MONASH University

Modelling Epigenetic Dysregulation in Rubinstein-Taybi Syndrome

Carmela Maria de Boer

Supervised by
Assoc. Prof. Jose M Polo
Dr. Stefan White
Dr. Patrick Western

A thesis submitted for the degree of Doctor of Philosophy
at
Monash University
2017

Table of Contents

Copyright Notice

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature:

Print Name: Carmela Maria de Boer

Date: 4th of April 2017

Acknowledgements

First of all I would like to thank my primary supervisors Associate Professor Jose Polo and Dr Stefan White. I started my PhD with Stefan at the Hudson Institute of Medical Research (formerly MIMR) after working as a research assistant for several years. With him I learned a lot about both genetics and epigenetics in combination with human diseases. I would like to thank him for getting me on track and making time for our weekly Skype meetings at often erratic hours (for him). Halfway through my PhD, I joined the laboratory of Jose Polo and I would like to thank him for taking me on and providing me with an easy transition. I would like to thank my co-supervisor Patrick Western for all the feedback during my yearly milestone reviews.

A massive thanks to the Polo lab, for welcoming me when I joined the group. A special thanks to Dr Fernando Rossello, for helping me with all the NGS data and Dr Christian Nefzger and Joseph, for helping me set up all the iPSC work. A very warm thanks to Amber and the other lab members, for making me countless coffees with art and have helped me to get through my PhD (and life in general). You turned my last couple of months in the lab into a little party.

I would like to say thanks to people in the Center of Genetic Diseases at the Hudson Institute of Medical Research. Especially Lexie and Kirsty, who have been great during my whole PhD and I see as friends for life. I would also like to give a special thanks to Amanda who has been a great role model and made my time in the White lab very enjoyable. Thanks to my office and lab neighbour Will, who patiently listened to all my banter while trying to work and finish his own PhD. Last but not least, thank you Sanna, for our weekly phone chats. You have been a great life mentor and I will always be grateful for the times when I needed it most.

I would like to thank my Australian families. A special thanks to Kate and David Morrow, who have been tremendous. They have been the greatest support and I will be forever grateful for the love and kindness that they have shown before, during and after my PhD. Thanks to my aunt Toos and cousins for being so kind and wonderful.

Thanks to all my Australian friends (both crazy scientists and grounded ones) I made along the way. Special thanks to Ingrid, for being my counsellor and keeping me sane. Thank you for always listening and cracking me up when I needed it the most.

And finally thank you to my family and friends back in The Netherlands. Thanks to my parents, for the weekly calls and not always asking how my PhD was going or

when I would come home. You have been a great support, despite being so far away. A big thanks to my sister Maaike and brothers Ruud and Loek, who again have been a great silent support and I am very fortunate to have these amazing siblings in my life. Last but not least, thanks to all my Dutch friends, especially Bernadette, Annemiek and Annet.

This experience has been a massive roller coaster and I am so glad I made it. I cannot wait to finally get off and I am confident that I will look back with fond memories in the near future.

Abbreviation List

Abbreviation	Definition
ALL	Acute Lymphoblastic Leukaemia
AML	Acute Monocytic Leukaemia
ASD	Autism Spectrum Disorder
BD	Bromo Domain
BMP	Bone Morphogenetic Protein
CBP	CREB-Binding Protein
CdLS	Cornelia de Lange Syndrome
CF	Cystic Fibrosis
CH	Cysteine-Histidine
ChIP-Seq	Chromatin Immunoprecipitation Sequencing
CMA	Chromosomal Microarray Analysis
CNV	Copy Number Variation
CREBBP	CREB Binding Protein
CTCL	Cutaneous T-cell Lymphoma
DAPI	4',6-diamidino-2-fenylindool
DLBCL	Diffuse Large B-Cell Lymphoma
DMD	Duchenne Muscular Dystrophy
DMSO	Dimethylsulfoxide
DNMT 3B	DNA Methyltransferase 3B
dNTP	Deoxynucleotide
ESC	Embryonic Stem Cell
FHS	Floating Harbour Syndrome
FISH	Fluorescent *In Situ* Hybridisation
GE	Genome Editing
GFP	Green Fluorescent Protein
HAT	Histone Acetyltransferase
HDAC	Histone deacetylase
HDFa	Human Dermal Fibroblast adult
HDFn	Human Dermal Fibroblast neonatal
H&E	Haematoxylin and Eosin
HEK	Human Embryonic Kidney
HSC	Hematopoietic Stem Cell
ICF1	Immunodeficiency Centromeric instability and Facial anomalies type I
ID	Intellectual Disability
IGV	Integrative Genomics Viewer
iPSC	induced Pluripotent Stem Cell
IQ	Intelligence Quotient
KAT	Lysine Acetyltransferase
MEF	Mouse Embryonic Fibroblast
MLL	Myeloid/Lymphoid Leukaemia
MLPA	Multiplex Ligation-dependent Probe Amplification
MOI	Multiplicity of Infection
MOZ	Monocytic Leukaemia Zinc Finger Protein
MRI	Magnetic Resonance Imaging
MT	Mitochondria

NGS	Next Generation Sequencing
NPC	Neuronal Progenitor Cell
NR	Nuclear Receptor
NSC	Neuronal Stem Cell
OSKM	OCT3/4, SOX2, KLF4 and MYC
PSC	Pluripotent Stem Cell
PBD-ZSD	Peroxisome Biogenesis Disorder - Zellweger Spectrum Disorder
PBMC	Peripheral Blood Mononuclear Cell
RTS	Rubinstein-Taybi Syndrome
SeV	Sendai Virus
shRNA	Short Hairpin RNA
SID	Steroid receptor coactivator-1 Domain
SMA	Spinal Muscular Atrophy
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SRCAP	SNF2-Related CREBBP Activated Protein
SV	Structural Variants
TALEN	Transcription Activator-Like Effector Nuclease
TS	Timothy Syndrome
VEP	Variant Effect Predictor
VPA	Valporic Acid
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
ZNF	Zinc-Finger Nuclease

Figure List

Chapter 4 – Aim 3 Unravelling the genetic background of Rubinstein-Taybi Syndrome using Next Generation Sequencing

Chapter 5 - General discussion and future directions of human genetic and (neuro)developmental disease research

Abstract

Rubinstein-Taybi Syndrome (RTS) is a rare congenital disease, caused by heterozygous mutations in either *CREBBP* or *EP300*, which code for CBP and p300 respectively. These proteins have multiple important overlapping roles in (epigenetic) gene expression control, development and neurogenesis as co-activators and acetyltransferases. Most studies of these two proteins have been performed with animal models. Several RTS mouse models have been generated in which either *CREBBP* or *EP300* are knocked down, and these mice show both physical and neurological similarities to RTS. However, human studies researching both proteins and RTS are limited, and this restricts the improvements in therapeutic developments for RTS.

Fibroblasts from RTS patients with known mutations in *CREBBP* were reprogrammed into induced pluripotent stem cells (iPSCs) and differentiated towards the neuronal lineage. Even though differences in reprogramming potential between RTS and WT cell lines were not detected, fewer mature neurons were generated from RTS iPSCs, which could explain the underlying neurological symptoms seen in these patients. Addition of an HDAC inhibitor during differentiation appeared to rescue this phenotype. These results are promising and imply HDAC inhibitors can be applied in RTS and other neurodevelopmental diseases as a potential therapeutic.

More research is needed in the genetic background of RTS, as only ~65% of clinically diagnosed RTS cases show a variant in either *CREBBP* or *EP300*, meaning 35% of cases lack a genetic diagnosis. It is very important for both patient and family to know the underlying cause of the disease, which could also lead to therapeutic possibilities. Whole exome sequencing (WES) was performed on six RTS patients to either detect variants in the two known RTS genes and/or to find additional candidate genes for this syndrome. An in house filtering strategy was designed for the generated sequencing data and detected a variant in *CREBBP*, which results in a premature stop codon. However, no other potential variants that could lead to a novel RTS causing gene were discovered in the other cases. This type of analysis could be improved by expansion of the cohort but also by performing whole genome sequencing (WGS). This approach includes screening of variants in noncoding DNA and regulatory elements, which could be involved in the onset of RTS.

Together, this study might lead to a better understanding of RTS and other developmental disorders, and will hopefully improve the quality of life for these patients and their families.

CHAPTER 1 – Introduction

CREBBP/EP300 and Rubinstein-Taybi Syndrome

Rubinstein-Taybi Syndrome (RTS) is a congenital condition, affecting ~1 in 100,000-125,000 individuals [1]. Characteristic features include intellectual disability, growth delay, specific facial dysmorphisms (beaked nose, down-slanting palpebral fissures, arched palate), and broad thumbs and big toes. Other clinical issues seen in a percentage of RTS cases include cardiac, eye and kidney malformations, urogenital abnormalities such as hypospadias and cryptorchidism (in males), keloid (scarring) formation, and seizures [2, 3]. As RTS patients enter adulthood there can be significant mood swings, behavioural problems and autistic traits and RTS has also been linked to an increased risk for certain cancers [4].

RTS is an autosomal dominant condition, and is caused in ~65% of cases by heterozygous mutations in either *CREBBP* [5] or *EP300* [6], which code for CREB binding protein (CBP) and p300 respectively. CBP was named as it was initially identified as a binding partner of CREB [7], whereas p300 was shown to bind adenovirus E1A protein and is therefore also known as E1A-associated 300-kD protein [8]. These homologous proteins are ubiquitously expressed [9] and act in chromatin remodelling, are transcriptionally co-activators, have intrinsic lysine acetytransferase activity of both histones (HAT) and non-histone (KAT) proteins, and are known to interact with over 400 proteins [10] (Figure 1). They play roles in multiple signal transduction pathways and participate in basic cellular processes such as DNA repair, cell growth and differentiation, apoptosis and tumour suppression [9].

CREBBP is located on chromosome 16p13, and *EP300* on 22q13 and comparing amino acid sequences throughout the two proteins reveal numerous regions of near identity. They are multidomain proteins and consist of three cysteine-histidine (CH) rich regions, a KIX domain, which is the CREB binding domain, a bromodomain to recognize acetylated lysine residues, an acetyltransferase (HAT/KAT) domain and a steroid receptor coactivator-1 domain (SID) [11, 12]. Although similar for over 70%, there are functional differences between the two proteins and have shown to have unique effects *in vitro* and *in vivo* [13] (Figure 2 and Table 1).
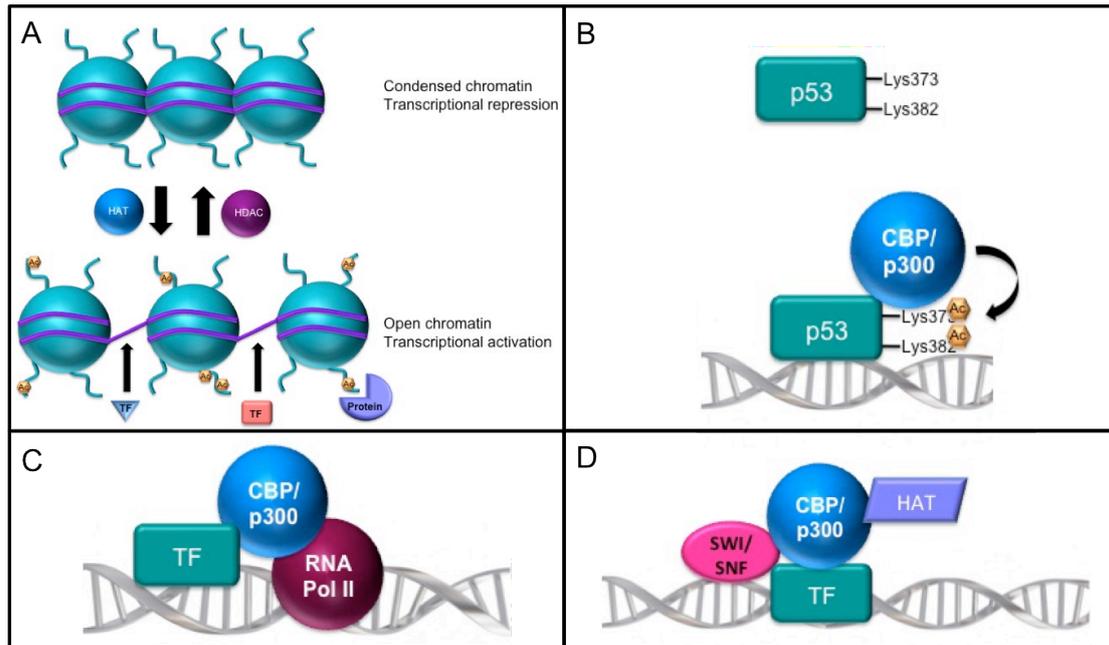
Figure 1. Functions of CBP and p300. (A) CBP and p300 acetylate lysine residues on proteins such as histones, hence are known as histone acetyl transferases (HATs). Addition of an acetyl group onto the histone tails will loosen and open up the chromatin structure, making the DNA more accessible to transcription factors (TF) and other protein interactions, leading to active gene transcription. Histone deacetyl transferases (HDACs), reverse this reaction, and removing the acetyl groups will condense the chromatin, preventing transcription. (B) CBP and p300 also acetylate non-histone proteins, which is why recently they have been referred to as lysine acetyl transferases (KATs). Acetylation of proteins will mostly lead to activation of these proteins. (C&D) CBP and p300 are co-activators but do not specifically bind to DNA. They interact with more than 400 proteins by forming bridges/scaffolds and also interact with different components of the transcription machinery, all leading to control of gene expression.

Figure 2. Homology between *CREBBP* and *EP300*. Adapted from Roelfsema et al., 2007 [13]. CBP and p300 are homologous genes and proteins and are overall 66% identical. CBP consist of 2441 amino acids whereas p300 is made up of 2414 amino acids. This shows a linear schematic representation of CBP and p300 homologous regions and functional domains in percentages. NR, nuclear receptor; CH1-3, cysteine and histidine-rich regions 1-3; KIX, binding site of CREB; BD, bromodomain; HAT/KAT, acetyltransferase domain.

Table 1. Functional Differences between *CREBBP* and *EP300*

Function	Description	Reference
Lymphocyte development	*Crebbp* conditional KO mouse showed increased percentage of CD+ single-positive thymocytes; not observed in Ep300 KO mice.	[14]
Cell cycle	Induction of p21(Cip1) required normal levels of p300 but not Cbp; opposite observed for p27(Kip1).	[15]
Adipocyte development	CBP and p300 each had unique functions in differentiation of mature adipocytes.	[16]
Embryogenesis	Growth retardation and craniofacial abnormalities observed in $Crebbp^{+/-}$ mice; not observed in $Ep300^{+/-}$ mice.	[17, 18]
Embryogenesis	$Ep300^{-/-}$ embryos showed impaired heart development; not observed in $Crebbp^{-/-}$ embryos.	[17, 19]
Cancer	$Crebbp^{+/-}$ mice showed increased incidence of hematological malignancies; not observed in $Ep300^{+/-}$ mice.	[18]
Gene regulation	CBP had a greater involvement in regulating transcription inhibition than P300.	[20]
DNA damage repair	P300 involved in p53-mediated growth arrest following ionizing radiation, CBP was not.	[21]
DNA damage repair	P300 played a role in cellular sensitivity to ionizing radiation; CBP did not.	[22]

Virtually all cases of RTS are *de novo*, although rare instances of transmission have been reported [1, 23, 24]. Different types of mutations have been detected, ranging from single base pair changes to whole gene deletions, with truncating and small rearrangements being the most common [25, 26]. The occurrence of deletions that remove the entire gene suggests that RTS is caused by haploinsufficiency [13]. Even though missense mutations in the HAT domain are associated with the classical RTS phenotype, there is no genotype-phenotype correlation detected.

The mutation distribution is uneven, with *CREBBP* mutations being significantly more frequent than *EP300* mutations (~60% and <10% of clinically diagnosed cases respectively) [26-28]. Even though the small number of reported RTS cases with *EP300* mutations makes direct comparisons difficult, it appears that RTS patients with a mutation in *EP300* are less severely affected. This is particularly seen in the skeletal abnormalities, such as the unaffected thumbs and big toes [6, 27, 29, 30]. This is also reflected in postnatal growth retardation in patients with *EP300* mutations, which may be associated with the frequent pre-eclampsia seen during pregnancy [26]. As a result, skull circumference and brain growth are also more disturbed in these patients compared to those with a *CREBBP* mutation. Remarkably, cognitive function is typically higher in *EP300* mutated patients, but behavioural problems, autism and autistic related traits are present in similar frequencies between *CREBBP* and *EP300* RTS cases [26]. Again, these observations support different roles in development for CBP and p300.

Models of RTS

Several animal models targeting either *Crebbp* or *Ep300* have been generated [31]. The first *Crebbp* mouse model was a null mutant, created by substituting ~230 amino acids in the N terminal of the CBP protein with a targeting vector [17]. Mice homozygous for this mutation died at 8-10 days post conception (dpc). Heterozygous mice showed features consistent with RTS, including growth retardation and skeletal abnormalities. Interestingly, the genetic background of the mice influenced the frequency of the different RTS-like features, suggesting that other genetic variants modulate the severity of the condition. A mouse model producing a truncated CBP (lacking HAT activity) showed a similar phenotype to the null mutant [32]. This supports the theory that loss of HAT activity is responsible for the developmental

defects, although a dominant negative effect is possible. Mice homozygous for this truncated variant also died at 8-10 dpc, with abnormalities of the haematopoietic and vascular systems [32].

Similar results were observed in p300 null mice, where loss of both p300 copies resulted in embryonic lethality, despite the presence of normal quantities of CBP [19]. These embryos showed defects in morphogenesis, and cell differentiation and proliferation. Compared to $Crebbp^{+/-}$ mice, there is increased embryonic lethality for $Ep300^{+/-}$ mice seen. However, $Ep300^{+/-}$ mice that were born showed a less severe phenotype in respect to growth retardation and craniofacial abnormalities when compared to $Crebbp^{+/-}$ mice [19] consistent with RTS in humans. Differences in cognitive function are also mirrored in mouse models, where p300$^{+/-}$ mice show similar but less marked deficits compared to cbp$^{+/-}$ mice [33].

Because of this consistency, the increased embryonic mortality may partly explain why fewer RTS cases are identified with *EP300* mutations compared to *CREBBP*. Interestingly, double heterozygous knockout mice for *Crebbp* and *Ep300* genes were not viable either, showing the same type of malformations as observed in the homozygous knockout mice, which suggests that the combined levels of CBP and p300 are critical during development.

Even though animal studies have given an enormous insight into the molecular background, RTS research using human cells and/or tissues has been limited. This is due to the rarity of the disease and the lack of available disease specific cells and/or tissues. This has been overcome by the discovery and development of (human) induced pluripotent stem cells (iPSCs), creating a disease in a dish model. These cells can be differentiated into disease specific cells, such as neuronal or cardiac cells, and used in combination with drug screens. This will give better insights into RTS and other (neuro)developmental diseases (Figure 3).
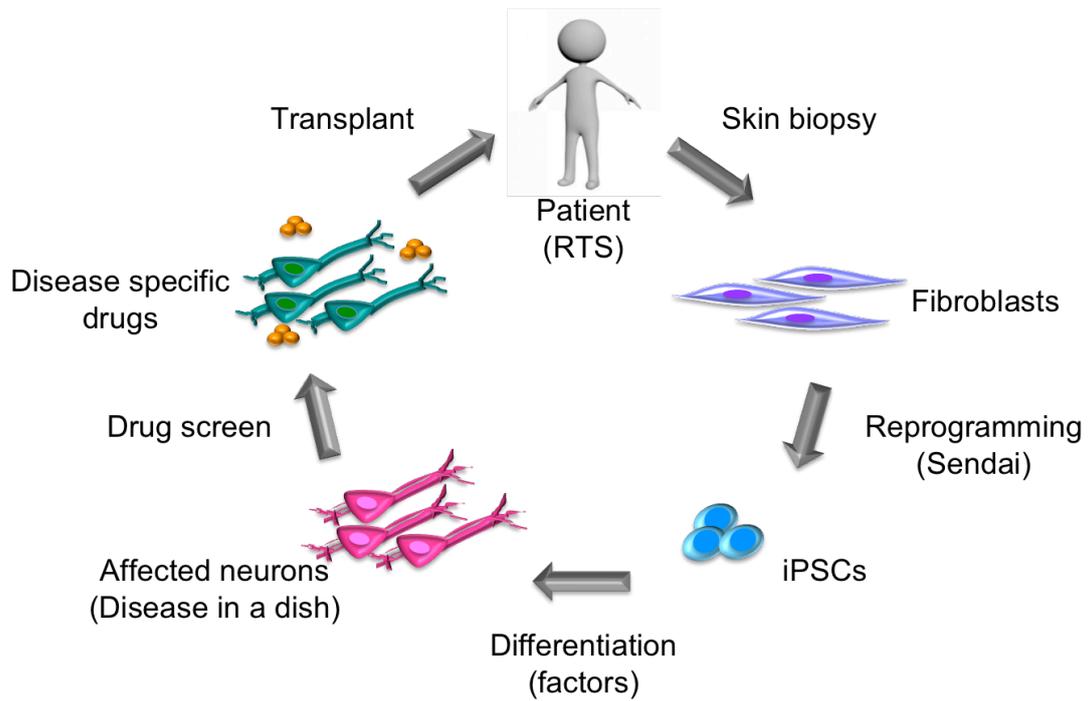
Figure 3. Disease in a dish model using iPSCs. Somatic cells from RTS patients can be reprogrammed into iPSCs. Differentiating these into neurons, will create a disease in a dish model, which can be used in drug screens. This can lead to potential therapeutics for RTS and other (neuro)developmental disorders.

How do *CREBBP* and *EP300* mutations lead to RTS?

Presumably RTS-causing missense mutations in CBP cluster within the HAT domain [13], consistent with loss of HAT activity being sufficient to cause the majority of the clinical features seen in RTS [6]. Acetylation of histones is generally associated with active chromatin, and most studies to date have focused on histone H3, especially H3K9ac and H3K27ac [34]. H3K27ac has been described as a marker of active enhancers [35, 36]. Consistent with acetylation being found at active genomic regions, p300 has been used to identify enhancers in multiple studies [35, 37, 38]. It should be noted however, that not all enhancers are marked by p300 [39].

CBP and p300 can both acetylate several lysine residues in the histone tails (H3: K9, K14, K18, and K27, and H4: K5, K8, K12 and K16 [9, 40-43]). However, there are differences in specificity and requirements for certain histone marks between the two proteins [9, 17, 19] and this could also underlie the slight differences seen between cases with a mutation in *CREBBP* compared to *EP300* [33, 44, 45].

There is conflicting data in the literature regarding the specific changes in histone acetylation as a consequence of reduced CBP and/or p300 levels. Histone acetylation levels were assayed in nine lymphoblastoid cell lines derived from RTS patients [44]. All lines had heterozygous *CREBBP* mutations, ranging in severity from a whole gene deletion to missense and splicing defects. Histones H2A and H2B showed the greatest reduction in acetylation levels across all samples, although a significant reduction in H3 acetylation levels was also observed in the sample with the *CREBBP* deletion. These results contrast with previous findings in a mouse model, where fibroblasts lacking both CBP and p300 have at least 90% reduction in global levels of H3K18ac and H3K27ac [46, 47].

One possibility is that p300 is able to substitute for CBP in acetylating H3, although no significant increase in p300 mRNA levels was observed in the RTS cell lines. In addition, the human study looked at global histone changes, whereas the mouse model was used to examine specific acetylation marks on the histone tails. Which, and how many, of these marks are affected in RTS cells is still to be determined, but will give further insight into the role of those specific modifications in gene regulation.

As mentioned before, CBP and p300 have intrinsic acetyltransferase activity, meaning they have the ability to also acetylate non-histone proteins, such as transcription factors [48]. Acetylation has a variety of effects on proteins, such as increased or

decreased DNA binding affinity [49, 50], protein stability [51] and changes in protein-protein interactions [52]. Although there are more than 400 proteins known to interact with CBP and/or p300, and a subset of these are required to be acetylated, there are no known reports on acetylation levels of non-histone proteins in RTS models.

Histone Deacetylase (HDAC) inhibition

Normal cell function depends on the balance between the opposing activities of HATs and HDACs, and each class of enzyme can be influenced by different compounds [53]. HDAC inhibitors can either be synthesized or isolated as natural products and interfere with the activities of HDACs. This inhibition may restore or increase the level of histone acetylation and can be an approach in treating a number of clinical disorders. There are different classes of HDAC inhibitors, based on structure, which have been used as mood stabilizers and anti-epileptics [54, 55] and are more recently being tested as possible treatments for cancers and inflammatory diseases [56-62].

Targeting of reversible epigenetic components altered in RTS with HDAC inhibitors might be an ideal strategy as a potential therapy. Studies in *Crebbp* knockout mouse models have already demonstrated that HDAC inhibitors have a positive effect on memory function [63, 64]. Wang et al. [65] showed with both *in vitro* and *in vivo* knockdown models of RTS that differentiation of embryonic cortical precursors into all three neural lineages was inhibited, due to decreased CBP binding and histone acetylation at promoters of neuronal and glial genes. Indeed, the addition of a HDAC inhibitor (trichostatin A (TSA)) rescued this phenotype and showed increased neuronal differentiation.

Are their additional genes that cause RTS?

Currently, less than 70% of RTS cases has a genetical diagnosis with a causative mutation in either *CREBBP* or *EP300*. Mutation screening usually focuses on coding sequence, looking for either small sequence variants or deletions/duplications of one or more exons. Therefore, mutations in non-coding DNA and regulatory elements, such as promoters and enhancers, are being missed. However, variants in these regions can disturb the expression of either *CREBBP* or *EP300* sufficiently, resulting in RTS.

As an initial RTS diagnosis is typically based on the clinical presentation, it is possible that cases with milder and/or atypical RTS characteristics may be misdiagnosed. This would especially be relevant for cases with an *EP300* mutation, where the clinical features are often less severe when compared to those seen in RTS cases with *CREBBP* mutations. Indeed, *EP300* variants have been identified in individuals not originally diagnosed with RTS, but instead were suspected of having Cornelia de Lange syndrome (CdLS), a genetically heterogeneous syndrome similar to RTS [66, 67]. A diagnosis for this syndrome is molecularly confirmed for ~65% of cases and genes involved in this disease are *NIPBL* (55%) [68, 69], *SMC1A, SMC3, RAD21* and *HDAC8* (together accounting for <10% of CdLS cases) [70]. As with CBP/p300, many functions of these proteins are linked with chromatin structure [71].

It is not surprising that syndromes showing significant clinical overlap with RTS are caused by mutations in genes encoding CBP and/or p300-interacting proteins. One example is Floating-Harbor Syndrome (FHS), which is characterized by short stature, learning disabilities (especially language), and specific facial features such as a triangular face, and bulbous nose [72]. The causative gene was identified in 2012 and analysis of 13 individuals with FHS found mutations in the *SRCAP* (SNF2-related CREBBP activated protein) gene [73]. *SRCAP* encodes a chromatin-remodeling protein that contains several functional domains and has, like CBP/p300, multiple co-activator roles, with CBP being one of its targets [74, 75]. All mutations identified in this study were truncating (either frame-shift or nonsense), and were clustered in the 3' end of the gene. Following the identification of *SRCAP* mutations in FHS, a cohort of nine patients with FHS diagnostic criteria was analysed. A SRCAP truncating mutation could only be identified in six cases, supporting genetic heterogeneity of FHS [76].

Even though these syndromes share common characteristics with RTS, they are unique conditions with distinctive symptoms. However, it is clear that chromatin regulations play major roles in these diseases and is possible that CBP and p300 are involved in shared pathways with these disease-causing genes, producing overlapping phenotypes.

Causative RTS genes

Diagnosis of RTS is essentially based on clinical presentation, and can be confirmed by genetic screening. However, for the remaining ~35% of clinically diagnosed RTS cases, the cause of the syndrome remains unknown at present. *CREBBP* was originally identified as an RTS-causing gene due to structural aberrations on chromosome 16 [5, 77]. The homology of *CREBBP* made *EP300* an obvious candidate gene, although it took an additional 10 years for the first *EP300* mutations in RTS patients to be identified [6]. Implementing screening methods such as exome and whole genome sequencing has led to the identification of genes responsible for several different syndromes [78, 79] and can be used to find more genes responsible for the onset of RTS.

Mosaicism can also be a possible explanation for the failure of diagnosis, as this has been reported in a range of conditions [80, 81]. The ability to detect mosaic changes depends on the degree of mosaicism and the technique used for genetic analysis. A study of 42 RTS patients identified three mosaic deletions using Fluorescent *In Situ* Hybridisation (FISH) analysis [82]. One of the mosaic cases, presenting an average of 18.5% of mutated cells in two different tissues, had previously been missed with microsatellite analysis, and would also have been difficult to detect with most other PCR-based methodologies e.g. Multiplex Ligation-dependent Probe Amplification (MLPA) [83].

Furthermore, the origin of the DNA sample can be important in mosaicism. This has been clearly demonstrated for RTS, where multiple studies have identified causative sequence variants in DNA derived from buccal cells, which were not detected in DNA from a matching blood sample [84, 85]. This has also been shown for other genetic diseases such as CdLS [86, 87] and Filippi syndrome [85].

There are several sequencing techniques to detect disease-causing variants and recently, Next Generation Sequencing has made major developments, making it more available and affordable to use. Whole exome sequencing (WES) has become important to use in the screening of disease causing mutations to conform or exclude clinical diagnosis and for the discovery of novel disease causing genes, such as for RTS. As the name suggests, this technique sequences the coding DNA (the exome), in contrast to whole genome sequencing (WGS) that sequences the majority of the DNA, including regulatory elements. Mutations in these regions, such as promoters and

enhancers, can have an influence on gene regulation and expression, which can also underlie diseases such as RTS. Therefore, the development of techniques like these can have a tremendous effect on the efficiency and certainty of the genetically diagnosis of RTS and other genetic diseases [105].

What are the roles of *CREBBP/EP300* mutations in cancer?

Germ line mutations

CREBBP and *EP300* have been considered to be tumor suppressor genes [88-91], and there is data supporting a role for *CREBBP/EP300* mutations in cancer. RTS has been associated with certain solid cancers of neural and developmental origin, including medulloblastoma, neuroblastoma and nasopharyngeal rhabdomyosarcoma [4]. It is noteworthy that these associations were made prior to the identification of either *CREBBP* or *EP300* as being causative for RTS. It is therefore possible that the RTS cases in which cancer was observed were a subset of RTS, defined by mutations in a specific gene. There is evidence in mice supporting this hypothesis. $Crebbp^{+/-}$ mice showed defects in B lymphoid development and an increased propensity for developing haematopoietic tumours [18]. In contrast, no such predisposition was observed in $Ep300^{+/-}$ mice. This discrepancy highlights the importance of being able to provide an accurate genetic diagnosis for RTS.

Somatic mutations

Somatic mutations in either *CREBBP* or *EP300* in different cancer types have been reported, but are extremely rare in solid tumours [92, 93]. Somatic translocations, affecting either *CREBBP* or *EP300*, have been described for several haematological malignancies [94] such as acute lymphoblastic leukaemia (ALL) and acute monocytic leukaemia (AML). Both genes can fuse with either *MLL* (Myeloid/Lymphoid Leukaemia) [95, 96] or *MOZ* (Monocytic Leukaemia Zinc Finger Protein) [97, 98], which encode histone-modifiers. These resulting fusion proteins can lead to abnormal regulation of specific transcription factors such as p53, leading to cancer [99].

Somatic sequence mutations in *CREBBP* and *EP300* have also been identified in leukaemia. A study of relapsed ALL showed that 18% (13/71) of cases contained a mutation in the *CREBBP* gene (by comparison, 1/71 contained an *EP300* mutation) [100]. Furthermore, a study in B-cell non-Hodgkin's lymphoma [101] identified a

high frequency of *CREBBP* and *EP300* mutations in diffuse large B-cell (DLBCL) and follicular lymphoma. Somatic *CREBBP/EP300* genomic deletions and/or point mutations were found in about 40% of lymphoma cases, with the majority clustering in exons encoding the HAT domain. This observation led to the analysis of p53 and BCL6 acetylation by CBP and p300, as BCL6 is inactivated by acetylation [102] whereas p53 requires acetylation for its role in apoptosis [103]. CBP and p300 with mutations affecting the HAT domain were indeed deficient in their ability to acetylate p53 and BCL6. It was anticipated that this would lead to a reduced ability to respond to DNA damage, which may be linked to the cancer.

Overall, CBP and p300 play important roles in oncogenesis, as somatic mutations in these genes are frequently detected in lymphoma and leukemia. Despite this link, the degree to which RTS patients are predisposed to cancer is still unclear. This has been discussed for other chromatin modifiers, such as members of the SWI/SNF complex, where the type of mutation plays a role in whether there is an associated susceptibility to cancer [104-106]. The potential role of *CREBBP* and *EP300* in cancer predisposition has important clinical implications. If it is indeed the case that only a subset of RTS cases (due to specific types of variants in a single RTS gene) has an increased risk of developing cancer, then any screening can be targeted to those most likely to benefit.

Conclusion

Although two genes have been identified as being involved in RTS, causative mutations cannot be identified in all cases. Additionally, the mechanisms linking reduced CBP and p300 levels with the clinical features of RTS are still unclear. This project will attempt to unravel the neurological abnormalities seen in RTS patients with a disease in a dish model. Induced pluripotent stem cells will be generated from somatic cells taken from RTS patients and differentiated into neuronal cells. In combination with an HDAC inhibitor during differentiation it will be investigate if this has a positive effect on neuronal differentiation in RTS cells. Whole exome sequencing on a cohort of clinically diagnosed RTS patients will be used to identify mutations in either *CREBBP* or *EP300*, and potentially discover novel RTS-causing genes. Together, this study might lead to a better understanding of RTS and related

developmental disorders, and will hopefully improve the quality of life for these patients and their families.

References

1. Hennekam, R.C., et al., *Rubinstein-Taybi syndrome in a mother and son.* Eur J Pediatr, 1989. **148**(5): p. 439-41.
2. Hennekam, R.C., *Rubinstein-Taybi syndrome.* Eur J Hum Genet, 2006. **14**(9): p. 981-5.
3. Schorry, E.K., et al., *Genotype-phenotype correlations in Rubinstein-Taybi syndrome.* Am J Med Genet A, 2008. **146A**(19): p. 2512-9.
4. Miller, R.W. and J.H. Rubinstein, *Tumors in Rubinstein-Taybi syndrome.* Am J Med Genet, 1995. **56**(1): p. 112-5.
5. Petrij, F., et al., *Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP.* Nature, 1995. **376**(6538): p. 348-51.
6. Roelfsema, J.H., et al., *Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease.* Am J Hum Genet, 2005. **76**(4): p. 572-80.
7. Chrivia, J.C., et al., *Phosphorylated CREB binds specifically to the nuclear protein CBP.* Nature, 1993. **365**(6449): p. 855-9.
8. Eckner, R., et al., *Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor.* Genes Dev, 1994. **8**(8): p. 869-84.
9. Kalkhoven, E., *CBP and p300: HATs for different occasions.* Biochem Pharmacol, 2004. **68**(6): p. 1145-55.
10. Bedford, D.C., et al., *Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases.* Epigenetics, 2010. **5**(1): p. 9-15.
11. Kwok, R.P., et al., *Nuclear protein CBP is a coactivator for the transcription factor CREB.* Nature, 1994. **370**(6486): p. 223-6.
12. Lundblad, J.R., et al., *Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP.* Nature, 1995. **374**(6517): p. 85-8.
13. Roelfsema, J.H. and D.J. Peters, *Rubinstein-Taybi syndrome: clinical and molecular overview.* Expert Rev Mol Med, 2007. **9**(23): p. 1-16.
14. Kasper, L.H., et al., *Conditional knockout mice reveal distinct functions for the global transcriptional coactivators CBP and p300 in T-cell development.* Mol Cell Biol, 2006. **26**(3): p. 789-809.
15. Kawasaki, H., et al., *Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation.* Nature, 1998. **393**(6682): p. 284-9.
16. Takahashi, N., et al., *Overexpression and ribozyme-mediated targeting of transcriptional coactivators CREB-binding protein and p300 revealed their indispensable roles in adipocyte differentiation through the regulation of peroxisome proliferator-activated receptor gamma.* J Biol Chem, 2002. **277**(19): p. 16906-12.
17. Tanaka, Y., et al., *Abnormal skeletal patterning in embryos lacking a single Cbp allele: a partial similarity with Rubinstein-Taybi syndrome.* Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10215-20.
18. Kung, A.L., et al., *Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP.* Genes Dev, 2000. **14**(3): p. 272-7.

19. Yao, T.P., et al., *Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300.* Cell, 1998. **93**(3): p. 361-72.

20. Ramos, Y.F., et al., *Genome-wide assessment of differential roles for p300 and CBP in transcription regulation.* Nucleic Acids Res, 2010. **38**(16): p. 5396-408.

21. Yuan, Z.M., et al., *Role for p300 in stabilization of p53 in the response to DNA damage.* J Biol Chem, 1999. **274**(4): p. 1883-6.

22. Yuan, Z.M., et al., *Function for p300 and not CBP in the apoptotic response to DNA damage.* Oncogene, 1999. **18**(41): p. 5714-7.

23. Marion, R.W., D.M. Garcia, and J.B. Karasik, *Apparent dominant transmission of the Rubinstein-Taybi syndrome.* Am J Med Genet, 1993. **46**(3): p. 284-7.

24. Lopez, M., et al., *First case report of inherited Rubinstein-Taybi syndrome associated with a novel EP300 variant.* BMC Med Genet, 2016. **17**(1): p. 97.

25. Petrij, F., et al., *Diagnostic analysis of the Rubinstein-Taybi syndrome: five cosmids should be used for microdeletion detection and low number of protein truncating mutations.* J Med Genet, 2000. **37**(3): p. 168-76.

26. Fergelot, P., et al., *Phenotype and genotype in 52 patients with Rubinstein-Taybi syndrome caused by EP300 mutations.* Am J Med Genet A, 2016. **170**(12): p. 3069-3082.

27. Bartholdi, D., et al., *Genetic heterogeneity in Rubinstein-Taybi syndrome: delineation of the phenotype of the first patients carrying mutations in EP300.* J Med Genet, 2007. **44**(5): p. 327-33.

28. Negri, G., et al., *Clinical and molecular characterization of Rubinstein-Taybi syndrome patients carrying distinct novel mutations of the EP300 gene.* Clin Genet, 2014.

29. Tsai, A.C., et al., *Exon deletions of the EP300 and CREBBP genes in two children with Rubinstein-Taybi syndrome detected by aCGH.* Eur J Hum Genet, 2011. **19**(1): p. 43-9.

30. Bartsch, O., et al., *Two patients with EP300 mutations and facial dysmorphism different from the classic Rubinstein-Taybi syndrome.* Am J Med Genet A, 2010. **152A**(1): p. 181-4.

31. Josselyn, S.A., *What's right with my mouse model? New insights into the molecular and cellular basis of cognition from mouse models of Rubinstein-Taybi Syndrome.* Learn Mem, 2005. **12**(2): p. 80-3.

32. Oike, Y., et al., *Truncated CBP protein leads to classical Rubinstein-Taybi syndrome phenotypes in mice: implications for a dominant-negative mechanism.* Hum Mol Genet, 1999. **8**(3): p. 387-96.

33. Viosca, J., et al., *Syndromic features and mild cognitive impairment in mice with genetic reduction on p300 activity: Differential contribution of p300 and CBP to Rubinstein-Taybi syndrome etiology.* Neurobiol Dis, 2010. **37**(1): p. 186-94.

34. Liang, G., et al., *Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome.* Proc Natl Acad Sci U S A, 2004. **101**(19): p. 7357-62.

35. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.

36.   Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010.

37.   Visel, A., et al., *ChIP-seq accurately predicts tissue-specific activity of enhancers.* Nature, 2009. **457**(7231): p. 854-8.

38.   Xi, H., et al., *Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome.* PLoS Genet, 2007. **3**(8): p. e136.

39.   Krebs, A.R., et al., *SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers.* Mol Cell, 2011. **44**(3): p. 410-23.

40.   Liu, X., et al., *The structural basis of protein acetylation by the p300/CBP transcriptional coactivator.* Nature, 2008. **451**(7180): p. 846-50.

41.   Luebben, W.R., N. Sharma, and J.K. Nyborg, *Nucleosome eviction and activated transcription require p300 acetylation of histone H3 lysine 14.* Proc Natl Acad Sci U S A, 2010. **107**(45): p. 19254-9.

42.   Schiltz, R.L., et al., *Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates.* J Biol Chem, 1999. **274**(3): p. 1189-92.

43.   Xue, K., et al., *Synchronous behaviors of CBP and acetylations of lysine 18 and lysine 23 on histone H3 during porcine oocyte first meiotic division.* Mol Reprod Dev, 2010. **77**(7): p. 605-14.

44.   Lopez-Atalaya, J.P., et al., *Histone acetylation deficits in lymphoblastoid cell lines from patients with Rubinstein-Taybi syndrome.* J Med Genet, 2012. **49**(1): p. 66-74.

45.   Henry, R.A., Y.M. Kuo, and A.J. Andrews, *Differences in specificity and selectivity between CBP and p300 acetylation of histone H3 and H3/H4.* Biochemistry, 2013. **52**(34): p. 5746-59.

46.   Kasper, L.H., et al., *CBP/p300 double null cells reveal effect of coactivator level and diversity on CREB transactivation.* EMBO J, 2010. **29**(21): p. 3660-72.

47.   Jin, Q., et al., *Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation.* EMBO J, 2011. **30**(2): p. 249-62.

48.   Sterner, D.E. and S.L. Berger, *Acetylation of histones and transcription-related factors.* Microbiol Mol Biol Rev, 2000. **64**(2): p. 435-59.

49.   Gu, W. and R.G. Roeder, *Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain.* Cell, 1997. **90**(4): p. 595-606.

50.   Munshi, N., et al., *Acetylation of HMG I(Y) by CBP turns off IFN beta expression by disrupting the enhanceosome.* Mol Cell, 1998. **2**(4): p. 457-67.

51.   Martinez-Balbas, M.A., et al., *Regulation of E2F1 activity by acetylation.* EMBO J, 2000. **19**(4): p. 662-71.

52.   Zhang, W., et al., *Site-specific acetylation by p300 or CREB binding protein regulates erythroid Kruppel-like factor transcriptional activity via its interaction with the SWI-SNF complex.* Mol Cell Biol, 2001. **21**(7): p. 2413-22.

53.   Wade, P.A., *Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin.* Hum Mol Genet, 2001. **10**(7): p. 693-8.

54. Phiel, C.J., et al., *Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen.* J Biol Chem, 2001. **276**(39): p. 36734-41.

55. Reynolds, M.F., E.C. Sisk, and N.L. Rasgon, *Valproate and neuroendocrine changes in relation to women treated for epilepsy and bipolar disorder: a review.* Curr Med Chem, 2007. **14**(26): p. 2799-812.

56. Reynolds, E.H., *Jackson, Todd, and the concept of "discharge" in epilepsy.* Epilepsia, 2007. **48**(11): p. 2016-22.

57. Mann, B.S., et al., *Vorinostat for treatment of cutaneous manifestations of advanced primary cutaneous T-cell lymphoma.* Clin Cancer Res, 2007. **13**(8): p. 2318-22.

58. Marks, P.A., *Discovery and development of SAHA as an anticancer agent.* Oncogene, 2007. **26**(9): p. 1351-6.

59. Barbarotta, L. and K. Hurley, *Romidepsin for the Treatment of Peripheral T-Cell Lymphoma.* J Adv Pract Oncol, 2015. **6**(1): p. 22-36.

60. Libby, E.N., et al., *Panobinostat: a review of trial results and future prospects in multiple myeloma.* Expert Rev Hematol, 2015. **8**(1): p. 9-18.

61. Damaskos, C., et al., *Histone Deacetylase Inhibitors: A Novel Therapeutic Weapon Against Medullary Thyroid Cancer?* Anticancer Res, 2016. **36**(10): p. 5019-5024.

62. Munster, P.N., et al., *A phase II study of the histone deacetylase inhibitor vorinostat combined with tamoxifen for the treatment of patients with hormone therapy-resistant breast cancer.* Br J Cancer, 2011. **104**(12): p. 1828-35.

63. Korzus, E., M.G. Rosenfeld, and M. Mayford, *CBP histone acetyltransferase activity is a critical component of memory consolidation.* Neuron, 2004. **42**(6): p. 961-72.

64. Alarcon, J.M., et al., *Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration.* Neuron, 2004. **42**(6): p. 947-59.

65. Wang, J., et al., *CBP histone acetyltransferase activity regulates embryonic neural differentiation in the normal and Rubinstein-Taybi syndrome brain.* Dev Cell, 2010. **18**(1): p. 114-25.

66. Solomon, B.D., et al., *Expanding the phenotypic spectrum in EP300-related Rubinstein-Taybi syndrome.* Am J Med Genet A, 2015.

67. Woods, S.A., et al., *Exome sequencing identifies a novel EP300 frame shift mutation in a patient with features that overlap Cornelia de Lange syndrome.* Am J Med Genet A, 2014. **164A**(1): p. 251-8.

68. Musio, A., et al., *X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations.* Nat Genet, 2006. **38**(5): p. 528-30.

69. Rohatgi, S., et al., *Facial diagnosis of mild and variant CdLS: Insights from a dysmorphologist survey.* Am J Med Genet A, 2010. **152A**(7): p. 1641-53.

70. Mannini, L., et al., *Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome.* Hum Mutat, 2013. **34**(12): p. 1589-96.

71. Deardorff, M.A., et al., *HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle.* Nature, 2012. **489**(7415): p. 313-7.

72. Lacombe, D., et al., *Floating-Harbor syndrome: description of a further patient, review of the literature, and suggestion of autosomal dominant inheritance.* Eur J Pediatr, 1995. **154**(8): p. 658-61.

73. Hood, R.L., et al., *Mutations in SRCAP, encoding SNF2-related CREBBP activator protein, cause Floating-Harbor syndrome.* Am J Hum Genet, 2012. **90**(2): p. 308-13.

74. Johnston, H., et al., *Identification of a novel SNF2/SWI2 protein family member, SRCAP, which interacts with CREB-binding protein.* J Biol Chem, 1999. **274**(23): p. 16370-6.

75. Monroy, M.A., et al., *Regulation of cAMP-responsive element-binding protein-mediated transcription by the SNF2/SWI-related protein, SRCAP.* J Biol Chem, 2001. **276**(44): p. 40721-6.

76. Le Goff, C., et al., *Not all floating-harbor syndrome cases are due to mutations in exon 34 of SRCAP.* Hum Mutat, 2013. **34**(1): p. 88-92.

77. Breuning, M.H., et al., *Rubinstein-Taybi syndrome caused by submicroscopic deletions within 16p13.3.* Am J Hum Genet, 1993. **52**(2): p. 249-54.

78. Hoischen, A., et al., *De novo mutations of SETBP1 cause Schinzel-Giedion syndrome.* Nat Genet, 2010. **42**(6): p. 483-5.

79. Santen, G.W., et al., *Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome.* Nat Genet, 2012. **44**(4): p. 379-80.

80. Notini, A.J., J.M. Craig, and S.J. White, *Copy number variation and mosaicism.* Cytogenet Genome Res, 2008. **123**(1-4): p. 270-7.

81. Biesecker, L.G. and N.B. Spinner, *A genomic view of mosaicism and human disease.* Nat Rev Genet, 2013. **14**(5): p. 307-20.

82. Gervasini, C., et al., *High frequency of mosaic CREBBP deletions in Rubinstein-Taybi syndrome patients and mapping of somatic and germ-line breakpoints.* Genomics, 2007. **90**(5): p. 567-73.

83. Aten, E., et al., *Methods to detect CNVs in the human genome.* Cytogenet Genome Res, 2008. **123**(1-4): p. 313-21.

84. Chiang, P.W., et al., *Somatic and germ-line mosaicism in Rubinstein-Taybi syndrome.* Am J Med Genet A, 2009. **149A**(7): p. 1463-7.

85. de Vries, T.I., et al., *Mosaic CREBBP mutation causes overlapping clinical features of Rubinstein-Taybi and Filippi syndromes.* Eur J Hum Genet, 2016. **24**(9): p. 1363-6.

86. Huisman, S.A., et al., *High rate of mosaicism in individuals with Cornelia de Lange syndrome.* J Med Genet, 2013. **50**(5): p. 339-44.

87. Braunholz, D., et al., *Hidden mutations in Cornelia de Lange syndrome limitations of sanger sequencing in molecular diagnostics.* Hum Mutat, 2015. **36**(1): p. 26-9.

88. Chan, H.M. and N.B. La Thangue, *p300/CBP proteins: HATs for transcriptional bridges and scaffolds.* J Cell Sci, 2001. **114**(Pt 13): p. 2363-73.

89. Goodman, R.H. and S. Smolik, *CBP/p300 in cell growth, transformation, and development.* Genes Dev, 2000. **14**(13): p. 1553-77.

90. Giles, R.H., D.J. Peters, and M.H. Breuning, *Conjunction dysfunction: CBP/p300 in human disease.* Trends Genet, 1998. **14**(5): p. 178-83.

91. Wang, F., C.B. Marshall, and M. Ikura, *Transcriptional/epigenetic regulator CBP/p300 in tumorigenesis: structural and functional versatility in target recognition.* Cell Mol Life Sci, 2013. **70**(21): p. 3989-4008.

92. Iyer, N.G., H. Ozdag, and C. Caldas, *p300/CBP and cancer.* Oncogene, 2004. **23**(24): p. 4225-31.

93. Kishimoto, M., et al., *Mutations and deletions of the CBP gene in human lung cancer.* Clin Cancer Res, 2005. **11**(2 Pt 1): p. 512-9.

94.   Yang, X.J., *The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases.* Nucleic Acids Res, 2004. **32**(3): p. 959-76.

95.   Sobulo, O.M., et al., *MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3).* Proc Natl Acad Sci U S A, 1997. **94**(16): p. 8732-7.

96.   Ida, K., et al., *Adenoviral E1A-associated protein p300 is involved in acute myeloid leukemia with t(11;22)(q23;q13).* Blood, 1997. **90**(12): p. 4699-704.

97.   Kitabayashi, I., et al., *Fusion of MOZ and p300 histone acetyltransferases in acute monocytic leukemia with a t(8;22)(p11;q13) chromosome translocation.* Leukemia, 2001. **15**(1): p. 89-94.

98.   Borrow, J., et al., *The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein.* Nat Genet, 1996. **14**(1): p. 33-41.

99.   Troke, P.J., et al., *MOZ fusion proteins in acute myeloid leukaemia.* Biochem Soc Symp, 2006(73): p. 23-39.

100.   Mullighan, C.G., et al., *CREBBP mutations in relapsed acute lymphoblastic leukaemia.* Nature, 2011. **471**(7337): p. 235-9.

101.   Pasqualucci, L., et al., *Inactivating mutations of acetyltransferase genes in B-cell lymphoma.* Nature, 2011. **471**(7337): p. 189-95.

102.   Bereshchenko, O.R., W. Gu, and R. Dalla-Favera, *Acetylation inactivates the transcriptional repressor BCL6.* Nat Genet, 2002. **32**(4): p. 606-13.

103.   Tang, Y., et al., *Acetylation is indispensable for p53 activation.* Cell, 2008. **133**(4): p. 612-26.

104.   Santen, G.W., M. Kriek, and H. van Attikum, *SWI/SNF complex in disorder: SWItching from malignancies to intellectual disability.* Epigenetics, 2012. **7**(11): p. 1219-24.

105.   Sun, Y., et al., *Next-generation diagnostics: gene panel, exome, or whole genome?* Hum Mutat, 2015. **36**(6): p. 648-55.

106.   Avior, Y., I. Sagi, and N. Benvenisty, *Pluripotent stem cells in disease modelling and drug discovery.* Nat Rev Mol Cell Biol, 2016. **17**(3): p. 170-82.

107.   Stadtfeld, M., et al., *Induced pluripotent stem cells generated without viral integration.* Science, 2008. **322**(5903): p. 945-9.

