOG were observed in the nuclei of the cells while SSEA4 and TRA-1-60 were observed in the cytoplasm in accordance with the protein localizations of each marker. We also quantified the ratio of cells which expressed the pluripotency markers, i.e., TRA-1-60 and OCT4, by FACS analysis, and this analysis revealed that more than 90% of cells of each clone expressed the two markers (Figure 3.19). Overall, the generated KO iPSC clones maintained their pluripotency after genome editing and were considered to be

suitable for further functional characterization.

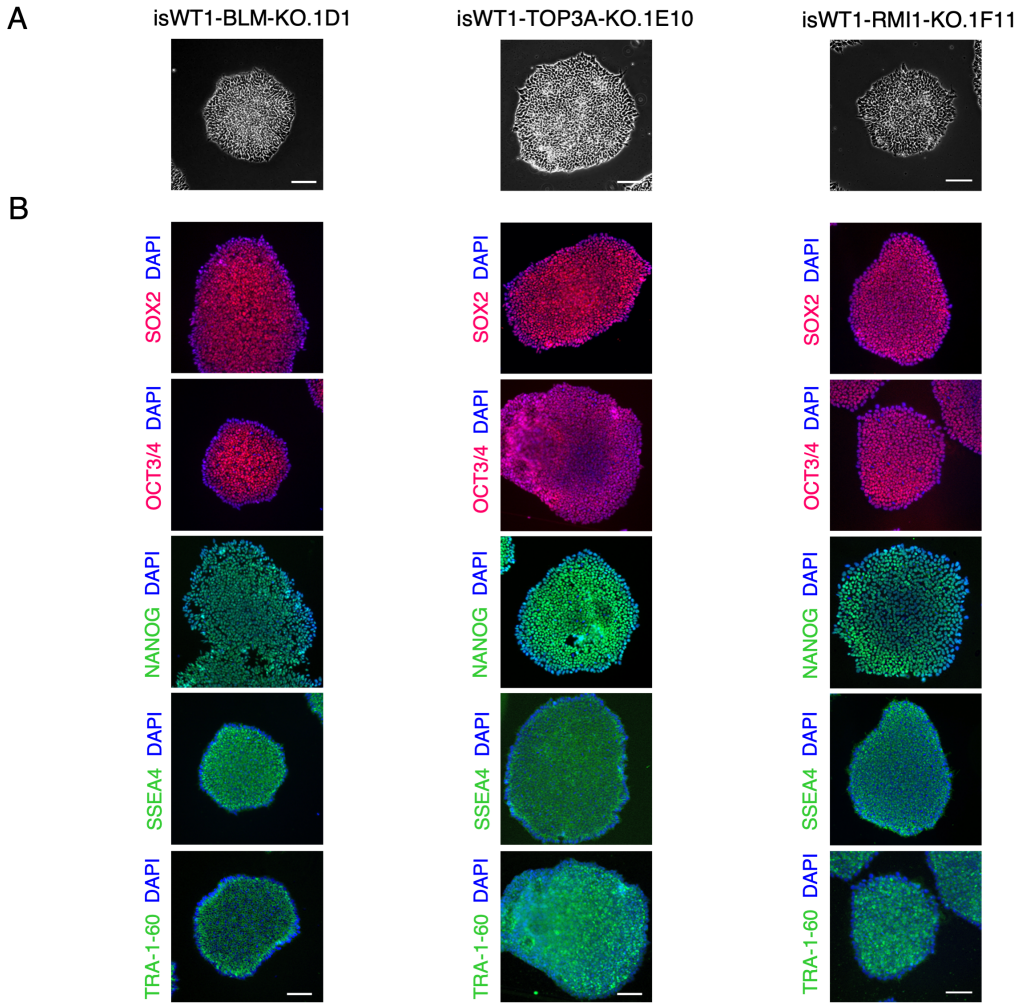


Figure 3.18. Morphological characterization and pluripotency assessments of the generated KO iPSC clones. (A) Representative images of all three KO iPSC clones confirming a typical human stem cell-like colony morphology. Scale bar 100 µm. (B) Immunofluorescence staining for protein expression of pluripotency markers SOX2 (red), OCT3/4 (red), NANOG (green), SSEA4 (green), and TRA160 (green) of the revealed expression of all corresponding pluripotency markers in all three generated iPSC clones. Nuclei were counterstained with DAPI (blue). Scale bar: 100 µm.

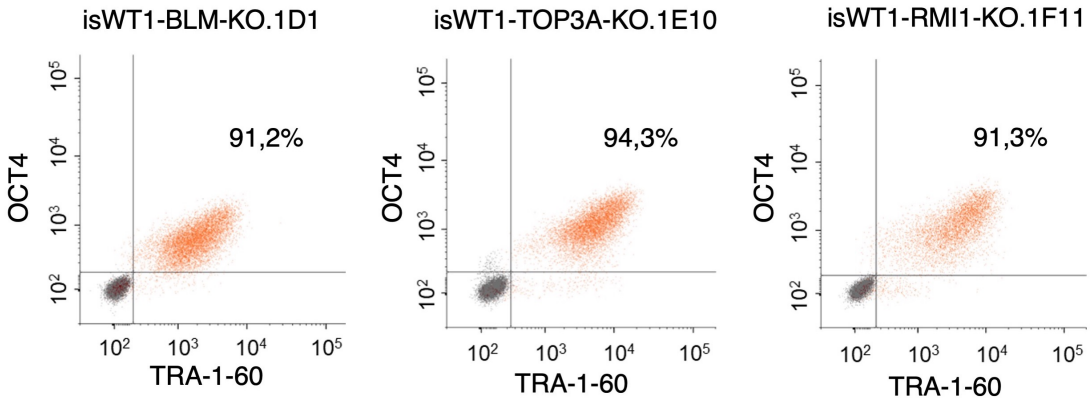


Figure 3.19. Flow cytometry analyses of protein expression of pluripotency markers OCT4 and TRA-1-60. KO iPSC clones for BLM (left panel), TOP3A (middle panel), and RMI1 (right panel) were stained with antibodies against OCT4 and TRA-1-60. More than 90% of analyzed cells were positive for both pluripotency markers. In total at least 6500 cells were analyzed in each experiment.

3.3.2 Cytogenetic and molecular cytogenetic characterizations of the BTRR complex-deficient iPSC lines

One cellular hallmark of BS is pronounced genomic instability leading to chromosomal aberrations including both structural and whole chromosomal instability (S-CIN and W-CIN, respectively) [57]. Therefore, we started our systematical analyses of the knockout cell lines for the members of the BTRR complex (BLM, TOP3A, and RMI1) by molecular cytogenetic analysis of chromosomal architecture and systematic characterization of molecular signatures of BTRR complex deficiency on gDNA.

3.3.2.1 Analyses of chromosomal integrity and CNVs in KO iPSC clones

Validation of the genomic integrity of iPSCs is a crucial step to execute after performing genome editing approaches in these cell lines. First, we started by analyzing the karyotypes of the generated iPSC clones. The chromosome preparation and analysis were performed in the course of routine genetics testing together with the cytogenetics group of the Institute of Human Genetics using protocols established for routine diagnostics settings. At least eight cells were analyzed for each different iPSC clone and

karyotype analyses revealed that all cells had a normal 46,XY karyotype without any chromosomal irregularity (Figure 3.20).

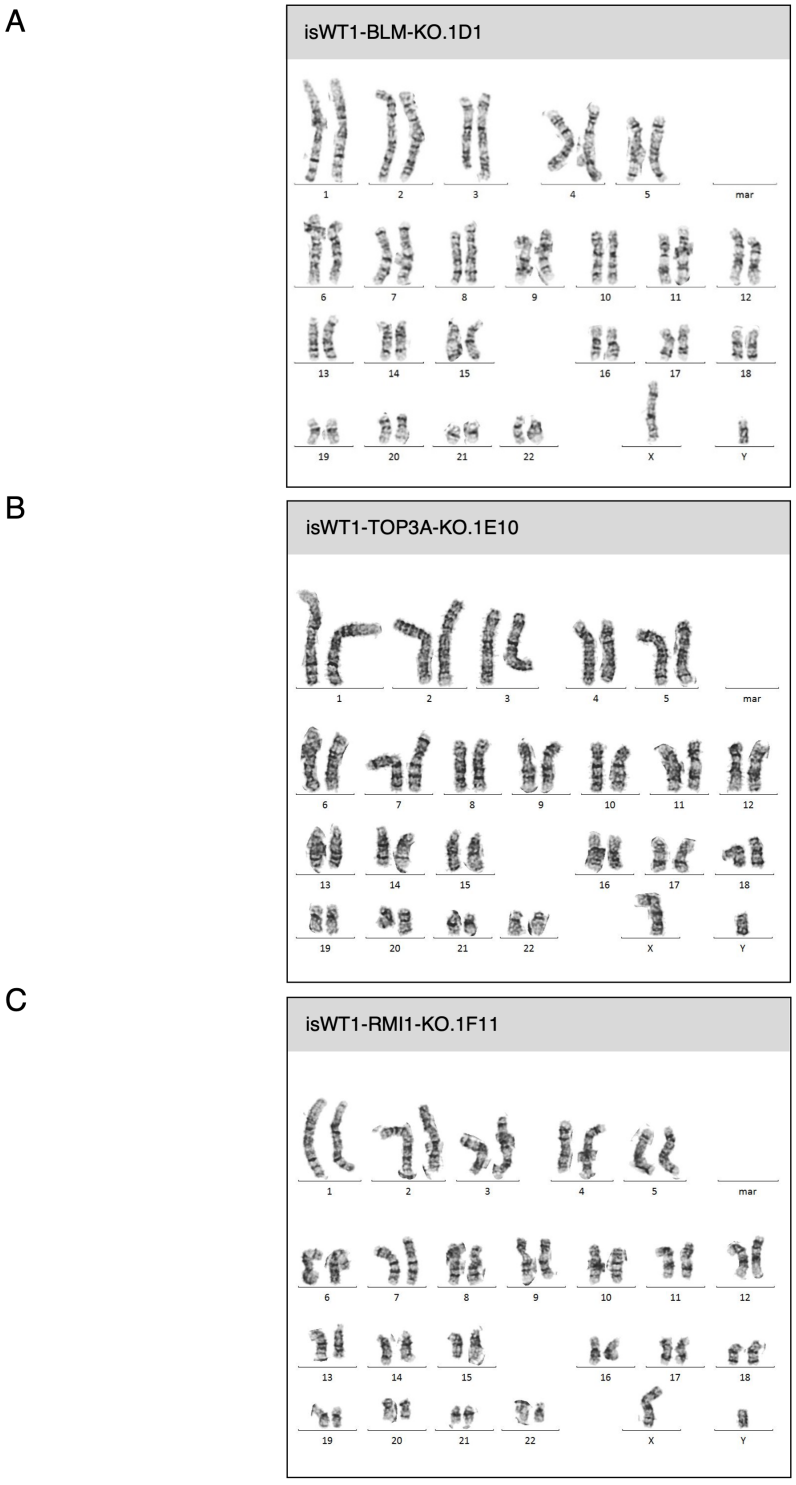


Figure 3.20. Metaphase karyotype analyses of the generated KO iPSC clones. Representative karyograms of (A) BLM-knockout, (B) TOP3A-knockout, and (C) RMI1-knockout iPSC clones are shown confirming normal karyotype (46, XY) after CRISPR/Cas9-based genome editing in all three clones. At least eight metaphases for each iPSC clone were analyzed according to standard procedures established for cytogenetic diagnostics.

Next, we analyzed the generated iPSC clones for possible copy number variations (CNVs). Array-based Comparative Genomic Hybridization (arrayCGH) is a tool used for the detection of segmental copy number variations on a higher resolution than classical cytogenetics [178]. It is based on hybridization of the differentially labeled genetic material of patient and control samples to reference genomic DNA targets on a microarray chip [179]. After the co-hybridization, the signals are detected by a SureScan Dx Microarray Scanner (Agilent), and the data was analyzed by bioinformatics software (Agilent Cytogenomics) used in routine diagnostics. The theoretical resolution of the array platform used in this thesis was 20 kb. All detected CNVs were compared to reference databases, such as dbVar genome browser, and/or the Database of Genomic Variants, and an internal database. Overall, I performed arrayCGH on all the three generated knockout iPSC clones. The highest number of CNVs were detected in the TOP3A-KO clone followed by RMI1-KO and BLM-KO clone. Most of the CNVs were identified in at least two of the iPSC clones tested. This could point out that the detected CNVs were originating from the parental cell line, which was the same for all three of the clones. The obtained CNVs were evaluated by comparison to the general population databases and all of them were classified as benign. CNVs smaller than 20 kb were not evaluated, which could be artifacts due to small size. These CNVs were excluded from further analysis. Overall, no pathogenic CNVs were detected in any of the three KO iPSC lines according to the diagnostic settings applied for the analysis of the arrayCGH data.

3.3.2.2 Establishment of increased sister chromatid exchange rates

Chromosomal instabilities in BS cells can mainly result from aberrant chromosome segregation events leading to uneven chromosome numbers, translocations, or chromatid breaks ([180, 181]. One of the prominent cytogenetic characteristics of BS cells is the increased rate of sister chromatid exchanges (SCE) due to unresolved double Holliday junction intermediates [141]. Sister chromatids, which contain high rates of SCEs, are called “harlequin-like” chromosomes and these are frequently observed, especially in

BS cells [182]. To determine SCE rates, I established a SCE assay using all three generated KO iPSC clones and the parental wild-type iPSC line as control. The SCE assay is based on incorporation of bromodeoxyuridine (BrdU) to the newly synthesized DNA followed by staining of metaphase spreads with Hoechst 33258 dye, UV-irradiation, and Giemsa labeling resulting in visualization of the sister chromatids with different color intensities. We successfully established a protocol for the analysis of SCE rates in iPSC lines. Initially, cells are incubated with BrdU during two cell cycles so that the BrdU is incorporated into both strands of newly synthesized DNA instead of thymidine. The duration of BrdU incubation has to be determined for every experiment separately and it depends on the cell cycle length of the cell lines used. First, all iPSC clones were incubated with BrdU for 24 hours with incubation according to a previous report of a BS patient-derived iPSC line [183]. The SCE assay did not reveal differentially stained sister chromatids after 24 hours of incubation; therefore, I increased the BrdU treatment duration to 40 hours. After BrdU incubation of 40 hours, sister chromatids were visualized differentially with the help of Hoechst 33258 and Giemsa dyes. Indeed, the metaphase spreads of all three KO iPSC clones showed easily recognizable 'harlequin' chromatids (Figure 3.21A). Thirteen metaphase spreads were analyzed for SCEs for each cell line, and I quantified the SCEs in the imaged metaphases and plotted the statistics against the exchange counts of the parental WT iPSC line. Highly significant increased rates of SCEs in all three generated KO clones were obtained when compared to the WT control (Figure 3.21B). The median of the SCE of the BLM-KO clone was 58 exchanges per metaphase, while it was only 3 for the WT control. Similarly, analysis of metaphase spreads of TOP3A-KO clone revealed a median of 63, and RMI1-KO clone had a median of 25 exchanges per metaphase indicating that metaphases spreads in the RMI1-KO clone was lower than in BLM-KO and TOP3A-KO clones, but were still highly significantly higher when compared to WT control cells.

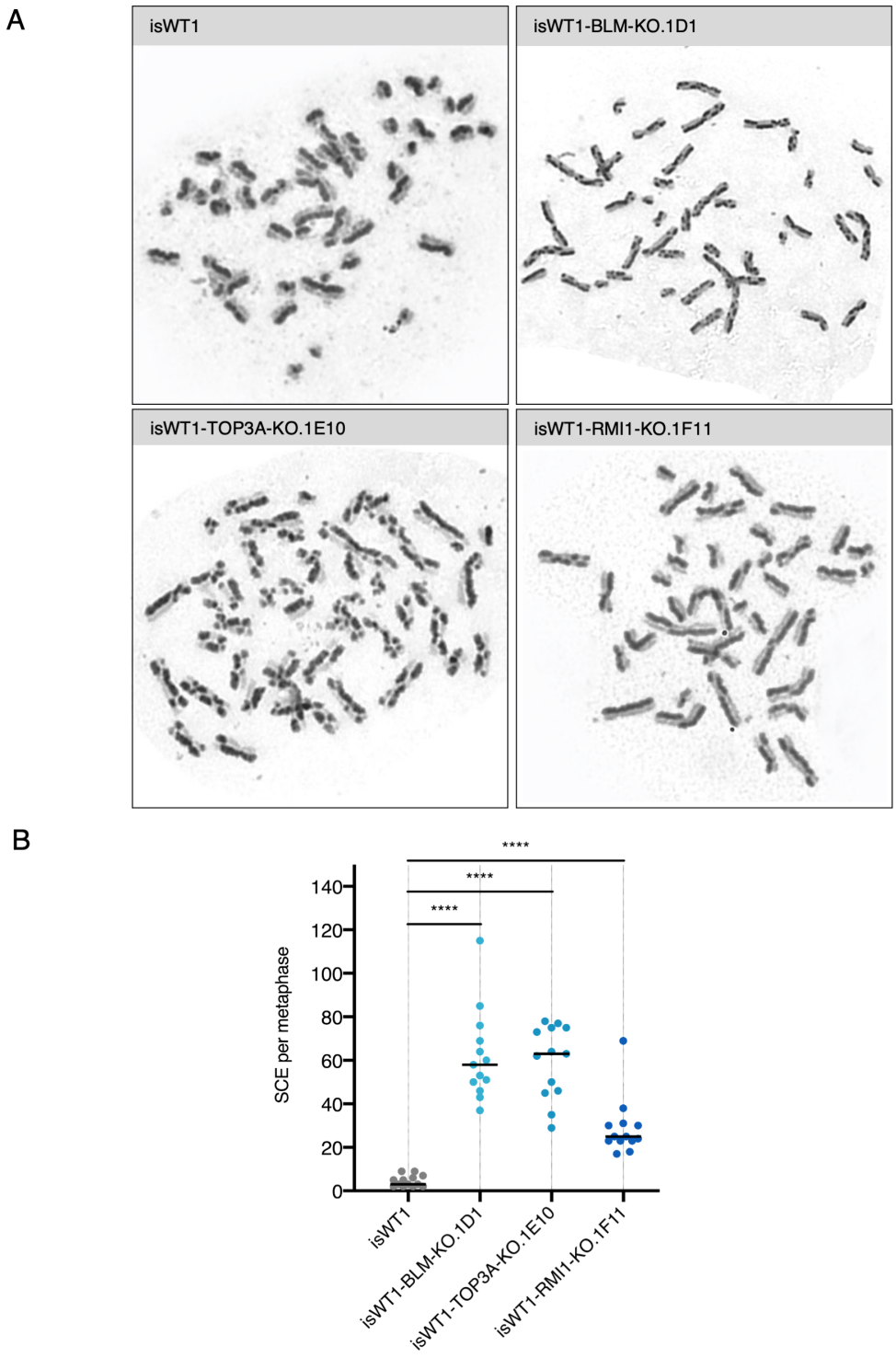


Figure 3.21. Analyses of sister chromatid exchange rates in KO iPSC clones and wild-type control cells. (A) Representative metaphase spreads of WT iPSCs (isWT1) and the generated three KO iPSC clones show BrdU strand-specific labeling of sister chromatids stained as dark and bright. (B) Quantification of the SCE rates reveals significant increases in all three generated KO iPSC clones. In total, thirteen metaphase spreads were analyzed for each sample. Unpaired non-parametric Mann-Whitney test was performed for each sample against the wild-type (isWT1). ****p<0.0001.

3.3.2.3 Aberrant chromosome segregation events in the BTRR complex deficiency

One source of genomic instability observed in BS patient-derived cells is caused by chromatin bridges, which can result from the unresolved DNA intermediates. Chromatin bridges occur as DAPI-positive thin lines between the sister chromatids that are being pulled to the opposite spindle poles during anaphase. To detect and quantify aberrant chromosome segregation events, immunofluorescence analysis of wild-type and KO iPSC clones was performed. DNA in unsynchronized cells was stained with DAPI, cells were fixed and subsequently imaged with a fluorescent microscope. At least 50 anaphases were analyses per experiment, each experiment was performed in triplicates, and chromatin bridges and lagging DNA in wild-type and the three KO cell clones were analyzed for each sample (Figure 3.22). Quantification showed that BLM-KO and TOP3A-KO iPSC clones had significantly higher amounts of chromatin bridges during anaphase. The mean percentage of anaphases showing chromatin bridges in WT control cells was 1.2%, while this value increased to 9.9% in BLM-KO and 13.4% in TOP3A-KO clones. Interestingly, the number of chromatin bridges observed in the RMI1-KO iPSC clone was increased, but not significantly higher than in the WT control (9.5%, p=0.122) (Figure 3.22).

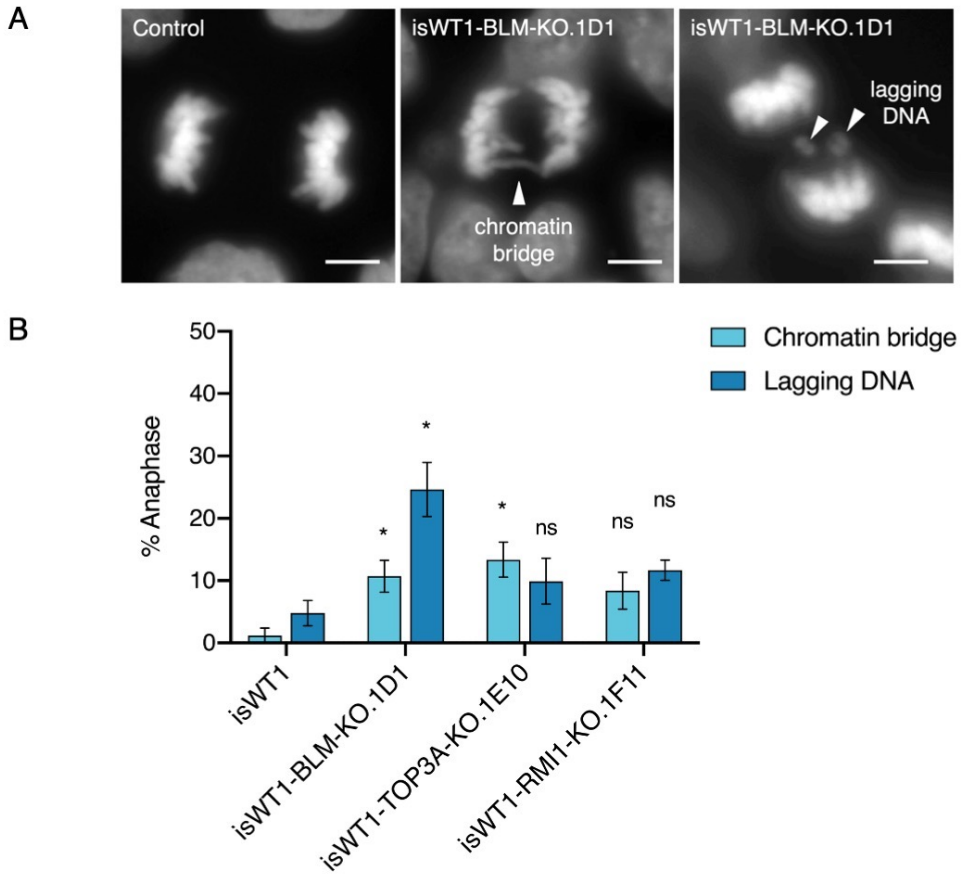


Figure 3.22. Determination of chromosome segregation defects during anaphase in BLM-, TOP3A- and RMI1-KO iPSC samples. (A) Representative images of DAPI-stained anaphases in control iPSC samples (left panel), positive chromatin bridges (middle panel) and lagging DNA (right panel) during anaphase of isWT1BLM-KO.1D1 cells. Scale bars: 2µm. (B) Quantification of anaphase defects in wild-type and three KO iPSC samples. Chromatin bridges and lagging DNA experiments were performed in triplicates with >50 cells analyzed per experiment. A two-tailed t-test was performed against the WT iPSCs for each sample. Mean ± SEM; *p<0.03.

