UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

DEPARTAMENTO DE GENÉTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

AVALIAÇÃO DO GENE *HAND2* E SELEÇÃO DE NOVOS GENES ALVOS NA SUSCETIBILIDADE GENÉTICA À EMBRIOPATIA DA TALIDOMIDA

Bruna Duarte Rengel

Orientadora: Profa. Dra. Fernanda Sales Luiz Vianna

Coorientador: Prof. Dr. Lucas Rosa Fraga

Porto Alegre, junho de 2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

DEPARTAMENTO DE GENÉTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

AVALIAÇÃO DO GENE *HAND2* E SELEÇÃO DE NOVOS GENES ALVOS NA SUSCETIBILIDADE GENÉTICA À EMBRIOPATIA DA TALIDOMIDA

Bruna Duarte Rengel

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para obtenção do título de **Mestre em Genética e Biologia Molecular.**

Orientadora: Profa. Dra. Fernanda Sales Luiz Vianna

Coorientador: Prof. Dr. Lucas Rosa Fraga

Porto Alegre, junho de 2021

Esse trabalho foi realizado no Laboratório de Genética Populacional e Evolução (Laboratório 113) do Departamento de Genética da UFRGS e no Laboratório de Medicina Genômica no Hospital de Clínicas de Porto Alegre, com apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). A aluna recebeu bolsa de estudos concedida pelo CNPq, vinculada ao Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM/UFRGS).

Agradecimentos

Em primeiro lugar gostaria de agradecer de todo o coração à minha orientadora Prof. Dra. Fernanda e ao meu coorientador, que considero também como meu orientador, o Prof. Dr. Lucas. Não poderia ter pedido melhores orientadores. Vocês foram imprescindíveis para o meu crescimento, tanto pessoal quanto profissional. Me auxiliaram da melhor forma durante o curso do mestrado. Tiveram paciência e entendimento com minhas dificuldades. Agradeço muito por ter os dois como meus mentores.

Agradeço também à Thayne, que foi minha salvadora em diversos momentos durante o mestrado. Manteve-se aberta para conversas e dúvidas e sempre disposta a ajudar. Muito obrigada também à todas as oportunidades que têm me proporcionado. Se não fosse por ti, que me apresentou o laboratório e a Fernanda, talvez hoje eu não estivesse aqui.

Agradeço ao João Matheus, que foi um grande amigo e colega de laboratório durante esses dois anos. Sempre me auxiliando e lembrando-me do meu valor. Que a distância não apague nossa amizade e que possamos continuar sendo sempre próximos.

Obrigada ao lab 133 que mesmo longe se manteve presente durante esse período de pandemia. E ao laboratório GBRD, que também se manteve presente. Não poderia escolher dois grupos melhores para trabalhar. Obrigada por tudo!

Agradeço a todos os professores e colegas que de alguma forma fizeram parte e auxiliaram nesta jornada. Ao Elmo e ao Gabriel, sempre dispostos a ajudar e respostas as dúvidas. Foi um tempo atípico devido à pandemia, mas certamente sem a colaboração deles teria sido muito mais difícil.

Obrigada ao CNPq pelo auxílio financeiro.

À Débora, minha amiga da vida, que durante oito meses foi minha colega de trabalho, compartilhando nossas inseguranças e dificuldades com o mestrado e com a vida.

Obrigada aos meus pais que permitiram que eu tivesse meus estudos e trabalho sem nenhuma dificuldade. Sempre dispostos a ajudar. A minha mãe que diversas vezes me colocou em seu colo nos momentos de estresse. E também a minha irmã, que me fazia rir mesmo durante os momentos de estresses. Amo vocês família.

Ao meu parceiro de vida, Mayk, que mesmo sem nenhuma experiência na área leu minha dissertação, ouviu todas as minhas preocupações, explicações e se manteve forte em todos os momentos as quais eu estive fraca. Fez comida para mim quando não tive tempo de cozinhar, que lembrou que eu precisava dar uma espairecida e que cuidou de mim quando estive hospitalizada por covid. Te amo.

Agradeço imensamente à minha psicóloga, Cris, que me manteve sã durante todo esse tempo, auxiliando-me na aceitação dos meus problemas, a lidar com a vida para que eu pudesse continuar trabalhando na minha dissertação.

E finalmente, agradeço à Deus e aos meus mentores espirituais, pai Seta Branca, Mãe Yara e Vozinha Marilu, que através de muita prece me auxiliaram na calma e força para vencer todas as dificuldades que uma pandemia e problemas de saúde puderam trazer.

Sumário

LISTA DE ABREVIATURAS 7
RESUMO 10
ABSTRACT 11
CAPÍTULO I 12
1. INTRODUÇÃO 13
1.1 TERATOGÊNESE 13
1.2 TALIDOMIDA 14
1.2.1 Histórico 14
1.2.2 Talidomida no Brasil 17
1.2.3 Embriopatia da talidomida 18
1.2.4 Mecanismos de teratogênese 21
1.2.5 Suscetibilidade genética à Embriopatia da Talidomida 25
1.3 TBX5 27
1.4. HAND2 29
1.4.1 HAND2, TBX5 e talidomida 32
CAPÍTULO II 34
2. JUSTIFICATIVA 35
3. OBJETIVOS 36
3.1. Objetivo geral 36
3.2 Objetivos específicos 36
CAPÍTULO III 37
Genetic Evaluation of *HAND2* gene and its Effects on Thalidomide Embryopathy 38
CAPÍTULO IV 60
Possible new targets of Thalidomide related to limbs and cardiac defects: a systems biology approach 61
CAPÍTULO V 116
4. CONSIDERAÇÕES FINAIS 117
REFERÊNCIAS BIBLIOGRÁFICAS 125
ANEXO I 137

LISTA DE ABREVIATURAS

AIDS Síndrome da imunodeficiência adquirida

ANVISA Agência Nacional de Vigilância Sanitária

BCL6 *BCL6 Transcription Repressor* – repressor de transcrição BCL6

bHLH *basic Helix-loop-helix* – hélice-alça-hélice básico

BMP4 *Bone morphogenic protein 4* – proteína morfogênica óssea 4

CBP *CREBB binding protein* – proteína de ligação à CREBB

CRBN *Cereblon* - cereblon

CUL4A *Cullin 4A* – culina 4A

DDB1 *Damaged DNA Binding Portein 1* - proteína de ligação específica ao DNA danificado 1

DL50 Dose Letal Aguda

ENH Eritema Nodoso Hansênico

eNOS *Endothelial Nitric Oxide Synthase* - Enzima Óxido Nítrico Sintase Endotelial

EP300 *E1A Binding Protein P300* – E1A proteína ligadora P300

ESCO2 *Establishment Of Sister Chromatid Cohesion N-Acetyltransferase 2* – estabelecimento de coesão de cromátide irmã n-acetiltransferase 2

FANCB *FA Complementation Group B* - FA grupo complementação B

FANCM *FA complementation group M* – FA grupo complementação M

FGF *Fibroblast growth fator* - fator de crescimento de fibroblasto

FGF10 *Fibroblast growth factor 10* – fator de crescimento de fibroblasto 10

FGF8 *Fibroblast growth factor 8* – fator de crescimento de fibroblasto 8

GATA4 *GATA Binding Protein 4* - Proteína Ligadora de GATA 4

GLI1 *GLI Family zinc finger 1* - dedo de zinco da família Gli 3

GLI3 *Gli family zinc finger 3* – dedo de zinco da família Gli 3

HAND1 *Heart and Neural Crest Derivatives Expressed 1* – derivados do coração e crista neural expressos 1

HAND2 *Heart and Neural Crest Derivatives Expressed 2* – derivados do coração e crista neural expressos 2

IKZF1 *Ikaros zinc finger family 1* – família Ikaros dedo-de-zinco 1

IKZF3 *Ikaros zinc finger family 3* – família Ikaros dedo-de-zinco 3

IMiDs *Immunomodulatory Drugs* – fármacos imunomoduladores

JUN *Jun Proto-Oncogene* – proto-oncogene Jun

MeCP2 Methyl-CpG Binding Protein 2 – proteína ligadora de metil-CpG 2

MEF2C *Myocyte Enhancer Factor 2C* - Fator Potenciador de Miócito 2C

miRNA *Micro ribonucleic acid* – micro ácido ribonucleico

mRNA *Messenger ribonucleic acid* – ácido ribonucleico mensageiro

NF-KB *Nuclear factor kappa B* – fator nuclear kappa B

NKX2-5 *NK2 homeobox 5* – NK2 homeobox 5

NO *Nitric oxide* - óxido nítrico

NOS3 *Nitric oxide synthase 3* – sintase de óxido nítrico 3

NPPA *Natriuretic Peptide A* - Peptídeo Natriurético A

PRDM16 *PR/SET Domain 16* – PR/SET domínio 16

PTGS2 *Prostaglandin synthase 2* – prostaglandina sintase 2

RDC Resolução da Diretoria Colegiada

ROS *Reactive oxygen species* - Espécies reativas de oxigênio

SALL4 *Spalt Like Transcription Factor 4* - fator de transcrição do tipo *spalt* 4

SHH *Sonic hedgehog*

SRCAP Snf2 Related CREBBP Activator Protein – proteína ativadora Snf2 relacionada CREBBP

TBX18 *T-box 18* – t-box 18

TBX2 *T-box 2* – t-box 2

TBX3 *T-box 3* – t-box 3

TBX5 *T-box 5* – t-box 5

TCF4 *Trancription factor 4* – fator de transcrição 4

TE Embriopatia da talidomida

TP53 *Tumor protein p53* – proteína tumoral p53

TP63 *Tumor protein p63* – proteína tumoral p63

TP73 *Tumor protein p73* – proteína tumoral p73

UBB *Ubiquitin B* – ubiquitina B

VEGF *Vascular endothelium growth factor* – fator de crescimento do endotélio vascular

VEGFA *Vascular endothelium growth factor A* – fator de crescimento do endotélio vascular A

VNTR Repetição em tandem de número variável

XRCC2 *X-Ray Repair Cross Complementing 2* – complementação cruzada de reparo de raio-x 2

ZIC3 *Zic family member 3* – membro da família Zic 3

ZNF *Zinc Finger Protein* – proteína dedo de zinco

RESUMO

A exposição da talidomida durante o desenvolvimento embrionário gera um conjunto de anomalias conhecido como Embriopatia da Talidomida (TE), caracterizado principalmente por malformações no coração e membros. Ainda hoje, os mecanismos de teratogenicidade da talidomida não estão completamente elucidados. Recentemente, o fator de transcrição TBX5 foi descrito como um alvo da talidomida. Além disso, uma interação entre TBX5 e outro fator de transcrição, HAND2, foi observada, sendo inibida na presença de talidomida. HAND2 é importante para o desenvolvimento do coração de membros. Dessa forma, o objetivo deste trabalho foi avaliar o gene *HAND2* e suas associações com a TE e buscar novos alvos para análises de susceptibilidade genética à embriopatia da talidomida. A partir de uma caracterização molecular *in silico, HAND2* se mostrou um gene muito conservado, sendo a região 3'UTR a mais polimórfica do gene. A avaliação de uma variante, p.P51, encontrada no éxon 1 de três pessoas portadoras de TE mostrou que ela possui potencial de afetar ilhas CpG subjacentes. Contudo, análise de metilação em indivíduos com TE não indicou alteração nos padrões de metilação deste gene. Também, análises *in silico* indicaram que p.P51 pode alterar a estrutura do mRNA de *HAND2*. Para a seleção de novos alvos, realizamos análises de redes de interação proteína-proteína da interação entre HAND2 e TBX5 e malformações de coração e membros da TE. Observamos que a proteína EP300 foi um alvo relevante em todas as redes analisadas. Além disso, procurou-se por proteínas com domínio C2H2 que estariam envolvidas na interação de HAND2 e TBX5 e nas malformações de coração e membros da TE. Observamos cinco proteínas com domínio C2H2 nas redes de malformações de coração e de membros da TE. Nós também investigamos genes diferencialmente expressos em células tratadas com talidomida e observamos três genes com expressão diminuída e presente nas redes de malformação de coração e membros da TE; FANC2, ESCO2 e XRCC2. Neste estudo observamos a conservação do gene *HAND2,* além de caracterizar a variante p.51 observada em indivíduos com TE. Finalmente, elencamos novos alvos de interesse para serem futuramente estudados na teratogênese da talidomida.

ABSTRACT

Thalidomide exposure during development generates a set of anomalies known as Thalidomide Embryopathy (TE), characterized mainly by congenital anomalies in the heart and limbs. Even today, thalidomide's teratogenicity mechanisms are not fully understood. Recently, the transcription factor TBX5 was described as a target of thalidomide. Furthermore, an interaction between TBX5 and another transcription factor, HAND2, was observed, being inhibited in the presence of thalidomide. HAND2 is important for the development of the limbs and heart. Thus, this study aimed to evaluate the *HAND2* gene and its associations with TE and to search for new targets for analyzes of genetic susceptibility to TE. From *in silico* molecular characterization, *HAND2* showed to be a very conserved gene, with the 3'UTR being the most polymorphic part. The evaluation of a variant, p.P51, found in exon 1 of three people with TE showed that it has the potential to affect CpG islands near the variant. However, methylation analysis in subjects with TE did not indicate changes in the methylation patterns of this gene. Also, *in silico* analyzes indicated that p.P51 can alter the structure of the HAND2 mRNA. For the selection of new targets, we performed analyzes of protein-protein interaction networks of the interaction between HAND2 and TBX5 and heart and limb malformations of TE. We observed that EP300 was a relevant target in all analyzed networks. In addition, we looked for proteins with a C2H2 domain that would be involved in the interaction of HAND2 and TBX5 and malformations of the heart and members of TE. We observed five proteins with a C2H2 domain in the networks of heart and limbs malformations of TE. We also investigated genes differentially expressed in cells treated with thalidomide and observed three genes with decreased expression and present in the networks of heart and limbs malformation of TE, FANC2, ESCO2, and XRCC2. In this study, we observed the conservation of the HAND2 gene in addition to characterizing a variant p.51 observed in individuals with TE. Finally, we list new targets of interest to be further studied in the thalidomide teratogenesis.

CAPÍTULO I

INTRODUÇÃO

1. INTRODUÇÃO

1.1 TERATOGÊNESE

O conhecimento de que agentes externos poderiam atravessar a placenta e causar danos ao embrião/feto em desenvolvimento surgiu na década de 40 com o estudo de Gregg (1941), o qual observou o desenvolvimento de catarata em embriões/fetos expostos à rubéola. A partir desse estudo e com o avanço da ciência, percebeu-se que um número considerável de agentes tinha a capacidade de atravessar a placenta e levar a algum efeito ao embrião/feto em desenvolvimento, causando uma variedade de anomalias congênitas (Moore and T.V.N. Persaud 2000).

Anomalias congênitas são alterações morfológicas, metabólicas, funcionais ou outras presentes no nascimento, podendo serem diagnosticadas pré-natalmente, ao nascimento ou depois (Moore and T.V.N. Persaud 2000; World Health Organization 2020a). Anomalias congênitas são a principal causa de mortalidade e morbidade no primeiro ano de vida e representam de 15% a 39% das mortes infantis na América Latina (World Health Organization 2020b). As causas das anomalias congênitas se dividem em fatores genéticos e ambientais, ou ainda uma combinação destes, sendo que os fatores ambientais representam 10% do total das anomalias observadas (Moore and T.V.N. Persaud 2000; Brent 2001). Os fatores ambientais que causam anomalias congênitas são conhecidos como teratógenos, os quais podem gerar anormalidades estruturais ou funcionais permanentes no embrião/feto que tenha sido exposto. (Moore and T.V.N. Persaud 2000). O processo pelo qual esses agentes podem perturbar o desenvolvimento de um embrião/feto é chamado de teratogênese (Gilbert-Barness 2010) e a ciência responsável por estudar as causas e mecanismos da teratogênese e teratógenos é conhecida como teratologia (“*teratos*”= monstro, “*logos*”=estudo), sendo o termo utilizado primeiramente por I.G. de Saint-Hillaire no século XIX (Ujházy et al. 2012).

O teratógeno pode levar a quatro principais desfechos, sendo eles a morte do concepto (abortamentos), anomalias congênitas, restrição de crescimento intrauterino e deficiências funcionais. Tais desfechos dependerão dos princípios

básicos da teratologia: período crítico do desenvolvimento; dose da exposição; e genótipo-materno fetal (Moore and T.V.N. Persaud 2000). O período crítico do desenvolvimento será o período no qual o agente possui efeitos no produto da gestação, sendo esse período denominado de janela de teratogênese. Além disso, o desfecho dependerá da dosagem usada da droga ou composto químico, sendo, no geral, quanto maior a dosagem maior a gravidade do desfecho. O *background* genético materno-fetal é considerado responsável pela variabilidade de fenótipos observados, e uma das razões que explicam o porquê nem todos os expostos apresentam os efeitos teratogênicos de medicamentos ou quaisquer outras exposições. Esses princípios de teratogênese foram estabelecidos e corroborados por vários agentes teratogênicos, mas, sem dúvidas, o principal teratógeno conhecido e que gerou conhecimentos sobre as bases da teratologia é a talidomida.

1.2 TALIDOMIDA

1.2.1 Histórico

A talidomida (α-ftalimidoglutarimida) é um medicamento sintetizado na década de 50 na Alemanha Ocidental pela empresa Chemie Grünenthal. Foi inicialmente desenvolvida como um anticonvulsivante, contudo, foram identificados baixos resultados para esse efeito (Smithells and Newman 1992). Mais à frente foi identificado um efeito interessante de depressão do sistema nervoso central, sendo indicada como um potente e seguro sedativo, já que, ao contrário dos sedativos utilizados naquela época, altas doses da talidomida levavam a um quadro de sono profundo, sem risco de morte aos usuários (Smithells and Newman 1992; Shardein 1993; Kang and Ghassemzadeh 2018).

A talidomida foi comercializada em mais de 46 países (Matthews and McCoy 2003). Interessantemente, nos Estados Unidos da América (EUA), a talidomida não foi aprovada para uso e portanto não foi comercializada devido a

preocupações a respeito de efeitos adversos irreversíveis de neuropatia periférica (Calabrese and Fleischer 2000). Ela passou a ser vendida combinada com outros fármacos com o intuito de tratar as mais diversas condições como asma, hipertensão e dor de cabeça, por exemplo (Smithells and Newman 1992). Além disso, ela também se mostrou eficaz em tratar doenças e comorbidades como doenças infecciosas, gastrite, hipertireoidismo, irritabilidade e enjoos (Lenz 1988; Saldanha 1994).

Estudos toxicológicos realizados na época não conseguiram definir a dose letal aguda (DL50) da talidomida, levando-a a ser considerada de uso extremamente seguro (SOMERS 1960). Por esse motivo ela era vendida sem a necessidade de prescrição médica em diversos estabelecimentos, inclusive para gestantes como antiemético para tratar enjoos comuns no início da gravidez (Shardein 1993).

Apesar da comprovada segurança toxicológica da talidomida, na época não era de rotina realizar estudos de toxicologia reprodutiva e do desenvolvimento para avaliar a segurança dos medicamentos para uso por gestantes. Assim, no início de 1960, começaram a surgir inúmeros relatos de crianças nascidas com anomalias congênitas raras e graves, especialmente de membros. Dois médicos, Dr. Lenz e Dr. McBride, independentemente sugeriram, em 1961, que a grande quantidade de crianças nascidas com tais malformações era devido ao uso indiscriminado da talidomida (McBride 1961; Lenz 1988). No final de 1961 a talidomida foi retirada do mercado mundial. Havendo uma subsequente queda dos números de nascimentos dessas crianças, foi então estabelecida a relação causal entre o uso da talidomida durante a gestação e o aparecimento de malformações congênitas nos conceptos (Figura 1).

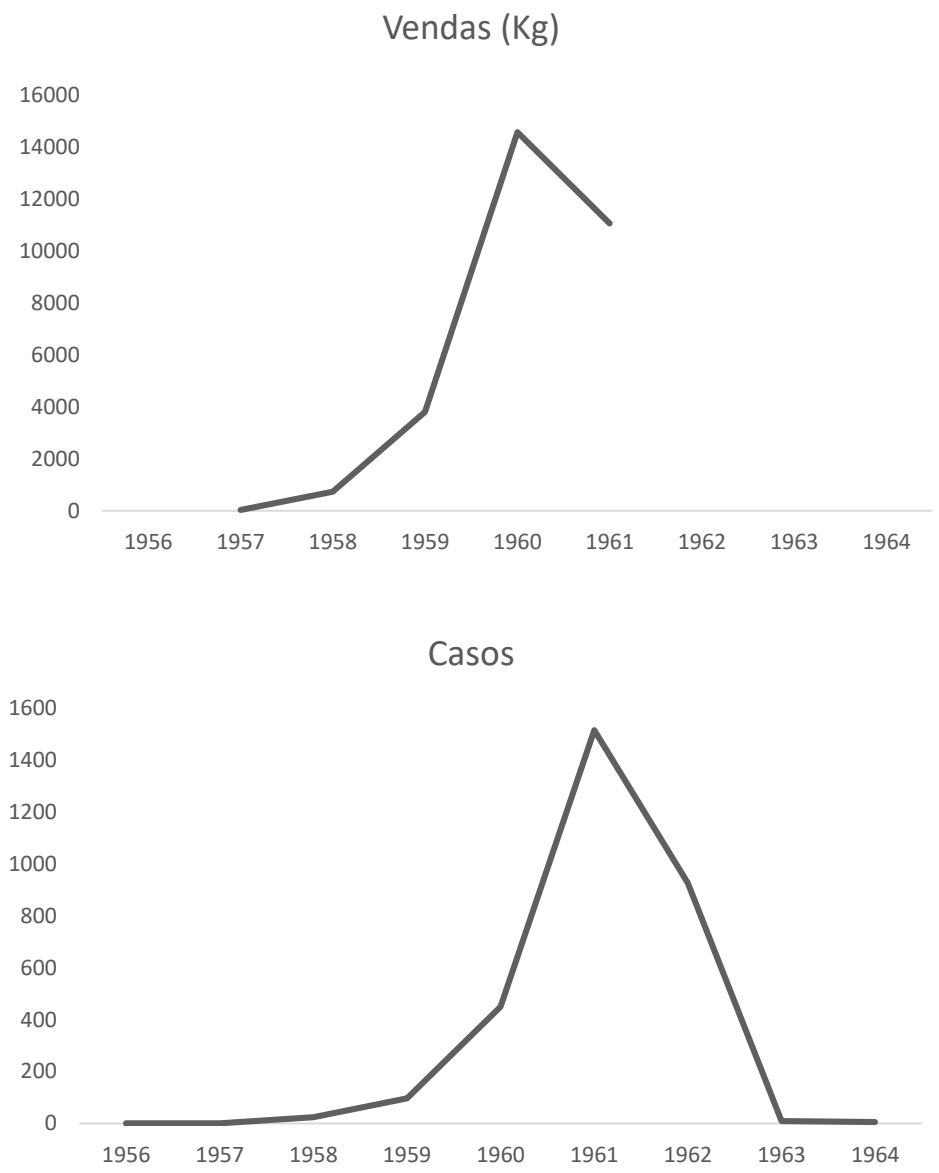


Figura 1: Número de medicamentos à base de talidomida vendidos e casos de malformações ocorridos entre 1956-1964 na Alemanha. É possível observar o pico de vendas no ano de 1960 e o pico de casos no ano de 1961 na Alemanha, decaindo os casos alguns meses após o banimento do medicamento. Fonte: Dados obtidos de Smithells and Newman 1992.

Durante o período do uso indiscriminado da talidomida, estima-se que mais de dez mil crianças tenham sido afetadas pelo conjunto de anomalias congênitas causadas pelo fármaco, que ficaram conhecidas como embriopatia da talidomida (TE) (Lenz 1988; Kim and Scialli 2011). Esta época, em decorrência do grande número de caso, a história ficou conhecida como a tragédia da talidomida.

Apesar de ter sido retirada do mercado devido à sua potente ação teratogênica, em 1965, houve a descoberta acidental do efeito da talidomida no tratamento de uma condição denominada de Eritema Nodoso Hansênico (ENH) (Sheskin 1965). O ENH é uma complicação inflamatória da Hanseníase (Darlong et al. 2016). Além disso, foi observado que a talidomida, devido às propriedades anti-inflamatórias e antiangiogênicas, teria efeito no tratamento de outras condições como o mieloma múltiplo, doença do enxerto *versus* hospedeiro, lúpus eritematoso sistêmico e reações ulcerativas da síndrome da imunodeficiência adquirida (AIDS) (Sheskin 1965; Sauer et al. 2000; Rajkumar and Blood 2006; Vianna et al. 2011). Assim, após a descoberta de sua ação para tratamento de ENH, a talidomida voltou a ser comercializada.

1.2.2 Talidomida no Brasil

A talidomida chegou ao Brasil no final da década de 50 (Saldanha 1994). Após a associação entre o uso da talidomida e seus efeitos durante a gestação em 1961, levou algum tempo até que o Brasil acompanhasse o restante do mundo e retirasse a talidomida de circulação. Foi apenas em 1962 que ela foi retirada do mercado no Brasil, três anos antes de ela ser novamente recolocada para uso devido à sua função no tratamento de ENH em 1965 (Saldanha 1994; Oliveira et al. 1999). Pelo fato de o Brasil ser um país com alta prevalência de ENH e a talidomida ser utilizada para tratamento dessa condição no país, ela ainda é bastante utilizada, sendo responsável por 63% do motivo para dispensação do fármaco no Brasil (Sales Luiz Vianna et al. 2015).

Apesar da dispensação da talidomida ter sido reiniciada logo após a descoberta se seus efeitos no tratamento de ENH em 1965, a primeira legislação de controle para o uso do medicamento no Brasil surgiu apenas em 2003. Atualmente, o uso e dispensação da talidomida são de responsabilidade da Agência Nacional de Vigilância Sanitária (ANVISA), conforme Lei 10.651 de 2003, regulamentada pela Resolução da Diretoria Colegiada da ANVISA (RDC) n°11, 22 de março de 2011. No restante do mundo também existem leis restritas

para a dispensação e uso da talidomida. Nos EUA, por exemplo, existe o programa S.T.E.P.S (*System for Thalidomide Education and Prescribing Safety in USA* – Sistema para educação à talidomida e segurança na prescrição nos EUA). Esse programa regula o uso da talidomida para seu uso seguro. As mulheres que fazem uso da medicação devem utilizar contraceptivos e realizar testes de gravidez regularmente (Uhl et al. 2006).

A demora na regulamentação da dispensação da talidomida pode ter sido o causador do fato de o Brasil ser o único país que relatou novos casos após a década de 60, sendo relatados seis casos nas década de 70, vinte casos em 80 e seis casos em 90 (Castilla et al. 1996; Sales Luiz Vianna et al. 2016). Foram relatados três novos casos em 2005-2006 e dois casos em 2007 (Schuler-Faccini et al. 2007; Vianna et al. 2011). Mais recentemente, um boletim epidemiológico do Ministério da Saúde relatou cinco novos casos de 2011 a 2019 (Ministério da saúde 2021). No total, foram relatados 42 novos casos após a década de 60 no Brasil.

1.2.3 Embriopatia da talidomida

A embriopatia da talidomida (TE) é o nome dado ao conjunto de anomalias causadas no embrião/feto em decorrência do uso de talidomida durante o desenvolvimento embrionário. Essas malformações são consideradas raras e graves, sendo bilaterais e geralmente simétricas (Smithells and Newman 1992).

As malformações causadas pela talidomida são observadas principalmente nos membros, coração, olhos e orelhas, podendo afetar outros órgãos. A principal anomalia da TE é a focomelia, que se caracteriza por encurtamento dos ossos longos dos membros com preservação das extremidades (Figura 2). Ainda nos membros, os achados mais observados são redução ou perda total de ossos longos. As anomalias ocorrem em uma ordem regular e pré-axial, iniciando pelo polegar, seguido pelo rádio, úmero e ulna, voltando à mão pelos dedos médio, anelar e mínimo. Pode ocorrer a preservação

dos dígitos ou amelia, que se caracteriza pela perda total do membro. Além disso, pode ocorrer fusão de ossos e desenvolvimento anormal dos músculos, dificultando a movimentação do membro (Smithells and Newman 1992).



Figura 2: Indivíduo portador de TE. É possível observar a focomelia nos membros superiores e inferiores. Fonte: Adaptado de Smithells and Newman 1992.

Os membros superiores são mais comumente afetados quando comparados com os membros inferiores. Além disso, os membros inferiores tem a simetria bilateral menos marcada. Nos membros inferiores os ossos longos são mais afetados, iniciando pela extremidade superior do fêmur, no local de conexão com o quadril. A amelia em membros inferiores é extremamente rara (Smithells and Newman 1992).

Seguido dos membros inferiores, a face é a mais afetada pela ação teratogênica da talidomida, em especial os olhos e as orelhas. São comumente observados problemas e inervação dos músculos oculares externos, músculos faciais e glândulas lacrimais, além de presença de fissura palatina, úvula bífida e atresia de coanas. A bilateralidade não é normalmente observada em defeitos

faciais e as malformações oculares são muitas vezes observadas em conjunto com os defeitos de orelha (Smithells and Newman 1992).

Dentre os órgãos internos o mais afetado é o coração, podendo ser observados persistência do canal arterial, defeito do septo ventricular, defeito do septo atrial e estenose pulmonar, principalmente. Anomalias cardíacas cono-truncais são observadas em afetados pela TE e associadas a óbitos desses casos (Smithells and Newman 1992). Defeitos em órgãos internos não são tão comumente observados em sobreviventes da TE a alta mortalidade. Foi descrito que em torno de 40% dos casos observados na década de 60 faleceram em decorrência das anomalias da TE (Tabela 1; Lenz 1988).

Tabela 1: Número de casos, óbitos e percentual de mortes neonatais entre indivíduos diagnosticados com embriopatia da talidomida em diferentes países na década de 1960.

Localização	Número de casos	Número de mortes	% de mortes
Suécia	153	66	43
Hamburgo	121	46	38
Canadá	116	41	35
Nova Zelândia	25	9	36
Total	415	162	39

Fonte: Adaptado de Lenz 1988b.

É importante ressaltar que, a partir de relatos e avaliações de casos, estimou-se um período durante o desenvolvimento embrionário que parece ser mais suscetível à ação teratogênica do fármaco. Esse período é chamado de janela teratogênica. A janela teratogênica da talidomida encontra-se, aproximadamente, entre o 20° ao 34° dia pós fertilização, havendo períodos mais suscetíveis para cada órgão (Figura 3; Miller and Strömland 1999; Kim and Scialli 2011). Nos períodos mais iniciais da gravidez, a exposição à talidomida gera alterações em membros superiores e a anomalias faciais (Smithells and Newman 1992). Membros inferiores são afetados em período após os membros superiores. Apesar disso, exposições antes e depois da janela teratogênica já

foram associados com algumas alterações (James 1965; Kajii et al. 1973). Exposições antes do 20° dia pós fertilização foram associadas com abortamentos espontâneos em humanos. Dessa forma, parece não haver período seguro para a exposição à talidomida (Vargesson 2015).

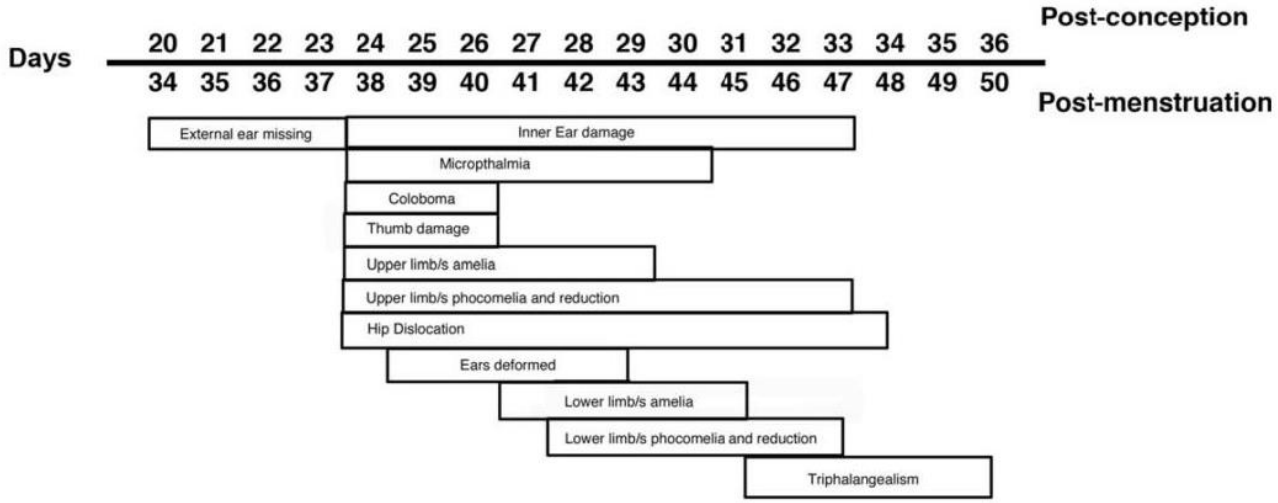


Figura 3: Janela de teratogênese da talidomida. Fonte: Vargesson 2015.

1.2.4 Mecanismos de teratogênese

Diversas hipóteses de teratogênese da talidomida foram propostas com o passar dos anos, contudo o mecanismo ainda não está completamente elucidado. Algumas hipóteses sugeridas envolvem desde a acetilação de macromoléculas até a desregulação de receptores de adesão celular (Hansen and Harris 2004). Dentre todas, as hipóteses mais bem aceitas e estudadas são relacionadas à indução de estresse oxidativo, ação antiangiogênica e ligação à proteína Cereblon (D'Amato et al. 1994; Parman et al. 1999; Ito et al. 2010).

Estudos observaram que o tratamento com talidomida leva à formação de espécies reativas de oxigênio (ROS) causando consequentemente estresse oxidativo às células (Parman et al. 1999; Hansen and Harris 2004). Foi observado que o tratamento de coelhos com talidomida levou à oxidação do DNA, sendo que essa oxidação voltava aos níveis normais após pré-tratamento com um agente de captura de radical livre (fenil-alfa-tetra-butil nitrona). Isso

mostrou que a talidomida gera a oxidação do DNA por meio da produção de ROS (Parman et al. 1999). Além disso, o tratamento com talidomida gerou malformações na prole semelhantes às observadas em humanos, incluindo focomelia. O fenótipo normal foi resgatado com o pré-tratamento com o agente de captura de radical livre (Parman et al. 1999).

O fator nuclear Kappa B (NF-KB) foi relatado como alterado durante esse processo de produção de ROS após tratamento com talidomida. Após a exposição ao fármaco no broto do membro de coelhos, observou-se a diminuição da quantidade de NF-KB, a qual foi resgatada após o uso de agentes de captura de radical livre (N-acetilcisteína e fenil-alfa-tetra-butil nitrona) (Hansen et al. 2002). Além disso, também observaram a diminuição da expressão de *Fgf8* (Fator de crescimento de fibroblasto 8)*, Fgf10* (Fator de crescimento de fibroblasto 10) e *Twist,* genes importantes durante o desenvolvimento de membros (Hansen et al. 2002).

Além da produção de ROS, foi observada capacidade antiangiogênica da talidomida. O tratamento com talidomida em broto de membro de galinha indicou bloqueio no crescimento de novos vasos sanguíneos (Jurand 1966). Em coelhos, se observou a inibição da área vascularizada da córnea após tratamento com talidomida no local (D'Amato et al. 1994). Em embriões de z*ebrafish*, a talidomida demonstrou possuir efeitos antiangiogênicos especialmente através da depleção de receptores do Fator de Crescimento Vascular Endotelial (VEGF; Yabu et al. 2005).

Um análogo sintético a um metabólito da talidomida, CPS49, foi administrado no broto do membros em embriões de galinha e foi observado um importante efeito antiangiogênico e perda de sinalização de fatores importantes para o desenvolvimento de membros como o Fator de Crescimento de Fibroblasto (FGF) e Sonic Hedgehog (Shh) (Therapontos et al. 2009). Além disso, os autores foram capazes de observar malformações nos membros que ocorreram em decorrência da perda de novos vasos em formação. Dessa forma, demonstraram que o potencial antiangiogênico levaria a defeitos nos membros. Ademais, em modelo de angiogênese em membrana do saco vitelínico do ovo de galinha, foi observado a inibição da angiogênese mediada por óxido nítrico

(NO) e inibição da formação de células endoteliais, corroborando o potencial antiangiogênico da talidomida (Tamilarasan et al. 2006).

Dentre o mecanismo mais recentemente descrito está a identificação da proteína Cereblon (CRBN) como um alvo da talidomida no organismo (Ito et al. 2010). CRBN faz parte de um complexo E3-ubiquitina ligase juntamente com as proteínas *Cullin* 4A (CUL4A) e *Damaged DNA Binding Protein 1 (*DDB1), formando o complexo $CRL4^{CRBN}$. Foi relatado que a talidomida se liga à CRBN na porção C-terminal, região mais conservada da proteína. Também foi observado que essa ligação da talidomida ao complexo $CRL4^{CRBN}$ levava a diminuição da expressão de *fgf8* e *fgf10* em embriões de galinhas e *zebrafish* (Ito et al. 2010).

A partir do reconhecimento da ligação talidomida-$CRL4^{CRBN}$, foi observado que CRBN tem um papel na ação de fármacos imunomoduladores (IMiDs) no tratamento de Mieloma Múltiplo (Zhu et al. 2011). A talidomida possui propriedades imunomoduladoras, sendo incluída neste grupo de IMiDs (Haslett et al. 1998; Davies et al. 2001). A partir desses estudos, observou-se que uma vez que IMiDs se ligam ao complexo $CRL4^{CRBN}$, o complexo passava a ubiquitinar os fatores de transcrição Ikaros (IKZF1) e Aiolos (IKZF3) (Krönke et al. 2014; Lu et al. 2014). Outros substratos do complexo IMiDs-$CRL4^{CRBN}$ foram identificados (Krönke et al. 2015; An et al. 2017). A partir disso surgiu a hipótese de que a ligação de IMiDs, inclusive da talidomida, ao complexo $CRL4^{CRBN}$ através de CRBN induziria à ubiquitinação de novos substratos e sua subsequente degradação por proteossomo (Figura 4). Mais recentemente, novos substratos do complexo $CRL4^{CRBN}$ ligado à talidomida foram observados, incluindo o Fator de Transcrição do tipo *Spalt* 4 (SALL4) e outras proteínas com domínio dedo de zinco C2H2 (Donovan et al. 2018; Sievers et al. 2018).

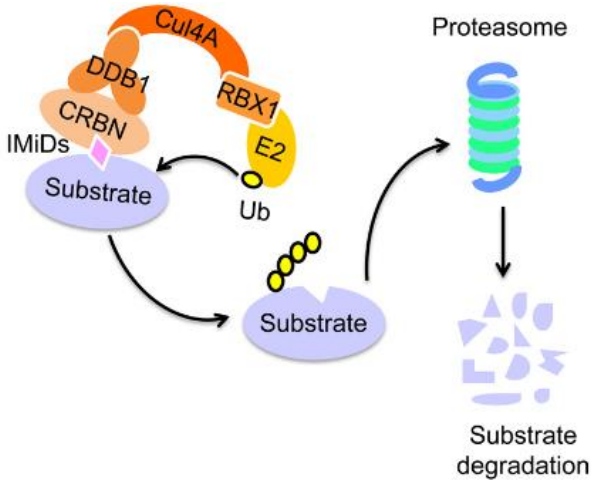


Figura 4: Ligação de fármacos imunomoduladores ao complexo $CRL4^{CRBN}$ e potencias neosubstratos de ubiquitinação. IMiDs se ligam ao complexo $CRL4^{CRBN}$ através da proteína CRBN, tendo por consequência neosubstratos que serão ubiquitinado e degradados via proteossomo, afetando, por exemplo o Fator de Transcrição do tipo *Spalt* 4 (SALL4). IMiDs: fármacos imunomoduladores; Ub: ubiquitinação. Fonte: Adaptado de Gao et al. 2020.

SALL4, um fator de transcrição com domínio dedo de zinco C2H2, foi relatado ser ubiquitinado pela complexo $CRL4^{CRBN}$-talidomida no domínio dedo de zinco, assim como outros neosubstratos do complexo, mais especificamente Ikaros e ZFP91 (Matyskiela et al. 2018). Observou-se também a degradação de SALL4 após tratamento com talidomida em coelhos, sendo que a degradação foi inibida com o *knockout* de CRBN em células tronco pluripotente induzidas (células iPS). Além disso, a degradação de SALL4 também foi diminuída na presença de um inibidor de proteossomo, mostrando que o mecanismo de degradação ocorre de fato via proteossomo (Matyskiela et al. 2018). Outro estudo, o qual partiu de uma abordagem *in vitro* em células tronco embrionárias humanas (células hESC), observou que SALL4 encontrava-se com expressão diminuída após tratamento com talidomida e de seus análogos, pomalidomida e lenalidomida (Donovan et al. 2018). De maneira interessante, SALL4 é um fator de transcrição expresso durante o desenvolvimento, e variantes patogênicas no gene de mesmo nome que codifica SALL4 causam uma síndrome conhecida como síndrome de Okihiro ou síndrome de Anomalia Radial e Duane (Cox et al. 2002). Esta síndrome possui diversos fenótipos semelhantes com os observados

na TE, desta forma a TE é conhecida como uma fenocópia da síndrome de Okihiro.

Outras proteínas com domínio dedo de zinco C2H2 foram vistas serem neosubstratos do complexo $CRL4^{CRBN}$ na presença de talidomida. Um estudo observou seis proteínas que foram completamente degradadas na presença da talidomida, sendo elas IKZF3, CFP91, ZNF276, ZNF653, ZNF692 e ZNF827 (Sievers et al. 2018). IKZF3 atua principalmente no sistema imune no desenvolvimento de linfócitos, enquanto o restante, em sua maioria, atua como ligador de DNA e regulação de expressão gênica. Interessantemente, ZNF276 está associado com Anemia de Fanconi e ZNF653 com a síndrome de Okihiro, síndromes às quais a TE é considerada fenocópia.

