

Nonsynonymous Single-Nucleotide Variations as Cardiovascular System Disease Biomarkers and Their Roles in Bridging Genomic and Proteomic Technologies

Ayman Abunimer, Hayley Dingerdissen, John Torcivia-Rodriguez, Phuc VinhNguyen Lam, and Raja Mazumder

Contents

Key Facts of Non-synonymous Single-Nucleotide Variation in Cardiovascular Diseases	823
Key Facts of Genomic Variant Discovery	823
Definitions	824
Introduction	824
Technologies Used in Variant Detection	826
Mapping of Reads (Generating an Alignment)	827
SNV/SNP Calling	828
Identifying nsSNVs	828
From Genomic to Proteomic Identification	829
Potential Applications to Prognosis, Other Diseases, or	
Conditions: Cardiovascular Diseases and Associated nsSNVs	831
Ischemic Stroke	831
Coronary Artery Disease	832
Sudden Cardiac Death	833
Congestive Heart Failure	834
Myocardial Infarction	835
Congenital Heart Defects	835
Hypertension	836
Arrhythmia	837
Cardiomyopathy	838

A. Abunimer • H. Dingerdissen • J. Torcivia-Rodriguez • P.V.N. Lam
 Department of Biochemistry and Molecular Medicine, George Washington University, Washington, DC, USA
 e-mail: ayman.abunimer@gmail.com; hmhamilt@gwmail.gwu.edu; torcivia@gwmail.gwu.edu; phuclam87@gmail.com

R. Mazumder (✉)
 Department of Biochemistry and Molecular Medicine, George Washington University, Washington, DC, USA

McCormick Genomic and Proteomic Center, George Washington University, Washington, DC, USA
 e-mail: mazumder@gwu.edu

The Future of nsSNVs and Cardiovascular Diseases	839
Genomic and Proteomic Projects Worldwide Associated with Cardiac Diseases	839
Workflow and Results	839
Sample Workflow	839
Sample Workflow Results	840
Summary Points	840
References	841

Abstract

Non-synonymous single-nucleotide variations (nsSNVs) are mutations in the coding regions of the genome which ultimately lead to amino acid alterations. nsSNVs represent potential diagnostic or therapeutic targets when associated with susceptibility to specific diseases or conditions. The emergence of next-generation sequencing (NGS) technologies has streamlined the process of identifying nsSNVs and offers an avenue for the robust study of disease genetics. This chapter examines the existing roles of nsSNVs in cardiovascular diseases and highlights their values as biomarkers given the current state of research. NGS technologies hold promise for a future of medicine built on understanding the genome and proteome, and the associations of each with disease susceptibility and progression. The chapter also provides an overview of NGS technologies currently available, as well as a sample workflow for harnessing the bio-informational value of nsSNVs.

Keywords

Non-synonymous single-nucleotide variations • Next-generation sequencing technology • Cardiovascular diseases • Biomarker identification • Amino acid variation • Proteomics applications

Abbreviations

CAD	Coronary artery disease
CHD	Congenital heart defects
DCM	Dilated cardiomyopathy
KDR	Kinase insert domain-containing receptor
LOF	Loss-of-function
LQTS	Long QT syndrome
LVEF	Left ventricular ejection fraction
MI	Myocardial infarction
MS	Mass spectrometry
NGS	Next-generation sequencing
nsSNV	Non-synonymous single-nucleotide variations
ORF	Open reading frame
rsIDs	dbSNP database identifier
SCD	Sudden cardiac death
SIDS	Sudden infant death syndrome
SNP(s)	Single-nucleotide polymorphism(s)

SNV	Single-nucleotide variation
VCF	Variant call format
VEGF	Vascular endothelial growth factor
VGSCs	Voltage-gated sodium channel
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

Key Facts of Non-synonymous Single-Nucleotide Variation in Cardiovascular Diseases

- DNA is composed of two strands of repeating nucleotide bases (adenine, guanine, cytosine, and thymine) that make up a “sequence.”
- Certain coding regions of DNA encode instructions for proteins such that three adjacent nucleotide bases comprise a codon and determine one amino acid, the building block of proteins.
- Non-synonymous single-nucleotide variations (nsSNV) are changes of a single base in the DNA sequence that result in a different amino acid being produced, and therefore a different, sometimes dysfunctional, protein being produced.
- nsSNVs are known to be associated with human disease, including a number of cardiac diseases.
- Cardiovascular diseases are a set of conditions which affect the structure or function of the heart.
- In addition to lifestyle, obesity, diet, and smoking, genetics are an important risk factor in the development of cardiovascular diseases and conditions.
- Genomics is the study of the entire human genome, or the collection of all genes belonging to a single human.
- Genome sequencing is the process by which the specific composition and order of nucleotide bases in an individual’s DNA can be determined.
- A reference genome is an example of a standard genome that is used for comparison purposes – observation of positional differences between an experimentally obtained sequence from an individual and the reference is how nsSNV is discovered.
- Proteomics is the study of the entire human proteome, or the collection of all proteins produced by a single human.

Key Facts of Genomic Variant Discovery

- Next-generation sequencing (NGS) methods are used to discover novel nsSNVs.
- There are many different NGS platforms and new and improved methods are continually being developed.
- The cost of NGS has dropped rapidly from over three billion dollars per human genome to around one thousand dollars currently.

- FASTA and FASTQ file formats are the standards used for recording genome and sequence read information.
- Researchers align short reads to a reference genome in order to generate the genomic sequence of their subject from short reads.
- Coverage depth for a position is the number of short reads resulting from an NGS experiment that cover this position.
- Contigs are continuous regions of the subject's genome that are able to be assembled in the alignment process due to parts of short reads overlapping each other.
- SAM and BAM file formats are the standards used for recording alignment information.
- Single-nucleotide polymorphism (SNP) or single-nucleotide variant (SNV) calling is the process of comparing a given genome (or DNA segment) with a reference genome to determine nucleotide differences.
- The variant call format (VCF) file format is the standard used most often for recording variations (SNVs and larger variations).

Definitions

BAM files Compressed Sequence Alignment/Map (SAM) files.

Biomarker A biological characteristic associated with disease.

Cardiovascular disease Any disease affecting the structure or function of the heart.

Codon Unit of three nucleotides which encodes a specific amino acid based on the nucleotide composition.

FASTA or FASTQ file Next-generation sequencing (NGS) output data of read names and nucleotides, the Q indicates the presence or absence of quality information for each nucleotide read.

nsSNV Variations in coding regions of a genome which result in amino acid substitution.

SAM files Sequence Alignment/Map; Human readable output files of all the read sequences, where they map to the reference genome, and their mapping score.

Introduction

Non-synonymous single-nucleotide variations (nsSNV) are mutations in the exonic or coding regions of the genome which, when transcribed and then translated, lead to substituted amino acids (missense mutations) or truncated proteins

(nonsense mutations). These alterations in the amino acid sequence may influence protein folding, disrupt protein-protein interactions, or even directly modify the active site (Dingerdissen et al. 2013). nsSNVs are not the only type of genetic mutation, but they are particularly valuable biomarkers and, due to their potential effects on protein function, represent a starting point for investigating biochemical pathways. Although this chapter focuses primarily on nsSNVs of the missense type, it is important to note that both missense and nonsense variations cause changes in the protein sequence, with respect to the normally translated protein, and should therefore be detectable by the proteomic technologies discussed below.

Next-generation sequencing (NGS) methods are essential in the search for nsSNVs as biomarkers for all aspects of physiology, including the cardiovascular system. There are several platforms that generate NGS data, and there is the promise of new, so-called ultrarapid technologies like nanopore sequencing (Deamer and Akeson 2000) on the horizon. Major software developments have addressed the complex computational challenges which stemmed from the extra-large scale of genomic data generated by NGS technologies. These tools facilitate the assembly and alignment of NGS data, the subsequent calling of single-nucleotide polymorphisms (SNPs), or the identification of other types of genomic variation in a sample. Determination of biomarkers from the pool of variation requires the integration of additional software developments with statistical analysis and a detailed consideration of disease-related annotations.

While genomic strategies have provided a broad foundation for the cataloging of disease-associated nucleotide variation, newly developed high-throughput proteomic technologies (Branca et al. 2014) can further elucidate biological and physiological understanding of amino acid variation at a molecular resolution (National Research Council 2006). Quantitative and structural proteomic approaches have already been applied to variant-based biomarker discovery in a number of human diseases (Nie et al. 2014; Marrocco et al. 2010) and hold the same promise for cardiovascular biomarker identification.

