Correlations between phenotype and gene region-specific episignatures in Rubinstein-Taybi syndrome and Menke-Hennekam syndrome

Yanan Tang
Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University
Xiantao Ye
Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University
Yongkun Zhan
Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University
Kaichuang Zhang
Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University
Wenjuan Qiu
Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University
WenQing Yang
Department of Pediatrics, The Second Affiliated Hospital of Fujian Traditional Chinese Medical University
Xuefan Gu
Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University
Yongguo Yu
Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University
Bing Xiao
Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University
Yu Sun (sunyu@xinhuamed.com.cn)
Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University

Research Article

Keywords: DNA methylation, CREBBP, Rubinstein-Taybi syndrome, Menke-Hennekam syndrome, human induced pluripotent stem cells

Posted Date: March 15th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2671798/v1

Abstract

Background: Rubinstein-Taybi syndrome (RSTS) and Menke-Hennekam syndrome (MKHK) are two rare Mendelian disorders presented with variable degrees of intellectual disability and different facial dysmorphism. They are caused by loss-of-function (LOF) variants or missense/inframe deletion variants in the exon 30 and 31 of the *CREBBP* gene respectively, which is involved in histone modification and chromatin remodeling. Genetic defects in numerous genes have been found to disrupt epigenomic profiles including DNA methylation (DNAm) patterns (referred as episignature) in affected individuals. To further investigate the mechanism of *CREBBP*related disorders, human induced pluripotent stem cells (hiPSCs) are applied to study the DNAm alteration.

Results: We presented RSTS and MKHK individuals with distinct clinical features. Detailed phenotype analysis showed that RSTS patients with nonsense-mediated mRNA decay evasion (NMD-evasion) variants had atypical facial dysmorphism and severer medical problems compared to the classical RSTS caused by LOF *CREBBP* variants. MKHK patients with variants in intrinsically disordered region (IDR) showed resemblant features. Further investigations elucidated these clinical conditions in methylation change. Genome-wide DNAm analysis of 9 RSTS and 8 MKHK patients and 33 controls identified two specific peripheral blood episignatures: RSTS and MKHK_IDR compared to matched normal controls. Methylation alterations in RSTS cases with NMD-evasion variants were mildly different from that of classical RSTS. MKHK subjects with variants outside the IDR did not obey the MKHK_IDR episignature. By interrogating DNAm in hiPSCs of 5 RSTS, 4 MKHK compared with 12 controls, we observed hypermethylated DNAm profiles of RSTS and MKHK in embryonic stage. Different methylation regions (DMRs) overlapping genes in hiPSCs of RSTS and MKHK play a role in embryonic development and organogenesis. Furthermore, DNAm patterns for hiPSCs of RSTS and MKHK were enriched for genes relevant for multicellular organismal homeostasis or transcriptional binding.

Conclusions:

We identified the type and locus of variants in the *CREBBP* gene as responsible for the RSTS and MKHK episignatures, consistent with phenotype analysis. DNAm profile analysis of hiPSCs revealed meaningful biological processes associated with embryonic development.

Background

Dozens of developmental delay and intellectual disability (DD/ID) disorders are caused by pathogenic variants in genes involved in epigenetic regulations(1). *CREBBP* is one of many genes that regulate the epigenome. *CREBBP* encodes the cAMP response element-binding protein (CREBBP or CBP), which is a histone transferase and acts as a transcriptional coactivator that interacts with multiple transcriptional factors and histones(2). The characteristic structure of CREBBP includes: two TAZ-type zinc finger (ZNF1 and ZNF3), a CREB-interacting kinase-inducible domain (KIX), a bromodomain (Br), a histone acetyltransferase domain (HAT), a ZZ-type zinc finger domain (ZNF2), and a nuclear receptor coactivator (NR) linked by several Intrinsically disordered protein regions (IDRs) (Fig. 1) (3, 4).

Loss-of-function (LOF) mutations in the human *CREBBP* gene cause Rubinstein-Taybi syndrome 1 (RSTS1, OMIM #180849). RSTS is a well-recognized syndrome characterized primarily by broad and often angulated thumbs and hallux, short statue, ID, and distinctive facial dysmorphism (including downslanted palpebral fissures, beaked nose, low-hanging columella, and grimacing smile)(5, 6). Missense and inframe deletion variants in the last part of exon 30 and the beginning of exon of 31 of *CREBBP* (NM_004380.3:c.5128_5614) (ZNF2, ZNF3, IDR) were detected in individuals with non-RSTS phenotypes, which led to a new clinical entity: the Menke-Hennekam syndrome 1 (MKHK1, OMIM #618332)(3, 7–11). However, although different types and locations of pathogenic variants with *CREBBP* causing two distinct ID syndromes have been clarified, the heterogenous or atypical phenotypes in RSTS and MKHK still need to be elucidated. For instance, a patient carrying NM_004380.3:c.6169C T (p.Gln2057*) variant was found atypical from the classical RSTS (broad thumbs or big toes, hirsutism) but resembled Floating-Harbor syndrome(12). Other studies also mentioned misleading initial diagnosis in RSTS patients with pathogenic variants located at the last part of *CREBBP* (13, 14). Since these variants are frameshift variants located nearby the last exon, we speculated that these atypical cases may be due to nonsense-mediated decay (NMD) evasion (15). Moreover, previous study(3) suggested that MKHK patients with mutations distal to codon 5595 show resemblances. We further asked whether individuals with variations located in the IDR could be grouped to a subtype of MKHK.

DNA methylation (DNAm) signatures (also known as “episignatures”) has been emerging as a predictive tool to identify neurodevelopmental disorders. So far, over 50 episignatures associated with more than 60 syndromes have been established(16). Previous studies have reported a distinct DNAm signature of RSTS(16, 17). Nevertheless, researchers failed to assess the episignature in all 31 MKHK individuals, but succeeded to identify a unique sub-episignature for 13 MKHK individuals with variants in IDR(16). There is compelling evidence that IDR variants in *CREBBP* may lead to a region-specific sub-signature of MKHK. For further clarification, clinical manifestations and comparisons are necessary to explore the association between specific phenotype with episignatures.

To elucidate the mechanism of neurodevelopmental disorders, human induced pluripotent stem cells (hiPSCs) emerged to be powerful tools because they represent the *in vitro* counterparts of embryonic cells, and are able to renew themselves as well as differentiate into neural lines which is unobtainable upon most occasions(18). DNA methylation plays critical roles in the reprogramming and redifferentiation of hiPSCs, and the active turnover of DNA methylation states might facilitate key lineage decisions(19).

In this study, we presented the clinical and genetic finding of RSTS and MKHK patients from China. By analyzing the phenotypes of patients from this report and literatures, we investigate the different clinical features between classical RSTS and RSTS individuals with NMD-evasion variants, likewise, MKHK individuals with IDR variants and MKHK with variants outside IDR. We further performed genome-wide methylation analysis, and reported two distinct blood-derived episignatures of RSTS and MKHK_IDR. In these two profiles, we detected atypical RSTS cases and region-specific cases in MKHK. Eventually hiPSCs were applied to explore the mechanisms by which interactions between epigenetic modifications and DNAm drive gene expression. The study workflow was illustrated in Fig. 2.

Results

Genetic spectrum of individuals with CREBBP variants

This study included 17 unrelated subjects with Pathogenic (P)/Likely pathogenic (LP) *CREBBP* variants: eight MKHK patients (five males, three females) and nine RSTS patients (four males, five females)(Fig. 1). MKHK patients carry variants located at ZNF2 (n = 2), ZNF3 (n = 2) and IDR (n = 4). Seven MKHK individuals had missense variants and one had in-frame deletion, among which c.5602C > T was recurrent. RSTS individuals harbor LOF variants across *CREBBP* gene. Two of them had the variants downstream the last 50 nt of the penultimate exon that may escape NMD, and were further referred as "RSTS_non-NMD". Four RSTS individuals carried different frameshift variants; two carried exonic deletions; one person had a missense variant in the HAT domain; and two had different nonsense mutations. All variants arose *de novo*. Detailed clinical and molecular information was listed in Table 1.

Table 1
The MKHK and RSTS patients enrolled in the study and their genetic finding

Patient ID	Age at blood draw	Sex	CREBBP variant (NM_004380.3)	Variant domain	MKHK episignature (blood)	RSTS episignature (blood)	hiPSC model
MKHK1	4y2m	M	c.5225T > A, p.Met1742Lys	ZNF2	Test	Test	Yes
MKHK2	9m	M	c.5602C > T, p.Arg1868Trp	IDR	Test	Test	Yes
MKHK2	7y	M	c.5602C > T, p.Arg1868Trp	IDR	Train	Test	Yes
MKHK3	4y	F	c.5614A > G, p.Met1872Val	IDR	Train	Test	Yes
MKHK4	3y3m	M	c.5602C > T, p.Arg1868Trp	IDR	Train	Test	Yes
MKHK5	1y6m	M	c.5357G > A, p.Arg1786His	ZNF3	Test	Test	No
MKHK6	Cord blood	F	c.5401G > A, p.Val1801Met	ZNF3	Test	Test	No
MKHK7	5y3m	F	c.5218C > T, p.His1740Tyr	ZNF2	Test	Test	No
MKHK8	10m	M	c.5595_5597del, p.Met1865_Arg1866delinsIle	IDR	Test	Test	No
RSTS1	4y	F	c.4888dupG, p.Glu1630Glyfs*30	HAT	Test	Train	Yes
RSTS2	3y2m	M	c.3752delT, p.Leu1251Argfs*25	RING	Test	Train	Yes
RSTS4	1y2m	F	chr16:3788557–3790594_del	/	Test	Train	Yes
RSTS5	7y	M	chr16:3777622–3929934_del	/	Test	Train	Yes
RSTS6	7y	F	c.201_202delTA, p.His67Glnfs*14	IDR	Test	Test	No
RSTS7	52d	M	c.5905C > T, p.Gln1969*	IDR	Test	Test	No
RSTS8	6m	M	c.5686C > T, p.Gln1896*	IDR	Test	Test	No
RSTS10	5y	F	c.4507T > C, p.Tyr1503His	HAT	Test	Test	No
RSTS11	10y	F	c.1153_1170delinsGTGT, p.Cys385Valfs*37	Br	Test	Train	No

Clinical Features Of Rsts And Mkhk Patients

To comprehensively evaluate phenotype similarities and differences, we further collected clinical information from our patients and literatures published previously(3, 7–14, 20–34). In total, 37 MKHK and 151 RSTS patients (including 115 classical RSTS and 36 patients with non-NMD variants) were enrolled (Table 2). Clinical information of individual patients was listed in Additional file 2: Table S1 and S2. The facial and digital dysmorphism of our patients were illustrated in Additional file 1: Figure S1.