Chromatid bridges can break by the oppositely applied forces on the segregating chromatids and this can result in genetic material unattached to the spindles and lagging chromosomes during later stages of mitosis. Lagging chromosomes (“laggards”) can fail to segregate together with the rest of the chromatids of the daughter cells leading to micronuclei (MN), which has been observed in BS patient-derived cells [148]. To quantify the amount of lagging chromosomes, we next analyzed the anaphase images for signs of lagging DNA and chromosomes. Quantification of lagging DNA at anaphase revealed that the BLM-KO clone had a significantly increased amount of lag-

ging DNA (Figure 3.22). The mean percentage of anaphases containing lagging chromosomes/DNA in the BLM-KO samples was approximately 25%, whereas only 4.8% anaphases of wild-type controls showed lagging chromatin. In TOP3A-KO and RMI1-KO anaphases, we observed an increased mean rate of anaphases containing lagging DNA (9.9% and 11.7%, respectively), still statistical analysis revealed no significance when compared to wild-type controls (p=0.308 and p=0.059, respectively).

3.3.2.4 Prevalence of the ultrafine anaphase bridges

Ultrafine bridges (UFB) are generally double stranded DNA bridges that occur between the segregating sister chromatids during anaphase [184, 185]. The UFB formation is a mitotic event that can occur in each dividing cell as a normal process. UFBs are caused by incomplete DNA replication in regions such as centromeres, common fragile sites or telomeres [106], and UFBs are resolved by helicases and other DNA repair proteins to protect the genome integrity. The BTRR complex as well as DNA resolvases such as GEN1 or MUS81 are known to be responsible for the clearance of UFBs. Due to the lack of histones and their stretched nature, UFBs cannot be visualized with regular DNA-staining dyes like DAPI [144]. Initially, they were detected by immunostaining for DNA helicases such as PICH or BLM [105, 186]. UFBs have been studied as a part of cellular characteristics of BS as the BLM helicase is strongly associated with the dissolution of the UFBs [145]. In collaboration with Markus Räschle and Angela Wieland from TU Kaiserslautern, I performed immunofluorescence stainings of anaphases to detect UFBs. As this study includes BTRR complex-deficient cell lines, we used PICH for UFB visualization (Figure 3.23A). At least 50 anaphases were imaged in biological triplicates for each cell line. Quantification of UFBs and statistical testing against the parental wild-type clone showed that all three of the KO iPSC clones had significantly increased rates of UFB occurrence (Figure 3.23B). I observed UFBs in approximately 28.4% of the imaged anaphases of wild-type control as well. For the BLM-KO cells, the mean rate of UFB containing anaphases increased to 57.4%. Similarly, iPSC clones lacking TOP3A or RMI1 also showed a significant increase in UFB-containing anaphases (52.4% for

TOP3A-KO and 47.9% for the RMI1-KO) indicating that all three components of the BTRR complex are essentially needed for efficient dissolution of UFBs (Figure 3.23B).

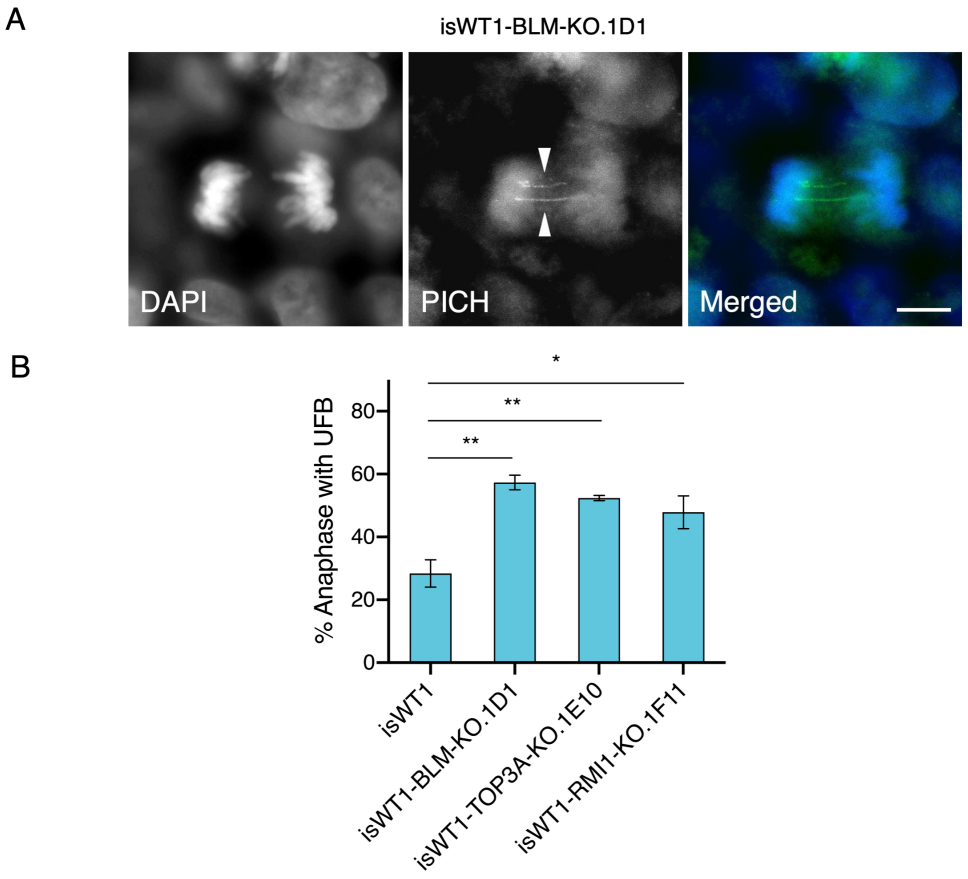


Figure 3.23. Analysis of ultrafine bridges in BLM-, TOP3A- and RMI1-KO iPSC samples. (A) Representative images of DAPI-negative, PICH-positive ultrafine bridges in BLM-deficient iPSCs. Scale bar: 2µm. (B) Quantification of ultrafine bridges in wild-type and BLM-, TOP3A- and RMI1-KO iPSCs. Experiments were performed in triplicates with >50 cells analyzed per experiment. A two-tailed t-test was performed against the wild-type clone for each sample. Mean ± SEM; *p<0.03, **p<0.002.

3.3.3 Determination of gene expression profiles of the BTRR complex-deficient iPSCs on a single-cell level

Next, we performed scRNAseq of isogenic BLM-KO, TOP3A-KO, and RMI1-KO iPSC clones to evaluate the transcriptional changes caused by the deficiency of each of these BTRR complex members. Additionally, we included the parental wild-type iPSC line (isWT1) and a second wild-type iPSC control line (isWT7). Further, isWT1, BLM-KO,

and TOP3A-KO iPSC clones were treated with aphidicolin for 24 hours at a concentration of 300 nM to induce replication stress prior to scRNAseq analysis. All eight samples were processed using ICELL8 (Takara Bio) system and sequenced on an Illumina HiSeq4000 as described in section 2.2.4.1. The raw sequencing data was obtained as a fastq file and analyzed using CogentAP and an in-house built scRNAseq data analysis pipeline in collaboration with Alexander Wolff (Institute of Human Genetics, University Medical Center Göttingen).

3.3.3.1 Quality control assessments of the scRNAseq data

We applied the quality control elements for scRNAseq data using the CogentAP pipeline (Takara) as described in Section 3.2.2.1.1 for scRNAseq data of BS patient-derived fibroblasts. Analysis of barcoded reads showed that the exonic reads were corresponding to 59.47% of all reads (Figure 3.24). In total, 720 out of 866 cells passed the quality control parameters, which included removal of cells with less than 10,000 total reads or less than 300 genes sequenced. Further quality control parameters revealed a robust quality of the scRNAseq data from the iPSC samples (Figure 3.25). For all samples, the mean number of reads per cell was approximately 200,000 and the total gene number per cell was between 6000 and 7000 which was within an optimal value range (Figure 3.25A). The ribosomal RNA percentage among all samples was almost zero. The detected mitochondrial RNA amounts were around 5% for all samples which was acceptably low for further analyses. Next, the gene expression profiles were assessed by the UMAP analysis (Figure 3.26). We observed homogeneous distribution for the wild-type iPSC (isWT1) and KO clones derived from this wild-type iPSC clone. Aphidicolin treated samples were also scattered among these samples and showed no distinct clustering. We observed separate clustering of the isWT7 iPSC clone which was included as an additional control cell line. Based on this, isWT7 was excluded from further analysis and the KO clones were compared to the isogenic parental wild-type iPSC control, isWT1.

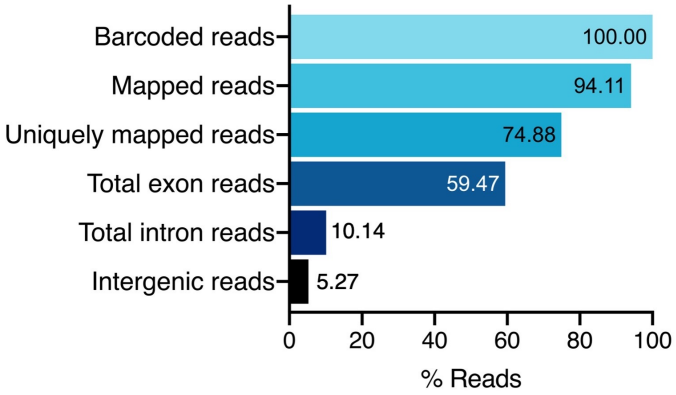


Figure 3.24. Classification of reads generated by scRNA sequencing of the wild-type and BLM-, TOP3A-, and RMI1-KO iPSC samples. Evaluation of the barcoded reads revealed that after quality filtering, more than 59% of all barcoded reads mapped to exonic regions. Approximately 15% of all barcoded reads mapped to intronic or intergenic regions and, therefore, were excluded.

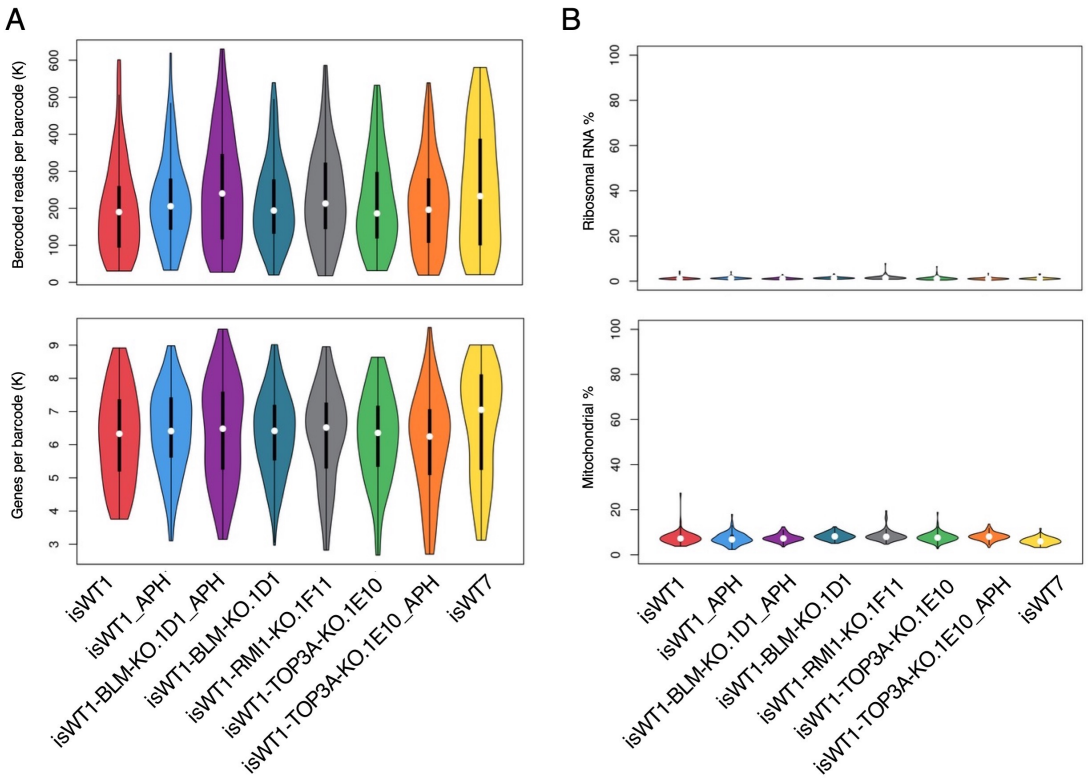


Figure 3.25. Quality control assessment of scRNAseq data of wild-type and BLM-, TOP3A-, and RMI1-KO iPSC samples. (A) After removing cells with <10,000 total reads or <300 genes, the average number of reads per barcode was around 200,000 for all samples (upper graph). Overall, expression of approximately 6.500 genes per cell was detected after removing genes with <100 total reads or genes expressed in fewer than three cells (lower row). (B) Analyses of reads mapping to ribosomal RNA (upper panel) and mitochondrial RNA (lower panel) showed almost no reads indicative of ribosomal RNA, and approximately 10% of reads mapping to mitochondrial RNA indicating robust scRNAseq data quality. Plots were generated using CogentAP (Cogent NGS Analysis Pipeline, Takara Bio).

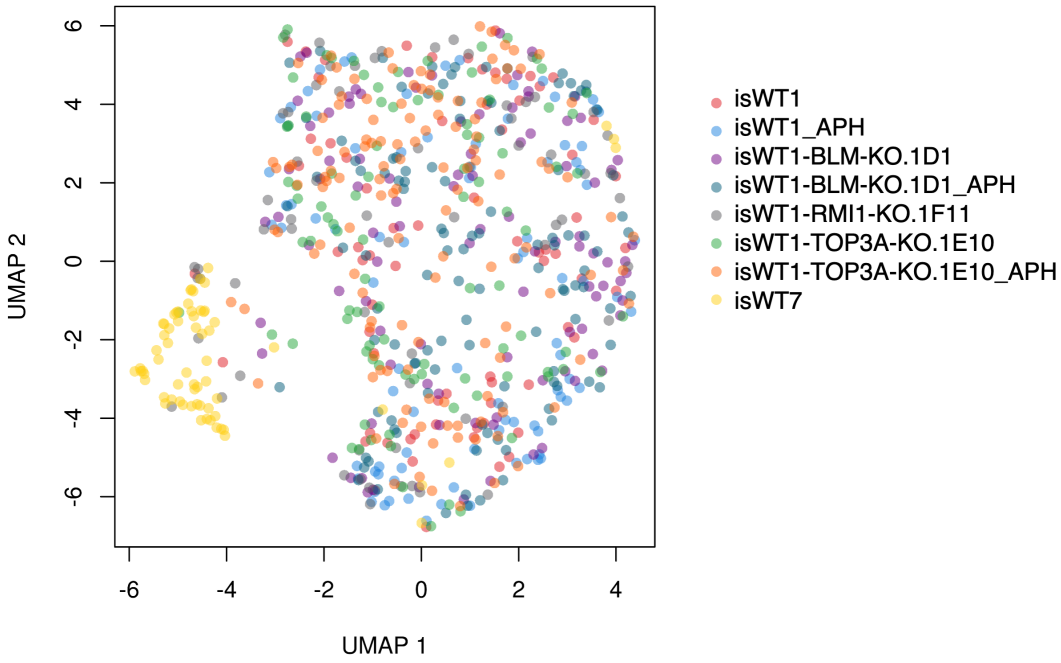


Figure 3.26. The UMAP analysis showing gene expression profiles of wild-type and BLM-, TOP3A-, and RMI1-KO iPSC samples. Cells were treated with 300 nM aphidicolin for 24 hours or left untreated as control. UMAP analysis reveals a homogeneous distribution of all generated KO clones, the parental wild type (isWT1), and all aphidicolin-treated samples. isWT7 (yellow), an additional, external wild-type control, clusters separately from the other samples. UMAP analysis was performed using CogentAP (Cogent NGS Analysis Pipeline, Takara Bio).

3.3.3.2 Analysis of differentially expressed genes in the three KO iPSC lines

Analysis of the single-cell gene expression profiles of wild-type iPSCs and KO-derivatives revealed similar gene expression patterns in all clones as already suggested by the UMAP analysis. Overall, differences within each sample as well as between the wild-type control and BTRR-KO clones were very small. We compared gene expression levels of each KO clone to the parental WT cell line and observed only a few numbers of genes as significantly deregulated in each KO iPSC clone (Figure 27). The highest number of significantly deregulated genes among the non-treated samples was obtained from the comparison of the RMI1-KO clone to the parental wild-type cells, which revealed 38 significantly deregulated genes. The comparison of the data obtained from the TOP3A-KO iPSC clone and the wild-type control showed 29 significantly deregu-

lated genes, and 25 genes were significantly deregulated when comparing the BLM-KO clone to the parental wild-type iPSC line. We observed some similarly deregulated genes in the KO iPSCs when compared to the WT. *HIST1H1A* was the most severely downregulated gene in the BLM-KO sample and also significantly downregulated in the RMI1-KO cells (Figure 3.27A,C). *HIST1H1A* codes for H1.1, a member of the histone family that belongs to a replication-dependent subtype of linker histone whose expression levels peaks during S-phase [187, 188]. In order to exclude that detection of *HIST1H1A* transcript was caused by differences in cell cycle distribution between WT and BLM-KO or RMI1-KO iPSC clones, we next analyzed cell cycle distribution in all samples using bioinformatics tests. Seurat toolkit was used for the cell cycle profiling of the iPSC data [166]. This package assigns a cell cycle score to every cell depending on the expression of cell cycle marker proteins. According to the assessed scores, cells are classified to be in either G1, S, or G2/M. With this analysis, we did not observe a significant difference between the cell cycle stages of iPSCs (data not shown), which validated that the observed alterations in *HIST1H1A* gene expression was not affected by the cell cycle stages. The second most downregulated transcript in the BLM-KO iPSCs was the *BLM* gene confirming the knockout of this cell line on a single-cell level transcriptome level. *TOP3A* and *RMI1* transcripts were not detected as significantly deregulated in the corresponding KO iPSC lines, i.e., TOP3A-KO and RMI1-KO, respectively. This could be due to higher mRNA stability of the truncated transcripts as these transcripts had premature stop codons towards the end of each corresponding protein which could result in stable mRNAs. Next, the most downregulated gene was *CHCHD2* in both TOP3A-KO and RMI1-KO iPSCs. *CHCHD2* encodes a transcription factor that is involved in transcription regulation in response to oxidative stress and its downregulation sensitizes cells to apoptosis [189]; however, a transcriptional regulation of *CHCHD2* by the BTRR complex has not been described yet. Of note, none of the downregulated genes were off targets of the guideRNAs used in the genome editing.

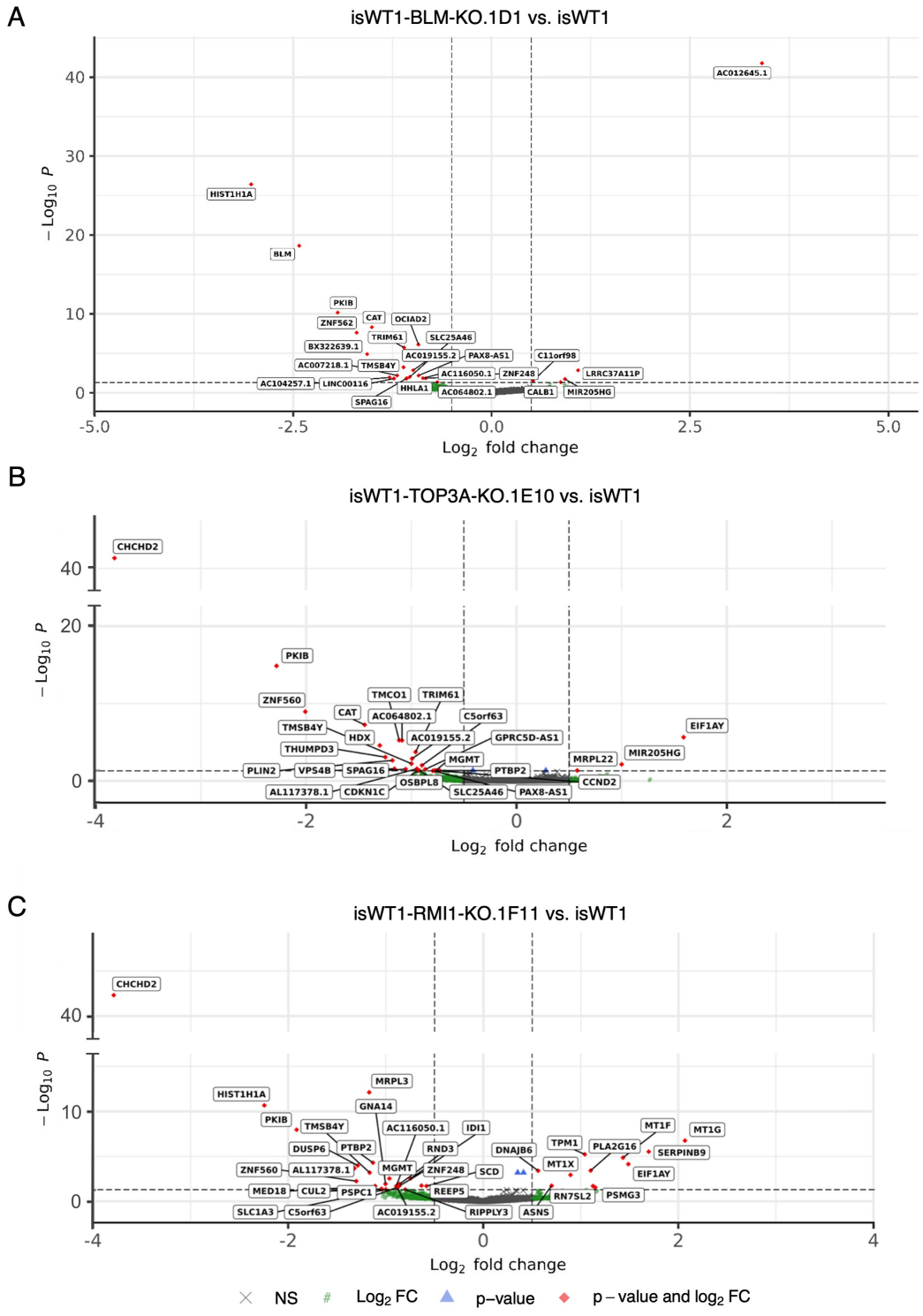


Figure 3.27. Differentially expressed genes in BLM-, TOP3A-, and RMI1-KO iPSCs. Volcano plot analysis of gene expression profiles of BLM- (A), TOP3A- (B), and RMI1-deficient (C) iPSC clones compared to the wild-type iPSCs. Significantly deregulated genes (Log2 fold changes >0.5 or <-0.5) are named and indicated by red rhombi. Volcano plot analysis was performed using the EnhancedVolcano package from Bioconductor for R-programming language.

