Genome-Wide Analysis of the Nucleosome Landscape in Individuals with Coffin-Siris Syndrome

Alexander Kalmbach^a Christopher Schröder^b Ludger Klein-Hitpass^c
Karl Nordström^d Peter Ulz^e Ellen Heitzer^e Michael R. Speicher^e
Sven Rahmann^b Dagmar Wieczorek^f Bernhard Horsthemke^a
Nuria C. Bramswig^a

^aInstitute of Human Genetics, University Hospital Essen, University Duisburg-Essen, Essen, Germany; ^bGenome Informatics, Institute of Human Genetics, University Duisburg-Essen, Essen, Germany; ^cBiochip Lab, Institute of Cell Biology, University Duisburg-Essen, Essen, Germany; ^dInstitute of Genetics/Epigenetics, Saarland University, Saarbrücken, Germany; ^eInstitute of Human Genetics, Diagnostic and Research Center for Molecular Biomedicine, Medical University of Graz, Graz, Austria; ^fInstitute of Human Genetics, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Keywords

Coffin-Siris syndrome · Epigenetics · Monocytes · NOME-seq · SWI/SNF

Abstract

The switch/sucrose non-fermenting (SWI/SNF) complex is an ATP-dependent chromatin remodeller that regulates the spacing of nucleosomes and thereby controls gene expression. Heterozygous mutations in genes encoding subunits of the SWI/SNF complex have been reported in individuals with Coffin-Siris syndrome (CSS), with the majority of the mutations in *ARID1B*. CSS is a rare congenital disorder characterized by facial dysmorphisms, digital anomalies, and variable intellectual disability. We hypothesized that mutations in genes encoding subunits of the ubiquitously expressed SWI/SNF complex may lead to alterations of the nucleosome profiles in different cell types. We performed the first study on CSS-patient samples and investigated the nucleosome land-

scapes of cell-free DNA (cfDNA) isolated from blood plasma by whole-genome sequencing. In addition, we studied the nucleosome landscapes of CD14⁺ monocytes from CSS-affected individuals by nucleosome occupancy and methylome-sequencing (NOME-seq) as well as their expression profiles. In cfDNA of CSS-affected individuals with heterozygous *ARID1B* mutations, we did not observe major changes in the nucleosome profile around transcription start sites. In CD14⁺ monocytes, we found few genomic regions with different nucleosome occupancy when compared to controls. RNA-seq analysis of CD14⁺ monocytes of these individuals detected only few differentially expressed genes, which were not in proximity to any of the identified differential nucleosome-depleted regions. In conclusion, we show that heterozygous mutations in the human SWI/SNF subunit *ARID1B* do not have a major impact on the nucleosome landscape or gene expression in blood cells. This might be due to functional redundancy, cell-type specificity, or alternative functions of *ARID1B*.

© 2019 S. Karger AG, Basel

Fig. 1. Mutations in the SWI/SNF complex may alter the nucleosome landscape in individuals with Coffin-Siris syndrome (CSS). **a** The SWI/SNF complex remodels the chromatin landscape by shifting nucleosomes to enable transcription factor binding and gene expression. The subunit ARID1B (red circle, left) is mutated in individuals with CSS (interrupted red circle, right) which may lead to impaired remodelling capability and may have an effect on nucleosome positioning and the subsequent gene expression. CTRL, control; exemplary SWI/SNF subunits in descending order of molecular size: SMARCA4 (yellow), SMARCC2 (blue), SMARCC1 (green), BRD9 (magenta), SMARCD1 (yellow green),

SMARCE1 (orange), ACTL6A (grey), SMARCB1 (light brown), SS18 (purple), DPF2 (dark green), ACTB (dark brown), BCL7A (black). **b** Workflow illustrating the methods used to investigate the proposed hypothesis. After blood draw, nucleosome profiles were generated from sequencing cell-free DNA from blood plasma (right). In addition, monocytes were isolated for chromatin accessibility and DNA methylation analysis by NOMe-seq (middle), as well as for RNA extraction and transcriptome sequencing (left). DMR, differentially methylated region; DNR, differential nucleosome region.

For spatial reasons, about 2 metres of human genomic DNA are tightly wound around nucleosomes to fit in the approximately 10- μ m cell nucleus. These nucleosomes do not only have a packaging role, but the nucleosome positioning also plays an important role in gene regulation with a nucleosome-depleted region (NDR) at the transcription start site (TSS) of expressed genes to enable transcription factor binding [Yuan et al., 2005]. So-called chromatin remodelling complexes facilitate the placement of nucleosomes. One of the first-discovered and most-studied of these complexes is the switch/sucrose non-fermenting (SWI/SNF) complex, an ATP-dependent chromatin remodeller with a helicase-like ATPase domain [Bork and Koonin, 1993]. This complex was first identified in yeast and named for the observed phenotype as mating type switching-deficient SWI1 [Haber and Garvik, 1977] and sucrose non-fermenting SNF1 [Carlson et al., 1981]. The human orthologue of the SWI/SNF complex, sometimes referred to as BRG1-

associated factors (BAF) complex, has conserved subunits from yeast but also incorporates distinct mammalian subunits. The core of the SWI/SNF complex is formed on the basis of a dimer of SMARCC1 and SMARCC2 and one of the 3 SMARCD subunits [Mashtalir et al., 2018]. Addition of SMARCE1 and SMARCB1 then leads to the formation of the SWI/SNF core module. Similar to the yeast complex, the human SWI/SNF complex remodels chromatin structure in an ATP-dependent manner and thereby enables transcription factor binding [Kwon et al., 1994] and activation of promoters [Imbalzano et al., 1994]. The SWI/SNF complex is ubiquitously expressed, but tissue-specific expression of certain subunits and complex compositions have been reported [Olave et al., 2002; Wu et al., 2007; Ho et al., 2009]. This specificity is also evident as several reports demonstrated the importance of the SWI/SNF complex in the differentiation of several lineages, e.g. haematopoietic [Chi et al., 2002; Gebuhr et al., 2003]

and neural differentiation [Lessard et al., 2007]. The largest and broadly expressed accessory subunit is the AT-rich interactive domain protein (ARID1) subunit, which has 2 very similar but mutually exclusive versions: ARID1A and ARID1B [Inoue et al., 2002]. ARID1A has been shown to be more abundant than its counterpart ARID1B and therefore might incorporate disproportionately into the SWI/SNF complex [Kelso et al., 2017]. ARID1B has been shown to affect cell growth and cell cycle regulation and to target related genes including *c-Myc* and *p21* [Inoue et al., 2011].