Table 2
Phenotypic analysis in MKHK and RSTS cohorts

	MKHK(3,7–11,34)					RSTS(10,12–14,20–33)		Group analysis_P value					
								MKHK vs RSTS_classical		MKHK_IDR vs non-IDR		RSTS_classical vs non-NMD	
	ZZ2	ZZ3	IDR	non-IDR	total	non-NMD	classical	*P*-value	Adj. *P*-value	*P*-value	Adj. *P*-value	*P*-value	Ad va
Individuals included	6	9	22	15	37	36	115						
Age at last examination	1.5–17	2–24	0.8–57	1.5–24	0.8–57	5m-30y	N/A						
Gender (Male)	5/6	6/9	12/22	11/15	23/37	15/27	N/A						
Gender (Female)	1/6	3/9	10/22	4/15	14/37	12/27	N/A						
Growth													
Intrauterine growth retardation	1/6	2/9	11/20	3/15	14/35	8/15	20/95	0.0292	1	0.0461	1	0.0189	0.
Post growth retardation	3/6	4/7	12/21	7/13	19/34	13/17	32/90	0.0402	1	1	1	0.0017	0.
Short statue	4/6	4/7	9/20	8/13	17/33	12/16	38/101	0.1590	1	0.4813	1	0.0050	0.
Microcephaly	1/6	6/8	11/20	7/14	18/34	21/21	27/37	0.0801	1	1	1	0.0240	0.
Development and Behavior													
Intellectual disability													
Mild	3/5	1/7	7/16	4/12	11/28	3/14	31/88	0.6971	1	0.7047	1	0.4764	1
Moderate	2/5	1/7	3/16	3/12	6/28	3/14	43/88	0.0105	0.4188	1	1	0.0553	1
Severe	0/5	5/7	9/16	5/12	14/28	8/14	10/88	1.10E-05	0.0004	0.7036	1	0.0001	0.
Speech delay (first word > 1y or delayed)	4/5	6/7	9/13	10/12	19/25	12/16	24/31	0.9005	1	0.6447	1	1	1
Motor delay (first walk > 1y or delayed)	3/4	6/7	9/11	9/11	18/22	10/14	13/18	0.7320	1	1	1	1	1
Autism/autism-like behavior	2/5	6/9	6/16	8/14	14/30	4/11	5/26	0.0306	1	0.4642	1	0.4038	1
Senses													
Hypermetropi	3/4	3/5	1/10	6/9	7/19	0/7	3/34	0.0328	1	0.0198	0.9488	1	1
Strabismus	1/3	0/4	7/11	1/7	8/18	4/10	42/86	0.7345	1	0.0656	1	0.8453	1
Hearing impairment	2/4	4/7	7/15	6/11	13/26	2/10	6/27	0.0350	1	1	1	1	1
Malformations and other health problems													
Cerebral anomaly	1/4	3/6	11/18	4/10	15/28	6/9	11/34	0.0920	1	0.4328	1	0.1365	1
Epilepsy/Seizures	0/5	1/9	3/20	1/14	4/34	5/8	18/92	0.3059	1	0.6272	1	0.0198	0.
Cardiac anomaly	0/6	0/9	7/20	0/15	7/35	9/20	14/39	0.1299	1	0.0125	0.5993	0.4974	1
Genital malformation	1/4	4/6	4/13	5/10	9/23	5/16	18/45	0.9447	1	0.4173	1	0.5351	1
Scoliosis / kyphosis	1/6	2/9	8/21	3/15	11/36	9/13	13/45	0.8703	1	0.2951	1	0.0206	0.
Recurrent infections	1/3	2/4	5/13	3/7	8/20	8/9	17/40	0.8531	1	1	1	0.0319	1
Feeding problems	3/4	4/6	12/16	7/10	19/26	10/14	74/96	0.6703	1	1	1	0.8977	1
Facial and distal limb morphology													
Highly arched eyebrows	3/5	2/7	7/13	5/12	12/25	14/14	12/36	0.2488	1	0.6951	1	2.27E-05	0.

N/A: Not available.

Fisher's exact and chi-square tests with Yates correction were used for the statistical analysis. All tests for multiple testing were adjusted by using the Bonferr method.

	MKHK(3,7–11,34)					RSTS(10,12–14,20–33)		Group analysis_P value					
								MKHK vs RSTS_classical		MKHK_IDR vs non-IDR		RSTS_classical vs non-NMD	
Telecanthi (T)	4/6	1/9	13/21	4/13	18/34	0/2	0/40	1.22E-07	4.89E-06	0.1571	1	1	1
Epicanthi (E)	2/6	1/9	7/21	2/13	10/34	2/2	12/40	0.9560	1	0.4267	1	0.1057	1
Hypertelorism	3/3	1/3	9/9	4/6	13/15	5/5	2/26	4.26E-07	0.0000	0.1429	1	0.0001	0.
Palp fiss upslant (U)	3/5	2/9	16/21	5/14	21/35	3/20	0/102	1.87E-17	7.47E-16	0.0332	1	0.0039	0.
Palp fiss downslant (D)	1/5	4/9	0/21	5/14	5/35	15/20	91/102	6.68E-17	2.67E-15	0.0062	0.2960	0.1739	1
Ptosis (P)	0/2	1/4	12/17	1/6	13/23	3/4	2/26	0.0002	0.0086	0.0515	1	0.0093	0.
Blepharophimosis (B)	2/2	0/4	7/17	2/6	9/23	1/4	0/26	0.0016	0.0630	1	1	0.1333	1
Palp fiss, short	4/5	1/8	17/21	5/13	22/34	2/2	N/A	N/A	N/A	0.0248	1	N/A	N/
Long eyelashes	2/5	4/9	6/20	6/14	12/34	7/7	N/A	N/A	N/A	0.4870	1	N/A	N/
Nasal ridge depressed	2/5	1/9	16/20	4/14	20/34	0/21	0/75	1.97E-13	7.89E-12	0.0046	0.2209	1	1
Nasal ridge convex	0/5	1/9	0/20	0/14	1/34	18/21	69/75	2.56E-19	1.02E-17	1	1	0.6527	1
Short nose	1/6	1/9	19/21	2/15	21/36	0/2	N/A	N/A	N/A	4.02E-06	0.0002	N/A	N/
Short columella	1/5	0/9	17/20	1/14	18/34	1/2	N/A	N/A	N/A	9.53E-06	0.0005	N/A	N/
Anteverted nares	2/5	0/9	14/17	2/14	16/31	N/A	N/A	N/A	N/A	0.0002	0.0117	N/A	N/
Broad nasal tip	0/5	5/9	18/20	5/14	23/34	2/3	N/A	N/A	N/A	0.0020	0.0954	N/A	N/
Philtrum short (S)	1/6	3/9	0/21	4/15	4/36	0/2	1/26	0.5727	1	0.0232	1	1	1
Philtrum long (L)	2/6	2/9	18/21	4/15	22/36	2/2	1/26	4.10E-06	0.0002	0.0005	0.0257	0.0079	0.
Philtrum deep (D)	0/6	1/9	6/21	1/15	7/36	0/2	0/26	0.0476	1	0.2003	1	1	1
Everted vermilion of upper lip	0/5	1/7	8/15	1/12	9/27	N/A	N/A	N/A	N/A	0.0192	0.9204	N/A	N/
Thin vermilion of upper lip	5/5	2/7	7/16	7/12	14/28	1/1	N/A	N/A	N/A	0.7036	1	N/A	N/
High palate	2/4	1/8	8/20	3/12	11/32	10/12	41/44	5.17E-08	2.07E-06	0.4647	1	0.6245	1
Micro/retrognathia	3/6	5/9	16/22	8/15	24/37	13/18	35/49	0.5161	1	0.3003	1	0.9491	1
Ears low-set (L)	4/6	4/9	16/21	7/15	24/36	8/11	8/40	3.88E-05	0.0016	0.0895	1	0.0030	0.
Ears short (S)	1/6	0/9	4/21	1/15	5/36	0/11	0/40	0.0482	1	0.3761	1	1	1
Fibular deviation distal halluces	1/4	4/9	10/15	5/13	15/28	N/A	6/20	0.1046	1	0.2545	1	N/A	N/
Halluces Broad (B)	0/4	1/9	2/19	1/13	3/32	24/26	100/106	3.61E-22	1.44E-20	1	1	1	1
Angulated thumbs/halluces	N/A	N/A	N/A	N/A	N/A	6/15	20/62	N/A	N/A	N/A	N/A	0.5694	1
Hypertrichosis/hirsutism	1/4	3/7	0/6	4/11	4/17	6/12	30/43	0.0011	0.0451	0.2374	1	0.3524	1
Grimacing smile	N/A	N/A	N/A	N/A	N/A	9/12	23/26	N/A	N/A	N/A	N/A	0.3567	1

N/A: Not available.