Diseases of the cardiovascular system affect the structure and/or function of the heart: they include conditions such as heart failure, sudden cardiac death (SCD), and coronary artery disease (CAD). Altogether, cardiovascular diseases are the leading cause of death for both men and women in the United States (Mozaffarian et al. 2015). While the causes and risk factors behind specific conditions are varied and multifaceted, it is agreed that genetics plays an influential role in susceptibility. Consequently, the ability to identify nsSNVs quickly and accurately is valuable toward the further study of the origins and outcomes of these often fatal conditions. As NGS technologies continue to improve, nsSNVs may play an increasingly important role as therapeutic and diagnostic biomarkers in cardiovascular system diseases. This chapter will offer a brief introduction to the roles of nsSNVs across conditions and diseases under the umbrella term of cardiovascular systems diseases.

Technologies Used in Variant Detection

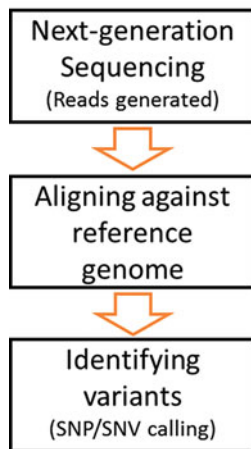
The first full human genome was sequenced by a chain-terminated (Sanger) sequencing method (Sanger et al. 1977) and cost approximately three billion dollars. The high cost and time-intensive nature of sequencing prevented widespread use of the technique until the discovery and development of new massively parallel sequencing methods (Metzker 2010; Grada and Weinbrecht 2013), later termed next-generation sequencing methods. Massively parallel throughput systems take advantage of the speed and efficiency of sequencing genetic fragments in parallel and then reassemble them via computational alignment algorithms. Although initially very expensive, the costs of these sequencing methods have fallen drastically, approaching \$1,000 per sample, bringing the goal of personalized genomic medicine closer than ever before.

NGS methods generally produce large series of short reads, often between 75 and 300 bases in length, depending on the machine used (Metzker 2010). It is not uncommon to produce over one billion short reads in a single experimental run. Although this massive volume of data presents computational challenges, finding efficient solutions is essential as the number of entities, both research and clinical, generating and using NGS data is rapidly expanding (Metzker 2010).

Several major platforms are currently available for next-generation sequencing including Pacific Biosciences, Ion Torrent, Roche/454, Illumina/Solexa, and SOLid, while exciting new techniques such as nanopore technology are undergoing development. This improved technology shows promise in producing very long read lengths (~10 kb or higher) to address current limitations to *de novo* assembly and alignment of sequence reads to a reference genome (Wang et al. 2014).

The basic pipeline of variant detection is shown in Fig. 1. The pipeline becomes increasingly complicated when augmented with additional quality control and analysis steps, but the schematic presented herein represents the core of the variant calling process.

Fig. 1 Basic variant calling pipeline starting from NGS



Mapping of Reads (Generating an Alignment)

A NGS experiment usually produces a FASTQ file which can then be mapped to a reference genome in FASTA format. A FASTA file contains only the read names and the nucleotide sequence of that read with a single file containing records for up to millions of reads. A FASTQ file contains the same ID and read information plus quality information for each nucleotide position as determined by the machine used. This quality information represents the confidence that a particular nucleotide was correctly identified by the sequencing machine.

Since short genomic reads are produced with variable coverage depth at any given position, reads are mapped, or aligned, to a reference genome. Generally, the human reference genome published by the Genome Reference Consortium (<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/>) is used for human samples. After specifying the reference genome, software maps each read to the genome via a computational alignment algorithm that takes a read and determines the most likely coordinates from the genome from which the experimental read was obtained. Coverage is determined for each nucleotide location that has been sequenced and can be matched to a read. Various software packages have been developed for this task, including BLAST (Schuler et al. 1991), TopHat (Trapnell et al. 2009), BWA (Li and Durbin 2009), HIVE-hexagon (Santana-Quintero et al. 2014), and others.

After alignment is complete, the total number of reads that were successfully aligned to a given position make up the coverage depth for that position, with full coverage of the experiment represented by the average coverage across all positions (see Fig. 2). Ideally, the genome will be fully covered such that overlapping reads map from one end of the genome (or chromosome) to the other without any gaps. However, this is frequently not the case, so the alignment software will also report the number of contigs – continuous regions of coverage provided by the reads. Some positions will have greater depth than others as an artifact of sequencing chemistry. This is an important consideration for assessing nsSNV as coverage is inferred to provide direct evidence for the presence of a variant in a sample.

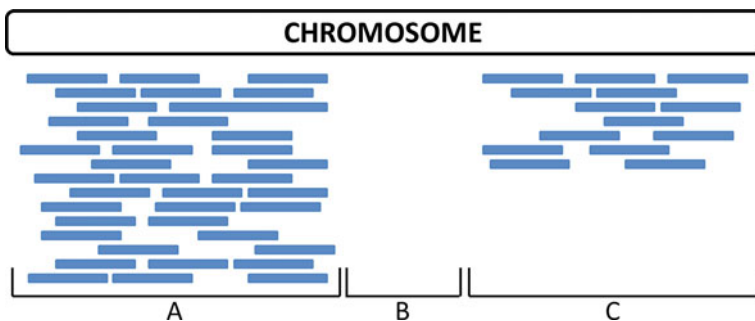


Fig. 2 Read mapping to a reference genome. (a) High coverage area of the chromosome; (b) Area of chromosome with no coverage; (c) Area of genome with low coverage

The output formats for this process vary widely, but the most common formats are SAM/BAM alignment files. Sequence Alignment/Map (SAM) files are human readable, whereas BAM files are compressed versions of the same information. Both of these files contain all the read sequences as well as where they map to the reference genome and their mapping scores (a measure of how well they mapped).

SNV/SNP Calling

Software is then used to “call” the variants at reported positions. Variant positions are those where the mapped nucleotides differ from the expected reference nucleotide. Common software used for variant calling includes SAMtools (Li et al. 2009a), HIVE-heptagon (Simonyan and Mazumder 2014), and SOAP2 (Li et al. 2009b). The variant calling process is more complicated in diploid and other polyploid organisms which can have two different nucleotides at a single position due to multiple copies of chromosomes, or by sequencing errors inherent in the process. A clever algorithm can utilize a higher coverage level to discard erroneous variations and also report the proportions of nucleotides in specific positions. For a human heterozygous at the position of interest, one would anticipate two different nucleotides each appearing 50 % of the time throughout the coverage. For a human homozygous at the position of interest, one would expect a single nucleotide to be represented. It is, however, possible to have a mosaic set of DNA from an individual or for nucleotides to be inserted or deleted relative to the reference genome.

After variant calling, a file is produced cataloging the variations found in the alignment. The most common format is the Variant Call Format (VCF) file. This is a human readable file which contains information about the position of each call as well as the reference nucleotide(s), the variation(s) noted including insertions and deletions (commonly called indels), and the frequencies of each variation. Additional optional, user-defined information can be included depending on the specifications of the researcher.

Identifying nsSNVs

Once variants are called, it is possible to categorize each single-nucleotide variation (SNV) as either non-synonymous or synonymous. Software is used to look at each variant’s position and compare that to a database of coding regions. The database contains information regarding open reading frames (ORFs) of the coding region which host the SNV. With information about the open reading frame, the software is then able to determine the new codon when the reference nucleotide is replaced by the variant one. This three-letter nucleotide set codes for the amino acid that will be included in the protein. Depending on the location of the nucleotide change, the amino acid might also change (e.g., often if the variation is in the first position of the codon) or it might remain the same (most commonly when the change occurs in the final position of the codon).

If the amino acid changed due to the variation, then the SNV is called a non-synonymous variation (see Fig. 3). Non-synonymous variations have the

a Original Sequence	...CGCTGC A GACCAC...
Variant Sequence	...CGCTGC C GACCAC...
b Reading Frame (Original)	...CG CTG CA G ACC AC...
Reading Frame (Variant)	...CG CTG CC G ACC AC...
c Amino Acid Sequence (Original)	...Leu Gln Thr...
Amino Acid Sequence (Variant)	...Leu Pro Thr...

Fig. 3 (a) The original DNA sequence of the sense (coding) strand in the 5' to 3' direction followed by the sequence with the SNV (the variant sequence) in the same orientation; (b) The sequences from A with the ORF information added; each space separates the nucleotide codons, or set of three nucleotides that code for an amino acid; (c) The amino acid sequence translated from mRNA transcribed from the original and variant sequences; here, the original sequence results in the amino acid chain leucine – glutamine – threonine. The variant sequence results in the amino acid sequence leucine – proline – threonine. In this example, the SNV would be considered non-synonymous since leucine was changed to a completely different amino acid, proline, due to the variation

potential to affect the function of the protein by direct interruption of active or binding sites or by indirect effects such as steric hindrances, charge modifications, and others. Disruption of the protein can also happen when the new amino acid changes the protein's three-dimensional structure. Synonymous variations, on the other hand, are generally innocuous. They do not directly change the shape or function of the protein, but can have regulatory effects by changing the rate that RNA polymerase is able to transcribe the region or by altering binding properties of that portion of the DNA.