We obtained several transcripts, which were deregulated in at least two of the KO-iPSC samples, such as *HIST1H1A* or *CHCHD2* as described above. In addition, we observed overlapping significant differential expression of genes in all three KO iPSCs, such as *PKIB, TMSB4Y,* and *AC019155.2* (Figure 3.28). These transcripts were detected as downregulated in three KO iPSCs. We further compared the differentially expressed genes in iPSC data to the BS patient-derived fibroblasts data. Two genes, i.e., *OCIAD2* and *PAX8-AS1*, were significantly downregulated in both BLM-KO iPSCs and BS patient-derived fibroblasts. *PAX8-AS1* transcript was also downregulated in the TOP3A-KO iPSCs. This transcript is a lncRNA, which might regulate the gene expression *PAX8*, that is a transcription factor from the PAX gene family. Although shared gene expression differences between both iPSC samples and the BS patient-derived fibroblasts were an interesting finding for the KO-iPSCs, a clear biological explanation in the context of BTRR complex deficiency was challenging to conduct.

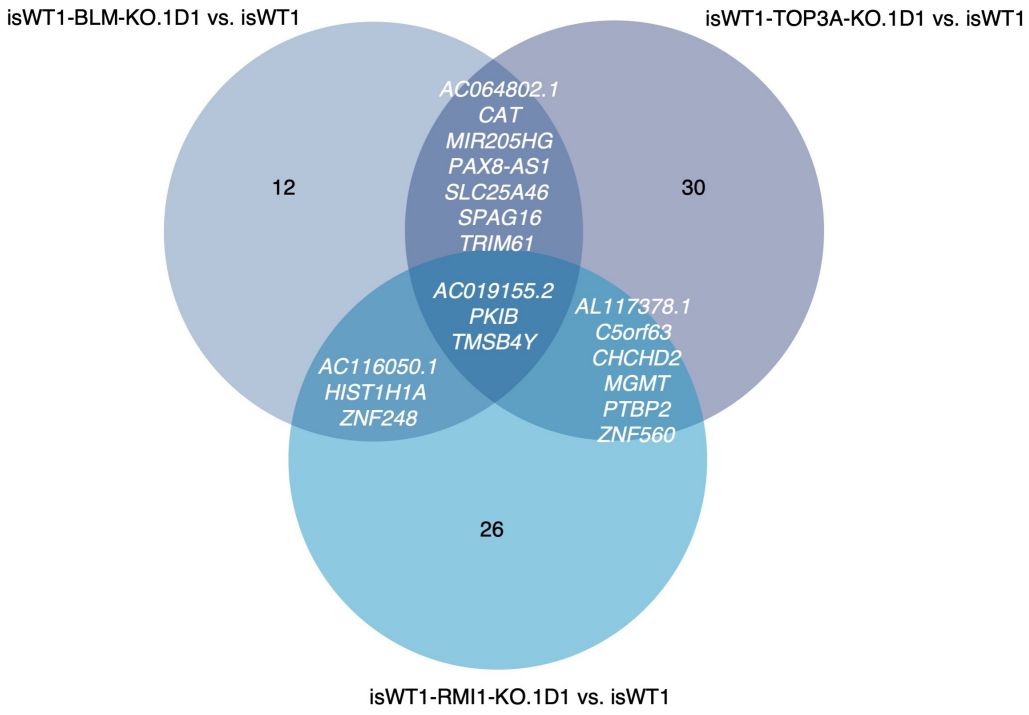


Figure 3.28. Overlapping significant genes in the comparisons of DEGs in KO iPSCs against the parental WT control. All given genes except *MIR205HG* were significantly downregulated in the KO-iPSCs (FDR<0.05). The numbers of the genes belonging to the rest of the comparisons are given in the chart.

3.3.3.3 Pathway analyses on the single-cells' transcriptome

Similar to scRNAseq data obtained from BS patient-derived fibroblast (section 3.2.2.1), I next performed a pathways analysis of differentially regulated genes in KO iPSC clones compared to the parental wild-type iPSC line. The pathway analyses using ORA did not reveal any significant deregulated pathways in the BLM-KO cells. Therefore, we reanalyzed the scRNAseq data using gene set enrichment analysis (GSEA) which resulted deregulated pathways in the BTRR complex-deficient cell lines. According to pathway analysis using GO terms for biological processes, only two terms, 'nucleobase metabolic process' and 'telomeric loop disassembly', were significantly deregulated (FDR<0.05) in the BLM-KO iPSC clone when compared to WT control cells. 'Telomeric loop disassembly' included genes e.g., *POT1, WRN*, and *BLM* as core enrichment, and was most probably detected as significant due to highly significant downregulation of *BLM* in the BLM-KO cells (Supplementary Table S2). Similarly, pathway analysis revealed 58 significant denominations for TOP3A-KO clone including 'SRP-dependent cotranslational protein targeting to membrane', 'mitochondrial respiratory chain complex assembly', and 'cotranslational protein targeting to membrane' which were the most significantly deregulated terms. The top 10 most significant GO terms for TOP3A-KO clone are given in Supplementary Table S3. The GO pathway analysis of the DEGs in RMI1-KO clone in comparison to the WT revealed only four significant GO terms for biological processes including 'cellular response to cadmium ion', 'pigment metabolic process', and 'response to copper ion' as top three results (Supplementary Table S4).

3.3.3.4 Determination of the effect of aphidicolin on gene expression

To determine the effect of aphidicolin on gene expression, we first compared single-cell gene expression profiles of aphidicolin-treated samples to the non-treated wild-type sample in order to analyze, if BTRR complex-deficient cells were more sensitive to replication stress than the wild-type cells. In total, we obtained a higher number of significantly deregulated genes in aphidicolin treated BTRR complex-deficient cells

when compared to their non-treated comparisons, which implied a possible effect of aphidicolin on gene expression. To decide whether the deregulation of these genes was a general effect of aphidicolin treatment on WT cells or a specific effect of aphidicolin on the BTRR complex-deficient cells, we also conducted the same analysis for the aphidicolin-treated wild-type cells in comparison to their non-treated counterparts. The highest number of differentially expressed genes in aphidicolin-treated samples was detected in TOP3A-KO cells while the lowest number was detected in BLM-KO cells (Table 3.6).

Table 3.6. Numbers of significant DEGs and GO pathways of from each comparison.

Compared cell line	Control	DEG number	Significant GO terms
isWT1-BLM-KO.1D1	isWT1	25	2
isWT1-BLM-KO.1D1_APH	isWT1	79	43
isWT1-TOP3A-KO.1E10	isWT1	117	29
isWT1-TOP3A-KO.1E10_APH	isWT1	368	218
isWT1_APH	isWT1	241	165

GSEA enrichement analysis was performed for GO terms of biological processes. Significant terms were counted for FDR<0.05.

Next, we performed the GSEA using the identified DEGs for a pathway analysis via GO terms for biological processes. This analysis revealed an enrichment of deregulated genes in pathways related to mitochondrial functions, and an enrichment of downregulated genes in chromatin-related pathways in all three comparisons (Supplementary Table S5, S6, and S7). Furthermore, ‘nucleosome assembly’ and ‘chromosome condensation’ pathways were significantly enriched for downregulated genes in all three aphidicolin treated samples, i.e., WT, BLM-KO, and TOP3A-KO when compared to the non-treated wild-type cells. Chromosome condensation was particularly interesting as this pathway is directly related to altered condensin expression that was observed in scRNAseq data of BS patient-derived fibroblasts. Therefore, we concentrated on the analysis of cellular pathways in which SMC2/4 take part, and we obtained various denominations for aphidicolin-treated TOP3A-KO, BLM-KO, and wild-type samples (Table

3.7). Aphidicolin treated BLM-KO cells revealed only ‘chromosome condensation’ pathway as deregulated including *SMC2/4* enrichment while TOP3A-KO aphidicolin sample revealed 15 pathways with *SMC2/4* enrichment out of 218 significant GO terms. To further define, whether this was a general effect of aphidicolin treatment or specifically caused by an increased sensitivity of BTRR complex-deficient cells to aphidicolin, we compared the obtained results of the pathway analyses with the comparison of aphidicolin-treated wild-type cells vs. untreated wild-type cells. We observed that the only pathway which was shared among the three aphidicolin treated samples (WT, BLM-KO, and TOP3A-KO) was chromosome condensation. Therefore, the effect of aphidicolin on BLM-KO sample on deregulation of chromosome condensation pathway was a global effect on the parental cell line but not due to the BLM knockout condition. Next, we compared the stressed TOP3A samples to stressed wild-type cells and observed that pathways such as ‘nuclear division’, ‘chromosome segregation’, and ‘sister chromatid segregation’ were shared between these analysis pairs pointing towards a general effect of aphidicolin on the deregulation of the *SMC2/4*-including pathways in iPSCs, (Table 3.7). Pathways, which were detected only in the aphidicolin-treated TOP3A-KO cells, were mainly related to meiosis which was unconclusive since the cells were not germline cells. Terms such as ‘cell division’ and ‘DNA confirmation change’ were only detected in the aphidicolin-treated TOP3A-KOs indicating a possibly increased sensitivity of TOP3A-KO cells to aphidicolin.

Overall, analysis of single-cell transcriptome data of the knockout iPSC clones resulted in only a low number of deregulated genes. Analyses of aphidicolin-treated samples revealed a possible aphidicolin effect on genes related to chromatin and chromosome condensation in iPSCs. The BTRR complex-deficient cell lines showed a possible sensitivity to replication stress, and analysis of scRNAseq data revealed significant deregulation of pathways related to chromatin as a specific effect of the aphidicolin treatment, particularly in TOP3A-deficient cells.

Table 3.7. *SMC2/4* containing deregulated GO terms in aphidicolin treated iPSCs and the respective enrichment scores according to GSEA.

GO terms (biological processes)	BLM-KO_APH vs. WT	TOP3A-KO_APH vs. WT	WT_APH vs. WT
Chromosome condensation	-0.533 *	-0.698 **	-0.629 *
Nuclear division	ns	-0.491 ****	-0.397 **
Organelle fission	ns	-0.365 ****	-0.373 **
Mitotic nuclear division	ns	-0.499 ****	-0.427 **
Nuclear chromosome segregation	ns	-0.495 ****	-0.421 **
Chromosome segregation	ns	-0.462 ****	-0.401 **
Sister chromatid segregation	ns	-0.492 ***	-0.439 **
Mitotic cell cycle process	ns	-0.376 ***	ns
DNA packaging	ns	-0.515 ***	-0.421 *
Mitotic sister chromatid segregation	ns	-0.498 ***	-0.432 **
Cell division	ns	-0.387 **	ns
Meiotic cell cycle	ns	-0.507 **	ns
DNA conformation change	ns	-0.421 **	ns
Meiotic cell cycle process	ns	-0.521 **	ns
Meiotic nuclear division	ns	-0.537 **	ns
Meiotic chromosome segregation	ns	-0.600 **	ns
Reproductive process	ns	-0.346 **	ns
Reproduction	ns	-0.345 **	ns
Chromosome organization involved in meiotic cell cycle	ns	-0.578 **	ns

The terms are listed according to FDR values of TOP3A-KO_APH vs WT comparison, except the first row. ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05, ns: non-significant.

4 Discussion

The human genome is constantly exposed to deleterious genotoxic events that, if not repaired properly, can give rise to genomic instability. Among the different types of DNA damage, DNA double strand breaks (DSB) are considered one of the most deleterious types of DNA damage, and two major cellular pathways have been developed in higher eukaryotes to repair DSB and protect the genomic integrity of cells: non-homologous end joining (NHEJ) and homologous recombination (HR). Different processes in the HR-based DSB repair are mediated by the BTRR protein complex which is formed by BLM, TOP3A, RMI1, and RMI2 proteins. Pathogenic variants in genes encoding the members of the BTRR complex have been associated with Bloom syndrome, a rare autosomal recessive, genomic instability syndrome characterized by primary microcephaly, developmental delay, photosensitivity, immunodeficiency, and a predisposition to specific cancer types.

During my PhD research, I was able to expand the phenotypic and mutational spectrum of Bloom syndrome. I assessed the phenotypic characteristics of eight BS patients with typical mutational variants including one novel pathogenic variant in BLM. Secondly, I defined the transcription profile changes caused by BLM deficiency on a single-cell level using patient-derived fibroblasts. Significant overexpression levels of various genes from the Fanconi anemia (FA) pathway, genes associated with primary microcephaly, and genes encoding condensin I/II complex were detected in BLM-deficient cells providing novel insights into molecular pathomechanisms of BS. Moreover, protein expression profiles of BS patient-derived fibroblasts were analyzed and the obtained data were compared to the findings of scRNAseq. Finally, I generated isogenic iPSC

lines harboring biallelic loss-of-function mutations in the genes encoding the members of the BTRR complex, namely, *BLM, TOP3A,* and *RMI1*. I systematically characterized the BTRR complex-deficient iPSCs in terms of impaired molecular functions such as sister chromatid exchange rates, chromosome segregation errors, and transcriptional changes. These cellular systems provide excellent sources for further elaboration of the BTRR complex deficiency and BS pathomechanisms in terms of impaired genomic integrity and altered transcription profiles.

4.1 Phenotypic distinctions of BS patients

Thanks to rapidly evolving NGS-based technologies, disease-causing variants can be more efficiently and rapidly detected when compared to the past. These technologies can include multigene panel testing and WES, which are commonly used in both diagnostic and research settings. With the advantage of WES, mutational spectra of disorders can be expanded like in BS. Initially, only autosomal recessive mutations in *BLM* were associated with BS phenotype; however, by the application of WES in the diagnoses of the patients presenting with BS phenotypes, LoF mutations in two additional genes, namely, *TOP3A* and *RMI1*, were also identified recently [74]. The *RMI1* variant as well as several pathogenic variants in *BLM* were identified by RG Wollnik before this doctoral project.

4.1.1 Genotype-phenotype correlation in BS

Defining phenotypic and mutational signatures of rare monogenic diseases can be challenging due to the limited number of affected individuals. BS is an autosomal recessive rare disease with approximately 300 affected individuals reported so far. The majority of the reported BS patients carry compound-heterozygous or homozygous LoF mutations in the gene [54]; yet, patients with pathogenic variants in *TOP3A*, *RMI1*, and *RMI2* have been described very recently as well [73, 74]. Depending on the causative

mutation and the affected gene, the BS phenotype can slightly vary among individuals diagnosed with this rare disease. Therefore, the molecular diagnosis has high importance in BS for better genetic counselling and patient management.

Eight patients presenting with a BS/BS-like phenotype were clinically evaluated in cooperation with Prof. Dr. Nursel Elcioglu (Marmara University, Istanbul, Turkey) and molecularly diagnosed at the Institute of Human Genetics (UMG, Göttingen) before the start of this doctoral thesis. Homozygous LoF mutations were identified in 6/8 patients in *BLM* and 2/8 patients in *RMI1*. One of the identified *BLM* variants (c.572_573del; p.Arg191Lysfs*4) was a novel variant that was not reported in any database of pathogenic human variants. This mutation was predicted to introduce a premature stop codon close to the N-terminal end of BLM, resulting in a truncated protein lacking the domains important for the enzyme activity of the helicase. Therefore, this novel variant was predicted as disease-causing by various in silico prediction tools and reported in a scientific report during the course of this doctoral thesis [158]. We scrutinized the clinical presentations of each patient and compared the phenotypic characteristics together with the respective mutational signatures of these eight BS patients. Primary microcephaly and global developmental delay were shared clinical features between the patients harboring homozygous *BLM* mutations and patients harboring the homozygous *RMI1* mutation. On the other hand, dermatological characteristics and immunodeficiency were only present in patients with pathogenic variants in *BLM* (Table 1.1). Hence, an overall milder phenotype was observed in *RMI1*-associated patients in comparison to *BLM*-associated patients. So far, our two patients are the only patients reported with a BS-like phenotype having a homozygous mutation in the *RMI1* gene. Therefore, further phenotypic comparisons between patients with *RMI1* mutations will be important in the future, when additional patients will be described.

Still, the milder phenotype observed in the *RMI1*-associated patients was striking. BLM and RMI1 are both essential parts of the BTRR complex that plays an important role in the dissolution of HR DNA repair intermediates [99, 190]. Yet, BLM has additional functions in the maintenance of genomic stability, which are independent of its role within the BTRR complex. For instance, BLM recruits FANCM to the stalled replication forks

for the proper management of the replication fork [191]. The localization of BLM to the stalled forks is dependent on FANCD2 on the same pathway where both work together for the restart of the stalled forks and suppression of new origin firing [140]. In contrast, the known functions of RMI1 are strongly dependent on the BTRR complex. RMI1 protein binds to BLM and TOP3A to stimulate the decatenation activity while interacting with RMI2 through their OB domains as well [72, 126]. Hence, RMI1 is important for the stability of the BTRR complex [130]. Taken together, the mild phenotype observed in *RMI1*-associated patients in terms of immunodeficiency and skin lesions might be explained then by the additional roles of the BLM, independent of the BTRR complex. Similar to RMI1 deficiency, patients with homozygous loss of RMI2 were reported to present with a milder phenotype than the canonical BS phenotype [73]. Hudson et al. speculated that the overall moderate BS phenotype observed in patients was due to the non-enzymatic roles of RMI2 within the BTRR complex, which similarly applies to RMI1 and the two affected patients we identified.

In addition, previously reported *TOP3A*-associated patients presented with prenatal-onset growth retardation, café-au-lait spots, and dilated cardiomyopathy [74]. Mitochondrial DNA depletion in muscle tissue was also detected in two of the *TOP3A*-associated patients. TOP3A is a part of the BTRR complex, however, like BLM, it has additional functions independent of the BTRR complex, i.e., in mitochondria [192]. TOP3A is important for mtDNA replication and maintenance through its decatenation activity [122]. The separation of replicated mtDNA is considered to be dependent on the mitochondrial TOP3A isoform which putatively interacts with Twinkle helicase, a mitochondrial equivalent of BLM helicase [193]. It will be interesting in the future to specifically analyze the effects of TOP3A dysfunction on mtDNA stability and somatic mtDNA mutation load. This might explain the cardiac phenotype and the mtDNA depletion in muscle tissue of *TOP3A*-associated patients. Overall, the comparison of the clinical features of eight BS/BS-like patients with previously published cases of BS phenotype indicates that the clinical features and the severity of BS strongly depend on the affected gene within the BTRR complex, which is particularly important for genetic counseling of affected families and clinical management of patients with BS.

4.1.2 Genomic instability and cancer predisposition in BS patients

Cancer predisposition is a hallmark of BS that has a severe impact on the lifetime and quality of the affected individuals. Approximately half of BS patients die from cancer based on the clinical data provided by the BS Registry [194]. Hematologic malignant neoplasms (leukemia and lymphoma) were observed in BS patients more commonly than solid tumors, among which colorectal, breast, and oropharyngeal tumors were commonly detected [68]. Among the eight BS patients described in this thesis, malignancies were not observed so far. This might be due to the young age of the patients which varied between 1.5 years and 16 years by the time of physical examination. One of our patients developed a Wilms tumor at the age of 5 which was operated subsequently. Although the incidence of Wilms tumor is very low in BS patient cohorts with eight reported patients [57], it is more frequently observed in BS patients than in the rest of the affected children where its incidence is 1 in every 10,000 children below age 5 [195–197]. Here, besides this single patient, no additional malignancies were observed in our patient cohort. Generally, recurrent cancer screens are advised for BS patients starting in adolescence as the average age at which malignancy occurs in BS patients is around mid-twenties [66]. Cancer treatment and cancer therapy in BS are further complicated by the fact that cancer therapeutics often includes radiation therapy and alkylating drugs [56]. These therapeutic options themselves affect the DNA integrity and induce DNA alterations. As BS manifests an impairment of DNA repair mechanisms, classical cancer management in BS might act as a double-edged sword and therefore should be advised carefully in BS patients [57]. Interestingly, in BS patients with pathogenic variants in *TOP3A* as well as *RMI1* and *RMI2*, cancer was not diagnosed. Currently, we do not know if this observation is caused by the limited number of *TOP3A*, *RMI1*, and *RMI2*-associated patients and the young age of affected individuals, or based on the different functions of these proteins within and beyond the BTRR complex.

Cancerous cells usually show high levels of genomic instability and their survival is dependent on DNA repair mechanisms [198]. Recent advances in cancer therapy sug-

gested drugs that chemically target DNA replication and repair mechanisms [199]. After the discovery of PARP and topoisomerase inhibitors as therapeutic agents, novel targets of the DNA repair pathway such as DNA helicases start to emerge [200]. A small molecule inhibiting BLM, ML216, was first described by Nguyen et al. as a potential anticancer agent [201]. They reported that inhibition of BLM by ML216 caused a significant decrease in cell proliferation and increased the sensitivity to aphidicolin resulting in DNA damage in vitro. Following this, another chemical compound (isaindigotone-derivative compound 29) was identified with a stronger inhibitory effect on BLM activity by disrupting the binding of BLM to DNA [202]. Recently, BLM was targeted in breast cancer cells by a similar small-molecule approach and cell expansion was decreased [203].

Interestingly, overexpression of BLM helicase was detected in various types of malignancies of lymphoid and epithelial origin correlating with the proliferative state of cells [204]. In breast cancer, upregulation of BLM was linked to poor survival and increased amounts of *BLM* in the cytoplasm was associated with aggressive phenotypes [205]. The BTRR complex activity was also observed during alternative lengthening of telomeres, which is an HR DNA repair-based pathway for telomere maintenance, used in approximately 10 % of cancers [206]. Loss of BLM function then can drive instability and apoptosis in cancerous cells. In addition to the expression and roles of BLM in cancer, overexpression of TOP3A was also correlated with worse overall survival in non-small-cell lung and lung adenocarcinoma patients [207]. Gene knockdown of *TOP3A* in gastric cancer cells resulted in a decrease in cell viability and proliferation rates [208]. Therefore, targeting TOP3A in cancerous cells could be further elaborated as a therapeutic strategy as well. Overall, we believe that our research on the BTRR complex is not limited to BS, but it might provide additional insights to define novel targets for therapeutic strategies and translational approaches in cancer.

4.2 Alteration of gene and protein expression signatures in BS

During the course of this doctoral thesis, I aimed to characterize the aberrated cellular signatures in BS caused by BLM deficiency. Therefore, I evaluated gene and protein expression profiles of three different BS patient-derived fibroblast samples in comparison to wild-type control fibroblasts. Single-cell transcriptome sequencing and proteomics experiments were conducted for defining and confirming the gene expression alterations in BS, respectively, and I further conducted Western blots to confirm the results obtained from the scRNAseq and proteomics experiments. The results that were obtained from these experiments will be discussed in the following sections of this chapter.

