



**UNIVERSIDADE FEDERAL DE PERNAMBUCO
LABORATÓRIO DE IMUNOPATOLOGIA KEIZO ASAMI
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA APLICADA À SAÚDE**

KATYANA KALINE SILVA FERREIRA

**ANÁLISE COMPUTACIONAL E GÊNICA DE FATORES DE
TRANSCRIÇÃO DA FAMÍLIA *KRUPPEL-LIKE FACTORS*
ASSOCIADOS A CARDIOPATIAS**

**Recife
2017**

KATYANA KALINE SILVA FERREIRA

**Análise computacional e gênica de fatores de transcrição
da família *Kruppel-like Factors* associados a cardiopatias**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Biologia Aplicada à Saúde, do Laboratório de Imunopatologia Keizo Asami da Universidade Federal de Pernambuco, como parte dos requisitos para obtenção do título de Mestre em Biologia Aplicada à Saúde.

Orientadora:

Prof^a. Dr^a. Danyelly Bruneska Gondim Martins

Departamento de Bioquímica, Centro de Biociências/UFPE;

Laboratório de Imunopatologia Keizo Asami – LIKA/UFPE.

Coorientador:

Prof. Dr. Enéas Ricardo de Moraes Gomes

Departamento de Biotecnologia / UFPB.

**Recife
2017**

Catálogo na Fonte:
Bibliotecário Bruno Márcio Gouveia, CRB-4/1788

Ferreira, Katyana Kaline Silva

Análise computacional e gênica de fatores de transcrição da família *Kruppel-like Factors* associados a cardiopatias / Katyana Kaline Silva Ferreira. – Recife, 2017.

63 f. : il.

Orientadores: Danyelly Bruneska Gondim Martins, Enéas Ricardo de Moraes Gomes

Dissertação (mestrado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Biologia Aplicada à Saúde, Recife, 2017.

Inclui referências e anexos

1. Coração – Doenças 2. Doenças hereditárias I. Martins, Danyelly Bruneska Gondim (orient.) II. Gomes, Enéas Ricardo de Moraes III. Título.

616.12

CDD (22.ed.)

UFPE/CB-2017-515

KATYANA KALINE SILVA FERREIRA

Análise computacional e gênica de fatores de transcrição da família *Kruppel-like Factors* associados a cardiopatias

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Biologia Aplicada à Saúde, do Laboratório de Imunopatologia Keizo Asami da Universidade Federal de Pernambuco, como parte dos requisitos para obtenção do título de Mestre em Biologia Aplicada à Saúde.

Aprovada em: 18 de julho de 2017.

COMISSÃO EXAMINADORA

Prof^a. Dr^a. Danyelly Bruneska Gondim Martins (Orientadora)
Departamento de Bioquímica
Universidade Federal de Pernambuco

Prof. Dr. José Luiz de Figueiredo
Departamento de Cirurgia
Universidade Federal de Pernambuco

Dr^a. Danielly Santos Campos Ferreira
Laboratório de Imunopatologia Keizo Asami – LIKA/UFPE

Aos meus pais, Kátia e Marcos, que sempre me incentivaram, fazendo o possível e impossível para que eu alcançasse novos horizontes e me realizasse profissionalmente.

*“Conheça todas as teorias,
domine todas as técnicas, mas ao
tocar uma alma humana, seja
apenas outra alma humana. ”*

Carl Jung

AGRADECIMENTOS

Primeiramente à Deus, por ter me concebido sabedoria, paciência e calma em todo o caminho percorrido até aqui, nos momentos estressantes, de pressão, de nervosismo...

Aos meus pais, Kátia e Marcos, que me incentivaram a chegar até aqui e me deram apoio em todos os momentos. Meus bens maiores e meu alicerce, sem eles não estaria neste mundo.

À minha Orientadora, Danyelly Bruneka, por ter me orientado nas madrugadas, porque a noite é uma criança! “E trabalho nunca está faltando!”. Por ter me incentivado a sempre buscar o crescimento científico.

Ao meu Coorientador Enéas, que em todos os momentos que precisei de ajuda sempre esteve presente e a disposição, e foi quem me fez abrir os olhos e a mente para o novo e para novos horizontes na pesquisa.

Aos professores do PPGBAS-LIKA, os quais puderam contribuir de forma direta ou indireta para a ampliação do meu conhecimento. Aliás, “O professor é aquele que faz duas ideias crescerem onde antes só havia uma”.

Ao meus amigos e colegas do Laboratório de Prospecção Molecular e Bioinformática - ProspecMol, pela troca de conhecimento, momentos de distração no laboratório e fora do laboratório, sem vocês o lab não seria tão anormal “Prospecmolianos”!

As “Mizerávis” do LIKA que se tornaram amigas para todos os momentos de convívio diário dentro e fora do lab, pelos momentos de distrações compartilhados e pela amizade recíproca criada nesses dois anos, migas amo vocês! Gisnayle, Renata, Fátima e Bárbara.

A minha amiga acadêmica (desde o LBCM – UFPB) e irmã de coração Bruna Braga, por me escutar em todos os momentos de estresses, ansiedade... e sempre ter um: “Calma Katy Katy”. Por todas as conversas produtivas e de descontração, porque não tem como não rir quando ela fala “Aiaiaiai...”.

Aos meus amigos mais que “Xegados”, por terem me feito sentir a presença deles mesmo estando em um Estado do lado, porque viajar no fim de semana entre Pernambuco e Paraíba não seria tão divertido se eu não os tivesse para encontrar e fazer aquela “farra” de sempre, sem contar nas visitas que me fizeram nos carnavais! Amo todos!

A CAPES pelo apoio financeiro.

RESUMO

Os fatores *Kruppel-like* (KLF) são responsáveis pela ativação e repressão de diversos genes envolvidos em processos fisiológicos e patológicos, como desenvolvimento cardíaco, adipogênese e diabetes. A família KLF possui um domínio conservado de ligação ao DNA contendo no qual mutações missense podem afetar a eficiência de ligação e regulação da transcrição. Apesar de alguns fatores de transcrição já serem conhecidos por estarem relacionados a síndrome de Down (SD), o papel das KLFs ainda é desconhecido para o desenvolvimento de defeitos cardíacos congênitos (DCC) na SD. O objetivo do estudo foi realizar análises de bioinformática dos efeitos funcionais e estruturais de mutações missense em KLFs envolvidas no desenvolvimento cardíaco e doenças cardíacas, assim como, avaliar o perfil de expressão gênica da KLF2 e KLF10 em crianças com SD e DCC. Dezesete algoritmos foram utilizados para prever o efeito das mutações sobre a estrutura e função de sete KLFs. Foram selecionadas as mutações Top3 em cada categoria: mutações no domínio dedo de zinco em resíduos conservados e não-conservados. Um total de 850 mutações missense foram encontradas em sete KLFs envolvidas no desenvolvimento cardíaco e doenças cardíacas. Apenas a KLF5 p.Cys410Phe, mutação de resíduo conservado de interação ao zinco, foi predita como deletéria em todos os algoritmos, enquanto apenas a KLF15 p.Arg364Pro afeta a ligação ao DNA. Para a análise de expressão da KLF2 e KLF10, amostras de sangue de 42 crianças de João Pessoa e Recife, com e sem SD e DCC, foram coletadas e analisadas por qPCR. A expressão relativa da KLF2 mostrou uma menor expressão em todos os grupos estudados, enquanto a KLF10 não mostrou expressão em crianças saudáveis, porém, foi expressa em crianças com SD e crianças com SD/DCC, mostrando uma diferença significativa ($p=0.0111$). Nossos resultados reforçam a importância da avaliação computacional através de diferentes algoritmos para determinar as mutações KLF mais importantes a serem estudadas *in vitro* e *in vivo*. Além disso, sugerem a KLF2 e KLF10 como fatores de transcrição que podem estar envolvidos não apenas em DCC, bem como em SD/DCC, destacando a importância do entendimento desses defeitos na SD.

Palavras-chave: *Kruppel-like factors*. Mutações não sinônimas. Algoritmos de predição. KLF2. KLF10. Síndrome de Down. Defeitos cardíacos congênitos.

ABSTRACT

Kruppel-like Factors (KLF) are responsible for activating and repressing many genes involved in physiologic and pathological processes, like cardiac development, adipogenesis, and diabetes. KLF family shows three conserved zinc-fingers, so non-synonymous variants in this domain can affect the binding efficiency and transcription regulation. Although some transcription factors have already been known related to Down syndrome (DS), the role of KLFs in the congenital heart defects (CHD) development in DS has not yet known. The aim of this study was to perform a computational evaluation of the structural and functional effects of non-synonymous variants in KLFs involved in cardiac development and diseases, as well as, to evaluate the expression profile of KLF2 and KLF10 in DS children with CHD. Seventeen algorithms were used to predict the effect of the non-synonymous variants over the structure and function of seven KLFs. The Top3 variants were selected from each category: conserved and non-conserved residues in the zinc-finger domain. A total of 850 non-synonymous variants were found in KLFs involved in the cardiac development and diseases. Only KLF5 p.Cys410Phe, mutation of conserved residue of zinc interaction, was predicted as deleterious in all algorithms, and only KLF15 p.Arg364Pro affects DNA-binding. For analysis of KLF2 and KLF10 expression, blood samples of 42 children from Pernambuco and Paraíba, with and without DS and CHD, were collected and analyzed by qPCR. The relative expression showed KLF2 downregulated in all groups studied, while KLF10 showed no expression in healthy children, however, was expressed in DS and in DS/CHD with a significant difference ($p=0.0111$). Our results reinforce the importance of performing computational evaluations through different algorithms to determine the most important KLF variants to be studied *in vitro* and *in vivo*. In addition, suggest KLF2 and KLF10 as transcription factors involved not only in CHD as well as in DS/CHD, highlighting the importance of understanding these congenital defects in DS.

Keywords: Kruppel-like factors. Non-synonymous mutations. Prediction algorithms. KLF2. KLF10. Down syndrome. Congenital heart defects.

LISTA DE FIGURAS (FUNDAMENTAÇÃO TEÓRICA)

- Figura 1.** Causas de morte no nascimento em todo o mundo. Adaptado de: OMS, 2016..... **19**
- Figura 2.** Defeitos cardíacos congênitos mais comuns na síndrome de Down. **22**
- Figura 3.** Estrutura 3D do domínio dedo de zinco das KLFs ligado ao DNA alvo, mostrando os resíduos conservados de interação ao zinco no DZ. DZ1= dedo de zinco 1; DZ2= dedo de zinco 2; DZ3= dedo de zinco 3; Cis= cisteína; His= histidina e Zn^{2+} = zinco..... **25**
- Figura 4.** Alinhamento da sequência do domínio dedo de zinco dos fatores de transcrição da família KLF. Resíduos destacados representam resíduos conservados de interação ao zinco e de ligação ao DNA. Em verde= cisteínas de interação ao Zn^{2+} ; azul= histidinas de interação ao Zn^{2+} ; roxo= resíduos de ligação ao DNA. **26**
- Figura 5.** Estrutura da proteína dos membros da família KLF em humanos. Domínios regulatórios transcricionais N-terminal de ligação a coativadores, correpresores transcricionais e sítios de ligação a histonas acetiltransferases. Adaptado de: MCCONNELL & YANG, 2010.. **28**
- Figura 6.** KLFs reguladoras chaves da adipogênese. Via das KLFs que levam a diferenciação terminal de pré-adipócitos em adipócitos. Adaptado de: MCCONNELL & YANG, 2010... **29**

LISTA DE FIGURAS (MANUSCRITOS)

MANUSCRITO 1

Figure 1. The pipeline of the KLF variants evaluated through prediction algorithms and 3D modeling by Swiss-model and MolProbity software.....74

Figure 2. Distribution of mutations according to the binary score in the seven KLFs related to cardiac development and diseases. Each dots correspond to one variant in the KLF according to the colors.....75

Figure 3. Alignment of the zinc finger domain sequences of the six KLFs variants more deleterious, showing conserved residues in each ZF. In green are represented Cysteine; in blue Histidine; in purple conserved residues in specific positions that bind to DNA target; and in red are presented the variants. WT: wild-type sequence; M: mutated sequence.....76

Figure 4. 3D models of the KLF wild-type and variants in the zinc finger domain and their zinc atom, showing the loss of interaction to zinc and destabilization of the ZF motif, generated and visualized by Swiss-model and Chimera, respectively. Residues in blue represent Histidine; in green Cysteine and red represent mutated residues. Gray sphere represents zinc atom. (A) Shows the KLF5 p.Cys410Phe and (B) KLF3 p.His314Asp.....77

Figure 5. 3D models and visualization of the KLF15 p.Arg364Pro, showing the protein hydrophobicity surface, evidencing the change from a hydrophilic to the neutral residue, generated and visualized by Swiss-model and Chimera. Red color indicates a hydrophobic surface; white color indicates neutral, and blue color a hydrophilic surface. (A) KLF15 p.Arg364 DNA-binding conserved residue in the -1 position on the ZF2, (B) KLF15 p.Arg364Pro.....78

MANUSCRITO 2

Figure 1. KLF2 and KLF10 mRNA levels and expression profile in DS children without and with CHD. **A)** mRNA levels in DS children without CHD * $p=0.0020$). **B)** mRNA levels in

DS children with CHD. *** $p < 0.0001$. **C)** Relative expression of KLF2 in CHD children with and without DS. **D)** KLF10 mRNA levels in children with and without DS. DS children with CHD non-AVSD present higher levels of KLF10 mRNA. * $p = 0.0111$**92**

LISTA DE TABELAS (MANUSCRITOS)

MANUSCRITO 1

Table 1. Algorithms and deleterious threshold used for computational analysis of KLF. NI= not identified; T: Tolerated; N: Neutral; D: Deleterious; U: Unknown; P: Probably damage; L: Low; M: medium; H: High; B: Benign and RI: Reliability Index.
*Algorithms.....67

Table 2. Reference table of the KLFs reported in the literature to be expressed in the heart tissue, involved in cardiac development and diseases, and adipogenesis process.....69

Table 3. A binary score of the Top3 most deleterious variants of conserved and non-conserved residues in the zinc-finger domain of five KLFs reported to cardiac diseases. 1 was assigned for deleterious prediction, while 0 was assigned for neutral prediction.....71

Table 4. Data of the six variants most deleterious evaluated by 3D modeling. ZF: Zinc Finger72

Table 5. mCSM and MAESTRO prediction of protein stability upon mutation in each KLF.....73

MANUSCRITO 2

Table 1. Type of congenital heart defects found in the patients with and without DS. *patients without DS.....91

LISTA DE ABREVIACÕES E SIGLAS

ASD	– <i>atrial septal defects</i>
AVSD	– <i>atrioventricular septal defect</i>
cDNA	– <i>complementary DNA</i>
C/EBP- α	– <i>CCAAT/Enhancer Binding Protein Alpha</i>
C/EBP- β	– <i>CCAAT/Enhancer Binding Protein Beta</i>
CHD	– <i>congenital heart defects</i>
CtBP	– <i>C-terminal binding Protein</i>
C-terminal	– Carboxi-terminal
DBD	– <i>DNA-binding domain</i>
dbNSFP	– <i>database for Non-Synonymous Functional Prediction</i>
DCC	– Defeitos cardíacos congênitos
DCV	– Doenças cardiovasculares
DLD	– domínio de ligação ao DNA
DNA	– Ácido Desoxirribonucleico
DSA	– Defeito septal atrial
DS	– <i>Down syndrome</i>
DSAV	– Defeito septal atrioventricular
DSV	– Defeitos septais ventricular
DZ	– Dedo de zinco
EUA	– Estados Unidos da América
GATA1	– <i>GATA binding protein 1</i>
GATA4	– <i>GATA binding protein 4</i>
HCM	– <i>Hypertrophic cardiomyopathy</i>
KLF	– <i>Kruppel-like factors</i>
Nkx2	– <i>NK2 Homeobox</i>
Nkx2.5	– <i>NK2 Homeobox 5</i>
N-terminal	– Amino-terminal
OMS	– Organização Mundial de Saúde
RNA	– Ribonucleic acid
Sin3A	– <i>SIN3 - Transcription Regulator Family Member A</i>
SNPs	– <i>Single Nucleotide Polymorphisms</i>
Sp1	– <i>Transcription factor Sp1</i>

PDA	– <i>Patent ductal arteriosus</i>
PPAR- γ	– <i>Peroxisome proliferator-activated receptors gamma</i>
SIM2	– <i>Single-minded 2</i>
TBX5	– <i>T-box transcription factor 5</i>
TF	– <i>Tetralogia de Fallot (TF)</i>
ToF	– <i>Tetralogy of Fallot</i>
VSD	– <i>Ventricular septal defects</i>
ZF	– <i>zinc-finger</i>
Zic3	– <i>Zic family member 3</i>
$\Delta\Delta G$	– <i>Gibbs free energy</i>

SUMÁRIO

1	INTRODUÇÃO	17
2	FUNDAMENTAÇÃO TEÓRICA.....	19
2.1	DOENÇAS CARDÍACAS	19
2.1.1	Defeitos cardíacos congênito	20
2.1.2	Defeitos cardíacos congênitos na síndrome de down.....	21
2.2	FATORES DE TRANSCRIÇÃO DA FAMÍLIA <i>KRUPPEL-LIKE FACTORS</i>	24
2.2.1	Estrutura e conservação das KLFs	25
2.2.2	Via das KLFs na adipogênese e doenças cardíacas.....	28
2.3	BIOINFORMÁTICA E BIOLOGIA COMPUTACIONAL	30
3	OBJETIVOS	34
3.1	OBJETIVO GERAL.....	34
3.2	OBJETIVOS ESPECÍFICOS.....	34
4	REFERÊNCIAS	35
5	MANUSCRITO 1	43
5.	MANUSCRITO 2	80
6.	CONSIDERAÇÕES FINAIS.....	94
ANEXO A - COMITÊ DE ÉTICA EM PESQUISA ENVOLVENDO SERES		
HUMANOS		95

1 INTRODUÇÃO

Em 2015 cerca de 17,7 milhões de pessoas morreram de doenças cardiovasculares, sendo elas as principais causas de morte e um dos maiores problemas de saúde pública em todo o mundo. Os defeitos cardíacos congênitos (DCC) são uma das principais causas de morbimortalidade no período neonatal, com uma estimativa anual de 303.000 recém-nascidos em todo o mundo (“WHO | Cardiovascular diseases (CVDs)”, 2017, “WHO | Congenital anomalies”, 2016). Dessa forma, crianças nascidas com DCC necessitam de um sistema de cuidado especial que facilite o diagnóstico e tratamento desses defeitos cardíacos.

A Síndrome de Down (SD), também conhecida como Trissomia 21, é o transtorno de deficiência intelectual e cromossômica mais comum (PUESCHEL, 1990), com uma frequência de 40-50% dos recém-nascidos portadores da SD apresentando algum tipo de DCC, com o defeito septal atrioventricular (DSAV) sendo o mais comum entre eles (43% dos casos) (ANTONARAKIS, 2016; FREEMAN et al., 2008). Nos EUA, cerca de 500.000 adultos possuem DCC e cerca de uma em cada 100 crianças nascem com DCC sendo causados por fatores genéticos ou anormalidades cromossômicas como a SD (SUN et al., 2015). Nos últimos anos grandes avanços têm sido realizados com o estudo do sequenciamento do genoma de pacientes com DCC, demonstrando genes e mutações relacionadas a esses defeitos (FAHED et al., 2013).

Mais de 50 genes humanos foram identificados associados a DCC ou síndromes genéticas, contudo, um estudo em camundongos mostrou mutações em mais de 500 genes levando ao desenvolvimento de defeitos cardíacos (ANDERSEN; TROELSEN; LARSEN, 2014), indicando que mais genes necessitam ser avaliados em humanos. Diversos fatores de transcrição já têm sido reportados estarem relacionados ao desenvolvimento de cardiopatias e DCC, tais como GATA4, Tbx5, Nkx2-5, Zic3 e alguns fatores da família KLF, assim como, mutações nesses fatores de transcrição também têm sido relacionadas ao desenvolvimento de doenças cardíacas (BOS et al., 2012; TONG, 2016; WARE et al., 2004).

A família de fatores de transcrição *Kruppel-like factors* (KLF) compreende 17 membros KLF, os quais em humanos desempenha um papel crítico na regulação de diversos processos fisiológicos e patológicos, sendo expressas em diversos tecidos e em diferentes

fases do desenvolvimento, ativando ou reprimindo genes alvo dependendo do tecido em que são expressas (NAGAI; FRIEDMAN; KASUGA, 2009). Esses fatores de transcrição são caracterizados por possuírem em sua estrutura um domínio C-terminal de ligação ao DNA conservado, contendo três motivos dedo de zinco (DZ) do tipo C₂H₂. Além disso, também possuem resíduos altamente conservados de ligação ao DNA (KACZYNSKI; COOK; URRUTIA, 2003; MCCONNELL; YANG, 2010).

