Lucas Alvizi Cruz

Mecanismos genéticos e epigenéticos na etiologia das fissuras orofaciais

Genetic and epigenetic mechanisms in the aetiology of orofacial clefts

São Paulo

2017

Lucas Alvizi Cruz

Mecanismos genéticos e epigenéticos na etiologia das fissuras orofaciais

Genetic and epigenetic mechanisms in the aetiology of orofacial clefts

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Doutor em Ciências, na Área de Biologia/Genética.

Orientadora: Profa. Dra. Maria Rita dos Santos e Passos-Bueno.

São Paulo

2017

Ficha Catalográfica

Cruz, Lucas Alvizi

Mecanismos genéticos e epigenéticos na etiologia das fissuras orofaciais / Lucas Alvizi Cruz; orientadora Maria Rita dos Santos Passos Bueno. -- São Paulo, 2017.

243 f.

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo, Departamento de Genética e Biologia Evolutiva.

1. Desenvolvimento craniofacial. 2. Fissura de lábio e/ou palato. 3. Transcriptoma. 4. Epigenoma. 5. Modelos celulares e animais. I. Passos Bueno, Maria Rita dos Santos, orient. II. Título.

Comissão Julgadora:

_____ _____

Prof(a). Dr(a) Prof(a). Dr(a)

_____ _____

Prof(a). Dr(a) Prof(a). Dr(a)

_____

Profa. Dra. Maria Rita dos Santos Passos Bueno

Orientadora

A todos os pacientes e familiares que conheci durante este projeto

“By changing what man knows about the world, he changes the world he knows; and by changing the world in which he lives, he changes himself. Herein lies a danger and a hope; a danger because random changes of the biological nature are likely to produce deterioration rather than improvement; a hope because changes resulting from knowledge can also be directed by knowledge.”

Theodosius Dobzhansky, 1961

Agradecimentos

Agradeço à minha família, sem os quais nada do que fiz e faço seria possível ou faria sentido. Agradeço a eles, minha mãe Sandra, meu pai Luiz, meus irmãos Thiago e Mateus, minha irmã Deda, meu irmão Vitor, pelo apoio constante nessa jornada e pelo amor que jamais me faltou. Agradeço à minha tia Maria e prima-irmã Déia, sempre interessadas e atentas aos meus passos e ao meu trabalho. Agradeço à minha tia Lúcia e Ana, que me deram todo o suporte e carinho, desde o vestibular, para avançar nessa carreira.

Agradeço à Rita, que desde o primeiro dia no laboratório tratou-me com atenção e acreditou em mim, abrindo as portas do laboratório para as minhas ideias e compartilhando as suas comigo. Obrigado pela sua dedicação na minha formação, pelos anos de orientação e disponibilidade comigo.

Agradeço aos amigos de laboratório, que são minha segunda família: Roberto, Karina, Erika Yeh, Erika Kague, Felipe, Bruno, May, Dani Moreira, Vanessa, Carol, Camila Musso, Camila Lovaglio, Camila Bassi, Ligia, Atique, Bela, Naila, Duda, Clarice, Ágatha, Cláudia, Giovanna Pontinho, Gabi Junia, Lucas Junior, Nicolli, Dani Bueno, Simone, Toninha e Andressa. Em especial, agradeço ao Luciano e ao Gerson, que foram e são meus amigos e mestres desde minha entrada no laboratório e que me ensinaram praticamente tudo que sei.

Gracias a todos los amigos que hice en el laboratório en Argentina, en especial a Antonela, a Seba y a Ezequiel. Muchas gracias a Nora Calcaterra por recibirme y supervisionarme en su laboratório y por toda la atención com mi llegada y estadía em Rosário.

Thanks to all my lab friends in the UK, in special to Emma, Dale, Cristina, Susanne, Nita, Pambos, Sayeda, Miro, Xiayi, Kevin, Karina e Lydia. Also, I am very greatful to Phil Stanier and Gudrun Moore for having me in their lab, supervisioned me and had been so kind to me during all my stay in London.

Agradeço a toda equipe do Centro de Pesquisas do Genoma Humano e Células-Tronco e do Departamento de Genética e Biologia Evolutiva pela disponibilidade, atenção e também pelo aprendizado que obtive com eles.

Agradeço a todos os voluntários e organizadores da Operação Sorriso, os quais foram sempre abertos à colaboração com nosso grupo e com os quais aprendi muito.

Agradeço a todos os meus amigos, que também sempre me apoiaram e fizeram esse percurso mais feliz. Em especial ao Pedrinho, Ju Bueno, Kawano e Rafinya Fontenelle, sempre muito presentes com muito carinho nestes últimos anos.

Este trabalho contou com o apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e do Ministério da Ciência, Tecnologia e Inovação do Brasil.

Notas

Esta tese de doutorado compreende um trabalho desenvolvido durante os anos de 2012 a 2017 no Laboratório de Genética do Desenvolvimento do Centro de Estudos do Genoma Humano e Células Tronco, Instituto de Biociências, Universidade de São Paulo.

A tese foi redigida no modelo de artigos e capítulos, no idioma inglês. Três artigos publicados em periódicos internacionais foram incluídos no corpo principal da tese e seguem a formatação original dos manuscritos conforme cada revista científica nas quais foram publicados. Publicações em co-autoria e não relacionadas ao tema principal da tese encontram-se sumarizadas nos Apêndices, ao final da tese.

O projeto que resultou na presente tese foi aprovado no Comitê de Ética em Pesquisa do Instituto de Biociências da Universidade de São Paulo (Número 132/2011 – FR 488465).

List of Abbreviations

BSAS	Bisulfite amplicon sequencing
ChIP	Chromatin Immunoprecipitation
CLO	Cleft lip only
CLP	Cleft lip and palate
CNC	Cranial neural crest
CPO	Cleft palate only
DEG	Differentially expressed gene
dpf	days post-fertilization
DPSC	Dental pulp stem cell
DSB	Double strand break
EJC	Exon junction comples
EMT	Epithelial-mesenchymal transition
FDR	False discovery rate
GWAS	Genome-wide association study
HDR	Homology-dependent repair
hpf	hours post-fertilization
IPA	Ingenuity pathway analysis
Md	mandibular primordia
MeDIP-seq	Methylated Immunoprecipitated DNA sequencing
meQTL	Methylation-quantitative trait loci
methWAS	Methylome-wide association study
MO	Morpholino
MVP	Methylation variable position
Mx	maxillary primordia
NCC	Neural crest cells
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NMD	Nonsense-mediated decay
NSCL/P	Non-syndromic cleft lip and/or palate
NSCPO	Non-syndromic cleft palate only
OMIM	Online mendelian inheritance in man
ps	palatal shelves
RCPS	Richieri-Costa-Pereira Syndrome
SAM	Significance analysis of microarrays
SHED	Stem cells from human exfoliated deciduous teeth
SNP	Single nucleotide polymorphism
Spl-MO	Splicing-blocking morpholino
TRA-MO	Translation-blocking morpholino
UTR	Untranslated region
WES	Whole-exome sequencing

Table of contents

Chapter 1

General introduction

The craniofacial development and the arisen of orofacial clefts

The human face is a crucial part to a person's identity and fundamental to human inter-recognition and social interaction. Therefore, defects affecting the head and face, so called craniofacial malformations, interfere directly into one of the most important feature of a person's self-identity and psychosocial behavior (CALDER; YOUNG, 2005; JACK; SCHYNS, 2015; LITTLE; JONES; DEBRUINE, 2011) . To understand how craniofacial malformations arise, it is fundamental to understand the basic biological processes of craniofacial development.

The human craniofacial development consists of a series of molecular and cellular mechanisms leading to ontogenetic processes and resulting into the development of facial and cranial structures. Cellular functions as proliferation, migration and differentiation play a role in growth, positioning and fusion of primordial structures of the face, so called facial prominences. At the molecular level, those functions are precisely coordinated in a time-spatial fashion for the proper development of such structures (ADAMEYKO; FRIED, 2016; WILKIE; MORRISS-KAY, 2001). Indeed, several gene signalling pathways have been described in the orchestration of such processes as *FGF*, *BMP*, *TGFB*, *SHH* and *WNT* pathways, in which their synchronization and regulation depend on a complex interaction of both genetic and environmental factors (CLOUTHIER et al., 2000; GOU; ZHANG; XU, 2015; WILKIE; MORRISS-KAY, 2001). Therefore, interferences in the regulation of such signalling pathways and cellular

functions may lead to an improper development of craniofacial structures, thus resulting in craniofacial malformation, as orofacial clefts, affecting lip and or palate development (LESLIE; MARAZITA, 2013)

Regarding the development of structures affected by oral clefts, lip morphogenesis begins at the 4th gestational week, when neural crest cells (NCCs), a highly migratory and multipotent cell population, delaminate from the neural folds and migrate to the developing craniofacial region (Figure 1A). NCCs migration gives rise to the five facial primordia or prominences: the frontonasal prominence, one pair of mandibular prominences and one pair of maxillary prominences (Figure 1B). As the development proceeds, those prominences grow, subdividing into other processes, and start positioning and fusing among themselves. Frontonasal prominence, for example, originates, at the lower portion, the median and lateral nasal prominences (1 pair each, Figure 1B). Upper lip and primary palate are a result of maxillary and medial nasal prominences fusion, occurring until the 7th week (Figure 1B) (JIANG; BUSH; LIDRAL, 2006). On the other hand, secondary palate morphogenesis begins at the 6th gestational week, when the pair of maxillary prominences originate, in the oral cavity, a pair of palatal shelves, taking place laterally to the developing tongue (Figure 2). Initially, the palatal shelves grow vertically and next, at the 7th gestational week, they advance horizontally above the tongue, where they grow and fuse at the midline (Figure 2). The fusion of facial prominences and palatal shelves is dependent on cell death and epithelial-mesenchymal transition, in which the epithelial cells from those contacting structures differentiate to a continuous mesenchymal tissue around the 10*th* gestational week (KERRIGAN et al., 2000; MOSSEY et al., 2009; TWIGG; WILKIE, 2015). At this time, the primordial lip and palate are formed and oral and nasal cavities are completely separated, although failures in those events may lead to cleft lip and or cleft palate.

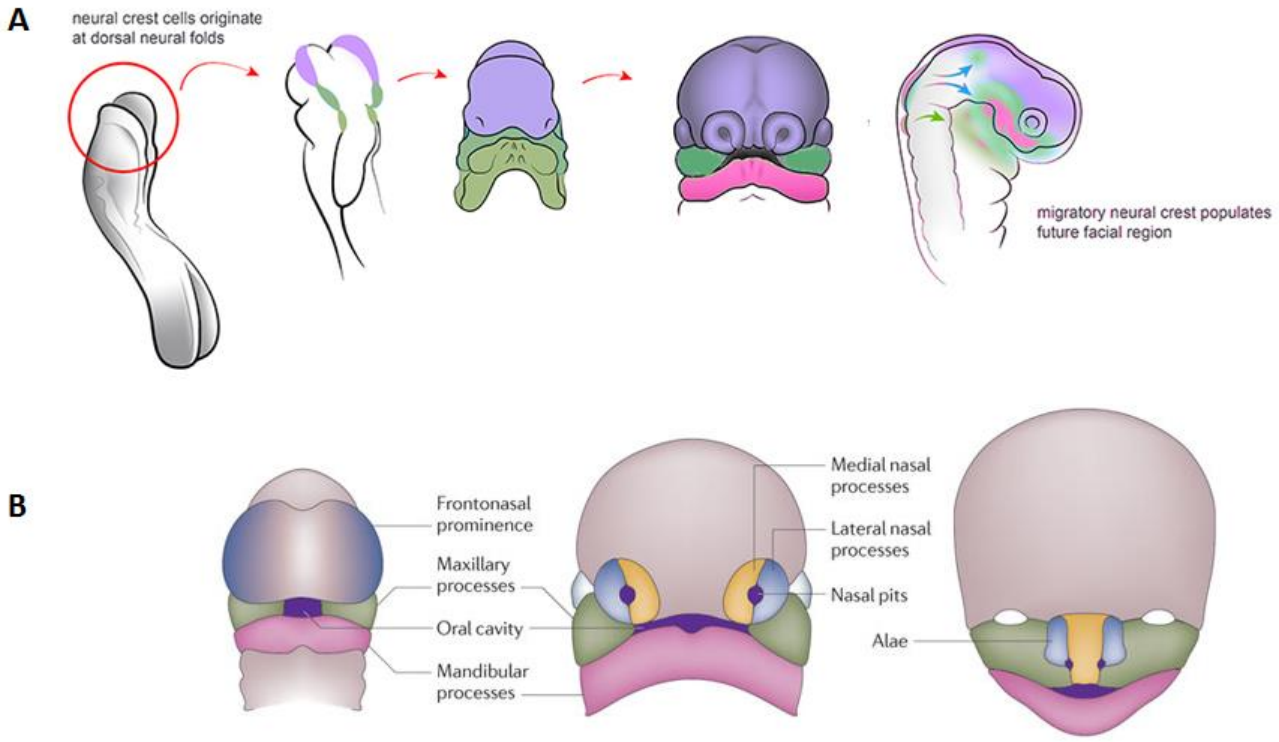


Figure 1: Scheme of craniofacial development in humans. **A**: Craniofacial development starts with Neural Crest Cells (NCCs) from dorsal neural folds, which populates the anterior-frontal part and future head of the embryo. **B**: NCCs populating the forming face originate prominences/processes (frontonasal, maxillary and mandibular). Frontonasal prominence subdivides into lateral and medial nasal prominences, at lower portion. Upper lip is formed from the fusion of maxillary and medial nasal prominences. Adapted from (ADAMEYKO; FRIED, 2016; DIXON et al., 2011).

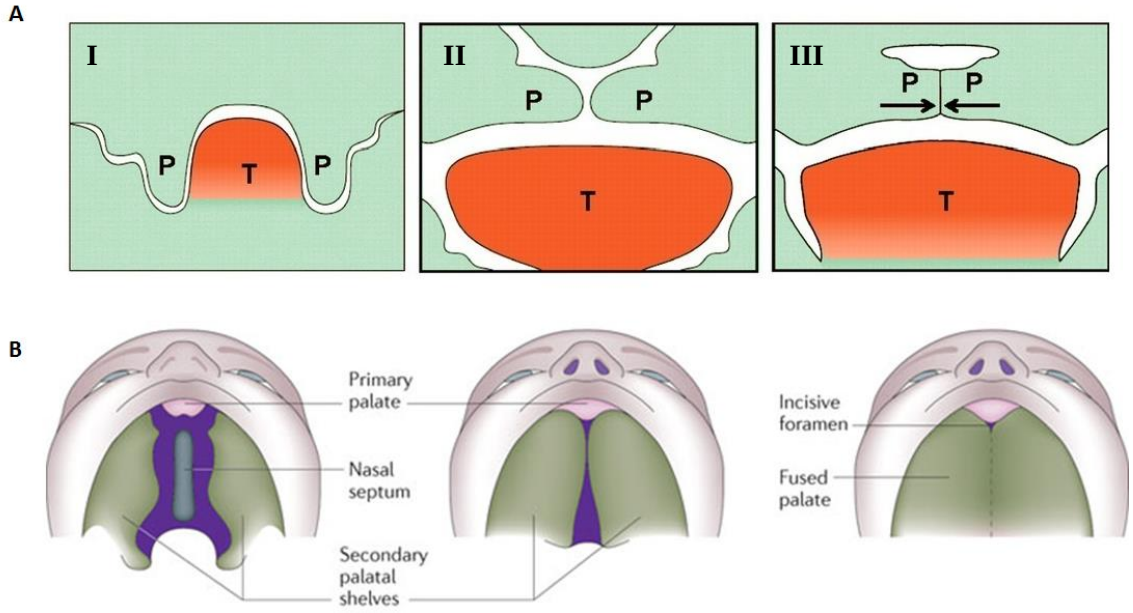


Figure 2: Scheme of secondary palate development. **A**:(I) Maxillary prominences in the oral cavity originate a pair of palatal shelves (P), (II) which grow, position above the tongue (T) and (III) fuse. **B**: Transversal view of palatal shelves growth and fusion, which starts in the anterior part to the posterior. Adapted from (DIXON et al., 2011; RAY; NISWANDER, 2012).

Clinical and Epidemiological Aspects of Orofacial Clefts

Orofacial clefts are clinically characterised by the discontinuity of the structures forming the lip and or palate and their clinics vary largely. For example, cleft lip may be unilateral or bilateral, restricted to the lip (cleft lip only, CLO) or even affect the dental-alveolar region (gum) and the pre or post-incisive foramen palate (cleft lip and palate, CLP). Also, cleft palate can occur without cleft lip (cleft palate only, CPO) and is frequently restricted to the post-incisive foramen palate (Figure 3) (SCHUTTE; MURRAY, 1999). Such defects in lip and or palate compromise basic and fundamental functions for the patient since birth as, for example, new-born feeding and, in later life, speech and psychosocial interaction. In this way, evaluation and rehabilitation of

orofacial cleft patients require a multidisciplinary team with plastic surgeons, dentists, paediatrics, speech therapist, psychotherapists, nurses and geneticists, with interventions starting from the first months of life until adulthood (BRITO et al., 2012; MOSSEY et al., 2009; TAIB et al., 2015; TURNER; RUMSEY; SANDY, 1998).

Epidemiologically, orofacial clefts are the most common congenital craniofacial malformations, affecting 1:700 live-births worldwide (MOSSEY et al., 2009). However, both the differences in lip and palate development, as above discussed, and epidemiological findings support the division of orofacial clefts in two independent malformations: cleft lip with or without cleft palate (CL/P) and CPO (FOGH-ANDERSEN, 1993; FRASER, 1955). In this sense, the prevalence of CL/P changes across populations, varying from 0,3:1000 in Africans, passing by and intermediate prevalence in Europeans as 1:1000, to the highest prevalence observed in Amerindians as 3,6:1000 (CARINCI et al., 2007; VANDERAS, 1987). In Brazil, such prevalence is estimated as 1:1000 (MENEGOTTO; SALZANO, 1991). Differently, the observed prevalence of CPO is frequently estimated as 1:2000 across populations, with no evident ethnic heterogeneity. Also, differences are observed in sex ratio in which CL/P exhibits a 2:1 male-to-female ratio, whereas CPO is more frequent in women.

Because orofacial clefts can occur in the presence of other malformations or comorbidities, they have been classified as non-syndromic or syndromic, in which non-syndromic forms correspond to 70% of CL/P cases and 50% of CPO cases, named as non-syndromic CL/P (NSCL/P) and non-syndromic CPO (NSCPO), respectively (JUGESSUR; FARLIE; KILPATRICK, 2009; STANIER; MOORE, 2004). Despite of the higher proportion of non-syndromic cases, there are more than 500 catalogued syndromes displaying orofacial clefts as part of the phenotype according to OMIM

database (Online Mendelian Inheritance in Man).

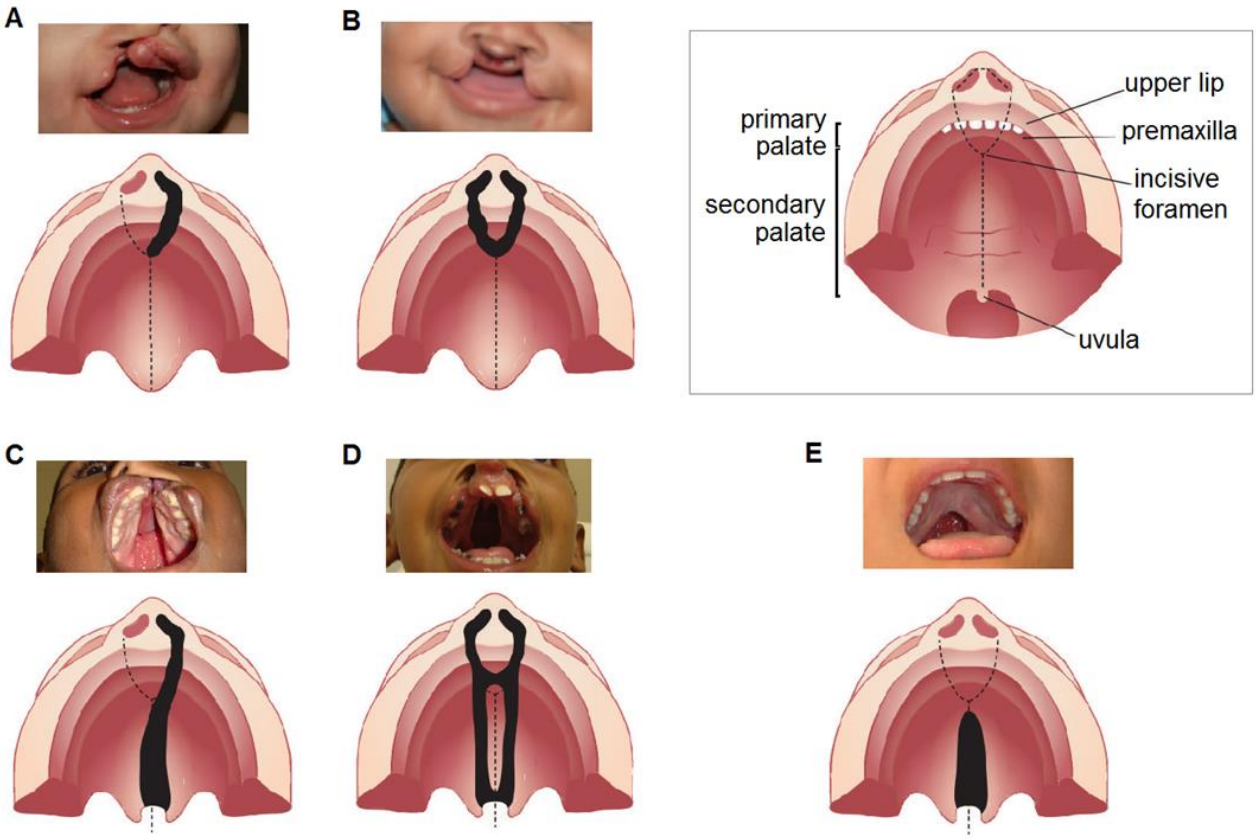


Figure 3: Most common types of clefts affecting lip and palate. **A:** Unilateral cleft lip with alveolar involvement; **B**: Bilateral cleft lip with alveolar involvement; **C**: Unilateral cleft lip with complete cleft palate; **D**: Bilateral cleft lip with complete cleft palate; **E**: Cleft palate only. Adapted from (BRITO et al., 2012).

Considering the high prevalence of orofacial clefts and the continuous treatment in rehabilitation, orofacial clefts are considered a worldwide burden to the public health system. Therefore, understanding the aetiology of such malformations is key to the development of preventive strategies.

Etiology of orofacial clefts: genetics, environment and missing pieces

Syndromic orofacial clefts may arise from gene point mutations, chromosomal abnormalities or environmental factors, such as exposure to teratogens during the first trimester of pregnancy (DIXON et al., 2011). For example, Van der Woude Syndrome (OMIM #119300) and Velocardiofacial Syndrome (or DiGeorge Syndrome, OMIM #188400), the most common syndromic forms of orofacial clefts, are caused by mutations at the *IRF6* gene (*Interferon Regulatory Receptor 6*) and deletions at the 22q11.21 region, respectively (DRISCOLL; BUDARF; EMANUEL, 1992; KONDO et al., 2002). Besides, the identification of mutations associated with syndromic forms of orofacial clefts have greatly contributed to the understanding of critical molecular pathways for the origin of NSCL/P. Therefore, studying rare syndromic orofacial clefts are key to identify causing mutated genes and molecular mechanisms which might be also involved in more frequent forms and NSCL/P.

Regarding teratogenic effects, an example is alcohol foetal syndrome, a syndrome caused by alcohol exposure during early pregnancy, which may include orofacial clefts (MILLER et al., 2006). Many syndromic orofacial clefts, however, still remain with unsolved aetiology.

On the other side, NSCL/P and NSCPO are considered complex disorders, with most cases displaying multifactorial inheritance, in which both genetic and environmental susceptibility factors play a role (DIXON et al., 2011). In this model, susceptibility genetic variants are transmitted to the embryo in which, in the presence of an environmental insult, would lead to the unbalance of the molecular and cellular mechanisms involved in lip and palate morphogenesis.

Recurrence risk and heritability studies have suggested a strong genetic component for both NSCL/P and NSCPO. Considering NSCPO, such studies have been limited for few populations and recurrence risk to first-degree relatives has been

estimated as 2.74%, while heritability has been estimated up to 49% (CHRISTENSEN; MITCHELL, 1996; NORDSTRÖM et al., 1996). On the other, most studies on recurrence risk and heritability of orofacial clefts have focused on NSCL/P, probably due to its higher prevalence. In this sense, NSCL/P recurrence risk for first-degree relatives has been estimated as 4% and heritability ranging from 45% to 85%, depending on the observed population (BRITO et al., 2011; CALZOLARI et al., 1988; CHRISTENSEN; FOGH-ANDERSEN, 1993; HU et al., 1982; MITCHELL et al., 2002).

Several attempts have been made to uncover the molecular basis of orofacial clefts, mostly focused on NSCL/P, with the use of different approaches as linkage analysis, association studies, sequencing and animal modeling. Linkage analysis, and sequencing strategies have been powerful tools in the identification of rare variants of high impact, especially in NSCL/P familial cases, in which mutations in *FOXE1, FGFR1*, *FGF8*, *MSX1*, *IRF6*, *CDH1* e *ARHGAP29* have been identified (ARDINGER et al., 1989; BRITO et al., 2015; MORENO et al., 2009; RAHIMOV et al., 2008; RILEY; MURRAY, 2007; SAVASTANO et al., 2017; VIEIRA et al., 2005). Complementarily, animal studies modeling have demonstrated the importance of such genes and gene pathways as WNT-Beta-Catenin signaling and Epithelial-Mesenchymal transition pathways (KUROSAKA et al., 2014). On the other hand, Genome-Wide Association Studies (GWAS) have been applied to the identification of low-impact common variants conferring risk to NSCL/P, mostly in isolated cases. Those GWAS identified loci are single-nucleotide polymorphisms (SNPs) in the gene *IRF6* and 8q24.21, 1p36 and 10q25 among other candidate regions outros (BEATY et al., 2010; BRITO et al., 2012; GRANT et al., 2009; LUDWIG et al., 2012; MANGOLD et al., 2010). However, such common variants do not contribute expressively to the NSCL/P risk once Odds-ratios are frequently estimated lower than 2 (BIRNBAUM et al., 2009; LUDWIG et al., 2012; MANGOLD et al., 2010; SUN et al., 2015) and therefore are not sufficient to explain the observed high heritability. Thus NSCL/P falls into the common problem described in complex diseases

denominated as “the missing heritability”, in which variants identified for traits with considerable high genetic contribution are not enough to justify such heritability (BRITO et al., 2012; MAHER, 2008; SCHIERDING; CUTFIELD; O’SULLIVAN, 2014). This deficiency in uncover the NSCL/P genetic component is given by the complexity of the interaction among genes and environment in a multifactorial system. Such scenario reveals the importance of complementing those approaches with the investigation of complementary molecular mechanisms integrating genetic variation, at DNA and transcriptome level, with environmental factors. In this sense, by studying NSCL/P transcriptome, dysregulated genes and/or pathways can be identified and lead to the discovery of altered regulatory mechanisms, such as epigenetic factors, of great importance to craniofacial development.

Regarding the investigation of environmental factors for orofacial clefts, several epidemiological studies have suggested malnutrition, low socioeconomic status, maternal smoking, anti-epileptic drug exposure, retinoic acid and alcohol consumption as risk factors (ACUÑA-GONZÁLEZ et al., 2011; BEATY et al., 2001; DEROO et al., 2008; JIA et al., 2011; MCKINNEY et al., 2013). Despite of presenting associations, those studies do not focus on the molecular and cellular mechanisms by which the exposure to such factors would lead to orofacial clefts. Notably, many of those environmental factors are known to influence gene expression through epigenetic changes (ANGRISANO et al., 2011; HOU et al., 2012; LEE; PAUSOVA, 2013). Epigenetic factors are heritable mechanisms that drive changes in gene activity without altering the DNA sequence, and can involve DNA methylation, chromatin modifications and expression of non-coding RNAs (BIRD, 2007). Considering gene regulation, epigenetic changes are a connection between environmental insults and gene function, as they act as a gene-environment regulation switch controlling chromatin access by the transcriptional machinery. Thus, epigenetic changes may have direct impact in gene expression and reflect in gene pathways and networks dysregulation and impairment in

cellular processes. (JAENISCH; BIRD, 2003). Many common diseases are known to be influenced or caused by epigenetic disturbances, as described for some types of cancer, heart failure and diabetes (DUYGU; POELS; DA COSTA MARTINS, 2013; FLORATH et al., 2016; GILBERT; LIU, 2012; KANWAL; GUPTA, 2012; SORIANO-TÁRRAGA et al., 2016). Also, epigenetic control has been demonstrated to be critical for proper craniofacial development in mice, especially during palatal development (SEELAN et al., 2013a, 2013b). However, no epigenetic study involving orofacial cleft patients has been performed to date and the association of epigenetic factors to orofacial clefts has not been demonstrated yet.

Therefore, understanding the variation in gene expression and molecular mechanisms controlling gene regulation, as those in epigenetics, are essential to better clarify the aetiology of orofacial clefts, which is our main interest here. Elucidating how gene dysregulation, and by which mechanisms they act, contribute to the risk for orofacial clefts could be also of great value to identify the factors under the “missing heritability”. We expect that solving those aetiological mechanisms will be important to the development of preventive strategies in the future.

Objectives

The main objective of this study was to investigate the molecular mechanisms involved in the aetiology of orofacial clefts, which was focused in gene expression and epigenetic analysis in NSCL/P as well as genetic, gene expression, animal modeling and epigenetics in Richieri-Costa-Pereira Syndrome, a rare syndromic form of orofacial cleft.

The specific objectives were:

- To identify differentially expressed genes and pathways by the study of transcriptome in NSCL/P cells;
- To verify whether epigenetic mechanisms were involved in the dysregulation of gene expression in NSCL/P cells;
- To investigate DNA methylation at 8q24.21 region, one of the most significant GWAS regions for NSCL/P;
- To investigate DNA methylation at a genomic level in NSCL/P;
- To identify the genetic cause of Richieri-Costa-Pereira Syndrome and its effects at the expression level;
- To investigate the involvement of DNA methylation in association with the Richieri-Costa-Pereira Syndrome genetic cause.

References

ACUÑA-GONZÁLEZ, G. et al. Family history and socioeconomic risk factors for non-syndromic cleft lip and palate: a matched case-control study in a less developed country. **Biomedica: Revista Del Instituto Nacional de Salud**, v. 31, n. 3, p. 381-391, set. 2011.

ADAMEYKO, I.; FRIED, K. The nervous system orchestrates and integrates craniofacial development: A review. **Frontiers in physiology**, v. 7, p. 49, 19 fev. 2016.

ANGRISANO, T. et al. Chromatin and DNA methylation dynamics during retinoic acid-induced RET gene transcriptional activation in neuroblastoma cells. **Nucleic Acids Research**, v. 39, n. 6, p. 1993-2006, mar. 2011.

ARDINGER, H. H. et al. Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. **American Journal of Human Genetics**, v. 45, n. 3, p. 348-353, set. 1989.

BEATY, T. H. et al. A case-control study of nonsyndromic oral clefts in Maryland. **Annals of Epidemiology**, v. 11, n. 6, p. 434-442, ago. 2001.

BEATY, T. H. et al. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. **Nature Genetics**, v. 42, n. 6, p. 525-529, jun. 2010.

BIRD, A. Perceptions of epigenetics. **Nature**, v. 447, n. 7143, p. 396-398, 24 maio 2007.

BIRNBAUM, S. et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. **Nature Genetics**, v. 41, n. 4, p. 473-477, abr. 2009.

BRITO, L. A. et al. Genetic contribution for non-syndromic cleft lip with or without cleft palate (NS CL/P) in different regions of Brazil and implications for association studies. **American Journal of Medical Genetics. Part A**, v. 155A, n. 7, p. 1581-1587, jul. 2011.

BRITO, L. A. et al. Genetics and management of the patient with orofacial cleft. **Plastic surgery international**, v. 2012, p. 782821, 1 nov. 2012.

BRITO, L. A. et al. Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate. **Human Mutation**, v. 36, n. 11, p. 1029-1033, nov. 2015.

CALDER, A. J.; YOUNG, A. W. Understanding the recognition of facial identity and facial expression. **Nature Reviews. Neuroscience**, v. 6, n. 8, p. 641-651, ago. 2005.

CALZOLARI, E. et al. Epidemiological and genetic study of 200 cases of oral cleft in the Emilia Romagna region of northern Italy. **Teratology**, v. 38, n. 6, p. 559-564, dez. 1988.

CARINCI, F. et al. Human genetic factors in nonsyndromic cleft lip and palate: an update. **International Journal of Pediatric Otorhinolaryngology**, v. 71, n. 10, p. 1509-1519, out. 2007.

CHRISTENSEN, K.; FOGHAAA501ANDERSEN, P. Cleft lip (+/- cleft palate) in Danish twins, 1970-1990. **American Journal of Medical Genetics**, v. 47, n. 6, p. 910-916, 1 nov. 1993.

CHRISTENSEN, K.; MITCHELL, L. E. Familial recurrence-pattern analysis of nonsyndromic isolated cleft palate--a Danish Registry study. **American Journal of Human Genetics**, v. 58, n. 1, p. 182-190, jan. 1996.

CLOUTHIER, D. E. et al. Signaling pathways crucial for craniofacial development revealed by endothelin-A receptor-deficient mice. **Developmental Biology**, v. 217, n. 1, p. 10-24, 1 jan. 2000.

DEROO, L. A. et al. First-trimester maternal alcohol consumption and the risk of infant oral clefts in Norway: a population-based case-control study. **American Journal of Epidemiology**, v. 168, n. 6, p. 638-646, 15 set. 2008.

DIXON, M. J. et al. Cleft lip and palate: understanding genetic and environmental influences. **Nature Reviews. Genetics**, v. 12, n. 3, p. 167-178, mar. 2011.

DRISCOLL, D. A.; BUDARF, M. L.; EMANUEL, B. S. A genetic etiology for DiGeorge syndrome: consistent deletions and microdeletions of 22q11. **American Journal of Human Genetics**, v. 50, n. 5, p. 924-933, maio 1992.

DUYGU, B.; POELS, E. M.; DA COSTA MARTINS, P. A. Genetics and epigenetics of arrhythmia and heart failure. **Frontiers in genetics**, v. 4, p. 219, 30 out. 2013.

FLORATH, I. et al. Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. **Diabetologia**, v. 59, n. 1, p. 130-138, jan. 2016.

FOGH-ANDERSEN, P. Inheritance of harelip and cleft palate; contribution to the elucidation of the etiology of the congenital clefts of the face. 1993.

FRASER, F. C. Thoughts on the etiology of clefts of the palate and lip. **Acta genetica et statistica medica**, v. 5, n. 4, p. 358-369, 1955.

GILBERT, E. R.; LIU, D. Epigenetics: the missing link to understanding β-cell dysfunction in the pathogenesis of type 2 diabetes. **Epigenetics**, v. 7, n. 8, p. 841-852, ago. 2012.

GOU, Y.; ZHANG, T.; XU, J. Transcription factors in craniofacial development: from receptor signaling to transcriptional and epigenetic regulation. **Current Topics in Developmental Biology**, v. 115, p. 377-410, 6 out. 2015.

GRANT, S. F. A. et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. **The Journal of Pediatrics**, v. 155, n. 6, p. 909-913, dez. 2009.

HOU, L. et al. Environmental chemical exposures and human epigenetics. **International Journal of Epidemiology**, v. 41, n. 1, p. 79-105, fev. 2012.

HU, D. N. et al. Genetics of cleft lip and cleft palate in China. **American Journal of Human Genetics**, v. 34, n. 6, p. 999-1002, nov. 1982.

JACK, R. E.; SCHYNS, P. G. The human face as a dynamic tool for social communication. **Current Biology**, v. 25, n. 14, p. R621-34, 20 jul. 2015.

JAENISCH, R.; BIRD, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. **Nature Genetics**, v. 33 Suppl, p. 245-254, mar. 2003.

JIA, Z. L. et al. Maternal malnutrition, environmental exposure during pregnancy and the risk of non-syndromic orofacial clefts. **Oral Diseases**, v. 17, n. 6, p. 584-589, set. 2011.

JIANG, R.; BUSH, J. O.; LIDRAL, A. C. Development of the upper lip: morphogenetic and molecular mechanisms. **Developmental Dynamics**, v. 235, n. 5, p. 1152-1166, maio 2006.

JUGESSUR, A.; FARLIE, P. G.; KILPATRICK, N. The genetics of isolated orofacial clefts: from genotypes to subphenotypes. **Oral Diseases**, v. 15, n. 7, p. 437-453, out. 2009.

KANWAL, R.; GUPTA, S. Epigenetic modifications in cancer. **Clinical Genetics**, v. 81, n. 4, p. 303-311, abr. 2012.

KERRIGAN, J. J. et al. Palatogenesis and potential mechanisms for clefting. **Journal of the Royal College of Surgeons of Edinburgh**, v. 45, n. 6, p. 351-358, dez. 2000.

KONDO, S. et al. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. **Nature Genetics**, v. 32, n. 2, p. 285-289, out. 2002.

KUROSAKA, H. et al. Disrupting hedgehog and WNT signaling interactions promotes cleft lip pathogenesis. **The Journal of Clinical Investigation**, v. 124, n. 4, p. 1660-1671, abr. 2014.

LEE, K. W. K.; PAUSOVA, Z. Cigarette smoking and DNA methylation. **Frontiers in genetics**, v. 4, p. 132, 17 jul. 2013.

LESLIE, E. J.; MARAZITA, M. L. Genetics of cleft lip and cleft palate. **American journal of medical genetics. Part C, Seminars in medical genetics**, v. 163C, n. 4, p. 246-258, nov. 2013.

LITTLE, A. C.; JONES, B. C.; DEBRUINE, L. M. The many faces of research on face perception. **Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences**, v. 366, n. 1571, p. 1634-1637, 12 jun. 2011.

LUDWIG, K. U. et al. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. **Nature Genetics**, v. 44, n. 9, p. 968-971, set. 2012.

MAHER, B. Personal genomes: The case of the missing heritability. **Nature**, v. 456, n. 7218, p. 18-21, 6 nov. 2008.

MANGOLD, E. et al. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. **Nature Genetics**, v. 42, n. 1, p. 24-26, jan. 2010.

MCKINNEY, C. M. et al. Micronutrients and oral clefts: a case-control study. **Journal of Dental Research**, v. 92, n. 12, p. 1089-1094, dez. 2013.

MENEGOTTO, B. G.; SALZANO, F. M. Epidemiology of oral clefts in a large South American sample. **The Cleft Palate-Craniofacial Journal**, v. 28, n. 4, p. 373-6; discussion 376, out. 1991.

MILLER, L. C. et al. Fetal alcohol spectrum disorders in children residing in Russian orphanages: a phenotypic survey. **Alcoholism, Clinical and Experimental Research**, v. 30, n. 3, p. 531-538, mar. 2006.

MITCHELL, L. E. et al. Guidelines for the design and analysis of studies on nonsyndromic cleft lip and cleft palate in humans: summary report from a Workshop of the International Consortium for Oral Clefts Genetics. **The Cleft Palate-Craniofacial Journal**, v. 39, n. 1, p. 93-100, jan. 2002.

MORENO, L. M. et al. FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. **Human Molecular Genetics**, v. 18, n. 24, p. 4879-4896, 15 dez. 2009.

MOSSEY, P. A. et al. Cleft lip and palate. **The Lancet**, v. 374, n. 9703, p. 1773-1785, 21 nov. 2009.

NORDSTRÖM, R. E. et al. Cleft-twin sets in Finland 1948-1987. **The Cleft Palate-Craniofacial Journal**, v. 33, n. 4, p. 340-347, jul. 1996.

RAHIMOV, F. et al. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. **Nature Genetics**, v. 40, n. 11, p. 1341-1347, nov. 2008.

RAY, H. J.; NISWANDER, L. Mechanisms of tissue fusion during development. **Development**, v. 139, n. 10, p. 1701-1711, maio 2012.

RILEY, B. M.; MURRAY, J. C. Sequence evaluation of FGF and FGFR gene conserved non-coding elements in non-syndromic cleft lip and palate cases. **American Journal of Medical Genetics. Part A**, v. 143A, n. 24, p. 3228-3234, 15 dez. 2007.

SAVASTANO, C. P. et al. Impact of rare variants in ARHGAP29 to the etiology of oral clefts: role of loss-of-function vs missense variants. **Clinical Genetics**, v. 91, n. 5, p. 683-689, maio 2017.

SCHIERDING, W.; CUTFIELD, W. S.; O'SULLIVAN, J. M. The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell. **Frontiers in genetics**, v. 5, p. 39, 18 fev. 2014.

SCHUTTE, B. C.; MURRAY, J. C. The many faces and factors of orofacial clefts. **Human Molecular Genetics**, v. 8, n. 10, p. 1853-1859, 1999.

SEELAN, R. S. et al. Epigenetic regulation of Sox4 during palate development. **Epigenomics**, v. 5, n. 2, p. 131-146, abr. 2013a.

SEELAN, R. S. et al. Developmental profiles of the murine palatal methylome. **Birth Defects Research. Part A, Clinical and Molecular Teratology**, v. 97, n. 4, p. 171-186, 3 abr. 2013b.

SORIANO-TÁRRAGA, C. et al. Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. **Human Molecular Genetics**, v. 25, n. 3, p. 609-619, 1 fev. 2016.

STANIER, P.; MOORE, G. E. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. **Human Molecular Genetics**, v. 13 Spec No 1, p. R73-81, 1 abr. 2004.

SUN, Y. et al. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. **Nature Communications**, v. 6, p. 6414, 16 mar. 2015.

TAIB, B. G. et al. Cleft lip and palate: diagnosis and management. **British Journal of Hospital Medicine**, v. 76, n. 10, p. 584-5, 588, out. 2015.

TURNER, S. R.; RUMSEY, N.; SANDY, J. R. Psychological aspects of cleft lip and palate. **European Journal of Orthodontics**, v. 20, n. 4, p. 407-415, ago. 1998.

TWIGG, S. R. F.; WILKIE, A. O. M. New insights into craniofacial malformations. **Human Molecular Genetics**, v. 24, n. R1, p. R50-9, 15 out. 2015.

VANDERAS, A. P. Incidence of cleft lip, cleft palate, and cleft lip and palate among races: a review. **The Cleft palate journal**, v. 24, n. 3, p. 216-225, jul. 1987.

VIEIRA, A. R. et al. Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. **PLoS Genetics**, v. 1, n. 6, p. e64, 2 dez. 2005.

WILKIE, A. O.; MORRISS-KAY, G. M. Genetics of craniofacial development and malformation. **Nature Reviews. Genetics**, v. 2, n. 6, p. 458-468, jun. 2001.

Chapter 2

Susceptibility to DNA Damage as a Molecular Mechanism for Non-Syndromic Cleft Lip and Palate

*Gerson Shigeru Kobayashi[1¥], Lucas Alvizi[1¥], Daniele Yumi Sunaga[1], Philippa Francis-West[2], Anna Kuta[2], Bruno Vinícius Pimenta Almada[1], Simone Gomes Ferreira[1], Leonardo Carmo de Andrade-Lima[3], Daniela Franco Bueno[1, 4], Cássio Eduardo Raposo-Amaral[4], Carlos Frederico Menck[3] and Maria Rita Passos-Bueno[1]**

1 *Human Genome Research Center, Institute for Biosciences, University of São Paulo, São Paulo, Brazil;* ***2*** *Dental Institute, Department of Craniofacial Development and Stem Cell Biology, King's College London, London, United Kingdom;* ***3*** *Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil;* ***4*** *SOBRAPAR Institute, Campinas, São Paulo, Brazil.*

(¥) These authors contributed equally to this work

Key-words: cleft lip and palate, gene expression, differential expression, DNA damage repair

This chapter was published in PLoS ONE in June 2013:

https://doi.org/10.1371/journal.pone.0065677

Abstract

Non-syndromic cleft lip/palate (NSCL/P) is a complex, frequent congenital malformation, determined by the interplay between genetic and environmental factors during embryonic development. Previous findings have appointed an aetiological overlap between NSCL/P and cancer, and alterations in similar biological pathways may underpin both conditions. Here, using a combination of transcriptomic profiling and functional approaches, we report that NSCL/P dental pulp stem cells exhibit dysregulation of a co-expressed gene network mainly associated with DNA double-strand break repair and cell cycle control ($p = 2.88x10^{-2} – 5.02x10^{-9}$). This network included important genes for these cellular processes, such as *BRCA1*, *RAD51*, and *MSH2*, which are predicted to be regulated by transcription factor E2F1. Functional assays support these findings, revealing that NSCL/P cells accumulate DNA double-strand breaks upon exposure to H_2O_2. Furthermore, we show that *E2f1*, *Brca1* and *Rad51* are co-expressed in the developing embryonic orofacial primordia, and may act as a molecular hub playing a role in lip and palate morphogenesis. In conclusion, we show for the first time that cellular defences against DNA damage may take part in determining the susceptibility to NSCL/P. These results are in accordance with the hypothesis of aetiological overlap between this malformation and cancer, and suggest a new pathogenic mechanism for the disease.

Resumo

As fissuras lábio-palatinas não sindrômicas (FL/P NS) são malformações complexas e frequentes, ocasionadas pela interação de fatores genéticos e ambientas durante o desenvolvimento embrionário. Estudos anteriores apontam para uma intersecção etiológica entre as FL/P NS e câncer, de forma que alterações em vias biológicas em comum podem levar a ambas condições. Neste trabalho, usando-se uma combinação de análise de transcriptomas e abordagens funcionais, nós descrevemos que células tronco de polpa de dente de FL/P NS exibem uma desregulação de uma rede gênica co-expressa associada ao reparo de duplas quebras de DNA e controle de ciclo celular ($p = 2.88x10^{-2} – 5.02x10^{-9}$). Esta rede inclui genes importantes para estes processos celulares, como *BRCA1*, *RAD51*, e *MSH2*, que são preditos a serem regulados pelo fator de transcrição E2F1. Ensaios Funcionais apoiam estes achados, revelando que células FL/P NS acumulam quebras duplas de DNA sob exposição ao H_2O_2. Além disso, nós demonstramos que *E2f1*, *Brca1* e *Rad51* são co-expressos nos primórdios orofaciais em desenvolvimento e podem atuar como um *hub* molecular desempenhando um papel na morfogênese dos lábios e palato. Em conclusão, nós demonstramos pela primeira vez que respostas celulares ao dano de DNA podem ter participação na susceptibilidade às FL/P NS. Estes resultados estão de acordo com a hipótese de intersecção entre esta malformação e câncer, sugerindo-se um novo mecanismo patogênico para a doença.

Introduction

Non-syndromic cleft lip with or without cleft palate (NSCL/P [OMIM %119530]) is one of the most common congenital defects. Its birth prevalence is variable, ranging from 3.4 to 22.9 per 10,000 births world-wide, depending upon factors such as ethnic background, geographical location, and socio-economic status [1].

The interplay between genetic and environmental factors during embryonic development is thought to be determinant in the aetiology of NSCL/P. Genome-wide association studies (GWAS) have enabled the consistent identification of several candidate *loci*. However, the low attributed risk for each variant fails to explain the estimated heritability for NSCL/P (as high as 85% in some populations [2]), and little is known about their functional role in the pathogenesis of the disease. The fact that non-coding genomic regions harbour many of these variants (which include an enhancer upstream of *IRF6*, and intronic and intergenic regions [3,4,5,6] is indicative that they might play a transcriptional regulatory role in the embryo, in agreement with the idea that alterations in gene expression may be a relevant mechanism underlying susceptibility to complex diseases [7]. Therefore, if disease susceptibility is shaped by transcriptional anomalies, which in turn are driven by the individual's genetic constitution, a feasible approach to investigate the aetiology of NSCL/P is the gene expression analysis of cells from affected individuals. This strategy not only allows the screening of candidate biological pathways contributing to the disease, but also enables the investigation of environmental agents and how they affect these pathways. Previous data derived from fibroblasts and mesenchymal stem cells support this rationale, as they revealed alterations in gene networks functionally relevant for orofacial development, such as collagen metabolism and extracellular matrix remodelling [8,9,10,11].

NSCL/P is thought to arise from anomalies in cell migration, proliferation, transdifferentiation and apoptosis [12,13,14], all of which are known to be involved in cancer progression. The relationship between orofacial clefts and cancer is subject of debate; however, several studies have reported co-occurrence of orofacial clefts and a variety of cancer types.[15,16,17,18]. Accordingly, alterations in genes that are known to play diverse roles during carcinogenesis, such as *CDH1*, *TP63*, *NBS* and *AXIN2*, have been related to both syndromic and non-syndromic cleft lip/palate [19,20,21,22,23,24]. Moreover, both are common diseases with significant genetic heterogeneity; therefore, an aetiological overlap is more likely to occur when compared to other diseases. Given these observations, similar biological pathways may be underpinning susceptibility to both conditions.

Thus, our objective was to search for dysregulated gene networks associated with tumourigenesis, using NSCL/P and control dental pulp stem cell cultures. We chose this cell source because it comprises populations of mesodermal and neural crest-derived cells, and therefore it shares the same origin with the cells that populate the mesenchyme of the craniofacial structures involved in lip and palate morphogenesis [25,26,27]; in addition, we have previously demonstrated the applicability of using these cells in detecting gene networks important for NSCL/P aetiology [9]. We expect that the results will not only assist in elucidating the aetiology of NSCL/P, but also provide more information on the mechanisms through which it relates to cancer.

Results

Overview of the differentially expressed genes (DEGs)

By comparing the microarray expression data from NSCL/P (n = 7) and control (n = 6) dental pulp stem cell cultures, we obtained 126 and 211 DEGs using the SAM (Significance Analysis of Microarrays) and Rank Products algorithms, respectively. Combining these two gene sets, we observed an overlap of 109 genes from both methodologies, with a final number of 228 DEGs (72 up-regulated, 156 down-regulated; Table S1), which were submitted to the subsequent analyses.

A transcriptional network associated with response to DNA damage and cell cycle control is dysregulated in NSCL/P cells

To functionally characterise the DEGs and to determine possible interactions between them, an IPA analysis (Interactive pathways analysis of complex 'omics data; Ingenuity Systems) was performed. The highest-scoring network assembled by IPA (score = 61) harboured DEGs associated with the following functions: Cell Cycle; DNA Replication, Recombination, and Repair; and Cellular Compromise ($p = 2.88x10^{-2}$ – $5.02x10^{-9}$; Table S2), in which the gene *BRCA1* occupied a central node, functionally connected to a variety of other molecules associated with DNA repair and cell cycle regulation (e.g. *MSH2*, *BLM*, *RAD51*, *CDC6, CLSPN*; Fig. 1A). Moreover, the top canonical pathway enriched in our gene set was 'Role of BRCA1 in DNA Damage Response' (p-value = $3.92x10^{-8}$; Fig. 1B and 1C). Taken together, the results indicated that NSCL/P cells feature transcriptional dysregulation of genes involved in cell cycle control and DNA damage repair mediated by BRCA1. This motivated us to further investigate this molecule and its relationship with the remaining DEGs.

Assuming that co-expressed genes may partake in the same biological process, we performed a similarity-based clustering analysis, based on *BRCA1* transcript levels. We obtained a highly homogeneous cluster harbouring 30 genes of similar expression patterns across samples (average homogeneity = 0.974, Fig. 2A), including several genes pertaining to the IPA interaction network and BRCA1-mediated DNA repair canonical pathway, such as *MSH2*, *RAD51*, and *BLM*. In accordance with the observations derived from IPA, this cluster exhibited Gene Ontology enrichment for attributes related to the DNA repair machinery and regulation of cell cycle ($p = 0.001$; Fig. 2B). Subsequent transcription factor binding site enrichment analysis revealed E2F1 as a putative regulator for this co-expression cluster ($p < 0.05$; Fig. 2B). To corroborate this result, we used ChIP-chip information available on the FANTOM4 (Functional ANnoTation of the Mammalian genome) database, demonstrating that E2F1 is experimentally proven to interact with 23 out of all 30 clustered genes (Fig. 2C).

Validation of the microarray assays using quantitative real time PCR (qRT-PCR)

We carried out the qRT-PCR validation using RNA extracted from independent cell cultures from the same individuals submitted to the microarray assays. We applied this strategy as an attempt to avoid biased interpretation of the transcriptomic data, as the expression of many of the genes detected is cell-cycle dependent and therefore may be subject to fluctuations in asynchronous cultures.

Twenty-four DEGs, including 10 genes present in the *BRCA1* similarity cluster, were submitted to validation by qRT-PCR. We observed that one NSCL/P sample (F4243.1) exhibited a discordant expression pattern for 15 out of the 24 genes as compared to the rest of the NSCL/P group, with expression values distant at least 3

standard deviations from the mean (data not shown). Based on this observation and the fact that RNA aliquots did not correspond to the original samples hybridised on the microarray chips, we classified this sample as an outlier and excluded it from the subsequent analyses. Differences in expression values between NSCL/P and control samples were tested for statistical significance, which revealed that mRNA levels for 17 out of the 24 DEGs selected for validation were significantly different between groups ($p < 0.05$). Moreover, we confirmed the differential expression of all genes that were submitted to validation and that also pertained to the *BRCA1* similarity cluster (Table S3).

NSCL/P cells accumulate double-strand breaks (DSBs) upon exposure to H_2O_2

In the presence of DNA double-strand breaks (DSBs), the BRCA1 pathway is responsible for cell cycle checkpoint regulation, DNA damage sensor signalling, and participates in DNA repair through homologous recombination [28]. Moreover, oxidative stress plays a major role in carcinogenesis and in teratogen-induced congenital malformations through DSBs induced by reactive oxygen species [29]. Based on the microarray results, we asked if the transcriptional dysregulation of the DSB response system would result in observable cellular phenotypes in NSCL/P cells upon oxidatively generated DNA damage. By flow cytometry quantitation of γ-H2AX (phospho-histone H2AX), we assessed DSB formation in cell cultures exposed to H_2O_2, using 7 NSCL/P and 5 control samples. We observed that a significantly greater proportion of NSCL/P cells were positively stained for γ-H2AX (γ-H2AX$^+$ cells; quantitated in relation to untreated cells for each individual) after 6 and 24 hours of H_2O_2 treatment, when compared to controls ($p < 0.05$; Fig. 3A, smaller graph). In fact, NSCL/P samples exhibited a heterogeneous response to H_2O_2 treatment, which could be divided into three subgroups: some individuals exhibited a very high percentage of γ-H2AX $^+$ cells after 6 hours (subgroup I - F4243.1, F4245.1, F4293.1); a second subgroup maintained

a higher frequency of γ-H2AX$^+$ for longer (24 hours; subgroup II - F4244.1 and F4293.1); and a third subgroup exhibited a lower frequency of γ-H2AX$^+$ cells, similar to the control cells (subgroup III - F4294.1, F4311.1 and F4388.1), as shown in figure 3A and 3B. For subgroup II, we observed an accumulation of γ-H2AX$^+$ cells in G1 and early S that coincided with the still elevated γ-H2AX staining at 24 hours. This subgroup also exhibited a significant increment of sub-G1 cells in this time point, compared to controls and the other subgroups ($p < 0.005$; Fig. 3C). Therefore, subgroup II had problems to resume the cycle even at 24 hours, and these cellular responses to H_2O_2 treatment indicate that at least part of the NSCL/P cells (subgroup II) presents a defective repair of DSBs, and this leads to increased cell death.

Next, using an expanded sample of NSCL/P (n = 11) and control (n = 10) cells, we performed qRT-PCR experiments to assess if exposure to H_2O_2 affects the transcriptional behaviour of DEGs related to cell cycle control and DNA repair that were co-expressed with *BRCA1*, and also the putative upstream regulator *E2F1*. We found that after 6 hours of treatment *E2F1*, *CDC6*, *BRCA1*, *BRIP1*, *RAD51*, *RAD51AP1*and *BLM* did not exhibit significant differences in fold-expression compared to untreated samples, for both NSCL/P and control groups. These genes were transcriptionally repressed at 24 hours, possibly as a cellular response to the treatment. On the other hand, *CDC45L*, which is essential for new DNA synthesis, was more expressed in control cells after 6 hours and 24 hours when compared to NSCL/P cells ($p < 0.05$, Fig. 3D).

In light of these findings, we proposed that NSCL/P cells also feature abnormalities in the response to oxidatively generated DNA damage. Therefore, we went back to the transcriptomic data and searched for DEGs related to oxidative stress and subsequent steps of oxidatively generated DNA damage repair. We found

differential expression of several molecules directly or indirectly related to these processes, including genes involved in oxidative stress (*GSTM2* [30]*, NOX4* [31], *PTGS2* [32], *SMAD3* [33], *BRCA1* [33]), stabilisation of the replication fork (*CLSPN* [34], *TIPIN* [34]), regulation of base-excision repair (*BRCA1* [35]), and new DNA synthesis during homologous recombination repair (*CDC45L* [36]*, GINS1* [36], *MCM10* [37]; Table S4), in addition to those already found to be involved in homologous recombination (e.g. *RAD51, MSH2,* BLM; Table SI and Fig. 1C).

Murine palatal shelves express DNA damage repair genes

As the transcriptome analysis and functional assays were conducted in adult stem cells, we considered it crucial to verify if genes involved in DNA repair and cell cycle regulation are expressed in the embryonic structures important for lip and palate morphogenesis. Hence, using qRT-PCR, we first assessed expression of some of these genes, namely *Brca1*, *Brip1*, *Msh2*, *Rad51*, *Rad51ap1*, and *Blm*, in murine palatal shelves extracted at E11.5 (n = 2), E14.5 (n = 3), and E17.5 (n = 3). We observed a positive expression during palatal shelf growth, followed by a trend towards down-regulation of these genes after palate formation had been completed (E17.5), albeit with statistical significance only for *Rad51ap1* ($p < 0.05$; Fig. S1), probably due to sample size.

We also analysed *Brca1*, *Rad51*, and *E2f1* expression by whole-mount RNA *in situ* hybridisation of E10.5-E13.5 murine embryos. This revealed that *Brca1*, *Rad51* and *E2f1* transcripts are expressed in the mesenchyme of all the facial primordia (Fig. 4A-C and data not shown). Analysis of E11.5 embryos showed that all 3 genes are expressed in similar domains within the mesenchyme of the maxillary primordia, lateral and medial nasal processes which contribute to the development of the upper lip. Low levels of expression were also detected in the ectoderm at the point of fusion. At E12.5

and E13.5, *Brca1*, *Rad51* and *E2f1* transcripts were clearly co-expressed in the mesenchyme and in some regions of the ectoderm within the developing palatal shelves which arise in the maxillary primordia (Fig. 4E-F and data not shown). Negative sense hybridisation control showed no staining in the facial primordia (Figure S3).

Discussion

Orofacial morphogenesis is dependent upon tightly regulated spatio-temporal patterns of cell migration, proliferation, transdifferentiation, and apoptosis [12,13,14]. Dysregulation of biological pathways orchestrating these processes is thus presumed to play a role in the pathogenesis of NSCL/P. Importantly, the identification of these pathways and how they interact with environmental agents will not only provide insight into the molecular basis of NSCL/P, but will also enable the development of more effective preventive strategies.

Using dental pulp stem cells from NSCL/P individuals, we identified a dysregulated transcriptional network mainly associated with response to DNA damage and cell cycle control. The functional interaction network assembled with the DEGs had a central node occupied by tumour suppressor *BRCA1* which, in combination with other key genes, plays a pivotal role in the cellular response to DNA damage and cell cycle control [38]. We confirmed that many genes found to be functionally connected to *BRCA1* in this network not only had a similar expression pattern, but also have well-established roles in the aforementioned cellular functions (e.g. *CDC6*, *CDC25A*, *MSH2*, *BLM, RAD51* [39,40]). E2F1, the putative upstream regulator identified for this dysregulation block, is a transcription factor that acts in conjunction with its repressor pRB (encoded by *RB1*) and is responsible for up-regulating a variety of genes necessary for the transition from G1 to S in the cell cycle, being essential for cell cycle

progression and DNA damage response [41,42]. Also, it has been shown that E2F1 may play a role during murine palatogenesis [43]. Since *E2F1* and *RB1* were not differentially expressed in NSCL/P cells, we believe that disturbances at the protein level (i.e. affecting protein function but not necessarily the expression) could be responsible for altering the activity of these regulators, resulting in the expression patterns detected here.

The cellular functions attributed to the dysregulated gene network found here are strongly associated with tumourigenesis and risk of cancer [44]. Notably, mutations in BRCA1 have been implicated in risk of hereditary cancers, such as breast, ovarian, pancreatic, and prostate cancer [45], whereas reduced levels of BRCA1 mRNA and protein have also been associated with sporadic tumours [46,47,48]. Additionally, other types of cancer have been ascribed to alterations in many DEGs detected in our analysis, such as *CDC6*, *BLM*, *RAD51*, and *MSH2* [39,49,50,51]. These observations are in agreement with the proposed hypothesis of aetiological overlap between cancer and NSCL/P [15,16,17,18]

We showed that transcriptional dysregulation of BRCA1 and its co-operators is associated with an accumulation of DSBs in NSCL/P cells, compared to controls. NSCL/P samples exhibited a heterogeneous behaviour, in which we observed three cellular phenotypes: individuals with increased DSB formation but efficient repair (subgroup I); those with increased DSB formation and deficient repair (subgroup II); and those with a similar pattern to the one observed among controls (subgroup III), which did not exhibit significant changes in the DSB profile. To better understand this variation, the mechanism by which H_2O_2 induces DSBs has to be taken into account. One possibility is that H_2O_2 generates the hydroxyl free radical $OH^{\cdot}$, a highly reactive oxygen species that afterwards induces formation of DNA single-strand breaks (SSBs)

which, in turn, result in DSBs upon collapse of the replication fork during the S phase of the cell cycle. In this situation, cells accumulate one-ended DSBs, and repair likely occurs by homologous recombination [52,53,54,55]. The detection of DEGs associated with processes presumed to prevent the oxidative generation of this type of DNA lesion and ensure its repair (i.e., oxidative stress, SSB repair, stabilisation of the replication fork, and DSB sensor signalling and repair via homologous recombination) explain the accumulation of DSBs observed in some of the NSCL/P samples. Therefore, we propose that the concurrent accumulation of G1/early S cells and elevated DSB signals at 24 hours, observed in NSCL/P subgroup II, reflects the inability of these cells to undergo homology-directed DSB repair. This is supported by the observed H_2O_2-dependent increment in apoptotic cells in this subgroup, and by the fact that NSCL/P cells that are able to repair the DSBs (subgroup I) did not display such accumulation in G1 and early S after 24 hours of treatment, being able to progress past early S and further into later stages of the cell cycle. Accordingly, controls and NSCL/P subgroup III did not show appreciable changes in DSB signal nor cell cycle distribution in the presence of H_2O_2, possibly because these cells do not possess alterations in anti-oxidative response and repair of DNA lesions. The finding that *CDC45L* fails to undergo up-regulation in NSCL/P cells exposed to H_2O_2 further suggests that deficiency in this type of DNA repair could play a role in the manifestation of the observed NSCL/P cellular phenotypes, as this DEG is involved in new DNA synthesis during homologous recombination [36] repair. H_2O_2 can also directly induce DSBs irrespective of cell cycle phase; consequently, deficiency of other repair pathways, such as non-homologous end-joining in G1 and G2 [56], would also be important. Yet, this does not seem to be the case, as the contribution of H_2O_2 in direct DSB formation is low [57]. Unexpectedly, in H_2O_2-treated cells we did not detect expression differences for the other DSB repair-related genes neither at 6 or 24 hours. This may have occurred because their transcriptional modulation could be required before the time points investigated in our experiments; for example, BRCA1 is known to act as an

early detector and mediator in response to DSBs, and RAD51 is required for homology searching before DNA repair takes place [52,58]. Nevertheless, we were able to confirm that NSCL/P cells display an impaired response to DNA damage, and the results suggest that oxidatively-generated DSBs may play an important role in this mechanism, providing a possible connection between this type of genotoxic insult and the aetiology of NSCL/P. Furthermore, the variation in response to H_2O_2 observed for the NSCL/P samples is in agreement with the genetic heterogeneity associated with the disease, indicating that these alterations are present in only some of the NSCL/P cases.

Previous research has reported a positive association between occurrence of NSCL/P and oxidative stress-initiating environmental factors [59]. Oxidative generation of DSBs and other types of DNA lesions have been reported to arise from many cleft-related environmental factors, such as maternal exposure to alcohol, nicotine, phenytoin, and valproic acid [59,60,61,62,63,64,65,66,67]. Among these, valproic acid has been reported to down-regulate homologous recombination DNA repair genes (e.g. *BRCA1*, *RAD51*, *BLM*) by decreasing E2F1 recruitment to its target promoters [68], which supports the hypothesis that this transcription factor could be at least in part responsible for dysregulating downstream genes through dysfunctional protein activity in the NSCL/P cells, as previously discussed. Moreover, folate deficiency has been related to orofacial clefts [69], and folate is essential for DNA biosynthesis, replication and repair [70,71]; accordingly, it has been shown that genes related to DNA repair and cell cycle regulation, many of which were detected in our analysis, are differentially expressed in response to this molecule [72]. Importantly, if orofacial clefts are related to oxidative/genotoxic factors, they must act during embryonic development and affect the structures responsible for lip and palate formation, where abnormal cellular responses

to these environmental insults are expected to play an important role in shaping susceptibility to NSCL/P.

