Universidade Federal do Rio Grande do Sul

Instituto de Ciências Básicas da Saúde

Departamento de Bioquímica

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS: BIOQUÍMICA

BUSCA DE NOVOS ALVOS TERAPÊUTICOS NO TRATAMENTO DE NEUROBLASTOMA UTILIZANDO FERRAMENTAS DE BIOLOGIA DE SISTEMAS

Mestrando: Ricardo D'Oliveira Albanus

Orientador: José Cláudio Fonseca Moreira

Co-orientador: Rodrigo Juliani Siqueira Dalmolin

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas: Bioquímica como requisito para obtenção do grau de Mestre em Bioquímica

Porto Alegre,

Abril de 2014

CIP - Catalogação na Publicação

Albanus, Ricardo D'Oliveira
Busca de Novos Alvos Terapêuticos no Tratamento de Neuroblastoma Utilizando Ferramentas de Biologia de Sistemas / Ricardo D'Oliveira Albanus. -- 2014.
66 f.

Orientador: José Cláudio Fonseca Moreira.
Coorientador: Rodrigo Juliani Siqueira Dalmolin.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Instituto de Ciências Básicas da Saúde, Programa de Pós-Graduação em Ciências Biológicas: Bioquímica, Porto Alegre, BR-RS, 2014.

1. Neuroblastoma. 2. Bioinformática. 3. Biologia de Sistemas. I. Moreira, José Cláudio Fonseca, orient. II. Dalmolin, Rodrigo Juliani Siqueira, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

SUMÁRIO

Para minha irmã, Silvia.

E um futuro onde os tijolos ficaram para trás.

"*We chose this particular moment in the program… to* tango*"*

Tom Waits – *The Tango (Shadows of Intolerance)*

AGRADECIMENTOS

Minha estadia no Departamento de Bioquímica não pode ser lembrada sem que algumas pessoas muito importantes recebam o devido reconhecimento por todo o auxílio que deram ao meu crescimento.

Em primeiro lugar, ao professor **José Cláudio Fonseca Moreira**, que desde o segundo semestre de graduação me mostrou como a biologia é fascinante. Depois, no final do curso, me ensinou que o processo de fazer ciência é ainda mais envolvente. Espelho-me em ti como pessoa, professor, orientador, cientista e (por que não?) músico. Três anos é muito pouco tempo para aprender tudo que tens a ensinar.

Ao colega, co-orientador e irmão mais velho **Rodrigo Dalmolin**, por toda a paciência, amizade e conhecimento compartilhados ao longo desses anos.

Ao co-co-orientador e Mestre dos Magos **Mauro Castro**, pelo contato fascinante com uma mente excepcionalmente criativa e uma companhia extremamente agradável.

Aos colegas de bioinformática **Jesus (Luís Henrique)**, **Ivaine**, **Mariana** e **Mariel**, que tornaram o trabalho em equipe um grande prazer. Em pouco tempo vocês passaram de colegas a grandes amigos.

Aos **colegas de Laboratório 32**, que fizeram da vivência neste departamento uma fonte de risadas e *insights* científicos.

A todos os **funcionários do Departamento de Bioquímica**. Em especial, aos funcionários da secretaria, **Cléia**, **Douglas** e **Giordano**, por toda a ajuda que prestam aos alunos da Bioquímica.

A **Rosaura Lembert**. Um porto seguro em águas outrora turbulentas.

E por último, servindo de alicerce para todo esse corpo de texto, a meus pais, **Ricardo** e **Celia**. Nada disso seria possível sem o apoio incondicional deles. Obrigado de coração por serem quem vocês são. Amo vocês.

RESUMO

Neuroblastoma é um tumor do sistema nervoso periférico e uma das principais causas de mortalidade infantil por câncer no Brasil e no mundo. A característica marcante desses tumores é sua heterogeneidade clínica – as apresentações variam desde formas benignas e tratáveis até variantes extremamente agressivas, que levam o paciente a óbito em poucos meses. A fim de determinar as melhores estratégias para o tratamento desse câncer, um grande número de pesquisadores tem procurado por novos biomarcadores e alvos terapêuticos. Atualmente, o melhor marcador biológico para a progressão de neuroblastoma é a amplificação do gene *MYCN*, que ocorre na maioria dos casos mais agressivos. Entretanto, esse marcador não é capaz de discriminar entre todos os tipos de pacientes, demonstrando a necessidade de encontrarmos outros marcadores. Neste trabalho, reconstruímos a rede de regulação gênica de neuroblastoma utilizando ferramentas de bioinformática e biologia de sistemas a fim de buscar novos biomarcadores e alvos terapêuticos para esta doença. Através dessa rede, analisamos uma assinatura de genes diferentemente expressos em metástases agressivas, buscando fatores de transcrição que pudessem estar envolvidos na progressão tumoral. Através dessa análise, observamos que *MAX* é um dos reguladores mestres do processo, e variações em sua expressão estavam correlacionadas com o desfecho de pacientes, sugerindo seu potencial como biomarcador. Além disso, também observamos o aumento do conteúdo de MAX em células de linhagem de neuroblastoma humano durante um protocolo de diferenciação experimental, sugerindo que essa proteína tem um papel muito mais central no controle do balanço entre proliferação e diferenciação do que previamente descrito.

ABSTRACT

Neuroblastoma is a peripheral nervous system tumor and one of the main causes of children cancer mortality in Brazil and in the rest of the world. The hallmark of these tumors is their clinical heterogeneity – presentations vary from benign and treatable to extremely aggressive variants, the latter leading the patient to death in just a few months. In order to come up with the best strategies for treatment and management of this cancer, a great number of researchers have been searching for novel biomarkers and therapeutic targets. To this date, the best neuroblastoma biomarker is the amplification of *MYCN* oncogene, which occurs in the majority of aggressive cases. However, this biomarker alone does not suffice to discriminate among all patients types, illustrating the need for finding other biomarkers. In this work, we reverse engineered the neuroblastoma gene regulatory network using systems biology and bioinformatics tools in order to find novel biomarkers and therapeutic targets. Using this network, we analyzed an aggressive metastasis gene signature to find transcription factors that could be involved in tumor progression. Through this analysis, we observed that *MAX* was a master regulator of this process, and that changes in its expression were associated with patient survival, suggesting its role as potential biomarker. Additionally, we observed that MAX content was increased in human neuroblastoma cell lines undergoing experimental differentiation. These results suggest that this protein has a much more central role in regulating the balance between proliferation and differentiation than previously described.

PARTE I

1. INTRODUÇÃO

1.1 Neuroblastoma

1.1.1 Características gerais

Neuroblastoma é um tumor infantil que afeta o sistema nervoso simpático. Do ponto de vista epidemiológico, este tumor é responsável por uma significativa parcela da mortalidade infantil por câncer no mundo (Howman-Giles *et al.*, 2007; Maris *et al.*, 2007; Parise *et al.*, 2006; Spix *et al.*, 2006). Em função de sua importância, um grande número de investigações está sendo feita a fim de trazer avanços no entendimento de sua biologia e tratamento. Os maiores grupos de estudo e classificação de neuroblastoma atualmente são o *International Neuroblastoma Risk Group* (INRG) e o *Children's Oncology Group* (COG) (Monclair *et al.*, 2009).

A característica mais marcante dos neuroblastomas é sua heterogeneidade clínica. Os desfechos possíveis variam desde a remissão espontânea até um avanço agressivo, que leva o paciente a morte mesmo com terapias multimodais avançadas. Notavelmente, um dos desfechos atribuídos a esse tumor é sua diferenciação espontânea em ganglioneuromas benignos, um processo único deste tipo de câncer. Devido a sua excepcional capacidade de metástase, podem gerar processos de invasão em praticamente todos os tecidos do corpo humano, incluindo ossos, fígado e encéfalo (Howman-Giles *et al.*, 2007; Mills, 2002). Do ponto de vista clínico, o foco de origem mais comum de neuroblastoma é o tecido simpático das glândulas adrenais (35%), que é originário de elementos embrionários da crista neural. Igualmente comum, são os gânglios paraespinhais (35%), mas também são descritos casos onde esta malignidade se apresenta incialmente em tecidos simpáticos do abdome e tórax (20%), pescoço (<5%) e pélvis (<5%) (Howman-Giles *et al.*, 2007). Além disso, aproximadamente

metade dos pacientes apresentam metástases já no diagnóstico, o que torna necessário um tratamento de urgência logo que a doença é detectada.

A faixa etária de incidência mais comum ocorre entre seis meses e três anos de idade, e está diretamente correlacionada com o desfecho da doença. Metade dos pacientes apresentam os sintomas antes dos dois anos de idade, e mais de 90% antes dos cinco anos. A apresentação dos sintomas antes dos 18 meses de vida está relacionada com um prognóstico positivo, correspondendo a uma taxa de sobrevivência em seis anos superior a 90%. Entretanto, pacientes diagnosticados após 18 meses apresentam menos de 25% de sobrevivência durante o mesmo período (Monclair *et al.*, 2009; Schmidt *et al.*, 2000, 2005).

1.1.2 Diagnóstico clínico e a busca por biomarcadores

As apresentações clínicas mais comuns desta doença são o aparecimento de massas abdominais causadas pela presença do tumor. Também comum é o aparecimento de feridas em volta dos olhos, causadas pelo frequente acometimento dos ossos e tecidos moles ao redor das órbitas. Algumas crianças podem apresentar sintomas similares ao de leucemia, como anemia, febre e dor nos ossos, e crianças mais velhas também podem apresentar dificuldades de locomoção sugestivas de artrite (Howman-Giles *et al.*, 2007). Dependendo da localização, os tumores podem causar obstrução da saída urinária e também causar hipertensão e irritabilidade devido à secreção de catecolaminas. Esse aumento de catecolaminas pode induzir diarreias, que também podem ser usadas para detectar a doença. A forma de diagnóstico mais utilizada na clínica, além da inspeção pelo pediatra, é o exame rotineiro de raios X na região torácica (Schwab *et al.*, 2003).

Os tratamentos para neuroblastoma são baseados nas classificações de subtipos, que dependem de dados de imagem, parâmetros biológicos e bioquímicos (Ambros *et al.*, 2009). Vários esforços têm sido feitos para classificar os neuroblastomas em diferentes grupos de risco e, atualmente, o sistema de classificação mais utilizado é o *International Neuroblastoma Staging System* (INSS). Este sistema classifica os tumores levando em conta a localização do tumor e o grau de comprometimento de outros tecidos (Tabela 1).

Tabela 1: Sistema de classificação INSS

Estágio	Características
1	Tumor unilateral localizado, sem comprometimento de outros órgãos e nódulos linfáticos. Pode ser removido por ressecção cirúrgica apenas.
2A	Tumor unilateral localizado, mas que não pode ser totalmente removido cirurgicamente. Sem comprometimento de outros órgãos e nódulos linfáticos.
2B	Idem ao estágio 2A, mas com comprometimento de nódulos linfáticos ipsilaterais ao tumor primário.
3	O tumor não comprometeu para partes distantes do corpo, mas pode: a) Não ser completamente removido cirurgicamente e cruzar a linha medial, comprometendo ou não nódulos linfáticos próximos. b) Ainda estar no local primário sem atravessar a linha medial, mas comprometendo nódulos linfáticos contralaterais próximos. c) Estar no centro do corpo, crescendo para os dois lados (diretamente ou através de nódulos linfáticos). Não pode ser removido por ressecção cirúrgica apenas.
4	O tumor comprometeu áreas distantes, como nódulos linfáticos distantes, pele, fígado, ossos, medula óssea ou outros órgãos.
4S	Similar ao estágio 4, mas a criança tem menos de um ano de vida. O tumor se apresenta unilateralmente e comprometeu nódulos ipsilaterais distantes, mas não contralaterais. O neuroblastoma pode comprometer fígado, pele e medula óssea, entretanto, menos de 10% das células da medula podem ser cancerosas.

Do ponto de vista molecular, o uso de informações genômicas específicas desses tumores tem sido incluído nas abordagens de estratificação de riscos. Entretanto, ainda são poucos os marcadores capazes de oferecer pistas sobre o desfecho do paciente.

Estudos citogenéticos em biopsias sugerem que alterações segmentais ou ganhos cromossômicos inteiros são um importante separador de desfecho em pacientes (Ooi *et al.*, 2012). Alterações cromossômicas específicas estão associadas com mau prognóstico. Destas, destaca-se a perda de 11q, ganhos na região 17q e perda de heterozigosidade em 1p. Ganhos de cromossomos que resultam em hiperploidias, entretanto, são indicadores de bom prognóstico e baixo risco para os pacientes (Maris e Matthay, 1999; Schwab *et al.*, 2003).

A aplicação de técnicas citogenéticas em nível clínico, todavia, é restrita devido ao custo e às dificuldades do protocolo. Uma alternativa que surge é o uso de genes marcadores, que podem ser facilmente detectados com técnicas rotineiras de biologia molecular. Alguns avanços recentes permitiram melhorar o entendimento da predisposição de neuroblastoma familiar, uma variante mais rara da doença que corresponde a cerca de 1-2% dos casos (Tonini *et al.*, 2003). Apesar de sua baixíssima incidência entre essa classe de tumores, a identificação dos genes de suscetibilidade guiou estudos correlacionados em neuroblastoma esporádico. Os casos esporádicos são provavelmente decorrentes da interação entre polimorfismos relativamente comuns em múltiplos alelos, mas com baixa penetrância, caracterizando um típico caso de uma doença geneticamente complexa. A fim de esclarecer os fatores de suscetibilidade para as variantes esporádicas, estudos de associação genômica (*genome-wide association study*, GWAS) estão sendo aplicados em neuroblastoma. Em uma revisão recente (Deyell e Attiyeh, 2011), foi apontado que mutações deletérias em alelos dos genes *PHOX2B* e *ALK* correspondem a 90% dos casos de neuroblastoma familiar. Para os casos esporádicos, entretanto, o repertório de polimorfismos apontados é consideravelmente mais extenso (Tabela 2).

Os padrões de expressão gênica em neuroblastoma espelham sua heterogeneidade clínica. São descritas inúmeras assinaturas de genes diferencialmente expressos em várias análises de microarranjos. Alguns estudos apontam expressão aberrante de membros da via de sinalização das neurotrofinas, que regulam processos de crescimento e diferenciação celular. Nesses estudos, foi sugerido que a alta expressão de *TRKA* (*NTRK1*) e *TRKC* (*NTRK3*) está associada com prognósticos favoráveis, enquanto que a alta expressão de *TRKB* (*NTRK2*) está envolvida com desfechos desfavoráveis (Deyell e Attiyeh, 2011; Schramm *et al.*, 2005).

Tabela 2: Resultados de GWAS em neuroblastoma esporádico (adaptado de Deyell e Attiyeh, 2011)

Gene	Localização cromossômica	Fenótipo de neuroblastoma
FLJ22536	6p22	Alto risco
BARD1	2q35	Alto risco
LMO1	11p15	Alto risco
DUSP12	1q23	Baixo risco
DDX4	5q11	Baixo risco
IL31RA	5q11	Baixo risco
HSD17B12	11p11	Baixo risco
NBPF23	1q21	Não-específico

Atualmente, o melhor marcador genético conhecido para a progressão de neuroblastoma é a amplificação do oncogene *MYCN* (Bagatell *et al.*, 2009; Tonini *et al.*, 1997; Westermann e Schwab, 2002). Estudos realizados durante as últimas décadas detectaram correlações significativas do número de cópias desse gene com o desfecho negativo de pacientes. Amplificações menores (*i.e.* menores que três cópias) podem apresentar desfechos positivos, mas dificilmente pacientes com mais de 10 cópias sobrevivem à doença. Em casos extremos, já foram detectados números superiores a 400 cópias (Oberthuer *et al.*, 2010, material suplementar). Esses pacientes apresentam tumores metastáticos resistentes a praticamente todos os tratamentos conhecidos.

Apesar de ser um relativo bom marcador, a amplificação de *MYCN* não pode ser amplamente aplicada na clínica, uma vez que as existem variantes de neuroblastoma que não apresentam amplificação deste oncogene e se apresentam com os mais variados graus de letalidade. Algumas dessas, inclusive, são extremamente agressivas e igualmente resistentes a tratamentos multimodais (Asgharzadeh *et al.*, 2006).

1.1.3 Relevância da pesquisa e a situação brasileira

A busca de novos marcadores para os casos citados anteriormente é de grande importância para o bem-estar dos pacientes. Por serem crianças ou recém-nascidos, os pacientes de neuroblastoma estão ainda passando por etapas fundamentais de seu desenvolvimento. O uso de tratamentos antiproliferativos em crianças com neuroblastoma está associado a sequelas permanentes como surdez parcial ou total e danos no sistema respiratório, sendo imperativo que seu tratamento seja feito com a maior parcimônia e precisão possível (Wagner e Danks, 2009). A fim de buscar esses novos marcadores, inúmeros estudos estão sendo feitos no mundo todo. Esses estudos têm por objetivo buscar relações entre mutações e alterações na expressão de diversos genes com a apresentação clínica dos tumores e seu desfecho. Em um conjunto de estudos recentes, que utilizou técnicas de mineração de dados em biópsias de neuroblastomas, foi inferida uma assinatura de 59 genes com uma boa capacidade preditiva, mas apenas quando utilizada em conjunto com outros dados clínicos, como estágio INSS, idade dos pacientes e amplificação do próprio *MYCN* (De Preter *et al.*, 2010). Apesar desses esforços, a biologia desse câncer continua elusiva e ainda não foi possível detectar novos marcadores com o poder de discriminação atribuídos à amplificação de *MYCN*.

No Brasil, neuroblastoma ocupa uma posição relevante nas estatísticas oficiais. Esses tumores são responsáveis por cerca de metade dos casos de neoplasias malignas em infantis lactentes[1] e por até 10% de todos os cânceres em crianças com idade inferior a 15 anos (Cartum, 2012). Embora pouco documentados no resto do país de uma maneira geral, existem estudos demonstrando o perfil epidemiológico dessa doença na Região Sul, que sugerem índices de incidência semelhantes ao resto do mundo (Parise *et al.*, 2006; Pianovski *et al.*, 2006).

1.2 Abordagens *high-throughput* para o estudo patologias humanas

1.2.1 A biologia de sistemas

Com a popularização das técnicas de alto desempenho (*high-throughput*) para geração de dados genômicos, transcriptômicos e interatômicos, um novo paradigma tem surgido na biologia moderna. O desenvolvimento de ferramentas de coletas de alta densidade de dados genômicos e metabólicos (*microarrays*, *yeast two-hybrid screens*, etc.) permitiu, nas duas últimas décadas, acesso a uma parcela significativa dos componentes de uma célula em qualquer dado instante, tornando possível vislumbrar a intrincada dinâmica das redes genômicas e proteicas (Figura 1). Para analisar este grande volume de dados é necessária a elaboração de avançados modelos matemáticos e estatísticos que compreendam as mudanças topológicas e propriedades dinâmicas oriundas dessas interações tão multíplices, de forma a criar uma visão quantitativa e, portanto, comparável das complexas redes que regulam e determinam o comportamento celular (Harrington, Jensen e Bork, 2008). Do ponto de vista da pesquisa, se torna um

[1] Crianças entre 28 dias e dois anos de idade.

grande desafio analisar e interpretar essa sempre crescente quantidade de dados a fim de se buscar resposta em perguntas biológicas de grande relevância.

Uma disciplina que tem galgado um lugar de destaque nesta busca é a biologia de sistemas, que, auxiliada por ferramentas de bioinformática, possibilita uma interpretação coerente desse imenso volume da informação. Esse foco de estudo surgiu a partir da compreensão de que as características estruturais das redes de interações moleculares de uma célula são compartilhadas em grande grau com outros sistemas complexos já estudados (Hartwell *et al.*, 1999; Yamada e Bork, 2009). Esta universalidade inesperada indica que leis similares podem reger a formação e evolução de redes de alta complexidade na natureza, permitindo que utilizemos o conhecimento adquirido no estudo de sistemas não biológicos bem mapeados para a caracterização das relações intrincadas contidas na organização das funções celulares (Barabási e Oltvai, 2004).

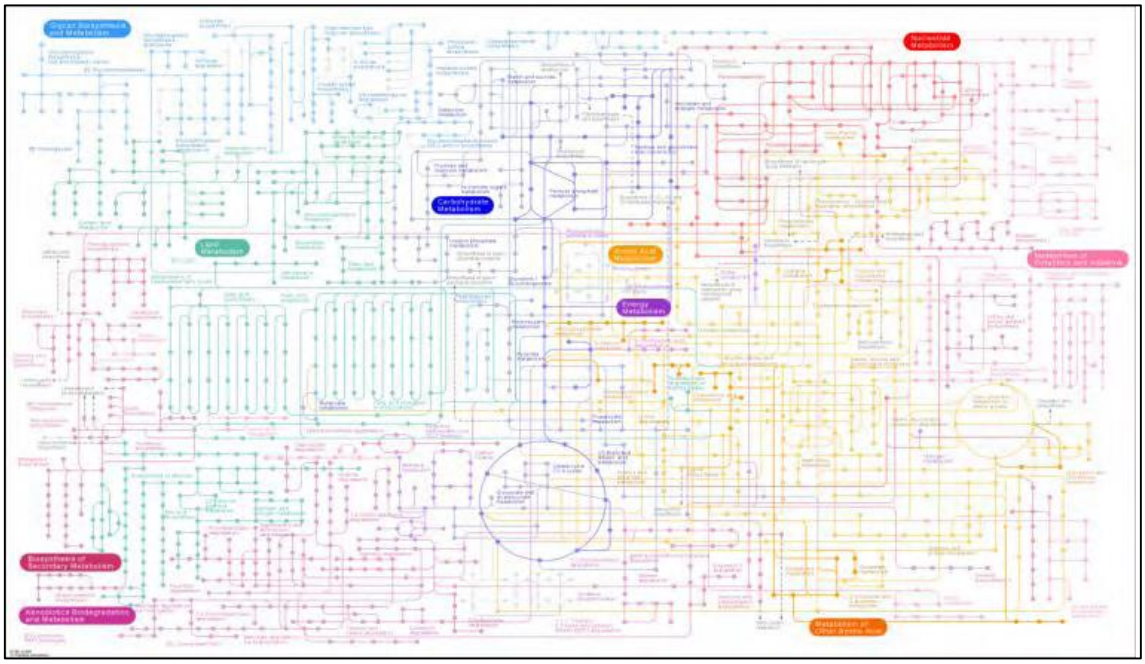

Figura 1. Mapa das redes bioquímicas de referência para *Homo sapiens* no repositório de dados interatômicos KEGG (http://genome.jp).