A análise de mutações e polimorfismos por meio de ferramentas de bioinformática têm sido usadas para a identificação de alterações nas sequências gênicas que possam levar ao desenvolvimento de doenças, assim como, também para favorecer o desenvolvimento de novas ferramentas e estratégias de diagnóstico, e terapêutica de doenças e anomalias. O estudo de mutações não-sinônimas em regiões conservadas têm sido de grande importância para a avaliação da função da proteína, uma vez que elas possuem uma grande probabilidade de levarem a perda de função (HINDORFF et al., 2009; KIEZUN et al., 2013; MACARTHUR et al., 2012). Assim, a descoberta de novos genes e mutações por meio de análises computacionais é de grande importância para a validação experimental de mutações que possam estar associadas a doenças (PIRES; ASCHER; BLUNDELL, 2014), visto que esse tipo de estudo tem se tornado cada vez mais efetivo e de baixo custo, economizando tempo e dinheiro para a realização da validação experimental.

Apesar de genes como GATA1, SIM2 e ZNF295, já terem sido reportados como relacionados a SD (ROBERTS; IZRAELI, 2014; SPELLMAN et al., 2013) e fatores de transcrição da família KLF já demonstrarem uma relação com a regulação gênica de processos que levam ao desenvolvimento cardíaco e de doenças, até o momento nada é conhecido sobre KLFs relacionadas ao desenvolvimento de DCC na SD. Assim, faz-se necessário o estudo *in silico* e *in vitro* desses fatores por meio de abordagens de bioinformática e moleculares para a busca e validação de potenciais fatores que possam estar associados a cardiopatias.

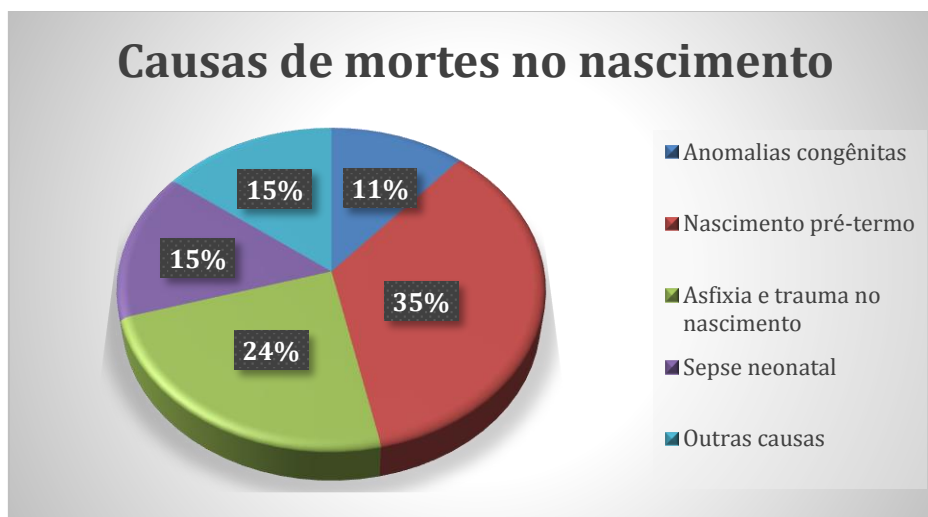
2 FUNDAMENTAÇÃO TEÓRICA

2.1 DOENÇAS CARDÍACAS

As doenças cardiovasculares são as principais causas de morte e um dos maiores problemas de saúde pública em todo o mundo. Segundo a Organização Mundial de Saúde (OMS) cerca de 17,7 milhões de pessoas morreram de doenças cardiovasculares em 2015 (“WHO | Cardiovascular diseases (CVDs)”, 2017). Dentre as DCV, os defeitos cardíacos congênitos (DCC) são uma das principais causas de morbimortalidade no período neonatal (FARIA et al., 2014) e os mais comuns dentre as anomalias congênitas severas, junto com a Síndrome de Down e defeitos do tubo neural (“WHO | Congenital anomalies”, 2016).

Segundo dados da OMS 2016, dentre as 2,68 milhões de mortes durante o período neonatal em 2015, cerca de 303.000 recém-nascidos (correspondente a 11%) morreram dentro de quatro semanas após o nascimento devido ao desenvolvimento de anomalias congênitas, sendo essa a estimativa anual em todo o mundo (Fig. 1) (“WHO | Congenital anomalies”, 2016). Dessa forma, os DCC ainda são um grande problema de saúde pública e necessitam de um sistema de cuidado especial que facilite o diagnóstico e terapêutica dos mesmos.

Figura 1. Causas de morte no nascimento em todo o mundo.



Fonte: Adaptado de: OMS, 2016.

As anomalias congênitas podem ocorrer por diversas causas e fatores de risco, sendo eles: fatores ambientais como exposição materna a agentes químicos, bem como o uso de certos medicamentos, álcool, tabaco e radiação durante a gravidez. Fatores socioeconômicos e

demográficos como a baixa renda podem ser considerados determinantes indiretos devido a uma possível falta de acesso a alimentos nutritivos por mulheres grávidas, ou ao menor acesso aos cuidados com a saúde, e a idade materna avançada, que pode contribuir como um fator de risco para o desenvolvimento fetal anormal aumentando o risco de anormalidades cromossômicas, incluindo a Síndrome de Down (SD). Já os fatores genéticos desempenham um papel importante, uma vez que os genes por hereditariedade podem codificar para uma anomalia ou ser resultado de mudanças no DNA que levam a mutações (“WHO | Congenital anomalies”, 2016).

Mais de 50 genes humanos foram identificados associados a DCC ou síndromes genéticas, as quais também podem apresentar como um dos fenótipos algum tipo de cardiopatia congênita, contudo, um estudo em camundongos mostrou mutações em mais de 500 genes levando ao desenvolvimento de defeitos cardíacos (ANDERSEN; TROELSEN; LARSEN, 2014), indicando que mais genes necessitam ser avaliados em humanos.

2.1.1 Defeitos cardíacos congênitos

Os DCC são definidos como defeitos de malformação estrutural do coração ou dos vasos sanguíneos que se desenvolvem antes do nascimento (BENJAMIN et al., 2017; SUN et al., 2015). Essas malformações podem ser detectadas ainda no útero, no nascimento ou durante os primeiros anos de vida. Na maioria dos casos de DCC, a criança não tem outro tipo de defeito do nascimento, porém, esses defeitos cardíacos podem fazer parte do fenótipo de síndromes genéticas e cromossômicas como a Síndrome de Down, Síndrome de Holt-Oram, Síndrome DiGeorge, Síndrome de Noonan, Síndrome de Turner e entre outras (KO, 2015; WEINRAUCH; ZIEVE; OGILVIE, 2015), com a SD sendo a síndrome cromossômica mais comum apresentando DCC (“Improved National Prevalence Estimates for 18 Selected Major Birth Defects --- United States, 1999--2001”, 2006).

Os DCC variam de simples a complexos classificados em 3 categorias: simples ou média, moderada e severa ou complexa. Os DCC simples são defeitos cardíacos cirurgicamente simples de serem resolvidos ainda na infância sem aparentes complicações. Os DCC moderados são defeitos nos quais são cirurgicamente corrigidos, porém, podem ainda haver algum tipo de problemas residuais. Enquanto que os DCC complexos compreendem qualquer doença cardíaca cianótica, as quais mesmo após intervenção cirúrgica podem causar

consideráveis complicações futuras (HOFFMAN; KAPLAN; LIBERTHSON, 2004). Para o melhor entendimento de como esses defeitos impactam na fisiologia do indivíduo eles também podem ser classificados de acordo com a passagem sanguínea (“Congenital Heart Disease”, [s.d.]) em:

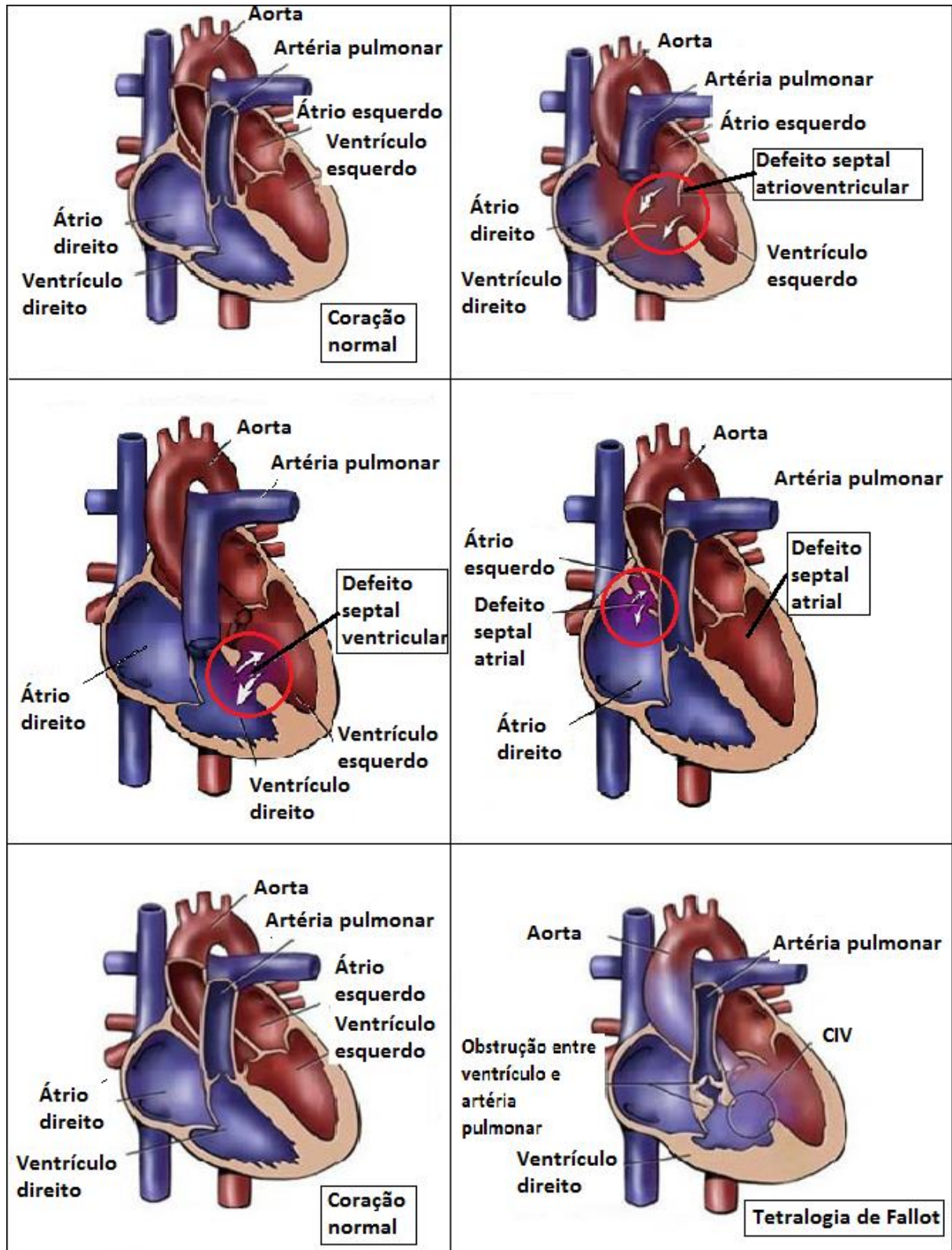
- Defeitos que levam à passagem de muito sangue através dos pulmões, os quais fazem com que o sangue rico em oxigênio que deveria passar pelo corpo retorne para os pulmões causando maior pressão e estresse nos pulmões. Estes defeitos incluem: canal arterial persistente, defeito septal atrial, defeito septal ventricular e canal atrioventricular.
- Defeitos que levam à passagem de pouco sangue através do pulmão, os quais fazem com que o sangue seja pouco oxigenado pelos pulmões (e, portanto, pobre em oxigênio) para seguir através do corpo, e assim, fazendo com que o corpo não receba oxigênio suficiente e o bebê seja cianótico, ou apresente uma coloração azul. Estes defeitos incluem: atresia tricúspide, atresia pulmonar, transposição das grandes artérias, Tetralogia de Fallot, dupla saída do ventrículo direito e Tronco arterioso.
- Defeitos que levam a passagem de pouco sangue através do corpo, os quais resultam em câmaras subdesenvolvidas do coração bloqueando os vasos sanguíneos de transportarem uma quantidade de sangue adequada para o corpo manter suas necessidades. Estes defeitos incluem: Coarctação da aorta, estenose aórtica e síndrome do coração esquerdo hipoplásico.

2.1.2 Defeitos cardíacos congênitos na síndrome de down

A Síndrome de Down é o transtorno de deficiência intelectual, causada por alterações cromossômicas, mais comum com uma frequência de aproximadamente um em cada 700 nascidos vivos em todo o mundo (PUESCHEL, 1990). Cerca de 40-50% dos recém-nascidos portadores da SD possuem algum tipo de defeito cardíaco congênito (STOLL et al., 2015), com o defeito septal atrioventricular (DSAV) sendo o mais comum entre eles (43% dos casos), em seguida, os defeitos septais ventriculares (DSV) (32%), defeito septal atrial (DAS) (19%), e por último com menor frequência, a Tetralogia de Fallot (TF) (6%) (Fig. 2) (ANTONARAKIS, 2016; FREEMAN et al., 2008). No Brasil, apesar de um estudo no Estado Paraná ter apresentado o DSAV como DCC mais comum na SD (CRISTINA et al., 2009), um estudo recente realizado do Sul do Brasil mostrou números diferentes para a prevalência dos

tipos de DCC na SD, onde o tipo mais comum foi o DSA (42,1%), seguidos do DSAV (15,1%), DSA e DSV (14,6%), DSV (12,7%) e dentre as menos comuns, a TF com apenas 2% dos casos (BERMUDEZ et al., 2015). Contudo, os DSAV têm uma prevalência na população de 0,83 em cada 10.000 nascidos vivos (HARTMAN et al., 2011), no entanto, em nascidos com SD o risco é 1.000 vezes maior, corroborando com a ideia de que os DSAV são peculiares na SD (PELLERI et al., 2017).

Figura 2. Defeitos cardíacos congênitos mais comuns na síndrome de Down.



Fonte: Ferreira, K. K. S. (2017).

Nos EUA, cerca de 500.000 adultos possuem DCC e cerca de uma em cada 100 crianças nascem com DCC sendo causados por fatores genéticos ou anormalidades cromossômicas como a SD (SUN et al., 2015). Nos últimos anos de pesquisas genéticas, grandes avanços têm sido realizados com o estudo do sequenciamento do DNA em amostras de sangue de pacientes com DCC. Mutações relacionadas a esses defeitos impactam em um conjunto de moléculas que orquestram o desenvolvimento cardíaco, alteram a dosagem de expressão gênica e causam uma variedade de malformações (FAHED et al., 2013). Assim, a identificação de novos genes e vias metabólicas associadas a doenças que levam ao desenvolvimento de DCC é de grande importância para o entendimento e ampliação do conhecimento acerca dos mecanismos moleculares no desenvolvimento cardíaco e patológico. Esse conhecimento por sua vez leva ao desenvolvimento de novas estratégias de prevenção e terapêuticas (ANDERSEN; TROELSEN; LARSEN, 2014).

Genes que codificam para fatores de transcrição como GATA1, SIM2 e ZNF295, já têm sido reportados estarem relacionados a SD (ROBERTS; IZRAELI, 2014; SPELLMAN et al., 2013). No entanto, apesar de vários estudos apontarem os fatores de transcrição da família KLF como importantes reguladores do desenvolvimento cardíaco, nenhum relato foi reportado até o momento sobre a relação desses fatores com o desenvolvimento de DCC nesta síndrome.

2.2 FATORES DE TRANSCRIÇÃO DA FAMÍLIA *KRUPPEL-LIKE FACTORS*

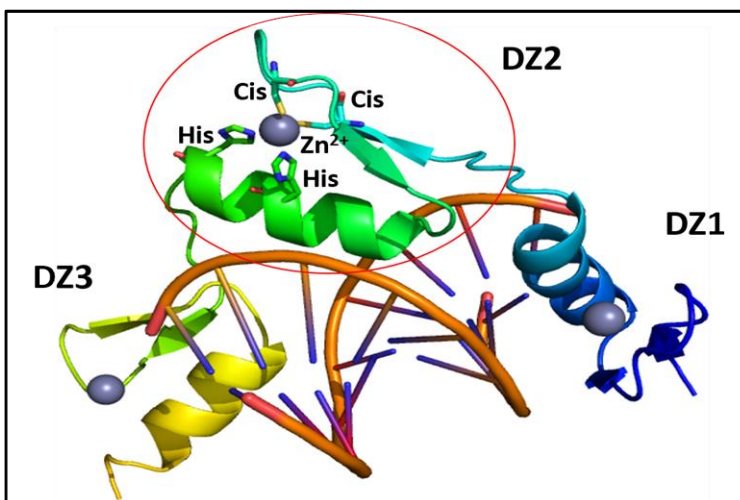
A família de fatores de transcrição *Kruppel-like factors* (KLF) compreende 17 membros KLF denominados de KLF1-KL17 de acordo com sua ordem de descoberta. O gene KLF codifica para a proteína Kruppel na *Drosophila melanogaster*, a qual é um dos fatores de transcrição responsáveis pelo desenvolvimento e segmentação de seu tórax e abdômen (MCCONNELL; YANG, 2010; NAGAI; FRIEDMAN; KASUGA, 2009). Em humanos a família KLF possui papel crítico na regulação de diversos processos fisiológicos e patológicos, como: proliferação e diferenciação celular, adipogênese, doenças metabólicas, inflamação, apoptose, diabetes, oncogênese, doenças cardíacas e entre outros processos, sendo expressas em diversos tecidos e em diferentes fases do desenvolvimento, com cada uma delas atuando na ativação ou repressão de genes alvo dependendo do tecido e do contexto em que são expressas (NAGAI; FRIEDMAN; KASUGA, 2009).

As KLFs possuem homologia com os fatores de transcrição da família Sp1, os quais também possuem em sua estrutura um domínio C-terminal conservado contendo três motivos dedo de zinco (DZ) repetidos e consecutivos do tipo C_2H_2 . Os fatores de transcrição KLF possuem propriedades de ligação ao DNA similares aos fatores Sp1, uma vez que eles também são responsáveis por reconhecer sequências ricas em GC em regiões promotoras (KACZYNSKI; COOK; URRUTIA, 2003; MCCONNELL; YANG, 2010; NAGAI; FRIEDMAN; KASUGA, 2009).

2.2.1 Estrutura e conservação das KLFs

O domínio de ligação ao DNA (DLD) das KLFs compreende três motivos DZ do tipo C_2H_2 que são caracterizados por dois resíduos conservados de Cisteína localizados em duas folhas- β antiparalelas e duas Histidinas conservadas em uma α -hélice ancorando um átomo de zinco (Fig. 3). O DLD é altamente conservado entre os membros da família KLF, apresentando uma homologia de mais de 65% de identidade da sequência C-terminal (KACZYNSKI; COOK; URRUTIA, 2003). Os motivos DZ1, DZ2 e DZ3 das KLFs possuem 25, 25 e 23 amino ácidos, respectivamente, com uma sequência ligadora entre os DZ que compreende cinco aminoácidos altamente conservados em todos os membros KLF (NAGAI; FRIEDMAN; KASUGA, 2009).

Figura 3. Estrutura 3D do domínio dedo de zinco das KLFs ligado ao DNA alvo, mostrando os resíduos conservados de interação ao zinco no DZ. DZ1= dedo de zinco 1; DZ2= dedo de zinco 2; DZ3= dedo de zinco 3; Cis= cisteína; His= histidina e Zn^{2+} = zinco.



Fonte: Ferreira, K. K. S. (2017).

O modo de reconhecimento ao DNA desses DZ foi predito pela análise estrutural de proteínas não KLF (WOLFE; NEKLUDOVA; PABO, 2000), os quais descobriram que cada DZ possui quatro resíduos conservados que se ligam diretamente ao DNA alvo para permitir a transcrição. Esses resíduos conservados ficam localizados na parte inicial da α -hélice em quatro posições específicas denominadas -1, 2, 3 e 6, com o resíduo -1 correspondendo ao resíduo anterior ao primeiro resíduo da α -hélice (NAGAI; FRIEDMAN; KASUGA, 2009; RAZIN et al., 2012).

Estes resíduos específicos se ligam diretamente a três pares de bases do DNA e desempenham um papel crítico no reconhecimento ao DNA alvo, sendo variáveis dentre os membros KLF (NAGAI; FRIEDMAN; KASUGA, 2009; RAZIN et al., 2012; WOLFE; NEKLUDOVA; PABO, 2000). O DZ1 tem como sequência de resíduos conservados KXSHXXA (com X sendo qualquer aminoácido) presente nas KLF1-13 e KLF15, enquanto a sequência KXSHXXS está presente nas KLF14, KLF16 e KLF17, com ambas as sequências se ligando as bases 5'-NGN-3' no DNA. No DZ2 os resíduos específicos são RXDEXXR, sendo completamente conservados em todas as KLFs e reconhecendo como alvo no DNA a sequência 5'-GCG-3'. No DZ3 os resíduos conservados são RXDHXXL, presentes nas KLF1-8 e KLF12, reconhecendo 5'-NGG-3' como alvo. Já nas KLF9-11 e KLF13-16 esses resíduos conservados são RXDHXXK e reconhecem a sequência 5'-(G/T)GG-3', enquanto a sequência RXDHXXQ está presente na KLF17, reconhecendo o alvo 5'-AGG-3' (KACZYNSKI; COOK; URRUTIA, 2003; NAGAI; FRIEDMAN; KASUGA, 2009) (Fig. 4).

