Gerson Shigeru Kobayashi

Análise do transcriptoma de células-tronco mesenquimais para o estudo da etiologia das fissuras lábio-palatinas não-sindrômicas

Transcriptome analysis of mesenchymal stem cells to investigate the aetiology of non-syndromic cleft lip and palate

Instituto de Biociências
Universidade de São Paulo
São Paulo
2011

Gerson Shigeru Kobayashi

Análise do transcriptoma de células-tronco mesenquimais para o estudo da etiologia das fissuras lábio-palatinas não-sindrômicas

Transcriptome analysis of mesenchymal stem cells to investigate the aetiology of non-syndromic cleft lip and palate

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

Orientadora: Maria Rita dos Santos e Passos-Bueno

São Paulo

2011

Ficha Catalográfica

Kobayashi, Gerson

Análise do transcriptoma de células-tronco mesenquimais para o estudo da etiologia das fissuras lábio-palatinas não-sindrômicas
93 páginas

Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Fissura lábio-palatina 2. Expressão gênica 3. células-tronco mesenquimais. I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

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

Orientador

“ Never be bullied into silence.
Never allow yourself to be made a victim.
Accept no one's definition of your life;
define yourself. ”

—Harvey Fierstein

AGRADECIMENTOS

À Rita, que além de ser uma orientadora e pesquisadora excepcional, sempre cuidou de todos nós. Seu empenho em relação aos alunos vai muito além da ciência. Obrigado pela inspiração, pelo carinho, e pelo aprendizado.

À Dani Bueno, que me iniciou no estudo das fissuras e na cultura de células. Obrigado pela amizade, e pelo companheirismo tanto nos momentos de stress e fadiga, como também nos momentos de diversão.

A todos os colegas de laboratório, pelos bons momentos, pelas cervejas, pelo auxílio e incentivo. À Erika Yeh e à Meire, pelos ensinamentos e ajuda nos momentos difíceis. Aprendi muito com vocês! À Dani Yumi, por toda a ajuda com a bioinformática, pelas conversas, discussões, e risadas. Ao Lucas, novo companheiro de projeto, cuja paixão pela ciência é contagiante. Ao Luciano, pelo sincero coleguismo. Ao Bruno, Felipe, Simone e Larissa, pelo auxílio com as culturas celulares, citometria, e arrays. À Karina, pelas longas discussões e reflexões. À minha amiga May, pela companhia e diversão. Ao pessoal do Genoma e da sala 211, e à Nati, pela paciência e ajuda. Agradeço também à Constância e a todos os funcionários da Genética e do Genoma. Obrigado por tudo.

Agradeço aos velhos e novos amigos: Andrews, Rafa, Elaine, Alexandre, Fábio, Leo, Mel, Emiliano, Maneco, Carol, Eduardo, Renata, Cris, Mari, Sofia, e muitos outros, cuja amizade foi e sempre será incondicional. Em especial, agradeço à Camila, pela grande amizade e apoio. Sem vocês por perto não sei se estaria aqui hoje, e estou certo de que sempre poderei contar com vocês.

Aos meus pais, sem os quais não teria a oportunidade de me dedicar aos estudos. Seu apoio foi crucial para minha formação.

Agradeço também a todos os pacientes e doadores de dentes, sem os quais não haveria motivo ou meios para a realização do trabalho.

E finalmente, agradeço à FAPESP/CEPID, PROAP, CNPq e MCT, pelo apoio financeiro.

ÍNDICE

NOTAS

Esta dissertação de Mestrado compreende um trabalho inédito desenvolvido durante os anos de 2008 a 2011 no Laboratório de Genética do Desenvolvimento do Centro de Estudos do Genoma Humano do Instituto de Biociências da Universidade de São Paulo (CEGH/IB-USP).

A apresentação do trabalho seguiu o modelo de divisão em capítulos/artigos, modelo este aceito atualmente pelos programas de pós-graduação do Instituto de Biociências. Dessa maneira, são apresentados apenas os estudos que resultaram em um artigo científico publicado em 2010, e outro que será submetido em breve. Por esse motivo, optei por redigí-lo em inglês. As seções referentes à introdução, metodologia e discussão geral estão escritas em português, conforme especificado pelas regras deste Instituto.

Embora o novo acordo ortográfico da língua portuguesa esteja em vigência desde Janeiro de 2009, ficou estabelecido um período de transição até 2012, durante o qual são válidas tanto a regra nova quanto a antiga. Sendo assim, optei por utilizar as normas antigas, mantendo o uso de crases, das regras de acentuação anteriores e uso de trema.

As referências bibliográficas referentes a cada um dos artigos estão relacionadas ao final dos mesmos. Já as referências bibliográficas que constam na introdução, metodologia e discussão encontram-se ao final da tese. A norma bibliográfica escolhida foi a estipulada pela ABNT, exceto para o artigo já publicado.

Visto que cada artigo científico possui descrição de suas metodologias, o capítulo que se refere à metodologia geral se aterá apenas a uma breve explicação acerca de procedimentos não detalhados nos artigos.

RESUMO

A fissura lábio-palatina não-sindrômica (FLP NS) é uma doença multifatorial, determinada pela interação entre fatores genéticos e ambientais e sua incidência é estimada entre 0,34 e 2,29 a cada 1000 nascimentos. Trata-se de uma embriopatia causada por erros durante a morfogênese orofacial, a qual depende de uma fina regulação de mecanismos como proliferação celular, remodelagem de matriz extracelular, e transição epitélio-mesenquimal. Apesar de intensos esforços para se determinar fatores genéticos e ambientais de susceptibilidade, a etiologia desta malformação permanece pouco compreendida. Vários *loci* associados às FLP NS vêm sendo identificados por meio de estudos de mapeamento gênico convencionais, entretanto, a grande maioria dos resultados não se replica em diferentes estudos, e não há clareza acerca do efeito funcional das variantes detectadas. Neste contexto, uma abordagem interessante para investigar a etiologia da doença é a análise de expressão gênica, que pode ser utilizada para identificar alterações de vias biológicas que convergem na manifestação do quadro clínico. Em vista disso, neste trabalho nós utilizamos a análise do transcriptoma de células-tronco de polpa dental de pacientes portadores de FLP NS, com o intuito de identificar padrões de expressão relacionados a mecanismos biológicos relevantes para a embriopatogênese da doença. Obtivemos padrões de expressão que sugerem desregulação de mecanismos associados à remodelagem de matriz extracelular e à transição epitélio-mesenquimal. Além disso, ao utilizarmos diferentes condições de cultura celular, verificamos em uma nova amostra de pacientes a desregulação de vias biológicas relacionadas ao reparo de DNA e *checkpoint* do ciclo celular. Nossos dados revelam a aplicabilidade das células-tronco de polpa dental para este tipo de abordagem, e indicam que tais perfis de expressão podem levar ao acometimento da morfogênese lábio-palatina. Além disso, mostramos pela primeira vez uma conexão entre desregulação de expressão gênica e a documentada maior incidência de formas esporádicas de câncer em famílias segregando a FLP NS. Nossos resultados abrem novas possibilidades para a investigação da etiologia das FLP NS, e ajudarão na compreensão dos eventos embrionários que predispõem a essa malformação.

ABSTRACT

Non-syndromic cleft lip and palate (NSCL/P) is a multifactorial disease determined by the interplay between genetic and environmental factors, with a variable incidence of 0.34-2.29:1000 births. This malformation arises from errors during lip and palate morphogenesis, which requires tight regulation of biological mechanisms such as cellular proliferation, extracellular matrix remodelling, and epithelial-mesenchymal transition. Albeit much effort has been put into determining the genetic and environmental factors underlying disease susceptibility, the aetiology of NSCL/P remains obscure. Many candidate *loci* have been identified through conventional gene mapping strategies, however, there is a general lack of reproducibility across studies, and there is no consensus with regard to the functional implications of the identified genetic variants. In this context, an alternative approach resides in assessing differential expression patterns to identify alterations in biological networks that could lead to phenotype manifestation. Here, we analysed the transcriptome of dental pulp stem cells from NSCL/P patients in order to pinpoint dysregulated pathways involved in the embryopathogenesis of the disease. We encountered expression patterns related to dysregulation of extracellular matrix remodelling and epithelial mesenchymal transition. Moreover, by subjecting a novel NSCL/P sample to differential cell culture conditions, we observed abnormal transcription of genes partaking in DNA repair and cell cycle checkpoint pathways. Our results show the applicability of dental pulp stem cells to this strategy and suggest that the observed expression patterns could lead to impairment of lip and palate morphogenesis. Moreover, we described for the first time a connection between abnormal gene expression in these individuals and the elevated occurrence of sporadic cancer types in NSCL/P families. Our results open new possibilities to investigate the aetiology of NSCL/P and provide further insight into the ontogenetic events underlying disease predisposition.

CAPÍTULO I - Introdução e objetivos

1. Considerações gerais

As fissuras lábio-palatinas (FLPs) representam um grupo de anomalias faciais extremamente importante, correspondendo a cerca de 25% de todas as malformações congênitas (Slavkin, 1992). Para as formas não-sindrômicas (FLP NS), a incidência mundial varia de 0,34 a 2,29 em 1000 nascimentos, dependendo de fatores como: localização geográfica, condição sócio-econômica e etnia (Gorlin *et al*., 2001; Mossey *et al*., 2009). No Brasil, as estimativas variam entre 0,28 e 1,54 em 1000 (Menegotto *et al*., 1991; Loffredo *et al*., 2001). Dada a alta incidência e ao fato de que seus portadores necessitam de um extenso tratamento por uma equipe multidisciplinar (médicos, dentistas, fonoaudiólogos, geneticistas, psicólogos, nutricionistas, entre outros) para sua reabilitação, as FLPs representam um problema de saúde pública relevante (Kasten *et al*., 2008).

2. Classificação e etiologia

As FLPs podem ser classificadas em sindrômicas e não-sindrômicas. Aproximadamente 30% dos casos de FLP são sindrômicos, ou seja, ocorrem associados a outras anomalias, sendo os restantes 70% não-sindrômicos, onde a FLP é a única malformação presente (Schutte & Murray, 1999).

As FLPs sindrômicas, relacionadas a pelo menos 350 síndromes, podem estar associadas a doenças mendelianas já descritas, ser decorrentes de anomalias cromossômicas, ou ainda terem sua origem na ação de teratógenos durante a embriogênese (Gorlin *et al.*, 2001). O estudo das formas sindrômicas de FLP tem permitido a identificação de alguns genes candidatos envolvidos em sua etiologia, como por exemplo, o gene *IRF6* (síndrome de van der Woude) (Kondo *et al.*, 2002), o gene *PVRL1* (síndrome da displasia ectodérmica) (Suzuki *et al.*, 2000; Avila *et al.*, 2006), e o gene *TBX22* (fissura palatina e anquiloglossia ligada ao X) (Braybrook *et al.*, 2001; Marcano *et al.*, 2004). Além disso, foi evidenciado que muitos dos genes já identificados para as formas sindrômicas de FLP também podem contribuir para a etiologia das formas não-sindrômicas (Stanier & Moore, 2004).

Se por um lado muitos agentes etiológicos responsáveis por doenças mendelianas associadas às FLPs já foram identificados, por outro, o estudo dos fatores de predisposição às FLP NS é dificultado pela natureza dessa condição. A FLP NS é considerada uma doença de etiologia multifatorial, tendo influência de vários *loci*, com diferentes níveis de penetrância, e ainda associados a fatores ambientais (Murray, 1995). Em decorrência da alta complexidade desta patologia, há um esforço mundial na tentativa de identificar tanto os fatores genéticos como ambientais de predisposição às FLP NS. Visto que o entendimento da etiologia dessa forma de fissura ainda é relativamente pobre, e que esta apresenta uma alta incidência, no presente trabalho estudamos a variante não-sindrômica de FLP.

Há várias evidências de que fissuras de lábio acompanhadas ou não de fissura de palato possuem etiologia distinta quando comparadas às fissuras palatinas isoladas. Dados epidemiológicos mostram que FLP NS e fissuras de palato isoladas não segregam em uma mesma família (Fraser, 1970). Entretanto, em alguns casos, há sobreposição de etiologias, a exemplo da síndrome de van der Woude, causada por mutações no gene *IRF6*, na qual ambas as formas de fissura podem se manifestar na mesma genealogia (Gorlin *et al*., 2001; Kondo *et ai*., 2002; Carinci *et al*., 2007).

2.1. Fatores ambientais

A contribuição de fatores ambientais para a etiologia das FLP NS vem sendo amplamente estudada. Como o ambiente é maleável, dados gerados por tais estudos representam uma oportunidade a curto prazo para a aplicação de medidas a fim de se prevenir a recorrência desta malformação (Dixon *et al*., 2011). Neste sentido, estudos já identificaram alguns fatores de risco agindo durante a gravidez, por exemplo: tabagismo, uso de ácido retinóico, consumo de álcool, uso de anticonvulsivantes, ocorrência de infecções, deficiência de vitaminas, entre outros (Wyszynski & Beaty, 1996; Shaw *et al*., 1996; Houdayer & Bahuau, 1998, Métneki *et al*., 2005; Krapels *et al*., 2006).

3. Aspectos clínicos

As FLP NS são tipicamente caracterizadas pela presença de fenda no lábio superior com ou sem acometimento do palato, e sem associação a

outras anomalias (Fig. 1). Essa alteração leva à separação incompleta entre as cavidades nasal e oral, provocando distúrbios de alimentação, respiração e comunicação (Kasten *et al*., 2008).

Há uma grande variabilidade clínica entre os casos de FLP NS, ocorrendo desde fenótipos bastante amenos, que incluem anomalias subclínicas no músculo *orbicularis oris*, onde há uma descontinuidade das fibras musculares (Klotz *et al*., 2010), até as fissuras completas e bilaterais, que envolvem além do lábio, os palatos primário e secundário (Gorlin *et al.*, 2001) (Fig. 1), e que representam os quadros clínicos mais graves (Johnston & Bronsky, 1995). Nesses casos, além das alterações musculares, nas regiões afetadas pela fissura há uma ruptura ou quebra de continuidade do osso alveolar, na maxila, e do osso palatino. Sendo assim, há comprometimento tanto do tecido ósseo, como do muscular (De Mey *et al*., 1989; Kerrigan *et al*., 2000; Wijayaweera *et al.*, 2000).

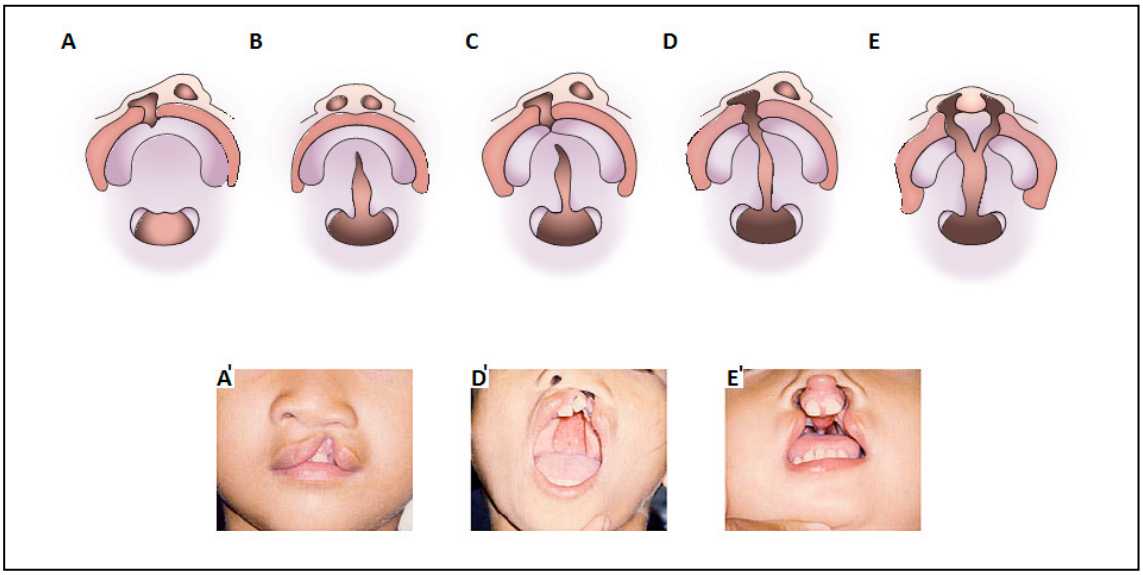


Figura 1: Variabilidade fenotípica das fissuras de lábio e palato. **A, A')** Fissura labial unilateral; **B)** Fissura palatina isolada; **C)** Fissura de lábio e palato unilateral incompleta; **D, D')** Fissura de lábio e palato unilateral completa; **E, E')** Fissura de lábio e palato bilateral completa. (Modificado de: Mossey *et al*., 2009; Dixon *et al*., 2011)

3. Desenvolvimento embrionário

A formação da face se inicia na 4ª semana de gestação, com a migração de células da crista neural craniana que, combinadas a células mesenquimais derivadas da mesoderme, formam os primórdios da face. Estes são constituídos por 5 proeminências, dispostas da seguinte maneira em relação ao estomodeu (precursor da cavidade oral): a) proeminência frontonasal, disposta rostralmente; b) 2 proeminências maxilares, dispostas lateralmente; c) 2 proeminências mandibulares, dispostas caudalmente. Entre a 5ª e 7ª semana da embriogênese, a mandíbula é formada a partir das proeminências maxilares, ao passo em que ocorre o crescimento da proeminência frontonasal, dando origem à proeminência nasal medial. Esta então cresce comcomitantemente às proeminências maxilares, ocorrendo a fusão e formando assim o lábio superior e o palato primário (Fig. 2A) (Jiang *et al*., 2006). A morfogênese do palato secundário se inicia na 6ª semana de vida intra-uterina, com o aparecimento de estruturas a partir das proeminências maxilares, chamadas de lâminas palatinas. Inicialmente, as lâminas palatinas crescem verticalmente, se posicionando em cada lado da língua em desenvolvimento. Então, ocorre a elevação e crescimento horizontal dessas estruturas, que entram em contato e subsequentemente se fundem acima do dorso da língua para formar o palato secundário, ao final da 12ª semana (Fig. 2B) (Kerrigan *et al*., 2000).

3.1 Mecanismos morfogenéticos

Levando-se em conta os dados da embriologia, acredita-se que as fissuras de lábio e palato são decorrentes de quaisquer fatores genéticos e ambientais que perturbem os mecanismos responsáveis pela correta orientação, crescimento, e fusão das proeminências faciais e das lâminas palatinas (Fig. 2C). A ontogenia orofacial se apoia no paradigma de que uma sinalização espaço-temporal precisa entre os tecidos embrionários é fundamental para a regulação de processos biológicos durante o desenvolvimento, dentre os quais podemos citar: o metabolismo de matriz extracelular, transição epitélio-mesenquimal, proliferação e migração celular, e apoptose (Jiang *et al*., 2006; Yu *et al*., 2009; Greene & Pisano, 2010).

Neste contexto, um dos mecanismos mais estudados é o de transição epitélio-mesenquimal (*epithelial-mesenchymal transition*, *EMT*). A *EMT* é um fenômeno biológico no qual células de fenótipo epitelial são reprogramadas, passando assim a manifestar um fenótipo tipicamente mesenquimal. Esta transformação é caracterizada por uma reprogramação celular que leva à perda de adesões intercelulares e aquisição de alta capacidade proliferativa, invasiva e migratória (Acloque *et al*., 2009). No contexto do desenvolvimento do lábio e palato, a *EMT* modula o processo de fusão das estruturas embrionárias orofaciais. Uma vez em contato, o epitélio dessas estruturas deve se desintegrar para que seja estabelecido um mesênquima contínuo. Sendo assim, as células epiteliais se transdiferenciam em células mesenquimais por *EMT* e atravessam a lâmina basal, invadindo o tecido mesenquimal (Sun *et al*., 2000; Nakazawa *et al*., 2008; Yu *et al*., 2009). É interessante ressaltar que existem evidências de que estas células epiteliais também sofrem apoptose, sendo que este mecanismo já foi considerado o principal responsável pela fusão do

lábio e do palato (Dudas *et al.*, 2007). No entanto, dados mais recentes indicam que tanto a morte celular programada como a *EMT* são processos envolvidos nesta etapa do desenvolvimento embrionário (Nakazawa *et al.*, 2008; Nawshad *et al.*, 2008; Iseki, 2011). Além da *EMT*, estudos relacionando capacidade de proliferação celular e metabolismo de matriz extracelular à etiologia das FLPs também vêm ganhando destaque (Sasaki *et al.*, 2004; Marinucci *et al.*, 2009; Dhulipala *et al.*, 2010; Enomoto *et al.*, 2010). Picos de proliferação celular, aliados a remodelagem de matriz extracelular, são necessários para o crescimento e posicionamento adequado das proeminências faciais e lâminas palatinas durante a morfogênese orofacial (Kerrigan *et al.*, 2000; Dhulipala *et al.*, 2006; Meng *et al.*, 2009).

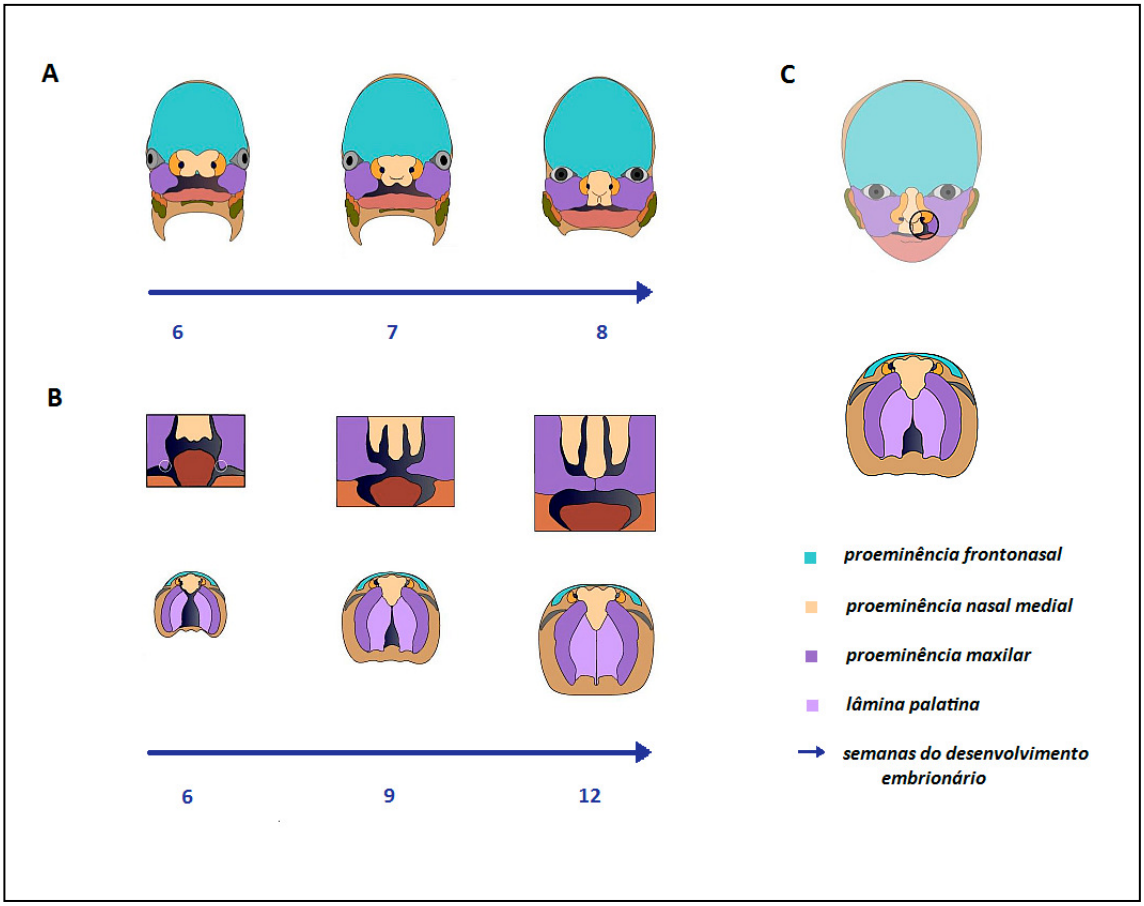


Figura 2: Morfogênese orofacial ao longo do desenvolvimento embrionário, e estruturas envolvidas. **A)** Morfogênese do lábio superior; **B)** Morfogênese do palato, em corte transversal (diagramas superiores), e em vista ventro-dorsal (diagramas inferiores); **C)** Fissura de lábio e palato. (Modificado de http://www.indiana.edu/~anat550/hnanim/face/face.html)

4. Estratégias para estudo de genes de predisposição às FLPs

4.1 Estudos de ligação e associação

Genes de susceptibilidade às FLP NS têm sido identificados em várias populações, inclusive na população brasileira (Gaspar *et al.*, 2002 & 2004; Dixon *et al.*, 2011). Para tanto, as abordagens de mapeamento genético mais utilizadas têm sido os estudos de ligação e associação.

Os estudos de ligação são baseadas na observação de co-segregação de marcadores genéticos conhecidos e o fenótipo da doença em uma ou mais famílias. Visto que *loci* próximos tendem a ser co-transmitidos por estarem geneticamente ligados, é possivel detectar regiões do genoma que potencialmente abrigam o agente etiológico da doença (Altshuler *et al.*, 2008). Este tipo de estudo requer um modelo de herança bem estabelecido, e são necessárias várias gerações com indivíduos afetados. Além disso, variantes genéticas de grande efeito são melhor detectadas com essa abordagem, a exemplo de doenças mendelianas. No entanto, segundo o modelo multifatorial de herança, as FLP NS são determinadas por vários alelos de pequeno efeito, sendo que a ocorrência de vários afetados em uma mesma família é rara.

Já nos estudos de associação de regiões candidatas e de varredura genômica (*genome-wide association study*), geralmente mais utilizados para doenças multifatoriais, é analisada a freqüência de polimorfismos entre casos e controles, estabelecendo-se assim uma associação estatística entre

presença de variante e manifestação de fenótipo. Porém, além de necessitarem de muitos indivíduos, esses estudos estão sujeitos a viés de estratificação populacional, distorcendo os resultados devido a freqüências alélicas mais freqüentes em uma determinada etnia; ou viés de desequilíbrio de ligação, no qual a variante detectada está mapeada próximo ao real agente etiológico (Altshuler *et al*., 2008).

Muitos *loci* de predisposição às FLP NS foram identificados seguindo as estratégias supracitadas (Dixon *et al*., 2011). Contudo, com exceção do gene *IRF6* e do *locus 8q24* (Ghassibé *et al.*, 2005; Park *et al*., 2007; Birnbaum *et al*., 2009; Beaty *et al*., 2010; Rojas-Martinez *et al*., 2010), não há consenso na literatura acerca da real relevância de cada um destes genes na etiologia das FLP NS, e sobretudo, de qual é o papel biológico dessas variantes na embriopatogênese da doença.

4.2 Análise do transcriptoma

Frente às dificuldades em se identificar *loci* de predisposição por meio do estudo de variantes no genoma de indivíduos acometidos por FLP NS, outra abordagem que pode ser empregada é a análise do perfil de expressão gênica. Esta estratégia permite a identificação de padrões de co-expressão gênica, associados a vias biológicas que, quando desreguladas, podem conferir predisposição à doença (Allocco *et al*., 2004; Cookson *et al*., 2009).

Tendo em vista a complexidade dos mecanismos embrionários que orquestram o desenvolvimento do lábio e palato, a análise de expressão

gênica é bastante promissora. Alguns estudos já utilizaram essa estratégia em modelos animais, onde foi analisado o perfil de expressão gênica global durante a embriogênese, levando à identificação de genes e vias de sinalização cruciais para o desenvolvimento orofacial adequado (Brown *et al*., 2003; Mukhopadhyay *et al*., 2004; Mukhopadhyay *et al*, 2010).