1.2.2 Um novo paradigma no estudo de patologias

A pesquisa de novos tratamentos para as mais variadas patologias humanas vem se beneficiando significativamente com a biologia de sistemas e a bioinformática nos últimos anos. Com a capacidade de analisar um sistema biológico simultaneamente em seus diferentes níveis de complexidade (*i.e.* organismo, órgãos, tecidos, células e biomoléculas), estudos utilizando as ferramentas de biologia de sistemas e bioinformática puderam trazer avanços sólidos no tratamento e diagnóstico de diversas doenças. Uma das estratégias mais comuns é a busca por conjuntos de biomarcadores capazes de discernir entre estados patológicos e seus respectivos estados normais, o que permitiria estratificar pacientes usando dados de genômica funcional. Destaca-se a busca de genes que podem ser determinantes para um processo particular, como por exemplo, o desenvolvimento de agressividade em um tumor ou diferenciação de tipos celulares distintos. Para esse tipo de estudo, usualmente são coletados dados de expressão globais de um número suficiente de amostras representativas de cada estado. Essas amostras são então comparadas por meio de estatísticas apropriadas para expressão diferencial (*e.g.* regressão linear) a fim de se procurar genes diferencialmente expressos entre os estados.

A pesquisa em câncer, particularmente, tem avançado muito com essas técnicas. Existem atualmente inúmeros repositórios de dados específicos para cada tipo de tumor, oferecendo informações sobre mutações genômicas (*i.e.* dados de SNPs, deleções e ganhos cromossomais), perfis de expressão gênica associados a dados de sobrevivência ou informações clínicas, etc. Esses repositórios podem ser públicos, apoiados ou até gerenciados por agências como o NCBI e o EMBL. Essa quantidade de dados, entretanto, ainda carece de ferramentas suficientemente poderosas para utilizar plenamente seu potencial de melhorar o diagnóstico e o tratamento de pacientes. A fim

de tentar reverter esse quadro, uma comunidade crescente de pesquisadores, profissionais e desenvolvedores está se unindo para gerar novas ferramentas de análise. Das várias iniciativas existentes atualmente, destaca-se o grupo *Bioconductor* (Gentleman *et al.*, 2004), que tem se tornado um dos mais importantes diretórios de desenvolvimento de ferramentas para bioinformática. A contrapartida tecnológica dessas questões também é muito relevante, já que a validação de novas ferramentas de bioinformática para estudos de associação genômica é mister para a descoberta de novos fármacos, novas abordagens terapêuticas e novos testes diagnósticos.

Recentemente, estudos utilizando técnicas de alto desempenho possibilitaram o desenvolvimento de metodologias para gerar redes transcricionais (Babu, Lang e Aravind, 2009; Basso *et al.*, 2005; Margolin *et al.*, 2006). Essas técnicas utilizam modelos estatísticos robustos para determinar relações direcionais entre fatores de transcrição e genes-alvos. Somente nos últimos anos foi possível a geração de dados em quantidade suficiente para esse tipo de análise, uma vez que se necessita de um volume consideravelmente alto de informação para diminuir a relação ruído/sinal dos perfis de expressão. Uma das metodologias mais consolidadas na reconstrução de redes regulatórias envolve o uso de ferramentas de teoria da informação para inferir essas redes através de correlações sutis entre os padrões de co-expressão de milhares de genes (Margolin *et al.*, 2006). A partir de uma rede modelada, é possível fazer questionamentos sobre quais fatores de transcrição potencialmente influenciam determinadas assinaturas gênicas, possibilitando estabelecer relações causais entre estados patológicos e um ou mais genes reguladores. Essa análise aprofundada permite não só o descobrimento de novos biomarcadores e alvos terapêuticos como também tem ampliado significativamente o nosso entendimento sobre os mecanismos de ação de alguns fatores de transcrição.

O grupo de pesquisa liderado por Andrea Califano, da Universidade de Columbia, nos Estados Unidos, tem aplicado com sucesso essa metodologia no estudo de diversos tipos de câncer. Dentre as aplicações bem-sucedidas, destaca-se a reconstrução das redes regulatórias de glioma (Carro *et al.,* 2010) e de células B (Lefebvre *et al.*, 2010), que permitiram descobertas relevantes a respeito da regulação dos processos de transformação celular e malignidade. Essas abordagens apontam novos alvos gênicos para o tratamento de glioma e linfoma. Seguindo as mesmas premissas, Fletcher e colaboradores (Fletcher *et al.,* 2013) reconstruíram a rede regulatória de câncer de mama a partir de dados de expressão de biópsias humanas e explicaram o papel do fator FGFR2 nesta doença. Este gene é constantemente trazido à tona em estudos de associação genômica, entretanto, seu papel permanecia completamente desconhecido. Utilizando células tratadas com moléculas perturbadoras de FGFR2, esses autores conseguiram determinar quais os fatores de transcrição que respondem a essa sinalização, explicando os mecanismos biológicos pelos quais esse fator atua na progressão tumoral do câncer mamário.

2. OBJETIVOS

Tendo em vista o exposto na Introdução, este trabalho teve por objetivo ampliar o conhecimento atual sobre neuroblastoma, visando propor novos alvos terapêuticos ou marcadores para a progressão desta doença. Mais especificamente, postulamos que a utilização de ferramentas modernas de biologia de sistemas e bioinformática seriam capazes de oferecer uma abordagem eficaz no diagnóstico desse tumor. Sendo assim, os objetivos específicos deste trabalho foram:

1) Reconstruir a rede regulatória de neuroblastoma utilizando dados públicos de expressão gênica;
2) Sobre a rede inferida, analisar uma assinatura de tumores agressivos a fim de buscar unidades regulatórias que possam estar enriquecidas no fenótipo tumoral;
3) Verificar em conjuntos de dados independentes a validade das unidades regulatórias na discriminação entre coortes de pacientes de neuroblastoma;
4) Validar *in vitro*, utilizando cultura de células de neuroblastoma humano (SH-SY5Y), os efeitos das unidades regulatórias sobre parâmetros de progressão tumoral.

PARTE II

REVERSE ENGINEERING THE NEUROBLASTOMA REGULATORY NETWORK UNCOVERS MAX AS ONE OF THE MASTER REGULATORS OF TUMOR PROGRESSION[2]

Ricardo D'Oliveira Albanus, Rodrigo Juliani Siqueira Dalmolin, Mauro Antônio Alves Castro, Matheus Augusto de Bittencourt Pasquali, Vitor de Miranda Ramos, Daniel Pens Gelain, José Cláudio Fonseca Moreira

[2] Artigo publicado no periódico *PLoS One* (doi: 10.1371/journal.pone.0082457) em dezembro de 2013; 8(12):e82457.

OPEN ACCESS Freely available online

PLOS ONE

Reverse Engineering the Neuroblastoma Regulatory Network Uncovers MAX as One of the Master Regulators of Tumor Progression

Ricardo D'Oliveira Albanus*, Rodrigo Juliani Siqueira Dalmolin, Mauro Antônio Alves Castro, Matheus Augusto de Bittencourt Pasquali, Vitor de Miranda Ramos, Daniel Pens Gelain, José Cláudio Fonseca Moreira

Departamento de Bioquímica, Núcleo de Bioinformática, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brasil

Abstract

Neuroblastoma is the most common extracranial tumor and a major cause of infant cancer mortality worldwide. Despite its importance, little is known about its molecular mechanisms. A striking feature of this tumor is its clinical heterogeneity. Possible outcomes range from aggressive invasion to other tissues, causing patient death, to spontaneous disease regression or differentiation into benign ganglioneuromas. Several efforts have been made in order to find tumor progression markers. In this work, we have reconstructed the neuroblastoma regulatory network using an information-theoretic approach in order to find genes involved in tumor progression and that could be used as outcome predictors or as therapeutic targets. We have queried the reconstructed neuroblastoma regulatory network using an aggressive neuroblastoma metastasis gene signature in order to find its master regulators (MRs). MRs expression profiles were then investigated in other neuroblastoma datasets so as to detect possible clinical significance. Our analysis pointed MAX as one of the MRs of neuroblastoma progression. We have found that higher MAX expression correlated with favorable patient outcomes. We have also found that MAX expression and protein levels were increased during neuroblastoma SH-SY5Y cells differentiation. We propose that MAX is involved in neuroblastoma progression, possibly increasing cell differentiation by means of regulating the availability of MYC:MAX heterodimers. This mechanism is consistent with the results found in our SH-SY5Y differentiation protocol, suggesting that MAX has a more central role in these cells differentiation than previously reported. Overexpression of MAX has been identified as anti-tumorigenic in other works, but, to our knowledge, this is the first time that the link between the expression of this gene and malignancy was verified under physiological conditions.

Citation: Albanus RD, Juliani Siqueira Dalmolin R, Alves Castro MA, Augusto de Bittencourt Pasquali M, de Miranda Ramos V, et al. (2013) Reverse Engineering the Neuroblastoma Regulatory Network Uncovers MAX as One of the Master Regulators of Tumor Progression. PLoS ONE 8(12): e82457. doi:10.1371/journal.pone.0082457

Editor: Javier S Castresana, University of Navarra, Spain

Received August 26, 2013; **Accepted** October 23, 2013; **Published** December 5, 2013

Funding: This paper was funded by CNPq (http://www.cnpq.br/), CAPES (http://www.capes.gov.br/) and FAPERGS (http://www.fapergs.rs.gov.br/) funding agencies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: ricardo.albanus@ufrgs.br

Introduction

Neuroblastoma is the most common extra-cranial solid tumor and one of the leading causes of cancer mortality in children worldwide [1–4]. These tumors are originated from embryonic elements of the neural crest and sympathetic nervous system, usually developing in the adrenal glands, but also arising in nervous tissues of the neck, thorax, abdomen, and pelvis. The metastatic capacity of this type of cancer is notable, being able of compromising almost any tissue in the human body [2,5]. Another distinguishing feature of this disease is its clinical heterogeneity. The possible endpoints span from complete remission to patient death, even with advanced multimodal therapy [6]. Numerous efforts have been made in order to sort neuroblastoma patients in separate risk groups, such as the International Neuroblastoma Risk Group (INRG) staging system [7], which is based on clinical factors and imaging studies, and the Children's Oncology Group (COG) risk stratification schema, which complements the former with molecular aspects. The molecular approaches for assessing patient clinical outcome are made by detecting DNA copy number alterations or by searching for more specific segmental aberrations, such as chromosomes 11q loss and 17q gain [8,9]. The most reliable molecular classifier, however, is the amplification of the MYCN oncogene, which is linked to grim prognosis in the majority of cases [10–12].

Lower grade neuroblastoma patients can be treated with surgical resection alone, and may even be subject to spontaneous remission without any intervention. Patients with metastatic MYCN amplified tumors have the highest mortality rate and are usually unresponsive to advanced multimodal treatment [6,7,13,14]. Patients with MYCN non-amplified metastatic tumors, however, present highly variable outcomes, but few endpoint predictors have been described for this group. Currently, the best known prognostic indicator is the age at which the tumor is diagnosed [7]. Patients younger than 18 months usually have better prognosis, with more than 90% 6-year event-free survival for patients younger than one year. Patients older than 18 months, on the other hand, suffer higher mortality rates and may present less than 25% 6-years event-free survival [15,16]. Several high-throughput studies have been made over the last years to understand the biology underlying this clinical variability, and at least one of them was aimed exclusively at these metastatic MYCN non-amplified patients [17]. Despite intensive study, few predictors have been brought forward, demanding increased efforts to understand this remarkable disease.

In recent years, the availability of gene expression studies has allowed novel strategies for understanding cancer biology. One of such is the use of mutual information models for inferring the regulatory networks of transcription factors (TFs) and their transcriptional targets (or regulons) in the gene expression profile of a given set of samples [18–20]. This methodology allows the detection of potential causal relationships between TFs and specific cancer signatures. As these statistics depend on rather large sample sizes (n>80), only recently are they being successfully applied in biology [21–23]. In this paper, we have reconstructed the neuroblastoma regulatory network in order to find TFs involved in the transition from primary tumors to highly aggressive bone marrow metastasis. We have found evidences that MAX is one of the master regulators of tumor progression, possibly by being an additional regulatory step for the availability of MYC:MAX heterodimers to regulate transcription and increase proliferation. We have also found evidences that MAX plays a more prominent role in SH-SY5Y cells differentiation than previously described.

Results

Neuroblastoma regulatory network and master regulator analysis

Through our regulatory network reconstruction workflow (Figure 1), we have identified 15,713 targets for 1,363 TFs in the first reconstructed neuroblastoma regulatory network (GSE16476), and 4,039 targets for 705 TFs in the second (GSE3960) (Figure 2). We have found eight master regulators (MRs) common to both networks (Table S1) using as query an aggressive neuroblastoma metastatic gene signature. We have chosen only regulons common to both networks so as to improve the specificity of our MRs analysis. However, by this criterion alone, there was still an elevated chance of finding nonspecific TFs. To overcome this problem, we have used the metastatic neuroblastoma gene signature to query a healthy mammary tissue network (GSE10780) [24] and use it as a positive control to detect MRs that are not tissue-specific to neuroblastoma. We then re-ran the MR analysis within the two neuroblastoma networks using as query the meta-PCNA proliferative gene signature [25]. This last step was performed in order to exclude MRs that were directly involved with proliferation. Of the eight MRs common to both networks, only MAX, TFEC, and ZNF101 remained after these specificity tests.

Regulon overlap and motif analysis

In order to validate the power of our pipeline for detecting biologically significant regulons, we have compared the composition of each regulon across the two neuroblastoma networks and verified whether their members were enriched with the binding site motif of its MR. We have found less than 1% overlap in the genes potentially regulated by MAX and ZNF101 in both networks, and 25% for TFEC. We proceeded to search for regulatory motifs in the targets sequences flanking regions. Because there are no available motifs for TFEC and ZNF101 in public databases, we could only query for the known MAX motif V$MAX, obtained at the JASPAR database [26,27]. We have found significant enrichment for V $MAX motif in both networks (p=0.038), indicating that our pipeline predicted correctly both MAX regulons, albeit their different compositions. Although there is no public TFEC motif available, this transcription factor is a member of the MITF bHLH family, which is closely related to MAX and MYC families [28]. Members of these families have similar CACGTG binding motifs, meaning that we can query the TFEC regulon using V $MAX motif to determine whether it is regulated by a bHLH transcription factor. TFEC regulon was significantly enriched with this motif (p=0.003), suggesting that it is indeed regulated by a bHLH factor.

Given that the MYC/MAX/MAD network may regulate up to 15% of the human genome [29–31], one would not expect finding substantial overlap in both networks, particularly when taking into account the significant genetic heterogeneity across tumors of the same type, which could greatly affect the specificity of such an important regulatory system. Taking this into consideration and the fact that both MAX regulons were significantly enriched with V$MAX binding motif, we chose to carry our analysis with this gene. ZNF101 was not analyzed further because it did not have corresponding probes the datasets below.

Clinical relevance of MAX and TFEC expression in neuroblastoma

To access whether MAX and TFEC expression could be related to patient outcome, we have analyzed the GSE3446 dataset [17]. This dataset consists of expression profiles of primary tumor biopsies obtained at diagnosis from neuroblastoma patients who (i) either had relapse after five years (n=46), (ii) did not present disease progression in the same period (n=56), and (iii) from tumors obtained at progression (n=12 + 3 obtained both at diagnosis and relapse). All patients in this study had untreated metastatic MYCN non-amplified tumors. 74 of the 102 patients studied at diagnosis

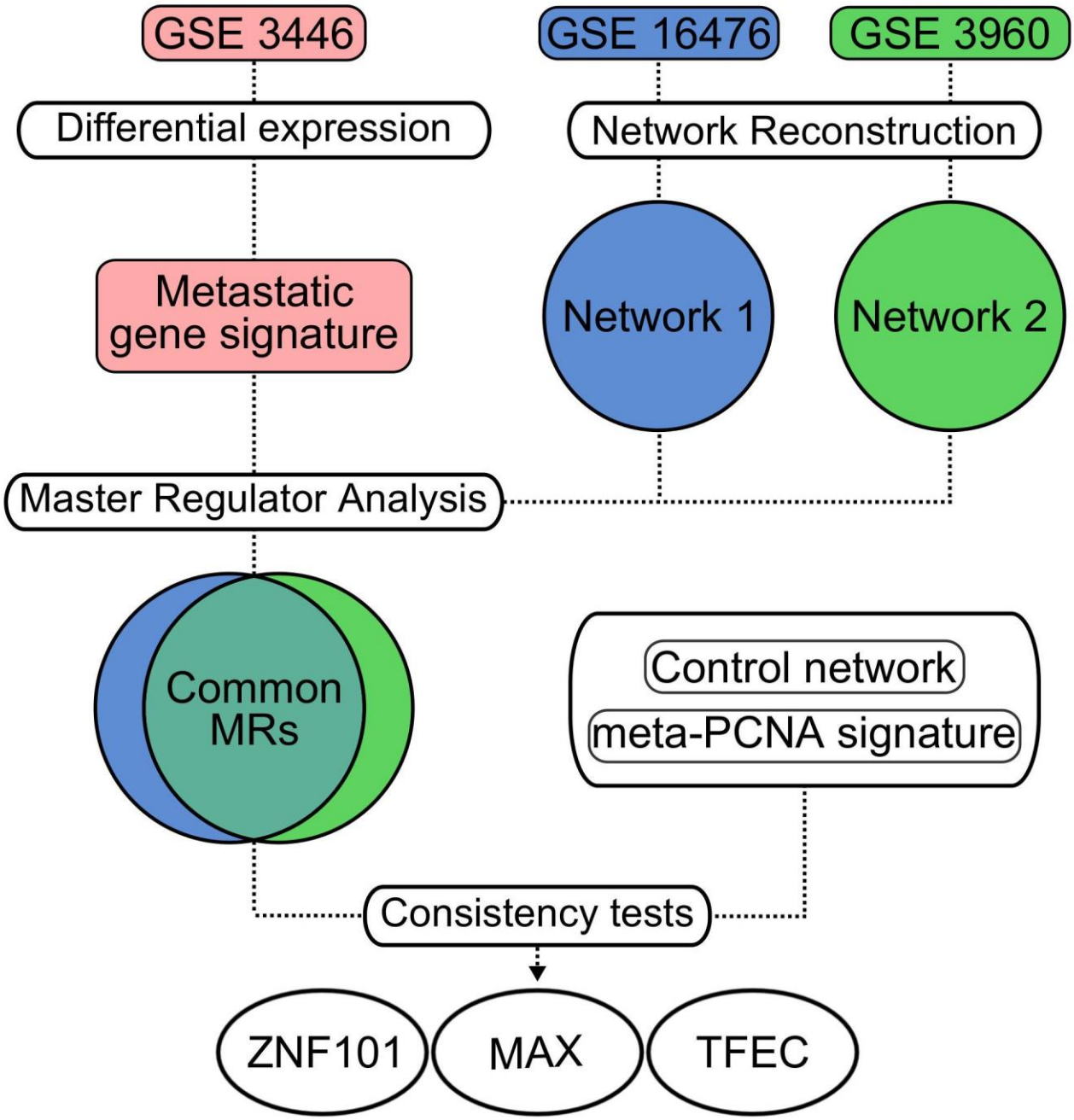

Figure 1: Schematic representation of the workflow used for reconstructing the neuroblastoma network and searching for master regulators of a metastatic gene signature.

Figure 1. Schematic representation of the workflow used for reconstructing the neuroblastoma network and searching for master regulators of a metastatic gene signature.

doi: 10.1371/journal.pone.0082457.g001

were considered high risk by COG risk stratification, and the rest was ranked as intermediate risk. We have found that MAX expression was significantly higher in patients who did not present disease setback in five years as opposed to those who had or were already in relapse (Figure 3), indicating that this

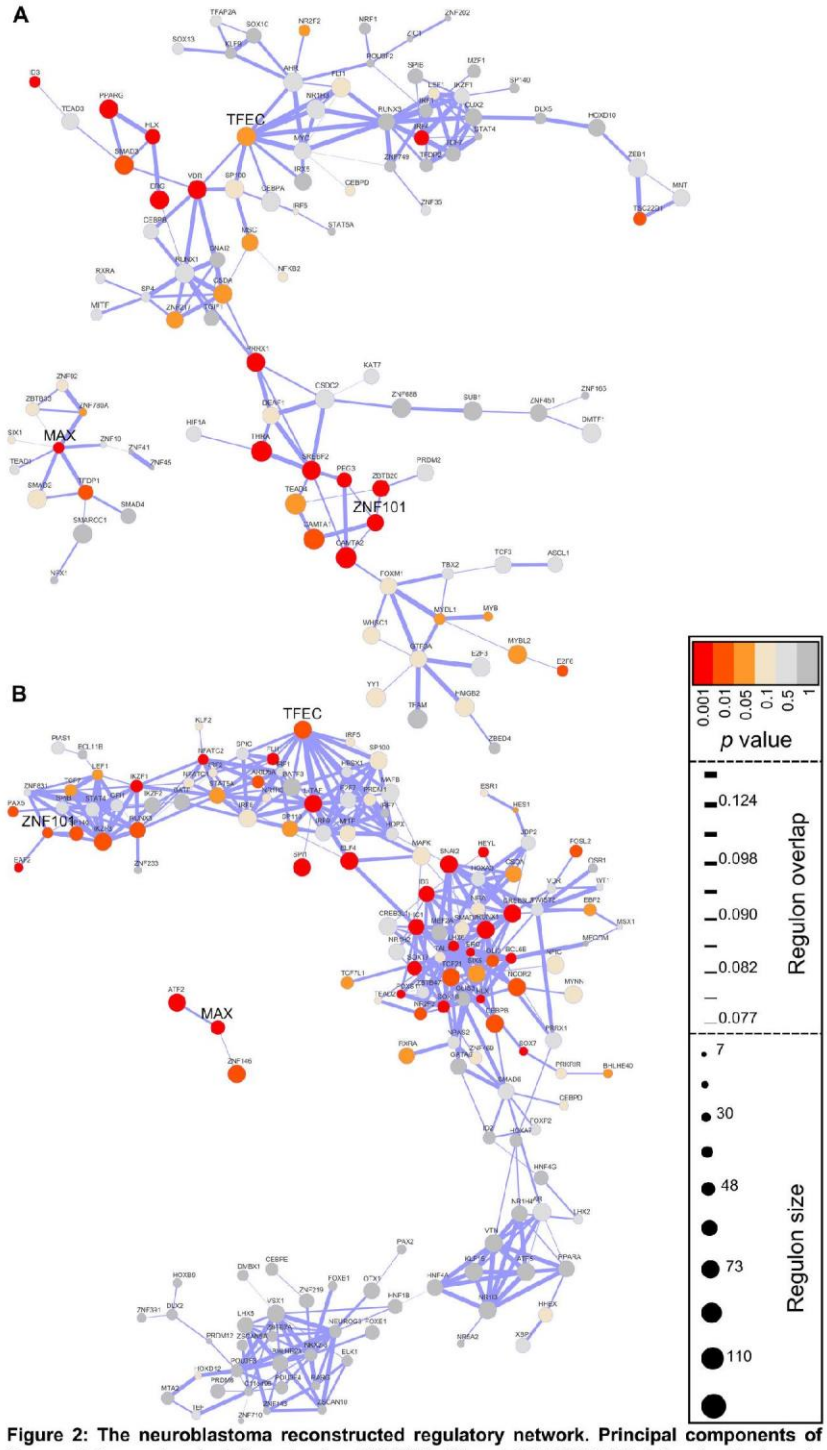

Figure 2: The neuroblastoma reconstructed regulatory network. Principal components of the regulatory networks inferred using GSE3996 (A) and GSE16476 (B) datasets. Each node represents a regulon, which is named by its regulator transcription factor. Node sizes are proportional to the number of regulon members, node colors are representative of enrichment significance, and edges widths are proportional to regulon overlap (using Jaccard similarity).