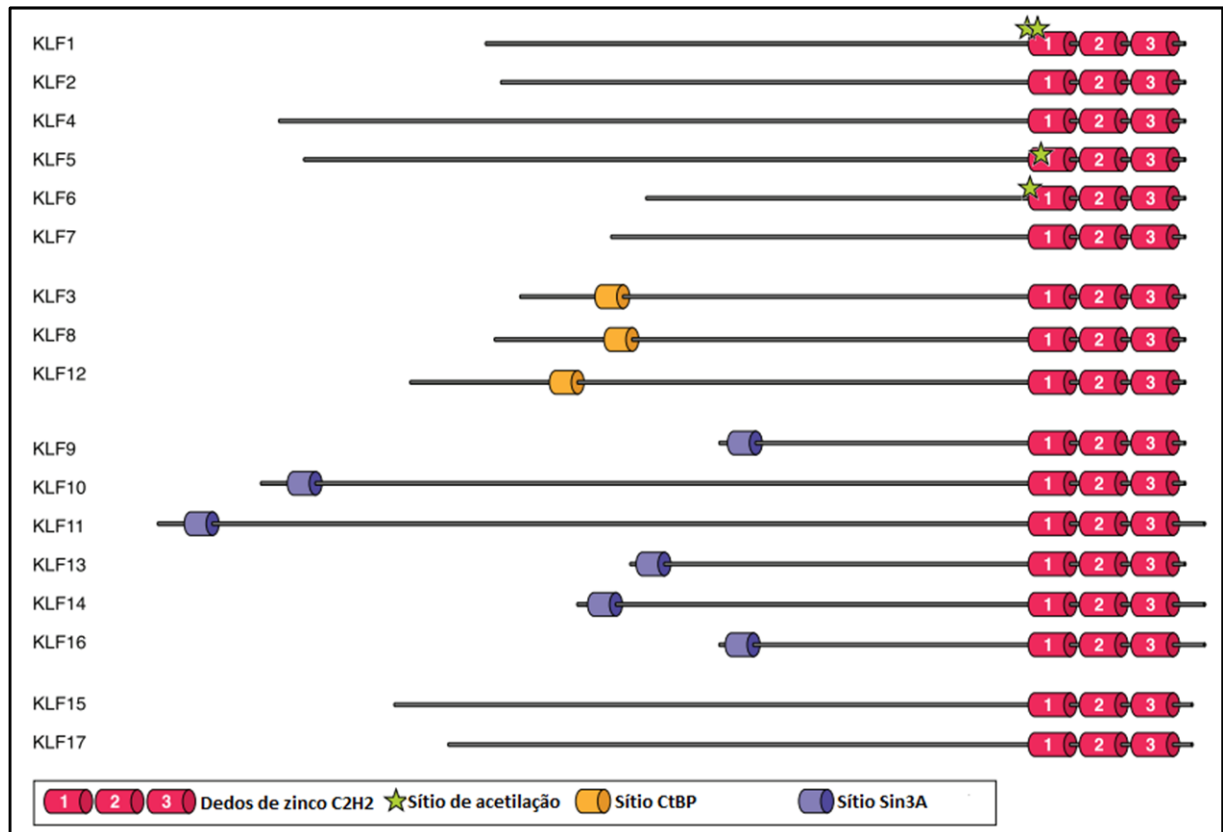
Figura 4. Alinhamento da sequência do domínio dedo de zinco dos fatores de transcrição da família KLF. Resíduos destacados representam resíduos conservados de interação ao zinco e de ligação ao DNA. Em verde= cisteínas de interação ao Zn^{2+} ; azul= histidinas de interação ao Zn^{2+} ; roxo= resíduos de ligação ao DNA.



Fonte: Ferreira, K. K. S. (2017).

A região N-terminal das proteínas KLF é mais variável do que a região C-terminal. Essa região contém domínios de ligação a coativadores ou correpressores transcripcionais específicos que são responsáveis por ativar ou reprimir a atividade transcrricional das KLFs (Fig. 5) (KACZYNSKI; COOK; URRUTIA, 2003; MCCONNELL; YANG, 2010; NAGAI; FRIEDMAN; KASUGA, 2009). Esse domínio regulatório transcrricional N-terminal permite a classificação das KLFs em três grupos. O grupo 1 compreende os repressores transcripcionais KLF3, KLF8 e KLF12, os quais possuem motivos de ligação aos correpressores CtBP (CtBP - *C-terminal binding Protein*), que são proteínas regulatórias que se ligam a sequências específicas de proteínas de ligação ao DNA e ajudam a reprimir a expressão gênica. O grupo 2 consiste predominantemente de ativadores transcripcionais, sendo eles as KLF1, KLF2, KLF4-7, com algumas delas possuindo um sítio de ligação a correpressores que contém atividade acetiltransferase. O grupo 3 compreende os repressores transcripcionais que atuam em conjunto com o correpessor Sin3A (SIN3 - *Transcription Regulator Family Member A*), sendo eles KLF9-11, KLF13, KLF14 e KLF16. Até o momento, KLF15 e KLF17 não foram incluídas nesses grupos uma vez que não apresentam nenhum motivo de interação definido (MCCONNELL; YANG, 2010; NAGAI; FRIEDMAN; KASUGA, 2009).

Figura 5. Estrutura da proteína dos membros da família KLF em humanos. Domínios regulatórios transcripcionais N-terminal de ligação a coativadores, correpressores transcripcionais e sítios de ligação a histonas acetiltransferases.

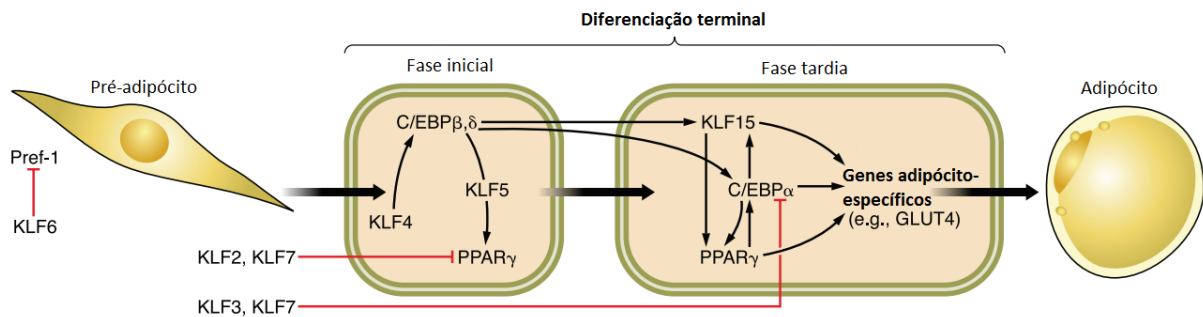


Fonte: adaptado de: MCCONNELL & YANG, 2017.

2.2.2 Via das KLFs na adipogênese e doenças cardíacas

Os fatores de transcrição da família KLF têm sido associados a diversos processos fisiológicos e patológicos, com grande parte da família estando diretamente relacionada ao processo de adipogênese (diferenciação de adipócitos). Oito membros KLF de mamíferos foram identificados como componentes chave do controle transcricional da adipogênese e obesidade, sendo elas as KLFs: KLF2, KLF3, KLF4, KLF5, KLF6, KLF7, KLF11 e KLF15 . (BIRSOY; CHEN; FRIEDMAN, 2008; EISENSTEIN et al., 2014; MCCONNELL; YANG, 2010; MORI et al., 2005; WU; WANG, 2013) (Fig. 6).

Figura 6. KLFs reguladoras chaves da adipogênese. Via das KLFs que levam a diferenciação terminal de pré-adipócitos em adipócitos.



Fonte: Adaptado de: MCCONNELL & YANG, 2010.

A diferenciação desordenada de adipócitos pode levar ao excesso de tecido adiposo e consequentemente à obesidade, condição patológica na qual aumenta o risco do desenvolvimento de doenças cardiovasculares. Um estudo nos EUA demonstrou que 33% de todos os pacientes com defeitos cardíacos congênitos, tanto crianças quanto adultos, apresentam sobrepeso ou são obesos (MOODIE, 2014). Outros estudos demonstraram a relação do desenvolvimento de sobrepeso/obesidade em crianças com defeitos cardíacos congênitos, onde não há diferença quando comparado a crianças saudáveis (SHUSTAK et al., 2012). Também tem sido reportado que outros preditores de risco comuns para a obesidade como gênero e etnia apresentam dados similares entre crianças saudáveis e com cardiopatias congênitas (BARBIERO et al., 2014), reforçando a relação direta da obesidade com cardiopatias congênitas, além das cardiopatias adquiridas (BABAOĞLU et al., 2015). Tais dados sugerem o envolvimento de moléculas reguladoras chaves de processos metabólicos que levam a obesidade, como por exemplo as KLFs, com o desenvolvimento cardíaco e de cardiopatias.

Algumas KLFs estão envolvidas no desenvolvimento cardíaco, com algumas delas levando a interação direta com GATA4, um importante fator de transcrição que atua como regulador do desenvolvimento cardíaco, hipertrofia cardíaca, cardiomiopatia hipertrófica e doenças cardíacas congênitas (HEINEKE; MOKKENTIN, 2006; HUANG et al., 2008; TONG, 2016). A exemplo, as KLF2, KLF4 e KLF15, que participam do processo de adipogênese e no coração interagem de forma direta com GATA4 para regular sua expressão, levando a inibição ou ativação do processo de hipertrofia cardíaca pela interação com GATA4 (CHIPLUNKAR et al., 2013; HALDAR; IBRAHIM; JAIN, 2007; NAGAI; FRIEDMAN;

KASUGA, 2009; XUDONG LIAO, SAPTARSI M. HALDAR, YUAN LU, DARWIN JEYARAJ, KAAVYA PARUCHURI; NAHORI, YINGJIE CUI, KLAUS H. KAESTNER, 2010).

Mutações em genes que codificam para fatores de transcrição já tem sido associadas ao desenvolvimento de doenças cardíacas e defeitos cardíacos congênitos, como por exemplo, nos fatores TBX5, Nkx2.5, GATA4, ZIC3 e alguns fatores da família KLF (BASSON et al., 1997; BERSELL et al., 2009; BOS et al., 2012; DRESSE et al., 2016; KELSEY et al., 2013; MCCONNELL; YANG, 2010; TONG, 2016; WARE et al., 2004). Mutações nas KLF10 e KLF3, por exemplo, levam ao desenvolvimento de cardiomiopatia hipertrófica em camundongos neonatais e hipertrofia cardíaca, respectivamente (BOS et al., 2012; KELSEY et al., 2013). Outros estudos também relacionam essas KLFs com o processo de adipogênese, ou sua expressão durante esse processo, como a KLF10 (BREY et al., 2009; KELSEY et al., 2013; MCCONNELL; YANG, 2010; PROSDOCIMO; SABEH; JAIN, 2015; WU; WANG, 2013). Até o momento pouco tem sido estudado com relação a mutações nesta família de fatores de transcrição e sua associação a doenças cardíacas, o que pode estar totalmente relacionado, uma vez que esses fatores de transcrição são altamente conservados em sua sequência de aminoácidos e domínios estruturais da proteína.

2.3 BIOINFORMÁTICA E BIOLOGIA COMPUTACIONAL

De modo geral, a bioinformática é definida como o estudo da aplicação de técnicas computacionais e matemáticas para a geração e gerenciamento de bioinformação (ARAÚJO et al., 2008). A Bioinformática e a biologia computacional têm suas raízes nas ciências da vida, bem como, nas tecnologias das ciências da informação e computação. Ambas são abordagens interdisciplinares de disciplinas específicas como a matemática, física, biologia, ciência do comportamento, ciência da computação e engenharia. Enquanto a bioinformática aplica princípios das ciências da informação e tecnologias para tornarem dados complexos das ciências da vida em dados mais úteis e entendíveis, a biologia computacional utiliza abordagens computacionais e matemáticas para resolver questões teóricas e experimentais da biologia (NATIONAL INSTITUTE OF HEALTH - BIOINFORMATICS DEFINITION COMMITTEE, 2000).

Devido à grande necessidade de tratar e entender dados biológicos brutos, a bioinformática e biologia computacional têm se consolidado como novas áreas do conhecimento. Programas computacionais para diversos fins biológicos têm desenvolvido, tais como: análise de sequências de genes e proteínas, predição tridimensional de proteínas, alinhamento de sequências biológicas, organização de dados biológicos, construção de árvores filogenéticas, análise de dados de expressão gênica, predição de impacto funcional e estrutural de SNPs (*Single Nucleotide Polymorphisms*) em proteínas, e entre outros (ARAÚJO et al., 2008; CATTLEY; ARTHUR, 2007).

Com o desenvolvimento dos projetos de sequenciamento do genoma humano e de outros organismos, diversas bases de dados biológicos têm sido criadas pela bioinformática para compilar os dados obtidos sobre genes e proteínas. Graças a essas bases de dados, o acesso a essas informações tem se tornado mais fácil e contribuído para a descoberta de novos genes associados a doenças, assim como, a novos alvos terapêuticos e desenvolvimento de medicamentos. dbSNP (*Single Nucleotide Polymorphism Database*) é uma arquivo público gratuito de variantes genética encontradas dentro e entre diferentes espécies, desenvolvida e pelo *National Centre of Biotechnological Information* (NCBI) em colaboração com o *National Human Genome Research Institute* (NHGRI) (SHERRY, 2001). O projeto Ensembl (YATES et al., 2016) é um sistema para anotações do genoma, análise, repositório e disseminação para facilitar o acesso a anotações genômicas de Chordata e modelos de organismos. Uma outra classe de base de dados são as bases de dados de proteínas, como o RCSB PDB (JIN et al., 2015), o qual desenvolve ferramentas e recursos que fornecem uma visão estrutural para da biologia para a pesquisa e educação, além da disponibilização de dados macromoleculares 3D, e a base de dados Uniprot, a qual é uma importante coleção de sequências de proteínas e suas anotações (BATEMAN et al., 2015).

A genômica funcional é o ramo da genômica que utiliza dados produzidos pelas análises genômicas dos projetos de sequenciamento do genoma para descrever a função e interação de genes e proteínas (HILTON; GERSBACH, 2015). Desta forma, a análise de mutações e polimorfismos por meio de abordagens e ferramentas da bioinformática e biologia computacional têm sido cada vez mais utilizadas. Essas análises têm sido usadas para a identificação de alterações nas sequências gênicas que possam levar ao desenvolvimento de doenças, assim como, para favorecer o desenvolvimento de novas ferramentas e estratégias de diagnóstico e terapêutica de doenças e anomalias. Alterações nas sequências nucleotídicas,

que podem ser resultado de mutações, têm sido submetidas a ferramentas estatísticas que predizem a probabilidade dessas alterações levarem ao desenvolvimento de doenças ou não no indivíduo (ARAÚJO et al., 2008).

Os SNPs são as formas mais comuns de variação genética em humanos, podendo estarem relacionados a doenças ou não (COOPER, 1998; MOTTAZ et al., 2010; ONTOLOGY, 2011). SNPs são variações de um único nucleotídeo na sequência de DNA causados por mutações. Eles podem ser classificados em diferentes tipos de acordo com o tipo de mutação, sendo classificados em mutações não sinônimas, ou não sinônimas, quando levam a troca de um aminoácido na sequência da proteína, que por sua vez podem levar a alterações na função da proteína; *nonsense*, ou sinônimas, quando a troca do nucleotídeo não leva a troca do aminoácido; *stop gain*, quando a troca leva a um códon de terminação da proteína, e entre outros tipos de mutações.

Enquanto os efeitos na transcrição ou tradução são atualmente difíceis de prever, vários grupos de pesquisa têm desenvolvido métodos de predição, a exemplo os algoritmos de predição, no intuito de avaliar o impacto de mutações não-sinônimas e a predição de fenótipos (WEI; XU; DUNBRACK, 2013). Diferentes algoritmos têm tomado como base a sequência da proteína para prever o quão deletério as mutações podem ser na estrutura e função da proteína (NG; HENIKOFF, 2003; SHIHAB et al., 2013), tomado como base a sequência ou a estrutura tridimensional (3D) da proteína para prever esses efeitos (CAPRIOTTI et al., 2013; CHENG; RANDALL; BALDI, 2006), ou mesmo dados evolutivos para prever o quanto uma mutação impacta a conservação de proteínas (DONG et al., 2015; NG; HENIKOFF, 2006; REVA; ANTIPIN; SANDER, 2011). A bioinformática tem se tornado cada vez mais específico e, mais recentes, algoritmos têm sido desenvolvidos para prever os efeitos das mutações na estabilidade da proteína, nos permitindo estudar mais a fundo físico-quimicamente os efeitos das mutações (CHENG; RANDALL; BALDI, 2006; LAIMER et al., 2015; PIRES; ASCHER; BLUNDELL, 2014).

Uma outra classe de métodos de predição tem nos permitido construir modelos tridimensionais de proteínas e visualizar mutações em posições específicas, graças a métodos de modelagem por homologia, os quais se baseiam na similaridade da sequência de aminoácidos de proteínas com estruturas conhecidas para construir os modelos (BORDOLI et al., 2009; FISER, 2010; PETTERSEN et al., 2004), e softwares de visualização e análises

desses modelos, como o PyMOL e Chimera softwares (PETTERSEN et al., 2004). Assim, a junção de todos esses métodos de bioinformática nos permite fazer uma melhor análise e seleção de dados, nos dando um melhor suporte para a validação experimental.

O estudo de mutações não-sinônimas em regiões conservadas tem sido importante para a avaliação da função da proteína, uma vez que mutações em regiões conservadas possuem uma grande probabilidade de levarem a perda de função da proteína (HINDORFF et al., 2009; KIEZUN et al., 2013; MACARTHUR et al., 2012). Assim, é de grande interesse o estudo desse tipo de mutação em fatores de transcrição, uma vez que eles são responsáveis pela regulação da transcrição gênica e necessitam de uma alta especificidade para se ligar a regiões promotoras, intensificadoras e de ligação a outras proteínas. Portanto, a busca de mutações não-sinônimas em motivos conservados que levam a perda de função ou especificidade de ligação ao DNA através de análises computacionais tem sido útil, contribuindo para a validação experimental de mutações que possam estar associadas a doenças (PIRES; ASCHER; BLUNDELL, 2014).

3 OBJETIVOS

3.1 OBJETIVO GERAL

Realizar análise computacional dos efeitos estruturais e funcionais de mutações não-sinônimas nos fatores de transcrição da família KLF que possam estar associados a doenças cardíacas, assim como, avaliar o perfil de expressão gênica dessas KLFs em pacientes com cardiopatias congênitas.

3.2 OBJETIVOS ESPECÍFICOS

- Selecionar KLFs que possam estar relacionadas ao desenvolvimento de doenças cardíacas;
- Identificar as mutações não-sinônimas presentes nos fatores de transcrição da família KLF;
- Realizar análise de predição funcional e estrutural através de algoritmos de predição;
- Avaliar o impacto das mutações potencialmente deletérias na estrutura 3D das proteínas KLF localizadas no domínio conservado de ligação ao DNA;
- Avaliar o impacto das mutações na estabilidade da proteína;
- Analisar o perfil de expressão de KLFs expressas em amostra de sangue de pacientes cardiopatas e não cardiopatas;
- Correlacionar os achados laboratoriais com os dados clínicos disponíveis.

4 REFERÊNCIAS

ANDERSEN, T. A.; TROELSEN, K. DE L. L.; LARSEN, L. A. Of mice and men: molecular genetics of congenital heart disease. **Cellular and molecular life sciences**, v. 71, n. 8, p. 1327–1352, 2014.

ANTONARAKIS, S. E. Down syndrome and the complexity of genome dosage imbalance. **Nature Reviews Genetics**, 2016.

ARAÚJO, N. D. DE et al. A Era da Bioinformática : seu potencial e suas implicações para as ciências da saúde. **Estudos de Biologia**, v. 30, p. 143–148, 2008.

BABAOĞLU, K. et al. Prevalence of overweight and obesity among patients with congenital and acquired heart disease in Kocaeli, Turkey. **Cardiology in the Young**, v. 25, n. 3, p. 533–538, 2015.

BARBIERO, S. M. et al. Overweight and obesity in children with congenital heart disease: combination of risks for the future? p. 1–6, 2014.

BASSON, C. T. et al. Mutations in human cause limb and cardiac malformation in Holt-Oram syndrome. **Nature genetics**, v. 15, n. 1, p. 30–35, 1997.

BATEMAN, A. et al. UniProt: A hub for protein information. **Nucleic Acids Research**, v. 43, n. D1, p. D204–D212, 2015.

BENJAMIN, E. J. et al. **Heart Disease and Stroke Statistics???2017 Update: A Report From the American Heart Association**. [s.l: s.n.].

BERMUDEZ, B. E. B. V. et al. Down syndrome: Prevalence and distribution of congenital heart disease in Brazil. **Sao Paulo Medical Journal**, v. 133, n. 6, p. 521–524, 2015.