4.2.1 Single-cell transcriptome alterations observed in BS patients

Although the roles of the BLM helicase on DNA repair and replication processes as well as the pathomechanisms and altered molecular pathways in BS have been well-described, the effect of BLM deficiency on transcriptional profiles has not been equally defined. Montenegro et al. reported on gene expression profiles in BS patient-derived samples using bulk RNA sequencing and observed various deregulated genes that take part in immunological pathways and apoptosis control [209]. Additionally, overexpression of interferon-stimulated genes in BS patient-derived fibroblasts was detected by RNA microarray [210]. An extensive study using BS patient-derived and, additionally, shRNA-mediated BLM depletion in fibroblasts, reported transcriptional changes in genes involved in cellular pathways including ‘cell proliferation’, ‘survival’, and ‘molecular mechanisms of cancer’ [161]. The authors observed that the transcriptional changes of genes were related to the presence of G4 structures at the transcription start sites which serve as targets of the BLM helicase. With the recent advances in single-cell sequencing techniques, G4 motifs in transcribed genes were determined as sister chromatid exchange locations in BS cells [143]. In the reported study, RNAseq was ap-

plied; however, the authors neither commented on the biological significance of the transcribed genes nor compared the transcriptional changes in BS cells to control cells. Therefore, the transcriptional profiles of BLM-deficient cells have not been extensively studied thus far.

During this doctoral thesis, we performed single-cell transcriptome sequencing (scRNAseq) on BS patient-derived fibroblasts and defined the transcriptional profiles in BLM deficiency. Single-cell RNA sequencing not only allows the detection of mild transcriptional changes but also can define the cell cycle stages according to specific markers; which are two aspects of this technique that are not possible to achieve using bulk RNAseq. It is important to achieve high-quality data, especially in a scRNAseq experiment for proper statistics due to the nature of the experiment. In collaboration with the NGS Integrative Genomics Core Unit (NIG, UMG), we used the ICELL8 single-cell system for cell sorting which was an established system performing approximately 2000 single cells for sequencing in one experiment. Single cells were dispensed to a nanowell-containing microchip and the content of the wells was assessed by the imaging system of ICELL8 [20]. This technology allowed us to visually select the wells containing only single cells and not duplicates/triplicates which cannot be differentiated in other single-cell platforms such as the 10x Genomics Chromium platform. The number of wells containing single cells could be further improved using a different type of cell sorting instrument such as CellenONE which evaluates the content of drops containing single cells by imaging before dispensing [20]. This system allows the detection of full-length RNA as well as several types of RNA molecules such as long non-coding- or microRNAs.

The quality of the sequencing data was robust and suitable for further analyses. The bioinformatics analyses revealed more than 2600 significantly differentially expressed genes (DEG) in the BS fibroblasts when compared to WT fibroblasts. The analyses of DEGs further showed deregulated pathways that were highly related to BS pathomechanisms such as ‘mitotic sister chromatid segregation’ and ‘cell division’. We focused on the expression of known interaction partners of BLM, e.g., from the Fanconi anemia pathway, which was a highly deregulated pathway detected by the ORA analysis. Addi-

tionally, we also observed an upregulation of various genes that are clinically associated with microcephaly, and overexpression of condensin I/II complex genes which was, so far, not observed in BLM deficiency. The results provided novel insights into changes in the gene expression profiles caused by BLM dysfunction in BS patient-derived cells.

4.2.2 Gene expression profiles of Fanconi anemia genes in BS

Biallelic LoF mutations in several genes encoding proteins of the Fanconi anemia (FA) pathway result in FA disorder, which is an autosomal recessive disorder, presenting overlapping clinical characteristics with BS such as growth retardation, microcephaly, immunodeficiency, and cancer predisposition. We obtained highly significant deregulation of the FA pathway in our scRNAseq data of BS patient-derived fibroblasts. This could point out commonly affected pathways in BS and FA, supported by the overlapping clinical phenotypes. It has been described that BLM has roles in both upstream and downstream of the FA pathways during the management of stalled replication forks [211]. FA proteins such as FANCM and FANCD2 further interact with the BTRR complex to promote HR DNA repair of ICLs [139]. Therefore, BLM has numerous interactions with FA proteins both dependent and independent of the BTRR complex within the DNA repair pathways.

As it was reported previously that the expression levels of FA pathway genes increase in S-phase [212], we first analyzed the cell cycle stages in our scRNAseq data using the Seurat model for bioinformatics analyses. We assessed the cell cycle stages of each sample and observed that there was no significant difference in the expression levels of cell-cycle specific markers between BS and control samples. All samples had similar ratios of cells at each cell cycle stage, and overall, most of the cells were in the G2/M phase. Hence, we eliminated a possible cell cycle effect on the differences in the gene expression levels of FA proteins in BS samples through bioinformatics analyses of scRNAseq data.

Several FA genes had significantly higher gene expression levels in BS single-cells than in WT cells. (Supplementary Table S1). This could point out to a possible regulatory

effect of BLM on the expression of FA genes. The most upregulated gene from the FA pathway was the *RPA3* gene which encodes a subunit of the heterotrimeric RPA protein. Also, the two additional members of the RPA family, *RPA1* and *RPA2*, were significantly upregulated in BS single cells. RPA is a single-stranded DNA binding protein that has essential roles in DNA replication, repair, recombination, and cell cycle checkpoints [213]. RPA interacts with both BLM and RMI1, stimulating the dHJ dissolution by the BTRR complex [214]. Furthermore, binding of RPA to BLM is required for proper localization of BLM to stalled replication forks induced by replication stress [131]. Taken together, BLM and RPA not only interact at the protein level but BLM deficiency might also result in the upregulation of RPA genes' expression based on our scRNAseq analyses.

FANCD2 was also among the most upregulated genes of the FA group. FANCD2 is a peculiar member of the FA pathway through its independent functions in DNA repair pathways than the rest of the FA proteins. In the presence of replication stress, several interactions between BLM and FANCD2 were documented for suppressing new replication origin firing and for the proper restart of the stalled replication forks [140]. Additionally, FANCD2/FANCI foci were observed at the tips of the UFBs originating from common fragile sites which are frequently observed in BS cells [184]. Common fragile sites can be difficult to replicate due to the secondary structures where their resolution is dependent on BLM activity [215]. At these sites replication forks might stall and require FANCD2/FANCI as protection from degredation during the S-phase, which in turn can result in trapped FANCD2/FANCI foci related to UFBs in mitosis due to chromosome condensation [216]. Stalled replication forks are further managed by FA core complex, especially FANCM, which is a known interaction partner of BLM. In our scRNAseq data, FANCM was significantly overexpressed in BLM deficient cells, supporting the known interaction of BLM with FANCM [139].

In general, the upregulation of genes involved in DNA repair pathways such as the FA pathway could be caused by the overall genomic instability profile of BS cells. Impaired DNA repair mechanisms, as a result of the BLM deficiency, could induce the accumulation of somatic mutations and DNA adducts which might trigger, in turn, higher expres-

sion levels of proteins involved in DNA repair such as the FA proteins. The upregulation of the FA pathway could also be interpreted as a compensation mechanism based on the impairment of BLM-dependent DNA repair pathways. A compensation mechanism was suggested for BRCA1- or BRCA2-deficient cells as they were hyperdependent on other DNA repair mechanisms [217–219]. The known interactions of BLM with the FA proteins could further explain the deregulation of the FA pathway; however, the exact answer for the overexpression of FA genes in BLM deficiency remains elusive. Overall, our scRNAseq data confirmed a link between the BLM and the expression of FA genes, especially RPA and FANCD2, but the exact interactions and mechanisms remain as remarkable questions for BLM deficiency.

4.2.3 Expression of genes clinically associated with microcephaly

One of the most prominent clinical characteristics of BS is primary microcephaly. Primary microcephaly is characterized by an occipitofrontal head circumference (OFC) of more than 3 SD below the mean of the control group [220]. The OFC values of our BS patients deviated from normal values by approximately 5 SD below normal. Furthermore, most BS patients do not have intellectual disability although microcephaly, in general, is accompanied by mild to severe intellectual disability with approximately 65% of microcephaly cases presenting with intellectual disability in several other disorders [221]. Based on the causative gene and cellular roles of the encoded protein, various pathomechanisms have been identified to cause microcephaly.

One hypothesis on how primary microcephaly is caused is based on an imbalance between symmetric and asymmetric cell division of neural precursor cells leading to premature neuronal cell maturation and early depletion of the pool of neural progenitor cells [43]. Further, the differentiation process of neurons during brain development can also be directly affected, thereby leading to severe microcephaly [38]. All these mechanisms are affected by the impairment of various factors related to cell division, i.e., centriole biogenesis, spindle assembly, or kinetochore attachment [38]. Mutations in genes affecting these cellular processes such as *CDK5RAP2, ASPM*, and *CENPE*

were described in patients presenting with primary microcephaly. Additionally, defective chromosome condensation can also result in microcephaly as observed in patients with microcephaly carrying mutations in, e.g., members of the condensin I/II complexes [173]. Currently, 27 genes are clinically associated with primary microcephaly [38]. I analyzed the expression profiles of these genes in BS patient-derived fibroblasts and I could show that a substantial number of these microcephaly-associated genes were significantly deregulated in BS single-cells based on our scRNAseq data. Moreover, most of the deregulated genes were overexpressed except for one gene, *CDK6*. Previously, downregulation of genes with G4 structures near the transcription start site in intron 1 was suggested [161]. Hence, we searched for G4 structures at the transcription start site of *CDK6*; however, no G4 structures as potential BLM helicase substrates were found in this region.

Most of the genes that are clinically associated with primary microcephaly encode proteins, which have roles in mitotic processes as well as DNA replication and repair processes. Here again, the overexpression of these genes could be a result of compensation mechanisms against the BLM deficiency. In general, we hypothesize common affected molecular pathways in both BS and MCPH although exact links between BLM deficiency and the deregulation of microcephaly-associated gene expression remain to be further elucidated.

4.2.4 Novel genetic interaction of condensin I/II complex genes and BLM

Single-cell transcriptome sequencing is a powerful tool for discovering novel gene expression signatures and thereby providing novel insights into pathomechanisms of diseases. Apart from deregulated gene expressions of known interaction partners of BLM, we analyzed the scRNAseq data for deregulated genes that, so far, have not been identified as BLM interaction partners. Among the genes analyzed, we detected significant overexpression of genes encoding the members of the condensin I/II complexes in BLM deficiency (Table 3.2). The SMC proteins, namely, SMC2 and SMC4 are both members of the condensin I/II complexes, and their gene expression values were the most highly

increased ones among the members of condensin I/II complexes in BS single cells. As confirmation, we were able to show a tendency of higher protein expression for SMC2 and SMC4 in the BS patient-derived cells by Western blot analyses. Interestingly, three of the non-SMC members of condensin I/II complexes, i.e., *NCAPD2, NCAPH*, and *NCAPD3* were already detected as highly expressed in BS when we previously analyzed the scRNAseq data for the microcephaly-associated genes. In total, seven out of eight genes encoding the members of the condensin I/II complexes were significantly overexpressed in BS single cells according to the scRNAseq analysis, which was a novel observation in BLM deficiency.

Condensin I/II complexes are multiprotein complexes with major roles in chromosome condensation and segregation [222]. They fold the DNA into loops forming the mitotic chromosomes while controlling the organization of the chromosomes and regulating DNA transcription during interphase [172]. Interestingly, chromatin bridges and UFBs were observed in cells of patients with primary microcephaly harboring biallelic mutations in genes encoding condensin I/II complex components such as *NCAPD2, NCAPH*, and *NCAPD3*, a genetic syndrome called 'condensinopathy' by Martin et al. [173]. Similarly, *NCAPG2* was identified as the causative gene for severe neurodevelopmental defects; and patient fibroblasts showed increased rates of chromatin bridges [174]. Interestingly, these kinds of anaphase bridges were also documented in BS and the BTRR complex deficiency. Of note, there is currently no patient phenotype described with *SMC2* or *SMC4* mutations.

Overlapping cellular phenotypes observed in condensinopathies and BLM deficiency are important remarks which could point towards common molecular pathways affected in both disorders. We evaluated our scRNAseq data such that BLM deficiency regulates the gene expression of condensin I/II complexes. However, the exact molecular link between the condensins and the BLM functions remains an unanswered question. One possible explanation could be based on the roles of condensin I/II complexes on DNA damage repair. Condensin I has roles in repair of the SSBs while condensin II is involved in DSB repair through HR [223–225]. Increased somatic mutation load in BS cells could require upregulation of DNA repair mechanisms which could affect the

expression of condensins as well. Furthermore, the role of condensin II complex in HR DNA repair could explain the overexpression of condensin genes as a compensation mechanism against BLM deficiency although the exact role of condensin II in HR is yet unknown.

4.2.5 Strategies for confirmation of the scRNAseq results

We obtained highly significant expression differences of genes related to BS pathogenesis by scRNAseq analyses. To confirm the observed differences and their consequences, we analyzed the effects of BLM deficiency on protein level using Western blot and mass spectrometry analyses. The protein expression of distinct proteins were directly analyzed via Western blot analysis using BS patient-derived and WT fibroblast cell lysates aiming to detect overexpression of various proteins detected in scRNAseq data. In Western blot analyses, we were successful in detecting SMC2 and SMC4 proteins in the cell lysates of the fibroblasts providing a verification of the overexpression of SMC2 and SMC4 transcripts as detected in scRNAseq data (Figure 3.13). On the other hand, various trials for the detection of several FA proteins such as FANCD2, FANCM, and BRCA1 failed to give results on WB analyses of BS fibroblasts. This could be due to the low protein amounts in the cell lysates concerning the respective cell cycle stages since FA proteins are mainly expressed in S-phase. It could be possible that the number of cells in S-phase at the time of harvest were low and not enough protein was extracted from unsynchronized cells. A cell synchronization assay could be used to test this in the future. Antibody specificity could also provide a further explanation for the challenges in detection of the FA proteins.

Further, as single-cell proteomics is a challenging field and high-throughput methods for detecting the protein content of single cells are not feasible yet [226], we assessed the translational profiles of samples using bulk cells of fibroblasts by mass spectrometry (MS) analysis in collaboration with Markus Räschle and Angela Wieland (Technical University of Kaiserslautern). This analysis was performed to determine global protein expression alterations in the BS patient-derived fibroblasts and used to confirm the gene

expression differences obtained by the scRNAseq data as well as to find novel targets of BLM-dependent protein expression. The analyses of the three biological replicates of each sample revealed 149 significantly deregulated proteins in BS samples. This result was a lower number than DEGs obtained from scRNAseq data, which could be due to the different chemistry and technology between the two experiments. Besides, a higher DEG number in transcriptional level could be plausible considering not all the transcripts would be translated into proteins. We analyzed the MS data for the proteins from the FA gene group, microcephaly-associated genes, and genes encoding for members of the condensin I/II complexes, and observed only a minor number of the detected peptides linked to proteins of these groups (Figure 3.14). The expression levels of most of the detected proteins from the respective gene groups were not significantly altered in BS samples when compared to the WT samples. Three subunits of the RPA protein were detected, although not significant in the MS data, while these transcripts were among the top most significantly overexpressed genes from the FA group in the scRNAseq data. SMC2 and SMC4 proteins were detected in the MS data as well as NCAPD2, NCAPH, and NCAPG, although the differences in the abundance of these proteins were not significant in contrast to the significant overexpression of transcripts of the corresponding genes. 8 proteins out of 27 were detected from the microcephaly-associated gene group and only one of them, the CIT protein, was significantly overexpressed in BS samples according to the MS data. CIT protein is a serine/threonine-protein kinase that has roles in cell division and cytokinesis; therefore, its abundance could likely be affected by BLM deficiency.

Overall, we observed considerable differences when comparing the scRNAseq results to the MS analyses. Currently, the reason for this discrepancy is unclear. One possible reason might be the comparison of single-cell RNAseq data to bulk MS data. Single-cell RNAseq technologies are particularly powerful when analyzing heterogeneous cell populations and detecting mild gene expression changes. As we aimed to detect transcriptional alteration in BLM deficient cells independent of possible cell cycle influences, scRNAseq enabled us to determine the cell cycle status of each analyzed cell by detection of certain cell cycle marker gene expressions, and simultaneously identify targets

of BLM-dependent transcriptional regulation. In the case of the MS analysis, we were not able to perform this assessment on a single-cell level, but only determined protein expression profiles of bulk samples. Hence, varying cell cycle stages might have caused no detection of various cell cycle-dependent proteins such as the FA proteins or BLM itself. Nevertheless, bulk proteomics is an important quantitative high-throughput method that can shed light on protein-protein interactions as well as post-translational modifications [227].

4.3 Isogenic BTRR complex-deficient iPSCs as a cellular model for BS

Although phenotype-genotype correlations are eased by the advances in NGS-based technologies, the functional studies of identified genes and mutations still present a bottleneck for gene associations with the corresponding disease phenotypes. Nevertheless, the association of the BTRR complex deficiency to BS is solid from the genetic point of view, and functional studies have been well established for BS. However, modeling rare diseases such as BS might face challenges in animal models or patient-derived models due to various issues. In this thesis, the induced pluripotent stem cell (iPSC)-based models provided novel insights into disease modeling applications and various aspects of the current state of BS models are discussed in the following sections

4.3.1 iPSC-based disease modeling and current BS models

The discovery of reprogramming somatic cells into a stem cell-like state by the introduction of transcription factors was a breakthrough in medical research [228]. These transcription factors, also known as Yamanaka factors (Oct3/4, Sox2, Klf4, and c-Myc), are introduced into differentiated cells such as fibroblasts or peripheral blood mononuclear cells for the induction of stemness [229]. This discovery brought to Yamanaka a Nobel Prize in Physiology or Medicine in 2012, and the applications of iPSC technology

expanded greatly in the next decade. The possibility of generating iPSCs from differentiated tissues offered a better understanding of cellular mechanisms in health and disease, and created an advantageous cellular model, especially when compared to primary patient-derived cell models such as fibroblasts.

Disease modeling is, hence, one of the widely used applications of iPSC technology. Drug screening approaches are eased by the use of patient-derived iPSC and tissue cultures derived from the iPSCs [230]. Furthermore, disease modeling via iPSCs provides advantages when compared to disease modeling in animal models, especially mice. It is less time-consuming, easier, and cheaper in most cases when compared to mice models. Besides, genetic differences between mouse strains and humans present challenges for various disease models. For instance, initial trials to produce a *Blm* knockout mouse were not successful as the mice died by embryonic day 13.5 [231]. Later on, Luo and colleagues reported on a Blm-negative viable mouse, and one-third of the mice developed cancer by 20 months [232]. Mice for knockout *Top3a* died at the embryo stage, [233]. Depletion of both *Rmi1* alleles in mouse also resulted in early embryonic lethality as reported by two groups [132, 133]. Therefore, currently, there are no mouse models for members of the BTRR complex except for BLM.

Disease modeling by iPSC culture can be initiated by two general approaches, i.e., reprogramming of patient-derived cells into iPSCs or genome editing of isogenic control iPSCs. Patient-derived iPSCs allow studying complex disease models with multiple affected loci as well as specific patient mutations [234]. However, genetic background differences present a challenge for patient-derived iPSCs in terms of possible biased disease-related phenotypes [235]. Genome editing approaches, in general, can be more feasible for a knockin or knockout approach than the generation of patient-derived iPSCs; although genes causing late-onset disorders or aging-related disorders could be difficult to study by genome editing on iPSCs as well. Still, differentiation opportunities of iPSC culture can provide insights into disease mechanisms using tissues of interest.

iPSC-based disease modeling can be challenged further by several additional factors. First, reprogramming approaches result in a loss of the epigenetic signatures in iPSCs,

which can create a bottleneck for the study of complex epigenetic disorders such as Parkinson's disease [236]. The reset of epigenetic markers of somatic cells in iPSCs can be also partial, which could further complicate differentiation and overall disease modeling [237, 238]. Nevertheless, successful differentiation approaches can recreate the epigenetic state of cells in some cases. For instance, patient-derived fibroblasts of Hutchinson-Gilford progeria syndrome (HGPS) that were reprogrammed to iPSCs showed almost the same levels of pluripotency, transcriptional, and epigenetic profiles as the non-affected reprogrammed iPSCs; yet, they recapitulated disease profile once they were differentiated into muscle cells [239]. Therefore, iPSC-based disease models should be considered in combination with a differentiation strategy although this can be still challenged by the initial epigenetic markers of the donor cells.

4.3.2 Genome editing by CRISPR/Cas9 on iPSCs to obtain isogenic KO cell lines

In the field of rare diseases, disease modeling strategies through reprogramming of patient-derived material into iPSCs can be challenging due to the low numbers of affected individuals and restricted availability of patient material. To overcome these limitations, genome editing technologies can be performed to introduce patient-specific mutations, i.e., knockin or general knockout mutations to study genetic disorders on an isogenic base. From this perspective, CRISPR/Cas9 technology has brought considerable benefits to the genome editing field of iPSCs within the last years.

Introducing patient-specific or truncating mutations into control iPSCs is feasible and greatly facilitated by the power of CRISPR/Cas9 technologies. However, efficiencies of CRISPR/Cas9-based genome editing can vary and depend on the targeted gene as well as the specific alteration that has to be introduced. In some cases, e.g., when dealing with LoF variants in genes of interest, the introduction of truncating (mainly frameshifting) variants in the targeted genes is easier than generating a model with the patient-specific variant. A knockout strategy was applied in the course of my PhD thesis to generate iPSC lines with biallelic LoF mutations, which result in the dysfunction of

each of the four members of the BTRR complex. As previously described, there are no mouse models for members of the BTRR complex except for BLM. In addition, previously, only a single study has reported a BS patient-derived iPSC line and confirmed the cellular BS phenotype by increased SCE rates [183]. However, further studies including BS patient-derived or genome-edited BTRR complex-deficient iPSC lines do not exist in the literature.

In the course of my doctoral work, I aimed to generate an experimental system that allows the analysis. Although eight BS patients were molecularly diagnosed in our institute, patient material was not possible to obtain for the generation of BS patient-derived iPSCs. Therefore, I aimed to introduce biallelic LoF mutations in the genes that are associated with a BS/BS-like phenotype via CRISPR/Cas9 technology. Those genes were the members of the BTRR complex, namely, *BLM, TOP3A, RMI1*, and *RMI2*. We designed guideRNAs for each gene at close positions to the described patient mutations. After two rounds of transfections and analyses of a total of 284 individual cell clones, I was able to obtain at least one iPSC clone harboring biallelic LoF mutations in the separate genes of the BTRR complex, i.e., *BLM, TOP3A*, and *RMI1*. Using Western blot analysis, I observed decreased amounts of full-length protein in the heterozygous KO clones, and none of the compound-heterozygous KO iPSC clones expressed any full-length protein of the respective BTRR complex member after the genome editing (Figure 3.17).

According to the results of deep amplicon sequencing of the targeted region, the highest transfection efficiencies were obtained from the transfection of *BLM* guideRNA (first transfection, 55%; heterozygous clone transfection, 92%, Table 3.4). However, analyses of iPSC clones revealed that none of the analyzed clones in culture harbored biallelic LoF mutations after the first transfection, and most of the detected alleles in heterozygous state were causing in-frame deletions. After the second round of the transfection which was performed using a heterozygous clone with a LoF *BLM* mutation, the newly introduced alleles were again mostly in-frame deletions except for one clone. This could be due to the important cellular functions of BLM during DNA replication and repair processes. In other words, complete BLM deficiency could be poorly

tolerated in genome edited cells. Overall, we were able to obtain only one BLM-KO iPSC clone after genome editing experiments which was used in the following characterization experiments. Currently, it is unclear if the lack of a fully functioning BLM protein might have caused challenges for the cells to survive the transfection or single-cell state after the singularization procedure. Although patients with homozygous or biallelic LoF mutations in *BLM* can survive to life in the general population as BS phenotype, the total number of affected BS individuals does not exceed 300 patients, which was observed by James German as a lower proportion than expected in an autosomal recessive disorder [50]. Besides, miscarriages in families with affected BS patients were reported although a controlled study has not been conducted yet. Of note, the *Blm* knockout in mice was initially reported to result in embryonic lethality by Chester et al. in 1998 [231]. Therefore, we could speculate a tendency of a low cell survival rate considering the low amount of biallelic LoF mutations in *BLM* in laboratory settings. It could be summarized that the generation of LoF mutations in *BLM* similar to patient mutations using the laboratory settings was challenging due to severe impact of BLM deficiency on cells.