É importante notar que não necessariamente cada hipótese de mecanismo teratogênico exclui as demais postuladas. Sauer et al. (2000) mostrou que o processo de antiangiogênese produzido pela talidomida pode estar ligado à produção de ROS. Devido ao fato de a teratogênese da talidomida produzir uma grande variedade de fenótipos e afetar diversos sistemas, é mais provável que diversos fatores auxiliem nos mecanismos de sua teratogenicidade. Nesse sentido, muitas são as vias e genes-alvo estudados para explicar a TE. Uma das abordagens que tenta explicar os mecanismos de surgimento das anomalias se baseia na variabilidade genética como mecanismo de susceptibilidade e modificador de fenótipo.

1.2.5 Suscetibilidade genética à Embriopatia da Talidomida

Nem todos os indivíduos expostos à talidomida durante a janela teratogênica desenvolvem a TE. Estima-se que 20-50% das crianças expostas desenvolvem as malformações referentes à teratogênese da talidomida. Essa variação é decorrente da susceptibilidade genética materno-fetal (Cassina et al. 2012; Vargesson and Fraga 2017). Desta forma, uma das maneiras de melhor compreender os mecanismos moleculares da teratogênese da talidomida é

identificando genes que influenciam na susceptibilidade ao desenvolvimento da TE.

São escassos os estudos realizados com o objetivo de identificar genes/variantes para a susceptibilidade da TE. O gene *NOS3*, o qual codifica a Enzima Óxido Nítrico Sintase Endotelial (eNOS), foi estudado em pessoas afetadas pela TE e se observou a presença de duas variantes, uma repetição em tandem de número variável (VNTR) no intron 4 e um polimorfismo no promotor, que em conjunto conferiam susceptibilidade genética à TE (Kowalski et al. 2016). No estudo, os indivíduos diagnosticados com TE apresentaram maior frequência do alelo que está relacionado a níveis menores de eNOS, o que potencialmente poderia ser um fator de risco adicional em um contexto de exposição a um antiangiogênico (Vianna et al. 2013; Kowalski et al. 2016). Variantes em outros genes também foram avaliados, como *PTGS2*, *VEGFA*, *FGF8, FGF10, BMP4, SHH, TP53, TP63* e *TP73*, porém os estudos não observaram diferença nas frequências dos genótipos entre indivíduos com TE e grupo controle (Kowalski et al. 2017; Gomes et al. 2018). Tais genes atuam em processos angiogênicos e de desenvolvimento embrionário, principalmente dos membros.

Dentre os genes que integram o complexo $CRL4^{CRBN}$, *CRBN, DDB1* e *CUL4A* foram sequenciados juntamente com os genes *IKZF1 e IKZF3* em portadores da TE para avaliar o impacto das variantes na susceptibilidade e nos fenótipos da embriopatia (Kowalski et al. 2020). Foram observadas um total de 145 variantes nos cinco genes avaliados, sendo que 7,5% das variantes estavam localizadas em regiões codificantes. Interessantemente, *CRBN* e *IKZF1* tiveram a maior parte das variantes em região regulatória 3'UTR, sendo essa frequência estatisticamente diferente quando comparadas à frequência na população em geral. Além disso, foram realizados estudos de predição funcional com as variantes identificadas. Vale ressaltar que 10 variantes observadas em *CRBN* causavam perturbação de sítios de ligação no local que codifica a cauda C-terminal de Cereblon, responsável pela ligação da proteína à talidomida (Kowalski et al. 2020). Finalmente, variantes genéticas podem influenciar na variabilidade de fenótipos. Duas variantes em *CRBN* foram observadas estar associadas a determinados fenótipos. As variantes rs1045433 e rs1620675

foram associadas com malformações pré-axiais de membros e anormalidades neurológicas, respectivamente (Vianna et al. 2016; Kowalski et al. 2020).

Dentre os genes já avaliados também estão *ESCO2 (Establishment Of Sister Chromatid Cohesion N-Acetyltransferase 2*), *SALL4 e TBX5 (T-BOX 5)*, os quais estão associados a síndromes as quais TE é considerada uma fenocópia devido à similaridade de fenótipos (Gomes et al. 2019). Este estudo observou variantes nos três genes que estavam mais frequentes em indivíduos com TE do que na população em geral, sendo três variantes em *ESCO2,* em região intrônica e 3'UTR; duas em *SALL4* em região exônica; e três em *TBX5*, em região intrônica e exônica (Gomes et al. 2019).

1.3 TBX5

TBX5 é um fator de transcrição com domínio T-box importante no desenvolvimento embrionário. Ele está expresso nos olhos, membros superiores e coração em embriões (Liberatore et al. 2000; Steimle and Moskowitz 2017). TBX5 está principalmente envolvido na formação do sistema cardiovascular, sendo um dos primeiros fatores a ser associado com anomalias cardíacas congênitas (Clowes et al. 2014). TBX5 expressa-se, no coração, no epicárdio, miocárdio e endocárdio, estando também expresso na parede das câmaras cardíacas, com expressão mais significativa nos átrios de embriões em desenvolvimento (Hatcher et al. 2000). Desta forma, ele atua na formação das câmaras cardíacas e na regulação do desenvolvimento atrial e de septo cardíaco (Clowes et al. 2014).

Estudos em camundongos observaram que o *knockout* de *Tbx5* levava a morte da prole ainda *in utero*, sendo observadas diversas anomalias cardíacas e total ausência dos brotos dos membros superiores (Bruneau et al. 2001). Isso aponta que Tbx5 teria um papel imprescindível no início da formação dos brotos dos membros.

Foi demonstrado que TBX5 atua regulando positivamente e negativamente em diversos alvos, como por exemplo *NPPA (*Peptídeo Natriurético A)*,* um marcador de diferenciação de câmara cardíaca (Bruneau et al. 2001; Steimle and Moskowitz 2017) (Figura 5). Além disso, atua interagindo com diversos outros fatores de transcrição importantes no desenvolvimento cardíaco, como Proteína Ligadora de GATA 4 (GATA4), NK2 *Homeobox 5 (*NKX2-5) e Fator Potenciador de Miócito 2C (MEF2C; Steimle and Moskowitz 2017).

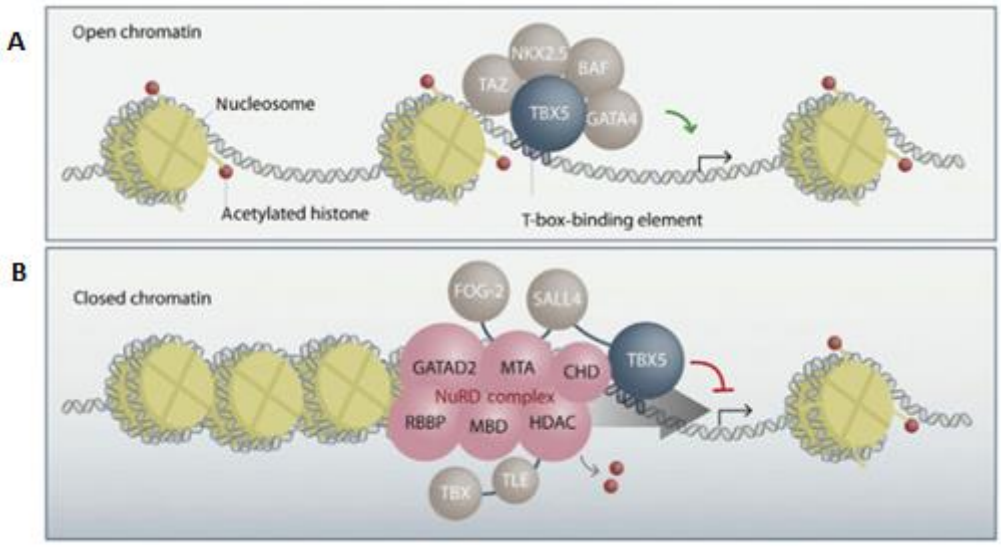


Figura 5: Regulação positiva e negativa de TBX5. A. Atuação de TBX5 na regulação positiva de alvos no DNA de forma sinérgica a outros fatores, como GATA4 e NKX2-5. B. Regulação negativa, de fechamento da cromatina através da interação com outros alvos, como SALL4. Tbx5: T-BOX 5; GATA4: Proteína Ligadora de GATA 4; NKX2-5: NK2 *Homeobox 5*; Sall4: Fator de Transcrição do tipo *Spalt* 4. Fonte: Steimle and Moskowitz 2017.

O gene de TBX5, que leva o mesmo nome que a proteína, está localizado no cromossomo 12 e em fita reversa. Alterações no gene *TBX5* são causadoras de uma síndrome conhecida como Holt-Oram. A síndrome de Holt-Oram se caracteriza por uma doença autossômica dominante rara em que em torno de 70% dos afetados possuem variantes patogênicas em heterozigose em *TBX5.* Ademais, 85% dos indivíduos desenvolvem a doença devido a variantes patogênicas *de novo* em *TBX5* (McDermott et al. 1993).

As principais características observadas em indivíduos com a síndrome de Holt-Oram são defeitos cardíacos e de membros superiores. Os defeitos cardíacos se caracterizam principalmente por defeitos de septo e de condução, já os defeitos de membros são observadas especialmente anomalias pré-axiais (Steimle and Moskowitz 2017). Muitas das anomalias observadas são similares às observadas na TE, desta forma, a TE é conhecida como uma fenocópia da síndrome de Holt-Oram, sendo a principal diferença entre as malformações a possibilidade de os defeitos serem unilaterais na síndrome de Holt-Oram, diferente da TE, na qual os defeitos observados são bilaterais (Basson et al. 1997).

Ademais, TBX5 também foi visto ser um alvo da talidomida. Além disso, foi observada que uma interação entre TBX5 e outro fator de transcrição HAND2 (*Heart and Neural Crest Derivatives Expressed 2)* era inibida em até 80% na presença do medicamento (Khalil et al. 2017).

1.4. HAND2

Heart and Neural Crest Derivatives Expressed 2 (HAND2) é um fator de transcrição parte do grupo *basic Helix-loop-helix* (bHLH), da família Twist e subfamília Hand (Figura 6; Tamura et al. 2014). Esse fator é muito importante durante o desenvolvimento embrionário, sendo necessário para a formação normal do sistema cardiovascular e dos membros (Srivastava et al. 1995; Dai and Cserjesi 2002).

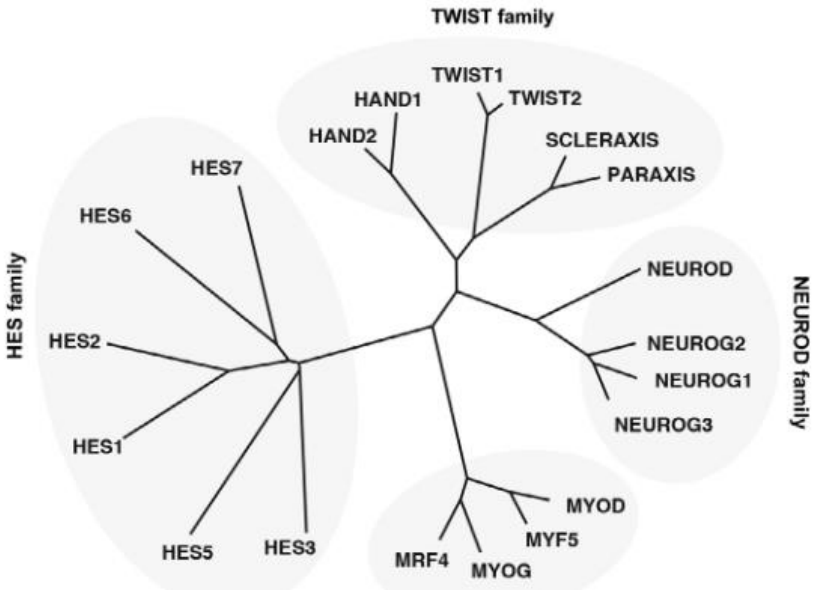


Figura 6 – Árvore demonstrando as famílias de fatores de transcrição bHLH. Nota-se que HAND2 (*Heart and Neural Crest Derivatives Expressed 2)* faz parte da família Twist e mais especificamente da subfamília Hand. Fonte: Tamura et al. 2014.

Proteínas da família bHLH são muito conservadas evolutivamente, estando presentes desde leveduras até humanos (Atchley and Fitch 1997). Possuem domínios muito conservados de ligação ao DNA e interação proteína-proteína. O motivo no DNA ao qual se ligam é conhecido como E-box e possui a sequência consenso CANNTG. Esse motivo é encontrado em grandes quantidades no genoma e as duas bases do meio permitem maior especificidade de ligação para cada grupo dentro da família bHLH (Atchley and Fitch 1997).

A ligação de fatores bHLH depende da dimerização deste fatores com outros da mesma família (Dai and Cserjesi 2002). HAND2 foi visto se dimerizar com HAND1, que faz parte da mesma subfamília (Figura 6), formando um heterodímero. Além disso, HAND2 também pode formar um homodímero *in vitro,* porém foi observado que o homodímero de HAND2 não é funcional *in vivo* (Firulli et al. 2000; Dai and Cserjesi 2002). Ademais, foi observado *in vivo* que HAND2 forma heterodímero com proteínas-E a fim de regular a expressão gênica (Dai and Cserjesi 2002). O mesmo estudo mostrou que o heterodímero HAND2 com a proteína E12 se ligava preferencialmente ao motivo E-box CATCTG.

Alterações em HAND2 já foram associadas com anormalidades do sistema cardiovascular. A deleção de *Hand2* em camundongos resulta em letalidade embrionária por falência cardíaca (Srivastava et al. 1997; Yamagishi et al. 2000). Ainda em camundongos-nulo para Hand2, foi observado a ausência do ventrículo direito com formação incorreta do ventrículo esquerdo (Yamagishi et al. 2000). Além disso, foi observado um sistema vascular defeituoso, sugerindo participação de Hand2 na formação da vasculatura e angiogênese. Os defeitos na vasculatura foram parecidos com os observados em modelos nulos para *VEGF (*fator de crescimento endotelial vascular) (Yamagishi et al. 2000).

HAND2 regula diversos fatores importantes no desenvolvimento, além de ser regulado por GATA4, um fator de transcrição dedo de zinco importante na formação do coração (Zeisberg et al. 2005). Camundongos *knockout* para *Gata4* apresentaram fenótipos similares aos camundongos *knockout* para *Hand2,* como hipoplasia do ventrículo direito. Além disso, relata-se que o *knockout* de *Gata4* diminui a expressão de *Hand2* (Zeisberg et al. 2005). Ainda no coração, HAND2 faz a regulação de VEGF (Yamagishi et al. 2000). Ademais, variantes patogênicas em *HAND2* foram associadas com anomalias cardiovasculares, como tetralogia de Fallot (Lu et al. 2016; Khalil et al. 2017).

HAND2 também atua na formação dos membros, sendo expresso no mesoderma lateral do embrião, no local onde surgirá o brotamento do membro superior e parte posterior do broto superior (Srivastava et al. 1995; Charité et al. 2000; Tamura et al. 2014). Foi observado que HAND2 realiza a regulação de GLI3 (*Gli Family zinc finger 3)* durante a formação dos membros (Figura 7; Yamagishi et al. 2000), além de outros fatores, como ativação de *TBX3* e *TBX2* e repressão de *TBX18* (Osterwalder et al. 2014).

HAND2 possui expressão, durante o desenvolvimento embrionário, no coração e membros, estando expresso também na decídua (tecido extraembrionário), células da crista neural cardíaca, ventrículo direito, trato de saída ventricular, pericárdio, células da crista neural cranial, glândula adrenal e estômago (Srivastava et al. 1995; Charité et al. 2000; Tamura et al. 2014).

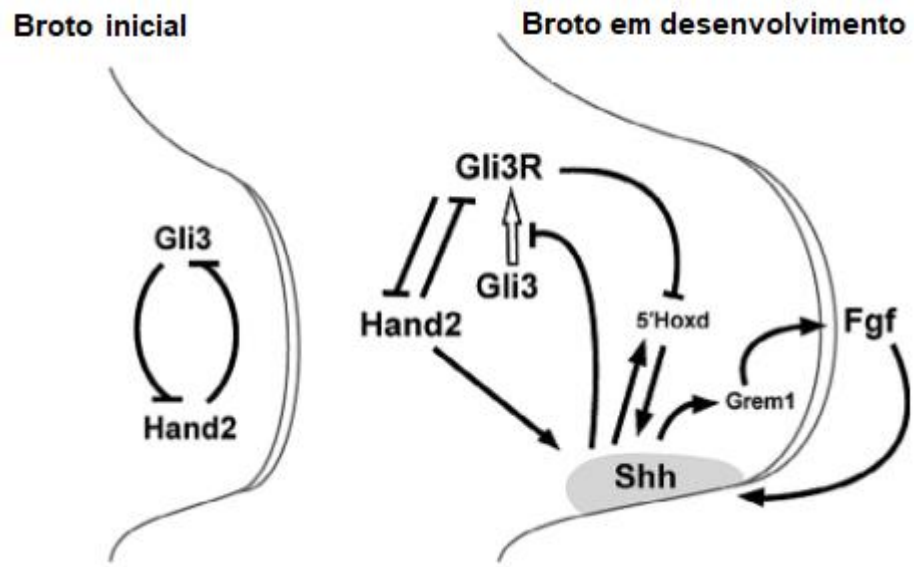


Figura 7: Esquema de regulações durante o desenvolvimento do broto do membro. É possível observar a interação entre Hand2 (*Heart and Neural Crest Derivatives Expressed 2)* e Gli3 (*Gli Family zinc finger 3)* durante a formação inicial do broto do membro e de Hand2 com outros fatores durante o crescimento do broto. Fonte: Adaptado de Tamura et al. 2014.

1.4.1 HAND2, TBX5 e talidomida

Mais recentemente, um novo alvo de ligação da talidomida foi descrito. Khalil *et al.* (2017) descreveu uma nova ligação entre TBX5 e talidomida. Essa ligação ocorre no domínio T-box da proteína, mesmo local pelo qual essa se liga a seus alvos no DNA. Desta forma, foi sugerido que a ligação da talidomida à TBX5 impediria suas funções normais no organismo. Além disso, o estudo descreveu uma nova interação entre os dois fatores de transcrição TBX5 e HAND2. Os autores observaram que após a exposição de talidomida essa interação era inibida em 80%.

Apesar de ter sido o primeiro estudo identificando uma interação direta entre os dois fatores de transcrição, HAND2 e TBX5 já são conhecidos por serem importantes durante o desenvolvimento embrionário, principalmente da formação do sistema cardiovascular e dos membros. Outra interação documentada entre os HAND2 e TBX5 foi em conjunto com mais duas proteínas,

MEF2C e GATA-4, conjunto o qual foi capaz de reprogramar fibroblastos a cardiomiócitos *in vitro* (Song et al. 2012).

Essa interação entre HAND2 e TBX5 pode ser importante para o desenvolvimento correto dos membros e coração, já que os fatores de transcrição estão associados à formação normal das duas estruturas separadamente. Desta forma, a inibição da interação entre HAND2 e TBX5 poderia afetar a formação adequada do coração e membros, podendo estar associados às malformações observadas após exposição à talidomida nesses dois órgãos.

CAPÍTULO II

JUSTIFICATIVA E OBJETIVOS

2. JUSTIFICATIVA

Mesmo mais de 60 anos após a constatação do potencial teratogênico da talidomida e de crianças ainda nascendo com TE, não se conhece por completo os mecanismos pelos quais o fármaco é capaz de gerar as malformações observadas na embriopatia. Mais recentemente, um estudo descreveu uma interação entre dois fatores de transcrição importantes para o desenvolvimento, principalmente do coração e membros, sendo eles TBX5 e HAND2. O mesmo estudo observou que na presença de talidomida essa interação estava quase completamente inibida, levantando a hipótese de participação desses dois fatores na TE.

Levando em consideração que nem todos os expostos à talidomida durante o desenvolvimento irão desenvolver a TE e apenas uma porcentagem realmente é afetada, a variabilidade genética em genes relevantes para desenvolvimento de membros e coração podem ser explorados. *TBX5* já foi avaliado para suscetibilidade genética à TE. Neste sentido, a avaliação de *HAND2* pode fornecer dados interessantes sobre os efeitos teratogênicos da talidomida, inclusive a partir da caracterização de variantes identificadas. Além disso, levando em consideração que os alvos de suscetibilidade genética e de teratogenicidade da talidomida ainda não estão completamente descritos, estudos de bioinformática que avaliam amplamente as redes de interação nas quais a talidomida pode estar atuando durante o desenvolvimento podem ser uma forma de identificar novos alvos a serem estudados explorados na suscetibilidade genética à TE e seus mecanismos de ação.

A recente descoberta de HAND2 sendo afetado pela talidomida nos dá pistas para investigá-lo como alvo e também as cascatas possivelmente afetadas, podendo inclusive identificar novos alvos para suscetibilidade. Dessa forma, avaliar o gene molecularmente, bem como suas interações, ajudaria a compreender melhor os mecanismos de teratogênese da talidomida. Portanto, a caracterização de variantes observadas em indivíduos portadores de TE e a seleção de novos genes de interesse para estudos de suscetibilidade genética à TE são importantes para ajudar a preencher as lacunas da teratogenicidade da talidomida.

3. OBJETIVOS

3.1. Objetivo geral

Avaliar o gene *HAND2* e suas possíveis associações com a embriopatia da talidomida e selecionar novos alvos para análises de susceptibilidade genética à embriopatia da talidomida a partir de estudos *in silico*.

3.2 Objetivos específicos

A. Realizar a caracterização molecular do gene *HAND2* por meio de ferramentas de bioinformática.

B. Predizer os efeitos da variante rs59621536 do gene *HAND2* no mRNA por meio de ferramentas *in silico.*

C. Avaliar os efeitos da variante rs59621536 de *HAND2* na metilação do gene em indivíduos com TE.

D. Investigar as vias de conexão/sinalização nas quais HAND2 e TBX5 estão envolvidas para entender sua ação nas anomalias observadas na TE e selecionar genes de interesse para serem avaliados na susceptibilidade genética à TE por meio de ferramentas de bioinformática e biologia de sistemas.

CAPÍTULO III

ARTIGO I

Genetic Evaluation of *HAND2* gene and its Effects on Thalidomide Embryopathy

Manuscrito em preparação

Gene

Genetic Evaluation of *HAND2* gene and its Effects on Thalidomide Embryopathy

Bruna Duarte Rengel[A-D] Thayne Woycinck Kowalski[A-E, G, H], João Matheus Bremm[B], Julia do Amaral Gomes[A-E], Lavínia Schüler-Faccini[A-C, E], Lucas Rosa Fraga[A,C,D,F,I], Fernanda Sales Luiz Vianna[A-E, I]

[A]Laboratory of Medical Genetics and Evolution, Genetics Department, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; [B]Postgraduate Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil; [C]Brazilian Teratogen Information Service (SIAT), Medical Genetics Service of Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; [D]Genomic Medicine Laboratory at Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, Brazil; [E]National Institute of Population Medical Genetics (INAGEMP), Porto Alegre, Brazil; [F]Department of Morphological Sciences, Institute of Health Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.; [G]Bioinformatics Core, Hospital de Clínicas de Porto Alegre, HCPA, Porto Alegre, Brazil; [H]Centro Universitário CESUCA, Cachoeirinha, Brazil.[I]Postgraduate Program in Medical Sciences, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil.

Corresponding authors: Fernanda Sales Luiz Vianna, Postgraduate Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul (UFRGS), Bento Gonçalves Avenue, 9500 – 43312 building, CEP: 91501-970, Porto Alegre, Brazil. Tel: +55 51 3359-7670. Email: fslvianna@gmail.com. AND Lucas Rosa Fraga, Department of Morphological Sciences, Institute of Health Sciences, Universidade Federal do Rio Grande do Sul (UFRGS), Rua Sarmento Leite, 500, CEP: 90035-190, Porto Alegre, RS, Brazil. Tel: +55 51 33083452. Email: lrfraga@ufrgs.br

Abstract

HAND2 is a transcription factor important for embryonic development, being required for limbs and cardiovascular system development. Thalidomide is a drug responsible to a spectrum of congenital anomalies known as Thalidomide Embryopathy (TE), which includes mainly limb, eye, and heart defects. Recently, a study observed an interaction between HAND2 and TBX5, which was inhibited in the presence of Thalidomide. The aim of this study was to evaluate and characterize *HAND2* gene and to evaluate its variability in a sample of TE individuals. *HAND2* was evaluated in 35 individuals with TE. DNA from TE subjects was extracted from saliva samples and PCR was performed for amplification and Sanger sequencing of *HAND2* gene. Sequencing analyses of *HAND2* coding sequence showed only one variant. This is a synonymous variant p.P51 (rs59621536) in exon 1 and was found in 3 individuals. As result, through *in silico* evaluation, *HAND2* was observed to be very conserved, being the 3'UTR the most polymorphic region of the gene. Next, i*n silico* analyses classified the variant as neutral, without alteration in splicing sites and miRNA sites. *In silico* predictions pointed to alteration of two CpG islands adjacent to the variant; however, we did not observe any alterations on the methylation pattern of *HAND2* gene in the TE individuals. In addition, alteration of the binding site of MeCP2, a nuclear protein involved in DNA methylation, was predicted along with alteration in HAND2 mRNA structure. In conclusion, our study was not able to associate *HAND2* to genetic susceptibility to TE. However, we observed one variant in three individuals which was predicted to alter mRNA structure and methylation patterns. Furthermore, we observed that *HAND2* is a well conserved gene. Further studies with a larger sample should be performed to evaluate the role of the variant in cardiac outcomes.

Keywords: HAND2; Polymorphism; Teratogenesis; Thalidomide

1. INTRODUCTION

Heart and Neural crest derivatives-expressed 2 (HAND2) is an important transcription factor for embryonic development, being expressed in the heart, limbs and neural crest derivatives (Dai and Cserjesi, 2002; Srivastava et al., 1995). HAND2 regulates several important factors for the formation of these organs, such as Gli family zinc finger 3 (GLI3) and vascular endothelial growth factor (VEGF) (Yamagishi et al., 2000). Altered HAND2 function, due to changes in expression patterns or gene deletion have been associated to cardiovascular malformations, embryonic lethality by cardiac failure and lack or incorrect formation of left ventricle in animal models (Dai and Cserjesi, 2002; Srivastava et al., 1997, 1995; Yamagishi et al., 2000) . In addition, in mice, overdosage of Hand2 causes heart and limb abnormalities (Tamura et al., 2013).

Recently, it was observed an interaction between HAND2 and the transcription factor T-box 5 (TBX5), which was inhibited by 80% in the presence of Thalidomide (Khalil et al., 2017). Moreover, they observed a strong interaction between thalidomide and the TBX5, which reduces its binding to its DNA targets (Khalil *et al.*, 2017). TBX5 is very important during cardiac development, mainly in the formation of the cardiac septum, but it is also involved in the development of the upper limbs (Clowes et al., 2014). Thalidomide is a drug well known for its teratogenic potential and it was associated with the birth of thousands of children with severe and rare congenital malformations in several organs in 1960s. The set of malformations resulting from Thalidomide exposure is known as Thalidomide Embryopathy (TE) and consists of bilateral and symmetrical limbs abnormalities, as well as heart, eyes and face anomalies (McBride, 1961; Smithells and Newman, 1992). Although more than 60 years have passed after Thalidomide was first synthesized, its teratogenic mechanisms have not yet been fully elucidated.

Our research group has been extensively worked on understanding the molecular targets of thalidomide through genetic approaches in TE samples (Gomes et al., 2019, 2018; Kowalski et al., 2020, 2017, 2016; Vianna et al., 2016, 2013). We have previously evaluated *TBX5* gene in TE individuals (Gomes et al., 2019). From the recent findings on TBX5 and HAND2 interaction and thalidomide, a molecular investigation on *HAND2* could bring novelty on how this gene influences limbs and heart defects. Therefore, here we evaluated the genetic

susceptibility of *HAND2* to TE by analyzing its coding sequences. We found only a single variant, P.p51, in three individuals with TE. Further, *in silico* analysis showed HAND2 being very conserved and a potential effect of pP51 on mRNA structure and the methylation in the gene, which was not observed in TE samples.

2. METHODS

2.1 Ethical Aspects

This study was approved by the Ethics and Research Committee of Porto Alegre Clinical Hospital (CAAE 21583913600005327). All subjects included in this study signed the Informed Consent Form.

2.2 Sample

Thirty-five individuals with Thalidomide embryopathy participated in this study. They were recruited from the Brazilian Association of Thalidomide Syndrome (ABPST) and had previous confirmation of the phenotype by clinical evaluations of the congenital anomalies. The clinical evaluation of the individuals with TE analyzed in this study was performed and described by Kowalski *et al.*(2015). We considered as cardiac conditions, outcomes presented later in life related to cardiovascular system (for instance, hypertension). Congenital anomalies were considered the anomalies presented at birth. It is important to notice that we do not have access to the dose or time of exposure to each individuals.

2.3 Molecular Analysis

Saliva samples were collected from individuals using the DNA Oragene Kit (Genotek). DNA extraction from saliva samples was performed according to the manufacturer's instructions. Sequencing of all the coding region of *HAND2* gene was performed on subjects with TE by the Polymerase Chain Reaction (PCR) technique using Taq Polimerase Platinum followed by Sanger sequencing. Analyses of exons 1 and 2, which corresponds to the coding part of the gene, were performed with amplification of 688bp (exon 1) and 547bp (exon 2) fragments using the following primer pairs 5'-AATGAGTCTGGTAGGTGGTTTTCC-3 '(forward) and 5'-

GTGCAAACAACCTTGAAGCG-3' (reverse); and 5 'CCGCTCCTCCATAAACGACT-3' (forward) and the primer 5 'GATGCCTCAAAGGGGGTCAA-3' (reverse), respectively. Confirmation of fragment amplification as well as samples purification were carried out prior to sequencing.

2.4 *In silico* Analyses

Chromatograms were aligned with the reference sequence (RefSeq NP_068808) obtained by Ensembl database (ensemble.org). Analyses were performed with CodonCode® Aligner v.3.0.1 program (CodonCode Corporation).

Molecular *in silico* characterization, which is a characterization of variants located in *HAND2,* was performed using RStudio package biomaRt (R v3.5.2). In order to evaluate the impact of genetic variants identified, *in silico* predictions were performed using the PredictSNP2 (https://loschmidt.chemi.muni.cz/predictsnp2/) platform and the Mutation Taster software (http://www.mutationtaster.org/). The Human splicing Finder® tool (Genetics and Bioinformatics Team) was used to evaluate the role of the variants at splicing sites. Changes in miRNA binding sites were analyzed using the miRBase tool (http://www.mirbase.org/). Changes in CpG islands by the variant were evaluated by the MethPrimer® (The Li Lab) tool. The Enhancer Atlas 2.0 database (http://enhanceratlas.org/) was used for changes in enhancer regions. The RStudio package *motifbreakR* (Coetzee, Coetzee e Hazelett, 2015) (R v3.5.2) was used for the analysis of possible alterations in transcription factors binding sites. RNAfold Web server (University of Vienna) was used to perform analysis of changes in mRNA.

2.5 DNA Methylation Analyses

DNA samples were subjected to bisulfite conversion using Cells-To-CpG Bisulfite Conversion kit (Applied Biosystems). Converted samples were used to amplify fragments of 513pb in exon 1 through the technique of PCR using the following primer pair 5'-TTGGTAGGTGGTTTTTTTTATTATT-3' (forward) and 5'-CTTAATCTCTACCTTAAAAACCTCC-3' (reverse). The sequences were obtained by Sanger sequencing and were analyzed in BIQ Analyzer© (Max Planck Institute for Informatics).

2.6 Statistical Analysis

Allelic and genotypic frequencies were calculated and tested for the Hardy-Weinberg equilibrium, which was performed using Haploview v. 4.2 software (IBM, USA). Data from the ABraOM Database (http://abraom.ib.usp.br/index.php) was used as a reference sample to compare our results with genetic variant frequencies from the general population. Comparison of the frequencies between individuals with TE and the reference database was performed by Fisher's Exact Test. Statistical tests were performed using SPSS® v. 18 (IBM, USA), considering as significant a value of $p < 0.05$.

3. RESULTS

3.1 One variant (rs59621536) was observed in HAND2 in TE individuals

In order to investigate a possible role of *HAND2* on genetic susceptibility to TE, we sequenced the coding region, exon 1 and exon 2, of the gene in 35 individuals with TE.

The clinical characteristics of the individuals affected by TE included in this study are described in Table 1. Among the individuals evaluated, 12% (3 out of 25 evaluated individuals) presented congenital cardiac defects. Of the 19 individuals evaluated for cardiac conditions, 10 (52%) presented cardiovascular diseases. It has been observed the development of hypertension in 7 out of 19 (36.8%) individuals, acute myocardial infarction in 2 out of 17 (11.8) and hemorrhagic stroke in 1 out of 17 (5.9%) subjects.

The two exons of the *HAND2* gene were sequenced in TE sample. Exon 1 and exon 2 were analyzed in 31 and 35 individuals, respectively. Sequencing analyses in both exons identified only one variant present in three individuals (9.7%; Table 2). The variant p.P51 (rs59621536) is located in exon 1 and leads to the exchange of a guanine to a cytosine at position 4: 173529137 (Figure 1C). The distribution of the variant p.P51 is in Hardy-Weinberg equilibrium. No difference of genotypic and allelic frequencies of this variant was found between TE and reference group of ABraOM ($p = 1.000$; Table 2).

3.2 p.P51 was not related to TE cardiac anomalies or diseases

Taking into consideration that HAND2 regulates VEGF, which is important for angiogenesis, variants in *HAND2* could lead to cardiovascular phenotypes through inadequate regulation of VEGF. Therefore, we next evaluated phenotypes related to cardiac malformations or diseases in the individuals carrying p.P51 variant (Supplementary Table 1). None individual with cardiac anomaly presented the variant. However, we did not have availability of such data for all subjects. There are clinical data of congenital anomalies of two individuals and information of late cardiac conditions of only one of the three individuals with the variant. We did not have access about congenital or cardiovascular condition of one subject presenting p.P51. The other individual with p.P51 has only upper limb anomalies, without neurological, ocular, auditory or cardiac congenital anomalies, and presents hypertension and one episode of acute myocardial infarction.

3.3 *In silico* analysis of *HAND2* variability confirms as a very conserved gene

We wanted to observe the number of variants in HAND2 and their association with diseases. Therefore, we performed an *in silico* molecular characterization of the gene *HAND2*.

We found a small number of variants located in *HAND2* gene, being few of them with minor allele frequency (MAF) higher than 0.01 (Figure 1A). The 3'UTR was the region with most variants and with the variants with higher MAF. In addition, the 5'UTR was the less variable region.

Most of the variants observed in *HAND2* were characterized with consequence as Non-sense mediated decay (NMD) transcript variants (29%) and intronic variants (29%; Figure 1B). Only few variants were characterized with consequences into the coding sequence (2%). In addition, no *de novo* mutations were described in *HAND2 (*denovo data base; denovo-db.gs.washington.edu*).* These data points to *HAND2 as* extremely conserved gene.

As HAND2 is important to cardiovascular development, we looked for variants previously associated to human phenotype (i.e.: malformations, late onset conditions). We found only two catalogued variants that were associated to any human phenotype. Both of them were associated to cardiovascular alterations. Variant rs10520260 (Figure 1C) is associated to atrial fibrillation and the variant rs1553974835 (Figure 1C) is associated to dilated cardiomyopathy 1A, located at 3'UTR and exon 1, respectively. Interestingly, we did not find in the databases, the missense variant in the position 202 of the protein (p.G202V) found by Khalil et al. (2017) (Figure 1C).

3.4 *In silico* analysis points p.P51 to possibly affecting mRNA structure and methylation site

Although p.P51 was not associated to neither susceptibility to TE nor cardiovascular conditions in this sample, we studied the functional consequences of the variant through *in silico* prediction as p.P51 could be associated to other diseases and conditions. In functional prediction, p.P51 variant was classified as neutral. The evaluations of splicing sites were not concordant between the two predictors used, one predicted that the variant affects splicing site, however, in the other no significant splicing site changes were detected by the variant. The predictions did not point to possible alterations of miRNA binding sites. In addition, the variant was observed to be present in the promoter region of another gene: *HAND2-AS1* (HAND2 antisense RNA 1), and it may alter some important region of the promoter of this gene.

It was observed alterations in two CpG islands adjacent to the variant, with the increase of one nucleotide in each CpG island. Moreover, in the analysis of changes in the binding site of transcription factors by variants, we observed the modification of the binding sites of the Methyl-CpG Binding Protein 2 (MeCP2) and the Pleomorphic Adenoma Gene Protein 1 (PLAG1) in the presence of the p.P51 polymorphism.

Prediction of mRNA structure showed a difference in the minimum free energy (MFE) of the secondary structure of *HAND2* mRNA with p.P51 variant (MFE = - 264.20 kcal/mol) when compared with the secondary structure of

HAND2 mRNA without p.P51 (MFE = - 264.50 kcal/mol). This difference points to a small gain of thermodynamic stability as the stability increases with the decreasing of MFE (ΔMFE = - 0.3 kcal/mol). Also, an alteration in a hairpin of *HAND2* mRNA can be observed in the presence of the variant p.P51 (Figure 2).

3.5 p.P51 did not appear to change methylation patterns in individuals with TE

Considering the potential of p.P51 in affecting two CpG islands, we conducted a DNA methylation analysis in DNA samples from TE individuals in the region near the variant. Thus, DNA methylation was compared between TE individuals carrying the variant p.P51 (n=3) with TE individuals without the variant (n=23).

The variant p.P51 is located in the 15th CpG dinucleotide of the sequence. Neither carriers nor individuals without this variant presented any DNA methylation in this dinucleotide (Figure 3A). In addition, we did not observe any alteration of the DNA methylation regions near the variant (Figure 3). From our results, the variant did not seem to alter the methylation pattern of the CpG islands. Still, we were able to detect methylation in other dinucleotides. DNA methylations were present in 11 of the 23 individuals without the variant (Figure 3B). Only 8 CpG dinucleotides were seen to be methylated in the 11 individuals without the variant (Figure 3B).

4. DISCUSSION

Despite HAND2 has already been associated with limb and cardiovascular malformations in animal models (Dai and Cserjesi, 2002; Srivastava et al., 1995), its link to limbs, heart and other Thalidomide outcomes is poorer investigated. So far, only one study encompassed HAND2 and TE. They observed an interaction between HAND2 and TBX5, which was suppressed in the presence of Thalidomide (Khalil et al., 2017). TBX5, which is also able to interact with Thalidomide (Khalil et al., 2017), is a transcription factor very important in heart and upper limp development. Thus, it is possible that HAND2 can play a part in the defects caused by Thalidomide, especially in the

susceptibility to the heart and limbs malformations observed in TE as the individuals are differentially affected by thalidomide.

We evaluated *HAND2* in individuals with TE and observed the presence of one variant, p.P51, located in exon 1. Through *in silico* analysis we observed that p.P51 could alter HAND2 mRNA structure and alter CpG islands near the variants, as well alter the binding site of MeCP2. Furthermore, *HAND2* showed to be a well conserved gene, presenting only a few variants in coding region and most of the variants located in regulatory regions, especially in 3'UTR. In addition, we found just two catalogued variants in *HAND2* previously associated to human phenotype, both being associated to cardiovascular diseases (Lu et al., 2016; Nielsen et al., 2018; Roselli et al., 2018).

The coding region of HAND2 was sequenced since previously the three variants associated to cardiovascular conditions were located in that region of the gene (Khalil et al., 2017; Lu et al., 2016; Sun et al., 2016). Furthermore, the connection between HAND2 and TBX5 occurs through the basic helix-loop-helix (Bhlh) domain located in HAND2, domain which is codified by the exon 1 of the gene (Khalil et al., 2017). We found the presence of only one variant, p.P51, in three individuals with TE. During the clinical data collection performed by Kowalski *et al.* (2015), many of the individuals reported not having been examined for internal abnormalities, which demonstrates that the observed frequency of cardiac abnormalities may have been underestimated. Also, it is estimated that about 40% of the affected by Thalidomide died before the first year of life, probably due to defects in internal organs such as the heart (Lenz, 1988; Smithells and Newman, 1992). It is possible that many cases of cardiac malformations and serious cardiac problems led to premature death or even miscarriages (Vargesson, 2009).

In accordance to the low variability we observed in *HAND2* in TE individuals, the gene showed to be highly conserved through the *in silico* molecular characterization. It is important from the evolution point of view that genes related to embryonic development to be highly conserved (Voth et al., 2009). Therefore, the limited variability of *HAND2* was expected since it plays an important role in the embryo development.

HAND2 is essential for normal formation especially of the heart. In fact, HAND2 acts regulating VEGF, a factor important to angiogenesis and that plays

a known role in the antiangiogenesis mechanism of thalidomide (Yabu et al., 2005; Yamagishi et al., 2000). Furthermore, the only two variants associated to human phenotype in *HAND2* were associated to cardiovascular conditions (Lu et al., 2016; Nielsen et al., 2018; Roselli et al., 2018). Therefore, HAND2 could act in the cardiovascular anomalies observed in TE, although we were not able to observe any association between p.P51 and cardiovascular phenotypes in the TE individuals.

Despite the fact that p.P51 did not present any association to TE individuals we characterized the possible effects of the variant. The p.P51 variant was observed to affect the MeCP2 binding site in HAND2. MeCP2 is part of a family of nuclear proteins with Methyl-CpG binding domain (MBD), which are related to recognition and binding to CpG islands and remodeling of chromatin (Bienvenu and Chelly, 2006). Mutations in *MeCP2* are the cause of most cases of Rett Syndrome, a disorder of neuronal development, which has a high mortality rate that are suspected to be due cardiac dysfunction (Bienvenu and Chelly, 2006; Goll and Bestor, 2005; Hara et al., 2015). The PLAG1 binding site was also altered by p.P51. However, PLAG1 binds on the forward strand, most likely acting on the regulation of *HAND-AS1*, unlike MeCP2, which has its binding site on the reverse strand, where *HAND2* is located.