The first direct link between *ARID1B* and intellectual disability (ID) was made in a patient with ID, agenesis of the corpus callosum, and speech delay who had a translocation between chromosomes 6 and 14, t(6;14)(q25.3;q13.2) [Backx et al., 2011]. Heterozygous mutations of *ARID1B* leading to haploinsufficiency have since been found to be a frequent cause of ID [Hoyer et al., 2012]. Coffin-Siris syndrome (CSS) is a rare congenital disorder and was first described in 1970 [Coffin and Siris, 1970]. CSS-affected individuals present with varying degrees of ID, hypo- or aplasia of fifth finger- or toenails and distal phalanges, coarse facial features with a flat nasal bridge and a wide mouth with thick lips, hirsutism, and short stature [Fleck et al., 2001]. In 2012, the genetic link between *ARID1B* and CSS was first reported [Santen et al., 2012; Tsurusaki et al., 2012]. Subsequently, heterozygous mutations in several other genes encoding subunits of the SWI/SNF complex, e.g. *SMARCB1*, *SMARCE1*, *SMARCA4*, and *ARID1A*, have been described, with heterozygous *ARID1B* mutations being the most common cause for CSS [Santen et al., 2013; Wieczorek et al., 2013; Tsurusaki et al., 2014].

Impaired function of the SWI/SNF complex has been shown to affect the nucleosome positioning at several promoters leading to changes in gene expression in primary mouse cells [Tolstorukov et al., 2013]. The goal of this study was to investigate whether heterozygous mutations in genes coding for subunits of the SWI/SNF complex in individuals with CSS lead to alterations of their nucleosome profiles (Fig. 1a). We analysed the nucleosome landscape of blood cells by inferring nucleosome occupancy from cell-free DNA (cfDNA) utilising whole-genome sequencing. In addition, we generated genome-wide nucleosome profiles of CD14⁺ monocytes from CSS-affected individuals using nucleosome occupancy and methylation sequencing (NOME-seq). We then focused on differentially occupied regions and examined the possible implications on gene regulation by RNA-seq analysis (Fig. 1b).

Table 1. Control individuals and CSS-affected individuals

Individual	Age	Sex	Mutation in <i>ARID1B</i>
<i>Controls</i>			
K001	10 y 2 mo	f	–
K002	11 y 10 mo	f	–
K004	8 y 5 mo	m	–
K010	10 y 11 mo	f	–
K012	15 y	m	–
K013	13 y 11 mo	f	–
K014	21 y	m	–
K015	18 y 7 mo	f	–
K017	9 y	m	–
K018	5 y 2 mo	f	–
K019	6 y	m	–
K020	9 y 7 mo	m	–
K021	7 y 7 mo	m	–
<i>CSS-affected</i>			
C002	3 y	m	c.5547dupC; p.Ser1851Lysfs*5
C003	3 y 9 mo	m	c.5728delG; p.Ala1910Profs*64
C004 ^a	10 y 3 mo	f	dup ex5–6
C005	6 y	m	c.3919C>T; p.Gln1307Ter
C006	19 y 9 mo	f	c.3255dupT; p.Pro1086Serfs*32

mo, months; y, years.

^a Previously published by Hoyer et al. [2012].

Material and Methods

Patient Cohort

Blood was collected from individuals presenting with clinically diagnosed CSS (CXXX) during routine examination. The control cohort (KXXX) was collected from individuals without any signs of ID or developmental delay during routine examination at the University Hospital Essen. An overview of individuals and the analyses is supplied in Table 1 and online supplementary Table 1 (see www.karger.com/doi/10.1159/000503266 for all online suppl. material).

Isolation and Sequencing of cfDNA

Blood for cfDNA analysis was collected in Cell-Free DNA BCT[®] tubes (Streck, Inc., La Vista, NE, USA), which stabilizes nucleated blood cells and prevents the release of genomic DNA. cfDNA was isolated using the QIAamp Circulating Nucleic Acid Kit (QIAGEN, Hilden, Germany). Libraries for shallow whole-genome sequencing were prepared as described previously [Heitzer et al., 2013] and sequenced on an Illumina NextSeq 550 (Illumina[®], Inc., San Diego, CA, USA) (on average 24,600,000 reads per sample, ranging from 19,200,000 to 43,400,000 reads) according to Ulz et al. [2016].

Monocyte Isolation

Monocytes were isolated from 6 mL whole blood using Straight-From[™] Whole Blood CD14 MicroBeads (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) according to the manufacturer's protocol. In short, whole blood was mixed with beads immediately after blood draw, incubated at 4°C, mixed with 5 volumes separation buffer (2 mM EDTA, 0.5% BSA in 1× PBS, pH 7.2) and

spun at 433 g for 10 min in a swinging-bucket rotor without brake. The magnetically labelled suspension was applied on a Whole Blood Column in a MidiMACS™ Separator (Miltenyi Biotec GmbH) and, after washing twice with separation buffer, the cells were eluted in Whole Blood Column Elution Buffer.

The monocytes were centrifuged at 300 g at 4°C for 5 min and resuspended in cold DPBS without calcium and magnesium. The cells were counted and aliquoted into ~400,000 cells; surplus cells were added to the aliquot designated for RNA isolation. Cells were pelleted at 800 g at 4°C for 5 min. The pellet for RNA isolation was immediately flash frozen with dry ice/ethanol, whereas the remaining pellets were fixed with 1% formaldehyde in DPBS for 10 min. The fixation was quenched with 125 mM glycine in H₂O for 10 min, and pelleted cells were frozen with dry ice/ethanol and stored at -80°C until further use.