From Genomic to Proteomic Identification

Since the advent of NGS technology, identification of genomic variation through whole-genome sequencing (WGS) and whole-exome sequencing (WES) has greatly improved, enhancing the ability to study genotype-phenotype disease associations. The International HapMap Project (International HapMap et al. 2007) has contributed to the identification of approximately ten million common DNA variants, primarily SNVs. Despite this accomplishment, however, the project asserts that current knowledge of human genetic variation is incomplete due to lack of information about rarer variants, such as minor allele frequency variants and copy number variants, which are not as well studied (International HapMap et al. 2010). Pilot results of the 1000 Genomes Project (Genomes Project et al. 2012) also demonstrate the limitations of genomic approaches, indicating that, while much common variation has been captured, significant phenotypic variation can be attributed to variants missed by commonly used genotyping arrays.

Thus, while genomic strategies have laid the foundation for the cataloging of variation, disease-associated and otherwise, there is still much left to be discovered. Furthermore, drawbacks of whole-genome approaches include high costs and provision of an overwhelmingly vast amount of data, increasing the difficulty of discerning benign variants from those that may be pathogenic (Royer-Bertrand and Rivolta 2015). Technical aspects of NGS data also present significant challenges including storage and maintenance, quality control, and analysis that is both reliable and efficient (Xuan et al. 2013). Despite disadvantages of genomic strategies of variant detection, the knowledge that can be deduced from such studies is imperative to a complete understanding of certain disease states.

Similarly, proteomic technologies used to discover disease-associated amino acid variation biomarkers are greatly beneficial. High-throughput proteomic technologies have only recently been developed (Branca et al. 2014), but have the potential to enable an understanding of biological and disease processes with increased granularity as compared to genomic technologies. While these strategies may pose similar challenges to technical logistics as described for the genomic approaches, the knowledge gained from proteomic studies is closer to the pathology of complex disease states and therefore closer to disease detection and therapy (National Research Council 2006). To this end, several proteomic databases have emerged over the last decade including GPMDB (Craig et al. 2004), PeptideAtlas (Deutsch 2010), MassIVE, Chorus, PRIDE (Cote et al. 2012), and more. Many of these databases belong to the ProteomeXchange consortium (www.proteomexchange.org) which facilitates central access to shared data across resources and maintains guidelines for acceptable data formatting. Despite best efforts, a number of unique challenges exist such as MS/MS spectral and peptide database matching, incomplete sequence databases with missing or incorrect annotations, the need for optimization, and the lack of standardized preparation and validation protocols (Omenn et al. 2005). As databases evolve to include a more biologically representative set of viable proteins and synthetic constructs of potential variant including peptides, the quality of resultant peptide libraries will increase tremendously. In turn, it will become easier to analyze the presence and statistical importance of nsSNVs as potential disease biomarkers.

Some current applications of proteomics to nsSNV-based biomarker discovery include quantitative analysis of pancreatic cancer-associated single-amino-acid variant peptides (Nie et al. 2014), identification of cancer-related splice variants and validation via custom library (Hatakeyama et al. 2011), and discovery of a novel hepatitis B-related candidate biomarker (Marrocco et al. 2010). There is a great emphasis in the literature and interest in the community on best interpretation of quantitative analysis, methods for identifying low-abundance peptides, and custom-built, purpose-specific peptide databases. With respect to cardiology, a combined tandem mass spectrometry and sequence homology approach was used to identify a novel, single-amino-acid variation resulting from nsSNV in swine cardiac troponin I (Zhang et al. 2010). These cases demonstrate the enormous potential of proteomics to further resolve mechanisms of various cardiovascular diseases and identify single-amino-acid variation resulting from nsSNVs as diagnostic biomarkers or potential therapeutic targets.

Potential Applications to Prognosis, Other Diseases, or Conditions: Cardiovascular Diseases and Associated nsSNVs

The following sections explore the different cardiovascular diseases and associated nsSNVs. While each of these conditions can be characterized by a wide range of biomarkers, symptoms, and risk factors, the nsSNVs reported were found to be associated with the disease, either through increased susceptibility or even decreased susceptibility. The nsSNVs are potential points for further investigation and do not yet represent definite clinical diagnostic markers. In the following text, specific variants are referred to by the rsID, or the reference SNP cluster ID, which is the accession number for a given variant in the dbSNP database.

Ischemic Stroke

An ischemic stroke is a lack of blood reaching the brain and is caused by narrowing or clogging of blood vessels with plaque (American Stroke Association 2013). According to stroke.org, someone dies from stroke every 4 min in the United States, and stroke is also the leading cause of adult disability. Ischemic stroke is associated with high mortality and severe morbidity: victims often experience permanent neurological disability following an episode (Lee et al. 2010). The main risk factors of ischemic stroke are high blood pressure, high cholesterol, and diabetes, but research suggests that genetic variations are another important factor (Flossmann et al. 2004; Gretarsdottir et al. 2003). While the exact mechanisms by which genetic variations influence the likelihood of ischemic stroke are poorly understood, the associations are significant (Guo et al. 2013).

In a recent study of 1,209 patients with stroke and 1,174 controls from a Chinese population, researchers found that rs2230500 is significantly associated with both the risk of ischemic stroke (age- and sex-adjusted odds ratio = 1.37; 95 % CI, 1.12–1.67; $P = 0.0019$) and cerebral hemorrhage (age- and sex-adjusted odds ratio = 1.96; 95 % CI, 1.21–3.19; $P = 0.0064$) (Wu et al. 2009). This result confirmed previous studies finding the variant significantly associated with stroke in Japanese populations (Kubo et al. 2007; Serizawa et al. 2008). Note that both of these are Asian populations where the minor allele frequencies of this nsSNV are 0.239 for Japanese in Tokyo and 0.178 for Han Chinese in Beijing. According to the HapMap database, the minor allele frequencies for Utah residents with Northern and Western European origins was 0.008, and 0.00 for Yoruba in Ibadan, Nigeria (Kubo et al. 2007). The polymorphism is a G to A substitution in exon 9 at position 1425 of PRKCH, a gene located in position 61457521 of chromosome 14q22–q23 in humans. The variant causes an amino acid substitution from valine to isoleucine in position 374 of the protein (Shimizu et al. 2007).

The residue change occurs in the ATP-binding site of the serine-threonine kinase (Wu et al. 2009). PRKCH is known to be involved in a variety of signaling pathways

and regulates cellular functions such as proliferation and apoptosis (Kubo et al. 2007). Expressed mainly in endothelial cells, the kinase plays a role in human atherosclerosis (Kubo et al. 2007). The nsSNV was found to significantly increase autophosphorylation and kinase activity after stimuli (Kubo et al. 2007). This agrees with the biological plausibility of the assertion that if a protein involved in atherosclerosis, a risk factor of ischemic stroke, is overly activated due to a genetic mutation, there will consequently be a higher risk of stroke.

Coronary Artery Disease

Coronary artery disease (CAD) is the most common type of heart disease and is responsible for the most deaths in the United States among men and women every year (National Heart Lung and Blood Institute 2014). The disease is characterized by the accumulation of plaque in the coronary arteries (National Heart Lung and Blood Institute 2014). This process, called atherosclerosis, gradually deprives the heart of oxygen-rich blood over time. If incoming blood is sufficiently blocked, a heart attack will occur. The major risk factors of coronary artery disease include dyslipidemia, smoking, hypertension, and diabetes (Achari and Thakur 2004). Unfortunately, due to the complexity of CAD, the influence of genetic factors on disease susceptibility is not completely understood. Pathogenesis is believed to be caused by the interactions of multiple genetic and environmental influences. The major role family history plays as an indicator of CAD susceptibility strengthens the idea that a genetic component is important (Wang 2005).

One potential genetic biomarker is the nsSNV rs2305948 on chromosome 4 at position 55113391 (Sherry 2001). The role of the polymorphism as a risk indicator for CAD was confirmed in two independent case-control studies. The first study was comprised of 655 patients with coronary heart disease and 1,015 controls, whereas the second study was based on 369 subjects and 625 controls (Wang et al. 2007). The two studies found that rs2305948 is associated with risk of coronary heart disease with an odds ratio of 1.41 ($P = 0.011$) in the first cohort and an odds ratio of 1.75 ($P = 0.003$) in the second cohort (Wang et al. 2007). The polymorphism is a C to T substitution in exon 7 of the kinase insert domain-containing receptor/fetal liver kinase-1 (KDR) gene. KDR is a receptor for the vascular endothelial growth factor (VEGF); together, they play a critical role in angiogenesis and vascular repair. The variant in the KDR gene results in an amino acid substitution from valine to isoleucine in position 297 in the third NH₂-terminal Ig-like domain within the extracellular region (Wang et al. 2007). As a key component of the VEGF-binding domain, the nsSNV decreases the efficiency of VEGF and KDR binding. This inhibits KDR function and dampens the resulting signaling pathway (Wang et al. 2007). While recent experiments in animal models have shown that VEGF promotes atherosclerosis, the exact mechanism by which KDR influences disease development is still unknown (Wang et al. 2007).