Figure 2. The neuroblastoma reconstructed regulatory network. Principal components of the regulatory networks inferred using GSE3996 (A) and GSE16476 (B) datasets. Each node represents a regulon, which is named by its regulator transcription factor. Node sizes are proportional to the number of regulon members, node colors are representative of enrichment significance, and edges widths are proportional to regulon overlap (using Jaccard similarity).

doi: 10.1371/journal.pone.0082457.g002

gene plays an important role in disease progression. We did not find any significant alterations in TFEC expression.

We have also analyzed two neuroblastoma survival cohorts with gene expression data. In the first cohort (E-TABM-38, *n*=130) [32], we have found that lower MAX expression correlated with decreased patient survival (Figure 4a), corroborating our previous result. However, this was not the case with the second cohort (E-MTAB-179, *n*=478) [33], in which higher MAX expression significantly correlated with poor survival (Figure 4b). We could not analyze TFEC because these studies were made with custom array platforms that did not contain probes for this gene. This was also the case with the regulons members themselves, which were poorly represented in these platforms and could not be analyzed further. We chose to analyze these genes in other types of cancers so as to detect if there was a general trend for their expression which could confirm our previous results with MAX and shed some light at TFEC. Using the Kaplan-Meier Plotter web tool [34,35], we have found that higher MAX expression was associated with improved prognosis in breast cancer (Figure S1), in lung cancer (Figure S2), and not related to prognosis in ovarian cancer cohorts (Figure S3). As for TFEC, we have found association with higher expression of this gene and improved lung cancer survival (Figure S4). We have not found associations with TFEC expression in breast and ovarian cancer outcome (Figures S5 and S6, respectively). Detailed results from this analysis are presented in Table 1.

Role of MAX and TFEC in SH-SY5Y differentiation

To understand whether the alterations detected previously in MAX expression could be involved in neuroblastoma cells differentiation and, as such, provide an explanation to why there seems to be a correlation with patient outcome, we chose to study a dataset of neuroblastoma SH-SY5Y cells undergoing differentiation (GSE9169) [36]. In this study, the authors differentiated cells by treatment with retinoic acid for 8 days, with further addition of brain-derived neurotrophic factor (BDNF) after the 5th day. We have found a significant increase in MAX expression starting at the 3rd and 5th days, which lasted until the end of the differentiation protocol (Figure 5). We have not found any significant alterations in TFEC expression during the course of this experiment (data not shown).

We have also performed a similar SH-SY5Y differentiation protocol in our laboratory and quantified MAX protein content during differentiation. We have found an increased amount of MAX after the 4th day of experiment (Figure 6), demonstrating that the changes in mRNA expression are indeed reflected at protein level in neuroblastoma cells.

Discussion

Neuroblastoma is clinically very heterogenic, and one of the possible outcomes for this disease is spontaneous tumor regression [1,2]. There are reported cases of differentiation of these tumors into benign ganglioneuromas, but the mechanisms by which this processes can occur are still unknown [8,37,38]. We have queried our reconstructed neuroblastoma network with a genetic signature related to bone marrow metastases, which are considered grim clinical prognosis for neuroblastoma patients [3,39]. We have observed that MAX regulon was significantly enriched with differentially expressed genes, suggesting its role as a master regulator of the transition from primary tumor to metastasis. Also, this regulon was not enriched in a healthy control network, neither involved with a nonspecific proliferative signature, corroborating that it is related to neuroblastoma-specific pathways.

MAX expression was significantly altered in three independent gene expression studies. In the first, we have observed that neuroblastoma patients who did not present disease relapse in five years had significant higher levels of MAX expression than patients who either had or were already at relapse. Of important note is that only 3 of the 15 patients studied at relapse were from those included at the beginning of the study, meaning that we can treat this last group independently. This result suggests that higher MAX levels are associated to improved patient prognosis in patients with high risk, MYCN non-amplified metastatic neuroblastoma. Corroborating this data, we have analyzed a survival cohort of neuroblastoma patients and found that individuals with lower MAX expression had significantly decreased survival rates than the others. However, we have found the opposite results in a second neuroblastoma cohort. In this study, patients with higher MAX expression levels presented lower survival rates than the rest. It is possible that this incongruence is brought by the great variability in the two survival cohorts. Patients in these studies, albeit having mostly MYCN non-amplified tumors, were sorted across all risk stages and were subject to different types of therapy, thus making difficult our attempts at understanding this data. Nevertheless, we have also found positive association of MAX expression and improved prognosis in two independent cohorts of breast and lung cancer patients, corroborating its role in disease progression.

To address whether MAX expression could affect neuroblastoma by activating a differentiation pathway, we have analyzed a dataset of MYCN non-amplified neuroblastoma cells SH-SY5Y subjected to a differentiation protocol. Strikingly, we have found a peak of MAX expression between the third and fifth days of differentiation, lasting to the end of the experiment. This data makes us believe that MAX is one the late effectors SH-SY5Y differentiation, suggesting that its expression is implicated in favorable outcomes for MYCN non-amplified neuroblastoma patients by means of enhancing cell differentiation and/or impairing proliferation. These results are further corroborated by the increase of MAX protein levels we have observed in our SH-SY5Y cells differentiation protocol.

The MYC/MAX/MAD network is a key axis in the cell decision-making for differentiation and proliferation, and it is altered in several types of cancer [40–42]. Much has been said about the oncogenic roles of MYC family genes, but, other than being the obligatory heterodimer of MYC and MAD families of proteins, little is known about how variations in MAX expression can affect cell-cycle progression. In earlier cell differentiation studies, it was found that MAX expression remains constant throughout the process [43] (for a review, see [44] and references therein). One study demonstrates that

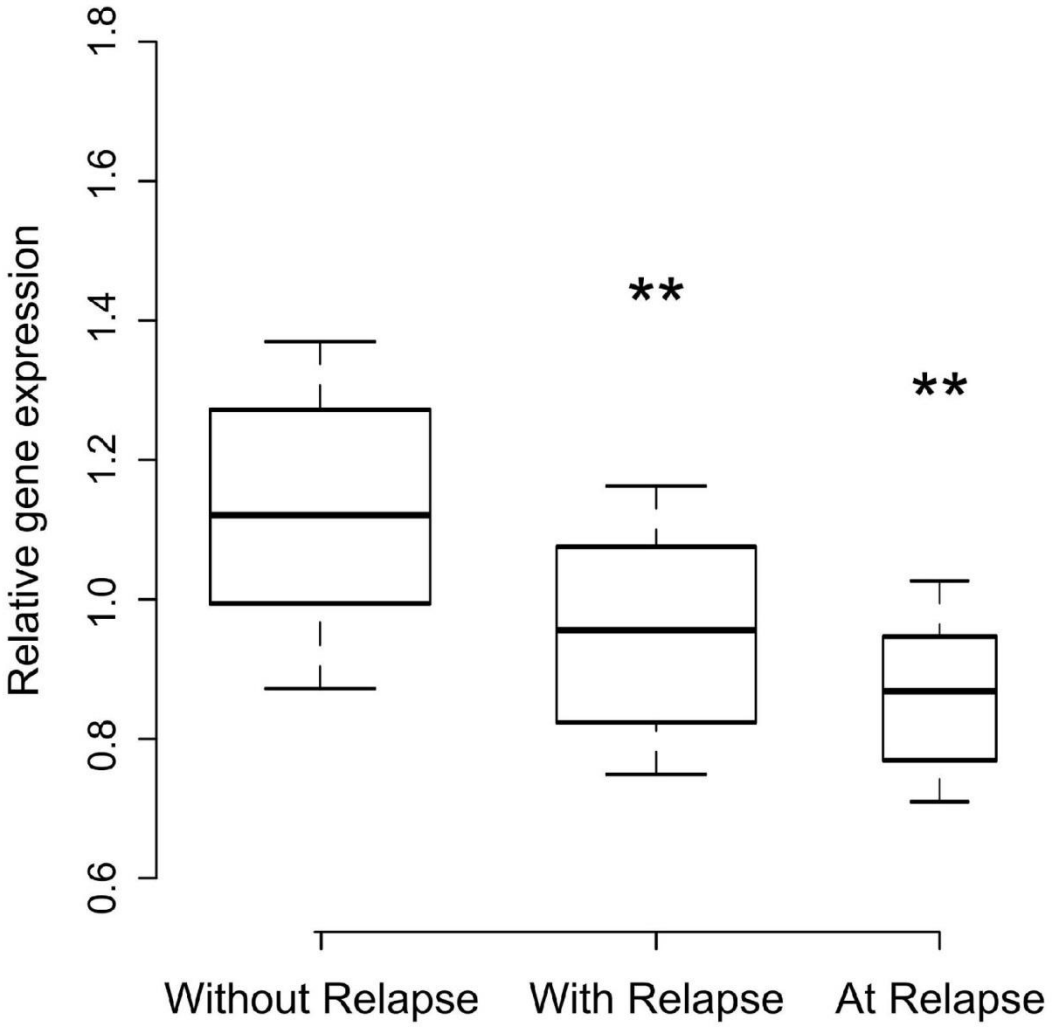

Figure 3: MAX expression in the GSE3446 patient cohort. Boxplots representations of patients who did not present relapse in five years, and patients that either had or were already at relapse, respectively. y-axis represents relative MAX expression (normalized sample value divided by MAX mean across all samples). Boxplot bars indicate lower and upper quartiles, central bar indicates mean, whiskers indicate one standard deviation of mean, and box widths are proportional to sample size. Double asterisks indicates significant differences from the first group (p<0.001, two-tailed pairwise t-tests with Bonferroni correction).

Figure 3. MAX expression in the GSE3446 patient cohort. Boxplots representations of patients who did not present relapse in five years, and patients that either had or were already at relapse, respectively. *y*-axis represents relative MAX expression (normalized sample value divided by MAX mean across all samples). Boxplot bars indicate lower and upper quartiles, central bar indicates mean, whiskers indicate one standard deviation of mean, and box widths are proportional to sample size. Double asterisks indicates significant differences from the first group ($p<0.001$, two-tailed pairwise *t*-tests with Bonferroni correction).

doi: 10.1371/journal.pone.0082457.g003

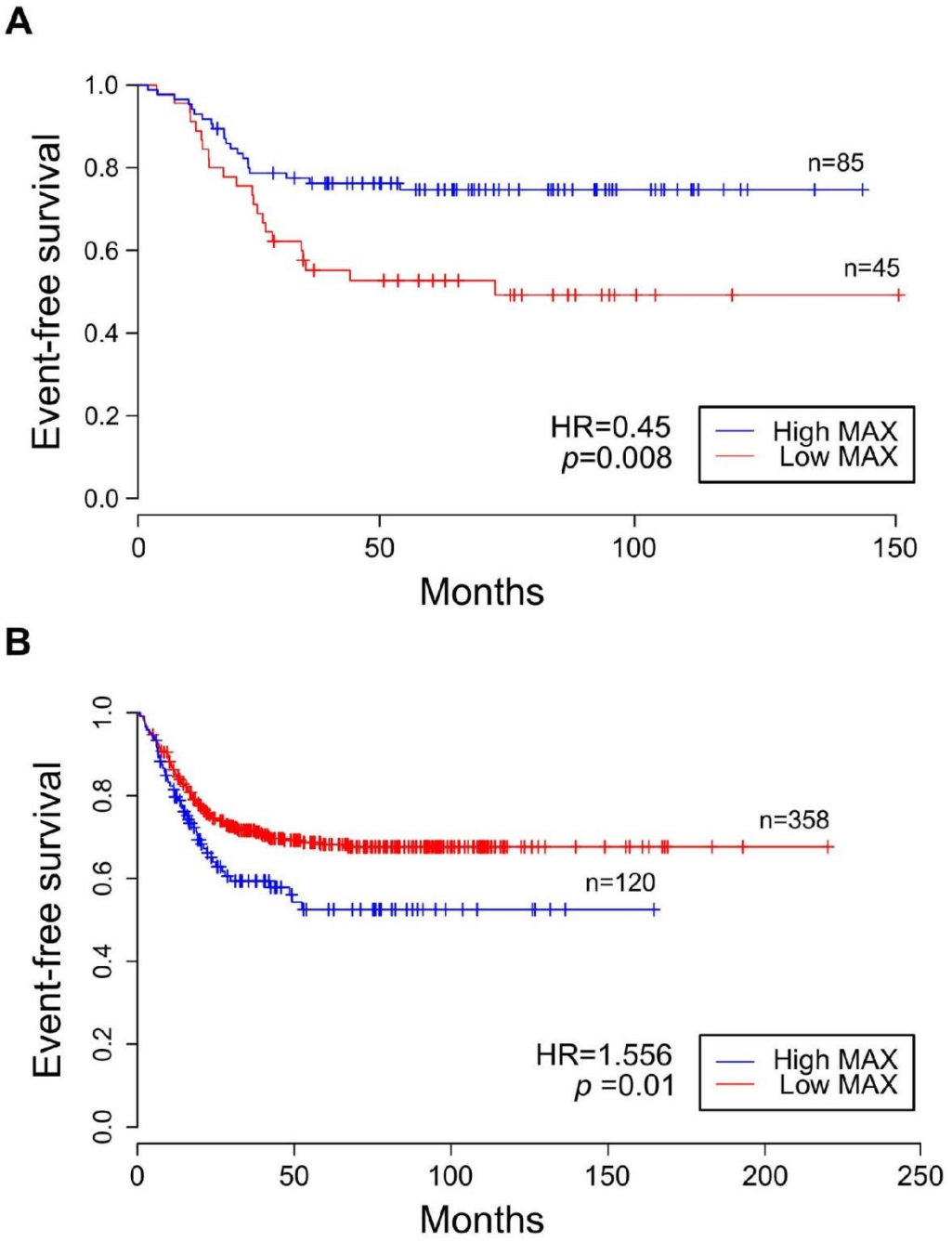

Figure 4: Kaplan-Meier plot of E-TABM-38 (A) and E-MTAB-179 (B) patient cohorts. x-axis indicates event-free survival time. y-axis represents the percentage of patients event-free survival. The red line represents patients with lower MAX expression, and the blue line, patients with higher MAX expression. Crosses mark censored data.

Figure 4. MAX expression and patient survival. Kaplan-Meier plots of E-TABM-38 (A) and E-MTAB-179 (B) patient cohorts. *x*-axis indicates event-free survival time. *y*-axis represents the percentage of patients event-free survival. The red line represents patients with lower MAX expression, and the blue line, patients with higher MAX expression. Crosses mark censored data.

doi: 10.1371/journal.pone.0082457.g004

Table 1. Survival statistics for MAX and TFEC in other type of cancers.

Gene	Cancer	Hazard Ratio	*p* value	*n*
MAX	Breast	0.75	1.E-06	2,878
	Lung	0.45	1.E-14	1,404
	Ovarian	ns	ns	1,171
TFEC	Breast	ns	ns	2,878
	Lung	0.57	1.E-09	1,404
	Ovarian	ns	ns	1,171

doi: 10.1371/journal.pone.0082457.t001

MAX $^{-/-}$ mice are subject to early embryonic lethality [45], indicating that this gene plays a vital role in development. These results suggest that this gene is indispensable to differentiation and proliferation processes, albeit only as an accessory part of its network. Lindeman and colleagues [46], however, have elegantly demonstrated that MAX has a more active role decreasing the size and frequency of tumors when overexpressed in lymphoma susceptible mice. In their study, authors have transfected mice lineages with two different MAX transgenes. They observed that both strains presented impaired lymphoproliferation and delayed tumor onset when co-expressed with a highly active MYC transgene. Interestingly, the authors noted that the MAX transgene with higher activity presented more pronounced tumor impairing capacity than the less active one, indicating that MAX has indeed tumor suppressing properties, and that the latter may be dose-dependent. These results are in accordance to related papers that link MAX overexpression with increased differentiation in other cell lines [47,48], and particularly, the work of Peverali and colleagues using neuroblastoma cells [49]. This author demonstrated that retinoic acid-treatment in SK-N-BE neuroblastoma cells overexpressing MAX induced differentiation more than twice as fast as with retinoic acid alone. Curiously, this cells are MYCN amplified, suggesting that MAX expression is sufficient to revert proliferation even in more aggressive neuroblastoma cell variants.

Because these studies were made using overexpression techniques, one may argue that they are not accurate reflections of biological processes occurring at more physiological conditions. The only currently known association with MAX and tumor biology in humans comes from studies of pheochromocytoma and paraganglioma patients. Several independent research centers have observed correlations regarding possible inactivating mutations in MAX exons and the appearance of these tumors [50–54]. These results further suggest the tumor suppressing properties of MAX. To our notice, the spontaneous increase in MAX expression during SH-SY5Y retinoic acid-induced differentiation was the first of this kind to be reported, and leads us to question whether this gene could have a more prominent role in neuroblastoma progression and neuronal cells progenitors development than previously described.

It is known that the MAX:MYC heterodimer induces cell proliferation by interacting with the TRRAP complex and histone acetyltransferases to transcriptionally activate cell cycle activators such as cyclins, and possibly repressing cell cycle inhibitors [55–57]. The MAX:MAD/MNT complex and the MAX homodimer, on the other hand, are proliferation repressors and inducers of terminal differentiation in various cell types [58–60]. The heterodimerization of MYC and MAX is preferential because the stabilization of their bHLH is thermodynamically more favorable than that of MAD and MNT with MAX [61]. Being the limiting factor of its network, one possible mechanism for how changes in MAX expression can interfere directly with proliferation/differentiation fate in cell is by regulating the amount of MYC that can activate gene expression. Lower quantities of MAX would bind preferentially with MYC and change the network balance towards proliferation. As MAX levels rise, all the available MYC binding sites would be occupied and more MAD:MAX, MNT:MAX and MAX:MAX dimers would be formed, switching the balance towards differentiation (Figure 7). In MYCN amplified tumors, a larger fraction of MAX would be used to form pro-proliferative MAX:MYCN heterodimers, explaining how these cells are more resistant to exit division [47]. In line with this, there have been studies with compounds that are able to disrupt the MYC:MAX complex in order to decrease malignancy [62,63]. Of particular relevance is the recent work of Montagne and colleagues [64], that used MAX bHLH as a protein transduction domain for decreasing MYC availability to interact with MAX, thus impairing proliferation in HeLa cells. These results demonstrate MAX capacity of directly interfering with MYC activity. Put together, this data bring MAX forward as a central player in its network, and urge us to dedicate more research in this intriguing subject.

There are questions made during this work that must be addressed in the near future. It would be interesting to compare whether MAX levels are different in stage 4s patients in order to clinically corroborate our findings with SH-SY5Y cells. There is also the putative role of ZNF101 and TFEC, which were pointed as master regulators but could not be implicated in this disease by means of their expression. We have found an indication that the latter is associated in lung cancer outcome, suggesting its importance may be greater than previously thought. The accurate roles of these genes in neuroblastoma progression are an open question and additional assays are needed to address this matter.

Conclusions

We have found evidences that MAX expression plays a role in neuroblastoma biology that has not been previously described, possibly as an additional regulator of the availability of MYC:MAX heterodimers and the balance of proliferation/differentiation. Our analyses also point that this gene may be used as a candidate predictor for positive clinical outcomes of MYCN non-amplified neuroblastoma patients. We have detected significant associations between higher MAX expression and improved survival rates for breast and lung cancer patients, suggesting that the clinical predictor potential can also be extended to other types of tumors. Lastly, we have observed that MAX expression was significantly altered during SH-SY5Y retinoic acid-induced differentiation, providing a

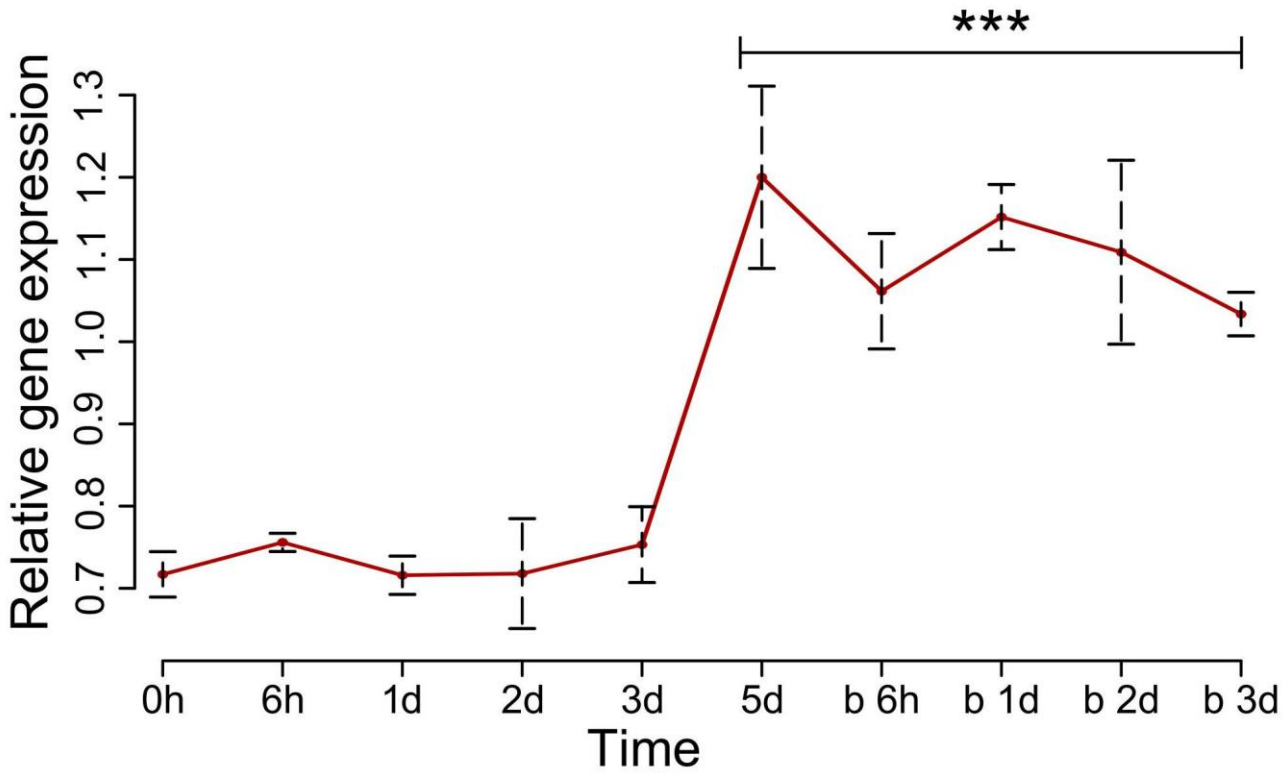

Figure 5: MAX expression in SH-SY5Y cells differentiation. x-axis indicates the time course of differentiation (times preceded by "b" indicate the addition of BDNF in the experimental protocol). y-axis indicates relative MAX expression (normalized sample value divided by MAX mean across all samples). Triple asterisks indicates statistically significant differences from the first three days (p<10-5, two-tailed pairwise t-tests with BH correction).