Methyltransferase Reaction (NOME-seq)

NOME-seq was performed essentially as described by Kelly et al. [2012] and Wallner et al. [2016]. Nuclei were extracted from fixed cells using the nuclei extraction with sonication (NEXSON) method [Arrighoni et al., 2016].

Cell pellets were resuspended in nuclei isolation buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% IGEPAL, protease inhibitor cocktail). Sonication was carried out in 130 µL on a Covaris™ S220 instrument (Applied Biosystems™, part of Thermo Fisher Scientific Inc., Waltham, MA, USA) using the following parameters: 75W peak incident power, duty factor 2%, 200 cycles per burst for 90 s.

After sonication, the nuclei were inspected under a light microscope. After this quality assessment, the nuclei were pelleted by centrifugation (1,000 g, 5 min, 4°C) and resuspended in NOME reaction mix (50 mM Tris-HCl pH 8.5, 50 mM NaCl, 10 mM DTT, 280 mM sucrose, 300 µM S-adenosylmethionine). The reaction mix was supplemented with 60 U of the GpC-methyltransferase M.CviPI, resulting in a total volume of 160 µL. The reaction was incubated for 3 h at 37°C, and 100 µM SAM was added every hour.

After incubation, an equal volume of NOME stop buffer (20 mM Tris-HCl pH 8.0, 600 mM NaCl, 1% SDS, 10 mM EDTA) supplemented with proteinase K was added. The reaction was stopped by overnight incubation at 65°C and shaking at 500 rpm.

Then, 20 µL of this crude reaction mix was directly used for bisulfite conversion and chromatin control amplicon PCR.

Genomic DNA was extracted with phenol/chloroform and ethanol precipitated.

Genomic DNA was bisulfite-converted with the EZ Methylation-Gold™ Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's protocol.

The libraries for whole-genome bisulfite sequencing were prepared with the TruSeq® DNA Methylation Kit (Illumina) according to the manufacturer's protocol.

Briefly, up to 50 ng of bisulfite-converted genomic DNA was used for library preparation. The final PCR was done with 1 µL index primer (TruSeq® DNA Methylation Indexes; Illumina) instead of the reverse primer. The resulting library was purified with 0.85X Agencourt® AMPure® XP beads (Beckman Coulter GmbH, Krefeld, Germany).

Quality control of the libraries was carried out with a High Sensitivity DNA Chip (Agilent Technologies, Santa Clara, CA, USA) on a 2100 Bioanalyzer Instrument (Agilent Technologies) according to the manufacturer's protocol.

Each library was sequenced in 1 flowcell-lane (100 bp paired-end reads) on a HiSeq 2500 System (Illumina). Raw data were aligned using *bwa-meth* [Pedersen et al., 2014] and processed as described by Wallner et al. [2016] to retrieve methylation values in CpG and GpC context.

RNA-Seq Analysis

RNA was isolated from cells using the Quick-RNA™ Mini-Prep-Kit (Zymo Research) according to the manufacturer's protocol. RNA-seq libraries were prepared using the Ovation® RNA-Seq FFPE System (NuGEN Technologies, Inc., San Carlos, CA, USA) and sequenced on a HiSeq 2500 System (Illumina) resulting in an average of 50,000,000 paired-end reads per sample.

Bioinformatics

The workflow depicting the bioinformatics pipeline is supplied in online supplementary Figure 1.

cfDNA Profiles

Analysis of cfDNA sequences was done as described by Ulz et al. [2016]. Briefly, raw sequencing reads were trimmed to contain only bases 53–113. This should include the central 60 bp of a theoretical 166-bp fragment in order to amplify the nucleosome signal. Trimmed reads were aligned to the human hg19 genome using *bwa-meth* [Pedersen et al., 2014], which is based on *bwa-mem* [Li and Durbin, 2009]. Then, coverage values were extracted for -1,000 and +1,000 bp around the TSS of housekeeping genes [Eisenberg and Levanon, 2013] using *pysam* (<https://github.com/pysam-developers/pysam>) [Li et al., 2009] and in-house scripts. Coverage values are averaged for every position in the 2,000-bp window.

Moreover, genes were ordered in quintiles according to their expression in cell-free RNA of 4 non-pregnant females [Koh et al., 2014]. Coverage profiles around TSS were generated by averaging coverage values per position for every TSS in the respective quintile.

Detection of Differential Nucleosome Regions

For each sample, NDRs were called as described by Wallner et al. [2016]. The union of NDRs in all samples was used to limit the detection of differential nucleosome regions (DNRs). Our approach involved a modified analysis with *metilene* [Jühling et al., 2016], usually used to investigate endogenous DNA methylation. In our case, we studied the artificially introduced GpC-methylation to find DNRs. Reported regions were set to encompass at least 5 GpCs, coverage of 5×, and a *q*-value cut-off of 0.05. Pair-wise qq-plots revealed mostly consistent behaviour between GpC methylation distributions (online suppl. Fig. 2). The most common deviation was one-sided, increasing at low methylation and constant at high methylation. Although the effect was small, each sample was normalised to the third quartile in order to avoid effects based on potentially varying efficiencies in the M.CviPI treatment in accessible regions.

Resulting regions were compared with the imputed ChromHMM [Ernst and Kellis, 2015; Roadmap Epigenomics Consortium, 2015] states for CD14⁺ monocytes viewed in the WashU EpiGenome Browser v46.1 based on the hg19 genome assembly. The regions were checked for annotations with the Ensembl Regulatory Build [Zerbino et al., 2016] for CD14⁺ monocytes in the Ensembl browser for the GRCh37 genome assembly.

Resulting regions were validated with targeted deep bisulfite amplicon sequencing using the MiSeq system (Illumina). Amplicons were analysed with the in-house developed *Amplifyer2* Software [Leitão et al., 2018].