BERSELL, K. et al. Neuregulin1/ErbB4 Signaling Induces Cardiomyocyte Proliferation

and Repair of Heart Injury. **Cell**, v. 138, n. 2, p. 257–270, 2009.

BIRSOY, K.; CHEN, Z.; FRIEDMAN, J. Transcriptional Regulation of Adipogenesis by KLF4. **Cell Metabolism**, v. 7, n. 4, p. 339–347, 2008.

BORDOLI, L. et al. Protein structure homology modeling using SWISS-MODEL workspace. **Nature protocols**, v. 4, n. 1, p. 1–13, 2009.

BOS, J. M. et al. TGFB-inducible early gene-1 (TIEG1) mutations in hypertrophic cardiomyopathy. **Journal of Cellular Biochemistry**, v. 113, n. 6, p. 1896–1903, 2012.

BREY, C. W. et al. Krüppel-like family of transcription factors: An emerging new frontier in fat biology. **International Journal of Biological Sciences**, v. 5, n. 6, p. 622–636, 2009.

CAPRIOTTI, E. et al. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. **BMC genomics**, v. 14, n. Suppl 3, 2013.

CATTLEY, S.; ARTHUR, J. W. BioManager: The use of a bioinformatics web application as a teaching tool in undergraduate bioinformatics training. **Briefings in Bioinformatics**, v. 8, n. 6, p. 457–465, 2007.

CHENG, J.; RANDALL, A.; BALDI, P. Prediction of protein stability changes for single-site mutations using support vector machines. **Proteins**, v. 62, n. 4, p. 1125–32, 2006.

CHIPLUNKAR, A. R. et al. Kruppel-Like Factor 2 Is Required for Normal Mouse Cardiac Development. **PLoS ONE**, v. 8, n. 2, 2013.

Congenital Heart Disease. Disponível em:
<<http://www.stanfordchildrens.org/en/topic/default?id=congenital-heart-disease-90-P02346>>. Acesso em: 3 jul. 2017.

COOPER, D. The human gene mutation database. **Nucleic Acids Research**, v. 26, n. 1, p. 285–287, 1998.

CRISTINA, T. et al. Características e Prevalência de Cardiopatias Congênitas em Crianças com Síndrome de Down Submetidas à Cirurgia Cardíaca em um Hospital na Região Norte do Paraná Characteristics and Prevalence of Congenital Heart Disease in Children with Down Syndrome Und. p. 11–16, 2009.

DONG, C. et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. **Human Molecular Genetics**, v. 24, n. 8, p. 2125–2137, 2015.

DRESSE, M. et al. A novel de novo TBX5 mutation in a patient with Holt-Oram syndrome leading to a dramatically reduced biological function. **Molecular genetics & genomic medicine**, v. 4, n. 5, p. 557–67, 2016.

EISENSTEIN, A. et al. An adenosine receptor-Kruppel-like factor 4 protein axis inhibits adipogenesis. **The Journal of Biological Chemistry**, v. 289, n. 30, p. 21071–21081, 2014.

FAHED, A. C. et al. Genetics of congenital heart disease: The glass half empty. **Circulation Research**, v. 112, n. 4, p. 707–720, 2013.

FARIA, P. F. et al. Associação entre cardiopatias congênitas e infecções graves em crianças com síndrome de Down. **Revista Portuguesa de Cardiologia**, v. 33, n. 1, p. 15–18, 2014.

FISER, A. Computational Biology. v. 673, p. 1–20, 2010.

FREEMAN, S. B. et al. Ethnicity, sex, and the incidence of congenital heart defects: a report from the National Down Syndrome Project. **Genetics in Medicine**, v. 10, n. 3, p. 173–180, 2008.

HALDAR, S. M.; IBRAHIM, O. A.; JAIN, M. K. Kruppel-like Factors (KLFs) in

muscle biology. **Journal of Molecular and Cellular Cardiology**, v. 43, p. 1–10, 2007.

HARTMAN, R. J. et al. The Contribution of Chromosomal Abnormalities to Congenital Heart Defects : A Population-Based Study. **Pediatric Cardiology**, v. 32, p. 1147–1157, 2011.

HEINEKE, J.; MOLKENTIN, J. D. Regulation of cardiac hypertrophy by intracellular signalling pathways. **Nature Reviews Molecular Cell Biology**, v. 7, n. 8, p. 589–600, 2006.

HILTON, I. B.; GERSBACH, C. A. Enabling functional genomics with genome engineering Genome engineering technologies. **Genome Research**, n. 25, p. 1442–1455, 2015.

HINDORFF, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. **Proceedings of the National Academy of Sciences**, v. 106, n. 23, p. 9362–9367, 2009.

HOFFMAN, J. I. E.; KAPLAN, S.; LIBERTHSON, R. R. Prevalence of congenital heart disease. **American Heart Journal**, v. 147, n. 3, p. 425–439, 2004.

HUANG, Y. et al. GATA4 is a survival factor in adult cardiac myocytes but is not required for alpha1A-adrenergic receptor survival signaling. **American journal of physiology. Heart and circulatory physiology**, v. 295, n. 2, p. H699-707, 2008.

Improved National Prevalence Estimates for 18 Selected Major Birth Defects --- United States, 1999--2001. Disponível em: <<https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5451a2.htm>>. Acesso em: 20 jun. 2017.

JIN, X. et al. PDB-Explorer: a web-based interactive map of the protein data bank in shape space. **BMC Bioinformatics**, v. 16, n. 1, p. 339, 2015.

KACZYNSKI, J.; COOK, T.; URRUTIA, R. Protein family review Sp1- and Krüppel-

like transcription factors. p. 1–8, 2003.

KELSEY, L. et al. ENU-induced Mutation in the DNA-binding Domain of KLF3 Reveals Important Roles for KLF3 in Cardiovascular Development and Function in Mice. **PLoS Genetics**, v. 9, n. 7, 2013.

KIEZUN, A. et al. Exome sequencing and the genetic basis of complex traits. **NIH Public Access**, v. 44, n. 6, p. 623–630, 2013.

KO, J. M. Genetic syndromes associated with congenital heart disease. **Korean Circulation Journal**, v. 45, n. 5, p. 357–361, 2015.

LAIMER, J. et al. MAESTRO - multi agent stability prediction upon point mutations. **BMC bioinformatics**, v. 16, n. 116, 2015.

MACARTHUR, D. G. et al. Europe PMC Funders Group Europe PMC Funders Author Manuscripts A systematic survey of loss-of-function variants in human protein-coding genes. v. 335, n. 6070, p. 823–828, 2012.

MCCONNELL, B. B.; YANG, V. W. Mammalian Krueppel-Like Factors in Health and Diseases. **Physiological Reviews**, v. 90, n. 4, p. 1337–1381, 2010.

MOODIE, D. Obesity Is Everywhere — Even in Patients with Congenital. **Congenit Heart Dis.**, v. 9, p. 367, 2014.

MORI, T. et al. Role of kruppel-like factor 15 (KLF15) in transcriptional regulation of adipogenesis. **Journal of Biological Chemistry**, v. 280, n. 13, p. 12867–12875, 2005.

MOTTAZ, A. et al. Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. **Bioinformatics**, v. 26, n. 6, p. 851–852, 2010.

NAGAI, R.; FRIEDMAN, S. L.; KASUGA, M. **The Biology of Kruppel-like Factors**. 1. ed. [s.l: s.n.]. v. 1.

NATIONAL INSTITUTE OF HEALTH - BIOINFORMATICS DEFINITION COMMITTEE. Nih Working Definition of Bioinformatics and Computational Biology. p. 1, 2000.

NG, P. C.; HENIKOFF, S. SIFT: Predicting amino acid changes that affect protein function. **Nucleic Acids Research**, v. 31, n. 13, p. 3812–3814, 2003.

NG, P. C.; HENIKOFF, S. Predicting the Effects of Amino Acid Substitutions on Protein Function. **Annual Review of Genomics and Human Genetics**, v. 7, n. 1, p. 61–80, 2006.

ONTOLOGY, G. DATA SET: DISEASE-RELATED MUTATIONS As a complementary experiment , we assess mCSM performance in predicting disease-related mutations as a classification task . This Classification tasks The following tables summarize the mCSM performance on classificat. p. 1–13, 2011.

PELLERI, M. C. et al. **Genotype-phenotype correlation for congenital heart disease in Down syndrome through analysis of partial trisomy 21 cases**. [s.l.] Elsevier Inc, 2017.

PETTERSEN, E. F. et al. UCSF Chimera — A Visualization System for Exploratory Research and Analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, 2004.

PIRES, D. E. V; ASCHER, D. B.; BLUNDELL, T. L. mCSM: Predicting the effects of mutations in proteins using graph-based signatures. **Bioinformatics**, v. 30, n. 3, p. 335–342, 2014.

PROSDOCIMO, D. A.; SABEH, M. K.; JAIN, M. K. Kruppel-like factors in muscle health and disease. **Trends in Cardiovascular Medicine**, v. 25, n. 4, p. 278–287, 2015.

PUESCHEL, S. M. Clinical Aspects of Down-Syndrome From Infancy To Adulthood. **American Journal of Medical Genetics**, v. 56, n. 7, p. 52–56, 1990.

RAZIN, S. V. et al. Cys2His2 zinc finger protein family: Classification, functions, and major members. **Biochemistry (Moscow)**, v. 77, n. 3, p. 277–288, 2012.

REVA, B.; ANTIPIN, Y.; SANDER, C. Predicting the functional impact of protein mutations: Application to cancer genomics. **Nucleic Acids Research**, v. 39, n. 17, p. 1–14, 2011.

ROBERTS, I.; IZRAELI, S. Haematopoietic development and leukaemia in Down syndrome. **British Journal of Haematology**, v. 167, n. 5, p. 587–599, 2014.

SHERRY, S. T. dbSNP: the NCBI database of genetic variation. **Nucleic Acids Research**, v. 29, n. 1, p. 308–311, 2001.

SHIHAB, H. A. et al. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. **Human Mutation**, v. 34, n. 1, p. 57–65, 2013.

SHUSTAK, R. J. et al. Prevalence of obesity among patients with congenital and acquired heart disease. **Pediatric Cardiology**, v. 33, n. 1, p. 8–14, 2012.

SPELLMAN, C. et al. Expression of trisomic proteins in Down syndrome model systems. **Gene**, v. 512, n. 2, p. 219–225, 2013.

STOLL, C. et al. Associated congenital anomalies among cases with Down syndrome. **European Journal of Medical Genetics**, v. 58, n. 12, p. 674–680, 2015.

SUN, R. et al. Congenital Heart Disease: Causes, Diagnosis, Symptoms, and Treatments. **Cell Biochemistry and Biophysics**, v. 72, n. 3, p. 857–860, 2015.

TONG, Y.-F. Mutations of NKX2.5 and GATA4 genes in the development of congenital heart disease. **Gene**, v. 588, n. 1, p. 86–94, 2016.

WARE, S. M. et al. Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. **American journal of human genetics**,

v. 74, n. 1, p. 93–105, 2004.

WEI, Q.; XU, Q.; DUNBRACK, R. L. Prediction of phenotypes of missense mutations in human proteins from biological assemblies. **Proteins: Structure, Function and Bioinformatics**, v. 81, n. 2, p. 199–213, 2013.

WEINRAUCH, L. A.; ZIEVE, D.; OGILVIE, I. **Congenital heart disease: MedlinePlus Medical Encyclopedia**. Disponível em: <<https://medlineplus.gov/ency/article/001114.htm>>. Acesso em: 19 jun. 2017.

WHO | Cardiovascular diseases (CVDs). **WHO**, 2017.

WHO | Congenital anomalies. **WHO**, 2016.

WOLFE, S. A.; NEKLUDOVA, L.; PABO, C. O. DNA Recognition by Cys2His2 Zinc Finger Proteins. **Annu. Rev. Biophys. Biomol. Struct.**, v. 29, p. 183–212, 2000.

WU, Z.; WANG, S. Role of kruppel-like transcription factors in adipogenesis. **Developmental Biology**, v. 373, n. 2, p. 235–243, 2013.

XUDONG LIAO, SAPTARSI M. HALDAR, YUAN LU, DARWIN JEYARAJ, KAAVYA PARUCHURI, M.; NAHORI, YINGJIE CUI, KLAUS H. KAESTNER, AND M. K. J. Krüppel-like factor 4 regulates pressure-induced cardiac hypertrophy. **J Mol Cell Cardiol.**, v. 49, n. 2, p. 334–338, 2010.

YATES, A. et al. Ensembl 2016. **Nucleic Acids Research**, v. 44, n. D1, p. D710–D716, 2016.

5 MANUSCRITO 1

Bioinformatics analysis of non-synonymous variants in the KLF genes related to cardiac diseases

Manuscrito submetido no formato *Research Article* à revista *Gene* (Elsevier).

Fator de impacto: 2.415

Periódico Qualis (Plataforma Sucupira): Ciências Biológicas 1 - B2

ISSN: 0378-1119

Link: <https://www.journals.elsevier.com/gene/>

Data da submissão: 19 de Setembro de 2017.

Elsevier Editorial System(tm) for Gene
Manuscript Draft

Manuscript Number:

Title: Bioinformatics analysis of non-synonymous variants in the KLF genes related to cardiac diseases

Article Type: Research paper

Keywords: Kruppel-like Factors; Non-synonymous variants; Bioinformatics analyses; Structural and functional impact; Cardiac diseases.

Corresponding Author: Ms. Katyana Kaline Silva Ferreira, MSc.

Corresponding Author's Institution: Federal University of Pernambuco - UFPE

First Author: Katyana Kaline Silva Ferreira, MSc.

Order of Authors: Katyana Kaline Silva Ferreira, MSc.; Enéas M Gomes, PhD.; José L Lima-Filho, PhD.; Carlos Henrique M Castelletti, PhD.; Danyelly G Martins, PhD.

Manuscript Region of Origin: BRAZIL

Abstract: Kruppel-like Factors (KLF) are responsible for regulating many genes involved in physiological and pathological processes. They are characterized by three conserved zinc-fingers in the DNA-binding domain, wherein mutations could affect the binding efficiency and transcription regulation. This study aimed to perform bioinformatics analysis to determine the most deleterious non-synonymous variants in KLFs involved in cardiac development and diseases, and their effects over the protein structure and stability. Eight hundred and fifty non-synonymous variants were found in seven KLFs related to cardiac diseases. Seventeen algorithms were used to predict the effect of selected variants over the structure and function of seven KLFs. The Top3 variants were selected in each category of conserved and non-conserved residues in the zinc-finger (ZF) domain. KLF5 p.Cys410Phe was the only variant predicted as deleterious in all algorithms, occurring in a conserved residue of zinc ion interaction. KLF15 p.Arg364Pro was the only variant predicted to affect the DNA-binding, and also occurs in a conserved ZF-domain. Our bioinformatics analysis determined potential variants that may lead to development of cardiac diseases, as well as reinforced the importance of KLF analysis in vitro and in vivo.

Bioinformatics analysis of non-synonymous variants in the KLF genes related to cardiac diseases

Katyana Kaline Silva Ferreira^{a*}; Enéas Ricardo de Moraes Gomes^b; José Luiz de Lima Filho^{c,d}; Carlos Henrique Madeiros Castelletti^{a,e}; Danyelly Bruneska Gondim Martins^{a,c}

Molecular Prospection and Bioinformatics Group (ProspecMol), Laboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil.

^a Molecular Prospection and Bioinformatics Group (ProspecMol), Laboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil.

^b Department of Biotechnology, Center of Biotechnology, Federal University of Paraíba (UFPB), Cidade Universitária, Campus I s/n, João Pessoa – PB, 58051-900, Brazil.

^c Laboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil.

^d Biochemistry Department, Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil.

^e Agronomic Institute of Pernambuco (IPA), Av. General San Martin 1371, Bongi, Recife – PE, 50761-000, Brazil.

* *Corresponding author at* Laboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil, Phone: +55 81 321268484, e-mail address: kferreira@prospecmol.org.

E-mail addresses: kferreira@prospecmol.org (Katyana Kaline Silva Ferreira); eneasricardo@cbiotec.ufpb.br (Enéas Ricardo de Moraes Gomes);

joseluiz60@gmail.com (José Luiz de Lima Filho); hcastelletti@prospecmol.org (Carlos Henrique Madeiros Castelletti); bruneska@prospecmol.org (Danyelly Bruneska Gondim Martins).

Abstract

Kruppel-like Factors (KLF) are responsible for regulating many genes involved in physiological and pathological processes. They are characterized by three conserved zinc-fingers in the DNA-binding domain, wherein mutations could affect the binding efficiency and transcription regulation. This study aimed to perform bioinformatics analysis to determine the most deleterious non-synonymous variants in KLFs involved in cardiac development and diseases, and their effects over the protein structure and stability. Eight hundred and fifty non-synonymous variants were found in seven KLFs related to cardiac diseases. Seventeen algorithms were used to predict the effect of selected variants over the structure and function of seven KLFs. The Top3 variants were selected in each category of conserved and non-conserved residues in the zinc-finger (ZF) domain. KLF5 p.Cys410Phe was the only variant predicted as deleterious in all algorithms, occurring in a conserved residue of zinc ion interaction. KLF15 p.Arg364Pro was the only variant predicted to affect the DNA-binding, and also occurs in a conserved ZF-domain. Our bioinformatics analysis determined potential variants that may lead to development of cardiac diseases, as well as reinforced the importance of KLF analysis *in vitro* and *in vivo*.

Keywords: Kruppel-like Factors; Non-synonymous variants; Bioinformatics analyses; Structural and functional impact; Cardiac diseases.

Abbreviations: KLF, Kruppel-like Factors; ZF, zinc-finger; HCM, hypertrophic cardiomyopathy; Nkx2, NK2 Homeobox; Nkx2.5, NK2 Homeobox 5; TBX5, T-box transcription factor 5; Sp1, transcription factor Sp1; CtBP, Carboxy-terminal binding protein; Sin3A, SIN3 Transcription Regulator Family Member A; DBD, DNA-binding

domain; dbNSFP, database for Non-Synonymous Functional Prediction; $\Delta\Delta G$, Gibbs free energy; C/EBP- α , CCAAT/Enhancer Binding Protein Alpha; C/EBP- β , CCAAT/Enhancer Binding Protein Beta; PPAR- γ , Peroxisome proliferator-activated receptors gamma.

1. INTRODUCTION

Worldwide, cardiac diseases are the leading cause of death and referred as one of the major problems for public health. According to World Health Organization, 17.7 million people died from cardiovascular diseases in 2015, and cardiac defects are one of the most common within severe congenital anomalies (“WHO | Cardiovascular diseases (CVDs),” 2017, “WHO | Congenital anomalies,” 2016). Studies in mice already revealed variants in more than 500 genes related to cardiac defects, but only 50 genes were associated in humans (Andersen et al., 2014). It indicates that more genes need further evaluation in humans.

Transcription factors are essential genes related to cardiac diseases, such as KLF, Nkx2, TBX5, and GATA4 (Basson et al., 1997; Dreßen et al., 2016; Kelsey et al., 2013; McConnell and Yang, 2010; Tong, 2016). In human, the Kruppel-like Factors (KLF) family has 17 members involved in the activation and repression of many genes on physiological and pathological processes, which some of these factors are related to cardiac development. KLFs have homology with the Sp1 transcriptional factors, showing a conserved C-terminal region with three zinc finger (ZF) motifs. This domain is responsible for regulating GC-boxes in promoters and enhancer regions of the genes (Kaczynski et al., 2003; McConnell and Yang, 2010; Nagai et al., 2009). Therefore, members of KLF family are classified into three groups according to its activity of gene repression or activation. Group 1 consists of the transcriptional repressors KLF3, KLF8 and KLF12 that interacts with carboxy-terminal binding protein (CtBP); group 2 is composed of the transcriptional activators KLF1, KLF2, KLF4-7; and group 3 consists of KLF9-11, KLF13-14, and KLF16 that acts as repressor together with the transcriptional corepressor Sin3A. KLF15 and KLF17 could not be included in any group since no motifs of interaction were defined yet (McConnell and Yang, 2010).

This ZF-motif type-C2H2 shows two residues of Cysteine and two residues of Histidine that anchors the zinc atom at the DNA-binding domain (DBD) (McConnell and Yang, 2010; Nagai et al., 2009; Razin et al., 2012). Each ZF consists of one α -helix and two antiparallel β -strands that bind to a zinc ion. There are four conserved residues responsible for binding KLF to DNA, playing a critical role in the nucleotide recognition. The residues occur at the initial portion of the ZF α -helix, in position -1 (before the first amino acid of α -helix), 2, 3, and 6 (Nagai et al., 2009; Razin et al., 2012; Wolfe et al., 2000). They are well conserved in each KLF but can be variable among the KLFs. ZF1 shows the sequence KXSHXXA and occurs in KLF1-13 and KLF15, while KXSHXXS occurs in KLF14, KLF16, and KLF17; both targets 5'-NGN-3' codon. In the ZF2, the RXDEXXR specific residues are conserved in all KLFs, recognizing 5'-GCG-3' target. In the ZF3, the conserved residues RXDHXXL are present in KLF1-8 and KLF12, recognizing 5'-NGG-3' as the target. RXDHXXXK sequence is found in KLF9-11 and KLF13-16, targeting the sequence 5'-(G/T)GG-3'; while RXDHXXQ sequence is present only in KLF17 and recognizes 5'-AGG-3' (Kaczynski et al., 2003; Nagai et al., 2009).

