Design of a common pathway drug for all types of cardiovascular diseases: A network biology approach

Al-Mustanjid Arif1, Chandan Mandal1, Md. Habibur Rahman1,2
1Department of Software Engineering, Daffodil International University (DIU), 102/1, Sukrabad, Dhanmondi, Dhaka-1207, Bangladesh
2Department of Information and Communication Technology, Mawlana Bhasani Science and Technology University (MBSTU), Santosh, Tangail-1902, Bangladesh
E-mail: mustanjid.se@gmail.com, chandan.swe@gmail.com, mdhabibur.r.bd@ieee.org

Received 26 March 2019; Accepted 10 April 2019; Published 1 June 2019

Abstract
More than an era among all non-communicable diseases, cardiovascular diseases have become a major concern worldwide. Cardiovascular diseases have occurred around the world because of some common risk factors. Diseases have a genetic association indirectly or directly resulting from similar risk factors. A disease is caused when a gene misses out its normal activity and affects the body negatively. Several research works have revealed the ways of how a structure-based drug from key biomolecule or protein can be designed for diseases using modern bioinformatics techniques and tools in network biology. This study evaluates protein-protein interaction network and designs a common pathway drug for all types of cardiovascular diseases. The data mining application called knowledge discovery in database (KDD) has been applied and genes are filtered, pre-processed, transformed and mined to identify common cardiovascular disease genes. Cardiovascular disease genes are collected using R from the National Center for Biotechnology Information gene database. UniHi is used as a tool for achieving the goal.

Keywords cardiovascular disease; data mining; R; protein-protein interaction network; drug design.

1 Introduction
Bioinformatics is the solicitation of information technology that analyzes living data and addresses biological problems. Integration of tools, databases and methods are endeavoring to solve biological problems (Zhang, 2016b). The invention and absorption of new drugs are the main concern in pharmaceutical and biomedical research (Rask-Andersen et al., 2011). Genomics, proteomics as well as drug design are the united working areas in bioinformatics to lead drug discovery (Anderson, 2003; Burbaum and Tobal, 2002; Iqbal et al., 2014). At present, network biology figures out new common pathways or interrelation mechanisms among various diseases affected by disease genes based on their sub-networks (Zhang, 2011, 2016a-b, 2018; Zhang and...
A common pathway drug reduced the dimension of drug absorption for diseases (Habib, 2017). Cardiovascular disease caused 17.9 million people to die each year (World Health Organization, 2013). It turns up to 31% of the total deaths globally. Premature deaths are caused by CVDs worldwide extremely (Nichols et. al., 2016). Further, CVDs are increasing in low-middle income countries due to insufficient health care (Cardiovascular, 2017). In addition, CVDs is quitting more living people than cancer and other chronic lower respiratory diseases all over the earth (American Heart Association (AHA), 2017; Mendis et al., 2011). CVDs is categorized into 5 types (Mendis et al., 2011). These types are Ischecmic heart disease (IHD), Cerebrovascular disease (CD), Hypertensive heart disease (HHD), Inflammatory heart disease, Rheumatic heart disease (RHD).

Atherosclerosis is one of the main cause of CVDs and leading to diabetes directly or indirectly (American Diabetes Association). It is a blockage or plaque in arteries that make the arteries thick and stiff which stops blood supplying. IHD is constructed due to atherosclerosis and ultimately leads to myocardial infarction, stroke, peripheral arterial disease or heart attack. It is the worst killer among all types. IHD is also called as coronary heart disease or coronary artery disease. Nichols et al., 2016 represented close to 2000 deaths were responsible for IHD in Australia. An American will be faced heart attack in every 40 seconds (AHA, 2017). About 1.8 million people died of coronary heart disease across Europe (Dégano et al., 2015).