Differential Gene Expression Analysis

Differential expression of genes was analysed on the Galaxy server (<https://usegalaxy.eu/>) using 2 approaches. First, fastq reads were mapped with *STAR* (Galaxy version: 2.5.2b-0) [Dobin et al., 2013] using default settings against the Ensembl primary genome assembly (GRCh37.75). Mapped reads were counted with *htseq-count* (Galaxy version: 0.6.1galaxy2) [Anders et al., 2015] and Ensembl*gtf (GRCh37.75) with default settings. Count tables were fed into *DESeq2* (Galaxy version: 2.11.39) [Love et al., 2014] and run with default settings.

In the second approach, fastq reads were quasi-mapped using *Salmon* (Galaxy version: 0.8.2) [Patro et al., 2015] with fragment GC bias correction against Ensembl transcriptome, which was a concatenation of noncoding RNA- and cDNA-sequences (GRCh37.75). The GC bias correction was implemented because a slight bias was discovered with *computeGCBias* (Galaxy Version 3.0.2.0). The resulting *Salmon* output was converted and fed into *DESeq2* via *tximport*, which was an available option in the Freiburg Galaxy instance. In both approaches, controls were compared with the patient cohort; where applicable a *DESeq2*-implemented correction for sex was applied. Significantly differentially expressed genes in the *DESeq2* output were determined by padj < 0.05.

The results of both pipelines were combined and the overlap considered robustly differentially expressed genes. Ensembl-IDs were converted to HUGO nomenclature with the Galaxy tool *join two datasets* (Galaxy version: 2.0.2) using a table containing Ensembl-IDs and corresponding HUGO names. Regulatory pathway analysis of significantly deregulated genes was performed in *DAVID* [Huang et al., 2009a, b] and *GeneTrail2* [Stöckel et al., 2016].

Results

Analysis of nucleosome profiles was performed on whole blood cfDNA, an easily accessible source of DNA from a heterogeneous cell population, and on isolated CD14⁺ monocytes for cell type-specific nucleosome profiles.

Nucleosome Profiles from cfDNA

We analysed the nucleosome landscapes of whole blood cfDNA from 5 individuals with clinically diagnosed CSS and compared them with whole blood cfDNA from 8 age- and sex-matched control individuals (Table 1). We plotted the nucleosome occupancy derived from cfDNA around the TSS of housekeeping genes in CSS and control individuals and observed no difference of NDR and conservatively positioned nucleosomes (−1, +1, +2, +3) (Fig. 2a). The nucleosome profile of K019

showed a peak in the NDR, which was not present in the other samples. Next, we separated the nucleosome profiles in quintiles according to the mean expression in blood, but the coverage around the TSS and the resulting nucleosome profile did not differ between controls and individuals with CSS (online suppl. Fig. 3).

Influence of Mutated SWI/SNF Subunits on the Nucleosome Landscape in Patient-Derived Monocytes

Whole blood cfDNA contains DNA of a combination of apoptotic cells from different haematopoietic lineages. Hence, it represents a mixture of chromatin from a heterogeneous cell population. To exclude any effect of cell type-dependent epigenomic variability, we sought to investigate the genome-wide nucleosome profile of an easily accessible, homogeneous cell population by isolating CD14⁺ monocytes from whole blood. For this, we used NOME-seq. Sequencing resulted in an average genome coverage of approximately 4×, ranging from 1× to 6×. Due to the chemistry of the library preparation, GC-rich regions are covered more deeply (up to 100×) than other regions. For analysing cell type-specific nucleosome profiles, we derived monocyte-specific nucleosome profiles from the NOME-seq data of 2 male patients (C002 and C003) and 3 male controls (K004, K012, K017). We plotted the whole-genome methylation data around CpG island (CGI)-containing TSS and found no major changes in the nucleosome occupancy in monocytes between controls and CSS-affected individuals (Fig. 2b).

To assess small/localised GpC-methylation differences, we normalised the whole-genome bisulfite sequencing data to the third quartile after pairwise qq-plot comparison (online suppl. Fig. 2) and searched for DNRs using the following thresholds: at least 5 GpCs, coverage of 5×, and a *q*-value cut-off of 0.05. We revealed 3 differentially occupied regions with about 20% methylation difference (Table 2; online suppl. Fig. 4). One 245-bp region was identified on chromosome 16 in the first intron of *USP7* (methylation difference: 21%, *q*-value 0.9×10^{-4}) and is annotated as open chromatin in the Ensembl regulatory build for monocytes. This region was reported as “more accessible” in the controls, i.e. the chromatin was more closed in CSS individuals. A second region with a size of 238 bp was found on chromosome 4 in the promoter of *AFAP1* (methylation difference: 26%, *q*-value 0.038), which is usually not active in monocytes according to Ensembl. The third region encompassed 103 bp (methylation difference: 22%, *q*-value 0.013) and is an intergenic region predicted to be a CTCF- or TF-binding site locat-

Fig. 2. Genome-wide profiles of the nucleosome occupancy around the transcription start sites (TSS). **a** Read depth of sequenced cfDNA around TSS of housekeeping genes showing the nucleosome-depleted region (NDR) at the TSS as well as the conservatively positioned nucleosomes, most prominently +1, +2, and +3 nucleosomes. No difference in nucleosome profiles was observed

between controls and CSS. The profile of K019 showed a peak in the NDR, which was not present in the other samples. **b** Nucleosome profile around the CpG island-containing TSS presenting with the NDR and +1 nucleosome, indicating no differences in the genome-wide mean nucleosome profile in patient-derived CD14⁺ monocytes.