The role of KLF in activating and repressing genes is so important that the study of variants in conserved regions have great importance for evaluating the protein function and the relation with diseases. Therefore, bioinformatics analysis can be performed to predict the effect of a non-synonymous variant can have over structural stability and protein function through DNA-binding. This analysis can contribute to the evaluation of the impact of the non-synonymous variants in the disease development (Pires et al., 2014). This study aimed to use different algorithms to predict the impact of non-synonymous variants over KLF protein structure and function related to the cardiac development and diseases.

2. MATERIALS AND METHODS

2.1 Screening of non-synonymous variants

All 17 KLF genes were evaluated for non-synonymous variants at Ensembl genome browser, which has resources that integrate dbSNP, HGMD and ClinVar databases (Yates et al., 2016). Then, it was selected the KLF genes reported in cardiac development, cardiac diseases, and adipogenesis regulation.

2.2 Analysis of structural and functional impact

The prediction of structural and functional impact was performed using 16 among 20 prediction algorithms in dbNSFP (database for Non-Synonymous Functional Prediction) (v 3.4) (Liu et al., 2016). SNPs&GO algorithm was included as an individual web-based platform (Capriotti et al., 2013), based on PANTHER and PhD-SNP algorithms. Each algorithm has a threshold that determines if the variant is deleterious or neutral (Table 1). The thresholds of VEST3, CADD, and DANN algorithms were assessed after contacting the developers. Four algorithms were excluded from the final evaluation because the threshold was not available. The algorithm outputs were converted into a binary system, assigning 0 (zero) for no structural/functional effect, while 1 (one) was assigned for deleterious prediction over the protein function/structure. Therefore, the final score obtained by the sum of the binary system reaches to the maximum of 17.

2.3 Modeling and structural analysis

The effect of the non-synonymous variants over the tridimensional (3D) structure of the zinc-finger domain was evaluated through the 3D model built at the

SWISS-MODEL (Bordoli et al., 2009). KLF4 (PDB ID: 2WBS) (Schuetz et al., 2011) and KLF5 (PDB ID: 2EBT) were used for template-based modeling with at least 60% homology with other KLFs. 2WBS is a crystal structure with X-Ray resolution of 1.7Å, R-Value Free 0.232, R-Value Work 0.200; while the 2EBT resolution was under NMR method (Nuclear Magnetic Resonance). Hydrogen was inserted in the structure using MolProbity to determine the quality of the 3D model. Ramachandran Plot evaluation was performed in the final structure to determine the presence of more than 95% of residues in the allowed region. Models were approved by MolProbity when the final score was near to 100th, as result of the combined evaluation of the clash score, rotamer, and Ramachandran data (Chen et al., 2010). The UCSF Chimera software was used for 3D visualization of the protein structure (Pettersen et al., 2004), and Kyte & Doolittle hydrophobicity scale was used to analyze the hydrophobicity surface (Kyte and Doolittle, 1982).

2.4 Protein stability upon mutation

The impact of the variants over the protein stability was evaluated through the prediction of the $\Delta\Delta G$ (Gibbs free energy) upon mutation and the (de)stabilization level of the protein. mCSM-Stability (mutation Cutoff Scanning Matrix) predicts the impact that relies on graph-based signatures, using the atom distances to construct environments, trained through experimental data. $\Delta\Delta G$ value higher than 0 is assumed as stabilizing, while lower than 0 are destabilizing (Pires et al., 2014). On the other hand, MAESTRO predicts the impact based on the multi-agent system through statistical scoring functions, combining Multiple Linear Regression, Neural Network approach and Support Vector Machine to improve the predictive power. It gives the confidence estimation (Cpred) about prediction accuracy in a range of 0 (lowest) to 1

(highest). MAESTRO predicts a stabilizing variant when the $\Delta\Delta G$ value is lower than 0, while higher values are predicted as destabilizing (Laimer et al., 2015).

3. RESULTS AND DISCUSSION

3.1. KLFs involved in cardiac development and heart diseases

Variants in some transcription factors have been related to the development of cardiac diseases, as reported for TBX5 in Holt-Oram Syndrome, NKX2.5 and GATA4 in congenital heart diseases (Basson et al., 1997; Dreßen et al., 2016; Tong, 2016). KLFs can interact with other transcription factors, as reported for GATA4 (Brey et al., 2009; Fisch et al., 2007), then variants in KLFs could also have some effect in the cardiac development. However, the impact of non-synonymous variants in the structure and function of the KLF is not fully understood.

KLF genes are expressed in several tissues in different stages of the development and throughout life. KLF3-5 and KLF15 were reported to participate in the development of cardiac diseases, like hypertrophic cardiomyopathy (HCM) and cardiac hypertrophy (Table 2). These KLFs also participate in adipogenesis process through regulation of C/EBP- α , C/EBP- β and PPAR- γ (Brey et al., 2009; McConnell and Yang, 2010; Wu and Wang, 2013). KLF13 is related to the 15q13.3 microdeletion (microde15q13.3) syndrome (Tropeano et al., 2014), characterized by a broad spectrum of neurodevelopmental disorders, but not related to the rare occurrence of congenital heart defect in these patients. KLF10 is involved HCM and cardiac hypertrophy, but its role in adipogenesis remains unclear. Although KLF2 participates in the adipogenesis and KLF13 is expressed during this process, neither cardiac diseases nor cardiac defects were related to these transcriptional factors in humans. However, both seem to be

critical for cardiac development through GATA4 upregulation, which is one of the key transcriptional mediators of the cardiac hypertrophy (Heineke and Molkentin, 2006). KLF4 and KLF15 also have GATA4 gene as a target, so deregulation of those KLFs may also play a role in the development of cardiac diseases, like HCM or cardiac hypertrophy (Huang et al., 2008). In fact, KLF shows great importance in regulating critical effectors in the development of cardiac diseases.

3.1 Deleterious variants in KLF

KLF family has 2,048 non-synonymous variants reported, being 850 variants observed in the seven KLFs reported in cardiac diseases and adipogenesis process. Only 125 variants occurred in the ZF-domain (Fig. 1). KLF4, the most studied members of the family, showed the highest number of 218 variants. Then, KLF10, KLF15, KLF5, and KLF3 showed 163, 157, 112 and 108, respectively. Regarding to KLF2 and KLF13, only 49 and 43 variants were reported for each gene, respectively (Fig. 2). KLF13 is one of the newest in the family, showing just a few studies relating to diseases.

Six variants were selected for further analysis, the Top3 variants in conserved residues of the ZF-domain, and the Top3 in non-conserved residues of the ZF-domain (Table 3). Only KLF5 p.Cys410Phe achieved the maximum score of 17, indicating that was deleterious in all algorithms (Table 3). In non-conserved residues group, KLF2 p.Gly310Cys showed score 14, together with KLF10 p.Arg438Glu and KLF10 p.Arg438Gly. Only the variant that showed the deleterious effect predicted by Polyphen2-HDIV and MutPred, functional/structural algorithms, were selected for 3D analysis.

All of the variants selected were located at ZF2 (Table 4), except for KLF10 that showed variants in the ZF3 (Fig. 3). ZF2 seems to be the most conserved ZF in the

KLF family, once it has the same conserved residues of DNA-binding in all KLF members (Nagai et al., 2009). Thus, variants in ZF2 would be more likely to show a deleterious prediction on the algorithms. In humans, KLF1 (p.Glu339Asp) variant was related to neonatal anemia. It occurs in a conserved residue of ZF2 involved in DNA-binding, leading to a loss of DNA-binding specificity and KLF1 binding to degenerate motifs at genomic loci, not usually occupied by wild-type KLF1 (Gillinder et al., 2017).

3.2 Top3 deleterious variants in conserved residues

KLF5 p.Cys410Phe was predicted as deleterious in all algorithms tested, including PolyPhen2-HDIV and MutationTaster, suggesting this variant as significant for KLF function. Only these two algorithms were used to predict TBX5 p.Pro85Thr as high probability to be damaging, being present in a patient with congenital heart defects and upper limb malformations associated with Holt-Oram Syndrome (Dreßen et al., 2016). KLF5 p.Cys410Phe occurs in the ZF2, changing a conserved Cysteine that binds to zinc ion by a Phenylalanine (Fig. 4A). A destabilization effect was predicted for KLF5 p.Cys410Phe by mCSM and MAESTRO algorithms (Table 5), reinforcing the great importance of this variant in the protein functionality.

Variants that change conserved residues are more likely to be predicted as deleterious and destabilizing by the algorithms (Razin et al., 2012). As observed for the transcription factor Gal4, 90% of the variants that affect the coordination of zinc ion showed association with the phenotype of cell growth reduction, due to a lower range of transcription mediated by Gal4 (Carter et al., 2013). KLF5 is expressed in fibroblasts, acting as a cardioprotective factor through paracrine regulation involving IGF-1. As important regulator of cardiac remodeling, KLF5 modulates the cardiomyocytes hypertrophy, establishing a bridge between the fibroblasts and cardiomyocytes that lead

to pathogenesis regulation of cardiac hypertrophy and cardiac remodeling (TAKAYUKI SHINDO et al., 2002; Takeda et al., 2010). So, KLF5 p.Cys410Phe seems to reduce the chance of zinc interaction, as observed for Gal4, supporting the prediction of loss of functionality.

KLF3 p.His314Asp showed a score of 15, changing a conserved residue in the ZF2 that is responsible for zinc interaction (Table 4) (Fig. 4B). This shift from a basic to an acid residue also results in destabilizing prediction for the 3D structure (Table 5). The KLF3 p.His257Arg variant in conserved Histidine was previously related to reduced blood pressure with enlarged chambers and valvular stenosis, biventricular cardiac hypertrophy in heterozygotes neonatal, and death in homozygotes mice embryos (Kelsey et al., 2013). Once it occurs in a conserved residue in the DBD, it seems to be critical for the transcription regulation of development of cardiac genes and could become a therapeutic target for congenital heart diseases and cardiovascular disorders (Kelsey et al., 2013; Prosdocimo et al., 2015).

KLF15 p.Arg364Pro showed a score of 12, changing a conserved Arginine in the -1 position of DNA-binding in ZF2, highly conserved in all KLF members (Nagai et al., 2009). The shift from a positively charged state to a neutral residue reduces its interaction with the negative DNA backbone. Besides, this variant replaces a polar for a nonpolar neutral residue, increasing the hydrophobicity from -4.5 to -1.6, modifying the molecule surface (Fig. 5). KLF15 was demonstrated to be a negative regulator of cardiac hypertrophy, through inhibition of a set of genes including GATA4, which is essential for cardiac development. It is downregulated in cardiac tissue from rodent models of hypertrophy as well in biopsy samples from patients with pressure-overload state induced (Shihab et al., 2013b). Thus, KLF15 p.Arg364Pro is a potential variant to

cause damage in the protein, leading to the loss of its negative regulator function in cardiac diseases.

3.3 Top3 deleterious variants in non-conserved residues

KLF15 p.Phe362Tyr achieved a score of 16, which is higher than KLF15 p.Arg364Pro found in the Top3 variants in conserved residues (Table 3). This result reinforces the idea that both variants in DNA-binding and non-DNA-binding positions can be likely to disrupt the protein function, once no difference was observed in the number of deleterious variants located in these positions from zinc fingers C₂H₂-type (Lockwood et al., 2014). KLF15 p.Phe362Tyr reduces the hydrophobicity from 2.8 to -1.3, which could lead to the protein destabilization and be related to the development of cardiac diseases also linked to GATA4 regulation.

KLF10 p.Phe427Val occurs at ZF3 and showed a score of 15, with an increment in the hydrophobicity levels from 2.8 to 4.2, predicted as destabilizing (Table 5). This variant showed the highest destabilization value predicted by mCSM ($\Delta\Delta G = -2.628$ Kcal/mol), which indicates a high chance of protein malfunction and association to cardiac disease phenotype. Values of $\Delta\Delta G < -2$ Kcal/mol were associated with a significant effect on the protein stability in Gal4, highly associated with the reduced cell growth (Pires et al., 2016), which suggest KLF10 p.Phe427Val could lead to protein malfunction and development of HCM. It may result in loss of its suppressor function over the Pituitary Tumor Transforming Gene (Pttg1), a gene that regulates the chromatid separation during mitosis and is involved in HCM and tumor cells hypertrophy (Bos et al., 2012; Tfelt-hansen et al., 2006). Other non-synonymous variants in KLF10 (p.Ala12Thr, p.Met27Thr, p.Glu137Lys, p.Ala204Thr, p.Thr216Ala,

and p.Ser225Asn) were related to HCM (Bos et al., 2012), though none was occurring in ZF-domain.

KLF2 p.Gly310Cys changes a nonpolar for a polar residue, showing a score of 14. The analysis of protein stability revealed a divergent between the prediction of mCSM and MAESTRO (Table 5). Despite the high score, no great change was observed for $\Delta\Delta G$ values and 3D models. Therefore, it is not possible to determine the role of this variant in KLF2 functionality, that is known to lead to a range of defects including thinning myocardium, high output heart failure, cardiac functional defects and death in KLF2 knockout mice (Lee et al., 2006). Ablation of KLF2 in mice also resulted in mRNA reduction of Sox9, UDP-glucose dehydrogenase (Ugdh), Tbx5 and GATA4 in atrioventricular canals, indicating that KLF2 interact on promoters of these genes and regulates their expression (Chiplunkar et al., 2013).

4. CONCLUSION

The present study shows that KLF variants can be related to cardiac diseases after evaluating them in 17 algorithms, giving higher confidence due to the different sets of criteria to predict deleterious variants. Our results pointed out six non-synonymous variants with potential for destabilizing the protein, and experimental studies should be performed to understand their effect in cardiac diseases. Therefore, they might become biomarkers for early detection of cardiac diseases.

Conflicts of interest: none.

Acknowledgments: The authors are grateful to Renato P. Melo-Neto and Vinícius Albertim Tigre da Costa for the contributions in the design of the figure.

Funding: This work was supported by the *Coordination for the Improvement of Higher Education Personnel (CAPES)*. This research did not receive another specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCE

- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi:10.1038/nmeth0410-248
- Andersen, T.A., Troelsen, K. de L.L., Larsen, L.A., 2014. Of mice and men: molecular genetics of congenital heart disease. *Cell. Mol. life Sci.* 71, 1327–1352. doi:10.1007/s00018-013-1430-1
- Basson, C.T., Bachinsky, D.R., Lin, R.C., Levi, T., Elkins, J.A., Soultz, J., Grayzel, D., Kroumpouzou, E., Traill, T.A., Leblanc-Straceski, J., Renault, B., Kucherlapati, R., Seidman, J.G., Seidman, C.E., 1997. Mutations in human cause limb and cardiac malformation in Holt-Oram syndrome. *Nat. Genet.* 15, 30–35. doi:10.1038/ng0197-30
- Birsoy, K., Chen, Z., Friedman, J., 2008. Transcriptional Regulation of Adipogenesis by KLF4. *Cell Metab.* 7, 339–347. doi:10.1016/j.cmet.2008.02.001
- Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., Schwede, T., 2009. Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* 4, 1–13. doi:10.1038/nprot.2008.197
- Bos, J.M., Subramaniam, M., Hawse, J.R., Christiaans, I., Rajamannan, N.M., Maleszewski, J.J., Edwards, W.D., Wilde, A.A.M., Spelsberg, T.C., Ackerman,

- M.J., 2012. TGF β -inducible early gene-1 (TIEG1) mutations in hypertrophic cardiomyopathy. *J. Cell. Biochem.* 113, 1896–1903. doi:10.1002/jcb.24058
- Brey, C.W., Nelder, M.P., Hailemariam, T., Gaugler, R., Hashmi, S., 2009. Krüppel-like family of transcription factors: An emerging new frontier in fat biology. *Int. J. Biol. Sci.* 5, 622–636. doi:10.7150/ijbs.5.622
- Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P.L., Altman, R.B., Casadio, R., 2013. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics* 14. doi:10.1186/1471-2164-14-S3-S6
- Carter, H., Douville, C., Stenson, P.D., Cooper, D.N., Karchin, R., 2013. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* 14, S3. doi:10.1186/1471-2164-14-S3-S3
- Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 12–21. doi:10.1107/S0907444909042073
- Chiplunkar, A.R., Lung, T.K., Alhashem, Y., Koppenhaver, B.A., Salloum, F.N., Kukreja, R.C., Haar, J.L., Lloyd, J.A., 2013. Kruppel-Like Factor 2 Is Required for Normal Mouse Cardiac Development. *PLoS One* 8. doi:10.1371/journal.pone.0054891
- Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P., 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One* 7. doi:10.1371/journal.pone.0046688
- Chun, S., Fay, J.C., 2009. Identification of deleterious mutations within three human genomes. *Genome Res.* 19, 1553–1561. doi:10.1101/gr.092619.109.2001

- De Assuncao, T.M., Lomberk, G., Cao, S., Yaqoob, U., Mathison, A., Simonetto, D.A., Huebert, R.C., Urrutia, R.A., Shah, V.H., 2014. New role for Kruppel-like factor 14 as a transcriptional activator involved in the generation of signaling lipids. *J. Biol. Chem.* 289, 15798–15809. doi:10.1074/jbc.M113.544346
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., Liu, X., 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* 24, 2125–2137. doi:10.1093/hmg/ddu733
- Dreßen, M., Lahm, H., Lahm, A., Wolf, K., Doppler, S., Deutsch, M.-A., Cleuziou, J., Pabst von Ohain, J., Schön, P., Ewert, P., Malcic, I., Lange, R., Krane, M., 2016. A novel de novo TBX5 mutation in a patient with Holt-Oram syndrome leading to a dramatically reduced biological function. *Mol. Genet. genomic Med.* 4, 557–67. doi:10.1002/mgg3.234
- Fisch, S., Gray, S., Heymans, S., Haldar, S.M., Wang, B., Pfister, O., Cui, L., Kumar, A., Lin, Z., Sen-Banerjee, S., Das, H., Petersen, C.A., Mende, U., Burleigh, B.A., Zhu, Y., Pinto, Y.M., Pinto, Y., Liao, R., Jain, M.K., 2007. Kruppel-like factor 15 is a regulator of cardiomyocyte hypertrophy. *PNAS* 104, 7075–7079. doi:10.1073/pnas.0701981104
- Gillinder, K.R., Ilsley, M.D., Danitza, N., Sachidanandam, R., Lajoie, M., Magor, G.W., Tallack, M.R., Bailey, T., Landsberg, J., Mackay, J.P., Parker, M.W., Miles, L.A., Graber, J.H., Peters, L.L., Bieker, J.J., Perkins, A.C., 2017. OUP accepted manuscript. *Cereb. Cortex* 45, 1130–1143. doi:10.1093/cercor/bhw393
- Haldar, S.M., Ibrahim, O.A., Jain, M.K., 2007. Kruppel-like Factors (KLFs) in muscle biology. *J. Mol. Cell. Cardiol.* 43, 1–10. doi:10.1016/j.yjmcc.2007.04.005
- Heineke, J., Molkentin, J.D., 2006. Regulation of cardiac hypertrophy by intracellular

- signalling pathways. *Nat. Rev. Mol. Cell Biol.* 7, 589–600. doi:10.1038/nrm1983
- Huang, Y., Wright, C.D., Kobayashi, S., Healy, C.L., Elgethun, M., Cypher, A., Liang, Q., O’Connell, T.D., 2008. GATA4 is a survival factor in adult cardiac myocytes but is not required for alpha1A-adrenergic receptor survival signaling. *Am. J. Physiol. Heart Circ. Physiol.* 295, H699-707. doi:10.1152/ajpheart.01204.2007
- Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-albright, L.A., Teerlink, C.C., Stanford, J.L., Isaacs, W.B., Xu, J., Cooney, K.A., Lange, E.M., Schleutker, J., Carpten, J.D., Powell, I.J., Cussenot, O., Cancel-tassin, G., Giles, G.G., Macinnis, R.J., Maier, C., Hsieh, C., Wiklund, F., Catalona, W.J., Foulkes, W.D., Mandal, D., Eeles, R.A., Kote-jarai, Z., Bustamante, C.D., Schaid, D.J., Hastie, T., Ostrander, E.A., Bailey-wilson, J.E., Radivojac, P., Thibodeau, S.N., Whittemore, A.S., 2016. REVEL : An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* 99, 877–885. doi:10.1016/j.ajhg.2016.08.016
- Jagadeesh, K.A., Wenger, A.M., Berger, M.J., Guturu, H., Stenson, P.D., Cooper, D.N., Bernstein, J.A., Bejerano, G., 2016. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* 48, 1581–1586. doi:10.1038/ng.3703
- Kaczynski, J., Cook, T., Urrutia, R., 2003. Protein family review Sp1- and Krüppel-like transcription factors 1–8.
- Kelsey, L., Flenniken, A.M., Qu, D., Funnell, A.P.W., Pearson, R., Zhou, Y.Q., Voronina, I., Berberovic, Z., Wood, G., Newbigging, S., Weiss, E.S., Wong, M., Quach, I., Yeh, S.Y.S., Deshwar, A.R., Scott, I.C., McKerlie, C., Henkelman, M., Backx, P., Simpson, J., Osborne, L., Rossant, J., Crossley, M., Bruneau, B.,