Além de elucidar melhor o papel de redes gênicas em situações normais de desenvolvimento, o estudo do transcriptoma pode ser aplicado à investigação das FLP e de outras doenças complexas. O perfil de expressão gênica nada mais é do que um reflexo do genoma e de suas variantes, observado sob a forma de padrões de expressão. Desta maneira, essa abordagem pode levar à confirmação de genes candidatos às FLP NS, caso suas variantes de risco estejam associadas a menores níveis de transcritos, e também à identificação de novos agentes etiológicos. Estes, por sua vez, podem estar diferencialmente expressos e exercer um efeito funcional, como no caso anterior, ou ainda, podem estar agindo como elementos regulatórios modulando o padrão de expressão ou co-regulação observado. Tendo em vista a suspeita de que perturbações em mecanismos de sinalização entre os tecidos embrionários possam ser os principais responsáveis pelas FLPs, consideramos essa estratégia adequada para investigar esta malformação.

Alguns grupos já relataram a existência de um perfil de expressão gênica associado às FLP NS. Jakobsen e colaboradores (2009), utilizando biópsias extraídas da região afetada pela fissura, encontraram um perfil de expressão distinto ao compararem indivíduos portadores de FLP NS e fissura palatina isolada. Do mesmo modo, foi observada uma desregulação de genes que

codificam moléculas de matriz extracelular, em culturas de fibroblastos de pacientes expostas a agentes teratogênicos (Bosi *et al*., 1998; Marinucci *et al*., 2009; Baroni *et al*., 2010). Esses dados confirmam a aplicabilidade de culturas celulares de pacientes para o estudo de expressão gênica e processos biológicos relevantes para a etiologia das FLP NS.

4.3 Uso de células-tronco mesenquimais (CTMs)

As células-tronco mesenquimais (CTMs) estão presentes em vários tecidos pós-natais e constituem populações celulares com capacidade de auto-renovação, se replicando por muitas gerações sem perder as características originais. *In vitro*, na presença de fatores de crescimento adequados, elas podem se diferenciar em osso, músculo, tecido adiposo, e cartilagem (Caplan, 2005). Já foi possível isolar células-tronco de diversos tecidos, como medula óssea, pele, tecido neural, retina, músculo esquelético e epitélio dentário (Gronthos *et al.*, 2002; Erices *et al.*, 2000; Williams *et al.*, 1999; Miura *et al.*, 2003; Bueno *et al*., 2009).

Sabendo que as FLPs são decorrentes de perturbações durante a embriogênese, o uso de CTMs de polpa de dente decíduo para investigar a etiologia das fissuras pode ser bastante interessante e promissor. Além de serem obtidas de forma não-invasiva, a partir de dentes em fase de esfoliação, foi demonstrado que populações celulares extraídas de polpa de dente possuem origem em células derivadas da crista neural e células mesenquimais (Chai *et al*., 2000; Waddington *et al*., 2009), as quais compõem o mesênquima das proeminências faciais embrionárias. Além disso, visto que

há comprometimento de tecido ósseo e muscular na região da fissura, o potencial de diferenciação osteogênica e miogênica das CTMs de polpa de dente poderá ser de grande valia em estudos futuros.

5. Objetivos principais

Frente às dificuldades para se elucidar a etiologia das FLP NS por meio de estratégias convencionais de mapeamento genético, propomos neste trabalho a abordagem de estudo do transcriptoma de culturas celulares oriundas de indivíduos acometidos pela doença. Para tanto, os objetivos gerais são:

a) Verificar se as CTM extraídas de polpa de dente decíduo de indivíduos acometidos por FLP NS exibem uma assinatura de expressão específica;

b) Identificar genes desregulados e padrões de expressão relacionados ao perfil transcricional dessas células;

c) A partir dos padrões de expressão observados, identificar vias de sinalização e mecanismos biológicos candidatos à embriopatogênese da doença.

CAPÍTULO II – Metodologia geral

1. Estabelecimento e manutenção das culturas celulares

A autorização para o isolamento de CTMs de polpa de dente foi obtida do Comitê de Ética em Pesquisa do Instiuto de Biociências da Universidade de São Paulo (protocolo 037/2005). O termo de consentimento livre e esclarecido foi assinado pelos pais ou responsáveis.

Polpas de dentes decíduos de indivíduos com idade entre 5 e 14 anos foram extraídas com auxílio de limas odontológicas estéreis, e lavadas 4 vezes com PBS *(Phosphate-Buffered Saline* 0,01M, PH = 7,4*)* suplementado com 4% de antibiótico *(Penicilin 5.000 U/mL, Streptomycin 5.000 μg/mL, Invitrogen)* para remoção de possíveis contaminantes e células sanguíneas. Os tecidos foram cortados em fragmentos menores e então incubados em solução de tripsina *(TrypLE™, Invitrogen)* diluída em PBS (1:10, v/v) por 1 hora a 37°C. Após inativação com *DMEM-F12 (Dubbelco's Modified Eagle Medium Nutrient Mixture F12 1:1, Invitrogen)* suplementado com 15% de soro fetal bovino (*HyClone*), os fragmentos foram centrifugados a 400xg por 5 minutos e então incubados em garrafas de cultura de 25 cm^2 contendo 5 mL de *DMEM-F12 (Invitrogen)* 15% soro fetal bovino (*HyClone*), 1% *Non-essential amino acids solution (Invitrogen)* e 1% antibiótico *(Invitrogen)*.

Todas as culturas são mantidas em estufa com ambiente umedecido a 37°C e 5% de CO_2, com troca de meio ocorrendo de 3 a 4 dias. Ao atingirem

80% de confluência, as culturas são lavadas com PBS, dissociadas com solução de tripsina, a qual é em seguida submetida à inativação com DMEM-F12/15% soro fetal bovino e posterior centrifugação. A partir deste ponto o *pellet* pode ser utilizado para extração de RNA, ou ressuspendido para semeadura em novas garrafas.

Para o congelamento, as células são ressuspendidas em meio contendo 40% de soro fetal bovino, 50% de DMEM-F12 e 10% de dimetilsulfóxido (DMSO, *LGC*), numa concentração de 5,0 x 10^5 células/mL, e acondicionadas em tubos próprios para o congelamento. A temperatura é então diminuída gradualmente, numa taxa de 1 °C por minuto, até que se atinja a temperatura final de -70 °C, na qual elas são mantidas por 24 horas antes de serem transferidas para o nitrogênio líquido.

2. Ensaios de *microarrays* de expressão

As análises da expressão das culturas foram realizadas por *microarray* com o *chip Human Gene 1.0 ST*. Todos os reagentes e equipamentos aqui descritos são produzidos pela *Affymetrix*. O *GeneChip® Whole Transcript Sense Target Labeling* é designado para gerar DNAs senso amplificados e marcados com biotina, resultantes da expressão de todo o genoma, sem nenhum tipo de viés. O procedimento de todo o experimento encontra-se melhor detalhado na Figura 1.

A partir de 300 ng de RNA total, o cDNA dupla-fita é sintetizado com hexâmeros randômicos contendo uma cauda de T7, que age como uma seqüência promotora. O cDNA dupla-fita é subsequentemente utilizado como

molde e amplificado pela polimerase T7 RNA, é realizada uma reação de transcrição *in vitro*, produzindo muitas cópias de cRNA anti-senso. No segundo ciclo de síntese do cDNA, hexâmeros randômicos são utilizados como oligonucleotídeos iniciadores para a transcrição reversa do cRNA, resultando em DNA simples-fita senso. Durante o segundo ciclo da transcrição reversa, é também adicionado dUTP à reação. O DNA simples-fita é então tratado com uma combinação de DNA uracil glicosilase (UDG) e endonuclease apurínica/apiridímica 1 (APE1), as quais reconhecem especificamente os resíduos de dUTP e quebram a fita de DNA, fragmentando-a. Em seguida, o DNA é marcado pela deoxinucleotídeo transferase terminal (TdT) com *DNA Labeling Reagent*, que é covalentemente ligado à biotina.

Este cDNA é hibridado por 17 horas ao *chip Human Gene 1.0 ST* com o auxílio de um forno de hibridação a 45 ˚C e 60 rpm. O processo de marcação por *Streptavidina* e de lavagem do *chip* ocorre na estação fluídica (*GeneChip® Fluidics Station 450*), comandada pelo *software* GCOS (*GeneChip® Console Operating Software*). Por fim, os *chips* são escanedos pelo *GeneChip® Scanner 3000 7G System* e os dados são gerados pelo *Affymetrix® Expression Console™ Software*.

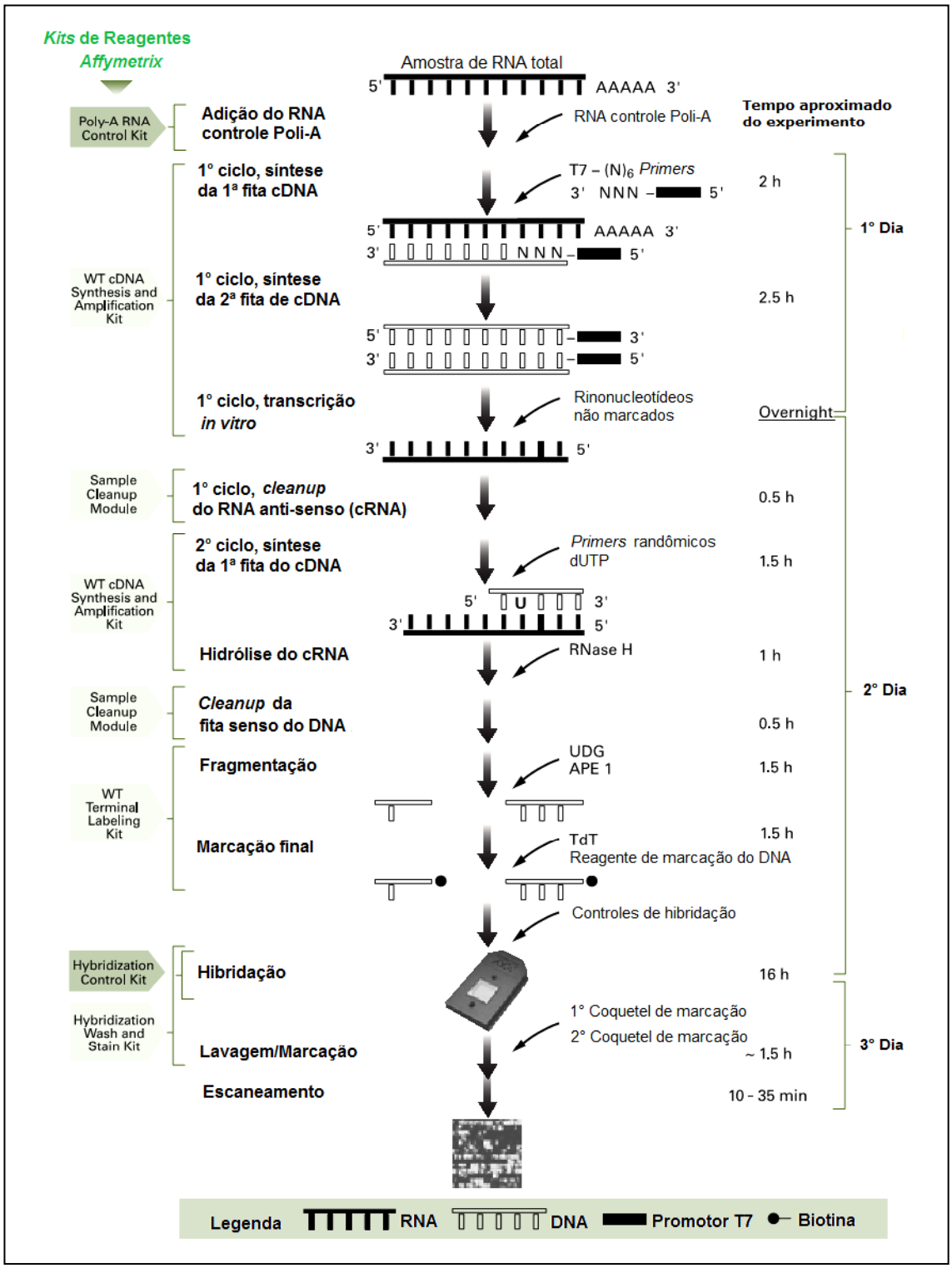


Figura 1. Esquema do procedimento de *microarray* utilizando o *chip Human Gene 1.0 ST.*

CAPÍTULO III

Human stem cell cultures from cleft lip/palate patients show enrichment of transcripts involved in extracelular matrix modeling by comparison to controls

Daniela Franco Bueno & Daniele Yumi Sunaga, Gerson Shigeru Kobayashi, Meire Aguena, Cassio Eduardo Raposo do Amaral, Cibele Masotti, Lucas Alvizi Cruz, Peter Lees Pearson, and Maria Rita Passos-Bueno

Stem Cell Rev and Rep. Jun;7(2):446-57 (2011). Published online: 30 Oct 2010

RESUMO

A fissura lábio-palatina não-sindrômica (FLP NS) é uma doença complexa que resulta da falha de fusão dos primórdios faciais, um processo que inclui a transição epitélio-mesenquimal (*EMT*). No presente trabalho, comparamos o transcriptoma de culturas de células-tronco de polpa dental (*DPSC*) entre portadores de FLP NS e controles. A rede de interação mais significativa compreendeu 13 dos 87 genes diferencialmente expressos, incluindo as proteínas extracelulares *ACAN, COL4A1, COL4A1, GDF15, IGF2, MMP1, MMP3, e PDGFA*; dentre as quais *MMP3*, *ACAN*, *COL4A1* e *COL4A2* apresentavam co-expressão. *MMP3*, um gene já associado à etiologia da doença, é responsável pela clivagem de uma variedade de proteínas extracelulares, como colágenos, proteoglicanas, fibronectina e laminina. Os resultados foram replicados em uma amostra maior, confirmando a presença padrões de expressão que diferem entre indivíduos com FLP NS e controles. Em conclusão, essas células exibem uma assinatura de expressão envolvendo genes relacionados a mecanismos de *EMT* e modelagem de matriz extracelular. Esta abordagem poderá levar à identificação de redes gênicas de predisposição a esta malformação de maneira mais rápida se comparado a estratégias convencionais de mapeamento gênico.

Stem Cell Rev and Rep (2011) 7:446–457
DOI 10.1007/s12015-010-9197-3

Human Stem Cell Cultures from Cleft Lip/Palate Patients Show Enrichment of Transcripts Involved in Extracellular Matrix Modeling By Comparison to Controls

Daniela Franco Bueno · Daniele Yumi Sunaga · Gerson Shigeru Kobayashi · Meire Aguena · Cassio Eduardo Raposo-Amaral · Cibele Masotti · Lucas Alvizi Cruz · Peter Lees Pearson · Maria Rita Passos-Bueno

Published online: 30 October 2010

Abstract Nonsyndromic cleft lip and palate (NSCL/P) is a complex disease resulting from failure of fusion of facial primordia, a complex developmental process that includes the epithelial-mesenchymal transition (EMT). Detection of differential gene transcription between NSCL/P patients and control individuals offers an interesting alternative for investigating pathways involved in disease manifestation. Here we compared the transcriptome of 6 dental pulp stem cell (DPSC) cultures from NSCL/P patients and 6 controls. Eighty-seven differentially expressed genes (DEGs) were identified. The most significant putative gene network comprised 13 out of 87 DEGs of which 8 encode extracellular proteins*: ACAN, COL4A1, COL4A2, GDF15, IGF2, MMP1, MMP3* and *PDGFa.* Through clustering analyses we also observed that *MMP3, ACAN, COL4A1 and COL4A2* exhibit co-regulated expression. Interestingly, it is known that *MMP3* cleavages a wide range of extracellular proteins, including the collagens IV, V, IX, X, proteoglycans, fibronectin and laminin. It is also capable of activating other MMPs. Moreover, *MMP3* had previously been associated with NSCL/P. The same general pattern was observed in a further sample, confirming involvement of synchronized gene expression patterns which differed between NSCL/P patients and controls. These results show the robustness of our methodology for the detection of differentially expressed genes using the RankProd method. In conclusion, DPSCs from NSCL/P patients exhibit gene expression signatures involving genes associated with mechanisms of extracellular matrix modeling and palate EMT processes which differ from those observed in controls. This comparative approach should lead to a more rapid identification of gene networks predisposing to this complex malformation syndrome than conventional gene mapping technologies.

Keywords Nonsyndromic cleft lip and palate · Gene expression profile · Dental pulp · Stem cell · Epithelial-mesenchymal transition · Extracellular matrix

Daniela Franco Bueno and Daniele Yumi Sunaga contributed equally to this work

Electronic supplementary material The online version of this article (doi:10.1007/s12015-010-9197-3) contains supplementary material, which is available to authorized users.

D. F. Bueno · D. Y. Sunaga · G. S. Kobayashi · M. Aguena · C. Masotti · L. A. Cruz · P. L. Pearson · M. R. Passos-Bueno
Human Genome Research Center,
Biosciences Institute of University of Sao Paulo (USP),
Sao Paulo, Sao Paulo, Brazil

C. E. Raposo-Amaral
Sobrapar Hospital,
Campinas, Sao Paulo, Brazil

M. R. Passos-Bueno (✉)
Depto. Genética e Biologia Evolutiva, Instituto de Biociências,
Universidade de São Paulo,
Rua do Matão, 277,
São Paulo, SP 05508-900, Brazil
e-mail: passos@ib.usp.br

Introduction

Nonsyndromic cleft lip and palate (NSCL/P [MIM 119530]), a complex multifactorial disorder, is one of the most common congenital malformations, with a prevalence of 0.69 to 2.35 per 1,000 births in the Caucasian population [1]. Taking account of the complexities of this orofacial malformation and the long rehabilitation period following surgery, cleft lip and palate is considered to be a major psychosocial and economic burden for families and society. Gaining insight

into the genetic causes of NSCL/P should lead to future improvement of genetic counseling, preventive and curative measures.

The development of the human face begins with migration of neural crest cells that combine with the core mesoderm and pharyngeal ectoderm, establishing the facial primordia, which in turn give rise to structures associated with upper lip and palate formation [2, 3]. The growth, differentiation and fusion of these structures are genetically determined, and it is likely that disturbances in genetic pathways orchestrating these processes result in facial abnormalities, such as cleft lip and palate [2–5]. In this context, the epithelial-mesenchymal transition (EMT) plays a central role in generating the cranial neural crest cells as well as ensuring palate and lip fusion [6–8].

The large number of genes known or suspected to be involved in clefting probably reflects the diversity of embryological events that contribute to the formation of these facial structures [4, 9–16]. Although gene mapping approaches, such as Genome-Wide Association Studies (GWAS), appeared to offer an option to identify at-risk alleles associated with NSCL/P, with better reproducibility among different studies [10, 11, 17, 18] than candidate genes, lack of progress over the last decade suggests that GWAS are still unlikely to provide sufficient information on the genetic etiology underlying the disease. However, genome-wide expression analyses based on differential gene expression associated with NSCL/P, as proposed here; present a viable and challenging alternative, since patterns of co-expression can be used to identify biological pathways or gene networks associated with disease predisposition. The current data supporting this supposition can be summarized as follows: transcriptome analysis in tissues of cleft palate (CP) patients showed a distinct gene expression signature when compared to CL/P [19]. It has been reported that a few genes coding for extracellular matrix proteins, such as *TGFB3* and *MMP3*, are differentially expressed in fibroblasts of NSCL/P patients as compared to controls [20, 21], suggesting that the transcriptome of NSCL/P cells might exhibit a specific expression signature irrespective of origin of the cells concerned.

The use of mesenchymal stem cells (MSCs) or induced pluripotent stem cells (iPSs) has been shown to be a promising new approach to study gene function and signaling pathways in genetic disorders [22–24]. MSCs constitute a long-lived population of cells possessing self-renewal and differentiation properties [25–29]. Accordingly, these cells are a good model to study the *in vitro* characteristics of NSCL/P, since in addition to gene expression, they can be tested for proliferation, migration and differentiation properties, including EMT, functions that are presumed to be altered in cells of NSCL/P patients during embryonic development. MSCs were originally isolated from bone marrow, and subsequently, similar cell populations were isolated from other tissues, such as adipose tissue [27], dental pulp [28, 30], *orbicularis oris* muscle [26], umbilical cord blood, and umbilical cord tissue [31]. Moreover, MSCs can be easily obtained from non-invasive sources, such as exfoliated teeth, both from NSCL/P patients and control individuals [28].

The main aim of this study was to verify if there are consistent gene expression profile differences between mesenchymal stem cells from NSCL/P patients and controls. We chose to study stem cells from dental pulp as they can be obtained relatively easy from both NSCL/P patients and controls. In addition, these stem cells, as for any other cells obtained from craniofacial tissues, are derived from cranial neural crest cells which play an important role in craniofacial development, including lip and palate [32, 33]. Our results will provide a base line for further investigation and insights into genetic pathway irregularities associated with craniofacial clefts.

Materials and Methods

Sample: NSCL/P Patients and Controls

Ethical approval to obtain stem cells from dental pulp of deciduous teeth was obtained from the Biosciences Institute Research Ethics Committee (Protocol 037/2005). Samples were included only after signed informed consent by the parents or legal guardians.

Deciduous teeth from controls were obtained from odontopediatric clinics in Sao Paulo, while teeth from NSCL/P patients were excised during the exfoliation period by Dr. Bueno D.F. at Sobrapar, Campinas, Sao Paulo. An individual was classified as NSCL/P if no other malformations than clefting of both lip and palate were present.

We analyzed mRNA of dental pulp stem cell cultures (DPSC) obtained from 6 controls and 6 NSCL/P patients (Table 1, supplementary material) for microarray expression analysis and for validation of 4 genes by quantitative Real-Time PCR (qRT-PCR). Validation of the microarray expression analysis by qRT-PCR for a larger number of genes was also done in mRNA obtained from DPSC cultures of 16 additional controls and 13 NSCL/P patients.

Stem Cell Culture

Stem cell cultures obtained from DPSC of deciduous teeth were established according to previously published protocols [28]. Cells were cultured at 37°C with 5% CO_2 in DMEM-F12 (Invitrogen, UK) supplemented with 15% Fetal Bovine Serum (FBS, HyClone, USA) and frozen in

Table 1 List of 87 differentially expressed genes sorted out by comparing controls and NSCL/P patients with the RankProd analysis

AFFY ID	Gene symbol	FC[a]	Pfp[b]	Cytoband
8176375	RPS4Y1	5.1099	0.0007	Yp11.3
8176624	DDX3Y	4.9358	0.0000	Yq11
8041206	LBH	4.6707	0.0000	2p23.1
8118509	PPT2 [c]	4.2212	0.0008	6p21.3
7972750	COL4A1 [d c]	3.6284	0.0006	13q34
7932254	ITGA8 [d c]	3.2852	0.0037	10p13
8125537	HLA-DMA [c]	3.2541	0.0005	6p21.3
7998927	–	3.0562	0.0053	–
8108370	EGR1	3.0266	0.0124	5q31.1
8176719	EIF1AY	2.9797	0.0055	Yq11.222
7952205	MCAM [d c]	2.9129	0.0123	11q23.3
7953200	CCND2	2.7894	0.0089	12p13
8056491	SCN9A [c]	2.7724	0.0223	2q24
7964388	NDUFA4L2	2.7670	0.0112	12q13.3
8113504	C5orf13	2.7473	0.0077	5q22.1
8176578	USP9Y	2.7360	0.0219	Yq11.2
8104663	CDH6 [d c]	2.7285	0.0263	5p15.1-p14
7985786	ACAN [d c]	2.6889	0.0042	15q26.1
8121838	TPD52L1	2.6688	0.0106	6q22-q23
8177137	UTY	2.6455	0.0280	Yq11
7970033	COL4A2 [d c]	2.6123	0.0210	13q34
7965573	NTN4 [c]	2.5853	0.0129	12q22-q23
8003298	SLC7A5 [c]	2.5767	0.0174	16q24.3
8156783	COL15A1 [c]	2.5452	0.0250	9q21-q22
8090565	SNORA7B	2.5214	0.0225	3q21.3
7954293	PDE3A	2.5063	0.0121	12p12
8137670	PDGFA [d c]	2.4802	0.0179	7p22
8049888	ATG4B	2.4474	0.0115	2q37.3
7974895	FLJ43390	2.4260	0.0427	14q23.2
7958262	TCP11L2	2.4172	0.0253	12q23.3
7989985	ITGA11 [c]	2.4038	0.0447	15q23
8104035	SORBS2	2.3912	0.0291	4q35.1
8102800	SLC7A11 [c]	2.3697	0.0390	4q28-q32
7912157	ERRFI1	2.3596	0.0308	1p36
7981728	–	2.3590	0.0407	–
7985493	TM6SF1	2.3535	0.0442	15q24-q26
7981962	SNORD116-5	2.3518	0.0419	15q11.2
7931977	ITIH5	2.3354	0.0486	10p14
7928308	DDIT4	2.3277	0.0467	10pter-q26.12
8034940	NOTCH3 [c]	2.3234	0.0396	19p13.2-p13.1
7966089	CMKLR1 [c]	2.2946	0.0455	12q24.1
8068024	JAM2 [d c]	2.2538	0.0397	21q21.2
8023152	TCEB3CL	2.2139	0.0407	18q21.1
8109159	LOC728264	2.2134	0.0459	5q33.1
8117018	JARID2	2.2051	0.0408	6p24-p23
8086752	RNU13P3	2.2046	0.0452	3p21.31
7982868	CHAC1	2.2041	0.0273	15q15.1
8139207	INHBA [c]	2.1839	0.0405	7p15-p13
8027352	–	2.1608	0.0399	–

Table 1 (continued)

AFFY ID	Gene symbol	FC[a]	Pfp[b]	Cytoband
8027002	GDF15 [c]	2.1594	0.0285	19p13.11
8048749	KCNE4 [c]	2.1436	0.0184	2q36.3
8007850	LRRC37A	2.0206	0.0244	17q21.31
8160138	NFIB	1.9790	0.0182	9p24.1
8131867	–	1.8907	0.0477	–
8064978	JAG1 [d] [c]	1.8529	0.0178	20p12.1-p11.23
8042788	ACTG2	1.7905	0.0414	2p13.1
8045804	–	1.7730	0.0492	–
8092726	CLDN1 [d]	1.4872	0.0445	3q28-q29
7937772	IGF2	−1.8634	0.0165	11p15.5
7915592	–	−1.8745	0.0195	–
8124391	HIST1H2AB	−2.0165	0.0426	6p21.3
8163202	SVEP1	−2.0970	0.0214	9q32
8037240	PSG1 [c]	−2.1467	0.0273	19q13.2
8044605	LOC654433	−2.2491	0.0468	–
8117594	HIST1H2BM	−2.2734	0.0128	6p22-p21.3
7951271	MMP1 [d] [c]	−2.3465	0.0026	11q22.3
8138888	PDE1C	−2.4348	0.0263	7p15.1-p14.3
8140668	SEMA3A [c]	−2.5040	0.0042	7p12.1
7951284	MMP3 [d] [c]	−2.5261	0.0433	11q22.3
8107044	ERAP2	−2.5407	0.0291	5q15
8003667	SERPINF1 [c]	−2.5688	0.0109	17p13.1
8110916	LOC442132	−2.5879	0.0258	5p15.31
7904293	PTGFRN [c]	−2.6619	0.0164	1p13.1
8113800	FBN2 [c]	−2.7034	0.0000	5q23-q31
8116418	GFPT2	−2.8180	0.0212	5q34-q35
8152617	HAS2 [d]	−2.8501	0.0038	8q24.12
8129573	MOXD1 [c]	−2.8809	0.0181	6q23.1-q23.3
7925929	AKR1C3	−3.1055	0.0165	10p15-p14
7917850	ARHGAP29	−3.1793	0.0031	1p22.1
8037251	PSG7 [c]	−3.2618	0.0006	19q13.2
8180291	–	−3.2801	0.0006	
8165808	XG [c]	−3.6591	0.0065	Xp22.33
8037272	PSG5 // PSG5 [c]	−3.7775	0.0000	19q13.2 // 19q13.2
8083887	CLDN11 [d]	−4.2221	0.0000	3q26.2-q26.3
7909730	KCNK2 [c]	−4.7183	0.0000	1q41
8037283	PSG4 [c]	−4.9049	0.0000	19q13.2
8152522	ENPP2 [d] [c]	−5.3472	0.0000	8q24.1

[a] FC = Fold change

[b] Pfp = estimated percentage of false positive predictions.