CHAPTER 2 – Aim 1 Reprogramming RTS somatic cells and generating disease-specific induced Pluripotent Stem Cells with reduced levels of CBP

Introduction

The majority of RTS cases are caused by heterozygous mutations in either *CREBBP* or *EP300*. These genes code for CBP and p300, which are homologous and ubiquitous expressed proteins. CBP and p300 have multiple overlapping functions as coactivators with intrinsic lysine acetyltransferase (KAT) activity as they can acetylate both histone and non-histone proteins. They can interact with over 400 proteins and play major roles in gene regulation. Both proteins act in different signal transduction pathways and are involved in the control of cell growth, cellular differentiation, DNA repair, apoptosis and tumour suppression [1] and play an important role in the development of the skeletal and central nervous systems [2].

Most knowledge about CBP and p300 is from research involving mouse models. These studies showed that CBP and p300 are ubiquitously expressed throughout development [3] and that homozygous and double heterozygous knockouts are lethal during embryogenesis [4]. This demonstrates that the level and balance of both proteins are fundamental during development and additionally shows that each protein has unique functions.

Even though these proteins are crucial throughout the stages of development, limited research has been performed looking at their importance in embryonic stem cells (ESCs). Fang et al. [5] characterized the function of p300 and Cbp in mouse ESCs and found that, even though they play important roles in the maintenance of the undifferentiated state of mESCs, they are functionally redundant. This study also showed that Nanog recruits both p300 and Cbp to ESC-specific enhancer regions to promote gene activation. In addition, it was demonstrated that both proteins maintain ESC-specific gene expression by the formation of long-range looping structures. However, a study with hematopoietic stem cells (HSCs) showed that Cbp also regulate self-renewal and differentiation in these cells and is essential for the balance in HSC homeostasis [6].

HAT activity seems to play an important role in maintaining pluripotency states of ESCs [7], and this is conserved in both mouse and human. Deficiency in HAT activity of CBP or p300 during development and adulthood seems to be the most likely cause of RTS symptoms. Besides this, little is known about the specific epigenetic modifications associated to RTS as well as its significance in terms of transcriptional defects and how this affects their differentiation.

Transcription factors in ESCs, such as OCT3/4, SOX2, and NANOG have essential roles in early development and are required for the maintenance of undifferentiated ESCs in culture [8-14]. These factors induce and regulate histone posttranslational changes, such as methylation and acetylation, through interactions with histone modifying proteins and enzymes [15]. These epigenetic mechanisms and modifications control accessibility of the chromatin structure and play key regulatory roles in gene expression [16]. Histone modifications are reversible and undergo dynamic global and local changes during development, cell division and cell activation [17]. Acetylation of the histones leads to a more open chromatin structure, which makes the DNA more accessible to transcription factors. Acetylation is therefore generally associated with active transcription of genes and is highly enriched in ESCs compared with differentiated cells [18-22].

Studies have shown that various families of HATs are active in ESCs to maintain the self-renewal capacity and pluripotency of these cells. KAT8 (also know as MOF or MYST) acetylates H4K16, which is important for ESC identity, and deletion of this HAT abolishes the self-renewal capacity and pluripotency of ESCs [20]. Studies with double knockout show that loss of Cbp and p300 lead to ~90% global loss of H3K18 and H3K27 acetylation in mouse embryonic fibroblasts [23, 24]. However, many genes in these double knockouts only show partial loss of expression and in some cases no difference was detected.

To date the underlying basis for the cell-type specificity of RTS is not fully known. Besides mouse models and a number of drosophila studies [25-29] there have been few reports looking at CBP and p300 in human cells, either in general or during development. To obtain ESCs from RTS embryos requires genetic screening at the blastocyst stage, which makes it, in combination with the rarity of the disease, nearly impossible to collect.

Lopez-Atalaya et al. [30] generated lymphoblastoid cell lines from RTS patients with different types of mutations in *CREBBP,* and looked at acetylation deficits. These lines were produced from RTS patients and healthy individuals by Epstein-Barr virus transformation of peripheral blood mononuclear cells [31]. They found a general reduction in the bulk acetylation of the four core histones, especially in H2A and H2B, which was similar to what was observed in the brain of mice with lost or reduced Cbp activity [32]. This reduction in histone acetylation was particularly seen in the cell line from a patient bearing a CBP null allele, showing severe intellectual

disability. However, variable acetylation levels were observed in cell lines harbouring missense and splicing mutations. Even though there might be a trend or correlation seen between cognitive impairments and epigenetic alterations in RTS, overall it is difficult to link the severity of the mutations and specific symptoms of this syndrome. This was also shown in a study using mouse embryonic fibroblasts from Crebbp/Ep300 cre-deleted double knockouts [33], which were transduced with mutant Crebbp alleles. When comparing these dKO MEFs to WT Cbp, results show a diminished acetylation of H3K18ac, which is a known target of CBP [34]. This was shown with various *CREBBP* mutations known to cause RTS [35], and even though acetylation was affected, it was not blocked completely. However, when introducing a specific known RTS causing mutation, Q1500P, which has been predicted to disrupt a key alpha helix in the HAT domain, they observed only a modest reduction in acetylation. This suggests that, besides mutations in the HAT domain, there are additional effects that impair histone acetylation.

Even though animal models, especially mouse models, have been invaluable tools for modelling human disease, there are differences between the species in several aspects of embryonic development [36]. Therefore, it is preferable to conduct biomedical research in humans, which is mostly limited to *in vitro* systems. Additionally, disease phenotypes are frequently cell-type specific, which might be difficult to isolate, and/or cannot be continuously grown in culture, such as neuronal cells. However, the availability of human pluripotent stem cells (PSCs), such as ESCs, has been an important development in the limitations of animal models for certain diseases. These cells are capable of self-renewal and have the potential to differentiate into almost any cell type. In 2006, further development lead to the technique of driving mouse somatic cells into a pluripotent state through the expression of a defined set of four transcription factors, OCT3/4, SOX2, KLF4 and MYC (OSKM factors), called induced pluripotent stem cells (iPSCs) [37]. The expression of the exogenous OSKM factors is required for iPSC generation, until the cells are committed to the pluripotent state. The generation of iPSCs from human fibroblasts was achieved in 2007 [38], and several different cell types, such as blood [39] and keratinocytes [40], have since been reprogrammed into iPSCs.

iPSCs have been a tremendous addition in the study of human development and disease [41], as they are primary cell lines, have self-renewal capacity, and can become virtually any cell type through differentiation. These properties enable to

study genotype-phenotype relationships in a broad range of human cell types, in combination with additional purposes such as drug screening, development and cell therapy.

The first-generation of iPSCs were induced with the use of retroviral vectors to deliver the reprogramming factors [37]. These vectors integrate into the genome of the host cell, potentially causing disruption of neighbouring genes. Occasionally the cassette may integrate in a gene or regulatory element, which will lead to a mutagenic footprint in the host genome and can lead to disruptions in cell function. The introduction of efficient integration-free methods for cell reprogramming resolved this and alternative induction approaches have been developed since, such as adenoviruses [42], plasmids [43, 44], transposons [45], synthetic mRNAs [46], recombinant proteins [47] and Sendai viruses [48, 49]. These non-integrating systems have increased not only the safety, but also the efficiency of the reprogramming process [50, 51]. Currently, episomal vectors, synthetic mRNAs and Sendai viruses are the main choice for the generation of integration-free iPSCs.

Additionally, since 2009, xeno-free conditions have been developed to overcome the limitations associated with traditional culture methods and to eliminate animal components. Currently, combinations of chemically defined media and recombinant matrix proteins, such as vitronectin and laminin, are widely used for the generation and maintenance of human iPSCs [52-54].

However, the mechanisms of OSKM-mediated reprogramming are still not fully understood. Reprogramming seems to be a multi-step process, with an initial, stochastic early phase of reprogramming leading to the generation of partially reprogrammed cells, which can then enter a second late phase, resulting in fully pluripotent cells. In the early phase of reprogramming, OSKM occupy many genomic loci, including loci these factors do not bind to in ESCs. These loci include enhancers and promoters of genes that determine the somatic identity of the cell, and binding will lead to the silencing of somatic genes needed to gain a pluripotent state [55, 56].

However, exactly how the ectopic expression of OSKM induces the transition to a pluripotent state remains an area of investigation. There are some reports indicating that OSKM can bind to chromatin regions that are not accessible to other factors, leading to the remodelling of specific chromatin regions, and thereby activating or repressing gene expression [55].

As mentioned previously, reprogramming is associated with epigenetic changes, and epigenetic modifiers have been studied with respect to their involvement in reprogramming [57]. These enzymes can influence the up or down regulation of both pluripotency-associated, as well as somatic genes. In this way cell fate is driven in a particular direction, and this effect depends on the nature of the histone marks. Besides transcriptional changes, somatic cells going through reprogramming also need to reset their epigenome to the ESC-like state. This is accomplished by changing the DNA methylation and post-translational histone modifications patterns [56, 58-60].

However, chromatin-modifying factors involved in this process, have not been identified completely. Studies using HDAC inhibitors have shown that they can improve the reprogramming process, which is likely to be due to an increase of global histone acetylation levels, but can also have an effect on the acetylation states of non-histone proteins involved in the reprogramming process [61-64]. CBP and p300 are important epigenetic proteins and in somatic cells, c-Myc is known to recruit p300 to enhancer regions to regulate gene expression. However, in ESCs this interaction seems to be absent and p300 is rather targeted to enhancers by Oct4-Sox2-Nanog interactions [65]. This suggests an important role for p300, and potentially for CBP too, in the reprogramming process.

It is clear that epigenetic mechanisms not only play important roles during the generation of iPSC, but also affect the properties of reprogrammed iPSCs. However, more research is needed to find the specific acetyltransferases involved in reprogramming and to reveal the mechanism underlying this process. Understanding the roles of various epigenetic factors in iPSC generation contributes to our knowledge of the reprogramming mechanisms.

Research with ESCs has shown that pluripotent cells are suitable models for the study of developmental or early onset disorders [66]. Soon after the first reports of human iPSCs, these cells were used to generate models of genetic disorders [67], creating so-called 'disease in a dish' models. The first studies were simply demonstrating that it is possible to generate disease specific iPSCs and a growing library of human disease specific lines has been established since then. This *in vitro* disease modelling has had a major impact on understanding biological pathways but also for therapeutic implications, such as pharmaceutical screening and therapies.

Wang et al. generated and characterized iPSCs from patient fibroblasts with Zellweger spectrum disorder (PBD-ZSD) [68]. They compared gene expression profiles of the PBD-ZSD fibroblasts and iPSCs and found that the gene-expression profiles of the patient-specific iPSCs, but not skin fibroblasts, reflected the proposed phenotype seen in this disorder.

Copy number variations in 7q11.23 can cause William-Beuren syndrome (deletions) and 7q-microduplication syndrome (duplication), which are rare neurodevelopmental disorders, characterized by developmental delay and specific facial features. Through analysis of these patient derived iPSCs, Adamo et al. found that 7q11.23 dosage imbalance disrupts transcriptional pathways beginning in the pluripotent state, which are further exacerbated upon differentiation of these iPSCs into disease relevant lineages [69].

Besides genetic diseases, several cancers have been modelled with iPSCs [70] to explore the molecular mechanisms of cancer progression. However, there have been relatively few reports demonstrating successful reprogramming of cancer cells. Even though knockdown of tumour suppressor genes is known to enhance reprogramming efficiency [71-73] reprogramming of human primary cancer cells has been inefficient. It is thought this is because of cancer-specific mutations, epigenetic modifications and/or accumulation of DNA damage [74-76]. Despite these difficulties, there are several reports of successful generation of iPSC lines from existing cancer cell lines, such as melanoma, breast cancer and glioblastoma [77-80].

In summary, iPSCs are providing valuable *in vitro* model systems to investigate molecular mechanisms and creating disease models. Combined with genetic and epigenetic modifiers these are useful tools for drug screening and evaluating targeted therapeutic interventions.

Even though the generation of human iPSCs is a technically simple technique and frequently applied, it is a time consuming and inefficient process. There are several elements that can have an impact on the reprogramming process, such as the somatic cell type, the choice of reprogramming factors and method to deliver these factors. But also aspects like culture conditions and methods to characterize reprogrammed and pluripotent cells will influence the outcome [81]. However, a concerning feature of pluripotent cells is their potential to turn into tumorigenic cells. Therefore, if applied in a clinical setting, methods need to be developed for safe regenerative medicine therapies to exclude the tumorigenicity of these cells [82].

Despite their different origins, iPSCs are extremely similar to ESCs in that they have been shown to express pluripotency markers and support the differentiation into cell types of all three germ layers [83, 84]. However, recently here has been a lot of discussion as whether iPSCs are truly ESCs. Several differences have been detected between these two types of PSCs [85], including the persistence of epigenetic memory from the somatic cells of origin [86, 87], differential DNA methylation signatures [88, 89], a different extent of genetic aberrations [90], and differences in gene expression profiles between cells [91, 92]. There are also studies that identified persistent donor cell-specific gene expression patterns in iPSCs produced from different cell types, suggesting an influence of the somatic cells of origin on the molecular properties of derived iPSCs [93].

In addition, there are few reports where significant genome-wide difference between iPSCs and ESCs remain undetected [59, 91, 94]. Although ESC and iPSC models may vary for certain disorders, iPSC models for most disorders are expected to mimic ESC models, thus highlighting the utility of reprogramming of patient cells.

Even though iPSCs retain some transcriptional and epigenetic memory of their cell of origin, the majority is only present in early iPSC passage numbers. These aberrations are lost upon continuous passaging, and can therefore be considered a transient epigenetic memory [95, 96]. It is known that reprogramming takes longer than previously thought, and continues for several passages even after the appearance of ESC morphological features and expression of pluripotency markers. This has been shown in mouse iPSCs at a higher passage number (p16). These cells lose the differences in gene expression associated with the cell of origin that is observed at earlier passages [94, 96].

Besides these expression and epigenetic aberrations, it has been proposed that iPSCs also bear genomic mutations, which can range from chromosomal aneuploidy to single base mutations [97-100]. These can result from the reprogramming process itself or arise during *in vitro* expansion [101-103] and it is though that one out of five clones will have gross chromosomal aberrations [97] and/or copy number variations (CNVs) [98]. However, when comparing these studies, there is no correlation between the extent of genetic aberrations and the method of reprogramming used and therefore these studies are still controversial.

Combined, iPSCs have and will contribute tremendously to developmental biology and disease modelling. Differentiating these cells to generate a disease in a dish

model will present unique opportunities for clinical applications with personalized medicine.

It is clear that both CBP and p300 play fundamental roles in epigenetic regulation of gene expression through various mechanism, e.g. as coactivators and their transient acetylation activity. Therefore, cells with a mutation in one of these genes will represent powerful tools for analyzing disturbed gene regulation in a range of pathological conditions. This study will reprogram primary fibroblast cell lines with known mutations in the *CREBBP* gene. Since epigenetic dynamics are very important in the reprogramming process this will also shade light on the role of these proteins in reprogramming and the acquisition of pluripotency. These RTS-iPSCs will be characterized both molecularly and functionally.

In summary, the lack of access to disease tissues of patients with RTS severely limits the translation of molecular observations in animal models to the clinic. The use of iPSCs can however provide some initial insight into the molecular etiology of the disease and this chapter will focus on the generation and characterization of patient specific iPSCs.

Material & Methods

Cell culture

Somatic cell lines used for reprogramming were primary human dermal fibroblasts. Two RTS primary fibroblast cell lines with known mutations in *CREBBP* were obtained from Leiden University, the Netherlands. RTS fibroblast cell lines were grown in AmnioMAX C-100 basal medium (Gibco, 17001074), supplemented with AmnioMAX C-100 Supplement (Gibco, 12556023) and 1% 10,000 U/ml Penicillin-Streptomycin (Pen/Strep) (Gibco, 15140122). WT cell lines used are commercially available and grown under standard culture conditions (37°C, 5% CO_2). HDFa and HDFn lines were grown in Medium 106 (Gibco, M-106) supplemented with 2% Low Serum Growth Supplement (Gibco, S-003-10) and 1% Pen/Strep. The fibroblast cell line from ATCC was grown in standard fibroblast medium (DMEM, 10% fetal bovine serum (FBS), 1% GlutaMAX 100X (Gibco, 25030081), 1% MEM Non-Essential Amino Acids Solution 100X (10mM) (NEAA) (Gibco, 11140050), and 1% Pen/Strep. Cells were harvested with Trypsin-EDTA (0.25%) (Gibco, 25200056) for 3-4 minutes at room temperature and collected with culture medium.

Whole cell protein lysate

Cells from a confluent T25 were washed twice with cold PBS and cold lysis buffer (250μl) was added while keeping the flask on ice. Lysis buffer (RIPA) used was 25mM Tris-HCl (pH 7.6) (Sigma-Aldrich, T1503), 150mM NaCl (Sigma-Aldrich, S7653), 1% NP-40 (Sigma-Aldrich, I8896), 1% sodium deoxycholate (Sigma-Aldrich, D6750) and 0.1% SDS (Sigma-Aldrich, L3771), supplemented with 7X Complete Protease Inhibitor Cocktail (Roche, 04693116001), 1mM PMSF (Sigma-Aldrich, P7626) and 20mM Sodium Butyrate (NaBu) (Sigma-Aldrich, B5887). Cells were scraped and collected on ice. DNA was sheared using a sonicator (2 cycles of 30 seconds ON, 30 seconds OFF) and samples were kept on ice throughout. Samples were clarified by spinning at 14,000rpm for 15 minutes at 4°C and supernatant was collected and frozen at -80°C. Protein concentrations were determined using the Qubit with the Qubit Protein Assay Kit (Invitrogen, Q33211).

SDS PAGE and Western Blot

Protein extract (30μg) was combined with 5μl 4X NuPAGE LDS sample buffer (Invitrogen NP0007), and 2μl 10X NuPAGE Reducing agent (Invitrogen, NP0004) in a total volume of 20μl in water. Samples were boiled at 70°C for 10 minutes, spun down and loaded onto a 4-12% gradient gel (Invitrogen, NP0335). Gel was run in 1X SDS MOPS running buffer (Invitrogen, NP0001) at 180V for ~50 minutes. Gel, membrane, sponges and filter paper were soaked in 1X NuPAGE transfer buffer (NP0006-1) and run for 2 hours at 70V.

Membrane was blocked for 1 hour at room temperature in Odyssey Block reagent (Li-Cor Biosciences, 927-400000) and incubated overnight with primary antibodies diluted in Odyssey Block reagent at 4°C. After washing with PBS/Tween-20, the membrane was incubated with secondary antibodies diluted in Odyssey Block reagent and incubated for 1 hour at room temperature in the dark.

The following primary antibodies were used: CBP (SantaCruz, sczsc-369), H3K9ac (Abcam, ab4441), H3K18ac (Abcam, ab1191), H3K27ac (Abcam, ab4729), H3 (Abcam, ab12079), and Actin (Abcam, ab8227) at 1:1,000. For secondary antibodies, donkey anti-rabbit IgG 680 (Li-Cor Biosciences 926-32223) and donkey anti-goat IgG 800 (Li-Cor Biosciences, 925-32214) were used at 1:10,000. The membrane was scanned on the Odyssey Imaging System CLx using the Odyssey Application Software.

Reprogramming

Fibroblast cell lines were reprogrammed using the CytoTune-iPS 2.0 Sendai reprogramming kit (Invitrogen, A16518). On day -2, fibroblasts were plated in a 12-well plate format and seeded at different densities ranging from 20,000-100,000 cells per well. On day 0 wells that showed roughly 80% cell confluence were selected. Virus (hKOS, hc-Myc and hKlf4) was added at a Multiplicity of Infection (MOI) of 5-5-3 respectively in the standard fibroblast culture medium for each cell line and this was refreshed 24 hours after infection (without virus). Cells were cultured for another 6 days in normal fibroblast medium, which was replenished every second day. On day 5 mouse embryonic fibroblasts (MEFs) were plated in 10cm dishes coated with gelatin (400,000 MEFs per plate), and on day 7 cells were harvested with 0.25% Trypsin and plated on the MEF dishes. Cells were cultured in human iPS (hiPS)

medium (DMEM/F12, GlutaMAX supplement (Gibco, 10565018), KnockOut Serum Replacement (KSR) (20%) (Gibco 10828028), NEAA (1%), β-mercaptoethanol (55mM) (0.1%) (Gibco, 21985023), Pen/Strep (1%), hFGF (10ug/ml) (Gibco, PHG0021)). Starting on day 8, reprogramming plates were observed and medium was replaced every day thereafter.

Picking of iPSC colonies

Clumps started to emerge between roughly 12-18 days post transduction. However, depending on the cell type, colonies might not appear until 4 weeks of growing. Normally, three to four weeks after transduction, colonies should have an appropriate size for transfer. Colonies should resemble a hESC-like morphology, characterized by a flat cobblestone-like appearance with individual cells clearly defined from each other in the colonies. Colonies should be picked closer to three weeks to avoid differentiation. The day before transferring the colonies, MEFs were plated in 60mm center well culture dishes (200,000 MEFs/dish). Colonies were picked manually and cut up in 6-8 pieces, depending on the size of the colony, with a 25 gauge $1\frac{1}{2}$ inch needle. The cut pieces were transferred to the freshly prepared MEF dishes in iPSC medium. The dishes were incubated under normal conditions and cells were allowed to attach for 48 hours before replacing with fresh hiPS medium, which was changed every day thereafter. Colonies were manually expanded from here on. When freezing down, cells were harvested with collagenase I (300 U/ml) (Worthington, CLS1 LS004194) and frozen down in iPSC freezing medium, which consist of freezing medium A (hiPSC medium and KSR (50%)) and freezing medium B (hiPSC medium and DMSO (20%)) in a 1:1 ratio.

Feeder-free system

After expanding clones manually, cultures were changed to a feeder-free system. Culture vessels were coated with Vitronectin (100X, 0.5ug/cm^2) (Gibco, A14700) and MEFs were plated at half the usual density. Cells were plated in normal hiPSC medium and 48 hours after replating, medium was changed to 50% Essential 8 (E8) medium (Gibco, A1517001). This was increased by 25% each medium change, until 100% E8 medium was reached. Cells were harvested and made into a single cell solution with 0.5mM EDTA in PBS (0.5M UltraPure, Gibco, 15575020) and plated

directly onto Vitronectin-coated vessels, without MEFs, and grown in E8 medium. From this point cells were enzymatically expanded in single cell solution with EDTA. Rho Kinase inhibitor 1000X (ROCKi, Y-27632 dihydrochloride) (Abcam, 120129) was added to the medium for the first 24 hours after replating to improve colony formation.

Sorting of iPSC by FACS

To purify the iPSCs, cell pellets were resuspended in labelling buffer containing 2% FBS, 1000X ROCKi, 100X EPCAM-BV650 (BioLegend, 324226) and 100X TRA-1-60-BUV395 (BD BioSciences, 563878) in PBS and incubated for 15 minuted on ice in the dark. Cell were washed with PBS and resuspended in FACS buffer (PBS, 2% FBS, 1000X ROCKi and 500X propidium iodide (PI) (Sigma-Aldrich, P4864)). Double positive cells were collected in E8 medium supplemented with 1000X ROCKi and after 24 hours in culture, medium was replaced with E8 medium. FACS sorting was performed by FlowCore at Monash University Melbourne, Australia.

Genomic DNA extraction

Frozen iPSC pellets (one confluent well of a 6-well plate) were thawed on ice from -20°C, and resuspended in 187.5µl lysis buffer (50mM Tris-HCl pH 8.0, 100mM NaCl, 10mM EDTA pH 8.0, 0.5% SDS) and 4.7µl Proteinase K (20 mg/ml). Samples were incubated for 2 hours at 55°C and mixed on an Eppendorf mixer for 5 minutes after. Then, 62.5µl of 5M NaCl was added and the samples were mixed again for 5 minutes before being centrifuged for 10 minutes at full speed (13,000 rpm). Supernatant (200-250µl) was transferred to a new tube and 125µl of isopropanol was added. Samples were mixed for 2 minutes and spun for 5 minutes at full speed. Samples were washed with 375µl of 70% ethanol and centrifuged for 5 minutes at full speed. The supernatant was taken off, and the pellet resuspended in 30µl of TE buffer. Samples were incubated for 2 hours at 37°C on a shaker. Samples were then quantified and stored at 4°C or -20°C.

Sanger Sequencing

DNA from iPSCs was sequenced using the following primers to confirm the respective *CREBBP* variants found in the fibroblasts. Primers used were Exon 29 (RTS-NL1) Forward 5' GCGACAGCAAGAATGCCAAG 3' and Reverse 5'

ACGCCTACCTCCTTGTGCTTCTC 3', Exon 3 (RTS-NL2) Forward 5' CAAGTCCATTTGGACAGCCC 3' and Reverse 5' ATTTGGCACGTTGGTGACTG 3'. Sanger sequencing was performed by Micromon, at Monash University Melbourne, Australia.

RNA extraction and cDNA preparation

RNA was extracted using the RNeasy Mini Kit (Qiagen, 74104) on the QIAcube including DNAse I treatment (Qiagen). RNA was quantified and SuperScript III First-Strand Synthesis System (Invitrogen, 18080051) was used to produce cDNA.

SeV genome expression

PCR on cDNA from several iPSC clones per line, including an early passage iPSC clone (HDFa p4) as a positive control, was performed with primers for the SeV genomes following the CytoTune protocol (Forward: 5' GGATCACTAGGTGATATCGAGC 3', Reverse: 5' ACCAGACAAGAGTTTAAGAGATATGTATC 3', 181 bp product) and *GAPDH* (5' TGAAGGTCGGAGTCAACGGA 3' and 5' CCATTGATGACAAGCTTCCCG 3') as a positive control. The PCR program was as follows: 35 cycles of 95C for 30 sec, 55C for 30 sec, 72C for 30 sec. To confirm, PCR product were run on a 1.5% agarose gel (100V for 1 hour).

CREBBP expression

Quantitative PCR (qPCR) was used to look at *CREBBP* expression in the iPSC lines. Primers used were Primer A Exon 1 Forward 5' CAACCCCAAAAGAGCCAAAC 3' and reverse 5' GGTTCCCACTGTTTAAAAGGC 3' and Primer B Exon 16 Forward 5' CTCTCAGTCAACATCTCCTTCG 3' and reverse 5' CGGAAAGGTAATGACTCTGGG 3'. Samples were run in triplicate and each reaction contained 3.5µl 2X SYBR Green Master Mix (Roche Applied Science, 04707516001), 0.35µl of each primer (10µM), 1.4µl water and 1.4µl cDNA (total volume 7µl). Samples were run in 384 plates on a LightCycler 480 system (Roche Applied Science), with the following cycling condition: 94°C for 30s; 60°C 30s and 72°C for 30s for 45 cycles followed by 72°C for 10min. Gene expression was standardized to housekeeper gene *GAPDH* (5' TGAAGGTCGGAGTCAACGGA 3'

and 5' CCATTGATGACAAGCTTCCCG 3'). Statistics were performed in GraphPad Prism using the one-way ANOVA with Dunnet's multiple comparisons test.

Karyotyping
iPSCs were grown to 80% confluence in a T25 under normal conditions and karyotyping was performed at the Monash Health Cytogenetics Unit, Monash Medical Centre, Melbourne Australia.

Fixing and staining of iPSC colonies
iPSC colonies were washed with PBS and fixed with 4% PFA for 10 minutes in the dark at room temperature. Cells were lysed with 0.3% Triton X-100 (Sigma-Aldrich, T9284) and incubated for 30 minutes at room temperature. Colonies were stained overnight with primary antibodies (NANOG 1:100 (Abcam, ab21624), OCT3/4 1:100 (SantaCruz, sc5279), SSEA4 1:100 (BioLegend, 330404), TRA-1-60 1:300 (BD BioScience, 562711), DAPI 1:1000 (Invitrogen, D1306)) at 4°C and incubated with secondary antibodies (1:400 dilution) for 2 hours at room temperature in the dark. Secondary antibodies used were AF488 goat anti-rabbit IgG (A-11008), AF555 goat anti-mouse IgG2b (A-21147), AF555 goat anti-mouse IgM (A-21426) all from Invitrogen, and Streptavidin-FITC (eBioscience, 11-4317-87). iPSC colonies were visualized with an inverted fluorescent microscope (Nikon Eclipse Ti-S, NIS-Elements software BR 3.20.00).

Pluripotency characterization (Teratoma formation)
Teratoma formation was monitored after injecting iPSCs into the testis of NODSIL2R mice (~1 million cells/testis). These were done in duplicate and performed and monitored by the animal facility at Monash University, Melbourne Australia. After teratoma formation was observed, the testes were harvested and a haematoxylin and eosin (H&E) staining was performed by the histology facility at Monash University, Melbourne Australia.

Ethics
All experiments involving human samples and animals were conducted with approval by and in accordance with the Monash University Ethics agreement (MUHREC 2012/3462/3271 and MARP2012/052).

Results

Two RTS fibroblast cell lines with known *CREBBP* mutations were reprogrammed into iPSCs, as well as three WT primary fibroblast cell lines. The two patients are both male, with fibroblasts taken from the sternum or upper arm. As controls, HDFa, which is a female dermal fibroblast cell line taken from breast tissue and two male neonatal fibroblast cell lines, both from foreskin were reprogrammed (Table 1 and 2).

Table 1. RTS cell line information

Cell line	Sex	Tissue	Passage	Frameshift mutation *CREBBP*
RTS-NL1 (Published as 213-1)	M	Sternum	P4	c.4837 del G, Exon 29
RTS-NL2 (Published as 199-3)	M	Forearm	P8	c.904_905del AG, Exon 3

Table 2. Wild type cell line information

Cell line	Sex	Tissue	Passage	Details
Human Dermal Fibroblast Adult (WT-HDFa)	F	Breast tissue	P6	Gibco (C-013-5)
Human Dermal Fibroblast Neonatal (WT-HDFn)	M	Foreskin	P3	Gibco (C-004-5C)
Human Dermal Fibroblast Neonatal (WT-ATCC)	M	Foreskin	P4	American Type Culture Collection (ATCC) (CRL-2097)

RTS fibroblasts were obtained from Leiden University, the Netherlands and grown in AmnioMax, a well-defined medium, which has been used for culturing human amniotic fluid cells and fibroblasts [104].

Because CBP is ubiquitous expressed and plays crucial roles in general cellular processes, the first thing to investigate was to see if this is reflected in the fibroblast phenotype. Cells were grown under standard conditions (37°C, 5% CO_2) and monitored daily under the microscope (Figure 1).

Next, specific histone marks that are known to have an association with CBP were looked at. Due to the reduced CBP levels in RTS fibroblasts and therefore acetylation levels, these histone marks might be affected. Proteins were extracted from the two RTS lines and two control lines (HDFa and HDFn). A Western Blot was performed with specific antibodies against histone marks H3K9ac, H3K27ac and H3K18ac (Figure 2) to establish a biochemical characterization, with antibodies against total H3 and Actin as positive controls. Please note that this Western blot has been cropped and modified and will be displayed in total in Chapter 3 (Figure 5).
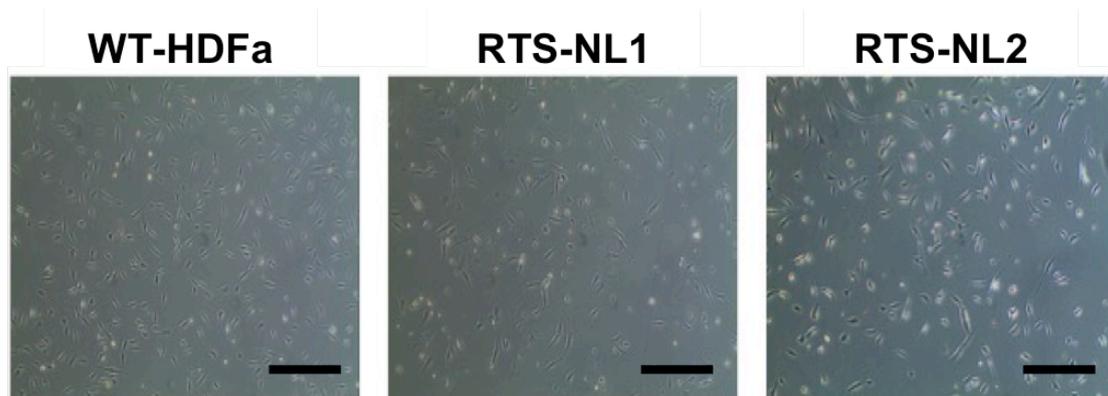
Figure 1. Primary fibroblast cell lines. Cell lines were grown under specified conditions and images were taken one day after replating. Passage numbers for WT-HDFa, RTS-NL1 and RTS-NL2 were p3, p6 and p8 respectively. Magnification of images 4X, scale bars 100 μm.
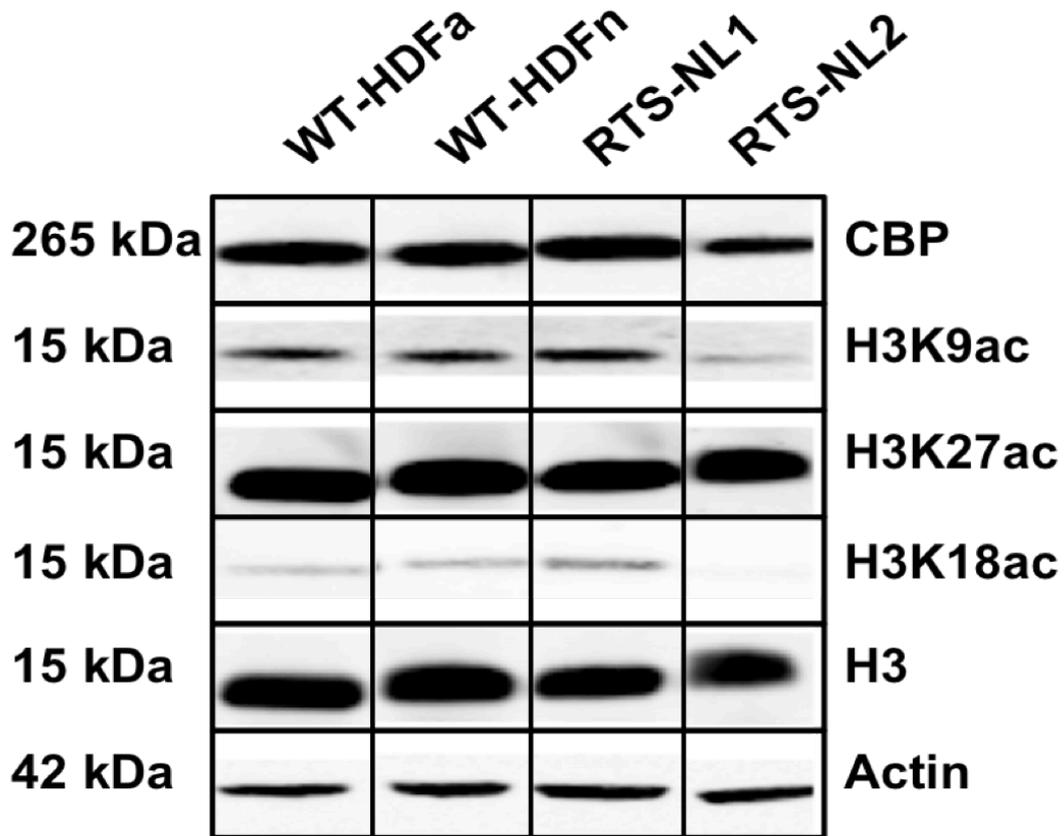
Figure 2. Western Blot for epigenetic factors. Western blot showing whole protein extracts from fibroblast taken from two control cell lines (WT-HDFa and WT-HDFn) and two RTS cell lines (RTS-NL1 and RTS-NL2). Proteins shown are CBP and several histone acetylation marks known to be affected by CBP (H3K9ac, H3K27ac and H3K18ac). Histone 3 (H3 total) and Actin were taken as controls. Note that this Western Blot has been modified and is show in full format in Chapter 3, Figure 5A (page 90).

51

When looking at protein levels of the epigenetic factors, there was a clear reduction in CBP, H3K9ac and H3K18ac in RTS-NL2 detected compared to the two wild type cell lines (WT-HDFa and WT-HDFn). However, there was no decrease seen in RTS-NL1 for any of the proteins and histone marks tested. This could be the result of the locations of the frameshift mutations in the RTS cell lines. The mutation in RTS-NL2 is situated at the start of the protein, in exon 3, whereas the variant in RTS-NL1 is located towards the end of the protein, in exon 29. Even though both variants are frameshift mutations, due to the location (in exon 29 of 31 exons in total), RTS-NL1 might still produces a transcript that may lead to a truncated protein. It seems that this does not directly influence histone acetylation levels, but probably has an effect on different targets of CBP, leading to an RTS phenotype.

In the next step, these fibroblast cell lines were reprogrammed into iPSCs using the CytoTune-iPS 2.0 Sendai reprogramming kit (Invitrogen). This method is non-integrating and uses the Sendai virus (SeV) to deliver the four reprogramming factors into the fibroblasts. The feeder-dependent method as described in the manual was followed when establishing several clones per cell line.

The experiment was set up in a 12-well plate format and the viral vectors were added to the wells plated on day -2 with 100,000 cells (~80% confluent on day 0). The virus was added at specific volumes according to the MOI (Table 3).

Table 3. Cell counts and volumes of vector with reprogramming factors

		hKOS MOI=5	hc-Myc MOI=5	hKlf4 MOI=3
Cell line	Cell count	8.5×10^7 CIU/ml	8.0×10^7 CIU/ml	1.0×10^8 CIU/ml
WT-ATCC	1.1×10^5	6.5 µl	6.9 µl	3.3 µl
WT-HDFn	1.0×10^5	5.9 µl	6.3 µl	3.0 µl
WT-HDFa	9.8×10^4	5.8 µl	6.1 µl	2.9 µl
RTS-NL1	9.9×10^4	5.8 µl	6.2 µl	3.0 µl
RTS-NL2	1.1×10^5	6.5 µl	6.9 µl	3.3 µl

$$\text{Volume of virus (µl)} = \frac{\text{MOI (CIU/cell) x number of cells}}{\text{Titer of virus (CIU/ml) x } 10^{-3} \text{ (ml/µl)}}$$

The protocol stated colonies should be starting to form on day 12 post-transduction (BJ cell line). However, the first colonies did not emerge until day 15-18 for the cultures and the first colonies were picked from day 32 onwards (3-4 weeks post transduction in the manual). This might have been due to differences between cell lines and passage numbers, slightly different culture conditions, or viral related.

Per cell line, up to 12 colonies were picked and the clones that continued to show ESC-like characteristics, were expanded and cryopreserved. After several expansion rounds, a minimal of five clones per cell line were generated that showed typical hESC-like morphology, such as round, tightly packed cells in round colonies with sharp edges (Table 4).

Table 4. Clones generated for each iPSC cell line used in this study

iPSC line	Expanded clone numbers						
WT-ATCC	2	3	6	7	9		
WT-HDFn	3	5	7	11	12		
WT-HDFa	1	2	5	7	8	10	13
RTS-NL1	2	3	4	6	8	9	11
RTS-NL2	1	3	5	7	9	12	

The culture system was then converted from a feeder set up to a feeder-free environment, which involved Vitronectin as coating substrate and Essential 8 (E8) medium as culture medium. To transit between systems, vessels were coated with Vitronectin and plated with half the number of MEFs (750,000 MEFs/T25) for the first split. The iPSCs were grown in normal hiPSC medium and 48 hours after replating, the medium was changed to 25% Essential 8 medium. This was increased with each daily medium change until 100% E8 medium was reached. When vessels were confluent again, cells were plated directly on Vitronectin (without MEFs) and grown in E8 medium. A schematic timeline is displayed in Figure 3.

Sendai virus is a non-integrating approach to deliver reprogramming factors into the host cell, in which the virus stays in the cytoplasm of the cell and is eliminated by passaging. To confirm the absence of the reprogramming vectors, a qPCR was performed to detect expression of the Sendai virus genome (SeV). cDNA was generated from 5×10^6 cells for two clones per iPSC line. As is shown in Figure 4, all clones, except ATCC 1p11, showed a negative result for SeV.

After confirming the generation of SeV free clones that resemble ESC-like morphology, the iPSCs were characterized. Firstly, several of the iPSC clones were karyotyped and demonstrated a normal profile for the clones as shown in the karyograms in Figure 5. WT cell line HDFa is a female cell line, together with four male cell lines (HDFn, ATCC, RTS-NL1 and RTS-NL2). Secondly, the pluripotency of the generated iPSCs was confirmed by looking at the expression of specific pluripotency markers. An immuno-fluorescence stain assay for SSEA4, TRA-1-60,

NANOG and OCT3/4 was performed and as is shown in Figure 6, all clones were positive for the pluripotency markers tested.

Then, a functional test was performed to demonstrate pluripotency and several iPSC clones were injected into the testes of NODSIL2R mice. Tumour growth was monitored and when of appropriate size, between 10-14 weeks after injection, tumours were seized. Teratomas were sectioned and an H&E staining was performed, showing differentiation into the three germ layers (Figure 7).
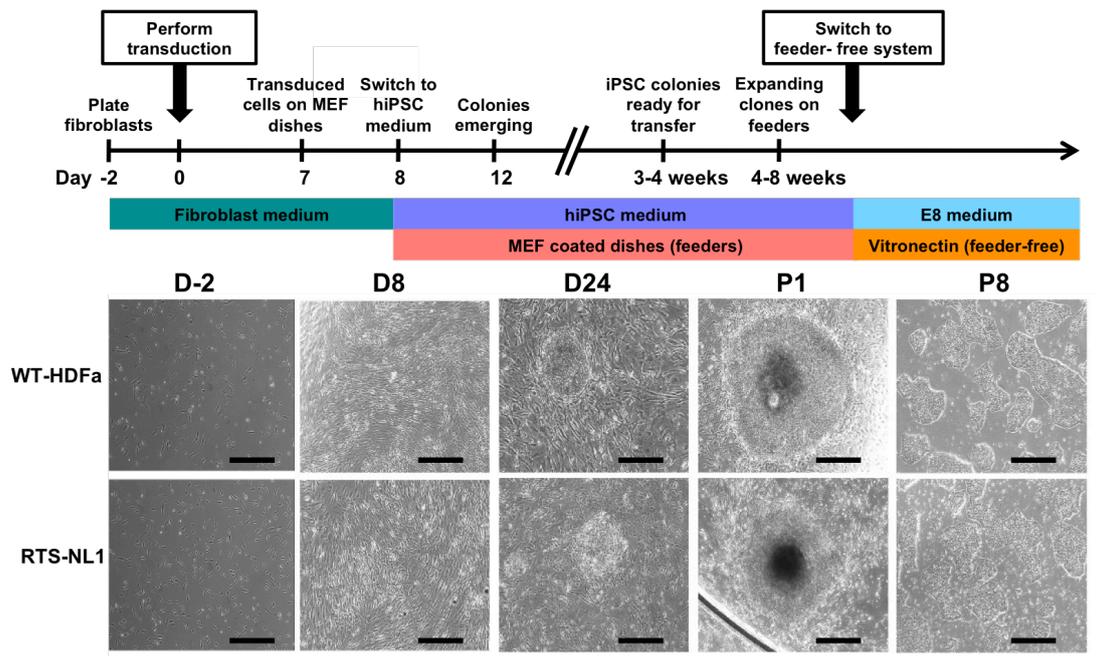
Figure 3. Timeline for the generation of human iPSCs using the CytoTune reprogramming kit. Fibroblasts were reprogrammed using Sendai virus to deliver the factors on day 0. On day 7 the fibroblasts were plated onto MEF feeder cells and cultured in iPSC medium. Colonies started to form after 3-4 weeks and several clones were expanded. When established, clones were transitioned into a feeder free system and grown in vitronectin-coated vessels in E8 medium. Pictures were taken during differentiating at days 2, 8, 24 and after passaging at P1 and P8 (vitronectin/E8) for both WT (top) and RTS (bottom) iPSC clones. Magnification 10X, scale bars 100 μm, n = 5-7 clones per cell line.
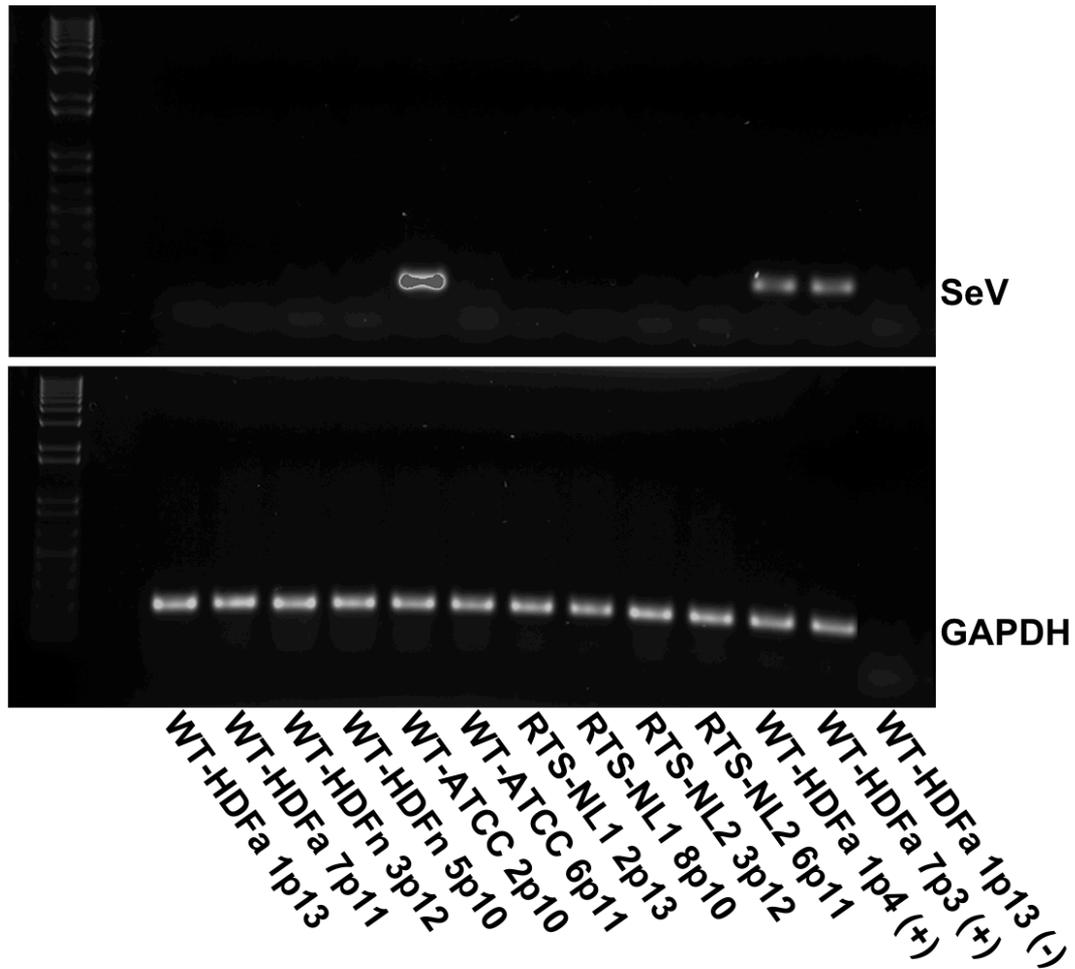
Figure 4. PCR with primers for SeV. Two clones per cell line were tested for the SeV construct. All samples were negative, except clone WT-ATCC 2p11. As a positive control two WT-HDFa clones were used at passage 4 and 3 respectively, and cDNA without SSIII as a negative control. *GAPDH* expression was used as a positive control (n = two clones per cell line).
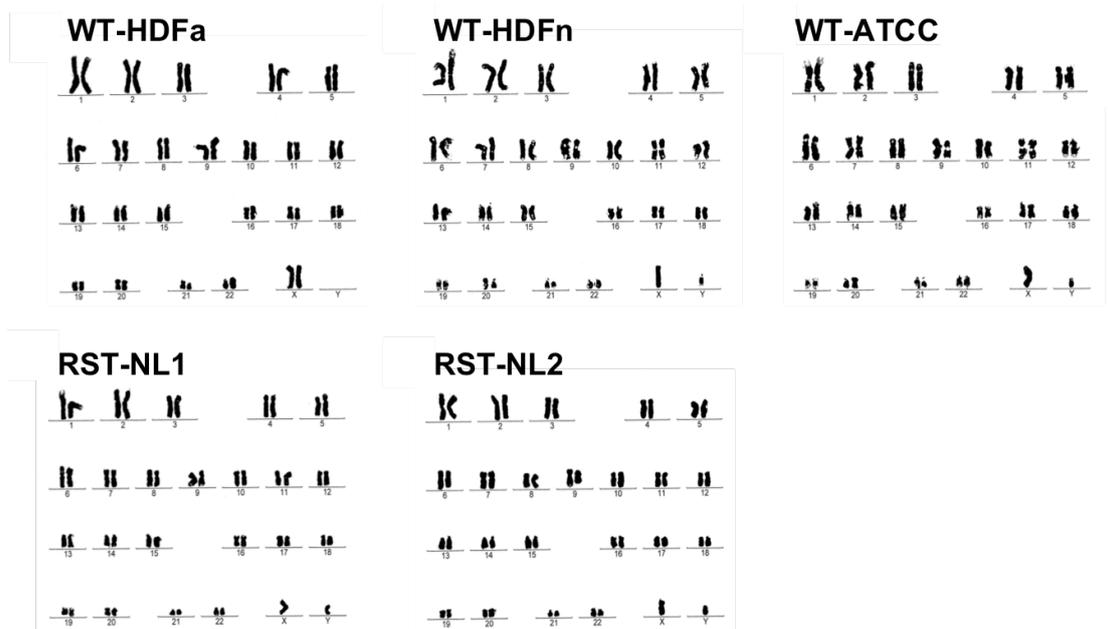
Figure 5. Karyograms for iPSCs. One iPSC clone per cell line was karyotyped. The clones tested have a normal karyotype, containing 22 pairs of autosomal chromosomes and one pair of sex chromosomes (46,XX for WT-HDFa, and 46,XY for WT-ATCC, WT-HDFn, RTS-NL1 and RTS-NL2, n = one clone per cell line, HDFa 7p11, HDFn 3p12, ATCC 6p11, RTS-NL1 2p13, RTS-NL2 3p12).
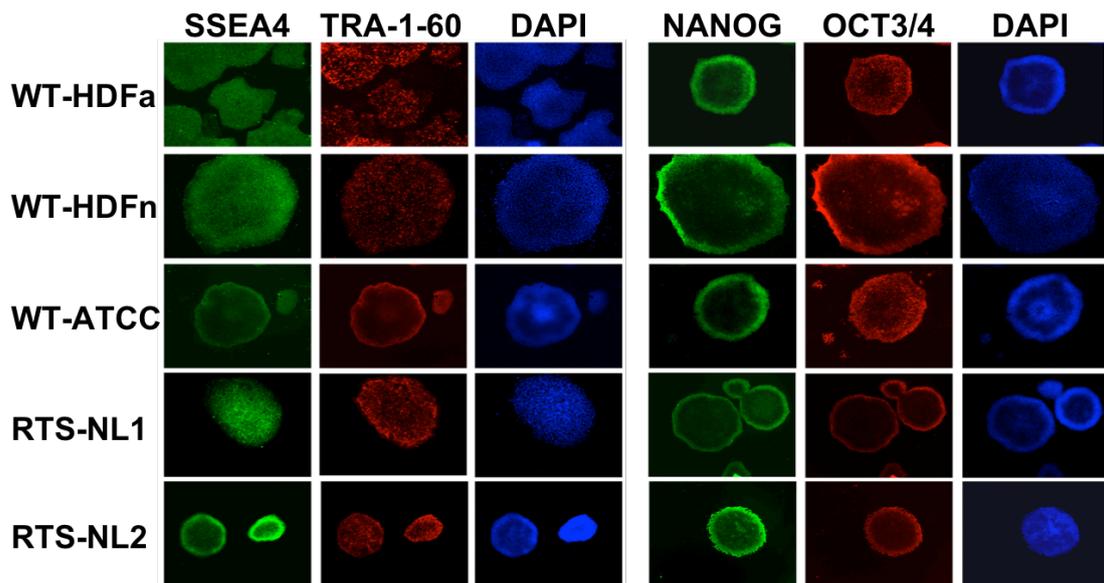
Figure 6. Staining pluripotency markers in iPSC colonies. Immunofluorescence for the pluripotency markers SSEA4, TRA-1-60, NANOG and OCT3/4 in iPSCs. DAPI, 4′,6-diamidino-2-phenylindole. Magnification 10X, n = one or two clones per cell line, HDFa 1p13 and 7p11, HDFn 3p12 and 5p10, ATCC 6p11, RTS-NL1 2p13 and 8p10, RTS-NL2 3p12 and 6p11.
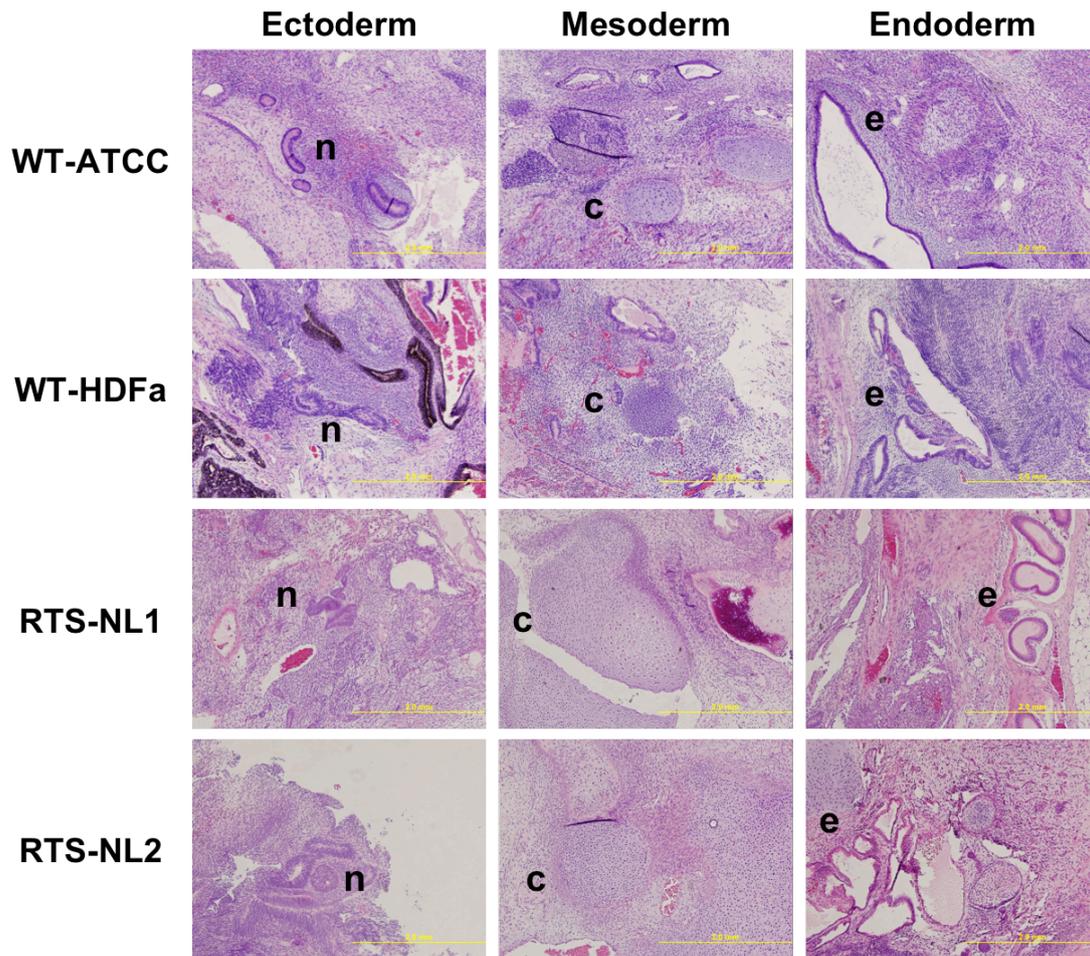
Figure 7. *In vivo* **differentiation of human iPSCs into teratomas.** Clones of all iPSC lines showed tumour growth when injected into testis of NODSIL2R mice over a period of 10-14 weeks. Teratomas were sectioned and an H&E stain shows differentiation into all three germ layers (Ectoderm: neuronal rosette like structures (n), Mesoderm: cartilage (c), Endoderm: epithelium (e)). Scale bars 2mm, n = one or two clones per cell line, HDFa 1p13 and 7p11, HDFn 3p12 and 5p10, ATCC 6p11, RTS-NL1 2p13 and 8p10, RTS-NL2 3p12 and 6p11.