- Adamson, S.L., 2013. ENU-induced Mutation in the DNA-binding Domain of KLF3 Reveals Important Roles for KLF3 in Cardiovascular Development and Function in Mice. *PLoS Genet.* 9. doi:10.1371/journal.pgen.1003612
- Kircher, M., Witten, D.M., Jain, P., O’Roak, B.J., Cooper, G.M., Shendure, J., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi:10.1038/ng.2892
- Kyte, J., Doolittle, 1982. A Simple Method for Displaying the Hydropathic Character of a Protein 157, 105–132.
- Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S., Lackner, P., 2015. MAESTRO - multi agent stability prediction upon point mutations. *BMC Bioinformatics* 16. doi:10.1186/s12859-015-0548-6
- Lavallée, G., Andelfinger, G., Nadeau, M., Lefebvre, C., Nemer, G., Horb, M.E., Nemer, M., 2006. The Kruppel-like transcription factor KLF13 is a novel regulator of heart development. *EMBO J.* 25, 5201–5213. doi:10.1038/sj.emboj.7601379
- Lee, J.S., Yu, Q., Shin, J.T., Sebzda, E., Bertozzi, C., Chen, M., Mericko, P., Stadtfeld, M., Zhou, D., Cheng, L., Graf, T., MacRae, C.A., Lepore, J.J., Lo, C.W., Kahn, M.L., 2006. Klf2 Is an Essential Regulator of Vascular Hemodynamic Forces In Vivo. *Dev. Cell* 11, 845–857. doi:10.1016/j.devcel.2006.09.006
- Li, B., Krishnan, V.G., Mort, M.E., Xin, F., Kamati, K.K., Cooper, D.N., Mooney, S.D., Radivojac, P., 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 25, 2744–2750. doi:10.1093/bioinformatics/btp528
- Liu, X., Wu, C., Li, C., Boerwinkle, E., 2016. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.* 37, 235–241. doi:10.1002/humu.22932

- Lockwood, S.H., Guan, A., Yu, A.S., Zhang, C., Zykovich, A., Korf, I., Rannala, B., Segal, D.J., 2014. The functional significance of common polymorphisms in zinc finger transcription factors. *G3 (Bethesda)*. 4, 1647–55.
doi:10.1534/g3.114.012195
- McConnell, B.B., Yang, V.W., 2010. Mammalian Krueppel-Like Factors in Health and Diseases. *Physiol. Rev.* 90, 1337–1381.
doi:10.1152/physrev.00058.2009.Mammalian
- Mori, T., Sakaue, H., Iguchi, H., Gomi, H., Okada, Y., Takashima, Y., Nakamura, K., Nakamura, T., Yamauchi, T., Kubota, N., Kadowaki, T., Matsuki, Y., Ogawa, W., Hiramatsu, R., Kasuga, M., 2005. Role of kruppel-like factor 15 (KLF15) in transcriptional regulation of adipogenesis. *J. Biol. Chem.* 280, 12867–12875.
doi:10.1074/jbc.M410515200
- Nagai, R., Friedman, S.L., Kasuga, M., 2009. The Biology of Kruppel-like Factors, 1st ed, Chapter 2: Molecular Structures of Kruppel-like Factor Function.
doi:10.1017/CBO9781107415324.004
- Ng, P.C., Henikoff, S., 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi:10.1093/nar/gkg509
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera — A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 25, 1605–1612.
doi:10.1002/jcc.20084
- Pires, D.E. V, Ascher, D.B., Blundell, T.L., 2014. mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 30, 335–342.
doi:10.1093/bioinformatics/btt691
- Pires, D.E. V, Chen, J., Blundell, T.L., Ascher, D.B., 2016. In silico functional

- dissection of saturation mutagenesis: Interpreting the relationship between phenotypes and changes in protein stability, interactions and activity. *Sci. Rep.* 6, 1–11. doi:10.1038/srep19848
- Prosdocimo, D.A., Sabeh, M.K., Jain, M.K., 2015. Kruppel-like factors in muscle health and disease. *Trends Cardiovasc. Med.* 25, 278–287. doi:10.1016/j.tcm.2014.11.006
- Quang, D., Chen, Y., Xie, X., 2015. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* 31, 761–763. doi:10.1093/bioinformatics/btu703
- Razin, S. V., Borunova, V. V., Maksimenko, O.G., Kantidze, O.L., 2012. Cys2His2 zinc finger protein family: Classification, functions, and major members. *Biochem.* 77, 277–288. doi:10.1134/S0006297912030017
- Reva, B., Antipin, Y., Sander, C., 2011. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.* 39, 1–14. doi:10.1093/nar/gkr407
- Schuetz, A., Nana, D., Rose, C., Zocher, G., Milanovic, M., Koenigsmann, J., Blasig, R., Heinemann, U., Carstanjen, D., 2011. The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. *Cell. Mol. Life Sci.* 68, 3121–3131. doi:10.1007/s00018-010-0618-x
- Schwarz, J.M., Rödelberger, C., Schuelke, M., Seelow, D., 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 7, 575–576. doi:10.1038/nmeth0810-575
- Shihab, H.A., Gough, J., Cooper, D.N., Day, I.N.M., Gaunt, T.R., 2013a. Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics* 29, 1504–1510. doi:10.1093/bioinformatics/btt182
- Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L.A., Edwards, K.J.,

- Day, I.N.M., Gaunt, T.R., 2013b. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Hum. Mutat.* 34, 57–65. doi:10.1002/humu.22225
- TAKAYUKI SHINDO, I., FUKUSHIMA, Y., TOBE, K., AIZAWA, K., SAKUMIYAMOTO¹, KAWAI-KOWASE⁶, K., NOBUOMORIYAMA⁴, IMAI¹, YASUSHIKAWAKAMI⁷, H., NISHIMATSU⁴, H., ISHIKAWA², T., , TORU SUZUKI¹ , HIROYUKIMORITA¹ , KOJIMAEMURA¹ , MASATAKA SATA¹ , YASUNOBU HIRATA¹ , MASAYUKI KOMUKAI⁸ , HIROYUKI KAGECHIKA⁵ , TAKASHI KADOWAKI³, M.K.& R.N., 2002. Krüppel-like zinc-finger transcription factor KLF5/BTEB2 is a target for angiotensin II signaling and an essential regulator of cardiovascular remodeling. *Nat. Med.* 8, 856–863. doi:10.1038/nm
- Takeda, N., Manabe, I., Uchino, Y., Eguchi, K., Matsumoto, S., Nishimura, S., Shindo, T., Sano, M., Otsu, K., Snider, P., Conway, S.J., Nagai, R., 2010. Cardiac fibroblasts are essential for the adaptive response of the murine heart to pressure overload. *J. Clin. Invest.* 120, 254–265. doi:10.1172/JCI40295DS1
- Tfelt-hansen, J., Kanuparthi, D., Chattopadhyay, N., 2006. The Emerging Role of Pituitary Tumor Transforming Gene in Tumorigenesis. *Clin. Med. Res.* 4, 130–137.
- Tong, Y.-F., 2016. Mutations of NKX2.5 and GATA4 genes in the development of congenital heart disease. *Gene* 588, 86–94. doi:10.1016/j.gene.2016.04.061
- Tropeano, M., Andrieux, J., Vassos, E., Collier, D.A., 2014. Clinical utility gene card for: 15q13.3 microdeletion syndrome. *Hum Genet* 22, 1338. doi:10.1038/ejhg.2014.88
- Uchida, S., Tanaka, Y., Ito, H., Saitoh-Ohara, F., Inazawa, J., Yokoyama, K.K., Sasaki,

S., Marumo, F., 2000. Transcriptional regulation of the CLC-K1 promoter by myc-associated zinc finger protein and kidney-enriched Krüppel-like factor, a novel zinc finger repressor. *Mol. Cell. Biol.* 20, 7319–7331.

doi:10.1128/MCB.20.19.7319-7331.2000

WHO | Cardiovascular diseases (CVDs), 2017. . WHO.

[http://www.who.int/mediacentre/factsheets/fs317/en/#.WS-](http://www.who.int/mediacentre/factsheets/fs317/en/#.WS-EGoJ4MYI.mendeley&title=Cardiovascular diseases)

[EGoJ4MYI.mendeley&title=Cardiovascular diseases](http://www.who.int/mediacentre/factsheets/fs317/en/#.WS-EGoJ4MYI.mendeley&title=Cardiovascular diseases)

(CVDs)&description=WHO cardiovascular diseases fact sheet providing key facts and information on risk factors, symptoms, rheumatic heart disease, treatment and prevention, WHO response. (accessed June 1, 2017).

WHO | Congenital anomalies, 2016. . WHO.

[http://www.who.int/mediacentre/factsheets/fs370/en/#.WNPZ11PrHJo.mendeley](http://www.who.int/mediacentre/factsheets/fs370/en/#.WNPZ11PrHJo.mendeley&title=Congenital anomalies&description=WHO fact sheet on congenital anomalies, an important cause of childhood death, chronic illness, and disability in many countries. (accessed March 23, 2017).)

[&title=Congenital anomalies&description=WHO fact sheet on congenital](http://www.who.int/mediacentre/factsheets/fs370/en/#.WNPZ11PrHJo.mendeley&title=Congenital anomalies&description=WHO fact sheet on congenital anomalies, an important cause of childhood death, chronic illness, and disability in many countries. (accessed March 23, 2017).)

anomalies, an important cause of childhood death, chronic illness, and disability in many countries. (accessed March 23, 2017).

Wolfe, S.A., Nekludova, L., Pabo, C.O., 2000. DNA Recognition by Cys2His2 Zinc Finger Proteins. *Annu. Rev. Biophys. Biomol. Struct.* 29, 183–212.

Wu, Z., Wang, S., 2013. Role of kruppel-like transcription factors in adipogenesis. *Dev. Biol.* 373, 235–243. doi:10.1016/j.ydbio.2012.10.031

Xudong Liao, Saptarsi M. Haldar, Yuan Lu, Darwin Jeyaraj, Kaavya Paruchuri, M., Nahori, Yingjie Cui, Klaus H. Kaestner, and M.K.J., 2010. Krüppel-like factor 4 regulates pressure-induced cardiac hypertrophy. *J Mol Cell Cardiol.* 49, 334–338. doi:10.1161/CIRCULATIONAHA.111.087940.The

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Gir??n, C.G., Gordon, L.,

- Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R., Flicek, P., 2016. Ensembl 2016. *Nucleic Acids Res.* 44, D710–D716. doi:10.1093/nar/gkv1157
- Yoshida, T., Gan, Q., Franke, A.S., Ho, R., Zhang, J., Chen, Y.E., Hayashi, M., Majesky, M.W., Somlyo, A. V, Owens, G.K., 2010. Smooth and Cardiac Muscle-selective Knock-out of Krüppel-like Factor 4 Causes Postnatal Death and Growth Retardation. *J. Biol. Chem.* 285, 21175–21184. doi:10.1074/jbc.M110.112482

Algorithm	Method	Database/algorithms	Deleterious Threshold	Reference ⁶⁸
Structural				
SIFT	NI	Swiss-Prot, Swiss-Prot with Trembl, UniRef-50, UniRef-90 and UniRef-100; PSI-BLAST.	Prediction: T/D	(Ng and Henikoff, 2003)
LRT	Likelihood Ratio Test of codon neutrality	OMIM database; Multiple Sequence Alignment (MSA).	Prediction: N/U/D	(Chun and Fay, 2009)
MutationTaster2	Naïve Bayes classifier	Ensembl, SwissProt / Uniprot database, dbSNP and HapMap.	Prediction: N/P/D	(Schwarz et al., 2010)
MutationAssessor	Combinatorial entropy formalism	COSMIC database; MSA.	Prediction: N/L/M/H	(Reva et al., 2011)
FATHMM	Hidden Markov Models	Human Gene Mutation Database (HGMD) and UniProt.	Prediction: T/D	(Shihab et al., 2013b)
Fathmm-MKL	Multiple Kernel Learning	HGMD and 1000 Genomes Project (1000G).	Prediction: N/D	(Shihab et al., 2013a)
PROVEAN	Delta alignment score	SwissProt / UniProt database; PSI-BLAST	Prediction: T/D	(Choi et al., 2012)
VEST3	Random Forest	HGMD and Exome Sequencing Project.	Score ≥ 0.95	(Carter et al., 2013)
CADD	Linear kernel SVM	Ensembl Variant Effect Predictor and UCSC genome browser; Grantham, SIFT and PolyPhen.	Phred – C score ≥ 20	(Kircher et al., 2014)
DANN	Neural Network	ClinVar database; CADD.	Score ≥ 0.96	(Quang et al., 2015)
Meta SVM	Radial kernel SVM	SIFT, PolyPhen2, MutationTaster, Mutation Assessor, FATHMM, LRT, GERP++, SiPhy*, PhyloP* and Minor Allele Frequency	Prediction: T/D	(Dong et al., 2015)

(MAF) in 1000G.				
Meta LR	Logistic Regression	SIFT, PolyPhen2, MutationTaster, Mutation Assessor, FATHMM, LRT, GERP++*, SiPhy*, PhyloP* and MAF in 1000G.	Prediction: T/D	(Dong et al., 2015)
M-CAP	Gradient boosting tree classifier	HGM and ExAC; SIFT, PolyPhen2, CADD, MutationTaster, MutationAssessor, FATHMM, LRT, MetaLR, MetaSVM. PhyloP*, PhastCons*, GERP++*, SiPhy*, RVIS, PAM250, BLOSUM62.	Prediction: T/D	(Jagadeesh et al., 2016)
REVEL	Random Forest	HGMD; MutPred, FATHMM, VEST3, Polyphen2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++*, SiPhy*, PhyloP*, and PhastCons*.	Score >0.375	(Ioannidis et al., 2016)
SNP&GO	SVM	Gene Ontology; PhD-SNP and PHANTER.	Score >0.5 and RI ≥ 5	(Capriotti et al., 2013)
Structural and functional				
Polyphen2-HDIV	Naïve Bayes	UniProt database; MSA and PSIC score (position-specific independent counts).	Prediction: B/P/D	(Adzhubei et al., 2010)
MutPred	Random Forest	PFAM database; SIFT and PSI-BLAST.	g (general score) >0.75	(Li et al., 2009)

Table 1. Algorithms and deleterious threshold used for computational analysis of KLF. NI= not identified; T: Tolerated; N: Neutral; D: Deleterious; U: Unknown; P: Probably damage; L: Low; M: medium; H: High; B: Benign and RI: Reliability Index. *Algorithms of conservation score. ¹Prediction assumed as deleterious and assigned as 1 in the binary matrix.

Protein	Function/Remarks	Reference
KLF2	Inhibits the adipogenesis process; Interacts with C/EBP- α , C/EBP- β and PPAR- γ ; It is important to cardiac development; Interact with GATA4 activating its expression.	(Brey et al., 2009; Chiplunkar et al., 2013; Wu and Wang, 2013)
KLF3	Inhibits the adipogenesis process; Interacts and inhibits C/EBP- α and C/EBP- β ; It is important to cardiac development being expressed in the embryonic stage; Involved with development of hypertrophic cardiomyopathy.	(Brey et al., 2009; Kelsey et al., 2013; McConnell and Yang, 2010; Wu and Wang, 2013)
KLF4	Activates the adipogenesis process; Interacts and activates C/EBP- α and C/EBP- β ; Expressed in initial stage of adipogenesis and in embryonic stage in the heart, increasing until adulthood; It is important to cardiac development, regulating GATA4 expression and inhibits the cardiac hypertrophy.	(Birsoy et al., 2008; Brey et al., 2009; Haldar et al., 2007; Wu and Wang, 2013; Xudong Liao, Saptarsi M. Haldar, Yuan Lu, Darwin Jeyaraj, Kaavya Paruchuri and Nahori, Yingjie Cui, Klaus H. Kaestner, 2010; Yoshida et al., 2010)
KLF5	Activates the adipogenesis process; Interacts and activates C/EBP- α , C/EBP- β and PPAR- γ ; Expressed in initial stage of adipogenesis and in the heart is expressed just in cardiac fibroblast; Reported leads to cardiac hypertrophy.	(Brey et al., 2009; TAKAYUKI SHINDO et al., 2002; Takeda et al., 2010; Wu and Wang, 2013)
KLF10	Expressed during the adipogenesis; It is important to cardiac development and inhibits the cardiac hypertrophy;	(De Assuncao et al., 2014; McConnell and Yang, 2010; Mori et al., 2005; Prosdocimo et al., 2015)

Reported leads to cardiac hypertrophy and hypertrophic cardiomyopathy.

KLF13	Expressed during the adipogenesis; It is important to cardiac development interacting with GATA4 activating its expression; Expressed in embryonic stage and decrease during adulthood.	(Fisch et al., 2007; Lavallée et al., 2006; Mori et al., 2005)
KLF15	Activates the adipogenesis process; Interacts and activates C/EBP- α and PPAR- γ ; It is important to cardiac development and inhibits the cardiac hypertrophy through inhibition of GATA4; Expressed in fetal period, postnatal and adult.	(Brey et al., 2009; Fisch et al., 2007; McConnell and Yang, 2010; Mori et al., 2005; Uchida et al., 2000)

Table 2. Reference table of the KLFs reported in the literature to be expressed in the heart tissue, involved in cardiac development and diseases, and adipogenesis process.

Algorithms	Top3 conserved variants			Top3 non-conserved variants		
	KLF5	KLF3	KLF15	KLF10	KLF2	
	C410F	H314D	R364P	F362Y	F427V	G310C
SIFT	1	1	1	1	1	1
LRT	1	1	1	1	1	0
MutationTaster2	1	1	1	1	1	1
MutationAssessor	1	1	1	1	1	1
FATHMM	1	0	0	1	1	0
fathmm-MKL	1	1	1	1	1	1
PROVEAN	1	1	1	1	1	1
VEST3	1	1	0	0	0	0
CADD	1	1	1	1	1	1
DANN	1	1	1	1	1	1
MetaSVM	1	1	0	1	1	1
MetaLR	1	1	0	1	1	1
M-CAP	1	1	1	1	1	1
REVEL	1	1	1	1	1	1
SNPs&GO	1	0	1	1	1	1
Polyphen2-HDIV	1	1	1	1	1	1
MutPred	1	1	0	1	0	1
Binary sum	17	15	12	16	15	14

Table 3. A binary score of the Top3 most deleterious variants of conserved and non-conserved residues in the zinc-finger domain of five KLFs reported to cardiac diseases. 1 was assigned for deleterious prediction, while 0 was assigned for neutral prediction.

Protein	Mutation	dbSNP	Accession Number	Ensembl ID	Change	Zinc Finger site
KLF2	p.Gly310Cys	rs747983105	NM_016270	ENST00000248071	c.928G>T	ZF2
KLF3	p.His314Asp	rs778933698	NM_016531	ENST00000261438	c.940C>G	ZF2
KLF5	p.Cys410Phe	rs762217397	NM_001730	ENST00000377687	c.1229G>T	ZF2
KLF10	p.Phe427Val	rs76769513	NM_001032282	ENST00000395884	c.1279T>G	ZF3
KLF15	p.Phe362Tyr	rs779753572	NM_014079	ENST00000296233	c.1085T>A	ZF2
	p.Arg364Pro	rs755719419	NM_014079	ENST00000296233	c.1091G>C	ZF2

Table 4. Data of the six variants most deleterious evaluated by 3D modeling. ZF: Zinc Finger.

Protein	Mutation	mCSM Prediction	MAESTRO Prediction
KLF2	G310C	Stabilizing	Destabilizing
KLF3	H314D	Destabilizing	Destabilizing
KLF5	C410F	Destabilizing	Destabilizing
KLF10	F427V	Destabilizing	Destabilizing
KLF15	F362Y	Destabilizing	Destabilizing
	R364P	Stabilizing	Stabilizing

Table 5. mCSM and MAESTRO prediction of protein stability upon mutation in each KLF.

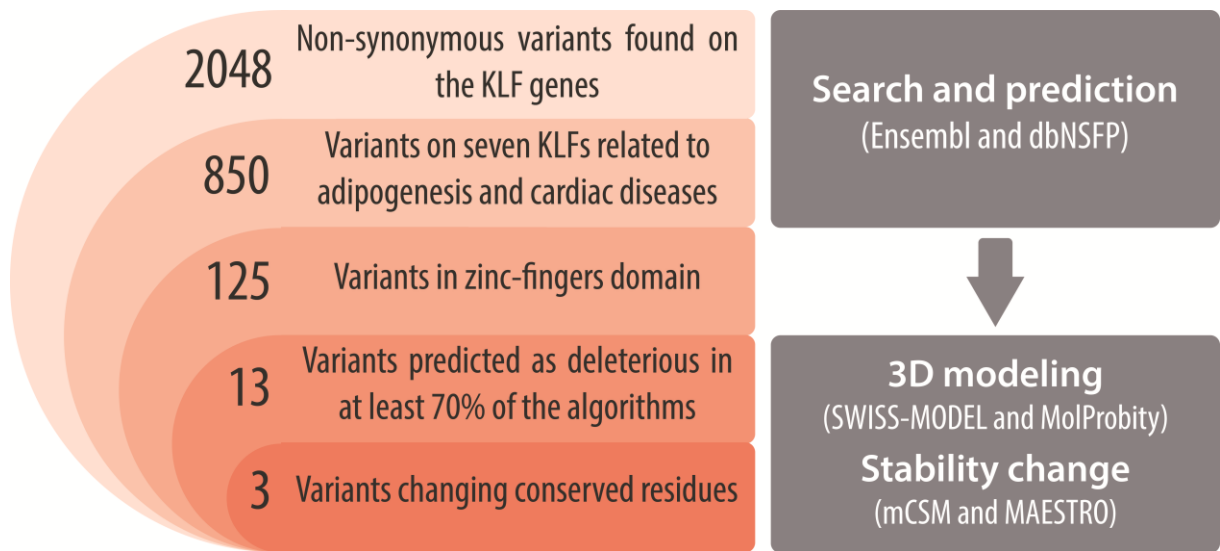
FIGURES

Figure 1. The pipeline of the KLF variants evaluated through prediction algorithms and 3D modeling by Swiss-model and MolProbity software.