Furthermore, the *in silico* predictions of the p.P51 variant showed that it can alter a specific hairpin in *HAND2* mRNA second structure with the addition of an internal loop. Internal loops are known as destabilizing elements which were seen to allow the connection of miRNAs into the mRNA (Zhao et al., 2005). Interestingly, it has been described a miRNA target site in HAND2 mRNA, although it is in the 3'UTR region, not near the internal loop observed (Zhao et al., 2005). Even though the internal loop is considered a destabilizing element, the small alteration in the MFE shows a small gain of thermodynamic stability in the structure of HAND2 mRNA (Trotta, 2014). Modifications in the structure of mRNAs can also affect the protein through disruption of protein expression and protein folding (Faure et al., 2017; Mauger et al., 2019). There is no data regarding alterations of the second structure of *HAND2* mRNA, therefore studies are needed to understand if this predicted alteration would have a biological impact.

According to *in silico* predictions, p.P51 could later CpG islands near the variant, therefore we wanted to observed if these alterations were seen in our sample of TE individuals. Although we did not observe any alteration in methylation patterns associated with the presence of the variant p.P51 it is important to notice that methylation patterns are time and tissue dependent (Reik et al., 2001). The samples used in this study were collected later in the life of the TE individuals, therefore the methylation alterations caused by exposure to thalidomide may no longer be observed. Moreover, p.P51 could cause changes in methylation patterns in others cardiovascular diseases that are not observed in TE individuals.

Our study faced some limitations specially regarding the small sample size, with few cases of cardiovascular conditions, which became difficult to establish a genotype-phenotype correlation. Another limitation was the absence of the ideal comparison group for the analyses. Furthermore, we were not able to validate the predicted alterations in HAND2 mRNA by the p.P51 in TE individuals.

5. CONCLUSION

We observed that *HAND2* is a well conserved gene through the *in silico* molecular characterization and evaluation in a sample of TE individuals, in which we observed only one variant. We hypothesize that the p.51 variant may lead to modifications in the epigenetic regulation of the *HAND2* gene, since alteration of the CpG islands and the MeCP2 binding site have been observed. The variant did not appear to affect the protein structure of HAND2, although it was predicted to alter mRNA structure. Although it is a synonymous variant and predicted as neutral, it might lead to an additional risk factor in the presence of Thalidomide because of the interaction between the genotype and the teratogen, however, more studies are necessary to verify this hypothesis.

5. FUNDINGS

This work was supported by National Institute of Population Medical Genetics (INAGEMP), Grant/Award Number: (CNPq 573993/2008-4); National Council for Scientific and Technological Development (CNPq 423249/2016-9)

and Research and Events Incentive Fund of Hospital de Clínicas de Porto Alegre - FIPE/HCPA (10-0244).

6. ACKNOWLEGMENTS

The authors acknowledge INAGEMP – National Institute of Population Medical Genetics [grant CNPq 573993/2008- 4], National Council for Scientific and Technological Development, FIPE/HCPA (10-0244) and Brazilian Association of People Affected by Thalidomide Syndrome - ABPST for the support provided for this project.

7. CONFLICT OF INTEREST

None declared.

8. REFERENCES

ALVAREZ-SAAVEDRA, M. *et al.* Elevated expression of MeCP2 in cardiac and skeletal tissues is detrimental for normal development. **Human Molecular Genetics**, v. 19, n. 11, p. 2177–2190, 2010.

BIENVENU, T.; CHELLY, J. Molecular genetics of Rett syndrome: When DNA methylation goes unrecognized. **Nature Reviews Genetics**, v. 7, n. 6, p. 415–426, 2006.

CLOWES, C. *et al.* The functional diversity of essential genes required for mammalian cardiac development. **genesis**, v. 52, n. 8, p. 713–737, 1 ago. 2014.

COETZEE, S. G.; COETZEE, G. A.; HAZELETT, D. J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. **Bioinformatics (Oxford, England)**, v. 31, n. 23, p. 3847–9, 1 dez. 2015.

D'AMATO, R. J. *et al.* Thalidomide is an inhibitor of angiogenesis. **Proceedings of the National Academy of Sciences**, v. 91, n. 9, p. 4082–4085, 1994.

DAI, Y.-S.; CSERJESI, P. The basic helix-loop-helix factor, HAND2, functions as a transcriptional activator by binding to E-boxes as a heterodimer. **The Journal of biological chemistry**, v. 277, n. 15, p. 12604–12, 12 abr. 2002.

FAURE, G. *et al.* **Adaptation of mRNA structure to control protein foldingRNA Biology**Taylor and Francis Inc., , 2 dez. 2017. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/28722509/>. Acesso em: 15 fev. 2021

GOLL, M. G.; BESTOR, T. H. Eukaryotic Cytosine Methyltransferases. **Annu. Rev.Biochem.**, v. 74, p. 481–514, 2005.

GRUNERT, M. *et al.* Comparative DNA methylation and gene expression analysis identifies novel genes for structural congenital heart diseases. **Cardiovascular Research**, v. 112, p. 464–477, 2016.

HARA, M. *et al.* Disturbance of cardiac gene expression and cardiomyocyte structure predisposes Mecp2-null mice to arrhythmias. **Scientific Reports**, v. 5, n. June, p. 1–17, 2015.

JING-BIN, H. *et al.* Molecular mechanisms of congenital heart disease. **Cardiovascular Pathology**, v. 19, n. 5, p. e183–e193, 1 set. 2010.

JURAND, A. Early changes in limb buds of chick embryos after thalidomide treatment. **Embryol. exp. Morph**, v. 16, n. 2, p. 289–300, 1966.

KHALIL, A. *et al.* A HAND to TBX5 Explains the Link Between Thalidomide and Cardiac Diseases. **Scientific Reports**, v. 7, n. 1, p. 1416, 2017.

KOWALSKI, T. W. *et al.* Thalidomide embryopathy: Follow-up of cases born between 1959 and 2010. **Birth Defects Research Part A - Clinical and Molecular Teratology**, v. 103, n. 9, p. 794–803, 2015.

KYPREOS, K. E. *et al.* Pharmacological Management of Dyslipidemia in Atherosclerosis : Limitations , Challenges , and New Therapeutic Opportunities. 2018.

LENZ, W. A short history of thalidomide embryopathy. **Teratology**, v. 38, n. 3, p. 203–215, 1988.

MATTHEWS, S. J.; MCCOY, C. Peginterferon alfa-2a: A review of approved and investigational uses. **Clinical Therapeutics**, v. 26, n. 7, p. 991–1025, 2003.

MAUGER, D. M. *et al.* mRNA structure regulates protein expression through changes in functional half-life. **Proceedings of the National Academy of Sciences of the United States of America**, v. 116, n. 48, p. 24075–24083, 26 nov. 2019.

MCBRIDE, W. Thalidomide and congenital abnormalities. **Lancet**, v. 2, p. 1358, 1961.

MCDERMOTT, D. A.; FONG, J. C.; BASSON, C. T. **Holt-Oram Syndrome**. [s.l.] University of Washington, Seattle, 1993.

OSTERWALDER, M. *et al.* HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. **Developmental cell**, v. 31, n. 3, p. 345–57, 10 nov. 2014.

PARMAN, T.; WILEY, M. J.; WELLS, P. G. Free radical-mediated oxidative DNA damage in the mechanism of thalidomide teratogenicity. **Nature medicine**, v. 5, n. 5, p. 582–5, 1 maio 1999.

RAJKUMAR, S. V.; BLOOD, E. Lenalidomide and venous thrombosis in multiple myeloma. **The New England journal of medicine**, v. 354, n. 19, p. 2079–80, 11 maio 2006.

REIK, W.; DEAN, W.; WALTER, J. **Epigenetic reprogramming in mammalian developmentScience**American Association for the Advancement of Science, , 10 ago. 2001. Disponível em:

<https://science.sciencemag.org/content/293/5532/1089>. Acesso em: 16 abr. 2021

SALDANHA, P. H. A tragédia da Talidomida e o advento da teratologia experimental. **Revista Brasileira de Genética**, v. 17, p. 449–464, 1994.

SHARDEIN, J. Psychotropic Drugs. **Chemically Induced Birth Defects**, n. 2nd, p. 208–270, 1993.

SHENG, W. *et al.* LINE-1 methylation status and its association with tetralogy of fallot in infants. **BMC medical genomics**, v. 5, p. 20, 6 jun. 2012.

SHESKIN, J. THALIDOMIDE IN THE TREATMENT OF LEPRA REACTIONS. **Clinical pharmacology and therapeutics**, v. 6, p. 303–6, 1965.

SMITHELLS, R. W.; NEWMAN, C. G. Recognition of thalidomide defects. **Journal of medical genetics**, v. 29, n. 10, p. 716–23, out. 1992.

SRIVASTAVA, D.; CSERJESI, P.; OLSON, -F-ERIC N. A Subclass of bHLH Proteins Required for Cardiac Morphogenesis. **Science**, v. 270, p. 1995–1999, 1995.

THERAPONTOS, C. *et al.* Thalidomide induces limb defects by preventing angiogenic outgrowth during early limb formation. **PNAS**, v. 106, n. 21, p. 8573–8578, 2009.

TROTTA, E. On the Normalization of the Minimum Free Energy of RNAs by Sequence Length. **Citation: Trotta E**, v. 9, n. 11, p. 113380, 2014.

VARGESSON, N. Thalidomide-induced limb defects: Resolving a 50-year-old puzzle. **BioEssays**, v. 31, p. 1327–1336, 2009.

VIANNA, F. S. L. *et al.* Epidemiological surveillance of birth defects compatible with thalidomide embryopathy in Brazil. **PloS one**, v. 6, n. 7, p. e21735, 2011.

VIANNA, F. S. L. *et al.* Recognition of the phenotype of thalidomide embryopathy in countries endemic for leprosy: New cases and review of the main dysmorphological findings. **Clinical Dysmorphology**, v. 22, n. 2, p. 59–63, 2013.

VOTH, H. *et al.* Co-regulated expression of HAND2 and DEIN by a bidirectional promoter with asymmetrical activity in neuroblastoma. **BMC molecular biology**, v. 10, p. 28, 6 abr. 2009.

YAMAGISHI, H.; OLSON, E. N.; SRIVASTAVA, D. The basic helix-loop-helix transcription factor, dHAND, is required for vascular development. **The Journal of clinical investigation**, v. 105, n. 3, p. 261–70, fev. 2000.

ZHAO, Y.; SAMAL, E.; SRIVASTAVA, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. **Nature**, v. 436, p. 214–220, 2005.

Table 1: Clinical characterization of individuals with thalidomide embryopathy.

Characteristics		n	%
Gender			
	Male	22/35	(63)
Congenital Abnormalities			
	Upper limb	24/25	(96)
	Lower limb	12/25	(48)
	Neurological	3/24	(12.5)
	Ocular	8/25	(32)
	Auditory	4/23	(17.4)
	Cardiac*	3/25	(12)
Cardiovascular Diseases			
	All	10/19	(52)
	Hypertension	7/19	(36.8)
	Myocardial Infaction	2/17	(11.8)
	Hemorragic stroke	1/17	(5.9)

*Clinical data are not available for all individuals included in this study.

Table 2: Genotypic and allelic frequencies of rs59621536 variant located in exon 1 of *HAND2* gene of individuals affected by TE compared to the frequencies obtained from ABraOM database.

rs59621536	**TE**		**ABraOM[1]**		
	N	%	n	%	p-Value[2]
GG	28	(90.3)	537	(88.4)	1.000
GC	3	(9.7)	67	(11.03)	
CC	0	(0)	3	(0.5)	
G	59	(95.2)	1,141	(94)	1.000
C	3	(4.8)	73	(6)	

[1]ABraOM Database (http://abraom.ib.usp.br/index.php).
[2]Exact Fisher Test.

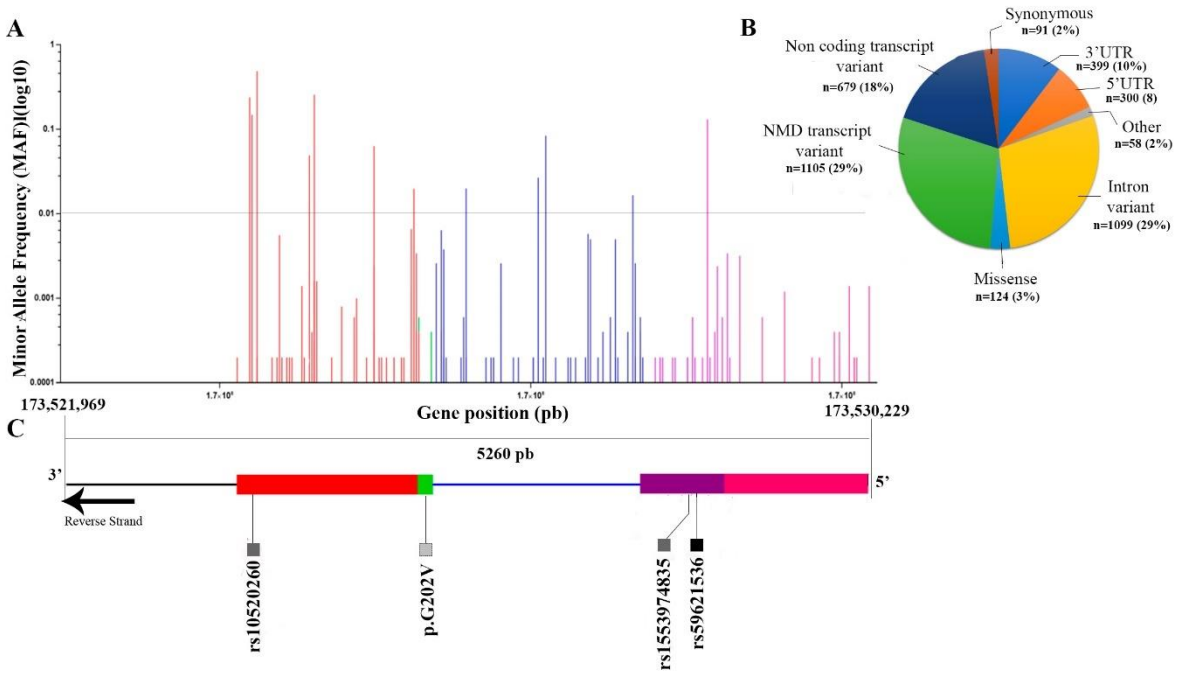


Figure 1: *In silico* molecular characterization of *HAND2* gene. Minor Allele Frequency (MAF) of variants observed in *HAND2* gene. Above the line in the middle of the graphic are the variants with MAF higher than 0,01 (polymorphisms) **(A).** Consequence of variants located in *HAND2.* Among 'Others' are: coding sequence (n=5), frameshift (n=6), inframe (n=15), splice variant (n=27), stop gained (n=3) and stop retained (n=2) **(B).** Scheme representing *HAND2* gene with variants associated to some phenotype (dark grey), the variant observed by Khalil *et al. (2017)* (light grey) and the variant observed in individuals with TE, p.P51 (black) **(C)**. 5'UTR (pink); exon 1 (purple); intron 1 (blue); exon 2 (green); 3'UTR (red).

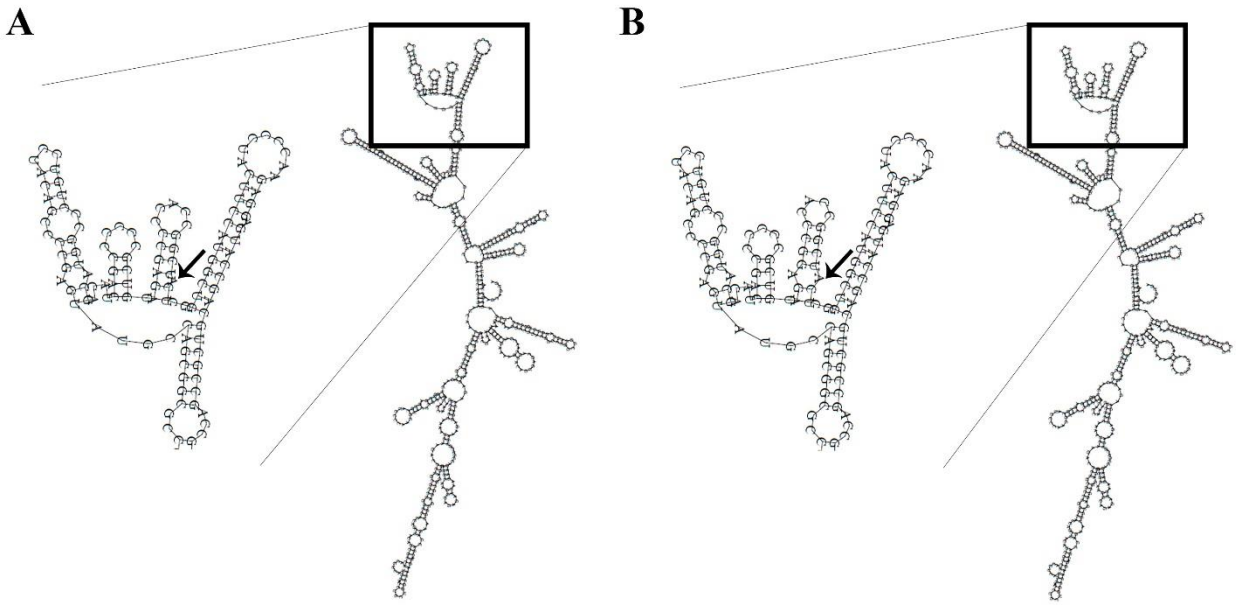


Figure 2: Functional prediction analysis of the impact of p.P51 variant in the mRNA of *HAND2*. Minimum free energy secondary structure of *HAND2* mRNA (MFE = - 264.20 kcal/mol) **(A).** Minimum free energy secondary structure of *HAND2* mRNA with the p.P51 variant (MFE = - 264.50 kcal/mol) **(B).** It can be observed an alteration of the structure of one hairpin in the presence of the variant (**arrow**).

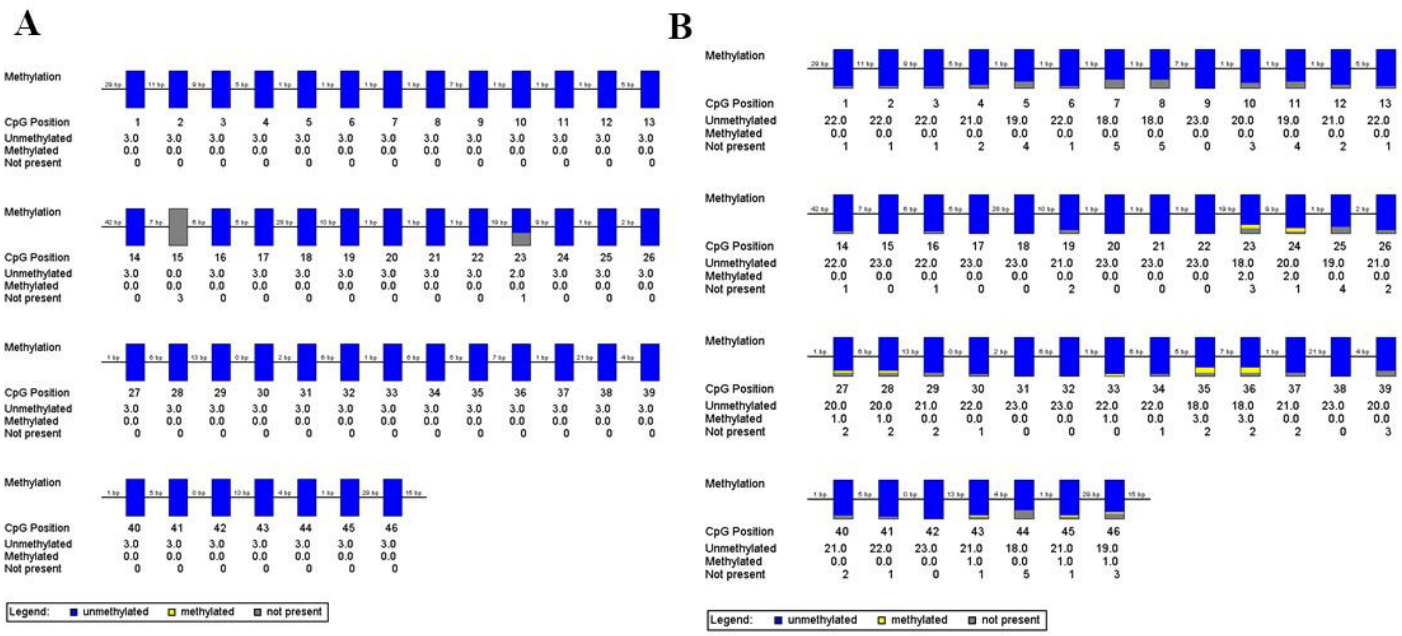


Figure 3: DNA Methylation analysis of CpG islands in the sequence near the variant p.P51 in exon 1 of *HAND2* in individuals with **(A)** and without the variant p.P51 **(B).** The variant p.P51 is located in the 15th CpG dinucleotide of the analyzed sequence.

Supplementary material

Supplementary Table 1: Cardiac clinical characterization of individuals with the variant p.P51.

Individuals	1	2	3
Cardiac Congenital Anomalies	No anomalies	No data	No data
Late Cardiac conditions	No data	Hypertension; Hemorragic stroke	No data

CAPÍTULO IV

ARTIGO II

Possible new targets of Thalidomide related to limbs and cardiac defects: a systems biology approach

Manuscrito em preparação

BMC Medical Genomics

Possible new targets of Thalidomide related to limbs and cardiac defects: a systems biology approach

Bruna Duarte Rengel[1,2,4,6]; Lavínia Schuler-Faccini[1-4]; Lucas Rosa Fraga[1,3,4,5]; Fernanda Sales Luiz Vianna[1-4,6]; Thayne Woycinck Kowalski[1,2,3,6-8]

[1]Laboratory of Medical Genetics and Evolution, Genetics Department, Federal University of Rio Grande do Sul, Porto Alegre, Brazil; [2]Postgraduate Program in Genetics and Molecular Biology, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil; [3]National Institute of Population Medical Genetics (INAGEMP), Porto Alegre, Brazil; [4]Brazilian Teratogen Information Service (SIAT), Medical Genetics Service of Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; [5]Department of Morphological Sciences, Institute of Health Sciences, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.[6]Genomic Medicine Laboratory at Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, Brazil; [7]Bioinformatics Core, Hospital de Clínicas de Porto Alegre, HCPA, Porto Alegre, Brazil; [8]Centro Universitário CESUCA, Cachoeirinha, Brazil.

Corresponding Author: Thayne Woycinck Kowalski, Genomic Medicine Laboratory, Hospital de Clínicas de Porto Alegre (HCPA), Ramiro Barcelos Street, 2350, CEP: 90035-903, Porto Alegre, Brazil. Tel: +55 51 984196181. E-mail: thaynewk@gmail.com.

Abstract

Background: Thalidomide is a well-known teratogen that causes the thalidomide embryopathy (TE), which consists in malformations especially in the heart and limbs. Its mechanism of teratogenicity is still not fully elucidated. Recently, a new target of thalidomide was described, TBX5, and was observed a new interaction between HAND2 and TBX5 that is disrupted in the presence of thalidomide. Therefore, our study aimed to select new genes of interest, through systems biology approaches evaluating HAND2 and TBX5 interaction and heart and limbs malformations of TE, which could play a role in the TE teratogenesis.

Methods: The genes and proteins were selected through TF2DNA, REACTOME, Human Phenotype Ontology and InterPro databases. The networks were assembled using STRING © database v11.0. Network analysis were performed in Cytoscape © software v.3.7.1 and R v3.6.2.

Results: We constructed a network for HAND2 and TBX5 interaction (Mater Network); a network for heart and limbs malformations of TE (Cardiac and Limbs network, respectively); and two joined networks Mater/Cardiac and Mater/Limbs network. We observed that EP300 protein seemed to be important in all the networks. We also looked for proteins containing C2H2 domain in the assembled networks that could be marked for degradation in the presence of thalidomide. ZIC3, GLI1, GLI3, ZNF148 and PRDM16 were the ones present in both heart an limbs malformations of TE networks. Furthermore, we investigated gene differentially expressed after treatment with thalidomide, and observed three genes, *FANCB*, *ESCO2* and *XRCC2*, to be downregulated and present in Cardiac and Limbs networks.

Conclusions: Our study was able to point to different new proteins and genes to be further evaluated in the TE teratogenicity. We used different strategies, and selected specially EP300, which was important in all the analyzed networks, as a good candidate to be further studied in TE teratogenicity.

Ketwords: HAND2; TBX5; Systems biology; thalidomide embryopathy; teratogenesis.

1. Background

Thalidomide is a drug well known for its high teratogenicity potential. It was synthesized in Germany in the 1950s and caused the birth of more than 10,000 children with rare congenital anomalies (1–3). The malformations caused by thalidomide are known as thalidomide embryopathy (TE) and are characterized especially by limbs malformations, also affecting eyes, ears, heart and other organs (1).

In the 1960s, after its use being associated with the cause of the anomalies, thalidomide was withdrawn from the world market. However, still in the 1960's it was shown thalidomide had the potential to treat an inflammatory condition called erythema nodosum leprosum (ENL) (4). From this discovery, further studies found thalidomide being effective on the treatment of several other conditions, including multiple myeloma and graft versus host disease (5,6). Unfortunately, despite strict regulations, the use of thalidomide is still leading to report of recent cases of TE in Brazil (7–9).

Although the teratogenic potential of thalidomide is well known, its molecular mechanisms are not yet fully understood. Many hypotheses have been raised over the years, being three the most accepted: disturbance of angiogenesis (10), induction of oxidative stress (11), and its binding to Cereblon protein, modifying the ubiquitination of some proteins from $CRL4^{CRBN}$ complex (12). In the recent years, new affected proteins and pathways have been proposed. Transcription factors (TFs) containing a C2H2 domain were observed to be neosubstrates of the complex thalidomide-$CRL4^{CRBN}$ (13). Recently, TBX5 (T-box 5) was shown to be a target of thalidomide (14). TBX5 was observed to bind to Heart- and neural crest derivatives-expressed 2 (HAND2), and this connection is lost in the presence of thalidomide (14). Interestingly, pathogenic variants in *TBX5* gene are associated with Holt-Oram Syndrome, of which TE is a phenocopy because of its similarities in heart and limbs malformations (15). TBX5 and HAND2 are important TFs for embryo development, enrolled in heart and limbs development (16–18). HAND2 has been associated specially with cardiovascular anomalies (19–21). Moreover, in a knockout study it was demonstrated mice null-

homozygotes for *HAND2* had impaired ventricle development and died because of cardiac failure (20).

HAND2 and TBX5 action in development is already well described, especially in heart and limbs (16–18). However, it is still not known how the impairment of HAND2 and TBX5 interaction caused by thalidomide could explain TE. Thus, molecular pathways affected by HAND2-TBX5 interaction disruption remains to be clarified. Therefore, our aim was to enlighten how the connection between HAND2 and TBX5 could be involved in TE's limbs and heart malformations by evaluating their biological network through systems biology approaches and to point to new targets of interest that could be further studied in TE.

2. Methods

2.1 Gene and Protein selection

Genes and proteins were retrieved from selection in TF2DNA (22), REACTOME (23), Human Phenotype Ontology (HPO) (24) and InterPro © (EMBL-EBI), always applying filter to *Homo sapiens*. HUGO Gene Nomenclature Committee (HGNC) was used to access the group of each gene obtained from the analysis. We selected genes and proteins that interact, regulate or are regulated by TBX5 and HAND2; genes associated to phenotypes observed in TE; and proteins with C2H2 domain.

2.2 Network assembling

Protein-protein interaction networks for proteins of interest were assembled on STRING © database v11.0 (STRING Consortium 2020), comprising only *Homo sapiens* interactions and selecting only experimental data (filter "experiment"). A confidence level of 0.4 (intermediate) was used, since this is an exploratory study.

2.3 Network analysis

The analysis of the networks was performed using Cytoscape © software v.3.7.1 (2001-2018 Cytoscape Consortium). Network metrics analysis included the evaluation of the betweenness centrality and closeness centrality for all the

nodes (proteins). The closeness centrality is a metric that demonstrates how fast the information flows in the network; nodes with high closeness centrality are in favorable location to the information flow. Merging of the networks through intersection and enrichment ontology analysis of communities were performed using Cytoscape © software v.3.7.1 (2001-2018 Cytoscape Consortium) through *DyNet* (25) and *ClueGo* (26) apps, respectively. All the other analysis were performed using R v3.6.2, as described in sequence. Enrichment ontology analysis was performed using the *ClusterProfiler* package (27). The mainhub was analyzed through *keyplayer* package. This metric indicates the most central node to the network, responsible to quickly diffuse and receive information (28). Hub, Authority and Community analyses were performed using the *igraph* package (29). Hub analysis points to the nodes that direct the information to the authorities, which contain the important information of the network.

2.4 Differential gene expression analysis

We selected the study GSE63935 through Gene Expression Omnibus (GEO) database (National Center of Biotechnology and Information, NCBI) to perform differential gene expression analysis since this study included thalidomide treatment in human cells, more specifically human embryonic stem cells (30). Differential gene expression analysis was performed in R, applying the robust multiaverage (RMA) normalization, using *edgeR* package. It was considered as upregulated the genes with logFC > 1.5 and an adjusted p-value for false discovery rate (FDR) < 0.05, and downregulated the genes with logFC < -1.5 and the same adjusted p-value for FDR.

3. Results

To enlighten how the connection between HAND2 and TBX5 could play a role in TE's heart and limbs malformations, we evaluated biological networks and proteins/genes selected thought different approaches (Figure 1). First, we assembled the networks of HAND2 and TBX5 interaction and of heart and limbs TE malformation individually. Secondly, we performed the analysis in each network.

With Mater network we wanted to evaluate HAND2 and TBX5 biological network in a context of interaction between them. Then we introduced TE in the context by assembling a network for cardiac malformations and for limbs malformations, independently, and after, analyzing together with Mater network. We then performed network analysis of the final networks; expression analysis to see the differential expressed genes in each network; and looked for C2H2 domain proteins in each network (Figure 1).

3.1 BAIAP2, TBX5, HEY2, HAND2 and TCF4 proteins are common points between HAND2 and TBX5 networks

In order to evaluate the genes/proteins important to maintain the network connections of HAND2 and TBX5 when they are interacting, we assembled protein-protein interaction networks, evaluating all genes and transcription factors that could regulate or be regulated by HAND2 and TBX5.

First, two networks were assembled: one containing genes selected through TF2DNA and REACTOME databases (TR network; Figure 2A) and other resulting from genes and proteins present in HAND2 and TBX5 interaction network (HTI network; Figure 2B). These databases give us genes and proteins related to the biological actions of HAND2 and TBX5. The HTI network was assembled directly on STRING using only HAND2 and TBX5 as an entry.

In order to direct the analysis to the interaction of HAND2 and TBX5 and to connect TBX5 to the main network we unified TR and HTI networks, forming a network named Mater, which represented the biological network of HAND2 and TBX5 in a context of interaction between them (Figure 2C). This network was used in all subsequent steps (Figure 1). We observed that the proteins BAIAP2, TBX5, HEY2, HAND2 and TCF4 were the common nodes of the networks. HEY2 and TCF4 are of the same group of proteins as HAND2, they are part of Basic helix-loop-helix proteins (Additional File 1).

Enrichment ontology analysis was performed on Mater network to verify whether the proteins selected were acting in process of development and related to TE. From GO, 205 ontologies were enriched, the 20 most statistically significant are listed in Figure 2D. From the 205 ontologies, 30 had HAND2 and

TBX5 acting together and includes “stem cell differentiation”, processes of heart development, processes of limb development and “pattern specification process” (Additional File 2). Moreover, the ontologies observed are mostly related to the beginning of development. This show us that the networks are adequate and have biological plausibility. In addition, KEGG analysis pointed to ontologies specially in cancer (Additional File 2).

These results show that other than the direct connection between HAND2 and TBX5 described by Khalil(14), the two proteins and its biological networks can also interact through BAIAP2, HEY2 and TCF4. Furthermore, the biological networks of HAND2 and TBX5, called here as Mater, had ontologies related to important developmental processes that can also be important to the teratogenicity of thalidomide, causing TE.

3.2 TCF4, UBB and JUN proteins are important to maintain the connection of HAND2 and TBX5 network

We further analyzed the Mater network to observe which were the nodes important to maintain its connection (Figure 1). Community analysis was performed. This analysis gives us the communities of genes clustered together by a topological value. The analysis showed the presence of 8 communities (Supplementary Figure 1A, Additional File 3), the connector hubs, nodes connecting the different communities, are show in Figure 2E. Enrichment ontology analysis in the communities found in the network showed only three communities having an ontology enriched, which included “regulation of transcription regulatory region DNA binding”, “ligand-activated transcription factor” and “translation regulator activity” (Supplementary Figure 1B, Additional File 3).

The proteins that stood out by high scores in the network analysis are shown in Figure 2E and the results of closeness centrality and betweenness centrality are in Supplementary Figure 1C-D (Additional File 3). TCF4, JUN and UBB presented high values of betweenness and closeness centralities. TCF4 was the main-hub and hub of the network, which means it is an important node to quickly diffuse and receive information pointing to the authority of the network, also was a node important to the connection between the networks HTI and TR.

The authority was the node of CTNNB1; therefore, it contains most of the information of the network, and was also a connector hub between the communities. Nonetheless, none of the proteins belonged to the same group of proteins (Additional File 1).

Taken all together, TCF4, JUN and UBB showed to be highly connected nodes and important for the control of information flow of the network. Their disruption would impact in the network connection, disbalancing the biological network of HAND2 and TBX5. Interestingly, HAND2 and TBX5 metrics did not stood out in the analysis performed. Although HAND2 has the second highest score of authority (0.59), it is still low when compared to the authority of the network CTNNB1 (1.00). Moreover, HAND2 is not a central node, nor responsible to transmit great part of the network information, so as TBX5. The two proteins do not have much influence in maintain the connections of the network or in the transmission of information when compared with JUN, UBB and TCF4. They also are not responsible for connecting the communities observed. Therefore, according to the criteria utilized to select the nodes, HAND2 and TBX5 did not appear to be important nodes to maintain the structure of their own network.

3.3 EP300 is important to both networks of cardiac and limbs malformations of TE

We also evaluated how HAND2 and TBX5 could be playing a part in the heart and limbs malformations observed in TE (Figure 1). For that, we selected the phenotypes of heart and limbs malformations of TE using literature (Table 1) (1,31,32). Using HPO database, the genes associated with those phenotypes were selected and used to assemble two different networks: Cardiac network and Limbs network, using only the proteins selected as an entry on STRING.

Network analysis were performed in both networks to find which are the nodes important to the connection of the different heart and limbs malformations compatible with TE (Figure 3C; Supplementary Figure 2, Additional File 3).

We found EP300 as the main-hub of Cardiac network and the node with higher closeness centrality on Limbs network, being important to information flow in both networks. The hub and authority of Cardiac network were NDUFA13 and NDUFS7, respectively, both proteins are part of the group of NADH:ubiquinone

oxidoreductase proteins (Additional File 1). The node with higher closeness centrality in Cardiac network was MED12. In Limbs network, hub and authority was BRCA1 and FANCM, respectively, both are part of FA complementation groups of proteins (Additional File 1). SMARCA4 was the main-hub. SMARCA4 and EP300 are part of Bromodomain containing group of proteins (Additional File 1). EP300 was the only node important in both networks, Cardiac and Limbs. An abnormal function of EP300 could impact in both networks at the same time, therefore it is a good protein of interest to be studied for the teratogenicity of thalidomide in the heart and limbs.

3.3 EP300 and FANCM are key nodes to disrupt the Mater/Cardiac and Mater/Limbs networks.

We next analyzed how the biological network of the interaction between HAND2 and TBX5 (Mater network) could affect in the cardiac and limbs malformations of TE (Cardiac and Limbs networks) (Figure 1). For that, we merged the Mater network with Cardiac or Limbs networks, forming two new networks; Mater/Cardiac and Mater/Limbs (Figure 1 and Figure 3A-B). The connection between Cardiac and Mater network was through the nodes of FANCM, SRCAP, EP300, GATA4, TBX5 e NKX2-5 (Figure 4A). Interestingly, FANCM, SRCAP, EP300 and TBX5 were also connecting Mater network with Limbs network, together with SMARCA4, MECP2 e RUNX2 (Figure 4B). FANCM, SRCAP, EP300 and TBX5 could represent the link between the network of interaction of HAND2 and TBX5 and the malformations in the heart and in the limbs of TE.

We performed community analysis in both Mater/Cardiac and Mater/Limbs, to look for the connector hubs that are connecting the different communities, and main-hub analysis (Figure 4C). EP300 was the main-hub of both Cardiac/Mater and Limbs/Mater networks, seeming to be important to maintain the information flow between the networks of malformations with HAND2 and TBX5 network. In community analysis, Cardiac/Mater network resulted in 13 communities with 49 connector hubs, among them were FANCM, SRCAP and EP300. For Limbs/Mater were 17 communities with 82 connector hubs, among them were all

the nodes that connected Mater network with Limbs network. It is important to notice that EP300 and FANCM also resulted as important to maintain the network of heart and limbs malformations independent of HAND2 and TBX5.

Taken all together EP300 and FANCM can represent key proteins to disrupt the network of HAND2 and TBX5 along with heart and limbs malformations of TE because they connected Mater network to Cardiac and also Limbs network, together with being important to maintain the connection of both Mater/Cardiac and Mater/Limbs. Therefore, they are interesting candidates to be evaluated in the connection of HAND2 and TBX5 to the malformations of heart and limbs of TE.

3.4 Mater, Cardiac and Limbs networks contains proteins with C2H2 domain

It is known that one of the main targets of thalidomide is the protein Cereblon, that is part of $CRL4^{CRBN}$ E3 ubiquitin ligase complex (12,33). Some studies have pointed out that when thalidomide binds to Cereblon, proteins containing C2H2 domain become new targets of ubiquitination of the complex, therefore are marked for degradation (13,34). Based on that, we wanted to know if there were proteins containing C2H2 domain in Cardiac, Limbs and Mater networks that could be marked for degradation by the $CRL4^{CRBN}$ complex and therefore impact in the normal functioning of that biological network.

In Mater network the proteins BCL6, ZNF133 and ZNF281 contains C2H2 domains (Table 2; Figure 2). None of these proteins showed to be relevant in the network analysis of Mater network nor were important to its connection to Cardiac or Limbs networks. However, BCL6 and ZNF281 were in the same community as CTNNB1 and JUN, which resulted to be important to maintain the integrity of the connections of Mater network. BCL6 is a protein regulated by HAND2 and was present in the enriched ontologies of Mater network: regulation of hemopoiesis; positive regulation of neurogenesis; lymphocyte differentiation; B cell differentiation; positive regulation of neuron differentiation; myeloid cell differentiation; B cell activation; regulation of Notch signaling pathway; notch signaling pathway. ZNF281 regulates HAND2 and participates in the enriched

ontology of stem cell differentiation. ZNF133 did not appear in a community of any important node, nor participates of an enriched ontology of the network.

In Cardiac and Limbs networks there were 14 proteins containing C2H2 domains (Table 2; Figure 3). ZIC3, GLI1, GLI3, ZNF148, and PRDM16 were in common between the two networks. None of these proteins showed to be significant in the network analysis of Cardiac or Limbs networks. Despite that, because they are part of cardiac and limbs network, these proteins represent possible neosubstrates of $CRL4^{CRBN}$-Thalidomide complex that could impact in the heart and in the limbs at the same time.

To understand in which ontologies these C2H2 proteins are included, we performed enrichment ontology analysis of the 17 C2H2 proteins presented in the networks. There were 143 enriched ontologies (Additional File 4). Among the enriched ontologies were: “regulation of cartilage development”, “chondrocyte differentiation”, “embryonic limb morphogenesis”, “limb morphogenesis”, “limb development” and “cartilage development”. Ontologies related to the heart development were also enriched: “positive regulation of cardiac muscle cell proliferation”; “positive regulation of cardiac muscle tissue growth”; “positive regulation of heart growth”; “cardiac muscle cell proliferation”. This implies that if these proteins were degraded it could cause impairments on to limb and heart development.

3.5 FACNB, ESCO2 and XRCC2 were downregulated and its proteins are part of both Cardiac and Limbs network.

Finally, we saw how thalidomide affects the expression of the genes present in the analyzed networks. To this, we selected data (GSE63935) from Gene Expression Omnibus (GEO) database (National Center of Biotechnology and Information, NCBI)(30). The study has performed gene expression analysis in human embryonic stem cells treated with thalidomide and compared to control group. We selected the samples analyzed two days after treatment. All the characteristics of the assay is available in the GEO database.

From our secondary analysis, it was found that among the proteins of the networks Cardiac, Limbs and Mater, seven genes were differentially expressed (Table 2). In Mater network, there was no protein in which its gene was found to be differentially expressed.

In Cardiac network, the genes *FANCB, ESCO2* and *XRCC2* were downregulated. FANCB and XRCC2 proteins are part of FA complementation groups of proteins together with FANCM and BRCA1 (Additional File 5). BRCA1 appeared to be important in Limbs network and FANCM were important to maintain the connection of Mater/Cardiac and Mater/Limbs. ESCO2 is part of Lysine acetyltransferases group of proteins together with EP300, KAT2B, MECOM and PRDM16 (Additional File 1).

CDC45, CDC6 and *ORC1* were downregulated and its proteins were part of Limbs network. The protein ORC1 is part of the group of Origin recognition complex proteins. CDC45 is part of the CMG helicase complex proteins group. Only one gene was upregulated and had its protein in some network, *COMP*, and it was part of Limbs network. CDC6 and COMP had none group of proteins predicted.