Sudden Cardiac Death

Sudden cardiac death (SCD) is estimated to be involved in a quarter of all human deaths globally each year (Abhilash and Namboodiri 2014). SCD describes an unexpected death within an hour of symptom onset due to cardiac causes without any extra cardiac event having occurred within the previous 24 h (Havmoller and Chugh 2012). While most instances of SCD are caused by ventricular fibrillation (Abhilash and Namboodiri 2014), other risk factors include coronary heart disease, physical stress, structural changes in the heart, and inherited disorders (National Heart Lung and Blood Institute 2011). Low survival rates have catalyzed the effort to identify improved risk markers (Havmoller and Chugh 2012). While the current widely used risk markers include QT interval and LVEF, the addition of potential biomarkers such as plasma and inflammatory markers has yet to provide adequate predictive value (Havmoller and Chugh 2012). However, the use of genomic or proteomic technologies may supply novel diagnostic and therapeutic targets.

One potential marker is the variant rs7626962 found on chromosome 3 in position 38579416. Although the variant has a minor allele frequency of approximately 13 % in African American populations (Cheng et al. 2011), it is difficult to conduct a genome-wide association study on deceased patients. Consequently, many variations are discovered in postmortem genetic testing. One association was found in a genetic analysis of a 23-year-old African American male who died suddenly (Cheng et al. 2011). The variant was also found in three affected members of a white family but not found in the non-affected family members (Chen et al. 2002). This finding is especially significant as the polymorphism was understood to have negligible prevalence in populations of white European ancestry (Splawski et al. 2002). Furthermore, two separate studies confirmed the association of rs7626962 with sudden death. The first examined 133 cases of sudden infant death syndrome (SIDS) and 1,056 controls and found that infants with two copies of the polymorphism have a 24-fold increased risk for SIDS (Plant et al. 2006). The second study also found a significant association between the nsSNV and SIDS in a cohort of 71 African American SIDS victims (Van Norstrand et al. 2008).

The variant is a C to A mutation in position 3308 of the SCN5A gene and causes an amino acid change from serine to tyrosine in position 1103 of the protein (Cheng et al. 2011). SCN5A is a voltage-gated, type V, alpha subunit sodium channel (Sherry 2001). In addition to sudden cardiac death, mutations in this gene are known to cause Brugada syndrome, long QT syndrome (LQTS), and arrhythmias (Abunimer et al. 2014; Plant et al. 2006). Although experiments showed that mutant and wild-type variants of the sodium channel behave identically at pH 7.4, functional differences were observed when tested under conditions that would be expected *in vivo*. When pH was decreased from 7.4 to 7.0 and then 6.7, as would be expected in acidosis, the Y1103 variants experienced progressive shifts in the voltage dependence of steady-state inactivation. In addition, the mutant channels had shortened recovery times from inactivation. This suggests that, in conditions of low internal pH, mutant SCN5A channels may activate during unanticipated periods of the

cardiac cycle compared to wild-type channels. This hypothesis was confirmed when the variant channels were found to abnormally reopen during depolarization at pH 6.7 compared to wild-type channels which remained inactive (Plant et al. 2006). This unexpected opening of sodium channels, which play a crucial role in cardiac cycles, may explain the association of the nsSNV and SCD, as well as provide further evidence to the value of rs7626962 as a biomarker in assessing SCD preventative therapy.

Congestive Heart Failure

In 2009, one in nine deaths in the United States was partially linked to heart failure. Today, there are approximately 5.1 million people living with heart failure, and nearly half of people who develop heart failure die within 5 years of diagnosis (Go et al. 2013). Risk factors for the disease include coronary heart disease, high blood pressure, diabetes, smoking, poor diet, sedentary lifestyle, and obesity. While it is known that there is a strong hereditary component, this component is poorly defined in common forms of the disease (Cappola et al. 2011). One promising genetic marker is a loss-of-function (LOF) variant in the CLCNKA chloride channel.

The nsSNV rs10927887 was found to be positively associated with heart failure in three independent Caucasian heart failure populations. The variant on chromosome 1 in position 16024780 is an A to G substitution in the CLCNKA gene, which leads to an arginine to glycine change in position 83 (exon 3) of the protein (Sherry 2001). The variant was found to be present in 50 % of the 625 unaffected controls and in 56 % of 1,117 Caucasian heart failure cases. These frequencies were similar in examination of another independent cohort of 857 subjects and 311 controls. The association was robust enough to be statistically significant in a subgroup analysis for heart failures of any type. Independent of age, gender, and hypertension, the risk of heart failure increases by 27 % and 54 % for heterozygotes and homozygotes of the nsSNV, respectively (Cappola et al. 2011). These associations are likely a result of the functional differences in the CIC-K_a channel as a result of the amino acid substitutions. The glycine 83 mutant channels evoked currents with smaller amplitudes across tested potentials compared to wild-type channels. In addition, the efficiency of the mutant channels was less sensitive to extracellular chloride ion concentration compared to wild type. An immunoblot analysis used as a control found no difference between expression levels of the two channels in the cellular model, suggesting that any differences in efficacy was due to the inherent characteristics of the mutant channels (Cappola, Matkovich et al. 2011). Ostensibly, a nsSNV reducing the chloride currents through a renal CIC-K_a chloride channel would not cause congestive heart failure. However, a known variant, Cys 80 CIC-K_a mutation, with a similar LOF profile was found to cause a Bartter-like syndrome in conjunction with the disruption of the related CLCNKB gene (Schlingmann et al. 2004). This syndrome is a salt-wasting disorder of which one abnormality is hyperreninemia, an established risk factor for heart failure (Modlinger et al. 1973; Bongartz et al. 2005).

Myocardial Infarction

Every year in the United States, an estimated 785,000 people will have a new myocardial infarction (MI). With approximately a death every minute in the United States, MI is a major cause of morbidity globally (Jneid et al. 2013) and the leading cause of death among all cardiovascular diseases (Sahoo and Losordo 2014). While the exact definition of a myocardial infarction includes patient symptoms, echocardiogram changes, and sensitive cTN biochemical markers, it is, in essence, a condition in which inadequate blood flow to heart muscles disrupts cardiac function and prompts necrosis (Jneid et al. 2013; National Heart Lung and Blood Institute 2013). Risk factors for MI include controllable risk factors such as smoking, hypertension, high cholesterol, obesity, a sedentary lifestyle, and uncontrollable factors such as age and genetics.

One possible biomarker in assessing the risk for myocardial infarction is the SNP rs73184536. While most of the nsSNVs explored in this chapter increase risk of a cardiovascular disease or condition, this variant offers protection. Found on chromosome 13 in position 37636968, the variant codes for a T to C allelic substitution in the gene for the transient receptor potential cation channel, subfamily C, member 4 (TRPC4). This mutation in exon 11 results in an isoleucine to valine substitution at position 957 of the protein (Jung et al. 2011). In a sample of 3,899 controls and 1,025 patients with a first MI, the variant was associated with decreased risk of MI (odds ratio = 0.61; 95 % CI (0.40–0.95); $P = 0.02$) when adjusted by age, sex, hypertension, and antihypertensive therapy.

The gene belongs to a family of nonselective ion channels and is expressed in vasculature (Yip et al. 2004) where it facilitates intracellular Ca^{2+} signaling. Intracellular Ca^{2+} signals are critical in the regulation of endothelial permeability (Tiruppathi et al. 2002), smooth muscle proliferation (Zhang et al. 2004), and endothelium- and nitric oxide (NO)-dependent vasorelaxation (Freichel et al. 2001). As mentioned before, the crux of the problem in MI is the inhibition of blood supply to the myocardium. As blood is a liquid, flow is inversely related to the resistance from the myocardial vascular bed (Jung et al. 2011). This resistance is dependent on the vascular smooth muscle and consequently on calcium signaling (Jaggar et al. 2000). TRPC4 activity is regulated through kinase phosphorylation of a tyrosine in position 959 that, once activated, inserts additional channels into the plasma membrane (Jung et al. 2011). A single-channel analysis revealed a threefold increase in active TRPC4-I957V channels compared to wild-type channels following carbachol stimulation. The enhanced channel activity of the TRPC4 variant increases Ca^{2+} signaling which may facilitate endothelium- and NO-dependent vasorelaxation. This process may ultimately decrease resistance in the myocardial vascular bed and explain the MI risk protection offered by the nsSNV rs73184536 (Jung et al. 2011).