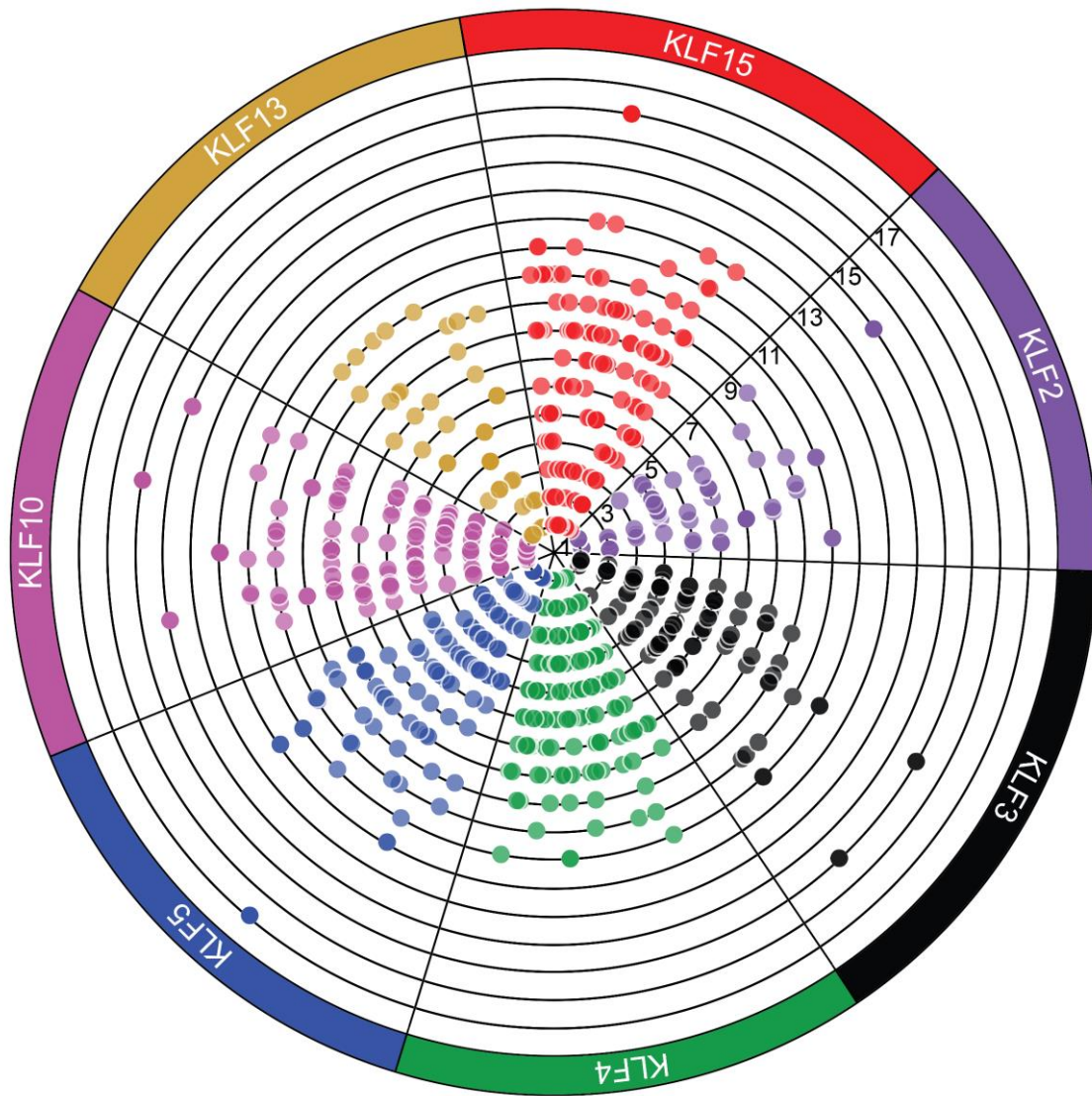


Figure 2. Distribution of mutations according to the binary score in the seven KLFs related to cardiac development and diseases. Each dots correspond to one variant in the KLF according to the colors.

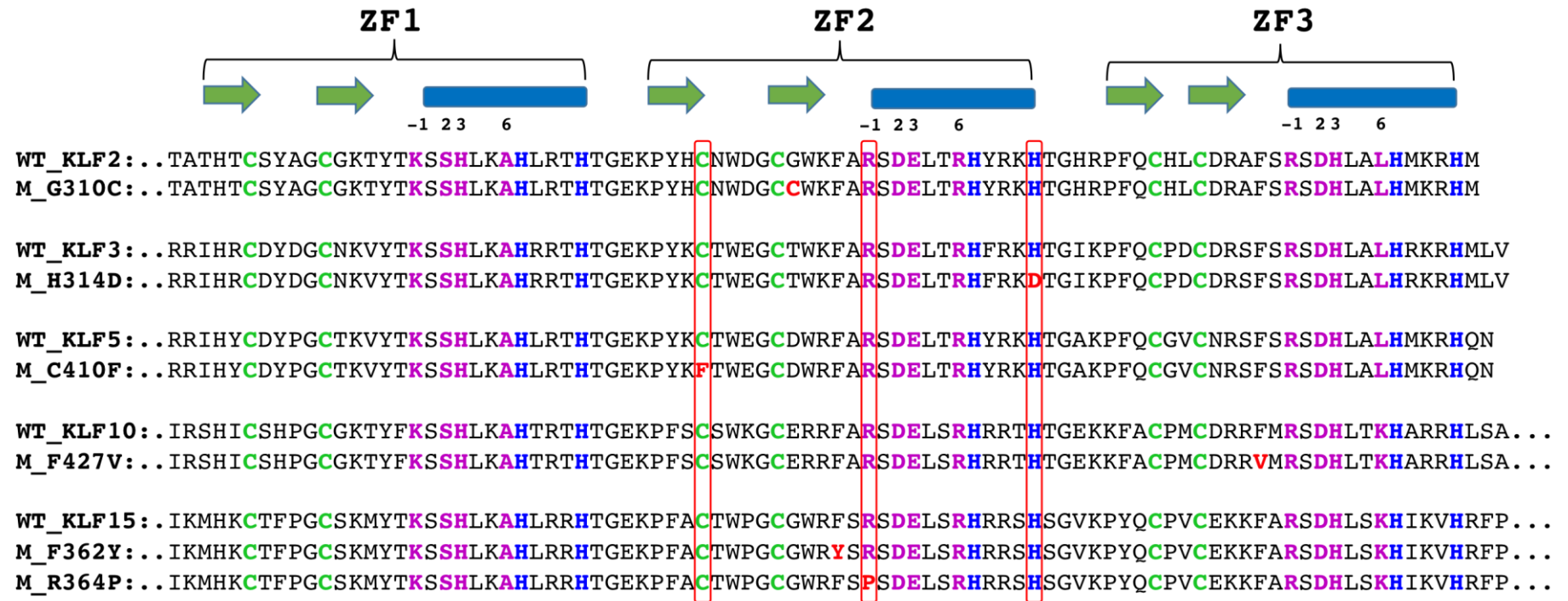


Figure 3. Alignment of the zinc finger domain sequences of the six KLFs variants more deleterious, showing conserved residues in each ZF. In green are represented Cysteine; in blue Histidine; in purple conserved residues in specific positions that bind to DNA target; and in red are presented the variants. WT: wild-type sequence; M: mutated sequence.

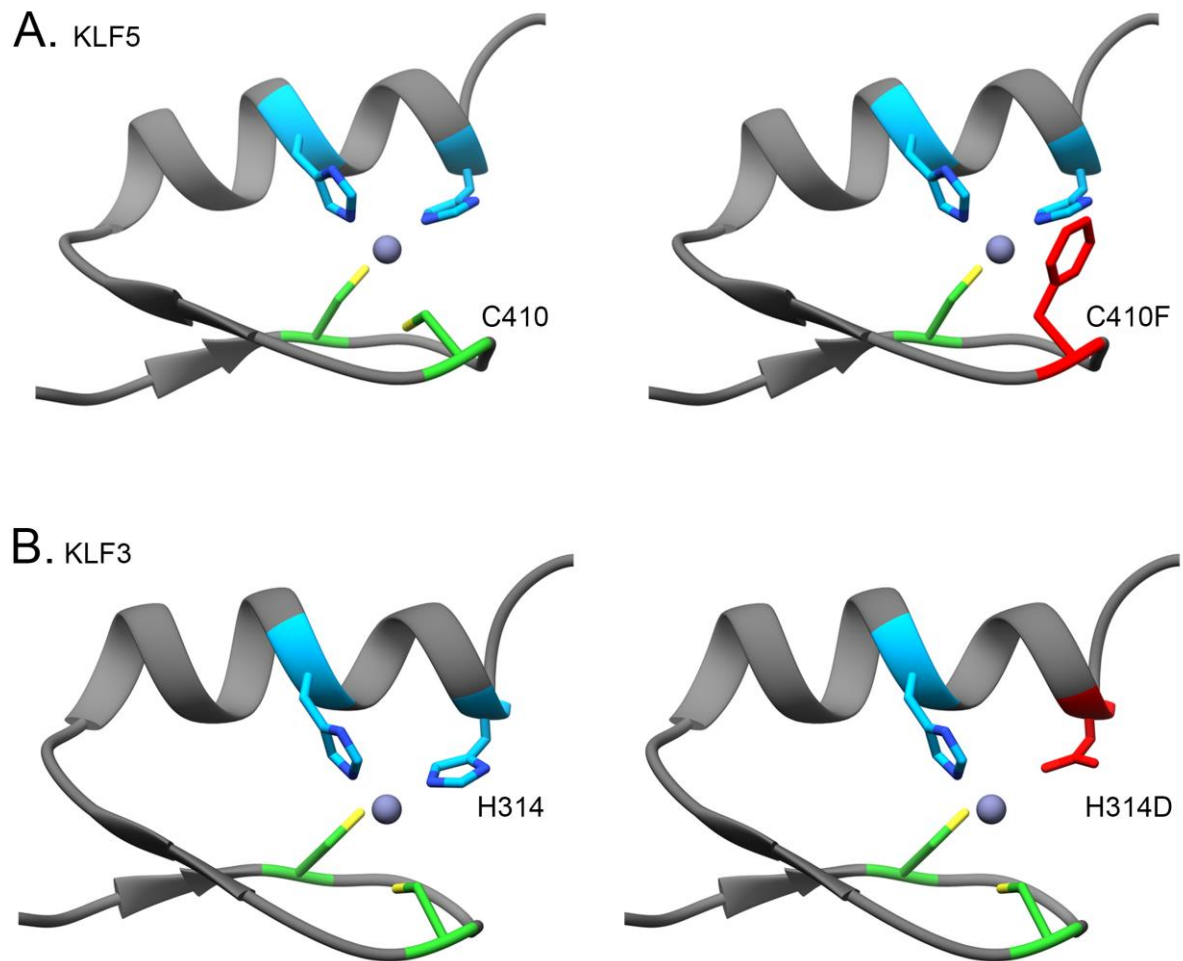


Figure 4. 3D models of the KLF wild-type and variants in the zinc finger domain and their zinc atom, showing the loss of interaction to zinc and destabilization of the ZF motif, generated and visualized by Swiss-model and Chimera, respectively. Residues in blue represent Histidine; in green Cysteine and red represent mutated residues. Gray sphere represents zinc atom. (A) Shows the KLF5 p.Cys410Phe and (B) KLF3 p.His314Asp.

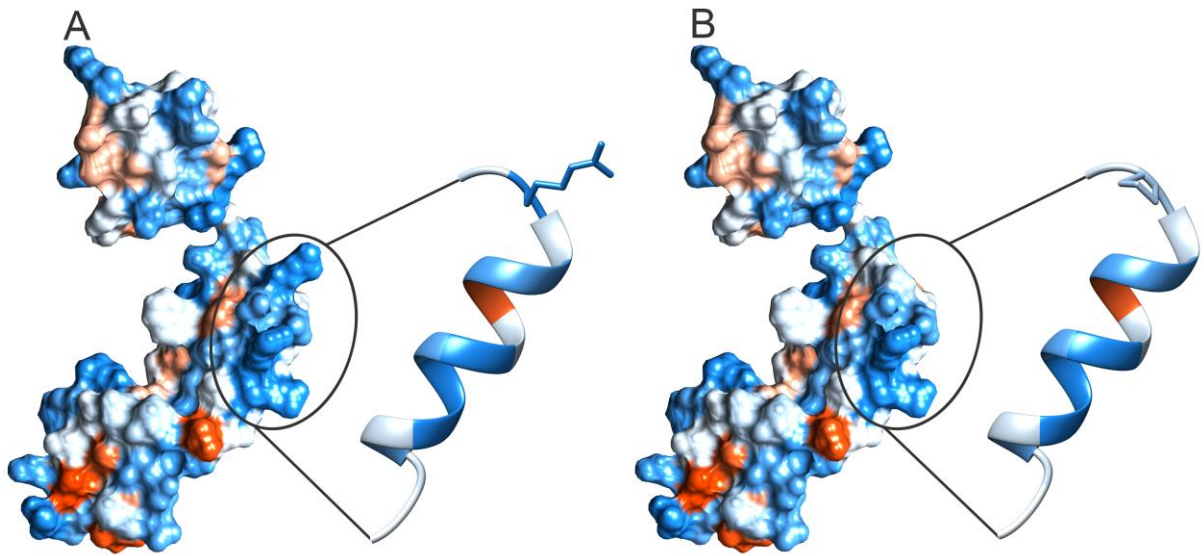


Figure 5. 3D models and visualization of the KLF15 p.Arg364Pro, showing the protein hydrophobicity surface, evidencing the change from a hydrophilic to the neutral residue, generated and visualized by Swiss-model and Chimera. Red color indicates a hydrophobic surface; White color indicates neutral, and blue color a hydrophilic surface. (A) KLF15 p.Arg364 DNA-binding conserved residue in the -1 position on the ZF2, (B) KLF15 p.Arg364Pro.

5. MANUSCRITO 2

Transcriptional imbalance of transcription factors: Could KLF2 and KLF10 be deregulated in congenital heart defects in Down syndrome?

Manuscrito no formato *Short Communication* submetido à revista *International Journal of Cardiology* (Elsevier).

Fator de impacto: 6.189

Periódico Qualis (Plataforma Sucupira): Ciências Biológicas I – A2

ISSN: 0167-5273

Link: <https://www.journals.elsevier.com/international-journal-of-cardiology>

Transcriptional imbalance of transcription factors: Could KLF2 and KLF10 be deregulated in congenital heart defects in Down syndrome?

Katyana Kaline Silva Ferreira^{a*}; Tereza Cristina Pinheiro Diogenes^b; Natália Didier Nunes Moser^b; Enéas Ricardo de Moraes Gomes^c; Sandra da Silva Mattos^b; José Luiz de Lima Filho^{d,e}; Danyelly Bruneska Gondim Martins^{a,d,e}

Molecular Prospection and Bioinformatics Group (ProspecMol), Laboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil.

^aMolecular Prospection and Bioinformatics Group (ProspecMol), Laboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil.

^bHeart Circle of Pernambuco, Recife-Pernambuco-Brazil.

^cDepartment of Biotechnology, Center of Biotechnology, Federal University of Paraíba (UFPB), Cidade Universitária, Campus I s/n, João Pessoa – PB, 58051-900, Brazil.

^dLaboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil.

^eBiochemistry Department, Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil.

**Corresponding author at Laboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco (UFPE). Av. Prof. Moraes Rego s/n, Recife – PE, 50670-901, Brazil, Phone: +55 81 321268484, e-mail address: kferreira@prospecmol.org.*

E-mail addresses: kferreira@prospecmol.org (Katyana Kaline Silva Ferreira); tereza.diogenes@hotmail.com (Tereza Cristina Pinheiro Diogenes); natidnmoser@gmail.com (Natália Didier Nunes Moser); eneasricardo@cbiotec.ufpb.br (Enéas Ricardo de Moraes Gomes); ssmattos@gmail.com (Sandra da Silva Mattos); joseluiz60@gmail.com (José Luiz de Lima Filho); bruneska@prospecmol.org (Danyelly Bruneska Gondim Martins).

Abstract

Down Syndrome (DS) is one of the most common genetic disorders, with more than 40% of the birth afflicted by some congenital heart defects (CHD). At least 50 human genes were reported associated with CHD or genetic syndromes, including Kruppel-like factors (KLF) genes. However, the role of KLFs in CHD development in DS children has not still known. The aim of this study was to evaluate the expression profile of KLF2 and KLF10 in DS children with CHD. Forty-two children were recruited and divided into four groups: DS, CHD, DS/CHD and healthy group. Among DS children, CHD prevalence was 63.63%, and atrioventricular septal defect (AVSD) was the most common defect (42.85%). The evaluation of gene expression in blood cells showed KLF2 downregulated in all groups. KLF10 showed no expression in healthy children but was expressed in DS and DS/CHD non-AVSD group ($p=0.0111$). Our results suggest KLF2 and KLF10 are deregulated in blood cells from children with CHD as well in DS/CHD, suggesting an important role in congenital defects in DS.

Keywords: Kruppel-like factors; KLF2; KLF10; Down syndrome; Congenital heart defects.

1. INTRODUCTION

Down Syndrome (DS) is one of the most common genetic disorders, with a frequency of one in 700 births in worldwide and more than 40% of the birth with DS presenting some congenital heart defects (CHD) [1]. It is the most common cardiovascular disease within several congenital anomalies, together with Down syndrome and neural tube defects [2]. Newborns diagnosed with DS show 43% atrioventricular septal defect (AVSD), followed by 32% ventricular septal defects (VSD), 19% atrial septal defects (ASD) and 6% Tetralogy of Fallot (ToF) [3].

Cardiac anomalies could occur by many risk factors, but the congenital anomalies lead to 303,000 newborns death in up to four weeks after the birth all year in worldwide [2]. Many types of genes were reported associated to CHD or genetic syndromes and some transcription factors were related to DS, like GATA1, SIM2, and ZNF295 [4,5].

In humans, Kruppel-like factors (KLF) Family contains 17 members involved in the regulation of many physiological and pathological process, being expressed in different stages of development in many tissues, activating or repressing target genes [6]. KLF2-5, KLF10, KLF13, and KLF15 were related to the cardiac development and diseases [7,8]. KLF2 is a transcriptional activator that recognizes and binds to GC-rich regions of target genes [6], being also reported as a regulator required for cardiac development, through interaction with GATA4 [8]. By another side, KLF10 is a transcriptional repressor that acts together with the corepressor Sin3A [6]. Reported as critical to cardiac development, KLF10 leads to cardiac diseases, like cardiac hypertrophy and hypertrophic cardiomyopathy [9], but its relationship with CHD is still unknown.

The aim of this study was to evaluate the expression profile of KLF2 and KLF10 in DS patients to understand their role in the CHD development, once these transcription factors could change the balance of transcriptional networks in this syndrome.

2. MATERIAL AND METHODS

2.1. Patients and samples

Samples were collected from 42 children in Pernambuco and Paraiba states from Brazil, aged from 0 to 12 years. Peripheral blood was collected from children in a 4ml K2-EDTA tubes containing 2mL of RNAlater for preserving the RNA content in the sample. Tubes were mixed gently, split in a smaller volume and maintained at -80°C until molecular processing. The patients were divided into four study groups: children with DS (n=12); children with CHD (n=5); children with DS and CHD (n=21); and healthy children (n=4). Parents of all children signed the inform consent of the approved project (UFPE/CCS No.1172909). All procedures were conducted according to the guidelines of the Declaration of Helsinki (1964).

2.2. RNA processing

Vials containing 600µl de peripheral blood with RNAlater were thawed once and used for RNA purification with QIAamp RNA Blood Mini Kit® (Qiagen, USA), following the manufacturer's protocol. RNA concentration was determined using NanoDrop® – 2000 Spectrophotometer (Thermo Fisher Scientific, USA). RNA purified was converted to complementary DNA (cDNA) with QuantiNova Reverse Transcription Kit® (Qiagen, USA), following manufacturer's instructions. RNA and cDNA concentration were determined using NanoDrop® – 2000 Spectrophotometer (Thermo Fisher Scientific, USA) and maintained at -80°C until further analysis.