Table 2. Differential nucleosome regions

DNR	Genomic region, hg19	Size, bp	ChromHMM state	Annotation, Ensembl regulatory build	Methylation level difference	q-value
USP7-DNR	chr16:9,035,680–9,035,924	245	Weak enhancer	Open chromatin	0.21	9.1×10^{-5}
AFAP1-DNR	chr4:7,941,465–7,941,702	238	Active TSS	Promoter	–0.26	0.038
IGLV3-10-DNR	chr22:23,152,591–23,152,693	103	DNase	CTCF-/TF-binding	–0.22	0.013

Male individuals C002 and C003 carrying *ARID1B* mutations were compared with male controls K004, K012, and K017. DNR, differential nucleosome regions; TSS, transcription start site.

ed upstream of *IGLV3-10*. The second and third regions were reported to be more accessible in individuals with CSS. We named the DNRs according to the gene closest to the region and checked the corresponding imputed ChromHMM states of the Roadmap Epigenomics Project for CD14⁺ monocytes. As expected from the genomic context, the *AFAP1*-DNR is considered an active TSS. The DNR close to *IGLV3-10* is predicted to be DNase sensitive, whereas ChIP-seq data from ENCODE suggest a weak enhancer for the region in *USP7*. Also, the Gencode annotation (v19) predicted a splice-variant starting about 6 kb downstream of this differential locus. Due to inher-

ent difficulties in primer design for NOME-seq sequences, only the *USP7*-DNR could be validated with targeted deep bisulfite amplicon sequencing (online suppl. Fig. 4d). The analysis of endogenous DNA methylation at the 3 identified regions did not reveal differential methylation (data not shown).

Differential Gene Expression in Monocytes of Individuals with CSS

Nucleosome occupancy and DNA methylation at the TSS are generally connected to regulation of gene expression [Kwon et al., 1994]. Next, we investigated a possible

Fig. 3. Minor effect of mutations in *ARID1B* on gene expression in CD14⁺ monocytes. **a, b** Logarithmic fold change from *DESeq2* plotted against the mean expression of genes after mapping with *STAR* (**a**) and quasi-mapping with *Salmon* (**b**). The red dots represent fold changes of significantly deregulated genes in monocytes of individuals with CSS. **c, d** Principal component analysis of the gene expression in the mapping (**c**) and quasi-mapping (**d**) approach. No clear clustering for principal components was observed.

regulatory effect of the DNRs on gene expression in the patient-derived monocytes.

Analysis of the expression pattern of *ARID1B* in our patient cohort revealed similar expression levels in samples C002 and C003, but we observed reduced expression in C004 and C005 (online suppl. Table 2). Closer analysis of the RNA-seq data on a single nucleotide level detected that the mutated allele was expressed in C002 and C003, whereas this was not the case in C004 and C005. The *ARID1B* mutations of C002 and C003 are located in the last exon; the mutated transcripts therefore probably circumvent nonsense-mediated decay.

For expression analysis by RNA-seq, we initially sequenced 4 controls and 4 individuals with CSS. To achieve a comparable analysis to the NOME-seq data, we again constrained our analysis to the male individuals with *ARID1B* mutations. For sequencing analysis, we applied 2 different approaches. First, we mapped the reads to the reference genome and counted the reads per gene (online suppl. Table 3). In a second approach, we used a more recent method, namely quasi-mapping with *Salmon* (online suppl. Table 4). Differential gene expression was determined with *DESeq2* in both approaches and revealed 36 significantly differentially expressed genes in the mapping approach and 50 genes in the quasi-mapping ap-

Table 3. Consistently deregulated genes between male individuals with *ARID1B* mutations and male controls

Gene ID	Direction	Mean IFC
<i>CRYZ</i>	Upregulation	1.64
<i>TRGV5P</i>	Upregulation	1.29
<i>TRGV5</i>	Upregulation	1.15
<i>TSPAN33</i>	Upregulation	1.13
<i>TPPP3</i>	Upregulation	1.10
<i>SAMD9L</i>	Upregulation	0.80
<i>DDX60</i>	Upregulation	0.80
<i>FMN1</i>	Downregulation	-0.98
<i>PER1</i>	Downregulation	-1.21
<i>MIR3648</i>	Downregulation	-1.54
<i>GSTM1</i>	Downregulation	-3.21

IFC, logarithmic fold change.

Patients C002 and C003 were compared with controls K004, K012, and K017.

proach (Fig. 3a, b). However, principal component analysis plots of gene expression based on mapping (Fig. 3c) and quasi-mapping (Fig. 3d) showed no clear clustering of the samples into the expected groups. Therefore, we only considered genes differentially regulated if they showed up in both approaches (Table 3, $n = 11$ genes; online suppl. Table 5). *GeneTrail2* and *DAVID* pathway analysis did not reveal any common regulatory pathway. To assess any regulatory influence of the DNRs on gene expression, we compared the genomic regions found in the NOME-seq analysis with the differentially expressed genes but found no overlap.

In summary, our data showed no major changes in the nucleosome landscape in blood cells of individuals with CSS, but we identified a minor effect on localised nucleosome occupancy and gene expression in patient-derived monocytes.

Discussion

In this study, we assessed whether mutations in subunits of the SWI/SNF complex in individuals with CSS and heterozygous *ARID1B* mutations lead to disease-relevant alterations in their nucleosome profile and gene expression (Fig. 1a). To our knowledge, this is the first study to investigate the effects of human SWI/SNF complex mutations on the nucleosome landscape in patient-derived material. We examined the effect of mutations in *ARID1B* in individuals with CSS and age- and sex-

matched control individuals on the genome-wide nucleosome landscape in cfDNA and patient-derived CD14⁺ monocytes (Fig. 1b). We found no difference in the overall nucleosome positioning around the TSS in cfDNA (Fig. 2a) and only a few regions with significantly different chromatin accessibility (DNRs) in CD14⁺ monocytes from CSS-affected individuals and controls (Table 2; online suppl. Fig. 4). We also discovered a few differentially expressed genes (Table 3), but they did not overlap with the DNRs. None of the differentially expressed genes showed an obvious link to CSS.