The CD is raised from deep vein thrombosis and atherosclerosis. Stroke is a prime disorder of all CD that prevents blood flow within the brain. Stroke is second and the cause of 11.8% of deaths among all CVDs (AHA, 2017). Nichols et al. (2014) indicated that 10% of men and 15% of women in Europe were killed by stroke in 2014. Stroke incidents have been found abundantly for men and women in Kazakhstan on the continent of Asia (Thrift et al., 2017).

HHD is the leading reason for death associated with high blood pressure. Tightness or pressure in the chest, fatigue, pain in the neck, back, arms, or shoulders, loss of appetite, leg or ankle swelling are the symptoms of HHD. In sub-Saharan Africa, the number of HHD deaths increased between 1990 and 2013 from 37525 to 86035 (Mensah et al., 2015).

Inflammatory heart disease is also known as myocarditis. The muscle layer of the heart wall known as myocardium is infected by virus, bacteria, fungi etc. After infection, it becomes inflamed and damages the effectiveness of the blood pump which ultimately leads to stroke, heart failure and heart attack. In England, 11824 patients were admitted with acute myocarditis from 1998 to 2016 (Lota et al., 2018). Myocarditis is one of the main reasons for around 62% of heart failure (Vos et al., 2015) and total global myocarditis outbreak was 22 out of 100 000 patients per year (Fung et al., 2016).

RHD is caused by rheumatic fever. Phillips et al. (2016) suggested that RHD is one of the major contributors to heart failure movement. It has become the largest burden of CVDs among youth due to RHD’s negligence in developing countries. Currently, RHD has caused 250,000 deaths worldwide annually (Marijon et al., 2012).

The present research focused on the analysis of the genetic relationship between all CVDs types. The liable genes for CVDs are recovered from the gene database of the National Center for Biotechnology Information (NCBI) using R. Regulatory interactions are identified in the network of protein-protein interaction (PPI) which leads to finding common pathways shared by common genes. Thus, a common drug is designed for all types of CVDs.

This article is categorized into 5 sections. Section 2 describes about background and previously related works. Section 3 discusses about the proposed methodology and further working procedures, section 4 describes and analyzes the results and lastly section 5 constructs the conclusion and future resolution.
2 Background and Related Work

Realization of risk factors can provide a high degree of etiology, prevention and treatment in CVDs (Thayer et al., 2010). The article by Thayer et al., 2010 identified frequently six mostly responsible risk factors for occurring CVDs. As risk factors are correlated, genetic interrelationships may be available including common pathways between diseases. High blood pressure (HBP) or hypertension, high cholesterol, smoking, physical inactivity or obesity and diabetes are the most common risk factors for CVDs (Thayer et al., 2010; Mendis et al., 2011; Berry et al., 2012; Koene et al., 2016).

Day after day, unnatural blood pressure and hypertension among humans lead to CVDs. Growth of CVDs is at high risk due to HBP (Lackland et al., 2015; Rahimi et al., 2015). Also in 2015, Lackland et al. considered that 1.56 billion people globally have hypertensive characteristics. Hypertension is the main cause of myocardial infarction, renal failure and death (James et al., 2014). Research studies have suggested that low blood pressure (< 130 mm Hg) and blood pressure mitigation treatment helps to reduce the risk of CVDs (Ettehad et al., 2016; Williamson et al., 2016).

Cholesterol makes the new cell and generates hormone in our body that is needed to run our lives. While the number of cholesterol in the blood is increasing, atherosclerosis is formed by plaques or walls. The existence of different cholesterols has also been found to increase CVDs (Ray et al., 2017; Varbo et al., 2019).

Atherosclerosis is mostly caused by smoking and toxins in tobacco that inaugurate CVDs are also difficult to spot out (Messner and Bernhard, 2014). Mons et al. (2015) came to the conclusion that smoking is a strong sovereign risk factor for CVDs and smoking remission was highly beneficial to the elderly people.