If appropriate transcriptional regulation of the genes detected by our analyses is critical for normal lip and palate development, spatially and temporally co-ordinated expression of these genes would be expected in the embryonic precursors of these structures. Indeed, RNA *in situ* hybridisation assays using mouse embryos confirmed that *Brca1*, *Rad51* and *E2f1* are co-expressed within the mesenchyme of the facial primordia that contribute to the lip and the developing palatal shelves. *Brca1*, *Rad51* and *E2f1* were also detected in small domains within the ectoderm where they may act together with *Irf6* which is expressed throughout the facial epithelium [73,74]. Therefore, *Brca1*, *Rad51* and *E2f1* are expressed during the critical stages of facial morphogenesis and dysregulation would be expected to impact on facial development. Their co-expression strongly supports the possibility that they function together as a molecular hub involved in facial growth and development. Another fact that supports this idea is that stem cells have increased DNA repair compared to differentiated cells [75]; thus, the decreased expression of these genes in NSCL/P cells may reflect problems during development. Furthermore, DNA repair and cell cycle-related genes are progressively down-regulated during orofacial morphogenesis, as suggested by the qRT-PCR experiments during palatogenesis here, and as reported in other works investigating the growth and fusion of facial prominences, and migrating neural crest cells [76,77]. If these cellular systems are more active prior to differentiation of the craniofacial tissue, they must be important in a context of intense cellular proliferation or migration during the establishment, growth, and fusion of the embryonic facial structures, before differentiation takes place. Consequently, these embryonic structures must be more susceptible to the action of environmental DNA-damaging agents, and the effects would be expected to be exacerbated if dysregulation of the biological DNA

repair processes revealed by the transcriptomic and functional assays here are taken into account. Therefore, the combined effects of transcriptional dysregulation and of environmental factors must be critical in a tissue-and time-specific manner, which can explain why orofacial clefts have not been observed in knockout animal models for pivotal DNA repair genes, including *E2F1*. Under these circumstances, we hypothesise that the inability to appropriately deal with DNA damage would result in disturbances in cell proliferation and/or lead to apoptosis, disrupting lip and palate morphogenesis.

In conclusion, we report here that gene networks governing cellular defences against DNA damage may play a role in the aetiology of NSCL/P, in accordance with the idea that orofacial clefts and cancer may have overlapping aetiologies. The identification of E2F1 as a putative regulator behind the expression profiles detected in this work reinforces the existence of one or a few upstream elements underlying dysregulation in NSCL/P cells. It is not yet possible to determine if the few NSCL/P-associated variants previously identified through GWAS [3,4,5,6] can be accountable for dysregulating entire cellular functions as seen here; additionally, none of these variants are mapped to any of the DEGs identified in this work. Therefore, we speculate that alterations in a few unidentified upstream genetic or epigenetic regulators, combined with the effects of disease-associated variants, could be responsible for disturbances in regulatory or signalling events, such as those modulating activity of transcription factors like E2F1, or directly regulating entire pathways. Importantly, we do not presuppose that dysregulation of the biological processes described here is fully responsible for the pathogenesis of NSCL/P; instead, we believe that they are part of a variety of mechanisms, such as perturbations in extracellular matrix biology [8,9], ultimately impairing orofacial morphogenesis. If regulatory anomalies are behind these disturbances, future research must focus on identifying such underlying genetic or epigenetic alterations that, upon interaction with environmental factors, result in cleft lip and palate. Consequently, a better

understanding of the impact of these environmental agents, particularly those with genotoxic properties, will enable the development of preventive strategies in the future.

Material and Methods

Ethics Statement

Ethical approval to extract stem cells from the dental pulp of deciduous teeth was obtained from the Biosciences Institute Research Ethics Committee (Protocol 037/2005) in the University of São Paulo. Samples were included only after signed informed consent by the parents or legal guardians. Those who declined to participate or otherwise did not participate were not disadvantaged in any way by not participating in the study. Care and use of mice were in compliance with the animal welfare guidelines approved by the Institute for Biosciences' Animal Care and Use Committee and the King's College Research Ethics Committee.

Cell cultures

Deciduous teeth were non-invasively obtained from children in exfoliation period. Specimens were kept in DMEM/High Glucose supplemented with 1% penicillin-streptomycin solution (Life Technologies), and taken to the laboratory to be processed. Control teeth were obtained from donors attending odontopaediatric clinics in São Paulo, Brazil, while NSCL/P teeth were obtained from patients enrolled for surgical treatment at SOBRAPAR Institute, Campinas, Brazil. We considered an individual to be affected by NSCL/P if no malformations other than clefting of the upper lip with or without cleft palate were present. RNA extracted from cell cultures derived from a total

of 6 controls and 7 NSCL/P patients was used for microarray assays and quantitative real-time PCR. The same 7 NSCL/P patients and 2 novel control cell cultures in addition to 3 of the 6 aforementioned control samples were submitted to flow cytometry quantitation of γ-H2AX. Additional 4 NSCL/P and 6 control samples were used for qRT-PCR during exposure to H_2O_2 (See Table S5 for more details regarding the samples).

Dental pulp stem cell cultures were established according to previously published protocols. The primary culture establishment protocols used in our laboratory are reproducible and consistent with regard to the immunophenotype and differentiation potential of the cell populations [9,78]. Cells were cultured in DMEM-F12 (Life Technologies) supplemented with 15% Foetal Bovine Serum (HyClone), 1% Non-essential aminoacids solution (Life Technologies), 1% penicillin-streptomycin solution (Life Technologies), in a humidified incubator at 37^{o}C and 5% CO_2. For storage, cells were frozen in medium containing 90% FBS and 10% DMSO (LGC Biotecnologia). For RNA extraction, frozen cells were thawed and grown until 80% confluent in a 75 cm^2 culture flask. Microarray, qRT-PCR and flow cytometry experiments were conducted using cells between the 4th and 8th passage. During routine culture, cell populations exhibited spindle-shaped morphology and did not show significant morphological changes or cell death. We used asynchronous cells in all experiments.

In order to ensure that the transcriptional profiles were not biased by proliferative differences between controls and NSCL/P cells, 3 NSCL/P (F4243.1; F4244.1; F4293.1) and 3 control (F4217.1; F6119.1; F6032.1) samples were randomly chosen for proliferation assays. A total of 10^4 cells/cm^2 were seeded into 12-well plates (Corning). The following day, medium was changed and cells were harvested at days 0, 2, 3, 4, and 5 post-seeding, fixed in 1% paraformaldehyde, and counted using a flow cytometer (Guava). The results did not reveal significant differences between NSCL/P

and controls (repeated measures two-way ANOVA, no interaction between factors; $F = 0.3701$, $p > 0.05$; Fig. S2).

RNA extraction and microarray hybridisation

Total RNA isolation was performed with NucleoSpin RNA II kit (Macherey-Nagel), following manufacturer's recommendations. RNA quality and concentration were assessed using Nanodrop 1000 and agarose gel electrophoresis. Only RNA samples with absorbance ratio 260/280 > 1.8, preserved rRNA ratio (28S/18S) and no signs of degradation were used.

Expression measurements were performed using the Affymetrix Human Gene 1.0 ST array, which interrogates 28,869 transcripts, followed by RNA labelling and hybridisation protocols as recommended by the manufacturer. After array scanning, quality control was performed with the GCOS software (Affymetrix) according to the manufacturer's recommendations. Raw gene expression data are available at http://www.ncbi.nlm.nih.gov/geo/, under accession code GSE42589.

Microarray data processing and mining

Gene expression values were obtained using the three-step Robust Multi-array Average (RMA) pre-processing method, implemented in the Affy package in R/Bioconductor [79]. DEGs were acquired using two algorithms: SAM and Rank Products both included in the MeV (MultiExperiment Viewer) software. SAM is a t-test based method in which mean and variance are taken into account in DEG selection [80]. In contrast, Rank Products is a ranking-based method, which enables the identification of consistent differences, even if only in a subgroup of samples under analysis [81]. Genes selected by Rank Products do not necessarily exhibit

homogeneous expression levels within test and control groups, and therefore, this analysis is suitable for detecting differential expression in complex diseases [9,82]. Due to the complexity of the disease studied in this work, we decided to use both approaches in order to interrogate genes that are altered in all affected individuals analysed as well as genes altered in only a subgroup of them. Since SAM is a more conservative method, its p-value threshold was set at 0.05 while Rank Products' was set at 0.01. Both were adjusted for multiple testing with the FDR (False Discovery Rate) method [83]. As the fold change calculation differs between the SAM and Rank Products methods, we calculated it for each gene by subtracting the average of the (log) control values from the average of the (log) case values (Avg(cases)-Avg(controls)).

Transcriptome analysis

We performed functional annotation and network analysis using IPA (http://www.ingenuity.com). We used the following parameters: Molecules per Network = 35; Networks per Analysis = 25; direct relationships only; "Ingenuity Expert Information" and "Ingenuity Supported Third Party Information" (including "miRNA-mRNA interactions", "protein-protein interactions", and "additional information") data sources.

Supervised clustering was performed using EXPANDER (EXpression Analyzer and DisplayER - http://acgt.cs.tau.ac.il/expander/).We selected the probe matching the *BRCA1* gene and set "expected cluster size" to 30. Gene Ontology (GO) enrichment analysis was executed with TANGO (Tool for ANalysis of GO enrichment), with the whole genome as the background set, and bootstrap-adjusted p-value = 0.001. Transcription factor binding site enrichment analysis was carried out using PRIMA (PRomoter Integration in Microarray Analysis), avoiding coding regions, with hit range

between –1000 and +200, all genes as background, and Bonferroni-adjusted p-value threshold < 0.05. These tools are also available in the EXPANDER software.

FANTOM4 (http://fantom.gsc.riken.jp/4/) was used to validate transcription factor-gene interactions. The FANTOM4 database contains transcriptomic and deep-CAGE information of differentiating THP-1 cell lines, as well as other published data, such as ChIP-chip [84]. We searched only for ChIP-chip experimental data, at t= 0 hours of differentiation.

qRT-PCR assays performed on cell cultures

Two micrograms of total RNA extracted from each cell culture were converted into cDNA using Superscript II, according to the manufacturer's recommendations. qRT-PCR reactions were performed in duplicates with final volume of 25 µL, using 20 ng cDNA, 2X SYBR Green PCR Master Mix, and 50 nM – 200 nM of each primer. Fluorescence was detected using ABI Prism 7500 Sequence Detection System, under standard temperature protocol. Primer pairs were designed with Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/; primer sequences in Table S6), and their amplification efficiencies (E) was determined by serial cDNA dilutions expressed in $\log_{10}$ in which $E = 10^{-1/slope}$. Expression of target genes was assessed relative to a calibrator cDNA (ΔCt). Finally, GeNorm v3.4 [85] was used to determine the most stable endogenous controls (among *GAPDH*, *HPRT1*, *SDHA*, *HMBS*), and calculate normalisation factors for each sample. The final expression values were determined based on a previous method [86], dividing $E^{-\Delta Ct}$ by the corresponding normalisation factor. To compare the expression between control and NSCL/P groups, we applied a Student's t-test with Welch's correction. Primers and reagents were supplied by Life Technologies.

DSB quantitation and assessment of cell cycle distribution

NSCL/P and control cells *were* seeded into 6-well plates (Corning), in duplicates (10^4 cells/cm^2). The next day, cells were rinsed with PBS, and the culture medium was replaced by medium containing 100uM of freshly diluted H_2O_2 (Merck), followed by incubation for 6 and 24 hours in the dark, at 37ºC/5% CO_2, to stimulate DSB formation. Cells were harvested by trypsin incubation and fixed in 4% paraformaldehyde on ice for 15 minutes, followed by fixation in 70% ethanol overnight at -20ºC. DSB quantification was performed based on a previously published protocol [87]. After complete fixation, cells were double-stained with PI (Propidium Iodide) and anti-γ-H2AX (Anti-phospho-H2A.X Ser139 clone JBW301 FITC conjugate, Millipore), in order to ascertain cell cycle distribution and DSB formation, respectively. Appropriate calibrators were applied for each individual (unstained sample; PI-stained- and anti-γ-H2AX -stained-only samples), and at least 5000 events were acquired. Data were analysed with Guava Express PRO software (Millipore) and gated to remove debris and cell clumps. To sort cells positively stained for γ-H2AX, we established a threshold using untreated cells of the same individual, below which ~98% of the cells expressed γ-H2AX (intrinsic DSB formation); cells exposed to H_2O_2 and located above this threshold were considered positive for γ-H2AX. Sub-G1 events were quantified to estimate the number of apoptotic cells. H2AX profiles were compared using a Student's t-test with Welch's correction, whilst differences in sub-G1 cells were assessed using two-way ANOVA (subgroups x treatment) with Bonferroni post-tests for multiple comparisons.

qRT-PCR and RNA in situ hybridisation studies in mouse embryos

Palatal shelves were dissected from CD1 mouse embryos at different stages of development: E11.5 (initial growth), E14.5 (period of fusion), and E17.5 (after complete

formation), and divided into pools which were submitted to RNA extraction and conversion to cDNA. Gene expression levels were assessed by qRT-PCR, using appropriate mouse endogenous controls for normalisation (*B2m*, *Tbp*, *Tubb5* and *Ywhaz*; primer sequences in Table S6). Procedures were carried out as previously described. One-way ANOVA with Bonferroni post-tests was applied to compare mean expression values.

Whole-mount *in situ* hybridisation studies were carried out to E10.5, E11.5, E12.5 and E13.5 CD1 strain mouse embryos as previously described [88]. At E12.5 and E13.5 the palatal shelves were isolated to increase probe penetration. The embryonic tissues were treated with 10 µg/mL proteinase K for 20 (E10.5), 30 (E11.5), 45 (E12.5) and 60 minutes (E13.5). Whole-mounted embryos were fixed, embedded in 20% gelatin and were vibratome-sectioned at 40 µm. 900bp-1kb cDNA templates for ribroprobe synthesis were generated by PCR using the following primers: *Brca1* (5'GTCCTCGGCGCTTGGAAGTACG3', 5'AACGACAGGCAGGTTCCCAGC3'), *Rad51* (5'GTGAGGATTTGGCGGGATTTCC3', 5'CACTACTCAGGGCGGGGAGAGC3'); *E2f1* (5'CGCTGGTAGCAGTGGGCCAT3', 5'ACCCCACGAGGCCCTTGACT3') and were cloned in the pCRII-TOPO Vector. Antisense ribroprobe transcripts were synthesised with either T7 (*E2f1*) or Sp6 (*Bcra1* and *Rad51*) RNA polymerases.

Acknowledgements

The authors would like to express their gratitude to: Mrs. Constância Urbani, for secretarial assistance; Vanessa Sato and Wagner Falciano, for financial management; Prof. Luis Eduardo Soares Netto, PhD and Erika Yeh, PhD, for their suggestions

regarding the manuscript; Angela May Suzuki, and André Veloso Rueda, MSc, for their assistance with flow cytometry and statistical analyses, respectively. Finally, Mrs. Regina Bueno for editing the figures.

References

1. *Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC (2009) Cleft lip and palate. Lancet 374: 1773-1785.*
2. *Brito LA, Cruz LA, Rocha KM, Barbara LK, Silva CB, et al. (2011) Genetic contribution for non-syndromic cleft lip with or without cleft palate (NS CL/P) in different regions of Brazil and implications for association studies. Am J Med Genet A 155A: 1581-1587.*
3. *Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, et al. (2009) Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 41: 473-477.*
4. *Grant SF, Wang K, Zhang H, Glaberson W, Annaiah K, et al. (2009) A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J Pediatr 155: 909-913.*
5. *Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, et al. (2009) Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet 42: 24-26.*
6. *Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, et al. (2009) A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet 42: 525-529.*
7. *Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. Nat Rev Genet 10: 184-194.*
8. *Baroni T, Bellucci C, Lilli C, Pezzetti F, Carinci F, et al. (2010) Human cleft lip and palate fibroblasts and normal nicotine-treated fibroblasts show altered in vitro*

expressions of genes related to molecular signaling pathways and extracellular matrix metabolism. J Cell Physiol 222: 748-756.

9. Bueno DF, Sunaga DY, Kobayashi GS, Aguena M, Raposo-Amaral CE, et al. (2010) Human stem cell cultures from cleft lip/palate patients show enrichment of transcripts involved in extracellular matrix modeling by comparison to controls. Stem Cell Rev 7: 446-457.

10. Marinucci L, Balloni S, Bodo M, Carinci F, Pezzetti F, et al. (2009) Patterns of some extracellular matrix gene expression are similar in cells from cleft lip-palate patients and in human palatal fibroblasts exposed to diazepam in culture. Toxicology 257: 10-16.

11. Pezzetti F, Carinci F, Palmieri A, Vizzotto L, Moscheni C, et al. (2009) Diphenylhydantoin plays a role in gene expression related to cytoskeleton and protein adhesion in human normal palate fibroblasts. Pathology 41: 261-268.

12. Jiang R, Bush JO, Lidral AC (2006) Development of the upper lip: morphogenetic and molecular mechanisms. Dev Dyn 235: 1152-1166.

13. Greene RM, Pisano MM (2010) Palate morphogenesis: Current understanding and future directions. Birth Defects Res C Embryo Today 90: 133-154.

14. Yu W, Serrano M, Miguel SS, Ruest LB, Svoboda KK (2009) Cleft lip and palate genetics and application in early embryological development. Indian J Plast Surg 42 Suppl: S35-50.

15. Zhu JL, Basso O, Hasle H, Winther JF, Olsen JH, et al. (2002) Do parents of children with congenital malformations have a higher cancer risk? A nationwide study in Denmark. Br J Cancer 87: 524-528.

16. Taioli E, Ragin C, Robertson L, Linkov F, Thurman NE, et al. (2010) Cleft lip and palate in family members of cancer survivors. Cancer Invest 28: 958-962.

17. Bille C, Winther JF, Bautz A, Murray JC, Olsen J, et al. (2005) Cancer risk in persons with oral cleft--a population-based study of 8,093 cases. Am J Epidemiol 161: 1047-1055.

18. *Vieira AR, Khaliq S, Lace B (2012) Risk of cancer in relatives of children born with isolated cleft lip and palate. Am J Med Genet A 158A: 1503-1504.*

19. *Rinne T, Brunner HG, van Bokhoven H (2007) p63-associated disorders. Cell Cycle 6: 262-268.*

20. *Restivo G, Nguyen BC, Dziunycz P, Ristorcelli E, Ryan RJ, et al. (2011) IRF6 is a mediator of Notch pro-differentiation and tumour suppressive function in keratinocytes. EMBO J 30: 4571-4585.*

21. *Letra A, Bjork B, Cooper ME, Szabo-Rogers H, Deleyiannis FW, et al. (2012) Association of AXIN2 with non-syndromic oral clefts in multiple populations. J Dent Res 91: 473-478.*

22. *Krenzlin H, Demuth I, Salewsky B, Wessendorf P, Weidele K, et al. (2012) DNA damage in Nijmegen Breakage Syndrome cells leads to PARP hyperactivation and increased oxidative stress. PLoS Genet 8: e1002557.*

23. *Ferone G, Thomason HA, Antonini D, De Rosa L, Hu B, et al. (2012) Mutant p63 causes defective expansion of ectodermal progenitor cells and impaired FGF signalling in AEC syndrome. EMBO Mol Med 4: 192-205.*

24. *Vogelaar IP, Figueiredo J, van Rooij IA, Simoes-Correia J, van der Post RS, et al. (2013) Identification of germline mutations in the cancer predisposing gene CDH1 in patients with orofacial clefts. Hum Mol Genet 22: 919-926.*

25. *Cordero DR, Brugmann S, Chu Y, Bajpai R, Jame M, et al. (2011) Cranial neural crest cells on the move: their roles in craniofacial development. Am J Med Genet A 155A: 270-279.*

26. *Komada M (2012) Sonic hedgehog signaling coordinates the proliferation and differentiation of neural stem/progenitor cells by regulating cell cycle kinetics during development of the neocortex. Congenit Anom (Kyoto) 52: 72-77.*

27. *Janebodin K, Horst OV, Ieronimakis N, Balasundaram G, Reesukumal K, et al. (2011) Isolation and characterization of neural crest-derived stem cells from dental pulp of neonatal mice. PLoS One 6: e27526.*

28. Yarden RI, Papa MZ (2006) BRCA1 at the crossroad of multiple cellular pathways: approaches for therapeutic interventions. Mol Cancer Ther 5: 1396-1404.

29. Wells PG, McCallum GP, Chen CS, Henderson JT, Lee CJ, et al. (2009) Oxidative stress in developmental origins of disease: teratogenesis, neurodevelopmental deficits, and cancer. Toxicol Sci 108: 4-18.

30. Sharma R, Yang Y, Sharma A, Awasthi S, Awasthi YC (2004) Antioxidant role of glutathione S-transferases: protection against oxidant toxicity and regulation of stress-mediated apoptosis. Antioxid Redox Signal 6: 289-300.

31. Weyemi U, Dupuy C (2012) The emerging role of ROS-generating NADPH oxidase NOX4 in DNA-damage responses. Mutat Res 751: 77-81.

32. Yu H, Liu Z, Zhou H, Dai W, Chen S, et al. (2012) JAK-STAT pathway modulates the roles of iNOS and COX-2 in the cytoprotection of early phase of hydrogen peroxide preconditioning against apoptosis induced by oxidative stress. Neurosci Lett 529: 166-171.

33. Li H, Sekine M, Seng S, Avraham S, Avraham HK (2009) BRCA1 interacts with Smad3 and regulates Smad3-mediated TGF-beta signaling during oxidative stress responses. PLoS One 4: e7091.

34. Uno S, Masai H (2011) Efficient expression and purification of human replication fork-stabilizing factor, Claspin, from mammalian cells: DNA-binding activity and novel protein interactions. Genes Cells 16: 842-856.

35. Saha T, Rih JK, Roy R, Ballal R, Rosen EM (2010) Transcriptional regulation of the base excision repair pathway by BRCA1. J Biol Chem 285: 19092-19105.

36. Lydeard JR, Lipkin-Moore Z, Sheu YJ, Stillman B, Burgers PM, et al. (2010) Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. Genes Dev 24: 1133-1144.

37. Maiorano D, Lutzmann M, Mechali M (2006) MCM proteins and DNA replication. Curr Opin Cell Biol 18: 130-136.

38. Linger RJ, Kruk PA (2010) BRCA1 16 years later: risk-associated BRCA1 mutations and their functional implications. FEBS J 277: 3086-3096.

39. Seifert M, Reichrath J (2006) The role of the human DNA mismatch repair gene hMSH2 in DNA repair, cell cycle control and apoptosis: implications for pathogenesis, progression and therapy of cancer. J Mol Histol 37: 301-307.

40. Yata K, Esashi F (2009) Dual role of CDKs in DNA repair: to be, or not to be. DNA Repair (Amst) 8: 6-18.

41. Biswas AK, Johnson DG (2011) Transcriptional and nontranscriptional functions of E2F1 in response to DNA damage. Cancer Res 72: 13-17.

42. Iaquinta PJ, Lees JA (2007) Life and death decisions by the E2F transcription factors. Curr Opin Cell Biol 19: 649-657.

43. Kusek JC, Greene RM, Nugent P, Pisano MM (2000) Expression of the E2F family of transcription factors during murine development. Int J Dev Biol 44: 267-277.

44. Dixon K, Kopras E (2004) Genetic alterations and DNA repair in human carcinogenesis. Semin Cancer Biol 14: 441-448.

45. Rosen EM, Fan S, Ma Y (2006) BRCA1 regulation of transcription. Cancer Lett 236: 175-185.

46. Thompson ME, Jensen RA, Obermiller PS, Page DL, Holt JT (1995) Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression. Nat Genet 9: 444-450.

47. Lee WY, Jin YT, Chang TW, Lin PW, Su IJ (1999) Immunolocalization of BRCA1 protein in normal breast tissue and sporadic invasive ductal carcinomas: a correlation with other biological parameters. Histopathology 34: 106-112.

48. Mueller CR, Roskelley CD (2003) Regulation of BRCA1 expression and its relationship to sporadic breast cancer. Breast Cancer Res 5: 45-52.

49. Calin G, Herlea V, Barbanti-Brodano G, Negrini M (1998) The coding region of the Bloom syndrome BLM gene and of the CBL proto-oncogene is mutated in

genetically unstable sporadic gastrointestinal tumors. Cancer Res 58: 3777-3781.

50. Borlado LR, Mendez J (2008) CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. Carcinogenesis 29: 237-243.

51. Gao LB, Pan XM, Li LJ, Liang WB, Zhu Y, et al. (2010) RAD51 135G/C polymorphism and breast cancer risk: a meta-analysis from 21 studies. Breast Cancer Res Treat 125: 827-835.

52. Helleday T, Lo J, van Gent DC, Engelward BP (2007) DNA double-strand break repair: from mechanistic understanding to cancer treatment. DNA Repair (Amst) 6: 923-935.

53. Llorente B, Smith CE, Symington LS (2008) Break-induced replication: what is it and what is it for? Cell Cycle 7: 859-864.

54. Pardo B, Gomez-Gonzalez B, Aguilera A (2009) DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. Cell Mol Life Sci 66: 1039-1056.

55. Langerak P, Russell P (2011) Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair. Philos Trans R Soc Lond B Biol Sci 366: 3562-3571.

56. Mao Z, Bozzella M, Seluanov A, Gorbunova V (2008) Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair (Amst) 7: 1765-1771.

57. Letavayova L, Markova E, Hermanska K, Vlckova V, Vlasakova D, et al. (2006) Relative contribution of homologous recombination and non-homologous end-joining to DNA double-strand break repair after oxidative stress in Saccharomyces cerevisiae. DNA Repair (Amst) 5: 602-610.

58. Wang Y, Cortez D, Yazdi P, Neff N, Elledge SJ, et al. (2000) BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev 14: 927-939.

59. *Bahado-Singh RO, Schenone M, Cordoba M, Shieh WS, Maulik D, et al. (2011) Male gender significantly increases risk of oxidative stress related congenital anomalies in the non-diabetic population. J Matern Fetal Neonatal Med 24: 687-691.*

60. *Nulman I, Scolnik D, Chitayat D, Farkas LD, Koren G (1997) Findings in children exposed in utero to phenytoin and carbamazepine monotherapy: independent effects of epilepsy and medications. Am J Med Genet 68: 18-24.*

61. *Winn LM, Kim PM, Nickoloff JA (2003) Oxidative stress-induced homologous recombination as a novel mechanism for phenytoin-initiated toxicity. J Pharmacol Exp Ther 306: 523-527.*

62. *Alsdorf R, Wyszynski DF (2005) Teratogenicity of sodium valproate. Expert Opin Drug Saf 4: 345-353.*

63. *DeRoo LA, Wilcox AJ, Drevon CA, Lie RT (2008) First-trimester maternal alcohol consumption and the risk of infant oral clefts in Norway: a population-based case-control study. Am J Epidemiol 168: 638-646.*

64. *Leite IC, Koifman S (2009) Oral clefts, consanguinity, parental tobacco and alcohol use: a case-control study in Rio de Janeiro, Brazil. Braz Oral Res 23: 31-37.*

65. *Sha K, Winn LM (2010) Characterization of valproic acid-initiated homologous recombination. Birth Defects Res B Dev Reprod Toxicol 89: 124-132.*

66. *Tung EW, Winn LM (2011) Valproic acid increases formation of reactive oxygen species and induces apoptosis in postimplantation embryos: a role for oxidative stress in valproic acid-induced neural tube defects. Mol Pharmacol 80: 979-987.*

67. *Park SC, Lim JY, Jeen YT, Keum B, Seo YS, et al. (2012) Ethanol-induced DNA damage and repair-related molecules in human intestinal epithelial Caco-2 cells. Mol Med Report 5: 1027-1032.*

68. *Kachhap SK, Rosmus N, Collis SJ, Kortenhorst MS, Wissing MD, et al. (2010) Downregulation of homologous recombination DNA repair genes by HDAC*

inhibition in prostate cancer is mediated through the E2F1 transcription factor. PLoS One 5: e11208.

69. Wehby GL, Murray JC (2010) Folic acid and orofacial clefts: a review of the evidence. Oral Dis 16: 11-19.

70. Joshi R, Adhikari S, Patro BS, Chattopadhyay S, Mukherjee T (2001) Free radical scavenging behavior of folic acid: evidence for possible antioxidant activity. Free Radic Biol Med 30: 1390-1399.

71. Gliszczynska-Swiglo A, Muzolf M (2007) pH-Dependent radical scavenging activity of folates. J Agric Food Chem 55: 8237-8242.

72. Bliek BJ, Steegers-Theunissen RP, Blok LJ, Santegoets LA, Lindemans J, et al. (2008) Genome-wide pathway analysis of folate-responsive genes to unravel the pathogenesis of orofacial clefting in man. Birth Defects Res A Clin Mol Teratol 82: 627-635.

73. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, et al. (2002) Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet 32: 285-289.

74. Washbourne BJ, Cox TC (2006) Expression profiles of cIRF6, cLHX6 and cLHX7 in the facial primordia suggest specific roles during primary palatogenesis. BMC Dev Biol 6: 18.

75. Rocha CR, Lerner LK, Okamoto OK, Marchetto MC, Menck CF (2012) The role of DNA repair in the pluripotency and differentiation of human stem cells. Mutat Res.

76. Feng W, Leach SM, Tipney H, Phang T, Geraci M, et al. (2009) Spatial and temporal analysis of gene expression during growth and fusion of the mouse facial prominences. PLoS One 4: e8066.

77. Albino D, Brizzolara A, Moretti S, Falugi C, Mirisola V, et al. (2011) Gene expression profiling identifies eleven DNA repair genes down-regulated during mouse neural crest cell migration. Int J Dev Biol 55: 65-72.

78. de Mendonca Costa A, Bueno DF, Martins MT, Kerkis I, Kerkis A, et al. (2008) Reconstruction of large cranial defects in nonimmunosuppressed experimental design with human dental pulp stem cells. J Craniofac Surg 19: 204-210.

79. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249-264.

80. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98: 5116-5121.

81. Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett 573: 83-92.

82. Raj A, Rifkin SA, Andersen E, van Oudenaarden A (2010) Variability in gene expression underlies incomplete penetrance. Nature 463: 913-918.

83. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) 57: 289-300.

84. Severin J, Waterhouse AM, Kawaji H, Lassmann T, van Nimwegen E, et al. (2009) FANTOM4 EdgeExpressDB: an integrated database of promoters, genes, microRNAs, expression dynamics and regulatory interactions. Genome Biol 10: R39.

85. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: RESEARCH0034.

86. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45.

87. Huang X, Darzynkiewicz Z (2006) Cytometric assessment of histone H2AX phosphorylation: a reporter of DNA damage. Methods Mol Biol 314: 73-80.

88. *Dastjerdi A, Robson L, Walker R, Hadley J, Zhang Z, et al. (2007) Tbx1 regulation of myogenic differentiation in the limb and cranial mesoderm. Dev Dyn 236: 353-363.*

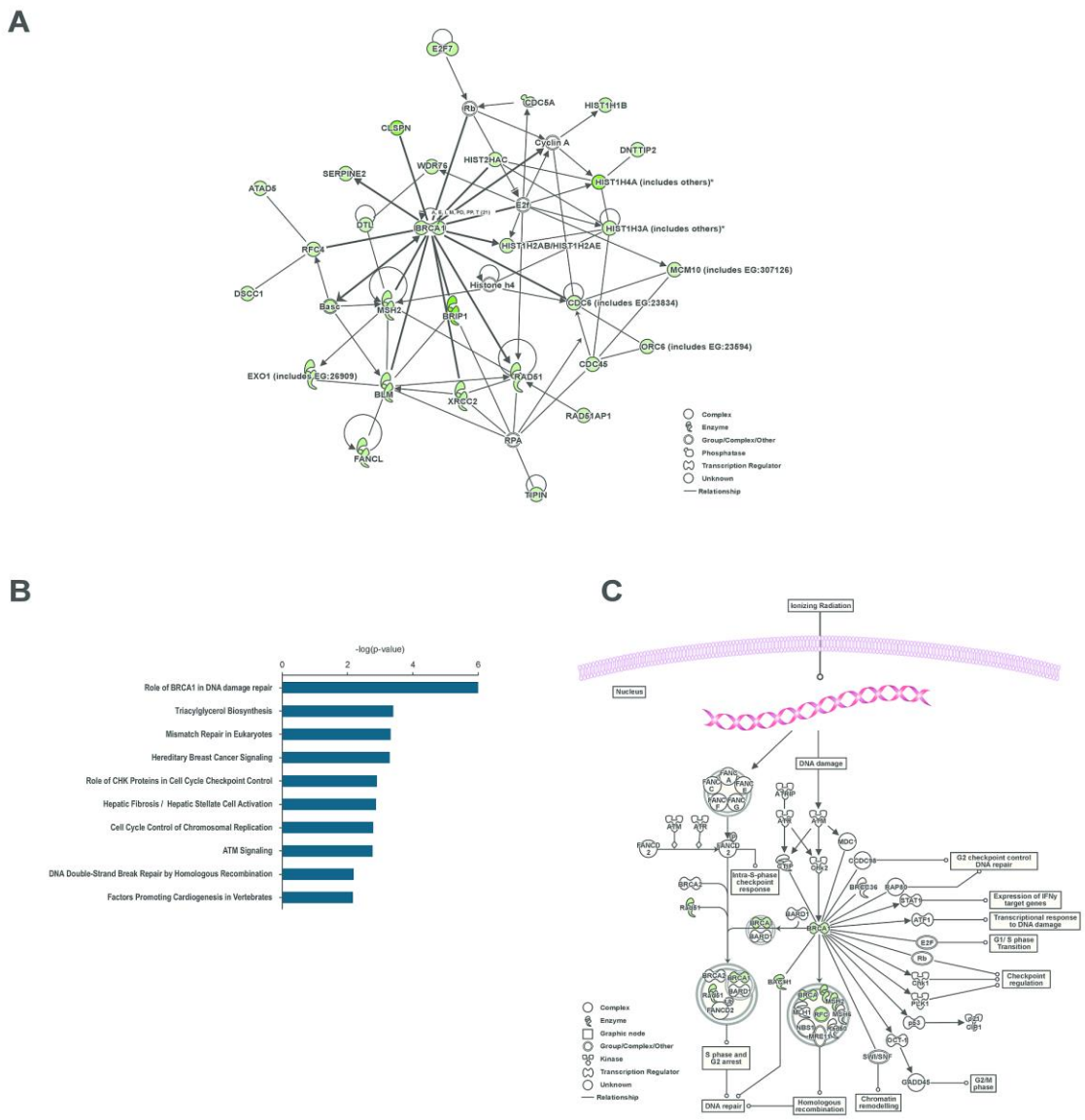


Figure 1. NSCL/P cells exhibit a gene expression profile associated with dysregulation of DNA repair and cell cycle control. (A) IPA network. DEGs were used to assemble a functional interaction map. Lines with arrowheads represent that one molecule acts on another, while regular lines represent protein interactions. Down-regulated genes are shown as green nodes, whereas genes without differential expression are shown as blank nodes. (B) Top 10 Canonical Pathways significantly enriched among the 228 DEGs (Fisher's Exact Test, p-value < 0.01). (C) 'Role of BRCA1 in DNA Damage Response' Canonical Pathway.

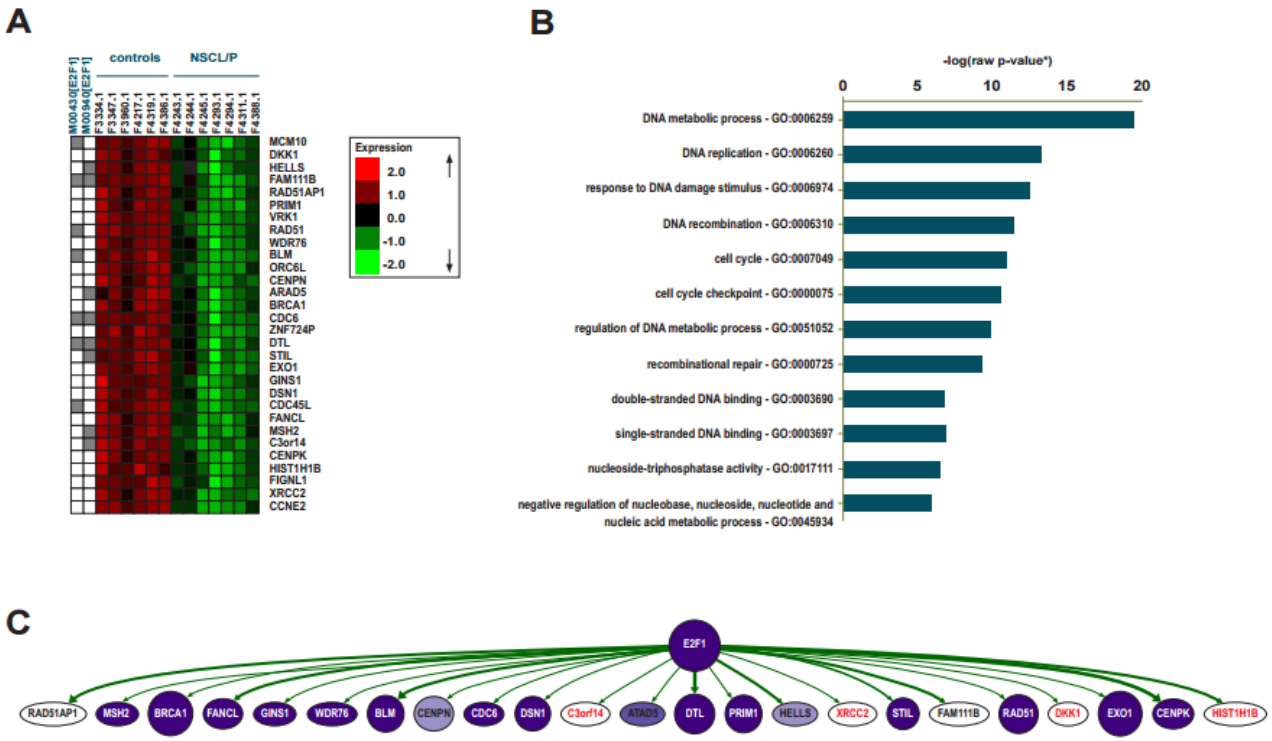


Figure 2. NSCL/P patterns of co-expression are associated with DNA repair, and suggest E2F1 as a regulator. (A) Supervised similarity cluster based on *BRCA1* expression (avg. homogeneity = 0.974). Transcription factor binding sites significantly over-represented in the cluster are marked in grey, for each motif identified (Bonferroni-adjusted p-value < 0.05). (B) GO attributes enriched in the similarity cluster and their respective representation among the 30 clustered genes, expressed in percentages ([*] Bootstrap-adjusted p-values = 0.001, raw p-values were used in the chart). (C) Analysis of transcription factor-gene interactions. ChIP-chip data from FANTOM4 were used to validate the interaction between E2F1 and 23 out of the 30 genes of the similarity cluster. The thickness of the arrows indicates how often the interaction has been experimentally detected; node colours represent the level of expression in the cell lines used to assemble the database, from low (light) to high (dark).

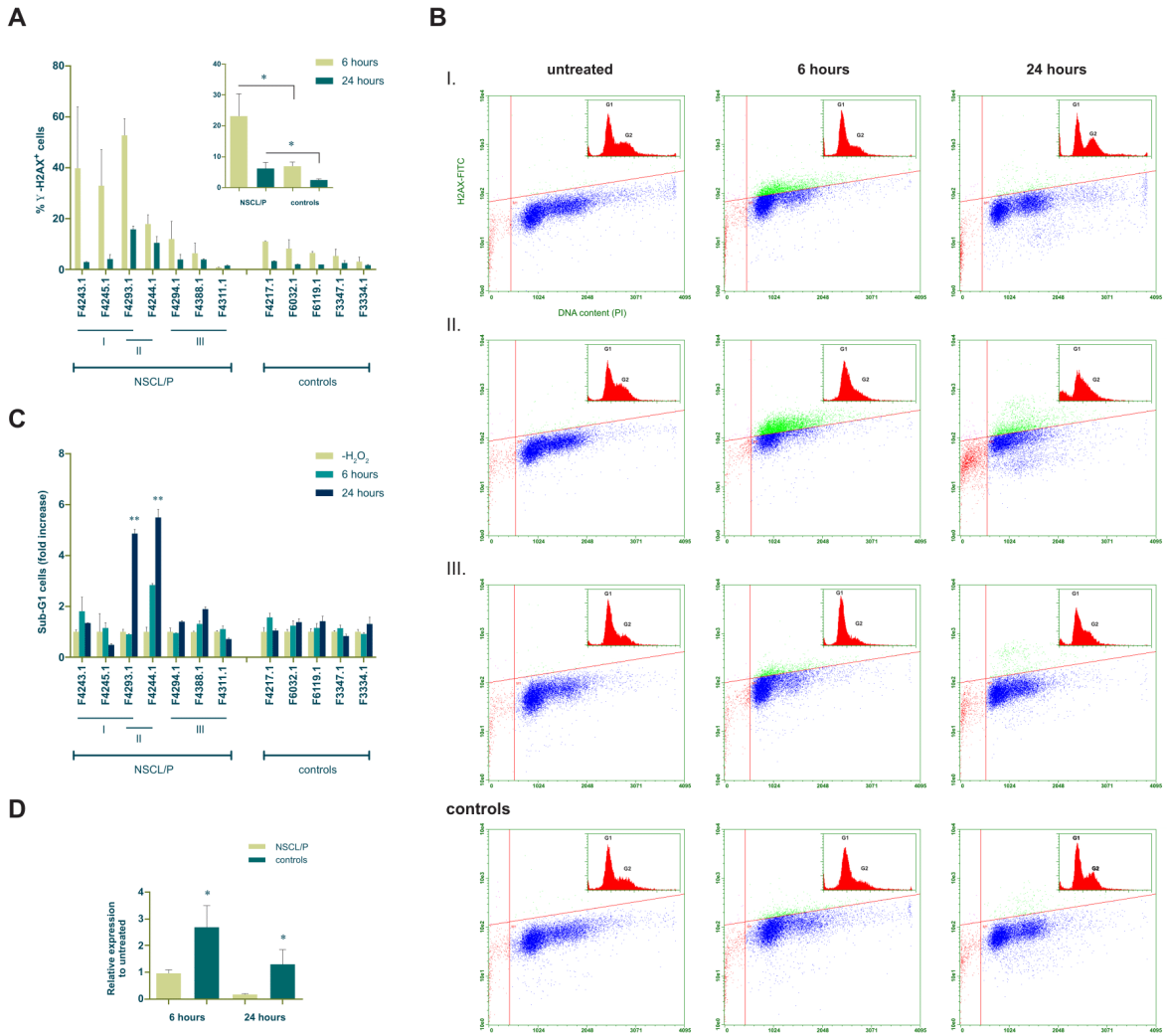


Figure 3. NSCL/P cells exhibit a heterogeneous response to H_2O_2, with distinct patterns of DSB accumulation. (A) Smaller graph - Comparison of the fraction of γ-H2AX $^{+}$ cells between NSCL/P and control cells, at 6 and 24 hours of treatment with H_2O_2. Large graph – Individual quantitation of γ-H2AX^{+} cells, revealing the subgroups (I-III) within the NSCL/P samples. (*) $p < 0.05$. (B) Representative γ-H2AX and PI profiles depicting DSB and cell cycle distribution, for each NSCL/P subgroup and controls. (C) Quantitation of sub-G1 cells, revealing a significant increment in NSCL/P subgroup II, at 24 hours of treatment, as compared to the other subgroups and the controls. (**) $F = 6.04$; $p < 0.005$. (D) Relative expression of *CDC45L* following treatment with H_2O_2, revealing a significant decrease in NSCL/P samples at both time points, by comparison to controls. (*) $p < 0.05$.

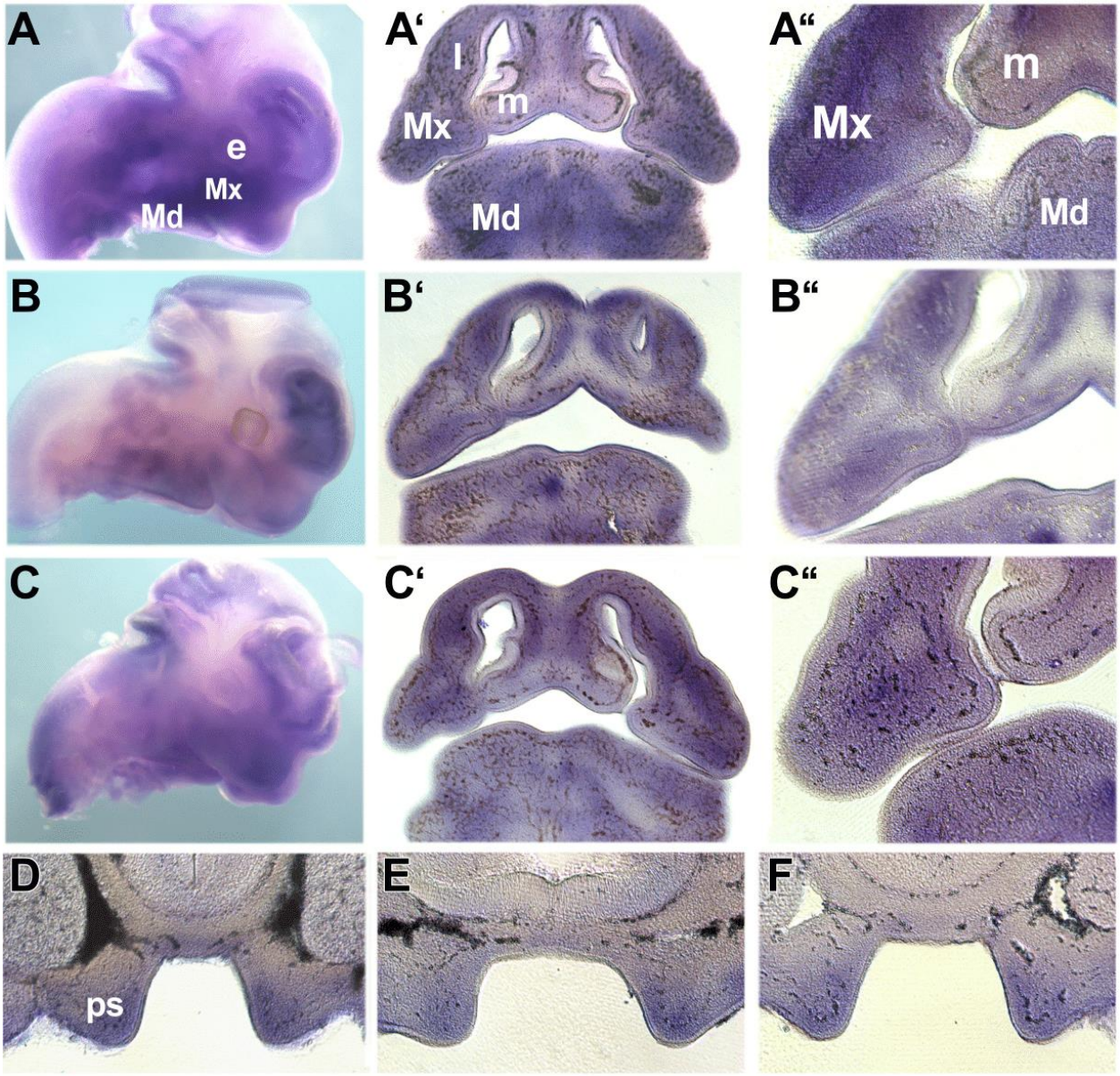


Figure 4. Expression of *Brca1*, *Rad51* and *E2f1* in the developing facial primordia. Whole-mount *in situ* hybridisation (A-F) showing the expression of *Brca1* (A, D), *Rad51* (B, E) and *E2F1* (C, F) in E11.5 (A-C) and E12.5 (D-F) embryos. Expression is indicated by the blue/purple staining. A-C are sagittal views of the developing head whilst A'-C' are frontal sections through the embryos shown in A-C. A''-C'' show high power views through the developing maxillary primordia, lateral and medial nasal processes. E-F are frontal sections through the developing E12.5 palatal shelves. *e*, eye; *l*, lateral nasal process; *m*, medial nasal process; *Md*, mandibular primordia; *Mx*, maxillary primordia; *ps*, palatal shelves.

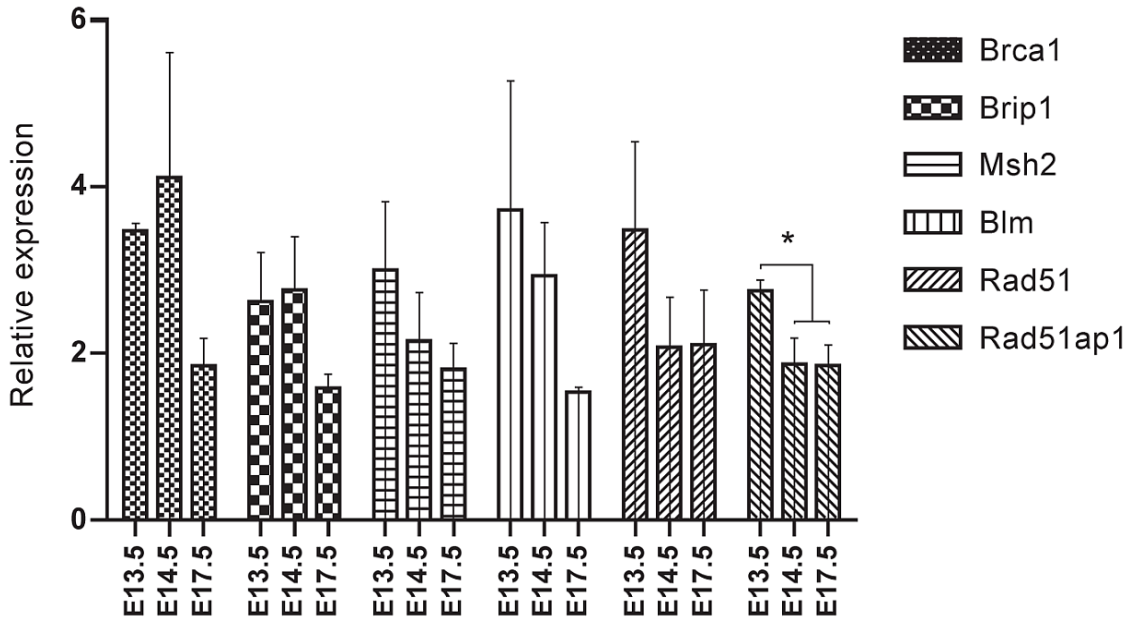


Figure S1. DNA repair genes are expressed in the developing palatal shelves. Gene expression of key DNA repair genes was assessed in murine palatal shelves at various stages of development, using qRT-PCR. (*) $p < 0.05$.

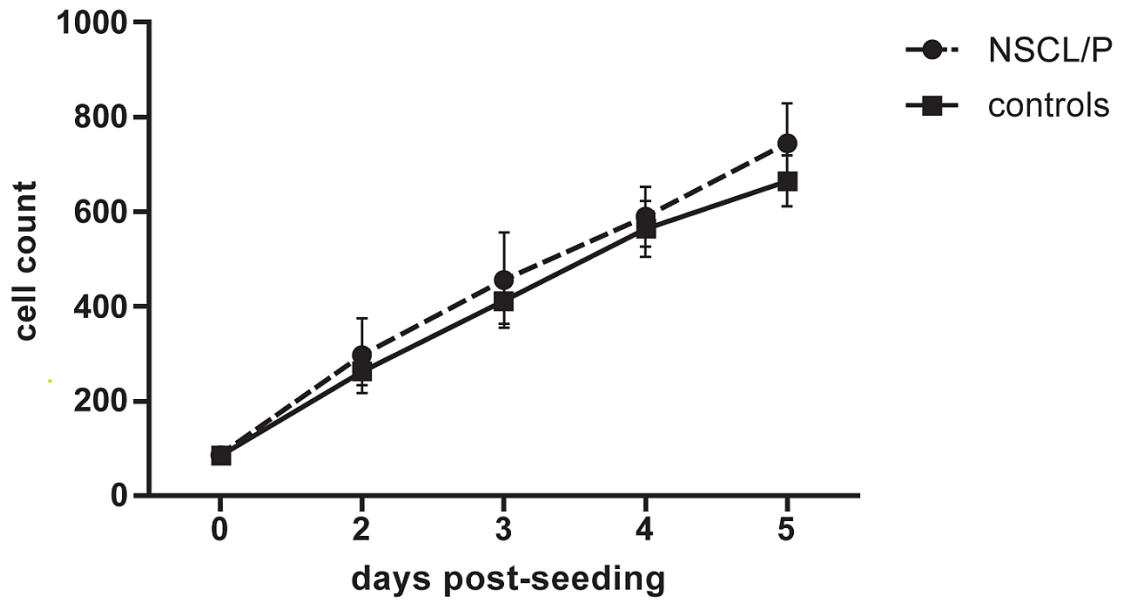


Figure S2. NSCL/P and control cells exhibit similar proliferation profiles. Proliferation assays were performed in 3 NSCL/P and 3 control cells, and revealed no significant differences (repeated measures two-way ANOVA, $p > 0.05$)

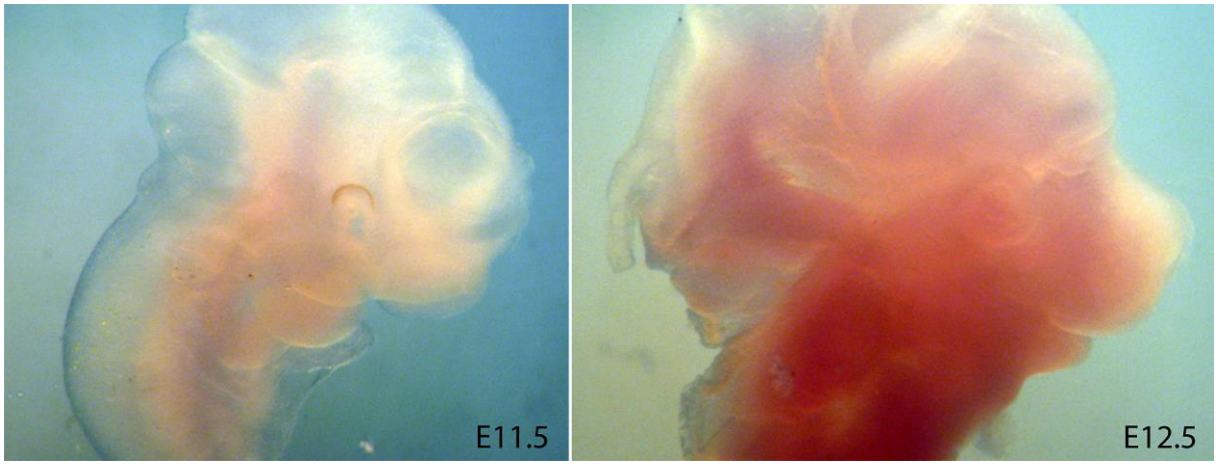


Figure S3. Negative control for the *in situ* hybridisation studies. Sagittal views of negative sense controls performed on E11.5 and E12.5 mouse embryos, showing no staining.

Table SI: List of DEGs obtained by comparing NSCL/P and control cells

Affy.I	Symb				
Entrez Gene	**Name**		**Fold change**		**P-value**
8112045 ° *	ESM1	endothelial cell-specific molecule 1	-5.330	±2.83	0.000
7922976 ° *	PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	-3.709	±2.11	0.000
8067233 ° *	PMEPA1	prostate transmembrane protein, androgen induced 1	-3.259	±2.01	0.000
8142981 ° *	PODXL	podocalyxin-like	-2.920	±1.89	0.000
8116780 ° *	DSP	desmoplakin	-2.736	±1.8	0.000
8100798 ° *	SULT1B1	sulfotransferase family, cytosolic, 1B, member 1	-2.660	±2.37	0.000
7915612 ° *	PTCH2	patched 2	-2.598	±1.51	0.000
8124527 ° *	HIST1H1B	histone cluster 1, H1b	-2.587	±1.51	0.000
7973067 ° *	PNP	purine nucleoside phosphorylase	-2.578	±1.45	0.000
8151871 ° *	CCNE2	cyclin E2	-2.577	±1.44	0.000
8091243 ° *	PCOLCE2 (includes EG:2657	procollagen C-endopeptidase enhancer 2	-2.568	±1.58	0.000
7924461 ° *			-2.567	±1.54	0.000
7939237 ° *	C11orf41	chromosome 11 open reading frame 41	-2.485	±1.39	0.000
8059376 ° *	SERPINE2	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2	-2.463	±1.56	0.000
8105267 ° *	ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	-2.440	±1.69	0.000
8045688 ° *	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	-2.436	±1.45	0.000
7952785 ° *	OPCML	opioid binding protein/cell adhesion molecule-like	-2.431	±1.67	0.000
7948902 ° *	SNHG1	small nucleolar RNA host gene 1 (non-protein coding)	-2.418	±1.49	0.000
7939215 ° *	C11orf41	chromosome 11 open reading frame 41	-2.410	±1.34	0.000
8180255 ° *			-2.404	±1.33	0.000
8180321 ° *			-2.404	±1.33	0.000
8145793 ° *	SNORD13	small nucleolar RNA, C/D box 13	-2.383	±1.45	0.000
7936968 ° *	ADAM12	ADAM metallopeptidase domain 12	-2.339	±1.67	0.000
7999754 ° *	XYLT1	xylosyltransferase I	-2.317	±1.43	0.000
7981084 ° *	SERPINA9	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 9	-2.307	±1.28	0.000
7974689 ° *	DACT1	dapper, antagonist of beta-catenin, homolog 1 (Xenopus laevis)	-2.304	±1.47	0.000
8086752 ° *			-2.279	±1.5	0.001
8124388 ° *	HIST1H3A (includes others)	histone cluster 1, H3a	-2.273	±1.29	0.001
7970793 ° *	SLC46A3	solute carrier family 46, member 3	-2.271	±1.47	0.000
7909568 ° *	DTL	denticleless E3 ubiquitin protein ligase homolog (Drosophila)	-2.268	±1.35	0.000
7897801 ° *	RNU5E-1	RNA, U5E small nuclear 1	-2.240	±1.32	0.001
7996260 ° *			-2.233	±1.63	0.000
8007071 ° *	CDC6 (includes EG:23834)	cell division cycle 6 homolog (S. cerevisiae)	-2.208	±1.31	0.001
7976812 ° *	SNORD113-4	small nucleolar RNA, C/D box 113-4	-2.195	±1.61	0.001
8154233 ° *	CD274	CD274 molecule	-2.184	±1.49	0.001
8156043 ° *	PSAT1	phosphoserine aminotransferase 1	-2.151	±1.28	0.001
7940147 ° *	FAM111B	family with sequence similarity 111, member B	-2.129	±1.24	0.001
7953218 ° *	RAD51AP1	RAD51 associated protein 1	-2.120	±1.26	0.001
8126798 °	GPR116	G protein-coupled receptor 116	-2.119	±2.36	0.000
7919642 ° *	HIST2H2AB	histone cluster 2, H2ab	-2.102	±1.2	0.002
8124391 ° *	HIST1H2AB/HIST1H2AE	histone cluster 1, H2ae	-2.098	±1.24	0.002
8077499 ° *	LINC00312	long intergenic non-protein coding RNA 312	-2.086	±1.51	0.001
7926259 ° *	MCM10 (includes EG:307126	minichromosome maintenance complex component 10	-2.078	±1.2	0.002
8160431 ° *	MIR31HG	MIR31 host gene (non-protein coding)	-2.060	±1.13	0.002
7950391 ° *	PGM2L1	phosphoglucomutase 2-like 1	-2.060	±1.17	0.002
7943158 ° *	SCARNA9	small Cajal body-specific RNA 9	-2.060	±1.27	0.002
7981978 ° *	SNORD116-15	small nucleolar RNA, C/D box 116-15	-2.054	±1.23	0.002
8035838 ° *	ZNF724P	zinc finger protein 724, pseudogene	-2.037	±1.15	0.002
8092640 ° *	RFC4	replication factor C (activator 1) 4, 37kDa	-2.022	±1.08	0.002
8080847 ° *	C3orf14	chromosome 3 open reading frame 14	-2.018	±1.11	0.002
8112376 ° *	CENPK	centromere protein K	-2.018	±1.16	0.003
7981982 ° *	SNRPN	small nuclear ribonucleoprotein polypeptide N	-2.004	±1.2	0.003
7981986 ° *	SNRPN	small nuclear ribonucleoprotein polypeptide N	-2.004	±1.2	0.003
8083876 ° *	SKIL	SKI-like oncogene	-2.002	±1.25	0.002
7927631 ° *	DKK1	dickkopf 1 homolog (Xenopus laevis)	-2.002	±1.26	0.003
7922162 ° *	SLC19A2	solute carrier family 19 (thiamine transporter), member 2	-2.001	±1.26	0.001
7998722 ° *	SNORD60	small nucleolar RNA, C/D box 60	-1.997	±1.13	0.003
7906919 °	RGS4	regulator of G-protein signaling 4	-1.989	±1.76	0.002
7929438 ° *	HELLS	helicase, lymphoid-specific	-1.976	±1.18	0.004
7981976 ° *	SNORD116-14	small nucleolar RNA, C/D box 116-14	-1.967	±1.25	0.004
8071212 ° *	CDC45	cell division cycle 45 homolog (S. cerevisiae)	-1.962	±1.08	0.004
7910997 ° *	EXO1 (includes EG:26909)	exonuclease 1	-1.952	±1.14	0.004
8061471 ° *	GINS1	GINS complex subunit 1 (Psf1 homolog)	-1.945	±1.12	0.004
8138527 ° *	STEAP1B	STEAP family member 1B	-1.939	±1.2	0.004
8030978 ° *	ZNF845	zinc finger protein 845	-1.938	±1.23	0.004
7919269 ° *	GSTM2	glutathione S-transferase mu 2 (muscle)	-1.937	±1.09	0.004
7919349 ° *	GSTM2	glutathione S-transferase mu 2 (muscle)	-1.937	±1.09	0.004
8124385 ° *	HIST1H4A (includes others)	histone cluster 1, H4a	-1.937	±1.11	0.004
7930980 °	PPAPDC1A	phosphatidic acid phosphatase type 2 domain containing 1A	-1.934	±1.35	0.003
8140534 °	SEMA3C	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	-1.933	±1.41	0.003
8004144 ° *	MIS12	MIS12, MIND kinetochore complex component, homolog (S. pombe)	-1.930	±1.05	0.004
7952339 ° *	SNORD14C	small nucleolar RNA, C/D box 14C	-1.929	±1.21	0.005
8081241 ° *	C3orf26	chromosome 3 open reading frame 26	-1.927	±1.13	0.004
7995354 ° *	ORC6 (includes EG:23594)	origin recognition complex, subunit 6	-1.923	±1.08	0.004
8127989 ° *	SNORD50B	small nucleolar RNA, C/D box 50B	-1.923	±1.1	0.005
7948904 ° *	SNORD28	small nucleolar RNA, C/D box 28	-1.921	±1.28	0.005
7986068 ° *	BLM	Bloom syndrome, RecQ helicase-like	-1.920	±1.08	0.004
7973896 ° *	GSTM2	glutathione S-transferase mu 2 (muscle)	-1.918	±1.05	0.005
7978568 ° *	GSTM2	glutathione S-transferase mu 2 (muscle)	-1.918	±1.05	0.005
7915926 ° *	STIL	SCL/TAL1 interrupting locus	-1.915	±1.15	0.004
7908072 °	LAMC2	laminin, gamma 2	-1.914	±1.34	0.004
8141150 ° *	ASNS	asparagine synthetase (glutamine-hydrolyzing)	-1.913	±1.11	0.004
7979710 ° *	PLEK2	pleckstrin 2	-1.913	±1.12	0.004
7964271 ° *	PRIM1	primase, DNA, polypeptide 1 (49kDa)	-1.909	±1.11	0.004
8015268 ° *	KRT34	keratin 34	-1.900	±1.1	0.006

