

Construção de um plasmídeo para *knockdown* do gene *TCF4*, ligado a um tipo sindrômico e monogênico de autismo

Palavras-Chave: Autismo, Biologia Molecular, Genética, Manipulação da expressão

Autores(as):

Nícolás Gasparine Figueiredo, Instituto de Biologia – UNICAMP

Prof. Dr. Fábio Papes, Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas - UNICAMP

Prof. Dr^a. Katlin B. Massirer, Centro de Química Medicinal (CQMED) - UNICAMP

INTRODUÇÃO:

Os Transtornos do Espectro Autista são um conjunto de distúrbios do neurodesenvolvimento agrupados em uma denominação clínica comum conhecida como “autismo”¹. A grande maioria dos casos de Transtorno do Espectro Autista (TEA) são idiopáticos, ou seja, não apresentam etiologia esclarecida (sem bases genéticas definidas), sendo em sua maioria fruto de uma ampla gama de interações entre fatores de risco genético^{2,3}. Dessa forma, a grande variabilidade genotípica dos indivíduos dificulta o estudo de mecanismos subjacentes relacionados às alterações no neurodesenvolvimento. Contudo, ainda existem raros casos monogênicos de autismo que permitem uma análise mais direcionada para a elucidação dos fenótipos neurobiológicos.

Um desses casos raros de TEA com bases genéticas bem determinadas é a Síndrome de Pitt-Hopkins (PTHS-MIM #610954) um tipo monogênico e sindrômico de autismo. A PTHS é uma doença autossômica dominante, causada geralmente por mutações *de novo* em alelo do gene *TCF4*^{4,5}. Pacientes PTHS apresentam alterações significativas no desenvolvimento, marcadas por deficiência intelectual e na linguagem. Os portadores também apresentam disfunções marcantes no sistema nervoso autônomo, como disfunções respiratórias e descompasso da motilidade intestinal, além de poderem manifestar atrasos motores debilitantes⁴.

O gene *TCF4*, localizado no *locus* 18q21.2, codifica um fator de transcrição bHLH que se liga ao DNA por meio de um domínio E-box identificando sequência consenso CANNTG. Esse fator de transcrição está relacionado ao processo de diferenciação neuronal, além de ser relacionado com outros distúrbios neuropsiquiátricos como a esquizofrenia⁵. Estudos recentes publicados pelo nosso time, demonstraram que a perda de função do gene *TCF4* associada a expressão desse fator estão associados com os fenótipos aberrantes do neurodesenvolvimento observados em PTHS, sendo o mecanismo a haploinsuficiência um dos principais mecanismos para a manifestação das alterações neurobiológicas significativas⁵. Dessa forma, a expressão diferencial de *TCF4* funcional parece impactar diretamente processos do neurodesenvolvimento.

Frente a essa realidade, o controle da expressão do *TCF4* se mostra uma ferramenta experimental vantajosa para estudos neurobiológicos porque permite elucidar mecanismos correlacionados com a expressão do gene. Portanto, a manipulação da expressão deste fator de transcrição em modelos experimentais é crucial para investigar os mecanismos patológicos associados à PTHS e ao TEA. O presente estudo teve como objetivo fornecer dados para consolidar uma estratégia de manipulação da expressão do gene *TCF4* em modelos *in vitro*, por meio do desenvolvimento de vetores lentivirais para realizar o *knockdown* do gene *TCF4* via shRNA em células progenitoras neurais (NPCs), com o objetivo de simular os efeitos da haploinsuficiência do gene *in vivo* e elucidar mecanismos associados.

METODOLOGIA:

O método de lentiviral para *knockdown* via shRNA é um processo que tem como base a transdução de células eucarióticas para adição de uma sequência *trans*-reguladora via ação de siRNA. Nosso método envolve um vetor com um cassete de expressão que inclui o promotor hU6 para recrutar a RNA polimerase III, um gene shRNA, um cPPT para transporte nuclear, além de promotor e gene de seleção (Figura 1). As LTRs flanqueiam o cassete, facilitando a integração no genoma eucariótico. Partículas lentivirais são produzidas pela transfecção do vetor shRNA e dos plasmídeos psPAX e pMD2.G em células HEK293. As NPCs são transduzidas pelo lentivírus, inserindo a sequência de interesse entre as LTRs no genoma. Isso permite a introdução de um repressor *trans* via siRNA para o transcrito TCF4-201. Os vetores contêm genes de resistência à puromicina (PuroR) para selecionar células transduzidas. A expressão do gene shRNA forma um siRNA que, pela via Drosha/Dicer, se torna um siRNA, o qual se liga ao mRNA alvo TCF4-201 via AGO/RISC, causando degradação ou repressão da tradução. Esse mecanismo promove o *knockdown* do gene TCF4, cuja eficiência real deve ser medida empiricamente. O sistema de vetor lentiviral utilizado por esse estudo foi o pLKO.1, específico para esse tipo de controle da expressão.