Fisher's exact and chi-square tests with Yates correction were used for the statistical analysis. All tests for multiple testing were adjusted by using the Bonferr method.

Table 3
The differentially methylated Regions (hg19) in hiPSC of RSTS and MKHK samples compared with normal controls

RSTS vs control DMRs (hiPSC)									
Chrom	Start	End	Size (bp)	geneSymbol	Probe count	FDR	Methylation_RSTS	Methylation_Control	delta beta (Δβ)
19	52391367	52391605	239	*ZNF577*	4	0.0309	0.8462	0.3028	0.5434
MKHK vs Control DMRs (hiPSC)									
Chrom	Start	End	Size (bp)	geneSymbol	Probe count	FDR	Methylation_MKHK	Methylation_Control	delta beta (Δβ)
1	2983926	2987645	3720	*PRDM16, PRDM16-DT*	23	1.84E-22	0.3381	0.0532	0.2849
4	5709858	5710767	910	*EVC2*	11	3.85E-06	0.4260	0.0904	0.3356
1	2979582	2980163	582	*PRDM16-DT*	4	0.0001	0.4133	0.0582	0.3552
15	26107382	26109614	2233	*ATP10A*	32	0.0026	0.5556	0.3503	0.2053

Both classical RSTS and MKHK patients present with varying degrees of short stature, ID, several shared nonspecific malformations (such as microcephaly, small jaw, bone deformity) and medical problems (such as intrauterine hypoplasia, audio-visual impairment, epilepsy, cerebral anomaly, cardiac and genital malformation, feeding problems, recurrent infection). Typical dysmorphisms of classical RSTS were downslanted palpebral fissures (91/102), grimacing smile (23/26), convex nasal ridge (69/75), hirsutism (30/43) and broad thumbs/halluces (100/106); facial features of MKHK included upslanted palpebral fissures (21/35), telecathi (18/34), hypertelorism (13/15), ptosis (13/23), depressed nasal bridge (20/34), and long philtrum (22/36), low-set ears (24/36) (adj. *p* value < 0.05), which was consistent with previous studies (3, 7). In terms of ID, the analysis did not discriminate two cohorts in mild and moderate ID, while MKHK cohort has a higher rate of severe ID (adj. *p* value = 0.0004). These results indicate again that MKHK and RSTS are two distinct rare disorders.

Based on previous DNAm results and clinical descriptions(3, 16), we presumed that variations located at IDR would lead to a unique sub-phenotype together with a region-specific sub-signature, and conjectured that individuals with IDR mutations would suffer more severe and focused clinical manifestations. Consequently, we used extensively and retrospectively collected data for two groups, non-IDR group (n = 15) and IDR group (n = 22). Folks in both groups display certain MKHK-like facial dysmorphism. However, we observed that IDR group chiefly presented with upslanted palpebral fissures (16/21), hypertelorism (9/9), short palp fissure (17/21), depressed nasal ridge (16/20), broad nasal tip (18/20), everted vermilion of upper lip (8/15); short nose (19/21), short columella (17/20), anteverted nares (14/17), long philtrum (18/21) (adj. *p* value < 0.05). Conversely, the incidence of downslanted palp fissure is higher in non-IDR group (5/14). On medical problems, both groups displayed similar frequency of ID, speech and motor delay, and autism disorder, while there was also evidence for higher prevalence of cardiac anomaly and intrauterine dysplasia in IDR group, but lacked statistical significance. Overall, there was subtle association between IDR variation of *CREBBP* gene with unique sub-phenotype of MKHK.

Next, we attempted to distinguish RSTS_non-NMD individuals (n = 36) from classical RSTS individuals (n = 115). There are subtle differences of facial dysmorphism: RSTS_non-NMD folks may display high arched eyebrows (14/14), hypertelorism (5/5) (adj. *p* value < 0.05); upslanted palp fiss (3/20), ptosis (3/4), long philtrum (2/2), low-set ears (8/11), which is in concordance with the fact that some RSTS_non-NMD patients were initially diagnosed with other types of ID reported previously (13, 14). Notably, RSTS_non-NMD folks have a higher rate of severe ID (8/14) (adj. *p* value < 0.05); and post growth retardation (13/17), recurrent infections (8/9). Apart from above distinctions, they also have consistent features, including classical RSTS dysmorphism (grimacing smile, broad and angulated thumbs/halluces), developing delay, feeding problems, visual and hearing impairment. Therefore, we observed that RSTS_non-NMD patients may be atypical from classical RSTS patients on account of several minor differences. Since the accessible information of RSTS_non-NMD folks was insufficient, more evidence needs to be discovered and gathered to support this hypothesis.

RSTS/MKHK DNAm signatures associated with CREBBP **in blood**

DNA methylation analysis was performed on cases with confirmed clinical and molecular diagnoses of MKHK (n = 8) or RSTS (n = 9). A comparison was conducted between these affected subjects with 33 age- and sex- matched healthy controls.

We first compared RSTS subjects with controls and identified 379 significant differentially methylated probes (306 hypomethylated and 73 hypermethylated) (Fig. 3a). All the differentially methylated probes (DMPs) identified in this analysis are listed in Additional file 2: Table S3. Heatmap of methylation level of above sites for each individual was created, and all nine RSTS samples clustered together. Methylation profile of the two non-NMD RSTS is distant from the classical RSTS in the clustering tree (Fig. 3b). This is in accord with the atypical characteristics.

Then we investigated whether MKHK cohort can be distinguished from healthy controls by setting up a MKHK episignature. Comparison of MKHK with controls identified 137 differentiated methylated probes (98 hypomethylated and 39 hypermethylated) (Fig. 3c). Detailed information of these sites was listed in Additional file 2: Table S4. Heatmap showed that six subjects (five peripheral blood samples with IDR variants, one umbilical cord blood sample with ZNF3 variant), clustered together on the right and were clearly separated from controls, while two patients with ZNF2 variants and one with ZNF3 variant clustered on the left (Fig. 3d).

Development Of Classification Models For Rsts/mkhk_idr

As shown above, the presence of RSTS and MKHK DNAm patterns suggested that we might establish two classification models for each cohort to distinguish each affected patient. We trained two support vector machine (SVM) models on the DNAm data from RSTS and MKHK episignatures. Methylation variant pathogenicity (MVP) scores generated from RSTS and MKHK_IDR SVM models were listed in Additional file 2: Table S5 and S6.

In the RSTS SVM model, all RSTS tests were classified positive, and MKHK cohort as well as additional controls classified negative, demonstrating 100% sensitivity and specificity. RSTS patients with NMD-evasion variants were classified as RSTS with a score of 0.85 and 0.82, which are slightly lower than others. In addition, the cord blood sample (MKHK6) with a slightly high score of about 0.42 was noticed in RSTS SVM model (Fig. 4a).

As MKHK subjects with IDR variants give rise to a region-specific DNAm pattern while subjects with ZNF2 and ZNF3 variants were slightly deviated from the former(16), we selected IDR subjects (referred as “MKHK_IDR”) for model training. In MKHK_IDR SVM model, except for one IDR subject at nine-month-old was classified negative (scored 0.42), four IDR subjects were classified positive. The ZNF2 and ZNF3 subjects were appropriately classified negative, with the score of one umbilical cord blood subject carrying ZNF3 variant (scored 0.46) is higher than those of other ZNF2 and ZNF3 subjects (scored around 0.23). Therefore, the classification was consistent with the heterogeneity of MKHK. We also speculated that MKHK subjects with IDR variants may lead to a region-specific DNAm signature. In addition, all the controls in our study and from public datasets were classified negative. Using MKHK_IDR SVM model, all the RSTS subjects were classified negative. Two non-NMD RSTS samples were scored higher (around 0.46) than others (around 0.13) (Fig. 4b).

Dnam Changes In Hipscs Of Rsts And Mkhk

To investigate the mechanism of the two *CREBBP* related disorders, we induced hiPSCs from four MKHK individuals, five RSTS individuals and twelve healthy donors. All the hiPSCs originated from peripheral blood lymphocytes, reprogramed using the same protocol. We first compared three passages (P4, P6, P8) of undifferentiated hiPSCs from one sample to analyze DNAm concordance of different cell passages. The correlation analysis of gene-linked CpG methylation revealed at least 98% concordance (Additional file 1: Figure S2). The inter-passage variance was small, then we used one passage of each subject in the following analysis.