8095574 ° *	DCK	deoxycytidine kinase							-1.899	±1.07	0.005
7964733 ° *	RPSAP52	ribosomal protein SA pseudogene 52							-1.899	±1.22	0.005
8025402 ° *	ANGPTL4	angiopoietin-like 4							-1.898	±1.06	0.004
8022674 ° *	CDH2	cadherin 2, type 1, N-cadherin (neuronal)	-1.895	±1.26	0.003	7990345 °	SEMA7A				
semaphorin 7A, GPI membrane anchor (John Milton Hagen blood group)			-1.883	±1.37	0.002	8142975 °	mir-29	microRNA			
29a	-1.883	±1.31	0.005								
7981990 °									-1.881	±1.27	0.005
8102787 ° *									-1.878	±1.05	0.006
8041867 ° *	MSH2	mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)							-1.872	±1.05	0.006
7965094 ° *	E2F7	E2F transcription factor 7							-1.868	±1.08	0.005
8052382 ° *	FANCL	Fanconi anemia, complementation group L							-1.860	±1.06	0.007
8079153 ° *	ABHD5	abhydrolase domain containing 5							-1.853	±1.02	0.007
7982792 ° *	RAD51	RAD51 homolog (S. cerevisiae)							-1.851	±1.03	0.007
7898375 ° *	GSTM2	glutathione S-transferase mu 2 (muscle)							-1.851	±1.03	0.007
7898411 ° *	GSTM2	glutathione S-transferase mu 2 (muscle)							-1.851	±1.03	0.007
7912800 ° *	GSTM2	glutathione S-transferase mu 2 (muscle)							-1.851	±1.03	0.007
7912850 ° *	GSTM2	glutathione S-transferase mu 2 (muscle)							-1.851	±1.03	0.007
7919576 ° *	GSTM2	glutathione S-transferase mu 2 (muscle)							-1.851	±1.03	0.007
7974882 ° *	SYT16	synaptotagmin XVI							-1.839	±1.13	0.005
8144036 ° *	XRCC2	X-ray repair complementing defective repair in Chinese hamster cells 2							-1.832	±1.04	0.007
7976621 ° *	VRK1	vaccinia related kinase 1							-1.831	±1	0.007
8139632 ° *	FIGNL1	fidgetin-like 1							-1.824	±1.06	0.008
8097449 °	PCDH10	protocadherin 10							-1.823	±1.64	0.005
8003298 ° *	SLC7A5	solute carrier family 7 (amino acid transporter light chain, L system), member 5							-1.819	±1.09	0.007
8006187 ° *	ATAD5	ATPase family, AAA domain containing 5							-1.816	±1.11	0.007
7974337 *									-1.814	±1.14	0.000
8086880 ° *	CDC25A	cell division cycle 25 homolog A (S. pombe)							-1.807	±0.99	0.009
8015769 ° *	BRCA1	breast cancer 1, early onset							-1.802	±1.08	0.009
7983306 ° *	WDR76	WD repeat domain 76							-1.801	±1.09	0.009
8127987 °	SNORD50A	small nucleolar RNA, C/D box 50A							-1.801	±1.15	0.009
8117368 °	HIST1H4A (includes others)	histone cluster 1, H4a							-1.800	±1.13	0.007
8109830 ° *	CCDC99	coiled-coil domain containing 99							-1.798	±1.11	0.009
8132843 ° *	HAUS6	HAUS augmin-like complex, subunit 6							-1.797	±1.08	0.007
8055672 ° *	MMADHC	methylmalonic aciduria (cobalamin deficiency) cblD type, with homocystinuria							-1.794	±1.05	0.010
7934979 °	ANKRD1	ankyrin repeat domain 1 (cardiac muscle)							-1.790	±2.18	0.000
8114287 ° *	SPOCK1	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1							-1.784	±1.08	0.007
8152703 °	FBXO32	F-box protein 32							-1.784	±1.26	0.006
7917771 *	DNTTIP2	deoxynucleotidyltransferase, terminal, interacting protein 2							-1.782	±1	0.000
8017262 °	BRIP1	BRCA1 interacting protein C-terminal helicase 1							-1.779	±1.11	0.008
7914878 °	CLSPN	claspin							-1.773	±1.12	0.010
8152582 *	DSCC1	defective in sister chromatid cohesion 1 homolog (S. cerevisiae)							-1.772	±0.97	0.000
8053797 *	ANKRD36C	ankyrin repeat domain 36C							-1.768	±1.06	0.000
7922846 ° *	FAM129A	family with sequence similarity 129, member A							-1.763	±1.06	0.009
8124537 *	HIST1H3A (includes others)	histone cluster 1, H3a							-1.756	±1.05	0.000
7917976 *	SASS6	spindle assembly 6 homolog (C. elegans)							-1.754	±0.99	0.000
8115490 °	ADAM19	ADAM metallopeptidase domain 19							-1.751	±1.32	0.004
7997381 *	CENPN	centromere protein N							-1.748	±0.99	0.000
8067029 °	KCNG1 (includes EG:241794)	potassium voltage-gated channel, subfamily G, member 1							-1.746	±1.28	0.005
7981953 *									-1.746	±0.99	0.000
7981966 *									-1.746	±0.99	0.000
8034512 *	SNORD41	small nucleolar RNA, C/D box 41							-1.737	±1.03	0.000
8151684 °	MMP16	matrix metallopeptidase 16 (membrane-inserted)							-1.721	±1.68	0.003
8054217 *	TXNDC9	thioredoxin domain containing 9							-1.719	±0.94	0.000
8001197 °	NETO2	neuropilin (NRP) and tolloid (TLL)-like 2							-1.715	±1.12	0.009
8042830 *	MTHFD2	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydro							-1.711	±0.96	0.000
7905088 *	HIST2H2AC	histone cluster 2, H2ac							-1.709	±0.96	0.000
7970763 °	FLT1	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)							-1.708	±1.31	0.008
8066074 *	DSN1 (includes EG:10000291)	DSN1, MIND kinetochore complex component, homolog (S. cerevisiae)							-1.701	±0.98	0.000
8042588 *	MPHOSPH10	M-phase phosphoprotein 10 (U3 small nucleolar ribonucleoprotein)							-1.700	±0.96	0.000
7962579 °	AMIGO2	adhesion molecule with Ig-like domain 2							-1.699	±1.19	0.008
7971077 °	POSTN	periostin, osteoblast specific factor							-1.691	±1.67	0.005
7968351 °	C13orf33	chromosome 13 open reading frame 33							-1.688	±1.4	0.006
7989915 *	TIPIN	TIMELESS interacting protein							-1.688	±0.95	0.000
7950933 °	NOX4	NADPH oxidase 4							-1.688	±1.48	0.007
7936322 *	GPAM	glycerol-3-phosphate acyltransferase, mitochondrial							-1.664	±0.93	0.000
7917182 °	ELTD1	EGF, latrophilin and seven transmembrane domain containing 1							-1.642	±1.35	0.007
8039484 °	IL11	interleukin 11							-1.622	±1.35	0.008
8145361 °	NEFM	neurofilament, medium polypeptide							-1.388	±2.23	0.003
7951271 °	MMP1 (includes EG:300339)	matrix metallopeptidase 1 (interstitial collagenase)							-1.228	±2.5	0.004
7985786 °	ACAN aggrecan	-1.204	±2.3	0.006	8112971 °	HAPLN1 hyaluronan and proteoglycan link protein 1	-0.939	±2.16	0.009		
7976567 °	BDKRB1	bradykinin receptor B1							1.129	±2.02	0.009
8045533 °									1.354	±1.46	0.007
8015179 °									1.384	±1.4	0.002
8019588 °									1.384	±1.4	0.002
7903214 °	LPPR4	lipid phosphate phosphatase-related protein type 4							1.391	±2.01	0.001
8176719 °	EIF1AY	eukaryotic translation initiation factor 1A, Y-linked							1.412	±2.09	0.002
8176375 °	RPS4Y1	ribosomal protein S4, Y-linked 1							1.431	±2.26	0.001
7999909 °	GPRC5B	G protein-coupled receptor, family C, group 5, member B							1.434	±1.38	0.009
7938225 °	OLFML1	olfactomedin-like 1							1.441	±1.04	0.009
7969202 °									1.464	±1	0.009
8176655 °	NLGN4Y	neuroligin 4, Y-linked							1.466	±1.7	0.003
8037240 °	PSG1	pregnancy specific beta-1-glycoprotein 1							1.480	±1.28	0.009
8013521 °									1.482	±1.08	0.009
8097288 °	FAT4	FAT tumor suppressor homolog 4 (Drosophila)							1.487	±1.07	0.009
7934185 °	C10orf54	chromosome 10 open reading frame 54							1.487	±0.87	0.008
7950005 °	MRGPRF	MAS-related GPR, member F							1.494	±1.14	0.007
8068361 °	SLC5A3	solute carrier family 5 (sodium/myo-inositol cotransporter), member 3							1.512	±0.97	0.009
8104758 °	NPR3 (includes EG:18162)	natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)							1.518	±1.16	0.005
7903358 °	VCAM1	vascular cell adhesion molecule 1							1.528	±1.56	0.009
8058664 °									1.536	±1.05	0.006
8037657 °	DMPK	dystrophia myotonica-protein kinase							1.563	±0.9	0.004

7896748 °			1.570	±1.89	0.001
7911335 °			1.571	±1.59	0.002
8165694 °			1.571	±1.59	0.002
8176578 °	USP9Y	ubiquitin specific peptidase 9, Y-linked	1.578	±1.99	0.001
7975076 °	HSPA2	heat shock 70kDa protein 2	1.588	±1.23	0.005
8150962 °	TOX	thymocyte selection-associated high mobility group box	1.635	±1.39	0.002
7932254 °	ITGA8	integrin, alpha 8	1.637	±1.06	0.004
8045835 °	GALNT5	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5 (GalNAc-T5)	1.639	±1.38	0.002
8176624 °	DDX3Y	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	1.644	±2.36	0.000
7926875 °	BAMBI	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)	1.651	±1.02	0.003
8141140 °	DLX5	distal-less homeobox 5	1.660	±1.25	0.003
8013523 °			1.662	±1.43	0.002
7962058 °	TMTC1	transmembrane and tetratricopeptide repeat containing 1	1.665	±1.66	0.002
7896750 °			1.669	±1.55	0.002
7984364 °	SMAD3	SMAD family member 3	1.689	±0.93	0.002
8156358 °			1.691	±1.9	0.001
8171297 °	MID1 (includes EG:10033095	midline 1 (Opitz/BBB syndrome)	1.691	±1.2	0.003
8113120 °			1.701	±1.22	0.002
8165707 °			1.701	±1.22	0.002
8129666 °	SLC2A12	solute carrier family 2 (facilitated glucose transporter), member 12	1.703	±1.04	0.002
8112668 °	GCNT4	glucosaminyl (N-acetyl) transferase 4, core 2	1.708	±1	0.002
7950810 °	SYTL2	synaptotagmin-like 2	1.718	±0.99	0.002
8095110 °	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	1.722	±1.22	0.002
7911343 °			1.727	±1.38	0.002
8165703 °			1.727	±1.38	0.002
7970565 °			1.731	±1.06	0.002
8008965 °			1.738	±1.13	0.002
7962559 °	SLC38A4	solute carrier family 38, member 4	1.738	±1.23	0.002
8052355 °	EFEMP1	EGF containing fibulin-like extracellular matrix protein 1	1.759	±2.09	0.001
8013987 °			1.824	±1.21	0.001
8160346 °	PTPLAD2	protein tyrosine phosphatase-like A domain containing 2	1.831	±1.2	0.002
7899615 °	SERINC2	serine incorporator 2	1.843	±1.13	0.001
8165663 °			1.863	±1.74	0.001
7911337 °			1.865	±1.92	0.000
7973871 °			1.865	±1.92	0.000
8165696 °			1.865	±1.92	0.000
8102831 °	MGARP	mitochondria-localized glutamic acid-rich protein	1.874	±1.21	0.001
8165709 °			1.914	±1.26	0.001
7911339 °			1.952	±2.12	0.000
8165698 °			1.952	±2.12	0.000
8045804 °			2.012	±1.98	0.000
8102532 °	PDE5A	phosphodiesterase 5A, cGMP-specific	2.024	±1.4	0.001
8104746 °	NPR3 (includes EG:18162)	natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)	2.032	±1.38	0.000
7985317 °	KIAA1199	KIAA1199	2.088	±1.9	0.000
8057677 °	SLC40A1	solute carrier family 40 (iron-regulated transporter), member 1	2.169	±1.51	0.000
8100310 °			2.221	±1.88	0.000
7933204 °	C10orf10	chromosome 10 open reading frame 10	2.301	±1.43	0.000
8051583 °	CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1	2.318	±1.84	0.000
8083887 °	CLDN11	claudin 11	2.430	±1.6	0.000
7912537 °	DHRS3	dehydrogenase/reductase (SDR family) member 3	2.441	±1.43	0.000
8006433 °	CCL2	chemokine (C-C motif) ligand 2	2.913	±2.11	0.000

(*) Genes selected by SAM; (°) Genes selected by Rank Products

Table SII: Top enriched functions in the IPA interaction network

Function	Function Annotation	p-value	# Molec.
Cell Cycle			
recombination	DNA recombination	5.02E-09	7
recombination	recombination of cells	5.66E-05	3
homologous recombination	homologous recombination of DNA	4.42E-07	4
homologous recombination	homologous recombination of plasmid DNA	6.58E-05	2
homologous recombination	homologous recombination of cells	3.96E-02	1
checkpoint control	checkpoint control	4.96E-07	6
checkpoint control	checkpoint control of fibroblast cell lines	1.82E-04	2
DNA replication checkpoint	DNA replication checkpoint	1.99E-06	3
G2/M phase	G2/M phase	2.16E-05	6
G2/M phase	G2/M phase of tumor cell lines	6.16E-04	3
G2/M phase	G2/M phase of breast cancer cell lines	1.21E-03	2
G2/M phase	G2/M phase of cervical cancer cell lines	2.32E-02	1
G2/M phase	arrest in G2/M phase of breast cancer cell lines	3.32E-02	1
S phase checkpoint control	S phase checkpoint control of cervical cancer cell lines	1.08E-04	2
interphase	interphase	2.96E-04	9
interphase	arrest in interphase of fibroblast cell lines	1.79E-03	3
interphase	interphase of breast cancer cell lines	2.22E-03	3
interphase	arrest in interphase	1.19E-02	5
interphase	interphase of cervical cancer cell lines	1.29E-02	2
interphase	interphase of tumor cell lines	2.64E-02	4
interphase	arrest in interphase of tumor cell lines	4.69E-02	3
sister chromatid exchange	sister chromatid exchange	6.67E-04	2
sister chromatid exchange	sister chromatid exchange of DNA	3.70E-03	1
sister chromatid exchange	sister chromatid exchange of chromosomes	4.52E-03	1
sister chromatid exchange	sister chromatid exchange of cervical cancer cell lines	6.16E-03	1
G1/S phase transition	arrest in G1/S phase transition	9.41E-04	3
G1/S phase transition	arrest in G1/S phase transition of breast cancer cell lines	5.75E-03	1
meiosis	meiosis of germ cells	1.27E-03	3
meiosis	meiosis of primordial germ cells	2.47E-03	1
meiosis	meiosis of male germ cells	7.36E-03	2
morphology	morphology of spindle fibers	2.22E-03	2
S phase	S phase	2.33E-03	5
S phase	entry into S phase of tumor cell lines	8.63E-03	2
S phase	arrest in S phase of keratinocytes	9.43E-03	1
S phase	arrest in S phase	1.11E-02	2
S phase	entry into S phase	1.42E-02	3
S phase	S phase of fibroblast cell lines	1.55E-02	2
S phase	entry into S phase of cervical cancer cell lines	2.00E-02	1
S phase	re-entry into S phase of bone cancer cell lines	2.32E-02	1
S phase	entry into S phase of cancer cells	3.12E-02	1
S phase	arrest in S phase of fibroblast cell lines	3.72E-02	1
S phase	entry into S phase of breast cancer cell lines	4.71E-02	1
G1 phase	arrest in G1 phase of breast cancer cell lines	3.30E-03	2
G1 phase	arrest in G1 phase	1.13E-02	4
G1 phase	G1 phase of fibroblast cell lines	1.92E-02	2
G1 phase	arrest in G1 phase of carcinoma cell lines	4.79E-02	1
endomitosis	endomitosis of leukemia cell lines	3.70E-03	1
cell cycle progression	arrest in cell cycle progression of embryonic stem cell lines	4.11E-03	1
mitosis	delay in mitosis of breast cancer cell lines	4.11E-03	1
mitosis	entry into mitosis of embryonic cell lines	8.61E-03	1
mitosis	entry into mitosis of epithelial cell lines	8.61E-03	1

mitosis	entry into mitosis of kidney cell lines	1.15E-02	1
mitosis	entry into mitosis of colon cancer cell lines	1.51E-02	1
mitosis	mitosis of tumor cell lines	2.96E-02	2
mitosis	entry into mitosis of cervical cancer cell lines	3.00E-02	1

progression	progression of chromosomes	4.11E-03	1
DNA damage checkpoint	DNA damage checkpoint	5.26E-03	2
G2 phase	arrest in G2 phase of endometrial cancer cell lines	5.34E-03	1
G2 phase	arrest in G2 phase of embryonic stem cells	7.38E-03	1
G2 phase	arrest in G2 phase	1.26E-02	3
G2 phase	arrest in G2 phase of tumor cell lines	3.63E-02	2
G2 phase	arrest in G2 phase of lung cancer cell lines	4.19E-02	1
G2 phase	arrest in G2 phase of carcinoma cell lines	4.59E-02	1
organization	organization of chromosomes	6.19E-03	2
illegitimate recombination	illegitimate recombination of plasmid DNA	7.79E-03	1
senescence	senescence of fibroblasts	8.58E-03	2
aneuploidy	aneuploidy of fibroblasts	8.61E-03	1
aneuploidy	aneuploidy of cells	9.83E-03	2
abnormal morphology	abnormal morphology of meiotic spindles	9.43E-03	1
G2/M phase transition	arrest in G2/M phase transition	1.33E-02	2
reorganization	reorganization of chromatin	1.47E-02	1
homologous pairing	homologous pairing of DNA	1.76E-02	1
formation	formation of chromosomes	2.08E-02	1
segregation	segregation of chromosomes	2.34E-02	2
DNA Replication, Recombination, and Repair			
recombination	DNA recombination	5.02E-09	7
recombination	recombination of cells	5.66E-05	3
homologous recombination	homologous recombination of DNA	4.42E-07	4
homologous recombination	homologous recombination of plasmid DNA	6.58E-05	2
homologous recombination	homologous recombination of cells	3.96E-02	1
checkpoint control	checkpoint control	4.96E-07	6
checkpoint control	checkpoint control of fibroblast cell lines	1.82E-04	2
DNA replication checkpoint	DNA replication checkpoint	1.99E-06	3
repair	repair of DNA	7.53E-06	7
repair	repair of gene	1.31E-02	1
DNA damage response	DNA damage response of cells	8.46E-06	6
damage	damage of chromosomes	1.86E-05	4
metabolism	metabolism of DNA	4.47E-05	8
homologous recombination repair	homologous recombination repair of DNA	7.07E-05	3
formation	formation of chiasmata	8.55E-05	2
formation	formation of nuclear foci	8.94E-05	3
formation	formation of chromosome components	6.21E-04	3
formation	formation of sex bodies	4.11E-03	1
formation	formation of chromatin	5.34E-03	2
formation	formation of RAD51 nuclear focus	1.23E-02	1
formation	formation of chromosomes	2.08E-02	1
S phase checkpoint control	S phase checkpoint control of cervical cancer cell lines	1.08E-04	2
progression	progression of replication fork	1.14E-04	2
progression	progression of chromosomes	4.11E-03	1
replication	DNA replication	1.17E-04	6
replication	initiation of replication of DNA	2.51E-03	2
modification	modification of gene	2.22E-04	2
somatic hypermutation	somatic hypermutation	3.04E-04	2
breakage	breakage of chromosomes	3.91E-04	3
sister chromatid exchange	sister chromatid exchange	6.67E-04	2

sister chromatid exchange	sister chromatid exchange of DNA	3.70E-03	1
sister chromatid exchange	sister chromatid exchange of chromosomes	4.52E-03	1
sister chromatid exchange	sister chromatid exchange of cervical cancer cell lines	6.16E-03	1
breakdown	breakdown of chromosomes	6.97E-04	2
mutation	mutation of gene	6.97E-04	2
unwinding	unwinding of DNA	1.45E-03	2
unwinding	unwinding of DNA fragment	1.51E-02	1
mismatch repair	mismatch repair	1.71E-03	2
chromosomal instability	chromosomal instability	1.87E-03	2
double-stranded DNA break repair	double-stranded DNA break repair	2.36E-03	3

double-stranded DNA break repair	double-stranded DNA break repair of B-lymphocyte	3.70E-03	1
double-stranded DNA break repair	double-stranded DNA break repair of cells	7.41E-03	2
double-stranded DNA break repair	double-stranded DNA break repair of epithelial cells	9.02E-03	1
conversion	conversion of gene	2.88E-03	1
exchange	exchange of chromosomes	2.88E-03	1
exchange	exchange of sister chromatids	1.59E-02	1
joining	joining of DNA	4.09E-03	2
DNA damage checkpoint	DNA damage checkpoint	5.26E-03	2
organization	organization of chromosomes	6.19E-03	2
illegitimate recombination	illegitimate recombination of plasmid DNA	7.79E-03	1
quantity	quantity of nucleoprotein filaments	8.20E-03	1
abnormal morphology	abnormal morphology of meiotic spindles	9.43E-03	1
reorganization	reorganization of chromatin	1.47E-02	1
excision repair	excision repair	1.68E-02	2
homologous pairing	homologous pairing of DNA	1.76E-02	1
catabolism	catabolism of ATP	2.14E-02	2
segregation	segregation of chromosomes	2.34E-02	2
instability	instability of DNA	2.44E-02	1
ligation	ligation of DNA	3.24E-02	1
elongation	elongation of DNA	4.86E-02	1
Cellular Compromise			
damage	damage of chromosomes	1.86E-05	4
breakage	breakage of chromosomes	3.91E-04	3
breakdown	breakdown of chromosomes	6.97E-04	2
micronucleation	micronucleation of breast cancer cell lines	4.11E-03	1
atrophy	atrophy of motor neurons	8.20E-03	1
oxidative stress response	oxidative stress response of fibroblast cell lines	2.88E-02	1

Table SIII: Validation of the microarray assays

Gene	p-value
*CDC45L**	0.0058
PCOLCE2	0.0062
*BRCA1**	0.0083
*BRIP1**	0.0101
*DTL**	0.0113
*CDC25A**	0.0117
*RAD51AP1**	0.0118
E2F7	0.0124
*DKK1**	0.0166
CDH2	0.0169
LAMC2	0.0182
CCDC99	0.02
*BLM**	0.0209
*CDC6**	0.0238
DCK	0.0244
*RAD51**	0.0279
*HIST1H1B**	0.0496
PODXL	0.1062
MTHFD2	0.1113
ADAM12	0.1616
PCDH10	0.1679
HIST1H4B	0.2584
AMIGO2	0.3369
ITGA2	0.3459

(*) Genes pertaining to the BRCA1 similarity cluster

Table SIV: DEGs involved in the oxidative generation and repair of DSBs

Symbol	Function	Reference
GSTM2	ROS detoxification, glutathione metabolism	[26]
NOX4	Production of ROS	[27]
PTGS2	Peroxidase activity	[28]
SMAD3	oxidative stress response/TGFB-mediated growth inhibition	[29]
BRCA1	oxidative stress response/TGFB-mediated growth inhibition; Activation of baseexcision repair/homologous recombination repair	[29], [31]
CLSPN	Stabilisation of the replication fork	[30]
TIPIN	Stabilisation of the replication fork	[30]
CDC45L	New DNA synthesis during DNA repair	[32]
MCM10	New DNA synthesis during DNA repair	[33]
GINS1	New DNA synthesis during DNA repair	[32]

Table SV: Cell cultures used in the study

Code	Gender	Clinical Status*	Microarray	qRT-PCR validation	Flow cytometry	H2O2 qRT-PCR
F3334	female	control	X	X	X	X
F3347	male	control	X	X	X	X
F3960	female	control	X	X		
F4217	male	control	X	X	X	X
F4319	female	control	X	X		X
F4386	female	control	X	X		
F4243	male	CLP/UR	X	X	X	X
F4244	male	CLP/UR	X	X	X	X
F4245	male	CLP/UR	X	X	X	X
F4293	female	CL/UL	X	X	X	X
F4294	female	CL/UL	X	X	X	X
F4311	male	CL/UL	X	X	X	X
F4388	male	CLP/UR	X	X	X	X
F3333	male	control				X
F5541	male	control				X
F5703	male	control				X
F6032	female	control			X	X
F6119	male	control			X	X
F6681	male	control				X
F4281	male	CLP/UL				X
F4282	male	CLP/UL				X
F5640	female	CL/UL				X
F5720	male	CLP/UR				X

(*) CL = Cleft Lip; CLP = Cleft Lip and Palate; UL = Unilateral Left; UR = Unilateral Right

Table SVI: Primer sequences used for qRT-PCR experiments

Gene	Primer Forward (5'-3')	Primer Reverse (5'-3')
ADAM12	TCTTGCTGCCGGATTTGTGGTT	AAGGGCGCACACACCTTAGTTT
AMIGO2	AGGTGCAAGAGTGACAGACACGG	CTCTTTCCTGGGTTAGTCGGTGC
BLM	AGCACATTGTGTCAGTCAGTGGG	CTGTGGCCGTAAGAGCCATCAC
BRCA1	AGGTGAAGCAGCATCTGGGTGTG	TGCATGGTATCCCTCTGCTGAGTGG
BRIP1	AGTTCAGCTTCGGTTTGCTCGGG	AGCACACAGCTCGTAGGGGTTCA
CCDC99	GCCCTAGAGAAAGCTCGTGTAGC	CCACCTCTGCAAACAAAGAGTTGC
CDC25A	TTCCCTACCTCAGAAGCTGTTGGG	AGTGCAGGCAGCCACGAGATAC
CDC45L	GTGTGTGACACCCATAGGCCAG	TCTAACCGTGTGCGCTTCTCAG
CDC6	TCAGGAAGAGGTATCCAGGCCAG	AATACCAACACAATCATGGGGCCC
CDH2	CGCAGTGTACAGAATCAGTGGCGG	AGTCGATTGGTTTGACCACGGTG
DCK	TCAAGCCACTCCAGAGACATGCTT	TGCAGGAGCCAGCTTTCATGTT
DKK1	ATTGACAACTACCAGCCGTACCCG	AGTAATTCCCGGGGCAGCACATAG
DTL	TATGGAAGGTCTCCACACCCTGG	GTGAAGTCAGATGGACACCAGCAC
E2F7	CCTGTGCCAGAAGTTTCTAGCTCG	GCGTCTCCTTTCCACACCAAGAC
HIST1H1B	TCTTGCCACCATGTCGGAAACCG	CGGCAGCCTTCTTAGTTGCCTTC
HIST1H4B	ACCGAAAAGTGCTGCGGGATAAC	ACCGGAAATTCGCTTAACCCCAC
ITGA2	TGTTAGCGCYCAGTCAAGGCATTT	TGCTGCACTGCATAGCCAAACTGT
LAMC2	AGGTTGATACCAGAGCCAAGAACGCT	TCATCTACACTGAGAGGCTGGTCCAT
MTHFD2	TCTCTAATGTCTGCTTTGGCTGCC	CAACAGCTTCATTTCGAACTGCCG
PCDH10	CCAACGAGACTAAACACCAGCGAG	CTATGTCGGCTTCCTGGAATGCAG
PCOLCE2	TGCTGAACCAAACGAAAGAGGGGA	ATGTGCCACACACAAGTGACTCCT
PODXL	TTCCAGGAAGTCAGACCGTGGTC	ACTGACCCCTGCCTCCTTTAGTTC
RAD51	GGCAATGCAGATGCAGCTTGAAGT	TTATGCCACACTGCTCTAACCGTG
RAD51AP1	GCAGTGCCTTGTACAAAGATGGCT	GTGGTGACTGTTGGAAGTTCCTTCA
E2F1	CCGCCATCCAGGAAAAGGTGTG	TTCAGGTCGACGACACCGTCAG
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGA
HPRT1	TGACACTGGCAAAACAATGC	GGTCCTTTTCACCAGCAAGCT
SDHA	TGGGAACAAGAGGGCATCTG	CCACCACTGCATCAAATTCA
HMBS	GGCAATGCGGCTGCAA	GGGTACCCACGCGAATCAC
Brca1	AGAAGAAAGGGCCTTCACAATGTCC	ACCATTTGTAAGCTGCATTCCCGTG
Brip1	GGGACGCGGACACACAAGCAG	ACTTCCCGTGCTTGCCATCCAG
Msh2	TGGCATTTAAGGCTTCTCCCGGC	ACCCATAACGCCAACGGAAGCTG
Blm	AGTGCTGCAACGACCCCTCG	AGCCATGATCCTCATCTGGCATCC
Rad51	AGATGGAGCAGCCATGTTCGCTG	TCTCAGGTACAGCCTGGTGGTTG
Rad51ap1	AAGTCACCATCTCAAGCCAAGGC	TTTCTGAACCTTCACCAGTGGCG
B2m	TCGCGGTCGCTTCAGTCGTC	TTCTCCGGTGGGTGGCGTGA
Tbp	CCACACCAGCTTCTGAGAGC	GACTGCAGCAAATCGCTTGGG
Tubb5	GACAGTGTGGCAACCAGATCG	CTGCAGGTCGCTGTCACCGT
Ywhaz	TGAGCAGAAGACGGAAGGTGC	GCGAAGCATTGGGGATCAAGA

Chapter 3

Differences in promoter methylation status are associated with *BRCA1* downregulation in non-syndromic cleft lip/palate

Lucas Alvizi Cruz, Gerson Shigeru Kobayashi, Camila Lovaglio Campos, Maria Rita dos Santos Passos-Bueno

Centro de Pesquisas Sobre o Genoma Humano e Células-Tronco, Instituto de Biociências, Universidade de São Paulo

Key-words: cleft lip, cleft palate, DNA damage repair, DNA methylation, epigenetics

Abstract

Non-syndromic cleft lip/palate (NSCL/P) is the most common craniofacial malformation and follows a multifactorial model of inheritance. Several candidate *loci* for NSCL/P have been identified by different strategies and both functional and association studies suggest the involvement of the DNA damage repair pathway in susceptibility to the malformation. Previous studies have demonstrated *BRCA1* downregulation and DNA damage repair deficiency in NSCL/P cells, however no causal mechanism have been elucidated in those cases. We have investigated *BRCA1* promoter methylation in 18 NSCL/P cells with confirmed *BRCA1* downregulation and deficiency in DNA damage repair in comparison to 12 controls. We observed a significant increase in total *BRCA1* promoter methylation in NSCL/P (p=0.0355) and a distinct methylation distribution in the analysed CpGs in comparison to controls. We also investigated H3K9Ac and H3K27Me levels in *BRCA1* promoter as well as *BRCA1*-targeting miRNAs levels and observed no significant differences, suggesting DNA methylation as the most potential mechanism. We also demonstrated a significant negative correlation between expression levels from *BRCA1* and *IRF6*, a gene in which mutations lead to syndromic forms of cleft lip and palate, which also suggests the involvement of other genes in *BRCA1* downregulation. In conclusion, *BRCA1* promoter methylation can be involved in NSCL/P *BRCA1* deficiency and can contribute to DNA damage repair impairment observed in those cells.

Resumo

As fissuras lábio-palatinas não sindrômicas (FL/P NS) são as malformações craniofaciais mais comuns e possuem um padrão multifatorial de herança. Muitos *loci* candidatos para as FL/P NS tem sido identificados por diferentes estratégias e tanto estudos funcionais e de associação sugerem o envolvimento da via de reparo de dano de DNA na susceptibilidade a esta malformação. Estudos anteriores demonstraram tanto a redução de expressão do *BRCA1* quanto a deficiência no reparo ao dano de DNA em células FL/P NS, no entanto os mecanismos causais responsáveis por tais fenômenos não foram elucidados. Nós investigamos a metilação no promotor do *BRCA1* em 18 células FL/P NS com redução da expressão do *BRCA1* e deficiência de reparo de DNA já confirmados em comparação a 12 células controles. Nós observamos um aumento significativo na metilação total do promotor do *BRCA1* nas amostras FL/P NS (p=0.0355) e uma distribuição de metilação ao longo das CpGs analisadas distinta em relação aos controles. Nós também investigamos os níveis relativos de H3K9Ac e H3K27Me no promotor do BRCA1, além dos níveis de expressão de miRNAs reguladores de *BRCA1* e não observamos diferenças significativas em ambos os casos, sugerindo-se, portanto, a metilação de DNA como a mecanismo mais provável. Nós também demonstramos uma correlação negativa significativa entre os níveis de expressão de *BRCA1* e *IRF6*, um gene cujas mutações resultam em formas sindrômicas de fissura de lábio e palato, o que também sugere a participação de outras moléculas na redução da expressão do *BRCA1*. Em conclusão, a metilação do promotor do *BRCA1* pode estar envolvida na deficiência de expressão do *BRCA1* em FL/P NS e pode contribuir à desregulação do reparo ao dano de DNA observados nessas células.

Introduction

Craniofacial development is a complex process dependent on several molecules and pathways under precise spatio-temporal regulation. Disruption in regulation of such pathways by either environmental or genetic factors may lead to craniofacial abnormalities, including non-syndromic cleft lip/palate (NSCL/P), the most common craniofacial malformation (CLOUTHIER et al., 2000; GOU; ZHANG; XU, 2015; MOSSEY et al., 2009; WILKIE; MORRIS-KAY, 2001). NSCL/P affects 1:700 live-births worldwide and is a relevant burden to health systems due to costs in treatment which requires a multidisciplinary approach (BRITO et al., 2012c).

NSCL/P is considered a multifactorial malformation in which both environmental and genetics factors play a role in its aetiology (MOSSEY et al., 2009; STANIER; MOORE, 2004). Attempts to uncover the genetic basis of NSCL/P have been made by several studies, and several candidate and at-risk loci have been uncovered under distinct approaches, such as linkage analysis, genome-wide association studies and whole-exome and whole-genome sequencing (BEATY et al., 2010; BIRNBAUM et al., 2009; BRITO et al., 2011, 2012a, 2012b; GRANT et al., 2009; LUDWIG et al., 2012; MANGOLD et al., 2010). Meanwhile, functional studies in cellular and animal models have also identified altered gene pathways implicated in orofacial clefting (BUENO et al., 2011; KOBAYASHI et al., 2013; KUROSAKA et al., 2014). In this manner, deficiency in DNA repair pathway genes has been suggested as a molecular mechanism for NSCL/P by transcriptome analysis and association studies. It has been shown that *BRCA1* is downregulated in cells from NSCL/P individuals, resulting in impaired DNA double strand break (DSB) repair in comparison to controls (BROOKLYIN et al., 2014; KOBAYASHI et al., 2013; MOSTOWSKA et al., 2014; XAVIER et al., 2017). *BRCA1* (breast cancer 1, OMIM *110735) is a gene in which loss-of-function mutations increase the risk for breast cancer, and plays a central function in DSB repair pathway alongside *BRCA2*, *RAD51*, *BLM*, *MSH2* and *BRIP1* (LINGER; KRUK, 2010; ROY; CHUN; POWELL, 2011). *BRCA1* downregulation can be a result of genetic

variants, epigenetic alterations or a combination of both mechanisms. Indeed, it has been demonstrated in breast cancer individuals carrying *BRCA1* mutation that epigenetic alterations such as promoter hypermethylation act as a second hit, leading to tumourigenesis (BEN GACEM et al., 2012; BIRGISDOTTIR et al., 2006; BOSVIEL et al., 2012; IWAMOTO et al., 2011; RICE; MASSEYAAA501BROWN; FUTSCHER, 1998).

Although variants in DNA repair genes have been identified in association with NSCL/P, no epigenetic factors have been studied in *BRCA1*-deficient NSCL/P cells. In this study we replicate previous findings on *BRCA1* downregulation and DNA damage repair deficiency in a new set of NSCL/P cells. We have also investigated *BRCA1* promoter methylation, histone acetylation and methylation marks, as well as the involvement of miRNAs in *BRCA1* dysregulation in those cells.

Methods

Ethics

This study was approved by the Ethics Committee of the Instituto de Biociências (Universidade de São Paulo, Brazil). Biological samples were collected after signed informed consent by the parents or legal guardians. All experiments were performed in accordance with relevant guidelines and regulations.

Patients' and controls' samples

Deciduous teeth were non-invasively obtained from children in exfoliation period. Specimens were kept in DMEM/High Glucose supplemented with 1% penicillin-streptomycin solution (Thermo Fisher Scientific), and taken to the laboratory to be processed. Control teeth were obtained from donors attending odontopaediatric clinics in São Paulo, Brazil, while NSCL/P

teeth were obtained from patients enrolled for surgical treatment at SOBRAPAR Institute, Campinas, Brazil. We considered an individual to be affected by NSCL/P if no clinical alterations other than clefting of the upper lip with or without cleft palate were present.

SHED (Stem Cells from Human Exfoliated Deciduous teeth) cultures were established according to previously published protocols. The primary culture establishment protocols used in our laboratory are reproducible and consistent with regard to the immunophenotype and differentiation potential of the cell populations (BUENO et al., 2011). Cells were cultured in DMEM/F12 (Thermo Fisher Scientific) supplemented with 15% Foetal Bovine Serum (HyClone), 1% Non-essential aminoacids solution (Thermo Fisher Scientific), 1% penicillin-streptomycin solution (Thermo Fisher Scientific), in a humidified incubator at 37°C and 5% CO_2. For storage, cells were frozen in medium containing 90% FBS and 10% DMSO (LGC Biotecnologia). We obtained a new batch of 18 NSCL/P and 12 control SHED independently from those previously studied (KOBAYASHI et al., 2013),used for DNA and RNA extractions as well as H_2O_2 exposure and y-H2AX flow cytometry. DNA and RNA were simultaneously extracted using TRIzol Reagent (Thermo Fisher Scientific) according to the manufacturer's recommended protocol. DNA and RNA integrity were checked via agarose gel electrophoresis. DNA samples were used for methylation analysis and RNA for gene expression, as described herein.

Gene expression analysis

cDNAs were synthesised from 1 µg of total RNA obtained from each NSCL/P and control sample using SuperScript II cDNA synthesis kit (Thermo Fisher Scientific). We used 20 ng of cDNA in real-time quantitative PCRs (RT-qPCR) for *BRCA1* and *IRF6* expression. We also quantified *GAPDH*, *TBP* and *HPRT1* as endogenous controls. For miRNA quantification, cDNAs were synthesised from 1 µg of total RNA from each sample using Ncode miRNA cDNA synthesis kit (Thermo Fisher Scientific). 20 ng of cDNA were used in RT-qPCRs for hsa-mir-638, hsa-mir-

15.a.5p, hsa-mir-16.1.3p, and hsa-mir-16.5p expression, using *RNU48*, *RNU44* and *RNU6B* as endogenous controls. Primers for all analysed coding genes and miRNAs are available under request. All RT-qPCRs were performed with the use of Fast SYBR-Green Master Mix (ThermoFisher) in an ABI 7500 Fast Sequence Detection System (Thermo Fisher Scientific). Expression of target genes or miRNas was assessed relative to a calibrator cDNA (ΔCt). Next, GeNorm v3.4 (VANDESOMPELE et al., 2002) was used to calculate normalisation factors for each sample based on endogenous controls. The final expression values were determined based on a previous method (PFAFFL, 2001), dividing $E^{-\Delta Ct}$ by the corresponding normalisation factor. To compare the expression between control and NSCL/P groups, we applied a Student's t-test with Welch's correction considering significant differences when p-value<0.05.

H_2O_2 exposure and y-H2AX quantification

We performed H_2O_2 exposure assays and y-H2AX quantification as previously published (KOBAYASHI et al., 2013). In summary, NSCL/P and control cells were seeded into 6-well plates (Corning), in duplicates (10^4 cells/cm^2). The next day, cells were rinsed with PBS, and the culture medium was replaced by medium containing 100 μM of freshly diluted H_2O_2 (Merck), followed by incubation for 6 hrs in the dark to stimulate DSB formation. After complete fixation, cells were double-stained with PI (Propidium Iodide) and anti-γ-H2AX (Anti-phospho-H2A.X Ser139 clone JBW301 FITC conjugate, Millipore), in order to ascertain cell cycle distribution and DSB formation, respectively. Data were analysed with Guava Express PRO software (Millipore) and γ-H2AX profiles were compared using a Student's t-test with Welch's correction.

BRCA1 promoter methylation analysis

We performed bisulfite sequencing analysis for *BRCA1* promoter methylation quantification. 1 µg of DNA from each sample was bisulfite-converted with the use of EpiTect Bisulfite Kit (QIAGEN), and bisulfite-specific PCR was performed for *BRCA1* promoter region using previously described primers and conditions (RICE; MASSEYAAA501BROWN; FUTSCHER, 1998). PCR amplicons were cloned using TOPO TA Cloning Kit (Thermo Fisher Scientific) and 10 clones per sample were sequenced with the ABI 3730 DNA Analyser (Thermo Fisher Scientific). Sequencing files were analysed with the online tool BISMA (Bisulfite Sequencing DNA Methylation Analysis - http://services.ibc.uni-stuttgart.de/BDPC/BISMA/) (ROHDE et al., 2010) for total and individual CpG methylation quantification and methylation distribution. We performed Chi-square tests using methylation values from both groups to detect significant differences.

Chromatin ImmunoPrecipitation (ChIP) for H3K9Ac and H3K27Me enrichment analysis

SHED from NSCL/P and controls were used for chromatin extraction followed by immunoprecipitation using antibodies for H3K9Ac and H3K27Me histone marks. 10^6 cells per sample were used for chromatin extraction with the Chromatin Extraction Kit (Abcam) and the recommended protocol. Chromatin extracts were next used for ChIP with the EpiSeeker ChIP Kit (Abcam) and the antibodies anti-H3K9Ac and anti-H3K27Me (Abcam). We performed RT-qPCR in ChIP samples using primers for 2 regions of the *BRCA1* promoter (referred as *BRCA1*-1 and *BRCA1*-2) in order to quantify H3K9Ac and H3K27Me enrichment in those regions. We also quantified *GAPDH* and *H19ICR* enrichment as endogenous calibrators. Primers for ChIP-RT-qPCR are described in Supplementary Table I. Enrichment values were calculated as described above in the ‘Gene expression analysis topic’. Average fold enrichment for each analysed histone

mark and region was compared between NSCL/P and controls with a Student's t-test with Welch's correction.

Results

NSCL/P SHED exhibit downregulation of BRCA1 and DSB repair impairment

We quantified *BRCA1* transcript levels in 18 NSCL/P and 12 control SHED in order to confirm the previously observed *BRCA1* downregulation (KOBAYASHI et al., 2013). We observed significant reduction of *BRCA1* transcripts (Mean Controls = 4.27 ± 0.32, Mean NSCL/P = 2.21 ± 0.31; 1.93-fold reduction, p=0.0001, Student's t-test with Welch's correction) in NSCL/P in comparison to controls (Figure 1A). Once *BRCA1* downregulation has been associated with impaired DSB repair, we quantified the DSB marker y-H2AX in cells upon H_2O_2 exposure. After 6 hours, we found a significant increase of y-H2AX-positive cells in NSCL/P (Figure 1B). Altogether, *BRCA1* expression reduction and y-H2AX increase in NSCL/P corroborate previous findings (KOBAYASHI et al., 2013) and confirm DSB repair deficiency and DNA damage accumulation in NSCL/P cells.

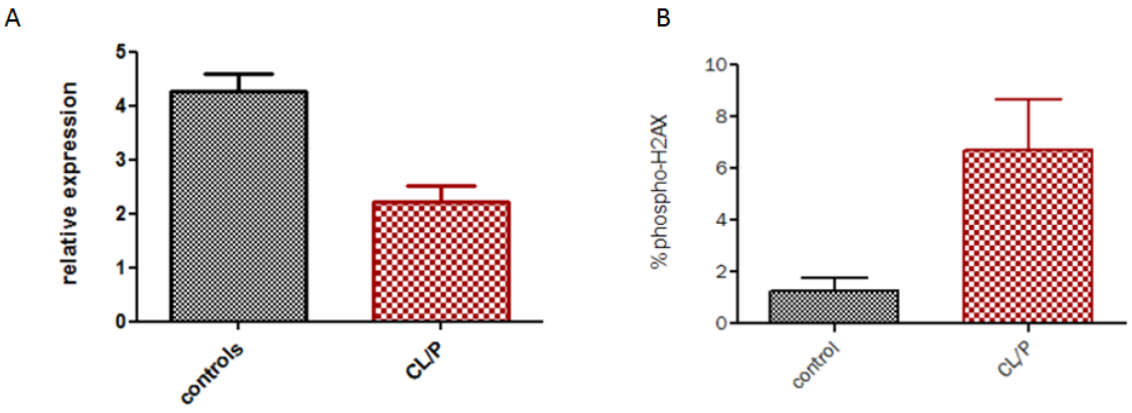


Figure 1: NSCL/P cells exhibit deficiency in DNA damage repair pathway. A: *BRCA1* expression in SHED from controls (grey) and NSCL/P individuals (red), showing significant expression reduction in NSCL/P samples (p=0.0001, Student's t-test with Welch's correction). B: y-H2AX quantification in SHED from controls (grey) and NSCL/P individuals (red) showing significant increase of y-H2AX^{+} cells in NSCL/P samples (p=0.037, Student's t-test with Welch's correction).

NSCL/P SHED show BRCA1 promoter hypermethylation and differences in methylation distribution

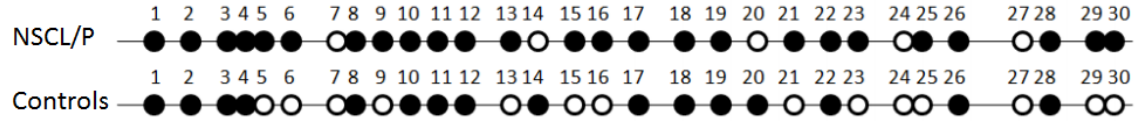
We have hypothesised *BRCA1* promoter methylation levels as a possible mechanism for *BRCA1* downregulation in the studied NSCL/P cells, as this is a frequently observed phenomenon in breast cancer cells. To test this hypothesis we performed bisulfite sequencing in 18 NSCL/P and 12 control DNA samples from SHED. We were able to analyse a CpG island containing 30 CpGs at the *BRCA1* promoter region (Chr17:43125325:43125983, GRCh38.p7). We observed a 1.2% increase in total methylation of *BRCA1* promoter in NSCL/P in comparison to control samples (Table I), revealing a discrete but significant methylation gain at the analysed region (p=0.0355, Chi-square test with Yates' correction).

Table I: Total DNA methylation quantification from 30 CpGs within the *BRCA1* promoter in NSCL/P and control samples, showing a 1.2% methylation increase in NSCL/P in comparison to controls (p=0.0355, Chi squared=3.261, 1 degree of freedom, Chi-square test with Yates' correction).

	NSCL/P		Controls		
	Number of CpGs	%	Number of CpGs	%	p-value
Methylated CpGs	125	4.7	60	3.5	0.0355
Non-methylated CpGs	2531	95.3	1641	96.5	
Total	2656		1701		

By analysing CpGs individually we were able to verify differences in methylation distribution and methylation architecture of the *BRCA1* promoter between NSCL/P and controls. Considering CpG sites fully unmethylated (methylation % = 0), controls displayed 14 sites completely unmethylated in contrast with 4 sites in NSCL/P (p=0.0101, Fisher's Exact Test) (Figure 2A). In this manner, 11 fully unmethylated sites in controls presented a gain of methylation up to 8.7% in NSCL/P (Methylation level at CpG 6, NSCL/P= 8.7%, Controls = 0%), while only 2 fully unmethylated sites in NSCL/P were observed to have a gain of methylation up to 2% in controls (Methylation level at CpG 14, NSCL/P=0%, Controls=2%) (Figure 2B). Taking into account also partially methylated sites, we observed in total 18 CpG sites with gain of methylation in NSCL/P in comparison to 9 in controls.

A



B

CpG		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
% methylation	NSCL/P	17.8	30.2	7.9	2.3	2.5	8.7	0	1.1	3.6	4.3	23.1	1.1	1.1	0	4.2	8.5	1.1	2.1	1.1	0	4.4	3.2	1.1	0	1.6	6.5	0	1.1	2.2	2.4
	CTRLS	42.2	22	7.8	15.2	0	0	0	7.4	0	3.6	13.3	1.5	0	2	0	0	4.2	1.4	3.5	1.4	0	1.4	0	0	0	1.5	0	4.3	0	0

Figure 2: Differences in distribution and composition of methylation values among the 30 CpGs in *BRCA1* promoter in NSCL/P and control samples. A: Distribution of CpGs with detectable levels of methylation (black circles) and CpGs with undetectable levels of methylation - fully unmethylated - (white circles) revealing 14 CpG sites fully unmethylated in controls whilst only 4 are observed in NSCL/P. B: Methylation levels per CpG reveals 11 sites with up to 8.7% gain of methylation in NSCL/P and are fully unmethylated in controls (grey cells), while only 2 sites present a up to 2% gain of methylation in controls and are fully unmethylated in NSCL/P.

Transcriptional dysregulation of BRCA1 in NSCL/P cells is not correlated to histone marks H3K9ac or H3K27me

Because gene expression may also depend on changes in histone marks (COSGROVE; BOEKE; WOLBERGER, 2004; COSGROVE; WOLBERGER, 2005; ZENTNER; HENIKOFF, 2013), we tested whether the specific histone marks H3K9ac and H3K27me, respectively associated with gene activation and repression (CALO; WYSOCKA, 2013; EZPONDA; LICHT, 2014; KARMODIYA et al., 2012), varied in NSCL/P and controls for two regions at the *BRCA1* promoter (herein referred to as *BRCA1*-1 and *BRCA1*-2). We observed no significant differences in H3K9ac or H3K27me for both tested regions, suggesting that changes in those histone marks are not involved in *BRCA1* dysregulation in the tested cells and regions (Figure 3).

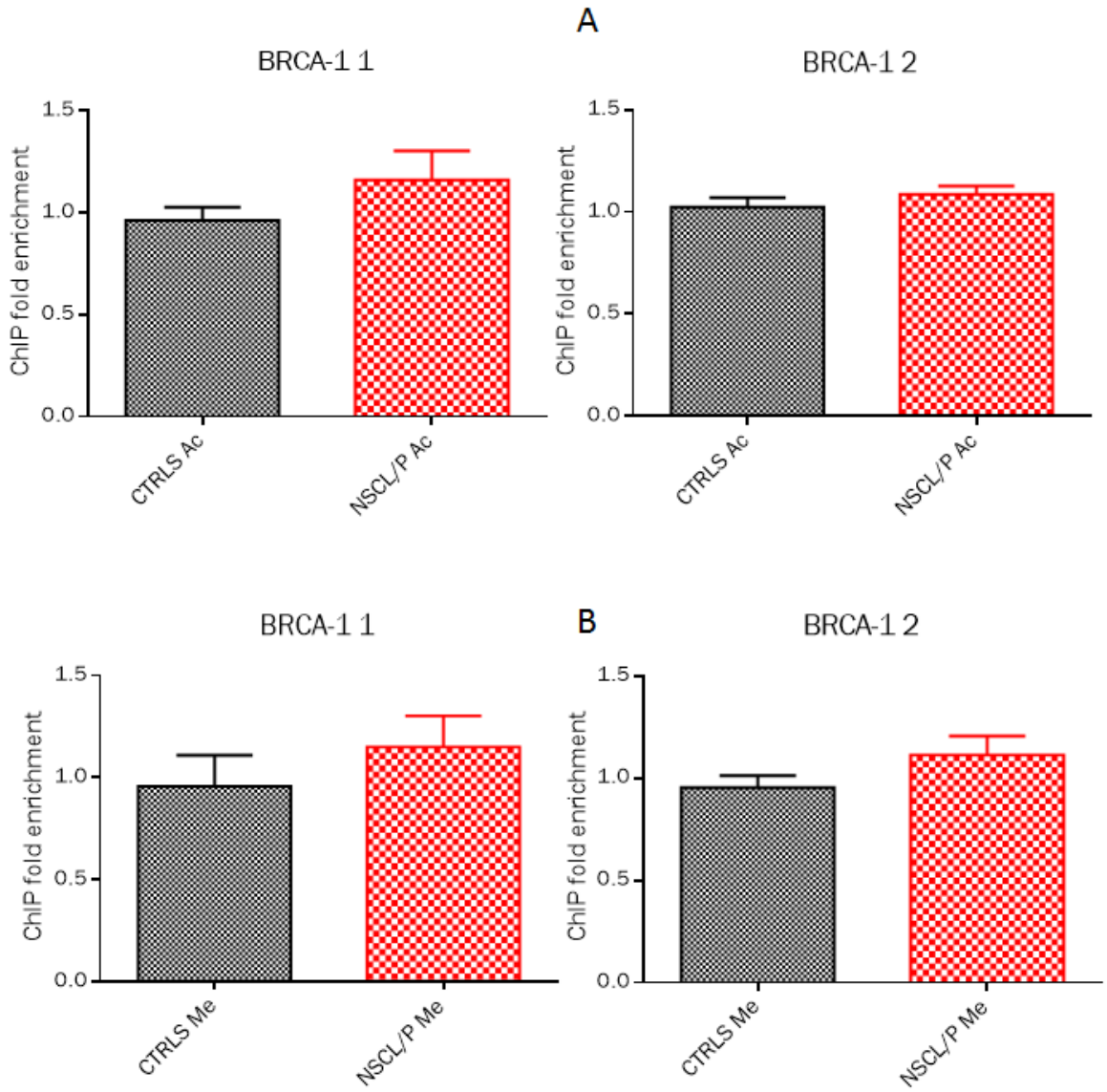


Figure 3: Chromatin Immunoprecipitation (ChIP) analysis for H3K9Ac (Ac) and H3K27Me (Me) for two regions of *BRCA1* promoter (*BRCA1*-1 and *BRCA1*-2) performed in NSCL/P (red) and controls (CTRLS, black) samples. A: ChIP fold enrichment analysis using the histone acetylation mark H3K9Ac (Ac) reveals no significant differences between controls and NSCL/P samples (*BRCA1*-1 p=0.26, *BRCA1*-2 p=0.35; t-test with Welch's correction). B: ChIP fold enrichment analysis using the histone methylation mark H3K27Me (Me) reveals no significant differences between

controls and NSCL/P samples (*BRCA1*-1 p=0.40, *BRCA1*-2 p=0.19; t-test with Welch's correction).

miRNAs targeting BRCA1 show no significant expression differences in NSCL/P samples

Besides DNA methylation and histone acetylation and methylation marks, we also investigated whether miRNAs could be involved in *BRCA1* downregulation in NSCL/P cells. We selected four well-characterised miRNAs known to target *BRCA1* mRNA: hsa-mir-638, hsa-mir-15.a.5p, hsa-mir-16.1.3p and hsa-mir-16.5p (CHANG; SHARAN, 2012). We quantified expression levels of those miRNAs in both NSCL/P and control cell samples and we observed no significant differences between the tested groups (Supplementary Figure 1).

BRCA1 downregulation correlates to IRF6 expression

IRF6 is a transcription factor that belongs to the Interferon Regulatory Factor family, and mutations in this gene lead to syndromic cleft lip/palate (KONDO et al., 2002). Because *BRCA1* deficiency in NSCL/P cells could also be consequence of changes in expression of regulatory transcription factors interacting with *BRCA1*, we tested if *IRF6* expression was correlated to *BRCA1* levels. We thus quantified *IRF6* expression in NSCL/P and control cells and performed a correlation test against *BRCA1* expression levels. We observed a significant and inverse correlation of expression between *BRCA1* and *IRF6* (Figure 4), suggesting that high levels of *IRF6* are in co-occurrence of low levels of *BRCA1* or *vice-versa*.

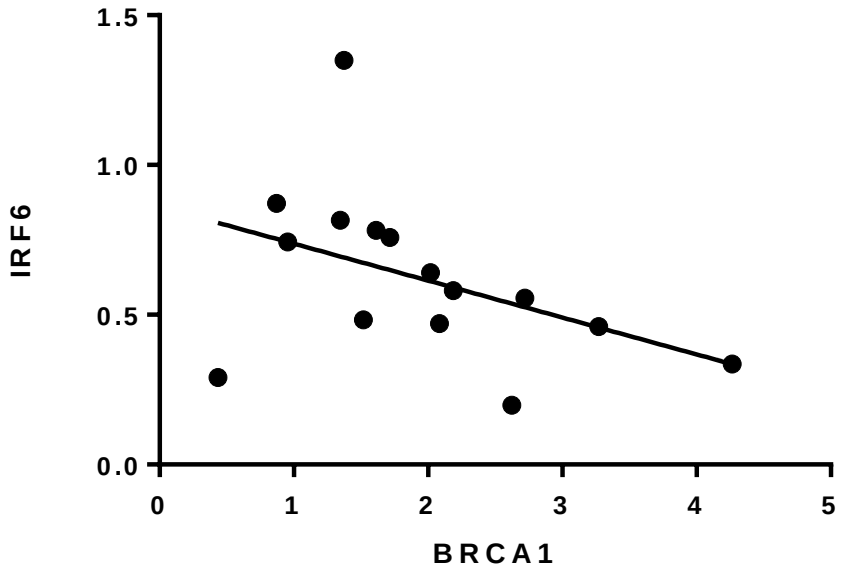


Figure 4: Correlation of expression levels for *BRCA1* and *IRF6* in NSCL/P and control cells, showing a significant negative correlation (p=0.0268, Spearman's correlation test).

Discussion

Deficiency in DNA damage repair due to *BRCA1* pathway downregulation has been suggested as a molecular mechanism for NSCL/P by functional studies (KOBAYASHI et al., 2013; XAVIER et al., 2017). Even though genetic variants in DNA damage repair genes have been found in significant association with NSCL/P, Odds Ratio estimates for these variants remain low (MOSTOWSKA et al., 2014), failing to explain the consistency of such phenomenon. These observations thus prompt redirecting efforts towards studying alternative mechanisms. Besides, given the multifactorial scenario, investigating whether such pathway dysregulation is related to epigenetic variation is necessary. Indeed, recent work on NSCL/P epigenetics has demonstrated the importance of investigating such mechanisms for this malformation (Alvizi et al, in press). In this study, we investigated the potential role of DNA methylation and histone mark changes at the

BRCA1 promoter, *BRCA1*-targeting miRNAs, and the transcription factor *IRF6* in *BRCA1* pathway downregulation in NSCL/P cells. We demonstrate a consistent deficiency of *BRCA1* expression and DNA damage repair in NSCL/P in independent batch of samples, which sustains the hypothesis that such mechanism is involved in the aetiology and susceptibility to NSCL/P, as a primary or secondary effect.

BRCA1 promoter hypermethylation is a common event in breast cancer tissue leading to *BRCA1* deficiency and tumourigenesis (BEN GACEM et al., 2012; BIRGISDOTTIR et al., 2006; BOSVIEL et al., 2012; IWAMOTO et al., 2011; RICE; MASSEYAAA501BROWN; FUTSCHER, 1998). In an equivalent manner, we investigated *BRCA1* promoter methylation in NSCL/P and control cells and found a significant increase of methylation in patient's samples as well as a distinct distribution and composition of CpG methylation in the analysed region. Despite its small magnitude, such methylation differences are significant and may interfere in gene regulation, as previously demonstrated (LEENEN; MULLER; TURNER, 2016). Additionally, we show that the analysed *BRCA1* promoter region is a variable methylation region (VMR) in architecture of methylation by comparing NSCL/P to controls. Changes as VMRs are also associated with changes and variation in gene regulation, which could ultimately lead to changes in gene expression (RAKYAN et al., 2011); thus, those variations could be implicated in *BRCA1* expression variation. On the other hand, considering DNA methylation, alternative regulatory regions comprising CpG islands that may interact with *BRCA1* promoter may also be involved in modulating *BRCA1* expression. Although we were unable to establish a causal effect between *BRCA1* promoter methylation and expression, we suggest those observed methylation variations as a potential mechanism for *BRCA1* deficiency in the studied cells.

We also attempted to investigate the association of both activation and repression histone marks (H3K9Ac and H3K27Me) as well as miRNAs in *BRCA1* regulation and differences between NSCL/P and controls. We found no significant association for both mechanisms with *BRCA1* expression and no differences for the tested marks and miRNAs in the comparison between

groups. We do not exclude, however, the possibility of other histone marks and miRNAs not here studied in taking part into the *BRCA1* dysregulation observed in NSCL/P.

As an alternative mechanism, *BRCA1* expression could also be affected by the expression of regulatory molecules such as transcription factors. The IRF family, consisting of transcription factors involved in several cellular processes, has been demonstrated to interact with tumour suppression genes, including *BRCA1* (TAKAOKA; TAMURA; TANIGUCHI, 2008; TAMURA et al., 2008). *IRF6*, a member of the IRF family in which variants are either associated with NSCL/P susceptibility or cause van der Woude and popliteal pterygium syndromes, two allelic syndromic forms of cleft lip and palate (KONDO et al., 2002), has been demonstrated to regulate tumourigenesis and potentially interact with *BRCA1* (TAKAOKA; TAMURA; TANIGUCHI, 2008; TAMURA et al., 2008). We therefore investigated the correlation of *BRCA1* expression with *IRF6* expression levels and found a significant inverse correlation, which suggests that the occurrence of *IRF6* transcripts coincides with downregulation of *BRCA1*, under a putative regulatory cross-talk between those genes. Such significant correlation does not indicate whether *IRF6* could downregulate *BRCA1* or *BRCA1* could downregulate *IRF6*, although dysregulation in both processes could contribute to the clefting phenotype. We also acknowledge the existence of several molecules taking part in the *BRCA1* DNA damage repair pathway, in which genetic or epigenetic alterations could lead to *BRCA1* deficiency, a possibility that could not be ruled out by this work.

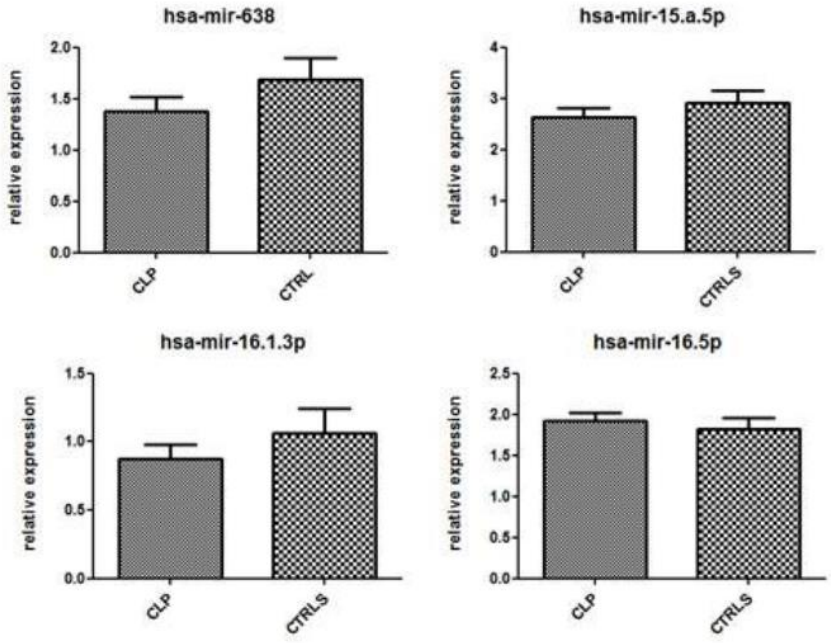
In summary we demonstrated the replication of both *BRCA1* and DNA damage repair deficiencies in NSCL/P cells and show *BRCA1* promoter methylation as a putative mechanism for *BRCA1* downregulation. Additionally, alternative regulatory genes as *IRF6* may play a role in this process.

References

BEATY, T. H. et al. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. **Nature Genetics**, v. 42, n. 6, p. 525-529, jun. 2010.
BEN GACEM, R. et al. Contribution of epigenetic alteration of BRCA1 and BRCA2 genes in breast carcinomas in Tunisian patients. **Cancer epidemiology**, v. 36, n. 2, p. 190-197, abr. 2012.
BIRGISDOTTIR, V. et al. Epigenetic silencing and deletion of the BRCA1 gene in sporadic breast cancer. **Breast Cancer Research**, v. 8, n. 4, p. R38, 2006.
BIRNBAUM, S. et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. **Nature Genetics**, v. 41, n. 4, p. 473-477, abr. 2009.
BOSVIEL, R. et al. BRCA1 promoter methylation in peripheral blood DNA was identified in sporadic breast cancer and controls. **Cancer epidemiology**, v. 36, n. 3, p. e177-82, jun. 2012.
BRITO, L. A. et al. Genetic contribution for non-syndromic cleft lip with or without cleft palate (NS CL/P) in different regions of Brazil and implications for association studies. **American Journal of Medical Genetics. Part A**, v. 155A, n. 7, p. 1581-1587, jul. 2011.
BRITO, L. A. et al. Region 8q24 is a susceptibility locus for nonsyndromic oral clefting in Brazil. **Birth Defects Research. Part A, Clinical and Molecular Teratology**, v. 94, n. 6, p. 464-468, jun. 2012a.
BRITO, L. A. et al. IRF6 is a risk factor for nonsyndromic cleft lip in the Brazilian population. **American Journal of Medical Genetics. Part A**, v. 158A, n. 9, p. 2170-2175, set. 2012b.
BRITO, L. A. et al. Genetics and management of the patient with orofacial cleft. **Plastic surgery international**, v. 2012, p. 782821, 1 nov. 2012c.
BROOKLYIN, S. et al. Assessment of folic Acid and DNA damage in cleft lip and cleft palate. **Clinics and practice**, v. 4, n. 1, p. 608, 27 mar. 2014.
BUENO, D. F. et al. Human stem cell cultures from cleft lip/palate patients show enrichment of transcripts involved in extracellular matrix modeling by comparison to controls. **Stem Cell Reviews**, v. 7, n. 2, p. 446-457, jun. 2011.
CALO, E.; WYSOCKA, J. Modification of enhancer chromatin: what, how, and why? **Molecular Cell**, v. 49, n. 5, p. 825-837, 7 mar. 2013.
CHANG, S.; SHARAN, S. K. BRCA1 and microRNAs: emerging networks and potential therapeutic targets. **Molecules and Cells**, v. 34, n. 5, p. 425-432, nov. 2012.
CLOUTHIER, D. E. et al. Signaling pathways crucial for craniofacial development revealed by endothelin-A receptor-deficient mice. **Developmental Biology**, v. 217, n. 1, p. 10-24, 1 jan. 2000.
COSGROVE, M. S.; BOEKE, J. D.; WOLBERGER, C. Regulated nucleosome mobility and the histone code. **Nature Structural & Molecular Biology**, v. 11, n. 11, p. 1037-1043, nov. 2004.
COSGROVE, M. S.; WOLBERGER, C. How does the histone code work? **Biochemistry and Cell Biology (Biochimie et Biologie Cellulaire)**, v. 83, n. 4, p. 468-476, ago. 2005.
EZPONDA, T.; LICHT, J. D. Molecular pathways: deregulation of histone h3 lysine 27 methylation in cancer-different paths, same destination. **Clinical Cancer Research**, v. 20, n. 19, p. 5001-5008, 1 out. 2014.

GOU, Y.; ZHANG, T.; XU, J. Transcription factors in craniofacial development: from receptor signaling to transcriptional and epigenetic regulation. **Current Topics in Developmental Biology**, v. 115, p. 377-410, 6 out. 2015.
GRANT, S. F. A. et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. **The Journal of Pediatrics**, v. 155, n. 6, p. 909-913, dez. 2009.
IWAMOTO, T. et al. BRCA1 promoter methylation in peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. **Breast Cancer Research and Treatment**, v. 129, n. 1, p. 69-77, ago. 2011.
KARMODIYA, K. et al. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. **BMC Genomics**, v. 13, p. 424, 24 ago. 2012.
KOBAYASHI, G. S. et al. Susceptibility to DNA damage as a molecular mechanism for non-syndromic cleft lip and palate. **Plos One**, v. 8, n. 6, p. e65677, 12 jun. 2013.
KONDO, S. et al. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. **Nature Genetics**, v. 32, n. 2, p. 285-289, out. 2002.
KUROSAKA, H. et al. Disrupting hedgehog and WNT signaling interactions promotes cleft lip pathogenesis. **The Journal of Clinical Investigation**, v. 124, n. 4, p. 1660-1671, abr. 2014.
LEENEN, F. A. D.; MULLER, C. P.; TURNER, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? **Clinical epigenetics**, v. 8, p. 92, 6 set. 2016.
LINGER, R. J.; KRUK, P. A. BRCA1 16 years later: risk-associated BRCA1 mutations and their functional implications. **The FEBS Journal**, v. 277, n. 15, p. 3086-3096, ago. 2010.
LUDWIG, K. U. et al. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. **Nature Genetics**, v. 44, n. 9, p. 968-971, set. 2012.
MANGOLD, E. et al. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. **Nature Genetics**, v. 42, n. 1, p. 24-26, jan. 2010.
MOSSEY, P. A. et al. Cleft lip and palate. **The Lancet**, v. 374, n. 9703, p. 1773-1785, 21 nov. 2009.
MOSTOWSKA, A. et al. Genetic variants in BRIP1 (BACH1) contribute to risk of nonsyndromic cleft lip with or without cleft palate. **Birth Defects Research. Part A, Clinical and Molecular Teratology**, v. 100, n. 9, p. 670-678, set. 2014.
PFAFFL, M. W. A new mathematical model for relative quantification in real-time RT-PCR. **Nucleic Acids Research**, v. 29, n. 9, p. e45, 1 maio 2001.
RAKYAN, V. K. et al. Epigenome-wide association studies for common human diseases. **Nature Reviews. Genetics**, v. 12, n. 8, p. 529-541, 12 jul. 2011.
RICE, J. C.; MASSEYAAA501BROWN, K. S.; FUTSCHER, B. W. Aberrant methylation of the BRCA1 CpG island promoter is associated with decreased BRCA1 mRNA in sporadic breast cancer cells. **Oncogene**, v. 17, n. 14, p. 1807-1812, 8 out. 1998.
ROHDE, C. et al. BISMA--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. **BMC Bioinformatics**, v. 11, p. 230, 6 maio 2010.
ROY, R.; CHUN, J.; POWELL, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. **Nature Reviews. Cancer**, v. 12, n. 1, p. 68-78, 23 dez. 2011.