After double transfections using the crRNAs targeting *TOP3A* and *RMI1*, I obtained three TOP3A- and two RMI1-knockout iPSC clones with biallelic truncating mutations. These clones were growing normally in cell culture and had a high pluripotency state. On the other hand, the single BLM-KO clone was growing slower than the other two KO clones. We did not establish the cell cycle lengths for the generated KO iPSCs clones, but we could expect an overall increase in the cell cycle length of the KO clones, partially due to a longer S-phase. The impaired DNA repair mechanisms might cause an accumulation of somatic mutations leading to an increased length of S-phase for appropriate repair of these mutations. Increased rates of somatic mutations could affect apoptosis rates as well, which has been previously reported for BLM depleted cells [240] and could explain the slow growth of BLM-KO iPSC in culture within this doctoral thesis.

4.3.3 Genomic instability signatures of the BTRR complex-deficient iPSCs

BS is one of the well-documented genomic instability syndromes based on impairment of the DNA replication and repair processes [139]. The genomic instability in BS often manifests as chromosome breaks and segregation errors [241]. One of the well-established cellular phenotypes of BS cells is the increased sister chromatid exchange (SCE) rates due to the unresolved recombination intermediates between the homologous regions [141, 142]. SCE is a specific cytogenetic characteristic peculiar to BS as the BTRR complex is the most efficient protein complex that is capable of dissolving HR-mediated DNA repair intermediates into non-crossover products [72].

During the course of this doctoral thesis, I established a protocol based on the previously defined SCE assay methodologies for differential staining of the sister chromatids of iPSCs. I performed this SCE assay using the three KO-iPSC clones with the parental WT cell line and evaluated the exchange rates on each visualized metaphase. The SCE rates were almost 20 times increased in the BLM- and TOP3A-KO clones in comparison to the WT iPSCs. The increase in SCE rates in the RMI1-KO clone was approximately 8 times increased when compared to the WT cells. This indicated that the SCE rates of the RMI1-KO clone were not as high as in the BLM- and TOP3A-KO clones implying a milder effect of RMI1 deficiency on SCE rates. This condition could orginate from two possible explanations: (i) by the specific functions and roles of each component of the BTRR complex or (ii) by the existence a putative truncated RMI1 protein detected on the Western blot of RMI1-KO clone.

The first explanation is based on the fact that BLM and TOP3A have major enzymatic roles in dissolving DNA intermediates through the BTRR complex while RMI1 contributes to the maintenance of genomic instability by interacting with the rest of the BTRR complex members [241]. RMI1 interacts with BLM and TOP3A stimulating the decatenation activity, and is an essential component of the BTRR complex [125]. Further, BLM and TOP3A have additional cellular roles independent of the BTRR complex such as management of the stalled replication forks or mitochondrial DNA maintenance,

respectively. Hence, the cellular phenotypes of RMI1 knockout could be observed as milder than the other two knockout conditions, BLM- and TOP3A-KOs, due to varying roles of the components of the BTRR complex for different cellular functions.

Still, a previous study reported increased amounts of SCEs by RNA interference-mediated depletion of RMI1, and observed that the SCE rates of RMI1-depleted HeLa cells were similar to the SCE rates of BLM-depleted cells in the same manner [77]. Also, *Rmi1*-knockout in mouse had resulted in embryonical lethality [132, 133]. Therefore, we could secondly hypothesize a milder effect on SCE rates in the RMI1-KO due to a partially functioning truncated RMI1 protein which was observed in the Western blot analysis of these cells (Figure 3.17). The patients described in this thesis carrying homozygous RMI1 mutations had truncating mutations at a close position to the *RMI1* mutations introduced to the RMI1-KO iPSC clone. Due to the lack of patient material, we can only speculate if this mutation also leads to translation of a truncated but stable RMI1 protein, which is partially functional and might explain the milder phenotype observed in these *RMI1*-associated patients. In the literature, so far, no additional patients with biallelic *RMI1* mutations are described, which makes the characterization of the phenotypic differences between patients carrying BLM and RMI1 mutations currently challenging.

Chromosome segregation errors are another cellular characteristic of BTRR complex-deficient cells resulting from impairment of DNA repair and replication pathways. BLM, TOP3A, and RMI2 deficiencies were documented to introduce chromatin bridges and micronuclei in patient cells [73, 74, 141, 148]. Patient cells carrying pathogenic variants in *RMI1* were not characterized so far. Still, RMI1 depletion or knockout in established cell lines revealed increased chromatin bridges as well as increased mitotic errors [77]. We analyzed isogenic BTRR complex-deficient iPSC lines for chromosome segregation errors such as chromatin bridges that were previously reported for the BTRR complex deficient cells. We focused on the characterization and quantification of chromatin bridges and lagging chromatin that might occur during anaphase. Chromatin bridges are DAPI-positive DNA threads that are formed between the segregating sister chromatids. Lagging chromatin or laggards can occur once these chromatin bridges break or via chromosomes that are not correctly attached to microtubule spindles. However,

the exact mechanisms for lagging chromatin formation remain elusive [242].

Using BTRR-deficient iPSC clones, we observed that the numbers of events of both chromatin bridges and lagging chromatin were significantly increased in the BLM-KO iPSC clone compared to the WT control. Chromatin bridge counts of the TOP3A-KO clone were also significantly increased compared to WT cells, while this mitotic event was not significantly increased in the RMI1-KO clone. The lagging chromatin events during anaphase were not significantly higher in either TOP3A- or RMI1-KO iPSC clones. The chromatin bridges and the laggards were not significantly higher in the RMI1-KO clone although the p-value for both events was 0.07. These results could be interpreted as a milder effect of RMI1 deficiency in cellular phenotypes, which could be due to the issues raised in the previous sections. We would also expect a significant increase in the laggard events in the TOP3A-KO clone based on the previous report of Martin et al. which used the patient-derived fibroblast cells for their analysis [74]. However, this was not the case in our analyses of iPSCs, which might point to a cell type-specific effect.

Next, we quantified the rates of occurrence of the ultrafine DNA bridges (UFB). The UFBs are DAPI-negative DNA bridges that occur during anaphase and they were first discovered in normal cells by fluorescence staining for the target BLM or PICH helicases [105, 106, 186]. Most of the UFBs originate from centromeric regions, which include double-stranded DNA catenates that can be resolved by topoisomerase II and PICH protein [243]. UFBs can further originate from under-replicated DNA such as common fragile sites, and their resolution depends on the FA proteins, namely, FANCD2 and FANCI [106]. BLM might also play a role in UFB dissolution or management as the BLM protein decorates the UFB threads as detected by immunofluorescence techniques [145]. Hence, cells cope with UFBs, mainly via topoisomerases, PICH, FA proteins, and/or the BTRR complex [106]. We evaluated the UFB events in the BTRR complex-deficient iPSC lines via immunofluorescence visualization of the PICH protein. The rates of UFBs in three KO iPSC clones were significantly increased when compared to the rates observed in the WT control iPSC line, and the three KO iPSC lines had similar rates of UFB occurrence. This could imply that the BTRR complex deficiency results

in unresolved UFB events independent of which member of the BTRR complex is impaired. Supporting this, a previous study on patients with loss of *RMI2* also reported increased UFB events in RMI2 deficiency [73]. Here, the isogenic KO-iPSCs provide an excellent cell model for studying the UFBs. Still, there are many unanswered questions regarding the UFB resolution and molecular functions of each UFB-associated protein including BLM. Moreover, the exact genomic sequences of UFBs are currently unknown and remain to be discovered. Taken together, the KO iPSC lines generated during this doctoral thesis could be used for further discoveries regarding the ultrafine bridge formation driven by the BTRR complex deficiency.

4.3.4 Transcription profiles of the KO iPSC samples

We performed single-cell transcriptome sequencing on the generated BTRR complex-deficient iPS cell lines to define their gene expression profiles and to compare the results with the scRNAseq data of the BS fibroblasts. Three KO iPSC samples, namely, BLM-, TOP3A-, and RMI1-KO iPSCs, and two wild-type iPSC controls including the parental cell line isWT1 that was used for the generation of the BTRR complex-deficient iPSCs were analyzed. Additionally, we performed scRNAseq analysis with BLM-KO, TOP3A-KO, and WT control samples after treatment of cells with aphidicolin to induce replication stress. Initial data analysis revealed that the second non-isogenic WT cell line isWT7, which was added as an additional WT control, showed severe differences in gene expression profile compared to the isogenic control isWT1 and all KO-iPSC samples, and was therefore excluded from further analyses.

We compared the gene expression levels on KO iPSC lines to the WT parental control single cells and observed mainly downregulated transcripts in the KO iPS cells. Of note, none of the downregulated genes in the KO iPSC clones were off-targets of the guideRNAs, which were analyzed and given by the CRISPOR tool during the design of the guideRNAs. Furthermore, a heterogenous gene expression profile was obtained among the WT and the KO iPSC lines based on the UMAP analysis. Here, the differentiation potential and almost unlimited capacity of cell proliferation of iPSCs might

have resulted in the heterogeneous transcriptome profile. In this sense, subpopulations within the iPSC culture could be expected to lead to heterogeneous gene expression profiles as observed in the UMAP.

Next, the differentially expressed genes of the BLM-KO iPSCs were evaluated. We observed that the second most downregulated transcript was the BLM transcript in the generated BLM-KO clone. This served as an additional confirmation of the KO condition for this clone on a single-cell transcriptome level apart from previous WB confirmation on the translational level. There were 24 additional transcripts deregulated in this clone, most of which were downregulated with fold changes varying around 1.0. These candidate genes could be associated with the KO condition, as we knew from the scRNAseq data of the BS fibroblasts that the gene expression differences of 1.0-fold change could be highly associated with BS pathogenesis. The *HIST1H1A* gene was detected as the most downregulated (logFC= -3) in BLM-KO iPSC cells. This gene encodes the histone H1 protein which binds to the linker DNA between the nucleosomes. *HIST1H1A* transcript was also detected to be downregulated in RMI1-KO iPSCs. Downregulation of histone proteins including this linker histone was observed after DNA damage by IR [244]. Furthermore, there are seven linker histone genes in human cells with varying prevalence among tissues and specific roles [245]. We returned to our scRNAseq data from the BS fibroblasts for the histone expressions and observed significant upregulation of various histone proteins such as H3F3B, H2AFZ, and HIST1H4C. Currently, there is no direct link between expression of histones and BS. The significant downregulation of *HIST1H1A* in BLM- and RMI1-KO iPSCs could then be a stem cell-specific effect of the BTRR complex deficiency having an impact on chromosome organization in this specific cell model.

Next, the differentially expressed genes in the TOP3A-KO iPSC clone were evaluated. *CHCHD2* was the most downregulated transcript in this clone. Interestingly, this gene was also the most downregulated gene in the RMI1-KO iPSC clone when compared to the WT control. *CHCHD2* encodes a transcription factor binding to responsive elements in the promotor of cytochrome c oxidase subunit 4I2, which is part of the cytochrome c oxidase (COX) complex [246]. CHCHD2 is responsible for mitochondrial

metabolism as well as mitochondrial apoptosis inhibition. Therefore, its prevalence in the mitochondria decreases by apoptotic stimuli and this starts the apoptotic cascade of mitochondria [189]. The highly significant downregulation of *CHCHD2* in TOP3A- and RMI1-KO clones could point towards an apoptotic signal of the cells which could be then linked to BTRR complex deficiency and DNA damage. Interestingly, this gene was not differentially expressed in BS patient-derived fibroblast samples when compared to wild-type cells.

In the list of the downregulated genes in the three KO-iPSC clones, we observed numerous undefined noncoding RNA transcripts. Some of these transcripts were detected in two or more KO-iPSCs such as *AC019155.2*, which was downregulated in all of the three KO clones. Long noncoding RNAs (lncRNA) are longer than 200 nucleotides and are not translated into functional proteins [247]. It is now known that these lncRNAs have specific regulatory roles in various processes including transcription, mRNA degradation, or translation; and more complex functions such as dosage compensation or genomic imprinting [248, 249]. We analyzed the genomic regions of each lncRNA and searched for coding genes in those regions. For instance, the transcript *AC019155.2*, which was downregulated in all three of the KO-iPSC samples was located in a region where the *POT1-AS1* gene was located. Therefore, we could expect an impairment in the regulation of the *POT1* gene, which encodes a nuclear protein involved in telomere maintenance by protecting the end of the chromosomes from illegitimate recombination and abnormal chromosome segregation. The aberration of telomere maintenance could then create a link to the BTRR complex through this lncRNA which was downregulated on each KO-iPSC sample.

Although the introduced LoF mutations are likely to impair DNA repair and replication mechanisms since the KO iPSCs showed cellular phenotypes related to BS such as increased SCE rates, the overall single-cell transcriptome changes were mild. These changes were also very subtle when compared to the scRNAseq analyses of BS fibroblasts. This issue could be explained by various factors: (i) A short time, which was around 10 passages, had passed between the genome editing experiments and the scRNAseq experiment. Therefore, the gross transcriptional changes still needed

more time for detection in the case of the isogenic cell lines. (ii) The nature of iPSCs did not allow detectable transcriptome changes in response to the KO conditions as cells might have alternative DNA repair mechanisms to compensate for the KO conditions. Here, the power of single-cell RNAseq should be emphasized again since the mild changes might not even be detected in the bulk RNA sequencing due to small expression changes between the KO iPSCs and the parental WT. These two potential reasons for the mild gene expression changes could have been assessed by either longer cultivating times of iPSCs or by differentiation strategies for the KO iPSCs. The scRNAseq experiment, then, could be repeated and the results could be compared to BS fibroblast data for a more accurate comparison.

Overall, single-cell transcriptome profiles of the BTRR complex-deficient iPSC lines were similar to the parental wild-type control iPSCs. Nevertheless, we were able to obtain various down- or upregulated genes. Inducing replication stress in iPSCs also resulted in further transcriptional changes in BLM-KO, TOP3A-KO, and WT cells. In the future, it will be important to use these KO iPSC models for differentiating into somatic cells to confirm and further elucidate the gene expression profiles affected by the BTRR complex deficiency in different types of human cells.

4.4 Replication stress in BS and the BTRR complex deficiency

Proper replication of DNA that is managed by the replication fork machinery is a highly regulated, crucial process for the cells. This process is initiated with the timely firing of the licensed origins, which leads to fork progression with constant accuracy, speed, and sufficient resources [250]. Yet, the replication forks can encounter DNA damage resulting in the stoppage of the replication fork progression. These obstacles on the DNA can be DNA lesions caused by endogenous and exogenous factors as well as RNA transcription, or further proteins that are tightly bound to the DNA [251]. The replication fork stalling can trigger several pathways to cope with the delays, such as dormant ori-

gin firing mechanism to finish the replication on time, or DNA repair pathways to repair the damaged region. Cells defective of proper DNA repair processes can suffer from increased levels of replication stress due to higher amounts of DNA damage leading to overall genomic instability. The BLM helicase is one of the important proteins for the management and restart of the stalled replication forks along with other DNA repair proteins. Hence, Bloom syndrome cells suffer from replication stress, which contributes to the overall genomic instability and cancer predisposition in BS [151].

We aimed to induce further replication stress in BS patient-derived fibroblast cells to compare the transcriptional responses on a single cell level. We made use of a mild concentration of aphidicolin which was enough to stall the forks by inhibiting DNA polymerase without blocking the cell cycle intensively [164]. We treated BS fibroblasts and one WT control for 24 hours with aphidicolin and compared the gene expression differences based on the replication stress. Here, the transcription profiles of the BS cells were already highly different than the WT control cells and this could overshadow the effect of aphidicolin-based replication stress on gene expression, which was why we did not continue with the comparison of aphidicolin-treated BS samples to WT cells. To identify the effect of replication stress specifically in BS and in WT cells separately, we compared the aphidicolin-treated samples to their non-treated sample pair. This analysis revealed 152 deregulated GO terms for BS cells while WT cells had 123 deregulated pathways depending on the replication stress. Further, BS cells revealed terms such as ‘mitotic nuclear division’, ‘sister chromatid segregation’, ‘nuclear division’, and ‘nuclear chromosome segregation’ as significantly deregulated pathways while these terms were not significant for the aphidicolin-treated WT cells. Interestingly, these denominations included several condensin genes as core enrichment. Overall, these results imply that the BS cells showed a tendency to be more susceptible to replication stress. In addition, previous studies showed impairment of mitotic processes in BS leading to changes when replication stress was induced, such as increased UFB formation [252].

We further performed a scRNAseq experiment with the CRISPR/Cas9-generated knockout iPSC lines including replication stress induction. The three knockout cell lines were

deficient of the members of the BTRR complex, namely, BLM, TOP3A, and RMI1. To induce replication stress and observe the transcription changes of the KO cells, we treated BLM-KO, TOP3A-KO, and WT iPSCs with the same concentration of aphidicolin as the previous scRNAseq experiment. After scRNAseq data analyses, the transcription profiles of the KO cells and the WT cells were similar to each other, probably due to the isogenic background of the cells. Here, the aphidicolin treatment did not reveal gross transcriptional changes either. We would expect the BLM-KO cells to be the most susceptible to replication stress; however, this was not the case. We could speculate that the BLM-KO cells did not show many transcriptional changes to replication stress because they were already familiar with the replication stress due to BLM deficiency, and might have already been utilizing additional pathways to cope with replication fork stalling. Amor-Guéret speculates an opinion about the existence of alternative responses of BLM deficient cells to replication errors which make the cells escape apoptosis [153]. This could also explain the mild differences between the transcription profiles of BLM-KO and the parental WT. Here TOP3A-KO iPSC was the most affected cell line by the aphidicolin. For a stronger effect of replication stress on gene expression, the concentration of aphidicolin could be increased as a further experimental setting. Although we detected overlapping GO terms with the replication stress-induced WT cells as a sign of a global effect of aphidicolin, additional deregulated pathways in TOP3A-KO cells could point out to a replication stress sensitivity in TOP3A deficiency. Furthermore, we observed several denominations which included deregulated *SMC2/4* genes as an effect of replication stress. We noted these pathways especially since we previously defined altered transcription of condensin I/II complex genes in BS cells. The altered pathways in replication stress-induced TOP3A-KO and WT samples could reveal a possible link between replication fork stalling, the BTRR complex, and the condensins. It is important to note that the association of condensins to BLM deficiency was a novel finding and additional strategies should be developed to elaborate the exact link between replication stress, chromosome condensation, and the cellular functions of the BTRR complex.

4.5 Future perspectives and concluding remarks

The recent advances in NGS-based technologies provide novel tools for analyzing genotype-phenotype correlations. Expansion of the patient cohort of BS and the mutational spectrum of associated disease-causing variants in BS is facilitated by the help of modern diagnostic tools, and is of future interest. Further, identification of additional variants in the *RMI1* gene is crucial for a better association of *RMI1* with BS phenotype. We hypothesize a milder phenotype of *RMI1*-associated patients when compared to *BLM*- or *TOP3A*-associated patients, but proving this point would be ideally possible through characterization of further *RMI1*-associated patients and patient-derived cell models. Moreover, additional gene identification in BS phenotypes without mutations in the BTRR complex genes is of further interest for broadening the mutational signatures and cellular pathomechanisms in BS.

Our study on single-cell transcription profiles on BS fibroblasts provided an exceptional pilot study for expanding the application of single-cell transcriptomics in disease modeling. Additional disorders could be further molecularly characterized through single-cell transcriptomics with the possibility of discoveries in disease-associated mechanisms. We detected highly significant deregulation of microcephaly-associated genes as well as overexpression of condensin I/II complex genes in BS single cells which were novel findings. The link between condensin overexpression in BLM deficiency could be elucidated in future experiments. Analyses of DNA packaging in BLM-deficient cells could further shed light on this link. Moreover, localization of SMC2 or SMC4 proteins during anaphase could be studied via immunofluorescence assays as there could be a link between chromosome condensation, and UFB formation. In addition, cell lines derived from primary microcephaly patients having mutations in the condensin I/II complex genes could be analyzed for further overlapping cellular phenotypes observed in BS and condensinopathies.

Our isogenic iPSC clones provide an excellent cellular model for studying the specific roles of each member of the BTRR complex, namely, BLM, TOP3A, RMI1, and RMI2.

RMI2 is known only to interact with RMI1 within the BTRR complex; therefore, its functions might as well be studied through the KO iPSCs generated in this study. Further analyses of the generated cell lines in terms of impaired chromosome segregation and mitotic events are of particular interest for the expansion of the molecular roles of the BTRR complex. For instance, UFBs were discovered in 2007 and there are many open questions such as the mechanisms of UFB resolution as well as their exact genomic locations and DNA content. We showed that the isogenic KO iPSCs showed significantly increased amounts of UFB formation during anaphase. Therefore, these cell lines can be further used for the determination of the UFB content, and specific roles of the members of the BTRR complex for UFB management. Antibody-based isolation of UFBs followed by NGS techniques could shed light on the DNA content and sequence of the UFBs once a proper protocol is established similar to ChIPseq experiments.

The iPSC culture has the advantage of differentiating into almost any kind of human cell which makes it suitable for further disease modeling of affected tissues. Our generated iPSCs can be then used for differentiating into tissues for extensive characterization of BS and the BTRR complex deficiency. One of the most prominent clinical characteristics of BS is primary microcephaly. Therefore, generated KO iPSCs could be differentiated into neural progenitor cells, and cell proliferation and differentiation could be studied in these somatic cell types. The cell lines could be further differentiated into brain organoids to model the microcephaly profile in BS. *TOP3A*-associated patients presented a dilated cardiomyopathy profile and mitochondrial DNA depletion in the muscle cells, probably due to the mitochondrial functions of TOP3A independent of the BTRR complex. Regarding this characteristic of TOP3A, KO iPSC could be differentiated into cardiomyocytes to test the mitochondrial phenotypes as well. Since the BTRR complex is essential to every human cell, most kinds of iPSC differentiation could lead to further interesting evaluations of the BTRR complex deficiency.

Applications of iPSC also go beyond disease modeling and can influence translational approaches. As discussed above, enhancement of our understanding of the impaired molecular mechanisms in BS could have a translational effect on BS-associated clinical features including increased cancer predisposition. Elucidating the DNA repair mech-

anisms and the involvement of the BTRR complex members in these processes might help the clinical management of BS patients in terms of cancer occurrence, and therapy possibilities. Additionally, the generated knockout iPSC lines can be used for therapeutic approaches and specific drug screening, which is of particular importance since BS patients are extremely sensitive to conventional cancer treatment therapies using genotoxic reagents. From a wider perspective, the generated cell lines, the obtained results from high-throughput experiments, and the cellular characterizations of the cell models used in this thesis are of great interest for future studies for elucidating better the DNA replication and repair mechanisms as well as novel therapeutic approaches for genomic instability disorders.