We identified 105 significant differentially methylated probes (75 hypermethylated and 30 hypomethylated) in RSTS group (Fig. 5a). The DMR in hiPSC of RSTS covers chr19: 52391367–52391605 overlapping the intron part of *ZNF577*. The *ZNF577* gene, coding for a zinc-finger protein, is involved in transcriptional regulation. In the hiPSC of MKHK group 66 differentially methylated probes (50 hypermethylated and 16 hypomethylated) were identified (Fig. 5b). Four DMRs discovered in hiPSC of MKHK are: chr1:2983926–2987645 (covering the 5’utr and exon 1 of *PRDM16*, and *PRDM16-DT*, a long non-coding RNA), chr4:5709858–5710767 (covering the 5’utr and exon 1 of *EVC2*), chr1:2979582–2980163 (covering the exon 3 of *PRDM16-DT*), and chr15: 26107382–26109614 (covering the 5’utr and exon 1 of *ATP10A*). The *PRDM16* gene, encoding PR/SET domain 16, functions as a transcriptional regulator and displays histone methyltransferase activity. During mouse embryonic development, *PRDM16* was first detected on E9.5, and expressed in a broad range of developing tissues (brain, lung, kidney, etc) on E14.5 (35). Previous study also uncovered that *PRDM16* mediated trans-differentiation of hiPSCs towards brown adipocytes(36). The *EVC2* gene, encoding EvC ciliary complex subunit 2, plays a critical role in bone formation and skeletal development, and is possibly involved in early embryonic morphogenesis(37, 38). The *ATP10A* gene, encoding ATPase phospholipid transporting 10A, showed expression in numerous tissues with high level in the brain(39, 40). Meanwhile, the expression of *ATP10A* was relatively low in undifferentiated human embryonic stem cell (hESC) lines whereas to be up-regulated in differentiated tissues(41). Therefore, methylation alterations on these genes may impact the embryonic development and organogenesis.

Eventually, we performed gene-set enrichment analysis on DMPs identified in each group. There were 71 genes overlapping the RSTS DNAm patterns and 18 genes for the MKHK DNAm patterns. 50 and 9 Gene ontology (GO) terms were identified in RSTS and MKHK hiPSCs, respectively (Additional file 2: Table S9 and S10). For RSTS hiPSCs, “multicellular organismal homeostasis” is the most highly enriched term in biological process, of which the key genes (*BBS1*, *ITPKB*, *NOS3*, *CFLAR*, etc) are closely related to cell signaling (Fig. 5c). For MKHK hiPSCs, “DNA-binding transcription factor binding” and “transcription factor binding” are the only terms in molecular functions, suggesting MKHK hiPSCs might have impaired capacity in transcription (Fig. 5d).

Discussion

In this study, we summarized the phenotypes of patients carrying *CREBBP* variants and investigated the RSTS and MKHK DNAm signature in blood and hiPSCs. The RSTS and the MKHK group displayed different methylation profiles and can be distinguished from each other. It was also found that methylation alterations of two RSTS subjects with NMD-evasion variants were mildly different from that of other RSTS subjects, and MKHK subjects with IDR variants also showed unique methylation profiles from those with variants in the ZNF2 and ZNF3 region. Meanwhile, we applied hiPSCs and discovered disease-associated DNA methylation alterations in hiPSCs. Importantly, the DMPs were annotated to, or close to genes enriched in nervous system development. Furthermore, DMRs related genes in hiPSCs of RSTS and MKHK play a role in embryonic development and organogenesis. These results indicate that variants in different regions of *CREBBP* gene may result in diverse methylation alterations, that might further hamper protein functions, and cause clinical phenotypes ultimately.

The nine RSTS cases displayed clustered DNA methylation profile in comparison with controls and the SVM model derived from RSTS episignature positively classified all the RSTS affected subjects, consistent with previous work(16, 17). Interestingly, though the two RSTS that might escape NMD (RSTS7 and RSTS8) are not predicted as MKHK in the MKHK_IDR model, they show relatively high MVP score (0.47 and 0.48). In the t-SNE analysis of MKHK episignature, they are mapped closer with MKHK patients than with other RSTS cases (Additional file 1: Figure S3). Similar to other reported RSTS_non-NMD individuals, we observed atypical RSTS phenotype of these two patients. RSTS7 showed atypical RSTS facial features (Additional file 1: Figure S1) and both RSTS7 and RSTS8 had severe medical problems compared to classical RSTS. They also manifested facial dysmorphisms overlapping with MKHK including high arched

eyebrows, hypertelorism, upslanted palp fissure, and ptosis, that might explain why some patients were misdiagnosed as Cornelia de Lange syndrome and Floating-Harbor syndrome(12, 14). This might be due to the frameshift or nonsense variants that would escape NMD and cause effects other than merely loss of CREBBP function(15).

MKHK variants located at the end of exon 30 and the beginning of exon 31 of the *CREBBP* gene, distributed in three domains, more frequently located in the IDR and less in ZNF2 and ZNF3 region. Heatmap showed that four patients with IDR variants clustered together and clearly separated themselves from control subjects, with two patients with ZNF2 variants and one patient with ZNF3 variant clustering in the left (Fig. 3b). In MKHK_IDR SVM model, MKHK_IDR subjects (except one with younger age) were classified positive correctly (Fig. 4b), MKHK_ZNF2/ZNF3 subjects were classified negative, indicating region-specific episignatures for the condition tested. For four patients with variants in the IDR, they have specific clinical manifestations separated from those with variants in ZNF2 and ZNF3 domains, consistent with other individuals reported in the literature(16). Statistical analysis suggests specific recognizable performances in patients with IDR variants: short nose, short columella, anteverted nares, long philtrum, etc. Although the sample size is small, it still suggested MKHK with IDR variants can give rise to a unique condition, while MKHK with ZNF2 and ZNF3 variants might be associated with another distinct episignature and phenotype. By face scan, Menke *et al.* found that a IDR MKHK patient's face resembled 16p13.3 duplication patients and inverted "mean RSTS face". They suggested IDR MKHK might be caused by gain of function effect. While by 3D modeling of the ZNF2 domain, a molecular dynamics analysis(9) demonstrated Glu1724Lys might alter protein–protein interactions due to the difference in electrostatic potential. Thus, variants in IDR, ZNF2 and ZNF3 might determine the phenotype via different molecular biological mechanisms, and cause the heterogenous phenotype and pathogenesis of MKHK.

In MKHK_IDR SVM model, MKHK2 at 9-month-old was wrongly classified negative, even if the same patient at seven-year-old was classified positive. Meanwhile, MKHK6 with ZNF3 variant from umbilical cord blood clustered with the MKHK_IDR variants in the heatmap although it was classified negative in MKHK_IDR SVM model. This may due to the limit number of training samples in SVM model generation.

hiPSCs displayed hyper-methylated alterations in MKHK and RSTS. Clusters of correlated CpGs sites, discovered several biologically meaningful DNAm alterations in RSTS and MKHK hiPSCs, possibly affecting embryonic development and subsequent differentiation. The enrichment analysis also uncovered multiple differentially methylated genes within transcriptional factor binding or multicellular organismal homeostasis, similarities and differences in MKHK and RSTS may partly be related to methylation differences of these genes. Further study on hiPSC-derived neurons might help to clarify the pathogenic mechanism of the two disorders.

Conclusions

Overall, we established two distinct blood-derived episignatures in a single gene, which are beneficial to study genotype-epigenotype-phenotype correlation of diseases caused by *CREBBP* variants in diverse locations. DNAm profiles in hiPSCs also showed some guiding significance for studying mechanism of RSTS and MKHK.

Methods

Patient enrollment

Seventeen patients with pathogenic/likely pathogenic *CREBBP* variants were enrolled in the study. Exome sequencing was performed to draw molecular diagnosis of them. The clinical information was collected and followed up by experienced pediatricians. All patients provided informed consent for genetic studies. hiPSC induction was performed under the written informed consent from four MKHK and five RSTS patients. The Institutional Review Board of Xinhua Hospital, Shanghai Jiaotong University School of Medicine approved the study protocol (Approval no. XHEC-D-2023-021).

Hipsc Induction

Peripheral blood mononuclear cells (PBMCs) were collected via gradient centrifugation from blood of five RSTS and four MKHK patients. PBMCs were infected for 7–10 days with EBV in transformation medium (RPMI-1640 (10-041-CV, Corning Cellgro), 20% FBS (10099141C, Gibco), 1% Penicillin-Streptomycin solution (15140122, Gibco), 2 μg/mL Cyclosporin A (MB1068, Meilunbio) in a 37°C humidified incubator with 5% CO2. Then the lymphocytes were electroporated with Yamanaka's factors plasmid (OCT4, SOX2, KLF4, and MYC) and cultured on Mouse Embryonic Fibroblasts (MEF) feeder layers in mTeSR plus medium (100–0276, STEMCELL). Approximately 30 days later, hiPSC colonies were hand-picked and expanded on Matrigel-coated plates. The hiPSC lines expressed stem cell markers for pluripotency, and could be *in vivo* (teratoma) differentiated into three germ layers(44, 45).