To confirm the known *CREBBP* mutations were still present in the iPSC lines, DNA was extracted from several iPSC clones and a PCR was performed. The specific frameshift variants in *CREBBP* in the two RTS iPSC clones were confirmed with Sanger sequencing, which were absent in the control lines (Figure 8). Finally, it was shown by qPCR that the expression of *CREBBP* was reduced by ~50% in the RTS iPSC lines, compared to the WT lines (Figure 9).
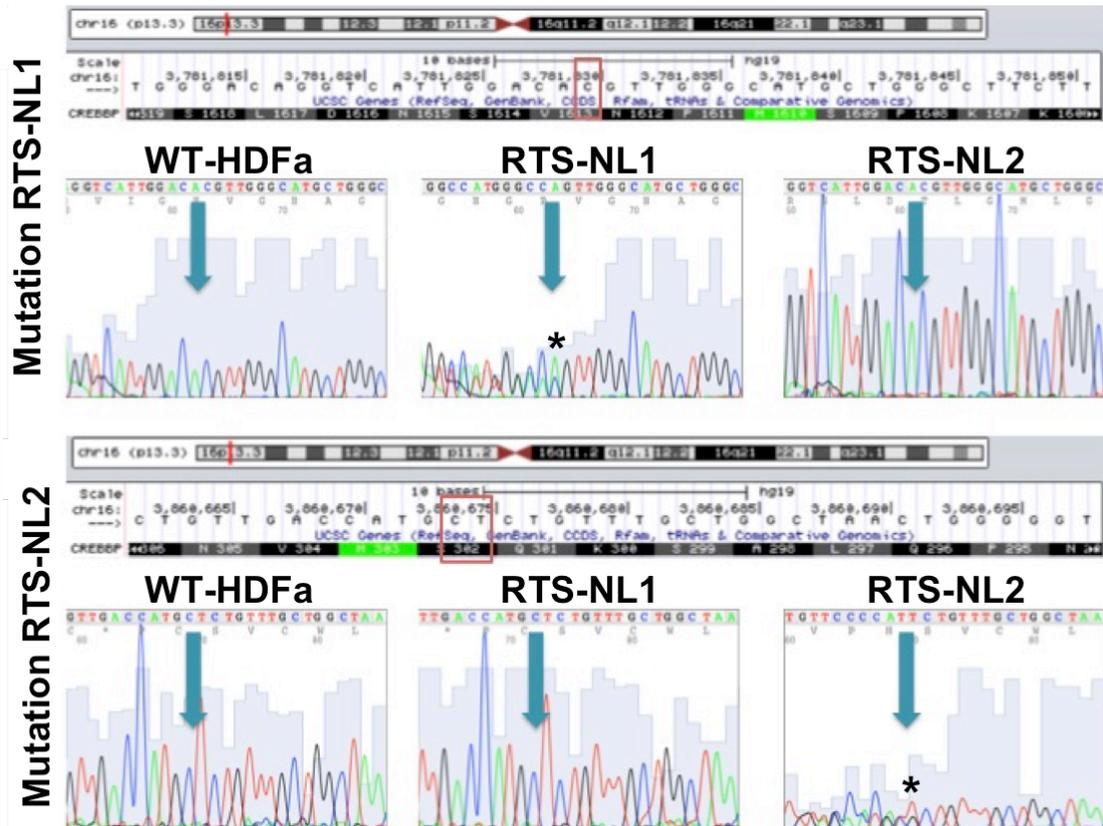
Figure 8. Sanger sequencing of *CREBBP* variants in iPSC lines. Sequencing traces showing the specific *CREBBP* mutations for RTS-NL1 (top) and RTS-NL2 (bottom). The location of the frameshift deletions in RTS-NL1 is chr16:3781830 and for RTS-NL2 is chr16:3860674-3860675, which are shown in the red boxes in the UCSC genome browser shots. The green arrows show the location of the deletions in the sequencing traces for RTS-NL1 and RTS-NL2, which are absent in the WT line (WT-HDFa). Mutation locations are based on the Genome Reference Consortium human genome build 37, human genome 19, GRCh37/hg19. (n = two clones per cell line, HDFa 1p13 and 7p11, RTS-NL1 2p13 and 8p10, RTS-NL2 3p12 and 6p11.)
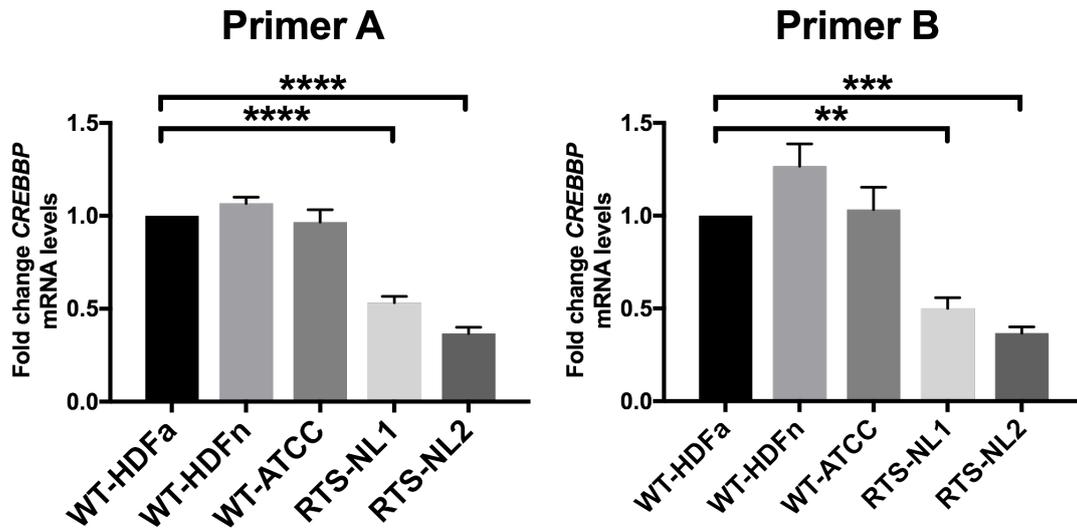
Figure 9. *CREBBP* expression iPSC lines. *CREBBP* expression was reduced in the RTS iPSC lines compared to the WT lines. CT values were normalized against *GAPDH* expression and the ratios were calculated setting WT-HDFa at 1. Mean and SEM were plotted with comparison by one-way ANOVA with Dunnet's multiple comparisons test, n= two clones per cell line, 3 technical replicates, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Discussion

In summary, it was shown that somatic cells with reduced levels of CBP can be reprogrammed into iPSCs. iPSC clones were generated from two RTS fibroblast cell lines with known frameshift mutations in *CREBBP*. These lines, in parallel with WT iPSC lines, were shown to be molecularly and karyotypically normal and displayed the characteristics of pluripotency. The lines were positive for several pluripotency markers tested such as NANOG and TRA-1-60 and most importantly they were able to form teratomas, showing tissues of all three germ layers.

Human models are needed to understand molecular and cellular processes, but also to develop therapies. Animal models are frequently used for drug screening; however, many disorders lack a suitable animal model or are not feasible to use for these experiments. Therefore, human iPSCs as models have become the new standard for certain disease modelling setups and in consequence in drug discovery.

Although these results are encouraging line-to-line variation due to genetic background between the fibroblast lines could not be completely discarded. However, genome-editing tools, such as CRISPR-Cas9, can introduce specific mutations into the genome of WT lines, providing an isogenic control for molecular studies. This will be further discussed in the next chapter. Furthermore, a study reprogramming several fibroblasts lines showed that there is a combined negative effect of donor age and time in culture on the reprogramming efficiency. Although all 11 lines gave rise to iPSC colonies, the efficiency declined with donor age and in addition, the late passage fibroblasts gave less reprogrammed colonies than the early passage lines. Therefore, care should be taken when comparing samples [104].

Nevertheless, parts of the reprogramming process remain unclear, and more research is needed. Studying the downstream targets of the OSKM may help understand the molecular mechanisms of reprogramming towards pluripotency and this might also increase the efficiency of reprogramming, which remains low.

The low efficiency of OSKM-mediated reprogramming is due to the fact that the majority of cells fail to complete the process, and only a small number will become iPSCs. Successful reprogramming depends on the occurrence of multiple events during different stages of this process [105] that must occur to complete reprogramming toward iPSCs [60, 106, 107]. Because of this low efficiency, it is difficult to obtain a pure pluripotent cell population, and cells can be isolated from the

pool of undifferentiated cells that failed to reprogram. This can be either done manually, or by sorting with the use of specific cell surface markers, such as TRA-1-60 [107-109].

The reprogramming process can cause mutations in the genome due to factor-induced reprogramming stress, also associated with oncogene expression during reprogramming [110]. These mutations can bring certain advantages for the reprogramming process and in combination with the low efficiency, this may represent a mutagenic factor, which could influence downstream experiments [101-103]. In addition, the Sendai method used here requires a replating step seven days after viral transduction. Therefore, the evaluation of the reprogramming efficiency was inhibited and may have differed between cell lines, but unable to be measured.

Although changes in the epigenome have fundamental roles during reprogramming, the decreased expression of *CREBBP* and levels of CBP did not effect the reprogramming of these cells. However, recently a study looked at epigenetic profiles, reprogramming fibroblasts taken from patients with Immunodeficiency centromeric instability and facial anomalies type I (ICF1) syndrome, caused by mutations in DNA methyltransferase (DNMT) 3B. The generated iPSCs exhibited global loss of methylation at gene promoters and enhancers compared to control iPSCs [111]. This suggests that histone acetylation in the RTS iPSCs could be abberant and should be looked at in future experiments.

Recently, there have been attempts to discover new methods and strategies to improve the efficiency of the reprogramming process. Besides different combinations of the four reprogramming factors, research has been done with additional genes that are highly expressed in ESCs, but also in attempts to discover new molecules to generate pluripotent cells, such as reprogramming enhancers [83]. In addition, the discovery that cell fate can be modulated by the addition of the four reprogramming factors and the generation of iPSCs, has lead to the development of direct reprogramming. With this technique, somatic cells can be differentiated into other cell types by the overexpression of tissue-specific transcription factors, without going through a state of pluripotency [112]. This would eliminate the inefficient process of iPSC generation, and may lead to a rapid and efficient way of generating the desired cell types.

Overlapping characteristics of ESCs and iPSCs are the expression of pluripotency genes and the ability to differentiate into cell types of all three germ layers. Teratoma

formation is the gold standard assay to test the pluripotency of human iPSCs. However, recently quantitative assays have been developed, such as the TeratoScore [113]. These algorithms use *in vivo* expression profiles of teratomas to calculate and provide each cell line with a quantitative measure for pluripotency. This array-based technique can also be used to make comparisons between expression profiles of teratomas with normal and aberrant karyotypes, which is useful in disease modeling.

It is important to note that there are subtle differences between iPSCs and ESCs, such as the somatic epigenetic memory, which can potentially affect their differentiation capacity and their use as disease models. Therefore, it is recommended when possible, to generate disease models in both ESCs and iPSCs to rule out any differences that can influence the phenotype [114]. However, for human disease-specific iPSCs it appears impossible to obtain both pluripotent cell types from the same individual.

Even though the mutation rate in human iPSCs pose a risk for their clinical use, ESCs have their own limitations. The probability cell transplants are genetically identical to the recipients is limited, which might lead to immunological reactions. Therefore, despite the fact that certain steps need optimization, the generation of iPSCs from somatic patient cells is a powerful technology.

Pluripotent stem cells have already been used for drug screening and validation, such as the transplantation of human ESC-derived retinal pigment epithelium in macular degeneration [115, 116]. Continuous developments like this will lead to 'bench to bedside' and will eventually result in patient-specific therapies using iPSCs {117-118].

Taken together, there are still some imperfections and hurdles to be taken before it is safe to use iPSCs in a clinical setting. Therefore, extensive genetic screening should become a standard procedure to ensure safety before clinical use [100]. However, the generation of patient-specific iPSCs are an ideal tool for disease modelling. Differentiation of these cells in disease specific cells, such as neurons or cardiac cells, could help in the understanding of human diseases and will be discussed in the next chapter.

References

1. Kalkhoven, E., *CBP and p300: HATs for different occasions.* Biochem Pharmacol, 2004. **68**(6): p. 1145-55.
2. Tanaka, Y., et al., *Abnormal skeletal patterning in embryos lacking a single Cbp allele: a partial similarity with Rubinstein-Taybi syndrome.* Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10215-20.
3. Partanen, A., J. Motoyama, and C.C. Hui, *Developmentally regulated expression of the transcriptional cofactors/histone acetyltransferases CBP and p300 during mouse embryogenesis.* Int J Dev Biol, 1999. **43**(6): p. 487-94.
4. Yao, T.P., et al., *Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300.* Cell, 1998. **93**(3): p. 361-72.
5. Fang, F., et al., *Coactivators p300 and CBP maintain the identity of mouse embryonic stem cells by mediating long-range chromatin structure.* Stem Cells, 2014. **32**(7): p. 1805-16.
6. Chan, W.I., et al., *The transcriptional coactivator Cbp regulates self-renewal and differentiation in adult hematopoietic stem cells.* Mol Cell Biol, 2011. **31**(24): p. 5046-60.
7. Chitilian, J.M., et al., *Critical components of the pluripotency network are targets for the p300/CBP interacting protein (p/CIP) in embryonic stem cells.* Stem Cells, 2014. **32**(1): p. 204-15.
8. Avilion, A.A., et al., *Multipotent cell lineages in early mouse development depend on SOX2 function.* Genes Dev, 2003. **17**(1): p. 126-40.
9. Chambers, I., et al., *Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells.* Cell, 2003. **113**(5): p. 643-55.
10. Hart, A.H., et al., *Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human.* Dev Dyn, 2004. **230**(1): p. 187-98.
11. Lee, J.H., S.R. Hart, and D.G. Skalnik, *Histone deacetylase activity is required for embryonic stem cell differentiation.* Genesis, 2004. **38**(1): p. 32-8.
12. Mitsui, K., et al., *The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells.* Cell, 2003. **113**(5): p. 631-42.
13. Nichols, J., et al., *Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4.* Cell, 1998. **95**(3): p. 379-91.
14. Scholer, H.R., et al., *Oct-4: a germline-specific transcription factor mapping to the mouse t-complex.* EMBO J, 1990. **9**(7): p. 2185-95.
15. Apostolou, E. and K. Hochedlinger, *Chromatin dynamics during cellular reprogramming.* Nature, 2013. **502**(7472): p. 462-71.
16. Myers, F.A., et al., *Acetylation of histone H2B mirrors that of H4 and H3 at the chicken beta-globin locus but not at housekeeping genes.* J Biol Chem, 2003. **278**(38): p. 36315-22.
17. Singh, P., et al., *Coordinated allele-specific histone acetylation at the differentially methylated regions of imprinted genes.* Nucleic Acids Res, 2010. **38**(22): p. 7974-90.
18. Efroni, S., et al., *Global transcription in pluripotent embryonic stem cells.* Cell Stem Cell, 2008. **2**(5): p. 437-47.
19. Krejci, J., et al., *Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation.* J Cell Physiol, 2009. **219**(3): p. 677-87.

20. Li, X., et al., *The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network.* Cell Stem Cell, 2012. **11**(2): p. 163-78.

21. Meshorer, E., et al., *Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells.* Dev Cell, 2006. **10**(1): p. 105-16.

22. Liang, G. and Y. Zhang, *Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective.* Cell Res, 2013. **23**(1): p. 49-69.

23. Jin, Q., et al., *Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation.* EMBO J, 2011. **30**(2): p. 249-62.

24. Kasper, L.H., et al., *CBP/p300 double null cells reveal effect of coactivator level and diversity on CREB transactivation.* EMBO J, 2010. **29**(21): p. 3660-72.

25. Akimaru, H., et al., *Drosophila CBP is a co-activator of cubitus interruptus in hedgehog signalling.* Nature, 1997. **386**(6626): p. 735-8.

26. Akimaru, H., D.X. Hou, and S. Ishii, *Drosophila CBP is required for dorsal-dependent twist gene expression.* Nat Genet, 1997. **17**(2): p. 211-4.

27. Lilja, T., et al., *The acetyltransferase activity of Drosophila CBP is dispensable for regulation of the Dpp pathway in the early embryo.* Dev Biol, 2007. **305**(2): p. 650-8.

28. Philip, P., et al., *CBP binding outside of promoters and enhancers in Drosophila melanogaster.* Epigenetics Chromatin, 2015. **8**: p. 48.

29. Waltzer, L. and M. Bienz, *Drosophila CBP represses the transcription factor TCF to antagonize Wingless signalling.* Nature, 1998. **395**(6701): p. 521-5.

30. Lopez-Atalaya, J.P., et al., *Histone acetylation deficits in lymphoblastoid cell lines from patients with Rubinstein-Taybi syndrome.* J Med Genet, 2012. **49**(1): p. 66-74.

31. Gilbert, J., *Establishment of permanent cell lines by Epstein-Barr virus transformation.* Curr Protoc Hum Genet, 2001. **Appendix 3**: p. Appendix 3H.

32. Alarcon, J.M., et al., *Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration.* Neuron, 2004. **42**(6): p. 947-59.

33. Mullighan, C.G., et al., *CREBBP mutations in relapsed acute lymphoblastic leukaemia.* Nature, 2011. **471**(7337): p. 235-9.

34. Kasper, L.H., et al., *Conditional knockout mice reveal distinct functions for the global transcriptional coactivators CBP and p300 in T-cell development.* Mol Cell Biol, 2006. **26**(3): p. 789-809.

35. Schorry, E.K., et al., *Genotype-phenotype correlations in Rubinstein-Taybi syndrome.* Am J Med Genet A, 2008. **146A**(19): p. 2512-9.

36. Strachan, T., Lindsay, S. & Wilson, D. I., *Molecular Genetics of Early Human Development* 1997: Academic Press.

37. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.

38. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.

39. Staerk, J., et al., *Reprogramming of human peripheral blood cells to induced pluripotent stem cells.* Cell Stem Cell, 2010. **7**(1): p. 20-4.

40. Aasen, T., et al., *Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes.* Nat Biotechnol, 2008. **26**(11): p. 1276-84.

41.   Avior, Y., I. Sagi, and N. Benvenisty, *Pluripotent stem cells in disease modelling and drug discovery.* Nat Rev Mol Cell Biol, 2016. **17**(3): p. 170-82.

42.   Stadtfeld, M., et al., *Induced pluripotent stem cells generated without viral integration.* Science, 2008. **322**(5903): p. 945-9.

43.   Okita, K., et al., *Generation of mouse induced pluripotent stem cells without viral vectors.* Science, 2008. **322**(5903): p. 949-53.

44.   Jia, F., et al., *A nonviral minicircle vector for deriving human iPS cells.* Nat Methods, 2010. **7**(3): p. 197-9.

45.   Kaji, K., et al., *Virus-free induction of pluripotency and subsequent excision of reprogramming factors.* Nature, 2009. **458**(7239): p. 771-5.

46.   Warren, L., et al., *Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA.* Cell Stem Cell, 2010. **7**(5): p. 618-30.

47.   Kim, D., et al., *Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins.* Cell Stem Cell, 2009. **4**(6): p. 472-6.

48.   Fusaki, N., et al., *Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome.* Proc Jpn Acad Ser B Phys Biol Sci, 2009. **85**(8): p. 348-62.

49.   Nishimura, K., et al., *Development of defective and persistent Sendai virus vector: a unique gene delivery/expression system ideal for cell reprogramming.* J Biol Chem, 2011. **286**(6): p. 4760-71.

50.   Schlaeger, T.M., et al., *A comparison of non-integrating reprogramming methods.* Nat Biotechnol, 2015. **33**(1): p. 58-63.

51.   Kim, J.J., *Applications of iPSCs in Cancer Research.* Biomark Insights, 2015. **10**(Suppl 1): p. 125-31.

52.   Chen, G., et al., *Chemically defined conditions for human iPSC derivation and culture.* Nat Methods, 2011. **8**(5): p. 424-9.

53.   Miyazaki, T., et al., *Laminin E8 fragments support efficient adhesion and expansion of dissociated human pluripotent stem cells.* Nat Commun, 2012. **3**: p. 1236.

54.   Nakagawa, M., et al., *A novel efficient feeder-free culture system for the derivation of human induced pluripotent stem cells.* Sci Rep, 2014. **4**: p. 3594.

55.   Soufi, A., G. Donahue, and K.S. Zaret, *Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome.* Cell, 2012. **151**(5): p. 994-1004.

56.   Sridharan, R., et al., *Role of the murine reprogramming factors in the induction of pluripotency.* Cell, 2009. **136**(2): p. 364-77.

57.   Buganim, Y., D.A. Faddah, and R. Jaenisch, *Mechanisms and models of somatic cell reprogramming.* Nat Rev Genet, 2013. **14**(6): p. 427-39.

58.   Hawkins, R.D., et al., *Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.* Cell Stem Cell, 2010. **6**(5): p. 479-91.

59.   Maherali, N., et al., *Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution.* Cell Stem Cell, 2007. **1**(1): p. 55-70.

60.   Mikkelsen, T.S., et al., *Dissecting direct reprogramming through integrative genomic analysis.* Nature, 2008. **454**(7200): p. 49-55.

61.   Huangfu, D., et al., *Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds.* Nat Biotechnol, 2008. **26**(7): p. 795-7.

62. Huangfu, D., et al., *Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2.* Nat Biotechnol, 2008. **26**(11): p. 1269-75.

63. Liang, G., et al., *Butyrate promotes induced pluripotent stem cell generation.* J Biol Chem, 2010. **285**(33): p. 25516-21.

64. Mali, P., et al., *Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes.* Stem Cells, 2010. **28**(4): p. 713-20.

65. Chen, X., et al., *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.

66. Morris, J.K., N.J. Wald, and H.C. Watt, *Fetal loss in Down syndrome pregnancies.* Prenat Diagn, 1999. **19**(2): p. 142-5.

67. Park, I.H., et al., *Disease-specific induced pluripotent stem cells.* Cell, 2008. **134**(5): p. 877-86.

68. Wang, X.M., et al., *Induced pluripotent stem cell models of Zellweger spectrum disorder show impaired peroxisome assembly and cell type-specific lipid abnormalities.* Stem Cell Res Ther, 2015. **6**: p. 158.

69. Adamo, A., et al., *7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages.* Nat Genet, 2015. **47**(2): p. 132-41.

70. Lee, D.F., et al., *Modeling familial cancer with induced pluripotent stem cells.* Cell, 2015. **161**(2): p. 240-54.

71. Kareta, M.S., et al., *Inhibition of pluripotency networks by the Rb tumor suppressor restricts reprogramming and tumorigenesis.* Cell Stem Cell, 2015. **16**(1): p. 39-50.

72. Krizhanovsky, V. and S.W. Lowe, *Stem cells: The promises and perils of p53.* Nature, 2009. **460**(7259): p. 1085-6.

73. Liao, J., et al., *Inhibition of PTEN tumor suppressor promotes the generation of induced pluripotent stem cells.* Mol Ther, 2013. **21**(6): p. 1242-50.

74. Banito, A. and J. Gil, *Induced pluripotent stem cells and senescence: learning the biology to improve the technology.* EMBO Rep, 2010. **11**(5): p. 353-9.

75. Banito, A., et al., *Senescence impairs successful reprogramming to pluripotent stem cells.* Genes Dev, 2009. **23**(18): p. 2134-9.

76. Ramos-Mejia, V., M.F. Fraga, and P. Menendez, *iPSCs from cancer cells: challenges and opportunities.* Trends Mol Med, 2012. **18**(5): p. 245-7.

77. Corominas-Faja, B., et al., *Nuclear reprogramming of luminal-like breast cancer cells generates Sox2-overexpressing cancer stem-like cellular states harboring transcriptional activation of the mTOR pathway.* Cell Cycle, 2013. **12**(18): p. 3109-24.

78. Lin, S.L., et al., *Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state.* RNA, 2008. **14**(10): p. 2115-24.

79. Stricker, S.H., et al., *Widespread resetting of DNA methylation in glioblastoma-initiating cells suppresses malignant cellular behavior in a lineage-dependent manner.* Genes Dev, 2013. **27**(6): p. 654-69.

80. Utikal, J., et al., *Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells.* J Cell Sci, 2009. **122**(Pt 19): p. 3502-10.

81.   Maherali, N. and K. Hochedlinger, *Guidelines and techniques for the generation of induced pluripotent stem cells.* Cell Stem Cell, 2008. **3**(6): p. 595-605.

82.   Knoepfler, P.S., *Deconstructing stem cell tumorigenicity: a roadmap to safe regenerative medicine.* Stem Cells, 2009. **27**(5): p. 1050-6.

83.   Takahashi, K. and S. Yamanaka, *A decade of transcription factor-mediated reprogramming to pluripotency.* Nat Rev Mol Cell Biol, 2016. **17**(3): p. 183-93.

84.   Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.

85.   Miura, K., et al., *Variation in the safety of induced pluripotent stem cell lines.* Nat Biotechnol, 2009. **27**(8): p. 743-5.

86.   Bar-Nur, O., et al., *Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells.* Cell Stem Cell, 2011. **9**(1): p. 17-23.

87.   Ohi, Y., et al., *Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells.* Nat Cell Biol, 2011. **13**(5): p. 541-9.

88.   Lister, R., et al., *Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.* Nature, 2011. **471**(7336): p. 68-73.

89.   Ruiz, S., et al., *Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells.* Proc Natl Acad Sci U S A, 2012. **109**(40): p. 16196-201.

90.   Weissbein, U., N. Benvenisty, and U. Ben-David, *Quality control: Genome maintenance in pluripotent stem cells.* J Cell Biol, 2014. **204**(2): p. 153-63.

91.   Guenther, M.G., et al., *Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells.* Cell Stem Cell, 2010. **7**(2): p. 249-57.

92.   Newman, A.M. and J.B. Cooper, *Lab-specific gene expression signatures in pluripotent stem cells.* Cell Stem Cell, 2010. **7**(2): p. 258-62.

93.   Ghosh, Z., et al., *Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells.* PLoS One, 2010. **5**(2): p. e8975.

94.   Chin, M.H., et al., *Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures.* Cell Stem Cell, 2009. **5**(1): p. 111-23.

95.   Kim, K., et al., *Epigenetic memory in induced pluripotent stem cells.* Nature, 2010. **467**(7313): p. 285-90.

96.   Polo, J.M., et al., *Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells.* Nat Biotechnol, 2010. **28**(8): p. 848-55.

97.   Mayshar, Y., et al., *Identification and classification of chromosomal aberrations in human induced pluripotent stem cells.* Cell Stem Cell, 2010. **7**(4): p. 521-31.

98.   Laurent, L.C., et al., *Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture.* Cell Stem Cell, 2011. **8**(1): p. 106-18.

99.   Hussein, S.M., et al., *Copy number variation and selection during reprogramming to pluripotency.* Nature, 2011. **471**(7336): p. 58-62.

100. Gore, A., et al., *Somatic coding mutations in human induced pluripotent stem cells.* Nature, 2011. **471**(7336): p. 63-7.

101. Imreh, M.P., et al., *In vitro culture conditions favoring selection of chromosomal abnormalities in human ES cells.* J Cell Biochem, 2006. **99**(2): p. 508-16.

102. Draper, J.S., et al., *Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells.* Nat Biotechnol, 2004. **22**(1): p. 53-4.

103. Baker, D.E., et al., *Adaptation to culture of human embryonic stem cells and oncogenesis in vivo.* Nat Biotechnol, 2007. **25**(2): p. 207-15.

104. Staunstrup, N.H., et al., *Development of transgenic cloned pig models of skin inflammation by DNA transposon-directed ectopic expression of human beta1 and alpha2 integrin.* PLoS One, 2012. **7**(5): p. e36658.

104. Trokovic, R., et al., *Combined negative effect if donor age and time in culture on the reprogramming efficiency into induced pluripotent stem cells.* Stem Cell Res, 2015. **15**(1): p. 254-62.

105. Samavarchi-Tehrani, P., et al., *Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming.* Cell Stem Cell, 2010. **7**(1): p. 64-77.

106. Deng, J., et al., *Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming.* Nat Biotechnol, 2009. **27**(4): p. 353-60.

107. Polo, J.M., et al., *A molecular roadmap of reprogramming somatic cells into iPS cells.* Cell, 2012. **151**(7): p. 1617-32.

108. Chan, E.M., et al., *Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells.* Nat Biotechnol, 2009. **27**(11): p. 1033-7.

109. O'Malley, J., et al., *High-resolution analysis with novel cell-surface markers identifies routes to iPS cells.* Nature, 2013. **499**(7456): p. 88-91.

110. Ji, J., et al., *Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells.* Stem Cells, 2012. **30**(3): p. 435-40.

111. Huang, K., et al., *Selective demethylation and altered gene expression are associated with ICF syndrome in human-induced pluripotent stem cells and mesenchymal stem cells.* Hum Mol Genet, 2014. **23**(24): p 6448-57.

112. Baker, D.E., et al., *Adaptation to culture of human embryonic stem cells and oncogenesis in vivo.* Nat Biotechnol, 2007. **25**(2): p. 207-15.

113. Nakagawa, M., et al., *Promotion of direct reprogramming by transformation-deficient Myc.* Proc Natl Acad Sci U S A, 2010. **107**(32): p. 14152-7.

114. Avior, Y., J.C. Biancotti, and N. Benvenisty, *TeratoScore: Assessing the Differentiation Potential of Human Pluripotent Stem Cells by Quantitative Expression Analysis of Teratomas.* Stem Cell Reports, 2015. **4**(6): p. 967-74.

115. Halevy, T. and A. Urbach, *Comparing ESC and iPSC-Based Models for Human Genetic Disorders.* J Clin Med, 2014. **3**(4): p. 1146-62.

116. Schwartz, S.D., et al., *Embryonic stem cell trials for macular degeneration: a preliminary report.* Lancet, 2012. **379**(9817): p. 713-20.

117. Trounson, A. and N.D. DeWitt, *Pluripotent stem cells progressing to the clinic.* Nat Rev Mol Cell Biol, 2016. **17**(3): p. 194-200.

118. Warren, L., et al., *Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA.* Cell Stem Cell, 2010. **7**(5): p. 618-30.

CHAPTER 3 – Aim 2 Differentiation of RTS induced Pluripotent Stem Cells into disease-specific cells of the neuronal lineage, creating a disease in a dish model

Introduction

Roughly 3% of the general population has intellectual disability, which is characterized by significant limitations in adaptive behaviour and intellectual functioning (Intelligence Quotient (IQ) below 70) [1]. These developmental disorders, such as RTS, are highly heterogeneous and can be combined with congenital abnormalities, such as skeletal defects and organ malformations.

RTS patients are born with a moderate to severe intellectual disability, with IQ ranging from <25 to 79, and it is unclear whether or not this progresses through life. Some RTS cases experience seizures and/or epilepsy, and entering adulthood there can be significant mood swings and behavioural problems. RTS has also been associated with certain solid cancers of neural and developmental origin, including medulloblastoma, neuroblastoma and nasopharyngeal rhabdomyosarcoma [2]. Although it has been reported that there is an increased risk for these cancers, the rarity of RTS makes it difficult to determine whether there is an actual susceptibility for tumour formation.

More importantly, there is limited knowledge about the neurological background of RTS patients, besides structural anomalies such as postnatal microcephaly, Chiari malformation type I, thinning of the corpus callosum and delayed myelination [3, 4]. In 1995, a cranial magnetic resonance imaging (MRI) study revealed abnormal cortical infoldings in the rolandic regions and diminishing of the white matter, which might underlie the intellectual disability seen in RTS [5]. However, there are no significant differences in the delayed development and/or cognitive impairment seen between RTS cases with and without normal brain images. These MRI abnormalities are more frequently observed in patients with a *CREBBP* mutation compared to *EP300*, which overall, show a less severe RTS phenotype. Even though there is no obvious correlation between the type of mutation and the degree of intellectual disability, there is a trend seen toward lower IQ and autistic features in cases with large deletions in *CREBBP* [6].

Developmental processes require continuous control and regulation to maintain specific programs and this is achieved through epigenetic regulators such as transcription factors and histone modifiers [7, 8]. Disturbed epigenetic control plays a role in human disease, but their involvement in neurological disorders remain relatively unknown due to the complexity of the human brain. From this viewpoint,

RTS can provide insights into the role of chromatin modifiers in neurogenesis. Most research into RTS has been performed with animal studies, and from these it has been suggested that the reduced levels of CBP or p300, which play a critical role in memory, addiction, neurogenesis and other forms of neuroplasticity [9, 10] might be responsible for the neurological phenotypes seen in RTS.

There are several different mouse models in which Cbp or p300 function is altered by complete or partial gene knockout, and these mice show certain similarities to the RTS phenotype. A study with *ep300*[-/-] mouse embryos showed that the density of mesenchymal cells in the hindbrain and forebrain was reduced in the homozygous knockout embryo compared to the wild type. These homozygous mice also showed defective neuronal tube closure, and died at E10 [11]. Studies have also shown that adult mouse brains contain the highest p300 HAT activity compared to other tissues, suggesting this protein may play an important role in gene regulation in the brain [12]. Indeed, *in vitro* studies have shown that p300 interacts with transcription factors that are known to play a role in learning and memory [13-16]. Long-term memory, in contrast to short-term memory, depends mainly on gene expression, which is regulated by chromatin remodelling, induced by the HAT activity of CBP/p300 [17]. This was seen in the lower acetylation levels in hippocampal tissue extracts of homozygous conditional *Crebbp* knockout mice compared to WT. These mutant mice showed deficits in synaptic plasticity and memory that is linked to aberrant histone 2B (H2B) acetylation [18]. It also showed that lacking the HAT domain will lead to impaired hippocampal long-term potential, implying that CBP is involved in long-term memory formation [19]. In addition, mouse studies revealed that decreased CBP levels in neuronal progenitor cells (NPCs), lead to reduced differentiation of these progenitors into interneurons in the cortex [20]. In conclusion, these observations in mouse studies imply possible explanations and mechanisms for the cognitive abnormalities seen in RTS but also for the epilepsy diagnosed in some RTS cases [21].

In addition, these RTS models revealed that, besides being important for the genesis of interneurons, CBP also promotes differentiation of embryonic cortical precursors by modifying chromatin at the promoters of neural genes and thereby enhancing transcription of those genes. In this mouse model, CBP was haploinsufficient, which caused decreased generation of both glia and cortical neurons, and showed behavioural abnormalities during early postnatal life [20].

It is clear that epigenetics not only play an important role in reprogramming but also in cell differentiation processes. When cells transit from a pluripotent to a differentiated state, there is an increase in DNA methylation, repressive histone marks and chromatin compaction [22, 23]. These epigenetic changes are essential to acquire a specific cell identity [24], such as neuronal phenotypes [25, 26] and chromatin remodeling and transition of histone marks is crucial during brain development [27-29].

Even though these animal models have made tremendous contributions, deeper insights into epigenetic patterns and global transcriptome alterations of the developing human brain remain poorly explored. However, Spiers et al. quantified genome-wide patterns of DNA methylation in human fetal brain samples (spanning 23 to 184 days post-conception) and showed significant changes in DNA methylation across fetal brain development, with an enrichment of loci becoming hypomethylated with fetal age [30].

A major challenge of neurological disorders has been the inaccessibility of human neuronal cells targeted by disease [31]. These are usually available only in post-mortem state and in combination with the rarity of RTS, obtaining neuronal cells from RTS patients is a nearly unfeasible task. Although both animal studies and transformed cell lines have contributed, they are often not representative due to the complex neurological structures seen in humans.

However, the process of reprogramming somatic cells into pluripotent cells has contributed to overcome the obstacles of collecting patient-specific disease cells. The generation of patient-specific iPSCs in combination with differentiation protocols gives insights into the pathological background and this has been done for both neurodegenerative diseases such as Huntington's Disease (HD) [32] but also neurodevelopmental diseases like Timothy Syndrome (TS) [33]. Especially in neurodevelopmental disorders, iPSC-based models can mimic the early steps of neuronal differentiation and can contribute in studying cellular and molecular mechanisms of these disorders. In addition, iPSC models are ideal for drug development due to the self-renewal capacity and differentiation properties into disease-specific cell types. Epigenetic drugs are currently already being explored in human disease models such as DNA methylation inhibitors, histone acetylation activators and histone deacetylation (HDAC) inhibitors [34].

Both HATs and HDACs are known to play important roles in the regulation of a variety of biological processes, including cell cycle progression, differentiation and development [35]. The balance between histone acetylation and deacetylation is crucial for correct gene regulation for cell survival and homeostasis, whereas imbalances are related to pathological conditions, such as RTS [36, 37].

In contrast to genetic mutations, epigenetic modifications are stable but reversible. Therefore, targeting and modulating these epigenetic components might be a promising therapeutic in syndromes like RTS. In neurodegenerative disorders, reports have suggested that the deregulation of histone acetylation levels can be modulated with epigenetic drugs such as HDAC inhibitors [38-40]. These inhibitors, such as valproic acid, have a long history of use in psychiatry and neurology as mood stabilizers and anti-epileptics [41]. More recently they are being investigated as possible treatments for cancers and inflammatory diseases [42, 43].

HDAC inhibitors induce hyperacetylation of both histone and non-histone HDAC targets, which will have an effect on a range of cellular and molecular aspects. It is suggested that this lead to the anti-tumour, immunological and neurological responses observed in both experimental and clinical settings [44]. The reaction to HDAC inhibition is specific and likely to be determined by a combination of factors such as cell of origin and the genetic and epigenetic profiles present in the cell [45].

To date, there is one study in literature looking at acetylation levels in cells from RTS patients. Lopez-Atalaya et al. [46] generated lymphoblastoid cell lines out of peripheral blood mononuclear cells from RTS patients with mutations in *CREBBP*. They showed that histone acetylation was primarily affecting H2A and H2B and this deficit was rescued by treatment with the HDAC inhibitor Trichostatin A (TSA). Importantly, they also showed that TSA did not affect cell viability. HDAC inhibitors have also been used in several RTS mouse models and treatment of *Crebbp*-knockout mice, increased the acetylation levels of hippocampal extracts, improving the long-term potential defect [19, 47-50]. In addition, epigenetic drugs modulating histone acetylation activity have been shown to alleviate pathological symptoms in experimental models of other neurological diseases, such as Parkinson's, Alzheimer's and Huntington's diseases, by reverting abnormal gene repression associated with disease [51-54].

There are several different HDAC inhibitors and Panobinostat (LBH589), which is a non-selective HDAC inhibitor, has been tested as a drug for oncological,

inflammatory and viral diseases [55-59], and has been approved by the U.S. Food and Drug Administration for the treatment of patients with multiple myeloma and cutaneous T-cell lymphoma (CTCL) [60, 61]. This compound has potent inhibitory activity at low concentrations compared to other HDAC inhibitors, which makes it clinically achievable to use as a potential drug [61].

The generation and differentiation of patient-specific iPSCs from generally easy accessible somatic cells, such as skin fibroblasts or blood cells, into disease specific cells will give a 'disease in a dish' model. This is an excellent approach to study specific aspects of disease, such as general cell biology, and additionally in combination with drug screening. This has been done for several diseases, such as Timothy Syndrome (TS) [62], which is a monogenic condition associated with developmental delay and autism spectrum disorder. iPSCs generated from individuals with TS were differentiated into cortical neuronal precursor cells and neurons and these cells show a TS phenotype with defects in calcium signalling and activity-dependent gene expression [33], making them an excellent model to study disease.

This chapter will focus on the differentiation of the generated RTS-iPSCs into neuronal progenitor cells. Neuronal cell structures and compositions will be studied, along with the effect of the HDAC inhibitor Panobinostat (LBH589) during differentiation. This compound could potentially be used in the treatment of RTS, with the intention of reducing certain neuronal symptoms seen in RTS, such as epilepsy, mood swings and/or behavioural problems.

Material & Methods

Western Blot

The following primary antibodies were used: CBP (SantaCruz, sczsc-369), H3K9ac (Abcam, ab4441), H3K18ac (Abcam, ab1191), H3K27ac (Abcam, ab4729), H3 (Abcam, ab12079), and Actin (Abcam, ab8227) at 1:1,000. For secondary antibodies, donkey anti-rabbit IgG 680 (Li-Cor Biosciences 926-32223) and donkey anti-goat IgG 800 (Li-Cor Biosciences, 925-32214) were used at 1:10,000. The membrane was scanned on the Odyssey Imaging System CLx using the Odyssey Application Software. Band intensities were calculated using ImageJ.

iPSC differentiation into neuronal progenitor cells (NPCs)

This protocol was adapted from Espundy-Camacho [63]. On day -2, cells were dissociated using EDTA and plated onto matrigel-coated plates at low confluence (5,000-10,000 cells/cm^2) in E8 medium supplemented with 1000X ROCKi. On day 0, the medium was changed to DDM (DMEM/F12 + GlutaMax supplemented with 1% N2 supplement, 0.1 mM NEAA, 1 mM sodium pyruvate, 500 ug ml-1 BSA, 0.1 mM 2-mercaptoethanol and 50 U ml-1 P/S) supplemented with 2% B27 (without vitamin A) and 100 ng/ml LDN193189. This medium was replenished every second day. On day 16, the medium was changed to DDM supplemented with B27 (without LDN193189) and changed every second day. On day 24, the progenitors were dissociated with Trypsin and plated into wells coated with Poly-D-lysine/Laminin (33.3 ug/ml and 3.3 ug/ml). Cells were cultured with DDM, supplemented with B27 and ROCKi for the first 24 hours and continued for another 7 days without ROCKi.

HDAC inhibitor

The HDAC inhibitor Panobinostat (LBH589), developed by Novartis Pharmaceuticals (Basel, Switzerland), was provided by the laboratory of Professor Neil Watkins, The Hudson Institute of Medical Research, Melbourne, Australia. The inhibitor was prepared in 100% DMSO at a stock concentration of 5mM and added to fresh DDM medium at concentrations ranging from 0.2-50nM diluted in DMSO.

Immunostaining

On day 31 NPCs were fixed in 4% paraformaldehyde for 10 minutes and permeabilized with 0.3% Triton X-100 in PBS for 30 minutes. Cells were stained with Tubulin β 3 (TUJ1) 1:1000 (BioLegend, 801202), SOX1 1:500 (R&D Systems, AF3369), MAP2 1:1000 (Merck Millipore, MAB3418 or Abcam, ab5392), GFAP 1:500 (Abcam, ab7260), NeuN 1:500 (Merck Millipore, MAB377) and 4',6-diamidino-2-phenylindole (DAPI) 1:1000. To calculate percentages of total cells, four images per well were taken and image quantification was analyzed using ImageJ, Version 2.0.0-rc-43/1.5. A minimum of 1,000 DAPI$^+$ cells per sample were scored and relative amounts of positive cells were calculated as a percentage of total DAPI$^+$ cells. All statistic analysis was performed in GraphPad Prism, using Two-way ANOVA or One-way ANOVA (Figure 13) with Bonferroni's multiple comparisons test.

shRNA

Lentiviral Expression Vectors (pZIP-hEF1a-ZsGreen) containing a GFP reporter, were obtained from Transomic (Integrated Sciences, TLHSU1400-1387) and contained three vectors with shRNA targeted to different regions in the human *CREBBP* gene (ULTRA-3214088, ULTRA-3214091 and ULTRA-3214092) including a non-targeting shRNA control vector (Table 1). Vectors were kept on ice and a fraction was scraped off and plated on agar-ampicillin plates. Cultures were grown overnight at 37°C and colonies were picked the next day. Clones were grown in LB medium with 50μg/ml ampicillin (total volume of 15ml per clone), overnight at 37°C with constant shaking. Plasmids were purified using GenElute HP Plasmid Midiprep Kit (Sigma-Aldrich, NA0200) and concentrations were measured using a nanodrop and stored at -20°C.

Table 1. shRNA vectors for *CREBBP*

Lentiviral expression vector pZIP-hEF1a-ZsGreen	
Clone ID	
ULTRA-3214088	*CREBBP* shRNA 1
ULTRA-3214091	*CREBBP* shRNA 2
ULTRA-3214092	*CREBBP* shRNA 3
Non-targeting shRNA control	Control shRNA

Viral particle production

One day prior to transfection, 293T HEK cells were seeded a $1x10^5$ cells/cm^2 in the absence of antibiotics. Immediately before transfection, the serum-containing medium was changed to un-supplemented, serum free Advanced Dulbecco's Modified Eagle Medium. Preparation of the lipofectamine-DNA complexes for one T175 vessel went as follows: in 3ml of OptiMEM medium, transfer vector (13.5μg), packaging plasmids (Tronolab, Switzerland) psPax2 (8.4μg) and pMD2G (5.4μg) were combined with 30μl of Lipofectamine PLUS component. After incubation for 5 minutes, 60μl of Lipofectamine LTX was added and mixed. Complexes were allowed to form (at room temperature) for 30 minutes and then added to the T175 vessel. Five hours after transfection, the serum free medium was exchanged with supplemented Advanced Dulbecco's Modified Eagle Medium containing 2% FBS, non essential amino acids, 2mM L-Glutamine and 30μM cholesterol (Sigma-Aldrich). Viral supplements were harvested after 24 and 36 hours and frozen down at -80°C.

Concentration of primary viral supernatants

Viral supernatant (~40ml/plasmid) was concentrated using Amicon Ultra 15 centrifugal filters (Millipore, UFC910096). Tubes were centrifuged at 4000rpm down to ~250μl per plasmid and stored in aliquots at -80°C until used.

Titer determination of viral concentrates

WT-HDFa iPSCs (clone 1p16) were plated in a 24 well plate at $7x10^4$ cells/well in E8 supplemented with ROCKi (1000x). The next day virus was added in different concentration (1:100, 1:1000, 1:10,000 and 1:100,000) plus a negative control in E8 medium with polybrene (1:1700). Plates were spun at 1900rpm for 1 hour at room temperature and incubated at 37°C for 24 hours before medium was changed to E8.

FACS analysis to determine efficiency/MOI

Cells were harvested on day 4 with trypsin and stained with PI (1:500) and DAPI (1:1000) in PBS, 2% FBS. Samples were analysed on the LSR II a.

shRNA knockdown of CREBBP

WT-HDFa 1p16 iPSCs were plated at $7.5x10^5$ cells per 6 well and cultured in E8 medium supplemented with ROCKi. Two days after plating, a control well was

counted and the viruses were added at a MOI of 10 in 1.5ml E8 medium with polybrene (1:1700). Plates were centrifuged at 1900rpm for 1 hour at room temperature and incubated at 37°C for 24 hours before medium is changed to E8.

RNA extraction and cDNA

RNA was extracted using the RNeasy Mini Kit (Qiagen, 74104) on the QIAcube. RNA was quantified and SuperScript III First-Strand Synthesis System (Invitrogen, 18080051) was used to produce cDNA.

CREBBP expression

Quantitative PCR (qPCR) was used to look at *CREBBP* expression in the iPSC lines. Primers used were (A) *CREBBP* Exon 2 5' GTGCTGGCTGAGACCCTAAC 3' and reverse 5' GGCTGTCCAAATGGACTTGT 3'; and (B) *CREBBP* exon 30 Forward 5' GACCGCTTTGTTTATACCTGC 3' and reverse 5' TCTTATGGGTGTGGCTCTTTG 3'. Samples were run in triplicate, with each reaction containing 3.5µl 2X SYBR Green Master Mix (Roche Applied Science, 04707516001), 0.35µl of each primer (10µM), 1.4µl H2O, 1.4µl cDNA (Total volume 7µl). 384 well plates were run on a LightCycler 480 system (Roche Applied Science), with the following cycling conditions: 94°C for 30s; 60°C 30s and 72°C for 30s for 45 cycles; followed by 72°C for 10min. Gene expression was standardized to housekeeper gene *GAPDH* (5' TGAAGGTCGGAGTCAACGGA 3' and 5' CCATTGATGACAAGCTTCCCG 3'). Statistics were performed in GraphPad Prism using the one-way ANOVA with Dunnett's multiple comparisons test.