([c]) Genes that were functionally categorized as glycoproteins by DAVID (p=8,0E-6).

([d]) Genes involved in EMT.

40% FBS for storage. Frozen cells were thawed and grown until 80% confluency in a 75 cm^2 flask and submitted to serum starvation (12 h) prior to RNA extraction. All experiments were conducted with cells between the 4th and 8th subculture.

Characterization of Mesenchymal Stem Cell Populations

Cell populations used in the microarray assays were analyzed in a flow cytometer for specific cell surface markers to evaluate homogeneity. Cells were harvested

with TrypLe (Invitrogen, UK), washed with PBS, and incubated at 4°C for 30 min with the following antibodies: CD29-PE CY5, CD90 (Thy-1), CD45-FITC, CD31-PE (Becton Dickinson, USA), SH2, SH3, and SH4 (Case Western Reserve University, USA). After a second wash, samples incubated with unconjugated primary antibodies were then incubated with anti-mouse-PE secondary antibody (Guava Technologies, Hayward, CA) for an additional 15 min at 4°C. Finally, the cell suspension was washed with PBS, and signals from 10^5 cells were acquired with an EasyCyte Flow cytometer (Guava Technologies). Control samples for determining background noise were incubated with PBS instead of primary antibody followed by incubation with anti-mouse-PE secondary antibody. All plots generated were analyzed with Guava ExpressPlus software (Guava Technologies).

The *in vitro* differentiation into bone, cartilage, muscle and fat had previously been demonstrated in 2 of the NSCL/P patients (F4280, F4285) and 3 of the control (F3363, F4271, F4272) samples included in this study [28, 29]. Because of the high reproducibility of our protocols, stem cell cultures are currently characterized only with respect to their immunophenotype.

RNA Processing

Total RNA was isolated using TRIzol (Invitrogen, UK) according to the manufacturer's protocol and purified with RNeasy Mini-kit (QIAGEN). RNA quality and concentration were assessed using Nanodrop 1000 and gel electrophoresis. Only samples with a preserved rRNA ratio (28S/18S) and no evidence of RNA degradation were used in the microarray hybridization and qRT-PCR.

Microarray Processing

Expression measurements were performed using the Affymetrix Human Gene 1.0 ST array, which interrogates 28,869 genes, following RNA labeling and hybridization protocols as recommended by the manufacturer. After array scanning, quality control was performed with GCOS software (Affymetrix, USA) according to the manufacturer's recommendations.

Transcriptome Analysis

Gene expression values were obtained using the three-step Robust Multi-array Average (RMA) pre-processing method, implemented in the Affy package from R/Bioconductor [34].

We employed the RankProd method for the selection of differentially expressed genes (DEGs), considering a p-value cut-off of 0.05 adjusted for FDR (False Discovery Rate) [35]. RankProd is a rank-based nonparametric procedure [36]. The method has the advantage of being able to deal with few samples for identifying biologically relevant expression changes [37]. Genes selected by RankProd do not necessarily need to present homogenous expression levels across all the samples of the test and control groups. Accordingly, RankProd seems to be a good choice for identifying differential gene expression in complex diseases, in which altered expression of a given candidate gene is expected in just a subgroup of patients due to both multilocus genetic heterogeneity and the stochastic nature of gene expression in complex systems [38].

Functional Annotation

We performed functional annotation analysis of the differentially expressed genes with the DAVID (Database for Annotation, Visualization and Integrated Discovery, http://david.abcc.ncifcrf.gov) and IPA (Ingenuity Pathway Analysis, http://www.ingenuity.com) tools. In IPA, we considered the default parameter Molecules per Network=35, Networks per Analysis=25, only direct relationships between genes and the "Ingenuity Expert Information" Data Source, including the new option of "Ingenuity ExpertAssist Findings", in which the information has been manually reviewed and curated from full-text scientific publications.

Validation of Microarray Expression Using Quantitative Real-Time PCR (qRT-PCR)

Initially we performed qRT-PCR for 4 genes (*COL15A1, ERAP2, PPT2 and EGFL8*) on the same samples used in the microarray assay. These genes were randomly selected, but they were amongst those with the highest fold change. *PPT2* and *EGFL8* are represented by a common probe set on the Affymetrix Human Gene 1.0 ST array and therefore both genes were tested for qRT-PCR with gene-specific primers. Next, we performed qRT-PCR for further 12 genes (*ACAN, CDH6, CLDN1, CLDN11, COL4A1, COL4A2, ENPP2, HAS2, ITGA8, JAG1, MCAM* and *MMP3*) plus the genes *COL15A1, ERAP2 and PPT2* in 16 controls and 13 NSCL/P patients. Only 4 of the control and 4 of the NSCL/P patients samples were the same as those used in the microarray assay due to unavailability of RNA from the remaining 4 individuals (Table 1 in supplementary material).

One microgram of total RNA from each cell culture was reverse transcribed with Superscript II (Invitrogen, UK), according to manufacturer's recommendations. Quantitative Real-Time PCR reactions were performed in duplicates with a final volume of 20 µl, using 20 ng of cDNA, 1X SYBR Green PCR Master Mix (*Applied Biosystems*) and 100–400 nM of each primer. We used

ABI Prism 7500 Sequence Detection System (*Applied Biosystems*) with standard temperature protocol. Primers were designed with Primer Express software V.2.0 (*Applied Biosystems*; primers sequence in supplementary Table 2) and the amplification efficiency (E) of each primer was calculated according to the equation $E=10^{(-1/slope)}$. The expression data of the target transcripts were determined by relative quantification in comparison to a pool of RNAs (4 controls and 4 patients). GeNorm v3.4 was used to determine the most stable endogenous controls (*SDHA, HPRT1* and *GAPDH*) and to calculate the normalization factors for each sample [39]. Expression values were calculated according to reference [40].

To compare the expression of *COL15A1, ERAP2, PPT2 and EGFL8,* obtained by qRT-PCR and microarray assay in the same samples, we used an unpaired t-test with Welch's correction.

To compare the results obtained by microarray with those obtained by qRT-PCR in the novel samples, for which we do not have microarray data, we performed the following strategy: for each of the 15 genes, we calculated the average (avg) expression for controls and for NSCL/P patients obtained by both methods (Table 3 in supplementary material). The correlation between the ratios "avg_patients/avg_controls" from microarray and qRT-PCR assays was calculated using Spearman's correlation test.

Clustering Analysis

The cluster analysis of the differentially expressed genes was performed using GEDI (Gene Expression Dynamics Inspector) software. This tool creates a 2 dimensional gene expression image for each sample in which each gene retains exactly the same position in the image of each sample and in which the gene positions are computed to give the most parsimonious gene arrangement for depicting expression level differences between the patient and control groups for all differentially expressed genes [41]. In the analysis, a 10×11 grid configuration of SOM (Self-Organizing Map) was used. Inspection of GEDI images allows a straightforward classification of the samples into subgroups without the aid of a clustering algorithm, but simply based on the visual differences in the patterns [42].

We also used two other conventional clustering methods: K-means and Hierarchical, both available in the MeV (MultiExperiment Viewer) software [43], with Spearman's correlation as the distance metric. The clustering analysis of qRT-PCR data followed two criteria: a) only DEGs from network 1 (Fig. 4) and b) only DEGs that showed the same tendency of expression in the qRT-PCR and microarray assays (Fig. 2 in supplementary material).

Results

Characterization of Mesenchymal Stem Cell Populations

The cell cultures presented positive labeling for cell adhesion (CD29, CD90) and mesenchymal stem cell markers (SH2, SH3, SH4) in most of the cells (>90%) and were negative for endothelial and hematopoietic cell markers (Fig. 2 in supplementary material). Moreover, 2 NSCL/P patients and 3 control cell cultures used in the microarray analyses had been previously shown to differentiate into bone, muscle, cartilage and fat upon *in vitro* induction [28, 29]. Therefore the cell populations used in this study had the main properties of stem cells.

Controls versus NSCL/P Patients

We identified a total of 87 differentially expressed genes (DEGs; 58 upregulated and 29 downregulated) in the comparison between controls and NSCL/P patients (adjusted $p \leq 0.05$; Table 1).

Next, in order to visualize the expression behavior of these DEGs, we performed clustering analysis with the GEDI software. An image was created reflecting the 87 DEGs' transcriptional behavior for each individual and where the gene position was fixed to give the most parsimonious arrangement to show differential gene expression between controls and NSCL/P patients. Upon visual inspection of the GEDI images, we observed that 4 of the NSCL/P patients showed a similar expression pattern (F4280, F4281, F4282, F4283; Fig. 1).

The clustering analysis with the k-means method resulted in 9 gene clusters. Four of them exhibited a similar expression profile among NSCL/P patients, most particularly in the afore-mentioned group (F4280, F4281, F4282, F4283; Fig. 2). The similarities and dissimilarities in expression levels observed between NSCL/P patients were similar for both clustering methods. The expression pattern found in 4 out of 6 NSCL/P patients illustrates the characteristic of RankProd of being capable of selecting genes with differential expression in just a subgroup of samples.

Functional Annotation

Differentially expressed genes were functionally annotated and analyzed with two different tools. First, the DAVID tool led to identification of three main canonical pathways from the KEGG Database: Focal adhesion, Cell adhesion molecules and ECM-receptor interaction (Fig. 3, 4 and 5 in supplementary material). Moreover, the most relevant functional category identified through DAVID was that of Glycoproteins (p=8.0E-6), which included 36 of the 87

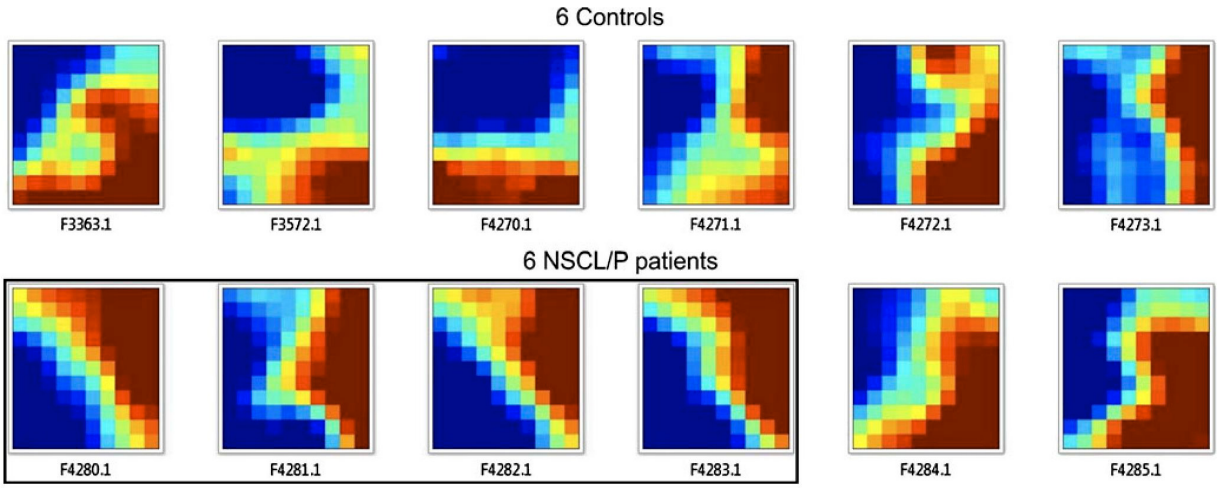


Fig. 1 Clustering of 87 DEGs resulted from the comparison between 6 controls and 6 NSCL/P patients. Each GEDI map (or mosaic) represents a gene expression profile of a single individual. The blue color represents the lowest expression level and red color represents the highest expression level on a scale of −4.70 to 7.98, respectively. The black frame highlights four patients with similar gene expression profile

DEGs (Table 1). Subsequently, the IPA tool was used to characterize the 87 DEGs regarding possible biological functions. We observed 3 relevant functions enriched with a significant number of our genes: Cellular movement (20 genes, p=6.4E-06–2.76E-02), Cellular growth and proliferation (27 genes, p=3.11E-05–2.68E-02) and Cellular development (27 genes, p=3.3E-05–2.47E-02) (Table 4 in supplementary material). A putative network with the largest number of DEGs built by IPA (13 DEGs; Fig. 4) suggests functional relationship among several extracellular proteins: *ACAN, COL4A1, COL4A2, GDF15, IGF2, MMP1, MMP3* and *PDGFa*. All of these 8 genes are DEGs.

Validation of the Microarray Analysis

The reproducibility of gene expression assayed by Affymetrix microarrays is high [44] and 4 genes (*COL15A1, ERAP2, PPT2* and *EGFL8*) were initially selected for validation through qRT-PCR. Except for *ERAP2* (p=0.0397), that showed lower expression levels, the other genes showed higher transcript levels in NSCL/P patients than in controls: *PPT2* (p=0.0087) and *COL15A1* (p=0.0355) (Fig. 3), which confirms the expression patterns observed in the microarray assays. *EGFL8* is represented by the same probe set of *PPT2*. Considering that we did not find significant differential expression levels through qRT-PCR with specific primers

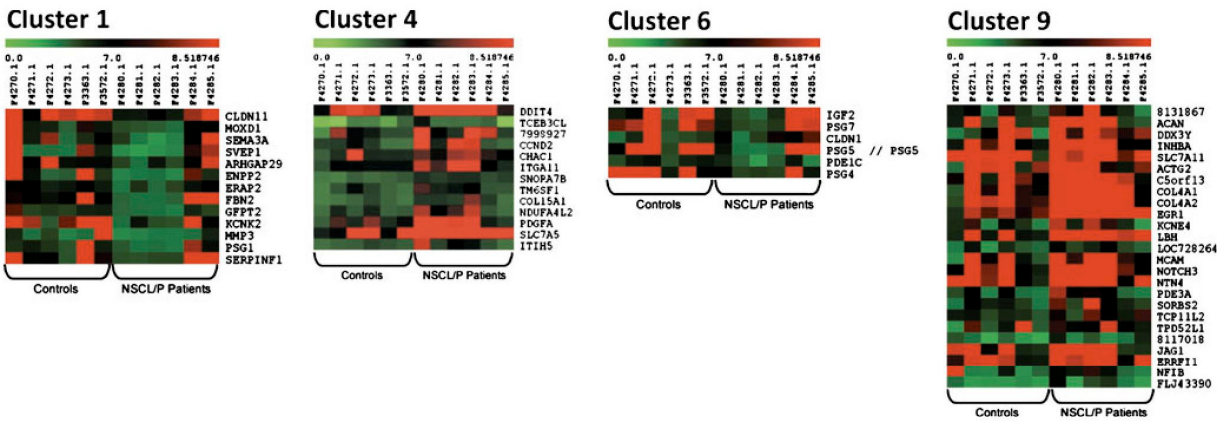


Fig. 2 Gene clusters 1, 4, 6 and 9 resulted from k-means method (k=9). In both clusters it is possible to observe a similar gene expression profile among 4 out of 6 patients (F4280, F4281, F4282 and F4283), indicating that many of the 87 selected DEGs are co-regulated in these 4 NSCL/P patients

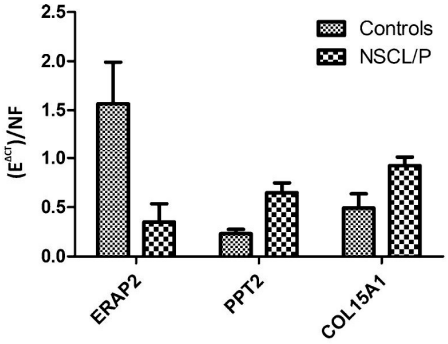


Fig. 3 Quantitative Real-Time PCR (qRT-PCR) initial analysis of NSCL/P patients and control samples for *ERAP2, PPT2, COL15A1*. E = primer amplification efficiency; CT = cycle threshold; delta_CT = sample's average_CT normalized by pool's average_CT; NF = normalization factor

for *EGFL8* (p=0.1781), we kept only *PPT2* among our candidate genes.

For further validation of our results, in addition to *COL15A1, PPT2* and *ERAP2*, we analyzed a further 12 other genes known to be involved with matrix remodeling in a novel sample of individuals (16 controls and 13 NSCL/P patients). By comparing the ratios of the average expression from patients/controls, we observed that *ENPP2* exhibited the highest discordance between the microarray and qRT-PCR ratios (0.645 and 4.324, respectively), therefore we considered this gene expression as not validated. However, we observed a significant positive correlation between microarray and qRT-PCR expression data (p=0.0114, r=0.653) for the 14 genes chosen for validation (Table 3 in supplementary material). Accordingly, the analyses thus show that the data obtained from microarrays and qRT-PCR are consistent with each other, even in an enlarged novel sample, attesting to the reliability of the microarray analysis.

Using NSCL/P patients expression data from qRT-PCR (13 NSCL/P patients) and microarray assays (6 NSCL/P patients), we also performed a clustering analysis for the following DEGs: *ACAN*, *COL4A1*, *COL4A2*, *ERAP2*, *HAS2*, and *MMP3*. These genes belong to network 1 (Fig. 4) and exhibited the same expression tendency in both experiments (Fig. 2 in supplementary material). The hierarchical clustering performed with qRT-PCR data

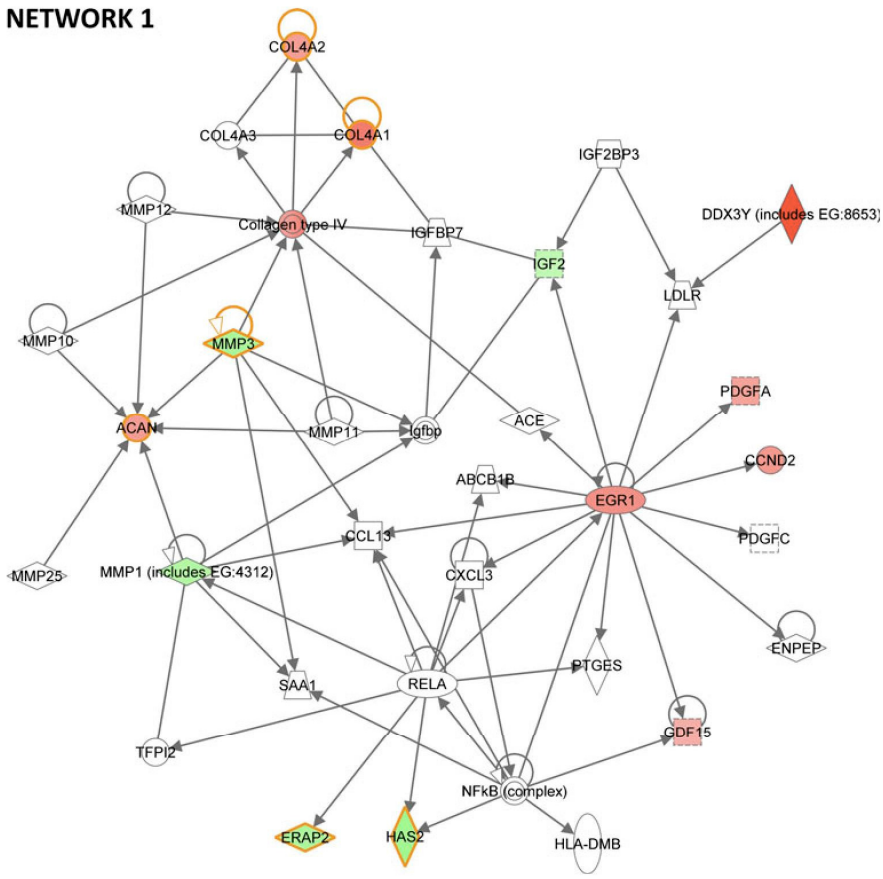
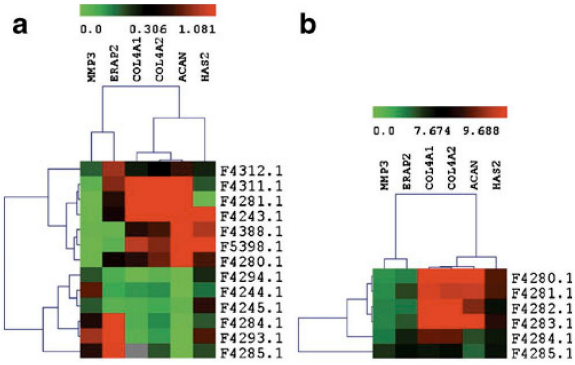


Fig. 4 The most significant network built by IPA tool with the highest number of differentially expressed genes. The upregulated genes in NSCL/P patients are indicated in red and the downregulated genes in green. The blank symbols pertain to genes that were either not present in our array or not differentially expressed. Solid lines indicate a direct linkage among two genes. Lines with arrows indicate that one gene acts on the other, and lines without arrows indicate that the corresponding proteins interact with each other. The 6 genes circled in orange were used in clustering analysis of qRT-PCR and microarray



c

	MMP3		COL4A1		COL4A2		ACAN		HAS2	
	Spearman r	p	Spearman r	p	Spearman r	p	Spearman r	p	Spearman r	p
MMP3	-	-	-	-	-	-	-	-	-	-
COL4A1	- 0.811	<0.05	-	-	-	-	-	-	-	-
COL4A2	- 0.791	<0.05	0.979	<0.05	-	-	-	-	-	-
ACAN	- 0.889	<0.05	0.921	<0.05	0.872	<0.05	-	-	-	-
HAS2	- 0.275	0.364	0.196	0.542	0.187	0.541	0.231	0.448	-	-
ERAP2	0.368	0.216	- 0.056	0.863	- 0.0055	0.986	- 0.239	0.431	- 0.390	0.188

Fig. 5 Hierarchical clustering of all the patients analyzed by qRT-PCR and microarray, considering only the genes from IPA network 1 (Fig. 4). **a** Clustering of expression values obtained by qRT-PCR. **b** Clustering of expression values obtained by microarray. **c** Correlations (Spearman's correlation test, r and p-values) between each co-regulated gene from qRT-PCR clustering. It is possible to observe that the gene *MMP3* is significantly and inversely correlated to *ACAN, COL4A1, COL4A2* genes in a subgroup of patients from qRT-PCR. This same pattern of co-regulation is also present in another group of patients analyzed by microarray, which includes two individuals (F4280 and F4281) from the mentioned qRT-PCR subgroup. In controls expression data of both assays, *MMP3* is upregulated and *ACAN, COL4A1* and *COL4A2* downregulated

revealed a homogeneous cluster with 7 out of 13 NSCL/P patients (F4312, F4311, F4281, F4243, F4388, F5398 and F4280), in which *MMP3* is downregulated and *ACAN, COL4A1* and *COL4A2* upregulated. *ERAP2* and *HAS2* did not exhibit a consistent expression pattern in these patients (Fig. 5a). Interestingly, the hierarchical clustering using microarray data showed the same transcriptional behavior among these genes in 4 out of 6 NSCL/P patients (F4280, F4281, F4282 and F4283), although in this case, *ERAP2* and *HAS2* seem to be co-regulated with *MMP3* and collagens (Fig. 5b). Spearman's correlation test (Fig. 5c) calculated for each relationship among the genes from Fig. 5a confirmed that *MMP3* and *ACAN* are negatively correlated ($r=-0.889$; $p<0.05$) while *ACAN* exhibited positive correlation with *COL4A1* and *COL4A2* ($r=0.921$ and $r=0.872$, respectively; $p<0.05$), corroborating our findings thus far. On the other hand, we did not achieve significance for *ERAP2* and *HAS2*, confirming the lack of a well-defined expression pattern for these genes, as observed in Fig. 5a.

Discussion

In this study, we conducted the first comparative transcriptome analysis between dental pulp stem cells from NSCL/P patients and controls to obtain more information on the genetic etiology of this malformation and explore new possibilities to identify pathways associated with disease pathology.

Genome expression microarray analysis is a powerful tool for assessing changes in the transcription patterns of related genes and identification of signaling pathways associated with specific cell types, culture conditions or disease states [45, 46]. Considering that the cell populations from NSCL/P patients and controls were established and maintained under similar identical conditions, it is very likely that a large proportion of the DEGs identified is related to the genetic constitutional differences between cells from NSCL/P patients and controls. The observations that NSCL/P disease-specific expression profiles have also been previously reported in tissue biopsies and fibroblast cultures [19, 21] suggests that such profiles may be ubiquitous; in this respect our findings suggest that the disease-specific transcription profile is also present in mesenchymal stem cells. Accordingly, transcriptome analysis of stem cells represents an additional option to the study of the transcriptome and genetics of NSCL/P.

Of the 87 DEGs obtained in our microarray analysis, we noted that *MMP3* had previously been proposed as a candidate gene for NSCL/P [47]. A further 2 genes, *PDGFa* and *ERRFI1*, had previously been indirectly associated with NSCL/P, since their receptors, respectively *PDGFR* and *EGRF*, were identified as predisposing loci for this form of clefting [48–50]. These observations provide further confirmation that the transcriptome of DPSCs from NSCL/P patients can also be used to identify genetic variations associated with the disease.

The functional annotation through DAVID showed that nearly half of the 87 DEGs correspond to glycoprotein molecules, including collagens, metalloproteinases, integrins, and adhesion proteins, which are important to orchestrate the signaling between the extracellular and intracellular compartment. Indeed, the three canonical pathways we identified through DAVID are mainly related to extracellular matrix components and signaling molecules located on the cell membrane (Fig. 3, 4 and 5 in supplementary material). Functional relationships among several extracellular proteins that are deregulated in our studies are also suggested by the putative network built by IPA (Fig. 4). These analyses suggest that a large proportion of the DEGs in the DPSC from NSCL/P patients are extracellular matrix components (ECM) involved in several cellular processes in facial development, such as extracellular remodeling during the epithelial-mesenchymal transition (EMT).

EMT is a fundamental mechanism behind palatal fusion. This process occurs through a regulated sequence of events determined both by the extracellular environment and the gene expression program of the cell, leading to loss of cell-cell adhesion, breakdown of basal laminae, and increased cell invasion and mobility [51, 52].

Fifteen genes enrolled in EMT are within the 87 DEGs (Table 1). Further, we observed the enrichment of biological functions involving cell proliferation, movement and invasion (Table 4 in supplementary material), all of which are expected phenotypic outcomes for genes involved with EMT.

In the most relevant IPA network we observed that 8 out of 13 DEGs (*ACAN, COL4A1, COL4A2, GDF15, IGF2, MMP1, MMP3* and *PDGFa,* Fig. 4) are extracellular matrix components, suggesting that an abnormal expression behavior of these genes may affect EMT during palate development.

Clustering analyses showed that *MMP3, ACAN, COL4A1* and *COL4A2* transcripts are co-regulated in 4 out of 6 NSCL/P patients analyzed by microarray as well as in 7 out of 13 NSCL/P patients from qRT-PCR analysis. Therefore, it seems that in NSCL/P mesenchymal cells, the down-regulation of *MMP3* is associated with up-regulation of *ACAN*, *COL4A1* and *COL4A2*. Such a deregulated pathway probably has serious consequences in embryonic development, since it is known that the *MMP3* protein cleaves a wide range of ECM proteins, including the collagens IV, V, IX, X, proteoglycans, fibronectin and laminin. It is also capable of activating other MMPs, as shown in network 1 (Fig. 4), and playing a key role in ECM degradation and remodeling [47]. In this context it has already been experimentally shown that lower levels of MMPs can block palatal fusion [7]. Therefore, it is possible that failure of lip and palate fusion in these groups of patients is at least partly associated with down-regulation of MMPs and up-regulation of collagens. It will be important in the near feature to identify the leading causative mechanism of altered expression of MMPs in these individuals. These results also show the robustness of RankProd to detect DEGs on a limited and heterogeneous group of samples, in contradistinction to a popular method SAM [53] which appears to require larger sample sizes. Moreover, RankProd is able to identify expression patterns in just a subgroup of affected samples, which is the ideal, considering that this is the expected to occur in a complex disease such as NSCL/P. Notwithstanding our favorable impression of RankProd, we are acutely aware of the small sample sizes employed in detecting DEGs in the initial microarray assay. Future studies must be directed towards confirming and/or expanding the pattern of DEGs using novel and much larger sample sizes.