Figure 5. MAX expression in SH-SY5Y cells differentiation. *x*-axis indicates the time course of differentiation (times preceded by "b" indicate the addition of BDNF in the experimental protocol). *y*-axis indicates relative MAX expression (normalized sample value divided by MAX mean across all samples). Triple asterisks indicates statistically significant differences from the first three days ($p<10^{-5}$, two-tailed pairwise *t*-tests with BH correction).

doi: 10.1371/journal.pone.0082457.g005

mechanism for which this gene could be linked to tumor progression.

Materials and Methods

Neuroblastoma regulatory network reconstruction and master regulator analysis

The neuroblastoma regulatory network was reconstructed and plotted using, respectively, the RTN [65] and RedeR [66] packages for R Statistical Computing, available at Bioconductor [67]. This analysis uses the information-theoretic content (i.e. mutual information) of the gene expression profile for inferring relevant pairwise interactions among genes. Regulatory networks were reconstructed from the neuroblastoma biopsies datasets GSE16476 [68] and GSE3960 [69]. Human transcription factors were gathered from the Animal Transcription Factor Database [70]. Master regulator analyses were made using as query a genetic

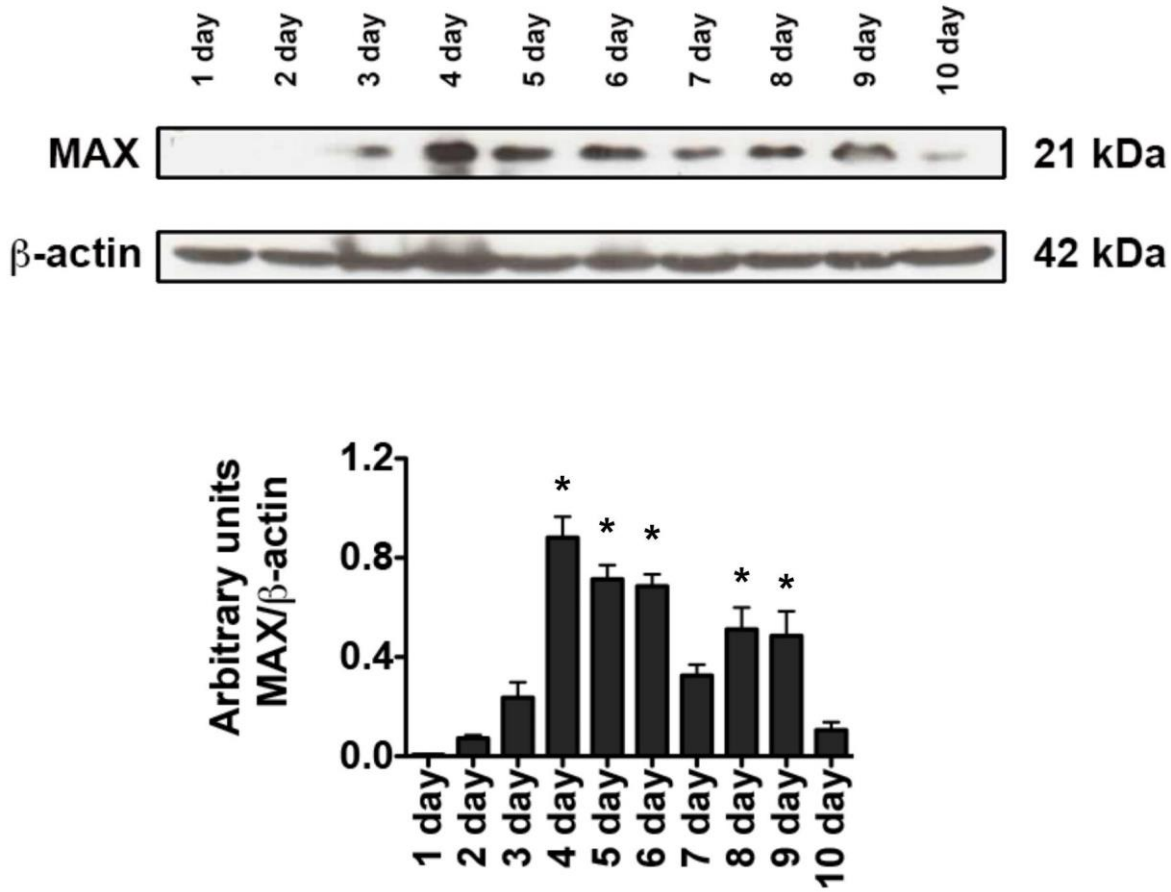

Figure 6: MAX immunocontent in differentiating SH-SY5Y neuroblastoma cells treated with retinoic acid. Asterisks indicate statistically significant ($p<0.001$) changes in the MAX/β-actin ratio in relation to the first day.

Figure 6. MAX immunocontent in differentiating SH-SY5Y neuroblastoma cells treated with retinoic acid for 10 days. Asterisks indicate statistically significant ($p<10^{-3}$) changes in the MAX/β-actin ratio in relation to the first day.
doi: 10.1371/journal.pone.0082457.g006

signature obtained from aggressive metastases. For assessing this metastatic signature, we have used Limma R package [71] on the GSE25623 dataset [72] in order to discover differentially expressed genes between neuroblastoma primary tumors and bone marrow metastases (for a schematic description of this workflow, we once again refer the reader to Figure 1). Master regulators DNA-binding motifs were searched using the FIMO tool for transcription factor binding site prediction [73]. We have considered target genes only those that had at least one motif occurrence in their flanking regions (either 2.5 Kb up and downstream from the gene start codon [74]). To assess whether regulons were significantly enriched with genes regulated by its transcription factor, we verified if our prediction was significantly higher than randomly sampling all human genes 100,000 times using regulon-sized samples. Statistical significance was calculated using one-tailed *z*-tests.

SH-SY5Y differentiation and Western blot analysis

Human neuroblastoma SH-SY5Y cells were obtained from the American Type Culture Collection (ATCC, Rockville, MD). Cells were cultivated using Dulbecco's Modified Eagle's Medium supplemented with 10% heat-inactivated fetal bovine serum, L-glutamine (2mM), and 0.28 mg/ml of gentamycin sulfate in a 5% CO_2 humidified incubator at 37°C. Differentiation protocol consisted in reducing fetal bovine serum concentration to 1% and adding retinoic acid (10μM) during 10 days. The culture medium was replaced every three days. Cells were grown in 6-well cluster dishes.

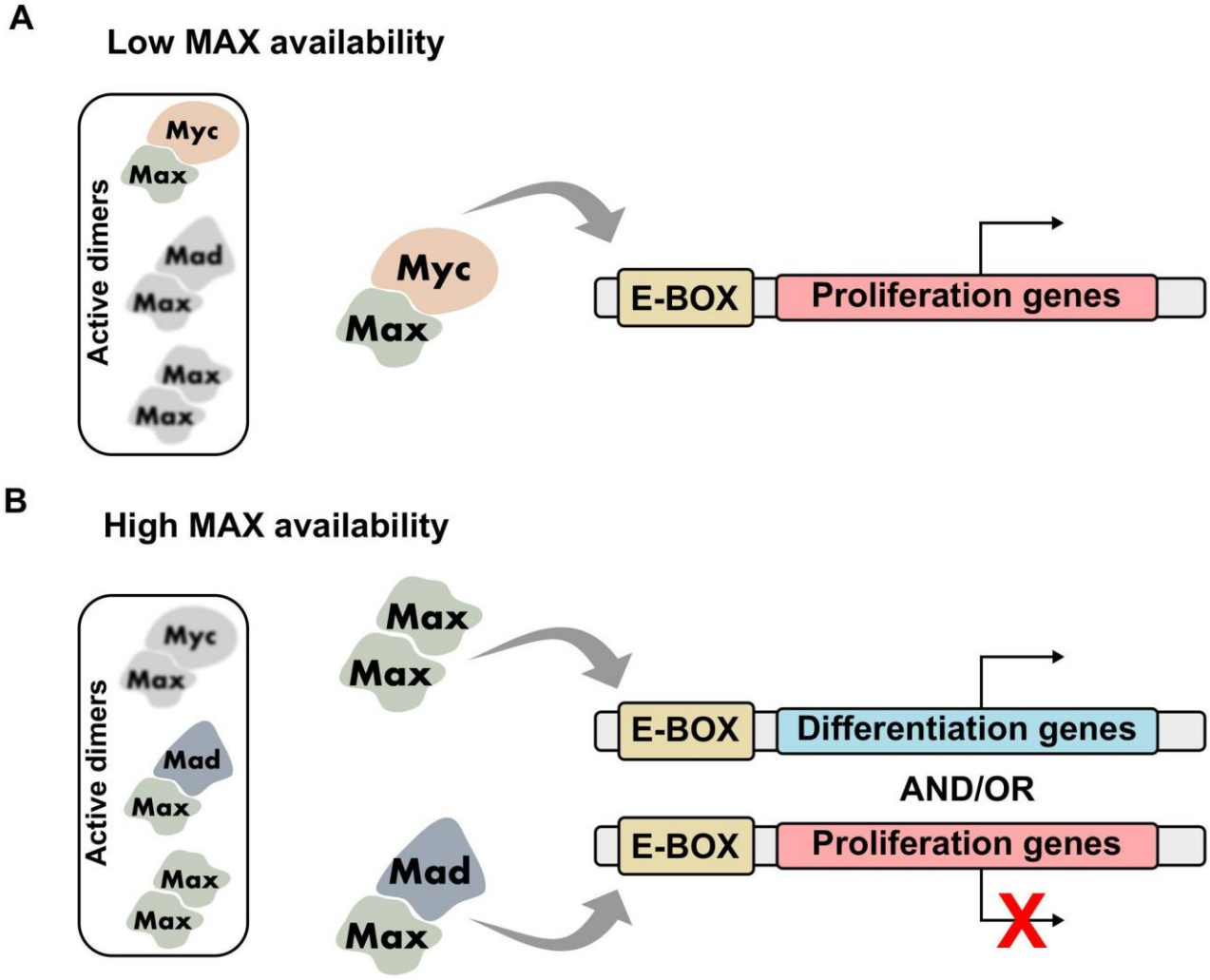

Figure 7: Schematic representation of MAX expression in neuroblastoma cell behavior. Cells with lower MAX concentration preferentially form pro-proliferative MYC:MAX heterodimers, leading to increased proliferation and de-differentiation (A). With higher MAX levels, there is no more free MYC available and more MAD:MAX heterodimers and MAX homodimers are formed, shifting balance towards differentiation (B).

Figure 7. Schematic representation of MAX expression in neuroblastoma cell behavior. Cells with lower MAX concentration preferentially form pro-proliferative MYC:MAX heterodimers, leading to increased proliferation and de-differentiation (A). With higher MAX levels, there is no more free MYC available and more MAD:MAX heterodimers and MAX homodimers are formed, shifting balance towards differentiation (B).

doi: 10.1371/journal.pone.0082457.g007

Proteins (20µg) were separated using SDS-PAGE – 10% (w/v) acrylamide, 0.275% (w/v) bisacrylamide gels – and electrotransferred onto nitrocellulose membranes. Membranes were then incubated in Tris-buffered saline Tween-20 [TBS-T; 20 mM Tris–HCl, pH 7.5, 137 mM NaCl, 0.05% (v/v) Tween 20] for 1h at room temperature. Subsequently, membranes were incubated for 12h with polyclonal rabbit anti-MAX (1:1.000 dilution; Cell Signaling). After washing in TBST, blots were incubated with rabbit peroxidase-linked anti-immunoglobulin G (IgG) antibodies (1:10.000 dilution) for 1.5h at room temperature. Chemiluminescent bands were detected, and densitometric analysis was performed by Image-J® software. All analyses were performed in triplicate.

Statistical analyses

Unless stated otherwise, all statistical analyses were made using two-tailed *t*-tests. Pairwise *t*-tests were made with Bonferroni and BH *p*-value corrections when more than two groups were analyzed. The latter was used when testing for more than four groups. Survival curves for gene expression were drawn using Kaplan-Meier plot and tested with Cox proportional hazards model using the Survival [75] and Survcomp [76] packages for R statistical computing. The best cutoff for samples split was chosen by plotting all percentiles of gene expression between the upper and lower quartiles and selecting the best performing threshold. For breast, lung and ovarian cancers, statistics were made using the Kaplan-Meier Plotter web tool. Parameters used in all analyses were auto select split cutoff and only the JetSet best probes were selected [77]. All gene expression profiles used in this paper were obtained from Gene Expression Omnibus and ArrayExpress public databases.

Supporting Information

Figure S1. Kaplan-Meier plot of the 2012 breast cancer patient cohort available at the Kaplan-Meier Plotter web tool. *x*-axis indicates event-free survival time. *y*-axis represents the percentage of patients event-free survival. The black line represents patients with lower MAX expression, and the red line, patients with higher MAX expression. Crosses mark censored data.
(TIFF)

Figure S2. Kaplan-Meier plot of the unified lung cancer patient cohort available at the Kaplan-Meier Plotter web tool. *x*-axis indicates event-free survival time. *y*-axis represents the percentage of patients event-free survival. The black line represents patients with lower MAX expression, and the red line, patients with higher MAX expression. Crosses mark censored data.
(TIFF)

Figure S3. Kaplan-Meier plot of the 2013 ovarian cancer patient cohort available at the Kaplan-Meier Plotter web tool. *x*-axis indicates event-free survival time. *y*-axis represents the percentage of patients event-free survival. The black line represents patients with lower MAX expression, and the red line, patients with higher MAX expression. Crosses mark censored data.
(TIFF)

Figure S4. Kaplan-Meier plot of the 2012 breast cancer patient cohort available at the Kaplan-Meier Plotter web tool. *x*-axis indicates event-free survival time. *y*-axis represents the percentage of patients event-free survival. The black line represents patients with lower TFEC expression, and the red line, patients with higher TFEC expression. Crosses mark censored data.
(TIFF)

Figure S5. Kaplan-Meier plot of the lung cancer patient cohort available at the Kaplan-Meier Plotter web tool. *x*-axis indicates event-free survival time. *y*-axis represents the percentage of patients event-free survival. The black line represents patients with lower TFEC expression, and the red line, patients with higher TFEC expression. Crosses mark censored data.
(TIFF)

Figure S6. Kaplan-Meier plot of the 2013 ovarian cancer patient cohort available at the Kaplan-Meier Plotter web tool. *x*-axis indicates event-free survival time. *y*-axis represents the percentage of patients event-free survival. The black line represents patients with lower TFEC expression, and the red line, patients with higher TFEC expression. Crosses mark censored data.
(TIFF)

Table S1. Overview of all Master Regulators found in our study and detailed composition of MAX, TFEC and ZNF101 regulons.
(XLSX)

Author Contributions

Conceived and designed the experiments: RD'OA RJSD MAAC JCFM. Performed the experiments: RD'OA MABP VMR. Analyzed the data: RD'OA RJSD MAAC JCFM. Contributed reagents/materials/analysis tools: MAAC DPG JCFM. Wrote the manuscript: RD'OA. Designed the software used in the analysis: MAAC.

References

1. Maris JM, Hogarty MD, Bagatell R, Cohn SL (2007) Neuroblastoma. Lancet 369: 2106–2120. doi:10.1016/S0140-6736(07)60983-0. PubMed: 17586306.
2. Howman-Giles R, Shaw PJ, Uren RF, Chung DKV (2007) Neuroblastoma and other neuroendocrine tumors. Semin Nucl Med 37: 286–302. doi:10.1053/j.semnuclmed.2007.02.009. PubMed: 17544628.
3. Spix C, Pastore G, Sankila R, Stiller CA, Steliarova-Foucher E (2006) Neuroblastoma incidence and survival in European children (1978-1997): report from the Automated Childhood Cancer Information System project. Eur J Cancer 42: 2081–2091. doi:10.1016/j.ejca.2006.05.008. PubMed: 16919772.
4. Parise IZS, Haddad BR, Cavalli LR, Pianovski MAD, Maggio EM et al. (2006) Neuroblastoma in southern Brazil: an 11-year study. J Pediatr Hematol Oncol 28: 82–87. doi:10.1097/01.mph.0000199601.35010.52. PubMed: 16462579.
5. Mills SE (2002) Neuroectodermal neoplasms of the head and neck with emphasis on neuroendocrine carcinomas. Mod Pathol 15: 264–278. doi:10.1038/modpathol.3880522. PubMed: 11904342.
6. Modak S, Cheung N-KV (2010) Neuroblastoma: Therapeutic strategies for a clinical enigma. Cancer Treat Rev 36: 307–317. doi:10.1016/j.ctrv.2010.02.006. PubMed: 20227189.
7. Monclair T, Brodeur GM, Ambros PF, Brisse HJ, Cecchetto G et al. (2009) The International Neuroblastoma Risk Group (INRG) staging system: an INRG Task Force report. J Clin Oncol 27: 298–303. doi:10.1200/JCO.2008.16.6876. PubMed: 19047290.
8. Maris JM, Matthay KK (1999) Molecular biology of neuroblastoma. J Clin Oncol 17: 2264–2279. PubMed: 10561284.

9. Schwab M, Westermann F, Hero B, Berthold F (2003) Neuroblastoma: biology and molecular and chromosomal pathology. Lancet Oncol 4: 472–480. doi:10.1016/S1470-2045(03)01166-5. PubMed: 12901961.
10. Tonini GP, Boni L, Pession A, Rogers D, Iolascon A et al. (1997) MYCN oncogene amplification in neuroblastoma is associated with worse prognosis, except in stage 4s: the Italian experience with 295 children. J Clin Oncol 15: 85–93.
11. Westermann F, Schwab M (2002) Genetic parameters of neuroblastomas. Cancer Lett 184: 127–147. doi:10.1016/S0304-3835(02)00199-4. PubMed: 12127685.
12. Bagatell R, Beck-Popovic M, London WB, Zhang Y, Pearson ADJ et al. (2009) Significance of MYCN amplification in international neuroblastoma staging system stage 1 and 2 neuroblastoma: a report from the International Neuroblastoma Risk Group database. J Clin Oncol 27: 365–370. PubMed: 19047282.
13. Kushner BH, Cheung NK, LaQuaglia MP, Ambros PF, Ambros IM et al. (1996) International neuroblastoma staging system stage 1 neuroblastoma: a prospective study and literature review. J Clin Oncol 14: 2174–2180. PubMed: 8683252.
14. Bhatnagar SN, Sarin YK (2012) Neuroblastoma: a review of management and outcome. Indian J Pediatr 79: 787–792. doi:10.1007/s12098-012-0748-2. PubMed: 22528697.
15. Schmidt ML, Lukens JN, Seeger RC, Brodeur GM, Shimada H et al. (2000) Biologic factors determine prognosis in infants with stage IV neuroblastoma: A prospective Children's Cancer Group study. J Clin Oncol 18: 1260–1268. PubMed: 10715296.
16. Schmidt ML, Lal A, Seeger RC, Maris JM, Shimada H et al. (2005) Favorable prognosis for patients 12 to 18 months of age with stage 4 nonamplified MYCN neuroblastoma: a Children's Cancer Group Study. J Clin Oncol 23: 6474–6480. doi:10.1200/JCO.2005.05.183. PubMed: 16116154.
17. Asgharzadeh S, Pique-Regi R, Sposto R, Wang H, Yang Y et al. (2006) Prognostic significance of gene expression profiles of metastatic neuroblastomas lacking MYCN gene amplification. J Natl Cancer Inst 98: 1193–1203. doi:10.1093/jnci/djj330. PubMed: 16954472.
18. Babu MM, Lang B, Aravind L (2009) Methods to reconstruct and compare transcriptional regulatory networks. Methods Mol Biol 541: 163–180. doi:10.1007/978-1-59745-243-4_8. PubMed: 19381525.
19. Margolin A, Wang K, Lim WK, Kustagi M, Nemenman I et al. (2006) Reverse engineering cellular networks. Nat Protoc 1: 662–671.
20. Mehra S, Hu W-S, Karypis G (2004) A Boolean algorithm for reconstructing the structure of regulatory networks. Metab Eng 6: 326–339. doi:10.1016/j.ymben.2004.05.002. PubMed: 15491862.
21. Segal E, Shapira M, Regev A, Pe'er D, Botstein D et al. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet 34: 166–176. doi:10.1038/ng1165. PubMed: 12740579.
22. Basso K, Margolin A, Stolovitzky G, Klein U, Dalla-Favera R et al. (2005) Reverse engineering of regulatory networks in human B cells. Nat Genet 37: 382–390.
23. Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X et al. (2010) The transcriptional network for mesenchymal transformation of brain tumours. Nature 463: 318–325. doi:10.1038/nature08712. PubMed: 20032975.
24. Chen D-T, Nasir A, Culhane A, Venkataramu C, Fulp W et al. (2010) Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. Breast Cancer Res Treat 119: 335–346. doi:10.1007/s10549-009-0344-y. PubMed: 19266279.
25. Venet D, Dumont JE, Detours V (2011) Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput Biol 7: e1002240. PubMed: 22028643.
26. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32: D91–D94. doi:10.1093/nar/gnh092. PubMed: 14681366.
27. Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X et al. (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res 38: D105–D110. doi:10.1093/nar/gkp950. PubMed: 19906716.
28. Ledent V, Paquet O, Vervoort M (2002) Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol 3: RESEARCH0030. PubMed: 12093377.
29. Fernandez PC, Frank SR, Wang L, Schroeder M, Liu S et al. (2003) Genomic targets of the human c-Myc protein. Genes Dev 17: 1115–1129. doi:10.1101/gad.1067003. PubMed: 12695333.
30. Li Z, Van Calcar S, Qu C, Cavenee WK, Zhang MQ et al. (2003) A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc Natl Acad Sci U S A 100: 8164–8169. doi:10.1073/pnas.1332764100. PubMed: 12808131.
31. Yap C-S, Peterson AL, Castellani G, Sedivy JM, Neretti N (2011) Kinetic profiling of the c-Myc transcriptome and bioinformatic analysis of repressed gene promoters. Cell Cycle 10: 2184–2196. doi:10.4161/cc.10.13.16249. PubMed: 21623162.
32. Oberthuer A, Berthold F, Warnat P, Hero B, Kahlert Y et al. (2006) Customized oligonucleotide microarray gene expression-based classification of neuroblastoma patients outperforms current clinical risk stratification. J Clin Oncol 24: 5070–5078. doi:10.1200/JCO.2006.06.1879. PubMed: 17075126.
33. Oberthuer A, Juraeva D, Li L, Kahlert Y, Westermann F et al. (2010) Comparison of performance of one-color and two-color gene-expression analyses in predicting clinical endpoints of neuroblastoma patients. Pharmacogenomics J 10: 258–266. doi:10.1038/tpj.2010.53. PubMed: 20676065.
34. Gyorffy B, Lánczky A, Szállási Z (2012) Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. Endocr Relat Cancer 19: 197–208. doi:10.1530/ERC-11-0329. PubMed: 22277193.
35. Györffy B, Lanczky A, Eklund AC, Denkert C, Budczies J et al. (2010) An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat 123: 725–731. doi:10.1007/s10549-009-0674-9. PubMed: 20020197.
36. Nishida Y, Adati N, Ozawa R, Maeda A, Sakaki Y et al. (2008) Identification and classification of genes regulated by phosphatidylinositol 3-kinase- and TRKB-mediated signalling pathways during neuronal differentiation in two subtypes of the human neuroblastoma cell line SH-SY5Y. BMC Res Notes 1: 95. doi:10.1186/1756-0500-1-95. PubMed: 18957096.
37. D'Angio G, Evans A, Koop CE (1971) special pattern of widespread neuroblastoma with a favourable prognosis. Lancet 297: 1046–1049. doi:10.1016/S0140-6736(71)91606-0.
38. Bénard J, Raguénez G, Kauffmann A, Valent A, Ripoche H et al. (2008) MYCN-non-amplified metastatic neuroblastoma with good prognosis and spontaneous regression: a molecular portrait of stage 4S. Mol Oncol 2: 261–271. doi:10.1016/j.molonc.2008.07.002. PubMed: 19383347.
39. Garaventa A, Parodi S, De Bernardi B, Dau D, Manzitti C et al. (2009) Outcome of children with neuroblastoma after progression or relapse. A retrospective study of the Italian neuroblastoma registry. Eur J Cancer 45: 2835–2842. doi:10.1016/j.ejca.2009.06.010. PubMed: 19616426.
40. O'Hagan RC, Schreiber-Agus N, Chen K, David G, Engelman JA et al. (2000) Gene-target recognition among members of the myc superfamily and implications for oncogenesis. Nat Genet 24: 113–119. doi:10.1038/72761. PubMed: 10655054.
41. Hurlin PJ, Dezfouli S (2004) Functions of myc:max in the control of cell proliferation and tumorigenesis. Int Rev Cytol 238: 183–226. doi:10.1016/S0074-7696(04)38004-6. PubMed: 15364199.
42. Dang CV (2012) MYC on the path to cancer. Cell 149: 22–35. doi:10.1016/j.cell.2012.03.003. PubMed: 22464321.
43. Pulverer B, Sommer A, McArthur GA, Eisenman RN, Lüscher B (2000) Analysis of Myc/Max/Mad network members in adipogenesis: inhibition of the proliferative burst and differentiation by ectopically expressed Mad1. J Cell Physiol 183: 399–410. doi:10.1002/(SICI)1097-4652(200006)183:3. PubMed: 10797315.
44. Rottmann S, Lüscher B (2006) The Mad side of the Max network: antagonizing the function of Myc and more. Curr Top Microbiol Immunol 302: 63–122. doi:10.1007/3-540-32952-8_4. PubMed: 16620026.
45. Shen-Li H, O'Hagan RC, Hou H, Horner JW, Lee HW et al. (2000) Essential role for Max in early embryonic growth and development. Genes Dev 14: 17–22. PubMed: 10640271.
46. Lindeman GJ, Harris AW, Bath ML, Eisenman RN, Adams JM (1995) Overexpressed max is not oncogenic and attenuates myc-induced lymphoproliferation and lymphomagenesis in transgenic mice. Oncogene 10: 1013–1017. PubMed: 7898919.
47. Gu W, Cechova K, Tassi V, Dalla-Favera R (1993) Opposite regulation of gene transcription and cell proliferation by c-Myc and Max. Proc Natl Acad Sci U S A 90: 2935–2939. doi:10.1073/pnas.90.7.2935. PubMed: 8385351.
48. Cañelles M, Delgado MD, Hyland KM, Lerga A, Richard C et al. (1997) Max and inhibitory c-Myc mutants induce erythroid differentiation and resistance to apoptosis in human myeloid leukemia cells. Oncogene 14: 1315–1327. doi:10.1038/sj.onc.1200948. PubMed: 9178892.
49. Peverali FA, Orioli D, Tonon L, Ciana P, Bunone G et al. (1996) Retinoic acid-induced growth arrest and differentiation of neuroblastoma cells are counteracted by N-myc and enhanced by max overexpressions. Oncogene 12: 457–462. PubMed: 8570225.