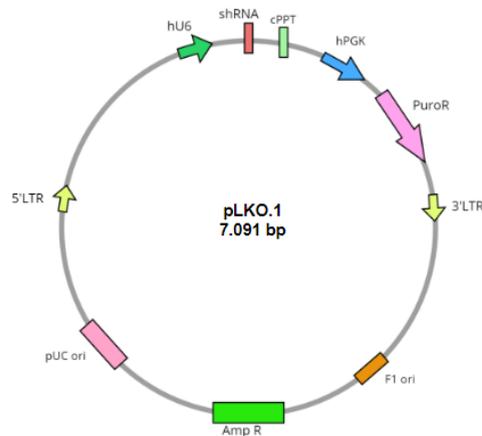


Figura 1 - Representação do plasmídeo pLKO 1-TRC cloning vector (Plasmid #10878) - Observa-se o vetor utilizado para a clonagem do gene shRNA para *knockdown* do transcrito TCF4-201. Pode-se observar um promotor U6, conhecido por sua confiabilidade na expressão de pequenos RNAs nucleares não codificantes. Apresenta o hPGK, que é responsável pelo gene puroR. Possui duas origens de replicação: uma pUC ori e uma F1 ori. Inclui o gene de resistência à ampicilina (AmpR). O trato central de polipurina (cPPT) aumenta a eficiência da transdução ao facilitar a importação nuclear. Observa-se também as extremidades 5'LTR e 3'LTR auto-inativadores.

Para que seja possível realizar o *knockdown* via shRNA de modo efetivo, é necessário selecionar alvos no transcrito seguindo a indicação de softwares específicos para esse método. O gene *TCF4* apresenta 41 exons e por meio do processo de *splicing* e influência de promotores alternativos, sendo a isoforma TCF4-B a mais expressa nos tecidos e células neurais⁶. A sequência dos transcritos desse gene foram obtidas por meio do banco de dados do NCBI (<https://www.ncbi.nlm.nih.gov>) por meio da entrada para o gene *TCF4* (Gene ID:6925). O transcrito TCF4-201 (isoforma a; TCF4-B+; NCBI: NM_001083962.2) foi utilizado como base para a predição dos alvos e análises *in silico*. Fizemos a análise da predição dos alvos utilizando dois algoritmos de predição, a ferramenta **Design Hairpins** do **Genetic Perturbation Platform** (GPP: <https://portals.broadinstitute.org/gpp/public/>) e o **siDirect** (<https://sidirect2.rnai.jp/>).

O algoritmo da função **Design Hairpins** do **GPP** contava com regras específicas já estabelecidas previamente pelo algoritmo conhecidas como "Rule set 9". Candidatos são excluídos se começarem com "AA" ou contiverem uma sequência de quatro bases idênticas consecutivas. Além disso, são desconsiderados aqueles com percentual de GC extremo ($\leq 25\%$ ou $> 60\%$), enquanto candidatos com GC entre 25-55% são promovidos, e aqueles entre 55-60% recebem uma pontuação neutra. Qualquer sequência contendo bases ambíguas, como "N", ou com sítios de restrição é excluída da seleção de alvos. Sequências com sete bases consecutivas de C/G também são eliminadas. Candidatos capazes de formar estruturas de haste-alça interna, com comprimento mínimo de haste de 5 e alça de 4, são penalizados. Por fim, favorece-se candidatos com pareamento de bases mais fraco na região 3' terminal (posições 15-20), sendo a pontuação máxima para aquelas totalmente compostas por A ou T, reduzindo-se rapidamente se houver predominância de G/C.