In summary our results suggest that NSCL/P patients exhibit deregulation of genes participating in either EMT or embryonic processes that depend on extracellular modeling. Abnormal expression of some genes encoding for extracellular matrix proteins in NSCL/P fibroblasts has also been reported by others, reinforcing the concept that disease expression signatures for NSCL/P are present in various tissues and not necessarily confined to a specific cell type. Moreover, the penetrance of the phenotype can depend on exposure to environmental factors and it is of interest that a recent report claimed a positive association between nicotine exposure and the CL/P phenotype in conjunction with deregulation of gene expression involved in ECM synthesis and degradation [21].

The observation that the DEGs in NSCL/P patients are associated with ECM component signaling suggests that further analysis of these cells presents unique opportunities to study the complexity of molecular pathways and alleles involved in NSCL/P etiology. However, arguably the major advantage of DPSCs above other cell types, such as fibroblasts, will be their potential to study *in vitro* differentiation into muscle, bone and cartilage that are affected tissues in NSCL/P. In this context, it will be possible to integrate genomic and transcriptome analysis under different experimentally induced environmental insults on identical cell cultures. Our results open new avenues for the study of novel candidate genes for NSCL/P, since most of the 87 DEGs have not been previously associated with NSCL/P. In particular, the potential offered by our approach can be best visualized in the gene network 1 (Fig. 4), in which several of our DEGs are ECM components, suggesting that these genes might be enrolled in EMT during the development of NSCL/P phenotype.

If wide-spread differences in co-regulated gene transcription, as observed in our experiments, are indeed a primary cause of NSCL/P, then the research emphasis must move away from naively cataloging which genes are being differentially expressed to defining the central transcription factors and regulatory elements that are driving the disease specific transcription patterns. In this respect, identifying the putative gene networks involved, as in our observations, will be an initial crucial step towards defining which regulatory elements are the most important.

Acknowledgements The authors would like to express their gratitude to: Mrs. Constancia Urbani for secretarial assistance, and the following colleagues: Rodrigo Atique, BSc, for helping us with figures; Eder Zucconi, PhD, and Natassia Vieira, BSc, for helping us with flow cytometry; Andrea Sertié, PhD, and Keith Okamoto, PhD for their helpful reviews and suggestions regarding the manuscript. FAPESP/CEPID, CNPq, MCT, FINEP sponsored this study.

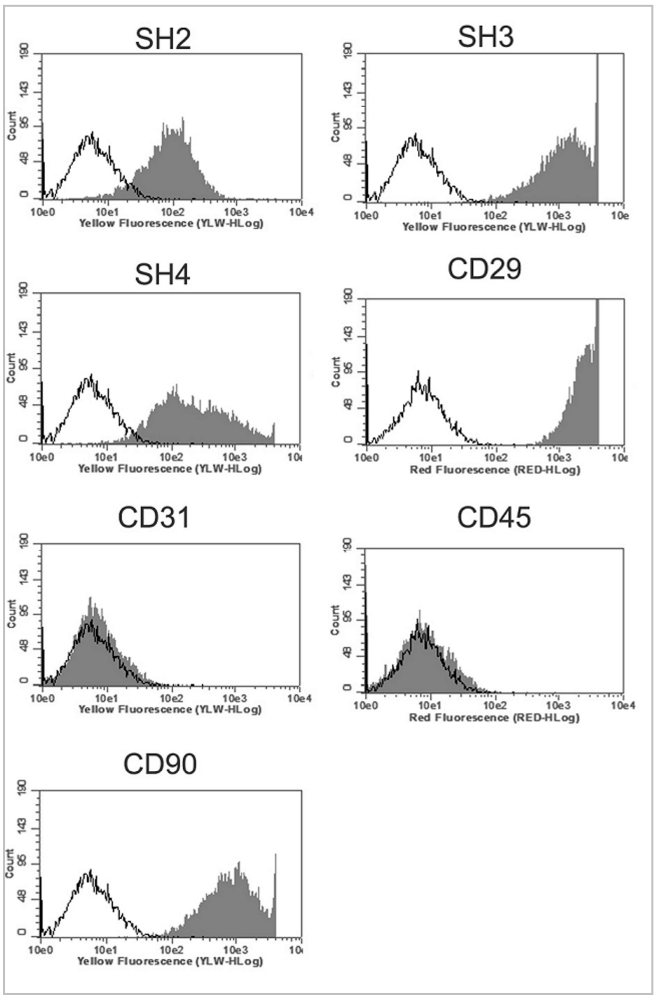
Disclosures The authors indicate no potential conflicts of interest.

References

1. Gundlach, K. K. H., & Maus, C. (2006). Epidemiological studies on the frequency of clefts in Europe and world-wide. *Journal of Cranio-Maxillo-Facial Surgery, 34*(2), 1–2.
2. Kerrigan, J. J., Mansell, J. P., Sengupta, A., Brown, N., & Sandy, J. R. (2000). Palatogenesis and potential mechanisms for clefting. *Journal of the Royal College of Surgeons of Edinburgh, 45*(6), 351–358.
3. Jiang, R., Bush, J. O., & Lidral, A. C. (2006). Development of the upper lip: morphogenetic and molecular mechanisms. *Developmental Dynamics, 235*(5), 1152–1166.
4. Jugessur, A., Farlie, P. G., & Kilpatrick, N. (2009). The genetics of isolated orofacial clefts: from genotypes to subphenotypes. *Oral Diseases, 15*(7), 437–453.
5. Suzuki, K., Hu, D., Bustos, T., et al. (2000). Mutations of PVRL1, encoding a cell-cell adhesion molecule/herpesvirus receptor, in cleft lip/palate-ectodermal dysplasia. *Nature Genetics, 25*(4), 427–430.
6. Acloque, H., Adams, M. S., Fishwick, K., Bronner-Fraser, M., & Nieto, M. A. (2009). Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. *The Journal of Clinical Investigation, 119*(6), 1438–1449.
7. Yu, W., Ruest, L. B., & Svoboda, K. K. (2009). Regulation of epithelial-mesenchymal transition in palatal fusion. *Experimental Biology and Medicine, 234*(5), 483–491.
8. Kang, P., & Svoboda, K. K. (2005). Epithelial-mesenchymal transformation during craniofacial development. *Journal of Dental Research, 84*(8), 678–690.
9. Juriloff, D. M., & Harris, M. J. (2008). Mouse genetic models of cleft lip with or without cleft palate. *Birth Defects Research. Part A: Clinical and Molecular Teratology, 82*(2), 63–77.
10. Mangold, E., Ludwig, K. U., Birnbaum, S., et al. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nature Genetics, 42*(1), 24–26.
11. Mangold, E., Reutter, H., Birnbaum, S., et al. (2009). Genome-wide linkage scan of nonsyndromic orofacial clefting in 91 families of central European origin. *American Journal of Medical Genetics. Part A, 149A*(12), 2680–2694.
12. Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Steffens, M., Rubini, M., et al. (2009). Key susceptibility locus for non-syndromic cleft lip with or without cleft palate on chromosome 8q24. *Nature Genetics, 41*, 473–477.
13. Marazita, M. L., Murray, J. C., Lidral, A. C., Arcos-Burgos, M., Cooper, M. E., Goldstein, T., et al. (2004). Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32–35. *American Journal of Human Genetics, 75*, 161–173.
14. Cai, J., Chen, J., Liu, Y., et al. (2006). Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells, 24* (3), 516–530.
15. Gong, S. G., Gong, T. W., & Shum, L. (2005). Identification of markers of the midface. *Journal of Dental Research, 84*(1), 69–72.
16. Brown, N. L., Knott, L., Halligan, E., Yarram, S. J., Mansell, J. P., & Sandy, J. R. (2003). Microarray analysis of murine palatogenesis: temporal expression of genes during normal palate development. *Development, Growth & Differentiation, 45*(2), 153–165.
17. Nikopensius, T., Birnbaum, S., Ludwig, K. U., et al. (2010). Susceptibility locus for non-syndromic cleft lip with or without cleft palate on chromosome 10q25 confers risk in Estonian patients. *European Journal of Oral Sciences, 118*(3), 317–319.
18. Beaty, T. H., Murray, J. C., Marazita, M. L., et al. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nature Genetics, 42* (6), 525–529. Erratum in: *Nature Genetics*, 42(8), 727, 2010.
19. Jakobsen, L. P., Borup, R., Vestergaard, J., et al. (2009). Expression analyses of human cleft palate tissue suggest a role for osteopontin and immune related factors in palatal development. *Experimental & Molecular Medicine, 41*, 77–85.
20. Brown, N. L., Yarram, S. J., Mansell, J. P., & Sandy, J. R. (2002). Matrix metalloproteinases have a role in palatogenesis. *Journal of Dental Research, 81*, 826–830.
21. Baroni, T., Bellucci, C., Lilli, C., 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. *Journal of Cellular Physiology, 222*(3), 748–756.
22. Gunaseeli, I., Doss, M. X., Antzelevitch, C., Hescheler, J., & Sachinidis, A. (2010). Induced pluripotent stem cells as a model for accelerated patient- and disease-specific drug discovery. *Current Medicinal Chemistry, 17*(8), 759–766.
23. Masotti, C., Ornelas, C., Splendore-Gordonos, A., Moura, R., Felix, T., Nivaldo, A., et al. (2009). Reduced transcription of TCOF1 in adult cells of Treacher Collins syndrome patients. *BMC Medical Genetics, 10*(1), 136.
24. Tsai, M.-S., Hwang, S.-M., Chen, K.-D., Lee, Y.-S., et al. (2007). Functional network analysis of the transcriptomes of mesenchymal stem cells derived from amniotic fluid, amniotic membrane, cord blood, and bone marrow. *Stem Cells, 25*(10), 2511–2523.
25. Zucconi, E., Vieira, N. M., Bueno, D. F., et al. (2010). Mesenchymal stem cells derived from canine umbilical cord vein—a novel source for cell therapy studies. *Stem Cells and Development, 19*(3), 395–402.
26. Bueno, D. F., Kerkis, I., Costa, A. M., et al. (2009). New source of muscle-derived stem cells with potential for alveolar bone reconstruction in cleft lip and/or palate patients. *Tissue Engineering, 15*, 427–435.
27. Vieira, N., Bueno, C. R., Jr., Brandalise, V., et al. (2008). Dystrophic mice express a significant amount of human muscle proteins following systemic delivery of human adipose-derived stromal cells without immunosuppression. *Stem Cells, 26*(9), 2391–2398.
28. Costa, A. M., Bueno, D. F., Martins, M. T., et al. (2008). Reconstruction of large cranial defects in nonimmunosuppressed experimental design with human dental pulp stem cells. *The Journal of Craniofacial Surgery, 19*, 204–210.
29. Bueno, D. F. (2007). The use of adult stem cells to study the etiopathogeny of cleft lip and palate and tissue engineering. Phd in human genetics, Institute of Biosciences, University of Sao Paulo (USP), Rua do Matao, trav. 14, n321—Cidade Universitaria—Sao Paulo—Brazil.
30. Gronthos, S., Brahim, J., Li, W., Fisher, L. W., et al. (2002). Stem cell properties of human dental pulp stem cells. *Journal of Dental Research, 81*(8), 531–535.
31. Secco, M., Moreira, Y., Zucconi, E., et al. (2009). Gene expression profile of mesenchymal stem cells from paired umbilical cord units: Cord is different from blood. *Stem Cell Reviews and Reports*. Epub ahead of print.
32. Widera, D., Zander, C., Heidbreder, M., et al. (2009). Adult palatum as a novel source of neural crest-related stem cells. *Stem Cells, 27*(8), 1899–1910.

33. Grenier, J., Teillet, M. A., Grifone, R., Kelly, R. G., & Duprez, D. (2009). Relationship between neural crest cells and cranial mesoderm during head muscle development. *PLoS ONE, 4*(2), e4381.
34. Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., & Speed, T. P. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research, 31*(4), e15.
35. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57*, 289–300.
36. Hong, F., Breitling, R., McEntee, C. W., Wittner, B. S., Nemhauser, J. L., & Chory, J. (2006). Rankprod: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics, 22*(22), 2825–2827.
37. 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 Letters, 573*, 83–92.
38. Raj, A., Rifkin, S. A., Andersen, E., & van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature, 463*(7283), 913–918.
39. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, *3*(7), research0034.1–0034.11.
40. Pfaffl, M. W. (2001). A new mathematical model for quantification in real-time-RT-PCR. *Nucleic Acids Research, 9*, 2002–2007.
41. Eichler, G. S., Huang, S., & Ingber, D. E. (2003). Gene expression dynamics inspector (GEDI): for integrative analysis of expression profiles. *Bioinformatics, 19*, 2321–2322.
42. Guo, Y., Eichler, G. S., Feng, Y., Ingber, D. E., & Huang, S. (2006). Towards a holistic, yet gene-centered analysis of gene expression profiles: a case study of human lung cancers. *Journal of Biomedicine and Biotechnology*, 69141.
43. Saeed, A. I., Sharov, V., White, J., et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques, 34*(2), 374–378.
44. Bochukova, E., Soneji, S., Wall, S., & Wilkie, A. O. (2009). Scalp fibroblasts have a shared expression profile in monogenic craniosynostosis. *Journal of Medical Genetics*, [Epub ahead of print].
45. LaGamba, D., Nawshad, A., & Hay, E. D. (2005). Microarray analysis of gene expression during epithelial-mesenchymal transformation. *Developmental Dynamics, 234*(1), 132–142.
46. Fanganiello, R. D., Sertié, A. L., Reis, E. M., et al. (2007). Apert pSer252Trp mutation in FGFR2 alters osteogenic potential and gene expression of cranial periosteal cells. *Molecular Medicine, 13*(7,8), 422–442.
47. Letra, A., Silva, R. A., Menezes, R., et al. (2007). MMP gene polymorphisms as contributors for cleft lip/palate: association with MMP3 but not MMP1. *Archives of Oral Biology, 52*(10), 954–960.
48. Xu, X., Bringas, P., Jr., Soriano, P., & Chai, Y. (2005). PDGFR-alpha signaling is critical for tooth cusp and palate morphogenesis. *Developmental Dynamics, 232*(1), 75–84.
49. Ferby, I., Reschke, M., Kudlacek, O., et al. (2006). Mig6 is a negative regulator of EGF receptor–mediated skin morphogenesis and tumor formation. *Nature Medicine, 12*, 568–573.
50. Miettinen, P. J., Chin, J. R., Shum, L., Slavkin, H. C., Shuler, C. F., Derynck, R., et al. (1999). Epidermal growth factor receptor function is necessary for normal craniofacial development and palate closure. *Nature Genetics, 22*, 69–73.
51. Yu, W., Ruest, L. B., & Svoboda, K. K. (2009). Regulation of epithelial-mesenchymal transition in palatal fusion. *Experimental Biology and Medicine, 234*(5), 483–491.
52. Khudyakov, J., & Bronner-Fraser, M. (2009). Comprehensive spatiotemporal analysis of early chick neural crest network genes. *Developmental Dynamics, 238*(3), 716–723.
53. Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America, 98*(9), 5116–5121.

Supplementary Material



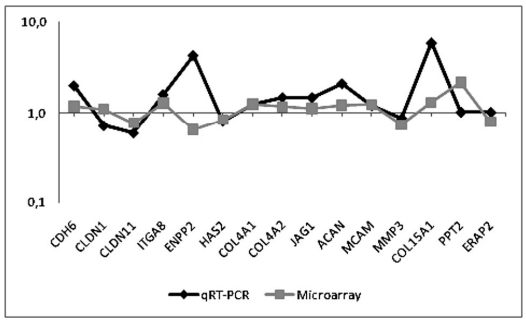


Figure 2: Comparison between expression ratios (log10 avg_patients/avg_controls) obtained by qRT-PCR (black line) and microarray (grey line) of the 15 genes chosen for validation.

Figure 1: Flow cytometry analysis of dental pulp stem cell (DPSC) cultures. The graphs display the immunophenotype of adherent DPSC isolated from NSCL/P patients. The control assay (not labeled) is represented by a black line and the experimental population (labeled with specific antibodies) is in gray. Values represent the mean percentage of all assessed cells positively stained for the indicated antigens: SH2 (97.43%), SH3 (97.02%), SH4 (93.92%), CD29 (99.9%), CD31 (0.08%), CD45 (3.84%) and CD90 (97.96%). Abbreviations: CD, Cluster of Differentiation.

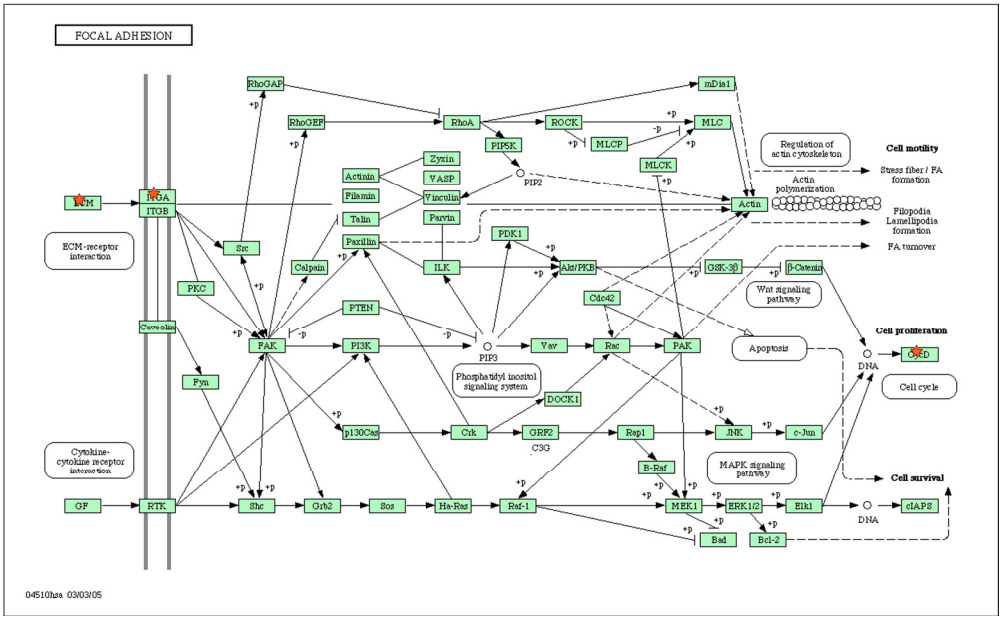


Figure 3: Focal Adhesion pathway from the KEGG Database sorted out by DAVID tool. The red stars correspond to genes deregulated in the comparison between NSCL/P patients and controls (ECM, ITGA and CycD).

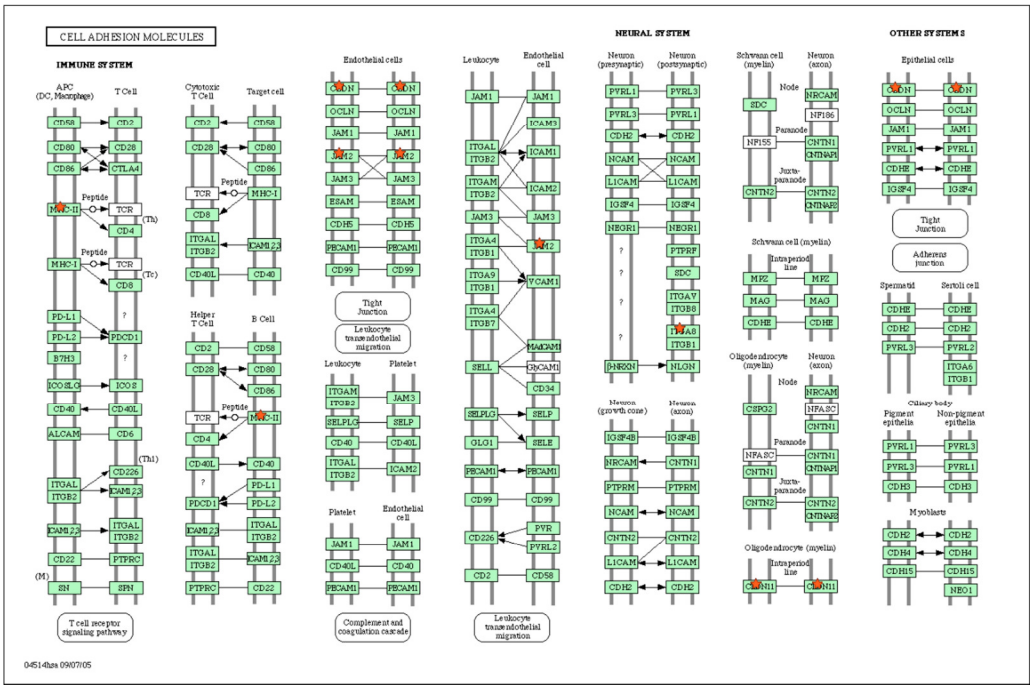


Figure 4: Cell Adhesion Molecules pathway from the KEGG Database sorted out by DAVID tool. The red stars correspond to genes deregulated in the comparison between NSCL/P patients and controls (MHC-II, CLDN, JAM2, ITGA8 and CLDN11).

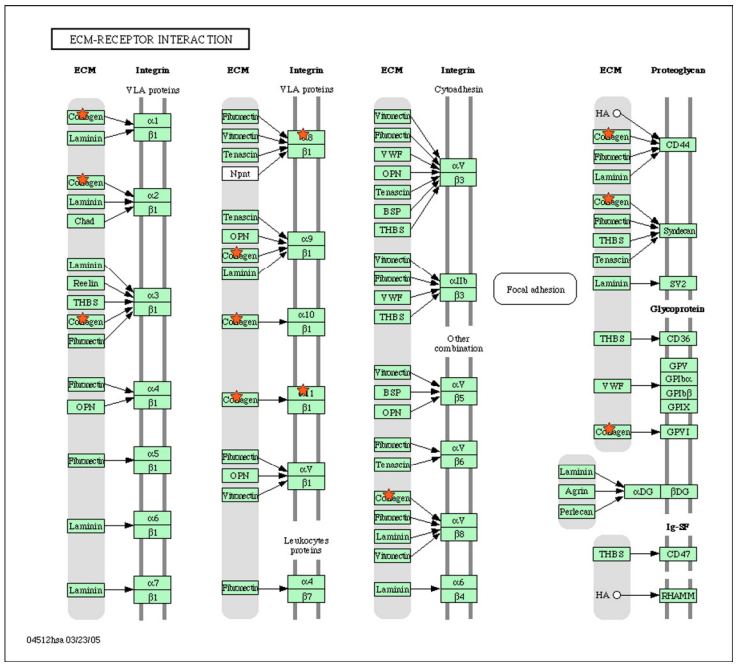


Figure 5: ECM-receptor interaction pathway from the KEGG Database sorted out by DAVID tool. The red stars correspond to genes deregulated in the comparison between NSCL/P patients and controls (Collagen, alpha8 and alpha12).

Table 1: Age, gender, and clinical status of the samples analyzed by qRT-PCR and/or microarray.

Code		Age	Gender	Clinical Status*	Method
1)	F4293.1	12	female	CL/UL	qRT-PCR
2)	F4311.1	12	male	CL/UL	qRT-PCR
3)	F4294.1	10	female	CL/UL	qRT-PCR
4)	F4285.1	8	female	CLP/B	microarray/qRT-PCR
5)	F4280.1	7	male	CLP/B	microarray/qRT-PCR
6)	F4282.1	8	male	CLP/UL	microarray
7)	F4281.1	8	male	CLP/UL	microarray/qRT-PCR
8)	F4284.1	8	male	CLP/UL	microarray/qRT-PCR
9)	F4283.1	8	male	CLP/UL	microarray
10)	F5398.1	8	male	CLP/UL	qRT-PCR
11)	F4312.1	8	female	CLP/UL	qRT-PCR
12)	F4243.1	5	male	CLP/UR	qRT-PCR
13)	F4244.1	12	male	CLP/UR	qRT-PCR
14)	F4245.1	10	male	CLP/UR	qRT-PCR
15)	F4388.1	14	male	CLP/UR	qRT-PCR
16)	F3363.1	6	female	control	microarray/qRT-PCR
17)	F3572.1	10	female	control	microarray/qRT-PCR
18)	F4273.1	8	male	control	microarray
19)	F4270.1	9	male	control	microarray/qRT-PCR
20)	F4271.1	8	male	control	microarray/qRT-PCR
21)	F4272.1	8	male	control	microarray/qRT-PCR
22)	F3334.1	8	female	control	qRT-PCR
23)	F3959.1	7	male	control	qRT-PCR
24)	F4215.4	9	female	control	qRT-PCR
25)	F4232.1	10	male	control	qRT-PCR
26)	F4264.1	6	female	control	qRT-PCR
27)	F4319.1	9	female	control	qRT-PCR
28)	F4386.1	11	female	control	qRT-PCR
29)	F4387.1	8	female	control	qRT-PCR
30)	F3960.1	7	female	control	qRT-PCR
31)	F4233.1	8	male	control	qRT-PCR
32)	F4217.1	8	male	control	qRT-PCR
33)	F3347.1	9	male	control	qRT-PCR

* CLP/UL = Unilateral Left Cleft Lip and Palate, CLP/UR = Unilateral Right Cleft Lip and Palate, CLP/B = Bilateral Cleft Lip and Palate.

Table 2: Primer sequences used for qRT-PCR.

Genes	Primer Forward (5'-3')	Primer Reverse (5'-3')
HPRT1	TGACACTGGCAAAACAATGC	GGTCCTTTTCACCCAGCAAGC
SDHA	TGGGAACAAGAGGGCATCTG	CCACCACTGCATCAAATTCA
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGA
COL15A1	AGGCAGACGGAGTAGCTGAGAT	GGATGGAGTTGGAGGTGTGTTT
ERAP2	TCCAAGCCAAGTTAGGTGATGG	CAACGGTTCATCAAAGCAAGG
PPT2	CATCTCATCTGCTACTCGCAG	ACTGTCCCATCTGTGGAGAGG
EGFL8	TAAGAAGCCTCTACAGCACCCC	TGCTCTTTGGTGGCTGATAGC
CDH6	AGCTGCAGTTTCAGCCGCGA	AGGATATCTCTGCTCGCCTTCCTG
ITGA8	AGGTGGGCTGGCCTTTCTCT	TGGGGAGGCAGCAGGCTTTA
CLDN1	TGCCGGCGACAACATCGTGA	GGGTTGCTTGCAATGTGCTGCT
CLDN11	CTGGGTCTGCCGGCCATTTT	GCGCAGAGAGCCAGCAGAATGA
ENPP2	TGCTGCGGAAACTCGTCAGGA	TCTTCCATTCCATGGTCTCCGACA
COL4A1	GGATGCAATGGCACAAAGGGGGA	TCAACAGCATCCCGGGCACA
COL4A2	AGTGGACGAACCGCCAGCAT	TTCACACCCGCCAAGACGCT
MMP3	GCAGGCAAGACAGCAAGGCA	GCAAACTGCCACGCACAGCA
JAG1	AGTGGTGCCAAGTGCCAGGA	TCGAGGGCCACACCAGACCTTT
ACAN	ACTGCTGCAGACCAGGAGGT	TGTGACTGCTGCTCCCGACA
HAS2	AGCTCGCAACACGTAACGCA	TCGTCCCAGTGCTCTGAAGGCT
MCAM	ACGTCAACGGCACGGCAAGT	GGAGGCCGTGCATTCAACACC

Table 3: Ratios of the average expression from patients/controls for 15 candidate genes chosen for validation.

Candidate Genes for Validation	CDH6	CLDN1	CLDN11	ITGA8	ENPP2	HAS2	COL4A1	COL4A2	JAG1	ACAN	MCAM	MMP3	COL15A1	PPT2	ERAP2
Number of tested patients/controls	13/16	13/16	8/15	13/16	13/16	13/16	12/16	13/14	12/16	13/15	13/15	13/16	13/14	13/16	13/16
qRT-PCR avg_patients/avg_controls	2.009	0.717	0.593	1.603	4.324	0.803	1.261	1.491	1.486	2.109	1.216	0.876	5.921	1.025	1.018
Number of tested patients/controls	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6	6/6
microarray avg_patients/avg_controls	1.184	1.102	0.771	1.278	0.645	0.835	1.243	1.176	1.117	1.214	1.241	0.736	1.306	2.163	0.792

Table 4: Top 3 (out of 70) functional categories identified by Ingenuity Pathways Analysis as being significantly overrepresented by differentially expressed genes.