50. Burnichon N, Cascón A, Schiavi F, Morales NP, Comino-Méndez I et al. (2012) MAX mutations cause hereditary and sporadic pheochromocytoma and paraganglioma. Clin Cancer Res 18: 2828–2837. doi:10.1158/1078-0432.CCR-12-0160. PubMed: 22452945.
51. Comino-Méndez I, Gracia-Aznárez FJ, Schiavi F, Landa I, Leandro-García LJ et al. (2011) Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. Nat Genet 43: 663–667. doi: 10.1038/ng.861. PubMed: 21685915.
52. Welander J, Söderkvist P, Gimm O (2011) Genetics and clinical characteristics of hereditary pheochromocytomas and paragangliomas. Endocr Relat Cancer 18: R253–R276. doi:10.1530/ERC-11-0170. PubMed: 22041710.
53. Kirmani S, Young WF (2008) Hereditary Paraganglioma-Pheochromocytoma Syndromes. University of Washington, Seattle.
54. Galan SR, Kann PH (2013) Genetics and molecular pathogenesis of pheochromocytoma and paraganglioma. Clin Endocrinol (Oxf) 78: 165–175. doi:10.1111/cen.12071. PubMed: 23061808.
55. McMahon SB, Wood MA, Cole MD (2000) The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc. Mol Cell Biol 20: 556–562. doi:10.1128/MCB.20.2.556-562.2000. PubMed: 10611234.
56. Park J (2001) The ATM-related domain of TRRAP is required for histone acetyltransferase recruitment and Myc-dependent oncogenesis. Genes Dev 15: 1619–1624. doi:10.1101/gad.900101. PubMed: 11445536.
57. Bouchard C, Dittrich O, Kiermaier A, Dohmann K, Menkel A et al. (2001) Regulation of cyclin D2 gene expression by the Myc/Max/Mad network: Myc-dependent TRRAP recruitment and histone acetylation at the cyclin D2 promoter. Genes Dev 15: 2042–2047. doi:10.1101/gad.907901. PubMed: 11511535.
58. Ayer DE, Lawrence QA, Eisenman RN (1995) Mad-max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. Cell 80: 767–776. doi: 10.1016/0092-8674(95)90355-0. PubMed: 7889570.
59. Hurlin PJ, Quéva C, Eisenman RN (1997) Mnt, a novel Max-interacting protein is coexpressed with Myc in proliferating cells and mediates repression at Myc binding sites. Genes Dev 11: 44–58. doi:10.1101/gad.11.1.44. PubMed: 9000049.
60. Ayer DE, Eisenman RN (1993) A switch from Myc:Max to Mad:Max heterocomplexes accompanies monocyte/macrophage differentiation. Genes Dev 7: 2110–2119. doi:10.1101/gad.7.11.2110. PubMed: 8224841.
61. Banerjee A, Hu J, Goss DJ (2006) Thermodynamics of protein-protein interactions of cMyc, Max, and Mad: effect of polyions on protein dimerization. Biochemistry 45: 2333–2338. doi:10.1021/bi0522551. PubMed: 16475822.
62. Wang H, Hammoudeh DI, Follis AV, Reese BE, Lazo JS et al. (2007) Improved low molecular weight Myc-Max inhibitors. Mol Cancer Ther 6: 2399–2408. doi:10.1158/1535-7163.MCT-07-0005. PubMed: 17876039.
63. Follis AV, Hammoudeh DI, Daab AT, Metallo SJ (2009) Small-molecule perturbation of competing interactions between c-Myc and Max. Bioorg Med Chem Lett 19: 807–810. doi:10.1016/j.bmcl.2008.12.025. PubMed: 19114306.
64. Montagne M, Beaudoin N, Fortin D, Lavoie CL, Klinck R et al. (2012) The Max b-HLH-LZ can transduce into cells and inhibit c-Myc transcriptional activities. PLOS ONE 7: e32172. doi:10.1371/journal.pone.0032172. PubMed: 22384171.
65. Fletcher MNC, Castro MAA, Wang X, de Santiago I, O'Reilly M et al. (2013) Master regulators of FGFR2 signalling and breast cancer risk. Nat Commun 4: 2464. PubMed: 24043118.
66. Castro MA, Wang X, Fletcher MN, Meyer KB, Markowetz F (2012) RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. Genome Biol 13: R29. doi:10.1186/gb-2012-13-4-r29. PubMed: 22531049.
67. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80. doi:10.1186/gb-2004-5-10-r80. PubMed: 15461798.
68. Molenaar JJ, Koster J, Zwijnenburg Da, van Sluis P, Valentijn LJ et al. (2012) Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature 483: 589–593. doi:10.1038/nature10910. PubMed: 22367537.
69. Wang Q, Diskin S, Rappaport E, Attiyeh E, Mosse Y et al. (2006) Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number. Cancer Res 66: 6050–6062. doi: 10.1158/0008-5472.CAN-05-4618. PubMed: 16778177.
70. Zhang HM, Chen H, Liu W, Liu H, Gong J et al. (2012) AnimalTFDB: a comprehensive animal transcription factor database. Nucleic Acids Res 40: D144–D149. doi:10.1093/nar/gkr965. PubMed: 22080564.
71. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article3
72. Morandi F, Scaruffi P, Gallo F, Stigliani S, Moretti S et al. (2012) Bone marrow-infiltrating human neuroblastoma cells express high levels of calprotectin and HLA-G proteins. PLOS ONE 7: e29922. doi:10.1371/journal.pone.0029922. PubMed: 22253825.
73. Grant CE, Bailey TL, Noble WS (2011) FIMO: scanning for occurrences of a given motif. Bioinformatics 27: 1017–1018. doi:10.1093/bioinformatics/btr064. PubMed: 21330290.
74. Kristiansson E, Thorsen M, Tamás MJ, Nerman O (2009) Evolutionary forces act on promoter length: identification of enriched cis-regulatory elements. Mol Biol Evol 26: 1299–1307. doi:10.1093/molbev/msp040. PubMed: 19258451.
75. Therneau T, Grambsch P (2000) Modeling survival data: extending the Cox model. Springer.
76. Haibe-Kains B, Desmedt C, Sotiriou C, Bontempi G (2008) A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? Bioinformatics 24: 2200–2208. doi:10.1093/bioinformatics/btn374. PubMed: 18635567.
77. Li Q, Birkbak NJ, Gyorffy B, Szallasi Z, Eklund AC (2011) Jetset: selecting the optimal microarray probe set to represent a gene. BMC Bioinformatics 12: 474. doi:10.1186/1471-2105-12-474. PubMed: 22172014.

PARTE III

4. DISCUSSÃO

Neuroblastomas podem ser considerados o exemplo clássico de uma doença multifatorial complexa (Tonini *et al.*, 2003). Cada caso possui suas próprias particularidades moleculares e clínicas, de forma que um mesmo fenótipo pode ser causado por inúmeras alterações genéticas. Entretanto, é sabido que todos os tumores necessitam realizar alguns passos-chaves para se estabelecer com sucesso no corpo e evitar o complexo conjunto de mecanismos de defesa que possuímos; e os neuroblastomas não são uma exceção. Partindo dessa premissa, acreditamos que a busca de marcadores efetivos para a progressão e tratamento dessa doença pode ser auxiliada com as técnicas robustas de análise de dados oferecidas pela biologia de sistemas.

Através da reconstrução da rede regulatória de neuroblastoma e de uma bateria rigorosa de validações, conseguimos verificar a existência de pelo menos uma unidade regulatória que estava enriquecida com os genes diferencialmente expressos entre tumores primários e metástases de medula óssea – consideradas uma das formas mais agressivas e terminais de neuroblastoma. Nossa análise demonstrou que o fator de transcrição MAX estava alterado nas metástases, sugerindo sua participação na progressão tumoral. Similarmente, encontramos indicativos de que os fatores TFEC e ZNF101 também estavam enriquecidos nesse processo. A fim de determinar de fato qual a importância desses fatores no desenvolvimento tumoral, analisamos todos os conjuntos de dados independentes de pacientes disponíveis em bancos de dados públicos, de forma a buscar correlações entre a expressão desses genes e o fenótipo tumoral. Infelizmente, a maioria das plataformas analisadas não possuíam sondas para os genes *TFEC* e *ZNF101*, impedindo que estes fossem analisados rigorosamente no contexto da progressão de neuroblastoma. Sendo assim, nossa discussão se focará

apenas nos aspectos analisados do gene *MAX*, cujo provável mecanismo de atuação em neuroblastoma será postulado no decorrer deste trabalho.

O primeiro conjunto de dados analisado corresponde a uma coorte de pacientes, cujas biópsias foram classificadas *a posteriori* referentes à recidiva ou não em um período de cinco anos pós-diagnóstico. Interessantemente, observamos que a expressão de *MAX* estava aumentada nos sujeitos que não tiveram recidiva do tumor no período, sugerindo um efeito protetor desse gene na doença. Complementarmente, esse conjunto de dados também possuía uma terceira coorte de pacientes, cujas biopsias foram retiradas de pacientes que retornavam ao hospital (*i.e.* da própria recidiva). Apenas três dos 15 pacientes desse grupo faziam intersecção com a coorte dos pacientes com as biópsias analisadas no diagnóstico, de forma que podem ser considerados um terceiro grupo independente. Quando analisamos a expressão dessa coorte, também observamos que a expressão de *MAX* estava diminuída nesses pacientes em relação aos que se curaram, reforçando a ideia que este gene pode estar envolvido na progressão de neuroblastoma.

A etapa seguinte deste estudo consistiu em verificar se a expressão de *MAX* era suficiente para discriminar curvas de sobrevivência de pacientes de neuroblastoma. Para esse fim, analisamos duas coortes independentes de pacientes de neuroblastoma que possuíam informações de sobrevivência. Em uma dessas coortes, observamos uma separação de curvas de sobrevivência reminiscentes aos resultados anteriores – os pacientes com maior expressão de *MAX* apresentavam um quadro de evolução da doença melhor que os outros. Esse resultado, entretanto, não foi corroborado pela segunda coorte; os pacientes com alta expressão desse gene apresentaram desfechos piores que os demais. Esses resultados sugerem que, embora a expressão de *MAX* tenha

influência no desfecho de pacientes, ela não pode ser utilizada por si só como marcador de prognóstico.

A literatura científica aponta o neuroblastoma como um dos únicos tumores que pode sofrer diferenciação espontânea. São descritos casos de cura de pacientes que apresentavam avançadas doenças metastáticas e cujos tumores se diferenciaram em ganglioneuromas benignos (Bhatnagar e Sarin, 2012; D'Angio, Evans e Koop, 1971). Entretanto, o mecanismo desse processo permanece desconhecido. Neste trabalho, postulamos que o papel de MAX na progressão de neuroblastoma poderia envolver uma via de controle de proliferação/diferenciação. A fim de determinarmos se esta hipótese era verdadeira, analisamos um experimento de diferenciação de células de neuroblastoma humano SH-SY5Y sofrendo diferenciação por tratamento com ácido retinoico[3]. Surpreendentemente, observamos que a expressão de *MAX* estava aumentada a partir do 5º dia de tratamento. Esse resultado foi posteriormente corroborado pela imunoquantificação de MAX, que apontou um aumento do conteúdo desta proteína a partir do 4º dia de tratamento com ácido retinoico.

4.1 MAX e a rede MYC/MAX/MAD

O gene *MAX* foi descoberto na era pré-genômica como um dos fatores que interagiam com MYC (Blackwood e Eisenman, 1991). Os primeiros estudos com esse gene desvendaram as diretrizes básicas de seu funcionamento – quando este formava dímeros MYC:MAX, era capaz de induzir a célula a um estado proliferativo através da ativação de genes ligados ao ciclo celular. Poucos anos depois, foi descoberto um novo

[3] Algumas linhagens de células de neuroblastoma, quando tratadas com esse composto, ativam uma via de diferenciação diminui significativamente sua taxa de proliferação e induz mudanças morfológicas e bioquímicas relacionadas a maturação neuronal (*e.g.* prolongamentos dendríticos, expressão de marcadores neuronais e produção de neurotransmissores).

parceiro funcional desse gene – o fator MAD[4] (Ayer, Kretzner e Eisenman, 1993). Começava assim o entendimento de que MYC, que há décadas estava associado com o surgimento de tumores, não era um fator solitário. Muito pelo contrário, ele integra uma complexa rede que é responsável pela tomada de decisões acerca do destino celular. No centro dessa rede, está a proteína MAX, que, dependendo de seu parceiro funcional, poderia guiar a célula para um estado proliferativo ou induzir à diferenciação (Ayer e Eisenman, 1993).

Devido ao papel central de MYC em quase todos os tipos de tumores (Dang, 2012), este é de longe o mais estudado dos elementos da rede MAX. Basta uma busca na base de dados *Pubmed – Gene* de *Homo sapiens* (acessado em 03/01/2014, às 13h55) para observarmos que existem 1.288 entradas para MYC (incluindo sua isoforma MYCN[5]), e apenas 120 entradas para MAX e 145 para MAD (isoformas MXD1, MXI1, MXD3 e MXD4). Esse resultado demonstra quão desbalanceado é nosso conhecimento sobre essa importante rede regulatória. Entretanto, já existe corpo de evidência para delinearmos de uma maneira geral tanto os papéis dessas proteínas no funcionamento sadio do organismo quanto seu envolvimento na progressão de algumas patologias.

Tanto MAX quanto MYC e MAD são fatores de transcrição da família MITF bHLH, caracterizados pela presença de um domínio hélice-volta-hélice. Essa estrutura secundária permite que essas proteínas se liguem ao DNA em sítios específicos CACGTG para regular a transcrição gênica (Ledent, Paquet e Vervoort, 2002). Em MAX, essa afinidade é modulada dependendo dos cofatores associados: quando ligado com MYC, ele permite a ativação de genes proliferativos (Hurlin e Dezfouli, 2004); e,

[4] Mais tarde se determinou que o fator MAD era, na verdade, uma subfamília composta por quatro isoformas com funções distintas (revisado em Rottmann e Lüscher, 2006). Para fins de simplificação, tratarei todas apenas como MAD, a não ser quando a distinção se torne necessária.

[5] Da mesma forma que com as isoformas de MAD, não farei distinção entre MYC e MYCN ao discutir a rede MAX neste trabalho. Ao leitor que deseja se aprofundar nas diferenças entre essas isoformas, sugiro a leitura da excelente revisão de Pirity (2006) e suas referências.

quando ligado às isoformas de MAD ou homodimerizado (*i.e.* MAX:MAX), ele inibe a transcrição de genes ativadores do ciclo celular tanto por inibição direta quanto pela ativação de outros genes inibidores, como *TP53* (Rottmann e Lüscher, 2006). Alguns estudos recentes determinaram que somente o dímero MYC:MAX é capaz de regular de 10 a 15% do genoma humano (Fernandez *et al.*, 2003; Yap *et al.*, 2011). Isso torna claro que se trata de uma rede de extrema importância para o desenvolvimento celular. Em acordo com isso, existem vários estudos que tentaram, ao longo das últimas décadas, determinar as origens e funções dessa rede.

Do ponto de vista evolutivo, a rede MAX é bastante antiga – sua apresentação usual (*i.e.* MYC, MAX e MAD) surge antes da divergência do nosso último ancestral comum com os coanoflagelados (McFerrin e Atchley, 2011; Young *et al.*, 2011). Os primeiros estudos com alterações nessa rede revelaram um papel fundamental no controle do desenvolvimento em diversos organismos modelos. A deleção experimental do ortólogo de *MYC* em *Drosophila* acarreta um fenótipo diminuto, muito pouco viável (Gallant *et al.*, 1996). Em camundongos, essa mesma ablação acarreta a morte embrionária entre o 9º e 10º dias (Davis *et al.*, 1993). Similarmente, a deleção de *MAX* também é responsável por ocasionar morte embrionária (Shen-Li *et al.*, 2000), o que sugere seu papel essencial no desenvolvimento. De fato, supõe-se que o único motivo pelo qual esses mutantes murinos sobrevivam nos primeiros dias seja as reservas maternas da proteína MAX presentes no zigoto. Os estudos com ablação das isoformas de *MAD*, no entanto, não demonstraram um papel essencial no desenvolvimento (Baudino e Cleveland, 2001), o que condiz com seu surgimento relativamente tardio no panorama evolutivo (Dalmolin *et al.*, 2011).

Acredita-se que o papel principal da rede MAX seja integrar informações das mais variadas cascatas de sinalização a fim de decidir o destino da célula entre

proliferação e diferenciação. Essa decisão culmina no tipo de dimerização que será feita com MAX: com MYC, para proliferar; ou com MAD, para diferenciar. Esse controle é feito basicamente através da regulação da expressão desses dois parceiros funcionais. Alguns estudos com células de linhagem humanas e murinas observaram que tanto a entrada quanto a saída do ciclo celular é acompanhada por alterações no perfil de expressão de *MYC* e *MAD* (*e.g.* Ayer and Eisenman, 1993; Pulverer *et al.*, 2000). Quando a célula está proliferando, a expressão das isoformas de MAD estão reprimidas, o que permite a formação preferencial de dímeros MYC:MAX para manutenção do estado proliferativo. Quando a célula é induzida a diferenciar ou entrar em G0 por algum fator interno ou externo, ocorre uma diminuição da expressão de MYC seguido de um aumento da das isoformas MAD[6], de forma a suspender a replicação e ativar processos envolvidos na diferenciação. Em todos os tipos celulares estudados até agora, a expressão de MAX permanece constante durante todo o processo, sugerindo que, embora seu papel seja essencial, ele não está sujeito a variações[7]. Os fatores que levam a formação desses dímeros ainda não são completamente conhecidos – supõe-se que somente o maquinário que acompanha o dímero MYC:MAX seja composto por pelo menos mais de duas dezenas de proteínas, cada uma com diferentes características regulatórias e responsivas aos mais diversos tipos de sinais (Cole e Nikiforov, 2006). Esse maquinário intricado permite uma grande versatilidade na capacidade de respostas dessa rede, que, em última análise, regula o destino de quase todas as células nos organismos complexos.