To understand in which ontologies these differentially expressed genes were included, we performed enrichment ontology analysis for the downregulated genes (Additional File 5). There were 38 enriched ontologies and they were related especially to mitotic process, DNA repair, DNA replication process and Cell cycle process. Because there was only one gene upregulated, we looked directly on *amigo* © database and observed that *COMP* has ontologies related to limbs development, skeletal and chondrocyte development, apoptotic process and others.

It is important to notice that none of the differentially expressed genes had its proteins as important nodes in any of the network analysis performed, nor are proteins containing C2H2 domain.

4 Discussion

The mechanism of action of thalidomide's teratogenicity is not completely described, some gaps are in need to be filled and computational approaches can help to look for candidate to be further evaluated in experimental studies. Knowing that the connection between HAND2 and TBX5, two transcription factors important in development, is lost in the presence of thalidomide(14), raise the question of how this connection could be playing a role in TE. Moreover, new candidates can be pointed out when considering this connection and biological network of HAND2 and TBX5 when evaluated in malformations of TE. Our study proposed to use systems biology approaches to understand how the link between HAND2 and TBX5 could be playing a role in heart and limbs malformations related to TE and to seek for new candidate genes/proteins possibly involved to the occurrence of TE.

The main findings of these study were candidate targets that could be further evaluated in differential susceptibility to TE. Among the ones that are part of the interaction between HAND2 and TBX5 is EP300, that appeared to be important in almost all the network analysis. We found C2H2 domain proteins that are part if this interaction and could be a neosubstrate for $CRL4^{CRBN}$-thalidomide complex, for example, ZIC3, GLI1, GLI3, ZNF148 and PRDM16, that were in acting in heart and limbs malformation network. Moreover, we found three genes*, FACNB, ESCO2* and *XRCC2*, that were downregulated after treatment with thalidomide and were part of heart and limbs malformations networks.

When only evaluating HAND2 and TBX5 network (Mater) we observed that TCF4, JUN and UBB were important to maintain the network. Interestingly, JUN, a proto-oncogene, was seen to be ubiquitinated and degraded by Cereblon, which is a target of thalidomide (12,36,37). JUN is related to regulation of expression and was seen to play roles on early embryonic development, heart and bone formation (38–40). TCF4 also plays a part in bone development, together with the regulation of expression through interaction with other basic helix-loop-helix proteins, of which HAND2 and TCF4 are part of (17,41,42). The gene *UBB* will codified a part of the final protein Ubiquitin, which will link to protein via the E1-E2-E3 enzyme cascade(43,44). This cascade is the same used by the complex of ubiquitination of Cereblon (12). Interestingly, TBX5 and HAND2 did not appear to be important nodes to maintain the connection of its own network.

It is important to notice that the data banks used in the analysis obtains its information from published studies. Therefore, HAND and TBX5 may not have appeared with high scores because of the lack of studies about them, in contrast to CTNNB1, the authority of Mater network, that is widely studied.

We also found that FANCM, SRCAP, EP300 and TBX5 were the proteins responsible for the connection between HAND2 and TBX5 network (Mater) with heart and limbs malformation network (Cardiac and Limbs) and EP300 and FANCM were responsible to maintain that connection. Therefore, EP300 and FANCM could represent the link between HAND2 and TBX5 interaction with heart and limbs malformations of TE being the ones that if altered could result in the most disbalance of that connection. Interestingly, FANCM is related to Fanconi anemia, which TE is considered a phenocopy (45). Also, FANCB, another protein related to Fanconi anemia was present in heart and limbs malformations network and we observed that its gene was downregulated after treatment with thalidomide.

EP300 represents the proteins that stood out the most in the analysis. It was presented connecting Mater network to Cardiac and Limbs networks, was important to maintain this connection, was important to Cardiac and Limbs networks individually and to Mater network. In addition, EP300 is part of Lysine acetyltransferases group of proteins together with *ESCO2*, which was seen to be downregulated in our analysis. ESCO2 is associated with Roberts Syndrome, of which TE is a phenocopy and was already studied for genetic susceptibility of TE (35,46). EP300 was seen to be important for cell proliferation during embryogenesis(47). The same study observed cardiovascular malformations in mice lacking p300 and associated those malformations with the embryonic lethality they observed in p300 lacking mice(47). Indeed, it is estimated that about 40% of embryos exposed to thalidomide died before one year of life and those deaths and miscarriages are probably due to malformations in internal organs, especially in the heart (1,2,48).

Moreover, mutations in *EP300* are associated with the occurrence of Rubinstein-Taybi syndrome (RSTS)(49,50). RSTS can cause facial abnormalities, skeletal anomalies, growth retardation and intellectual disability,

among others(49,50). Facial and skeletal anomalies are also seen in TE(1). Furthermore, cardiovascular malformations were described in RSTS individuals, including some malformations also observed in TE, as ventricular septal defects and pulmonary stenosis(1,50).

It is important to notice that although there is many molecular information about thalidomide's use in multiple myeloma, the data for thalidomide teratogenicity are limited. Therefore, the design of the study was performed in the best way to give reliable information and results given the small amount of information available causing some limitations to the study. In the gene selection through malformations compatible to TE we were not able to select only genes associated to TE due to the scarce of data. To amplify the number of genes selected we included all genes associated to the malformations observed in TE but that could also be associated to other conditions. We did not select a specific tissue to analyze, we were evaluating all the tissues at the same time. A tissue-specific analysis would be more accurate as HAND2 and TBX5 are not expressed constitutively across all tissues; however, this was not possible because the data from databases were not sorted by tissue. Another limitation of the present study is the selection of C2H2-domain proteins. We cannot affirm that all the C2H2-domain proteins will be affected by the CRL4 ^{CRBN} -thalidomide complex. Furthermore, we analyzed only protein-protein interaction networks, and most of the proteins that stood out in the analysis were transcription factors, which have as targets genes. We were not able to evaluated theses different level of complexity including the interaction of protein-DNA.

5 Conclusions

We have been restricted to targets of thalidomide and dependents of findings in humans, biochemical and *in vivo* approaches. In this study we could integrate what we know about thalidomide with information about the interaction of proteins with related conditions to TE. When we integrated these data, we were able to find potential candidates with biological sense. We showed that EP300, which appeared to be important in all the analyzed networks, could represent a link between limbs and cardiac malformations of TE, being a good candidate to be further evaluated in the teratogenicity of TE. Furthermore, FANCM and FANCB,

proteins associated to Fanconi anemia, also are good candidates to be evaluated in TE, they were in limbs and cardiac networks and differentially expressed, respectively. We have to look not only to specific genes, but to its surroundings, networks and pathways as potential candidates to cause what we observe in TE. This study does not limit but open a door to further investigations on TE through approaches *in vivo* and in humans so, in the future, we can design safe drugs. Therefore, experimental analysis *in vitro* and *in vivo* are needed to understand if the suggested genes in this study does really play a part in TE.

6 References

1. Smithells RW, Newman CG. Recognition of thalidomide defects. J Med Genet [Internet]. 1992 Oct [cited 2018 Mar 20];29(10):716–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1433232

2. Lenz W. A short history of thalidomide embryopathy. Teratology. 1988;38(3):203–15.

3. Kim JH, Scialli AR. Thalidomide: the tragedy of birth defects and the effective treatment of disease. Toxicol Sci [Internet]. 2011 Jul [cited 2017 Jan 20];122(1):1–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21507989

4. Sheskin J. THALIDOMIDE IN THE TREATMENT OF LEPRA REACTIONS. Clin Pharmacol Ther [Internet]. 1965 [cited 2018 Mar 22];6:303–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14296027

5. Rajkumar SV, Blood E. Lenalidomide and venous thrombosis in multiple myeloma. N Engl J Med [Internet]. 2006 May 11 [cited 2018 Mar 22];354(19):2079–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16696148

6. Vianna FSL, Lopez-Camelo JS, Leite JCL, Sanseverino MTV, da Graça Dutra M, Castilla EE, et al. Epidemiological surveillance of birth defects

compatible with thalidomide embryopathy in Brazil. PLoS One. 2011;

7. Castilla EE, Ashton-Prolla P, Barreda-Mejia E, Brunoni D, Cavalcanti DP, Correa-Neto J, et al. Thalidomide, a current teratogen in South America. Teratology [Internet]. 1996 Dec [cited 2018 Jun 19];54(6):273–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9098920

8. Schuler-Faccini L, Soares RCF, de Sousa ACM, Maximino C, Luna E, Schwartz IVD, et al. New cases of thalidomide embryopathy in Brazil. Birth Defects Res Part A Clin Mol Teratol [Internet]. 2007 Sep [cited 2018 Jun 19];79(9):671–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17676592

9. Vianna FSL, Schüler-Faccini L, Leite JCL, De Sousa SHC, Da Costa LMM, Dias MF, et al. Recognition of the phenotype of thalidomide embryopathy in countries endemic for leprosy: New cases and review of the main dysmorphological findings. Clin Dysmorphol. 2013;22(2):59–63.

10. D'Amato RJ, Loughnan MS, Flynn E, Folkman J. Thalidomide is an inhibitor of angiogenesis. Proc Natl Acad Sci [Internet]. 1994;91(9):4082–5. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.91.9.4082

11. Parman T, Wiley MJ, Wells PG. Free radical-mediated oxidative DNA damage in the mechanism of thalidomide teratogenicity. Nat Med [Internet]. 1999 May 1 [cited 2017 Jan 20];5(5):582–5. Available from: http://www.nature.com/doifinder/10.1038/8466

12. Ito T, Ando H, Suzuki T, Ogura T, Hotta K, Imamura Y, et al. Identification of a primary target of thalidomide teratogenicity. Science [Internet]. 2010 Mar 12 [cited 2017 Jan 20];327(5971):1345–50. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1177319

13. Sievers QL, Petzold G, Bunker RD, Renneville A, Słabicki M, Liddicoat BJ, et al. Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. Science (80-) [Internet]. 2018 Nov 2;362(6414):eaat0572. Available from: file:///C:/Users/Carla Carolina/Desktop/Artigos para acrescentar na qualificação/The impact of

birth weight on cardiovascular disease risk in the.pdf

14. Khalil A, Tanos R, El-Hachem N, Kurban M, Bouvagnet P, Bitar F, et al. A HAND to TBX5 Explains the Link Between Thalidomide and Cardiac Diseases. Sci Rep [Internet]. 2017;7(1):1416. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28469241%0Ahttp://www.nature.com/articles/s41598-017-01641-3

15. McDermott DA, Fong JC, Basson CT. Holt-Oram Syndrome [Internet]. GeneReviews. University of Washington, Seattle; 1993 [cited 2018 May 15]. 1–22 p. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20301290

16. Srivastava D, Thomas T, Lin Q, Kirby ML, Brown D, Olson EN. Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. Nat Genet. 1997;16(2):154–60.

17. Dai Y-S, Cserjesi P. The basic helix-loop-helix factor, HAND2, functions as a transcriptional activator by binding to E-boxes as a heterodimer. J Biol Chem [Internet]. 2002 Apr 12 [cited 2017 Nov 10];277(15):12604–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11812799

18. Clowes C, Boylan MGS, Ridge LA, Barnes E, Wright JA, Hentges KE. The functional diversity of essential genes required for mammalian cardiac development. genesis [Internet]. 2014 Aug 1 [cited 2018 Mar 19];52(8):713–37. Available from: http://doi.wiley.com/10.1002/dvg.22794

19. Dai Y-S, Cserjesi P, Markham BE, Molkentin JD. The transcription factors GATA4 and dHAND physically interact to synergistically activate cardiac gene expression through a p300-dependent mechanism. J Biol Chem [Internet]. 2002 Jul 5 [cited 2017 Nov 10];277(27):24390–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11994297

20. Yamagishi H, Olson EN, Srivastava D. The basic helix-loop-helix transcription factor, dHAND, is required for vascular development. J Clin Invest [Internet]. 2000 Feb [cited 2018 Mar 22];105(3):261–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10675351

21. Tsuchihashi T, Maeda J, Shin CH, Ivey KN, Black BL, Olson EN, et al. Hand2 function in second heart field progenitors is essential for cardiogenesis. Dev Biol [Internet]. 2011 Mar 1 [cited 2017 Nov 10];351(1):62–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21185281

22. Pujato M, Kieken F, Skiles AA, Tapinos N, Fiser A. Prediction of DNA binding motifs from 3D models of transcription factors; identifying TLX3 regulated genes. Nucleic Acids Res. 2014;42(22):13500–12.

23. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. Nucleic Acids Res [Internet]. 2020 Jan 1 [cited 2021 Mar 4];48(D1):D498–503. Available from: https://pubmed.ncbi.nlm.nih.gov/31691815/

24. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The human phenotype ontology in 2021. Nucleic Acids Res [Internet]. 2021 Jan 8 [cited 2021 Mar 4];49(D1):D1207–17. Available from: https://hpo.jax.org/

25. Goenawan IH, Bryan K, Lynn DJ. DyNet: Visualization and analysis of dynamic molecular interaction networks. In: Bioinformatics [Internet]. Oxford University Press; 2016 [cited 2021 May 14]. p. 2713–5. Available from: https://pubmed.ncbi.nlm.nih.gov/27153624/

26. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics [Internet]. 2009 [cited 2021 May 14];25(8):1091–3. Available from: https://pubmed.ncbi.nlm.nih.gov/19237447/

27. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. Omi A J Integr Biol [Internet]. 2012 May 1 [cited 2021 Apr 7];16(5):284–7. Available from: https://www.liebertpub.com/doi/abs/10.1089/omi.2011.0118

28. An W, Liu YH. Keyplayer: An R package for locating key players in social

networks. R J. 2016;8(1):257–68.

29. Csárdi G, Nepusz T. The igraph software package for complex network research. InterJournal. 2006;Complex Sy:1695.

30. Schwartz MP, Hou Z, Propson NE, Zhang J, Engstrom CJ, Costa VS, et al. Human pluripotent stem cell-derived neural constructs for predicting neural toxicity. Proc Natl Acad Sci U S A [Internet]. 2015 Oct 6 [cited 2021 Mar 2];112(40):12516–21. Available from: https://pubmed.ncbi.nlm.nih.gov/26392547/

31. Newman CGH. The thalidomide syndrome: Risks of exposure and spectrum of malformations. Clin Perinatol. 1986;13(3):555–73.

32. Vianna FSL, Schüler-Faccini L, Leite JCL, de Sousa SHC, da Costa LMM, Dias MF, et al. Recognition of the phenotype of thalidomide embryopathy in countries endemic for leprosy. Clin Dysmorphol. 2013 Mar 1;22(2):59–63.

33. Fischer ES, Böhm K, Lydeard JR, Yang H, Stadler MB, Cavadini S, et al. Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. Nature. 2014;512(7512).

34. Donovan KA, An J, Nowak RP, Yuan JC, Fink EC, Berry BC, et al. Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane radial ray syndrome. Elife [Internet]. 2018 [cited 2021 May 11];7:1–25. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6156078/

35. Vega H, Waisfisz Q, Gordillo M, Sakai N, Yanagihara I, Yamada M, et al. Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet [Internet]. 2005 May [cited 2021 Mar 2];37(5):468–70. Available from: https://pubmed.ncbi.nlm.nih.gov/15821733/

36. Yang J, Huang M, Zhou L, He X, Jiang X, Zhang Y, et al. Cereblon suppresses the lipopolysaccharide-induced inflammatory response by

promoting the ubiquitination and degradation of c-Jun. J Biol Chem [Internet]. 2018 Jun 29 [cited 2021 Mar 16];293(26):10141–57. Available from: https://pubmed.ncbi.nlm.nih.gov/29748389/

37. Meng Q, Xia Y. c-Jun, at the crossroad of the signaling network [Internet]. Vol. 2, Protein and Cell. Higher Education Press; 2011 [cited 2021 Mar 16]. p. 889–98. Available from: https://pubmed.ncbi.nlm.nih.gov/22180088/

38. Yu C, Zhang X, Wang L, Liu Y, Li N, Li M, et al. Interleukin-6 regulates expression of Fos and Jun genes to affect the development of mouse preimplantation embryos. J Obstet Gynaecol Res [Internet]. 2018 Feb 1 [cited 2021 Mar 16];44(2):253–62. Available from: https://pubmed.ncbi.nlm.nih.gov/29171142/

39. Wagner EF. Functions of AP1 (Fos/Jun) in bone development. In: Annals of the Rheumatic Diseases [Internet]. BMJ Publishing Group; 2002 [cited 2021 Mar 16]. p. ii40–2. Available from: http://ard.bmj.com/

40. Eferl R, Sibilia M, Hilberg F, Fuchsbichler A, Kufferath I, Guertl B, et al. Functions of c-Jun in liver and heart development. J Cell Biol [Internet]. 1999 May 31 [cited 2021 Mar 16];145(5):1049–61. Available from: http://www.jcb.org

41. Ah Cho E, Dressler GR. TCF-4 binds β-catenin and is expressed in distinct regions of the embryonic brain and limbs. Mech Dev. 1998;77(1):9–18.

42. Forrest MP, Hill MJ, Quantock AJ, Martin-Rendon E, Blake DJ. The emerging roles of TCF4 in disease and development. Vol. 20, Trends in Molecular Medicine. Elsevier Ltd; 2014. p. 322–31.

43. Wiborg O, Pedersen MS, Wind A, Berglund LE, Marcker KA, Vuust J. The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. EMBO J. 1985;4(3):755–9.

44. Scheffner M, Nuber U, Huibregtse JM. Protein ubiquitination involving an

E1–E2–E3 enzyme ubiquitin thioester cascade. Nature. 1995 Jan 5;373(6509):81–3.

45. Titus TA, Yan Y-L, Wilson C, Starks AM, Frohnmayer JD, Canestro C, et al. The Fanconi anemia/BRCA gene network in zebrafish: Embryonic expression and comparative genomics. Mutat Res [Internet]. 2009;668(2):117–32. Available from: http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi

46. Gomes J do A, Kowalski TW, Fraga LR, Macedo GS, Sanseverino MTV, Schuler-Faccini L, et al. The role of ESCO2, SALL4 and TBX5 genes in the susceptibility to thalidomide teratogenesis. Sci Rep [Internet]. 2019 Dec 6 [cited 2019 Aug 28];9(1):11413. Available from: http://www.nature.com/articles/s41598-019-47739-8

47. Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, et al. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell [Internet]. 1998 May 1 [cited 2021 Apr 13];93(3):361–72. Available from: http://www.cell.com/article/S0092867400811654/fulltext

48. Vargesson N. Thalidomide-induced limb defects: Resolving a 50-year-old puzzle. BioEssays. 2009;31:1327–36.

49. Roelfsema JH, White SJ, Ariyürek Y, Bartholdi D, Niedrist D, Papadia F, et al. Genetic heterogeneity in Rubinstein-Taybi syndrome: Mutations in both the CBP and EP300 genes cause disease. Am J Hum Genet [Internet]. 2005 [cited 2021 Apr 13];76(4):572–80. Available from: https://pubmed.ncbi.nlm.nih.gov/15706485/

50. Fergelot P, Van Belzen M, Van Gils J, Afenjar A, Armour CM, Arveiler B, et al. Phenotype and genotype in 52 patients with Rubinstein-Taybi syndrome caused by *EP300* mutations. Am J Med Genet Part A [Internet]. 2016 Dec 1 [cited 2021 Apr 13];170(12):3069–82. Available from: http://doi.wiley.com/10.1002/ajmg.a.37940

7 Declarations

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by Hospital de Clínicas de Porto Alegre (HCPA) - Fundo de Incentivo a Pesquisa e Eventos (FIPE), grant numbers 2010-0244 and 2017-0248. FSLV is recipient of a Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) scholarship grant (grant number CNPq 312993/2017-0). TWK was first recipient of a Coordination of Improvement of Higher Education Personnel (CAPES) scholarship (grant 88881.132344/2016-01) and then of a CNPq scholarship (grant 156158/2018-3). BDR is recipient of a CNPq scholarship (grant 133335/2019-4).

Authors' contributions

All authors contributed to the development of the idea and construction of the workflow of the analysis. BDR performed the analysis with help of TWK. BDR wrote the first draft of the manuscript. LSF reviewed the phenotypes selected for cardiac and limbs malformations of TE. TWK, FSLV and LRF contributed to the writing and rational of the analysis. All authors read and approved the final manuscript.

Acknowledgments

We acknowledge Têmis Maria Felix for helping and reviewing the phenotypes selected for cardiac and limbs malformations of TE.

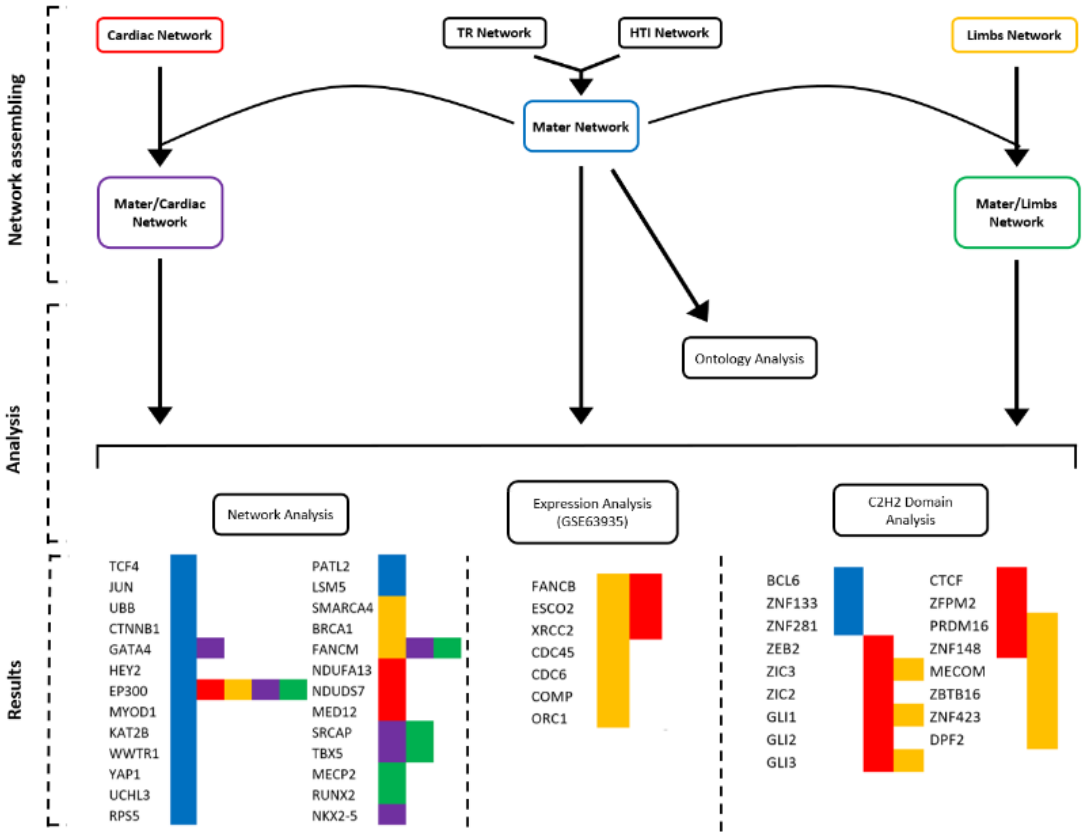


Figure 1: Scheme showing the steps of the analysis, the analysis performed and a summary of the results for each step and network. First, we assembled the networks. We constructed a network for HAND2 and TBX5 (Mater network) and networks for heart and limbs malformations of TE (Cardiac and Limbs network, respectively). In the sequence, it was performed network analysis, expression analysis and C2H2 domain analysis of each final network. The genes resulted in each analysis are represented for each network according to the colors. Red for Cardiac network. Blue for Mater network. Purple for Mater/Cardiac network. Yellow for Limbs network. Green for Mater/Limbs network. TR network: TF2DNA/REACTOME network. HTI network: HAND2 and TBX5 interaction network.

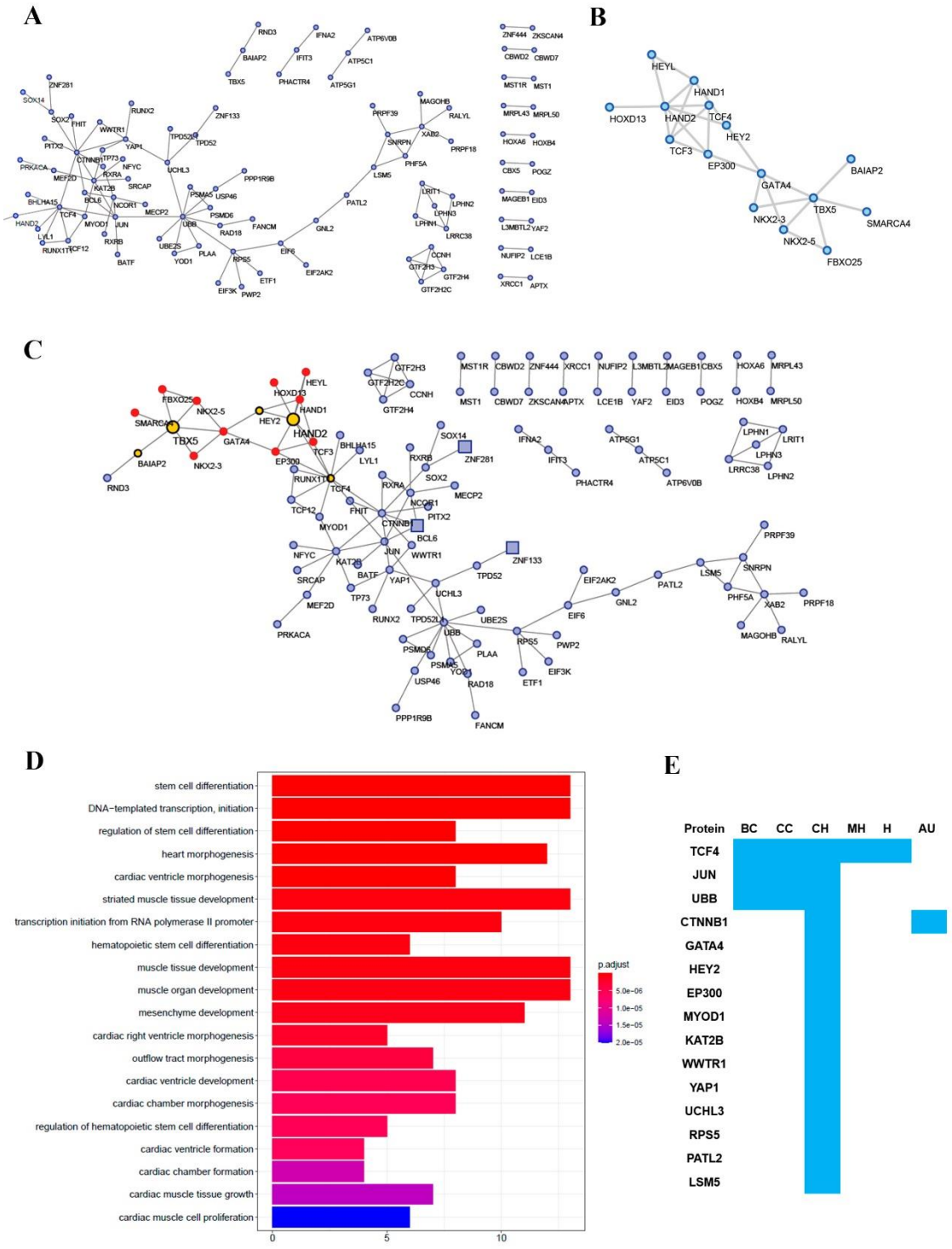


Figure 2: Analysis performed on Step1. A. TF2DNA/REACTOME (TR) network. In total 528 genes for HAND2 and 98 genes for TBX5 were selected. Only two genes were in common between the two databases search: *KCNE1* and *ZNF436*. From all the genes selected, only 98 had some interaction described and were present in the assembled network. *KCNE1* and *ZNF436* did not have any interaction described, therefore, they were not in the network. B. HAND2 and TBX5 Interaction (HTI) network. This network contains 15 proteins and represents the interactions already known between the two proteins. C. Mater network. In red the proteins from HTI network, in blue the protein from TR network

and in yellow the proteins in common. HAND2 and TBX5 are in evidence. C2H2 proteins are represented as a square. D. 20 enriched ontologies with smaller p-value from Mater network. E. Nodes that stood out in maintaining the structure of Mater network with each analysis performed. BC: betweenness centrality. CC: closeness centrality. CH: connector hub. MH: mainhub. H: hub. AU: authority.

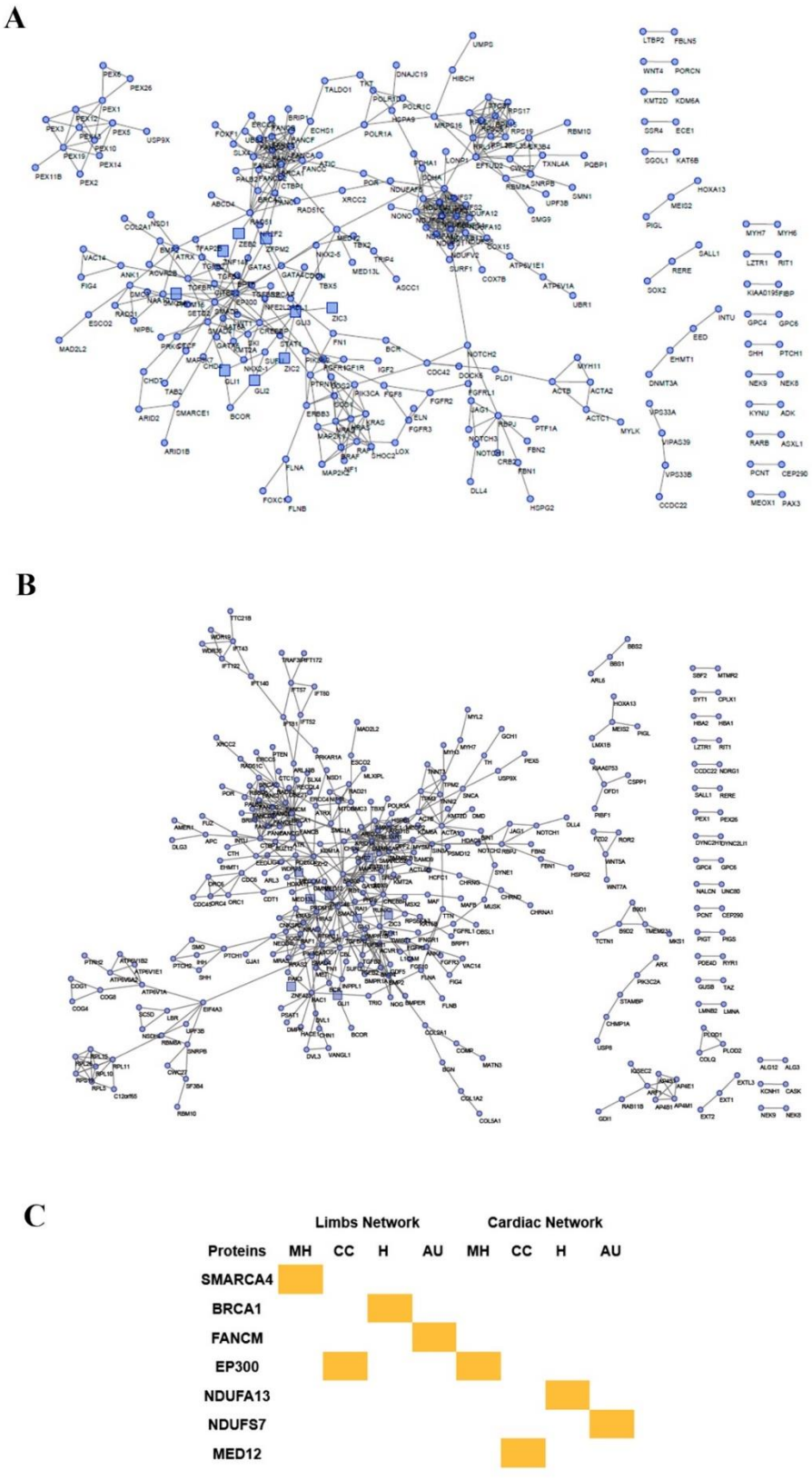


Figure 3: Network and resulted nodes of network analyses of heart and limbs malfomations of TE, first part of step 2. A. Cardiac Network. Cardiac network was formed by 254 proteins from the 475 genes selected on HPO B. Limbs network.

Limbs network was formed by 326 proteins from the 651 genes selected on HPO. C2H2 proteins are represented in a square. C. Nodes resulted from the network analysis of each network.

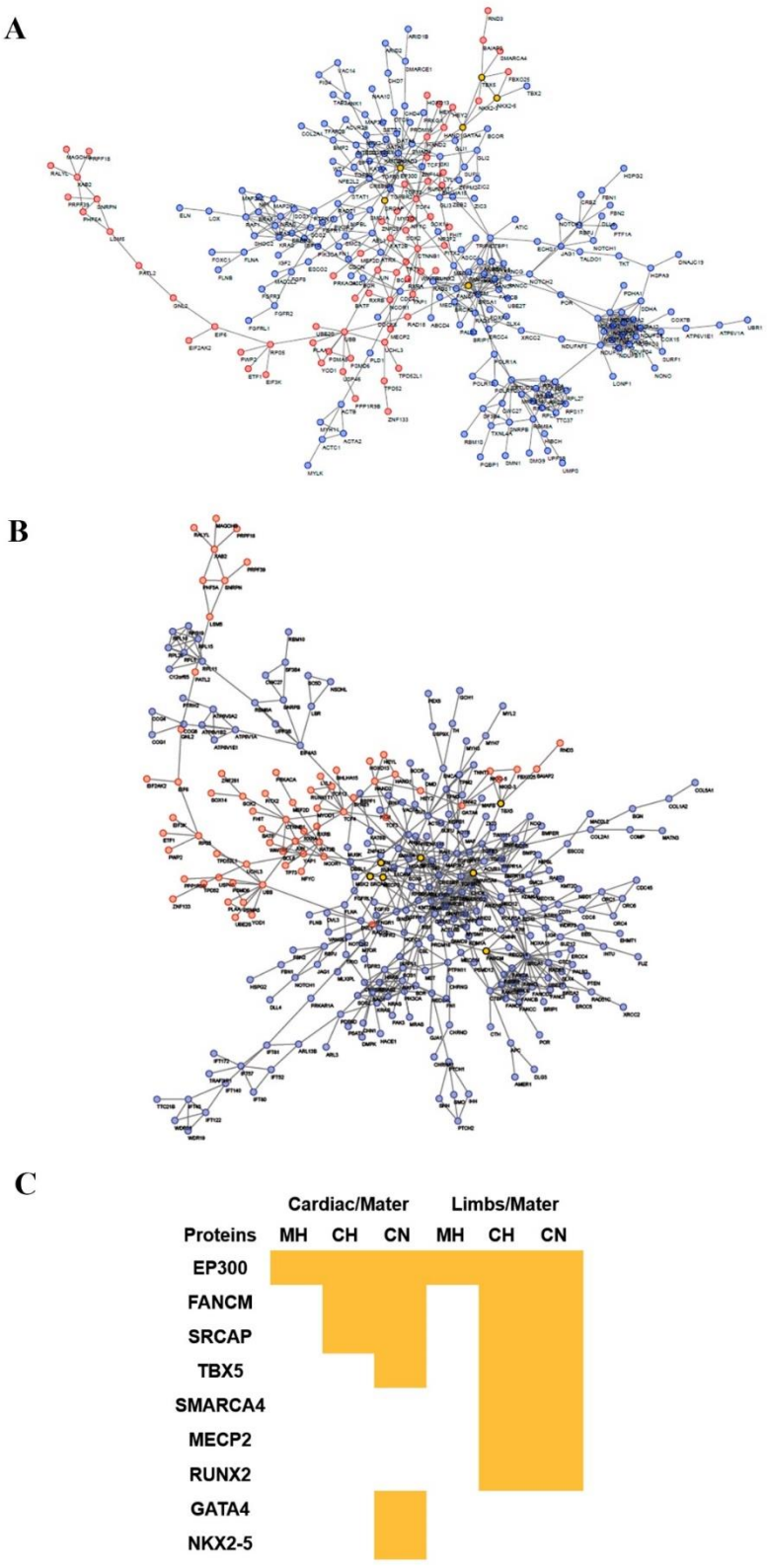


Figure 4: Results of second part of Step 2. A. Cardiac/Mater network. In blue is Cardiac network, in red is Mater network and in yellow the nodes connecting both networks B. Limbs/Mater network. In blue is Limbs networks, in red is Mater network and in yellow the nodes connecting both networks. C. Nodes resulted

from network analysis of Cardiac/Mater and Limbs/Mater networks. MH: mainhub; CH: connector hub; CN: connector between the networks.

Table 1: Phenotypes selected for limbs and heart malformations compatible with TE.

	HPO ID	Fenótipos
Heart		
	HP:0001629	Ventricular Septal Defect
	HP:0001631	Atrial Septal Defect
	HP:0001642	Pulmonic Stenosis
	HP:0001643	Patent Duct Arteriosus
	HP:0001710	Conotruncal Defect
Limbs		
	HP:0006713	Aplasia/Hypoplasia of the scapulae
	HP:0006710	Aplasia/Hypoplasia of the clavicles
	HP:0009829	Phocomelia
	HP:0006507	Aplasia/Hypoplasia of the humerus
	HP:0009823	Aplasia involving bones of the upper limbs
	HP:0003939	Humeroulnar synostosis
	HP:0002974	Radioulnar synostosis
	HP:0002984	Hypoplasia of the radius
	HP:0003022	Hypoplasia of the ulna
	HP:0006265	Aplasia/Hypoplasia of fingers
	HP:0005725	Nonopposable triphalangeal thumb
	HP:0001199	Triphalangeal thumb
	HP:0009601	Aplasia/Hypoplasia of the thumb
	HP:0005613	Aplasia/Hypoplasia of the femur
	HP:0002999	Patellar dislocation
	HP:0005736	Short tibia
	HP:0003038	Fibular hypoplasia
	HP:0001762	Talipes equinovarus (clubfoot)
	HP:0001829	Foot polydactyly
	HP:0009827	Amelia
	HP:0004059	Radial club hand
	HP:0006379	Proximal tibial hypoplasia
	HP:0003173	Hypoplastic pubic bone

HP:0005772	Aplasia/hypoplasia of tibia
HP:0006426	Rudimentary to absent tibiae
HP:0009556	Absent tibia
HP:0006492	Aplasia/Hypoplasia of the fibula
HP:0002990	Fibular aplasia

Table 2: Proteins resulted from analysis of C2H2 domain and differentially expression analysis.