Functional category / number of genes	***P- value****	***Genes***
Cellular Movement (20)	6.4E-06 - 2.76E-02	*CMKLR1, CLDN11, COL4A1, MMP3, PDGFA, ACAN, EGR1, GDF15, PPT2, COL4A2, INHBA, SEMA3A, IGF2, MCAM, HAS2, ERRFI1, ENPP2, JAG1, C5ORF13, MMP1*
Cellular Growth and Proliferation (27)	3.11E-05 - 2.68E-02	*EIF1AY, CLDN11, SLC7A11, AKR1C3, MMP3, PDGFA, GDF15, COL4A2, ITGA11, MCAM, SLC7A5, ERRFI1, ENPP2, COL4A1, NOTCH3, EGR1, KCNK2, SERPINF1, INHBA, SEMA3A, IGF2, CCND2, HAS2, NFIB, JAG1, C5ORF13, FBN2*
Cellular Development (27)	3.3E-05 - 2.47E-02	*CLDN11, SLC7A11, AKR1C3, MMP3, NTN4, PDGFA, ITGA8, GDF15, COL4A2, MCAM, MMP1, HLA-DMA, CMKLR1, NOTCH3, COL4A1, EGR1, SERPINF1, INHBA, DDX3Y, SEMA3A, IGF2, CCND2, UTY, HAS2, JAG1, C5ORF13, USP9Y*

* The significance value associated with each biological function is a measure of the likelihood of that function being associated with the dataset due to random chance. The range of p-values is calculated with the right-tailed Fisher's Exact Test, which compares the number of user-specified genes to the total number of occurrences of these genes in the respective functional annotations stored in the Ingenuity Pathways Knowledge Base.

CAPÍTULO IV

Gene expression analysis of cleft lip/palate stem cells reveals dysregulation of DNA damage repair and cell cycle checkpoint control pathways

Gerson Shigeru Kobayashi, Lucas Alvizi Cruz, Daniele Yumi Sunaga, Daniela Franco Bueno, Simone Gomes Ferreira, Meire Aguena, and Maria Rita Passos-Bueno

RESUMO

A fissura lábio-palatina não-sindrômica (FLP NS) é uma doença multifatorial causada por erros durante a morfogênese orofacial embrionária. Apesar de intensos esforços para se determinar fatores genéticos e ambientais de susceptibilidade, a etiologia desta malformação ainda é pouco compreendida. No presente trabalho, por meio da análise de expressão gênica de células-tronco de polpa dental de indivíduos portadores de FLP NS, identificamos padões transcricionais associados à desregulação de mecanismos de controle do *checkpoint* celular e reparo de DNA dependente de *BRCA1*. Também identificamos o fator de transcrição *E2F1* como um possível elemento regulatório modulando o perfil de expressão observado. Nossos resultados sugerem que deficiências na progressão do ciclo celular e reparo de DNA podem estar relacionadas à embriopatogênese da FLP NS, além de prover uma ligação entre a desregulação desses mecanismos e a documentada maior incidência de formas esporádicas de câncer em famílias segregando a doença.

Gene expression analysis of cleft lip/palate stem cells reveals dysregulation of DNA damage repair and cell cycle checkpoint control pathways

Gerson Shigeru Kobayashi, Lucas Alvizi Cruz, Daniele Yumi Sunaga, Daniela Franco Bueno, Simone Gomes Ferreira, Meire Aguena, and Maria Rita Passos-Bueno

ABSTRACT

Non-syndromic cleft lip/palate (NSCL/P) is a multifactorial disease that arises from errors during embryonic orofacial morphogenesis. Although much effort has been put into identifying genetic and environmental factors underlying disease susceptibility, the aetiology of this complex malformation remains obscure. Here, using global gene expression analysis of NSCL/P dental pulp stem cell cultures, we identified expression patterns associated with a general dysregulation of *BRCA1*-dependent DNA repair and cell cycle checkpoint pathways. Moreover, we identified transcription factor *E2F1* as a putative upstream regulator driving the observed expression signature. Our results suggest that deficient cell cycle progression and DNA repair could be related to impairment of embryonic lip and palate morphogenesis, and also provide a link between dysregulation of these pathways and the documented increased occurrence of sporadic cancer in NSCL/P families.

INTRODUCTION

Non-syndromic cleft lip with or without cleft palate (NSCL/P [MIM 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 (Mossey *et al*., 2009). NSCL/P generates significant medical and psychological burdens for affected children and their families (Christensen & Mortensen, 2002; Kasten *et al*., 2008). Therefore, gaining insight into the causes of this malformation may lead to improved preventive and curative measures in the future.

NSCL/P is a complex disease, determined by the interplay between environmental and genetic agents during embryonic orofacial development. Although experimental and epidemiological studies have been able to identify genetic variants and environmental factors associated with the disease, there is a general lack of consistency across studies, and the aetiology of this complex malformation remains unclear (Dixon *et al*., 2011). Nevertheless, it has been thoroughly demonstrated that gene-environment interactions play a critical role in the aetiological heterogeneity underlying NSCL/P (Lidral *et al*., 2008; Zhu *et al*., 2009; Marinucci *et al*., 2011).

Abnormal transcript levels - which may be determined by a combination of factors, such as environmental insults and functionally deficient regulatory elements - is considered to be an important mechanism underlying susceptibility to complex diseases (Cookson *et al*., 2009). Tightly regulated

spatio-temporal patterns of cell proliferation, apoptosis and transdifferentiation are pivotal to ensure appropriate lip and palate morphogenesis. Consequently, dysregulated gene expression may cause disturbances in these mechanisms, leading to the development of orofacial clefts. In this context, it has been demonstrated that genome-wide expression profiling to assess patterns of differential expression is a viable alternative to identify biological pathways associated with disease predisposition (Brown *et al*., 2003; Jakobsen *et al*., 2009; Bueno & Sunaga *et al*., 2010).

In a previous work, our group reported that dental pulp stem cells from NSCL/P patients display a disease-specific expression signature with dysregulation of genes involved in extracellular matrix remodelling and epithelial-mesenchymal transition (Bueno & Sunaga *et al*., 2010). Noteworthy is that, prior to gene expression profiling, these cell cultures were subject to serum withdrawal, which may represent a different environment when compared to cells grown in serum-supplemented culture medium. The use of cell culture modifiers, such as teratogens associated with occurrence of orofacial clefting, has been thoroughly explored to study gene expression in the context of NSCL/P (Bosi *et al*., 1998; Young *et* al., 2000; Zhang *et al*., 2003; Marinucci *et al*., 2009; Pezzetti *et al*., 2009; Baroni *et al*., 2010). Thus, we deemed necessary to examine the transcriptome of NSCL/P cells without serum deprivation.

METHODS

Cell cultures

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). Samples were included only after signed informed consent by the parents or legal guardians. Deciduous teeth from controls were obtained from odontopaediatric clinics in Sao Paulo, while teeth from NSCL/P patients were excised during the exfoliation period by Dr. Bueno D.F. at the Sobrapar Institute, Campinas, São Paulo. 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 was present. RNA extracted from cell cultures derived from a total of 6 controls and 7 NSCL/P patients (Table I, Supplementary Material) was used for microarray assays and quantitative real-time PCR.

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 (de Mendonça Costa A & Bueno *et al*., 2008; Bueno *et al*., 2009; Bueno & Sunaga *et al*., 2010). Cells were cultured in DMEM-F12 (Invitrogen) supplemented with 15% Foetal Bovine Serum (FBS; HyClone), 1% Non-essential aminoacids solution (GIBCO), 1% penicillin-streptomycin solution (GIBCO), in a humidified incubator at 37°C and 5% CO_2. For storage, cells were frozen in medium containing 40% FBS, 50% DMEM-F12, and 10% DMSO (LGC). For RNA extraction, frozen cells were

thawed and grown until 80% confluent in a 75 cm^2 culture flask. All experiments were conducted using cells between the 4th and 8th passage.

RNA extraction and microarray hybridisation

Total RNA was isolated using 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 genes, followed by RNA labelling and hybridisation protocols as recommended by the manufacturer. After array scanning, quality control was performed with the GCOS software (Affymetrix, USA) according to the manufacturer's recommendations.

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 (Irizarry *et al.*, 2003).

Differentially expressed genes (DEGs) were acquired using two algorithms: SAM (Significance Analysis of Microarray) and RankProd (RankProducts), both included in the MeV (MultiExperiment Viewer) software.

SAM is a t-test based method in which only mean differences are taken into account in DEG selection (Tusher *et al*., 2001). In contrast, RankProd is a rank and permutation-based algorithm and it is able to identify differences within groups and subgroups under analysis (Breitling *et al*., 2004). Genes selected by RankProd do not necessarily exhibit homogeneous expression levels within test and control groups, thus being suitable for detecting differential expression in complex diseases (Bueno & Sunaga *et al*., 2010, Raj *et al*., 2010). Therefore, due the complexity of the disease studied in this work, we decided to appoint as a DEG molecules selected by both algorithms. Since SAM is a very strict method, its p-value threshold was 0.05 while RankProd's was 0.0001.

Transcriptome analysis

We performed functional annotation and network analysis using IPA (Ingenuity Pathway Analysis, Ingenuity Systems - 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.

Clustering procedures were performed using EXPANDER (EXpression Analyzer and DisplayER - http://acgt.cs.tau.ac.il/expander/), with Spearman's correlation as the distance metric. Unsupervised clustering of the differentially expressed genes was performed with the k-means method (k=10). For the similarity clustering, 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. For pathway enrichment analysis we used KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathways, with all genes as the background, and Bonferroni-adjusted p-value <0.01. Transcription factor binding site prediction was performed with PRIMA (PRomoter Integration in Microarray Analysis), with hit range of –1000 to +200, all genes as background, and Bonferroni-adjusted p-value <0.05. These tools are also available in the EXPANDER software.

FANTOM 4 (Functional ANnoTation Of the Mammalian genome - http://fantom.gsc.riken.jp/4/) queries considered only ChIP-chip published data.

Quantitative Real-Time PCR

Two micrograms of total RNA extracted from each cell culture were converted into cDNA using Superscript II (Invitrogen), according to the manufacturer's recommendations. Quantitative Real-Time PCR (qRT-PCR) reactions were performed in duplicates with final volume of 25 µL, using 20 ng cDNA, 2X SYBR Green PCR Master Mix (Applied Biosystems), and 100 – 200 nM of each primer. Fluorescence was detected using ABI Prism 7500 Sequence Detection System (Applied Biosystems), under standard temperature protocol. Primer pairs were designed with Primer-BLAST

(http://www.ncbi.nlm.nih.gov/tools/primer-blast/; primer sequences on Supplementary Table II), and their amplification efficiency (E) was determined by 5-point series of 10-fold cDNA dilutions in which $E = 10^{-1/slope}$. Expression of target genes was assessed relatively to a calibrator cDNA pool (3 controls and 3 patients). Finally, GeNorm v3.4 (Vandesompele *et al*., 2002) was used to determine the most stable endogenous controls (*SDHA, HPRT1, HMBS*), and normalisation factors for each sample. The final expression values were calculated according to Pfaffl (2001). To compare the expression levels of genes selected for validation between control and NSCL/P groups, we applied an unpaired Student's t-test with Welch's correction.

RESULTS

Overview of the differentially expressed genes (DEGs)

By comparing the microarray expression data from NSCL/P and control DPSC cultures, we obtained 125 and 210 differentially expressed genes (DEGs) using the SAM and RankProd algorithms, respectively. There was an overlap of 109 genes from both methodologies, which resulted in a final amount of 228 DEGs (72 up-regulated, 156 down-regulated; Supplementary Table III).

First, to functionally characterise the DEGs and determine possible interactions between them, an Ingenuity Pathways Analysis (IPA; Ingenuity Systems) was performed. We observed enrichment with several genes in regard to the following functions: DNA Replication, Recombination, and Repair (26 genes; p=4.59E-06 - 2.09E-02); Cell Cycle (32 genes; p=1.49E-05 - 2.13E-02);

Cellular Compromise (11 genes; p=6.50E-05 - 2.13E-02) (Supplementary Table IV). Moreover, the most representative network assembled by IPA (containing 27 DEGs) was associated with these functions (Fig. 1).

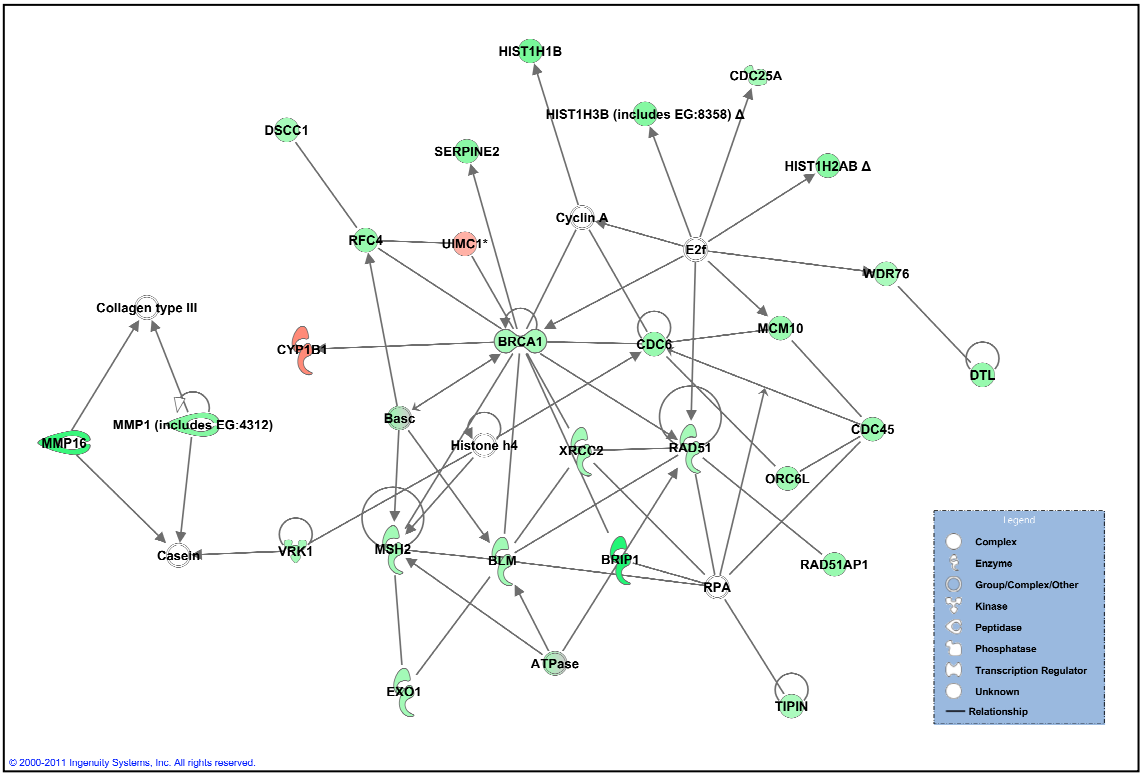


Figure 1: IPA Network 1. Several differentially expressed genes were used to assemble a functional association map. Solid lines indicate direct reationships among genes. This network was associated with 'DNA Replication, Recombination, and Repair'; ' Cell Cycle'; and 'Cellular Compromise' functions. Lines with arrows indicate that one gene acts on the other, and lines without arrows indicate that the corresponding proteins interact with each other. Up-regulated genes are shown in red, and down-regulated genes are shown in green. Blank nodes represent molecules without differential expression.

Cell cycle checkpoint and BRCA1-mediated DNA repair pathways are dysregulated in NSCL/P cells

In order to assess the expression behaviour of DEGs across samples, we performed a clustering analysis using k-means. Ten gene clusters were generated, among which 4 representative clusters exhibited high average homogeneity (h > 0.8; Fig. 1, supplementary material). Assuming that co-expressed genes could be functionally related or partake in the same biological process, we used TANGO (Tool for ANalysis of Gene Ontology enrichment) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis to investigate if there were functional classifications and pathways enriched for each expression cluster. Cluster 4 exhibited a significant enrichment for categories mainly related to cell cycle control and DNA damage repair (corrected p-value <0.01; Supplementary Tables V and VI). Moreover, 16 genes from the aforementioned cluster are present in Network 1 (data not shown), supporting the existence of functional interplay among these genes.

Next, to further investigate the biological attributes and pathways associated with the expression patterns obtained thus far, we used IPA to search for enrichment of canonical pathways related to the progression of cell cycle and repair of DNA damage. Strikingly, pathways encompassing cell cycle checkpoint regulation and DNA repair functions, including the *BRCA1* pathway, were found to be significantly enriched in our gene set (Fig. 2, Supplementary Table VII). Furthermore, network overlay revealed a great overlap between these canonical pathways and interactants of Network 1 (Fig.3).

Following our preliminary findings regarding the enrichment of cell cycle and DNA repair pathways, we assembled a similarity-based cluster in order to fine-group genes with similar transcript levels across samples. Hence, based on the expression pattern of *BRCA1*, we obtained a homogeneous cluster harbouring 30 genes of similar expression (average homogeneity = 0.974; Fig. 4). In accordance with the results thus far, this cluster exhibited GO and KEGG enrichment for cell cycle and DNA repair mechanisms (Tables I and II). There was a large overlap between genes from this cluster, Network 1, and *BRCA1* and checkpoint control canonical pathways (Fig. 3).

Table I: Gene Ontology (GO) attributes significantly enriched in the *BRCA1* similarity cluster (adjusted p-value = 0.001).

GO category	#genes	raw p-value	frequency in set (%)	gene list
DNA metabolic process - GO:0006259	15	2.72E-20	50	*[EXO1, GINS1, XRCC2, RAD51AP1, BLM, DTL, MSH2, MCM10, BRCA1, RAD51, PRIM1, FANCL, CDC45L, ORC6L, HELLS]*
DNA replication - GO:0006260	9	4.91E-14	30	*[GINS1, PRIM1, CDC45L, BLM, DTL, ORC6L, MCM10, BRCA1, RAD51]*
response to DNA damage stimulus - GO:0006974	10	3.02E-13	33.33	*[EXO1, FANCL, RAD51AP1, XRCC2, BLM, DTL, MSH2, ATAD5, BRCA1, RAD51]*
DNA recombination - GO:0006310	7	3.72E-12	23.33	*[EXO1, RAD51AP1, XRCC2, BLM, MSH2, BRCA1, RAD51]*
cell cycle - GO:0007049	11	1.08E-11	36.66	*[EXO1, CCNE2, CDC6, CDC45L, XRCC2, BLM, DSN1, MSH2, BRCA1, HELLS, RAD51]*
cell cycle checkpoint - GO:0000075	6	2.50E-11	20	*[CCNE2, CDC6, CDC45L, BLM, MSH2, BRCA1]*
regulation of DNA metabolic process - GO:0051052	6	1.21E-10	20	*[CDC6, CDC45L, BLM, MSH2, BRCA1, RAD51]*
recombinational repair - GO:0000725	4	4.96E-10	13.33	*[RAD51AP1, BLM, BRCA1, RAD51]*
double-stranded DNA binding - GO:0003690	4	1.44E-07	13.33	*[RAD51AP1, BLM, MSH2, RAD51]*
single-stranded DNA binding - GO:0003697	4	1.44E-07	13.33	*[RAD51AP1, BLM, MSH2, RAD51]*
nucleoside-triphosphatase activity - GO:0017111	8	3.13E-07	26.66	*[CDC6, XRCC2, BLM, MSH2, FIGNL1, ATAD5, HELLS, RAD51]*
negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process - GO:0045934	6	1.16E-06	20	*[CDC6, CDC45L, BLM, MSH2, BRCA1, HELLS]*

Table II: KEGG pathways significantly enriched in the *BRCA1* expression similarity cluster (Bonferroni-corrected p-value < 0.01).

KEGG pathway	#genes	raw p-value	corrected p-value	enrichment factor	gene list
Homologous recombination	3	3.40E-06	9.86E-05	101.538	*[XRCC2, BLM, RAD51]*
Cell cycle	4	1.00E-05	0.001	29.385	*[CCNE2, CDC6, CDC45L, ORC6L]*
Mismatch repair	2	2.68E-04	0.006	82.407	*[EXO1, MSH2]*

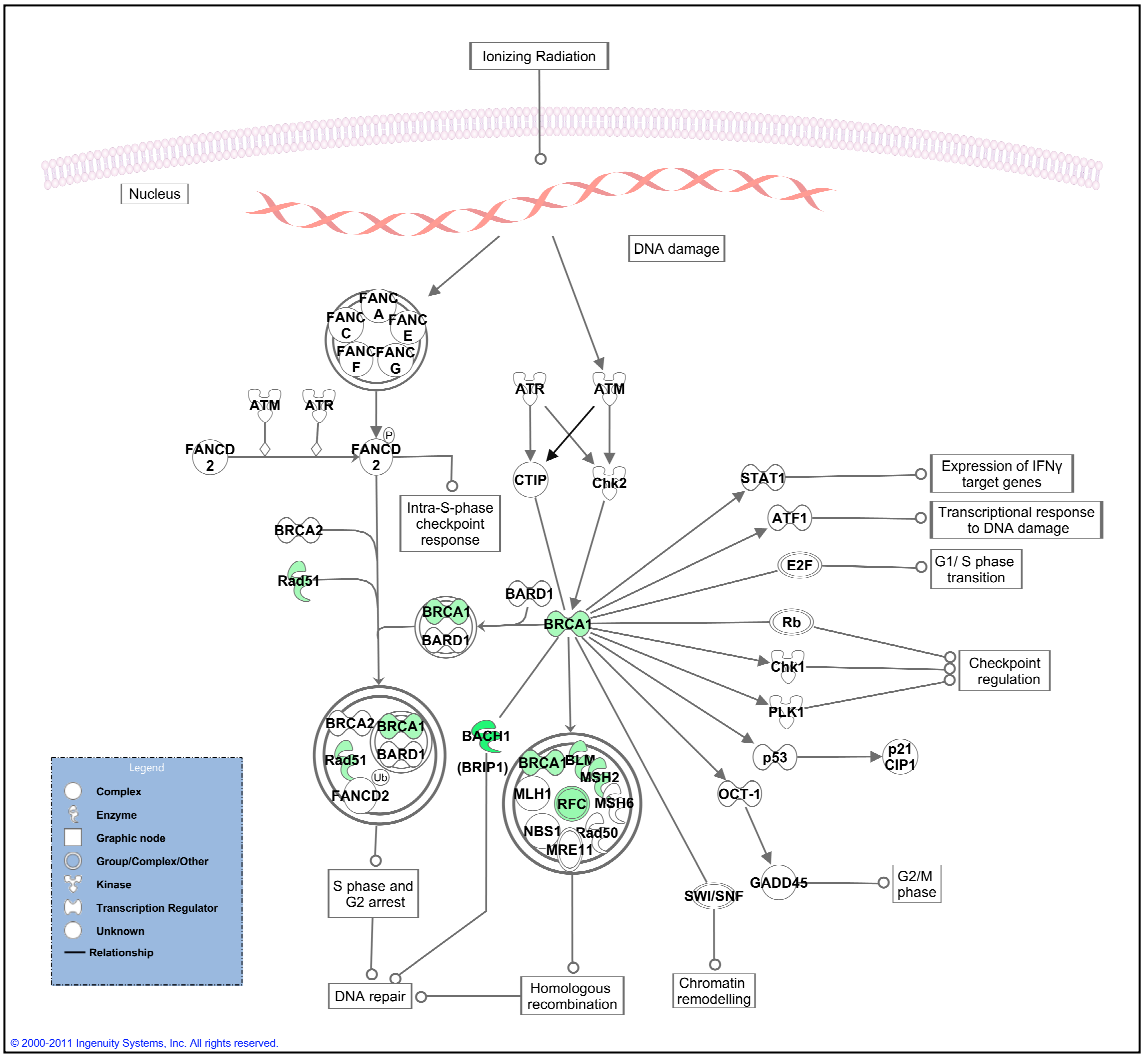


Figure 2: 'Role of *BRCA1* in DNA damage response' canonical pathway. Up-regulated genes are shown in red, and down-regulated genes are shown in green. Blank nodes represent molecules without differential expression.

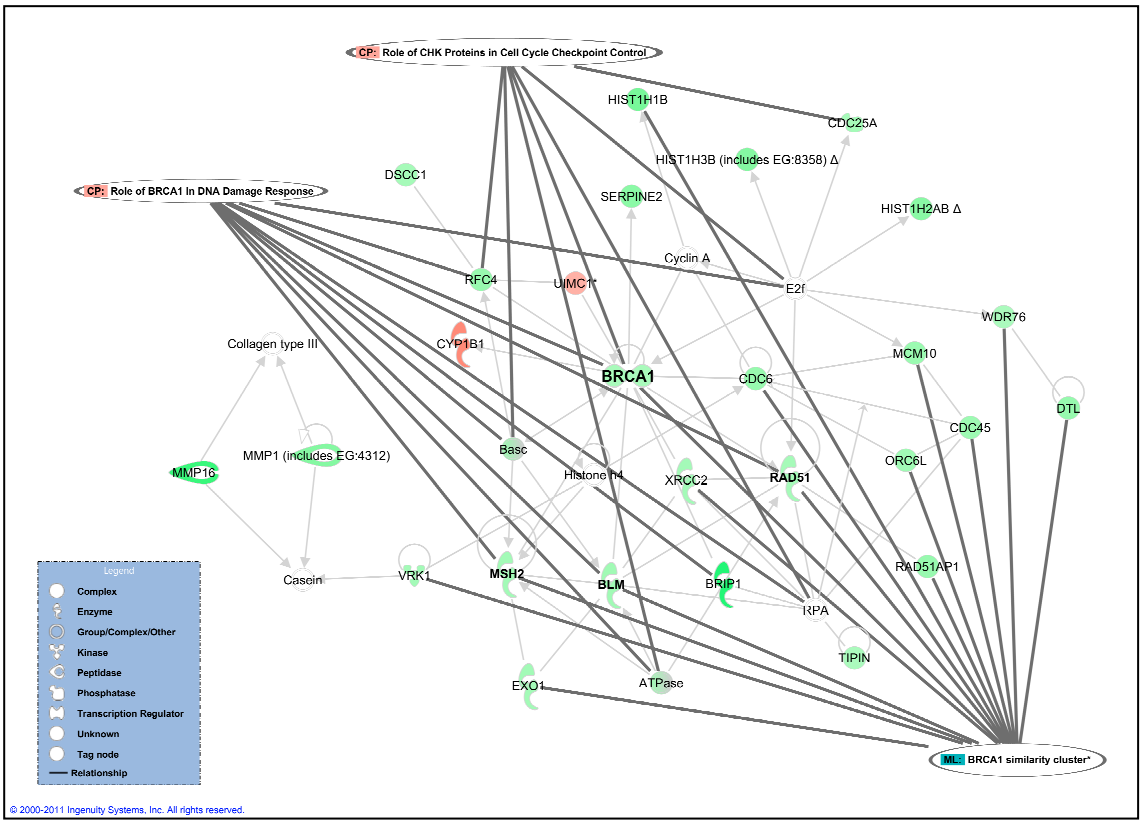


Figure 3: Network 1 overlay of *BRCA1*-dependent DNA repair and cell cycle checkpoint control canonical pathways, and genes from the *BRCA1*-based similarity cluster, revealing a great overlap among these sets. Dark lines indicate co-occurrence of molecules in the sets.