[6] É interessante ressaltar a isoforma MAD3 (MXD3) é a única cujo efeito da dimerização com MAX é proliferativo. Essa proliferação, curiosamente, não está associada com a manutenção da fase de replicação, mas sim com um *burst* proliferativo que é necessário ocorrer antes de alguns tipos celulares entrarem efetivamente na diferenciação terminal (em algumas outras células, o pico de MAD3 pode ser substituído por uma expressão transiente de MYC) (Hurlin e Huang, 2006; Rottmann e Lüscher, 2006).

[7] Neste trabalho observamos uma possível nova etapa regulatória associada diretamente à expressão de MAX, que será abordada nos próximos tópicos.

4.2 Alterações na rede MAX e suas implicações

Seria de se esperar, dada a complexidade de fatores acima e abaixo da cascata causal da rede MAX, que esta seja uma rede particularmente sensível a perturbações externas. Na seção anterior, vimos que sua regulação se dá primariamente pelo controle da expressão dos genes *MAD* e *MYC*, que levam a célula a um caminho de diferenciação ou proliferação. Em tumores, uma das primeiras alterações que ocorre nesta rede é a perda da capacidade da atividade de MYC ser inibida pelo aumento de seu próprio conteúdo (Penn *et al.*, 1990). Em tecidos sadios, altas quantidades[8] da proteína MYC ativam uma alça de retroalimentação negativa que impede a transcrição desse gene e previne a célula de entrar em um estado proliferativo descontrolado. Simultaneamente, praticamente todas as principais rotas de sinalização conhecidas potencialmente atuam na região *cis*-regulatória de *MYC*, conferindo um terreno fértil para o estabelecimento de mutações oncogênicas (revisado em Liu e Levens, 2006). A expressão de *MAD*, por outro lado, está diminuída em alguns tumores, de forma a impedir a diferenciação celular e manter a célula permanentemente se dividindo (Han *et al.*, 2000).

Os níveis de MAX parecem se manter constantes nos processos fisiológicos de diferenciação e proliferação. Entretanto, essa afirmação não parece ser verdadeira para alguns tumores. Estudos envolvendo o sequenciamento de amostras de feocromocitomas[9] detectaram associação dessa doença com mutações possivelmente inativantes de *MAX*, sugerindo que aberrações em seu funcionamento possam induzir um estado proliferativo (Burnichon *et al.*, 2012)[10]. Essa observação servirá de ponto de

[8] Em células sadias, os níveis de MYC são de fato muito baixos. Algumas estimativas sugerem que existe apenas uma única molécula do RNAm de MYC em 40% das células adultas, enquanto que em células embrionárias, este número chegue a cinco moléculas. Em células terminalmente diferenciadas, MYC está irreversivelmente silenciado (Liu e Levens, 2006; Warrington *et al.*, 2000).

[9] Tumores de glândula adrenal.

[10] Ver também Comino-Méndez *et al.*, 2011; Galan e Kann, 2013; Welander, Söderkvist e Gimm, 2011.

partida para o restante desta discussão, onde analisarei os resultados de meu trabalho à luz de alguns estudos realizados ao longo das duas últimas décadas, que sugerem um papel mais complexo para MAX do previamente descrito.

4.3 O papel de MAX na regulação do destino celular

Até agora, discutimos a regulação da rede MAX como dependente unicamente das flutuações dos conteúdos de MYC e MAD. Embora aparentemente verdade para alguns tipos celulares, o papel regulatório de MAX parece ser mais complexo do que simplesmente servir de ponto de apoio para a heterodimerização de seus parceiros funcionais.

As primeiras investigações acerca do papel regulatório de MAX começaram na década de 1990. Lindeman e colaboradores, em um artigo publicado no periódico *Oncogene* (1995), observaram que a superexpressão de *MAX* em camundongos geneticamente suscetíveis a linfomas era capaz de diminuir a frequência de aparecimento e o tamanho dos tumores. Para isso, eles compararam os efeitos da superexpressão *MYC* pareada ou não com a de *MAX*. Como esperado, os camundongos com apenas a expressão de *MYC* em células linfoides apresentaram um aumento na taxa de surgimento de tumores. Entretanto, quando pareado com uma superexpressão de *MAX*, o efeito de MYC foi drasticamente diminuído; os tumores ocorreram em menor número, com apresentações mais brandas e mais tardias. Esse efeito, inclusive, foi capaz de afetar positivamente a sobrevivência dos camundongos. Ainda referente a esse estudo, é importante destacar que foram testadas duas variantes de transgenes com atividades distintas. Interessantemente, a supressão dos efeitos de MYC era diretamente

correlacionada com a atividade do transgene, sugerindo um efeito supressor tumoral[11] dose-dependente.

No ano seguinte, Peverali e colaboradores publicam no mesmo periódico (Peverali *et al.*, 1996) o relato de um estudo envolvendo o uso de um transgene de *MAX* durante o processo de diferenciação de células de neuroblastoma humano SK-N-BE com ácido retinoico[12]. O trabalho desse autor utilizou um protocolo de transfecção de um transgene de *MAX*, que foi capaz reduzir pela metade o tempo necessário para a diferenciação dessas células. Esse resultado, no entanto, parece ser específico para células de neuroblastoma e alguns outros tipos celulares, mas não para todos. Cañelles e colaboradores (1997) observaram que a superexpressão de *MAX* induzia a diferenciação de eritrócitos, mas o mesmo protocolo experimental em outro estudo não surtiu o menor efeito sobre a diferenciação de precursores de macrófagos (Lerga *et al.*, 1999). Esses resultados sugerem que a rede MAX, e mais especificamente o próprio MAX, está sujeito a um complexo sistema de regulação que varia dependendo do tipo celular em que está sendo expresso.

4.4 MAX e neuroblastoma

Os resultados de Lindeman, apresentados na seção anterior, demonstraram que MAX pode ter um papel supressor tumoral em alguns tipos de tumores. O trabalho subsequente de Peverali aproxima ainda mais a discussão aos nossos resultados ao demonstrar o papel desse gene na diferenciação de células de neuroblastoma. Uma

[11] Lindeman afirma no final de seu artigo:

"*The effects of MAX overexpression observed here were antiproliferative in nature. No evidence was found that* max *can act as an oncogene."*

Embora ele não postule exatamente com essas palavras, esse é o primeiro estudo que define categoricamente *MAX* como um potencial gene supressor tumoral. Essa definição só será retomada efetivamente nos últimos anos.

[12] *Idem* a nota nº 2.

crítica que poderia ser feita a esses estudos quanto a sua implicação é que eles envolvem a superexpressão de *MAX* e, portanto, correspondem a alterações que não condizem com o que é observado em nível de organismo (analogamente, seria uma situação similar aos trabalhos expressão de *PAX6*[13] em drosófilas – importantes do ponto de vista mecanicista, mas não relacionados a nenhum fenótipo observável na natureza[14]). Nossos resultados, no entanto, fazem a primeira ponte desses resultados com um possível fenômeno que ocorre *in vivo*. A regulação da própria expressão de *MAX* pode ser um dos passos-chaves necessários para o estabelecimento de neuroblastomas, sugerindo um novo campo para busca de alvos terapêuticos. Sendo assim, esses dois trabalhos formarão o alicerce para discutirmos o papel de MAX na progressão de neuroblastoma no restante desta discussão.

Apesar de não relacionado diretamente com a progressão de neuroblastoma anteriormente, a ideia de usar MAX como alvo terapêutico não é exatamente nova. Montagne e colaboradores (2012) utilizaram o domínio hélice-volta-hélice de MAX para inibir o crescimento de células HeLa[15]. O princípio da técnica é relativamente simples – a superexpressão desse domínio ocuparia os sítios ativos das moléculas de MYC, impedindo que este se ligasse com MAX para induzir a proliferação. Esse *insight* parece ser contraintuitivo com o que observamos em nosso trabalho, uma vez que a ideia dos pesquisadores é justamente impedir a heterodimerização de MYC e MAX. Para entendermos melhor como a alta expressão de MAX poderia ser benéfica para o

[13] A expressão de *PAX6* é o sinal homeótico para iniciar a cascata de formação de estruturas oculares em bilatérios. Os mutantes com *PAX6* expresso ectopicamente desenvolvem olhos rudimentares em outras partes do corpo (Bonini *et al.*, 1997).

[14] Embora o mau funcionamento de ortólogos de *PAX6* esteja sabidamente relacionado com erros no desenvolvimento de olhos e outras estruturas sensoriais, não há, para conhecimento deste autor, um único caso descrito na literatura de superexpressão ectópica desse gene em organismos que não seja de forma experimental.

[15] A primeira célula de linhagem humana, desenvolvida nos anos 1950. Oriunda de câncer cervical.

organismo, é necessário nos debruçarmos sobre mais um estudo importante realizado na última década.

Banerjee e colaboradores (2006) determinaram as propriedades termodinâmicas das ligações MYC:MAX, MAD:MAX e MAX:MAX. Eles observaram que o heterodímero MYC:MAX é o mais estável de todos, e sua formação é preferencial na presença dos outros parceiros funcionais. Pensando em um cenário onde a expressão de MAX parece sofrer regulação, podemos postular que níveis baixos desta proteína favoreceriam a formação dos dímeros proliferativos e, à medida que os níveis de MAX aumentam, todos os sítios de MYC se tornam ocupados e se muda o balanço para a formação dos outros dímeros não proliferativos. Partindo dessa premissa, poderíamos esperar que, levando somente em conta a expressão de MAX, pacientes com baixos níveis desta proteína apresentariam desfechos piores do que os com níveis mais altos. É claro que em um cenário real, a quantidade de fatores que interagem diretamente nessa dinâmica é muito alta, e essa afirmativa talvez não se traduziria para a clínica. Nossos resultados, entretanto, demonstram exatamente o contrário; a alta expressão de *MAX* correlacionou com um prognóstico favorável em duas coortes independentes de pacientes. Isso sugere que este gene está de fato envolvido diretamente na progressão de neuroblastoma.

Ao analisarmos o conteúdo de MAX durante a diferenciação de células de neuroblastoma, descrevemos pela primeira vez que ele se altera espontaneamente durante esse processo. Esses resultados fortalecem as duas premissas citadas anteriormente, de que 1) MAX está de fato sofrendo algum tipo de regulação em células

de neuroblastoma e 2) os altos níveis desta proteína ocorrem na etapa de efetivação[16] do fenótipo neuronal dessas células, implicando sua correlação com um fenótipo clinicamente favorável. Dessa forma, acreditamos ter chegado em uma base sólida para implicar o envolvimento desse gene na progressão de neuroblastoma.

[16] Por efetivação, refiro-me a apresentação *de facto* do fenótipo neuronal (*i.e.* prolongamentos dendríticos, etc.), que só ocorre a partir do quarto dia, justamente quando foi observado o aumento da expressão de MAX.

5. CONCLUSÕES

5.1 Cumprimento dos objetivos

Tendo em vista os quatro objetivos expostos no início deste trabalho, podemos afirmar que todos foram cumpridos. Sendo assim, este trabalho foi capaz de responder todas as perguntas que se propusera inicialmente:

1) Fomos capazes de reconstruir não apenas uma, mas duas redes regulatória de neuroblastoma utilizando dados de expressão públicos, 2) que puderam ser utilizadas para analisarmos uma assinatura metastática e definir MAX como um dos reguladores mestres do fenótipo tumoral. 3) Observamos que a expressão de *MAX* correlacionou com o desfecho de pacientes de neuroblastoma, validando seu papel como possível biomarcador. 4) Por último, verificamos que o conteúdo de MAX estava aumentado durante a diferenciação de células de neuroblastoma, indicando seu provável papel na progressão da doença.

5.2 Considerações finais e perspectivas

Ao longo deste trabalho, observamos o possível papel de MAX na progressão de neuroblastoma. Entretanto, algumas perguntas ficaram em aberto e deverão ser respondidas em um futuro próximo. Primeiro, se o mecanismo que postulamos está correto, por que a expressão de MAX não correlacionou com o desfecho dos pacientes da terceira coorte de sobrevivência? Outros fatores não tão evidentes devem ser levados em conta antes de adotarmos esse gene como um marcador. Além disso, também há o papel dos fatores TFEC e ZNF101, que foram apontados como reguladores mestres da progressão dessa doença. A expressão de *TFEC* foi capaz de discriminar pacientes de

câncer de pulmão quanto a sua sobrevivência, sugerindo que esse gene deve ser alvo de estudos futuros. Por último, devemos entender exatamente qual o papel de MAX na diferenciação neuronal. Vimos que ele está aumentado no decorrer do processo, mas não há como implicar seu papel diretamente. Serão necessários estudos de ganho e perda de função para que entendamos mais profundamente seu funcionamento.

5. REFERÊNCIAS

AMBROS, P. F. *et al.* International consensus for neuroblastoma molecular diagnostics: report from the International Neuroblastoma Risk Group (INRG) Biology Committee. **British journal of cancer**, v. 100, n. 9, p. 1471–82, 5 maio 2009.

ASGHARZADEH, S. *et al.* Prognostic significance of gene expression profiles of metastatic neuroblastomas lacking MYCN gene amplification. **Journal of the National Cancer Institute**, v. 98, n. 17, p. 1193–203, 6 set. 2006.

AYER, D. E.; EISENMAN, R. N. A switch from Myc:Max to Mad:Max heterocomplexes accompanies monocyte/macrophage differentiation. **Genes & development**, v. 7, n. 11, p. 2110–9, nov. 1993.

AYER, D. E.; KRETZNER, L.; EISENMAN, R. N. Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. **Cell**, v. 72, n. 2, p. 211–22, 29 jan. 1993.

BABU, M. M.; LANG, B.; ARAVIND, L. Methods to reconstruct and compare transcriptional regulatory networks. **Methods in molecular biology (Clifton, N.J.)**, v. 541, p. 163–80, jan. 2009.

BAGATELL, R. *et al.* Significance of MYCN amplification in international neuroblastoma staging system stage 1 and 2 neuroblastoma: a report from the International Neuroblastoma Risk Group database. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 27, n. 3, p. 365–70, 20 jan. 2009.

BANERJEE, A.; HU, J.; GOSS, D. J. Thermodynamics of protein-protein interactions of cMyc, Max, and Mad: effect of polyions on protein dimerization. **Biochemistry**, v. 45, n. 7, p. 2333–8, 21 fev. 2006.

BARABÁSI, A.-L.; OLTVAI, Z. N. Network biology: understanding the cell's functional organization. **Nature reviews. Genetics**, v. 5, n. 2, p. 101–13, fev. 2004.

BASSO, K. *et al.* Reverse engineering of regulatory networks in human B cells. **Nature genetics**, v. 37, n. 4, p. 382–90, abr. 2005.

BAUDINO, T. A.; CLEVELAND, J. L. The Max network gone mad. **Molecular and cellular biology**, v. 21, n. 3, p. 691–702, fev. 2001.

BHATNAGAR, S. N.; SARIN, Y. K. Neuroblastoma: a review of management and outcome. **Indian journal of pediatrics**, v. 79, n. 6, p. 787–92, jun. 2012.

BLACKWOOD, E. M.; EISENMAN, R. N. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. **Science (New York, N.Y.)**, v. 251, n. 4998, p. 1211–7, 8 mar. 1991.

BONINI, N. M. *et al.* The Drosophila eyes absent gene directs ectopic eye formation in a pathway conserved between flies and vertebrates. **Development (Cambridge, England)**, v. 124, n. 23, p. 4819–26, dez. 1997.

BURNICHON, N. *et al.* MAX mutations cause hereditary and sporadic pheochromocytoma and paraganglioma. **Clinical cancer research : an official journal of the American Association for Cancer Research**, v. 18, n. 10, p. 2828–37, 15 maio 2012.

CAÑELLES, M. *et al.* Max and inhibitory c-Myc mutants induce erythroid differentiation and resistance to apoptosis in human myeloid leukemia cells. **Oncogene**, v. 14, n. 11, p. 1315–27, 20 mar. 1997.

CARRO, M. S. *et al.* The transcriptional network for mesenchymal transformation of brain tumours. **Nature**, v. 463, n. 7279, p. 318–25, 21 jan. 2010.

CARTUM, J. Neuroblastoma: o enigmático tumor da infância. **Pediatria Moderna**, v. 48, n. 8, p. 296–301, 2012.

COLE, M. D.; NIKIFOROV, M. A. Transcriptional activation by the Myc oncoprotein. **Current topics in microbiology and immunology**, v. 302, p. 33–50, jan. 2006.

COMINO-MÉNDEZ, I. *et al.* Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. **Nature genetics**, v. 43, n. 7, p. 663–7, jul. 2011.

D'ANGIO, G.; EVANS, A.; KOOP, C. E. SPECIAL PATTERN OF WIDESPREAD NEUROBLASTOMA WITH A FAVOURABLE PROGNOSIS. **The Lancet**, v. 297, n. 7708, p. 1046–1049, maio 1971.

DALMOLIN, R. J. S. *et al.* Evolutionary plasticity determination by orthologous groups distribution. **Biology direct**, v. 6, n. 1, p. 22, jan. 2011.

DANG, C. V. MYC on the path to cancer. **Cell**, v. 149, n. 1, p. 22–35, 30 mar. 2012.

DAVIS, A. C. *et al.* A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. **Genes & development**, v. 7, n. 4, p. 671–82, abr. 1993.

DEYELL, R. J.; ATTIYEH, E. F. Advances in the understanding of constitutional and somatic genomic alterations in neuroblastoma. **Cancer genetics**, v. 204, n. 3, p. 113–21, mar. 2011.

FERNANDEZ, P. C. *et al.* Genomic targets of the human c-Myc protein. **Genes & development**, v. 17, n. 9, p. 1115–29, 1 maio 2003.

FLETCHER, M. N. C. *et al.* Master regulators of FGFR2 signalling and breast cancer risk. **Nature communications**, v. 4, p. 2464, 17 set. 2013.

FROTA JUNIOR, M. L. C. DA *et al.* In vitro optimization of retinoic acid-induced neuritogenesis and TH endogenous expression in human SH-SY5Y neuroblastoma cells

by the antioxidant Trolox. **Molecular and cellular biochemistry**, v. 358, n. 1-2, p. 325–34, dez. 2011.

GALAN, S. R.; KANN, P. H. Genetics and molecular pathogenesis of pheochromocytoma and paraganglioma. **Clinical endocrinology**, v. 78, n. 2, p. 165–75, fev. 2013.

GALLANT, P. *et al.* Myc and Max homologs in Drosophila. **Science (New York, N.Y.)**, v. 274, n. 5292, p. 1523–7, 29 nov. 1996.

GENTLEMAN, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. **Genome biology**, v. 5, n. 10, p. R80, jan. 2004.

HAIBE-KAINS, B. *et al.* A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? **Bioinformatics (Oxford, England)**, v. 24, n. 19, p. 2200–8, 1 out. 2008.

HAN, S. *et al.* Clinical implication of altered expression of Mad1 protein in human breast carcinoma. **Cancer**, v. 88, n. 7, p. 1623–32, 1 abr. 2000.

HARRINGTON, E. D.; JENSEN, L. J.; BORK, P. Predicting biological networks from genomic data. **FEBS letters**, v. 582, n. 8, p. 1251–8, 9 abr. 2008.

HARTWELL, L. H. *et al.* From molecular to modular cell biology. **Nature**, v. 402, n. 6761 Suppl, p. C47–52, 2 dez. 1999.

HOWMAN-GILES, R. *et al.* Neuroblastoma and other neuroendocrine tumors. **Seminars in nuclear medicine**, v. 37, n. 4, p. 286–302, jul. 2007.

HURLIN, P. J.; DEZFOULI, S. Functions of myc:max in the control of cell proliferation and tumorigenesis. **International review of cytology**, v. 238, p. 183–226, jan. 2004.

HURLIN, P. J.; HUANG, J. The MAX-interacting transcription factor network. **Seminars in cancer biology**, v. 16, n. 4, p. 265–74, ago. 2006.

LEDENT, V.; PAQUET, O.; VERVOORT, M. Phylogenetic analysis of the human basic helix-loop-helix proteins. **Genome biology**, v. 3, n. 6, p. RESEARCH0030, jan. 2002.

LEFEBVRE, C. *et al.* A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. **Molecular systems biology**, v. 6, n. 377, p. 377, 8 jun. 2010.

LERGA, A. *et al.* Regulation of c-Myc and Max in megakaryocytic and monocytic-macrophagic differentiation of K562 cells induced by protein kinase C modifiers: c-Myc is down-regulated but does not inhibit differentiation. **Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research**, v. 10, n. 9, p. 639–54, set. 1999.

LINDEMAN, G. J. *et al.* Overexpressed max is not oncogenic and attenuates myc-induced lymphoproliferation and lymphomagenesis in transgenic mice. **Oncogene**, v. 10, n. 5, p. 1013–7, 2 mar. 1995.

LIU, J.; LEVENS, D. Making myc. **Current topics in microbiology and immunology**, v. 302, p. 1–32, jan. 2006.

MARGOLIN, A. A *et al.* Reverse engineering cellular networks. **Nature protocols**, v. 1, n. 2, p. 662–71, jan. 2006.

MARIS, J. M. *et al.* Neuroblastoma. **Lancet**, v. 369, n. 9579, p. 2106–20, 23 jun. 2007.

MARIS, J. M.; MATTHAY, K. K. Molecular biology of neuroblastoma. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 17, n. 7, p. 2264–79, jul. 1999.

MCFERRIN, L. G.; ATCHLEY, W. R. Evolution of the Max and Mlx networks in animals. **Genome biology and evolution**, v. 3, p. 915–37, jan. 2011.

MILLS, S. E. Neuroectodermal neoplasms of the head and neck with emphasis on neuroendocrine carcinomas. **Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc**, v. 15, n. 3, p. 264–78, mar. 2002.

MONCLAIR, T. *et al.* The International Neuroblastoma Risk Group (INRG) staging system: an INRG Task Force report. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 27, n. 2, p. 298–303, 10 jan. 2009.

MONTAGNE, M. *et al.* The Max b-HLH-LZ can transduce into cells and inhibit c-Myc transcriptional activities. **PloS one**, v. 7, n. 2, p. e32172, jan. 2012.

OBERTHUER, A. *et al.* Comparison of performance of one-color and two-color gene-expression analyses in predicting clinical endpoints of neuroblastoma patients. **The pharmacogenomics journal**, v. 10, n. 4, p. 258–66, 1 ago. 2010.

OOI, W. F. *et al.* Segmental Chromosome Aberrations Converge on Overexpression of Mitotic Spindle Regulatory Genes in High-Risk Neuroblastoma. v. 556, n. February, p. 545–556, 2012.

PARISE, I. Z. S. *et al.* Neuroblastoma in southern Brazil: an 11-year study. **Journal of pediatric hematology/oncology**, v. 28, n. 2, p. 82–7, fev. 2006.

PENN, L. J. *et al.* Negative autoregulation of c-myc transcription. **The EMBO journal**, v. 9, n. 4, p. 1113–21, abr. 1990.

PEVERALI, F. A. *et al.* Retinoic acid-induced growth arrest and differentiation of neuroblastoma cells are counteracted by N-myc and enhanced by max overexpressions. **Oncogene**, v. 12, n. 2, p. 457–62, 18 jan. 1996.