Congenital Heart Defects

According to the American Heart Association, congenital heart defects are a common form of birth defects and comprise a long list of heart malformations, including

aortic valve stenosis and atrial septal defect. Every year in the United States, nearly 1 % of births are affected by congenital heart defects (CHD). While not all cases are fatal, CHDs are responsible for 4.2 % of all neonatal deaths. In addition, while 95 % of babies born with a noncritical CHD are expected to survive to adulthood, this increases the number of adults living with CHD (Center for Disease Control and Prevention 2014). The exact mechanisms behind each type of defect vary, but CHDs are generally understood to be a result of multiple environmental and genetic factors (Arrington et al. 2012).

One potential genetic marker is a non-synonymous mutation in the pre-B-cell leukemia homeobox 3 (PBX3) gene. The rs145687528 variant is found on chromosome 9 in position 125915818 and is a C to T substitution which results in an alanine to valine amino acid substitution in position 136 of the protein (Sherry 2001). The variant is positioned in a conserved polyalanine track and was present in 5.2 % of the 95 heart defect patients, compared to only 1.3 % of the race and ethnicity-matched control patients (Arrington et al. 2012). This significant overrepresentation of the variant reveals rs145687528 as a valuable risk allele for congenital heart defects (Arrington et al. 2012).

The gene in question codes for a pre-B-cell leukemia homeobox (PBX) protein and belongs to the pre-B-cell leukemia (PBC) transcription factor family and shares a three-amino-acid loop extension in the homeodomain with other members of the TALE superfamily (Arrington et al. 2012). The variant is the seventh alanine in a nine-alanine motif in PBX3. That the amino acid sequence is highly conserved bolsters conclusions from *in silico* analysis which showed a high probability that the mutation is deleterious (Arrington et al. 2012). Polyalanine tracts are thought to be involved in transcription factor repression or facilitation of DNA binding in a transcription complex (Brown and Brown 2004). While the exact mechanism by which this mutation leads to a congenital heart defect is not understood, it does provide a new avenue for further investigation.

Hypertension

Hypertension is a chronic condition where elevated blood pressure slowly damages blood vessels and organs. The increasing rate of hypertension is a cause for concern as it is a major risk factor for cardiovascular disease and leads to higher mortality globally (Lawes et al. 2008; Xi et al. 2012, 2013). Currently, an estimated 26.4 % of the world's adult population are afflicted with hypertension (Kearney et al. 2005). While obesity, stress, and excess salt in the diet are known causes of hypertension, there are also genetic factors that interact and play a role (Medicine 2015). Genetic factors contribute approximately 20–40 % of the variance in blood pressure among the general population (Choh et al. 2005). Another approximation attributed 65 % of variation in blood pressure over a 24 h period to genetic factors (Tobin et al. 2005).

One potential genetic factor is the nsSNV rs7565062 in the gene SCN7A. Found in exon 25 on chromosome 2 in position 166477575, the variant is a G to T substitution that leads to a threonine to asparagine change in position 41 of the

sodium channel, voltage-gated, type VII, alpha subunit (Sherry 2001). In a study of 1,232 unrelated subjects from the Northern Han population of China, 615 with hypertension and 617 controls, the T allele in rs7565062 had significantly higher prevalence in the hypertensive cohort ($P = 0.045$). This association with hypertension signifies that the T allele acts as a risk factor for the condition. Through logistic regression analysis, rs7565062 was found to be significantly associated with essential hypertension in both the additive (TT vs. TG vs. GG: $P = 0.024$, OR = 1.283, 95 % CI: [1.033–1.592]) and dominant ((TT + TG) vs. GG: $P = 0.013$, OR = 1.203, 95 % CI: [1.040–1.392]) genetic models (Zhang et al. 2015).

The sodium channel, voltage-gated, type VII, α -subunit (SCN7A) belongs to the gene family encoding the α -subunit of voltage-gated sodium channels (VGSCs). Although this is the official classification of the channel, one study found the channel encoded in part by SCN7A is sodium concentration gated rather than voltage-gated (Hiyama et al. 2002). The channel was also identified to function as a sodium-level sensor in blood flow (Shimizu et al. 2007) and regulate sodium intake (Hiyama et al. 2010). The mechanism in which this variant induces hypertension may be found through its connection with Na_x , which is an isoform of the α -subunit found in voltage-gated sodium channels (Zhang et al. 2015). Na_v2 is a member of the SCN7A-encoded Na_x and is expressed in the neurons and ependymal cells in circumventricular organs involved in body-fluid homeostasis (Watanabe et al. 2000). Experiments in a mouse model showed that Na_x -null mice had abnormal intake of hypertonic saline. The finding suggests that Na_x monitors sodium concentration and is involved in sodium intake regulation (Zhang et al. 2015). These findings offer a biological context that reinforces the association between the mutation and elevated risk of hypertension.

Arrhythmia

Arrhythmias, including atrial fibrillation, tachycardia, and bradycardia, are a set of conditions defined by abnormal electrical activity of the heart and are a major cause of stroke and sudden cardiac arrest (Abunimer et al. 2014). The role of arrhythmias in these sudden adverse cardiac events is such that hereditary arrhythmias are responsible for over half of sudden cardiac deaths in young individuals (Beckmann et al. 2011). Despite the low prevalence of hereditary arrhythmias in populations, early detection of the condition is essential to beginning early preventative measures. Consequently, understanding the genetic causes underlying the various conditions categorized as arrhythmias is imperative for improving diagnosis and therapy and ultimately identifying individuals who may be at a higher risk for severe cardiac events associated with arrhythmias.

A candidate for further genetic study of arrhythmias is the nsSNV rs6795970. A study found the mutation strongly associated with QRS duration, which measures cardiac intraventricular conduction and is a common indicator of arrhythmias (Ritchie et al. 2013). The exact mutation is an A to G substitution in the sodium channel, voltage-gated type X, alpha subunit (SCN10A) gene, in chromosome 3 at

position 38725184. This exonic polymorphism corresponds to a valine to alanine amino acid substitution at position 1073 (Sherry 2001). In a genome-wide association study of nearly 14,000 European-American subjects, this particular SNP on chromosome 3 was found to be significantly associated with cardiac arrhythmias, atrial fibrillation and flutter, arterial embolism and thrombosis, and many other conditions. While the association with cardiac arrhythmias was strongest, the association of rs6795970 with altered QRS duration and with cardiac arrhythmia were not dependent, which suggests that while the SNP may influence QRS duration and susceptibility to arrhythmia development, their pathways are divergent (Ritchie et al. 2013).

The gene in question, SCN10A, is a voltage-gated sodium channel labeled $Na_v1.8$ and codes for a protein more commonly known for cold perception in afferent nociceptive fibers (Blasius et al. 2011). While the exact mechanism through which the mutated SCN10A gene leads to arrhythmias is unknown, the three predominant theories are that it affects conduction directly via cardiomyocytes, indirectly via intracardiac neurons, or, more recently proposed, as an enhancer of SCN5A gene expression. A recent study discovered that while SCN10A expression is negligible in human and murine hearts, a T-box enhancer within SCN10A drives SCN5A expression in cardiomyocytes (Park and Fishman 2014). This third theory is further evidenced by previously inconclusive studies of attempting to characterize the role of the SCN10A protein in heart physiology (Akopian et al. 1996). Despite a yet uncharacterized pathway, the nsSNV rs6795970 is definitively associated with cardiac arrhythmias, and further study on the SNP is necessary to further elucidate potential therapeutic or diagnostic targets.

Cardiomyopathy

Cardiomyopathies are diseases of the myocardium classified by structural and functional abnormalities (Sisakian 2014). In most cases, heart muscle becomes thicker or more rigid than normal. While patients with cardiomyopathy may live long healthy lives, it is a major cause of heart failure which is a leading cause of death (Simonson et al. 2010). As with other conditions and diseases, genetic biomarkers are playing an increasingly important role in classification and diagnosis (Sisakian 2014).

One potential marker for identifying susceptibility for dilated cardiomyopathy (DCM) is the cytotoxic T-lymphocyte antigen 4 (CTLA4) (Ruppert et al. 2010). The receptor belongs to the CD28-B7 immunoglobulin superfamily of immune regulatory molecules which downregulate T-cell activation. CTLA4 is expressed on the plasma membrane of activated T cells and functions as an inhibitory signal for T-cell proliferation after binding to B7 receptor molecules on antigen-presenting cells (Ruppert et al. 2010). Ostensibly, a receptor in the immune system should not be involved in the development of DCM. However, a major factor in DCM pathogenesis is known to be autoimmune-mediated damage to cardiac tissue (Ruppert et al. 2010).