STANIER, P.; MOORE, G. E. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. **Human Molecular Genetics**, v. 13 Spec No 1, p. R73-81, 1 abr. 2004.
TAKAOKA, A.; TAMURA, T.; TANIGUCHI, T. Interferon regulatory factor family of transcription factors and regulation of oncogenesis. **Cancer Science**, v. 99, n. 3, p. 467-478, mar. 2008.
TAMURA, T. et al. The IRF family transcription factors in immunity and oncogenesis. **Annual Review of Immunology**, v. 26, p. 535-584, 2008.
VANDESOMPELE, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. **Genome Biology**, v. 3, n. 7, p. RESEARCH0034, 18 jun. 2002.
WILKIE, A. O.; MORRISSAAA501KAY, G. M. Genetics of craniofacial development and malformation. **Nature Reviews. Genetics**, v. 2, n. 6, p. 458-468, jun. 2001.
XAVIER, L. A. DA C. et al. Analysis of genome instability biomarkers in children with non-syndromic orofacial clefts. **Mutagenesis**, v. 32, n. 2, p. 313-321, 1 mar. 2017.
ZENTNER, G. E.; HENIKOFF, S. Regulation of nucleosome dynamics by histone modifications. **Nature Structural & Molecular Biology**, v. 20, n. 3, p. 259-266, mar. 2013.



Supplementary Figure 1: Quantification of *BRCA1*-targeting miRNAs in NSCL/P and control cells. We observed no significant differences between NSCL/P and controls samples for all four tested miRNAs ($p>0.05$, Student's t-test with Welch's correction).

Bibliography

BEATY, T. H. et al. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nature Genetics, v. 42, n. 6, p. 525-529, jun. 2010.

BEN GACEM, R. et al. Contribution of epigenetic alteration of BRCA1 and BRCA2 genes in breast carcinomas in Tunisian patients. Cancer epidemiology, v. 36, n. 2, p. 190-197, abr. 2012.

BIRGISDOTTIR, V. et al. Epigenetic silencing and deletion of the BRCA1 gene in sporadic breast cancer. Breast Cancer Research, v. 8, n. 4, p. R38, 2006.

BIRNBAUM, S. et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nature Genetics, v. 41, n. 4, p. 473-477, abr. 2009.

BOSVIEL, R. et al. BRCA1 promoter methylation in peripheral blood DNA was identified in sporadic breast cancer and controls. Cancer epidemiology, v. 36, n. 3, p. e177-82, jun. 2012.

BRITO, L. A. et al. Genetic contribution for non-syndromic cleft lip with or without cleft palate (NS CL/P) in different regions of Brazil and implications for association studies. American Journal of Medical Genetics. Part A, v. 155A, n. 7, p. 1581-1587, jul. 2011.

BRITO, L. A. et al. Region 8q24 is a susceptibility locus for nonsyndromic oral clefting in Brazil. Birth Defects Research. Part A, Clinical and Molecular Teratology, v. 94, n. 6, p. 464-468, jun. 2012a.

BRITO, L. A. et al. IRF6 is a risk factor for nonsyndromic cleft lip in the Brazilian population. American Journal of Medical Genetics. Part A, v. 158A, n. 9, p. 2170-2175, set. 2012b.

BRITO, L. A. et al. Genetics and management of the patient with orofacial cleft. Plastic surgery international, v. 2012, p. 782821, 1 nov. 2012c.

BROOKLYIN, S. et al. Assessment of folic Acid and DNA damage in cleft lip and cleft palate. Clinics and practice, v. 4, n. 1, p. 608, 27 mar. 2014.

BUENO, D. F. et al. Human stem cell cultures from cleft lip/palate patients show enrichment of transcripts involved in extracellular matrix modeling by comparison to controls. Stem Cell Reviews, v. 7, n. 2, p. 446-457, jun. 2011.

CALO, E.; WYSOCKA, J. Modification of enhancer chromatin: what, how, and why? Molecular Cell, v. 49, n. 5, p. 825-837, 7 mar. 2013.

CHANG, S.; SHARAN, S. K. BRCA1 and microRNAs: emerging networks and potential therapeutic targets. Molecules and Cells, v. 34, n. 5, p. 425-432, nov. 2012.

CLOUTHIER, D. E. et al. Signaling pathways crucial for craniofacial development revealed by endothelin-A receptor-deficient mice. Developmental Biology, v. 217, n. 1, p. 10-24, 1 jan. 2000.

COSGROVE, M. S.; BOEKE, J. D.; WOLBERGER, C. Regulated nucleosome mobility and the histone code. Nature Structural & Molecular Biology, v. 11, n. 11, p. 1037-1043, nov. 2004.

COSGROVE, M. S.; WOLBERGER, C. How does the histone code work? Biochemistry and Cell Biology (Biochimie et Biologie Cellulaire), v. 83, n. 4, p. 468-476, ago. 2005.

EZPONDA, T.; LICHT, J. D. Molecular pathways: deregulation of histone h3 lysine 27 methylation in cancer-different paths, same destination. Clinical Cancer Research, v. 20, n. 19, p. 5001-5008, 1 out. 2014.
GOU, Y.; ZHANG, T.; XU, J. Transcription factors in craniofacial development: from receptor signaling to transcriptional and epigenetic regulation. Current Topics in Developmental Biology, v. 115, p. 377-410, 6 out. 2015.
GRANT, S. F. A. et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. The Journal of Pediatrics, v. 155, n. 6, p. 909-913, dez. 2009.
IWAMOTO, T. et al. BRCA1 promoter methylation in peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. Breast Cancer Research and Treatment, v. 129, n. 1, p. 69-77, ago. 2011.
KARMODIYA, K. et al. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics, v. 13, p. 424, 24 ago. 2012.
KOBAYASHI, G. S. et al. Susceptibility to DNA damage as a molecular mechanism for non-syndromic cleft lip and palate. Plos One, v. 8, n. 6, p. e65677, 12 jun. 2013.
KONDO, S. et al. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nature Genetics, v. 32, n. 2, p. 285-289, out. 2002.
KUROSAKA, H. et al. Disrupting hedgehog and WNT signaling interactions promotes cleft lip pathogenesis. The Journal of Clinical Investigation, v. 124, n. 4, p. 1660-1671, abr. 2014.
LEENEN, F. A. D.; MULLER, C. P.; TURNER, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? Clinical epigenetics, v. 8, p. 92, 6 set. 2016.
LINGER, R. J.; KRUK, P. A. BRCA1 16 years later: risk-associated BRCA1 mutations and their functional implications. The FEBS Journal, v. 277, n. 15, p. 3086-3096, ago. 2010.
LUDWIG, K. U. et al. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nature Genetics, v. 44, n. 9, p. 968-971, set. 2012.
MANGOLD, E. et al. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nature Genetics, v. 42, n. 1, p. 24-26, jan. 2010.
MOSSEY, P. A. et al. Cleft lip and palate. The Lancet, v. 374, n. 9703, p. 1773-1785, 21 nov. 2009.
MOSTOWSKA, A. et al. Genetic variants in BRIP1 (BACH1) contribute to risk of nonsyndromic cleft lip with or without cleft palate. Birth Defects Research. Part A, Clinical and Molecular Teratology, v. 100, n. 9, p. 670-678, set. 2014.
PFAFFL, M. W. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research, v. 29, n. 9, p. e45, 1 maio 2001.
RAKYAN, V. K. et al. Epigenome-wide association studies for common human diseases. Nature Reviews. Genetics, v. 12, n. 8, p. 529-541, 12 jul. 2011.
RICE, J. C.; MASSEYAAA501BROWN, K. S.; FUTSCHER, B. W. Aberrant methylation of the BRCA1 CpG island promoter is associated with decreased BRCA1 mRNA in sporadic breast cancer cells. Oncogene, v. 17, n. 14, p. 1807-1812, 8 out. 1998.

ROHDE, C. et al. BISMA--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. BMC Bioinformatics, v. 11, p. 230, 6 maio 2010.
ROY, R.; CHUN, J.; POWELL, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nature Reviews. Cancer, v. 12, n. 1, p. 68-78, 23 dez. 2011.
STANIER, P.; MOORE, G. E. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. Human Molecular Genetics, v. 13 Spec No 1, p. R73-81, 1 abr. 2004.
TAKAOKA, A.; TAMURA, T.; TANIGUCHI, T. Interferon regulatory factor family of transcription factors and regulation of oncogenesis. Cancer Science, v. 99, n. 3, p. 467-478, mar. 2008.
TAMURA, T. et al. The IRF family transcription factors in immunity and oncogenesis. Annual Review of Immunology, v. 26, p. 535-584, 2008.
VANDESOMPELE, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology, v. 3, n. 7, p. RESEARCH0034, 18 jun. 2002.
WILKIE, A. O.; MORRIS-KAY, G. M. Genetics of craniofacial development and malformation. Nature Reviews. Genetics, v. 2, n. 6, p. 458-468, jun. 2001.
XAVIER, L. A. DA C. et al. Analysis of genome instability biomarkers in children with non-syndromic orofacial clefts. Mutagenesis, v. 32, n. 2, p. 313-321, 1 mar. 2017.
ZENTNER, G. E.; HENIKOFF, S. Regulation of nucleosome dynamics by histone modifications. Nature Structural & Molecular Biology, v. 20, n. 3, p. 259-266, mar. 2013.

Chapter 4

DNA methylation analysis at the non-syndromic cleft lip/palate associated region 8q24.21

Lucas Alvizi Cruz, Camila da Silva Bassi, Luciano Abreu Brito, Maria Rita dos Santos Passos-Bueno

Centro de Pesquisas Sobre o Genoma Humano e Células-Tronco, Instituto de Biociências, Universidade de São Paulo

Key-words: cleft lip, cleft palate, DNA methylation, 8q24.21, single nucleotide polymorphism

Abstract

Among all congenital craniofacial malformations, non-syndromic cleft lip/palate (NSCL/P) is the most common and exhibits a multifactorial pattern of inheritance. Several at-risk loci have been identified for NSCL/P by different strategies and 8q24.21 has emerged as the most replicated region by Genome-Wide Association Studies (GWAS). Accordingly, associated single nucleotide polymorphisms (SNPs) at 8q24.21 fall into a so called gene desert which function as a distal regulatory element with enhancer features for *MYC* has been demonstrated. The mechanisms by which genetic variation at 8q24.21 enhancer region influences gene expression in human, however, have not been demonstrated so far. On the other hand, DNA methylation at enhancers is a known mechanism that can interfere in gene expression leading to phenotype variation and disease. In this study we have investigated methylation differences in a 600bp region flanking the NSCL/P associated SNP at 8q24.21 rs987525 in 34 NSCL/P and 44 control blood samples. We observed this region to be highly hypermethylated both in NSCL/P and controls (methylation level>95%) and we found no significant methylation differences between those groups. Attempting to investigate if rs987525 would act as a methylation quantitative trait loci (meQTL), we correlated NSCL/P rs987525 genotypes (AA, AC or CC) with respective methylation values. No significant difference was observed, although CC group, the highest at-risk genotype, presented the lowest methylation values. In conclusion, methylation at the analysed region seems not to be a molecular mechanism involved in NSCL/P aetiology for 8q24.21 region and rs987525 variation is not suggestive to influence methylation of proximal CpGs in a meQTL level.

Resumo

Dentre todas as malformações craniofaciais, as fissuras lábio-palatinas não sindrômicas (FL/P NS) são as mais frequentes e exibem um padrão de herança multifatorial. Diversos *loci* de risco para as FL/P NS tem sido identificados por diferentes estratégias e a região 8q24.21 é a mais replicada por *Genome-Wide Association Studies* (GWAS). Desta forma, os polimorfismos de base única (*single nucleotide polymorphisms* – SNPs) na região 8q24.21 pertencem a um deserto gênico cuja função tem sido demonstrada como um elemento de regulação distal do tipo *enhancer* para o gene *MYC*. No entanto, os mecanismos pelos quais a variação genética na região do *enhancer* no 8q24.21 influenciam a expressão gênica em humanos não foram demonstrados até o momento. Por outro lado, a metilação de DNA em *enhancers* é um mecanismo que sabidamente pode interferir na expressão gênica e levar a variação fenotípica e possivelmente doenças. Neste trabalho, nós investigamos diferenças de metilação de DNA em uma região de 600bp franqueando um dos SNPs associados às FL/P NS na região 8q24.21, rs987525, em 34 amostras de sangue de FL/P NS e 44 controles. Nós observamos que esta região é altamente hipermetilada tanto em FL/P NS quanto em controles (níveis de metilação > 95%) e não encontramos diferenças significativas de metilação entre os grupos. Na tentativa de investigar se o rs987525 atuaria como um *methylation quantitative trait loci* (meQTL), nós correlacionamos os possíveis genótipos do rs987525 em amostras FL/P NS (AA, AC, CC) com seus respectivos valores de metilação. Nenhuma diferença significativa foi observada, apesar das amostras CC, genótipo o qual confere maior risco às FL/P NS, terem apresentado os menores níveis de metilação. Em conclusão, a metilação de DNA na região analisada não é um provável mecanismo envolvido no risco conferido pela região 8q24.21 para as FL/P NS e a variação do rs987525 não sugere sua influência na metilação das CpGs proximais como em um meQTL.

Introduction

Non-syndromic cleft lip/palate (NSCL/P) is the most frequent congenital craniofacial malformation, affecting 1:700 live-births worldwide (DIXON et al., 2011; WHO REGISTRY MEETING ON CRANIOFACIAL ANOMALIES (2001 : BAURU, BRAZIL) et al., 2003). NSCL/P incidence and heritability, however, varies geographically indicating that its causes also vary among populations (CARINCI et al., 2007; JUGESSUR; FARLIE; KILPATRICK, 2009). In this sense, NSCL/P is considered a multifactorial trait, with a genetic contribution estimated by heritability studies ranging from 40 to 85% depending on the studied population (BRITO et al., 2011).

Attempts to uncover NSCL/P genetic causes have been made using distinct approaches from linkage analysis to association studies and next-generation sequencing. Linkage analysis and mostly sequencing studies have been applied to the identification of rare variants especially in familial cases, in which TGFA, FOXE1, FGFR1, FGF8, MSX1, CDH1 and ARHGAP29 (BRITO et al., 2015; RILEY; MURRAY, 2007; SAVASTANO et al., 2017) have been brought up as candidate genes for NSCL/P (ARDINGER et al., 1989; MARAZITA et al., 2009; MORENO et al., 2009; VIEIRA et al., 2005). On the other hand, common variants conferring susceptibility to NSCL/P have been mostly identified by Genome-Wide Association Studies (GWAS) and the most replicated genomic region lies in a gene desert at 8q24.21, in which rs987525 emerges as the most significant SNP, including in the Brazilian population (BIRNBAUM et al., 2009; BRITO et al., 2012a; GRANT et al., 2009). Except for SNPs in *IRF6* promoter disrupting *AP2a* transcription factor interaction, NSCL/P GWAS identified variants are frequently in non-coding regions, with minimum demonstrated explanation on how such associated regions could contribute to the phenotype (BRITO et al., 2012b; GRANT et al., 2009; RAHIMOV et al., 2008). Nevertheless, those GWAS associated regions are believed to exhibit regulatory features such as enhancers, as suggested by ENCODE epigenetic data and other functional studies in cells or animal models,

which is the case of 8q24.21 NSCL/P associated region (AHMADIYEH et al., 2010; SOTELO et al., 2010; USLU et al., 2014). Moreover, in the attempt to explore molecular mechanisms involving variation at enhancer regions in disease, DNA methylation studies have demonstrated how methylation disturbances in enhancer can act and contribute to cellular phenotype (ARAN; HELLMAN, 2014; ARAN; SABATO; HELLMAN, 2013; HEYN et al., 2016), although no such study has been performed to NSCLP GWAS hits.

In this studied we evaluated whether DNA methylation differences at 8q24.21 region harboring rs987525 could be suggested as a molecular mechanism in this NSCL/P associated region and whether rs987525 genotypes could influence methylation levels of neighbour CpGs.

Methods

Patient and Control samples

All blood samples from NSCL/P patients and controls were collected after signed informed consent forms by parents or legal guardians, approved by the Ethics Committee from the Instituto de Biociências - Universidade de Sao Paulo. NSCL/P samples (n=34) were collected during Operation Smile Brazil missions, while control samples (n=44) were collected at the Centro de Pesquisas Sobre o Genoma Humano e Células-Tronco (CEGH-CEL) from Universidade de Sao Paulo. Blood samples were submitted to automated DNA extraction (Autopure LS -Gentra Systems).

Methylation analysis

One microgram of genomic DNA from each sample were submitted to bisulfite conversion using EpiTect Bisulfite Conversion Kit (QIAgen). Bilsulfite converted DNA was subsequently used in PCR for the rs987525 600bp flanking region, in which primers were designed with MethPrimer (www.urogene.org/methprimer/) (LI; DAHIYA, 2002) (Forward Primer: 5'-TAGGAGTATGAGAATAAGGTT-3'; Reverse: 5'-CAACAAATTCAATATAACTCTC-3'). Amplicons were checked by agarose electrophoresis and cloned using the TOPO TA Cloning Kit for Sequencing (ThermoFisher) and we Sanger sequenced 10 clones per sample using the BigDye® Terminator v3.1 Cycle Sequencing Kit and the ABI 3730 DNA Analyser. Sequencing files were then analysed for methylation quantification using the online tool BISMA - Bisulfite Sequencing DNA Methylation Analysis (ROHDE et al., 2010). Methylation values were computed and differences tested between groups using Student's T-Test with Welch's correction.

rs987525 SNP genotyping and meQTL analysis

NSCL/P samples were genotyped for rs987525 using TaqMan (ThermoScientific). Genotype groups (AA, AC and CC) methylation levels were compared to inspect methylation quantitative trait loci (meQTL) using a Kruskal-Wallis test (α=0.05)

Results

We first virtually inspected the rs987525 region at UCSC Genome Browser (chr8:129,945,904-129,946,404; GRCh37/hg19) and observed positive methylation signals reported from both MeDIP-seq CpG Score and MeDIP-seq Raw signal, indicating methylation sites nearby the associated SNP (Figure 1). Those methylation signals are also in superposition with DNaseI Hypersensitivity Clusters, indicating this segment as an open/active chromatin region (Figure 1).

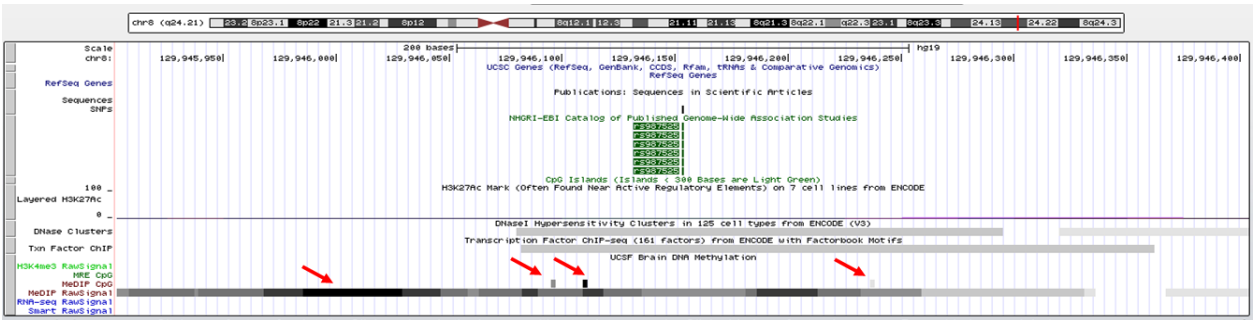


Figure 1: rs987525 region at UCSC Genome Browser (501bp window). Rs987525 SNP position is showed in green at the center. Methylation signals by MeDIP-seq CpG Score and MeDIP-seq Raw Signal are pointed by red arrows, in which black segments indicate stronger signals than gray or white segments. DNaseI Hypersensitivity Clusters are above methylation signals, displayed in a grey segment.

We next interrogated methylation status of a 600bp region flanking rs987525 (300bp upstream and 300 bp downstream) in 34 NSCL/P and 44 control samples. This 600bp region harboured 3 CpGs and, therefore, was not characterised as a CpG island due to low CpG density (observed CpG/expected CpG ratio: < 0.65). We observed this region to be highly methylated (>95% methylation level) both in NSCL/P and controls samples. When methylations levels were

compared between NSCL/P and controls, no significant methylation differences was observed (*P*=0.872, Fisher's Exact Test;Table I).

Table I: Total methylation level quantification of 3 CpG sites in the region flanking rs987525, showing no significant methylation differences between NSCL/P and control samples (p=0.872, Fisher's Exact Test).

	NSCL/P		Controls		
	Number of CpGs	%	Number of CpGs	%	*p-value*
Methylated CpGs	644	96.1	877	95.95	0.872
Non-methylated CpGs	26	3.9	37	4.05	
Total	**670**		**914**		

Further, to address whether rs987525 could act as a methylation quantitative trait loci (meQTL) to the investigated CpGs, we correlated genotypes from NSCL/P samples with methylation values. NSCL/P rs987525 genotypes consisted of 12 AA, 10 AC and 12 CC. Despite of the CC group presented highest variance and lowest methylation values, no significant methylation differences was observed among genotype groups (Figure 2) and therefore rs987525 was not a meQTL in those samples for the analysed 3 CpGs.

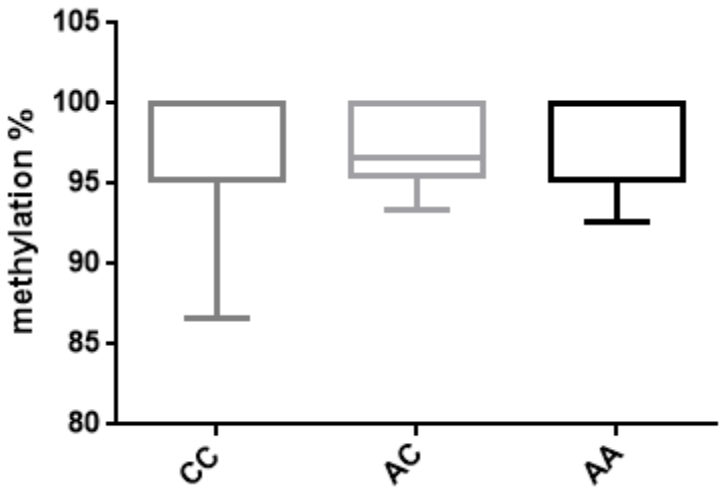


Figure 2: Methylation levels from the 3 CpGs flanking rs987525 according to rs987525 genotypes AA, AC and CC in NSCL/P samples. Mean, standard deviation and coefficient of variation for each genotype group were: AA = 97,84 , 3,00 and 3,07%; AC = 97,16 , 2,42 and 2,50%; CC = .96,67 , 4,79 and 4,96%. No significant differences or correlation was observed among genotype groups and their respective methylation levels.

Discussion

Unravelling function of non-coding variation at gene deserts on gene expression and ultimately phenotype is one goal of the genomic era, especially when several disease associated SNPs fall into such regions (MANOLIO; BROOKS; COLLINS, 2008; SCHIERDING; CUTFIELD; O'SULLIVAN, 2014). As for many of those non-coding regions, they are believed to act in a distal regulatory manner as enhancers or insulators and variation could thus affect regulation of neighbour genes (SANYAL et al., 2012). In this context, 8q24.21 NSCL/P associated region has been demonstrated to harbour enhancer elements for craniofacial development and more recently demonstrated by functional studies in human cells and mice embryos as an enhancer for *Myc*, in which disruption of such regulatory region may lead to cancer or clefts (AHMADIYEH et al., 2010; BRITO et al., 2012b; GRANT et al., 2009; SOTELO et al., 2010; USLU et al., 2014; WRIGHT; BROWN; COLE, 2010). However, how 8q24.21 variation contributes, at the molecular level, to the clefting phenotype has not been demonstrated so far. On the other hand, DNA methylation of enhancers has been demonstrated to be altered in cancer and to contribute to cell phenotype (ARAN; HELLMAN, 2014; ARAN; SABATO; HELLMAN, 2013; CALO; WYSOCKA, 2013; HEYN et al., 2016). Therefore DNA methylation and other epigenetic factors emerge as possible molecular mechanisms involved in non-coding regulatory regions as for the 8q24.21 NSCL/P associated region.

We investigated DNA methylation in a 600bp region flanking the NSCL/P associated SNP rs987525 which contains 3 CpGs and was not considered a CpG island due to low CpG density. We did not observe any significant or suggestive hyper or hypomethylation difference of total or individual CpG methylation levels between NSCL/P and control groups. Our experiments are conclusive therefore that DNA methylation in this region does not vary between NSCL/P and control samples. It is possible, however, that DNA methylation could be involved in other regions adjacent to rs987525 out of the 600bp window here analysed. Despite of no CpG island was

found in the association region, CpGs as “open sea” are spread on the genome and have the potential to impact regulatory elements (MA et al., 2013). Also, alternative epigenetic mechanisms as histone modifications are potentially involved factors for enhancer activity regulation in which were not evaluated here. Indeed, enhancers are best characterised by specific histone marks of active and open chromatine as H3K9 and H3K14 acetylation (CALO; WYSOCKA, 2013; KARMODIYA et al., 2012) and investigating variation in quantity of such marks at an enhancer region between NSCL/P and controls could indicate whether enhancer activity in one group changes in comparison to the other. We also acknowledge that enhancer activity is cell-type dependent (HEINZ et al., 2015) and since we analysed enhancer methylation only in whole-blood DNA, eventual differences would not be detected due tissue limitations.

We also investigated whether methylation in the analysed region would correlate with genotypes in a meQTL manner. We did not observe methylation differences among AA, AC and CC groups and a meQTL effect for rs987525 could not be considered in our sample. Nevertheless, we did observe the lowest methylation values in CC group, although insufficiently to cause significant differences in comparison to the other genotypes. We acknowledge, however, that observing QTLs is highly dependent on large sample sizes, and our small sample number in the meQTL analysis (n=34) could be a limiting factor to detect such effects. We thus believe that analysing a potential meQTL for rs987525 in a robust sample size is necessary to either exclude or confirm this phenomenon.

In conclusion, methylation differences in the 600pb region flanking rs987525, evaluated from blood cell DNA, seems not to be a major etiological contributor to 8q24-associated risk, suggesting alternative genetic or epigenetic mechanisms to be involved in NSCL/P susceptibility in this region.

References

AHMADIYEH, N. et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 21, p. 9742-9746, 25 maio 2010.

ARAN, D.; HELLMAN, A. Unmasking risk loci: DNA methylation illuminates the biology of cancer predisposition: analyzing DNA methylation of transcriptional enhancers reveals missed regulatory links between cancer risk loci and genes. **Bioessays: News and Reviews in Molecular, Cellular and Developmental Biology**, v. 36, n. 2, p. 184-190, fev. 2014.

ARAN, D.; SABATO, S.; HELLMAN, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. **Genome Biology**, v. 14, n. 3, p. R21, 12 mar. 2013.

ARDINGER, H. H. et al. Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. **American Journal of Human Genetics**, v. 45, n. 3, p. 348-353, set. 1989.

BIRNBAUM, S. et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. **Nature Genetics**, v. 41, n. 4, p. 473-477, abr. 2009.

BRITO, L. A. et al. Genetic contribution for non-syndromic cleft lip with or without cleft palate (NS CL/P) in different regions of Brazil and implications for association studies. **American Journal of Medical Genetics. Part A**, v. 155A, n. 7, p. 1581-1587, jul. 2011.

BRITO, L. A. et al. Region 8q24 is a susceptibility locus for nonsyndromic oral clefting in Brazil. **Birth Defects Research. Part A, Clinical and Molecular Teratology**, v. 94, n. 6, p. 464-468, jun. 2012a.

BRITO, L. A. et al. Genetics and management of the patient with orofacial cleft. **Plastic surgery international**, v. 2012, p. 782821, 1 nov. 2012b.

BRITO, L. A. et al. Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate. **Human Mutation**, v. 36, n. 11, p. 1029-1033, nov. 2015.

CALO, E.; WYSOCKA, J. Modification of enhancer chromatin: what, how, and why? **Molecular Cell**, v. 49, n. 5, p. 825-837, 7 mar. 2013.

CARINCI, F. et al. Human genetic factors in nonsyndromic cleft lip and palate: an update. **International Journal of Pediatric Otorhinolaryngology**, v. 71, n. 10, p. 1509-1519, out. 2007.

DIXON, M. J. et al. Cleft lip and palate: understanding genetic and environmental influences. **Nature Reviews. Genetics**, v. 12, n. 3, p. 167-178, mar. 2011.

Global registry and database on craniofacial anomalies. [s.d.].

GRANT, S. F. A. et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. **The Journal of Pediatrics**, v. 155, n. 6, p. 909-913, dez. 2009.

HEINZ, S. et al. The selection and function of cell type-specific enhancers. **Nature Reviews. Molecular Cell Biology**, v. 16, n. 3, p. 144-154, mar. 2015.

HEYN, H. et al. Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. **Genome Biology**, v. 17, p. 11, 26 jan. 2016.

JUGESSUR, A.; FARLIE, P. G.; KILPATRICK, N. The genetics of isolated orofacial clefts: from genotypes to subphenotypes. **Oral Diseases**, v. 15, n. 7, p. 437-453, out. 2009.
KARMODIYA, K. et al. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. **BMC Genomics**, v. 13, p. 424, 24 ago. 2012.
LI, L.-C.; DAHIYA, R. MethPrimer: designing primers for methylation PCRs. **Bioinformatics**, v. 18, n. 11, p. 1427-1431, nov. 2002.
MA, X. et al. DNA methylation data analysis and its application to cancer research. **Epigenomics**, v. 5, n. 3, p. 301-316, jun. 2013.
MANOLIO, T. A.; BROOKS, L. D.; COLLINS, F. S. A HapMap harvest of insights into the genetics of common disease. **The Journal of Clinical Investigation**, v. 118, n. 5, p. 1590-1605, maio 2008.
MARAZITA, M. L. et al. Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype-specific differences in linkage and association results. **Human Heredity**, v. 68, n. 3, p. 151-170, 11 jun. 2009.
MORENO, L. M. et al. FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. **Human Molecular Genetics**, v. 18, n. 24, p. 4879-4896, 15 dez. 2009.
RAHIMOV, F. et al. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. **Nature Genetics**, v. 40, n. 11, p. 1341-1347, nov. 2008.
RILEY, B. M.; MURRAY, J. C. Sequence evaluation of FGF and FGFR gene conserved non-coding elements in non-syndromic cleft lip and palate cases. **American Journal of Medical Genetics. Part A**, v. 143A, n. 24, p. 3228-3234, 15 dez. 2007.
ROHDE, C. et al. BISMA--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. **BMC Bioinformatics**, v. 11, p. 230, 6 maio 2010.
SANYAL, A. et al. The long-range interaction landscape of gene promoters. **Nature**, v. 489, n. 7414, p. 109-113, 6 set. 2012.
SAVASTANO, C. P. et al. Impact of rare variants in ARHGAP29 to the etiology of oral clefts: role of loss-of-function vs missense variants. **Clinical Genetics**, v. 91, n. 5, p. 683-689, maio 2017.
SCHIERDING, W.; CUTFIELD, W. S.; O'SULLIVAN, J. M. The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell. **Frontiers in genetics**, v. 5, p. 39, 18 fev. 2014.
SOTELO, J. et al. Long-range enhancers on 8q24 regulate c-Myc. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 7, p. 3001-3005, 16 fev. 2010.
USLU, V. V. et al. Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. **Nature Genetics**, v. 46, n. 7, p. 753-758, jul. 2014.
VIEIRA, A. R. et al. Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. **PLoS Genetics**, v. 1, n. 6, p. e64, 2 dez. 2005.
WRIGHT, J. B.; BROWN, S. J.; COLE, M. D. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. **Molecular and Cellular Biology**, v. 30, n. 6, p. 1411-1420, mar. 2010.

Chapter 5

Differential methylation is associated with non-syndromic cleft lip and palate and contributes to penetrance effects

Lucas Alvizi[1], Xiayi Ke[2], Luciano Abreu Brito[1], Rimante Seselgyte[2], Gudrun E. Moore[2], Philip Stanier[2], Maria Rita Passos-Bueno[1].

1- Centro de Pesquisas Sobre o Genoma Humano e Células-Tronco, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil.

2- Genetics and Genomic Medicine, Institute of Child Health, University College of London, London, UK.

Key-words: cleft lip and palate, craniofacial malformation, epigenomics, methylome profiling

This chapter was published in Scientific Reports – Nature: DOI: 10.1038/s41598-017-02721-0

Abstract

Non-syndromic cleft lip and/or palate (NSCL/P) is a common congenital malformation with a multifactorial model of inheritance. Although several at-risk alleles have been identified, they do not completely explain the high heritability. We postulate that epigenetic factors as DNA methylation might contribute to this missing heritability. Using a Methylome-wide association study in a Brazilian cohort (67 NSCL/P, 59 controls), we found 578 methylation variable positions (MVPs) that were significantly associated with NSCL/P. MVPs were enriched in regulatory and active regions of the genome and in pathways already implicated in craniofacial development. In an independent UK cohort (171 NSCL/P, 177 controls), we replicated 4 out of 11 tested MVPs. We demonstrated a significant positive correlation between blood and lip tissue DNA methylation, indicating blood as a suitable tissue for NSCL/P methylation studies. Next, we quantified *CDH1* promoter methylation levels in *CDH1* mutation-positive families, including penetrants, non-penetrants or non-carriers for NSCL/P. We found methylation levels to be significantly higher in the penetrant individuals. Taken together, our results demonstrated the association of methylation at specific genomic locations as contributing factors to both non-familial and familial NSCL/P and altered DNA methylation may be a second hit contributing to penetrance.

Resumo

As fissuras lábio-palatinas não sindrômicas (FL/P NS) são malformações craniofaciais congênitas comuns com um modelo de herança multifatorial. Apesar de vários alelos de risco terem sido identificados, eles ainda não explicam completamente a alta herdabilidade observada. Nós postulamos que fatores epigenéticos, como metilação de DNA, podem contribuir para a herdabilidade perdida. Usando-se um estudo de associação de varredura de metiloma (*Methylome-wide association study*) em uma coorte brasileira (67 amostras FL/P NS e 59 controles), nós encontramos 578 sítios de metilação variável (*methylation variable positions* – MVPs) significativamente associados às FL/P NS. Estas MVPs estão enriquecidas em regiões ativas e regulatórias do genoma e também em vias já implicadas no desenvolvimento craniofacial. Em uma coorte independente do Reino Unido (171 amostras FL/P NS e 177 controles), nós replicamos 4 de 11 MVPs testados. Nós também demonstramos uma correlação positiva e significante entre os níveis de metilação em sangue e tecido de lábio, indicando que amostras de sangue são apropriadas para o estudo de metilação em FL/P NS. Em seguida, nós quantificamos os níveis de metilação do promotor do *CDH1* em famílias segregando mutações nesse gene, incluindo indivíduos penetrantes, não-penetrantes e não portadores da mutação para FL/P NS. Nós encontramos que os níveis de metilação no promotor de *CDH1* estão significativamente maiores em indivíduos penetrantes. Em conjunto, nossos resultados demonstram a associação de metilação de DNA em sítios específicos do genoma como fatores contribuintes para tanto FL/P NS não familial quanto familial e que alterações de metilação de DNA podem atuar como segundo *hit* contribuindo para penetrância.

Introduction

Craniofacial development is a tightly regulated event that requires expression of many genes at a precise space-temporal specificity. These processes are carefully orchestrated by several well-documented molecular pathways, including FGF, BMP, TGF-β, SHH and WNT signalling pathways. Interference in the regulation of these pathways is known to lead to abnormal phenotypes affecting the face and cranium [1–3]. Regulation of these pathways is further complicated by interaction between genetic and environmental factors such that disturbance to either may result in craniofacial malformation, including non-syndromic forms of cleft lip and/or palate (NSCL/P).

NSCL/P is one of the most common congenital malformations in humans, affecting 1:700 live-births worldwide[4]. In broad terms, NSCL/P is considered to follow a multifactorial model[5,6], which has been supported by heritability studies with a genetic contribution estimated to vary from 45% to 85% depending on the population[7]. Common and rare variants identified through genomic analysis have successfully revealed several at-risk cleft alleles in distinct populations[8–12]. However, as for many other common diseases, the total sum of known variants still only explains a small percentage of the genetic contribution and NSCL/P falls within the missing heritability category[12,13]. Studies on the environmental contribution for NSCL/P have mostly involved epidemiology of maternal exposures to factors such as malnutrition, alcohol, tobacco, folate and anti-epileptic drugs[14–19], often with little explanation as to how those agents might disrupt molecular and cellular mechanisms that ultimately lead to a clefting phenotype. One possible explanation is their effect on epigenetic mechanisms, particularly methylation of DNA or histones. The contribution of differential DNA methylation to disease states can be assessed using genome wide analysis of CpG dinucleotides and MethWAS has been successfully applied in disease such as diabetes, rheumatoid arthritis and schizophrenia[20–24]. In this study, we have investigated the epigenetic contribution to NSCL/P by performing a MethWAS, first in a

Brazilian cohort and followed by a replication study in an independent cohort from the UK. We also compare methylation levels detected in blood to those from matched lip tissues obtained during surgery, in order to investigate differences arising from alternative, and perhaps more developmentally relevant tissues. We have further investigated methylation differences in familial NSCL/P displaying incomplete penetrance upon *CDH1* mutations in order to verify whether methylation differences correlate to phenotype penetrance.

Results

NSCL/P presents a distinct methylation signature enriched in regions of open and active chromatin

In this study, we therefore performed a MethWAS using a Brazilian age-matched cohort of 67 NSCL/P and 59 control samples using the Infinium Human Methylation 450K platform (Illumina). We found 578 MVPs at the genomic level of significance ($p<10^{-7}$, FDR adjusted) (Supplementary Table I), suggesting a different methylation signature in NSCL/P samples. Next all 578 MVPs were subjected to an exploratory process in which a significant enrichment of MVPs was found for regions of open and active chromatin as marked by H3K4me1, H3K4me3, DNaseI sites and Gene Promoters (Figure 1 and Supplementary Figure 1). Sixty-nine percent of MVPs belonged to promoters, in comparison to 47% from the 450K array, which represented a significant enrichment (p<0,0001, Chi-square test with Yates correction). Co-methylation of the identified MVPs was verified by comparing to their neighbouring CpG sites in a 500bp window, both up and downstream. Sixty percent of these CpGs had the same methylation as the MVPs (Fisher´s exact test, p=0.0043). When selected MVPs with methylation differences greater than 7% were chosen, the co-methylation rate was increased to 70%. Next MVPs belonging to genomic regions and genes that

either associated with NSCL/P or were involved in some aspect of craniofacial development were analysed (list from Jugessur et al, 2009[25]). Eighteen MVPs were found in candidate genes or regions for CL/P representing a significant enrichment in clefting-associated regions (Fisher´s exact test, p=0.00044). As a final step of the exploratory analysis, Ingenuity Pathway Analysis (IPA, QIAGen) was used to look for enriched canonical pathways in the 578 MVPs. Among the 5 top-ranked canonical pathways were the "Regulation of the Epithelial-Mesenchymal Transition Pathway" and "WNT Beta-Catenin Signaling" (Figure 2). Both pathways are strongly associated with craniofacial development and also CL/P etiology[26–30]. Altogether, these results suggested the robustness of our data, and encouraged us to move forward with a validation step.

Methylation differences replicates in an independent cohort

To validate the MethWAS findings, a replication study was carried out using an independent cohort of individuals with non-familial NSCL/P recruited in the London area in the UK, which represented an ethnically and environmentally distinct population. Selection of candidate MVPs was based on the following combined criteria: the MVP with the lowest p-value and highest methylation difference (*C11orf58* MVP, probe ID cg10633981), two MVPs with high methylation differences (*chr1* MVP, probe ID cg15897635 and *chr17* MVP, probe ID cg09319020) and eight MVPs in candidate regions/genes for NSCL/P (*FAT1* MVP, probe ID cg00405769; *FGF8* MVP, probe ID cg11706469; *FGFR1* MVP, probe ID cg20913106; *MYC* MVP, probe ID cg00611675; *PVRL1* MVP, probe ID cg06391300; *WHSC1* MVP, probe ID cg03150409; *WNT2B MVP*, probe ID cg11806528 and *WNT7A* MVP, probe ID cg13602813. In total, 11 MVPs were selected for replication in 171 NSCL/P and 177 age and ethnicity-matched controls samples using a Bisulfite Amplicon Sequencing (BSAS) approach. In the replication cohort, significant differential methylation was

found for 4 out of the 11 candidate MVPs (*chr1* MVP p=0.03; *FAT1* MVP p=0.0002; *MYC* MVP p<0.0001; *WHSC1* MVP p=0.04) (Table I). Two of these (*chr1* MVP and *WHSC1* MVP) presented the opposite methylation direction than in the methylome analysis. All mean methylation values in the case group (NSCL/P) were lower than those in the control group, even in non-significant MVPs, suggesting an apparent hypomethylation in the NSCL/P group. In order to test if the overall methylation in the NSCL/P group compared to controls was significantly reduced, methylation values from all CpG sites in the 11 analysed amplicons covered by the BSAS were investigated. No significant differences were found between methylation means from all tested amplicons (p=0.7131, mean in NSCL/P = 0.3377 ± 0.008810; mean in controls = 0.3423 ± 0.008716) (Supplementary Figure 2). Our results suggest that differential methylation is specific at those tested MVPs and not a generalised hypomethylation in those amplicons in the UK NSCL/P cohort.

Methylation in blood correlates to methylation in lip tissue DNA

Both the MethWAS and the replication study were performed using whole-blood DNA. Clearly blood is not closely representative of the tissue types directly affected by the clefting phenotype during craniofacial development. However, it is not possible to obtain tissue samples at the time of lip formation from affected individuals. Acknowledging these deficiencies, we instead collected lip tissue samples from the time of first cleft repair surgery (n = 18) as the best available alternative. To examine possible tissue specific variability, we tested how well blood and lip tissues correlated for methylation findings. BSAS methylation values from the 11 tested amplicons were obtained in paired lip tissue samples and whole-blood DNA from 18 individuals. A linear regression analysis was performed to evaluate the correlation between the two sample types. This resulted in high and significant similarity (R-square=0.9028, p<0.0001) (Figure 3).

Promoter hypermethylation correlates to penetrance in familial NSCL/P

We have recently shown that missense pathogenic mutations in *CDH1* play a role in the aetiology of NSCL/P, particularly familial cases. However, *CDH1* mutations alone do not seem to be the only causative factor, as incomplete penetrance (~60%) was observed in these families[31]. We, therefore, hypothesised whether methylation could represent a second hit. Methylation levels of 33 CpG sites at the *CDH1* promoter in penetrant (n=8), non-penetrant (n=7) and non-carrier individuals (n=3) were quantified using targeted bisulfite sequencing. *CDH1* promoter methylation was found to be higher in penetrant individuals than in non-penetrant and non-carriers (p=0,0112, Kruskal-Wallis test, Figure 4), suggesting methylation as a possible second hit to explain the cleft phenotype.

Discussion

We hypothesised that NSCL/P has a different epigenetic signature compared to controls. Indeed, after correcting our data for batch and confounding effects, our preliminary findings showed that we could detect significant enrichment of MVPs, which mostly lie in potential regulatory and active regions of the genome. We were also able to demonstrate a co-methylation pattern within our data which is compatible with methylation variation reported elsewhere[32]. Also, we showed that our MVPs were significantly enriched at NSCL/P candidate regions and in canonical pathways reported to have relevance for the malformation such as WNT-β-catenin signalling and Regulation of the Epithelial-Mesenchymal-Transition pathway[26–30]. These results suggested the robustness of our data, and encouraged us to move forward with a validation step.

Replication of the significant methylation differences (in the same direction of difference) for 2 MVPs (*FAT1* and *MYC*) strongly suggests that methylation at those sites may well have importance for NSCL/P in both populations. Indeed, the *MYC* locus represents a very strong candidate region, since it has been replicated by several NSCL/P GWASs[9,10,33], including patients from the Brazilian population[12]. The SNP at 8q24.21 region lies in an enhancer for *Myc* where disruption of the syntenic region in a murine model has been shown to result in clefts[34]. The *MYC* MVP we studied is located at the *MYC* promoter, which may potentially interfere with *MYC* expression levels. With respect to the *FAT1* locus, this gene encodes an atypical cadherin and is associated with the planar-cell-polarity pathway (PCP), which has been suggested to be important for craniofacial development[35,36]. Differential methylation was also found for 2 candidate MVPs that were not in the same methylation direction as in the initial study (*chr1 MVP* and *WHSC1 MVP*). The chr1 MVP, selected by its large methylation difference in the Brazilian cohort, falls into a gene desert (1q41). We do not know if this is a regulatory region for genes in the vicinity, however, the closest gene *MARK1*, is involved in microtubule array and regulation of cell polarity and cell shape[37]. At the *WHSC1* MVP is located the *WHSC1* promoter for the Wolf-Hirschhorn syndrome candidate 1 gene. Interestingly this syndrome is characterised by multiple closure defects including cleft lip and cleft palate (OMIM#194190). Despite the inversion in the methylation direction, general methylation disturbances at those loci are possible mechanisms and either hypo- or hypermethylation could contribute to the phenotype. Therefore, we were able to replicate 36% of the tested MVPs in the UK cohort, despite the fact that these two populations are ethnically different and exposed to distinct environments. Taking into account the large social differences and potential genetic background variation between Brazil and the UK, we considered the validation of those MVPs as a high rate of concordance, which suggests a potential role for those methylation sites in the NSCL/P etiology. Our results also suggest that there might be either shared environmental factors in these two populations or even distinct

environmental factors leading to common epigenetic changes. In this regard, we were able to compare our 578 MVPs with 2965 sites differentially methylated associated to smoking during pregnancy[38] and found two MVPs in common (cg14087168 and cg26015973), suggesting smoking as a contributing but not major environmental factor for our findings.

We also showed DNA methylation correlating to NSCL/P penetrance in families displaying *CDH1* mutations. This would be the expected mechanism based on studies with *CDH1* mutated gastric cancer in which abnormal cell behaviour leading to cancer depends on both mutation in one allele and hypermethylation of the other[39,40]. The same phenomena is true, for example, in the A/WySn mouse strain which exhibits spontaneous cleft lip/palate in 15% of neonates in which the *Wnt9b* gene is mutated on one allele and methylation of the other acts as a metastable epiallele, thus leading to the clefting phenotype[41]. Epigenetic mechanisms therefore may well play a role in the clinical expression of NSCL/P for both familial and non-familial NSCL/P.

It is notable that we could detect significant methylation differences both in non-familial and familial cases, despite investigating tissues from a different cell lineage from the malformation. It is possible that such differences may have been more pronounced in biologically relevant lip or palatal tissue DNA. Moreover, as we are studying a developmental defect from early gestation using postnatal tissues, we also realise that methylation differences and levels might not accurately reflect the relevant tissues at the time of lip and palatal formation. However, the methylation differences identified here might remain as a mark that occurred during development. Indeed, the biological significance of our findings are supported by the high correlation between methylation sites in blood and lip tissue DNA, and suggest that blood can serve as a proxy tissue for methylation studies involving craniofacial malformations.

In the longer term, it will be very important to better understand the function of the detected methylation differences. We can speculate that one possible effect might be to confer subtle differences on gene expression, which contributes to the

phenotype. However, since mRNA is not available from the relevant samples, it is not possible in this study to examine for a potential correlation between mRNA levels and DNA methylation. Therefore, changes at the detected MVPs should be considered a marker of NSCL/P, requiring further exploration to determine a direct or indirect role in the pathogenesis.

Although there are various epigenetic mechanisms known to be involved in gene regulation and disease, in this study we chose to analyse DNA methylation due to its well-known effect and the availability of suitable platforms previously developed for this approach. However, we acknowledge that other epigenetic mechanisms such as histone modifications could also be involved in NSCL/P aetiology. Nevertheless, our approach also makes a good starting point because of the known correlation between DNA methylation and some histone marks across the human genome, especially in unmethylated regions of DNA, which tend to be enriched for acetylated histones whereas methylated regions tend to lack acetylated histones and correlate to H3K9 methylation[42,43]. Additionally, environmental factors associated with NSCL/P have also been reported to interfere with histone acetylation, methylation and phosphorylation[44–47]. Therefore, in the future it will be interesting to examine if the methylation differences we report here may also indicate alterations in histone modifications within the same regions.

In summary, we found a number of differentially methylated sites associated with NSCL/P, providing evidence that epigenetic factors may play a role in the aetiology of this malformation. We also provide evidence that DNA methylation may represent a second hit in individuals with *CDH1* mutations. Further studies investigating how environmental factors interfere with the normal methylation pattern at these and other sites will be important. Investigation of how different methylation levels might functionally contribute to the clefting phenotype, such as misregulation of normal gene expression, will be required to explain our findings and to provide therapeutic targets in the future.

Material and Methods

Ethics

This study was approved by the Ethics Committee of the Instituto de Biociências (Universidade de São Paulo, Brazil) and Great Ormond Street Hospital for Children NHS Trust Ethics Committee. Biological samples were collected after signed informed consent by the parents or legal guardians. All experiments were performed in accordance with relevant guidelines and regulations.

Patients and controls

The MethWAS cohort included 126 Brazilian samples, in which 67 were cases from non-familial NSCL/P individuals (males=37, females=30; average age at sampling=5.29±0.53 yrs) and 59 age and sex-matched controls from healthy individuals (males=28, females=31, average age at sampling=6.45±0.51 yrs). Brazilian samples were ascertained either at the Hospital das Clínicas of Universidade de São Paulo (São Paulo, Brazil), Centro de Pesquisas Sobre o Genoma Humano e Células-Tronco of Universidade de São Paulo (São Paulo, Brazil) or during missions or during missions of Operation Smile Brazil, in the Brazilian state of Ceará. The replication cohort was composed by 348 samples from the UK, in which 171 were non-familial NSCL/P (males=107, females=64; average age at sampling=3.30±0.46 yrs) and 177 age and sex-matched non-cleft controls with no family history of CLP (males=100, females=77; average age at sampling=3.96±0.01 yrs) recruited from the North East Thames Regional Genetics Service. Populations from both cohorts are genetically admixed, in which applying the concept of race is difficult. However, both populations are largely enriched for white European ancestry (~70%). Whole-blood DNA was extracted from all samples by the North Thames Regional Genetics Service. In addition to blood, lip tissue samples recovered as discard material at time of the first corrective

surgery were stored in RNAlater (Thermo Scientific, USA) and available for 18 of these patients. DNA and RNA were extracted from these tissues using TRIzol Reagent (Thermo Scientific, USA) following the manufacturer's protocol.

MethWAS analysis

Samples from the MethWAS cohort (Brazilian cohort) were subjected to bisulfite conversion using 1ug of DNA in the EpitectBisulfite Kit (QIAGen). The samples were analysed on the Illumina HumanMethylation 450K Bead-Array platform (Illumina) according to the recommendations at the sequencing and array facility DeoxiBiotecnologia (Araçatuba, Brazil). Because samples were split into 3 batches for hybridization, we re-submitted 12 samples (6 NSCL/P, 6 controls) (here called as re-batch samples) to identify and correct for batch-effects between runs. Analysis consisted of two major steps: first to identify batch-effect markers and a second for differential methylation analysis. In the first step we compared the 12 paired samples, initial batch versus re-batch, using a t-test aiming to find differentially methylated sites. As the same sample was compared in two different batches, methylation differences could be assumed to be due to either batch or confounding effects. Selecting differentially methylated sites at the genomic level (t-test, $p<10^{-7}$), 270 sites were identified and these sites were then used as batch-effect markers for data correction. Using an interactive ranking method, which selected the lowest p-values among those 270 batch-effect markers, loci independent effects were identified in a total of 5 sites (cg20154084, cg00196012, cg23948080, cg22345647, cg09307868). When these sites were subsequently used as covariates for p-value correction, none of the remaining 265 sites showed significant p-values. Therefore these 5 sites were chosen as batch-effect markers and used as covariates for the next analysis step. The second step consisted of filtering, normalisation and differential methylation analysis using the RnBeads pipeline[48]. We filtered out probes affected by SNPs, on sex chromosomes, probes with a p-value detection >0.05 (Greedycut), probes with non-CpG methylation

pattern and probes from sites with a p-value >0.01 in the paired batch analysis. Data was normalised using the SWAN method. Differential methylation analysis was performed using the RefFreeEWAS method implemented in the RnBeads pipeline, which is a LIMMA based method and corrects for blood cellular heterogeneity. We also used sex, age and the 5 batch-effect markers as covariates for differential methylation analysis correction. We selected as MVPs those probes with differential methylation p-value $<10^{-7}$ after FDR adjustment.

Investigation of MVPs in a replication cohort

To quantify methylation at candidates MVPs for replication, we used the Bisulfite Amplicon Sequencing (BSAS) method, which relies on bisulfite PCR, library preparation and DNA sequencing with a NGS sequencer[49]. In total, we selected 11 MVPs and designed bisulfite-specific PCR primers using the online tool MethPrimer (http://www.urogene.org/methprimer/). Primers sequences are available under request. Samples from the replication cohort and palatal tissue sample DNAs were submitted for bisulfite conversion using 2 ug of DNA in the e EZ-96 Methylation Kit (Zymo Research). Converted DNA was used as a template for bisulfite-specific PCR for each primer pair with the HotStartTaq Plus (QIAGen) standard protocol and products were checked by agarose gel electrophoresis. All 11 products were pooled per sample and sizes were checked using a DNA TapeStation prior to library preparation. During the library preparation indexes were added in one PCR step for each pooled sample (Access Array Barcode Library, Fluidigm). Libraries were purified by Ampure XP Beads in a magnetic column and checked again in the DNA TapeStation for peak shift visualization.Finally libraries were submitted for sequencing with the MiSeq Reagent V2 Kit 250bp pair-endedrun on a MiSeq Sequencer (Illumina). We performed de-multiplexing of sequences using the FASTX Barcode Splitter program in the FastX Toolkit R package (http://hannonlab.cshl.edu/fastx_toolkit/). Following this, we filtered

out reads of low quality, selecting only reads with at least 50% of bases with Q>30 using the FASTQ Quality Filter program, also part of the FastX Toolkit R package. Next FASTQ files were converted to FASTA files using the FASTQ-to-FASTA program in the same package. For the quantification of methylation levels in the 11 candidate MVPs we used the BiQAnalyzer HT software[50]. The BiQAnalyzer HT also has some additional quality filters including sequence identity, conversion rate and gaps allowed in CpG sites. In this case we set the parameters to a minimal reference sequence identity to 90%, a minimal bisulfite conversion rate of 90% and a maximum of 10% gaps allowed in CpG sites. In order to achieve a high sensitivity, we also accepted samples with a minimum of 1000 reads after all filtering steps. Following these parameters, analysis was conducted per gene-region in all samples and we performed a comparison of methylation levels in each MVP site between NSCL/P and control groups using an unpaired two-tailed t-test with Welch's correction, for differential methylation analysis with level of accepted significance =<0.05.

Correlation of methylation in lip tissue and blood

DNA samples from lip tissue as well as whole blood were available from the same patients and were included in the BSAS work-flow. A linear regression analysis was performed to evaluate methylation differences between lip tissue and whole-blood at each of the tested sites (11 MVPs).

CDH1 promoter methylation in familial NS CL/P

DNA samples from two families segregating CDH1 mutations with NS CL/P previously described by Brito et al, 2015[31] were used for CDH1 promoter methylation quantification by conventional bisulfite sequencing. Those samples included 8 affected individuals, 7 non-penetrant individuals and 3 non-carriers. DNAs were submitted to bisulfite conversion with the EpiTect Bisulfite Kit (QIAGen) and PCR using primers for the bisulfite converted CDH1 promoter, covering 33 CpGs. Amplicons were cloned

using the TOPO Cloning Kit (ThermoFisher) and 32 clones per sample were submitted to Sanger Sequencing. Sequencing results were analysed with the online tool BISMA (http://services.ibc.uni-stuttgart.de/BDPC/BISMA/) and the total methylation values were compared among affected, non-penetrant and non-carriers groups using Kruskal-Wallis test. Level of significance accepted = <0.05.

References

1. Wilkie, A. O. & Morriss-Kay, G. M. Genetics of craniofacial development and malformation. *Nat Rev Genet* **2,** 458–468 (2001).
2. Clouthier, D. E. *et al.* Signaling Pathways Crucial for Craniofacial Development Revealed by Endothelin-A Receptor-Deficient Mice. *Dev. Biol.* **217,** 10–24 (2000).
3. Gou, Y., Zhang, T. & Xu, J. *Transcription Factors in Craniofacial Development: From Receptor Signaling to Transcriptional and Epigenetic Regulation. Current topics in developmental biology* **115,** (Elsevier Inc., 2015).
4. Organization, W. H. Global registry and database on craniofacial anomalies Main editors : *Hum. Genet.* 4–6 (2001).
5. Stanier, P. & Moore, G. E. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Hum. Mol. Genet.* **13 Spec No,** R73–R81 (2004).
6. Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J. & Shaw, W. C. Cleft lip and palate. *Lancet* **374,** 1773–1785 (2009).
7. Brito, L. A. *et al.* Genetic contribution for non-syndromic cleft lip with or without cleft palate (NS CL/P) in different regions of Brazil and implications for association studies. *Am. J. Med. Genet. Part A* **155,** 1581–1587 (2011).
8. Grant, S. F. a *et al.* A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J. Pediatr.* **155,** 909–

913 (2009).

9. Mangold, E. *et al.* Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet.* **42,** 24–6 (2010).

10. Beaty, T. H. *et al.* A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat. Genet.* **42,** 525–9 (2010).

11. Brito, L. A. *et al.* IRF6 is a risk factor for nonsyndromic cleft lip in the Brazilian population. *Am. J. Med. Genet. Part A* **158 A,** 2170–2175 (2012).

12. Brito, L. A., Meira, J. G. C., Kobayashi, G. S. & Passos-Bueno, M. R. Genetics and management of the patient with orofacial cleft. *Plast. Surg. Int.* **2012,** 782821 (2012).

13. Schierding, W., Cutfield, W. S. & O'Sullivan, J. M. The missing story behind Genome Wide Association Studies: Single nucleotide polymorphisms in gene deserts have a story to tell. *Front. Genet.* **5,** 1–7 (2014).

14. Jia, Z. L. *et al.* Maternal malnutrition, environmental exposure during pregnancy and the risk of non-syndromic orofacial clefts. *Oral Dis.* **17,** 584–589 (2011).

15. Acuña-González, G. *et al.* Family history and socioeconomic risk factors for non-syndromic cleft lip and palate: a matched case-control study in a less developed country. *Biomedica* **31,** 381–91 (2011).

16. Meyer, K. a., Williams, P., Hernandez-Diaz, S. & Cnattingius, S. Smoking and the Risk of Oral Clefts. *Epidemiology* **15,** 671–678 (2004).

17. Puh??, E. H., Szunyogh, M., M??tneki, J. & Czeizel, A. E. Drug treatment during pregnancy and isolated orofacial clefts in Hungary. *Cleft Palate-Craniofacial J.* **44,** 194–202 (2007).

18. DeRoo, L. a, Wilcox, A. J., Drevon, C. a & Lie, R. T. First-trimester maternal alcohol consumption and the risk of infant oral clefts in Norway: a population-based case-control study. *Am. J. Epidemiol.* **168,** 638–46 (2008).

19. Deck, G. M., Nadkarni, N., Montouris, G. D. & Lovett, A. Congenital malformations in infants exposed to antiepileptic medications in utero at Boston Medical Center from 2003 to 2010. *Epilepsy Behav.* **51,** 166–169 (2015).

20. Chambers, J. C. *et al.* Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: A nested case-control study. *Lancet Diabetes Endocrinol.* **3,** 526–534 (2015).

21. Florath, I. *et al.* Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia* **59,** 130–138 (2016).

22. Soriano-Tárraga, C. *et al.* Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum. Mol. Genet.* **25,** 1–11 (2015).

23. Dempster, E. L. *et al.* Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum. Mol. Genet.* **20,** 4786–4796 (2011).

24. Wockner, L. F. *et al.* Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Transl. Psychiatry* **4,** e339 (2014).

25. Jugessur, a, Farlie, P. G. & Kilpatrick, N. The genetics of isolated orofacial clefts: from genotypes to subphenotypes. *Oral Dis.* **15,** 437–53 (2009).

26. Lan, Y. *et al.* Expression of Wnt9b and activation of canonical Wnt signaling during midfacial morphogenesis in mice. *Dev. Dyn.* **235,** 1448–1454 (2006).

27. Song, L. *et al.* Lrp6-mediated canonical Wnt signaling is required for lip formation and fusion. *Development* **136,** 3161–3171 (2009).

28. Menezes, R. *et al.* Studies with Wnt genes and nonsyndromic cleft lip and palate. *Birth Defects Res. Part A - Clin. Mol. Teratol.* **88,** 995–1000 (2010).

29. Ke, C.-Y., Xiao, W.-L., Chen, C.-M., Lo, L.-J. & Wong, F.-H. IRF6 is the mediator of TGFβ3 during regulation of the epithelial mesenchymal transition and palatal

fusion. *Sci. Rep.* **5,** 12791 (2015).

30. Kang, P. & Svoboda, K. K. H. CRITICAL REVIEWS IN ORAL BIOLOGY & MEDICINE Epithelial-Mesenchymal Transformation during Craniofacial Development. *Crit. Rev. Oral Biol. Med.* 678–690 (2005).
31. Brito, L. A. *et al.* Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate. *Hum. Mutat.* **36,** 1029–1033 (2015).
32. Finer, S. *et al.* Maternal gestational diabetes is associated with genome-wide DNA methylation variation in placenta and cord blood of exposed offspring. *Hum. Mol. Genet.* **24,** 3021–9 (2015).
33. Ludwig, K. U. *et al.* Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat. Genet.* **44,** 968–971 (2012).
34. Uslu, V. V. *et al.* Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat. Genet.* **46,** 1–24 (2014).
35. Topczewski, J., Dale, R. M. & Sisson, B. E. Planar cell polarity signaling in craniofacial development. *Organogenesis* **7,** 255–9 (2011).
36. Yang, T. *et al.* Analysis of PRICKLE1 in human cleft palate and mouse development demonstrates rare and common variants involved in human malformations. *Mol. Genet. genomic Med.* **2,** 138–51 (2014).
37. Macara, I. G. Parsing the polarity code. *Nat. Rev. Mol. Cell Biol.* **5,** 220–231 (2004).
38. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98,** 680–696 (2016).
39. Grady, W. M. *et al.* Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nat. Genet.* **26,** 16–17 (2000).
40. Machado, J. C. *et al.* E-cadherin gene (CDH1) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. *Oncogene* **20,** 1525–1528

(2001).

41. Juriloff, D. M., Harris, M. J., Mager, D. L. & Gagnier, L. Epigenetic Mechanism Causes Wnt9B Deficiency and Nonsyndromic Cleft Lip and Palate in the A/WySn Mouse Strain. *Birth Defects Res. Part A - Clin. Mol. Teratol.* **100,** 772–788 (2014).

42. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **16,** 519–532 (2015).

43. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10,** 295–304 (2009).

44. Pan, B. *et al.* Alcohol Consumption During Gestation Causes Histone3 Lysine9 Hyperacetylation and an Alternation of Expression of Heart Development-Related Genes in Mice. *Alcohol. Clin. Exp. Res.* **38,** 2396–2402 (2014).

45. Ungerer, M., Knezovich, J. & Ramsay, M. In utero alcohol exposure, epigenetic changes, and their consequences. *Alcohol Res* **35,** 37–46 (2013).

46. Sundar, I. K., Nevid, M. Z., Friedman, A. E. & Rahman, I. *Cigarette smoke induces distinct histone modifications in lung cells: Implications for the pathogenesis of COPD and lung cancer*. *Journal of Proteome Research* **13,** (2014).

47. Krämer, O. H. *et al.* The histone deacetylase inhibitor valproic acid selectively induces proteasomal degradation of HDAC2. *EMBO J.* **22,** 3411–3420 (2003).

48. Assenov, Y. *et al.* Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* **11,** 1138–40 (2014).

49. Masser, D. R., Berg, A. S. & Freeman, W. M. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenetics Chromatin* **6,** 1 (2013).

50. Lutsik, P. *et al.* BiQ Analyzer HT: Locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing. *Nucleic Acids Res.* **39,** 551–556 (2011).

Acknowledgements

We thank all of the participants and families for taking part in this study. We thank Operation Smile Brasil and Dr. Daniela Bueno for the productive collaboration and Suzana Ezquina for assistance in the analysis. We thank Lucy Jenkins and Deborah Morrow for their assistance with collecting UK control samples. Thanks also go to Mark Swan, Chris Theopold and members of the North Thames Cleft team for their assistance with recruitment and sample collection and to Charles Mein at the Barts and The London Genome Centre, QMUL for assisting with the Bisulfite Amplicon Sequencing. LA, LAB and MRPB are supported by FAPESP (São Paulo Research Foundation, CEPID 2013/08028-1) and CNPq (National Counsel of Technological and Scientific Development, CNPq-PQ10/2011). PS and GEM are supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London and are also grateful to grant support from Cleft and the MRC.

Author contribution

LA designed experiments, performed experiments and analyses and wrote the paper. XK analysed data and discussed experimental design. LAB performed experiments. RS collected samples. GEM discussed experimental design. PS and MRPB discussed experimental design and results and wrote the paper.

Competing interests

The authors declare no competing financial interests.

Corresponding authors

Correspondence to Maria Rita Passos Bueno or Philip Stanier.