Dna Methylation Array And Quality Control Of The Experiment

DNA was isolated from peripheral blood of seven MKHK patients and nine RSTS patients, umbilical blood of one MKHK individual, and human lymphocyte-induced pluripotent stem cells (hiPSCs) from four MKHK patients and five RSTS patients by using DNeasy Blood and Tissue Kit (69506, Qiagen, Germany). The purity and concentration of DNA was estimated using Nanodrop one (840-317400, Thermo Fisher, USA) / Quibt3.0 (Q33216, Thermo Fisher, USA). 500 ng DNA of each sample was used to bisulfite converted using EZ DNA Methylation Kits (D5002, Zymo Research, USA), then converted products were put into the Illumina Infinium methylation EPIC BeadChips (850k) following the manufacturer's protocol (Illumina). The EPIC array data was imported into R 3.5.2 for analysis. The minfi(46) and ChAMP(47) Bioconductor package were used to preprocess data including quality control, normalization and background subtraction, followed by extraction of β values. β values represent DNA methylation level / the proportion of methylated each CpG site, ranging between 0 (no methylation) and 1 (full methylation). Then we filtered probes with detection p-value ≥ 0.01, probes with < 3 beads in at least 5% of samples, non-CpG probes,

multi-hit probes, probes located in chromosome X and Y and (SNP-related probes), 727553 probes from blood sample and 733939 probes from hiPSC sample remained for analysis. The β value matrix was normalized using BMIQ(48) for adjusting type I and type II probe bias. Next, we used Singular Value Decomposition Analysis (SVD)(49) to analysis the batch effect caused by BeadChip Slide and Array, then applied Combat(49) to correct batch effect.

Dna Methylation Signature Identification

To assess DNAm patterns, we identified differentially methylated sites in two aspects: one DNAm distribution for MKHK cases (n = 9) and one for the RSTS cases (n = 9) versus age- and sex-matched control subjects (n = 33). We could not generate a signature for MKHK cases with variants in each domain (ZNF2, ZNF3, IDR) due to small sample size, so we opted to classify the whole MKHK cohort and observe inner differences instead. Differential Methylated CpGs Position (probes) were identified by limma package(50) and *p* values were adjusted for multiple testing using the Benjamini-Hochberg method(51). Optimal adj. *p* value and average DNAm difference (Δβ) thresholds were selected using volcano plots. For episignature establishment, probes with adj. *p* value < 0.05 and a methylation difference greater than 10% ($|\Delta\beta| >= 0.1$) were considered significant. Heatmap demonstrated that the selected probes separated the patients from controls. Using DMRcate(52), we further prioritized DMRs on blood and hiPSC methylation data with a false discovery rate (FDR) of < 0.05.

Construction And Validation Of Classification Models

Genome-wide DNA methylation data from above samples with various age and ethnicity were used for mapping of episignatures and model training. For every iteration aiming at the identification of the methylation profiles or feature selection, the ratio of training versus testing cohort is diverse in each comparison for the assessment of the performance of the classification models. We developed two machine learning models for MKHK and RSTS, respectively, one using each blood-derived DNAm signature (adj. *p* value < 0.05, $|\Delta\beta| >= 0.1$, RSTS episignature, n = 379; MKHK episignature, n = 137). Using the R package caret, CpG sites with correlations equal to or greater than 90% to other signature CpGs were removed by findCorrelations function of caret(53). This led to a set of n = 311 from the RSTS signature and n = 113 non-redundant CpGs from MKHK signature. The t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis(54) was used to map high-dimensional data into a two-dimensional space to visualize and explore the selected methylation data. To test the specificity and sensitivity of RSTS and MKHK episignature, we developed two support vector machine (SVM) models with linear kernel trained on differentially methylated CpG sites. Training was done using the e1071 R package(55).

Each model was trained using the methylation values for discovery cases compared to discovery controls, i.e., for RSTS SVM model, RSTS cases (n = 5, classical) versus controls (n = 26); for MKHK SVM model, MKHK cases (n = 3, IDR variants) versus controls (n = 26)(Table 1). Then we performed validation on each model with the remaining subjects (RSTS n = 4, MKHK n = 6, control n = 7). Both RSTS (n = 9) and MKHK (n = 9) samples were used to validate the other model. We also obtained 17 controls downloaded from GEO database for classification (GSE111165). Methylation variant pathogenicity (MVP) score was generated from SVM model, ranging between 0 and 1 (0–100%), classifying samples as “positive” (score > 0.5) or “negative” (score < 0.5). A negative classification typically indicates a benign result for the condition tested or another signature distinct from this model. As DNAm can be tissue and cell-type specific, we tested undifferentiated hiPSC-derived DNA from five RSTS and four MKHK patients with *CREBBP* variants in comparison to twelve hiPSC controls as well.

Genotype-phenotype Correlation Analysis

Due to limitations of small sample size in this report, we gathered cases in this study and those in previous literatures or databases to expand the samples' quantity. As we hypothesized that region-specific phenotype maybe significant, MKHK cohort were divided into IDR and non-IDR groups (including ZNF2 and ZNF3 variants), and RSTS cohort were divided into non-NMD and classical groups for further analysis. RSTS_non-NMD group carries NMD-escape variants located downstream the last 50 nt of the penultimate exon of *CREBBP*. RSTS_classical group carries variants outside of the NMD-sensitive region. Clinical data are derived from: 1) this report ; 2) previously published literature (MKHK(3, 7–11, 34), RSTS_classical(20–23), RSTS_non-NMD(10, 12–14, 20, 22, 24–33)); 3) ClinVar and HGMD. Inclusion criteria are: i) P/LP variants found in *CREBBP* gene; and ii) clinial data is available. A total of 37 MKHK patients, 36 non-NMD RSTS patients and 115 classical RSTS patients were enrolled in this research. Comparisons in different groups (MKHK vs RSTS_classical, MKHK_IDR vs non-IDR, RSTS_classical vs non-NMD) were analyzed using chi-square tests (with Yates correlation) or Fisher’s exact test. Both unadjusted and Bonferroni-corrected *p* values were calculated. Adjust. *p* value < 0.05 was considered statistically significant.

Abbreviations

DD/ID
Developmental delay and intellectual disability
DMR
Differentially methylated regions
DNAm
DNA methylation
Episignature
DNA methylation signature
FDR
False discovery rate
GOF

Gain of function
hiPSCs
human induced pluripotent stem cells
hESC
human embryonic stem cell
IDRs
Intrinsically disordered protein regions
LOF
Loss of function
LP
likely pathogenic
MEF
Mouse Embryonic Fibroblasts
MKHK
Menke-Hennekam syndrome
MVP
Methylation variant pathogenicity
NMD
Nonsense-mediated mRNA decay
P
Pathogenic
PBMCs
Peripheral blood mononuclear cells
PD
Parkinson's disease
PTCs
Premature termination codons
RSTS
Rubinstein-Taybi syndrome
SVD
Singular value decomposition analysis
SVM
Support vector machine
t-SNE
t-Distributed Stochastic Neighbor Embedding
VUSs
Variants of uncertain significance
ZNF
Zinc finger domain.

Declarations

Ethics approval and consent to participate

This study has been approved by the ethic review commitment of Xinhua Hospital, Shanghai Jiaotong University School of Medicine approved the study protocol (Approval no. XHEC-D-2023-021). All of the samples and records were de-identified before experiments and analysis. Consent was obtained from the parents of all patients for participation of the study.

Consent for publication

The legal guardians of all patients have provided consent for the publication of the patients' data.

Availability of data and materials

The datasets analyzed in this study are available from the corresponding authors on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The work is funded by Natural Science Foundation of Shanghai (20ZR1472700 and 22ZR1451400), National Natural Science Foundation of China (81873724), and the National Key R&D Program of China (2022YFC2703400), and Shanghai Municipal Nature Science Foundation (19ZR1442100 to BX).

Authors' contributions

Tang YN performed data analysis and summarized the clinical information of patients. Ye XT and Zhan YK contributed to hiPSC reprogramming, induction and culture. Ye XT and Tang YN created images and tables. Zhang KC, Yang WQ, Qiu WJ, Xiao B, Gu XF, Yu YG recruited the patients, contributed to clinical assessment and interpretation. Sun Y analyzed the sequencing data. Yu YG, Xiao B and Sun Y conceived the project, supervised all aspects of the study. Tang YN, Xiao B and Sun Y wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the patients and their family members for their participation in the study. We would like to thank the staff, molecular geneticists, and other clinical geneticists for the evaluation and diagnosis of the patients with neurodevelopmental conditions presented in this study. The Chip experiments were performed by Sinotech Genomics (Shanghai, China).