Study Limitations

Our discovery of a limited number of DNRs and differentially expressed genes could be due to different reasons as discussed below. First, it could be due to the small patient cohort, which was limited mainly for 2 reasons: (1) CSS is a very rare disease, and (2) the monocytes had to be isolated shortly after blood draw making it impossible to include CSS-affected individuals from other far-away centres. Second, the detection of minor effects on the nucleosome and transcriptional landscape in our study could arise from the analysis of a non-disease relevant tissue. We chose to analyse cfDNA from whole blood and CD14⁺ monocytes for several reasons: (1) *ARID1B* is ubiquitously expressed, (2) blood cells are easily accessible, and (3) the disease-relevant tissue, i.e. neurons, cannot be obtained. We refrained from using patient-specific, induced pluripotent stem cell-derived cell types (e.g., neurons) because of the known effects of reprogramming on the epigenomic landscape.

The influence of impaired SWI/SNF function on cell type-specific nucleosome occupancy at the TSS has been described: deficiency of 2 SWI/SNF complex subunits was shown to reduce nucleosome occupancy around TSS in primary mouse cells at both CGI- and non-CGI-TSS and alter the phasing of nucleosomes downstream of the +1 position [Tolstorukov et al., 2013]. In this study, we used a medium coverage approach of heterogeneous cfDNA sequencing to investigate whether there are any major changes in the nucleosome landscape of whole blood in individuals with CSS (Fig. 2a; online suppl. Fig. 3). Our limited coverage might hide potential DNRs below the detection limit and/or potential DNRs were not found due to the heterogeneity of our cell population. To limit the possible influence of this heterogeneity, we isolated monocytes as they were also under investigation in the German Epigenome Program (DEEP) consortium. In monocytes, we investigated the nucleosome landscape around the TSS, the locus-specific nucleosome occupan-

cy, as well as the gene expression. We did not observe differences in the nucleosome landscape around the CGI-containing TSS (Fig. 2b). Possible explanations are discussed below.

Possible Functional Redundancy of SWI/SNF Subunits

The number of DNRs and differentially expressed genes could be limited by compensatory mechanisms, i.e. subunits of the SWI/SNF complex being encoded by mutually exclusive versions of each other (e.g., *ARID1A* and *ARID1B*, *SMARCA2* and *SMARCA4*). Although these subunits are present in different complexes, there are reports of a certain degree of functional redundancy, which could lead to a reduced effect of mutations in specific subunits [Raab et al., 2015; Kelso et al., 2017]. Our data could suggest that, at least in blood, there seems to be a redundant function of the ARID1 subunits, as we did not observe changes in the genome-wide nucleosome positioning around the TSS in cfDNA or in monocytes of individuals with *ARID1B* variants. The few detected regions might be especially sensitive to impaired function of one subunit resulting in differential nucleosome occupancy in monocytes (Table 2; online suppl. Fig. 4). In support of functional redundancy, in vitro knockdown of *ARID1B* in HCT116 wild-type cells also revealed only a limited influence on chromatin accessibility in a recent study [Kelso et al., 2017], emphasising interplay of ARID1A- and ARID1B-containing SWI/SNF complexes. In the human liver cancer cell line HepG2, Raab et al. [2015] showed that the SWI/SNF subunits ARID1B and ARID2 regulate a largely overlapping gene network. This overlapping regulation may explain functional redundancy in absence of differential expression of SWI/SNF subunits as in our study (data not shown). A knockout mouse model of *Brm* (homologue of *SMARCA2*) was described to be viable possibly due to an upregulation of its counterpart *Brg1* (homologue of *SMARCA4*) [Reyes et al., 1998]. In cardiovascular tissue, viability of *Brg1*-deficient cells in vivo was dependent on *Brm*, which implicates a functional compensation of these 2 subunits [Willis et al., 2012]. Therefore, functional redundancy of SWI/SNF subunits might reduce effects on the nucleosome landscape to specific sensitive regions.

Tissue Specificity of the SWI/SNF Complex

There is a well-known tissue specificity in subunit composition of the SWI/SNF complex, which may alter the effects of mutations in subunits in a cell type-dependent manner. SWI/SNF complex subunits are especially important in neurons and the formation of dendrites [Parrish et al., 2006]. It has been shown that a switch in

subunit composition is mandatory for the differentiation of neural progenitor cells into mature neurons [Lessard et al., 2007]. The subunit necessary for the neuronal switch, ACTL6B, has also been reported to be critical for long-term potentiation, implicating an important role for nucleosome remodelling in synaptic plasticity, learning, and memory [Vogel-Ciernia et al., 2013]. Taken together, mutations in the SWI/SNF complex subunits possibly play a more pronounced role in neurons of the affected individuals due to tissue-specific differences in complex composition.

Alternative Functions of the SWI/SNF Complex

Alternative functions of the SWI/SNF complex apart from nucleosome positioning in gene regulation have been suggested. It was shown that the SWI/SNF complex plays a role in the eviction of the opposing polycomb repressive complexes [Stanton et al., 2017]. In this study, complexes with a mutated ATPase subunit were not able to evict the polycomb repressive complex at specific targets, and the authors suggested a possible role of this mechanism in the aetiology of neurodevelopmental disorders [Stanton et al., 2017]. It is also possible that disease-relevant transcriptional changes occur independent of active nucleosome remodelling, e.g. through alternative splicing [Batsché et al., 2006] or changes in the histone modification landscape. These changes might still underlie tissue-specific differences. A recent study found a defining DNA methylation signature in blood cells of individuals with CSS, further supporting the idea of non-chromatin-remodelling functions of the SWI/SNF complex [Aref-Eshghi et al., 2018]. We did not identify major DNA methylation changes in monocytes of individuals with CSS (data not shown). In addition, neither our DNRs nor the differentially expressed genes showed any overlap with the differentially methylated regions identified by Aref-Eshghi et al. [2018].