The mutation in question is rs231775, an A to G substitution in position 49 of exon 1 of CTLA4 on chromosome 2 in position 203867991 (Sherry 2001). The nsSNV was confirmed in a study of two independent cohorts of dilated cardiomyopathy patients ($n = 251$ and 223) and a sample of 591 healthy controls (Ruppert et al. 2010). The G/G genotype of the variant was found in 14.7 % of subjects compared with only 7.4 % of controls, ($P = 0.005$). The mutation codes for a threonine to alanine substitution in position 17 of the protein (Sherry 2001). This position corresponds to the peptide leader sequence of the CTLA4 receptor. This specific mutation was shown to increase expression of cell-surface CTLA4 receptors on stimulation of T cells, as well as associate with autoimmunity in general (Ligers et al. 2001). These findings further strengthen the interrelatedness of autoimmune disorders and cardiomyopathies as well as present an additional risk marker in DCM pathogenesis.

The Future of nsSNVs and Cardiovascular Diseases

Genomic and Proteomic Projects Worldwide Associated with Cardiac Diseases

There are a high number of institutes and centers worldwide that have recently published papers investigating cardiac systems diseases and conditions through genomic or proteomic means. The high number of international institutes displays that the value of these technologies in identifying potential biomarkers and nsSNV as potential therapeutic and diagnostic targets is globally appreciated. The over 2,000 departments, schools, labs, and centers reinforce the theme of this chapter: namely, that genomic and proteomic technologies are an excellent method of identifying potential therapeutic and diagnostic biomarkers in cardiovascular diseases. In particular, nsSNVs and their associations with cardiovascular disease susceptibility and protection represent value opportunities for further study.

Workflow and Results

Sample Workflow

We present a sample workflow which may be applied to diverse datasets to harness the nsSNVs associated with cardiovascular (or other) diseases as biomarkers. The following workflow was performed for S-nitrosylation but may be repeated and expanded for other features. The importance of S-nitrosylation stems from nitric oxide's role as a relaxation factor derived in the endothelium – where nitric oxide (NO) is largely controlled by S-nitrosylation (Lima et al. 2010). The first step was to

retrieve the human proteome and nine other species (*Mus musculus*, *Bos taurus*, *Canis familiaris*, *Equus caballus*, *Xenopus tropicalis*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*) from the UniProtKB/Swiss-Prot database, which is available online.

Next, using the protein BLAST tool, we performed pairwise alignments between the human proteome and nine other species. From the alignment results, all conserved cysteine positions, i.e., the positions which exist in human protein sequences and were mapped at least to one species, were extracted. Cysteine positions were specifically targeted because a cysteine thiol is covalently modified by an NO group to produce S-nitrosothiol (SNO) and thus plays a central role in S-nitrosylation (Lima et al. 2010). Then, the table containing conserved nsSNVs cysteine positions was generated by mapping the conserved cysteine positions among species and human nsSNVs positions from SNVDis (Karagiannis et al. 2013). The GPS-SNO tool (Xue et al. 2010) was used to predict S-nitrosylation sites for the conserved cysteine positions. The rsIDs and swissvarIDs (variation identifier from Swiss-Prot database) obtained from the table of the conserved cysteines and also predicted to be S-nitrosylation sites were used in order to get the information about diseases caused by the variation.

Sample Workflow Results

Results of this workflow are counts, positions, and amino acid variations of observed and predicted disease-related nsSNVs occurring at a conserved cysteine residue of the reference human genome. Please see Tables 1 and 2 for a summary of the results.

Summary Points

- Non-synonymous single-nucleotide variations (nsSNVs) are changes in the genome which ultimately lead to amino acid substitutions and possible changes in biochemical pathways or protein structure or activity.
- Next-generation sequencing (NGS) technologies represent an opportunity to rapidly, cheaply, and efficiently identify nsSNVs as biomarkers and potential therapeutic or diagnostic targets associated with diseases and conditions.

Table 1 Summary counts of impacted proteins, sites, and disease relatedness of genome-wide nsSNV at conserved cysteine residues

	Number of proteins	Number of sites	Diseases
Predicted sites	44	49	84
Experimental sites	7	7	15

This table shows the distribution of nsSNVs among all genomic cysteine positions and the corresponding abundance of potentially affected proteins containing the site and resulting disease relatedness. Although experimental evidence for clinical significance is limited, this omics approach identifies several predicted candidate nsSNVs for further study.

Table 2 Summary of different types of variation and their frequencies

Variation	Predicted sites	Experimental sites
c → r	11	4
c → *	5	
c → y	13	1
c → w	2	
c → s	4	1
c → f	7	1
c → g	6	
c → a	1	

This table shows the distribution of the number of specific amino acid substitutions resulting from nsSNV at a reference cysteine codon. Although several possible substitutions exist, it is interesting to note that only four such substitutions have been observed to date.

- NGS technologies and major software developments have expedited the discovery and analysis of genomic and proteomic data which are essential to the identification of nsSNVs.
- Mapping of the genomic data is facilitating the process of identifying diagnostic and therapeutic targets for a number of diseases and conditions.
- Proteomic approaches can enable cheaper, more rapid, and more robust identification of variation biomarkers and validation of genomic targets against amino acid variants.
- Cardiovascular diseases are an increasing global public health problem, and nsSNVs are playing an integral part in their further study.

References

- Abhilash SP, Namboodiri N. Sudden cardiac death – historical perspectives. *Indian Heart J.* 2014;66 Suppl 1:S4–9.
- Abunimer A, Smith K, Wu TJ, Lam P, Simonyan V, Mazumder R. Single-nucleotide variations in cardiac arrhythmias: prospects for genomics and proteomics based biomarker discovery and diagnostics. *Genes (Basel).* 2014;5(2):254–69.
- Achari V, Thakur AK. Association of major modifiable risk factors among patients with coronary artery disease – a retrospective analysis. *J Assoc Physicians India.* 2004;52:103–8.
- Akopian AN, Sivilotti L, Wood JN. A tetrodotoxin-resistant voltage-gated sodium channel expressed by sensory neurons. *Nature.* 1996;379(6562):257–62.
- American Stroke Association T. Ischemic strokes (Clots). 2015, from http://www.strokeassociation.org/STROKEORG/AboutStroke/TypesofStroke/IschemicClots/Ischemic-Stroke-Clots_UCM_310939_Article.jsp (7 Nov 2013).
- Arrington CB, Dowse BR, Bleyl SB, Bowles NE. Non-synonymous variants in pre-B cell leukemia homeobox (PBX) genes are associated with congenital heart defects. *Eur J Med Genet.* 2012;55(4):235–7.
- Beckmann BM, Pfeufer A, Kaab S. Inherited cardiac arrhythmias: diagnosis, treatment, and prevention. *Dtsch Arztebl Int.* 2011;108(37):623–33; quiz 634.
- Blasius AL, Dubin AE, Petrus MJ, Lim BK, Narezkina A, Criado JR, Wills DN, Xia Y, Moresco EM, Ehlers C, Knowlton KU, Patapoutian A, Beutler B. Hypermorphic mutation of the voltage-