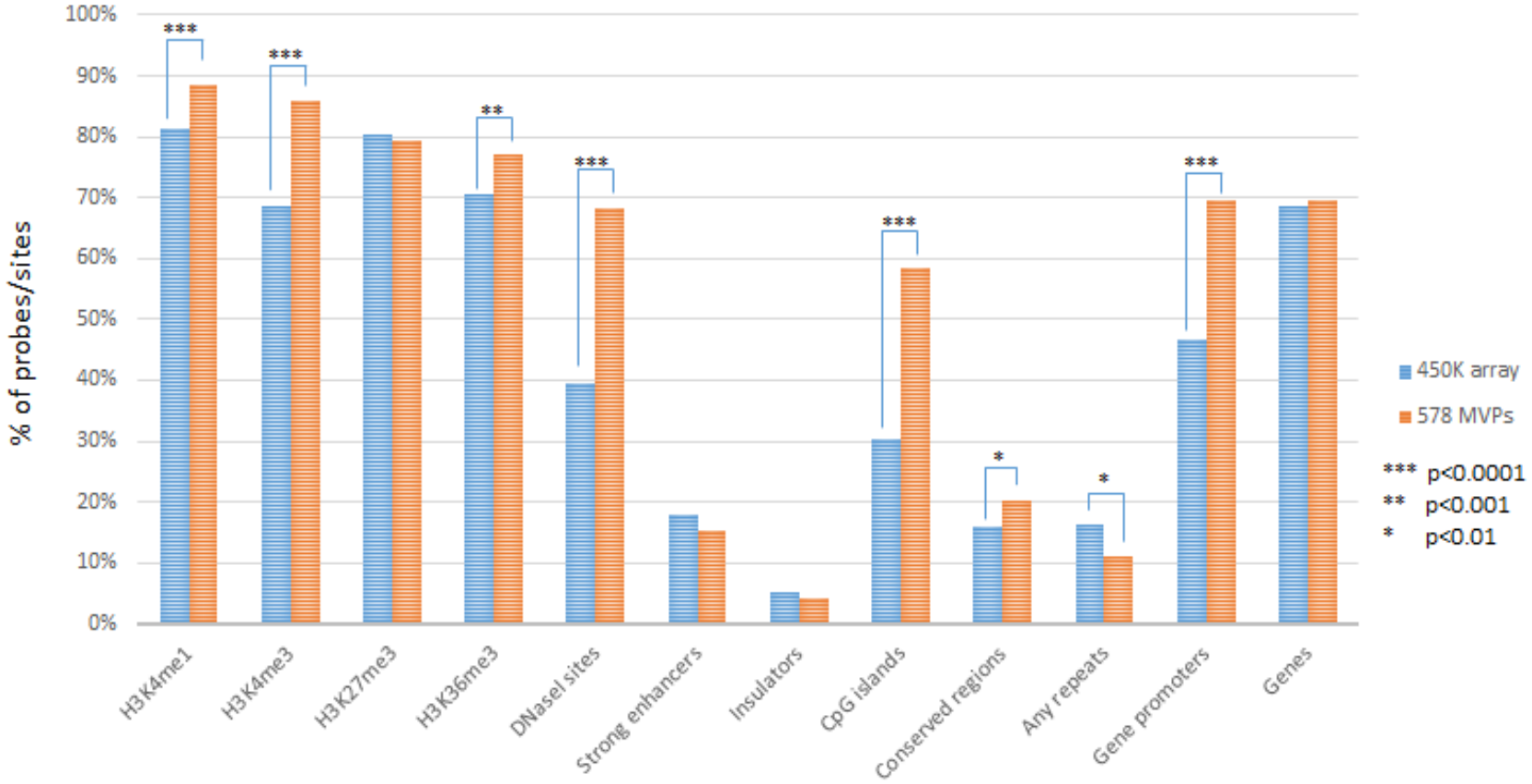


Figure 1: MVPs are enriched at active regions of the genome. Genomic distribution of EpiExplorer classes comparing the 578 MVPs and the Infinium Human Methylation 450K Bead-Array (450K array) filtered probes showing significant enrichment in the MVPs for active regions of the genome, including gene promoters (Chi-Square Test with Yates correction). Genomic coordinates from the 578 MVPs and 450K filtered probes were used as input in the EpiExplorer online tool (http://epiexplorer.mpi-inf.mpg.de/) for genomic distribution and chromatin segments comparisons.

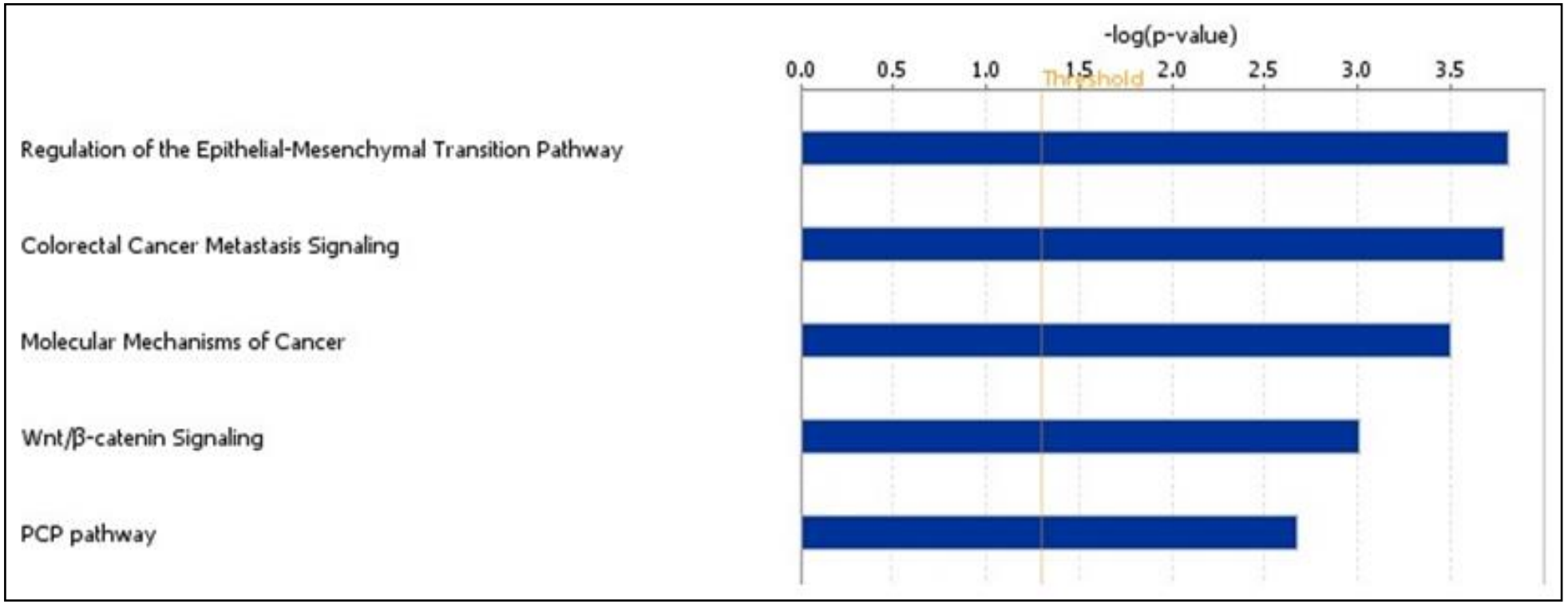


Figure 2: Enrichment of MVPs in canonical pathways related to craniofacial development. Top five canonical pathways on Ingenuity Pathway Analysis (IPA) using the 578 MVPs shows “Regulation of the Epithelial-Mesenchymal Transition Pathway”, “Wnt/B-catenin Signaling” and “PCP pathway” as significantly enriched in those MVPs. Those pathways are extensively related to craniofacial development, including lip and palate morphogenesis.

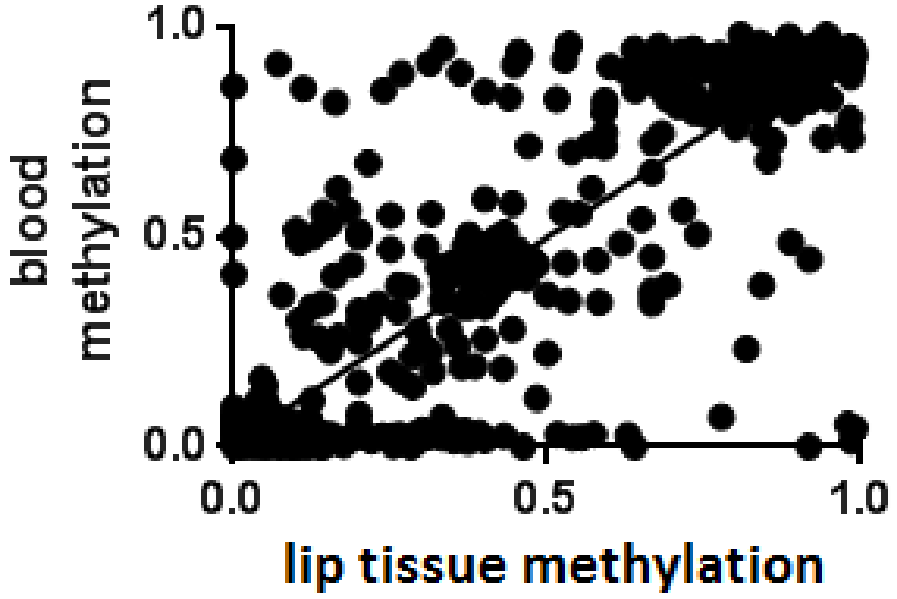


Figure 3: Whole-blood DNA methylation correlates to lip tissue DNA methylation. Linear regression between methylation values from all analysed MVPs and flanking regions covered by BSAS showing significant correlation between whole-blood and lip tissue methylation levels (R-square=0.9028, $p<0.0001$).

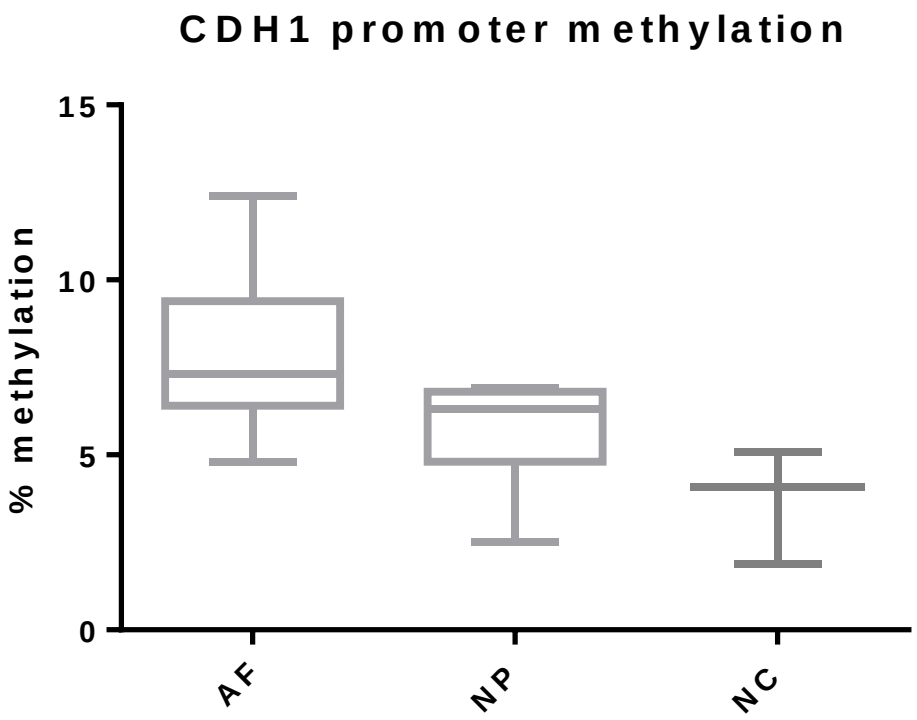
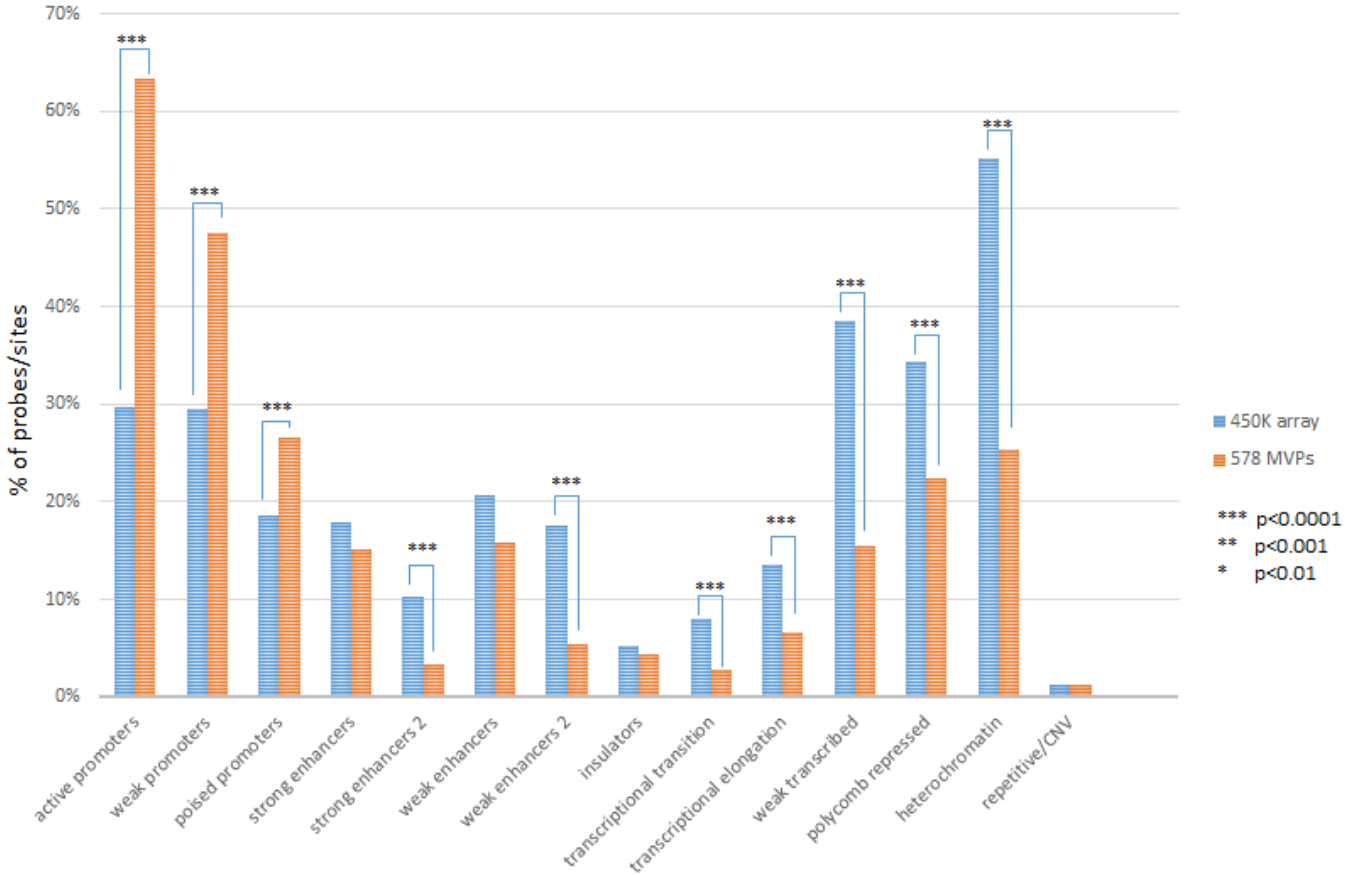


Figure 4: *CDH1* promoter methylation correlates to penetrance. Boxplot of *CDH1* promoter methylation levels in familial NS CL/P segregating *CDH1* mutations. Affected individuals (AF, n=8) exhibit significantly higher *CDH1* promoter methylation levels in comparison to non-penetrant (NP, n=7) and non-carriers (NC, n=3) (p=0,0112, Kruskal-Wallis test). We quantified methylation levels from 33 CpGs at *CDH1* promoter using conventional bisulfite sequencing in which 32 clones were analysed per sample.

Table I: Methylation differences for the 11 tested MVPs in the methWAS (Brazilian cohort, NSCL/P N=68, control N = 59) and Replication study (British cohort, NSCL/P N=171, control N=177). We found significant methylation difference in the replication study for four MVPs (*FAT1* MVP, *MYC* MVP, *WHSC1* MVP and chr1 MVP). Considering direction of methylation, only *FAT1* MVP and *MYC* MVP were replicated with the same pattern as observed in the methWAS. (Genomic coordinates on hg18; Meth Diff: methylation difference).

			methWAS		Replication study			
	450K ID	Genomic coordinate	Meth Diff	adjusted p-value	Mean Methylation NSCL/P	Mean Methylation Control	Meth Diff	p-value
FAT1 MVP	cg00405769	chr4:187539852	-0,075	4,66E-09	0.9244 ± 0.002172	0.9331 ± 0.0007216	-0,00871	0,0002
MYC MVP	cg00611675	chr8: 128748464	-0,02626	1,52E-09	0.003672 ± 0.0003027	0.009375 ± 0.0008236	-0,0057	<0,0001
WHSC1 MVP	cg03150409	chr4: 1892317	0,07564	1,02E-08	0.3055 ± 0.009068	0.3302 ± 0.008440	-0,02473	0,04
PVRL1 MVP	cg06391300	chr11: 119600292	-0,0419	3,01E-13	0.001697 ± 0.0001290	0.001795 ± 0.0001241	-0,0001	0,58
chr17 MVP	cg09319020	chr17: 7304467	-0,11844	2,05E-09	0.8406 ± 0.001630	0.8428 ± 0.0007958	-0,00216	0,23
C11orf58 MVP	cg10633981	chr11: 16779768	0,12801	1,85E-16	0.8587 ± 0.005899	0.8683 ± 0.007230	-0,00963	0,3
FGF8 MVP	cg11706469	chr10:103535362	-0,04486	0,000000064	0.007407 ± 0.0002551	0.007412 ± 0.0002048	-0,00001	0,98
WNT2B MVP	cg11806528	chr1:113051977	0,0321	4,58E-08	0.005564 ± 0.0001993	0.005658 ± 0.0003072	-0,00009	0,79
WNT7A MVP	cg13602813	chr3:13920840	-0,03251	9,88E-08	0.01223 ± 0.0004003	0.01312 ± 0.0004133	-0,00088	0,12
chr1 MVP	cg15897635	chr1:220697615	0,10364	1,82E-14	0.7349 ± 0.003714	0.7450 ± 0.003034	-0,0101	0,03
FGFR1 MVP	cg20913106	chr8:38324522	0,0243	2,95E-12	0.004360 ± 0.0002625	0.004482 ± 0.0002628	-0,00012	0,74



Supplementary Figure 1: MVPs are enriched in regulatory regions of the genome. Comparison of Chromatin segments from EpiExplorer between the 578 MVPs and the Infinium Human Methylation 450K Bead-Array (450K array) filtered probes showing significant enrichment in the MVPs for active regions as “active promoters”, “weak promoters”, “poised promoters”. Alternatively, regions as “weak transcribed”, “polycombrepressed” and “heterochromatin” are significantly less represented in our MVPs (Chi-Square Test with Yates correction).

Supplementary Table I: Differentially methylated MVPs and respective probe Ids, methylation values, standard deviations and associated p-values

CGID	Chromosome	Start	Strand	Mean NSCL/P	Mean Control	Methylation Difference	SD NSCL/P	SD Control	Adjusted p-value
cg00040446	chr1	40367965	+	0,084100494	0,118857687	-0,034757193	0,016324646	0,027398075	2,18E-12
cg00087005	chr10	126848218	+	0,05749841	0,047330967	0,010167443	0,009106939	0,008223907	6,04E-08
cg00185336	chr7	73153699	+	0,058819827	0,039841398	0,018978429	0,014521135	0,012432393	7,93E-10
cg00297767	chr21	40721426	+	0,140729609	0,188332486	-0,047602876	0,02293198	0,033912163	1,50E-10
cg00319567	chr13	60971258	-	0,078773029	0,109171343	-0,030398315	0,012437483	0,022563358	1,16E-10
cg00345230	chr4	151936469	+	0,083749423	0,121743021	-0,037993598	0,013496632	0,031898587	1,13E-10
cg00366471	chr15	71184982	+	0,068477725	0,097584455	-0,02910673	0,012192961	0,024398131	8,62E-10
cg00405769	chr4	187539852	+	0,803752767	0,878755594	-0,075002827	0,061397463	0,064216338	4,66E-09
cg00498163	chr2	242821619	-	0,866086495	0,824853172	0,041233323	0,025999108	0,035753343	3,11E-08
cg00611675	chr8	128748464	+	0,071423065	0,097679093	-0,026256028	0,011539241	0,020695635	1,52E-09
cg00645593	chr10	35930499	-	0,147944973	0,177793701	-0,029848728	0,020826516	0,030794041	5,87E-08
cg00774457	chr2	58479214	+	0,759287729	0,709340679	0,049947051	0,031233968	0,041382169	2,91E-08
cg00787537	chr1	211307839	-	0,054668363	0,073252637	-0,018584274	0,008380661	0,015295657	1,04E-08
cg00811382	chr7	97821315	-	0,936193586	0,925414071	0,010779515	0,016561789	0,016937067	4,33E-08
cg00959118	chr1	169455055	+	0,107691474	0,158328181	-0,050636707	0,017660968	0,033863232	6,39E-13
cg01092528	chr4	76861103	-	0,091457324	0,049627163	0,04183016	0,021820811	0,022226612	1,22E-15
cg01101773	chr3	15901650	-	0,039373277	0,033075145	0,006298132	0,004231328	0,008396769	1,28E-09
cg01123250	chr2	210673545	+	0,713745742	0,808840809	-0,095095067	0,093951893	0,120432545	9,66E-08
cg01176458	chr3	27763828	+	0,11407498	0,165527094	-0,051452113	0,01955723	0,039384967	5,75E-12
cg01213331	chr16	15069449	+	0,045719392	0,062340969	-0,016621577	0,009413064	0,012502858	1,41E-09
cg01334186	chr1	151372572	+	0,089064328	0,060984261	0,028080067	0,013864016	0,027230207	2,11E-12
cg01338834	chr21	40720919	+	0,184071381	0,236625189	-0,052553808	0,027160223	0,041000392	4,63E-08
cg01356463	chr12	112280204	+	0,100825206	0,081815774	0,019009433	0,012602918	0,01995035	5,01E-09

cg01374398	chr3	47516975	-	0,156118525	0,205139108	-0,049020583	0,021587054	0,033968436	1,86E-10
cg01411982	chr11	134123293	+	0,054912701	0,071465895	-0,016553195	0,008697628	0,015696694	5,16E-08
cg01460276	chr8	18871817	+	0,077253292	0,10062992	-0,023376629	0,013348632	0,02414445	5,66E-09
cg01519350	chr3	137906342	-	0,054464931	0,038152838	0,016312093	0,01052709	0,012108475	1,18E-10
cg01550055	chr1	119543216	+	0,131562817	0,162506896	-0,030944079	0,020660214	0,028453329	8,48E-09
cg01614598	chr2	38152364	-	0,053462422	0,082127738	-0,028665315	0,009465548	0,020559137	3,87E-13
cg01684248	chr16	86536239	-	0,430343205	0,480502172	-0,050158968	0,043952247	0,052105128	1,13E-08
cg01722584	chr20	5451695	-	0,718574908	0,780163928	-0,06158902	0,043866646	0,058209371	2,36E-08
cg01736133	chr4	120605519	+	0,818647174	0,747742374	0,0709048	0,045164091	0,066011439	1,90E-09
cg01737507	chr5	122372093	+	0,261059795	0,233590338	0,027469456	0,01631816	0,034041869	9,63E-09
cg01775802	chr14	72945461	+	0,459308844	0,360067443	0,099241401	0,066971252	0,062925492	1,36E-08
cg01910272	chr3	52489346	+	0,088968277	0,125385948	-0,03641767	0,019030561	0,029454266	1,69E-09
cg01930417	chr11	122855023	+	0,056746078	0,045825315	0,010920763	0,008672946	0,004893127	2,41E-09
cg01941274	chr14	50234650	-	0,068337961	0,096275884	-0,027937923	0,012584086	0,022166908	3,24E-10
cg01969796	chr4	750851	+	0,920589729	0,897795463	0,022794267	0,012355842	0,018338544	5,26E-09
cg01976034	chr3	57313955	+	0,860284976	0,829590912	0,030694064	0,03314653	0,063538278	9,30E-08
cg01988500	chr16	3030808	+	0,066791694	0,094636318	-0,027844624	0,012289514	0,023845165	1,62E-08
cg02038173	chr21	34852701	+	0,073492837	0,105785193	-0,032292356	0,013835654	0,027262209	1,31E-10
cg02158565	chr7	150788256	+	0,847185829	0,811751721	0,035434108	0,025908388	0,027702914	5,73E-08
cg02180424	chr14	48095779	-	0,694727972	0,77542347	-0,080695497	0,067150507	0,085707173	4,49E-08
cg02308712	chr12	52598204	-	0,624245214	0,520903661	0,103341553	0,058845728	0,064321234	2,63E-12
cg02317763	chr16	67753188	-	0,03726701	0,03060295	0,006664059	0,003417482	0,005509199	9,40E-09
cg02339752	chr16	86549705	+	0,181373776	0,216141341	-0,034767565	0,020443156	0,031750641	2,05E-08
cg02394512	chr11	94245735	-	0,078438058	0,060700429	0,017737629	0,012504716	0,016434391	2,98E-08
cg02467765	chr4	78978708	+	0,094720918	0,073469806	0,021251112	0,014117412	0,022619627	1,86E-09
cg02511809	chr5	112073544	+	0,13531237	0,178788397	-0,043476027	0,019982841	0,034992225	7,83E-11
cg02548780	chr16	22020026	+	0,072507464	0,101730395	-0,029222931	0,012588613	0,023135395	2,94E-08
cg02551910	chr10	12110991	+	0,068893567	0,0952367	-0,026343133	0,009289941	0,018610977	2,55E-10

cg02559264	chr11	73490474	-	0,091474215	0,12167173	-0,030197515	0,014919081	0,025000021	3,56E-08
cg02732134	chr10	101732382	-	0,542350277	0,466020503	0,076329775	0,042814169	0,038834122	1,39E-14
cg02867857	chr16	88941494	+	0,838800819	0,811980968	0,02681985	0,0312948	0,020384068	1,45E-08
cg02895995	chr19	7554069	+	0,082395425	0,049245566	0,033149859	0,013081049	0,020951209	6,09E-17
cg02905964	chr19	29704262	+	0,19898055	0,245578228	-0,046597678	0,021699899	0,034307614	1,25E-10
cg02934500	chr4	142557404	+	0,068991491	0,086160026	-0,017168534	0,011373576	0,015860886	3,82E-08
cg02996269	chr4	106394692	-	0,136067964	0,180293848	-0,044225884	0,018166624	0,034866007	1,36E-11
cg03043160	chr13	76056444	-	0,037621502	0,030247869	0,007373633	0,004919194	0,010278188	1,16E-11
cg03147627	chr4	83956202	+	0,050667548	0,040049076	0,010618472	0,008512024	0,008147407	5,36E-08
cg03148428	chr7	156433416	-	0,161812548	0,130495069	0,031317479	0,031361683	0,055156344	1,21E-08
cg03149093	chr19	12035973	+	0,079691797	0,113460433	-0,033768636	0,012045694	0,023692373	3,18E-12
cg03150409	chr4	1892317	+	0,604464495	0,528821594	0,075642901	0,067375568	0,065225063	1,02E-08
cg03178691	chr11	66629022	-	0,881752447	0,841806165	0,039946282	0,029414179	0,042568142	9,39E-08
cg03245111	chr7	76253113	-	0,50127631	0,457995215	0,043281095	0,063387509	0,057681656	1,28E-08
cg03258011	chr17	54857409	+	0,063928942	0,043603815	0,020325127	0,013286319	0,013936341	4,05E-11
cg03284310	chr10	124894415	+	0,172464234	0,215271733	-0,042807499	0,022341938	0,034453929	4,25E-08
cg03312572	chr7	17979929	+	0,125991764	0,16848292	-0,042491156	0,017047804	0,026289709	3,88E-11
cg03410231	chr7	143042798	+	0,099883942	0,124757371	-0,024873429	0,012384972	0,022246268	2,07E-08
cg03447967	chr2	45163099	+	0,141547791	0,177553333	-0,036005542	0,017714413	0,032676348	2,57E-11
cg03462975	chr6	33172517	-	0,165649754	0,212604138	-0,046954384	0,020196282	0,036195137	3,51E-09
cg03465513	chr13	28550046	-	0,123028686	0,156296992	-0,033268305	0,019818777	0,030346955	6,03E-08
cg03486485	chr11	14927004	+	0,044847797	0,061188387	-0,016340589	0,010538859	0,023804964	6,59E-08
cg03515980	chr2	105488847	-	0,074442433	0,107196614	-0,032754181	0,012191905	0,028950219	1,54E-10
cg03637614	chr7	129691233	+	0,066912705	0,096250307	-0,029337602	0,013209328	0,02517	7,73E-10
cg03648631	chr4	103749307	-	0,127723979	0,176270678	-0,048546699	0,019622176	0,041217284	2,13E-09
cg03758392	chr3	184870408	+	0,035785172	0,02642394	0,009361232	0,00584647	0,009350557	8,41E-09
cg03766703	chr19	46393454	+	0,928592595	0,915965593	0,012627002	0,011716411	0,019930789	4,94E-11
cg04152921	chr16	47495445	-	0,070367003	0,096996838	-0,026629835	0,01254299	0,021448459	1,76E-10

cg04193218	chr4	20701691	-	0,057821527	0,083180366	-0,025358839	0,011452307	0,018925872	1,89E-11
cg04205664	chr7	139209511	-	0,341838842	0,266613354	0,075225489	0,068192056	0,081375122	7,68E-10
cg04240971	chr2	202031356	-	0,68429284	0,743511886	-0,059219046	0,036222531	0,043353763	5,08E-09
cg04250181	chr12	20966524	-	0,843513868	0,782102616	0,061411252	0,038530611	0,052307513	8,29E-13
cg04335562	chr19	2945000	+	0,17830578	0,229884651	-0,051578871	0,021054732	0,029079334	1,68E-12
cg04358685	chr3	12598421	-	0,139126742	0,179573297	-0,040446554	0,028479322	0,034446548	2,42E-08
cg04364463	chr1	37498270	+	0,164370694	0,208863932	-0,044493238	0,018605706	0,037556654	1,23E-08
cg04371633	chr14	65381025	+	0,044798605	0,03255078	0,012247825	0,007597079	0,011900642	1,11E-10
cg04414295	chr14	102606318	+	0,075554382	0,103637158	-0,028082775	0,011195151	0,02406803	2,70E-09
cg04459753	chr16	28858035	+	0,051642266	0,071448596	-0,01980633	0,007368754	0,014770618	4,31E-10
cg04468334	chr7	149414901	-	0,828101928	0,77272344	0,055378489	0,035320429	0,044486122	5,68E-09
cg04558166	chr1	210001279	+	0,181914395	0,162472574	0,019441821	0,015995697	0,02126811	5,28E-08
cg04581293	chr3	156889197	-	0,760833878	0,714172412	0,046661467	0,033331786	0,047249602	3,43E-08
cg04600792	chr7	29256702	+	0,907160174	0,879547048	0,027613126	0,017195409	0,027391025	6,80E-08
cg04623165	chr2	10372402	-	0,726806159	0,657086548	0,069719611	0,048568356	0,051512129	5,68E-12
cg04684553	chr8	89339819	+	0,08867378	0,067968222	0,020705558	0,014700173	0,022760133	3,11E-08
cg04699668	chr2	132109931	+	0,755424352	0,698349798	0,057074554	0,031546133	0,046321921	4,18E-08
cg04721582	chr2	178257292	-	0,054044605	0,073498674	-0,019454069	0,00824811	0,014920809	1,51E-10
cg04871364	chr1	6501062	+	0,176831555	0,23594288	-0,059111325	0,055448917	0,072887378	1,14E-08
cg04925841	chr12	24715491	+	0,135883042	0,174983435	-0,039100394	0,016144086	0,033345387	7,23E-09
cg04930982	chr6	149202328	-	0,936495121	0,912044425	0,024450695	0,014936112	0,036747536	5,17E-09
cg04948475	chr4	69810056	-	0,653191634	0,736465249	-0,083273615	0,050983663	0,078254533	8,30E-08
cg05003666	chr11	8102717	-	0,289536156	0,238626178	0,050909978	0,037363779	0,058861758	7,70E-08
cg05052194	chr1	2160249	+	0,084072406	0,119398946	-0,03532654	0,015365963	0,025886162	1,42E-10
cg05106659	chr4	152682899	-	0,916603651	0,887789184	0,028814467	0,022688355	0,037379749	1,41E-09
cg05190577	chr11	61062843	+	0,058869312	0,076430226	-0,017560914	0,009902553	0,017007928	3,34E-08
cg05246190	chr22	50330022	+	0,124793266	0,166517871	-0,041724604	0,02189094	0,030815715	1,32E-09
cg05301818	chr8	125385252	-	0,114442811	0,146172487	-0,031729676	0,017729601	0,028356988	4,53E-08

cg05303728	chr8	32078368	+	0,080132996	0,107983012	-0,027850015	0,016911253	0,026299295	1,02E-08
cg05393509	chr6	30294475	+	0,132889211	0,176001249	-0,043112038	0,023158726	0,028741991	6,39E-10
cg05426700	chr16	998082	+	0,722309519	0,655947349	0,06636217	0,033922539	0,064461909	8,99E-08
cg05478497	chr12	4274287	-	0,051350805	0,070715859	-0,019365054	0,008008071	0,016009782	3,39E-09
cg05485520	chr3	45883628	+	0,079668821	0,054296219	0,025372602	0,013867338	0,019600798	2,54E-13
cg05492810	chr11	78129288	+	0,187819374	0,234603822	-0,046784448	0,021482618	0,037690054	2,74E-10
cg05492964	chr12	53835262	+	0,1164075	0,153101078	-0,036693578	0,019532228	0,031899266	2,80E-08
cg05494465	chr13	74710406	+	0,082490675	0,110534037	-0,028043363	0,01307899	0,024188901	8,65E-08
cg05590156	chr5	14183868	+	0,839587551	0,783710486	0,055877065	0,027906765	0,070006808	9,95E-09
cg05654404	chr1	45452166	+	0,102436032	0,147407115	-0,044971083	0,01828182	0,037226311	3,11E-08
cg05681148	chr13	28195812	-	0,115716112	0,145767755	-0,030051643	0,017011589	0,021372926	1,67E-08
cg05725489	chr7	600658	-	0,682868161	0,659196674	0,023671487	0,047782798	0,051360581	7,76E-09
cg05732725	chr18	59712034	+	0,847963885	0,799689384	0,048274502	0,029878713	0,038817264	7,20E-10
cg05736393	chr2	74775957	-	0,127605963	0,165328032	-0,037722069	0,017582604	0,024439219	1,84E-11
cg05796321	chr1	92950348	-	0,097890218	0,137917502	-0,040027284	0,01913218	0,032245504	9,21E-10
cg05975845	chr20	47538463	-	0,044497631	0,058050385	-0,013552753	0,007280225	0,012520644	7,04E-08
cg05977109	chr10	52178181	-	0,051204388	0,068848967	-0,017644579	0,009452869	0,015533958	6,54E-12
cg06124224	chr1	99470709	-	0,057896059	0,079193616	-0,021297557	0,008305803	0,018153716	5,12E-09
cg06194536	chr1	220701829	-	0,0734931	0,102221558	-0,028728459	0,012814953	0,025902512	1,96E-10
cg06210054	chr1	36173658	-	0,224682151	0,173757101	0,05092505	0,029930495	0,043118735	3,56E-09
cg06391300	chr11	119600292	+	0,085637051	0,127533516	-0,041896465	0,011425248	0,030426724	3,01E-13
cg06425136	chr16	31040922	-	0,734009853	0,681508157	0,052501697	0,069913839	0,077655243	8,60E-08
cg06510410	chr3	196594695	-	0,157479949	0,134147152	0,023332797	0,017348083	0,034890659	9,26E-10
cg06513149	chr13	33002431	-	0,240093951	0,283013802	-0,042919851	0,017541797	0,029906808	1,36E-10
cg06551493	chr7	77166702	-	0,217329869	0,160740959	0,05658891	0,03957474	0,059283103	3,59E-08
cg06679760	chr3	49027156	+	0,052194629	0,07664371	-0,024449081	0,008931152	0,019538352	5,06E-12
cg06742800	chr2	27255708	+	0,085389362	0,125182303	-0,039792941	0,014164892	0,030307901	2,71E-11
cg06744057	chr14	73493715	-	0,129996408	0,169919621	-0,039923214	0,020003781	0,030945694	7,21E-09

cg06766273	chr11	62521983	+	0,040276493	0,02904236	0,011234133	0,011154696	0,00873099	9,21E-10
cg06812693	chr4	26323246	-	0,168855325	0,215801754	-0,046946429	0,016898646	0,033941894	9,61E-12
cg06824641	chr8	22298455	-	0,070399309	0,058532076	0,011867234	0,012485519	0,016216151	5,48E-08
cg06974483	chr4	124317831	+	0,051543813	0,039115826	0,012427987	0,010115411	0,009914378	9,43E-08
cg06994813	chr7	82792357	+	0,053327484	0,070986098	-0,017658614	0,013340415	0,019710198	1,93E-10
cg07022953	chr1	27527861	+	0,935211741	0,924122723	0,011089018	0,012942304	0,019000366	3,66E-09
cg07047601	chr22	21337040	+	0,149324374	0,194661911	-0,045337537	0,023186799	0,033930375	6,67E-08
cg07113562	chr10	52834093	+	0,080338538	0,105129335	-0,024790797	0,010603281	0,02213189	1,79E-08
cg07161888	chr16	53537348	-	0,087189564	0,125804065	-0,038614501	0,013052933	0,02977105	4,82E-12
cg07237590	chr20	62526788	-	0,126812414	0,1008253	0,025987114	0,022329739	0,037253663	1,88E-09
cg07243700	chr1	100731457	+	0,036991312	0,030398249	0,006593064	0,00628276	0,008409744	8,81E-08
cg07283595	chr17	5372003	-	0,127166692	0,178153131	-0,050986438	0,017156583	0,034080326	8,29E-13
cg07284286	chr16	58662935	-	0,141331603	0,181729687	-0,040398084	0,016840232	0,034422677	1,88E-08
cg07291923	chr2	28113263	-	0,109821717	0,145031096	-0,035209379	0,013023954	0,025139773	1,04E-09
cg07326438	chr11	57194562	+	0,034659767	0,029055508	0,005604258	0,003507189	0,005316752	4,43E-09
cg07449447	chr14	55228808	-	0,800601712	0,759800595	0,040801118	0,036675394	0,047981703	1,89E-12
cg07450219	chr17	78428858	+	0,230250697	0,183209338	0,047041359	0,040017561	0,053061497	1,48E-10
cg07482222	chr19	3465714	+	0,938370798	0,928569635	0,009801163	0,010319813	0,021060951	1,49E-08
cg07647108	chr3	107150444	+	0,07361571	0,056957279	0,016658431	0,007915058	0,019125992	4,44E-12
cg07700962	chr1	174968950	-	0,045155787	0,058114238	-0,01295845	0,007993314	0,011607543	1,70E-08
cg07757959	chr15	79165367	+	0,05956985	0,081868158	-0,022298308	0,011043694	0,019381168	1,48E-09
cg07791468	chr2	31492359	+	0,55375295	0,650799572	-0,097046622	0,065657993	0,096378433	6,51E-09
cg07796520	chr18	5296200	-	0,03126074	0,026847707	0,004413033	0,003364154	0,004081739	4,07E-09
cg07830086	chr19	18107778	+	0,077812908	0,104166958	-0,02635405	0,011651001	0,021995439	5,18E-09
cg07899451	chr1	157108248	+	0,074098909	0,058675366	0,015423543	0,009389111	0,014866303	1,84E-10
cg07919028	chr12	4554480	-	0,810071309	0,787773968	0,022297341	0,021223567	0,01943286	6,19E-08
cg07949612	chr2	69664562	+	0,089495383	0,123537633	-0,034042251	0,014226762	0,027451157	1,42E-08
cg08005411	chr12	69753742	+	0,055478401	0,077426725	-0,021948324	0,0090513	0,017544119	6,72E-10

cg08031094	chr2	54013912	-	0,075702703	0,108912373	-0,03320967	0,010411185	0,028357492	1,37E-09
cg08040170	chr16	18812977	+	0,051422672	0,071591459	-0,020168786	0,009175878	0,016118966	3,07E-08
cg08052428	chr9	135996421	-	0,086029226	0,051664219	0,034365007	0,016875606	0,028259521	1,10E-13
cg08092446	chr19	58987653	-	0,063373316	0,088485011	-0,025111695	0,009060122	0,020851967	9,56E-11
cg08125401	chr15	74726457	+	0,132085553	0,178561456	-0,046475903	0,02204915	0,03554899	6,65E-10
cg08173533	chr4	106756925	-	0,854940953	0,821942318	0,032998635	0,022809739	0,031649304	3,04E-08
cg08204369	chr17	73150733	-	0,094263182	0,07059556	0,023667622	0,015049566	0,019891085	1,38E-10
cg08257600	chr19	41199188	+	0,814683097	0,767104386	0,047578712	0,025685822	0,046687629	2,91E-10
cg08258788	chr8	109455865	+	0,069767436	0,100014805	-0,030247369	0,011110065	0,023203557	9,14E-13
cg08269986	chr1	35735167	+	0,049649832	0,067927828	-0,018277996	0,010273488	0,016779282	5,44E-10
cg08344081	chr6	41339785	-	0,05494867	0,04579991	0,00914876	0,010167896	0,012352195	7,50E-08
cg08409451	chr16	2273123	+	0,189261313	0,235162963	-0,04590165	0,023747694	0,033202027	8,33E-10
cg08440556	chr4	83718952	-	0,100212468	0,126596579	-0,026384111	0,018225045	0,027390578	4,20E-08
cg08452348	chr5	134303703	+	0,86700701	0,828010456	0,038996554	0,027686498	0,033635947	3,62E-08
cg08466051	chr2	166810428	-	0,154704078	0,118906056	0,035798022	0,028521091	0,039874768	7,75E-08
cg08502652	chr6	168592971	-	0,893115491	0,832156745	0,060958746	0,021413712	0,089189223	3,89E-08
cg08516222	chr3	153840255	+	0,096947136	0,125188249	-0,028241113	0,018696719	0,031257582	2,86E-08
cg08557876	chr16	81130217	-	0,065157237	0,05376359	0,011393646	0,012120426	0,021447078	7,75E-08
cg08642237	chr15	45694386	-	0,075532843	0,116504498	-0,040971656	0,014459211	0,033627724	3,58E-10
cg08722104	chr11	47448306	+	0,037032131	0,026268028	0,010764103	0,007543084	0,011297598	5,30E-09
cg08723977	chr5	85577255	-	0,742664683	0,681448963	0,06121572	0,027833765	0,055625349	3,00E-09
cg08795320	chr16	66982803	+	0,468391104	0,414895133	0,053495971	0,038839816	0,056086669	3,37E-08
cg08820231	chr12	58013687	-	0,205553326	0,24472498	-0,039171654	0,01831427	0,02899782	7,68E-10
cg08843001	chr8	30891486	-	0,068361734	0,093928034	-0,0255663	0,010239146	0,023991116	1,05E-09
cg08905496	chr1	165414332	-	0,074444307	0,047358336	0,027085971	0,014699292	0,018997126	1,84E-10
cg08924777	chr16	48278166	-	0,058539513	0,080999068	-0,022459555	0,010249752	0,018715417	4,48E-09
cg08935299	chr12	123011730	+	0,063544978	0,080604794	-0,017059817	0,013756604	0,016808704	2,71E-08
cg08948898	chr15	74428991	-	0,095375434	0,079248893	0,016126541	0,013929286	0,019363621	2,90E-08

cg08995662	chr19	17137689	+	0,16841795	0,122945334	0,045472616	0,043882831	0,063837559	1,56E-08
cg09012784	chr1	109584868	-	0,486925131	0,437805878	0,049119254	0,037284546	0,050125464	1,03E-08
cg09015484	chr9	96929106	+	0,05093632	0,0384706	0,012465721	0,009549177	0,013649929	3,56E-08
cg09047471	chr8	56792345	+	0,062285312	0,090281476	-0,027996164	0,009612433	0,019183977	1,46E-11
cg09098825	chr3	47866946	-	0,062358963	0,085753109	-0,023394146	0,010725713	0,020993179	9,00E-10
cg09112371	chr6	33240090	-	0,087691237	0,111532809	-0,023841572	0,014216131	0,022064717	8,65E-08
cg09302904	chr17	62340561	-	0,118453602	0,156217074	-0,037763472	0,014326545	0,02726743	1,93E-10
cg09314708	chr11	132812991	+	0,080541228	0,056293238	0,02424799	0,015536424	0,021334724	4,63E-08
cg09319020	chr17	7304467	+	0,510572627	0,62901059	-0,118437964	0,072090362	0,082235086	2,05E-09
cg09320595	chr6	31478822	+	0,921072807	0,904187855	0,016884952	0,013727113	0,020754545	5,59E-08
cg09337544	chr22	38857452	-	0,120999011	0,165742907	-0,044743896	0,020887067	0,032339817	1,19E-09
cg09372808	chr1	167791030	-	0,80765468	0,738420548	0,069234133	0,040496437	0,052869524	2,48E-09
cg09379601	chr19	12992224	-	0,068181642	0,039247921	0,028933721	0,013259691	0,01958603	4,48E-17
cg09403165	chr15	90118814	+	0,067202541	0,093444416	-0,026241875	0,010900817	0,021803323	5,21E-13
cg09460490	chr13	113829357	+	0,441499119	0,510517774	-0,069018655	0,060648719	0,060454711	8,52E-08
cg09477292	chr4	6577027	+	0,168680597	0,21215958	-0,043478983	0,020689814	0,030117569	3,36E-09
cg09556825	chr13	31736209	-	0,044699814	0,066143763	-0,021443949	0,008474987	0,016834577	4,68E-12
cg09562174	chr16	87811505	+	0,176810401	0,21359288	-0,036782479	0,021516215	0,029143429	2,64E-08
cg09656541	chr10	115932823	+	0,401558163	0,327183793	0,07437437	0,065514599	0,070803822	3,57E-08
cg09705531	chr12	133525532	+	0,713144567	0,685823799	0,027320768	0,026062362	0,023520915	3,89E-09
cg09717979	chr4	103790325	-	0,093979943	0,137257958	-0,043278015	0,012095597	0,029489924	5,08E-12
cg09757127	chr1	3810632	-	0,898514068	0,870178897	0,028335171	0,019602799	0,022710235	1,85E-08
cg09766628	chr17	4634728	+	0,08940852	0,127382157	-0,037973637	0,013432787	0,02947126	1,67E-10
cg09890653	chr11	113931305	+	0,114752132	0,148246031	-0,033493898	0,016691108	0,030872997	6,23E-09
cg09897639	chr1	178063342	+	0,075804227	0,100091232	-0,024287005	0,015466659	0,026098782	3,75E-08
cg09938490	chr15	44088809	+	0,92709899	0,914499293	0,012599697	0,014010602	0,01735762	7,45E-08
cg09975850	chr8	65711396	+	0,132280092	0,168085287	-0,035805195	0,020840151	0,035216703	6,68E-09
cg10059171	chr1	147400607	-	0,046658197	0,062755555	-0,016097358	0,007584902	0,013340865	5,64E-09

cg10227327	chr14	62584037	+	0,100987831	0,130321439	-0,029333609	0,013550665	0,021575417	7,63E-11
cg10300729	chr10	65284527	+	0,597321624	0,677416093	-0,080094469	0,056386385	0,079148802	8,41E-08
cg10335230	chr11	134628434	+	0,720273275	0,684528339	0,035744936	0,030325593	0,028705083	3,92E-08
cg10459717	chr5	6633577	-	0,069644338	0,09521728	-0,025572942	0,01056983	0,019633379	1,97E-10
cg10466987	chr17	76921528	+	0,154000057	0,202666649	-0,048666592	0,021206112	0,038351315	7,17E-09
cg10505247	chr1	207226475	+	0,049313399	0,069263351	-0,019949952	0,007496429	0,016271312	5,00E-11
cg10515414	chr1	15650268	+	0,07820974	0,112950697	-0,034740957	0,01364166	0,027270401	1,48E-09
cg10548355	chr11	70129868	-	0,892425214	0,868657274	0,02376794	0,016500363	0,023835176	6,43E-08
cg10570662	chr22	19159548	-	0,090008038	0,114705141	-0,024697102	0,01524187	0,021610121	4,03E-08
cg10581632	chr12	56360715	+	0,100409349	0,146836486	-0,046427137	0,012590574	0,034544246	9,04E-12
cg10592245	chr1	215256451	+	0,046224013	0,037155367	0,009068646	0,005935763	0,009237183	6,15E-08
cg10625758	chr17	73851514	-	0,076138013	0,104029803	-0,02789179	0,010555379	0,01933827	3,16E-11
cg10633147	chr2	68384753	-	0,067919608	0,056249886	0,011669722	0,008147624	0,011409349	1,24E-08
cg10633981	chr11	16779768	-	0,723240697	0,595226869	0,128013828	0,060693226	0,065709755	1,85E-16
cg10850054	chr5	55290839	+	0,036638394	0,030369876	0,006268517	0,00429029	0,004820925	3,69E-08
cg10880177	chr5	1729256	+	0,93864449	0,926774256	0,011870234	0,011035968	0,023339638	1,03E-08
cg10905877	chr3	42845934	+	0,251877677	0,212378122	0,039499556	0,039302888	0,02672771	8,51E-09
cg10909163	chr13	41635251	+	0,055495741	0,077344465	-0,021848724	0,010141045	0,019407769	7,30E-08
cg10921517	chr11	1520881	+	0,636372899	0,612572064	0,023800836	0,022903173	0,014575861	9,10E-08
cg10981770	chr1	38272531	+	0,550162939	0,497895787	0,052267152	0,028951449	0,057028216	3,79E-08
cg10983056	chr13	113344077	-	0,079434402	0,101013484	-0,021579082	0,01158739	0,018101192	5,68E-12
cg10988041	chr14	38060564	+	0,169180591	0,216249116	-0,047068524	0,019816094	0,048467343	1,07E-08
cg11064073	chr4	164253938	+	0,059281212	0,038922931	0,02035828	0,013230059	0,011580509	2,11E-12
cg11079354	chr6	27740361	-	0,0705116	0,04709562	0,02341598	0,014146819	0,017537849	1,03E-10
cg11227075	chr10	52750930	+	0,178922048	0,224721988	-0,045799939	0,021204161	0,040351272	8,23E-09
cg11553596	chr7	128338932	+	0,891398092	0,86932652	0,022071572	0,019488273	0,020203276	1,17E-10
cg11564483	chr6	37401398	-	0,090592441	0,120163616	-0,029571175	0,01288435	0,026669309	2,20E-08
cg11651961	chr2	54197711	+	0,090539234	0,12117924	-0,030640006	0,011771227	0,019138749	9,62E-13

cg11688874	chr10	28822482	+	0,052334183	0,03265112	0,019683062	0,011637317	0,016158198	4,07E-13
cg11706469	chr10	103535362	-	0,245091443	0,289955838	-0,044864394	0,025499557	0,048068459	6,40E-08
cg11725182	chr1	5968678	-	0,67924884	0,739483523	-0,060234683	0,046878738	0,056803017	8,51E-08
cg11733294	chr11	118123053	-	0,106019578	0,14299585	-0,036976272	0,017307618	0,031708765	4,26E-09
cg11788103	chr19	12833533	-	0,083716473	0,113387015	-0,029670542	0,014946996	0,025324226	4,99E-08
cg11806528	chr1	113051977	+	0,174579207	0,142478755	0,032100452	0,029569114	0,047197675	4,58E-08
cg11815914	chr18	59992350	+	0,042831026	0,036190991	0,006640035	0,006828967	0,003984089	3,02E-08
cg11888381	chr3	184053532	+	0,167218506	0,214361163	-0,047142657	0,022157486	0,028698693	3,52E-10
cg11932640	chr4	4861200	-	0,061195818	0,083774487	-0,022578669	0,009479001	0,018119642	1,88E-09
cg12027761	chr17	38279750	+	0,100813033	0,138515818	-0,037702785	0,01449368	0,026166651	2,08E-13
cg12046629	chr8	59466299	-	0,049397561	0,063014135	-0,013616574	0,008231808	0,014148772	3,78E-08
cg12098645	chr12	69326716	+	0,06209175	0,087647097	-0,025555348	0,010895772	0,020222189	2,04E-09
cg12162201	chr21	48055631	-	0,022168017	0,01891663	0,003251388	0,002738886	0,003081822	2,54E-08
cg12223960	chr6	32063403	-	0,724906418	0,68911422	0,035792199	0,027765099	0,018995231	6,19E-08
cg12241367	chr10	13203349	+	0,141747914	0,192680283	-0,050932368	0,02070157	0,033662699	2,37E-10
cg12250883	chr12	63328707	+	0,217752521	0,184961775	0,032790746	0,021396365	0,035758297	6,14E-09
cg12260146	chr3	33840237	+	0,231048126	0,276655232	-0,045607106	0,016669867	0,036346912	2,99E-12
cg12261117	chr19	70918	+	0,059487838	0,049757376	0,009730462	0,010263485	0,029190704	2,18E-09
cg12444684	chr1	65431340	-	0,073160075	0,102480234	-0,029320159	0,012029911	0,023360205	2,35E-10
cg12457721	chr3	129822854	-	0,715752251	0,684705841	0,03104641	0,03772546	0,028706249	4,97E-08
cg12619612	chr7	100428744	-	0,945440054	0,929696168	0,015743885	0,013171434	0,023081369	1,26E-08
cg12627537	chr22	43485414	+	0,151625907	0,121121227	0,03050468	0,029805005	0,036959183	1,38E-08
cg12666107	chr1	31381697	+	0,122673598	0,105190047	0,017483552	0,012272455	0,019275353	8,93E-08
cg12832228	chr8	28244096	+	0,078586968	0,104575031	-0,025988063	0,011036036	0,023564581	2,58E-08
cg12878213	chr20	17550481	+	0,041196119	0,052976846	-0,011780727	0,006497954	0,009982092	2,80E-08
cg12922032	chr2	32582209	+	0,102684244	0,144802788	-0,042118543	0,015014237	0,030807622	7,75E-12
cg12929040	chr8	56792373	-	0,085220414	0,117480978	-0,032260564	0,012662297	0,026477368	7,96E-10
cg12950166	chr12	122517149	+	0,091313258	0,121217269	-0,029904011	0,017149453	0,028962656	3,30E-08

cg12963096	chr13	39612675	+	0,067843091	0,092855955	-0,025012864	0,010064978	0,020732268	1,17E-10
cg12989020	chr6	157098657	-	0,070883289	0,057722033	0,013161256	0,008360004	0,014474389	7,50E-09
cg13088939	chr1	155659111	+	0,039571348	0,055273621	-0,015702273	0,007412591	0,013977723	4,98E-09
cg13093793	chr14	35099518	-	0,250390913	0,294303128	-0,043912215	0,023608614	0,032694351	4,66E-09
cg13241473	chr6	41703039	-	0,035126333	0,025966589	0,009159744	0,00912971	0,011447112	1,77E-08
cg13286318	chr8	57906058	+	0,135437197	0,180695084	-0,045257888	0,019008383	0,037281786	9,62E-13
cg13325330	chr6	44355541	+	0,055914707	0,077361638	-0,021446931	0,00944938	0,016795086	2,51E-08
cg13326048	chr6	28641863	-	0,061664062	0,04308142	0,018582642	0,013783764	0,013978222	5,92E-08
cg13501090	chr1	115212659	-	0,192040137	0,234193415	-0,042153278	0,018212276	0,028023574	9,52E-12
cg13501181	chr3	167967446	-	0,048929362	0,064300145	-0,015370783	0,00796375	0,01550869	2,68E-08
cg13512204	chr9	78506874	+	0,245735655	0,287549527	-0,041813872	0,020957683	0,029408448	1,61E-11
cg13531588	chr14	75725958	+	0,052617119	0,072852466	-0,020235347	0,008003185	0,015170614	2,07E-11
cg13554667	chr10	62538283	-	0,105884451	0,084029146	0,021855305	0,022118153	0,022699376	3,89E-08
cg13565624	chr19	913277	-	0,247276206	0,17969684	0,067579366	0,036522025	0,052199359	1,90E-08
cg13602813	chr3	13920840	+	0,119675287	0,152184815	-0,032509527	0,017762981	0,028125931	9,88E-08
cg13645767	chr9	97766849	-	0,07192067	0,094770267	-0,022849597	0,009820605	0,019064555	1,32E-10
cg13758310	chr4	1834493	-	0,880112761	0,84480846	0,035304302	0,020848246	0,0276273	6,47E-08
cg13777502	chr17	66031814	-	0,094907562	0,132249653	-0,03734209	0,016707938	0,026092502	1,32E-10
cg13787029	chr2	42721631	-	0,096151866	0,140465321	-0,044313454	0,013745144	0,030917837	2,95E-12
cg13810766	chr7	151542452	+	0,386653447	0,328808475	0,057844972	0,03275894	0,045915202	2,83E-11
cg13817083	chr7	159026609	+	0,897473496	0,877389595	0,020083902	0,018527487	0,023229087	6,16E-08
cg13938963	chr4	68411314	-	0,104059279	0,139004164	-0,034944885	0,016416401	0,024726011	1,76E-10
cg13960352	chr19	45826863	+	0,1081368	0,137185935	-0,029049135	0,016630756	0,0271215	1,12E-08
cg13977374	chr9	86152905	+	0,076071309	0,099544156	-0,023472847	0,013461424	0,020135449	9,88E-08
cg13997788	chr2	46593223	-	0,856914755	0,817953535	0,03896122	0,0238486	0,038044723	1,27E-08
cg14087168	chr1	156450669	+	0,112526307	0,152111931	-0,039585624	0,020130393	0,035984189	2,95E-09
cg14183907	chr9	136114344	+	0,058914847	0,083676398	-0,024761551	0,008931427	0,019808621	4,88E-10
cg14212738	chr11	82783003	-	0,059603003	0,087708837	-0,028105833	0,011683479	0,023241356	8,56E-09

cg14291009	chr22	18560618	-	0,049660532	0,07434219	-0,024681658	0,011187307	0,021552104	1,87E-08
cg14312959	chr17	79008933	-	0,061475411	0,048042284	0,013433127	0,008571964	0,014613318	4,70E-08
cg14347670	chr6	41908995	-	0,047830042	0,033164351	0,01466569	0,012251478	0,013301289	2,45E-08
cg14397416	chr15	78219273	+	0,84497317	0,8017306	0,04324257	0,022971543	0,040688134	8,90E-09
cg14497545	chr4	140754475	+	0,490508765	0,392434954	0,098073812	0,043196065	0,049287279	4,56E-15
cg14530304	chr17	77777792	-	0,107452171	0,140782918	-0,033330747	0,017323434	0,028656258	6,35E-11
cg14728310	chr19	9879317	+	0,086302445	0,113567101	-0,027264656	0,013602965	0,021326099	6,55E-08
cg14733031	chr19	17904427	-	0,506121451	0,394844442	0,111277009	0,096748044	0,086636857	1,34E-08
cg14770570	chr17	73975813	-	0,055223837	0,080966725	-0,025742888	0,009036058	0,021910776	1,09E-09
cg14837635	chr1	116943616	-	0,950393095	0,936224917	0,014168177	0,007687349	0,011326741	2,22E-09
cg14848555	chr1	171750797	-	0,110279618	0,094611967	0,015667651	0,01285554	0,0109381	1,69E-09
cg14905632	chr17	38333933	+	0,078283322	0,108231548	-0,029948227	0,015321461	0,025268584	4,42E-09
cg14911132	chr11	61596755	+	0,077747023	0,10662564	-0,028878617	0,014104537	0,02436289	3,63E-12
cg14932796	chr17	55038330	-	0,03669357	0,028291211	0,008402358	0,008497044	0,009504359	7,86E-08
cg14935078	chr19	17414399	-	0,080210555	0,108901022	-0,028690467	0,017079039	0,027831605	5,99E-09
cg14957430	chr2	64995219	-	0,083719868	0,112266797	-0,028546929	0,01205619	0,023165035	8,60E-09
cg15044954	chr7	89874365	+	0,053825456	0,040847288	0,012978169	0,010520532	0,011118773	9,34E-08
cg15100630	chr6	30524468	+	0,110826118	0,141736082	-0,030909964	0,013953205	0,025967072	3,29E-09
cg15103181	chr19	35454479	-	0,165487501	0,210662605	-0,045175104	0,018849774	0,035846483	1,38E-10
cg15153914	chr20	44563668	-	0,045437504	0,03714152	0,008295983	0,006810302	0,005396961	1,52E-08
cg15192143	chr17	77787315	-	0,27509461	0,222829422	0,052265188	0,033005616	0,040999612	1,11E-08
cg15242061	chr7	90226535	+	0,149661855	0,193619607	-0,043957751	0,022076295	0,034721042	2,35E-10
cg15248989	chr19	14682900	+	0,186252704	0,237229537	-0,050976833	0,021371477	0,029796877	4,35E-13
cg15260268	chr12	53613892	+	0,0946248	0,131645339	-0,037020539	0,01358853	0,029237399	2,79E-08
cg15321108	chr14	102227943	-	0,1093023	0,085648114	0,023654186	0,018017361	0,024498991	3,24E-09
cg15352431	chr2	226896008	+	0,090682117	0,120998495	-0,030316379	0,013899444	0,027773392	2,45E-08
cg15421363	chr5	131826865	-	0,088594098	0,115593591	-0,026999492	0,013041388	0,023579881	1,77E-08
cg15558129	chr1	219347279	+	0,117981243	0,161928508	-0,043947264	0,016169617	0,030061913	1,19E-12

cg15656470	chr6	29134247	-	0,69578853	0,753113843	-0,057325313	0,046514004	0,069901529	4,49E-10
cg15684724	chr8	67875033	+	0,05300318	0,070684325	-0,017681146	0,008867105	0,016662433	4,75E-08
cg15739717	chr3	136537790	+	0,04749629	0,040371931	0,00712436	0,004642672	0,005485525	1,31E-08
cg15799226	chr15	68115192	+	0,10444221	0,135042734	-0,030600525	0,020320026	0,027729326	8,58E-09
cg15859278	chr4	39770439	-	0,917836125	0,897807755	0,020028369	0,010926618	0,021857295	2,10E-08
cg15869463	chr17	62777690	-	0,115494603	0,158727914	-0,043233311	0,017796139	0,032983247	4,38E-10
cg15879179	chr4	164088111	+	0,059606793	0,047452553	0,012154239	0,009101526	0,008328964	4,83E-09
cg15889012	chr2	44059266	-	0,116194864	0,15050593	-0,034311066	0,017135633	0,030533807	1,33E-08
cg15897635	chr1	220697615	-	0,512699203	0,409060051	0,103639152	0,050497254	0,058458135	1,82E-14
cg16079347	chr9	100265181	-	0,149009199	0,124550712	0,024458486	0,017249454	0,026142263	5,33E-09
cg16101278	chr6	6648823	+	0,377224865	0,295445995	0,08177887	0,06618896	0,076817719	7,21E-12
cg16141325	chr8	90914397	-	0,054576128	0,076828531	-0,022252403	0,008771485	0,017903799	7,63E-11
cg16154073	chr3	32433121	-	0,076425315	0,11374837	-0,037323055	0,013554328	0,028302325	4,39E-10
cg16163697	chr1	113498742	+	0,051309698	0,041082157	0,01022754	0,009723406	0,016120038	1,40E-08
cg16261730	chr1	24742018	-	0,055738639	0,076488874	-0,020750236	0,008600458	0,015770813	4,28E-11
cg16361301	chr1	218519549	-	0,048248232	0,070108601	-0,021860369	0,010689887	0,018565127	5,13E-08
cg16404530	chr2	98964691	-	0,188421886	0,237841663	-0,049419777	0,026992053	0,042213998	6,62E-11
cg16422343	chr6	36564708	-	0,78583997	0,720071154	0,065768816	0,029453528	0,059825389	1,41E-09
cg16435539	chr3	39851009	+	0,054775483	0,043833212	0,010942271	0,007887695	0,008810848	5,55E-08
cg16449659	chr4	153701590	-	0,070384442	0,097176119	-0,026791677	0,012653515	0,023962786	4,63E-08
cg16574871	chr8	97165770	+	0,235948216	0,271415666	-0,035467449	0,027290667	0,047133936	3,42E-08
cg16593921	chr8	6693344	+	0,092716004	0,125170692	-0,032454688	0,018944924	0,026903173	1,27E-10
cg16656864	chr17	27718261	-	0,084711547	0,120871274	-0,036159727	0,016785887	0,02852659	7,93E-10
cg16758970	chr13	28527699	-	0,108594568	0,141750859	-0,033156291	0,015980974	0,028141333	7,93E-10
cg16822387	chr11	31831591	-	0,127928397	0,167167176	-0,039238779	0,019064353	0,035794851	3,47E-08
cg16864700	chr12	31226536	-	0,110137365	0,149202535	-0,039065169	0,015033125	0,029658517	1,18E-10
cg16944159	chr2	62132759	-	0,042370842	0,031628859	0,010741983	0,006484605	0,009889324	5,66E-09
cg16963286	chr7	33102443	+	0,075619505	0,059710481	0,015909024	0,019792828	0,025629446	5,30E-08