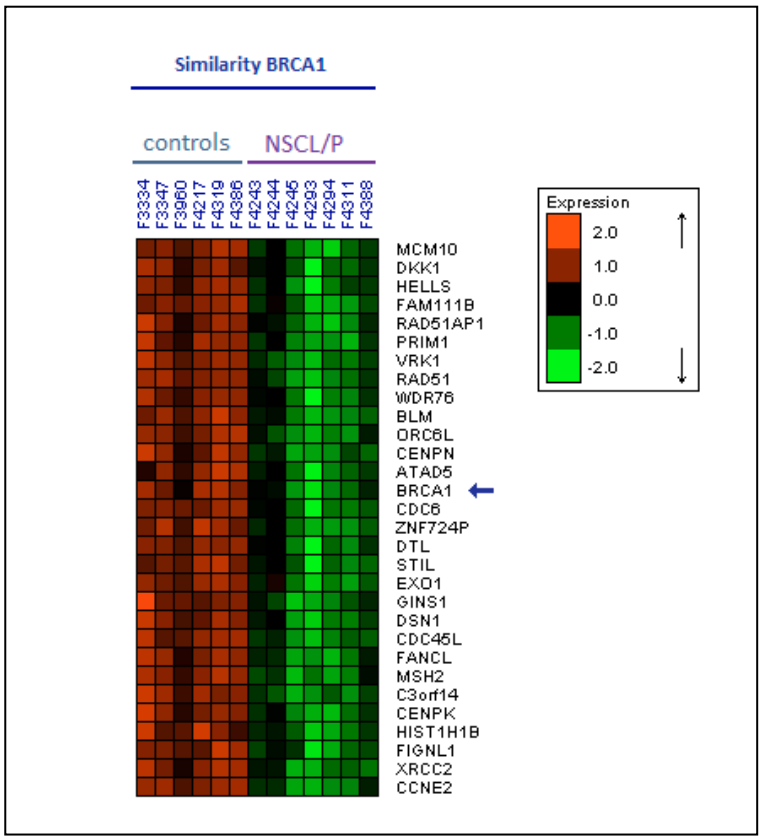


Figure 4: Supervised cluster based on *BRCA1* expression, grouping genes with highly similar patterns among samples (average homogeneity = 0.974).

E2F1 is a candidate regulator driving the expression patterns

In addition to participating in the same biological process, it is accepted that co-expressed genes are likely to be regulated by a cohesive mechanism (Allocco *et al*., 2004). Therefore, we used PRIMA (PRomoter Integration in Microarray Analysis) to perform a transcription factor binding site prediction and enrichment analysis on the highly homogeneous *BRCA1* similarity cluster. This led to the identification of transcription factor *E2F1*, whose binding sites were significantly over-represented in this gene set (12 targets; Bonferroni-corrected p-value <0.05, Table III). Next, in order to validate this *in silico* approach, we used FANTOM 4 (Functional ANnoTation of the Mammalian genome), a database containing experimental data (e.g. ChIP-chip) of mouse and human genomes. Strikingly, *E2F1* is a proven regulator of 23 out of all 30 clustered genes (Fig. 5).

Table III: Transcription factor binding site prediction analysis. Binding sites for *E2F1* motifs significantly over-represented in the *BRCA1* similarity cluster, identified by the PRIMA algorith. Numbers in parenthesis indicate the localisation of the binding site, relative to transcription start site (Bonferroni-corrected p-value < 0.05).

Enriched with	#genes	corrected p-value	enrichment factor	gene list
M00430[E2F-1]	7	0.025	7.687	*[MCM10(-18), CDC6(-43), RAD51(4), FAM111B(-672), CDC45L(182), BLM(-392), DTL(-37)]*
M00940[E2F-1]	8	0.049	4.122	*[HELLS(-38), STIL(-12), C3orf14(-139), CDC6(-9), MSH2(0), FAM111B(-505,-53), ATAD5(-25), DTL(-50,-38)]*

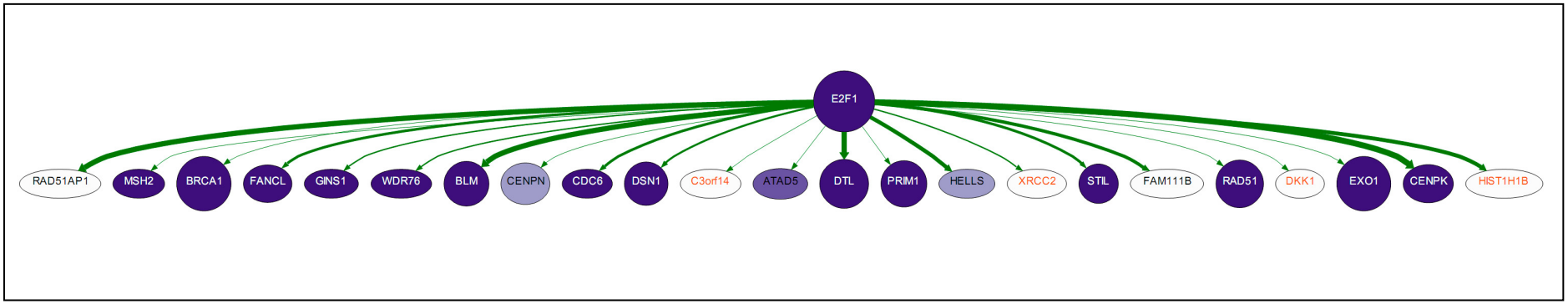


Figure 5: Analysis of transcription factor-gene interactions. FANTOM 4 (a ChIP-chip database) was used to validate the interaction of *E2F1* and 23 of the 30 genes from the *BRCA1*-based similarity cluster. Thickness of the arrows indicates how often the interaction was detected experimentally. The colour of the node represents the level of expression detected in experiments from the database, from light (low) to dark (high).

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

Twenty-four DEGs were submitted to validation by qRT-PCR. Taking account of the fact that the RNA aliquots used for qRT-PCR did not correspond to the original samples hybridised on the microarray chips, we observed that one NSCL/P sample (F4243.1) exhibited a discordant expression pattern as compared to the rest of the NSCL/P group, for 15 out of the 24 genes. The expression values were distant at least 3 standard deviations from the mean, thus, this sample was classified as an outlier, and excluded from the subsequent analyses. Differences in expression values between NSCL/P and control samples were tested for statistical significance, and the mRNA levels of 16 out of the 24 DEGs selected for validation were significantly different between groups (Student's t-test; $p<0.05$; Supplementary Table VIII). Notably, we confirmed the differential expression of all genes that were submitted to validation, and that also pertained to the *BRCA1* similarity cluster (10 genes, Fig. 6).

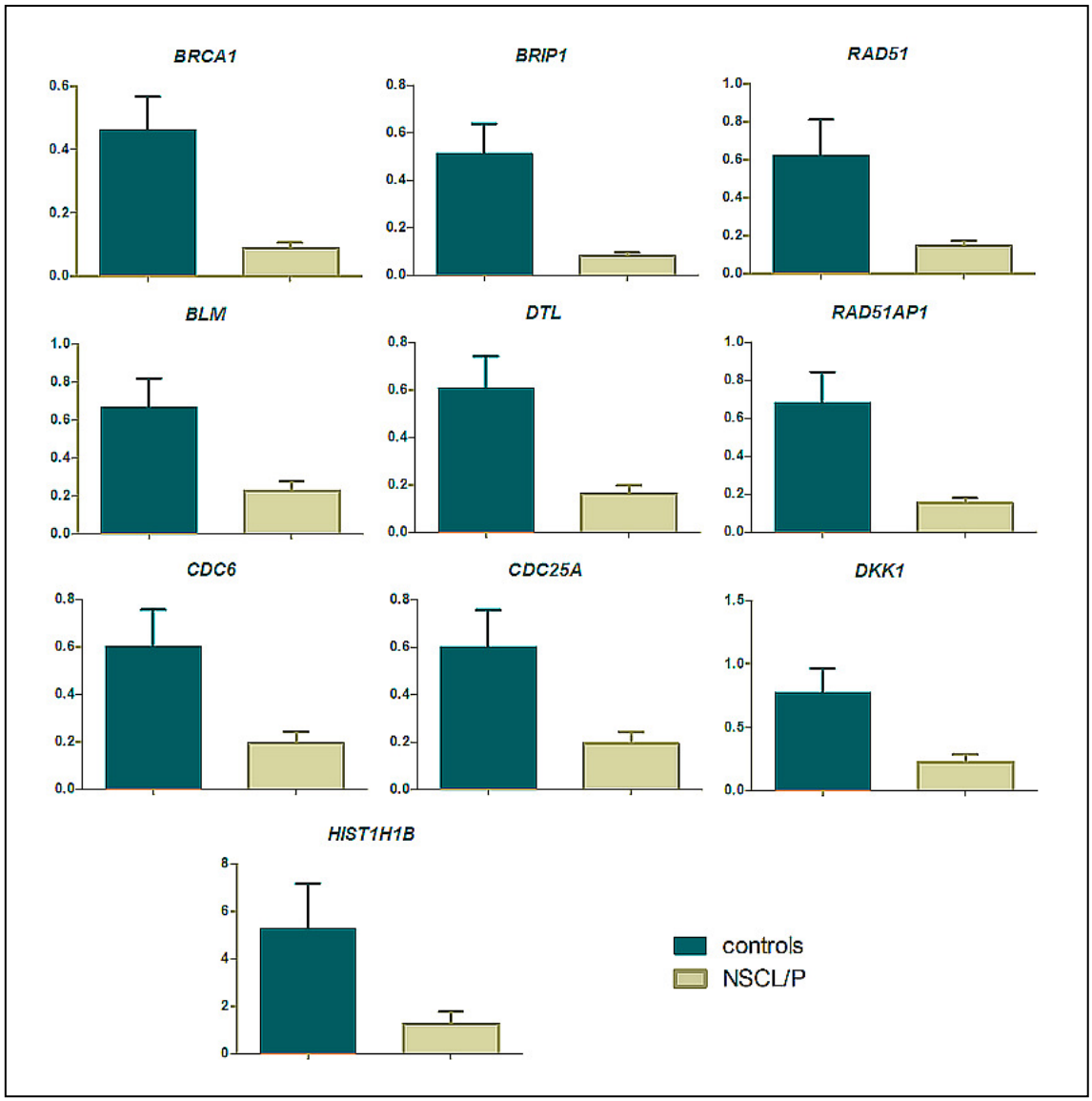


Figure 6: Quantitative real-time PCR validation of 10 genes from the *BRCA1* similarity cluster. Expression levels confirmed the microarray results, with p-value < 0.05. Y axis = relative units of expression, normalised to endogenous controls.

DISCUSSION

We have previously demonstrated the applicability of dental pulp stem cells from NSCL/P patients to investigate gene networks associated with disease pathology (Bueno & Sunaga *et al*., 2010). Here, we also identified an expression profile that is particular to the NSCL/P cells, comprising 228 DEGs. We identified 9 genes in common with our previous microarray analysis (including *CLDN11*, *ACAN*, *ITGA8*, *MMP1*, *HIST1H2AB*, and *PSG1*), which supports the findings by Bueno and colleagues. However, the microarray analysis depicted here resulted in a significantly larger amount of DEGs (2567 vs. 87, data not shown) when the data mining was performed using the same settings (RankProd alone, $p<0.05$). The fact that, even under more stringent settings (applying SAM; $p<0.05$ and RankProd; $p<0.0001$), we still obtained a larger amount of DEGs, suggests that different culture conditions could have induced different responses from NSCL/P cells, being that full serum supplementation presumably led to the manifestation of such expression profile. This favours the hypothesis that the genome of NSCL/P patients might harbour specific variants or epigenetic modifications that respond differently to the environment as compared to non-clefting individuals. Hence, we speculate that different environmental conditions, simulated by serum supplementation in this case, may drive different responses from NSCL/P dental pulp stem cells. If that is the case, then these cells express not only a disease-specific expression pattern, but are also subject to an environment-specific influence on such pattern.

IPA and clustering analyses enabled us to identify functions and pathways associated with the observed expression arrangement among DEGs.

Preliminary unsupervised clustering combined with IPA network assembly suggested enrichment of functional categories mainly related to DNA repair and cell cycle control, with the *BRCA1* node occupying a central position in Network 1. Accordingly, in-depth analysis of this network revealed a great overlap between *BRCA1* and cell cycle checkpoint control canonical pathways. The gene *BRCA1*, down-regulated in our analysis, encodes a tumour suppressor protein that plays a crucial role in the cellular response to DNA damage and regulation of cell cycle progression, acting to maintain the integrity of the genome (Linger & Kruk, 2010). Hence, if co-expressed genes are likely to participate in the same biological process, the attainment of a cluster of higher resolution, based on expression similarity to this central molecule, would allow us to determine which genes are the most accountable for dysregulation of *BRCA1*-mediated DNA repair mechanisms and checkpoint control in these cells. Indeed, many genes accompanying *BRCA1* expression patterns have well-established roles in cell cycle/checkpoint control and response to DNA damage (e.g. *MSH2*, *BLM*, *RAD51*, CDC6, *CDC25A*; Seifert & Reichrath, 2006; Yata & Esashi, 2009). Moreover, the relevance of these genes in this context is further corroborated by the large co-occurrence of members of this similarity cluster and interactants of Network 1.

Besides allowing for a more precise dissection of the relationship between the NSCL/P transcriptome and biological processes, the aforementioned strategy provided us a unique opportunity to detect transcription factors that could modulate the activity of the grouped genes. This led to the identification of *E2F1* as a candidate transcriptional regulator of *BRCA1* and its accompanying DEGs, implying that functional impairment of this

transcription factor is likely to be responsible for their abnormal expression levels. *E2F1* is a key molecule that controls the activation of several genes that contribute to G1/S phase progression in the cell cycle (Chen *et al*., 2009), including cyclins, cyclin-dependent kinases, and checkpoint genes (Inoshita *et al*., 1999; Deng *et al*., 2006; Yang *et al*., 2008), while also promoting apoptosis control (Bell & Ryan, 2003). Notably, a role for *E2F1* in palatogenesis has been proposed through immunobloting using mouse embryonic tissues (Kusek *et al*., 2000). Thus, in the context of NSCL/P, it is plausible to infer that disarrangement of cell cycle progression and *BRCA1*-mediated DNA repair could contribute to the cleft phenotype, as accurate control of cell proliferation is necessary to ensure appropriate growth of embryonic structures responsible for lip and palate formation (Muraoka *et al*., 2005; Jiang *et al*., 2006). If true, this relationship exemplifies how genetic susceptibility to a complex disease (i.e. involving different *loci*) may converge onto undermined regulation of a single biological pathway. Conversely, we acknowledge that the involvement of *E2F1/BRCA1* mechanisms in the aetiology NSCL/P cannot be confirmed as yet; nevertheless, the relationship between such mechanisms and NSCL/P proposed here is supported by epidemiological findings, as discussed next.

Pathogenic mutations in *BRCA1* are associated with increased risk of developing a diversity of hereditary tumour types, including breast, ovarian, pancreatic, and prostate cancer (Yarden & Papa, 2006; Rosen *et al*., 2006). Additionally, reduced levels of *BRCA1* mRNA and protein has been reported in sporadic forms of breast cancer, suggesting that compromised function of this gene by decreased expression is a critical step in cancer progression (Thompson *et al*., 1995; Lee *et al*., 1999; Mueller *et al*., 2003). Notably,

association between occurrence of NSCL/P and a variety of cancer types, including breast cancer, has been repeatedly reported (Zhu *et al*., 2002; Menezes *et al*., 2009; Taioli *et al*., 2010). Our results support this association, as they indicate that many genes partaking in *BRCA1* response to DNA damage are down-regulated in NSCL/P cells. Additionally, non-hereditary cancer progression has also been ascribed to alterations in some of these genes, such as *CDC6*, *BLM*, *RAD51*, *MSH2* (Borlado & Mendes, 2008; Calin *et al*., 1998; Gao *et al*., 2011; Seifert & Reichrath, 2006), among others. Finally, loss of pro-apoptotic activity, or exacerbated proliferation modulated by *E2F1* have also been implicated in cancer biology (Bell & Ryan, 2003). These findings suggest that, if abnormal behaviour of *E2F1* and/or transcriptional dysregulation of *BRCA1* and its co-operators contribute to disruption of lip and palate morphogenesis in the embryo, such mechanisms may also predispose to cancer during adulthood.

In summary, global gene expression analysis of NSCL/P dental pulp stem cells identified dysregulation of genes pertaining to cell cycle checkpoint and *BRCA1*-dependent DNA repair pathways. To the best of our knowledge, this is the first experimental evidence linking transcriptional alterations of these genes to the documented increase in occurrence of sporadic cancer in NSCL/P families. To explain this association, much effort has been put in the search for polymorphisms in genes linked to cancer. However, in the light of our results, the focus must instead be directed at upstream regulatory elements, and not their downstream targets. In order to confirm that deficient function of *E2F1* and/or *BRCA1*/checkpoint pathways are mechanisms underlying susceptibility to NSCL/P *de facto*, future research should also concentrate on examining the

phenotypic outcome of these alterations in NSCL/P cells, and on subsequent study of animal embryo models. In addition, if cell culture conditions have such an impact on the transcriptome of NSCL/P cells, these studies must also assess the influence of environmental insults, especially by teratogens that are known to predispose to clefting. Finally, further dissection of upstream regulatory pathways will be required, both to confirm the involvement of *E2F1* in the pathophysiology of NSCL/P, and to identify other possible regulatory agents that, when altered, may confer predisposition to this orofacial malformation.

SUPPLEMENTARY MATERIAL

Supplementary Figures

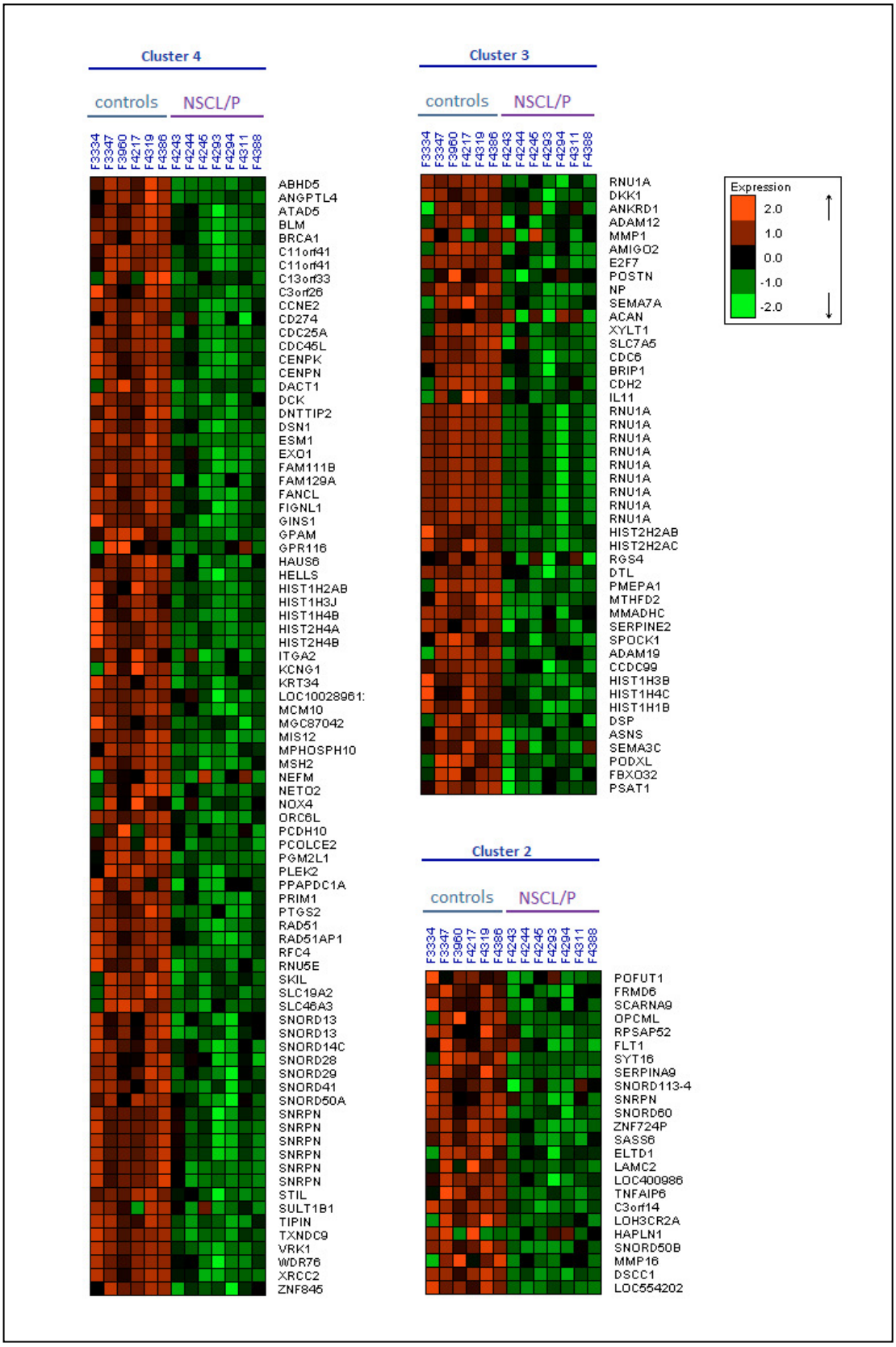


Figure 1: Three representative clusters generated by the k-means method, exhibiting high homogeneity levels (h>0.8)

Supplementary Tables

Table I: Age, gender and clinical status of samples analysed by microarray and qRT-PCR

Code	Age	Gender	Clinical Status*
F3334	8	female	control
F3347	10	male	control
F3960	10	female	control
F4217	12	male	control
F4319	11	female	control
F4386	10	female	control
F4243	8	male	CLP/UL
F4244	13	male	CLP/UL
F4245	11	male	CLP/UR
F4293	12	female	CL/UL
F4294	10	female	CLP/UR
F4311	11	male	CL/UL
F4388	12	male	CLP/UR

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

Table II: Primer sequences used for qRT-PCR.

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	GGCAATGCAGATGCAGCTTGAAG T	TTATGCCACACTGCTCTAACCGTG
RAD51AP1	GCAGTGCCTTGTACAAAGATGGCT	GTGGTGACTGTTGGAAGTTCCTTCA

Table III: List of differentially expressed genes obtained by comparing NSCL/P and control cells.

Probe Set ID	Gene Symbol	FC*	Cytoband	Probe Set ID	Gene Symbol	FC*	Cytoband
8112045	*ESM1*	-5.5397	5q11.2	7981084	*SERPINA9*	-2.4325	14q32.13
8126798	*GPR116*	-4.3447	6p12.3	8100798	*SULT1B1*	-2.4124	4q13.3
7906919	*RGS4*	-3.9706	1q23.3	8124388	*HIST1H3B*	-2.3586	6p21.3
7930980	*PPAPDC1A*	-3.8212	10q26.12	7939215	*C11orf41*	-2.3585	11p13
8140534	*SEMA3C*	-3.8175	7q21-q31	7996260	*POFUT1*	-2.3504	---
7908072	*LAMC2*	-3.7681	1q25-q31	8145793	*SNORD13*	-2.3477	8p12
7990345	*SEMA7A*	-3.6893	15q22.3-q23	7951271	*MMP1*	-2.3419	11q22.3
8142975	---	-3.6874	---	7985786	*ACAN*	-2.3042	15q26.1
7981990	*SNORD116-21*	-3.6830	15q11.2	7948902	*SNORD29*	-2.2990	11q13
7922976	*PTGS2*	-3.6406	1q25.2-q25.3	8124391	*HIST1H2AB*	-2.2738	6p21.3
8097449	*PCDH10*	-3.5370	4q28.3	8059376	*SERPINE2*	-2.2573	2q33-q35
8067233	*PMEPA1*	-3.5160	20q13.31|20q13.31-q13.33	7936968	*ADAM12*	-2.2565	10q26.3
8127987	*SNORD50A*	-3.4841	6q14.3	7897801	*RNU5E*	-2.2447	1p36.22
8117368	*HIST1H4C*	-3.4817	6p21.3	7919642	*HIST2H2AB*	-2.1893	1q21
7934979	*ANKRD1*	-3.4591	10q23.31	7922162	*SLC19A2*	-2.1662	1q23.3
8152703	*FBXO32*	-3.4433	8q24.13	7950391	*PGM2L1*	-2.1467	11q13.4
8017262	*BRIP1*	-3.4328	17q22-q24	8086752	*SNORD13*	-2.1292	8p12
7914878	*LOC100289612*	-3.4174	1p34.3	8160431	*LOC554202*	-2.1183	9p21.3
8115490	*ADAM19*	-3.3659	5q32-q33	8105267	*ITGA2*	-2.1111	5q23-q31
8142981	*PODXL*	-3.3573	7q32-q33	8083876	*SKIL*	-2.1046	3q26
8067029	*KCNG1*	-3.3543	20q13	8138527	*MGC87042*	-2.0841	7p15.3
8151684	*MMP16*	-3.2960	8q21.3	8156043	*PSAT1*	-2.0621	9q21.2
8001197	*NETO2*	-3.2835	16q11	7909568	*DTL*	-2.0604	1q32.1-q32.2
7970763	*FLT1*	-3.2679	13q12	8081241	*C3orf26*	-2.0585	3q12.1
7962579	*AMIGO2*	-3.2461	12q13.11	7953218	*RAD51AP1*	-2.0309	12p13.2-p13.1
7971077	*POSTN*	-3.2279	13q13.3	8092640	*RFC4*	-2.0249	3q27
7968351	*C13orf33*	-3.2224	13q12.3	8007071	*CDC6*	-2.0144	17q21.3
7950933	*NOX4*	-3.2217	11q14.2-q21	8124385	*HIST1H4B*	-2.0124	6p21.3
7917182	*ELTD1*	-3.1215	1p33-p32	7943158	*SCARNA9*	-2.0099	11q21
8039484	*IL11*	-3.0789	19q13.3-q13.4	8080847	*C3orf14*	-2.0073	3p14.2
7952785	*OPCML*	-3.0484	11q25	7974882	*SYT16*	-2.0047	14q23.2
7915612	*PTCH2*	-2.9210	1p33-p34	8035838	*ZNF724P*	-1.9868	19p12
8116780	*DSP*	-2.8423	6p24	7998722	*SNORD60*	-1.9828	16p13.3
7924461	---	-2.8344	---	8025402	*ANGPTL4*	-1.9752	19p13.3
8091243	*PCOLCE2*	-2.7404	3q21-q24	7940147	*FAM111B*	-1.9684	11q12.1
8045688	*TNFAIP6*	-2.6672	2q23.3	8112376	*CENPK*	-1.9664	5p15.2-q12.3
8145361	*NEFM*	-2.6167	8p21	8022674	*CDH2*	-1.9646	18q11.2
7974689	*DACT1*	-2.5540	14q23.1	7926259	*MCM10*	-1.9639	10p13
8124527	*HIST1H1B*	-2.5535	6p22-p21.3	7979710	*PLEK2*	-1.9562	14q23.3
7973067	*NP*	-2.5424	14q13.1	8071212	*CDC45L*	-1.9326	22q11.21

Table III (continued)