PIANOVSKI, M. A. D. *et al.* Mortality rate of adrenocortical tumors in children under 15 years of age in Curitiba, Brazil. **Pediatric blood & cancer**, v. 47, n. 1, p. 56–60, jul. 2006.

PIRITY, M.; BLANCK, J. K.; SCHREIBER-AGUS, N. Lessons learned from Myc/Max/Mad knockout mice. **Current topics in microbiology and immunology**, v. 302, p. 205–34, jan. 2006.

PRETER, K. DE *et al.* Accurate outcome prediction in neuroblastoma across independent data sets using a multigene signature. **Clinical cancer research : an official journal of the American Association for Cancer Research**, v. 16, n. 5, p. 1532–41, 1 mar. 2010.

PULVERER, B. *et al.* Analysis of Myc/Max/Mad network members in adipogenesis: inhibition of the proliferative burst and differentiation by ectopically expressed Mad1. **Journal of cellular physiology**, v. 183, n. 3, p. 399–410, jun. 2000.

ROTTMANN, S.; LÜSCHER, B. The Mad side of the Max network: antagonizing the function of Myc and more. **Current topics in microbiology and immunology**, v. 302, p. 63–122, jan. 2006.

SCHMIDT, M. L. *et al.* Biologic factors determine prognosis in infants with stage IV neuroblastoma: A prospective Children's Cancer Group study. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 18, n. 6, p. 1260–8, mar. 2000.

SCHMIDT, M. LOU *et al.* Favorable prognosis for patients 12 to 18 months of age with stage 4 nonamplified MYCN neuroblastoma: a Children's Cancer Group Study. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 23, n. 27, p. 6474–80, 20 set. 2005.

SCHRAMM, A. *et al.* Biological effects of TrkA and TrkB receptor signaling in neuroblastoma. **Cancer letters**, v. 228, n. 1-2, p. 143–53, 18 out. 2005.

SCHWAB, M. *et al.* Neuroblastoma: biology and molecular and chromosomal pathology. **The lancet oncology**, v. 4, n. 8, p. 472–80, ago. 2003.

SHEN-LI, H. *et al.* Essential role for Max in early embryonic growth and development. **Genes & development**, v. 14, n. 1, p. 17–22, 1 jan. 2000.

SPIX, C. *et al.* Neuroblastoma incidence and survival in European children (1978-1997): report from the Automated Childhood Cancer Information System project. **European journal of cancer (Oxford, England : 1990)**, v. 42, n. 13, p. 2081–91, set. 2006.

THERNEAU, T.; GRAMBSCH, P. **Modeling survival data: extending the Cox model**. [s.l.] Springer, 2000. p. 350

TONINI, G. P. *et al.* MYCN oncogene amplification in neuroblastoma is associated with worse prognosis, except in stage 4s: the Italian experience with 295 children.

Journal of clinical oncology : official journal of the American Society of Clinical Oncology, v. 15, n. 1, p. 85–93, jan. 1997.

TONINI, G. P. *et al.* Familial neuroblastoma: a complex heritable disease. **Cancer Letters**, v. 197, n. 1-2, p. 41–45, jul. 2003.

WAGNER, L. M.; DANKS, M. K. New therapeutic targets for the treatment of high-risk neuroblastoma. **Journal of cellular biochemistry**, v. 107, n. 1, p. 46–57, 1 maio 2009.

WARRINGTON, J. A. *et al.* Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. **Physiological genomics**, v. 2, n. 3, p. 143–7, 27 abr. 2000.

WELANDER, J.; SÖDERKVIST, P.; GIMM, O. Genetics and clinical characteristics of hereditary pheochromocytomas and paragangliomas. **Endocrine-related cancer**, v. 18, n. 6, p. R253–76, dez. 2011.

WESTERMANN, F.; SCHWAB, M. Genetic parameters of neuroblastomas. **Cancer letters**, v. 184, n. 2, p. 127–47, 28 out. 2002.

YAMADA, T.; BORK, P. Evolution of biomolecular networks: lessons from metabolic and protein interactions. **Nature reviews. Molecular cell biology**, v. 10, n. 11, p. 791–803, nov. 2009.

YAP, C.-S. *et al.* Kinetic profiling of the c-Myc transcriptome and bioinformatic analysis of repressed gene promoters. **Cell cycle (Georgetown, Tex.)**, v. 10, n. 13, p. 2184–96, 1 jul. 2011.

YOUNG, S. L. *et al.* Premetazoan ancestry of the Myc-Max network. **Molecular biology and evolution**, v. 28, n. 10, p. 2961–71, out. 2011.

ANEXO I – Artigo publicado durante o período do Mestrado

DIFFERENTIAL EVOLUTIONARY CONSTRAINTS IN THE EVOLUTION OF CHEMORECEPTORS: A MURINE AND HUMAN CASE STUDY[17]

Ricardo D'Oliveira Albanus, Rodrigo Juliani Siqueira Dalmolin, José Luíz Rybarczyk-Filho, Mauro Antônio Alves Castro, José Cláudio Fonseca Moreira

[17] Artigo publicado no periódico *The Scientific World Journal* (doi: http://dx.doi.org/10.1155/2014/696485) em janeiro de 2014; vol. 2014, Article ID 696485.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 696485, 9 pages
http://dx.doi.org/10.1155/2014/696485

Research Article

Differential Evolutionary Constraints in the Evolution of Chemoreceptors: A Murine and Human Case Study

Ricardo D'Oliveira Albanus,[1] Rodrigo Juliani Siqueira Dalmolin,[1] José Luiz Rybarczyk-Filho,[2] Mauro Antônio Alves Castro,[1] and José Cláudio Fonseca Moreira[1]

[1] *Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Rua Ramiro Barcelos 2600, 90040-180 Porto Alegre, RS, Brazil*
[2] *Departamento de Física e Biofísica, Universidade Estadual Paulista, Distrito de Rubião Júnior, S/N, 18618-970 Botucatu, SP, Brazil*

Correspondence should be addressed to Ricardo D'Oliveira Albanus; ricardo.albanus@ufrgs.br

Received 7 August 2013; Accepted 23 October 2013; Published 23 January 2014

Academic Editors: M. Frank, B. Gantenbein-Ritter, B.-Y. Liao, and M. Shimoyama

Chemoreception is among the most important sensory modalities in animals. Organisms use the ability to perceive chemical compounds in all major ecological activities. Recent studies have allowed the characterization of chemoreceptor gene families. These genes present strikingly high variability in copy numbers and pseudogenization degrees among different species, but the mechanisms underlying their evolution are not fully understood. We have analyzed the functional networks of these genes, their orthologs distribution, and performed phylogenetic analyses in order to investigate their evolutionary dynamics. We have modeled the chemosensory networks and compared the evolutionary constraints of their genes in *Mus musculus*, *Homo sapiens*, and *Rattus norvegicus*. We have observed significant differences regarding the constraints on the orthologous groups and network topologies of chemoreceptors and signal transduction machinery. Our findings suggest that chemosensory receptor genes are less constrained than their signal transducing machinery, resulting in greater receptor diversity and conservation of information processing pathways. More importantly, we have observed significant differences among the receptors themselves, suggesting that olfactory and bitter taste receptors are more conserved than vomeronasal receptors.

1. Introduction

The ability to evaluate the environment has always been of vital importance to all organisms. In order to find food, detect dangers, and search for reproductive partners, a constant appraisal of the outside world must be made by any organism. Chemosensory reception is one such tool for this task, and it is present in all life forms. Over the last decade, several studies were conducted in order to characterize the different chemosensory receptors (CR) genes [1–4]. In vertebrates, they are coded by six major multigene families: the trace amine-associated receptors (TAAR) [5], the olfactory receptors (OR) [6], the type I and II vomeronasal receptors (V1R and V2R) [3, 4, 7], and type I and II taste receptors (T1R and T2R) [1, 2]. All proteins coded by these genes are G protein-coupled proteins [8].

Different from other environmental appraisal systems such as vision and hearing, which remained relatively stable once they were formed, chemosensory reception must be constantly tuned to an ever-changing environment of odors and toxins. This need for variability is reflected in the organization of the CR genes in the genome. In all studied species, it was found that these genes occur in great numbers, and there are considerable numbers of CR pseudogenes [4, 9–11], suggesting that they are prone to duplication and inactivation events. There are theories to explain the evolution of CR genes [9, 11–14], but several gaps regarding this subject still remain. For instance, there are no currently available data regarding the evolutionary dynamics of the chemosensory apparatus as a whole (i.e., the CR and its signal transducing machinery). Equally unclear are the differences in evolutionary dynamics among the CR families.

In this work, we have tackled the evolution of the mammalian CR gene families and their signal transducing machinery from a systems biology-oriented approach. We have analyzed the orthologs distribution of the chemosensory machinery, their functional networks topologies, and their phylogenetic diversity in *Mus musculus*, *Rattus norvegicus*, and *Homo sapiens*. We have found evidences that there are distinct evolutionary dynamics in the CR genes and the signal transducing apparatus. More importantly, we have observed significant differences among the CR gene families, suggesting distinct evolutionary dynamics for each receptor type.

2. Methods

2.1. Data Collection. In order to determine which receptors are involved in each sensory modality, we have gathered data from the Gene Ontology (GO) Consortium [15] regarding *Homo sapiens, Mus musculus,* and *Rattus norvegicus*. GO groups used were 0004984—MF Olfactory receptor activity, 0007608—BP Sensory perception of smell, 0008527—MF Taste receptor activity, 0050909—BP Sensory perception of taste, 0016503—MF Pheromone receptor activity, 0019236—BP Response to Pheromone. We have chosen these three species for our study due to robustness of their genomic/proteomic data available in databases. Studied genes were sorted in groups according to their receptor modality: olfactory receptors; taste receptors; and vomeronasal receptors. We made one further division of the GO taste group to separate taste receptors type 1 and 2 and study them separately because of their functional differences. TAAR genes were withdrawn from our analysis due to lack of data in the databases. Also due to lack of available data, we have combined the two vomeronasal families (V1R and V2R) and studied them as a single group (VN). We have sorted all genes in GO groups into two functional categories: the first consisted of genes coding the proteins directly involved in binding chemical stimuli (the chemosensory receptors *per se*), and the second consisted of the rest of the genes related to signal transduction machinery (STM).

Functional network parameters of proteins coded by CR genes were gathered using STRING database (String-DB), version 8.3 [16], using their corresponding ENSEMBL IDs. To assemble these IDs, a cross-search was performed between GO, String-DB, HUGO Gene Nomenclature Consortium [17], Mouse Genome Informatics [18], Rat Genome Database [19], and BioMart [20] databases. Genes that presented ID divergences among databases were manually curated or removed from our analysis. String-DB analyses were made with a 0.7 combined score and only interactions generated from experiments and databases were used. This is a medium to high stringency parameter.

2.2. Topology and Evolutionary Plasticity Analysis. Topologies of the receptors functional networks were analyzed by connectivity [$k(i)$] and clusterization [$c(i)$] indexes of their components. $k(i)$ index is calculated by the number of neighbors that an i node has in a network, and $c(i)$ by the equation

$$c(i) = \frac{2n_i}{k_i(k_i - 1)}, \tag{1}$$

which represent general interactivity of i's neighbors, where n_i represents the number of their connections among each other. Evolutionary Plasticity Index (EPI) of the orthologous groups of these proteins was calculated by equation

$$\text{EPI} = 1 - \frac{H_\alpha}{\sqrt{D_\alpha}}, \tag{2}$$

where H_α is the ortholog diversity in the eukaryotic tree, calculated using how many species the ortholog is found, and D_α is its abundance, calculated by the number of ortholog members found in each species [21]. Orthology data of these proteins was also gathered using String-DB. All statistical analyses were made using one-way ANOVA with Tukey's test. $k(i)$ and $c(i)$ indexes were compared by the Shannon diversity (S) of their distribution, using equation

$$S = -\sum p \ln p, \tag{3}$$

where p is the probability of a value occurrence in any dataset. Entropy calculation was used in a complementary way in order to mathematically support or refute any observations in the connectivity and clusterization distribution behavior. In order to generate the graphical representations of the CR network, we have plotted String-DB interactions of all Gene Ontology groups proteins among each other using RedeR R package [22]. The list of all the genes analyzed in this work is presented in the Supplementary Material available online at http://dx.doi.org/10.1155/2014/696485.

2.3. Phylogenetic Analysis. Chemoreceptor genes sequences were gathered from the Chemosensory Receptor Database [23]. Alignments and trees were made with the MEGA 5.2 software [24], using, respectively, the Muscle alignment algorithm [25, 26] and the Tamura-Nei model [27]. Parameters used were the default for each algorithm. Branch reliability was calculated using bootstrap method. 100 bootstrap replications were performed for T1R, T2R, and VN and 50 replications for OR. For entropy analysis, we have calculated the Shannon diversity of the phylogenetic trees by subsetting each tree into n consecutives samples of w size, where n is the tree size and w is the maximum tree depth (i.e., number of levels). This was made to detect whether tree diversity was consistent throughout the entire tree radius. One-way ANOVA was used in order to compare these results. We have chosen w as the number of levels as a means for defining proportional windows for each tree.

3. Results

3.1. Differences between the Chemosensory Receptors and the Signal Transducing Machinery. We have calculated separately the Evolutionary Plasticity Index (EPI) [21] of the CR genes

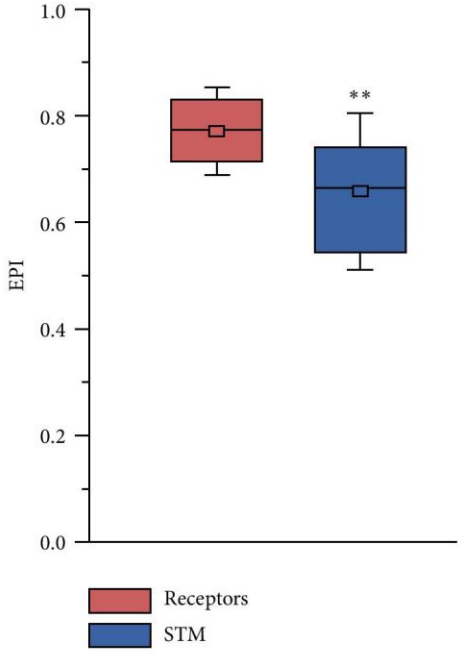

FIGURE 1: Mean EPI values of chemoreceptors (red) and signal transducing machinery (blue). The edges of the boxes indicate the upper and lower quartiles. The line at the center of each box indicates the median, the square represents the mean, and whiskers represent the standard deviation. Asterisks indicate statistically significant data ($P < 0.001$).

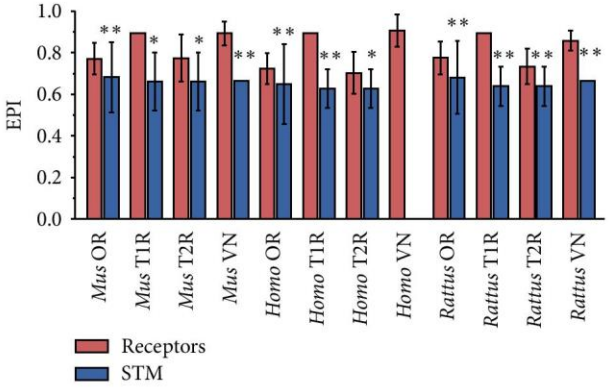

FIGURE 2: Mean EPI values of chemoreceptors families (red) and their respective signal transducing machineries (blue). Plasticity values are shown in the vertical axis and the different subgroups are listed on the horizontal axis. Whiskers represent the standard deviation. Statistically significant data are indicated by double ($P < 0.001$) and single ($P < 0.05$) asterisks.

and their signal transducing machinery (STM). We have observed that CR genes as a whole have significantly higher plasticity values than the STM (Figure 1), indicating that CR genes have a broader ortholog distribution than the STM, meaning that the latter was subject to less variation during the course of evolution. To further corroborate these findings, we have compared each CR family separately to its signal transducing machinery. We have found that, in all cases but one, the EPI of each CR family was significantly higher than its STM (Figure 2). The exception was the human vomeronasal (VN) genes, which lack their STM due to the loss of the TRPC2 channel [28, 29].

Next, we have compared the network topologies of each CR family and their STM. We have observed that most CR genes are functionally less connected than their STM. Most CR genes are connected only to their respective G proteins, indicating that they are located in the periphery of their functional networks (Figures 3 and 4). This assumption is further supported by analyzing the Shannon diversity of the connectivity and clusterization indexes. We have found that the STM has higher diversity values for these indexes ($P < 0.05$), suggesting that they occupy a broader range of niches in their network. Exceptions to this are some olfactory receptors, which presented higher connectivity and clusterization values among each other.

3.2. Differences among the Different Chemosensory Families. We have compared the EPI of the different CR families with themselves in order to identify differences in their orthologs distribution. Due to lack of data regarding the V1R and V2R, we have considered these genes as a single group in our analysis (VN). Strikingly, we have observed that CR families can be sorted in two groups regarding their plasticity. The OR and T2R have significantly lower plasticity than the T1R and VN in the three mammals we have studied, indicating that they had evolved under different constraints in these species (Figure 5). To further assess these differences, we have reconstructed the phylogenetic relationships among each CR family. We have observed that the OR, T1, and T2 genes form branches preferentially with their orthologs in other species, whereas the VN genes branches with their inparalogs (Figure 6). These results are further supported by calculating the Shannon diversity index stepwise for each CR tree. We have found that the VN tree had significantly lower diversity values ($P < 10^{-16}$) than the other CR, suggesting that the VN genes are less conserved than the other CR. The original trees with bootstrap replications confidence values can be found in the Supplementary Material.

3.3. The Functional Organization of the CR Genes Network. Finally, we have reconstructed the CR genes network in order to visualize its functional organization. We can observe that even though they form completely separate functional clusters, all the CR families, with the exception of VN, share the same STM cluster (Figure 7). This indicates that the STM machinery is essentially the same in every CR cell type.

4. Discussion

Chemosensory perception is one of the most important systems for appraisal of the environment. It is of vital necessity to every organism that the chemical species detected by each chemoreceptor are tuned to tastes or odorants which bring meaningful information from the outside world. Unlike physical sensory modalities, whose stimuli nature is constant (e.g., light, sound), chemical perception may be

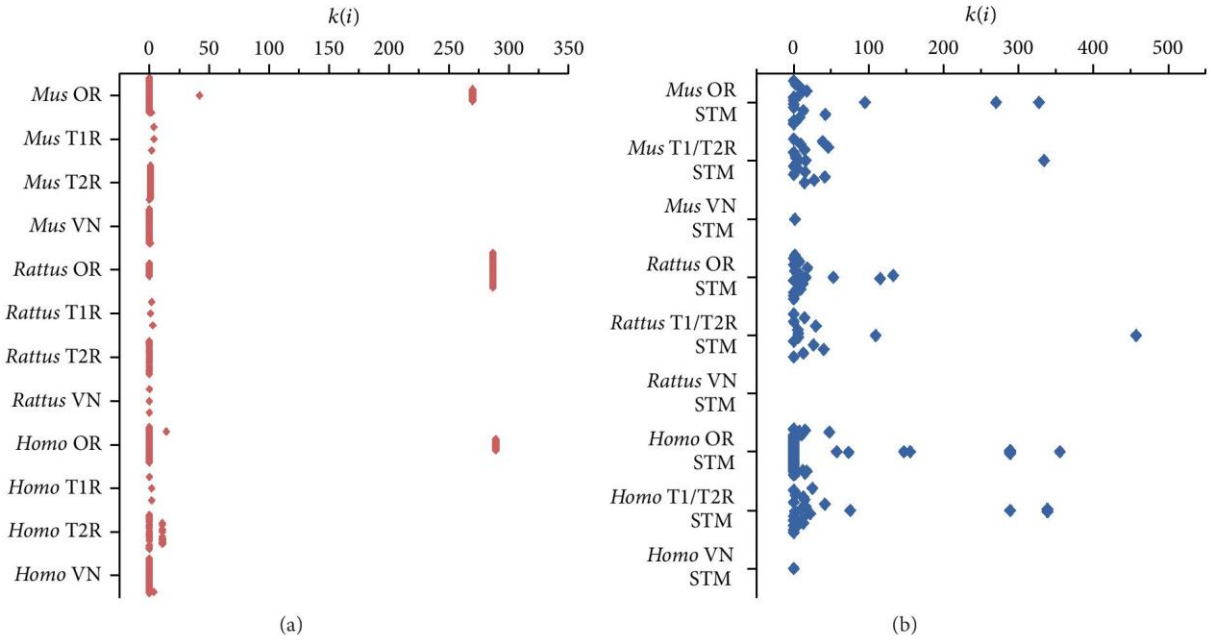

FIGURE 3: Connectivity values distribution for the chemoreceptors families (red) and their respective signal transducing machineries (blue). Values are shown in the vertical axis and the different subgroups are listed in the horizontal axis.

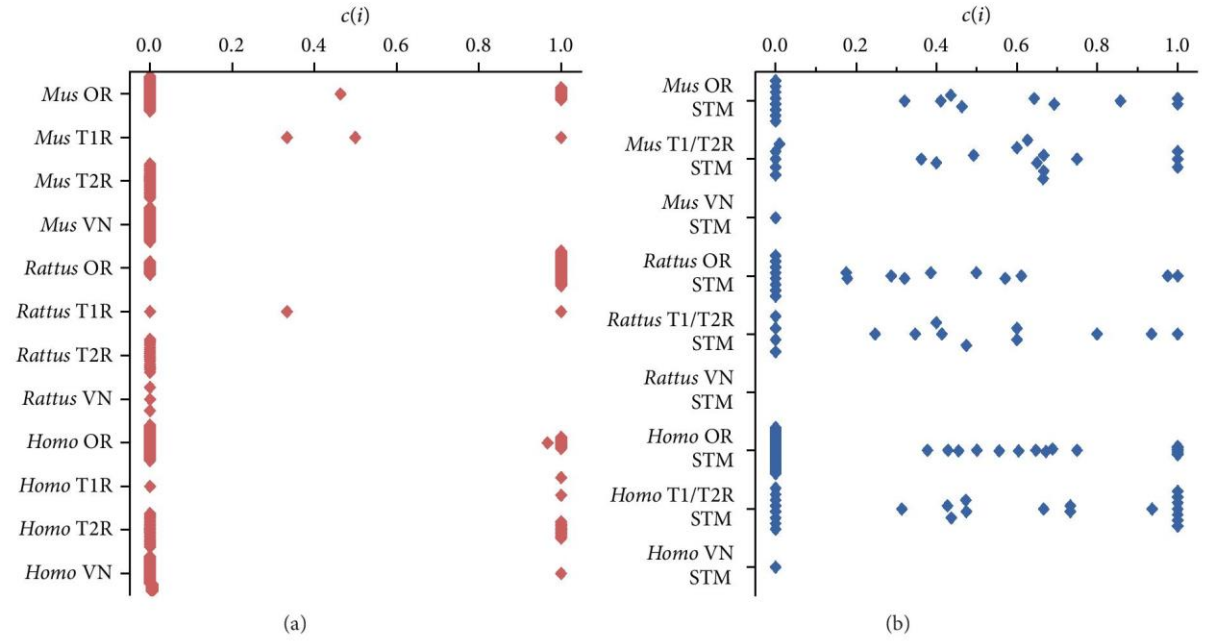

FIGURE 4: Clusterization values distribution for the chemoreceptors families (red) and their respective signal transducing machineries (blue). Values are shown in the vertical axis and the different subgroups are listed in the horizontal axis.

subject to radical changes in very short time windows. For instance, some plants are able to change their repertory of toxic secondary compounds in just a few generations [30], forcing herbivorous species that can potentially ingest these compounds to keep equally updated their ability for detecting these toxins. From an evolutionary point of view, this means that the genes coding these receptors must have a more relaxed behavior in order to accommodate novelties in the environment.