O **siDirect** permite a personalização dos parâmetros e selecionamos eles conforme descrito na literatura para desenho de shRNA e o protocolo da Addgene(<https://www.addgene.org/protocols/plko/>) relativo ao sistema pLKO.1, utilizado como vetor lentiviral para transdução de células humanas e de mamíferos^{7,8}. Com esse algoritmo editamos o parâmetro para um conteúdo GC entre 30% e 60%, excluimos qualquer sequência-alvo com mais de 4 nucleotídeos T em *tandem*, visto que funciona como sinal de parada para RNA Pol III. Além disso, o siDirect apresenta três métodos diferenciais de seleção de alvos. O método Ui-Tei *et.al* que prefere alvos iniciados em GC, com as 7 últimas bases próximas à extremidade 3' com altas frequências de AU⁹. O método Reynolds *et.al* que delimita a Tm dos alvos em menos de 20°C e enriquece os alvos com AU nas cinco últimas bases, além de selecionar candidatos com um U na base em posição mediana da sequência¹⁰. E por último, o método Amarzguioui *et.al* que é o menos restritivo apenas relutando excluindo candidatos U e G nas extremidades finais¹¹.

Dessa forma, conseguimos prever alvos no transcrito TCF4-201, observando quais exons e posições da sequência apresentam a maior frequência de alvos preditos, além de demonstrar quais regiões contém mais alvos e aquelas que não são tão vantajosas para predição de alvos. Além disso, foram selecionados alvos para darem origem aos oligos shRNA que passaram pelo processo de clonagem. Os oligos foram desenhados seguindo os protocolos do sistema pLKO.1 para formação de heteroduplex-inserto competente para a clonagem e formação dos vetores. Os clones gerados foram sequenciados para confirmação do inserto via sequenciamento Sanger.

RESULTADOS E DISCUSSÃO:

O transcrito TCF4-201 (isoforma a; TCF4-B+; NCBI: NM_001083962.2) é resultado de um processo de *splicing* com 20 exons, tendo um tamanho total de 8.041 bases com o exon 20 sendo o maior, com 5846 bases, ou seja, em torno de 72,2% das bases. Dentro do banco de dados do NCBI existem 46 transcritos anotados e revisados para o gene *TCF4* (Gene ID:6925). Analisamos o transcrito por meio ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) observamos que a ORF que traduz-se na isoforma a (TCF4-B+) nesse transcrito se estende entre a posição 176 até a 2191. Referente às análises de previsão de alvos no transcrito, observamos que a ferramenta do GPP selecionou com maior frequência alvos candidatos no exon 20, especificamente no intervalo da sequência entre a posição 4401 e 4500. Além disso, observamos que o algoritmo da GPP não selecionou alvos candidatos nos exons 4, 5, 6, 12, 13, 16 e 18 do transcrito (Figura 2.a). As previsões com o siDirect foram diferentes, tendo uma preferência geral, os três métodos, pelo exon 17 do transcrito (Figura 2.b-d).

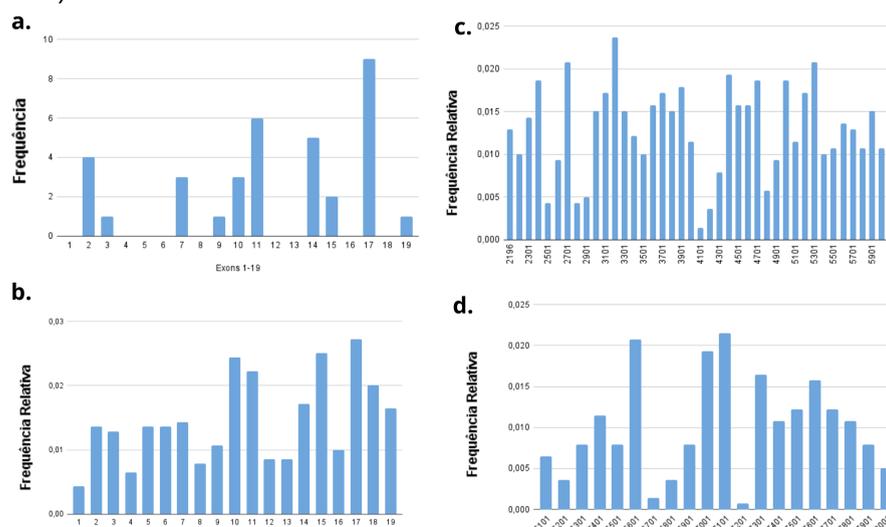


Figura 2 - Frequência de alvos encontrados ao longo da sequência do transcrito TCF4-201. (a) Gráfico representa o número de candidatos em cada um dos 19 primeiros exons do transcrito seguindo o algoritmo da ferramenta para alvos shRNA do GPP. (b) Frequência relativa dos alvos encontrados nos 19 primeiros exons do transcrito segundo a ferramenta siDirect levando em conta os três métodos de seleção de candidatos disponíveis. (c-d) São os alvos presentes no exon 20, subdividido em intervalos de aproximadamente 100 bases, sendo os números marcados na horizontal a primeira posição de cada intervalo.