Results

Differentiation

In normal neuronal development, neural stem cells (NSCs), which are self-renewing cells, will differentiate into neural progenitors cells (NPCs). These cells eventually give rise to neurons, as well as glia cells (astrocytes and oligodendrocytes) of the nervous system. An established protocol by Espuny-Camacho et al. [63] was followed, which uses DDM medium [64] (Figure 1). This medium was adjusted and Noggin was replaced by LDN193189, which are both signalling molecules required for correct nervous system development and bone morphogenetic protein (BMP) inhibitors [65]. This is crucial in neurogenesis, and *in vitro* it has been shown that Noggin will induce differentiation of human iPSCs into NPCs, that will convert to a cortical identity [63].

Firstly, the generated RTS and WT iPSC lines from Chapter 2 (HDFa 1p13 and 7p11, RTS-NL1 2p13 and 8p10, RTS-NL2 3p12 and 6p11) were differentiated into NPCs, to investigate if reduced levels of CBP have an effect on differentiation potential. Looking at the WT and RTS differentiation potential of the iPSCs under a light microscope during day 0 till 24 of differentiation (before replating), a clear difference was not detected due to the confluence of the wells. However, a staining was performed, using markers representing different stages of neurogenesis (Figure 2). SOX1 was used as a marker for NPCs, and this marker should decrease during the differentiation of NPCs towards neurons [66]. TUJ1, which is a tubulin marker for neuronal axons and is present in newly generated immature postmitotic neurons [67], was used to look at more mature neurons. Other mature markers stained for include GFAP, which is an astrocyte marker in differentiated cell cultures, and MAP2, (microtubule associated protein) which is a neuron specific protein that promotes assembly and stability of the microtubule network, located in the cell soma, whereas TUJ1 is found in the dendrites [68]. As seen in Figure 2, SOX1 expression was similar between WT and RTS lines on day 8, suggesting that RTS lines can differentiate into NSCs like the WT lines. However, differences in TUJ1[+] cells where observed, with RTS cell lines showing reduced numbers of cells positive for this marker during differentiation compared to WT on day 24 of differentiation.
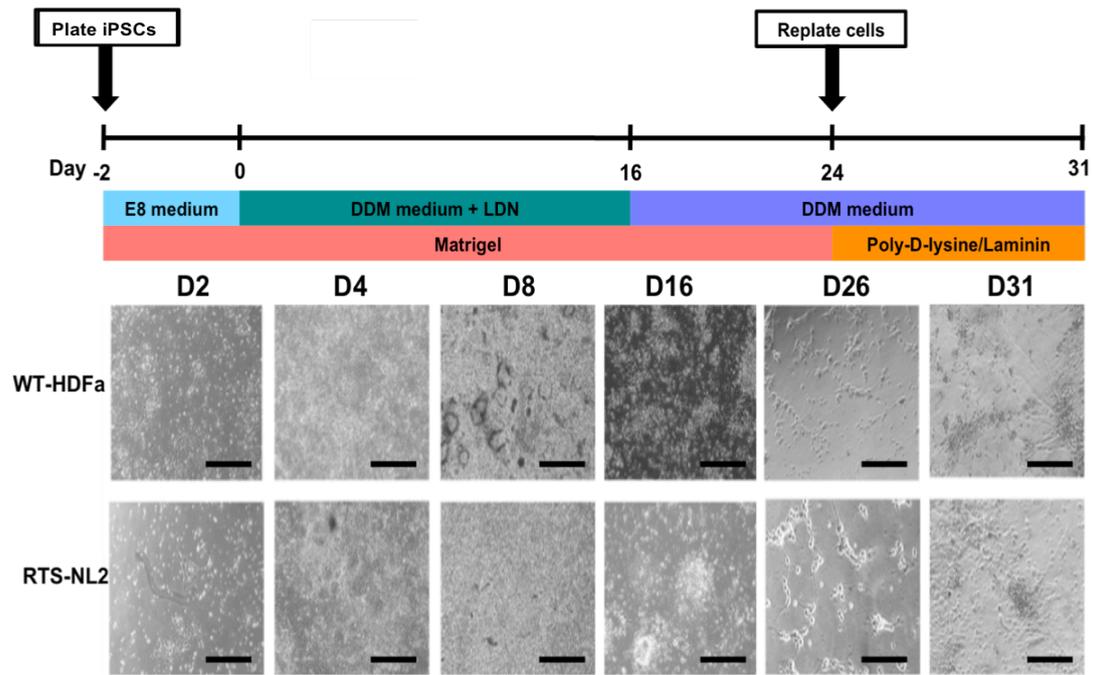
Figure 1. Timeline for neuronal differentiation. iPSCs were plated on Matrigel and grown in DDM medium supplemented with LDN193189 for 16 days. On day 24 of differentiation, cells were harvested and replated on Poly-Lysine/Laminin coated plates. Cells were grown for a further seven days (31 days in total) of differentiation. Magnification 10X, scale bars 100 μm, n = 2 clones per cell line, 3 biological replicates per clone.
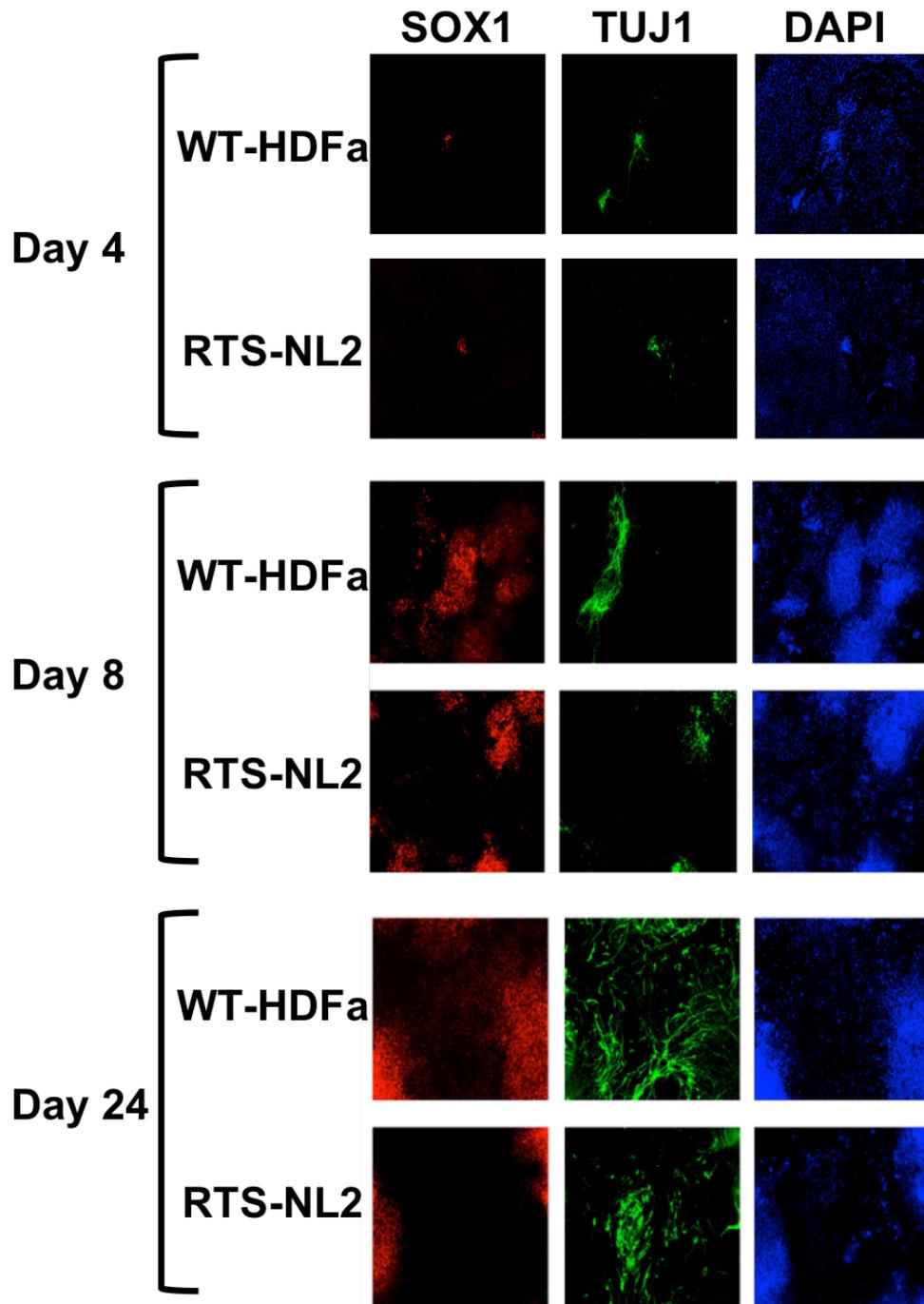
Figure 2. Neuronal differentiation potential of RTS iPSCs. These images show immunofluorescence staining during the neuronal differentiation process of WT and RTS iPSCs. Stainings were performed on days 4, 8 and 24 of differentiation for SOX1 (red) and TUJ1 (green). DAPI is shown in blue. Magnification 10X, n = 2 clones per cell line, 3 biological replicates per clone.

In order to determine if these results were significant, the number of SOX1 and TUJ1 positive cells per well were determined using ImageJ and are shown in Figure 3 for all time points as a percentage of DAPI$^+$ cells.

When quantifying, there was a clear pattern seen in SOX1 and TUJ1 expression for WT and RTS iPSCs during differentiation. SOX1 expression increases early during differentiation (from day 8 onwards) and decreases after day 24. TUJ1 is a more mature marker for neurogenesis, and expression increases after 16 days of differentiation and a significant difference on day 24 between WT and both RTS cell lines is seen. This significance was lost on day 31, even though there was still a clear increase in TUJ1 levels seen compared to earlier days. This could be explained by the fact the cells were dissociated and replated on day 24, and therefore needed longer to recover before a significant difference between WT and RTS cells is shown again.

In conclusion, these results could imply that even though NPC formation (SOX1) is normal in RTS cells, the ability to mature (TUJ1) is impaired which could correspond to the neurological abnormalities seen in these patients.
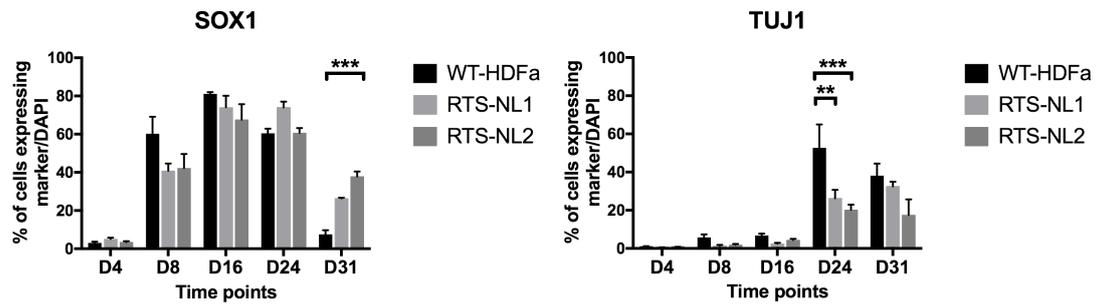
Figure 3. Quantification of neuronal differentiation markers. Stainings were quantified at different time points during neuronal differentiation. An increase of SOX1 during differentiation was observed, which reduced on day 31. There is a significant difference between WT (HDFa) and RTS-NL2 for SOX1 on day 31. TUJ1[+] cells increased in both WT and RTS cells later in the differentiation process, and there is a significant difference on day 24 between WT and both RTS lines. Mean and SEM were plotted with comparison by Two-way ANOVA with Bonferroni's multiple comparisons test, n = 3 biological replicates, ** $p<0.01$, *** $p<0.001$.

HDAC inhibitor

Next it was investigated if the addition of an HDAC inhibitor has an effect on acetylation levels in RTS patients and can rescue the observed phenotype. This has been shown for lymphoblastoid cell lines from RTS patients treated with the HDAC inhibitor TSA [46]. LBH589 was used in this study, which is a non-selective hydroxamic acid derived HDAC inhibitor [61].

In order to test if the acetylation pattern of histones can be modulated and recovered in RTS cells, the HDAC inhibitor was first tested in fibroblasts. The inhibitor was added to both RTS and WT fibroblast cell lines at a final concentration of 500nM to look at the molecular effect of the HDAC inhibitor. A study with fibroblasts from spinal muscular atrophy (SMA) patients showed that LBH589 induces up regulation of the *SMN* gene through H3K9 hyperacetylation of the promoter at nanomolar doses, which makes it a highly promising candidate for SMA therapy. The optimal expression was observed at a concentration of 500nM when incubated for 48 hours, and cell viability did not seem to be affected at concentrations of up to 10μM when incubated for 96 hours [70].

Fibroblasts were plated at 5×10^6 cells in T75 and on day 2 either the HDAC inhibitor at a final concentration of 500nM or dimethylsulfoxide (DMSO) as a control was added to the cells. Cells were harvested after 48 hours of growth, and fibroblasts did not show differences in morphology or cell viability (Figure 4).
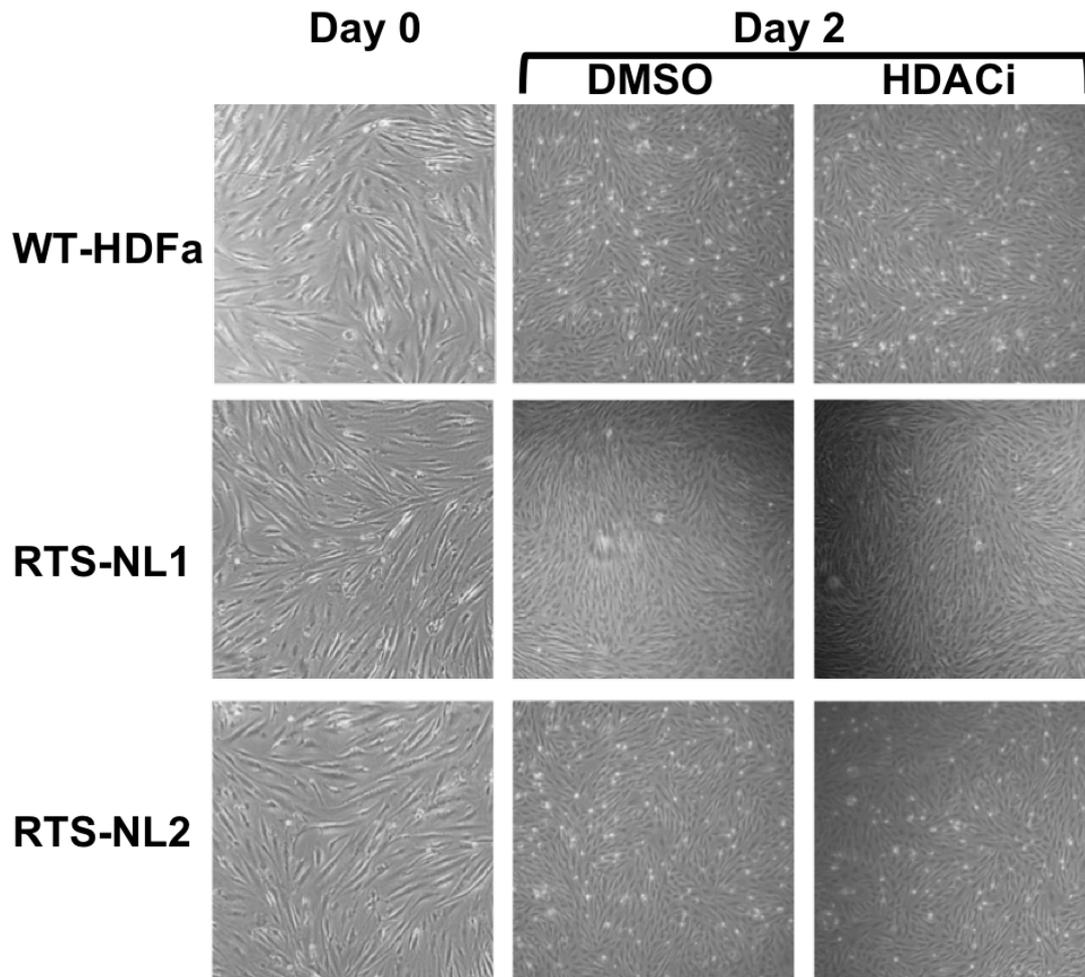
Figure 4. Fibroblasts cultured with HDAC inhibitor for 48 hours. Pictures were taken with a light microscope on day 0 (10X) and day 2 (4X) after addition of either DMSO as a control, or 500nM HDAC inhibitor (4X). No difference in morphology was observed between WT or RTS fibroblasts, with or without treatment. Magnification 10X, n = 3 biological replicates.

Next, whole cell extracts for the fibroblasts were prepared, and Western Blot analysis looking at H3K9ac, H3K18ac and H3K27ac was performed (Figure 5) to explore what effect an HDAC inhibitor has on histone acetylation. After quantification with Image J, most histone marks show an increase in acetylation (Figure 5B). Especially for RTS-NL2, a clear increase was seen, with acetyation levels for H3K9 and H3K18 close to normal control levels. This could imply that an HDAC inhibitor could potentially lead to hyperacetylation in RTS cells, bringing the levels closer to WT. There was no clear difference between WT and RTS-NL1 and this could potentially be explained by the location of the mutation. Whereas the mutation for RTS-NL1 is located at the end of the protein (exon 29), for RTS-NL2 the variant is at the start of the gene, in exon 2. This could mean that for RTS-NL1, there is a partially functional protein from the mutated allele, which is also shown in the amount of CBP. Therefore, the RTS phenotype in this patient could be due to deficiencies other than HAT function of CBP.
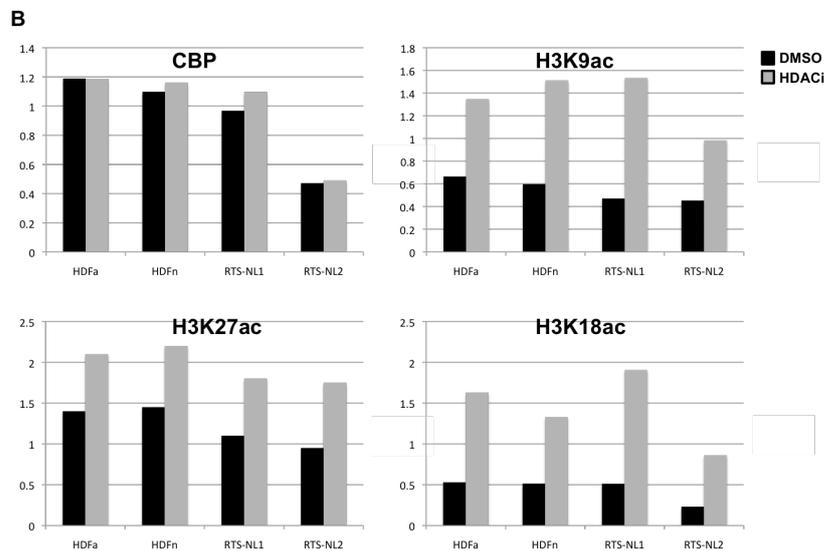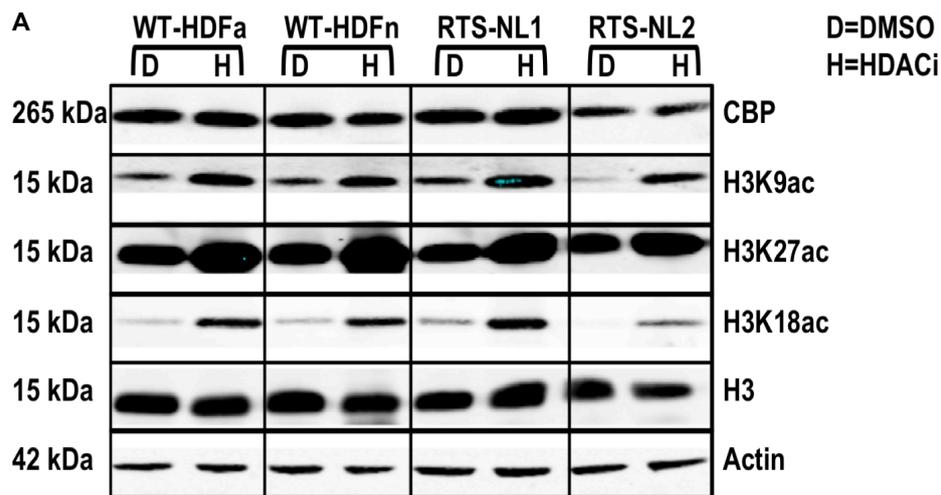
Figure 5. Western Blot for histone acetylation. Whole cell lysates were extracted from WT and RTS fibroblasts after incubation with either DMSO (control) or an HDAC inhibitor (500nM) for two days. A Western Blot was performed (A) and showed that addition of the inhibitor leads to an increase in histone acetylation levels. When quantifying the data, there is a difference seen in total amount of CBP between WT and RTS-NL2, which is not seen in RTS-NL1 (B). Histone 3 (H3) total protein and Actin were taken as controls, n = 1. Band intensities were calculated using ImageJ.

Next, the HDAC inhibitor was added during neuronal differentiation to see what effect this has on differentiation, and if this would improve the neuronal differentiation process of RTS iPSCs. The HDAC inhibitor or DMSO as a control was added to the cells at concentrations of 0.2, 1, 5, 10, and 100nM. It was noted that the cells started dying at concentrations of 10nM and higher. Therefore, it was decided to perform all subsequent experiments at a concentration of 5nM.

Figure 6 shows images taken with a light microscope and again, any clear differences in differentiation potential or morphology between WT and RTS lines, with or without HDAC inhibitor treatment was not detected. However, when stained for different neuronal markers, differences in differentiation potential were observed (Figure 7). Quantification of this data confirmed a significant increase in TUJ1 positive cells when the inhibitor is added to RTS cells, with levels almost similar to WT on day 24 (Figure 8). This indicates that an HDAC inhibitor might rescue the phenotype of delayed neuronal maturation seen in the RTS cell lines. Note that even though a similar trend is observed, percentages are lower compared to day 24 in Figure 3. This can be explained by technical variation due to a significant time lapse between experiments and thawing/freezing cycles of the cells.

Next, the cells were cultured for a further two weeks (d45) to see if this rescued phenotype was stable over time, (Figure 9). The HDAC inhibitor was removed on 31 of differentiating and cells were fixed and stained on day 45. This showed that the rescued phenotype remained stable over time (Figure 9). Cells were also stained for MAP2, which is a more mature marker and essential in neurogenesis. Since MAP2 comes up later during neuronal differentiation, it should show similar results compared to TUJ1, although slightly lower. When quantifying this data, there was a significant difference seen for TUJ1 in the RTS cell lines with the HDAC inhibitor. The same trend, although not statistically significant, was observed for MAP2 (Figure 10). Quantifying the immunofluorescence data when removing the HDAC inhibitor for 2 weeks (day 31-45) showed that the amounts of TUJ1 and MAP2 stay stable over time. Again, this is statistically significant for TUJ1, with the same, non-significant, trend observed for MAP2 (Figure 11). MAP2 shows slightly less positive numbers of cells compared to TUJ1 and this can be explained by the fact that MAP2 is a more mature marker and will come up later during differentiation compared to TUJ1 [69].

In conclusion, the addition of an HDAC inhibitor to RTS cells during differentiation increases the level of TUJ1 positive cells and seems to rescue the RTS deficient

neurodifferentation phenotype. It was also shown that removing the HDAC inhibitor after the TUJ1 cells have formed did not have an effect on the stability and numbers of TUJ1$^+$ cells. These results indicate that an HDAC inhibitor could potentially be used as a drug to ameliorate RTS symptoms in future.
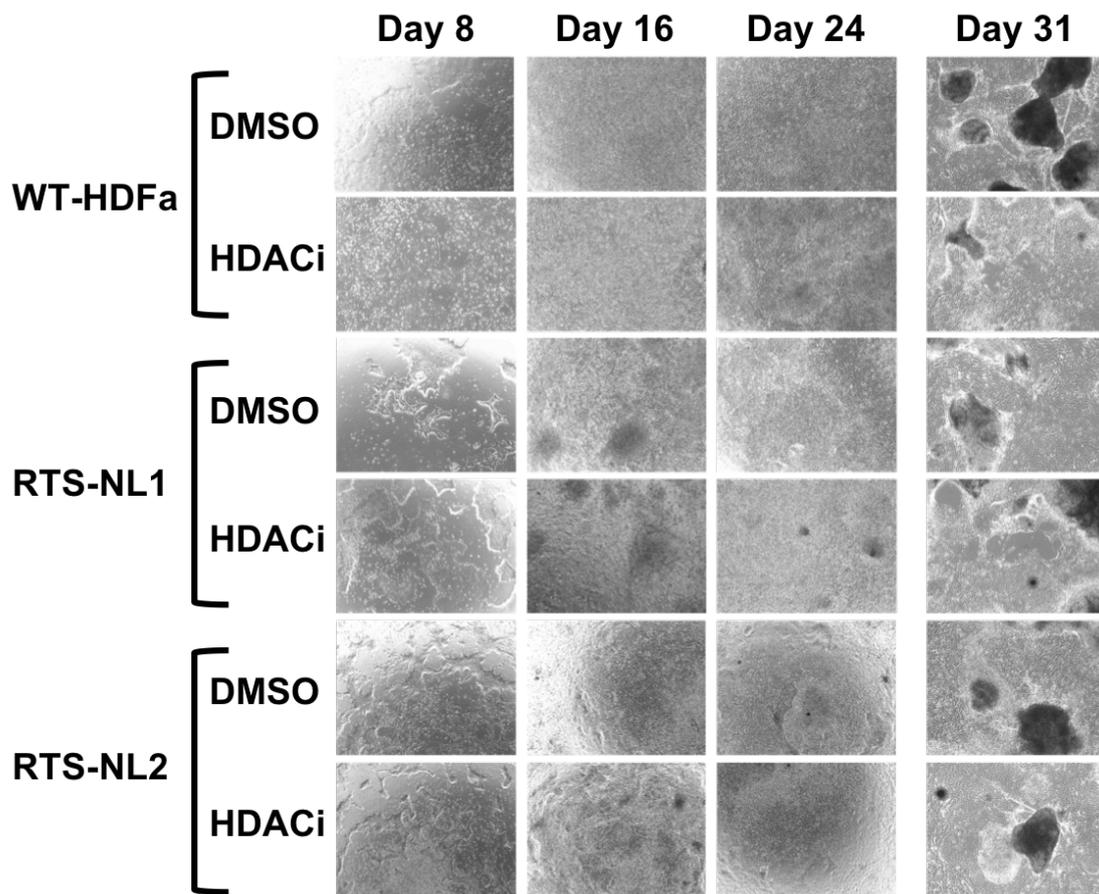
Figure 6. Differentiation of RTS iPSCs into neuronal cells with an HDAC inhibitor. Pictures were taken with a light microscope on day 8, 16, 24 and 31 with either DMSO as a control, or 5nM HDAC inhibitor (4X). No difference in morphology was observed between WT or RTS differentiation over time, with or without HDAC inhibitor treatment. Magnification 10X, n = 2 clones per cell line, 3 biological replicates per clone.
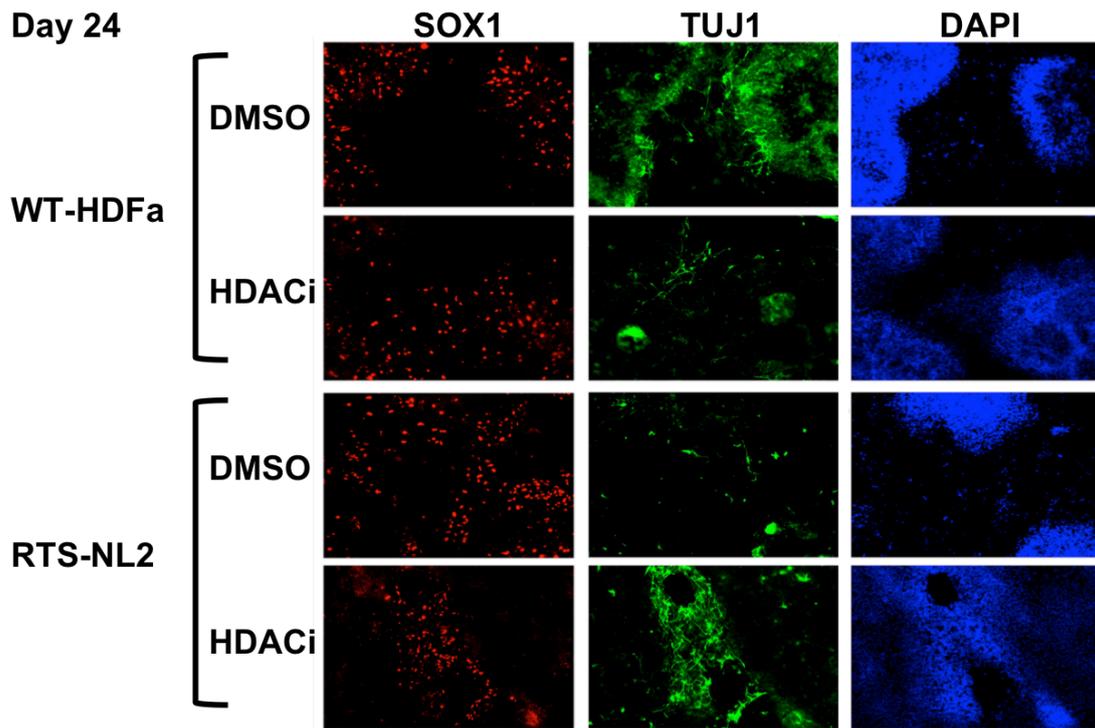
Figure 7. Neuronal markers for neuronal differentiation of iPSCs in the presence of an HDAC inhibitor. The HDAC inhibitor was added at 5nM during differentiation, and cells were fixed and stained on day 24. The addition of an HDAC inhibitor had no effect on the number of SOX1[+] cells. However, when looking at TUJ1[+] cells, there is an increase seen in the RTS lines during neuronal differentiation. Magnification 10X, n = 2 clones per cell line, 3 biological replicates per clone.
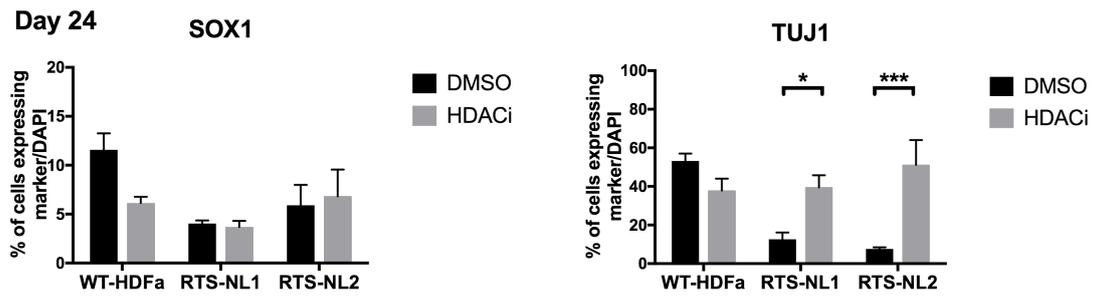
Figure 8. Quantification of neuronal markers with HDAC inhibitor. SOX1 expression is low and declining during differentiation as expected and there is no difference seen between WT and RTS lines. The TUJ1 expressing cells increases for RTS cells on day 24 following the addition of an HDAC inhibitor. Mean and SEM were plotted with comparison by Two-way ANOVA with Bonferroni's multiple comparisons test, n = 3 biological replicates, * p<0.05, *** p<0.001.
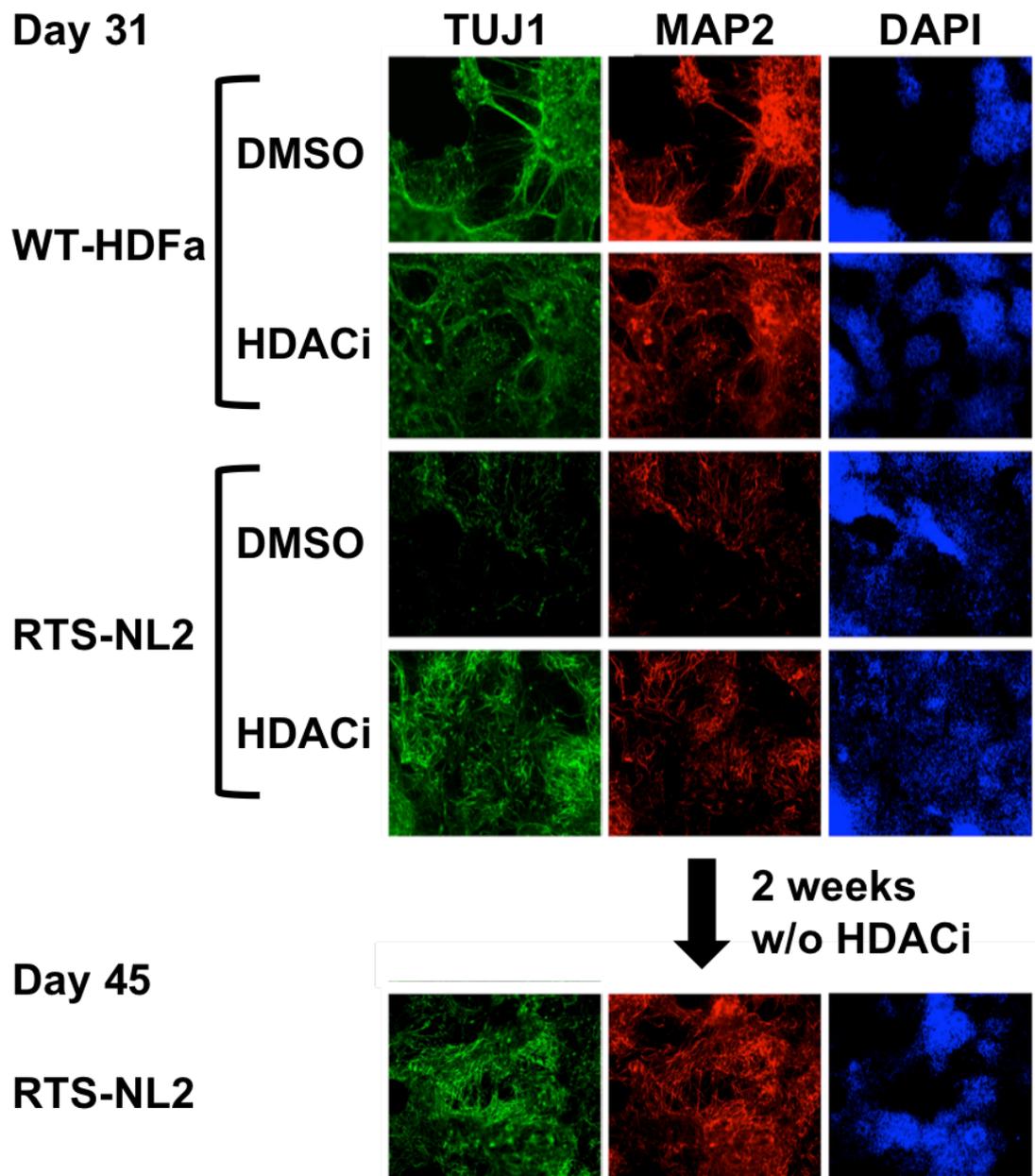
Figure 9. Neuronal differentiation on day 45. Immunofluorescence showed that both TUJ1+ and MAP2+ cells increase during differentiation in RTS lines when the HDAC inhibitor is added (A). When we removed the inhibitor for 2 weeks (day 31-45), both TUJ1 and MAP2 positive cell levels stayed stable. Magnification 10X, n = 2 clones per cell line, 3 biological replicates per clone.
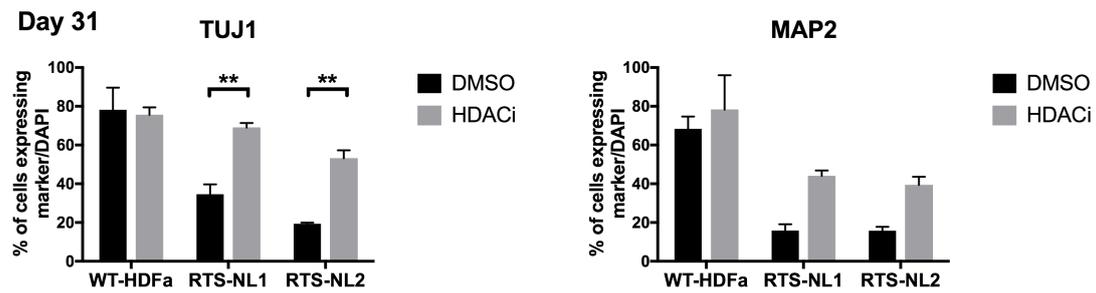
Figure 10. Quantification of neuronal marker day 31. There is a significant difference in TUJ1 positive cells on day 31 seen in the RTS lines following addition of the HDAC inhibitor. A similar trend is observed for MAP2. Mean and SEM were plotted with comparison by Two-way ANOVA with Bonferroni's multiple comparisons test, n = 3 biological replicates, ** $p < 0.01$.
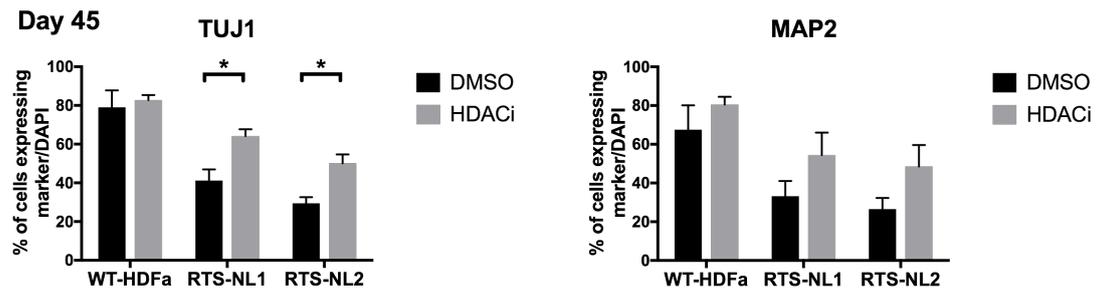
Figure 11. Differentiation potential remains stable when HDAC inhibitor is removed. When HDAC inhibitor is removed after 31 days of differentiation, TUJ1 and MAP2 positive cells stay stable in RTS lines when cultured for a further two weeks (day 45). There is a significant difference for TUJ1 positive cells when the inhibitor is added, and a trend is shown for MAP2. Mean and SEM were plotted with comparison by Two-way ANOVA with Bonferroni's multiple comparisons test, n = 3 biological replicates, * p<0.05.

shRNA knockdown of *CREBBP*

Next, to avoid the differences in somatic cell origin between the RTS and WT fibroblasts due to gender and tissue of origin, an isogenic control line was generated. This was achieved by knocking down *CREBBP* in a WT iPSC line to generate an ideal control-WT model. Short hairpin RNAs (shRNAs) were introduced into WT-HDFa iPSCs (HDFa 7p11) through infection with viral vectors with green fluorescent protein (GFP) reporters, to induce long-term knockdown of the target gene *CREBBP* [71] leading to lower CBP protein levels. Table 2 shows the efficiency of the control (negative) and three *CREBBP* shRNA vectors in human embryonic kidney (HEK) cells.

Table 2. FACS analysis to determine efficiency/MOI

Sample	Dilution	Efficiency (%)	MOI = 10 (ul/1.5ml med)
Control shRNA	1:100	91.8	
	1:1000	88.9	
	1:10000	32.6	
	1:100000	6.63	2.7
CREBBP shRNA 1	1:100	97.1	
	1:1000	93	
	1:10000	54.3	
	1:100000	13.8	1.3
CREBBP shRNA 2	1:100	96	
	1:1000	93.3	
	1:10000	44.4	
	1:100000	8.32	2.1
CREBBP shRNA 3	1:100	87.2	
	1:1000	90.7	
	1:10000	35.4	
	1:100000	9.18	1.9

Then WT-HDFa iPSCs were transfected with the vectors (1:100,000) and a GFP$^+$ cell sort was performed when colonies were established. Figure 12 shows the iPSC colonies at passage 10 showing the colonies through a light microscope (top) and a fluorescent microscope (bottom), which shows the GFP positive iPSCs within the colonies, showing efficient uptake of the shRNA vectors.
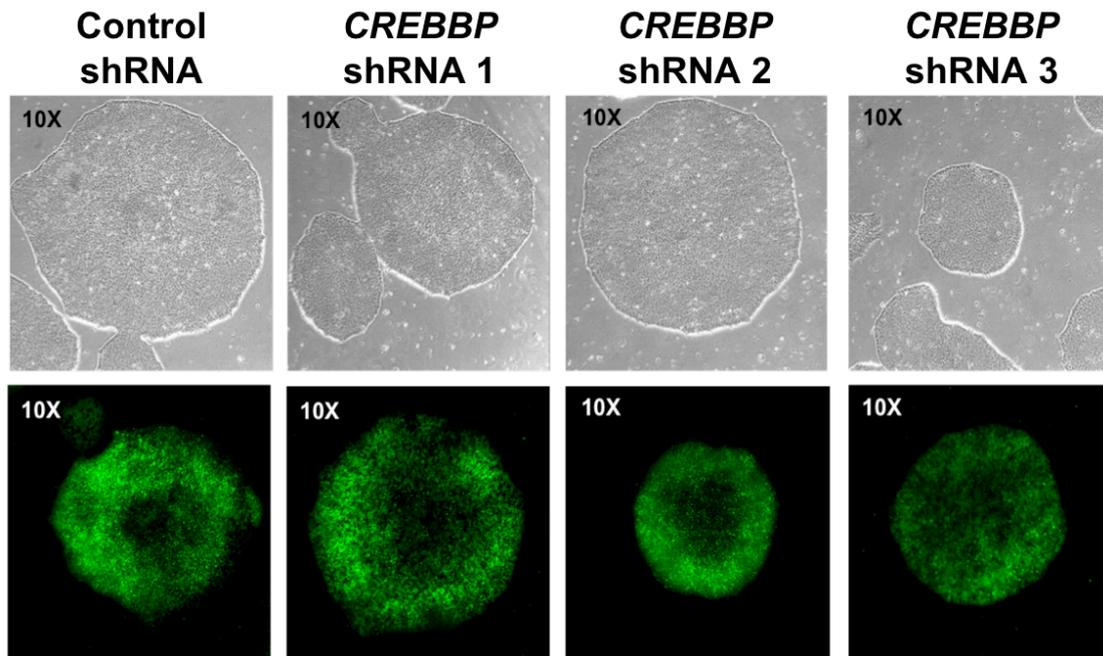
Figure 12. iPSC colonies transfected with shRNA vectors for *CREBBP* knockdown. iPSC colonies, showing normal morphology (top). GFP expression in the iPSC colonies, indicating efficient uptake of the shRNA vectors, added in a 100,000 dilution (bottom). Note that bright field and fluorescent images do not correspond. Magnification 10X, n = 3 biological replicates.

RNA was collected and the generated cDNA was used in a qPCR to confirm knockdown of the *CREBBP* gene (Figure 13). Vectors *CREBBP* shRNA 2 and 3 showed a 50% knockdown of *CREBBP*, which would be equivalent of the loss of one allele seen in RTS. Vector *CREBBP* shRNA 1 showed minimal knockdown of *CREBBP,* and was therefore excluded from the study.

These iPSCs containing the shRNA vectors were then differentiated into neuronal cells, using the same protocols as before. Again, there were no differences seen in morphology during differentiation with or without the HDAC inhibitor (Figure 14). However, when performing immunofluorescence staining on these cells on day 31 of neuronal differentiation, similar results were observed as the differentiated RTS iPSC lines (Figure 15). There is a reduction seen in TUJ1 positive cells in the *CREBBP* knockdown cells, and this is increased when the HDAC inhibitor is added during differentiation. Similar results are shown for MAP2. When quantifying this data, there was a significant difference between WT (control shRNA) and the *CREBBP* knockdown lines (Figure 16). This data matches the results observed with the RTS cell lines.

In conclusion, using shRNA to knockdown *CREBBP* can be used as a model for RTS. This method was used to better match control and disease cell lines, and by using cells with the same genetic background and generating an isogenic control system. Again, these results suggest that the addition of an HDAC inhibitor can alter neuronal protein levels seen in RTS cells *in vitro* and could potentially lead to a therapeutic option to reduce symptoms seen in RTS patients.
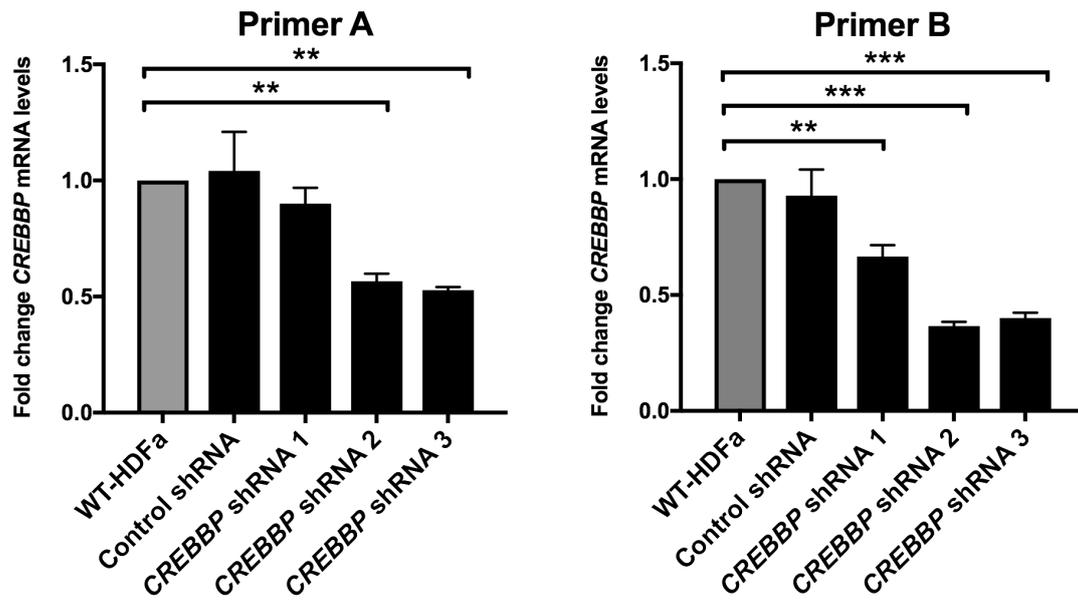
Figure 13. ***CREBBP* expression after shRNA knockdown.** qPCR data showing that both vectors *CREBBP* shRNA 2 and 3 leading to a 50% reduction in *CREBBP* expression. This would replicate the conditions seen in RTS. Mean and SEM were plotted with comparison by One-way ANOVA with Dunnet's multiple comparisons test, n = 3 biological replicates, ** p<0.01, *** p<0.001.
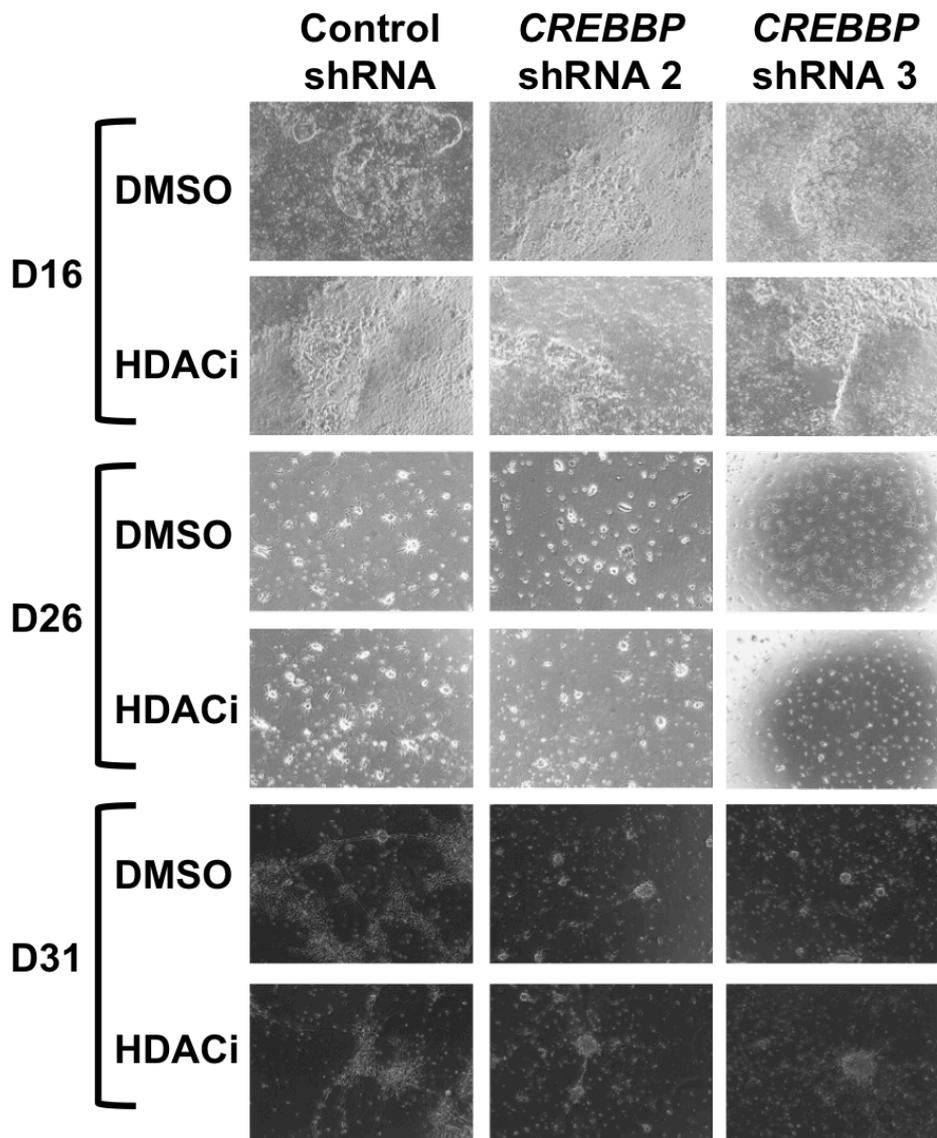
Figure 14. Differentiation of iPSCs with shRNA knockdown of *CREBBP*. Pictures where taken on days 16, 26 and 31 in the presence of either the HDAC inhibitor or DMSO as a control. No morphological differences where observed between WT and *CREBBP* knockdown cell lines. Magnification 10X, n = 2 clones per cell line, 3 biological replicates.
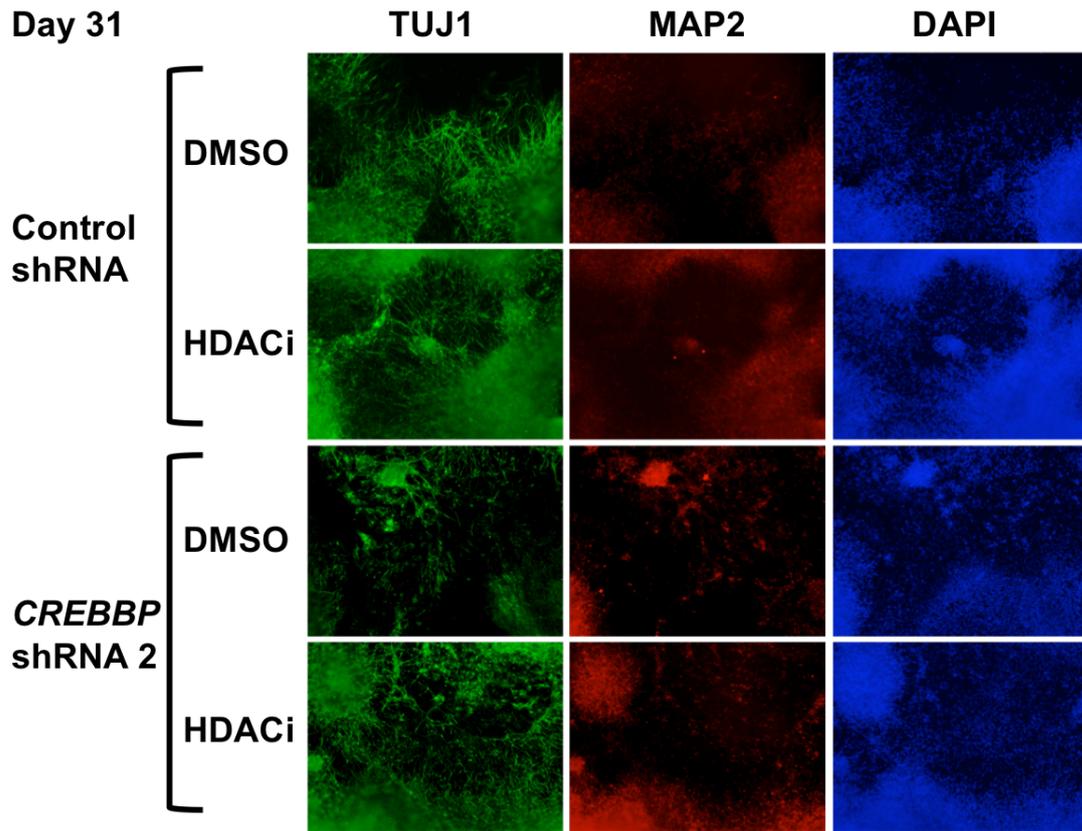
Figure 15. Neuronal differentiation staining day 31 shRNA *CREBBP* knockdown. Cells were fixed and stained on day 31 for TUJ1 and MAP2 markers. There was a reduction seen for both markers in the knockdown lines, which is reversed by the addition of the HDAC inhibitor. Magnification 10X, n = 3 biological replicates.
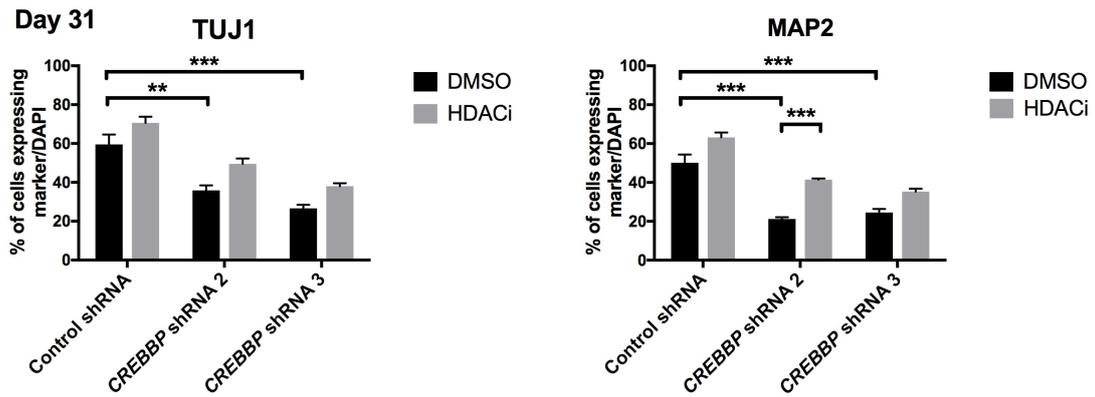
Figure 16. Quantification of Neuronal differentiation staining day 31 shRNA *CREBBP* knockdown. TUJ1 and MAP2 positive cells increase when an HDAC inhibitor is added during differentiation for the CREBBP knockdown lines. Mean and SEM were plotted with comparison by Two-way ANOVA with Bonferroni's multiple comparisons test, n = 3 biological replicates, ** $p < 0.01$, ***$p < 0.001$.