References

1. Sadikovic B, Aref-Eshghi E, Levy MA, Rodenhiser D. DNA methylation signatures in mendelian developmental disorders as a diagnostic bridge between genotype and phenotype. Epigenomics. 2019;11(5):563–75.
2. Dutto I, Scalera C, Prosperi E. CREBBP and p300 lysine acetyl transferases in the DNA damage response. Cell Mol Life Sci [Internet]. 2018;75(8):1325–38. Available from: https://doi.org/10.1007/s00018-017-2717-4
3. Menke LA, Gardeitchik T, Hammond P, Heimdal KR, Houge G, Hufnagel SB, et al. Further delineation of an entity caused by CREBBP and EP300 mutations but not resembling Rubinstein–Taybi syndrome. Am J Med Genet Part A. 2018;176(4):862–76.
4. Dyson HJ, Wright PE. Role of intrinsic protein disorder in the function and interactions of the transcriptional coactivators CREB-binding Protein (CBP) and p300. J Biol Chem. 2016;291(13):6714–22.
5. Allanson JE, Roberts AE. Rubinstein-Taybi Syndrome Summary Suggestive Findings. GeneReviews®. 2019;1–33.
6. Hennekam RCM. Rubinstein-Taybi syndrome. Eur J Hum Genet. 2006;14(9):981–5.
7. Menke LA, van Belzen MJ, Alders M, Cristofoli F, Ehmke N, Fergelot P, et al. CREBBP mutations in individuals without Rubinstein–Taybi syndrome phenotype. Am J Med Genet Part A. 2016;170(10):2681–93.
8. Banka S, Sayer R, Breen C, Barton S, Pavaine J, Sheppard SE, et al. Genotype–phenotype specificity in Menke–Hennekam syndrome caused by missense variants in exon 30 or 31 of CREBBP. Am J Med Genet Part A. 2019;179(6):1058–62.
9. Angius A, Uva P, Oppo M, Persico I, Onano S, Olla S, et al. Confirmation of a new phenotype in an individual with a variant in the last part of exon 30 of CREBBP. Am J Med Genet Part A. 2019;179(4):634–8.
10. Nishi E, Takenouchi T, Miya F, Uehara T, Yanagi K, Hasegawa Y, et al. The novel and recurrent variants in exon 31 of CREBBP in Japanese patients with Menke–Hennekam syndrome. Am J Med Genet Part A. 2022;188(2):446–53.
11. Sima A, Smădeanu RE, Simionescu AA, Nedelea F, Vlad A-M, Becheanu C. Menke–Hennekam Syndrome: A Literature Review and a New Case Report. Children. 2022;9(5):759.
12. Elalaoui SC, Smaili W, Van-Gils J, Fergelot P, Ratbi I, Tajir M, et al. Clinical description and mutational profile of a moroccan series of patients with rubinstein taybi syndrome. Afr Health Sci. 2021;21(2):960–7.
13. Cross E, Duncan-Flavell PJ, Howarth RJ, Hobbs JI, Thomas NS, Bunyan DJ. Screening of a large Rubinstein–Taybi cohort identified many novel variants and emphasizes the importance of the CREBBP histone acetyltransferase domain. Am J Med Genet Part A. 2020;182(11):2508–20.
14. Stojanovic JR, Miletic A, Peterlin B, Maver A, Mijovic M, Borlja N, et al. Diagnostic and Clinical Utility of Clinical Exome Sequencing in Children With Moderate and Severe Global Developmental Delay / Intellectual Disability. J Child Neurol. 2020;35(2):116–31.
15. Supek F, Lehner B, Lindeboom RGH. To NMD or Not To NMD: Nonsense-Mediated mRNA Decay in Cancer and Other Genetic Diseases. Trends Genet [Internet]. 2021;37(7):657–68. Available from: https://doi.org/10.1016/j.tig.2020.11.002
16. Levy MA, McConkey H, Kerkhof J, Barat-Houari M, Bargiacchi S, Biamino E, et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. Hum Genet Genomics Adv. 2022;3(1):100075.
17. Kerkhof J, Squeo GM, McConkey H, Levy MA, Piemontese MR, Castori M, et al. DNA methylation episignature testing improves molecular diagnosis of Mendelian chromatinopathies. Genet Med [Internet]. 2022;24(1):51–60. Available from: https://doi.org/10.1016/j.gim.2021.08.007
18. Bar S, Benvenisty N. Epigenetic aberrations in human pluripotent stem cells. EMBO J. 2019;38(12):1–18.
19. Parry A, Rulands S, Reik W. Active turnover of DNA methylation during cell fate decisions. Nat Rev Genet [Internet]. 2021;22(1):59–66. Available from: http://dx.doi.org/10.1038/s41576-020-00287-8
20. Spena S, Milani D, Rusconi D, Negri G, Colapietro P, Elcioglu N, et al. Insights into genotype-phenotype correlations from CREBBP point mutation screening in a cohort of 46 Rubinstein-Taybi syndrome patients. Clin Genet. 2015;88(5):431–40.
21. Rusconi D, Negri G, Colapietro P, Picinelli C, Milani D, Spena S, et al. Characterization of 14 novel deletions underlying Rubinstein–Taybi syndrome: an update of the CREBBP deletion repertoire. Hum Genet [Internet]. 2015;134(6):613–26. Available from: http://dx.doi.org/10.1007/s00439-015-1542-9

22. Wincent J, Luthman A, Van Belzen M, Van Der Lans C, Albert J, Nordgren A, et al. CREBBP and EP300 mutational spectrum and clinical presentations in a cohort of swedish patients with rubinstein–taybi syndrome. Mol Genet Genomic Med. 2016;4(1):39–45.
23. Schorry EK, Keddache M, Lanphear N, Rubinstein JH, Srodulski S, Fletcher D, et al. Genotype-phenotype correlations in Rubinstein-Taybi syndrome. Am J Med Genet Part A. 2008;146(19):2512–9.
24. Bentivegna A, Milani D, Gervasini C, Castronovo P, Mottadelli F, Manzini S, et al. Rubinstein-Taybi Syndrome: Spectrum of CREBBP mutations in Italian patients. BMC Med Genet. 2006;7:1–13.
25. Lee JS, Byun CK, Kim H, Lim BC, Hwang H, Choi JE, et al. Clinical and mutational spectrum in Korean patients with Rubinstein-Taybi syndrome: The spectrum of brain MRI abnormalities. Brain Dev [Internet]. 2015;37(4):402–8. Available from: http://dx.doi.org/10.1016/j.braindev.2014.07.007
26. Bartsch O, Schmidt S, Richter M, Morlot S, Seemanová E, Wiebe G, et al. DNA sequencing of CREBBP demonstrates mutations in 56% of patients with Rubinstein-Taybi syndrome (RSTS) and in another patient with incomplete RSTS. Hum Genet. 2005;117(5):485–93.
27. Enomoto Y, Yokoi T, Tsurusaki Y, Murakami H, Tominaga M, Minatogawa M, et al. Divergent variant patterns among 19 patients with Rubinstein-Taybi syndrome uncovered by comprehensive genetic analysis including whole genome sequencing. Clin Genet. 2022;101(3):335–45.
28. Bradford L, Ross MK, Minso J, Cernelc-Kohan M, Shayan K, Wong SS, et al. Interstitial lung disease in children with Rubinstein-Taybi syndrome. Pediatr Pulmonol. 2022;57(1):264–72.
29. Roelfsema JH, White SJ, Ariyürek Y, Bartholdi D, Niedrist D, Papadia F, 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):572–80.
30. Chiang PW, Lee NC, Chien N, Hwu WL, Spector E, Tsai ACH. Somatic and germ-line mosaicism in Rubinstein-Taybi syndrome. Am J Med Genet Part A. 2009;149(7):1463–7.
31. Pérez-Grijalba V, García-Oguiza A, López M, Armstrong J, García-Miñaur S, Mesa-Latorre JM, et al. New insights into genetic variant spectrum and genotype–phenotype correlations of Rubinstein-Taybi syndrome in 39 CREBBP-positive patients. Mol Genet Genomic Med. 2019;7(11):1–13.
32. I Coupry, C Roudaut, M Stef, M-A Delrue, M Marche, I Burgelin, L Taine, C Cruaud, D Lacombe BA. Molecular analysis of the CBP gene in 60 patients with Rubinstein-Taybi syndrome. J Med Genet. 2008;31(2):253–7.
33. Yu S, Wu B, Qian Y, Zhang P, Lu Y, Dong X, et al. Clinical exome sequencing identifies novel CREBBP variants in 18 Chinese Rubinstein–Taybi Syndrome kids with high frequency of polydactyly. Mol Genet Genomic Med. 2019;7(12):1–15.
34. Dyment DA, O'Donnell-Luria A, Agrawal PB, Coban Akdemir Z, Aleck KA, Antaki D, et al. Alternative genomic diagnoses for individuals with a clinical diagnosis of Dubowitz syndrome. Am J Med Genet Part A. 2021;185(1):119–33.
35. Horn KH, Warner DR, Pisano M, Greene RM. PRDM16 expression in the developing mouse embryo. Acta Histochem. 2011;113(2):150–5.
36. Ciano P Di, Grandy D, Foll B Le, Addiction T, Health M, Health O, et al. White-to-brown metabolic conversion of human adipocytes by JAK inhibition. Nat Cell Biol. 2015;17(1):1–18.
37. Badri MK, Zhang H, Ohyama Y, Venkitapathi S, Alamoudi A, Kamiya N et al. Expression of Evc2 in craniofacial tissues and craniofacial bone defects in Evc2 knockout mouse. Arch Oral Biol. 2016;176(4):139–48.
38. Kwon EK, Louie K, Kulkarni A, Yatabe M, Ruellas AC de O, Snider TN, et al. The Role of Ellis-Van Creveld 2(EVC2) in Mice During Cranial Bone Development. Anat Rec. 2018;301(1):46–55.
39. Kayashima T, Yamasaki K, Joh K, Yamada T, Ohta T, Yoshiura KI, et al. Atp10a, the mouse ortholog of the human imprinted ATP10A gene, escapes genomic imprinting. Genomics. 2003;81(6):644–7.
40. Kashiwagi A, Meguro M, Hoshiya H, Haruta M, Ishino F, Shibahara T, et al. Predominant maternal expression of the mouse Atp10c in hippocampus and olfactory bulb. J Hum Genet. 2003;48(4):194–8.
41. Bifari F. Epigenetic states and expression of imprinted genes in human embryonic stem cells. World J Stem Cells. 2010;2(4):97.
42. Nishizawa M, Chonabayashi K, Nomura M, Tanaka A, Nakamura M, Inagaki A, et al. Epigenetic Variation between Human Induced Pluripotent Stem Cell Lines Is an Indicator of Differentiation Capacity. Cell Stem Cell [Internet]. 2016;19(3):341–54. Available from: http://dx.doi.org/10.1016/j.stem.2016.06.019
43. Nishino K, Umezawa A. DNA methylation dynamics in human induced pluripotent stem cells. Hum Cell. 2016;29(3):97–100.
44. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. Science (80-). 2007;318(5858):1917–20.
45. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell. 2007;131(5):861–72.
46. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363–9.
47. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: Updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics. 2017;33(24):3982–4.
48. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189–96.
49. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–27.
50. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

51. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med. 1990;9(7):811–8.
52. Peters T, Buckley M, Clark S, Molloy P. The DMRcate package user's guide. 2015;2(Figure 1):1–7. Available from: https://bioconductor.org/packages/DMRcate/
53. Qian SS. Classification and Regression Tree. Environmental and Ecological Statistics with R. 2021. 237–268 p.
54. Donaldson J. T-Distributed Stochastic Neighbor Embedding for R (t-SNE). 2022;2–5. Available from: https://github.com/jdonaldson/rtsne/
55. Hornik K, Weingessel A, Leisch F, Davidmeyerr-projectorg MDM. Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien [Internet]. 2021. Available from: https://cran.r-project.org/package=e1071

Figures

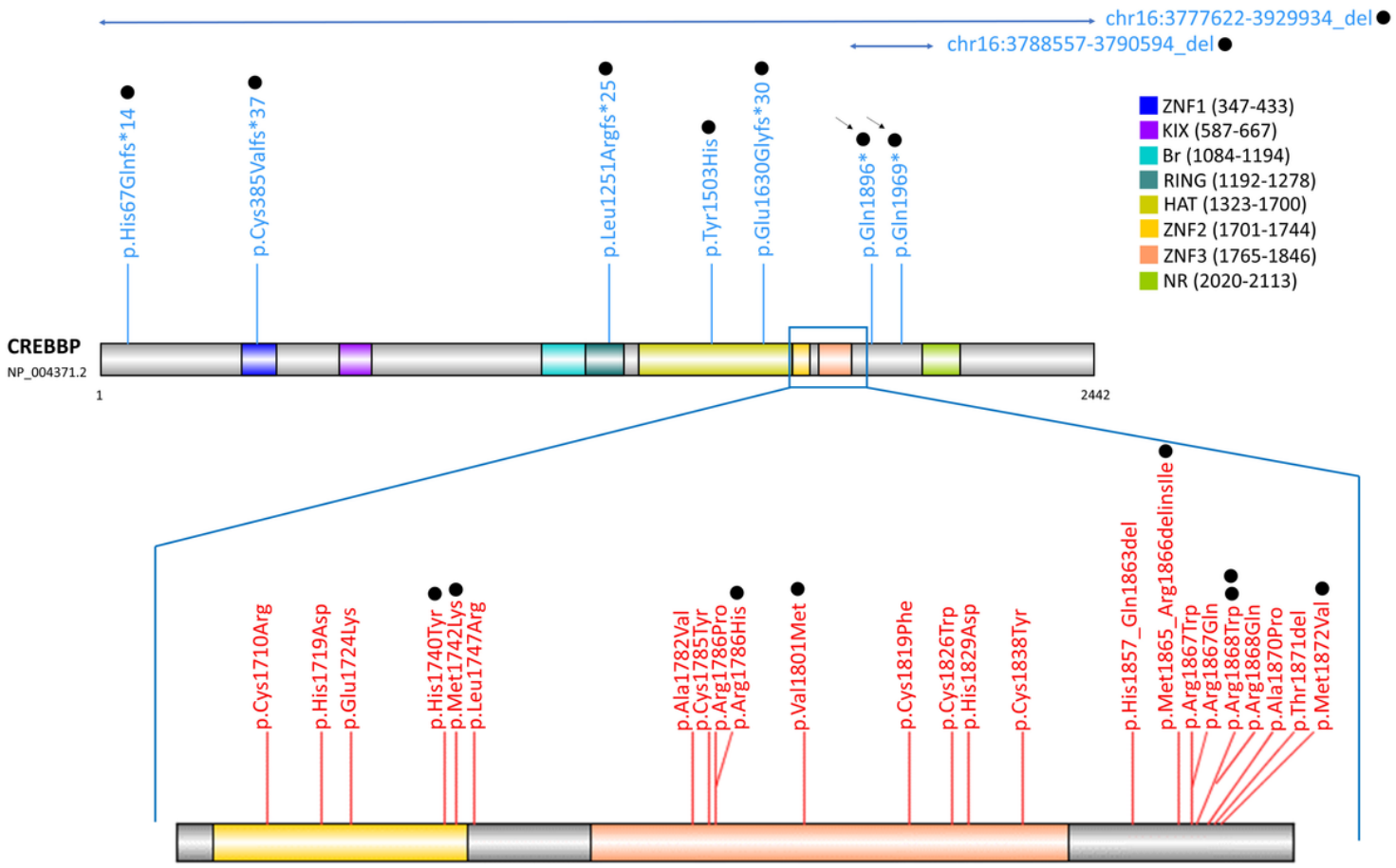

Figure 1

The *CREBBP* protein diagram and the MKHK and RSTS variants. The CREBBP protein (GenBank: NP_004371.2), its functional domains, and the variants used in this study are depicted schematically. Domains are labeled in different colors. ZNF2, ZNF3 and part of IDR is zoomed in for MKHK analysis at the bottom. MKHK and RSTS-causing variants are shown in red and blue, respectively (MKHK group involves all the variants we collected. RSTS group involves variants used in this article). NMD-evasion variants were marked by arrows. Black dots indicate the variants harbored by samples for DNA methylation analysis.

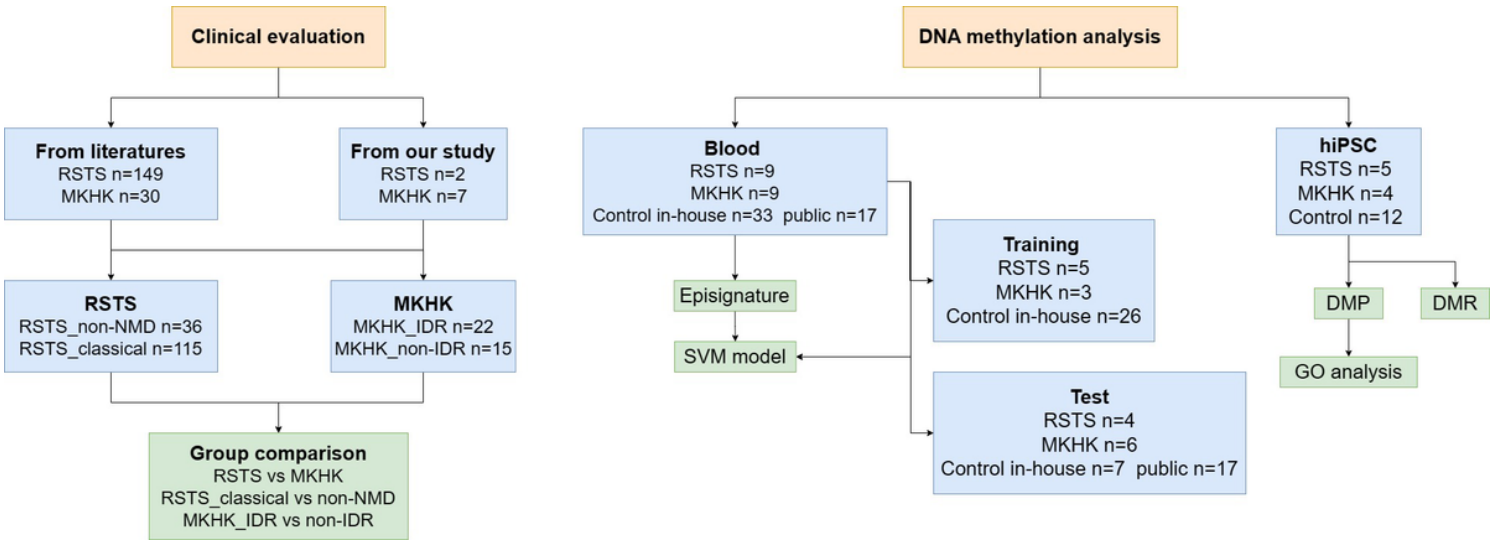

Figure 2

The workflow of the study.