Escolhemos três alvos e desenhamos oligos para a formação de um inserto shRNA para o sistema pLKO.1. Dois alvos têm sequências alvos no exon 17 e um alvo tem sua sequência como parte do exon 18 (Tabela 1). Acreditamos que os três são sequências alvos competentes, uma vez que todos apresentam 3 ou mais malpareamentos com sequências não correlacionadas, reduzindo o efeito off-target do sistema, além de estarem de acordo com todos os outros requisitos do sistema, tendo %GC entre 30%-60%, estão posicionados a mais de 25 bp do códon ATG da fase de leitura da ORF, não apresentam nucleotídeos T em *tandem* e extremidade 3' apresentasse baixa probabilidade de formação de pontes intracadeia pela presença de T ou A entre as posições 15-19.

Alvos	Sequências	bp	%C-G	(%)	Posição TCF4-B+	Exons
shV 1	CAC GAA ATC TTC GAA GGA CAA	21	0,48	100	1743-1765	17
shV 4	GAA AGG AAT CTG AAT CCG AAA	21	0,38	100	2045-2067	18
shV 6	TAG GTC AAG ATC ATG CAA TAA	21	0,33	100	1803-1825	17
(-)	CAA CAA GAT GAA GAG CAC CAA	21	0,43	61,9	N/A	N/A

Tabela 1 - Alvos (21-mers) escolhidos e previstos para *knockdown* via shRNA.

Dentre os vetores clonados com os alvos escolhidos, foi possível sequenciar e obter um clone shV4 positivo para a presença correta do inserto e que será utilizado em experimentos para avaliação de sua eficácia em culturas *in vitro* de NPCs (Figura 3).

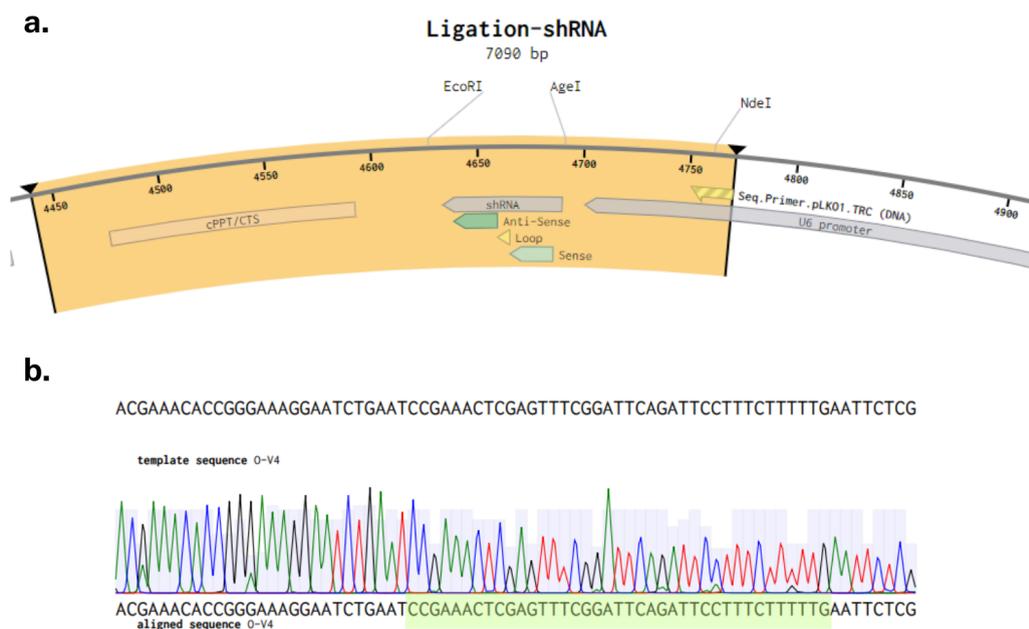


Figura 3 - Resultados do sequenciamento Sanger voltado para a validação do vetor shV4 que tem como alvo uma sequência do exon 18. (a) Região alvo do primer de sequenciamento selecionado para anelar com uma região terminal do promotor hU6. (b) Eletroferograma como resultado do sequenciamento do clone 0-V4, um clone que apresenta o um vetor com inserto shRNA positivo na posição correta downstream ao promotor.