Obesity arrives due to extra fat consumption as well as overweight of an individual. It biologically and ecologically ruins public health and increases the chances of CVDs (Nigro et al., 2014). Obesity pushes much more badly toward CVDs year after year (Lavie et al., 2016). Due to consuming junk foods vastly, teenagers are getting obese in Bangladesh (Salahuddin, 2018). Physical inactivity is a major malefactor for CVDs and study found that free time physical activity minimizes CVDs risk (Li and Siegrist, 2012).

Diabetes appears while blood sugar levels are high. The existence of diabetes in the human body gradually activates CVDs by hampering the vessels of the heart. Diabetes spreads more frequently among adults in the United States (Menke et al., 2015; Mendola et al., 2018). Ogurtsova et al. (2017) articulated that just by 2040, the presence of diabetes may rise to 10.4 %. Type 2 diabetes is a potent CVDs risk factor. Risk factors such as hypertension, cessation of smoking, weight loss must be controlled to reduce the likelihood of CVDs with type 2 diabetes (Martín-Timón et al., 2014).

Mostly with the support of bioinformatics tools and resources, multiple research works researched genetically associated diseases and even their own common pathways procured from the PPI network. Habib (2016) inspected correlation among bipolar disorder and its associated diseases. Data mining application was applied by the authors and 17 common interrelated genes were found using R. PPI network generated from common genes and found a common pathway that integrates the diseases to one another. The accountable genes gathered from the NCBI gene database and several PPI network generating tools were used to reach the goal. Moreover, following the common pathway, Habib (2017) formed a common structure-based drug. The conclusion was drawn from the UniHi tool. Further, the PPI networks and common pathways identified in the article of Akter (2018) for vector transmitted diseases.

It is clear from the above discussion that types of CVDs may interrelate with each other. There might also be a common path among them. For these reasons, the analysis and continuation of the inquiry necessitate common genes PPI network. This article explores the genetic connection retained from the types of CVDs and the PPI network. Finally, common pathways in the PPI network are pinpointed and a common CVDs drug is designed.
3 Materials and Methods
The design of drug is a sequential procedure. KDD steps are taken to achieve the desired objective. The following subsections from 3.1 to 3.8 are serially carried out to accomplish the goal.

3.1 Gene search and collection
All divergent biological data are stored by NCBI. “rEntrez” is a package in R which is accessed in various NCBI databases including gene. Utilize “rEntrez” package in R to search and collect directly responsible gene ids for all types of CVDs from the gene database.

3.2 Gene pre-processing and filtering
Clearing noise from obtained data is called preprocessing. Only human genes are essential. By inserting Homo sapiens to the search query using “rEntrez” again, previously collected data is filtered and only human genes are released. Then all gene ids for all types of CVDs are stored as vectors.

3.3 Cross-linked gene collection
The associated diseases are recognized and a total of 26 cross combinations among all types of CVDs have been measured. The interlinked genes are discovered in the search term using the "AND" alternative in combination with the “rEntrez” package. Genes are assembled in combinations of 2 types, 3 types, 4 types and 5 types as well as reserved in vectors.

3.4 Gene sorting
A gene consists of a lot of information like gene ids, aliases, gene symbols, etc. To sort and scrap only gene symbols (actually protein) from human and linkage gene ids, an open source software called "Bioconductor" with “Annotation” package is incorporated with R.

3.5 Gene mining
A large amount of data can commit a critical outcome and any error can spoil the entire output. Read vectors of all types of CVDs and cross-link gene symbols in R and implicated "reduce" function to distinguish common genes. Common mined genes are deposited in a database.

3.6 Verify and find common gene
The mined genes are checked by the Expasy database. Finally, there were 14 common validated genes discovered. The 14 verified common genes are namely “TP53”, “TNF”, “IL6”, “MTHFR”, “TGFB1”, “ACE”, “MMP9”, “CRP”, “TLR4”, “NPPB”, “HMOX1”, “AGTR1”, “MMP1” and “F3”.

