REVIEW ARTICLE

Big Data and Genome Editing Technology: A New Paradigm of Cardiovascular Genomics

Chayakrit Krittanawong^{1,3,*}, Tao Sun² and Eyal Herzog³

¹Department of Internal Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ²Department of Cardiovascular Disease, Mayo Clinic, Rochester, Minnesota, MN, USA; ³Department of Cardiovascular Disease, Icahn School of Medicine at Mount Sinai St' Luke, Mount Sinai Heart, New York, NY, USA

ARTICLE HISTORY

Received: April 24, 2017
Revised: July 25, 2017
Accepted: July 28, 2017

DOI:
10.2174/1573403X13666170804152432

Abstract: Opinion Statements: Cardiovascular diseases (CVDs) encompass a range of conditions extending from congenital heart disease to acute coronary syndrome most of which are heterogeneous in nature and some of them are multiple genetic loci. However, the pathogenesis of most CVDs remains incompletely understood. The advance in genome-editing technologies, an engineering process of DNA sequences at precise genomic locations, has enabled a new paradigm that human genome can be precisely modified to achieve a therapeutic effect. Genome-editing includes the correction of genetic variants that cause disease, the addition of therapeutic genes to specific sites in the genomic locations, and the removal of deleterious genes or genome sequences. Site-specific genome engineering can be used as nucleases (known as molecular scissors) including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated 9 (Cas9) systems to provide remarkable opportunities for developing novel therapies in cardiovascular clinical care. Here we discuss genetic polymorphisms and mechanistic insights in CVDs with an emphasis on the impact of genome-editing technologies. The current challenges and future prospects for genome-editing technologies in cardiovascular medicine are also discussed.

Keywords: Genome editing technology, genome, genome engineering, GWAS, genome-wide association studies.

1. INTRODUCTION

1.1. Genetics and Cardiovascular Diseases

The large majority of CVDs are complex, heterogeneous, and polygenic [1]. In general, genetic mutations can be classified as gain- or loss-of-function of protein, leading to mechanistic insights of CVD. To date, with advancements in genomics, bioinformatics and technologies, determining whether genetic polymorphisms are truly pathogenic and capable of causing CVD is feasible. Studies over the past decade have found the relationship between genetic polymorphisms and CVD (e.g., hypertrophic cardiomyopathy, dilated cardiomyopathy, arrhythmogenic cardiomyopathy, brugada syndrome, and familial hyperlipidemia). To date, studies found that genetic polymorphisms are associated with subclinical disease such CV risk factors in coronary artery disease (CAD) [2, 3].

Recent meta-analysis showed that ABO and ADAMTS7 genes were associated with angiographically confirmed

coronary atherosclerosis, while CNNM2 and APOA5 genes were associated with hypertension and hypertriglyceridemia, respectively [4].

In 1990, Rigat *et al.* [5] identified the polymorphism of the angiotensin-converting enzyme (ACE) gene based on the presence or absence of a 287-bp element on intron 16 on chromosome 17. In fact, ACE gene insertion/deletion (I/D) polymorphisms can influence the variability in systemic ACE levels [6]. To date, studies found that I/D and D/D polymorphism in the ACE gene was associated with CAD [7] and the development of LVH in patients with hypertension [8, 9]. In addition, meta analysis [10] showed that the D allele of ACE gene was significantly associated with an increased risk of CVD (CHD, CAD, and MI). Meiling *et al.* [11] found that ACE gene I/D polymorphisms and the CHD in Hainan Li and Han nationality *via* a mechanism of higher TG level and the lower HDL-C level. Interestingly, recent study suggested that the D allele of the ACE gene may increase the risk of developing heart failure with a preserved ejection fraction (HFpEF) *via* a mechanism of LVH in patients with hypertension [12]. However, The I/D polymorphism of ACE gene may be in linkage disequilibrium with other SNPs or allelic variant. Therefore, the analysis of multiple genetic markers in the context of linkage disequilibrium

*Address correspondence to this author at the Department of Internal Medicine, Icahn School of Medicine at Mount Sinai, 1000 10th Ave, New York, NY 10019, USA; Fax: 212-523-8605; Tel: 212-523-4000; E-mail: Chayakrit.Krittanawong@mountsinai.org

with other SNPs or allelic variant is required to increase the probability of providing clinically useful information.