	Proteins/Genes	Network
C2H2 domain		
	BCL6	Mater
	ZNF133	Mater
	ZNF281	Mater
	ZEB2	Cardiac
	ZIC3	Cardiac; Limbs
	ZIC2	Cardiac
	GLI1	Cardiac; Limbs
	GLI2	Cardiac
	GLI3	Cardiac; Limbs
	CTCF	Cardiac
	ZFPM2	Cardiac
	ZNF148	Cardiac; Limbs
	PRDM16	Cardiac; Limbs
	MECOM	Limbs
	ZBTB16	Limbs
	ZNF423	Limbs
	DPF2	Limbs
Differencialy expressed		
Downregulated		
	FANCB	Cardiac; Limbs
	ESCO2	Cardiac; Limbs
	XRCC2	Cardiac; Limbs
	CDC45	Limbs
	CDC6	Limbs
	ORC1	Limbs
Upregulated		
	COMP	Limbs

Supplementary Material

Additional File 1

Protein	Group (HGNC)
BAIAP2	I-BAR domain containing
ORC1	AAA ATPases
PATL2	Armadillo like helical domain containing
CTNNB1	Armadillo repeat containing
KAT2B	ATAC complex
DPF2	BAF complex
SMARCA4	BAF complex
HEY2	Basic helix-loop-helix proteins
MYOD1	Basic helix-loop-helix proteins
TCF4	Basic helix-loop-helix proteins
JUN	Basic leucine zipper proteins
CTNNB1	Beta catenins
BRCA1	BRCA1 A complex
BRCA1	BRCA1 B complex
BRCA1	BRCA1 C complex
EP300	Bromodomain containing
SMARCA4	Bromodomain containing
BCL6	BTB domain containing
ZBTB16	BTB domain containing
CTCF	Cilia and flagella associated
CDC45	CMG helicase complex
BRCA1	FA complementation groups
FANCB	FA complementation groups
FANCM	FA complementation groups
XRCC2	FA complementation groups
FANCB	FA core complex
FANCM	FA core complex
GATA4	GATA zinc finger domain containing
SMARCA4	GBAF complex
KAT2B	GCN5 related N-acetyltransferases
JUN	Jun transcription factor family
LSM5	LSm proteins
EP300	Lysine acetyltransferases
ESCO2	Lysine acetyltransferases
KAT2B	Lysine acetyltransferases
MECOM	Lysine methyltransferases
PRDM16	Lysine methyltransferases
MED12	Mediator complex

MECP2	Methyl-CpG binding domain containing
KAT2B	MicroRNA protein coding host genes
PRDM16	MicroRNA protein coding host genes
MYOD1	Myogenic regulatory family
NDUFS7	NADH:ubiquinone oxidoreductase core subunits
NDUFA13	NADH:ubiquinone oxidoreductase supernumerary subunits
NKX2-5	NKL subclass homeoboxes and pseudogenes
CDC6	none
COMP	none
UBB	none
WWTR1	none
YAP1	none
ORC1	Origin recognition complex
SMARCA4	PBAF complex
DPF2	PHD finger proteins
MECOM	PR/SET domain family
PRDM16	PR/SET domain family
BRCA1	Protein phosphatase 1 regulatory subunits
BRCA1	Ring finger proteins
FANCM	RNA helicases
RUNX2	Runt-related transcription factors
RPS5	S ribosomal proteins
KAT2B	SAGA complex
ZFPM2	SET domain containing
SRCAP	Small nucleolar RNA protein coding host genes
SRCAP	SRCAP complex
TBX5	T-box transcription factors
MED12	Trinucleotide repeat containing
UCHL3	Ubiquitin C-terminal hydrolases
CTNNB1	Wnt enhanceosome complex
ZEB2	ZF class homeoboxes and pseudogenes
BCL6	Zinc fingers C2H2-type
CTCF	Zinc fingers C2H2-type
GLI1	Zinc fingers C2H2-type
GLI2	Zinc fingers C2H2-type
GLI3	Zinc fingers C2H2-type
MECOM	Zinc fingers C2H2-type
PRDM16	Zinc fingers C2H2-type
ZBTB16	Zinc fingers C2H2-type
ZEB2	Zinc fingers C2H2-type
ZFPM2	Zinc fingers C2H2-type
ZIC2	Zinc fingers C2H2-type
ZIC3	Zinc fingers C2H2-type
ZNF133	Zinc fingers C2H2-type
ZNF148	Zinc fingers C2H2-type

ZNF281	Zinc fingers C2H2-type
ZNF423	Zinc fingers C2H2-type
ZFPM2	Zinc fingers C2HC-type
EP300	Zinc fingers ZZ-type

Additional File 2

	ID	Description	p.adjust	Genes
KEGG				
	hsa05216	Thyroid cancer	0.017844574	RXRA/RXRB/CTNNB1
	hsa03040	Spliceosome	0.017844574	LSM5/PHF5A/XAB2/MAGOHB/PRPF18
	hsa05223	Non-small cell lung cancer	0.067238931	FHIT/RXRA/RXRB
	hsa05200	Pathways in cancer	0.089957508	RUNX1T1/RXRA/EP300/RXRB/CTNNB1/JUN
	hsa04520	Adherens junction	0.089957508	BAIAP2/EP300/CTNNB1
	hsa04310	Wnt signaling pathway	0.089957508	PRKACA/EP300/CTNNB1/JUN
	hsa03008	Ribosome biogenesis in eukaryotes	0.089957508	EIF6/PWP2/GNL2
	hsa05222	Small cell lung cancer	0.089957508	FHIT/RXRA/RXRB
GO				
With TBX5				
	GO:0006352	DNA-templated transcription, initiation	1.3245E-08	HEY2/RXRA/YAP1/RUNX2/NKX2-5/WWTR1/TBX5/RXRB/SMARCA4/KAT2B/TCF4/CTNNB1/JUN
	GO:2000736	regulation of stem cell differentiation	1.28694E-07	EIF2AK2/YAP1/TP73/NKX2-5/TBX5/TCF3/GATA4/TCF12
	GO:0014706	striated muscle tissue development	8.26558E-07	MEF2D/HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5/EP300/HAND1/GATA4/HEYL/MYOD1/CTNNB1
	GO:0006367	transcription initiation from RNA polymerase II promoter	8.44311E-07	HEY2/RXRA/YAP1/RUNX2/NKX2-5/WWTR1/TBX5/RXRB/KAT2B/TCF4
	GO:0060537	muscle tissue development	9.0665E-07	MEF2D/HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5/EP300/HAND1/GATA4/HEYL/MYOD1/CTNNB1
	GO:0007517	muscle organ development	1.03794E-06	MEF2D/HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5/EP300/HAND1/HEYL/TCF12/MYOD1/CTNNB1
	GO:0055017	cardiac muscle tissue growth	1.44616E-05	HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5/GATA4
	GO:0060038	cardiac muscle cell proliferation	2.00971E-05	HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5
	GO:0060419	heart growth	2.08593E-05	HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5/GATA4
	GO:0033002	muscle cell proliferation	3.09206E-05	MEF2D/HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5/CTNNB1/JUN
	GO:0060413	atrial septum morphogenesis	3.37608E-05	HEY2/NKX2-5/TBX5/GATA4
	GO:0060411	cardiac septum morphogenesis	4.18429E-05	HEY2/NKX2-5/TBX5/HAND1/GATA4/HEYL
	GO:0014855	striated muscle cell proliferation	4.74433E-05	HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5
	GO:0003283	atrial septum development	6.53104E-05	HEY2/NKX2-5/TBX5/GATA4
	GO:0042692	muscle cell differentiation	0.000107128	MECP2/HEY2/BHLHA15/RXRA/NKX2-5/TBX5/TCF3/GATA4/MYOD1/CTNNB1
	GO:0016202	regulation of striated muscle tissue development	0.000108909	HEY2/YAP1/TP73/NKX2-5/TBX5/MYOD1/CTNNB1

GO:1901861	regulation of muscle tissue development	0.000116005	HEY2/YAP1/TP73/NKX2-5/TBX5/MYOD1/CTNNB1
GO:0003229	ventricular cardiac muscle tissue development	0.000120475	HEY2/RXRA/NKX2-5/TBX5/HAND1
GO:0048634	regulation of muscle organ development	0.000120475	HEY2/YAP1/TP73/NKX2-5/TBX5/MYOD1/CTNNB1
GO:0060043	regulation of cardiac muscle cell proliferation	0.000125832	HEY2/YAP1/TP73/NKX2-5/TBX5
GO:0048738	cardiac muscle tissue development	0.000132999	HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5/HAND1/GATA4
GO:0003279	cardiac septum development	0.000204542	HEY2/NKX2-5/TBX5/HAND1/GATA4/HEYL
GO:0003209	cardiac atrium morphogenesis	0.000230115	HEY2/NKX2-5/TBX5/GATA4
GO:0003281	ventricular septum development	0.000326021	HEY2/NKX2-5/TBX5/GATA4/HEYL
GO:0003230	cardiac atrium development	0.000326021	HEY2/NKX2-5/TBX5/GATA4
GO:0055021	regulation of cardiac muscle tissue growth	0.000344912	HEY2/YAP1/TP73/NKX2-5/TBX5
GO:0035265	organ growth	0.0003594	HEY2/RXRA/YAP1/TP73/NKX2-5/TBX5/GATA4
GO:0001570	vasculogenesis	0.000401266	HEY2/YAP1/NKX2-5/TBX5/CTNNB1
GO:0060420	regulation of heart growth	0.000463955	HEY2/YAP1/TP73/NKX2-5/TBX5
GO:0003214	cardiac left ventricle morphogenesis	0.00056099	HEY2/TBX5/HAND1
GO:0045844	positive regulation of striated muscle tissue development	0.000580548	HEY2/YAP1/TBX5/MYOD1/CTNNB1
GO:0048636	positive regulation of muscle organ development	0.000580548	HEY2/YAP1/TBX5/MYOD1/CTNNB1
GO:1901863	positive regulation of muscle tissue development	0.000599644	HEY2/YAP1/TBX5/MYOD1/CTNNB1
GO:0003179	heart valve morphogenesis	0.000914987	HEY2/TBX5/GATA4/HEYL
GO:0055024	regulation of cardiac muscle tissue development	0.000927527	HEY2/YAP1/TP73/NKX2-5/TBX5
GO:0010002	cardioblast differentiation	0.001177204	NKX2-5/TBX5/GATA4
GO:0003170	heart valve development	0.001261157	HEY2/TBX5/GATA4/HEYL
GO:2000738	positive regulation of stem cell differentiation	0.001320228	NKX2-5/TBX5/GATA4
GO:0046620	regulation of organ growth	0.001336462	HEY2/YAP1/TP73/NKX2-5/TBX5
GO:0003181	atrioventricular valve morphogenesis	0.001663658	HEY2/TBX5/HEYL
GO:0051146	striated muscle cell differentiation	0.00190778	HEY2/BHLHA15/RXRA/NKX2-5/TBX5/GATA4/MYOD1
GO:0003171	atrioventricular valve development	0.001997559	HEY2/TBX5/HEYL
GO:1905207	regulation of cardiocyte differentiation	0.002042342	HEY2/NKX2-5/TBX5/GATA4
GO:0055007	cardiac muscle cell differentiation	0.002515534	HEY2/RXRA/NKX2-5/TBX5/GATA4
GO:0001837	epithelial to mesenchymal transition	0.002874314	HEY2/WWTR1/TBX5/HEYL/CTNNB1
GO:0035115	embryonic forelimb morphogenesis	0.004166985	RUNX2/TBX5/CTNNB1
GO:0048638	regulation of developmental growth	0.004536774	HEY2/PLAA/YAP1/TP73/NKX2-5/TBX5/MYOD1
GO:0060045	positive regulation of cardiac muscle cell proliferation	0.005414329	HEY2/YAP1/TBX5

GO:0035136	forelimb morphogenesis	0.006670754	RUNX2/TBX5/CTNNB1
GO:1905209	positive regulation of cardiocyte differentiation	0.006670754	NKX2-5/TBX5/GATA4
GO:0003197	endocardial cushion development	0.007517222	TBX5/GATA4/HEYL
GO:0048639	positive regulation of developmental growth	0.007985798	HEY2/PLAA/YAP1/TBX5/MYOD1
GO:0055023	positive regulation of cardiac muscle tissue growth	0.009138929	HEY2/YAP1/TBX5
With HAND2			
GO:0003215	cardiac right ventricle morphogenesis	2.18168E-06	HEY2/NKX2-5/HAND1/GATA4/HAND2
GO:0003151	outflow tract morphogenesis	3.73296E-06	PITX2/HEY2/NKX2-5/HAND2/HEYL/CTNNB1/JUN
GO:0035050	embryonic heart tube development	3.41468E-05	YAP1/NKX2-5/HAND1/GATA4/HAND2/CTNNB1
GO:0051090	regulation of DNA-binding transcription factor activity	0.000204542	UBB/EIF2AK2/EP300/TCF3/HAND1/SMARCA4/HAND2/HEYL/CTNNB1/JUN
GO:0003143	embryonic heart tube morphogenesis	0.000225457	YAP1/NKX2-5/HAND1/GATA4/HAND2
GO:0009855	determination of bilateral symmetry	0.000326021	PITX2/NKX2-5/HAND1/GATA4/HAND2/CTNNB1
GO:0009799	specification of symmetry	0.000326021	PITX2/NKX2-5/HAND1/GATA4/HAND2/CTNNB1
GO:0042476	odontogenesis	0.000330935	PITX2/RUNX2/HAND1/NKX2-3/HAND2/CTNNB1
GO:0001649	osteoblast differentiation	0.00056099	SOX2/MEF2D/PRKACA/RUNX2/WWTR1/HAND2/CTNNB1
GO:0007512	adult heart development	0.00056099	MEF2D/NKX2-5/HAND2
GO:0042475	odontogenesis of dentin-containing tooth	0.000599644	RUNX2/HAND1/NKX2-3/HAND2/CTNNB1
GO:0048732	gland development	0.001056151	PITX2/SOX2/RXRA/NKX2-5/NKX2-3/HOXD13/HAND2/CTNNB1/JUN
GO:2000677	regulation of transcription regulatory region DNA binding	0.001056151	HEY2/HAND1/HAND2/CTNNB1
GO:0043392	negative regulation of DNA binding	0.001182541	HEY2/HAND1/HAND2/JUN
GO:0001947	heart looping	0.001552735	NKX2-5/HAND1/GATA4/HAND2
GO:0007368	determination of left/right symmetry	0.001778433	PITX2/NKX2-5/HAND1/GATA4/HAND2
GO:0061371	determination of heart left/right asymmetry	0.002042342	NKX2-5/HAND1/GATA4/HAND2
GO:0051101	regulation of DNA binding	0.002042342	HEY2/HAND1/HAND2/CTNNB1/JUN
GO:0048562	embryonic organ morphogenesis	0.002104978	YAP1/RUNX2/NKX2-5/HAND1/GATA4/HAND2/CTNNB1
GO:0061311	cell surface receptor signaling pathway involved in heart development	0.002996017	HEY2/HAND2/CTNNB1
GO:0060562	epithelial tube morphogenesis	0.003285892	PRKACA/YAP1/NKX2-5/HAND1/GATA4/HAND2/CTNNB1
GO:0048568	embryonic organ development	0.003484423	HEY2/YAP1/RUNX2/NKX2-5/HAND1/GATA4/HAND2/CTNNB1
GO:0014031	mesenchymal cell development	0.004256267	PITX2/HEY2/HAND2/HEYL
GO:0010667	negative regulation of cardiac muscle cell apoptotic process	0.004785635	HEY2/NKX2-5/HAND2
GO:0010664	negative regulation of striated muscle cell apoptotic process	0.005717697	HEY2/NKX2-5/HAND2
GO:0061448	connective tissue development	0.006810913	UBB/MEF2D/RUNX2/HAND1/HAND2/CTNNB1

GO:0051098	regulation of binding	0.00700108	PRKACA/HEY2/EP300/HAND1/HAND2/CTNNB1/JUN
GO:0001503	ossification	0.007985798	SOX2/MEF2D/PRKACA/RUNX2/WWTR1/HAND2/CTNNB1
GO:0061307	cardiac neural crest cell differentiation involved in heart development	0.009138929	PITX2/HAND2
GO:0061308	cardiac neural crest cell development involved in heart development	0.009138929	PITX2/HAND2
With TBX5 and HAND2			
GO:0048863	stem cell differentiation	1.61802E-09	PITX2/ZNF281/EIF2AK2/YAP1/TP73/RUNX2/NKX2-5/TBX5/TCF3/GATA4/HAND2/BATF/TCF12
GO:0003007	heart morphogenesis	1.39935E-07	PITX2/HEY2/RXRA/YAP1/NKX2-5/TBX5/HAND1/GATA4/HAND2/HEYL/CTNNB1/JUN
GO:0003208	cardiac ventricle morphogenesis	1.57961E-07	HEY2/RXRA/NKX2-5/TBX5/HAND1/GATA4/HAND2/HEYL
GO:0060485	mesenchyme development	1.43452E-06	PITX2/HEY2/YAP1/NKX2-5/WWTR1/TBX5/HAND1/GATA4/HAND2/HEYL/CTNNB1
GO:0003231	cardiac ventricle development	5.01128E-06	HEY2/RXRA/NKX2-5/TBX5/HAND1/GATA4/HAND2/HEYL
GO:0003206	cardiac chamber morphogenesis	5.52313E-06	HEY2/RXRA/NKX2-5/TBX5/HAND1/GATA4/HAND2/HEYL
GO:0003211	cardiac ventricle formation	5.83075E-06	NKX2-5/TBX5/HAND1/HAND2
GO:0003207	cardiac chamber formation	1.29041E-05	NKX2-5/TBX5/HAND1/HAND2
GO:0003205	cardiac chamber development	2.70845E-05	HEY2/RXRA/NKX2-5/TBX5/HAND1/GATA4/HAND2/HEYL
GO:0061323	cell proliferation involved in heart morphogenesis	4.74433E-05	PITX2/TBX5/HAND2/CTNNB1
GO:0007389	pattern specification process	6.02652E-05	PITX2/PRKACA/HEY2/NKX2-5/TBX5/EP300/HAND1/GATA4/HOXD13/HAND2/CTNNB1
GO:0035051	cardiocyte differentiation	0.000171036	PITX2/HEY2/RXRA/NKX2-5/TBX5/GATA4/HAND2
GO:0048736	appendage development	0.000230115	PITX2/RUNX2/TBX5/HOXD13/HAND2/KAT2B/CTNNB1
GO:0060173	limb development	0.000230115	PITX2/RUNX2/TBX5/HOXD13/HAND2/KAT2B/CTNNB1
GO:0003263	cardioblast proliferation	0.000275436	TBX5/HAND2/CTNNB1
GO:0003266	regulation of secondary heart field cardioblast proliferation	0.000275436	TBX5/HAND2/CTNNB1
GO:0003264	regulation of cardioblast proliferation	0.000400261	TBX5/HAND2/CTNNB1
GO:0002009	morphogenesis of an epithelium	0.000477811	PRKACA/RXRA/YAP1/NKX2-5/TBX5/HAND1/GATA4/HOXD13/HAND2/CTNNB1
GO:0048762	mesenchymal cell differentiation	0.000519905	PITX2/HEY2/WWTR1/TBX5/HAND2/HEYL/CTNNB1
GO:0060977	coronary vasculature morphogenesis	0.000914987	HEY2/TBX5/HAND2
GO:2000136	regulation of cell proliferation involved in heart morphogenesis	0.001038478	TBX5/HAND2/CTNNB1
GO:0060973	cell migration involved in heart development	0.001177204	PITX2/TBX5/HAND2
GO:0030326	embryonic limb morphogenesis	0.002236742	RUNX2/TBX5/HOXD13/HAND2/CTNNB1
GO:0035113	embryonic appendage morphogenesis	0.002236742	RUNX2/TBX5/HOXD13/HAND2/CTNNB1
GO:0060914	heart formation	0.002533176	TBX5/HAND2/CTNNB1

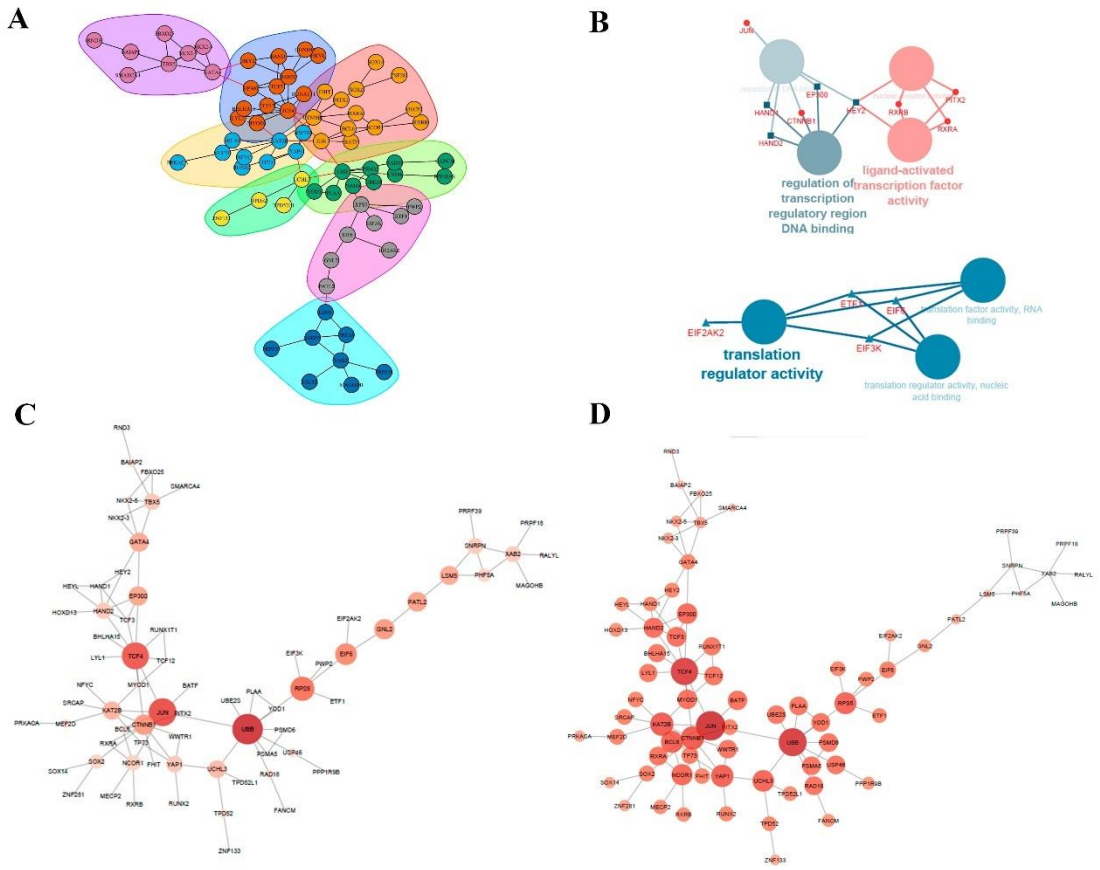
GO:2000027	regulation of animal organ morphogenesis	0.002578746	RXRA/RUNX2/TBX5/HOXD13/HAND2/CTNNB1
GO:0035107	appendage morphogenesis	0.003868897	RUNX2/TBX5/HOXD13/HAND2/CTNNB1
GO:0035108	limb morphogenesis	0.003868897	RUNX2/TBX5/HOXD13/HAND2/CTNNB1
GO:0003156	regulation of animal organ formation	0.005717697	TBX5/HAND2/CTNNB1
GO:2000826	regulation of heart morphogenesis	0.006068295	TBX5/HAND2/CTNNB1
Without HAND2 and/or TBX5			
GO:0060218	hematopoietic stem cell differentiation	9.0665E-07	EIF2AK2/YAP1/TP73/TCF3/BATF/TCF12
GO:1902036	regulation of hematopoietic stem cell differentiation	5.52313E-06	EIF2AK2/YAP1/TP73/TCF3/TCF12
GO:1901532	regulation of hematopoietic progenitor cell differentiation	2.70845E-05	EIF2AK2/YAP1/TP73/TCF3/TCF12
GO:1903706	regulation of hemopoiesis	4.74433E-05	EIF6/EIF2AK2/YAP1/TP73/BCL6/EP300/TCF3/KAT2B/TCF12/CTNNB1/JUN
GO:0071383	cellular response to steroid hormone stimulus	5.22162E-05	RXRA/NCOR1/YAP1/EP300/RXRB/SMARCA4/HEYL/MYOD1/CTNNB1
GO:0043401	steroid hormone mediated signaling pathway	5.40852E-05	RXRA/NCOR1/YAP1/EP300/RXRB/SMARCA4/HEYL/CTNNB1
GO:0050769	positive regulation of neurogenesis	8.63501E-05	BAIAP2/PLAA/TP73/NKX2-5/BCL6/EP300/TCF3/HEYL/TCF12/TCF4/CTNNB1
GO:0030522	intracellular receptor signaling pathway	0.000107128	UBB/RXRA/NCOR1/YAP1/EP300/RXRB/SMARCA4/HEYL/CTNNB1
GO:0009755	hormone-mediated signaling pathway	0.000204542	RXRA/NCOR1/YAP1/EP300/RXRB/SMARCA4/HEYL/CTNNB1
GO:0002244	hematopoietic progenitor cell differentiation	0.000216349	EIF2AK2/YAP1/TP73/TCF3/BATF/TCF12
GO:0060765	regulation of androgen receptor signaling pathway	0.000216349	NCOR1/EP300/SMARCA4/HEYL
GO:0051147	regulation of muscle cell differentiation	0.000276565	MECP2/HEY2/BHLHA15/NKX2-5/TCF3/MYOD1/CTNNB1
GO:0045165	cell fate commitment	0.000276565	SOX2/HEY2/RUNX2/NKX2-5/TCF3/BATF/MYOD1/CTNNB1
GO:0030098	lymphocyte differentiation	0.000285994	LYL1/RUNX2/BCL6/EP300/TCF3/NKX2-3/BATF/TPD52/CTNNB1
GO:0030521	androgen receptor signaling pathway	0.000297853	NCOR1/EP300/SMARCA4/HEYL/CTNNB1
GO:0030183	B cell differentiation	0.000326021	LYL1/BCL6/EP300/TCF3/NKX2-3/TPD52
GO:0045666	positive regulation of neuron differentiation	0.000330935	BAIAP2/PLAA/NKX2-5/BCL6/EP300/TCF3/HEYL/TCF12/TCF4
GO:2000142	regulation of DNA-templated transcription, initiation	0.000405542	HEY2/NKX2-5/CTNNB1/JUN
GO:0033143	regulation of intracellular steroid hormone receptor signaling pathway	0.00044233	NCOR1/YAP1/EP300/SMARCA4/HEYL
GO:0072182	regulation of nephron tubule epithelial cell differentiation	0.000477811	YAP1/WWTR1/CTNNB1
GO:0030518	intracellular steroid hormone receptor signaling pathway	0.00056099	NCOR1/YAP1/EP300/SMARCA4/HEYL/CTNNB1
GO:0048545	response to steroid hormone	0.00056099	RXRA/NCOR1/YAP1/EP300/RXRB/SMARCA4/HEYL/MYOD1/CTNNB1
GO:0060766	negative regulation of androgen receptor signaling pathway	0.000651052	NCOR1/SMARCA4/HEYL
GO:0072160	nephron tubule epithelial cell differentiation	0.000651052	YAP1/WWTR1/CTNNB1
GO:0055010	ventricular cardiac muscle tissue morphogenesis	0.000858069	HEY2/RXRA/NKX2-5/HAND1
GO:0043923	positive regulation by host of viral transcription	0.000914987	EP300/SMARCA4/JUN

GO:0055012	ventricular cardiac muscle cell differentiation	0.000914987	HEY2/RXRA/NKX2-5
GO:0048546	digestive tract morphogenesis	0.000948634	NKX2-3/GATA4/HOXD13/CTNNB1
GO:0007221	positive regulation of transcription of Notch receptor target	0.001038478	EP300/HEYL/KAT2B
GO:2000696	regulation of epithelial cell differentiation involved in kidney development	0.001038478	YAP1/WWTR1/CTNNB1
GO:0003002	regionalization	0.001177204	PITX2/PRKACA/HEY2/NKX2-5/EP300/GATA4/HOXD13/CTNNB1
GO:0043254	regulation of protein complex assembly	0.001320228	BAIAP2/MECP2/HEY2/EIF2AK2/NKX2-5/EP300/PATL2/CTNNB1/JUN
GO:0010498	proteasomal protein catabolic process	0.001329678	UBB/UBE2S/PSMA5/FHIT/PRKACA/PSMD6/PLAA/YOD1/WWTR1
GO:0050434	positive regulation of viral transcription	0.001493692	EP300/SMARCA4/JUN
GO:0008380	RNA splicing	0.001546234	LSM5/PHF5A/SNRPN/XAB2/MAGOHB/PRPF39/PRPF18/MYOD1/PPP1R9B
GO:0051817	modification of morphology or physiology of other organism involved in symbiotic interaction	0.001546234	RXRA/EIF2AK2/EP300/SMARCA4/JUN
GO:0072215	regulation of metanephros development	0.001874686	YAP1/WWTR1/CTNNB1
GO:0022618	ribonucleoprotein complex assembly	0.001886475	RPS5/EIF6/EIF3K/XAB2/PRPF39/PATL2/PRPF18
GO:0051148	negative regulation of muscle cell differentiation	0.00190778	MECP2/HEY2/BHLHA15/NKX2-5
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	0.00190778	LSM5/PHF5A/SNRPN/XAB2/MAGOHB/PRPF39/PRPF18/MYOD1
GO:0000398	mRNA splicing, via spliceosome	0.00190778	LSM5/PHF5A/SNRPN/XAB2/MAGOHB/PRPF39/PRPF18/MYOD1
GO:0000375	RNA splicing, via transesterification reactions	0.001996552	LSM5/PHF5A/SNRPN/XAB2/MAGOHB/PRPF39/PRPF18/MYOD1
GO:0022613	ribonucleoprotein complex biogenesis	0.001996552	RPS5/EIF6/EIF3K/PWP2/GNL2/XAB2/PRPF39/PATL2/PRPF18
GO:0071826	ribonucleoprotein complex subunit organization	0.002289967	RPS5/EIF6/EIF3K/XAB2/PRPF39/PATL2/PRPF18
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	0.002289967	UBB/UBE2S/PSMA5/FHIT/PSMD6/PLAA/YOD1/WWTR1
GO:0072202	cell differentiation involved in metanephros development	0.002377754	YAP1/WWTR1/CTNNB1
GO:0055008	cardiac muscle tissue morphogenesis	0.002533176	HEY2/RXRA/NKX2-5/HAND1
GO:0000956	nuclear-transcribed mRNA catabolic process	0.002533176	LSM5/RPS5/ETF1/MAGOHB/PATL2/PRPF18
GO:0043921	modulation by host of viral transcription	0.002533176	EP300/SMARCA4/JUN
GO:0052472	modulation by host of symbiont transcription	0.002533176	EP300/SMARCA4/JUN
GO:2000144	positive regulation of DNA-templated transcription, initiation	0.002533176	NKX2-5/CTNNB1/JUN
GO:0052312	modulation of transcription in other organism involved in symbiotic interaction	0.002783867	EP300/SMARCA4/JUN
GO:0030099	myeloid cell differentiation	0.002783867	EIF6/BCL6/EP300/NKX2-3/BATF/KAT2B/CTNNB1/JUN
GO:0042113	B cell activation	0.002787712	LYL1/BCL6/EP300/TCF3/NKX2-3/BATF/TPD52
GO:0006402	mRNA catabolic process	0.002979485	UBB/LSM5/RPS5/ETF1/MAGOHB/PATL2/PRPF18

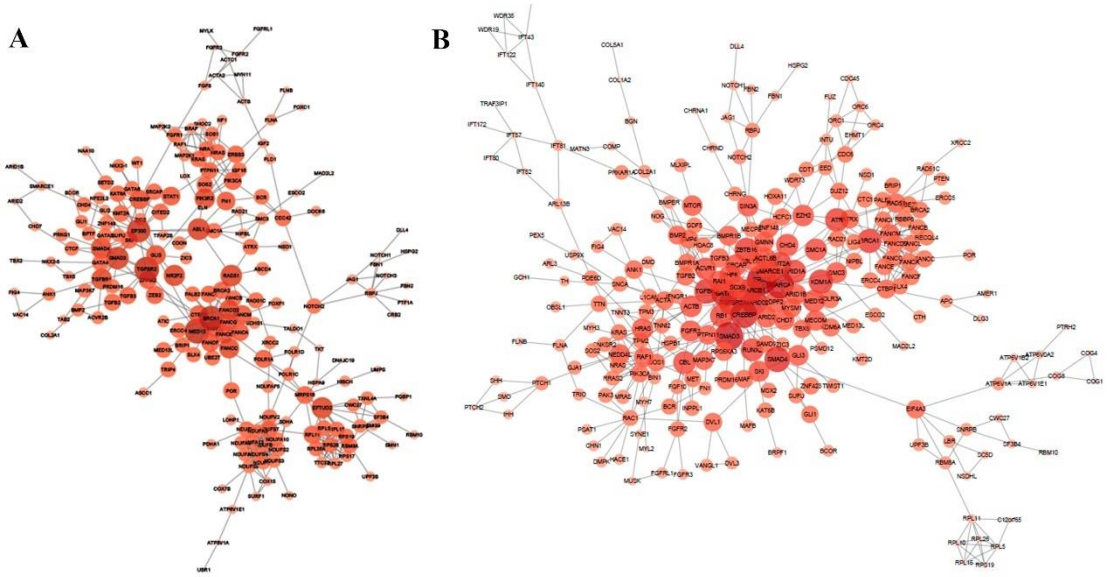
GO:1904837	beta-catenin-TCF complex assembly	0.003285892	EP300/SMARCA4/CTNNB1
GO:0072332	intrinsic apoptotic signaling pathway by p53 class mediator	0.003348013	UBB/FHIT/TP73/EP300
GO:0001655	urogenital system development	0.003827589	RXRA/YAP1/TP73/WWTR1/HOXD13/HEYL/CTNNB1
GO:1905314	semi-lunar valve development	0.003868897	HEY2/GATA4/HEYL
GO:0060415	muscle tissue morphogenesis	0.004256267	HEY2/RXRA/NKX2-5/HAND1
GO:0048536	spleen development	0.004470972	PITX2/NKX2-5/NKX2-3
GO:0016573	histone acetylation	0.004623118	SRCAP/MECP2/EP300/KAT2B/MYOD1
GO:0007050	cell cycle arrest	0.004896717	SOX2/PRKACA/TP73/EP300/KAT2B/PPP1R9B
GO:0006401	RNA catabolic process	0.004896717	UBB/LSM5/RPS5/ETF1/MAGOHB/PATL2/PRPF18
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	0.004896717	UBB/RUNX2/NKX2-5/WWTR1/EP300/GATA4/JUN
GO:0033144	negative regulation of intracellular steroid hormone receptor signaling pathway	0.005079939	NCOR1/SMARCA4/HEYL
GO:0018393	internal peptidyl-lysine acetylation	0.005166166	SRCAP/MECP2/EP300/KAT2B/MYOD1
GO:0048644	muscle organ morphogenesis	0.005241891	HEY2/RXRA/NKX2-5/HAND1
GO:0007519	skeletal muscle tissue development	0.005498961	MEF2D/EP300/HEYL/MYOD1/CTNNB1
GO:0035821	modification of morphology or physiology of other organism	0.005498961	RXRA/EIF2AK2/EP300/SMARCA4/JUN
GO:0006475	internal protein amino acid acetylation	0.005717697	SRCAP/MECP2/EP300/KAT2B/MYOD1
GO:0003015	heart process	0.005811619	PRKACA/HEY2/YAP1/NKX2-5/WWTR1/GATA4
GO:0018394	peptidyl-lysine acetylation	0.006068295	SRCAP/MECP2/EP300/KAT2B/MYOD1
GO:0051091	positive regulation of DNA-binding transcription factor activity	0.006068295	UBB/EIF2AK2/EP300/TCF3/SMARCA4/CTNNB1
GO:0072080	nephron tubule development	0.006068295	YAP1/WWTR1/HEYL/CTNNB1
GO:0031668	cellular response to extracellular stimulus	0.006372128	BAIAP2/BHLHA15/EIF2AK2/MYOD1/PPP1R9B/JUN
GO:0035315	hair cell differentiation	0.006372128	PITX2/HEY2/CTNNB1
GO:0048469	cell maturation	0.006433427	MECP2/PRKACA/BHLHA15/RUNX2/CTNNB1
GO:0061326	renal tubule development	0.006433427	YAP1/WWTR1/HEYL/CTNNB1
GO:0060538	skeletal muscle organ development	0.006548071	MEF2D/EP300/HEYL/MYOD1/CTNNB1
GO:0008593	regulation of Notch signaling pathway	0.007278775	HEY2/BCL6/EP300/KAT2B
GO:0021700	developmental maturation	0.007801966	LYL1/MECP2/PRKACA/BHLHA15/RUNX2/CTNNB1
GO:0007440	foregut morphogenesis	0.00785216	GATA4/CTNNB1
GO:0035850	epithelial cell differentiation involved in kidney development	0.007878084	YAP1/WWTR1/CTNNB1
GO:0046782	regulation of viral transcription	0.007878084	EP300/SMARCA4/JUN
GO:0060412	ventricular septum morphogenesis	0.008278629	HEY2/NKX2-5/HEYL

GO:0016579	protein deubiquitination	0.008278629	UBB/PSMA5/USP46/UCHL3/PSMD6/YOD1
GO:0060525	prostate glandular acinus development	0.009138929	RXRA/HOXD13
GO:0030850	prostate gland development	0.009138929	RXRA/HOXD13/CTNNB1
GO:0007219	Notch signaling pathway	0.009527734	HEY2/BCL6/EP300/HEYL/KAT2B
GO:0019083	viral transcription	0.009580912	EP300/SMARCA4/JUN
GO:0072009	nephron epithelium development	0.009580912	YAP1/WWTR1/HEYL/CTNNB1

Additional File 3



Supplementary Figure 1: Network Analysis of Mater network. Community analysis of Mater network. Each community is represented by a different color. A. Enrichment Ontology analysis of communities found in Mater Network. Only three communities were enriched. B. Betweenness Centrality of Mater network. C. Closeness Centrality of Mater network. D.



Supplementary Figure 2: Network Analysis of Cardiac and Limbs networks. Closeness centrality of Cardiac network. A. Closeness centrality of Limbs network. B.

Additional File 4

	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
GO:0030324	GO:0030324	lung development	5/17	169/18493	3.4E-07	9.49E-05	4.17E-05	ZIC3/GLI1/GLI2/GLI3/ZFPM2	5
GO:0030323	GO:0030323	respiratory tube development	5/17	173/18493	3.82E-07	9.49E-05	4.17E-05	ZIC3/GLI1/GLI2/GLI3/ZFPM2	5
GO:0003002	GO:0003002	regionalization	6/17	341/18493	3.92E-07	9.49E-05	4.17E-05	ZEB2/ZIC3/GLI1/GLI2/GLI3/ZBTB16	6
GO:0060541	GO:0060541	respiratory system development	5/17	194/18493	6.74E-07	0.000122	5.37E-05	ZIC3/GLI1/GLI2/GLI3/ZFPM2	5
GO:0045582	GO:0045582	positive regulation of T cell differentiation	4/17	83/18493	8.59E-07	0.000123	5.43E-05	BCL6/GLI2/GLI3/ZBTB16	4
GO:0009952	GO:0009952	anterior/posterior pattern specification	5/17	211/18493	1.02E-06	0.000123	5.43E-05	ZEB2/ZIC3/GLI2/GLI3/ZBTB16	5
GO:0045621	GO:0045621	positive regulation of lymphocyte differentiation	4/17	94/18493	1.42E-06	0.000146	6.39E-05	BCL6/GLI2/GLI3/ZBTB16	4
GO:0007389	GO:0007389	pattern specification process	6/17	434/18493	1.6E-06	0.000146	6.39E-05	ZEB2/ZIC3/GLI1/GLI2/GLI3/ZBTB16	6
GO:0009954	GO:0009954	proximal/distal pattern formation	3/17	32/18493	3.15E-06	0.000254	0.000112	GLI1/GLI2/GLI3	3
GO:0048565	GO:0048565	digestive tract development	4/17	131/18493	5.33E-06	0.000387	0.00017	ZIC3/GLI1/GLI2/GLI3	4
GO:0045580	GO:0045580	regulation of T cell differentiation	4/17	137/18493	6.37E-06	0.00042	0.000185	BCL6/GLI2/GLI3/ZBTB16	4
GO:1902107	GO:1902107	positive regulation of leukocyte differentiation	4/17	142/18493	7.34E-06	0.000421	0.000185	BCL6/GLI2/GLI3/ZBTB16	4
GO:0055123	GO:0055123	digestive system development	4/17	143/18493	7.54E-06	0.000421	0.000185	ZIC3/GLI1/GLI2/GLI3	4
GO:0030850	GO:0030850	prostate gland development	3/17	47/18493	1.02E-05	0.000525	0.000231	GLI1/GLI2/GLI3	3
GO:0032330	GO:0032330	regulation of chondrocyte differentiation	3/17	49/18493	1.16E-05	0.000525	0.000231	GLI2/GLI3/ZBTB16	3
GO:0048546	GO:0048546	digestive tract morphogenesis	3/17	49/18493	1.16E-05	0.000525	0.000231	GLI1/GLI2/GLI3	3
GO:0045619	GO:0045619	regulation of lymphocyte differentiation	4/17	166/18493	1.36E-05	0.000559	0.000246	BCL6/GLI2/GLI3/ZBTB16	4
GO:2000179	GO:2000179	positive regulation of neural precursor cell proliferation	3/17	52/18493	1.39E-05	0.000559	0.000246	GLI1/GLI2/GLI3	3
GO:1903708	GO:1903708	positive regulation of hemopoiesis	4/17	183/18493	2E-05	0.000763	0.000335	BCL6/GLI2/GLI3/ZBTB16	4
GO:0042733	GO:0042733	embryonic digit morphogenesis	3/17	61/18493	2.25E-05	0.000816	0.000358	GLI2/GLI3/ZBTB16	3
GO:0046637	GO:0046637	regulation of alpha-beta T cell differentiation	3/17	63/18493	2.48E-05	0.000856	0.000376	BCL6/GLI3/ZBTB16	3
GO:0050870	GO:0050870	positive regulation of T cell activation	4/17	202/18493	2.94E-05	0.000941	0.000413	BCL6/GLI2/GLI3/ZBTB16	4
GO:0048568	GO:0048568	embryonic organ development	5/17	422/18493	2.98E-05	0.000941	0.000413	ZIC3/GLI1/GLI2/GLI3/ZFPM2	5
GO:0061035	GO:0061035	regulation of cartilage development	3/17	68/18493	3.12E-05	0.000943	0.000414	GLI2/GLI3/ZBTB16	3
GO:0021936	GO:0021936	regulation of cerebellar granule cell precursor proliferation	2/17	10/18493	3.56E-05	0.000956	0.00042	GLI1/GLI2	2
GO:0048570	GO:0048570	notochord morphogenesis	2/17	10/18493	3.56E-05	0.000956	0.00042	GLI1/GLI2	2
GO:0060033	GO:0060033	anatomical structure regression	2/17	10/18493	3.56E-05	0.000956	0.00042	GLI1/GLI2	2
GO:1903039	GO:1903039	positive regulation of leukocyte cell-cell adhesion	4/17	214/18493	3.69E-05	0.000956	0.00042	BCL6/GLI2/GLI3/ZBTB16	4
GO:0016570	GO:0016570	histone modification	5/17	456/18493	4.32E-05	0.000988	0.000434	BCL6/CTCF/PRDM16/MECOM/DPF2	5
GO:0021536	GO:0021536	diencephalon development	3/17	76/18493	4.35E-05	0.000988	0.000434	ZEB2/GLI1/GLI2	3