3.7 Generate PPI network
UniHi or Unified Human Interactive is a trusted PPI visualization tool managed by SysBioLab in Bioinformatics. Common genes PPI networks are produced using UniHi tool even including common pathways.

3.8 Drug design
The desired goal of this research is to design a common pathway drug for all types of CVDs. A drug is an element used to treat, cure and prevent disease, to relieve pain or to tone down certain appointed processes in the body for a particular cure. The therapeutic reaction to the invention of the drug is called the root (Habib, 2017). Feng et al. (2017) argued that research into biomolecule proteins is the inventive process of finding new drugs. The drug should design in a way that only affects dysfunctional proteins or prohibits the protein to regulate with another protein and keep the supply of the normal chemical process in the body. Specifically targeted genes or proteins are required in order to establish a new drug for the disease. A fairly rapid and revolutionary improvement in proteomics, genomics and molecular biology has profound effects on the discovery of drugs (Blundell et al., 2002). Feng et al. (2017) also found that drug targets can be discovered by observing the PPI network from non-drug proteins. Identification of targets, drug discovery and design is guided by protein structure (Blundell, et al., 2006). The invention, display and assessment of PPI is therefore a
prerequisite for the design of drugs. UniHi illustrates the molecular mechanism of diseases and redefines the mark point of drug design structurally based on biomolecular pathways and networks of interaction (Kalathur et al., 2013; Fotis et al., 2018).

4 Results and Discussion
4.1 Gene search and collection
About 17000 primarily responsible genes were accounted for all types of CVDs. The genes were collected for all organisms.

![Fig. 1 Primary and filtered genes for all types of CVDs.](image)

<table>
<thead>
<tr>
<th>Linkage among 2 types</th>
<th>Gene Amount</th>
<th>Gene</th>
<th>Linkage among 3 types</th>
<th>Gene Amount</th>
<th>Gene</th>
<th>Linkage among 4 types</th>
<th>Gene Amount</th>
<th>Gene</th>
<th>Linkage among 5 types</th>
<th>Gene Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>I, C</td>
<td>93</td>
<td>I, C, H</td>
<td>46</td>
<td>I, C, H, In</td>
<td>44</td>
<td>I, C, H, In, R</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I, H</td>
<td>153</td>
<td>I, C, In</td>
<td>72</td>
<td>I, C, H, R</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C, H</td>
<td>51</td>
<td>I, H, In</td>
<td>135</td>
<td>I, C, In, R</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I, In</td>
<td>477</td>
<td>C, H, In</td>
<td>14</td>
<td>I, H, In, R</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C, In</td>
<td>97</td>
<td>I, C, R</td>
<td>16</td>
<td>C, H, In, R</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H, In</td>
<td>192</td>
<td>L, H, R</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I, R</td>
<td>47</td>
<td>C, H, R</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C, R</td>
<td>16</td>
<td>L, In, R</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H, R</td>
<td>31</td>
<td>C, In, R</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In, R</td>
<td>81</td>
<td>H, In, R</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Sub-total             | 1238        | 414   | 113       | 14          |
| Total                 | 1779        |       |           |             |

<table>
<thead>
<tr>
<th>Ischemic</th>
<th>Cerebrovascular</th>
<th>Hypertensive</th>
<th>Inflammatory</th>
<th>Rheumatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>C</td>
<td>H</td>
<td>In</td>
<td>R</td>
</tr>
</tbody>
</table>
4.2 Gene pre-processing, filtering and sorting
Only human genes are filtered and pre-processed from primary genes and counted 764 as for IHD, 128 for CD, 2476 for Inflammatory heart disease, 263 for HHD, 100 for RHD respectively. Gene symbols are picked out for preprocessed and filtered genes. Differences are shown in Fig. 1 between primary and filtered genes.

4.3 Cross-linkage gene collection
Total 26 associations have been found among all CVDs types. The counting of cross-link genes is shown in Table 1.