5 Supplementary Material

Differentially expressed gene analyses of BS patient-derived fibroblasts in comparison to WT fibroblasts from the scRNAseq data are given in Table S1. Adjusted p-values were calculated according to Benjamini-Hochberg correction.

Table S1. DEG analyses of BS patient derived-fibroblasts in comparison to WT single cells sorted according to adjusted p-value.

The BTRR complex				
Genes	LogFC	AveExpr	P-value	Adj. p-value
TOP3A	0,359424	3,305138	0,009482	0,095486
RMI2	0,287609	3,139070	0,037528	0,261001
RMI1	0,225762	3,765547	0,113959	0,531244
BLM	-0,081187	2,481933	0,447474	0,601904

Fanconi anemia gene group				
Symbol	LogFC	AveExpr	P-value	Adj. p-value
RPA3	1,089976	5,607849	1,63396E-17	5,40372E-15
FANCD2	0,887237	4,176586	2,96052E-08	1,71044E-06
RPA2	0,607171	7,001507	6,77979E-08	3,61908E-06
FANCI	0,803874	4,876541	1,15228E-06	4,49841E-05
UBE2T	0,659916	6,572938	9,83958E-06	0,000296134
CENPX	0,400359	7,160502	4,66218E-05	0,001178737
FANCB	0,539818	3,184149	5,32383E-05	0,001320494
FANCA	0,563059	3,441442	0,00013027	0,002854623
FAAP20	-0,473073	4,975949	0,000159986	0,003364859
UBC	-0,231467	10,455946	0,000177345	0,00369483
USP1	0,554388	5,352282	0,000217177	0,004409758
FANCM	0,495472	3,232642	0,000248313	0,004906913
FANCE	0,520677	3,272309	0,000339439	0,006345898

FANCC	0,418690	2,775033	0,001093584	0,01696204
UBB	0,142781	9,046506	0,003265694	0,041674955
RPS27A	-0,184924	7,534762	0,007011275	0,076009192
ATRIP	0,360668	3,426300	0,010358315	0,102839692
ATR	0,330815	3,590473	0,021124533	0,17237523
EME1	0,367057	3,591579	0,0272014	0,208370253
FAAP100	-0,330063	3,512093	0,029208944	0,218854098
DCLRE1B	0,363047	3,907719	0,029420696	0,219696926
UBA52	-0,097905	9,206339	0,084376393	0,446428388
CENPS	0,184356	2,778507	0,088273173	0,459858
ERCC4	0,208029	3,624949	0,139970827	0,576998186
POLN	0,155718	2,348609	0,148114664	0,576998186
RPA1	0,152998	6,896450	0,218722516	0,576998186
ERCC1	-0,112884	6,409482	0,259420604	0,576998186
FAN1	0,074406	3,068053	0,569254096	0,677584761
FAAP24	0,065106	3,055614	0,618441186	0,713613735
FANCG	0,077170	4,971221	0,66535341	0,748935813
MUS81	-0,061806	4,130212	0,683618725	0,763290787
WDR48	-0,049628	4,339427	0,742068517	0,808099442
FANCL	0,039957	3,947869	0,794884281	0,847887954
DCLRE1A	0,031734	2,814781	0,803862343	0,854814922
FANCF	0,028944	3,064415	0,820578593	0,867684208
SLX4	-0,001843	2,556047	0,987943777	0,991839155
EME2	-0,001597	2,383431	0,988376056	0,992169279

Genes clinically associated with primary microcephaly

Symbol	LogFC	AveExpr	P-value	Adj. p-value
ASPM	1,121367	6,220315	4,0148E-09	2,85219E-07
CDK5RAP2	0,802403	4,897030	8,33711E-08	4,36137E-06
LMNB1	0,940226	5,156482	3,32474E-07	1,50883E-05
KIF14	0,841766	4,243427	5,45088E-07	2,3269E-05
CENPE	0,910331	5,479966	2,43516E-06	8,76901E-05
KNL1	0,836474	4,876570	2,90544E-06	0,000102195

CDK6	-0,509100	4,992703	0,00023115	0,004637831
NUP37	0,418713	5,308375	0,000927795	0,014813047
SASS6	0,371065	2,944559	0,002236537	0,030701165
LMNB2	0,447537	4,018949	0,003305329	0,042080053
CEP152	0,422173	3,491547	0,006114552	0,068374616
MFSD2A	0,327717	2,773458	0,008961609	0,091819047
WDR62	0,359078	3,100731	0,00917411	0,093337167
CIT	0,343501	4,095577	0,023954578	0,18924248
STIL	0,348774	3,847276	0,024463813	0,192262452
ZNF335	0,280375	2,932904	0,033315768	0,239700248
CENPJ	0,239384	3,441942	0,100904378	0,496978828
COPB2	0,094655	9,056733	0,139663512	0,576998186
PHC1	0,128446	2,159436	0,164373784	0,576998186
ANKLE2	-0,198164	5,241692	0,16670934	0,576998186
MCPH1	0,158737	3,344398	0,215799992	0,576998186
C7orf43	0,117326	2,543166	0,298030997	0,576998186
WDFY3	-0,100430	3,711868	0,457652986	0,606471637
CEP135	0,079490	3,131434	0,545939992	0,661324749

Significantly deregulated GO terms for biological processes obtained by GSEA of BLM-KO iPSCs is given in Table S2.

Table S2. Significant deregulated GO terms in the BLM-KO clone obtained by the GSEA.

GO term (biological processes)	Normalized enrichment score (NES)	P-value	Adj. p-value
Nucleobase metabolic process	2,411191	0,000005	0,009282
Telomeric loop disassembly	-2,037025	0,000054	0,047665

The top 10 most significantly deregulated GO terms for biological processes obtained by analyses of TOP3A-KO iPSCs' gene expression is given in Table S3.

Table S3. Significant deregulated GO terms in the TOP3A-KO clone obtained by the GSEA.

GO term (biological processes)	NES	P-value	Adj. p-value
SRP-dependent cotranslational protein targeting to membrane	2,281557	1,92684E-08	0,000039
Mitochondrial respiratory chain complex assembly	2,221436	2,23935E-08	0,000039
Cotranslational protein targeting to membrane	2,109003	2,67214E-07	0,000266
Oxidative phosphorylation	2,035552	3,05823E-07	0,000266
NADH dehydrogenase complex assembly	2,196185	2,28563E-06	0,001325
Mitochondrial respiratory chain complex I assembly	2,196185	2,28563E-06	0,001325
Mitochondrial electron transport, NADH to ubiquinone	2,194055	7,82037E-06	0,003884
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	1,803689	3,18439E-05	0,013840
Electron transport chain	1,728718	4,7107E-05	0,018199
Cytochrome complex assembly	2,248139	5,95373E-05	0,020701

Signifcantly deregulated pathways in RMI1-KO iPSC clone are given in Table S4. Only 4 pathways were significant as determined by GSEA using GO terms for biological processes.

Table S4. Significant deregulated GO terms in the RMI1-KO clone obtained by the GSEA.

GO term (biological processes)	NES	P-value	Adj. p-value
Response to copper ion	2,292339	1,20804E-05	0,031477
Cellular response to cadmium ion	2,284061	3,03009E-05	0,038781
Pigment metabolic process	2,149724	4,46048E-05	0,038896
Response to cadmium ion	2,135307	7,06564E-05	0,040000

Top 20 most significantly deregulated terms in the aphidicolin-treated BLM-KO iPSCs in comparison to WT cells determined by GSEA are listed according to adjusted p-value in Table S5.

Table S5. Top 20 most significantly deregulated terms in aphidicolin-treated BLM-KO iPSCs in comparison to WT cells. DGEA was performed using GO terms for biological processes.

GO term	NES	P-value	Adj. p-value
Pigment metabolic process	2,586056	0,000001	0,000657
Mitochondrial electron transport, cytochrome c to oxygen	2,926988	0,000002	0,000871
Aerobic electron transport chain	2,926988	0,000002	0,000871
Porphyrin-containing compound metabolic process	2,557331	0,000002	0,000889
Heme metabolic process	2,713321	0,000003	0,001015
Pigment biosynthetic process	2,219103	0,000004	0,001015
Positive regulation of ubiquitin protein ligase activity	-2,146635	0,000007	0,001795
Chromatin organization involved in negative regulation of transcription	-1,699575	0,000023	0,005330
Alpha-amino acid biosynthetic process	-1,863513	0,000024	0,005330
Tetrapyrrole metabolic process	1,984096	0,000038	0,007498
Mitochondrial translational elongation	1,841472	0,000055	0,010139
Nucleosome assembly	-1,649971	0,000057	0,010139
Negative regulation of DNA-templated transcription, elongation	2,320911	0,000102	0,016649
Chromatin organization involved in regulation of transcription	-1,638725	0,000102	0,016649
Ribonucleoside triphosphate biosynthetic process	2,050582	0,000106	0,016852
Nucleosome positioning	-1,989083	0,000111	0,017207
Regulation of release of cytochrome c from mitochondria	2,099332	0,000122	0,018292
Negative regulation of DNA recombination	-1,885280	0,000125	0,018292
Phospholipid dephosphorylation	-1,858050	0,000125	0,018292
Nucleoside biosynthetic process	2,101870	0,000166	0,022432

Top 20 most significantly deregulated GO terms obtained by GSEA in aphidicolin-treated TOP3A-KO iPSCs in comparison to WT cells are listed in Table S6 according to adjusted p-value.

Table S6. Top 20 most significantly deregulated terms in aphidicolin-treated TOP3A-KO iPSCs. DGEA was performed using GO terms for biological processes.

GO term	NES	P-value	Adj. p-value
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	2,641021	1E-10	4,09909E-08
Protein targeting to membrane	2,438996	1E-10	4,09909E-08
Cotranslational protein targeting to membrane	3,082846	1E-10	4,09909E-08
SRP-dependent cotranslational protein targeting to membrane	3,193664	1E-10	4,09909E-08
Protein targeting to ER	2,836974	1E-10	4,09909E-08
Protein localization to endoplasmic reticulum	2,484792	1E-10	4,09909E-08
Establishment of protein localization to endoplasmic reticulum	2,751709	1E-10	4,09909E-08
Nuclear division	-1,888424	1,28705E-09	4,14522E-07
Organelle fission	-1,845575	1,71404E-09	4,83037E-07
Translational initiation	2,101229	1,98746E-09	5,15302E-07
Oxidative phosphorylation	2,333530	2,05709E-09	5,15302E-07
Mitotic nuclear division	-1,903323	1,79258E-08	4,04137E-06
Nuclear chromosome segregation	-1,870325	9,38387E-08	1,83965E-05
Regulation of protein modification process	-1,511664	2,12036E-07	3,98362E-05
Chromosome segregation	-1,778562	2,91639E-07	5,26E-05
Electron transport chain	2,002672	8,68907E-07	0,00014673
Mitochondrial respiratory chain complex assembly	2,131101	8,79184E-07	0,00014673
Sister chromatid segregation	-1,841545	9,11167E-07	0,00014673
Regulation of cell cycle process	-1,559244	9,67839E-07	0,000150482
Viral transcription	1,879179	1,16008E-06	0,000174361

Top 20 most significantly deregulated GO terms obtained by GSEA in aphidicolin-

treated TOP3A-KO iPSCs in comparison to WT cells are listed in Table S7 according to adjusted p-value.

Table S7. Top 20 most deregulated GO terms in aphidicolin treated versus non-treated WT iPSCs. DGEA was performed using GO terms for biological processes.

GO term	NES	P-value	Adj. p-value
Regulated exocytosis	1,939057	7,98811E-10	2,65162E-06
Immune effector process	1,805905	1,5213E-09	2,65162E-06
Exocytosis	1,846616	2,71016E-09	3,14921E-06
Neutrophil mediated immunity	1,949456	5,92929E-09	4,26176E-06
Granulocyte activation	1,914767	6,5874E-09	4,26176E-06
Myeloid leukocyte activation	1,934149	7,33522E-09	4,26176E-06
Neutrophil activation involved in immune response	1,946610	1,91363E-08	8,33864E-06
Neutrophil degranulation	1,946610	1,91363E-08	8,33864E-06
Neutrophil activation	1,916945	2,27778E-08	8,82259E-06
Myeloid cell activation involved in immune response	1,885526	4,65631E-08	1,58904E-05
NAD metabolic process	2,570165	5,35244E-08	1,58904E-05
Leukocyte degranulation	1,878164	5,47001E-08	1,58904E-05
Myeloid leukocyte mediated immunity	1,842664	1,4131E-07	3,78928E-05
Secretion by cell	1,613796	2,2365E-07	5,56889E-05
Leukocyte mediated immunity	1,732943	2,70387E-07	6,28378E-05
Nucleobase-containing small molecule metabolic process	1,705983	4,08191E-07	8,57422E-05
Export from cell	1,581558	4,18135E-07	8,57422E-05
NADH metabolic process	2,446453	4,47839E-07	8,67315E-05
Secretion	1,568588	5,96193E-07	0,000109386
Cell activation involved in immune response	1,701666	1,04566E-06	0,000182258
Small molecule catabolic process	1,904682	1,34484E-06	0,000195805

Bibliography

1. Ferreira, C. R. The burden of rare diseases. *American Journal of Medical Genetics Part A* **179,** 885–892 (2019).
2. Bhattacharjee, S. & Nandi, S. Rare genetic diseases with defects in DNA repair: Opportunities and challenges in orphan drug development for targeted cancer therapy. *Cancers* **10,** 298 (2018).
3. Angelis, A., Tordrup, D. & Kanavos, P. Socio-economic burden of rare diseases: A systematic review of cost of illness evidence. *Health Policy* **119,** 964–979 (2015).
4. Melnikova, I. Rare diseases and orphan drugs. *Nature Reviews Drug Discovery* **11,** 267–268 (2012).
5. Wright, C. F., FitzPatrick, D. R. & Firth, H. V. Paediatric genomics: diagnosing rare disease in children. *Nature Reviews Genetics* **19,** 253–268 (2018).
6. Raffan, E. & Semple, R. K. Next generation sequencing—implications for clinical practice. *British Medical Bulletin* **99,** 53–71 (2011).
7. Hansen, M. C., Haferlach, T. & Nyvold, C. G. A decade with whole exome sequencing in haematology. *British Journal of Haematology* **188,** 367–382 (2020).
8. Biesecker, L. G., Shianna, K. V. & Mullikin, J. C. Exome sequencing: The expert view. *Genome Biology* **12,** 128 (2011).
9. Ho, S. S., Urban, A. E. & Mills, R. E. Structural variation in the sequencing era. *Nature Reviews Genetics* **21,** 171–189 (2020).
10. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17,** 333–351 (2016).
11. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376,** 44–53 (2022).
12. Saliba, A.-E., Westermann, A. J., Gorski, S. A. & Vogel, J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Research* **42,** 8845–8860 (2014).
13. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6,** 377–382 (2009).
14. Regev, A. *et al.* The Human Cell Atlas. *eLife* **6,** e27041 (2017).
15. Regev, A. *et al.* The Human Cell Atlas White Paper. *ArXiv preprint arXiv:1810.05192* (2018).
16. Aldridge, S. & Teichmann, S. A. Single cell transcriptomics comes of age. *Nature Communications* **11,** 4307 (2020).
17. Dal Molin, A. & Di Camillo, B. How to design a single-cell RNA-sequencing experiment: Pitfalls, challenges and perspectives. *Briefings in Bioinformatics* **20,** 1384–1394 (2018).

18. Baran-Gale, J., Chandra, T. & Kirschner, K. Experimental design for single-cell RNA sequencing. *Briefings in Functional Genomics* **17,** 233–239 (2017).
19. Goldstein, L. D. *et al.* Massively parallel nanowell-based single-cell gene expression profiling. *BMC Genomics* **18,** 519 (2017).
20. Shomroni, O. *et al.* A novel single-cell RNA-sequencing approach and its applicability connecting genotype to phenotype in ageing disease. *Scientific Reports* **12,** 4091 (2022).
21. Hoeijmakers, J. H. J. Genome maintenance mechanisms for preventing cancer. *Nature* **411,** 366–374 (2001).
22. Spry, M., Scott, T., Pierce, H. & D'Orazio, J. A. DNA repair pathways and hereditary cancer susceptibility syndromes. *Frontiers in Bioscience* **12,** 4191–4207 (2007).
23. Vijg, J. & Suh, Y. Genome instability and aging. *Annual Review of Physiology* **75,** 645–668 (2013).
24. Bernstein, C., Bernstein, H., Payne, C. M. & Garewal, H. DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: Fail-safe protection against carcinogenesis. *Mutation Research/Reviews in Mutation Research* **511,** 145–178 (2002).
25. Hoeijmakers, J. H. Genome maintenance mechanisms for preventing cancer. *Nature* **411,** 366–374 (2001).
26. Tiwari, V. & Wilson, D. M. DNA damage and associated DNA repair defects in disease and premature aging. *The American Journal of Human Genetics* **105,** 237–257 (2019).
27. Ruggieri, V. *et al.* Loss of MUTYH function in human cells leads to accumulation of oxidative damage and genetic instability. *Oncogene* **32,** 4500–4508 (2013).
28. Iyer, R. R., Pluciennik, A., Burdett, V. & Modrich, P. L. DNA Mismatch Repair: Functions and Mechanisms. *Chemical Reviews* **106,** 302–323 (2006).
29. Sehgal, R. *et al.* Lynch syndrome: An updated review. *Genes* **5,** 497–507 (2014).
30. Park, D. J., Stoehlmacher, J. & Lenz, H.-J. Tailoring chemotherapy in advanced colorectal cancer. *Current Opinion in Pharmacology* **3,** 378–385 (2003).
31. Nouspikel, T. DNA repair in mammalian cells. *Cellular and Molecular Life Sciences* **66,** 994–1009 (2009).
32. Lehmann, A. R. DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *Biochimie* **85,** 1101–1111 (2003).
33. Woodbine, L., Gennery, A. R. & Jeggo, P. A. The clinical impact of deficiency in DNA non-homologous end-joining. *DNA Repair* **16,** 84–96 (2014).
34. De Bruin, C. *et al.* An XRCC4 Splice Mutation Associated With Severe Short Stature, Gonadal Failure, and Early-Onset Metabolic Syndrome. *The Journal of Clinical Endocrinology & Metabolism* **100,** E789–E798 (2015).
35. Kumrah, R. *et al.* Genetics of severe combined immunodeficiency. *Genes & Diseases* **7,** 52–61 (2020).
36. Gennery, A. R., Cant, A. J. & Jeggo, P. A. Immunodeficiency associated with DNA repair defects. *Clinical and Experimental Immunology* **121,** 1–7 (2000).

37. Woods, C. G. Human microcephaly. *Current Opinion in Neurobiology* **14,** 112–117 (2004).
38. Siskos, N., Stylianopoulou, E., Skavdis, G. & Grigoriou, M. E. Molecular genetics of microcephaly primary hereditary, an overview. *Brain Sciences* **11,** 581 (2021).
39. Jackson, A. P. *et al.* Identification of microcephalin, a protein implicated in determining the size of the human brain. *American Journal of Human Genetics* **71,** 136–142 (2002).
40. Trimborn, M. *et al.* Mutations in microcephalin cause aberrant regulation of chromosome condensation. *American Journal of Human Genetics* **75,** 261–266 (2004).
41. Xu, X., Lee, J. & Stern, D. F. Microcephalin is a DNA damage response protein involved in regulation of CHK1 and BRCA1. *Journal of Biological Chemistry* **279,** 34091–34094 (2004).
42. Nicholas, A. K. *et al.* WDR62 is associated with the spindle pole and is mutated in human microcephaly. *Nature Genetics* **42,** 1010–1014 (2010).
43. Wollnik, B. A common mechanism for microcephaly. *Nature Genetics* **42,** 923–924 (2010).
44. Nicholas, A. K. *et al.* The molecular landscape of ASPM mutations in primary microcephaly. *Journal of Medical Genetics* **46,** 249–253 (2009).
45. Bond, J. *et al.* Protein-truncating mutations in ASPM cause variable reduction in brain size. *American Journal of Human Genetics* **73,** 1170–1177 (2003).
46. Yigit, G., Rosin, N. & Wollnik, B. Molekulare Grundlagen der autosomal-rezessiven primären Mikrozephalie. *Medizinische Genetik* **27,** 345–350 (2015).
47. Jean, F., Stuart, A. & Tarailo-Graovac, M. Dissecting the genetic and etiological causes of primary microcephaly. *Frontiers in Neurology* **11** (2020).
48. Dehay, C. & Kennedy, H. Cell-cycle control and cortical development. *Nature Reviews Neuroscience* **8,** 438–450 (2007).
49. Bloom, D. Congenital telangiectatic erythema resembling lupus erythematosus in dwarfs; probably a syndrome entity. *A.M.A. American Journal of Diseases of Children* **88,** 754–758 (1954).
50. German, J. Bloom's syndrome. I. Genetical and clinical observations in the first twenty-seven patients. *American Journal of Human Genetics* **21,** 196–227 (1969).
51. German, J. Bloom's syndrome. *Dermatologic Clinics* **13,** 7–18 (1995).
52. Ellis, N. A. *et al.* Somatic intragenic recombination within the mutated locus BLM can correct the high sister-chromatid exchange phenotype of bloom syndrome cells. *American Journal of Human Genetics* **57,** 1019–1027 (1995).
53. Karow, J. K., Chakraverty, R. K. & Hickson, I. D. The Bloom's syndrome gene product is a 3'-5' DNA helicase. *Journal of Biological Chemistry* **272,** 30611–30614 (1997).
54. German, J., Sanz, M. M., Ciocci, S., Ye, T. Z. & Ellis, N. A. Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry. *Human Mutation* **28,** 743–753 (2007).
55. Ababou, M. Bloom syndrome and the underlying causes of genetic instability. *Molecular Genetics and Metabolism* **133,** 35–48 (2021).

56. Arora, H. *et al.* Bloom syndrome. *International Journal of Dermatology* **53,** 798–802 (2014).
57. Cunniff, C., Bassetti, J. A. & Ellis, N. A. Bloom's syndrome: Clinical spectrum, molecular pathogenesis, and cancer predisposition. *Molecular Syndromology* **8,** 4–23 (2017).
58. Woods, C. G. & Parker, A. Investigating microcephaly. *Archives of Disease in Childhood,* archdischild–2012 (2013).
59. Gretzula, J. C., Hevia, O. & Weber, P. J. Bloom's syndrome. *Journal of the American Academy of Dermatology* **17,** 479–488 (1987).
60. Keller, C., Keller, K. R., Shew, S. B. & Plon, S. E. Growth deficiency and malnutrition in Bloom syndrome. *The Journal of Pediatrics* **134,** 472–479 (1999).
61. Cottrell, E. *et al.* The value of whole exome sequencing for genetic diagnosis in a patient with Bloom syndrome. *Journal of Endocrinological Investigation* **44,** 1331–1334 (2021).
62. Flanagan, M. & Cunniff, C. M. *Bloom Syndrome* (University of Washington, Seattle, Seattle (WA), 1993).
63. Renes, J. S., Willemsen, R. H., Wagner, A., Finken, M. J. J. & Hokken-Koelega, A. C. S. Bloom syndrome in short children born small for gestational age: A challenging diagnosis. *Journal of Clinical Endocrinology and Metabolism* **98,** 3932–3938 (2013).
64. Sanz, M. M., German, J. & Cunniff, C. Bloom's syndrome. *GeneReviews®[Internet]. Seattle: University of Washington* (1993).
65. Masmoudi, A. *et al.* Clinical and laboratory findings in 8 patients with Bloom's syndrome. *Journal of Dermatological Case Reports* **6,** 29–33 (2012).
66. Cunniff, C. *et al.* Health supervision for people with Bloom syndrome. *American Journal of Medical Genetics Part A* **176,** 1872–1881 (2018).
67. Van Kerckhove, C. W. *et al.* Bloom's syndrome: Clinical features and immunologic abnormalities of four patients. *American Journal of Diseases of Children* **142,** 1089–1093 (1988).
68. Sugrañes, T. A. *et al.* Age of first cancer diagnosis and survival in Bloom syndrome. *Genetics in Medicine* (2022).
69. Ellis, N. A. *et al.* Linkage disequilibrium between the FES, D15S127, and BLM loci in Ashkenazi Jews with Bloom syndrome. *American Journal of Human Genetics* **55,** 453–460 (1994).
70. Ellis, N. A. *et al.* The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* **83,** 655–666 (1995).
71. Bahr, A., De Graeve, F., Kedinger, C. & Chatton, B. Point mutations causing Bloom's syndrome abolish ATPase and DNA helicase activities of the BLM protein. *Oncogene* **17,** 2565–2571 (1998).
72. Bythell-Douglas, R. & Deans, A. J. A structural guide to the Bloom syndrome complex. *Structure* **29,** 99–113 (2021).
73. Hudson, D. F. *et al.* Loss of RMI2 increases genome instability and causes a Bloom-like syndrome. *PLOS Genetics* **12,** 1–24 (2016).