cg17015704	chr6	119400380	+	0,135780795	0,184272391	-0,048491595	0,021958708	0,03838603	4,40E-09
cg17024643	chr1	214725274	+	0,12677183	0,168387785	-0,041615955	0,016606545	0,030121274	9,83E-12
cg17035366	chr12	40499623	-	0,100281188	0,130864755	-0,030583567	0,011144539	0,024125397	3,50E-10
cg17059624	chr11	85358760	+	0,046498655	0,033089454	0,013409201	0,008610188	0,01196005	3,99E-10
cg17144760	chr11	27384507	+	0,067625908	0,101242558	-0,03361665	0,012071526	0,025762297	1,88E-12
cg17150246	chr1	161068070	+	0,056436791	0,043204994	0,013231797	0,008165634	0,01011593	5,89E-09
cg17157516	chr1	35332203	+	0,034165412	0,026372877	0,007792535	0,006078852	0,008747699	4,83E-08
cg17246714	chr3	55519186	-	0,07670126	0,103493987	-0,026792727	0,014172079	0,026232299	3,60E-08
cg17502989	chr6	153451486	+	0,141546903	0,099828474	0,04171843	0,035141827	0,042449701	3,00E-09
cg17554502	chr17	44271559	+	0,035483098	0,047393192	-0,011910094	0,00573956	0,01111705	3,76E-09
cg17595353	chr2	176951078	+	0,108185174	0,142069137	-0,033883963	0,015810398	0,02330826	2,34E-14
cg17665931	chr19	38894173	+	0,059957028	0,07911091	-0,019153882	0,008425132	0,01716657	8,61E-08
cg17720267	chr12	6580010	+	0,050468716	0,074821123	-0,024352407	0,010564029	0,019475935	1,69E-08
cg17772342	chr5	134871686	-	0,121734904	0,156579539	-0,034844635	0,019372568	0,032618337	3,24E-09
cg17809276	chr1	172501547	+	0,055438046	0,070947785	-0,015509738	0,007633123	0,014747101	5,02E-09
cg17861791	chr1	113162073	+	0,098993169	0,135219063	-0,036225893	0,013718783	0,030204342	2,75E-09
cg17951138	chr5	92908392	+	0,11451637	0,159171309	-0,044654939	0,023880661	0,039235259	8,65E-10
cg18011078	chr7	1971098	-	0,825616669	0,76657276	0,059043909	0,038988734	0,052502949	1,46E-09
cg18080401	chr12	49365529	-	0,057152311	0,081515162	-0,024362852	0,010939097	0,019328132	2,06E-08
cg18082638	chr11	31827757	-	0,116046888	0,163334248	-0,04728736	0,017144622	0,035803012	3,84E-15
cg18122310	chr12	50236657	+	0,03962849	0,05049322	-0,01086473	0,005138787	0,009855409	4,97E-08
cg18159740	chr13	30881703	-	0,086624864	0,120120855	-0,03349599	0,014387324	0,025515887	2,36E-09
cg18169128	chr14	53619322	-	0,050842923	0,068210264	-0,017367341	0,008372998	0,015305244	7,73E-10
cg18187726	chr12	69080414	+	0,08620899	0,11510901	-0,028900019	0,016946062	0,02404891	4,35E-08
cg18219180	chr2	204800350	-	0,871869852	0,825418813	0,046451038	0,025525616	0,029156103	7,02E-12
cg18266479	chr6	30419005	+	0,131090568	0,181537159	-0,050446591	0,024292395	0,031966289	1,24E-09
cg18382893	chr6	30139641	+	0,057835362	0,074784117	-0,016948755	0,00989061	0,01442709	7,23E-09
cg18404039	chr20	3184938	-	0,052751206	0,074551317	-0,02180011	0,008259787	0,020238415	7,49E-10

cg18441082	chr5	16466057	-	0,104252898	0,147013542	-0,042760644	0,016412364	0,033578074	2,52E-09
cg18481683	chr19	39574911	-	0,107241257	0,137932416	-0,030691159	0,014323381	0,022338037	4,95E-10
cg18500431	chr7	1709489	-	0,106012629	0,072258552	0,033754077	0,020044687	0,022374764	2,95E-09
cg18538297	chr19	58892579	-	0,268139321	0,24185435	0,026284972	0,017255023	0,038299474	7,23E-10
cg18615184	chr17	61678500	-	0,048569812	0,06845218	-0,019882368	0,007226645	0,015448592	1,86E-10
cg18635645	chr1	85667791	+	0,048331005	0,031023701	0,017307304	0,011933639	0,01102681	1,65E-09
cg18782651	chr7	65969785	-	0,638766439	0,729799711	-0,091033272	0,055061021	0,08408876	1,37E-08
cg18802567	chr5	171433913	+	0,105047048	0,145371162	-0,040324114	0,014989966	0,030598873	4,35E-11
cg18827097	chr2	3522506	+	0,053829312	0,035838264	0,017991048	0,012825676	0,012048993	2,28E-09
cg18877862	chr22	43116932	+	0,279709941	0,240245325	0,039464616	0,032663316	0,047551618	8,96E-10
cg18987274	chr7	132766962	+	0,052436644	0,078663612	-0,026226969	0,008551308	0,019555617	2,99E-12
cg19171792	chr12	3427463	+	0,678606306	0,737457739	-0,058851433	0,039825163	0,052711584	1,14E-08
cg19226203	chr22	29833658	+	0,924002036	0,903928028	0,020074007	0,015108337	0,014477137	1,14E-08
cg19269264	chr2	238600693	-	0,16226307	0,117272765	0,044990305	0,044846548	0,052180906	3,41E-08
cg19344263	chr19	58919951	+	0,04072277	0,053752607	-0,013029837	0,008291831	0,012848371	3,69E-08
cg19387006	chr6	45345784	-	0,027130782	0,023400433	0,003730349	0,003215838	0,003171251	2,04E-09
cg19484645	chr6	4995292	-	0,883409776	0,859415521	0,023994255	0,017404982	0,025228449	2,49E-08
cg19534005	chr17	75939819	+	0,913992439	0,898697304	0,015295135	0,019873191	0,035663028	1,38E-08
cg19571046	chr19	12175719	-	0,051021037	0,073783383	-0,022762346	0,010715826	0,016067965	7,38E-12
cg19608455	chr18	48086580	+	0,023222247	0,020239903	0,002982344	0,00211749	0,004633462	8,53E-08
cg19613905	chr15	56757180	-	0,078083609	0,054081445	0,024002164	0,016248964	0,017443928	3,50E-10
cg19752602	chr5	153039227	+	0,624773444	0,727050092	-0,102276648	0,060803077	0,083131136	3,51E-08
cg19795556	chr10	60494185	+	0,926277813	0,909293294	0,016984519	0,011852986	0,024580249	5,58E-09
cg19819285	chr19	58545290	+	0,040029219	0,032175336	0,007853884	0,008162318	0,006573664	1,38E-08
cg19824334	chr17	73851399	-	0,054429865	0,075783091	-0,021353226	0,008700183	0,018284391	9,98E-08
cg19835839	chr12	120966954	-	0,045898995	0,065184001	-0,019285006	0,011898118	0,014586044	1,43E-08
cg19869610	chr19	45996498	-	0,154174037	0,201953645	-0,047779608	0,020253151	0,039989197	9,41E-09
cg19890431	chr11	3659810	-	0,666271542	0,630987647	0,035283894	0,038618487	0,032496062	5,21E-11

cg19977691	chr22	50480018	+	0,925603492	0,905895177	0,019708315	0,020412751	0,031873509	7,42E-09
cg20000718	chr4	41869433	+	0,071030581	0,099423128	-0,028392547	0,012786139	0,024265334	1,05E-10
cg20015168	chr14	57046681	-	0,05570279	0,073745736	-0,018042946	0,008345633	0,015272099	2,27E-08
cg20034209	chr17	66287369	-	0,093041878	0,128560733	-0,035518855	0,013391484	0,02692884	1,52E-11
cg20089531	chr14	92573973	-	0,877604705	0,843705525	0,03389918	0,027202494	0,042826496	8,96E-10
cg20164915	chr1	149871167	-	0,089424479	0,119585002	-0,030160523	0,012970029	0,022803036	1,66E-10
cg20179222	chr3	120461682	-	0,036895198	0,048963906	-0,012068708	0,005816298	0,010383207	1,62E-08
cg20237920	chr15	76225808	-	0,739847883	0,812950938	-0,073103055	0,04864511	0,069018238	1,28E-09
cg20250396	chr14	65880036	+	0,090323324	0,113696856	-0,023373531	0,011600232	0,02062963	8,27E-09
cg20264566	chr11	10316295	-	0,081406472	0,106495564	-0,025089092	0,012751694	0,023431123	5,57E-08
cg20318726	chr4	13549621	-	0,069453572	0,094612546	-0,025158975	0,013934114	0,02358903	5,67E-09
cg20494553	chr4	71554112	-	0,102811092	0,142858313	-0,040047221	0,015595231	0,030943104	5,83E-11
cg20518154	chr20	39318705	+	0,088721396	0,120943566	-0,032222169	0,014090868	0,025381814	5,71E-08
cg20547653	chr2	108603456	+	0,155180481	0,183860992	-0,028680511	0,025535569	0,03129199	3,45E-08
cg20604317	chr20	4129314	+	0,098157334	0,134064913	-0,035907579	0,016040829	0,025999355	1,89E-09
cg20623868	chr19	15666885	+	0,139652642	0,183238179	-0,043585537	0,018598402	0,030100831	2,43E-12
cg20683799	chr1	63989117	+	0,044456076	0,060900814	-0,016444738	0,007568954	0,015327971	6,29E-08
cg20709868	chr19	12992305	-	0,094152862	0,134722459	-0,040569598	0,012393691	0,028870352	2,87E-12
cg20730634	chr19	37709434	+	0,106368147	0,13783487	-0,031466723	0,013722101	0,025558652	8,97E-09
cg20884636	chr2	172290779	-	0,075562544	0,104434818	-0,028872274	0,012532547	0,026224122	2,43E-08
cg20913106	chr8	38324522	-	0,072263304	0,047967423	0,024295881	0,013330672	0,016249338	2,95E-12
cg20956633	chr20	19997839	-	0,074699021	0,102585462	-0,027886441	0,012351158	0,023939922	7,69E-09
cg20973743	chr15	49447782	-	0,124277152	0,178745429	-0,054468277	0,016795295	0,036835644	1,76E-12
cg21099776	chr17	17739742	-	0,061640417	0,083538725	-0,021898308	0,010147028	0,017421569	3,56E-09
cg21144462	chr17	42082647	+	0,20813842	0,253023665	-0,044885244	0,021379796	0,034889288	1,36E-10
cg21153962	chr11	57102758	+	0,058504623	0,077630443	-0,01912582	0,009274479	0,015015594	1,77E-08
cg21171314	chr11	16628640	+	0,060955681	0,046032165	0,014923517	0,011638653	0,012893027	7,50E-09
cg21242584	chr8	25041785	-	0,137192056	0,172777033	-0,035584977	0,01844267	0,028354733	1,12E-08

cg21282054	chr17	76127608	-	0,072105011	0,101665771	-0,02956076	0,015000767	0,03657123	6,73E-09
cg21284370	chr13	42621865	+	0,142188754	0,189909774	-0,047721021	0,026239889	0,03853687	1,78E-10
cg21292957	chr21	34602922	+	0,046881435	0,065840884	-0,01895945	0,0076394	0,015839555	1,34E-09
cg21304062	chr17	78075816	+	0,061616996	0,08104306	-0,019426064	0,010583616	0,019241409	9,69E-08
cg21304211	chr1	112281877	-	0,219425229	0,273104361	-0,053679133	0,020727528	0,039649981	3,90E-12
cg21397124	chr4	11386127	-	0,455777939	0,390831827	0,064946112	0,04630257	0,042387111	9,99E-12
cg21423973	chr7	72298667	+	0,171820925	0,222576788	-0,050755863	0,018696952	0,031567513	1,82E-14
cg21515243	chr18	21033072	+	0,077994749	0,05366088	0,024333869	0,012603986	0,022075531	1,49E-09
cg21625563	chr22	39152448	-	0,069359643	0,094356898	-0,024997254	0,007840504	0,019219528	1,11E-09
cg21638374	chr5	147763491	-	0,112608293	0,154454703	-0,041846409	0,018688681	0,026170142	1,48E-11
cg21714212	chr6	30647104	-	0,091773845	0,122371863	-0,030598017	0,019698677	0,028987542	4,10E-08
cg21761427	chr6	28220064	+	0,052469588	0,067699906	-0,015230318	0,008576525	0,014052167	7,96E-08
cg21776568	chr13	95254007	+	0,075540816	0,111151861	-0,035611045	0,013412434	0,029126174	1,97E-10
cg21846305	chr12	45270304	+	0,133518863	0,179366619	-0,045847756	0,02057722	0,037738282	7,15E-08
cg21942256	chr3	156892528	+	0,07167584	0,104620703	-0,032944863	0,011607813	0,026713245	1,54E-11
cg21950287	chr19	54385441	+	0,121903979	0,164444552	-0,042540573	0,020724859	0,033864914	2,46E-08
cg21980338	chr4	331929	+	0,356608427	0,310045067	0,04656336	0,032755998	0,043462969	1,42E-08
cg22062659	chr8	70747431	+	0,057955582	0,077625543	-0,019669961	0,009266188	0,016770977	6,04E-09
cg22270384	chr19	12848240	+	0,056197849	0,08072172	-0,024523871	0,009289276	0,020818527	2,73E-09
cg22325196	chr17	47492792	+	0,052375574	0,074601516	-0,022225942	0,010900204	0,020387526	6,91E-13
cg22433398	chr13	27825486	+	0,181393742	0,224545664	-0,043151923	0,022920007	0,028462611	4,93E-09
cg22435132	chr17	59312558	+	0,813277701	0,765935833	0,047341868	0,034699421	0,054162891	3,22E-08
cg22471230	chr3	154797779	+	0,047983379	0,065112432	-0,017129053	0,014581402	0,020660245	3,21E-08
cg22488745	chr15	53096763	-	0,077074659	0,104663004	-0,027588345	0,012194543	0,02163723	2,87E-08
cg22798468	chr1	225150341	+	0,640714189	0,706274655	-0,065560466	0,036767751	0,057827567	1,16E-08
cg22805909	chr14	96967825	+	0,054766067	0,037176705	0,017589362	0,011277559	0,015777504	2,45E-08
cg22854549	chr19	5904785	-	0,059341753	0,077121263	-0,017779511	0,011469966	0,014674715	8,48E-09
cg22870994	chr11	71340352	-	0,525123836	0,452035779	0,073088057	0,032953276	0,071346088	1,38E-09

cg23055081	chr6	30181315	-	0,108176172	0,150259091	-0,042082919	0,017413681	0,024378636	7,76E-13
cg23164678	chr6	83775487	-	0,094432608	0,127292804	-0,032860196	0,010738832	0,023886661	1,50E-09
cg23176270	chr14	23476067	-	0,047756366	0,066789797	-0,019033432	0,007754625	0,016803029	3,19E-09
cg23215275	chr13	19301347	+	0,913055807	0,869914555	0,043141251	0,012146233	0,03370814	1,82E-08
cg23218665	chr1	11159850	+	0,087700127	0,119558621	-0,031858495	0,014239669	0,029164917	3,23E-08
cg23223044	chr19	49990882	-	0,146535721	0,188342945	-0,041807224	0,017625629	0,029972301	8,92E-12
cg23243902	chr11	117049849	-	0,107902146	0,14987213	-0,041969984	0,017030006	0,031239756	3,06E-10
cg23274247	chr7	38358683	+	0,050004729	0,037672587	0,012332142	0,009345653	0,009021442	9,79E-08
cg23359048	chr12	2921757	+	0,071506635	0,104682392	-0,033175757	0,012566132	0,027523937	6,56E-11
cg23378609	chr5	178054312	-	0,055251946	0,046455734	0,008796212	0,008218251	0,01361448	1,47E-08
cg23453437	chr7	148892747	-	0,06491803	0,089765252	-0,024847222	0,01026678	0,021239149	3,11E-08
cg23598419	chr2	25391042	+	0,086268177	0,107964445	-0,021696268	0,015680696	0,022520469	4,63E-08
cg23606758	chr1	52869977	+	0,062400228	0,088428806	-0,026028578	0,010633159	0,020935764	6,37E-10
cg23616126	chr1	247463678	+	0,919516013	0,903814815	0,015701198	0,014097624	0,014797481	9,64E-08
cg23642001	chr8	67025419	-	0,124475613	0,177336453	-0,05286084	0,019093731	0,034312033	2,93E-13
cg23705487	chr8	91013396	+	0,051898198	0,043934689	0,007963509	0,009605487	0,005097971	9,28E-09
cg23783653	chr6	27059850	+	0,089656194	0,117045829	-0,027389636	0,018678439	0,025857349	3,11E-08
cg23797200	chr17	40174494	-	0,487559331	0,404614873	0,082944458	0,078029087	0,063099757	1,88E-12
cg23850707	chr15	55582494	+	0,057579112	0,044623686	0,012955427	0,009372657	0,014546923	8,23E-08
cg23872756	chr20	42543407	+	0,04076504	0,034860354	0,005904686	0,00420668	0,006320685	5,19E-08
cg23886783	chr2	113403248	-	0,046422615	0,064452171	-0,018029555	0,007588999	0,01527117	1,77E-09
cg23922560	chr14	37117375	-	0,115994271	0,160568655	-0,044574384	0,023386811	0,040059816	5,35E-11
cg24009109	chr3	47018482	+	0,041869333	0,030663469	0,011205864	0,009890249	0,007282616	6,51E-08
cg24038058	chr6	1611745	+	0,062636647	0,077928124	-0,015291477	0,010924396	0,016861273	2,87E-08
cg24061580	chr7	151573966	+	0,119135376	0,16201072	-0,042875343	0,018333067	0,035475453	2,10E-08
cg24095222	chr5	161178787	+	0,711432356	0,658308782	0,053123574	0,055765951	0,037287434	1,18E-08
cg24105280	chr12	6797846	+	0,060467624	0,078790697	-0,018323073	0,00882346	0,016032554	7,45E-08
cg24127861	chr14	24640947	+	0,062950786	0,083609182	-0,020658395	0,012108271	0,018702722	1,89E-08

cg24258125	chr22	44287772	-	0,150175512	0,195065648	-0,044890136	0,021495757	0,034570554	1,99E-08
cg24293614	chr22	43807429	-	0,147454864	0,190007062	-0,042552198	0,022351973	0,033857775	1,30E-11
cg24303698	chr19	46367351	-	0,063642405	0,085069274	-0,02142687	0,009631846	0,015344534	2,24E-09
cg24321688	chr19	39393347	+	0,931292329	0,920975507	0,010316821	0,012398457	0,01745955	5,26E-09
cg24377330	chr3	149375961	+	0,047960214	0,071943185	-0,023982971	0,00859892	0,021439157	4,10E-08
cg24396566	chr4	52904500	+	0,110192443	0,089655564	0,020536878	0,026949055	0,014372136	9,05E-08
cg24413235	chr2	27851648	+	0,171394673	0,220484755	-0,049090082	0,024966701	0,033633881	1,82E-08
cg24430189	chr2	128173685	-	0,06505932	0,046873114	0,018186206	0,012911485	0,013963306	4,94E-08
cg24451422	chr1	23751210	-	0,086311432	0,066382617	0,019928815	0,014388724	0,019706799	5,10E-09
cg24485820	chr12	117536718	+	0,060002467	0,087239183	-0,027236716	0,011222854	0,020949631	2,11E-12
cg24517252	chr1	249200781	-	0,1170673	0,093372303	0,023694997	0,020650522	0,032036271	3,05E-08
cg24657816	chr4	1140440	+	0,699981167	0,672575207	0,027405959	0,039219817	0,024759878	8,30E-08
cg24806732	chr17	7608859	-	0,105204182	0,143531489	-0,038327307	0,015016207	0,033243696	1,34E-10
cg24824725	chr10	97453973	-	0,03270635	0,028278203	0,004428147	0,003979957	0,003716038	3,14E-08
cg24863706	chr4	120549163	+	0,08748878	0,108642449	-0,021153669	0,011994638	0,019391688	3,62E-08
cg24883642	chr15	40226588	+	0,120483539	0,166099747	-0,045616208	0,017670932	0,036697775	1,73E-08
cg24902113	chr13	33160718	+	0,053214024	0,071914262	-0,018700238	0,010955019	0,016338429	1,06E-08
cg24950336	chr3	113464879	+	0,087708423	0,116733215	-0,029024793	0,014028104	0,021811716	3,25E-08
cg25009327	chr7	148820504	-	0,275702395	0,204575156	0,071127239	0,038486688	0,032488438	9,19E-14
cg25053798	chr4	89206063	-	0,034480428	0,028901037	0,005579391	0,004202905	0,006615298	2,27E-08
cg25135004	chr1	167522579	+	0,15191227	0,115579965	0,036332305	0,023933842	0,03493586	1,57E-08
cg25142100	chr6	151547095	+	0,817122712	0,861242422	-0,04411971	0,028644626	0,042736319	4,84E-08
cg25228510	chr10	62538235	-	0,130724446	0,115078193	0,015646253	0,011739976	0,022935654	9,44E-08
cg25231972	chr10	31608136	+	0,176162293	0,215714643	-0,03955235	0,017713289	0,033488508	3,50E-09
cg25355465	chr20	36576591	-	0,541088692	0,623112142	-0,08202345	0,050668226	0,07295278	7,95E-08
cg25376393	chr8	143957809	-	0,587786744	0,639040817	-0,051254074	0,02814286	0,044833264	7,15E-09
cg25428810	chr11	49839836	-	0,58015232	0,646993317	-0,066840998	0,040022579	0,054521214	6,59E-08
cg25561913	chr10	77161129	+	0,07496209	0,101136541	-0,026174451	0,010223323	0,024615139	9,43E-10

cg25591664	chr5	79551865	+	0,143556222	0,099105263	0,044450959	0,021305246	0,045591982	1,29E-10
cg25613101	chr22	38577875	+	0,052514152	0,074583471	-0,022069319	0,009830883	0,018563337	2,20E-10
cg25614094	chr8	54164374	+	0,108410751	0,081690275	0,026720476	0,031704589	0,024736744	8,98E-08
cg25645657	chr5	52776244	+	0,055356519	0,084022034	-0,028665516	0,010038411	0,02200893	9,97E-13
cg25683204	chr20	34129839	+	0,06819471	0,092673462	-0,024478752	0,009163062	0,018668235	3,30E-11
cg25757017	chr12	10306462	+	0,092423471	0,134519988	-0,042096517	0,014441617	0,035609232	3,74E-12
cg25787812	chr6	31478830	+	0,907106524	0,882111317	0,024995206	0,01368704	0,031009334	9,81E-08
cg25828872	chr2	179058612	-	0,072195046	0,097137575	-0,024942529	0,011200716	0,018884757	1,87E-11
cg25837350	chr11	61584442	+	0,179833629	0,228016513	-0,048182884	0,024805794	0,035805932	4,35E-09
cg25837738	chr7	148725785	+	0,073407317	0,098268236	-0,024860918	0,012813015	0,021252649	6,01E-09
cg25856120	chr17	46970223	-	0,064346437	0,082871251	-0,018524814	0,009840926	0,015507005	2,33E-09
cg25857710	chr3	139654740	+	0,069280216	0,096690403	-0,027410186	0,011143452	0,020403207	1,91E-09
cg25876975	chr15	77363192	+	0,402473109	0,357138622	0,045334487	0,040115157	0,055413104	6,54E-09
cg25921418	chr6	31621008	+	0,068973371	0,088050055	-0,019076685	0,010313802	0,013927729	2,87E-09
cg26015973	chr10	103599705	+	0,152253731	0,188044315	-0,035790584	0,020254374	0,027971863	2,29E-08
cg26057840	chr20	62677629	+	0,856528424	0,823452833	0,033075591	0,021003771	0,028837962	3,56E-09
cg26067897	chr5	41904347	-	0,054407085	0,075248541	-0,020841456	0,009458363	0,017085715	9,44E-13
cg26105015	chr3	122745099	-	0,053256881	0,069692959	-0,016436078	0,010292936	0,017454556	2,68E-09
cg26135012	chr11	110583599	-	0,05786965	0,078608455	-0,020738805	0,010723172	0,021174282	7,29E-08
cg26144629	chr1	207926000	-	0,169423237	0,113836221	0,055587017	0,042290304	0,050168473	5,78E-08
cg26151079	chr11	58345687	-	0,085239008	0,116487296	-0,031248288	0,014232079	0,01965849	3,56E-09
cg26156860	chr19	34287218	-	0,034859659	0,029780075	0,005079584	0,004724328	0,004594681	8,07E-08
cg26164151	chr11	44749353	-	0,476192407	0,421233257	0,05495915	0,032405846	0,058768191	1,03E-09
cg26168400	chr2	202316215	+	0,034765869	0,029095467	0,005670402	0,003838113	0,003587657	2,90E-09
cg26180732	chr17	13972513	+	0,398500095	0,360690331	0,037809764	0,028627293	0,043612622	4,66E-09
cg26323130	chr2	149632683	-	0,073913414	0,098852355	-0,024938941	0,013406219	0,020074376	1,59E-09
cg26326615	chr3	18485765	-	0,085488926	0,119372698	-0,033883773	0,013381845	0,024978929	1,08E-12
cg26348180	chr3	193852754	-	0,244211639	0,287706177	-0,043494538	0,0194235	0,031937989	9,37E-12

cg26420824	chr1	28562883	-	0,088686199	0,116827653	-0,028141454	0,012182125	0,022081401	8,78E-10
cg26434548	chr7	129074607	-	0,080793266	0,104952102	-0,024158836	0,013382635	0,020775029	6,35E-09
cg26506725	chr6	31587990	-	0,074920801	0,105284301	-0,0303635	0,014160899	0,023430737	6,88E-10
cg26576155	chr15	85201829	-	0,057772249	0,082165928	-0,024393679	0,009352434	0,018929858	4,67E-09
cg26600954	chr7	23719792	+	0,094877602	0,132998733	-0,038121131	0,015201257	0,030035721	6,14E-10
cg26622699	chr17	46628303	-	0,833944045	0,785388462	0,048555583	0,022502157	0,068275794	4,60E-10
cg26639561	chr12	58239492	-	0,048499447	0,065912411	-0,017412964	0,00778838	0,013986695	8,81E-10
cg26683639	chr6	30653755	+	0,130496166	0,179265943	-0,048769777	0,022580275	0,053991583	4,20E-09
cg26770187	chr3	14693171	-	0,077197733	0,055488069	0,021709664	0,013343474	0,021114342	2,65E-08
cg26817334	chr2	119613580	+	0,053809491	0,076432494	-0,022623003	0,009267467	0,018229689	1,03E-10
cg26820911	chr4	83720044	+	0,033471774	0,04155921	-0,008087436	0,00473304	0,00880142	9,60E-09
cg26880762	chr12	123943063	+	0,085575009	0,124985971	-0,039410962	0,013811066	0,030948776	1,36E-10
cg26907082	chr1	18971840	-	0,173927818	0,221370914	-0,047443095	0,021702483	0,031980063	4,12E-12
cg27128322	chr18	43547338	+	0,047762646	0,038370471	0,009392174	0,008249975	0,0067799	8,56E-10
cg27161463	chr11	118123074	-	0,101129083	0,1383749	-0,037245817	0,014480986	0,030094071	9,52E-09
cg27179111	chr16	3627489	-	0,183721899	0,244272252	-0,060550354	0,018917235	0,070117418	3,24E-09
cg27183454	chr6	28547741	+	0,862409471	0,830571418	0,031838053	0,023653852	0,034374667	7,70E-10
cg27220681	chr17	685915	-	0,120175015	0,160824822	-0,040649807	0,018297543	0,024562301	4,44E-12
cg27247697	chr2	87018054	+	0,07249815	0,09557356	-0,02307541	0,011346341	0,017893403	7,23E-10
cg27282530	chr3	125898680	-	0,086906617	0,113669924	-0,026763307	0,013288881	0,022964369	5,03E-09
cg27379715	chr2	875990	+	0,690140883	0,623631443	0,066509441	0,041196318	0,046195738	3,40E-10
cg27404272	chr2	240447293	-	0,859702822	0,79902894	0,060673882	0,025878406	0,035570763	2,08E-13
cg27473538	chr4	15780522	+	0,089369123	0,127025593	-0,03765647	0,012403913	0,031487415	1,12E-10
cg27650678	chr11	66636318	+	0,080191218	0,109260535	-0,029069317	0,012861473	0,028841523	1,34E-08

Chapter 6

A non-coding expansion in *EIF4A3* causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects

Francine P. Favaro[1,10], Lucas Alvizi[2,10], Roseli M. Zechi-Ceide[1], Debora Bertola[2], Temis M. Felix[3], Josiane de Souza[4], Salmo Raskin[5], Stephen R. F. Twigg[6], Andrea M.J. Weiner[7], Pablo Armas[7], Ezequiel Margarit[7], Nora B. Calcaterra[7], Gregers R. Andersen[8], Simon McGowan[9], Andrew O. M. Wilkie[6], Antonio Richieri-Costa[1], Maria L. G. de Almeida[1], Maria Rita Passos-Bueno[2]

[1]*Departamento de Genética Clínica, Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo (HRAC-USP), Bauru, São Paulo, Brasil*
[2]*Centro de Pesquisas Sobre o Genoma Humano e Células-Tronco, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil*
[3]*Hospital das Clínicas de Porto Alegre, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brasil*
[4]*Centro de Atendimento Integral ao Fissurado Lábio Palatal, Curitiba, Paraná, Brasil*
[5]*Núcleo de Investigação Molecular Avançada, Pontifícia Universidade Católica do Paraná, Curitiba, Paraná, Brasil*
[6]*Clinical Genetics Group, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.*
[7]*Instituto de Biología Molecular y Celular de Rosario (IBR), Universidad Nacional de Rosario, Rosario, Argentina*
[8] *Department of Molecular Biology and Genetics. Aarhus University, Aarhus, Denmark*
[9]*Computational Biology Research Group, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK*

[10] *These authors contributed equally to this work*

Key-words: cleft palate, Richieri-Costa-Pereira Syndrome, acrofacial dysostosis, repeat expansion, EIF4A3.

This chapter was published in The American Journal of Human Genetics: DOI: http://dx.doi.org/10.1016/j.ajhg.2013.11.020

Abstract

Richieri-Costa-Pereira syndrome is an autosomal recessive acrofacial dysostosis characterized by mandibular median cleft associated with other craniofacial anomalies and severe limb defects. Learning and language disabilities are also prevalent. We mapped the mutated gene to a 122 kb region at 17q25.3 through identity-by-descent analysis in 17 genealogies. Sequencing strategies identified an expansion of a region with several repeats of 18- or 20-nucleotide motifs in the 5´ untranslated region (5´ UTR) of *EIF4A3*, which contained from 14 to 16 repeats in the affected individuals and from 3 to 12 repeats in 520 healthy individuals. A missense substitution of a highly conserved residue likely to affect the interaction of eIF4AIII with the UPF3B subunit of the exon junction complex in *trans* with an expanded allele was found in an unrelated individual with an atypical presentation, thus expanding mutational mechanisms and phenotypic diversity of RCPS. *EIF4A3* transcript abundance was reduced in both white blood cells and mesenchymal cells of RCPS individuals as compared to controls. Notably, targeting the orthologous *eif4a3* in zebrafish led to underdevelopment of several craniofacial cartilage and bone structures, in agreement with the craniofacial alterations seen in RCPS. Our data thus suggest that RCPS is caused by mutations in *EIF4A3* and show that *EIF4A3,* a gene involved in RNA metabolism*,* plays a role in mandible, laryngeal, and limb morphogenesis.

Resumo

A síndrome de Richieri-Costa-Pereira (*Richieri-Costa-Pereira Syndrome* - RCPS) é uma disostose acrofacial autossômica recessiva caracterizada por fissura média de mandíbula associada a outras anomalias craniofaciais e defeitos graves de membros. Outros sinais clínicos como deficiência de aprendizagem e fala também são prevalentes. Nós mapeamos o gene mutado em uma região de 122 kb no 17q25.3 por meio de análise de segmentos idênticos entre descendentes em 17 genealogias. Por meio de estratégias de sequenciamento, observou-se uma expansão de uma região com muitas repetições de motivos com 18 e 20 nucleotídeos na região 5' não traduzida (*5' untranslated region* – UTR) do *EIF4A3*, a qual contem de 14 a 16 repetições em indivíduos afetados e de 3 a 12 repetições em 520 indivíduos não afetados. Uma substituição *missense* em um resíduo altamente conservado, que provavelmente afeta a interação entre eIF4AIII e a subunidade UPF3B do complexo de junção exon-exon, foi encontrada em *trans* com um alelo expandido em um indivíduo com forma atípica da síndrome, mostrando-se, portanto, uma maior diversidade de mecanismos mutacionais e fenótipos em RCPS. Observou-se também a redução de transcritos do *EIF4A3* tanto em leucócitos quanto células mesenquimais de indivíduos RCPS em comparação a controles. Notavelmente, reduzindo-se a expressão do ortólogo *eif4a3* em *zebrafish* levou ao subdesenvolvimento de diversas cartilagens e ossos craniofaciais, o que está de acordo com as alterações observadas em RCPS. Portanto, nossos dados sugerem que RCPS é causada por mutações no *EIF4A3* e mostram que este gene, envolvido no metabolismo de RNA, tem papel na morfogênese da mandíbula, laringe e membros.

Report

Richieri-Costa-Pereira syndrome (RCPS [MIM 268305]) is an autosomal recessive syndrome characterized by a midline cleft mandible in addition to Robin sequence, laryngeal abnormalities, and radial and tibial deficiencies associated with clubfeet[1] (**Figure 1**). Learning and language disabilities have been reported in more than 50% of the affected individuals[1]. All but one of the RCPS families described to date are from Brazil, suggesting a founder effect[1]. Previous attempts to identify the genetic cause of RCPS were unsuccessful[2], possibly because the causative mutation is ancient and the shared flanking region has been reduced by multiple recombinations. We performed homozygosity mapping analysis with the Affymetrix 50K Xba1 SNP array in seven individuals belonging to four consanguineous families (**Table S1**, individuals 1, 5, 6, 7, 9, 12, 19 and **Figure S1**) and one unaffected sibling for each of the probands 6, 9 and 19[1]. Although these families came from the same geographic region, they have no knowledge of shared ancestry. Approval for research on human subjects was obtained from the ethics committee of the HRAC – USP-Bauru. All samples were collected after individuals or their relatives had provided informed consent. The region of homozygosity was selected with the Homozygosity Mapper online program assuming a rare-recessive model of inheritance. The only extended region of homozygosity unique to affected individuals was observed at 17q25.3 (**Figure S1**). This region was further genotyped with nine microsatellite markers and 10 SNPs (**Table S2**) in 20 affected individuals from 17 apparently unrelated families and 46 unaffected relatives (31 parents and 15 unaffected siblings). LOD scores were estimated by the Merlin[3] software under a parametric model including data from 8 families (families of the probands 2, 3, 4, 5, 7, 9, 10, and 13). Segregation analysis with the above mentioned markers confirmed linkage to 17q25.3 (**Figure S2;** Maximum LOD score: 9.533 at θ = 0.0 for marker rs2289534). Additional recombinants at rs2289534 (patients 8 and 17) and rs3829612 (patient 11) (**Figure S1b**) refined the disease locus to a 122 kb region

(GRCh37/hg19 – chr17: 78,039,369 - 78,161,152), containing the genes *CCDC40* (MIM 613799)*, GAA* (MIM 606800)*, EIF4A3* (MIM 608546) and *CARD14* (MIM 607211). Mutations in *GAA* cause glycogen storage disease type II (GSD2 [MIM 232300]), and hence *GAA* was excluded as candidate for RCPS. Most recently, heterozygous mutations in *CARD14* have been associated with psoriasis (PSOR2 [MIM 602723])[4] and bi-allelic *CCDC40* loss-of-function mutations were associated with ciliary dyskinesia, primary 15 (CIDL5 [MIM 613808])[5]. Since there is no clinical overlap between the phenotypes in RCPS and those above syndromes, *EIF4A3* is the most likely candidate for RCPS.

Except for exon 1 of *EIF4A3* which proved difficult to sequence, Sanger sequencing of the remaining exons of *EIF4A3,* and all exons of *CCDC40, CARD14* and their exon-intron boundaries did not reveal any non-conservative coding substitutions or obvious splice site disruptions (**Table S3**). Whole-exome sequencing (WES) was performed in affected individual 17 according to previously standardized protocols[6]. WES revealed only 7 homozygous non-synonymous changes not described in the available SNP database at that time [NCBI dbSNP (132)]; however, none of them mapped to 17q25.3.

By re-examining the WES data for the candidate region, we found that exon 1 of *EIF4A3* had low coverage (mean coverage = 3.6) compared to the other exons (mean coverage of the remaining exons = 17.2); in addition, the *EIF4A3* 5´ UTR is GC-rich, which could explain our technical difficulties in sequencing this region. PCR amplification of exon 1 showed a larger homozygous allele only among affected individuals (**Figure S3**). This larger 5´ UTR allele was not observed in any of the 520 Brazilian control individuals, further confirming its association with the disease. To better characterize the 5´ UTR of *EIF4A3*, we sequenced 140 control alleles from 70 unrelated Brazilian individuals and discovered multiple allelic patterns, which varied in size and organization of motifs containing 18- or 20-nucleotides (nt). These motifs could be divided in three types: I) 20-nt motif, TCGGCAGCGGCACAGCGAGG, termed

‘**CACA-20-nt’;** II) 18-nt motif, TCGGCAGCGGCAGCGAGG, termed ‘**CA-18-nt’**; III) another 20-nt motif which possessed a G instead of an A, TCGGCAGCGGC**G**CAGCGAGG, termed ‘**CGCA-20-nt**’. The most prevalent (97%) allelic pattern among controls was characterized by an initial CACA-20-nt repeated between 2 and 9 times, followed by one CA-18-nt, another CACA-20-nt, and one final CA-18-nt (total repeats = 5 to 12), ending 43 bases upstream of the first ATG (**Figure 2**). In turn, affected individuals exhibited the following pattern: an initial CACA-20-nt, followed by 12 to 13 repeats of CGCA-20-nt, one CACA-20-nt and one final CA-18-nt (total number of repeats = 15 or 16). To maintain clarity, the identified alleles are hereafter referred by the absolute number of repeats, which varied from 3 to 12 in controls and were either 15 or 16 in these RCPS individuals. We observed that 17 affected probands were homozygous for the 16-repeat allele, and 3 apparently unrelated affected individuals (6, 14 and 18; **Figure S2**) were compound heterozygotes (15/16 repeats). All tested parents were heterozygous for the 16-repeat allele, and unaffected siblings either lacked the expanded allele or were heterozygotes (**Figure S2**). Accordingly, the expanded alleles segregated perfectly with the disease, following an autosomal recessive model.

Comparative analysis of the 5´ UTR of *EIF4A3* mammalian orthologs showed that only anthropoid primates, which present mandible fusion as one of their morphological autapomorphies[7], share a repetitive sequence highly similar to the human CA-18-nt motif (**Figure S4**). It is thus possible that the most ancient allele in humans might have had one or few CA-18-nt motifs and the 20-nt motif has arisen more recently by a CA, or more rarely, by a CG insertion. We speculate that an increased instability of the region is possibly being driven by the CGCA-20-nt motif; however, current data are not sufficient to predict the mechanisms responsible for the appearance of the 16-repeat allele linked to the disorder. Even though the 16-repeat allele seems to be relatively stable through meiosis and with a unique pattern of organization, we classified this mutation as an expansion and added RCPS to the

growing group of disorders caused by non-coding repeat expansions, which includes Friedreich's Ataxia[8] (FRDA [229300]), myotonic dystrophies[9] (DM1 [MIM 60900]; DM2 [MIM 602668]) fragile X-associated tremor/ataxia syndromes[10] (Fragile X syndrome [MIM 300624], FXTAS [MIM 300623], several spinocerebellar ataxias[11-14], frontotemporal dementia and amyotrophic lateral sclerosis[15] (FTDALS [MIM 105550]) (reviewed in **Table S4**).

EIF4A3 encodes a DEAD box helicase (eIF4AIII), the core protein of the exon junction complex (EJC). It co-ordinates the control of downstream processes of mRNA splicing and nonsense-mediated mRNA decay (NMD), and is also involved in rRNA biogenesis[16-20]. eIF4AIII interacts directly with mRNA and forms the minimally stable core of the EJC by interacting with Y14, MAGOH, UPF3, MLN51 and RNPS1[21, 22].

Sequencing of *EIF4A3* in five additional Brazilian affected individuals ascertained elsewhere[23, 24] revealed that four of them (individuals 21 to 24, including 2 siblings) were homozygous for the 16-repeat allele. In contrast, the fifth (individual 25) was a compound heterozygote, possessing a 14-repeat allele in *trans* with a nucleotide change, c.809A>G (*EIF4A3* transcript, NM_014740.3), which leads to an amino acid substitution at Asp270 (p.Asp270Gly) (**Figure 3**). The 14-repeat allele showed a distinct pattern of motifs to that of the typical 16-repeat allele (**Figure 2**). These two mutations were not identified in any database nor in our control sample (n = 520/expansion allele; n = 285/substitution mutation). Strong functional constraint of Asp270 was suggested by its location in the C-terminal helicase RecA2 domain and by its complete conservation not only in eIF4AIII orthologs from 7 divergent vertebrate species, plant and yeast but also in eIF4A paralogs (**Figure 3**). SIFT and *PolyPhen* analyses predicted that Gly at position p.270 perturbs the structure and function of eIF4AIII (**Figure 3**). The side chain of aspartate 270 forms a hydrogen bond with Tyr429 in UPF3B (**Figure 3**), an EJC protein involved in NMD[25-27] and mutation of this residue also affects eIF4AIII recruitment by CWC22 to the spliceosome[28]. Furthermore glycine strongly destabilizes helices compared to aspartate[29], so the helical structure

downstream of the mutated residue may also be affected (**Figure 3C**). Together, these observations suggest that p.Asp270Gly is likely pathogenic and RCPS is caused by different mutational mechanisms: expansion of the 5´ UTR (14 to 16 repeats) or missense mutation.

The 15- and 16-repeat allele haplotypes spanning *EIF4A3* are consistent with a common origin (**Figures S2, S5**) and corroborate our previous founder effect hypothesis for most Brazilian RCPS individuals. The 15-repeat alleles might represent a retraction of the 16-repeat allele or alternatively might have arisen by unequal crossing over (Figure S2), as suggested in polyalanine expansion disorders[30]. The 14-repeat allele is embedded in the same 42 kb haplotype observed in the recombinant 16-repeat alleles (Figure S5); however, its origin remains unclear given its distinct motif organization and structure (**Figure 2**). In contrast, the c.809A>G mutation is embedded in a distinct haplotype (**Figures S2, S5**), suggesting multiple pathogenic mutational origins in *EIF4A3*.

The phenotype of RCPS individuals varies in expressivity even within families. The main clinical craniofacial characteristic fully penetrant in all 16-repeat allele homozygous or 15/16-repeat allele compound heterozygous individuals is midline mandibular involvement, ranging from abnormal fusion observed radiologically to complete lack of fusion leading to a wide gap at the mandibular symphysis (**Figure 1** and **Table S1**)[1]. The three affected compound heterozygous individuals for the 15/16-repeat alleles presented a very similar phenotype to the homozygous 16/16 affected individuals (**Table S1**). In contrast, individual 25 who is a compound heterozygote for a 14-repeat allele and c.809A>G (p.Asp270Gly) was the only individual with mandible fusion (**Table S1**; **Figure 1).** The milder phenotype in individual 25 might be due to a weaker effect of the amino acid substitution mutation or of the 14-repeat allele in comparison to the 16-repeat allele. Atypical or milder phenotypes have been associated with substitution mutations in Friedreich's Ataxia[31], an expansion disorder with an autosomal recessive inheritance pattern (**Table S4**).

To further understand the functional effect of the *EIF4A3* expanded alleles, we first investigated *EIF4A3* transcripts. Through cDNA sequencing analysis, we showed that the 5´ UTR repeats are included in the *EIF4A3* transcripts from white blood cells from both controls and affected individuals (data not shown). We did not observe any evidence of *EIF4A3* alternative splicing either in controls or affected individuals (**Figure S6**). Next, we performed real time quantitative PCR (RT-qPCR) for *EIF4A3* with mRNA obtained from white blood cells (4 affected and 9 control individuals) and from mesenchymal cells (2 affected and 11 control individuals), using standardized methods[32,33]. *EIF4A3* transcript abundance was about 30-40% lower in affected individuals than in controls in both cell types tested (white blood cells, $P = 0.0193$; mesenchymal cells, $P = 0.001$; **Figure S6)**. It is likely that the expanded allele does not alter the splicing of *EIF4A3* mRNA but it seems to reduce its abundance in the RCPS cell types investigated. Further studies, including quantification of expressed protein in different cell types, will be necessary to confirm whether this expansion represents a partial loss-of-function mutation. We also cannot exclude the possibility that this expanded allele leads to RNA toxicity affecting other proteins, as has been shown for full expansions at *DMPK* (MIM 605377) and at *CNBP1* (MIM 116955)[34]. The substitution p.Asp270Gly is likely to weaken the interaction between eIF4AIII and UPF3B, thus resulting in a less efficient NMD and transcriptional or translation regulation. It is thus possible that the mechanism behind the RCPS phenotype might be a partial loss-of-function of *EIF4A3* (hypomorphic mutation), which would be the expected model considering the autosomal recessive inheritance pattern of the disease.

To further investigate the role of *EIF4A3* in craniofacial development as well as its deficiency as a putative mechanism for RCPS, we modeled *eif4a3* deficiency in zebrafish embryos by using specific morpholinos (MO). Three different MOs along with their corresponding mispaired controls were purchased from Gene Tools, LLC (USA) (MO sequences in **Figures 4, S7 and S8**). MOs were designed to block either *eif4a3*

mRNA translation (TRA1-MO and TRA2-MO) or *eif4a3*-pre-mRNA splicing (Spl-MO). Zebrafish embryos at 1-4 cell stage were injected with 5 nl of each MO at the indicated concentration (ranging from 0.01 to 0.1 mM) depending on the MO, and development was allowed to proceed[35,36]. As a slight delay in morphant development was observed, developing fish were staged taking into account the presence of typical developmental structures. Morphants with consistent and reproducible craniofacial phenotypes were scored as a unique category (**Table S5 and Figures 4, S7 and S8**). Similar craniofacial phenotypes were observed regardless of the nature of the injected MO. Morphants staged at 24 hours post fertilization (hpf) showed eyes reduced in size and dark and opaque zones in all brain structures. In addition, the otic vesicle and the midbrain/hindbrain border regions were barely detectable. Anomalies in the trunk and tail were observed in some embryos, however these anomalies were not further characterized because they were variable and inconsistent among different MOs. (**Figures 4, S7 and S8**). Acridine orange staining[37] performed on 24-hpf staged morphant and control embryos suggested extensive apoptosis in morphants in comparison to controls, which was even more intense in the anterior-most regions of morphants (**Figures 4, S7 and S8**). Cartilage and bone staining revealed underdevelopment of craniofacial cartilage, bone alterations and clefting of the lower jaw (**Figures 5, S7 and S8**). Furthermore, the 3rd through 6th pharyngeal arches were underdeveloped in morphants. Therefore, morphant fish displayed multiple defects in craniofacial structures analogous to those affected in RCPS individuals. Importantly, most of the abnormalities were rescued when an *in vitro* synthesized-mRNA encoding for the zebrafish eIF4AIII translationally fused to EGFP was co-injected with each of the three MOs tested (**Table S5 and Figures 4, 5, S7, S8 and S9**). The TRA1-MO and TRA2-MO anneal on two different regions of the 5′ UTR and, thus, do not anneal to the injected mRNA, ruling out MO off-targeting effects. As craniofacial cartilage and bones mainly derive from the cranial neural crest (CNC), we assessed the expression of typical CNC marker genes in treated and control 24-hpf staged embryos. RT-qPCR

revealed that *eif4a3* knockdown adversely affects the transcription of typical neural crest gene markers, such as *sox9b*, *foxd3*, *sox10*, and *tbx2* (**Figure S10**). Therefore, *eif4a3* depletion might result in failure of the EJC assembly, which ultimately would lead to CNC cell death and underdevelopment of the pharyngeal arches.

In summary, our findings suggest that *EIF4A3* deficiency leads to abnormal development of most pharyngeal arches, resulting in altered mandible and laryngeal morphogenesis. Although not explored in this manuscript, deficiency of *EIF4A3* also interferes in limb development and is associated with learning and language disabilities observed in a high proportion of RCPS patients. The altered neurodevelopmental phenotype in RCPS, which should be further studied, is unsurprising given the role of eiF4AIII in regulating transcription abundance of neuronal effector genes that underlie learning and memory processes[38, 39]. In addition, heterozygous deletions of *EIF4A3* associated with intellectual disability and autism have recently been reported[40]. RCPS, belongs to the growing list of craniofacial syndromes caused by loss-of-function mutations of genes encoding proteins involved in RNA metabolism and ribosome biosynthesis, such as *TCOF1* (MIM 606847; Treacher Collins syndrome, TCS1 [MIM 154500])[41], *SF3B4* (MIM 605593; Nager syndrome, AFD1 [MIM 154400])[42] and *EFTUD2* (MIM 603892; Mandibular dysostosis and microcephaly, MFDM [MIM 610536])[43]. Apoptosis of CNC cells might be a common mechanism underlying all these syndromes[44, 45]. However, considering the complexity of RNA metabolism, it is possible that dysregulation of distinct pathways might explain their specific phenotypes. Of these other syndromes, only *EFTUD2*-mutated individuals present cognitive impairment; notably, the U5-116-kD spliceosomal GTPase protein encoded by *EFTUD2* directly interacts with eIF4AIII[45]. Further studies will be necessary to investigate the functional relationship between *EIF4A3* and the above mentioned genes and to understand how mutations in *EIF4A3* lead to the pleiotropic phenotype of RCPS.

Web Resources

International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/

National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/;

NCBI dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/;

PolyPhen-2 prediction of functional effects of human nsSNPs,

http://genetics.bwh.harvard.edu/pph2/

Sequence accession numbers: Human *EIF4A3* cDNA (ENST00000269349) c.1 > c.295.

SIFT (Sorting Intolerant from Tolerant), http://sift.bii.a-star.edu.sg/

Acknowledgments

We are thankful to Alessandra Splendore and Gerson S. Kobayashi for critical discussion on the manuscript, to Mrs. Regina Bueno for art support and to the Human Genome Research Center sequencing service for technical support. This work was funded by CEPID/FAPESP, CNPq and INCT/CNPq-FAPESP and by the Wellcome Trust (093329 to A.O.M.W. and S.R.F.T.). The authors declare no conflict of interest.

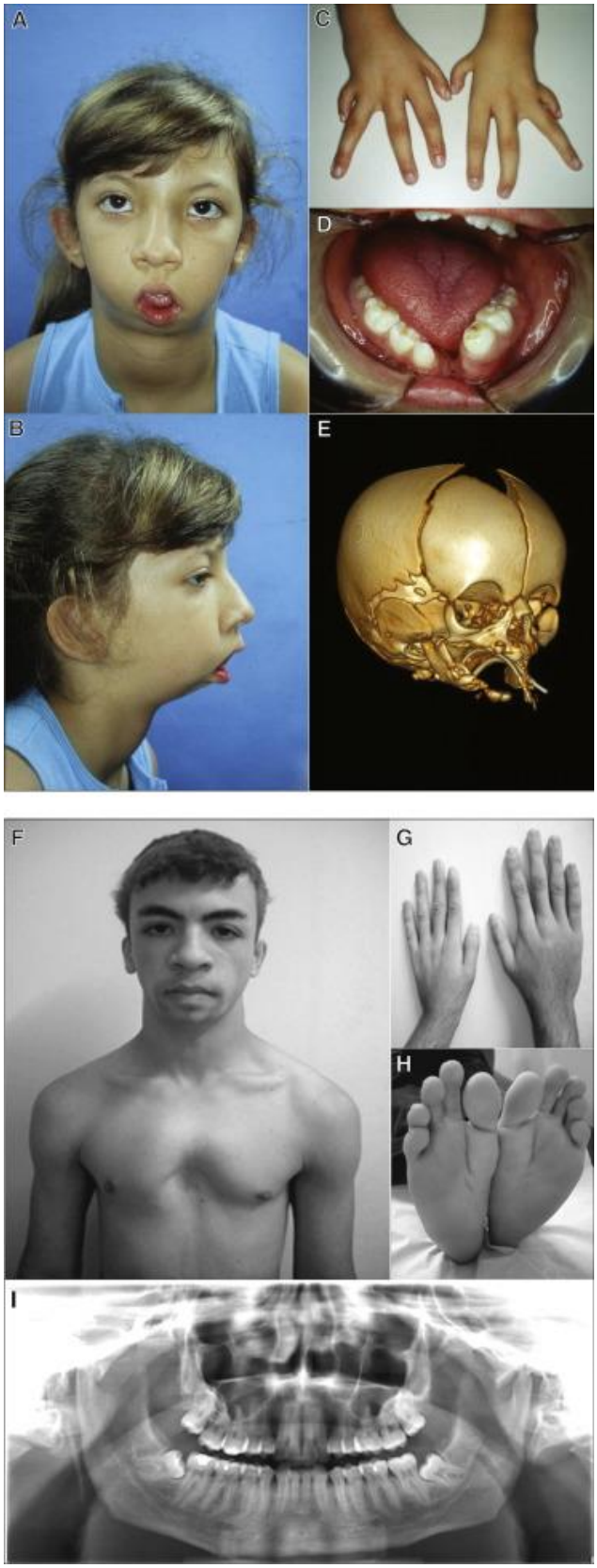


Figure 1. Clinical and radiographic aspects of typical (probands 2 and 23) and a milder form (proband 25) of RCPS. **(a-d)** Affected individual 2 at 9 years of age, illustrating **(a,**

b) typical facial features of the syndrome including micrognathia and microstomia, **(c)** hypoplasia of fingers and clinodactyly and **(d)** absence of lower central incisors and median mandibular cleft**; (e)** CT scan of the skull of affected individual 23 at 12 days of age. Note the very rudimentary mandibular formation with large medial cleft, micrognathia, incomplete zygomatic arches. **(f-i)** Affected individual 25, illustrating **(f)** absence of microstomia and presence of pectus excavatum, (**g**) short left hand with hypoplastic thumbs, (**h**) feet with characteristically abnormal shape, (**i**) normal fusion of mandible.

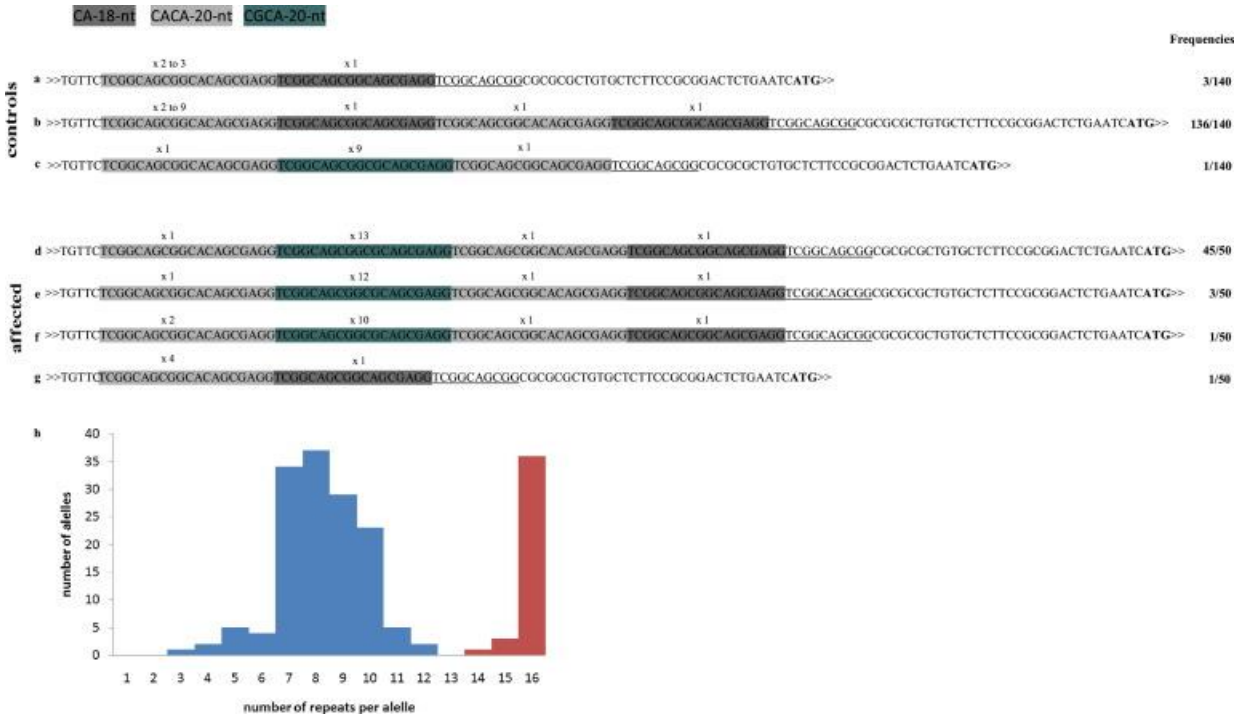


Figure 2. Representation of the pattern of the repeats at the 5´ UTR of *EIF4A3,* and their distribution frequency. **(a, b, c)** Control alleles: note that the number of CA-18-nt motifs varied among the different allele patterns and that the CGCA-20-nt motif was present in only one out of 140 alleles. **(d, e, f)** Affected alleles: the 15- and 16- repeat alleles have the same underlying structure, whereas the 14-repeat allele contains an additional copy of the CACA-20-nt motif; **(g)** the non expanded 5´ UTR in *cis* with the nucleotide substitution c.809A>G (p.Asp270Gly) allele in individual 25 corresponds to one of the patterns of the control individuals. The first “Thymidine (T)” at the 5´ end corresponds to position +38 of the transcription initiation site while the **ATG** corresponds to the translation initiation codon. Underlined sequence downstream of the last 18-nt motif represents a partially conserved motif in all alleles. **(h)** Distribution of the alleles per repeat size. Control sample alleles are showed in blue and affected alleles showed in red. The counting of the alleles in affected individuals took into consideration relatedness

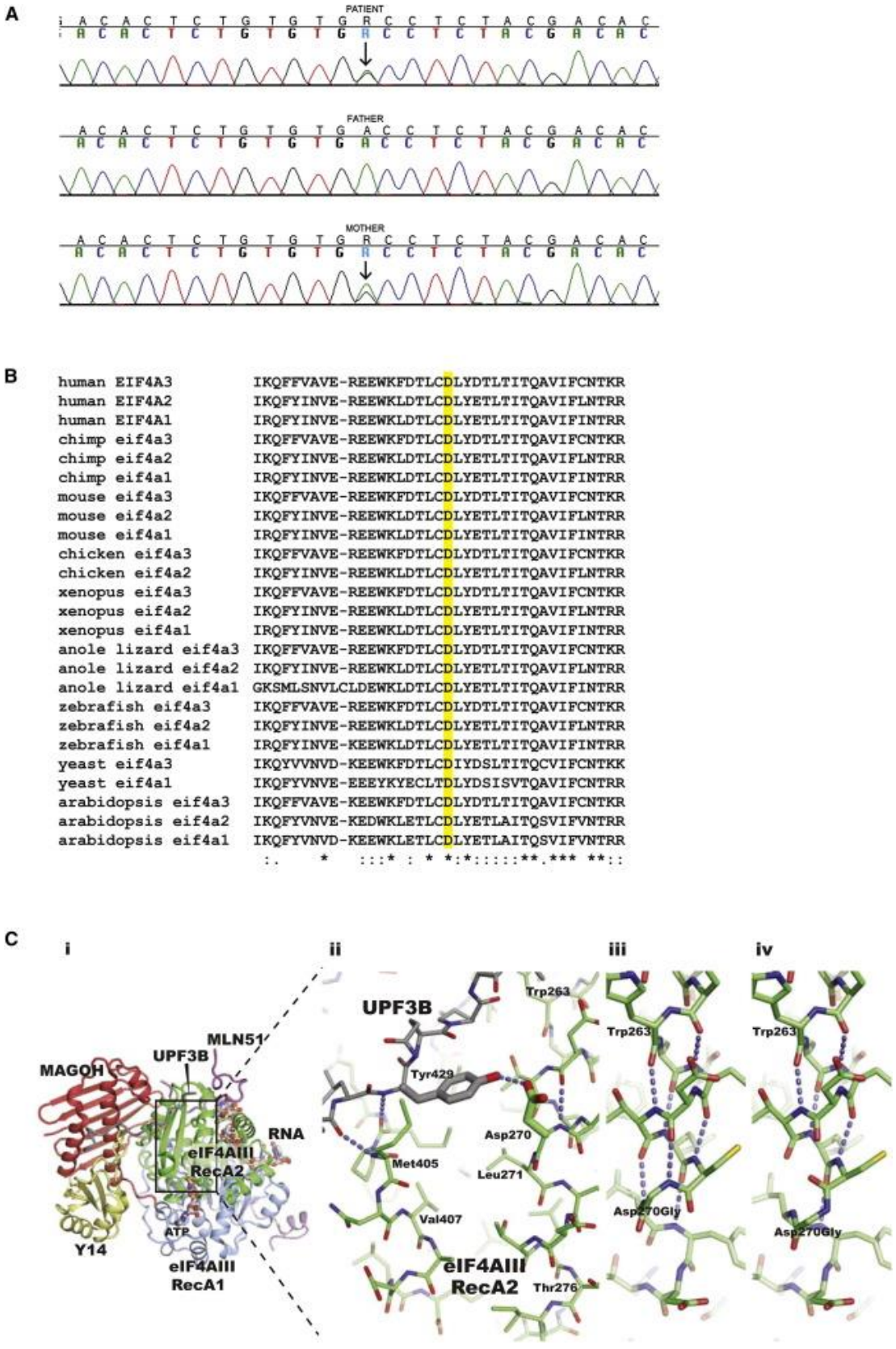


Figure 3. Description of the missense mutation in *EIF4A3*, evolutionary comparative analysis and structural analysis. **(a)** Sequence analysis of exon 8 showing the A to G substitution at position c.809 (indicated by an arrow), leading to the amino acid

substitution p.Asp270Gly. The online tool *PolyPhen-2 prediction of functional effects of human nsSNPs* predicts this mutation as possibly damaging with a score of 0.860 (sensitivity: 0.72 and specificity: 0.89). **(b)** Comparative sequence analysis of *EIF4A3* encoded orthologous proteins and its paralogs, showing that the Asp 270 (D) is highly conserved throughout evolution. **(c)** Structure of the core human exon junction complex bound to a C-terminal fragment of UPF3B (RCSB entry 2XB2). (**i)** Overall view, the RNA- and ATP binding eIF4A3 is shown with its RecA1 and RecA2 domains colored blue and green, respectively. A fragment of MLN51 (magenta) encloses both RecA domains and also forms part of the binding site for the 5' end of the RNA. MAGOH (red) contacts both domains of eIF4AIII, the MLN51 fragments and positions Y14 (yellow) in the complex. The C-terminal UPF3B fragment (grey) interacts with both Y14 and the eIF4AIII RecA2 domain. (**ii)** Close-up on the UPF3B Tyr429 interaction with the Asp270 in the eIF4A3 RecA2 domain. Selected putative hydrogen bonds of relevance for the Asp-Glycine mutation are indicated with blue dashed lines. **iii**) Selected model of the Glycine mutant with a secondary structure similar to that of the experimental structures of wildtype eIF4AIII. (**iv**) Selected model of the Glycine mutant with a disrupted secondary structure compared to the experimental structures of wildtype eIF4AIII. Models were prepared with Modeller[29].

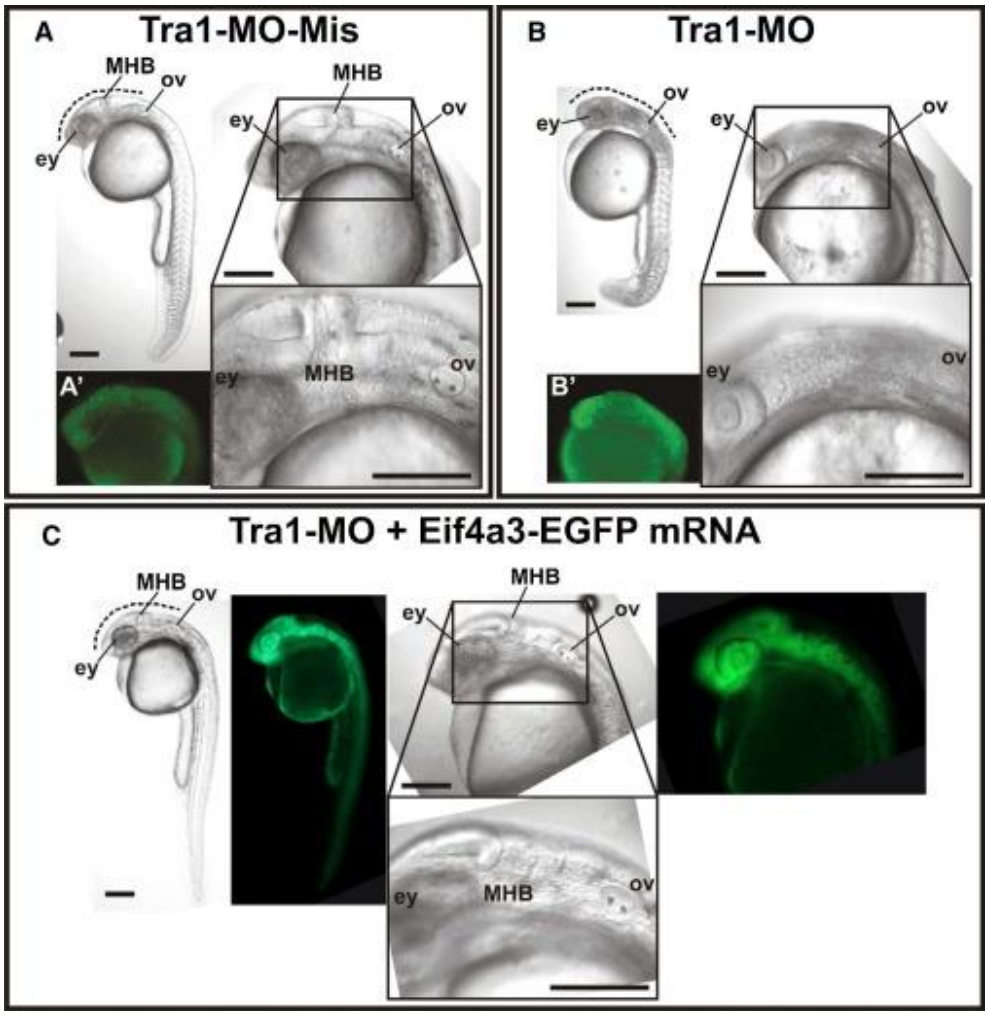


Figure 4. Zebrafish *eif4a3* knockdown using morpholinos (MOs). Microinjected zebrafish embryos staged at 24 hours post-fertilization (hpf) presented an observable phenotype. **(a)** Control embryo microinjected with mispaired MO (TRA1-MO-Mis). **(b)** Morphant embryo microinjected with MO designed to block *eif4a3*-mRNA translation (TRA1-MO). **(c)** Rescued embryo microinjected with TRA1-MO and mRNA coding for *eif4a3* fused to EGFP (*eif4a3*-EGFP mRNA). Lateral views of embryos were registered under stereoscopic microscope (whole body at the upper left of each panel) or with two different magnifications under differential interference contrast microscope (anterior-most region at the upper and lower right of each panel). Alterations in the morphants' trunk were observed but not reproducible and only phenotyping at the craniofacial level was done. Dotted lines mark regions between eyes (ey) and otic vesicle (ov), which include the midbrain-hindbrain border (MHB) and showed darkened tissue in

morphants, suggesting the presence of apoptotic cells. **(A')** and **(B')** show lateral views of 24-hpf embryos after acridine orange staining, using standardized protocols[35]. Green fluorescence panels in **(c)** show expression of eif4a3-EGFP in rescued embryo. Scale bars represent 200 μm. TRA1-MO sequence: TGTGACGGATTTCGGTGTAAATTAC. TRA1-MO-Mis sequence: TGTCACCGATTTCCGTCTAAAATAC.