Probe Set ID	Gene Symbol	FC*	Cytoband	Probe Set ID	Gene Symbol	FC*	Cytoband
8077499	*LOH3CR2A*	-2.5084	3p25.3	7964733	*RPSAP52*	-1.9311	12q14.3
8180255	*HIST2H4B*	-2.4964	---	7981978	*SNORD116-15*	-1.9238	15q11.2
8180321	*HIST2H4A*	-2.4964	---	8061471	*GINS1*	-1.9215	20p11.21
7939237	*C11orf41*	-2.4825	11p13	8141150	*ASNS*	-1.9199	7q21.3
8151871	*CCNE2*	-2.4690	8q22.1	8112971	*HAPLN1*	-1.9168	5q14.3
7970793	*SLC46A3*	-2.4625	13q12.3	8004144	*MIS12*	-1.9120	17p13.2
7999754	*XYLT1*	-2.4398	16p12.3	8015268	*KRT34*	-1.9050	17q12-q21
8124537	*HIST1H3J*	-1.8990	6p22-p21.3	7981966	*SNORD116-3*	-1.7342	15q11.2
8154233	*CD274*	-1.8974	9p24	8015769	*BRCA1*	-1.7286	17q21
8127989	*SNORD50B*	-1.8879	6q14.3	7905088	*HIST2H2AC*	-1.7277	1q21-q23
8079153	*ABHD5*	-1.8874	3p21	7974337	*FRMD6*	-1.7244	---
7981982	*SNORD116-17*	-1.8814	15q11.2	7936322	*GPAM*	-1.7180	10q25.2
7981986	*SNORD116-17*	-1.8814	15q11.2	8054217	*TXNDC9*	-1.7145	2q11.2
7995354	*ORC6L*	-1.8720	16q12	8042588	*MPHOSPH10*	-1.7036	2p13.3
7986068	*BLM*	-1.8707	15q26.1	8003298	*SLC7A5*	-1.6987	16q24.3
7973896	*RNU1A*	-1.8653	1p36.1	8053797	*LOC400986*	-1.6909	2q11.1
7978568	*RNU1A*	-1.8653	1p36.1	7989915	*TIPIN*	-1.6836	15q22.31
7929438	*HELLS*	-1.8483	10q24.2	7983306	*WDR76*	-1.6800	15q15.3
8095574	*DCK*	-1.8462	4q13.3-q21.1	8055672	*MMADHC*	-1.6732	2q23.2
8030978	*ZNF845*	-1.8454	19q13.42	7922846	*FAM129A*	-1.6557	1q25
7919269	*RNU1A*	-1.8445	1p36.1	8066074	*DSN1*	-1.6440	20q11.23
7919349	*RNU1A*	-1.8445	1p36.1	8109830	*CCDC99*	-1.6380	5q35.1
7927631	*DKK1*	-1.8428	10q11.2	7976567	*BDKRB1*	2.1870	14q32.1-q32.2
7964271	*PRIM1*	-1.8378	12q13	8045533	---	2.5570	---
8041867	*MSH2*	-1.8289	2p22-p21	8015179	*KRTAP1-5*	2.6106	17q12-q21
7917771	*DNTTIP2*	-1.8218	1p22.1	8019588	*KRTAP1-5*	2.6106	17q12-q21
7910997	*EXO1*	-1.8166	1q42-q43	7903214	*LPPR4*	2.6232	1p21.2
8052382	*FANCL*	-1.8152	2p16.1	8176719	*EIF1AY*	2.6602	Yq11.223
7976812	*SNORD113-4*	-1.8131	14q32.31	8176375	*RPS4Y1*	2.6958	Yp11.3
7976621	*VRK1*	-1.8091	14q32	7999909	*GPRC5B*	2.7015	16p12
8144036	*XRCC2*	-1.8037	7q36.1	7938225	*OLFML1*	2.7143	11p15.4
8114287	*SPOCK1*	-1.7965	5q31	7969202	---	2.7578	---
7915926	*STIL*	-1.7934	1q32|1p32	8176655	*NLGN4Y*	2.7629	Yq11.221
8102787	*RNU1A*	-1.7920	---	8037240	*PSG1*	2.7901	19q13.2
7952339	*SNORD14C*	-1.7901	11q23.3-q25	8013521	---	2.7933	---
7982792	*RAD51*	-1.7873	15q15.1	8097288	*FAT4*	2.8028	4q28.1
7898375	*RNU1A*	-1.7849	1p36.1	7934185	*C10orf54*	2.8031	10q22.1
7898411	*RNU1A*	-1.7849	1p36.1	7950005	*MRGPRF*	2.8157	11q13.3
7912800	*RNU1A*	-1.7849	1p36.1	8068361	*SLC5A3*	2.8511	21q22.12
7912850	*RNU1A*	-1.7849	1p36.1	8104758	*C5orf23*	2.8643	5p13.3

Table III (continued)

Probe Set ID	Gene Symbol	FC*	Cytoband	Probe Set ID	Gene Symbol	FC*	Cytoband
7919576	*RNU1A*	-1.7849	1p36.1	7903358	*VCAM1*	2.8833	1p32-p31
8152582	*DSCC1*	-1.7790	8q24.12	8058664	---	2.9007	---
7981976	*SNORD116-14*	-1.7779	15q11.2	8037657	*DMPK*	2.9540	19q13.3
8086880	*CDC25A*	-1.7774	3p21	7896748	---	2.9701	---
7948904	*SNORD28*	-1.7742	11q13	7911335	---	2.9706	---
8132843	*HAUS6*	-1.7591	9p22.1	8165694	---	2.9706	---
8006187	*ATAD5*	-1.7583	17q11.2	8176578	*USP9Y*	2.9865	Yq11.2
8139632	*FIGNL1*	-1.7576	7p12.1	7975076	*HSPA2*	3.0061	14q24.1
8042830	*MTHFD2*	-1.7520	2p13.1	8150962	*TOX*	3.1055	8q12.1
7997381	*CENPN*	-1.7446	16q23.2	7932254	*ITGA8*	3.1105	10p13
8034512	*SNORD41*	-1.7412	19p13.2	8045835	*GALNT5*	3.1139	2q24.1
7965094	*E2F7*	-1.7409	12q21.2	8176624	*DDX3Y*	3.1253	Yq11
7917976	*SASS6*	-1.7348	1p21.2	7926875	*BAMBI*	3.1399	10p12.3-p11.2
7981953	*SNORD116-3*	-1.7342	15q11.2	8141140	*DLX5*	3.1609	7q22
8013523	---	3.1638	---	7899615	*SERINC2*	3.5882	1p35.1
7962058	*TMTC1*	3.1704	12p11.22	8165663	---	3.6368	---
7896750	---	3.1802	---	7911337	---	3.6419	---
7984364	*SMAD3*	3.2242	15q22.33	7973871	---	3.6419	---
8156358	---	3.2296	---	8165696	---	3.6419	---
8171297	*MID1*	3.2297	Xp22	8102831	*C4orf49*	3.6651	4q31.1
8113120	*TOB2*	3.2502	22q13.2-q13.31	8165709	---	3.7686	---
8165707	*TOB2*	3.2502	22q13.2-q13.31	7911339	---	3.8696	---
8129666	*SLC2A12*	3.2565	6q23.2	8165698	---	3.8696	---
8112668	*GCNT4*	3.2672	5q12	8045804	---	4.0321	---
7950810	*SYTL2*	3.2894	11q14	8102532	*PDE5A*	4.0664	4q25-q27
8095110	*KIT*	3.2996	4q11-q12	8104746	*NPR3*	4.0886	5p14-p13
7911343	*UIMC1*	3.3096	5q35.2	7985317	*KIAA1199*	4.2513	15q24
8165703	*UIMC1*	3.3096	5q35.2	8057677	*SLC40A1*	4.4972	2q32
7970565	---	3.3203	---	8100310	---	4.6636	---
8008965	---	3.3361	---	7933204	*C10orf10*	4.9268	10q11.21
7962559	*SLC38A4*	3.3364	12q13	8051583	*CYP1B1*	4.9870	2p21
8052355	*EFEMP1*	3.3858	2p16	8083887	*CLDN11*	5.3906	3q26.2-q26.3
8013987	---	3.5398	---	7912537	*DHRS3*	5.4306	1p36.1
8160346	*PTPLAD2*	3.5566	9p21.3	8006433	*CCL2*	7.5308	17q11.2-q12

(*) FC = fold change

Table IV: Ingenuity Pathways Analysis (IPA). Top 10 enriched functions among the 228 differentially expressed genes.

Enriched category	p-value	Molecules
Cell Cycle*	1.28E-06-1.67E-02	*SMAD3,DSN1,UIMC1,ASNS,SASS6,RAD51,VRK1,XRCC2,CCDC99,KIT,MCM10,CD274,BRCA1,BLM,CDC25A,CDC45,STIL,ABHD5,CCNE2,NOX4,FLT1,CDC6,ITGA2,ORC6,TIPIN,HSPA2,CYP1B1,MSH2,HELLS,SKIL,MIS12,PTGS2,IL11*
DNA Replication, Recombination, and Repair*	3.87E-06-1.39E-02	*PRIM1,SMAD3,UIMC1,SASS6,RAD51,XRCC2,KIT,BLM,BRCA1,BRIP1,EXO1,CDC25A,CDC45,CCNE2,NOX4,DCK,CDC6,ORC6,TIPIN,GINS1,RFC4,MSH2,HELLS,MIS12,PTGS2,RAD51AP1*
Carbohydrate Metabolism	4.14E-05-1.35E-02	*SERINC2,GPAM,ABHD5,GALNT5,ACAN,FLT1,SMAD3,XYLT1,RGS4,DKK1,PTGS2,TNFAIP6*
Cellular Compromise*	7.04E-05-9.06E-03	*VCAM1,CDH2,MSH2,ITGA2,KIT,FBXO32,BLM,BRIP1,BRCA1*
Cellular Movement	7.45E-05-1.67E-02	*CLDN11,MMP16,SMAD3,ANGPTL4,CCL2,PCDH10,ADAM19,POSTN,KIT,MMP1,PODXL,VCAM1,NOX4,DCK,ACAN,FLT1,ITGA2,RGS4,LAMC2,CYP1B1,SERPINE2,CDH2,ADAM12,PTGS2,DKK1,SEMA3C,BDKRB1,IL11*
Cell Morphology	8.15E-05-1.1E-02	*SMAD3,FLT1,ITGA2,DLX5,KIT,RGS4,CD274,PTGS2,BRCA1,BLM*
Cellular Function and Maintenance	8.15E-05-9.06E-03	*SMAD3,DKK1,BRCA1,BLM,HSPA2,DSP*
Cellular Assembly and Organization	1.95E-04-1.62E-02	*ABHD5,PODXL,CLDN11,SMAD3,CDC6,SASS6,XRCC2,CDH2,HELLS,KIT,MIS12,BRCA1,BLM,SYTL2,DSP*
Cell-To-Cell Signaling and Interaction	2.82E-04-1.67E-02	*PODXL,VCAM1,CLDN11,MMP16,SMAD3,FLT1,OPCML,ITGA2,LAMC2,CDH2,ADAM12,CCL2,PCDH10,POSTN,KIT,SLC7A5,DKK1,SPOCK1,TNFAIP6,BDKRB1,MMP1,IL11*
Gene Expression	3.74E-04-1.26E-02	*SMAD3,BAMBI,UIMC1,PTGS2,SKIL,BRCA1,IL11*
Cellular Development	4.57E-04-1.61E-02	*SMAD3,ITGA8,CCL2,POSTN,KIT,CD274,BLM,BRCA1,DSP,MMP1,VCAM1,NOX4,DCK,FLT1,ACAN,ITGA2,DLX5,RGS4,HSPA2,MSH2,HELLS,PNP,PDE5A,DKK1,PTGS2,TNFAIP6,TOX,IL11*

(*) Functions associated with IPA Network 1

Table V: Gene Ontology (GO) attribute enrichment for Cluster 4 (adjusted p-value = 0.001). No significant results were retrieved for the remaining clusters.

GO category	#genes	raw p-value	frequency in set (%)	gene list
DNA metabolic process - GO:0006259	17	7.07E-16	20.48	*[EXO1, GINS1, RAD51AP1, XRCC2, BLM, MSH2, TIPIN, MCM10, BRCA1, CDC25A, RAD51, FANCL, PRIM1, CDC45L, RFC4, ORC6L, HELLS]*
DNA replication - GO:0006260	11	4.10E-13	13.25	*[GINS1, PRIM1, CDC45L, RFC4, BLM, ORC6L, TIPIN, MCM10, BRCA1, CDC25A, RAD51]*
response to DNA damage stimulus - GO:0006974	11	3.56E-10	13.25	*[EXO1, FANCL, RAD51AP1, XRCC2, RFC4, BLM, MSH2, TIPIN, ATAD5, BRCA1, RAD51]*
DNA-dependent DNA replication - GO:0006261	6	1.85E-09	7.22	*[PRIM1, CDC45L, RFC4, BLM, TIPIN, RAD51]*
DNA recombination - GO:0006310	7	3.43E-09	8.43	*[EXO1, RAD51AP1, XRCC2, BLM, MSH2, BRCA1, RAD51]*
cell cycle - GO:0007049	13	5.00E-09	15.66	*[EXO1, XRCC2, BLM, MSH2, DSN1, TIPIN, BRCA1, CDC25A, MIS12, RAD51, CCNE2, CDC45L, HELLS]*
cell cycle checkpoint - GO:0000075	6	8.28E-09	7.22	*[CCNE2, CDC45L, BLM, MSH2, TIPIN, BRCA1]*
recombinational repair - GO:0000725	4	2.24E-08	4.81	*[RAD51AP1, BLM, BRCA1, RAD51]*
regulation of DNA metabolic process - GO:0051052	6	3.89E-08	7.22	*[CDC45L, BLM, MSH2, TIPIN, BRCA1, RAD51]*

Table VI: KEGG pathways significantly enriched in clusters derived by k-means (Bonferroni-corrected p-value < 0.01).

Set	KEGG pathway	#genes	raw p-value	corrected p-value	enrichment factor	gene list
Cluster 4	Mismatch repair	3	3.06E-05	0.001	49.106	*[EXO1, RFC4, MSH2]*
Cluster 4	Homologous recombination	3	5.62E-05	0.002	40.337	*[XRCC2, BLM, RAD51]*
Cluster 3	Systemic lupus erythematosus	4	3.54E-05	0.005	21.529	*[HIST2H2AB, HIST2H2AC, HIST1H3B, HIST1H4C]*
Cluster 7	O-Glycan biosynthesis	2	2.31E-04	0.009	88.654	*[GCNT4, GALNT5]*

Table VII: Top 10 canonical pathways enriched in our gene set, detected by Ingenuity Pathway Analysis.

Ingenuity Canonical Pathways	-log(p-value)	Molecules
Role of BRCA1 in DNA Damage Response	4.96E+00	*RAD51,MSH2,RFC4,BRIP1,BLM,BRCA1*
Mismatch Repair in Eukaryotes	3.43E+00	*MSH2,RFC4,EXO1*
Hepatic Fibrosis / Hepatic Stellate Cell Activation	2.76E+00	*VCAM1,CCL2,FLT1,SMAD3, BAMBI,MMP1*
Cell Cycle Control of Chromosomal Replication	2.74E+00	*CDC45,CDC6,ORC6*
Role of CHK Proteins in Cell Cycle Checkpoint Control	2.49E+00	*RFC4,BRCA1,CDC25A*
Hereditary Breast Cancer Signaling	2.34E+00	*RAD51,MSH2,RFC4,BLM,BRCA1*
DNA Double-Strand Break Repair by Homologous Recombination	2.16E+00	*RAD51,BRCA1*
Factors Promoting Cardiogenesis in Vertebrates	2.09E+00	*CCNE2,NOX4,CDC6,DKK1*
ATM Signaling	1.93E+00	*RAD51,BRCA1,CDC25A*
Cell Cycle: G1/S Checkpoint Regulation	1.85E+00	*CCNE2,SMAD3,CDC25A*

Table VIII: Quantitative Real Time-PCR validation of 24 differentially expressed genes detected by microarray. Unpaired t-test with Welch's correction, $p<0.05$.

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

(*) genes pertaining to the *BRCA1* similarity cluster

REFERENCES

1. ALLOCCO, D. J., KOHANE, I. S. & BUTTE, A. J. 2004. Quantifying the relationship between co-expression, co-regulation and gene function. BMC bioinformatics, 5, 18.

2. BARONI, T., BELLUCCI, C., LILLI, C., PEZZETTI, F., CARINCI, F., LUMARE, E., PALMIERI, A., STABELLINI, G. & BODO, M. 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. Journal of cellular physiology, 222, 748-56.

3. BELL, L. A. & RYAN, K. M. 2004. Life and death decisions by E2F-1. Cell death and differentiation, 11, 137-42.

4. BORLADO, L. R. & MENDEZ, J. 2008. CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. Carcinogenesis, 29, 237-43.

5. BOSI, G., EVANGELISTI, R., VALENO, V., CARINCI, F., PEZZETTI, F., CALASTRINI, C., BODO, M. & CARINCI, P. 1998. Diphenylhydantoin affects glycosaminoglycans and collagen production by human fibroblasts from cleft palate patients. Journal of dental research, 77, 1613-21.

6. 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 letters, 573, 83-92.

7. BROWN, N. L., KNOTT, L., HALLIGAN, E., YARRAM, S. J., MANSELL, J. P. & SANDY, J. R. 2003. Microarray analysis of murine palatogenesis: temporal expression of genes during normal palate development. Development, growth & differentiation, 45, 153-65.

8. BUENO, D. F., KERKIS, I., COSTA, A. M., MARTINS, M. T., KOBAYASHI, G. S., ZUCCONI, E., FANGANIELLO, R. D., SALLES, F. T., ALMEIDA, A. B., DO AMARAL, C. E., ALONSO, N. & PASSOS-BUENO, M. R. 2009. New source of muscle-derived stem cells with potential for alveolar bone reconstruction in cleft lip and/or palate patients. Tissue engineering. Part A, 15, 427-35.

9. BUENO, D. F., SUNAGA, D. Y., KOBAYASHI, G. S., AGUENA, M., RAPOSO-AMARAL, C. E., MASOTTI, C., CRUZ, L. A., PEARSON, P. L. & PASSOS-BUENO, M. R. 2011. 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, 7, 446-57.

10. 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 research, 58, 3777-81.

11. CHEN, H. Z., TSAI, S. Y. & LEONE, G. 2009. Emerging roles of E2Fs in cancer: an exit from cell cycle control. Nature reviews. Cancer, 9, 785-97.

12. CHRISTENSEN, K. & MORTENSEN, P. B. 2002. Facial clefting and psychiatric diseases: a follow-up of the Danish 1936-1987 Facial Cleft cohort. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association, 39, 392-6.

13. COOKSON, W., LIANG, L., ABECASIS, G., MOFFATT, M. & LATHROP, M. 2009. Mapping complex disease traits with global gene expression. Nature reviews. Genetics, 10, 184-94.

14. DE MENDONCA COSTA, A., BUENO, D. F., MARTINS, M. T., KERKIS, I., KERKIS, A., FANGANIELLO, R. D., CERRUTI, H., ALONSO, N. & PASSOS-BUENO, M. R. 2008. Reconstruction of large cranial defects in nonimmunosuppressed experimental design with human dental pulp stem cells. The Journal of craniofacial surgery, 19, 204-10.

15. DENG, C. X. 2006. BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. Nucleic acids research, 34, 1416-26.

16. DIXON, M. J., MARAZITA, M. L., BEATY, T. H. & MURRAY, J. C. 2011. Cleft lip and palate: understanding genetic and environmental influences. Nature reviews. Genetics, 12, 167-78.

17. GAO, L. B., PAN, X. M., LI, L. J., LIANG, W. B., ZHU, Y., ZHANG, L. S., WEI, Y. G., TANG, M. & ZHANG, L. 2011. RAD51 135G/C polymorphism and breast cancer risk: a meta-analysis from 21 studies. Breast cancer research and treatment, 125, 827-35.

18. INOSHITA, S., TERADA, Y., NAKASHIMA, O., KUWAHARA, M., SASAKI, S. & MARUMO, F. 1999. Regulation of the G1/S transition phase in mesangial cells by E2F1. Kidney international, 56, 1238-41.

19. JAKOBSEN, L. P., BORUP, R., VESTERGAARD, J., LARSEN, L. A., LAGE, K., MAROUN, L. L., KJAER, I., NIEMANN, C. U., ANDERSEN, M., KNUDSEN, M. A., MOLLGARD, K. & TOMMERUP, N. 2009. Expression analyses of human cleft palate tissue suggest a role for osteopontin and immune related factors in palatal development. Experimental & molecular medicine, 41, 77-85.

20. JIANG, R., BUSH, J. O. & LIDRAL, A. C. 2006. Development of the upper lip: morphogenetic and molecular mechanisms. Developmental dynamics : an official publication of the American Association of Anatomists, 235, 1152-66.

21. KASTEN, E. F., SCHMIDT, S. P., ZICKLER, C. F., BERNER, E., DAMIAN, L. A., CHRISTIAN, G. M., WORKMAN, H., FREEMAN, M., FARLEY, M. D. & HICKS, T. L. 2008. Team care of the patient with cleft lip and palate. Current problems in pediatric and adolescent health care, 38, 138-58.

22. KUSEK, J. C., GREENE, R. M., NUGENT, P. & PISANO, M. M. 2000. Expression of the E2F family of transcription factors during murine development. The International journal of developmental biology, 44, 267-77.

23. LEE, W. Y., JIN, Y. T., CHANG, T. W., LIN, P. W. & SU, I. J. 1999. Immunolocalization of BRCA1 protein in normal breast tissue and sporadic invasive ductal carcinomas: a correlation with other biological parameters. Histopathology, 34, 106-12.

24. LIDRAL, A. C., MORENO, L. M. & BULLARD, S. A. 2008. Genetic Factors and Orofacial Clefting. Seminars in orthodontics, 14, 103-114.

25. LINGER, R. J. & KRUK, P. A. 2010. BRCA1 16 years later: risk-associated BRCA1 mutations and their functional implications. The FEBS journal, 277, 3086-96.

26. MARINUCCI, L., BALLONI, S., BODO, M., CARINCI, F., PEZZETTI, F., STABELLINI, G., CONTE, C. & LUMARE, E. 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, 10-6.

27. MARINUCCI, L., BALLONI, S., CARINCI, F., LOCCI, P., PEZZETTI, F. & BODO, M. 2011. Diazepam effects on non-syndromic cleft lip with or without palate: epidemiological studies, clinical findings, genes and extracellular matrix. Expert opinion on drug safety, 10, 23-33.

28. MENEZES, R., MARAZITA, M. L., GOLDSTEIN MCHENRY, T., COOPER, M. E., BARDI, K., BRANDON, C., LETRA, A., MARTIN, R. A. & VIEIRA, A. R. 2009. AXIS inhibition protein 2, orofacial clefts and a family history of cancer. Journal of the American Dental Association, 140, 80-4.

29. MOSSEY, P. A., LITTLE, J., MUNGER, R. G., DIXON, M. J. & SHAW, W. C. 2009. Cleft lip and palate. Lancet, 374, 1773-85.

30. MUELLER, C. R. & ROSKELLEY, C. D. 2003. Regulation of BRCA1 expression and its relationship to sporadic breast cancer. Breast cancer research : BCR, 5, 45-52.

31. MURAOKA, N., SHUM, L., FUKUMOTO, S., NOMURA, T., OHISHI, M. & NONAKA, K. 2005. Transforming growth factor-beta3 promotes mesenchymal cell proliferation and angiogenesis mediated by the enhancement of cyclin D1, Flk-1, and CD31 gene expression during CL/Fr mouse lip fusion. Birth defects research. Part A, Clinical and molecular teratology, 73, 956-65.

32. PEZZETTI, F., CARINCI, F., PALMIERI, A., VIZZOTTO, L., MOSCHENI, C., VERTEMATI, M., CALASTRINI, C., PELLATI, A. & STABELLINI, G. 2009. Diphenylhydantoin plays a role in gene expression related to cytoskeleton and protein adhesion in human normal palate fibroblasts. Pathology, 41, 261-8.

33. PFAFFL, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic acids research, 29, e45.

34. RAJ, A., RIFKIN, S. A., ANDERSEN, E. & VAN OUDENAARDEN, A. 2010. Variability in gene expression underlies incomplete penetrance. Nature, 463, 913-8.

35. ROSEN, E. M., FAN, S. & MA, Y. 2006. BRCA1 regulation of transcription. Cancer letters, 236, 175-85.

36. 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. Journal of molecular histology, 37, 301-7.

37. TAIOLI, E., RAGIN, C., ROBERTSON, L., LINKOV, F., THURMAN, N. E. & VIEIRA, A. R. 2010. Cleft lip and palate in family members of cancer survivors. Cancer investigation, 28, 958-62.

38. THOMPSON, M. E., JENSEN, R. A., OBERMILLER, P. S., PAGE, D. L. & HOLT, J. T. 1995. Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression. Nature genetics, 9, 444-50.

39. TUSHER, V. G., TIBSHIRANI, R. & CHU, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America, 98, 5116-21.

40. VANDESOMPELE, J., DE PRETER, K., PATTYN, F., POPPE, B., VAN ROY, N., DE PAEPE, A. & SPELEMAN, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology, 3, RESEARCH0034.

41. YANG, S. Z., LIN, F. T. & LIN, W. C. 2008. MCPH1/BRIT1 cooperates with E2F1 in the activation of checkpoint, DNA repair and apoptosis. EMBO reports, 9, 907-15.

42. YARDEN, R. I. & PAPA, M. Z. 2006. BRCA1 at the crossroad of multiple cellular pathways: approaches for therapeutic interventions. Molecular cancer therapeutics, 5, 1396-404.

43. YATA, K. & ESASHI, F. 2009. Dual role of CDKs in DNA repair: to be, or not to be. DNA repair, 8, 6-18.

44. YOUNG, D. L., SCHNEIDER, R. A., HU, D. & HELMS, J. A. 2000. Genetic and teratogenic approaches to craniofacial development. Critical reviews in oral biology and medicine : an official publication of the American Association of Oral Biologists, 11, 304-17.

45. ZHANG, Y., MORI, T., ISEKI, K., HAGINO, S., TAKAKI, H., TAKEUCHI, M., HIKAKE, T., TASE, C., MURAKAWA, M., YOKOYA, S. & WANAKA, A. 2003. Differential expression of decorin and biglycan genes during palatogenesis in normal

and retinoic acid-treated mice. Developmental dynamics : an official publication of the American Association of Anatomists, 226, 618-26.

46. ZHU, H., KARTIKO, S. & FINNELL, R. H. 2009. Importance of gene-environment interactions in the etiology of selected birth defects. Clinical genetics, 75, 409-23.

47. ZHU, J. L., BASSO, O., HASLE, H., WINTHER, J. F., OLSEN, J. H. & OLSEN, J. 2002. Do parents of children with congenital malformations have a higher cancer risk? A nationwide study in Denmark. British journal of cancer, 87, 524-8.

CAPÍTULO V - Discussão geral e conclusões

O presente trabalho consistiu na análise do perfil de expressão gênica de culturas de células-tronco mesenquimais (CTM) de polpa de dente decíduo, extraídas de pacientes acometidos por fissura lábio-palatina não-sindrômica (FLP NS). Demonstramos que estas culturas celulares apresentam uma assinatura de expressão gênica que é específica à doença. Além disso, identificamos padrões de desregulação transcricional que podem levar ao acometimento de processos biológicos relevantes para o desenvolvimento do lábio e palato.

No trabalho descrito no Capítulo III, identificamos desregulação de genes principalmente associados à remodelagem de matriz extracelular e transição epitélio-mesenquimal (*epithelial-mesenchymal* transition, *EMT*), dois processos biológicos extremamente importantes para a morfogênese orofacial. Para que haja o posicionamento correto das estruturas embrionárias que formarão o lábio e palato, é necessário um reprocessamento da matriz extracelular, permitindo a motilidade dessas estruturas. Tal atividade é caracterizada principalmente pela degradação de colágeno e outras moléculas de matriz pela ação de metaloproteinases (Kerrigan *et al*., 2000). Além disso, a *EMT*, que modula a fusão das proeminências faciais e lâminas palatinas, também depende desse mecanismo, uma vez que as células transdiferenciadas por *EMT* devem desintegrar a lâmina basal e invadir o tecido mesenquimal através do meio extracelular (Sun *et al*., 2000; Nakazawa *et al*., 2008; Yu *et al*., 2009). A observação de que, em CTMs de pacientes,

moléculas de matriz exibem níveis altos de expressão, enquanto que metaloproteinases exibem níveis menores, sugere que a desregulação desses componentes pode levar à manifestação de FLP durante a embriogênese.

Utilizando novas culturas de CTMs de pacientes e controles submetidas a uma condição de cultivo diferente (Capítulo IV), descrevemos um perfil de expressão distinto em relação ao obtido no Capítulo III. Observamos uma maior quantidade de genes diferencialmente expressos, e uma evidente desregulação de genes envolvidos nos mecanismo de reparo de DNA associado à proteína *BRCA1* e *checkpoint* do ciclo celular. Além disso, identificamos o fator de transcrição *E2F1* como elemento regulatório candidato à modulação do padrão de expressão observado. Esses resultados indicam que mecanismos relacionados principalmente à proliferação e progressão do ciclo celular também podem estar comprometidos em indivíduos portadores de FLP NS. A desregulação desses mecanismos pode estar relacionada à patogênese da doença, uma vez que um fino controle de proliferação celular é crucial para o crescimento das estruturas da face do embrião (Kerrigan *et al*., 2000; Dhulipala *et al*., 2006; Meng *et al*., 2009). Além disso, dados da literatura apontam que *E2F1* exerce um papel funcional durante a formação do palato (Kusek *et al*., 2000). A maior ocorrência de certos tipos de câncer em famílias segregando FLP NS também apoia a hipótese de desregulação de ciclo celular e reparo de DNA no contexto das fissuras, uma vez que detectamos transcrição anormal de muitos genes associados a formas hereditárias de câncer e em uma mesma via de interação, a exemplo do *BRCA1* (câncer de mama, ovário, e próstata), *MSH2* (câncer colorretal), *BLM* (síndrome de Bloom), entre outros. Esta é a primeira

evidência do envolvimento de uma rede gênica diretamente relacionada à tumorigênese com a etiologia das FLP NS.