- gated sodium channel encoding gene *Scn10a* causes a dramatic stimulus-dependent neurobehavioral phenotype. *Proc Natl Acad Sci U S A*. 2011;108(48):19413–8.
- Bongartz LG, Cramer MJ, Doevendans PA, Joles JA, Braam B. The severe cardiorenal syndrome: 'Guyton revisited'. *Eur Heart J*. 2005;26(1):11–7.
- Branca RM, Orre LM, Johansson HJ, Granholm V, Huss M, Perez-Bercoff A, Forshed J, Kall L, Lehtio J. HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat Methods*. 2014;11(1):59–62.
- Brown LY, Brown SA. Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet*. 2004;20(1):51–8.
- Cappola TP, Matkovich SJ, Wang W, van Booven D, Li M, Wang X, Qu L, Sweitzer NK, Fang JC, Reilly MP, Hakonarson H, Nerbonne JM, Dorn 2nd GW. Loss-of-function DNA sequence variant in the *CLCNKA* chloride channel implicates the cardio-renal axis in interindividual heart failure risk variation. *Proc Natl Acad Sci U S A*. 2011;108(6):2456–61.
- Center for Disease Control and Prevention, T. National Center on Birth Defects and Developmental Disabilities (NCBDDD). 2014. Congenital Heart Defects (CHDs). 2015, from <http://www.cdc.gov/ncbddd/heartdefects/data.html> (9 July 2015).
- Chen S, Chung MK, Martin D, Rozich R, Tchou PJ, Wang Q. SNP S1103Y in the cardiac sodium channel gene *SCN5A* is associated with cardiac arrhythmias and sudden death in a white family. *J Med Genet*. 2002;39(12):913–5.
- Cheng J, Tester DJ, Tan BH, Valdivia CR, Kroboth S, Ye B, January CT, Ackerman MJ, Makielski JC. The common African American polymorphism *SCN5A*-S1103Y interacts with mutation *SCN5A*-R680H to increase late Na current. *Physiol Genomics*. 2011;43(9):461–6.
- Choh AC, Czerwinski SA, Lee M, Demerath EW, Wilson AF, Towne B, Siervogel RM. Quantitative genetic analysis of blood pressure response during the cold pressor test. *Am J Hypertens*. 2005;18(9 Pt 1):1211–7.
- Cote RG, Griss J, Dianas JA, Wang R, Wright JC, van den Toorn HW, van Breukelen B, Heck AJ, Hulstaert N, Martens L, Reisinger F, Csordas A, Ovelleiro D, Perez-Rivevol Y, Barsnes H, Hermjakob H, Vizcaino JA. The PRoteomics IDentification (PRIDE) Converter 2 framework: an improved suite of tools to facilitate data submission to the PRIDE database and the ProteomeXchange consortium. *Mol Cell Proteomics*. 2012;11(12):1682–9.
- Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res*. 2004;3(6):1234–42.
- Deamer DW, Akeson M. Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends Biotechnol*. 2000;18(4):147–51.
- Deutsch EW. The peptide Atlas project. *Methods Mol Biol*. 2010;604:285–96.
- Dingerdissen H, Motwani M, Karagiannis K, Simonyan V, Mazumder R. Proteome-wide analysis of nonsynonymous single-nucleotide variations in active sites of human proteins. *FEBS J*. 2013;280(6):1542–62.
- Flossmann E, Schulz UG, Rothwell PM. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke*. 2004;35(1):212–27.
- Freichel M, Suh SH, Pfeifer A, Schweig U, Trost C, Weissgerber P, Biel M, Philipp S, Freise D, Droogmans G, Hofmann F, Flockerzi V, Nilius B. Lack of an endothelial store-operated Ca^{2+} current impairs agonist-dependent vasorelaxation in TRP4^{-/-} mice. *Nat Cell Biol*. 2001;3(2):121–7.
- Genomes Project, C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56–65.
- Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Franco S, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Huffman MD, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Magid D, Marcus GM, Marelli A, Matchar DB, McGuire DK, Mohler ER, Moy CS, Mussolino ME, Nichol G, Paynter NP, Schreiner PJ, Sorlie PD, Stein J, Turan TN, Virani SS, Wong ND, Woo D, Turner MB,

- C. American Heart Association Statistics and S. Stroke Statistics. Heart disease and stroke statistics – 2013 update: a report from the American Heart Association. *Circulation*. 2013;127(1):e6–245.
- Grada A, Weinbrecht K. Next-generation sequencing: methodology and application. *J Invest Dermatol*. 2013;133(8), e11.
- Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, Bjarnadottir SM, Einarsson OB, Gudjonsdottir HM, Hawkins M, Gudmundsson G, Gudmundsdottir H, Andrason H, Gudmundsdottir AS, Sigurdardottir M, Chou TT, Nahmias J, Goss S, Sveinbjornsdottir S, Valdimarsson EM, Jakobsson F, Agnarsson U, Gudnason V, Thorgeirsson G, Fingerle J, Gurney M, Gudbjartsson D, Frigge ML, Kong A, Stefansson K, Gulcher JR. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet*. 2003;35(2):131–8.
- Guo L, Zhou X, Guo X, Zhang X, Sun Y. Association of interleukin-33 gene single nucleotide polymorphisms with ischemic stroke in north Chinese population. *BMC Med Genet*. 2013;14:109.
- Hatakeyama K, Ohshima K, Fukuda Y, Ogura S, Terashima M, Yamaguchi K, Mochizuki T. Identification of a novel protein isoform derived from cancer-related splicing variants using combined analysis of transcriptome and proteome. *Proteomics*. 2011;11(11):2275–82.
- Havmoller R, Chugh SS. Plasma biomarkers for prediction of sudden cardiac death: another piece of the risk stratification puzzle? *Circ Arrhythm Electrophysiol*. 2012;5(1):237–43.
- Hiyama TY, Watanabe E, Ono K, Inenaga K, Tamkun MM, Yoshida S, Noda M. Na(x) channel involved in CNS sodium-level sensing. *Nat Neurosci*. 2002;5(6):511–2.
- Hiyama TY, Matsuda S, Fujikawa A, Matsumoto M, Watanabe E, Kajiwarra H, Niimura F, Noda M. Autoimmunity to the sodium-level sensor in the brain causes essential hypernatremia. *Neuron*. 2010;66(4):508–22.
- International HapMap, C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z,

- Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449(7164):851–61.
- C. International HapMap, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghorri MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52–8.
- Jagger JH, Porter VA, Lederer WJ, Nelson MT. Calcium sparks in smooth muscle. *Am J Physiol Cell Physiol*. 2000;278(2):C235–56.
- Jneid H, Alam M, Virani SS, Bozkurt B. Redefining myocardial infarction: what is new in the ESC/ACCF/AHA/WHF Third Universal Definition of myocardial infarction? *Methodist Debaque Cardiovasc J*. 2013;9(3):169–72.
- Jung C, Gene GG, Tomas M, Plata C, Selent J, Pastor M, Fandos C, Senti M, Lucas G, Elosua R, Valverde MA. A gain-of-function SNP in TRPC4 cation channel protects against myocardial infarction. *Cardiovasc Res*. 2011;91(3):465–71.
- Karagiannis K, Simonyan V, Mazumder R. SNVDis: a proteome-wide analysis service for evaluating nsSNVs in protein functional sites and pathways. *Genom Proteome Bioinforma*. 2013;11(2):122–6.
- Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. *Lancet*. 2005;365(9455):217–23.
- Kubo M, Hata J, Ninomiya T, Matsuda K, Yonemoto K, Nakano T, Matsushita T, Yamazaki K, Ohnishi Y, Saito S, Kitazono T, Ibayashi S, Sueishi K, Iida M, Nakamura Y, Kiyohara Y. A nonsynonymous SNP in PRKCT (protein kinase C eta) increases the risk of cerebral infarction. *Nat Genet*. 2007;39(2):212–7.
- Lawes CM, Vander Hoorn S, Rodgers A, H. International Society of. Global burden of blood-pressure-related disease, 2001. *Lancet*. 2008;371(9623):1513–8.
- Lee JS, Hong JM, Moon GJ, Lee PH, Ahn YH, Bang OY, S. collaborators. A long-term follow-up study of intravenous autologous mesenchymal stem cell transplantation in patients with ischemic stroke. *Stem Cells*. 2010;28(6):1099–106.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, S. Genome Project Data Processing. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009a;25(16):2078–9.
- Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009b;25(15):1966–7.
- Ligers A, Teleshova N, Masterman T, Huang WX, Hillert J. CTLA-4 gene expression is influenced by promoter and exon 1 polymorphisms. *Genes Immunol*. 2001;2(3):145–52.
- Lima B, Forrester MT, Hess DT, Stamler JS. S-nitrosylation in cardiovascular signaling. *Circ Res*. 2010;106(4):633–46.
- Marrocco C, Rinalducci S, Mohamadkhani A, D'Amici GM, Zolla L. Plasma gelsolin protein: a candidate biomarker for hepatitis B-associated liver cirrhosis identified by proteomic approach. *Blood Transfus*. 2010;8 Suppl 3:s105–12.