2.3. KLF expression

For assessing KLF2 and KLF10 expression levels, KLF2 primers (Fw: 5'-GCG TGA GCT CGT CTG AG-3' and Rv: 5'-GCA AGA CCT ACA CCA AGA CTT-3') and KLF10

primers (Fw: 5'-CAA ATT TCT TCT CAC CCG TGT-3' and Rv: 5'-CTT TAA AAG TTC CCA TCT GAA GGC-3') were tested for qPCR experiments, which were performed in RotorGene Q® (Qiagen, USA) using QuantiNova Sybr Green Mastermix® (Qiagen, USA) following manufacturer's protocol. Melting curve (T_m) was generated and analyzed to determine the quality of the reaction. The samples were studied in duplicates using RPLP0 as a reference gene, and relative expression was calculated using the $2^{-\Delta\Delta C_t}$ method.

2.4. Statistical analysis

GraphPad Prism (V. 7.0 – Software, CA) was used for statistical analysis. The data are expressed as median and comparisons of parameters were performed using the non-parametric test Mann–Whitney and Wilcoxon. Were considered *p*-value <0.05 as statistical significance.

3. RESULTS AND DISCUSSION

3.1. Congenital heart defects

A total of 42 children samples were evaluated, being 78.57% (33/42) affected by DS. In this group, 63.63% was affected by at least one type of CHD, corroborating with the majority of the studies worldwide [1,3]. AVSD was the most common type of CHD in DS group (42.85%), followed by ASD (19.04%), VSD (19.04%) and patent ductus arteriosus (PDA) (19.04%) (Table 1). No child without DS was affected by AVSD. Our results showed higher prevalence than in Paraná-Brazil that can vary from 17.1% to 27.75% [10, 11]. However, was similar to the prevalence of AVSD worldwide (43%) [3]. Some children showed more than one type of CHD: one DS child affected by AVSD and ToF; two DS non-AVSD children affected by Patent ductus arteriosus together with other CHD; and one CHD non-DS group affected by VSD, pulmonary atresia, and systemic-pulmonary shunt.

3.2. KLF2 and KLF10 expression

KLF2 and KLF10 mRNA levels showed different profile between the four groups analyzed. KLF2 was expressed in all groups, while KLF10 expression was not observed in healthy children. DS group showed higher mRNA level of KLF2 compared to KLF10 ($p=0.0020$) (Fig. 1A), as well in children with DS/CHD ($p<0.0001$) (Fig. 1B).

KLF2 is highly expressed in endothelial cells during the cardiac development, in the atrioventricular canal [14]. In animals model, KLF2 regulates the expression of vasoprotective genes and inhibit the expression of pro-inflammatory genes preventing atherosclerosis [13]. The comparison of KLF2 mRNA levels among the four groups showed that DS children have a similar expression to DS/CHD children with AVSD. Otherwise, CHD and DS/CHD non-AVSD showed a similar profile of low expression, comparable to healthy children. KLF2 relative expression was downregulated in all groups (Fig. 1C), as observed in dendritic cells of patients with atherosclerosis [12]. The KLF2^{-/-} mice embryos showed septation abnormalities, with significant downregulation of transcription factors related to cardiac development, Tbx5, GATA4, and Sox9, indicating that these genes are positively regulated by KLF2 [8]. Therefore, the downregulation of KLF2 could be related to septation defects in CHD group studied. Besides, the decrease in the KLF2 mRNA levels could also play a role in atherosclerosis development in adult life of DS patients.

KLF10 expression showed significance between DS and DS/CHD non-AVSD children group ($p=0.0111$), but was absent in blood of healthy children (Fig. 1D). KLF10 was reported to have a low expression in the heart tissue, but the functional role in the myocardium remained poorly understood during years. KLF10 null mice develops cardiac hypertrophy, and KLF10 mutations lead to hypertrophic cardiomyopathy [9,15]. Our results suggest that KLF10 could be related to the phenotype of DS, regulating important genes involved in the development of CHD others than AVSD-type. However, it is important to

note that KLF10 expression in blood cells can be related to other processes than heart tissue development, indicating greater impact of this gene in cardiac diseases.

4. CONCLUSION

The transcriptional factors KLF2 and KLF10 seem to be involved in CHD development in DS children, and could be potential markers for cardiac diseases. The finding of mRNA levels in blood samples could indicate these genes can be involved in other processes than heart tissue development. Our analysis highlights the importance of studying these genes, once many questions remain to be elucidated. Further studies with a larger cohort and a follow-up evaluation can contribute with new insights for understanding congenital defects and cardiac conditions associated to Down syndrome.

Study limitations

An important limitation was the small number of healthy children and also in the CHD group. Although KLF2 and KLF10 are expressed in leukocytes, analysis of the expression in heart tissue can be required for direct association with CHD.

Acknowledgments: The authors are grateful to the children and relatives that agreed in participating in this study.

Funding: This work was supported by the *Coordination for the Improvement of Higher Education Personnel* (CAPES). This research did not receive another particular grant from funding agencies in the public, commercial, or not-for-profit sectors.

5. Reference

- [1] C. Stoll, B. Dott, Y. Alembik, M.P. Roth, Associated congenital anomalies among cases with Down syndrome, *Eur. J. Med. Genet.* 58 (2015) 674–680.
doi:10.1016/j.ejmg.2015.11.003.
- [2] WHO | Congenital anomalies, WHO. (2016).
[http://www.who.int/mediacentre/factsheets/fs370/en/#.WNPZ11PrHJo.mendeley&title=Congenital anomalies&description=WHO fact sheet on congenital anomalies, an important cause of childhood death, chronic illness, and disability in many countries.](http://www.who.int/mediacentre/factsheets/fs370/en/#.WNPZ11PrHJo.mendeley&title=Congenital+anomalies&description=WHO+fact+sheet+on+congenital+anomalies,+an+important+cause+of+childhood+death,+chronic+illness,+and+disability+in+many+countries.+(accessed+March+23,+2017).)
(accessed March 23, 2017).
- [3] S.B. Freeman, L.H. Bean, E.G. Allen, S.W. Tinker, A.E. Locke, C. Druschel, C.A. Hobbs, P.A. Romitti, M.H. Royle, C.P. Torfs, K.J. Dooley, S.L. Sherman, Ethnicity, sex, and the incidence of congenital heart defects: a report from the National Down Syndrome Project, *Genet. Med.* 10 (2008) 173–180.
doi:10.1097/GIM.0b013e3181634867.
- [4] I. Roberts, S. Izraeli, Haematopoietic development and leukaemia in Down syndrome, *Br. J. Haematol.* 167 (2014) 587–599. doi:10.1111/bjh.13096.
- [5] C. Spellman, M.M. Ahmed, D. Dubach, K.J. Gardiner, Expression of trisomic proteins in Down syndrome model systems, *Gene.* 512 (2013) 219–225.
doi:10.1016/j.gene.2012.10.051.
- [6] B.B. McConnell, V.W. Yang, Mammalian Krueppel-Like Factors in Health and Diseases, *Physiol. Rev.* 90 (2010) 1337–1381.
doi:10.1152/physrev.00058.2009.Mammalian.
- [7] D.A. Prosdocimo, M.K. Sabeh, M.K. Jain, Kruppel-like factors in muscle health and disease, *Trends Cardiovasc. Med.* 25 (2015) 278–287. doi:10.1016/j.tcm.2014.11.006.
- [8] A.R. Chiplunkar, T.K. Lung, Y. Alhashem, B.A. Koppenhaver, F.N. Salloum, R.C.

- Kukreja, J.L. Haar, J.A. Lloyd, Kruppel-Like Factor 2 Is Required for Normal Mouse Cardiac Development, *PLoS One*. 8 (2013). doi:10.1371/journal.pone.0054891.
- [9] J.M. Bos, M. Subramaniam, J.R. Hawse, I. Christiaans, N.M. Rajamannan, J.J. Maleszewski, W.D. Edwards, A.A.M. Wilde, T.C. Spelsberg, M.J. Ackerman, TGF β -inducible early gene-1 (TIEG1) mutations in hypertrophic cardiomyopathy, *J. Cell. Biochem.* 113 (2012) 1896–1903. doi:10.1002/jcb.24058.
- [10] T. Cristina, K. Doná, B. Lawin, C. Simone, J. Marques, Características e Prevalência de Cardiopatias Congênitas em Crianças com Síndrome de Down Submetidas à Cirurgia Cardíaca em um Hospital na Região Norte do Paraná Characteristics and Prevalence of Congenital Heart Disease in Children with Down Syndrome Und, (2009) 11–16.
- [11] B.E.B.V. Bermudez, S.L. Medeiros, M.B. Bermudez, I.M. Novadzki, N.I.R. Magdalena, Down syndrome: Prevalence and distribution of congenital heart disease in Brazil, *Sao Paulo Med. J.* 133 (2015) 521–524. doi:10.1590/1516-3180.2015.00710108.
- [12] H. Fang, J. Lin, L. Wang, P. Xie, X. Wang, J. Fu, W. Ai, S. Chen, F. Chen, F. Zhang, Y. Su, D. Li, Kruppel-like factor 2 regulates dendritic cell activation in patients with acute coronary syndrome, *Cell. Physiol. Biochem.* 32 (2013) 931–941. doi:10.1159/000354496.
- [13] G.B. Atkins, Y. Wang, G.H. Mahabeleshwar, H. Shi, H. Gao, D. Kawanami, V. Natesan, Z. Lin, D.I. Simon, M.K. Jain, Hemizygous deficiency of krüppel-like factor 2 augments experimental atherosclerosis, *Circ. Res.* 103 (2008) 690–693. doi:10.1161/CIRCRESAHA.108.184663.
- [14] J.S. Lee, Q. Yu, J.T. Shin, E. Sebzda, C. Bertozzi, M. Chen, P. Mericko, M. Stadtfeld, D. Zhou, L. Cheng, T. Graf, C.A. MacRae, J.J. Lepore, C.W. Lo, M.L. Kahn, Klf2 Is

- an Essential Regulator of Vascular Hemodynamic Forces In Vivo, *Dev. Cell.* 11 (2006) 845–857. doi:10.1016/j.devcel.2006.09.006.
- [15] N.M. Rajamannan, M. Subramaniam, T.P. Abraham, V.C. Vasile, M.J. Ackerman, D.G. Monroe, T.L. Chew, T.C. Spelsberg, TGF β inducible early gene-1 (TIEG1) and cardiac hypertrophy: Discovery and characterization of a novel signaling pathway, *J. Cell. Biochem.* 100 (2007) 315–325. doi:10.1002/jcb.21049.

Tables

Congenital heart defect	Amount of patients
Atrioventricular septal defect	9
Atrial septal defect	4
Ventricular septal defect	4
Patent ductus arteriosus	4
Transposition of the great arteries	2*
Pulmonary atresia	1
Aortic obstruction	1
Systemic-pulmonary shunt	1
Tetralogy of Fallot	1
Pulmonary valve stenosis	1*
Truncus arteriosus	1*

Table 1. Type of congenital heart defects found in the patients with and without DS. *patients without DS.

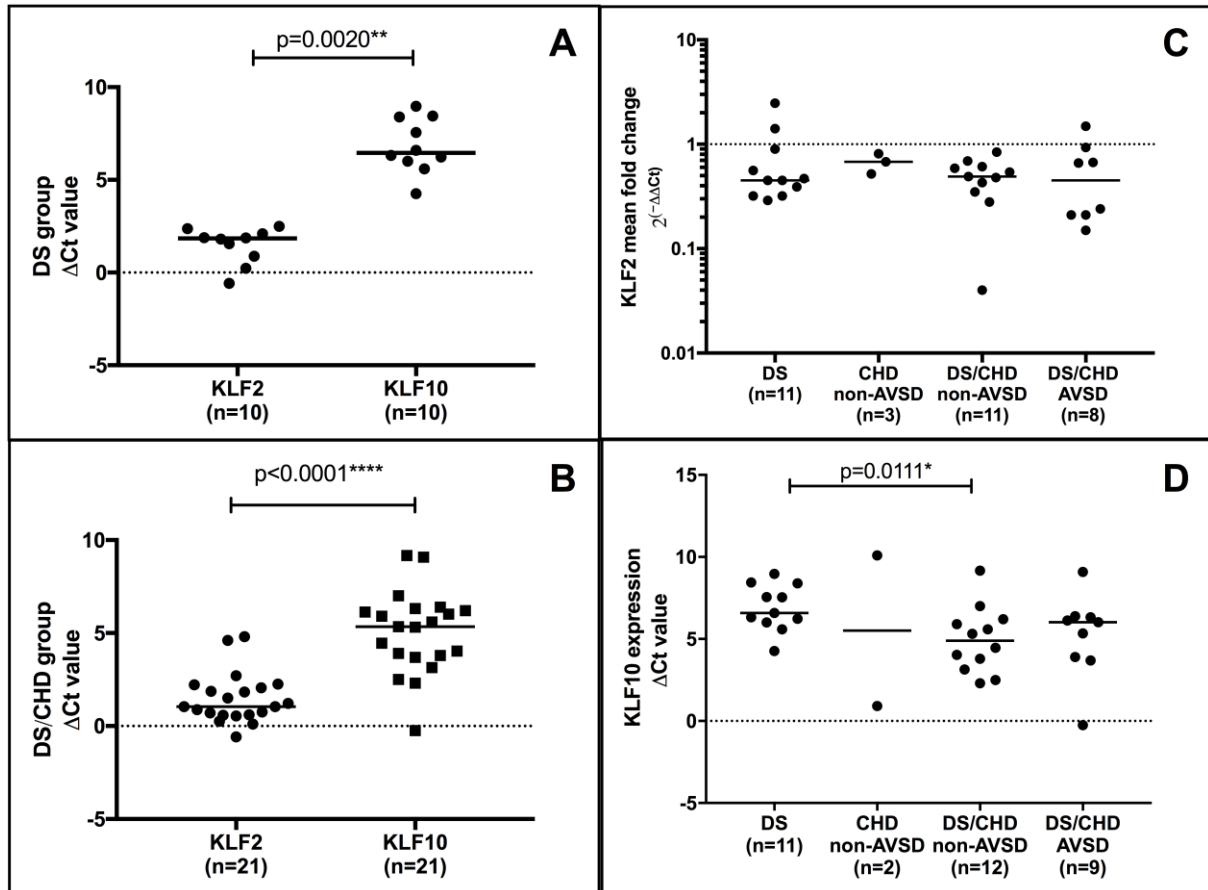


Figure 1. KLF2 and KLF10 mRNA levels and expression profile in DS children without and with CHD. **A)** mRNA levels in DS children without CHD (* $p=0.0020$). **B)** mRNA levels in DS children with CHD. *** $p<0.0001$. **C)** Relative expression of KLF2 in CHD children with and without DS. **D)** KLF10 mRNA levels in children with and without DS. DS children with CHD non-AVSD present higher levels of KLF10 mRNA. * $p=0.0111$.

6. CONSIDERAÇÕES FINAIS

Nossos resultados reforçam a importância da aplicação de ferramentas computacionais com diferentes algoritmos para predição de potenciais mutações a serem estudadas e validadas experimentalmente *in vitro* e *in vivo*. Em nosso estudo, as mutações das KLF e os estudos de expressão da KLF2 e KLF10 em pacientes com síndrome de Down demonstraram um grande potencial de estarem relacionadas ao desenvolvimento de doenças cardíacas e defeitos cardíacos congênitos. Contudo, estudos de validação das mutações KLF potencialmente deletérias e a análise de uma maior coorte são necessários para elucidar questões ainda não esclarecidas.

**ANEXO A - COMITÊ DE ÉTICA EM PESQUISA ENVOLVENDO SERES
HUMANOS**

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Em busca da origem das cardiopatias congênitas em portadores de síndrome de Down.

Pesquisador: Sandra da Silva Mattos

Área Temática: Genética Humana:

(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP;);

Versão: 1

CAAE: 47407615.4.0000.5208

Instituição Proponente: CIRCULO DO CORACAO DE PERNAMBUCO

Patrocinador Principal: CIRCULO DO CORACAO DE PERNAMBUCO

DADOS DO PARECER

Número do Parecer: 1.172.909

Data da Relatoria: 05/08/2015

Apresentação do Projeto:

Trata-se de projeto de pesquisa com desenho de caso-controle envolvendo 200 indivíduos com Síndrome de Down para avaliar a associação entre alterações gênicas e cardiopatias.

Objetivo da Pesquisa:

Verificar a associação entre alterações na expressão gênica e presença de polimorfismos na origem das cardiopatias congênitas em portadores de síndrome de Down.

Avaliação dos Riscos e Benefícios:

Riscos de uma coleta sanguínea. Benefícios: Os pacientes e seus responsáveis terão maiores informações sobre a genética que levou ao defeito cardíaco, fato este que pode até ser usado como fator prognóstico.

Comentários e Considerações sobre a Pesquisa:

Trata-se de pesquisa envolvendo 200 portadores de S. de Down divididos em 2 grupos, com e sem anormalidades cardíacas detectadas por ecocardiograma. Não consta a proporção de indivíduos em cada grupo. Serão submetidos a coleta de 20ml de sangue para análise genética que está detalhada no protocolo. Não está descrito onde os pacientes serão recrutados.

Endereço: Av. da Engenharia s/nº - 1º andar, sala 4, Prédio do CCS

Bairro: Cidade Universitária

CEP: 50.740-600

UF: PE

Município: RECIFE

Telefone: (81)2126-8588

E-mail: cepccs@ufpe.br

Continuação do Parecer: 1.172.909

Considerações sobre os Termos de apresentação obrigatória:

Constam: Termo de assentimento (TALE), TCLE para maiores de 18 anos, TCLE para analfabetos e TCLE para pais ou responsáveis, o projeto e suas informações básicas, Folha de rosto, termo de confidencialidade, autorização de uso de dados, carta de anuência (2).

Recomendações:

Adequar a linguagem dos TALE e TCLEs ao nível de compreensão dos participantes, especialmente no 1º parágrafo de informações sobre a pesquisa.

Conclusões ou Pendências e Lista de Inadequações:

Adequar a linguagem do TALE e TCLEs ao nível de compreensão dos participantes, especialmente no 1º parágrafo de informações sobre a pesquisa. Por exemplo, em vez de "Descrição da pesquisa: o objetivo desta pesquisa é analisar a expressão gênica e pesquisar polimorfismos e padrão de metilação em genes localizados na região entre Tiam e Kcnj6 e correlaciona-los com a presença de cardiopatia congênita em portadores de síndrome de Down. Para tal, será obtida uma amostra sanguínea dos participantes para análise, assim como pesquisa do seu histórico médico." Sugiro : O objetivo desta pesquisa é avaliar se alterações ou diferenças em alguns genes estão associados a doenças congênitas (de nascença) do coração em pessoas com Síndrome de Down.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

O Protocolo foi avaliado na reunião do CEP e está APROVADO para iniciar a coleta de dados. Informamos que a APROVAÇÃO DEFINITIVA do projeto só será dada após o envio do Relatório Final da pesquisa. O pesquisador deverá fazer o download do modelo de Relatório Final para enviá-lo via "Notificação", pela Plataforma Brasil. Siga as instruções do link "Para enviar Relatório Final", disponível no site do CEP/CCS/UFPE. Após apreciação desse relatório, o CEP emitirá novo Parecer Consubstanciado definitivo pelo sistema Plataforma Brasil.

Informamos, ainda, que o (a) pesquisador (a) deve desenvolver a pesquisa conforme delineada neste protocolo aprovado, exceto quando perceber risco ou dano não previsto ao voluntário participante (item V.3., da Resolução CNS/MS Nº 466/12).

Eventuais modificações nesta pesquisa devem ser solicitadas através de EMENDA ao projeto, identificando a parte do protocolo a ser modificada e suas justificativas.

Endereço: Av. da Engenharia s/nº - 1º andar, sala 4, Prédio do CCS

Bairro: Cidade Universitária

CEP: 50.740-600

UF: PE

Município: RECIFE

Telefone: (81)2126-8588

E-mail: cepccs@ufpe.br

Continuação do Parecer: 1.172.909

Para projetos com mais de um ano de execução, é obrigatório que o pesquisador responsável pelo Protocolo de Pesquisa apresente a este Comitê de Ética relatórios parciais das atividades desenvolvidas no período de 12 meses a contar da data de sua aprovação (item X.1.3.b., da Resolução CNS/MS Nº 466/12). O CEP/CCS/UFPE deve ser informado de todos os efeitos adversos ou fatos relevantes que alterem o curso normal do estudo (item V.5., da Resolução CNS/MS Nº 466/12). É papel do/a pesquisador/a assegurar todas as medidas imediatas e adequadas frente a evento adverso grave ocorrido (mesmo que tenha sido em outro centro) e ainda, enviar notificação à ANVISA – Agência Nacional de Vigilância Sanitária, junto com seu posicionamento.

RECIFE, 06 de Agosto de 2015

Assinado por:
LUCIANO TAVARES MONTENEGRO
(Coordenador)

Endereço: Av. da Engenharia s/nº - 1º andar, sala 4, Prédio do CCS

Bairro: Cidade Universitária

CEP: 50.740-600

UF: PE

Município: RECIFE

Telefone: (81)2126-8588

E-mail: cepccs@ufpe.br