Minor allele carriers of several genetic polymorphisms in TGFB1, MMP3, GJA4, APOE ϵ 4 allele, R92H allele in PLA2G7 gene were associated with an increased risk of developing CHD [13, 14]. However, further studies comparing the effect of those genes in CHD are needed in order to use predominant genes in genome editing. For example, El-Lebedy *et al.* [15] found that apoE gene polymorphisms associated with CVD and identified apoE as an independent risk factor for both T2DM and CVD. Therefore, APOE gene polymorphisms should be targeted for genome editing to reduce incident T2DM and CVD.

In addition, Panahloo *et al.* [16] showed that T2DM subjects with the DD polymorphism in the ACE gene had increased insulin sensitivity, leading to lower concentrations of insulin level. Polymorphisms in CDKN2A/2B and FTO may be associated with T2DM in Chinese populations [17], while CDKAL1 gene rs7756992 A/G polymorphism was significantly associated with T2DM susceptibility in the Caucasian [18]. The alpha subunit of the type V voltage-gated sodium channel (SCN5A) mutations was associated with an inherited cardiac arrhythmia [19], dilated cardiomyopathy [20], and long QT syndrome [21, 22].

Over the last ten years, GWAS have evolved as a powerful tool for investigating the genetic architecture of human disease. To date, GWAS was achieved by genotyping families affected by CVDs using a collection of genetic markers across the genome using linkage analysis technique to examine how those genetic markers segregate with the disease across multiple families.

A recent meta-analysis of GWAS involving more than 30,000 case and control subjects showed that CAD was associated with LIPA, PDGFD, ADAMTS7-MORF4L1, and KIAA1462 in multiple ethnic groups [3]. A GWAS found that idiopathic dilated cardiomyopathy was associated with genetic polymorphisms in the HSPB7 gene [23]. Moreover, GWAS have also identified genetic polymorphisms for arrhythmias, including atrial fibrillation [24], ventricular fibrillation [25], sudden cardiac death [26], and the sick sinus syndrome [27].

In fact, the analysis of GWAS is a very specific, clinically relevant disease phenotype. GWAS might need to be specifically evaluated for genome editing. For example, the analysis of GWAS to identify predominantly non-coding sequences which could ultimately turn out to be in key regulatory regions of the genome, such as enhancers, which help turn genes on and off.

1.2. Genome Editing Based Therapy

Genome-editing technology is rapidly growing and being applied into cardiovascular medicine and research to facilitate a greater understanding the pathogenesis of CVD (*i.e.* lipid metabolism, electrophysiology, cardiomyopathies, HFpEF) to open the door to novel therapies.

Genome-editing based therapy includes correction or inactivation of deleterious mutations that cause CVDs, the addition of therapeutic genes to specific sites in the genome,

the removal of deleterious genes or disruption of specific genome sequences. Genome editing can be used for therapeutics in monogenic CVD (*i.e.* familial hypercholesterolemia, sudden cardiac death, long QT syndrome, Marfan syndrome) or prevention of polygenic CVDs (*i.e.* CAD, hypertension or HFpEF). To date, the developments in site-specific genome engineering using 3 major classes of nucleases (molecular scissors) include zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated 9 (Cas9) systems have been implemented in cardiovascular care to enable site-specific genome-editing based therapy. Briefly, these molecular scissors can be custom-engineered to create double-stranded DNA (dsDNA) breaks at a targeted sequence in specific genomic location. Next, dsDNA break can be repaired by either the non-homologous end joining pathway or the homologous recombination pathway. Using molecular scissors to modify these polymorphisms is promising and can be tested in animal models before testing in human. This genome-editing approach may have therapeutic potential for the prevention of CVD (Table 1). Using molecular scissors in I/D and D/D polymorphisms in the ACE gene could potentially have important prevention in insulin insensitivity, HFpEF and CAD. For example, genetic mutations in the cases of the D allele of the ACE gene for HFpEF may be corrected by using gene disruption method to specifically inactivate the pathogenic allelic variants. PCSK9 plays a major role in the LDL receptor pathway and LDL-C metabolism. To date, CRISPR-Cas9 genome editing approach, a widespread popularity in cardiovascular genomics and research, has been implemented in mice. Recent study showed that genome editing with the CRISPR-Cas9 system disrupts the PCSK9 gene in vivo with high efficiency and reduces LDL-C in mice [28].