It is well-established that chromatin accessibility is part of the regulation of gene expression [reviewed in Cairns, 2009], and differences in chromatin accessibility may result in gene deregulation. Lopez-Serra et al. [2014] showed that chromatin remodelling by the yeast RSC complex (related to the human SWI/SNF complex) is necessary to recruit cohesion loader and to regulate target gene expression. They proposed that defects in nucleosome positioning might have an effect on disease-relevant gene expression. In our data, we found a very mild effect on gene expression (Fig. 3; Table 3). In line with our finding, Tolstorukov et al. [2013] observed that the average transcription rate of most genes remained steady with

altered nucleosome occupancy. Therefore, the SWI/SNF complex seems to regulate gene expression by mechanisms other than active chromatin remodelling.

In summary, this is the first study on possible effects of heterozygous mutations in the human SWI/SNF complex on the nucleosome landscape and transcriptome in patient-derived material. We demonstrated minor, localised effects in CD14⁺ monocytes. At this point, we cannot exclude a significant tissue-specific effect on the nucleosome landscape in disease-relevant cells, i.e. neural progenitor cells and/or neurons. Our findings might point to an alternate, e.g. tissue-specific transcriptional pathogenic mechanism leading to SWI/SNF-related diseases.

Acknowledgements

We are sincerely grateful to the individuals and their families for participating in this study.

Statement of Ethics

This study was approved by the Ethics Committee of the University Hospital Essen (Ethikkommission der Medizinischen Fakultät der Universität Duisburg-Essen, proposal 15-6641-BO). Written informed consent was obtained from the individuals/families of all affected and unaffected individuals for participation in this study.

References

- Anders S, Pyl PT, Huber W: HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169 (2015).
- Aref-Eshghi E, Bend EG, Hood RL, Schenkel LC, Carere DA, et al: BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. *Nat Commun* 9:4885 (2018).
- Arrigoni L, Richter AS, Betancourt E, Bruder K, Diehl S, et al: Standardizing chromatin research: a simple and universal method for ChIP-seq. *Nucleic Acids Res* 44:e67 (2016).
- Backx L, Seuntjens E, Devriendt K, Vermeesch J, Van Esch H: A balanced translocation t(6;14)(q25.3;q13.2) leading to reciprocal fusion transcripts in a patient with intellectual disability and agenesis of corpus callosum. *Cytogenet Genome Res* 132:135–143 (2011).
- Batsché E, Yaniv M, Muchardt C: The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol* 13:22–29 (2006).
- Bork P, Koonin EV: An expanding family of helicases within the “DEAD/H” superfamily. *Nucleic Acids Res* 21:751–752 (1993).
- Cairns BR: The logic of chromatin architecture and remodelling at promoters. *Nature* 461:193–198 (2009).
- Carlson M, Osmond BC, Botstein D: Mutants of yeast defective in sucrose utilization. *Genetics* 98:25–40 (1981).
- Chi TH, Wan M, Zhao K, Taniuchi I, Chen L, et al: Reciprocal regulation of CD4/CD8 expression by SWI/SNF-like BAF complexes. *Nature* 418:195–199 (2002).
- Coffin GS, Siris E: Mental retardation with absent fifth fingernail and terminal phalanx. *Am J Dis Child* 119:433–439 (1970).
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al: STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21 (2013).
- Eisenberg E, Levanon EY: Human housekeeping genes, revisited. *Trends Genet* 29:569–574 (2013).
- Ernst J, Kellis M: Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat Biotechnol* 33:364–376 (2015).
- Fleck BJ, Pandya A, Vanner L, Kerkering K, Bodurtha J: Coffin-Siris syndrome: review and presentation of new cases from a questionnaire study. *Am J Med Genet* 99:1–7 (2001).
- Gebuhr TC, Kovalev GI, Bultman S, Godfrey V, Su L, Magnuson T: The role of *Brg1*, a catalytic subunit of mammalian chromatin-remodeling complexes, in T cell development. *J Exp Med* 198:1937–1949 (2003).
- Haber JE, Garvik B: A new gene affecting the efficiency of mating-type interconversions in homothallic strains of *Saccharomyces cerevisiae*. *Genetics* 87:33–50 (1977).
- Heitzer E, Ulz P, Belic J, Gutsch S, Quehenberger F, et al: Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med* 5:30 (2013).
- Ho L, Jothi R, Ronan JL, Cui K, Zhao K, Crabtree GR: An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci USA* 106:5187–5191 (2009).

Disclosure Statement

P.U. is a technical consultant to Freenome Inc. (San Francisco, USA). All other authors declare that they have no conflicts of interest.

Funding Sources

This work was supported by the German Federal Ministry of Research and Education (01GM1520B to N.C.B.; 01GM1520E to D.W.) as part of the Chromatin-Net Consortium and the DEEP consortia. K.N. was supported by the German Federal Ministry of Research and Education grant for de.NBI (031L0101D).

Author Contributions

N.C.B. and B.H. designed the project. N.C.B. and D.W. acquired and interpreted patient data. A.K. performed monocyte isolations, NOME-seq, and RNA-seq experiments. P.U., E.H., and M.R.S. performed cfDNA experiments and data analysis. A.K., C.S., L.K.-H., and K.N. performed other genomic data analyses. A.K., N.C.B., B.H., P.U., S.R., and D.W. wrote the manuscript. All authors read and approved the manuscript.

Data Availability

All genomic and transcriptomic data will be accessible under ENA Accession No. PRJEB32258.