- Medicine, U. S. N. L. o. PubMed Health. Hypertension (High Blood Pressure), 2015, from <http://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0024199/>
- Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet.* 2010;11(1):31–46.
- Modlinger RS, Nicolis GL, Krakoff LR, Gabrilove JL. Some observations on the pathogenesis of Bartter's syndrome. *N Engl J Med.* 1973;289(19):1022–4.
- Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Liu S, Mackey RH, Matchar DB, McGuire DK, Mohler 3rd ER, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner MB, on behalf of the American Heart and A. S. C. a. S. S. Subcommittee. Heart disease and stroke statistics – 2015 update: a report from the American Heart Association. *Circulation.* 2015;131:e29–e322.
- National Research Council (US). Committee on intellectual property rights in genomic and protein research and innovation. Merrill SA, Mazza AM, editors. Reaping the benefits of genomic and proteomic research: Intellectual property rights, innovation, and public health. Washington (DC): National Academies Press (US); 2006. 2, Genomics, Proteomics, and the Changing Research Environment. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK19861/>
- National Heart Lung and Blood Institute T. Who is at risk for sudden Cardiac Arrest? 2011. 2015, from <http://www.nhlbi.nih.gov/health/health-topics/topics/scda/atrisk> (1 Apr 2011).
- National Heart Lung and Blood Institute T. What is a heart attack? 2013. 2015, from <http://www.nhlbi.nih.gov/health/health-topics/topics/heartattack> (17 Dec 2013).
- National Heart Lung and Blood Institute T. What is coronary heart disease? 2014. 2015, from <http://www.nhlbi.nih.gov/health/health-topics/topics/cad> (29 Sept 2014).
- Nie S, Yin H, Tan Z, Anderson MA, Ruffin MT, Simeone DM, Lubman DM. Quantitative analysis of single amino acid variant peptides associated with pancreatic cancer in serum by an isobaric labeling quantitative method. *J Proteome Res.* 2014;13(12):6058–66.
- Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS, Kapp EA, Moritz RL, Chan DW, Rai AJ, Admon A, Aebersold R, Eng J, Hancock WS, Hefta SA, Meyer H, Paik YK, Yoo JS, Ping P, Pounds J, Adkins J, Qian X, Wang R, Wasinger V, Wu CY, Zhao X, Zeng R, Archakov A, Tsugita A, Beer I, Pandey A, Pisano M, Andrews P, Tammen H, Speicher DW, Hanash SM. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics.* 2005;5(13):3226–45.
- Park DS, Fishman GI. Nav-igating through a complex landscape: SCN10A and cardiac conduction. *J Clin Invest.* 2014;124(4):1460–2.
- Plant LD, Bowers PN, Liu Q, Morgan T, Zhang T, State MW, Chen W, Kittles RA, Goldstein SA. A common cardiac sodium channel variant associated with sudden infant death in African Americans, SCN5A S1103Y. *J Clin Invest.* 2006;116(2):430–5.
- Ritchie MD, Denny JC, Zuvich RL, Crawford DC, Schildcrout JS, Bastarache L, Ramirez AH, Mosley JD, Pulley JM, Basford MA, Bradford Y, Rasmussen LV, Pathak J, Chute CG, Kullo IJ, McCarty CA, Chisholm RL, Kho AN, Carlson CS, Larson EB, Jarvik GP, Sotoodehnia N, H. Cohorts for, Q. R. S. G. Aging Research in Genomic Epidemiology, Manolio TA, Li R, Masys DR, Haines JL, Roden DM. Genome- and phenome-wide analyses of cardiac conduction identifies markers of arrhythmia risk. *Circulation.* 2013;127(13):1377–85.
- Royer-Bertrand B, Rivolta C. Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer. *Cell Mol Life Sci.* 2015;72(8):1463–71.
- Ruppert V, Meyer T, Struwe C, Petersen J, Perrot A, Posch MG, Ozcelik C, Richter A, Maisch B, Pankuweit S, N. German Heart Failure. Evidence for CTLA4 as a susceptibility gene for dilated cardiomyopathy. *Eur J Hum Genet.* 2010;18(6):694–9.

- Sahoo S, Losordo DW. Exosomes and cardiac repair after myocardial infarction. *Circ Res*. 2014;114(2):333–44.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463–7.
- Santana-Quintero L, Dingerdissen H, Thierry-Mieg J, Mazumder R, Simonyan V. HIVE-hexagon: high-performance, parallelized sequence alignment for next-generation sequencing data analysis. *PLoS ONE*. 2014;9(6), e99033.
- Schlingmann KP, Konrad M, Jeck N, Waldegger P, Reinalter SC, Holder M, Seyberth HW, Waldegger S. Salt wasting and deafness resulting from mutations in two chloride channels. *N Engl J Med*. 2004;350(13):1314–9.
- Schuler GD, Altschul SF, Lipman DJ. A workbench for multiple alignment construction and analysis. *Proteins*. 1991;9(3):180–90.
- Serizawa M, Nabika T, Ochiai Y, Takahashi K, Yamaguchi S, Makaya M, Kobayashi S, Kato N. Association between PRKCH gene polymorphisms and subcortical silent brain infarction. *Atherosclerosis*. 2008;199(2):340–5.
- Sherry ST, Ward M, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308–11.
- Shimizu H, Watanabe E, Hiyama TY, Nagakura A, Fujikawa A, Okado H, Yanagawa Y, Obata K, Noda M. Glial Nax channels control lactate signaling to neurons for brain [Na⁺] sensing. *Neuron*. 2007;54(1):59–72.
- Simonsen TS, Zhang Y, Huff CD, Xing J, Watkins WS, Witherspoon DJ, Woodward SR, Jorde LB. Limited distribution of a cardiomyopathy-associated variant in India. *Ann Hum Genet*. 2010;74(2):184–8.
- Simonyan V, Mazumder R. High-Performance Integrated Virtual Environment (HIVE) tools and applications for big data analysis. *Genes (Basel)*. 2014;5(4):957–81.
- Sisakian H. Cardiomyopathies: evolution of pathogenesis concepts and potential for new therapies. *World J Cardiol*. 2014;6(6):478–94.
- Spławski I, Timothy KW, Tateyama M, Clancy CE, Malhotra A, Beggs AH, Cappuccio FP, Sagnella GA, Kass RS, Keating MT. Variant of SCN5A sodium channel implicated in risk of cardiac arrhythmia. *Science*. 2002;297(5585):1333–6.
- Tirupathi C, Freichel M, Vogel SM, Paria BC, Mehta D, Flockerzi V, Malik AB. Impairment of store-operated Ca²⁺ entry in TRPC4(-/-) mice interferes with increase in lung microvascular permeability. *Circ Res*. 2002;91(1):70–6.
- Tobin MD, Raleigh SM, Newhouse S, Braund P, Bodycote C, Ogleby J, Cross D, Gracey J, Hayes S, Smith T, Ridge C, Caulfield M, Sheehan NA, Munroe PB, Burton PR, Samani NJ. Association of WNK1 gene polymorphisms and haplotypes with ambulatory blood pressure in the general population. *Circulation*. 2005;112(22):3423–9.
- Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105–11.
- Van Norstrand DW, Tester DJ, Ackerman MJ. Overrepresentation of the proarrhythmic, sudden death predisposing sodium channel polymorphism S1103Y in a population-based cohort of African-American sudden infant death syndrome. *Heart Rhythm*. 2008;5(5):712–5.
- Wang Q. Molecular genetics of coronary artery disease. *Curr Opin Cardiol*. 2005;20(3):182–8.
- Wang Y, Zheng Y, Zhang W, Yu H, Lou K, Zhang Y, Qin Q, Zhao B, Yang Y, Hui R. Polymorphisms of KDR gene are associated with coronary heart disease. *J Am Coll Cardiol*. 2007;50(8):760–7.
- Wang Y, Yang Q, Wang Z. The evolution of nanopore sequencing. *Front Genet*. 2014;5:449.
- Watanabe E, Fujikawa A, Matsunaga H, Yasoshima Y, Sako N, Yamamoto T, Saegusa C, Noda M. Nav2/NaG channel is involved in control of salt-intake behavior in the CNS. *J Neurosci*. 2000;20(20):7743–51.
- Wu L, Shen Y, Liu X, Ma X, Xi B, Mi J, Lindpaintner K, Tan X, Wang X. The 1425G/A SNP in PRKCH is associated with ischemic stroke and cerebral hemorrhage in a Chinese population. *Stroke*. 2009;40(9):2973–6.

- Xi B, Liang Y, Reilly KH, Wang Q, Hu Y, Tang W. Trends in prevalence, awareness, treatment, and control of hypertension among Chinese adults 1991–2009. *Int J Cardiol.* 2012;158(2):326–9.
- Xi B, Liang Y, Mi J. Hypertension trends in Chinese children in the national surveys, 1993–2009. *Int J Cardiol.* 2013;165(3):577–9.
- Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett.* 2013;340(2):284–95.
- Xue Y, Liu Z, Gao X, Jin C, Wen L, Yao X, Ren J. GPS-SNO: computational prediction of protein S-nitrosylation sites with a modified GPS algorithm. *PLoS ONE.* 2010;5(6), e11290.
- Yip H, Chan WY, Leung PC, Kwan HY, Liu C, Huang Y, Michel V, Yew DT, Yao X. Expression of TRPC homologs in endothelial cells and smooth muscle layers of human arteries. *Histochem Cell Biol.* 2004;122(6):553–61.
- Zhang S, Remillard CV, Fantozzi I, Yuan JX. ATP-induced mitogenesis is mediated by cyclic AMP response element-binding protein-enhanced TRPC4 expression and activity in human pulmonary artery smooth muscle cells. *Am J Physiol Cell Physiol.* 2004;287(5):C1192–201.
- Zhang J, Dong X, Hacker TA, Ge Y. Deciphering modifications in swine cardiac troponin I by top-down high-resolution tandem mass spectrometry. *J Am Soc Mass Spectrom.* 2010;21(6):940–8.
- Zhang B, Li M, Wang L, Li C, Lou Y, Liu J, Liu Y, Wang Z, Wen S. The association between the polymorphisms in a sodium channel gene SCN7A and essential hypertension: a case-control study in the Northern Han Chinese. *Ann Hum Genet.* 2015;79(1):28–36.