Discussion

In this chapter a disease in a dish model was generated and a small drug screen was performed as a proof of concept. The neuronal differentiation potential of RTS iPSCs was determined by looking at an NPC marker (SOX1) and more mature neural markers (TUJ1 and MAP2). It was shown that the amount of SOX1$^+$ cells is similar between WT and RTS iPSCs, but there is a reduction seen in cells expressing TUJ1 during neuronal differentiation. This could mean that RTS patients/iPSCs can generate the early neuronal lineages (NPCs) but have reduced (or slowed down) neurogenesis into more mature neural cells. The addition of an HDAC inhibitor (LBH589/Panobinostat) at a concentration of 5nM shows a potential increase in TUJ1 positive cells when differentiating the RTS cells, which is also shown for the MAP2 marker. To better match the genetic background of WT and disease cells for this model, shRNA was induced in a WT iPSC line to knockdown *CREBBP* expression, generating an isogenic model. This showed a similar phenotype of reduced TUJ1 and MAP2 positive cells in the differentiated cell lines with decreased levels of CBP and again, this phenotype appeared to be rescued when Panobinostat was added during the differentiation process. It should be noted that more experiments and analysis are needed to support these conclusions further. First of all, the quantification of the data was performed by visual analysis, and a more robust quantification, such as Imaris which is 3D imaging software (BitPlane, Belfast), should be used in future to confirm the data.

In conclusion, both RTS cells and WT isogenic controls show similar results in these experiments. Therefore, depending on the situation and availability of cells, both models are good alternatives. In addition, shRNA can also be used to generate an *EP300* knockdown model, as mutations in this gene are rare in RTS patients. Even though more research is needed, Panobinostat could be a potential compound to reduce specific RTS symptoms. However, this experimental set up can be easily expanded into a larger drug screen cohort, using a variety of different HDAC inhibitors, such as more selective ones. Panobinostat is a non-selective inhibitor, and might induce overall hyperacetylation in the cells, which could consequently lead to an increase in abnormal gene expression. Chromatin immunoprecipitation (ChIP) assays could be used to look at these differences in expression profiles between patient and WT, and/or in addition of a HDAC inhibitor. If time and money were not

issues, this can be combined with Next Generation Sequencing (ChIP-seq) to generate a vast amount of data on histone acetylation profiles between these samples.

There is limited knowledge about RTS brains besides structural anomalies. However, mouse studies have shown that cbp/p300 HAT activity plays an important role in memory and other forms of neuroplasticity. Based on these studies the generated iPSCs from Chapter 2 were differentiated towards the neuronal lineage. This has been done for other neurodevelopmental disorders such as Timothy Syndrome (TS). Paşca et al. generated cortical neuronal precursor cells and neurons from iPSCs derived from individuals with TS [33]. The cerebral cortex is the brains outer layer of neural tissue and is composed of grey and white matter. It also contains the neocortex, which include the hippocampus and plays a key role in memory [72].

Using human iPSCs as a disease in a dish model has opened up many possibilities, especially in the neurological field. The complexity of the human brain makes it difficult to study many neurological disorders in model organisms, highlighting the need for an *in vitro* model of human brain development. The generation of human iPSCs and differentiation protocols used to generate various cell lineages has made the collection of cells from foetuses, human remains, and discarded surgical waste unnecessary. Generating patient-specific neuronal cells from iPSCs has promising applications in drug screening as well as for *in vitro* modelling of neurodegenerative and neuro-developmental disorders. These iPSC derived neuronal cell lines have the advantage to genetically and physiologically resemble the *in vivo* natural environment of the human body more closely compared to alternative research models, such as (transformed) cancer-derived human neuronal cell lines.

Even though animal models have been frequently used to generate models for human diseases and for drug screening to develop therapies, many diseases lack a suitable animal model and high-throughput screening is usually not feasible. For example, mutations that lead to microcephaly in humans do not severely reduce the size of a mouse brain and the exact reason behind this remains unknown [73]. However, it is thought that mouse neural progenitors do not undergo expansion to the same extent as in humans before the onset of neurogenesis. Therefore, a disruption would not be as severe as in humans [74]. Consequently, neurons generated from iPSCs may provide a vital tool for the study of human neurodegenerative and neurodevelopmental diseases in a controlled *in vitro* environment [75].

However, more research is needed into the use of iPSCs as a model. It is known that

residual somatic epigenetic memory may persist in iPSCs [76-78], which can potentially affect their differentiation capacity. However, due to the small cohorts, more research is needed in order to form strong conclusions. Nevertheless, this is likely to improve over time, when more patient derived cell lines will become available [79].

Even though it is possible to differentiate human iPSCs into almost any desired cell type, there are still improvements needed and the ability to model cell-specific disorders depends on efficient and robust differentiation protocols. Recently, patient-derived disease specific cell types have been generated from iPSCs using well-established protocols. This has been done for a variety of neuronal degenerative diseases such as Parkinson and Huntington's diseases [80].

The generation of iPSCs in combination with neuronal differentiation can be a very time consuming and costly process. To differentiate fibroblasts into neural cells via iPSCs, typically requires 4–6 months before the first functional neurons are generated, with 2–6 weeks for neuronal differentiation. However, high proportions of truly functional cells may require extended culture periods for e.g. cortical neurons, which takes a remarkably long time to reach functional maturation in the embryonic brain, which is reflected in culture [81]. Therefore, the optimization of *in vitro* maturation protocols will be important to make neuronal differentiation more efficient and to select different subtypes of neurons that are disease specific.

There is a large number of novel differentiation and conversion protocols available to generate different types of neuronal cells. Recently, a protocol using direct differentiation from fibroblasts to NPCs has been developed, which will shorten the complete differentiation process considerably [82, 83] and functional human neurons can be generated from fibroblasts within 1–3 weeks plus maturation time [82, 84].

However, there are technical as well as conceptual differences between iPSC differentiation and direct neuronal conversion that may influence the choice of system for specific *in vitro* projects [85, 86]. Whereas conventional differentiation with iPSCs involves the use of a variety of growth factors and mitogens to develop neuronal lineages [87, 88], direct conversion uses the over-expression of cell type specific transcription factors. This will stimulate lineage changes and direct the identity towards the desired cell type, thereby bypassing most developmental stages [86].

When differentiating human iPSCs into neural tissue, these pluripotent cells will

follow a developmental pathway that simulates several cellular stages of human neural development [27]. This is particularly interesting and relevant for neurodevelopmental diseases such as Rett syndrome and RTS [89] but also for neuropsychiatric diseases such as autism spectrum disorders and schizophrenia [90, 91]. The pathology of these syndromes might have neurodevelopmental components that can be closer reflected through iPSC differentiation rather than direct conversion. If the disease phenotype is thought to originate during the NPC stage of neurodevelopment, it is better to avoid direct differentiation. This method skips all developmental precursor cell stages, and therefore generates neurons that have never been in an NPC-like cell stage.

When screening drugs in an *in vitro* assay, it is vital that optimized differentiation protocols are developed to generate a homogeneous and stable neural cell population. This can be achieved by characterization of cell populations but also by optimizing differentiation protocols for better reproducibility and consistency, which is essential to serve as a high-predicatively drug-screening tool. Although these protocols are continuously being improved due to more efficient reprogramming and differentiation methods, cell heterogeneity remains an issue in these models when generating disease-specific cell types that can act as a potential confounder [92, 93]. By performing FACS isolation assays, this cell heterogeneity can be controlled by sorting pure cell population prior to the epigenetic analyses [94, 95].

Conversely, the human brain is a complex organ composed of immense network of different (neuronal) cell types. It is therefore important to know whether iPSC-derived neurons will mimic the neuronal cell type of interest completely. This can be of great relevance for the expression and epigenetic profiles of neuronal cells, as these are affected by the combination of genetic background, environmental factors, and cell-cell interactions in the brain. Therefore, culturing these cells *in vitro* might incorrectly reflect the affected brain areas in disease [95].

Even though the cerebral cortex is one of the most complex areas of the human body, the developmental origin stems from a single layer of neuroepithelial progenitors, which give rise to all different cell types of the cortex. Despite this complexity, several groups have demonstrated that PSCs grown *in vitro* have the tendency to form polarized neuroepithelial structures that resemble progenitor zones of the embryonic cortex [62, 96]. Even though, *in vitro* neural cells are typically cultured as a monolayer, the generation of three-dimensional (3D) structures represents a more

natural microenvironment of the brain. These studies show that under the right conditions, iPSCs are able to differentiate and develop into 3D structures, resembling miniature organs. This has been shown for the intestine, thyroid, retina, and intestine [97-99] and it is thought that interactions between different cell types, as is the case in the brain, can be better understood by 3D tissue structures.

As mentioned before, there are significant differences between mice and humans, especially in brain development and therefore, methods that can simulate this (*in vitro*) as closely as possible are in high demand. These 'mini brains' develop various brain regions and include a cerebral cortex containing progenitor populations that organize and produce mature cortical neuron subtypes. Besides looking at (molecular) interactions, these 3D structures have also been used to uncover disease related phenotypes. Lancaster et al. generated cerebral organoids or mini brains from human iPSCs [73] and made a model for human microcephaly, which lacks a suitable animal model. By using RNA interference (RNAi) in combination with patient specific iPSCs, they demonstrated premature neuronal differentiation in patient organoids, a defect that could help to explain the microcephaly phenotype. In conclusion, these 3D organoids can mimic development and disease even in the most complex human tissue.

Even though the HDAC inhibitor used in this study, Panobinostat, did not influence viability to fibroblasts at concentrations of 500nM, as a previous report has shown [46], the iPSCs showed reduced viability at concentrations of 10nM and higher. This may be explained by the fact that iPSCs (and ESCs) share cellular and molecular phenotypes with tumour cells and cancer cell lines [100-102], such as rapid cell proliferation [103], tendency for genomic instability [104, 105] and gene expression and epigenetic profiles [106-109]. It was shown with several human cancer cell lines (osteosarcomas and rhabdoid tumour cell lines) that continuous exposure to LBH589 resulted in dose-dependent reduction of cell viability. Concentrations of 15nM or higher resulted in a marked reduction in cell growth, whereas concentrations over 30nM lead to cell death [110].

Due to the limitations of the neuronal cell culture system and set up in this study, there were not sufficient cells to perform a Western Blot to reflect the effects of different concentrations of the HDAC inhibitor on the levels of histone acetylation in RTS-iPSCs and the differentiated neuronal cells. However, a HIV study with human

peripheral blood mononuclear cells (PBMCs) treated with 10nM Panobinostat showed three times more histone acetylation than untreated PBMCs [111].

HDAC inhibitors have a history in neurological and psychiatric fields and more recently are being tested in clinical trial as cancer and inflammatory treatments. There are several classes of HDAC inhibitors, both chemically synthesized or isolated as natural products. The most commonly used HDAC inhibitors are pan-HDAC inhibitors, which target multiple HDACs. However, there is limited knowledge about the mechanisms of these non-selective HDAC inhibitors and they have varying target specificity and properties *in vitro* and *in vivo* [112]and in clinical settings [113]. Due to the non-selective nature of these HDAC inhibitors, it is difficult to determine if the promising results in drug screening are due to inhibition of a specific HDAC, a combined effect of the inhibition of several HDACs or in respect to protein complexes that include HDACs. However, alternative methods are needed to target CBP and/or p300 directly. General HDAC inhibition would also have an effect on non-CBP/p300 target and this could lead to hyperacetylation at non-specific regulatory elements and histone marks. Despite of the promising results and the fact that HDAC inhibitors have made it to clinical trials, more research is needed to unravel the mechanisms of these compounds.

The availability of the human genome and the generation and differentiation of human iPSCs has revolutionized disease research. This in combination with genome editing will give opportunities to develop potential therapies and treatments [114]. Genome editing tools (GE), such as ZFNs, TALENs, and CRISPR-Cas9, are used to introduce targeted genetic changes, which can be used to correct mutations in genes causing disease and have been applied in a variety of fields, such as biology, agriculture and medicine. When applied to patient specific iPSCs, this can be used to correct the variants and in combination with differentiating these cells could potentially be transplanted back into the patient, leading to an improved phenotype while avoiding a rejection reaction. These genome-editing techniques also allow isogenic controls to be used, making comparisons more relevant.

However, GE tools are limited by off-target mutations, which might lead to negative side effects such as the induction of apoptosis and chromosomal rearrangements, which can be of major concern in medical and clinical studies [115]. Nevertheless, it has been shown that with WGS and WES that off-target mutations are rarely induced by the CRISPR-Cas9 technique [116, 117], which strengthen the potential for

therapeutic purposes [120]. However, GE tools need to be completely safe to be used in gene therapies and in clinical studies to avoid any unintended consequences arising from off-target mutations.

GE tools in combination with iPSCs have already been applied in several mouse model studies. Mice with sickle cell anaemia have been treated with iPSC-derived haematopoietic progenitors [121] and neural cells derived from mouse iPSCs have been used to cure Parkinson disease and spinal cord injury [122, 123]. Further progress has been made with a study treating spinal cord injury with engraftments of human iPSC-derived neural stem and progenitor cells in non-human primates [124].

In conclusion, modelling human diseases using iPSCs has already been successfully applied from bedside to bench by reprogramming and differentiation of patient specific somatic cells. These disease in a dish models, are being extensively used in drug screening and validation studies for the development of therapeutics. However, now is the time to move from bench back to bedside and, ideally, find new treatments and development of personalized medicine for a wide range of (neurological) disorders, such as RTS.

References

1.	*American Association of Intellectual and Developmental Disabilities.* Available from: https://aaidd.org/.
2.	Miller, R.W. and J.H. Rubinstein, *Tumors in Rubinstein-Taybi syndrome.* Am J Med Genet, 1995. **56**(1): p. 112-5.
3.	Marzuillo, P., et al., *Novel cAMP binding protein-BP (CREBBP) mutation in a girl with Rubinstein-Taybi syndrome, GH deficiency, Arnold Chiari malformation and pituitary hypoplasia.* BMC Med Genet, 2013. **14**: p. 28.
4.	Parsley, L., et al., *Identical twin sisters with Rubinstein-Taybi syndrome associated with Chiari malformations and syrinx.* Am J Med Genet A, 2011. **155A**(11): p. 2766-70.
5.	Sener, R.N., *Rubinstein-Taybi syndrome: cranial MR imaging findings.* Comput Med Imaging Graph, 1995. **19**(5): p. 417-8.
6.	Schorry, E.K., et al., *Genotype-phenotype correlations in Rubinstein-Taybi syndrome.* Am J Med Genet A, 2008. **146A**(19): p. 2512-9.
7.	Panning, B. and R. Jaenisch, *RNA and the epigenetic regulation of X chromosome inactivation.* Cell, 1998. **93**(3): p. 305-8.
8.	Reik, W. and J. Walter, *Genomic imprinting: parental influence on the genome.* Nat Rev Genet, 2001. **2**(1): p. 21-32.
9.	Borrelli, E., et al., *Decoding the epigenetic language of neuronal plasticity.* Neuron, 2008. **60**(6): p. 961-74.
10.	Zocchi, L. and P. Sassone-Corsi, *Joining the dots: from chromatin remodeling to neuronal plasticity.* Curr Opin Neurobiol, 2010. **20**(4): p. 432-40.
11.	Yao, T.P., et al., *Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300.* Cell, 1998. **93**(3): p. 361-72.
12.	Li, Q., H. Xiao, and K. Isobe, *Histone acetyltransferase activities of cAMP-regulated enhancer-binding protein and p300 in tissues of fetal, young, and old mice.* J Gerontol A Biol Sci Med Sci, 2002. **57**(3): p. B93-8.
13.	Cesena, T.I., et al., *CCAAT/enhancer-binding protein (C/EBP) beta is acetylated at multiple lysines: acetylation of C/EBPbeta at lysine 39 modulates its ability to activate transcription.* J Biol Chem, 2007. **282**(2): p. 956-67.
14.	Li, Q.J., et al., *MAP kinase phosphorylation-dependent activation of Elk-1 leads to activation of the co-activator p300.* EMBO J, 2003. **22**(2): p. 281-91.
15.	Lundblad, J.R., et al., *Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP.* Nature, 1995. **374**(6517): p. 85-8.
16.	Romano, A., et al., *Evolutionarily-conserved role of the NF-kappaB transcription factor in neural plasticity and memory.* Eur J Neurosci, 2006. **24**(6): p. 1507-16.
17.	Oliveira, A.M., et al., *Transgenic mice expressing an inhibitory truncated form of p300 exhibit long-term memory deficits.* Learn Mem, 2007. **14**(9): p. 564-72.
18.	Barrett, R.M., et al., *Hippocampal focal knockout of CBP affects specific histone modifications, long-term potentiation, and long-term memory.* Neuropsychopharmacology, 2011. **36**(8): p. 1545-56.

19.   Wood, M.A., et al., *Transgenic mice expressing a truncated form of CREB-binding protein (CBP) exhibit deficits in hippocampal synaptic plasticity and memory storage.* Learn Mem, 2005. **12**(2): p. 111-9.

20.   Tsui, D., et al., *CBP regulates the differentiation of interneurons from ventral forebrain neural precursors during murine development.* Dev Biol, 2014. **385**(2): p. 230-41.

21.   Hallam, T.M. and R. Bourtchouladze, *Rubinstein-Taybi syndrome: molecular findings and therapeutic approaches to improve cognitive dysfunction.* Cell Mol Life Sci, 2006. **63**(15): p. 1725-35.

22.   Xie, W., et al., *Epigenomic analysis of multilineage differentiation of human embryonic stem cells.* Cell, 2013. **153**(5): p. 1134-48.

23.   Lister, R., et al., *Global epigenomic reconfiguration during mammalian brain development.* Science, 2013. **341**(6146): p. 1237905.

24.   Hawkins, R.D., et al., *Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.* Cell Stem Cell, 2010. **6**(5): p. 479-91.

25.   Cantone, I. and A.G. Fisher, *Epigenetic programming and reprogramming during development.* Nat Struct Mol Biol, 2013. **20**(3): p. 282-9.

26.   Gifford, C.A., et al., *Transcriptional and epigenetic dynamics during specification of human embryonic stem cells.* Cell, 2013. **153**(5): p. 1149-63.

27.   Ziller, M.J., et al., *Dissecting neural differentiation regulatory networks through epigenetic footprinting.* Nature, 2015. **518**(7539): p. 355-9.

28.   Mertens, J., et al., *Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience.* Nat Rev Neurosci, 2016. **17**(7): p. 424-37.

29.   Merkle, F.T. and A. Alvarez-Buylla, *Neural stem cells in mammalian development.* Curr Opin Cell Biol, 2006. **18**(6): p. 704-9.

30.   Spiers, H., et al., *Methylomic trajectories across human fetal brain development.* Genome Res, 2015. **25**(3): p. 338-52.

31.   Breier, J.M., et al., *Neural progenitor cells as models for high-throughput screens of developmental neurotoxicity: state of the science.* Neurotoxicol Teratol, 2010. **32**(1): p. 4-15.

32.   An, M.C., et al., *Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells.* Cell Stem Cell, 2012. **11**(2): p. 253-63.

33.   Pasca, S.P., et al., *Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome.* Nat Med, 2011. **17**(12): p. 1657-62.

34.   Hamm, C.A. and F.F. Costa, *The impact of epigenomics on future drug design and new therapies.* Drug Discov Today, 2011. **16**(13-14): p. 626-35.

35.   Haberland, M., R.L. Montgomery, and E.N. Olson, *The many roles of histone deacetylases in development and physiology: implications for disease and therapy.* Nat Rev Genet, 2009. **10**(1): p. 32-42.

36.   Graff, J., et al., *An epigenetic blockade of cognitive functions in the neurodegenerating brain.* Nature, 2012. **483**(7388): p. 222-6.

37.   Saha, R.N. and K. Pahan, *HATs and HDACs in neurodegeneration: a tale of disconcerted acetylation homeostasis.* Cell Death Differ, 2006. **13**(4): p. 539-50.

38.   Feng, Y., J. Jankovic, and Y.C. Wu, *Epigenetic mechanisms in Parkinson's disease.* J Neurol Sci, 2015. **349**(1-2): p. 3-9.

39.   Konsoula, Z. and F.A. Barile, *Epigenetic histone acetylation and deacetylation mechanisms in experimental models of neurodegenerative disorders.* J Pharmacol Toxicol Methods, 2012. **66**(3): p. 215-20.

40.   Pirooznia, S.K. and F. Elefant, *Targeting specific HATs for neurodegenerative disease treatment: translating basic biology to therapeutic possibilities.* Front Cell Neurosci, 2013. **7**: p. 30.

41.   Reynolds, M.F., E.C. Sisk, and N.L. Rasgon, *Valproate and neuroendocrine changes in relation to women treated for epilepsy and bipolar disorder: a review.* Curr Med Chem, 2007. **14**(26): p. 2799-812.

42.   Miller, T.A., D.J. Witter, and S. Belvedere, *Histone deacetylase inhibitors.* J Med Chem, 2003. **46**(24): p. 5097-116.

43.   Blanchard, F. and C. Chipoy, *Histone deacetylase inhibitors: new drugs for the treatment of inflammatory diseases?* Drug Discov Today, 2005. **10**(3): p. 197-204.

44.   Tough, D.F., et al., *Epigenetic drug discovery: breaking through the immune barrier.* Nat Rev Drug Discov, 2016. **15**(12): p. 835-853.

45.   Falkenberg, K.J. and R.W. Johnstone, *Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders.* Nat Rev Drug Discov, 2014. **13**(9): p. 673-91.

46.   Lopez-Atalaya, J.P., et al., *Histone acetylation deficits in lymphoblastoid cell lines from patients with Rubinstein-Taybi syndrome.* J Med Genet, 2012. **49**(1): p. 66-74.

47.   Alarcon, J.M., et al., *Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration.* Neuron, 2004. **42**(6): p. 947-59.

48.   Korzus, E., M.G. Rosenfeld, and M. Mayford, *CBP histone acetyltransferase activity is a critical component of memory consolidation.* Neuron, 2004. **42**(6): p. 961-72.

49.   Vecsey, C.G., et al., *Histone deacetylase inhibitors enhance memory and synaptic plasticity via CREB:CBP-dependent transcriptional activation.* J Neurosci, 2007. **27**(23): p. 6128-40.

50.   Haettig, J., et al., *HDAC inhibition modulates hippocampus-dependent long-term memory for object location in a CBP-dependent manner.* Learn Mem, 2011. **18**(2): p. 71-9.

51.   Harrison, I.F. and D.T. Dexter, *Epigenetic targeting of histone deacetylase: therapeutic potential in Parkinson's disease?* Pharmacol Ther, 2013. **140**(1): p. 34-52.

52.   Jia, H., et al., *HDAC inhibition imparts beneficial transgenerational effects in Huntington's disease mice via altered DNA and histone methylation.* Proc Natl Acad Sci U S A, 2015. **112**(1): p. E56-64.

53.   Jin, H., et al., *Histone hyperacetylation up-regulates protein kinase Cdelta in dopaminergic neurons to induce cell death: relevance to epigenetic mechanisms of neurodegeneration in Parkinson disease.* J Biol Chem, 2014. **289**(50): p. 34743-67.

54.   Johnson, A.A., et al., *Increasing Tip60 HAT levels rescues axonal transport defects and associated behavioral phenotypes in a Drosophila Alzheimer's disease model.* J Neurosci, 2013. **33**(17): p. 7535-47.

55.   Duvic, M., et al., *Panobinostat activity in both bexarotene-exposed and -naive patients with refractory cutaneous T-cell lymphoma: results of a phase II trial.* Eur J Cancer, 2013. **49**(2): p. 386-94.

56.   Hogh Kolbaek Kjaer, A.S., et al., *The histone deacetylase inhibitor panobinostat lowers biomarkers of cardiovascular risk and inflammation in HIV patients.* AIDS, 2015. **29**(10): p. 1195-200.

57.   Rathkopf, D.E., et al., *A phase 2 study of intravenous panobinostat in patients with castration-resistant prostate cancer.* Cancer Chemother Pharmacol, 2013. **72**(3): p. 537-44.

58.   Tan, P., et al., *Dual epigenetic targeting with panobinostat and azacitidine in acute myeloid leukemia and high-risk myelodysplastic syndrome.* Blood Cancer J, 2014. **4**: p. e170.

59.   Younes, A., et al., *Panobinostat in patients with relapsed/refractory Hodgkin's lymphoma after autologous stem-cell transplantation: results of a phase II study.* J Clin Oncol, 2012. **30**(18): p. 2197-203.

60.   Garnock-Jones, K.P., *Panobinostat: first global approval.* Drugs, 2015. **75**(6): p. 695-704.

61.   Atadja, P., *Development of the pan-DAC inhibitor panobinostat (LBH589): successes and challenges.* Cancer Lett, 2009. **280**(2): p. 233-41.

62.   Mariani, J., et al., *Modeling human cortical development in vitro using induced pluripotent stem cells.* Proc Natl Acad Sci U S A, 2012. **109**(31): p. 12770-5.

63.   Espuny-Camacho, I., et al., *Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits in vivo.* Neuron, 2013. **77**(3): p. 440-56.

64.   Gaspard, N., et al., *Generation of cortical neurons from mouse embryonic stem cells.* Nat Protoc, 2009. **4**(10): p. 1454-63.

65.   McMahon, J.A., et al., *Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite.* Genes Dev, 1998. **12**(10): p. 1438-52.

66.   Suter, D.M., et al., *A Sox1 to Pax6 switch drives neuroectoderm to radial glia progression during differentiation of mouse embryonic stem cells.* Stem Cells, 2009. **27**(1): p. 49-58.

67.   von Bohlen Und Halbach, O., *Immunohistological markers for staging neurogenesis in adult hippocampus.* Cell Tissue Res, 2007. **329**(3): p. 409-20.

68.   Lyck, L., et al., *Immunohistochemical markers for quantitative studies of neurons and glia in human neocortex.* J Histochem Cytochem, 2008. **56**(3): p. 201-21.

69.   Gusel'nikova, V.V. and D.E. Korzhevskiy, *NeuN As a Neuronal Nuclear Antigen and Neuron Differentiation Marker.* Acta Naturae, 2015. **7**(2): p. 42-7.

70.   Garbes, L., et al., *LBH589 induces up to 10-fold SMN protein levels by several independent mechanisms and is effective even in cells from SMA patients non-responsive to valproate.* Hum Mol Genet, 2009. **18**(19): p. 3645-58.

71.   Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.

72.   Kwon, H.B., et al., *Neuroligin-1-dependent competition regulates cortical synaptogenesis and synapse number.* Nat Neurosci, 2012. **15**(12): p. 1667-74.

73.   Lancaster, M.A., et al., *Cerebral organoids model human brain development and microcephaly.* Nature, 2013. **501**(7467): p. 373-9.

74.   Lui, J.H., D.V. Hansen, and A.R. Kriegstein, *Development and evolution of the human neocortex.* Cell, 2011. **146**(1): p. 18-36.

75.   Marchetto, M.C., et al., *Induced pluripotent stem cells (iPSCs) and neurological disease modeling: progress and promises.* Hum Mol Genet, 2011. **20**(R2): p. R109-15.

76.   Bar-Nur, O., et al., *Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells.* Cell Stem Cell, 2011. **9**(1): p. 17-23.

77.   Ohi, Y., et al., *Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells.* Nat Cell Biol, 2011. **13**(5): p. 541-9.

78.   Kim, K., et al., *Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells.* Nat Biotechnol, 2011. **29**(12): p. 1117-9.

79.   Avior, Y., I. Sagi, and N. Benvenisty, *Pluripotent stem cells in disease modelling and drug discovery.* Nat Rev Mol Cell Biol, 2016. **17**(3): p. 170-82.

80.   Yap, M.S., et al., *Neural Differentiation of Human Pluripotent Stem Cells for Nontherapeutic Applications: Toxicology, Pharmacology, and In Vitro Disease Modeling.* Stem Cells Int, 2015. **2015**: p. 105172.

81.   Nicholas, C.R., et al., *Functional maturation of hPSC-derived forebrain interneurons requires an extended timeline and mimics human neural development.* Cell Stem Cell, 2013. **12**(5): p. 573-86.

82.   Pang, Z.P., et al., *Induction of human neuronal cells by defined transcription factors.* Nature, 2011. **476**(7359): p. 220-3.

83.   Vierbuchen, T., et al., *Direct conversion of fibroblasts to functional neurons by defined factors.* Nature, 2010. **463**(7284): p. 1035-41.

84.   Liu, M.L., et al., *Small molecules enable neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons.* Nat Commun, 2013. **4**: p. 2183.

85.   Kelaini, S., A. Cochrane, and A. Margariti, *Direct reprogramming of adult cells: avoiding the pluripotent state.* Stem Cells Cloning, 2014. **7**: p. 19-29.

86.   Vierbuchen, T. and M. Wernig, *Direct lineage conversions: unnatural but useful?* Nat Biotechnol, 2011. **29**(10): p. 892-907.

87.   Chambers, S.M., et al., *Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling.* Nat Biotechnol, 2009. **27**(3): p. 275-80.

88.   Maury, Y., et al., *Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes.* Nat Biotechnol, 2015. **33**(1): p. 89-96.

89.   Marchetto, M.C., et al., *A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells.* Cell, 2010. **143**(4): p. 527-39.

90.   Rapoport, J.L., J.N. Giedd, and N. Gogtay, *Neurodevelopmental model of schizophrenia: update 2012.* Mol Psychiatry, 2012. **17**(12): p. 1228-38.

91.   Samaco, R.C., A. Hogart, and J.M. LaSalle, *Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3.* Hum Mol Genet, 2005. **14**(4): p. 483-92.

92.   De Jager, P.L., et al., *Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci.* Nat Neurosci, 2014. **17**(9): p. 1156-63.

93.   Lunnon, K., et al., *Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease.* Nat Neurosci, 2014. **17**(9): p. 1164-70.

94.   Poulin, J.F., et al., *Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling.* Cell Rep, 2014. **9**(3): p. 930-43.

95.   Roessler, R., et al., *Detailed analysis of the genetic and epigenetic signatures of iPSC-derived mesodiencephalic dopaminergic neurons.* Stem Cell Reports, 2014. **2**(4): p. 520-33.

96.   Gaspard, N., et al., *An intrinsic mechanism of corticogenesis from embryonic stem cells.* Nature, 2008. **455**(7211): p. 351-7.

97.   Sato, T., et al., *Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche.* Nature, 2009. **459**(7244): p. 262-5.

98.   Eiraku, M., et al., *Self-organizing optic-cup morphogenesis in three-dimensional culture.* Nature, 2011. **472**(7341): p. 51-6.

99.   Sasai, Y., *Next-generation regenerative medicine: organogenesis from stem cells in 3D culture.* Cell Stem Cell, 2013. **12**(5): p. 520-30.

100.   Dreesen, O. and A.H. Brivanlou, *Signaling pathways in cancer and embryonic stem cells.* Stem Cell Rev, 2007. **3**(1): p. 7-17.

101.   Knoepfler, P.S., *Deconstructing stem cell tumorigenicity: a roadmap to safe regenerative medicine.* Stem Cells, 2009. **27**(5): p. 1050-6.

102.   Blum, B. and N. Benvenisty, *The tumorigenicity of human embryonic stem cells.* Adv Cancer Res, 2008. **100**: p. 133-58.

103.   Amit, M., et al., *Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture.* Dev Biol, 2000. **227**(2): p. 271-8.

104.   Baker, D.E., et al., *Adaptation to culture of human embryonic stem cells and oncogenesis in vivo.* Nat Biotechnol, 2007. **25**(2): p. 207-15.

105.   Harrison, N.J., D. Baker, and P.W. Andrews, *Culture adaptation of embryonic stem cells echoes germ cell malignancy.* Int J Androl, 2007. **30**(4): p. 275-81; discussion 281.

106.   Sperger, J.M., et al., *Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13350-5.

107.   Ben-Porath, I., et al., *An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors.* Nat Genet, 2008. **40**(5): p. 499-507.

108.   Wong, D.J., et al., *Module map of stem cell genes guides creation of epithelial cancer stem cells.* Cell Stem Cell, 2008. **2**(4): p. 333-44.

109.   Calvanese, V., et al., *Cancer genes hypermethylated in human embryonic stem cells.* PLoS One, 2008. **3**(9): p. e3294.

110.   Cain, J.E., et al., *Sustained Low-Dose Treatment with the Histone Deacetylase Inhibitor LBH589 Induces Terminal Differentiation of Osteosarcoma Cells.* Sarcoma, 2013. **2013**: p. 608964.

111.   Tsai, P., et al., *In vivo analysis of the effect of Panobinostat on cell-associated HIV RNA and DNA levels and latent HIV infection.* Retrovirology, 2016. **13**(36): p. 3560-70.

112.   Johnstone, R.W., *Histone-deacetylase inhibitors: novel drugs for the treatment of cancer.* Nat Rev Drug Discov, 2002. **1**(4): p. 287-99.

113.   West, A.C. and R.W. Johnstone, *New and emerging HDAC inhibitors for cancer treatment.* J Clin Invest, 2014. **124**(1): p. 30-9.

114.   Dreesen, O. and A.H. Brivanlou, *Signaling pathways in cancer and embryonic stem cells.* Stem Cell Rev, 2007. **3**(1): p. 7-17.

115.   Garnock-Jones, K.P., *Panobinostat: first global approval.* Drugs, 2015. **75**(6): p. 695-704.

116. Hockemeyer, D., et al., *Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases.* Nat Biotechnol, 2009. **27**(9): p. 851-7.

117. Hendel, A., et al., *Quantifying on- and off-target genome editing.* Trends Biotechnol, 2015. **33**(2): p. 132-40.

118. Iyer, V., et al., *Off-target mutations are rare in Cas9-modified mice.* Nat Methods, 2015. **12**(6): p. 479.

119. Smith, C., et al., *Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs.* Cell Stem Cell, 2014. **15**(1): p. 12-3.

120. Trounson, A. and N.D. DeWitt, *Pluripotent stem cells progressing to the clinic.* Nat Rev Mol Cell Biol, 2016. **17**(3): p. 194-200.

121. Hanna, J., et al., *Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin.* Science, 2007. **318**(5858): p. 1920-3.

122. Tsuji, O., et al., *Therapeutic potential of appropriately evaluated safe-induced pluripotent stem cells for spinal cord injury.* Proc Natl Acad Sci U S A, 2010. **107**(28): p. 12704-9.

123. Wernig, M., et al., *Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease.* Proc Natl Acad Sci U S A, 2008. **105**(15): p. 5856-61.

124. Kobayashi, Y., et al., *Pre-evaluated safe human iPSC-derived neural stem cells promote functional recovery after spinal cord injury in common marmoset without tumorigenicity.* PLoS One, 2012. **7**(12): p. e52787.

CHAPTER 4 – Aim 3 Unravelling the genetic background of Rubinstein-Taybi Syndrome using Next Generation Sequencing

Introduction

Rubinstein-Taybi Syndrome (RTS) is a congenital disease affecting ~1 in 100,000 live births [1]. It is characterized by intellectual disability, growth delay and specific facial features [2] in addition to other malformations and behavioural problems that can vary greatly between RTS patients. The majority of cases are due to *de novo* germline mutations affecting a single copy of either *CREBBP* [3] or *EP300* [4]. RTS is autosomal dominant and a mutation in one of these genes is seen in roughly 65% of cases. Nearly all mutations in *CREBBP* and *EP300* are *de novo*, however, even though very rare, there are reports of inheritance and recurrence in families [1, 5-9]. The mutation distribution is uneven, with *CREBBP* mutations being significantly more frequent than *EP300* mutations (~60% and <10% of cases respectively) [10, 11].

CREBBP and *EP300* code for CBP [12] and p300 [13] respectively, which are highly homologous proteins that play major roles in gene regulation. These two proteins act as histone acetyltransferases (HATs) and coactivators, interact with transcription factors and are able to form a complex with RNA polymerase II [14]. Both CBP and p300 interact with many genes and proteins, with >400 interactions described [15, 16].

A variety of mutations have been detected in RTS cases, ranging from single nucleotide variants to large deletions that remove can remove the entire gene [17]. However, the majority of variants are point mutations, small deletions or duplications of one to several base pairs, which may cause premature translation stops, amino acid substitutions or splicing defects [18]. Analysis has shown that most RTS-causing missense mutations in *CREBBP* are located within the HAT domain, which indicates that disruption of the HAT domain is sufficient to cause RTS [19, 20]. This suggests that the levels of CBP and/or p300 are important and that happloinsufficiency of these proteins is at the basis of RTS [21].

Even though the small number of RTS cases with mutations in this gene makes direct comparisons difficult, it appears that RTS patients with a mutation in *EP300* are less severely affected. This is particularly seen in the skeletal abnormalities, such as the absence of the broad thumbs and big toes and can lead to misdiagnosis of RTS patients [4, 7, 10, 22, 23]. As an initial RTS diagnosis is typically based on the clinical presentation, it is possible that cases with milder and/or atypical RTS features

may be misdiagnosed which might be more relevant for RTS cases involving *EP300*. *EP300* variants have been identified in individuals not originally diagnosed with RTS, but instead were suspected of having Cornelia de Lange syndrome (CdLS) [24, 25]. CdLS is a genetically heterogeneous syndrome with many functions of the different proteins involved linked with chromatin structure, like CBP and p300. Another example is Floating Harbour Syndrome (FHS), caused by mutations in the *SRCAP* (SNF2-related CREBBP activated protein) gene. *SRCAP* encodes a chromatin-remodeling protein that contains several functional domains and is a known CBP and/or p300-interacting protein.

There are several mouse models with knockouts of either *Crebbp* or *Ep300* to mimic RTS and these have shown that the HAT domain is highly conserved between species, and mutations in this domain of the mouse *Crebbp* gene leads to RTS like symptoms [4]. Mouse studies have also shown that p300 heterozygous knockouts have an increased incidence of lethality compared to wild type mice, whereas cbp heterozygotes do not or at a much lower rate, depending on genetic background [21, 26, 27]. This could explain the relatively low frequency of *EP300* mutations seen in RTS, as mutations in this gene could lead more often to a miscarriage in humans too. Even though CBP and p300 have numerous overlapping functions and targets, they also exhibit unique roles [16, 27, 28], especially during embryogenesis [29, 30]. Therefore, another hypothesis is that partial loss of p300 due to a mutation is compensated for by the recruitment of CBP, leading to overall depletion of CBP in the cell, resulting in RTS [4].

Somatic and germline mosaicism [6, 8] in RTS have been reported, in which the parents are clinically unaffected or mildly affected, but offspring diagnosed with RTS [31-33]. These mosaics mutations in either *CREBBP* or *EP300*, suggests these individuals tend to have a less severe phenotype and this could lead to misdiagnosis. These findings of mosaicism highlight the variable phenotypes of RTS [6] and can be a cause of failure to diagnose in a selection of RTS cases, which can have considerable impact on both patients and their families [34]. The identification of causative mutations and genes not only confirms diagnosis but also enables the understanding of gene functions and biological pathways [35] and may be a first critical step towards proper therapeutic intervention and family counselling [36].

Even though most cases of RTS are based on clinical presentation, the disease can be confirmed by genetic screening [2]. The first genetic tests for RTS were done in 1991,

where a *de novo* reciprocal translocation with breakpoints in chromosomal region 16p13.3 was identified in a group of patients [37-39]. Subsequently, in 1995, Fluorescent *In Situ* Hybridization (FISH) was performed in combination with additional research and this led to the discovery of RTS causing mutations in the *CREBBP* gene [3]. Additionally, mutations in *EP300* were identified in the first three RTS cases in 2005 [4].

However, diagnosing diseases just by phenotypic features and conventional diagnostic testing remains a challenge; in an average clinical setting the diagnostic rate is ~50%, and even lower for rare diseases [40]. Similar numbers are seen for diseases for which genetic testing is available, where a disease causing variant can be identified in only 52% of cases overall (called the molecular diagnostic rate) [41, 42].

The golden standard in genetic testing has been Sanger sequencing, which was first described in 1977 [43]. This technique has been used for many different applications and has undergone continuous improvements over time. The development of more advanced DNA sequencing methods, such as parallel sequencing, has allowed Sanger sequencing to be used in larger scale productions of genomic sequences and projects [44]. In combination with better instruments, techniques and bioinformatics this has led to increased efficiency and accuracy. Regardless of this progress, it seems unlikely further substantial improvements in Sanger sequencing will result in increases in throughput and decreases in cost [34].

Besides Sanger sequencing, other alternative techniques are frequently used in genetic diagnostics, such as FISH. This method is regularly used in RTS diagnostics, and relies on large rearrangements at the chromosomal level. Multiplex ligation-dependent probe amplification (MLPA) is a technique that screens and detects copy number variations (CNVs) [45] and was used in the discovery of the *EP300* gene in RTS [4]. Finally, chromosomal microarray analysis (CMA) is an array-based form of comparative genomic hybridization and can detect clinically significant structural changes, particular deletions and duplications in the genome.

Although traditional gene mapping approaches (including karyotyping [46], linkage analysis [47], homozygosity mapping [48] and CNV analysis [49]) have made enormous contributions to insights in genetic diseases, they are unable to detect all forms of genomic variation and are mainly focused on gene coding regions of the genome [34]. These techniques detect relatively large rearrangements, and are not likely to identify smaller variants in the DNA sequence, such as point mutations,

which are the cause of the majority of RTS cases. In addition, these genetic tests are chosen based on specific clinical presentations of individual patients and most include only one or a small set of genes known to be associated with a particular syndrome. However, in the last decade, disease gene identifying research has received a major boost by the introduction of Next Generation Sequencing (NGS). NGS enables thousands of genes to be analyzed simultaneous at substantial higher throughput and lower cost than previously possible. There are different applications of NGS, such as whole genome sequencing (WGS) and whole exome sequencing (WES). These sequencing techniques, will make the entire protein coding DNA sequence available, which allows for the analysis of the ~20,000 genes in the human genome in one assay. There is also the possibility to sequence smaller subset of genes with NGS, such as a 'mini-genome' or a panel of genes, relevant to a disease phenotype. With the exception of WGS, these applications involve targeted enrichment of DNA in combination with massively parallel sequencing. Besides being lower in costs this will lead to a higher sequencing coverage/depth than WGS and, therefore, highly accurate DNA variant calling for the region of interest [50]. The reduction in costs, both in sequencing, and reduced storage and analysis expenses, makes it feasible to increase the number of samples to be sequenced. This enables large population based comparisons and makes WES the most commonly applied approach at present.

Nonetheless, with unlimited resources and time, WGS is the obvious choice as it sequences both protein coding and non-coding sequences (~1-2% and ~99% of the human genome, respectively). This allows studying all single-nucleotide variants (SNVs), indels, structural variants (SVs) and CNVs, in the whole genome, including regulatory regions such as promoter and enhancers, which are excluded in WES. However, due to technically challenging regions of the genome for sequencing, such as high GC content, large repeat regions, centromeres and telomeres, in reality WGS covers 95-98% of the genome. Nonetheless, due to the elimination of the targeting process, WGS is more reliable and generates a more uniform coverage of the genome. Differences in the hybridization efficiency of capture probes, can result in regions of the genome with little or no coverage and off target capture effects. However, WES frequently requires an amplification step during library preparation, as the input needed to capture is generally ~1μg DNA, which can increase the potential of GC bias.

NGS technologies include a number of steps that are grouped as template preparation, sequencing, imaging, and data analysis. The unique combination of these specific protocols determines the type of data produced from each platform. As mentioned before, WES requires a targeting step using a library called baits, which usually correspond to most exonic sequences and can be either DNA or RNA. Relatively little input DNA is needed (a few micrograms) to produce a library, and to generate this, gDNA isolated from patients is firstly fragmented into small segments by sonication and linked to artificial tags. These fragments are hybridized to a library of baits and enriched in a pull down process with the resulting fragments representing most of the protein-coding portion of the genome. Both ends of the DNA fragments are ligated to specific adaptor oligos and sequencing of these fragments will occur in parallel using an *in situ* amplification method. This typically involves fluorescent dyes, which are read by a laser. The generated sequence reads are assembled as a series of overlapping fragments and aligned to a reference sequence, in this case the human genome [51-53]. Differences between patient and reference genomes (variants) can then be identified.

The main advantage of NGS is the production of an enormous quantity of sequencing data for a relatively low cost. Furthermore, by using unique adaptors, it is possible to pool several patient samples and run them together in a single lane. This addition of specific tags or barcodes will lower running costs, while pooled samples can be analyzed biologically independent [54].

An important consideration for NGS is coverage or sequencing depth, which refers to the average number of reads in which each nucleotide is sequenced and provides an estimation of how likely a variant at a specific position would be detected. These variants will not always be pathogenic mutations, as most of them represent benign polymorphisms (SNPs). There are several effect prediction algorithms, such as SIFT and PolyPhen, that will calculate the pathogenic likelihood of a variant. Besides this, several filtering techniques can be used to narrow down the list with variants to identify potential candidate genes. Even though NGS results are considered of high quality, in practice, identified variants are still being validated using Sanger sequencing.

As the cost of NGS has fallen significantly, it becomes more feasible for smaller laboratories to adopt this technique as a routine procedure. However, there are still initial costs such as the sequencing machines themselves, and the associated cost of

data analysis (bioinformatics expertise) and storage. In addition, because of the enormous amount of data and potential disease-causing variants, geneticists need appropriate training in how to interpret the data and how to report this to patients.

In conclusion, the availability of the human genome sequence and NGS has made the identification of disease-causing variants with little clinical association possible [55]. WES in particular has led to the identification of causative mutations in a large number of analysis, and has given new insight into rare genomic variation [56, 57].

Despite this progress, finding relevant mutations remains a complex task. The diagnostic rate reported with NGS is currently only ~25-30%, which however is still substantially higher than both karyotyping (<5%) and array CGH (~15-20%) [42]. Nevertheless, the diagnostic rate for NGS is expected to increase over the coming period as research into genes and variants is continuing.

The main objective of this chapter will be the combination of NGS and exome analysis to identify the genetic basis of RTS cases without a mutation in either *CREBBP* or *EP300*. DNA samples taken from RTS patients will be processed using whole exome sequencing and the data generated will be analyzed to identify variants that may be causative for RTS.

Genomic DNA samples of RTS patients clinically diagnosed with RTS but without a genetic diagnosis, were collected and sequenced, and it is expected to identify one or more samples with an RTS causing variant in *CREBBP* and/or *EP300*. For samples without a variant in the two known RTS genes, genes that are known to be involved in related diseases will be screened for variants. Since RTS has overlapping clinical features with other syndromes and diseases, variants in genes will be selected that cause conditions with similar phenotypes, with particular interest in genes involved in epigenetic mechanisms.

The next step is to detect variants in genes known to interact with CBP and/or p300. There are more than 400 interaction described with these two proteins, listed in the Interactome (Appendix 2). These proteins are either regulated through direct interactions with CBP and/or p300 or by gene expression control. After whole exome sequencing, variants in genes will be compared with the Interactome and with a list composed of genes known to be involved in intellectual disability and/or epilepsy. Finally, genes found with variants in more than one patient are considered strong candidates for causative RTS genes. Combining this data, candidate gene list for RTS based on these parameters will be generated. This list will be mainly based on non-

synonymous variations, such as frameshift insertions and deletions and stop gain SNVs, and PolyPhen and SIFT scores will be taken in consideration for missense variants. Finally, interesting variants will be confirmed by Sanger Sequencing in combination with parental DNA samples when available. The pipeline designed for finding candidate genes for RTS is summarized as a schematic overview in Figure 1. However, exome analysis is primarily focused on sequence variants, but RTS can also be caused by CNVs. Therefore, the RTS DNA samples are also screened by MLPA.

Figure 1. Analysis procedure to select candidate genes. Sequencing data from RTS patients will go through this pipeline to determine the genetic cause in these cases. Firstly, RTS causing variants in either *CREBBP* or *EP300* are excluded and thereafter genes known to interact with these two proteins (the Interactome), and also genes involved in intellectual disability and/or epilepsy will be screened. The effect that each specific variant will have (SIFT and PolyPhen scores) will also be taken in consideration. With this information, a candidate gene list for RTS is generated which will be validated with Sanger Sequencing.

Material & Methods

Genomic DNA samples

In total, 12 samples from individuals clinically diagnosed with RTS were collected (Table 1). DNA buccal swabs were send to RTS cases and their parents, and returned for genomic DNA extraction using the IsoHelix kit (Astral Scientific DDK-50SK1S). In four cases it was also possible to receive a blood-derived DNA sample. Due to the low quality of the gDNA isolated from buccal samples, we were only able to sequence two of these samples (RTS3 and 7) in addition to the four blood gDNA samples (RTS1, 10, 11 and 12). DNA concentrations ranged from 1030-56700ng.

Table 1. RTS patient DNA samples

RTS case number	gDNA sample	DNA yield (ng)
RTS1*	Blood	6250
RTS2	Buccal	1250
RTS3*	Buccal	2380
RTS4*	Buccal	8260
RTS5	Buccal	3290
RTS6	Buccal	5600
RTS7*	Buccal	1296
RTS8	Buccal	5830
RTS9	Buccal	1030
RTS10	Blood	6200
RTS11	Blood	39000
RTS12	Blood	56700

*Parental buccal DNA was collected when available.

In addition, the two RTS primary fibroblast cell lines from Chapter 2 with known mutations in *CREBBP* were included in this analysis. Cells were grown under standard conditions and genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, 69504). These cases and corresponding mutations have been published previously (Table 2).