GO:0021521	GO:0021521	ventral spinal cord interneuron specification	2/17	11/18493	4.35E-05	0.000988	0.000434	GLI2/GLI3	2
GO:0060573	GO:0060573	cell fate specification involved in pattern specification	2/17	11/18493	4.35E-05	0.000988	0.000434	GLI2/GLI3	2
GO:0008589	GO:0008589	regulation of smoothened signaling pathway	3/17	78/18493	4.7E-05	0.001035	0.000455	GLI1/GLI2/GLI3	3
GO:0016569	GO:0016569	covalent chromatin modification	5/17	470/18493	4.99E-05	0.001065	0.000468	BCL6/CTCF/PRDM16/MECOM/DPF2	5
GO:0030217	GO:0030217	T cell differentiation	4/17	234/18493	5.22E-05	0.001084	0.000476	BCL6/GLI2/GLI3/ZBTB16	4
GO:2000177	GO:2000177	regulation of neural precursor cell proliferation	3/17	84/18493	5.87E-05	0.001184	0.00052	GLI1/GLI2/GLI3	3
GO:0021520	GO:0021520	spinal cord motor neuron cell fate specification	2/17	13/18493	6.17E-05	0.00121	0.000532	GLI2/GLI3	2
GO:0022409	GO:0022409	positive regulation of cell-cell adhesion	4/17	251/18493	6.86E-05	0.001299	0.000571	BCL6/GLI2/GLI3/ZBTB16	4
GO:0009953	GO:0009953	dorsal/ventral pattern formation	3/17	89/18493	6.98E-05	0.001299	0.000571	GLI1/GLI2/GLI3	3
GO:0060831	GO:0060831	smoothened signaling pathway involved in dorsal/ventral neural tube patterning	2/17	14/18493	7.19E-05	0.001305	0.000573	GLI2/GLI3	2
GO:0046634	GO:0046634	regulation of alpha-beta T cell activation	3/17	91/18493	7.46E-05	0.00132	0.00058	BCL6/GLI3/ZBTB16	3
GO:0060579	GO:0060579	ventral spinal cord interneuron fate commitment	2/17	15/18493	8.29E-05	0.001397	0.000614	GLI2/GLI3	2
GO:0060581	GO:0060581	cell fate commitment involved in pattern specification	2/17	15/18493	8.29E-05	0.001397	0.000614	GLI2/GLI3	2
GO:1902105	GO:1902105	regulation of leukocyte differentiation	4/17	265/18493	8.47E-05	0.001397	0.000614	BCL6/GLI2/GLI3/ZBTB16	4
GO:0046632	GO:0046632	alpha-beta T cell differentiation	3/17	101/18493	0.000102	0.001641	0.000721	BCL6/GLI3/ZBTB16	3
GO:0021514	GO:0021514	ventral spinal cord interneuron differentiation	2/17	17/18493	0.000107	0.001693	0.000744	GLI2/GLI3	2
GO:0048562	GO:0048562	embryonic organ morphogenesis	4/17	284/18493	0.000111	0.00171	0.000751	ZIC3/GLI1/GLI2/GLI3	4
GO:0006275	GO:0006275	regulation of DNA replication	3/17	107/18493	0.000121	0.001826	0.000802	BCL6/GLI1/GLI2	3
GO:1903037	GO:1903037	regulation of leukocyte cell-cell adhesion	4/17	293/18493	0.000125	0.00185	0.000813	BCL6/GLI2/GLI3/ZBTB16	4
GO:0050863	GO:0050863	regulation of T cell activation	4/17	305/18493	0.000146	0.002089	0.000918	BCL6/GLI2/GLI3/ZBTB16	4
GO:0030903	GO:0030903	notochord development	2/17	20/18493	0.00015	0.002089	0.000918	GLI1/GLI2	2
GO:0043586	GO:0043586	tongue development	2/17	20/18493	0.00015	0.002089	0.000918	GLI3/PRDM16	2
GO:0002062	GO:0002062	chondrocyte differentiation	3/17	117/18493	0.000157	0.002155	0.000947	GLI2/GLI3/ZBTB16	3
GO:0060562	GO:0060562	epithelial tube morphogenesis	4/17	313/18493	0.000161	0.002164	0.000951	ZEB2/ZIC3/GLI2/GLI3	4
GO:0032332	GO:0032332	positive regulation of chondrocyte differentiation	2/17	21/18493	0.000165	0.002182	0.000959	GLI3/ZBTB16	2
GO:0022612	GO:0022612	gland morphogenesis	3/17	120/18493	0.00017	0.002198	0.000966	GLI1/GLI2/GLI3	3
GO:0001655	GO:0001655	urogenital system development	4/17	322/18493	0.000179	0.002275	0.001	GLI1/GLI2/GLI3/ZBTB16	4
GO:0021513	GO:0021513	spinal cord dorsal/ventral patterning	2/17	22/18493	0.000182	0.002275	0.001	GLI2/GLI3	2
GO:0007159	GO:0007159	leukocyte cell-cell adhesion	4/17	327/18493	0.00019	0.002277	0.001	BCL6/GLI2/GLI3/ZBTB16	4
GO:0030326	GO:0030326	embryonic limb morphogenesis	3/17	125/18493	0.000191	0.002277	0.001	GLI2/GLI3/ZBTB16	3
GO:0035113	GO:0035113	embryonic appendage morphogenesis	3/17	125/18493	0.000191	0.002277	0.001	GLI2/GLI3/ZBTB16	3
GO:0030318	GO:0030318	melanocyte differentiation	2/17	23/18493	0.000199	0.002316	0.001017	ZEB2/GLI3	2

GO:0051251	GO:0051251	positive regulation of lymphocyte activation	4/17	333/18493	0.000204	0.002316	0.001017	BCL6/GLI2/GLI3/ZBTB16	4
GO:0045598	GO:0045598	regulation of fat cell differentiation	3/17	128/18493	0.000205	0.002316	0.001017	ZFPM2/PRDM16/ZBTB16	3
GO:0007224	GO:0007224	smoothened signaling pathway	3/17	129/18493	0.00021	0.002316	0.001017	GLI1/GLI2/GLI3	3
GO:0021511	GO:0021511	spinal cord patterning	2/17	24/18493	0.000217	0.002316	0.001017	GLI2/GLI3	2
GO:0021904	GO:0021904	dorsal/ventral neural tube patterning	2/17	24/18493	0.000217	0.002316	0.001017	GLI2/GLI3	2
GO:0046639	GO:0046639	negative regulation of alpha-beta T cell differentiation	2/17	24/18493	0.000217	0.002316	0.001017	BCL6/GLI3	2
GO:0030098	GO:0030098	lymphocyte differentiation	4/17	344/18493	0.000231	0.00241	0.001059	BCL6/GLI2/GLI3/ZBTB16	4
GO:0046631	GO:0046631	alpha-beta T cell activation	3/17	134/18493	0.000235	0.00241	0.001059	BCL6/GLI3/ZBTB16	3
GO:0021952	GO:0021952	central nervous system projection neuron axonogenesis	2/17	25/18493	0.000236	0.00241	0.001059	ZEB2/GLI2	2
GO:0016571	GO:0016571	histone methylation	3/17	135/18493	0.00024	0.00242	0.001063	CTCF/PRDM16/MECOM	3
GO:0001889	GO:0001889	liver development	3/17	137/18493	0.000251	0.002493	0.001095	ZIC3/GLI1/GLI3	3
GO:0021846	GO:0021846	cell proliferation in forebrain	2/17	26/18493	0.000255	0.002503	0.0011	ZEB2/GLI3	2
GO:0061008	GO:0061008	hepaticobiliary system development	3/17	140/18493	0.000267	0.002586	0.001136	ZIC3/GLI1/GLI3	3
GO:0018205	GO:0018205	peptidyl-lysine modification	4/17	361/18493	0.000278	0.002655	0.001166	CTCF/PRDM16/MECOM/DPF2	4
GO:0035107	GO:0035107	appendage morphogenesis	3/17	146/18493	0.000302	0.002813	0.001236	GLI2/GLI3/ZBTB16	3
GO:0035108	GO:0035108	limb morphogenesis	3/17	146/18493	0.000302	0.002813	0.001236	GLI2/GLI3/ZBTB16	3
GO:0002696	GO:0002696	positive regulation of leukocyte activation	4/17	372/18493	0.000312	0.002863	0.001258	BCL6/GLI2/GLI3/ZBTB16	4
GO:0030900	GO:0030900	forebrain development	4/17	374/18493	0.000318	0.002886	0.001268	ZEB2/GLI1/GLI2/GLI3	4
GO:1903707	GO:1903707	negative regulation of hemopoiesis	3/17	150/18493	0.000327	0.002933	0.001288	BCL6/GLI3/ZBTB16	3
GO:0001503	GO:0001503	ossification	4/17	380/18493	0.000338	0.002981	0.00131	GLI1/GLI2/GLI3/ZBTB16	4
GO:0021696	GO:0021696	cerebellar cortex morphogenesis	2/17	30/18493	0.000341	0.002981	0.00131	GLI1/GLI2	2
GO:0050867	GO:0050867	positive regulation of cell activation	4/17	384/18493	0.000352	0.003038	0.001335	BCL6/GLI2/GLI3/ZBTB16	4
GO:0022407	GO:0022407	regulation of cell-cell adhesion	4/17	388/18493	0.000366	0.003123	0.001372	BCL6/GLI2/GLI3/ZBTB16	4
GO:0021522	GO:0021522	spinal cord motor neuron differentiation	2/17	32/18493	0.000388	0.003239	0.001423	GLI2/GLI3	2
GO:0061036	GO:0061036	positive regulation of cartilage development	2/17	32/18493	0.000388	0.003239	0.001423	GLI3/ZBTB16	2
GO:0045785	GO:0045785	positive regulation of cell adhesion	4/17	397/18493	0.000399	0.003284	0.001443	BCL6/GLI2/GLI3/ZBTB16	4
GO:0021915	GO:0021915	neural tube development	3/17	161/18493	0.000403	0.003284	0.001443	ZEB2/GLI2/GLI3	3
GO:0021955	GO:0021955	central nervous system neuron axonogenesis	2/17	33/18493	0.000413	0.003295	0.001447	ZEB2/GLI2	2
GO:0050931	GO:0050931	pigment cell differentiation	2/17	33/18493	0.000413	0.003295	0.001447	ZEB2/GLI3	2
GO:0048566	GO:0048566	embryonic digestive tract development	2/17	34/18493	0.000439	0.003387	0.001488	GLI2/GLI3	2
GO:0048665	GO:0048665	neuron fate specification	2/17	34/18493	0.000439	0.003387	0.001488	GLI2/GLI3	2
GO:0097421	GO:0097421	liver regeneration	2/17	34/18493	0.000439	0.003387	0.001488	GLI1/GLI3	2
GO:1903706	GO:1903706	regulation of hemopoiesis	4/17	418/18493	0.000485	0.003629	0.001594	BCL6/GLI2/GLI3/ZBTB16	4

GO:0006479	GO:0006479	protein methylation	3/17	174/18493	0.000505	0.003629	0.001594	CTCF/PRDM16/MECOM	3
GO:0008213	GO:0008213	protein alkylation	3/17	174/18493	0.000505	0.003629	0.001594	CTCF/PRDM16/MECOM	3
GO:0021953	GO:0021953	central nervous system neuron differentiation	3/17	175/18493	0.000514	0.003629	0.001594	ZEB2/GLI2/GLI3	3
GO:0048736	GO:0048736	appendage development	3/17	175/18493	0.000514	0.003629	0.001594	GLI2/GLI3/ZBTB16	3
GO:0060173	GO:0060173	limb development	3/17	175/18493	0.000514	0.003629	0.001594	GLI2/GLI3/ZBTB16	3
GO:0021587	GO:0021587	cerebellum morphogenesis	2/17	37/18493	0.00052	0.003629	0.001594	GLI1/GLI2	2
GO:0045740	GO:0045740	positive regulation of DNA replication	2/17	37/18493	0.00052	0.003629	0.001594	GLI1/GLI2	2
GO:0046636	GO:0046636	negative regulation of alpha-beta T cell activation	2/17	37/18493	0.00052	0.003629	0.001594	BCL6/GLI3	2
GO:0060045	GO:0060045	positive regulation of cardiac muscle cell proliferation	2/17	37/18493	0.00052	0.003629	0.001594	GLI1/ZFPM2	2
GO:0048608	GO:0048608	reproductive structure development	4/17	428/18493	0.00053	0.003665	0.00161	GLI1/GLI2/GLI3/ZFPM2	4
GO:0061458	GO:0061458	reproductive system development	4/17	431/18493	0.000544	0.003727	0.001637	GLI1/GLI2/GLI3/ZFPM2	4
GO:0048732	GO:0048732	gland development	4/17	434/18493	0.000559	0.00379	0.001665	ZIC3/GLI1/GLI2/GLI3	4
GO:0021532	GO:0021532	neural tube patterning	2/17	40/18493	0.000608	0.004086	0.001795	GLI2/GLI3	2
GO:0021575	GO:0021575	hindbrain morphogenesis	2/17	41/18493	0.000639	0.004253	0.001869	GLI1/GLI2	2
GO:0042110	GO:0042110	T cell activation	4/17	451/18493	0.000645	0.00426	0.001871	BCL6/GLI2/GLI3/ZBTB16	4
GO:0050769	GO:0050769	positive regulation of neurogenesis	4/17	453/18493	0.000656	0.004292	0.001886	BCL6/ZEB2/GLI2/GLI3	4
GO:0045581	GO:0045581	negative regulation of T cell differentiation	2/17	42/18493	0.00067	0.004344	0.001908	BCL6/GLI3	2
GO:0030278	GO:0030278	regulation of ossification	3/17	193/18493	0.000683	0.004389	0.001928	GLI1/GLI3/ZBTB16	3
GO:0021983	GO:0021983	pituitary gland development	2/17	43/18493	0.000702	0.004474	0.001965	GLI1/GLI2	2
GO:0048066	GO:0048066	developmental pigmentation	2/17	44/18493	0.000736	0.004644	0.00204	ZEB2/GLI3	2
GO:0051249	GO:0051249	regulation of lymphocyte activation	4/17	472/18493	0.000766	0.004792	0.002105	BCL6/GLI2/GLI3/ZBTB16	4
GO:0021517	GO:0021517	ventral spinal cord development	2/17	46/18493	0.000804	0.004935	0.002168	GLI2/GLI3	2
GO:0060443	GO:0060443	mammary gland morphogenesis	2/17	46/18493	0.000804	0.004935	0.002168	GLI2/GLI3	2
GO:0002009	GO:0002009	morphogenesis of an epithelium	4/17	479/18493	0.000809	0.004935	0.002168	ZEB2/ZIC3/GLI2/GLI3	4
GO:0051216	GO:0051216	cartilage development	3/17	206/18493	0.000825	0.004994	0.002194	GLI2/GLI3/ZBTB16	3
GO:0046638	GO:0046638	positive regulation of alpha-beta T cell differentiation	2/17	47/18493	0.000839	0.004994	0.002194	GLI3/ZBTB16	2
GO:0055023	GO:0055023	positive regulation of cardiac muscle tissue growth	2/17	47/18493	0.000839	0.004994	0.002194	GLI1/ZFPM2	2
GO:0021695	GO:0021695	cerebellar cortex development	2/17	48/18493	0.000875	0.005166	0.002269	GLI1/GLI2	2
GO:0001649	GO:0001649	osteoblast differentiation	3/17	214/18493	0.000922	0.005397	0.002371	GLI1/GLI2/GLI3	3
GO:0045444	GO:0045444	fat cell differentiation	3/17	217/18493	0.00096	0.005573	0.002448	ZFPM2/PRDM16/ZBTB16	3
GO:0045620	GO:0045620	negative regulation of lymphocyte differentiation	2/17	51/18493	0.000988	0.005646	0.00248	BCL6/GLI3	2
GO:0060421	GO:0060421	positive regulation of heart growth	2/17	51/18493	0.000988	0.005646	0.00248	GLI1/ZFPM2	2
GO:0021515	GO:0021515	cell differentiation in spinal cord	2/17	52/18493	0.001027	0.005822	0.002558	GLI2/GLI3	2

GO:0060043	GO:0060043	regulation of cardiac muscle cell proliferation	2/17	55/18493	0.001148	0.00646	0.002838	GLI1/ZFPM2	2
GO:0035019	GO:0035019	somatic stem cell population maintenance	2/17	60/18493	0.001364	0.00762	0.003347	ZIC3/PRDM16	2
GO:0046635	GO:0046635	positive regulation of alpha-beta T cell activation	2/17	61/18493	0.00141	0.007754	0.003407	GLI3/ZBTB16	2
GO:0055025	GO:0055025	positive regulation of cardiac muscle tissue development	2/17	61/18493	0.00141	0.007754	0.003407	GLI1/ZFPM2	2
GO:0046622	GO:0046622	positive regulation of organ growth	2/17	62/18493	0.001456	0.007889	0.003466	GLI1/ZFPM2	2
GO:0048645	GO:0048645	animal organ formation	2/17	62/18493	0.001456	0.007889	0.003466	GLI2/GLI3	2
GO:0045600	GO:0045600	positive regulation of fat cell differentiation	2/17	64/18493	0.001551	0.008278	0.003637	PRDM16/ZBTB16	2
GO:0060038	GO:0060038	cardiac muscle cell proliferation	2/17	64/18493	0.001551	0.008278	0.003637	GLI1/ZFPM2	2
GO:0061448	GO:0061448	connective tissue development	3/17	263/18493	0.001668	0.008841	0.003884	GLI2/GLI3/ZBTB16	3
GO:0048663	GO:0048663	neuron fate commitment	2/17	67/18493	0.001698	0.008932	0.003924	GLI2/GLI3	2
GO:0006260	GO:0006260	DNA replication	3/17	268/18493	0.001761	0.009196	0.00404	BCL6/GLI1/GLI2	3
GO:0034504	GO:0034504	protein localization to nucleus	3/17	269/18493	0.001779	0.009228	0.004054	BCL6/GLI3/ZBTB16	3
GO:0033077	GO:0033077	T cell differentiation in thymus	2/17	69/18493	0.0018	0.009266	0.004071	GLI2/GLI3	2
GO:0001822	GO:0001822	kidney development	3/17	271/18493	0.001818	0.009293	0.004082	GLI2/GLI3/ZBTB16	3
GO:0070227	GO:0070227	lymphocyte apoptotic process	2/17	72/18493	0.001958	0.009938	0.004366	BCL6/GLI3	2

Additional File 5

	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
GO:0000076	GO:0000076	DNA replication checkpoint	3/6	16/18493	1.06E-08	1.53E-06	3.8E-07	CDC45/CDC6/ORC1	3
GO:0000083	GO:0000083	regulation of transcription involved in G1/S transition of mitotic cell cycle	3/6	29/18493	6.91E-08	4.5E-06	1.12E-06	CDC45/CDC6/ORC1	3
GO:0006270	GO:0006270	DNA replication initiation	3/6	32/18493	9.38E-08	4.5E-06	1.12E-06	CDC45/CDC6/ORC1	3
GO:0006302	GO:0006302	double-strand break repair	4/6	236/18493	3.8E-07	1.37E-05	3.4E-06	FANCB/ESCO2/XRCC2/CDC45	4
GO:0006260	GO:0006260	DNA replication	4/6	268/18493	6.32E-07	1.82E-05	4.53E-06	ESCO2/CDC45/CDC6/ORC1	4
GO:0033314	GO:0033314	mitotic DNA replication checkpoint	2/6	10/18493	3.94E-06	9.46E-05	2.35E-05	CDC6/ORC1	2
GO:0000724	GO:0000724	double-strand break repair via homologous recombination	3/6	123/18493	5.66E-06	0.000104	2.59E-05	FANCB/XRCC2/CDC45	3
GO:0000725	GO:0000725	recombinational repair	3/6	124/18493	5.8E-06	0.000104	2.59E-05	FANCB/XRCC2/CDC45	3
GO:0006261	GO:0006261	DNA-dependent DNA replication	3/6	148/18493	9.87E-06	0.000158	3.93E-05	CDC45/CDC6/ORC1	3
GO:0031570	GO:0031570	DNA integrity checkpoint	3/6	158/18493	1.2E-05	0.000173	4.3E-05	CDC45/CDC6/ORC1	3
GO:0000075	GO:0000075	cell cycle checkpoint	3/6	216/18493	3.06E-05	0.000401	9.96E-05	CDC45/CDC6/ORC1	3
GO:0044818	GO:0044818	mitotic G2/M transition checkpoint	2/6	31/18493	4.06E-05	0.000487	0.000121	CDC6/ORC1	2
GO:0000082	GO:0000082	G1/S transition of mitotic cell cycle	3/6	263/18493	5.51E-05	0.00061	0.000152	CDC45/CDC6/ORC1	3
GO:0006310	GO:0006310	DNA recombination	3/6	282/18493	6.78E-05	0.000651	0.000162	FANCB/XRCC2/CDC45	3
GO:0044843	GO:0044843	cell cycle G1/S phase transition	3/6	282/18493	6.78E-05	0.000651	0.000162	CDC45/CDC6/ORC1	3
GO:1900087	GO:1900087	positive regulation of G1/S transition of mitotic cell cycle	2/6	43/18493	7.87E-05	0.000709	0.000176	CDC45/CDC6	2
GO:0010972	GO:0010972	negative regulation of G2/M transition of mitotic cell cycle	2/6	48/18493	9.83E-05	0.000833	0.000207	CDC6/ORC1	2
GO:1902808	GO:1902808	positive regulation of cell cycle G1/S phase transition	2/6	53/18493	0.00012	0.00096	0.000239	CDC45/CDC6	2
GO:1902750	GO:1902750	negative regulation of cell cycle G2/M phase transition	2/6	60/18493	0.000154	0.001138	0.000283	CDC6/ORC1	2
GO:1901990	GO:1901990	regulation of mitotic cell cycle phase transition	3/6	375/18493	0.000158	0.001138	0.000283	CDC45/CDC6/ORC1	3
GO:0051052	GO:0051052	regulation of DNA metabolic process	3/6	412/18493	0.000209	0.001367	0.00034	FANCB/ESCO2/CDC6	3
GO:1901987	GO:1901987	regulation of cell cycle phase transition	3/6	412/18493	0.000209	0.001367	0.00034	CDC45/CDC6/ORC1	3
GO:1901992	GO:1901992	positive regulation of mitotic cell cycle phase transition	2/6	85/18493	0.000309	0.001937	0.000482	CDC45/CDC6	2
GO:1901989	GO:1901989	positive regulation of cell cycle phase transition	2/6	100/18493	0.000428	0.002569	0.000638	CDC45/CDC6	2
GO:0006275	GO:0006275	regulation of DNA replication	2/6	107/18493	0.00049	0.002714	0.000675	ESCO2/CDC6	2
GO:0044774	GO:0044774	mitotic DNA integrity checkpoint	2/6	107/18493	0.00049	0.002714	0.000675	CDC6/ORC1	2

GO:0051053	GO:0051053	negative regulation of DNA metabolic process	2/6	152/18493	0.000985	0.005133	0.001276	FANCB/CDC6	2
GO:0010389	GO:0010389	regulation of G2/M transition of mitotic cell cycle	2/6	153/18493	0.000998	0.005133	0.001276	CDC6/ORC1	2
GO:0045931	GO:0045931	positive regulation of mitotic cell cycle	2/6	157/18493	0.001051	0.005216	0.001296	CDC45/CDC6	2
GO:0007093	GO:0007093	mitotic cell cycle checkpoint	2/6	166/18493	0.001173	0.00545	0.001355	CDC6/ORC1	2
GO:2000045	GO:2000045	regulation of G1/S transition of mitotic cell cycle	2/6	166/18493	0.001173	0.00545	0.001355	CDC45/CDC6	2
GO:1902749	GO:1902749	regulation of cell cycle G2/M phase transition	2/6	170/18493	0.00123	0.005535	0.001376	CDC6/ORC1	2
GO:0000819	GO:0000819	sister chromatid segregation	2/6	185/18493	0.001454	0.006159	0.001531	ESCO2/CDC6	2
GO:1902806	GO:1902806	regulation of cell cycle G1/S phase transition	2/6	185/18493	0.001454	0.006159	0.001531	CDC45/CDC6	2
GO:1901991	GO:1901991	negative regulation of mitotic cell cycle phase transition	2/6	191/18493	0.001549	0.006373	0.001584	CDC6/ORC1	2
GO:0000086	GO:0000086	G2/M transition of mitotic cell cycle	2/6	199/18493	0.00168	0.006719	0.00167	CDC6/ORC1	2
GO:1901988	GO:1901988	negative regulation of cell cycle phase transition	2/6	211/18493	0.001886	0.007339	0.001824	CDC6/ORC1	2
GO:0044839	GO:0044839	cell cycle G2/M phase transition	2/6	218/18493	0.002011	0.007621	0.001894	CDC6/ORC1	2

CAPÍTULO V

CONSIDERAÇÕES FINAIS

4. CONSIDERAÇÕES FINAIS

Essa dissertação teve como objetivo principal avaliar o papel do gene *HAND2* na TE, através da caracterização molecular do gene e de seu envolvimento da TE. Avaliamos o papel do gene na suscetibilidade genética à embriopatia. Caracterizamos *in silico* a variante p.P51 observada em indivíduos portadores da TE. Finalmente, selecionamos novos alvos de interesse considerando a interação entre HAND2 e TBX5, a qual é inibida na presença da talidomida, e inserimos esta interação no contexto das malformações de coração e de membros compatíveis com a TE.

A caracterização molecular de *HAND2* nos mostrou que ele é um gene muito bem conservado, tendo poucas variantes presentes em sua sequência e um número menor ainda quando consideramos apenas variantes polimórficas. A caracterização revelou que a maior parte das variantes não são localizadas em regiões codificantes. A região mais polimórfica observada foi a região 3'UTR. A alta conservação do gene era de certa forma esperada, considerando que *HAND2* é um gene de desenvolvimento, os quais são, em sua maioria, bem conservados evolutivamente (Voth et al. 2009). Além disso, o grupo bHLH, ao qual HAND2 faz parte, também é um grupo conservado, sendo bem conservado desde leveduras até humanos (Atchley and Fitch 1997).

Mutações em *HAND2* são associadas a doenças cardiovasculares. A variante p.L47P (rs1553974835) foi observada em um homem com tetralogia de Fallot (Lu et al. 2016), enquanto a variante p.G202V foi identificada em cinco indivíduos que possuíam tetralogia de Fallot, coarctação da aorta e estenose aórtica (Khalil et al. 2017). Ademais, um estudo associou uma variante *missense* com casos familiares de defeito do septo ventricular e estenose pulmonar (Sun et al. 2016). No nosso estudo, observamos a presença de uma variante polimórfica sinônima, p.P51 (rs59621536), em indivíduos com TE. No entanto, não foi possível associar a variante a fenótipos específicos da TE, provavelmente devido à escassez de dados dos indivíduos assim como o tamanho amostral avaliado. Contudo, a caracterização funcional *in silico* mostrou que a variante apresenta a

capacidade de alterar ilhas CpGs localizadas próximas à variante, além de alterar o sítio de ligação de uma proteína com capacidade de reconhecer regiões metiladas, MeCP2. Isso sugere que a variante pode ter efeitos epigenéticos. Apesar disso, não observamos alterações nos padrões de metilação em indivíduos com TE. Isso não exclui a possibilidade de modificações epigenéticas pela variante que possam estar associadas com outras anomalias cardíacas que não as observada na TE. Além disso, é importante ressaltar que padrões de metilação são tecido e tempo específicas (Reik et al. 2001), podendo ser a razão de não termos observados alterações epigenéticas em nossa amostra de saliva de indivíduos portadores da TE.

A partir da caracterização funcional *in silico* da variante p.P51 também foi possível observar que ela leva à alteração na estrutura do mRNA de HAND2. A variante leva à modificação de um grampo no mRNA com a adição de uma alça interna. Alterações deste tipo são desestabilizadoras da sequência e são conhecidas por permitirem a ligação de sítios de miRNA (Zhao et al. 2005). No entanto, não observamos nenhuma alteração de sítios de ligação de miRNA na presença da variante p.P51. Não localizamos estudos descrevendo sítios de miRNA no local da variante ou possíveis consequências da alteração da estrutura do mRNA de HAND2. Mais estudos são necessários para avaliar se essa alteração na estrutura afeta a estabilidade do mRNA ou a tradução da proteína.

De acordo com as análises realizadas, é possível observar que apesar da variante p.P51 não ter apresentado papel na suscetibilidade genética à TE, não se pode excluir sua participação na susceptibilidade ao aparecimento de outras malformações ou até mesmo condições de aparecimento tardio. Mais estudos são necessários para averiguar esta hipótese. Além disso, também deve-se levar em consideração o potencial epigenético e de alteração do mRNA da variante p.P51 em *HAND2*.

Poucos genes têm sido avaliados para suscetibilidade genética à TE. Desta forma, torna-se importante a seleção de novos genes de interesse para estudar seu possível papel na suscetibilidade genética à TE. Além disso, os mecanismos moleculares de teratogenicidade da talidomida não estão completamente

elucidados, tornando-se necessário a seleção de novas proteínas de interesse para estudos de mecanismo. Desta forma, buscamos selecionar novos alvos de interesse para estudos na TE através de diversas abordagens, utilizando a biologia de sistemas como método. A partir da interação de HAND2 e TBX5, avaliamos as malformações de coração e de membros compatíveis com as observadas na TE, buscando pelas proteínas importantes para manter as conexões da rede, além de proteínas que possuem domínio C2H2 e genes diferencialmente expressos após tratamento com talidomida. Vale lembrar que proteínas com domínio C2H2 foram observadas como neosubstratos do complexo $CRL4^{CRBN}$ quando ligado à talidomida, levando a marcação destas proteínas para degradação (Matyskiela et al. 2018; Sievers et al. 2018). Este, portanto, é um provável mecanismo que pode explicar como ocorre as alterações de expressão gênica causadas pela talidomida.

Avaliando a rede de interação entre HAND2 e TBX5 observamos que as proteínas UBB, JUN e TCF4 foram as proteínas mais importantes para manter as conexões. O gene *UBB* codifica uma proteína ubiquitina que faz parte da mesma cascata de ubiquitinação de Cereblon (Wiborg et al. 1985; Scheffner et al. 1995; Ito et al. 2010). Além disso, JUN já foi visto ser ubiquitinado e degradado por Cereblon (Meng and Xia 2011; Yang et al. 2018). Já TCF4 faz parte do grupo bHLH de proteínas, mesmo grupo de HAND2, sendo que ele atua em conjunto com outras proteínas bHLH no desenvolvimento do esqueleto (Ah Cho and Dressler 1998; Dai and Cserjesi 2002; Forrest et al. 2014).

Já para as redes de malformações de coração e membros (Redes Cardiac e Limbs, respectivamente) foi principalmente EP300 que apareceu como importante nas duas redes. A rede de interação de HAND2 e TBX5 (Rede Mater) foi conectada às redes de malformações de coração (Rede Cardiac) e de membros (Rede Limbs) formando duas novas redes, uma com interação de HAND2 e TBX5 e malformação de coração (Rede Mater/Cardiac) e outra com interação de HAND2 e TBX5 e malformação de membros (Rede Mater/Limbs). Para manter a conexão entre a interação de HAND2 e TBX5 com tanto as malformações de coração quanto com as de membros de TE, foram importantes as proteínas EP300, SRCAP, FANCM e TBX5. Interessantemente, EP300 apareceu como importante em todas as redes analisadas.

Foi relatado que EP300 teria um papel relevante durante o desenvolvimento, principalmente para proliferação celular (Yao et al. 1998). Também foram observadas malformações cardiovasculares associadas à letalidade em camundongos nocaute para p300. Interessantemente, também observou-se letalidade em torno de 40% dos embriões humanos expostos à talidomida, isso devido a anormalidades graves nos órgãos internos, especialmente no coração (Lenz 1988; Smithells and Newman 1992; Vargesson 2009). Além disso, EP300 é associada com a síndrome de Rubinstein-Taybi, que causa anormalidade faciais, restrição de crescimento, entre outras anormalidades (Vargesson 2009; Fergelot et al. 2016). Malformações cardiovasculares, como defeitos do septo ventricular e estenose pulmonar, já foram descritas em indivíduos com a síndrome de Rubistein-Taybi. Anomalias faciais, de esqueleto e as anomalias cardiovasculares citadas são também observadas em indivíduos com TE (Smithells and Newman 1992; Fergelot et al. 2016).

Além de *EP300* estar associado a fenótipos semelhantes com os observados na TE, a proteína p300, em conjunto com CBP (*CREBB binding protein)*, foi vista realizar a acetilação de um alvo (Glutamina Sintetase) do complexo $CRL4^{CRBN}$, permitindo o reconhecimento deste alvo pelo complexo. CBP/p300 funcionam como acetiltransferases de histonas e co-fatores de transcrição. Desta forma, CBP/p300 realiza a acetilação de um domínio de lisina KxxK, localizado na proteína Glutamina Sintetase, que quando acetilado se liga ao domínio carboxi-terminal de CRBN, mesmo domínio ao qual IMiDs se ligam. Ademais, essa ligação de IMiDs à CRBN promoveu uma maior interação entre CRBN e o domínio KxxK acetilado da Glutamina Sintetase (Nguyen et al. 2016). Desta forma, p300 poderia estar atuando nas acetilações de domínios KxxK, os quais são reconhecidos pelo complexo $CRL4^{CRBN}$-IMiDs.

SRCAP e FANCM também foram importantes para manter a conexão entre a interação de HAND2 e TBX5 tanto com as malformações de coração quanto com as de membros de TE. Mutações no gene de SRCAP estão associadas com a síndrome de Floating-Harbor, que é caracterizada por anormalidades craniofaciais, atraso de crescimento ósseo, entre outros (Nikkel et al. 2013). Ademais, um caso de síndrome de Floating-Harbor foi relatado estar acompanhado de uma

anormalidade cardiovascular, a cardiomiopatia dilatada (Hou et al. 2021). No caso de FANCM, ele está relacionado à anemia de Fanconi, que pode apresentar um conjunto diverso de malformações, incluindo anormalidades cardiovasculares, malformações faciais, do polegar e rádio, entre outras (Titus et al. 2009; Soulier 2011). Devido às similaridades das malformações da TE com a anemia de Fanconi, elas são consideradas fenocópias.

Vale ressaltar que MeCP2, proteína a qual o sítio de ligação em *HAND2* foi predito ser perdido a presença da variante p.P51, foi visto ser importante na rede de conexão entre a rede de interação de HAND2 e TBX5 e a rede de malformações de membros. MeCP2 é uma proteína nuclear com domínio ligador de metil-CpG. Ela realiza o reconhecimento e ligação à ilhas CpGs e remodelamento da cromatina (Bienvenu and Chelly 2006). Mutações em MeCP2 estão associadas com a síndrome de Rett, síndrome na qual a maior parte das mortes está relacionada a disfunções cardíacas (Hara et al. 2015). Além disso, desregulação da expressão de MeCP2 causa malformações cardíacas e esqueléticas em camundongos (Alvarez-Saavedra et al. 2010). A expressão de MeCP2 foi relatada ser necessária para a formação normal do coração (Alvarez-Saavedra et al. 2010; Hara et al. 2015) e a sua desregulação foi associada com malformações cardíacas e esqueléticas em camundongos (Alvarez-Saavedra et al. 2010). Ademais, MeCP2 parece regular a expressão cardíaca adulta e embriônicas de alguns fatores de transcrição importantes no desenvolvimento cardiovascular, como TBX5, o qual possui sítios de ligação para MeCP2, os quais a ligação de MeCP2 leva à diminuição da expressão de TBX5 (Hara et al. 2015).

Com relação às análises de proteínas com domínio C2H2, observamos três proteínas na rede de interação de HAND2 e TBX5. As proteínas BCL6, ZNF133 e ZNF281 podem representar alvos do complexo $CRL4^{CRBN}$-talidomida na rede de interação entre HAND2 e TBX5 por possuírem o domínio C2H2. Dentre as malformações de coração e de membros compatíveis com a TE, nós observamos cinco proteínas com domínio C2H2 em comum entre as duas redes: ZIC3, GLI1, GLI3, PRDM16 e ZNF148. Essas proteínas seriam neosubstratos que se afetados pelo complexo $CRL4^{CRBN}$-talidomida poderiam explicar o surgimento das malformações de coração e membros ao mesmo tempo.

Ademais, analisamos genes diferencialmente expressos em células tratadas com talidomida. Observamos sete genes diferencialmente expressos, sendo que três estavam presentes tanto na rede de malformações de coração quanto na rede de malformação de membros da TE. Os genes *FANCB*, *ESCO2* e *XRCC2* estavam com expressão menor após o tratamento com talidomida. Interessantemente, mutações em *ESCO2* são associadas com a síndrome de Roberts, a qual a TE é considerada uma fenocópia (Vega et al. 2005). Além disso, FANCB, assim como FANCM, está relacionada a anemia de Fanconi, que como já mencionado, a TE também é considerada uma fenocópia devido às similaridades de fenótipos (Titus et al. 2009).

As ferramentas de biologia de sistemas nos permitiram avaliar as redes de interação proteína-proteína de forma mais ampla, indo em busca de alvos importantes na manutenção dessas. Isso nos permitiu selecionar tais proteínas buscando também entender como elas se encaixariam no contexto da teratogênese da talidomida. Muitas estavam envolvidas com desenvolvimento embrionário e também com fenocópias da TE. Tal avaliação ampla das redes de interação torna-se importante já que a teratogênese da talidomida afeta diversos órgãos. Dessa forma, é mais provável que diversos alvos estejam sendo afetados e participem dos efeitos observados na TE. É importante salientar que os dados obtidos através da biologia de sistema podem e devem ser validados com estudos experimentais *in vitro* e *in vivo.* EP300, assim como os alvos da Anemia de Fanconi, FANCM e FANCB, tornaram-se alvos interessantes de estudo, sendo necessária a avaliação dos mesmos na suscetibilidade genética à TE assim como nos mecanismos de teratogênese da talidomida.

É importante salientar que mesmo não observando um papel de *HAND2* na suscetibilidade genética à TE, isso não exclui o papel de HAND2 nos mecanismos de teratogênese da talidomida. Isso, pois, algumas proteínas que sabidamente atuam nos mecanismos de teratogênese da talidomida não tiveram papel na suscetibilidade genética, como foi o caso para *VEGFA, FGF8* e *FGF10* (Kowalski et al. 2017; Gomes et al. 2018). HAND2, além de seu papel demonstrado a partir da inibição de sua interação com TBX5 pela talidomida, que faz com que alvos deste complexo não sejam regulados na presença do medicamento, pode estar

atuando no mecanismo de antiangiogênese da talidomida. HAND2 já foi visto atuar em processo de angiogênese e desenvolvimento vascular (Yamagishi et al. 2000). Além disso, fenótipos resultantes do nocaute de *HAND2* foram muito similares aos observados para o nocaute de *VEGF* (Yamagishi et al. 2000). VEGF sabidamente atua nos efeitos de antiangiogênese da talidomida, já que foram observados efeitos antiangiogênicos especialmente através da depleção de receptores de VEGF (Yabu et al. 2005). Além disso, HAND2 atua regulando VEGF no desenvolvimento do coração (Yamagishi et al. 2000). Portanto, HAND2 poderia atuar nos mecanismos de teratogênese da talidomida via processo de antiangiogênese através de VEGF.

Como limitações do estudo podemos citar a não avaliação da predição de alteração do mRNA de HAND2 pela variante p.P51 em indivíduos afetados pela TE. Fomos capazes de analisar a predição feita da metilação, porém não analisamos a predição de alteração da estrutura do mRNA. Dessa forma, não sabemos se essa alteração estaria modificando os níveis de expressão de HAND2 ou até auxiliando no provável papel de HAND2 na teratogênese da talidomida. Além disso, na seleção de novos genes alvos para estudos de suscetibilidade genética à TE, tivemos como principal limitação a escassez de dados relacionados à exposição da talidomida em humanos. Existe apenas um estudo avaliando a expressão após exposição a talidomida em humanos, sendo esta exposição em células (Schwartz et al. 2015). Dessa forma, necessitamos partir de dados de banco de dados e literatura para construir o fluxo de hipóteses e ao final validar os achados com dados de expressão.

Em conclusão, a partir dos resultados obtidos foi possível cumprir com os objetivos propostos dessa dissertação. Realizamos a caracterização molecular *in silico* do gene *HAND2* observando a alta conservação do gene. Avaliamos o gene *HAND2* nos indivíduos com TE e identificamos a variante p.P51 (rs59621536). Ao avaliar o impacto dessa variante, observamos o potencial de alteração na metilação do gene *HAND2* em indivíduos portadores da TE. No entanto, tal impacto não foi identificado quando avaliamos os padrões de metilação naqueles contendo a variante p.P51 comparado aos que não apresentam. Avaliamos os efeitos da variante p.P51 no mRNA do gene *HAND2* por meio de ferramentas *in silico* e

observamos a alteração da estrutura do mRNA na presença da variante, por uma adição de uma alça interna em um grampo. E por fim, selecionamos genes de interesse para serem avaliados na susceptibilidade genética à TE por meio de ferramentas de bioinformática e biologia de sistemas, observando principalmente EP300 como um bom candidato a estudos de suscetibilidade genética à TE.

REFERÊNCIAS BIBLIOGRÁFICAS

Ah Cho E and Dressler GR (1998) TCF-4 binds β-catenin and is expressed in distinct regions of the embryonic brain and limbs. Mech Dev 77:9–18. doi: 10.1016/S0925-4773(98)00131-2

Alvarez-Saavedra M, Carrasco L, Sura-Trueba S, Aiello VD, Walz K, Neto JX and Young JI (2010) Elevated expression of MeCP2 in cardiac and skeletal tissues is detrimental for normal development. Hum Mol Genet 19:2177–2190. doi: 10.1093/hmg/ddq096

An J, Ponthier CM, Sack R, Seebacher J, Stadler MB, Donovan KA and Fischer ES (2017) PSILAC mass spectrometry reveals ZFP91 as IMiD-dependent substrate of the CRL4 CRBN ubiquitin ligase. Nat Commun. doi: 10.1038/ncomms15398

Atchley WR and Fitch WM (1997) A natural classification of the basic helix-loop-helix class of transcription factors. Proc Natl Acad Sci U S A 94:5172–5176. doi: 10.1073/pnas.94.10.5172

Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, Soults J, Grayzel D, Kroumpouzou E, Traill TA, Leblanc-Straceski J et al. (1997) Mutations in human TBX5 cause limb and cardiac malformation in Holt-Oram syndrome. Nat Genet 15:30–5. doi: 10.1038/ng0197-30

Bienvenu T and Chelly J (2006) Molecular genetics of Rett syndrome: When DNA methylation goes unrecognized. Nat Rev Genet 7:415–426. doi: 10.1038/nrg1878

Brent RL (2001) The cause and prevention of human birth defects : What have we learned in the past 50 years ? Congenit Anom (Kyoto) 41:3–21.