- Hoyer J, Ekici AB, Ende S, Popp B, Zweier C, et al: Haploinsufficiency of *ARID1B*, a member of the SWI/SNF-A chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am J Hum Genet* 90:565–572 (2012).
- Huang DW, Sherman BT, Lempicki RA: Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37:1–13 (2009a).
- Huang DW, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57 (2009b).
- Imbalzano AN, Kwon H, Green MR, Kingston RE: Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* 370:481–485 (1994).
- Inoue H, Furukawa T, Giannakopoulos S, Zhou S, King DS, Tanese N: Largest subunits of the human SWI/SNF chromatin-remodeling complex promote transcriptional activation by steroid hormone receptors. *J Biol Chem* 277:41674–41685 (2002).
- Inoue H, Giannakopoulos S, Parkhurst CN, Matsumura T, Kono EA, et al: Target genes of the largest human SWI/SNF complex subunit control cell growth. *Biochem J* 434:83–92 (2011).
- Jühling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S: metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res* 26:256–262 (2016).
- Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA: Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res* 22:2497–2506 (2012).
- Kelso TWR, Porter DK, Amaral ML, Shokhirev MN, Benner C, Hargreaves DC: Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in *ARID1A*-mutant cancers. *Elife* 6:e30506 (2017).
- Koh W, Pan W, Gawad C, Fan HC, Kerchner GA, et al: Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proc Natl Acad Sci USA* 111:7361–7366 (2014).
- Kwon H, Imbalzano AN, Khavari PA, Kingston RE, Green MR: Nucleosome disruption and enhancement of activator binding by a human SWI/SNF complex. *Nature* 370:477–481 (1994).
- Leitão E, Beygo J, Zeschnigk M, Klein-Hitpass L, Bargull M, et al: Locus-specific DNA methylation analysis by targeted deep bisulfite sequencing. *Methods Mol Biol* 1767:351–366 (2018).
- Lessard J, Wu JI, Ranish JA, Wan M, Winslow MM, et al: An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* 55:201–215 (2007).
- Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760 (2009).
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079 (2009).
- Lopez-Serra L, Kelly G, Patel H, Stewart A, Uhlmann F: The Scc2–Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. *Nat Genet* 46:1147–1151 (2014).
- Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550 (2014).
- Mashtalir N, D'Avino AR, Michel BC, Luo J, Pan J, et al: Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell* 175:1272–1288.e20 (2018).
- Olave I, Wang W, Xue Y, Kuo A, Crabtree GR: Identification of a polymorphic, neuron-specific chromatin remodeling complex. *Genes Dev* 16:2509–2517 (2002).
- Parrish J, Kim M, Jan LY, Jan YN: Genome-wide analyses identify transcription factors required for proper morphogenesis of *Drosophila* sensory neuron dendrites. *Genes Dev* 20:820–835 (2006).
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C: Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. *bioRxiv*, DOI: <https://doi.org/10.1101/021592> (2015).
- Pedersen BS, Eyring K, De S, Yang IV, Schwartz DA: Fast and accurate alignment of long bisulfite-seq reads. <http://arxiv.org/abs/1401.1129> (2014).
- Raab JR, Resnick S, Magnuson T: Genome-wide transcriptional regulation mediated by biochemically distinct SWI/SNF complexes. *PLoS Genet* 11:e1005748 (2015).
- Reyes JC, Barra J, Muchardt C, Camus A, Babinet C, Yaniv M: Altered control of cellular proliferation in the absence of mammalian brahma (SNF2α). *EMBO J* 17:6979–6991 (1998).
- Roadmap Epigenomics Consortium: Integrative analysis of 111 reference human epigenomes. *Nature* 518:317–330 (2015).
- Santen GWE, Aten E, Sun Y, Almomani R, Gilissen C, et al: Mutations in SWI/SNF chromatin remodeling complex gene *ARID1B* cause Coffin-Siris syndrome. *Nat Genet* 44:379–380 (2012).
- Santen GWE, Aten E, Vulto-van Silfhout AT, Pottinger C, van Bon BWM, et al: Coffin-Siris syndrome and the BAF complex: genotype-phenotype study in 63 patients. *Hum Mutat* 34:1519–1528 (2013).
- Stanton BZ, Hodges C, Calarco JP, Braun SMG, Ku WL, et al: Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. *Nat Genet* 49:282–288 (2017).
- Stöckel D, Kehl T, Trampert P, Schneider L, Backes C, et al: Multi-omics enrichment analysis using the GeneTrail2 web service. *Bioinformatics* 32:1502–1508 (2016).
- Tolstorukov MY, Sansam CG, Lu P, Koellhoffer EC, Helming KC, et al: Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *Proc Natl Acad Sci USA* 110:10165–10170 (2013).
- Tsurusaki Y, Okamoto N, Ohashi H, Kosho T, Imai Y, et al: Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet* 44:376–378 (2012).
- Tsurusaki Y, Okamoto N, Ohashi H, Mizuno S, Matsumoto N, et al: Coffin-Siris syndrome is a SWI/SNF complex disorder. *Clin Genet* 85:548–554 (2014).
- Ulz P, Thallinger GG, Auer M, Graf R, Kashofer K, et al: Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat Genet* 48:1273–1278 (2016).
- Vogel-Ciernia A, Matheos DP, Barrett RM, Kramár EA, Azzawi S, et al: The neuron-specific chromatin regulatory subunit BAF53b is necessary for synaptic plasticity and memory. *Nat Neurosci* 16:552–561 (2013).
- Wallner S, Schröder C, Leitão E, Berulava T, Haak C, et al: Epigenetic dynamics of monocyteto-macrophage differentiation. *Epigenetics Chromatin* 9:33 (2016).
- Wieczorek D, Bögershausen N, Beleggia F, Steiner-Haldenstatt S, Pohl E, et al: A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. *Hum Mol Genet* 22:5121–5135 (2013).
- Willis MS, Homeister JW, Rosson GB, Annayev Y, Holley D, et al: Functional redundancy of SWI/SNF catalytic subunits in maintaining vascular endothelial cells in the adult heart. *Circ Res* 111:e111–e122 (2012).
- Wu JI, Lessard J, Olave IA, Qiu Z, Ghosh A, et al: Regulation of dendritic development by neuron-specific chromatin remodeling complexes. *Neuron* 56:94–108 (2007).
- Yuan G, Liu Y, Dion MF, Slack MD, Wu LF, et al: Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309:626–631 (2005).
- Zerbino DR, Johnson N, Juetteman T, Sheppard D, Wilder SP, et al: Ensembl regulation resources. *Database* 2016:bav119 (2016).