Table 2. RTS fibroblast cell lines

Individual as published	This study	*CREBBP*	Mutation (Deletions)	Amino Acid change
213-1 [20]	RTS-NL1	exon 29	c.4837 del G	V1613fsX1634
199-3 [4]	RTS-NL2	exon 3	c.904_905del AG	S302fsX348

The mutations in Table 2 are denoted as published in Roelfesma et al., 2005 [4]. This is in accordance with the nomenclature recommendations published in den Dunnen and Antonarakis [58]. CBP mutations are described in relation to GenBank sequence NM_004380, with the A of the ATG start codon counted as nucleotide 1. Changes on the protein level as stated in the last column of Table 2 are predictions.

Next Generation Sequencing

DNA samples were processed and sequenced by the Monash Health Translation Precinct (MHTP) Medical Genomics Facility, Monash University, Melbourne, Australia. Sequencing platforms used where the SOLiD (Applied Biosystems), HiSeq 1500 (Illumina) or the Ion Proton (Life Technologies). Samples were collected and processed over an extended period of time, therefore all sequencing preparation was performed with kits available at the time of sequencing. Sample RTS10 was sequenced three times, each time using a different targeting kit in combination with the corresponding sequencing platform (Table 3).

Table 3. WES sample sequencing platform and capture kit details

Sample	Sequencing platform	Library preparation kit / Targeting kit / Exome capture kit
RTS1	Proton	TargetSeq (Life Technologies)
RTS3	Proton	AmpliSeq (Life Technologies)
RTS7	Proton	AmpliSeq (Life Technologies)
RTS10 AS	Proton	AmpliSeq (Life Technologies)
RTS10 TS	Proton	TargetSeq (Life Technologies)
RTS10 HP	Illumina	HaloPlex (Agilent)
RTS11	Proton	AmpliSeq (Life Technologies)
RTS12	Proton	AmpliSeq (Life Technologies)
RTS-NL1	SOLiD	TargetSeq (Life Technologies)
RTS-NL2	Proton	AmpliSeq (Life Technologies)

RTS10 was sequenced three times, using different exome capture kits. AS = AmpliSeq, TS = TargetSeq, HP = HaloPlex.

Data analysis

After sequencing, reads were aligned to the human genome (Genome Reference Consortium human genome build 37, human genome 19, GRCh37/hg19, which was the current genome database when the samples were sequenced). Generated sequencing files were annotated with Ensembl, which uses the Variant Effect Predictor (VEP) as a default. VEP takes SNPs, insertions and deletions (indels), CNVs and structural variants into account and determines the effect of these variants on genes, transcripts, and protein structures. This will generate VCF output files with reported annotations on every genomic feature of the input variant and will show separate readings for genes with multiple alternative splicing variants (isoform transcripts). However, this depth of annotation was not required for this study and therefore the following annotation was used to reduce the amount of output produced. The generated VCF files were annotated using the following Ensembl annotation summary, called 'Pick' (http://asia.ensembl.org). This VEP selects one line of annotation per variant, using an ordered set of criteria and prioritizes if the transcript is canonical or not, with a preference for the canonical one (longest transcript). If the variant does not have the same effect (SIFT and PolyPhen), the noncanonical transcript is also included. Only protein coding transcripts are included and consequences are ranked according to Ensembl (Appendix 1).

Sanger sequencing

Primers were designed for regions in genes that contain specific variants and were amplified in a PCR reaction. Sanger sequencing was performed at the MHTP Medical Genomics Facility, Monash University, Melbourne, Australia, to confirm variants of interest in the RTS samples.

MLPA

Probes were designed for exons in *CREBBP* and *EP300* and evenly spaced throughout the gene (Table 4). *CREBBP* probes were labelled with the 5' GGGTTCCCTAAGGGTTGGA 3' and 5' TCTAGATTGGATCTTGCTGGCGC 3' upstream and downstream hybridizing sequences respectively, that includes a blue fluorescent label. For *EP300* 5' GGCCGCGGGAATTCGATT 3' and 5' CACTAGTGAATTCGCGGC 3' was added as hybridizing sequences, containing a

green fluorescent label. Per mix 3 to 4 control probes were added as shown in Table 5. Samples were separated on the ABI3700 capillary sequencer (Applied Biosystems) at the MHTP Medical Genomics Facility, Monash University, Melbourne, Australia. and analysed with Coffalyser software.

Table 4. MLPA probe sequences for *CREBBP* and *EP300*

Location	Upstream hybridizing sequence	Downstream hybridizing sequence	Total probe length
CREBBP 3' UTR	CAGCAGCGGATTCTGCAGCAACA	GCAGATGAAGCAGCAGATTGGGC	88
CREBBP exon 1	GCAGGTGAAAATGGCTGAGAACTTGCT	GGACGGACCGCCCAACCCCAAAAGAGCCAAA	100
CREBBP exon 2	CAAACATAAACAACTGTCGGAGCTTCTACGAGGA	GGCAGCGGCTCTAGTATCAACC	98
CREBBP exon 4	GAATTGTACCCACACAAGCAATTGCAACAG	GCCCCACTGCAGATCCTGAAAC	94
CREBBP exon 6	CCAGCTAGTGGAATTCAAAACACAATTGGTTCTGTTGGCACA	GGGCAACAGAATGCCACTTCTTTAAGTAAC	114
CREBBP exon 9	CAACACCTGATCCCGCAGCTCT	AAAGGATCGCCGCATGGAAAACCTCC	90
CREBBP exon 14	CGGCTGCTGGCATGCCATCTCTCCA	GCACACGACACCACCTGGGATGACC	92
CREBBP exon 16	CAAGTTAAAGAAGAAACAGACATAGCAGAGCAGAAATCAG	AACCAATGGAAGTGGATGAAAAGAAACCTGAAGTGAAAGT	122
CREBBP exon 18	GCTTGCAGAGGTCTTTGAGCAGGAAATTGACCCT	GTCATGCAGTCCCTTGGATATTGCTG	102
CREBBP exon 20	GTATCATTTCTGTGAGAAGTGTTTCACAG	AGATCCAGGGCGAGAATGTGACCCTGGGTGACGACCCTT	110
CREBBP exon 25	GTTTGTGGATTCTGGGGAAATGTCTGAATCTT	TCCCATATCGAACCAAAGCTCTGTTTGCTTTTGAGGAA	112
CREBBP exon 27	CCAAGTGAAGGAGATGATTACATCTTCCATTGCCACCCACCT	GATCAAAAAATACCCAAGCCAAAACGACTGCAGGAG	120
CREBBP exon 28	CAAGCAACTGAAGACAGGCTCACCAGT	GCCAAGGAACTGCCCTATTTTGAAGGT	96
EP300 exon 1	GCCGAAGAAGAGATTTCCTGAGGATTCT	GGTTTTCCTCGCTTGTATCTCCGC	88
EP300 exon 2	GGGACTAACCAATGGTGGTGATATTAATCAGCT	TCAGACAAGTCTTGGCATGGTACAAGATG	98
EP300 exon 3	CCATATACTCAGAATCCTGGACAGCAGATTGGAGCCA	GTGGCCTTGGTCTCCAGATTCAGACAAAC	102
EP300 exon 7	GTGCTAGTCCTATGGGAGTAAATGGAGGTGTAGG	AGTTCAAACGCCGAGTCTTCTTTCTGACTCAATGTTGCAT	110
EP300 exon 11	GCCCTCTACCTGACCCAAGTATGAT	CCGTGGCAGTGTGCCAAACCAGATGATGC	90
EP300 exon 14	CAGACACCTACACCACCAACAACACAACT	TCCCCAACAAGTGCAGCCTTCACTTCCTC	94
EP300 exon 17	CACTGATGCCAACTTTGGAGGCACTTTACC	GTCAGGATCCAGAATCCCTTCCCTTTCGTCAACCTGTC	104
EP300 exon 23	GGTTGCCATCTACCAGACTTGGCACCTTTCTAGAG	AATCGTGTGAATGACTTTCTGAGGCGACAGAATCC	106
EP300 exon 31	GGTATCAGCCCACTCAAACCAGGCACTGT	GTCTCAACAAGCCTTACAAAACCTTTTGCGC	96

Table 5. MLPA control probe sequences

Location	Upstream hybridizing sequence	Downstream hybridizing sequence	Total probe length
RSPOI exon 8	CATGCAGAAAGAGTTCAGTGCTACTCTGCGTGA	TTCAAGCTTTCCTGAACTGGAACGTCGGGGGCA	106
CYCLIN E exon 11	GTGCGACATAGAGAACTGTGTCAAGTGGATGGTT	CCATTTGCCATGGTTATAAGGGAGACGGGGAGCT	108
RSPOI exon 4	GCTGCTGCTCTCACCGCTGAGCGGATAAAGACCCCAC	GAGGACCTCCCTTTGAGCTGATGGAGAACTGG	110
EZH2 exon 9	GCTGCTGCTCTCACCGCTGAGCGGATAAAGACCCCAC	CAAAACGTCCAGGAGGCCGCAGAAGAGGACGGCTTCC	114

Results

Whole Exome Sequencing of RTS cell lines

Firstly, the two fibroblast cell lines with known RTS mutations in the *CREBBP* gene were sequenced. RTS-NL1 was sequenced on the SoLID whereas sample RTS-NL2 was processed on the Proton. The number of variants found in the samples differed widely and ranged between ~30,000 to over 80,000, which is sample specific but also due to the different exome capture techniques and sequencing platforms used (Table 6).

Table 6. Sequencing details for RTS fibroblast lines

General statistics	RTS-NL1	RTS-NL2
Variants processed (all)	82255	29229
Lines of output written (coding)	22501	20471
Novel	15554	1592
Known	66701 (81.1%)	27637 (94.6%)

The known RTS causing variants in the *CREBBP* gene were confirmed with WES, consistent with the previously published results for these two patients. The mutations were both deletions, of either 1 or 2 base pairs, predicted to cause frameshifts in the transcripts (Table 7).

Table 7. Variant annotation for RTS fibroblast cell lines with mutations in *CREBBP*

Sample	Location	Feature_type	Consequence	Amino_acids	Codons
RTS-NL1	16:3781830	Transcript	frameshift_variant	V/X	Gtg/tg
RTS-NL2	16:3860674 -3860675	Transcript	frameshift_variant	S/X	AGc/c

The Integrative Genomics Viewer (IGV), developed by the Broad Institute, was used to look at the coverage of specific variants. IGV is a visualization tool for genomic data sets generated by next generation sequencing [59]. The IGV plots in Figure 2 represent RTS-NL1 and RTS-NL2 with their respective mutations shown in the green boxes. The red and blue lines in the plots represent individual reads, forward and reverse respectively, with arrows indication the direction and end points of reads. White spaces indicate the absence of (overlapping) reads, and black bars in these

breaks represent deletions in the reads. Both heterozygous variants are recognized by the grey bars at the top of the graph as they have a reduced height of approximately 50% compared to the neighbouring nucleotides. This is also represented in the black bars, showing a deletion in roughly 50% of the reads compared to a C for RTS-NL1 and a C and T for RTS-NL2.

This software also shows the total read counts for each nucleotide. This is displayed for the fibroblast lines in Table 8. For RTS-NL1, 20% of reads (5/25) showed the deletion, and for RTS-NL2 this is 42% (37/88) and 44% (38/87) of the reads respectively.

Table 8. Coverage of variants in RTS fibroblast cell lines

Sample	RTS-NL1	RTS-NL2	
Total counts	25	88	87
A	1 (4%, 1+ 0-)	0	0
C	24 (96%, 17+ 7-)	88 (100%, 62+ 26-)	0
G	0	0	0
T	0	0	87 (100%, 62+ 25-)
N	0	0	0
del	5	37	38
ins	0	0	0

Next, primers for the region of the variants were designed to confirm the mutations with Sanger sequencing (Figure 3). Again, the Sanger sequence trace shows that both cell lines have heterozygous frameshift mutations as the traces are out of sync after the mutations occur (see arrows).
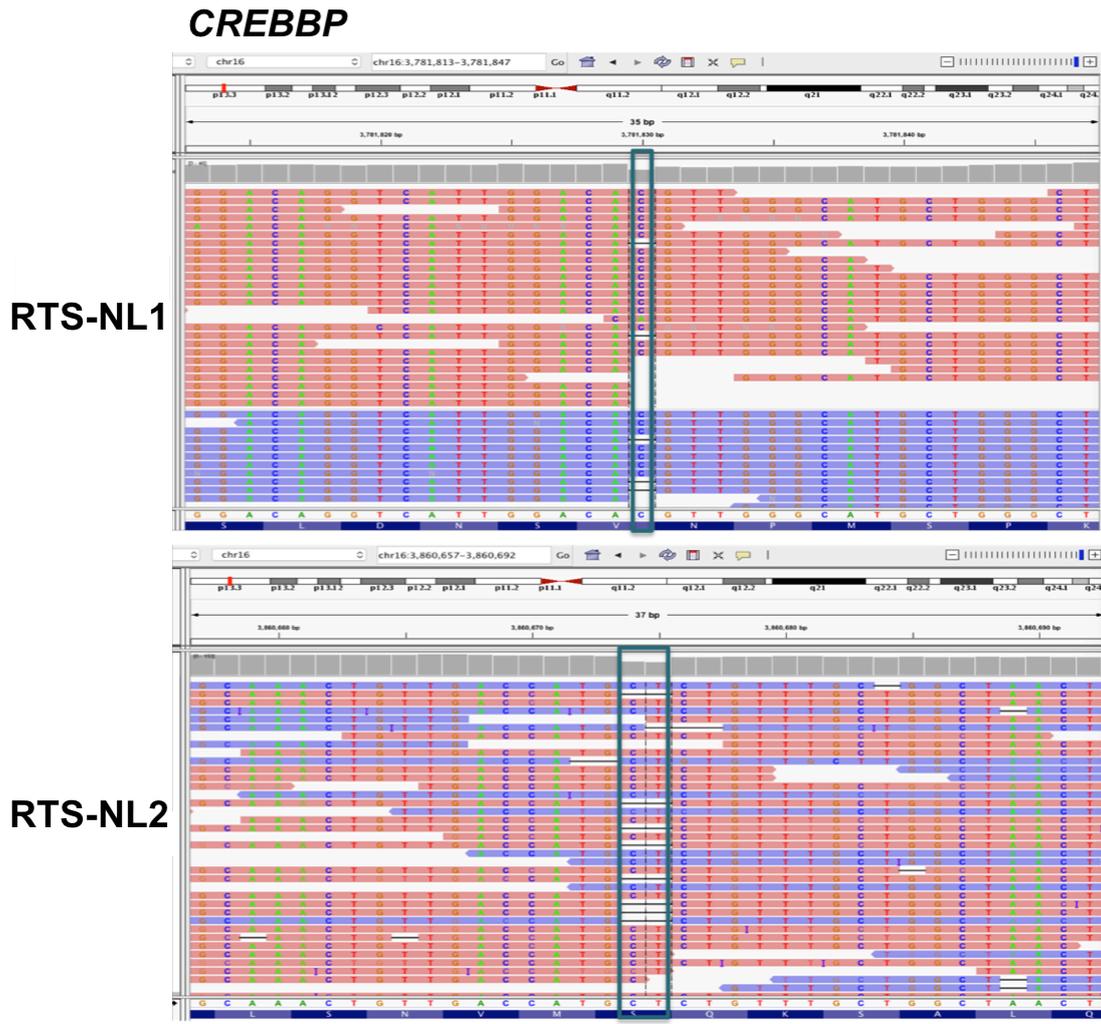
CREBBP

Figure 2. IGV plot for *CREBBP* frameshift mutation in the RTS fibroblast cell lines. IGV plots showing part of the *CREBBP* gene, located on chromosome 16, with the variants shown in the green boxes. The top plot shows the heterozygous frameshift deletion of 1 base pair (C), at location 16:3781,830 for RTS-NL1. The bottom plot shows the heterozygous frameshift deletion of 2 base pairs (CT), at location 16:3860674-3860675 for RTS-NL2. Red and blue rows represent forward and reverse reads respectively, and white breaks with black bars represent deletions, whereas white breaks alone represent the absence of reads.
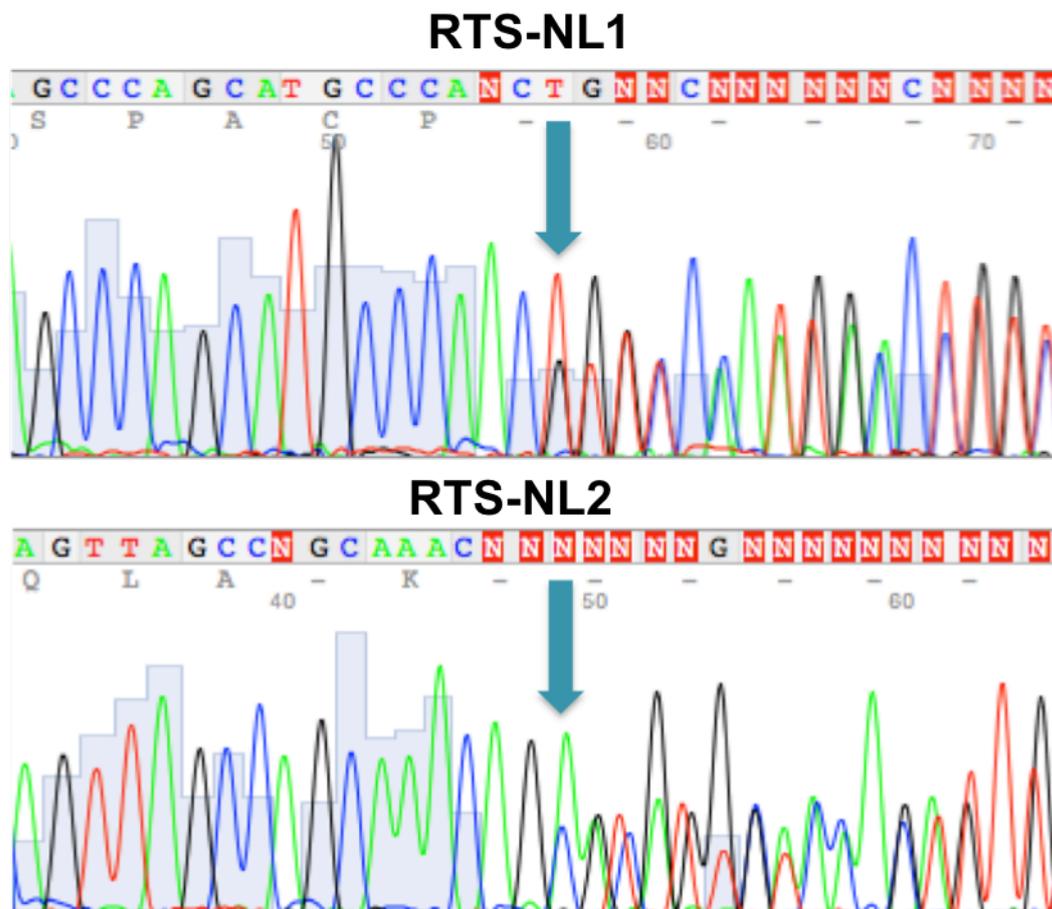
Figure 3. Sanger sequence traces for RTS fibroblast cell lines. These Sanger sequencing plots show traces for the heterozygous frameshift mutations indicated by the green arrows for RTS-NL1 (top) and RTS-NL2 (bottom).

Whole exome sequencing of RTS samples

Next, WES was performed on the collected DNA samples from patients who are clinically diagnosed with RTS, but without a genetic diagnosis. Of the 12 collected DNA samples, six samples were processed for whole exome sequencing as shown in Table 9 and these samples were selected on DNA quality. It appeared that even though the collection of buccal samples is a low invasive method, the quality of DNA extracted from these samples is generally too low for NGS.

Table 9. Whole exome sequencing details for RTS patients

General statistics	RTS1	RTS3	RTS7	RTS10	RTS11	RTS12
Sex	F	F	M	M	F	M
Variants processed	36323	32243	16679	52907	50608	50824
Lines of output written (after filtering, in coding DNA)	24349	21944	11242	16177	20887	20627
Novel	3500 (14.4%)	1960 (8.9%)	1416 (12.6%)	1361 (8.4%)	2354 (11.3%)	2318 (11.2%)
Existing	20849 (86.6%)	19984 (91.1%)	9826 (87.4%)	14816 (91.6%)	18533 (88.7%)	18309 (88.8%)

Table 9 displays the total number of variants per sample. The variants processed include all changes identified in each patients personal genome when aligned to the reference haploid human genome sequence, GRCh37/hg19. A filtering strategy as outlined in Figure 4 was used to process the variants.

The number of variants processed varies and this depends on the sample, in combination with the targeting kits and sequencing platforms used. The lines of output written include all the variants called in coding DNA (protein) whereas the variants processed include both coding and non-coding DNA. The number of novel variants (excluding all known SNPs) was approximately 10% of output variants (Figure 5).

Variants can be categorized according to the coding consequences and this is displayed in Table 10. Even though the number of variants for each sample varies considerably, missense variants were called most frequent in all samples sequenced, followed by synonymous and frameshift variants, which are graphed in Figure 6.
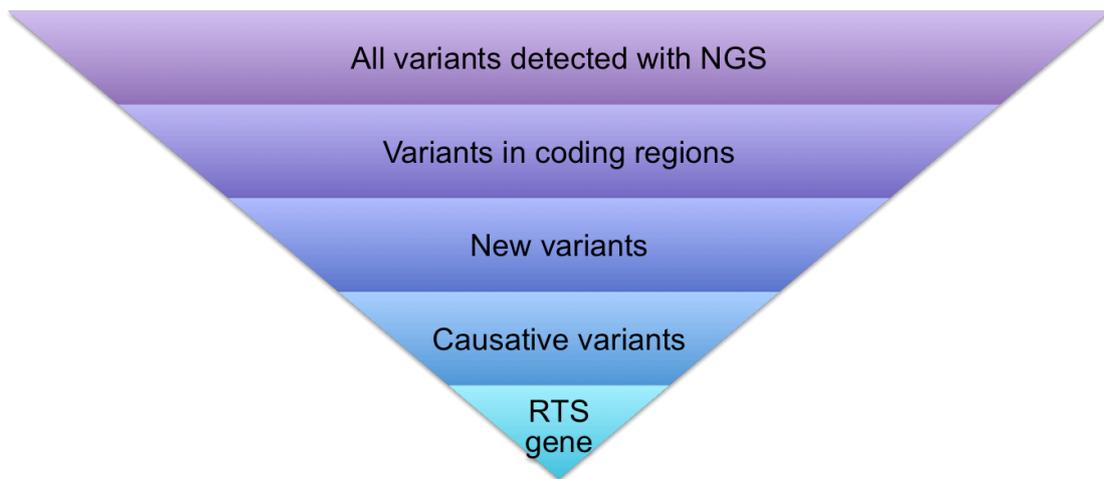
Figure 4. Filtering strategy after WES. After the generated WES data is aligned to the human genome, all variants detected will be listed. These will be filtered for coding regions, new variants (not reported as SNPs), causative variants (SIFT and PolyPhen scores) which will finally lead to a list of potential candidate RTS genes.
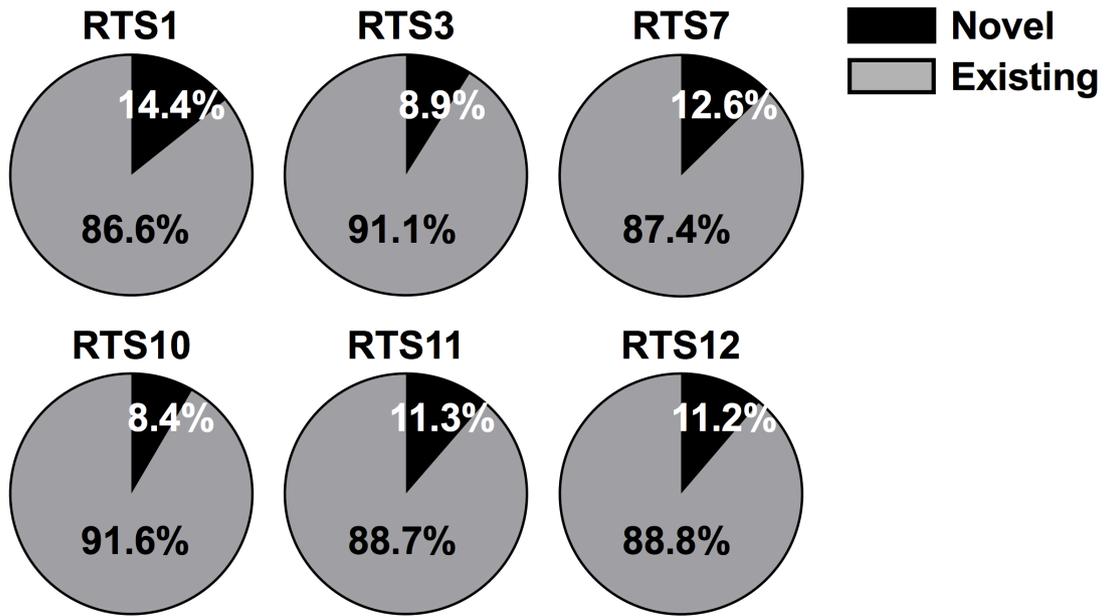
Figure 5. Novel and existing variant ratios per RTS sample. The pie charts show the ratio between novel (black) and existing (grey) variants per RTS sample sequenced. The novel variants represent approximately 10% of the total variants called per sample.

Table 10. Coding consequences for RTS samples

Coding consequences	RTS1	RTS3	RTS7	RTS10	RTS11	RTS12
Coding sequence variant	1	1	1	1	4	4
Frameshift variant	760	119	495	278	519	455
Inframe deletion	16	9	25	27	10	12
Inframe insertion	1	0	2	5	3	1
Missense variant	1537	995	461	512	1003	1040
Splice region	26	11	7	7	18	17
Start lost	10	2	3	1	1	2
Stop gained	22	17	5	15	24	33
Synonymous variant	1114	799	409	507	762	746

Coding consequences
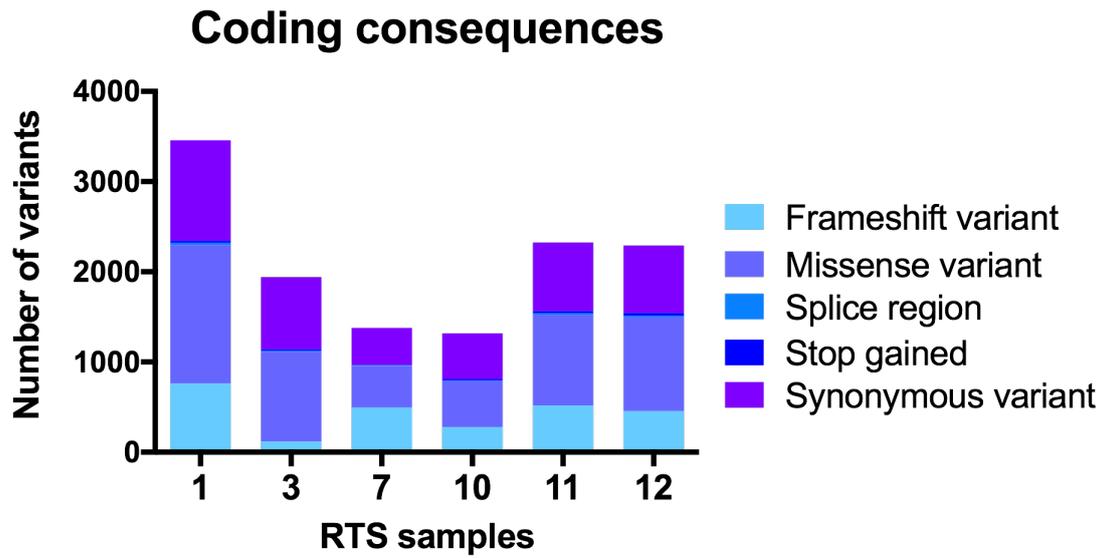
Figure 6. Coding consequences for RTS samples. The total number of variants found in the coding regions of the RTS samples varies. Most variants are non-synonymous mutations, such as frameshift and missense variants.

When looking at variants per chromosome a clear pattern between samples can be seen (Figure 7), which correlates with the gene density of the chromosomes. Chromosomes 13, 18 and 21 contain the lowest gene density, besides the two sex chromosomes, and this is reflected in the low number of called variants in these three chromosomes.

Contrary to the expectations, all female samples showed variants in the Y chromosome. However, these were in noncoding DNA and can be explained by the homology between the X and Y chromosome. Surprisingly though, one female RTS sample (RTS1) had two variants mapped the coding sequence of the Y chromosome. These variants were both in the *SLC9B1P1* gene. However, this is classified as a pseudogene which are not expressed and therefore do not code for protein. This gene shows 94.9% homology to an area on chromosome 4, which is part of the *SLC9B1* gene. In addition, RTS1 also shows two variants in this gene, which codes for a sodium/hydrogen exchanger and transmembrane protein (Table 11). However, the expression of this gene may be limited to testis [60] and therefore most likely variants in this protein will not have an effect in females.

Table 11. Variants mapped to the Y chromosome for sample RTS1 (Female)

Location	SYMBOL	Consequence	Amino_acids	Codons	Existing_variation	SIFT	PolyPhen
Y:13524639	SLC9B1P1	missense variant	E/K	Gaa/Aaa	rs373363143,COSM1491346	-	benign(0.001)
Y:13524664	SLC9B1P1	missense variant	L/F	ttG/ttT	COSM4157072,rs78220120	-	probably_damaging(0.995)
4:103822404	SLC9B1	missense variant	A/V	gCa/gTa	rs58374767	deleterious(0)	probably_damaging(0.955)
4:103911069	SLC9B1	missense_variant	Q/H	caG/caT	rs2715591	tolerated(0.4)	benign(0)
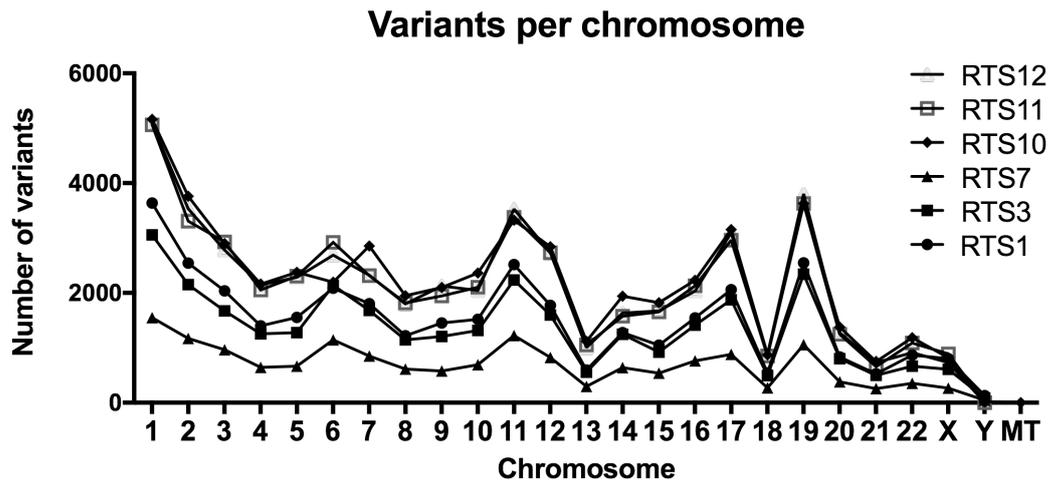
Figure 7. Variants per chromosome. Plotting the variants per chromosome shows a consistent pattern in the RTS samples. The number of variants appears to correlate with gene density per chromosome. MT is mitochondrial genome.

Next, all variants called in *CREBBP* and *EP300* were selected (Tables 12 and 13). When looking at these variants in detail, the majority of these could be excluded as disease causing genes.

Table 12. Variants found in *CREBBP* gene

Sample	Location	Consequence	Amino_acids	Codons	Comments
RTS1	16:3828061-3828062	frameshift_variant	P/X	cCG/c	5xC
RTS1	16:3843577	synonymous_variant	T	acA/acG	Synonymous
RTS10	16:3777726	frameshift_variant	G/X	gGc/gc	3xC
RTS10	16:3777798-3777799	frameshift_variant	R/TX	agg/aCgg	GC region
RTS10	16:3778599	missense_variant	P/L	cCg/cTg	GC region
RTS10	16:3900416	frameshift_variant	P/X	cCt/ct	3xG
RTS10	16:3900422	frameshift_variant	P/X	cCg/cg	CGGC
RTS10	16:3788629	frameshift_variant	P/X	cCa/ca	GC region
RTS10	16:3790471	frameshift_variant	A/X	gcC/gc	GC region
RTS10	16:3795282	frameshift_variant	S/X	Tca/ca	
RTS10	16:3900515	frameshift_variant	G/X	gGc/gc	GC region
RTS12	16:3860674-3860675	frameshift_variant	S/X	AGc/c	

Table 13. Variants found in *EP300* gene

Sample	Location	Consequence	Amino_acids	Codons	Comments
RTS1	22:41551002	missense_variant	F/S	tTc/tCc	4xT
RTS1	22:41551003	synonymous_variant	F	ttC/ttT	Synonymous
RTS10	22:41533693	frameshift_variant	L/X	ctG/ct	3xG
RTS10	22:41536157	frameshift_variant	F/X	Ttt/tt	4xT
RTS10	22:41537117	frameshift_variant	R/X	cgA/cg	
RTS10	22:41547960	frameshift_variant	K/X	Aaa/aa	4xA
RTS10	22:41556652	frameshift_variant	H/X	caT/ca	3xT
RTS10	22:41565581	frameshift_variant	E/X	gAa/ga	3xA
RTS10	22:41523637	frameshift_variant	C/X	tgC/tg	
RTS10	22:41537167	frameshift_variant	V/X	gTt/gt	
RTS10	22:41537173	frameshift_variant	V/X	gTt/gt	3xT
RTS10	22:41574548-41574549	frameshift_variant	M/IX	atg/atCg	

Most variants appeared to be in homopolymers, which is a repeat of the same nucleotide. The mutation in RTS12 was sequenced with Sanger sequencing but was not confirmed. The mutations in RTS10 were only detected with one of the capture kits and therefore excluded. After further analysis only one RTS causing variant was confirmed, in sample RTS3, which is shown in Table 14 and in an IGV plot in Figure 8. This mutation was confirmed with Sanger sequencing and is shown in Figure 9 with the arrow indicating the location of the stopgain variant in sample RTS3.

Table 14. Novel stop gain variant in *CREBBP* in RTS3

Sample	Location	SYMBOL	Feature_type	CANONICAL	Consequence	Amino_acids	Codons
RTS3	16:3786719	CREBBP	Transcript	YES	stop_gained	R/*	Cga/Tga

The information from WES combined with the IGV plot and Sanger sequence trace, shows that the variant involved is a heterozygous mutation in exon 26 of the *CREBBP* gene. This alteration of C>T will cause the change of an Arginine (CGA) into a stop codon (TGA). This premature stop codon will lead to a truncated mRNA transcript and if translated, into a truncated and assumed to be a non-functional CBP protein. In conclusion this variant is predicted to lead to reduced CBP, likely to be responsible for RTS seen in this case.

Genes with variants in the WES data were compared with the gene lists composed from the Interactome (Appendix 2), and genes known to be associated with intellectual disability and/or epilepsy (Appendix 3). In addition, genes with called variants between two or more samples were also selected. Frameshift, missense and stopgain variants found in genes that are included in either the Interactome and/or the ID/epilepsy lists are listed in Appendix 4.
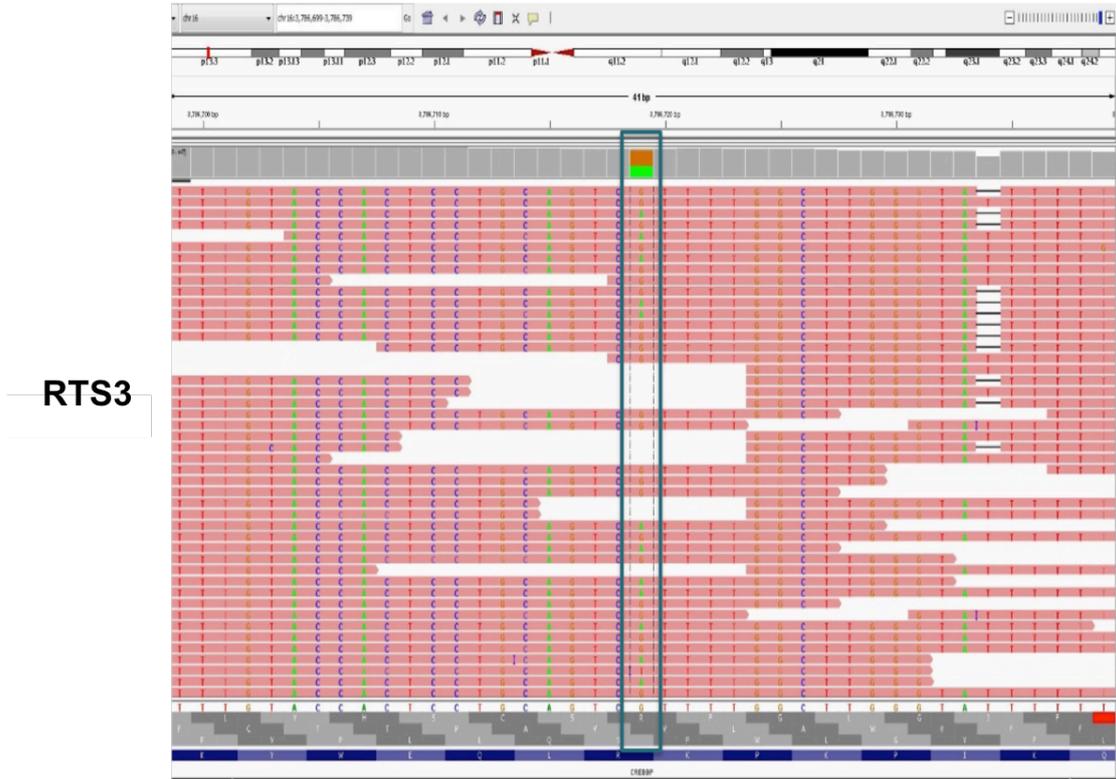
Figure 8. IGV plot showing the stop gain variant in *CREBBP* in RTS3. IGV plot showing the stop gain mutation in the *CREBBP* gene in sample RTS3. The variant is shown in the green box, located at 16:3786719.
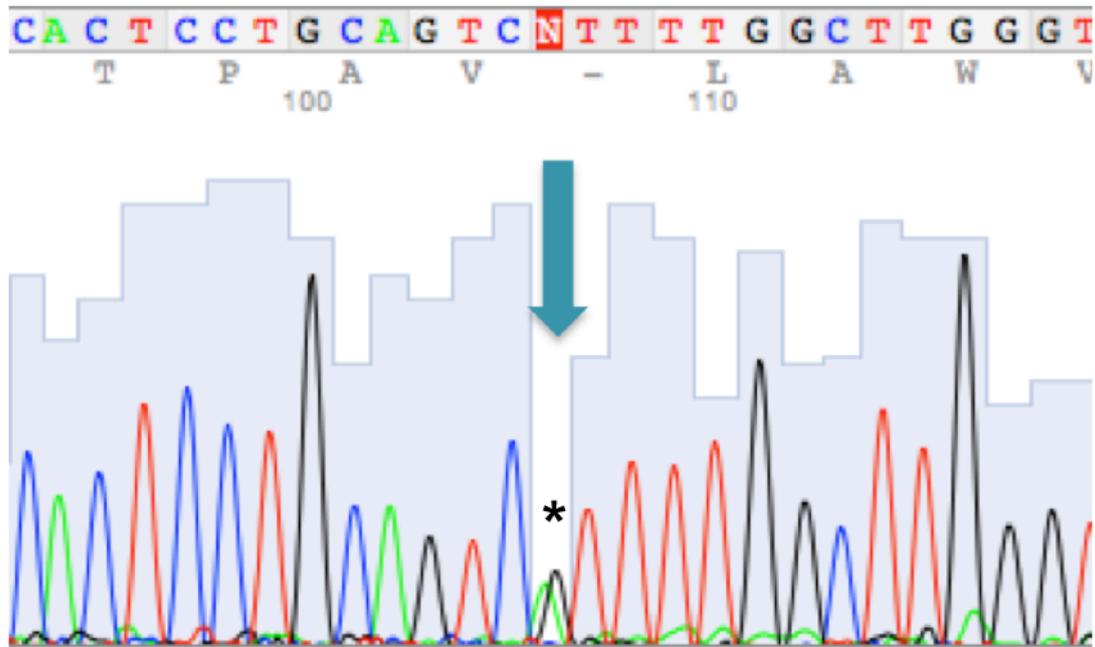
Figure 9. Sanger sequencing trace of the stop gain variant in *CREBBP*. Sanger sequencing trace of RTS3, showing part of the *CREBBP* gene. The green arrow indicating the position of the heterozygous point mutation, resulting in a stop codon, confirming the WES data.

However, genes that had variants in two or more RTS samples (Table 15) and genes with variants that are known to be involved in syndromes similar to RTS (Table 16) did not lead to a novel RTS causing gene. Confirmation of these variants called in the WES data was performed with Sanger sequencing with primers listed in Table 17.

Table 15. Genes showing variations in two or more RTS samples

Gene	Samples	Function
BCOR	RTS1 RTS10 RTS11	BCL6 Corepressor, interacts with specific Class I and II HDACs
C5orf42	RTS1 RTS10 RTS12	The protein encoded by this gene has putative coiled-coil domains and may be a transmembrane protein. Defects in this gene are a cause of Joubert syndrome (JBTS)
NOTCH3	RTS1 RTS7 RTS12	Homologues of the notch-ligands have also been identified in human, but precise interactions between these ligands and the human notch homologues remains to be determined. Mutations in NOTCH3 have been identified as the underlying cause of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL)

Table 16. Genetic diseases with RTS like symptoms

Gene	Disease	Symptoms
ARID1A ARID1B SMARCA4 SMARCB1 SMARCE1	Coffin-Siris syndromes (SWI/SNF complex)	Intellectual disability and typical somatic characteristics: sparse hair, low frontal hairline, large mouth with thick and everted lips, and hands and feet anomalies
SMARCA2	Nicolaides-Baraitser Syndrome (SWI/SNF complex)	Mild to severe intellectual disability, absent or delayed speech, growth delay, sparse hair, typical facial characteristics, brachydactyly, hand anomalies
SRCAP	Floating-Harbor syndrome	Mild intellectual disability, growth delay, characteristic facial features, delayed speech development
CKAP2L	Filippi syndrome	Intellectual disability, growth defects, microcephaly, facial features and syndactyly
ZEB2	Mowat-Wilson Syndrome	Intellectual disability, distinct facial phenotype, epilepsy

Table 17. Primer sequences for Sanger sequencing confirmation of WES

Gene	Forward	Reverse	Details
HDAC4	CAGAAAGGGGCCAGTGCTGAA	ACAGTCCTTGGTTGGTGCAGAC	Involved in chromatin remodelling/epigenetics
SMARCA4	TTTTAAACAAGCCTTGCGGGG	TTCTCGATCCGCTCGTTCTC	Coffin-Siris syndromes (SWI/SNF complex)
SMC1A	CCTGGCTGTCTGACTCTATGC	CCACCACCTTACCACCTAGC	SWI/SNF complex
SMARCA1	CTGGCATCCTTGTCTCCGAC	TTGCCTTTGCGAGATAGTGGA	SWI/SNF complex
SRCAP	TGGGAAGCTTGGGGGTATGG	CACTGGTCAACAGAGCTGGG	FHS
EVC	ACCCGTCTCTGCATGAAAACT	CACTCCAAGAGGACTGGAAGC	Ellis-van Creveld Syndrome
PGAP2	AGGGCGAGGCTGGATAGTT	TCTGCCCCCACTAATTTCGG	Hyperphosphatasia with mental retardation syndrome

Comparison of Exome targeting kits

There are a variety of library and targeting kits available using different capturing techniques, directed to specific sequencing platforms. Comparing the analyses using three different exome capture kits, will provide insight into different approaches used to sequence whole exomes. To see how these kits operate, one DNA sample has been sequenced three times, using different kits in combination with the appropriate sequencing platform. The kits used were the Ion AmpliSeq™2.0 Library and Ion TargetSeq™ exome kits, both from Life Technologies and sequenced using the Ion Proton and the HaloPlex Exome Target Enrichment kit from Agilent, which is sequenced on the HiSeq.

HaloPlex Exome Target Enrichment kit is a targeted approach and requires 200 ng of input DNA. This kit comprises about 2.5 million probes for comprehensive coverage of the coding regions of the human genome. The Ion AmpliSeq™2.0 Library Kit (Life Technologies) is a technique using multiplex PCR reactions for preparation of amplicon libraries. This kit only needs 10ng of starting DNA and uses panels and primer pools to amplify the genomic target regions. Ion TargetSeq™ exome kit (Life Technologies) technology uses barcoded fragment libraries for hybridization and exome capture. This kit capture region is covered by over 2 million capture probes, which is considerably higher than the other two kits.

Sample RTS10 was selected for this approach due to availability, good DNA quality (blood gDNA) and yield to be sequenced in triplicate. Looking at the general statistics, the overall number of variants per kit shows a wide range (Table 18 and Figure 10). The TargetSeq kit shows an extremely high number of processed variants, in addition to a high percentage of novel variants compared to known, which is also seen for the AmpliSeq kit. Normally the percentage of novel variants is around 10% per genome which is shown in the capturing data for the HaloPlex kit.

Table 18. General statistics using three different exome capture kits

General statistics	AmpliSeq	TargetSeq	HaloPlex
Variants processed	56964	170642	52907
Lines of output written	34331	65106	16176
Novel	20370	40447	786
Existing – SNPs (rs)	12619	21971	14808
Other Existing	1327	2599	579

Existing variants are the RefSNPs (rs) listed in The Single Nucleotide Polymorphism Database (dbSNP). Other existing include variants from databases such as COSMIC, the catalogue of somatic mutations in cancer.
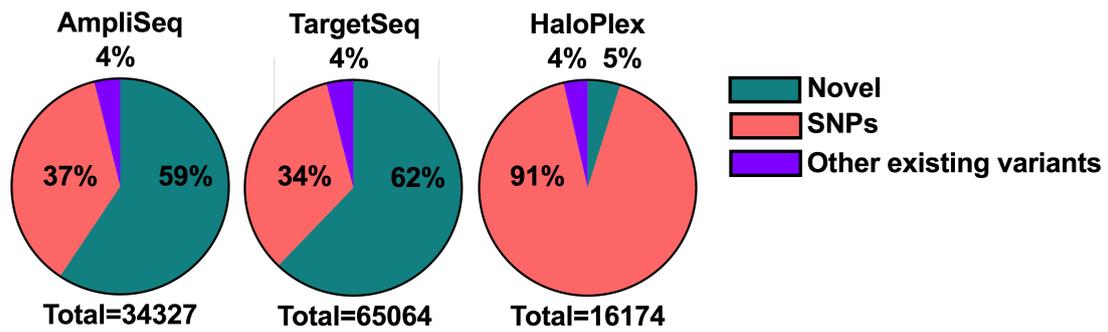
Figure 10. Variant coverage for RTS10 sequenced with three different capture kits. Novel and existing (SNPs and other) variants in RTS10 when sequenced with three different capture kits. The percentages vary considerably due to the different capture kits used in combination with the sequencing platforms. (AmpliSeq and TargetSeq sequenced on Ion Proton and HaloPlex sequenced on HiSeq)

The numbers in Table 18 and Figure 10 are all variants processed (both in coding and non-coding regions). Table 19 and Figure 11 show the number of variants in coding sequences.

Table 19. Novel variants found in RTS10 using 3 different capture kits

Coding Consequences (no SNPs)	AmpliSeq	TargetSeq	HaloPlex
Frameshift variant	20543	40997	279
Inframe deletion	26	57	30
Inframe insertion	2	4	5
Missense variant	586	1067	513
Stop gained	48	39	11
Stop lost	0	2	1
Synonymous variant	473	864	508
Splice region	14	39	7
Total	21692	43069	1354
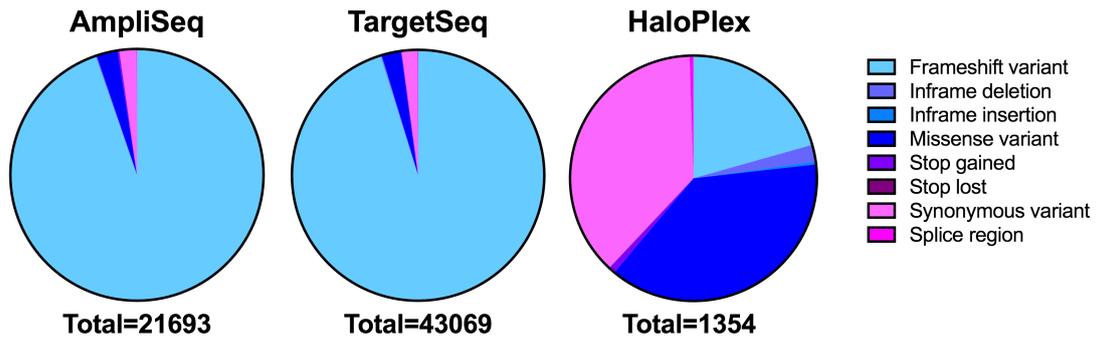
Figure 11. Coding consequences in RTS10 when using three different capture kits. These pie charts show the range in different variants in the coding sequence of RTS10, comparing 3 different capture kits.

When looking at frameshift, missense and stop gain variants (Table 19 and Figure 11), there is a significant increase in the number of frameshift mutations in both the AmpliSeq (20543) and TargetSeq (40997) kits compared to the HaloPlex (279) kit. Frameshift mutations are caused by insertions or deletions of a number of nucleotides that cannot be divided by three, leading to a change in the reading frame and resulting in an aberrant translation. However, besides differences in kits, the wide range in numbers can also be a consequence of the sequencing platforms used. The Ion Torrent system is based on the detection of hydrogen ions that are released during the polymerisation of DNA, as opposed to the optical methods used in other sequencing systems. When homopolymer repeats are present in the template sequence, multiple nucleotides will be incorporated in a single cycle. This leads to a corresponding number of released hydrogens and a proportionally higher signal which can lead to incorrect number of polymer repeats. Illumina HiSeq uses fluorescently labelled dNTPs and when polymerase elongates the strand and incorporates one of these dNTPs, the fluorescent label is excited and cleaved. This will unblock the site for the following nucleotide to be incorporated in the next cycle. Each cycle only permits the elongation of a single nucleotide at the time, therefore homopolymers can be determined precisely (Figure 12). This difference in signal detection is reflected in the amount of frameshift mutations called in samples sequenced with the Ion Proton.

This is indicated by the selection of 20 random frameshift mutations per kit. This shows 45% and 55% of called variants within a homopolymer for the AmpliSeq and TargetSeq kits respectively, whereas for the HaloPlex kit this is only 20% (Appendix 5). When looking at overlapping variants between kits (SNPs, genes and unique locations), there is a relative large number of variants that are not called in all three kits (Figure 13).