Bruneau BG, Nemer G, Schmitt JP, Charron F, Robitaille L, Caron S, Conner DA, Gessler M, Nemer M, Seidman CE et al. (2001) A murine model of Holt-Oram syndrome defines roles of the T-Box transcription factor Tbx5 in cardiogenesis and disease. Cell 106:709–721. doi: 10.1016/S0092-8674(01)00493-7

Calabrese L and Fleischer AB (2000) Thalidomide: current and potential clinical

applications. Am J Med 108:487–95.

Cassina M, Salviati L, Di Gianantonio E and Clementi M (2012) Genetic susceptibility to teratogens: State of the art. Reprod Toxicol 34:186–191. doi: 10.1016/j.reprotox.2012.05.004

Castilla EE, Ashton-Prolla P, Barreda-Mejia E, Brunoni D, Cavalcanti DP, Correa-Neto J, Delgadillo JL, Dutra MG, Felix T, Giraldo A et al. (1996) Thalidomide, a current teratogen in South America. Teratology 54:273–7. doi: 10.1002/(SICI)1096-9926(199612)54:6<273::AID-TERA1>3.0.CO;2-#

Charité J, McFadden DG and Olson EN (2000) The bHLH transcription factor dHAND controls Sonic hedgehog expression and establishment of the zone of polarizing activity during limb development. Development 127:2461–70.

Clowes C, Boylan MGS, Ridge LA, Barnes E, Wright JA and Hentges KE (2014) The functional diversity of essential genes required for mammalian cardiac development. genesis 52:713–737. doi: 10.1002/dvg.22794

Cox R, Bouzekri N, Martin S, Southam L, Hugill A, Golamaully M, Cooper R, Adeyemo A, Soubrier F, Ward R et al. (2002) Okihiro syndrome is caused by SALL4 mutations. Hum Mol Genet 11:2979–2987. doi: 10.1093/hmg/11.23.2979

D'Amato RJ, Loughnan MS, Flynn E and Folkman J (1994) Thalidomide is an inhibitor of angiogenesis. Proc Natl Acad Sci 91:4082–4085. doi: 10.1073/pnas.91.9.4082

Dai Y-S and Cserjesi P (2002) The basic helix-loop-helix factor, HAND2, functions as a transcriptional activator by binding to E-boxes as a heterodimer. J Biol Chem 277:12604–12. doi: 10.1074/jbc.M200283200

Darlong J, Govindharaj P, Charles DE, Menzies A and Mani S (2016) Experiences with Thalidomide for Erythema Nodosum Leprosum – a retrospective study. Lepr Rev 87:211–220.

Davies FE, Raje N, Hideshima T, Lentzsch S, Young G, Tai YT, Lin B, Podar K, Gupta D, Chauhan D et al. (2001) Thalidomide and immunomodulatory derivatives augment natural killer cell cytotoxicity in multiple myeloma. Blood 98:210–216. doi:

10.1182/blood.V98.1.210

Donovan KA, An J, Nowak RP, Yuan JC, Fink EC, Berry BC, Ebert BL and Fischer ES (2018) Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane radial ray syndrome. Elife 7:1–25. doi: 10.7554/eLife.38430

Fergelot P, Van Belzen M, Van Gils J, Afenjar A, Armour CM, Arveiler B, Beets L, Burglen L, Busa T, Collet M et al. (2016) Phenotype and genotype in 52 patients with Rubinstein-Taybi syndrome caused by *EP300* mutations. Am J Med Genet Part A 170:3069–3082. doi: 10.1002/ajmg.a.37940

Firulli BA, Hadzic DB, McDaid JR and Firulli AB (2000) The basic helix-loop-helix transcription factors dHAND and eHAND exhibit dimerization characteristics that suggest complex regulation of function*. J Biol Chem 275:33567–33573. doi: 10.1074/jbc.M005888200

Forrest MP, Hill MJ, Quantock AJ, Martin-Rendon E and Blake DJ (2014) The emerging roles of TCF4 in disease and development. Trends Mol Med 20:322–331. doi: 10.1016/j.molmed.2014.01.010

Gao S, Wang S, Fan R and Hu J (2020) Recent advances in the molecular mechanism of thalidomide teratogenicity. Biomed Pharmacother 127:110114. doi: 10.1016/j.biopha.2020.110114

Gilbert-Barness E (2010) Teratogenic causes of malformations. Ann Clin Lab Sci 40:99–114.

Gomes J do A, Kowalski TW, Fraga LR, Macedo GS, Sanseverino MTV, Schuler-Faccini L and Vianna FSL (2019) The role of ESCO2, SALL4 and TBX5 genes in the susceptibility to thalidomide teratogenesis. Sci Rep 9:11413. doi: 10.1038/s41598-019-47739-8

Gomes J do A, Kowalski TW, Fraga LR, Tovo-Rodrigues L, Sanseverino MTV, Schuler-Faccini L and Vianna FSL (2018) Genetic susceptibility to thalidomide embryopathy in humans: Study of candidate development genes. Birth Defects Res 110:456–461. doi: 10.1002/bdr2.1163

Gregg NM (1941) CONGENITAL CATARACT FOLLOWING GERMAN MEASLES

IN THE MOTHER. Trans Ophthalmol Soc Aust 3:35–46.

Hansen JM, Gong S-G, Philbert M and Harris C (2002) Misregulation of gene expression in the redox-sensitive NF-Kb-dependent limb outgrowth pathway by thalidomide. Dev Dyn 225:186–194. doi: 10.1002/dvdy.10150

Hansen JM and Harris C (2004) A Novel Hypothesis for Thalidomide-Induced Limb Teratogenesis: Redox Misregulation of the NF-κB Pathway. Antioxid Redox Signal 6:1–14. doi: 10.1089/152308604771978291

Hara M, Takahashi T, Mitsumasu C, Igata S, Takano M, Minami T, Yasukawa H, Okayama S, Nakamura K, Okabe Y et al. (2015) Disturbance of cardiac gene expression and cardiomyocyte structure predisposes Mecp2-null mice to arrhythmias. Sci Rep 5:1–17. doi: 10.1038/srep11204

Haslett PAJ, Corral LG, Albert M and Kaplan G (1998) Thalidomide costimulates primary human t lymphocytes, preferentially inducing proliferation, cytokine production, and cytotoxic responses in the CD8+ subset. J Exp Med 187:1885–1892. doi: 10.1084/jem.187.11.1885

Hatcher CJ, Goldstein MM, Mah CS, Delia CS and Basson CT (2000) Identification and localization of TBX5 transcription factor during human cardiac morphogenesis. Dev Dyn 219:90–95. doi: 10.1002/1097-0177(200009)219:1<90::AID-DVDY1033>3.0.CO;2-L

Hou C, Xie L, Qiu Q, Lin H, Liu W, Sun X, Zhang Y, Xu M, Li Y and Xiao T (2021) Generation of an induced pluripotent stem cell line from a Chinese Han infant with floating-harbor syndrome accompanied with dilated cardiomyopathy. Stem Cell Res 51:1873–5061. doi: 10.1016/j.scr.2021.102182

Ito T, Ando H, Suzuki T, Ogura T, Hotta K, Imamura Y, Yamaguchi Y and Handa H (2010) Identification of a primary target of thalidomide teratogenicity. Science (80-). doi: 10.1126/science.1177319

James WH (1965) Teratogenetic properties of thalidomide. Br Med J 2:1064. doi: 10.1136/bmj.2.5469.1064-b

Jurand A (1966) Early changes in limb buds of chick embryos after thalidomide

treatment. Embryol exp Morph 16:289–300.

Kajii T, Kida M and Takahashi K (1973) The effect of thalidomide intake during 113 human pregnancies. Teratology 8:163–166. doi: 10.1002/tera.1420080208

Kang M and Ghassemzadeh S (2018) Toxicity, Benzodiazepine.

Khalil A, Tanos R, El-Hachem N, Kurban M, Bouvagnet P, Bitar F and Nemer G (2017) A HAND to TBX5 Explains the Link Between Thalidomide and Cardiac Diseases. Sci Rep 7:1416. doi: 10.1038/s41598-017-01641-3

Kim JH and Scialli AR (2011) Thalidomide: The Tragedy of Birth Defects and the Effective Treatment of Disease. Toxicol Sci 122:1–6. doi: 10.1093/toxsci/kfr088

Kowalski TW, Fraga LR, Tovo-Rodrigues L, Sanseverino MTV, Hutz MH, Schuler-Faccini L and Vianna FSL (2016) New Findings in eNOS gene and Thalidomide Embryopathy Suggest pre-transcriptional effect variants as susceptibility factors. Sci Rep 6:1–6. doi: 10.1038/srep23404

Kowalski TW, Fraga LR, Tovo-Rodrigues L, Sanseverino MTV, Hutz MH, Schuler-Faccini L and Vianna FSL (2017) Angiogenesis-related Genes and Thalidomide Teratogenesis in Humans: An Approach on Genetic Variation and Review of Past In Vitro Studies. Reprod Toxicol. doi: 10.1016/j.reprotox.2017.01.012

Kowalski TW, Gomes J do A, Garcia GBC, Fraga LR, Paixao-Cortes VR, Recamonde-Mendoza M, Sanseverino MTV, Schuler-Faccini L and Vianna FSL (2020) CRL4-Cereblon complex in Thalidomide Embryopathy: a translational investigation. Sci Rep 10:1–13. doi: 10.1038/s41598-020-57512-x

Krönke J, Fink EC, Hollenbach PW, MacBeth KJ, Hurst SN, Udeshi ND, Chamberlain PP, Mani DR, Man HW, Gandhi AK et al. (2015) Lenalidomide induces ubiquitination and degradation of CK1α in del(5q) MDS. Nature 523:183–188. doi: 10.1038/nature14610

Krönke J, Udeshi ND, Narla A, Grauman P, Hurst SN, McConkey M, Svinkina T, Heckl D, Comer E, Li X et al. (2014) Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. Science (80-) 343:301–305. doi: 10.1126/science.1244851

Lenz W (1988) A short history of thalidomide embryopathy. Teratology 38:203–215. doi: 10.1002/tera.1420380303

Liberatore CM, Searcy-Schrick RD and Yutzey KE (2000) Ventricular Expression of tbx5 Inhibits Normal Heart Chamber Development. Dev Biol 223:169–180. doi: 10.1006/dbio.2000.9748

Lu C-X, Gong H-R, Liu X-Y, Wang J, Zhao C-M, Huang R-T, Xue S and Yang Y-Q (2016) A novel HAND2 loss-of-function mutation responsible for tetralogy of Fallot. Int J Mol Med 37:445–451.

Lu G, Middleton RE, Sun H, Naniong MV, Ott CJ, Mitsiades CS, Wong KK, Bradner JE and Kaelin WG (2014) The myeloma drug lenalidomide promotes the cereblon-dependent destruction of ikaros proteins. Science (80-) 343:305–309. doi: 10.1126/science.1244917

Matthews SJ and McCoy C (2003) Peginterferon alfa-2a: A review of approved and investigational uses. Clin Ther 26:991–1025. doi: 10.1016/S0149-2918(04)90173-7

Matyskiela ME, Couto S, Zheng X, Lu G, Hui J, Stamp K, Drew C, Ren Y, Wang M, Carpenter A et al. (2018) SALL4 mediates teratogenicity as a thalidomide-dependent cereblon substrate. Nat Chem Biol 14:981–987. doi: 10.1038/s41589-018-0129-x

McBride W (1961) Thalidomide and congenital abnormalities. Lancet 2:1358.

McDermott DA, Fong JC and Basson CT (1993) Holt-Oram Syndrome. University of Washington, Seattle

Meng Q and Xia Y (2011) c-Jun, at the crossroad of the signaling network. Protein Cell 2:889–898. doi: 10.1007/s13238-011-1113-3

Miller MT and Strömland K (1999) Teratogen update: thalidomide: a review, with a focus on ocular findings and new potential uses. Teratology 60:306–21. doi: 10.1002/(SICI)1096-9926(199911)60:5<306::AID-TERA11>3.0.CO;2-Y

Ministério da saúde (2021) Monitoramento dos casos de arboviroses urbanas

causados por vírus transmitidos pelo mosquito Aedes (dengue, chikungunya e zika), semanas epidemiológicas 1 a 14, 2021. Bol Epidemiológico 52:1–24.

Moore KL and T.V.N. Persaud (2000) Defeitos Congênitos Humanos. Embriologia Clínica, 6th ed. W.B. Saunders Company, Philadelphia, PA, pp 161–193

Nguyen T Van, Lee JE, Sweredoski MJ, Yang SJ, Jeon SJ, Harrison JS, Yim JH, Lee SG, Handa H, Kuhlman B et al. (2016) Glutamine Triggers Acetylation-Dependent Degradation of Glutamine Synthetase via the Thalidomide Receptor Cereblon. Mol Cell 61:809–820. doi: 10.1016/j.molcel.2016.02.032

Nikkel SM, Dauber A, De Munnik S, Connolly M, Hood RL, Caluseriu O, Hurst J, Kini U, Nowaczyk MJM, Afenjar A et al. (2013) The phenotype of Floating-Harbor syndrome: Clinical characterization of 52 individuals with mutations in exon 34 of SRCAP. Orphanet J Rare Dis 8:63. doi: 10.1186/1750-1172-8-63

Oliveira MA, Antônio J, Bermudez Z, Custódio A and De Souza M (1999) Talidomida no Brasil: Vigilância com responsabilidade compartilhada? Cad Saúde Pública 15:99–112.

Osterwalder M, Speziale D, Shoukry M, Mohan R, Ivanek R, Kohler M, Beisel C, Wen X, Scales SJ, Christoffels VM et al. (2014) HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. Dev Cell 31:345–57. doi: 10.1016/j.devcel.2014.09.018

Parman T, Wiley MJ and Wells PG (1999) Free radical-mediated oxidative DNA damage in the mechanism of thalidomide teratogenicity. Nat Med 5:582–5. doi: 10.1038/8466

Rajkumar SV and Blood E (2006) Lenalidomide and venous thrombosis in multiple myeloma. N Engl J Med 354:2079–80.

Reik W, Dean W and Walter J (2001) Epigenetic reprogramming in mammalian development. Science (80-) 293:1089–1093. doi: 10.1126/science.1063443

Saldanha PH (1994) A tragédia da Talidomida e o advento da teratologia experimental. Rev Bras Genética 17:449–464.

Sales Luiz Vianna F, de Oliveira MZ, Sanseverino MTV, Morelo EF, de Lyra Rabello Neto D, Lopez-Camelo J, Camey SA and Schuler-Faccini L (2015) Pharmacoepidemiology and thalidomide embryopathy surveillance in Brazil. Reprod Toxicol 53:63–67. doi: 10.1016/j.reprotox.2015.03.007

Sales Luiz Vianna F, Kowalski TW, Fraga LR, Sanseverino MTV and Schuler-Faccini L (2016) The impact of thalidomide use in birth defects in Brazil. Eur J Med Genet 60:12–15. doi: 10.1016/j.ejmg.2016.09.015

Sauer H, Günther J, Hescheler J and Wartenberg M (2000) Thalidomide inhibits angiogenesis in embryoid bodies by the generation of hydroxyl radicals. Am J Pathol 156:151–8. doi: 10.1016/S0002-9440(10)64714-1

Scheffner M, Nuber U and Huibregtse JM (1995) Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. Nature 373:81–83. doi: 10.1038/373081a0

Schuler-Faccini L, Soares RCF, de Sousa ACM, Maximino C, Luna E, Schwartz IVD, Waldman C and Castilla EE (2007) New cases of thalidomide embryopathy in Brazil. Birth Defects Res Part A Clin Mol Teratol 79:671–672. doi: 10.1002/bdra.20384

Schwartz MP, Hou Z, Propson NE, Zhang J, Engstrom CJ, Costa VS, Jiang P, Nguyen BK, Bolin JM, Daly W et al. (2015) Human pluripotent stem cell-derived neural constructs for predicting neural toxicity. Proc Natl Acad Sci U S A 112:12516–12521. doi: 10.1073/pnas.1516645112

Shardein J (1993) Psychotropic Drugs. Chem Induc Birth Defects 208–270.

Sheskin J (1965) THALIDOMIDE IN THE TREATMENT OF LEPRA REACTIONS. Clin Pharmacol Ther 6:303–6.

Sievers QL, Petzold G, Bunker RD, Renneville A, Słabicki M, Liddicoat BJ, Abdulrahman W, Mikkelsen T, Ebert BL and Thomä NH (2018) Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. Science (80-) 362:eaat0572. doi: 10.1126/science.aat0572

Smithells RW and Newman CG (1992) Recognition of thalidomide defects. J Med

Genet 29:716–23.

SOMERS GF (1960) PHARMACOLOGICAL PROPERTIES OF THALIDOMIDE (PHTHALIMIDO GLUTARIMIDE), A NEW SEDATIVE HYPNOTIC DRUG. Br J Pharmacol Chemother. doi: 10.1111/j.1476-5381.1960.tb01217.x

Song K, Nam Y-J, Luo X, Qi X, Tan W, Huang GN, Acharya A, Smith CL, Tallquist MD, Neilson EG et al. (2012) Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature 485:599–604. doi: 10.1038/nature11139

Soulier J (2011) Fanconi Anemia. Am Soc Hematol 492–497.

Srivastava D, Cserjesi P and Olson -F-Eric N (1995) A Subclass of bHLH Proteins Required for Cardiac Morphogenesis. Science (80-) 270:1995–1999.

Srivastava D, Thomas T, Lin Q, Kirby ML, Brown D and Olson EN (1997) Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. Nat Genet 16:154–160. doi: 10.1038/ng0697-154

Steimle JD and Moskowitz IP (2017) TBX5: A Key Regulator of Heart Development. Curr Top Dev Biol 122:195–221. doi: 10.1016/bs.ctdb.2016.08.008

Sun YM, Wang J, Qiu XB, Yuan F, Li RG, Xu YJ, Qu XK, Shi HY, Hou XM, Huang RT et al. (2016) A HAND2 loss-of-function mutation causes familial ventricular septal defect and pulmonary stenosis. G3 Genes, Genomes, Genet 6:987–992. doi: 10.1534/g3.115.026518

Tamilarasan KP, Kolluru GK, Rajaram M, Indhumathy M, Saranya R and Chatterjee S (2006) Thalidomide attenuates nitric oxide mediated angiogenesis by blocking migration of endothelial cells. BMC Cell Biol 7:17. doi: 10.1186/1471-2121-7-17

Tamura M, Amano T and Shiroishi T (2014) The Hand2 gene dosage effect in developmental defects and human congenital disorders, 1st ed. Curr Top Dev Biol. doi: 10.1016/B978-0-12-405943-6.00003-8

Therapontos C, Erskine L, Gardner ER, Figg WD and Vargesson N (2009) Thalidomide induces limb defects by preventing angiogenic outgrowth during early

limb formation. PNAS 106:8573–8578.

Titus TA, Yan Y-L, Wilson C, Starks AM, Frohnmayer JD, Canestro C, Rodriguez-Mari A, He X and Postlethwait JH (2009) The Fanconi anemia/BRCA gene network in zebrafish: Embryonic expression and comparative genomics. Mutat Res 668:117–132. doi: 10.1016/j.mrfmmm.2008.11.017

Uhl K, Cox E, Rogan R, Zeldis JB, Hixon D, Furlong LA, Singer S, Holliman T, Beyer J and Woolever W (2006) Thalidomide use in the US: Experience with pregnancy testing in the S.T.E.P.S.® programme. Drug Saf 29:321–329. doi: 10.2165/00002018-200629040-00003

Ujházy E, Mach M, Navarová J, Brucknerová I, Dubovický M and Ujházy AE (2012) Teratology – past, present and future. Interdiscip Toxicol 5:163–168. doi: 10.2478/v10102-012-0027-0

Vargesson N (2015) Thalidomide-induced teratogenesis: History and mechanisms. Birth Defects Res Part C - Embryo Today Rev 105:140–156. doi: 10.1002/bdrc.21096

Vargesson N (2009) Thalidomide-induced limb defects: Resolving a 50-year-old puzzle. BioEssays 31:1327–1336. doi: 10.1002/bies.200900103

Vargesson N and Fraga L (2017) Teratogenesis. eLS. John Wiley & Sons, Ltd, Chichester, UK, pp 1–7

Vega H, Waisfisz Q, Gordillo M, Sakai N, Yanagihara I, Yamada M, Van Gosliga D, Kayserili H, Xu C, Ozono K et al. (2005) Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet 37:468–470. doi: 10.1038/ng1548

Vianna FSL, Fraga LR, Tovo-Rodrigues L, Tagliani-Ribeiro A, Biondi F, Maximino CM, Sanseverino MTV, Hutz MH and Schuler-Faccini L (2013) Polymorphisms in the endothelial nitric oxide synthase gene in thalidomide embryopathy. Nitric Oxide - Biol Chem. doi: 10.1016/j.niox.2013.09.002

Vianna FSL, Kowalski TW, Tovo-Rodrigues L, Tagliani-Ribeiro A, Godoy BA,

Fraga LR, Sanseverino MTV, Hutz MH and Schuler-Faccini L (2016) Genomic and in silico analyses of CRBN gene and thalidomide embryopathy in humans. Reprod Toxicol 66:99–106. doi: 10.1016/j.reprotox.2016.10.003

Vianna FSL, Lopez-Camelo JS, Leite JCL, Sanseverino MTV, Dutra M da G, Castilla EE and Schüler-Faccini L (2011) Epidemiological surveillance of birth defects compatible with thalidomide embryopathy in Brazil. PLoS One 6:e21735. doi: 10.1371/journal.pone.0021735

Voth H, Oberthuer A, Simon T, Kahlert Y, Berthold F and Fischer M (2009) Co-regulated expression of HAND2 and DEIN by a bidirectional promoter with asymmetrical activity in neuroblastoma. BMC Mol Biol 10:28. doi: 10.1186/1471-2199-10-28

Wiborg O, Pedersen MS, Wind A, Berglund LE, Marcker KA and Vuust J (1985) The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. EMBO J 4:755–759. doi: 10.1002/j.1460-2075.1985.tb03693.x

World Health Organization (2020a) Congenital anomalies. https://www.who.int/news-room/fact-sheets/detail/congenital-anomalies. Accessed 6 Dec 2020

World Health Organization (2020b) Distribution of causes of death among children aged < 5 years. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/distribution-of-causes-of-death-among-children-aged-5-years-(-). Accessed 6 Dec 2020

Yabu T, Tomimoto H, Taguchi Y, Yamaoka S, Igarashi Y and Okazaki T (2005) Thalidomide-induced antiangiogenic action is mediated by ceramide through depletion of VEGF receptors, and is antagonized by sphingosine-1-phosphate. Blood 106:125–34. doi: 10.1182/blood-2004-09-3679

Yamagishi H, Olson EN and Srivastava D (2000) The basic helix-loop-helix transcription factor, dHAND, is required for vascular development. J Clin Invest 105:261–70. doi: 10.1172/JCI8856

Yang J, Huang M, Zhou L, He X, Jiang X, Zhang Y and Xu G (2018) Cereblon suppresses the lipopolysaccharide-induced inflammatory response by promoting the ubiquitination and degradation of c-Jun. J Biol Chem 293:10141–10157. doi: 10.1074/jbc.RA118.002246

Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston DM and Eckner R (1998) Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell 93:361–372. doi: 10.1016/S0092-8674(00)81165-4

Zeisberg EM, Ma Q, Juraszek AL, Moses K, Schwartz RJ, Izumo S and Pu WT (2005) Morphogenesis of the right ventricle requires myocardial expression of Gata4. J Clin Invest 115:1522. doi: 10.1172/JCI23769

Zhao Y, Samal E and Srivastava D (2005) Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. doi: 10.1038/nature03817

Zhu YX, Braggio E, Shi C-X, Bruins LA, Schmidt JE, Van Wier S, Chang X-B, Bjorklund CC, Fonseca R, Bergsagel PL et al. (2011) Cereblon expression is required for the antimyeloma activity of lenalidomide and pomalidomide. Blood 118:4771–9. doi: 10.1182/blood-2011-05-356063

ANEXO I

Artigo publicado durante o período do mestrado.

Assembling systems biology, embryo development and teratogenesis: What do we know so far and where to go next?

Reproductive Toxicology 88 (2019) 67–75

Contents lists available at ScienceDirect

Reproductive Toxicology

journal homepage: www.elsevier.com/locate/reprotox

Review

Assembling systems biology, embryo development and teratogenesis: What do we know so far and where to go next?

Thayne Woycinck Kowalski[a,b,c,d,e,*], Ágata de Vargas Dupont[b,c], Bruna Duarte Rengel[a,b,c,e], Eduarda Sgarioni[b], Julia do Amaral Gomes[a,b,c,d,e], Lucas Rosa Fraga[e,f], Lavínia Schuler-Faccini[a,b,d,e], Fernanda Sales Luiz Vianna[a,b,c,d,e,g,*]

[a] Post-Graduation Program in Genetics and Molecular Biology, PPGBM, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[b] Laboratory of Medical Genetics and Evolution, Genetics Department, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[c] Laboratory of Genomic Medicine, Center of Experimental Research, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
[d] National Institute of Medical Population Genetics, INAGEMP, Porto Alegre, Brazil
[e] Sistema Nacional de Informação sobre Agentes Teratogênicos, SIAT, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
[f] Department of Morphological Sciences, Institute of Health Sciences, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
[g] Group of Post-Graduation Research, GPPG, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

ARTICLE INFO

Keywords:
Network
Birth defects
Teratogen STITCH database
Interactome
Congenital anomalies
Malformation
Thalidomide
Zika

ABSTRACT

The recognition of molecular mechanisms of a teratogen can provide insights to understand its embryopathy, and later to plan strategies for the prevention of new exposures. In this context, experimental research is the most invested approach. Despite its relevance, these assays require financial and time investment. Hence, the evaluation of such mechanisms through systems biology rise as an alternative for this conventional methodology. Systems biology is an integrative field that connects experimental and computational analyses, assembling interaction networks between genes, proteins, and even teratogens. It is a valid strategy to generate new hypotheses, that can later be confirmed in experimental assays. Here, we present a literature review of the application of systems biology in embryo development and teratogenesis studies. We provide a glance at the data available in public databases, and evaluate common mechanisms between different teratogens. Finally, we discuss the advantages of using this strategy in future teratogenesis researches.

1. Background

The identification of a human teratogen is a complex and multidisciplinary process [1]. Nevertheless, the identification of the molecular mechanisms of a teratogen can be even more challenging. Basic research is only the first step to the understanding of these molecular mechanisms. However, the benefits after years of research might be translated to (1) the identification of susceptibility biomarkers, i.e. a genetic variant that provides information regarding the risk of exposure; (2) the implantation of prevention strategies, and (3) the pursuit for safe alternatives when in exposure.

The recognition of an embryopathy is extremely relevant in this scenario. It is clear that a phenotypic manifestation is the reflex of the teratogenic properties [2]. From a genetic approach, the impact of teratogens in the embryo is the downstream effect of the deregulation of one or more signaling pathways, which are involved on the embryonic development. However, studies in teratogenesis field can be difficult, since there are limitations that include ethical and financial barriers regarding the use of animal models, and the knowledge that the results encountered cannot be simply extrapolated to humans [3,4]. Hence, the study of the gene ontologies (GO), which are the biological processes in which these genes are involved, have helped in the recognition of the role of these signaling pathways in the embryopathy [5–7]. As an alternative, the rise of bioinformatic studies in this “big data” era can also be helpful in teratogenesis studies.

Systems biology is a branch of bioinformatics, and is based on the concept that the organism must be studied as a whole [8]. This notion is important in the identification of molecular mechanisms of teratogenesis in humans. Hence, here we intend to demonstrate the importance of assembling systems biology networks in embryology and teratology studies. We present a literature review of previous researches that were based upon this approach, and provide the data available in public

* Corresponding authors at: Hospital de Clínicas de Porto Alegre, Laboratory of Genomic Medicine - Center of Experimental Research, Av. Ramiro Barcelos, 2350, Porto Alegre, RS, Brazil.
E-mail addresses: tkowalski@hcpa.edu.br (T.W. Kowalski), fvianna@hcpa.edu.br (F.S.L. Vianna).

https://doi.org/10.1016/j.reprotox.2019.07.015
Received 1 February 2019; Received in revised form 28 June 2019; Accepted 19 July 2019
Available online 27 July 2019
0890-6238/ © 2019 Elsevier Inc. All rights reserved.

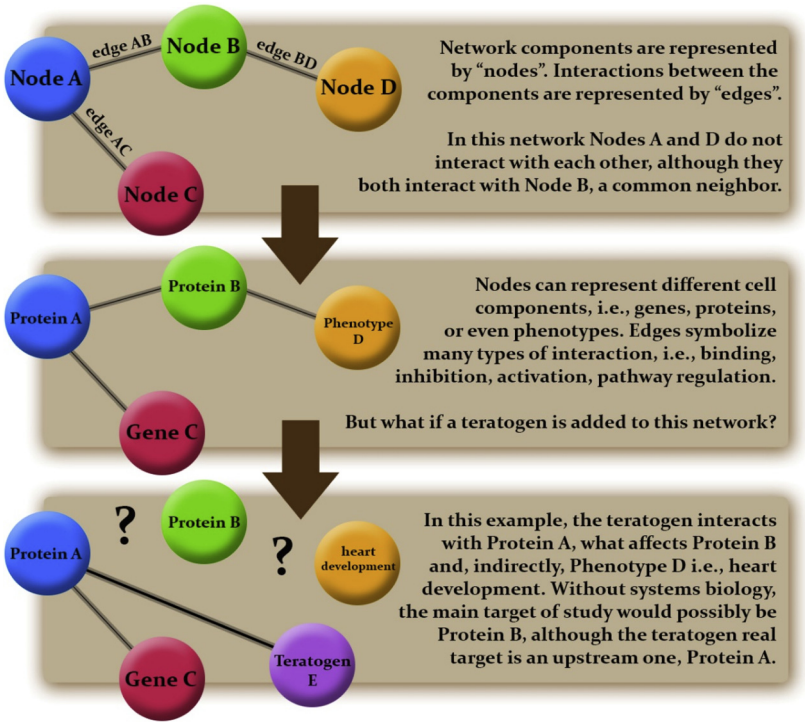


Fig. 1. Scheme representing a network and its main elements.
Legend: The scheme presents a representative network, showing the main graphical demonstrations of the "nodes" and the "edges", and what are the possibilities of a teratogen disrupt or affect this network.

databases. Finally, we discuss the importance of consolidating systems biology researches in teratogenesis.

2. A glance at systems biology

According to the Institute of Systems Biology, this field is based upon the understanding that the whole is bigger than the sum of the parts [9].Systems biology integrates experimental data through computational methodologies to understand how the molecular components of an organism interact in different scales [10], i.e. gene-gene, gene-protein, gene-drug.

Biological networks are the center of the systems biology researches [11,12]. Networks can be mistakenly thought as a simple graphic illustration when they actually area representation of the experimental prior knowledge for that biological process [11,13]. Networks are composed by nodes, which represent the components, and edges, that refer to the interactions between the nodes [14]. Edges can connect nodes of different classes i.e. gene-protein, protein-drug; hence, when studying a teratogenic process, both the teratogen and the proteins/genes deregulated by the exposure agent are considered nodes. The teratogen can affect the network, disrupting or modifying edges. Fig. 1 schematizes a network affected by a teratogen.

Computational modeling of experimental data can generate new hypotheses, by network inference and simulations [15]. In this context, a quantitative (gene expression) or even a logical model (phenotype ontology) can be used in this inference [11]. These models can be used for evaluation of the normal embryo development, and then it is possible to add a factor of interference responsible for a deregulation in this normal process (which will lead to the occurrence of a congenital anomaly). In most cases, researchers aim to identify genes that are mutated and cause congenital anomalies of genetic etiology. Here, the factor of interference to be added will always be a teratogen.

One of the main attempts in the field is the Virtual Embryo Project. It was proposed in 2008, by United States Environmental Protection Agency, as a research program [16]. The project aimed to develop a computational model of the mammalian embryo that could be used to (1) understand the risk of environmental chemicals on development and (2) to predict the impact of a chemical in the embryo by modeling development toxicity pathways and its genetic-cellular networks [16]. It aligns with other screening assays, including systems biology ones, such as ACToR and ToxCast™. The main results of the Virtual Embryo project regarding teratogenic exposure will be further presented. However, this robust, large-scale type of analysis is not always available, especially because of the costs of developing this type of project.

In this context, an alternative that can be used is a friendly online resource. STITCH database integrates data on protein-protein, protein-chemical and chemical-chemical interaction [17,18]. STITCH (http://stitch-db.org) was developed by the European Molecular Biology Laboratory (EMBL) and can be used in pharmacology, toxicology or teratology researches.

3. Methods

Literature research was performed in PubMed (National Library of Medicine, NLM, USA), using the terms "embryo AND development" OR "teratogen*" OR "birth defect" AND "systems" AND "biology". The term "teratogen" was later substituted by the name of the molecule, according to the list provided by Mazzu-Nascimento et al. [19]. Detailed number of publications encountered for each teratogen is available in Supplementary Table 1. Articles were read, and only the ones that included experimental data provided by systems biology were included in this review; experiments only evaluating therapeutic mechanisms or evaluation in adult cells or models were also ignored in this research.

From these data, generation of networks was executed in STRING (http://string-db.org) and STITCH (http://stitch-db.org) databases (EMBL). Networks were assembled using a medium confidence level of 0.400 (default); aiming to obtain only experimental data, textmining interactions were excluded from the analysis. The most enriched ontologies, according to STITCH database analysis, were annotated and

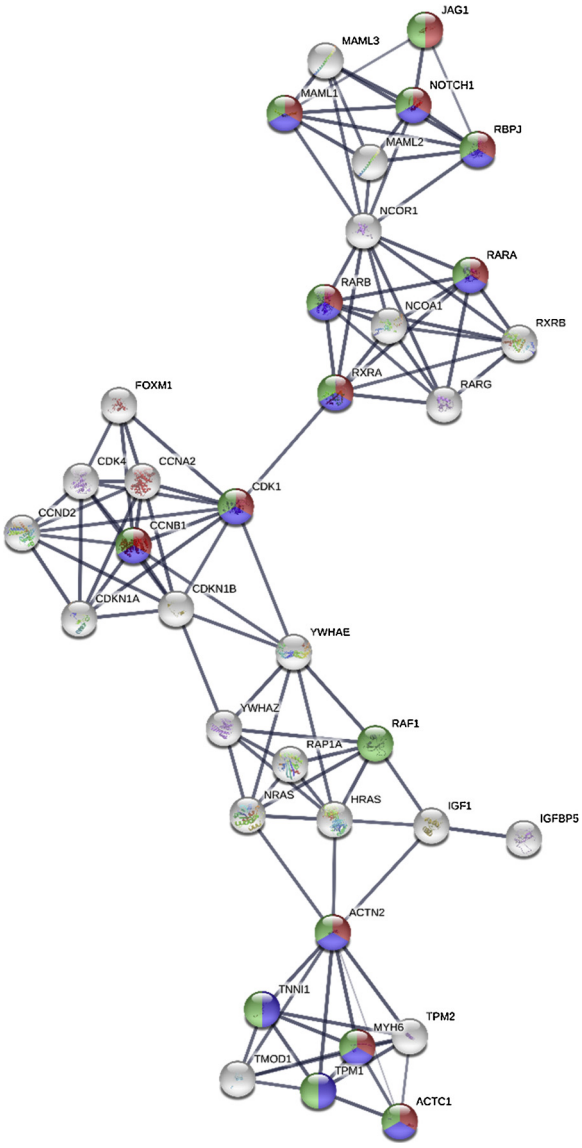


Fig. 2. Interaction network including the main proteins involved in heart development, according to the literature review. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).
Legend: Colors represent Gene Ontologies (GO) for biological processes: "cardiomyocyte differentiation" (red), "cardiac muscle tissue development" (blue) and "heart development" (green). Edges thickness represent the confidence level for the interaction (higher levels of confidence have thicker edges). All data presented was experimentally validated.

evaluated.

Network gathering was performed with DyNet app in Cytoscape v.3.6.1, using default parameters. Common proteins between different teratogens were annotated. The functions of these proteins were evaluated in GeneCards database [20].

All data here presented was generated from public online databases, which are in accordance with the ethical principles of its institutions.

We present below an integrated view of the literature research, comprising the main results for the teratogens that both (1) presented published research in systems biology and (2) could be assembled in networks provided in STITCH database. Furthermore, we present data regarding Zika virus congenital syndrome because of its current relevance in the teratogenesis field. We also present the main findings in the Virtual Embryo Project regarding teratogenic exposure.

4. Networking the embryo development

Systems biology field have studied networks involved in organs and systems development, helping to understand the interactions between developmental genes and proteins. Literature research demonstrated studies published in this field mainly including evaluation on the nervous system, eye development, and heart development [21–27]. Other researchers approached networks of development genes, such as *Sox* genes, without restricting to only one organ or system [28]. Here we present the main findings encountered for embryonic development evaluation through systems biology, with a main focus on cardiac system development, which is, by far, the most studied in this field [24].

Some recent researches have discussed the need to use systems biology to understand eye development [21,22]. According to the authors, there is a great number of genes involved in the process, however it lacks of organized signaling pathways data to understand how these genes interact during the embryonic development [21,22]. The main genes identified in this process are *Bmp4, Bmp7, Hes1, Pax6, Sox1, Sox2, Sox11, Oct1, Prox1, Pitx3,* and *Six3*.

Sox family members, including *Sox5, Sox6,* and *Sox9*, have been recently studied in systems biology [28]. These genes were evaluated in mouse embryos (stage E13.5). A network analysis demonstrated these genes interact with *Tgfb2, Tle3,* and *Fbxl18*, providing the recognition of new *Sox* genes targets during skeleton development and chondrogenesis [28].

Nervous system development has also been approached through systems biology methods. Autism especially has been evaluated, mainly due to its complex nature [23,29]. According to literature review, after network assembling, sixteen biologic processes were suggested to be associated to autism, including neurogenesis and neuron development [23,29]. From this evaluation, *RAC1* must be highlighted as a significant candidate for the understanding of autism [23,29]; this gene might have a role in neuropathologic events, probably because of its neuroimmune interaction with other molecules [23,29].

4.1. Cardiovascular system development

Systems biology has been mainly applied in embryology especially regarding the heart development and the cardiac congenital anomalies [24,25,30,31]. The huge interest in cardiovascular system development might be related to the heart early embryogenesis; the heart is the first organ to be completely formed, and the congenital cardiopathies are one of the most common observed defects [32,33]. Furthermore, experimental research in heart development is challenging; the rat is the main animal model of choice, although a high number of animals is needed, once the number of cardiomyocytes available for studying is scarce [24]. The high cost of *in vivo* studies, added to the ethical dilemma of using an elevated number of animals, justifies the rising interest on systems biology.

A study comprising systems biology and heart development has evaluated genetic risk factors and protein functional convergence [26]. Nineteen processes related to the heart development were analyzed, which included "transcriptional regulation", "BMP signaling pathway", and "focal adhesion". Interaction networks were assembled combining the processes and the risk factors [26]. Some genes were recognized in multiple processes, whilst other had not been previously associated to congenital cardiopathy. The latter comprises *CCND2, FOXM1, IGFBP5, PCSK6, PLCB1, TWIST1,* and *TMOD1* [26]. These results provide new targets to be further studied. Fig. 2 demonstrates how some of these proteins interact in processes such as "cardiomyocyte differentiation" (red), "cardiac muscle tissue development" (blue) and "heart development" (green).

The levels of gene expression in the cardiac tissue are also used in systems biology studies. In mice heart samples, hypertrophy and cardiac insufficiency were associated to developmental genes such as

MRPS5, NPPA, and *MYH7B*, which have been shown to be altered in these conditions [27]. Other approach evaluated the interaction of genes differentially methylated and Fallot tetralogy, a condition that comprises the four major congenital heart malformations [24]. This study led to the identification of *TDGF1, GATA6,* and *PDGFRA* as the main candidates for Fallot tetralogy occurrence [25].

These results demonstrate how systems biology can help in the understanding of embryonic development and in the occurrence of congenital anomalies of genetic etiology. A similar approach can be used in the investigation of congenital anomalies caused by teratogens.

5. The main findings in the virtual embryo project

Within the research program it was built a predictive model to identify chemicals with potential developmental toxicity; the interaction between the chemical and the biological systems is one of the controlling factors of a chemical capacity to disrupt normal embryogenesis [34]. Based on a high-throughput screening (HTS), they were able to link cellular targets and pathways with specific endpoint toxicity through systems biology tools. It was observed that fetal weight reduction (FWR) was associated to cranial malformations through inflammatory response in the rabbit embryos [34]. Prenatal loss (PNL) in rats was associated with FWR through blood vessel development, histone phosphorylation, Ras signaling, migration and apoptosis pathways [34]. They also demonstrated xenobiotic metabolism role in developmental toxicity [34].

Using in vitro HTS and animal studies database, Leung et al. studied developmental toxicity of the male reproductive system [35]. They observed chemicals target nuclear receptors, vascular remodeling proteins, cytochrome P450s and GPCRs (G-protein-coupled receptors), as well as estrogen and androgen receptors [35]. They were also able to associate peroxisome proliferator-activated receptors (PPARs), retinoic acid receptors (RARs) and retinoid X receptors (RXRs) with developmental regulation; and GPCRs with neural signaling, inflammation and angiogenesis [35].

6. Teratogens are network deregulators

According to Wilson (1959), the third principle of teratology sustains "Teratogenic Agents Act in Specific Ways (Mechanisms) on Developing Cells and Tissues to Initiate Sequences of Abnormal Developmental Events (Pathogenesis)" [36,37]. At the time of this definition, the knowledge on teratology was very limited, but soon to be worldwide diffused with Thalidomide Embryopathy recognition in 1961 [38,39]. Despite the growing interest on teratogenic exposure, many difficulties have been encountered in the complete establishment of its molecular mechanisms.