Interessantemente, obtivemos 2 tipos distintos de padrões de expressão, a depender das condições de cultivo celular. Sabendo da natureza multifatorial das FLP NS, esses resultados nos levam a especular que a presença ou ausência de soro fetal bovino pode ser um modificador ambiental importante, levando à manifestação de perfis de expressão refletindo desregulação de diferentes mecanismos biológicos. Ainda, com a remoção de soro, obtivemos resultados semelhantes ao de outros grupos de pesquisa, que utilizaram agentes teratogênicos (como a fenitoína e a nicotina) em culturas de fibroblastos de pacientes com FLP NS. Nesses trabalhos, também foi averiguada uma expressão anormal de componentes de matriz extracelular, incluindo alguns colágenos (Bosi *et al*., 1998; Marinucci *et al*., 2009; Baroni *et al*., 2010); no entanto, nunca foi se estabeleceu uma relação entre os genes desregulados e a *EMT*. Sendo assim, é possível que a retirada de soro represente um insulto ambiental de efeito semelhante ao descrito por esses pesquisadores, porém, com o adicional comprometimento da *EMT* no tipo celular utilizado no presente trabalho. Finalmente, nossos resultados estão de acordo com a hipótese de que o genoma de indivíduos portadores de FLP NS abriga variantes ou mecanismos epigenéticos específicos que respondem diferentemente ao ambiente, em relação a controles. Por outro lado, estamos cientes de que para confirmar o efeito ambiental proporcionado pela retirada de soro, mais estudos serão necessários, sobretudo expondo as mesmas culturas celulares às diferentes condições de cultivo.

Em conjunto, nossos dados sugerem que culturas celulares de indivíduos portadores de FLP NS apresentam desregulação de redes gênicas relacionadas à remodelagem de matriz extracelular, transição epitélio-mesenquimal, e reparo de danos ao DNA e *checkpoint* do ciclo celular. Tais funções estão diretamente relacionadas a mecanismos fundamentais para a correta morfogênese do lábio e palato. Sendo assim, se anomalias nesses mecanismos são responsáveis pelo fenótipo de fissura, podemos inferir que padrões de expressão anômalos em tecidos embrionários podem ser transmitidos a tecidos adultos. No caso, sabe-se que populações celulares isoladas de polpa de dente são formadas por células originadas do tecido mesenquimal e células da crista neural do embrião (Chai *et al*., 2000; Waddington *et al*., 2009). Esta observação indica que as CTM de polpa de dente decíduo são de grande valia para estudar a etiologia das FLP NS, pois essas células possivelmente refletem erros de sinalização ocorridos em seus tecidos de origem.

Em resumo, dado o padrão de herança multifatorial das FLP NS, nossos resultados indicam que o reflexo transcricional gerado por diferentes variantes genéticas presentes no genoma ou epigenoma de portadores de FLP NS pode convergir na desregulação de mecanismos-chave para a patogênese desta malformação. Estudos futuros serão direcionados ao esclarecimento de vias regulatórias *upstream* aos padrões de expressão observados, bem como a uma investigação mais cautelosa acerca dos mecanismos aqui relatados. Testes funcionais serão necessários para confirmar o comprometimento das funções celulares identificadas nos experimentos de expressão gênica, particularmente em populações celulares sujeitas a insultos ambientais. Os

dados gerados por nossos estudos serão importantes para o entendimento acerca dos complexos mecanismos que orquestram a etiologia das FLP NS.

BIBLIOGRAFIA

1. ACLOQUE, H., ADAMS, M. S., FISHWICK, K., BRONNER-FRASER, M. & NIETO, M. A. 2009. Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. The Journal of clinical investigation, 119, 1438-49.

2. ALTSHULER, D., DALY, M. J., LANDER, E. S. 2008. Genetic mapping in human disease. Science, 322, 881-888.

3. AVILA, J. R., JEZEWSKI, P. A., VIEIRA, A. R., ORIOLI, I. M., CASTILLA, E. E., CHRISTENSEN, K., DAACK-HIRSCH, S., ROMITTI, P. A. & MURRAY, J. C. 2006. PVRL1 variants contribute to non-syndromic cleft lip and palate in multiple populations. American journal of medical genetics. Part A, 140, 2562-70.

4. BARONI, T., BELLUCCI, C., LILLI, C., PEZZETTI, F., CARINCI, F., LUMARE, E., PALMIERI, A., STABELLINI, G. & BODO, M. 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. Journal of cellular physiology, 222, 748-56.

5. BEATY, T. H., MURRAY, J. C., MARAZITA, M. L., MUNGER, R. G., RUCZINSKI, I., HETMANSKI, J. B., LIANG, K. Y., WU, T., MURRAY, T., FALLIN, M. D., REDETT, R. A., RAYMOND, G., SCHWENDER, H., JIN, S. C., COOPER, M. E., DUNNWALD, M., MANSILLA, M. A., LESLIE, E., BULLARD, S., LIDRAL, A. C., MORENO, L. M., MENEZES, R., VIEIRA, A. R., PETRIN, A., WILCOX, A. J., LIE, R. T., JABS, E. W., WU-CHOU, Y. H., CHEN, P. K., WANG, H., YE, X., HUANG, S., YEOW, V., CHONG, S. S., JEE, S. H., SHI, B., CHRISTENSEN, K., MELBYE, M., DOHENY, K. F., PUGH, E. W., LING, H., CASTILLA, E. E., CZEIZEL, A. E., MA, L., FIELD, L. L., BRODY, L., PANGILINAN, F., MILLS, J. L., MOLLOY, A. M., KIRKE, P. N., SCOTT, J. M., ARCOS-BURGOS, M. & SCOTT, A. F. 2010. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nature genetics, 42, 525-9.

6. BIRNBAUM, S., LUDWIG, K. U., REUTTER, H., HERMS, S., DE ASSIS, N. A., DIAZ-LACAVA, A., BARTH, S., LAUSTER, C., SCHMIDT, G., SCHEER, M., SAFFAR, M., MARTINI, M., REICH, R. H., SCHIEFKE, F., HEMPRICH, A., POTZSCH, S., POTZSCH, B., WIENKER, T. F., HOFFMANN, P., KNAPP, M., KRAMER, F. J., NOTHEN, M. M. & MANGOLD, E. 2009a. IRF6 gene variants in Central European patients with non-syndromic cleft lip with or without cleft palate. European journal of oral sciences, 117, 766-9.

7. BIRNBAUM, S., LUDWIG, K. U., REUTTER, H., HERMS, S., STEFFENS, M., RUBINI, M., BALUARDO, C., FERRIAN, M., ALMEIDA DE ASSIS, N., ALBLAS, M. A., BARTH, S., FREUDENBERG, J., LAUSTER, C., SCHMIDT, G., SCHEER, M., BRAUMANN, B., BERGE, S. J., REICH, R. H., SCHIEFKE, F., HEMPRICH, A., POTZSCH, S., STEEGERS-THEUNISSEN, R. P., POTZSCH, B., MOEBUS, S., HORSTHEMKE, B., KRAMER, F. J., WIENKER, T. F., MOSSEY, P. A., PROPPING, P., CICHON, S., HOFFMANN, P., KNAPP, M., NOTHEN, M. M. & MANGOLD, E. 2009b. Key susceptibility locus for

nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nature genetics, 41, 473-7.

8. BOSI, G., EVANGELISTI, R., VALENO, V., CARINCI, F., PEZZETTI, F., CALASTRINI, C., BODO, M. & CARINCI, P. 1998. Diphenylhydantoin affects glycosaminoglycans and collagen production by human fibroblasts from cleft palate patients. Journal of dental research, 77, 1613-21.

9. BRAYBROOK, C., DOUDNEY, K., MARCANO, A. C., ARNASON, A., BJORNSSON, A., PATTON, M. A., GOODFELLOW, P. J., MOORE, G. E. & STANIER, P. 2001. The T-box transcription factor gene TBX22 is mutated in X-linked cleft palate and ankyloglossia. Nature genetics, 29, 179-83.

10. BROWN, N. L., KNOTT, L., HALLIGAN, E., YARRAM, S. J., MANSELL, J. P. & SANDY, J. R. 2003. Microarray analysis of murine palatogenesis: temporal expression of genes during normal palate development. Development, growth & differentiation, 45, 153-65.

11. BUENO, D. F., KERKIS, I., COSTA, A. M., MARTINS, M. T., KOBAYASHI, G. S., ZUCCONI, E., FANGANIELLO, R. D., SALLES, F. T., ALMEIDA, A. B., DO AMARAL, C. E., ALONSO, N. & PASSOS-BUENO, M. R. 2009. New source of muscle-derived stem cells with potential for alveolar bone reconstruction in cleft lip and/or palate patients. Tissue engineering. Part A, 15, 427-35.

12. CAPLAN, A. I. 2005. Review: mesenchymal stem cells: cell-based reconstructive therapy in orthopedics. Tissue engineering, 11, 1198-211.

13. CARINCI, F., SCAPOLI, L., PALMIERI, A., ZOLLINO, I. & PEZZETTI, F. 2007. Human genetic factors in nonsyndromic cleft lip and palate: an update. International journal of pediatric otorhinolaryngology, 71, 1509-19.

14. CHAI, Y., JIANG, X., ITO, Y., BRINGAS, P., JR., HAN, J., ROWITCH, D. H., SORIANO, P., MCMAHON, A. P. & SUCOV, H. M. 2000. Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. Development, 127, 1671-9.

15. CHRISTENSEN, K. & MORTENSEN, P. B. 2002. Facial clefting and psychiatric diseases: a follow-up of the Danish 1936-1987 Facial Cleft cohort. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association, 39, 392-6.

16. COBOURNE, M. T. 2004. The complex genetics of cleft lip and palate. European journal of orthodontics, 26, 7-16.

17. COOKSON, W., LIANG, L., ABECASIS, G., MOFFATT, M. & LATHROP, M. 2009. Mapping complex disease traits with global gene expression. Nature reviews. Genetics, 10, 184-94.

18. DE MEY, A., VAN HOOF, I., DE ROY, G. & LEJOUR, M. 1989. Anatomy of the orbicularis oris muscle in cleft lip. British journal of plastic surgery, 42, 710-4.

19. DHULIPALA, V. C., WELSHONS, W. V. & REDDY, C. S. 2006. Cell cycle proteins in normal and chemically induced abnormal secondary palate development: a review. Human & experimental toxicology, 25, 675-82.

20. DHULIPALA, V. C., MADDALI, K. K., RAY, B. K., WELSHONS, W. V. & REDDY, C. S. 2010. Role of p21 and cyclin E in normal and secalonic acid D-inhibited proliferation of human embryonic palatal mesenchymal cells. Human & experimental toxicology.

21. DIXON, M. J., MARAZITA, M. L., BEATY, T. H. & MURRAY, J. C. 2011. Cleft lip and palate: understanding genetic and environmental influences. Nature reviews. Genetics, 12, 167-78.

22. DUDAS, M., LI, W. Y., KIM, J., YANG, A. & KAARTINEN, V. 2007. Palatal fusion - where do the midline cells go? A review on cleft palate, a major human birth defect. Acta histochemica, 109, 1-14.

23. ENOMOTO, H., NELSON, C. M., SOMERVILLE, R. P., MIELKE, K., DIXON, L. J., POWELL, K. & APTE, S. S. 2010. Cooperation of two ADAMTS metalloproteases in closure of the mouse palate identifies a requirement for versican proteolysis in regulating palatal mesenchyme proliferation. Development, 137, 4029-38.

24. ERICES, A., CONGET, P. & MINGUELL, J. J. 2000. Mesenchymal progenitor cells in human umbilical cord blood. British journal of haematology, 109, 235-42.

25. FRASER, F. C. 1970. The genetics of cleft lip and cleft palate. American journal of human genetics, 22, 336-52.

26. GASPAR, D. A., MATIOLI, S. R., DE CASSIA PAVANELLO, R., ARAUJO, B. C., ALONSO, N., WYSZYNSKI, D. & PASSOS-BUENO, M. R. 2004. Maternal MTHFR interacts with the offspring's BCL3 genotypes, but not with TGFA, in increasing risk to nonsyndromic cleft lip with or without cleft palate. European journal of human genetics : EJHG, 12, 521-6.

27. GASPAR, D. A., MATIOLI, S. R., PAVANELLO, R. C., ARAUJO, B. C., ANDRE, M., STEMAN, S., OTTO, P. A. & PASSOS-BUENO, M. R. 2002. Evidence that BCL3 plays a role in the etiology of nonsyndromic oral clefts in Brazilian families. Genetic epidemiology, 23, 364-74.

28. GHASSIBE, M., BAYET, B., REVENCU, N., VERELLEN-DUMOULIN, C., GILLEROT, Y., VANWIJCK, R. & VIKKULA, M. 2005. Interferon regulatory factor-6: a gene predisposing to isolated cleft lip with or without cleft palate in the Belgian population. European journal of human genetics : EJHG, 13, 1239-42.

29. GORLIN, R. J., COHEN, M. M. & HENNEKAM, R. C. M. 2001. Syndromes of the head and neck, Oxford England ; New York, Oxford University Press.

30. GREENE, R. M. & PISANO, M. M. 2010. Palate morphogenesis: Current understanding and future directions. Birth defects research. Part C, Embryo today : reviews, 90, 133-54.

31. GRONTHOS, S., BRAHIM, J., LI, W., FISHER, L. W., CHERMAN, N., BOYDE, A., DENBESTEN, P., ROBEY, P. G. & SHI, S. 2002. Stem cell properties of human dental pulp stem cells. Journal of dental research, 81, 531-5.

32. HOUDAYER, C. & BAHUAU, M. 1998. Orofacial cleft defects: inference from nature and nurture. Annales de genetique, 41, 89-117.

33. ISEKI, S. 2011. Disintegration of the medial epithelial seam: is cell death important in palatogenesis? Development, growth & differentiation, 53, 259-68.

34. JAKOBSEN, L. P., BORUP, R., VESTERGAARD, J., LARSEN, L. A., LAGE, K., MAROUN, L. L., KJAER, I., NIEMANN, C. U., ANDERSEN, M., KNUDSEN, M. A., MOLLGARD, K. & TOMMERUP, N. 2009. Expression analyses of human cleft palate tissue suggest a role for osteopontin and immune related factors in palatal development. Experimental & molecular medicine, 41, 77-85.

35. JIANG, R., BUSH, J. O. & LIDRAL, A. C. 2006. Development of the upper lip: morphogenetic and molecular mechanisms. Developmental dynamics : an official publication of the American Association of Anatomists, 235, 1152-66.

36. JOHNSTON, M. C. & BRONSKY, P. T. 1995. Prenatal craniofacial development: new insights on normal and abnormal mechanisms. Critical reviews in oral biology and medicine : an official publication of the American Association of Oral Biologists, 6, 25-79.

37. KASTEN, E. F., SCHMIDT, S. P., ZICKLER, C. F., BERNER, E., DAMIAN, L. A., CHRISTIAN, G. M., WORKMAN, H., FREEMAN, M., FARLEY, M. D. & HICKS, T. L. 2008. Team care of the patient with cleft lip and palate. Current problems in pediatric and adolescent health care, 38, 138-58.

38. KERRIGAN, J. J., MANSELL, J. P., SENGUPTA, A., BROWN, N. & SANDY, J. R. 2000. Palatogenesis and potential mechanisms for clefting. Journal of the Royal College of Surgeons of Edinburgh, 45, 351-8.

39. KLOTZ, C. M., WANG, X., DESENSI, R. S., GRUBS, R. E., COSTELLO, B. J. & MARAZITA, M. L. 2010. Revisiting the recurrence risk of nonsyndromic cleft lip with or without cleft palate. American journal of medical genetics. Part A, 152A, 2697-702.

40. KONDO, S., SCHUTTE, B. C., RICHARDSON, R. J., BJORK, B. C., KNIGHT, A. S., WATANABE, Y., HOWARD, E., DE LIMA, R. L., DAACK-HIRSCH, S., SANDER, A., MCDONALD-MCGINN, D. M., ZACKAI, E. H., LAMMER, E. J., AYLSWORTH, A. S., ARDINGER, H. H., LIDRAL, A. C., POBER, B. R., MORENO, L., ARCOS-BURGOS, M., VALENCIA, C., HOUDAYER, C., BAHUAU, M., MORETTI-FERREIRA, D., RICHIERI-COSTA, A., DIXON, M. J. & MURRAY, J. C. 2002. Mutations in IRF6

cause Van der Woude and popliteal pterygium syndromes. Nature genetics, 32, 285-9.

41. KRAPELS, I. P., VERMEIJ-KEERS, C., MULLER, M., DE KLEIN, A. & STEEGERS-THEUNISSEN, R. P. 2006. Nutrition and genes in the development of orofacial clefting. Nutrition reviews, 64, 280-8.

42. LIDRAL, A. C., MORENO, L. M. & BULLARD, S. A. 2008. Genetic Factors and Orofacial Clefting. Seminars in orthodontics, 14, 103-114.

43. LOFFREDO, L. C. M., FREITAS, J. A. S., GRIGOLLI, A. A. G. 2001. Prevalence of oral clefts from 1975 to 1994, Brazil. Revista de Saúde Pública, 35.

44. MANGOLD, E., LUDWIG, K. U., BIRNBAUM, S., BALUARDO, C., FERRIAN, M., HERMS, S., REUTTER, H., DE ASSIS, N. A., CHAWA, T. A., MATTHEISEN, M., STEFFENS, M., BARTH, S., KLUCK, N., PAUL, A., BECKER, J., LAUSTER, C., SCHMIDT, G., BRAUMANN, B., SCHEER, M., REICH, R. H., HEMPRICH, A., POTZSCH, S., BLAUMEISER, B., MOEBUS, S., KRAWCZAK, M., SCHREIBER, S., MEITINGER, T., WICHMANN, H. E., STEEGERS-THEUNISSEN, R. P., KRAMER, F. J., CICHON, S., PROPPING, P., WIENKER, T. F., KNAPP, M., RUBINI, M., MOSSEY, P. A., HOFFMANN, P. & NOTHEN, M. M. 2010. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nature genetics, 42, 24-6.

45. MARCANO, A. C., DOUDNEY, K., BRAYBROOK, C., SQUIRES, R., PATTON, M. A., LEES, M. M., RICHIERI-COSTA, A., LIDRAL, A. C., MURRAY, J. C., MOORE, G. E. & STANIER, P. 2004. TBX22 mutations are a frequent cause of cleft palate. Journal of medical genetics, 41, 68-74.

46. MARINUCCI, L., BALLONI, S., BODO, M., CARINCI, F., PEZZETTI, F., STABELLINI, G., CONTE, C. & LUMARE, E. 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, 10-6.

47. MARINUCCI, L., BALLONI, S., CARINCI, F., LOCCI, P., PEZZETTI, F. & BODO, M. 2011. Diazepam effects on non-syndromic cleft lip with or without palate: epidemiological studies, clinical findings, genes and extracellular matrix. Expert opinion on drug safety, 10, 23-33.

48. MENEGOTTO, B. G. & SALZANO, F. M. 1991. Epidemiology of oral clefts in a large South American sample. The Cleft palate-craniofacial journal : official publication of the American Cleft Palate-Craniofacial Association, 28, 373-6; discussion 376-7.

49. MENEZES, R., MARAZITA, M. L., GOLDSTEIN MCHENRY, T., COOPER, M. E., BARDI, K., BRANDON, C., LETRA, A., MARTIN, R. A. & VIEIRA, A. R. 2009. AXIS inhibition protein 2, orofacial clefts and a family history of cancer. Journal of the American Dental Association, 140, 80-4.

50. MENG, L., BIAN, Z., TORENSMA, R. & VON DEN HOFF, J. W. 2009. Biological mechanisms in palatogenesis and cleft palate. Journal of dental research, 88, 22-33.

51. METNEKI, J., PUHO, E. & CZEIZEL, A. E. 2005. Maternal diseases and isolated orofacial clefts in Hungary. Birth defects research. Part A, Clinical and molecular teratology, 73, 617-23.

52. MIURA, M., GRONTHOS, S., ZHAO, M., LU, B., FISHER, L. W., ROBEY, P. G. & SHI, S. 2003. SHED: stem cells from human exfoliated deciduous teeth. Proceedings of the National Academy of Sciences of the United States of America, 100, 5807-12.

53. MOSSEY, P. A., LITTLE, J., MUNGER, R. G., DIXON, M. J. & SHAW, W. C. 2009. Cleft lip and palate. Lancet, 374, 1773-85.

54. MUKHOPADHYAY, P., BROCK, G., PIHUR, V., WEBB, C., PISANO, M. M. & GREENE, R. M. 2010. Developmental microRNA expression profiling of murine embryonic orofacial tissue. Birth defects research. Part A, Clinical and molecular teratology, 88, 511-34.

55. MUKHOPADHYAY, P., GREENE, R. M., ZACHARIAS, W., WEINRICH, M. C., SINGH, S., YOUNG, W. W., JR. & PISANO, M. M. 2004. Developmental gene expression profiling of mammalian, fetal orofacial tissue. Birth defects research. Part A, Clinical and molecular teratology, 70, 912-26.

56. MURAOKA, N., SHUM, L., FUKUMOTO, S., NOMURA, T., OHISHI, M. & NONAKA, K. 2005. Transforming growth factor-beta3 promotes mesenchymal cell proliferation and angiogenesis mediated by the enhancement of cyclin D1, Flk-1, and CD31 gene expression during CL/Fr mouse lip fusion. Birth defects research. Part A, Clinical and molecular teratology, 73, 956-65.

57. MURRAY, J. C. 1995. Face facts: genes, environment, and clefts. American journal of human genetics, 57, 227-32.

58. NAKAZAWA, M., MATSUNAGA, K., ASAMURA, S., KUSUHARA, H., ISOGAI, N. & MURAGAKI, Y. 2008. Molecular mechanisms of cleft lip formation in CL/Fr mice. Scandinavian journal of plastic and reconstructive surgery and hand surgery / Nordisk plastikkirurgisk forening [and] Nordisk klubb for handkirurgi, 42, 225-32.

59. NAWSHAD, A. 2008. Palatal seam disintegration: to die or not to die? that is no longer the question. Developmental dynamics : an official publication of the American Association of Anatomists, 237, 2643-56.

60. PARK, J. W., MCINTOSH, I., HETMANSKI, J. B., JABS, E. W., VANDER KOLK, C. A., WU-CHOU, Y. H., CHEN, P. K., CHONG, S. S., YEOW, V., JEE, S. H., PARK, B. Y., FALLIN, M. D., INGERSOLL, R., SCOTT, A. F. & BEATY, T. H. 2007. Association between IRF6 and nonsyndromic cleft lip with or without cleft palate in four populations. Genetics in medicine : official journal of the American College of Medical Genetics, 9, 219-27.

61. PEZZETTI, F., CARINCI, F., PALMIERI, A., VIZZOTTO, L., MOSCHENI, C., VERTEMATI, M., CALASTRINI, C., PELLATI, A. & STABELLINI, G. 2009. Diphenylhydantoin plays a role in gene expression related to cytoskeleton and protein adhesion in human normal palate fibroblasts. Pathology, 41, 261-8.

62. ROJAS-MARTINEZ, A., REUTTER, H., CHACON-CAMACHO, O., LEON-CACHON, R. B., MUNOZ-JIMENEZ, S. G., NOWAK, S., BECKER, J., HERBERZ, R., LUDWIG, K. U., PAREDES-ZENTENO, M., ARIZPE-CANTU, A., RAEDER, S., HERMS, S., ORTIZ-LOPEZ, R., KNAPP, M., HOFFMANN, P., NOTHEN, M. M. & MANGOLD, E. 2010. Genetic risk factors for nonsyndromic cleft lip with or without cleft palate in a Mesoamerican population: Evidence for IRF6 and variants at 8q24 and 10q25. Birth defects research. Part A, Clinical and molecular teratology, 88, 535-7.

63. SASAKI, Y., TANAKA, S., HAMACHI, T. & TAYA, Y. 2004. Deficient cell proliferation in palatal shelf mesenchyme of CL/Fr mouse embryos. Journal of dental research, 83, 797-801.

64. SCHUTTE, B. C. & MURRAY, J. C. 1999. The many faces and factors of orofacial clefts. Human molecular genetics, 8, 1853-9.

65. SHAW, G. M., WASSERMAN, C. R., LAMMER, E. J., O'MALLEY, C. D., MURRAY, J. C., BASART, A. M. & TOLAROVA, M. M. 1996. Orofacial clefts, parental cigarette smoking, and transforming growth factor-alpha gene variants. American journal of human genetics, 58, 551-61.

66. SLAVKIN, H. C. 1992. Incidence of cleft lips, palates rising. Journal of the American Dental Association, 123, 61-5.

67. STANIER, P. & MOORE, G. E. 2004. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. Human molecular genetics, 13 Spec No 1, R73-81.

68. SUN, D., BAUR, S. & HAY, E. D. 2000. Epithelial-mesenchymal transformation is the mechanism for fusion of the craniofacial primordia involved in morphogenesis of the chicken lip. Developmental biology, 228, 337-49.

69. SUZUKI, K., HU, D., BUSTOS, T., ZLOTOGORA, J., RICHIERI-COSTA, A., HELMS, J. A. & SPRITZ, R. A. 2000. Mutations of PVRL1, encoding a cell-cell adhesion molecule/herpesvirus receptor, in cleft lip/palate-ectodermal dysplasia. Nature genetics, 25, 427-30.

70. WADDINGTON, R. J., YOUDE, S. J., LEE, C. P. & SLOAN, A. J. 2009. Isolation of distinct progenitor stem cell populations from dental pulp. Cells, tissues, organs, 189, 268-74.

71. WIJAYAWEERA, C. J., AMARATUNGA, N. A. & ANGUNAWELA, P. 2000. Arrangement of the orbicularis oris muscle in different types of cleft lips. The Journal of craniofacial surgery, 11, 232-5.

72. WILLIAMS, J. T., SOUTHERLAND, S. S., SOUZA, J., CALCUTT, A. F. & CARTLEDGE, R. G. 1999. Cells isolated from adult human skeletal muscle capable of differentiating into multiple mesodermal phenotypes. The American surgeon, 65, 22-6.

73. WYSZYNSKI, D. F. & BEATY, T. H. 1996. Review of the role of potential teratogens in the origin of human nonsyndromic oral clefts. Teratology, 53, 309-17.

74. YOUNG, D. L., SCHNEIDER, R. A., HU, D. & HELMS, J. A. 2000. Genetic and teratogenic approaches to craniofacial development. Critical reviews in oral biology and medicine : an official publication of the American Association of Oral Biologists, 11, 304-17.

75. YU, W., RUEST, L. B. & SVOBODA, K. K. 2009. Regulation of epithelial-mesenchymal transition in palatal fusion. Experimental biology and medicine, 234, 483-91.

76. ZHANG, Y., MORI, T., ISEKI, K., HAGINO, S., TAKAKI, H., TAKEUCHI, M., HIKAKE, T., TASE, C., MURAKAWA, M., YOKOYA, S. & WANAKA, A. 2003. Differential expression of decorin and biglycan genes during palatogenesis in normal and retinoic acid-treated mice. Developmental dynamics : an official publication of the American Association of Anatomists, 226, 618-26.

77. ZHU, H., KARTIKO, S. & FINNELL, R. H. 2009. Importance of gene-environment interactions in the etiology of selected birth defects. Clinical genetics, 75, 409-23.