In addition, Claussnitzer *et al.* [29] performed a combination of GWAS and CRISPR-Cas9 genome editing in ARID5B, IRX3, IRX5 gene variants, and rs1421085 SNP in mice and found that those variants were associated with pro-obesity and anti-obesity effects via a mechanism of calorie storage in white fat and brown fat. Additional studies in human are needed to confirm the potential of the CRISPR-Cas9 system for manipulating the PCSK9 gene and obesity related genes. Liang *et al.* [30] performed CRISPR/Cas9-mediated genome editing technology in 2 patients with type 1 BrS carrying 2 different SCN5A variants compared to 2 healthy control subjects. They found that correcting the SCN5A variant (rs397514446) using CRISPR/Cas9-mediated genome editing lead to restoration of electrical properties, including normalization of beat-to-beat interval, a fast AP upstroke velocity, and Ca²⁺ transients that are as robust as those from the control myocytes. However, further studies using pluripotent stem cell-derived cardiomyocytes may provide insight into the cellular mechanisms of BrS and accelerate discovery of new therapeutic modalities.

However, each of ZFNs, TALENs, and CRISPR/Cas9 system has different roles and disadvantages. ZFNs are small in size, but they are difficult to design to bind a desired sequence, their target-site selection is limited, and there is often no potential targetable site in the genomic region of interest. TALENs are easy to design, and the capacity to easily

Table 1. A table summarizes the key genes involved in the CVDs.

Nucleases	Target Genes	Disorders
ZFN	ACE	Insulin insensitivity, HFpEF and CAD
CRISPR/Cas9	PCSK9, SORT1, ABCG8, SH2B3, LDLR	LDL-C metabolism
CRISPR/Cas9	ARID5B, IRX3, IRX5 variants, and rs1421085 SNP	Pro-obesity and anti-obesity effects
CRISPR/Cas9	SCN5A variant (rs397514446)	Electrical properties in cardiac myocytes
Future Applications	FCN1	Marfan syndrome
Future Applications	APOA1, APOA5, APOC3	HDL-C metabolism
Future Applications	LIPA	Endothelial function in ACS
Future Applications	HLA-C	Triglycerides metabolism
Future Applications	PTGS1	Prostaglandins metabolism in NSAIDs users
Future Applications	ABO	IL-6, E-selectin in ACS
Future Applications	CACNA2D2	Voltage-dependent calcium channel auxiliary subunit in HTN
Future Applications	PDE5	Phosphodiesterase 5A in HTN
Future Applications	CBLN2 rs2217560	Pulmonary arterial hypertension
Future Applications	HGC22, BAG3	Dilated cardiomyopathy
Future Applications	CDKN2A, CDKN2B	Cyclin-dependent kinase inhibitor 2A and 2B in atherosclerosis

generate longer DNA-binding domains (by simply adding extra repeats) allows for greater target-site specificity; however, TALENs are much larger than ZFNs, which complicates their delivery into cells. The CRISPR/Cas9 system has recently emerged as a potentially facile and efficient alternative to ZFNs and TALENs for inducing targeted genetic alterations. The CRISPR/Cas9 system is simple to design, cost-effectiveness, time-efficient, easy construction and lower toxicity in human cells and it is easy to target multiple genomic locations simultaneously by using multiple guide RNAs, not protein/DNA recognition [31]. In addition, guide RNAs can be designed easily and cheaply to target nearly any sequence in the genome specifically. To date, several biotech companies and academic institutions announced the launch of their first clinical trials using CRISPR-Cas. barriers. He *et al.* [32] compared TALENs with CRISPR/Cas9 and found that CRISPR/Cas9 is more efficient and precise than TALEN in context of induced targeted genomic deletion, but TALENs is more efficient at stimulating homology directed repair (HDR) more efficiently than CRISPR/Cas9 and caused fewer targeted genomic deletions, compared to CRISPR/Cas9. Several clinical trials have been using genome editing in various fields such as HIV infection and hematocarcinology. ZFNs, TALENs, and CRISPR/Cas9 system have the potential to revolutionize cardiovascular research and impact personalized medicine. However, CVDs are mainly heterogeneous with environmental factors involvement. Genome editing may be challenging biologically (notably the postmitotic nature of cardiomyocytes) and technically. In conclusion, in cardiovascular genomics, understanding genetic mechanisms using big data analytics, GWAS and applying genome-editing technologies to improve therapeutics towards personalized medicine. Genome-editing tech-