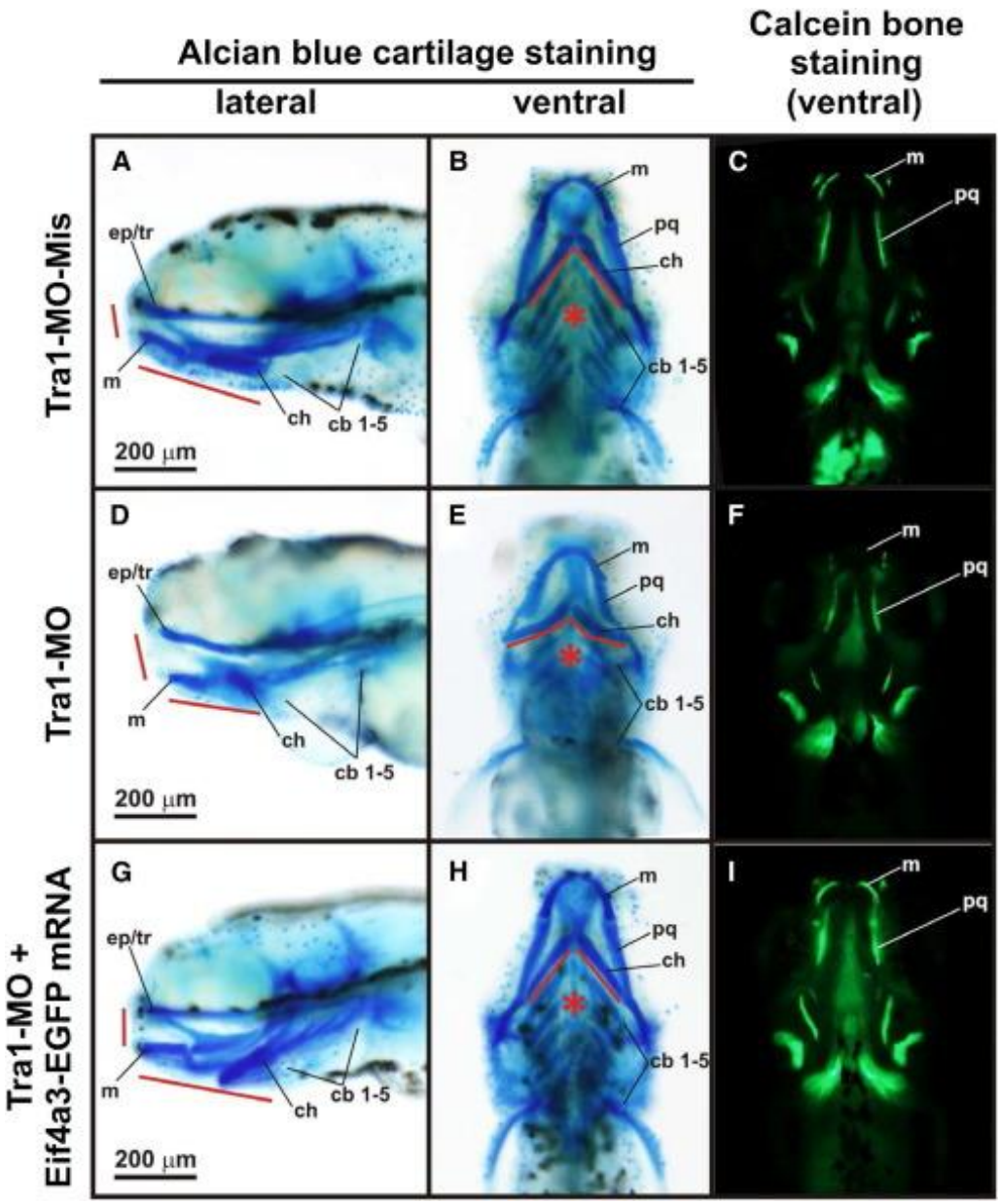


Figure 5. Craniofacial phenotype in zebrafish *eif4a3* morphant larvae. Five days post-fertilization (dpf) larvae microinjected with TRA1-MO-Mis **(a-c)**, TRA1-MO **(d-f),** and TRA1-MO + eif4a3-EGFP mRNA **(g-i)** were stained with alcian blue[36] to observe cartilage structures (**a, d, g,** lateral views; **b, e, h,** ventral views) or with calcein to observe bone structures (**c, f, i,** ventral views). Eyes were removed to register alcian blue stained larvae. Craniofacial precursors in morphant larvae were severely affected showing hypoplasia of numerous craniofacial cartilages (red lines), which is evident in the rostral and jaw elements. Impairment of pharyngeal arches development was also observed (asterisks). cb1–5 = ceratobranchial arches 1–5; ch = ceratohyal; ep/tr =

ethmoid plate/trabecula; m = Meckel's cartilage; pq = palatoquadrate. Scale bars represent 200 μm.

References

1. Favaro, F.P., Zechi-Ceide, R.M., Alvarez, C.W., Maximino, L.P., Antunes, L.F.B.B., Richieri-Costa, A., and Guion-Almeida ML. (2011). Richieri-Costa-Pereira syndrome: A unique acrofacial dysostosis type. An overview of the Brazilian cases. Am J Med Genet. *155*, 322-331.

2. Lima, R.L.L.F., Moretti-Ferreira, D., Richieri-Costa, A., and Murray, J.C. (2003). Identity by descent and candidate gene mapping of Richieri-Costa and Pereira syndrome. Am J Med Genet. *122*, 56-58.

3. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. *30*, 97-101.

4. Jordan, C.T., Cao, L., Roberson, E.D., Duan, S., Helms, C.A., Nair, R.P., Duffin, K.C., Stuart, P.E., Goldgar, D., Hayashi, G., et al. (2012) Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappaB, in psoriasis. Am J Hum Genet. 90, 796-808.

5. Becker-Heck, A., Zohn, I.E., Okabe, N., Pollock, A., Lenhart, K.B., Sullivan-Brown, J., McSheene, J., Loges, N.T., Olbrich, H., Haeffner, K., et al. (2011). The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis. Nat Genet. 43:79-84.

6. Twigg, S.R., Lloyd, D., Jenkins, D., Elçioglu, N.E., Cooper, C.D., Al-Sannaa, N., Annagür, A., Gillessen-Kaesbach, G., Hüning, I., Knight, S.J., et al (2012). Mutations in

multidomain protein MEGF8 identify a Carpenter syndrome subtype associated with defective lateralization. Am J Hum Genet. *91*, 897-905.

7. Scott, J.E., Lack, J.B., and Ravosa, M.J. (2012). On the reversibility of mandibular symphyseal fusion. Evolution. *66*, 2940-2952.

8. Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science. *271*, 1423–1427.

9. Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T., et al. (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell. *68*, 799-808.

10. Galloway, J.N., and Nelson, D.L. (2009). Evidence for RNA-mediated toxicity in the fragile X associated tremor/ataxia syndrome. Future Neurol. *4*, 785.

11. Daughters, R.S., Tuttle, D.L., Gao, W., Ikeda, Y., Moseley, M.L., Ebner, T.J., Swanson, M.S., and Ranum, L.P. (2009). RNA gain-of-function in spinocerebellar ataxia type 8. PLoS Genet. e1000600.

12. Kobayashi, H., Abe, K., Matsuura, T., Ikeda, Y., Hitomi, T., Akechi, Y., Habu, T., Liu, W., Okuda, H., and Koizumi, A. (2011). Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. Am J Hum Genet. *89*, 121-130.

13. Moseley, M.L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A.K., Daughters, R.S., Chen, G., Weatherspoon, M.R., Clark, H.B., Ebner, T.J., et al. (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nat Genet. *38*, 758-769.

14. Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., Takahashi, M., Matsuura, T., Flanigan, K.M., Iwasaki, S., et al. (2009). Spinocerebellar ataxia type 31 is associated with “inserted” pentanucleotide repeats containing (TGGAA)n. Am J Hum Genet. *85*, 544-557.

15. DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxe, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. *72*, 245-256.

16. Chan, C.C., Dostie, J., Diem, M.D., Feng, W., Mann, M., Rappsilber, J., and Dreyfuss, G. (2004). eIF4A3 is a novel component of the exon junction complex. RNA. *10*, 200-209.

17. Shibuya, T., Tange, T.O., Stroupe, M.E., and Moore, M.J. (2006). Mutational analysis of human eIF4AIII identifies regions necessary for exon junction complex formation and nonsense-mediated mRNA decay. RNA. *12,* 360-374.

18. Alexandrov, A., Colognori, D. and Steitz, J.A. (2011). Human eIF4AIII interacts with an eIF4G-like partner, NOM1, revealing an evolutionarily conserved function outside the exon junction complex. Genes Dev. *25*, 1078-1090.

19. Haremaki, T., Sridharan, J., Dvora, S. and Weinstein, D.C. (2010). Regulation of

vertebrate embryogenesis by the exon junction complex core component Eif4a3. Dev Dyn. *239*, 1977-1987.

20. Kunz, J.B., Neu-Yilik, G., Hentze, M.W., Kulozik, A.E., and Gehring, N.H. (2006). Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. RNA. *12*, 1015-1022.

21. Zhang, Z., and Krainer, A.R. (2007). Splicing remodels messenger ribonucleoprotein architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. Proc Natl Acad Sci. *104*, 11574–11579.

22. Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012) The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus. Cel. *151*, 750-764.

23. Souza, J., dalVesco, K., Tonocchi, R., Closs-Ono, M.C., Passos-Bueno, M.R., and Silva-Freitas, R. (2011). The Richieri-Costa and Pereira Syndrome: Report of Two Brazilian Siblings and Review of Literature. Am J Med Genet. *155A*, 1173-1177.

24. Raskin, S., Souza, M.B., Medeiros, M.C., Manfron, M., and Chong-Silva, D.C. (2013). Richieri-Costa and Pereira syndrome: Case Report of a severe phenotype. Am J Med Genet A.doi: 10.1002/ajmg.a.35989 [Epub ahead of print].

25. Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Séraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science. *313*, 1968-1972.

26. Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell. *126*, 713-725.

27. Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F., Le Hir, H., and Conti, E. (2010). Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3B complex. Proc Natl Acad Sci. *107*, 10050-10055.

28. Steckelberg, A.L., Boehm, V., Gromadzka, A.M., and Gehring, N.H. (2012). CWC22 connects pre-mRNA splicing and exon junction complex assembly. Cell Rep. *2*, 454-461.

29. Fiser, A., Do, R.K., and Sali, A. (2000). Modeling of loops in protein structures. Protein Sci. *9*, 1753-1773.

30. Warren, S.T. (1997). Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. Science. *275*, 408-409.

31. Martelli, A., Napierala, M., and Puccio, H. (2012). Understanding the genetic and molecular pathogenesis of Friedreich's ataxia through animal and cellular models. Dis Model Mech. *5,* 165-176.

32. Kobayashi, G.S., Alvizi, L., Sunaga, D.Y., Francis-West, P., Kuta, A., Almada, B.V., Ferreira, S.G., de Andrade-Lima, L.C., Bueno, D.F., Raposo-Amaral, C.E., Menck, C.F., Passos-Bueno, M.R. (2013). Susceptibility to DNA damage as a molecular mechanism for non-syndromic cleft lip and palate. Plos One 8 (6): e65677.

33. Masotti, C., Ornelas, C.C., Splendore-Gordonos, A., Moura, R., Félix, T.M., Alonso, N., Camargo, A.A., and Passos-Bueno, M.R. (2009). Reduced transcription of TCOF1 in adult cells of Treacher Collins syndrome patients. BMC Med Genet.*10*,136.

34. Udd, B., Krahe, R. (2012). The myotonic dystrophies: molecular, clinical, and therapeutic challenges. Lancet Neurol 11:891-895.

35. Eimon, P. and Ashkenazi, A. (2010) The zebrafish as a model organism for the study of apoptosis. Apoptosis 15 (3):331-349.

36. Nasevicius, A., and Ekker, S.C. (2000). Effective targeted gene 'knockdown' in zebrafish. Nat. Genet. *26*, 216-220.

37. Solomon, K.S.;Kudoh, T.;Dawid, I.B. and Fritz, A. (2003) Zebrafish foxi1 mediates otic placode formation and jaw development. Development. 130(5):929-940.

38. Giorgi, C., Yeo, G.W., Stone, M.E., Katz, D.B., Burge, C., Turrigiano, G., and Moore, M.J. (2007). The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. Cell. *130,* 179-191.

39. Barker-Haliski, M.L., Pastuzyn, E.D., and Keefe, K.A. (2012). Expression of the core exon-junction complex factor eukaryotic initiation factor 4A3 is increased during spatial exploration and striatally-mediated learning. Neuroscience. *226*, 51-61.

40. Nguyen, L.S., Kim, H.G., Rosenfeld, J.A., Shen, Y., Gusella, J.F., Lacassie, Y., Layman, L.C., Shaffer, L.G., and Gecz, J. (2013). Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders. Hum Mol Genet. Epub ahead of print.

41. Dauwerse, J.G., Dixon, J., Seland, S., Ruivenkamp, C.A., van Haeringen, A., Hoefsloot, L.H., Peters, D.J., Boers, A.C., Daumer-Haas, C., Maiwald, R., et al. (2011). Mutations in genes encoding subunits of RNA polymerases I and III cause Treacher Collins syndrome. Nat. Genet. *43*, 20-22.

42. Bernier, F.P., Caluseriu, O., Ng, S., Schwartzentruber, J., Buckingham, K.J., Innes, A.M., Jabs, E.W., Innis, J.W., Schuette, J.L., Gorski, J.L., et al. (2012). Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal Complex, Causes Nager Syndrome. Am J Hum Genet. *90*, 925-933.

43. Lines, M.A., Huang, L., Schwartzentruber, J., Douglas, S.L., Lynch, D.C., Beaulieu, C., Guion-Almeida, M.L., Zechi-Ceide, R.M., Gener, B., Gillessen-Kaesbach, G., et al. (2012). Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. Am J Hum Genet. *90*, 1-9.

44. Dixon J, Brakebusch C, Fässler R, Dixon MJ. Increased levels of apoptosis in the prefusion neural folds underlie the craniofacial disorder, Treacher Collins syndrome. Hum Mol Genet. 2000 Jun 12;9(10):1473-80.

45. Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell. *151*, 750-764.

46. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: RESEARCH0034.

47. Weiner A.M.J., Scampoli N.L., Calcaterra N.B. (2012) Fishing the Molecular Bases of Treacher Collins Syndrome. PLoS ONE 7(1): e29574. doi:10.1371/journal.pone.0029574

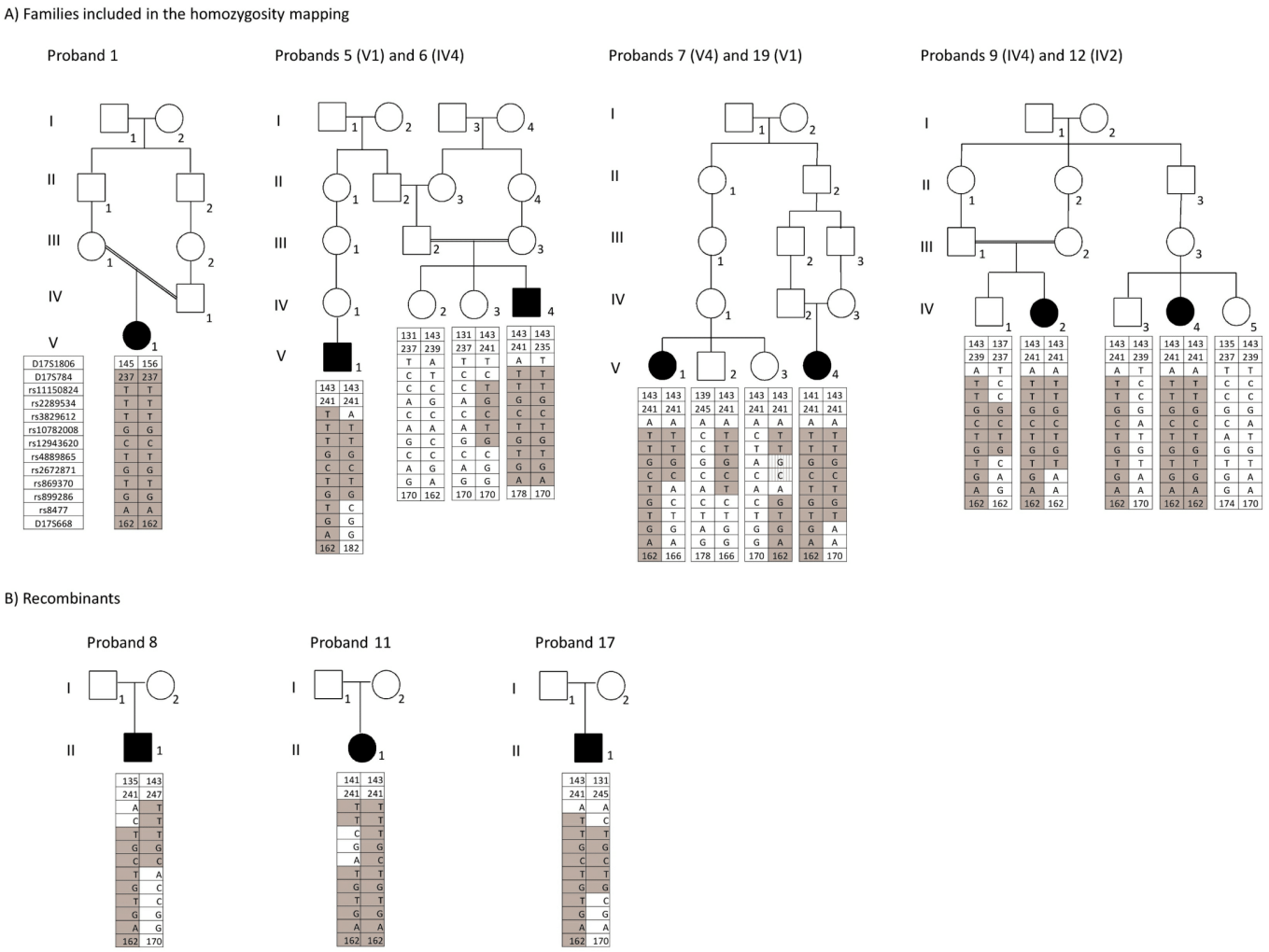


Figure S1: Schematic representation of haplotypes illustrating the homozygosity region and recombinants critical to define the genetic candidate

interval. Only 13 of the 19 polymorphic markers tested are shown. (a) Families used in the original homozygosity mapping showing the at-risk haplotypes (marked in grey). Haplotype of proband 1 is probably the ancestral one based on the analysis of all the probands (data not shown). Lighter grey indicates uninformative region; (b) Recombinants (probands 8, 11 and 17) that allowed the definition of the candidate region to 122 kb, between markers rs2289534 and rs3829612.

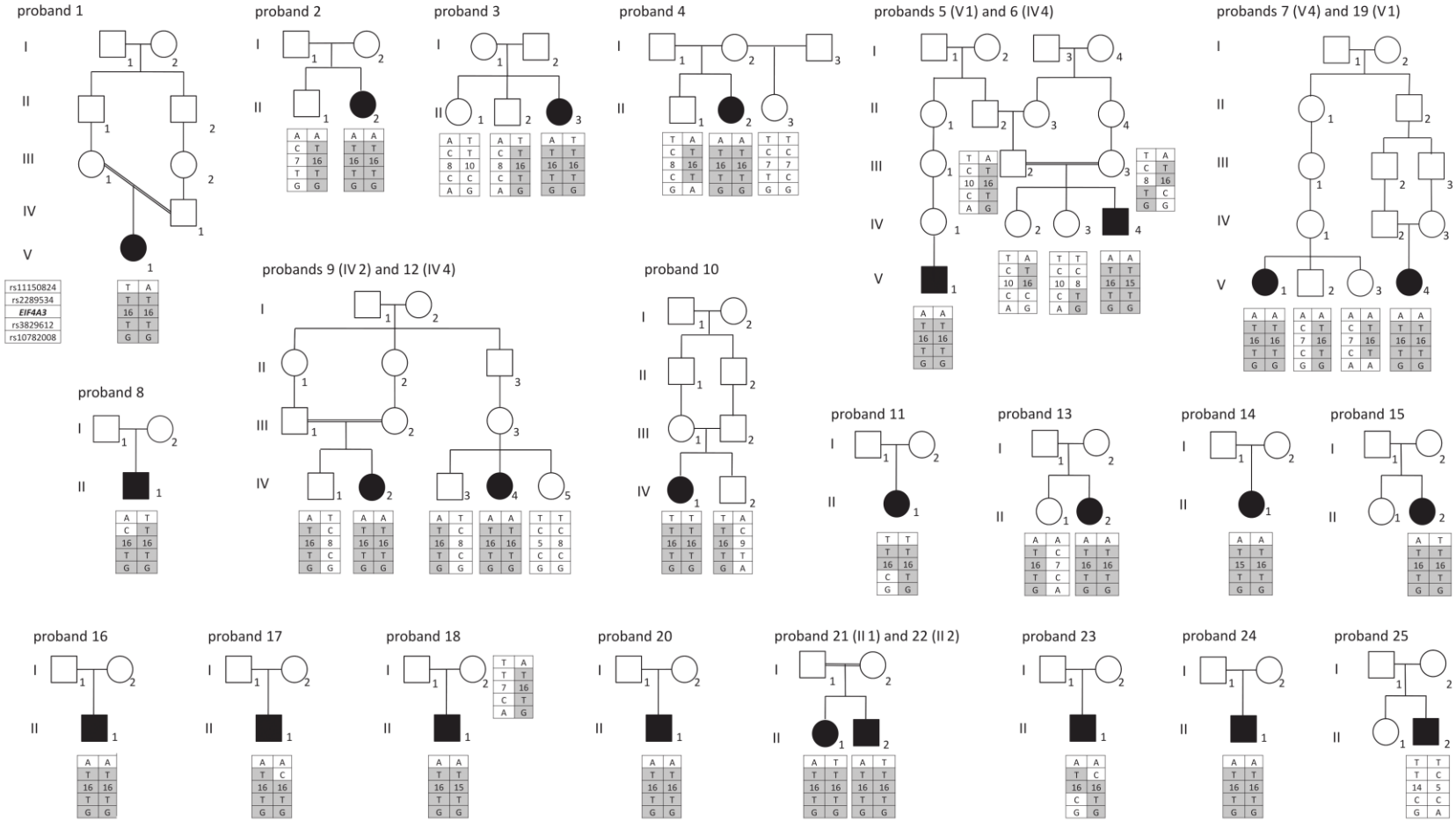


Figure S2: Schematic representation of the haplotypes from all the 25 affected probands and their families showing the segregation of the expanded allele with the disease following an autosomal recessive pattern of inheritance (co-segregating haplotype marked in grey). Note that

probands 6, 14 and 18 have one expanded 15-repeat allele while proband 25 is heterozygous for 5- and 14- repeat alleles. This 5-repeat allele is in cis with the nucleotide substitution c.809A>G (p.Asp270Gly).

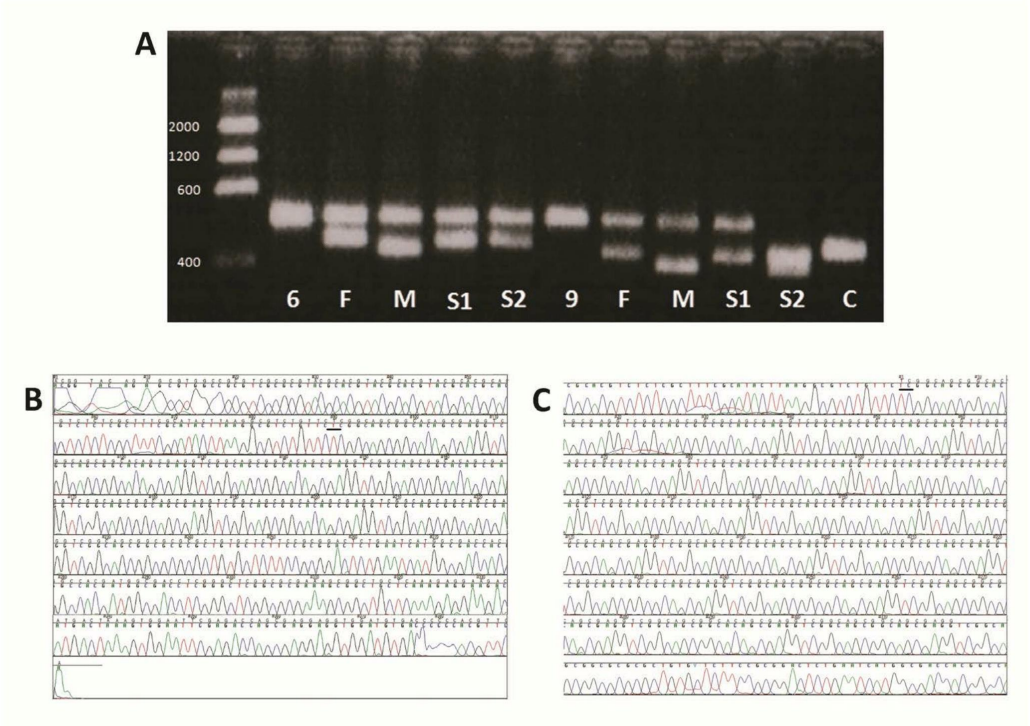


Figure S3: Visualization of the expanded alleles at the 5´ UTR of EIF4A3. (a) PCR products of the EIF4A3 5´ UTR containing the expanded alleles, showing homozygosity of the 16-repeat allele in affected individuals 6 and 9 and their unaffected relatives. M = Mother, F = Father; S1 = Sib 1 and S2 = Sib 2 of the affected individuals 6 and 9, respectively. (b) Chromatogram of the 5´ UTR of a control individual homozygous for a 7-repeat allele with the pattern 4 CACA-20nt/1 CA-18nt/1 CACA-20nt/1 CA-18nt. (c) Chromatogram of the 5´ UTR of an individual homozygous for the typical 16-repeat expanded allele. The first nucleotide T of the repeat is underlined.

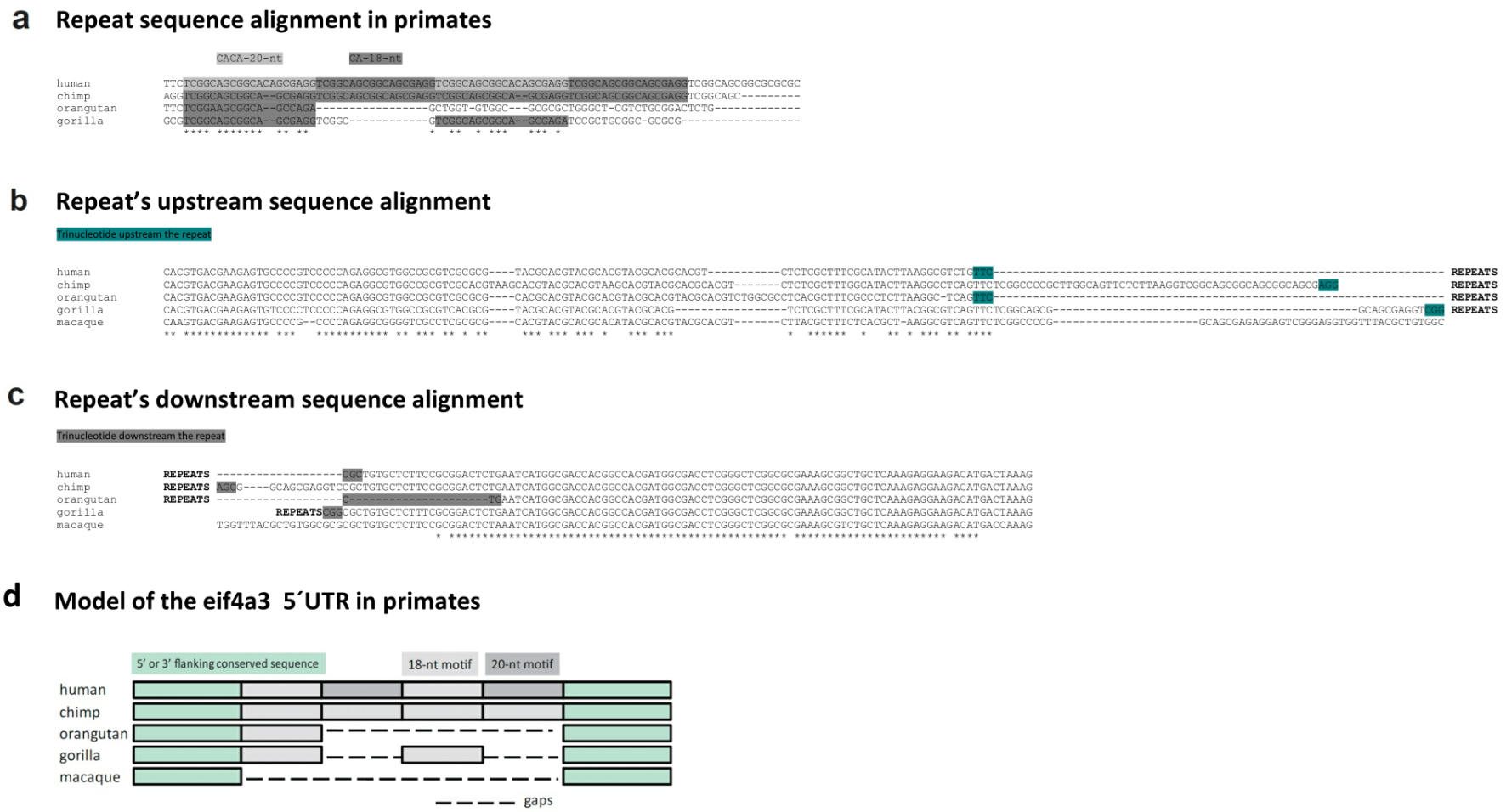


Figure S4: Comparative evolutionary analysis of the EIF4A3 5´ UTR in primates. (a) Alignment of primate 5´ UTR eif4a3 sequences showing a

similar repeat motif when compared to human sequence (first line) except for macaque. Note that only anthropoid primates (hominidae) share a highly conserved 18 nt motif (dark grey shaded) with the human EIF4A3 5´ UTR repeat. These repeats were not observed in other species (data not shown – used sequences listed in Table S6); (b) Alignment of primate upstream sequences of the EIF4A3 5´ UTR repeats showing that they are highly conserved in hominidae and non-hominidae primate (macaque). (c) Alignment of primate downstream sequences of the EIF4A3 5´ UTR repeats showing that they are highly conserved in hominidae and non-hominidae primate (macaque). (d) Illustrative scheme representing the evolutionary conserved sequences within the repeats of the 5´ UTR of EIF4A3 and their flanking regions.

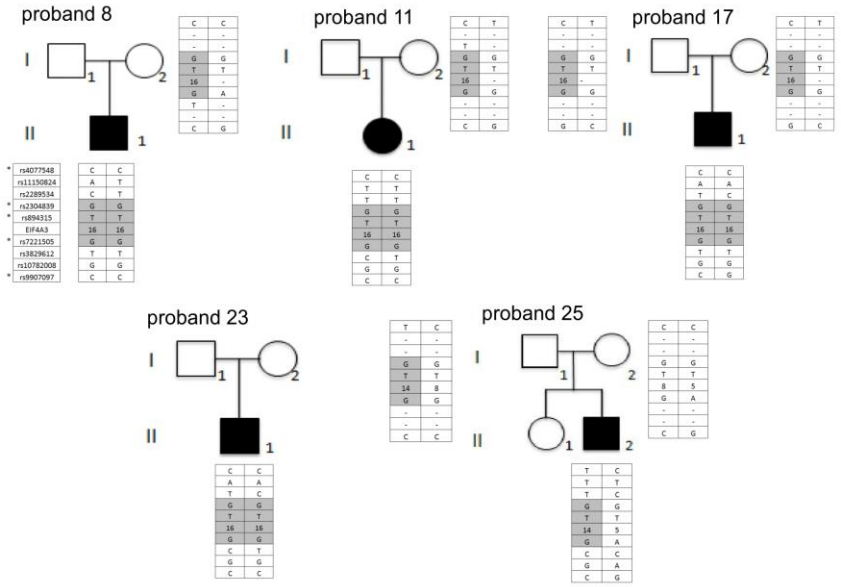


Figure S5: Haplotypes with five additional SNPs (rs4077548, rs2304839, rs894315, rs7221505 and rs9907097) spanning 600 kb of the EIF4A3. Based on HapMap, the markers rs2304839, rs894315, rs7221505 are within a linkage disequilibrium block. Note that all 16-repeat recombinant alleles as well as the 14-repeat allele share a common haplotype of markers rs2304839, rs894315, rs7221505, which span a region of approximately 42 kb (shaded). The missense mutation, in cis with the 5-repeat allele, is embedded in a different haplotype.

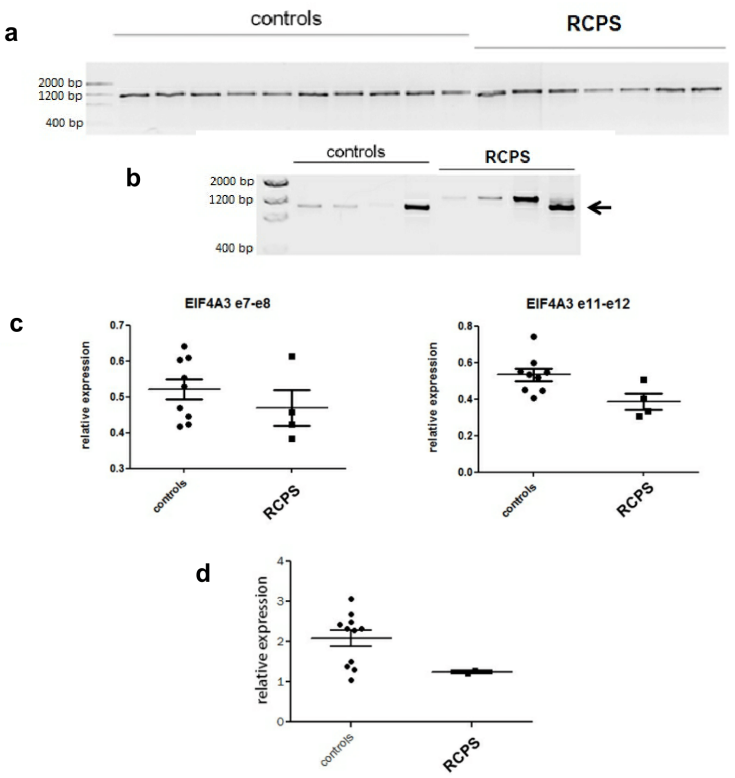


Figure S6: EIF4A3 expression in white blood cells and mesenchymal stem cells of affected individuals and controls. (a) Agarose gel electrophoresis of a PCR amplicon spanning exon 1 to exon 12 of EIF4A3 cDNA showing no evidence of mRNA splicing alterations (primers amplified the cDNA region from base +246 to +1452, with primers e1e2F and e11e12R); (b) No evidence of altered splicing was observed using the forward primer at the 5´UTR region (primers amplified cDNA region from base +2 to +957, with primers 5UTRF and e7e8R). The arrow indicates two different size products in affected individual 25 corresponding to the 5-repeat and 14-repeat alleles. (c) EIF4A3 real time quantitative PCR (RT-qPCR) was performed using 2 primer sets for exon junctions at e7-e8 and at e11-e12 (primers e7e8F, e7e8R and e11e12F, e11e12R, respectively: P= 0.0193, Student's t-test) in white blood cells. (d) EIF4A3 real time quantitative PCR (RT-qPCR) showed significant expression reduction in affected individuals when compared to controls in mesenchymal cells (primers e7e8F and

e7e8R; P = 0.001, Student's t-test). Error bars are relative to standard errors of the mean (SEM). Conditions for mesenchymal cell cultures and RT-qPCRs expression values calculations were described elsewhere32,33. Expression values were normalized using Genorm1 with the endogenous controls GAPDH, HPRT1, SDHA and TBP both in white blood cells and mesenchymal cells. Primer sequences: e1e2F: 5´ -GAAAGCGGCTGCTCAAAGAG-3´; e7e8F: 5´-TGGTTCTCATCAGTGCCACG-3´, e7e8R: 5´-TTCATCACGTTTCACCAAGATGC-3´, 5UTRF: 5´-CGCACGCACGTCTCTCGCTT-3´; e11e12F: 5´-GGAGATCAGGTCGATACGGC-3´, e11e12R: 5´-GATCAGCAACGTTCATCGGC-3´. 1: Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: RESEARCH0034.

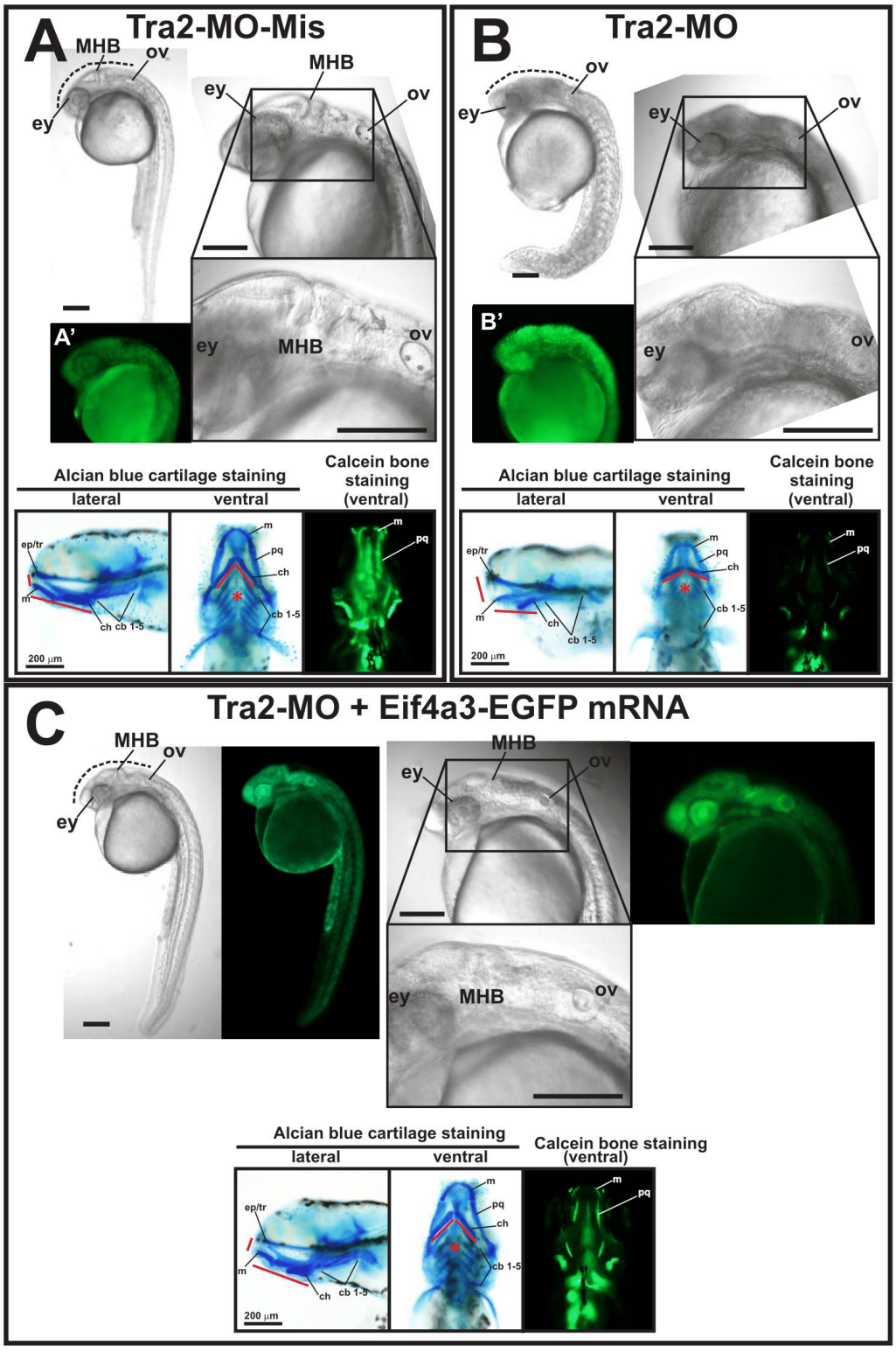


Figure S7: Zebrafish eif4a3 knock-down using Translation Morpholino (Tra2-MO). (a) 24-hpf control embryo microinjected with mispaired MO (Tra2-MO-Mis). (b) 24-hpf morphant embryo microinjected with MO designed to block eif4a3-mRNA translation

(Tra2-MO). (c) 24-hpf rescued embryos microinjected with Tra2-MO and mRNA coding for eif4a3 fused to EGFP (eif4a3-EGFP mRNA). Lateral views of embryos were registered under stereoscopic microscope (whole body at the upper left of each panel) or with two different magnifications under differential interference contrast microscope (anterior-most region at the upper and middle right of each panel). Dotted lines mark regions between eyes (ey) and otic vesicle (ov), which include midbrain-hindbrain border (MHB) and showed darkened tissue in morphants, suggesting the presence of apoptotic cells. (a') and (b') show lateral views of 24-hpf embryos after acridine orange staining. Green fluorescence panels in (c) show expression of eif4a3-EGFP in rescued embryo. At the bottom of each panel it is shown the craniofacial phenotype analysis (stained with alcian blue to observe cartilage structures or with calcein to observe bone structures) of 5 days post-feritlization (dpf) larvae corresponding to each treatment. Eyes were removed to register alcian blue stained larvae. Red lines and asterisks indicate mostly affected craniofacial cartilages (rostral portion, jaw elements). cb 1–5 = ceratobranchial arches 1–5; ch = ceratohyal; ey = eye; ep/tr = ethmoid plate/trabecula; m = Meckel's cartilage; MHB = midbrain-hindbrain border; ov = otic vesicle; pq = palatoquadrate. Scale bars represent 200 μm. TRA2-MO sequence: 5´-TTAAAATATGAAAGCGCGACTGTGC-3´. TRA2-MO-Mis sequence: 5´-TTAATAAATCAAAGCCCCACTGTGC-3´.

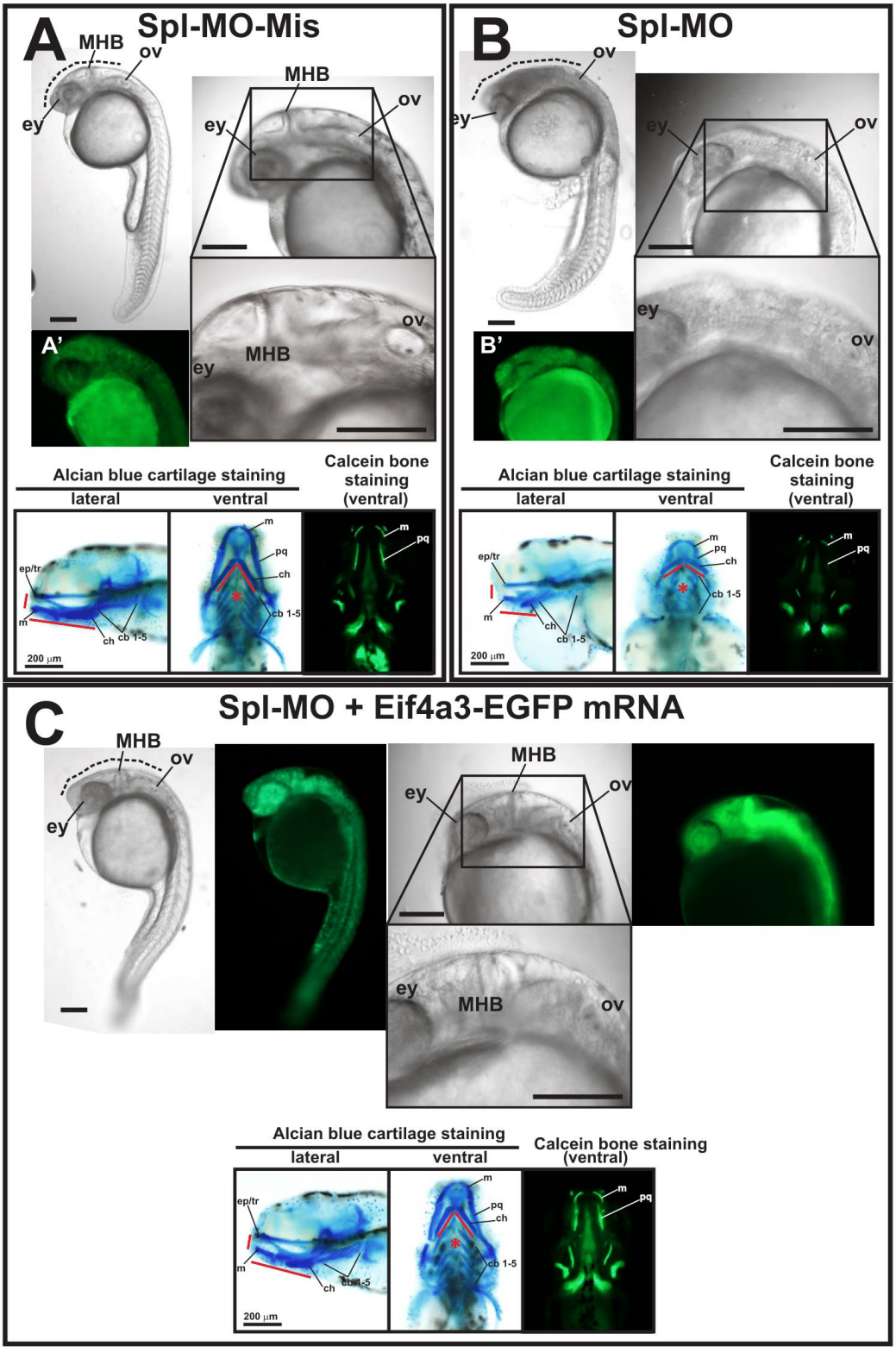


Figure S8: Zebrafish eif4a3 knock-down using Spl Morpholino (Spl-MO). (a) 24-hpf control embryo microinjected with mispaired MO (Spl-MO-Mis). (b) 24-hpf morphant embryo microinjected with MO designed to block eif4a3-pre-mRNA splicing (SplMO).

(c) 24-hpf rescued embryo microinjected with Spl-MO and mRNA coding for eif4a3 fused to EGFP (eif4a3-EGFP mRNA). Lateral views of embryos were registered under stereoscopic microscope (whole body at the upper left of each panel) or with two different magnifications under differential interference contrast microscope (anteriormost region at the upper and middle right of each panel). Dotted lines mark regions between eyes (ey) and otic vesicle (ov), which include midbrain-hindbrain border (MHB) and showed darkened tissue in morphants, suggesting the presence of apoptotic cells. (a') and (b') show lateral views of 24-hpf embryos after acridine orange staining. Green fluorescence panels in (c) show expression of eif4a3-EGFP in rescued embryo. At the bottom of each panel it is shown the craniofacial phenotype analysis (stained with alcian blue to observe cartilage structures or with calcein to observe bone structures) of 5 days post-fertilization (dpf) larvae corresponding to each treatment. Eyes were removed to register alcian blue stained larvae. Red lines and asterisks indicate mostly affected craniofacial cartilages (rostral portion, jaw elements). cb 1–5 = ceratobranchial arches 1–5; ch = ceratohyal; ey = eye; ep/tr = ethmoid plate/trabecula; m = Meckel's cartilage; MHB = midbrain-hindbrain border; ov = otic vesicle; pq = palatoquadrate. Scale bars represent 200 μm. Spl-MO sequence: 5´-GGATTTTTGTGGTGTTTTTTACCGT-3´. Spl-MO-Mis sequence: 5´-GGATTTATCTGGTCTTTTTTAGCCT-3´.

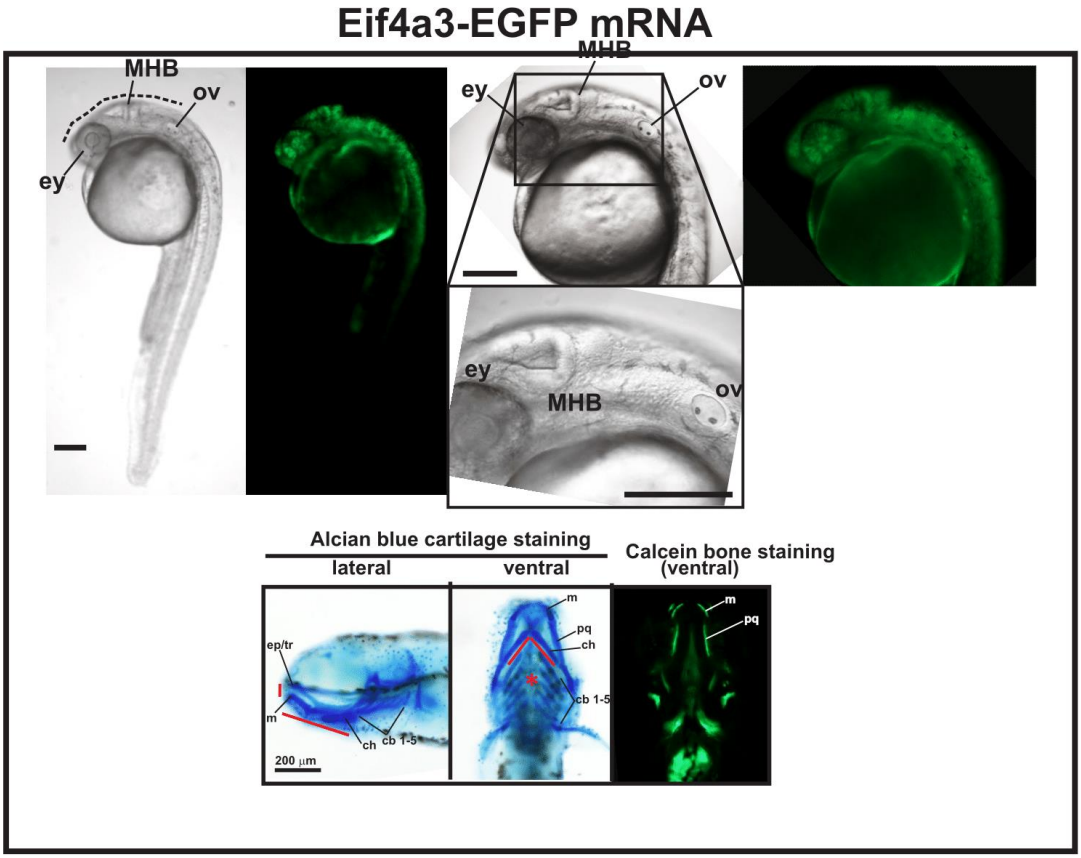


Figure S9: Zebrafish embryos microinjected with mRNA coding for eif4a3 fused to EGFP (eif4a3-EGFP mRNA). Lateral views of 24-hpf embryos microinjected with eif4a3-EGFP mRNA were registered under stereoscopic microscope (whole body at the upper left of each panel) or with two different magnifications under differential interference contrast microscope (anterior-most region at the upper and middle right of each panel). Dotted lines mark regions between eyes (ey) and otic vesicle (ov), which include midbrain-hindbrain border (MHB). Green fluorescence panels show expression of eif4a3-EGFP. At the bottom it is shown the craniofacial structures analysis (stained with alcian blue to observe cartilage structures or with calcein to observe bone structures) of 5 days post-fertilization (dpf) larvae microinjected with eif4a3-EGFP mRNA. Eyes were removed to register alcian blue stained larvae. Red lines and asterisks indicate rostral portion, jaw elements. cb 1–5 = ceratobranchial arches 1–5; ch

= ceratohyal; ey = eye; ep/tr = ethmoid plate/trabecula; m = Meckel's cartilage; MHB = midbrain-hindbrain border; ov = otic vesicle; pq = palatoquadrate. Scale bars represent 200 μm.

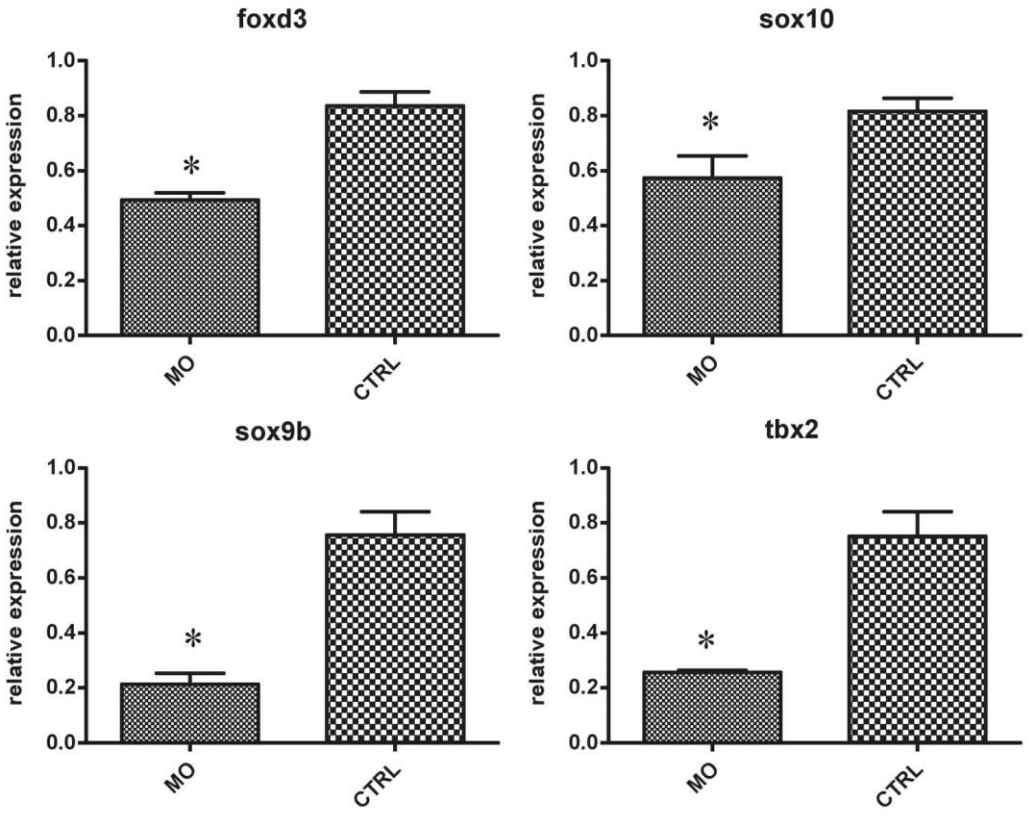


Figure S10: Expression of neural crest markers foxd3, sox10, sox9b and tbx2 are significantly reduced in zebrafish eif4a3 morphants (MO) in comparison to mismatch control morphants (CTRL) (P<0,05, Student's T-test). Error bars are relative to standard errors of the mean (SEM), calculated with technical triplicates. Conditions for embryo RNA extractions, RT-qPCRs expression values calculations and primer sequences were described elsewhere2 . 2: Weiner A.M.J., Scampoli N.L., Calcaterra N.B. (2012) Fishing the Molecular Bases of Treacher Collins Syndrome. PLoS ONE 7(1): e29574. doi:10.1371/journal.pone.0029574.

Table S1 Clinical and molecular data of affected individuals

Affected Individuals	Molecular alterations (repeats)	Clinical findings							
		Microstomia	Micrognathia	Cleft palate	Absent midline mandibular fusion	Abnormal larynx[5]	Radial ray abnormalities	Abnormal tíbia	Club foot
01	16/16	+	+	+	+	NE	+	+	+
02	16/16	+	+	+	+	NE	+	+	+
03	16/16	+	+	+	+	+	+	+	+
04	16/16	+	+	+	+	+	+	+	+
05[1]	16/16	+	+	-	+	+	+	+	+
06[1]	16/15	+	+	-	+	+	+	+	+
07[2]	16/16	+	+	+	+	+	+	+	+
08	16/16	+	+	-	+	+	+	+	+
09[3]	16/16	+	+	+	+	+	+	+	+
10	16/16	+	+	+	+	+	+	+	+
11	16/16	+	+	-	+	+	+	+	+
12[3]	16/16	+	+	+	+	+	+	+	+
13	16/16	+	+	+	+	+	+	+	+
14	16/15	+	+	+	+	+	+	+	+
15	16/16	+	+	+	+	+	+	+	+
16	16/16	+	+	+	+	+	+	+	+
17	16/16	+	+	+	+	+	+	+	+
18	16/15	+	+	+	+	+	+	+	+
19[2]	16/16	+	+	+	+	NE	+	+	+

20	16/16	+	+	+	+	+	+	+	+
21[4]	16/16	+	+	+	+	+	+	+	+
22[4]	16/16	+	+	+	+	+	+	+	+
23	16/16	+	+	-	+	+	+	+	+
24	16/16	+	+	NE	+	+	+	+	+
25	14/ c.809A>G	-	+	-	-	+	+	+	-

+: present

-: non present NE:

not evaluated

[1]Individuals 5 and 6 belong to the same family

[2]Individuals 7 and 19 belong to the same family

[3]Individuals 9 and 12 belong to the same family

Individuals 1-20 described in ref 1 (Individuals 1-3, 5-8, 10-13, 17, 19, 21-25, 27 and 28, respectively)

[4]Individuals 21 and 22 belong to the same family were described in ref 17 (Individuals 1 and 2, respectively)

Individual 23 described in ref 18

Individuals 24 and 25 are new cases

[5]Abnormal Larynx included: round shape, small size, absence of epiglottis, arytenoids hypertrophic, aryepiglottic folds reduced and hypertrophic

Table S2 Microsatellites and SNPs used to characterize the homozygous region of chromosome 17 (GRCh37).

Physical position	Marker	Source	Heterozygosity	Sense Strand primer	Antisense strand primer	Product size
68465456-68465676	**D17S949**	*ABI PRISM*â	0,8	ATAGAAACTCCACATTTGCATTA	CTTTCCCACNCGTGTC	207-221
74431373-74431569	**D17S785**	*ABI PRISM*â	0,83	ATCCCTGGAGAGTGAAAATG	AAGGCCAACCTGAAAACTAA	181-207
77802191-77802424	**D17S784**	*ABI PRISM*â	0,77	GAGTCTCCTAAATGCTGGGG	AGCTCCTGCACAGTTCTTAAATA	226-238
80252880-80253028	**D17S928**	*ABI PRISM*â	0,77	TAAAACGGCTACAACACATACA	ATTTCCCCACTGGCTG	135-165
75347271-75347415	**D17S937**	*NCBI*	0,7	CATGGAGGGACTTGCG	TTCCCAGAACCCGGTTT	125-149
76234610 76234787	¯ **D17S802**	*NCBI*	0,81	GCCACCTGCCCCTCAA	CTGCCAGCAGAGGCCA	166-188
77024861-77025022	**D17S1847**	*NCBI*	0,66	GATCACCAGGAACACCC	TCTTCAGAGCTTGCCAG	144-164
77445675-77445787	**D17S1806**	*NCBI*	0,72	GATGTGCTTATTTGAAACCTGC	TGTAACGTCCACCAGCAGAG	109-151
80028021-80028175	**D17S668**	*NCBI*	0,88	TGACAGAGCGAGACCCTGC	GAACGGGATTGGTTATACTAATGTC	151-174
77925092	**rs11150824**	*NCBI*	0.367 (HapMap-CEU)	TGCAGCTCAAAGACCAGGTA	AATAAGGGAAGCCAGCAGGA	395
78039619	**rs2289534**	*NCBI*	0.375 (HapMap-CEU)	AGGACACCCGGATTTTAAGG	AGGCCTTCTGTCACCCACTA	263
78161402	**rs3829612**	*NCBI*	0.460 (HapMap-CEU)	TCCCTTCCAGCTCTGACTTG	GACAGCAGCCCAAGAGAAAC	311
78305871	**rs10782008**	*NCBI*	0.442 (HapMap-CEU)	GCTGTGTGCTCTTGAGTTCG	TCTTCTTCGGGCTCACTCAG	387
78444658	**rs12943620**	*NCBI*	0.366 (HapMap-CEU)	TGAAGGACACTCTGGGTGGT	AGGGAGAGAAGAGACGCACA	310
78568447	**rs4889865**	*NCBI*	0.383 (HapMap-CEU)	GAGTTCGATACCAGCGTTTC	TTACCATGGAGCCACAGTGC	376
78806313	**rs2672871**	*NCBI*	0.310 (HapMap-CEU)	TGGCCACTGGAGCAACATT	CAGTCCTCAGTTGCAGTTGG	329
79165171	**rs869370**	*NCBI*	0.600 (AFD EUR PANEL)	CACACTGGGTCCTGACAACA	CAGTGGTTCATTGCCGTAGT	309
79417400	**rs899286**	*NCBI*	0.339 (HapMap-CEU)	TGCTGTGAGGTTATCGTTGC	ATTCCTCCATCCTCCCTCAG	571
79617489	**rs8477**	*NCBI*	0.363 (HapMap-CEU)	TCACAGTCATCTGCCCTTG	GGGCCTGTATCTCCAGATGT	315
77730756	**rs4077548**	*NCBI*	0.393 (1000 Genomes)	AGAGGTAAGGGGTGAAGCGA	GGCACTTCATCCCATCTGCT	352
78086003	**rs2304839**	*NCBI*	0.918 (1000 Genomes)	ACCTGATGCCTGTAGGTGAG	CAGCCACCAGTTTCTCTCCA	379
78098700	**rs894315**	*NCBI*	0.967 (1000 Genomes)	GTGGTTTCCAGTCAGTGGGT	GGATTATGGTGGGTGAGGGC	569
78128231	**rs7221505**	*NCBI*	0.271 (1000 Genomes)	GTTAAGGCAGTTGTGGGGGT	GCAACAGAGCGAGACTCCA	251
78396091	**rs9907097**	*NCBI*	0.480 (1000 Genomes)	GAGACAGTGCCCCCTTCTTG	AGAGGTAAGGGGTGAAGCGA	352

Table S3. Variants detected in genes at the candidate 17q25.3 region (GRCh37/hg19 – chr17: 78,039,369 - 78,161,152) by Sanger sequencing of all exons and introns boundaries. Their position within genes, the corresponding described SNPs and Poly-Phen2 prediction are indicated. *GAA* was not sequenced as it is associated to GSD2.

Gene	cDNA variant	Protein effect	SNP	Poly-Phen2
CCDC40	No alterations	-	-	-
CARD14	c.676-6G>A	unknown	rs286740001	Benign
	c.1641G>C	p.Arg547Ser	rs2066964	Benign
	c.2458C>T	p.Arg820Trp	rs11652075	Benign
EIF4A3	No alterations	-	-	-

Table S4- - Diseases caused by non-coding expansion. Except for Richieri-Costa-Pereira Syndrome (RCPS), all the others are neurological or neuromuscular diseases [8-11]. "Repeat expansion size" and "Minimum length (bp)" are referred respectively by the lowest number of repeats and the lowest total expansion length characteristic of each disease.

Disease	Inheritance*	Expanded repeat structure	Repeat expansion size	Minimum length (bp)	Repeat position-Gene	Molecular pathogenicity
Fragile-X syndrome (full mutation) [MIM 300624]	X-L-R	CGG	> 200	600	5'UTR-*FMR1*	Loss-of -function / Gain-of-function
Fragile-X syndrome (pre-mutation) (FXTAS [MIM 300623])	X-L-R	CGG	55-200	165	5'UTR-*FMR1*	Loss-of-function
Spinobulbar muscular atrophy, X-linked 1 (SMAX1 [MIM 313200])	X-L-D	CAG	38-62	186	5'UTR-*AR*	Gain-of-function
Myotonic dystrophy type 1 (DM1 [MIM 160900])	AD	CTG	>50	150	3'UTR – *DMPK*	Gain-of-function
Friedreich ataxia (FRDA [MIM 229300])	AR	GAA	>70	210	Intron 1 – *FXN*	Loss-of-function
Myotonic dystrophy type 2 (DM2 [MIM 602668])	AD	CCTG	>75	225	Intron 1 - *ZNF9*	Gain-of-function
Spinocerebellar ataxia type 10 (SCA10 [MIM 603516])	AD	ATTCT	>400	2000	Intron 9 - *ATXN*10	Loss-of -function / Gain –of-function
Progressive myoclonus epilepsy (PME [MIM 254800])	AR	C(4)GC(4)GCG	3	36	5'UTR- *CSTB*	Loss-of -function
Frontemporal dementia/ALS** (FTDALS [MIM 105550])	AD	GGGGCC	700-1600	4200	Exon 1a-1b - *C9ORF72*	Loss-of-function /Gain-of -function
Spinocerebellar ataxia type 36 (SCA36 [MIM 614153])	AD	GGCCTG	>1500	9000	Intron 1 - *NOP56*	Gain-of-function
Spinocerebellar ataxia type 8 (SCA8 [MIM 608768])	AD	CTG	>71	213	3'UTR - *ATXN8OS*	Gain-of-function
Spinocerebellar ataxia type 2 (SCA2 [MIM 183090])	AD	CAG	>31	93	Exon 1 - *ATXN2*	Gain-of-function
Spinocerebellar ataxia type 31 (SCA31 MIM 117210])	AD	TGGAA	>110	550	Intron 1 – *BEAN*	Gain-of-function
Richieiri-Costa-Pereira (RCPS [MIM 268305])	AR	TCGGCAGCGGCGC AGCGAGG	>14	240	5'UTR - *EIF4A3*	Probably loss-of-function

*X-L: X-linked; A: autosomal; D: Dominant; R: Recessive

**ALS amytropic Lateral

Table S5. Percentages of normal, affected (morphant eif4a3 phenotype), or dead 24-hpf zebrafish embryos after microinjection of TRA-1, TRA-2, SPL morpholinos and their respective mismatch control morpholinos (Mis) .Morpholinos were injected alone or combined with 200 ng/ml eif4a3-EGFP mRNA. Total number of treated embryos per condition are indicated in the last row.