In conclusion, the number of called variants depends not only on the sample, but also on the capture kits and sequencing platforms used. It is recommended to confirm all variants of interest by Sanger Sequencing.
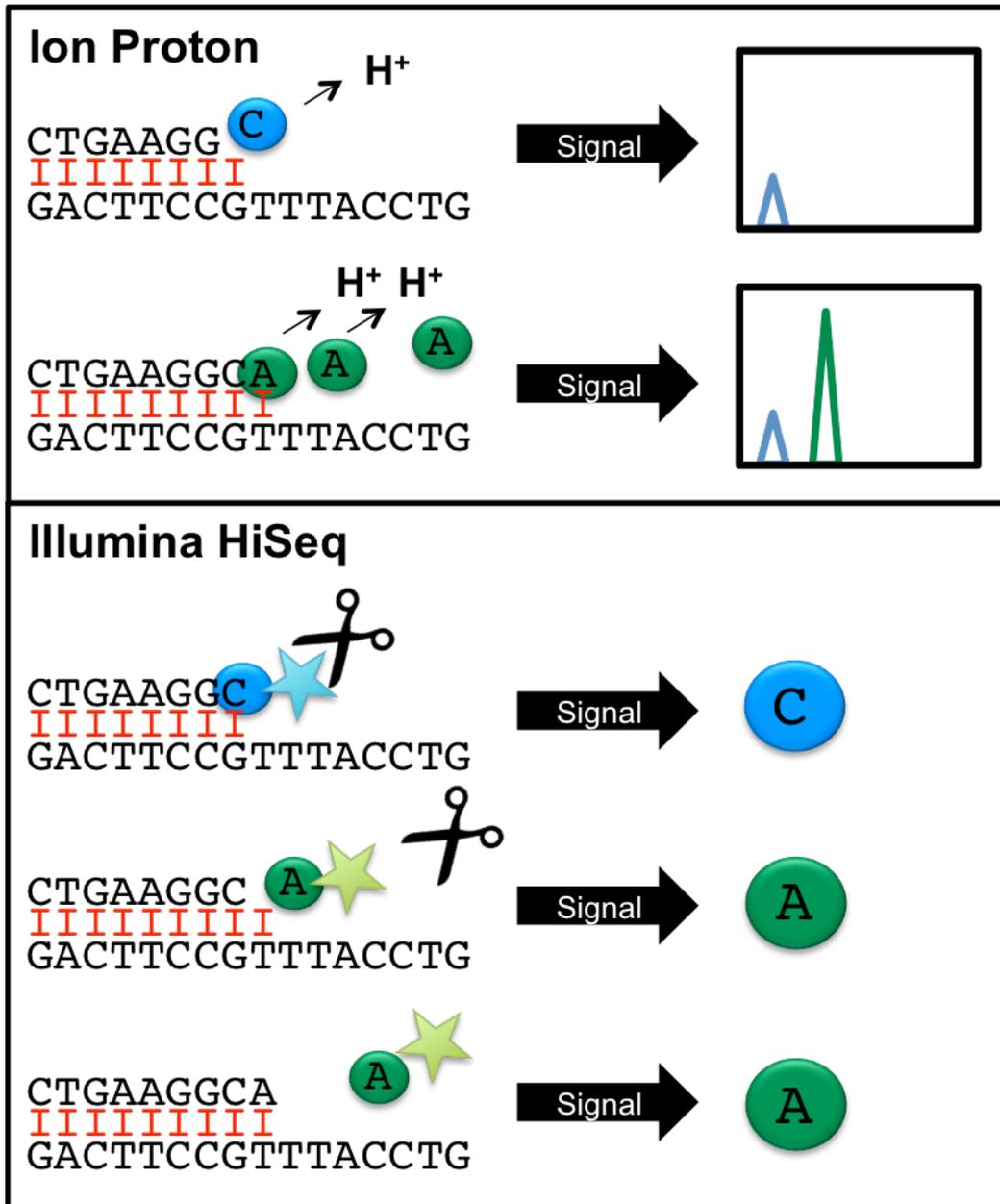
Figure 12. Sequencing chemistry. The Ion Proton system (top) releases hydrogen ions when nucleotides are incorporated during sequencing. When detecting homopolymers, multiple nucleotides are inserted in one cycle, leading to a relevant peak height. The Illumina HiSeq (bottom) incorporates one nucleotide per cycle, and when matching, the fluorescent tag is cleaved leading to a signal. This will unblock the next site for a nucleotide to be inserted in the following cycle.
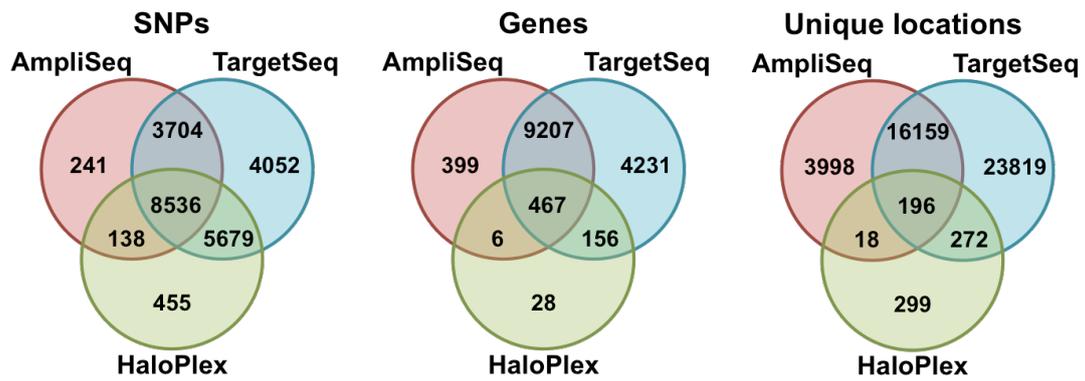
Figure 13. Overlap in DNA sample RTS10 sequenced with three different capture kits. These Venn diagrams show the overlap of SNPs, genes and unique locations of variants between the three capture kits used for WES in sample RTS10. AmpliSeq in red, TargetSeq in blue and HaloPlex in green.

MLPA with probe mix for *CREBBP* and *EP300*

A different method to detect genetic variation is with MLPA, which can be used for detecting copy number variation i.e. deletions and duplications. Probes were designed for exons evenly spread through the *CREBBP* and *EP300* genes.

MLPA was performed on all RTS samples, but unfortunately, the buccal samples did not give positive results, presumably due to the poor DNA quality. However, the four blood DNA samples gave a positive result for the *CREBBP* probe mix, whereas *EP300* probes worked for three out of the four samples (RTS11 failed). The samples analysed were RTS1, RTS10, RTS11 (only *CREBBP*) and RTS12, with peak heights normalized against control probes.

Results are shown in Figure 14, with probes for each exon on the X-axis and the ratios of the probes after normalisation on the Y-axis. *EP300* did not show any CNV for the analysed samples and probes. However, for *CREBBP* we detected a potential deletion in sample RTS1 in exon 4 was detected. This probe showed a normalized ratio of ~0.5, consistent with a heterozygous deletion of the corresponding exon.

When using BLAST on the UCSC browser for the probe we designed for *CREBBP* exon 4, it shows a SNP located near the ligation site of the MLPA probe (Figure 15). This variant was also detected in the WES results, as a synonymous variant. With MLPA, probes are solely amplified when the two half probes are ligated. It is assumed this SNP interferes with the ligation site of the probe, resulting in a false positive deletion in exon 4 of the *CREBBP* gene.

The SNP concerned is rs751666077, with an allele frequency of T: 99.997% and C: 0.003%, causing a synonymous variant. Figure 16 shows the IGV plot, with the probe ligation site.

In conclusion, these results show that MLPA is a sensitive technique that can be used as a time efficient, high throughput pre-screen to detect CNVs. However, positive results from a single probe need to be confirmed using other methods, such as Sanger sequencing.
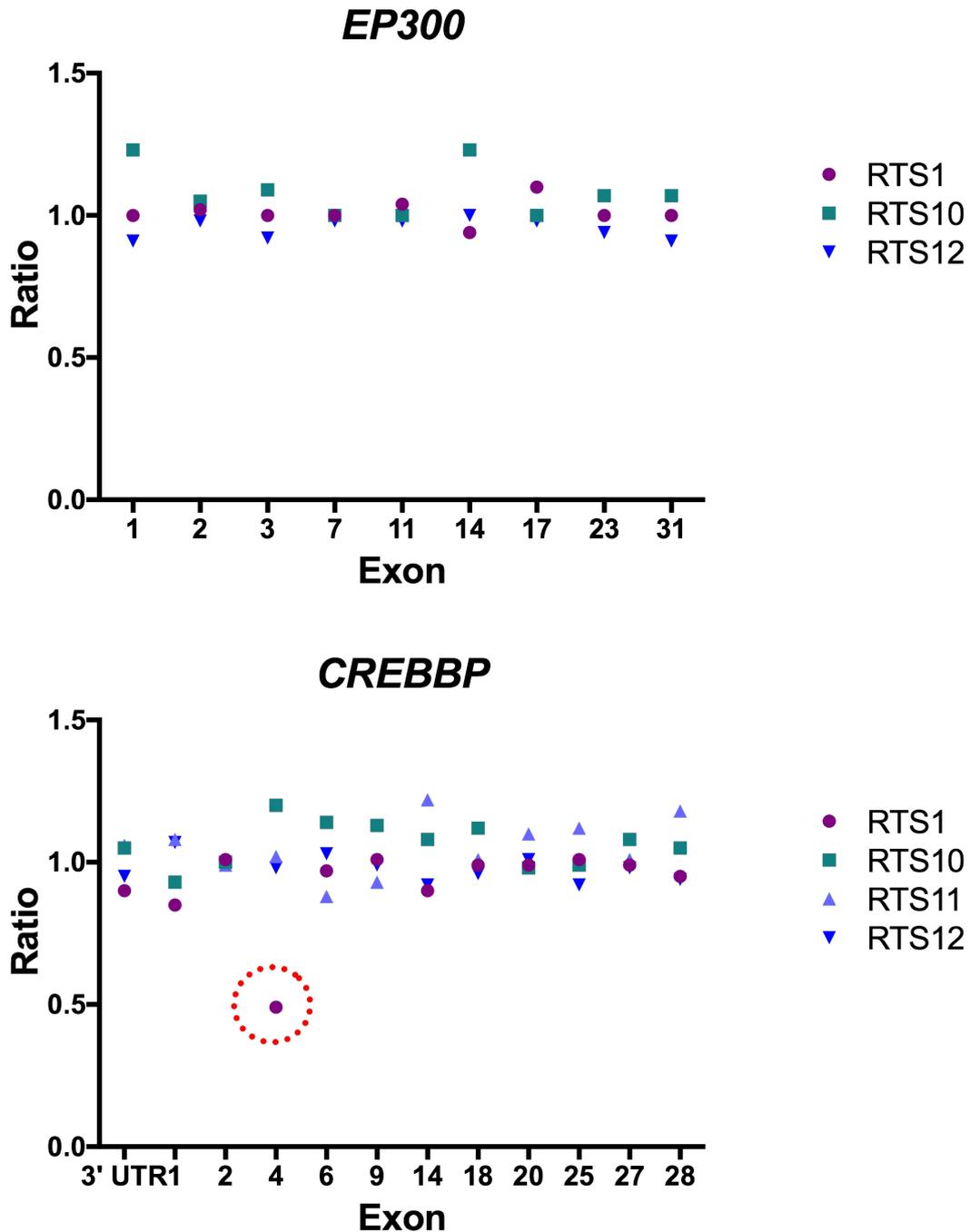
Figure 14. MLPA results for RTS samples. RTS samples did not show CNVs for the exons tested with probes in *EP300* (top). RTS1 showed a deletion after normalization in exon 4 of the *CREBBP* gene shown in the red circle, with a ratio of ~0.5, consistent with a heterozygous deletion (bottom). Ratios are calculated after normalization with control probes are shown in the Y-axis, and exon number of either *EP300* or *CREBBP* are shown on the X-axis.
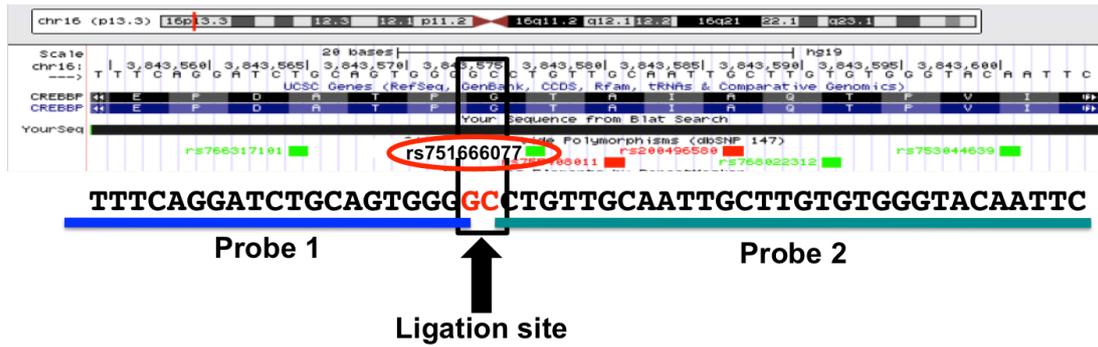
Figure 15. MLPA probe ligation site location. The ligation site (red CG) of the two half probes for exon 4 of the *CREBBP* gene, shown in blue and green, is located two base pairs from a known SNP (rs751666077), shown in the red circle.
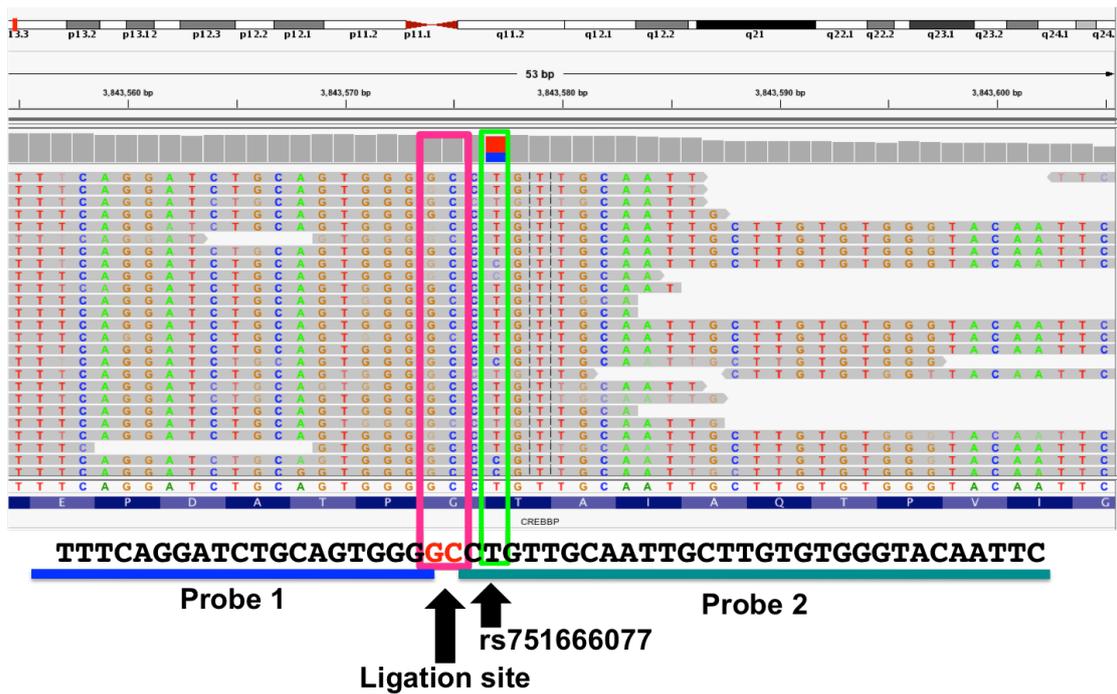
Figure 16. IGV plot showing the SNP and ligation site. Shown in the pink box is the ligation site of the two half probes for exon 4 in the *CREBBP* gene and the green box shows the known SNP.

Discussion

This chapter describes the analysis of the genome of a cohort of six RTS samples. These patients have been clinically diagnosed with RTS, without a genetic diagnosis. By performing WES on these DNA samples, a novel variant in the *CREBBP* gene has been detected. This variant is a stop gain mutation in exon 26 of the *CREBBP* gene, causing a premature stop codon, leading to a truncated and non-functional CBP protein. As expected, no RTS causing variant in *EP300* was detected, due to the fact that the mutation rate in this gene is low (<10%) in combination with our small cohort (six patient samples). The aim of this study was to detect additional RTS causing genes. However, after analyzing the WES data, a novel RTS gene could not be confirmed with Sanger sequencing.

Even though the majority of human diseases are caused by point mutations in specific genes, gene deletions and duplications (CNVs) represent a significant portion of all disease causing mutations (~5%) [61-63]. This has been reported for RTS, were cases had partial or complete *CREBBP* gene deletions and MLPA is a technique that can be used for the molecular analysis of the presence of gene CNVs in a diagnostic setting.

Therefore, MLPA was applied to screen RTS samples using probe mixes for both *CREBBP* and *EP300*. After analysis, a deletion in exon 4 of the *CREBBP* gene in one of the samples was detected. However, this deletion could not be identified in the probe target area in the WES data. When performing a BLAST search for the probe using UCSC, it appeared the ligation site of the two half probes was located near a SNP, two base pairs away. This is likely to interfere with the ligation process of the two half probes, which is essential for the probe to be amplified. Due to the absence of ligation and therefore amplification of the probe, the respective probe was detected as a false positive deletion. Hence, caution needs to be taken when designing probes, and SNPs in the genome of interest need to be considered and avoided when possible. Even though being a false positive result, this confirms MLPA is a sensitive technique for detecting variants in the genome and could be regarded as a cost effective and high throughput first screening method of candidate genes in a diagnostic setting. It is important, however, to validate apparent CNVs detected with a single probe using an alternative approach, such as Sanger sequencing.

There are some additional limitations for MLPA, in particular the DNA quality, as MLPA is sensitive to the type of sample used for DNA extraction (e.g., blood or

buccal swab) but also the method used to extract the DNA. This can have an effect on the result and therefore it is recommended to use the same tissue of origin and DNA extraction methods for all samples in a single experiment. MLPA is also more sensitive to DNA impurities than other techniques, such as PCR, and might not be able to identify mosaicism. However, MLPA is an efficient technique and can be applied in molecular diagnostics to detect and analyze CNVs in rare genetic conditions.

In the literature, there have been conflicting reports about results obtained using buccal DNA. However, buccal swabs are an inexpensive and non-invasive method to collect DNA samples and are preferred over blood samples when certain patient groups are involved, such as children or elderly people [64, 65]. Nevertheless, studies have shown that buccal swabs have several limitations that might interfere with NGS [66, 67]. The yield of buccal-derived DNA is lower, which can be due to the fact that the epithelial cells collected from the mouth are superficial and about 25% of these are in the process of apoptosis [68]. In addition, buccal DNA is prone to degradation, which may limit the success rate of further experiments. Finally, DNA isolated from the mouth has a high risk of containing bacteria, and this exogenous DNA contamination might interfere with downstream analysis [69]. However, due to the rarity of RTS and the geographical location of the participants in this study, buccal swabs were selected as DNA collection approach due to the fact they can be mailed out for self-collection.

No RTS causing mutations in *EP300* were detected, consistent with the rarity of mutations in this gene in RTS, and the fact that patients with a mutation in *EP300* are generally less severely affected. This is particularly seen in the skeletal abnormalities, such as the absence of the abnormal broad thumb and big toe. Initial RTS diagnosis is typically based on the clinical presentation, and therefore cases with milder and/or atypical RTS features may be clinically misdiagnosed and excluded for genetic testing for RTS. This is further supported by studies that identified *EP300* variants in individuals not originally diagnosed with RTS, but instead were suspected of having CdLS, which is a genetically heterogeneous syndrome with similar phenotypes [25, 70, 71]. In contrast, individuals that are initially clinically diagnosed for RTS but do not have a mutation in either *CREBBP* or *EP300*, could potentially be affected by other development disorders, such as CdLS and FHS. It is also known that CBP and p300 interact with many proteins and control expression of numerous genes.

Therefore a mutation located in genes either up or down-stream of CBP/p300 could lead to RTS.

Another frequent phenomenon that has been reported in a range of condition is mosaicism [72, 73]. The ability to detect this depends on the degree of mosaicism and the technique used. Gervasni et al. studied 42 RTS patients and identified three mosaic deletions using FISH analysis [32]. One of these cases showed a *CREBBP* mutation in 18.5% of cells tested. This mutation had been previously missed with microsatellite analysis and would have been difficult to detect with other techniques such as MLPA [74]. Besides the screening technique used, the origin of samples plays an important role when mosaicism is involved. DNA for diagnostic purposes is routinely isolated from blood samples. However, multiple studies with CdLS samples have identified causative sequence variants in DNA derived from buccal cells, which were not detected in DNA from matching blood samples [75, 76]. This has also been shown in a WES study by de Vries et al. [77] where a mutation in a known RTS gene could not be identified in the DNA from blood, but subsequent analysis of DNA isolated from buccal mucosa revealed a mosaic variant in *CREBBP*.

Additionally, WES and other previous used techniques for mutation screening mainly focus on the exome (or protein coding) sequences or structural variants, and therefore, regulatory elements, such as promoters and enhancers, will be missed. However, it is known that monogenic diseases can also be caused by mutations affecting the regulation of a specific gene. For some genes, these regulatory elements can stretch as far as 1 Mb in either direction [78, 79] and can even reside within the introns of neighbouring genes [80, 81]. The control of these genes is through long-range physical interactions where enhancer and insulator elements often engage in physical contacts with their target promoters. Not only mutations in these elements can result in disease but also disruption of the chromatin structure to enable the long-range interactions between regulatory proteins and *cis*-regulatory elements [82-85]. Mutations and disruptions like these could also lead to disturbed regulation and expression of *CREBBP* and *EP300* with RTS as a result. However, despite these (minimal) limitations, exome sequencing will remain an important experimental approach for rare-variant studies until WGS becomes less costly.

NGS has made significant advances in human genetic research towards the understanding of many molecular aspects underlying genetic diseases. Currently, multiple sequencing platforms are available, which can be combined with a variety of

methods involving template preparation, sequencing, imaging and data analysis. The unique combination of protocols determines the type of data produced, which presents challenges when comparing samples based on data quality and cost. NGS will generate a vast amount of data and applying filtering strategies can lead to loss of potential candidate genes, which might be biologically relevant for the onset of RTS. Another aspect to take into consideration is alignment. Correct alignment is crucial in variant detection, as incorrect aligned reads may lead to errors in SNP and genotype calling.

The Ion Proton technique has been the main sequencing platform in this study. It differs from other sequencing technologies and uses ion semiconductor sequencing, instead of modified nucleotides or optics. Even though updated recently, this technique is known to have difficulties with homopolymer repeats of the same nucleotide. After a certain number of homopolymer nucleotides (>3), the system cannot discriminate the amount of signal with mistakes in detection as a result. This will lead to false positive variants called as deletions or duplications. Even though most variants of interest are being validated with Sanger sequencing in practice, this is even more critical for variants located in homopolymer repeats. Another limitation of this sequencing platform is the short read length compared to other sequencing methods. This is not of concern for our study, but can be an issue when performing a *de novo* genome assembly [86].

For ethical and financial reasons, WES was limited to the exomes of the RTS patients. In a *de novo* approach, sequencing both patient and parents (trios) has been successfully used to identify causative genes of various diseases [87-89]. However, Sanger sequencing was performed on parental samples when available, to confirm *de novo* variants in the candidate gene list for RTS genes. As exome sequencing becomes more affordable, trio analysis will become the preferred approach in future.

RNA-seq is currently being applied to studying more complex diseases such as neurodegenerative diseases [90]. This technique could provide more detail on gene expression profiles in RTS. Especially since both CBP and p300 act as coactivators and HATs, they play key roles in epigenetic changes that are crucial in gene expression regulation. Rather than looking at DNA sequences, RNA-sequencing (RNA-seq), will give more insights in the transcriptome, such as gene expression, different types of transcripts, RNA editing and short and long non-coding RNAs [91, 92].

Completion of the human genome in combination with the development of sequencing into NGS, has allowed the identification of genes responsible for a specific phenotype/disease and to analyze their function, which has revolutionized the understanding of biology and medicine. Following on from the human genome project, the sequence of thousands of genomes are now available. It is likely that complete sequencing of all genomes will eventually occur in the future, resulting in a better understanding of these conditions, along with possible therapeutic directions.

References

1.	Hennekam, R.C., et al., *Rubinstein-Taybi syndrome in a mother and son.* Eur J Pediatr, 1989. **148**(5): p. 439-41.

2.	Hennekam, R.C., *Rubinstein-Taybi syndrome.* Eur J Hum Genet, 2006. **14**(9): p. 981-5.

3.	Petrij, F., et al., *Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP.* Nature, 1995. **376**(6538): p. 348-51.

4.	Roelfsema, J.H., et al., *Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease.* Am J Hum Genet, 2005. **76**(4): p. 572-80.

5.	Marion, R.W., D.M. Garcia, and J.B. Karasik, *Apparent dominant transmission of the Rubinstein-Taybi syndrome.* Am J Med Genet, 1993. **46**(3): p. 284-7.

6.	Bartsch, O., et al., *Inheritance and variable expression in Rubinstein-Taybi syndrome.* Am J Med Genet A, 2010. **152A**(9): p. 2254-61.

7.	Bartsch, O., et al., *Two patients with EP300 mutations and facial dysmorphism different from the classic Rubinstein-Taybi syndrome.* Am J Med Genet A, 2010. **152A**(1): p. 181-4.

8.	Chiang, P.W., et al., *Somatic and germ-line mosaicism in Rubinstein-Taybi syndrome.* Am J Med Genet A, 2009. **149A**(7): p. 1463-7.

9.	Tajir, M., et al., *Germline mosaicism in Rubinstein-Taybi syndrome.* Gene, 2013. **518**(2): p. 476-8.

10.	Bartholdi, D., et al., *Genetic heterogeneity in Rubinstein-Taybi syndrome: delineation of the phenotype of the first patients carrying mutations in EP300.* J Med Genet, 2007. **44**(5): p. 327-33.

11.	Negri, G., et al., *Clinical and molecular characterization of Rubinstein-Taybi syndrome patients carrying distinct novel mutations of the EP300 gene.* Clin Genet, 2015. **87**(2): p. 148-54.

12.	Kwok, R.P., et al., *Nuclear protein CBP is a coactivator for the transcription factor CREB.* Nature, 1994. **370**(6486): p. 223-6.

13.	Lundblad, J.R., et al., *Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP.* Nature, 1995. **374**(6517): p. 85-8.

14.	Goodman, R.H. and S. Smolik, *CBP/p300 in cell growth, transformation, and development.* Genes Dev, 2000. **14**(13): p. 1553-77.

15.	Bedford, D.C., et al., *Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases.* Epigenetics, 2010. **5**(1): p. 9-15.

16.	Kasper, L.H., et al., *Conditional knockout mice reveal distinct functions for the global transcriptional coactivators CBP and p300 in T-cell development.* Mol Cell Biol, 2006. **26**(3): p. 789-809.

17.	Petrij, F., et al., *Diagnostic analysis of the Rubinstein-Taybi syndrome: five cosmids should be used for microdeletion detection and low number of protein truncating mutations.* J Med Genet, 2000. **37**(3): p. 168-76.

18.	Rusconi, D., et al., *Characterization of 14 novel deletions underlying Rubinstein-Taybi syndrome: an update of the CREBBP deletion repertoire.* Hum Genet, 2015. **134**(6): p. 613-26.

19.   Murata, T., et al., *Defect of histone acetyltransferase activity of the nuclear transcriptional coactivator CBP in Rubinstein-Taybi syndrome.* Hum Mol Genet, 2001. **10**(10): p. 1071-6.

20.   Kalkhoven, E., et al., *Loss of CBP acetyltransferase activity by PHD finger mutations in Rubinstein-Taybi syndrome.* Hum Mol Genet, 2003. **12**(4): p. 441-50.

21.   Oike, Y., et al., *Truncated CBP protein leads to classical Rubinstein-Taybi syndrome phenotypes in mice: implications for a dominant-negative mechanism.* Hum Mol Genet, 1999. **8**(3): p. 387-96.

22.   Tsai, A.C., et al., *Exon deletions of the EP300 and CREBBP genes in two children with Rubinstein-Taybi syndrome detected by aCGH.* Eur J Hum Genet, 2011. **19**(1): p. 43-9.

23.   Zimmermann, N., et al., *Confirmation of EP300 gene mutations as a rare cause of Rubinstein-Taybi syndrome.* Eur J Hum Genet, 2007. **15**(8): p. 837-42.

24.   Solomon, B.D., et al., *Expanding the phenotypic spectrum in EP300-related Rubinstein-Taybi syndrome.* Am J Med Genet A, 2015. **167A**(5): p. 1111-6.

25.   Woods, S.A., et al., *Exome sequencing identifies a novel EP300 frame shift mutation in a patient with features that overlap Cornelia de Lange syndrome.* Am J Med Genet A, 2014. **164**A(1): p. 251-8.

26.   Kung, A.L., et al., *Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP.* Genes Dev, 2000. **14**(3): p. 272-7.

27.   Yao, T.P., et al., *Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300.* Cell, 1998. **93**(3): p. 361-72.

28.   Kalkhoven, E., *CBP and p300: HATs for different occasions.* Biochem Pharmacol, 2004. **68**(6): p. 1145-55.

29.   Kwok, R.P., X.T. Liu, and G.D. Smith, *Distribution of co-activators CBP and p300 during mouse oocyte and embryo development.* Mol Reprod Dev, 2006. **73**(7): p. 885-94.

30.   Partanen, A., J. Motoyama, and C.C. Hui, *Developmentally regulated expression of the transcriptional cofactors/histone acetyltransferases CBP and p300 during mouse embryogenesis.* Int J Dev Biol, 1999. **43**(6): p. 487-94.

31.   Bentivegna, A., et al., *Rubinstein-Taybi Syndrome: spectrum of CREBBP mutations in Italian patients.* BMC Med Genet, 2006. **19**(7): p. 77.

32.   Gervasini, C., et al., *High frequency of mosaic CREBBP deletions in Rubinstein-Taybi syndrome patients and mapping of somatic and germ-line breakpoints.* Genomics, 2007. **90**(5): p. 567-73.

33.   Schorry, E.K., et al., *Genotype-phenotype correlations in Rubinstein-Taybi syndrome.* Am J Med Genet A, 2008. **146A**(19): p. 2512-9.

34.   Gilissen, C., et al., *Unlocking Mendelian disease using exome sequencing.* Genome Biol, 2011. **12**(9): p. 228.

35.   Oti, M. and H.G. Brunner, *The modular nature of genetic diseases.* Clin Genet, 2007. **71**(1): p. 1-11.

36.   Antonarakis, S.E. and J.S. Beckmann, *Mendelian disorders deserve more attention.* Nat Rev Genet, 2006. **7**(4): p. 277-82.

37.   Imaizumi, K. and Y. Kuroki, *Rubinstein-Taybi syndrome with de novo reciprocal translocation t(2;16)(p13.3;p13.3).* Am J Med Genet, 1991. **38**(4): p. 636-9.

38. Lacombe, D., et al., *Confirmation of assignment of a locus for Rubinstein-Taybi syndrome gene to 16p13.3.* Am J Med Genet, 1992. **44**(1): p. 126-8.

39. Tommerup, N., C.B. van der Hagen, and A. Heiberg, *Tentative assignment of a locus for Rubinstein-Taybi syndrome to 16p13.3 by a de novo reciprocal translocation, t(7;16)(q34;p13.3).* Am J Med Genet, 1992. **44**(2): p. 237-41.

40. Shashi, V., et al., *The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders.* Genet Med, 2014. **16**(2): p. 176-82.

41. GeneReviews. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1116/.

42. Chong, J.X., et al., *The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities.* Am J Hum Genet, 2015. **97**(2): p. 199-215.

43. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.

44. Ronaghi, M., et al., *Real-time DNA sequencing using detection of pyrophosphate release.* Anal Biochem, 1996. **242**(1): p. 84-9.

45. Schouten, J.P., et al., *Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.* Nucleic Acids Res, 2002. **30**(12): p. e57.

46. Kurotaki, N., et al., *Haploinsufficiency of NSD1 causes Sotos syndrome.* Nat Genet, 2002. **30**(4): p. 365-6.

47. Kerem, B., et al., *Identification of the cystic fibrosis gene: genetic analysis.* Science, 1989. **245**(4922): p. 1073-80.

48. Lander, E.S. and D. Botstein, *Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children.* Science, 1987. **236**(4808): p. 1567-70.

49. Vissers, L.E., et al., *Mutations in a new member of the chromodomain gene family cause CHARGE syndrome.* Nat Genet, 2004. **36**(9): p. 955-7.

50. Nijman, I.J., et al., *Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples.* Nat Methods, 2010. **7**(11): p. 913-5.

51. Chaisson, M.J., D. Brinza, and P.A. Pevzner, *De novo fragment assembly with short mate-paired reads: Does the read length matter?* Genome Res, 2009. **19**(2): p. 336-46.

52. Pop, M. and S.L. Salzberg, *Bioinformatics challenges of new sequencing technology.* Trends Genet, 2008. **24**(3): p. 142-9.

53. Trapnell, C. and S.L. Salzberg, *How to map billions of short reads onto genomes.* Nat Biotechnol, 2009. **27**(5): p. 455-7.

54. Varley, K.E. and R.D. Mitra, *Nested Patch PCR enables highly multiplexed mutation discovery in candidate genes.* Genome Res, 2008. **18**(11): p. 1844-50.

55. Rauch, A., et al., *Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation.* Am J Med Genet A, 2006. **140**(19): p. 2063-74.

56. Cirulli, E.T. and D.B. Goldstein, *Uncovering the roles of rare variants in common disease through whole-genome sequencing.* Nat Rev Genet, 2010. **11**(6): p. 415-25.

57. Iossifov, I., et al., *The contribution of de novo coding mutations to autism spectrum disorder.* Nature, 2014. **515**(7526): p. 216-21.

58. den Dunnen, J.T. and S.E. Antonarakis, *Nomenclature for the description of human sequence variations.* Hum Genet, 2001. **109**(1): p. 121-4.

59. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.

60. Martins, A.D., et al., *Physiology of na+/h+ exchangers in the male reproductive tract: relevance for male fertility.* Biol Reprod, 2014. **91**(1): p. 11.

61. Armour, J.A., et al., *The detection of large deletions or duplications in genomic DNA.* Hum Mutat, 2002. **20**(5): p. 325-37.

62. Den Dunnen, J.T., et al., *Topography of the Duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications.* Am J Hum Genet, 1989. **45**(6): p. 835-47.

63. van der Steege, G., et al., *PCR-based DNA test to confirm clinical diagnosis of autosomal recessive spinal muscular atrophy.* Lancet, 1995. **345**(8955): p. 985-6.

64. Adriaanse, M.P., et al., *Human leukocyte antigen typing using buccal swabs as accurate and non-invasive substitute for venipuncture in children at risk for celiac disease.* J Gastroenterol Hepatol, 2016. **31**(10): p. 1711-1716.

65. Freeman, B., et al., *DNA by mail: an inexpensive and noninvasive method for collecting DNA samples from widely dispersed populations.* Behav Genet, 1997. **27**(3): p. 251-7.

66. Livy, A., et al., *Evaluation of quality of DNA extracted from buccal swabs for microarray based genotyping.* Indian J Clin Biochem, 2012. **27**(1): p. 28-33.

67. Yin, Y., et al., *Application of High-Throughput Next-Generation Sequencing for HLA Typing on Buccal Extracted DNA: Results from over 10,000 Donor Recruitment Samples.* PLoS One, 2016. **11**(10): p. e0165810.

68. Rudney, J.D. and R. Chen, *The vital status of human buccal epithelial cells and the bacteria associated with them.* Arch Oral Biol, 2006. **51**(4): p. 291-8.

69. Mahfuz, I., W. Cheng, and S.J. White, *Identification of Streptococcus parasanguinis DNA contamination in human buccal DNA samples.* BMC Res Notes, 2013. **22**(6): p. 481.

70. Rohatgi, S., et al., *Facial diagnosis of mild and variant CdLS: Insights from a dysmorphologist survey.* Am J Med Genet A, 2010. **152A**(7): p. 1641-53.

71. Solomon, B.D., et al., *Expanding the phenotypic spectrum in EP300-related Rubinstein-Taybi syndrome.* Am J Med Genet A, 2015. **167A**(5): p. 1111-6.

72. Biesecker, L.G. and N.B. Spinner, *A genomic view of mosaicism and human disease.* Nat Rev Genet, 2013. **14**(5): p. 307-20.

73. Notini, A.J., J.M. Craig, and S.J. White, *Copy number variation and mosaicism.* Cytogenet Genome Res, 2008. **123**(1-4): p. 270-7.

74. Aten, E., et al., *Methods to detect CNVs in the human genome.* Cytogenet Genome Res, 2008. **123**(1-4): p. 313-21.

75. Braunholz, D., et al., *Hidden mutations in Cornelia de Lange syndrome limitations of sanger sequencing in molecular diagnostics.* Hum Mutat, 2015. **36**(1): p. 26-9.

76. Huisman, S.A., et al., *High rate of mosaicism in individuals with Cornelia de Lange syndrome.* J Med Genet, 2013. **50**(5): p. 339-44.

77. de Vries, T.I., et al., *Mosaic CREBBP mutation causes overlapping clinical features of Rubinstein-Taybi and Filippi syndromes.* Eur J Hum Genet, 2016. **24**(9): p. 1363-6.

78. Pfeifer, D., et al., *Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region.* Am J Hum Genet, 1999. **65**(1): p. 111-24.

79.   Kimura-Yoshida, C., et al., *Characterization of the pufferfish Otx2 cis-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification.* Development, 2004. **131**(1): p. 57-71.

80.   Kleinjan, D.A., et al., *Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6.* Hum Mol Genet, 2001. **10**(19): p. 2049-59.

81.   Lettice, L.A., et al., *Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly.* Proc Natl Acad Sci U S A, 2002. **99**(11): p. 7548-53.

82.   Kleinjan, D.J. and V. van Heyningen, *Position effect in human genetic disease.* Hum Mol Genet, 1998. **7**(10): p. 1611-8.

83.   Jiang, G., et al., *Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q.* Hum Mol Genet, 2003. **12**(22): p. 2909-21.

84.   Saveliev, A., et al., *DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing.* Nature, 2003. **422**(6934): p. 909-13.

85.   Tufarelli, C., et al., *Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease.* Nat Genet, 2003. **34**(2): p. 157-65.

86.   Narzisi, G. and B. Mishra, *Comparing de novo genome assembly: the long and short of it.* PLoS One, 2011. **6**(4): p. e19175.

87.   Crow, J.F., *The origins, patterns and implications of human spontaneous mutation.* Nat Rev Genet, 2000. **1**(1): p. 40-7.

88.   Jouan, L., et al., *Rare variants in complex traits: novel identification strategies and the role of de novo mutations.* Hum Hered, 2012. **74**(3-4): p. 215-25.

89.   Veltman, J.A. and H.G. Brunner, *De novo mutations in human genetic disease.* Nat Rev Genet, 2012. **13**(8): p. 565-75.

90.   Costa, V., et al., *RNA-Seq and human complex diseases: recent accomplishments and future perspectives.* Eur J Hum Genet, 2013. **21**(2): p. 134-42.

91.   Sultan, M., et al., *A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome.* Science, 2008. **321**(5891): p. 956-60.

92.   Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

93.   Gilissen, C., et al., *Disease gene identification strategies for exome sequencing.* Eur J Hum Genet, 2012. **20**(5): p. 490-7.

94.   Sun, Y., et al., *Next-generation diagnostics: gene panel, exome, or whole genome?* Hum Mutat, 2015. **36**(6): p. 648-55.

Appendices

Appendix 1

Variant consequences (Ensembl)

SO term	SO description	Impact
Transcript ablation	A feature ablation whereby the deleted region includes a transcript feature	HIGH
Splice acceptor variant	A splice variant that changes the 2 base region at the 3' end of an intron	HIGH
Splice donor variant	A splice variant that changes the 2 base region at the 5' end of an intron	HIGH
Stop gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	HIGH
Frameshift variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	HIGH
Stop lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	HIGH
Start lost	A codon variant that changes at least one base of the canonical start codo	HIGH
Transcript amplification	A feature amplification of a region containing a transcript	HIGH
Inframe insertion	An inframe non synonymous variant that inserts bases into in the coding sequenc	MODERATE
Inframe deletion	An inframe non synonymous variant that deletes bases from the coding sequenc	MODERATE
Missense variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	MODERATE
Protein altering variant	A sequence_variant which is predicted to change the protein encoded in the coding sequence	MODERATE
Splice region variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron	LOW
Incomplete terminal codon variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed	LOW
Stop retained variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains	LOW
Synonymous variant	A sequence variant where there is no resulting change to the encoded amino acid	LOW
Coding sequence variant	A sequence variant that changes the coding sequence	MODIFIER
Mature miRNA variant	A transcript variant located with the sequence of the mature miRNA	MODIFIER
5 prime UTR variant	A UTR variant of the 5' UTR	MODIFIER
3 prime UTR variant	A UTR variant of the 3' UTR	MODIFIER
Non coding transcript exon variant	A sequence variant that changes non-coding exon sequence in a non-coding transcript	MODIFIER
Intron variant	A transcript variant occurring within an intron	MODIFIER
NMD transcript variant	A variant in a transcript that is the target of NMD	MODIFIER
Non coding transcript variant	A transcript variant of a non coding RNA gene	MODIFIER
Upstream gene variant	A sequence variant located 5' of a gene	MODIFIER
Downstream gene variant	A sequence variant located 3' of a gene	MODIFIER
TFBS ablation	A feature ablation whereby the deleted region includes a transcription factor binding site	MODIFIER
TFBS amplification	A feature amplification of a region containing a transcription factor binding site	MODIFIER
TF binding site variant	A sequence variant located within a transcription factor binding site	MODIFIER
Regulatory region ablation	A feature ablation whereby the deleted region includes a regulatory region	MODERATE
Regulatory region amplification	A feature amplification of a region containing a regulatory region	MODIFIER
Feature elongation	A sequence variant located within a regulatory region	MODIFIER
Regulatory region variant	A sequence variant located within a regulatory region	MODIFIER
Feature truncation	A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence	MODIFIER
Intergenic variant	A sequence variant located in the intergenic region, between genes	MODIFIER

The terms in the table are shown in order of severity (more severe to less severe) as estimated by Ensembl. The consequence terms are defined by the Sequence Ontology (SO), which is developed by the Gene Ontology Consortium.

Appendix 2

Interactome gene list [15]

ACTA2	CNOT3	FOXO1	HOXB7	MLL	ONECUT1	RPS6KA3	TAX1BP3
AHR	CNTN2	FOXO3	HOXB9	MN1	PAPOLA	RUNX1	TBP
AIRE	CREB1	FOXO4	HOXD10	MRE11A	PARP1	RUNX2	TBX21
AKT1	CREM	FUS	HOXD12	MSH2	PAX3	RUNX3	TCF12
ALX1	CRTC2	GABPA	HOXD13	MSH6	PAX5	RXRA	TCF3
ANAPC5	CRX	GAK	HOXD4	MSX1	PAX6	SATB1	TCF4
ANAPC7	CSE1L	GAPDH	HSF1	MSX2	PAX8	SERTAD1	TCLA
AP1B1	CSNK2A1	GATA1	HTT	MTF1	PCNA	SET	TCLB
APEX1	CSNK2A2	GATA2	IKBKG	MYB	PDX1	SGK1	TDG
AR	CTBP1	GATA3	ING1	MYBL1	PELP1	SH3GL1	TFAP2A
ARNT	CTBP2	GATA4	ING2	MYBL2	PGR	SIRT1	TGFB1I1
ARNTL	CTNNB1	GATA5	ING4	MYC	PHOX2B	SIRT2	TGS1
ASCL1	CUX1	GATA6	ING5	MYOD1	PIAS1	SKP2	THRB
ATF1	DACH1	GCM1	IRF1	NAP1L1	PIAS2	SMAD1	TLX1
ATF2	DAXX	GIPC1	IRF2	NAP1L4	PIAS3	SMAD2	TNP2
ATF4	DBP	GLI3	IRF3	NBN	PLAG1	SMAD3	TP53
ATG7	DCP1A	GMEB1	IRF7	NCOA1	PLAGL1	SMAD4	TP53BP1
ATN1	DDX24	GMEB2	JDP2	NCOA2	PLAGL2	SMAD7	TP63
ATR	DDX5	GPBP1	JMY	NCOA3	PML	SMARCA4	TP73
BAG6	DHX15	GPS2	JUN	NCOA6	POLB	SMARCB1	TRERF1
BAZ2A	DUSP1	GTF2B	JUNB	NCOR1	POLR2A	SND1	TRIP10
BCL6	E2F1	GTF2E2	KAT2A	NEDD1	POU1F1	SNIP1	TSG101
BRCA1	E2F5	GTF2F1	KAT2B	NEUROD1	POU2F3	SNW1	TTC5
BTRC	EBF1	GTF2F2	KHDRBS1	NEUROG1	POU3F	SOCS1	TWIST1
C9ORF156	EGR1	GTF3C1	KLF1	NF!A	POU6F1	SOX2	UBTF
CALCOCO1	EID1	H2A	KLF13	NFATC1	PPARA	SOX4	VDR
CARM1	EID2	H2B	KLF2	NFATC2	PPARG	SOX9	VHL
CCND1	EID3	H3	KLF4	NFATC4	PPARGC1A	SP1	VRK1
CCNE1	EIF2B1	H4	KLF5	NFE2	PPP2R5C	SP3	VRK2
CD44	ELF3	HAND2	KLF6	NFE2L2	PRMT5	SPI1	WRN
CDC25B	ELK1	HBZ	KPNA2	NFKB1	PROX1	SPIB	WT1
CDH2	ELK4	HDAC1	KPNA6	NFYA	PRPS1	SRC1	WWTR1
CDK2	EMB	HDAC6	LHX3	NFYB	PSMC5	SRCAP	XAF1
CDK5RAP3	ESR1	HERC1	MAF	NKX2-1	PSMD9	SREBF1	XRCC6
CDK8	ETS1	HES6	MAFG	NOTCH1	PTEN	SREBF2	YAP1
CDX2	ETS2	HIF1A	MAML1	NPAT	PTF1A	SRF	YBX1
CEBPA	ETV1	HIPK2	MAML2	NR1H4	PTMA	SS18	YY1
CEBPB	ETV4	HLF	MAP3K1	NR2F2	RAD23A	SS18L1	ZBTB16
CEBPD	EWSR1	HNF1A	MAPK1	NR3C1	RAD50	STAT1	ZBTB17
CENPJ	EYA1	HNF1B	MAPK14	NR3C2	RARA	STAT2	ZBTB2
CFDP1	EYA3	HNF4A	MDC1	NR4A1	RB1	STAT3	ZEB1
CHD4	FBXO3	HOXA10	MDM2	NR5A1	RBBP4	STAT4	ZFPM2
CHD9	FEN1	HOXA11	MECOM	NRIP1	RBCK1	STAT5A	ZNF148
CHUK	FGFR1	HOXA9	MED25	NUDT21	RBM14	STAT5B	ZNF639
CIITA	FHL2	HOXB1	MEF2C	NUP93	RECQL4	STAT6	
CITED1	FOS	HOXB2	MEF2D	NUP98	RELA	SUB1	
CITED2	FOXA2	HOXB3	METTL8	NUPR1	ROCK2	TAF15	
CITED4	FOXL1	HOXB4	MIER1	OGG1	RORA	TAL1	
CLTC	FOXM1	HOXB6	MITF	OLIG2	RPS6KA2	TAT	

Appendix 3

ID and epilepsy gene list [93, 94]

AAAS	ATP7A	CLCN7	ERCC3	GRIN2A	LARGE	NDUFS2	PLCB4	SDHA	TCF4
AARS2	ATR	CLIC2	ERCC4	GRIN2B	LARP7	NDUFS3	PLP1	SERAC1	TCOF1
ABCC9	ATRX	CLN8	ERCC5	GRM1	LCT	NDUFS4	PMM2	SETBP1	TECR
ABCD1	ATXN1	CNTNAP2	ERCC6	GSS	LIG4	NDUFS7	PNKP	SH3PXD2B	TFAP2A
ABCD4	AUH	COG1	ERCC8	GTDC2	LRP1	NDUFS8	PNP	SHANK2	TFAP2B
ABHD5	AUTS2	COG7	ERLIN2	GTF2H5	LRP2	NDUFV1	POLG	SHANK3	TGFBR1
ACAD9	B3GALNT2	COG8	ESCO2	GUSB	LRPPRC	NEDD4L	POLR1C	SHH	TGFBR2
ACO2	B3GALTL	COL2A1	ETFB	HAX1	MAGT1	NEU1	POLR1D	SHOC2	TGIF1
ACOX1	B3GNT1	COL4A1	ETHE1	HCCS	MAN1B1	NF1	POLR3A	SHOX	THRB
ACSF3	B4GALT1	COL4A2	EVC	HCFC1	MAN2B1	NFIA	POLR3B	SHROOM4	TIMM8A
ACSL4	B4GALT7	COLEC11	EVC2	HDAC4	MANBA	NFIX	POMGNT1	SIL1	TMCO1
ACTB	B9D1	COQ2	EXOSC3	HDAC8	MAP2K1	NHS	POMT1	SIX3	TMEM165
ACTG1	B9D2	COX15	EXT1	HEPACAM	MAP2K2	NIPBL	POMT2	SIX5	TMEM216
ACVR1	BBS1	COX7B	EYA1	HESX1	MAPT	NKX2-1	PORCN	SKI	TMEM231
ADAMTS10	BBS10	CRADD	EZH2	HLCS	MASP1	NKX21	POU1F1	SLC12A1	TMEM237
ADAR	BBS12	CRBN	FAM123B	HMGA2	MAT1A	NLGN1	POU3F4	SLC12A6	TMEM67
ADCK3	BBS2	CREBBP	FAM20C	HNRNPU	MBD5	NLGN4X	PPM1D	SLC16A2	TNK2
ADSL	BBS4	CTDP1	FANCB	HOXA1	MBTPS2	NLRP3	PPOX	SLC17A5	TOR1A
AFF2	BBS5	CTNNB1	FANCD2	HPD	MCCC1	NOG	PQBP1	SLC25A15	TPK1
AGA	BBS7	CUBN	FARS2	HPRT1	MCCC2	NOTCH3	PRODH	SLC26A9	TPO
AGPAT2	BBS9	CUL3	FBN1	HRAS	MCOLN1	NPHP1	PRPS1	SLC2A1	TRAPPC9
AGTR2	BCKDHA	CUL4B	FGD1	HSD17B10	MCPH1	NR0B1	PRSS12	SLC2A2	TREX1
AHCY	BCKDHB	CYB5R3	FGFR1	HSD17B4	MECP2	NR4A2	PSEN1	SLC33A1	TRIM32
AHI1	BCOR	D2HGDH	FGFR2	HUWE1	mecp2e1	NRAS	PTCH1	SLC35C1	TRPM6
AIFM1	BCS1L	DARS2	FGFR3	IDS	MED12	NRXN1	PTCHD1	SLC4A4	TSC1
AIMP1	BLM	DBT	FH	IDUA	MED17	NRXN1B	PTDSS1	SLC6A3	TSC2
AK1	BRAF	DCAF17	FKRP	IER3IP1	MED23	NSD1	PTEN	SLC6A8	TSPAN7
AKT3	BRAT1	ISPD	FKTN	IFT172	MEF2C	NSDHL	PTPN11	SLC7A7	TTC8
ALDH18A1	BRWD3	DCX	FLG	IGBP1	MEGF8	NSUN2	PVRL1	SLC9A6	TTN
ALDH3A2	BSCL2	DDHD2	FLNA	IGF1	MGAT2	NTRK1	PYCR1	SMAD4	TUBA1A
ALDH5A1	BUB1B	DDX11	FLNB	IKBKG	MICU1	OBSL1	RAB18	SMARCA2	TUBB2B
ALG1	C10orf2	DHCR24	FMR1	IL1RAPL1	MID1	OCA2	RAB27A	SMARCA4	TUSC3
ALG12	C5orf42	DHCR7	FOXG1	INPP5E	MIR17HG	OCLN	RAB39B	SMARCB1	TWIST1
ALG13	C7orf11	DHFR	FOXP1	IQSEC2	MKKS	OCRL	RAB3GAP1	SMARCE1	UBE2A
ALG2	CA2	DHTKD1	FOXP2	IRS1	MKS1	OFD1	RAB3GAP2	SMC1A	UBE3A
ALG3	CA8	DIP2B	FOXRED1	ITGB3	MLH1	OPHN1	RAB40AL	SMC3	UBE3B
ALG6	CACNA1A	DIS3L2	FRAS1	ITPR1	MLL	ORC1	RAD21	SMOC1	UBR1
ALG9	CACNA1C	DISC1	FREM2	JAG1	MLL2	OTC	RAD50	SMPD1	UPB1
AMT	CACNG2	DKC1	FTO	JAM3	MLL3	PACS1	RAF1	SMS	UPF3B
ANK2	CAMTA1	DLD	FTSJ1	KANK1	MLYCD	PAFAH1B1	RAI1	SNAP29	VLDLR
ANK3	CASK	DLG3	FUCA1	KANSL1	MMAA	PAK3	RARS2	SOBP	VPS13B
ANKH	CBS	DMD	GABRA1	KAT6B	MMACHC	PANK2	RASGEF1B	SOS1	VPS39
ANKRD11	CC2D1A	DNAH5	GABRB3	KCNA1	MMADHC	PARP1	RBBP8	SOX10	WDR11
ANO5	CC2D2A	DNAJC19	GAD1	KCNJ1	MNX1	PAX6	RBFOX1	SOX2	WDR19
ANTXR1	CCBE1	DNMT3B	GALE	KCNJ10	MOCS1	PC	RBM28	SOX3	WDR35
AP1S1	CCDC22	DOCK6	GALT	KCNJ11	MOCS2	PCDH19	RBM8A	SOX5	WDR45
AP1S2	CCDC39	DOCK8	GAMT	KCNK9	MPDU1	PCNT	RELN	SPG7	WDR62
AP3B1	CCDC78	DPAGT1	GATAD2B	KCNQ1OT1	MPDZ	PDE4D	RFT1	SPRED1	WDR81
AP4B1	CCDC88C	DPCD	GATM	KCNQ2	MRPS22	PDHA1	RFX3	SPTAN1	WNT5A
AP4E1	CD96	DPM1	GCH1	KCNQ3	MSH6	PEPD	RMND1	SRCAP	WNT7A
AP4S1	CDC6	DPM2	GCSH	KCNT1	MTMR2	PEX1	RNASEH2A	SRD5A3	XPA
APAF1	CDH15	DPYD	GDI1	KCTD7	MTR	PEX10	RNASEH2B	SRGAP3	YWHAE
APTX	CDK5RAP2	DYM	GFAP	KDM5C	MTRR	PEX11B	RNASEH2C	SRPX2	ZBTB16
ARFGEF2	CDKL5	DYNC1H1	GJC2	KDM6A	MUT	PEX13	ROGDI	ST3GAL3	ZBTB24
ARHGAP31	CDKN1C	DYNC2H1	GK	KIAA1279	MVK	PEX26	ROR2	STIL	ZDHHC9
ARHGEF6	CDON	DYRK1A	GLDC	KIF11	MYCN	PEX5	RP2	STRA6	ZEB2
ARHGEF9	CDT1	EFTUD2	GLI2	KIF7	MYH9	PEX7	RPGRIP1L	STS	ZFHX4