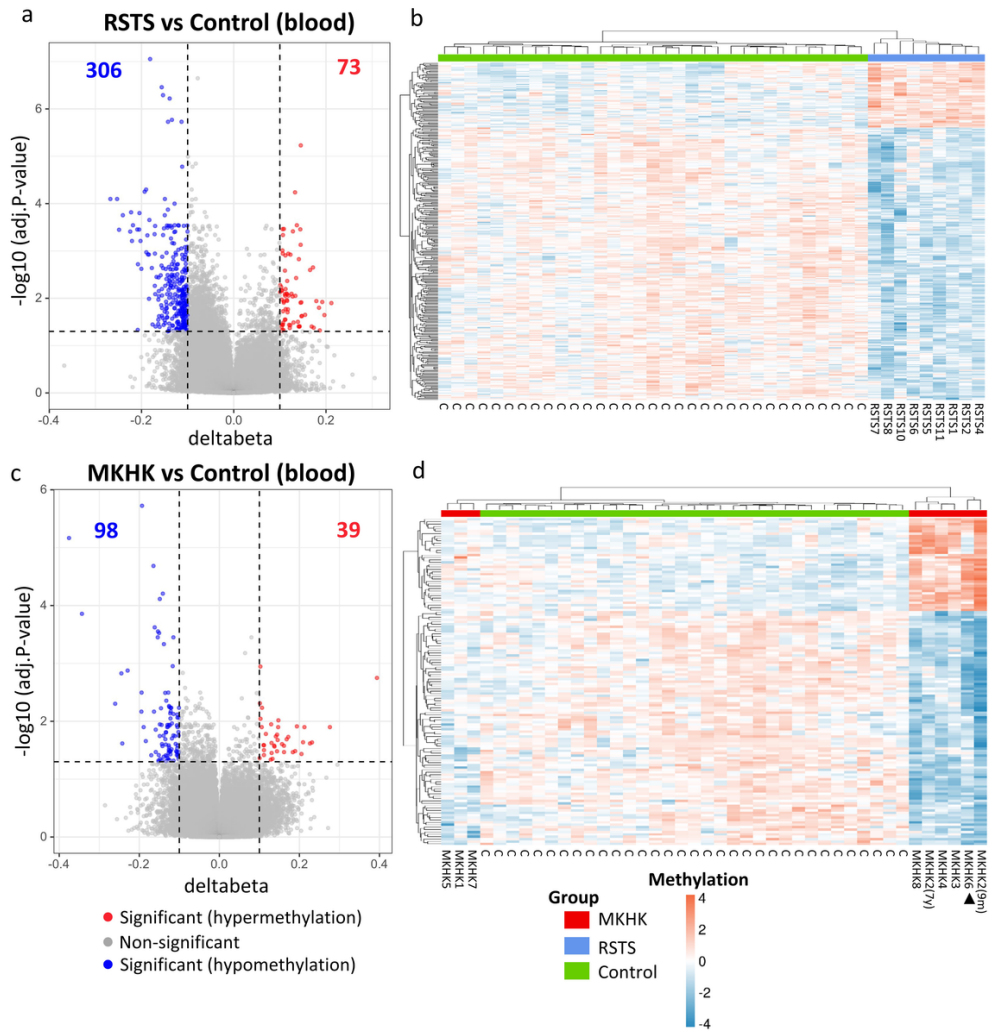

Figure 3

The blood-derived episignatures in RSTS and MKHK. (a) The volcano plot of the methylation difference between RSTS patients and healthy controls. 73 hypermethylation sites and 306 hypomethylation sites were identified. (b) Heatmap showed a cluster of RSTS subjects (n=9, blue) clearly separate from healthy controls (n=33, green) in the RSTS episignature. (c) The volcano plot of the methylation difference between MKHK patients and healthy controls. 39 hypermethylation sites and 98 hypomethylation sites were identified. (d) Heatmap showed two clusters of MKHK subjects (n=9, red) separateby healthy

controls (n=33, green). *CREBBP* variants and the corresponding region are annotated. The umbilical cord blood sample, marked by triangle, was group with the samples with IDR variants. Samples with ZNF2 and ZNF3 variants were grouped in the left.

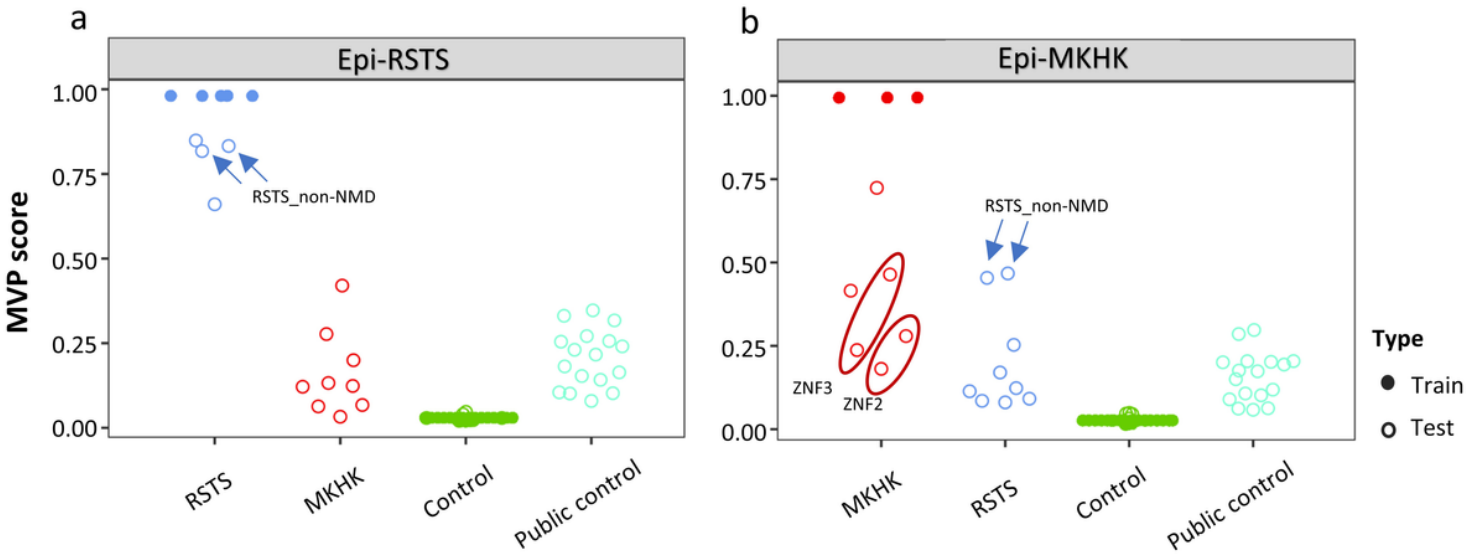

Figure 4

MVP scores based on RSTS and MKHK_IDR classification models Samples were scored using the RSTS SVM model and the MKHK_IDR SVM model. Scores between 0 and 1 with a cutoff of 0.5, calculated for each class on the *X*-axis, are shown on the *Y*-axis. Hollow dots indicate the training samples and filled dots indicate the testing samples. RSTS_NMD subjects are marked by arrows, MKHK_ZNF2/ZNF3 are enclosed by circles. (a) Epi-RSTS, the RSTS SVM model. All the RSTS testing subjects (n=4), MKHK subjects (n=9), testing controls (n=7) and published controls (n=17) were classified correctly, demonstrating 100% sensitivity and specificity of the model. (b) Epi-MKHK, the MKHK_IDR SVM model. one MKHK_IDR testing subject was classified positive, while another MKHK_IDR testing subject with younger age was classified negative, MKHK_ZNF2 (n=2) and MKHK_ZNF3 testing subjects (n=2) were classified negative. Testing controls (n=7), published controls (n=17) and RSTS subjects (n=9) were classified negative.

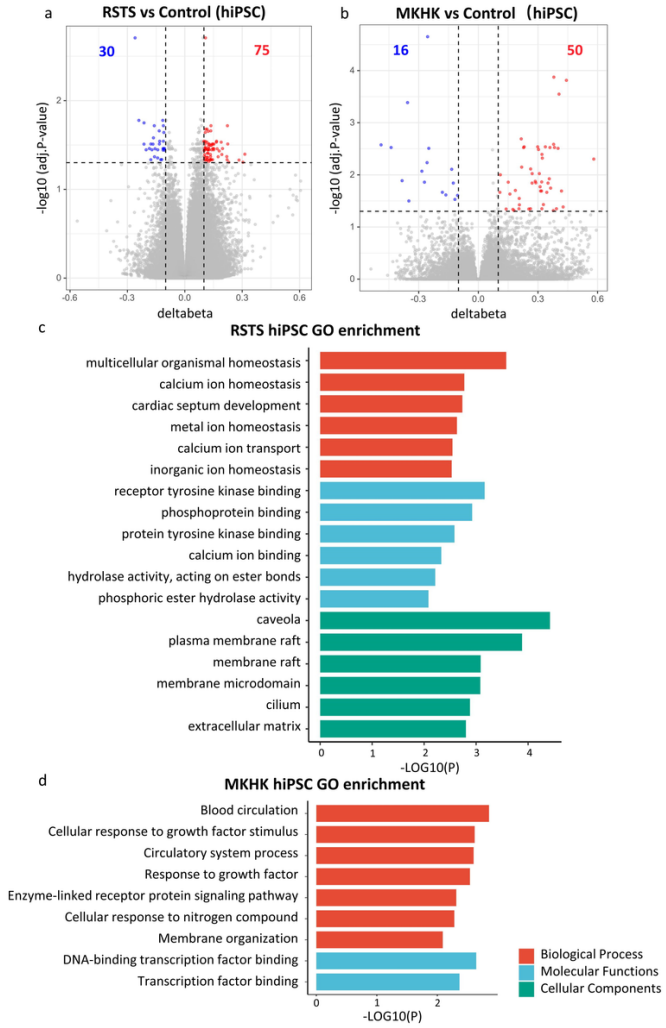

Figure 5

DNA methylation alterations in hiPSCs for MKHK/RSTS. (a and b) The volcano plot of the methylation difference identified 75 hypermethylated sites and 30 hypomethylated sites in RSTS hiPSC group, and 50 hypermethylated sites and 16 hypomethylated sites in MKHK hiPSC group. (c and d) GO terms enriched in the genes of RSTS/MKHK hiPSC DMPs. The *X*-axis represents the negative log of the p values of the enrichment of the corresponding gene ontology terms. All the GO terms of MKHK hiPSCs and the top six GO terms in each category of RSTS hiPSCs are displayed in the figure. Enriched genes from RSTS and MKHK hiPSC DNAm patterns involved in all the terms is shown in Additional file 2: Table S9 and S10.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- Additionalfile2.xlsx