When comparing all CR genes to their STM, we have observed that CR have higher evolutionary plasticity values, suggesting that they were more subject to variation in the course of evolution than the STM. This indicates that the STM has remained relatively unchanged since its appearance,

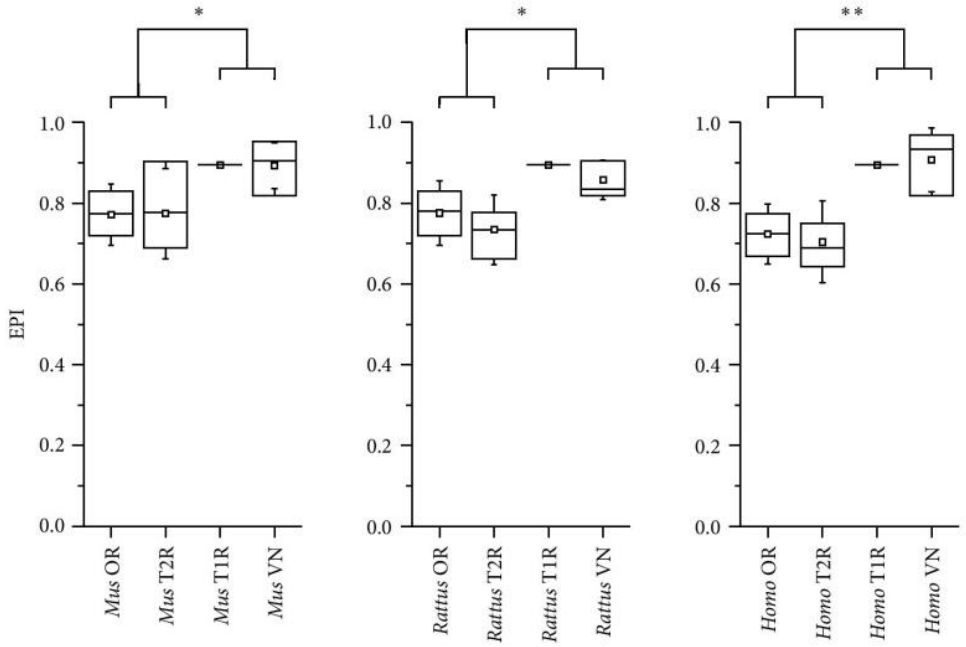

FIGURE 5: Mean EPI values of chemoreceptors families. Plasticity values are shown in the vertical axis and the different families are listed on the horizontal axis. The line at the center of each box indicates the median, the square represents the mean, and whiskers represent the standard deviation. Statistically significant data are indicated by double ($P < 0.001$) and single ($P < 0.05$) asterisks.

while the receptors themselves were free to experiment with the environment. By analyzing the network topology of the CR and STM, we have observed that CR occupy a peripheral position in their functional network. It has been proposed that proteins located in the periphery of their respective functional networks have elevated propensity to duplicate and undergo positive selection [31, 32]. This happens because poorly connected and loose clusters are able to more efficiently accommodate evolutionary novelties such as gene duplications, deletions, and changes of function, and thus they become the "evolutionary motors" of their biological networks [33–35]. D'Antonio and Ciccarelli have recently demonstrated evidences supporting this assertion [36]. In their paper, these authors have thoroughly analyzed network properties, sequences, and orthology data from *E. coli*, yeast, fly, and human. They observed that genes acquired during evolution encode less connected and less central proteins that are subject to more duplication events. Conversely, it has been observed in other types of signal transducing cascades that the receptors are more constrained than the intermediate elements of their networks [37–40]. These studies, however, were made with pathways such as those of insulin/TOR, which integrate information from inside the organism. As corporeal composition remained relatively the same throughout vertebrate evolution, intra- and extracellular components are not subject to radical variation, making necessary that internal signal transducing cascades must be more tightly constrained in order to consistently maintain their behavior. The environment, however, is constantly subject to changes, and the chemoreceptors cannot be too tightly constrained in order to accommodate these fluctuations. Our data support that CR are a special case of signal transducing pathways that have loosely constrained receptors.

Our subsequent insight into CR evolution was made when comparing the receptors with themselves. We have observed striking evidences suggesting that the vomeronasal receptors are less constrained than the other CR families. First, their plasticity is significantly higher than the other CR, suggesting that this gene family was probably more subject to duplications and deletions than the other receptors. This is further supported by their phylogenetic tree, which is grouped by inparalogs rather than orthologs, suggesting that these genes have arisen from recent duplications and, therefore, are probably less constrained. This finding is similar to what Grus and Zhang observed when studying the dynamics of vomeronasal and olfactory receptors in vertebrate species [41]. Lastly, we have observed that their functional network is completely detached from the other receptors, making them the most peripheral CR. From an evolutionary point of view, one would be tempted to think that the VN code the least important CR in terms of individual survival. The T2R are responsible for detection of bitter tastes. In general, these tastes are typically associated with toxic nitrogenated compounds, such as alkaloids and amines [42]. The perception of these toxins is a major issue in the survival of any organism that has chances of ingesting them. Equally important to their survival is the detection of food, predators, and members of the same species by OR. Conversely, the VN genes likely give clues about potential reproductive partners by detecting genetic likeness and even

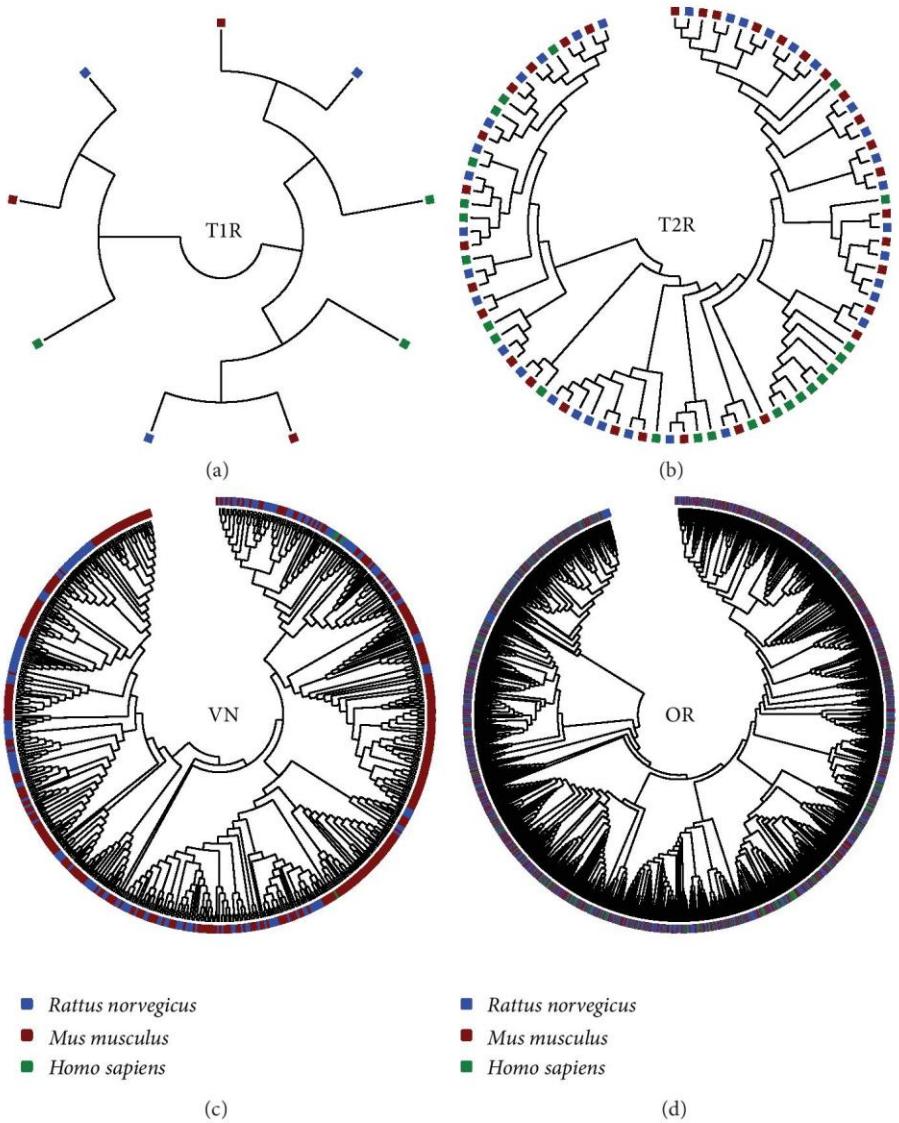

FIGURE 6: Reconstructed phylogenetic tree of the chemoreceptor families. Each square represents a CR gene. Blue, red, and green squares represent *Rattus norvegicus*, *Mus musculus*, and *Homo sapiens* genes, respectively. Phylogenetic trees were reconstructed with Tamura-Nei model. T1R: type I taste receptors; T2R: type II taste receptors; VN: vomeronasal receptors.

immune compatibility [43, 44]. These characteristics, albeit very important to long term adaptation and survival of the species as a whole, are not a major issue in direct survival of the individual.

An apparent contradiction in our analysis was the case of the T1R, which code sweet and *umami* receptors. From our phylogenetic analysis, these receptors are tightly constrained. All three species have the same number of these receptors, each branching more closely with its orthologs in other species rather than the others of the same species. This finding is supported by an earlier analysis that found the same pattern in all vertebrate species [45]. However, by their ortholog distribution, we have found high EPI values. These receptors are grouped in the KOG1056 group, which encompasses 1790 proteins in 52 species, with most varied functions (e.g., bride of sevenless, a homeotic gene). The high-plasticity values of T1R family are owed to the comprehensive reach of this orthologous group, suggesting that these receptors are constrained members of a larger and more dynamic family of proteins. Albeit instigating, these assumptions can only be confirmed with further in-depth study of this interesting orthologous group.

From a systems perspective, we have found evidences that the CR evolved through duplication events that resulted in gain of function. We have observed that all CR families share the same STM cluster, suggesting that the latter is an older transducing core that was reused in several cell types. The CR, on the other hand, are specific and only expressed in their appropriate cell type. The only CR family that diverges fromz this behavior is the vomeronasal receptors, which were adapted to convey their signal directly to an ion-channel. This

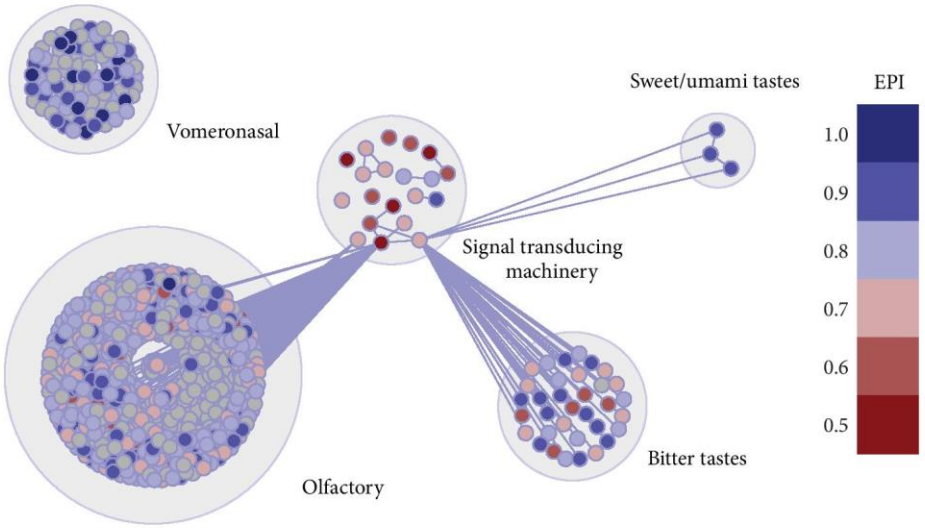

FIGURE 7: Graphical representation of *Mus musculus* chemosensory network. EPI values are plotted on each node by a color scale. Higher plasticity is indicated by bluish colors and lower plasticity by reddish colors. The other networks are not shown in this paper. Nodes represent protein coding genes and edges, functional interactions.

deviation may be the reason why these receptors are under different evolutionary constraints.

Our results suggest that genes coding chemoreceptors were subject to more variation in the course of evolution than those coding signal transducing machinery, reflecting their distinct functional roles in organisms. We have also found significant variation even among the different receptor modalities, suggesting, for the first time to our notice, that olfactory and bitter taste receptors are, albeit less constrained than the transduction machinery, more conserved than vomeronasal receptors. These differences are due to the distinct ecological roles played by the receptors, with the low-plasticity olfactory and bitter taste receptors taking major part in direct survival of the organism, whereas high-plasticity vomeronasal receptors contribute to overall adaptation of the species. Sweet/*umami* receptors cannot be analyzed by their orthologous distribution alone due to the large variability of their ortholog group, and further studies are needed in order to understand the selective pressures imposed on them. We believe that the chemoreceptor networks case is illustrative to demonstrate the generation of novelties through evolutionary tinkering. During the course of evolution, the chemosensory cells generated novel receptor clusters probably by duplicating older ones in order to perceive different sensory inputs. Even among these clusters, there is a great deal of evolutionary experimentation, so that the organisms can be kept up to date with their environment. The signal transduction machinery and other information pathways, however, remained essentially the same throughout generations.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] E. Adler, M. A. Hoon, K. L. Mueller, J. Chandrashekar, N. J. P. Ryba, and C. S. Zuker, "A novel family of mammalian taste receptors," *Cell*, vol. 100, no. 6, pp. 693–702, 2000.

[2] G. Nelson, M. A. Hoon, J. Chandrashekar, Y. Zhang, N. J. P. Ryba, and C. S. Zuker, "Mammalian sweet taste receptors," *Cell*, vol. 106, no. 3, pp. 381–390, 2001.

[3] I. Rodriguez, K. Del Punta, A. Rothman, T. Ishii, and P. Mombaerts, "Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors," *Nature Neuroscience*, vol. 5, no. 2, pp. 134–140, 2002.

[4] J. M. Young, M. Kambere, B. J. Trask, and R. P. Lane, "Divergent V1R repertoires in five species: amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs," *Genome Research*, vol. 15, no. 4, pp. 231–240, 2005.

[5] Y. Hashiguchi and M. Nishida, "Evolution of trace amine-associated receptor (TAAR) gene family in vertebrates: Lineage-specific expansions and degradations of a second class of vertebrate chemosensory receptors expressed in the olfactory epithelium," *Molecular Biology and Evolution*, vol. 24, no. 9, pp. 2099–2107, 2007.

[6] L. Buck and R. Axel, "A novel multigene family may encode odorant receptors: A molecular basis for odor recognition," *Cell*, vol. 65, no. 1, pp. 175–187, 1991.

[7] J. M. Young, H. F. Massa, L. Hsu, and B. J. Trask, "Extreme variability among mammalian V1R gene families," *Genome Research*, vol. 20, no. 1, pp. 10–18, 2010.

[8] P. Shi and J. Zhang, "Extraordinary diversity of chemosensory receptor gene repertoires among vertebrates," *Results and Problems in Cell Differentiation*, vol. 47, pp. 1–23, 2009.

[9] Y. Niimura, "Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents," *Human genomics*, vol. 4, no. 2, pp. 107–118, 2009.

[10] M. Nei, Y. Niimura, and M. Nozawa, "The evolution of animal chemosensory receptor gene repertoires: Roles of chance and

necessity," *Nature Reviews Genetics*, vol. 9, no. 12, pp. 951–963, 2008.

[11] Y. Niimura and M. Nei, "Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates," *Journal of Human Genetics*, vol. 51, no. 6, pp. 505–517, 2006.

[12] J. M. Eirín-López, L. Rebordinos, A. P. Rooney, and J. Rozas, "The birth-and-death evolution of multigene families revisited," *Genome Dynamics*, vol. 7, pp. 170–196, 2012.

[13] D. Dong, G. He, S. Zhang, and Z. Zhang, "Evolution of olfactory receptor genes in primates dominated by birth-and-death process," *Genome Biology and Evolution*, vol. 1, pp. 258–264, 2009.

[14] Y. Niimura, "Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents," *Human genomics*, vol. 4, no. 2, pp. 107–118, 2009.

[15] M. Ashburner, C. A. Ball, J. A. Blake et al., "Gene ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.

[16] L. J. Jensen, M. Kuhn, M. Stark et al., "STRING 8—A global view on proteins and their functional interactions in 630 organisms," *Nucleic Acids Research*, vol. 37, no. 1, pp. D412–D416, 2009.

[17] R. L. Seal, S. M. Gordon, M. J. Lush, M. W. Wright, and E. A. Bruford, "Genenames.org: The HGNC resources in 2011," *Nucleic Acids Research*, vol. 39, no. 1, pp. D514–D519, 2011.

[18] J. A. Blake, C. J. Bult, J. A. Kadin, J. E. Richardson, and J. T. Eppig, "The mouse genome database (MGD): Premier model organism resource for mammalian genomics and genetics," *Nucleic Acids Research*, vol. 39, no. 1, pp. D842–D848, 2011.

[19] M. R. Dwinell, E. A. Worthey, M. Shimoyama et al., "The rat genome database 2009: variation, ontologies and pathways," *Nucleic Acids Research*, vol. 37, no. 1, pp. D744–D749, 2009.

[20] S. Haider, B. Ballester, D. Smedley, J. Zhang, P. Rice, and A. Kasprzyk, "BioMart central portal: unified access to biological data," *Nucleic Acids Research*, vol. 37, no. 2, pp. W23–W27, 2009.

[21] R. J. S. Dalmolin, M. A. A. Castro, J. L. Rybarczyk Filho, L. H. T. Souza, R. M. C. de Almeida, and J. C. F. Moreira, "Evolutionary plasticity determination by orthologous groups distribution," *Biology Direct*, vol. 6, article 22, 2011.

[22] M. A. A. Castro, X. Wang, M. N. C. Fletcher, K. B. Meyer, and F. Markowetz, "RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations," *Genome Biology*, vol. 13, article R29, 2012.

[23] D. Dong, K. Jin, X. Wu, and Y. Zhong, "CRDB: Database of chemosensory receptor gene families in vertebrate," *PLoS ONE*, vol. 7, no. 2, Article ID e31540, 2012.

[24] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.

[25] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.

[26] R. C. Edgar, "MUSCLE: a multiple sequence alignment method with reduced time and space complexity," *BMC Bioinformatics*, vol. 5, article 113, 2004.

[27] K. Tamura and M. Nei, "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees," *Molecular Biology and Evolution*, vol. 10, no. 3, pp. 512–526, 1993.

[28] W. T. Swaney and E. B. Keverne, "The evolution of pheromonal communication," *Behavioural Brain Research*, vol. 200, no. 2, pp. 239–247, 2009.

[29] B. Nilius and G. Owsianik, "The transient receptor potential family of ion channels," *Genome Biology*, vol. 12, no. 3, article 218, 2011.

[30] T. Hartmann, "From waste products to ecochemicals: fifty years research of plant secondary metabolism," *Phytochemistry*, vol. 68, no. 22-24, pp. 2831–2846, 2007.

[31] P. M. Kim, J. O. Korbel, and M. B. Gerstein, "Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 51, pp. 20274–20279, 2007.

[32] T. Yamada and P. Bork, "Evolution of biomolecular networks lessons from metabolic and protein interactions," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 11, pp. 791–803, 2009.

[33] X. Zhu, M. Gerstein, and M. Snyder, "Getting connected: Analysis and principles of biological networks," *Genes and Development*, vol. 21, no. 9, pp. 1010–1024, 2007.

[34] R. Albert, "Scale-free networks in cell biology," *Journal of Cell Science*, vol. 118, no. 21, pp. 4947–4957, 2005.

[35] Y. I. Wolfa, P. S. Novichkovb, G. P. Kareva, E. V. Koonina, and D. J. Lipmana, "The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 18, pp. 7273–7280, 2009.

[36] M. D'Antonio and F. D. Ciccarelli, "Modification of gene duplicability during the evolution of protein interaction network," *PLoS Computational Biology*, vol. 7, no. 4, Article ID e1002029, 2011.

[37] D. Alvarez-Ponce, M. Aguadé, and J. Rozas, "Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 Drosophila genomes," *Genome Research*, vol. 19, no. 2, pp. 234–242, 2009.

[38] X. Wu, X. Chi, P. Wang, D. Zheng, R. Ding, and Y. Li, "The evolutionary rate variation among genes of HOG-signaling pathway in yeast genomes," *Biology Direct*, vol. 5, article 46, 2010.

[39] R. Jovelin and P. C. Phillips, "Expression level drives the pattern of selective constraints along the insulin/tor signal transduction pathway in caenorhabditis," *Genome Biology and Evolution*, vol. 3, no. 1, pp. 715–722, 2011.

[40] D. Alvarez-Ponce, M. Aguadé, and J. Rozas, "Comparative genomics of the vertebrate insulin/TOR signal transduction pathway: A network-level analysis of selective pressures," *Genome Biology and Evolution*, vol. 3, no. 1, pp. 87–101, 2011.

[41] W. E. Grus and J. Zhang, "Distinct evolutionary patterns between chemoreceptors of 2 vertebrate olfactory systems and the differential tuning hypothesis," *Molecular Biology and Evolution*, vol. 25, no. 8, pp. 1593–1601, 2008.

[42] J. I. Glendinning, "Is the bitter rejection response always adaptive?" *Physiology and Behavior*, vol. 56, no. 6, pp. 1217–1227, 1994.

[43] P. A. Brennan and K. M. Kendrick, "Mammalian social odours: Attraction and individual recognition," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, no. 1476, pp. 2061–2078, 2006.

[44] P. Chamero, T. Leinders-Zufall, and F. Zufall, "From genes to social communication: molecular sensing by the vomeronasal organ," *Trends in Neurosciences*, vol. 35, no. 10, pp. 597–606, 2012.

[45] P. Shi and J. Zhang, "Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes," *Molecular Biology and Evolution*, vol. 23, no. 2, pp. 292–300, 2006.