ARID1A	CELSR1	EHMT1	GLI3	KIRREL3	MYO1E	PGAP2	RPS6KA3	STXBP1	ZFYVE26
ARID1B	CENPJ	EIF2AK3	GLRA1	VPS13B	MYO3A	PGK1	RTEL1	SUCLA2	ZIC2
ARL13B	CEP135	EIF4G1	GMPPB	KIT	MYO5A	PHF21A	RUNX1	SUOX	ZNF335
ARL6	CEP152	ELOVL4	GNAS	KMT2D	MYT1L	PHF6	RYR1	SURF1	ZNF41
ARSE	CEP290	EMX2	GNPAT	KNCQ2	NAGA	PHF8	RYR2	SYN1	ZNF592
ARX	CEP41	EP300	GNRHR	KPTN	NAGLU	PHGDH	SALL1	SYNE1	ZNF674
ASL	CEP63	EPB41L1	GNS	KRAS	NBN	PIGA	SALL4	SYNGAP1	ZNF711
ASPA	CHAT	EPG5	GP1BB	KRBOX4	NDE1	PIGL	SATB2	SYP	ZNF81
ASPM	CHD2	EPHA5	GPC3	L1CAM	NDP	PIGN	SBF1	SYT14	
ASXL1	CHD7	EPHB2	GPHN	L2HGDH	NDUFA1	PIGO	SC5D	TAT	
ATP1A2	CHKB	ERBB3	GPR56	LAMA2	NDUFA10	PIGV	SCN1A	TBC1D24	
ATP2A2	CHMP1A	ERBB4	GRIA3	LAMB1	NDUFA11	PIK3R2	SCN2A	TBCE	
ATP6AP2	CHRND	ERCC1	GRIK2	LAMC3	NDUFA12	PLA2G6	SCN8A	TBR1	
ATP6V0A2	CHUK	ERCC2	GRIN1	LAMP2	NDUFS1	PLCB1	SCO2	TBX5	

Appendix 4

Genes with variants in RTS samples that show overlap with the Interactome and ID/epilepsy gene lists

RTS1		RTS7		RTS10	RTS11		RTS12	
AGPAT2	KANK1	AIRE	MPDU1	ANK2	ABCD1	L1CAM	ADAR	MDC1
ANK2	KANSL1	ATF4	MTF1	BBS4	ACAD9	LAMC3	ANKRD11	MECOM
ANKH	KAT6B	AUTS2	NCOA2	BCOR	ADAR	LRP1	C5orf42	MED12
AP1S1	LRP2	BRWD3	NOTCH3	C5orf42	ALDH18A1	MAPT	CACNA1C	MEGF8
APEX1	MAF	CACNA1C	PC	CDKN1C	ASXL1	MDC1	CCDC88C	MPDZ
ARID1A	MDC1	COL4A1	PEX5	CELSR1	ATXN1	MEGF8	CHD7	NDUFV1
ARID1B	MEF2D	CRADD	PHF21A	ETFB	BCOR	MN1	CRX	NFATC1
ATP7A	MEGF8	CTBP2	PIAS3	FOXA2	C5orf42	NAGLU	CSNK2A2	NFATC2
BAG6	MYO1E	CUL4B	PML	HERC1	CCDC78	NFATC2	DHCR7	NFATC4
BCOR	NCOA6	DACH1	POLR3B	HNF1B	CCDC88C	NFATC4	DNAH5	NOTCH3
C5orf42	NDP	DNAH5	RBFOX1	HOXD4	CIITA	NHS	ERCC6	PEX5
CACNA1C	NDUFA11	DPCD	RUNX1	IRF1	COL4A2	PEX5	EVC	PGAP2
CC2D1A	NF1	DYM	RUNX2	IRF7	CRX	PIK3R2	FLG	PIGO
CDH15	NOTCH3	FGD1	RUNX3	LARP7	CTBP2	PYCR1	FOS	PIGV
CDK5RAP2	NR4A2	GIPC1	SETBP1	LRP1	CUL4B	RAI1	FOXP2	POU2F3
CELSR1	OTC	GNRHR	SH3PXD2B	MTF1	DOCK8	RELA	FREM2	PYCR1
CIITA	PAX6	GRIK2	SHANK2	NOTCH1	EGR1	RP2	GNAS	RAB27A
CNOT3	PIAS3	HERC1	SLC12A6	PIGN	ETFB	RUNX3	GTF2F1	RELA
COL2A1	PIGO	HLCS	SLC9A6	PPARGC1A	EVC	SETBP1	HNF4A	RMND1
COL4A2	PLAGL1	HOXB3	SMAD2	PPM1D	FKRP	SLC12A1	HSD17B4	SHANK3
CREBBP	PORCN	HUWE1	SMC1A	SKI	FOS	SRCAP	KAT6B	TCF3
CTBP2	POU2F3	KLF1	SREBF1	SMARCA2	HCFC1	SREBF1	KMT2D	TP73
D2HGDH	RECQL4	KPNA2	SRF	ST3GAL3	HNF4A	STAT2	LAMC3	TSC1
DOCK6	RELA	LAMC3	TCF3	TGFBR2	HRAS	TBP	MAN1B1	VDR
EP300	RYR1	LHX3	TNK2	TTN	HSF1	TNK2		
FANCD2	SBF1	LRP1	TPO	WDR81	HTT	TP73		
FLNA	SCN8A	LRP2	TREX1		IRS1	TPK1		
FOS	SMARCA4	MBD5	TSG101		KMT2D	VDR		
FOXO1	SMS	MED25	TTN					
GAK	SOX2	MKS1	ZBTB17					
GTF3C1	SRGAP3							
HDAC4	SYNE1							
HOXB2	TMCO1							
HSD17B4	UBR1							
ING2	VDR							
IQSEC2								

Appendix 5

Random selected homopolymers

Variants shown in blue indicate homopolymers

AmpliSeq

Location	SYMBOL	Consequence	Amino_acids	Codons
1:203143609	MYBPH	frameshift_variant	A/X	Gct/ct
1:85561698	WDR63	frameshift_variant	G/X	Ggg/gg
11:114393802	NXPE1	frameshift_variant	I/X	Ata/ta
11:73850782	C2CD3	frameshift_variant	K/X	aAg/ag
12:6345381	CD9	frameshift_variant	E/X	gaG/ga
14:67940232	TMEM229B	frameshift_variant	R/X	Cgc/gc
16:1841068	IGFALS	frameshift_variant	H/X	Cac/ac
17:33310276	LIG3	frameshift_variant	G/X	ggC/gg
18:31599377	NOL4	frameshift_variant	I/X	Ata/ta
19:48806002	CCDC114	frameshift_variant	L/X	Ctt/tt
2:203620315	FAM117B	frameshift_variant	G/X	Ggc/gc
20:31669389	BPIFB4	frameshift_variant	S/X	agC/ag
22:47569181	TBC1D22A	frameshift_variant	W/X	tGg/tg
3:138402533	PIK3CB	frameshift_variant	F/X	ttT/tt
4:148555005-148555006	TMEM184C	frameshift_variant	EG/EX	gaGGgc/gagc
5:156918873	ADAM19	frameshift_variant	E/X	gaA/ga
6:99883711	USP45	frameshift_variant	A/X	Gct/ct
7:103048398	SLC26A5	frameshift_variant	G/X	gGc/gc
9:123202064	CDK5RAP2	frameshift_variant	L/X	tTa/ta
X:53592128	HUWE1	frameshift_variant	P/X	ccC/cc

TargetSeq

Location	SYMBOL	Consequence	Amino_acids	Codons
1:156812959	INSRR	frameshift_variant	G/X	gGc/gc
1:205132979	DSTYK	frameshift_variant	A/X	Gct/ct
10:100012153	LOXL4	frameshift_variant	G/X	ggG/gg
11:124120521	OR8G1	frameshift_variant	G/X	ggA/gg
12:52115400-52115401	SCN8A	frameshift_variant	R/X	aGG/a
13:32913435-32913436	BRCA2	frameshift_variant	A/AX	gca/gcAa
14:45605351	FANCM	frameshift_variant	K/X	aaG/aa
15:66794986	RPL4	frameshift_variant	A/X	Gct/ct
16:31088894	ZNF646	frameshift_variant	K/X	Aaa/aa
17:18034583	MYO15A	frameshift_variant	F/X	Ttc/tc
18:76886364	ATP9B	frameshift_variant	L/X	cTt/ct
3:64132969	PRICKLE2	frameshift_variant	E/X	gaG/ga
5:42711437	GHR	frameshift_variant	V/X	gtA/gt
6:43096970	PTK7	frameshift_variant	A/X	gCc/gc
8:39091451-39091452	ADAM32	frameshift_variant	YP/YX	taCCct/tact
9:34723632	FAM205A	frameshift_variant	K/X	aAg/ag
X:153044463	PLXNB3	frameshift_variant	A/X	gCc/gc
22:38130835	TRIOBP	frameshift_variant	P/X	Ccc/cc
20:31628069	BPIFB6	frameshift_variant	K/X	Aag/ag
4:79832573	BMP2K	frameshift_variant	P/X	Ccc/cc

HaloPlex

Location	SYMBOL	Consequence	Amino_acids	Codons
1:109198317	HENMT1	frameshift_variant	A/X	Gca/ca
1:183091263	LAMC1	frameshift_variant	G/X	Gga/ga
1:65690527	AK4	frameshift_variant	V/X	gtG/gt
11:18231607	RP11-113D6.10	frameshift_variant	V/X	gtG/gt
11:9090979	SCUBE2	frameshift_variant	G/X	gGc/gc
12:68051327	DYRK2	frameshift_variant	P/X	Ccc/cc
15:63953987	HERC1	frameshift_variant	A/X	gcC/gc
16:84529360	TLDC1	frameshift_variant	E/X	Gag/ag
17:67249739	ABCA5	frameshift_variant	D/X	Gat/at
19:33628630	WDR88	frameshift_variant	C/X	tgC/tg
2:177016517	HOXD4	frameshift_variant	Y/X	taC/ta
20:22562963	FOXA2	frameshift_variant	G/X	gGc/gc
8:145694204	KIFC2	frameshift_variant	E/X	gAg/gg
4:119653999	SEC24D	frameshift_variant	L/X	ctA/ct
5:134010900	SEC24A	frameshift_variant	N/X	Aac/ac
6:114279855	HDAC2	frameshift_variant	R/X	Aga/ga
7:135269725	NUP205	frameshift_variant	I/X	atC/at
3:130187865	COL6A5	frameshift_variant	L/X	ctG/ct
9:79259766-79259767	PRUNE2	frameshift_variant	AE/AX	gcCGaa/gcaa
X:100911718	ARMCX2	frameshift_variant	V/X	gTa/ga

CHAPTER 5 - General discussion and future directions of human genetic and (neuro)developmental disease research

In this study, iPSCs were generated from fibroblasts taken from RTS patients with a known heterozygous mutation in the *CREBBP* gene, which results in reduced CBP protein. CBP, which is ubiquitously expressed in most cells, is an important transcriptional co-activator with intrinsic lysine acetyltransferase activity, and plays important roles in basic cellular processes, also during development. This study demonstrated that somatic cells with reduced levels of CBP can be reprogrammed into iPSCs. This was performed alongside WT fibroblasts, showing a normal karyotype for both WT and RTS iPSCs lines. These iPSC lines were characterized by the expression of several pluripotent markers such as NANOG, TRA-1-60 and OCT3/4. Lines from all samples were able to generate teratomas, showing differentiation potential into all three germ layers, and thereby confirming functional pluripotency of the RTS iPSCs. In conclusion, the reduced amount of CBP causing RTS in humans does not appear to have an effect on reprogramming, demonstrated by the generation and characterization of patient specific iPSCs.

Several disease studies have been performed through the generation of patient-specific stem cells. Park et al. reprogrammed somatic cells, either dermal fibroblast or bone marrow-derived mesenchymal cells, from patients with a range of human genetic diseases [1]. These included Mendelian or complex diseases such as Parkinson disease, Duchenne muscular dystrophy and Down syndrome. They characterized these iPSCs and showed normal karyotypes, pluripotent markers and teratoma formation.

Even though these novel techniques are available, most research into CBP and RTS has been in animal models, and human cell studies are still limited. Therefore, this *in vitro* study can give additional information about RTS and CBP and can play a role in the development of therapeutics for RTS.

In addition to the small cohort of this study, another limitation was the compatibility of WT and RTS cell lines. As controls, fibroblast cell lines were used that did not completely resemble the origin and genetic background of the RTS samples. RTS fibroblasts were taken from the upper body (sternum and forearm) of two male patients, whereas the WT lines were either isolated from female dermal breast tissue, or male foreskin cells. These different backgrounds could have an influence on further experiments, such as differentiation potential and expression profiles.

Therefore, this study cohort should be expanded with both WT and RTS samples, and include samples from RTS patients with a mutation in *EP300*. Due to slight

differences in RTS phenotype and certain unique functions between the two proteins, it would be interesting to compare the reprogramming processes of cells with either reduced CBP or p300 protein levels. Because mutations in *EP300* are rare, it is difficult to obtain cells from these RTS cases. However, this and the background dissimilarities with control samples can be resolved by using gene-editing methods, such as the CRISPR-Cas9 system, to introduce specific mutations in WT cells and thereby generating isogenic disease and control pairs [2].

This has been done in a study researching early onset Alzheimer's disease. They introduced heterozygous and homozygous dominant mutations in human iPSCs in the *APP* and *PSEN1* genes using CRISPR-Cas9, generating isogenic controls to facilitate the study of human disease [3]. Combining these methods would give more insights into the role of CBP and p300 in reprogramming and development to expand human research towards RTS.

Next, the generated RTS iPSCs were differentiated into neuronal cells. There was a difference observed in mature neuronal markers throughout differentiation. These results showed that even though SOX1, a NPC marker, showed similar levels, the RTS iPSCs appeared to produce fewer cells positive for the mature neuronal markers TUJ1 and MAP2 compared to the WT lines. Addition of an HDAC inhibitor during neuronal differentiation, led to an increase in the number of TUJ1$^+$ cells in the RTS lines. This suggests that potentially an HDAC inhibitor can rescue the RTS phenotype seen during the neuronal differentiating process of these disease-specific iPSCs (Figure 1).

To reduce tissue of origin and genetic background differences between cell lines, shRNA was used to knockdown *CREBBP* in WT iPSCs to generate isogenic control cell lines with reduced CBP levels. These experiments showed a highly comparable outcome to the RTS differentiation results, both with and without the HDAC inhibitor.

Even though RTS iPSCs generate normal NPCs, which was reflected in the number of SOX1$^+$ cells, the reduced levels of mature neurons detected in RTS cultures could potentially explain the intellectual disability and/or behavioural problems seen in RTS. The increase in TUJ1$^+$ cells, induced by the HDAC inhibitor, could consequently lead to a novel treatment for RTS patients to reduce certain symptoms. These findings provide strong evidence that CBP regulates the differentiation of

neurons in humans, and offers new insights into the neurological phenotype seen in RTS patients.

RTS mouse model data, which formed the basis of these experiments, have shown similar results. Knockdown of CBP in cortical precursors caused a decrease in mature neurons, and this phenotype was rescued by the addition of an HDAC inhibitor. Other *in vitro* studies using iPSCs generated from patients with neurodevelopemental diseases also showed promising results. This has been done with iPSCs generated from cells taken from patients with Rett syndrome, which is characterized by developmental delays and autism. This study showed that even though there was no difference in neuronal survival, measured by MAP2 expression, there were fewer synapses and smaller soma size detected in the RTT neurons compared to WT [4].

Even though several mouse *Crebbp* knockdown models have been generated over the years that resemble the RTS phenotype, besides structural abnormalities, there is inadequate research into the neuronal basis of RTS patients. Therefore, neuronal differentiation of RTS iPSCs gives new insights into the disease and in combination with a HDAC inhibitor can lead to therapeutics.

An established protocol was used to differentiate iPSCs in a monolayer system into a cortical neuronal population. Ideally, more data needs to be generated by using other differentiation systems, screening more neuronal markers, potentially in combination with FACS and/or single cell gene expression analysis [5, 6].

To create a *CREBBP* knockdown system in WT iPSCs, shRNA was used. Generally, shRNA does not induce complete knockdown of the target gene [7]. Fortunately, this model required a knockdown of 50% for *CREBBP* to reflect the heterozygosity in RTS. Therefore, multiple shRNA constructs were tested to induce ~50% *CREBBP* knockdown. However, in future the CRISPR-Cas9 system would be ideally applied to make a more efficient and realistic RTS model. With this system heterozygous mutations published as RTS causing variants in both *CREBBP* and *EP300* genes can be introduced.

Since RTS patients have a variety of symptoms that include different organ malformations, it would be interesting to differentiate these RTS iPSCs into other cell types, such as cardiomyocytes. This system can also be combined with an extended drug screen cohort to discover more specific therapeutic options. This has been done for Alzheimer's disease, where iPSCs from these patients were differentiated into neuronal cells and a drug screen with three substances was performed. This showed

that each drug had specific susceptibilities during different stages of differentiation [8].

Currently, the diagnosis of RTS is postnatal and in some cases not diagnosed until later in life. However, even though RTS is rare, prenatal screening assays will eventually be based on complete genome sequencing. In future, when screening cohorts include both *CREBBP* and *EP300*, potentially, HDAC inhibitors can be administered to women during pregnancy expecting a child with RTS to either reduce or control symptoms. This has been done with valproic acid (VPA) in murine models and when prenatally exposed, this showed behavioural alterations similar to those observed in humans with autism. The HDAC inhibitor induced transient hyperactylation in the embryonic brain, which caused changes in autism-related molecules linked to a delay of neuronal maturation. However, this was seen when administered at E12.5, but not at E14.5 and this indicates that the time of drug administrating is very important during embryonic development for the outcome [9, 10]. These results suggests that more research is needed to focus on the side effects of HDAC inhibitors during embryonic development, which could lead to hypoacetylation resulting in deleterious effects, before administration to pregnant women.

In this study, a monolayer to generate neuronal cells was used, but other methods of neuronal differentiation could result in more realistic models of neurodevelopmental diseases. This has been achieved by the generation of mini brains [6], and it would be valuable to make these 3D structures with the RTS iPSCs which would allow for better analysis of synapse formation and function.

Even though most genetic mutations are irreversible and therapeutic approaches are limited, the development of the CRISPR-Cas9 technology has been introduced as an efficient genome-editing tool [11, 12]. Besides using this technique to create targeted mutations, excluding the need for patient samples and providing isogenic controls, CRISPR-Cas9 has also been used as a gene-correcting tool [13]. This has recently been done in combination with human iPSCs to correct defective genotypes *in vitro*, with DNA repair templates and homology directed repair (HDR) for several diseases and conditions [14-17]. In a cystic fibrosis (CF) model, iPSCs were generated from CF patients with a homozygous deletion in the *CFTR* gene. CRISPR-Cas9 was used to target sequences in this gene and to precisely correct the deletion causing CF [18].

More recently, this has been successfully applied in *in vivo* models. A Duchenne muscular dystrophy (DMD) mouse model showed improvements in muscle function after *in vivo* genome editing with CRISPR-Cas9 [19-23], which could be translated to humans in future to correct disease-causing mutations in the muscle tissue of DMD patients in future.

This shows that CRISPR-Cas9 based approaches can correct disease alleles in both *in vitro* and *in vivo* models. Combining these genome editing tools with 3D models of tissues and organs generated from patient specific iPSCs, would provide an *ex vivo* method, with potential to be transplanted back into the patients without rejection. In combining with drug screens, this can give a better insight into the effect of altered CBP/p300 function in clinically significant human cell types and tissues (Figure 2).

Even though genome editing sounds promising for genetic diseases, only ~65% of RTS cases can be genetically diagnosed with a mutation in either *CREBBP* or *EP300*, which means ~35% of cases is without a genetic diagnosis. Through whole exome sequencing (WES) of six RTS patients, one novel RTS causing variant in the *CREBBP* gene was detected. This heterozygous C>T substitution in exon 26, causes a premature stop codon, leading to a truncated transcript. These reduced levels of functional CBP are predicted to result in an RTS phenotype. Unfortunately, applying the in house designed filtering strategy to the WES data, did not result in the discovery and confirmation of an additional RTS causing gene.

However, a putative, single exon deletion was detected with MLPA in the *CREBBP* gene in a RTS sample. However, further examination in combination with the WES data and UCSC browser, showed that this deletion appeared to be a known, but rare, SNP near the ligation site of the MLPA probes. Even though it emerged to be a false positive, this shows that MLPA is a sensitive and high throughput screening method. However, especially single exon deletions or duplications need to be confirmed with other sequencing techniques.

Even though WES studies have promising results in disease gene identification, only one relevant novel mutation was identified. However, Iossifiv et al., performed WES on more than 2,500 families with a child diagnosed with autism spectrum disorder (ASD). This study discovered multiple *de novo* missense mutations and CNVs in ~30% of cases [24]. In addition, Zhu et al., analyzed 119 trios using WES and identified a genetic diagnosis for 29 (24%) of patients who were initially undiagnosed [25]. Even thought these numbers appear low, this is an enormous contribution to

disease gene discovery. Through both technical and bioinformatic developments, these numbers will only rise in future. This is of great significance, as diagnosing the disease both clinically and genetically is very important for patients and their families. Knowing the underlying cause of the disease can contribute heavily to deciding the most appropriate treatment options.

For future studies, a larger RTS patient cohort would be desirable. DNA samples were collected from 12 RTS patients, of which six samples showed sufficient DNA quality for whole exome sequencing. Although more costly, and with the ethical complications of unsolicited findings, ideally trio analysis for identifying rare genetic conditions would contribute to the efficiency. Whole genome sequencing, in comparison to WES, would also be expected to increase the diagnostic yield [26]. This allows the identification of genetic variants in non-coding DNA, such as regulatory elements. Mutations in these regions could be disease causing by disrupting expression levels of specific genes.

Together, these studies would provide a vast amount of information in correlation to disease mechanisms. In combination with drug screening they could lead to rapid drug development, resulting in relevant therapeutic options. Modern genome engineering platforms could add another layer of possibilities and is a promising area for genome research to improve human health [27-29]. In conclusion, this study connects genomics, therapeutic targets and disease phenotypes into one model. Even though more research is needed, hopefully this will lead to better diagnostic tools, not only for RTS but also for other (neuro)developmental disorders.

The work described in this thesis will be at the basis for two research papers. The first one will be reporting the generation of RTS-specific iPSCs and will be submitted to Stem Cell Research. The second paper will outline the neuronal differentiation process and the effect of an HDAC inhibitor on the differentiation potential of these RTS-iPSCs. Furthermore, the introduction will be adjusted to be published as a review paper. Finally, the book chapter in the Appendix, describing the MLPA method, has been published recently.
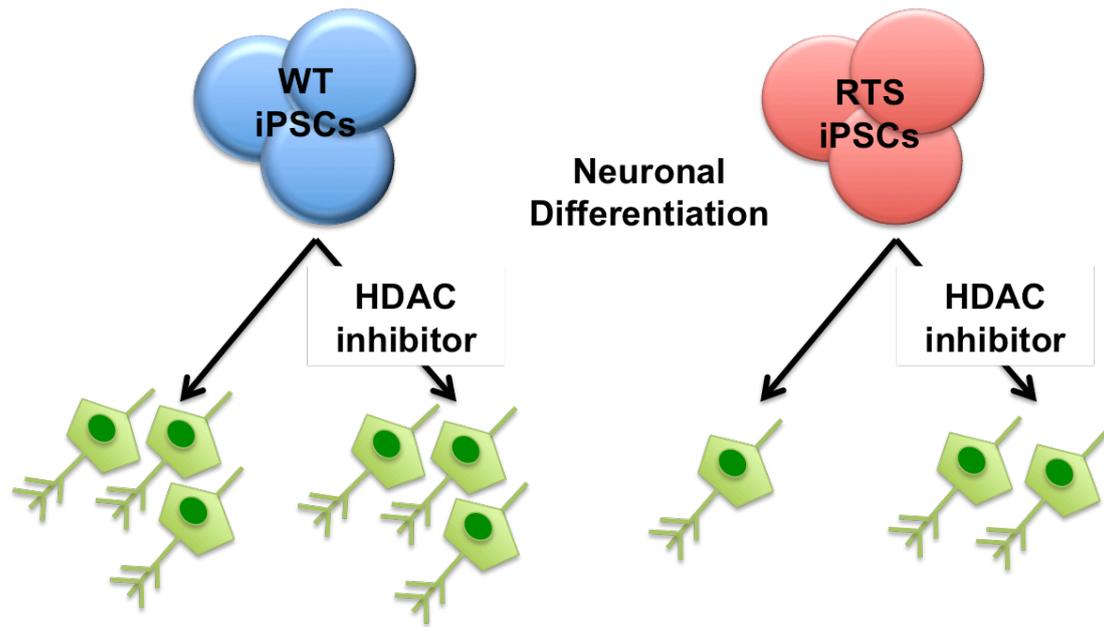
Figure 1. Addition of an HDAC inhibitor can stimulate neuronal differentiation in RTS cells. RTS iPSCs with a mutation in the *CREBBP* gene show a decline in neuronal differentiation potential compared to WT cells. Addition of an HDAC inhibitor can stimulate the formation of neuronal cells and bringing the number more towards WT levels.
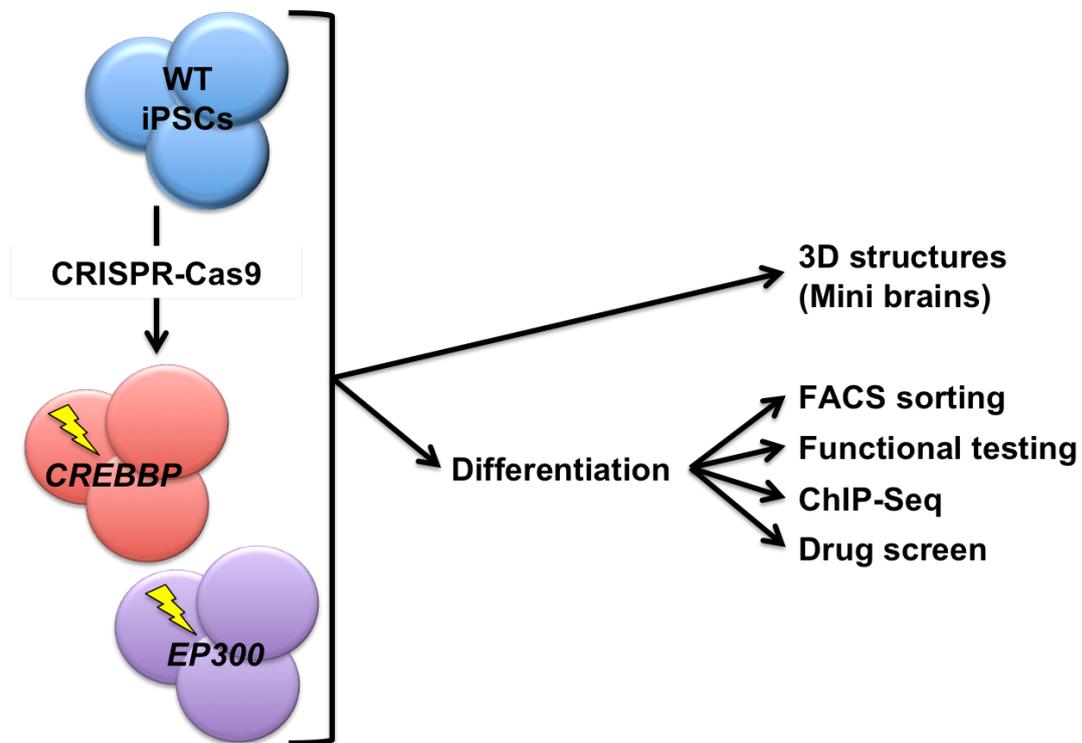
Figure 2. RTS disease modelling using genome editing tools. Genome editing tools, such as CRISPR-Cas9, can be used to introduce RTS specific mutations into WT iPSC lines, thereby creating isogenic controls. These can be directly used to generate 3D structures, such as mini brains or be differentiated into any cell type of interest, related to the studied disease. In addition, these differentiated disease-specific cells can be used in down stream experiments, such as FACS sorting to look at cell markers, functional testing, such as calcium clamping for neurons, ChIP-Seq to generate epigenetic profiles, and performing drug screens. Combining these methods, a complete disease in dish model can be generated, contributing to a better understanding and possible therapeutics for RTS and other diseases.

References

1. Park, I.H., et al., *Disease-specific induced pluripotent stem cells.* Cell, 2008. **134**(5): p. 877-86.
2. Heidenreich, M. and F. Zhang, *Applications of CRISPR-Cas systems in neuroscience.* Nat Rev Neurosci, 2016. **17**(1): p. 36-44.
3. Paquet, D., et al., *Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9.* Nature, 2016. **533**(7601): p. 125-9.
4. Marchetto, M.C., et al., *A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells.* Cell, 2010. **143**(4): p. 527-39.
5. Pasca, S.P., et al., *Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome.* Nat Med, 2011. **17**(12): p. 1657-62.
6. Mariani, J., et al., *Modeling human cortical development in vitro using induced pluripotent stem cells.* Proc Natl Acad Sci U S A, 2012. **109**(31): p. 12770-5.
7. Moore, C.B., et al., *Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown.* Methods Mol Biol, 2010. **629**: p. 141-58.
8. Yahata, N., et al., *Anti-Abeta drug screening platform using human iPS cell-derived neurons for the treatment of Alzheimer's disease.* PLoS One, 2011. **6**(9): p. e25788.
9. Kataoka, S., et al., *Autism-like behaviours with transient histone hyperacetylation in mice treated prenatally with valproic acid.* Int J Neuropsychopharmacol, 2013. **16**(1): p. 91-103.
10. Kawanai, T., et al., *Prenatal Exposure to Histone Deacetylase Inhibitors Affects Gene Expression of Autism-Related Molecules and Delays Neuronal Maturation.* Neurochem Res, 2016. **41**(10): p. 2574-2584.
11. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.
12. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.
13. Barrangou, R. and J.A. Doudna, *Applications of CRISPR technologies in research and beyond.* Nat Biotechnol, 2016. **34**(9): p. 933-941.
14. Song, B., et al., *Improved hematopoietic differentiation efficiency of gene-corrected beta-thalassemia induced pluripotent stem cells by CRISPR/Cas9 system.* Stem Cells Dev, 2015. **24**(9): p. 1053-65.
15. Xie, F., et al., *Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac.* Genome Res, 2014. **24**(9): p. 1526-33.
16. Chen, W., et al., *Corrigendum: Generation of the SCN1A epilepsy mutation in hiPS cells using the TALEN technique.* Sci Rep, 2015. **5**: p. 11174.
17. Yang, Y., et al., *Naive Induced Pluripotent Stem Cells Generated From beta-Thalassemia Fibroblasts Allow Efficient Gene Correction With CRISPR/Cas9.* Stem Cells Transl Med, 2016. **5**(2): p. 267.
18. Firth, A.L., et al., *Functional Gene Correction for Cystic Fibrosis in Lung Epithelial Cells Generated from Patient iPSCs.* Cell Rep, 2015. **12**(9): p. 1385-90.
19. Wu, Y., et al., *Correction of a genetic disease in mouse via use of CRISPR-Cas9.* Cell Stem Cell, 2013. **13**(6): p. 659-62.

20.   Long, C., et al., *Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA*. Science, 2014. **345**(6201): p. 1184-8.

21.   Tabebordbar, M., et al., *In vivo gene editing in dystrophic mouse muscle and muscle stem cells*. Science, 2016. **351**(6271): p. 407-11.

22.   Nelson, C.E., et al., *In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy*. Science, 2016. **351**(6271): p. 403-7.

23.   Ousterout, D.G., et al., *Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy*. Nat Commun, 2015. **6**: p. 6244.

24.   Iossifov, I., et al., *The contribution of de novo coding mutations to autism spectrum disorder*. Nature, 2014. **515**(7526): p. 216-21.

25.   Zhu, X., et al., *Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios*. Genet Med, 2015. **17**(10): p. 774-81.

26.   Goodwin, S., J.D. McPherson, and W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies*. Nat Rev Genet, 2016. **17**(6): p. 333-51.

27.   Hendel, A., et al., *Quantifying on- and off-target genome editing*. Trends Biotechnol, 2015. **33**(2): p. 132-40.

28.   Iyer, V., et al., *Off-target mutations are rare in Cas9-modified mice*. Nat Methods, 2015. **12**(6): p. 479.

29.   Smith, C., et al., *Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs*. Cell Stem Cell, 2014. **15**(1): p. 12-3.

APPENDIX MLPA BOOK CHAPTER

De Boer CM and White SJ, *Genotyping Complex Copy Number Variation with Multiplex Ligation-dependent Probe Amplification (MLPA).* Methods Mol Biol, 2017. 1492: p 147-153. (www.ncbi.nlm.nih.gov/pubmed/27822861)

Genotyping Complex Copy Number Variation with Multiplex Ligation-dependent Probe Amplification (MLPA)

Carmela M. de Boer[1] and Stefan J. White[2]

[1]Department of Anatomy & Developmental Biology Monash University, Clayton, Victoria 3800, Australia

[2]Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands.

Running title: CNV analysis with MLPA

Key words: Copy number variation, Deletion, Duplication, Capillary electrophoresis,

Abstract

Multiallelic copy number variants are genomic loci that can be present in a range of different copy numbers between individuals. High or low copy numbers of specific genes have been associated with different diseases. Accurate genotyping of these loci can be complicated, and relies on accurate assays. Multiplex Ligation-dependent Probe Amplification (MLPA) is a PCR-based approach that allows copy number determination of up to 50 genomic loci in a single reaction. In this chapter we outline the basic protocol, with a particular emphasis on the appropriate approach to accurately genotype complex copy numbers.

1. Introduction

There are many types of genetic variation in the human genome. One class is copy number variation, defined as a gain or a loss compared to the reference genome. A number of loci show a wide range of copy numbers between individuals, which collectively are known as multiallelic copy number variants, or mCNV.

A number of different methodologies have been applied to the analysis of mCNV (reviewed in [1]). One approach is Multiplex Ligation-dependent Probe Amplification (MLPA), a PCR-based technique first described in 2002 [2].

MLPA is based around the ligation of two half probes which recognise a specific sequence of interest (Figure 1). Ligation only will occur when both half probes are hybridized to their target sequences, and only ligated probes are amplified simultaneously during the PCR reaction. Because the probes contain identical ends, the ligated products can be amplified together with a single primer pair. One of the two primers in the PCR is fluorescently labelled, meaning that the amplified products can be visualized during fragment separation by capillary electrophoresis. Each probes is designed to have a unique length, and relative differences in peak heights correspond to changes in copy number (Figure 2).

The principle advantages of MLPA are that it allows for a rapid (results being available within 24 hours) and high-throughput quantification (96 samples can be handled simultaneously) of up to 50 sequences per DNA sample in a single reaction, by using a single PCR primer pair. It has been adapted to a range of different applications, including gene expression [3] and methylation analysis [4].
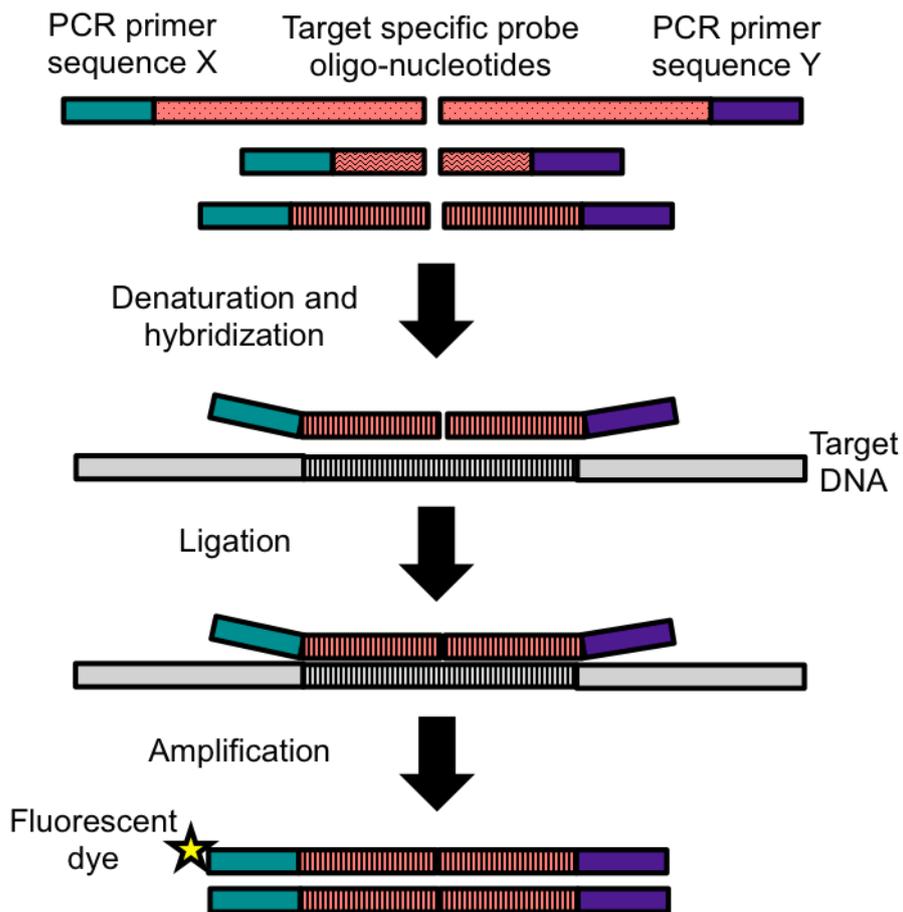
Figure 1. The basis of MLPA. Genomic DNA is denaturized, with the half-probes hybridizing to the single-stranded DNA. Only half-probes that hybridize adjacently can be ligated together, and only ligated products can amplified with PCR.
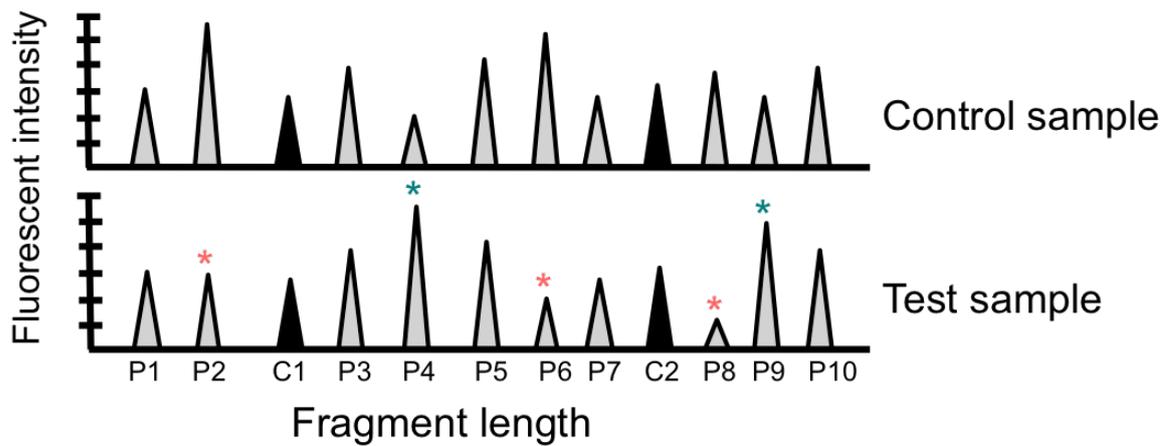
Figure 2. Fragment separation by capillary electrophoresis distinguishes probes by their unique length, and will generate a consistent peak pattern. The Peak heights of control probes (C1, C2) should be consistent between samples. Comparing relative differences in test probes (P1-P10) between the control sample and the test sample shows a decrease in copy number for probes P2, P6, and P8 (marked with *), and a gain in copy number for probes P4 and P9 (marked with #).
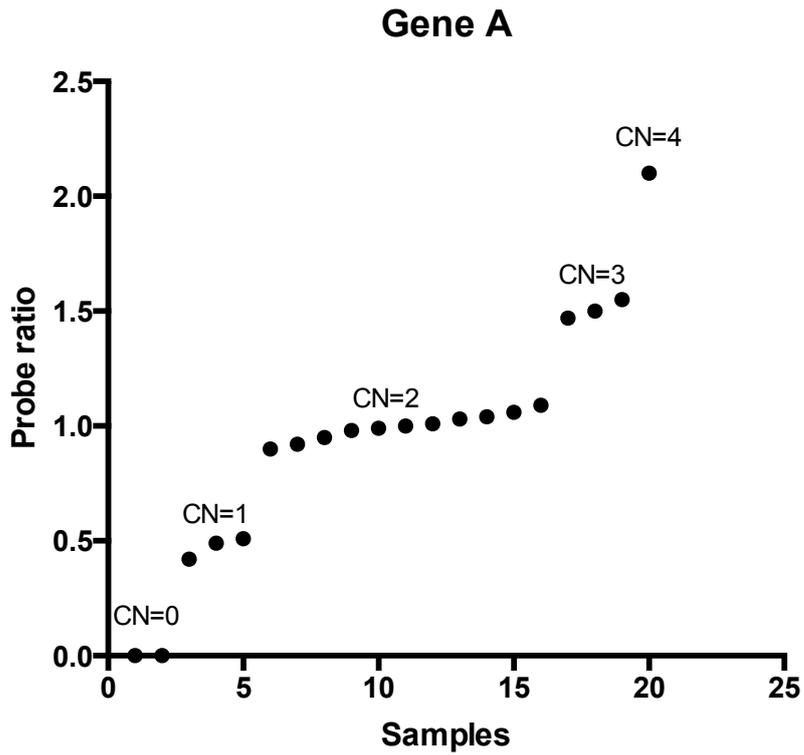
Figure 3. Tight clustering of normalized ratios for each locus simplifies data interpretation. In this case the average difference between groups is 0.5, supporting the conclusion that the copy numbers (CN) range from 0-4.

2. Materials

1. MLPA reagents.

All reagents for this MLPA protocol can be purchased from MRC-Holland, The Netherlands (www.mlpa.com). The different components can be recognised by a distinguishing cap colour.

SALSA MLPA buffer (yellow cap)

SALSA Ligase-65 (green cap)

Ligase Buffer A (transparent cap)

Ligase Buffer B (white cap)

SALSA Polymerase (orange cap)

SALSA PCR Primer Mix (brown cap)

The PCR Primer Mix contains the following primers

Forward 5'-GGGTTCCCTAAGGGTTGGA-3'

Reverse 5'-GTGCCAGCAAGATCCAATCTAGA-3'

The forward primer is fluorescent labelled at the 5' end, usually with FAM.

Probe mix (black cap)

MRC has a variety of ready to order probe mixes (black cap), or homemade probes can be developed using synthetic oligonucleotides [5]. When designing probes, the CG content of the hybridizing sequence of each half probe should be 35-60%, and the Tm should be greater than 66°C. Finally, to maximize signal strength, it has been shown that the first nucleotide of the unique sequence of the left half probe should be a C or a G. The right oligonucleotide should be phosphorylated at the 5' end, to allow ligation to take place.

MLPA probes are typically designed against unique sequences in the reference genome. Polymorphic loci are usually represented more than once, so extra care must be taken when choosing probe sequences. If a class of homologous genes is to be assayed then it is important to choose sequences that are identical across all genes. Conversely, if a specific gene is to be studied then the oligonucleotides should be chosen such that any sequence mismatches are at or near the ligation site. Although a

single mismatch may be sufficient to generate a specific product, it is preferable for multiple nucleotides to be different.

As probes are typically separated by capillary electrophoresis, it is essential that each probe has a different length. We have successfully used probes generating products within the size range of 80-150 bp.

2. Additional materials and reagents

Thermocycler with heated lid

Filter tips

PCR strip tubes with individual lids

Hi-Di Formamide (Applied Bioscience)

Size standard (Applied Bioscience)

3. Methods

The MLPA protocol below is an updated version of that described in the original publication [2]. It is also available at the website of MRC-Holland (www.mlpa.com).

1. Add 20-500 ng genomic DNA in a final volume of 5 μl to a PCR tube (Note 1).

2. The DNA is denatured at 98°C for 5 minutes and allowed to cool to room temperature for at least 5 minutes (Note 2).

3. To the denatured DNA add 1.5 μl MLPA probe mix and 1.5 μl SALSA MLPA buffer, and mix with care. Incubate for 1 minute at 95°C, then 16 hours at 60°C (Note 3).

4. Prepare the ligase mix at room temperature. Mix 3 μl Ligase-65 buffer A and 3 μl Ligase-65 buffer B in 25 μl H_2O. Add 1 μl Ligase-65 and mix again.

5. Reduce the temperature of the thermal cycler to 54°C. While keeping the PCR tubes in the thermal cycler, add 32 μl of the ligase mix to each tube and mix (Note 4). The reactions should be incubated for 10-15 minutes at 54°C, followed by 5 minutes at 98°C to inactivate the ligase (Note 5).

6. To make the polymerase master mix, prepare the following for each reaction (Note 6):

H_2O	7.5 μl
SALSA PCR primer mix	2 μl
SALSA Polymerase	0.5 μl

7. Store on ice until use. At room temperature, add 10 μl polymerase master mix to each tube containing the MLPA ligation reaction and mix by gently pipetting.

8. Place the tubes in the thermocycler and run the PCR reaction with the following settings

1 cycle: 1 minute 95°C

35 cycles: 30 seconds 95°C; 30 seconds 60°C; 30 seconds 72°C

1 cycle: 20 minutes 72°C

9. Prepare samples for fragment analysis on a capillary sequencer. Add 5 μl size standard to 1 ml Hi Di Formamide and mix. Into each well of a 96-well plate, add 9 μl of the Formamide/size standard mix, and then add 1 μl of PCR product to each well.

10. Data analysis. Fragment separation is usually performed on a capillary sequencer, which measures absolute fluorescence. Peaks generated by capillary sequencing require normalisation, which consist of two steps. First there is intrasample normalisation, where the height of each probe peak is compared to the peak heights of reference probes within a sample to produce a ratio. An intersample normalisation is then performed, by dividing each probe ratio by the median value of the matching probe ratios across all samples.

For typical diploid loci this normalised ratio will be 1.0, with deleted and duplicated loci within individual samples having normalised ratios of ~0.5 and 1.5 respectively. When analysing mCNP loci this will not be the case. There have been different approaches described for assigning specific copy numbers to samples when a range of copy numbers is expected. For high quality data it may be possible to identify distinct groups by eye. The copy number of each group can then be estimated by determining the proportional difference between the groups (Figure 3). Copy number grouping can be improved by having multiple probes per locus, and using the average value [6, 7]. For less clear data it is possible to bin samples into arbitrary groups based on predefined borders, however this has the chance of introducing bias.

4. Notes

1. High quality DNA, isolated in a consistent manner, is essential for a successful MLPA analysis. A degree of degradation can be tolerated, as the DNA sequence used as template for oligonucleotide hybridization is usually <200 bp. Sites of DNA breakage are unlikely to completely random, however, meaning that a commonly degraded locus may appear as a (somatic) deletion. Impurities such as phenol can influence the MLPA reaction. The method in which the DNA is isolated and purified can have a subtle impact on relative peak heights, meaning that otherwise high quality DNA samples, prepared using different protocols, may give spurious copy number differences. For this reason, it is best for all DNA samples within a study to be isolated using the same procedure. Similarly, if two or more study populations are being analysed, the samples should be randomised during the MLPA procedure, rather than processing the populations in separate batches on different days.

2. Incomplete denaturation of the DNA can lead to reduced probe access, which can be resolved by increasing the denaturation time.

3. This protocol uses an overnight hybridization step, although it has been shown that 2-3 hours is sufficient in some cases [8]. However, all probes have a variable annealing time, therefore a shorter hybridization step should be tested prior to routine implementation.

4. The ligation step requires to be performed at 54°C, therefore keep tubes in the thermocycler at 54°C when adding the ligation mix.

5. The ligated products are stable, and can be stored for months at -20°C.

6. The fluorescent labels are light sensitive. Minimize exposure of the primer mix to light during the pipetting steps, and PCR products should be stored in the dark.

8. This protocol assumes the use of capillary sequencer ABI3700 (Applied Biosystems). Other machines may have different requirements.

9. The optimal amount of PCR product needed may be different, depending on the sensitivity of the specific sequencer.

10. The use of a thermocycler with heated lid is essential, to prevent condensation disturbing the reaction.

11. Prior to use, briefly centrifuge all MLPA reagents (to remove drops from the inside of the lids). Mix buffers by briefly vortexing, and enzymes by gently pipetting up and down.

12. Prepare master mixes (ligation and PCR) <1 hour before use and store the PCR mix on ice. When making master mixes, enzymes should be added last.

13. It is recommended to use filter tips throughout the procedure to prevent contamination.

References

1.   Cantsilieris, S., Baird, P. N., and White, S. J. (2013). Molecular methods for genotyping complex copy number polymorphisms. *Genomics* **101**:86-93.

2.   Schouten, J. P., McElgunn, C. J., Waaijer, R., et al. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* **30**:e57.

3.   Eldering, E., Spek, C. A., Aberson, H. L., et al. (2003). Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. *Nucleic Acids Res* **31**:e153.

4.   Nygren, A. O., Ameziane, N., Duarte, H. M., et al. (2005). Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res* **33**:e128.

5.   den Dunnen, J. T., and White, S. J. (2006). MLPA and MAPH: sensitive detection of deletions and duplications. *Curr Protoc Hum Genet* **7**:1-14.

6.   Groth, M., Szafranski, K., Taudien, S., et al. (2008). High-resolution mapping of the 8p23.1 beta-defensin cluster reveals strictly concordant copy number variation of all genes. *Hum Mutat* **29**:1247-54.

7.   White, S. J., Vissers, L. E., Geurts van Kessel, A., et al. (2007). Variation of CNV distribution in five different ethnic populations. *Cytogenet Genome Res* **118**:19-30.

8.   Aten, E., White, S. J., Kalf, M. E., et al. (2008). Methods to detect CNVs in the human genome. *Cytogenet Genome Res* **123**:313-21.