	Control (not injected)	TRA1-MO	TRA1-MO-Mis	TRA2-MO	TRA2-MO-Mis	SPL-MO	SPL-MO-Mis	mRNA	TRA1-MO/mRNA	TRA2-MO/mRNA	SPL-MO/mRNA
Normal (%)	97	23.8	94.8	28.5	92.15	44	96.8	91.5	90.65	79.1	86
Morphant eif4a3 Phenotype (%)	0	70.7	0	69.5	2.9	56	0	0.5	7.05	17.4	15.7
Dead (%)	3	5.5	5.2	2	4.95	0	3.2	8	2.3	3.5	1.7
Total (n)	201	188	194	280	363	216	61	167	232	217	203

Table S6 Gene IDs and specific sequences coordinates used for motif conservation analysis and alignment in **Figure S4**.

Species	Ensembl Gene ID	Genome Version	Genomic Coordinates
Human	ENSG00000141543	GRCh37	17:78108413:78120575
Mouse	ENSMUSG00000025580	GRCm38	11:119287763:119299813
Dog	ENSCAFG00000005614	CanFam3	1:9:1581439:1592805
Cat	ENSFCAG00000008246	Felis_catus_62	E1:60532616:60542095
Opossum	ENSMODG00000002190	BROADO5	2:235817664:235842126
Hyrax	ENSPCAG00000002931	proCap1	GeneScaffold_3781:60245:74378
Elephant	ENSLAFG00000015628	loxAfr3	Scaffold_49:2786073:2802935
Panda	ENSAMEG00000015194	ailMel1	GL192862.1:276026:287365
Squirrel	ENSSTOG00000002664	spetri2	JH393404.1:4148019:4160038
Cow	ENSBTAG00000016023	UMD31	19:53084863:53096981
Pig	ENSSSCG00000017155	Sscrofa102	12:2266908:2294687
Horse	ENSECAG00000015925	EquCab2	11:2814386:2827608
Macaque	ENSMMUG00000009590	MMUL_1	16:75579204:75594242
Orangutan	ENSPPYG00000008714	PPYG2	17:70181996:70194118
Gorilla	ENSGGOG00000012918	gorGor31	5:3006126:3018309
Chimpanzee	ENSPTRG00000009737	CHIMP214	17:79318890:79439810

Table S7 Proteins IDs from *EIF4A3*, *EIF4A2* and *EIF4A1* orthologs used for alignment to check protein conservation at human eIF4AIII p.Asp270 (**Figure 3**). IDs are referred to Ensembl or NCBI protein sequences.

Sequence name	**Protein ID**
human *EIF4A3*	ENSP00000269349
human *EIF4A2*	ENSP00000326381
human *EIF4A1*	ENSP00000293831
chimp *eif4a3*	ENSPTRP00000016565
chimp *eif4a2*	ENSPTRP00000027042
chimp *eif4a1*	ENSPTRP00000014825
mouse *eif4a3*	ENSMUSP00000026667
mouse *eif4a2*	ENSMUSP00000023599
mouse *eif4a1*	ENSMUSP00000127034
chicken *eif4a3*	ENSGALP00000038179
chicken *ei4a2*	ENSGALP00000014119
xenopus *eif4a3*	ENSXETP00000030049
xenopus *eif4a2*	ENSXETP00000042005
xenopus *eif4a1*	ENSXETP00000043428
anole lizard *eif4a3*	ENSACAP00000014210
anole lizard *eif4a2*	ENSACAP00000007590
anole lizard *eif4a1*	ENSACAP00000005480
zebrafish *eif4a3*	ENSDARP00000027276
zebrafish *eif4a2*	ENSDARP00000104653
zebrafish *eif4a1*	ENSDARP00000058912
yeast *eif4a3*	NP_010304.3
yeast *eif4a1*	NP_012985.3
arabidopsis *eif4a3*	NP_188610.1

arabidopsis *eif4a2*	NP_175829.1
arabidopsis *eif4a1*	NP_177417.1

Chapter 7

DNA methylation in *EIF4A3* repeats is not a mechanism for gene downregulation in Richieri-Costa-Pereira syndrome

Lucas Alvizi Cruz, Lucas Vechiato de Melo, Maria Rita dos Santos Passos-Bueno

Centro de Pesquisas Sobre o Genoma Humano e Células-Tronco, Instituto de Biociências, Universidade de São Paulo

Key-words: Richieri-Costa-Pereira Syndrome, cleft palate, EIF4A3, DNA methylation, repeat expansion

Abstract

Richieri-Costa-Pereira syndrome (RCPS) is a rare autosomal recessive acrofacial dysostosis including both craniofacial and limb malformations. Complex repeat expansions at the Eukaryotic translation factor 4A3 (*EIF4A3*) 5'UTR have been described as the main mutational cause, leading to gene downregulation and compromised bone and cartilage differentiation. Because such repeats are CG rich, we hypothesised DNA hypermethylation in RCPS samples as a potential cause for *EIF4A3* repression. To test this hypothesis, we performed bisulfite sequencing in the *EIF4A3* repeat region in 6 RCPS patients and 7 controls using whole-blood DNA samples. We observed that RCPS expansions resulted in a gain of 37 CpG sites in comparison to controls, however with no significant hypermethylation. Again, inspecting methylation differences at CpG sites individually did not reveal any evident methylation increase. We therefore refuted the DNA hypermethylation at the *EIF4A3* as a mechanism for gene downregulation in RCPS. The investigation of additional transcriptional or post-transcriptional mechanisms involving repeat interaction with repressors, RNA transport or RNA-protein interaction may clarify how repeat expansions in *EIF4A3* interfere in gene expression.

Resumo

A síndrome de Richieri-Costa-Pereira (*Richieri-Costa-Pereira Syndrome* – RCPS) é uma disostose acrofacial rara de herança autossômica recessiva, a qual inclui tanto malformações craniofaciais quanto de membros. Expansões de repetições complexas na região 5'UTR do gene *EIF4A3* (*Eukaryotic translation fator 4A3*) foram descritas como a principal mutação causal da síndrome, levando à redução da expressão do gene e comprometendo a diferenciação óssea e condrogênica. Uma vez que tais repetições são ricas em dinucleotídeos CG, nós criamos a hipótese de que as amostras RCPS apresentariam hipermetilação de DNA na região das repetições como causa potencial para a repressão do *EIF4A3*. Para testar esta hipótese, nós realizamos sequenciamento de bissulfito para a região de repetição do *EIF4A3* em 6 amostras RCPS e 7 controles usando-se DNA de sangue total. Nós observamos que as expansões em RCPS resultaram em um ganho de 37 sítios CpGs em comparação aos controles, no entanto sem aumento significativo de metilação. Da mesma forma, analisando-se os valores de metilação por cada sítio CpG individualmente também não encontramos nenhum aumento evidente de metilação. Nós, portanto, refutamos a hipótese de hipermetilação de DNA nas repetições do *EIF4A3* como mecanismo para redução da expressão do gene em RCPS. Assim, a investigação de mecanismos regulatórios transcricionais ou pós-transcricionais que envolvam a interação de repetições com repressores, transporte de RNA ou interação RNA-proteína podem melhor explicar de que modo as repetições no *EIF4A3* interferem na expressão gênica.

Introduction

Richieri-Costa-Pereira syndrome (RCPS; OMIM #268305) is a rare acrofacial dysostosis characterised by cleft palate, median mandibular cleft and other craniofacial anomalies as well as limb defects (FAVARO et al., 2011). RCPS segregates in an autosomal recessive fashion and mutations at the *EIF4A3* gene have been described to cause this syndrome (FAVARO et al., 2014).

The most typical forms of RCPS are caused by a complex repeat expansion in the 5'UTR of *EIF4A3*, in which affected individuals are homozygous for the 16-repeat allele mainly consisting of a 20 nucleotide unit referred as 20nt-CGCA (TCGGCAGCGG**CGCA**GCGAGG) while non-affected individuals are frequently homozygous for a 8-repeat allele which is also predominantly consisted of a 20 nucleotide unit, however referred as 20nt-CACA (TCGGCAGCGG**CACA**GCGAGG) (Figure 1). Nevertheless, the repeat structure varies both in RCPS and control individuals, in which RCPS may display 14 or 15 repeat alleles and controls 3 to 12 repeat alleles. Also, *EIF4A3* missense mutation have been reported in one RCPS patient (FAVARO et al., 2014).

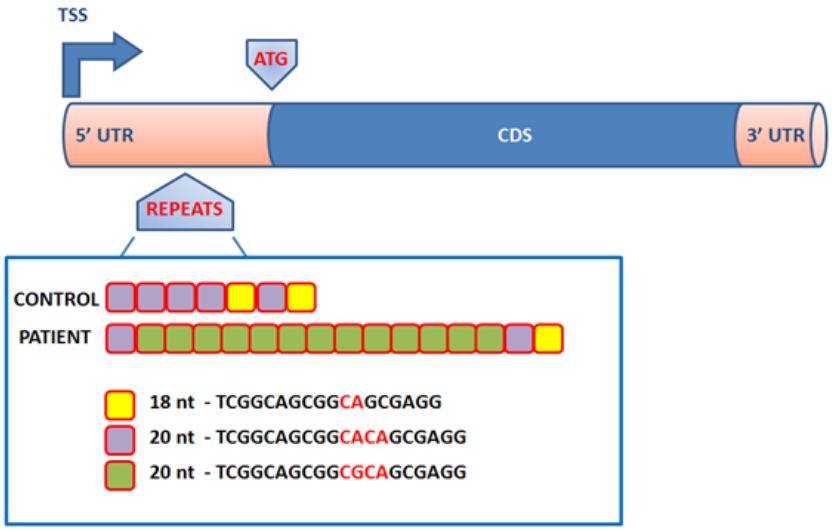


Figure 1: Simplified model of repeat structure in RCPS and controls. Repeats occur at the 5'UTR of *EIF4A3* and consist of 3 repeat types: 18nt-CA, 20nt-CACA and 20nt-CGCA. Expansion observed in RCPS patients is mainly given by the 20nt-CGCA repeat.

Recent work has importantly uncover cellular mechanisms by which *EIF4A3* loss-of-function lead to craniofacial abnormalities. *EIF4A3* deficiency results in decreased neural-crest cells migratory capacity and premature bone differentiation as well as compromised cartilage differentiation (MILLER et al., 2017). However, how repeat expansion in RCPS cause the observed *EIF4A3* downregulation still remains unclear. By closely inspecting the expansion structure, *EIF4A3* repeats are CG rich and presents 5 CpGs in the control 20nt-CACA unit, while 6 CpGs in the RCPS 20nt-CGCA unit. When considering the total expansion, this results in a notable CpG increase in RCPS repeats (Figure 2). Because CpG is the most studied target for DNA methylation and is involved to gene repression especially when methylation occurs at promoter/5'UTR (FEIL; FRAGA, 2012; SCHÜBELER, 2015; TAMMEN; FRISO; CHOI, 2013), we

hypothesised DNA hypermethylation at the RCPS repeats as a molecular consequence of the expansion leading to *EIF4A3* downregulation.

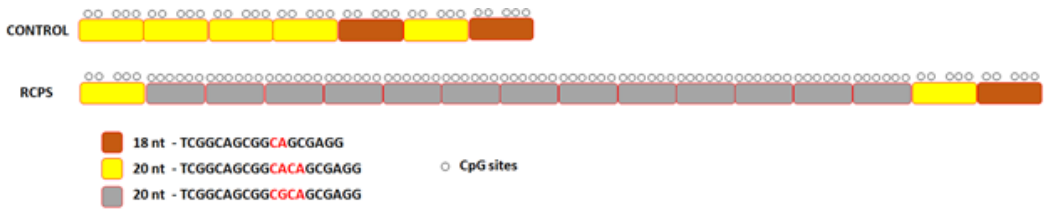


Figure 2: Model of CpG gain in RCPS patients in comparison to controls revealing an increase in CpG quantity in RCPS.

In this study we investigated DNA methylation at the *EIF4A3* 5'UTR repeat region as possible cause for gene downregulation in RCPS in comparison to controls.

Methods

Ethics

This study was approved by the Ethics Committee of the Instituto de Biociências (Universidade de São Paulo, Brazil). Biological samples were collected after signed informed consent by the parents or legal guardians. All experiments were performed in accordance with relevant guidelines and regulations.

Patients and Samples

RCPS samples (n=6) were previously described by our group (FAVARO et al., 2014) and were collected at the Hospital de Reabilitação de Anomalias Craniofaciais da Universidade de São Paulo (HRAC-USP) during patient's treatment and follow-ups. RCPS samples were from patients with the typical mutation consisting of homozygous 16 repeats alleles. Control samples were collected at the Centro de Pesquisas Sobre o Genoma Humano da Universidade de São Paulo (CPGH-CEL USP) and were also previously described by our group (FAVARO et al., 2014), consisting of 8 repeats allele individuals. All samples are from peripheral blood and DNAs were automated extracted (Autopure LS -Gentra Systems).

Methylation analysis

1ug of genomic DNA from each sample were submitted to bisulfite conversion using EpiTect Bisulfite Conversion Kit (QIAgen). Bilsulfite converted DNA was subsequently used in PCR for the *EIF4A3* repeat region, in which primers were designed with MethPrimer (www.urogene.org/methprimer/) (LI; DAHIYA, 2002) (Forward: 5'-GYGAGAGTAGAAATATTTTATTTTTT -3'; Reverse: 5'-TCATATCTTCCTCTTTAAACAACC-3'). Amplicons were checked by agarose electrophoresis and cloned using the TOPO TA Cloning Kit for Sequencing (ThermoFisher), and Sanger sequencing was carried out for 10 clones per sample using the BigDye® Terminator v3.1 Cycle Sequencing Kit and the ABI 3730 DNA Analyser. Sequencing files were then analysed for methylation quantification using the online tool BISMA - Bisulfite Sequencing DNA Methylation Analysis (ROHDE et al., 2010). Methylation values were computed and differences tested between groups using Fisher's Exact Test.

Results

We inspected methylation levels at the 5'UTR expansion region of *EIF4A3* in 6 RCPS and 7 control samples. That resulted in the analysis of 58 and 21 CpG sites respectively in RCPS and control samples, representing a 37 CpG sites gain in RCPS samples (Figure 3). We observed that both RCPS and controls presented highly unmethylated levels of the expansion region, with a significant reduction of methylation in RCPS of 1.1%, on the opposite of our hypothesis (Table I).

Table I: Total methylation levels in the analysed expanded regions in RCPS and control samples. We observed a slight and significant methylation reduction in RCPS samples in comparison to controls (p=0.0461, Fisher's Exact Test).

	RCPS		Controls		
	Number of CpGs	%	Number of CpGs	%	p-value
Methylated CpGs	40	1.7	29	2.8	0.0461
Non-methylated CpGs	2352	98.3	1011	97.2	
Total	2392		1040		

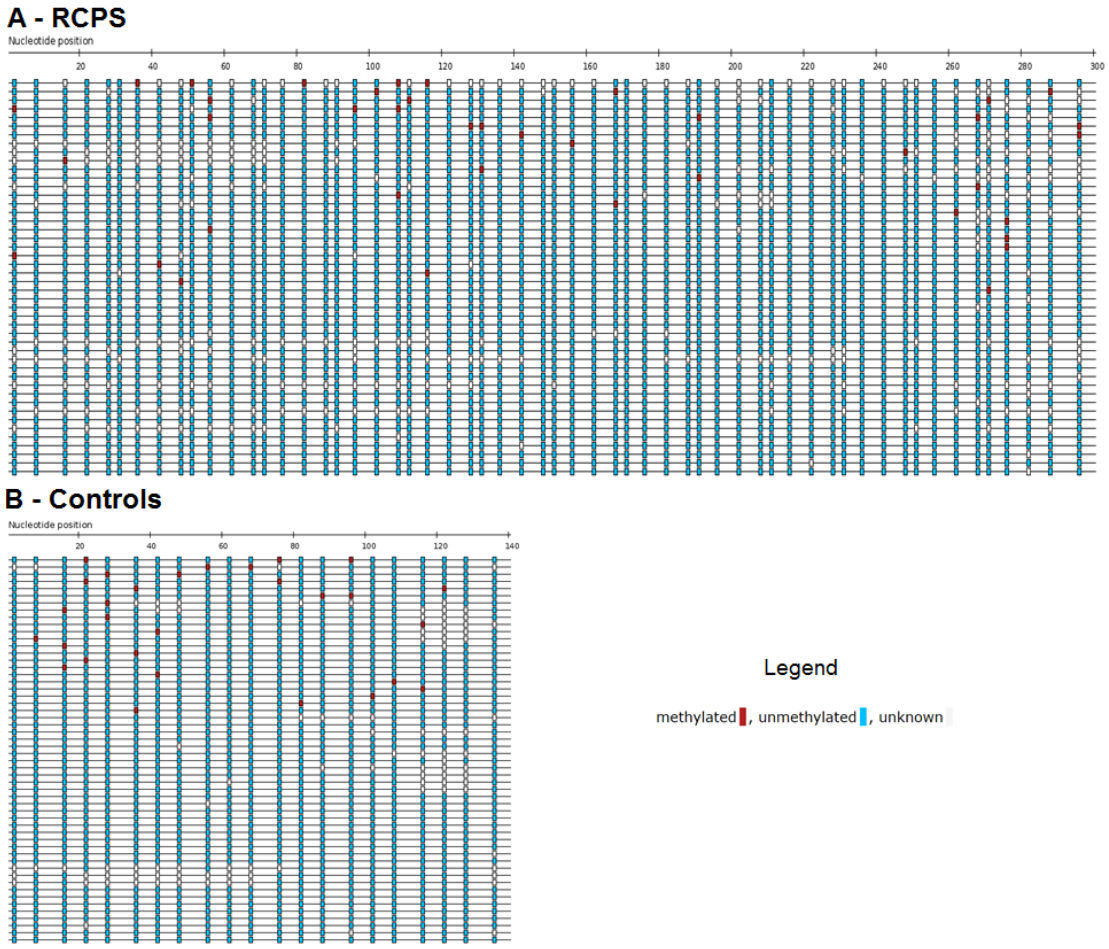


Figure 3: Representation of methylated and unmethylated cytosines observed in each sequence at the reference CpG sites in context of the sequence length. Sequences are represented from the most methylated (top) to the least methylated (bottom). RCPS sequences comprised 58 analysed CpG sites (A) while 21 in controls (B). Red boxes = methylated CpGs. Blue boxes = unmethylated CpGs. White boxes = unknown/undetermined methylation level.

Because there is no ideal alignment between RCPS and control sequences, especially after bisulfite conversion, we were not able to compare methylation levels per CpG between those groups. However, we can observe that there is no evident methylation variation at any specific CpG both in RCPS and controls, in which all presented low levels of methylation (Figure 4).

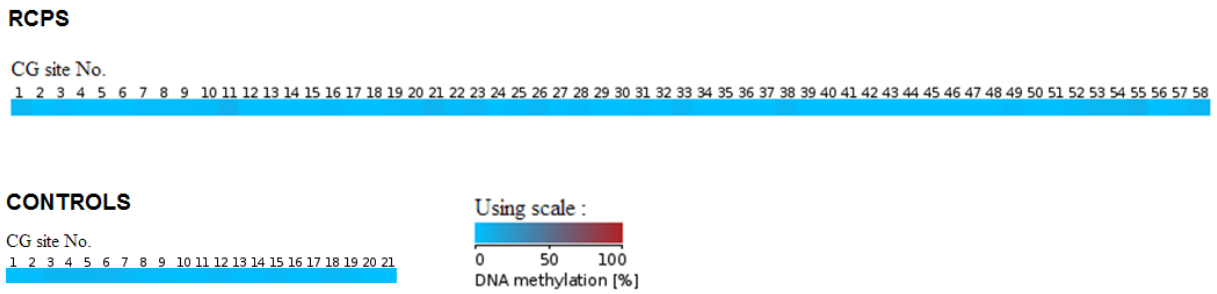


Figure 4: Methylation levels per CpG in RCPS and control samples revealing a high level of hypomethylation in all CpGs in both groups and no evident CpG methylation variation. The bluer, the more unmethylated a CpG is, while the redder, the more methylated a CpG is.

Discussion

Because repeat expansion at the 5'UTR region of *EIF4A3* are CpG dense, including an additional CpG in RCPS-causing expansions, we postulated that those expansions were submitted to hypermethylation leading to gene downregulation. Indeed, that is a mechanism reported for several repeat expansion diseases, including Fragile X Syndrome, Myotonic Dystrophy, and Spinocerebellar ataxias (NAGESHWARAN; FESTENSTEIN, 2015). As we have previously observed *EIF4A3* downregulation in whole-blood cells (FAVARO et al., 2014), we tested DNA methylation at the *EIF4A3* expansion region using this tissue in RCPS and non-expanded control samples.

Despite of our hypothesis, we observed no increased methylation at the *EIF4A3* expansion region in RCPS. On the contrary, we observed a small but significant methylation reduction in RCPS samples in comparison to controls. Once DNA hypomethylation at

promoter/5'UTR regions is associated to gene upregulation (SCHÜBELER, 2015), we refuted our hypothesis of DNA hypermethylation in the *EIF4A3* repeats as part of the molecular mechanism causing RCPS. Even looking at CpGs individually did not reveal any evident hypermethylation signal in our analysis. The observed methylation levels in the *EIF4A3* 5'UTR/promoter region both in RCPS and control samples are consistent with those reported for active genes, which are extensively highly hypomethylated (BAUBEC; SCHÜBELER, 2014; SCHÜBELER, 2015).

Once we refuted our hypothesis, we can speculate alternative scenarios by which the repeat expansions would lead to the observed *EIF4A3* downregulation. First, the expansion of the 20-nt CGCA repeat type could create bind sites for repressor proteins and act as a silencer. This would cause gene repression by directly interfering in the transcription machinery at the protein level, as reported in the study of gene regulation by such factors (COOPER, 2000; LODISH et al., 2000; OGBOURNE; ANTALIS, 1998; RILEY et al., 2008). Secondly, *EIF4A3* downregulation could be a result of post-transcriptional regulatory mechanisms. *EIF4A3* expanded repeats are transcript and could interfere in mRNA metabolism, including transport from nucleus to cytosol or even the interaction with translation factors, in which the interference of such processes has been reported (BURGUETE et al., 2015; FREIBAUM et al., 2015). Finally, the presence of those repeats in the final mRNA could also attract other RNA-binding proteins with cytotoxic effects, as also reported in Myotonic dystrophy in which MBNL1 proteins bind to DMPK RNA with expansions, leading to low or absent mRNA translation and also toxic effect of MBNL1 accumulation (LEE; COOPER, 2009).

In summary, we investigated the hypothesis of DNA hypermethylation in RCPS-causing expanded repeats in *EIF4A3* and observed no signals of hypermethylation. Alternative post-transcriptional regulatory mechanism may be involved in *EIF4A3* downregulation observed in RCPS.

References

BAUBEC, T.; SCHÜBELER, D. Genomic patterns and context specific interpretation of DNA methylation. **Current Opinion in Genetics & Development**, v. 25, p. 85-92, abr. 2014.

BURGUETE, A. S. et al. GGGGCC microsatellite RNA is neuritically localized, induces branching defects, and perturbs transport granule function. **eLife**, v. 4, p. e08881, 9 dez. 2015.

COOPER, G. M. Regulation of Transcription in Eukaryotes. 2000.

FAVARO, F. P. et al. Richieri-Costa-Pereira syndrome: a unique acrofacial dysostosis type. An overview of the Brazilian cases. **American Journal of Medical Genetics. Part A**, v. 155A, n. 2, p. 322-331, fev. 2011.

FAVARO, F. P. et al. A noncoding expansion in *EIF4A3* causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. **American Journal of Human Genetics**, v. 94, n. 1, p. 120-128, 2 jan. 2014.

FEIL, R.; FRAGA, M. F. Epigenetics and the environment: emerging patterns and implications. **Nature Reviews. Genetics**, v. 13, n. 2, p. 97-109, 4 jan. 2012.

FREIBAUM, B. D. et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. **Nature**, v. 525, n. 7567, p. 129-133, 3 set. 2015.

LEE, J. E.; COOPER, T. A. Pathogenic mechanisms of myotonic dystrophy. **Biochemical Society Transactions**, v. 37, n. Pt 6, p. 1281-1286, dez. 2009.

LI, L.-C.; DAHIYA, R. MethPrimer: designing primers for methylation PCRs. **Bioinformatics**, v. 18, n. 11, p. 1427-1431, nov. 2002.

LODISH, H. et al. Molecular Mechanisms of Eukaryotic Transcriptional Control. 2000.

MILLER, E. E. et al. *EIF4A3* deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira Syndrome. **Human Molecular Genetics**, 2 mar. 2017.

NAGESHWARAN, S.; FESTENSTEIN, R. Epigenetics and Triplet-Repeat Neurological Diseases. **Frontiers in neurology**, v. 6, p. 262, 21 dez. 2015.

OGBOURNE, S.; ANTALIS, T. M. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. **The Biochemical Journal**, v. 331 (Pt 1), p. 1-14, 1 abr. 1998.

RILEY, T. et al. Transcriptional control of human p53-regulated genes. **Nature Reviews. Molecular Cell Biology**, v. 9, n. 5, p. 402-412, maio 2008.

ROHDE, C. et al. BISMA--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. **BMC Bioinformatics**, v. 11, p. 230, 6 maio 2010.

SCHÜBELER, D. Function and information content of DNA methylation. **Nature**, v. 517, n. 7534, p. 321-326, 15 jan. 2015.

TAMMEN, S. A.; FRISO, S.; CHOI, S.-W. Epigenetics: the link between nature and nurture. **Molecular aspects of medicine**, v. 34, n. 4, p. 753-764, ago. 2013.

Chapter 8

General discussion and main conclusions

Because craniofacial development depends on a tight regulation of gene expression and gene expression depends on several regulatory mechanisms, from epigenetic changes to protein activation or inhibition (BELL et al., 2011; COULON et al., 2013; LEE; YOUNG, 2013; OGBOURNE; ANTALIS, 1998), disarrange in such regulation may lead to molecular, cellular and ultimately morphological phenotype. The current challenge is to identify which insults and genomes are more susceptible to unrepair those dysregulated processes thus resulting in an altered phenotype, so aetiology can be clarified and preventive strategies can be developed.

Attempting to identify dysregulated gene pathways in NSCL/P, we compared the transcriptome of NSCL/P and control cells and found the *BRCA1*-dependent DNA Damage repair pathway less expressed in patients. The validation of *BRCA1* downregulation as well as the quantification of DNA damage in those cells allowed us to suggest dysregulation in this pathway as a new susceptibility mechanism for NSCL/P. Indeed, after we published this work, some studies have also reported the dysregulation of DNA damage repair genes and the accumulation of DSB in NSCL/P cells and others found association of variants in such genes (BROOKLYIN et al., 2014; MOSTOWSKA et al., 2014). It is our belief, therefore, that the impairment of the DNA

damage repair pathways during craniofacial development would lead to compromised cellular processes and/or an increase in cell death, resulting in the incomplete growth and fusion of facial prominences and thus orofacial clefting.

Once we verified *BRCA1*-dependent DNA damage repair pathway downregulation in NSCL/P cells, we questioned whether epigenetic mechanisms would be responsible for the observed *BRCA1* expression reduction. Specifically, *BRCA1* promoter hypermethylation has been demonstrated as a molecular mechanism behind breast cancer tumourigenesis (BIRGISDOTTIR et al., 2006; BOSVIEL et al., 2012; IWAMOTO et al., 2011; RICE; MASSEY-BROWN; FUTSCHER, 1998), we thus postulated the similar mechanism in NSCL/P *BRCA1* downregulation. We were able to replicate our previous findings in *BRCA1* expression reduction and higher levels of DNA damage in NSCL/P in a new set of samples, suggesting us some generalisation of this mechanism. After evaluating methylation levels at the *BRCA1* promoter in NSCL/P and controls, we observed a small but significant increase of methylation in patients and also a distinct distribution of methylation among the analysed CpGs. We were also able to observe a significant inverse correlation between *BRCA1* and *IRF6* transcripts levels, suggesting the participation of *IRF6* in the pathway dysregulation. On the other hand, if *BRCA1* promoter methylation is indeed a cause for the gene downregulation in NSCL/P cells, at least two additional questions arise: first, what, or, more specifically, which environmental factor is driving differential methylation and, second, if there are DNA methylation changes at specific regions of the genome already associated with NSCL/P or even more broad epigenetic changes in a genomic level. In this work we verified the association of DNA methylation with NSCL/P in other regions, although not investigating the possible environmental factors leading to such alterations.

In order to investigate DNA methylation at NSCL/P GWAS associated regions, we analysed the flanking regions of rs987525, a SNP at 8q24.21, the most replicated GWAS NSCL/P associated region. After quantifying methylation levels in a 600bp region embedding the associated SNP in NSCL/P and controls, we observed no significant differences between groups

and no correlation of methylation levels with risk genotypes. We refuted therefore our hypothesis of a differential methylation in this region affecting its demonstrated long-range regulatory feature. Those results, however, do not invalidate the possibility of such mechanism occurring in a close but not analysed region nor even for other genomic regions.

We next questioned ourselves whether NSCL/P samples would exhibit a different epigenome in comparison to controls. We then performed methylome profiling in NSCL/P and controls and found several MVPs differentially methylated in regulatory regions of the genome and enriched in canonical pathways as WNT-Beta Catenin signaling and Regulation of Epithelial-Mesenchymal transition, already known of importance to orofacial clefts, and replicated those results in a different population. We were therefore able to answer the participation of epigenetic factors, in this case DNA methylation, in a genomic dimension and to establish in which set of genes those alterations are enriched. We speculate that environmental insults during development would lead to epigenetic changes as DNA methylation, which would persist until adult life. Because NSCL/P is a complex disease with still unexplained heritability, we believe that such alterations as DNA methylation can contribute to the missing heritability problem, which fits in the multifactorial scenario. Indeed, differential DNA methylation has been associated to some multifactorial diseases and such studies have been providing a bridge between “nature” and “nurture”, an old issue in biological sciences. Also, we have answered a very import question regarding phenotype penetrance in familial NSCL/P, which frequently displays incomplete penetrance. We showed a significant increase of promoter methylation in penetrant individuals from families segregating *CDH1* mutations, in comparison to non-penetrant individuals and controls. Such result is important for not only penetrance in NSCL/P, but could expanded for other NSCL/P *loci* with reported incomplete penetrance. Therefore, we have also demonstrated that DNA methylation can contribute to penetrance effects

Identifying genes and pathways for syndromic orofacial clefts is important not only for elucidating the molecular cause for a syndrome itself but also for clarifying general mechanisms

involved in craniofacial development that can be expanded for non-syndromic orofacial clefts, for example. In this manner, we studied the genetic basis of a very rare acrofacial dysostosis syndrome, Richieri-Costa Pereira syndrome (RCPS), which along with limb defects displays cleft palate and other craniofacial alterations. After unsuccessful results in exome analysis, we were able to identify by homozygosity mapping a region shared in all affected individuals and, by sequencing candidate genes, we found an expansion at the *EIF4A3* promoter/5'UTR consisting of a complex structure of 16 repeats in RCPS individuals. By studying gene expression in RCPS cells and animal modeling in zebrafish, we could demonstrate that those expansions result in a decrease of *EIF4A3* expression, which is therefore the most probable cause for the syndrome. More recently work from our group has better elucidate this mechanism by investigating *EIF4A3* conditional knock-out mice and RCPS IPSC (Induced Pluripotent Stem Cells) and cell differentiation, revealing that *EIF4A3* is essential for NCCs (Neural Crest Cells) migration, growth and differentiation into craniofacial tissues (MILLER et al., 2017). Our work identified the first craniofacial syndrome caused by an expansion, once repeat expansions have been extensively associated to neurodevelopmental and neurodegenerative disorders. Also, our group has been trying to identify the molecular mechanisms behind the origin of such expansions by genotyping a large cohort of non-affected individuals, in which we speculate unequal crossing-over as the most probable cause for the arise of such expansions.

Finally, we have also attempted to explore the role of DNA methylation in RCPS expansions, once they largely increased CpG number in that region. We demonstrated that DNA methylation is not a mechanism leading to *EIF4A3* downregulation and we suggest the participation of repressors, as in proteins interacting to those repeats, as plausible molecular mechanism involved in *EIF4A3* hypomorphism. Again, our group has been also investigating the role of such proteins by ChIP (Chromatin Immunoprecipitation) analysis and we believe that we will be able to also elucidate how those repeat expansions interfere in *EIF4A3* transcript levels.

Therefore, the main conclusions of this work are: *BRCA1*-dependent DNA damage repair pathway is altered in NSCL/P and such alteration is possibly driven by DNA methylation; NSCL/P displays a specific DNA methylation signature in important gene pathways for craniofacial development and DNA methylation can also contribute to penetrance in familial NSCL/P; RCPS is caused by a repeat expansion at *EIF4A3* promoter/5'UTR, however with no involvement of DNA methylation, but demonstrating that craniofacial malformation can be also caused by repeat expansions.

It is important to mention that this work, besides its relevance in the understanding of craniofacial development and malformation, has also resulted in scientific advances in the grounds of national and international collaboration and local technology development. For example, in the study of NSCL/P transcriptome we were able to collaborate with other institutes from our university and with researchers from the United Kingdom, as well as starting assays not applied in our laboratory before. In the case of methylation analysis, such epigenetic studies were developed for the first time in our laboratory and resulted in a collaboration with also researchers from the United Kingdom and the application of the technology in our centre thereafter, which will be of great importance for all future generations and work to be developed in our laboratory.

References

BELL, O. et al. Determinants and dynamics of genome accessibility. **Nature Reviews. Genetics**, v. 12, n. 8, p. 554-564, 1 ago. 2011.

BIRGISDOTTIR, V. et al. Epigenetic silencing and deletion of the BRCA1 gene in sporadic breast cancer. **Breast Cancer Research**, v. 8, n. 4, p. R38, 2006.

BOSVIEL, R. et al. BRCA1 promoter methylation in peripheral blood DNA was identified in sporadic breast cancer and controls. **Cancer epidemiology**, v. 36, n. 3, p. e177-82, jun. 2012.

BROOKLYIN, S. et al. Assessment of folic Acid and DNA damage in cleft lip and cleft palate. **Clinics and practice**, v. 4, n. 1, p. 608, 27 mar. 2014.

COULON, A. et al. Eukaryotic transcriptional dynamics: from single molecules to cell populations. **Nature Reviews. Genetics**, v. 14, n. 8, p. 572-584, ago. 2013.

IWAMOTO, T. et al. BRCA1 promoter methylation in peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. **Breast Cancer Research and Treatment**, v. 129, n. 1, p. 69-77, ago. 2011.

LEE, T. I.; YOUNG, R. A. Transcriptional regulation and its misregulation in disease. **Cell**, v. 152, n. 6, p. 1237-1251, 14 mar. 2013.

MILLER, E. E. et al. EIF4A3 deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira Syndrome. **Human Molecular Genetics**, 2 mar. 2017.

MOSTOWSKA, A. et al. Genetic variants in BRIP1 (BACH1) contribute to risk of nonsyndromic cleft lip with or without cleft palate. **Birth Defects Research. Part A, Clinical and Molecular Teratology**, v. 100, n. 9, p. 670-678, set. 2014.

OGBOURNE, S.; ANTALIS, T. M. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. **The Biochemical Journal**, v. 331 (Pt 1), p. 1-14, 1 abr. 1998.

RICE, J. C.; MASSEY-BROWN, K. S.; FUTSCHER, B. W. Aberrant methylation of the BRCA1 CpG island promoter is associated with decreased BRCA1 mRNA in sporadic breast cancer cells. **Oncogene**, v. 17, n. 14, p. 1807-1812, 8 out. 1998.

Chapter 9

Abstract

Craniofacial development is a tightly regulated event that requires expression of many genes at a precise space-temporal specificity. Interference in the regulation of such genes and their pathways is known to lead to abnormal phenotypes affecting the face and cranium. In this manner, regulation of these pathways is further complicated by interaction between genetic and environmental factors such that disturbance to either may result in craniofacial malformation, as orofacial clefts. Despite several at-risk *loci* have been identified, they do not completely explain the high heritability observed for the orofacial clefts and many questions remain open. For example, concerning the orofacial clefts transcriptome, the gene pathways which may be dysregulated and the affected cellular processes are still poorly understood. Further, if there is gene expression dysregulation in orofacial clefts, the causes leading to that need to be elucidated, such as the investigation of epigenetic factors. Also, since the multifactorial contribution makes environment relevant to this malformation, epigenetic and epigenomic differences in orofacial clefts should clarified. At last, rare syndromic forms of orofacial clefts with still unknown molecular cause and mechanisms should be elucidated in order to better understand craniofacial development and their impact in non-syndromic forms. Therefore, the main objective of this study was to investigate the molecular mechanisms involved in the aetiology of orofacial clefts, which was focused in gene expression and epigenetic analysis in non-syndromic cleft lip and/or palate (NSCL/P) as well as genetic, gene expression, animal modelling and epigenetics in Richieri-Costa-Pereira Syndrome (RCPS), a rare autosomal recessive syndromic form of orofacial cleft. We found significant transcriptome differences in NSCL/P in comparison to controls, revealing the *BRCA1*-dependent DNA damage repair pathway as compromised in NSCL/P cells

leading to DNA damage accumulation. Next, we studied the potential of DNA methylation in those cells and found a slight but significant increase of *BRCA1* promoter DNA methylation in NSCL/P cells and a distinct DNA methylation distribution, point to a possible epigenetic contribution in this phenomenon. We also evaluated the contribution of DNA methylation in 8q24.21 region, one of the most replicated regions in NSCL/P Genome-wide association studies and found no significant differences in our sample. Attempting to investigate DNA methylation in NSCL/P in an epigenomic level, we analysed methylomes and found 578 methylation variable positions in NSCL/P, highly enriched in regulatory regions and in relevant gene pathways for craniofacial development as Epithelial-Mesenchymal Transition pathway. We also studied effect of DNA methylation in familial NSCL/P displaying incomplete penentrance and found a significant increase of *CDH1* promoter hypermethylation in penetrant cases in comparison to non-penetrants. Finally, by the use of different sequencing strategies and identity-by-descent analysis we mapped the mutation region of RCPS to *EIF4A3* 5'UTR/promoter and found a complex structure of expanded repeats in RCPS patients leading to *EIF4A3* downregulation. We were also able to validate the phenotypes using an animal modelling strategy in zebrafish. Because those repeats are CG rich, we investigated whether they were submitted to DNA hypermethylation in RCPS patients as a cause for *EIF4A3* hypomorphism, however we found no evidence of methylation increase in RCPS. In conclusion, we were able to associate dysregulated pathways to NSCL/P susceptibility and DNA methylation differences to both non-familial and familial NSCLP. Besides, we were able to identify the genetic cause of RCPS, which now can be molecularly diagnosed. Altogether, our results add to the understanding of craniofacial development and the aetiology of orofacial clefts.

Key-words: craniofacial development, cleft lip and/or palate, transcriptomics, epigenomics, homozygosity mapping, cellular and animal models.

Resumo

O desenvolvimento craniofacial é um evento finamente regulado que requer a expressão de muitos genes em uma precisão espaço-temporal específica. A interferência na regulação de tais genes e suas respectivas vias é sabidamente causadora de fenótipos que afetam a face e o crânio. Neste sentido, a regulação destas vias é decorrente da interação entre fatores genéticos e ambientais, de tal forma que a perturbação de quaisquer destes fatores pode resultar em malformações craniofaciais, como as fissuras orofaciais. Apesar dos muitos *loci* de risco já identificados, estes não explicam completamente a alta herdabilidade observadas nas fissuras orofaciais e muitas questões permanecem em aberto. Por exemplo, em relação ao transcriptoma em fissuras orofaciais, as vias genéticas que podem estar desreguladas, assim como processos celulares afetados em decorrência, são ainda pouco compreendidos. Além disso, se há desregulação na expressão de genes em fissuras orofaciais, as causas que levam a essas diferenças necessitam ser elucidadas, como, por exemplo, por meio da investigação de fatores epigenéticos. Também, uma vez que o componente multifatorial torna a influência do ambiente relevante para esta malformação, diferenças epigenéticas e epigenômicas nas fissuras orofaciais devem ser melhor compreendidas. Por fim, formas raras e sindrômicas de fissuras orofaciais sem elucidação de causa moleculares devem ser estudadas para que melhor se compreenda o desenvolvimento craniofacial e o impacto destes mecanismos moleculares em formas não-sindrômicas. Portanto, nosso objetivo principal neste estudo foi investigar os mecanismos moleculares envolvidos na etiologia das fissuras orofaciais, com o foco na análise de expressão gênica e epigenètica em fissuras de lábio-palatinas não-sindrômicas (FL/P NS) e também o estudo genético, de expressão gênica, modelagem animal e epigenética na Síndrome de Richieri-Costa-Pereira (RCPS), uma forma sindrômica e autossômica recessiva de fissura orofacial. Nós encontramos diferenças significantes no transcriptoma de FL/P NS em comparação com controles, que revelaram o

comprometimento da via do *BRCA1* no reparo ao dano de DNA e o acúmulo de dano de DNA em células FL/P NS. Em seguida, nós estudamos o potencial da metilação de DNA nestas células e encontramos um pequeno, porém significante, aumento de metilação de DNA no promotor do *BRCA1* e uma distribuição diferente de metilação, apontando para uma possível contribuição epigenética na desregulação do gene. Nós também avaliamos a contribuição da metilação de DNA na região 8q24.21, uma das mais associadas às FL/P NS por meio de *Genome-wide association studies*, porém não encontramos diferenças significantes na nossa amostra. Com o intuito de investigar a metilação de DNA em FL/P NS em uma escala epigenômica, nós analisamos o perfil de metilomas e encontramos 578 sítios diferencialmente metilados nas FL/P NS, altamente enriquecidos em regiões regulatórias e em vias relevantes para o desenvolvimento craniofacial como a via de Transição Epitélio-Mesenquimal. Nós também estudamos o efeito da metilação de DNA em casos famílias de FL/P NS com penetrância incompleta e encontramos um aumento significativo de metilação do promotor do *CDH1* nos casos penetrantes em comparação aos não-penetrantes. Por último, por meio de diferentes estratégias de sequenciamento e análise de segregação de haplótipos nós mapeamos a mutação de RCPS na região 5'UTR/promotor do *EIF4A3* e encontramos uma estrutura complexa de expansão de repetições nos pacientes RCPS, ocasionando a diminuição da expressão do *EIF4A3*. Nós também reproduzimos fenótipos comparáveis aos da RCPS por meio de modelo animal em *zebrafish*. Uma vez que tais repetições são ricas em CG, nós investigamos se estas poderiam ser submetidas à metilação de DNA em pacientes RCPS como uma causa para a redução dos transcritos do *EIF4A3*, porém não encontramos evidências de aumento de metilação em RCPS. Em conclusão, nós conseguimos associar vias gênicas desreguladas à susceptibilidade para as FL/P NS e diferenças de metilação de DNA tanto em casos familiais como não-familiais de FL/P NS. Além disso, identificamos a causa genética de RCPS, sendo que a síndrome pode ser agora diagnosticada molecularmente. Em conjunto, nossos resultados adicionam ao conhecimento do desenvolvimento craniofacial e na etiologia das fissuras orofaciais.

Palavras chave: desenvolvimento craniofacial, fissura de lábio e/ou palato, transcriptoma, epigenoma, mapeamento por homozigose, modelos celulares e animais.

Appendix: Additional Publications

I- Griesi-Oliveira, K., Sunaga, D.Y., **ALVIZI, L.**, Vadasz, E., and Passos-Bueno, M.R. (2013). Stem cells as a good tool to investigate dysregulated biological systems in autism spectrum disorders. ***Autism Research*** 6, 354–361.

SHORT REPORT

Stem Cells as a Good Tool to Investigate Dysregulated Biological Systems in Autism Spectrum Disorders

Karina Griesi-Oliveira, Daniele Yumi Sunaga, Lucas Alvizi, Estevão Vadasz, and Maria Rita Passos-Bueno

Identification of the causes of autism spectrum disorders (ASDs) is hampered by their genetic heterogeneity; however, the different genetic alterations leading to ASD seem to be implicated in the disturbance of common molecular pathways or biological processes. In this scenario, the search for differentially expressed genes (DEGs) between ASD patients and controls is a good alternative to identify the molecular etiology of such disorders. Here, we employed genome-wide expression analysis to compare the transcriptome of stem cells of human exfoliated deciduous teeth (SHEDs) of idiopathic autistic patients ($n = 7$) and control samples ($n = 6$). Nearly half of the 683 identified DEGs are expressed in the brain ($P = 0.003$), and a significant number of them are involved in mechanisms previously associated with ASD such as protein synthesis, cytoskeleton regulation, cellular adhesion and alternative splicing, which validate the use of SHEDs to disentangle the causes of autism. Autistic patients also presented overexpression of genes regulated by androgen receptor (AR), and AR itself, which in turn interacts with *CHD8* (chromodomain helicase DNA binding protein 8), a gene recently shown to be associated with the cause of autism and found to be upregulated in some patients tested here. These data provide a rationale for the mechanisms through which *CHD8* leads to these diseases. In summary, our results suggest that ASD share deregulated pathways and revealed that SHEDs represent an alternative cell source to be used in the understanding of the biological mechanisms involved in the etiology of ASD. *Autism Res 2013, 6: 354–361.* © 2013 International Society for Autism Research, Wiley Periodicals, Inc.

Keywords: expression studies; androgen signaling; CHD8; stem cells of human exfoliated deciduous teeth

Introduction

Autism spectrum disorders (ASDs), characterized by impairments in communication and social skills and stereotyped behavior, have a strong genetic component. Their genetic architecture is still debated with clear evidence of allelic and locus heterogeneity [Betancur, 2011; El-Fishawy & State, 2010].

Transcriptome analysis in ASD has shown deregulation of genes and pathways related to neurological functions and central nervous system development [Baron, Liu, Hicks, & Gregg, 2006; Gregg et al., 2008; Hu, Frank, Heine, Lee, & Quackenbush, 2006; Hu et al., 2009a,b; Seno et al., 2010; Voineagu et al., 2011]. Most of these studies were carried out using lymphoblasts, lymphoblastoid cell lines (LCLs) or brain tissue. Lymphoblasts and LCL have shown several limitations due to their embryonic origin and cell differentiation degree. On the other hand, expression studies using postmortem brain tissue are hampered mainly by the potential biases from postmortem effects and the relatively small sample sizes. Therefore, it is of value the identification of other cell sources to study ASD.

Stem cells of human exfoliated deciduous teeth (SHEDs) are an easily accessible cell source, with an ectodermic origin and expression of neuronal markers upon differentiation, which makes them functionally and embryonically more related to nervous tissue cells than the previously used lymphocytes and LCL [d'Aquino et al., 2009; Miura et al., 2003]. To investigate the applicability of SHEDs to study the pathogenesis of ASD, we have asked if ASD candidate genes and biological processes/pathways involved in neuronal development and function were enriched upon functional annotation analysis of the differentially expressed genes identified. Next, we detected deregulation of the androgen signaling pathway. The relationship between androgen signaling and chromodomain helicase DNA binding protein 8 (*CHD8*), a candidate gene recently identified in ASD exome sequencing studies [Neale et al., 2012; O'Roak et al., 2012a,b; Sanders et al., 2012; Talkowski et al., 2012], is discussed.

From the Centro de Estudos do Genoma Humano, Departamento de Biologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil (K.G.-O., D.Y.S., L.A.C., M.R.P.-B.); Instituto de Psiquiatria do Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brasil (E.V.)

Received July 11, 2012; accepted for publication April 10, 2013

Address for correspondence and reprints: Maria Rita Passos-Bueno, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, 277, sala 200, São Paulo—SP 05508-090, Brazil. E-mail: passos@ib.usp.br

Published online 25 June 2013 in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/aur.1296

In this study we investigate the potential use of SHED to investigate autism transcriptome and we found several differentially expressed genes in relevant pathways to autism spectrum disorders.

II- Brito, L.A., Paranaiba, L.M.R., Bassi, C.F.S., Masotti, C., Malcher, C., Schlesinger, D., Rocha, K.M., **CRUZ, L.A**., Bárbara, L.K., Alonso, N., et al. (2012). Region 8q24 is a susceptibility locus for nonsyndromic oral clefting in Brazil. ***Birth Defects*** Res Part A Clin Mol Teratol 94, 464–468.

Birth Defects Research (Part A) 94:464–468 (2012)

Region 8q24 Is a Susceptibility Locus for Nonsyndromic Oral Clefting in Brazil

Luciano Abreu Brito,[1] Lívia Máris Ribeiro Paranaiba,[2,3] Camila Fernandes Silva Bassi,[1] Cibele Masotti,[1] Carolina Malcher,[1] David Schlesinger,[1,4] Katia Maria Rocha,[1] Lucas Alvizi Cruz,[1] Lígia Kobayashi Bárbara,[1] Nivaldo Alonso,[5] Diogo Franco,[6] Elizabete Bagordakis,[2] Hercílio Martelli Jr.,[3,7] Diogo Meyer,[8] Ricardo D. Coletta,[2] and Maria Rita Passos-Bueno[1]*

[1]Human Genome Research Center, Institute of Biosciences, University of São Paulo, São Paulo, Brazil
[2]Department of Oral Diagnosis, School of Dentistry, State University of Campinas, Piracicaba, São Paulo, Brazil
[3]Stomatology Clinic, Dental School, State University of Montes Claros, Montes Claros, Minas Gerais, Brazil
[4]Instituto do Cérebro, Instituto Israelita de Ensino e Pesquisa Albert Einstein, São Paulo, Brazil
[5]Department of Surgery, School of Medicine, University of São Paulo, São Paulo, Brazil
[6]Department of Plastic Surgery, University of Rio de Janeiro, Rio de Janeiro, Brazil
[7]Center for Rehabilitation of Craniofacial Anomalies, Dental School, University of Alfenas, Minas Gerais, Brazil
[8]Department of Genetics, Institute of Biosciences, University of São Paulo, São Paulo, Brazil

Received 28 January 2012; Revised 2 March 2012; Accepted 6 March 2012

BACKGROUND: Nonsyndromic cleft lip with or without cleft palate is a relatively common craniofacial defect with multifactorial inheritance. The association of the rs987525 single nucleotide variant, located in a gene desert at 8q24.21 region, has been consistently replicated in European populations. We performed a structured association approach combined with transcriptional analysis of the *MYC* gene to dissect the role of rs987525 in oral clefting susceptibility in the ethnically admixed Brazilian population. **METHODS**: We performed the association study conditioned on the individual ancestry proportions in a sample of 563 patients and 336 controls, and in an independent sample of 221 patients and 261 controls. The correlation between rs987525 genotypes and *MYC* transcriptional levels in orbicularis oris muscle mesenchymal stem cells was also investigated in 42 patients and 4 controls. **RESULTS**: We found a significant association in the larger sample ($p = 0.0016$; OR = 1.80 [95% confidence interval {CI}, 1.21–2.69], for heterozygous genotype, and 2.71 [95% CI, 1.47–4.96] for homozygous genotype). We did not find a significant correlation between rs987525 genotypes and *MYC* transcriptional levels ($p = 0.14$; $r = -0.22$, Spearman Correlation). **CONCLUSIONS**: We present a positive association of rs987525 in the Brazilian population for the first time, and it is likely that the European contribution to our population is driving this association. We also cannot discard a role of rs987515 in *MYC* regulation, because this locus behaves as an expression quantitative locus of *MYC* in another tissue. *Birth Defects Research (Part A) 94:464–468, 2012.* © 2012 Wiley Periodicals, Inc.

Key words: rs987525; rs1476165; rs2099897; heritability; cleft lip; cleft palate; 8q24 gene desert; *MYC* transcriptional levels; common disease-common variant; mesenchymal stem cells

INTRODUCTION

Nonsyndromic cleft lip with or without cleft palate (NS CL/P; OMIM 119530), which represents 70% of orofacial clefts and is the most common craniofacial anomaly, is always accompanied by physical and psychological impairments to affected children, with great family burden (Wehby and Cassell, 2010).

NS CL/P has a worldwide birth prevalence of approximately 1:700 live births (Dixon et al., 2011), varying according to ethnicity (African, 0.3:1000; European, 1.0:1000; Native American, 3.6:1000), geographic origin,

Additional Supporting Information may be found in the online version of this article.

Financial support from FAPESP/CEPID (98/14254-2), CNPq (401952/2010-8) and the Brazilian Ministry of Health.

*Correspondence to: Maria Rita Passos-Bueno, Rua do Matão, 277, sala 200, Cidade Universitária, São Paulo, SP 05508-090, Brazil. E-mail: passos@ib.usp.br

Published online 18 April 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/bdra.23011

In this study we performed an association study with SNP rs987525 (8q24.21) in a Brazilian cohort and found positive association for this region with NSCL/P, likely driven by the European contribution to our population.

III- Fanganiello, R.D., Ishiy, F.A.A., Kobayashi, G.S., **ALVIZI, L.**, Sunaga, D.Y., and Passos-Bueno, M.R. (2015). Increased In Vitro Osteopotential in SHED Associated with Higher IGF2 Expression When Compared with hASCs. ***Stem Cells Reviews and Reports*** *11*, 635–644.

Stem Cell Rev and Rep (2015) 11:635–644
DOI 10.1007/s12015-015-9592-x

Increased In Vitro Osteopotential in SHED Associated with Higher *IGF2* Expression When Compared with hASCs

Roberto Dalto Fanganiello[1] · Felipe Augusto Andre Ishiy[1] · Gerson Shigeru Kobayashi[1] · Lucas Alvizi[1] · Daniele Yumi Sunaga[1] · Maria Rita Passos-Bueno[1]

Published online: 1 May 2015

Abstract Mesenchymal stem cell (MSC) osteogenic differentiation potential varies according to factors such as tissue source and cell population heterogeneity. Pre-selection of cell subpopulations harboring higher osteopotential is a promising strategy to achieve a thorough translation of MSC-based therapies to the clinic. Here, we searched for novel molecular markers predictive of osteopotential by comparing MSC populations from two sources harboring different osteogenic potentials. We show that MSCs from human deciduous teeth (SHED) have an intrinsically higher osteogenic potential when compared with MSCs from human adipose tissue (hASCs) under the same in vitro controlled induction system. Transcriptome profiling revealed *IGF2* to be one of the top upregulated transcripts before and during early in vitro osteogenic differentiation. Further, exogenous IGF2 supplementation enhanced alkaline phosphatase activity and matrix mineralization, and inhibition of IGF2 lessened these parameters in SHED and hASCs, validating *IGF2* as an osteogenic factor in these MSCs. Further, we found *IGF2* to be biallelically expressed in SHED, but not in hASCs. We observed a 4 % methylation increase in the imprinting control region within the *IGF2-H19* locus in SHED, and this is mainly due to 2 specific CpG sites. Thus, we suggest that *IGF2* upregulation in SHED is due to loss of imprinting. This study unravels osteogenic properties in SHED, implying *IGF2* as a potential biomarker of MSCs with higher osteopotential, and unveils *IGF2* loss-of-imprinting in SHED.

Keywords SHED · hASCs · In vitro osteopotential · *IGF2* · Imprinting · Biallelic expression

Electronic supplementary material The online version of this article (doi:10.1007/s12015-015-9592-x) contains supplementary material, which is available to authorized users.

✉ Roberto Dalto Fanganiello
robertofanganiello@gmail.com

[1] Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, Rua do Matao, 277, sala 200, Sao Paulo, SP, Brazil 05508-090

Introduction

Studies on the use of mesenchymal stem cells (MSCs) for cellular therapies with the ultimate goal of bone regeneration are mainly based on their capacity of self-renewal, direct differentiation toward the osteogenic lineage and the ability to secrete trophic factors that may contribute to bone repair [1, 2]. These characteristics have been firstly explored in bone marrow mesenchymal stem cells (BMSCs) both in the lab and in different clinical scenarios [3–5]. However, BMSCs are scarce, and bone marrow harvesting remains painful and laborious. Therefore, there is an increasing interest in studying osteogenic cells from more convenient tissue sources, such as dental pulp and adipose tissue.

Stem cells from human exfoliated deciduous teeth (SHED), a specific type of dental pulp stem cell (DPSC), are potentially useful due to their differentiation plasticity, noninvasive isolation process, and similarity to osteoprogenitor cells [6–9]. SHED have been used to repair cranial defects in immunocompromised mice [10], to reconstruct trabecular bone in systemic lupus erythematosus-like MRL/lpr mice [8], and to aid critical-size calvarial defect regeneration in Wistar rats [11]. Human adipose tissue-derived stem cells (hASCs) represent another promising MSC type since they can be harvested from the stromal-vascular fraction of adipose tissue in abundant quantities by a simple surgical procedure. Their osteogenic potential has been demonstrated by the use of different culture

In this study we investigated the molecular mechanisms behind a sub-population of mesenchymal stem cells with high osteogenic potential and by the use of gene expression analysis we found IGF2 to be upregulated. Next we demonstrated a gain of methylation in the IGF2-H19 imprinting control region, possibly driving this upregulation.

IV- Brito, L.A., Yamamoto, G.L., Melo, S., Malcher, C., Ferreira, S.G., Figueiredo, J., **ALVIZI, L**., Kobayashi, G.S., Naslavsky, M.S., Alonso, N., et al. (2015). Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate. ***Human Mutation*** 36, 1029–1033.

BRIEF REPORT

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate

Luciano Abreu Brito,[1] Guilherme Lopes Yamamoto,[1] Soraia Melo,[2,3] Carolina Malcher,[1] Simone Gomes Ferreira,[1] Joana Figueiredo,[2,3] Lucas Alvizi,[1] Gerson Shigeru Kobayashi,[1] Michel Satya Naslavsky,[1] Nivaldo Alonso,[4] Temis Maria Felix,[5] Mayana Zatz,[1] Raquel Seruca,[2,3,6] and Maria Rita Passos-Bueno[1]*

[1]*Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil;* [2]*IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal;* [3]*Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal;* [4]*Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brasil;* [5]*Laboratório de Medicina Genômica, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brasil;* [6]*Faculty of Medicine, University of Porto, Porto, Portugal*

Communicated by Iain McIntosh

Received 2 February 2015; accepted revised manuscript 17 June 2015.
Published online 29 June 2015 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22827

ABSTRACT: Nonsyndromic orofacial cleft (NSOFC) is a complex disease of still unclear genetic etiology. To investigate the contribution of rare epithelial cadherin (*CDH1*) gene variants to NSOFC, we target sequenced 221 probands. Candidate variants were evaluated via in vitro, in silico, or segregation analyses. Three probably pathogenic variants (c.760G>A [p.Asp254Asn], c.1023T>G [p.Tyr341*], and c.2351G>A [p.Arg784His]) segregated according to autosomal dominant inheritance in four nonsyndromic cleft lip with or without cleft palate (NSCL/P) families (Lod score: 5.8 at θ = 0; 47% penetrance). A fourth possibly pathogenic variant (c.387+5G>A) was also found, but further functional analyses are needed (overall prevalence of *CDH1* candidate variants: 2%; 15.4% among familial cases). *CDH1* mutational burden was higher among probands from familial cases when compared to that of controls ($P = 0.002$). We concluded that *CDH1* contributes to NSCL/P with mainly rare, moderately penetrant variants, and *CDH1* haploinsufficiency is the likely etiological mechanism.
Hum Mutat 36:1029–1033, 2015. © 2015 Wiley Periodicals, Inc.

KEY WORDS: *CDH1*; oral clefts; gastric cancer; two-hit model; rare variant

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) and nonsyndromic cleft palate only (NSCPO) are two complex disorders within the nonsyndromic orofacial cleft (NSOFC) spectrum [Gorlin et al., 2001]. While the genetic etiology of NSCPO is largely unclear, genetic *loci* have been systematically implicated in NSCL/P, such as common low-risk 8q24, 10q25, and *IRF6* variants [Rahimov et al., 2008; Birnbaum et al., 2009; Mangold et al., 2010, Brito et al., 2012aa, 2012b]. However, given the high heritability attributed to NSCL/P [Hu et al., 1982; Calzolari et al., 1988; Brito et al., 2011], searching for alternative genetic variants or mechanisms is necessary to bridge the missing heritability gap of these malformations.

Rare germline variants in the gene encoding the adhesion molecule epithelial cadherin, *CDH1* (MIM# 192090), have long been associated with diffuse gastric cancer and lobular breast cancer [van Roy and Berx, 2008]. Most recently, *CDH1* mutations have been reported in OFC patients in association with gastric cancer [Frebourg et al., 2006; Kluijt, et al. 2012; Benusiglio et al., 2013] or not [Vogelaar et al., 2013; Bureau et al., 2014]. These findings raise the questions as to what the proportion of NSOFC cases underlain by *CDH1* variants and their attributed penetrance is, and which types of mutations or mechanisms lead to OFC, cancer, or both phenotypes.

Here, we performed a variant screening for *CDH1* (NM_004360.3) coding region in 221 NSOFC probands (affected by NSCL/P [$n = 189$] or NSCPO [$n = 32$], either from nonfamilial [$n = 138$] or familial cases [$n = 83$]; Supp. Table S1). Sequencing was performed by using next-generation sequencing (NGS, exome or targeted gene sequencing) and Sanger sequencing (SS), and applied for 65 and 156 probands, respectively. Additional NGS or SS was performed for extra members of familial cases, when available. When exome sequencing was performed in affected members of the same family, we filtered out variants with minor allele frequency greater than 1% in public databases (1000 Genomes Project, and NHLBI ESP exomes) and in our in-house database of 609 Brazilian control exomes (Supp. Methods). Among the 221 probands, we identified a total of 47 variants, of which 12 were absent in our controls (two missense, one nonsense, and nine noncoding or synonymous variants; Supp. Table S2). Variants were submitted to the LOVD database (at http://www.lovd.nl/CDH1).

The novel missense variant c.760G>A (p.Asp254Asn, exon 6) was the most likely causative variant among the main candidates detected by exome analysis (mean coverage of 60×; average of 25,140 variants called for each individual; Supp. Table S3) in families F3788

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Maria Rita dos Santos e Passos-Bueno, Rua do Matão 277, sala 200, Cidade Universitária, São Paulo, SP, 05508-090, Brasil. Tel.: 5511 3091-7740, E-mail: passos@ib.usp.br

Contract grant sponsor(s): Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; CEPID Project 2013/08028-1); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; 401952/2010-0); Fundos FEDER através do Programa Operacional Factores de Competitividade (COMPETE); Portuguese Foundation for Science and Technology (FCT; PTDC/SAU-ONC/110294/2009, PTDC/BIM-ONC/0171/2012, PEst-OE/EEI/LA0009/2013, and JF post-doctoral grant SFRH/BPD/87705/2012).

In this study we investigate the contribution of *CDH1* variants in 221 probands and found potential damaging variants in three families. We also observed a significant mutational burden of *CDH1* variants in familial *CDH1* in comparison to controls.

V- Savastano, C.P., Brito, L.A., Faria, Á.C., Setó-Salvia, N., Peskett, E., Musso, C.M., **ALVIZI, L.**, Ezquina, S.A.M., James, C., GOSgene, et al. (2017). Impact of rare variants in ARHGAP29 to the etiology of oral clefts: role of loss-of-function vs missense variants. ***Clinical Genetics*** 91, 683–689.

CLINICAL GENETICS An International Journal of Genetics, Molecular and Personalized Medicine

Clin Genet 2017: 91: 683–689
Printed in Singapore. All rights reserved

CLINICAL GENETICS
doi: 10.1111/cge.12823

Original Article

Impact of rare variants in *ARHGAP29* to the etiology of oral clefts: role of loss-of-function *vs* missense variants

Savastano C.P., Brito L.A., Faria Á.C., Setó-Salvia N., Peskett E., Musso C.M., Alvizi L., Ezquina S.A.M., James C., GOSgene , Beales P., Lees M., Moore G.E., Stanier P., Passos-Bueno M.R.. Impact of rare variants in *ARHGAP29* to the etiology of oral clefts: role of loss-of-function *vs* missense variants.
Clin Genet 2017: 91: 683–689. © John Wiley & Sons A/S. Published by John Wiley & Sons Ltd, 2016

Non-syndromic cleft lip with or without cleft palate (NSCL/P) is a prevalent, complex congenital malformation. Genome-wide association studies (GWAS) on NSCL/P have consistently identified association for the 1p22 region, in which *ARHGAP29* has emerged as the main candidate gene. *ARHGAP29* re-sequencing studies in NSCL/P patients have identified rare variants; however, their clinical impact is still unclear. In this study we identified 10 rare variants in *ARHGAP29*, including five missense, one in-frame deletion, and four loss-of-function (LoF) variants, in a cohort of 188 familial NSCL/P cases. A significant mutational burden was found for LoF (Sequence Kernel Association Test, $p = 0.0005$) but not for missense variants in *ARHGAP29*, suggesting that only LoF variants contribute to the etiology of NSCL/P. Penetrance was estimated as 59%, indicating that heterozygous LoF variants in *ARHGAP29* confer a moderate risk to NSCL/P. The GWAS hits in *IRF6* (rs642961) and 1p22 (rs560426 and rs4147811) do not seem to contribute to the penetrance of the phenotype, based on co-segregation analysis. Our data show that rare variants leading to haploinsufficiency of *ARHGAP29* represent an important etiological clefting mechanism, and genetic testing for this gene might be taken into consideration in genetic counseling of familial cases.

Conflict of interest

The authors have declared no conflicting interests.

C.P. Savastano[a], L.A. Brito[a], Á.C. Faria[a], N. Setó-Salvia[b,c], E. Peskett[b], C.M. Musso[a], L. Alvizi[a], S.A.M. Ezquina[a], C. James[b], GOSgene[b], P. Beales[b], M. Lees[d], G.E. Moore[b], P. Stanier[b] and M.R. Passos-Bueno[a]

[a]Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil, [b]Genetics and Genomic Medicine, UCL Institute of Child Health, London, UK, [c]Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK, and [d]Department of Clinical Genetics, Great Ormond Street Children's Hospital, London, UK

Key words: cleft lip and palate – GWAS – haploinsufficiency – IRF6 – nonsense mutations – penetrance – rare variants – 1p22

Corresponding authors: Maria Rita Passos-Bueno, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências da USP, Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, Universidade de São Paulo, Rua do Matão 277, sala 200, Cidade Universitária, São Paulo, SP 05508-090, Brazil.
Tel.: +55 11 3091 7740
e-mail: passos@ib.usp.br
and
Philip Stanier, Genetics and Genomic Medicine, UCL Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.
Tel.: +44 207 905 2867;
Fax: +44 (0) 207 905 2953
e-mail: p.stanier@ucl.ac.uk

Received 3 May 2016, revised and accepted for publication 16 June 2016

In this study we investigated the impact of rare variants in ARHGAP29 in familial NSCL/P cases and identified 10 rare variants in our cohort. Also, we observed a significant mutational burden for ARHGAP29 loss-of-function variants in NSCL/P.