4.4 Gene mining
After mining, 14 common genes found across all types of CVDs. The 14 common genes are TP53, TNF, IL6, MTHFR, TGFβ1, ACE, MMP9, CRP, TLR4, NPPB, HMOX1, AGTR1, MMP1 and F3. Expasy database used to certify these common genes.

4.5 Network of PPI including regulatory interaction
The UniHi tool is used for determining the PPI network between common genes and additional experiments. Uploading common genes to UniHi, PPI networks are established and these networks reveal the protein interactions of indirectly or directly related genes. Fig. 2 shows the PPI network among 14 common genes but Fig. 3 indicated only direct regulatory interacted genes. Fig. 3 also clearly represents that only 6 genes in two clusters have common pathway encounters. The first cluster has four genes: TP53, MMP1, TNF, TGFβ1 and the second cluster has two genes: MMP9 and IL6. Again, Fig. 4 and Fig. 5 visualize PPI networks among 4
common genes and between 2 common genes consecutively. Regulatory interactions in both Fig. 6 and Fig. 7 illustrate relatively straight aligned proteins sequentially with culpable disease genes.

**Fig. 3** Regulatory interaction among 6 common genes in two clusters.

### 4.6 Drug design

The conclusive goal of this study is to design common pathway drug for all type of CVDs. It is ineffective to attain successful disease medications from target acquisition. A drug must dominate the target proteins so that it does not coincide with normal impacts. There are several bioinformatics tools invented to accomplish protein activity and UniHi is one of them and very popular (Kohl et al., 2011). Common pathway drug is designed with UniHi tool for examined diseases. The drug design mapped for the ultimately responsible 6 common genes are shown in Fig. 8 and Fig. 9. The effected proteins are mapped through red color in both networks and yellow color proteins may interact in other contexts. The target proteins mapped in red color also have a clear and direct compound with the target genes.
Fig. 4 PPI network among 4 common genes in first cluster.

Fig. 5 PPI network between 2 common genes in second cluster.
A drug must attach to a specific point on aberrant protein or nucleotide to control the disease. This article must acknowledge and study the major composite that interferes directly against a disease. By using UniHi filtering strategies, inseparably linked proteins directly associated with the diseases under inquiry are properly identified and viewed in Fig. 10 and Fig. 11. The common structure-based drug is thus designed for all types of CVDs.
Fig. 8 Mapped drug target among 4 common genes in first cluster.

Fig. 9 Mapped drug target between 2 common genes in second cluster.
5 Conclusions
Advancing bioinformatics tools and databases reveal new areas of biological research and tasks are becoming easier to do. The integration of data mining and bioinformatics makes the design of computer-aided drugs
more effective. It is very important to identify disease affected genes or targets for designing a drug. Drug development is being accelerated by target recognition and designing a drug.

Common risk factors caused all types of CVDs. KDD the data mining application is performed to retrieve common genes from NCBI gene dataset. To design a common drug, PPI network is generated from common genes and interconnected genes are identified based on regulatory interactions. Finally, performing various actions a structure-based common pathway drug is designed that will heal all types of CVDs. This study's future resolution is to work on different other correlated diseases to design a common drug.

**Abbreviations**

Cardiovascular Disease = CVDs, Ischemic Heart Disease = IHD, Cerebrovascular disease = CD, Hypertensive heart disease = HHD, Rheumatic heart disease = RHD, Protein-protein interaction = PPI, NCBI = National center for biotechnology information

**References**

Akter T. 2018. Investigate to find common gene and design a PPI network for vector borne diseases(Malaria, Dengue and Chikungunya)-A bioinformatics approach. Network Biology, 8(3): 113-125


Anderson AC. 2003. The process of structure-based drug design. Chemistry and Biology, 10(9): 787-797


participant data from prospective cohort studies of the CHANCES consortium. BMJ, 350: h1551
Varbo A, Nordestgaard BG. 2019. Commentary: Triglycerides or HDL cholesterol in cardiovascular disease—which is the true culprit? International Journal of Epidemiology, dy292