Presently, the most common mechanisms of teratogenic actions are: (1) hyperacetilation of histones; (2) cholesterol imbalance; (3) alterations in folate pathway; (4) retinoic acid imbalance; (5) endocrine disruption; (6) vascular disruption; and (7) induction of oxidative stress [40]. Other specific signaling pathways, however, can be severely affected by teratogenic exposure. It is not completely comprehended whether all these mechanisms are interconnected and how they interfere in the normal embryonic development. Two main approaches have been used to try to answer these inquiries: Adverse Outcome Pathway (AOP) studies and systems biology strategies.

Adverse Outcome Pathway (AOP) is a concept that represents a link between a direct molecular initiating event and a relevant adverse effect [41]. AOPs usually comprise a series of sequential events, in different biological levels of organization [42]; the relationships between the different levels can be obtained from different approaches, such as *in vitro*, *in vivo*, or even using computational resources [41]. However, it must be highlighted AOPs are not systems biology models [43], although it can be affirmed the fields have a similar aim: to link the molecular mechanisms to a final outcome (phenotype).

The main differences between both approaches are (1) AOPs do not intend to be molecularly detailed [43]; and (2) it is a challenge to translate AOPs into a mathematical model [44], whilst systems biology lies on it [11]. Despite their differences, both approaches have been combined recently, generating AOP networks, based on the AOP framework and different scales of interaction [45,46]. This combination could be very useful in teratogenesis in order to provide a better understanding of teratogenic mechanisms.

Regarding our literature research, only “thalidomide”, “retinoic acid”, “alcohol” and “tobacco” presented more than ten published articles evaluating its teratogenesis through systems biology (Supplementary Table 1); however, a network with “tobacco” could not be assembled in STITCH. A summary of the main results for all the other teratogens which presented both a network assembly and literature data regarding teratogenic molecular mechanisms is presented in sequence.

6.1. Thalidomide

Thalidomide was released in the market as a safe, effective drug, without the application of any teratogenic studies [47,48]. More than ten thousand babies were born affected by Thalidomide Embryopathy before its teratogenesis property was discovered [38,39]; this tragedy led to many modifications in the law enforcements applied to new drug’s researches and releases, bringing into focus the importance of teratogenesis tests before any drug is released in the market as safe for pregnant women consumption [49]. Thalidomide Embryopathy is well characterized, causing especially limb defects, although this teratogen can affect almost every organ and system [48,50]. Despite that, many doubts still surround thalidomide, especially regarding its molecular mechanisms of action.

Systems biology approaches, however, represent a good strategy for this teratogen for two main reasons: (1) thalidomide has a primary target of interaction, which is Cereblon protein [51]; (2) thalidomide is a phenocopy of at least three genetic syndromes (Roberts, Holt-Oram, and Okihiro syndromes) that also cause limb defects [2], making it possible to evaluate the teratogen together with different phenotype ontologies.

Our group performed the only study that has evaluated thalidomide in a network interaction context in humans [52]. After a literature review of thalidomide antiangiogenic property, all the genes and proteins that had been shown to be experimentally affected by the drug were included in this network [52]. This analysis provided the identification of *NOS3* and *CTNNB1* genes as the main components of this network [52], bringing new insights to the researches on thalidomide’s molecular mechanisms.

According to STITCH database, thirty-three proteins interact directly with thalidomide (Fig. 3A). These proteins are mainly related to thalidomide’s immunomodulatory and anti-inflammatory properties; it is important to highlight the presence of Cereblon, cytochrome-P450 proteins (which are responsible for its metabolization), and the fibroblast growth factor FGF2 (bFGF).

6.2. Bisphenol A

Exposure to bisphenol A (BPA) at high concentrations during embryonic development results in teratogenic responses including: craniofacial abnormalities and edema and, at lower concentrations, behavioral impairments [53,54].

Saili et al. have investigated global gene expression changes in zebrafish exposed to a dosage of 0.1μM or 80μM of BPA [55]. They demonstrated that, at 0.1 μM of BPA exposure, downregulated genes were of ERK/MAPK and tight junction signaling pathways, both important to maintain nervous system function [55]. In addition, 14/54 biological functions significantly affected were related to nervous system development [55]. The authors also identified several canonical

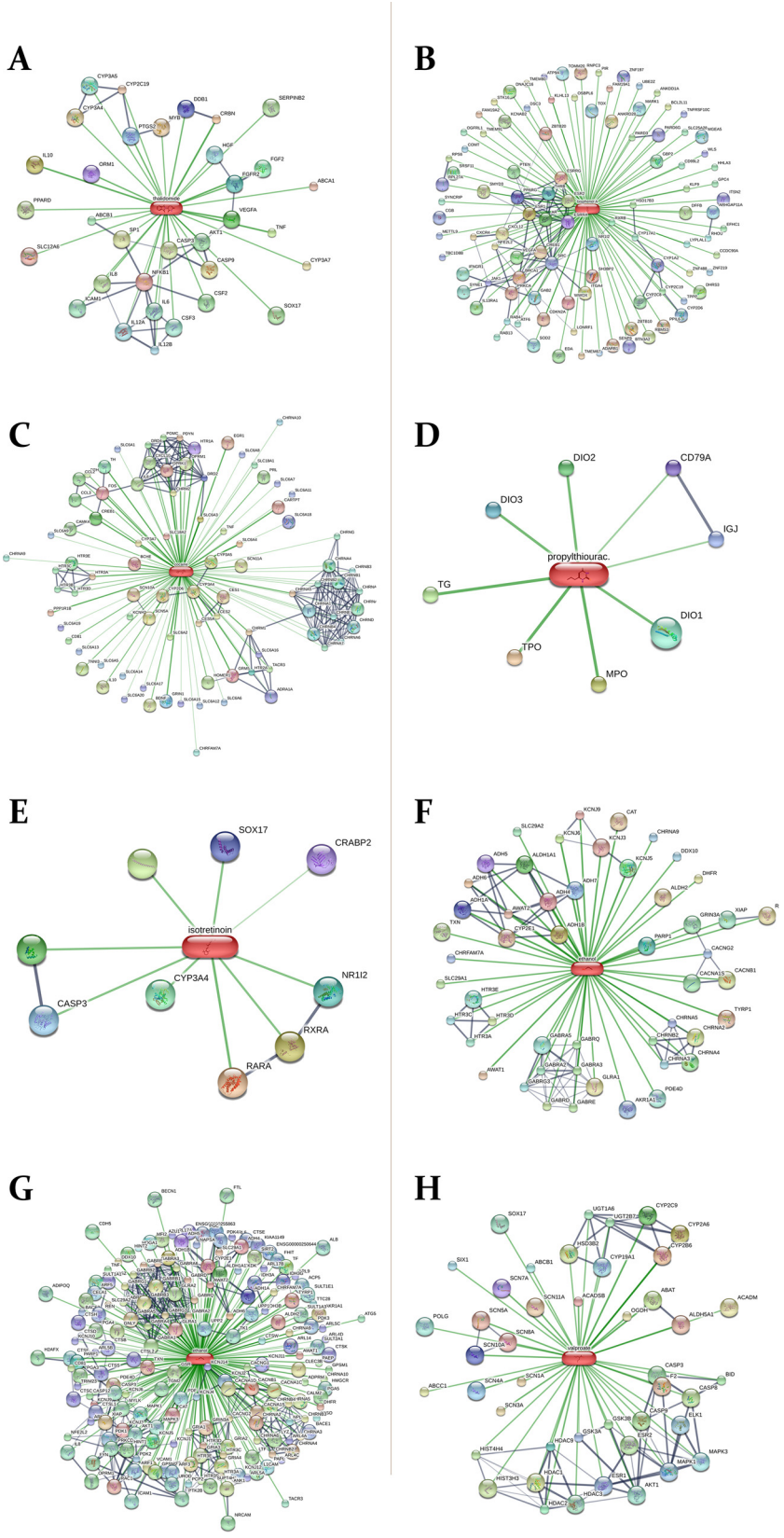


Fig. 3. STITCH networks for (A) thalidomide, (B) bisphenol A, (C) cocaine, (D) propylthiouracil, (E) isotretinoin, (F) ethanol (main interactions), (G) ethanol (complete), and (H) valproate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).
Legend: Teratogens are represented in red. Only protein-teratogen direct interactions are presented in the network. Edges thickness represent the confidence level for the interaction (higher levels of confidence have thicker edges). All data presented was experimentally validated.

pathways and biological functions related to embryonic development impacted by 80 μM of BPA exposure [55].

We have found that seventy-four proteins interact with BPA in STITCH analysis (Fig. 3B). GO analysis demonstrated these proteins are mainly related to "cell response to chemical stimulus" and "organ and system development", what demonstrates this teratogen may affect different organs and systems. Other ontologies demonstrate a significant enrichment in endocrine and reproductive systems impairment, what is in agreement with data regarding BPA teratogenesis.

6.3. Cocaine

Cocaine use during pregnancy causes maternal hypertension, vasoconstriction, and decreased uterine blood flow, leading to impaired nutrient and oxygen delivery for the fetus [56]. In addition, studies indicate maternal exposure to cocaine may significantly increase cell death in the fetal nervous system, what may interfere in brain development [57,58].

Novikova et al. employed a microarray profiling analysis in order to examine differential gene expression, using samples of the cerebral wall of cocaine-exposed mice fetuses [59]. They demonstrated the affected genes represent multiple pathways of apoptosis regulation and more than three quarters of these genes have large proapoptotic activity [59]. Therefore, the study suggests that chronic cocaine exposure may influence in cell survival by affecting a wide range of apoptosis-regulating mechanisms [59].

The network generated in STITCH presented 98 proteins connected to cocaine (Fig. 3C). Ontology analyses demonstrated many of these proteins related with cell signaling, signal transduction, and many neurologic processes, such as brain development. Some proteins were involved in the notorious ontologies such as cholinergic receptors (CHRM2, CHRNA1, CHRNA4, CHRNA7, CHRNB1, CHRND) and the dopamine receptor D2 (DRD2).

6.4. Propylthiouracil (PTU)

Benavides et al. investigated the teratogenic effects of PTU during mice embryogenesis [60]. According to the study, PTU exposure during critical periods of embryogenesis causes neural tube and cardiac abnormalities [60]. PTU altered cytoskeleton remodeling and keratin filaments signaling pathways; thus, potential teratogenic mechanisms may be related to these pathways once they play a role in the organization of the cytoskeleton, in the morphogenesis of the epithelium, and in neuronal survival [60].

Only eight proteins were identified when including PTU in STITCH database (Fig. 3D), hence few ontologies were enriched. However, it is possible to observe these proteins are involved with oxidative stress, hormone biosynthesis and metabolism.

6.5. Retinoic acid

Retinoic acid (RA) is considered one of the most potent teratogens in humans [3]; its embryopathy includes limb, craniofacial, and cardiac anomalies [61].

Berenguer et al. treated mouse embryos with RA and identified new candidate genes and pathways that could be involved in craniofacial development, which could lead to the observed craniofacial anomalies [62]. They also demonstrated that RA exposure may alter some cellular metabolic networks such as glycolysis and gluconeogenesis, as well as the insulin signaling pathway and the hypoxic pathway [62].

Isotretinoin molecule, which the chemical form of RA used in many dermatologic conditions, was available in STITCH for the obtention of PPI networks. Nine proteins were connected to isotretinoin (Fig. 3E), mainly related to epithelial tissue development and involved in the cardiovascular system. Ontologies for "hormonal and inorganic substances response" were enriched.

An AOP network centered around RA was performed by Tonk et al. [63]. The authors collected a list of genes affected by RA perturbation and observed their perturbation across the different levels of organization [63]. The approach driven here, regarding a systems biology network assemble and discussion, provided a set of different genes when comparing to the AOP analysis; i.e. metabolization was represented through Cyp26 family genes in the latter, whilst STITCH pointed to a role from CYP3A4 (Fig. 3E). These differences highlight the concept that both methods are important to understand the molecular mechanisms of teratogenic action, however they can be very contrasting.

6.6. Alcohol

Alcohol use during pregnancy can disturb neurodevelopment and lead to a variety of severe morphological, cognitive and behavioral disabilities known as fetal alcohol spectrum disorders (FASDs) [64,65]. The cognitive and behavioral changes observed are dose-dependent and are related to the period of exposure [65].

Kleiber et al. investigated the timing effects of ethanol administration in mice; neurodevelopmental time of exposure represented the three human gestational trimesters [66]. The study demonstrated that ethanol treatment affects biological activities according to the time of exposure: cell proliferation (first trimester), cell migration, differentiation and morphology (second trimester) and neurotransmission (third trimester) [66]. They also suggested that, although ethanol alters expression of different genes, according to the exposure moment, many of these are functionally related [66].

STITCH provided a PPI network of 190 proteins and ethanol. In order to provide a better visualization, Fig. 3F presents the 50 more centralized proteins. The complete network is available in Fig. 3G. The main ontologies encountered were related to oxidative stress and transmembrane transportation. Neurodevelopment proteins were also enriched.

6.7. Valproic acid

Valproic acid (VPA) is a known teratogen that induces especially neural tube defects and craniofacial anomalies [67]. Mainly used as an anticonvulsant, recently it was discovered that VPA is also a histone deacetylase (HDAC) inhibitor, hence this teratogen is a deregulator of the epigenetic mechanisms [68].

Karén et al. investigated the effects of VPA in pericytes proliferation, viability, migration and differentiation [69]. The study demonstrated HDAC inhibition through VPA affects microvascular pericytes, inhibiting their migration and differentiation [69]. Exposure to VPA also altered the expression of angiogenesis-related genes, what was already expected given the relevance of the pericytes in this process [69].

Valproate molecule was available in STITCH for network construction. The results provided a network with 44 proteins (Fig. 3H). Enriched ontologies mainly associated to molecular functions, such as catalytic activity and transmembrane transportation.

7. Gathering networks (and teratogens)

It is known that many teratogens share similar molecular mechanisms, once biochemical pathways can respond distinctly to different agents [70]. Based on this concept, we questioned whether the networks assembled in the present study, for the teratogens that also had studies evaluating systems biology (alcohol, BPA, cocaine, PTU, RA, thalidomide, and VPA) could be interconnected.

With the exception of PTU, the networks of the other six teratogens could be assembled into one (Fig. 4). Thalidomide and isotretinoin, two of the most known teratogens [50,71], are represented really close; both interact with CYP3A4, a cytochrome P450 protein. Considering only the teratogens here evaluated, alcohol and cocaine share the

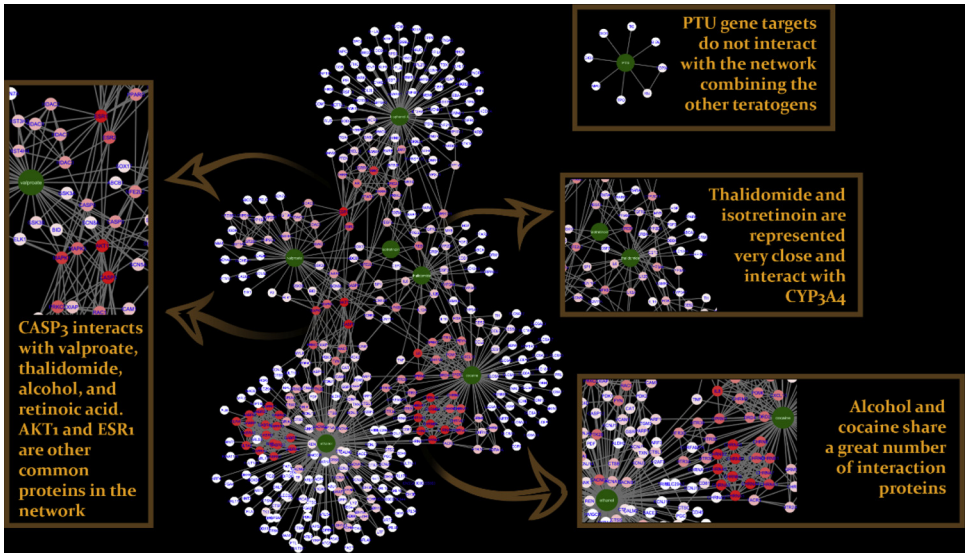


Fig. 4. Assemble of the protein-teratogen networks generated by STITCH. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).
Legend: White nodes: proteins interacting with only one teratogen; Pink and red nodes: proteins interacting with two or more teratogens; Green nodes: teratogens, which are PTU, thalidomide, isotretinoin, alcohol, cocaine, and valproate. Network assembling was generated in DyNet (Cytoscape app) from STITCH tool data. All data presented is experimentally validated.

higher number of targets; these two recreational chemicals have embryopathies with characteristic neurodevelopmental disorders.

Proteins ESR1, AKT1, and CASP3 are interaction targets of different chemicals. These proteins were evaluated in GeneCards database. Estrogen receptor 1 (ESR1) is essential for sexual development, reproductive function, and other tissues such as bones. AKT serine-kinase 1 (AKT1) is an oncogene of the PIK3 pathway; in the developing nervous system, AKT1 acts as a critical mediator of growth-factor induced neuronal survival. Caspase 3 (CASP3) interacts with four teratogens: retinoic acid, thalidomide, alcohol, and valproic acid. CASP3 plays a central role in cell apoptosis, a mechanism that has been previously demonstrated be induced to these chemicals teratogens [72–75]. Previous studies demonstrated the correct balance between pro-apoptotic and cell-survival mechanisms might be determinant to the outcome of a teratogenic exposure, leading to: the occurrence of the embryopathy, embryo/fetal death or normal development [76]. All these analyses demonstrate the systems biology results are according to the biological processes involved in the embryopathies studied.

In a different approach, non-teratogenic drugs were added to the comparative network assembled. Folic acid, azithromycin, liothyronine, levothyroxine and pyridoxine were added to the networks of the teratogens listed above (Supplementary File 1). In this new assemble, PTU was connected to the main network; pyridoxine was the only chemical that did not share interacting proteins with the other drugs evaluated.

Some important concepts must be highlighted when integrating teratogens and non-teratogens in a single network. First, similar target proteins and mechanisms of action are shared by both classes of drugs. This is an expected result once the chemicals act on similar signaling pathways, i.e. ethanol and folic acid act on neurodevelopment, the first disrupting and the latter stimulating it. Second, comparison of systems biology networks must be evaluated together with a strong research hypothesis, combined with centrality metrics for statistical purposes; i.e., PTU, an endocrine disruptor, is connected to the same proteins as levothyroxine and liothyronine, although the hormones network is more centralized. The direction of the protein-drug interactions must also be well characterized to comprehend their molecular mechanisms; however, this data was not available and networks were treated as undirected. Finally, STITCH presents drug-protein interactions not only during the embryonic development. Hence, processes that are more relevant in the adult organism or in the therapeutic response are also well represented in the networks provided by the tool; i.e. thalidomide and azithromycin both act on immune system proteins. Despite thalidomide being a potent immunomodulator, researches indicate this property is more strongly correlated to its therapeutic effect than its teratogenic one.

8. What can we learn from zika virus, a new human teratogen?

STITCH tool provides only protein-protein and protein-drug interactions, hence non-chemical teratogens networks were not available for comparison. The only non-chemical teratogen with data on systems biology was Zika virus (ZIKV). In order to demonstrate systems biology studies can be applied to a wide range of teratogens, not necessarily only the chemical ones. In order to provide a different view of systems biology in teratogenesis, literature review regarding ZIKV teratogenesis and systems biology is presented in sequence.

ZIKV is a very recent teratogen; hence, its molecular mechanisms are not very well comprehended, but they have always been accessed by recent advanced bioinformatics and systems biology tools. This was not possible when the teratogenic potential of i.e., thalidomide or alcohol, was discovered. Research strategies in Congenital Syndrome by Zika (CSZ) usually combine genomics and systems biology in the attempt to understand the neurodevelopmental disorders of the affected children. Regarding the infection, these approaches originate faster and large-scale results. Hence, has been discovered integrating systems biology and CSZ in the last few years.

Moni & Lio used systems biology to better understand the similarities and differences between differentially expressed genes after ZIKV infection, when compared to other infections, i.e., yellow fever, dengue, chikungunya, or the Guillain-Barré syndrome [77]. Comprising seven different datasets of human-cells *in vitro* studies, the authors constructed and analyzed gene networks and ontologies regarding biological processes and signaling pathways; it was observed that ZIKV infection resembles dengue fever regarding the expression of genes involved on the infectious process and in the symptomatology [77].

The identification of new targets for antivirals is also very important in ZIKV researches. This was the aim for the research of Tiwari et al. and they encountered genes associated to changes in the adaptative immune system, angiogenesis and host metabolic processes [78]. In a similar analysis, genes related to histones were identified in clusters [7]. The recognition of an impact on epigenetic mechanisms was also evidenced in a study that identified altered methylation of neural genes; the authors evaluated ZIKV infection in an brain organoid model [79].

CSZ molecular mechanisms are being widely evaluated in the big data era. It is obvious that this approach is a much faster hypothesis generator than the conventional methods used in the past. From interaction networks it is possible to extract data regarding the infection, the individual immune response and teratogenesis itself. With this

approach, it is more likely to elucidate the complex etiology of CSZ in a much shorter time, and also to provide future evaluation on antivirals and vaccines that are still to be developed.

9. Concluding remarks

Here we presented the findings in literature regarding systems biology and teratogenesis. We searched for more than forty teratogens known to affect humans, and observed the number of researches in this area is very scarce. Zika virus was the only non-chemical teratogen with systems biology studies available, an indication of how this approach can be much more explored.

Our network demonstrations already suggested many proteins that can be further explored in order to understand their role in the originated embryopathy. Retinoic acid is a classic teratogen, however only nine proteins are known to interact with the drug. Some of these targets include important proteins for the development of the cardiovascular system; once cardiac defects are observed in retinoic acid embryopathy [71], these proteins could help in the understanding of the cause of these anomalies.

We could also demonstrate many teratogens share common interaction proteins; hence, it is possible to suggest common molecular mechanisms between the chemicals here evaluated. Caspase 3, an important protein to apoptosis, was identified in the networks of four teratogens. Thalidomide, alcohol and valproic acid have demonstrated to act inducing apoptosis [72–75], although further studies are necessary to understand their specific effect on Caspase 3. The role of the protein could be evaluated not only in each embryopathy separately, but also comparing the common effects between those drugs, and even on studies of genetic susceptibility.

Clearly a computational method can never substitute experimental and epidemiological studies. However, here we present a cheap hypothesis generator alternative for teratogenic studies. Much time and effort can be saved if researches are focused in a specific, biologically relevant target, a strategy that can be later tested in experimental models.

Systems biology is an integrative field that aims to evaluate the organism as a whole, complex being, a feature that can be thoroughly studied through the simplest interaction between its components. Withal, teratogenesis refers to the whole impact of a simple molecule in a complex organism. It is feasible to understand how these two fields can be much more integrated in the future, and how much advantage can be taken from this interaction.

Declaration of Competing Interest

The authors declare no conflict of interests.

Acknowledgements

We deeply thank Professor Mariana Recamonde-Mendoza for the scientific support. The authors would like to acknowledge the financial support: INAGEMP (National Institute of Population Medical Genetics; Grant CNPq573993/2008-4, 465549/2014-4, FAPERGS17/2551.0000521-0 and CAPES) and National Council of Scientific and Technologic Development (CNPq); Grant 156158/2018-3, 423249/2016-9, and 312993/2017-0.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.reprotox.2019.07.015.

References

[1] L. Schuler-Faccini, M. Sanseverino, A. Abeche, F. Vianna, A. da Silva, Manual de Teratogênese em Humanos, FEBRASGO, 2011.
[2] T. Cassina, M. Licker, Outcome of geriatric patients admitted to intensive care unit after surgery: the burden of chronic diseases, Minerva Anestesiol. 83 (2017) 1236–1238.
[3] M. De Santis, G. Straface, B. Carducci, A.F. Cavaliere, L. De Santis, A. Lucchese, A.M. Merola, A. Caruso, Risk of drug-induced congenital defects, Eur. J. Obstet. Gynecol. Reprod. Biol. 117 (2004) 10–19.
[4] N. Vargesson, Thalidomide-induced limb defects: resolving a 50-year-old puzzle, Bioessays 31 (2009) 1327–1336.
[5] X. Gao, R.L. Sprando, J.J. Yourick, Transcriptomic changes in mouse embryonic stem cells exposed to thalidomide during spontaneous differentiation, Data Brief 4 (2015) 199–202.
[6] A.J. Rolfe, D.B. Bosco, J. Wang, R.S. Nowakowski, J. Fan, Y. Ren, Bioinformatic analysis reveals the expression of unique transcriptomic signatures in zika virus infected human neural stem cells, Cell Biosci 6 (2016) 42.
[7] R. Brahma, S. Gurumayum, L.D. Naorem, M. Muthaiyan, J. Gopal, A. Venkatesan, Identification of hub genes and pathways in zika virus infection using RNA-seq data: a network-based computational approach, Viral Immunol. 31 (2018) 321–332.
[8] N.D. Price, J.L. Reed, B. Palsson, Genome-scale models of microbial cells: evaluating the consequences of constraints, Nat. Rev. Microbiol. 2 (2004) 886–897.
[9] L. Hood, L. Rowen, D.J. Galas, J.D. Aitchison, Systems biology at the institute for systems biology, Brief Funct. Genomic Proteomic 7 (2008) 239–248.
[10] I. Tavassoly, J. Goldfarb, R. Iyengar, Systems biology primer: the basic methods and approaches, Essays Biochem. 62 (2018) 487–500.
[11] N. Le Novère, Quantitative and logic modelling of molecular and gene networks, Nat. Rev. Genet. 16 (2015) 146–158.
[12] J. Snider, M. Kotlyar, P. Saraon, Z. Yao, I. Jurisica, I. Stagljar, Fundamentals of protein interaction network mapping, Mol. Syst. Biol. 11 (2015) 848.
[13] H. de Jong, Modeling and simulation of genetic regulatory systems: a literature review, J. Comput. Biol. 9 (2002) 67–103.
[14] A.L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[15] Z. Ji, K. Yan, W. Li, H. Hu, X. Zhu, Mathematical and computational modeling in complex biological systems, Biomed Res. Int. 2017 (2017) 5958321.
[16] T.B. Knudsen, P.a.E.B.C.T.C. symposium (Ed.), The Virtual Embryo Project, 2008 Piscataway, NJ, USA.
[17] D. Szklarczyk, A. Santos, C. von Mering, L.J. Jensen, P. Bork, M. Kuhn, STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data, Nucleic Acids Res. 44 (2016) D380–4.
[18] M. Kuhn, C. von Mering, M. Campillos, L.J. Jensen, P. Bork, STITCH: interaction networks of chemicals and proteins, Nucleic Acids Res. 36 (2008) D684–D688.
[19] T. Mazzu-Nascimento, D.G. Melo, G.G. Morbioli, E. Carrilho, F.S.L. Vianna, A.A. Silva, L. Schuler-Faccini, Teratogens: a public health issue - a Brazilian overview, Genet. Mol. Biol. 40 (2017) 387–397.
[20] M. Safran, I. Dalah, J. Alexander, N. Rosen, T. Iny Stein, M. Shmoish, N. Nativ, I. Bahir, T. Doniger, H. Krug, A. Sirota-Madi, T. Olender, Y. Golan, G. Stelzer, A. Harel, D. Lancet, Gene cards version 3: the human gene integrator, Database (Oxford) 2010 (2010) baq020.
[21] S.A. Lachke, R.L. Maas, Building the developmental oculome: systems biology in vertebrate eye development and disease, Wiley Interdiscip. Rev. Syst. Biol. Med. 2 (2010) 305–323.
[22] D. Anand, S.A. Lachke, Systems biology of lens development: a paradigm for disease gene discovery in the eye, Exp. Eye Res. 156 (2017) 22–33.
[23] Y. Zhao, C. Fung, D. Shin, B.C. Shin, S. Thamotharan, R. Sankar, D. Ehninger, A. Silva, S.U. Devaskar, Neuronal glucose transporter isoform 3 deficient mice demonstrate features of autism spectrum disorders, Mol. Psychiatry 15 (2010) 286–299.
[24] S.R. Sperling, Systems biology approaches to heart development and congenital heart disease, Cardiovasc. Res. 91 (2011) 269–278.
[25] M. Grunert, C. Dorn, H. Cui, I. Dunkel, K. Schulz, S. Schoenhals, W. Sun, F. Berger, W. Chen, S.R. Sperling, Comparative DNA methylation and gene expression analysis identifies novel genes for structural congenital heart diseases, Cardiovasc. Res. 112 (2016) 464–477.
[26] K. Lage, S.C. Greenway, J.A. Rosenfeld, H. Wakimoto, J.M. Gorham, A.V. Segrè, A.E. Roberts, L.B. Smoot, W.T. Pu, A.C. Pereira, S.M. Mesquita, N. Tommerup, S. Brunak, B.C. Ballif, L.G. Shaffer, P.K. Donahoe, M.J. Daly, J.G. Seidman, C.E. Seidman, L.A. Larsen, Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 14035–14040.
[27] F.E. Dewey, M.V. Perez, M.T. Wheeler, C. Watt, J. Spin, P. Langfelder, S. Horvath, S. Hannenhalli, T.P. Cappola, E.A. Ashley, Gene coexpression network topology of cardiac development, hypertrophy, and failure, Circ. Cardiovasc. Genet. 4 (2011) 26–35.
[28] W.J. Lee, S. Chatterjee, S.P. Yap, S.L. Lim, X. Xing, P. Kraus, W. Sun, X. Hu, V. Sivakamasundari, H.Y. Chan, P.R. Kolatkar, S. Prabhakar, T. Lufkin, An integrative developmental genomics and systems biology approach to identify an in vivo sox trio-mediated gene regulatory network in murine embryos, Biomed Res. Int. 2017 (2017) 8932583.
[29] F. Zeidán-Chuliá, J.L. Rybarczyk-Filho, A.B. Salmina, B.H. de Oliveira, M. Noda, J.C. Moreira, Exploring the multifactorial nature of autism through computational systems biology: calcium and the Rho GTPase RAC1 under the spotlight, Neuromolecular Med. 15 (2013) 364–383.
[30] A.D. McCulloch, G. Paternostro, Cardiac systems biology, Ann. N. Y. Acad. Sci. 1047 (2005) 283–295.
[31] F. Ferrazzi, R. Bellazzi, F.B. Engel, Gene network analysis: from heart development to cardiac therapy, Thromb. Haemost. 113 (2015) 522–531.

[32] V.K. Khodiyar, D.P. Hill, D. Howe, T.Z. Berardini, S. Tweedie, P.J. Talmud, R. Breckenridge, S. Bhattarcharya, P. Riley, P. Scambler, R.C. Lovering, The representation of heart development in the gene ontology, Dev. Biol. 354 (2011) 9–17.
[33] T.A. Andersen, Ke.L. Troelsen, L.A. Larsen, Of mice and men: molecular genetics of congenital heart disease, Cell. Mol. Life Sci. 71 (2014) 1327–1352.
[34] N.S. Sipes, M.T. Martin, D.M. Reif, N.C. Kleinstreuer, R.S. Judson, A.V. Singh, K.J. Chandler, D.J. Dix, R.J. Kavlock, T.B. Knudsen, Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data, Toxicol. Sci. 124 (2011) 109–127.
[35] M.C. Leung, J. Phuong, N.C. Baker, N.S. Sipes, G.R. Klinefelter, M.T. Martin, K.W. McLaurin, R.W. Setzer, S.P. Darney, R.S. Judson, T.B. Knudsen, Systems toxicology of male reproductive development: profiling 774 chemicals for molecular targets and adverse outcomes, Environ. Health Perspect. 124 (2016) 1050–1061.
[36] J.G. WILSON, Experimental studies on congenital malformations, J. Chronic Dis. 10 (1959) 111–130.
[37] J.G. Wilson, Environment and Birth Defects, Academic Press, 1973.
[38] W. McBride, Thalidomide and congenital abnormalities, Lancet 2 (1961) 1358.
[39] W. LENZ, K. KNAPP, Thalidomide embryopathy, Arch. Environ. Health 5 (1962) 100–105.
[40] E. Giavini, E. Menegola, Biomarkers of teratogenesis: suggestions from animal studies, Reprod. Toxicol. 34 (2012) 180–185.
[41] G.T. Ankley, R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Hornung, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrrano, J.E. Tietge, D.L. Villeneuve, Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment, Environ. Toxicol. Chem. 29 (2010) 730–741.
[42] A. Bal-Price, M.E.B. Meek, Adverse outcome pathways: application to enhance mechanistic understanding of neurotoxicity, Pharmacol. Ther. 179 (2017) 84–95.
[43] M. Leist, A. Ghallab, R. Graepel, R. Marchan, R. Hassan, S.H. Bennekou, A. Limonciel, M. Vinken, S. Schildknecht, T. Waldmann, E. Danen, B. van Ravenzwaay, H. Kamp, I. Gardner, P. Godoy, F.Y. Bois, A. Braeuning, R. Reif, F. Oesch, D. Drasdo, S. Höhme, M. Schwarz, T. Hartung, T. Braunbeck, J. Beltman, H. Vrieling, F. Sanz, A. Forsby, D. Gadaleta, C. Fisher, J. Kelm, D. Fluri, G. Ecker, B. Zdrazil, A. Terron, P. Jennings, B. van der Burg, S. Dooley, A.H. Meijer, E. Willighagen, M. Martens, C. Evelo, E. Mombelli, O. Taboureau, A. Mantovani, B. Hardy, B. Koch, S. Escher, C. van Thriel, C. Cadenas, D. Kroese, B. van de Water, J.G. Hengstler, Adverse outcome pathways: opportunities, limitations and open questions, Arch. Toxicol. 91 (2017) 3477–3505.
[44] C. Obiol-Pardo, J. Gomis-Tena, F. Sanz, J. Saiz, M. Pastor, A multiscale simulation system for the prediction of drug-induced cardiotoxicity, J. Chem. Inf. Model. 51 (2011) 483–492.
[45] D. Knapen, M.M. Angrish, M.C. Fortin, I. Katsiadaki, M. Leonard, L. Margiotta-Casaluci, S. Munn, J.M. O'Brien, N. Pollesch, L.C. Smith, X. Zhang, D.L. Villeneuve, Adverse outcome pathway networks I: development and applications, Environ. Toxicol. Chem. 37 (2018) 1723–1733.
[46] D.L. Villeneuve, M.M. Angrish, M.C. Fortin, I. Katsiadaki, M. Leonard, L. Margiotta-Casaluci, S. Munn, J.M. O'Brien, N.L. Pollesch, L.C. Smith, X. Zhang, D. Knapen, Adverse outcome pathway networks II: network analytics, Environ. Toxicol. Chem. 37 (2018) 1734–1748.
[47] G.F. SOMERS, Pharmacological properties of thalidomide (alpha-phthalimido glutarimide), a new sedative hypnotic drug, Br. J. Pharmacol. Chemother. 15 (1960) 111–116.
[48] W. Lenz, A short history of thalidomide embryopathy, Teratology 38 (1988) 203–215.
[49] A. Daemmrich, A tale of two experts: thalidomide and political engagement in the United States and West Germany, Soc. Hist. Med. 15 (2002) 137–158.
[50] R.W. Smithells, C.G. Newman, Recognition of thalidomide defects, J. Med. Genet. 29 (1992) 716–723.
[51] T. Ito, H. Ando, T. Suzuki, T. Ogura, K. Hotta, Y. Imamura, Y. Yamaguchi, H. Handa, Identification of a primary target of thalidomide teratogenicity, Science 327 (2010) 1345–1350.
[52] T.W. Kowalski, L.R. Fraga, L. Tovo-Rodrigues, M.T.V. Sanseverino, M.H. Hutz, L. Schuler-Faccini, F.S.L. Vianna, Angiogenesis-related genes and thalidomide teratogenesis in humans: an approach on genetic variation and review of past in vitro studies, Reprod. Toxicol. 70 (2017) 133–140.
[53] K.S. Saili, M.M. Corvi, D.N. Weber, A.U. Patel, S.R. Das, J. Przybyla, K.A. Anderson, R.L. Tanguay, Neurodevelopmental low-dose bisphenol a exposure leads to early life-stage hyperactivity and learning deficits in adult zebrafish, Toxicology 291 (2012) 83–92.
[54] F. Perera, J. Vishnevetsky, J.B. Herbstman, A.M. Calafat, W. Xiong, V. Rauh, S. Wang, Prenatal bisphenol a exposure and child behavior in an inner-city cohort, Environ. Health Perspect. 120 (2012) 1190–1194.
[55] K.S. Saili, S.C. Tilton, K.M. Waters, R.L. Tanguay, Global gene expression analysis reveals pathway differences between teratogenic and non-teratogenic exposure concentrations of bisphenol a and 17β-estradiol in embryonic zebrafish, Reprod. Toxicol. 38 (2013) 89–101.
[56] L.M. Smith, L.S. Santos, Prenatal exposure: the effects of prenatal cocaine and methamphetamine exposure on the developing child, Birth Defects Res. C Embryo Today 108 (2016) 142–146.
[57] M.C. Nassogne, J. Louahed, P. Evrard, P.J. Courtoy, Cocaine induces apoptosis in cortical neurons of fetal mice, J. Neurochem. 68 (1997) 2442–2450.
[58] E.S. Mitchell, A. Snyder-Keller, Blockade of D1 dopaminergic transmission alleviates c-fos induction and cleaved caspase-3 expression in the brains of rat pups exposed to prenatal cocaine or perinatal asphyxia, Exp. Neurol. 182 (2003) 64–74.
[59] S.I. Novikova, F. He, J. Bai, I. Badan, I.A. Lidow, M.S. Lidow, Cocaine-induced changes in the expression of apoptosis-related genes in the fetal mouse cerebral wall, Neurotoxicol. Teratol. 27 (2005) 3–14.
[60] V.C. Benavides, M.K. Mallela, C.J. Booth, C.C. Wendler, S.A. Rivkees, Propylthiouracil is teratogenic in murine embryos, PLoS One 7 (2012) e35213.
[61] T.J. Cunningham, G. Duester, Mechanisms of retinoic acid signalling and its roles in organ and limb development, Nat. Rev. Mol. Cell Biol. 16 (2015) 110–123.
[62] M. Berenguer, M. Darnaudery, S. Claverol, M. Bonneu, D. Lacombe, C. Rooryck, Prenatal retinoic acid exposure reveals candidate genes for craniofacial disorders, Sci. Rep. 8 (2018) 17492.
[63] E.C. Tonk, J.L. Pennings, A.H. Piersma, An adverse outcome pathway framework for neural tube and axial defects mediated by modulation of retinoic acid homeostasis, Reprod. Toxicol. 55 (2015) 104–113.
[64] K.L. Jones, D.W. Smith, C.N. Ulleland, P. Streissguth, Pattern of malformation in offspring of chronic alcoholic mothers, Lancet 1 (1973) 1267–1271.
[65] K.L. Jones, H.E. Hoyme, L.K. Robinson, M. Del Campo, M.A. Manning, L.M. Prewitt, C.D. Chambers, Fetal alcohol spectrum disorders: extending the range of structural defects, Am. J. Med. Genet. A 152A (2010) 2731–2735.
[66] M.L. Kleiber, K. Mantha, R.L. Stringer, S.M. Singh, Neurodevelopmental alcohol exposure elicits long-term changes to gene expression that alter distinct molecular pathways dependent on timing of exposure, J. Neurodev. Disord. 5 (2013) 6.
[67] K. Kultima, A.M. Nyström, B. Scholz, A.L. Gustafson, L. Dencker, M. Stigson, Valproic acid teratogenicity: a toxicogenomics approach, Environ. Health Perspect. 112 (2004) 1225–1235.
[68] O.H. Krämer, P. Zhu, H.P. Ostendorff, M. Golebiewski, J. Tiefenbach, M.A. Peters, B. Brill, B. Groner, I. Bach, T. Heinzel, M. Göttlicher, The histone deacetylase inhibitor valproic acid selectively induces proteasomal degradation of HDAC2, EMBO J. 22 (2003) 3411–3420.
[69] J. Karén, A. Rodriguez, T. Friman, L. Dencker, C. Sundberg, B. Scholz, Effects of the histone deacetylase inhibitor valproic acid on human pericytes in vitro, PLoS One 6 (2011) e24954.
[70] M. Cassina, L. Salviati, E. Di Gianantonio, M. Clementi, Genetic susceptibility to teratogens: state of the art, Reprod. Toxicol. 34 (2012) 186–191.
[71] J. Pan, K.M. Baker, Retinoic acid and the heart, Vitam Horm 75 (2007) 257–283.
[72] E. Menegola, M.L. Broccia, F. Di Renzo, E. Giavini, Acetaldehyde in vitro exposure and apoptosis: a possible mechanism of teratogenesis, Alcohol 23 (2001) 35–39.
[73] J.M. Hansen, C. Harris, A novel hypothesis for thalidomide-induced limb teratogenesis: redox misregulation of the NF-kappaB pathway, Antioxid. Redox Signal. 6 (2004) 1–14.
[74] A. Okada, K. Kushima, Y. Aoki, M. Bialer, M. Fujiwara, Identification of early-responsive genes correlated to valproic acid-induced neural tube defects in mice, Birth Defects Res. A Clin. Mol. Teratol. 73 (2005) 229–238.
[75] N.J. Velez-Ruiz, K.J. Meador, Neurodevelopmental effects of fetal antiepileptic drug exposure, Drug Saf. 38 (2015) 271–278.
[76] A. Torchinsky, V. Toder, Mechanisms of the embryo's response to embryopathic stressors: a focus on p53, J. Reprod. Immunol. 85 (2010) 76–80.
[77] M.A. Moni, P. Lio', Genetic profiling and comorbidities of zika infection, J. Infect. Dis. 216 (2017) 703–712.
[78] S.K. Tiwari, J. Dang, Y. Qin, G. Lichinchi, V. Bansal, T.M. Rana, Zika virus infection reprograms global transcription of host cells to allow sustained infection, Emerg. Microb. Infect. 6 (2017) e24.
[79] S. Janssens, M. Schotsaert, R. Karnik, V. Balasubramaniam, M. Dejosez, A. Meissner, A. García-Sastre, T.P. Zwaka, Zika virus alters DNA methylation of neural genes in an organoid model of the developing human brain, mSystems 3 (2018).