74. Martin, C. A. *et al.* Mutations in TOP3A cause a Bloom syndrome-like disorder. *American Journal of Human Genetics* **103,** 221–231 (2018).

75. Wu, L. *et al.* The Bloom's syndrome gene product interacts with Topoisomerase III. *Journal of Biological Chemistry* **275,** 9636–9644 (2000).

76. Chang, M. *et al.* RMI1/NCE4, a suppressor of genome instability, encodes a member of the RecQ helicase/Topo III complex. *The EMBO Journal* **24,** 2024–2033 (2005).

77. Yin, J. *et al.* BLAP75, an essential component of Bloom's syndrome protein complexes that maintain genome integrity. *The EMBO Journal* **24,** 1465–1476 (2005).

78. Lu, H. & Davis, A. J. Human RecQ helicases in DNA double-strand break repair. *Frontiers in Cell and Developmental Biology* **9,** 640755 (2021).

79. Van Brabant, A. J., Stan, R. & Ellis, N. A. DNA helicases, genomic instability, and human genetic disease. *Annual Review of Genomics and Human Genetics* **1,** 409–459 (2000).

80. Croteau, D. L., Popuri, V., Opresko, P. L. & Bohr, V. A. Human RecQ helicases in DNA repair, recombination, and replication. *Annual Review of Biochemistry* **83,** 519–552 (2014).

81. Guo, R.-b., Rigolet, P., Zargarian, L., Fermandjian, S. & Xi, X. G. Structural and functional characterizations reveal the importance of a zinc binding domain in Bloom's syndrome helicase. *Nucleic Acids Research* **33,** 3109–3124 (2005).

82. Rong, S.-B., Väliaho, J. & Vihinen, M. Structural basis of Bloom syndrome (BS) causing mutations in the BLM helicase domain. *Molecular Medicine* **6,** 155–164 (2000).

83. Wu, L. *et al.* The HRDC domain of BLM is required for the dissolution of double Holliday junctions. *The EMBO Journal* **24,** 2679–2687 (2005).

84. Teng, F.-Y. *et al.* The HRDC domain oppositely modulates the unwinding activity of E. coli RecQ helicase on duplex DNA and G-quadruplex. *Journal of Biological Chemistry* **295,** 17646–17658 (2020).

85. Abu-Libdeh, B. *et al.* RECON syndrome is a genome instability disorder caused by mutations in the DNA helicase RECQL1. *The Journal of Clinical Investigation* **132** (2022).

86. Larsen, N. B. & Hickson, I. D. RecQ Helicases: Conserved guardians of genomic integrity. *Advances in Experimental Medicine and Biology* **767,** 161–184 (2013).

87. Gangloff, S., McDonald, J. P., Bendixen, C., Arthur, L. & Rothstein, R. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Molecular and Cellular Biology* **14,** 8391–8398 (1994).

88. Watt, P. M., Louis, E. J., Borts, R. H. & Hickson, I. D. Sgs1: a eukaryotic homolog of E. coli RecQ that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation. *Cell* **81,** 253–260 (1995).

89. Ralf, C., Hickson, I. D. & Wu, L. The Bloom's syndrome helicase can promote the regression of a model replication fork. *Journal of Biological Chemistry* **281,** 22839–22846 (2006).

90. Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *Journal of Biological Chemistry* **273,** 27587–27592 (1998).

91. Huber, M. D., Lee, D. C. & Maizels, N. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Research* **30,** 3954–3961 (2002).

92. Dutertre, S. *et al.* Cell cycle regulation of the endogenous wild type Bloom's syndrome DNA helicase. *Oncogene* **19,** 2731–2738 (2000).

93. Rao, V. A. *et al.* Phosphorylation of BLM, dissociation from topoisomerase IIIalpha, and colocalization with gamma-H2AX after topoisomerase I-induced replication damage. *Molecular and Cellular Biology* **25,** 8925–8937 (2005).

94. Leng, M. *et al.* MPS1-dependent mitotic BLM phosphorylation is important for chromosome stability. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 11485–11490 (2006).

95. Mankouri, H. W. & Hickson, I. D. The RecQ helicase–topoisomerase III–Rmi1 complex: a DNA structure-specific 'dissolvasome'? *Trends in Biochemical Sciences* **32,** 538–546 (2007).

96. Rodriguez, A. & D'Andrea, A. Fanconi anemia pathway. *Current Biology* **27,** R986–R988 (2017).

97. Ling, C. *et al.* Bloom syndrome complex promotes FANCM recruitment to stalled replication forks and facilitates both repair and traverse of DNA interstrand crosslinks. *Cell Discovery* **2,** 16047 (2016).

98. Yan, Z. *et al.* A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Molecular Cell* **37,** 865–878 (2010).

99. Manthei, K. A. & Keck, J. L. The BLM dissolvasome in DNA replication and repair. *Cellular and Molecular Life Sciences* **70,** 4067–4084 (2013).

100. Beamish, H. *et al.* Functional link between BLM defective in Bloom's syndrome and the ataxia-telangiectasia-mutated protein, ATM. *Journal of Biological Chemistry* **277,** 30515–30523 (2002).

101. Davies, S. L., North, P. S., Dart, A., Lakin, N. D. & Hickson, I. D. Phosphorylation of the Bloom's syndrome helicase and its role in recovery from S-phase arrest. *Molecular and Cellular Biology* **24,** 1279–1291 (2004).

102. Wu, L. Role of the BLM helicase in replication fork management. *DNA Repair* **6,** 936–944 (2007).

103. Naim, V. & Rosselli, F. The FANC pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities. *Nature Cell Biology* **11,** 761–768 (2009).

104. Ying, S. & Hickson, I. D. Fanconi anaemia proteins are associated with sister chromatid bridging in mitosis. *International Journal of Hematology* **93,** 440–445 (2011).

105. Chan, K. L., North, P. S. & Hickson, I. D. BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. *The EMBO Journal* **26,** 3397–3409 (2007).

106. Chan, K. L. & Hickson, I. D. On the origins of ultra-fine anaphase bridges. *Cell Cycle* **8,** 3065–3066 (2009).

107. Schoenaker, M. H. D. *et al.* Immunodeficiency in Bloom's syndrome. *Journal of Clinical Immunology* **38,** 35–44 (2018).

108. Hsieh, C. L., Arlett, C. F. & Lieber, M. R. V(D)J recombination in ataxia telangiectasia, Bloom's syndrome, and a DNA ligase I-associated immunodeficiency disorder. *Journal of Biological Chemistry* **268,** 20105–20109 (1993).

109. Babbe, H. *et al.* Genomic instability resulting from Blm deficiency compromises development, maintenance, and function of the B cell lineage. *The Journal of Immunology* **182,** 347–360 (2009).

110. Grabarz, A. *et al.* A role for BLM in double-strand break repair pathway choice: prevention of CtIP/Mre11-mediated alternative nonhomologous end-joining. *Cell Reports* **5,** 21–28 (2013).

111. Yamanishi, A. *et al.* Enhancement of microhomology-mediated genomic rearrangements by transient loss of mouse Bloom syndrome helicase. *Genome Research* **23,** 1462–1473 (2013).

112. Babbe, H., Chester, N., Leder, P. & Reizis, B. The Bloom's syndrome helicase is critical for development and function of the alphabeta T-cell lineage. *Molecular and Cellular Biology* **27,** 1947–1959 (2007).

113. Forterre, P., Gribaldo, S., Gadelle, D. & Serre, M.-C. Origin and evolution of DNA topoisomerases. *Biochimie* **89,** 427–446 (2007).

114. Champoux, J. J. DNA topoisomerases: Structure, function, and mechanism. *Annual Review of Biochemistry* **70,** 369–413 (2001).

115. Schoeffler, A. J. & Berger, J. M. DNA topoisomerases: Harnessing and constraining energy to govern chromosome topology. *Quarterly Reviews of Biophysics* **41,** 41–101 (2008).

116. Viard, T. & de la Tour, C. B. Type IA topoisomerases: A simple puzzle? *Biochimie* **89,** 456–467 (2007).

117. Bizard, A. H. & Hickson, I. D. The many lives of type IA topoisomerases. *Journal of Biological Chemistry* **295,** 7138–7153 (2020).

118. Yang, J., Bachrati, C. Z., Ou, J., Hickson, I. D. & Brown, G. W. Human topoisomerase IIIα is a single-stranded DNA decatenase that is stimulated by BLM and RMI1. *Journal of Biological Chemistry* **285,** 21426–21436 (2010).

119. Hodson, C. *et al.* Mechanism of Bloom syndrome complex assembly required for double Holliday junction dissolution and genome stability. *Proceedings of the National Academy of Sciences of the United States of America* **119,** e2109093119 (2022).

120. Li, W. & Wang, J. C. Mammalian DNA topoisomerase IIIalpha is essential in early embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 1010–1013 (1998).

121. Wang, Y., Lyu, Y. L. & Wang, J. C. Dual localization of human DNA topoisomerase IIIα to mitochondria and nucleus. *Proceedings of the National Academy of Sciences* **99,** 12114–12119 (2002).

122. Nicholls, T. J. *et al.* Topoisomerase 3α is required for decatenation and segregation of human mtDNA. *Molecular Cell* **69,** 9–23 (2018).

123. Jiang, W. *et al.* Predominant cellular mitochondrial dysfunction in the TOP3A gene-caused Bloom syndrome-like disorder. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1867,** 166106 (2021).

124. Singh, T. R. *et al.* BLAP18/RMI2, a novel OB-fold-containing protein, is an essential component of the Bloom helicase-double Holliday junction dissolvasome. *Genes & Development* **22,** 2856–2868 (2008).

125. Xu, D. *et al.* RMI, a new OB-fold complex essential for Bloom syndrome protein to maintain genome stability. *Genes & Development* **22,** 2843–2855 (2008).

126. Wang, F. *et al.* Crystal structures of RMI1 and RMI2, two OB-fold regulatory subunits of the BLM complex. *Structure* **18,** 1159–1170 (2010).

127. Séguéla-Arnaud, M. *et al.* RMI1 and TOP3α limit meiotic CO formation through their C-terminal domains. *Nucleic Acids Research* **45,** 1860–1871 (2017).

128. Liu, Y. & West, S. C. More complexity to the Bloom's syndrome complex. *Genes & Development* **22,** 2737–2742 (2008).

129. Hoadley, K. A. *et al.* Structure and cellular roles of the RMI core complex from the bloom syndrome dissolvasome. *Structure* **18,** 1149–1158 (2010).

130. Wu, L. *et al.* BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 4068–4073 (2006).

131. Shorrocks, A.-M. K. *et al.* The Bloom syndrome complex senses RPA-coated single-stranded DNA to restart stalled replication forks. *Nature Communications* **12,** 585 (2021).

132. Chen, H., You, M. J., Jiang, Y., Wang, W. & Li, L. RMI1 attenuates tumor development and is essential for early embryonic survival. *Molecular Carcinogenesis* **50,** 80–88 (2011).

133. Guiraldelli, M. F., Eyster, C. & Pezza, R. J. Genome instability and embryonic developmental defects in RMI1 deficient mice. *DNA Repair* **12,** 835–843 (2013).

134. San Filippo, J., Sung, P. & Klein, H. Mechanism of Eukaryotic Homologous Recombination. *Annual Review of Biochemistry* **77,** 229–257 (2008).

135. Rajvee, S. P., Jose, M. M., M., W. H. D., Wai, C. Y. & C., W. S. Resolution of single and double Holliday junction recombination intermediates by GEN1. *Proceedings of the National Academy of Sciences* **114,** 443–450 (2017).

136. Matos, J. & West, S. C. Holliday junction resolution: Regulation in space and time. *DNA Repair* **19,** 176–181 (2014).

137. Karow, J. K., Constantinou, A., Li, J.-L., West, S. C. & Hickson, I. D. The Bloom's syndrome gene product promotes branch migration of Holliday junctions. *Proceedings of the National Academy of Sciences* **97,** 6504–6508 (2000).

138. Räschle, M. *et al.* Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* **134,** 969–980 (2008).

139. Deans, A. J. & West, S. C. FANCM connects the genome instability disorders Bloom's syndrome and Fanconi anemia. *Molecular Cell* **36,** 943–953 (2009).

140. Chaudhury, I., Sareen, A., Raghunandan, M. & Sobeck, A. FANCD2 regulates BLM complex functions independently of FANCI to promote replication fork recovery. *Nucleic Acids Research* **41,** 6444–6459 (2013).

141. Chaganti, R. S., Schonberg, S. & German, J. A manyfold increase in sister chromatid exchanges in Bloom's syndrome lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America* **71,** 4508–4512 (1974).

142. German, J., Schonberg, S., Louie, E. & Chaganti, R. S. Bloom's syndrome. IV. Sister-chromatid exchanges in lymphocytes. *American Journal of Human Genetics* **29,** 248–255 (1977).

143. Van Wietmarschen, N. *et al.* BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nature Communications* **9,** 1–12 (2018).

144. Liu, Y., Nielsen, C. F., Yao, Q. & Hickson, I. D. The origins and processing of ultra fine anaphase DNA bridges. *Current Opinion in Genetics & Development* **26,** 1–5 (2014).

145. Chan, K. L. & Hickson, I. D. New insights into the formation and resolution of ultra-fine anaphase bridges. *Seminars in Cell and Developmental Biology* **22,** 906–912 (2011).

146. Suhasini, A. N. & Brosh, R. M. Fanconi anemia and Bloom's syndrome crosstalk through FANCJ-BLM helicase interaction. *Trends in Genetics* **28,** 7–13 (2012).

147. Yankiwski, V., Marciniak, R. A., Guarente, L. & Neff, N. F. Nuclear structure in normal and Bloom syndrome cells. *Proceedings of the National Academy of Sciences* **97,** 5214–5219 (2000).

148. Rosin, M. P. & German, J. Evidence for chromosome instability in vivo in bloom syndrome: Increased numbers of micronuclei in exfoliated cells. *Human Genetics* **71,** 187–191 (1985).

149. Giannelli, F., Benson, P. F., Pawsey, S. A. & Polani, P. E. Ultraviolet light sensitivity and delayed DNA-chain maturation in Bloom's syndrome fibroblasts. *Nature* **265,** 466–469 (1977).

150. Lönn, U., Lönn, S., Nylen, U., Winblad, G. & German, J. An abnormal profile of DNA replication intermediates in Bloom's syndrome. *Cancer Research* **50,** 3141–3145 (1990).

151. Rassool, F. V., North, P. S., Mufti, G. J. & Hickson, I. D. Constitutive DNA damage is linked to DNA replication abnormalities in Bloom's syndrome cells. *Oncogene* **22,** 8749–8757 (2003).

152. Davalos, A. R. & Campisi, J. Bloom syndrome cells undergo p53-dependent apoptosis and delayed assembly of BRCA1 and NBS1 repair complexes at stalled replication forks. *The Journal of cell biology* **162,** 1197–1209 (2003).

153. Amor-Guéret, M. Bloom syndrome, genomic instability and cancer: The SOS-like hypothesis. *Cancer Letters* **236,** 1–12 (2006).

154. Rössler, U. *et al.* Efficient generation of osteoclasts from human induced pluripotent stem cells and functional investigations of lethal CLCN7-related osteopetrosis. *Journal of Bone and Mineral Research* **36,** 1621–1635 (2021).

155. Gönenc, I. I. *et al.* Single-cell transcription profiles in Bloom syndrome patients link BLM deficiency with altered condensin complex expression signatures. *Human Molecular Genetics* **31,** 2185–2193 (2022).

156. Pudelko, K., Wieland, A., Hennecke, M., Räschle, M. & Bastians, H. Increased microtubule growth triggered by microvesicle-mediated paracrine signaling is required for melanoma cancer cell invasion. *Cancer Research Communications* **2,** 366–379 (2022).

157. Campbell, M. B. *et al.* Bloom syndrome: Research and data priorities for the development of precision medicine as identified by some affected families. *Molecular Case Studies* **4** (2018).

158. Gönenc, I. I. *et al.* Phenotypic spectrum of BLM- and RMI1–related Bloom syndrome. *Clinical Genetics* **101,** 559–564 (2022).

159. Ellis, N. A. *et al.* The Ashkenazic Jewish Bloom syndrome mutation blmAsh is Present in Non-Jewish Americans of Spanish ancestry. *The American Journal of Human Genetics* **63,** 1685–1693 (1998).

160. Li, L., Eng, C., Desnick, R. J., German, J. & Ellis, N. A. Carrier frequency of the Bloom syndrome blmAsh mutation in the Ashkenazi Jewish population. *Molecular Genetics and Metabolism* **64,** 286–290 (1998).

161. Nguyen, G. H. *et al.* Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 9905–9910 (2014).

162. Davies, S. L., North, P. S. & Hickson, I. D. Role for BLM in replication-fork restart and suppression of origin firing after replicative stress. *Nature Structural and Molecular Biology* **14,** 677–679 (2007).

163. Pan, X. *et al.* FANCM, BRCA1, and BLM cooperatively resolve the replication stress at the ALT telomeres. *Proceedings of the National Academy of Sciences* **114,** E5940–E5949 (2017).

164. Böhly, N., Kistner, M. & Bastians, H. Mild replication stress causes aneuploidy by deregulating microtubule dynamics in mitosis. *Cell cycle (Georgetown, Tex.)* **18,** 2770–2783 (2019).

165. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP, Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3,** 861 (2018).

166. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184,** 3573–3587.e29 (2021).

167. Qiu, P. Embracing the dropouts in single-cell RNA-seq analysis. *Nature Communications* **11,** 1–9 (2020).

168. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Research* **33,** 741–748 (2005).

169. De Winter, J. P. & Joenje, H. The genetic and molecular basis of Fanconi anemia. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **668,** 11–19 (2009).

170. Xue, X., Raynard, S., Busygina, V., Singh, A. K. & Sung, P. Role of Replication Protein A in double Holliday junction dissolution mediated by the BLM-Topo IIIα-RMI1-RMI2 protein complex. *Journal of Biological Chemistry* **288,** 14221–14227 (2013).

171. Paul, M. R., Hochwagen, A. & Ercan, S. Condensin action and compaction. *Current Genetics* **65,** 407–415 (2019).

172. Kalitsis, P., Zhang, T., Marshall, K. M., Nielsen, C. F. & Hudson, D. F. Condensin, master organizer of the genome. *Chromosome Research* **25,** 61–76 (2017).

173. Martin, C. A. *et al.* Mutations in genes encoding condensin complex proteins cause microcephaly through decatenation failure at mitosis. *Genes and Development* **30,** 2158–2172 (2016).

174. Khan, T. N. *et al.* Mutations in NCAPG2 cause a severe neurodevelopmental syndrome that expands the phenotypic spectrum of condensinopathies. *American Journal of Human Genetics* **104,** 94–111 (2019).

175. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13,** 731–740 (2016).

176. Concordet, J.-P. & Haeussler, M. CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research* **46,** W242–W245 (2018).

177. Park, J., Lim, K., Kim, J.-S. & Bae, S. Cas-analyzer: An online tool for assessing genome editing results using NGS data. *Bioinformatics (Oxford, England)* **33,** 286–288 (2017).

178. Shinawi, M. & Cheung, S. W. The array CGH and its clinical applications. *Drug Discovery Today* **13,** 760–770 (2008).

179. Bejjani, B. A. & Shaffer, L. G. Application of array-based comparative genomic hybridization to clinical diagnostics. *The Journal of Molecular Diagnostics* **8,** 528–533 (2006).

180. Honma, M. *et al.* Chromosomal instability in B-lymphoblasotoid cell lines from Werner and Bloom syndrome patients. *Mutation Research* **520,** 15–24 (2002).

181. Hustinx, T. W., Scheres, J. M., Weemaes, C. M., ter Haar, B. G. & Janssen, A. H. Karyotype instability with multiple 7/14 and 7/7 rearrangements. *Human Genetics* **49,** 199–208 (1979).

182. Wechsler, T., Newman, S. & West, S. C. Aberrant chromosome morphology in human cells defective for Holliday junction resolution. *Nature* **471,** 642–646 (2011).

183. Gatinois, V. *et al.* iPSC line derived from a Bloom syndrome patient retains an increased disease-specific sister-chromatid exchange activity. *Stem Cell Research* **43,** 101696 (2020).

184. Chan, Y. W., Fugger, K. & West, S. C. Unresolved recombination intermediates lead to ultra-fine anaphase bridges, chromosome breaks and aberrations. *Nature Cell Biology* **20,** 92–103 (2018).

185. Sarlós, K. *et al.* Reconstitution of anaphase DNA bridge recognition and disjunction. *Nature Structural and Molecular Biology* **25,** 868–876 (2018).

186. Baumann, C., Körner, R., Hofmann, K. & Nigg, E. A. PICH, a centromere-associated SNF2 family ATPase, is regulated by Plk1 and required for the spindle checkpoint. *Cell* **128,** 101–114 (2007).

187. Armstrong, C. & Spencer, S. L. Replication-dependent histone biosynthesis is coupled to cell-cycle commitment. *Proceedings of the National Academy of Sciences* **118,** e2100178118 (2021).

188. Hergeth, S. P. & Schneider, R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Reports* **16,** 1439–1453 (2015).

189. Liu, Y. & Zhang, Y. CHCHD2 connects mitochondrial metabolism to apoptosis. *Molecular & Cellular Oncology* **2,** e1004964 (2015).

190. Daley, J. M., Chiba, T., Xue, X., Niu, H. & Sung, P. Multifaceted role of the Topo IIIα–RMI1-RMI2 complex and DNA2 in the BLM-dependent pathway of DNA break end resection. *Nucleic Acids Research* **42,** 11083–11091 (2014).

191. Meetei, A. R. *et al.* A multiprotein nuclear complex connects Fanconi anemia and Bloom syndrome. *Molecular and Cellular Biology* **23,** 3417–3426 (2003).