nologies have enabled a new paradigm in which the sequence of the human genome can be precisely manipulated to achieve a therapeutic effect. In fact, the genome-editing field has already established itself as a powerful tool for the generation of new cellular and animal models to investigate pathophysiological mechanisms in medicine, particularly in infectious diseases and oncology. The genome-editing field in cardiovascular medicine is growing and we can undoubtedly expect remarkable opportunities for novel therapies for CVDs. However, potential barriers in genome editing technologies include biomolecular complexity, shortage of staffs, technical difficulty, time consuming and cost-effectiveness. In the future, genome editing based therapy can be used in T2DM, CAD, heart failure, metabolic syndrome and familial hypercholesterolemia and allow us to study the molecular mechanisms underlying the pathology of genetically based CVD. However, much work remains in addressing the current shortcomings of genome-editing technology.

CONCLUSION

In the past decades, numerous genetic polymorphisms have been implicated in the pathogenesis of cardiovascular diseases (CVDs). Large amount of data have increasingly been promoted as a revolutionary development in the genomic medicine. Progress in genomics and bioinformatics using big data analytics has facilitated the successful implementation of genome-wide association studies (GWAS) towards understanding the genetic basis of CVDs. GWAS have identified a large number of single nucleotide polymorphisms (SNPs) and genetic polymorphisms associated with CVD phenotypes. To date, genome-editing technologies are revolutionizing cardiovascular clinical care. In the future, big