CONCLUSÕES:

Os resultados deste estudo mostram a criação de um vetor plasmidial para o *knockdown* do gene *TCF4*, demonstrando-se um ferramenta experimental que poderá ser muito vantajosa para o estudo do papel da expressão do *TCF4* em modelos *in vitro*. Além disso, foi possível observar o comportamento de duas importantes de predição de alvos para sistemas siRNA/shRNA frente ao transcrito *TCF4-201*. O clone shV4 está armazenado no laboratório e será usado em experimentos para o *knockdown* do gene *TCF4*, visando a posição 2045-2067 do transcrito *TCF4-201* (Figura 04).

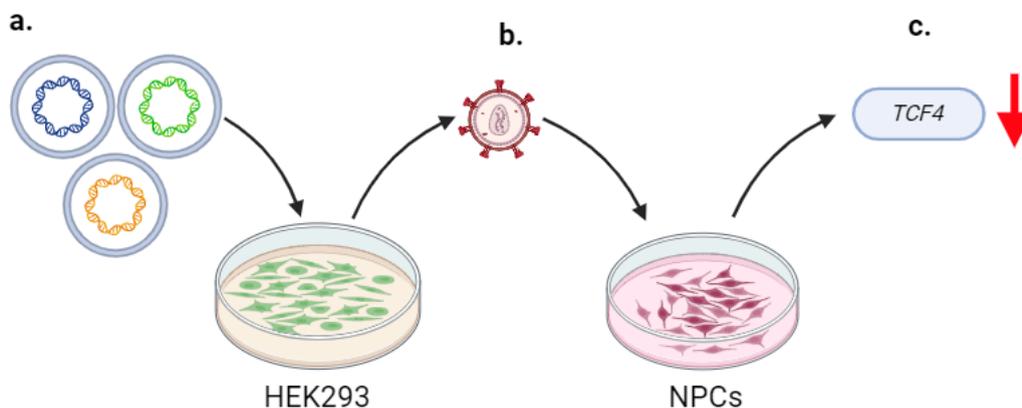


Figura 04 - Fluxo de trabalho para produção lentiviral usando o clone shV4 engenhado no estudo - (a) Os plasmídeos shV4, pMD2.G e psPAX.2 são transfectados em células HEK293. (b) A produção de partículas lentivirais permite a transdução das NPCs. (c) Após a transdução e seleção com puromicina, busca-se o *knockdown* da expressão de *TCF4*.

BIBLIOGRAFIA

1. JOHNSON, C. P. et al. Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 120, 1183–1215 (2007)
2. GROVE, Jakob et al. Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics*, v. 51, n. 3, p. 431-444, 2019
3. LORD, Catherine et al. Autism spectrum disorder. *The lancet*, v. 392, n. 10146, p. 508-520, 2018.
4. ZOLLINO, Marcella et al. Diagnosis and management in Pitt-Hopkins syndrome: First international consensus statement. *Clinical Genetics*, v. 95, n. 4, p. 462-478, 2019.
5. PAPES, Fabio et al. Transcription Factor 4 loss-of-function is associated with deficits in progenitor proliferation and cortical neuron content. *Nature Communications*, v. 13, n. 1, p. 2387, 2022.
6. SEPP, Mari et al. Functional diversity of human basic helix-loop-helix transcription factor *TCF4* isoforms generated by alternative 5' exon usage and splicing. *PloS one*, v. 6, n. 7, p. e22138, 2011.
7. KHVOROVA, Anastasia; REYNOLDS, Angela; JAYASENA, Sumedha D. Functional siRNAs and miRNAs exhibit strand bias. *Cell*, v. 115, n. 2, p. 209-216, 2003.
8. MOFFAT, Jason et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*, v. 124, n. 6, p. 1283-1298, 2006.
9. UI-TEI, Kumiko et al. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic acids research*, v. 32, n. 3, p. 936-948, 2004.
10. REYNOLDS, Angela et al. Rational siRNA design for RNA interference. *Nature biotechnology*, v. 22, n. 3, p. 326-330, 2004.
11. AMARZGUIOUI, Mohammed; PRYDZ, Hans. An algorithm for selection of functional siRNA sequences. *Biochemical and biophysical research communications*, v. 316, n. 4, p. 1050-1058, 2004.