192. Menger, K. E., Rodríguez-Luis, A., Chapman, J. & Nicholls, T. J. Controlling the topology of mammalian mitochondrial DNA. *Open Biology* **11,** 210168 (2022).

193. Sobek, S. & Boege, F. DNA topoisomerases in mtDNA maintenance and ageing. *Experimental Gerontology* **56,** 135–141 (2014).

194. German, J. Bloom's syndrome. XX. The first 100 cancers. *Cancer Genetics and Cytogenetics* **93,** 100–106 (1997).

195. Breslow, N., Olshan, A., Beckwith, J. B. & Green, D. M. Epidemiology of Wilms tumor. *Medical and Pediatric Oncology* **21,** 172–181 (1993).

196. Jain, D. *et al.* Bloom syndrome in sibs: First reports of hepatocellular carcinoma and Wilms tumor with documented anaplasia and nephrogenic rests. *Pediatric and Developmental Pathology* **4,** 585–589 (2001).

197. Szychot, E., Apps, J. & Pritchard-Jones, K. Wilms' tumor: Biology, diagnosis and treatment. *Translational Pediatrics* **3,** 12–24 (2014).

198. Zhu, H., Swami, U., Preet, R. & Zhang, J. *Harnessing DNA replication stress for novel cancer therapy* 2020.

199. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nature Reviews Cancer* **8,** 193–204 (2008).

200. Datta, A. & Brosh, R. M. New insights into DNA helicases as druggable targets for cancer therapy. **5** (2018).

201. Nguyen, G. H. *et al.* A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. *Chemistry I& Biology* **20,** 55–62 (2013).

202. Yin, Q.-K. *et al.* Discovery of isaindigotone derivatives as novel Bloom's syndrome protein (BLM) helicase inhibitors that disrupt the BLM/DNA interactions and regulate the homologous recombination repair. *Journal of Medicinal Chemistry* **62,** 3147–3162 (2019).

203. Zhang, W. *et al.* Screening antiproliferative drug for breast cancer from bisbenzylisoquinoline alkaloid tetrandrine and fangchinoline derivatives by targeting BLM helicase. *BMC Cancer* **19,** 1009 (2019).

204. Turley, H., Wu, L., Canamero, M., Gatter, K. & Hickson, I. The distribution and expression of the Bloom's syndrome gene product in normal and neoplastic human cells. *British journal of cancer* **85,** 261–265 (2001).

205. Arora, A. *et al.* Transcriptomic and protein expression analysis reveals clinicopathological significance of Bloom syndrome helicase (BLM) in breast cancer. *Molecular Cancer Therapeutics* **14,** 1057–1065 (2015).

206. Zhang, J.-M. M. & Zou, L. Alternative lengthening of telomeres: from molecular mechanisms to therapeutic outlooks. *Cell I& Bioscience* **10,** 30 (2020).

207. Hou, G.-X., Liu, P., Yang, J. & Wen, S. Mining expression and prognosis of topoisomerase isoforms in non-small-cell lung cancer by using Oncomine and Kaplan–Meier plotter. *PLOS ONE* **12,** e0174515 (2017).

208. Zeng, Z. *et al.* Identifying novel therapeutic targets in gastric cancer using genome-wide CRISPR-Cas9 screening. *Oncogene* **41,** 2069–2078 (2022).

209. Montenegro, M. M. *et al.* Gene expression profile suggesting immunological dysregulation in two Brazilian Bloom's syndrome cases. *Molecular Genetics and Genomic Medicine* **8,** 1–10 (2020).

210. Gratia, M. *et al.* Bloom syndrome protein restrains innate immune sensing of micronuclei by cGAS. *Journal of Experimental Medicine* **216,** 1199–1213 (2019).

211. Panneerselvam, J. *et al.* BLM promotes the activation of Fanconi anemia signaling pathway. *Oncotarget* **7,** 32351–32361 (2016).

212. Mjelle, R. *et al.* Cell cycle regulation of human DNA repair and chromatin remodeling genes. *DNA Repair* **30,** 53–67 (2015).

213. Zou, Y. U. E., Liu, Y., Wu, X. & Shell, S. M. Functions of human replication protein A (RPA): From DNA replication to DNA damage and stress responses. *Journal of Cellular Physiology* **208,** 267–273 (2006).

214. Yang, J., Bachrati, C. Z., Hickson, I. D. & Brown, G. W. BLM and RMI1 alleviate RPA inhibition of TopoIIIα decatenase activity. *PLOS ONE* **7,** e41208 (2012).

215. Wang, H. *et al.* BLM prevents instability of structure-forming DNA sequences at common fragile sites. *PLOS Genetics* **14,** 1–23 (2018).

216. Fernández-Casañas, M. & Chan, K. L. The unresolved problem of DNA bridging. *Genes* **9** (2018).

217. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434,** 913–917 (2005).

218. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434,** 917–921 (2005).

219. Kennedy, R. D. & D'Andrea, A. D. DNA repair pathways in clinical practice: Lessons from pediatric cancer susceptibility syndromes. *Journal of Clinical Oncology* **24,** 3799–3808 (2006).

220. Gilmore, E. C. & Walsh, C. A. Genetic causes of microcephaly and lessons for neuronal development. *Wiley Interdisciplinary Reviews: Developmental Biology* **2,** 461–478 (2013).

221. Der Hagen, M. *et al.* Diagnostic approach to microcephaly in childhood: A two-center study and review of the literature. *Developmental Medicine & Child Neurology* **56,** 732–741 (2014).

222. Hirano, T. Condensins: Organizing and segregating the genome. *Current Biology* **15,** R265–R275 (2005).

223. Heale, J. T. *et al.* Condensin I interacts with the PARP-1-XRCC1 complex and functions in DNA single-strand break repair. *Molecular cell* **21,** 837–848 (2006).

224. Wood, J. L., Liang, Y., Li, K. & Chen, J. Microcephalin/MCPH1 associates with the condensin II complex to function in homologous recombination repair. *Journal of Biological Chemistry* **283,** 29586–29592 (2008).

225. Wu, N. & Yu, H. The Smc complexes in DNA damage response. *Cell and Bioscience* **2,** 5 (2012).

226. Marx, V. A dream of single-cell proteomics. *Nature Methods* **16,** 809–812 (2019).

227. Arias-Hidalgo, C. *et al. Single-Cell Proteomics: The Critical Role of Nanotechnology* 2022.

228. Gingold, J., Zhou, R., Lemischka, I. R. & Lee, D.-F. Modeling cancer with pluripotent stem cells. *Trends in cancer* **2,** 485–494 (2016).

229. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).

230. Wiegand, C. & Banerjee, I. Recent advances in the applications of iPSC technology. *Current Opinion in Biotechnology* **60,** 250–258 (2019).

231. Chester, N., Kuo, F., Kozak, C., O'Hara, C. D. & Leder, P. Stage-specific apoptosis, developmental delay, and embryonic lethality in mice homozygous for a targeted disruption in the murine Bloom's syndrome gene. *Genes & Development* **12,** 3382–3393 (1998).

232. Luo, G. *et al.* Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nature Genetics* **26,** 424–429 (2000).

233. Li, W. & Wang, J. C. Mammalian DNA topoisomerase IIIα is essential in early embryogenesis. *Proceedings of the National Academy of Sciences* **95,** 1010–1013 (1998).

234. Avior, Y., Sagi, I. & Benvenisty, N. Pluripotent stem cells in disease modelling and drug discovery. *Nature Reviews Molecular Cell Biology* **17,** 170–182 (2016).

235. Soldner, F. & Jaenisch, R. iPSC Disease Modeling. *Science* **338,** 1155–1156 (2012).

236. Fernández-Santiago, R. *et al.* Aberrant epigenome in iPSC-derived dopaminergic neurons from Parkinson's disease patients. *EMBO Molecular Medicine* **7,** 1529–1546 (2015).

237. Kim, K. *et al.* Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nature Biotechnology* **29,** 1117–1119 (2011).

238. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).

239. Chen, Z. *et al.* Reprogramming progeria fibroblasts re-establishes a normal epigenetic landscape. *Aging Cell* **16,** 870–887 (2017).

240. Kaur, E., Agrawal, R. & Sengupta, S. Functions of BLM helicase in cells: Is it acting like a double-edged sword? *Frontiers in Genetics* **12** (2021).

241. Payne, M. & Hickson, I. D. Genomic instability and cancer: lessons from analysis of Bloom's syndrome. *Biochemical Society Transactions* **37,** 553–559 (2009).

242. Gisselsson, D. Classification of chromosome segregation errors in cancer. *Chromosoma* **117,** 511–519 (2008).

243. Chan, Y. W. & West, S. C. A new class of ultrafine anaphase bridges generated by homologous recombination. *Cell Cycle* **17,** 2101–2109 (2018).

244. Su, C. *et al.* DNA damage induces downregulation of histone gene expression through the G1 checkpoint pathway. *The EMBO Journal* **23,** 1133–1143 (2004).

245. Millán-Ariño, L. *et al.* Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2. *Nucleic Acids Research* **42,** 4474–4493 (2014).

246. Kee, T. R. *et al.* Mitochondrial CHCHD2: Disease-associated mutations, physiological functions, and current animal models. *Frontiers in Aging Neuroscience* **13,** 660843 (2021).

247. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (New York, N.Y.)* **316,** 1484–1488 (2007).

248. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics* **10,** 155–159 (2009).

249. Richard, J. L. C. & Eichhorn, P. J. A. Platforms for investigating lncRNA functions. *SLAS Technology* **23,** 493–506 (2018).

250. Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nature Cell Biology* **16,** 2–9 (2014).

251. Berti, M. & Vindigni, A. Replication stress: Getting back on track. *Nature Structural & Molecular Biology* **23,** 103–109 (2016).

252. Barefield, C. & Karlseder, J. The BLM helicase contributes to telomere maintenance through processing of late-replicating intermediate structures. *Nucleic Acids Research* **40,** 7358–7367 (2012).

List of Figures

1.1 DNA damaging agents creating different types of lesions and relevant DNA repair mechanisms. 6
1.2 Mutation spectrum of *BLM* reported by Bloom Syndrome Registry with respect to schematic representation of BLM helicase. 14
1.3 chematic representation and crystal structure of BLM helicase. 18
1.4 Schematic representation and crystal structure of TOP3A. 21
1.5 Schematic representations of RMI1 and RMI2 proteins and conserved domains. 22
1.6 Schematic representation of double Holliday junction (dHJ) dissolution by the BTRR complex. 24

3.1 “Bloom syndrome and the overlapping phenotypes” flyer. 53
3.2 Genetic confirmation of the BS patient-derived fibroblast cell lines. 59
3.3 Metaphase karyotype of the BS patient-derived fibroblast sample Goe1306 60
3.4 Well content assessment of the nanowell chip by fluorescent imaging . . 62
3.5 Classification of the reads generated by scRNA sequencing data of BS patient-derived and WT fibroblasts 63
3.6 Quality control assessment of scRNAseq data 64
3.7 UMAP analysis showing gene expression profiles of each different fibroblast sample 65
3.8 Differentially expressed genes in BS single cells. 67
3.9 Pathway analyses of DEGs in BS single cells compared to wild-type single cells. 69
3.10 KEGG-pathway analysis of the DEGs of Fanconi anemia pathway 71
3.11 Volcano plot analysis of gene expression in BS single cells visualizing genes of the condensin I/II complexes and microcephaly-associated genes. 72
3.12 Expression levels of SMC2 and SMC4 in control and BS patient-derived single fibroblasts 74
3.13 Analyses of SMC2 and SMC4 protein expression levels. 76
3.14 Mass spectrometry analyses of protein expression in BS patient-derived fibroblasts compared to WT cells. 78
3.15 Workflow of the CRISPR/Cas9-based genome editing strategy 80
3.16 Sanger sequencing analyses of the KO iPSC clones. 85
3.17 Western blot analyses of target protein expression in CRISPR/Cas9-based KO iPSC clones. 87
3.18 Morphological characterization and pluripotency assessments of the generated KO iPSC clones. 88
3.19 Flow cytometry analyses of protein expression of pluripotency markers OCT4 and TRA-1-60 89
3.20 Metaphase karyotype analyses of the generated KO iPSC clones 91

3.21 Analyses of sister chromatid exchange rates in KO iPSC clones and WT control cells. 94
3.22 Determination of chromosome segregation defects during anaphase in BLM-, TOP3A- and RMI1-KO iPSC samples. 96
3.23 Analysis of ultrafine bridges in BLM-, TOP3A- and RMI1-KO iPSC samples. 98
3.24 Classification of reads generated by scRNAseq of the wild-type, BLM-, TOP3A-, and RMI1-KO iPSCs 100
3.25 Quality control assessment of scRNAseq data of wild-type and BLM-, TOP3A-, and RMI1-KO iPSC samples 101
3.26 The UMAP analysis showing gene expression profiles of wild-type and BLM-, TOP3A-, and RMI1-KO iPSC samples. 102
3.27 Differentially expressed genes in BLM-, TOP3A-, and RMI1-KO iPSCs. . . 104
3.28 Overlapping significant genes in the comparisons of DEGs in KO iPSCs against the parental WT control. 105

List of Tables

1.1 Clinical chracteristics of BS (MIM 210900) 11

2.1 Antibodies used in Western blot experiments 28
2.2 Antibodies used for the immunocytochemistry assays 29
2.3 Fibroblast samples used in this study 29
2.4 The iPSC lines used in this thesis. 30
2.5 Reference sequences 30
2.6 Primer sequences and their applications 30
2.7 guideRNA sequences for target genes of the BTRR complex 32
2.8 Chemicals used during the doctoral thesis. 33
2.9 Kits used for each corresponding application. 34
2.10 Non-expandable equipments 35
2.11 Softwares used and their application 37
2.12 Databases and online tools 37

3.1 Clinical characteristics of molecularly diagnosed BS patients. 56
3.2 Differential expressions of condensin I/II complex genes in BS single cells. 74
3.3 BS patient mutations used as reference regions in genome editing. 81
3.4 Experimental flow and the transfection efficiencies of genome editing. . . 82
3.5 Allele details of the iPSC clones for the corresponding genes. 84
3.6 Numbers of significant DEGs and GO pathways of from each comparison 107
3.7 *SMC2/4* containing deregulated GO terms in aphidicolin treated iPSCs and the respective enrichment scores according to GSEA. 109

S1 DEG analyses of BS patient derived-fibroblasts in comparison to WT single cells. 142
S2 Significant deregulated GO terms in the BLM-KO clone obtained by the GSEA 144
S3 Significant deregulated GO terms in the TOP3A-KO clone obtained by the GSEA 145
S4 Significant deregulated GO terms in the RMI1-KO clone obtained by the GSEA. 145
S5 Top 20 most significantly deregulated terms in aphidicolin-treated BLM-KO iPSCs. 146
S6 Top 20 most significantly deregulated terms in aphidicolin-treated TOP3A-KO iPSCs 147
S7 Top 20 most deregulated GO terms in aphidicolin treated versun non-treated WT iPSCs. 148

Abbreviations

µg	Microgram
µl	Microliter
µm	Micrometer
A	Adenine
Adj. p-value	Adjusted p-value (Bejamini-Hochberg method)
Ala	Alanine
ALL	Acute lymphatic lymphoma
AML	Acute myeloid lymphoma
arrayCGH	Array-based comparative genomic hybridization
ATP	Adenosine triphosphate
AveExpr	Average expression
BCA	Bicinchoninic acid assay
BER	Base excision repair
BLM	BLM RecQ-like helicase
BLM-KO	isWT1-BLM-KO.1D1
bp	Base pair
BrdU	5-bromo-2'-deoxyuridine
BS	Bloom syndrome
BSA	Bovine serum albumin
BSR	Bloom Syndrome Registry
C	Cytosine
cDNA	Complementary DNA
cm	Centimeter
cm	Centimeter
CNV	Copy number variation
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR-RNA
Cys	Cysteine
DAPI	4´,6-diamidino-2-phenylindole
DBPS	Dulbecco's phosphate-buffered saline
DDR	DNA damage response
DEG	Differentially expressed gene
dHJ	Double Holliday junction
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid

dNTP	Deoxynucleotide triphosphate
DSB	Double-strand break
EDTA	Ethylenediaminetetraacetic acid
FA	Fanconi anemia
FACS	Fluorescence-activated cell sorting
FANCD2	FA complementation group D2
FANCM	FA Complementation Group M
FBS	Fetal bovine serum
FDR	False discovery rate
FIJI	Fiji Is Just ImageJ
G	Guanine
G4 DNA	G-quadruplexes
GO	Gene ontology
GSEA	Gene set enrichment analysis
HCA	The Human Cell Atlas
HGPS	Hutchinson-Gilford progeria syndrome
HR	Homologous recombination
HRDC	RNase D-like C-terminal
HRP	Horseradish peroxidase
ICL	Interstrand crosslink
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
iPSC	Induced pluripotent stem cell
J	Joule
kb	Kilobase
KCl	Potassium chloride
kg	Kilogram
KO	Knockout
Leu	Leucine
LIG4	DNA ligase 4
lncRNA	Long non-coding RNA
LoF	Loss-of-function
LogFC	Log2 fold change
Lys	Lysine
M	Molar
MCPH	Autosomal recessive primary microcephaly
Met	Methionine
min	Minute
miRNA	MicroRNA

ml	Mililiter
mm	Milimeter
MMR	Mismatch repair
MN	Micronuclei
mRNA	Messenger RNA
MS	Mass spectrometry
mtDNA	Mitochondrial DNA
NaN3	Sodium azide
NANOG	Nanog homeobox
NCAPD2	Non-SMC condensin I complex subunit D2
NCAPD3	Non-SMC condensin II complex subunit D3
NCAPG	Non-SMC condensin I complex subunit G
NCAPG2	Non-SMC condensin II complex subunit G2
NCAPH	Non-SMC condensin I complex subunit H
NCAPH2	Non-SMC condensin II complex subunit H2
NER	Nucleotide excision repair
NES	Normalized enrichment score
NGS	Next generation sequencing
NH4Cl	Ammonium chloride
NHEJ	Non-homologous end joining
nl	Nanoliter
NP40	Nonyl phenoxypolyethoxylethanol
OCT3/4	Octamer-binding transcription factor 3/4
OFC	Occipitofrontal head circumference
ORA	Over-representation analysis
P/S	Pen/Strep
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
PDVF	Polyvinylidene difluoride
PFA	Paraformaldehyde
PICH	ERCC6L, ERCC excision repair 6 like,
RMI1	RecQ mediated genome instability 1
RMI1-KO	isWT1-RMI1-KO.1F11
RMI2	RecQ mediated genome instability 2
RNA	Ribonucleic acid
RPA	Replication protein A
rpm	Round per minute
SCE	Sister chromatid exchange
scRNAseq	Single-cell RNA sequencing

SCU	Stem Cell Unit
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
sec	Second
Ser	Serine
SMC2	Structural maintenance of chromosomes 2
SMC4	Structural maintenance of chromosomes 4
SOX2	SRY-box transcription factor 2
SRP	Signal-recognition particle
SSB	Single-stranded break
SSC	Saline-sodium citrate
SSEA4	Stage specific embryo antigen 4
T	Thymidine
T2T	Telomere-to-Telomere
TBE	Tris-borate-EDTA
TBST	Tris-buffered saline, Tween 20
TCEP	TRIS(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TGS	Tris-Glycine-SDS
TOP3A	DNA topoisomerase III alpha
TOP3A-KO	isWT1-TOP3A-KO.1E10
TOPRIM	Topoisomerase-primase subdomain
TOUCH58	Touchdown-like 58 PCR program
tracrRNA	Trans-activating CRISPR RNA
TRIS	Tris(hydroxymethyl)aminomethane
TU	Technical University
TZV	Thiazovivin
UMAP	Uniform Manifold Approximation and Projection
UMG	University Medical Center Göttingen
UV	Ultraviolet
V	Volt
vol	Volume
WES	Whole exome sequencing
WGS	Whole genome sequencing

Acknowledgements

The final part of this thesis is the most important part for me since this work would not be possible without teamwork and support. First, I am extremely grateful to Prof. Dr. Bernd Wollnik for giving me the great opportunity to work on this project. I am further indebted to Bernd for the wonderful support, constant motivation, and professional mentorship and guidance throughout this journey which helped me greatly on the way to becoming a competent researcher.

I want to thank my thesis committee; Prof. Dr. Matthias Dobbelstein and Dr. Lukas Cyganek for supporting the project with helpful meetings. I would like to thank the members of the examination board Prof. Dr. Uwe Kornak, Prof. Dr. Holger Bastians, and Prof. Dr. Wolfram Zimmermann for assessing this work by participating in the extended committee.

I am grateful to all members of the Institute of Human Genetics. Many thanks to, especially Gökhan Yigit for the insightful scientific input for manuscript preparations, presentations, and editing process of this thesis. Further special thanks to the members of the Wollnik lab, thank you; Yun Li for teaching me exome data analysis and fruitful conversations; and Christian Müller for excellent technical support in the lab, fun talks, and training. I would like to thank the co-first-author in my first published article, Alexander Wolff for the great support of bioinformatics and for being available always for questions. Many thanks to Karin Boss for the great support on graphics and post-production in the scientific writings. I am grateful to Barbara Vona for the daily support and motivation at our shared office. I thank the team of Kornak lab for sharing the iPSC culture smoothly and for further friendships: especially Floriane Hennig, Regina Grün, Dario Gajewski, and Angelika Biedermann. Further special thanks to the valuable team of Institute of Human Genetics; Silke Kaulfuß, Peter Burfeind, Loukas Argyriou, Arne Zibat, Julia Schmidt, Nina Bögershausen, Maria Kuzyakova, Tobias Scherf de Almeida, Petra Albers, Manuela Manafas, Carolin Stecker,Ilona Eggert, Sabine Herold, Gabriela Salinas, Orr Shomroni, Maren Sitte, and Fabian Ludewig, and members of diagnostic team who helped processing the experiments.

I would like to extend my sincere thanks to the Stem Cell Unit team, especially Lukas Cyganek for accepting me as a student and Alexandra Busley for teaching me the iPSC

culture and for her supportive friendship over the years. Thanks, Mandy Kleinsorge and Óscar Gutiérrez-Gutiérrez for the singularization experiments; thanks Laura Cyganek and Yvonne Hintz for taking care of the cells, and many thanks to the rest of the SCU team for their hospitality.

I want to further thank the members of the FOR2800 research consortium. Many cordial thanks to Markus Räschle and Angela Wieland for several collaborations and their excellent hospitality at Kaiserslautern. Thank you Angela for being a great friend in the last years.

I am grateful to my friends who supported me on this journey from various places. Thank you Deniz Yüzak, Yunus Can Erol, and Dinah Burfeind for making the life easier during my first days in Göttingen and for the friendship in the following years. Many thanks to my friends from Istanbul, Gamze Topçu and Melis Türer, for their sincere support for many years, and Sila Unal for being a wonderful friend for more than 20 years.

Last but not least, I am extremely thankful to my family. Many cordial thanks Alev Heilbronn and Kurt Heilbronn for the wonderful support as a part of my family here in Germany. Further cordial thanks to my dearest aunt, Seher Başaran, for guiding me to the genetics field and for the solid support on personal and professional subjects. My amazing parents, Sedef Gönenç, and Kerem Gönenç, I cannot thank you enough for your infinite love and support, I am very fortunate. Finally, I am indebted to Antxon Smith López, for the patience and understanding over the years and especially last months while writing this thesis, thank you.