data will enhance genome-editing technologies using high functional computer to analyze GWAS to facilitate personalized medicine.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] Lee DS, Pencina MJ, Benjamin EJ, *et al.* Association of parental heart failure with risk of heart failure in offspring. *N Engl J Med* 2006; 355: 138-47.
- [2] Parikh NI, Hwang SJ, Larson MG, *et al.* Parental occurrence of premature cardiovascular disease predicts increased coronary artery and abdominal aortic calcification in the framingham offspring and third generation cohorts. *Circulation* 2007; 116: 1473-81.
- [3] Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in europeans and south asians identifies five new loci for coronary artery disease. *Nat Genet* 2011; 43: 339-44.
- [4] Schunkert H, König IR, Kathiresan S, *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011; 43: 333-8.
- [5] Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, Soubrier F. An insertion/deletion polymorphism in the angiotensin i-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest* 1990; 86: 1343-6.
- [6] Krittana Wong C, Namath A, Lanfear DE, Tang WH. Practical pharmacogenomic approaches to heart failure therapeutics. *Curr Treat Options Cardiovasc Med* 2016; 18: 60.
- [7] McNamara DM, Holubkov R, Postava L, *et al.* Pharmacogenetic interactions between angiotensin-converting enzyme inhibitor therapy and the angiotensin-converting enzyme deletion polymorphism in patients with congestive heart failure. *J Am Coll Cardiol* 2004; 44: 2019-26.
- [8] Gharavi AG, Lipkowitz MS, Diamond JA, Jhang JS, Phillips RA. Deletion polymorphism of the angiotensin-converting enzyme gene is independently associated with left ventricular mass and geometric remodeling in systemic hypertension. *Am J Cardiol* 1996; 77: 1315-9.
- [9] Celentano A, Mancini FP, Crivaro M, *et al.* Cardiovascular risk factors, angiotensin-converting enzyme gene i/d polymorphism, and left ventricular mass in systemic hypertension. *Am J Cardiol* 1999; 83: 1196-200.
- [10] You FJ, Shen DM. Association between angiotensin-converting enzyme insertion/deletion polymorphisms and the risk of heart disease: An updated meta-analysis. *Genet Mol Res* 2016; 15: 15017194.
- [11] Meiling Y, Yong Z, Jianghua Z, *et al.* A study the relation of ace gene polymorphisms and risk factor with coronary heart disease in hainan li and han nationality. *Heart* 2011; 97: A196.
- [12] Bahramali E, Rajabi M, Jamshidi J. Association of ace gene d polymorphism with left ventricular hypertrophy in patients with diastolic heart failure: A case-control study. *BMJ Open* 2016; 6: e010282.
- [13] Lu Y, Boer JMA, Barsova RM, *et al.* Tgfb1 genetic polymorphisms and coronary heart disease risk: A meta-analysis. *BMC Med Genet* 2012; 13: 39.
- [14] Hirashiki A, Yamada Y, Murase Y, *et al.* Association of gene polymorphisms with coronary artery disease in low- or high-risk subjects defined by conventional risk factors. *J Am Coll Cardiol* 2003; 42: 1429-37.
- [15] El-Lebedy D, Raslan HM, Mohammed AM. Apolipoprotein e gene polymorphism and risk of type 2 diabetes and cardiovascular disease. *Cardiovasc Diabetol* 2016; 15: 12.
- [16] Panahloo A, Andrs C, Mohamed-Ali V, *et al.* The insertion allele of the ace gene i/d polymorphism. *Circulation* 1995; 92: 3390.
- [17] Xiao S, Zeng X, Fan Y, *et al.* Gene polymorphism association with type 2 diabetes and related gene-gene and gene-environment interactions in a uighur population. *Med Sci Monit* 2016; 22: 474-87.
- [18] Li YY, Wang LS, Lu XZ, *et al.* Cdkal1 gene rs7756992 a/g polymorphism and type 2 diabetes mellitus: A meta-analysis of 62,567 subjects. *Sci Rep* 2013; 3: 3131.
- [19] Chen Q, Kirsch GE, Zhang D, *et al.* Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. *Nature* 1998; 392: 293-6.
- [20] McNair WP, Ku L, Taylor MR, *et al.* Scn5a mutation associated with dilated cardiomyopathy, conduction disorder, and arrhythmia. *Circulation* 2004; 110: 2163-7.
- [21] Schott JJ, Alshinawi C, Kyndt F, *et al.* Cardiac conduction defects associate with mutations in scn5a. *Nat Genet* 1999; 23: 20-1.
- [22] Wang Q, Shen J, Splawski I, *et al.* Scn5a mutations associated with an inherited cardiac arrhythmia, long qt syndrome. *Cell* 1995; 80: 805-11.
- [23] Villard E, Perret C, Gary F, *et al.* A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J* 2011; 32: 1065-76.
- [24] Ellinor PT, Lunetta KL, Glazer NL, *et al.* Common variants in kcnh3 are associated with lone atrial fibrillation. *Nat Genet* 2010; 42: 240-4.
- [25] Bezzina CR, Pazoki R, Bardai A, *et al.* Genome-wide association study identifies a susceptibility locus at 21q21 for ventricular fibrillation in acute myocardial infarction. *Nat Genet* 2010; 42: 688-91.
- [26] Arking DE, Junttila MJ, Goyette P, *et al.* Identification of a sudden cardiac death susceptibility locus at 2q24.2 through genome-wide association in european ancestry individuals. *PLoS Genet* 2011; 7: e1002158.
- [27] Holm H, Gudbjartsson DF, Sulem P, *et al.* A rare variant in myh6 is associated with high risk of sick sinus syndrome. *Nat Genet* 2011; 43: 316-20.
- [28] Ding Q, Strong A, Patel KM, *et al.* Permanent alteration of pcsk9 with in vivo crispr-cas9 genome editing. *Circ Res* 2014; 115: 488-92.
- [29] Claussnitzer M, Dankel SN, Kim K-H, *et al.* Fto obesity variant circuitry and adipocyte browning in humans. *N Engl J Med* 2015; 373: 895-907.
- [30] Liang P, Sallam K, Wu H, *et al.* Patient-specific and genome-edited induced pluripotent stem cell-derived cardiomyocytes elucidate single-cell phenotype of brugada syndrome. *J Am Coll Cardiol* 2016; 68: 2086.
- [31] Ding Q, Regan SN, Xia Y, Oostrom LA, Cowan CA, Musunuru K. Enhanced efficiency of human pluripotent stem cell genome editing through replacing talens with crisprs. *Cell Stem Cell* 2013; 12: 393-4.
- [32] He Z, Proudfoot C, Whitelaw CB, Lillico SG. Comparison of crispr/cas9 and talens on editing an integrated egfp gene in the genome of hek293ft cells. *Springerplus* 2016; 5: 814.